Palacký University Olomouc

Faculty of Science

Department of Geoinformatics

THE EVALUATION OF CARTOGRAPHIC VISUALIZATION METHODS AND INTERACTIVE MAP INTERFACES THROUGH EYE-TRACKING TECHNOLOGY

Habilitation thesis

Stanislav POPELKA

Olomouc 2024

Geoinformatics and cartography

I declare that I have prepared the submitted habilitation thesis independently using properly cited literary sources.

In Olomouc on 12. 11. 2024

RNDr. Stanislav Popelka, Ph.D.

I am deeply grateful to my parents for their unwavering support throughout my life. Their decision to enroll me in an eight-year gymnasium was the starting point of a long journey that ultimately led me here.

I am also thankful to Professor Vít Voženílek, whose timely decision to purchase an eye-tracker coincided perfectly with the start of my PhD studies. This choice set the direction of my career toward the fascinating field of eye-tracking, which I find both enriching and fulfilling.

I would also like to thank all my colleagues at the Department of Geoinformatics, whose friendship and good humor make the department a place I truly love to be, turning the duty of work into a pleasure and source of fun.

This thesis would not have been possible without the hard work of my co-authors. Special thanks go to Markéta Beitlová, Alžběta Brychtová, and Jitka Doležalová for their invaluable contributions. I am also grateful to Vilém Pechanec, Jaroslav Burian and Zdeněk Stachoň for their insightful discussions and feedback on the thesis text. And, of course, I thank ChatGPT for its assistance in polishing this text.

Work isn't everything, so I would also like to thank my friends, who have kept me grounded through it all.

Finally, I would like to express my gratitude to Markéta—not only for her support and our never-ending debates about cartography, but also for everything else she brings into my life.

TABLE OF CONTENTS

MOTIVATION **5**

1 CONCEPT **6**

1.1 Structure of the Habilitation 6

1.2 Goals of the Habilitation 7

1.3 Narrative of the Habilitation 8

2 INTRODUCTION **11**

2.1 Cognitive Cartography 11

2.2 History, Principle, and Use of Eye-tracking 14

2.3 Eye-tracking in Cognitive Cartography 19

3 METHODS **24**

3.1 Research Methods in Cognitive Cartography 24

3.2 Experiment Design 25

3.3 Apparatus & Software 27

3.4 Analytical Methods 28

4 RESULTS **32**

4.1 Paper 2D-3D (Popelka & Brychtová, 2013) 🗺️ 33

4.2 Paper ScanGraph (Doležalová & Popelka, 2016) 🧰 35

4.3 Paper EyeTribe (Popelka et al., 2016) 🧰 37

4.4 Paper Glyphs (Opach et al., 2017) 🗺️ 39

4.5 Paper UrbanPlans (Burian et al., 2018) 🗺️ 41

4.6 Paper WeatherMaps (Popelka, Vondrakova, et al., 2019) 🎮 43

4.7 Paper Author-Reader (Beitlova et al., 2022) 🧰 45

4.8 Paper Swipe-Multiple (Popelka et al., 2022) 🎮 47

4.9 Paper GazePlotter (Popelka, Kominek, et al., 2024) 🧰 49

4.10 Paper ET2Spatial (Sultan et al., 2022) 🧰 51

4.11 Paper Dashboards (Porti Suarez & Popelka, 2023) 🎮 53

5 DISCUSSION **55**

6 CONCLUSIONS AND FUTURE WORK **58**

6.1 Contributions to Cartographic Knowledge of Visualization Methods 🗺️ 58

6.2 Insights into Interactivity with Cartographic Interfaces 🎮 59

6.3 Advances in the Methodology of Cartographic Eye-tracking Research 🧰 60

6.4 Summary of Findings 61

REFERENCES **63**

SUPPLEMENTS **74**

MOTIVATION

George Frederick Jenks, one of the most influential cartographic researchers of the 20th century, published the first eye-tracking study in cartography over half a century ago (Jenks, 1973). Jenks described the eye-tracker as a tool that allows researchers to "**get inside the map reader's head."** However, he concluded this pioneering study by comparing the work to opening **Pandora's box**. He initially believed that the components of map reading would easily fall into place, but instead, he uncovered numerous challenges and complexities in thematic cartography. Additional information, often positioned at the edges of a map, tended to distract the reader and potentially distort the map's intended message. The placement of titles, legends, and scales was no longer just about achieving visual balance; it had become critical in shaping how the map was interpreted.

In the past fifty years, cartography has grown more intricate. Modern geographic information systems (GIS) generate complex visualizations, and users are now required to interact with dynamic, often multilayered interfaces. As the complexity of cartographic tools increases, so too does the need to understand how users navigate these interfaces and comprehend the information presented. This research is driven by an urgent need to overcome the challenges associated with both the design of intuitive, user-friendly maps and the exploration of user interactions with these cartographic products.

Since Jenks' time, eye-tracking technology has advanced significantly. No longer is it necessary to use a modified Kodak 8 mm camera to capture respondents' corneas and analyze the data frame by frame. Modern technology offers sophisticated tools for deeper analysis, yet challenges remain. This underscores the need for further methodological advancements in eye-tracking research to develop more precise and effective tools for data collection and analysis.

The ongoing complexity of cartography, particularly in the digital age, requires a deeper understanding of how users process visual information and engage with interactive maps. My research focuses on **analyzing these cognitive processes**, aiming to develop methods that enable cartographers to design maps that are both effective and accessible. By utilizing modern eye-tracking technologies—tools that allow us to "get inside the head of the map reader"—I aim to illuminate how users perceive, interpret, and interact with various cartographic elements, providing insights that can significantly enhance map design.

Through these efforts, I hope this research will bring us a step closer to closing the Pandora's box of complexities that Jenks opened over 50 years ago.

1 CONCEPT

The habilitation thesis "The Evaluation of Cartographic Visualization Methods and Interactive Map Interfaces Through Eye-tracking Technology" consists of **eleven research articles** representing this author's long-term eye-tracking research conducted in the field of cognitive cartography.

Eye-tracking is a unique technology that can help reveal the map-reading process. With its employment, it is possible to evaluate the quality and comprehensibility of maps, identify the strategies used to read, analyze, and interpret maps, and compare how these strategies vary among participants and their groups.

This thesis's author was the first to employ eye-tracking in Czechia in the geographic domain. This eye-tracking research, at the Department of Geoinformatics, Palacký University Olomouc, started in 2011.

After familiarizing himself with eye-tracking technology, the author designed his first eye-tracking experiments, made subsequent modifications, and applied them to the field of cognitive cartography. His PhD thesis (Popelka, 2015) focused on comparing 2D and 3D visualization in GIS using eye-tracking. Besides empirical findings about how users perceive 2D and 3D visualizations, one of the results of this thesis was a set of recommendations for eye-tracking research in cartography. Later, practical recommendations on how to design, prepare, conduct, and analyze eye-tracking experiments were summarized in the book "Eye-tracking nejen v kognitivní kartografii [Eye-tracking (not only) in cognitive cartography]"(Popelka, 2018a), which was the first and remains the only book about eye-tracking in Czech. This book serves as a practical guide on how to utilize eye-tracking in cognitive cartography, but also in other fields of research. In contrast to the non-theoretical topic of the book, this habilitation thesis comprises some of the research the author has conducted on **fundamental topics in cognitive cartography.** It summarizes research findings from the field of cognitive cartography that have been gained using eye-tracking over the years.

1.1 Structure of the Habilitation

The thesis is organized into five key chapters. The **introduction** describes the history of cognitive cartography, and its significance and importance are highlighted. The next sub-chapter is focused on the history and principles of eye-tracking. The last sub-chapter of the introduction explains the role of eye-tracking in cognitive cartography.

The **methods** section briefly introduces the research methods used in cognitive cartography, focusing on eye-tracking and its integration with other research methods. Next, the specifics of experiment design in eye-tracking research are described, and the number of participants used in the experiments is discussed. The apparatus and software that were used are also mentioned. Finally, the methods of analysis for the eye-tracking data used in the thesis are introduced.

After the methods section, in the **results** section, each of the eleven articles included in this habilitation thesis is described. The articles were selected according to their topics, and their order was chosen in such a way that they collectively form a coherent narrative and build upon each other. This sequence illustrates the long-term process of eye-tracking research in cognitive cartography.

The schema of interconnections between papers included in the habilitation thesis is depicted in Figure 1.

The **discussion** chapter synthesizes insights from each study, addressing broader challenges, such as article selection, thematic focus, and methodological constraints related to data collection and the technology used.

The **conclusions and future work** chapter summarizes the thesis's key contributions. This chapter emphasizes the ongoing nature of research in cognitive cartography, noting that these findings and developed tools offer foundational steps for future studies in this evolving field.

1.2 Goals of the Habilitation

With deliberate consideration and a comprehensive understanding of the field's needs, three primary research goals were established:

1. **Evaluation and comparison of cartographic visualization methods** 🗺️

The primary aim of the first goal is to assess the effectiveness and usability of different cartographic visualization techniques. By comparing methods such as 2D versus 3D visualizations, polyline versus star glyphs, and thematic mapping techniques, the research seeks to determine which visualization methods enhance user comprehension and accuracy in interpreting spatial data. The findings contribute to an understanding of the strengths and weaknesses of various cartographic techniques and provide guidance for future visualization design.

2. **Assessment of interactive map interfaces** 🎮

The second goal focuses on evaluating user interactions with modern, interactive map interfaces, particularly in the context of weather forecasts, urban planning, and comparative map designs such as swipe and multiple views. The usability of these interfaces is measured through eye-tracking data, combined with user feedback, to explore how design choices impact task efficiency, accuracy, and overall user experience. Insights gained from this research offer critical recommendations for the design of intuitive, user-friendly interactive map systems.

3. **Design and development of eye-tracking analysis and visualization tools** 🧰

Recognizing the limitations of existing tools, this goal focuses on the creation of new methods for analyzing and visualizing eye-tracking data. These methods allow for more effective analysis of eye movement data, including the comparison of participants' viewing strategies and visualizing the interactions of dynamic stimuli. By enabling the exploration of scanpath similarities, fixation sequences, and participant behavior across both static and interactive cartographic elements, these tools contribute to deeper insights into user interactions with maps and enhance the evaluation of cartographic products. Although primarily developed for use in cartography, these tools are versatile and find application in other research fields as well.

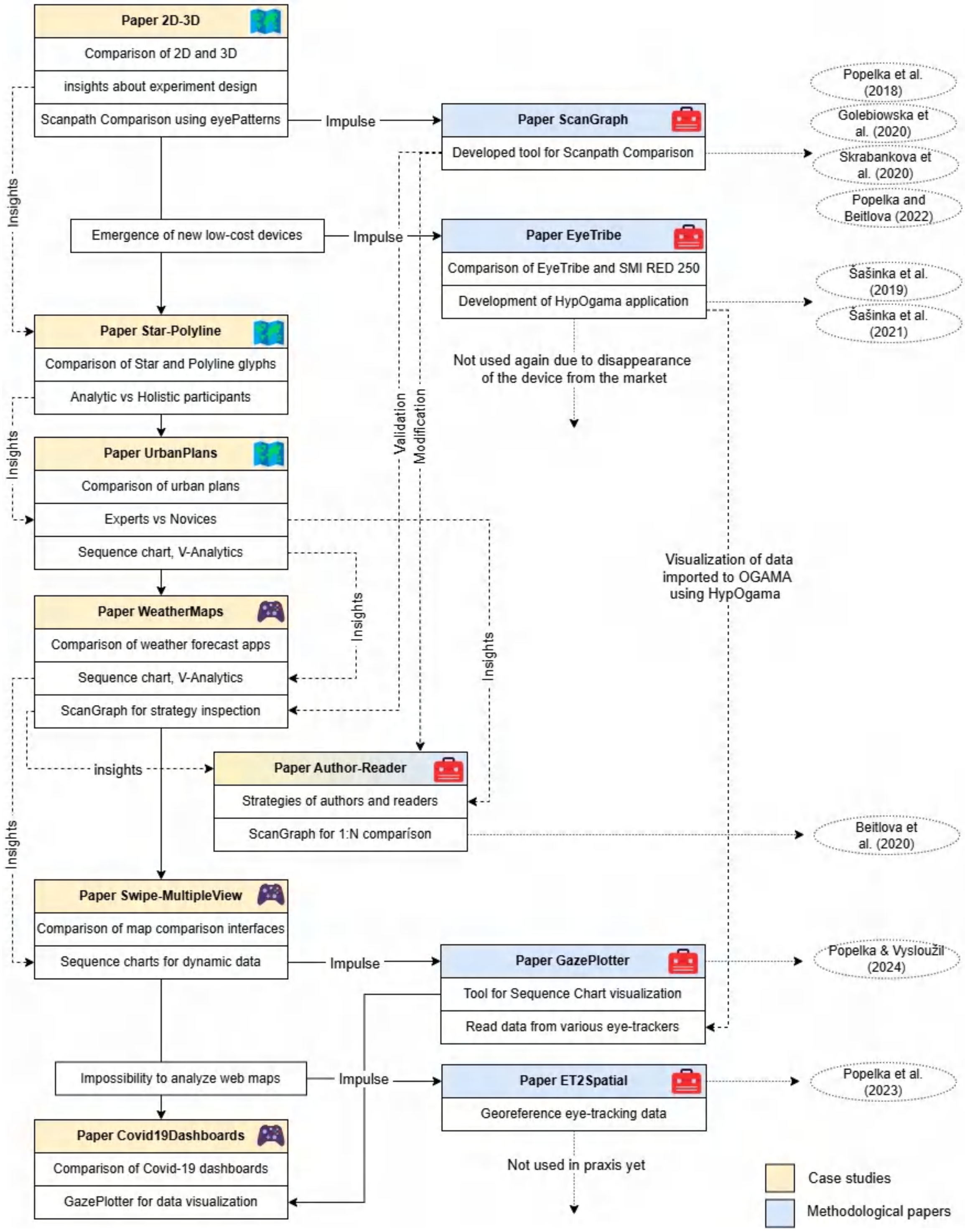


Figure 1. The interconnection of the papers included in the habilitation thesis.

1.3 Narrative of the Habilitation

The first paper included in this habilitation (Popelka & Brychtová, 2013) **[paper 2D-3D]** was written while the author was a graduate student; it focused on analyzing differences in the perception of 2D and 3D terrain visualizations. The original idea was to compare both visualizations side by side. However, it was found that the order of visualization in the stimulus has a greater influence on perception than its content does. Due to this discovery, all subsequent studies were designed as standalone visualizations presented in within-subject designed experiments. The majority of the analyses presented in this paper were quantitative, comparing metrics like dwell time or scanpath length. At the end of the article, an attempt was made to perform a scanpath comparison using the eyePatterns software (West et al., 2006). Although this software was used by recognized scientists (Coltekin et al., 2010), it was later found that the software displayed the similarities improperly, which led to the development of a new tool for scanpath comparison called ScanGraph.

The development of ScanGraph is described in the second paper **[paper ScanGraph]** (Doležalová & Popelka, 2016). This freely available tool employs String-Edit-Distance, where eye-movement trajectories are replaced with sequences of letters representing visited areas of interest. The output of a scanpath comparison is a simple graph, and groups of similar sequences/participants are displayed as cliques of this graph. In the article, the outputs of ScanGraph are illustrated on a cartographic example and compared to the ones from eyePatterns, showing that the solution implemented in ScanGraph is more reliable.

Another methodological paper **[paper EyeTribe]** (Popelka et al., 2016) was a response to the emergence of a new low-cost eye-tracker in 2014. This device, called the EyeTribe, cost a fraction of the price of professional eye-trackers (99 USD). The article aimed to compare the quality of recorded data between the EyeTribe and the SMI RED 250 eye-tracker. The results were promising; however, the EyeTribe company was bought by Oculus, which led to its tracker's disappearance from the market. An outgrowth of the work done for this paper was the development of the HypOgama application, which was later used in several studies (outside of this habilitation).

The next case study, **[paper Glyphs]** (Opach et al., 2017), compared the effectiveness of two types of glyphs – polyline and star glyphs. The design of the experiment was informed by the previous case study. The eye movement data were enhanced by information from a questionnaire addressing the subjective opinions of the participants about glyphs. Moreover, psychological tests were employed to identify the cognitive styles of the participants.

The urban plans of four different cities in the Czech Republic were used as stimulus material in the next case study **[paper UrbanPlans]** (Burian et al., 2018). In this study, the stimuli were partially static (maps) and partially dynamic (scrollable legends). For the visualization of time spent on the maps and on the legends, sequence charts available in SMI BeGaze software were used. In addition to the available tools in the software provided by the eye-tracker vendor, we used the functionality of V-Analytics – a geographic information system developed for spatiotemporal data (Andrienko et al., 2012) to create flow map visualizations of participants' gaze. As participants, experts (employees of urban planning departments) and novices (students) were recruited, and differences between these groups were investigated.

Fully interactive stimuli were used in **[paper WeatherMaps]**, a study comparing different weather forecast map applications (Popelka, Vondrakova, et al., 2019). In contrast to previous case studies, not only was the cartographic visualization evaluated, but also the user interface. Eye movement data recordings were enhanced by the think-aloud method to gain subjective feedback from the participants. Sequence charts and flow map visualizations were used again for data visualization. We used ScanGraph software, developed as an output of **[paper ScanGraph]**, to find similarities in the strategy of stimulus inspection.

Once the usability of the ScanGraph application was demonstrated in praxis, it was then utilized in a more complex way. Beitlova et al. (2022) **[paper Author-Reader]** lies on the borderline between a case study and a methodological paper. The aim of this study was to place map authors into the role of users of their own maps. The functionality of ScanGraph was modified to compare the similarities of eye movement strategies of one participant (the author) versus N participants (map readers).

The study on weather forecast apps **[paper WeatherMaps]** demonstrated that analyzing data recorded from dynamic stimuli is feasible. In the next case study (Popelka et al., 2022) **[paper Swipe-MultipleView]**, the focus was on evaluating two user interfaces for map comparison—swipe and multiple views—within fully interactive environments. During the data recording and interpretation process, emphasis was placed on visualizing the data using sequence charts. BeGaze software was utilized for this purpose; however, visualizing dynamic stimuli proved to be complex and required manual modifications in a graphic editor.

The process of sequence chart visualization from dynamic data was extremely time-consuming, leading to the decision to develop our own tool (Popelka, Kominek, et al., 2024; Vojtechovska & Popelka, n.d.) **[paper GazePlotter]**. The tool functions as a progressive web application that is accessible to all and can read data from various eye-trackers. This makes the process of visualizing sequence charts simple and straightforward.

Interactive maps have increasingly begun to be used as stimuli in the eye-tracking experiments run by the author of this habilitation thesis. Web maps are a special example of an interactive map, whose visible extent is dynamic, based on input by the user (pan, zoom). The tool ET2Spatial (Sultan et al., 2022) **[paper ET2Spatial]** was designed to log user interactions with a map and synchronize them with eye-tracking data. Eye-tracking data recorded over a web map (Google Maps) are placed into geographic coordinates, allowing for their comparison across multiple participants.

The last study included in this habilitation (Porti Suarez & Popelka, 2023) **[paper Dashboards]** focuses on the evaluation of design aspects for COVID-19 dashboards. The study compares four existing solutions using eye-tracking and subjective interviews. Based on design insights from this comparison, two self-developed dashboards were created. In the last step, these two dashboards were again compared. GazePlotter was used to visualize the recorded eye-tracking data.

2 INTRODUCTION

This chapter offers a brief introduction to cognitive cartography, emphasizing its historical development and its significance, illustrated by an analysis of the International Cartographic Association's research agendas. It then summarizes the history, principles, and applications of eye-tracking technology. Finally, the chapter connects these topics by discussing the use of eye-tracking in cognitive cartography, particularly highlighting the author's contributions to eye-tracking methodologies and their application in cartographic studies.

2.1 Cognitive Cartography

Cognitive map-design research aims to understand human cognition in order to **improve the design and use of maps** (Montello, 2002). The question of how people read, analyze, and interpret maps and why some are better at navigating and orienting themselves in the terrain than others is interesting to both cartographers and psychologists. Psychological cognitive studies began over 100 years ago, while cartographic studies were first conducted in the 1950s and 1960s (Lobben, 2004). During this time, many geographers adopted ideas first formulated by Wright (1942) in his article titled "Map Makers Are Human." He pointed out that maps, like any other human creation, reflect the subjectivity, biases, and limitations of their creators. At the same time, he emphasized that map users must apply critical thinking when interpreting maps, so as to understand their limitations and avoid unwarranted conclusions.

A significant milestone in the field of cognitive cartography was the publication of "The Look of Maps" (Robinson, 1952), which initiated research focused on map design and the cognitive aspects of cartography. Robinson urged cartographers to systematically observe and measure data on how people view and interpret maps. In the chapter dedicated to map design, he explicitly states that cartographers should design maps to "lead the eye in the direction and sequence necessary for the proper grasp of the complicated material" (Robinson, 1952, p. 68).

The task of cognitive cartography is to understand how users read the individual elements of a map and how the meaning assigned to these elements differs among various users. Petchenik (1977, p. 119) stated that "for there to be successful communication, the receiver of a message must be able to construct meaning from the physical stimulus in essentially the same way that the originator of the meaning constructed it."

A theoretical foundation for research on the process of map reading is provided by **cartographic communication models**. In their simplest form, these communication models describe maps as channels for transmitting information from the source (reality) to the receiver (reader). More complex models include additional processes, such as encoding between reality and the map, the process of decoding the map on the receiver's side, or noise during transmission (Montello, 2002). Various models exist, developed by different authors from diverse cartographic schools (Board, 1978; Koláčný, 1969; Morrison, 1977; Ratajski, 1978); however, all of them are based on Shannon's (1948) theories of communication. Most of these models share four common entities. The first entity is the map's author—the cartographer. On the opposite end of the model is the map user, the target consumer. The map itself acts as the communication instrument between these two entities. The final entity in this tetragon is the shared understanding of reality between the map author and the map user.

Perhaps the most significant communication model was created by the Czech cartographer Antonín Koláčný. Koláčný's (1969) model (Figure 2) specifies that the reality (Universum) to which the cartographic representation refers is not exactly the same for the cartographer (map maker) and the receiver (map reader). For the cartographer, it refers to a part of reality, U_1; for the map reader, it refers to a part of reality, U_2. The cartographer is represented in the diagram by the content of his knowledge (S_1), which is influenced by his tasks, aims, knowledge and experience, abilities, and other characteristics, as well as his psychological processes and external conditions (i.e., environmental influences). The map reader is represented analogously, by the content of his consciousness (S_2), including his needs, interests, and aims, his knowledge and experience, his abilities, and other characteristics, as well as his psychological processes and the external conditions of his environment. Both the cartographer and the map user know a cartographic language, i.e., a system of map symbols and rules for their use, denoted by L. The map (M) is considered to be a system of map features that embody cartographic information, I_c (Beitlova et al., 2022).

Experiments in cognitive cartography can help identify barriers and failures in cartographic communication from both the map maker's and the map reader's perspectives. Qualitative analysis of eye movement data can **reveal the strategies** used during the map reading process. The map reader's understanding of the depicted phenomenon is indicated by the overlap of entities U_1 and U_2. The simplest way to measure this overlap is by analyzing the accuracy of the map reader's answers in cognitive cartography experiments.

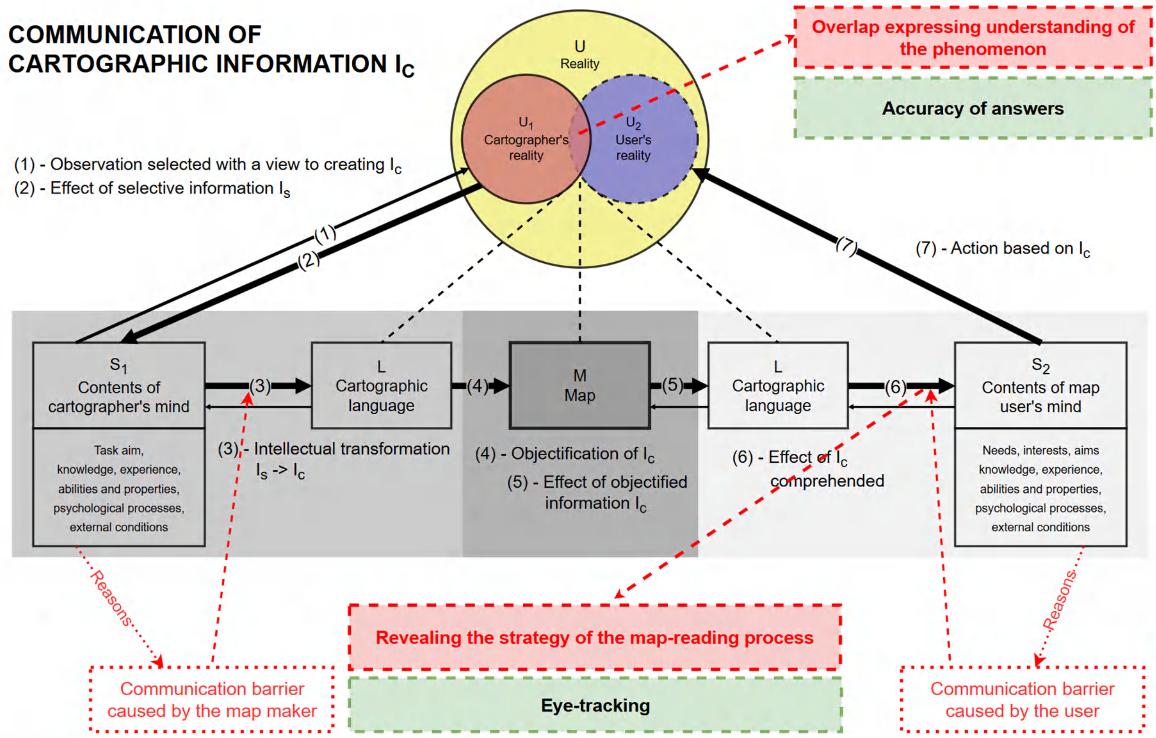


Figure 2. Cartographic communication model according to Koláčný (1969), supplemented by the notes of the author.

Robinson and Petchenik (1976) and Petchenik (1977) criticize these systematic cartographic communication models based on information theory. They present a Venn diagram (Figure 3) illustrating cognitive elements in cartographic communication. The outer rectangle represents all conceptions of the geographical environment, which can be correct (S_C) or incorrect (S_E). Area A shows the author's conceptions; area B shows the map user's conceptions. The green rectangle represents concepts marked on the map by the author. Area M_1 shows what the user already knew, M_2 represents new knowledge gained from the map, and M_3 indicates concepts the user did not understand, highlighting communication failures.

Petchenik (1977) suggests that research should focus primarily on cases of cartographic communication failure connected with area M_3. Area U shows an unplanned increment of the user's spatial understanding that was neither intended nor symbolized in the map in any way by the author.

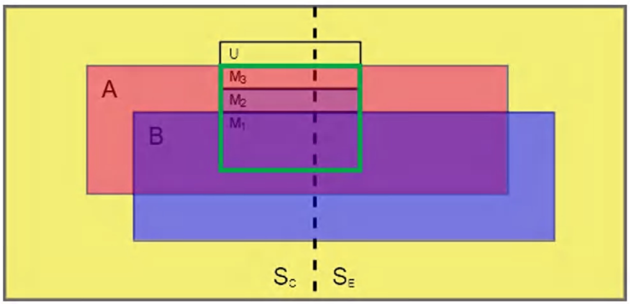


Figure 3. Venn diagram summarizing the cognitive elements in cartographic communication (Robinson & Petchenik, 1976).

The period between 1975 and 1982 was the golden age for map design research. Nearly 30% of articles in professional cartographic journals were dedicated to this topic, and a large number of dissertations were also written, focusing on the user aspects of maps. Before 1975, most cartographic articles were focused on historical themes, while after 1985, there was a significant increase in topics related to automated cartography (Gilmartin, 1992).

In the early 1980s, the use of computers in cartography increased significantly, leading to the rapid expansion of GIS technology. This made map production much faster and simpler. According to Montello (2002), this digital revolution led to a decline, if not a complete loss, of interest in cognitive research in cartography and even in cartography itself.

However, in 1995, the publication of "How Maps Work" (MacEachren, 1995) reignited interest in cognitive and semiotic research related to map design, interpretation, and geovisualization. During the 1990s, there was a renewed increase in the number of professional cartographic articles focusing on cognitive aspects. Since the establishment of GIScience in 1992, cognitive research has been a fundamental part of the field, and this continues to be true today (Montello, 2009).

The **necessity of cognitive research** in cartography can be illustrated by the proposal of the research agenda for geovisualization by MacEachren and Kraak (2001). The research agenda aimed to identify key topics for the International Cartographic Association (ICA) Commission on Visualization and Virtual Environments to address. Three main themes were selected: (1) Representation of Geospatial Information, which included semiotics, user interaction, dynamic visual tools, and immersive environments; (2) Integration with Knowledge Construction, focusing on data mining, computational analysis, pattern discovery, and human-machine collaboration; and (3) Interface Design for Geovisualization Environments, encompassing creative thinking, metaphors, cognitive challenges, virtual environments, and mobile displays. Cognitive aspects were integral to all three topics, which led to the creation of a fourth topic area focused on cognition and usability. This working group, led by Terry Slocum, addressed how geovisualization tools were used and perceived by different users. They explored questions such as the effectiveness of navigation tools, user reactions to immersive environments, and the factors that determined the success or failure of geovisualization. The goal was to develop cognitive theories and usability assessments that supported the effective use of dynamic, interactive geovisualization methods across various contexts.

However, the importance of cognitive and usability aspects extends beyond a single ICA Commission to the International Cartographic Association as a whole. Virrantaus et al. (2009) proposed a research agenda for ICA to guide the Commission's work and highlight the ICA's contribution to global scientific research. The ICA Research Agenda was developed through discussions in the early 2000s, structured planning, and brainstorming sessions using the Mind Map technique. Ten keywords were selected: geographical information, metadata and SDIs, geospatial analysis and modeling, usability, geovisualization, map production, cartographic theory, history of cartography, education, and society. Usability included terms such as user-centered map design, special map interfaces and augmented realities, usability testing, visual perception of maps, use of maps in specific situations, spatial thinking, understanding and cognition, location-based services, and adaptive maps. An online survey among the chairs of Commissions and Working Groups identified research overlaps and gaps, with usability issues being central to many Commissions' interests. Visual perception of maps was highlighted as the most important, with more than 60% of ICA Commissions and Working Groups showing interest (Figure 4).

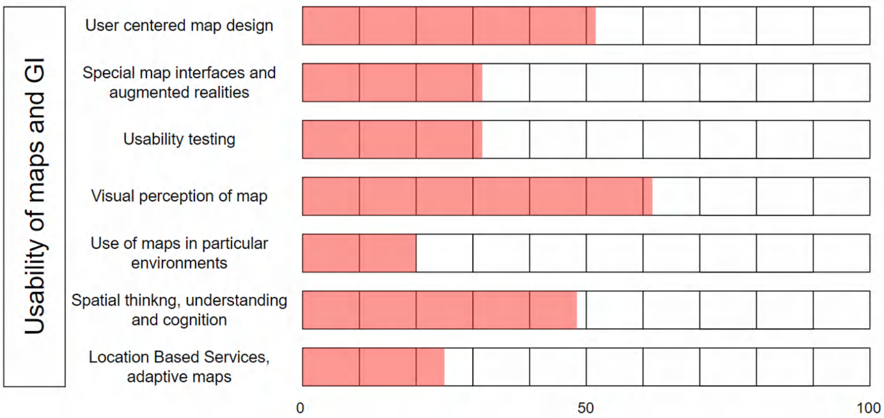


Figure 4. Percentage of ICA Commissions and Working Groups interested in usability of maps and geographical information. Modified according to Virrantaus et al. (2009).

Currently, within the International Cartographic Association, there are two commissions dedicated to the user aspect of maps, methods of evaluating map effectiveness, and their optimization: the UX: Designing the User Experience Commission and the Commission on Cognitive Visualization.

2.2 History, Principle, and Use of Eye-tracking

Petchenik (1977, p. 126) emphasized the importance of understanding how map users perceive specific elements and how the interpretation of these elements varies among individuals. Similarly, Montello (2002, p. 289) stated that "**one of the more significant empirical approaches to map psychology involved recording the eye movements of subjects as they viewed maps**". Eye-tracking, a technology that records an individual's eye movements, can be a valuable method for this purpose, providing detailed information about where, when, for how long, and in what sequence a person looked.

In 2009, Montello reviewed recent achievements in cognitive research within geographic information science and discussed future directions, suggesting exploring methodologies like eye-tracking. He noted that prior to 2009, technical difficulties with eye-trackers and inadequate theory development had led to the discontinuation of eye movement analysis in GIScience, despite the promise this technology was demonstrating in cognitive research outside GIScience. Montello emphasized the necessity for GIScientists to gain new insights into geographic information through eye movement recording.

Human vision is the most important and most utilized sense, providing a wide field of view of approximately 200° horizontally and 130° vertically (Biedert et al., 2010). The structure of the eye allows light to pass through the pupil, be inverted, and focus on the retina at the back of the eye, where photoreceptor cells called rods and cones convert the light into electrical impulses sent to the brain. The retina, however, has varying structure; sharp images are captured by cones, which are densely packed in a small central region called the **fovea centralis**. Due to its limited size, covering less than 2% of the visual field, the eyes must constantly move to position objects of interest within this area of the highest visual resolution (Synek & Skorovská, 2014). This continuous movement is essential for detailed vision and is facilitated by various types of eye movements, including fixations and saccades.

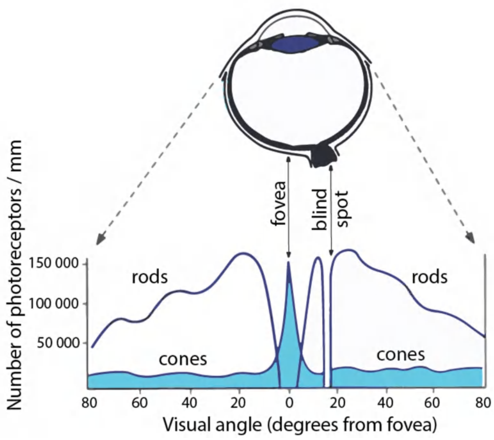


Figure 5. The number of photoreceptors in the human eye. Modified according to Snowden et al. (2012).

Fixations are one of the most crucial eye movements, allowing the eyes to stay focused on a single point for several milliseconds to a few seconds (Holmqvist et al., 2011). During fixations, the brain processes the visual information from the observed image. At the same time, the eye performs three types of micro-movements: tremor, drift, and microsaccades (Martinez-Conde & Macknik, 2008). These movements are unconscious and detectable only with high-frequency eye-trackers. While they are relevant in neurological and clinical research (van der Geest et al., 2001), they hold no significance in cartographic research. One primary focus of eye movement research in cartography is on the location, duration, and sequence of fixations.

In contrast, **saccades** are rapid, ballistic movements that shift the gaze from one fixation point to another, with visual perception significantly reduced during these movements due to a neurological process called saccadic suppression (Hammoud & Mulligan, 2008). Saccades are typically visualized as straight lines between fixations. However, saccades are rarely straight; they can vary in shape and curvature. Often, saccades don't end precisely at the target (the center of the next fixation). Instead, the eye wobbles before stopping, resulting in a movement known as a glissade. The eye performs 3-4 saccades per second, totaling around 200,000 per day (Bojko, 2013).

Figure 6 shows an example of the eye movements of one participant during a visual inspection of a map. Black lines represent raw data recorded with a frequency of 250 Hz. From these, fixations and saccades were identified. Fixations are represented with red circles, whose size corresponds to the length of the fixation. Saccades are visualized as straight red lines.

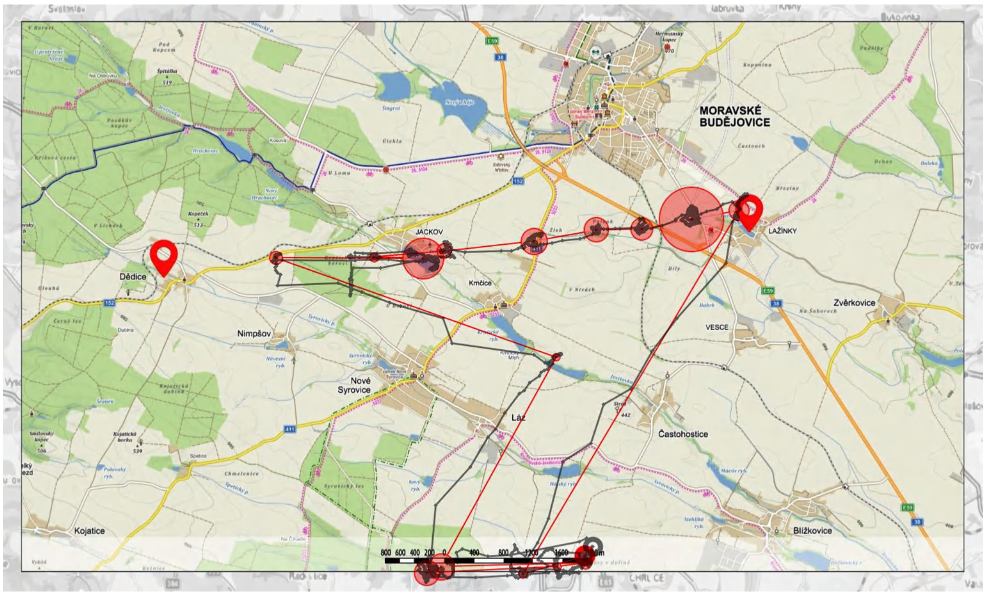


Figure 6. Fixations and saccades (red circles and red lines) and raw eye movement data (black lines) recorded over a map. (Source of the background map: Mapy.cz).

The **identification of fixations and saccades** from eye movement recordings is not trivial. Fixations can be identified manually (Harris et al., 1988), but there are also a large number of classification algorithms. Komogortsev et al. (2010) mention several studies that have addressed these algorithms, e.g. (McConkie, 1981; Salvucci & Goldberg, 2000; Sauter et al., 1991). Among the most commonly used are the I-VT (Velocity Threshold Identification) and I-DT (Dispersion Threshold Identification) algorithms. In the studies included in this habilitation thesis, an SMI RED 250 eye-tracker with a sampling frequency of 250 Hz was used. For these frequencies, the I-DT algorithm is most commonly used (Holmqvist et al., 2011).

This algorithm requires two parameters: a threshold for Dispersion and the Duration of the fixation. The I-DT algorithm focuses on detecting fixations based on the spatial and temporal proximity of measured eye positions. It defines a time window (Duration) that moves through the recorded data. The spatial proximity of points (Dispersion) within the time window is compared to a defined threshold value. If the dispersion is lower than the threshold, the points within the time window are marked as part of a fixation. Otherwise, the window shifts by one record and the first point of the previous window is classified as part of a saccade (Komogortsev et al., 2010). The main drawback of this algorithm is the high interdependence between Dispersion and Duration. A small dispersion threshold combined with a long duration can result in no fixations being identified. However, when the thresholds are set appropriately, the I-DT algorithm provides very good results (Salvucci & Goldberg, 2000). Therefore, it is crucial to correctly set the parameters for the I-DT algorithm. The optimal settings of thresholds for this algorithm for cartographic studies were determined in a study by Popelka (2014). These settings were then used in all subsequent studies.

The **first attempt to record eye movements** was made by Émile Javal, the founder of the ophthalmology laboratory at the Sorbonne in Paris. Javal (1878) was the first to develop a device for recording eye movements. He used the reflection of a mirror attached to the eye, which was then recorded on a photographic plate. Javal also introduced the term "saccades". Building on Javal's work, Delabarre (1898) replaced the mirror with a plaster cup attached to the eye with a wire. At the beginning of the 20th century, there was a rapid advancement in eye-tracking technology. Dodge and Cline (1901) developed a photographic device that did not require any physical attachment to the eye. This invention sparked a revolution in eye movement research and enabled researchers to determine the specific points where subjects focused their visual attention. A key person in the development of the area of visual attention was Buswell, who systematically examined the eye movements of respondents viewing complex stimuli, such as paintings (Buswell, 1935). Buswell's work revolutionized the field of eye-tracking. Another pioneer in eye-tracking research was Yarbus et al. (1967), who conducted an experiment showing the same image to respondents seven times, each time with a different task. This experiment confirmed Buswell's earlier observation that the tasks given to respondents could significantly alter their fixation points. Yarbus's work became a classic in eye-tracking research, often cited as clear evidence that "high-level" (task-related) factors can overshadow any "low-level" (stimulus-related) factors.

Eye-trackers developed in the late 19th and early 20th centuries revealed fundamental insights into the nature of eye movements. Technological advancements allowed for the exploration of new questions and identified unexpected issues in the psychology and physiology of eye movements and their relationship to cognition. In the 1970s, most work focused on technical improvements to eye trackers, primarily aimed at increasing their accuracy (Mohamed et al., 2007).

Duchowski (2007) defines four eye-tracking techniques:

- Electrooculography,
- Scleral contact lens,
- Photo-oculography and video-oculography,
- Pupil and corneal reflection tracking.

The first three of these techniques measure the eye's position relative to the head. The last one, **pupil and corneal reflection tracking** (P-CR), is the only one that measures the eye's orientation in space, known as the "Point of Regard" (L. R. Young & Sheena, 1975).

P–CR eye trackers were introduced by Merchant (1967). They contain one or more infrared illuminators and a high-speed infrared camera that records participants' eyes. The “P” in the name of the technique refers to the pupil center in the camera image, and the CR to one or more reflection center(s) in the cornea from infrared illuminators (Figure 7). P–CR eye trackers estimate gaze direction as a function of the relative positions of P and CR coordinates. P–CR eye trackers are prevalent due to their accuracy, the wide range of models on the market, and the user-friendly software available for data analysis (Holmqvist et al., 2022).

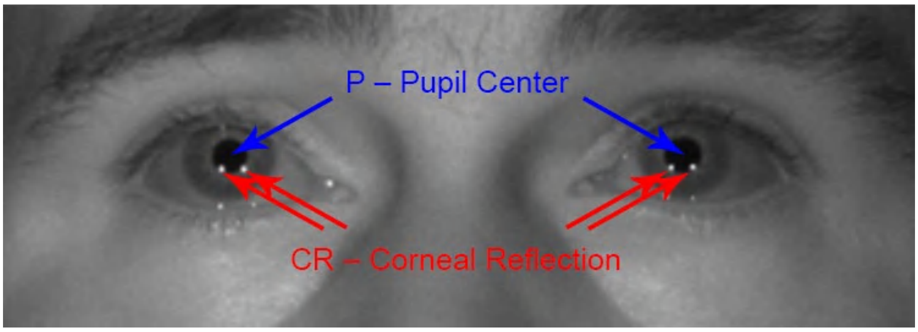


Figure 7. Pupil centers and corneal reflections recorded by a Tobii Pro Spectrum 300 eye-tracker.

Video-based P-CR eye-trackers dominate the contemporary market almost completely. Nevertheless, even if they all use the same measurement technology, they can be divided into two main categories – static (remote) eye-trackers and head-mounted (wearable) eye-trackers (Figure 8).

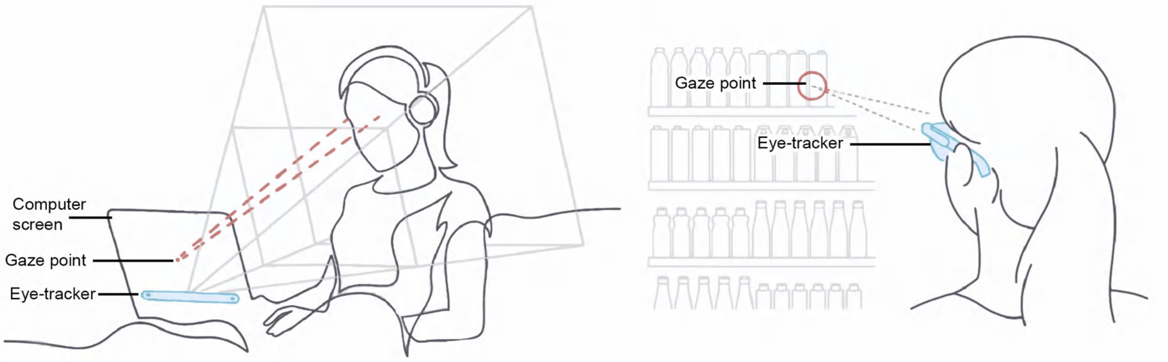


Figure 8. The principle of remote (left) and wearable (right) eye-trackers. Modified from Tobii.com.

The most common set-up is the **static eye-tracker**, where the cameras and infrared illuminators are placed in front of the participant. Holmqvist et al. (2011) categorized static eye-trackers into two types: tower-mounted and remote. Tower-mounted eye-trackers are positioned close to participants and restrict head movements, while remote eye-trackers operate from a distance. Nowadays, tower-mounted eye-trackers are rarely used, with most research relying on remote eye-trackers. Typically, stimuli are displayed on a computer screen with the eye-tracker mounted below it. However, the same hardware can also record data from real-world scenes or physical stimuli such as wall maps or paintings.

Remote eye-trackers are used in various fields. Duchowski (2007) divides the use of eye-tracking into two basic categories – interactive and diagnostic use. In the interactive use, the eye-tracker is employed to control the computer. The gaze replaces or supplements common peripherals such as the keyboard (Majaranta & Räihä, 2002) or mouse. This application of eye-tracking finds its use among disabled individuals (Caligari et al., 2013), but also in playing computer games (Sundstedt, 2012) and many other fields.

In diagnostic applications, eye-tracking is commonly used in clinical research to identify eye diseases, as well as to monitor and diagnose mental and neurological disorders such as autism spectrum disorder, ADHD, and Parkinson's disease. In various fields, eye movements are recorded to determine which parts of the stimuli captured the respondent's attention, with cognitive psychology being a traditional area of focus (Matlin, 2014). Eye-tracking research has significantly contributed to understanding how people read, with studies dating back to the late 19th century. Early research focused on the fundamental nature of eye movements (Tinker, 1946), and modern studies often focus on individuals with dyslexia (Dostálová et al., 2024; Sibert et al., 2000). One of the most common commercial uses of eye-tracking is in marketing. It is used to analyze consumer behavior when selecting specific products, in print and television advertisements, in political marketing, label design and branding, and most recently, in the evaluation of websites (Wedel & Pieters, 2008). In education and didactics, there are two basic approaches to utilizing eye-tracking: evaluating materials (Jarodzka et al., 2017) and evaluating users – teachers (Yamamoto & Imai-Matsumura, 2013) or students (Kekule, 2015; Skrabankova et al., 2020).

In the case of **wearable eye-trackers**, the cameras and illuminators are on the head of the participant, mounted on a helmet, cap, or pair of glasses. The main advantage of this system is that it allows participants maximum mobility. For this reason, wearable eye-trackers are usually used in real-world experiments. Outside of cartography, for example, in sports, wearable eye-trackers are used to record the eye movements of athletes in ball sports like basketball, football, or tennis. Most studies focus on sports where vision is crucial for both movement control and decision-making (Kredel et al., 2017). In the automotive and aerospace industries, eye-trackers are worn by drivers or pilots to analyze their visual and cognitive performance (Kircher, 2007; K. Young et al., 2007). Wearable eye-trackers have also been used to record the eye movements of infants (Franchak et al., 2010) or animals (Pelgrim et al., 2023), to analyze shopping behavior (Hummel et al., 2021) and identify hotspots of attention in museum exhibitions (Garbutt et al., 2020).

Moreover, in recent years, eye-trackers implemented into **VR headsets** have become more and more common; eye-tracking is used there for enhancing user interaction and also improving the realism and speed of image rendering. Eye-tracking enables foveated rendering, which renders high-resolution graphics only where the user is looking.

2.3 Eye-tracking in Cognitive Cartography

Early eye movement research focused on areas other than cartography and did not involve map reading. Initial studies concentrated on reading text and later expanded to various graphic stimuli, such as images, newspaper ads, and photographs. Researchers conducted numerous studies on visual exploration of aerial photographs, radar screens, X-rays, and scenes recorded from moving cars or airplanes. These studies were generally conducted by researchers specializing in the specific fields related to the stimuli being studied or by psychologists with a strong interest in applied research (Steinke, 1987). One of the first studies related to cartography was conducted by Enoch, who performed several experiments focused on the interpretation of aerial photographs (Enoch, 1959).

A milestone in eye-tracking research in cartography was the 1970 symposium, "Influence of the Map User on Map Design," at Queen's University, Ontario. Psychologist Leon Williams presented his eye-tracking research conducted at Honeywell, focusing on visual search processes (Williams, 1971). This presentation influenced several attendees to explore how people read maps using eye-tracking. One such cartographer was **George Jenks**, who was enthusiastic about applying eye-tracking to address cartographic problems. He later stated that his excitement stemmed from the technique's ability to "get inside of the map reader's head" and its potential to open up research possibilities in cartography that were previously considered impossible (Steinke, 1987).

Jenks and his students began studying how readers perform regionalization on a dot map. He was surprised by the differences among the students' results. Jenks used a modified Kodak 8 mm camera to photograph respondents' corneas. This inexpensive but time-consuming method resulted in scanpaths that revealed irregular and individual map reading patterns. He found that students concentrated on areas with a high density of information and overlooked those with less information. The results were published in the International Yearbook of Cartography (Jenks, 1973); this is believed to be the first published cartographic eye-tracking research.

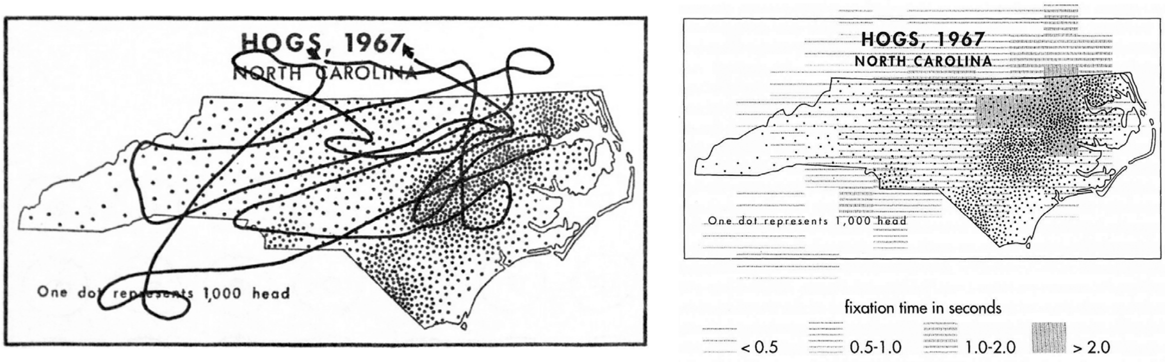


Figure 9. The results of the first cartographic eye movement study by Jenks (1973).

During the 1970s, eye-tracking research in cartography began to flourish, with several studies emerging that explored various aspects of map reading and user interaction. Steinke (1987) provided a comprehensive review of eye-tracking research up until the mid-1980s, noting the initial promise of this line of inquiry. However, despite these early advances, cartographic eye-tracking research experienced a **significant decline** in the following decades. From the mid-1980s until the early 2000s, there was a notable absence of studies in this field, due to limitations in both technology and theory. Only at the turn of the millennium did a few eye-tracking studies in cartography reappear.

One of the first significant contributions in the early 2000s was by Brodersen et al. (2002), who introduced a novel method for assessing the usability of topographic maps using both eye and head tracking. Their approach combined traditional usability techniques, such as think-aloud protocols, semi-structured interviews, and video analysis of non-verbal behavior. This study, involving ten participants, explored the relationship between perceived map complexity and cognitive behavior, as measured through eye movement data. The results were promising, laying the groundwork for future research with tools like the newly developed MapObs software, which enables the synchronized analysis of eye movements and verbal responses.

Building on this renewed interest, Fabrikant et al. (2008) conducted a controlled experiment focusing on small multiple map displays. They introduced the concept of "inference affordance" to complement traditional empirical measures such as task accuracy and completion time. By applying sequence alignment analysis techniques from bioinformatics, Fabrikant and her team presented a novel methodology for quantifying how well visual analytics tools support inference-making, further pushing the boundaries of eye-tracking research in cartography.

Similarly, Alacam and Dalci (2009) evaluated the usability of four popular web maps—Google Maps, Live Search Maps, MapQuest, and Yahoo Maps—analyzing the impact of iconic representation and pop-up windows on user performance. Their findings revealed significant differences in usability, with Google Maps outperforming the others in task completion and pop-up window usage. This study also highlighted how user experience levels influenced the effectiveness of different map designs.

Around the same time, Coltekin et al. (2009) integrated traditional usability methods with eye-tracking analysis in a comparative study of two online map interfaces. While standard usability metrics uncovered several issues, the eye movement data provided deeper insights into the cognitive demands placed on users, offering a more detailed understanding of how different interface designs affect task completion and overall usability.

More recently (after 2010), it has been suggested that eye tracking be used systematically as part of empirical methodology in both cartographic research and practice (Kiefer et al., 2017). The **current state of cartographic eye movement research** is summarized in Wang et al. (2016), Kiefer et al. (2017), and Krassanakis and Cybulski (2019).

Wang et al. (2016) performed a bibliometric analysis of Web of Science (WoS) and used multiple visual metaphors to illustrate the intellectual structure of eye movement research in cartography. The authors used co-citation analysis to identify classic literature in the field. According to their results, the earliest influential work was Buswell's *How People Look at Pictures* (1935), which provided the first comprehensive analysis of eye movement behavior. Rayner's work (1998) on eye movements in reading and information processing emerged as the most significant, focusing on visual attention and map usability. Other prominent works included studies by Coltekin et al. (2009) on interactive map designs and Itti's (1998) research on computational models of visual attention, highlighting the interdisciplinary nature of this field, integrating psychology, usability, and computer science.

The authors used a bibliographic coupling network to identify the clusters of research themes. In cluster 1, usability metrics are primarily applied to traditional cartography challenges, such as improving map label placement, map legend usability, and enhancing eye movement data analysis methods. Cluster 2 focuses on usability research on emerging technologies, including web mapping navigation, citizen-based mapping, mobile devices, and geo-portal user experience. Other research areas identified across clusters include 2D and 3D terrain visualization, differences in expert and novice behavior, animated maps, web mapping image enhancement, and volunteered geographic information (VGI).

According to the results of the co-occurrence network analysis, the author of this habilitation thesis was identified as the **most productive author** in the field of eye-tracking and cartography (Wang et al., 2016, p. 11). Palacký University Olomouc was named as a center of eye-tracking research in cartography, together with Zurich and Ghent Universities (Figure 10).

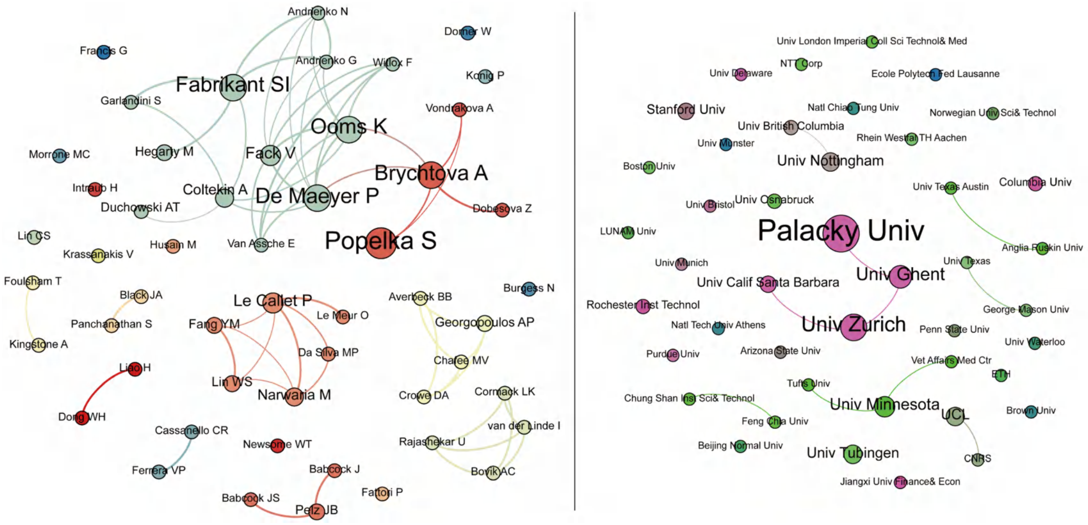


Figure 10. Co-occurrence network of authors (left) and institutions (right) active in cartographic eye-tracking research (Wang et al., 2016).

Kiefer et al. (2017) provided an overview of the research which utilizes eye-tracking methodology in human navigation and cognitive cartography, or assesses gaze as an input modality for geographic information. In cognitive cartography, they differentiate recent studies into four groups.

The first group consists of studies focused on design guidelines for **static maps**. In this type of research, stimuli are typically created by systematically varying a single cartographic design variable or altering the entire map design. Gaze data are used either as aggregated measures to infer cognitive states or as exploratory gaze visualizations to explain other dependent variables, such as completion times. Illustrative examples include comparisons of label placement methods (Ooms, Maeyer, et al., 2012), the effects of color distance and font size on map readability (Brychtova & Coltekin, 2016), evaluation of uncertainty visualization methods (Brus et al., 2019) or comparisons between 2D and 3D visualizations ((Liao et al., 2017), (Popelka, 2018b), **[paper 2D-3D]**).

The second approach to using eye-tracking in cognitive cartography focuses on varying the **user group** of a map. Fabrikant et al. (2010) explored how expertise influences map viewing and also evaluated the impact of saliency on weather maps. This approach, which considers user group differences, is crucial for effective cartographic communication (Thorndyke & Stasz, 1980) and was later adopted by Ooms, De Maeyer, et al. (2012), Stofer and Che (2014), **[paper UrbanPlans]** or **[paper Author-Reader]**.

As a third research topic, **animated maps** can be identified. While animations can serve as attention-grabbing elements, they are also used to convey complex spatio-temporal information, such as real-time spatial depictions in air traffic control (Maggi et al., 2016). Eye-tracking studies on animated maps typically focus on two main objectives: (1) determining how to design animations to effectively capture attention (e.g., considering timing (Krassanakis et al., 2016) or visual design (Dong et al., 2014)), and (2) investigating how viewers comprehend animations (Opach et al., 2014). The evaluation of eye-tracking data collected on animated stimuli is more challenging because most vendors' standard software packages do not support the automated analysis of visual attention on dynamic stimuli. To overcome this deficiency, a tool called GazePlotter was developed **[paper GazePlotter]**.

Even more challenging is the research conducted on **interactive maps**, which has increased in importance due to the widespread use of the internet and mobile technologies, leading to a surge in the popularity of interactive maps among the general public (A. Mendonça & Delazari, 2014).

The rise of more advanced visualization technologies has resulted in a greater cognitive load on individuals engaging with these visualizations, particularly when using maps in time-sensitive decision-making scenarios (Fuhrmann et al., 2015). By understanding how users interact with spatial data, we can identify their needs and enhance maps to support effective and efficient decision-making while minimizing cognitive overload (Fairbairn & Hepburn, 2023).

Usability testing of interactive maps was performed by Coltekin et al. (2009), who compared two informationally equivalent but differently designed online interactive map interfaces. Golebiowska et al. (2017) examined how users interact with a geographic visualization tool featuring coordinated and multiple views, including a map, parallel coordinate plot, and table. Despite the potential confusion from multiple visual components and interactive functions, the study found that users appreciated the flexibility to choose visualization methods. Manson et al. (2012) used eye-tracking and mouse metrics for the analysis of four different web map navigation schemes (pan zoom, double-clicking, zoom by rectangle, and wheel zoom); they found that the participants preferred rectangle zoom followed by wheel zoom.

Interactive maps have also served as stimuli in the research of the author of this habilitation thesis. In Popelka, Herman, et al. (2019), three map-based visual analytics tools were evaluated, leading to the development of a set of recommendations for graphical user interface design. To analyze user interactions with 3D environments, a tool called 3DGazeR was developed by Herman et al. (2017). The functionality of the tool was introduced in a case study using digital elevation models as stimuli. Moreover, the publications **[paper Glyphs]**, **[paper WeatherMaps]**, **[paper Swipe-MultipleView]**, and **[paper Dashboards]** are focused directly on the assessment of interactive map interfaces.

A special example of an interactive map is a **web map**, whose visible extent is dynamic, based on the input by the user. It is necessary to log all interactions with the map and synchronize them with the eye-tracking data (Ooms, Coltekin, et al., 2015). For this, the ET2Spatial tool was developed **[paper ET2Spatial]**.

While the work of Kiefer et al. (2017) was more broadly aimed, including sections about mobile eye-tracking for human navigation as well as the pervasive use of eye-tracking, Krassanakis and Cybulski (2019) provided a review specifically of eye movement studies which focused on the investigation of the map reading process. The authors divided the map reading eye-tracking research into **five categories** – Cartographic symbolization and design principles, Comparing 2D and 3D representations, Map users' expertise, Cartographic studies on various topics, and Eye-tracking analysis tools and methods delivered by the cartographic community. The author of this habilitation thesis is active in all these categories; Krassanakis and Cybulski refer to fourteen of his works in their review. The issue of cartographic symbolization and design principles is addressed in **[paper Glyphs]**, **[paper UrbanPlans]**, **[paper WeatherMaps]**, **[paper Swipe-MultipleView]**, and **[paper Dashboards]**. The comparison of 2D and 3D visualization was the content of the author's PhD thesis (Popelka, 2015) and is included in **[paper 2D-3D]**. Differences in map reading according to map users' expertise were studied in **[paper Author-Reader]**, **[paper Glyphs]**, and **[paper UrbanPlans]**. The cartographic community not only transfers and uses existing methods and tools for the investigation of maps, but it substantially contributes to the further extension of methodological approaches to eye-tracking, as in **[paper ScanGraph]**, **[paper EyeTribe]**, **[paper GazePlotter]**, **[paper ET2Spatial]**.

3 METHODS

This chapter provides an overview of the research methods and techniques used in cognitive cartography, with a particular focus on eye-tracking. It then outlines the design of eye-tracking experiments and details the software and hardware utilized in this habilitation. Finally, it introduces methods for analyzing eye-tracking data, with an emphasis on those developed during this habilitation.

3.1 Research Methods in Cognitive Cartography

Cognitive cartography employs a wide range of research methods and techniques that have been and continue to be developed by psychological disciplines. The methods and techniques used in cognitive cartography fall under the umbrella of user experience and usability evaluation testing.

Usability is primarily associated with websites and computer programs, but it is also essential for the evaluation of objects of everyday use as well as maps. The International Organization for Standardization defines usability as "The extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO, 2019).

Rohrer (2014) categorizes usability research methods according to three criteria:

- subjective (attitudinal) and objective (behavioral),
- qualitative and quantitative,
- by context of use.

Subjective methods reveal what respondents say, while objective methods show what respondents actually do. Svoboda (1999) considers objective methods to be those that prevent deliberate distortion by the respondent and ensure that the results are independent of the evaluator.

Qualitative studies provide data on respondents' behavior or attitudes based on direct observation, whereas quantitative studies collect data indirectly through measurement. Qualitative data analysis is typically non-mathematical, while quantitative data are precisely quantified and analyzed statistically. Interpretations are then based on these statistical results (Štěrba et al., 2015). Qualitative methods are more suitable for answering "why" or "how" questions, whereas quantitative methods address "how many" questions (Rohrer, 2014). Both approaches can effectively complement each other, particularly in testing the effectiveness of cartographic visualization (Štěrba et al., 2015).

Rohrer's final criterion for dividing research methods is the context of use, which refers to how respondents interact with the product during testing. This can be natural or nearly natural use, laboratory use, a combination of both, or evaluation without direct product interaction.

Figure 11 presents a selection of usability evaluation research methods applicable in cognitive cartography. However, this overview is incomplete; there are many more usability and user experience evaluation methods (Hanington & Martin, 2012).

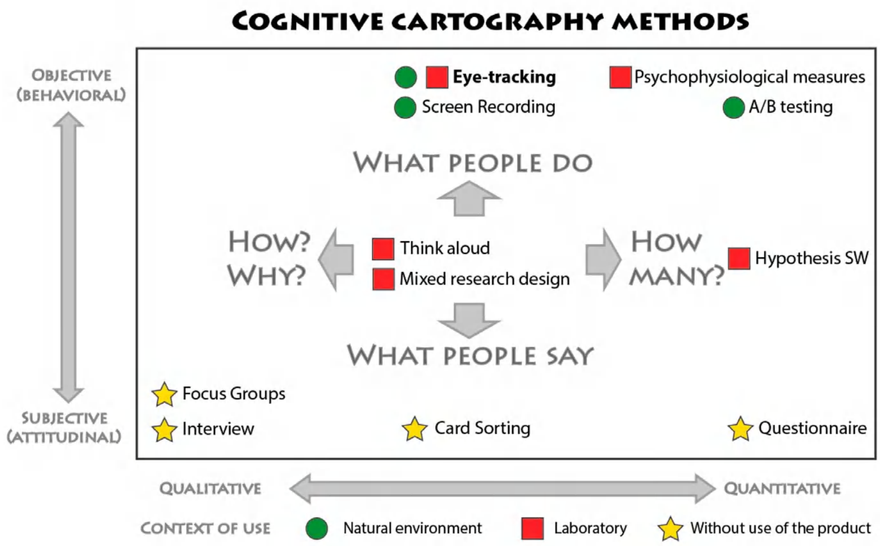


Figure 11. An overview of selected cognitive cartography methods and their classification according to three criteria. Burian et al. (2018), modified from Rohrer (2014).

This habilitation thesis uses eye-tracking to evaluate and compare map visualizations and interactive map interfaces. As is evident from Figure 11, eye-tracking is an objective method because it shows what people do, specifically where they look. Eye movement data can be analyzed qualitatively as well as quantitatively. The qualitative analysis consists of observing how the participant navigated the visual stimulus. In contrast, the quantitative analysis is focused on the statistical evaluation of eye-tracking metrics and their comparison among various stimuli (maps). From the context of use, eye-tracking studies can be conducted in the field and the laboratory. However, laboratory use is more common, especially when evaluating maps using a remote eye tracker. It is important to note that eye movement data can rarely be used independently. While eye-tracking objectively shows where people are looking, it does not explain why they are looking there. Therefore, combining eye-tracking with other methods, such as think-aloud protocols, interviews, or questionnaires, is highly beneficial.

3.2 Experiment Design

Eye-tracking can be used to evaluate the usability of products (maps) in two ways: formative (qualitative) and summative (quantitative) research (Bojko, 2013). However, both approaches can be combined.

In **formative (qualitative)** research, the goal is to identify areas for improvement in the product. Formative research can be conducted using a single product as a stimulus, focusing on identifying its weaknesses and potential enhancements. This approach was predominantly used in the studies **[paper Dashboards]** and **[paper WeatherMaps]**. Although both studies compared several maps, their primary aim was to identify design flaws and weaknesses. In **summative (quantitative)** research, multiple versions of a product are compared. Summative research uses experiments, and the collected data are analyzed using statistical methods to explore, describe, and verify relationships between observed variables (Hendl, 2008). Primarily summative research has been undertaken in **[paper 2D-3D]**, **[paper Glyphs]**, **[paper Urban Plans]**, and **[paper Swipe-Multiple]**.

Experiments are used in summative research to determine causal relationships between variables. A causal relationship means that changes in variable A cause changes in variable B

(Martin, 2007). According to Campbell (1980), three conditions must be met to establish a causal relationship: the cause must precede the effect in time, the cause and effect must covary, and no other variables should explain the changes in the effect. An experiment that fulfills these criteria is considered internally valid (Punch, 2008).

Key characteristics of an internally valid experiment include manipulating the **independent variable** (the presumed cause), measuring the **dependent variable** (the presumed effect), and controlling **external variables** that could provide alternative explanations (Ferjenčík, 2000). In the case of cartography, an independent variable might be the type of map (e.g., 2D vs. 3D), while the dependent variable could be task completion time or the number of fixations. Controlling for intervening variables, such as the different brightness of the maps, is crucial to ensure that the findings are due to the manipulated variable and not other factors.

An experiment's design must address how to meet the three conditions required to establish a causal relationship, as outlined by Campbell (1980). A key decision in experimental design is how to assign respondents to different levels of the independent variable. In a **between-subjects** design, each subject is exposed to only one level of the independent variable. In a **within-subjects** design, each subject experiences (sees) all levels of the independent variable.

When a **between-subjects** design is used, different groups of participants are exposed to different levels of the independent variable (the map). Each participant experiences only one variant of the map, which helps to prevent learning effects and reduces the risk of bias due to repeated measures. However, the crucial problem is inter-individual differences between participants. There exist techniques to minimize this problem, like random assignment, matching, pretesting, and block randomization; however, they are not very effective in experiments with a small number of respondents, including eye-tracking experiments. For that reason, within-subjects designs are more frequent in (cartographic) eye-tracking studies.

When a **within-subjects** design is used, participants are exposed to all levels of the independent variable (e.g., all map variants). This design is more economical as each participant serves as their own control, effectively addressing the issue of non-equivalent groups. However, the most significant problem in this design is the learning effect, where participants may improve or change their behavior simply due to repeated exposure to the stimuli. Techniques like randomization and counterbalancing can mitigate this issue by varying the order of stimuli presentation. In most studies included in this habilitation thesis, the within-subject design was used. To avoid or minimize the learning effect, randomization **[paper Urban Plans]** and sometimes a slight change in the task **[paper Swipe-Multiple]**, **[paper 2D-3D]**, or time delay between exposures to different levels of the independent variable **[paper Glyphs]** were used.

The number of participants needed for the study is connected to the type of research and the type of experimental design. For formative experiments, the number of participants was calculated using an online calculator (Sauro, 2023) based on the calculations of Sauro and Lewis (2016). In the summative experiments, the number of participants varied from 19 to 40. Although the number of participants in these studies is not ideal, it aligns with other eye-tracking research, where studies have used sample sizes ranging from 10 to 26 participants. An analysis of 15 quantitative cartographic eye-tracking studies (Popelka, 2015) revealed an average of 25 participants, with time constraints often limiting the number of participants in such experiments. A limited number of available participants is one of the reasons why qualitative analysis is often used in this habilitation thesis.

3.3 Apparatus & Software

In most studies included in this habilitation thesis, an **SMI RED 250 eye-tracker** with a sample frequency of 250 Hz was used (Figure 12). The only exception is the study **[paper Author-Reader]**, where SMI data were enhanced by data recorded by Tobii X2-60 eye-trackers. In the study **[paper EyeTribe]**, data recorded by the SMI RED 250 were compared with those recorded by an EyeTribe tracker. ET2Spatial **[paper ET2Spatial]** uses data from the SMI and Tobii eye-trackers. The same is valid for GazePlotter **[paper GazePlotter]**, which also uses data from GazePoint and Varjo eye-trackers.

Figure 12. The SMI RED 250 eye-tracker used in the habilitation thesis (Photo Karel Macků).

Stimuli in the case studies were presented using the vendor software SMI Experiment Center. Images or screen recording stimuli were used. Data were analyzed using SMI BeGaze, but also in other software like OGAMA (Voßkühler et al., 2008) or V-Analytics (Andrienko et al., 2012). Moreover, tools that the author developed, like **[paper ScanGraph]** or **[paper GazePlotter]**, were used.

3.4 Analytical Methods

Many methods and approaches exist to analyze and visualize eye-tracking data. Recorded raw data contain timestamps and X and Y coordinates of the point of regard (in the coordinate system of the computer screen). For most analyses, fixations and saccades are identified from the raw data using various algorithms. In this habilitation thesis, the I-DT algorithm was used with the settings described in Popelka (2014).

A visualization of the eye movement trajectory called **scanpath** is usually used for initial familiarization with the recorded data. Scanpath refers to the trajectories of saccades connecting fixation positions, which are superimposed over the stimulus, which serves as a background. The scanpath shows fixations as circles (or possibly crosses) of various sizes (their radius corresponds to the duration of the fixations) and saccades as lines connecting these circles (Raiha et al., 2005). An example of a scanpath is displayed in Figure 13 - right. On the left side of the figure, the recorded raw data are displayed, from which the scanpath was generated by identification of fixations and saccades.

Figure 13. Recorded raw data (left) from which the scanpath (right) was generated (Popelka, 2018a).

Attention maps, also known as heat maps (Figure 14), are tools for visualizing the quantitative characteristics of a user's gaze. From attention maps, it is evident which areas of the observed image the users examined more and which areas they did not pay attention to. Attention maps are very useful in eye-tracking for quickly providing an overview of which parts of a document users focus on and which parts are suitable for deeper analysis. Both visualization methods described above are more useful for familiarizing oneself with the data rather than for the actual interpretation of results.

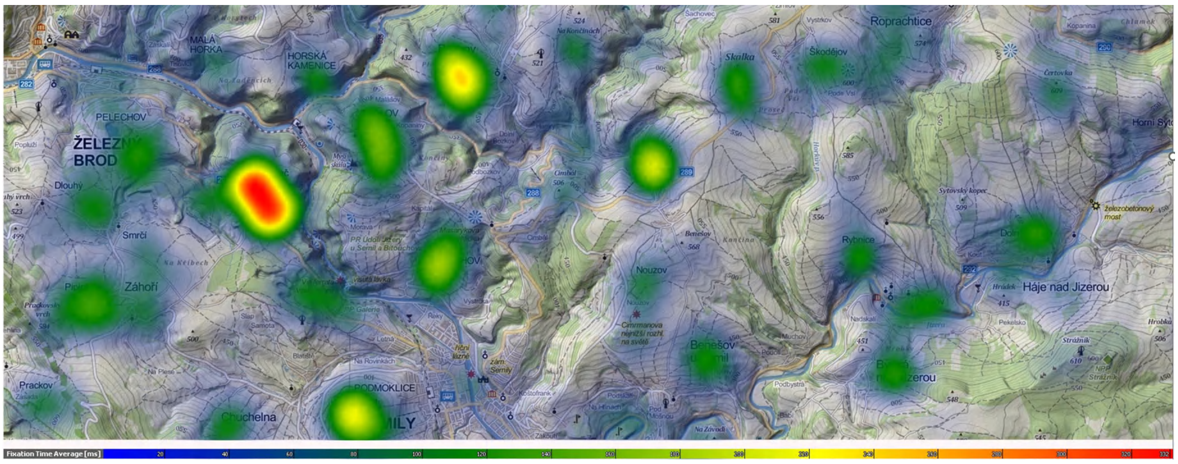


Figure 14. An attention map generated in SMI BeGaze with an accompanying legend.

A very useful method of data analysis is the creation of **areas of interest (AOI)**. AOIs are regions marked on the stimulus (i.e. around compositional elements of a map) for which it is determined how much they attracted the respondent's attention, how many fixations were recorded in specific areas of interest, the sequence in which these areas were visited, and so on. An example of AOIs is displayed in Figure 15. The creation of AOIs is simple for static stimuli, such as images. However, for dynamic stimuli, such as interactive maps, the process is much more time-consuming. This is because AOIs must be defined not only spatially but also temporally, as the stimulus changes over time.

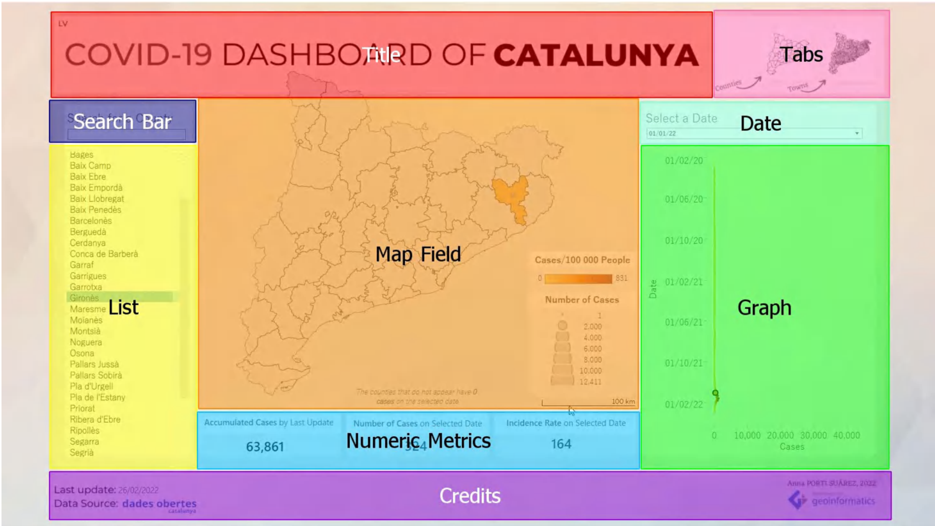


Figure 15. An example of areas of interest marked in the stimulus.

Eye movement metrics are commonly used to **quantitatively** compare participants' behavior. These metrics might be related to the stimulus as a whole, but they are often used to evaluate specific areas of interest. Holmqvist et al. (2011) mention more than 120 eye-tracking metrics and note that new ones are continually being developed. However, many of these metrics have only been used in a single study by one author. In practice, only a few of the most significant metrics are commonly used. In the author's dissertation (Popelka, 2015), the most important eye-tracking studies in the field of cartography were summarized, and the metrics used in these studies were identified.

The **fixation count** metric describes the number of fixations recorded while viewing the stimulus or within a specified area of interest. A higher number of fixations indicates a low level of search efficiency or an inappropriate user interface for the evaluated application (Goldberg & Kotval, 1999). The **fixation frequency** metric indicates the number of fixations recorded per second. It has the advantage of providing a relative value, so it is not affected by the varying lengths of trials for individual respondents. The **dwell time** metric is very important and frequently used. This metric is primarily used to analyze areas of interest, indicating how much time respondents spend looking at a defined area. The **scanpath length** metric describes the length of the eye movement trajectory within the stimulus. It can indicate the question's difficulty or the stimulus's clarity (Goldberg & Helfman, 2011). **Time to first fixation** is used almost exclusively to evaluate areas of interest. It indicates when a first fixation was recorded in a specific area. A shorter time suggests a higher ability to attract the user's attention.

The values of eye-tracking metrics for various visualization variants are **statistically** compared when assessing the effectiveness of different cartographic visualizations. Based on the author's experience, the measured data do not follow a normal distribution in most cartographic experiments. Therefore, non-parametric tests are used. The most commonly used tests are the **Wilcoxon test** (signed-rank and rank-sum) when comparing two visualizations and the **Kruskal-Wallis test**, a non-parametric equivalent of analysis of variance, for comparing three or more visualizations.

However, a quantitative comparison of eye movement metrics is often not sufficient. When analyzing eye-tracking data using statistical analysis of eye-tracking metrics, the order of fixations in different parts of the stimulus over time is ignored. However, this sequence is a rich source of information about respondent behavior (Anderson et al., 2014). In many cases, the most interesting insights have emerged from **qualitative analysis** of the distribution of fixations within stimuli (within different AOIs). **Sequence charts** (scarf plots) can be used to visualize these sequences of fixations. This visualization technique is available in SMI BeGaze, but it has limitations: it is a static raster image, it cannot display the chart for more than 20 participants, and the size of the X scale is dynamically generated based on the longest sequence. Despite these drawbacks, sequence charts have been effectively used in many studies. However, the issues described above led us to create our own tool for generating interactive vector sequence charts, published in **[paper GazePlotter]**. An example of an interactive sequence chart (scarf plot) generated in GazePlotter is depicted in Figure 16. Each line represents a sequence of fixations for one participant. The color of the segments corresponds to the AOI where the fixations were directed.

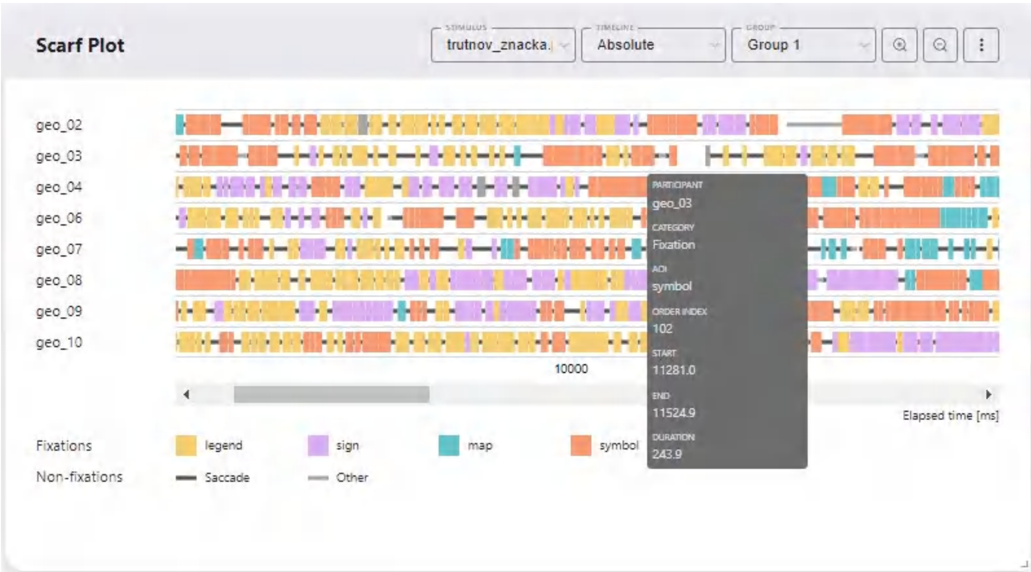


Figure 16. An example of an interactive scarf plot in the GazePlotter application.

Using sequence charts, it is possible to visually compare the similarity of stimuli inspection among multiple participants. The tool called ScanGraph **[paper ScanGraph]** was developed to quantify **scanpath similarity**. ScanGraph works on the principle of the String-Edit-Distance method, which was first used for eye-tracking data by Privitera and Stark (2000) but originates in biology, where it is used for measuring differences between protein sequences.

For the String-Edit-Distance method, AOIs must be defined on the studied stimulus. The scanpath is then replaced by a string of characters representing the sequence of fixations recorded in each AOI. This approach simplifies a complex trajectory into a sequence of characters. ScanGraph offers the calculation of String-Edit-Distance using three algorithms: the Levenshtein distance (Levenshtein, 1966), the Damerau-Levenshtein distance (Damerau, 1964), and the Needleman-Wunsch algorithm (Needleman & Wunsch, 1970). The output of these calculations is a simple graph. The user specifies the parameter as the desired degree of similarity. Groups of scanpaths which are similar according to this parameter are displayed as cliques of this graph. ScanGraph's interface is depicted in Figure 17. Five groups of participants with strategies which were at least 85% similar are displayed.

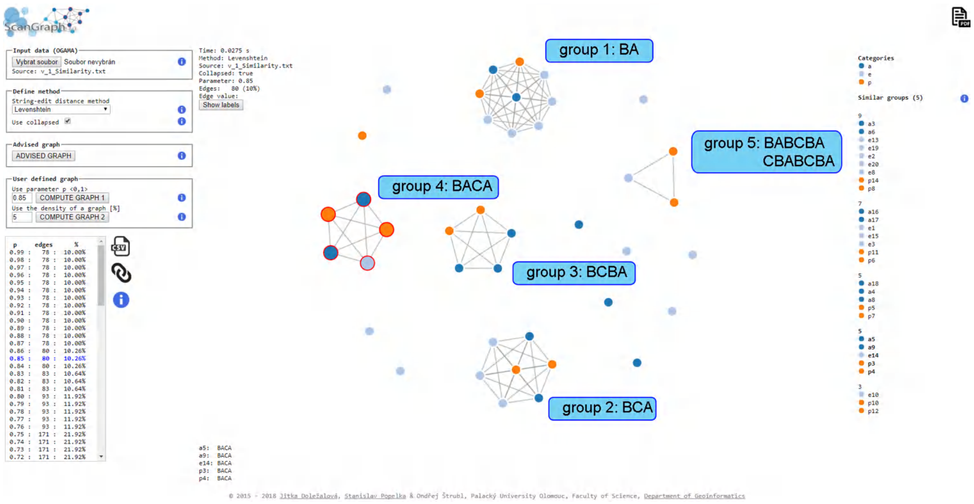


Figure 17. Interface of the ScanGraph software. Figure taken from (Popelka et al., 2018).

4 RESULTS

This chapter presents the **eleven articles** that make up this habilitation thesis, arranged chronologically to follow the thesis narrative (Chapter 1.3). Symbols indicate each article's relevance to specific goals: 🗺️ for evaluating cartographic visualization methods, 🎮 for assessing interactive map interfaces, and 🧰 for the development of eye-tracking tools.

The selection of articles in this habilitation thesis begins in the same way that research in cognitive cartography generally does—with the analysis and comparison of specific visualization methods 🗺️. The articles **[paper 2D-3D]**, **[paper Glyphs]**, and **[paper UrbanPlans]** describe case studies focused on comparing different cartographic visualization techniques.

Building on MacEachren's (1994) Cartography^3 framework, which emphasizes interactivity for effective visual thinking, **[paper WeatherMaps]**, **[paper Swipe-MultipleView]**, and **[paper Dashboards]**, explore the usability and evaluation of interactive map interfaces 🎮.

The development of these case studies highlighted gaps in existing tools for eye-tracking analysis, leading to the creation of new methods and tools 🧰, described in **[paper ScanGraph]**, **[paper EyeTribe]**, **[paper Author-Reader]**, **[paper GazePlotter]**, and **[paper ET2Spatial]**.

Table 1. Overview of papers included in the habilitation thesis.

[paper]	Year	Journal	IF/Q	Contribution
2D-3D 🗺️	2013	Cartographic Journal	0.299; Q4	70%; 1st; corr.
ScanGraph 🧰	2016	Journal of Eye Movement Research	1.295; Q3	50%; corr.
EyeTribe 🧰	2016	Computational Intelligence and Neuroscience	1.215; Q3	25%; 1st; corr.
Glyphs 🗺️	2017	Cartography and Geographic Information Science	1.785; Q2	40%
UrbanPlans 🗺️	2018	ISPRS International Journal of Geo-Information	1.840; Q3	40%
WeatherMaps 🎮	2019	ISPRS International Journal of Geo-Information	2.239; Q3	34%; 1st
Author-Reader 🧰	2022	The Cartographic Journal	1.000; Q4	40%; corr.
Swipe-Multiple 🎮	2022	Cartography and Geographic Information Science	2.500; Q2	50%; 1st
GazePlotter 🧰	2024	ETRA '24	0/0[1]	40%; 1st; corr.
ET2Spatial 🧰	2022	Earth Science Informatics	2.800; Q3	40%; corr.
Dashboards 🎮	2023	Int. Journal of Human–Computer Interaction	3.400; Q2	50%; corr.

In the following sub-chapters, each paper is briefly introduced, highlighting the motivation behind it, the methods employed, and the key findings. Additionally, its role within the overall narrative of the habilitation thesis and its significance for other studies are discussed. Each sub-chapter concludes with a short statement emphasizing the article's most important contribution.

[1] *The methodological paper describing the tool and its functionality is currently under review in Behavior Research Methods (Vojtechovska & Popelka, n.d.). (IF 4.600)*

4.1 Paper 2D-3D (Popelka & Brychtová, 2013)

The first paper of this habilitation thesis was the first paper the author published in a journal with impact factor. It described the first experiment of the author's dissertation thesis, which focused on comparing 2D and 3D visualization using eye-tracking. In this paper, two terrain visualization variants were created – a classical two-dimensional (2D) visualization with contour lines and a perspective (pseudo)3D view. The aim of the article was to prove Haeberling's (2005) assumption that perspective perception of a generalized and symbolized geographic space often offers a better understanding of spatial coherences.

In this paper, two experiments, named DualMap and SingleMap, were conducted so as to determine the most suitable experimental design (Figure 18). The concept behind DualMap was straightforward. The initial plan was to display both visualizations side by side, with the expectation that the experiment would reveal which visualization participants preferred. To eliminate the influence of map positioning (left or right), two groups of users were tested. Both groups were shown the same stimuli, but the positions of the 2D and 3D maps were swapped. In the first group, the 2D map was on the left and the 3D map on the right, while in the second group, the arrangement was reversed. Despite these efforts to address the issue of map positioning, it was found that the presentation order of the maps in the stimuli had a larger impact on the results than the visualization method itself.

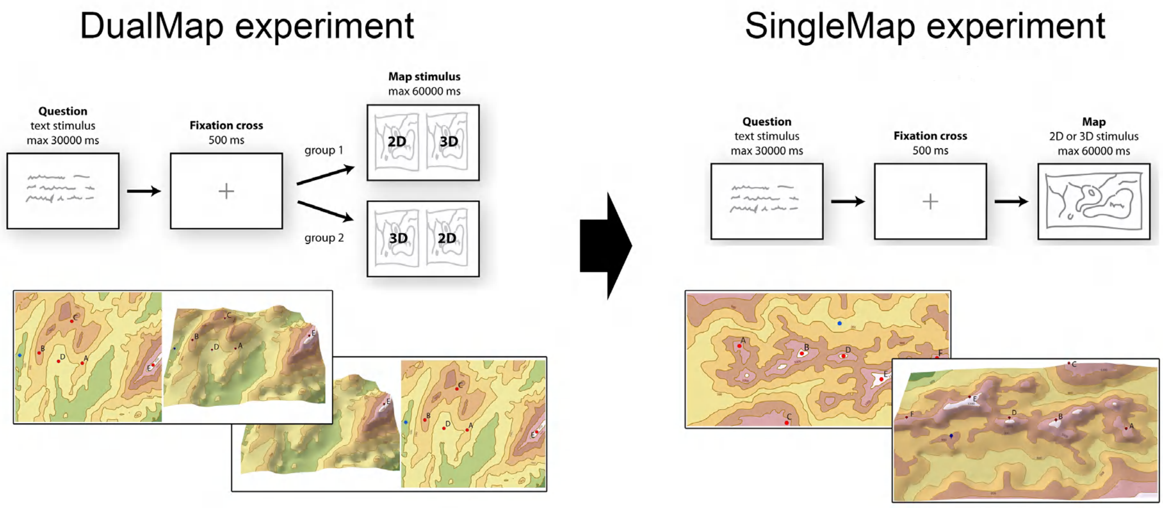


Figure 18. Two types of experiments created in [paper 2D-3D]. Figure modified from Popelka and Brychtová (2013).

In the second experiment, called SingleMap, each stimulus displayed only one map. The stimuli were presented using a within-subject design in a random order, and to mitigate the learning effect, the 3D visualization map was rotated, and the question was slightly modified. This approach was later found to be more effective than presenting stimuli side by side. However, in this experiment, the attempt to eliminate the learning effect introduced a new intervening variable: unequal task difficulty. For example, in the stimuli shown in the bottom-right corner of Figure 18, participants were asked to mark all the red points visible from a blue point. To prevent the learning effect, the map was rotated, but the position of the blue point was also changed, which inadvertently affected the task's difficulty.

The eye-tracking experiment was complemented by a questionnaire that assessed participants' subjective attitudes toward both visualization methods. The questionnaire results indicated a preference for 3D visualization among the participants. The subjective feedback proved valuable, leading to the decision to incorporate subjective questionnaires or interviews in all subsequent case studies. Based on these insights, the possibilities of mixed research design began to be explored. The experience from this experiment later led to the development of the HypOgama application, which combines the Hypothesis application (Šašinka et al., 2017) with eye-tracking **[paper EyeTribe]**.

Subsequent papers expanded on this work by comparing different types of 2D and 3D visualizations. For instance, Popelka (2018b) presented an experiment focused on terrain visualization using hill-shading, while Popelka & Dedkova (2014) and Popelka (2018d) evaluated 3D visualizations of urban areas. Additionally, 3D thematic maps were analyzed in studies by Popelka (2018c) and Popelka and Dolezalova (2016). All these experiments focused on 3D visualizations; they, along with others and those included in this habilitation thesis, benefited from the foundational work of the first paper. This initial study established the experiment design, tested eye movement data recording, and validated methods for data analysis.

Most of the analyses presented in this paper were quantitative, focusing on metrics such as dwell time and scanpath length. In the SingleMap experiment, reliance on the vendor's eye-tracking software was avoided, and the eyePatterns software was used for Scanpath Comparison analysis. This software, specifically designed for eye movement data, had been previously used by Cao and Nishihara (2013) and in the cartographic domain by Coltekin et al. (2010). However, it was later discovered that the tree graph visualization of similarities between participants was improperly integrated into the software, leading to the development of the ScanGraph software.

This article laid the foundation for the design of cartographic eye-tracking experiments and guided the further research direction of the author of this habilitation thesis.

4.2 Paper ScanGraph (Doležalová & Popelka, 2016)

The analysis of eye-tracking data often relies on specialized software tools to quantify and compare fixation patterns and their sequences. It was discovered that the software eyePatterns (West et al., 2006), used in the previous paper, was malfunctioning. This tool calculates scanpath similarity using the String-Edit-Distance method. According to Duchowski et al. (2010), scanpath comparison based on the String Edit Distance introduced by Privitera and Stark (2000) was among the first quantitative methods to compare not only the locations of fixations but also their sequence. This method requires defining areas of interest (AOIs) within the stimulus. The scanpath is then represented as a character string, with each character indicating the AOIs where fixations occur. The distances between these sequences are calculated, and the similarity between two scanpaths is determined by counting the number of transformations (insertions, deletions, and substitutions) needed to change one string into the other (Anderson et al., 2014). Various algorithms for this calculation can be used. However, eyePatterns uses Levenshtein distance (Levenshtein, 1966).

The flaws of eyePatterns lie in its hierarchical clustering process, which affects the accuracy with which sequence similarity can be visualized. Specifically, while the tool correctly starts by clustering sequences with the lowest Levenshtein distance, it then recalculates distances using the average distance of those newly formed clusters. This averaging distorts the actual distances between sequences. As a result, the tree graph visualization can incorrectly represent the similarity between sequences, leading to misleading conclusions. Hence, tree-graph visualization doesn't correspond to the statement used in the eyePatterns interface – "The fewer branches that are between two sequences, the more similar those sequences are".

In the paper, the inaccuracies of eyePatterns' tree graphs were demonstrated using eye movement data recorded with cartographic stimuli. The left part of Figure 19 shows an output from eyePatterns, in which participants are represented as bubbles. The color of each bubble indicates group membership, distinguishing between cartographers and non-cartographers. Participants S1 and S17 are positioned close to each other, suggesting they should have similar scanpaths. However, their Levenshtein distance is 13, indicating significant differences. In contrast, two participants positioned on opposite sides of the tree graph (S12 and S14) have a Levenshtein distance of 4, meaning only four changes are needed to transform one sequence into the other.

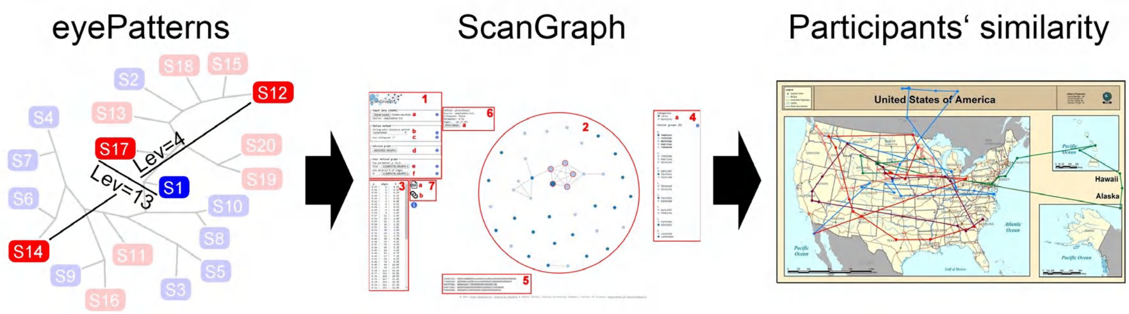


Figure 19. The inaccuracies of eyePatterns led to the development of ScanGraph for the calculation of the similarity of participants' eye movement strategies in [paper ScanGraph]. Figure modified from Doležalová and Popelka (2016).

Upon discovering the inaccuracies in eyePatterns, we decided to develop our own application, called ScanGraph, to identify similar sequences in eye movement data. Unlike eyePatterns, which includes all sequences in a tree graph, ScanGraph focuses only on sequences that are similar based on predefined parameter (desired minimal similarity). The results are displayed in a simple graph, with similar groups shown as cliques within this graph.

ScanGraph was developed using PHP, C#, and D3.js, and it can directly read data from the open-source eye-tracking application OGAMA (Voßkühler et al., 2008). The similarity calculation algorithms were modified to handle strings of different lengths more effectively. The functionality of ScanGraph was demonstrated using a simple cartographic case study. An example which identifies participants with similar behavior is shown in the right part of Figure 19. This figure displays five participants forming two cliques based on a similarity parameter of 0.75. The scanpaths of the three participants are shown in shades of red, while the blue and green participants are both similar to the red group but not to each other.

The functionality of ScanGraph was later enhanced by incorporating a new algorithm for String-Edit-Distance calculations and adding the ability to calculate the similarity between participants based on multiple files, each representing different stimuli (Popelka et al., 2018).

ScanGraph has been extensively applied in numerous studies co-authored by the author of this habilitation thesis (Brus et al., 2019; Dolezalova & Popelka, 2016; Popelka & Beitlova, 2022; Skrabankova et al., 2020). Additionally, it has been utilized in research conducted by other cartography experts, such as Havelková and Gołębiowska (2020) and Havelková and Hanus (2019). Furthermore, ScanGraph has been employed to visualize differences in eye movement strategies between individuals with autism and a control group as shown in studies by Eraslan et al. (2017, 2019).

In this habilitation thesis, ScanGraph was used to find similarities between participants as they worked with weather forecast maps (Popelka, Vondrakova, et al., 2019) **[paper WeatherMaps]**. Its functionality was later modified in Beitlova et al. (2022) **[paper Author-Reader]** to be able to compare the strategies of one participant (map author) versus multiple participants (map readers). This approach was later used to compare map reading strategies between students and their geography teacher (Beitlova et al., 2020).

The article describes the tool developed for scanpath comparison, which has been used in many other papers, not only by the author of this habilitation thesis and not only in cartography.

4.3 Paper EyeTribe (Popelka et al., 2016)

The impulse for this publication was driven by two main factors. The first factor was the emergence of a new, affordable eye-tracking device called the EyeTribe in 2014. The second factor was insights gained from the initial case study (**[paper 2D-3D]**), which highlighted the advantages of integrating qualitative and quantitative data. These insights led us to develop a tool called HypOgama, which combines a quantitative data acquisition platform named Hypothesis with eye-tracking technology.

Eye-trackers are typically expensive, so laboratories usually have only one device, limiting the number of participants that can be recorded individually. However, in 2014, a new, affordable eye-tracker called EyeTribe became available for just 99 USD, presenting an opportunity to record eye movement data from large populations. Our goal was to compare the accuracy of the EyeTribe's data with that of the SMI RED 250. If the data quality proved suitable, we planned to purchase several EyeTribe devices to record data from multiple participants simultaneously. To compare the two devices, we conducted concurrent recordings of the same participants using both devices. We prepared an experiment with six static images in the OGAMA software, connected to the EyeTribe tracker, while simultaneously recording ScreenRecording stimuli with the SMI RED 250. To ensure a successful comparison, we synchronized both datasets and divided the SMI data into segments corresponding to the stimuli in OGAMA. This process is illustrated in Figure 20 (left).

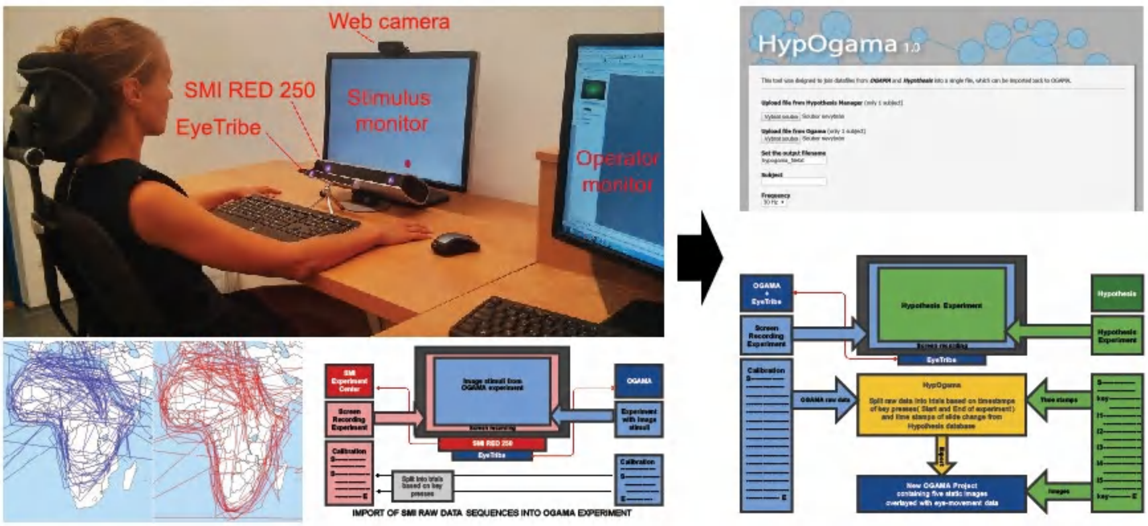


Figure 20. Concurrent eye movement recording for the comparison of the EyeTribe and SMI (left) and a similar approach used for developing the HypOgama application (right). Figure modified from Popelka et al. (2016).

The accuracy of the recordings from both devices was comparable, with the largest deviations observed in the middle-bottom area of the screen with the EyeTribe. The results indicated that the EyeTribe could be a valuable tool for cognitive cartography experiments and assessing user behavior during map reading. The author of this habilitation thesis also collaborated on another paper that compared the EyeTribe with the SMI. This paper reached similar conclusions, stating that, when used correctly, the EyeTribe tracker is a valuable tool for academic research (Ooms, Dupont, et al., 2015). Unfortunately, after Oculus (Facebook) acquired the EyeTribe company at the end of 2016, the devices were discontinued, and therefore, they could not be used in further studies.

The experiences gained from synchronizing and splitting datasets were instrumental in developing the HypOgama application, which acts as a bridge between the eye-tracker and the Hypothesis platform. Hypothesis is a web-based platform tailored for large-scale data collection, especially in psychological testing and the evaluation of cartographic works (Morong & Šašinka, 2014; Šašinka et al., 2017). It supports adaptive testing and modular plugins, facilitating extensive logging of user actions and event data in a controlled and secure environment. The HypOgama tool automates the synchronization and processing of data from the Hypothesis platform and the EyeTribe eye-tracker (or SMI eye-tracker) connected to the OGAMA eye-tracking software. It synchronizes timestamps from both systems using a key press, divides the recorded data into blocks corresponding to specific slides from the Hypothesis experiment, and prepares the data for direct import into OGAMA, effectively integrating eye-tracking and quantitative data.

Although the EyeTribe trackers were not used in praxis, the HypOgama application was utilized multiple times to combine eye-tracking data with data gathered using Hypothesis (Šašinka et al., 2021; Šašinka et al., 2019). Since HypOgama outputs an OGAMA project, the GazePlotter application **[paper GazePlotter]** could be employed to visualize data from the EyeTribe using sequence charts.

This paper compares the accuracy of professional and low-cost eye-tracking devices. Additionally, it describes the development of the HypOgama application, which was later used in numerous studies.

4.4 Paper Glyphs (Opach et al., 2017)

The next case study compared two types of glyphs—small geometric shapes commonly used in geovisualization to represent multidimensional spatial data. The two evaluated types, star glyphs and polyline glyphs, can convey the same information and offer similar functionality. However, it is unclear which type is more effective for participants in various tasks. To explore this, an empirical study was conducted to examine differences in user performance between polyline and star glyphs, displayed either in a grid plot or on a map.

Glyphs are a popular visualization technique which provide an overview of a dataset by representing its items as distinct, simplified graphical entities. These are displayed as small plots or charts placed side by side, typically without labels (Borgo et al., 2013; Ünlü & Malik, 2011). Star glyphs are one of the most commonly used types of glyphs. Despite their widespread implementation in many geovisualization environments (Gribov et al., 2006; Takatsuka & Gahegan, 2002), star glyphs have limitations due to their use of polar coordinates. Visual scanning in polar coordinates can be more time-consuming and error-prone compared to reading vertical and horizontal axes (Goldberg & Helfman, 2011). Consequently, Opach and Rød (2018) suggest using polyline glyphs, which resemble polylines from parallel coordinates, as an alternative to star glyphs.

During his internship abroad in Trondheim, Norway, the author of this habilitation was contacted by Opach and Rød. They were interested in exploring the use of eye-tracking to evaluate the usability of two types of glyphs and their display formats (either in maps or grids). This study aimed to answer three research questions: (RQ1) Are there specific tasks that are better suited to one of the glyph types? (RQ2) Are there specific tasks that are better suited to a particular display type? (RQ3) Does the cognitive style of the participants influence their performance?

The experiment was designed based on insights from an initial case study on 3D visualization **[paper 2D-3D]** and involved 26 participants. The study used a single-page web application accessible through a standard web browser, which presented data using four layout modes with either polyline or star glyphs arranged in a grid or on a map (Figure 21, left). The study was conducted in two phases, spaced at least three days apart, to mitigate learning effects. Two task order variants were used: Variant 1 featured a grid plot with polyline glyphs followed by a map with star glyphs, while Variant 2 reversed this order (Figure 21, right). Participants alternated between these variants across the two phases to ensure they experienced all layout modes and glyph types.

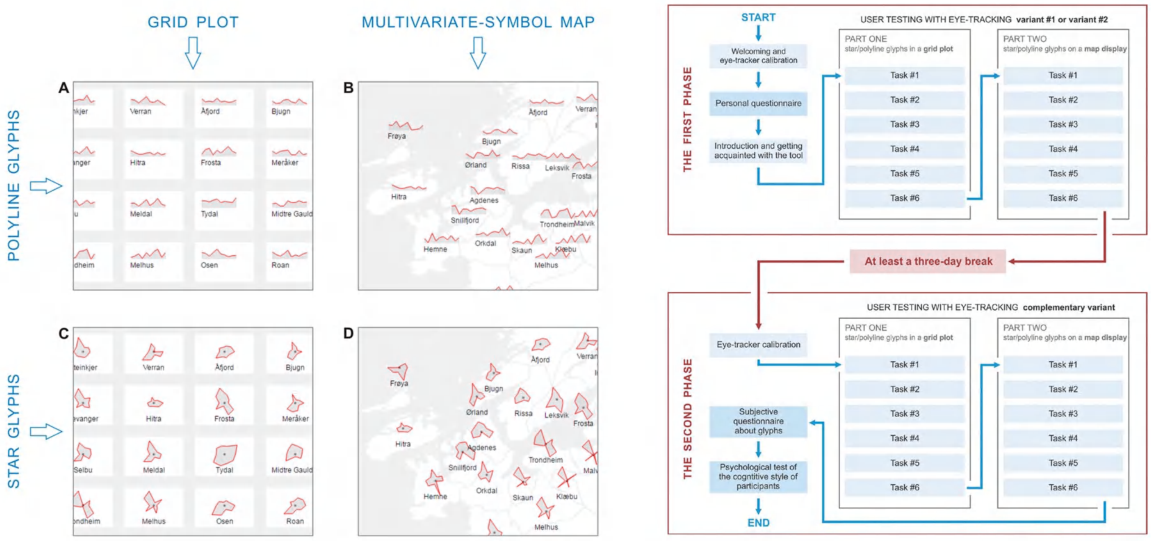


Figure 21. Stimulus types (left) and the design of the experiment (right) used in [paper Glyphs]. Figure modified from Opach et al. (2017).

In each phase, participants completed six tasks designed to assess different aspects of glyph usability, such as estimating values, identifying specific glyphs, finding similar or distinct glyphs, and identifying compact areas of similar glyphs. After the tasks, participants completed two questionnaires: a personal one to collect demographic information and a subjective one to capture their preferences regarding the two glyph types. Additionally, participants took a psychological test administered via the Hypothesis platform (Morong & Šašinka, 2014; Šašinka et al., 2017) based on Navon's cognitive style test (Navon, 1977) to determine whether their cognitive style—holistic or analytic—affected their performance with the different glyph types and layouts. For data analysis, the eye movement metrics trial duration, fixation counts, fixation duration, scanpath length, and revisit counts were analyzed. Moreover, the fixation counts recorded for AOIs around individual glyphs were visualized.

The study investigated three key research questions. For RQ1 (glyph comparison), star glyphs generally outperformed polyline glyphs, particularly in tasks requiring comparison and identification, achieving higher answer accuracy and faster completion times. However, polyline glyphs were preferred for tasks involving value estimation due to their linear arrangement, which facilitated interpretation. For RQ2 (display comparison), the map display proved more effective and efficient than the grid plot for most tasks, with participants performing better in finding similar or distinctive glyphs on maps. The geographical context provided by maps often aided in visual search tasks. RQ3 (cognitive style influence) revealed that analytic users had higher accuracy in tasks requiring detailed analysis, while holistic users were faster overall. However, the differences were not statistically significant across all tasks. The findings suggest that task type, display format, and individual cognitive preferences have strong effects on the usability of different glyph types and layouts. Additionally, the observations on the experimental methodology involving various groups of participants were utilized in the subsequent studies **[paper UrbanPlans]**, **[paper Author-Reader]**.

The main findings of this paper offer valuable insights into the effectiveness of different glyph types, their displays, and their usability by analytic and holistic users.

4.5 Paper UrbanPlans (Burian et al., 2018)

Urban plans are essential for effective spatial planning and development, as they visually represent the intended use and organization of space (Monmonier, 1996). Despite the importance of clear and precise cartographic visualization in urban planning, there has been limited research on standardizing map symbols and the overall quality of urban plan cartography (Dühr, 2004, 2007). This lack of standardization forces urban planners to make subjective decisions about the visual aspects of these plans, which can lead to inconsistent levels of clarity and accuracy (Burian et al., 2016). Consequently, the cartographic quality of urban plans—an important factor influencing user interpretation and decision-making—remains a critical yet underexplored area in urban planning research.

This case study examined the cartographic quality of urban plans in the Czech Republic, focusing on various cartographic styles. The research aimed to test three main hypotheses using the eye-tracking method: (1) that map symbology affects legibility and understanding, (2) that the clarity of the legend influences comprehension and interaction, (3) and that students and experts read plans differently, impacting task accuracy and duration.

To evaluate the usability and effectiveness of the urban plans, six tasks were designed to represent typical actions users might perform with urban plans. These tasks followed the simplest category from the Wehrend and Lewis (1990) objective-based taxonomy, the "identify" category. The tasks required participants to identify point, line, and polygon features on the maps, covering a broad range of cartographic symbols. Participants were asked to mark areas for housing, sports or recreation, proposed public services, railroads, wastewater treatment plants, and protected areas of water resources. The urban plans of four Czech cities (Figure 22, top) were presented as static images at a consistent scale of 1:5000. Each task included a legend that could be scrolled while keeping the map static, ensuring a uniform experience across all tasks and respondents. At the experiment's conclusion, a short questionnaire gathered respondents' subjective opinions on the plans. The experiment involved 34 participants, but eight were excluded due to calibration issues. This left 26 respondents: 20 students, chosen for their comparable level of knowledge about urban planning and cartography, and six experts from urban planning departments in Olomouc. These experts, working daily with urban plans, had similar backgrounds, ensuring consistency in experience and skills.

The recorded data were analyzed quantitatively, focusing on trial duration and fixation count metrics. This analysis identified task 5 for the urban plan of Jihlava and task 6 for the urban plan of Olomouc as the most challenging for participants. In the next phase, areas of interest were defined around the map and legend in the stimuli, and the number of fixations and dwell time in these AOIs were examined. Additionally, inspection strategies for the stimuli were visualized using sequence charts (Figure 22, middle). These methods revealed different reasons for the high trial durations and fixation counts in these two tasks. For Jihlava, participants could locate the symbol for the water treatment station in the legend but struggled to find it on the map due to inconsistencies between the map and the legend. Conversely, in task 6, participants had difficulty identifying the protected area of the water resource in the complex legend of the Olomouc urban plan. The data were also visualized using FlowMap method available in V-Analytics software (Figure 22, bottom).

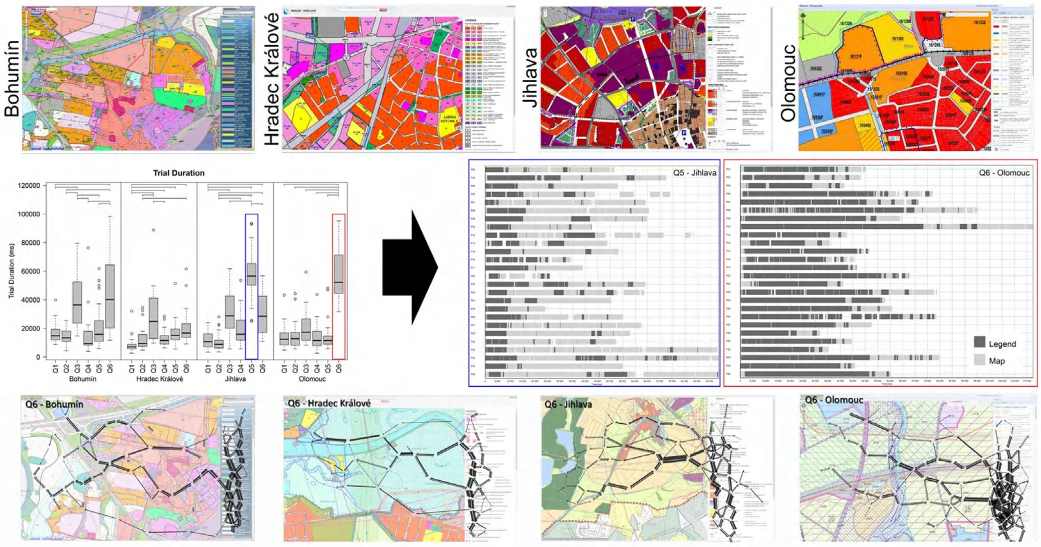


Figure 22. Types of urban plans used in [paper UrbanPlans] (top); the highest trial duration values explained by sequence charts (middle); the visualization of eye movements using the FlowMap method (bottom). Figure modified from Burian et al. (2018).

The study tested three hypotheses regarding map symbology, legend structure, and differences between students and experts on urban plan readability and usability. The analysis confirmed that map symbology significantly affects legibility and task accuracy, with the Jihlava plan being the most challenging due to numerous and inconsistent symbols. Legend structure was also crucial, as respondents often spent more time on the legend than the map, especially with unstructured legends. Differences between students and experts showed that experts, despite fewer fixations on common elements, made more errors due to overconfidence and familiarity with different styles. These findings highlight the need for standardization in urban planning to improve clarity and prevent misunderstandings.

The success of using sequence charts in this study to understand participants' strategies for inspecting stimuli led us to plan their use in a future case study on weather maps **[paper WeatherMaps]**. Additionally, the observed differences between experts and students prompted us to focus our next research study on the differences in map reading between map authors and map readers **[paper Author-Reader]**.

The empirical study described in this paper serves as an argument for the standardization of urban planning design.

4.6 Paper WeatherMaps (Popelka, Vondrakova, et al., 2019) 🎮

The widespread use of the internet and mobile technologies has led to a surge in the popularity of interactive maps among the general public (A. Mendonça & Delazari, 2014). Peterson (1998, p. 3) presciently observed that "The incorporation of interaction in the display of maps may be viewed as a major accomplishment of the computer era in cartography." These maps have revolutionized the production and distribution of spatial information (Sack & Roth, 2017) by enabling users to explore data at various scales and easily analyze it using features like filtering and layering (Roth, 2012). The primary advantage of interactive maps lies in their ability to be customized to meet the specific needs and preferences of users (MacEachren, 1994).

This interactivity requires digital cartographic visualizations to be clear and intuitive, demanding a deeper understanding from cartographers of design and user interaction. Despite these advancements, there is still a lack of comprehensive guidelines for the design of web maps (Cartwright et al., 2001; A. L. A. Mendonça & Delazari, 2012; Nivala, 2007). The rapid advancement of information technology highlights the need to ensure good usability, a positive user experience, and strong user orientation in these applications (Hennig & Vogler, 2016).

Focusing on this need, we conducted a study on interactive weather maps. The main reason for selecting weather maps was that these maps are considered complex visual displays (Hegarty et al., 2010). The primary objective of this study was to analyze and evaluate web-based weather maps in terms of their level of interactivity and user perception.

Five interactive weather maps were selected for this formative study. The selected maps included popular platforms such as Windy, DarkSky, In-Počasí, YR.no, and Wundermap, each offering varying levels of thematic detail. The study comprised both static and dynamic stimuli. In the static segment, participants worked with screenshots of the maps, while in the dynamic segment, they actively interacted with the maps. The dynamic tasks required participants to switch layers, search, and navigate, while the static section focused on assessing participants' understanding of the map layout without interaction. A think-aloud protocol was used alongside eye-tracking to better understand user behavior and interaction. The study included 34 participants, categorized as novices and experts based on their backgrounds and previous experience with web maps. The appropriate number of participants for the study was determined using the MeasuringU tool (Sauro, 2023).

In addition to trial duration, scanpath length, and fixation count eye movement metrics, the number of fixations were recorded in a regular grid overlaying the stimuli in the static segment. As in the previous study focusing on urban plans **[paper UrbanPlans]**, the flow map method was used to visualize the aggregated eye movements of all participants (Figure 23, upper left). One of the contributions of the study was the successful utilization of the ScanGraph application **[paper ScanGraph].** ScanGraph was used to analyze the strategies participants employed to inspect the stimuli, focusing on those who visited areas of interest (AOIs) in the same order. The analysis showed that there was generally low consistency in the strategies used by participants across different maps, with few respondents following the same approach (Figure 23, right). However, in some cases, clearer patterns emerged, indicating that certain maps guided users more effectively toward the correct answers. Overall, the ScanGraph analysis was extremely useful in highlighting the variability in user approaches, demonstrating that map design significantly influences how users interact with and interpret map information. The effectiveness of the tool was verified, leading to the decision to apply it in a more advanced manner for comparing map reading strategies between map authors and map readers in **[paper Author-Reader]**. Building on the successful use of sequence chart visualization in the previous case study on urban plans, sequence charts were applied again. Unlike the previous case, this study involved dynamic stimuli—interactive maps—requiring the marking of dynamic AOIs within the stimuli and manual modification of the resulting visualization in a graphic editor (Figure 23, lower left). This experience sparked the initial idea to develop our own tool for sequence chart visualization, which ultimately led to the creation of GazePlotter **[paper GazePlotter]**. Furthermore, this study provided early support for the idea of using the think-aloud protocol more thoroughly in cartographic research. The think-aloud protocol was later elaborated upon in a methodological paper by Vanicek and Popelka (2023).

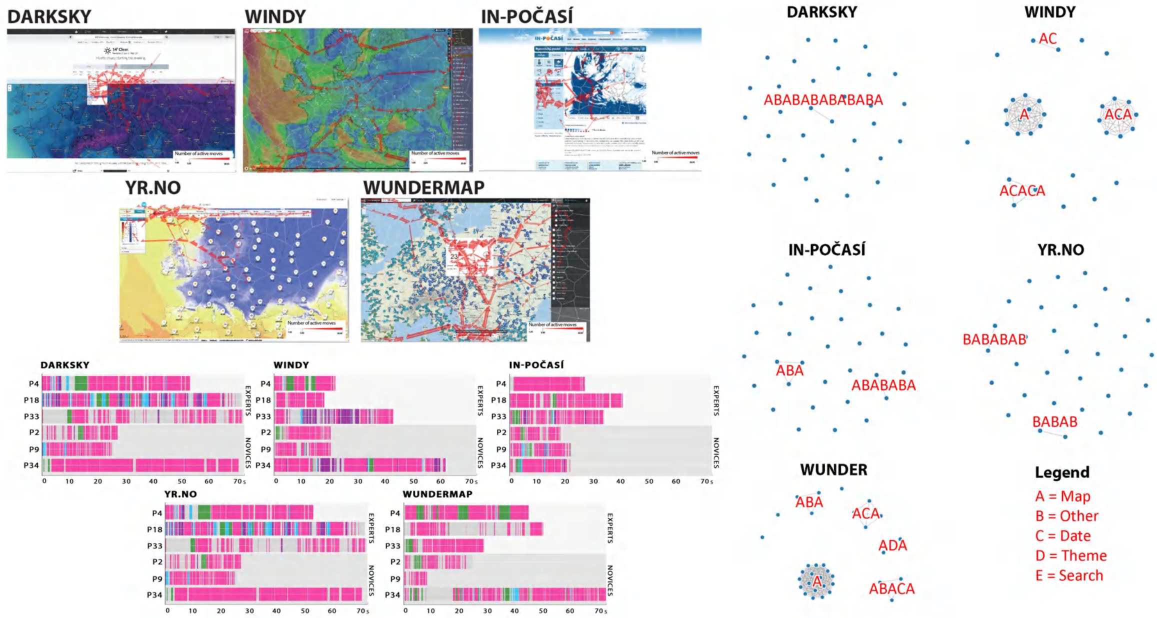


Figure 23. Stimuli from the study [paper WeatherMaps] overlaid by FlowMaps (upper left); sequence charts created for dynamic AOIs (lower left); the output of ScanGraph (right). Figure modified from Popelka, Vondrakova, et al. (2019).

This study found that users interacted with weather web maps in a straightforward manner, primarily focusing on the main screen controls and avoiding hidden or advanced features. Interactive elements were generally explored only after users became familiar with the map's basic layout. For novices, the main interest lies in utilizing the map's content, such as finding specific locations, rather than exploring functionalities. In contrast, experts delved into advanced features like thematic layers and analytical tools. The evaluation also revealed that users initially struggled with complex map compositions, especially when controls were scattered around the map. The assessment of user-friendliness indicated that respondents valued functionality, simplicity, and fast loading times over modern design aesthetics.

This paper evaluated the user experience with weather forecast maps. Additionally, the functionality of the ScanGraph application was demonstrated in practice.

4.7 Paper Author-Reader (Beitlova et al., 2022)

As outlined in chapter 2.1, cartographic communication models provide a theoretical foundation for studying the process of map reading. These models often focus on the relationship between the author and the user of the map, as well as the interaction between the map and its user. However, prior research has not thoroughly explored how map authors read their own maps and whether their approach is similar to that of typical map users. In this experiment, map authors were put in the role of readers, and their map-reading strategies were analyzed. The assumption was that the authors, being (or having been at the time of map creation) well-acquainted with the depicted phenomena, would have a straightforward map-reading strategy. The experiment also compared the authors' strategies with those of cartographers (participants with a cartographic education) and novices who were not familiar with the data and methods used in the maps. Thus, the experiment aimed to assess the degree to which the perspectives of the map authors and the map readers align, as represented by U_1 and U_2 in Koláčný's (1969) model (Figure 2) and A and B in the Venn diagram (Figure 3) by Robinson and Petchenik (1976). Additionally, it examined the map reading strategies by which this overlap is achieved.

This study aimed to verify cartographic communication models and explore differences in map reading strategies among three groups: map authors, cartography students, and novices. A key contribution of this paper was the introduction of a new method to calculate the similarity of participants' map reading strategies.

The study used eye-tracking technology to examine map-reading strategies across three groups: map authors, cartography students (referred to as cartographers), and novices. The map authors were second-year university students who created maps as part of their cartography courses. These students were familiar with the data and design decisions involved in creating these maps. In the experiment, each map had a specific author who also participated as a reader, allowing a comparison between their roles as authors and as readers. The experiment was conducted in two phases. The first phase involved 22 participants, including the map authors and other cartography students. The second phase compared the strategies of these cartographers with 17 novices who had no specific experience in cartography. Participants interacted with 44 maps in total. During the free viewing phase, they inspected 22 maps without specific tasks, allowing researchers to observe their natural viewing patterns. These maps were reproductions from various atlases created during general cartography classes. In the task completion phase, participants worked with another set of 22 thematic maps created as part of thematic cartography coursework. Each map had a designated author among the participants who had created it during their studies. When viewing their own maps, these students acted as map authors, while for other maps, they participated as readers with general cartographic knowledge but without specific familiarity with the data (Figure 24, left).

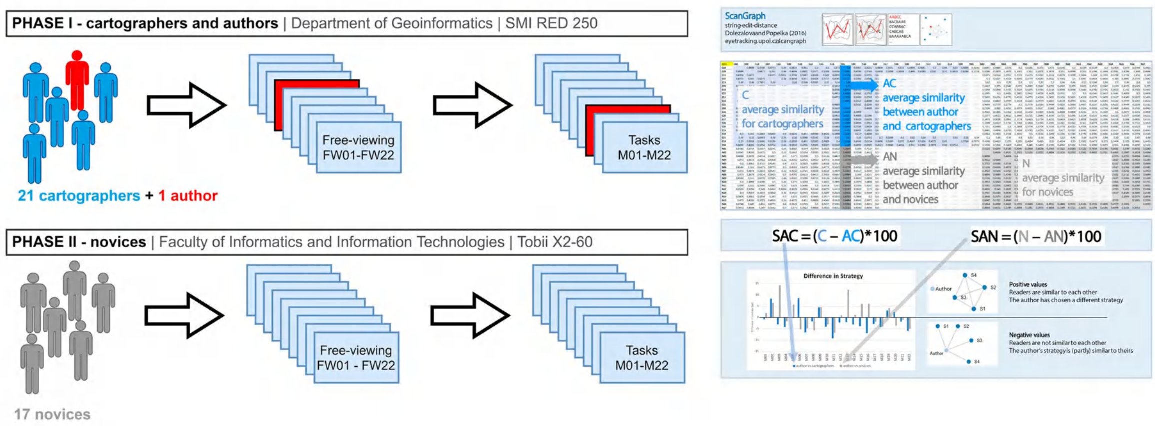


Figure 24. The summary of the design of the study (left) and the overview of the scanpath comparison calculation (right). Figure modified from Beitlova et al. (2022).

The study utilized both qualitative and quantitative methods to analyze the recorded data. For the task-completion phase, the accuracy of participants' answers was evaluated, and problematic tasks were further investigated using scanpath inspection and sequence chart visualization. The metrics trial duration, fixation count, and dwell time were statistically analyzed. The map reading strategies of participants were compared using ScanGraph **[paper ScanGraph]**. The study did not set a minimum similarity level (parameter p) for strategies but instead used modified matrices to compare the average similarity of strategies between the author and cartographers and between the author and novices. This approach identified cases where the author's strategy differed from that of other readers.

This experiment confirmed the hypothesis that there would be differences in the accuracy of answers, trial duration, and eye-tracking metrics between map authors and other participant groups. While map authors made no mistakes, cartographers and novices showed lower accuracy, with novices requiring more fixations during tasks but fewer during free viewing. Additionally, a new method for quantifying differences in map reading strategies was successfully tested, revealing unique strategies among map readers and highlighting situations where these differed significantly from the authors' strategies.

This method was later improved by Popelka and Beitlova (2022) by incorporating the Multimatch algorithm (Dewhurst et al., 2012) instead of using String-Edit-Distance. This enhancement makes it more suitable for analyzing stimuli where defining AOIs is challenging. The improved method was successfully applied in a study comparing the map reading strategies of students and their geography teacher (Beitlova et al., 2020).

This article won the Henry Johns Award for the Most Outstanding Paper of 2023. This award is given by the Editorial Board of The Cartographic Journal to a paper which makes a significant empirical contribution toward understanding map reading and invites fresh appraisals of the theoretical framework of cartographic communication.

4.8 Paper Swipe-Multiple (Popelka et al., 2022) 🎮

This study builds upon the previous one which used weather maps **[paper WeatherMaps]** and is focused on the interactive methods that allow the comparison of maps. Map comparison is a fundamental method that geographers apply to understand the world. The goal of comparison is to "enhance the likelihood that an analyst will see not only features but the relationships between features" (MacEachren, 2004, p. 401). The quantification of spatial distributions and patterns and comparison across regions or over time is central to many types of geographical research and application (Long & Robertson, 2018).

Map comparison techniques are based on two main principles: juxtaposition and superimposition (Gleicher, 2017). Juxtaposition involves displaying different representations of data in separate, non-overlapping windows. The most straightforward form of this technique is called multiple view (also known as juxtaposition or window juxtaposition), where two or more maps are placed side by side. These maps are synchronized, so any change in the coordinates on one map is mirrored on the other. This method is advantageous because it allows simultaneous consideration of two situations by shifting focus between the maps. However, it requires users to perceive the maps as complete images to detect changes in spatial patterns or object characteristics. A drawback of this method is that it can divide the user's attention (Harrison et al., 1995). On the other hand, superimposition involves overlaying layers and using various techniques to compare them. In this case, the divided attention of the user is not a problem; visual interference, however, presents difficulties. One of the techniques employed for map comparison of superimposed maps is swipe, which allows users to drag one map over another (Lobo et al., 2015).

This paper aimed to evaluate and compare user behavior when using two map interaction methods—multiple view and swipe—in the Esri environment, as illustrated in the upper left part of Figure 25. The study involved 25 participants who performed nine tasks using both visualization types. Pre-loaded land suitability maps created with Urban Planner software (Burian, Stachova, et al., 2018) within the Esri ArcGIS environment were used for the study. The number of suitability layers varied, with the first two tasks involving two layers and the last two involving four layers. Participants were tasked with identifying areas of highest suitability for different uses (Figure 25, upper right).

Unlike with static stimuli such as figures, the analysis of dynamic stimuli like interactive maps requires the marking of areas of interest (AOIs) both spatially and temporally. Although this process is time-consuming, it enables a detailed analysis of participant behavior, including metrics such as the duration each AOI was displayed and the time participants spent viewing each AOI. In a previous case study on weather maps **[paper WeatherMaps]**, sequence charts were introduced to visualize eye movement data collected from dynamic stimuli, proving their effectiveness. Therefore, sequence charts were used again in this study. However, because SMI BeGaze only exports raster sequence charts for individual participants, the sequence charts had to be manually created using graphical software (Figure 25, bottom). This process was very demanding, but the sequence charts offered valuable insights into participant behavior. This experience prompted us to develop an automated solution for generating sequence charts, now called GazePlotter **[paper GazePlotter]**.

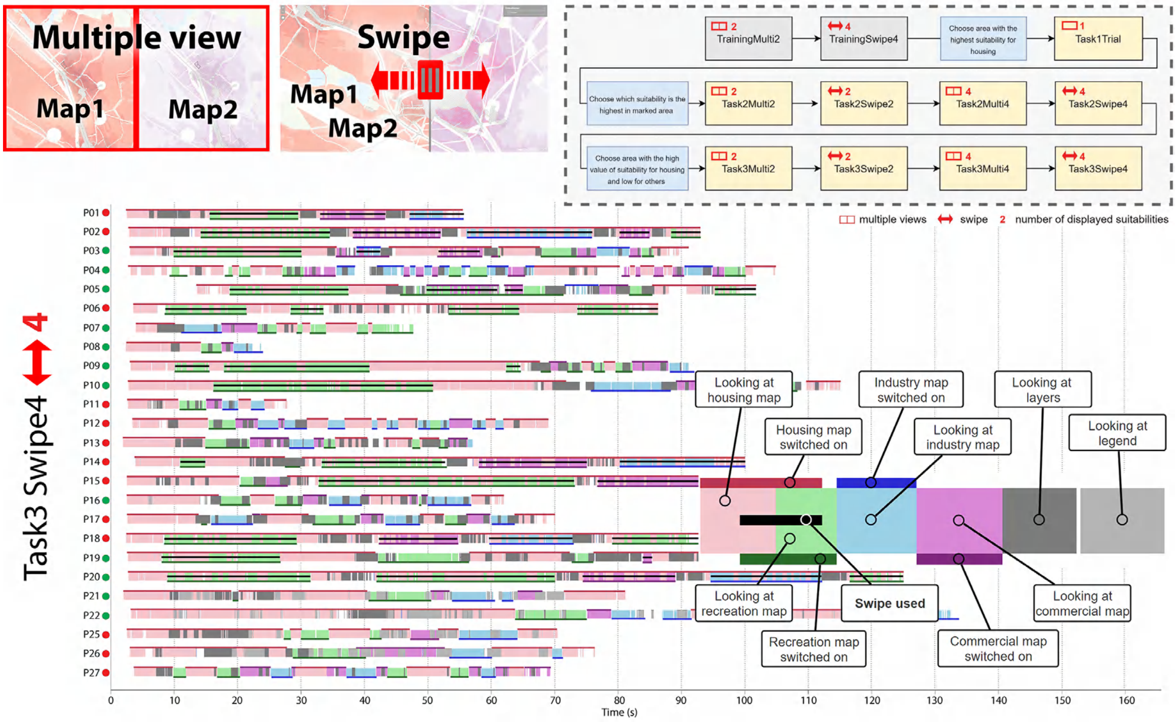


Figure 25. Two types of visualizations used in [paper Swipe-Multiple] (upper left); tasks used in the study (upper right); a sequence chart created in a graphic editor. Figure modified from Popelka et al. (2022).

The results showed that multiple views were generally more effective and intuitive, particularly for comparing four maps, as it required no additional settings and displayed all layers by default. In contrast, swipe was less intuitive, needing complex adjustments, especially for more than two maps, and did not allow simultaneous legend viewing. The most notable limitation was the informational inequivalence between methods; swipe required layer selection while multiple views did not, largely due to the manner of implementation in Esri's environment rather than the swipe method itself. The study highlights the need for improved swipe functionality and clearer overlay definitions in the Esri platform.

This paper reveals significant insights into the usability of swipe and multiple views methods, highlighting shortcomings in the swipe user interface provided by Esri.

4.9 Paper GazePlotter (Popelka, Kominek, et al., 2024) 🧰

As gaze patterns can reflect cognitive processes, preferences, and mental states, understanding how individuals visually explore stimuli has significant implications (Carter & Luke, 2020). The qualitative analysis of scanpaths and areas of interest provides important advantages. By examining the sequence in which different areas are viewed, this approach reveals how attention is distributed over time across a stimulus. It can uncover subtle patterns of engagement or disengagement that quantitative metrics may miss, offering deeper insights into participants' cognitive behavior. The benefits of the use of sequence charts (scarf plots) were demonstrated in the study **[paper UrbanPlans]** for static stimuli and later in the study **[paper WeatherMaps]** for the visualization of sequences of visits of dynamic AOIs. Finally, sequence charts were the main visualization method used in **[paper Swipe-MultipleView]**.

However, the possibility of generating sequence charts in the proprietary software offered by eye-tracker manufacturers is limited. Software such as Tobii Pro Lab and GazePoint Analysis lack this option. Only the now-discontinued SMI BeGaze allowed for their generation, but with caveats: SMI BeGaze allows the generation of sequence charts for a maximum of 20 participants, and its output is a raster image. However, when using dynamic AOIs, it is possible to use SMI BeGaze to generate a sequence chart for a single participant. In addition, there are some open-source applications capable of generating sequence charts, like SEQIT (Wu & Munzner, 2015), AlpScarf (Yang & Wacharamanotham, 2018), and GazeAlytics (Chen et al., 2023). However, these applications do not support the indication of AOI visibility, so they cannot be used to visualize dynamic AOI sequences.

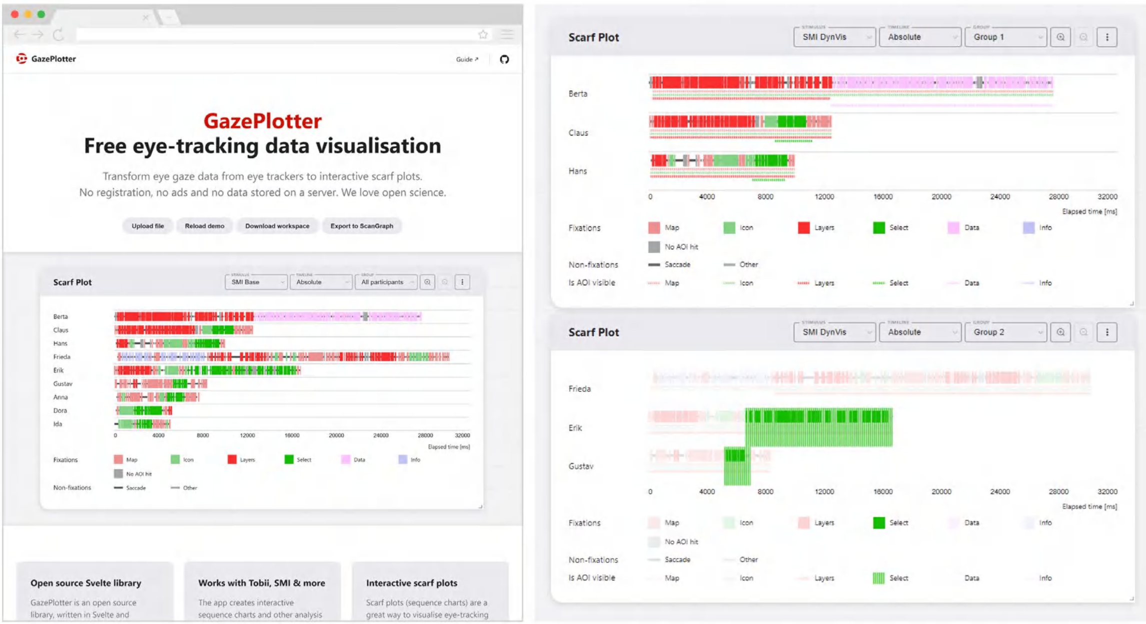


Figure 26. The interface of the GazePlotter application (left). Indication of AOI visibility using narrow lines (top right) and interactive highlighting (bottom right). Figure modified from (Vojtechovska & Popelka, n.d.).

Feeling the absence of a tool that would allow the simple generation of sequence charts from dynamic AOIs, we decided to develop our own solution, known as GazePlotter (Figure 26, left). This tool allows users to visualize data from multiple eye-trackers without needing further transformations and can be used online without registration or cost. It supports dynamic AOIs (Figure 26, top right), interaction with the output (Figure 26, bottom right), and offers advanced visualization customization options, ensures privacy through client-based processing, and is open-source to encourage collaboration.

GazePlotter was developed as an open-source progressive web app (PWA) designed to visualize eye-tracking data using interactive and adaptive scarf plots. Multiple software exports and CSV formats are supported, and preloaded demo data is provided for ease of use. Eye-tracking researchers were included throughout the development process, and extensive unit and cross-browser testing was conducted with real data. The app was iteratively refined based on user feedback, with a transition to a SvelteKit project with an automated GitHub CI/CD pipeline for improved feature development and maintainability.

The functionality of GazePlotter was demonstrated by an analysis of eye movement data collected from geological maps. The study utilized scanned maps, a paper map, and a web map application to compare the map-reading strategies of geologists, geographers, and geoinformatics professionals. The results showed that geologists were the most efficient, completing tasks faster and with fewer fixations due to their familiarity with geological symbols and features. Additionally, the study revealed usability challenges in the Czech Geological Survey's online map application, particularly with navigating layers and functions, highlighting the need for design improvements. These findings underscore the influence of professional background on map-reading strategies and suggest the necessity of tailored map designs for different user groups.

GazePlotter was instrumental in visualizing the eye-tracking data in this study, enabling the researchers to analyze fixation sequences and attention patterns across map elements. Three different types of datasets were visualized using GazePlotter: static stimuli from scanned maps, dynamic data from web map application testing, and data from eye-tracking glasses, converted into a CSV format. GazePlotter's versatility was further demonstrated in other studies, such as in the visualization of Tobii Pro Glasses 3 data from a science center (Popelka & Vysloužil, 2024). Moreover, it was used to visualize SMI eye-tracking data recorded during interactions with COVID-19 dashboards (Porti Suarez & Popelka, 2023) **[paper Dashboards]**.

This paper describes the possibilities of an open-source tool for scarf plot visualization of eye-movement data using geological maps as an example. However, the tool can be easily applied in any field of research.

** The methodological paper describing the tool and its functionality in detail is currently under review in Behavior Research Methods (Vojtechovska & Popelka, n.d.).*

4.10 Paper ET2Spatial (Sultan et al., 2022)

The development of GazePlotter [paper GazePlotter] facilitated the effective analysis of eye-tracking data from interactive maps with dynamic AOIs. However, creating these dynamic AOIs remains a time-consuming task. Tools like the assisted mapping function in Tobii Pro Lab can help streamline this process. This function uses automatic image recognition to align gaze data with snapshot images, which are typically photographs of real-world settings or screenshots of screen-based content. After the use of assisted mapping, it is possible to generate visualizations, calculate AOI-based metrics, or visualize sequence charts as if the stimuli were static.

Unfortunately, this method for the creation of dynamic AOIs is not well-suited for the interactive web maps commonly used in cartographic research. The rapidly changing content of these maps, where users frequently pan and zoom, makes manual annotation of AOIs impractical and labor-intensive, as each participant's video is unique. The above-mentioned approaches, like Assisted Mapping, also do not work properly because, due to generalization, the map's content differs for different zoom levels, and it is complicated to recognize similar features automatically (Vanicek, Beitlova, Vojtechovska, & Popelka, 2024.

For the effective eye-tracking analysis of participants' behavior with interactive web maps, the ET2Spatial tool was developed. The tool's main functionality is to convert the screen coordinates of the participant's gaze to real-world coordinates and allow exports in commonly used spatial data formats (Figure 27, top left). Gaze coordinates transformed into geo-coordinates can provide more information and feasible solutions to the existing issues with interactive web maps (Giannopoulos et al., 2012; Ooms, Coltekin, et al., 2015).

ET2Spatial was developed in Python. The tool takes three input files: the raw ET points, fixation points, and the user interaction data recorded using MapTrack (Růžička, 2012). These input datasets are pre-processed, synchronized based on timestamps, and stitched together. The main conversion of points relies on the Web Mercator projection formulas. The tool offers export in shapefile format along with CSV and GeoJSON formats (Figure 27, top right).

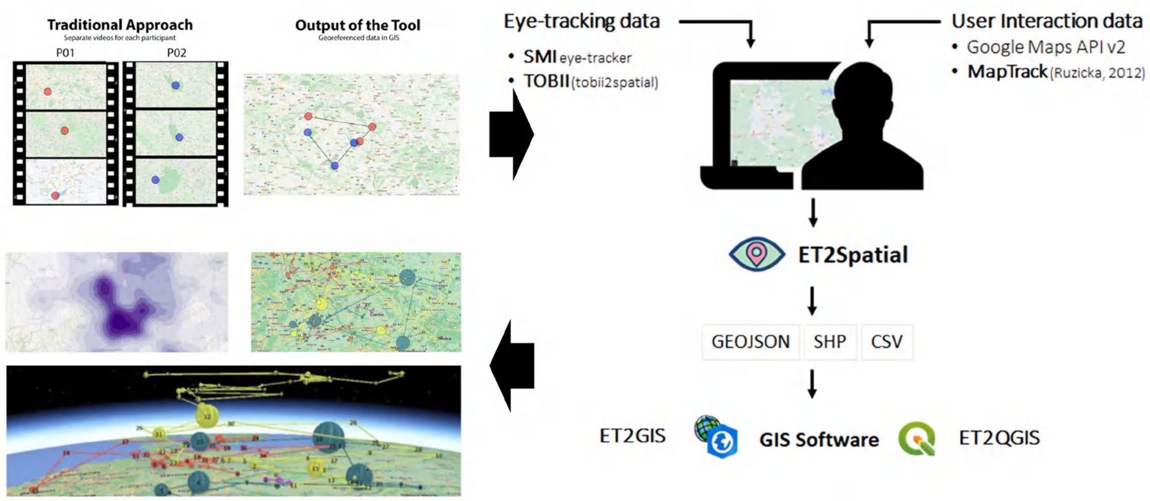


Figure 27. The principle of the ET2Spatial tool (top left); the process of eye movement data georeferencing (right); eye movement data visualizations in GIS (bottom left). Figure modified from Sultan et al. (2022).

A toolbox for Esri ArcGIS Pro 3.0+ called ET2GIS was developed to simplify working with the resulting shapefiles (Popelka et al., 2023). It includes a data import function for integrating data into the Esri environment and 12 visualization functions, such as scanpaths, heatmaps, kernel density, hexagons, zoom level clustering, space-time cube, and time visualization (Figure 27, bottom left). Concurrently, a similar application for the QGIS environment is under development (Popelka et al., 2023).

The georeferencing of eye movement data has not yet been applied to a real case study, but the entire framework is nearly complete and ready for practical use.

This paper introduced a tool for georeferencing eye movement data which will greatly simplify the analysis of eye movement data recorded on web maps.

4.11 Paper Dashboards (Porti Suarez & Popelka, 2023) 🎮

The final study in this habilitation thesis describes two case studies focused on user evaluations of COVID-19 dashboards containing geospatial information. A dashboard is a graphical user interface presenting key information in a consolidated and easily accessible format, typically on a single screen. While the definition of a dashboard varies, Few (2006) described it as a tool for displaying the most valuable information needed to achieve specific objectives at a glance. The design of a dashboard is crucial for effectively communicating information, and requires careful consideration of usability, visual perception, and human-centered design principles. Despite the increasing popularity of dashboards during the COVID-19 pandemic, there remains a notable lack of comprehensive design guidelines, making the evaluation and refinement of dashboard design aspects essential to ensure that they meet user needs and facilitate decision-making (Few, 2006; Monkman et al., 2021; Sedrakyan et al., 2019).

The study established three goals: first, to analyze user interactions with four existing dashboards to identify problematic elements and recommend improvements; second, to develop new dashboards based on these insights; and third, to verify the usability of the newly developed dashboards through empirical assessment.

Experiment I utilized four COVID-19 dashboards aligning with Few's definition of a dashboard(Figure 28, top left), each with different functionalities and geospatial visualizations. Twelve tasks of varying difficulty were assigned to these dashboards to evaluate user interaction, following methods similar to those used by Fan et al. (2023). Insights from this experiment directed the design of two self-developed dashboards showcasing COVID-19 data from Catalonia (Figure 28, top right). These dashboards were then used as stimuli in experiment II, where five tasks of increasing difficulty were assigned to each of them, following the Roth and MacEachren (2016) typology. As in the studies published in **[paper UrbanPlans]** and **[paper WeatherMaps]**, objective eye-tracking data was supplemented with subjective methods, including an interview focused on opinions about dashboard design (Figure 28, bottom left) and a brief questionnaire.

The results of Experiment I identified problematic aspects of the dashboards and provided recommendations for improvement. Users preferred light aesthetics, intuitive interfaces, and interactive features, while they found dashboards with too many elements, difficult-to-use features, and excessive explanatory text challenging. The study recommends designing dashboards with light, simple aesthetics, a choropleth map, a searchable country list, interactive numeric metrics, user-friendly graphs, and a clear title, while avoiding darker colors, static metrics, and overly complex features. Two dashboard variants for Catalonia, light and dark, were developed based on these insights.

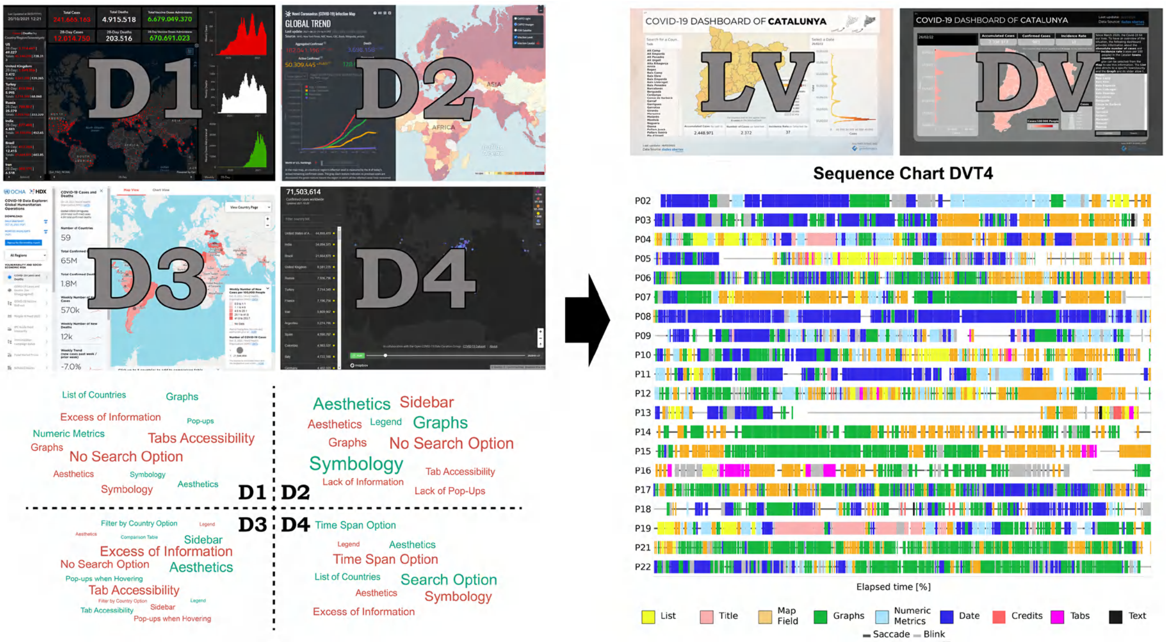


Figure 28. Four existing dashboards served as stimuli in Experiment I (top left); participants' preferences gathered from the interview (bottom left); two self-developed dashboards served as stimuli in Experiment II (top right) and a sequence chart generated using GazePlotter. Figure modified from Porti Suarez and Popelka (2023).

Experiment II found that users preferred the light dashboard for its user-friendly design and effective use of interactive elements. In addition, users were more accurate in the performance of their tasks when using the light dashboard. Sequence charts created using **[paper GazePlotter]** were used to visualize attention distribution among marked AOIs (Figure 28, bottom right). In the light version of the dashboard, the graph was readily used for date searches, and while the map field and numeric metrics were widely used, the tabs for changing administrative levels were often overlooked and confusing. In the dark version, the time slider drew more attention but was inefficient, and despite increased fixation on the tabs, the overall usability was lower.

This study evaluated user interaction with COVID-19 dashboards, established design improvements based on mixed research methods, and developed more user-friendly interfaces by analyzing eye-tracking data and insights on user preferences to enhance information communication.

The paper provides valuable insights into the design challenges of dashboards, which became widely used by the general public during the COVID-19 pandemic.

5 DISCUSSION

The discussions for each study are presented within their respective articles, each focusing on specific aspects. However, the purpose of this overall discussion is to address the **broader challenges and limitations** of the habilitation thesis as a whole. This includes topics such as the selection of articles for the habilitation, the thematic focus of the individual studies, and the methodological constraints or difficulties related to technology and data collection. The aim is to offer a comprehensive overview of the results and highlight any potential limitations.

The habilitation thesis comprises eleven articles chosen not for their impact factors or citation counts, but for how they **align with the overarching research narrative**. This work is highly interdisciplinary, integrating principles from cognitive science, technology, and cartography to address complex questions in user interaction and visualization. The goal is to present long-term research in cognitive cartography, supported by eye-tracking technology, prioritizing the coherence and progression of the research over individual metrics. The focus is on how these works collectively contribute to a unified and evolving research story, rather than their standalone impact.

While the overall narrative is central to the habilitation thesis, the selection of topics for individual case studies was not part of any grand plan aimed specifically at writing the thesis. Instead, these topics were often shaped by the interests of collaborators, who sought to validate or compare the visualization methods they were using. This approach, though seemingly diverse, accurately mirrors the fluid nature of cognitive cartography.

As noted in the Introduction (chapter 2.1), MacEachren and Kraak (2001) highlighted that user-related aspects were relevant across all the agenda topics for the Commission on Visualization and Virtual Environments. This insight led to the formation of a specialized working group focused on usability. Similarly, a cognitive cartographer does not necessarily need to select specific areas of exploration but should be prepared to address the needs of colleagues unfamiliar with user testing.

A clear example of this collaborative approach is found in the case study by Beran et al. (2021). Colleagues from the University of West Bohemia in Pilsen developed methods for visualizing noise and submitted an article for review. The reviewers, however, requested usability validation. The author of this thesis was then involved, and an experiment was collaboratively designed and conducted to test the methods with users.

Although collaborators often influenced the study topics, efforts were consistently made to structure the studies and analyses in a way that aligned with a broader research plan, now reflected in the thesis narrative.

Expecting a **single transformative discovery** is unrealistic in the field of cognitive cartography. This is illustrated by the progression of the author's dissertation research. Early in his scientific career, while focusing on 3D visualization in cartography, the author envisioned conducting a series of experiments comparing different types of 2D and 3D representations. The plan was to analyze the data and determine a clear recommendation for the use of 3D visualization methods in cartography. In retrospect, this approach was overly ambitious and naive. The scope was too broad, and even if the research had focused on a single aspect of 3D visualization, the findings would not have been universally applicable due to the many influencing factors—such as the design of the maps, the tasks given to participants, and even intercultural differences. Rather than yielding one definitive conclusion, the research provided a series of smaller, incremental insights into the usability of 3D visualization methods in cartography. Though modest on their own, these insights collectively improve the effectiveness and usability of maps. This same approach was adopted for the habilitation thesis—not aiming for a singular ultimate goal, but rather uncovering individual insights about specific visualizations and interactions with maps.

Through years of involvement in eye movement research, a comprehensive methodology has been developed, applicable to every phase of eye-tracking cognitive research—from experimental design through study execution to the analysis and interpretation of collected data. Multiple tools have been designed to streamline the research process and, while intended for cartographic experiments, they are versatile enough for interdisciplinary applications across various fields. However, making these tools publicly accessible poses certain challenges and potential risks. Issues, undiscovered during development and testing, may arise, requiring ongoing updates and user-driven modifications, which can be time-consuming. Despite this, most of the tools are available through a GitHub repository under the GNU GPL v3 license, encouraging public participation in their development. Some tools have commercial potential, but the decision was made to release them as **open source**. Given the relatively small size of the eye-tracking research community, making these tools freely available feels like the right choice, particularly as open-source tools from others have greatly benefited this work. By contributing to this collaborative environment, the goal is to enhance the research capabilities of those working with eye-tracking data.

A common limitation in empirical studies involving human subjects is **sample size**. In the studies presented, the number of participants ranged from 19 to 40. While these numbers are not ideal, they are consistent with other eye-tracking studies. For example, Alacam and Dalci (2009) used 26 participants in their study on map symbol identification, and Fuchs et al. (2009) used 21 participants in their research on flood maps. Some studies have even used smaller samples, such as Ooms et al. (2010) with 14 participants and Opach and Nossum (2011) with 10 participants. In the author's dissertation research, an analysis of 15 quantitative cartographic eye-tracking studies revealed that the average number of participants was 25 (Popelka, 2015). As Bojko (2013) noted, the final number of participants in a study often depends on available resources. In eye-tracking studies, the primary constraint is time, as typically only one participant can be tested at a time. Given the limited sample sizes, emphasis was often placed on a qualitative approach to data analysis, which does not require large numbers of participants.

Unlike in psychology and other fields in the social sciences or humanities, obtaining **ethical approval** for experiments involving human subjects, such as recording eye movements, has not historically been common practice in cartography. However, this has changed in recent years, with many cartographic journals now requiring ethical approval. At Palacký University Olomouc, the Faculty of Science's ethics committee was established on January 1, 2020. Since 2021, the cognitive research conducted at the Department of Geoinformatics has been approved by this committee, and all experiments adhere to the ethical standards of the institutional research committee and the 1964 Helsinki Declaration, along with its subsequent amendments.

This research described in the habilitation thesis is ongoing, representing an **evolving field** rather than a completed chapter. For instance, **[paper ET2Spatial]** introduces a tool for georeferencing eye-tracking data. Building on this work, a toolbox for ArcGIS was developed, offering a comprehensive framework for working with web maps. However, despite its availability, this toolbox has yet to be applied in practice and awaits future utilization.

Despite the above-mentioned limitations, it is evident that the long-term eye-tracking research conducted by the Department of Geoinformatics at Palacký University is of **world-class quality**. Since the first studies in 2011, substantial progress has been made. Initially, eye-tracking was used as a tool based primarily on manufacturer-supplied methods, but over time, the research evolved to develop custom tools for analysis, applicable far beyond cognitive cartography. A notable example of this broader impact was the invitation to collaborate with 46 leading scientists on formulating guidelines for reporting eye movement studies (Holmqvist et al., 2022). Unfortunately, this paper was later retracted. Officially, this was because a number of statements were supported by two references from the first author, which should not have been used. Practically, this was due to the criminal acts of the first author, Kenneth Holmqvist. Nevertheless, working with others across research domains was a rewarding experience; there were no other researchers from the geo-domain included in this collaboration.

In Wang et al.'s (2016) review, Palacký University Olomouc was named as a center of eye-tracking research in cartography, and the author was identified as the most prominent in the field of eye-tracking in cognitive cartography.

Further recognition came with the Henry Johns Award, which the **[paper Author-Reader]** received from the Editorial Board of *The Cartographic Journal* for being the most outstanding paper of 2023. The paper was praised for its significant empirical contribution to understanding map reading and for prompting fresh appraisals of the theoretical framework of cartographic communication.

In July 2024, the author's impact was again acknowledged outside the geosciences, when he was ranked as a Top Scholar by the ScholarGPS platform, placing him among the top 0.5% of scientists worldwide in the field of eye-tracking.

6 CONCLUSIONS AND FUTURE WORK

The habilitation thesis consists of eleven articles, six of which present case studies evaluating cartographic visualization methods or interactive map interfaces. The remaining five articles focus on methodological contributions to eye-tracking research, particularly in cognitive cartography.

The results of this work can be categorized into three areas corresponding to these specific goals:

1. **Contributions to cartographic knowledge of visualization methods (🗺️);**
2. **Insights into interactivity with cartographic interfaces (🎮);**
3. **Advances in the methodology of cartographic eye-tracking research (🧰).**

6.1 Contributions to Cartographic Knowledge of Visualization Methods 🗺️

The author would like to highlight a few results from this part of the work that he considers the most significant in advancing the field of cartography:

- The cartographic results of **[paper 2D-3D]** related to 3D visualization were not particularly strong due to the less-than-ideal design of the experiment. However, this article was essential for the design of all subsequent experiments. Additionally, the research revealed that a commonly used data analysis method was improperly implemented, leading to the development of a new approach for scanpath comparison.
- Another important case study is **[paper Glyphs]**, which gave the author his first opportunity to work within an international team. The design of this experiment was thoroughly planned, and the results are reliable and directly applicable to cartographic practice. The entire study was conducted with the practical aim of determining which visualization method used in weather risk visualization tools (Opach & Rød, 2013) is more suitable for users.
- The final case study, **[paper UrbanPlans]**, conducted as a part of the first goal of the habilitation thesis, focused on the evaluation of urban plans. The motivation for choosing this topic came from collaboration with Dr. Burian, who has a longstanding commitment to urban planning. He has consistently advocated for the standardization of zoning plan designs, emphasizing that poor cartographic quality in urban plans significantly impacts how they are interpreted and, in extreme cases, can lead to incorrect conclusions. The goal of this study was to empirically verify this assertion, which was successfully accomplished.
- Although only three case studies were part of the first sub-goal, it is important to mention **[paper Author-Reader]** here. While it is categorized as a methodological study due to the development of a new scanpath comparison method, its findings also hold significant value for the field of cartography. This was the first study to empirically verify Koláčný's (1969) model of cartographic communication by placing map authors in the role of map readers. The article won the Cartographic Journal's **Henry Johns Award for the most outstanding paper** of 2023 because it makes a significant empirical contribution toward understanding map reading and invites fresh appraisals of the theoretical framework of cartographic communication. The newly developed scanpath comparison method facilitated the identification of situations where the map reading strategies of authors and readers either matched or diverged.

The range of cartographic visualization methods available for evaluation is virtually endless. Eye-tracking has proven to be a valuable resource for understanding how participants interact with these visualizations, making it applicable in a wide variety of contexts, especially with the development of tools that simplify this analysis.

All of the case studies presented in this habilitation thesis employed a remote eye-tracker, a device placed under the monitor, which limits its use to maps displayed on a screen. Nonetheless, the author has already made initial attempts to use **different types of eye-trackers**. For example, mobile eye-tracking glasses were used to evaluate the attractiveness of a geographic exhibit at the science museum Pevnost Poznání (Popelka & Vysloužil, 2024).

The current trend is to integrate eye-tracking into VR headsets, where the eye-tracker might help to improve the realism and speed of image rendering. The tracker can also provide data about how users perceive various visualization methods displayed in the VR environment. The author is currently collaborating with colleagues at Masaryk University in Brno, who have developed a new platform for collaborative learning and teaching in virtual reality called eDIVE (Jochecová et al., 2022). This platform is being used in a VR eye-tracking experiment focused on evaluating 3D multivariate cartographic visualization methods. For the initial study, bar charts and Chernoff faces with varying levels of separability were selected (Kvarda et al., 2024).

6.2 Insights into Interactivity with Cartographic Interfaces

The first sub-objective focused on methods of cartographic visualization, specifically the visualization of static maps. Since static maps are simpler to evaluate using eye-tracking, this was a logical starting point. However, with interactive maps becoming prevalent, it has become evident that analyzing interactive maps is also necessary. As a result, the following case studies were dedicated to evaluating and comparing the interactivity with interfaces of these maps.

- In **[paper WeatherMaps]**, the primary contribution was the practical validation of the ScanGraph tool. From a cartographic perspective, the main result is the identification of how design choices impact user interaction with weather web maps. The study highlights that simpler, static menus improve user efficiency, while overly complex interactivity can hinder it.
- The results of the **[paper Swipe-Multiple]** study deserve special attention. The original idea was quite practical: to compare the map comparison methods of swipe and multiple views, which were being used for visualizing land suitability in the UrbanPlanner model (Burian et al., 2015). However, during the pilot testing, it became clear that the results would be influenced by the way users interact with the Esri Web App environment. In the case of swipe and comparison of more than two layers, this environment proved to be practically unusable. During the preparation of the experiment, a debate arose about whether to adhere to the original plan of comparing swipe and multiple views by adjusting the design or to shift the study's focus to analyzing the Esri interface itself. Ultimately, it was determined that, given the widespread use of the Esri environment, evaluating its shortcomings could have a much greater impact on cartography than merely comparing the two cartographic visualization methods.
- The final article in this section is **[paper Dashboards]**, chosen for logical reasons—the global COVID-19 pandemic led to the widespread use of dashboards among the general public, and it was noted that not all of them were cartographically well-designed. The key contribution of this study is the set of recommendations for improving the cartographic

design of dashboards. It was found that effective interactivity in COVID-19 dashboards significantly enhances user experience, with well-designed interactive elements like dynamic data selectors improving task accuracy and user satisfaction. In contrast, poorly implemented or non-interactive elements led to confusion and lower task success rates.

Much remains to be done in the evaluation of **cartographic interaction**, presenting significant opportunities for future research by cognitive cartographers. The doctoral thesis of PhD candidate Tomáš Vaníček, who is under the supervision of the author, is directly focused on the evaluation of dynamic eye-tracking data, ensuring continuity for future research. The gap in the evaluation of cartographic interaction motivated the application for a Czech Science Foundation project, which was successfully granted. The author is now serving as the principal investigator of the project "Identification of Barriers in the Process of Communication of Spatial Socio-Demographic Information," which focuses on eye-tracking investigations into the challenges of using interactive maps. In the project's first phase, qualitative interviews with experts identified and classified potential barriers. A key finding was that the growing use of maps on mobile devices negatively impacts interactivity, as it is challenging both to design for and to use interactive maps effectively on mobile devices.

Initial research into the usability challenges presented through the display of interactive maps on mobile device screens has already begun. It is expected to lead to a collaboration with colleagues at Adam Mickiewicz University in Poznań, which is equipped with a Gazepoint GP3 Mobile eye-tracker, capable of recording eye movements during mobile phone use. Additionally, a user action logging application is currently being developed as an alternative to eye-tracking (Popelka, Vojtechovska, et al., 2024).

Future research will focus on addressing these challenges and exploring map reading on mobile devices.

6.3 Advances in the Methodology of Cartographic Eye-tracking Research 🧰

A major area of contribution in this habilitation thesis is to the methodology of eye-tracking research, which has broad applications not only in the analysis and optimization of maps but also across various research fields.

From the methodological advancements in the field of cartography, the following results have been selected as the most important:

- The first significant contribution to the analytical tools for eye movement data is the software ScanGraph **[paper ScanGraph]**, developed as an alternative to eyePatterns (West et al., 2006), which was found to be unusable. ScanGraph enables scanpath comparison using the String-Edit-Distance method. The initial version of the software was published in Doležalová and Popelka (2016). However, its capabilities were later expanded to calculate similarities for multiple files representing various stimuli (Popelka et al., 2018) and by incorporating the Multimatch algorithm (Dewhurst et al., 2012) alongside String-Edit-Distance (Popelka & Beitlova, 2022). In **[paper Author-Reader]**, the ScanGraph functionality was further modified to compare map reading strategies between one author and multiple map readers. Over the years, ScanGraph has been instrumental in evaluating map reading (Beitlova et al., 2020; Brus et al., 2019; Havelková & Gołębiowska, 2020; Havelková & Hanus, 2019), education and didactics (Skrabankova et al., 2020) and autism research (Eraslan et al., 2017, 2019).

- In 2014, a new low-cost eye-tracker called the EyeTribe entered the market, offering the potential to analyze a larger number of participants simultaneously. To verify the accuracy and precision of these devices, we conducted a comparison with the professional eye-tracking device SMI RED 250 **[paper EyeTribe].** Although the results were promising, the EyeTribe device was soon discontinued, making it unavailable for future studies. However, as an offshoot of this comparison, we developed a tool called HypOgama, which allows eye movement data recorded on the Hypothesis platform to be imported into the OGAMA environment for further analysis. This tool has since been utilized in numerous studies (Šašinka et al., 2021; Šašinka et al., 2019), making it a valuable aid for eye-tracking research conducted on the Hypothesis platform.
- Another groundbreaking piece of software developed as part of this habilitation is GazePlotter **[paper GazePlotter]**. This online, open-source application enables the generation of sequence charts (scarf plots) from various eye-trackers, including the Tobii, SMI, GazePoint, and Varjo. Its key advantages are its interactivity, the ability to export data in vector format, and its capability to work with dynamic AOIs. Additionally, GazePlotter can convert data into the ScanGraph format. Although the software is relatively new, it has already been utilized in studies by Popelka, Kominek, et al. (2024), Popelka and Vysloužil (2024), and **[paper Dashboards]**. The tool was also presented at two of the most important conferences in the field of eye-tracking (ETRA 2024 and ECEM 2024), where it received significant attention.
- The final methodological contribution of this habilitation thesis is the ET2Spatial software **[paper ET2Spatial]**. This tool was specifically developed for use in cognitive cartography, as it converts gaze coordinates from the screen's pixel-based coordinate system into a geographic coordinate system (latitude and longitude). This functionality enables the analysis of eye movements recorded while using a web map as a stimulus. To further simplify data analysis, a toolbox for Esri ArcGIS Pro was subsequently created (Popelka et al., 2023), and a version for QGIS is currently in development. The entire system is ready for use and awaits future case studies where it can be applied.

Currently, the tools available for eye-tracking analysis seem to meet the needs of the scientific community, but it is expected that **new challenges will arise soon**. The functionality of GazePlotter could be further enhanced, for instance, by adding result quantification or integrating ScanGraph capabilities directly into GazePlotter. It would also be valuable to combine sequence charts with other data, such as user interactions or auditory feedback from think-aloud protocols. The ambitious goal is to develop fully functional open-source software for eye-tracking analysis as an alternative to OGAMA, which was discontinued in 2015 and remains the only open-source application capable of handling data from various eye-trackers.

The advancements in eye movement methodology developed within this habilitation have focused primarily on the diagnostic use of eye-tracking, where data is recorded and analyzed afterward. The second potential application of eye-tracking—**interactive use**—has been completely overlooked. This gap will be addressed by a PhD candidate under my supervision, Michaela Vojtěchovská, the main developer of GazePlotter, whose doctoral thesis is centered on real-time gaze-based interactions with maps.

6.4 Summary of Findings

Reflecting on the research presented, the habilitation thesis has made significant contributions to advancing the field of cartography, particularly in the context of eye-tracking methodologies. While the field of cartography continues to evolve, this research has uncovered **numerous incremental insights** into how users interact with maps, both static and dynamic, across various visualizations. Although there was no single transformative discovery, the methodological developments and case studies presented have collectively **deepened the understanding of cognitive cartography**. From developing custom tools such as ScanGraph and GazePlotter to investigating user interaction with increasingly complex cartographic interfaces, these contributions serve as essential building blocks for **future research**.

This habilitation thesis follows in the footsteps of George Frederick Jenks, whose groundbreaking eye-tracking research over 50 years ago opened a "**Pandora's box**" of cartographic complexity, revealing the intricate challenges of how users interact with maps. While the initial goal of this work was to help with closing that box, it has become clear that it would be a mistake to do so. Eye-tracking has provided so many invaluable insights into how maps are used, and attempting to resolve every challenge would limit the potential for future discoveries. New possibilities continue to emerge, from displaying maps on mobile devices and interacting with them in virtual reality to exploring gaze-based and multimodal interactions with cartographic outputs.

The research presented in this habilitation is far from complete—and that is where its greatest strength and beauty lie.

REFERENCES

Alacam, Ö., & Dalci, M. (2009). A usability study of WebMaps with eye tracking tool: the effects of iconic representation of information. In J. A. Jacko (Ed.), *New trends in Human-Computer interaction, Lecture Notes in Computer Science* (Vol. 5610, pp. 12-21). Springer. https://doi.org/10.1007/978-3-642-02574-7_2

Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2014). A comparison of scanpath comparison methods. *Behavior research methods*(47), 1377–1392. https://doi.org/10.3758/s13428-014-0550-3

Andrienko, G., Andrienko, N., Burch, M., & Weiskopf, D. (2012). Visual Analytics Methodology for Eye Movement Studies. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2889-2898. https://doi.org/10.1109/TVCG.2012.276

Beitlova, M., Popelka, S., Konopka, M., & Macku, K. (2022). Verification of cartographic communication models using detection of map reading strategies based on eye movement recording. *The Cartographic Journal*, *60*(1), 5-24. https://doi.org/10.1080/00087041.2022.2150362

Beitlova, M., Popelka, S., & Vozenilek, V. (2020). Differences in Thematic Map Reading by Students and Their Geography Teacher. *ISPRS International Journal of Geo-Information*, *9*(9), 492. https://doi.org/10.3390/ijgi9090492

Beran, D., Jedlička, K., Kumar, K., Popelka, S., & Stoter, J. (2021). The Third Dimension in Noise Visualization–a Design of New Methods for Continuous Phenomenon Visualization. *The Cartographic Journal*, *59*(1), 1-17. https://doi.org/10.1080/00087041.2021.1889450

Biedert, R., Buscher, G., & Dengel, A. (2010). The eyebook–using eye tracking to enhance the reading experience. *Informatik-Spektrum*, *33*(3), 272-281. https://doi.org/10.1007/s00287-009-0381-2

Board, C. (1978). Map reading tasks appropriate in experimental studies in cartographic communication. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *15*(1), 1-12. https://doi.org/10.3138/AG15-V252-3726-W346

Bojko, A. (2013). *Eye tracking the user experience: A practical guide to research*. Rosenfeld Media.

Borgo, R., Kehrer, J., Chung, D. H., Maguire, E., Laramee, R. S., Hauser, H., Ward, M., & Chen, M. (2013). *Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications*. Eurographics (state of the art reports).

Brodersen, L., Andersen, H. H., & Weber, S. (2002). *Applying eye-movement tracking for the study of map perception and map design*. National Survey and Cadastre.

Brus, J., Kucera, M., & Popelka, S. (2019). Intuitiveness of geospatial uncertainty visualizations: a user study on point symbols. *Geografie*, *124*(2), 163-185.

Brychtova, A., & Coltekin, A. (2016). An Empirical User Study for Measuring the Influence of Colour Distance and Font Size in Map Reading Using Eye Tracking. *The Cartographic Journal*, *53*(3), 202-212. https://doi.org/10.1179/1743277414Y.0000000103

Burian, J., Brychtová, A., Vávra, A., & Hladišová, B. (2016). Analytical material for planning in Olomouc, Czech Republic. *Journal of Maps*, *12*, 649-654. https://doi.org/10.1080/17445647.2015.1060177

Burian, J., Popelka, S., & Beitlova, M. (2018). Evaluation of the Cartographical Quality of Urban Plans by Eye-Tracking. *ISPRS International Journal of Geo-Information*, *7*(5), 192. https://doi.org/10.3390/ijgi7050192

Burian, J., Stastny, S., Brus, J., Pechanec, V., & Vozenilek, V. (2015). Urban planner: Model for optimal land use scenario modelling. *Geografie*, *120*(3), 330-353. https://doi.org/10.37040/geografie2015120030330

Buswell, G. T. (1935). *How people look at pictures*. University of Chicago Press Chicago.

Caligari, M., Godi, M., Guglielmetti, S., Franchignoni, F., & Nardone, A. (2013). Eye tracking communication devices in amyotrophic lateral sclerosis: impact on disability and

quality of life. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *14*(7-8), 546-552. https://doi.org/10.3109/21678421.2013.803576
Campbell, D. (1980). *Modely experimentov v socialnoj psichologii i prikladnych issledovanijach*. Progress.
Cao, J., & Nishihara, A. (2013). Viewing behaviors affected by slide features and learning style in slide video from a sequence analysis perspective. *The Journal of Information and Systems in Education*, *12*(1), 1-12.
Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49-62. https://doi.org/10.1016/j.ijpsycho.2020.05.010
Cartwright, W., Crampton, J., Gartner, G., Miller, S., Mitchell, K., Siekierska, E., & Wood, J. (2001). Geospatial Information Visualization User Interface Issues. *Cartography and Geographic Information Science - CARTOGR GEOGR INF SCI*, *28*, 45-60. https://doi.org/10.1559/152304001782173961
Coltekin, A., Fabrikant, S., & Lacayo, M. (2010). Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. *International Journal of Geographical Information Science*, *24*(10), 1559-1575. https://doi.org/10.1080/13658816.2010.511718
Coltekin, A., Heil, B., Garlandini, S., & Fabrikant, S. I. (2009). Evaluating the effectiveness of interactive map interface designs: a case study integrating usability metrics with eye-movement analysis. *Cartography and Geographic Information Science*, *36*(1), 5-17. https://doi.org/10.5167/uzh-10125
Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, *7*(3), 171-176. https://doi.org/10.1145/363958.363994
Delabarre, E. B. (1898). A method of recording eye-movements. *The American Journal of Psychology*, *9*(4), 572-574.
Dewhurst, R., Nyström, M., Jarodzka, H., Foulsham, T., Johansson, R., & Holmqvist, K. (2012). It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach. *Behavior research methods*, *44*(4), 1079-1100. https://doi.org/10.3758/s13428-012-0212-2
Dodge, R., & Cline, T. S. (1901). The angle velocity of eye movements. *Psychological Review*, *8*(2), 145-157. https://doi.org/10.1037/h0076100
Dolezalova, J., & Popelka, S. (2016). *Evaluation of user strategy on 2D and 3D city maps based on novel scanpath comparison method and graph visualization*. ISPRS 2016, Prague.
Doležalová, J., & Popelka, S. (2016). ScanGraph: A novel scanpath comparison method using graph cliques visualization. *Journal of Eye Movement Research*, *9*(4), 1-13. https://doi.org/10.16910/jemr.9.4.5
Dong, W., Liao, H., Xu, F., Liu, Z., & Zhang, S. (2014). Using eye tracking to evaluate the usability of animated maps. *Science China Earth Sciences*, *57*(3), 512-522. https://doi.org/10.1007/s11430-013-4685-3
Dostálová, N., Pátková Daňsová, P., Ježek, S., Vojtechovska, M., & Šašinka, Č. (2024). *Towards the Intervention of Dyslexia: a Complex Concept of Dyslexia Intervention System using Eye-Tracking*. Proceedings of the 2024 Symposium on Eye Tracking Research and Applications.
Dühr, S. (2004). The form, style, and use of cartographic visualisations in European spatial planning: examples from England and Germany. *Environment and Planning A*, *36*, 1961-1989. https://doi.org/10.1068/a35262
Dühr, S. (2007). The visual language of spatial planning. 224. https://doi.org/10.4324/9780203965818
Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice* (Vol. 373). Springer - Verlag.
Duchowski, A. T., Driver, J., Jolaoso, S., Tan, W., Ramey, B. N., & Robbins, A. (2010). *Scanpath comparison revisited*. Symposium on Eye-Tracking Research & Applications (ETRA 2010), Austin, Texas, USA.

Enoch, J. M. (1959). Effect of the size of a complex display upon visual search. *JOSA*, *49*(3), 280-285. https://doi.org/10.1364/JOSA.49.000280

Eraslan, S., Yaneva, V., Yesilada, Y., & Harper, S. (2017). *Do web users with autism experience barriers when searching for information within web pages?* Proceedings of the 14th International Web for All Conference.

Eraslan, S., Yaneva, V., Yesilada, Y., & Harper, S. (2019). Web users with autism: eye tracking evidence for differences. *Behaviour & Information Technology*, *38*(7), 678-700. https://doi.org/10.1080/0144929X.2018.1551933

Fabrikant, S. I., Hespanha, S. R., & Hegarty, M. (2010). Cognitively inspired and perceptually salient graphic displays for efficient spatial inference making. *Annals of the Association of american Geographers*, *100*(1), 13-29. https://doi.org/10.1080/00045600903362378

Fabrikant, S. I., Rebich-Hespanha, S., Andrienko, N., Andrienko, G., & Montello, D. R. (2008). Novel method to measure inference affordance in static small-multiple map displays representing dynamic processes. *Cartographic Journal, The*, *45*(3), 201-215. https://doi.org/10.1179/000870408X311396

Fairbairn, D., & Hepburn, J. (2023). Eye-tracking in map use, map user and map usability research: what are we looking for? *International Journal of Cartography*, *9*(2), 231-254. https://doi.org/10.1080/23729333.2023.2189064

Fan, M., Wang, Y., Xie, Y., Li, F. M., & Chen, C. (2023). Understanding How Older Adults Comprehend COVID-19 Interactive Visualizations via Think-Aloud Protocol. *International Journal of Human–Computer Interaction*, *39*(8), 1626-1642. https://doi.org/https://doi.org/10.48550/arXiv.2202.11441

Ferjenčík, J. (2000). *Úvod do metodologie psychologického výzkumu: jak zkoumat lidskou duši*. Portál.

Few, S. (2006). *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly Media, Inc. https://doi.org/10.5555/1206491

Franchak, J. M., Kretch, K. S., Soska, K. C., Babcock, J. S., & Adolph, K. E. (2010). *Head-mounted eye-tracking of infants' natural interactions: a new method*. Symposium on Eye-Tracking Research & Applications (ETRA 2010), Austin, Texas, USA.

Fuhrmann, S., Holzbach, M. E., & Black, R. (2015). Developing interactive geospatial holograms for spatial decision-making. *Cartography and Geographic Information Science*, *42*(sup1), 27-33. https://doi.org/10.3390/buildings14072030

Fuchs, S., Spachinger, K., Dorner, W., Rochman, J., & Serrhini, K. (2009). Evaluating cartographic design in flood risk mapping. *Environmental Hazards*, *8*(1), 52-70. https://doi.org/10.3763/ehaz.2009.0007

Garbutt, M., East, S., Spehar, B., Estrada-Gonzalez, V., Carson-Ewart, B., & Touma, J. (2020). The embodied gaze: Exploring applications for mobile eye tracking in the art museum. *Visitor Studies*, *23*(1), 82-100. https://doi.org/10.1080/10645578.2020.1750271

Giannopoulos, I., Kiefer, P., & Raubal, M. (2012). *GeoGazemarks: Providing gaze history for the orientation on small display maps*. Proceedings of the 14th ACM international conference on Multimodal interaction.

Gilmartin, P. (1992). Twenty-five years of cartographic research: a content analysis. *Cartography and Geographic Information Systems*, *19*(1), 37-47.

Gleicher, M. (2017). Considerations for visualizing comparison. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 413-423. https://doi.org/10.1109/TVCG.2017.2744199

Goldberg, J. H., & Helfman, J. (2011). Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information Visualization*, *10*(3), 182-195. https://doi.org/10.1177/1473871611406623

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, *24*(6), 631-645. https://doi.org/10.1016/S0169-8141(98)00068-7

Golebiowska, I., Opach, T., & Rød, J. K. (2017). For your eyes only? Evaluating a coordinated and multiple views tool with a map, a parallel coordinated plot and a table using an eye-tracking approach. *International Journal of Geographical Information Science*, *31*(2), 237-252. https://doi.org/10.1080/13658816.2016.1191636

Gribov, A., Unwin, A., & Hofmann, H. (2006). About glyphs and small multiples: Gauguin and the Expo. *Statistical Computing and Graphics Newsletter*, *17*(2), 14-17.

Haeberling, C. (2005). *Cartographic design principles for 3D maps–A contribution to cartographic theory*. 22nd International Cartographic Conference, Caruña, Spain.

Hammoud, R. I., & Mulligan, J. B. (2008). Introduction to Eye Monitoring (*Passive Eye Monitoring* (pp. 1-19). Springer. https://doi.org/10.1007/978-3-540-75412-1_1

Hanington, B., & Martin, B. (2012). *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions*. Rockport Publishers.

Harris, C. M., Hainline, L., Abramov, I., Lemerise, E., & Camenzuli, C. (1988). The distribution of fixation durations in infants and naive adults. *Vision Research*, *28*(3), 419-432. https://doi.org/10.1016/0042-6989(88)90078-4

Harrison, B. L., Ishii, H., Vicente, K. J., & Buxton, W. A. (1995). Transparent layered user interfaces: An evaluation of a display design to enhance focused and divided attention. In I. R. Katz (Ed.), *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 317-324). Association for Computing Machinery. https://doi.org/10.1145/223904.223945

Havelková, L., & Gołębiowska, I. M. (2020). What Went Wrong for Bad Solvers during Thematic Map Analysis? Lessons Learned from an Eye-Tracking Study. *ISPRS International Journal of Geo-Information*, *9*(1), 9. https://doi.org/10.3390/ijgi9010009

Havelková, L., & Hanus, M. (2019). Research into map-analysis strategies: Theory-and data-driven approaches. *Geografie*, *124*(2), 187-216. https://doi.org/10.37040/geografie2019124020187

Hegarty, M., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 37-53. https://doi.org/10.1037/a0017683

Hendl, J. (2008). *Kvalitativní výzkum: základní teorie, metody a aplikace*. Portál.

Hennig, S., & Vogler, R. (2016). User-Centred Map Applications Through Participatory Design: Experiences Gained During the 'YouthMap 5020' Project. *The Cartographic Journal*, *53*(3), 213-229. https://doi.org/10.1080/00087041.2016.1148217

Herman, L., Popelka, S., & Hejlova, V. (2017). Eye-tracking Analysis of Interactive 3D Geovisualization. *Journal of Eye Movement Research*, *10*(3), 1-15, Article 2. https://doi.org/10.16910/jemr.10.3.2

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Holmqvist, K., Örbom, S. L., Hooge, I. T. C., Niehorster, D. C., Alexander, R. G., Andersson, R., Benjamins, J. S., Blignaut, P., Brouwer, A.-M., Chuang, L. L., Dalrymple, K. A., Drieghe, D., Dunn, M. J., Ettinger, U., Fiedler, S., Foulsham, T., van der Geest, J. N., Hansen, D. W., Hutton, S. B., Kasneci, E., Kingstone, A., Knox, P. C., Kok, E. M., Lee, H., Lee, J. Y., Leppänen, J. M., Macknik, S., Majaranta, P., Martinez-Conde, S., Nuthmann, A., Nyström, M., Orquin, J. L., Otero-Millan, J., Park, S. Y., Popelka, S., Proudlock, F., Renkewitz, F., Roorda, A., Schulte-Mecklenbeck, M., Sharif, B., Shic, F., Shovman, M., Thomas, M. G., Venrooij, W., Zemblys, R., & Hessels, R. S. (2022). Eye tracking: empirical foundations for a minimal reporting guideline. *Behavior research methods*, *54*(5), 1600-1626. https://doi.org/10.3758/s13428-021-01762-8

Hummel, G., Maier, S., Baumgarten, M., Eder, C., Strubich, P. T., & Stroebele-Benschop, N. (2021). Visual attention towards food during unplanned purchases–A pilot study

using mobile eye tracking technology. *PLoS One*, *16*(3), e0247755. https://doi.org/10.1371/journal.pone.0247755
Chen, K.-T., Prouzeau, A., Langmead, J., Whitelock-Jones, R. T., Lawrence, L., Dwyer, T., Hurter, C., Weiskopf, D., & Goodwin, S. (2023). *Gazealytics: A unified and flexible visual toolkit for exploratory and comparative gaze analysis*. Proceedings of the 2023 Symposium on Eye Tracking Research and Applications.
ISO. (2019). Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. International Organization for Standardization.
Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *20*(11), 1254-1259. https://doi.org/10.1109/34.730558
Jarodzka, H., Holmqvist, K., & Gruber, H. (2017). Eye tracking in Educational Science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research*, *10*(1), 1-18. https://doi.org/10.16910/jemr.10.1.1
Javal, L. É. (1878). Essai sur la psychologie de la lecture. *Annales d'Oculistique*, *80*, 240-274.
Jenks, G. F. (1973). Visual integration in thematic mapping: fact or fiction? *International Yearbook of Cartography*, *13*, 27-35.
Jochecová, K., Černý, M., Stachoň, Z., Švedová, H., Káčová, N., Chmelík, J., Brůža, V., Kvarda, O., Ugwitz, P., & Šašinková, A. (2022). Geography education in a collaborative virtual environment: a qualitative study on geography teachers. *ISPRS International Journal of Geo-Information*, *11*(3), 180. https://doi.org/10.3390/ijgi11030180
Kekule, M. (2015). *Students' different approaches to solving problems from kinematics in respect of good and poor performance*. International Conference on Contemporary Issues in Education, ICCIE.
Kiefer, P., Giannopoulos, I., Raubal, M., Duchowski, A., Kiefer, P., Giannopoulos, I., Raubal, M., & Duchowski, A. (2017). Eye Tracking for Spatial Research: Cognition, Computation, Challenges. *Spatial Cognition & Computation*, *17*(1-2), 1-19. https://doi.org/10.1080/13875868.2016.1254634
Kircher, K. (2007). *Driver distraction: a review of the literature*. Statens väg-och transportforskningsinstitut.
Koláčný, A. (1969). Cartographic information—a fundamental concept and term in modern cartography. *The Cartographic Journal*, *6*(1), 47-49.
Komogortsev, O. V., Jayarathna, S., Koh, D. H., & Gowda, S. M. (2010). *Qualitative and quantitative scoring and evaluation of the eye movement classification algorithms*. Symposium on Eye-Tracking Research & Applications (ETRA 2010), Austin, Texas, USA.
Krassanakis, V., & Cybulski, P. (2019). A review on eye movement analysis in map reading process: the status of the last decade. *Geodesy and Cartography*, *68*(1), 191-209. https://doi.org/10.24425/gac.2019.126088
Krassanakis, V., Filippakopoulou, V., & Nakos, B. (2016). Detection of moving point symbols on cartographic backgrounds. *Journal of Eye Movement Research*, *9*(2), 1-16. https://doi.org/10.16910/jemr.9.2.2
Kredel, R., Vater, C., Klostermann, A., & Hossner, E.-J. (2017). Eye-Tracking Technology and the Dynamics of Natural Gaze Behavior in Sports: A Systematic Review of 40 Years of Research. *Frontiers in psychology*, *8*, 1845. https://doi.org/10.3389/fpsyg.2017.01845
Kvarda, O., Šašinka, Č., & Stachoň, Z. (2024). *Spatial data visualization in collaborative iVR*. 9th International Conference on Cartography and GIS, Nessebar, Bulgraria.
Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics doklady*, *10*(8), 707-710.
Liao, H., Dong, W., Peng, C., & Liu, H. (2017). Exploring differences of visual attention in pedestrian navigation when using 2D maps and 3D geo-browsers. *Cartography and Geographic Information Science*, *44*(6), 474-490. https://doi.org/10.1080/15230406.2016.1174886

Lobben, A. K. (2004). Tasks, strategies, and cognitive processes associated with navigational map reading: A review perspective. *The Professional Geographer*, *56*(2), 270-281. https://doi.org/10.1111/j.0033-0124.2004.05602007.x
Lobo, M.-J., Pietriga, E., & Appert, C. (2015). An evaluation of interactive map comparison techniques (*Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 3573-3582). Association for Computing Machinery. https://doi.org/10.1145/2702123.2702130
Long, J., & Robertson, C. (2018). Comparing spatial patterns. *Geography Compass*, *12*(2), e12356. https://doi.org/10.1111/gec3.12356
MacEachren, A. M. (1994). Visualization in modern cartography: setting the agenda. *Visualization in modern cartography*, *28*(1), 1-12.
MacEachren, A. M. (1995). *How maps work: representation, visualization and design*. Guilford Press.
MacEachren, A. M. (2004). *How maps work: representation, visualization, and design*. Guilford Press.
MacEachren, A. M., & Kraak, M.-J. (2001). Research challenges in geovisualization. *Cartography and Geographic Information Science*, *28*(1), 3-12. https://doi.org/10.1559/152304001782173970
Maggi, S., Fabrikant, S. I., Imbert, J.-P., & Hurter, C. (2016). How do display design and user characteristics matter in animations? An empirical study with air traffic control displays. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *51*(1), 25-37. https://doi.org/10.3138/cart.51.1.3176
Majaranta, P., & Räihä, K.-J. (2002). *Twenty years of eye typing: systems and design issues*. Symposium on Eye tracking research & applications (ETRA 2002), New Orleans, Louisiana, USA.
Manson, S. M., Kne, L., Dyke, K. R., Shannon, J., & Eria, S. (2012). Using eye-tracking and mouse metrics to test usability of web mapping navigation. *Cartography and Geographic Information Science*, *39*(1), 48-60. https://doi.org/10.1559/1523040639148
Martin, D. W. (2007). *Doing psychology experiments*. Wadsworth Cengage Learning.
Martinez-Conde, S., & Macknik, S. L. (2008). Fixational eye movements across vertebrates: Comparative dynamics, physiology, and perception. *Journal of Vision*, *8*(14), 28-28. https://doi.org/10.1167/8.14.28
Matlin, M. W. (2014). *Cognitive psychology*. John Wiley & Sons.
McConkie, G. W. (1981). Evaluating and reporting data quality in eye movement research. *Behavior Research Methods & Instrumentation*, *13*(2), 97-106. https://doi.org/10.3758/BF03207916
Mendonça, A., & Delazari, L. (2014). Testing Subjective Preference and Map Use Performance: Use of Web Maps for Decision Making in the Public Health Sector. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *49*, 114-126. https://doi.org/10.3138/carto.49.2.1455
Mendonça, A. L. A., & Delazari, L. S. (2012). Remote Evaluation of the Execution of Spatial Analysis Tasks with Interactive Web Maps: A Functional and Quantitative Approach. *Cartographic Journal The*, *49*, 7-20. https://doi.org/10.1179/1743277411Y.0000000020
Merchant, J. (1967). *The oculometer* NASA Contractor Report No. CR-805
Mohamed, A. O., Da Silva, M. P., & Courboulay, V. (2007). *A history of eye gaze tracking* Rapport Interne
Monkman, H., Martin, S. Z., Minshall, S., Kushniruk, A. W., & Lesselroth, B. J. (2021). Opportunities to Improve COVID-19 Dashboard Designs for the Public. In R. Marcilly, L. Dusseljee-Peute, C. E. Kuziemsky, X. Zhu, P. Elkin, & N. Christian (Eds.), *Context Sensitive Health Informatics: The Role of Informatics in Global Pandemics* (2021/11/11 ed., Vol. 286, pp. 16-20). IOS Press BV. https://doi.org/10.3233/shti210628
Monmonier, M. (1996). *How to Lie with Maps (2nd Edition)*. University Of Chicago Press.

Montello, D. R. (2002). Cognitive map-design research in the twentieth century: Theoretical and empirical approaches. *Cartography and Geographic Information Science*, *29*(3), 283-304. https://doi.org/10.1559/152304002782032443
Morong, K., & Šašinka, Č. (2014). *Hypothesis – online software platform for objective experimental testing*. Applying principles of cognitive psychology in practice, Brno.
Morrison, J. L. (1977). The science of cartography and its essential processes. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *14*(1), 58-71.
Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353-383. https://doi.org/10.1016/0010-0285(77)90012-3
Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, *48*(3), 443-453. https://doi.org/10.1016/0022-2836(70)90057-4
Nivala, A.-M. (2007). *Usability Perspectives for the Design of Interactive Maps* [Dissertation, Helsinki University of Technology].
Ooms, K., Coltekin, A., De Maeyer, P., Dupont, L., Fabrikant, S., Incoul, A., Kuhn, M., Slabbinck, H., Vansteenkiste, P., & Van der Haegen, L. (2015). Combining user logging with eye tracking for interactive and dynamic applications. *Behavior research methods*, *47*(4), 977-993. https://doi.org/10.3758/s13428-014-0542-3
Ooms, K., De Maeyer, P., & Fack, V. (2010). *Analyzing eye movement patterns to improve map design*. 25th International Cartographic Conference.
Ooms, K., De Maeyer, P., Fack, V., Van Assche, E., & Witlox, F. (2012). Interpreting maps through the eyes of expert and novice users. *International Journal of Geographical Information Science*, *26*(10), 1773-1788. https://doi.org/10.1080/13658816.2011.642801
Ooms, K., Dupont, L., Lapon, L., & Popelka, S. (2015). Accuracy and precision of fixation locations recorded with the low-cost Eye Tribe tracker in different experimental setups. *Journal of Eye Movement Research*, *8*(1). https://doi.org/10.16910/jemr.8.1.3
Ooms, K., Maeyer, P. D., Fack, V., Assche, E. V., & Witlox, F. (2012). Investigating the Effectiveness of an Efficient Label Placement Method Using Eye Movement Data. *Cartographic Journal, The*, *49*(3), 234-246. https://doi.org/10.1179/1743277412Y.0000000010
Opach, T., Golebiowska, I., & Fabrikant, S. I. (2014). How Do People View Multi-Component Animated Maps? *The Cartographic Journal*, *51*(4), 330-342. https://doi.org/10.1179/1743277413Y.0000000049
Opach, T., & Nossum, A. (2011). *Evaluating the usability of cartographic animations with eye-movement analysis*. 25th International Cartographic Conference, Paris.
Opach, T., Popelka, S., Dolezalova, J., & Rod, J. K. (2017). Star and Polyline Glyphs in a Grid Plot and on a Map Display: Which Perform Better? *Cartography and Geographic Information Science*, *45*(5), 400-419. https://doi.org/10.1080/15230406.2017.1364169
Opach, T., & Rød, J. K. (2013). Cartographic visualization of vulnerability to natural hazards. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *48*(2), 113-125. https://doi.org/10.3138/carto.48.2.1840
Opach, T., & Rød, J. K. (2018). Augmenting the usability of parallel coordinate plot: The polyline glyphs. *Information Visualization*, *17*(2), 108-127. https://doi.org/10.1177/1473871617693041
Pelgrim, M. H., Espinosa, J., & Buchsbaum, D. (2023). Head-mounted mobile eye-tracking in the domestic dog: a new method. *Behavior research methods*, *55*(4), 1924-1941. https://doi.org/10.3758/s13428-022-01907-3
Peterson, M. P. (1998). That interactive thing you do. *Cartographic Perspectives*(29), 3-5. https://doi.org/10.14714/CP29.671

Petchenik, B. B. (1977). Cognition in cartography. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *14*(1), 117-128.
Popelka, S. (2014). *Optimal eye fixation detection settings for cartographic purposes*. 14th SGEM GeoConference on Informatics, Geoinformatics and Remote Sensing, Albena, Bulgaria.
Popelka, S. (2015). *Hodnocení 3D vizualizací v GIS s využitím sledování pohybu očí* (Terra Notitia ed.). Univerzita Palackého v Olomouci.
Popelka, S. (2018a). *Eye-tracking (nejen) v kognitivní kartografii - praktický průvodce tvorbou a vyhodnocením experimentu*. Univerzita Palackého v Olomouci. https://doi.org/10.5507/prf.18.24453132
Popelka, S. (2018b). Eye-tracking analysis of the influence of relief shading on finding labels on tourist maps. *Geografie*, *123*(3), 353-378. https://doi.org/10.37040/geografie2018123030353
Popelka, S. (2018c). *Eye-tracking evaluation of 3D thematic maps*. Proceedings of the 3rd Workshop on Eye Tracking and Visualization.
Popelka, S. (2018d). Eye-Tracking Evaluation of Non-Photorealistic Maps of Cities and Photo-Realistic Visualization of an Extinct Village. In R. A. White, A. Coltekin, & R. R. Hoffman (Eds.), *Remote Sensing and Cognition Human Factors in Image Interpretation* (pp. 87-110). CRC Press. https://doi.org/10.1201/9781351040464-5
Popelka, S., & Beitlova, M. (2022). *Scanpath Comparison using ScanGraph for Education and Learning Purposes: Summary of previous educational studies performed with the use of ScanGraph*. 2022 Symposium on Eye Tracking Research and Applications, Seattle, WA, USA. https://doi.org/10.1145/3517031.3529243
Popelka, S., & Brychtová, A. (2013). Eye-tracking Study on Different Perception of 2D and 3D Terrain Visualisation [Article]. *The Cartographic Journal*, *50*(3), 240-246. https://doi.org/10.1179/1743277413y.0000000058
Popelka, S., Burian, J., & Beitlova, M. (2022). Swipe versus multiple view: a comprehensive analysis using eye-tracking to evaluate user interaction with web maps. *Cartography and Geographic Information Science*, *49*(3), 252-270. https://doi.org/10.1080/15230406.2021.2015721
Popelka, S., & Dedkova, P. (2014). *Extinct village 3D visualization and its evaluation with eye-movement recording*. Computational Science and Its Applications – ICCSA 2014. http://www.scopus.com/inward/record.url?eid=2-s2.0-84904893639&partnerID=40&md5=4bfbea1154288e61abd7dcc7a59edde4
Popelka, S., & Dolezalova, J. (2016). *Differences between 2D map and virtual globe containing point symbols an eye-tracking study*. Informatics, Geoinformatics and Remote Sensing Conference Proceedings, SGEM, Albena, Bulgaria. <Go to ISI>://WOS:000395499600023
Popelka, S., Dolezalova, J., & Beitlova, M. (2018). *New features of scangraph: a tool for revealing participants' strategy from eye-movement data*. ACM Symposium on Eye Tracking Research & Applications.
Popelka, S., Herman, L., Řezník, T., Pařilová, M., Jedlička, K., Bouchal, J., Kepka, M., & Charvát, K. (2019). User Evaluation of Map-Based Visual Analytic Tools. *ISPRS International Journal of Geo-Information*, *8*(8), 363. https://doi.org/10.3390/ijgi8080363
Popelka, S., Kominek, J., & Vojtechovska, M. (2024). *Exploring Geological Map Usability Through Sequence Chart Visualization*. Proceedings of the 2024 Symposium on Eye Tracking Research and Applications.
Popelka, S., Senkerik, J., Vanicek, T., & Beitlova, M. (2023). *Visualizing Eye Movement Data of Web Maps: The ET2Spatial & ET2GIS & ET2QGIS Tools*. 18th International Conference on Location Based Services.
Popelka, S., Stachoň, Z., Šašinka, Č., & Doležalová, J. (2016). EyeTribe Tracker Data Accuracy Evaluation and Its Interconnection with Hypothesis Software for Cartographic Purposes. *Computational Intelligence and Neuroscience*, *2016*, 1-14, Article 9172506. https://doi.org/10.1155/2016/9172506

Popelka, S., Vojtechovska, M., Ruzicka, O., Beitlova, M., & Stachon, Z. (2024). *Adapting to the mobile majority: A new approach to interactive map usability assessment*. EuroCarto2024 Workshop: Online User Experiments, Vienna.

Popelka, S., Vondrakova, A., & Hujnakova, P. (2019). Eye-Tracking Evaluation of Weather Web Maps *ISPRS International Journal of Geo-Information*, *8*(6), 29. https://doi.org/10.3390/ijgi8060256

Popelka, S., & Vysloužil, J. (2024). *Exploring visitor engagement through eye-tracking analysis at a science exhibition*. 2nd European Conference on Eye Movements, Maynooth, Ireland.

Porti Suarez, A., & Popelka, S. (2023). Design Aspects for COVID-19 Dashboards–Evidence from Eye-Tracking Evaluation. *International Journal of Human–Computer Interaction*, 1-19. https://doi.org/10.1080/10447318.2023.2272078

Privitera, C. M., & Stark, L. W. (2000). Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(9), 970-982. https://doi.org/10.1109/34.868698

Punch, K. (2008). *Základy kvantitativního šetření*. Portál.

Raiha, K.-J., Aula, A., Majaranta, P., Rantala, H., & Koivunen, K. (2005). Static visualization of temporal eye-tracking data (*Human-Computer Interaction-INTERACT 2005* (Vol. 3585, pp. 946-949). Springer. https://doi.org/10.1007/11555261_76

Ratajski, L. (1978). The main characteristics of cartographic communication as a part of theoretical cartography. *International Yearbook of Cartography*, *18*, 21-32.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372-422. https://doi.org/10.1037/0033-2909.124.3.372

Robinson, A. H. (1952). *The Look of Maps*. University of Wisconsin press.

Robinson, A. H., & Petchenik, B. B. (1976). *The nature of maps: Essays toward understanding maps and mapping*. University of Chicago Press

Rohrer, C. (2014). *When to Use Which User-Experience Research Methods*. Nielsen Norman Group https://www.nngroup.com/articles/which-ux-research-methods/

Roth, R. E. (2012). Cartographic interaction primitives: Framework and synthesis. *The Cartographic Journal*, *49*(4), 376-395. https://doi.org/10.1179/1743277412Y.0000000019

Roth, R. E., & MacEachren, A. M. (2016). Geovisual analytics and the science of interaction: an empirical interaction study. *Cartography and Geographic Information Science*, *43*(1), 30-54. https://doi.org/10.1080/15230406.2015.1021714

Růžička, O. (2012). *Profil uživatele webových map*, Palacký University Olomouc].

Sack, C. M., & Roth, R. E. (2017). Design and evaluation of an Open Web Platform cartography lab curriculum. *Journal of Geography in higher Education*, *41*(1), 1-23. https://doi.org/10.1080/03098265.2016.1241987

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In A. Duchowski, J. W. Senders, & K. S. Kam (Eds.), *Symposium on Eye tracking research & applications (ETRA 2000)* (pp. 71-78). Association for Computing Machinery,. https://doi.org/10.1145/355017.355028

Sauro, J. (2023). *Sample Size Calculator for Discovering Problems in a User Interface*. MeasuringU https://measuringu.com/calculators/problem_discovery/

Sauro, J., & Lewis, J. R. (2016). *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann.

Sauter, D., Martin, B., Di Renzo, N., & Vomscheid, C. (1991). Analysis of eye tracking movements using innovations generated by a Kalman filter. *Medical and biological Engineering and Computing*, *29*(1), 63-69. https://doi.org/10.1007/BF02446297

Sedrakyan, G., Mannens, E., & Verbert, K. (2019). Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. *Journal of Computer Languages*, *50*, 19-38. https://doi.org/10.1016/j.jvlc.2018.11.002

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, *27*(3), 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
Sibert, J. L., Gokturk, M., & Lavine, R. A. (2000). *The reading assistant: eye gaze triggered auditory prompting for reading remediation*. 13th annual ACM symposium on User interface software and technology, San Diego, CA, USA.
Skrabankova, J., Popelka, S., & Beitlova, M. (2020). Students' ability to work with graphs in physics studies related to three typical student groups. *Journal of Baltic Science Education*, *19*(2), 298-316. https://doi.org/10.33225/jbse/20.19.298
Snowden, R., Snowden, R. J., Thompson, P., & Troscianko, T. (2012). *Basic vision: an introduction to visual perception*. Oxford University Press.
Steinke, T. R. (1987). Eye movement studies in cartography and related fields. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *24*(2), 40-73. https://doi.org/10.3138/Ag15-V29X-4757-36J5
Stofer, K., & Che, X. (2014). Comparing experts and novices on scaffolded data visualizations using eye-tracking. *Journal of Eye Movement Research*, *7*(5). https://doi.org/10.16910/jemr.7.5.2
Sultan, M. N., Popelka, S., & Strobl, J. (2022). ET2Spatial–software for georeferencing of eye movement data. *Earth Science Informatics*, 1-19. https://doi.org/https://doi.org/10.1007/s12145-022-00832-5
Sundstedt, V. (2012). *Gazing at games: An introduction to eye tracking control* (Vol. 14). Morgan & Claypool Publishers.
Svoboda, M. (1999). *Psychologická diagnostika dospělých*. Portál.
Synek, S., & Skorovská, Š. (2014). *Fyziologie oka a vidění*. Grada Publishing, a.s.
Šašinka, Č., Morong, K., & Stachoň, Z. (2017). The Hypothesis platform: An online tool for experimental research into work with maps and behavior in electronic environments. *ISPRS International Journal of Geo-Information*, *6*(12), 407. https://doi.org/10.3390/ijgi6120407
Šašinka, Č., Stachoň, Z., Čeněk, J., Šašinková, A., Popelka, S., Ugwitz, P., & Lacko, D. (2021). A comparison of the performance on extrinsic and intrinsic cartographic visualizations through correctness, response time and cognitive processing. *PLoS One*, *16*(4), e0250164. https://doi.org/10.1371/journal.pone.0250164
Šašinka, Č., Stachoň, Z., Kubíček, P., Tamm, S., Matas, A., & Kukaňová, M. (2019). The Impact of global/local bias on task-solving in map-related tasks employing extrinsic and intrinsic visualization of risk uncertainty maps. *The Cartographic Journal*, *56*(2), 175-191. https://doi.org/10.1080/00087041.2017.1414018
Štěrba, Z., Šašinka, Č., Stachoň, Z., Štampach, R., & Morong, K. (2015). *Selected Issues of Experimental Testing in Cartography*. Masarykova Univerzita.
Takatsuka, M., & Gahegan, M. (2002). GeoVISTA Studio: A codeless visual programming environment for geoscientific data analysis and visualization. *Computers & Geosciences*, *28*(10), 1131-1144. https://doi.org/10.1016/S0098-3004(02)00047-0
Thorndyke, P. W., & Stasz, C. (1980). Individual differences in procedures for knowledge acquisition from maps. *Cognitive Psychology*, *12*(1), 137-175. https://doi.org/10.1016/0010-0285(80)90006-7
Tinker, M. A. (1946). The study of eye movements in reading. *Psychological bulletin*, *43*(2), 93-120. https://doi.org/10.1037/h0063378
Ünlü, A., & Malik, W. A. (2011). Interactive Glyph Graphics of Multivariate Data in Psychometrics. *Methodology*, *7*(4). https://doi.org/10.1027/1614-2241/a000031
van der Geest, J. N., Kemner, C., Camfferman, G., Verbaten, M. N., & van Engeland, H. (2001). Eye movements, visual attention, and autism: A saccadic reaction time study using the gap and overlap paradigm. *Biological Psychiatry*, *50*(8), 614-619. https://doi.org/10.1016/S0006-3223(01)01070-8
Vanicek, T., & Popelka, S. (2023). The think-aloud method for evaluating the usability of a regional atlas. *ISPRS International Journal of Geo-Information*, *12*(3), 95. https://doi.org/10.3390/ijgi12030095

Virrantaus, K., Fairbairn, D., & Kraak, M.-J. (2009). ICA research agenda on cartography and GIScience. *Cartography and Geographic Information Science*, *36*(2), 63-75. https://doi.org/10.1179/000870409X459824

Vojtechovska, M., & Popelka, S. (n.d.). GazePlotter: an open-source solution for the automatic generation of scarf plots from eye-tracking data. *Behavior research methods*, *Under review*.

Voßkühler, A., Nordmeier, V., Kuchinke, L., & Jacobs, A. M. (2008). OGAMA (Open Gaze and Mouse Analyzer): open-source software designed to analyze eye and mouse movements in slideshow study designs. *Behavior research methods*, *40*(4), 1150-1162. https://doi.org/10.3758/BRM.40.4.1150

Wang, S., Chen, Y., Yuan, Y., Ye, H., & Zheng, S. (2016). Visualizing the Intellectual Structure of Eye Movement Research in Cartography. *ISPRS International Journal of Geo-Information*, *5*(10), 1-22. https://doi.org/10.3390/ijgi5100168

Wedel, M., & Pieters, R. (2008). A review of eye-tracking research in marketing (*Review of marketing research* (Vol. 4, pp. 123-147). Emerald Group Publishing Limited. https://doi.org/10.1108/S1548-6435(2008)0000004009

Wehrend, S., & Lewis, C. (1990). *A problem-oriented classification of visualization techniques*. Proceedings of the 1st Conference on Visualization'90.

West, J. M., Haake, A. R., Rozanski, E. P., & Karn, K. S. (2006). *eyePatterns: software for identifying patterns and similarities across fixation sequences*. Symposium on Eye tracking research & applications (ETRA 2006), San Diego, California, USA.

Williams, L. (1971). The Role Of The User In The Map Communication Process: Obtaining Information From Displays With Discrete Elements. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *8*(2), 29-34. https://doi.org/10.3138/A724-2K5V-2887-P20

Wright, J. K. (1942). Map makers are human: Comments on the subjective in maps. *Geographical review*, *32*(4), 527-544. https://doi.org/10.2307/209994

Wu, M. M., & Munzner, T. (2015). SEQIT: visualizing sequences of interest in eye tracking data. *IEEE TVCG*, *22*(1), 449-458.

Yamamoto, T., & Imai-Matsumura, K. (2013). Teachers' gaze and awareness of students' behavior: using an eye tracker. *Comprehensive Psychology*, *2*, 1-7. https://doi.org/10.2466/01.IT.2.6

Yang, C.-K., & Wacharamanotham, C. (2018). *Alpscarf: Augmenting scarf plots for exploring temporal gaze patterns*. Extended abstracts of the 2018 CHI conference on human factors in computing systems.

Yarbus, A. L., Haigh, B., & Rigss, L. A. (1967). *Eye movements and vision* (Vol. 2). Plenum press

Young, K., Regan, M., & Hammer, M. (2007). *Driver distraction: A review of the literature*. International Conference on Driver Distraction and Inattention.

Young, L. R., & Sheena, D. (1975). Eye-movement measurement techniques. *American Psychologist*, *30*(3), 315-330. https://doi.org/10.1037/0003-066X.30.3.315

SUPPLEMENTS

The list of attached publications

Paper 2D-3D

Popelka, S.*, & Brychtova, A. (2013). Eye-tracking Study on Different Perception of 2D and 3D Terrain Visualisation. Cartographic Journal, 50(3), 240-246.
https://doi.org/10.1179/1743277413y.0000000058

[IF 0.299; Q4]

[63 citations on WoS; 56 citations without autocitations]

[Author 's contribution: 70%]

Abstract

The use of computer-generated perspective views, often named as three-dimensional (3D) maps, is growing. These terrain visualisations should be more understandable for users without cartographic education, which are not familiar with contour lines. Within the study, two eye-tracking experiments and online questionnaire were used for investigating the difference between user cognition of classical two-dimensional (2D) visualisation with contour lines and perspective 3D view. Questionnaire was focused on maps understandability, suitability and aesthetics. Results of the questionnaire shows, that the majority of participants prefer 3D visualisation. First eye-tracking experiment was designed as a pair of maps in one stimulus. One shows 2D visualisation, the other 3D visualisation. No significant differences between user preferences of 2D and 3D visualisation were found, but the results were influenced with the order of the maps in the stimuli. Because of that another experiment was designed. In this case stimuli contained only one of two possible visualisations (2D and 3D). ScanPath comparison of this experiment results confirmed that users have different strategies for cognition of 2D and 3D visualization, although a statistically significant difference between both types of visualization was found in the ScanPath length metric only.

The Cartographic Journal *Vol. 50 No. 3 pp. 240–246 International Cartographic Conference 2013, Dresden – Special Issue August 2013*

REFEREED PAPER

Eye-tracking Study on Different Perception of 2D and 3D Terrain Visualisation

Stanislav Popelka and Alzbeta Brychtova

Palacký University in Olomouc, Czech Republic
Email: standa.popelka@gmail.com

The use of computer-generated perspective views, often named as three-dimensional (3D) maps, is growing. These terrain visualisations should be more understandable for users without cartographic education, which are not familiar with contour lines. Within the study, two eye-tracking experiments and online questionnaire were used for investigating the difference between user cognition of classical two-dimensional (2D) visualisation with contour lines and perspective 3D view. Questionnaire was focused on maps understandability, suitability and aesthetics. Results of the questionnaire shows, that the majority of participants prefer 3D visualisation. First eye-tracking experiment was designed as a pair of maps in one stimulus. One shows 2D visualisation, the other 3D visualisation. No significant differences between user preferences of 2D and 3D visualisation were found, but the results were influenced with the order of the maps in the stimuli. Because of that another experiment was designed. In this case stimuli contained only one of two possible visualisations (2D and 3D). ScanPath comparison of this experiment results confirmed that users have different strategies for cognition of 2D and 3D visualisation, although statistically significant difference between both types of visualisation was found in the ScanPath length metric only.

Keywords: terrain, 3D, eye-tracking, visualisation

INTRODUCTION

Today, computer-generated perspective views of cartographic content, often named as three-dimensional (3D) maps, are widespread. The development of geoinformation technologies has facilitated the creation of graphic representations and therefore they are used not only by experts in the field of geoinformatics or cartography, but also by the general public. As the Haeberling (2005) states, perspective perception of a generalized and symbolized geographic space often offers a better understanding of spatial coherences. 3D maps could be seen as a supporting complement of classic orthogonal maps (Brychtová and Popelka, 2011).

The need to present 3D cartographic content on computer monitors is growing and the possibilities for these presentations are increasing (Buchroithner *et al.*, 2011).

In the history it can be traced tendency of using some typical graphic techniques aimed to express the relief of the landscape as realistic as it is possible (Petrovic and Masera, 2006). Usually it was perspective visualisation suggesting 3D effect applying shading, cross-hatching or hill symbols.

Many inexperienced map users have troubles to read two-dimensional (2D) topographic maps that typically depict landscape features with contour lines, shaded relief and height points. To help these users, cartographers have increasingly turned to 3D perspective maps, which allow users to more easily visualize 3D landscapes (Schobesberger and Patterson, 2008).

Three-dimensional map contains both semantic and geometric description of the captured area (Zebedin *et al.*, 2006) and its visualisation in 3D environment gives the user a better idea of space, especially height proportions.

Haeberling (2002) mentions that term '3D map' is not found in the cartographic literature. Although they possess cartographic characteristics, 3D maps should be considered a map-related representation, not a map in the classic sense.

EVALUATION OF DIFFERENCES BETWEEN 2D AND 3D MAPS

Several studies focused on finding user preferences of different visualisations of terrain were done so far. Most of them used questionnaire as the main evaluation tool. However, several another approaches of the research on user perception and evaluation of the applicability and effectiveness of maps exist. Among methods of so called usability studies the eye-tracking belongs.

Evaluation with use of traditional methods

In a research project at the Institute of Cartography, ETH Zürich, Haeberling (2004) evaluated design variables (the

DOI: 10.1179/1743277413Y.0000000058

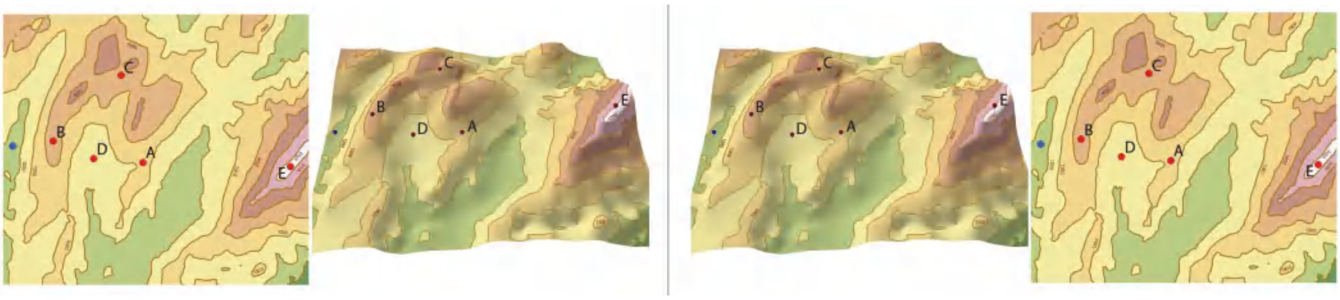


Figure 1. Two variants of the same stimuli with different orders of 2D and 3D map on it. 2D3D (left) and 3D2D (right)

inclination angle of the viewing direction; the viewing distance; the horizontal lighting direction; the sky structure and haze density) for 3D maps.

Schobesberger and Patterson (2008) investigated differences between 2D and 3D map of the Zion National Park in Utah. They used two methods for collecting data – trailheads exhibits and questionnaires. In the case of map attraction, 3D map was more successful.

Petrovic and Masera (2006) also used questionnaire to analyze user's preferences on 2D and 3D maps. Respondents were asked to decide, which type of map they will use for four different tasks: distance measurement, height difference measurement, defining North direction, impression about the route.

Savage *et al.* (2004) tried to answer a question, if the integration of all three dimensions in a perspective view provides an advantage for spatial visualisation over the traditional 2D topographic map. She used a questionnaire for two groups of randomly divided respondents. In the results of the study, there was no apparent advantage of 3D map for those tasks requiring elevation information, nor was there a disadvantage for integrated tasks which did not require elevation information.

Evaluation with use of eye-tracking

Eye-tracking has not been fully implemented in the field of 3D cartography. There exist some examples of studies, where cartographic outputs or maps were evaluated with use of eye-tracking.

Most of these studies are focused to evaluation or analysis of 'traditional' orthogonal maps (Steinke, 1987; Coltekin *et al.*, 2009; Fabrikant *et al.*, 2008; Ooms *et al.*, 2011; Opach and Nossum, 2011; Fuchs *et al.*, 2009).

The issue of 3D map visualisation was investigated by Fuhrmann *et al.* (2009). They have dealt with an assumption that 3D topographic maps provide more effective route planning, navigation, orientation, and way-finding results than traditional 2D representations. The eye-tracking metrics analysis indicates with a high statistical level of confidence that 3D holographic maps enable more efficient route planning.

Irvankoski *et al.* (2012) presents an eye-movement study on visualisation of elevation information on maps. Participants had to complete four map-related tasks (search a symbol, compare heights between two points, select a hiking area and plan a route between two points). Three types of elevation visualisation were available and authors investigate differences in fixation durations.

CASE STUDY

The aim of the case study was to analyze differences between cognition of classical orthogonal maps and their equivalents made with use of 3D visualisation. A questionnaire and two eye-tracking experiments were used for the analysis of user preferences.

Experiment design

For the study, remote eye-tracking device SMI RED 250, developed by SensoMotoric Instruments, was used. This device operates at a frequency of 120 Hz.

A total number of 40 respondents had participated within this eye-tracking study. Half of them were selected from the group of undergraduate students, who already attended cartography course. The rest of them were selected from students of different fields of study than cartography. Differences between cartographers and non-cartographers can be investigated.

DualMap experiment and questionnaire

In the first test, called DualMap, stimuli were designed as a pair of maps in 2D and 3D side by side. The aim was to reveal, which kind of visualisation will be preferred when searching for answer on spatial query.

A total of 11 image stimuli were used in the experiment. For purpose of this paper, just five of them (Stimuli 4, 5, 6, 7, 8) will be mentioned. Terrain visualisations were created in Esri ArcMap and ArcScene. The work of Savage *et al.* (2004) was used as an inspiration.

Before each stimulus, the respondents had 30 seconds to read and remember the task. After that, the fixation cross was presented for 600 ms to ensure that all respondents started from the centre of the stimuli. Then, the stimuli with the map were projected for 60 seconds. The respondents had to answer the question with use of mouse click directly into the map.

To avoid the influence of the location of maps within the stimulus (left, right), two groups of users were tested. Stimuli for both groups were the same, but the position of 2D and 3D maps within the stimulus was changed. On the first one, 2D map was presented on the left side, 3D on the right. On the second version vice versa. Example of the stimuli is shown in Figure 1.

After the eye-tracking experiment, each respondent was asked to fill an online questionnaire. Questions were focused on user subjective attitudes to both visualisation

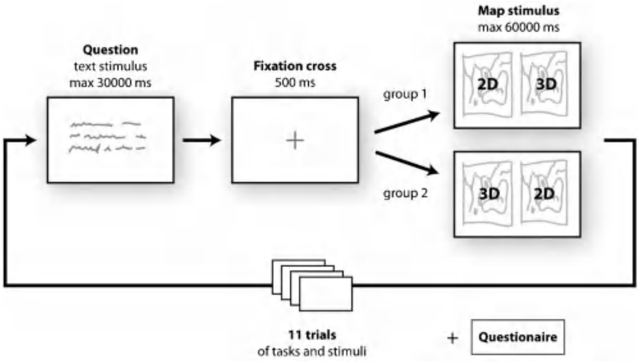


Figure 2. Process of testing. Respondents had 30 seconds to read the question and after short fixation cross, they had 60 seconds to find the answer. After finishing of eye-tracking testing, they fulfill short questionnaire about maps

methods. The whole process of testing is depicted in Figure 2.

SingleMap experiment

Second eye-tracking experiment (SingleMap) contains a total of 15 stimuli. Four of them were selected for an analysis, because they contain the similar terrain visualisation like stimuli presented in above mentioned DualMap experiment.

In this experiment, each stimulus contains one map. Stimuli 1, 2 and 3, 4 form pairs with similar maps. Stimuli 1 and 3 contain 2D visualisation, stimuli 2 and 4 3D visualisation (see Figure 3).

Stimuli were presented in random order. For avoiding the learning effect, map with 3D visualisation was inverted or the question was modified a bit. Whole process of experiment is depicted in Figure 4.

ANALYSES OF QUESTIONNAIRE

At the beginning of data analyses, the answers from questionnaire were investigated. The questionnaire contained images of all stimuli from the DualMap experiment together with the associated question. Respondents had to decide four different tasks for each map.

- Was the 2D map understandable?
- Was the 3D map understandable?
- Was the 3D map more suitable for finding the answer than 2D map?
- Was the 3D map more aesthetic than 2D map?

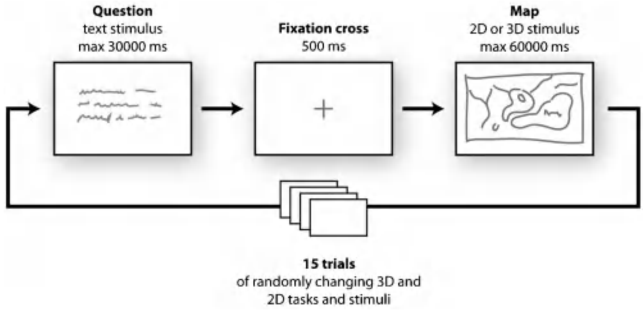


Figure 4. Process of testing. Respondents had 30 seconds to read the question and after short fixation cross, they had 60 seconds to find the answer

Questionnaire data were tested with a Wilcoxon rank sum test and statistically significant difference between 2D and 3D maps was observed on significance level $\alpha=0.05$ in all three cases (understandability, suitability and aesthetics). Results are shown in Table 1. and graphs in Figures 5–7.

In the third question, respondents had to decide, if 3D map was more suitable for finding the right answer than 2D map. In 10 cases from 11, majority of them choose answer 'Yes'. The only exception is question Q5, where they had to compare the distances between points. For this question, 2D visualisation was more suitable.

In the last question, in almost all cases, more than 60% of respondents choose that 3D map was more aesthetic than 2D map.

Data displayed in all three graphs are related to all respondents regardless of the membership to the group of cartographers and non-cartographers.

Differences between group of cartographers and non-cartographers were tested with use of Wilcoxon rank sum test. On the significance level $\alpha=0.05$, the difference between group of cartographers and non-cartographers was not statistically significant.

Investigation of questionnaire proved, that there exists statistically significant difference between preference of 2D and 3D map visualisation.

Table 1. Wilcoxon test of differences between 2D and 3D map visualisation

	Alpha	W	p-value	statement
Map Understandability	0.05	15	0.005389	Rejecting H0
Map Suitability	0.05	118	0.0001555	Rejecting H0
Map Aesthetic	0.05	211	6.777e-05	Rejecting H0

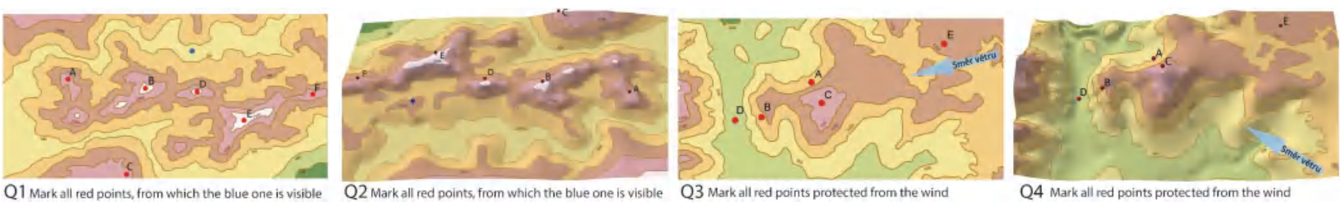


Figure 3. Experiment stimuli from the SingleMap experiment

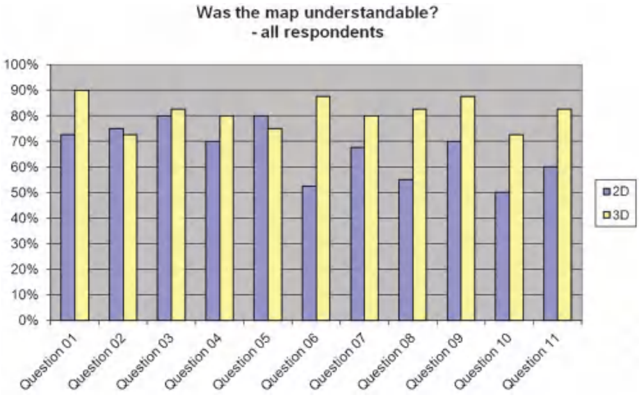


Figure 5. Graph summarizing percentage of positive answers about understandability of 2D and 3D visualisation

ANALYSES OF EYE-TRACKING METRICS

Next task was to compare this finding with eye-tracking metrics and investigate, if there is any difference between 2D and 3D visualisation cognition.

Eye-tracking metrics, such as fixation duration, saccade amplitude, ScanPath length or dwell time, are derived from basic eye-events (fixations and saccades). From statistical analysis of these quantitative characteristics is possible to indicate respondent's tactics or cognitive load during solving the task with a map.

Analysis of DualMap experiment

For the analysis of DualMap experiment, the 'Dwell Time' metric was chosen. Dwell time can be defined as the duration of one visit in concrete area of interest (AOI), from entry to the exit (Holmqvist *et al*., 2011). According to Mello-Thoms *et al.* (2004), dwell is the sum of all fixation durations within a prescribed area. It is best used to compare attention distributed between targets.

Within the stimuli only two AOI representing 2D and 3D map were marked. The statistical analysis was then used to examine whether there is a difference between how users viewed part of the stimulus with 3D and 2D map.

On the significance level α=0.05 no statistically significant difference for any of the studied stimulus was found (see results in Table 2).

In the next step, it was tested, if the value of dwell time is not influenced by order of the maps in the stimuli. Generally speaking, the image on the left side should be

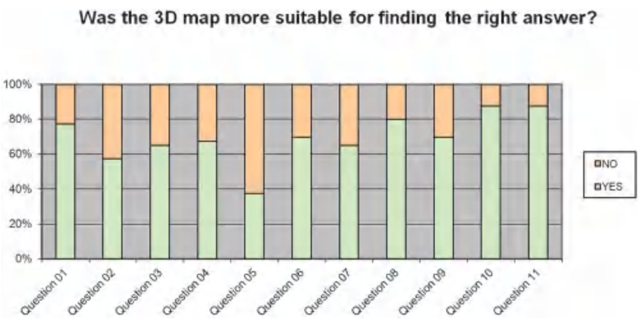


Figure 6. Graph summarizing percentage of positive answers about understandability of 2D and 3D visualisation

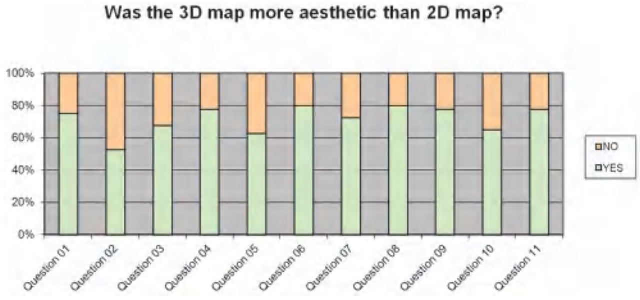


Figure 7. Graph summarizing respondent answers about an aesthetics of 3D map visualisation

preferred more, because participants are used to read text from the left side. The boxplot (Figure 8) shows dwell time values for 2D and 3D AOI for each stimuli. The order of the AOI in the stimuli is described by '2D-3D' and '3D-2D' label.

Differences between dwell time values based on the order of maps in the stimuli were also tested with use of Wilcoxon rank sum test. There was found a statistically significant difference in the half of observations.

Order of maps in the stimuli influenced the value of Dwell time in stimuli 4, 6 and 8(see Tables 3 and 4).

These results indicate that DualMap experiment design was not suitable, because respondent cognition is influenced with the order of the maps in the stimuli more than differences between 2D and 3D visualisation.

For further analyses, use of SingleMap experiment should be more appropriate.

Analysis of SingleMap experiment

For analysis of SingleMap experiment, four stimuli were selected and 'ScanPath length' was analyzed in this case. It is described as the length of the gaze fixation connections in pixels. According to Goldberg *et al.* (2002), a longer ScanPath indicates less efficient searching.

Contrast to this assertion, group of cartographers has longer ScanPaths than non-cartographers in all cases (see Figure 9).

Wilcoxon test was used again to investigate differences between ScanPath lengths for '2D' and '3D' maps (always the pair of maps – 1–2 and 3–4). Statistically significant difference was found between map 3 and 4 (in case of non-cartographers, and generally if neglecting groups; see Table 5).

No differences were found in case of using other eye-tracking metrics (fixation duration mean, fixation count, trial duration).

Table 2. Wilcoxon test of differences between dwell time in 2D and 3D map visualisation

2D vs. 3D AOI	Alpha	W	p-value	statement
Trial 004	0.05	842.0	0.6861	Failed to reject
Trial 005	0.05	943.5	0.1673	Failed to reject
Trial 006	0.05	748.5	0.6202	Failed to reject
Trial 007	0.05	772.0	0.7876	Failed to reject
Trial 008	0.05	740.5	0.5669	Failed to reject

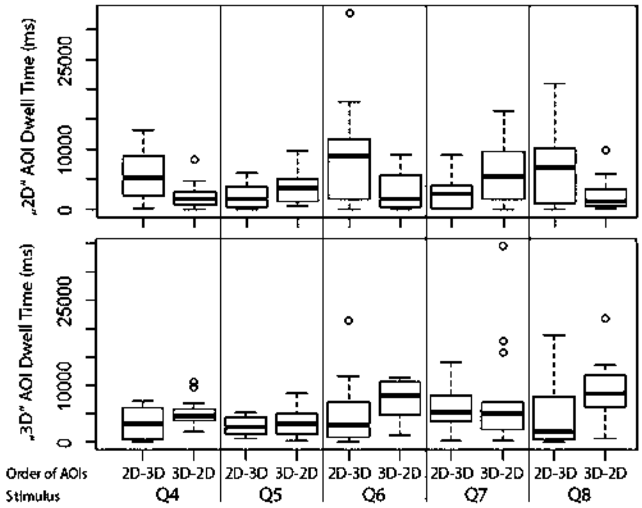


Figure 8. Boxplots of dwell time values for 2D and 3D AOI for each stimuli

Selected stimuli were very similar (both are maps, only with slightly different visualisations), so the eye-movement characteristics are almost the same in both examples (2D and 3D). If we use completely different stimuli (i.e. text versus image), the difference between these metrics will be perhaps more significant.

The fact, that there were found almost no differences, does not mean that participant cognition strategy is the same for both stimuli. User strategy can be revealed better with use of ScanPath comparison.

Privitera and Stark (2000) introduced ScanPath comparison based on string editing. Fixations are replaced with characters standing for the AOIs they hit and the ScanPath is represented as character string.

The principle of this method is the transformation of 2D data (*X*, *Y* coordinates of fixations) to one-dimensional data (character string). Two or more character strings are then compared and their similarity is measured. String edit algorithm determines the number of operations (insertions, deletions and substitutions) needed to transform one sequence to another (Popelka *et al.*, 2012).

Table 3. Wilcoxon test of differences between 2D3D and 3D2D order of maps. Values for '2D' AOI

2D: 2D3D vs. 3D2D	Alpha	W	p-value	statement
Trial 004	0.05	157	0.006702	Rejecting H0
Trial 005	0.05	57	0.06204	Failed to reject
Trial 006	0.05	144	0.03445	Rejecting H0
Trial 007	0.05	66	0.1409	Failed to reject
Trial 008	0.05	138	0.06905	Failed to reject

Table 4. Wilcoxon test of differences between 2D3D and 3D2D order of maps. Values for '3D' AOI

3D: 2D3D vs. 3D2D	Alpha	W	p-value	statement
Trial 004	0.05	52	0.03453	Rejecting H0
Trial 005	0.05	92	0.8036	Failed to reject
Trial 006	0.05	55	0.04974	Rejecting H0
Trial 007	0.05	97	0.982	Failed to reject
Trial 008	0.05	52	0.03504	Rejecting H0

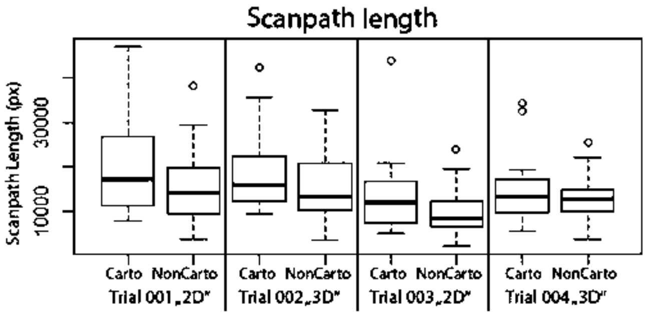


Figure 9. Boxplots of ScanPath length values for group of cartographers and non-cartographers for each stimuli

ScanPath comparison was used for stimuli 1 and 2 from the SingleMap experiment. In each stimuli, a rounded AOI around each point in the map was created (see Figure 10). ScanPath strings were generated with use of open-source application OGAMA and string comparison was visualized via tree-graph.

Software eyePatterns can visualize the string matrix by displaying clusters of similar sequences. Any 2 nodes that come off of the tree in a 'V' in Figure 11 were clustered together. At some point in the clustering process, those two nodes were the most similar items (Adjusted from West *et al.*, 2006).

Figure 11 shows visualisation of ScanPath similarity for 40 participants and two stimuli (1 and 2). On both maps, the same AOIs were created.

The colours in the graph are distinguished according to stimulus 1 (red) and stimulus 2 (blue). The groups with the same colour are visible from the graph. For example in the upper part of an image, there is a branch with only red labels. In the bottom part, another large branch contains 16 blue labels and only two red.

From this result, the different participant strategies for two stimuli are visible.

CONCLUSION

The study was focused to the finding of differences between 2D and 3D visualisation of the terrain. Two eye-tracking experiments, each with 40 respondents was created.

First experiment was called DualMap experiment and it contains both visualisation methods together in one

Table 5. Wilcoxon test of differences between pairs of maps (2D vs 3D)

Scanpath length	Alpha	W	p-value	statement
T01 vs. T02 Carto	0.05	198	0.968	Failed to reject
T03 vs. T04 Carto	0.05	168	0.3983	Failed to reject
T01 vs. T02 NonCarto	0.05	198	0.968	Failed to reject
T03 vs. T04 NonCarto	0.05	127	0.04909	Rejecting H0

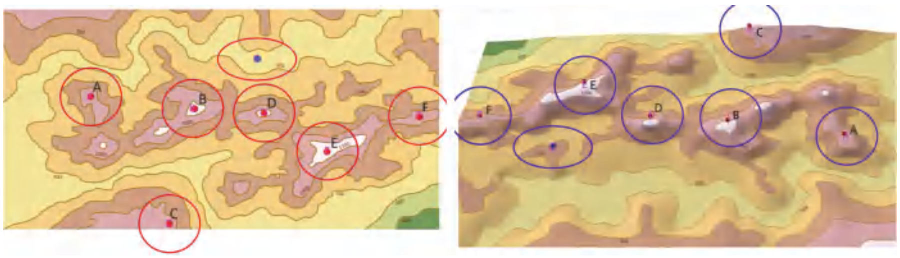


Figure 10. Location of AOIs in the stimuli 1 (left) and 2 (right). AOI has the same size in both maps

stimulus. In the second experiment, called SingleMap, each stimulus contains just one map with 2D or 3D visualisation.

Study participants also fill out the questionnaire, in which they answer the questions about each map. Questions were focusing on map understandability, suitability and aesthetics.

Analysis of the questionnaire results has shown that majority of participants prefer 3D visualisation. Next step was to verify this result with an eye-tracking data analysis.

Statistical analysis of DualMap experiment results has not proven significant differences between user perceiving of 2D and 3D visualisation. It was found that the order of maps in the stimuli influenced the results much more than the visualisation method.

In the SingleMap experiment, the differences between 2D and 3D visualisation were observed only in case of eye-tracking metric 'ScanPath length'. We assume that between both visualisation methods are so small visual differences, which will not be reflected on the length of the fixations or other metrics.

It does not mean that participant cognition strategy is the same for both visualisation methods. The different strategies for different visualisation methods were found with use of method of ScanPath comparison.

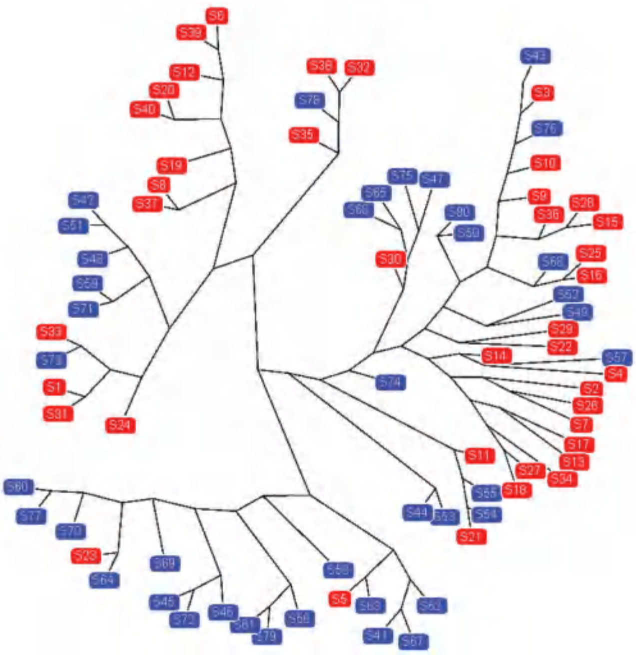


Figure 11. Tree graph of data from 40 participants and two stimuli – colours are distinguished according to the belonging to the stimulus (stimuli 1 – red; stimuli 2 – blue)

Nevertheless, further research will be needed to determine, which visualisation method is more suitable.

BIOGRAPHICAL NOTES

Stanislav Popelka is a PhD student at Department of geoinformatics, Palacký University in Olomouc, Czech Republic. His PhD thesis is focused on evaluation of various visualisation techniques with use of eye-tracking system. The main objective of his research is to analyze cartographical methods for visualisation of three-dimensional terrain data. He is a member of ICA Commission on Cognitive Visualisation.

ACKNOWLEDGEMENT

The paper has been completed within the project CZ.1.07/2.4.00/31.0010 'NeoCartoLink' and CZ.1.07/2.3.00/20.0170 'StatGisTeam' which are co-financed from European Social Fund and State financial resources of the Czech Republic

REFERENCES

Brychtová, A. and Popelka, S. (2011). 'Visualization of Spatio-Temporal Changes of The Olomouc City', **Proceedings of 6th International Workshop Digital Approaches to Cartographic Heritage ICC**, Haag, The Netherlands, Jul 3–8.

Buchroithner, M., Schenkel, R. and Kirschenbauer, S. (2000). '3D display techniques for cartographic purposes: Semiotic aspects', **International Archives of Photogrammetry and Remote Sensing**, XXXIII, Part B5.

Coltekin, A., Garlandini, S., Heil, B. and Fabrikant, S. I. (2009). 'Evaluating the Effectiveness of Interactive Map Interface Designs: A Case Study Integrating Usability Metrics with Eye-movement Analysis', **Cartography and Geographic Information Science**, 36, pp. 5–17.

Fabrikant, S. I., Rebich-Hespanha, S., Andrienko, N., Andrienko, G. and Montello, D. R. (2008). 'Novel method to measure inference affordance in static small-multiple map displays representing dynamic processes', **The Cartographic Journal**, 45, pp. 201–215.

Fuchs, S., Spachinger, K., Dorner, W., Rochman, J. and Serrhini, K. (2009). 'Evaluating cartographic design in flood risk mapping', **Environmental Hazards**, 8, pp. 52–70.

Fuhrmann, S., Komogortsev, O. and Tamir, D. (2009). 'Investigating Hologram-Based Route Planning', **Transactions in GIS**, pp. 177–196.

Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N. and Wichansky, A. M. (2002). 'Eye Tracking in Web Search Tasks: Design Implications', **Proceedings of the Eye Tracking Research and Applications Symposium**, pp. 51–58, New Orleans, LA, Mar 25–27.

Haeberling, Ch. (2002). '3D Map Presentation – A Systematic Evaluation of Important Graphic Aspects', **Proceedings of ICA Mountain Cartography Workshop 'Mount Hood'**, Mt. Hood, OR, May 15–19.

Haeberling, Ch. (2004). 'Topografische 3D-Karten ñ Thesen für kartografische Gestaltungsgrundsaetze', Dissertation, Institut für Kartographie der ETH Zuerich. Zuerich.

Haeberling, Ch. (2005). 'Cartographic Design Principles for 3D Maps – A Contribution to Cartographic Theory', **Proceedings of ICA Congress Mapping Approaches into a Changing World**, A Coruna, Spain, Jul 9–16.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., Van De Weijer, J. (2011). 'Eye tracking: A comprehensive guide to methods and measures', Oxford UP.

Irvankoski, K., Torniainen, J. and Krause, Ch. M. (2012). 'Visualisation of Elevation Information on Maps: An Eye Movement Study', **Conference Proceedings SWAET 2012, The Scandinavian Workshop on Applied Eye Tracking Karolinska Instituet**, Stockholm,May 2–4.

Mello-Thoms, C., Nodine, C. F. and Kundel, H. L. (2004). 'What Attracts the Eye to the Location of Missed and Reported Breast Cancers?' **Proceedings of the Eye Tracking Research and Applications Symposium**, pp. 111–117, San Antonio, TX, Mar 22–24.

Ooms, K., De Maeyer, K. and Fack, W. (2011). 'Can Experts Interpret a Map's Content More Efficiently?', **Proceedings of 25th International Cartographic Conference ICC**, Paris, France, Jul 3–8.

Opach, T. and Nossum, A. S. (2011). 'Evaluating the Usability of Cartographic Animations With Eye-Movement Analysis', **Proceedings of the 25th International Cartography Conference ICC**, Paris, France, Jul 3–8.

Popelka, S., Brychtová, A. and Brus, J. (2012). 'Advanced Map Optimalization Based on Eye-tracking', Cartography, InTech, Rieka, Croatia.

Petrovic, D. and Masera, P. (2006). 'Analysis of User'S Response on 3D Cartographic Presentations', **Proceedings of 5th ICA Mountain Cartography Workshop**, Bohinj, Slovenia, Mar 29–Apr 1.

Privitera, C. M. and Stark, L. W. (2000). 'Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations', **IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)**, 22, pp. 970–982.

Savage, D. M., Wiebe, E. N. and Devine, H. A. (2004). 'Performance of 2D versus 3D Topographic Representations for Different Task Types', **Proceedings of the Human Factors and Ergonomics Society Annual Meeting**, 48, pp. 1793–1797.

Schobesberger, D. and Patterson, T. (2008). 'Exploring of Effectiveness of 2D vs. 3D Trailhead Maps', **Proceedings of 6th ICA Mountain Cartography Workshop**, Lenk, Switzerland, Feb 11–15.

Steinke, T. R. (1987). 'Eye movement studies in cartography and related fields', **Cartographica: The International Journal for Geographic Information and Geovisualization**, 24, pp. 40–73.

West, J. M., Haake, A. H., Rozanski, E. P. and Karn, K. S. (2006). 'eyePatterns: Software for Identifying Patterns and Similarities Across Fixation Sequences', **Proceedings of the Symposium on Eye Tracking Research & Applications**, pp. 27–29, San Diego, CA, March 27–29.

Zebedin, L., Klaus, A., Gruber-Geymayer, B. and karner, K. (2006). 'Towards 3D map generation from digital aerial images', **ISPRS Journal of Photogrammetry and Remote Sensing**, 60, pp. 413–427.

<u>Paper ScanGraph</u>

Doležalová, J., & **Popelka, S.*** (2016). ScanGraph: A Novel Scanpath Comparison Method Using Visualisation of Graph Cliques. Journal of Eye Movement Research, 9(4), Article 5. https://doi.org/10.16910/jemr.9.4.5

[IF 1.295; Q3]

[36 citations on WoS; 22 citations without autocitations]

[Author's contribution: 50%]

Abstract

The article describes a new tool for analyses of eye-movement data. Many different approaches to scanpath comparison exist. One of the most frequently used approaches is String Edit Distance, where the gaze trajectories are replaced by the sequences of visited Areas of Interest. In cartographic literature, the most commonly used software for scanpath comparison is eyePatterns. During the analysis of eyePatterns functionality, we have found that tree-graph visualization of its results is not reliable. Thus, we decided to develop a new tool called ScanGraph. Its computational algorithms are modified to work better with the sequences with different lengths. The output is visualized as a simple graph, and similar groups of sequences are displayed as cliques of this graph. The article describes ScanGraph's functionality on the example of a simple cartographic eye-tracking study. Differences of the reading strategy of a simple map between cartographic experts and novices were investigated. The paper should serve to the researchers who would like to analyze differences between groups of participants, and who would like to use our tool - ScanGraph, available at www.eyetracking.upol.cz/scangraph.

Journal of Eye Movement Research
9(4):5, 1-13.

ScanGraph: A Novel Scanpath Comparison Method Using Visualisation of Graph Cliques

Jitka Dolezalova
Palacký University Olomouc,
Olomouc, Czech Republic

Stanislav Popelka
Palacký University Olomouc,
Olomouc, Czech Republic

The article describes a new tool for analysing eye-movement data. Many different approaches to scanpath comparison exist. One of the most frequently used approaches is String Edit Distance, where gaze trajectories are replaced by sequences of visited Areas of Interest. In cartographic literature, the most commonly used software for scanpath comparison is eyePatterns. During an analysis of eyePatterns functionality, we found that the tree-graph visualisation of its results is not reliable. Thus, we decided to develop a new tool called ScanGraph. Its computational algorithms are modified to work better with sequences of different length. The output is visualised as a simple graph, and similar groups of sequences are displayed as cliques of this graph. This article describes ScanGraph's functionality on a simple cartographic eye-tracking study example. Differences in the reading strategy of a simple map between cartographic experts and novices were investigated. The paper should aid researchers who would like to analyse the differences between groups of participants, and who would like to use our tool, available at www.eyetracking.upol.cz/scangraph.

Keywords: eye movement, eye tracking, scanpath, scanpath comparison, string edit distance, usability, cartography, application

Introduction

This article describes a new tool for scanpath comparison using visualisation of graph cliques. This tool will allow to find similarities between participants' process of presented stimuli observation. With information about their personal characteristics (age, sex, knowledge, etc.) it is possible to reveal if these groups are using a similar strategy. The output of the tool is a simple graph. In this graph cliques are identified. A clique is a subset of vertices in a graph where all vertices are connected by an edge with all of the others from that subset. These cliques represents participants with similar sequences of visited Areas of Interest (similar approaches to observing stimuli). The advantage over other scanpath comparison techniques is that the visualisation highlights only those participants that are similar according to the user-defined value of the degree of similarity.

Received April 4, 2016; Published August 5, 2016.
Citation: Dolezalova, J. & Popelka, S. (2016). ScanGraph: A novel scanpath comparison method using visualisation of graph cliques. Journal of Eye Movement Research, 9(4):5, 1-13.
Digital Object Identifier: 10.16910/jemr.9.4.5
ISSN: 1995-8692

In the introduction, the history and background of scanpath comparison is described with an emphasis on the most frequently used method – String Edit Distance. Also, the software eyePatterns is mentioned. During analysis of the eyePatterns outputs, it was found that results were not reliable. Its weaknesses are described in the first part of the methods section. Based on these findings, we decided to develop a new tool called ScanGraph, which calculates the similarities between scanpaths and visualises results in the form of graph cliques. The basic theory of simple graphs and cliques is also described in the methods section. The results section contains detailed information about ScanGraph. The functionality of the application is presented practically on an example of a model case study from the field of cartography. The results provide brief information about the model case study, and then an example of practical use of ScanGraph is presented. In the discussion section, the limitations of ScanGraph are described together with future proposals how to eliminate them.

During analysis of eye-tracking data using an average of eye-movement measures as fixation counts and durations, eye-movement behaviour unfolding a particular sequence over time is ignored. This sequence is a rich source of information (Anderson, Anderson, Kingstone, & Bischof, 2014). To analyse sequences of eye-movements, a large number of methods comparing scanpaths has been developed. These methods are collectively known as scanpath comparison.

The beginnings of interest in distinctive scanning patterns can be found in the study of Noton and Stark (1971), who reported a qualitative similarity in eye-movements when people viewed line drawings on multiple occasions. This observation was used to support the "Scanpath Theory", which proposed that visual features were encoded and stored alongside a motor memory of the scanpath made during perception. When a picture is seen again, it is recognised by executing the stored scanpath and matching the sequential features (Foulsham et al., 2012). The scanpath comprises sequences of alternating saccades and fixations that repeat themselves when a respondent is viewing stimuli. Scanpath comparison methods can be divided into six groups (String Edit Distance, ScanMatch, Sample-based Measures, Linear Distance, MultiMatch and Cross-recurrence Quantification Analysis). An overview of these methods and their comparison is described in Anderson et al. (2014).

One of the most frequently used methods is String Edit Distance, which is used to measure the dissimilarity of character strings. As Duchowski et al. (2010) mentions, scanpath comparison based on the String Edit Distance introduced by Privitera and Stark (2000) was one of the first methods to quantitatively compare not only the loci of fixations but also their order.

When using String Edit Distance, the grid or Areas of Interest (AOI) have to be marked in the stimulus. The gaze trajectory (scanpath) is then replaced by a character string representing the sequence of fixations with characters for AOIs they hit. Only 10 percent of the scanpath duration is taken up by the collective duration of saccadic eye-movements. Fixations in the created Areas of Interest took 90 percent of the total viewing period (Bahill & Stark, 1979). A sequence of transformations (insertions, deletions, and substitutions) is used to transform one string to another. Their similarity is represented as the number of transformation steps between two analysed strings (Anderson et al., 2014). Foulsham et al. (2012) pointed to the disadvantage of String Edit Distance, which is reducing distances to binary classification (because of the necessity of dividing stimuli on a grid or creating Areas of Interest). For some applications, as in cartography, this disadvantage can be turned into an advantage – for example, when analysing the behaviour of respondents to map composition elements.

One of the most used metrics calculating the distance between sequences is called Levenshtein distance, named after the Russian scientist Vladimir Levenshtein (Levenshtein, 1966). The Levenshtein distance between two strings $a = a_1 a_2 \ldots a_{|a|}$; $b = b_1 b_2 \ldots b_{|b|}$ of length $|a|$ a $|b|$ (let us denote $Lev(a, b)$) is the lowest number of deletions, insertions, or substitutions required to transform the source string into a target string. For example, the distance between sequences "gravitation" and "gravidity" is equal to 5 (three changes and two deletions). Hence, $Lev(a, b) \in \mathbb{N}_0$, $Lev(a, b) = 0$ if and only if the strings are equal and $Lev(a, b) = \max\{|a|, |b|\}$ if and only if there is any correspondence between the strings. The value of the Levenshtein distance increases with larger differences between the strings. The Levenshtein method was the first used for searchpath and scanpath analysis in the study of Choi, Mosley, and Stark (1995) and Stark and Choi (1996).

The other possible metric is called the Needleman-Wunsch algorithm with its scoring system. The Needleman-Wunsch algorithm (let us denote its value $NW(a, b)$) searches for concordant elements between two strings $a = a_1 a_2 \ldots a_{|a|}$; $b = b_1 b_2 \ldots b_{|b|}$ of the length $|a|$ a $|b|$. The basic scoring system used for our needs is given by Match reward $= 1$, Gap cost $= 0$ and Mismatch penalty $= -1$. For example, the distance between "gravitation" and "gravidity" is equal to 6 (six matches). Hence, $NW(a, b) \in \mathbb{N}_0$, $NW(a, b) = \min\{|a|, |b|\}$, when a is a subset of b or b is a subset of a. The value of $NW(a, b)$ increases with the similarity between the strings.

In our issue, we want to count the distances between each pair of sequences from a certain set. Both of these metrics properly work when all of the compared strings have the same length. But when the lengths of the sequences are not equal, a serious problem arises. Let us show an example with Levenshtein distance. Let $a = ABC, b = DEF, c = ABCDEFGHIJKLM, d = ABDGHJKNOP$. The distances are $Lev(a, b) = 3, Lev(c, d) = 7$, thus $Lev(c, d) >$

$Lev(a, b)$. But whereas the sequences c and d have similar parts, the sequences a and b are totally different.

String Edit Distance measurement was used for the evaluation of web page design (Josephson & Holmes, 2002). Areas of Interest were marked on the web page, and alphabetic code was assigned to each of them. Then, the eye-path sequence for each subject's viewing of each web page by recording the sequence of fixations was created. Sequences were compared with the use of the Optimal Matching Analysis tool (Chan, 1995). Fabrikant, Rebich-Hespanha, Andrienko, Andrienko, and Montello (2008) analysed eye-movement data recorded in controlled experiments on small-multiple map (a series of similar maps using the same scale, allowing them to be easily compared) displays with the use of ClustalG software (Wilson, Harvey, & Thompson, 1999). Clustal software packages are widely used for analysing gene sequences in DNA and proteins. ClustalG was developed specifically to analyse social-science data. Based on the results of visual geoanalytical approaches with sequence alignment analysis techniques, it was found that small-multiple displays cannot generally be computationally or informationally equivalent to non-interactive animations (animations which cannot be controlled – merely the playback of the video).

In 2006, West, Haake, Rozanski, and Karn (2006) introduced the software eyePatterns – software that uses well-established sequence analysis algorithms designed primarily to aid eye-movement researchers in comparing sequence patterns within and across experimental groups of subjects. Apart from String Edit Distance, eyePatterns also integrates transition frequency analysis, clustering, sequence alignment, and pattern discovery.

For research in the field of cognitive cartography, String Edit Distance is the most important feature. Based on eye-movement data, this method can answer questions such as "How is it possible that one person orientates themselves in a map very quickly, while it takes others a long time?", "Is there a difference in map reading between men and women?", or "Do all people look at maps the same way?" In some cartographic studies, sequence alignment methods were also used for non-eye-movement data. For example Joh, Arentze, Hofman, and Timmermans (2002) developed a new measure for similarity between activity patterns in activity-travel patterns data. Shoval and Isaacson (2007) used sequence alignment for analysing sequential aspects within the temporal and spatial dimensions of human activities.

String Edit Distance in eyePatterns was used, for example, in the study by Coltekin, Fabrikant, and Lacayo (2010), who analysed dynamic visual analytics displays. Levenshtein (Levenshtein, 1966) and Needleman-Wunsch (Needleman & Wunsch, 1970) algorithms implemented in eyePatterns were used to generate a distance matrix. Data were visualised in eyePatterns with a tree-graph constructed by a hierarchical clustering algorithm. In this tree-graph, clusters of participants were identified visually. Together with Path Similarity Analysis (Andrienko, Andrienko, Burch, & Weiskopf, 2012) eyePatterns was also used in the author's study by Popelka, Dvorsky, Brychtova & Hanzelka (2013). The aim of the study was to identify the typology of map readers (common behavioural characteristics identical or similar between more individuals) based on their eye-movements while solving geographical tasks with the use of a map.

Methods

eyePatterns and its disadvantages

West et al. (2006) states that eyePatterns uses hierarchical clustering for calculating sequence similarity. Clustering partitions data into subsets of items that share similar traits. Agglomerative hierarchical clustering builds a hierarchy of clusters, beginning with the two most closely related sequences, and ending with the most distant sequence or cluster. The hierarchy tree can be visualised, exposing outlying and the most similar sequences (West et al., 2006).

When we analysed outputs from eyePatterns, we found that sequences with the lowest value of Levenshtein distance have (correctly) the closest possible distance (two edges between them), because the algorithm starts with them. These two sequences now make a cluster and distances in the matrix are recalculated using this cluster instead of the original nodes (sequences). eyePatterns uses the average of distances between the pair of sequences making a new cluster. Due to this clustering, the distances between nodes in the tree-graph are distorted. The distance is now calculated for the average value for the whole cluster. The problem is that the distance between particular nodes inside the cluster towards the other node can be lower than the distance between this node and the average value for the whole cluster. From the tree-graph, it is not possible to distinguish in which cases the distances between original sequences were used and where the average for the cluster was used. Hence, tree-graph visualisation doesn't correspond to the statement used in the eyePatterns interface – "The fewer branches that are between two sequences, the more similar those sequences are", as is illustrated below in Figure 1.

Twenty scanpath strings (non-collapsed, marked as S1 – S20) from the stimulus used in model case study (see section "Model Case Study") were used to highlight the inaccuracy/errors of similarity calculation in eyePatterns. The tree-graph in Figure 1 displays the output of Levenshtein distance ("default scoring scheme") calculation from eyePatterns. Blue labels S1-S10 display participants GIS1-GIS10 (cartographers). Red labels S11-S20 stand for participants NOGIS1-NOGIS10 (non-cartographers).

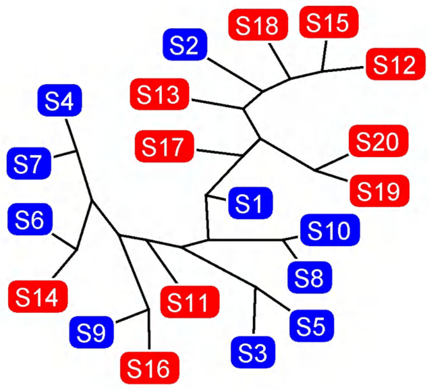


Figure 1. Output of eyePatterns – a tree-graph constructed by a hierarchical clustering algorithm

Figure 2 displays the tree-graph from Figure 1 with four highlighted sequences (participants). The closest highlighted pair in the tree-graph contains sequences S1 and S17. This should mean that the sequences are very similar. However, the Levenshtein distance between these two sequences was 13. Compare with the pair S12 and S14, lying on the opposite sides of the tree-graph. This should mean that the sequences differ a lot. The Levenshtein distance of these two sequences is only 4 – which means that only four changes are necessary to change one sequence to another. A similar situation is visible from the dendrogram (Figure 3) displaying the same data.

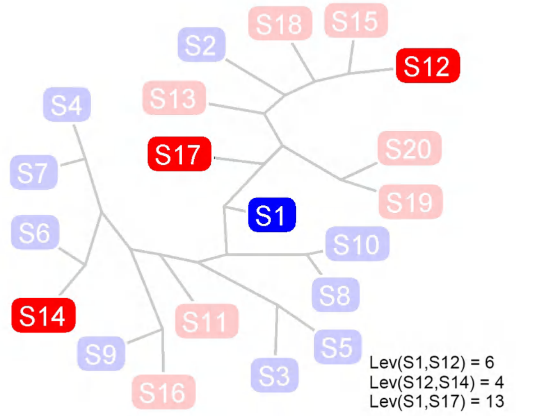


Figure 2. Tree-graph from eyePatterns with highlighted inaccuracies

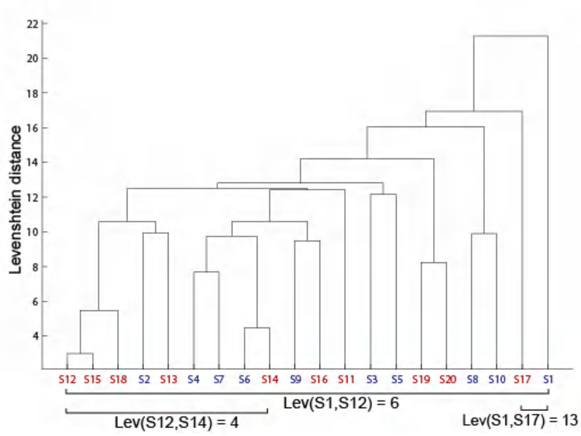


Figure 3. Tree-graph from eyePatterns redrawn to a dendrogram

Trajectories represented by sequences S1 and S17 were displayed in an OGAMA Scanpath module (Figure 4). It is evident that these two trajectories are very different as it is also obvious from their sequences (S1=EAAAAA AAAAAAAAAAAAACCC, S17=AAAAACCBBCCCC AAAA). Participant S1 (blue line) performed fewer fixations in the map. He also visited AOI E (Map title) and B (Map of Alaska), while no fixations from participant S17 (red line) were recorded there.

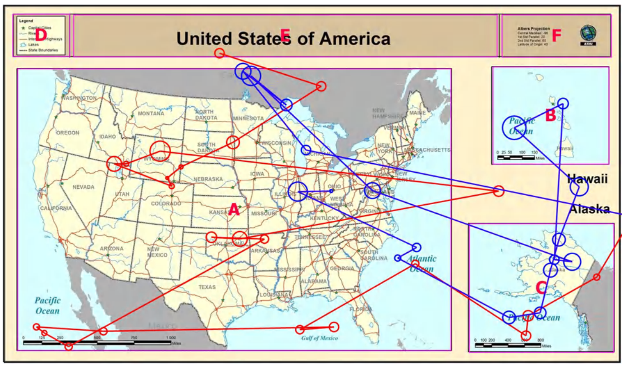


Figure 4. Comparison of trajectories S1 and S17, which were selected by eyePatterns as very similar

After discovering the inaccuracy of eyePatterns, we decided to develop our own application called ScanGraph for finding similar sequences in eye-tracking data. Compared with eyePatterns, where all sequences were included in the tree-graph, ScanGraph highlights only those sequences that are similar according to pre-set parameters. The result is displayed as a simple graph and similar groups are displayed as cliques of this graph.

Novel approach – ScanGraph

Our aim was to create a new tool that will work on the principle of binary relation. The task of finding groups of similar elements is equivalent to the task of seeking tolerance classes of a tolerance relation. This is also equivalent to the problem of finding cliques in a simple graph and can be easily and clearly visualised. The necessary terms are defined below.

Binary relations

A binary relation between two sets A and B is a subset of the Cartesian product $A \times B$. A binary relation on set A is a subset of $A \times A$.

When an element $a \in A$ is in a relation to an element $b \in B$ we write aRb.

Given a binary relation R on a set A we have the following definitions:

A relation R on a set A is called reflexive if and only if aRa for every element $a \in A$.

A binary relation R on a set A is called symmetric if and only if for any a and b in A, whenever aRb, then bRa.

A binary relation R on a set A is called transitive, if and only if for any a, b and c in A, whenever aRb, bRc, then aRc.

A binary relation R on A is set to be a relation of equivalence if it is reflexive, symmetric, and transitive.

A partition of a set A is by definition a union of subsets A_i that cover A but do not intersect each other: $A = \bigcup_{i=1}^{n} A_i, A_i \cap A_j = \emptyset, \forall i,j \in \{1, \ldots, n\}$. Given a relation of equivalence, we denote by $[[a]]$ the class of equivalence of an element a: $[[a]] = \{b \in A : aRb\}$. Two elements have the same class if and only if they are in relation: $[[a]] = [[b]] \Leftrightarrow aRb$. This is a direct consequence of transitivity and symmetry. Hence, given a relation of equivalence on a set A, its classes of equivalence form a partition of the set A.

A binary relation R on a set A is set to be a relation of tolerance if it is reflexive and symmetric.

The notion of tolerance relation is an explication of similarity or closeness.

As an analogy of equivalence classes and partitions, here we have tolerance classes and coverings. A set $B \subset A$ is called a tolerance preclass if it holds that for all $a, b \in B$, a and b are tolerant, i.e. aRb. A maximum preclass is called a tolerance class. So two tolerance classes can have common elements.

Given a non-empty set A, a collection $\prod_{i=1}^{n} A_i$ of non-empty subsets of A such that $\bigcup_{B \in \Pi} B$ is called a covering of A. Given a tolerance relation on a set A, the collection of its tolerance classes forms a covering of A.

Every partition is a covering; not every covering is a partition (Chajda, 2005).

Simple graphs

A graph $G = (V, E)$ is defined as a structure of two finite sets V and E. The elements of V are called vertices, and the elements of E are called edges. Each edge has a set of one or two vertices associated with it, which are called its endpoints.

An edge is said to join its endpoints. A vertex joined by an edge to a vertex v is said to be a neighbour of v.

A proper edge is an edge that joins two distinct vertices.

A self-loop is an edge that joins a single endpoint to itself.

A multi-edge is a collection of two or more edges having identical endpoints.

A simple graph has neither self-loops nor multi-edges (Gross & Yellen, 2005).

When we use the term graph without a modifier, we mean a simple graph.

Cliques

A subset S of $V(G)$ is called a clique if every pair of vertices in S is joined by at least one edge, and no proper superset of S has this property.

Thus, a clique of a graph G is a maximal subset of mutually adjacent vertices in G.

The clique number of a graph G is the number $\omega(G)$ of vertices in the largest clique in G.

A complete graph is a simple graph such that every pair of vertices is joined by an edge (Gross & Yellen, 2005).

Results

ScanGraph Application

The application was created using PhP and C# (Backend) and D3.js (Frontend). Its interface can be seen in Figure 5. When the web page www.eyetracking.upol.cz/scangraph/ is loaded, only the left column (1) is displayed. The user can select input data or try the functionality with a predefined source of data (1a). The application works with data exported from the application OGAMA (Voßkühler, Nordmeier, Kuchinke, & Jacobs, 2008). OGAMA (OpenGazeAndMouseAnalyzer) is an open-source application design to record and analyse eye and mouse movement data. It allows Levenshtein distances between sequences to be calculated, but the output is just a matrix with distance values. It is not possible to find groups of similar participants. Sequence similarity measures from OGAMA can be exported to a text file – and this text file can be imported directly to ScanGraph. This is one of the advantages over eyePatterns, which needs data in a specific format prepared in a text file or table processor as necessity.

Then, the user can specify the method of computation (1b) and select between collapsed or original strings (1c). In collapsed strings, there are no successive characters (AOIs) in the sequence. In the last step, the user can display an advised graph (1d) or construct a graph according to parameter p (1e) or percentage of edges (1f) (see below).

After clicking on the button “Advised graph” or “Compute graph”, elements $2-7$ are added to the display. In element 2, the points (vertices) representing all sequences of participants from the input dataset are shown. Different colours represent the affiliation to the category (e.g. male/female, expert/novice, etc.) The table on the left

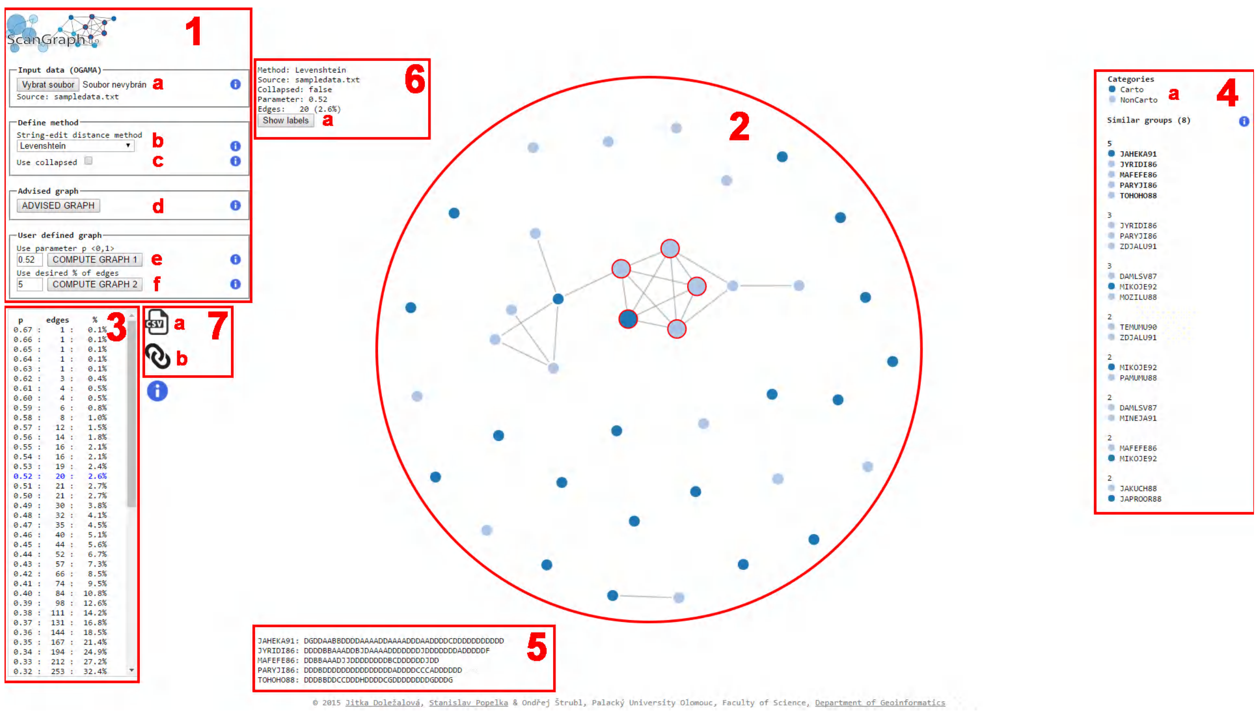


Figure 5. Interface of the ScanGraph application

(3) contains 100 pre-computed values of parameter p. In addition to parameter values, the number of edges is visible. The last column contains the percentage of edges from the complete graph. The user can click into the table to display the particular graph.

Cliques with two and more vertices found in these graphs are listed on the right side of the page (4). Colour points in front of the participant's name represents their group. An explanation of colours can be found in the upper part (4a). After clicking on any group of similar sequences in this section of the page, the clique is highlighted in the graph, and strings are also shown in the bottom part of the page (5). The user can visually inspect the sequence of the characters in the string. The overview of settings is shown in the upper part of the page (6), and the user can add labels with subject names to the graph (6a).

By clicking on the icon (7a), the user can download the table with all matrices (original matrix, modified matrix, and adjacency matrix), listed similarity groups with their character strings, input data, and overview of the settings. Clicking on (7b) allows a permanent link to the displayed graph to be created.

Computations and Visualisation

At first, the distance matrix (original matrix $D = (d_{ij})$) is constructed according to the Levenshtein or Needleman-Wunsch algorithm. Each element of the matrix d_{ij} is a distance between sequences i a j. The distances, however, are poorly comparable between themselves because of the different lengths of the sequences, as mentioned above. According to this, the value of a distance d_{ij} is divided by the higher value of the length of sequences i and j. It is the highest distance the two sequences could have (in the case of Levenshtein) or the value of the greatest similarity the two sequences could have (in the case of Needleman-Wunsch) (modified matrix $M = (m_{ij})$). Obviously, now the new elements could have a value from the interval $\langle 0,1 \rangle$ and still apply the higher the value is (in the case of Levenshtein), or the lower the value is (in the case of Needleman-Wunsch) the more different the strings are. The next step is up to the user.

The first option is to select the Advised graph button. This button returns a graph with 5% of the possible edges and a corresponding value of parameter p (see below). This graph is user-friendly and according to our experiences has a very high interpretive value about any similarities. This option is recommended for users with no experience with ScanGraph. The second option is to construct a user-defined graph. The graph is created according to parameter p or percentage of edges.

The parameter p takes its value from the interval $\langle 0,1 \rangle$ and represents the degree of similarity. The higher the value of p, the higher the similarity of the given sequences. Obviously, $p_{ij} = 1 - m_{ij}$ applies in the case of Levenshtein distance and $p_{ij} = m_{ij}$ in the case of the Needleman-Wunsch algorithm $\forall i, j \in \{1, \dots, n\}$ of the modified matrix, where n is the number of participants. The value of p constructs a new matrix (adjacency matrix $A = (a_{ij})$) according to these conditions:

$$a_{ij} = \begin{cases} 1, & \text{if } p \geq p_{ij}, \\ 0, & \text{otherwise.} \end{cases}$$

Hence, the adjacency matrix represents a simple graph, which is displayed. The second option is to set a percentage of edges. This number takes a value from the interval $< 0,100 >$. The algorithm finds a value of the parameter p for which the graph will have a given percentage of edges (eventually rounded to the nearest lower value) and displays the graph and parameter p.

Besides the graph itself, a table with three columns is displayed. Parameter p with its 100 possible values and number of edges, and the percentage of the corresponding graph.

Each graph is represented by its adjacency matrix. Using the matrix, all cliques contained in the graph can be found. Each clique represents a group of sequences which has the same or higher degree of similarity than the given parameter p.

The maximal clique problem (finding all maximal cliques in a graph) is an NP-complete decision problem.

Definitions of the decision problem according to (Gross & Yellen, 2005) follows.

A decision problem is a problem that requires only a yes or no answer regarding whether some element of its domain has a particular property.

A decision problem belongs to the class P if there is a polynomial-time algorithm to solve the problem.

A decision problem belongs to the class NP if there is a way to provide evidence of the correctness of a yes answer so that it can be confirmed by a polynomial-time algorithm.

A decision problem R is polynomially reducible to Q if there is a polynomial-time transformation of each instance I_R of problem R to an instance I_Q of problem Q such that instances I_R and I_Q have the same answer.

A decision problem is NP hard if every problem in class NP is polynomially reducible to it.

An NP-hard problem R is NP-complete if R is in class NP.

The algorithm default used by ScanGraph is based on an exhaustive algorithm and finds an optimal solution with all cliques with two or more vertices in the given graph. When the computational time is too large, ScanGraph uses a heuristic algorithm.

Figure 6 displays an example of the influence of the parameter p value. On each image, the largest clique is marked. When the parameter is set to 0, the graph is always a complete graph. As the value of a parameter is increased, vertices very different from each other are dropped from the largest clique. When the value of a parameter is set to 0.5, the first vertex is out of the graph – it is not similar to any other in the dataset. The largest value of the parameter in the figure is 0.9 and only two vertices are making a clique. The character strings of these vertices were exactly the same, so in this case, it was needless to include an image with the graph with parameter 1, because it will be exactly the same as the previous image.

Model Case Study

The functionality of the developed ScanGraph tool was presented in an example of a case study comparing different map compositions. The data were recorded as a part of work by students. The aim of the study was to reveal whether cartographers and non-cartographers perceive maps differently.

The experiment contained a total of 18 stimuli. Six types of maps were created and each of them was presented with three different map compositions. The distribution of map elements (map, legend, title, imprint, addition map) were placed at various positions in the stimuli.

A total number of 20 respondents participated in this eye-tracking study. Half of them were selected from a group of undergraduate students who had already attended a cartography course. The rest of them were selected from among students in fields of study other than cartography. The differences between cartographers and non-cartographers were investigated. For the case study, an eye-tracking device SMI RED 250 was used, and data were recorded with a frequency of 60Hz. Eye positions were recorded every 16ms. Eyes move in a number of different ways, simultaneously responding to commands from a number of different brain areas. One of the most important types of eye movement is no movement at all, but rather the ability to keep the eye trained on a fixed spot in the world. This is known as fixation. To get from one fixation to the next, the eyes make rapid, ballistic movements known as saccades (Hammoud & Mulligan, 2008). Plenty of algorithms for fixation detection exist, but the

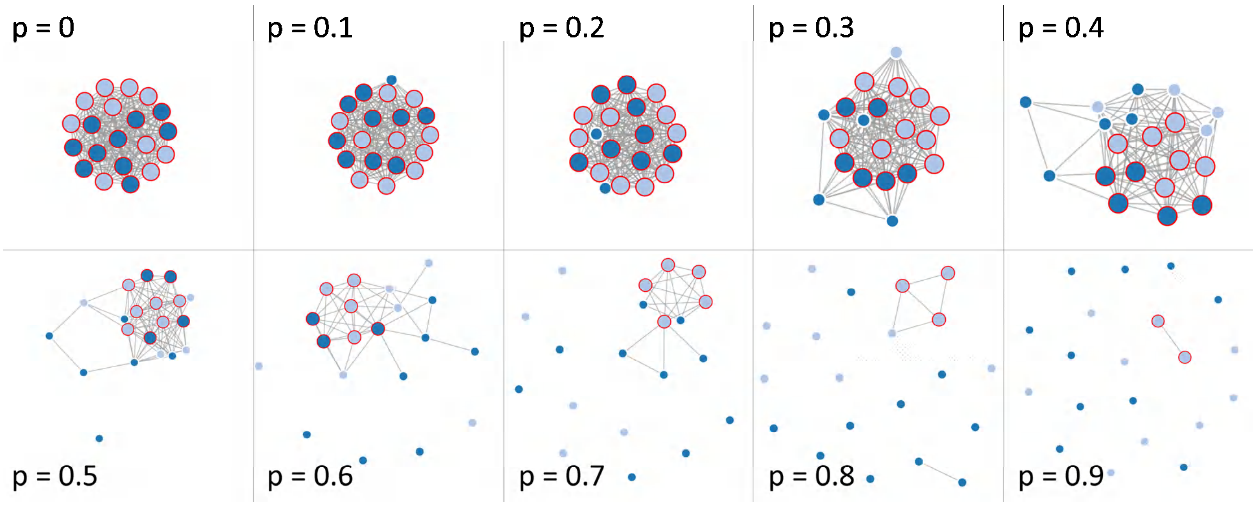


Figure 6. Example of the influence of parameter p value

most used for low-speed data (up to 250 Hz) is I-DT. I-DT takes into account the close spatial proximity of eye position points during eye movement trace (Salvucci & Goldberg, 2000).

For the case study, the software OGAMA was used with an ID-T algorithm for fixation detection. In OGAMA, the most important parameters for fixation detection are "Maximum distance" and "Minimum number of samples". Thresholds in OGAMA were set to 15px (distance) and 10 samples. More information about this setting is described in (Popelka & Doležalová, 2015).

Results of the Model Case Study

Recorded eye-movement data were visualised using the Sequence Chart method available in SMI BeGaze. The Sequence Chart shows the temporal sequence of the visited Areas of Interest. Figure 7 shows a Sequence Chart for all respondents for three different map compositions. Areas of Interest were marked around all map composition elements. The colour of the stripes under the maps represents a distribution of respondent attention between these AOIs. From a visual analysis of the Sequence Chart, a difference between the group of cartographers and non-cartographers can be seen. The most prominent difference is in the map element title (blue).

A fixation cross preceded each stimulus, so AOI representing the map field is always viewed in the first 500ms. Beyond this time, most cartography students automatically read the title of the map, or rather, noted fixations representing it in AOI. Non-cartographers did not do so. It is evident especially in the first column, where the stimulus was an "ideal" map composition. In the following columns, the composition did not obey cartographic rules. Despite this fact, students of cartography were trying to find the title of the map.

The Sequence Chart is illustrative and easy to interpret, but a deeper analysis of differences between the strategies of participant groups needs a more sophisticated method of analysis. In this paper, data from this short study were used for demonstrating the use and possibilities of the developed ScanGraph web application. More specifically, eye-movement data recorded during observation of map composition #1 (first column of Figure 7) were used.

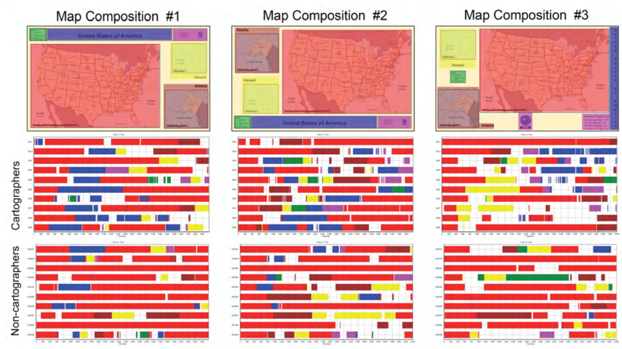


Figure 7. Sequence chart visualisation of participant reading strategy of three different map compositions. Map composition #1 was used for a model case study.

Demonstration of using the ScanGraph Application

The ScanGraph application was designed to work with data exported from the open-source application OGAMA (Voßkühler et al., 2008) – An open source software designed to analyse eye and mouse movements in slideshow study designs. OGAMA contains a tool called "Levenshtein Distance Calculation", which is capable of computing Levenshtein distances between the trajectories of participants. Sequence similarities can be calculated based on the regular grid or user defined Areas of Interest.

The output of this tool is a matrix of similarities between sequences and also the list of scanpath strings for each participant. When using ScanGraph, only the strings are important. The values of similarity calculated in OGAMA are not used, because the Levenshtein algorithm was modified to take into account different lengths of strings.

The first step of data analysis with ScanGraph is the creation of Areas of Interest above analysed stimuli. In our case, map composition #1 from Figure 7 was used, and Areas of Interest were marked around map composition elements (see Figure 8).

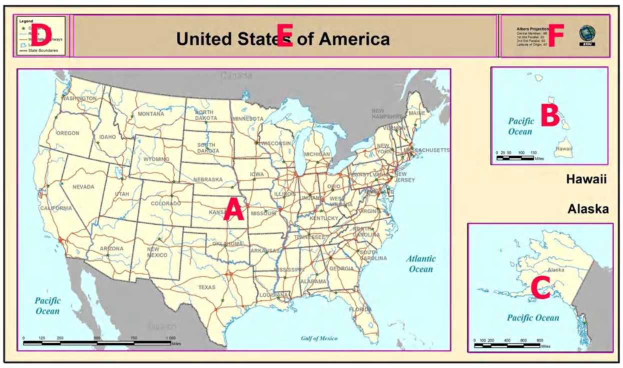


Figure 8. Areas of Interest marked in the stimulus with Map composition #1

In the next step, the Scanpath module in OGAMA was used to display the trajectories of participants and scanpath strings. The text file with sequence similarities can be exported from OGAMA and directly used as input data in ScanGraph.

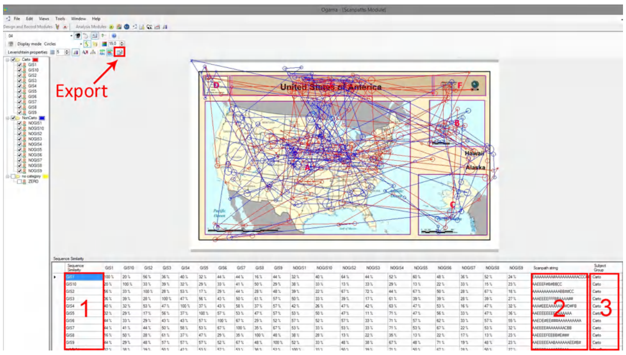


Figure 9. An environment of OGAMA's Scanpath module. Levenshtein distances between selected participants (and set of AOIs) are calculated and can be exported as a text file. Only Subject names (1), Scanpath strings (2), and affiliation to subject groups (3) is used in ScanGraph.

The exported text file was then opened with ScanGraph. The user can choose between the Levenshtein and Needleman-Wunsch algorithm and construct a graph.

In this particular example, the order of visited areas as well as their number of fixations was investigated. Therefore, the original (non-collapsed) data were analysed with the Levenshtein algorithm. The user can modify the value of parameter p or percentage of edges. In our case, we started with the parameter value 0.8, and two cliques were found (Figure 10).

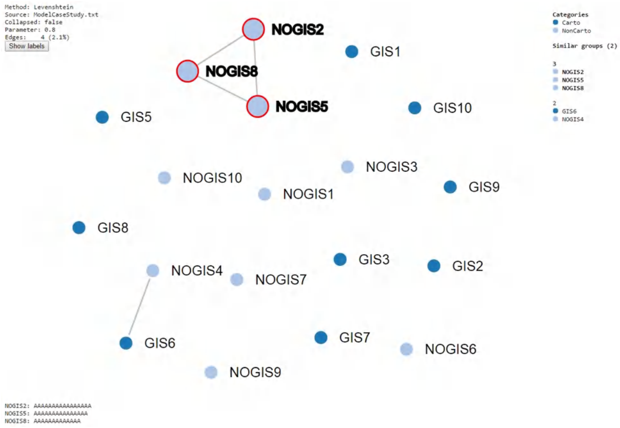


Figure 10. Output of ScanGraph showing two cliques based on parameter $p = 0.8$.

Figure 11 shows the trajectories of participants making similar groups with parameter $p = 0.8$ (Figure 10). The larger one contained three participants from the group of non-cartographers. All of them spent their whole observation time in the AOI A (containing the map field). The other clique comprised two participant sequences and is displayed in shades of blue. Both participants performed the same number of fixations (17), mainly in the map field (AOI A) and map title (AOI E). Participant GIS6 (dark blue) made an extra fixation in the map legend.

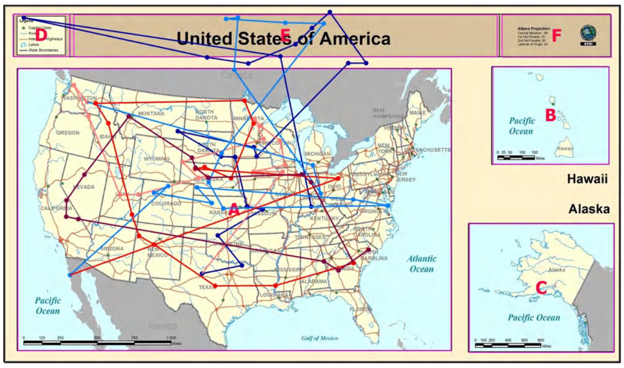


Figure 11. Trajectories of five participants making two cliques (based on parameter $p = 0.8$). Three participants displayed in shades of red spent the whole time in the map field. Participants GIS6 and NOGIS4 (shades of blue) visited the map field and map title.

When the value of the parameter was decreased to 0.75, a total of six cliques was found in the data (Figure 12). The group of three similar participants was (obviously) preserved, but it was extended by participant NOGIS4 (in the case of the first clique), respectively GIS2 (in the case of the second clique). This means that participants NOGIS4 and GIS2 are both similar to the group of three participants from Figure 10 (NOGIS2, NOGIS5, and NOGIS8), but are not similar to each other.

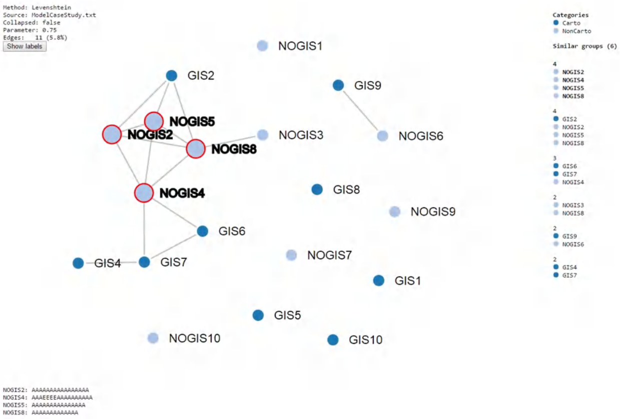


Figure 12. ScanGraph output showing two cliques based on parameter $p = 0.75$.

The trajectories from Figure 12 are displayed in Figure 13. Trajectories of participants NOGIS2, NOGIS5, and NOGIS8 shown in shades of red are again displayed in shades of red. The blue trajectory represents participant NOGIS4, and the green belongs to participant GIS2. Both trajectories are similar to the red ones, but they are not similar to each other.

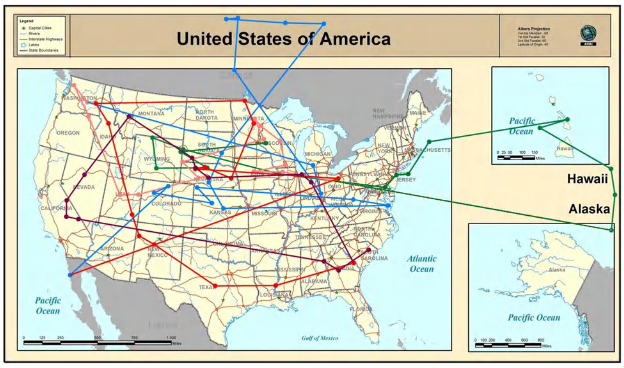


Figure 13. Trajectories of five participants making two cliques (based on parameter $p = 0.75$*). Participants NOGIS2, NOGIS5, and NOGIS8 are displayed in shades of red. Participants NOGIS4 (blue line) and GIS2 (green line) are both similar to the red ones, but not to each other.*

Discussion

As already mentioned in the introduction, plenty of scanpath comparison methods exist (Anderson et al., 2014). For cartographic research (and not only), there is an advantage in using String Edit Distance based on Areas of Interest marked around map composition elements.

Until now, scanpath comparison in cartography was performed by using the eyePatterns application (West et al., 2006), which offers a variety of functions, but in cartography, only evaluating similarity between sequences was previously used. We have found that the implementation of Levenshtein and Needleman-Wunsch algorithms in eyePatterns is correct, but it is not appropriate for comparing strings with different length. Visualisation of results via tree-graphs is inaccurate and misleading, so we decided to develop our own tool – ScanGraph. We believe that our tool offers more useful results than eyePatterns in this specific functionality, but the variety of functionality (i.e. search for patterns) in eyePatterns can still be used for some applications.

The case study presented in this article demonstrated the use of the ScanGraph application. The goal of the case study was not to find anything important from the trajectories of participants but to present ScanGraph functionality. For that reason, only one stimulus observed by 20 participants was investigated.

During the development of ScanGraph, we ran into several problems, but most of them were solved or bypassed. As was mentioned above, the exhaustive algorithm is NP-complete. Due to this, with an increasing number of edges, the computational time increases non-polynomially. In that case, a heuristic algorithm is used. The heuristic algorithm might not find an optimal solution, i.e. all maximal cliques (Vecerka, 2007). However, the graphs where a solution could be found by an exhaustive algorithm have a higher interpretative value for the experiment.

ScanGraph is still under development. The next step will be to add a matrix of differences between AOIs. For some experiments, it may be important to define the cost of transitions between each pair of AOI separately. For example, transition from AOI A to AOI B (e.g. map vs. legend) could mean a more important change than transition from AOI C to AOI D (e.g. two columns of the legend), so the Levenshtein distance should be different. For this case, the user will define his own matrix of differences between AOIs.

Another possible improvement could be the computation of similarities between participants for a whole experiment. The user will upload a compressed file containing character strings from all stimuli of the experiment. The global similarity between all participants throughout all stimuli will be again displayed as a graph.

Conclusion

This article describes the possibilities of a newly developed application for scanpath comparison called ScanGraph. The application performs scanpath comparison based on the String Edit Distance method, and its output is a graph. Groups of similar sequences/participants are displayed as cliques of this graph.

ScanGraph can be used for all studies where differences between gaze movements in different groups of participants are investigated. ScanGraph works with an exported file from the open-source application OGAMA. To use ScanGraph, it is necessary to create Areas of Interest around specific parts of analysed stimuli. OGAMA allows a text file containing character strings representing the order of visited AOIs to be exported. This file can be directly imported to ScanGraph, which is freely available at www.eyetracking.upol.cz/scangraph. In the ScanGraph web environment, the user can display groups of participants with similar character strings – participants with a similar strategy. The user can calculate an advised graph (containing 5% of edges) or a user defined graph based on parameter (percentage of similarity) or percentage of edges. Groups of similar participants are marked in this graph and the user can quickly inspect their character strings.

Until now, the eyePatterns application was commonly used for this purpose. We have found that the eyePatterns output, a tree-graph showing similarity between all sequences, does not reflect the similarity measured by the algorithms used. From the tree-graph, similar groups can only be found visually, which is very inaccurate. Our approach does not connect all the sequences (participants), but created groups of similar participants correspond to the computations. The user knows that an identified group is similar according to a given parameter – which is not possible in eyePatterns. The algorithms for calculating similarity were modified and work better with strings of different length.

The functionality of ScanGraph was presented in an example of a simple cartographic case study in detail. This paper can serve as a user manual for ScanGraph.

Acknowledgements

This paper is supported by the projects of Operational Program Education for Competitiveness – European Social Fund (projects CZ.1.07/2.3.00/20.0170), of the Ministry of Education, Youth, and Sports of the Czech Republic and the student project IGA_PrF_2016_008 of Palacky University.

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2014). A comparison of scanpath comparison methods. *Behavior research methods*, 1-16.

Andrienko, G., Andrienko, N., Burch, M., & Weiskopf, D. (2012). Visual Analytics Methodology for Eye Movement Studies. *Visualization and Computer Graphics, IEEE Transactions on, 18*(12), 2889-2898.

Bahill, A. T., & Stark, L. (1979). The trajectories of saccadic eye movements. *Scientific American, 240*(1), 108-117.

Coltekin, A., Fabrikant, S., & Lacayo, M. (2010). Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. *International Journal of Geographical Information Science, 24*(10), 1559-1575.

Duchowski, A. T., Driver, J., Jolaoso, S., Tan, W., Ramey, B. N., & Robbins, A. (2010). *Scanpath comparison revisited.* Paper presented at the Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications.

Fabrikant, S. I., Rebich-Hespanha, S., Andrienko, N., Andrienko, G., & Montello, D. R. (2008). Novel method to measure inference affordance in static small-multiple map displays representing dynamic processes. *Cartographic Journal, The, 45*(3), 201-215.

Foulsham, T., Dewhurst, R., Nyström, M., Jarodzka, H., Johansson, R., Underwood, G., & Holmqvist, K. (2012). Comparing scanpaths during scene encoding and recognition: A multi-dimensional approach.

Gross, J. L., & Yellen, J. (2005). *Graph theory and its applications*: CRC press.

Hammoud, R. I., & Mulligan, J. B. (2008). Introduction to Eye Monitoring *Passive Eye Monitoring* (pp. 1-19): Springer.

Chajda, I. (2005). *Úvod do algebry* (Vol. 1). Olomouc: Vydavatelství Univerzity Palackého.

Chan, T. W. (1995). Optimal matching analysis: a methodological note on studying career mobility. *Work and occupations, 22*(4), 467-490.

Choi, Y. S., Mosley, A. D., & Stark, L. W. (1995). String editing analysis of human visual-search. *Optometry and Vision Science, 72*(7), 439-451. doi:10.1097/00006324-199507000-00003

Joh, C.-H., Arentze, T., Hofman, F., & Timmermans, H. (2002). Activity pattern similarity: a multidimensional sequence alignment method. *Transportation Research Part B: Methodological, 36*(5), 385-403.

Josephson, S., & Holmes, M. E. (2002). *Visual attention to repeated internet images: testing the scanpath theory on the world wide web.* Paper presented at the Proceedings of the 2002 symposium on Eye tracking research & applications.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics doklady, 10*(8), 707-710.

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology, 48*(3), 443-453.

Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research, 11*(9), 929-IN928. doi:http://dx.doi.org/10.1016/0042-6989(71)90213-6

Popelka, S., & Doležalová, J. (2015). Non-photorealistic 3D Visualization in City Maps: An Eye-Tracking Study *Modern Trends in Cartography* (pp. 357-367): Springer.

Popelka, S., Dvorsky, J., Brychtova, A., & Hanzelka, J. (2013). User typology based on eye-movement paths *Geoconference on Informatics, Geoinformatics and Remote Sensing - Conference Proceedings, Vol I* (pp. 1041-1048).

Privitera, C. M., & Stark, L. W. (2000). Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22*(9), 970-982.

Salvucci, D. D., & Goldberg, J. H. (2000). *Identifying fixations and saccades in eye-tracking protocols*. Paper presented at the Proceedings of the 2000 symposium on Eye tracking research & applications.

Shoval, N., & Isaacson, M. (2007). Sequence alignment as a method for human activity analysis in space and time. *Annals of the Association of american Geographers, 97*(2), 282-297.

Stark, L. W., & Choi, Y. S. (1996). Experimental metaphysics: The scanpath as an epistemological mechanism. *Advances in psychology, 116*, 3-69.

Vecerka, A. (2007). Grafy a grafové algoritmy. *Univerzita Palackého Prırodovedecká Fakulta Katedra Informatiky, Olomouc*.

Voßkühler, A., Nordmeier, V., Kuchinke, L., & Jacobs, A. M. (2008). OGAMA (Open Gaze and Mouse Analyzer): open-source software designed to analyze eye and mouse movements in slideshow study designs. *Behavior research methods, 40*(4), 1150-1162.

West, J. M., Haake, A. R., Rozanski, E. P., & Karn, K. S. (2006). *eyePatterns: software for identifying patterns and similarities across fixation sequences.* Paper presented at the Proceedings of the 2006 symposium on Eye tracking research & applications.

Wilson, C., Harvey, A., & Thompson, J. (1999). *ClustalG: Software for analysis of activities and sequential events.* Paper presented at the IATUR Conference Proceedings.

Paper EyeTribe

Popelka, S.*, Stachon, Z., Sasinka, C., & Dolezalova, J. (2016). EyeTribe Tracker Data Accuracy Evaluation and Its Interconnection with Hypothesis Software for Cartographic Purposes. Computational Intelligence and Neuroscience, Article 9172506.
https://doi.org/10.1155/2016/9172506

[IF 1.215; Q3]

[48 citations on WoS; 43 citations without autocitations]

[Author's contribution: 25%]

Abstract

The mixed research design is a progressive methodological discourse that combines the advantages of quantitative and qualitative methods. Its possibilities of application are, however, dependent on the efficiency with which the particular research techniques are used and combined. The aim of the paper is to introduce the possible combination of Hypothesis with EyeTribe tracker. The Hypothesis is intended for quantitative data acquisition and the EyeTribe is intended for qualitative (eye-tracking) data recording. In the first part of the paper, Hypothesis software is described. The Hypothesis platform provides an environment for web-based computerized experiment design and mass data collection. Then, evaluation of the accuracy of data recorded by EyeTribe tracker was performed with the use of concurrent recording together with the SMI RED 250 eye-tracker. Both qualitative and quantitative results showed that data accuracy is sufficient for cartographic research. In the third part of the paper, a system for connecting EyeTribe tracker and Hypothesis software is presented. The interconnection was performed with the help of developed web application HypOgama. The created system uses open-source software OGAMA for recording the eye-movements of participants together with quantitative data from Hypothesis. The final part of the paper describes the integrated research system combining Hypothesis and EyeTribe.

Hindawi Publishing Corporation
Computational Intelligence and Neuroscience
Volume 2016, Article ID 9172506, 14 pages
http://dx.doi.org/10.1155/2016/9172506

Research Article

EyeTribe Tracker Data Accuracy Evaluation and Its Interconnection with Hypothesis Software for Cartographic Purposes

Stanislav Popelka,[1] Zdeněk Stachoň,[2] Čeněk Šašinka,[2] and Jitka Doležalová[1]

[1] *Palacký University, Olomouc, 77146 Olomouc, Czech Republic*
[2] *Masaryk University, 61137 Brno, Czech Republic*

Correspondence should be addressed to Stanislav Popelka; standa.popelka@gmail.com

Received 25 November 2015; Revised 8 February 2016; Accepted 22 February 2016

Academic Editor: Ying Wei

The mixed research design is a progressive methodological discourse that combines the advantages of quantitative and qualitative methods. Its possibilities of application are, however, dependent on the efficiency with which the particular research techniques are used and combined. The aim of the paper is to introduce the possible combination of Hypothesis with EyeTribe tracker. The Hypothesis is intended for quantitative data acquisition and the EyeTribe is intended for qualitative (eye-tracking) data recording. In the first part of the paper, Hypothesis software is described. The Hypothesis platform provides an environment for web-based computerized experiment design and mass data collection. Then, evaluation of the accuracy of data recorded by EyeTribe tracker was performed with the use of concurrent recording together with the SMI RED 250 eye-tracker. Both qualitative and quantitative results showed that data accuracy is sufficient for cartographic research. In the third part of the paper, a system for connecting EyeTribe tracker and Hypothesis software is presented. The interconnection was performed with the help of developed web application HypOgama. The created system uses open-source software OGAMA for recording the eye-movements of participants together with quantitative data from Hypothesis. The final part of the paper describes the integrated research system combining Hypothesis and EyeTribe.

1. Introduction

The paper presents methodological-technical approach combining quantitative and qualitative methods which are based on specific technical tools. The aim of this paper is to introduce the newly developed technical research system and results of its validation: specifically, the creation and empirical verification of an interconnection of a web-based platform Hypothesis with an EyeTribe eye-tracking system connected to open-source software OGAMA. The interconnection was done by the creation of a new web application HypOgama.

The introduction of the paper discusses the methodology and mixed-research design (combination of quantitative and qualitative, resp., explorative methods) in the area of cognitive visualization and cartography. The paper consists of three parts which are ordered due the logic and procedure of the research system creation and verification. The first part is focused on the presentation of a tool for mass data collection: web-based platform Hypothesis. The second part of the paper presents the new low-cost eye-tracking system EyeTribe, which allows efficient realization of qualitative, respectively, explorative studies. In this part, close attention is paid to empirical study verifying the truthfulness of the low-cost EyeTribe tracker in comparison with SMI RED 250 system. The final part of the paper describes the research system which combines and integrates above-mentioned tools. Part of this last section is also an illustration of possible empirical study, where the interconnection of Hypothesis and EyeTribe for cartographic and psychology research is presented. However this case study is only an example of how the integrated research system and HypOgama application works, and it should only illustrate the procedure of conducting a mixed-research design.

A significant portion of experimental studies in the area of cognitive visualization can be sorted into two main categories. The studies in the first category monitor and record the behaviour of individuals or, rather, their conscious actions and general work methods when completing tasks with a use of a map. The most common aspects of studies are completion speed, accuracy, and correctness or frequency of a given solution (see [1–5]). The mentioned studies use a quantitative approach and subsequent statistical methods of data analysis. A second significant category is the use of eye-tracking systems. Eye-tracking studies are in many cases combined with the recording of conscious behaviour, that is, user actions (see the first category), but the crucial activities recorded are eye-movements, which offer continuous data about (even unconscious) behaviour of the participant while solving a task. In other words, the focus of the user's attention is foregrounded [6]. Due to the high processing requirements, these studies are often performed on a small sample of participants and methods other than statistical data analysis are being used, for example, explorative data analysis [7].

Eye-tracking was used for the evaluation of maps for the first time already in the late 1950s [8], but it has been increasingly used in the last ten to fifteen years. The main reasons are the declining prices of the equipment and the development of computer technology that allows faster and more efficient analysis of measured data. For usability research, eye-tracking data should be combined with additional qualitative data, since eye-movements cannot always be clearly interpreted without the participant providing context to the data [9].

An example of comprehensive research in the field of cognitive visualization by using eye-tracking is the work of Alaçam and Dalcı [10], who compared four map portals (Google Maps, Yahoo Maps, Live Search Maps, and MapQuest). The basic assumption of the study was that lower average fixation duration indicates more intuitive map portal environment. The shortest average fixation duration was found in the case of Google Maps. Fabrikant et al. [11] used eye-tracking for the evaluation of map series expressing the evolution of the phenomenon over time, or for evaluation of user cognition of weather maps [12]. Ooms et al. [13] dealt with the suitability of map label positions and differences in map reading between experts and novices. Popelka and Brychtova [14] investigated the role of 2D and 3D terrain visualization in maps.

Olson [15] compared cognitive visualization and cognitive psychology, arguing that cartographers can adapt ideas and experiments in methodology from cognitive psychologists. Equally, psychologists can use maps as stimuli in their studies. Both disciplines can examine the cognitive processes while reading and understanding maps. However, cognitive psychologists are interested in different types of cognitive processes such as attention, visual perception, memorizing, or decision-making. A map is only a tool in this context. For a cognitive cartographer, the map is far more important.

The approach mentioned above is based on close cooperation between cartographers and psychologists and shows

Large-scale study
(i) Large research sample
(ii) Extensive and mass data collection on the Hypothesis platform
(iii) Event logging: user actions (conscious behaviour)

Small-scale study
(i) Limited research sample
(ii) Combination of Hypothesis and EyeTribe systems
(iii) Logging user actions and gaze tracking

Figure 1: The combination of large-scale and small-scale study.

the possibility of a connection between large-scale studies and small-scale studies based on gathering and analysing eye-tracking data. Differences between large-scale and small-scale studies are described in Figure 1.

As it is discussed in Štěrba et al. [16], using only a qualitative (explorative) or quantitative type of evaluation method is not sufficient. Therefore, it is necessary to combine those methods, enabling their suitable completion, obtaining more valid results, and achieving better interpretation. A combination of quantitative and qualitative methods was established as mixed-research design [17]. The key idea and innovation of our method are the interconnection of two approaches in the area of cognitive visualization and also finding a technological solution.

The Hypothesis platform serves primarily for the creation of experimental test batteries, online administration, and extensive data gathering. After connecting with the eye-tracking system, more detailed data on the experimental task processing methods are gathered, which allow deeper insight into the postulated cognitive processes that underlie the behavioural reactions.

Štěrba et al. [18] propose two variants of mixed-research design:

(i) Using the eye-tracking system for a pilot study examining a quality of experiment design with results from this pilot study being used for improvement of experiment design before large-scale data collection.

(ii) Using Hypothesis for large-scale quantitative approach and secondary using of eye-tracking method for the subsequent specification of certain results with adjusted or changed types of tasks.

Both approaches and technical specification of Hypothesis platform are described in detail in [18] and are available online in English.

2. A Tool for Mass Data Collection: Web-Based Platform Hypothesis

For the purposes of large-scale experimental investigation, the creation of psychological tests, and evaluation of cartographic works, new research software concept was designed within the project "Dynamic Geovisualization in Crisis Management" [19]. Subsequently, this concept has been realized, and original software MuTeP was developed [20, 21]. MuTeP

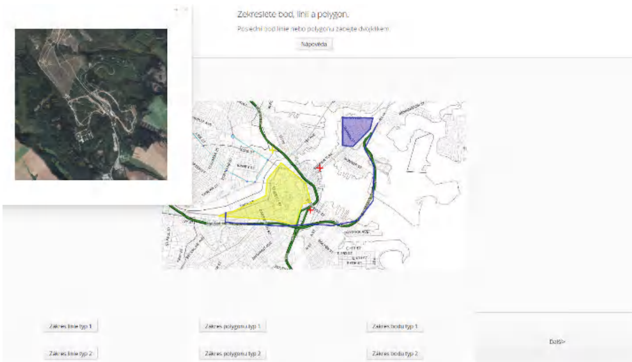


Figure 2: Example task on WMS interactive map. The user indicates the requested objects, draws lines, and marks out target areas by polygons. In the example shown, the user called up an orthophoto map in a dialogue-window. All the actions including the drawn point coordinates, lines, and polygons are saved in the database, and the correctness of the solution is automatically evaluated under preset conditions.

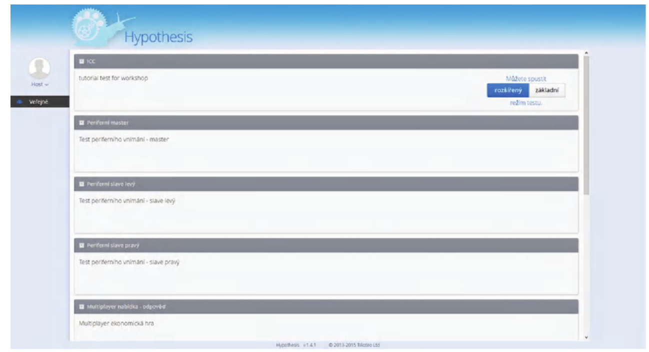


Figure 3: Management module in the Hypothesis platform. The user can launch the available tests in two modes: (a) legacy (launches in a normal browser) and (2) featured (launches in a controlled mode in SWT browser). The manager and the superuser have an extended access and can unlock the tests, create users, export results, and so forth.

was primarily created for the purposes of objective experimental exploration and evaluation of cartographic products in the perspective of user personality.

Although MuTeP was practically proven [22], it was clear that the conception used will soon reach its limits. Another impulse for the search for a more flexible solution was an effort to involve dynamic cartographic visualization as stimuli, randomization, nonlinear test batteries, connection with eye-tracking technology, and so forth, which were not possible to implement into MuTeP software.

Based on experience with MuTeP and in the context of current requirements, a new software concept was designed. This new software should have the potential for long-term growth and development [23]. Hypothesis has several important advantages in comparison with MuTeP. Above all, Hypothesis enables computer adaptive testing and offers a modular solution with plugin support (such as video or interactive animation plugins) and enables the work with interactive maps (such as web map services; see Figure 2).

The technology used for designing Hypothesis consists of the following: (1) the application core and user interface are built on framework Vaadin 7; work with the database is provided by ORM Hibernate; and (2) PostgreSQL in version 9.1 (and higher) is used as a primary database system [18].

The architecture of the system is three-layer: a client, server, and database. The client part is designed for communication and interaction with the user, and its operation is provided by standard web browsers (thin client) or a special browser distributed in the application package—special Hypothesis Browser. Hypothesis Browser is based on Standard Widget Toolkit (SWT) components and ensures more strict conditions and control over running tests [18, 24].

Hypothesis works as an event-logger application, which logs all user actions and events (coordinates and timestamps of clicks, key presses, start and end time of each presented slide, exposition time of every component such as a picture or dialogue-window, zoom of maps, rotation of 3D objects, etc.). Extensive logging of user actions and events is enabled through the structure of the final slides used for the test battery (package). The package comprises the hierarchical structure of branches which contain one or more tasks, and each task contains at least one slide. The slide consists of a template and content. Such structure enables nonlinear branching of the test slides or randomization of slides. All parts of the package are stored in structured XML format. After starting a test, a selected package is loaded from the database to the server application and a new test is created. Emphasis was placed on variability and range of software usability. Figure 2 shows an example of the slide using WMS. The slide consists of two layers. The underlaying image is created with a layer: ImageLayer. Above it, there is a transparent layer: FeatureLayer, which is designed to draw demanded points, polylines, or areas by mouse and store the events [18].

Hypothesis is also improved with two new key functionalities that are vital for the interconnection between eye-tracking systems (or other peripherals such as EEG) and enable the realization of experiments with high reliability. These functionalities involve the use of SWT browser that allows the client to monitor and control the testing process. In other words, when using the controlled mode (see Figure 3), the participant has no way to intentionally or unintentionally exit the test by, for example, pressing alt + F4. Other common functions of web browsers are also strictly disabled, such as page refreshing or opening menus by right-clicking the mouse. The second key functionality is the recording of two time sets in the database. To avoid the problem of slow internet connection, both server time and local PC time are recorded, which means that events on the client side can be accurately synchronized (e.g., synchronizing stimulus exposition with data from the eye-tracker).

Researchers can effectively create new test batteries thanks to a combination of a number of subfunctions and tools. Emphasis is also placed on the efficiency of the software. Researchers can effectively change the content of already finished test slides and create derivatives from sample

Table 1: Summary of calibration results for all participants.

Participant	SMI X	SMI Y	EyeTribe
P01	0,4	0,2	Good
P02	0,3	0,1	Poor
P03	0,4	0,6	Moderate
P04	0,4	0,4	Perfect
P05	0,9	0,5	Good
P06	0,3	0,5	Redo
P07	0,2	0,4	Moderate
P08	0,6	0,3	Moderate
P09	0,4	0,1	Perfect
P10	0,3	0,4	Poor
P11	0,6	0,3	Poor
P12	0,5	0,5	Moderate
P13	0,3	0,3	Moderate
P14	0,4	0,6	Poor

templates through the modules for user access administration and also export structured results.

Hypothesis software is freely available for collaboration on a various research topic in the Czech Republic and abroad. Access to the database and modules is provided after registration.

3. In-Depth Analysis of Cognitive Processes Using Eye-Tracking System

3.1. EyeTribe Tracker. Eye-tracking technology is becoming increasingly cheaper, both on the hardware and on the software front. Currently, the EyeTribe tracker is the most inexpensive commercial eye-tracker in the world, at a price of $99. More information about the device is available at the web page of the manufacturer (https://theeyetribe.com/). The low-cost makes it a potentially interesting resource for research, but no objective testing of its quality has been performed as of yet [25]. Dalmaijer in his study [25] with five participants compared the EyeTribe tracker with high-frequency EyeLink 1000. He states that concurrent tracking by both devices of the same eye-movements proved to be impossible, due to the mutually exclusive way in which both devices work. One of the reasons was that EyeLink uses only one eye for the recording. Delmaijer [25] also states that recording with both devices at the same time results in deterioration of results of both and often leads to a failure to calibrate at least one. Ooms et al. [26] compared EyeTribe with SMI RED 250 but also did not use the concurrent recording. In our study, we compared the EyeTribe tracker with SMI RED 250. In our case, we have not noticed any problems with calibration (see Table 1).

3.2. Methods of EyeTribe Accuracy Evaluation. For the comparison study, recording with SMI RED 250 and the EyeTribe tracker at the same time was performed. Laboratory setup is displayed in Figure 4. The EyeTribe tracker stands in front of the SMI device.

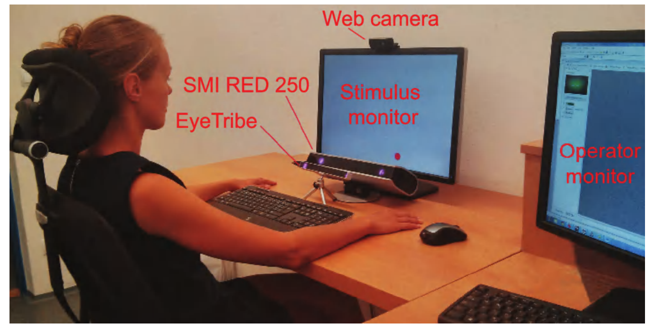


Figure 4: Laboratory setting for EyeTribe and SMI accuracy comparison.

EyeTribe tracker was connected with the OGAMA software [27], where the experiment with six static image stimuli was prepared. At the same time, screen recording experiment was created in SMI experiment center (sampling frequency was set up to 60 Hz, to be the same as EyeTribe). Both devices were calibrated separately (but the eye-trackers were at their positions and turned on).

After calibrations, recording with SMI started. After that, experiment with static images in OGAMA was performed. That means the SMI device recorded the experiment data as well (as a screen recording video). The whole experiment procedure was done with fourteen participants. The purpose of the study was to verify how trustworthy data from EyeTribe tracker are. Recorded fixations from both eye-trackers were compared qualitatively and quantitatively. A diagram of the whole recording procedure is displayed in Figure 5.

For the comparison of recorded data from both devices, the OGAMA environment was used. Data from EyeTribe were displayed in OGAMA directly; SMI data had to be converted. For this conversion, the tool smi2ogama developed by S. Popelka was used. The tool is available at http://eyetracking.upol.cz/smi2ogama/.

The recorded screen data were cropped according to the pertinence to individual stimuli. For that, recorded key presses (for a slide change) were used.

3.3. Participants. Total of 14 respondents participated in this part of the study (ten males and four females with an average age of 29.5). They were employees and postgraduate students of department of geoinformatics. 16-point calibration was used for both devices. Results of calibration are summarized in Table 1. With the EyeTribe, it was almost not possible to achieve perfect calibration result. Figure 6 shows the details of calibration results for participant P03. The results in OGAMA show calibration result for each of the 16 calibration points (with the use of colour); SMI shows only the average value in degrees of visual angle for axes X and Y.

For all recordings, I-DT fixation detection in OGAMA was used with the same settings. A value of 20 px was used as "maximum distance"; "minimum number of samples" was set up to 5. More information about fixation detection settings is available in [28, 29].

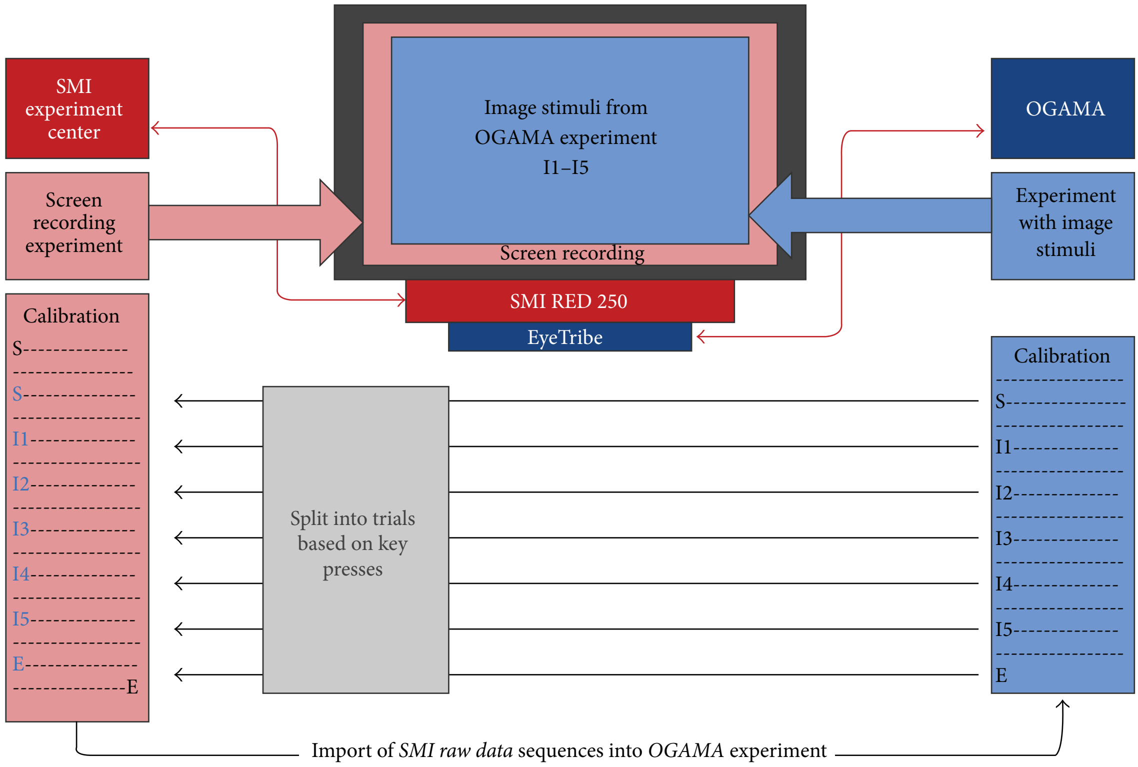


FIGURE 5: Diagram of concurrent eye-movements recording with SMI RED 250 and EyeTribe.

(a) (b)

FIGURE 6: Calibration results from EyeTribe (a) and SMI RED (b) for participant "P03."

3.4. Stimuli. The experiment contained six static images. The first one contained a grid with nine numbers; second one (Slide 2, Figure 7) contained sixteen numbers. The task of the participants was to read numbers in ascending order (from top to the bottom). Next three stimuli contained different types of maps, but the results of these stimuli are not described in this paper. The last stimulus (Slide 6, Figures 8 and 9) contained a map of the world and respondents' task was to move the eyes around Africa.

3.5. Results and Discussion of EyeTribe Evaluation. Eye-movement data recorded from participant P03 are displayed in Figure 7. Red points represent fixations from SMI, and blue points are fixations from EyeTribe. The task in this stimulus was only to read the numbers.

From Figure 7, it can be seen that both devices recorded around one or two fixations over each number. The accuracy of the recording is comparable. Accuracy reflects the eye-tracker's ability to measure the point of regard and is defined

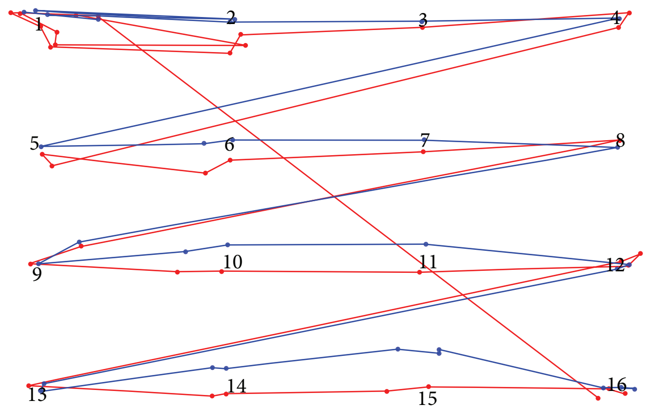


Figure 7: Comparison of recorded eye-movement data from participant P03 in Slide 2 from EyeTribe (blue) and SMI RED (red).

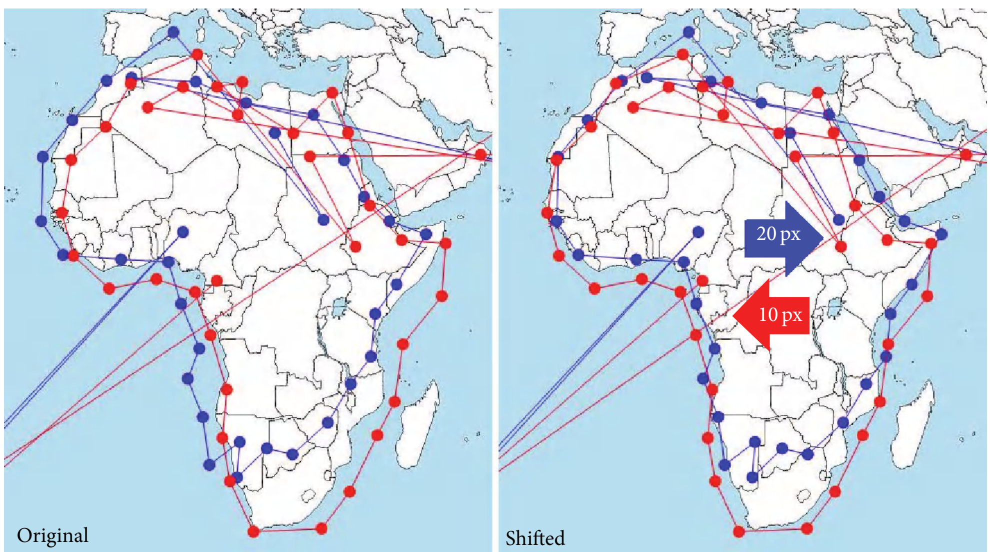


Figure 8: Comparison of recorded eye-movement data from participant P03 in Slide 6 from EyeTribe (blue) and SMI RED (red).

as the average difference between a test stimulus position and the measured gaze position [30]. The largest deviations of the EyeTribe tracker data were observed for two points in the middle of the bottom line. This situation was observed in almost all recorded data. The situation can be seen in Figure 7 in the case of points 14 and 15 (middle points in the lowest line of numbers). Gaze position recorded by EyeTribe is shifted upwards.

Another example is visible in Figure 8, which is the crop of Slide 6 stimuli. In this stimulus, the task was to move the eyes around the continent of Africa on the map. The data recorded by EyeTribe tracker were moved to the left by 20 px, but this systematic error can be corrected by a manual shift of fixations in OGAMA. This situation is depicted in Figure 8. On the left side, original data are displayed. On the right, data after horizontal shift (20 px to the right for EyeTribe and 10 px to the left for SMI) are depicted. Eye-movement data from EyeTribe for horizontally central fixations are shifted upwards, especially in the bottom part of the stimuli. See Figure 12 for more detailed analysis of fixation locations. The same issue was reported in all stimuli for most of the participants. Visualization of gaze trajectories of all participants is in Figure 9. The solution for dealing with this inaccuracy is to avoid placing important parts of the stimulus to the bottom of the screen. It will be possible to compare recorded raw data, but, in cartographic research, fixations are used for analysis, so it was more meaningful to compare fixations (identified with the same algorithm).

As an alternative for the comparison of raw data, comparison of data loss was performed. In the case of SMI recordings, average data loss (samples with coordinates 0, 0) was 0.57% of all recorded data. With the EyeTribe, the average data loss was 1.22%. Although the value is more than twice higher than in the case of SMI, it is still acceptable.

The graph in Figure 10 shows the percentage of data loss for Slide 2. It is evident that data loss is higher in the case of EyeTribe recordings, but, in most cases, less than 2% of data is missing. The highest values were observed for participants

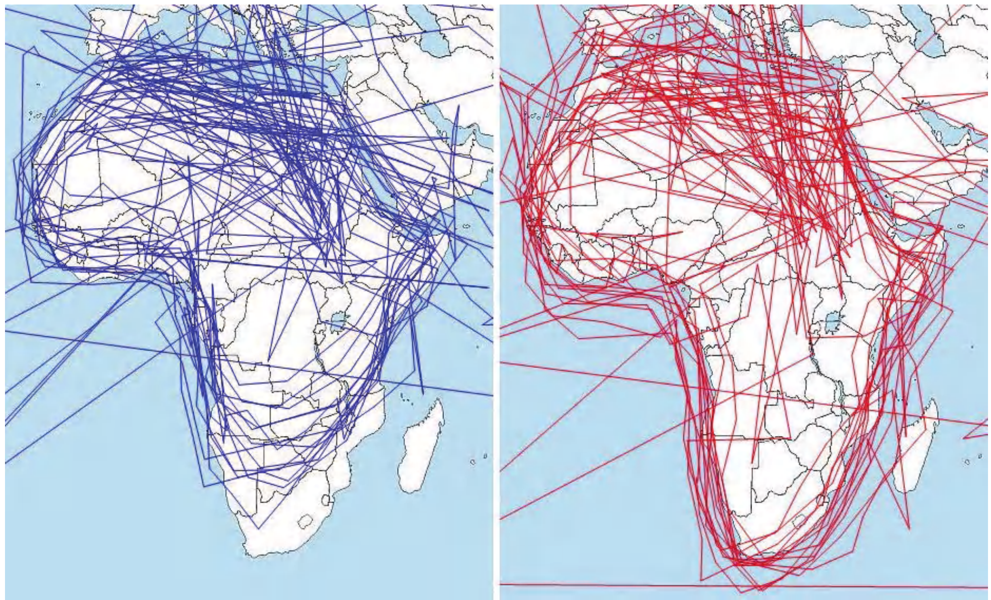


Figure 9: Problems with data recorded by EyeTribe (blue) at the bottom of the stimuli in comparison with SMI data (red).

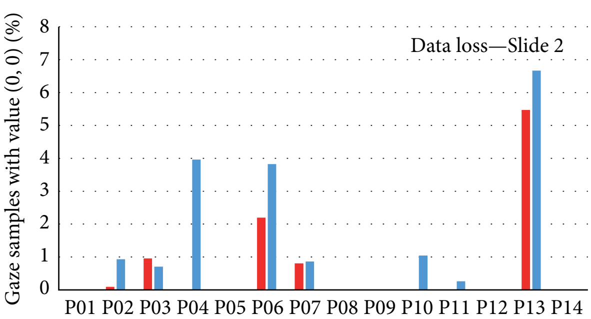


Figure 10: Comparison of data losses of fourteen participants during observation of Slide 2. Red bars represent SMI RED 250; blue ones represent EyeTribe tracker.

P06 and P13. Participant P06 had the worst calibration from all respondents. Participant P13 has worn glasses which can possibly cause the high data loss.

In the next step of accuracy evaluation, values of eye-tracking metric fixation count recorded by SMI RED 250 and the EyeTribe tracker were compared for all six stimuli in the experiment. A summary of the results is shown in Figure 11. The correlation between numbers of detected fixations was between 0.949 and 0.989 with the exception of participant P13 with the correlation of 0.808. The ratio between a number of recorded fixations with SMI device and EyeTribe was also investigated. On average, EyeTribe recorded 88.2% of fixations that were recorded by SMI device. The correlation and ratio values for each participant are presented as part of Figure 11.

Beside the number of fixations, their location was compared. For this evaluation, Slide 2 with a grid of 16 numbers was chosen (Figure 7). For each participant, the deviations between coordinates of the target (number) and closest fixation were calculated. The graphs in Figure 12 show the median size and direction of the deviation for each of the 16 targets in the stimuli. It is evident that the largest deviations (heading upwards) for EyeTribe were observed for the points in the bottom part of the image (numbers 14 and 15). Each graph contains the value of the Euclidean distance of median deviations from the origin. Average deviation was 26 px for EyeTribe and 22 px for SMI.

The evaluation of truthfulness was performed on fourteen participants. According to Nielsen [31], this number should be sufficient. The evaluation of qualitative (Figures 7, 8, and 9) and quantitative (Figures 10, 11, and 12) data indicates that accuracy of low-cost EyeTribe tracker is sufficient for the use in cartographic research. Similar results were found by Ooms et al. [26], who measured the accuracy by the distance between recorded fixation locations and the actual location.

The limitation of the low-cost device is the sampling frequency, which is only 60 Hz (compare with 250 Hz of SMI RED eye-tracker). Another problem is shift of fixation locations in the bottom part of the screen. Taking into account described limits of the device, the EyeTribe may be an appropriate tool for cartographic research.

4. Integrated Research System: Interconnection of Hypothesis Software and EyeTribe

As one of the practical applications of the mixed-research experiment design, the Hypothesis software interconnected with the EyeTribe tracker was chosen. For the recording of eye-tracking data, the OGAMA software was used because the EyeTribe tracker is intended for developers and contains no software for data recording and analysis. OGAMA has an inbuilt slide show viewer, but the range of functionality of this viewer in comparison with SW Hypothesis is quite

FIGURE 11: Comparison of fixation count eye-tracking metric for fourteen participants. EyeTribe data are displayed as blue line; SMI data are displayed as red line.

limited. Desktop application OGAMA principally does not allow working with web-based interactive maps and mouse clicks are recorded but not shown. Oppositely, Hypothesis visualizes clicks and allows drawing of lines and polygons. This functionality is crucial in the context of working with maps. Because of this functionality, Hypothesis connected to OGAMA via HypOgama was used.

4.1. Methods of Hypothesis and EyeTribe Interconnection. For the study, a simple Hypothesis experiment containing five stimuli (intro, three pairs of maps, and last slide) was used. Participants' task was to identify the differences between the maps. Coordinates of the clicks representing differences were also recorded.

OGAMA experiment was designed with only one screen recording stimulus. OGAMA in version 5.0 can record dynamic web stimuli, but it is not possible to use slides from Hypothesis as separate stimuli.

Recorded data were split according to their belonging to particular slides in the Hypothesis experiment. For the split, timestamps from Hypothesis indicating the slide change were used. The splitting and conversion of recorded data

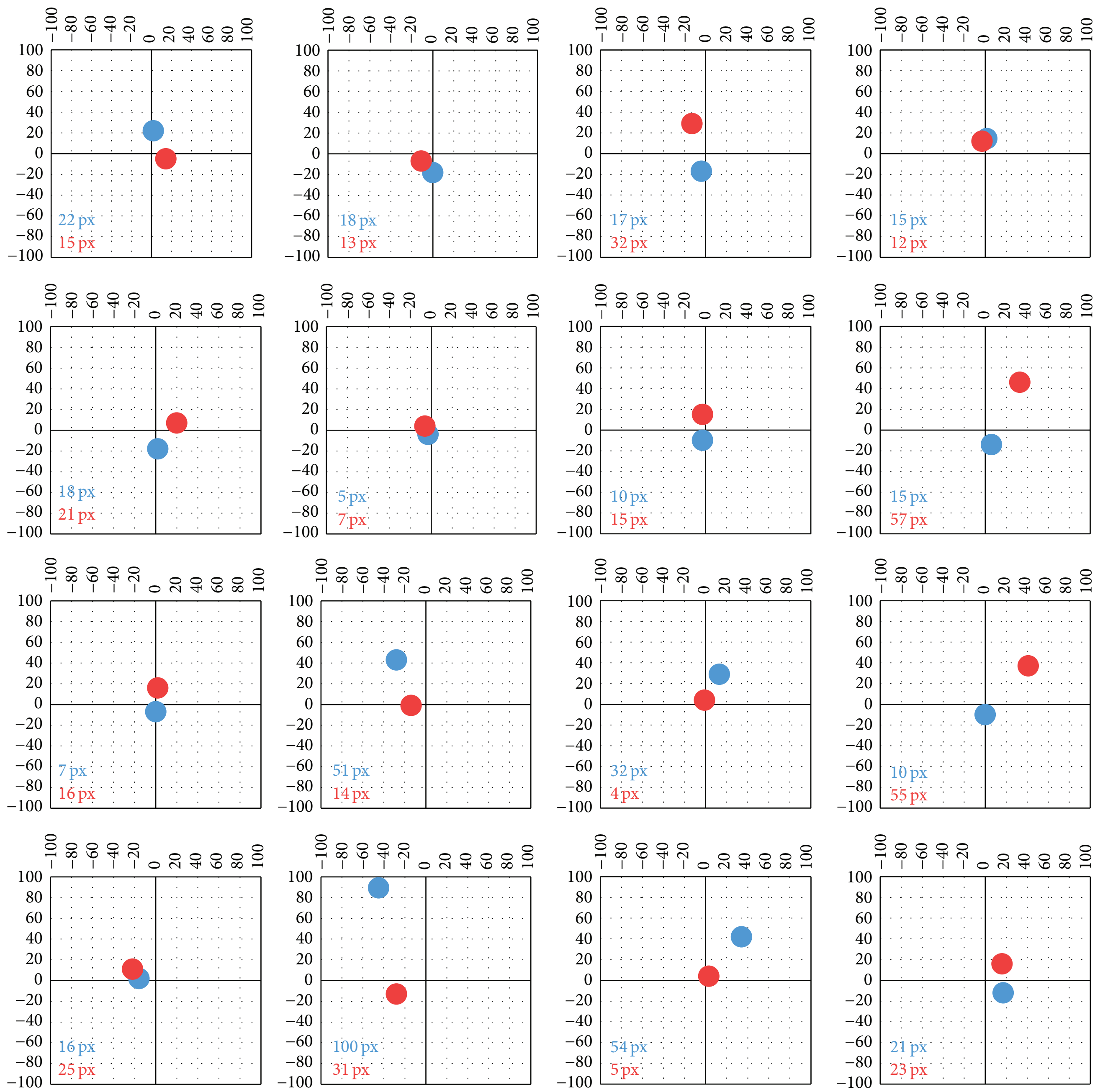


FIGURE 12: Comparison of fixation positions in Slide 2 for fourteen participants. Distance from the center of the image shows fixation deviation in pixels. EyeTribe data are displayed as blue dots; SMI data are displayed as red dots.

manually were time-consuming and not user-friendly. Thus, a web application called HypOgama was written in PHP for the automation of the process. The functionality of HypOgama application is illustrated in Figure 13.

The HypOgama application (Figure 14) is freely available at http://eyetracking.upol.cz/hypogama/.

The application synchronizes the Hypothesis time with the timestamp from the eye-tracking recording in OGAMA. The synchronization is processed by the key press that was used to start the Hypothesis experiment and which was recorded in both systems—in Hypothesis and OGAMA.

In the next step, the application scans the Hypothesis file and finds the timestamps of slide changes. These timestamps are then used for splitting raw eye-tracking data into blocks belonging to particular slides. The name of the relevant stimuli is added to all records from each block. In the final step, the data structure is modified for the direct import into a new OGAMA project.

The application contains six input fields:

(1) Exported file from Hypothesis manager containing data for one participant.

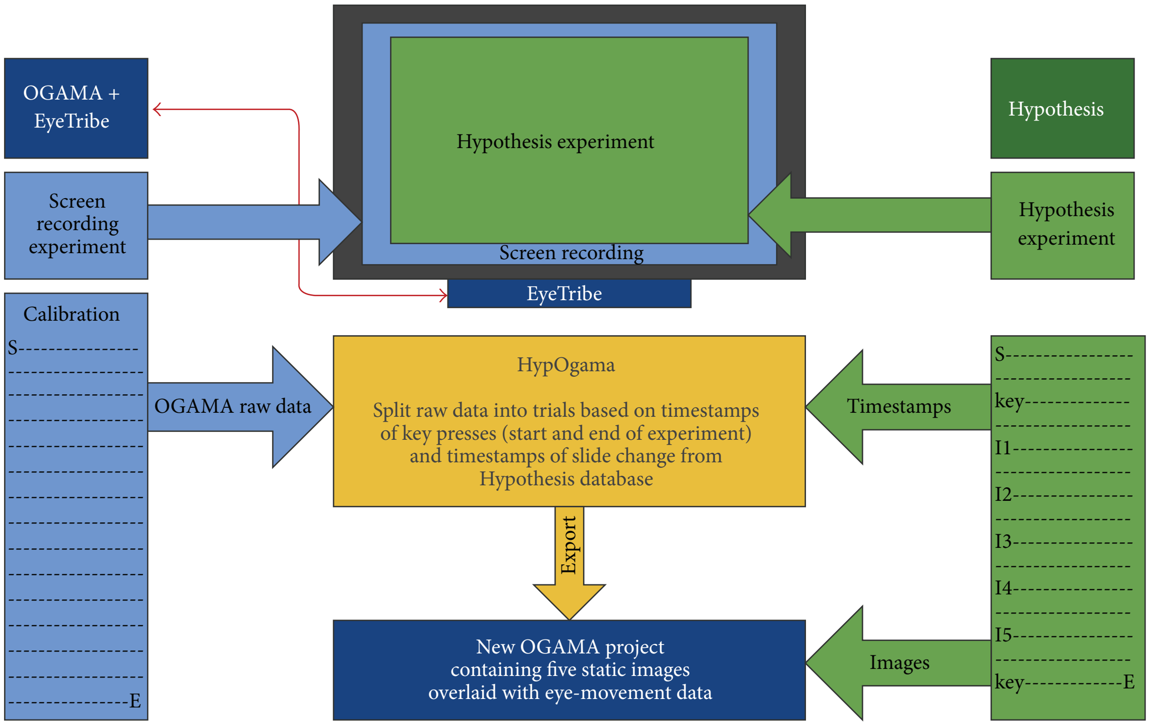


Figure 13: Process of splitting recorded data (screen recording) into trials with the use of HypOgama web application.

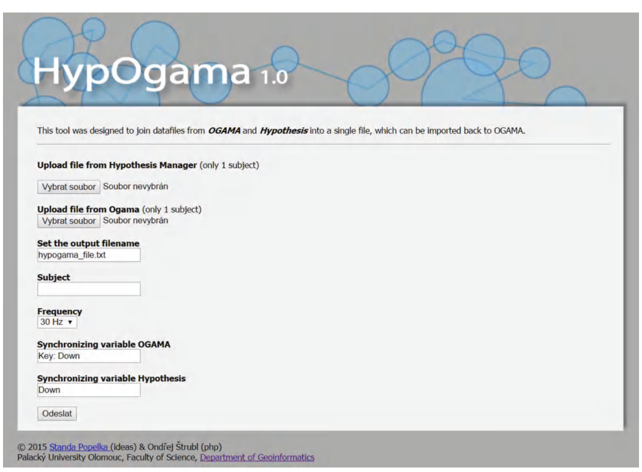


Figure 14: Environment of HypOgama web application.

(2) Exported raw data from the OGAMA application for one participant.

(3) Name of the output file.

(4) Subject name (if blank, the ID from Hypothesis will be used).

(5) Frequency of an eye-tracker (30 or 60 Hz).

(6) Synchronization variables: these values indicate which key was used for the synchronization of Hypothesis and OGAMA (default value is “Key: Down” in OGAMA format and “Down” in the format of Hypothesis application).

In the Hypothesis file (ad 1), HypOgama finds the row with the key press (default Key: Down) and the corresponding time, which corresponds to the beginning of the experiment. In the next step, the column containing the slide names is scanned and the time of the first occurrence of each slide is also stored. According to this time, OGAMA recording is split. The last information obtained from the Hypothesis file is the name of the subject, overwriting the subject name in the OGAMA file.

In OGAMA file, all records prior to the synchronization key press are erased. Stimuli names are replaced by those from Hypothesis file.

Outputs of the created script are raw eye-movement data for each slide that could be directly imported into the OGAMA project. The only one necessary thing is to put image files (stimuli) into OGAMA project folder. If it is the same filename as the one contained in the Hypothesis file, images will be automatically assigned to proper data. After the whole process, a user has OGAMA project containing static image stimuli with all corresponding eye and mouse movement data. The proposed concept was applied and verified through a selected case study described below. The purpose of this short study was to illustrate the functionality of interconnection of EyeTribe and OGAMA.

For the verification of the designed process of Hypothesis and EyeTribe combination, simple test battery was designed. For chosen procedure, Hypothesis was used for large-scale quantitative approach and eye-tracking method for the subsequent specification of certain results.

The test battery was established in the Hypothesis software and was focused on verification of Gestalt principles,

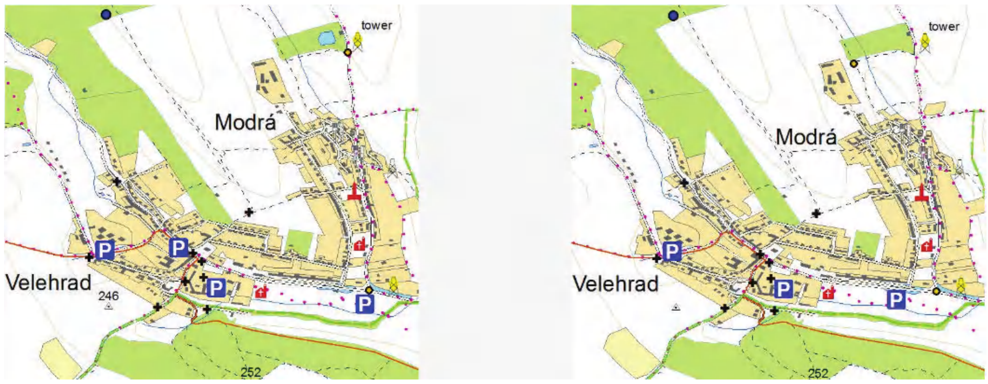


FIGURE 15: Example of stimuli—the first pair of topographic maps.

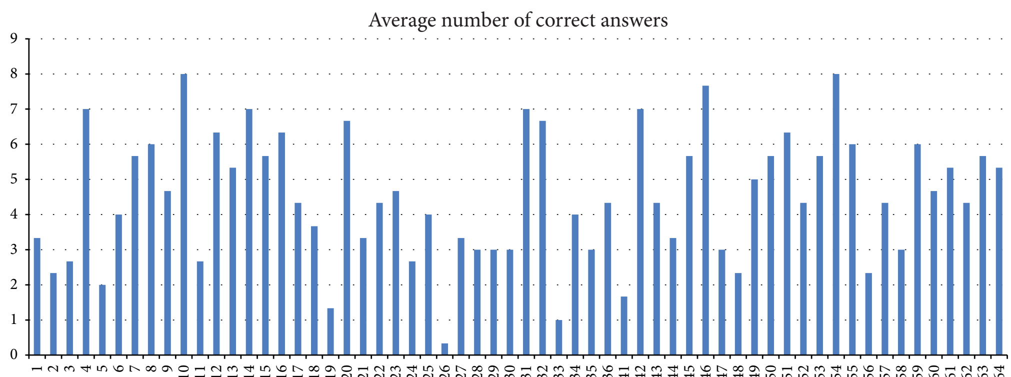


FIGURE 16: An example of results from Hypothesis. An average number of correct answers for each of the participants.

respectively, figure-ground organization, and on the cross-cultural comparison in the context of visual perception of cartographic stimuli [22, 32–35] on the example of specific cartographic products. The cartographic tasks were part of these more complex research batteries. The main purpose of this short cartographic study was the verification of HypOgama application and whole integrated research system for further research studies.

4.2. Participants. Participants of this illustrative case study were 64 students from the Masaryk University, Czech Republic, and 64 students from Wuhan University, China. In the first phase, participants were tested only on the web-based platform Hypothesis. Only a half of the dataset (Czech population) was further used in context of this particular study where the topographic and thematic maps were compared. In the second phase, the experiment was conducted with the use of eye-tracking system and the research sample is still continually extended.

4.3. Stimuli. The stimuli were represented by three pairs of maps that differed in 10 variables, for example, different colours of map signs, different position of the signs, and missing map signs. First two pairs of stimuli contained topographic maps. The third pair of the maps contained a thematic map.

The test was structured in three main parts. In the first part, participants filled out a personal questionnaire; in the second part, a representative example of the stimuli was presented to familiarize the participants with the environment of Hypothesis. In the third part, three tasks containing pairs of stimuli described above were presented. Participants were asked to mark the differences between presented maps. The time limit for each task was 45 seconds. An example of a topographic map (Slide 1) is displayed in Figure 15. On Slide 2, similar topographic map in different scale was shown. The last slide contained thematic map (see Figure 17).

4.4. Results and Discussion of Hypothesis and EyeTribe Interconnection. The performed study verified stability of proposed system on long distances and, at the same time, part of the test battery was used as a pilot study to verify the functionality of an integrated research system. Stimuli comparing the effectiveness of visual search between topographic and thematic maps were selected.

In the first phase, the test was performed in the Hypothesis application only. A number of differences identified between pairs of maps on Czech population were analysed (see Figure 16).

In the case of two pairs of topographic maps, the average number of correct answers was four. In the case of the stimuli

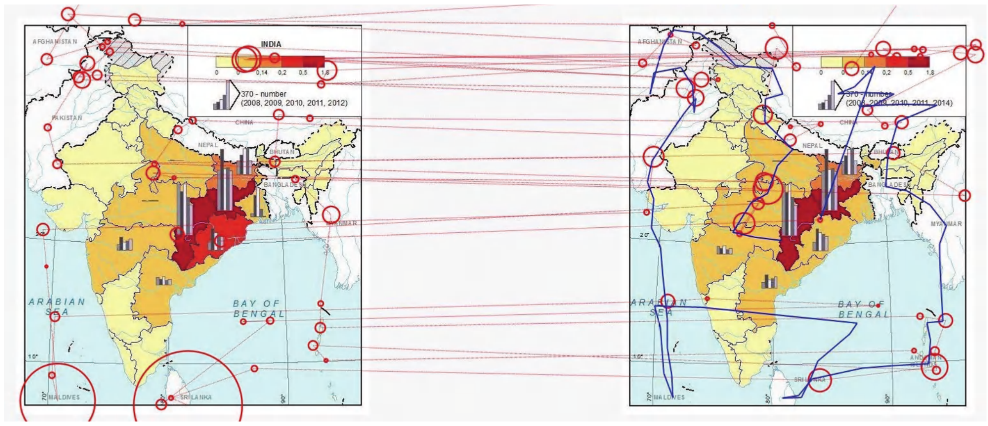


Figure 17: Example of eye-movement data recorded during the Hypothesis experiment. Circles represent fixations; blue line on the right is a mouse trajectory.

with a thematic map, the average number of correct answers was five.

To generalize the findings, an increase of the number of maps per condition would be necessary. However, this difference was the first clue to establish working hypotheses. Based on the data from the first phase of testing, hypotheses were established only at the level of stimulus-reaction. The way of task processing by users and their solving strategies were still a black-box; thus there was a need for more detailed procedural data, especially for information about distinct search strategies.

To explore differences in the visual search, eye-tracking can be used due to the ability to provide more detailed information (e.g., which kind of object was omitted, which kind of object could be found at first glance, and which areas attracts main attention).

Therefore, in the second phase, the already used experimental battery created in Hypothesis was interconnected with OGAMA through HypOgama application and the experiment was launched with the EyeTribe system. Cartographic stimuli and the eye-tracking data were linked together and further analysed with OGAMA.

The example in Figure 17 shows outputs from OGAMA-scan path and mouse trajectory of one participant over the stimulus with thematic maps. In this case, fixations are distributed mainly over the text labels in the map. Participant did not find the difference in the colour of the Odisha state (on the east coast of India) under the relatively large graph. At the same time, eye-tracking metrics (e.g., fixation count, dwell time for each map, and a number of saccades between these maps) can be statistically analysed. Based on findings from both types of analyses, the hypotheses for subsequent study can be established.

The functionality of the integrated research system has been fully verified in the above-mentioned pilot study. The experiment created on the Hypothesis platform was connected with OGAMA and EyeTribe via HypOgama. Data capture including eye-tracking recording continued and exploratory analyses of these data were performed.

5. Conclusion

The aim of the paper was to prove the concept of the mixed-research design through the interconnection of Hypothesis (software for experiment creation, experiment execution, and data collection) and the EyeTribe tracker (the most inexpensive commercial eye-tracker). This system could prove to be a valuable tool for cognitive cartography experiments and evaluation of user behaviour during map reading process.

The first necessary step was to evaluate the accuracy of the EyeTribe tracker with the use of concurrent recording together with the SMI RED 250 eye-tracker. The results of the comparison show that the EyeTribe tracker can be a valuable resource for cartographical research.

The next part of the study was focused on the interconnection of the EyeTribe with the Hypothesis platform, developed at Masaryk University in Brno. The connection was made through a newly created web application that modifies eye-movement data recorded during screen recording experiment in the OGAMA open-source application. The application is publicly available for the community of cartographers and psychologists at web page http://eyetracking.upol.cz/hypogama.

The interconnection advantages were illustrated on an example of simple case study containing three pairs of maps. The performed case study demonstrated the ability of the combined system of the Hypothesis platform and the EyeTribe tracker to support each other and to serve as an effective tool for cognitive studies in cartography.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This paper was supported by projects of Operational Program Education for Competitiveness (European Social Fund) (Projects CZ.1.07/2.3.00/20.0170 and CZ.1.07/2.3.00/30.0037) of the Ministry of Education, Youth and Sports of the Czech Republic and the Student Project IGA_PrF_2015_012 of the Palacky University.

References

[1] G. L. Allen, C. R. Miller Cowan, and H. Power, "Acquiring information from simple weather maps: influences of domain-specific knowledge and general visual-spatial abilities," *Learning and Individual Differences*, vol. 16, no. 4, pp. 337–349, 2006.

[2] D. Edler, A.-K. Bestgen, L. Kuchinke, and F. Dickmann, "Grids in topographic maps reduce distortions in the recall of learned object locations," *PLoS ONE*, vol. 9, no. 5, Article ID e98148, 2014.

[3] P. Kubíček, Č. Šašinka, and Z. Stachoň, "Vybrané kognitivní aspekty vizualizace polohové nejistoty v geografických datech," *Geografie*, vol. 119, no. 1, pp. 67–90, 2014.

[4] E. S. Nelson, "Using selective attention theory to design bivariate point symbols," *Cartographic Perspectives*, vol. 32, pp. 6–28, 1999.

[5] E. S. Nelson, "Designing effective bivariate symbols: the influence of perceptual grouping processes," *Cartography and Geographic Information Science*, vol. 27, no. 4, pp. 261–278, 2000.

[6] A. H. Duc, P. Bays, and M. Husain, "Eye movements as a probe of attention," in *Progress in Brain Research*, C. Kennard and R. J. Leigh, Eds., vol. 171, chapter 5.5, pp. 403–411, Elsevier, 2008.

[7] N. Andrienko and G. Andrienko, *Exploratory Analysis of Spatial and Temporal Data. A Systematic Approach*, Springer, New York, NY, USA, 2005.

[8] J. M. Enoch, "Effect of the size of a complex display upon visual search," *Journal of the Optical Society of America*, vol. 49, no. 3, pp. 280–286, 1959.

[9] A. Hyrskykari, S. Ovaska, P. Majaranta, K.-J. Räihä, and M. Lehtinen, "Gaze path stimulation in retrospective think-aloud," *Journal of Eye Movement Research*, vol. 2, no. 4, pp. 1–18, 2008.

[10] Ö. Alaçam and M. Dalcı, "A usability study of WebMaps with eye tracking tool: the effects of iconic representation of information," in *Human-Computer Interaction. New Trends*, vol. 5610 of *Lecture Notes in Computer Science*, pp. 12–21, Springer, Berlin, Germany, 2009.

[11] S. I. Fabrikant, S. Rebich-Hespanha, N. Andrienko, G. Andrienko, and D. R. Montello, "Novel method to measure inference affordance in static small-multiple map displays representing dynamic processes," *Cartographic Journal*, vol. 45, no. 3, pp. 201–215, 2008.

[12] S. I. Fabrikant, S. R. Hespanha, and M. Hegarty, "Cognitively inspired and perceptually salient graphic displays for efficient spatial inference making," *Annals of the Association of American Geographers*, vol. 100, no. 1, pp. 13–29, 2010.

[13] K. Ooms, P. De Maeyer, and V. Fack, "Study of the attentive behavior of novice and expert map users using eye tracking," *Cartography and Geographic Information Science*, vol. 41, no. 1, pp. 37–54, 2014.

[14] S. Popelka and A. Brychtova, "Eye-tracking study on different perception of 2D and 3D terrain visualisation," *Cartographic Journal*, vol. 50, no. 3, pp. 240–246, 2013.

[15] J. M. Olson, "Cognitive cartographic experimentation," *Cartographica*, vol. 16, no. 1, pp. 34–44, 1979.

[16] Z. Štěrba, Č. Šašinka, Z. Stachoň, P. Kubíček, and S. Tamm, "Mixed research design in cartography: a combination of qualitative and quantitative approaches," *Kartographische Nachrichten*, vol. 64, no. 5, pp. 262–269, 2014.

[17] J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, Sage, Thousand Oaks, Calif, USA, 2nd edition, 2003.

[18] Z. Štěrba, Č. Šašinka, Z. Stachoň, K. Morong, and R. Štampach, *Selected Issues of Experimental Testing in Cartography*, Masaryk University, 2015, http://munispace.muni.cz/index.php.

[19] M. Konečný, Š. Březinová, M. V. Drápela et al., *Dynamická Geovizualizace v Krizovém Managementu*, Masarykova Univerzita, Brno, Czech Republic, 2011 (Czech).

[20] P. Kubíček, Č. Šašinka, and Z. Stachoň, "Uncertainty visualisation testing," in *Proceedings of the 4th Conference on Cartography and GIS*, Sofia, Bulgaria, June 2012.

[21] Z. Stachoň, Č. Šašinka, P. Kubíček, and Z. Štěrba, "MuTeP—alternativní nástroj pro testování kartografických vizualizací a sběr dat," in *Proceedings of the 23rd Sjezd České Geografické Společnosti*, Prague, Czech Republic, 2014 (Czech).

[22] Z. Stachoň, Č. Šašinka, Z. Štěrba, J. Zbořil, Š. Březinová, and J. Švancara, "Influence of graphic design of cartographic symbols on perception structure," *Kartographische Nachrichten*, vol. 63, no. 4, pp. 216–220, 2013.

[23] K. Morong and Č. Šašinka, "Hypothesis—online software platform for objective experimental testing," in *Proceedings of the Applying Principles of Cognitive Psychology in Practice*, Brno, Czech Republic, May 2014.

[24] S. W. T. Widgets, https://www.eclipse.org/swt/.

[25] E. Dalmaijer, "Is the low-cost EyeTribe eye tracker any good for research?" *PeerJ PrePrints*, vol. 2, Article ID e585v1, 2014.

[26] K. Ooms, L. Dupont, L. Lapon, and S. Popelka, "Accuracy and precision of fixation locations recorded with the low-cost Eye Tribe tracker in different experimental set-ups," *Journal of Eye Movement Research*, vol. 8, no. 1, pp. 1–24, 2015.

[27] A. Voßkühler, V. Nordmeier, L. Kuchinke, and A. M. Jacobs, "OGAMA (Open Gaze and Mouse Analyzer): open-source software designed to analyze eye and mouse movements in slideshow study designs," *Behavior Research Methods*, vol. 40, no. 4, pp. 1150–1162, 2008.

[28] S. Popelka, "Optimal eye fixation detection settings for cartographic purposes," in *Proceedings of the 14th SGEM GeoConference on Informatics, Geoinformatics and Remote Sensing (SGEM '14)*, vol. 1, pp. 705–712, June 2014.

[29] S. Popelka and J. Doležalová, "Non-photorealistic 3D visualization in city maps: an eye-tracking study," in *Modern Trends in Cartography*, pp. 357–367, Springer, 2015.

[30] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye Tracking: A Comprehensive Guide to Methods and Measures*, Oxford University Press, Oxford, UK, 2011.

[31] J. Nielsen, *Why You Only Need to Test with 5 Users*, Nielsen Norman Group, 2000, http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/.

[32] J. Čenek, "Individualism and collectivism and their cognitive correlates in cross-cultural research," *The Journal of Education, Culture and Society*, vol. 2, pp. 210–225, 2015.

[33] R. E. Nisbett, I. Choi, K. Peng, and A. Norenzayan, "Culture and systems of thought: holistic versus analytic cognition," *Psychological Review*, vol. 108, no. 2, pp. 291–310, 2001.

[34] E. Rubin, "Figure and ground," in *Visual Perception*, S. Yantis, Ed., pp. 225–229, Psychology Press, Philadelphia, Pa, USA, 2001.

[35] J. Wagemans, J. H. Elder, M. Kubovy et al., "A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization," *Psychological Bulletin*, vol. 138, no. 6, pp. 1172–1217, 2012.

Paper Glyphs

Opach, T.*, **Popelka, S.**, Dolezalova, J., & Rød, J. K. (2017). Star and polyline glyphs in a grid plot and on a map display: which perform better? Cartography and Geographic Information Science, 1-20. https://doi.org/10.1080/15230406.2017.1364169

[IF 1.785; Q2]

[20 citations on WoS; 14 citations without autocitations]

[Author's contribution: 40%]

Abstract

Glyphs are small geometric shapes that in geovisualization are often used to represent multidimensional spatial data. The aim of this study is to investigate the effectiveness of their two types - star and polyline glyphs, as they can encode the same message and can provide similar functionality. Thus, if the two glyph types are similar and can be used for the same data, the question arises as to which of them better facilitates various user tasks. To address this question, an empirical study of 26 individual users is conducted to investigate differences in user performance for polyline and star glyphs shown either in a grid plot or on a map display. In this study, a task-based approach with eye-tracking is applied, as well as a subjective questionnaire and a psychological test of cognitive style. The finding is that polyline glyphs better facilitate tasks when datapoint values in glyphs are to be read, whereas star glyphs are better when a visual search among glyphs is to be done. Moreover, the results reveal that the map display works better than the grid plot. If star glyphs are to be used, the key (legend) needs to be better incorporated into a visual interface.

CARTOGRAPHY AND GEOGRAPHIC INFORMATION SCIENCE
2018, VOL. 45, NO. 5, 400–419
https://doi.org/10.1080/15230406.2017.1364169

Taylor & Francis
Taylor & Francis Group

ARTICLE

Star and polyline glyphs in a grid plot and on a map display: which perform better?

Tomasz Opach [a], Stanislav Popelka [b], Jitka Dolezalova [b] and Jan Ketil Rød [a]

[a]Department of Geography, Norwegian University of Science and Technology, Trondheim, Norway; [b]Department of Geoinformatics, Palacký University, Olomouc, Czech Republic

ABSTRACT
Glyphs are small geometric shapes that in geovisualization are often used to represent multi-dimensional spatial data. The aim of this study is to investigate the effectiveness of their two types – star and polyline glyphs, as they can encode the same message and can provide similar functionality. Thus, if the two glyph types are similar and can be used for the same data, the question arises as to which of them better facilitates various user tasks. To address this question, an empirical study of 26 individual users is conducted to investigate differences in user performance for polyline and star glyphs shown either in a grid plot or on a map display. In this study, a task-based approach with eye-tracking is applied, as well as a subjective questionnaire and a psychological test of cognitive style. The finding is that polyline glyphs better facilitate tasks when datapoint values in glyphs are to be read, whereas star glyphs are better when a visual search among glyphs is to be done. Moreover, the results reveal that the map display works better than the grid plot. If star glyphs are to be used, the key (legend) needs to be better incorporated into a visual interface.

ARTICLE HISTORY
Received 9 February 2017
Accepted 2 August 2017

KEYWORDS
Eye-tracking; glyphs; polyline glyphs; star glyphs; user study

Introduction

Glyphs are a commonly used visualization technique that provides an overview of a dataset by showing its items as separate and simplified graphical entities in the form of juxtaposed small plots or charts without any labels (Borgo et al., 2013; Gribov, Unwin, & Hofmann, 2006; Ünlü & Malik, 2011). Hence, despite the limited explanations (labeling), juxtaposition enables access to particular data items, which often is unfeasible in the case of superimposition (Opach & Rød, 2017), such as when using parallel coordinates. Glyphs can differ greatly with regard to their form (Ward, 2002, 2008). One of the most commonly used glyph types is star glyphs. Although they have been implemented in many geovisualization environments (Gribov et al., 2006; Takatsuka & Gahegan, 2002), their use is impeded by polar coordinates in which visual scanning is more time-consuming and more error-prone than reading vertical and horizontal axes (Goldberg & Helfman, 2011). Therefore, Opach and Rød (2017) propose the use of polyline glyphs that resemble polylines from parallel coordinates as an alternative to star glyphs. The two glyph types can encode the same message and can provide similar functionality. Thus, if star and polyline glyphs are so similar and can be used for the same data, the question arises as to which of them better facilitates various user tasks.

In this paper, we aim to contribute to the body of previous work by investigating the performance of a data display consisting of either star or polyline glyphs (Figure 1). Moreover, as glyphs are frequently used on map displays and grid plots such as tables or matrices, we investigate these two layout arrangements to see whether there are differences regarding their performance. Finally, we examine whether there are differences regarding user behavior (undertaken actions when interacting with a graphical interface) between those who use star glyphs and those who use polyline glyphs. If such differences exist, what lessons can be learned from an empirical study in which user behavior is investigated? Can any findings be of value to map-makers and practitioners of information visualization?

The study consists of a theoretical part and an empirical part, in which a task-based approach with eye-tracking is employed. Additionally, we use a subjective questionnaire and a psychological test to gain deeper insights into the behaviors and opinions of users of polyline and star glyphs. The paper is organized as follows. After the background section, in

CONTACT Tomasz Opach tomasz.opach@ntnu.no

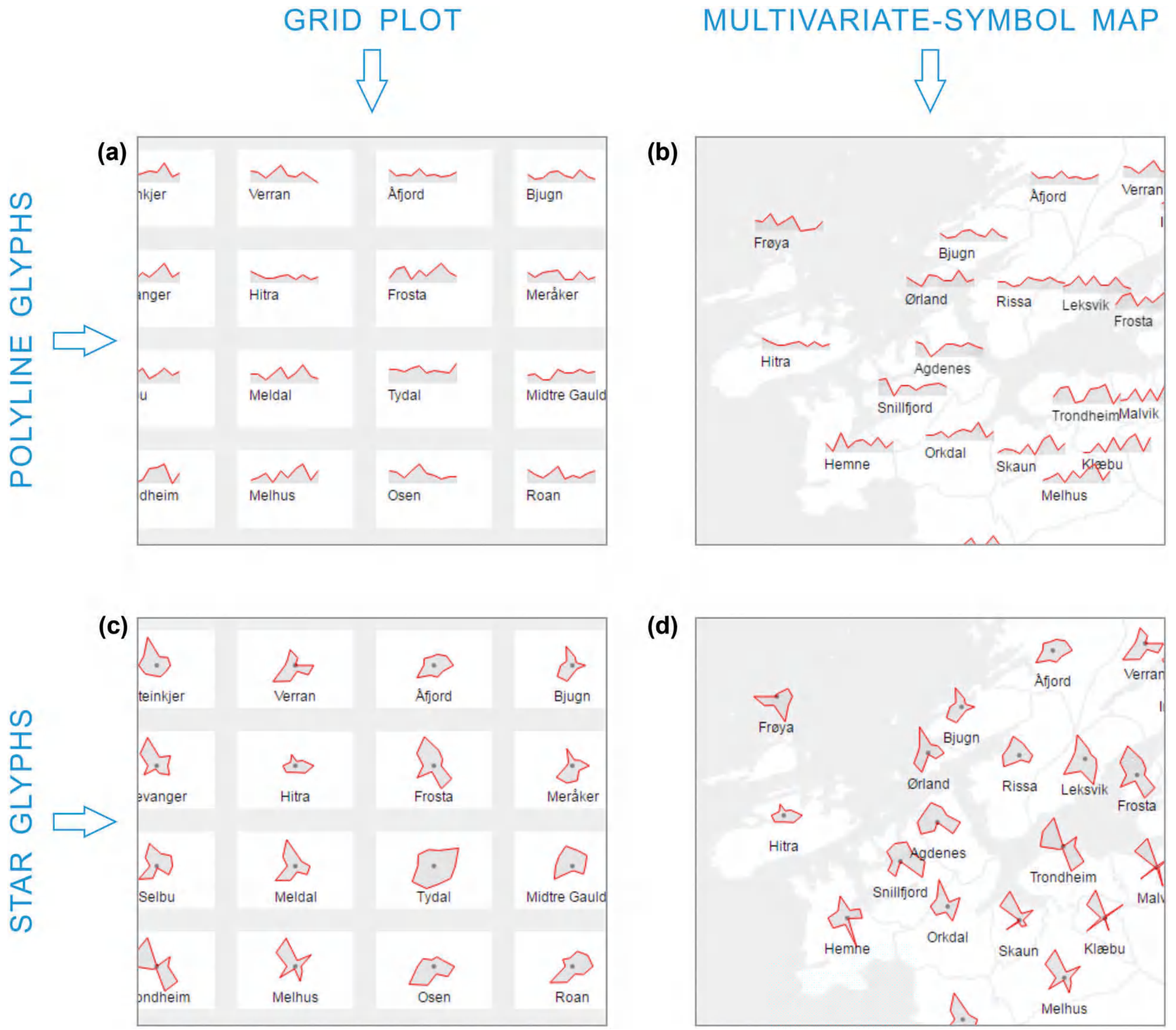


Figure 1. Polyline glyphs (a, b) and star glyphs (c, d) arranged as a grid plot (a, c) and as a multivariate symbol map (b, d).

which the state of the art in the glyph-based visualization is briefly discussed, we consider the advantages and weaknesses of star and polyline glyphs. Thereafter, we report the settings of our empirical study: its objectives, methods, and procedure. We then present and discuss the results, followed by our conclusions.

Background

Glyphs and multivariate data visualization

The growing role of information visualization in application areas such as information dashboards and business intelligence implies the need for a better understanding of various visualization techniques in general, and particularly techniques that, despite differing in form, can encode the same data. One such technique uses multivariate glyphs.

In order to encode hundreds of n-dimensional data items and show them in a limited space, the use of tiny graphical entities known as glyphs seems to be a sensible choice. Although there have been many studies of glyphs (Borgo et al., 2013), the areas in which they can be effectively used are still insufficiently studied (Ward, 2008). There are many ways in which glyphs can be used in information visualization in general and in geovisualization in particular. Glyphs seem to be suitable for visualization of multivariate vector fields (Forsberg, Chen, & Laidlaw, 2009). However, efforts needed to interpret such visualization may make glyph-based displays ineffective. Glyphs can also be embedded in a table view (Opach & Rød, 2013) or organized in a grid plot (Figure 1(a and c)) and, as part of coordinated and multiple views (CMVs), they can be dynamically linked with other visualization techniques, such as radar plots or parallel coordinates (Takatsuka & Gahegan, 2002). Additionally, glyphs can be superimposed on maps (Figure 1(b and d)) to form multivariate-symbol maps.

Glyphs in geovisualization: multivariate-symbol maps and grid plots

Glyphs have long been used within cartography and geovisualization. However, in these domains, glyphs are not a certain mapping technique to be studied. As Slocum, McMaster, Kessler, and Howard (2010, p. 337) claim, glyphs are "multivariate point symbols used to represent nonrelated attributes." Glyphs can also appear in the form of uncommon shapes such as peculiar Chernoff faces (Chernoff, 1973); such faces can be arranged as cartograms (Dorling, 1995). In general, glyphs on maps can simply be called symbols or multivariate symbols in the case in which they encode more than one variable.

Glyphs can encode geographic objects directly onto maps, and then the mapping technique can be attributed a specific name, such as a bar-chart map (Figure 2(a)), a radar-plot map, or a multivariate-symbol map. The potential of such mapping techniques has been well known in cartography for decades, since cartographers have long been concerned with ways to visualize multivariate or time-series data (Arnberger, 1977; Bertin, 1967; Ratajski, 1989; Slocum et al., 2010), and simplified multivariate symbols have long been used in thematic maps. For instance, Ostrowski and Uhorczak (1972) introduced cartotypograms as a mapping technique in which n-dimensional (typically four-dimensional) star-plots (typograms) without coordinates are used to indicate types [typical multivariate signatures, see Figure 2(b)]. This approach enables the differences between plot shapes and sizes to be clearly visible. Therefore, users are able to distinguish types among data items. Currently, the technique is known as star map, and its implementations, can be found in many geovisualization tools, such as the GeoViz Toolkit, an application derived from GeoVISTA Studio (Takatsuka & Gahegan, 2002).

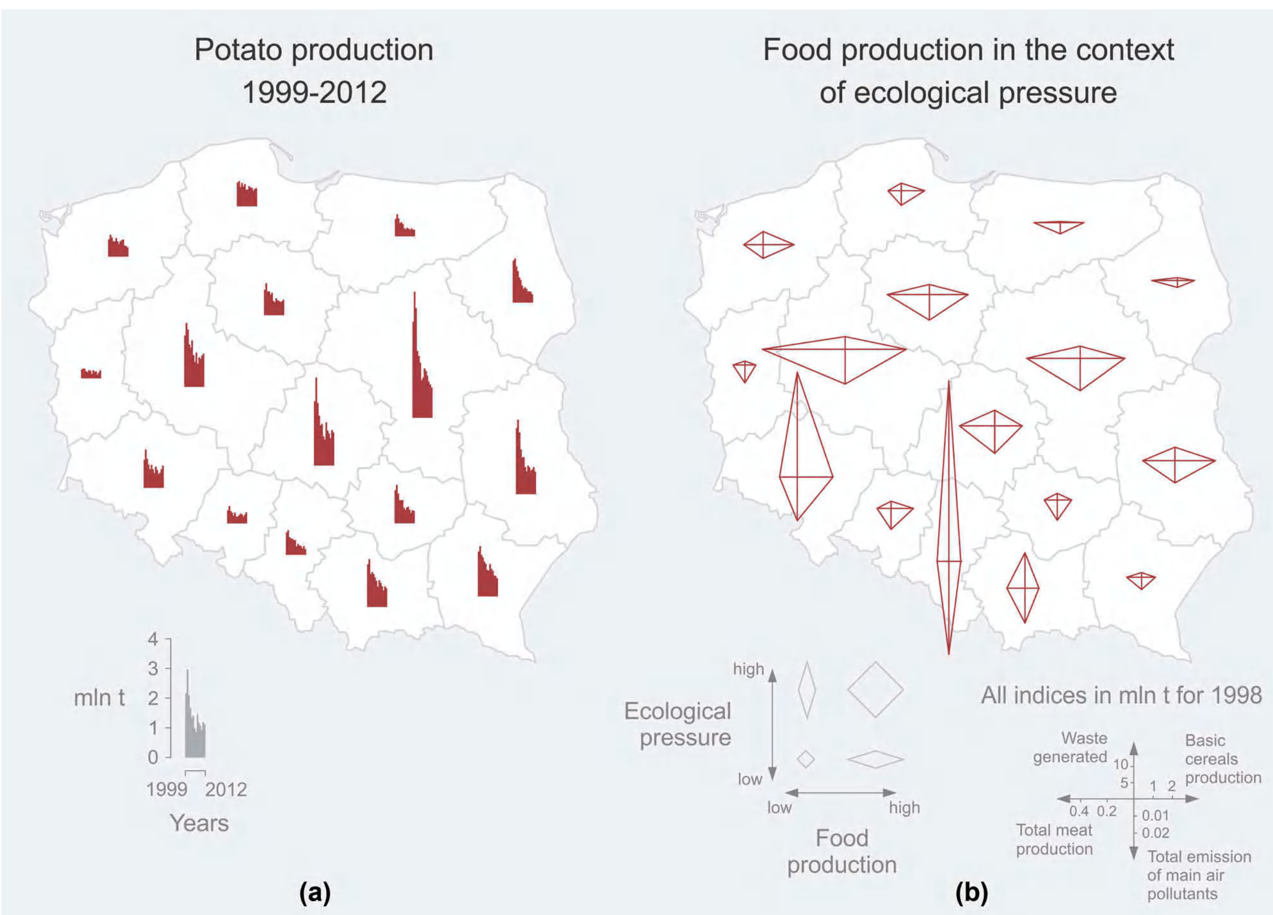


Figure 2. Two glyph maps of Poland: (a) the bar-chart map shows time-series data for potato production and (b) the cartotypogram map (star map) shows multivariate data on food-production types in the context of ecological pressure.

Recently, glyphs have been widely used in interactive map displays, but they are referred to differently, depending on the purpose of the displays [e.g. "utility symbols" (Andrienko, Andrienko, & Jankowski, 2003)]. However, in the broad geovisualization context, glyphs can be used more extensively than only as multivariate symbols on thematic map displays. They can be used differently, especially when they form part of CMV tools, whereby various visualization techniques are dynamically linked in order to facilitate information exploration and knowledge construction (Andrienko et al., 2002). In such cases, for more analytical purposes, glyphs can be grouped together and dynamically linked with other displays. They can be shown as scatterplot points (Chung et al., 2015; Gribov et al., 2006; Ünlü & Malik, 2011) or placed below or next to each other in a small multiple or a grid plot (matrix), in which "information slices are positioned within the eye span, so that viewers make comparisons at a glance–uninterrupted visual reasoning" (Tufte, 1990, p. 67). In this way, the similarities and differences between glyphs are likely to be identified more efficiently (Klippel, Hardisty, Li, & Weaver, 2009; Ward, 2008).

Star and polyline glyphs: advantages and weaknesses

Star glyphs are one of the most commonly used glyph types (Gribov et al., 2006; Ünlü & Malik, 2011). They show data items as graphic entities embedded in the polar coordinates context. Star glyphs can be thought of as a parallel coordinate plot in polar coordinates (Gribov et al., 2006; Ünlü & Malik, 2011). Therefore, in the CMV tools, in certain conditions, parallel coordinates can sometimes be replaced with star glyphs. It happens since visual attention can be shifted from parallel coordinates to star glyphs without adjusting visual reasoning (Klippel et al., 2009). Such adjusting is not needed if parallel coordinates are to be replaced with polyline glyphs (Opach & Rød, 2017) – graphical entities that resemble the polylines from parallel coordinates. Thus, glyphs can serve as an independent visualization component or they can support parallel coordinates.

While it can be assumed that polyline glyphs and star glyphs can be used interchangeably, since these two visualization techniques can encode the same data "payload," there is an essential difference in the way such encoded data are shown in the two glyph types. Since parallel coordinates are aligned in the polyline glyphs, users may find it easier to get datapoint values than with star glyphs, in which polar coordinates are used. Goldberg and Helfman's (2011) eye-tracking study revealed that visual scanning can be done more quickly along vertical and horizontal axes than circular scanning along rings, and the latter method is error-prone and not reliable. However, since star glyphs are more centered and compacted than polyline glyphs, they might perform better in tasks involving either similar or distinctive glyphs. However, such statements must be empirically tested. This raises the question as to whether star or polyline glyphs can be used for the same data, and if so, which glyph type performs better?

An example of a study in which line glyphs (similar to polyline glyphs) and star glyphs are compared has been published by Fuchs, Fischer, Mansmann, Bertini, and Isenberg (2013). Their study reveals that line glyphs are a good choice for tasks in which peak and trend detection is to be done when examining time-series data. By contrast, radial encoding of time in star glyphs works better if one has to find a particular temporal location. These findings contradict to some extent Goldberg and Helfman (2011) claim that linear graphs can better support the dimension-finding phase, since their linearly aligned dimensions support searches better than radial graphs. Lee, Reilly, and Butavicius (2003) compare four visualization techniques, including Chernoff faces and star glyphs, in terms of their usefulness in user tasks. Their study reveals that both types of glyph visualizations lead to slow, inaccurate answers being given, with a low degree confidence. In a more recent study, Chung et al. (2015) conclude that various interactive functions, such as glyph sorting, that support user exploration in glyph visualization can significantly enhance user performance.

Eye-tracking for evaluation of glyphs

Eye-tracking plays a particular role in empirical research on information visualization. According to Goldberg and Helfman (2011), to date, this technique has been underutilized as a method for understanding how individuals make use of information graphics. Although there has been a rapid increase in eye-tracking studies in geovisualization, studies of glyph-based visualization have been sparse and seldom. For example, Ho, Yey, Lai, Lin, and Cherng (2015) use the method to examine various 2D visualizations of flow, including a glyph-based technique, and Golebiowska, Opach, and Rød (2017) use eye-tracking to investigate a CMV interface consisting of a choropleth map, a parallel coordinate plot, and a table with polyline glyphs.

Eye-tracking is not necessary to examine the performance of glyph-based displays. Nevertheless, it enables a better sense of the differences between the visual behavior of different users – what and how long they look at. Similar empirical data can be obtained with other methods, such as mouse-tracking or recording task execution time. However, without insight into eye-movement data, it is unfeasible to analyze, for example, how often participants use a key, how many times they look at particular map symbols (revisit them), or which glyphs attract their attention most. Therefore, to gain a comprehensive insight into how polyline and star glyphs work, and then consider their advantages and weaknesses, we conduct an empirical study with eye-tracking as our main empirical technique.

Empirical study

Objectives

Since there might be differences in user performance for polyline and star glyphs shown either in a grid plot or on a map display, we examine four layout modes (M1–4). These are presented in Figure 1. We address three research questions:

(1) RQ1: Are there specific user tasks for which one of the two investigated glyph types outperforms the other?
(2) RQ2: Are there specific user tasks in which glyphs shown by means of one of the two investigated display types (grid plot or map) work better than those arranged in the other display type?
(3) RQ3: Are there certain user skills that influence task-solving strategies with star or polyline glyphs?

Participants

Total of 26 individuals (15 males and 11 females, average age 23 years) attend the study voluntarily. They are not paid any compensation for the attendance. All of them are either bachelor or master's students taking the geoinformatics course at Palacký University. Their skills and knowledge are considered representative of target users. Students from the first year of the bachelor's study program are excluded because they have not had any training in GIScience, and therefore their performance is likely to be worse than those who have had this training.

Study material

As study material, we design a single-page web application where regular web browser can be used to run the tool. We use costless JavaScript APIs such as D3.js and jQuery Sparklines to develop the tool. Labels and comments in the tool's interface are in Czech, with the exception of the key (legend) which is in English. The interface features four layout modes (see Figure 1) in which glyphs are either polyline glyphs (M1-PolyGrid, M4-PolyMap) or star glyphs (M2-StarMap, M3-StarGrid), and in which glyphs are either regularly distributed in a grid (M1-PolyGrid, M3-StarGrid) or geographically distributed on a map (M2-StarMap, M4-PolyMap). Additionally, for the purpose of the empirical study, the tool has an opening dialog box in which the four layout modes are grouped as follows:

- Variant 1, in which a grid plot with polyline glyphs (M1-PolyGrid) is followed by a map with star glyphs (M2-StarMap);
- Variant 2, which displays the supplementary modes [i.e. a grid plot with star glyphs (M3-StarGrid) followed by a map with polyline glyphs (M4-PolyMap)].

In all modes, the main panel is accompanied by both a task panel and a key that explains how data are encoded in glyphs (see Figure 3).

For the tool's data content, we use 10 socioeconomic indicators (variables) describing 48 municipalities in the counties of Sør-Trøndelag and Nord-Trøndelag in central Norway. The participants are Czech or Slovak and such data content is unknown to them. Hence, prior knowledge of the visualized variables cannot influence the participants' answers.

Methods

We gain scientific evidence through individual user sessions in which we ask participants to

- use the tool to solve six user tasks during an eye-tracking session,
- fill in a personal questionnaire and a subjective questionnaire on glyphs,
- perform a psychological test of the cognitive style of users.

We combine the methods above to get a comprehensive insight into participant choices and behavior.

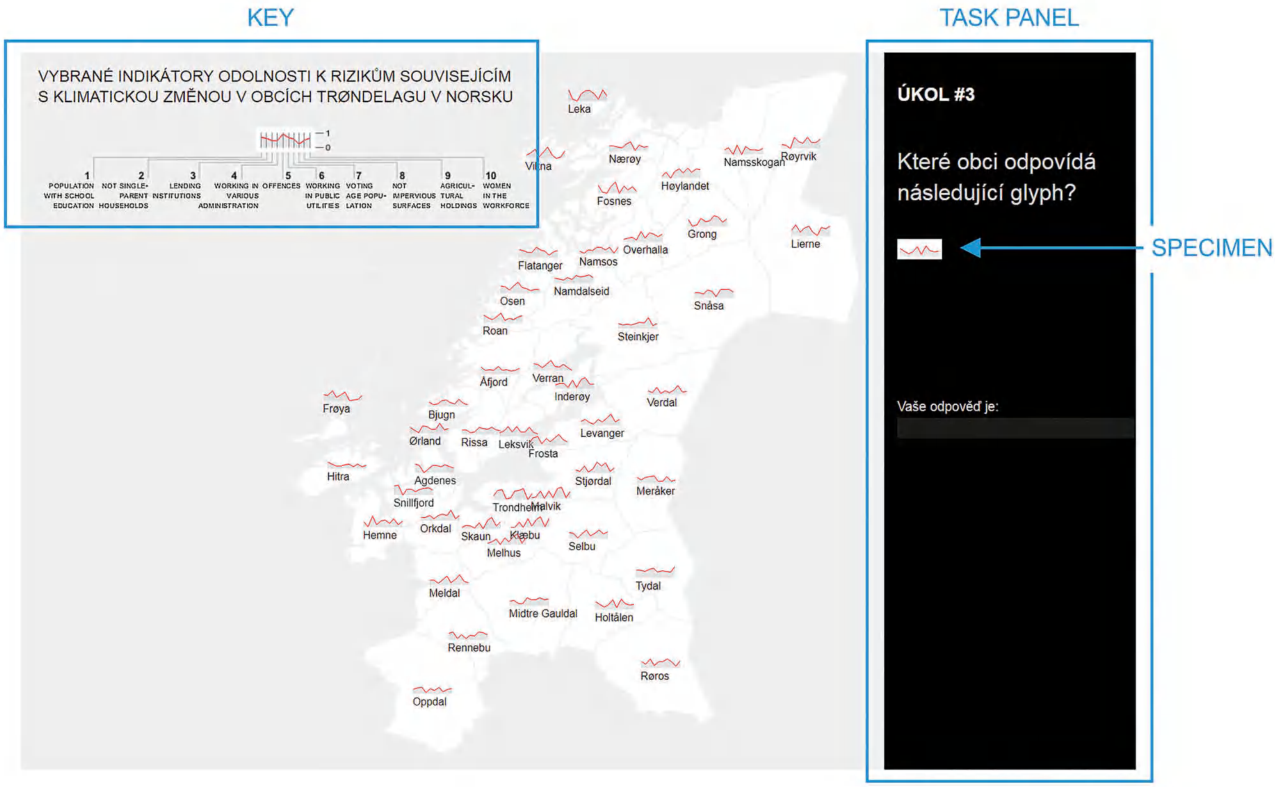


Figure 3. M#4-PolyMap (polyline glyphs on a map display) – one of the four layout modes used in the empirical study. It is presented here a bit enlarged with the task T#3-FindGlyph in which participants search for the same glyph as the one (specimen) shown in the task panel. The key explains the way how polyline glyphs encode multivariate data items.

User tasks

While participants execute the six user tasks, T#1–6 (Table 1), we record their eye-movements, oral comments, the screen, and their task answers. Regarding the tasks, these are designed to reveal how users interact with polyline and star glyphs. T#1-EstimVal is used to examine how users derive values from either polyline or star glyphs. T#2-IdenGlyph is complementary to T#1 as participants are expected to search for a glyph that encodes a certain variable score. In turn, in T#3-FindGlyph, users search for the same glyph as the one shown in the task panel (see Figure 3). We use this task to verify which of the four layout modes performs better regarding participants' visual searches. In T#4-SimilGlyphs, participants compare all glyphs between each other to find the two most similar glyphs. Regarding T#5-DistinctGlyphs, a common task in visual analytics is to find the most distinctive cases among graphic entities, and we therefore request users to do the same. Again, we investigate whether this task can be more effectively accomplished with polyline glyphs or with star glyphs. In the final task, T#6-CompArea, participants search for a compact area of three glyphs that are similar to each other.

Two questionnaires: personal and subjective about the glyph-based visualization

While the personal questionnaire will give information about participants' age and gender, the subjective questionnaire will inform about their subjective preferences regarding the two glyph types. The questionnaire consists of five questions (Q#1–5) that participants answer on a 7-point scale. The questions are as follows:

- Q#1 concerns overall feeling about the usability of glyphs.
- Q#2 is about the aesthetics of glyphs.
- In Q#3, participants are asked to specify a glyph type that is suitable for reading datapoint values; this question concerns T#1-EstimVal and T#2-IdenGlyph.
- Q#4 concerns a comparison of glyphs between themselves and refers to T#3-FindGlyph (search for a glyph), T#4-SimilGlyphs (find similar glyphs), and T#5-DistinctGlyphs (find distinctive glyphs).
- In Q#5, participants state which glyph type works best for getting an overview of all glyphs; it refers to T#6-CompArea, in which participants select a compact area of three similar glyphs.

Table 1. Six tasks used in the empirical study.

ID	Short name	Task question	Purpose of task
T#1-EstimVal	Glyph value estimation	Use the key shown in the upper left corner and estimate the datapoint values of the selected glyph on variables #5 and #10	Examine how users derive values from polyline and star glyphs
T#2-IdenGlyph	Search for a glyph featuring a concrete score on selected variable	Find a glyph with the score 1 on variable #5	Complementary to Task #1, examine how users search for a glyph that features a certain datapoint value
T#3-FindGlyph	Search for a glyph	Search for the same glyph as the one shown in the task panel	Examine how effectively users search for either a polyline or star glyph
T#4-SimilGlyphs	Point out the two most similar glyphs	Point out the two most similar glyphs	Examine how effectively users search for two similar glyphs (polyline glyphs or star glyphs)
T#5-DistinctGlyphs	Find two glyphs with the most distinctive cases	Find two glyphs with the most distinctive cases	Examine how effectively users search for distinctive glyphs (polyline glyphs or star glyphs)
T#6-CompArea	Find a compact area of three glyphs that are similar to each other	Indicate a compact area that consists of three glyphs of similar shape	Examine whether user performance regarding finding similarities is better for polyline or star glyphs

The scale used in the questionnaire is designed so that the values toward the left-hand end reflect the user's preference for polyline glyphs, whereas the values toward the right-hand end reflect their preference for star glyphs; the middle value means no preference (neutral choice).

Psychological test of the cognitive style of users

The last test – Navon's cognitive style test (Navon, 1977) – is one of the most frequently used tests for measurement of the global-analytic dimension of cognitive processing (Brand & Johnson, 2014). This test is designed to reveal whether an individual's preferred cognitive style is holistic or analytic – the distinction that is one of the most common among people's cognitive styles (Dewey, 2004). According to Dewey, in the analytic thinking, individuals comprehend a system, first, by recognizing its particular parts, and second, by understanding how they constitute a larger scale pattern. Whereas in the holistic thinking, of primary importance to individuals is to first recognize large-scale patterns of a system, not its particular elements. The fact that this cognitive style could affect work with a map, at least in the subprocesses, is well documented (Kubíček et al., 2016). We use a compound letter test, an adaptation of Navon's hierarchical figures test (Navon, 1977) developed as part of the GEOKRIMA project (Šašinka, 2013). In this test, big numbers composed of small numbers are displayed, and participants are requested to recognize either small or big numbers. We use the Hypothesis software (Morong & Šašinka, 2014) to perform the test. Its output will help to determine the cognitive style of participants and will be used to compare the affiliation to these cognitive styles with the strategy of solving the tasks.

Equipment

We use the eye-tracker SMI RED 250. The eye-tracker is arranged in the Eye-tracking Laboratory of the Department of Geoinformatics at Palacký University, Olomouc, in the Czech Republic. The stimulus is displayed on a 24-in screen with a resolution of 1920 × 1200 pixels. Eye positions are recorded at a frequency of 250 Hz. The eye-tracker is supplemented with a web camera that records participants during the sessions. We do this because audio and video recording can help to reveal the possible cause of missing data, participants' reactions to the stimuli, and their comments on the tasks.

Procedure

Individual user sessions are arranged as two-phase user testing (Figure 4) with a minimum of 3 days between each phase. This is done to avoid a learning effect, whereby, during the "later" stage of testing, participants may use their knowledge acquired in the "earlier" stage. Without two-phase user testing, the results might be influenced by already gathered experience. Moreover, a variant assigned to participants in each phase will shift: The first participant starts with variant #1, the next participant uses variant #2, and so on. In the second phase, participants use the complementary variant. For example, if a participant uses variant #1 in the first phase, they use variant #2 in the second phase.

The test sessions are performed using SMI Experiment Center. We collect participants' answers, eye-movement data complemented by audio and video recording of participants, screen recording, and task completion time. Each session is organized as follows (see Figure 4). In its first phase, after the participants have been welcomed, an eye-tracker is calibrated for each of them. The maximal allowed deviation is set as 1° of the visual angle. Next, each

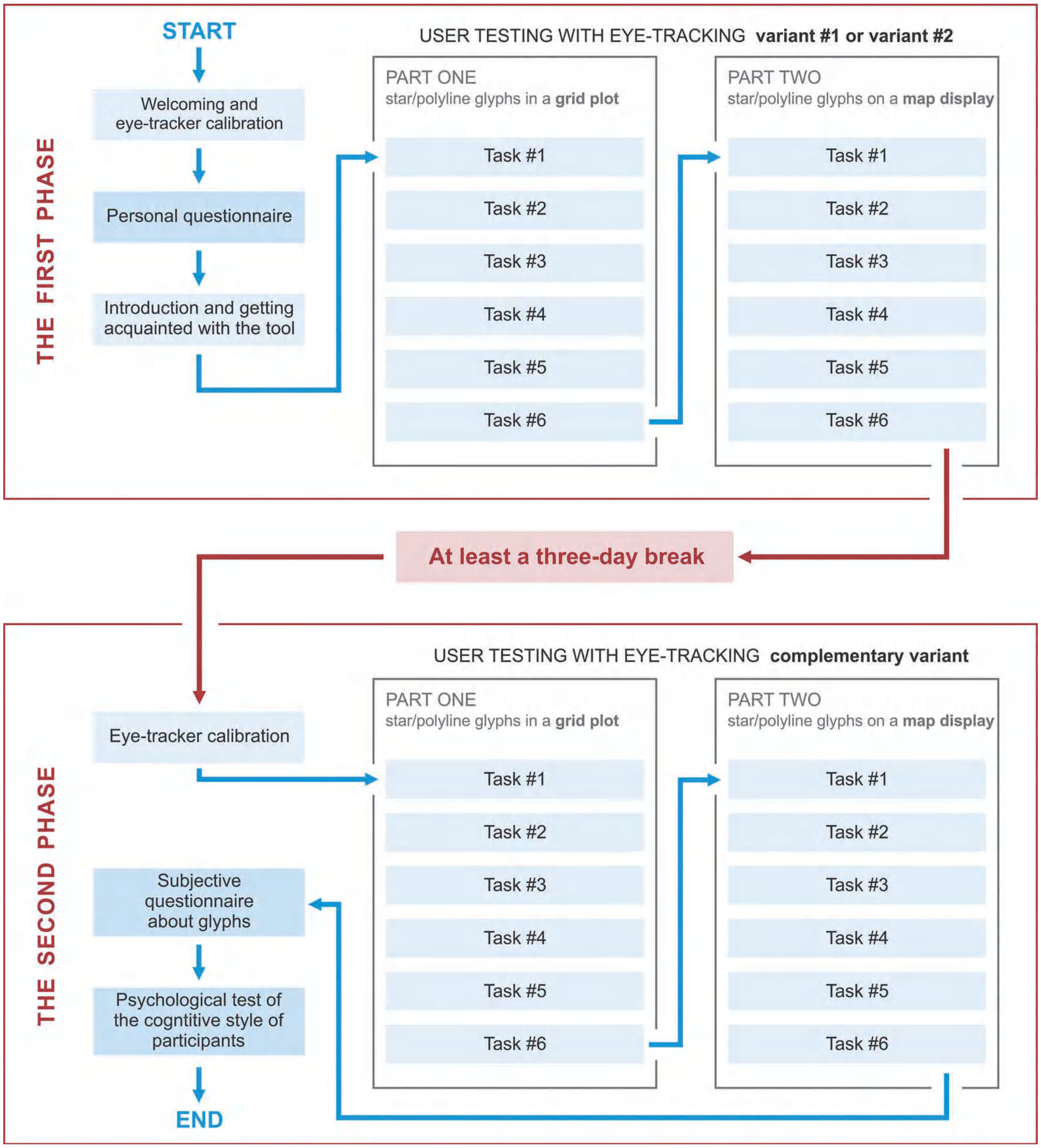


Figure 4. The design of the individual user session.

participant fills in the personal questionnaire. This is followed by a short instruction about glyph-based visualization and each participant is given 2 min to play around with the display. The testing consists of two parts. In both parts, participants solve six user tasks (see Table 1). However, in the first part, 48 glyphs constitute a grid plot, whereas in the second part, they constitute a map display with the coordinates of the municipalities they represent. In the second phase (a few days later), participants perform the same six tasks again, but with the complementary variant. Hence, within the entire session, each participant solves the six user tasks for each of the four layout modes. After the user testing in the second phase, participants fill in the subjective questionnaire and perform the psychological test on their cognitive style.

Data obtained

Eye-movement data are recorded as "screen recording" type of stimulus. Result of the recording is a separate video file obtained for each participant. Therefore, to analyze all recordings together, we use the Custom Trial Selector from the SMI BeGaze software to combine all videos by task. The custom trial is designed for each task as a screenshot, and then the corresponding part of each recording is assigned to it. In the analysis of eye-movement

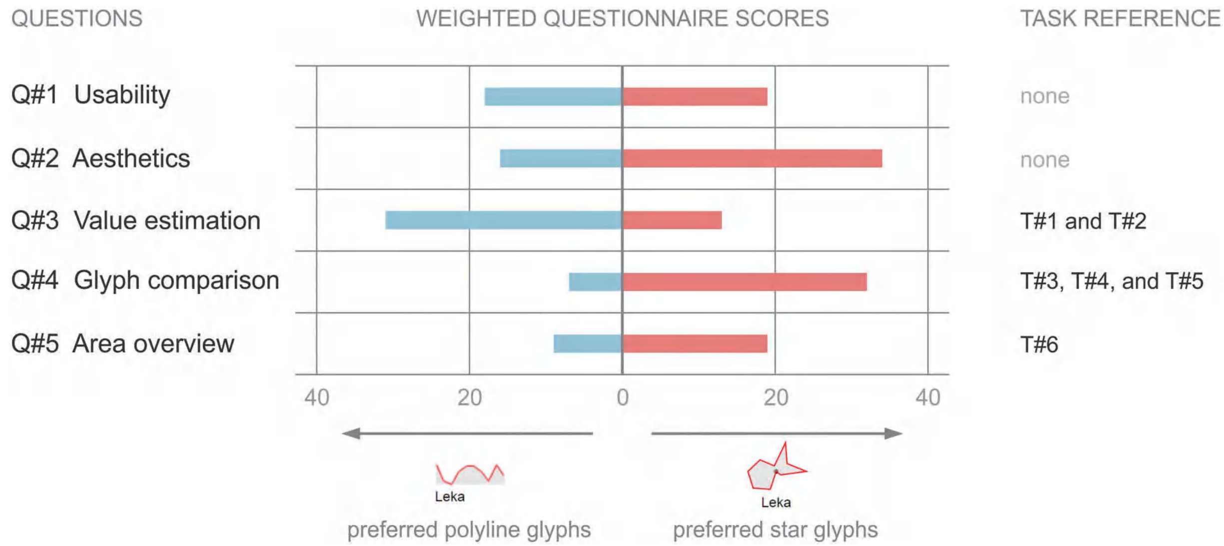


Figure 5. The outcomes of the subjective questionnaire about glyphs.

data, we use the I-DT algorithm for fixation detection (Salvucci & Goldberg, 2000). This algorithm is mostly used for low-frequency data (up to 250 Hz) and takes into account the close spatial proximity of the eye position points in the eye-movement trace. Threshold values in BeGaze are set to 80 ms for "duration threshold" and 50 pixels for "dispersion threshold," as these values are suggested as optimal (Popelka, 2014).

To analyze answer accuracy in T#4–6, scores on the similarity measure for all possible pairs of glyphs need to be calculated. To do this, we use the SimUrb software (eyetracking.upol.cz/simurb) – a derivative of the ScanGraph software (Dolezalova & Popelka, 2016). It calculates the similarity measure for all pair of glyphs – the Euclidean distance in n-dimensional space. All statistical tests are executed in RStudio at 0.05 significance level.

Data analysis

The subjective questionnaire about glyphs

Since the questionnaire is presented at the end of the second phase (see Figure 4), all participants use all four layout modes (M#1–4) ahead of the questionnaire. In its analysis, weighted scores are used to amplify higher ratings: The values leading to both ends of the scale have increasing weights, from 1 to 3 (the middle value, 0, is subtracted from the analysis). In all but Q#3, star glyphs receive more points (Figure 5); it is not a surprise, because Q#3 concerns value estimation (required in T#1-EstimVal and T#2-IdenGlyph), to which polyline glyphs are supposedly better tailored. Although the answers are almost balanced in Q#1 about usability, from the aesthetics point of view (Q#2), participants prefer star glyphs. Finally, more points are given to star glyphs in Q#4 and Q#5, about glyph comparison and area overview, respectively.

Eye-movement analysis: trial duration metric by task and layout mode

We analyze trial duration by task and layout mode. This metric shows how long it takes to solve a task. A quick look at the boxplots in Figure 6 will reveal that in the comparison of the grid plot and the map display (consisting of the same glyph types), all tasks are solved quicker if the map display is used. Statistically significant results of the Kruskal–Wallis test with the post hoc Nemenyi test are found for a number of configurations (marked with asterisks in Figure 6). However, statistically significant differences (p = 0.02) between polyline and star glyphs on the same display are found only for map display in T#3-FindGlyph: It is faster to find a star glyph.

In turn, the Wilcoxon rank-sum test reveals statistically significant differences between the grid plot and the map display (regardless of the glyph type) for T#1-EstimVal (W = 1871, $p < 0.001$), T#4-SimilGlyphs (W = 1766, p = 0.007), and T#6-CompArea (W = 2195, $p < 0.001$). In these three cases, the map display proves to be quicker. When it comes to the differences between polyline and star glyphs (regardless of the display type), the Wilcoxon test reveals statistically significant differences (star glyphs are quicker) for T#3-FindGlyph (W = 1927.5, $p < 0.001$) and T#4-SimilGlyphs (W = 1770, p = 0.007).

From the analysis above, it is apparent that trial duration metric is task dependent and in particular, tasks of higher values are observed for different

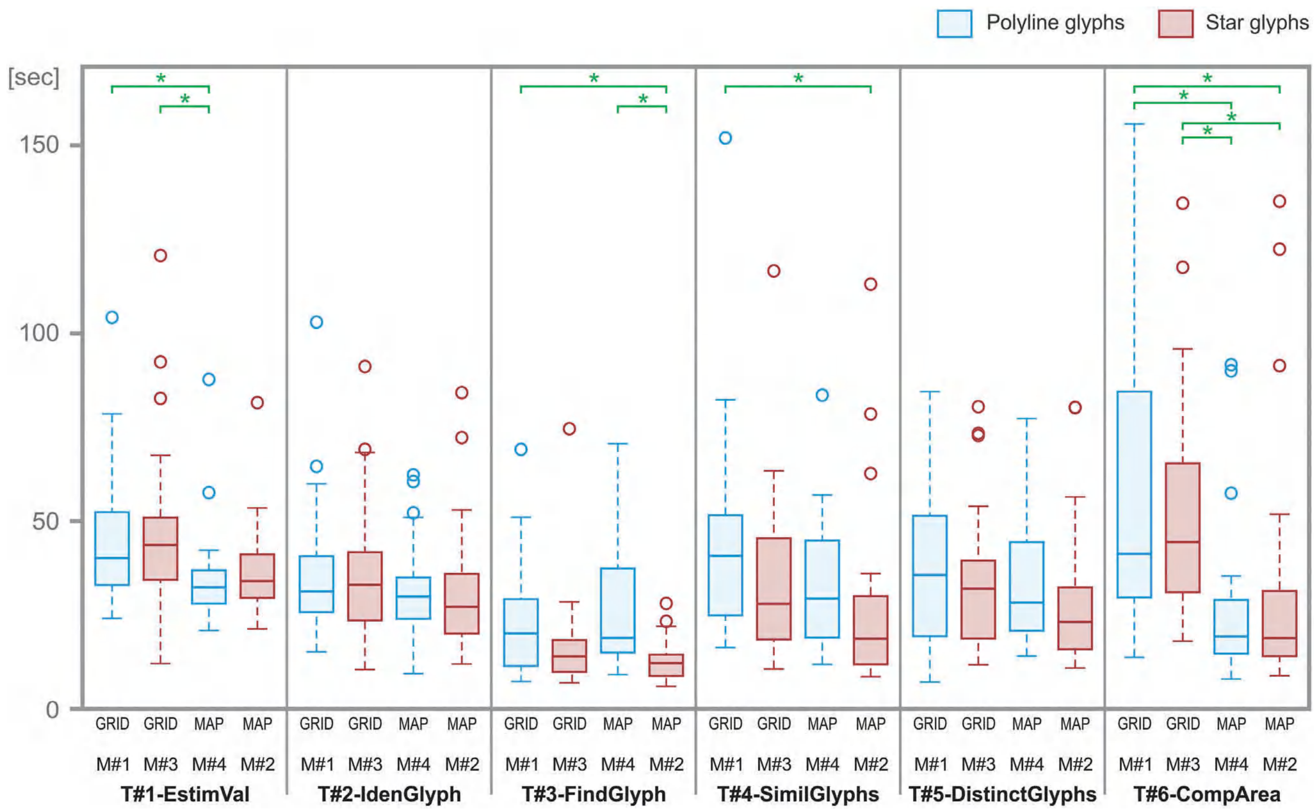


Figure 6. Trial duration metric by task and layout mode.

configurations, for either polyline or star glyphs. From this reason, user behavior during task execution is further analyzed separately for subsequent tasks.

Eye-movement analysis: fixation counts in glyphs by task and layout mode

If fixation counts in glyphs are compared in various tasks and layout modes (see Figure 7), more numerous fixations occur in the last three tasks in general, and in T#6-CompArea's grid plot in particular. These are caused by the intensive visual searches required for those tasks. The Kruskal–Wallis test with the post hoc Nemenyi test reveals statistically significant differences for a number of configurations (marked with asterisks in Figure 7). The significant differences are similar to those reported in the preceding section.

The glyphs that participants look at depend on their user tasks. For example, to read datapoint values encoded in a given glyph in T#1-EstimVal, participants can either use the key ("absolute interpretation") or compare the glyph with other glyphs ("relative interpretation"). In T#1-EstimVal, fixations are more numerous in the selected glyph and its neighborhood (the latter may be caused by the eye-tracker inaccuracy), because participants look at the nearest surroundings of the selected glyph and do not look at the bottom part of the grid plot (Figure 8). This may mean that participants do not tend to compare the glyph's shape with other glyphs. In more distant glyphs, only a few fixations occur and they probably accompany the gaze movements from the selected glyph to the key.

In the remaining tasks, fixations are scattered around the whole stimuli, as participants search for glyphs. In T#2-IdenGlyph, such behavior can lead to glyph decoding executed through comparison with other glyphs ("relative interpretation"). Furthermore, in T#2-IdenGlyph's grid plot, participants look mostly at the first two rows and finish solving the task just after localizing the first glyph fulfilling the requirement. In T#3–6, the displays are also fully covered by fixations; hence, all glyphs attract attention. Nevertheless, in T4-SimilGlyphs, most fixations are recorded for the glyphs selected by participants, mostly for Klæbu and Malvik, particularly in the map display.

Answer accuracy and visual behavior in particular tasks

T#1: glyph value estimation

Answer accuracy. In T#1-EstimVal, participants read datapoint values of two variables (#5 and #10) in a marked glyph. To avoid a learning effect, there

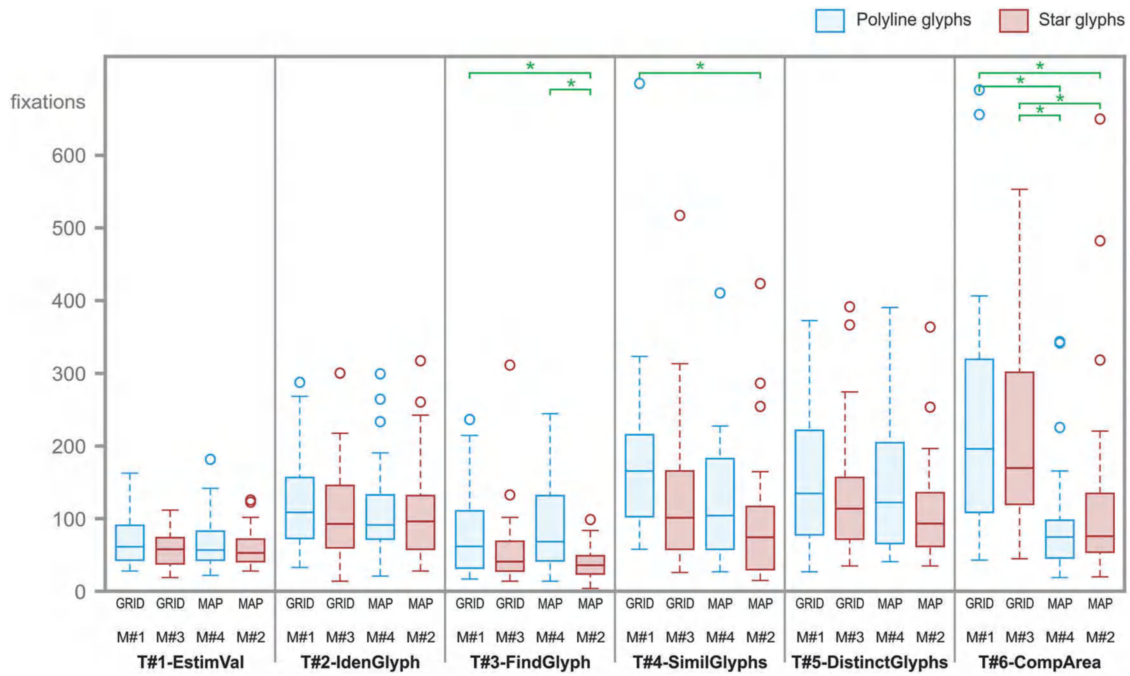


Figure 7. Fixation count in glyphs by task and layout mode.

are two different glyphs marked in two subsequent layout modes used in the same test session: Leksvik in the grid plot and Selbu in the map display. To examine estimation accuracy, an average difference between estimated values and correct scores is calculated. The same as claimed in Q#3 in the subjective questionnaire, better answer accuracy (lower average differences) is observed for polyline glyphs. However, the Kruskal–Wallis test with the post hoc Nemenyi test reveals no statistically significant differences in estimation accuracy between any pair of the four layout modes.

AOI analysis. Users need a key (legend) to solve T#1-EstimVal. We therefore examine how intensively the key is used in the four layout modes. The dwell time measure calculated for the areas of interest (AOIs) marked around various parts of the stimuli shows what portion (percentage) of the trial duration participants spend in particular AOIs. The key is used longer for star glyphs (medians of 9.5% and 12% for the grid plot in M#3-StarGrid and the map display in M#2-StarMap, respectively) than for polyline glyphs (8.5% for M#1-PolyGrid and 8.6% for M#4-PolyMap). However, the Kruskal–Wallis test with the post hoc Nemenyi test reveals no statistically significant difference. Slightly higher attention to the key when using star glyphs occurs in the revisits measure, which informs how many times participants revisit the AOI with the key during the trial duration. The numbers of revisits for star and polyline glyphs are, respectively, 4 and 3 (medians) in the grid plot, and 4 and 2 in the map display. Statistically significant difference is found between polyline and star glyphs (regardless the display type) using the Wilcoxon test ($p = 0.02$).

We also investigate the number of transitions between the AOI with the key and the AOI with either the grid plot or the map display. It turns out that participants look at the key more frequently if star glyphs are used: In the grid plot, participants move their visual attention from the main display to the key (total switches between the two AOIs) 146 times for star glyphs (M#3-StarGrid), and 142 times for polyline glyphs (M#1-PolyGrid). By contrast, in the map display, we again observe 146 such transitions for star glyphs (M#2-StarMap), and only 127 for polyline glyphs.

T#2: search for a glyph featuring a concrete score on selected variable

Answer accuracy. In T#2-IdenGlyph, participants search for a glyph that leads to the value of 1 on variable #5. The difference between the chosen glyph's score on variable #5 and the value of 1 is calculated for each participant. The participants' answers for star glyphs are slightly better [this contradicts the responses to the subjective questionnaire (Q#3) that polyline glyphs are more suitable than star glyphs for value estimation]; however, the Kruskal–Wallis test reveals no statistically significant differences.

AOI analysis. Although participants also need the key to solve T#2-IdenGlyph, they look at it a shorter time than in T#1-EstimVal. They spend 9.05% of the trial

Figure 8. Fixations by task and layout mode.

duration looking at the key in T#1, whereas in T#2, the corresponding percentage is only 1.55% (3623 and 496 ms, respectively). This may be due to a learning effect. In T#1, participants have already learned how to decode variables, and therefore they have less need for the key in T#2.

We also analyze the visual behavior of participants with poor answer accuracy. In most cases of poor answers, participants use the map display with polyline glyphs. Furthermore, poor answers appear especially common for the glyphs for Stjørdal and Trondheim that score high on variable #6. This may be the reason why the participants mix it up with variable #5.

T#3: search for a glyph

Answer accuracy. In T3-FindGlyph, participants search for the same glyph as the one – the specimen – shown in the task panel (see Figure 3). Although for both glyph types, the glyph for Holtålen municipality is used, the two glyph versions differ strongly with regard to shape.

The answer accuracy is high and all but three answers are correct. Three incorrect answers occur in the grid with polyline glyphs (M#1-PolyGrid). This finding may confirm the results of the questionnaires, in which (Q#3) star glyphs are claimed as more suitable for glyph comparison.

AOI analysis. Transitions between the display and the specimen reveal that participants move their attention to the specimen less frequently if a star glyph is found: 131 times (total switches between the two AOIs) when using the grid plot and 137 times for the map. For polyline glyphs, the corresponding number of times is 226 for the grid and 253 for the map. This may suggest that it is more difficult to remember the shape of a polyline glyph than of a star glyph since, in the latter case, participants do not need to bring back the shape of the specimen so often. This claim can be further backed up by the analysis of revisits of the specimen. The Wilcoxon rank-sum test reveals statistically significant differences (W = 2000, $p < 0.001$) between polyline glyphs and star glyphs for both the grid and the map (Figure 9). This means that the participants look at the specimen less frequently when searching for a star glyph. We thus interpret this that it is easier to remember a star glyph and that its specimen does not need to be checked so often.

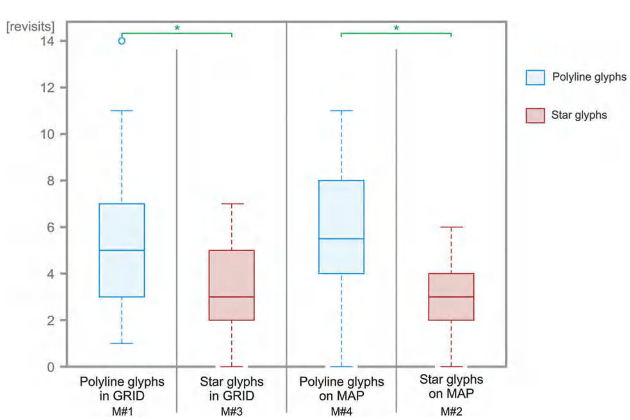


Figure 9. Revisits of the area of interest (AOI) marked around the specimen in the task panel.

We examine how quickly participants make the final decision after localizing the glyph in the display, and thus how certain they are. We also want to know what they do after localizing the glyph, how many times they revisit the specimen, and whether they check the remaining glyphs to be more confident. The analysis reveals that there is no difference between glyph and display types. Participants exhibit similar levels of confidence in all layout modes.

T#4: point out two most similar glyphs

Answer accuracy. In T#4-SimilGlyphs, participants point out the two most similar glyphs. To examine answer accuracy, the scores on the similarity measure need to be first calculated for all possible pairs of glyphs. To do this, we use the SimUrb tool. Then, we check the scores received for the pairs chosen by the participants. The highest similarity (0.89 in the range 0–1) features the pair Klæbu–Malvik, and this pair is selected 50 times in all 104 trials. The Kruskal–Wallis test reveals statistically significant differences (p = 0.001) between the accuracy of the answers given by those who use polyline and star glyphs in the map display. The test also returns statistically significant differences (p = 0.001) between polyline glyphs in the grid (M#1-PolyGrid) and star glyphs on the map (M#2-StarMap) (Figure 10(a)). In these cases, star glyphs perform better: Star glyphs selected by participants as pairs are more similar than pairs of polyline glyphs. This analysis correlates with the subjective questionnaire, in which star glyphs are claimed more suitable for comparisons of glyphs and hence for finding similar glyphs. Moreover, as expected from the trial duration analysis by layout mode (described earlier in this paper), the map gives better results than the grid plot for T#4-SimilGlyphs. For these two displays (regardless of the glyph type), the Kruskal–Wallis test reveals a statistically significant difference (p = 0.029).

Scanpath length analysis. We use scanpath length (the length of the gaze trajectory) in the data analysis because, as Holmqvist et al. (2011) claim, it measures the efforts needed for visual search, and therefore it might reflect task complexity. In the comparison of the grid plot and the map display, the Kruskal–Wallis test reveals statistically significant differences ($p < 0.001$): Shorter scanpaths are recorded for the map display (Figure 10(b)). In turn, in the comparison of polyline and star glyphs, although shorter scanpaths are

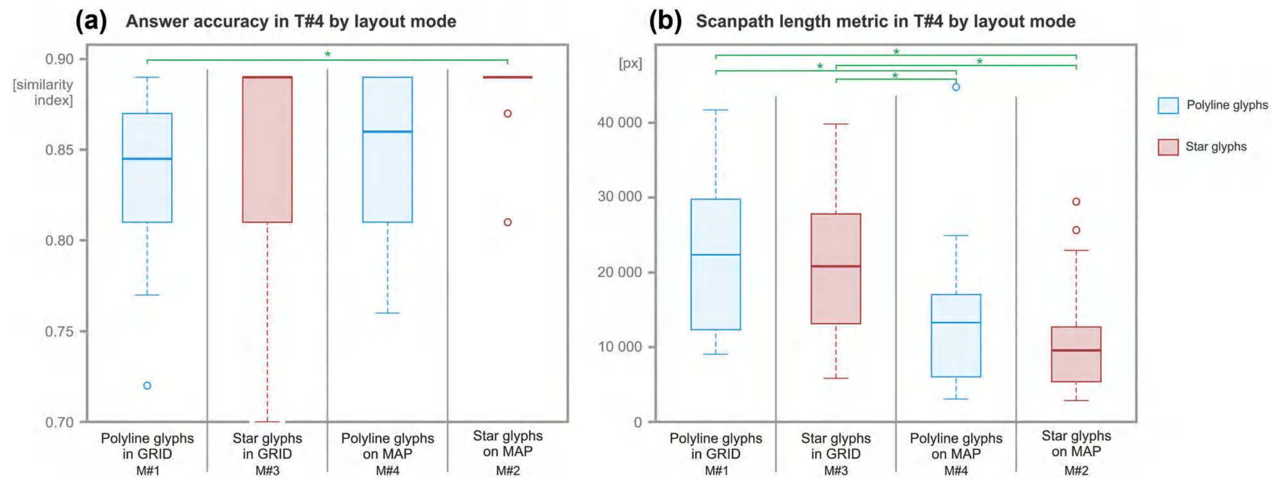


Figure 10. The outcomes of T#4-SimilGlyphs (point out two most similar glyphs) by layout mode: (a) scores on the similarity measure calculated for the pair of glyphs selected by participants and (b) the scanpath length metric.

observed for the latter, the difference is not statistically significant (W = 1569, p = 0.158).

T#5: find two glyphs with the most distinctive cases

Answer accuracy. T#5-DistinctGlyphs is similar to T#4-SimilGlyphs, except that participants in T#5-DistinctGlyphs point out two the most distinctive glyphs from all glyphs presented in the display. As in T#4-SimilGlyphs, the SimUrb tool is used in T#5-DistinctGlyphs for data analysis. However, we calculate an average similarity for each of two selected glyphs and all remaining glyphs. The lowest average similarity (i.e. most distinctive glyphs) is calculated for Leka (0.55) and Trondheim (0.59). These two glyphs are selected 30 and 8 times, respectively, in all 104 trials. However, a combination of both glyph types is selected only once, on the map display with polyline glyphs.

The Wilcoxon test reveals a statistically significant difference (W = 993.5, p = 0.018) between the answer accuracy of those who use star glyphs and those who use polyline glyphs (regardless of the display type). "Better" answers (lower average similarity of two selected glyphs) are given by those who use polyline glyphs. We also test the differences between polyline and star glyphs separately for the grid and the map. The Wilcoxon test reveals a statistically significant difference (W = 398, p = 0.028) between polyline and star glyphs only for the grid. In this analysis, too, lower average similarity is observed for polyline glyphs. The latter observation along with the previous one contradicts the observations from T#4-SimilGlyphs in which star glyphs are found better for finding similar (and thus different) glyphs.

Scanpath length analysis. In the scanpath length analysis, the Kruskal–Wallis test shows a clear tendency toward significance (p = 0.055) only for the difference between polyline glyphs in the grid and star glyphs on the map. This may indicate that although in T#5-DistinctGlyphs, polyline glyphs facilitate better answer accuracy, T#5-DistinctGlyphs can be more "easily" solved if star glyphs are used, particularly on the map display. Further research is however needed to better elaborate this.

T#6: find a compact area of three glyphs that are similar to each other

Answer accuracy. T#6-CompArea is similar to T#4-SimilGlyphs. However, participants search for a compact area of the three most similar glyphs. Although the purpose of this task is clear for the map display, since glyph positions are dependent upon the municipalities they represent, the purpose of this task may be questioned if the grid plot is to be used where glyph positions are random. We use T#6-CompArea for the grid plot, as well to ensure consistency in the testing.

Although it is stated that selected glyphs must be adjacent, some participants select glyphs (5 times in 104 trials) that are far away. As we suppose, they do this because they do not read the task question sufficiently carefully. We calculate average similarity scores for all selected glyph triplets and use them as the answer accuracy measure. The highest similarity score for three adjacent glyphs in the map display is found for the triplet Klæbu–Malvik–Melhus (an average of 0.87). This triplet is selected 19 times of 104 trials. The Kruskal–Wallis test reveals statistically significant differences [shown marked with asterisks in Figure 11 (a)] between four layout modes. Moreover, the Wilcoxon test reveals a statistically significant difference ($W = 371.5$, $p < 0.001$) between two display types: the map display gives better results. Regarding the difference between polyline and star glyphs, as visible in Figure 11(a), the average similarity of selected star glyphs seems to be higher than the similarity of polyline glyphs. However, the Wilcoxon test reveals no statistically significant difference ($W = 1136$, $p = 0.157$) between two glyph types.

Scanpath length analysis. For the scanpath length metric, the Kruskal–Wallis test gives statistically significant differences for the same combinations as in T#4-SimilGlyphs. In this case, too, there are no statistically significant differences between two glyph types in the same display ($p = 0.976$ for the grid plot, $p = 0.795$ for the map display). Longer scanpaths are observed for the grid plot than for the map display and slightly longer scanpaths are observed for star glyphs than for polyline glyphs (Figure 11(b)).

The test of the participants' cognitive style: analytic versus holistic users

In the cognitive style test, 32 images are displayed that show big numbers (hereafter referred to as BNs) composed of small numbers (SNs). Participants are asked about either BNs or SNs, in 16 images for each. From the Hypothesis software used to run the test, we obtain the average times for SNs and BNs for each participant. Then, we calculate averages for SNs and BNs for the whole sample ($n = 26$). The times obtained for BNs are 13% shorter than those obtained for SNs. Therefore, to eliminate the global precedence effect (Navon, 1977), we equalize both data samples by multiplying the SN times by coefficient 0.87 (as it reduces them by 13%). Finally, we calculate the quotient of the BN time and the SN time; the value of 1 represents the most balanced (analytic vs. holistic) participants. Participants who feature the smallest deviation from 1 (who deviate to less than 10%) are labeled neutral (Table 2). Remaining participants are labeled as either analytics (six participants with better performance for SNs) or holistics (seven participants with better performance for BNs).

We examine which participants – analytic or holistic – perform better regarding answer accuracy, trial duration, and fixation frequency. For answer accuracy, holistics perform better only in T#4-SimilGlyphs. In T#2-IdenGlyph and T#3-FindGlyph, the results are similar for both groups, whereas in T#1-EstimVal, 5, and 6, analytics have better results. However, the Wilcoxon test does not reveal any statistically significant differences for any of these results. With regard to trial duration, holistics are faster in almost all tasks (Figure 12(a)), as expected. The only exception is T1-EstimVal, in which participants do not need to search for any glyph but only estimate two datapoint values in a selected glyph. Again, no differences between analytic and holistic participants are statistically significant.

The fixation frequency measure shows the number of fixations per second. In T#1-EstimVal and T#2-IdenGlyph, fixation frequency is higher for analytics, but in the remaining tasks, holistics have higher fixation frequency (Figure 12(b)). Furthermore, in T#5-DistinctGlyphs and T#6-CompArea, the Wilcoxon rank-sum test reveals

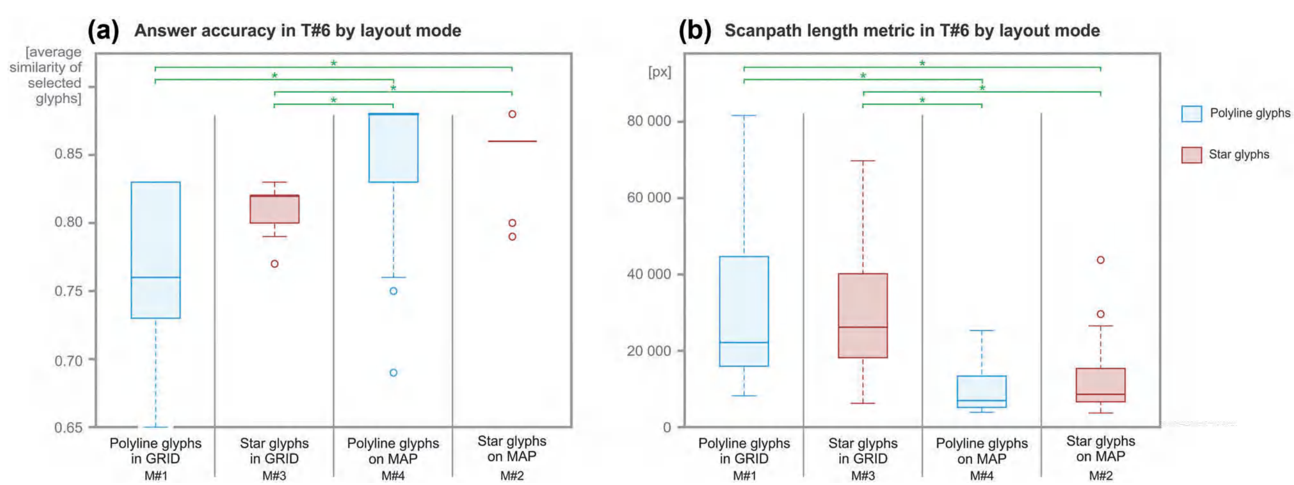


Figure 11. The outcomes of T#6-CompArea (find a compact area of three glyphs that are similar to each other) by layout mode: (a) average scores on the similarity measures calculated for triplets of glyphs selected by participants and (b) the scanpath length metric.

Table 2. The distinction between analytic and holistic participants based on the outcomes of the test of their cognitive style.

ID	Ratio[a]	Difference	Cognitive style
P04	1.470	0.470	Analytic users
P10	1.470	0.470	
P09	1.234	0.234	
P17	1.161	0.161	
P11	1.158	0.158	
P14	1.107	0.107	
P24	1.079	0.079	Neutral users
P19	1.056	0.056	
P15	1.035	0.035	
P07	1.033	0.033	
P12	1.023	0.023	
P08	0.995	−0.005	
P03	0.947	−0.053	
P06	0.939	−0.061	
P22	0.928	−0.072	
P23	0.916	−0.084	
P05	0.914	−0.086	
P18	0.911	−0.089	
P02	0.902	−0.098	
P16	0.870	−0.130	Holistic users
P13	0.861	−0.139	
P20	0.841	−0.159	
P21	0.838	−0.162	
P01	0.837	−0.163	
P26	0.818	−0.182	
P25	0.769	−0.231	

[a]The quotient of the BL time and the SL time.

statistically significant differences: $W = 201$, $p = 0.013$ and $W = 221.5$, $p = 0.035$, respectively.

Results and discussion

In this section, we relate the findings of the empirical part to the research questions (RQ#1–3). We then summarize our assessment of which glyph and display type is most effective (answer accuracy) and most efficient (trial duration and other eye-tracking metrics). Moreover, we present our conclusions about whether users find star glyphs or polyline glyphs more satisfying. We refer to these three aspects – effectiveness (task completion by users), efficiency (task in time), and satisfaction (responded by users in terms of experience) – because according to ISO (9241-11), they constitute usability in a given context of use (users, tasks, equipment, and environments).

RQ#1: polyline versus star glyphs

In general, star glyphs perform better than polyline glyphs. In most cases, their use leads to better answer accuracy and shorter task accomplishment time (Table 3). "Strong" results are obtained especially for T#3–5, in which participants need to compare glyphs in a display. In these three tasks, star glyphs receive better scores in the subjective questionnaire (Q#3). They also feature fewer revisits in T#3-FindGlyph (i.e. better efficiency) in which the participant needs to compare glyphs with the specimen. The analysis reveals no statistically significant differences for any of the results from T#2-IdenGlyph and T#6-CompArea. Nevertheless, polyline glyphs receive better ratings in the subjective questionnaire in T#2-IdenGlyph, whereas star glyphs receive better ratings in T6-CompArea.

Polyline glyphs receive better scores in T#1-EstimVal, in which participants read the datapoint values of two variables. In this task, better answer accuracy and shorter task accomplishment time are observed for those who use polyline glyphs; however, the analysis reveals no statistically significant differences. In T#1-EstimVal, the analysis reveals significant result for the revisits to the key. Participants do not need to check the key as frequently in the case of polyline glyphs compared with star glyphs. We suppose that better performance of polyline glyphs might be caused by the linear order of variables encoded in a polyline glyph, thus resulting in their easier interpretation. In star glyphs, users must find variable positions, and this needs more effort (i.e. more time and more numerous revisits). Lastly, in the subjective questionnaire (Q#1), participants' preferences for estimating values lean strongly toward polyline glyphs.

Polyline glyphs seem to perform better if they are used to read datapoint values encoded in glyphs. However, if comparisons are made among glyphs or if a compact area of glyphs is found with glyphs that are similar to each other, then star glyphs work better. These findings confirm, to some extent, those reported

Table 3. Polyline versus star glyphs.

Task	Better answer accuracy	Shorter task accomplishment time	Eye-movement data analysis: revisits of specific AOIs (R) or SL		Subjective questionnaire
T#1-EstimVal	Polyline glyphs	Polyline glyphs	R	Polyline glyphs[a]	Polyline glyphs
T#2-IdenGlyph	Star glyphs	Star glyphs	SL	Polyline glyphs	Polyline glyphs
T#3-FindGlyph	Star glyphs (no errors)[a]	Star glyphs[a]	R	Star glyphs[a]	Star glyphs
T#4-SimilGlyphs	Star glyphs[a]	IStar glyphs[a]	SL	Star glyphs	Star glyphs
T#5-DistinctGlyphs	Polyline glyphs[a]	Star glyphs	SL	Star glyphs	Star glyphs
T#6-CompArea	Star glyphs	Polyline glyphs	SL	Polyline glyphs	Star glyphs

SL: Scanpath length. R: Revisits. [a]Statistically significant or strong results.

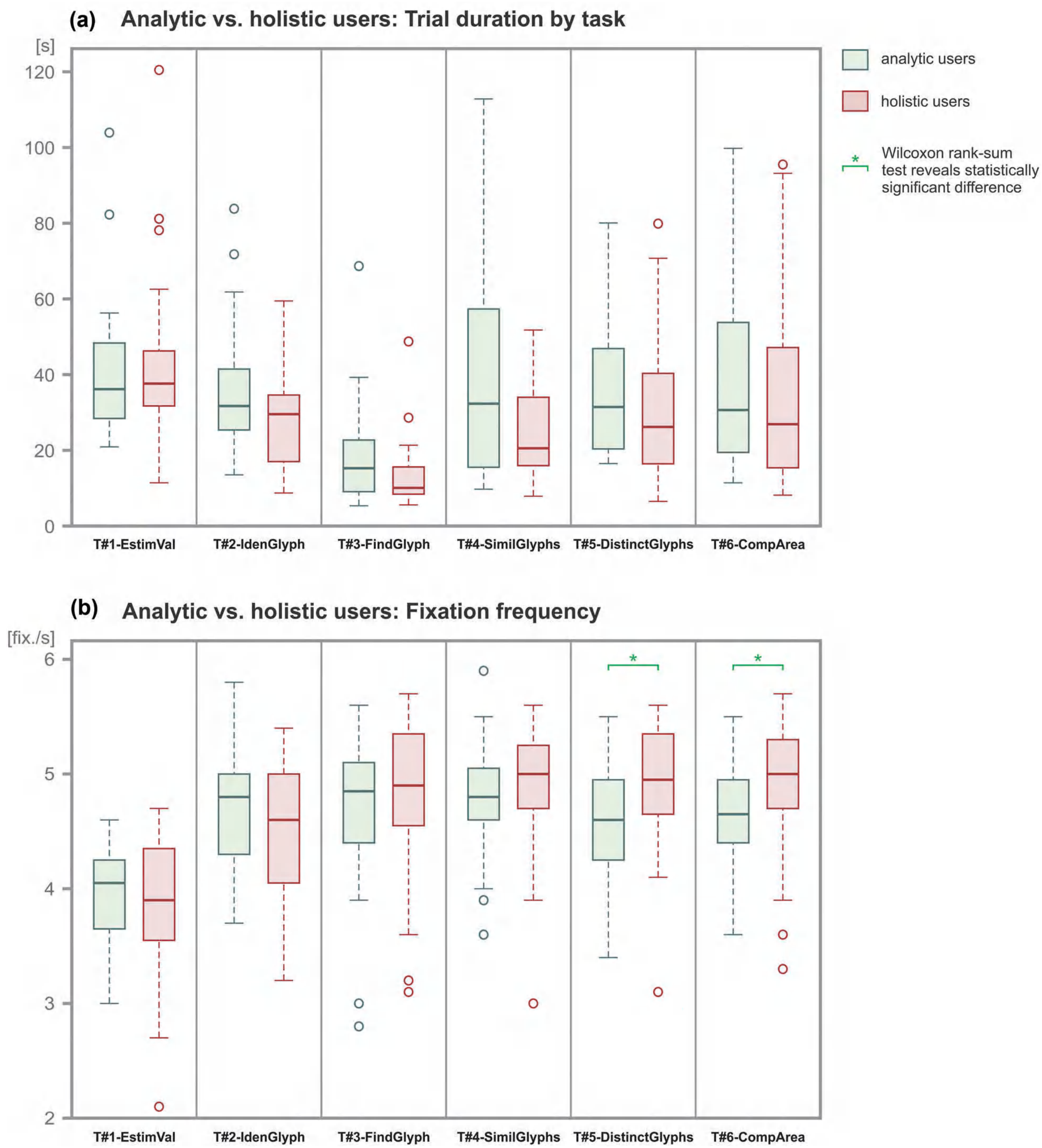


Figure 12. Analytic versus holistic participants by task: (a) the trial duration and (b) the fixation frequency measure.

by Goldberg and Helfman (2011). Indeed, linear graphs support the dimension-finding task better than do radial graphs.

RQ#2: grid plot versus map display

The outcomes are consistent when the grid plot is compared with the map display. Although the answer accuracy in T#2-IdenGlyph and the revisits measure in T#3-FindGlyph are inconclusive (Table 4) and median values of these metrics for the grid and the map are the same, in the majority of tasks, map works better than grid. The only exception is T#1-EstimVal, in which a glyph is marked and participants estimate its datapoint values. However, given the nature of this task, display type makes no difference to users.

It can be assumed that maps' geographical background may function as noise in user tasks not related to the spatial context, and therefore arranging glyphs in a grid may facilitate their decoding and increase user performance in tasks such as T#1-EstimVal and T#3-FindGlyph. However, on maps, similar glyphs are more likely to be near to each other, since adjacent municipalities may feature similar variable scores. Therefore, the map's better results in T#4-SimilGlyphs and T#6-CompArea do not surprise us. In other cases, the map's better results are not explained. It can be speculated that this might be due to the participants' lack of familiarity with grid plots and due to the use of the grid plot as first in the empirical study. However, ahead of the testing, the participants are explained what grid plots are and how they work. The participants are also given 2 min to freely examine the display used in the empirical study.

RQ#3: analytic versus holistic users

Apart from the differences between different glyph and display types, we examine whether there are certain features that influence user performance. We therefore verify whether different cognitive styles (analytic and holistic) influence user behavior. We take into account answer accuracy, task accomplishment time, and fixation frequency (Table 5). Although in T#2-IdenGlyph and T#3-FindGlyph, holistics and analytics show the same results, in other tasks analytics generally feature greater answer accuracy, whereas holistics need less time to solve tasks and feature higher fixation frequency. However, in the comparison between analytic and holistic users, statistically significant differences are revealed only for the fixation frequency in T#5-DistinctGlyphs and T#6-CompArea. These results are in accordance with those reported by Tang (2010): For the majority of tasks, fixation frequency is higher for holistics, because they are more proficient at sensing a system's large-scale patterns and reacting to them instead of investigating the system's parts.

Although the comparison between analytic and holistic users provides inconclusive results, they can serve as suggestions for further research. In certain tasks (e.g. T#1-EstimVal), analyzing glyph details is more important than sensing glyphs' large-scale patterns and reacting to them (Dewey, 2004). Therefore, it may explain the better answer accuracy observed for analytics. Furthermore, the shorter task accomplishment time and higher fixation frequency of holistic participants is not surprising, since this visual behavior is expected for such users, as they tend to act quicker and focus on general patterns (Kubíček et al., 2016; Navon, 1977). Nevertheless, it is necessary to take account of the fact that, in addition to cognitive style, the way glyphs are marked on a map may also be affected by cartographic knowledge and experience, the pursuit for innovative solutions, and certain personal aspects, which are not included in the analytical and holistic dimensions, such as care.

Conclusions

Glyphs facilitate visual analysis of multivariate geographical data. Star glyphs are particularly common in geovisualization; however, as they make use of polar coordinates, their decoding is impeded. To remedy this, in geovisualization tools – both in map displays and grid plots – star glyphs can be replaced by polyline glyphs. Our study reveals that if either polyline or star glyphs can be used, polyline glyphs are better for facilitating tasks in which datapoint values are to be read.

Table 4. Grid plot versus map display.

Task	Better answer accuracy	Shorter task accomplishment time	Eye-movement data analysis: revisits of specific AOIs (R) or SL	
T#1-EstimVal	Grid plot	Map display[a]	R	Map display
T#2-IdenGlyph	Inconclusive	Map display	SL	Map display[a]
T#3-FindGlyph	Map display (no errors)[a]	Map display	R	Inconclusive
T#4-SimilGlyphs	Map display[a]	Map display[a]	SL	Map display[a]
T#5-DistinctGlyphs	Map display	Map display	SL	Map display[a]
T#6-CompArea	Map display[a]	Map display[a]	SL	Map display[a]

SL: Scanpath length. R: Revisits. [a]Statistically significant or strong results.

Table 5. Analytic versus holistic users.

Task	Better answer accuracy	Shorter task accomplishment time	Higher fixation frequency
T#1-EstimVal	Analytic users	Analytic users	Analytic users
T#2-IdenGlyph	Inconclusive	Holistic users	Analytic users
T#3-FindGlyph	Inconclusive	Holistic users	Holistic users
T#4-SimilGlyphs	Holistic users	Holistic users	Holistic users
T#5-DistinctGlyphs	Analytic users	Holistic users	Holistic users[a]
T#6-CompArea	Analytic users	Holistic users	Holistic users[a]

[a]Statistically significant results.

By contrast, if the purpose is to facilitate visual search among glyphs (i.e. to find similar or distinctive glyphs), then star glyphs seem to be a better choice. Moreover, our study reveals that polyline and star glyphs arranged as a map display work generally better than glyphs grouped in a grid plot: Participants who use glyphs in the grid to solve user task perform worse than those who use glyphs shown on the map display. However, this finding needs more research in the future.

There are no particular differences in the visual behavior of participants who use polyline glyphs and participants who interact with star glyphs. One finding is that participants use the key (legend) more frequently if they read datapoint values from star glyphs than if they do so from polyline glyphs. Therefore, our research finding is that a key needs to be better incorporated in a visual interface if star glyphs are to be used to support such user tasks. Finally, glyphs are likely to be used more accurately by analytic users, although analytic users can take more time in comparison with holistic users.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the KLIMAFORSK program – the Norwegian Research Council's initiative for climate research [Grant Number 235490], "Climate change, and natural hazards: The geography of community resilience in Norway" (ClimRes) and by the Internal Grant Agency of the Palacký University in Olomouc (project IGA_PrF_2017_024, Cloud-based Platform for Integration and Visualisation of Different Kind of Geodata).

ORCID

Tomasz Opach http://orcid.org/0000-0002-3561-1691
Stanislav Popelka http://orcid.org/0000-0003-2127-2614
Jitka Dolezalova http://orcid.org/0000-0003-2694-5304
Jan Ketil Rød http://orcid.org/0000-0003-2935-6206

References

Andrienko, G., Andrienko, N., & Jankowski, P. (2003). Building spatial decision support tools for individuals and groups. *Journal of Decision Systems*, 12(2), 193–208. doi:10.3166/jds.12.193-208

Andrienko, N., Andrienko, G., Voss, H., Bernardo, F., Hipolito, J., & Kretchmer, U. (2002). Testing the usability of interactive maps in commonGIS. *Cartography and Geographic Information Science*, 29(4), 325–342. doi:10.1559/152304002782008369

Arnberger, E. (1977). *Thematische Kartographie* [Thematic Cartography]. Brunswick, Germany: Westermann.

Bertin, J. (1967). *Sémiologie Graphique: Les diagrammes, les réseaux, les cartes* [Semiology of graphics: Diagrams, networks, maps]. Paris: Gauthier-Villars.

Borgo, R., Kehrer, J., Chung, D. H. S., Maguire, E., Laramee, R. S., Hauser, H., … Chen, M. (2013). Glyph-based visualization: Foundations, design guidelines, techniques and applications. In M. Sbert & L. Szirmay-Kalos, *Eurographics 2013 - State of the Art Reports* (pp. 39–63). Geneva: The Eurographics Association.

Brand, J., & Johnson, A. P. (2014). Attention to local and global levels of hierarchical Navon figures affects rapid scene categorization. *Frontiers in Psychology*, 5(Article 1274). doi:10.3389/fpsyg.2014.01274

Chernoff, H. (1973). The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68(342), 361–368. doi:10.1080/01621459.1973.10482434

Chung, D. H. S., Legg, P. A., Parry, M. L., Bown, R., Griffiths, I. W., Laramee, R. S., & Chen, M. (2015). Glyph sorting: Interactive visualization for multi-dimensional data. *Information Visualization*, 14(1), 76–90. doi:10.1177/1473871613511959

Dewey, R. A. (2004). *Introduction to Psychology*. Belmont, CA: Wadsworth Publishing.

Dolezalova, J., & Popelka, S. (2016). ScanGraph: A novel scanpath comparison method using graph cliques visualization. *Journal of Eye Movement Research*, 9(4), 5, 1–13. doi:10.16910/jemr.9.4.5

Dorling, D. (1995). The visualization of local urban change across Britain. *Environment and Planning B: Planning and Design*, 22(3), 269–290. doi:10.1068/b220269

Forsberg, A. S., Chen, J., & Laidlaw, D. H. (2009). Comparing 3D vector field visualization methods: A user study. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 1219–1226. doi:10.1109/TVCG.2009.126

Fuchs, J., Fischer, F., Mansmann, F., Bertini, E., & Isenberg, P. (2013). Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3237–3246). New York: ACM. doi:10.1145/2470654.2466443

Goldberg, J., & Helfman, J. (2011). Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information Visualization*, 10(3), 182–195. doi:10.1177/1473871611406623

Golebiowska, I., Opach, T., & Rød, J. K. (2017). For your eyes only? Evaluating a coordinated and multiple views tool with a map, a parallel coordinated plot and a table using an eye-tracking approach. *International Journal of Geographical Information Science*, 31(2), 237–252. doi:10.1080/13658816.2016.1191636

Gribov, A., Unwin, A., & Hofmann, H. (2006). About glyphs and small multiples: Gauguin and the Expo. *Statistical Computing and Graphics Newsletter*, 17(2), 18–22.

Ho, H.-Y., Yeh, I.-C., Lai, Y.-C., Lin, W.-C., & Cherng, F.-Y. (2015). Evaluating 2D flow visualization using eye tracking. *Computer Graphics Forum*, 34(3), 501–510. doi:10.1111/cgf.12662

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van De Weijer, J. (2011). *Eye tracking: A* *comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Klippel, A., Hardisty, F., Li, R., & Weaver, C. (2009). Colour-enhanced star plot glyphs: Can salient shape characteristics be overcome? *Cartographica: The International Journal for Geographic Information and Geovisualization*, 44(3), 217–231. doi:10.3138/carto.44.3.217

Kubíček, P., Šašinka, Č., Stachoň, Z., Štěrba, Z., Apeltauer, J., & Urbánek, T. (2016). Cartographic design and usability of visual variables for linear features. *The Cartographic Journal*, 54(1), 91–102. doi:10.1080/00087041.2016.1168141

Lee, M. D., Reilly, R. E., & Butavicius, M. E. (2003). An empirical evaluation of Chernoff faces, star glyphs, and spatial visualizations for binary data. In *Proceedings of the Asia-Pacific symposium on Information visualization – Volume 24* (pp. 1–10). Darlinghurst: Australian Computer Society.

Morong, K., & Šašinka, Č. (2014, May). *Hypothesis – online software platform for objective experimental testing*. Paper presented at the Applying principles of cognitive psychology in practice, Brno, Czech Republic.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9(3), 353–383. doi:10.1016/0010-0285(77)90012-3

Opach, T., & Rød, J. K. (2013). Cartographic visualization of vulnerability to natural hazards. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 48(2), 113–125. doi:10.3138/carto.48.2.1840

Opach, T., & Rød, J. K. (2017). Augmenting the usability of parallel coordinate plot: The polyline glyphs. *Information Visualization*. Advance online publication. doi:10.1177/1473871617693041

Ostrowski, J., & Uhorczak, F. (1972). Typogramy F. Uhorczaka jako środek graficznej prezentacji zjawisk wielocechowych [Typograms of F. Uhorczak as a means of graphical presentation of multi-feature phenomena]. *Polski Przegląd Kartograficzny*, 4(4), 145–151.

Popelka, S. (2014). Optimal eye fixation detection settings for cartographic purposes. *Proceedings of the 14th SGEM GeoConference on Informatics, Geoinformatics and Remote Sensing (SGEM'14)*, 1, 705–712.

Ratajski, L. (1989). *Metodyka kartografii społeczno-gospodarczej* [The Methods of Socio-Economic Cartography] (2nd ed.). Warsaw: PPWK.

Salvucci, D. D., & Goldberg, J. H. (2000, November). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications* (pp. 71–78). New York: ACM. doi:10.1145/355017.355028

Šašinka, Č. (2013). *Interindividuální rozdíly v percepci prostoru a map* [Interindividual differences in perception of space and maps] (Doctoral dissertation). Masaryk University, Brno, Czech Republic.

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2010). *Thematic cartography and geographic visualization* (3rd ed.). Upper Saddle River, NJ: Pearson.

Takatsuka, M., & Gahegan, M. (2002). GeoVISTA *Studio*: A codeless visual programming environment for geoscientific data analysis and visualization. *Computers & Geosciences*, 28(10), 1131–1144. doi:10.1016/S0098-3004(02)00031-6

Tang, D. (2010). *Read between the eyes: Cognitive style difference between Americans and Chinese in a dual-display setting* (Unpublished master's thesis). Cornell University, Ithaca, NY. Retrieved from http://hdl.handle.net/1813/17776

Tufte, E. R. (1990). *Envisioning Information*. Cheshire, CT: Graphics Press.

Ünlü, A., & Malik, W. A. (2011). Interactive glyph graphics of multivariate data in psychometrics: The software package gauguin. *Methodology. European Journal of Research Methods for the Behavioral and Social Sciences*, 7(4), 134–144. doi:10.1027/1614-2241/a000031

Ward, M. O. (2002). A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1(3–4), 194–210. doi:10.1057/palgrave.ivs.9500025

Ward, M. O. (2008). Multivariate data glyphs: Principles and practice. In C.-H. Chen, W. K. Hardle, & A. Unwin (Eds.), *Handbook of data visualization* (pp. 179–198). Berlin: Springer. doi:10.1007/978-3-540-33037-0_8

Paper UrbanPlans

Burian, J.*, **Popelka, S.**, & Beitlova, M. (2018). Evaluation of the Cartographical Quality of Urban Plans by Eye-Tracking. ISPRS International Journal of Geo-Information, 7(5), 192. http://www.mdpi.com/2220-9964/7/5/192

[IF 1.840; Q3]

[21 citations on WoS; 17 citations without autocitations]

[Author's contribution: 40%]

Abstract

This paper describes a study of the evaluation of cartographic quality of urban plans in the Czech Republic using eye-tracking. Although map visualization is a crucial part of the urban planning process, only a few studies have focused on the evaluation of these maps. The plans of four Czech cities with different styles of visualization and legends were used in this eye-tracking experiment. Respondents were required to solve spatial tasks consisting of finding and marking a certain symbol on a map. Statistical analyses of various eye-tracking metrics were used, and the differences between experts and students and between the map and legend sections of the stimuli were explored. The study results showed that the quality of map symbols and the map legend significantly influence the legibility and understandability of urban plans. For correct decision-making, it is essential to produce maps according to certain standards, to make them as clear as possible, and to perform usability testing on them.

International Journal of *Geo-Information*

Article

Evaluation of the Cartographical Quality of Urban Plans by Eye-Tracking

Jaroslav Burian *, Stanislav Popelka and Marketa Beitlova

Department of Geoinformatics, Palacký University in Olomouc, 17. listopadu 12, 779 00 Olomouc, Czech Republic; stanislav.popelka@upol.cz (S.P.); marketa.beitlova01@upol.cz (M.B.)

* Correspondence: jaroslav.burian@gmail.com; Tel.: +42-072-471-2953

Received: 19 March 2018; Accepted: 16 May 2018; Published: 17 May 2018

Abstract: This paper describes a study of the evaluation of cartographic quality of urban plans in the Czech Republic using eye-tracking. Although map visualization is a crucial part of the urban planning process, only a few studies have focused on the evaluation of these maps. The plans of four Czech cities with different styles of visualization and legends were used in this eye-tracking experiment. Respondents were required to solve spatial tasks consisting of finding and marking a certain symbol on a map. Statistical analyses of various eye-tracking metrics were used, and the differences between experts and students and between the map and legend sections of the stimuli were explored. The study results showed that the quality of map symbols and the map legend significantly influence the legibility and understandability of urban plans. For correct decision-making, it is essential to produce maps according to certain standards, to make them as clear as possible, and to perform usability testing on them.

Keywords: urban planning; cartography; evaluation; urban plan; master plan; eye-tracking

1. Introduction

A key part of any urban planning process is the final output visualized by the map. Urban and regional planning would be completely chaotic without maps [1]. Churchill [2] stated that, although the term "urban cartography" only came into use after the Second World War, it is plausible to consider town plans as the oldest forms of urban maps. Because a plan is also a map, cartographic and geoinformation principles should also be considered [3].

Spatial planning policy documents in Europe involve a symbolic representation of the territory in the form of icons, diagrams, and maps. Cartographic visualization, or the conceptualization of territory, is an integral part of spatial planning [4]. Dühr [4] stated that not much research has been undertaken on the use of cartography in planning. Jarvis [5] even commented that planning theory hardly touches "drawing" at all. Neuman [6,7] for Spain, Lussault [8] for France, and Gabellini [9] for Italy, for example, investigated the communicative potential of visualizations in urban planning. Söderström's work [10] has concentrated on understanding how the structure of visualizations influences the activities of planners in Swiss towns cities (Bern and Zurich).

1.1. Urban Plan Standardization

The cartographical aspects of urban plans are represented by the processes of standardization. Map symbol standardization received early attention from academic and practicing cartographers beginning more than 150 years ago [11]. According to Robinson [11], the first printed discussion of map symbol standardization was introduced by Funkhouser [12]. Symbol standardization was applied to economic maps, topographic maps, and transportation maps, mostly during the 1970s, but urban and regional planning were not included. In the case of a map symbology for regional planning, many opinions and completely different outcomes with respect to the graphical aspect can be found [13].

The standardization of urban (master) and regional plans has been investigated by many authors, for example, in Poland [14], Hungary [15] and Norway [16]. The relevant documents adopt approaches that vary from complete indifference to very strict rules for a map symbology set. For example, in the Polish standardization, colors are used for various types of land use. These colors are very different from the colors commonly used in Czech urban planning. This discrepancy can lead to misunderstandings at borders where the two plans touch [17].

Urban planning has a long tradition in the Czech Republic. Urban plans have been created since the 1930s, and the importance of maps as a part of the urban plan has increased. Most of the maps included in the urban plan have become more detailed and more complicated, decreasing the level of understanding. These changes were visualized and described by Burian et al. [18] using the example of five urban plans for the city of Olomouc (1930, 1955, 1985, 1999, and 2010). Plans with more complicated structures and content led to a process of standardization. The Construction Act [19] can be seen as the first methodological approach. During the last 20 years and in connection with the implementation of GIS (Geographical Information Systems) techniques, many methodologies have appeared (e.g., [20]). Although the public sector has authored methodologies, many regional methodologies have also been created by private companies, such as Hydrosoft Veleslavín [21,22] or T-MAPY [23,24]. However, none of these methodologies address the issues of map symbology creation in detail. Several attempts to create a standardized set of map symbols for urban plans have been introduced by several Czech companies in connection with regional methodologies, but none of these symbol sets have been published.

Currently, no uniform approach to the cartographic visualization of urban plans in the Czech Republic [17] exists. With no sufficient standardization, the author-designer makes many decisions about the visual appearance of the symbology. This can be an issue because the designer might not have sufficient experience to make these decisoins. Therefore, flaws might arise in map composition, map symbology, and other important cartographic aspects. This can reduce legibility and understandability and even lead to misunderstandings. Many of these failures connected with missing standardization were elaborated by the authors of this manuscript in several publications in the past (e.g., [3,13]). The authors applied subjective research methods on more than 50 Czech urban plans and concluded that, in most of the actual plans, many technical and cartographical failures can be observed. Especially in the case of plans comparison or cross-border tasks, wrong decisions can appear. The authors also conclude that the cartographic quality or urban plans and missing unified standardization are the significant issues that should be solved. These conclusions were accepted only by som of the involved urban planners. The rest requested for the confirmation of subjective results using more objective approach.

For this reason, authors suggested using the new symbology that has been created in cooperation between Palacký University in Olomouc and the Regional Authority of the Olomouc Region (see [3,17] for details). The authors considered four aspects of symbology creation: users, cartographic correctness, established convention, and the use of a data model. The advantage of the use of this standardized symbology is that urban plans in different districts look uniform and can be compared. This symbology is currently in use in seven of the eleven districts in the Olomouc region.

1.2. Urban Plans Evaluation

According to Štěrba et al. [25], the quality of cartographic visualization influences the user's judgement and subsequently his or her decision. The overall usability of the map (as a strictly objective criterion) should be a determining aspect when evaluating the quality of cartographic visualization.

Various methods can be used to analyze maps and cartographic visualization in spatial planning. Rohrer [26] presented "A Landscape of User Research Methods", in which various user-experience research methods are shown on a three-dimensional framework with the following axes:

- Attitudinal vs. Behavioral (Subjective vs. Objective)

- Qualitative vs. Quantitative
- Context of Use

This scheme has been modified and is shown in Figure 1. According to this distribution [26], eye-tracking is considered as a behavioral (objective) method, because it shows "what people do", instead of "what people say". From the qualitative/quantitative point of view, eye-tracking lies in the middle, which means that recorded data could be analyzed qualitatively as well as quantitatively. From the perspective of "context of use", eye-tracking experiments could be designed as natural (in the real world, especially using eye-tracking glasses) or lab-based. This lab-based design was used in the presented case study with urban plans.

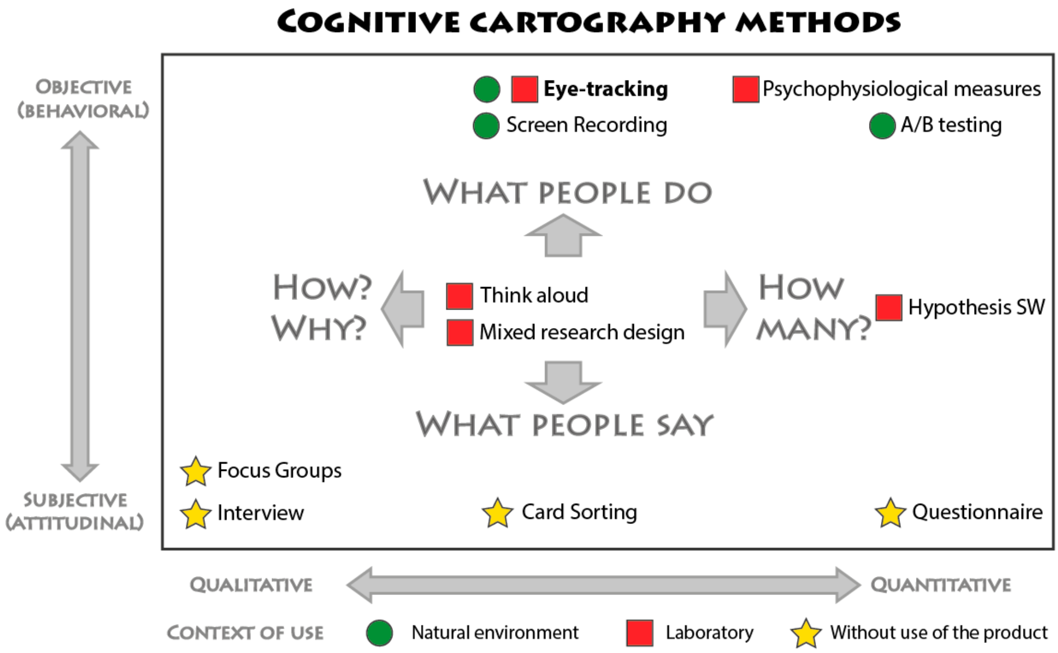


Figure 1. Cognitive cartography methods (modified from [26]).

Dühr [4,27] described several approaches to analyzing cartographic visualization in urban planning ([10,28–30]). Pickles [29] suggested an approach to map analysis not unlike discourse analysis, which treats maps as an expanded concept of text. Two interrelated internal structures of the map are considered: graphical and linguistic. According to Dühr [4], Söderström [10] proposed what he called a "visual circuit" for the analysis of spatial representations in planning. This visual circuit consists of four interrelated fields: the context of elaboration, the process of production, the context of usage of the visualization, and the "materialization" or implementation.

Based on these methods, Dühr [4,27] introduced a method of map analysis for strategic spatial plans in Germany, England, and Denmark. Because neither Pickles nor Harley provided a detailed list of criteria for analysis, Dühr [31] suggested three categories for a comparative analysis of cartographic representation: the level of abstraction, the level of complexity, and the use of associative colors and symbols "on the map". Each category consists of several criteria for analyzing cartographic representation (spatial positioning, visual hierarchy, complexity, map symbols, map style, etc.). Most of the selected criteria use the qualitative method, which is sometimes very subjective, and its application to different maps from different regions can be difficult. However, even between planning systems within a single "planning tradition", we see differences in approach to mapping. Dühr's findings show that the function of a plan ultimately determines the form and style of the visualization. A clear difference between the comprehensive and regulatory approach to planning is evident in the German and Dutch plans compared to England's less formal approach [17].

Another comparative analysis is that of Tang and Hurni [32], who provided an example comparing China and Switzerland at different planning tiers in respect of the plan contents and their symbolic modalities and visual styles. They conclude that the symbol system in Chinese cases has better logic and hierarchies but lacks the harmonious and exquisite vision of the Swiss cases. In general, the vertical comparison of visual styles of the planning maps in the two countries' planning systems show to a certain extent a substantial difference between federal and other levels in Switzerland and a resemblance between the visualization of thematic layers of all tiers in China despite the existence of diverse symbol faces. Next, the transverse cross-comparison between the two countries denotes that the cartographic style of Swiss federal spatial concepts is analogous to the urban system plans above the prefectural level in China, but the Swiss cantonal plans are widely divergent from the Chinese provincial plans [32].

Qualitative research described by Dühr [27,31,33], Tang and Hurni [32], Söderström [10] and Pickles [29] has not determined how people behave when solving tasks using urban plans. Those authors compared several spatial plans (plans at regional, country, national or transnational level) using qualitative methods and described the similarities and differences, but do not evaluate cartographical quality.

1.3. Eye-Tracking

For the reasons mentioned in Section 1.2, a behavioral research method (eye-tracking) was adopted to analyze urban plans as objectively as possible. Eye-tracking is one of the most precise and objective methods of usability studies because eye-movement recording does not rely on self-reporting [34]. With the use of eye-tracking, it is possible to gather information that is inaccessible using any other technique, particularly when information about people's behavior when solving tasks is difficult to observe by any other method (e.g., how much time participants spend on the different sections of urban plans, such as the map or legend). Other aspects of people's behavior, such as Trial Duration or the correctness of answers, can be investigated using other methods, for example, by direct observation or screen recording. In these examples, it would be much more difficult to distinguish what participants were really doing.

One of the first comprehensive publications to address the application of eye-tracking in cartography was that of Steinke [35], who summarized the results of the former research and stressed the importance of distinguishing between user groups according to their age and education. More recent studies using eye-tracking in cartography have focused on evaluating cartographic principles [36], interactive maps [37], small multiple map displays [38], graphical outputs from GIS [39], the differences between experts and novices [40], map uncertainty [41,42], 3D visualization in maps [43,44], and color schemes [45].

The examination of users' perception in static maps is clearly related with an examination of their reaction in visual variables change. The first cognitive studies in cartography were focused on the effectiveness of symbols used in thematic maps. An example of such study may be Taylor's work [46] where graphical dimensions of symbols such as length, area, or color were investigated. An eye-tracking was used for investigation of the visual variables for example in the classic study of Garlandini et al. [47], who investigated the influence of a change of four visual variables: size, color value, color hue, and orientation. Petchenik [48] stated that, for a successful transfer of information between a cartographer and a map reader, the reader must understand the map in the same way as the author has created it. The task of cognitive cartography is to reveal how users read the individual map elements and how the meaning assigned to these elements varies between different users.

Although cartographic user research of visual variables has a long tradition, the investigation of complex map works is relatively unexplored. A clear example of a complex map containing many various symbols is a geological map. Two legend designs (alphabetically ordered and color-ordered) of soil–landscape (geological) maps were compared in the study of Coltekin et al. [49]. Similar to geological maps, urban plans contain many layers represented by various symbols. According to our knowledge, eye-tracking (or another user-experience method) have not been previously used to evaluate urban plans. Only a few studies have addressed landscape perception, and studies focused on specific urban planning issues such as flooding maps [50]. One example is the study of Dupont et al. [51], who analyzed the observation patterns of 23 participants viewing different types of photographs of the landscape in Flanders (Belgium). The authors tested whether the degree of openness and heterogeneity of a landscape affects the observation pattern. The analysis clearly reveals that both landscape characteristics have an influence. Kim et al. [52] analyzed nightscape (night-time landscape) images using traditional survey methods (preference survey) and eye-movement analysis. The authors found a significant relationship between the results of the preference survey and recorded eye-movement data. Noland et al. [53] surveyed the eye-tracking visual preferences of 20 participants using a set of 40 images. The aim of the study was to qualitatively evaluate how individuals process and rank images in public settings for urban planning. For the data analyses, Areas of Interest around important parts of the images (cars, buildings, sidewalks, etc.) were marked. Time to First Fixation, Time Spent (Dwell Time), and Fixation Counts were investigated together with the qualitative ranking. The results showed that cars, parking, and advertisements are associated with negative rankings, but attract a participant's attention. All these studies used photographs for the experiments, and no study focusing on the eye-tracking evaluation of urban plans can be found.

In the study described in this paper, we focused on the analysis and evaluation of the cartographical quality of selected urban plans in the Czech Republic. For this purpose, the following hypotheses were made and investigated using the eye-tracking method:

1. Map symbology (number of colors, map symbols, and features/layers on the map) significantly influences the legibility and understandability of plans, which will impact the duration and correctness of the tasks.
2. Legend structure significantly influences the legibility and understandability of the plans, which will impact the number of fixations and length of dwell time in the Legend AOI.
3. Differences between students and experts in how these groups read plans will impact the duration and correctness of the tasks.

2. Materials and Methods

This section contains three main parts. The study's design is described along with how data were prepared for analyses (identifying fixations). Finally, the methods used for analyses are explained.

2.1. Study Design

2.1.1. Selection of Plans

In the Czech Republic, "Act No. 183/2006 Coll., the Construction Act" [54] provides two main urban and regional planning tools: "Analytical Material for Planning" and "Planning Documentation". Planning Documentation (regional plans and master plans) must be created, updated, and published online for each region (14 regions in the Czech Republic) and each municipality (6258 municipalities). The master plan is a set of specific text and graphic documents that regulate and propose construction in a designated area. The graphical part consists of several maps visualizing several aspects of city planning, for example, zoning plan, land-use limits map, water and waste management map, nature protection map, utility networks map, transportation map, etc. Each map consists of many thematic layers that are not easy to visualize together, even though a large scale of 1:5000 is used.

This study considers urban plans published using web applications. To analyze comparable urban plans, it was necessary to choose applications with similar layouts (large map, a legend on the right hand side, and no additional map features). Four urban plans created by different authors using different styles and published in different years were selected. The oldest plan is that of Jihlava (1999—revised in 2013). The Hradec Králové plan was created in 2012, and the Bohumín and Olomouc plans were created in 2014. All plans were created by private companies; each company adopted their approach to visualizing the urban plan (no methodology described in the Introduction was used).

Each plan represents a different map style. Map styles have been investigated by many authors (e.g., [55–57]). According to Beconyte [55], style in modern cartography can be defined as a set of parameters, some of which are determined by the map scale, theme and general purpose, whereas others are subject to the designer's free choice. Beconyte [55] defined the main parameters that allow a map's style to be defined: decorativeness, expressiveness, and originality. To select plans with different map styles, we used expressiveness (composition; proportion and colors of map symbols and text) as the main parameter.

The tasks were designed as static views, and it was therefore not possible to move the maps. The advantage of this approach is that the maps had similar scales and eye-tracking data recorded for all respondents and their answers could be directly compared. Nevertheless, it was impossible to use images as stimuli in the eye-tracking experiment. The legend for all the maps was longer than the height of the image. We created an HTML page for each city/task consisting of the image with the static map on the left and a legend in a window with a slider on the right. The results look exactly same as the original webpages with the plan, and, although it was possible to move the legend, the map was static. The same zoom level (scale 1:5000) was used for all plans because the printed version of each plan in this scale is also available. In this scale, only a small part of the city was displayed on the screen, which was useful to eliminate any participant knowledge of the cities.

Figure 2 shows an overview of the stimuli used. Tasks Q1 and Q3, which were used in our analysis, were omitted from this overview because these tasks were given on the same plan as for Q2 (Zoning map). Plans in higher resolution can be accessed via www.eyetracking.upol.cz/urban, where all the stimuli used in the eye-tracking experiment are displayed. All the stimuli were prepared to a resolution of 1920 × 1200 px.

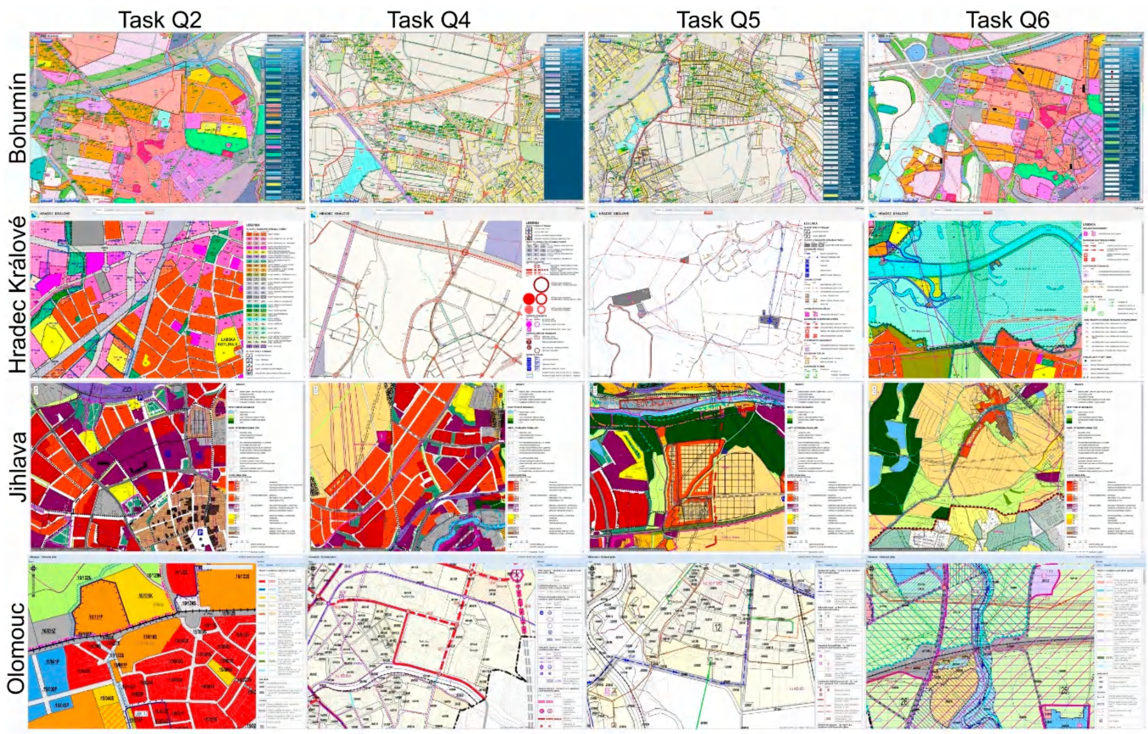


Figure 2. Overview of selected plans.

2.1.2. Selection of Tasks

To cover the most typical tasks during standard work with an urban plan (i.e., finding a new place for housing or identifying areas with proposed public services), analysis of the urban plans was based on six tasks. There are many different taxonomies of tasks (see [58] for a detailed description). According to the commonly used Wehrend and Lewis [59] objective-based taxonomy, we used the simplest task category (Identify). To prevent any misunderstandings, the tasks were created to be as simple as possible. The tasks focused on the point, line and polygon map features to cover as many cartographical symbols as possible. One task (Q3) involved proposals, not real conditions. The following questions were asked:

- Q1: Mark an area for housing (Zoning map)
- Q2: Mark an area for sports or recreation (Zoning map)
- Q3: Mark an area for proposed public services (Zoning map)
- Q4: Mark a railroad (Transportation map)
- Q5: Mark a wastewater treatment plant (Utility networks map)
- Q6: Mark a protected area of water resources (Natural protection map)

2.1.3. Procedure

Finally, six standalone webpages were obtained for each city. Eye-tracker SMI RED 250 (developed by SensoMotoric Instruments, Berlin, Germany) with a sampling frequency of 250 Hz was used in the study, and the experiment was created at the SMI Experiment Center. Stimuli were presented on a monitor with a resolution of 1920 $\times$ 1200 px. At the beginning of the experiment, the respondents completed a short questionnaire about their age and experience, and then calibration was performed. We set a calibration threshold of 1° of the visual angle. Respondents with a higher deviation were excluded from the results, along with those whose Tracking Ratio (the proportion of time that the eye tracker recorded point of gaze coordinates over the entire experiment) was higher than 90%. After calibration, respondents were informed about the purpose of the experiment and provided with some basic information about urban planning. The tasks were then presented. No time limits were set for respondents to read and remember the task. After pressing the F2 key, an Internet browser would open, displaying a webpage with one of the stimuli automatically. The tasks were presented sequentially from Q1 to Q6. The city plans were randomized for each task. At the end of the experiment, a short questionnaire about the respondents' subjective opinions on the presented plans was displayed. This questionnaire was created in Google Forms and contained an image of the representative plan for each city (to remind the respondent how it looked) and a scale from 1 (best) to 5 (worse). The experiment was ten to fifteen minutes long. A diagram showing the experiment's design is in Figure 3.

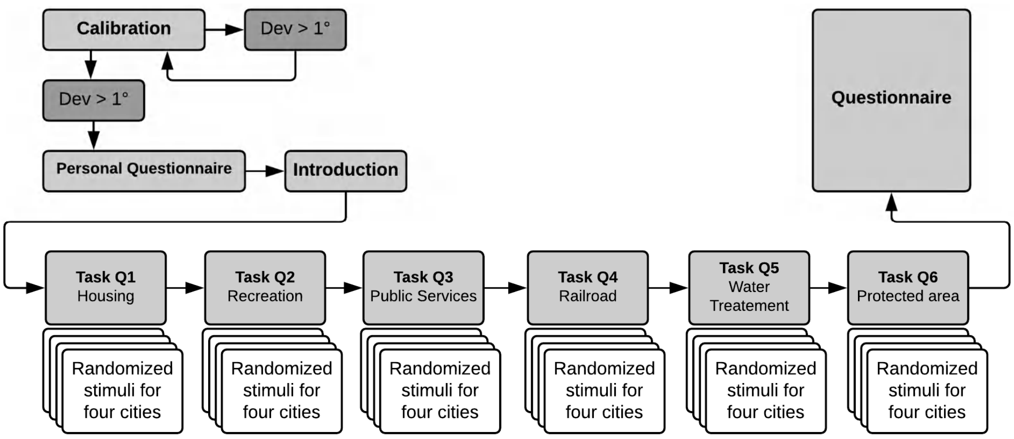


Figure 3. Design of the Experiment.

2.1.4. Participants

Thirty-four respondents participated in the experiment. Because of the inaccuracy of calibration or an insufficient Tracking Ratio, eight were excluded from the results analysis. The device SMI RED 250 used in the study is a remote eye-tracker. The errors were caused by problems with drooping eyelids and participants moving towards the screen. Quality over quantity was preferred and therefore the participants with tracking ratios of less than 1° of the visual angle were excluded. Finally, we obtained 26 respondents. Twenty were students (11 male and 9 female) in the third year of the Geoinformatics and Cartography bachelor's program or the Geoinformatics master's program. Students were used as the majority of the sample because they have studied urban planning and cartography courses and generally have a basic knowledge of urban plans and cartography. This ensured that their knowledge in these areas was at similar levels and the results would be comparable.

Six respondents (1 male and 5 female) were experts working in the departments of urban planning at the Municipality of the City of Olomouc or the Regional Authority of the Olomouc Region. These experts work with urban plans every day and had similar backgrounds. Thus, it was possible to consider them as a consistent group with relatively homogenous experience and skills.

According to their statements, all the participants had normal or corrected to normal vision and were not color blind.

2.2. Fixations and Their Detection

The eyes move in many ways, simultaneously responding to commands from several different brain areas. One of the most important types of eye movement, known as fixation, is not really a movement at all, but instead is the ability to keep the eye trained on a fixed spot in the world. It is generally considered that when we measure fixation, we also measure attention to that position [60].

Our visual experience consists of a series of fixations on different objects. To get from one fixation to the next, the eyes make rapid, ballistic movements known as saccades [61].

It is important to define the exact detection algorithm for detecting fixations and saccades because different parameterizations of an algorithm might lead to different results. Although many algorithms exist, for low-speed data (up to 250 Hz), the most commonly used algorithm is I-DT, which considers the close spatial proximity of eye position points in an eye movement trace [62]. The algorithm defines a temporal window that moves one point at a time, and the spatial dispersion created by the points within this window is compared against the threshold. For the case study, SMI BeGaze software and the ID-T algorithm were used to detect fixation. Threshold values in BeGaze were set to 80 ms for "Duration threshold" and 50 px for "Dispersion threshold". More information about this setting is described in [63].

2.3. Methods of Data Analyses

Because the data were recorded as screen recording stimuli, the tracking results provided separate video records for each respondent. To analyze them together, it was necessary to combine all the videos according to each task and city combination. To do so, a function of the BeGaze software called Custom Trial Selector was used. The custom trial was designed for each task and city as a screenshot from the video for all city-task combinations. Next, the corresponding part of each recording was assigned to it. With the Custom Trial Selector, whether respondents were looking at the map or the legend could be analyzed. A detailed analysis of eye movements in the map section of the stimulus was also possible. The only disadvantage of this approach was not being able to analyze detailed eye-movements in the legend because respondents moved it in different ways (the legend was long and it was necessary to move it). If detailed analysis of scanpaths is required, the original screen recordings for each participant can be used instead of the custom trials, where data from all participants are displayed together.

To analyze the results, statistical analysis using the Wilcoxon rank-sum test [64] and the Kruskal–Wallis test [65] was performed. These tests were chosen as non-parametric variants of *t*-test and ANOVA because (as in majority of eye-movement studies) the data recorded in our study did not have normal distribution. The purpose of both tests is to check null hypothesis that no difference exists between variables. All data were analyzed at the 95% confidence level. Statistically significant differences were marked directly into boxplots.

Trial Duration showing how long it took to solve a task and Fixation Count describing how many fixations were performed during a task were investigated. Next, the map and legend were analyzed to see in which part of the stimuli participant attention was focused. Sequence Chart visualization was selected to display respondent eye-movements between the map and legend stimuli. The Sequence Chart shows the temporal sequence of the visited Areas of Interest. From the visualization, it is clear where respondents looked first and where they looked later [66]. In addition to Sequence Chart that is included in the software from eye-trackers manufacturer, we visually analyed recorded data using a FlowMap method in V-Analytics software. V-Analytics [67] is intended for the visual analysis of spatio-temporal data and thus can also be used for the analysis of eye-movements [68]. The output of FlowMap shows aggregated eye-movement trajectory of all participants. In the first step, Voronoi polygons covering the whole stimulus are created based on the distribution of fixations recorded over this stimulus. Then, the arrows between these polygons are displayed, and their width represents the number of gaze movements between them. In our case, we constructed arrows with the settings 0; 0; 0; 0; 75 and filtered out arrows displaying less than three moves. With this setting, the output is illustrative and is not overfilled.

According to above mentioned indicators, the plans were qualified as either good or bad. For example, if high values of Trial Duration or Fixation Count were observed, the plan was considered bad. If the participants' answers contained many inaccuracies, the plan was also considered bad. Finally, these objective measurements were compared with the results of the subjective questionnaire.

3. Results

3.1. Trial Duration

The first part of analyzing the recorded eye-movement data focused on the Trial Duration metric (Figure 4). This metric shows how long it took to solve a task. Higher values of Trial Duration are expected for more complex tasks or plans with lower legibility. In these cases, participants may have problems finding the proper symbol in the legend or identifying it on the map. Statistically significant differences between cities for each task according to the Kruskal–Wallis test are shown in the upper part of the figure. From the boxplot, it is evident that some tasks were much more difficult than others. Participants were very fast (below 20 s) in Tasks Q1, Q2, and Q4. The most problematic tasks were Q3 and Q6. For Task Q5, the only problematic plan was for Jihlava.

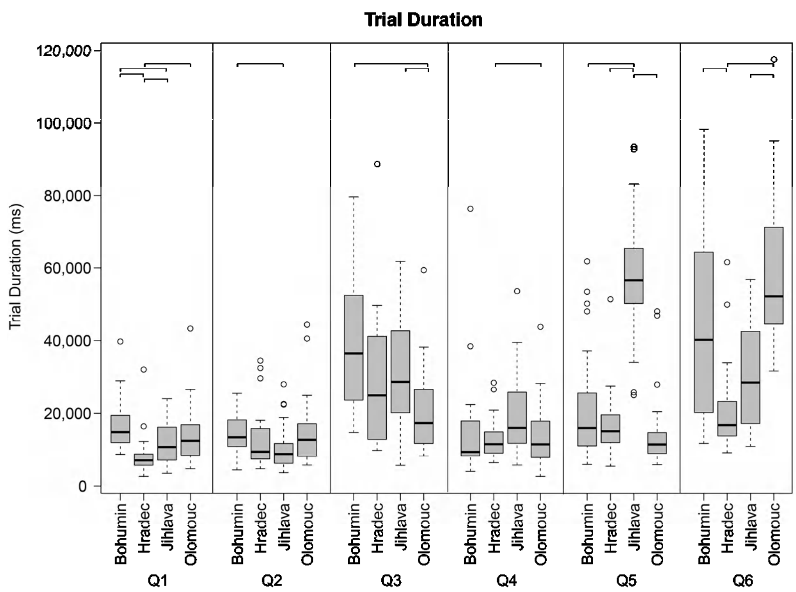


Figure 4. Boxplot of Trial Duration values. The smallest and most consistent values were observed for Hradec Králové and Olomouc.

For the first task, statistically significant differences were found between the four pairs of cities. The most efficient plan was the Hradec Králové plan. By contrast, the highest values of Trial Duration were observed for the Bohumín plan. In this case, the legend is not structured, and finding housing areas took a longer time. In Q2, the results were almost balanced. The best results were recorded for the Jihlava plan, the worst for the Bohumín plan. A statistically significant difference was observed between these two plans. In Q3, Trial Duration for the Bohumín plan was again the highest. The best results were obtained for the Olomouc plan. Balanced results were obtained in Q4, where overall Trial Durations were relatively low. The worst results were obtained for the Jihlava plan, the best for Bohumín. Again, a statistically significant difference was observed between the Trial Duration values of these plans. The highest value of Trial Duration from the whole experiment was recorded for the Jihlava plan in Q5. Statistically significant differences were found between this plan and all others. The respondents had to find the wastewater treatment plant, which was not easy because of the use of incorrect map symbols (the symbol in the legend was not equivalent to the symbol on the map). Very different values of Trial Duration between certain plans were observed in Q6. This task focused on finding a protected area of water resources. The legend of the natural protection map was very complex in the Olomouc case. It contained more than 200 symbols and the description of each symbol was also relatively long. Thus, it took a long time for respondents to complete this task with the Olomouc plan.

3.2. Fixation Count

For the next analysis, Fixation Count was chosen. A higher number of fixations indicates either a low level of efficiency during a search or an inconvenient user interface [34]. We presumed that respondents would perform fewer fixations on urban plans with structured legends and better legibility. However, the values of Trial Duration and Fixation Count are highly correlated, and the boxplot displayed in Figure 4 looks very similar to the Fixation Count.

Figure 5 shows the Fixation Count values for four analyzed cities (the tasks are aggregated). The smallest values of Fixation Count were observed for Hradec Králové (a clearly designed plan with a well-organized legend and a low number of map symbols and colors). Statistically significant differences were found between Hradec Králové and Bohumín and Hradec Králové and Jihlava. The median value for Olomouc was similar to the value of Hradec Králové; therefore, no significant difference between these two cities was found. It may seem that many outliers are in the boxplot, but it is necessary to consider that each city contained six tasks and was observed by 26 participants.

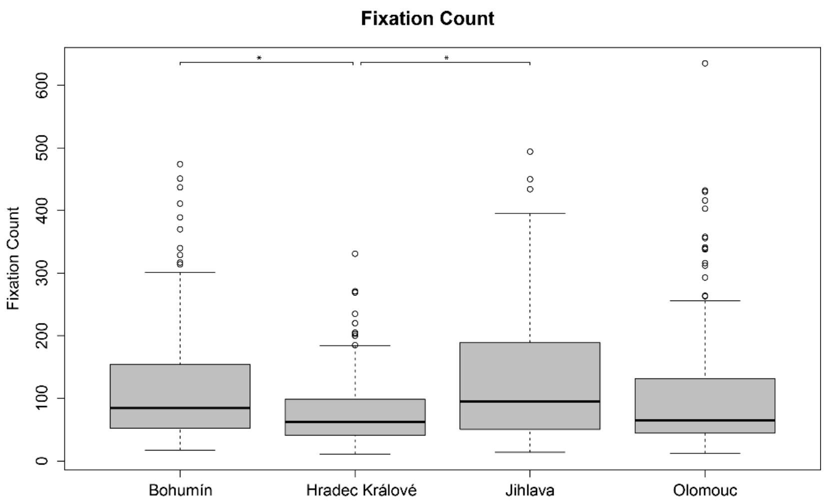


Figure 5. Boxplot of Fixation Count values. According to the results of the Kruskal–Wallis test, statistically significant differences were found between the plans of Hradec Králové and Bohumín and those of Hradec Králové and Jihlava.

In contrast to Figure 5, Figure 6 shows the Fixation Count values for all tasks (the cities are aggregated). Tasks Q1, Q2, and Q4 were easy to solve. Simple questions were aimed at typical tasks probably expected by the respondents. The higher values of Q5 are due to the problematic stimulus of Jihlava. Task Q6 and the stimulus of Olomouc is similar. The symbol used for protected areas of water resources was featureless and was not easy to find on either the map or legend. The high number of fixations for Q3 was due to the task type: respondents were looking for a proposed element (public areas).

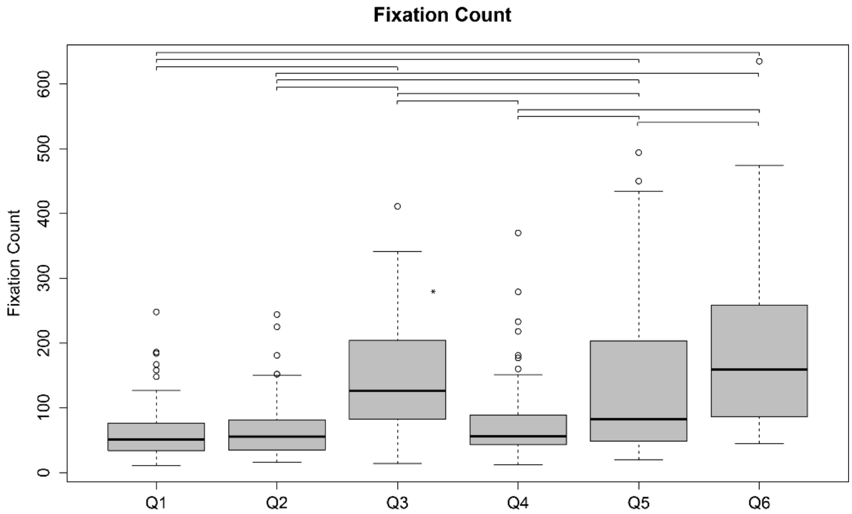


Figure 6. Boxplot of Fixation Count values.

3.3. Differences between Map and Legend

Previous analyses addressed entire stimuli, whereas subsequent processing focused on evaluating user perception of the map and legend sections of the stimuli. It is would be possible to divide the stimuli into more AOIs, but, for the purpose of this study, it was crucial whether participants were looking for the unknown symbol in the legend or searching for an already known symbol on the map. Another option would be to mark, for example, map-targeted symbols or legend-targeted symbols. Unfortunately, using the Custom Trial Selector (Section 2.3), it was not possible to create an AOI around specific symbols in the legend.

Areas of Interest (AOIs) were marked around the map and legend of each stimulus. Areas of Interest are regions in the stimulus in which the researcher is interested [60]. The number of fixations in each AOI were calculated and are shown in Figure 7. It was difficult to locate the correct symbol in the legend for Tasks Q3 and Q6. In Q3, the map symbol for proposed public services was in the legend in the second column, which influenced a high number of fixations in identifying the correct symbol. The most fixations in the legend were observed in Q6 (Olomouc). This result was due to the position of the correct symbol in the legend (in the lower part).

	Fixation Count in AOI "Map" and "Legend"							
	Bohumín		Hradec Králové		Jihlava		Olomouc	
	Map	Legend	Map	Legend	Map	Legend	Map	Legend
Q1	983	1150	707	358	688	843	1066	694
Q2	557	1312	513	1159	435	917	780	1109
Q3	1123	3889	655	2816	1773	2292	917	1658
Q4	829	911	739	874	1004	1370	899	850
Q5	1525	1074	736	1232	5137	2033	902	776
Q6	1470	3844	741	1847	1091	2633	1559	5919

Figure 7. The number of fixations in the AOI "Map" and "Legend" for each task and city combination.

The most important metric for describing user interactions with AOIs is Dwell Time. Dwell Time is calculated as a sum of all of the fixation durations within a prescribed area, in this case, the map and the legend [69]. Figure 8 shows the relative values of Dwell Time for each city. For each city, respondents spent statistically significantly more time in the legend than on the map. The biggest difference was observed for Bohumín, where the legend was not structured at all. This can be interpreted also as the result of low map legibility due to the low level of associativity of the map symbols. In the next part, a detailed analysis of Dwell Times according to the tasks was performed.

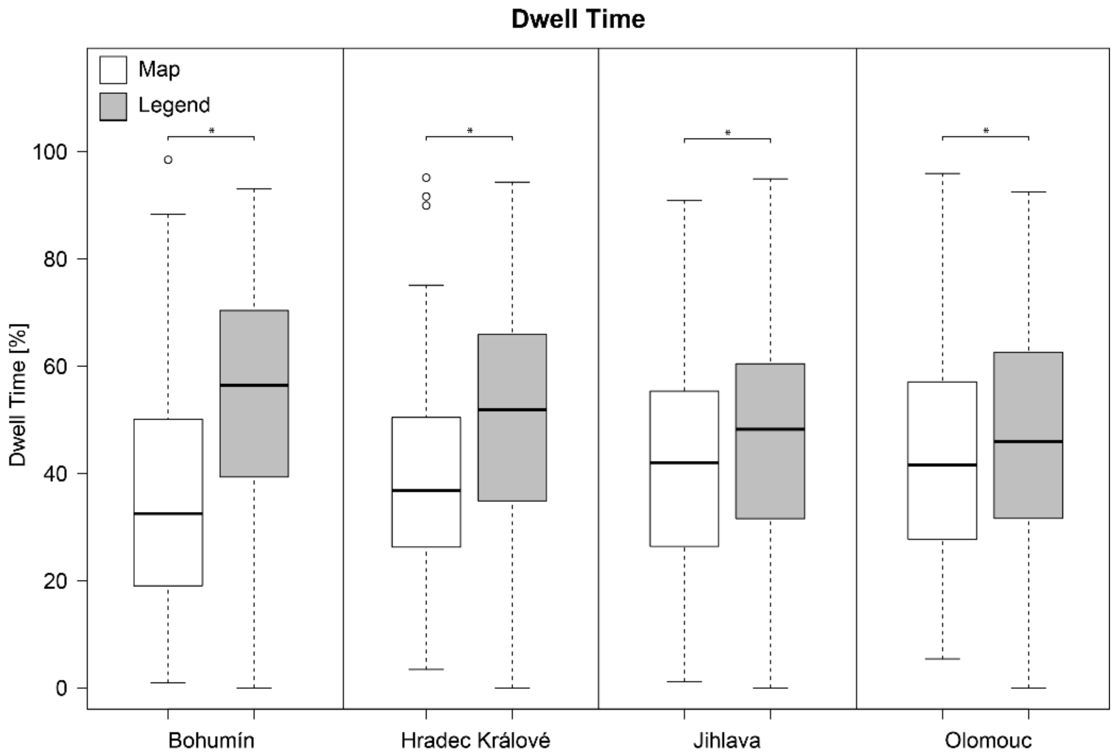


Figure 8. Boxplot of Dwell Time values for AOI around the Map and Legend. Statistically significant differences between maps and legends were found for all four cities.

Figure 9 shows that the legend was observed for a longer time than the map in Tasks Q2, Q3, and Q6. The largest difference was found in Q3, which focused on proposed public areas. To solve the task, participants had to consider this and search for the symbol for proposals in the legend. A similar situation can be seen in Q6, the correct map symbol being in the lower part of the legend, which influenced the time spent searching for it. Task Q2 focused on recreation areas, usually marked in a yellow or orange color. It took some time to find the symbol in the legend, but, due to the brightness of the color of recreation areas, it was easy to find them on the map. Recreation areas are often relatively large, so they were easily identifiable.

In Tasks Q1 and Q5, respondents spent more time on the map. Task Q1 was the clearest: respondents were required to find housing areas that were usually shown at the top of the legend. Task Q5 was influenced by an incorrect symbol in the Jihlava plan, and participants spent a lot of time on the map. Task Q4 was the most balanced, requiring respondents to find a railroad. This was the only task in which a statistically significant difference was not found.

In Figure 9, the ratio between time spent on the map and the legend is presented. Figure 10 displays the values of Dwell Time for the AOI "Legend" only. The results show that the most problematic search in the legend was in Tasks Q3 and Q6. High Dwell Time values for these tasks can be explained as the same as above—the proposal for Q3 and location of the symbol at the bottom of the legend for Task Q6.

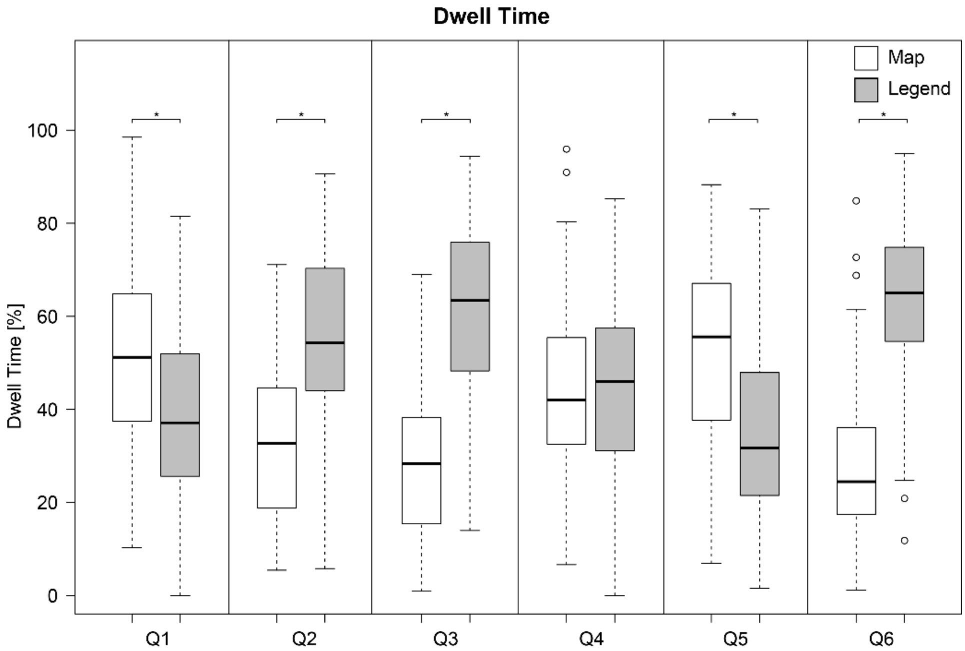


Figure 9. Boxplot of Dwell Time values for AOI around the Map and Legend. Statistically significant differences between the maps and legends were found for all tasks except Q4.

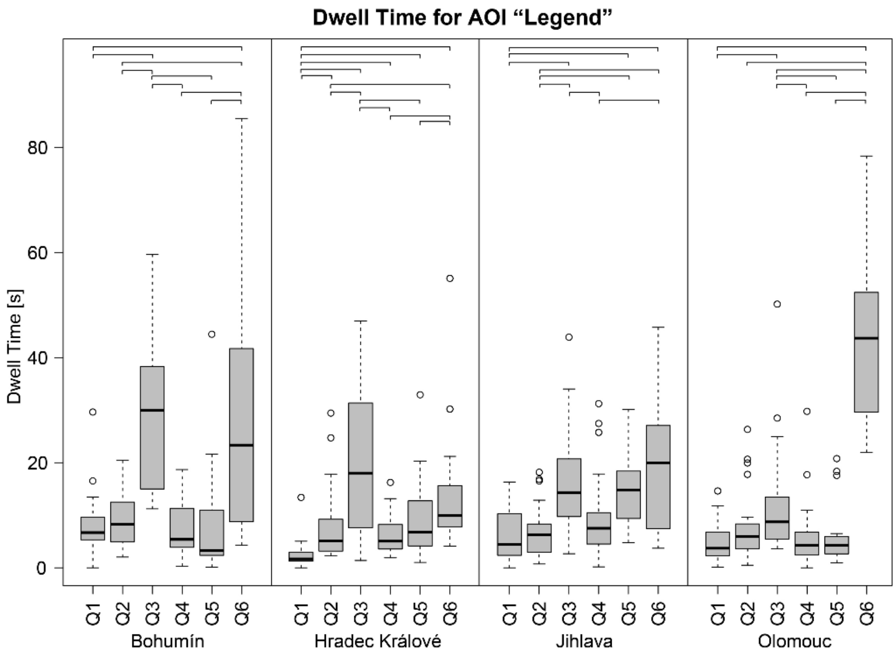


Figure 10. Boxplot of Dwell Time values in seconds for AOI "Legend".

The Trial Duration values for Tasks Q5 (Jihlava) and Q6 (Olomouc) were the highest (see Figure 4). Nevertheless, the reason for these tasks requiring a lot of time to solve is different in both cases, as can be seen in Figure 11. The recorded data were visualized using the Sequence Chart. In this chart, each row visualizes data for one participant looking at stimuli Q5 (Jihlava) and Q6 (Olomouc). The length of the color line corresponds to the time spent on the stimulus. Darker parts of the lines represent the time spent in the legend; brighter ones are associated with the map. In Q5 (Jihlava), it was easy to find the correct symbol in the legend, needing approximately ten seconds. Respondents then looked at the map and attempted to find the symbol. In some cases (i.e., respondents P09, P12, P22, P33, P36, and others) and after some time with the map, respondents looked back to the legend to verify the symbol for the wastewater treatment plant. In Task Q6 (Olomouc), the opposite pattern can be seen: respondents spent most of their time in the legend of the stimulus. After finding the symbol in the legend, they quickly marked it with a mouse click on the map.

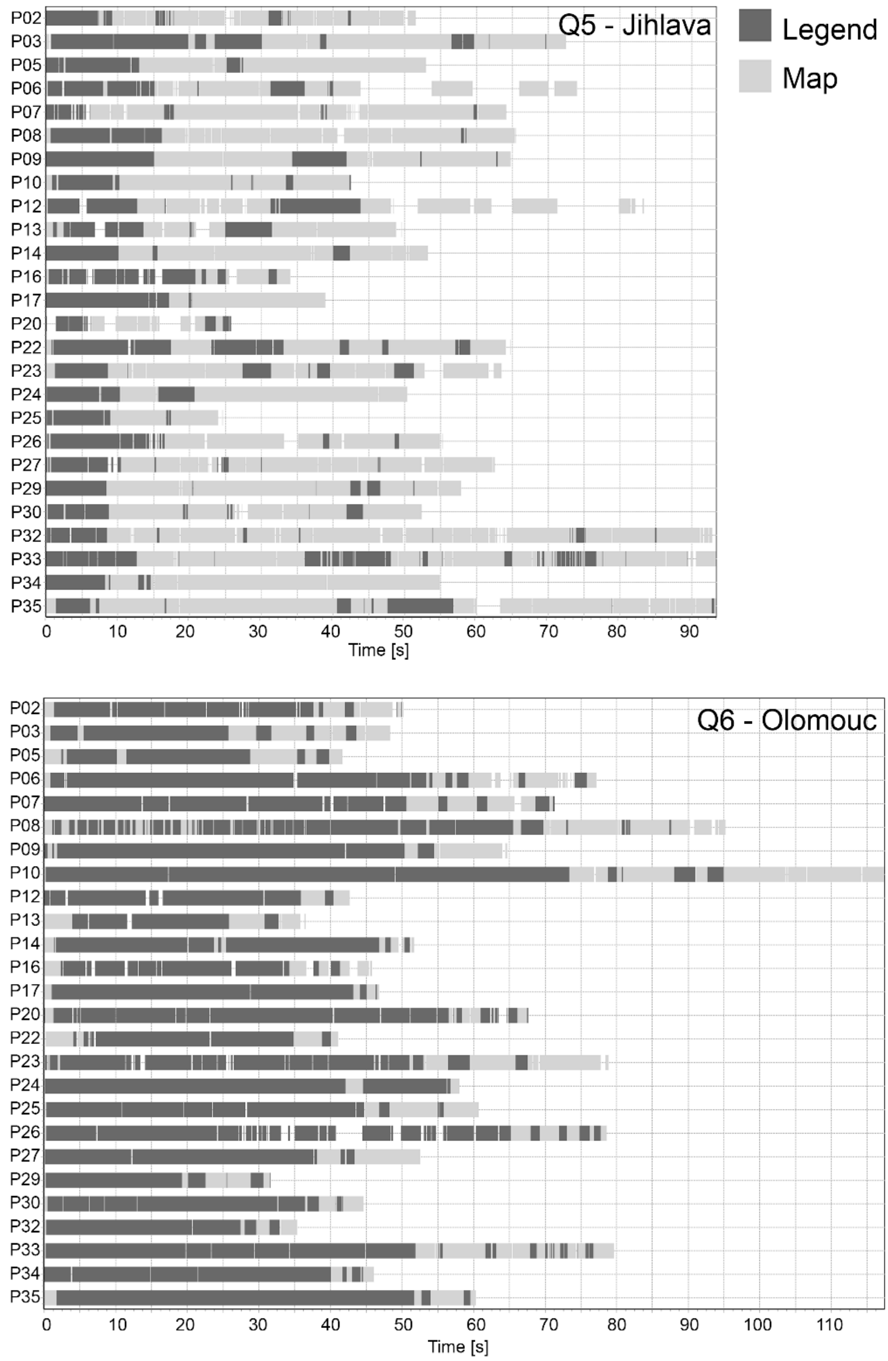


Figure 11. Sequence Chart showing the differences in participant strategy while solving different tasks. The two most complex tasks (Q5 (Jihlava) and Q6 (Olomouc)) were selected for comparison.

In addition to Sequence Chart, the data from the whole experiment were visualized using a FlowMap method in V-Analytics software [67]. The most illustrative results of a FlowMap method were obtained from the data recorded over the Task Q6 (Figure 12). As was already mentioned in the methods, FlowMap shows the aggregated eye-movements of all participants. The width of the arrows represents the number of gaze movements between Voronoi polygons in the stimulus. In accordance with previously described results, the most straightforward strategy of stimulus inspection was observed in the case of Hradec Králové. The symbol of a protected area of water resources was quite clear and distinctive, so the aggregated gaze trajectory displayed as a sequence of arrows leads directly to the correct place on the map. The legend was clearly structured, and participants did not need too many fixations for the finding of the proper symbol. On the other hand, the most complicated gaze trajectories were observed for Bohumín and Olomouc plans. The participants spent a lot of time in the legend trying to find the correct symbol in unstructured (Bohumín) or very comprehensive (Olomouc) legend. This can be observed in Figures 7 and 10 as well. In addition, the trajectories in the maps were leading to different places because the protected area of water resources was represented by the symbol that was difficult to distinguish (thin line similar to many other symbols).

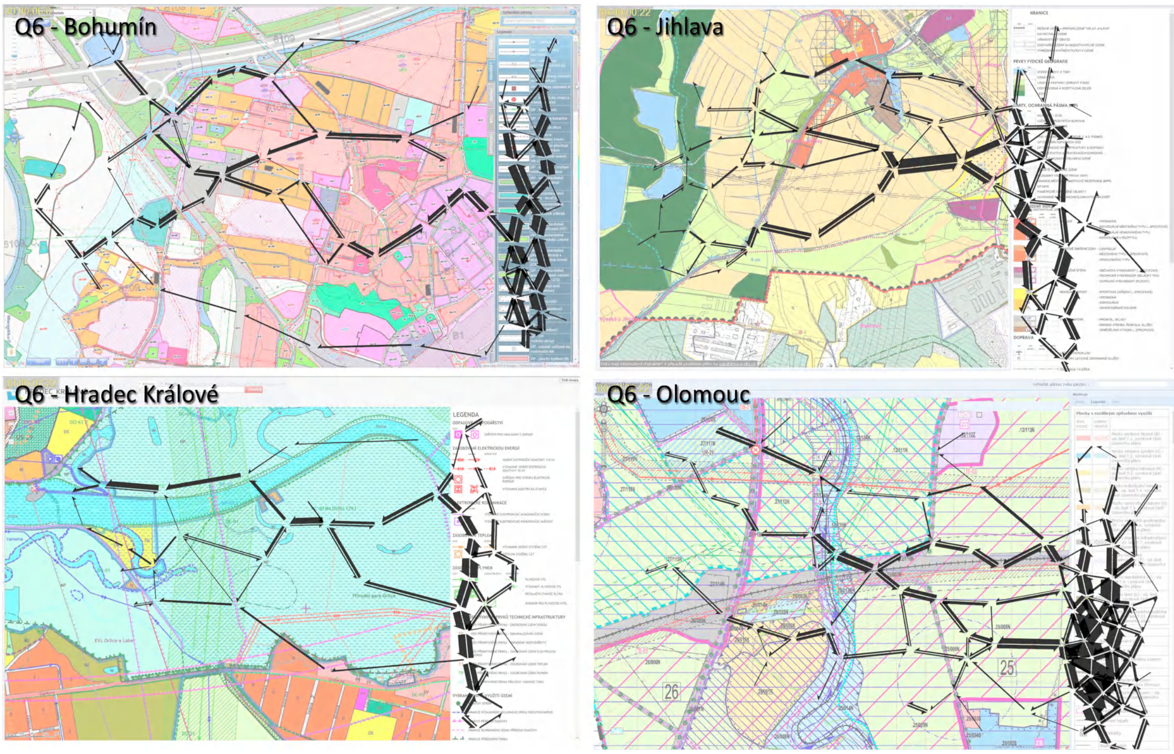


Figure 12. FlowMaps showing the aggregated moves of participants' gaze while Task Q6 solving. Only arrows representing more than three moves are displayed.

3.4. Differences between Experts and Students

In the following evaluation, the differences between the experts and students were investigated. The size of the group of experts was small (6), and statistically significant differences could have been influenced by this fact, which is why these results are presented as explanatory. The boxplot in Figure 13 shows Fixation Count values for cities (tasks are aggregated). The differences between experts and students were found for Jihlava and Olomouc. Experts needed fewer fixations to solve the tasks in these cases.

Figure 14 shows the Fixation Count metric values for the tasks (the cities are aggregated). The differences between students and experts for Tasks Q1, Q2, and Q3 are due to the commonness of these tasks. Tasks Q1, Q2, and Q3 focused on very common elements, and experts work with them much more frequently than with the elements in Tasks Q4, Q5, and Q6.

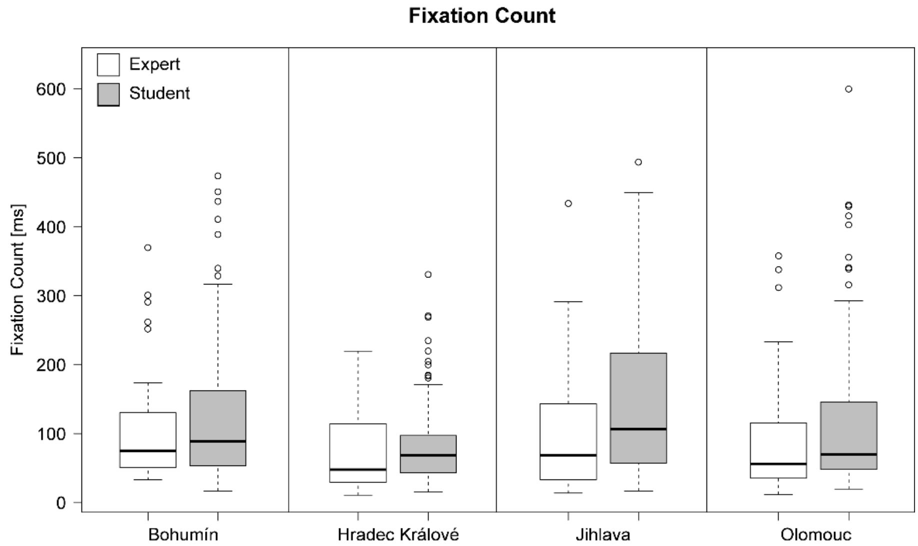


Figure 13. Boxplot of Fixation Count values. Differences between experts and students, especially for the Jihlava and the Olomouc plans, were found.

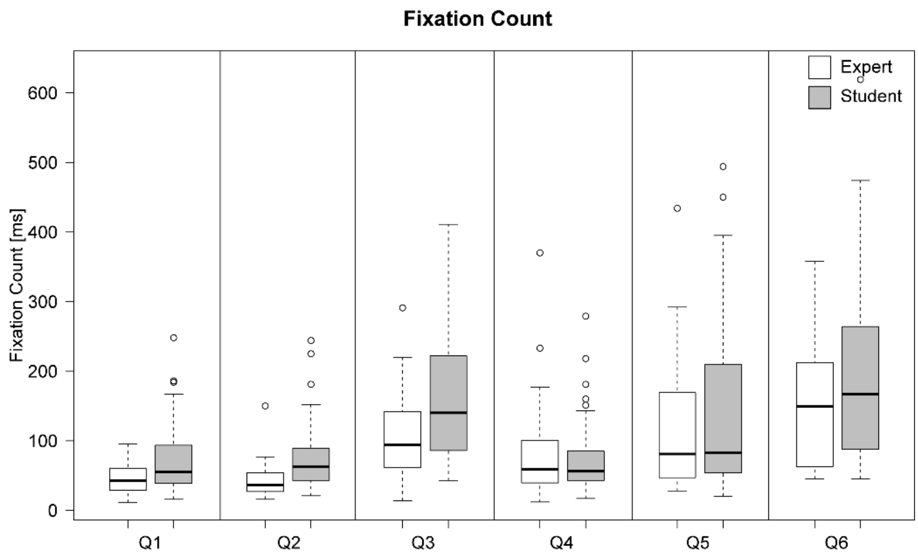


Figure 14. Boxplot of Fixation Count values. Differences between experts and students, especially for Tasks Q1, Q2, and Q3, were found. These tasks focused on frequently used map elements. Experts are more familiar with them, and they needed fewer fixations to solve the tasks.

3.5. Accuracy of Answers

In the next part of the data analysis, the accuracy of answers was investigated. The results are summarized in Figure 15. As mentioned above, the first two questions were easy and consisted of standard task solving while working with the plan. Only one incorrect answer (detected by mouse click coordinates) was recorded.

By contrast, solving Q3 was problematic in all four cities. Interestingly, of the six experts participating in the study, four of them were incorrect in three of the four cities. In this task, respondents were required to find the area for proposed public services (marked as a purple cross-hatched area). Four of the six experts marked the grey cross-hatched area instead (Figure 16: red diamonds represent mouse clicks by students, and black diamonds with a dot represent clicks by experts). This grey symbol corresponds to the proposed traffic infrastructure element. The experts were familiar with these symbols, explaining that, in their work, they deal primarily with technical or traffic infrastructure, which is usually marked with a grey symbol.

Another interesting result was found for Task Q4 (finding the railroad). For each city except Olomouc, all answers were correct. For Olomouc, three incorrect answers by experts were found. This is interesting, because all the experts work with the Olomouc plan every day. The problem was the black line symbol (at the bottom of Figure 17). Although this symbol can be mistaken for a railroad, in reality it represents a boundary between city districts. Experts were confident of their answers being correct and did not check the symbol in the legend.

In Task Q5 (Jihlava), in which respondents were required to find the wastewater treatment plant, the incidence of incorrect answers was highest. Ten students and three experts marked the incorrect area. Additionally, five participants were unable to mark anything, and they skipped this task. The symbol in the legend (the small, black ČOV caption) did not correspond to the symbol on the map (the large, red ÈOV caption) (Figure 18). The difference between captions was due to a coding error of diacritics in the Jihlava plan.

	Incorrect Answers											
	Bohumín			Hradec Králové			Jihlava			Olomouc		
	Students	Experts	Missing	Students	Experts	Missing	Students	Experts	Missing	Students	Experts	Missing
Q1	0	0	0	0	0	0	0	0	0	0	0	0
Q2	0	0	0	0	0	0	0	0	0	1	0	0
Q3	13	4	0	4	4	0	7	1	3	10	4	1
Q4	0	0	0	0	0	0	0	0	0	0	3	0
Q5	2	1	0	0	1	0	10	3	5	10	2	0
Q6	1	0	3	0	1	0	0	0	3	8	1	1

Figure 15. A summary of incorrect answers throughout the experiment.

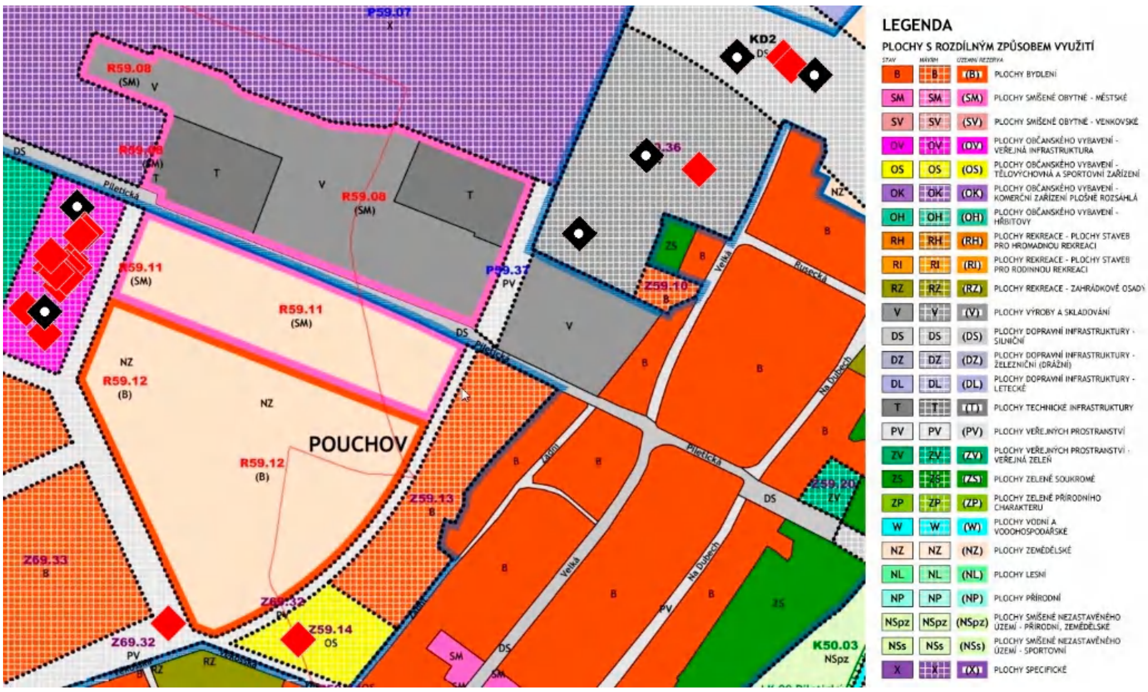


Figure 16. Student answers for Task Q3 (Hradec Králové) are marked with diamonds. Expert answers are marked with diamonds with a dot. The correct answer is the purple cross-hatched area on the left.

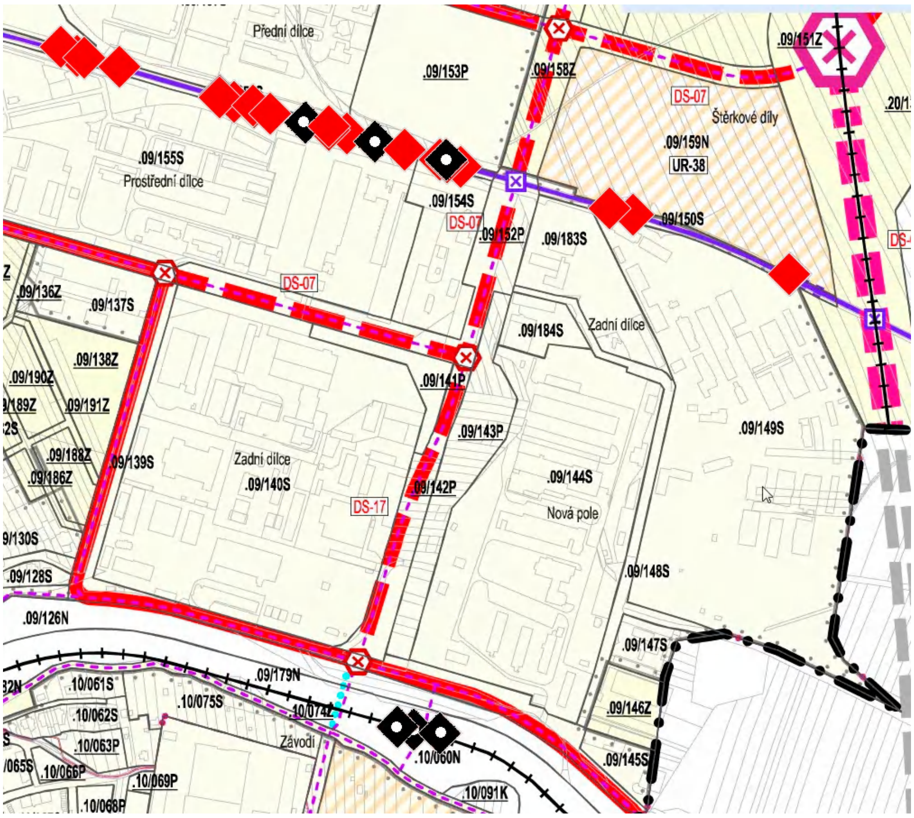


Figure 17. Student answers for Task Q4 (Olomouc) are marked with diamonds. Expert clicks are diamonds with a dot. The correct answer is the line at the top.

Figure 18. A clip from Task Q5 (Jihlava). The symbol in the legend (bottom) did not correspond to the symbol on the map.

3.6. Results of the Questionnaire

The final part of the evaluation addressed the questionnaire, in which respondents rated the legibility of the plans of the four cities according to their subjective opinion. The main aim of this part of the data analyses was to compare the objective results of eye-movement and Trial Duration analysis with the subjective attitude of the participants. The questionnaire was filled out immediately after the eye-tracking experiment, and respondents ranked all the plans on a five-point scale from 1 (best) to 5 (worst). The results correspond with objective eye-tracking measurements. Olomouc and Hradec Králové (average rank 2 and 2.04), which used correct, clear and structured legends with a low number of map symbols and colors, were ranked the highest, followed by Bohumín (average rank 2.54). Respondents ranked Jihlava as the worst plan with an average rank of 4.07 (Figure 19).

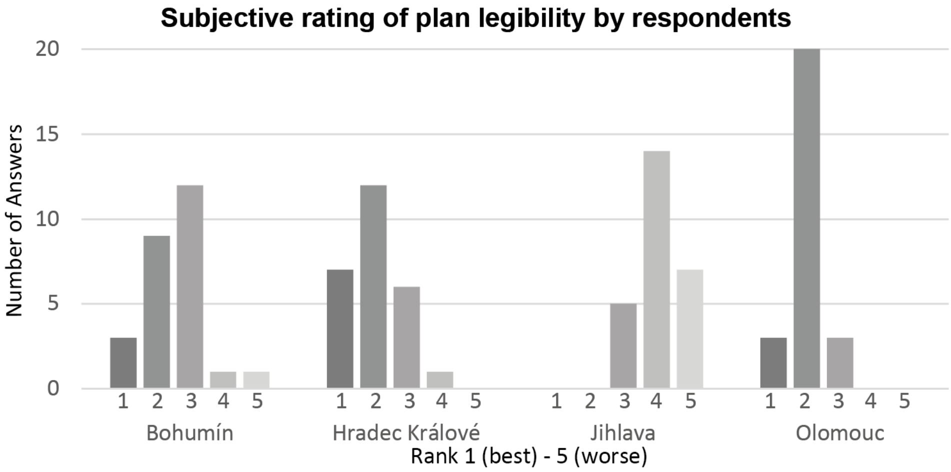


Figure 19. Results of the subjective questionnaire. The respondents ranked all plans on a scale from 1 (best) to 5 (worst).

4. Discussion

The Introduction mentions that most previous papers (e.g., [10,29,31]) have focused on the analysis or evaluation of urban plans, but not from a cartographical point of view. All authors use a qualitative method, which is sometimes very subjective, and its application to different maps from different regions can be difficult. The authors evaluate the plans of various types and scales (regional, country, and national). However, their evaluations are more widely focused than the research presented in this paper. In previous studies, cartographical aspects or urban plans were analyzed only very superficially (e.g., [13]), mostly in relation to standardization or spatial policy. For this reason, an additional value of this paper lies mainly in the application of objective research firstly focused on the quality of urban plans.

The problematic part of this study is the unequal and small representation of experts participating in the study. Although user groups with the same sample size would be better, the number of experts in the field of urban planning who might be willing to participate in the study is limited. In analyzing the work of experts in specific fields, it is common to use fewer participants. For example, Bianchetti's doctoral dissertation [70] investigates the cognitive tasks and fundamental visual stimuli used in the interpretation of aerial imagery. She analyzed the work of seven analysts in the domain of forest management. Kiefer et al. [71] included only five participants in their study using a mobile eye-tracking device. For this study, all employees of the urban planning departments in Olomouc were asked, and all agreed to being recorded. Many of them, though, were older, and recording eye-tracking was problematic because they had either drooping eyelids or wore glasses. Because of the different group sizes and the low number of experts, the part of the study comparing the two groups of participants can be considered explanatory, serving only as a brief insight into the differences between the behavior of experts and students.

The number of overall participants (26) in the experiment is not ideal, but it is in accordance with other eye-tracking studies. Alacam and Dalci [72] also used 26 participants in their study about finding map symbols, and Fuchs et al. [50] used 21 respondents in their study about flood maps. Many studies have also used a much smaller sample, for example, Ooms et al. [73] (14 participants), Opach and Nossum [74] (10 participants), etc. In the paper of Coltekin et al. [49], this issue is discussed, and authors concluded that it is quite difficult to recruit a large number of people with expertise in geovisualization and/or specific thematic domain (in their case soil maps, in our case urban planning). Similar to Coltekin et al., we argue that, despite the limitations regarding the sample size, we provide many interesting insights about user's interaction with different urban plans.

Statistical analysis of Trial Duration and Fixation Count was used in this study, as well as visualization using Sequence Charts and FlowMaps. Other methods exist that could also be used to analyze data. A comprehensive description of visualization methods was introduced by Blascheck [75]. Ambient and focal visual attention can be analyzed using methods proposed by Krejtz et al. [76], or similarities in stimuli reading strategy can be investigated with a tool proposed by Dolezalova and Popelka [77]. For the purposes of confirming the hypotheses, the employed methods were sufficient. Urban plans were analyzed in general. The aim was not to evaluate each symbol, but to find maps on which symbols were readable and understandable, which could be discovered from effectiveness (accuracy of answers) and efficiency (time to answer). To analyze symbols or symbology in detail, other analyses and visualization methods would be used.

Some of the results were influenced by a mistake in the Jihlava plan. The symbols in the legend did not match those on the map, and a problem also existed with diacritics. The study used four plans with different map designs. The study was created following a procedure of design, plan selection, task preparation, and stimuli preparation. In the stimuli preparation stage, the Jihlava plan using an incorrect map symbol was detected. This was not exceptional, as many mistakes of this type are in the Jihlava plan. This task was not eliminated because it permitted observation of a situation causing problems during task solving and provided evidence of the importance of creating correct sets of map symbols.

At the beginning of the study, three hypotheses dealing with map symbology, legend structure, and the differences between students and experts were proposed.

According to the first hypothesis, map symbology significantly influences the legibility and understandability of the plans, which will impact the duration and correctness of the tasks. According to the Trial Duration and Fixation Count, the worst plan is the Jihlava plan, the most time being needed to find an answer in Tasks Q4 and Q5. The problem encountered with this plan was the high number of map symbols in the legend and particularly the errors in those map symbols. In some cases, the symbol in the legend did not match the symbol used on the map. Another problem was with diacritics. The associativity of the map symbols was also low, and line symbols were too thick. The Hradec Králové plan can be considered the best plan, followed by the Olomouc plan. Both plans needed the least time in two of the six tasks (Q1 and Q6 for Hradec Králové and Q3 and Q5 for Olomouc), and also recorded the least number of fixations. The Hradec Králové plan contained the least number of map symbols as well as clearly distinguished colors as was for example illustrated on the Figure 12 showing output of FlowMap.

In addition to the eye-tracking experiments, the subjective questionnaire gathered respondents' opinions about the plans. The plan seen as worst was the Jihlava plan with incorrect symbols and a low level of legibility. As the best plan, respondents identified Olomouc and Hradec Králové. These results are consistent with the objective results from the eye-tracking data analysis.

According to the second hypothesis, legend structure significantly influences the legibility and understandability of the plans, which will impact the number of fixations and length of Dwell Time in the Legend AOI. Respondents spent statistically significantly more time in the legend than on the map. This can be interpreted as a result of low map legibility due to the low level of associativity of map symbols and the low quality of the map legend. Clearly designed maps with well-organized and correct legends and an adequate number of map symbols and colors would increase the legibility and understandability of the plan. The highest value of Dwell Time for Legend AOIs was observed for the Bohumín plan, which has an unstructured legend with a high number of symbols. In most cases, respondents spent more time in the legend than on the map. The exception is Task Q1 concerning housing areas. Symbols for this type of element are usually found in the upper part of the legend. Thus, respondents observed the legend for a statistically smaller proportion of time. Another explanation could be that all the plans used red for the housing category. This color is commonly used in Czech urban planning for this category, which was also expected by the respondents. Another exception is Task Q5, which was influenced by the long observation time on the Jihlava map due to an incorrect symbol being used. The most time spent in the Legend AOI was observed in Q3 and Q6. In Q3, respondents had to find proposed public areas. Proposed elements were in the second column of the legend, and it was more difficult to find the correct answer. In Q6, the map symbol used for the protected area of water resources was in the lower part of the legend on each plan, which prolonged the time needed to solve the task (see Figure 12).

The third hypothesis proposes that differences between students and experts will be seen in how these groups read plans, which will impact the duration and correctness of the tasks. The experts needed fewer fixations to solve the first three tasks in the study, which focused on common elements. Experts work with these elements daily. Greater differences between students and experts were found in the Olomouc and Jihlava plans. The experts were from the Olomouc region, so they may have been more familiar with the Olomouc plan. The difference in the Jihlava plan may have been due to its style. The Jihlava plan was from 1999 and older visualization styles might be better-known to experts than students. Some map symbols were very similar to symbols used in the methodologies described in the Introduction [23,24]. The surprising finding is that the proportion of incorrect answers was much higher for the group of experts. In many cases, they were accustomed to another map style (different colors, different map symbols) and were confident of their (incorrect) answer even though they did not check the symbols in the legend. This demonstrates that even experienced users can be misled by the incorrect use of map symbols. If the map symbols used in urban plans had been standardized, the experts' answers probably would not have been incorrect. This argument supports the importance of standardization in urban planning. Similar conclusions appear in Dühr's [4,27] examination of Dutch plans, which are not standardized. However, Dühr mentioned the high level of standardization and uniformity in the German planning system, which means that the established rules for cartographic representations are almost impossible to change.

5. Conclusions

No existing studies investigate either the cartographical quality of urban plans or the cognitive aspect of working with urban plans. Only one subjective study focuses on cartographic failures in urban plans [13]. For this reason, an objective eye-tracking experiment focusing on the analysis of four urban plans of cities in the Czech Republic was performed. The eye-tracking method is considered as objective, and with its use, it is possible to perform analyses that are not possible with any other method of evaluation. One example could involve an analysis of time spent on the map and the legend sections, along with eye-movement transitions between these two parts.

Four urban plans created by different authors, having different styles and published in different years were selected. To cover the most typical tasks involved in standard work with an urban plan, the analysis was based on six tasks. Twenty-six respondents (20 students and 6 experts working in urban-planning departments) participated in the study.

We conclude that two crucial factors influence the legibility of plans and significantly impact the understandability of maps. Those are the quality of map symbology (number of colors, the design of symbols, and features/layers on the map) and logical hierarchy/structure of the legend (number of symbols, legend size, legend structure, and legend order). Use of incorrect map symbol (similar but not the same) in the legend can cause a dramatical change in the duration and correctness of task solving.

Based on our study, we also conclude that the increasing quality of Czech urban plan symbology can be observed. Older plans (Jihlava and Bohumin) were designed based on ten-year-old symbology (many map symbols, similar colors, low level of symbols associativity, and line symbols were too thick).

On the opposite side, plans of Olomouc and Hradec Králové used newer symbology with a clearly structured legend, a low number of map symbols, and clearly distinct colors. According to the results of the eye-tracking data analysis, plans of Jihlava and Bohumín have lower cartographic quality (proved by longer Trial Duration, higher Fixation Count, longer Dwell Time in the Legend AOI and by more incorrect answers).

A similar conclusion follows from the respondents' average ranking of plans (Olomouc, 2; Hradec Králové, 2.04; Bohumín, 2.54; and Jihlava, 4.07). Plans that used correct, clear and structured legends with a small number of map symbols and colors were ranked much higher.

During the explanatory observation of the different behavior between experts and students, the accuracy of answers was found to be dependent on many factors, such as the position of the symbol in the legend, previous user experience, and self-confidence in the correct answer. Prior knowledge of more different urban plans symbology can lead to faster task solving but also to incorrect answers (if the knowledge of varying symbology is applied to another one). In the case of a complicated task or complicated legend (symbology), a group of experts is not faster than other users (students in this study).

To avoid misunderstandings, urban planners should be aware of quality issues in urban plans. For correct decision-making, it is essential to produce maps according to certain standards, make maps as clear as possible, and perform usability testing on maps. Standardization of urban plans, as the most complex thematic maps, should be in the focus of cartographers and urban planners more than today.

Supplementary Materials: Plans in higher resolution can be accessed via www.eyetracking.upol.cz/urban.

Author Contributions: J.B. was responsible for the literature review on urban planning issues and collaborated on the experiment's design. S.P. was responsible for execution of the eye-tracking experiment and analysis of recorded data. M.B. collaborated on data analysis. All authors cooperated in interpretation and discussion of the results.

Acknowledgments: This paper was supported by the Internal Grant Agency of Palacký University Olomouc (project IGA_PrF_2018_028—Innovative evaluation methods and advanced analysis of spatially based systems) and by the ERASMUS+ project no. 2016-1-CZ01-KA203-024040.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Monmonier, M. *How to Lie with Maps*, 2nd ed.; University of Chicago Press: Chicago, IL, USA, 1996; p. 207.
2. Churchill, R. Urban cartography and the mapping of Chicago. *Geogr. Rev.* **2004**, *94*, 1–22. [CrossRef]
3. Burian, J.; Brychtová, A.; Vávra, A. Proposal of urban plan symbology standardization: Olomouc region case study. In Proceedings of the 15th International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, Albena, Bulgaria, 16–25 June 2015; Volume 2, pp. 1051–1062.
4. Dühr, S. The form, style, and use of cartographic visualisations in european spatial planning: Examples from england and germany. *Environ. Plan. A* **2004**, *36*, 1961–1989. [CrossRef]
5. Jarvis, B. The arts of town planning. *Plan. Pract. Res.* **1995**, *10*, 111–120. [CrossRef]
6. Neuman, M. Images as institution builders: Metropolitan planning in madrid. *Eur. Plan. Stud.* **1996**, *4*, 293–312. [CrossRef]
7. Neuman, M. Communicate this! Does consensus lead to advocacy and pluralism? *J. Plan. Educ. Res.* **2000**, *19*, 343–350. [CrossRef]
8. Lussault, M. La ville clarifiée. Essai d'analyse de quelques usages carto- et iconographiques en œuvre dans le projet urbain [the clarified city: An analytical essay on the use and impact of carto- and iconographics in the urban development project]. In *La Cartographie en Débat, Représenter ou Convaincre*; Cambrézy, L., Maximy, R., Eds.; Karthala-Orstom: Paris, France, 1995; pp. 157–193.
9. Gabellini, P. *Il Disegno Urbanistico [The Urban Drawing]*; Nuova Italia Scientifica: Roma, Italy, 1996.
10. Söderström, O. *Des Images pour Agir: Le Visuel en Urbanisme*; Payot: Lausanne, Switzerland, 2000; p. 139.
11. Robinson, A.C.; Roth, R.E.; Blanford, J.; Pezanowski, S.; MacEachren, A.M. Developing map symbol standards through an iterative collaboration process. *Environ. Plan. B Plan. Des.* **2012**, *39*, 1034–1048. [CrossRef]
12. Funkhouser, H.G. *Historical Development of the Graphical Representation of Statistical Data*; Saint Catherine Press: London, UK, 1937; p. 404.
13. Burian, J.; Šťávová, Z. Kartografické a geoinformatické chyby v územních plánech [The difficulties of using urban plans for cartographers and geoinformatists]. *Geografie* **2009**, *114*, 179–191.
14. Infrastructure, M.O. Regulation of the Minister of Infrastructure [Rozporządzenie Ministra Infrastruktury z Dnia 26 Sierpnia 2003 r. W Sprawie Wymaganego Zakresu Projektu Miejscowego Planu Zagospodarowania Przestrzennego]. Available online: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwiOjtCsz4vbAhWID8AKHeD2DGoQFggqMAA&url=http%3A%2F%2Fdziennikustaw.gov.pl%2FD1995029015101.pdf&usg=AOvVaw0fFOoHWp8KmsSXTEggvLSb (accessed on 10 April 2018).

15. Pál, S. Different cartographic visualizations of spatial structure in the documents of hungarian spatial policies. In Proceedings of the RSA 2010 Annual International Conference, Regional Responses to Global Shifts: Actors, Institutions and Organisations, Pécs, Hungary, 24–26 May 2010; p. 15.
16. Norvay, M.O.E.O. Reguleringsplan Bebyggelsesplan. Available online: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjQn9av0IvbAhVFF8AKHX0cD2EQFggnMAA&url=https%3A%2F%2Fwww.arkitektur.no%2F%3Fnid%3D157760%26lcid%3D1044%26pid%3DNAL-EcoPublication-Attachment&usg=AOvVaw0qYIQZ5raMczDSavJ1Ipzh (accessed on 10 April 2018).
17. Burian, J.; Brychtová, A.; Vávra, A.; Hladišová, B. Analytical material for planning in olomouc, Czech Republic. *J. Maps* **2016**, *12*, 649–654. [CrossRef]
18. Burian, J.; Brus, J.; Voženílek, V. Development of olomouc city in 1930–2009: Based on analysis of functional areas. *J. Maps* **2013**, *9*, 64–67. [CrossRef]
19. Ministry of Regional Development of the Czechoslovakia. *Act No.50/1976 Coll. on town & country planning and on building regulations (the Building Act)*; Ministry of Regional Development of the Czechoslovakia: Prague, Czechoslovakia, 1976.
20. Ministry of Regional Development of the Czech. *Uniform Standard of Map Symbology for a Main Regional Municipality Plan and the Regulatory Plan*; Ministry of Regional Development of the Czech: Prague, Czech Republic, 1999; p. 12.
21. *Methodology of a Digital Processing of a Regional Municipality Plan for GIS in Public Administration for a District Council*; Hydrosoft Veleslavín: Prague, Czech Republic, 2000; p. 52.
22. *The Minimal Standard for Digital Processing of Regional Municipality and Town Plans in GIS*; Hydrosoft Veleslavín: Prague, Czech Republic, 2005; p. 52.
23. T-MAPY. *Methodology of Digital Processing of Rpd for Gis*; T-MAPY: Hradec Králové, Czech Republic, 2001.
24. T-MAPY. *Integrated Approach for Digital Processing of Regional Municipality Plan for Gis*; T-MAPY: Hradec Králové, Czech Republic, 2004.
25. Štěrba, Z.; Šašinka, Č.; Stachoň, Z.; Štampach, R.; Morong, K. *Selected Issues of Experimental Testing in Cartography*; Masarykova Univerzita: Brno, Czech Republic, 2015.
26. Rohrer, C. When to Use Which User-Experience Research Methods. Available online: https://www.nngroup.com/articles/which-ux-research-methods/ (accessed on 10 April 2018).
27. Dühr, S. *The Visual Language of Spatial Planning*; Routledge: London, UK, 2007; p. 224.
28. Harley, J.B. Deconstruction the map. *Cartogr. Int. J. Geogr. Inf. Geovis.* **1989**, *26*, 1–20.
29. Pickles, J. Texts, hermeneutics and propaganda maps. In *Writing Worlds. Discourse, Text and Metaphor in the Representation of Landscape*; Barnes, J.T., Duncan, J.S., Eds.; Routledge: London, UK, 1992; pp. 193–230.
30. Söderström, O. Paper cities: Visual thinking in urban planning. *Cult. Geogr.* **1996**, *3*, 249–281. [CrossRef]
31. Dühr, S. Analysing cartographic representations in spatial planning. In *The Routledge Handbook of Planning Research Methods*; Silva, E.A., Healey, P., Harris, N., Eds.; Routledge, Taylor & Francis Group: London, UK, 2015; pp. 192–201.
32. Tang, X.; Hurni, L. Regional spatial planning maps: A sino-swiss comparison of cartographic visualization methodologies. In Proceedings of the 24th International Cartographic Conferenc—The World's GeoSpatial Solutions, Santiago, Chile, 15–21 November 2009; International Cartographic Association: Vienna, Austria, 2009; p. 14.
33. Dühr, S. Spatial policies for regional sustainable development: A comparison of graphic and textual representations in regional plans in england and germany. *Reg. Stud.* **2005**, *39*, 1167–1182. [CrossRef]
34. Goldberg, J.H.; Kotval, X.P. Computer interface evaluation using eye movements: Methods and constructs. *Int. J. Ind. Ergon.* **1999**, *24*, 631–645. [CrossRef]
35. Steinke, T.R. Eye movement studies in cartography and related fields. *Cartogr. Int. J. Geogr. Inf. Geovis.* **1987**, *24*, 40–73. [CrossRef]
36. Brychtova, A.; Popelka, S.; Dobesova, Z. Eye—Tracking methods for investigation of cartographic principles. In Proceedings of the 12th International Multidisciplinary Scientific GeoConference and EXPO, Varna, Bulgaria, 17–23 June 2012; Volume 2, pp. 1041–1048.
37. Coltekin, A.; Heil, B.; Garlandini, S.; Fabrikant, S.I. Evaluating the effectiveness of interactive map interface designs: A case study integrating usability metrics with eye-movement analysis. *Cartogr. Geogr. Inf. Sci.* **2009**, *36*, 5–17. [CrossRef]

38. Fabrikant, S.I.; Rebich-Hespanha, S.; Andrienko, N.; Andrienko, G.; Montello, D.R. Novel method to measure inference affordance in static small-multiple map displays representing dynamic processes. *Cartogr. J.* **2008**, *45*, 201–215. [CrossRef]
39. Brychtova, A.; Popelka, S.; Vozenilek, V. The analysis of eye movement as a tool for evaluation of maps and graphical outputs from gis. In Proceedings of the 19th International Conference on Geography and Geoinformatics: Challenge for Practise and Education; Svobodova, H., Ed.; Masarykova Univerzita: Brno, Czech Republic, 2012; pp. 154–162.
40. Ooms, K.; De Maeyer, P.; Fack, V. Study of the attentive behavior of novice and expert map users using eye tracking. *Cartogr. Geogr. Inf. Sci.* **2014**, *41*, 37–54. [CrossRef]
41. Kubíček, P.; Šašinka, Č.; Stachoň, Z. Uncertainty visualization testing. In Proceedings of the 4th Conference on Cartography and GIS; Bandrova, T., Konečný, M., Zhelezov, G., Eds.; Bulgarian Cartographic Association: Sofia, Bulgaria, 2012; pp. 247–256.
42. Kubíček, P.; Šašinka, Č.; Stachoň, Z.; Štěrba, Z.; Apeltauer, J.; Urbánek, T. Cartographic design and usability of visual variables for linear features. *Cartogr. J.* **2017**, *54*, 91–102. [CrossRef]
43. Popelka, S.; Brychtova, A. Eye-tracking study on different perception of 2d and 3d terrain visualisation. *Cartogr. J.* **2013**, *50*, 240–246. [CrossRef]
44. Popelka, S.; Dedkova, P. Extinct village 3d visualization and its evaluation with eye-movement recording. In *Computational Science and Its Applications—ICCSA 2014*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: New York, NY, USA, 2014; Volume 8579 LNCS, pp. 786–795.
45. Brychtova, A.; Coltekin, A. Discriminating classes of sequential and qualitative colour schemes. *Int. J. Cartogr.* **2015**, *1*, 62–78. [CrossRef]
46. Taylor, R. Information theory and map evaluation. *Int. Yearb. Cartogr.* **1975**, *15*, 165–180.
47. Garlandini, S.; Fabrikant, S.I. Evaluating the effectiveness and efficiency of visual variables for geographic information visualization. In *Spatial Information Theory*; Springer: New York, NY, USA, 2009; pp. 195–211.
48. Petchenik, B.B. Cognition in cartography. *Cartogr. Int. J. Geogr. Inf. Geovis.* **1977**, *14*, 117–128. [CrossRef]
49. Çöltekin, A.; Brychtová, A.; Griffin, A.L.; Robinson, A.C.; Imhof, M.; Pettit, C. Perceptual complexity of soil-landscape maps: A user evaluation of color organization in legend designs using eye tracking. *Int. J. Dig. Earth* **2017**, *10*, 560–581. [CrossRef]
50. Fuchs, S.; Spachinger, K.; Dorner, W.; Rochman, J.; Serrhini, K. Evaluating cartographic design in flood risk mapping. *Environ. Hazards* **2009**, *8*, 52–70. [CrossRef]
51. Dupont, L.; Antrop, M.; Van Eetvelde, V. Eye-tracking analysis in landscape perception research: Influence of photograph properties and landscape characteristics. *Landsc. Res.* **2013**, *39*, 417–432. [CrossRef]
52. Kim, M.; Kang, Y.; Bakar, S.A. A nightscape preference study using eye movement analysis. *Int. J. Sustain. Trop. Des. Res. Pract.* **2013**, *6*, 85–99.
53. Noland, R.B.; Weiner, M.D.; Gao, D.; Cook, M.P.; Nelessen, A. Eye-tracking technology, visual preference surveys, and urban design: Preliminary evidence of an effective methodology. *J. Urban. Int. Res. Placemak. Urban Sustain.* **2016**, *10*, 98–110. [CrossRef]
54. Act No. 183/2006 Coll., Construction Act. Available online: https://www.cka.cz/en/chamber/legislation/building_act/building_act_aj (accessed on 10 April 2018).
55. Beconyte, G. Cartographic styles: Criteria and parameters. In Proceedings of the 25th International Cartographic Conference (ICC 2011), Paris, France, 3–8 July 2011; pp. 3–8.
56. Benová, A.; Pravda, J. Map style. In *Cartography and Art*; Springer: New York, NY, USA, 2009; pp. 1–10.
57. Christophe, S. Cartographic styles between traditional and original (towards a cartographic style model). In Proceedings of the AutoCarto Conference, Columbus, OH, USA, 16–18 September 2012; pp. 16–18.
58. Roth, R.E. Cartographic interaction primitives: Framework and synthesis. *Cartogr. J.* **2012**, *49*, 376–395. [CrossRef]
59. Wehrend, S.; Lewis, C. A problem-oriented classification of visualization techniques. In Proceedings of the 1st Conference on Visualization'90, San Francisco, CA, USA, 23–26 October 1990; IEEE Computer Society Press: Washington, DC, USA, 1990; pp. 139–143.
60. Holmqvist, K.; Nyström, M.; Andersson, R.; Dewhurst, R.; Jarodzka, H.; Van de Weijer, J. *Eye Tracking: A Comprehensive Guide to Methods and Measures*; Oxford University Press: Oxford, UK, 2011; p. 537.

61. Hammoud, R.I.; Mulligan, J.B. Introduction to eye monitoring. In *Passive Eye Monitoring*; Springer: New York, NY, USA, 2008; pp. 1–19.
62. Salvucci, D.D.; Goldberg, J.H. Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA 2000), Palm Beach Gardens, FL, USA, 6–8 November 2000; ACM: Palm Beach Gardens, FL, USA, 2000; pp. 71–78.
63. Popelka, S. Optimal eye fixation detection settings for cartographic purposes. In Proceedings of the 14th SGEM GeoConference on Informatics, Geoinformatics and Remote Sensing, Albena, Bulgaria, 17–26 June 2014; Volume 1, pp. 705–712.
64. Mann, H.B.; Whitney, D.R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **1947**, *18*, 50–60. [CrossRef]
65. Kruskal, W.H.; Wallis, W.A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **1952**, *47*, 583–621. [CrossRef]
66. Popelka, S. Decision scheme for the selection of the most appropriate method for analyses and visualization of eye-tracking data in cartographic research. In Proceedings of the 15th International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, Albena, Bulgaria, 16–25 June 2015; Volume 2, pp. 803–810.
67. Andrienko, N.; Andrienko, G. *Exploratory Analysis of Spatial and Temporal Data*; Springer: Berlin, Germany, 2006; p. 704.
68. Andrienko, G.; Andrienko, N.; Burch, M.; Weiskopf, D. Visual analytics methodology for eye movement studies. *IEEE Trans. Vis. Comput. Graph.* **2012**, *18*, 2889–2898. [CrossRef] [PubMed]
69. Mello-Thoms, C.; Nodine, C.F.; Kundel, H.L. What attracts the eye to the location of missed and reported breast cancers? In Proceedings of the 2002 Symposium on Eye Tracking Research & Applications, New Orleans, LA, USA, 25–27 March 2002; ACM: New York, NY, USA, 2002; pp. 111–117.
70. Bianchetti, R.A. *Looking Back to Inform the Future: The Role of Cognition in Forest Disturbance Characterization from Remote Sensing Imagery*; The Pennsylvania State University: State College, PA, USA, 2014.
71. Kiefer, P.; Straub, F.; Raubal, M. Location-aware mobile eye tracking for the explanation of wayfinding behavior. In Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, France, 24–27 April 2012; p. 5.
72. Alacam, Ö.; Dalci, M. Alacam, Ö.; Dalci, M. A usability study of webmaps with eye tracking tool: The effects of iconic representation of information. In *New Trends in Human-Computer Interaction*; Springer: New York, NY, USA, 2009; pp. 12–21.
73. Ooms, K.; De Maeyer, P.; Fack, V. Analyzing eye movement patterns to improve map design. In Proceedings of the 18th International Research Symposium on Computer-Based Cartography and GIScience (AutoCarto 2010): Geospatial Data and Geovisualization for the Environments, Security, and Society, Orlando, FL, USA, 15–18 November 2010.
74. Opach, T.; Nossum, A. Evaluating the usability of cartographic animations with eye-movement analysis. In Proceedings of the 25th International Cartographic Conference, Paris, France, 3–8 July 2011; pp. 1–11.
75. Blascheck, T.; Kurzhals, K.; Raschke, M.; Burch, M.; Weiskopf, D.; Ertl, T. State-of-the-art of visualization for eye tracking data. In *EuroVis*; The Eurographics Association: Swansea, UK, 2014; Volume 2014, pp. 1–20.
76. Krejtz, K.; Coltekin, A.; Duchowski, A.; Niedzielska, A. Using coefficient k to distinguish ambient/focal visual attention during map viewing. *J. Eye Mov. Res.* **2017**, *10*. [CrossRef]
77. Dolezalova, J.; Popelka, S. Scangraph: A novel scanpath comparison method using visualisation of graph cliques. *J. Eye Mov. Res.* **2016**, *9*, 1–13.

Paper WeatherMaps

Popelka, S., Vondrakova, A.*, & Hujnakova, P. (2019). Eye-Tracking Evaluation of Weather Web Maps ISPRS International Journal of Geo-Information, 8(6), 29. https://doi.org/10.3390/ijgi8060256

[IF 2.239; Q3]

[22 citations on WoS; 19 citations without autocitations]

[Author's contribution: 34%]

Abstract

Weather is one of the things that interest almost everyone. Weather maps are therefore widely used and many users use them in everyday life. To identify the potential usability problems of weather web maps, the presented research was conducted. Five weather maps were selected for an eye-tracking experiment based on the results of an online questionnaire: DarkSky, In-Poasi, Windy, YR.no, and Wundermap. The experiment was conducted with 34 respondents and consisted of introductory, dynamic, and static sections. A qualitative and quantitative analysis of recorded data was performed together with a think-aloud protocol. The main part of the paper describes the results of the eye-tracking experiment and the implemented research, which identify the strengths and weaknesses of the evaluated weather web maps and point out the differences between strategies in using maps by the respondents. The results include findings such as the following: users worked with web maps in the simplest form and they did not look for hidden functions in the menu or attempt to find any advanced functionality; if expandable control panels were available, the respondents only looked at them after they had examined other elements; map interactivity was not an obstacle unless it contained too much information or options to choose from; searching was quicker in static menus that respondents did not have to switch on or off; the graphic design significantly influenced respondents and their work with the web maps. The results of the work may be useful for further scientific research on weather web maps and related user issues.

International Journal of
Geo-Information

MDPI

Article

Eye-tracking Evaluation of Weather Web Maps

Stanislav Popelka, Alena Vondrakova * and Petra Hujnakova

Department of Geoinformatics, Faculty of Science, Palacký University Olomouc, 17. listopadu 50, 77146 Olomouc, Czech Republic; stanislav.popelka@upol.cz (S.P.); petra.hujnakova@gmail.com (P.H.)
* Correspondence: alena.vondrakova@upol.cz; Tel.: +420585634517

Received: 30 November 2018; Accepted: 28 May 2019; Published: 30 May 2019

check for updates

Abstract: Weather is one of the things that interest almost everyone. Weather maps are therefore widely used and many users use them in everyday life. To identify the potential usability problems of weather web maps, the presented research was conducted. Five weather maps were selected for an eye-tracking experiment based on the results of an online questionnaire: DarkSky, In-Počasí, Windy, YR.no, and Wundermap. The experiment was conducted with 34 respondents and consisted of introductory, dynamic, and static sections. A qualitative and quantitative analysis of recorded data was performed together with a think-aloud protocol. The main part of the paper describes the results of the eye-tracking experiment and the implemented research, which identify the strengths and weaknesses of the evaluated weather web maps and point out the differences between strategies in using maps by the respondents. The results include findings such as the following: users worked with web maps in the simplest form and they did not look for hidden functions in the menu or attempt to find any advanced functionality; if expandable control panels were available, the respondents only looked at them after they had examined other elements; map interactivity was not an obstacle unless it contained too much information or options to choose from; searching was quicker in static menus that respondents did not have to switch on or off; the graphic design significantly influenced respondents and their work with the web maps. The results of the work may be useful for further scientific research on weather web maps and related user issues.

Keywords: web maps; cartography; user issues; eye-tracking technology; weather; meteorology

1. Introduction

Maps have been popular for centuries, moreover, crafted for several millennia. With the development of technologies in the twentieth century, digital forms have become popular. At the beginning of the twenty-first century, a variety of web applications have gradually become modern trends and a regular part of everyday life. Web maps as another form of cartographic work have become popular [1,2].

Many people see the Internet as a revolution for cartography because of new approaches and new technologies. While previously published maps were tied to a paper medium and expensive large-format colour print technology and had limited distribution and use, the Internet has made it possible not only to distribute maps to a much larger audience but also to incorporate interaction and animation [3–5]. These maps are becoming progressively more suitable, as some traffic and weather maps are updated every few minutes [3].

Numerous web map studies have been performed. Research topics have varied from theoretical foundations to purely applied studies: how web maps provide users with information [6], how the use of web-based maps could be made easier for users [7], what problems are associated with web map design [8], what the usability problems are, and others [9–15]. In the conclusions of those studies, problems related to the map field are often mentioned. One of the most significant conclusions that can be made is that a large number of web maps have a small map field and unnecessarily large web map controls or legends.

Many modern studies use the eye-tracking technology. Eye-tracking is used for usability tests, evaluation of interactive map interfaces [16], evaluation of animated maps of traffic flows [17], analyses of maps and plans [18], analyses of 3D geovisualizations [19], analyses of dynamic stimuli [20–24], and in general to evaluate the process of map reading and map use [25]. User studies in cartography provide enormous opportunities for further development of maps [26].

The research presented below focuses on the question of what problems users have with some of the most commonly used thematic maps on the Internet: weather web maps.

1.1. Weather Maps

Maps containing meteorological characteristics and meteorological data, generally known as "weather maps", were selected for the study from a wide range of available web maps, the main reason for this choice being that these maps are considered complex visual displays [23] and are one of the most used thematic web maps.

The term "weather maps" loosely refers to any cartographic depiction of weather or weather-related phenomena, including climatic maps. There are four main types of weather web maps [24]. Maps in the first three categories (weather maps in a narrower sense, satellite image maps, and radar image maps) describe recent or current conditions in the atmosphere. The fourth type consists of maps that predict future conditions. The development of web maps has been enormous, and most weather web maps now combine all of these types.

Web maps may be static or dynamic. Each of these categories is further subdivided into view only and interactive maps [4]. The current trend is to use animation to show the natural development of weather. A method called semi-static animation was introduced by Nossum [27] using an example of temperature forecast for four days. The concept's core was to make all information visually available to the user at any given time of the animation. The bottlenecks encountered in animated (web) maps are no longer because of hardware, software or data, but in the limited visual and cognitive processing skills of the map reader [28].

Semi-static and dynamic animations have been analysed in numerous studies [27,29]. Analysis of eye-tracking data revealed that the viewing behaviour of respondents for both map types (animation and semi-static animation) are surprisingly similar [30]. Weather maps have also been the subject of studies focusing other issues, such as map reading and visual salience [23], communication issues concerning climate forecasts [31], and methods of cartographic visualization [32].

1.2. Geovisualization Methods

There is great variability in the range of maps produced, not only because of the potential availability of technologies. The differences are in the specifics of the presented phenomena, the chosen methods of cartographic expression, graphic design, and many other aspects.

The suitability of selected methods of spatial data visualization and particular implementations significantly affect a user's ability to determine the correct information from a map. The quantification and evaluation of different factors affecting how information in a map is perceived by different user groups is the main task in many types of research [33–35]. Addressing modern trends in cognitive cartography and cartographic visualization methods can lead to insights in, and improvement to, cartographic production.

Methods of geovisualization represent a set of rules to express the spatial characteristics in a map. Methods of geovisualization are also described as cartographic visualization methods, methods of representation, means of expression, interpretive methods, graphical representation, mapping expression, and others. In most cases, they are not universally standardized, and the methods of geovisualization used generally depend on the personality and expertise of the author creating the visualization. Although there are a number of textbooks that describe map creation, approaches vary and, thus, the map designs of individual cartographers also vary [36]. However, the approach is uniform in the evaluation of map symbology through visual variables, which describe the graphic dimensions across which a map or other visualization is varied to encode information [37]. The methods of cartographic visualization on weather web maps are analyzed and described in the selected weather web map sub-descriptions.

The spatial data visualization process (geovisualization) can result in different levels of processing of the final visualized product, from a simple data view (graphical representation of spatial data layers) to a map (cartographic visualization with all the features and compositional elements). While maps are produced for different target groups for different purposes and present many different topics, the approaches to geovisualization are also changing in many aspects [38]. One possible aspect is time. The preferred methods of geovisualization may differ in the various age groups of map-makers as well as age groups of target users. Differences also exist in national approaches and various cartographic schools. Nevertheless, the selection of appropriate geovisualization methods and appropriate parameters of each method is the main task for a person with an education in cartography.

To ensure the correct map communication goals, user testing should be conducted, complex, non-technological aspects, user and usability issues can be addressed and evaluated during the map production analysis.

There are fewer methods of cartographic visualization which express meteorological elements. A basic point method is used to visualize stations, measurement points or other point-located variables. Area symbols are also often used to visualize presented phenomena, because meteorological indicators have the characteristic of being continuous data and are mostly presented as continuous surfaces. The most used methods include isolines, graduated symbols (diagrams), points, line symbols, and areas (area patterns) [39].

The composition of weather web maps varies greatly. Some authors see an advantage in the most uncomplicated map composition with the most basic controls so that the map is not overcrowded with information or options and that users can work and control it as simply as possible, other authors attach more importance to very interactive maps with a complex composition and large number of controls and visualization options [4,5,16,40].

The primary objective of the study presented in this article is to analyse how web-based weather maps are used and perceived. Only design aspects were evaluated, not the accuracy of the predictive model, rate of data update or other aspects. A more detailed analysis of weather web map functionality and interactivity may be considered for future work. The results of the presented research may be useful to ordinary users or for further scientific research on web maps and related user issues.

2. Materials and Methods

2.1. Materials – Description of Selected Web Maps

Many web portals contain web maps with meteorological content. These include Windy, In-Počasí, DarkSky, Wundermap, YR.no, PovodnovyPlan, Meteoearth, Ventusky, Meteoblue, Rainviewer, Weather, and many others. Web maps, portals, and weather apps take different forms. For the most part, they have very similar content; namely the visualization of meteorological phenomena. Some maps contain a large number of thematic layers and some contain only basic meteorological indicators, such as temperature, precipitation, wind or frontal systems.

For the eye-tracking experiment, five web weather maps were selected. No study was found with a complete comparison of weather maps, even though weather web maps are widely used by the public almost every day. Their use is not limited by previous knowledge or expertise and they can be used by almost anyone. The evaluated maps were: DarkSky (https://darksky.net/), Windy (https://www.windy.com), In-Počasí (https://www.in-pocasi.cz), YR.no (https://www.yr.no/kart), and Wundermap (https://www.wunderground.com/wundermap). This number was selected so that the time required for the entire experiment did not exceed 30 minutes and to minimize the fatigue and disorientation experienced by respondents.

The set of the evaluated web maps was designed so that the selected weather web maps included both foreign and Czech web-based weather maps and maps both known and unknown to the public. Our laboratory conducted an online survey to gauge the familiarity of a Czech audience to a range of web-based weather maps, both of Czech and foreign origin. Of 140 respondents, 34% indicated that they commonly use weather maps. The three most frequently indicated weather web maps were YR.no, In-Počasí, and Windy. Due to their differences in visualization methods, the less commonly used maps, DarkSky and Wundermap, were also selected for testing. These maps are not all representatives of the different types of maps, but as described, they are the three most frequently used maps (according to online survey) and two maps that the respondents know but do not ordinarily use.

2.1.1. Dark Sky

DarkSky (Figure 1) is a start-up established in 2011 in England and includes web and mobile weather forecast applications for the world. The map is directly loaded, and the composition is organized into horizontal blocks. A search field and basic information about the location the user is looking for is at the top, with the current meteorological indicator values and a timeline. The next block is a map with additional graphs and temperature forecasts for the coming days. Switching thematic layers is easily accessed in the popup menu. The application contains data about temperature, wind speed and direction, clouds, precipitation, dew point, UV index, ozone, and a layer with emoticons. The application is simple and has no advanced map features compared to the other evaluated sample weather web maps.

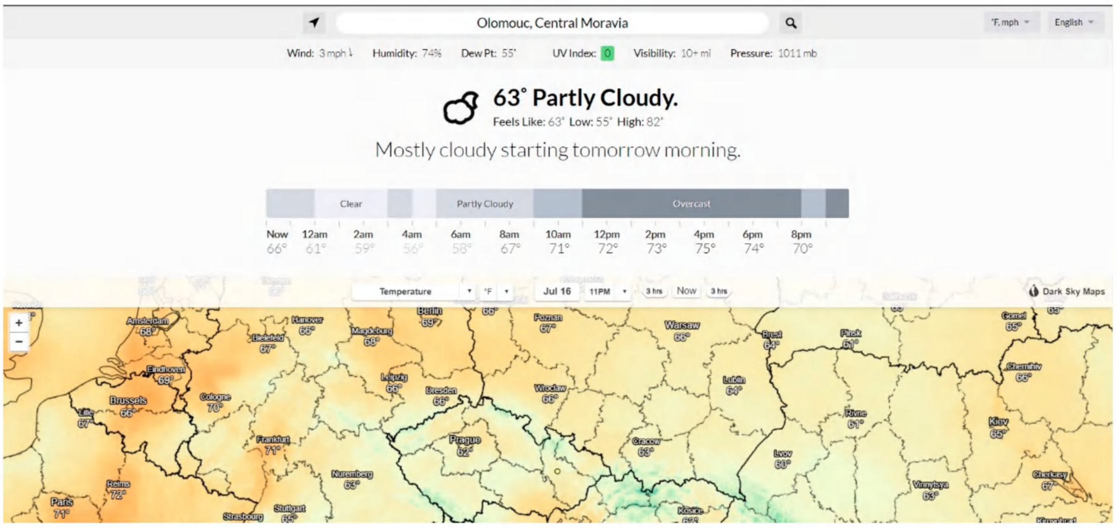


Figure 1. DarkSky map preview. The map has a simple composition, highlighting the basic information based on the location that is in the top field. The sample map in the preview is presenting temperatures.

Area patterns combined with the isoline method are implemented in this map. One thematic layer also offers emoticons in the point method. The map does not contain a legend. Methods are used correctly; nevertheless, a legend is missing. The application's design is simple and easy to use.

2.1.2. Windy

Windy (Figure 2) is an application developed by the Czech company Seznam.cz and is very detailed. For example, a user can choose his or her particular altitude. The map contains up-to-date weather information and forecasts for nine days. The map is loaded directly and has all controls in the map field. The composition is divided into four distinct areas: search, timeline, display options menu, and information menu. The layout of controls is logical and intuitive.

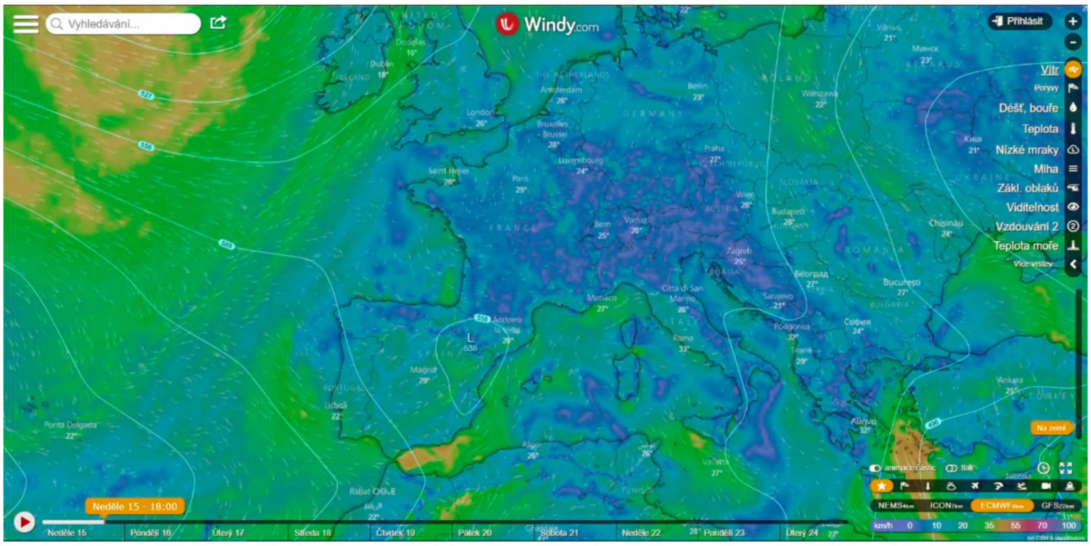


Figure 2. Windy map preview. The map has a composition divided into logical elements (search, timeline, display options, information menu). The sample map in the preview is presenting temperatures.

The map offers many options, for example, a wind conditions view for surfing, kiting and paragliding. The user can also choose layers for different activities (not only sports), such as aeroplane cloud elevation, sea currents for boats and snow elevation for skiers. It allows the user to choose the prediction model to calculate the prediction. Isoline and area pattern methods are used. The legend is in the lower-right corner of the map field.

2.1.3. In-Počasí

The In-Počasí web portal (Figure 3) is produced by the Czech Hydrometeorological Institute and contains a detailed weather forecast for the Czech Republic and a less detailed forecast for Europe. The portal is extensive and contains abundant weather information, including a map showing six meteorological phenomena in partial thematic layers. The map design is simple and intuitive. The composition of the map is balanced. At the top is a date bar with thematic layer information, while the section at the left allows thematic content and time options to be set. Below the map is a legend and supplementary information. Map controls are located outside the map field. The map does not show any interactive elements.

Only the isolines method is used, and it is used correctly. The only exception is the use of a colour scale, which can sometimes be confusing to users; the presented amount of cloud cover is maximum in white and minimal in dark blue colour, which is not usual. The legend is located below the map field.

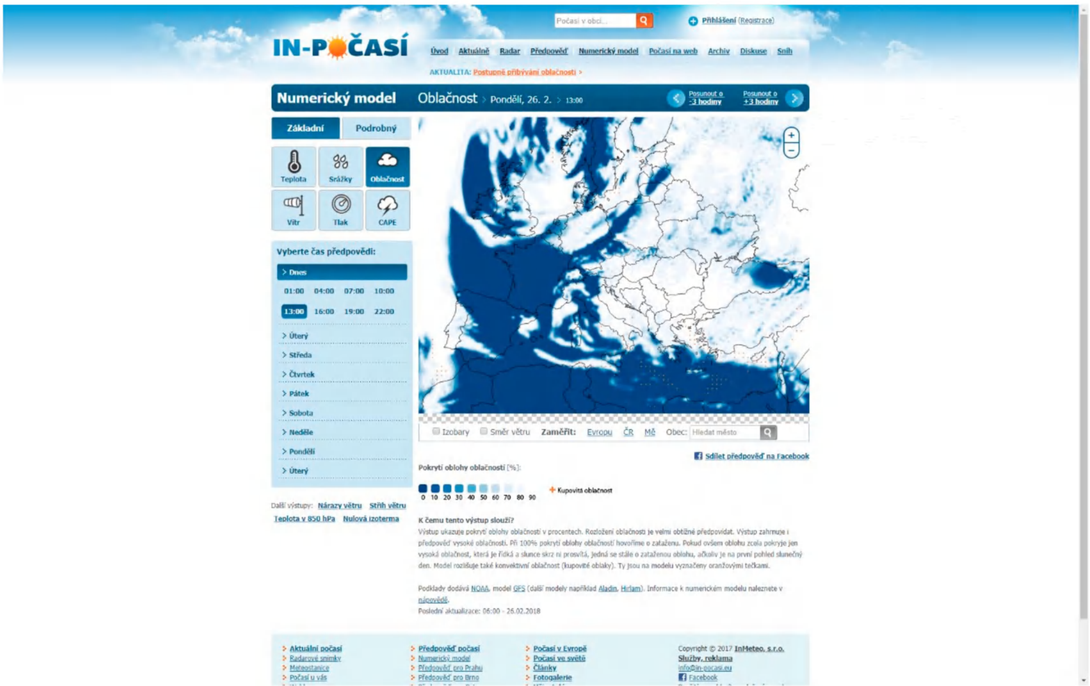


Figure 3. In-Počasí map preview. The map is part of a web-page layout and has additional controls as web-page composition elements. The sample map presents the amount of cloud cover.

2.1.4. YR.no

YR.no (Figure 4) is a Norwegian web portal and mobile app, including a map showing weather information. The user can choose to see the weather in the form of text, symbols, and diagrams or as a map. The meteorological map is available for Europe and Asia. The map's composition is divided into three areas: a layer switching menu, timeline, and the top bar dedicated to search.

Figure 4. YR.no map preview. The map composition elements are clearly graphically separated. The sample map presents the cloud coverage and precipitation.

Controls are embedded in the map. Switching thematic layers is different for Nordic countries and the rest of the world. Besides the basic meteorological indicators, other indicators are available, for example, UV radiation, sea currents or wave heights. Advanced and special features are not provided.

The main method used is the isoline method, which is deployed correctly, and the legend is well placed. The overall simple design permits easy searching. A positive feature of this web map is its interactivity when searching for interest areas—other charts and forecasts with relatively detailed values are displayed in this map. The disadvantage of this map is the difference in detail when displaying data for either Nordic countries or the rest of the world.

2.1.5. Wundermap

Wundermap (Figure 5) is produced by German Weather Underground. The map is loaded directly and the data are provided for the whole world. The page is divided into several blocks that are somewhat chaotically deployed: at the top is a search box and a button for sharing or switching on/off the thematic layer menu, in the right corner is a panel to control the thematic layers themselves, at the bottom is the timeline. Controls are standard and located in a menu at the edge of the map.

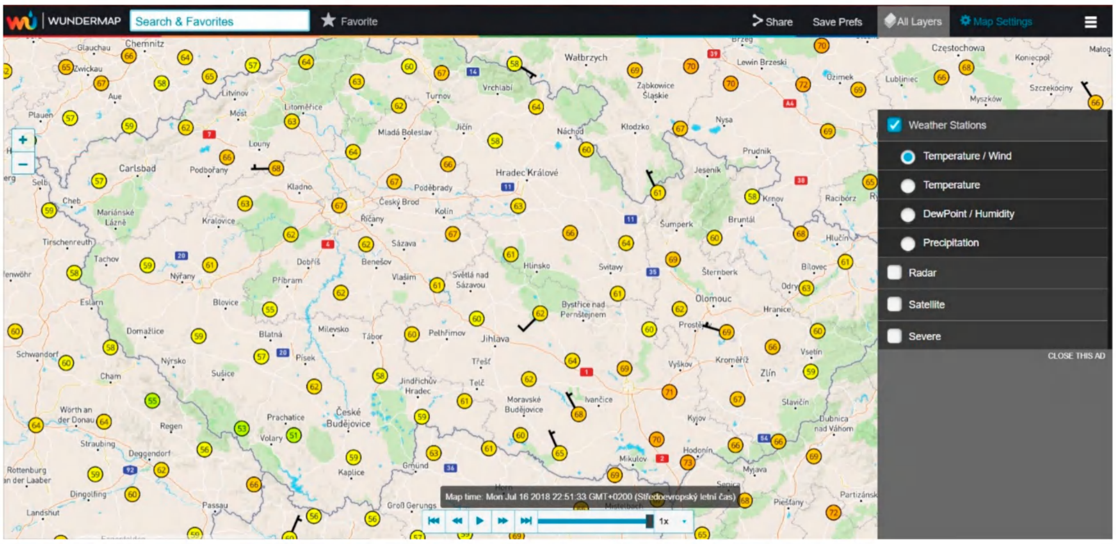


Figure 5. Wundermap map preview. The map has a simple composition with switching between the tabs (all layers, map settings). The sample map presents the temperature and wind.

The map contains both basic and advanced map features, although the thematic content is not interpolated, and the point method is used for "weather stations", which are irregularly scattered across the displayed territory. The map legend is hidden in the map settings tab and is incomplete.

2.2. Methods

The five maps described above were tested in an eye-tracking experiment complemented by think-aloud analysis. The eye-tracking experiment was performed at the Department of Geoinformatics, Palacký University Olomouc, Czech Republic, between 19 February 2018 and 23 March 2018. The eye-tracking laboratory is specifically designed for conducting eye-tracking experiments and is equipped with an SMI RED 250 eye-tracker with an operating frequency of 250 Hz. The eye-tracking data recordings were supplemented by audio and video recordings of the respondents. These data were used for further think-aloud analysis. When the eye-tracking experiment was completed, the results were analysed, evaluated and interpreted.

2.2.1. Design of the Experiment

The SMI Experiment Center™ software was used to design the experiment. The eye-tracking test was divided into introductory, dynamic and static sections (Figure 6).

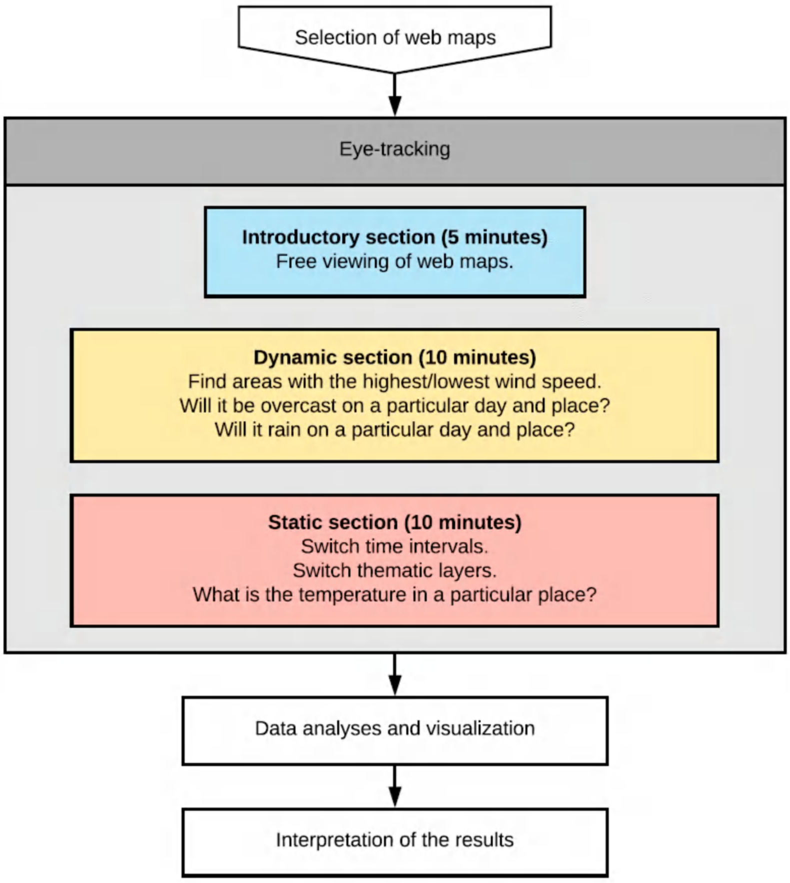


Figure 6. Map evaluation process.

The introductory section consisted of free viewing of selected web maps, one minute for each map. Users could work with the map and learn about its functionality. This section took five minutes. In the dynamic section, each map always had three questions (three rounds of questions for five evaluated weather maps). This section of the test was designed to take no more than ten minutes. Questions in each round were defined differently for each web map so that the respondent was prevented from memorizing the correct answer and forced to work with the evaluated weather web map. The dynamic section of the test was presented "live" – so each respondent saw different weather pattern because they were looking at different days.

The static section of the test also provided three rounds of questions and was designed to take no more than ten minutes. In static testing, the respondent was prevented from interacting with the elements in the map and could only view the static image (screenshot) of the evaluated weather web map. For the last question (*what is the temperature in a particular place?*), locations were changed so that respondents did not memorize the answer.

The first round of questions in the dynamic section addressed wind speed. The respondent was required to answer two questions concerning which area of the Czech Republic currently had the highest or lowest wind speeds. The second round of questions consisted of five questions concerning cloud cover. Respondents were asked to respond whether clouds were at a specific location and time. Questions in the third round concerned precipitation. The respondents answered whether rain would occur at a particular place and time.

The questions were defined so that users had to switch thematic layers, use the search or scroll map, switch timelines and be able to work with the legend in order to answer them correctly. Responses were recorded using a webcam with audio recording and logging of mouse clicks. The test was not devised to elicit the correct answer but to analyse how users worked with the map and whether they could find the required features to accomplish the task. The correctness of the response was therefore only an accompanying indicator of whether the user had correctly understood the phenomenon displayed.

As mentioned, the static section of the test also had three rounds of questions and took no more than ten minutes. The difference between static and dynamic testing is significant. In static testing, the respondent was not permitted to interact with the elements in the map and could only view the static image (a screenshot clip) of the evaluated weather web map. This type of testing cannot be used to determine whether a user can actively use a web map as a whole. Static testing evaluates whether a user understands the phenomenon and can find the basic web map controls and understand the map layout.

The respondents were asked the same questions about all web maps. Respondents were required to indicate in the static picture where to switch the weather forecast to another day, where to switch thematic layers, or to answer what the temperature was at a given location. For these questions, each map presented was of a different location to prevent the user from memorizing the same answer.

2.2.2. Respondents

Web maps showing the weather and phenomena associated with weather are usually up to date and accessible to anyone. The target user group of these maps is therefore extensive and not limited by age, employment, literacy or nationality. Weather information is available to everyone around the world. This suggests that weather web maps should be adapted to a large number of user groups. Therefore, the user interface of a weather web map and level of adaptation to user needs should be tailored to the comprehensive needs of different target user groups.

Testing was therefore targeted at multiple user groups. Thirty-four respondents participated in the eye-tracking experiment (14 males and 20 females, median age 23 years). These respondents were separated into two groups of users: novices (16) and experts (18). Students who had not studied Earth Sciences and other respondents without a more in-depth knowledge of meteorology, geoinformatics or cartography were included in the group of novices. This separation may not always be tangible. Nevertheless, a non-geographic student may understand maps and have more experience than a student in Earth Sciences. For a more reliable separation, respondents were asked whether they had any previous experience with web maps, and if so, were included in the expert group. All respondents were from the Czech Republic or Slovakia and the instructions were in Czech. The respondents participated in the study voluntarily and were not paid for the experiment.

To obtain representative test results, testing a predetermined number of respondents is appropriate. This number depends on the nature of the test data, specifically on the number of problems that may arise when solving tasks. Therefore, ten users were tested in the first stage and six problems were identified during testing, these being difficulties in navigating the web map, inability to find an answer without assistance, a poorly recognizable colour scale, inability to find where to switch thematic layers, misunderstanding of the presented phenomena, and inability to find where to switch time intervals.

The online calculator *MeasuringU* [41] (https://measuringu.com/problem_discovery/), which calculates an estimated sample size from the given occurrence of problems, was used to help estimate the ideal number of respondents. This calculator is based on normalization and the binomial probability equation. Problems recorded from the sample of respondents (in this case, the first ten respondents) were entered into the matrix. The calculator estimates how many respondents would be appropriate for testing to detect at least 99% of the problems encountered (Figure 7). In this case, the result was 26 respondents. As mentioned above, a total of 34 respondents participated in the test, which was more than recommended.

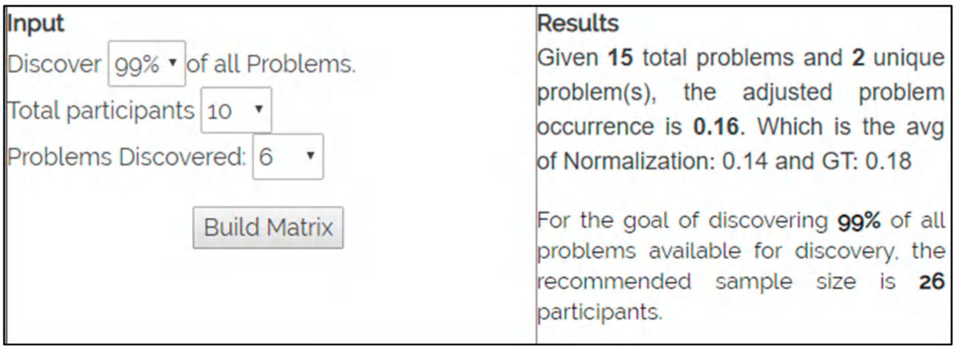


Figure 7. Estimated number of respondents based on the occurrence of problems (MeasuringU).

2.2.3. Analytical Methods

Before the recorded data were statistically evaluated and analysed, data pre-processing was performed. This included a data check and quality control and the exclusion of respondents where a recording error appeared during the experiment. DataLoss, or a percentage of incorrectly measured records, was less than 1%, and the rated data, therefore, retained a high reporting value, as only two user records were removed from the experiment. Fixations and saccades were identified using the I-DT algorithm with dispersion = 80 px and duration = 50 ms. Popelka [42] explains this setting in more detail.

The first step of data analysis was to evaluate the accuracy of respondents' answers. This analysis was not straightforward. In the first task of the dynamic section of the experiment, answers were recorded by clicking on the map. The analysis of this kind of data was a lengthy process. Testing was performed over several weeks and with screen recording (dynamic eye-tracking test). The data displayed on the weather map were therefore continually updated, and each respondent saw different values. The accuracy of answers in the dynamic section was done manually based on recorded videos or using notes created during the testing.

The eye-tracking experiment was divided into three parts– the introductory test section, dynamic test section and static test section. The methods of analyses vary due to the different nature of the recorded data in these three parts.

In the Introductory Section of the experiment, the results were gained based on the video recordings of respondents' work with the map overlayed by eye-movements. After viewing all recorded videos, a fundamental insight applying to all the web maps used in the experiment was gained.

Processing the results of the Dynamic Section was very time-consuming, as it was necessary to analyse data using dynamic Areas of Interest. Since each respondent worked with the map individually, dynamic Areas of Interest were created for each web map and each respondent separately. These areas of interest (AOIs) were: map fields, timer switching, switching of thematic layers and other information such as legends and supplementary charts. These layers were not active throughout testing and appeared according to how respondents clicked on them. Creation of dynamic AOIs is highly time-consuming, so only six respondents were chosen for this type of analysis. Data were visualized using Sequence Chart method, which displays each respondent's eye-movement data in time as rows. The colour of these rows corresponds to the visited AOIs.

Analysis of the Static Section was much easier, since all respondents were looking on the same stimuli – screenshots of the web maps. The first method, called Gridded AOI is implemented using the open-source OGAMA. The image was divided into a regular grid, each grid segment displaying how many fixations were recorded there.

Another method utilized in eye-movement data visualization is called FlowMap and is implemented in V-Analytics software. FlowMaps use Thiessen polygons generated based on the fixation distribution. Arrows between these polygons display the number of moves between them.

ScanGraph was another method used to study the above task. This method was developed to identify differences in the stimulus reading strategy of different groups of respondents [43]. Before analysing the data, areas of interest over the stimulus must be created and marked, for example, A, B, C, etc. The Scanpath of each respondent can then be replaced by a string of letters expressing the order of the visited areas of interest. ScanGraph calculates the similarity of these strings by employing three different algorithms: Levenshtein distance, Needleman-Wunsch algorithm and Damerau-Levenshtein distance. Individual respondents are visualized as nodes in the graph, and ScanGraph searches the so-called "cliques" in this graph - a group of respondents who are similar to each other at least to a specified degree. The tool can be used to determine, for example, whether the stimulus was read differently by men and women or experts and novices.

Both the Dynamic and Static Sections were also analysed statistically using the Wilcoxon rank sum test, since the data did not have a normal distribution. Statistically significant differences are marked by an asterisk in the figures below. We chose three eye-tracking metrics to analyse data – Trial Duration, Fixation Count and Scanpath Length. Description of these metrics and their meanings is in Table 1.

Table 1. Description of the eye-tracking metrics used and their meanings.

Trial Duration	Longer time needed to solve a task indicates a problem with user interface or higher complexity of the task.
Fixation Count	A higher number of fixations indicates a low level of search efficiency or an inappropriate user interface of the evaluated application [44].
Scanpath Length	A longer scanpath indicates less efficient searching (perhaps due to a sub-optimal layout) [45].

In addition to eye-tracking, the Think-Aloud method was also used to analyse respondents' behaviour during the experiment. Unfortunately, the majority of respondents had problems with verbalizing their actions. They were therefore given the required silence during testing to fully concentrate. For this reason, the Think-Aloud method was only employed with some of the more experienced respondents.

3. Results

3.1. Accuracy of Answers

The first step in evaluating the eye-tracking experiment was to analyse the accuracy of respondents' answers. All responses recorded during the test are listed in Table 2.

The first task of dynamic testing was to identify areas with the highest or lowest wind intensity by clicking on the map. From the table, it is evident that this task was highly problematic in the case of Wundermap. The information about wind speed is combined with the information about the temperature. Temperature was expressed by colour and number (degrees), but the wind speed was displayed using the symbol shape. This was confusing for the respondents. For the rest of the maps, fewer users responded with incorrect answers.

In the second and third task of the dynamic testing, respondents answered whether it would be cloudy (task 2) or rainy (task 3) in a particular place. It was found that if respondents knew how to find the answers, their responses were correct in most cases. In Table 1, red indicates situations when a respondent chose the wrong answer or gave up (chose to answer the question with "No Answer"). The bold in the table refers to situations when a little assistance from the researcher was needed. The most significant problems in tasks 2 and 3 were encountered with the Wundermap map, in which respondents were not able to orient themselves.

Table 2. Responses to questions and tasks given by the respondents in the dynamic and static sections of the test. Green indicates correct answers, red is incorrect answers, and bold plus exclamation mark indicates answers where a small amount of assistance was required.

	Dynamic part															Static Part														
	Task 1					Task 2					Task 3					Task 1					Task 2					Task 3				
Participant	Darksky	Windy	Inpocasi	YR.no	Wundermap	Darksky	Windy	Inpocasi	YR.no	Wundermap	Darksky	Windy	Inpocasi	YR.no	Wundermap	Darksky	Windy	Inpocasi	YR.no	Wundermap	Darksky	Windy	Inpocasi	YR.no	Wundermap	Darksky	Windy	Inpocasi	YR.no	Wundermap
P01	OK	OK	OK	OK	X	YES	YES	YES	YES	YES	NO	YES	NO	NO	NO	OK	OK	OK	OK	X	OK	OK	OK	OK	OK	NA	NA	-8	-3	21
P02	OK	OK	OK	OK	X	YES	YES	NO	YES	X	NO	NO	NO	YES	X	OK	OK	OK	OK	X	OK	OK	OK	OK	OK	16	-9	-6	-4	16
P03	OK	X	OK	X	X	YES	YES	NO	YES	X	NO	NO	NO	YES	X	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	18	-9	-6	-4	20
P04	OK	X	OK	OK	X	YES	YES	YES	YES	YES	NO	NO	NO	NO	YES	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	18	-9	-6	-1	19
P05	X	X	X	OK	X	YES	YES	YES	YES	X	YES	NO	NO	YES	X	X	OK	OK	OK	X	OK	OK	OK	OK	OK	21	-9	-6	-3	16
P06	X	OK	X	X	X	NO	YES	YES	NO	YES	NO	NO	NO	YES	NO	X	X	OK	OK	X	OK	OK	OK	OK	OK	20	-9	-30	-5	20
P07	OK	OK	OK	X	X	NO	NO	NO	NO	X	NO	NO	YES	YES	X	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	18	-9	-6	-3	20
P08	OK	OK	OK	X	X	NO	NO	NO	NO	NO	NO	YES	NO	NO	NO	OK	OK	X	OK	OK	OK	OK	OK	OK	OK	18	-11	-15	-2	19
P09	OK	OK	OK	OK	X	**YES!**	**NO!**	NO	NO	X	NO	YES	NO	YES	X	X	OK	OK	OK	OK	OK	OK	OK	OK	OK	18	-9	-15	-8	16
P10	X	OK	OK	OK	X	YES	NO	NO	YES	**X!**	NO	YES	YES	YES	NO	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	22	-9	8	-3	20
P11	OK	OK	OK	OK	X	NO	NO	NO	NO	X	NO	X	YES	YES	X	OK	OK	X	OK	X	OK	OK	OK	OK	OK	18	-9	-15	-4	20
P12	X	OK	OK	OK	X	NO	NO	NO	X	NO	NO	NO	YES	YES	X	OK	OK	X	OK	OK	OK	OK	OK	OK	OK	18	-9	-4	-3	21
P13	OK	X	OK	OK	X	YES	YES	NO	NO	**NO!**	NO	**NO!**	YES	**NO!**	X	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	21	-9	-6	-3	21
P14	OK	OK	OK	OK	X	YES	YES	NO	NO	**X!**	NO	YES	YES	NO	NO	OK	OK	X	OK	X	OK	OK	OK	OK	OK	18	-9	-4	-4	21
P15	OK	OK	OK	OK	OK	YES	YES	YES	X	X	NO	NO	YES	NO	X	OK	OK	OK	OK	X	OK	OK	OK	OK	OK	18	-9	-15	-3	21
P16	OK	OK	OK	OK	OK	YES	NO	YES	NO	X	NO	YES	YES	YES	X	OK	OK	OK	OK	X	OK	OK	OK	OK	OK	18	-9	-15	-9	19
P17	OK	X	X	OK	X	YES	YES	NO	NO	X	NO	NO	NO	NO	**X!**	OK	OK	OK	OK	X	OK	OK	OK	OK	OK	20	-9	-6	-4	16
P18	OK	OK	OK	OK	X	NO	NO	NO	NO	**X!**	NO	NO	YES	NO	X	OK	OK	OK	OK	X	OK	OK	OK	OK	OK	20	-9	-6	-3	19
P19	OK	OK	X	X	X	YES	NO	NO	NO	X	NO	NO	YES	YES	**X!**	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	22	-9	-4	-4	20
P20	OK	OK	OK	OK	X	YES	NO	YES	YES	NO	YES	NO	NO	YES	X	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	18	-9	-6	NA	23
P21	OK	OK	OK	OK	OK	NO	NO	YES	YES	X	YES	YES	YES	NO	NO	OK	OK	X	OK	OK	OK	OK	OK	OK	OK	18	-9	-8	-4	18
P22	OK	OK	OK	OK	OK	NO	NO	NO	NO	YES	NO	YES	YES	NO	X	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	18	-9	-6	-4	19
P23	OK	OK	OK	OK	X	YES	YES	YES	NO	YES	YES	NO	YES	X	X	OK	OK	OK	OK	X	OK	OK	OK	OK	OK	18	-9	-6	-3	19
P24	OK	OK	OK	OK	X	NO	NO	**NO!**	YES	**X!**	X	YES	YES	YES	**X!**	OK	OK	OK	OK	X	OK	OK	OK	OK	OK	NA	-9	-6	-3	16
P25	OK	OK	X	OK	OK	YES	NO	**X!**	YES	**X!**	YES	YES	YES	YES	**X!**	OK	OK	OK	OK	X	OK	OK	OK	OK	OK	23	-9	-6	-3	19
P26	OK	OK	OK	OK	X	YES	YES	YES	YES	**X!**	NO	YES	YES	YES	**X!**	OK	OK	X	OK	OK	OK	OK	OK	OK	OK	2	-9	-15	-3	20
P27	OK	OK	OK	OK	X	YES	YES	YES	NO	**X!**	NO	YES	YES	NO	**X!**	OK	OK	OK	OK	OK	OK	OK	OK	OK	OK	21	-9	-6	-4	16
P28	OK	X	X	OK	X	NO	NO	YES	**YES!**	**X!**	YES	NO	YES	YES	**X!**	OK	OK	X	OK	X	OK	OK	OK	OK	X	16	-9	-8	-3	16
P29	OK	OK	OK	OK	X	NO	NO	YES	**NO!**	X	NO	NO	NO	NO	**X!**	OK	OK	OK	OK	X	OK	OK	OK	OK	OK	24	-9	-6	-3	21
P30	OK	X	OK	OK	X	YES	NO	YES	YES	X	NO	YES	YES	YES	X	OK	OK	OK	OK	X	OK	OK	OK	OK	OK	18	-9	-4	-3	16
P31	OK	OK	OK	OK	X	**YES!**	**NO!**	**YES!**	**NO!**	**X!**	**NO!**	**NO!**	**NO!**	**YES!**	**X!**	OK	X	OK	OK	X	OK	OK	OK	OK	OK	21	-9	-6	0	21
P32	OK	OK	OK	OK	X	NO	YES	NO	YES	**X!**	YES	YES	NO	YES	**NO!**	OK	OK	OK	OK	X	OK	OK	OK	OK	OK	21	-9	-8	-4	16
P33	OK	OK	OK	OK	OK	NO	YES	**X!**	NO	**X!**	NO	NO	YES	YES	**X!**	OK	OK	OK	OK	X	OK	OK	OK	OK	OK	18	-9	-15	-3	16
P34	OK	X	OK	OK	X	YES	NO	YES	NO	**X!**	NO	YES	YES	YES	**X!**	X	OK	OK	OK	X	OK	OK	OK	OK	OK	23	-9	-6	-3	18
Correct	30	26	28	29	6	34	34	32	32	9	33	33	34	33	8	30	33	27	34	14	34	34	34	34	33	16	32	17	27	0
Wrong	4	8	6	5	28	0	0	2	2	25	1	1	0	1	26	4	1	7	0	20	0	0	0	0	1	18	2	17	7	34

LEGEND Correct answer Wrong answer **Bold** Assistance needed

In the Static section of the experiment, respondents were required to first indicate where the time interval on the map could be switched (task 1). To evaluate the correctness of the responses, areas of interest in the stimulus around the correct answers had to be created to detect whether respondents had clicked on the field. The most significant problems again occurred with Wundermap (Figure 8). In the second question, respondents were required to indicate where the thematic layer could be switched. In this situation, almost all of the answers on all maps were correct; only one respondent (P28) on the Wundermap answered incorrectly. In the final question of the static section, respondents answered what temperature it would be at certain times in certain cities.

It was immediately apparent that users had the most significant problems finding the correct answers in the Wundermap weather web map; all the respondents' answers were incorrect due to the unreadability and misstatement of the presented phenomenon. The authors of the map had chosen an inappropriate cartographic method for visualizing temperature, and respondents were not able to state the temperature in a given city with any certainty. Figure 8 shows a screen capture of the Wundermap in which test respondents were asked to find the temperature in Olomouc.

Most of the respondents correctly responded to the tasks on YR.no and Windy weather web maps and could orient themselves to find the correct answer quickly. On the In-Počasí web map, respondents had problems finding the interval to which the correct answer belonged. The colour range of the displayed values is extensive, and the colour spacing between individual colours is difficult to discern. Respondents found it difficult to assign the colours depicted on the map to the correct interval in the legend. On the DarkSky web map, respondents had to make a greater effort than on previous maps to find the temperature, which was not highlighted in the map but only indicated in the information text located above the map field.

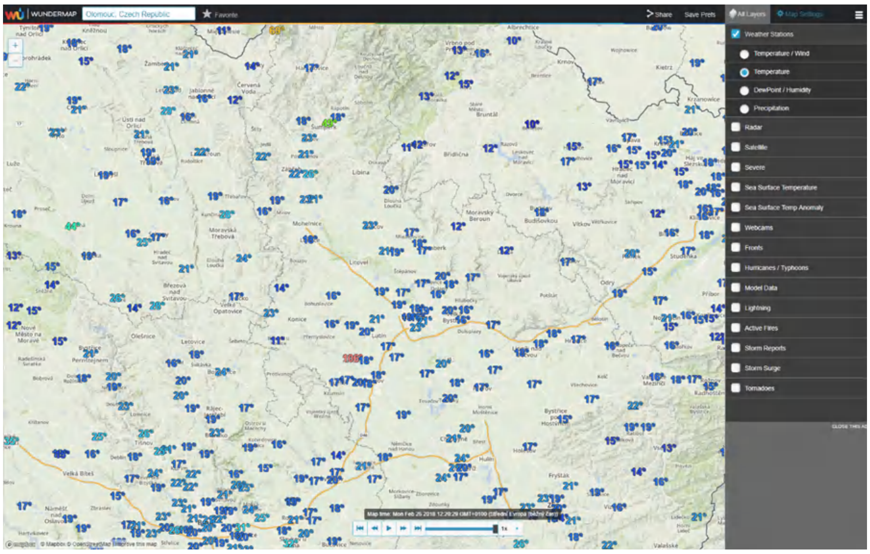


Figure 8. Image of the Wundermap weather web map from the test.

3.2. Eye-tracking Results

3.2.1. Introductory Test Section Results

As described above, the first part of the eye-tracking test was free viewing of selected weather web maps. In this section, respondents were required to view the maps they would work with throughout the test in five minutes. This section was not evaluated in detail, as it was aimed at orientating primary users with the selected web maps. For this task, only the essential characteristics of each evaluated web map were summarized and are explained below.

Respondents in the novice group worked differently with the maps. Novices viewed the map itself, zoomed in on their place of residence, viewed the contents of the map and then focused on switching thematic layers, etc. Respondents in the group of experts, however, immediately focused on map functionality after the maps were loaded. They looked for available thematic layers, switched timescales and attempted to find out whether it was possible to switch units where the forecast was displayed and whether it was possible to look into the legend. These basic findings confirmed the appropriate separation of respondents into groups of novices and experts. More than 70 percent of respondents thus had typical behaviors corresponding to their inclusion in the group of novices/experts, and less than 30 percent of respondents did not demonstrate this typical behavior.

During free viewing of the DarkSky web map, respondents focused mainly on switching thematic layers, switching the time for displaying the forecast and observing the headline of the web map, where the current temperature was written with large digits (set by default to Fahrenheit). While browsing, respondents had no problems finding basic controls.

The Windy web map is the most attractive at a glance. Each respondent navigated differently through the map, as it was possible to select and display many different thematic layers and show different units and time intervals. The possibilities are almost countless, and respondents, therefore, moved around the map field with considerable difference. Interestingly, most respondents used the mouse wheel to zoom in/out, not the button specified in the map box.

Respondents did not encounter any problems while viewing the In-Počasí web map. Control and understanding of the map were intuitive, and free viewing therefore did not present any unexpected conclusions. Test respondents attempted switching thematic layers, zooming in and out, switching predictions and looked for primary or detailed viewing.

Free viewing of the Norwegian web map YR.no also demonstrated that respondents had no problems handling the map. As in other maps, they attempted basic web map control. Some respondents selected interactive map features, mainly graphs showing additional weather information. The unique feature of this map is the possibility of displaying different thematic layers for Scandinavian countries than other European countries. No peculiarities in controlling the map were observed.

As the final map in the free viewing section, the Wundermap web map provided the most significant difficulties for respondents. All of the respondents attempted switching thematic layers, but over 50% experienced problems with loading thematic layers (slow loading of content during zoom in/out). Problems were also encountered with switching prediction timing, and some respondents mentioned that they did not understand the method of data visualization, suggesting that their interpretation of the map's information was problematic.

No unpredictable conclusions were discovered from the free viewing. Respondents always explored the basic functionality of the web maps, how to control them and the possibility of displaying thematic layers or additional functions. As mentioned above, the main reason for the free viewing section was to for respondents to gain familiarity with the maps. Respondents who had worked with web maps previously (experts) focused more on the functionality of the web map and the display options the web map offered. By contrast, users with less experience of web maps (novices) were primarily interested in the map's content (viewing places on the map or attempting to find their place of residence).

3.2.2. Dynamic Test Section Results

The dynamic section of the eye-tracking experiment immediately followed the introductory section. The objective of this section was to monitor and identify how respondents worked with the maps, whether they used all the available elements, used the map interactively or otherwise. In this part of the test, each web map consisted of three tasks.

The first task required: *Locate and click to highlight the area with the lowest/highest real-time wind speed in the Czech Republic*. This question was evaluated by creating dynamic AOIs and then visualized using an AOI Sequence Chart. Sequence Charts were created for six respondents—three experts and three novices. In the following charts (Figure 9), six respondents and their work with the web maps to find an answer to the given question can be seen. One chart was created for each test web map.

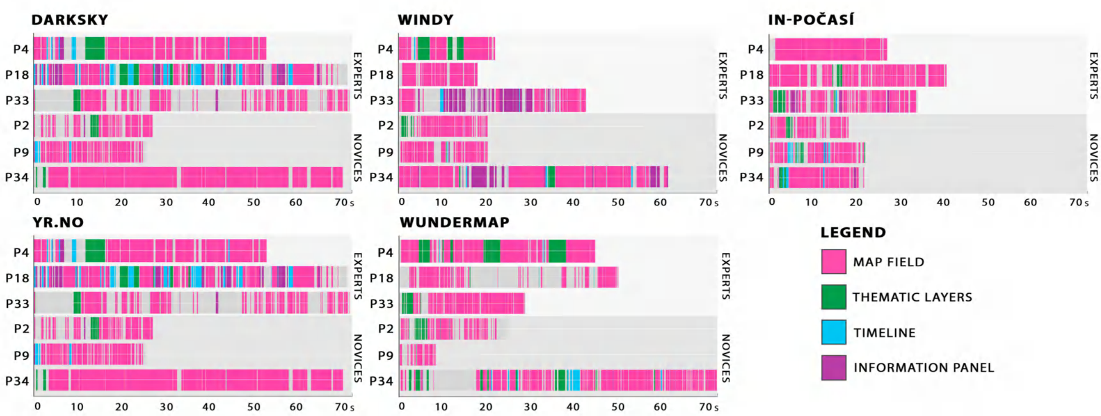


Figure 9. Areas of Interest (AOI) Sequence Chart for all evaluated maps for six selected respondents (selection of illustrative examples).

While searching for a solution to this task, respondents spent the most time on the DarkSky web map. Novices had significantly longer response times than experts (except P34) (Figure 9). P34 was not sure of the answer and thoroughly explored the map to properly identify the place he wanted. All of the maps showed that experts needed much less time to find the answers than novices. Ideally, a respondent would orient themselves, look into the thematic layers, activate the thematic layer for wind and then look back to the map field to find the desired area, i.e., depicted in a sequence of pink-green-pink colours. This sequence was observed for the YR.no web map, where searching was most effective. If several colours alternated in the graph in succession, it indicated that the respondent was confused about finding the correct answer on the screen or that the map controls were inappropriately divided. The AOI Sequence Chart of a Wundermap web map could be misled by assuming that searching on this map would be efficient and fast. However, in reality, it was different. Respondents mentioned that searching on this map was too complicated and did not even attempt to locate the right answer on it.

In addition to visualizing the Sequence Chart of selected respondents, three eye-tracking metrics were analysed—Fixation Count, Trial Duration and Scanpath Length. In all three cases, the identified trend was similar. The In-Počasí web map offered the fastest solution, respondents very quickly finding the button to switch to the thematic layer for wind information and reading the scale colour. The least effective in terms of Scanpath Length, though, were observed with the Wundermap web map, where information about wind speed and direction was incomprehensible. Surprisingly, a relatively low Scanpath Length value on the Windy web map was observed (Figure 10). Wind speed and direction information in this map are processed in a very detailed way using animation. Therefore, the solution to the given task was more demanding than in a static visualization. It is important to note that wind information was presented in a much more detailed and accurate manner than information in other maps on the Windy web portal.

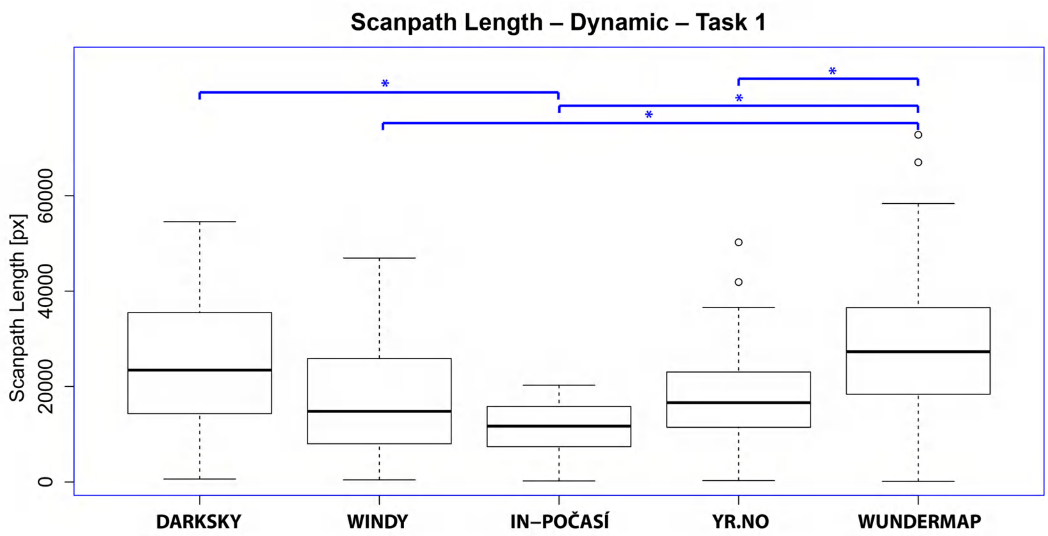


Figure 10. Scanpath Length analysis for the Dynamic Section—Task 1.

The second task asked: *Will it be cloudy today at [time] in [location]?* The objective of this task was to analyse the amount of time a respondent needed to find the answer (Figure 11). The longest response time was observed with the Wundermap web map, where respondents spent on average 65 s. (median 54.4 s) The shortest response time was observed with the In-Počasí web map, respondents spending on average 30 s (median 27 s), which is approximately half that of the Wundermap web map. As mentioned above, some respondents refused to use the Wundermap web map to find the correct answer because of thematic layers loading slowly and not being able to understand the cartographic method of the web map.

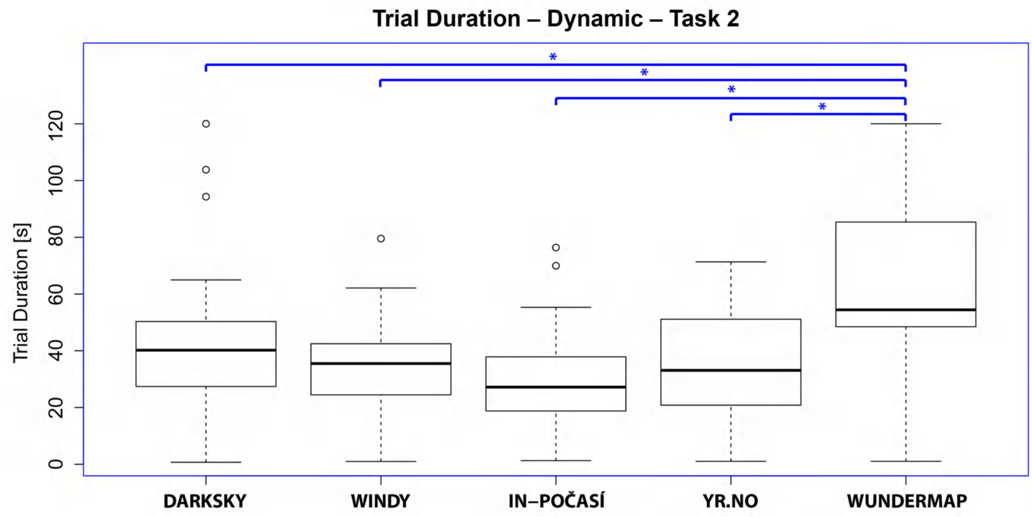


Figure 11. Trial Duration to find the answer in the Dynamic Section—Task 2.

The third and final task in the dynamic section asked: *Will it rain tomorrow in [location]?* This task concerned the occurrence of precipitation on the next day. The first evaluation method used was Fixation Count or the average number of fixations (Figure 12). This method shows how effective a user's search is in the stimulus, or whether the user interface of the tested stimulus (web map) is poorly defined. The greater the number of fixation counts, the less user-friendly the web map and the less effective the user search. The highest fixation median values were observed with the Wundermap web map (145) and DarkSky web map (126). The In-Počasí web map achieved the best results (84).

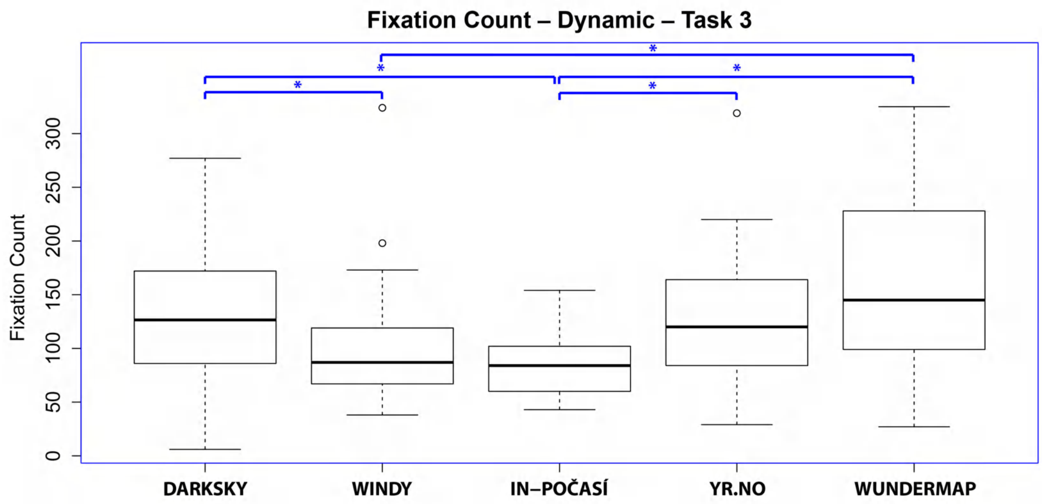


Figure 12. Fixation Count analysis in the Dynamic Section—Task 3.

From these conclusions and evaluations, it is clear that the Wundermap web map was the worst of all evaluation means and procedures, while the In-Počasí web map had the best features for interactivity, user-friendliness, convenience and adaptation to different user groups.

3.2.3. Static Test Section Results

The following section evaluates the static section of the eye-tracking experiment and provides corresponding conclusions. Processing this section was not dynamic or time-consuming. Each web map is evaluated and then compared to others at the end of this section.

The first task required: *Find and click where the weather forecast can be switched to another day*. A visual evaluation of this question was performed using the Gridded AOI method. This method was selected to facilitate the comparison of stimuli, regardless of their content. The resulting output is shown in Figure 13. Analysis showed that the Windy and In-Počasí web maps were intuitive to respondents, as they almost immediately found the required location on the map. In contrast, respondents searched for the required button on Wundermap. This analysis showed that despite the very colourful and graphically rich content of this map, the button to switch the weather forecast to another day is not conveniently or intuitively positioned. In the case of the In-Počasí web map, a simple and clean design with basic content and no unnecessary features proved to be user-friendly.

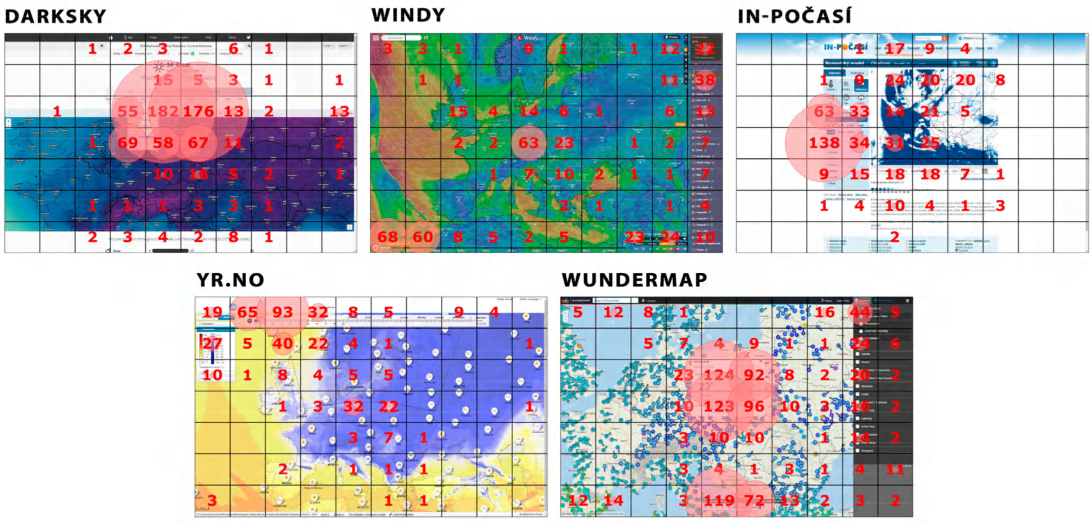


Figure 13. Number of fixations recorded in the regular grid overlaying the stimuli in Static – Task 1. The size of the circle and number represent the number of fixations recorded in each cell of the grid.

Another interesting indicator is Trial Duration (Figure 14). From the box plot, it is evident that respondents spent the most time finding the correct answer on the Wundermap web map and the least time on the Norwegian YR.no. Statistically significant differences were found between the Wundermap and all other maps.

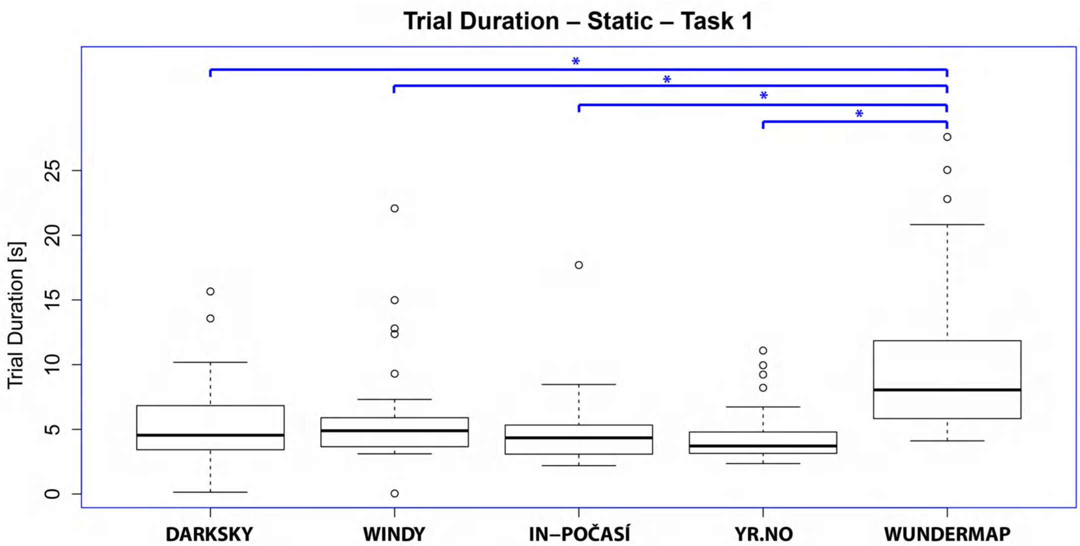


Figure 14. Trial Duration analysis in the Static Section—Task 1.

The other method of visualizing the results mentioned above was carried is a FlowMap (Figure 15). Similar lines of sight of respondents' eyes were observed. Only arrows with five or more moves between places were plotted. Where arrows are thicker and closer together, respondents were more efficient in searching for a result and knew where to search for an element to find and pinpoint it accurately. This is evident on the DarkSky web site, where respondents did not search the whole screen and found the time switching layer directly. The most confused searches were seen on the Windy and Wundermap web maps. The arrows on the Wundermap web map show high-frequency eye movements across the entire map. This means that respondents searched the entire screen and were distracted by the other elements displayed on the map, their search, therefore, being ineffective

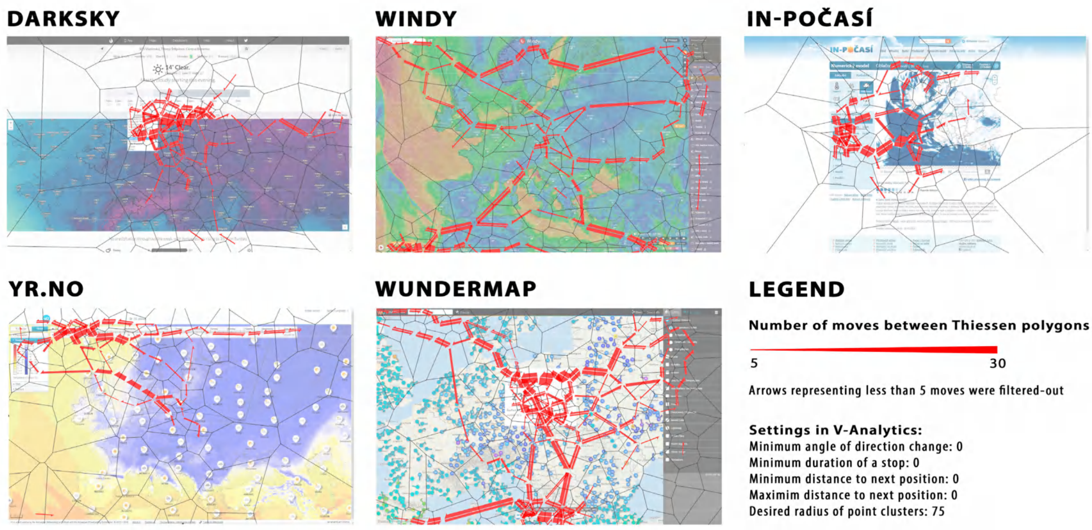


Figure 15. FlowMap analysis of evaluated web maps. The visualization displays aggregated eye-movements of all respondents as arrows between generated Thiessen polygons.

When combined with the two above-mentioned evaluation methods, DarkSky demonstrated the best-defined control for switching web map time scales. Both evaluations showed that this web map was best. The Windy web map is a little ambiguous, having the most significant number of fixations in the right place, although respondents only found it by searching the entire screen. The Windy web map is well-arranged, and the controls are intuitive. This contradictory evaluation could be attributed to the arrangement of its controls. These are located in the corners and sides of the map field; the controls are spaced apart, and a user has to navigate the entire screen to find out where the desired element is. A comparison of the average time respondents needed to find the data switch control showed that despite the far-reaching controls on the Windy map, the time required to find them was still less than on the Wundermap web map.

The second task also dealt with the web map and its controls, asking: *Find the place where the theme layers can be switched and click to mark it*. This presented a very similar situation to the third task of the dynamic section. The least fixations required to solve the task was recorded for the In-Počasí web map (Figure 16). A statistically significant difference was found between this map and the DarkSky, Wundermap and YR.no maps. Thematic layer switching on the In-Počasí web map is implemented through intuitive symbols, and respondents found it less complicated. Similar symbols are also used on Windy web map, but they are located in the top right corner of the screen, and are much smaller and less pronounced.

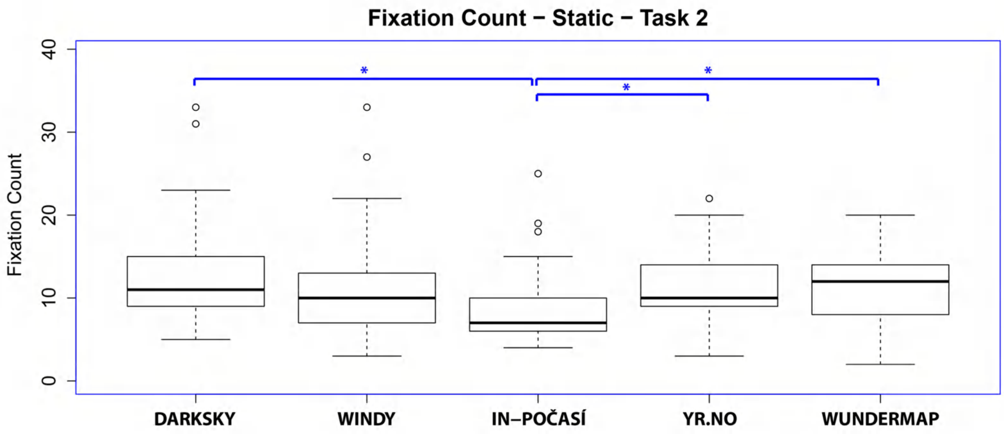


Figure 16. Fixation Count analysis in the Static Section—Task 2.

The final task in the static section of the eye-tracking test asked: *What is the temperature (in Celsius) in [location]?* The Windy web map provided the quickest solution, listing temperature values directly at each city. Between this map and all others, statistically significant differences were found in the Fixation Count and Trial Duration metrics (Figure 17). On the DarkSky web map, however, the temperature near Prague was missing, and a large number of incorrect answers were therefore recorded, and the time required to solve this task was the longest of all evaluated maps.

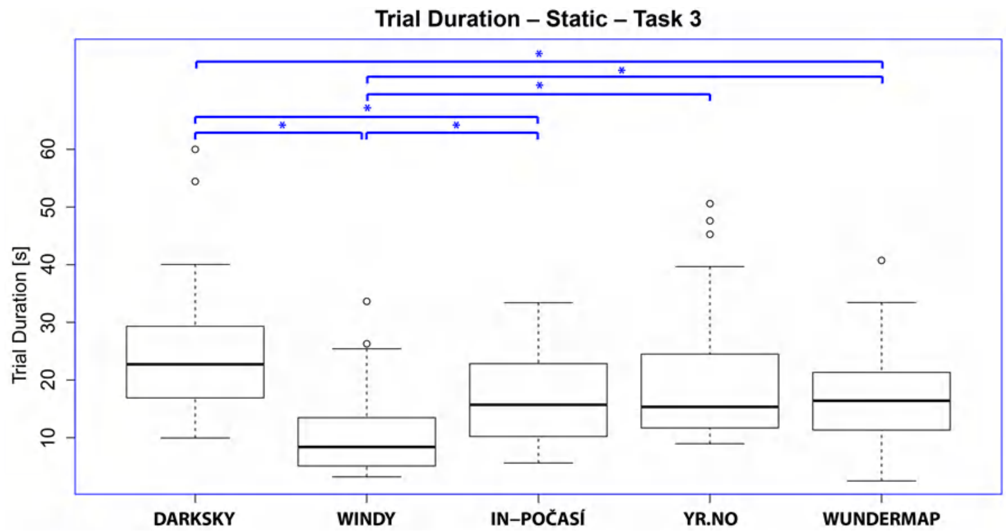


Figure 17. Trial Duration in the Static Section—Task 3.

ScanGraph was another method used to study the above task, because it can help to find similarities in the strategy of stimulus inspection. In this case, no significant differences were found between the expert and novice groups. ScanGraph was nevertheless used, only in a slightly different manner, in order to tease out similarities and differences in strategy between respondents. Distribution into groups of experts and novices was not considered, and parameter p indicating the degree of similarity was set to 100%; therefore, only those respondents whose order of visited areas of interest were the same became visible. At the same time, "collapsed" was selected so that repeated fixations in one area of interest were not considered. The resulting graphs for all five maps are shown in Figure 18. Each dot represents one respondent. The order of visited areas of interest is shown in red letters.

This analysis can show how difficult it was to find the right answer on individual maps and whether respondents chose the same strategy. On the DarkSky web map, only two respondents were observed utilizing the same strategy. This was at the transition between the areas of interest around the map field and the information text above the map that indicated the temperature for

Prague, which represented the correct solution to the task. Respondents could, therefore, look at this element (labelled B) to find the correct answer. However, it is clear that respondents did not realize this and did not apply this strategy. In a more in-depth data analysis, respondents needed to see at least three AOI (P17 – BAB), but no more than 60 (P24 – BABABABACBCDABADBABAC BACACABACABACBABABCDABABACABABABABCABAC), with an average of 15 respondents in the area of interest.

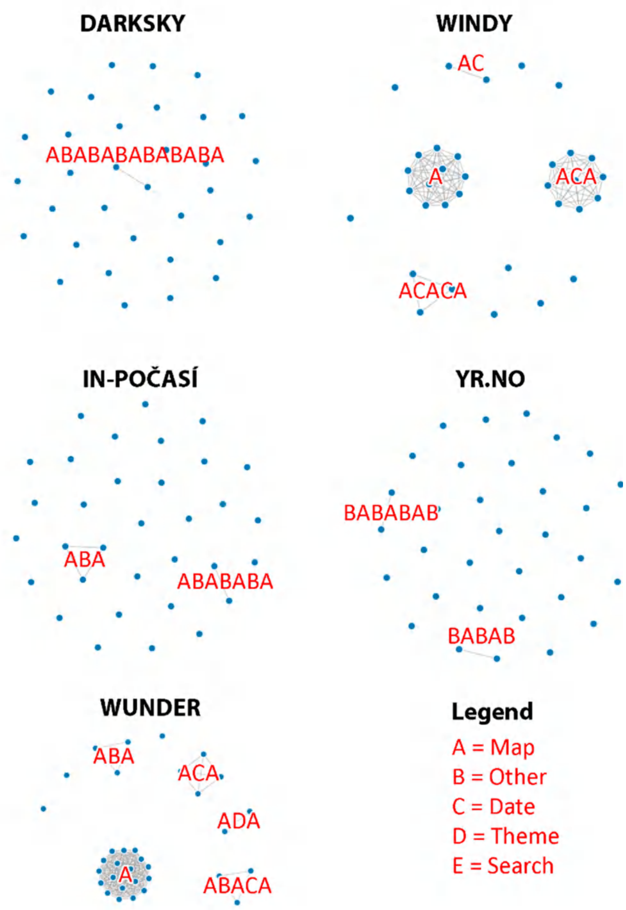


Figure 18. Graphic outcome of the ScanGraph analysis. Each dot represents one respondent Dots connected by an edge form a group based on the same order of visited Areas of Interest.

A diametrically different situation occurred with the Windy web map, with clearly identified groups of respondents applying the same strategy. The largest group consisted of eleven respondents, who looked at only the map field. Other groups of nine, three and two members looked at the date switch panel in addition to the map, where the weather forecast for the following days was also presented in the form of a meteogram.

On the In-Počasí web map, the most crucial area of interest was the colour gamut contained in the B-marked element. As shown in the boxplot in Figure 18, the third task on the In-Počasí web map was more demanding than the previous two, and respondents took longer to find the answer. This is confirmed by the ScanGraph analysis, which showed only two small groups of respondents with the same strategy.

A similar situation was observed with the YR.no web map, where the forecast for the following days was also displayed as a meteogram (labelled B). Even with this map, only two groups of respondents adopted the same strategy.

The largest group of respondents who adopted the same strategy to complete the task was observed with the Wundermap web map. This group consisted of 18 respondents, all of whom only looked at the map field. Unfortunately, the visualization of temperature on this map is very unclear, and in Olomouc, the temperature data overlapped, and it was difficult for respondents to find the correct answer.

3.3. Think-Aloud Results

The Think-Aloud method is one of the oldest research methods [46]. Since the analysis of eye-tracking data alone does not provide an answer to the question "Why does the user behave as he/she behaves", the application of the Think-Aloud method can bring new insights and justification of the acquired findings.

Although this method was planned to be applied to all respondents, most of the data collected were not relevant since, as already mentioned, respondents said it was difficult for them to describe what they were doing and why, and it interfered with their concentration. Therefore, they were often interviewed after the experiment was completed, so that the information gained could still be used for further analysis (but not as Think-Aloud results). In the text below is an example of a respondent who was able to cooperate 100%. It was an expert who commented on his actions and his reasons without any problems.

Because it is not a synthesis of knowledge based on the data from all respondents, but merely an illustrative example of how to use the method, it does not present the majority respondents' opinions or approaches. The commentary of one of the respondents was translated and transcribed into text as follows:

Question 1: Locate and click to highlight the area with the highest/lowest real-time wind speed in the Czech Republic.

> DarkSky: *"I'm trying to zoom in, but it's not possible using a mouse. Well, here I found some information on the map that's in Hradec Králové, because there are lower numbers than everywhere else, but I can't read it from the colours. However, do you want general information or rather point values? As a map user, I would go after that number, so I clicked on Hradec Kralove."*

> Windy: *"I'll find the wind. So, the information here is in degrees, and it's in the cities, and I would have to go here by colour and click somewhere near Olomouc. Also, when I click on it, I'll get the information with the exact number."*

> In-Počasí: *"Here, I would go to the border of the three regions, finding it by colour."*

> YR: *"Here, I have to study this strange colour scale for a long time to find out that it's the lightest green, and I'd like to see it somewhere near Zlín."*

> Wundermap: *"The wind is hidden somehow strangely here in the layer. So, I want to find the highest number, but what does that number show? Well, it's according to Fahrenheit, but it is the temperature, yet it's strange. Trying to right-click the legend, but I just can't see it. Well, look at this, I'm missing the legend, and it's been redrawn on another layer. Well, I can't find the highest one, so let's say here, because it's so green and there has to be that temperature. It's totally stupid to me."*

The quotes clearly show where the user found the answer quickly and where not. For example, the knowledge of the user's ability to read and understand colours on the maps is very beneficial.

Question 2: Will it be cloudy today at [time] in [location]?

> DarkSky: *"So, cloud cover can be clicked here. I wrote Prague into the search here, it's even listed here. So today at 10 pm, it will be cloudy and partly cloudy. I ignored the map and found it up there in that information."*

> Windy: *"So I switch to clouds. Here, I switched to clouds, and here found ten o'clock, and now I'm going to look at the map. Well, the answer is that it won't be overcast, but there will be some cloud."*

In-Počasí: *"Clouds are already there, here, it doesn't lead me to a location, but when I load it, there is not much of a base layer here, so I'm looking for a location not very well and estimate it will be covered by 50% or so, clouds will be there."*

YR: *"So precipitation, here, I'm misled and cannot find the clouds, but it is right in the icons, so that I can find the time and the place. Now I thought it would hit me and the meteogram would start, and at 8 o'clock it won't be cloudy. Clouds only arrive later."*

Wundermap: *"Help me. Here there might be a clue, I could find it there. Why does it load so slowly that I have to wait? We want it for 7 pm, but the map only shows now. How does the timescale change there, probably not. It may be because it is slow. So, when I click on the map, it will probably not work out anything. It stopped me from looking, so I won't even look for it."*

The task required work with layers/topics and quotes show that this has appropriately verified user credibility within map functionality. Sub-comments lead to knowledge of shortcomings occurring in the evaluated maps and provide the basis for better interpretation of the results.

Question 3: Will it rain tomorrow in [location]?

DarkSky: *"I see that I'm not the first to find it through search. So here I am, switching the date to tomorrow after I found the place. Well, it will rain there, but I'm not quite sure now if that's tomorrow. Well, I'm only a little bit confident that the contents of the map will switch to tomorrow, but not at all with the strip above with information. So, I have to look at the map, and it won't rain tomorrow. However, I'm not quite sure."*

Windy: *"So we want tomorrow again. So, it makes me think of the maps as they move, and yes it will rain."*

In-Počasí: *"I'm clicking on Friday and crashing. I know roughly where Paris is, but I would rather write it, and now I see it. Again, the times go through, and I can see that the showers will come, and more rain will come in the evening."*

YR: *"I'm starting to move here on that timeline, and I see that tomorrow it should be raining."*

Wundermap: *"I have to find it here, but there is a very slow server here. I was trying to select it, but the menu has been stuck. So, disappear. This is a pain. I want to know if it's going to rain. Well, here I can see only the current, and here's just a chart for today. I won't find out about tomorrow. I have a feeling I'm not going to find out about tomorrow."*

Again, the ability to work with web map features, including change of layer/topic and time, was evaluated in this task. Quotes show how the user obtained the information (from map movement, layer switching and time change, etc.) and what was easier for him.

3.4. Summary of Results

Users worked with web maps in the simplest form; they did not look for hidden functions in the menu or attempt to find any advanced functionality. They primarily looked at the controls on the main screen of the web map. If expandable control panels were available, the respondents only looked at them after they had examined other elements. Therefore, interactive map elements were only explored by respondents after they had become acquainted with the map. Map interactivity was not an obstacle unless it contained too much information or options to choose from. Searching was still quicker in static menus that respondents did not have to switch on or off. Static menus were available on Windy, In-Počasí and Yr.No; it was necessary to switch the menu on/off in DarkSky and Wundermap; Figures 10–13 show the better results for the maps with static menus. For example, the average value of Scanpath Length (Figure 10) for maps with static menus was 14,350px; for those with dynamic menus the average Scanpath Length value was 25,355px.

After evaluating how users worked with weather web maps, novices were identified as being disinterested in web map functionality and primarily interested in map content and what they could see on the map (for example, whether they could find their place of residence). Experts, though, were interested in exploring web map functionality, such as display capabilities, thematic layers, additional analysis, zooming in/out, switching timescales and other features (based on the comparison of ScanGraph analyses and qualitative evaluation). Mapmakers (cartographers and GIScience experts) should, therefore, consider the target user group when designing a map. Given that weather information is accessed by complete cartographic novices with minimal web map experience, weather web maps should be as simple as possible. The importance of this statement is paramount if the map is intended for the public. If mapmakers expect the map to be mainly for experts and the web map content will contain not only basic weather indicators but also extensive meteorological indicators and indexes of meteorological phenomena, the choice of more sophisticated interactive elements is advisable.

User issues are of relevance to many aspects of mapmaking, such as historical, sociological, psychological, conceptual, and others. One of the most important issues is adapting to the needs of different user groups. User issues in cartography are determined by map users and represent the most important influence in the process of map creation [47]. It is such an important aspect that map makers should pay great attention to it.

Much research is involved in discovering user interests and preferences. In some studies, however, user preferences have been shown as not very accurate regarding the quality of assessed geovisualizations and maps and the suitability of their respective purpose and user target group [48]. This finding was confirmed in this study, specifically in the combined evaluation of the Think-Aloud method with the results of eye-tracking testing (despite the limitations and problems that accompanied the use of the Think-Aloud). For example, one respondent liked a certain map at first glance (mentioned during map viewing in the Think-Aloud record), but it was difficult for him to complete the task. Conversely, in a map that the respondent did not take any interest in at first and would be rated as average in the preference rating, the correct solution was much more accessible. Unfortunately, because this respondent needed to concentrate on solving the task and did not attempt to comment on the process, it was not possible to substantiate this claim with statistical indicators.

The differences between experts and novices are evident from the evaluation of the experiment. The group of experts worked much more efficiently and could find the correct answers to the required tasks. The differences between respondents were also visible in the Sequence Chart evaluation (Figure 9), where it is clear that experts moved their eyes with more concentration on the goal and did not revise or search. Despite the striking differences in the individual evaluation methods compared to the similarity of fixation strings, novices and experts did not differ significantly. No significant differences in trajectories and movements between AOI areas were found (Figure 18).

Evaluation of the dynamic section of eye-tracking testing clearly showed that respondents had a complex map composition problem, mainly in that controls were on different sides of the map field rather than in one place. This problem arose, though, only during the first use of a web map. As soon as the respondents learned a map's functionality, they found this map element easily. Assessing the factors influencing a new user is very different from assessing the factors influencing an experienced user. This was detected while respondents were monitored as they worked with the Windy web map. At first, respondents had great difficulty finding the required controls, as the elements were distributed along the sides of the map field and set very far apart. In the final task, users no longer demonstrated the problem of finding map composition elements and used them more efficiently than in the first task.

The user aspect was mainly measured as user-friendliness and showed how a respondent felt while using a web map, what worked best for the respondent and what their preferences were. This assessment was subjective and very much depended on the respondent and their habits. This user aspect is closely related to all other user issues mentioned above. During the Think-Aloud assessment, some respondents mentioned the map that was best for them to control and which one they would like to use. Testing also showed that the concept of user comfort introduces the notion of intuitive map control and modern map design. Some respondents did not need a modern design, but they required functionality, simplicity and high-speed web map loading. For this reason, it was very complicated to evaluate the user aspect. Testing showed, however, that if a web map did not contain modern visualization elements, had very complex layouts and was very slow to load, it was very inconvenient to the respondent (for example, the Wundermap web map).

Lastly, graphic design significantly influenced respondents and their work with the web maps. Modern depiction enhances the attractiveness of maps and empowers a user's vision, even if they do not have flawless control and cartographic visualization methods are sometimes incorrect. Graphic map design, therefore, adds to the overall impression of a web map, portal or application. Respondents identified the Windy web map as attractive, but after the final task, some described this webpage as excessively detailed and cluttered with unnecessary information and suggested the possibility of changing thematic content. Most respondents identified the In-Počasí web map as balanced in map content and graphic design.

4. Discussion

This study assessed selected aspects of weather maps and focused on the degree of interactivity of these maps and user perception. Several works have already evaluated web maps, but only in a few cases at the level of user interpretation, perception, and cognition or general analysis of selected web maps.

The evaluated maps were selected based on an online survey, which was used to garner information on the most frequently used maps and adding a selection of different map types (known but less used maps). As most of the respondents had used international web resources in their work and personal life, the selection included the very frequently used weather web map YR.no. Another important aspect considered was that the respondents in the present study would be of Czech nationality. Therefore, the frequently used Czech weather web map In-Počasí, which includes only the territory of the Czech Republic, was included in the selection. Another Czech weather web map included was Windy map, developed by the owner of the most popular web map application in the Czech Republic Mapy.cz. The final maps selected were the Wundermap map and the DarkSky map, because their interfaces differ from the interfaces of the other maps.

The stimuli were presented in a fixed order because the analysis of dynamic stimuli combined with random order would be problematic. The analysis of dynamic stimuli will be very problematic when they will be randomized. In the static section of the experiment, it is possible to randomize the stimuli, but we did not do so in order to remain consistent within the experimental structure. We hope that the learning effect did not affect results, since different maps have different control mechanisms and are use different cartographic methods.

The eye-tracking experiment dataset was also analysed with Sequence Charts, using dynamic areas of interest. For the analysis, only six respondents (three experts and three novices) were selected. This was due to the clarity of the resulting visualization and extreme demands on time for creating dynamic areas of interest. The authors are aware that viewing the order of visited areas of interest for all 34 respondents might be interesting, but it would be necessary to manually create dynamic areas of interest for all respondents and all stimuli. However, this question may be a part of future research that could address weather web maps and their use.

User issues in map creation are determined by the target users and represent the most significant influence in the process of geovisualization. Therefore, considerable attention is addressed toward the user's needs, requirements, and preferences. Experiments, as presented in this article, allow inspecting in more detail the specifics that relate to different types of maps. Closely related topics enable detailed analysis of the experimental data and permit to draw relevant conclusions.

It is necessary to evaluate geovisualizations not only in terms of the correctness of the methods used and their compliance to cartographic principles, but also in their aesthetics and the user perception and interpretation of perceived information. The results above demonstrate that user preferences and user needs can be different. This conclusion is based on the Think-Aloud data analysis. The research outcomes show that it is crucial to implement map user testing into the geovisualization process, including a functional evaluation of interactive maps.

5. Conclusions

Weather maps were evaluated by combining research methods with a core eye-tracking experiment that focused on analysing the behaviour of respondents as they worked with the selected maps. The experiment was divided into three parts: a free viewing section, a dynamic section, and a static section. Five selected web maps with meteorological themes were employed in testing. Thirty-four respondents performed the test, separated into two map user groups of experts and novices.

The main aim of the presented research was to find out how users work with selected weather web maps. There are many map characteristics and parameters that affect the metrics being evaluated. All weather web maps are complex cartographic works; they differ in map composition, map symbology, map interactivity, map content, etc. Therefore, it is not possible to conclude which weather web maps were the best and worst overall. It can only be concluded that some maps are easier to understand and use (Windy, In-Počasí, YR.no) and some maps are not (Wundermap).

Partial results are presented in the task evaluation (Section 3.2). The acquired knowledge can be used to further discussion of weather web maps and their implementation. Our results include the findings that if expandable control panels were available, the respondents only looked at them after they had examined other elements; map interactivity was not an obstacle unless it contained too much information or too many options to choose from; searching was quicker in static menus that respondents did not have to switch on or off; and that the Think-Aloud method has significant limits in the case of dynamic testing due to high user demands.

Each web map is different, and both major and minor differences were identified. Further related research may focus on the impact of these differences on the user perception and cognition. Analysis can also be focused on different thematic maps and, thus, differences in attitudes of experts and general public (novices) can be evaluated. To that need, one of the planned future experiments will focus on the analysis of web maps intended for archaeologists.

Author Contributions: Conceptualization, Alena Vondrakova; Methodology, Stanislav Popelka and Petra Hujnakova; Data collection, Petra Hujnakova; Investigation, Petra Hujnakova and Stanislav Popelka; Writing—original draft preparation, Petra Hujnakova, Stanislav Popelka and Alena Vondrakova; Writing—review and editing, Stanislav Popelka and Alena Vondrakova.

Funding: The APC was funded by the Czech Science Foundation (Grant No.: 18-05432S).

Acknowledgments: The work on the paper was supported by the Ministry of Culture of the Czech Republic (NAKI II research programme; project No. DG18P02OVV017) and Czech Science Foundation (Grant No.: 18-05432S). Authors would like to thank for all respondents of the experiment.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Haklay, M.; Singleton, A.; Parker, C. Web mapping 2.0: The neogeography of the GeoWeb. *Geogr. Compass* **2008**, *2*, 2011–2039. [CrossRef]
2. Roth, R.E.; Donohue, R.G.; Sack, C.M.; Wallace, T.R.; Buckingham, T. A process for keeping pace with evolving web mapping technologies. *Cartogr. Perspect.* **2015**, 25–52. [CrossRef]
3. Peterson, M.P. *Maps and the Internet*; Elsevier: Amsterdam, The Netherlands, 2003.
4. Kraak, M.-J.; Brown, A. *Web Cartography: Developments and Prospects*; Taylor & Francis: New York, NY, USA, 2001.
5. Nivala, A.-M.; Brewster, S.; Sarjakoski, T.L. Usability evaluation of web mapping sites. *Cartogr. J.* **2008**, *45*, 129–138. [CrossRef]
6. Harrower, M.; Keller, C.P.; Hocking, D. Cartography on the Internet: Thoughts and a preliminary user survey. *Cartogr. Perspect.* **1997**, 27–37. [CrossRef]

7. MacEachren, A.M.; Boscoe, F.P.; Haug, D.; Pickle, L.W. Geographic visualization: Designing manipulable maps for exploring temporally varying georeferenced statistics. In Proceedings of the IEEE Symposium on Information Visualization, Triangle Park, NC, USA, 19–20 October 1998; pp. 87–94.
8. Arleth, M. Problems in Screen Map Design. In Proceedings of the 19th International Cartographic Conference, Ottawa, ON, Canada, 14–21 August 1999; pp. 849–857.
9. Schnur, S.; Bektaş, K.; Çöltekin, A. Measured and perceived visual complexity: A comparative study among three online map providers. *Cartogr. Geogr. Inf. Sci.* **2018**, *45*, 238–254. [CrossRef]
10. Wu, H.; Li, Z.; Zhang, H.; Yang, C.; Shen, S. Monitoring and evaluating the quality of Web Map Service resources for optimizing map composition over the internet to support decision making. *Comput. Geosci.* **2011**, *37*, 485–494. [CrossRef]
11. Alacam, Ö.; Dalci, M. A Usability Study of WebMaps with Eye Tracking Tool: The Effects of Iconic Representation of Information. In *New trends in Human-Computer Interaction*; Springer: Berling, Germany, 2009; pp. 12–21.
12. You, M.; Chen, C.-W.; Liu, H.; Lin, H. A usability evaluation of web map zoom and pan functions. *Int. J. Des.* **2007**, *1*, 15–25.
13. Buscher, G.; Dumais, S.T.; Cutrell, E. The Good, the bad, and the Random: An Eye-Tracking Study of Ad Quality in Web Search. In Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Athens, Greece, 24–28 July 2000; pp. 42–49.
14. Juřík, V.; Herman, L.; Kubíček, P.; Stachoň, Z.; Šašinka, Č. Cognitive aspects of collaboration in 3d virtual environments. *Int. Archives Photogram. Remote Sens. Spatial Inf. Sci.* **2016**, *41*, 663. [CrossRef]
15. Šašinka, Č.; Morong, K.; Stachoň, Z. The Hypothesis Platform: An Online Tool for Experimental Research into Work with Maps and Behavior in Electronic Environments. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 1–22.
16. Coltekin, A.; Heil, B.; Garlandini, S.; Fabrikant, S.I. Evaluating the effectiveness of interactive map interface designs: A case study integrating usability metrics with eye-movement analysis. *Cartogr. Geogr. Inf. Sci.* **2009**, *36*, 5–17. [CrossRef]
17. Dong, W.; Liao, H.; Xu, F.; Liu, Z.; Zhang, S. Using eye tracking to evaluate the usability of animated maps. *Sci. China Earth Sci.* **2014**, *57*, 512–522. [CrossRef]
18. Burian, J.; Popelka, S.; Beitlova, M. Evaluation of the Cartographical Quality of Urban Plans by Eye-Tracking. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 192. [CrossRef]
19. Herman, L.; Popelka, S.; Hejlova, V. Eye-tracking Analysis of Interactive 3D Geovisualization. *J. Eye Movement Res.* **2017**, *10*, 1–15. [CrossRef]
20. Kurzhals, K.; Weiskopf, D. Space-Time Visual Analytics of Eye-Tracking Data for Dynamic Stimuli. *Vis. Comput. Gr. IEEE Trans.* **2013**, *19*, 2129–2138. [CrossRef] [PubMed]
21. Ramloll, R.; Trepagnier, C.; Sebrechts, M.; Beedasy, J. Gaze data visualization tools: Opportunities and challenges. In Proceedings of the Eight International Conference on Information Visualisation, London, UK, 11–14 July 2017; pp. 173–180.
22. Tsang, H.Y.; Tory, M.; Swindells, C. eSeeTrack—Visualizing Sequential Fixation Patterns. *Vis. Comput. Gr. IEEE Trans.* **2010**, *16*, 953–962. [CrossRef] [PubMed]
23. Hegarty, M.; Canham, M.S.; Fabrikant, S.I. Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *J. Exp. Psychol. Learn. Memory Cognit.* **2010**, *36*, 37. [CrossRef]
24. Blok, C. *Chapter 11: Web maps and weather. Web Cartography*; Taylor & Francis: London, UK, 2003; pp. 145–162.
25. Opach, T.; Popelka, S.; Dolezalova, J.; Rod, J.K. Star and Polyline Glyphs in a Grid Plot and on a Map Display: Which Perform Better? *Cartogr. Geogr. Inf. Sci.* **2017**, 1–20. [CrossRef]
26. Roth, R.E.; Çöltekin, A.; Delazari, L.; Filho, H.F.; Griffin, A.; Hall, A.; Korpi, J.; Lokka, I.; Mendonça, A.; Ooms, K. User studies in cartography: Opportunities for empirical research on interactive maps and visualizations. *Int. J. Cartogr.* **2017**, *3*, 61–89. [CrossRef]
27. Nossum, A.S. Semistatic animation–Integrating past, present and future in map animations. *Cartogr. J.* **2012**, *49*, 43–54. [CrossRef]
28. Harrower, M. The cognitive limits of animated maps. *Cartogr. Int. J. Geogr. Inf. Geovis.* **2007**, *42*, 349–357. [CrossRef]
29. Maggi, S.; Fabrikant, S.I.; Imbert, J.-P.; Hurter, C. How do display design and user characteristics matter in animations? An empirical study with air traffic control displays. *Cartogr. Int. J. Geogr. Inf. Geovis.* **2016**, *51*, 25–37. [CrossRef]

30. Opach, T.; Nossum, A. Evaluating the Usability of Cartographic Animations with Eye-Movement Analysis. In Proceedings of the 25th International Cartographic Conference, Paris, France, 3–8 July 2011; pp. 1–11.
31. Ishikawa, T.; Barnston, A.G.; Kastens, K.A.; Louchouarn, P.; Ropelewski, C.F. Climate forecast maps as a communication and decision-support tool: An empirical test with prospective policy makers. *Cartogr. Geogr. Inf. Sci.* **2005**, *32*, 3–16. [CrossRef]
32. Kerpedjiev, S.M. Transformation of weather forecasts from textual to cartographic form. *Comput. Phys. Commun.* **1990**, *61*, 246–256. [CrossRef]
33. Herman, L.; Juřík, V.; Stachoň, Z.; Vrbík, D.; Russnák, J.; Řezník, T. Evaluation of User Performance in Interactive and Static 3D Maps. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 415. [CrossRef]
34. Fabrikant, S.I.; Lobben, A. Introduction: Cognitive issues in geographic information visualization. *Cartogr. Int. J. Geogr. Inf. Geovis.* **2009**, *44*, 139–143. [CrossRef]
35. Staněk, K.; Friedmannová, L.; Kubíček, P.; Konečný, M. Selected issues of cartographic communication optimization for emergency centers. *Int. J. Digital Earth* **2010**, *3*, 316–339. [CrossRef]
36. Field, K. *Cartography*; Esri Press: Redlands, CA, USA, 2018; 576p.
37. Roth, R.E. Cartographic interaction primitives: Framework and synthesis. *Cartogr. J.* **2012**, *49*, 376–395. [CrossRef]
38. Vondrakova, A.; Popelka, S. The use of eye-tracking for the evaluation of various cartographic tasks. In Proceedings of the Geoconference on Informatics, Geoinformatics and Remote Sensing, Albena, Bulgaria, 17–26 June 2014; Volume III, pp. 981–988.
39. Voženílek, V.; Kaňok, J. *Metody Tematické Kartografie: Vizualizace Prostorových Jevů*; Univerzita Palackého v Olomouci pro katedru geoinformatiky: Olomouc, Czech Republic, 2011.
40. Muehlenhaus, I. *Web Cartography: Map Design for Interactive and Mobile Devices*; CRC Press: Boca Racaton, FL, USA, 2013.
41. Lewis, J.R. Evaluation of procedures for adjusting problem-discovery rates estimated from small samples. *Int. J. Human-Comput. Interact.* **2001**, *13*, 445–479. [CrossRef]
42. Popelka, S. Optimal eye fixation detection settings for cartographic purposes. In Proceedings of the 14th SGEM GeoConference on Informatics, Geoinformatics and Remote Sensing, Albena, Bulgaria, 17–26 June 2014; pp. 705–712.
43. Dolezalova, J.; Popelka, S. ScanGraph: A Novel Scanpath Comparison Method Using Visualisation of Graph Cliques. *J. Eye Mov. Res.* **2016**, *9*. [CrossRef]
44. Goldberg, J.H.; Kotval, X.P. Computer interface evaluation using eye movements: Methods and constructs. *Int. J. Ind. Ergon.* **1999**, *24*, 631–645. [CrossRef]
45. Goldberg, J.H.; Stimson, M.J.; Lewenstein, M.; Scott, N.; Wichansky, A.M. Eye tracking in web search tasks: Design implications. In Proceedings of the 2002 Symposium on Eye Tracking Research & Applications, New Orleans, LA, USA, 25–27 March 2002; pp. 51–58.
46. Aanstoos, C. The think aloud method in descriptive research. *J. Phenomenol. Psychol.* **1983**, *14*, 243. [CrossRef]
47. Popelka, S.; Brychtova, A.; Brus, J. *Advanced Map Optimalization Based on Eye-Tracking*; IntechOpen: London, UK, 2012; p. 38305.
48. Vondráková, A.; Voženílek, V. User preferences in image map using. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B4*, 599–603. [CrossRef]

Paper Author-Reader

Beitlova, M., **Popelka, S.***, Konopka, M., & Macku, K. (2022). Verification of cartographic communication models using detection of map reading strategies based on eye movement recording. The Cartographic Journal. https://doi.org/10.1080/00087041.2022.2150362

[IF 1.0; Q4]

[0 citations on WoS; 0 citations without autocitations]

[Author's contribution: 40%]

Abstract

The aim of this paper is to introduce a method that can be used for the verification of cartographic communication models. The authors of the maps were placed in the role of the users of their maps. Then, eye-tracking was utilized to reveal their map reading strategies and compare them with the strategies of other cartographers and cartographic novices. The crucial part of the data analysis was scanpath comparison using the sequence of visited Areas of Interest, which helped quantify map reading strategies' similarity. The use of the same strategy as the map author used might be a prerequisite for users' proper understanding of a map reflected by the overlap of the author's and users' realities in Kolacny's model. The overlap was considerable in most cases; however, exceptions in which authors used a different map reading strategy were identified.

THE CARTOGRAPHIC JOURNAL
2023, VOL. 60, NO. 1, pp. 5–24
https://doi.org/10.1080/00087041.2022.2150362

REFEREED PAPER

Verification of Cartographic Communication Models Using Detection of Map Reading Strategies Based on Eye Movement Recording

Marketa Beitlova, Stanislav Popelka, Martin Konopka and Karel Macku

Marketa Beitlova, Stanislav Popelka, Department of Geoinformatics, Palacký University Olomouc, Olomouc, Czech Republic
Martin Konopka, Kempelen Institute of Intelligent Technologies, Bratislava, Slovakia
Karel Macku, Department of Geoinformatics, Palacký University Olomouc, Olomouc, Czech Republic

ABSTRACT
The aim of this paper is to introduce a method that can be used for the verification of cartographic communication models. The authors of the maps were placed in the role of the users of their maps. Then, eye-tracking was utilized to reveal their map reading strategies and compare them with the strategies of other cartographers and cartographic novices. The crucial part of the data analysis was scanpath comparison using the sequence of visited Areas of Interest, which helped quantify map reading strategies' similarity. The use of the same strategy as the map author used might be a prerequisite for users' proper understanding of a map reflected by the overlap of the author's and users' realities in Koláčný's model. The overlap was considerable in most cases; however, exceptions in which authors used a different map reading strategy were identified.

ARTICLE HISTORY
Received 7 July 2021
Accepted 17 November 2022

KEYWORDS
Cognitive cartography; cartographic communication models; map reading; eye-tracking; scanpath comparison; expert-novice

Introduction

The development of society is realized through the constant increase in the amount and depth of knowledge, based on the evolution and improvement of communication processes. These processes consist of an endless cycle of creating, transmitting, and utilizing information, which becomes the raw material of thought. Cartographic works form an important part of society-wide information sources (Hojovec *et al.*, 1987).

The first documented ideas about applying psychology to map perception appeared in the first half of the twentieth century (Wright, 1942). Nevertheless, it took ten more years until Arthur Robinson's 1952 publication, *The Look of Maps*, provided the impetus for cartographers to explore the process of working with maps in a way that was grounded in psychology and from a user-driven perspective.

According to Robinson (1952), the cartographer's primary task is to provide information to the map user. The efficiency of message transmission depends on the method by which it is presented. To make map communication more effective, deeper knowledge about cartographic visualization methods as well as a means of assessing their impact on the map user are necessary. Robinson's thoughts have been followed by many research studies on the perception of symbols used on maps (Castner, 1964; Wood, 1968; Crawford, 1973; Potash, 1977; Clarke, 1989; Griffin, 2014; Voženílek *et al.*, 2014; Kubíček *et al.*, 2017; Opach *et al.*, 2018; Stachoň *et al.*, 2019; Klettner, 2020; Stachoň *et al.*, 2020).

With the gradually increasing influence of psychology on cartography, it is becoming clear that cognitive processes play an important role in our interaction with maps. As the primary consumer of cartographic information, the user is placed at the forefront of interest. The use of eye-tracking in cognitive science is well established and it has become a popular method for investigating research questions related to spatial cognition, GIScience and cartography (Kiefer *et al.*, 2017). With the use of this technology, it is possible to reveal the details of the cartographic communication process. These details may aid in the subsequent optimization of cartographic works and the evaluation of map users' cartographic skills. Eye-tracking provides insight into user behaviour during the information gathering process. Revealing map use strategies allows for observation of the process of interaction with the map and communication through the map. The process of cartographic communication is described using cartographic communication models.

Cartographic communication models

Cartographic communication models are a theoretical starting point for the research concerning the process of map reading. There are a number of them from different authors across different cartographic schools. In most of

CONTACT Stanislav Popelka stanislav.popelka@upol.cz

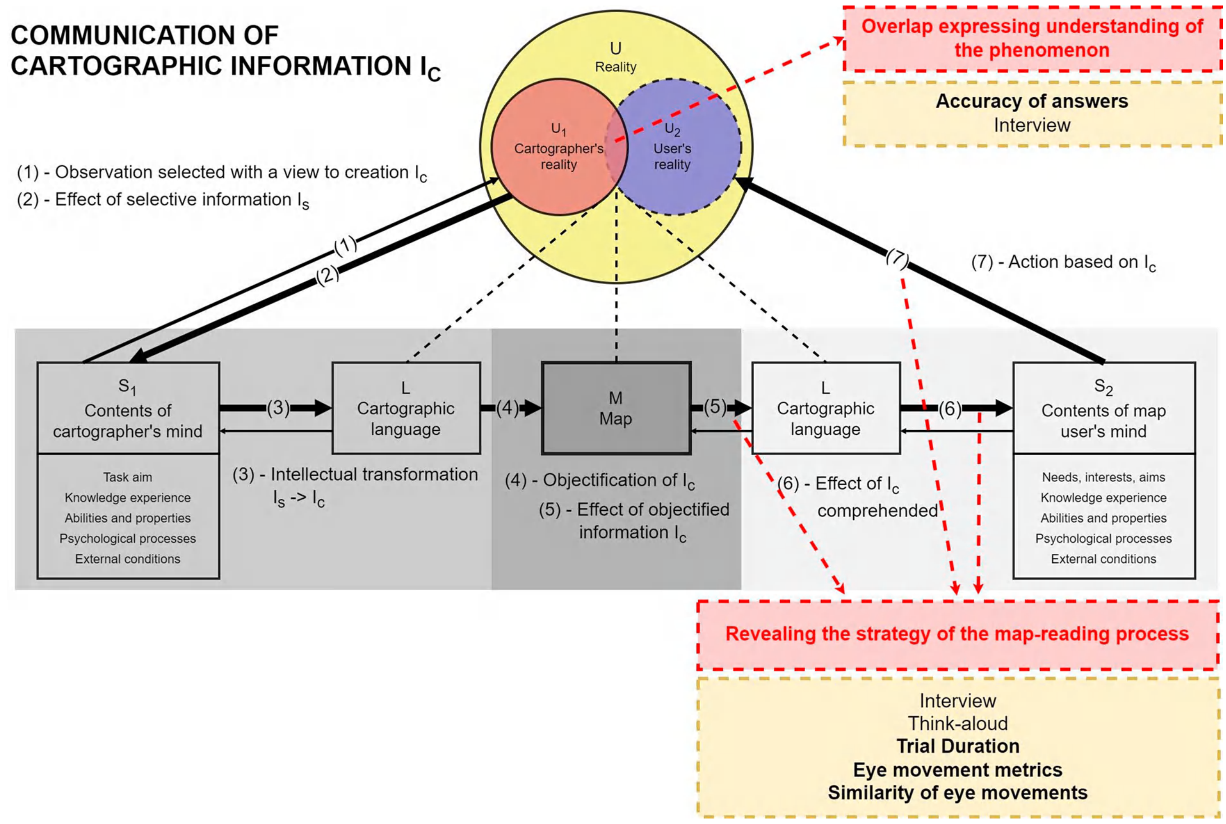


Figure 1. Cartographic communication model according to Koláčný (1969) and annotated by the authors.

these models, four common entities can be traced. The first of them is the author of the map – the cartographer. On the other side of the model, there is a map user as a target consumer of a map. The map serves as a communication instrument between these two entities. The last entity of this tetragon is the overlap of the map author's and map user's realities (understanding of a phenomenon).

In the early 1950s, Robinson (1952) pioneered research on cartographic communication and many authors later followed. Koláčný (1969), Morrison (1977), Board (1978) and Ratajski (1978) developed the first models of cartographic communication based on the theories of Shannon and Weaver (1948). Koláčný's (1969) model (Figure 1) specifies that the reality (Universum) to which the cartographic representation refers is not exactly the same reality for the cartographer (author) and the map user. For the cartographer, it refers to a part of reality U_1; for the map user, it is a part of reality U_2. The cartographer is represented in the diagram by the content of his knowledge S_1, which is influenced by his tasks, aims, knowledge and experience, abilities and other characteristics, his psychological processes and external conditions, i.e., environmental influences. The map user is represented analogously, by the content of his consciousness (S_2), his needs, interests and aims, his knowledge and experience, his abilities and other characteristics, his psychological processes and the external conditions of his environment. Both the cartographer and the map user know a cartographic language, i.e., a system of map marks and rules for their use, denoted by L. The map M is considered to be a system of map features that embody cartographic information, I_c.

The excessive simplicity of the model was later criticized (e.g. Olson, 1976; Robinson and Petchenik, 1976; Petchenik, 1977; MacEachren, 1995; Keates, 1996). The main reason for this critique was that knowledge is not transmitted in the sense of a closed packet carrying unchanged information from the transmitter to the receiver (Rieger, 1996). According to MacEachren (2004), an increasing number of scientists do not perceive the map as an objective representation of reality, which excludes the idea of objective research.

Robinson and Petchenik (1976) and Petchenik (1977) oppose systematic cartographic communication models based on information theory and offer a Venn diagram summarizing the cognitive elements in cartographic communication (Figure 2). The outer rectangle in Figure 2 defines the set of all conceptions of the geographical environment, and these conceptions can be either correct (S_C) or incorrect (S_E). Area A indicates a subset of the author's conceptions; area B, a subset of the map user's conceptions. The diagram shows the usual (desirable) state – the relative size and position of rectangles A and B show that the subset of the

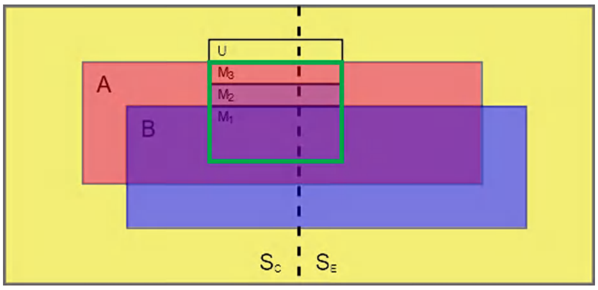


Figure 2. Venn diagram summarizing the cognitive elements in cartographic communication (Robinson and Petchenik, 1976) and annotated by the authors.

author's conceptions is larger, and a relatively larger part of it falls into the area of correct ideas. However, at the same time, the areas of the conceptions of author and map user overlap significantly.

The green rectangle defines the set of concepts marked by the author on the map. Area M_1 shows the part of the map's conceptions that were already known to the user. Area M_2 contains conceptions that the user did not know before and thus constitutes a direct increment to his spatial understanding. Area M_3 indicates a subset of concepts that the user did not understand from the map – the discrepancy between input and output of the communication system. According to Petchenik (1977), cartographic research should focus primarily on cases of cartographic communication failure connected with area M_3. Area U shows an unplanned increment of the user's spatial understanding that was neither intended nor symbolized in any way by the author.

According to King (1982), the research of Robinson and Petchenik reflects the trends of behavioural and cognitive psychology. It deals with the description and understanding of the processes by which the map reader creates his idea of the relationships between depicted objects and phenomena, based on the map.

The work of Robinson and Petchenik (1976) and later MacEachren (1995) is thus in contrast to the prior understanding of the map as a channel for the transmission of cartographic information. These authors understand the map as one of the possible representations of objects or phenomena in space, and their work emphasizes the need to study perceptual and cognitive processes during map reading and spatial information processing. MacEachren (1995) emphasizes the role of all aspects of maps, from the initial data gathering to their final visualization that enables communication. He also highlights the increase of the user's knowledge due to the usage of the map and integration of the information presented on the map. The fact that the development of cartographic communication models is still ongoing is demonstrated by Kent (2018), who offers a critical discussion of these models and presents a new version for cartographic communication in an age of social media.

Relationship between author and user

In cartographic communication models, the emphasis is often placed on the relationship between the author and the map user. Furthermore, the interaction between the map and its user is often analysed. However, previous works have not yet determined how maps are read by their authors and to what extent their approach is similar to map users' approach. In the presented experiment, the authors of the maps are placed in the role of the readers, and the strategy of their map reading is analysed. The assumption is that the authors are (or were at the time of map creation) familiar with the displayed phenomena in detail, so their map reading strategy should be straightforward. Thus, the experiment examines the similarity of the authors' map reading strategy compared with cartographers (participants with cartographic education) and compared with novices who are not familiar with the data and methods used on the maps. Thus, the experiment focuses on estimating the extent to which the author's and map reader's realities overlap (U_1 and U_2 in Koláčný's (1969) model, and A and B in Robinson and Petchenik's Venn diagram (1976)) and the way how the overlap was achieved.

The most straightforward way to measure the overlap between the author's and map reader's realities is to analyse the accuracy of answers. This analysis directly shows how the readers understand the map (at least to answer the given task), and the results can be easily quantified. Alternatively, an interview might be used to verify if the reader understands the map. More demanding is to uncover readers' strategy to obtain information from maps. It might be possible to employ qualitative methods like interviews or think-aloud to understand this process. The time needed to answer the given questions might also be a valuable source of information. However, the use of eye-tracking can give us much more detailed insight into the way how the readers work with the map. It allows us to monitor their visual attention and get information on where, when, and for how long they looked. Eye movement data can be analysed in several ways, from the qualitative (manual) evaluation of each scanpath to more sophisticated methods for quantifying eye movement's similarity. The latter approach was used in this research study. Dashed boxes in Figure 1 show how these measures are incorporated into Kolacny's model of cartographic communication.

The study is aimed at the verification of cartographic communication models and discovering whether there is a difference in map reading strategies among three groups of participants:

- authors of the maps (map creators);
- cartographers (students of cartography); and
- novices.

One of the contributions of this paper is the description of a method for the calculation of the similarity of participants' strategies based on the order of visited AOIs. This procedure is described in detail in the methods section.

Methods

Eye-tracking was used to analyse the behaviour of participants as they worked with maps and the approach is detailed in the sections below.

Experiment design

The experiment was designed as within-subject with group affiliation (author, cartographer, novice) as an independent variable. The accuracy of participants' answers, trial duration, and the eye movement metrics fixation count and dwell time were dependent variables. In addition, the task strategy – how the task was conducted – was analysed qualitatively (using scanpath visualization) as well as quantitatively (using scanpath comparison methods).

The hypothesis was that there would be differences in dependent variables between groups of participants and that the strategy for stimuli inspection would differ within these groups. If so, the cartographic models would be verified, since the map authors' reality would differ from the map readers'.

Participants

The data were captured in two phases which were conducted at different universities. The same set of experiments was used in both phases at both universities. The first data recording was conducted at the Department of Geoinformatics, Palacký University Olomouc, Czech Republic. In this phase, the eye-movement data of 22 participants marked as cartographers were recorded (14 males and 8 females). These were university students in their second year (~20 years old) of studying geoinformatics and geography, who were enrolled in two cartography courses. These students were also the authors of the analysed maps. Thus, one of them was always the author of one of the maps in each set. For example, participant C01 was an author of map M01. For all other maps, he was in the role of the map reader (cartographer, experienced mapper). For the map M02, participant C02 was in the role of the author (and in the role of the cartographers for all the others). See Figure 3 for the illustration.

The second phase aimed to compare the map reading strategy of cartographers with cartographic novices. In this phase, the participants were students enrolled at the Faculty of Informatics and Information Technologies at the Slovak University of Technology in Bratislava, and these students had no specific cartographic experience. A total of 17 novices participated in the experiment (17 males, 4 of which were undergraduate engineering students; the rest were PhD students). Each map was observed by its author ($n = 1$), the group of cartographers ($n = 21$) and the group of novices ($n = 17$).

Materials

The experiment contained a total of 44 maps. The experiment was divided into two parts – free viewing and task completion. Each of these parts contained 22 maps. All maps were created by students of geoinformatics during their first year of study as required classwork within their general cartography and thematic cartography classes.

In the free viewing segment, participants were not asked to perform any task, only to inspect the maps as they wished. This part of the experiment was used to reveal the importance of particular map composition elements and to discover whether the relative importance of these elements differs across the analysed groups of

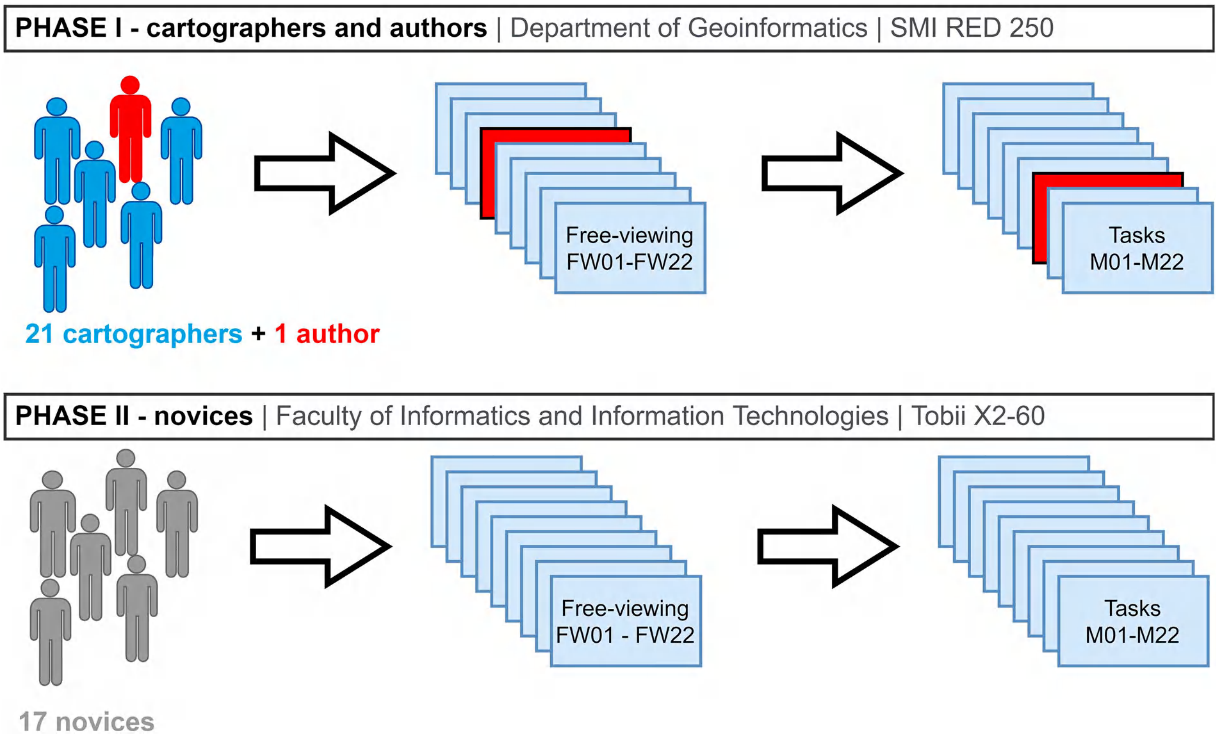


Figure 3. Summary of the design of the study.

participants. According to Opach *et al.* (2014), free viewing (free examination) can enable insight into how users intuitively direct their attention to the various map components, perhaps mostly driven by bottom-up (stimuli-driven) processes. Free viewing stimuli were previously used in the cartographic studies of Opach *et al.* (2014), Krassanakis *et al.* (2016), Popelka *et al.* (2019) and others. Moreover, free viewing stimuli are commonly used in saliency studies (Koehler *et al.*, 2014).

Maps for free viewing were reproductions of maps from different atlases (Figure 4). These maps were created by students in the general cartography classes as their semester projects six months before their participation in this eye-tracking experiment. Strictly speaking, they are not actual authors (due to reproducing). However, we expect they will behave differently than other groups because they already know these maps. Each of these maps was presented for 12 s.

The task completion part of the experiment contained 22 maps as well (Figure 5). In this case, the maps were thematic maps primarily created using statistical data (except M03, M14, M20, M21) as semester project in thematic cartography classes. These maps were created by the students 12 months before their participation in this experiment.

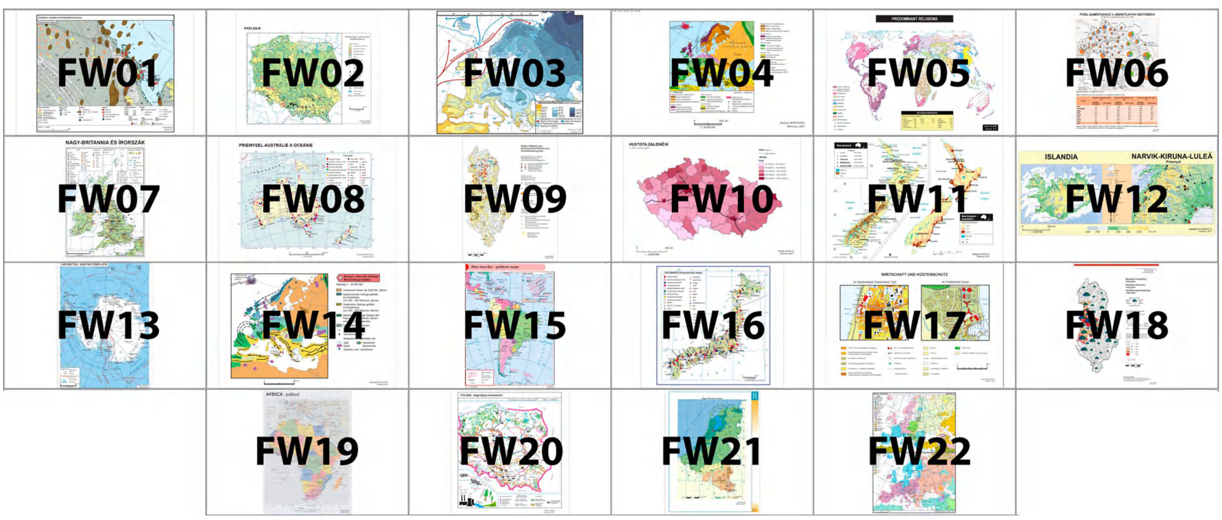


Figure 4. Stimuli used in the free viewing part of the experiment 10.17632/595g8mp82d.2.

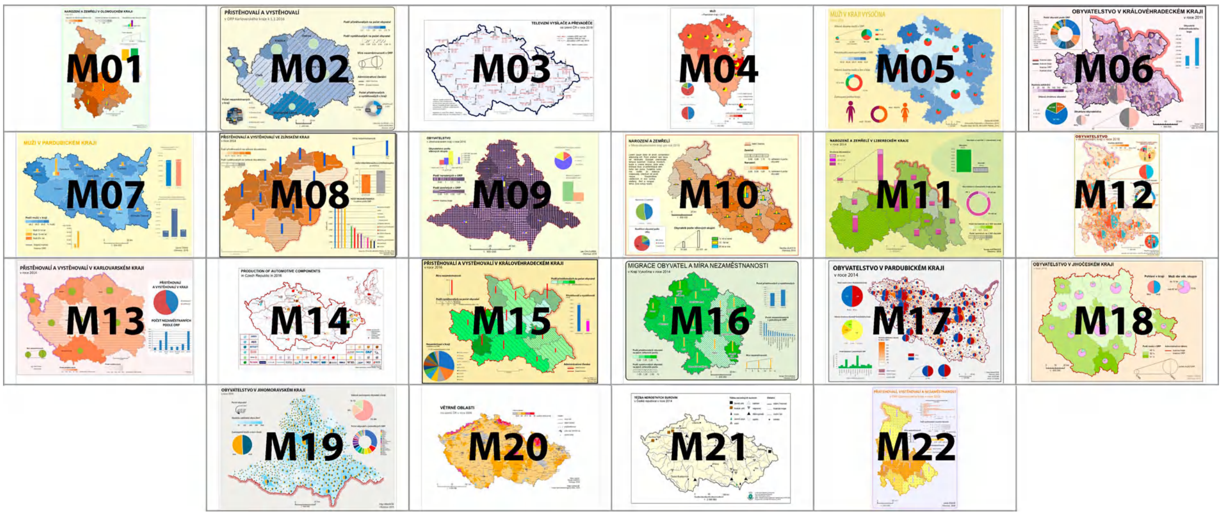


Figure 5. Stimuli used in the task completion part of the experiment 10.17632/595g8mp82d.2.

All stimuli were presented in random order on a screen with a resolution of 1920 × 1200 px. Answers in the task completion part of the experiment were provided via mouse click onto the map. Two different types of eye-trackers were used for the recording of eye-movement data. In the first phase, an SMI RED 250 eye-tracker with a sampling frequency 250 Hz was used. Data were visualized in the proprietary software made by SensoMotoric Instruments, SMI BeGaze and the open-source application OGAMA (Open Gaze and Mouse Analyzer) (Voßkühler *et al.*, 2008). In the second phase, data were recorded in the user-experience classroom at the User Experience and Interaction Research Center (UXI) (Bielikova *et al.*, 2018) equipped with 20 Tobii X2-60 eye-trackers with a sampling frequency of 60 Hz. Recorded data were again converted to OGAMA.

For both systems, point-by-point calibration was selected with nine calibration points. Fixations in both datasets (from the SMI eye-tracker as well as from the Tobii) were identified using the same I-DT (Identification of Dispersion Threshold) in OGAMA. The thresholds for fixation identification were defined according to Popelka (2014). The only difference was in the parameter 'minimum number of samples that can be considered as fixation' which is dependent on the sampling frequency. The threshold for the SMI (250 Hz) was selected to be 20 samples, whereas for the Tobii (60 Hz), it was 5 samples.

Procedure

As described above, participants were recruited at two universities in two different countries (the Czech Republic and Slovakia). They participated voluntarily and they did not receive any reward for their participation. Upon entering the eye-tracking laboratory, they were acquainted with the procedure and principle of eye movement

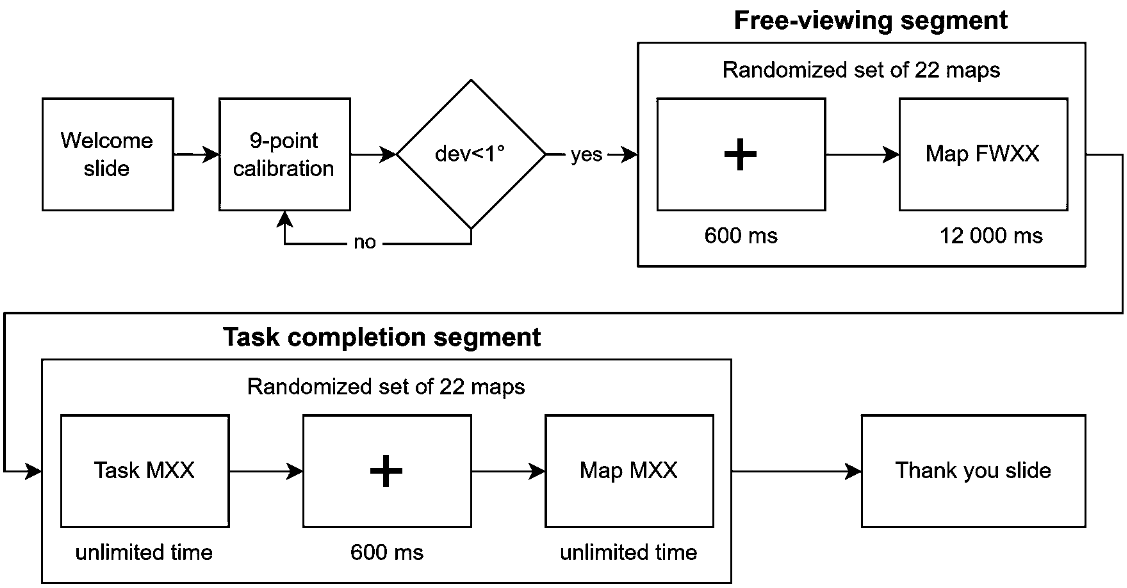


Figure 6. A scheme showing the eye-tracking experiment's procedure.

recording. Then, they were seated in front of the eye-tracker and the experiment started. See Figure 6 with the schematic outline of the experiment.

At the start of the experiment, a welcome slide was displayed (translation from the Czech original): 'Hello. Welcome to this experiment, which is focused on map reading. The experiment will be divided into two parts – free viewing and task completion. Here, you will mark answers via mouse click'. Then, calibration was begun. Calibration with a deviation of less than 1° was considered successful. Otherwise, the calibration process was repeated. After calibration, the free viewing segment of the experiment started. It contained 22 maps (FW01–FW22) displayed in random order. Each of them was displayed for 12 s. A fixation cross was displayed for 600 ms before the presentation of each map.

After the free viewing segment, the task completion part of the experiment started. Participants had unlimited time to read and remember the task. Then, after participants clicked the spacebar to indicate they were ready, a fixation cross was displayed for 600 ms, followed by a map (M01–M22). Participants had unlimited time for the completion of these tasks. The triplets of task information – fixation cross – map were also displayed in random order. In all tasks, participants were asked to identify specific information on the map according to the legend. The list of tasks is displayed in Table 1.

Methods of analysis

Qualitative as well as quantitative methods were used for the analysis of the recorded data. For the experiment's task-completion segment, the accuracy of answers was analysed, and the problematic tasks were thoroughly investigated using scanpath inspection or sequence chart visualization. Then, eye-tracking metrics were analysed statistically. The Kruskal–Wallis post hoc Nemenyi test was applied in RStudio at a significance level of 0.05. For the experiment's free viewing segment, Areas of Interest (AOIs) were marked around map composition elements, and the dwell time metric was analysed. Dwell time shows the percentage of time spent in defined Areas of Interest (AOIs). The final comparison of visual inspection strategies was made using ScanGraph software (Dolezalova and Popelka, 2016; Popelka *et al.*, 2018; Popelka and Beitlova, 2022). ScanGraph is an online tool using string-edit distance to calculate scanpath similarity using Levenshtein distance (Levenshtein, 1966). Usually, the use of ScanGraph starts with the selection of parameter p. The parameter p takes its value from the interval $\langle 0, 1\rangle$ and represents the desired degree of participants' similarity. Value 1 represents the absolute similarity – the sequences of visited AOIs were exactly the same. In contrast, value 0 means completely different sequence. After the selection of this parameter, individual participants are displayed as nodes in a simple graph. Groups of participants with a similar strategy (at a given level p) are visualized as cliques in this graph. A clique is a subset of vertices in a graph where all vertices are connected by an edge with all of the others from that subset (Gross and Yellen, 2005).

In this study, the parameter p was not used, so the minimal level of strategies' similarity was not defined and participants' cliques were not sought. Modified matrices were used instead. The modified matrix contains the normalized values of similarity. Specifically, the average similarity for all participants' pairs was calculated separately for the group of cartographers (C) and novices (N). The average similarity between the author and

Table 1. The complete list of tasks used in the experiment (translated from Czech).

M01	Identify the municipality with the 3rd highest number of inhabitants in the age group 15–64 years.
M02	Identify the region which has the lowest share of both immigrants and emigrants.
M03	Identify the location with an effective radiated power of the converter above 100 W.
M04	Identify the region with the lowest and then the region with the highest share of men in the total population.
M05	Identify the region with the lowest population.
M06	Identify a male-dominated municipality.
M07	Identify the region where males aged 0–14 years exceed men aged 65 and over.
M08	Identify the region with the highest unemployment rate.
M09	Identify the region with the highest share of both births and deaths.
M10	Identify the smallest municipality in the region.
M11	Identify the region with the lowest population under 15 years of age.
M12	Identify the area within the region which has the lowest population density.
M13	Identify the region with the highest number of emigrants per 1000 inhabitants.
M14	Identify all Benteler offices.
M15	Identify the region with the highest share of immigrants.
M16	Identify the region with the lowest unemployment rate.
M17	Identify the municipality with the highest population in the region.
M18	Identify the region with a 51% male share.
M19	Identify the area with the lowest population.
M20	Identify the area with a wind speed of 27.5–30 m/s.
M21	Identify the locations where kaolin is mined.
M22	Identify the region with the lowest share of emigrants per population.

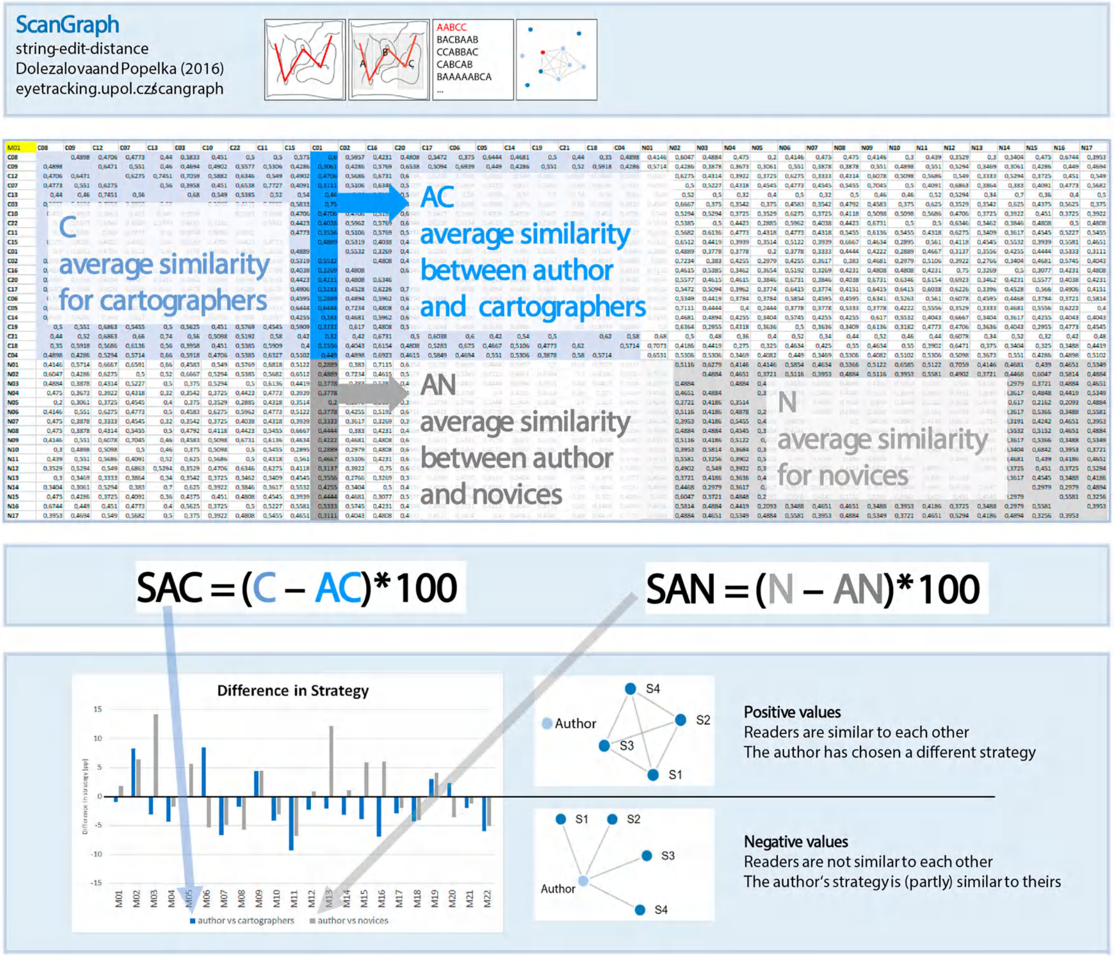


Figure 7. An overview of scanpath comparison.

all cartographers (AC) and novices (AN) was calculated. Finally, these two averages were subtracted to find out the similarity of the author's strategy to that of cartographers (SAC) and to that of novices (SAN):

$$SAC = (C - AC) \times 100$$

$$SAN = (N - AN) \times 100$$

This method enabled the identification of cases in which the author used a different strategy than the readers. The whole process of scanpath comparison is depicted in Figure 7. A similar approach was previously used in Beitlova *et al.* (2020).

Results

The results are divided into two main parts: task completion and free viewing. The accuracy of answers, eye-tracking metrics, and similarity of strategies were assessed.

Task completion

Accuracy of answers

To start, the accuracy of answers for three groups of participants was analysed for each map. The authors did not make any mistakes. The percentage of correct answers for the group of cartographers ranges between 59% and 100%, with an average of 91%. The group of novices was slightly less accurate. Their accuracy of answers for each map varied between 65% and 100%, but the average was 87%. The differences in the accuracy of answers were not statistically significant. The overview of answers for all maps is displayed in Figure 8.

The lowest level of accuracy was observed for map M19; the errors were due to ambiguities in the task phrasing. For that reason, this task was removed from the evaluation of answers' accuracy.

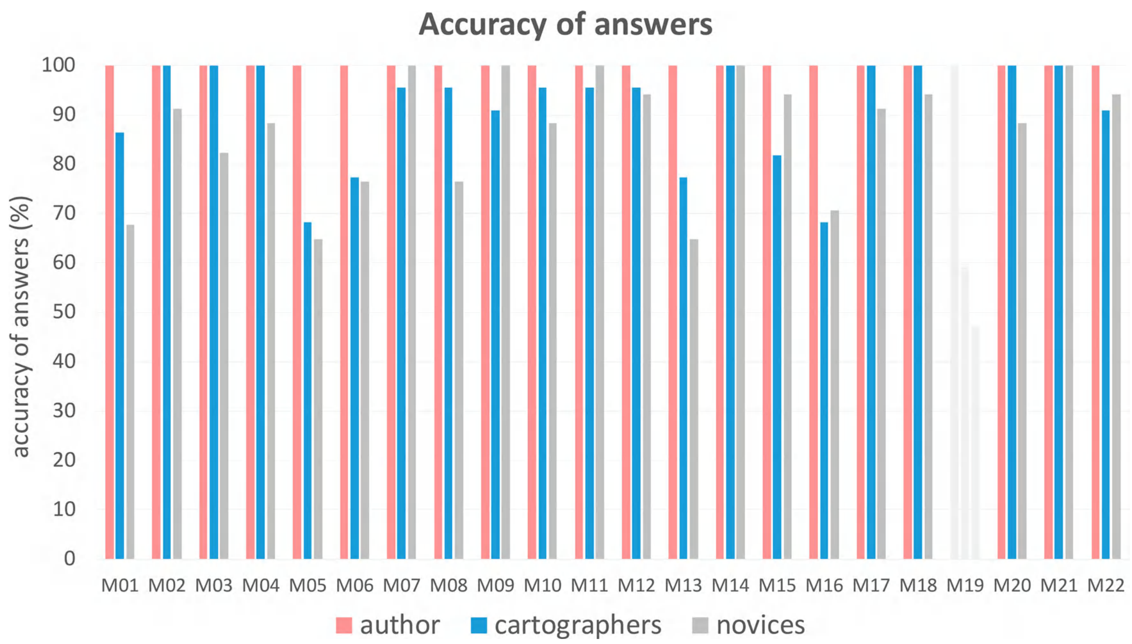


Figure 8. Accuracy of answers.

The highest numbers of incorrect answers were recorded for maps M05, M13 and M16. A deeper qualitative eye-tracking analysis helped to identify the reasons for these errors.

In map M05, the task was: 'Identify the regions with the lowest population'. This information was not explicitly stated in the legend, but it was possible to find the answer using the value scale of 'age groups in regions'. Participants probably did not realize this, because only a small percentage of fixations were aimed at this legend element (18.5% for cartographers and 22.1% for novices).

For map M13, the task was to identify the district with the highest number of emigrants per 1,000 inhabitants. The number of emigrants in each region was shown using horizontal lines; the closer the lines were, the higher number of emigrants per 1000 inhabitants (Figure 9). Cartographers' inattention probably

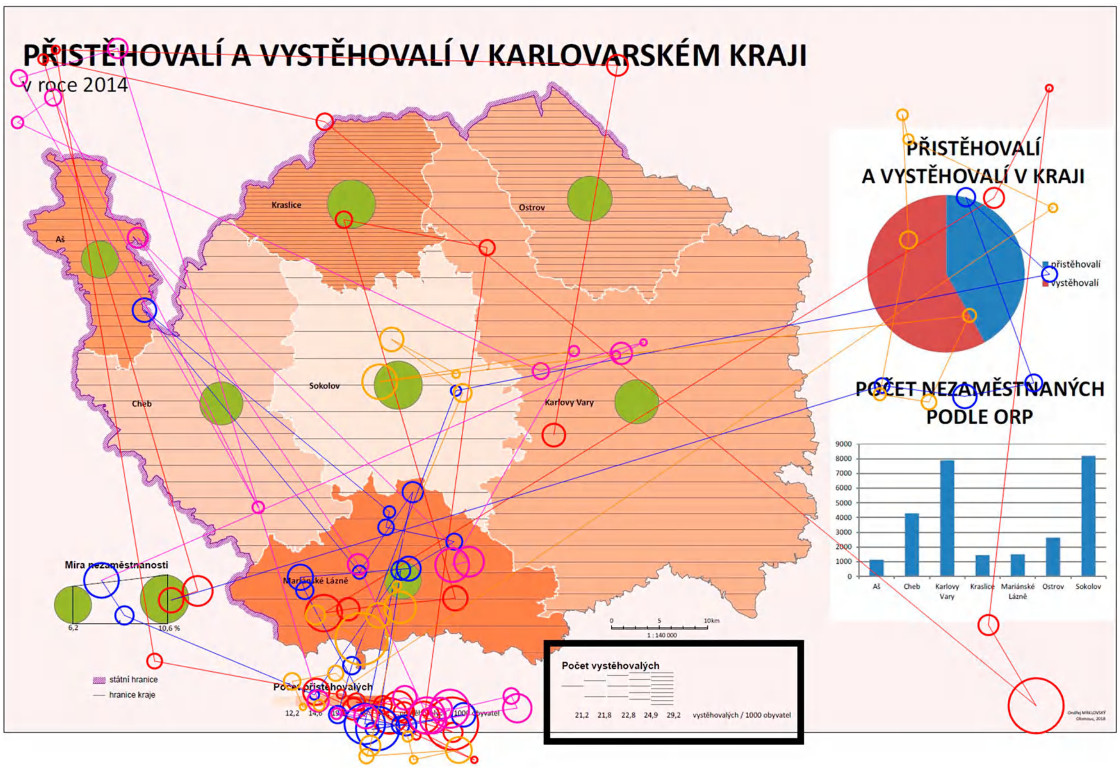


Figure 9. Scanpaths of four cartographers with an incorrect answer on M13 who did not look at the correct part of the legend (marked with a rectangle).

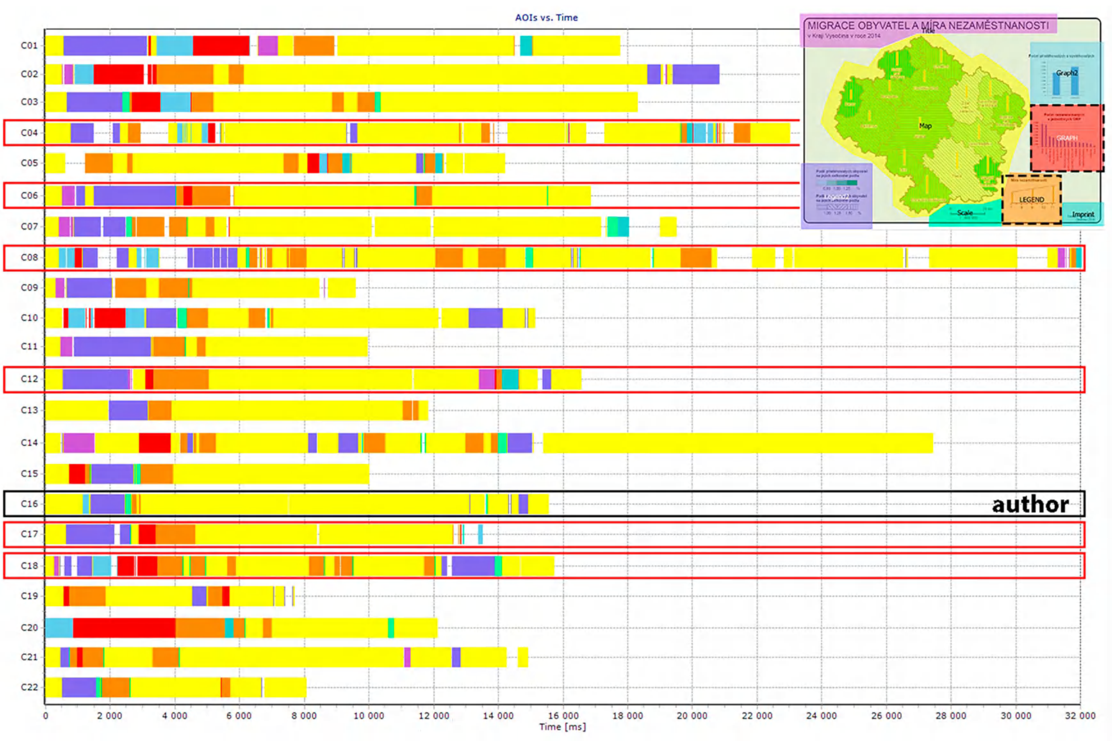


Figure 10. Sequence chart for map M16.

caused their higher error rate. Of five cartographers with an incorrect answer, four did not look at the correct legend at all. The situation for the group of novices was different. From six incorrect answers, only two participants did not fixate on this part of the legend.

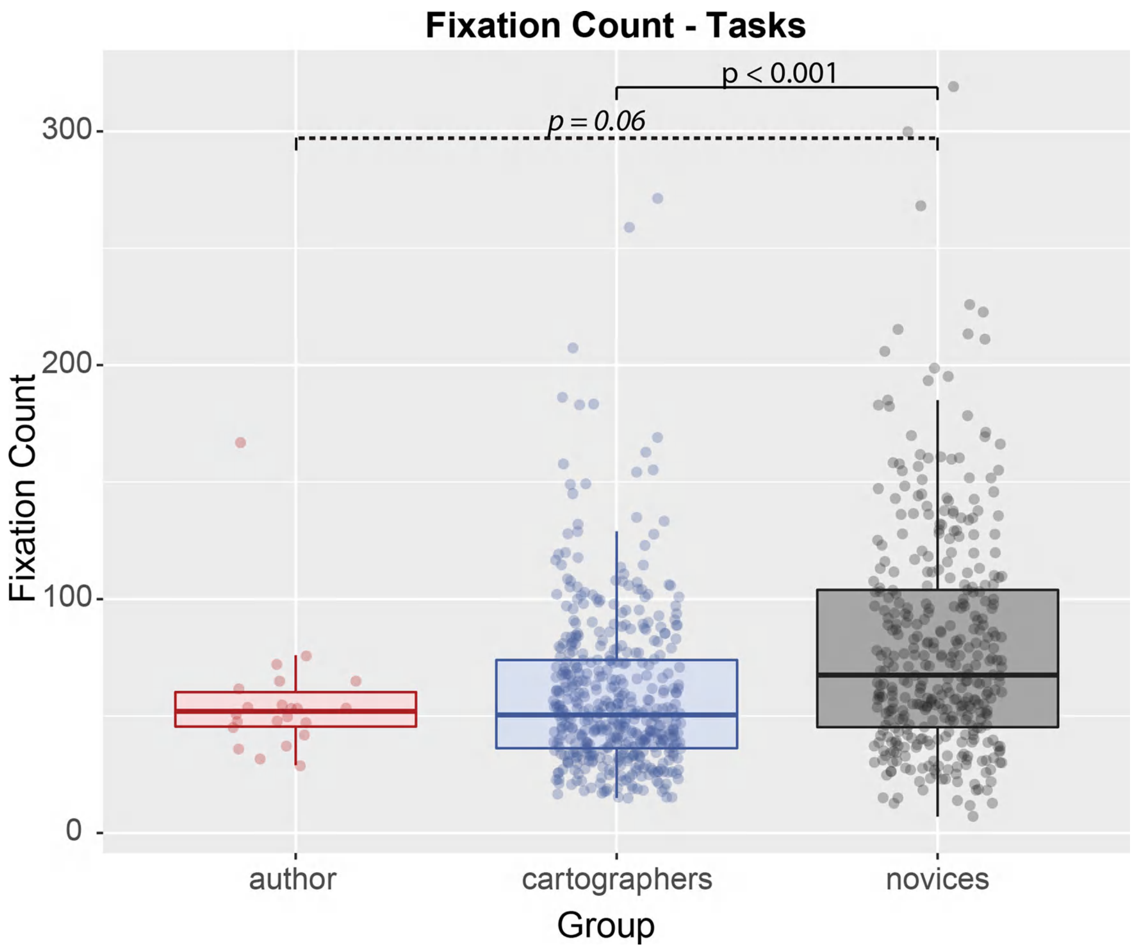


Figure 11. Fixation count for the three participant groups for the task completion part of the experiment.

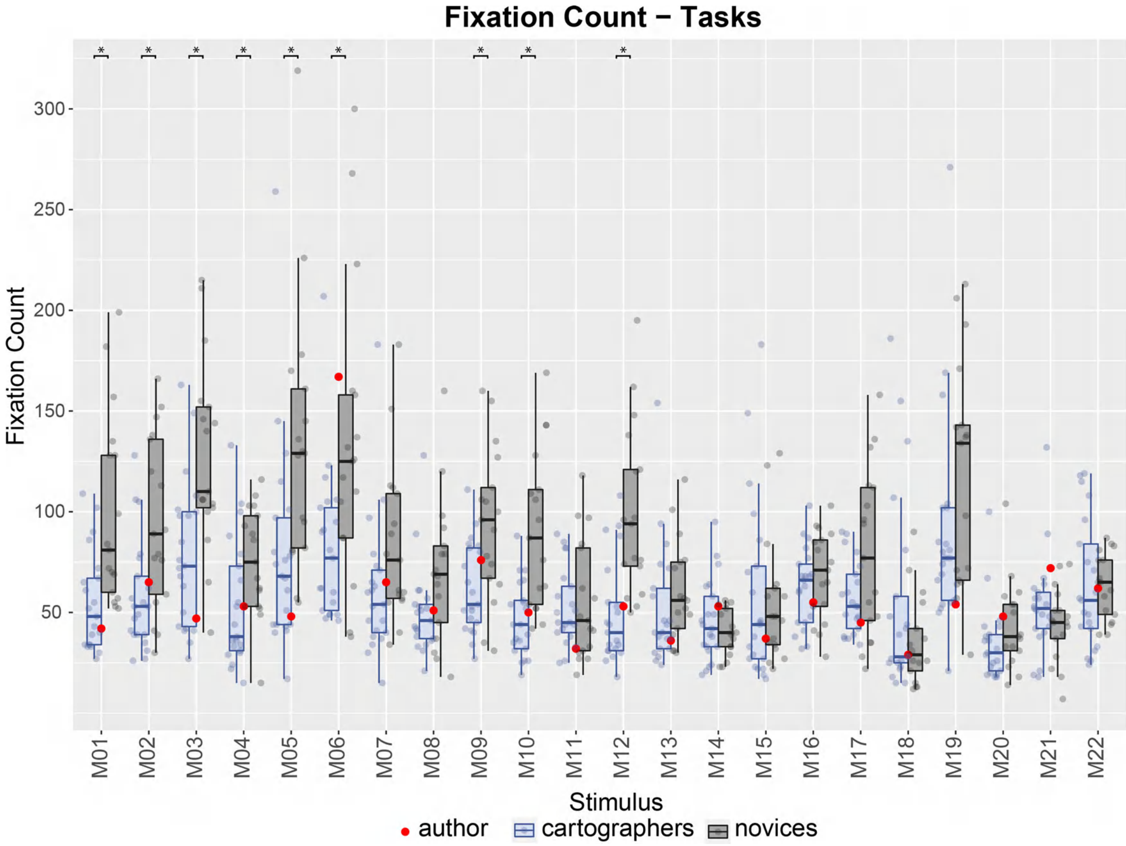


Figure 12. Fixation count for all tasks of the experiment.

In the case of map M16, the task was to identify the region with the lowest unemployment rate. The correct answer depended on the identification of the smallest columnal map diagram. The size of value scale was not chosen ideally, and the differences between columns were not sufficiently distinguishable. However, the

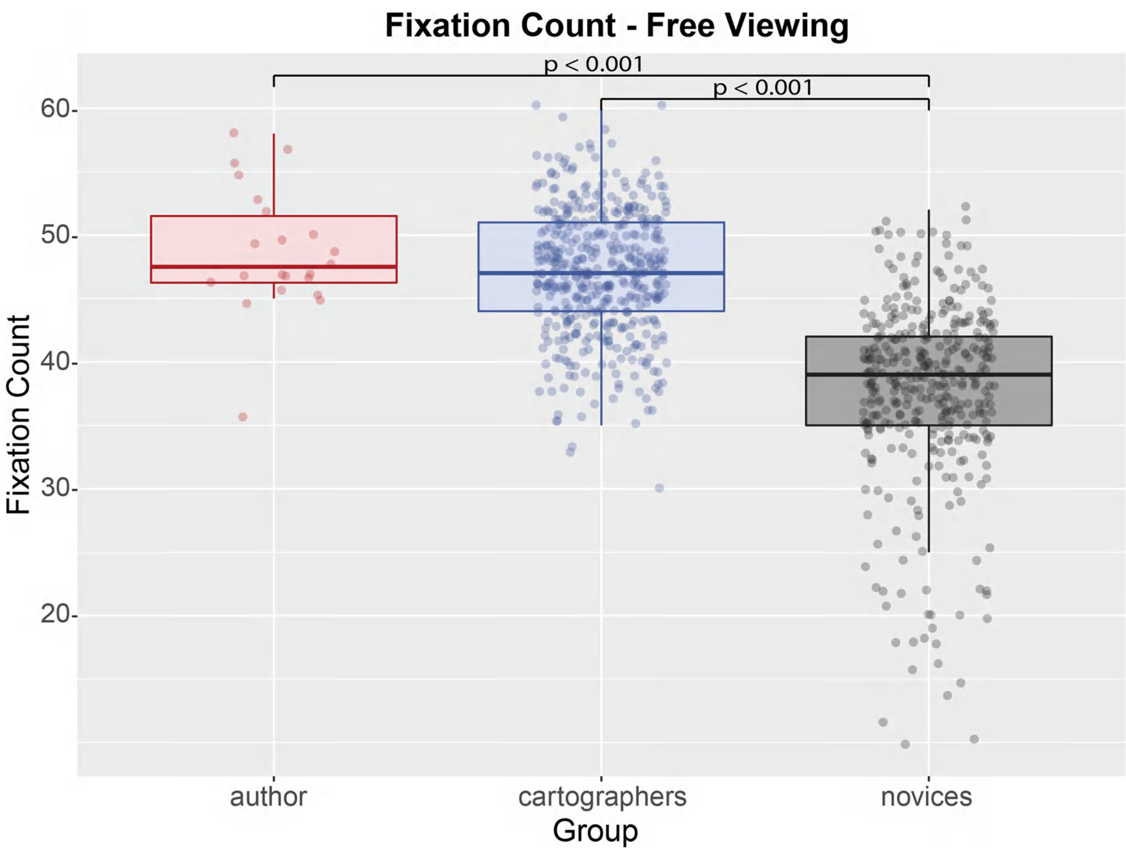


Figure 13. Fixation count for the free viewing part of the experiment.

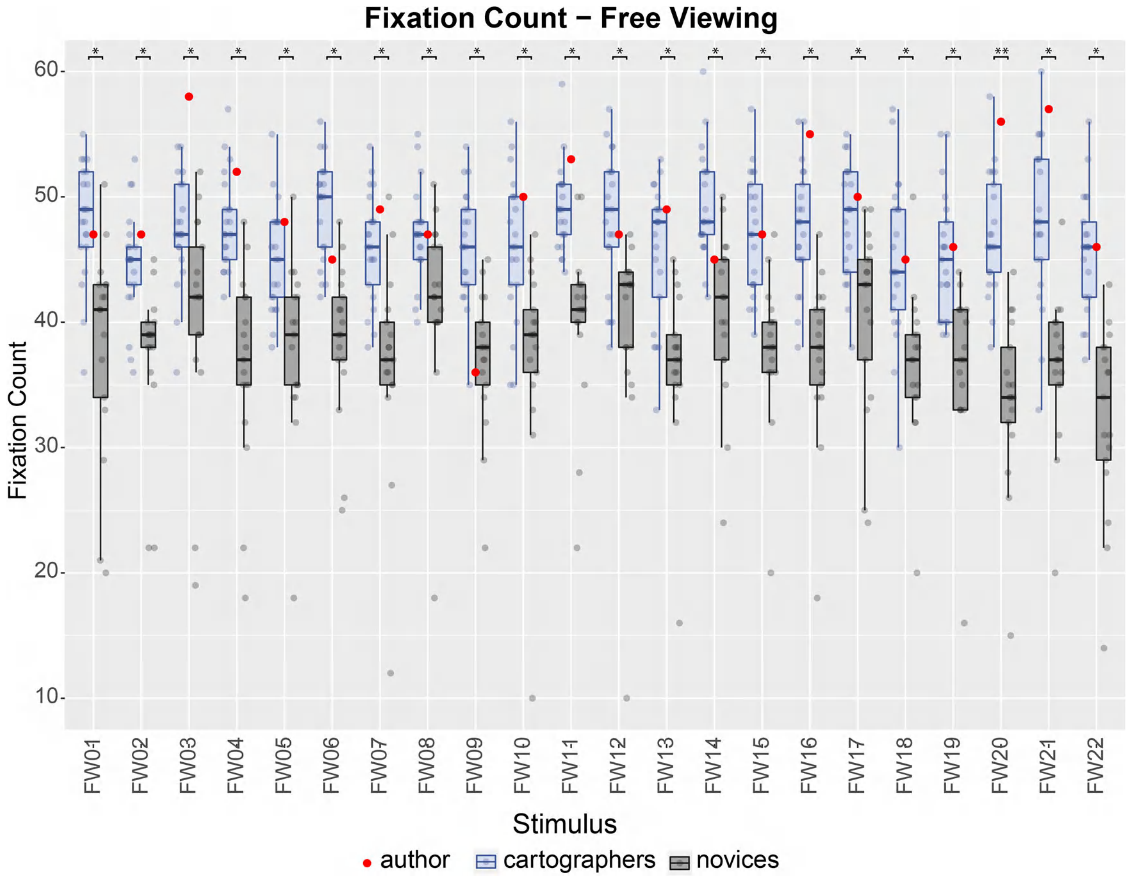


Figure 14. Fixation count for individual maps of the free viewing part of the experiment.

answer can also be found in the graph (marked red in the small inset in Figure 10), but almost no participant looked there. The sequence chart in Figure 10 shows that the author looked neither at the legend nor the graph and immediately selected (the correct) answer.

Trial duration and fixation count

Two metrics were statistically evaluated – trial duration, representing the total time needed to solve a task, and fixation count, representing the number of recorded fixations. The first analysis was focused on the groups of participants. No statistically significant differences were observed for the trial duration metric. The median value of this metric was 20–21 s for all three groups.

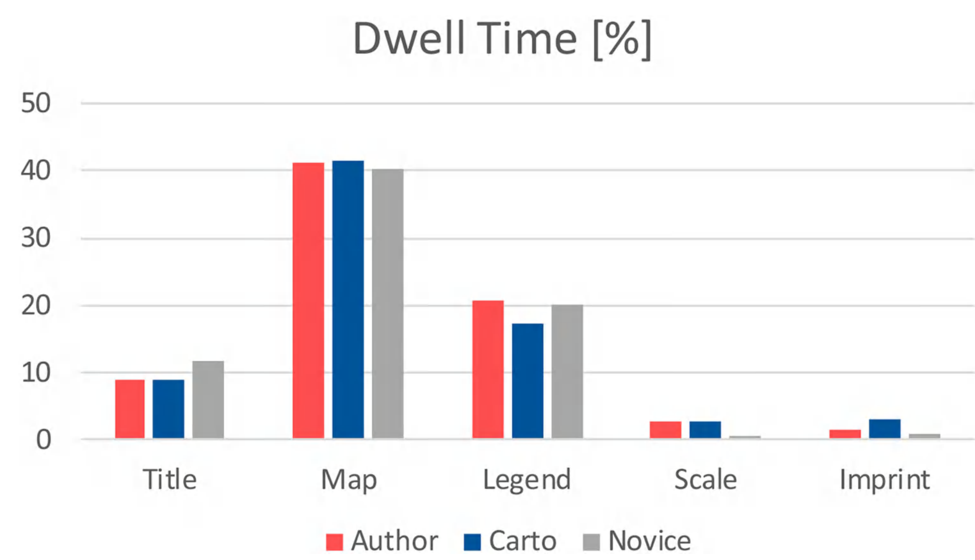


Figure 15. Dwell Time averages for authors (red), cartographers (blue) and novices (grey).

	Difference Author - Cartographers					Difference Author - Novices				
	Title	Map	Legend	Scale	Imprint	Title	Map	Legend	Scale	Imprint
FW01	-5.1	-16.7	16.4	-1.8	1.5	-5.6	-13.2	14.2	-1.0	4.0
FW02	2.4	-4.4	12.6	-0.1	-1.4	2.2	-1.8	3.2	1.5	1.8
FW03	-0.7	-3.5	10.5	0.1	-3.0	-3.9	-7.6	14.0	1.8	-0.5
FW08	-2.7	5.4	-13.2	7.6	1.6	-3.2	3.6	-13.1	9.0	3.4
FW09	7.3	14.2	-1.1	-3.0	-2.1	4.3	12.9	-3.2	-1.2	-1.5
FW10	5.4	-17.6	10.8	4.7	-2.6	6.0	-28.7	11.1	12.5	0.6
FW14	-7.7	11.0	7.0	-1.6	-0.4	-13.8	26.2	-3.0	-0.9	2.3
FW15	-4.9	11.0	-5.4	-1.0	-0.7	-3.6	3.0	-6.6	0.0	0.0
FW17	-3.2	-6.0	7.7	-2.6	-3.1	-11.9	6.0	4.1	-0.3	-1.0
FW18	9.4	3.9	-13.7	-1.9	-3.8	3.0	7.7	-17.1	-1.0	-1.2

Figure 16. The difference in Dwell Time for particular AOIs between the author and averages of cartographers (left) and novices (right).

The fixation count metric gave different results. Boxplots in Figure 11 show the fixation count values. The smallest number of fixations was observed for the group of cartographers; however, the difference between authors and cartographers was minimal. In contrast, the difference between cartographers and novices was statistically significant ($p < .001$), and the difference between authors and novices approached significance ($p = .06$).

Next, fixation count values were evaluated for separate tasks (and thus for separate maps). Data for cartographers are depicted in blue, novices are displayed in grey, and the value for the map's author is represented by a red dot (Figure 12).

Statistically significant differences between cartographers and novices were observed for nine maps (M01, M02, M03, M04, M05, M06, M09, M10 and M12). In all significant cases, the number of fixations performed by the cartographers was lower than the number of fixations of the novices. A statistically significant difference between authors and cartographers or authors and novices was not observed.

Free viewing segment

As the next step of data analysis, the free viewing part of the experiment was analysed. In this case, the accuracy of the answers was not investigated because the participants were not asked to complete any task – they were only instructed to look at the maps. Similarly, the trial duration metric was not analysed because the length of time each stimulus was presented was a constant 12 seconds.

Fixation count

As in the experiment's task segment, the fixation count for participants groups was analysed first (Figure 13).

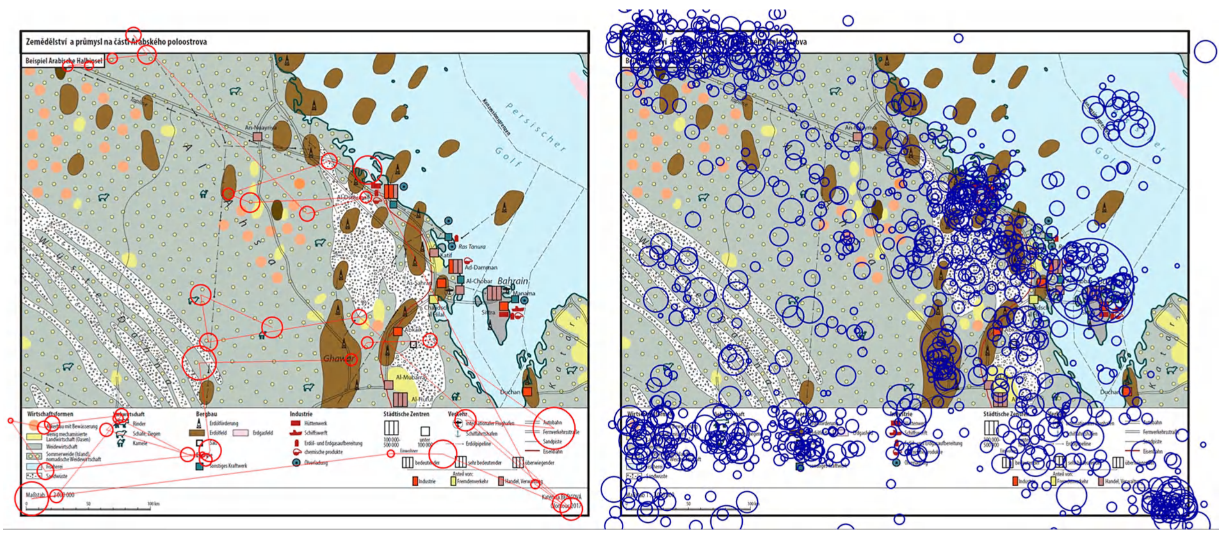


Figure 17. Scanpath for participant C01 – author of map FW01 (red) and all cartographers' fixations (blue).

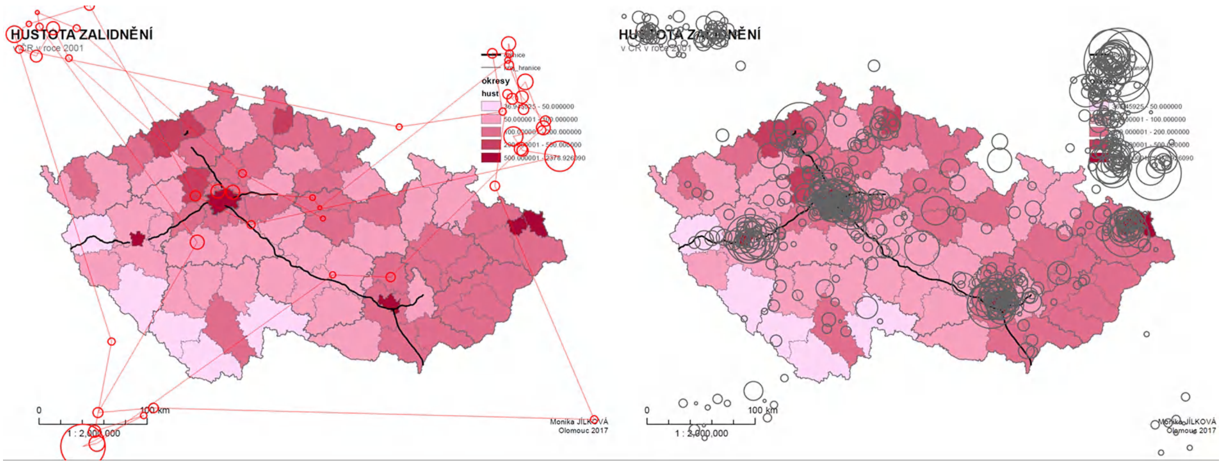


Figure 18. Scanpath for participant C10 – author of map FW10 (red) and fixations of all novices (grey).

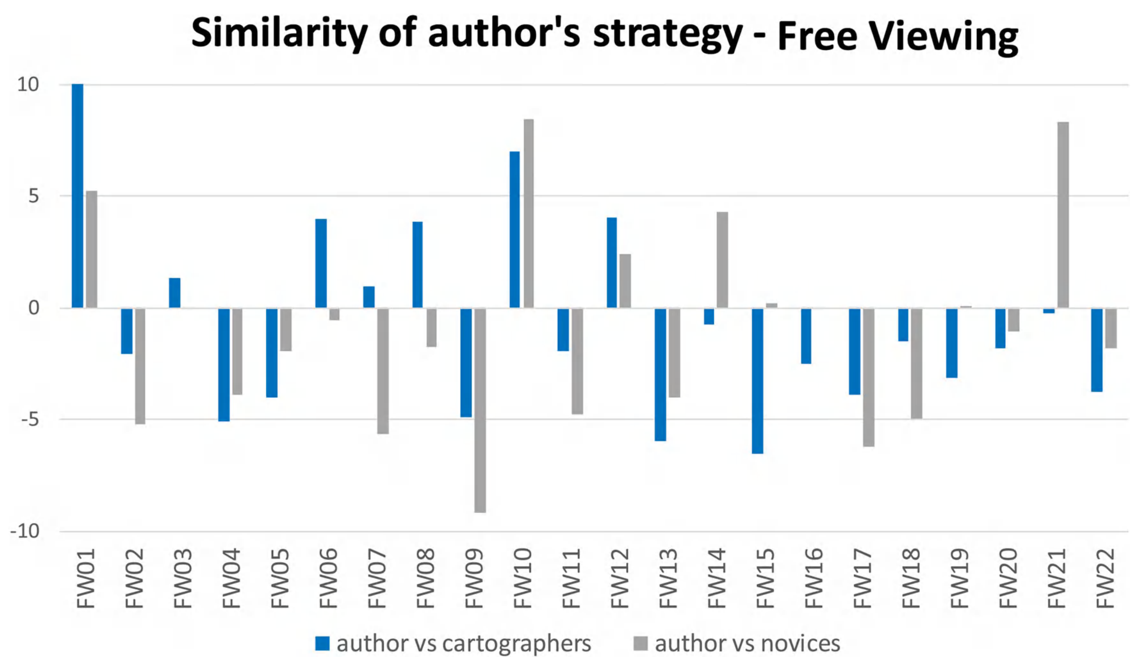


Figure 19. The similarity of the author's free viewing strategy towards the strategies of cartographers (blue) and novices (grey). Extreme values indicate a unique strategy.

In this case, the situation was the opposite of the task part of the experiment. The lowest number of fixations was observed for the group of novices. The differences between novices and both other groups were statistically significant ($p < .001$).

In the next step, the fixation count for individual maps was analysed. Statistically significant differences between cartographers and novices were observed for all maps in the experiment. In all cases, the values were higher for cartographers (Figure 14). Significant differences between other groups were not observed. The only exception is map FW20, where a statistically significant difference between the author and novices was found ($p = .04$). A total of 56 fixations were recorded for the author (C20) whereas the median of the fixation count for novices was only 34. The author focused his attention on point symbols and the corresponding part of the legend while novices inspected the title of the map. In several cases, the difference between the map author and novices approached significance (FW03 $p = .08$; FW04 $p = .089$; FW11 $p = .063$; FW16 $p = .063$; FW21 $p = .055$). In all these cases, the author's fixation count was higher than the value for other groups.

Dwell Time analysis

In the next step, the Dwell Time metric for free viewing was evaluated for all three groups of participants. These AOIs were marked around map compositional elements (title, map field, legend, scale, imprint). Ten maps containing just these five AOIs were included in the analysis.

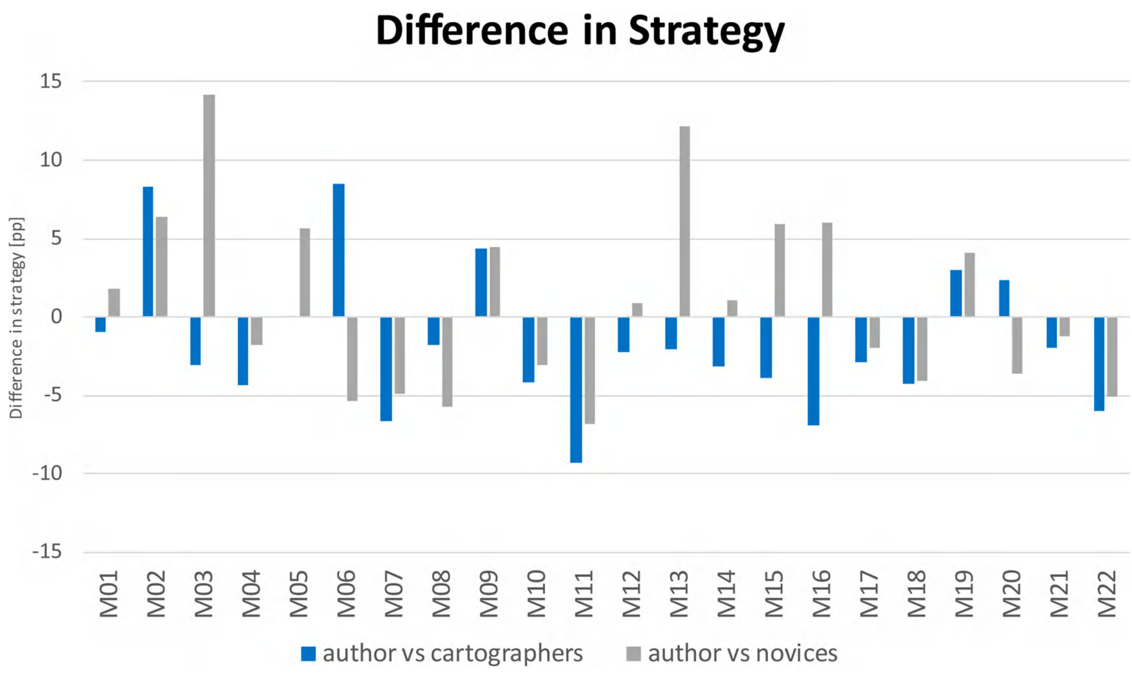


Figure 20. The similarity of each author's task completion strategy compared to cartographers (blue) and novices (grey). Extreme values indicate a unique strategy.

As is clear from Figure 15, the differences between groups were minimal. The only statistically significant difference was observed for the scale and imprint AOIs between groups of cartographers and novices. Novices fixated in these areas very rarely; they spent less than 1% of their observation time in these areas.

To investigate the variances between individual maps, the differences between authors and both other groups were calculated for each map (Figure 16).

Some interesting differences were found from the qualitative analysis for individual maps. For example, when comparing the author of map FW01 (participant C01) and cartographers, the author spent less time on the map but more time in the legend. The scanpath of participant C01 is depicted in the left part of Figure 17. It is evident that most fixations were aimed at the legend. The cartographers (displayed in blue in the right part of Figure 17) fixated intensively in the AOI around the map title. In contrast, the author of the map looked there only marginally.

The largest difference between authors and novices was observed for map FW10. The author (participant C10) almost overlooked the map field and aimed most fixations at the legend AOI. A large proportion of fixations was also recorded around the scale of the map (Figure 18).

Although the average distribution of fixations among map composition elements does not vary significantly between groups, some interesting differences were found for individual maps.

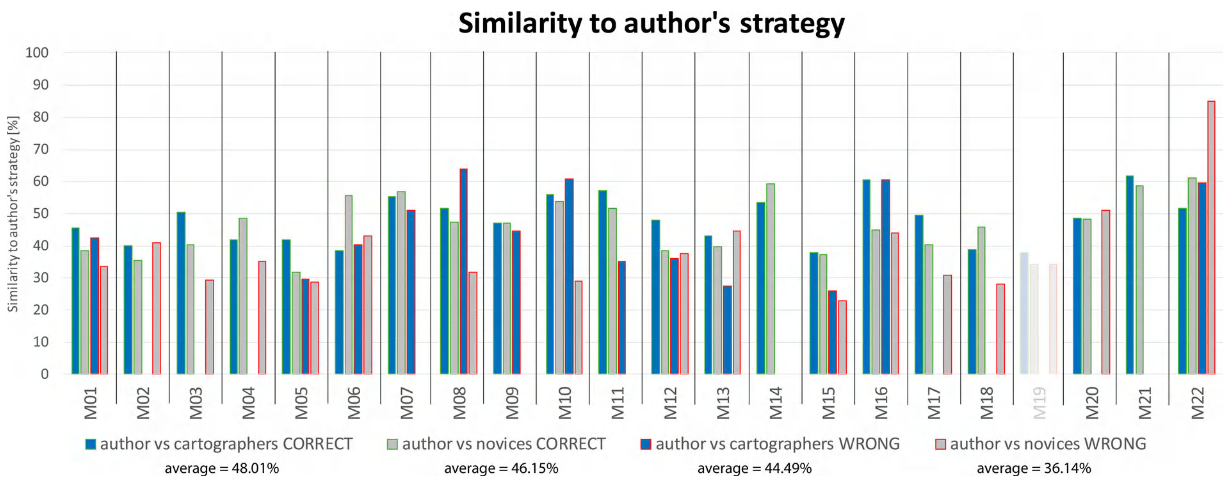


Figure 21. The similarity of participants' strategies to the strategy of the author, grouped by accuracy of answers.

Similarity of strategies

The last step of the analysis focused on the order that Areas of Interest were visited and different stimuli inspection strategies. For this evaluation, the ScanGraph tool was used.

The results for free viewing are displayed in Figure 19, for task completion in Figure 20. Extreme values represent the situations when the author used a unique strategy (different from the other participants). In contrast, values around 0 mean that the author's strategy was very similar to other participants' strategies. Positive extremes show when the readers' strategies are similar to each other, and the author used a different strategy. For negative values, the situation is more complicated. The author's strategy is more similar to the readers' strategies than the readers' strategies are among themselves, which makes the author's strategy unique also, but differently than in the case of positive values (as is shown in Figure 7). In this case, the difference between strategies cannot reach 100%.

This approach highlighted the maps where the authors used strategies that were either the most similar or the most different to those of the other groups. Regarding the unique strategies for free viewing, the largest values were observed for maps FW01 and FW10. This result is in line with Dwell Time analysis findings, where one of the largest differences was found for these maps. The reasons for these differences were described above (see Figure 17 and Figure 18). An interesting situation occurred for map FW21, where the author's strategy was unique in comparison to novices but similar to the cartographers. Map FW21 was not included in the comparison of Dwell Times because it contains more than five AOIs. In most cases, cartographers visually checked all Areas of Interest, whereas novices very rarely fixated on the legend, scale, imprint and information box AOIs. Cartographers spent 12.2% of their time in these four AOIs, whereas novices spent only 4.6%.

In the case of the task completion section, two extreme differences were observed between the author and the group of novices. For map M03, this difference was caused by the different number of fixations. The author (C03) needed 47 fixations to solve the task, whereas novices needed more than twice as many (110, see Figure 12). For map M13, the difference was that the author (C13) focused his attention on the correct part of the legend (see Figure 9). For this task, one of the biggest differences in accuracy of answers was also found.

Finally, we took into account the accuracy of the answers and the relationship between this metric and the respondents' strategy. We compared the similarity of strategies between each author and both groups of participants, which were further divided into those who answered correctly and incorrectly. The results are displayed in Figure 21. The authors' strategies were on average the most similar to the strategies of cartographers who answered correctly. The most variation in strategy was observed between the authors and the set of novices with incorrect answers. The average differences between these groups are not very distinctive, since they were influenced by the variable number of participants in the groups. Consider map M22 as an example. The task was to identify the region with the lowest share of emigrants per population. The key to finding the answer was in the legend with a point grid. In the group of novices, only one incorrect answer was recorded. Participant N13 inspected the correct part of the legend; however, he focussed his attention on the part with the highest density of points and answered by clicking into the region with the highest share of emigrants. The similarity of his stimulus inspection strategy with the author's was very high (85%), which then affects the average value of the whole group of novices with incorrect answers.

Discussion

Well-known cartographic communication models describe the relationship between the map author and the map user in general. The map serves as a communication instrument between these two individuals. Much effort has been made so far to investigate the quality of this communication. In this eye-tracking experiment, authors were placed in the role of users of their maps to prove the process of cartographic communication. The behaviour of map authors reading their own maps was compared with the map reading behaviour of a group of cartographers and a group of novices. We were inspired by two models that we consider the most balanced – the cartographic communication model introduced by Koláčný (1969) and the Venn diagram of cognitive elements which was proposed by Robinson and Petchenik (1976). Both models consider map information retrieval to be a process which is needed for completion of a given task. The cartographic communication process was verified by analysing the overlap between a map author's reality (U_1) and a map user's reality (U_2) according to Koláčný's (1969) model, or the same entities designated as A and B in the Venn diagram of cognitive elements (Robinson and Petchenik, 1976). This overlap can be quantified by accuracy of answers. The process by which the map was understood and the overlap in realities was achieved might be described by trial duration, eye-tracking metrics and analysis of Areas of Interest. One of the main contributions of this paper is a proposal of a new scanpath comparison method suitable for uncovering the strategies that participants used during map reading.

The experiment contained two segments – free viewing and task completion. For the task completion segment, the accuracy of answers was used to assess the overlap of the realities. All authors completed the tasks on their maps correctly, which indicates a significant overlap and that part M_3 in the cartographic communication model of Robinson and Petchenik (1976) was empty for them (at least when completing the given task). The cartographers and novices were similarly (in)accurate in their answers. All the tasks were focused on simple map reading (identification of the value of socioeconomic phenomena on maps containing diagrams and choropleth maps based on their legend). A higher difference between cartographers and novices might occur if the tasks were more complex.

In the next step, the fixation count metric was evaluated statistically. The results showed that novices needed more fixations to solve a task, whereas authors' and cartographers' results were comparable. Novices probably did not know where to find the correct answer, and thus they need a higher number of fixations to inspect the maps than participants with cartographic knowledge.

In the free viewing segment of the experiment, participants were asked to observe maps freely without performing any task. In contrast with the previous finding, the lowest number of fixations was recorded for novices. Participants with cartographic knowledge inspected maps more systematically. They automatically started with reading the map title and then focussed their attention on all compositional elements. In contrast, novices did not inspect all parts of the maps – they almost entirely ignore the scale and imprint elements of the maps. Regarding the imprint, this behaviour might be caused by the fact that the cartographers knew the authors personally and wanted to check who created the map.

The final step in data analysis was the evaluation of similarities in map reading strategies. These strategies were identified through the analysis of the similarities in the order of visited Areas of Interest using the self-developed online tool ScanGraph, which calculates the similarities and differences according to sequences of visited AOIs.

This approach has its limitations. Even the initial subdivision of the map into areas of interest can potentially be problematic. Defining areas of interest around the compositional elements of the map is logical, but problematic in terms of the incomparable amount of information contained in these AOIs (e.g. map field versus scale). Substituting a comprehensive map field with a single Area of Interest may lead to over-generalisation and loss of information about the details of map reading behaviour (which parts of the map were in the centre of participant's attention). The use of the proposed approach using ScanGraph is beneficial in the situation when participants inspected the stimuli for a limited amount of time (up to one minute) and for stimuli that contain a limited number (up to ten) of clearly identifiable AOIs (like map composition elements). The tool calculates the similarities in other cases as well, but they are usually too small to distinguish the strategies.

Using the same strategy as the map author used might be a prerequisite for the proper understanding of a map. However, the analysis of map reading strategies' similarity between authors and groups of participants (which were divided according to the accuracy of answers) did not lead to clear results. Perhaps this is due to the effect of inconsistent group size for individual maps, especially in situations when only one participant answered incorrectly. Nevertheless, the authors believe that this method for quantification of strategies' similarity can be used to identify problematic issues in cartography which might later be investigated thoroughly.

Another limitation of the study was the low number of participants. We were limited by the total number of students who created the maps. According to the results of G*Power (Faul *et al.*, 2007), the data for the free viewing segment of the experiment are reliable. However, the number of participants in the task completion segment of the experiment is not sufficient. Any interpretation of the results garnered from the statistical evaluation of this data should be made with restraint. Participants were recruited at two universities in two different countries. However, we did not expect any cross-cultural differences in map-reading (Lee *et al.*, 2016; Stachoň *et al.*, 2019; Lacko *et al.*, 2020) due to high proximity of the two countries (Czechia and Slovakia) both in geographical and cultural context. The analysis did not include individual differences of participants like gender, working memory capacity or brain lateralization (Lloyd and Bunch, 2008).

The results helped identify the situations when map users used similar or different map reading strategies than the authors. For now, we are unable to say if the cases when authors used unique strategies depend on particular psychological attributes of the author (such as memory) or the properties of the map. Exploring users' strategies during map interaction using eye-tracking can reveal users' cognitive schemas – relationships between cognitive processes. A deeper study of these schemata and the development of analytical methods for their detection may help to advance cartography in two directions. Developing methods for visualizing spatial data and facilitating the use of maps by users.

Conclusion

This paper describes an eye-tracking experiment in which map authors were placed in the role of users of their maps. The map reading behaviour of the authors during free viewing and task completion was compared with the behaviour of two other groups of participants – cartographers (students who had completed two cartography courses), and novices.

The purpose of this experiment was the verification of cartographic communication models using quantification of the overlap between map authors' and map users' realities. The overlap could be expressed by the accuracy of the answers. None of the authors made any mistakes in the task completion segment of the experiment, but the average accuracy of answers for the two other groups was lower. Analysis of the accuracy of answers highlighted problematic tasks, which were then qualitatively analysed. The reason for mistakes was further investigated using eye movement visualization methods like scanpath, sequence charts, and analysis of eye-tracking metrics.

Fixation count values were evaluated separately for tasks and free viewing segments. For tasks, a similar number of fixations was recorded for the group of authors and the group of cartographers. In contrast, novices needed significantly more fixations to complete the tasks. In the free viewing segment, the situation was different. The lowest number of fixations was observed for the group of novices. Dwell Time analysis revealed that novices did not focus their attention on the maps' scale and imprint elements.

Finally, we developed and successfully tested a procedure for quantification of the differences in map reading strategies based on the order of visited Areas of Interest. This approach might offer a window into the map reading processes by which the map was understood and the overlap in realities was achieved. Moreover, this analysis highlighted the cases where the users used a unique map reading strategy (different from the authors). The cases where the highest difference in map reading was detected were in line with previous experimental findings (accuracy of answers, eye-tracking metrics) and were analysed qualitatively.

The hypothesis that there would be a difference in the accuracy of answers, trial duration and eye-tracking metrics between the author and other groups of participants was confirmed. In addition, a new method for the quantification of the differences in map reading strategies was introduced which pointed to situations where the strategies between groups differed.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Palacký University Olomouc [grant number IGA_PrF_2022_027] and [grant number IGA_FF_2022_036].

Notes on the contributor

Marketa Beitlova holds a PhD in Geoinformatics and Cartography. She is using eye-tracking technology combined with other qualitative methods in cognitive cartography with a focus on school world atlases. Her research examines differences in map reading between the specific groups of participants based on scanpath comparison methods.

References

Beitlova, M., Popelka, S. and Vozenilek, V. (2020) "Differences in Thematic Map Reading by Students and Their Geography Teacher" *ISPRS International Journal of Geo-Information* 9 (9) 492 DOI: 10.3390/ijgi9090492.

Bielikova, M., Konopka, M., Simko, J., Moro, R., Tvarozek, J., Hlavac, P. and Kuric, E. (2018) "Eye-Tracking en Masse: Group User Studies, Lab Infrastructure, and Practices" *Journal of Eye Movement Research* 11 (3) 6 DOI: 10.16910/jemr.11.3.6.

Board, C. (1978) "Map Reading Tasks Appropriate in Experimental Studies in Cartographic Communication" *Cartographica* 15 (1) pp.1–12 DOI: 10.3138/AG15-V252-3726-W346.

Castner, H.W. (1964) "The Role of Pattern in the Visual Perception of Graded Dot Area Symbols in Cartography" *Dissertation Abstracts* 25 (6) p.3696.

Clarke, L.M. (1989) "An Experimental Investigation of the Communicative Efficiency of Point Symbols on Tourist Maps" *The Cartographic Journal* 26 (2) pp.105–110 DOI: 10.1179/caj.1989.26.2.105.

Crawford, P.V. (1973) "The Perception of Graduated Squares as Cartographic Symbols" *The Cartographic Journal* 10 (2) pp.85–88 DOI: 10.1179/caj.1973.10.2.85.

Dolezalova, J. and Popelka, S. (2016) "ScanGraph: A Novel Scanpath Comparison Method Using Visualisation of Graph Cliques" *Journal of Eye Movement Research* 9 (4) 5 DOI: 10.16910/jemr.9.4.5.

Faul, F., Erdfelder, E., Lang, A.-G. and Buchner, A. (2007) "G* Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences" *Behavior Research Methods* 39 (2) pp.175–191 DOI: 10.3758/bf03193146.

Griffin, A.L. (2014) "Reflections on 'Visual Perception and Map Design'" In *Landmarks in Mapping: 50 Years of the Cartographic Journal* London: Routledge, p.45 DOI: 10.4324/9781351191234-5.

Gross, J.L. and Yellen, J. (2005) *Graph Theory and Its Applications* New York: Chapman and Hall/CRC.

Hojovec, V., Daniš, M., Hájek, M. and Veverka, B. (1987) *Kartografie* Prague: Geodetický a kartografický podnik.

Keates, J. (1996) *Understanding Maps* New York: Routledge.

Kent, A.J. (2018) "Form Follows Feedback: Rethinking Cartographic Communication" *Westminster Papers in Communication and Culture* 13 (2) pp.96–112 DOI: 10.16997/wpcc.296.

Kiefer, P., Giannopoulos, I., Raubal, M., Duchowski, A., Kiefer, P., Giannopoulos, I., Raubal, M. and Duchowski, A. (2017) "Eye Tracking for Spatial Research: Cognition, Computation, Challenges" *Spatial Cognition & Computation* 17 (1-2) pp.1–19 DOI: 10.1080/13875868.2016.1254634.

King, R. (1982) "On Geography, Cartography and the Fourth Language" *Geographical Research Forum* 5 pp.42–56.

Klettner, S. (2020) "Affective Communication of Map Symbols: A Semantic Differential Analysis" *ISPRS International Journal of Geo-Information* 9 (5) 289 DOI: 10.3390/ijgi9050289.

Koehler, K., Guo, F., Zhang, S. and Eckstein, M.P. (2014) "What Do Saliency Models Predict?" *Journal of Vision* 14 (3) 14 DOI: 10.1167/14.3.14.

Koláčný, A. (1969) "Utilitární kartografie–cesta k optimální účinnosti kartografické informace" *Geodeticky a kartograficky obzor* 15 (10) pp.239–244.

Krassanakis, V., Filippakopoulou, V. and Nakos, B. (2016) "Detection of Moving Point Symbols on Cartographic Backgrounds" *Journal of Eye Movement Research* 9 (2) pp.1–16 DOI: 10.16910/jemr.9.2.2.

Kubíček, P., Šašinka, Č., Stachoň, Z., Štěrba, Z., Apeltauer, J. and Urbánek, T. (2017) "Cartographic Design and Usability of Visual Variables for Linear Features" *The Cartographic Journal* 54 (1) pp.91–102 DOI: 10.1080/00087041.2016.1168141.

Lacko, D., Šašinka, Č., Čeněk, J., Stachoň, Z. and Lu, W.-l. (2020) "Cross-Cultural Differences in Cognitive Style, Individualism/Collectivism and Map Reading Between Central European and East Asian University Students" *Studia Psychologica* 62 (1) pp.23–43 DOI: 10.31577/sp.2020.01.789.

Lee, D., Shin, H.J., Weldon, R.B. and Sohn, M.-H. (2016) "Cross-Cultural Differences in Consistent Mapping Effects of Targets and Distractors" *Journal of Cross-Cultural Psychology* 47 (4) pp.525–539 DOI: 10.1177/0022022116631823.

Levenshtein, V.I. (1966) "Binary Codes Capable of Correcting Deletions, Insertions, and Reversals" *Soviet physics doklady* 10 (8) pp.707–710.

Lloyd, R.E. and Bunch, R.L. (2008) "Explaining Map-Reading Performance Efficiency: Gender, Memory, and Geographic Information" *Cartography and Geographic Information Science* 35 (3) pp.171–202 DOI: 10.1559/152304008784864677.

MacEachren, A. (1995) *How Maps Work: Representation, Visualization and Design* New York: Guilford Press.

MacEachren, A. (2004) *How Maps Work: Representation, Visualization, and Design* New York: Guilford Press.

Morrison, J.L. (1977) "The Science of Cartography and Its Essential Processes" *Cartographica* 14 (1) pp.58–71.

Olson, J.M. (1976) "A Coordinated Approach to Map Communication Improvement" *The American Cartographer* 3 (2) pp.151–160 DOI: 10.1559/152304076784080177.

Opach, T., Golebiowska, I. and Fabrikant, S.I. (2014) "How Do People View Multi-Component Animated Maps?" *The Cartographic Journal* 51 (4) pp.330–342 DOI: 10.1179/1743277413Y.0000000049.

Opach, T., Popelka, S., Dolezalova, J. and Rød, J.K. (2018) "Star and Polyline Glyphs in a Grid Plot and on a Map Display: Which Perform Better?" *Cartography and Geographic Information Science* 45 (5) pp.400–419 DOI: 10.1080/15230406.2017.1364169.

Petchenik, B.B. (1977) "Cognition in Cartography" *Cartographica* 14 (1) pp.117–128 DOI: 10.3138/97R4-84N4-4226-0P24.

Popelka, S. (2014) "Optimal Eye Fixation Detection Settings for Cartographic Purposes" *14th SGEM GeoConference on Informatics, Geoinformatics and Remote Sensing* Albena, Bulgaria: International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM Vol. 1, pp.705–712 DOI: 10.5593/SGEM2014/B21/S8.091.

Popelka, S. and Beitlova, M. (2022) "Scanpath Comparison Using ScanGraph for Education and Learning Purposes: Summary of Previous Educational Studies Performed with the Use of ScanGraph" *2022 Symposium on Eye Tracking Research and Applications* Seattle: Association for Computing Machinery, Article 51 DOI: 10.1145/3517031.3529243.

Popelka, S., Dolezalova, J. and Beitlova, M. (2018) "New Features of Scangraph: A Tool for Revealing Participants' Strategy from Eye-Movement Data" *ACM Symposium on Eye Tracking Research & Applications: ACM* 103 DOI: 10.1145/3204493.3208334.

Popelka, S., Herman, L., Řezník, T., Pařilová, M., Jedlička, K., Bouchal, J., Kepka, M. and Charvát, K. (2019) "User Evaluation of Map-Based Visual Analytic Tools" *ISPRS International Journal of Geo-Information* 8 (8) 363.

Potash, L.M. (1977) "Design of Maps and Map-Related Research" *Human Factors: The Journal of the Human Factors and Ergonomics Society* 19 (2) pp.139–150 DOI: 10.1177/001872087701900203.

Ratajski, L. (1978) "The Main Characteristics of Cartographic Communication as a Part of Theoretical Cartography" *International Yearbook of Cartography* 18 pp.21–32.

Rieger, M.K. (1996) *A Cognitive Analysis of Map Users' Understanding of Geographic Information Systems' Images* Calgary: University of Calgary.

Robinson, A. (1952) *The Look of Maps* Madison: University of Wisconsin Press.
Robinson, A. and Petchenik, B. (1976) *The Nature of Maps: Essays Toward Understanding Maps and Mapping* Chicago: University of Chicago Press.
Shannon, C.E. and Weaver, W. (1948) "A Mathematical Theory of Communication" *The Bell System Technical Journal* 27 (3) pp.379–423 DOI: 10.1002/j.1538-7305.1948.tb01338.x.
Stachoň, Z., Šašinka, Č., Čeněk, J., Štěrba, Z., Angsuesser, S., Fabrikant, S.I., Štampach, R. and Morong, K. (2019) "Cross-Cultural Differences in Figure–Ground Perception of Cartographic Stimuli" *Cartography and Geographic Information Science* 46 (1) pp.82–94.
Stachoň, Z., Šašinka, Č., Konečný, M., Popelka, S. and Lacko, D. (2020) "An Eye-Tracking Analysis of Visual Search Task on Cartographic Stimuli" *8th International Conference on Cartography and GIS* Nessebar, Bulgaria: Bulgarian Cartographic Association 1 pp.36–41.
Voßkühler, A., Nordmeier, V., Kuchinke, L. and Jacobs, A.M. (2008) "OGAMA (Open Gaze and Mouse Analyzer): Open-Source Software Designed to Analyze Eye and Mouse Movements in Slideshow Study Designs" *Behavior Research Methods* 40 (4) pp.1150–1162 DOI: 10.3758/BRM.40.4.1150.
Voženílek, V., Morkesová, P. and Vondráková, A. (2014) "Cognitive Aspects of Map Symbology in the World School Atlases" *Procedia-Social and Behavioral Sciences* 112 pp.1121–1136 DOI: 10.1016/j.sbspro.2014.01.1277.
Wood, M. (1968) "Visual Perception and Map Design" *The Cartographic Journal* 5 (1) pp.54–64 DOI: 10.1179/caj.1968.5.1.54.
Wright, J.K. (1942) "Map Makers Are Human: Comments on the Subjective in Maps" *Geographical Review* 32 (4) pp.527–544 DOI: 10.2307/209994.

Paper Swipe-MultipleView

Popelka, S., Burian, J.*, & Beitlova, M. (2022). Swipe versus multiple view: a comprehensive analysis using eye-tracking to evaluate user interaction with web maps. Cartography and Geographic Information Science, 1-19. https://doi.org/10.1080/15230406.2021.2015721

[IF 2.5; Q2]

[5 citations on WoS; 3 citations without autocitations]

[Author’s contribution: 50%]

Abstract

The comparison of multiple maps is a common fundamental process used by geographers to explore the world. The most frequently applied interactive methods for the comparison of maps are multiple view and swipe. Swipe allows the user to interactively drag and overlap two different maps. Multiple view is based on the simultaneous side-by-side display of several maps. The current paper presents an analysis of the use of these two map comparison techniques in an Esri environment using an eye-tracking study which involved 25 participants. The participants completed two different tasks which compared land suitability using two or four maps. Based on an analysis of the recorded data, we compared the effectiveness of these methods through the accuracy of answers, the trial duration, and eye-tracking metrics of the individual compositional elements of the interactive maps. Cognitive processing was investigated through the analysis of dynamic areas of interest. This labor-intensive analysis yielded results which could be visualized using sequence charts. Based on these analyses, we concluded that the participants worked more effectively with multiple views, especially in comparing four maps. Working with swipe in the Esri environment is non-intuitive in comparisons of more than two maps. Many participants instead preferred simple toggling between layers instead of interactive swipe comparisons. However, when swipe was used to compare two maps, the method was more efficient, especially during cognitively demanding tasks.

CARTOGRAPHY AND GEOGRAPHIC INFORMATION SCIENCE
2022, VOL. 49, NO. 3, 252–270
https://doi.org/10.1080/15230406.2021.2015721

Swipe versus multiple view: a comprehensive analysis using eye-tracking to evaluate user interaction with web maps

Stanislav Popelka, Jaroslav Burian and Marketa Beitlova

Department of Geoinformatics, Palacký University in Olomouc, Czech Republic

ABSTRACT
The comparison of multiple maps is a common fundamental process used by geographers to explore the world. The most frequently applied interactive methods for the comparison of maps are multiple view and swipe. Swipe allows the user to interactively drag and overlap two different maps. Multiple view is based on the simultaneous side-by-side display of several maps. The current paper presents an analysis of the use of these two map comparison techniques in an Esri environment using an eye-tracking study which involved 25 participants. The participants completed two different tasks which compared land suitability using two or four maps. Based on an analysis of the recorded data, we compared the effectiveness of these methods through the accuracy of answers, the trial duration, and eye-tracking metrics of the individual compositional elements of the interactive maps. Cognitive processing was investigated through the analysis of dynamic areas of interest. This labor-intensive analysis yielded results which could be visualized using sequence charts. Based on these analyses, we concluded that the participants worked more effectively with multiple views, especially in comparing four maps. Working with swipe in the Esri environment is non-intuitive in comparisons of more than two maps. Many participants instead preferred simple toggling between layers instead of interactive swipe comparisons. However, when swipe was used to compare two maps, the method was more efficient, especially during cognitively demanding tasks.

ARTICLE HISTORY
Received 30 June 2021
Accepted 5 December 2021

KEYWORDS
Urban planner; eye-tracking; swipe; web map; multiple view; map comparison

Introduction

Interactivity in map applications

All types of maps (except printed maps) require some interaction with map users. Digital maps and web map applications allow many options to interact, especially for map comparison. However, web map applications can be so complex that the large selection of available tools creates confusion on the user's side.

Roth (2013, p. 64) defines cartographic interaction as "the dialog between human and map, mediated through a computing device," and further states that this interaction is "essential in the research of interactive cartography, geovisualization, and geovisual analytics." It is important to note that in connection with the increasing use of web mapping applications, the demands of their users are growing and so is the functionality of those applications. Peterson (1998, p. 3) presciently stated that "The incorporation of interaction in the display of maps may be viewed as a major accomplishment of the computer-era in cartography." Indeed, many applications in current use no longer serve only as data viewers, but offer a host of other functions, not only in terms of visualization and interactivity, but also with regards to data collection, data management, and analysis. Finally, user experience cannot be discounted; interactivity is important for its own sake, because it "increases people's enjoyment of maps and helps them retain greater amounts of information." (Cartwright, 1997; Krygier et al., 1997) in Manson et al. (2012, p. 49).

The increase in interactivity, especially in map applications, is quite significant and, in addition to increased interaction between the user and the map, also offers new possibilities for visual data analysis or map comparison. User opinions about interactivity and cartography were investigated by Roth (2015) who performed a semi-structured interview study with 21 geospatial professionals. Qualitative data analysis depicted the current trends in interactive mapping according to the view of experts.

Yi et al. (2007) and Roth (2012) summarized the different taxonomies of interaction in the field of information visualization and cartography. The most detailed taxonomy of possible types of interactivity which can be applied in computer cartography or GIS was presented by Crampton (2002), who proposed four interactivity types (upper part of Figure 1). The final category in this division

CONTACT Jaroslav Burian jaroslav.burian@upol.cz
Supplemental data for this article can be accessed here

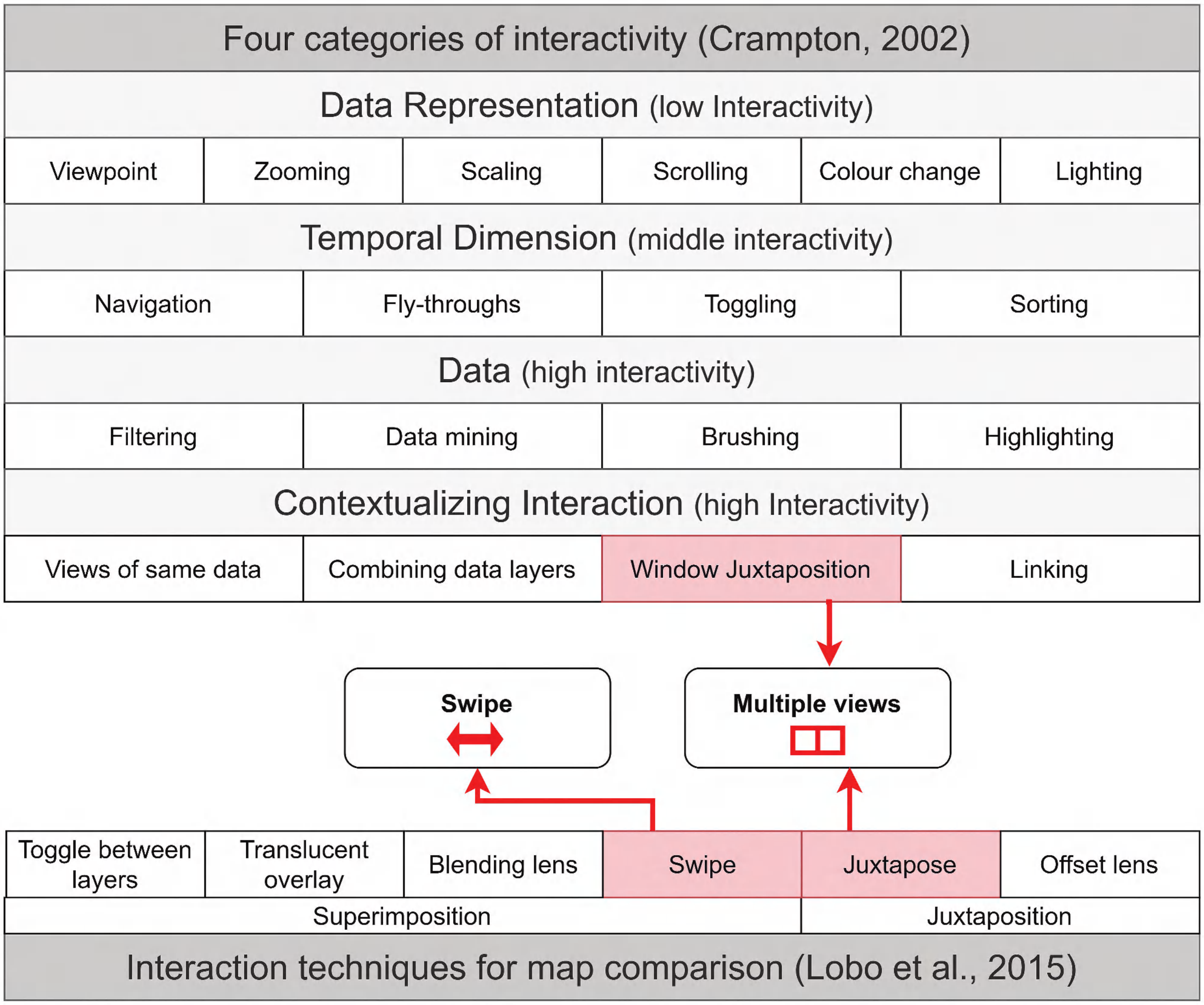


Figure 1. Four categories of interactivity proposed by Crampton (2002) and the interaction techniques for map comparison proposed by Lobo et al. (2015), including their relationships to the multiple view and swipe methods.

contains interactive methods which allow the comparison of maps. Map comparison is a fundamental method which geographers apply to understand the world. Methods of interactivity designed for map comparison were mentioned in seven out of nine of the taxonomies summarized by Roth (2012). The goal of comparison is to "enhance the likelihood that an analyst will see not only features but the relationships between features" (MacEachren, 2004, p. 401). The quantification of spatial distributions and patterns and comparison across regions or over time is central to many types of geographical research and application (Long & Robertson, 2018). A long historical tradition exists in creating visual designs which better support comparison (Gleicher, 2017).

Lobo et al. (2015) surveyed various techniques of interaction which enable map comparison and described the six representative methods applied (lower part of Figure 1). Map comparison techniques employ two key principles: juxtaposition and superimposition (Gleicher, 2017). In the case of juxtaposition, multiple non-overlapping windows depict different representations of data. By contrast, in superimposition, layers are overlaid, and different techniques are employed for their comparison.

The first **juxtaposition** method is called simply **juxtapose** and places two (or more) maps side by side. Both maps are linked, and any change in the first map's coordinates is immediately reflected in a change in the second map. According to Andrienko et al. (2003), juxtaposition appears to be the most suitable method for detecting any changes. It's advantage is in the method's capability of considering two situations simultaneously by arbitrarily shifting the focus of attention from one map to the other. The detection of changes in the overall spatial distribution pattern or characteristics of objects requires the maps to be perceived in their entirety as integral images. The disadvantage of this method is the divided (split) attention of the user (Harrison et al., 1995). Divided attention is when a user is asked to process more than one piece of information at a time and is synonymous with multi-tasking (Greenlee, 2000; Najdowski et al., 2014). The issue of divided (split) attention had been discussed in detail by Harrower (2007), Greenlee (2000), Lobo et al. (2015), and Najdowski et al. (2014) also mentioned the second method from the juxtaposition group called **offset lens**, which combines juxtaposition and the blending lens.

The simplest case of **superimposition** uses two overlaying layers and enables users to **toggle between** them. All the other more elaborate methods allow both maps to be displayed simultaneously. The superimposition method uses **translucent overlays** to display two superimposed maps and enables the user to change the opacity of the upper map. In this case, divided attention of the user is not a problem; visual interference, however, presents difficulties. The **blending lens** also uses two superimposed maps and allows the user to apply a lens to show the lower layer in a locally bounded region around the cursor. The final superimposition technique is called **swipe** and enables the user to drag one map across another map. Swipe minimizes both divided attention of the user and visual interference.

The paper presents an analysis of the difference between multiple view and swipe. Multiple view is mentioned in the study by Lobo et al. (2015) and the taxonomy of Crampton (2002), but the method is referred to as "Window Juxtaposition" or "Juxtapose." The swipe method is only mentioned in the study by Lobo et al. (2015), perhaps because when Crampton's taxonomy was compiled in 2002, the use of this method was not common.

Many private companies and the user community have a major influence on the terminology which is used in web map applications. Since the term "multiple view" is used by Esri, a leading company in Geographic Information Systems (Esri, 2021a), we therefore decided to apply the term throughout our research. Multiple view and swipe are implemented in both desktop GIS and online web map applications. The multiple view method is based on two or more (six maximum) synchronized map windows which display the same area but different layers. Each window is equipped with a zooming and panning function, legend and tool to define the main map window for synchronization. The standard Esri swipe method (horizontal or vertical) is based on an one layer which swipes (overlays) over the second layer. This simple method works well with two layers, but with more layers, the setting is also dependent on the order of the layers. In some cases, the swipe layer can be hidden (behind another layer) and the entire setting becomes confusing. A more intuitive swipe solution is implemented in World Imagery Wayback (Esri, 2021b) and allows a clear definition of two layers for comparison. Nevertheless, this approach is not simple to apply in the standard ArcGIS Online environment since it requires some programming skills. Another limitation is that only two layers may be displayed simultaneously. In the case of more layers, this method is not usable.

The present paper aims to evaluate and compare user behavior in the application of these two map interaction methods. These methods are the most commonly applied methods of comparison for two or more maps and are widely used in Esri products. The principle of both methods is illustrated in Figure 2.

We employed eye-tracking to analyze user interaction with the web maps. Eye-tracking enables a better sense of the differences between the visual behavior of different users (Opach et al., 2017). It can be considered an objective method since it indicates directly what the participants do. In contrast with simple observation or screen recording or the evaluation of user action logs, eye-movement analysis may reveal more detailed information about the participant's perception during map comparison tasks.

Related works on map comparison methods

In 2001, MacEachren and Kraak (2001) proposed a list of research challenges in geovisualization. In the section which focused on the evaluation of interfaces, the authors proposed the development of a comprehensive, user-centered design approach in geovisualization. Moreover, they highlighted the need to better understand how ordinary users interact with geospatial displays.

Roth et al. (2017) noted that interactivity in maps has transformed the passive role of map readers into the active role of map users who can create representations to best suit their needs. As Lobo et al. (2015, p. 3573)

Figure 2. Principle of multiple view (left) and swipe (right) methods of map comparison.

stated, "from a human-computer interaction perspective, one of the main challenges is to design interactive compositions that optimize the legibility of the resulting map and that ease layer comparison." In order to fulfil this challenge, it is necessary to perform usability testing of these environments, because design problems in layout and interaction can be detected through the evaluation of user interfaces (Freitas et al., 2002). Usability evaluations apply efficiency, effectiveness, and user satisfaction to measure human-computer interaction performance (Scholtz, 2006). Thorough descriptions of usability testing methods including practical advices can be found in Nielsen (1994), Rubin and Chisnell (2008), Barnum (2010), and Tullis (2007). Applications of usability testing in the field of geovisualization are outlined in Slocum et al. (2001), Robinson et al. (2005), Bartling et al. (2021), and Lobo et al. (2015) compared all five of the above-mentioned map comparison methods. Lobo et al. (2015) hypothesized that techniques which apply superimposed maps (including swipe) perform better than those which juxtaposed maps (e.g. multiple view). In this user study, 15 participants compared orthophoto images with a topographic map to determine differences (missing roads, modified roads, etc.). However, swipe performed poorly in this study. Lobo et al. (2015) explained that the problems were caused by tight coupling between motor actions and visual comparison, which forced users to adopt a very constrained scanning strategy. In the summary of results, the authors suggested reconsidering swipe use in web mapping applications.

Luz and Masoodian (2014) studied the use of translucent overlays and evaluated the readability of a background map layer superimposed with a translucent layer with sliders. The authors found significant differences in the use of different transparency levels. Bonanni et al. (2009) also investigated the use of translucent overlays and developed a tool called Wetpaint. The tool was designed to explore multi-layered images and find meaningful relationships by scraping areas. A pilot study showed that Wetpaint permits an intuitive comparison of arbitrary areas and is superior in performance than the use of a slider to change opacity. Plumlee and Ware (2006) designed an eye-tracking experiment which compared a zooming user interface and a multi-window interface. In the experiment, twenty respondents solved a multiscale pattern-matching task. Analysis of the eye-tracking data showed that the respondents performed many more gaze transitions between multiple windows than with a zooming interface.

Comparison of juxtaposition and superimposition is related to the extrinsic (visually separable variables – juxtaposition) and intrinsic (visually inseparable variables – superimposition) visualizations. Brügger et al. (2017) compared three linear symbols in bicycle maps using eye-tracking to demonstrate split attention. Two visualization methods (color hue and color-coded arrows) depicted elevation intrinsically while the third one (elevation profile) was extrinsic to the map. The authors hypothesized that the intrinsic visualization methods would outperform the extrinsic one. Their results confirmed this hypothesis. The opposite results were found by Šašinka et al. (2021), who compared bivariate intrinsic and extrinsic cartographic visualizations of soil moisture and soil depth. Processing of the intrinsic bivariate cartographic visualization method was cognitively more demanding, required more time and its use led to higher error rates. Šašinka et al. (2019) hypothesized a connection between a holistic cognitive style and extrinsic visualization, and a link between an analytic cognitive style and intrinsic visualization. Lee et al. (2021) found that holistic thinkers process divided (split) attention tasks faster than analytic thinkers.

The literature review revealed the study by Lobo et al. (2015) as the only comprehensive study of map comparison methods. Other studies addressed the topic only minimally. Eye-tracking as a method of investigation was used just once in a study by Plumlee and Ware (2006).

In the current paper, we are following up on the study by Lobo et al. (2015), while focusing on the two most frequently used map comparison approaches – multiple views and swipe. This paper performs user testing of a working map comparison solution created in an Esri environment which, critically, is actually used in practice. Eye-tracking was employed as the main investigation method, since it allows unique information about the behavior of the participants to be revealed.

Motivation and research questions

The present research was motivated primarily by the authors' previous experiences with the use of web map applications for map comparison. We observed that users have problems with the use of swipe and instead toggle layers on and off. Based on our previous research (Burian et al., 2018, 2015), we selected urban planning topics which require the comparison of several maps. Comparing several options or scenarios for one location is one of the most common tasks in spatial planning. For this purpose, we applied the results from the Urban Planner model (land suitability maps), which has been used in several regions in the Czech Republic (e.g. Olomouc Region, Ostrava Region, Vysocina Region) as support material for planning decisions conducted at urban planning offices (Burian et al., 2018, 2015). User feedback revealed that almost all users had problems comparing the different types of land suitability (e.g. suitability for housing and suitability for light industry). Based on our observations of the map-use behavior of urban planners in comparing land suitability maps, we formulated the main research question to explore the different methods which users apply during map comparison in Esri environment. The Esri environment was chosen because of its clear dominance in the environment of Czech planning offices. All regional offices currently have Esri software products and run most of their mapping applications on this system. Tasks aimed at comparing multiple maps (even in the case of four maps) are therefore almost always solved using swipe or multiple view tools.

RQ: Swipe or multiple view: which method of user interaction in Esri environment is more suitable for map comparison in web map applications?

Methods

Empirical study design

The study was designed to investigate the differences between multiple view and swipe visualizations. The study contained nine tasks, plus two training tasks to familiarize the participants with the environment and the process of recording eye movements. The study's design was within-subject; all participants worked with both visualization types (multiple view and swipe). The study has been performed in accordance with the ethical standards.

The study commenced with an introduction which acquainted the participants with the purpose of the study and the principle of recording eye movements. Participants then answered a questionnaire with demographic questions. Participants filled in their age and on three-point likert scale they stated their cartographic experience; familiarity with web map applications, multiple view, swipe, and Urban Planner. Summary of the demographic questions is a part of supplementary data, which can be found in Figure S2. The eye-tracker was then calibrated. The calibration error threshold was set to 1°; however, most participants were calibrated with a lower deviation (average of 0.4°). This procedure ensured that the recorded data was sufficiently accurate for areas of interest analysis. Recording the eye movements then began. The study was prepared in the SMI Experiment Center (version 3.7) environment, which employs screen recording stimuli.

Web maps require a considerable amount of time to load, and ad hoc loading of each web map is therefore problematic. To eliminate this problem, we pre-loaded all the web maps into Google Chrome tabs. Blank tabs were inserted between each web map tab (Figure S1). Participants read the task in the Experiment Center. Screen recording then began, and the Google Chrome window was displayed. Participants began by closing the first tab (blank), causing the first web map to be displayed (because it was prepared on the second tab). Participants ended their work with each web map by closing the tab, thus revealing the next blank tab prepared for the next task.

Participants responded with their solutions in a form created in the SMI Experiment Center environment. Participants were able to either select one of the answers or add a comment. The scheme of the study is detailed in Figure 3. Details of the tasks are shown in the lower part of the chart.

Stimuli

Calculation of Land Suitability Maps

The present study is based on the authors' previous experiences and cooperation with spatial planning offices. We therefore decided to use land suitability maps for comparison. Land suitability maps are used as supporting layers to create various urban planning documents in Czechia (e.g. analytical materials for planning or urban plans (Burian et al., 2016). A typical task with these maps is the comparison of land suitability values in several areas. To simplify comparison, the multiple view tool or swipe tools is used. Land suitability maps can be created in Urban Planner software (Burian et al., 2015, 2018), which is an analytic extension for Esri ArcGIS for Desktop designed to evaluate land suitability and detect the most suitable spatial development areas. The software uses multi-criteria analysis, respects the principles of sustainable development, and allows the creation of several land use and land suitability scenarios. The core of Urban Planner focuses on the evaluation of land suitability according to input data, its values, and weights. Land suitability is analyzed on three levels (pillars, factors, and layers) for the five predefined land use categories. Only four types of land suitability were used in the present study (housing, recreation, commercial activities, and industry).

Web map applications

We created Web Map Applications in the Esri ArcGIS Online environment. First, raster layers of land suitability (housing, recreation, commercial activities, and industry) were published as the tile layers. Vector layers

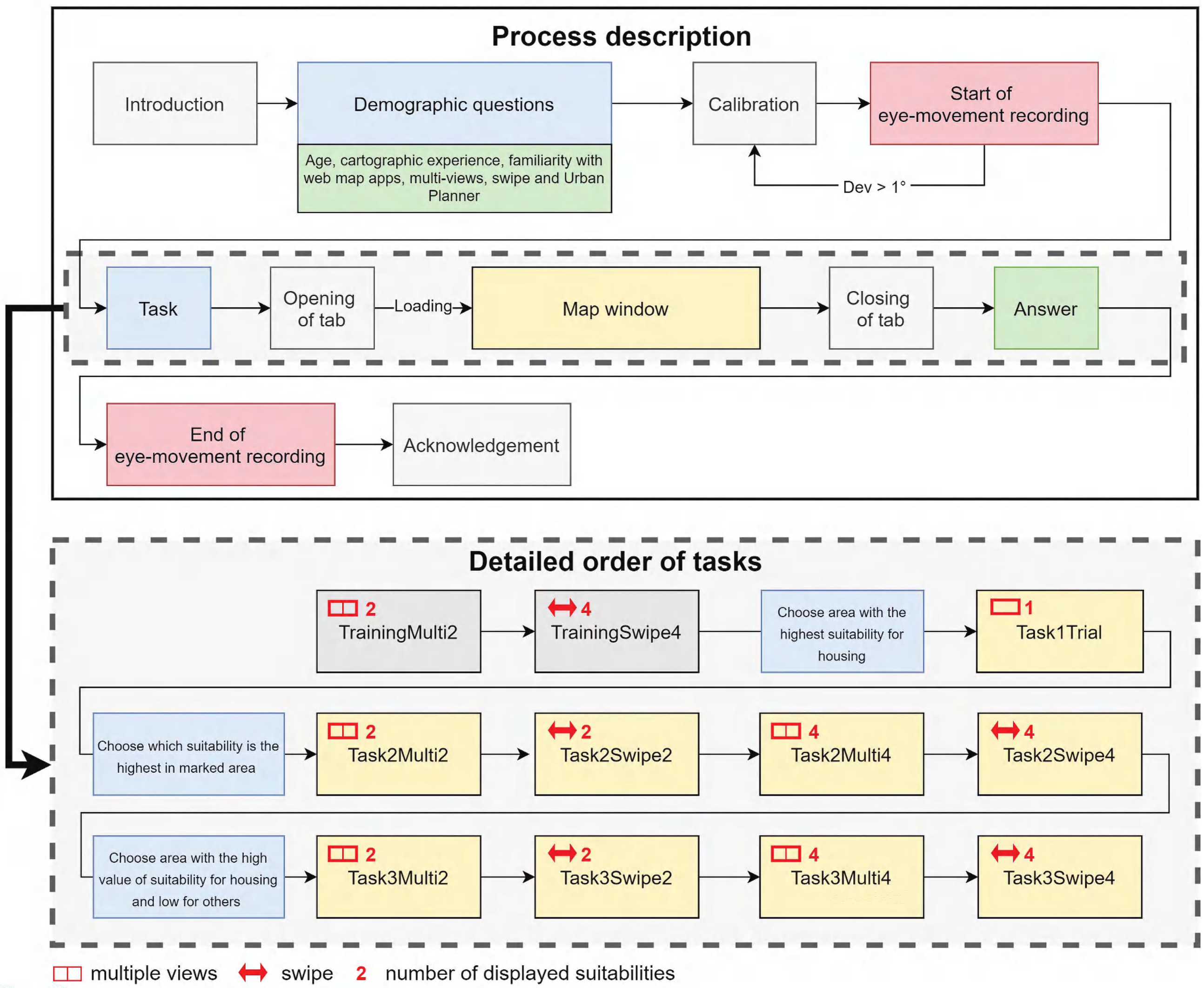


Figure 3. Scheme of the study.

(polygons of built-up areas and polygons showing the areas for comparison) were published as feature layers. We then created nine web maps and nine web map applications to cover all the tasks for the study. This principle follows the logic of ArcGIS Online, in which web map application is based on the web map and the web map contains several layers. Using Web App Builder, which allows customization of the basic template, we created web map applications for the study. Two types of application were created: 1) a single map application with a swipe tool, and 2) a multiple view application (two or four windows). We designed the applications to be as simple possible to eliminate any other tools which could distract users while they performed the tasks. For this reason, only a few basic functions were enabled: zoom, pan, switch layers on and off, and show legend. The default extent of the map was set to all polygons for comparison. Each area (small polygons) which displayed a different location for comparison was numbered 1–4 in each task.

Tasks

The tasks were presented in a fixed order. Participants began with the two training tasks. The first training task contained multiple view with two windows, while the second contained swipe with four maps. The environment was introduced to the participants and they were permitted to explore its use for two minutes. They then commenced solving the tasks. In each task, the respondents were asked to determine the location (polygon) according to the question. Answers were represented by the polygon's number. The first task was to select one of four areas which had the highest

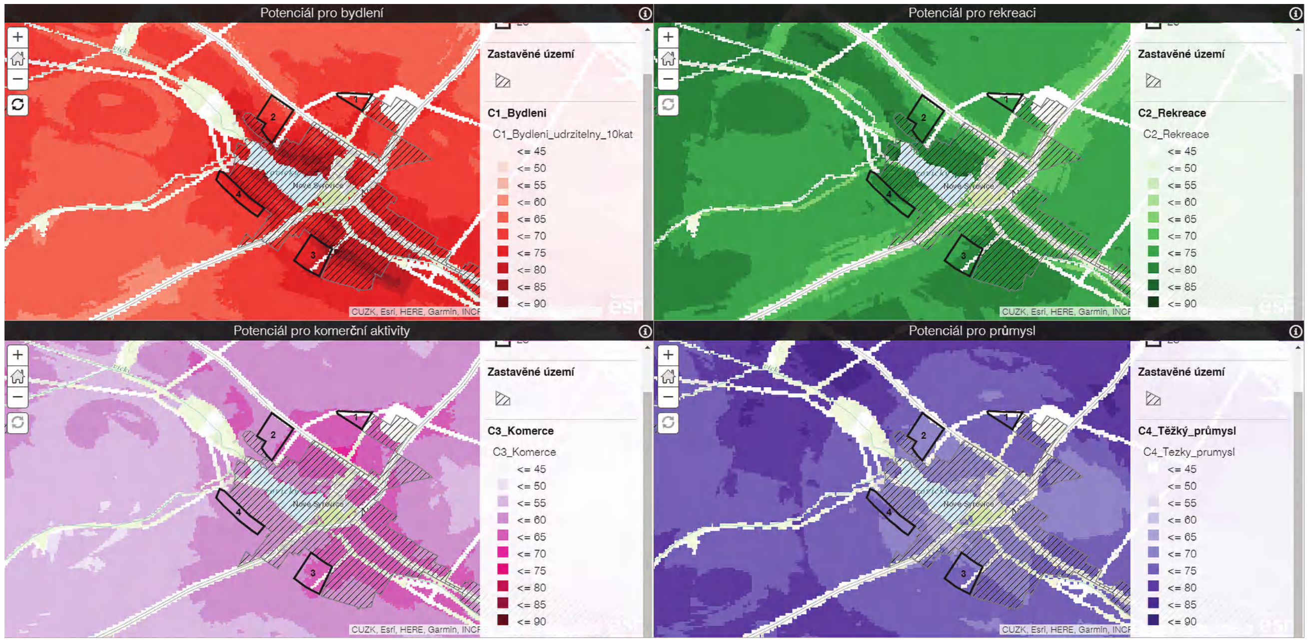


Figure 4. Task3Multi4 with multiple view tool.

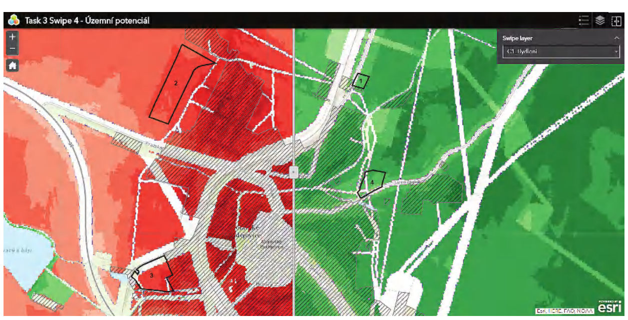


Figure 5. Task3Multi4 with multiple view tool.

suitability for housing. This task was different from the others and contained only one window. This task was not analyzed in any detail and only used to familiarize the participants with the environment and the process of the study. Two sets of tasks were then displayed. For the first four views, the task was to select the highest suitability in the marked area. Two tasks (Task2Multi2 and Task2Multi4) contained multiple view interaction, while the other two (Task2Swipe2 and Task2Swipe4) contained swipe interaction. Another variable which changed during the study was the suitability number (and, therefore, the number of layers). The first two tasks contained two layers of suitability, and the latter two contained four layers of suitability (Figures 4 and 5). In the second group of tasks, the participants selected the area with a high value of suitability for housing and a low value for all the other types of suitability. The order of tasks was the same as in the previous set. The type of visualization and the number of displayed layers of suitability are indicated with red symbols in Figure 3. All applications are available at the links listed in Table 1.

Apparatus

Eye-movement data were recorded using remote eye-tracker SMI RED 250 at a sampling frequency of 250 Hz. The eye-tracker was located in the eye-tracking laboratory, which is a specially equipped room with covered windows where participants can focus on task solving without disturbance. The study was created in SMI Experiment Center 3.7 and applied the screen-recorded stimuli as described above. Web maps were presented in the Google Chrome environment and presented on a 24″ IPS monitor.

Participants

A total of 27 participants participated in the study; however, two were excluded because of calibration problems. Data from 25 participants were therefore analyzed (16 males and nine females). The power of the tests was verified using G*Power software (Faul et al.,

Table 1. Links to web map applications used in the study.

	2	↔ 2	4	↔ 4
TASK 2	Task2Multi2	Task2Swipe2	Task2Multi4	Task2Swipe4
TASK 3	Task3Multi2	Task3Swipe2	Task3Multi4	Task3Swipe4

2007). The average age of the participants was 28.3 years. Since the Urban Planner application is not designed for use by members of the general public, we selected participants who had some cartographic background for the study. The participants were divided into two groups: students and experts. The first group consisted of 15 master's and doctorate students of geoinformatics, and the latter of 10 employees from the urban planning departments of Olomouc regional authorities. The group of experts was included because user testing with real users is the most fundamental usability method (Nielsen, 1994). As Liao et al. (2021) stated, multiple studies have reported significant group differences in users' eye movements during map reading. The overview of participants' characteristics is displayed in Figure S2.

At the beginning of the study, participants answered several questions concerning their familiarity with web map applications. Twenty participants answered that they worked with web map applications almost every day. The remaining five participants estimated that they used web map apps once a week. Fifteen participants considered their cartography knowledge "average-level," and the remaining ten participants answered "high-level." The next two questions concerned familiarity with multiple view and swipe applications. The answers to these questions were very similar. Only two participants had never previously encountered multiple view applications. Four participants did not have any experience with swipe. Around three-quarters of participants answered that they had used these features several times. The final question concerned familiarity with Urban Planner. Nine participants had no experience with this tool.

Data pre-processing

The recorded eye-movement data required pre-processing before it could be analyzed. The first step filtered participants with insufficient calibration quality or a high amount of data loss. For these reasons, the data of two participants were excluded from the analysis. The excluded participants (P23; P24) were from the group of experts and older persons who wore glasses, which may have been a factor in affecting the data quality. The average tracking ratio (the number of non-zero gaze positions divided by the sampling frequency multiplied by the run duration, expressed in percent) was 95.4%. The average calibration deviation was 0.4°.

The next step in data pre-processing was setting the fixation detection algorithm. During a fixation, the eyes look at one spot in the visual scene relatively steadily. The duration of a fixation is from a few tens of milliseconds to seconds (Holmqvist et al., 2011). Visual perception consists of a series of fixations focused on individual elements in a scene. Many fixation detection algorithms are available (Salvucci & Goldberg, 2000), but the two most commonly used are I-DT and I–VT. For the detection of fixations in the present study, we used the I-DT (identification by dispersion threshold) algorithm. I-DT takes into account the close spatial proximity of the eye's points of position in the eye-movement trace. The algorithm applies two thresholds: minimum fixation duration and maximum dispersion. These thresholds were set to 80 ms and 50 px, respectively, according to the recommendation by Popelka (2014).

Screen-recording stimuli were applied in the study, and the output was a screen-recording video overlaid with the eye-movement trajectory. In the analysis of this type of data, we applied two methods. The first method works with custom trials, and the second works with dynamic areas of interest, which are marked in the screen-recording stimuli. Both methods require a large amount of manual work and are time-consuming.

Custom trials featured in SMI BeGaze 3.7 allows the segmentation of video recordings into trials which represent individual tasks. Instead of individual videos for each participant, the data appears as if all users had looked at a single image (screenshot of the task). This method facilitated analysis. However, the method can only be used for the analysis of eye-tracking metrics for the entire task, not for a detailed analysis of user behavior within each task (analysis of areas of interest). Custom trial data were used to analyze trial duration and other eye-tracking metrics. The value of trial duration from custom trials was later used to calculate the ratio of swipe method use.

The second method of data analysis involved the creation of dynamic areas of interest. In contrast to static (figure) stimuli, screen-recording data requires areas of interest to be marked not only in space but also in time. Using the AOI editor in SMI BeGaze 3.7, we created dynamic areas of interest which covered individual maps, legends, and layer tabs. Key frames in the software had to be inserted at points when the AOI appeared (i.e. the participant opened the legend) or disappeared (i.e. the participant closed the legend). It is also possible to change the AOIs dynamically in time (i.e. the participant used the swipe function). As Dong et al. (2020) stated, manual methods like drawing dynamic areas of interest or mapping of fixations to a static reference image are effective but labor-intensive. In the present study, the procedure took approximately ten times longer than the length of time of the video recording. However, dynamic Areas of

Interest allows a detailed analysis of participant behavior. It is possible to calculate the length of time each AOI was displayed, how much time participants spent looking at them, and so on.

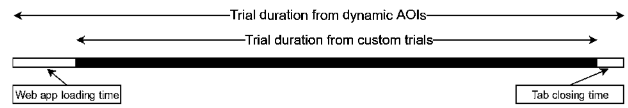


Figure 6. Task3Swipe4 with swipe tool.

Methods of analysis

During the data analysis we focused on effectiveness (accuracy of answers) and efficiency (response time) for the tested visualization methods. The last aspect of usability, user satisfaction, was not analyzed directly, but many insights were found through qualitative analysis and, additionally, based on the user comments during the study.

First, we evaluated the accuracy of the participants' answers. Then, we analyzed the time each participant required to solve a task by applying the trial duration metric. We then examined the participants' behaviors in detail. We analyzed the proportion of time the participants spent using the swipe function and performed a detailed investigation of the participants' eye movements using sequence chart visualization.

Trial duration, which illustrates the time required to solve a task, was investigated to determine which tasks and visualization methods the participants found challenging. As described above, the trial duration values were obtained from the custom trials export. Data were visualized using boxplots, and the differences between multiple view and swipe were statistically tested using the Wilcoxon signed-rank test. In the next step of data analysis, we applied the Wilcoxon rank-sum test to compare the participants who answered correctly. Both tests were executed in RStudio with a 0.05 significance level.

The final step of data analysis was a detailed description of user behavior using dynamic AOIs. For each task, we visualized visible and dwell times for each Area of Interest in graphs. Visible time refers to the time when the AOI was active. Dwell time refers to the time the participant spent gazing at the AOI.

We created additional AOIs representing the use of the swipe function to measure the duration of swipe use. The value of trial duration from custom trials export was divided by the swipe duration to obtain the proportion of swipe use. The export from the dynamic AOIs contained the trial duration value, but this value was affected by the application's loading time. The difference between these two metrics is depicted in Figure 6.

Finally, we visualized the AOI data in sequence charts, which offered illustrative overviews of the participants' work with the stimuli. Unfortunately, the sequence charts had to be manually created in the graphical software. SMI BeGaze allows the export of raster sequence charts only for individual participants. Sequence charts are very complex and present various types of information, such as which layers were displayed, where participants looked, how they used the legend, and so on. We therefore created a graphical legend for each sequence chart.

Results

Correctness of answers

An overview of the correct and incorrect answers given by the participants is shown in Figure 7. In almost all cases, the answers were correct, and only 1–2 participants answered incorrectly. We can observe a significant exception in Task3Swipe4, which was the most difficult task. The respondents were required to compare four land suitability maps with the swipe tool. The differences in suitability in all the polygons were small, therefore comparison of the maps and subsequent selection of the correct answer was challenging. Both the students and experts responded with fewer correct answers than in the previous cases. Two participants were unsure which answer was correct and responded with "I do not know."

Trial duration

The complexity and difficulty of individual tasks may be illustrated using the trial duration metric, which represents the time needed to solve each task. The

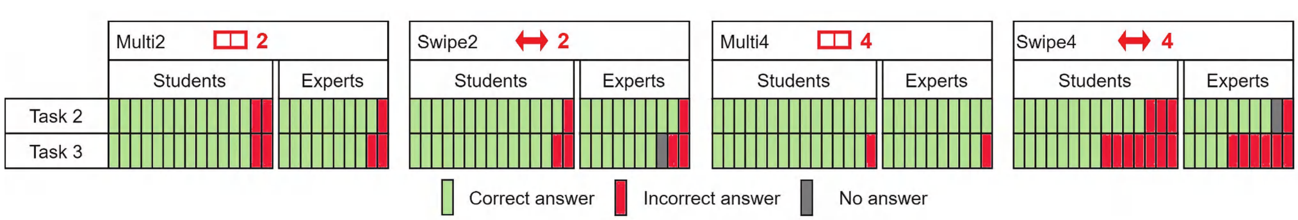


Figure 7. Overview of correct and incorrect answers.

results are charted in Figure 8. Serving as a training task, Task1Trial was one of the quickest tasks solved. This is logical since the web map app, in this case, contained only one window. The comparison between multiple view and swipe, however, is more relevant. Task 2 (selecting the highest suitability) revealed that participants spent more time with the swipe version of the application. The results were statistically significant ($V = 6$, $p < 0.001$). Task2Swipe4, which required participants to select four layers for swipe comparison, took the participants almost three times longer to complete than the multiple view version of the task (mean 76.4 seconds (Std Dev 40.9) versus 28.4 seconds (Std Dev 13) for Task2Multi4). A comparison of this with the results for the accuracy of answers is essential. In the swipe version of the app, participants made five errors. In contrast, using the multiple view version of the app, all participants answered correctly.

Task 3, which required determining an area with a large value of one type of suitability and areas with small values of other types of suitability, produced different results. In the tasks which compared just two types of suitability, the participants were quicker with the swipe version of the web map app. These results suggest that swipe may have been an advantage in this type of task. However, in the tasks involving all four suitability maps, participants struggled to select the layers for swipe comparison, and solving the task with the swipe version required significantly more time. The difference in the correctness of answers was even more significant than in the previous example. With the multiple view version, only two participants were incorrect. In the case of the swipe version, almost one-half (12) of participants made an error.

We calculated the values for trial duration for each task, including the tasks where participants provided incorrect answers. In the variants which required changing the layers for swipe comparison, some participants did not complete the task because they did not know how to select the proper combination of layers. This behavior may have affected the trial duration values. Figure 9 shows the trial duration values only for those tasks where participants answered correctly. The results were very similar to the results of all other tasks. For correct answers only, the difference between Task3Multi2 and Task3Swipe2, where swiping was faster than multiple view, was statistically significant ($V = 6$; $p = 0.037$).

A more thorough qualitative analysis of the participants' behavior is described in the next section.

AOI based analysis

For the AOI analysis, we calculated the average visible times of all AOIs. The results are visualized in the graph in Figure 10. In the multiple view tasks (App2, App4), the ratio of visible times of all map related AOIs was constant because the maps were visible for the entire duration of the task. In contrast, the visible times of legend related AOIs may vary because participants switched legends separately for each suitability map. However, the average visible times of the legend related AOIs were very similar. As the sequence charts will show, many participants switched on the legends almost simultaneously and kept them visible until the end of the task.

For the tasks which involved swipe, the legend could be displayed in two different ways: the first was via the "legend" button, the second via "layers'." Each layer could be expanded in the layer tab to display the legend (indicated as "Layers legend" in Figure 10). In Swipe2, the layers were expanded by default, whereas in Swipe4, participants had to expand them manually. Figure 10 shows that the visible time of "Layers" AOI is much longer than the visible time of "Layers legend" AOI, revealing that many participants did not or could not expand the legend in the layer tab in Task2Swipe4 and Task3Swipe4.

The chart for dwell time (Figure 11) confirms the trend we discovered during the trial duration analysis. Dwell times in swipe tasks were significantly longer than in multiple view tasks. The only exception was the Task3Multi2 and Task3Swipe2 pair. The average time spent in legend related AOIs was almost the same in both variants. Dwell time for map related AOIs was even shorter in the swipe version of the task. This result might suggest that the use of swipe was beneficial in Task3 and two maps only. However, the participants found that the use of swipe with four maps was challenging, and they focused their attention on the Layers (or Layers legend) AOI instead. Task2Swipe4 produced an extreme result: the participants spent half of the time (30 seconds) in each of the layer and legend AOIs. The results of Task3Swipe4 were unclear since many of the participants abandoned the use of swipe with four maps.

Some participants had problems with the selection of the correct layers during swipe comparison. Many switched the layers on and off instead of using the swipe functionality. Figure 12 shows the percentage of time the participants spent using the swipe function.

The results indicate that swipe was most used to compare two layers in Task 3 (Task3Swipe2). On average, this function was used for 49.7% of the time needed to solve a task. Nevertheless, this result was the highest

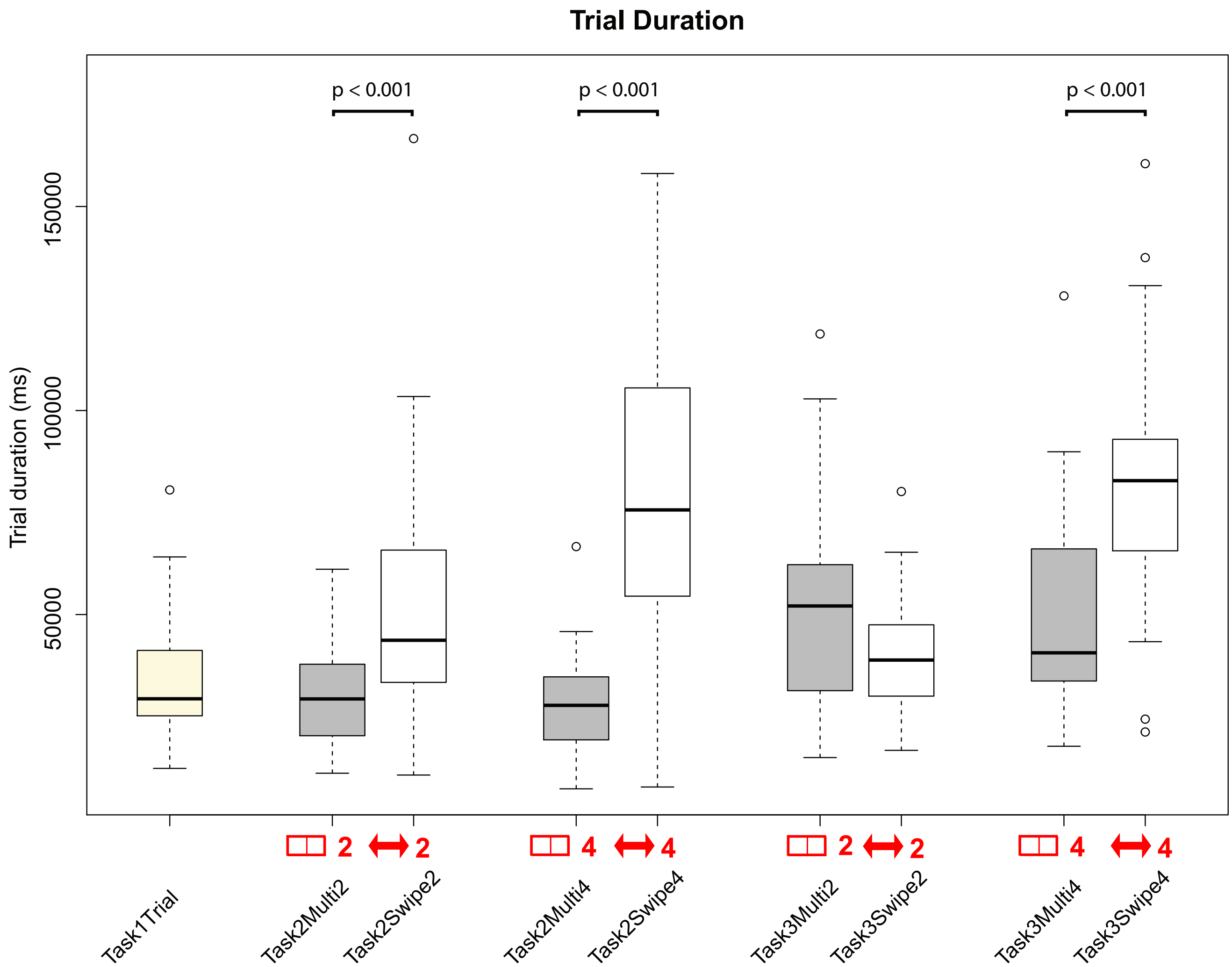


Figure 8. Trial duration values for each task in the study.

figure obtained in the entire study. We expected that this function would be used much more extensively. Figure 14 also shows that some participants never used the swipe function.

Dynamic AOIs can be displayed effectively as sequence charts to indicate the behavior of individual participants.

Task2Multi2 – Multiple view comparison of two maps

The sequence chart for Task2Multi2 (Figure 13) shows that the majority of participants switched on both legends after approximately 10 seconds of inspecting the stimuli and kept them switched on until the end of the task. The exceptions were participants P10, P15, and P26, who switched the layers on and off during the task. Four participants (P07, P08, P20, and P27) did not switch on the legends. Surprisingly, some incorrect answers were recorded: participant P20 answered incorrectly; P13 only switched on the legend for commercial suitability, nevertheless answered incorrectly.

Task2Swipe2 – Swipe comparison of two maps

The accuracy of answers was slightly better in the swipe variant of the task (Task2Swipe2); only two incorrect answers were recorded. More than half of the participants (13) did not use the swipe function during this task. These participants switched layers on and off instead and visually compared them separately. An illustrative example of this behavior is the sequence chart for P21 or P22 (Figure S4). Some participants attempted to combine both methods. For example, P15 switched layers on and off, and after 50 seconds, also attempted to swipe.

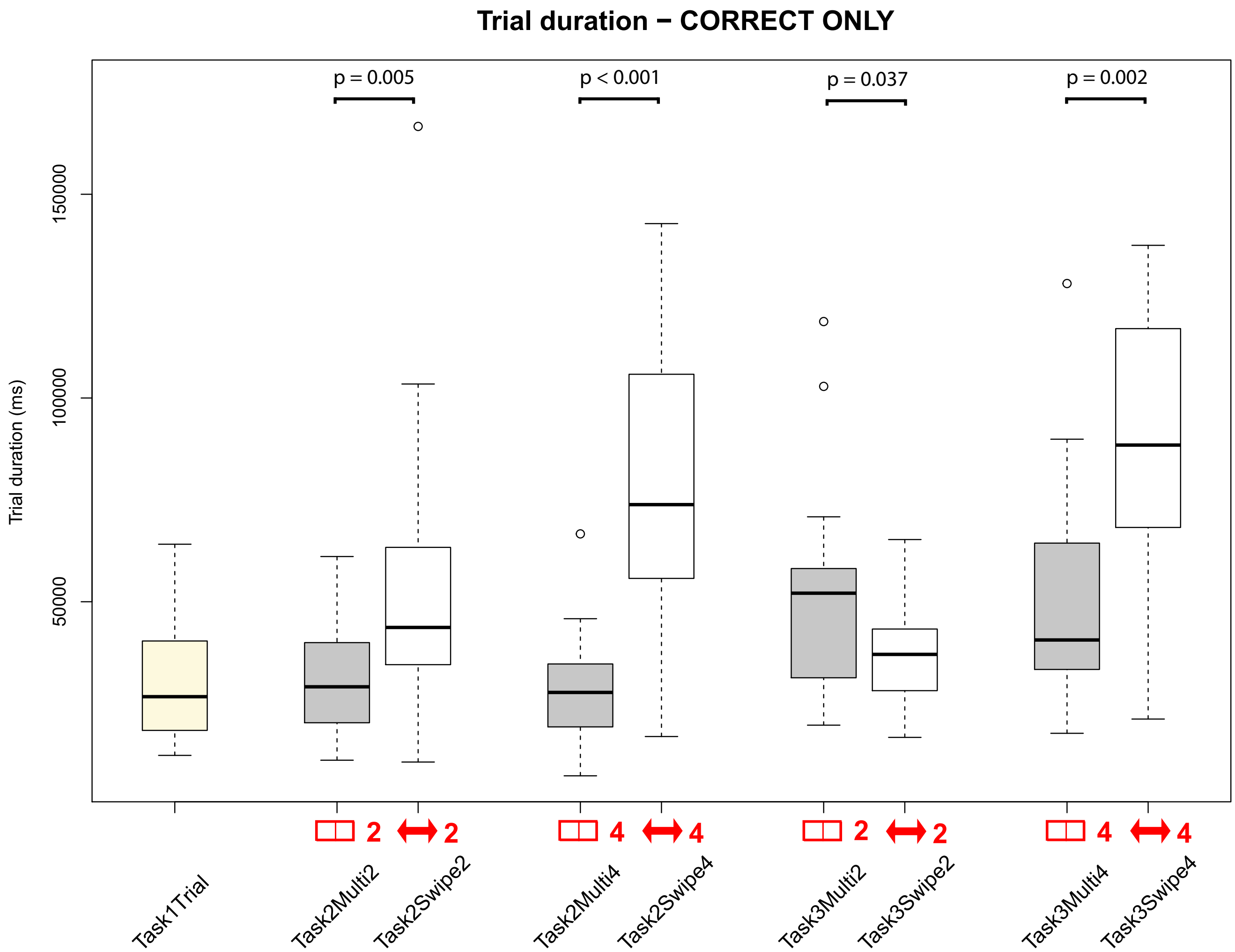


Figure 9. Trial duration values for correct answers.

Task2Multi4 – Multiple view comparison of four maps

In Task2Multi4, participants were required to select the highest suitability for a marked location from four options. This task proved to be the simplest of the entire study: all the participants' answers were correct, and the participants spent the least amount of time to complete the task. Interestingly, five participants did not switch on the legends (Figure S5). Four of these participants were the same as those who did not display the legends in the previous task (Task2Multi2). Three other participants (P14, P15, P19) switched on only the legends for suitability maps they had not yet seen during the study.

Task2Swipe4 – Swipe comparison of four maps

The task with swipe and four maps proved challenging to the participants. Swipe was used the least in Task2Swipe4 (less than 10% of the observation time). As in the previous swipe task, thirteen participants did not use this function (Figure S6). However, it is important to mention that these were not the same participants as those indicated by the results of the previous task. Only three participants switched on swipe with all four suitability maps (P01, P02, P20).

Task3Multi2 – Multiple view comparison of two maps

Task3 appeared even more challenging than the previous task. Participants were required to select one of four locations where the value of suitability for housing was high and all other types of suitability were low. The participants therefore had to compare the suitability values for four locations very thoroughly. Figure S7 indicates that the participants most likely relied on memory since 10 of them did not switch on the legends while they solved the task. We can also observe that even the participants who displayed the legend did not look at them for any significant amount of time. The average dwell time for each legend's AOI was only slightly more than 3 seconds.

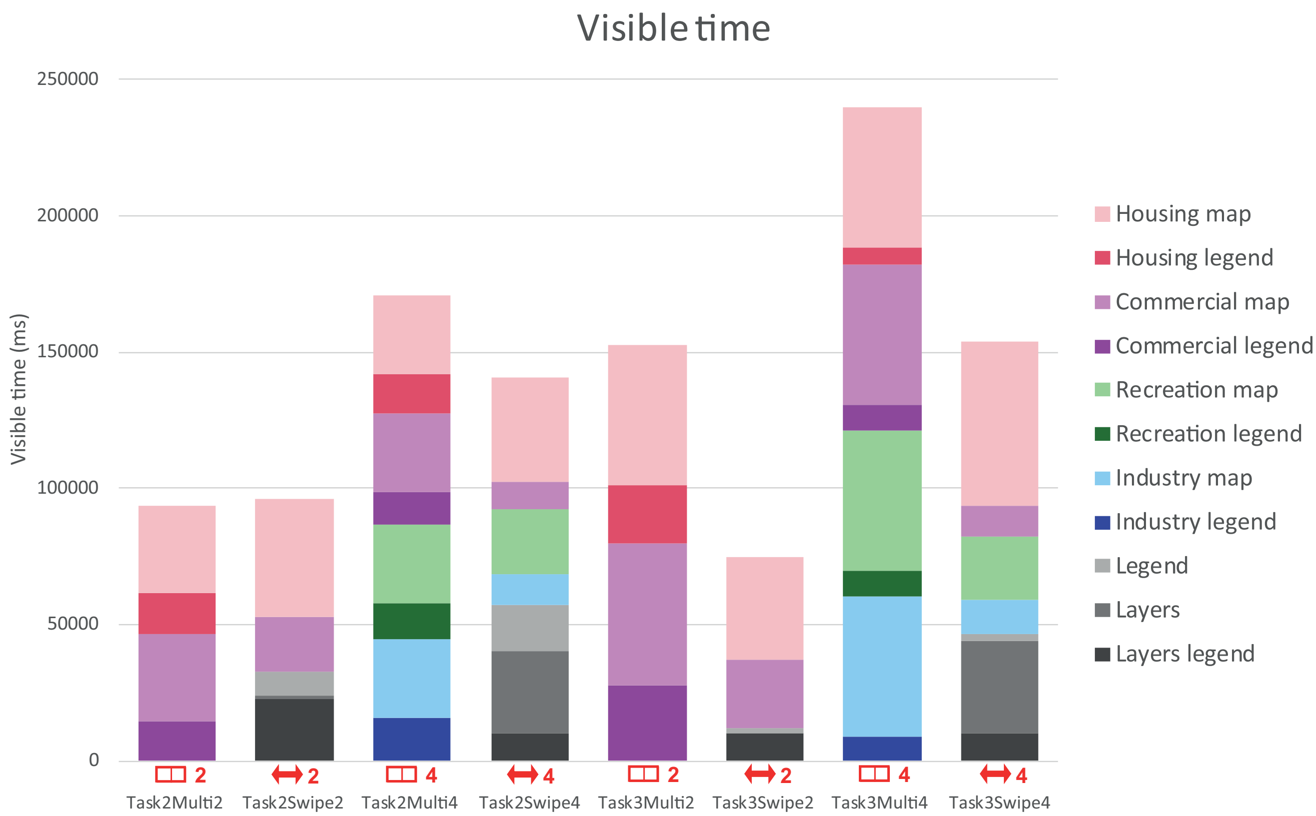


Figure 10. Visible times for dynamic AOIs marked in the stimuli. The figure shows for how long each AOI was displayed for each task.

Task3Swipe2 – Swipe comparison of two maps

Task3Swipe2 was the only task where the use of swipe was beneficial. As indicated in Figure S8, participants who responded with correct answers solved this task much more quickly using swipe than using multiple view ($p = 0.037$). Five participants answered incorrectly. Six participants did not use the swipe function in this task (Figure S8); five of these participants had not yet used swipe in any task. Furthermore, as indicated in Figure 12, participants used swipe for almost 50% of the time.

Task3Multi4 – Multiple view comparison of four maps

In the final pair of tasks, participants compared values of four types of suitability. The sequence chart in Figure 9 reveals that most of the participants did not display the legends. Only three participants switched on all four legends (P14, P21, P26). At least one legend was displayed by only seven participants. Nevertheless, the accuracy of answers in this task was one of the highest, with only two incorrect answers. This might be caused by the fact that the stimuli were presented in the fixed order and the legends were still the same and participants might remember them.

Task3Swipe4 – Swipe comparison of four maps

In Task3Swipe4, the correctness of answers was the lowest in the entire study, and the task required the most time to be solved. As in Task2Swipe2 and Task2Swipe4, again, 13 participants did not use the swipe function. Many of the participants complained about the manner in which the layers for swipe were selected. Compared to Task2Swipe4, more participants used swipe with all four types of suitability. Six participants (P01, P02, P03, P14, P18 and P20) apparently understood the principle of layer selection (Figure S10).

Discussion

Although interactive map comparison is a fundamental task for geographers, an analysis of the effectiveness, efficiency, and user satisfaction of various interactive methods has not previously been attempted. The only exception is a study by Lobo et al. (2015), who compared five interactive methods, including multiple view and swipe, which we examined in the present paper. Lobo et al. (2015, p. 3581) discovered that swipe performed poorly with map comparison. The authors tentatively attributed this to “tight coupling between motor actions and visual comparison, forcing participants to adopt a very constrained scanning strategy if they wanted to

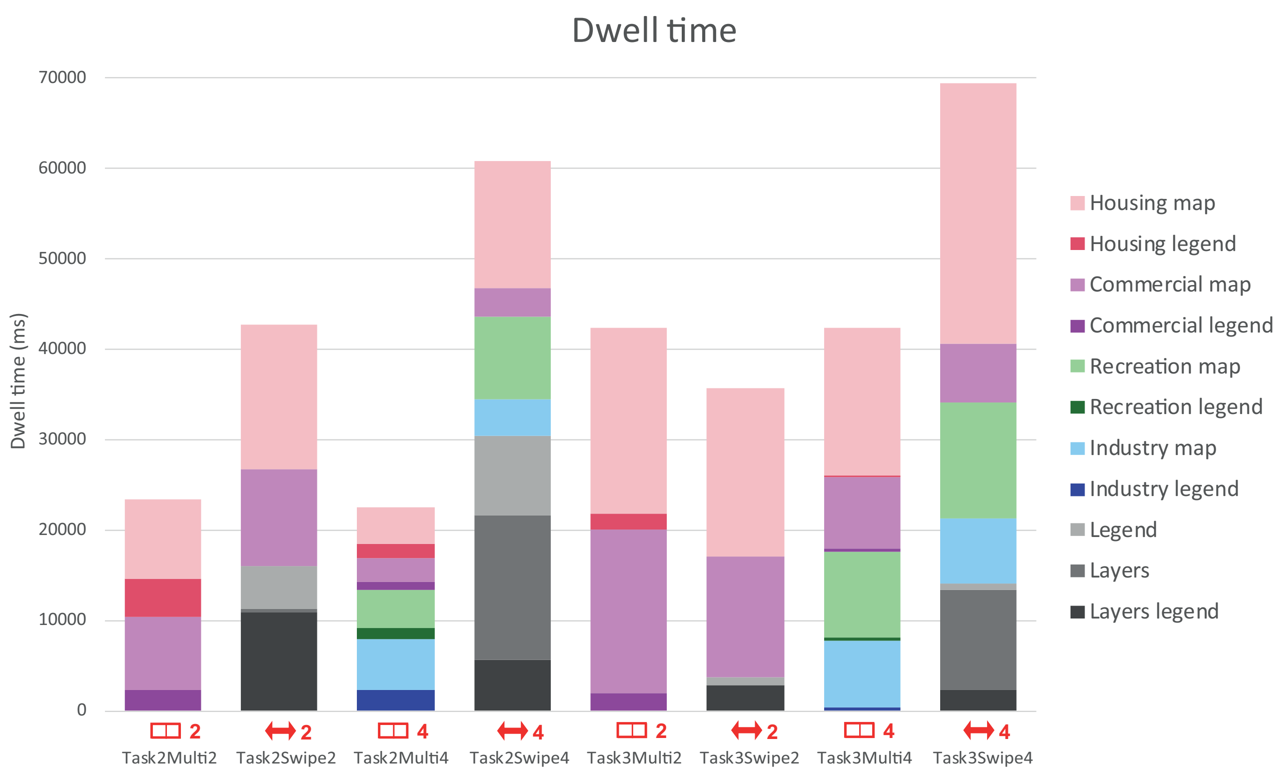


Figure 11. Dwell times for dynamic AOIs marked in the stimuli. The figure shows the length of time participants gazed at individual AOIs. The figure shows abnormal use of the legend and layers AOI in the case of swipe tasks.

avoid too many micro-swipes back and forth." However, this assumption was not supported by any objective data since only the ratio of correct answers and time needed to solve the task were evaluated. Our findings support these results and are verified by both quantitative and qualitative analyses of eye-movement data.

We employed eye-tracking to gain detailed insight into behavior of participants during performing of the map comparison tasks. Without its use, we would have to rely solely on the correctness of answers and trial duration. Direct observation of the participants during the work could provide insight into the use of swipe, but it it would not be possible to detect which map composition elements were observed and which were not. The disparity between map composition elements' (AOIs) visible time and dwell time is evident from Figures 10 and 11.

The results obtained from the eye-tracking analysis in the present study are consistent with those of Lobo et al. (2015). Tasks which employed swipe required more time to be solved and produced more incorrect answers. However, the qualitative analysis revealed that the most serious problems which occurred with swipe were caused by the selection of layers during the comparison of four maps. When swipe was correctly set up and used to compare only two maps, the results were comparable with multiple view, especially in more complex tasks (Task3Swipe2).

In the study by Lobo et al. (2015), participants in the experiment were not required to set up anything in the experiment's visualizations. The visualizations were completely ready, and participants simply had to analyze the content presented. In the present study, we tested a functional solution. We applied a real web map environment (Esri ArcGIS Online – Web App Builder) and real tasks which revealed the participants' problems in layer selection. Alternative methods for layer selection are emerging, but this issue is yet to be solved, especially when more than two layers are compared. A potential solution to this problem is similar to the solution applied in World Imagery Wayback, but

Percentage of time swipe function used

		Task2Swipe2	Task2Swipe4	Task3Swipe2	Task3Swipe4
P01	Expert	24.9	24.0	48.1	61.4
P02	Student	43.9	59.9	82.6	70.8
P03	Student	38.5	6.6	54.9	35.2
P04	Student	24.8	20.4	92.8	0.0
P05	Student	0.0	16.9	84.7	45.5
P06	Student	26.7	9.2	64.5	51.1
P07	Student	56.8	0.0	44.4	0.0
P08	Student	0.0	0.0	0.0	0.0
P09	Student	0.0	6.9	84.9	23.4
P10	Student	35.1	0.0	64.8	30.9
P11	Student	0.0	0.0	82.1	0.0
P12	Student	0.0	13.1	87.5	0.0
P13	Student	0.0	0.0	0.0	0.0
P14	Student	35.9	7.3	53.5	65.2
P15	Student	12.4	13.7	76.6	59.0
P16	Student	0.0	0.0	0.0	0.0
P17	Expert	0.0	5.8	81.9	0.0
P18	Expert	58.0	0.0	48.2	67.4
P19	Expert	0.0	0.0	77.6	34.0
P20	Expert	64.2	60.4	82.1	74.6
P21	Expert	0.0	0.0	14.6	0.0
P22	Expert	0.0	0.0	0.0	0.0
P25	Expert	0.0	0.0	17.5	0.0
P26	Expert	8.5	0.0	0.0	0.0
P27	Expert	0.0	0.0	0.0	0.0
average		**17.2**	**9.8**	**49.7**	**24.7**

Figure 12. Percentage of time when the swipe function was used for individual participants and individual tasks.

with the option to re-order multiple layers on each side of the swipe window. The challenge in future research is to develop an environment where the swipe function can be used to simultaneously compare more than two maps.

Limitations

In the study, data from 25 participants were evaluated. The number of participants is consistent with other usability studies of GIS interfaces. Unrau and Kray (2019) summarized 39 usability studies in the field of geographic information systems and calculated that the median number of participants in these studies was 18. Moreover, the power of the test was validated using G*Power post hoc analysis (Faul et al., 2007).

The most notable limitation of the study is that for the comparison of four maps, the use of swipe and multiple views is not informationally equivalent. For swipe, it is necessary to select layers to be compared, while for multiple views, all of the layers are visible in the default view. Moreover, the legend cannot be displayed during swipe comparison. We are aware that these drawbacks are caused by the implementation in the Esri environment, not by the swipe method itself. However, the Esri solution is widely used (not only) in public administration. In Czechia, Esri is the primary software used by all regional planning authorities. The use of an inappropriate map comparison tool can have far-reaching consequences (not only) in urban planning, but anywhere where multiple layers need to be compared. It could be interesting to perform a similar study with different technology (like Esri Wayback or Juxtapose.js).

Presenting the tasks in a fixed order can also be assessed as a limitation of the study. We chose this approach for technical reasons (all tasks were preloaded in Google Chrome tabs) and also to allow respondents to complete tasks from simpler to more complex. We acknowledge that it would have been more appropriate to vary the order of visualizations (multiple views and swipe) to reduce potential bias. However, we do not think that this procedure affected the results of the study. There was no learning effect because the tasks were not exactly the same.

Before the study, we performed a pretest to examine the design of the study and validate the whole process. Nevertheless, the difficulty of the tasks may have been set too low (see correctness of answers – Figure 7). This might have resulted in a ceiling effect, when some tasks were too easy, and all participants had correct answer.

Working with legends for multiple view and swipe differs significantly in the Esri Web App Builder user environment. In multiple view, it is possible to switch the legend on or off and work with multiple maps simultaneously. However, when swipe is used, it is not possible to display a legend. In this case, when the user clicks on the legend button, the swipe function is disabled and only one layer is displayed, preventing a visual comparison of the colors in the map and legend.

Using a layer button in the swipe web apps permits display of the legend indicated by the "Layers legend" during qualitative analysis. However, in Swipe2, which compared two maps, the layers were expanded so that the participants could see the legends directly. In Swipe4, which compared four maps, users had to expand the layers manually to display the legend. This was set by default because of the legend's size. In Swipe2, the legend fit the map's height, therefore it was expanded. In Swipe4, the legend for all the layers would exceed the map's height, therefore it was collapsed.

The software which is available for eye-tracking data analysis does not provide tools for the automated generation of sequence charts based on dynamic Areas of

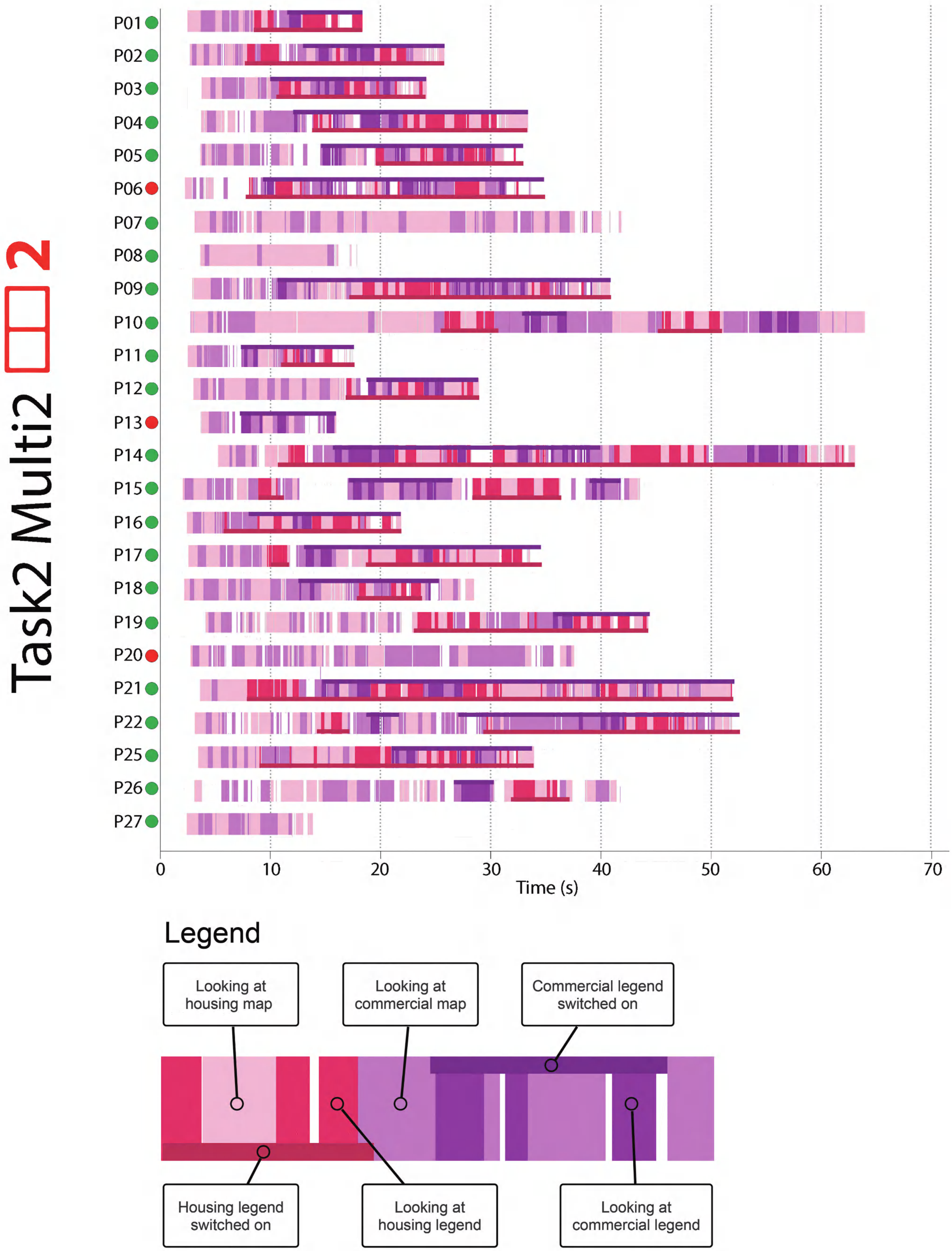


Figure 13. Sequence chart for Task2Multi2.

Interest. The development of such a tool would ease and accelerate future studies which explore interactivity with (not only) maps.

Conclusions

The main aim of the present study was to analyze the use of the swipe and multiple view method applied by users in map comparison tasks. For the comparison tasks, we used the ArcGIS Online environment to create web map applications with maps which depicted land suitability. Twenty-five participants (15 students and 10 experts) were asked to solve tasks similar to real map comparison situations. The user behavior was monitored with an eye-tracking method which allowed us to compare the accuracy of the participants' answers with the trial duration metrics related to the individual map elements. Based on our findings, we conclude that:

(1) Multiple view is a better method of map comparison than swipe, especially in a task which compares four maps. Multiple view is intuitive and more effective, especially in the case of fully synchronized windows (maps). No settings are required, and applications based on multiple view are ready to use immediately once they have loaded into the environment.
(2) The settings for swipe in the Esri environment are not intuitive and require additional settings. If more than two maps are compared, users not only have to change the swipe settings but also change the map (layer) order. In this case, users are confused which layer is the overlaid layer. Users tended to prefer a simple layer switching toggle. In this case, switching layers on or off was more effective than swipe comparison.
(3) The only situation where swipe outperformed multiple view was in Task 3, where two maps were displayed, and the participants had to compare four areas. This result may indicate that the swipe method is an advantage in more complex tasks but requires correct setting up since the displayed selection of layers appeared to be very complicated.
(4) The challenge in future research is to improve the swipe function for the comparison of more than two maps simultaneously. Moreover, the setup of the swipe tool in the Esri environment is not intuitive; users are not always able to clearly identify which layer overlays which. The problem could be solved by defining the overlay layer more clearly and also by programming a function that allows for the comparison of up to four different maps using the swipe tool (e.g. by moving the cursor in four basic directions).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This paper was supported by the project "Application of Geospatial Technologies for Spatial Analysis, Modelling, and Visualization of Spatial Phenomena" [IGA_PrF_2021_020] supported by the Internal Grant Agency of Palacký University Olomouc) and by the project [No. DG18P02OVV017] of the NAKI II research programme supported by the Ministry of Culture of the Czech Republic.

ORCID

Stanislav Popelka http://orcid.org/0000-0003-2127-2614
Jaroslav Burian http://orcid.org/0000-0003-0729-9757
Marketa Beitlova http://orcid.org/0000-0002-3935-5824

Data availability statement

The data that support the findings of this study are openly available in Mendeley Data at anonymised. The data is available at: http://dx.doi.org/10.17632/29wvkzj377.1

References

Andrienko, N., Andrienko, G., & Gatalsky, P. (2003). Exploratory spatio-temporal visualization: An analytical review. *Journal of Visual Languages & Computing*, *14*(6), 503–541. https://doi.org/10.1016/S1045-926X(03)00046-6

Barnum, C. M. (2010). *Usability testing essentials*. Morgan Kaufmann. https://doi.org/10.1016/C2009-0-20478-8

Bartling, M., Robinson, A. C., Resch, B., Eitzinger, A., & Atzmanstorfer, K. (2021). The role of user context in the design of mobile map applications. *Cartography and Geographic Information Science*, *48*(5), 432–448. https://doi.org/10.1080/15230406.2021.1933595

Bonanni, L., Xiao, X., Hockenberry, M., Subramani, P., Ishii, H., Seracini, M., & Schulze, J. (2009). Wetpaint: Scraping through multi-layered images. In D. R. Olsen & R. B. Arthur (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 571–574). New York: Association for Computing Machinery. https://doi.org/10.1145/1518701.1518789

Brügger, A., Fabrikant, S. I., & Çöltekin, A. (2017). An empirical evaluation of three elevation change symbolization methods along routes in bicycle maps. *Cartography and Geographic Information Science*, *44*(5), 436–451. https://doi.org/10.1080/15230406.2016.1193766

Burian, J., Brychtová, A., Vávra, A., & Hladišová, B. (2016). Analytical material for planning in Olomouc, Czech Republic. *Journal of Maps*, *12*(4), 649–654. https://doi.org/10.1080/17445647.2015.1060177

Burian, J., Stachova, M., & Vondrakova, A. (2018). Land suitability assessment of the Olomouc region: An application of an Urban Planner model. *Journal of Maps*, *14*(1), 73–80. https://doi.org/10.1080/17445647.2018.1493407

Burian, J., Stastny, S., Brus, J., Pechanec, V., & Vozenilek, V. (2015). Urban planner: Model for optimal land use scenario modelling. *Geografie*, *120*(3), 330–353. https://doi.org/10.37040/geografie2015120030330

Cartwright, W. (1997). New media and their application to the production of map products. *Computers & Geosciences*, *23*(4), 447–456. https://doi.org/10.1016/S0098-3004(97)00008-3

Crampton, J. W. (2002). Interactivity types in geographic visualization. *Cartography and Geographic Information Science*, *29*(2), 85–98. https://doi.org/10.1559/152304002782053314

Dong, W., Liao, H., Liu, B., Zhan, Z., Liu, H., Meng, L., & Liu, Y. (2020). Comparing pedestrians' gaze behavior in desktop and in real environments. *Cartography and Geographic Information Science*, *47*(5), 432–451. https://doi.org/10.1080/15230406.2020.1762513

Esri. (2021a). *Link multiple views*. https://pro.arcgis.com/en/pro-app/latest/help/mapping/navigation/link-multiple-views.htm

Esri. (2021b). *World imagery wayback*. https://livingatlas.arcgis.com/wayback/

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Freitas, C. M., Luzzardi, P. R., Cava, R. A., Winckler, M., Pimenta, M. S., & Nedel, L. P. (2002). On evaluating information visualization techniques. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (pp. 373–374). New York: Association for Computing Machinery. https://doi.org/10.1145/1556262.1556326

Gleicher, M. (2017). Considerations for visualizing comparison. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 413–423. https://doi.org/10.1109/TVCG.2017.2744199

Greenlee, M. W. (2000). Human cortical areas underlying the perception of optic flow: Brain imaging studies. *International Review of Neurobiology*, *44*, 269–292. https://doi.org/10.1016/s0074-7742(08)60746-1

Harrison, B. L., Ishii, H., Vicente, K. J., & Buxton, W. A. (1995). Transparent layered user interfaces: An evaluation of a display design to enhance focused and divided attention. In I. R. Katz (Ed.), *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 317–324). New York: Association for Computing Machinery. https://doi.org/10.1145/223904.223945

Harrower, M. (2007). The cognitive limits of animated maps. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *42*(4), 349–357. https://doi.org/10.3138/carto.42.4.349

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Krygier, J. B., Reeves, C., DiBiase, D., & Cupp, J. (1997). Design, implementation and evaluation of multimedia resources for geography and earth science education. *Journal of Geography in Higher Education*, *21*(1), 17–39. https://doi.org/10.1080/03098269708725407

Lee, L. Y., Talhelm, T., Zhang, X., Hu, B., & Lv, X. (2021). Holistic thinkers process divided-attention tasks faster: From the global/local perspective. *Current Psychology*, 1–13. https://doi.org/10.1007/s12144-021-01879-1

Liao, H., Dong, W., & Zhan, Z. (2021). Identifying map users with eye movement data from map-based spatial tasks: User privacy concerns. *Cartography and Geographic Information Science*, *49*(1), 1–20. https://doi.org/10.1080/15230406.2021.1980435

Lobo, M.-J., Pietriga, E., & Appert, C. (2015). An evaluation of interactive map comparison techniques. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 3573–3582). New York: Association for Computing Machinery. https://doi.org/10.1145/2702123.2702130

Long, J., & Robertson, C. (2018). Comparing spatial patterns. *Geography Compass*, *12*(2), e12356. https://doi.org/10.1111/gec3.12356

Luz, S., & Masoodian, M. (2014). Readability of a background map layer under a semi-transparent foreground layer. In F. Garzotto, P. Paolini, A. De Angeli, G. Jacucci, & A. Malizia (Eds.), *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces* (pp. 161–168). New York: Association for Computing Machinery. https://doi.org/10.1145/2598153.2598174

MacEachren, A. M., & Kraak, M.-J. (2001). Research challenges in geovisualization. *Cartography and Geographic Information Science*, *28*(1), 3–12. https://doi.org/10.1559/152304001782173970

MacEachren, A. M. (2004). *How maps work: Representation, visualization, and design*. Guilford Press.

Manson, S. M., Kne, L., Dyke, K. R., Shannon, J., & Eria, S. (2012). Using eye-tracking and mouse metrics to test usability of web mapping navigation. *Cartography and Geographic Information Science*, *39*(1), 48–60. https://doi.org/10.1559/1523040639148

Najdowski, A., Persicke, A., & Kung, E. (2014). Executive functions. In D. Granpeesheh, J. Tarbox, A. Najdowski, & J. Kornack (Eds.), *Evidence-Based Intervention for Children with Autism: The CARD Model* (pp. 353–385). Academic Press.

Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.

Opach, T., Popelka, S., Dolezalova, J., & Rod, J. K. (2017). Star and polyline glyphs in a grid plot and on a map display: Which perform better? *Cartography and Geographic Information Science*, *45*(5), 1–20. https://doi.org/10.1080/15230406.2017.1364169

Peterson, M. P. (1998). That interactive thing you do. *Cartographic Perspectives*, *29*(3–5), 2. https://doi.org/10.14714/CP29.671

Plumlee, M. D., & Ware, C. (2006). Zooming versus multiple window interfaces: Cognitive costs of visual comparisons. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *13*(2), 179–209. https://doi.org/10.1145/1165734.1165736

Popelka, S. (2014). Optimal eye fixation detection settings for cartographic purposes. In B. Oszczak (Ed.), *14th SGEM GeoConference on Informatics, Geoinformatics and Remote Sensing* (Vol. 1, pp. 705–712). Albena, Bulgaria: International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM.

Robinson, A. C., Chen, J., Lengerich, E. J., Meyer, H. G., & MacEachren, A. M. (2005). Combining usability techniques to design geovisualization tools for epidemiology. *Cartography and Geographic Information Science*, *32*(4), 243–255. https://doi.org/10.1559/152304005775194700

Roth, R. E., Çöltekin, A., Delazari, L., Filho, H. F., Griffin, A., Hall, A., Korpi, J., Lokka, I., Mendonça, A., Ooms, K., & van Elzakker, C. P. J. M. (2017). User studies in cartography: Opportunities for empirical research on interactive maps and visualizations. *International Journal of Cartography*, *3*(sup1), 61–89. https://doi.org/10.1080/23729333.2017.1288534

Roth, R. E. (2012). Cartographic interaction primitives: Framework and synthesis. *The Cartographic Journal*, *49*(4), 376–395. https://doi.org/10.1179/1743277412Y.0000000019

Roth, R. E. (2013). Interactive maps: What we know and what we need to know. *Journal of Spatial Information Science*, *2013*(6), 59–115. https://doi.org/10.5311/JOSIS.2013.6.105

Roth, R. E. (2015). Interactivity and cartography: A contemporary perspective on user interface and user experience design from geospatial professionals. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *50*(2), 94–115. https://doi.org/10.3138/cart.50.2.2427

Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing: How to plan, design and conduct effective tests*. Wiley Publising Inc.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In A. Duchowski, J. W. Senders, & K. S. Kam (Eds.), *Symposium on Eye tracking research & applications (ETRA 2000)* (pp. 71–78). New York: Association for Computing Machinery. https://doi.org/10.1145/355017.355028

Šašinka, Č., Stachoň, Z., Čeněk, J., Šašinková, A., Popelka, S., Ugwitz, P., Lacko, D., & Gao, S. (2021). A comparison of the performance on extrinsic and intrinsic cartographic visualizations through correctness, response time and cognitive processing. *PLoS One*, *16* (4), e0250164. https://doi.org/10.1371/journal.pone.0250164

Šašinka, Č., Stachoň, Z., Kubíček, P., Tamm, S., Matas, A., & Kukaňová, M. (2019). The impact of global/local bias on task-solving in map-related tasks employing extrinsic and intrinsic visualization of risk uncertainty maps. *The Cartographic Journal*, *56*(2), 175–191. https://doi.org/10.1080/00087041.2017.1414018

Scholtz, J. (2006). Beyond usability: Evaluation aspects of visual analytic environments. In *2006 IEEE Symposium On Visual Analytics Science And Technology* (pp. 145–150). Piscataway, N.J.: IEEE. https://doi.org/10.1109/VAST.2006.261416

Slocum, T. A., Blok, C., Jiang, B., Koussoulakou, A., Montello, D. R., Fuhrmann, S., & Hedley, N. R. (2001). Cognitive and usability issues in geovisualization. *Cartography and Geographic Information Science*, *28*(1), 61–75. https://doi.org/10.1559/152304001782173998

Tullis, T. S. (2007). Older adults and the web: Lessons learned from eye-tracking. In C. Stephanidis (Ed.), *Universal acess in human computer interaction. coping with diversity* (pp. 1030–1039). Springer. https://doi.org/10.1007/978-3–540-73279-2_115

Unrau, R., & Kray, C. (2019). Usability evaluation for geographic information systems: A systematic literature review. *International Journal of Geographical Information Science*, *33*(4), 645–665. https://doi.org/10.1080/13658816.2018.1554813

Yi, J. S., Ah Kang, Y., Stasko, J., & Jacko, J. A. (2007). Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, *13*(6), 1224–1231. https://doi.org/10.1109/TVCG.2007.70515

Table 1. Links to web map applications used in the study.

	2			
TASK 2	Task2Multi2	Task2Swipe2	Task2Multi4	Task2Swipe4
TASK 3	Task3Multi2	Task3Swipe2	Task3Multi4	Task3Swipe4

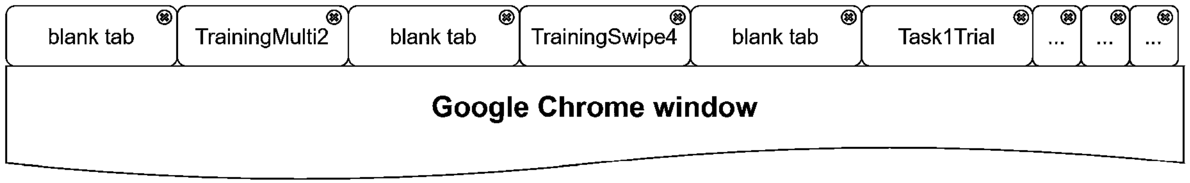


Figure 1. Preloaded Google Chrome tabs with tasks.

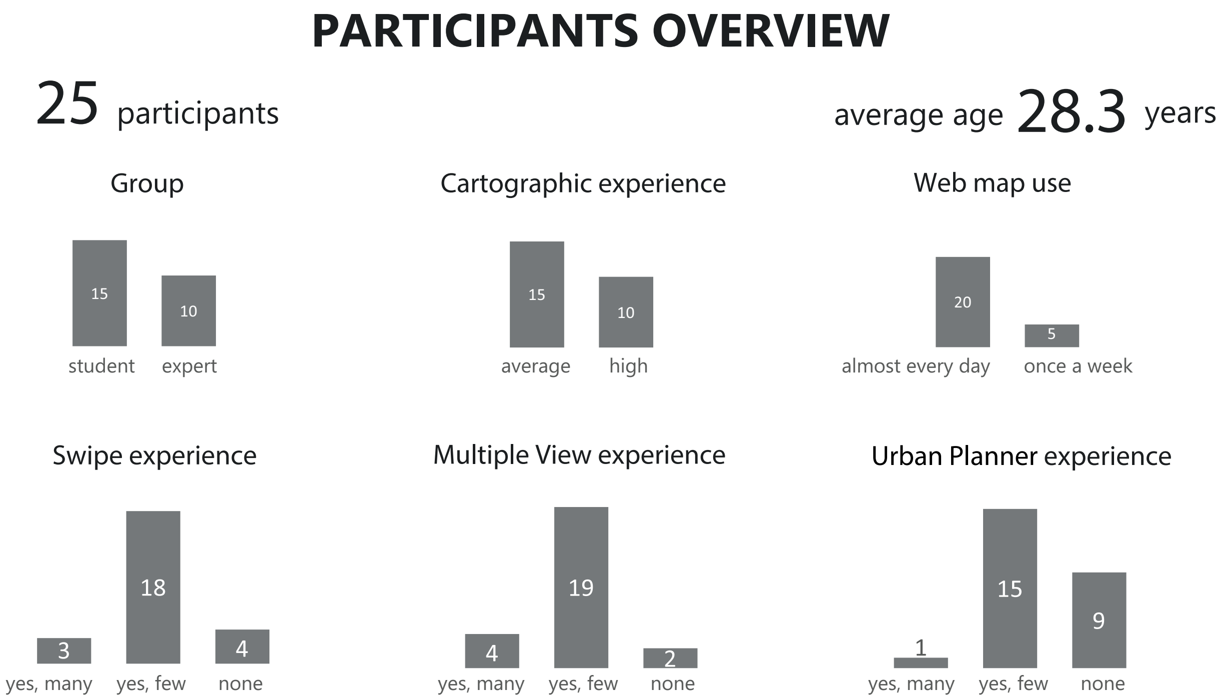


Figure 2. Overview of the participants' characteristics.

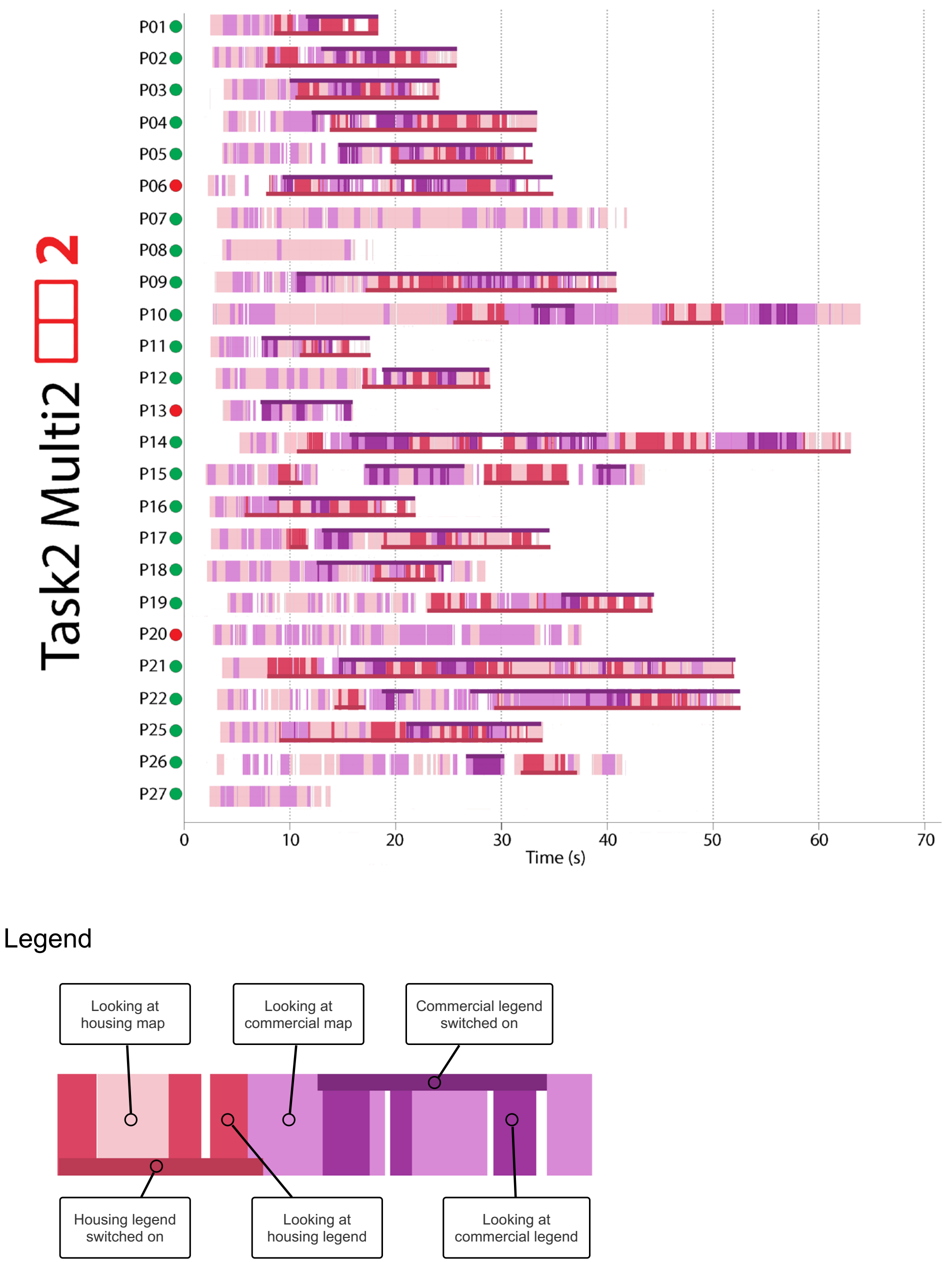
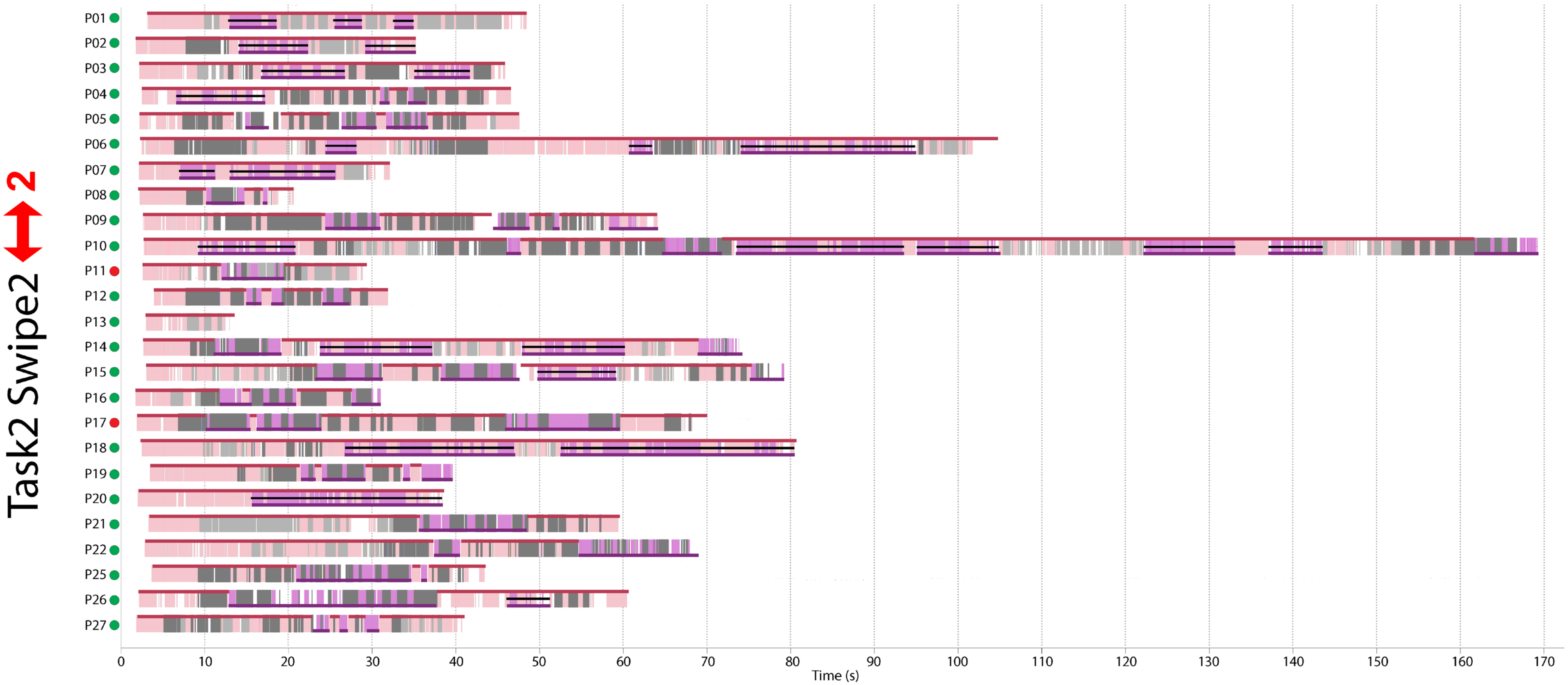


Figure 3. Sequence chart for Task2Multi2.



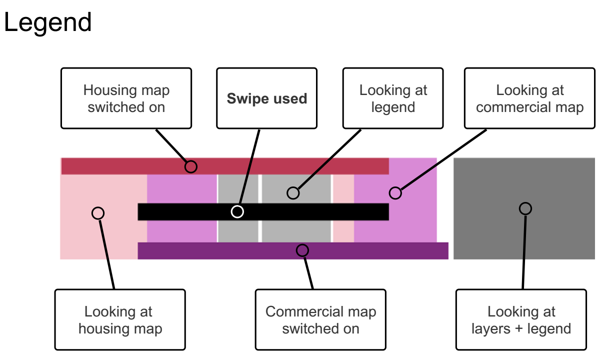


Figure 4. Sequence chart for Task2Swipe2.

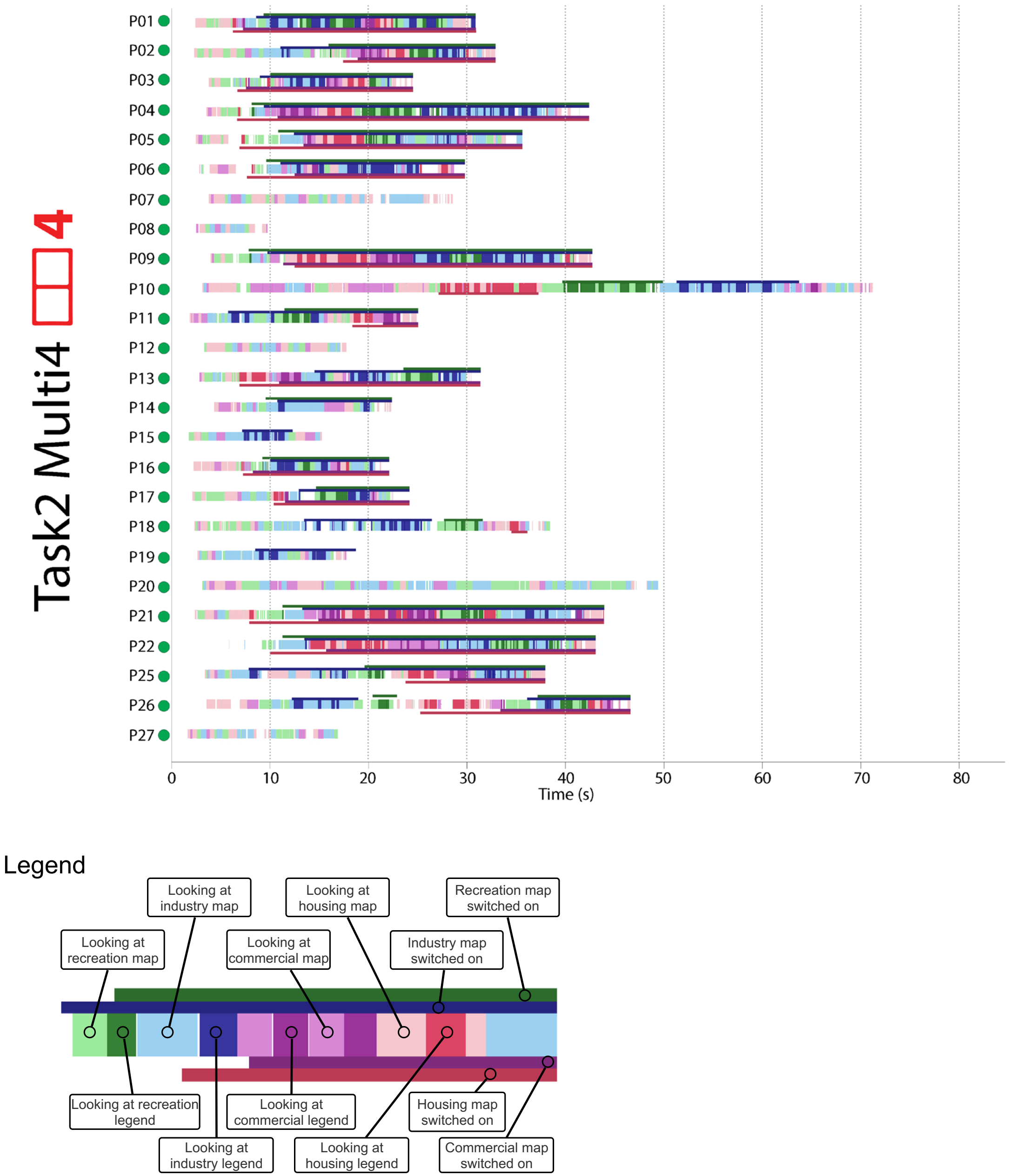


Figure 5. Sequence chart for Task2Multi4.

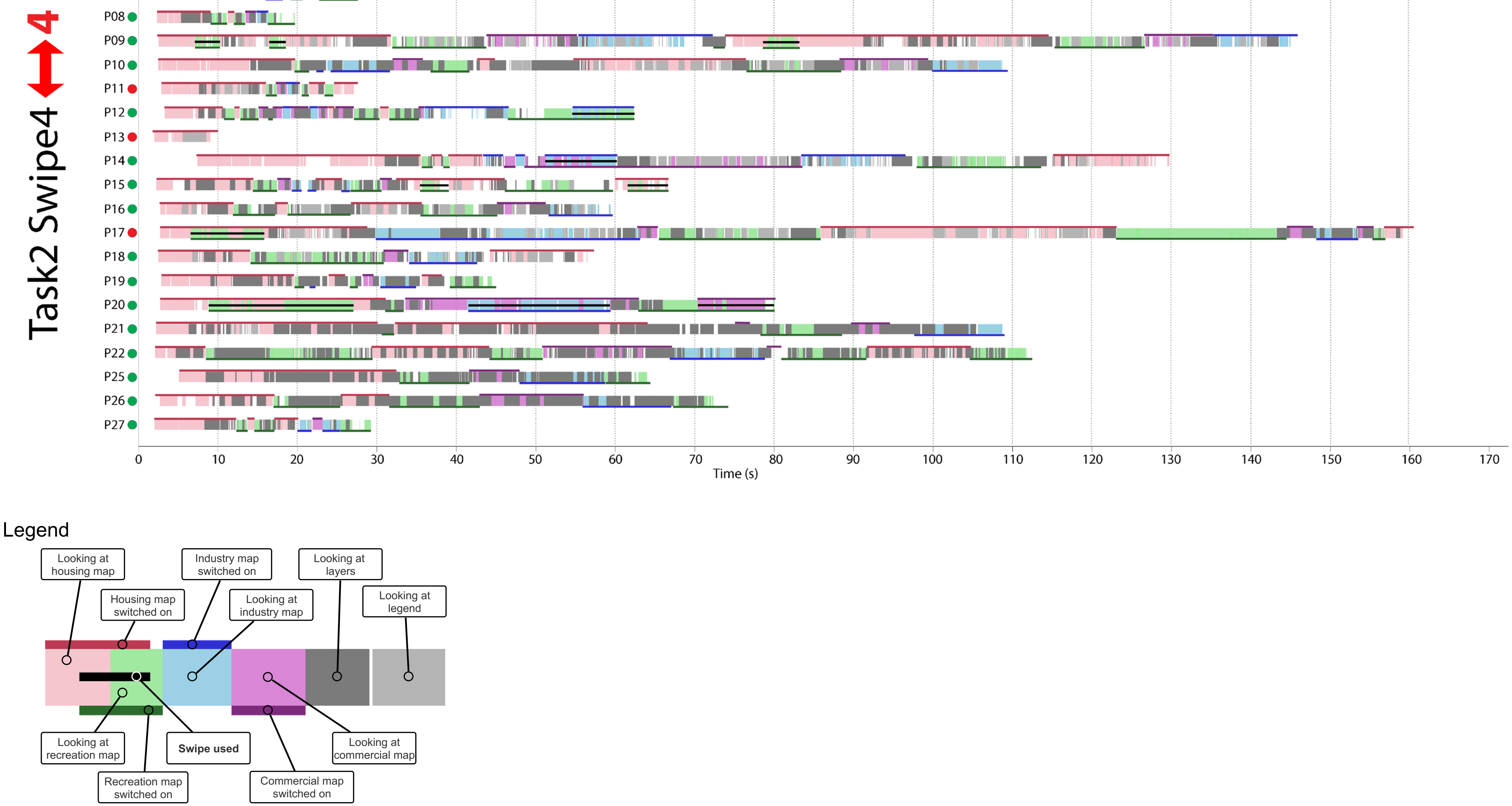


Figure 6. Sequence chart for Task2Swipe4.

Figure 7. Sequence chart for Task3Multi2.

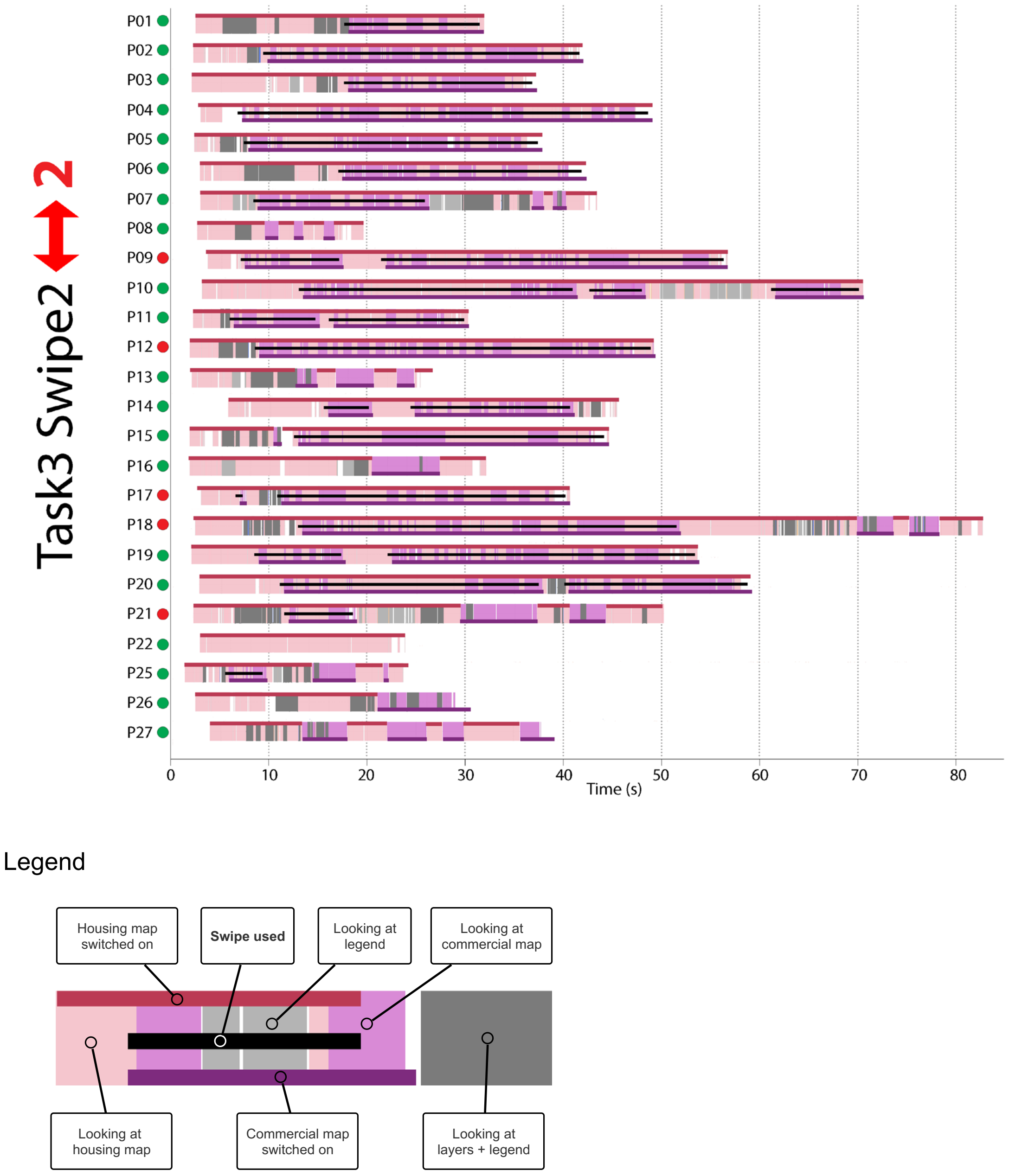


Figure 8. Sequence chart for Task3Swipe2.

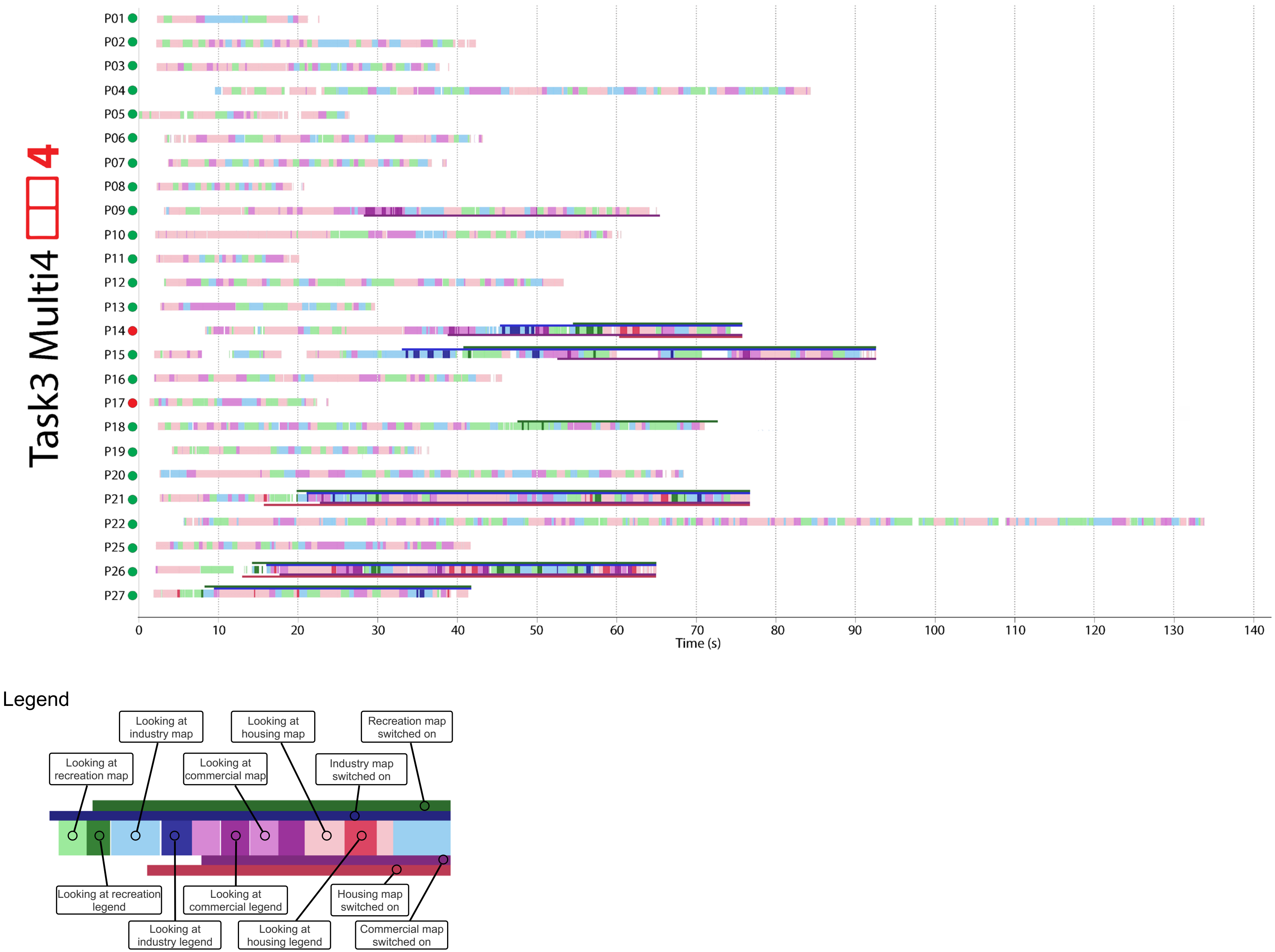


Figure 9. Sequence chart for Task3Multi4.

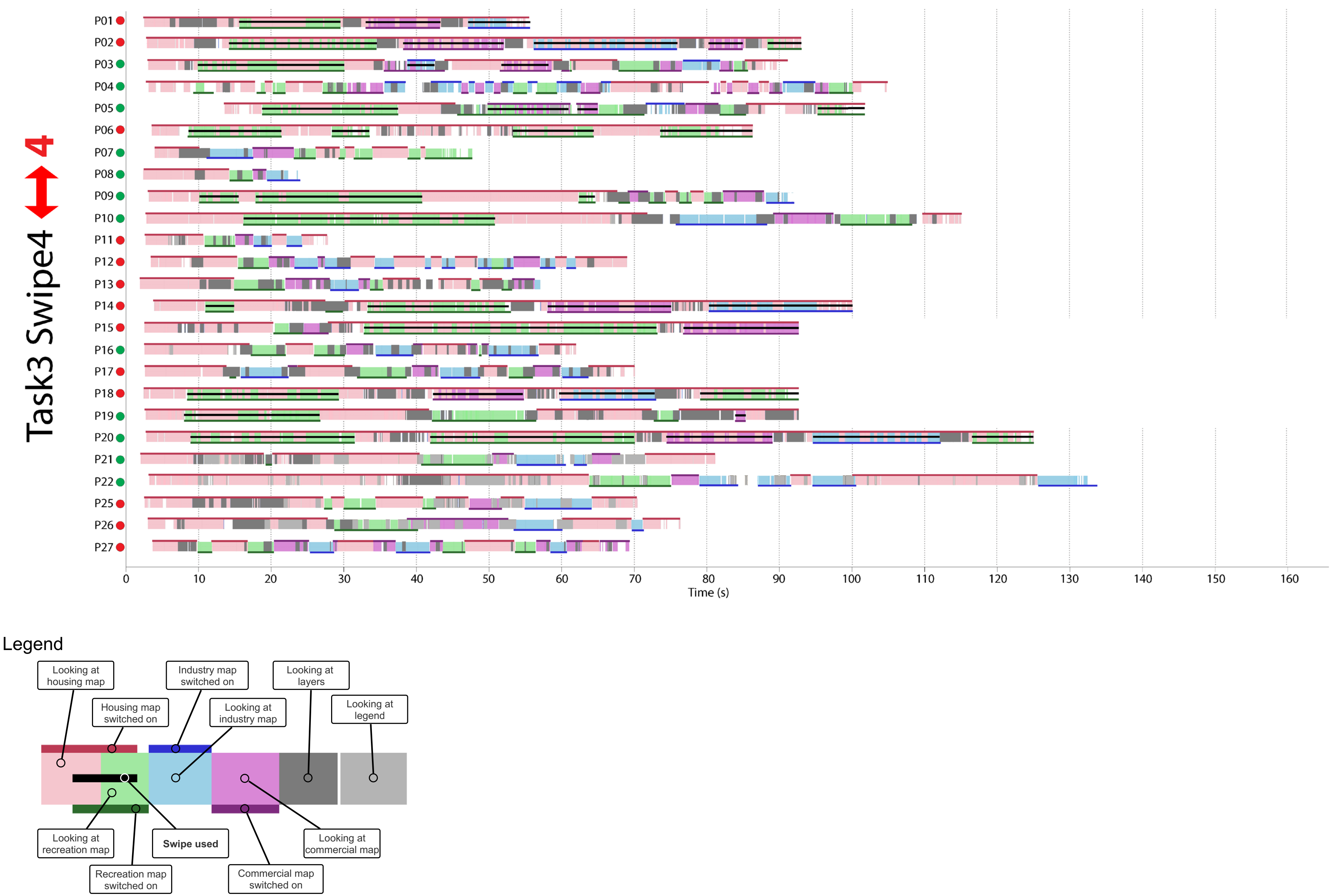


Figure 10. Sequence chart for Task3Swipe4.

Paper GazePlotter

Popelka, S.*, Komínek, J., & Vojtechovska, M. (2024). Exploring geological map usability through sequence chart visualization: Expert vs. novice perspectives. In *2024 Symposium on Eye Tracking Research and Applications (ETRA '24), June 04–07, 2024, Glasgow, United Kingdom* (pp. 1–7). ACM.

[IF 0[2]]

[0 citations on WoS; 0 citations without autocitations]

[Author's contribution: 40%]

Abstract

The paper presents a detailed analysis of how geological maps are read and interpreted differently by experts in geology and those new to the field. It primarily focuses on three eye-tracking experiments. The first experiment employs a remote eye-tracker to evaluate user reading of scans of geological maps presented on the screen. In the second experiment, the usability of the Czech Geological Survey's (CGS) online map application was also evaluated using a remote eye-tracker. In the last experiment, paper geological maps were used as stimuli, and eye movement data were recorded using eye-tracking glasses. Participants in the study were categorized into three groups: geologists, geographers, and geoinformatics professionals. Recorded data were visualized using a newly developed open-source online tool called GazePlotter to visualize sequence charts. A variety of experiments helped us to showcase the possibilities of GazePlotter. Findings indicate that geologists tend to concentrate more on the map content than the other groups and generally spend less time completing tasks. The study also uncovers several usability issues within the CGS online map application.

[2] *The methodological paper describing the tool and its functionality is currently under review in Behavior Research Methods (Vojtechovska & Popelka, n.d.). (IF 4.600)*

Exploring Geological Map Usability Through Sequence Chart Visualization

Expert vs. Novice Perspectives

Stanislav Popelka*
stanislav.popelka@upol.cz
Department of Geoinformatics,
Palacký University Olomouc
Olomouc, Czechia

Jiří Komínek
kominekjiri@seznam.cz
Department of Geoinformatics,
Palacký University Olomouc
Olomouc, Czechia

Michaela Vojtechovska
mail@vojtechovska.com
Department of Geoinformatics,
Palacký University Olomouc
Olomouc, Czechia

ABSTRACT

The paper presents a detailed analysis of how geological maps are read and interpreted differently by experts in geology and those new to the field. It primarily focuses on three eye-tracking experiments. The first experiment employs a remote eye-tracker to evaluate user reading of scans of geological maps presented on the screen. In the second experiment, the usability of the Czech Geological Survey's (CGS) online map application was also evaluated using a remote eye-tracker. In the last experiment, paper geological maps were used as stimuli, and eye movement data were recorded using eye-tracking glasses. Participants in the study were categorized into three groups: geologists, geographers, and geoinformatics professionals. Recorded data were visualized using a newly developed open-source online tool called GazePlotter to visualize sequence charts. A variety of experiments helped us to showcase the possibilities of GazePlotter. Findings indicate that geologists tend to concentrate more on the map content than the other groups and generally spend less time completing tasks. The study also uncovers several usability issues within the CGS online map application.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

geological maps, eye-tracking, user experience, map usability

ACM Reference Format:
Stanislav Popelka, Jiří Komínek, and Michaela Vojtechovska. 2024. Exploring Geological Map Usability Through Sequence Chart Visualization: Expert vs. Novice Perspectives. In *2024 Symposium on Eye Tracking Research and Applications (ETRA '24), June 04–07, 2024, Glasgow, United Kingdom.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3649902.3653520

ETRA '24, June 04–07, 2024, Glasgow, United Kingdom

ACM ISBN 979-8-4007-0607-3/24/06
https://doi.org/10.1145/3649902.3653520

1 INTRODUCTION

Geological maps are often regarded as some of the most complex types of maps to construct and interpret in terms of content analysis [Sayidov et al. 2020]. Geological maps show the distribution at the earth's surface of different kinds of rocks. Maltman [2012] describes a geological map as one created by adding geological elements to a simplified topographic base. Recently, web services and web mapping applications have increasingly emerged, gradually pushing paper maps into the background. Over the years, geological mapping has also undergone development. With the advancement of technologies, geological maps are penetrating a variety of fields and are becoming increasingly accessible to both professionals and the general public. Given these developments, understanding how readers from different educational backgrounds engage with these maps becomes essential. Conducting empirical studies on the map reading process can illuminate how users comprehend these complex cartographic works. Such insights could lead to better map design and dissemination strategies, ultimately enhancing public and professional engagement with geological maps.

Kübler and Voisard [1999] emphasize the importance of a user-friendly interface for geologic maps. Maltese et al. [2013] described a study where eye-tracking glasses were used in the field of geology. As part of the research, several selected students were given glasses to record their work during fieldwork in geology. The aim was to reveal what and how students pay attention during exercise. For example, how often they use the map. For relevant results and good accuracy, the user must be calibrated several times during testing. As a result, the authors watched the user's view through the recorded video rather than the eye-movement data itself.

Çöltekin et al. [2017] compared two variants of the legend of soil maps. A total of 19 participants were divided into three groups: experts, occasional users, and novices. The maps contained two variants of legend sorting: alphanumeric characters and colors. It was found that the organization of the legend does not play a very large role, and the users' preference for map legends was distributed almost equally.

This paper focuses on contrasting how geologists (experts) and non-geologists (novices) engage with geological maps, representing a comparison between those with specialized training and general audiences. Previous studies [Beitlova et al. 2020; Brychtová and Coltekin 2016; Ooms et al. 2014] have explored the disparities in map reading skills between expert and novice groups through eye-tracking methodologies. Utilizing the innovative tool GazePlotter, which creates sequence charts from eye-tracking data, this research visually represents how these two groups read maps. Furthermore, the study aims to showcase the capabilities of GazePlotter in conducting qualitative analyses of eye movement patterns.

The research questions for this paper were developed in collaboration with the Czech Geological Survey and are as follows: RQ1: How does prior geological education influence the efficiency of using geological maps?; RQ2: What are the usability challenges of the Czech Geological Survey's online geological map?; RQ3: How does the sequence chart visualization enhance the understanding of how users work with geological maps?

2 METHODS

The study was conducted in three segments. In the first part, we utilized an SMI RED 250 remote eye-tracker to observe participants' responses to static stimuli, specifically scanned images of geological maps. The second part employed the same eye-tracker but focused on dynamic stimuli, using an online geological map. For the third segment, we switched to using SMI Glasses 2 to examine how participants interacted with a physical, paper geological map as stimuli.

In each part, the participants had to solve tasks above geological maps. Only a few tasks were selected for this paper. However, the complete list of tasks is attached as supplementary material. The expanded open-source tool GazePlotter [Vojtechovska and Popelka 2023] was used to generate sequence charts to visualize the results of all three parts. The tool is freely available at https://gazeplotter.com. We analyzed the time needed to solve the task, fixation count in AOI, and AOI entry time. For statistical testing, Nemenyi's non-parametric all-pairs comparison test was used on Kruskal-type ranked data, with a significance level of $\alpha = 0.05$.

2.1 Participants

Individuals selected as participants were those who currently work with geological maps or are anticipated to work with them in the future. In total, three groups were selected: students of geoinformatics (GIS; n=24), students and staff of the geology department (GEOL; n=10), and students of the geography department (GEO; n=10). The GEOL group thus represented respondents with geological education. The GIS group represented non-geologists, as the respondents had not undergone any geological course. The GEO group was chosen so that the test subjects had only completed a basic course in geology. All the respondents participated voluntarily and were involved in all three performed studies.

2.2 Scanned maps testing

The first part of the experiment used cutouts of scanned geological maps as stimuli. Participants were solving map reading and analysis tasks above these stimuli. The tasks were focused on feature identification, reading information from the map, working with the accompanying diagram, and estimating distances. Respondents recorded their answers using mouse clicks into maps or voice.

2.3 Web map application testing

Testing of map applications is a vital part of both informatics and cartography development. The results should contribute to the improvement and enhancement of services for end-users. The CGS Geological Map 1:50,000 is the most used product in the field of geological web maps. Therefore, users must be able to work with it quickly and efficiently.

The experiment focused on a map application and employed screen-recording stimuli. SMI RED 250 eye-tracker with a sampling frequency of 250 Hz was employed in the study. The study included 15 tasks chosen to cover the widest range of possible activities that could occur during a user's interaction with the application.

The first task was pre-training; participants had unlimited time to explore the map's functionality. Before each task, the instructions were displayed for an unlimited amount of time. The tasks were focused on different map functions, searching for areas, making comparisons, etc. All stimuli were presented without a time limit, so it was up to the respondent to decide when to move on to the next stimulus. Respondents who were still looking for a solution within 60 seconds were helped to prevent discouragement. Moreover, some of the tasks were interconnected, and failure in one might affect the results of others. Following the testing, respondents were asked about their experience, focusing on any challenges they faced and notable aspects they observed while using the application.

2.4 Paper maps testing

The last experiment was conducted using ET glasses due to their ability to evaluate paper geological maps in their entirety. This was essential since map sheets are often large, and evaluating them on a monitor requires cropping, which does not allow for a complete map assessment. SMI Glasses 2 with a sampling frequency of 60 Hz were used in the study. A notable drawback of this device is the complexity of data analysis compared to that obtained through a stationary eye-tracker or using more advanced glasses allowing the use of snapshots.

The study included six tasks; however, only the first task is discussed in this paper. This initial task involved free-viewing without any specific objectives for an unlimited time. This open-ended observation aimed to identify which sections of the map garnered the most interest across all participant groups.

3 RESULTS

3.1 Scanned maps

The experiment with scanned geological maps presented as static (image) stimuli contained a total of nine tasks. In these tasks, the differences between the three studied groups of participants were sought. For most tasks, the GEOL group was the fastest. It was also confirmed that GEOL participants were faster at reading legends than the other groups. Their speed in completing tasks and reading legends more efficiently than other groups can be attributed to their extensive training and experience in geology. This background likely gives them an intuitive understanding of geological symbols, terminology, and map layouts, enabling them to navigate and interpret map legends faster and more accurately.

The largest differences between the groups were observed in a task where participants had to mark the geological cross-section, which was depicted in a schema below the map. Fixations in the Cross section in map AOI indicated the proper approach. In contrast, fixations to the rest of the map indicated that participants do not know where the cross section is located. GEOLs needed significantly fewer fixations than the other groups. The median of the fixation count for GEOL was 84, contrasting with 301 for GEO and 206 for GIS. The difference between GEOL and GEO is statistically significant (p=0.041). This data is visualized in Figure 1, where it is evident that, in addition to a higher resolution speed, GEOLs also looked much less (except for participant geol_08) at those parts of the map where the cross section was not present (red segments in Figure 1). Moreover, the accuracy of the answers was much higher for GEOL than for the other groups.

This efficiency suggests that GEOLs possess a more targeted approach to navigating maps, likely due to their familiarity with geological features and their representation on maps. Their training and experience enable them to focus quickly on relevant areas, reducing the need for extensive searching and thereby increasing their task efficiency.

3.2 Web map application

The second part of the experiment reveals insights into user interactions with a geological web map application, highlighting challenges and efficiency in task completion. Participants encountered varying difficulty levels across different tasks, from locating towns to identifying geological features and adjusting map layers. The average times to complete tasks indicated that while some tasks were completed swiftly, others took longer, reflecting the complexity and user familiarity with specific features.

The study also pointed out challenges in navigating the application and managing map layers. One of the most important issues was the difficulty in finding a function for changing base maps. The application contained an icon for switching the base maps in the top-right corner of the interface (AOI "Icon"). The icon looked like four squares. After clicking that icon,the window with options for selecting base maps appeared (AOI "Select")." However, many participants tried to find this option in the panel on the left (i.e., AOI "Layers"). The content of that panel can be changed from Layers to Data, Legend, SearchBar, or Info. The proper strategy to switch the basemap to orthophoto was to find the proper Icon (purple segments in Figure 2) and then select orthophoto in the Select window (green segments in Figure 2).

Due to the dynamic nature of the tested stimulus, dynamic AOIs had to be defined in the BeGaze software. Definitions of these AOIs were then exported as XML files. GazePlotter has the functionality to display information about the visibility of Areas of Interest for individual participants according to these XML files. This information is depicted as a dashed line in the color corresponding to AOI in Figure 2. This functionality saved a significant amount of manual work in preparing visualizations, which until now had to be done by hand (see [Popelka et al. 2022]).

The average gaze entry time for the icon for switching base maps was 8.9s for GIS, 12.9s for GEO, and 20.8s for GEOL. It is interesting that for metric time to a first mouse click, the differences are much smaller (GIS 20.6s, GEOL 24.7s, and GEO 27.9s). However, many participants spent considerable amounts of time in the Layers, visualized as orange segments in Figure 2. The GEO group looked there for 22% of the time, and the other two groups spent 15% of the stimulus observation time). Participants also tried to find this function by clicking on different panels on the left (visualized by dashed lines and segments with corresponding colors in Figure 2).

The insights suggest that enhancing user interface design, improving application performance, and providing clearer navigational cues could significantly improve usability and efficiency for users of the mapping application. All the results were provided to the Czech Geological Survey.

3.3 Paper maps

In the last experiment, participants freely observed the paper map while wearing eye-tracking glasses. To analyze recorded data, it was necessary to assign data to Areas of Interest marked on the map. The semantic Gaze Mapping extension, which involves the operator reviewing recorded videos while assigning fixations to predefined AOIs by clicking on a reference image, was unavailable in the laboratory. An alternative was creating dynamic areas of interest, which was extremely time-consuming. Processing 30 seconds of footage took nearly an hour. Therefore, the final videos were manually reviewed, and for predefined 2-second intervals, they were recorded into a CSV file, which was the part of the map the participants focused on. Using one respondent as an example, we compared the results obtained from this manual assignment with data from dynamic areas of interest. It was found that the results are comparable.

As a result of this manual assignment, we obtained table data indicating the specific parts of the map each respondent focused on during the predefined 2-second intervals. Although the data were not exported from any eye-tracking software, it is possible to display them in GazePlotter since it can read data from custom CSV. The resulting visualization is displayed in Figure 3.

The analysis showed that GEO participants dedicated the longest period to freely exploring the paper map, taking up to 90.8 seconds. In contrast, GIS inspected the map for only 44.3 seconds. Significant differences between groups were observed regarding the distribution of attention among Areas of Interest marked on the map. GEOL and GEO investigated the map element about 55% of the time. However, GIS spent only 32% of their time there. In contrast, they looked into additional composition elements of the map (schemas) more than twice as long as the other two groups (19 % vs 8 % of observation time). Moreover, they spent more time inspecting the legend.

The study's focus on a geological map, a type less familiar to GIS, explains their different approach. Skilled in digital mapping, GIS spent less time on the map, focusing more on schemas and legends, indicative of their training in extracting key information from such elements. In contrast, GEO and GEOL regularly use geological maps and spend more time examining the map's detailed geological features. This difference indicates how professional background influences map interpretation, suggesting the need for tailored training and design in map-based systems to cater to various user groups' expertise and needs.

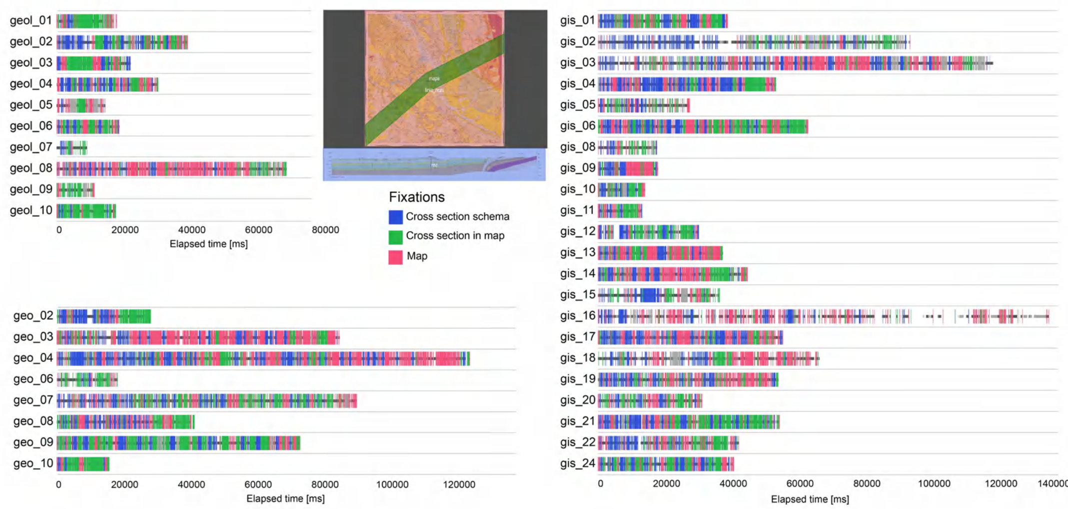


Figure 1: Visualization of the work with scanned geological map using sequence charts generated by GazePlotter.

Figure 2: Visualization of the work with geological web map application using sequence charts generated by GazePlotter.

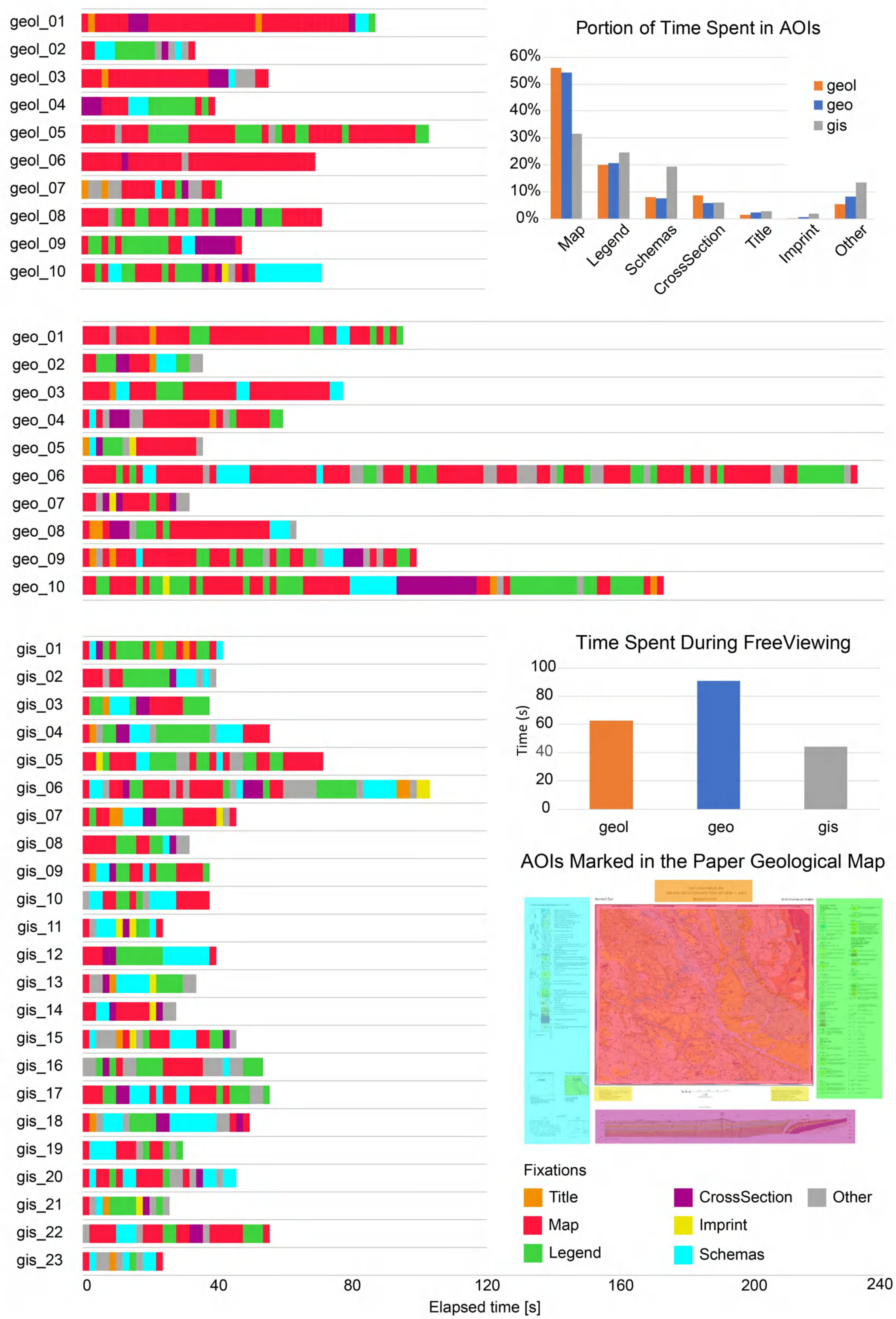


Figure 3: Visualization of the results of geological paper map free viewing using sequence charts generated by GazePlotter.

4 LIMITATIONS

The studies' methodology demonstrated limitations, particularly concerning the variety of participants and the data analysis methods employed. The study's sample size is modest, with an unequal distribution of participants across the selected groups of geologists, geographers, and geoinformatics professionals, which could skew results and limit their generalizability. Moreover, all participants had some education in Earth sciences. Incorporating individuals from distinctly different fields, such as the humanities, might have offered insights into how diverse backgrounds influence map reading, enriching the study's breadth. The tasks chosen for analysis do not encompass all potential interactions with geological maps, omitting scenarios that could reveal additional usability challenges or cognitive processes. Lastly, the approach to analyzing paper map interactions relied on manual assignment of data to Areas of Interest (AOIs), a method vulnerable to human error. Adopting image recognition software for this task could streamline data analysis and reduce potential biases.

5 DISCUSSION AND CONCLUSION

The culmination of this article underscores the differences in map reading and interpretation abilities between geologists and non-geologists, shedding light on the proficiency and challenges each group faces when engaging with geological maps. Through a detailed examination using both static and dynamic stimuli alongside traditional paper maps, this study showcases the capabilities of the innovative GazePlotter tool in visualizing and analyzing eye-tracking data across various map formats. The findings reveal that GEOL, with their specialized training and familiarity with geological symbols and layouts, demonstrates a more efficient and focused approach to navigating maps, particularly evident in tasks requiring the identification of specific geological features. This efficiency is attributed to their reduced need for fixations and their ability to bypass irrelevant map sections, a testament to their expertise.

Moreover, the exploration into web map applications highlights significant user interface challenges, pointing towards a need for improved design and functionality to enhance user experience and performance. The difficulties encountered in navigating the application and managing map layers emphasize the importance of intuitive design and accessible features to accommodate both expert and novice users.

The paper map analysis further illustrates how professional backgrounds influence engagement with geological maps. GEOL and GEO, accustomed to the intricate details of geological maps, dedicated more time to examining map features. In contrast, GIS, with an affinity for digital mapping, allocated more time to understanding schemas and legends. This variation underlines the importance of considering user-specific needs and backgrounds in the development and design of map-based systems.

In summary, this article not only highlights the differences in map reading skills between experts and novices but also presents GazePlotter as a valuable tool for qualitative eye movement analysis. The insights gained from this research offer valuable contributions to the fields of cartography and geoinformatics, suggesting pathways for enhancing the usability and accessibility of geological maps for a diverse range of users. The collaboration with the Czech Geological Survey ensures that these findings will contribute to the ongoing improvement of geological mapping tools, ultimately benefiting both professional and general audiences in their interaction with geological data.

ACKNOWLEDGMENTS

The article was supported by the project of the Czech Science Foundation, "Identification of Barriers in the Process of Communication of Spatial Socio-Demographic Information" (Grant No. 23-06187S).

REFERENCES

Marketa Beitlova, Stanislav Popelka, and Vit Vozenilek. 2020. Differences in Thematic Map Reading by Students and Their Geography Teacher. *ISPRS International Journal of Geo-Information* 9, 9 (2020), 492. https://doi.org/10.3390/ijgi9090492

Alžběta Brychtová and Arzu Coltekin. 2016. The effect of spatial distance on the discriminability of colors in maps. *Cartography and Geographic Information Science* (2016), 229–245. https://doi.org/10.1080/15230406.2016.1140074

Stefanie Kübler and Agnès Voisard. 1999. GeoHyp: an adaptive human interface for geologic maps and their databases. *ISPRS journal of photogrammetry and remote sensing* 54, 4 (1999), 234–243.

Adam V Maltese, Russell N Balliet, and Eric M Riggs. 2013. Through their eyes: Tracking the gaze of students in a geology field course. *Journal of Geoscience Education* 61, 1 (2013), 81–88.

Alex Maltman. 2012. *Geological maps: an introduction.* Springer Science & Business Media.

Kristien Ooms, Philippe De Maeyer, and Verlee Fack. 2014. Study of the attentive behavior of novice and expert map users using eye tracking. *Cartography and Geographic Information Science* 41, 1 (2014), 37–54.

Stanislav Popelka, Jaroslav Burian, and Marketa Beitlova. 2022. Swipe versus multiple view: a comprehensive analysis using eye-tracking to evaluate user interaction with web maps. *Cartography and Geographic Information Science* (2022), 1–19. https://doi.org/10.1080/15230406.2021.2015721

Azimjon Sayidov, Meysam Aliakbarian, and Robert Weibel. 2020. Geological map generalization driven by size constraints. *ISPRS International Journal of Geo-Information* 9, 4 (2020), 284.

Michaela Vojtechovska and Stanislav Popelka. 2023. GazePlotter – tool for eye movement sequences visualization. *Abstracts of the ICA* 6 (08 2023), 1–2. https://doi.org/10.5194/ica-abs-6-264-2023

Arzu Çöltekin, Alžběta Brychtová, Amy Griffin, Anthony C Robinson, Mark Imhof, and Chris Pettit. 2017. Perceptual complexity of soil-landscape maps: a user evaluation of color organization in legend designs using eye tracking. *International Journal of Digital Earth* 10, 6 (2017), 560–581.

A SUPPLEMENTARY MATERIAL

A.1 Scanned maps testing

Aparatus:

- SMI RED 250

Tasks:

(1) Find the abandoned quarry on the map. (Map S1)
(2) Locate the identified fault on the map. (Map S1)
(3) Find the detachment edge of the landslide. (Map S1)
(4) Mark the course of the geological cross-section. (Map S1)
(5) Explain what the marked symbol denotes. (Map S2)
(6) Determine the number of the adjacent map sheet at the bottom edge. (Map S3)
(7) Determine which of the indicated rocks is older. (Map S4)
(8) Mark the community waste landfill. (Map S5)
(9) Examine the map. (Map 2)
(10) Find Kozákov hill (center of the geological section). (Map 2)

Stimuli:

- Map S1 - Scanned geological map "ZM_25_03-431 Lomnice nad Popelkou"
- Map S2 – Scanned geological map "ZM_25_03-342 Turnov"

- Map S3 – Scanned geological map "Zm_tvary_25_03-413 Semily"
- Map S4 – Scanned geological map "Zm_25_03-342_ Rovensko pod Troskami"
- Map S5 – Scanned geological map "Zm_25_geofaktory_32_23 Černá v Pošumaví"

A.2 Web map application testing

Apparatus:

- SMI RED 250

Tasks:

(1) Examine the map and try its functions.
(2) Locate the town of Moravská Třebová.
(3) Mark the body of water near Moravská Třebová.
(4) Turn on the borehole exploration layer.
(5) Find the village of Chvalkov and determine the number of the map sheet it lies on.
(6) Display the legend and click into the map's legend.
(7) Determine the dominant rock in the selected area (Chvalkov).
(8) Search for the village of Brteč.
(9) How far apart are the centers of Brteč and Svareň?
(10) Change the base map to orthophoto.
(11) Add a soil map to the map.
(12) Set the soil map's transparency to 50
(13) Move the soil map below the geological one.
(14) Search the map for areas with occurrences of loess.
(15) Print the requested area including the legend.

Stimuli:

- Geological web map https://mapy.geology.cz/geocr50/

A.3 Paper maps testing

Apparatus:

- SMI Glasses 2

Tasks:

(1) Free-viewing for unlimited time (Map P1)
(2) In what coordinate system was the map created? (Map P2)
(3) What two geological units are located on the territory of the map sheet? (Map P2)
(4) Which rock predominates in the smaller unit? (Map P2)
(5) Find the area on the map with a frequent occurrence of landslides. (Map P2)
(6) Which rock predominates in the area with the highest amplitude of geomagnetic anomalies? (Map P2)

Stimuli:

- Map P1 - Paper geological map "ZM_25_03-342 Turnov"
- Map P2 - Paper geological map "ZM_25_03-413 Semily"

<u>Paper ET2Spatial</u>

Sultan, M. N., **Popelka, S.***, & Strobl, J. (2022). ET2Spatial–software for georeferencing of eye movement data. Earth Science Informatics, 1-19.
https://doi.org/https://doi.org/10.1007/s12145-022-00832-5

[IF 2.8; Q3]
[0 citations on WoS; 0 citations without autocitations]
[Author's contribution: 40%]

Abstract

The paper focuses on the development of an open-source utility tool for the analysis of eye-tracking data recorded on interactive web maps. The tool simplifies the labor-intensive task of frame-by-frame analysis of screen recordings with overlaid eye-tracking data in the current eye-tracking systems. The tool's main functionality is to convert the screen coordinates of the participant's gaze to real-world coordinates and allow exports in commonly used spatial data formats. The paper explores the existing state-of-art in an eye-tracking analysis of dynamic cartographic products as well as the research and technology aiming at improving the analysis techniques. The developed software, called ET2Spatial, is tested in-depth in terms of performance and accuracy. The capabilities of GIS software for visualizing and analyzing recorded eye-tracking data are investigated. The tool aims to enhance the research capabilities in the field of eye-tracking in geovisualization.

Earth Science Informatics (2022) 15:2031–2049
https://doi.org/10.1007/s12145-022-00832-5

SOFTWARE ARTICLE

ET2Spatial – software for georeferencing of eye movement data

Minha Noor Sultan[1] · Stanislav Popelka[1] · Josef Strobl[2]

Received: 23 February 2022 / Accepted: 13 June 2022 / Published online: 24 June 2022

Abstract
The paper focuses on the development of an open-source utility tool for the analysis of eye-tracking data recorded on interactive web maps. The tool simplifies the labor-intensive task of frame-by-frame analysis of screen recordings with overlaid eye-tracking data in the current eye-tracking systems. The tool's main functionality is to convert the screen coordinates of the participant's gaze to real-world coordinates and allow exports in commonly used spatial data formats. The paper explores the existing state-of-art in an eye-tracking analysis of dynamic cartographic products as well as the research and technology aiming at improving the analysis techniques. The developed software, called ET2Spatial, is tested in-depth in terms of performance and accuracy. The capabilities of GIS software for visualizing and analyzing recorded eye-tracking data are investigated. The tool aims to enhance the research capabilities in the field of eye-tracking in geovisualization.

Keywords Utility · Eye-tracking · Georeferencing · Interactivity · User-logging · GIS

Introduction

The number of users and researchers who use eye-tracking (ET) systems is growing tremendously. The current technological development in eye-tracking software and hardware is on a trajectory where no advanced technical skills are needed to adopt and apply these systems. As such, the user base is expected to grow even further (Holmqvist et al. 2011). This, combined with the frequency of data generation, requires efficient and faster ways for data analysis.

Cartography has been no alien to employing eye-tracking in the usability study of maps. One of the earliest examples dating back to the evaluation of simple drawn maps and aerial images is a study by Enoch (1959). Cartographers started with eye-tracking research in the early 1970s, when Jenks incorporated eye movements in his user studies (Jenks 1973, 1974).

The maps as visual stimuli have changed since the beginning of cartographic eye movement research, from static to dynamic displays. Słomska (2018) investigated the types of stimuli used in cartographic empirical research. According to the analysis of more than one hundred papers, she found that more than one-third of used stimuli are interactive. Moreover, her research showed a steady increase in the number of papers published in the field of cognitive cartography. Similar results were provided to the field of eye-tracking in cartography by Krassanakis and Cybulski (2019). In 2019, Unrau and Kray (2019) provided a comprehensive literature review of usability studies in the field of geographic information systems. This review was exploring 39 studies. The most frequently used analysis methods were task scores, subjective rankings, and user comments. Eye-tracking was included in 15.4% of the studies. The authors concluded the review by defining key challenges and opportunities for future research in this domain. They proposed the methods for evaluating GIS to be capable of capturing the observations during complex tasks and combining map interaction with a traditional user interface. Eye-movement analysis was one of the recommended methods.

The cartographic and visual analysis community has shown long-standing efforts to improve the possibilities of

Communicated by: H. Babaie

Stanislav Popelka
stanislav.popelka@upol.cz

Minha Noor Sultan
minha.noor16@gmail.com

Josef Strobl
josef.strobl@plus.ac.at

1 Department of Geoinformatics, Palacký University in Olomouc, 17. listopadu 50, 779 00 Olomouc, Czech Republic

2 Department of Geoinformatics, Paris Lodron University of Salzburg, Schillerstraße 30, 5020 Salzburg, Austria

eye-tracking data analysis. Researchers delivered several tools, methods, and algorithms. Burch et al. (2018) developed a tool for sequence alignment analysis called EyeMSA. The tool allows finding common subsequences in scanpaths. Scanpaths of multiple participants can be compared using ScanGraph (Dolezalova and Popelka 2016; Popelka et al. 2018). This tool has already been used not only in the field of cartography but also in geographic education (Beitlova et al. 2020; Havelková and Gołębiowska 2020; Popelka and Beitlova 2022), physics education (Skrabankova et al. 2020) or for the investigation of the eye movements of participants with an autism spectrum disorder (Eraslan et al. 2019). Krassanakis et al. (2014) developed a MATLAB analysis toolbox, EyeMMV, which fully supports the analysis of eye-tracking data. As a result of cooperation between geographers and psychologists, the testing platform Hypothesis (Šašinka et al. 2017) was created and used for user data recording (Šašinka et al. 2021).

Eye-tracking and analysis of eye-movement data on interactive stimuli are costly in terms of time and data storage since the standard mechanisms produce video recordings with overlaid gaze points (Pfeiffer 2012). Another drawback of traditional eye-tracking mechanisms for interactive media is the effort required to create areas of interest (AOI) during analysis (Dong et al. 2020). As opposed to static maps, where the researcher can create polygons on stimuli for easier statistical evaluation. Dynamic AOIs are a possible approach suggested by Holmqvist et al. (2011), where the AOI can be established for non-static stimuli such as screen recording of a moving map. Manual creation of dynamic areas of interest is labor-intensive. Popelka et al. (2022) mention that drawing of such AOIs in the study comparing multiple views and swipe map comparison environments took approximately ten times longer than the length of the screen recording. A tool developed by Papenmeier and Huff (2010) facilitates drawing these dynamic AOIs. However, the rapidly changing content of the interactive web media where users perform pan, zoom and identify clicks makes it rather impractical to manually annotate the areas of interest and makes it very labor-intensive because the video for each participant is different.

Although research in the intersectional domain of eye-tracking and interactive web maps is nowhere near saturated, it is not non-existent. Efforts have been made to understand cognitive processes involving interactive dynamic screen maps, but very little knowledge has been gathered until the last decade. Burch (2019) discuss the challenges for visual analysis of the recorded data. Moreover, he described a data model which leads to interactive graphs – one of the possible ways of for analysis and visualization of interactive eye-tracking data. A study carried out by Ooms et al. (2012) deployed a technique for understanding user behavior with interactive maps. This technique involved a standard eye-tracking apparatus combined with a joystick for tactile input. The dynamic map, however, was pre-recorded with its pan actions. The recordings were used as stimuli in the eye-tracking experiment, and users logged a signal as soon as they identified the object. The study combined recordings from both input mediums to measure response times of different user groups and conclude results about interaction. The experiments pertain to larger research aiming at insight into users' cognitive processes while reading interactive maps. Content-dependent analysis, contrary to content-independent analysis for eye-tracking in cartography, needs sophisticated methods for evaluation (Ooms et al. 2012). The study was pioneering for evaluating a dynamic medium; however, a dedicated framework to assess the content was not used.

Gaze coordinates that can be transformed into geo-coordinates can provide more information and feasible solution to the existing issues with interactive web maps (Giannopoulos et al. 2012; Ooms et al. 2015). Gaze Map Matching was introduced, taking into account the aforementioned issue of content-dependent analysis. The gaze metrics such as gaze sequences and gaze fixation points are studied with respect to the underlying vector data to inspect the geographic features as subjects (Kiefer and Giannopoulos 2012).

When it comes to non-conventional usage of eye-tracking data in the field of cartography and GIS, a handful of studies have been carried out so far. Considering eye-tracking data as spatial features allows dedicated spatial functions to be applied to the eye-tracking metrics such as scan paths, fixation, and raw gaze points. The traditional methods and tools offered by eye-tracking software, however, lack the capability of studying the gaze data metrics from a spatio-temporal perspective. The dataset has a very similar structure to movement datasets of real-world features in geographic spaces. Hence techniques designed for the evaluation of spatial movement data can also be resourceful in the evaluation of gaze points data (Andrienko et al. 2012).

While these attempts at understanding the content of the interactive mediums have been less, the research on open-source solutions for eye-tracking mechanisms on dynamic maps has been far fewer. One such attempt at an evaluation mechanism of an interactive medium was the exploration of 3D models. 3DgazeR helps in a less cumbersome way with the analysis of interactive 3D models. The tool works by calculating 3D coordinates (x,y,z) for each point of view in a 3D scene. These coordinates are derived from the orientation and location of virtual cameras as well as the screen coordinates of eye movements. The output generates gaze points referenced to the 3D model, which can be visualized in QGIS (Herman et al. 2017). The tool addresses the problem of eye-tracking in an interactive medium but is constrained to a 3D environment.

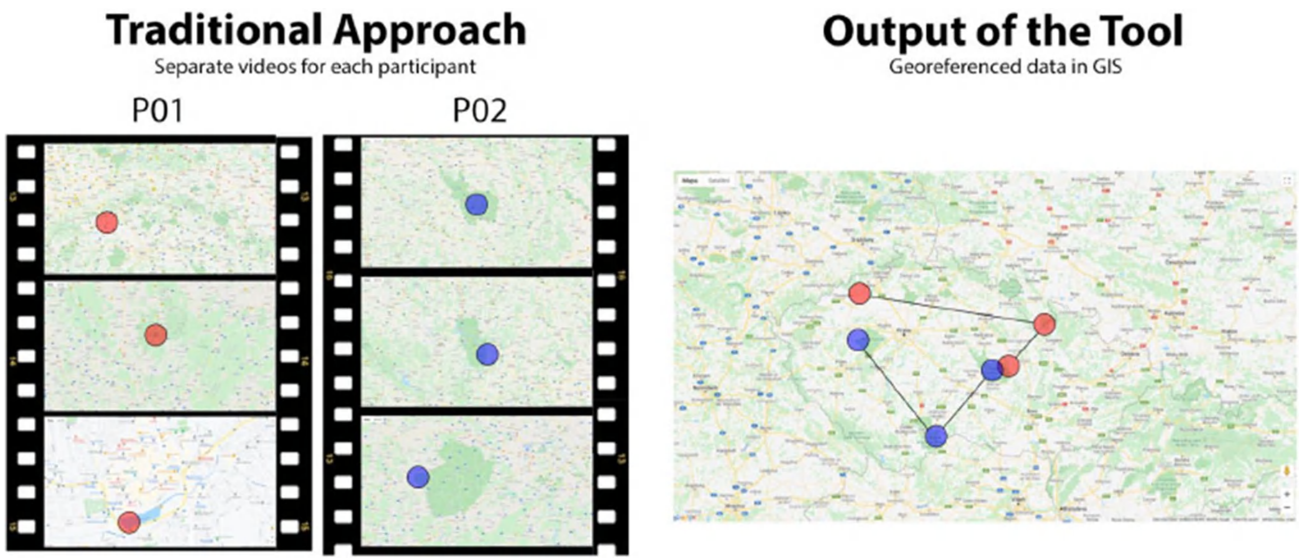


Fig. 1 Main use-case scenario

A framework suggested by Ooms et al. (2015) captures the essence of the problem in the best way. The study aims to build an application that is compliant with interactive map mediums and logs user data and raw gaze samples. Several approaches are tested in the study to determine which ones fare best and can be used independently with a variety of software. The study explores both desktop-based and online-based user data logging solutions and settles on the desktop-based user-logging approach using PyHook as the preferable one because of its ability to work with a wide array of applications. This approach logged all user mouse clicks and key presses on web pages which were eventually synchronized with eye-tracking data through an imposed mouse-click command. The conversion of the ET screen coordinates to geographic coordinates is done by calculating the map extent, distances, and direction of the user-interaction movement. The principle that the scale remains constant during pan operation helps in estimating the new center coordinates of the viewing window in reference to the previous one. The mathematical conversion itself is done through forward and inverse map projection. The use-cases focus on applying the methodology in other domains such as marketing, psychology, and traffic science.

FeatureEyeTrack, a tool developed at ETH Zurich, measures the real-world coordinates from a user's screen coordinates for interactive web maps. It is the approach closest to the tool described in this paper and shares the same goal of easing analysis of eye-tracking data for dynamic online mediums. The framework involves an eye-tracker, a logger, and a web map service. The logger records all the mouse clicks, user inputs, the extent of the map, and the zoom level are fetched and stored in an SQLite database through a web page featuring an interactive web map. The tool uses Mapbox API, and the main program written in Java which receives the gaze data stream, which is then combined with the user logged data, converted, and then stored in the database (Göbel et al. 2019).

Objective

The aim of this paper is to document the development of a tool that will allow recording eye-movement data observed during interaction with a web map and store them as georeferenced data. The main functionality of the tool will be to convert user's gaze positions on a map to real-world coordinates. One of the main driving forces behind this study is the inability of the traditional ET software to overlay multiple participant data for interactive media on one file. Figure 1 shows the main problem the ET2Spatial tool addresses.

The tool aims at allowing easier analysis of eye-tracking data in the realm of interactive online maps. Through individual spatial data generated by the tool, the visualization and comparison of multiple users' data simultaneously on a basemap will become feasible. In summary, the project aims to simplify the task of eye-tracking evaluation techniques on web maps by creating an application that takes factors like zoom level and pan operations on the map into account.

The tool will be evaluated in terms of functionality as well as the quality of output. The usability of the tool will be tested through the example of multiple visualizations. The study will also tackle a non-conventional approach of inspecting eye-movement data as spatial features, subjecting them to traditional spatial operations and comparing them to standard techniques provided by eye-tracking software

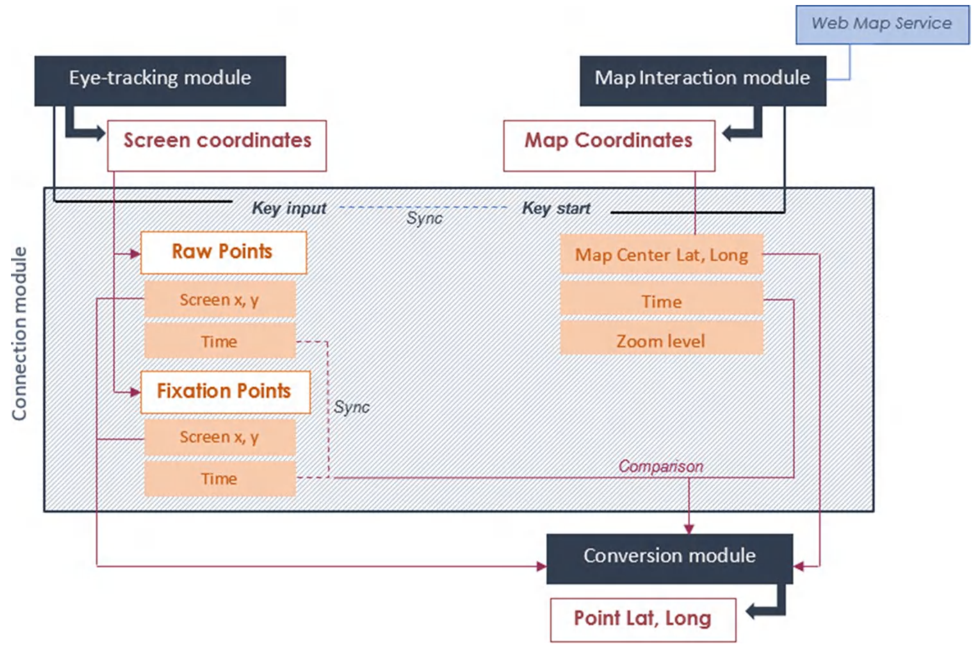


Fig. 2 Schema of ET2Spatial tool

and gauging the applicability of spatial functions on the eye-tracking data. The final output will be an open-source desktop application with a simple GUI.[1]

Design and implementation

The architecture of the solution comprises four main modules; the eye-tracking module, map interaction module, connection module, and conversion module. Figure 2 provides an illustrative view of the solution.

The eye-tracking module consists of an eye-tracking device connected to a 1920 × 1200 px screen. This module is responsible for sensing, recording, and exporting the gaze locations of the user. The map interaction module consists of a framework for extracting the user's interaction with the web map. In this study, MapTrack (Růžička 2012) is used. However, an application-independent approach is also discussed. The map interaction module outputs the active map coordinates as latitude and longitude. The connection module, which is one of the main components of the tool, is responsible for data synchronization between the ET data and the user interaction data. After the connection module, data gets passed on to the conversion module. The conversion module accounts for the calculation of geographic coordinates from the screen coordinates.

From a user perspective, a typical workflow activity would involve conducting ET experiments on a web map (Google Maps) displayed through the MapTrack application on the screen of an ET setup. The data from both, ET device and the MapTrack application would be fed to the ET2Spatial tool. The tool then converts the points in the datasets into spatial features that can be imported into any GIS software and overlaid for multiple participants on a basemap (Fig. 3).

User interaction data

User interaction data implies the logged interaction of a user with a web application or a website. These interactions can be registered by deploying custom JavaScript code on a proxy server to detect different mouse and keyboard events on the client-side. Several dedicated applications and tools are also available for these tasks. However, one downside of using this approach is that some web map applications block certain user input events from being registered, for instance, a mouse down event (Ooms et al. 2015).

On the other hand, Web Map APIs can potentially overcome this as they offer built-in functions for registering user map events. For the scope of this study, the focus is on the content of the web map itself and not the layout of the webpage as a whole. The primary data needed from a user interaction framework, recorded during an ET experiment, was the geographic coordinates of the current center of the web map with an associated timestamp, the current zoom level, and a map pan event registration.

MapTrack (Růžička 2012) was chosen as the suitable application for this study primarily because of its ability to record the aforementioned web map events on the client side. In addition, MapTrack being open-source

1 https://github.com/minha94/ET2Spatial

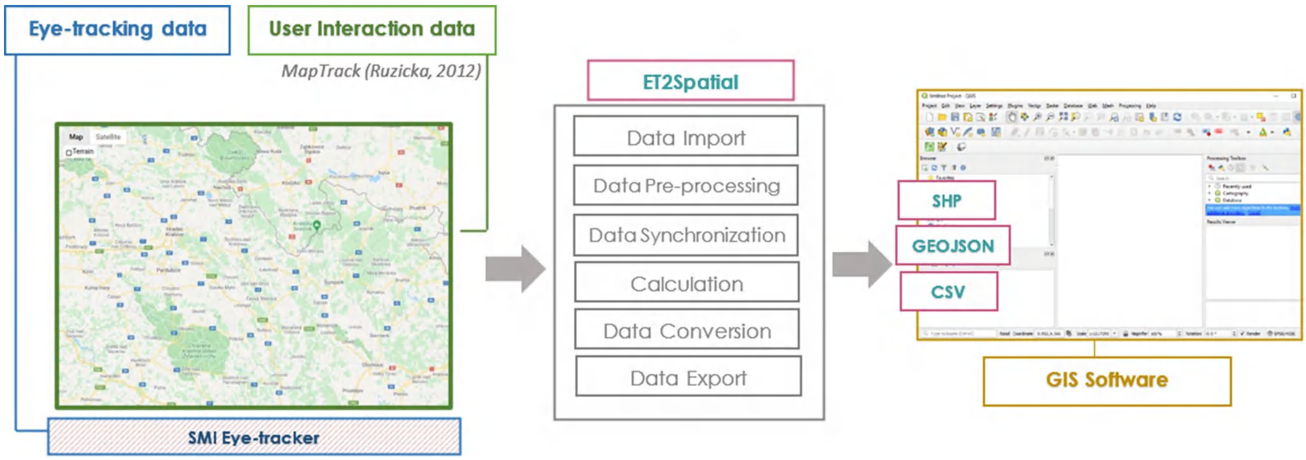


Fig. 3 User workflow diagram

allowed access to the source code and hence the possibility of small additions or tweaks if needed to create a harmonized integration with the developed tool.

MapTrack is an online application[2] that creates a registry of user activity with the web map. MapTrack is configured to work with Google Maps API v2, which allows adding various types of overlays.

An initial experiment was done to collect sample data to work with. The MapTrack application was displayed on the screen connected to the ET device and used as a (screen-recording) stimulus. Every time the user scrolls or pans, the zoom level and map center are changed and registered. For the sample experiment, it was stored into an xml file that was available to download. The experiment was done in full-screen mode, which removed the need to filter out ET points outside the web map.

For collecting the sample data, SMI RED 250 eye-tracker with a sampling frequency of 250 Hz was involved. The data was exported through SMI BeGaze software. For this study, raw gaze points were used as well as identified fixations. The raw points data stream exported by the eye tracker comprises of triples in the form of screen x, y & t. These triples are typically aggregated spatio-temporally in reference to the fixations. A fixation is registered when the eye rests on a screen location for a certain amount of time compared to saccades which denote the quick eye movements between fixations. The number of properties, as columns, for both raw points and fixations was kept to basic minimum i.e. additional details such as age, gender etc. were omitted during export.

[2] https://github.com/ondrejruzicka/maptrack

Raw gaze points

During the export of a raw gaze points file from the SMI BeGaze software, the timestamp, recording time in milliseconds, point of regard x and y, and participant information were deemed necessary to work with. The file was exported from the software as a text file and read through a CSV python library within the script. The file was also converted to a pandas data frame that allowed for easier visualization of data sheets and offered efficient methods to manipulate rows and columns in the data. The pre-processing steps in the script for the raw data points involved:

- Data slicing vertically; this implies column selection. Although the user can select only required columns during ET data exports from an ET software, specific columns mentioned above are extracted explicitly through a script to reduce active data volume in case of a large number of columns in the original export.
- Data slicing horizontally; this implies row selection. The row selection is an important step which plays a part in further data synchronization. ET systems have the ability to record key inputs in addition to eye movement data. One feature of MapTrack initiation and termination mechanism is the 'F2' key press which also gets registered by the eye-tracker in the file output. This input is used to crop the raw points dataset such that only the points recorded during the usage of the web map are considered.
- Cleaning: the text file contained special characters for missing values. The rows with these characters for any one of the columns are removed to avoid errors in later calculations.
- Indexing: many functions in organized data structures such as pandas df rely on proper indexing. The index

for the pre-processed raw data is rebuilt after subjection to slicing and deletion operations.

- Time column formatting: After the raw points data is synchronized with the user interaction data through key input, the recording time need adjustment as well. To achieve that, time length *length1* is calculated by subtracting the initial time stamp in the unsliced data from the timestamp at index 0 in the newly sliced dataset. This length is then subtracted from each entry in the *RT* column to create a formatted time column *Format_RT*.

Fixation points

The raw gaze points file is sufficient on its own for an independent export, but the fixations file depends on the time calculations from the raw points file for synchronization with the user interaction data. To prepare the fixations points data following steps were implemented in the code:

- Data slicing vertically: Similar to the procedure in raw points pre-processing, the columns from the original text file are selected to only the necessary ones for processing i.e., Fixation start time, Duration, Fixation position x, y, and Participant number.
- Renaming, Cleaning, Indexing: Similar to the raw points data processing. Time column formatting: The fixation start time in the original fixations file are the relevant time stamps needed for calculations; however, these times are not synchronized to the time window for the actual user web interaction. Only the fixations that occurred during the participant's usage of the web map are needed. It meant that the fixations occurring during the prompt slides and instructions reading had to be excluded. As mentioned, the fixations file exported from SMI BeGaze does not include a record of the user's key inputs. Hence the method followed for the raw points file could not be applied here. The time length calculated for the raw points was also utilized here by subtracting it from the fixation start times. Consequently, the negative values from the results were filtered out as it implied those fixations happened before the user initiated MapTrack. These results were stored in a new column *Sync_time*. The data was now synchronized at the same starting time as the user interaction. To remove the fixations that happed after the MapTrack application was closed, another length *length2* was calculated. This value was computed by subtracting the first and last timestamps of raw gaze points which yielded the total gaze time on a web map. All the values in *Sync_time* that were greater than length2 were filtered out.

Data synchronization (stitching)

The concept of synchronization is important for this study because it contributes to the temporal and spatial accuracy of the output of the tool. It implies that all the input datasets have the same starting and ending time. As mentioned earlier, an imposed tactile user input from the web map recorded by the ET can significantly help. In the raw data file, F2 was used to synchronize it with MapTrack data. The fixations data was synchronized to the MapTrack data through variables in the raw data file. Once these three datasets were pre-processed and synced, the next step was to combine map interaction properties such as map center Lat Long and zoom level to raw data and fixations data individually. This step would stitch the datasets based on timestamps necessary for later coordinate calculation.

Data stitching required iterating over each row in both the ET raw dataframe(df) and the MT df. Pandas provide efficient data handling functions especially accessing df values through multiple options. However, the iteration over rows on both tables through loops proved to be complicated in a pandas df. The iterrows() function gives a series in return for every row it is iterated over and which is susceptible to data type changes. In addition, iteration over df is tricky, especially if the values are being modified because the iterrows() returns a copy instead of a view (pandas documentation). Hence, the data frames were converted to dictionary data structures to ease the process of looping and modifying values.

The dictionary was oriented by 'records' which meant that every item would have a column name and its associated value for the specific row. It is a list-like structure *[{column->value}]*.

Figure 4 shows the logic for the stitching operation. Every row in the time column of the fixations data table is compared to the corresponding row in the time column of the MapTrack data table. The *sync_time* values are compared to the MapTrack_RT. If the time in row *i* of fixations data is greater than or equal to the time in row *i* of user interaction data but less than the time of the succeeding row in the user interaction data table, the map center coordinates and zoom level from the MT table get appended to the fixations table for that row.

Data conversion

Once the raw points data table and the fixation points data table were populated with a zoom level and map

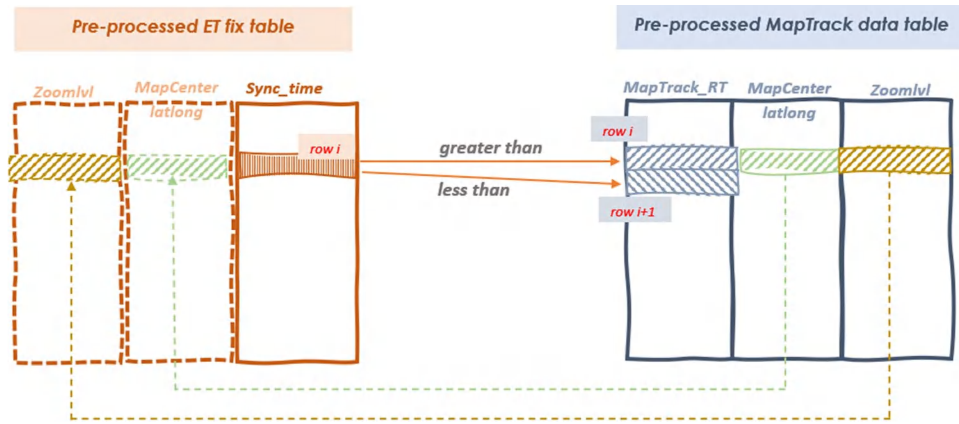


Fig. 4 Data synchronization schema

center coordinates for every time stamp, the next step was to calculate the real-world point coordinates. Several approaches were researched and tested for the said calculations, out of which only the ones with higher accuracy and thus relevance have been discussed here.

The first approach revolved around finding the screen distance between the gaze point and the screen center. The screen center was calculated through screen size stored during the MT data import. The screen distance was then multiplied by distance per pixel to get the distance in meters. Since the latitude of the ET gaze point was the subject in question and unknown, the known map center latitude was substituted in the formula instead. After calculating distance in meters, the output was converted to decimal degrees and added to the map center latitude and longitude. The resulting coordinates obtained through this method were off by one degree in both directions. In practice, especially at smaller zoom levels, this would lead to inaccuracies in visualizations and analyses.

The second approach used map extent coordinates and screen coordinates of points as inputs. The idea behind this approach was the transformation of the points from one coordinate system to another coordinate system, i.e. from the screen coordinate system to the geographical coordinate system. The formula remaps values from one range to another. The maximum and minimum values of both the screen and the map were used in addition to the screen point coordinate. The output was then latitude for the y screen coordinate and longitude for x screen coordinate as input. The y coordinate had to be adjusted since the origin of a traditional screen coordinate system is on the top left, and hence the y values progress in a direction opposite to a geographic coordinate or Cartesian coordinate system. The accuracy of this approach was 5–6 s which was very good for the scope of this study.

The third approach was the one taken forward to implement in the tool for this study. The inputs were the same as the first approach; zoom level, map center coordinates, and screen coordinates. These inputs were used in the Web Mercator projection formula shown below. Web Mercator is a variation of spherical projection, which is the de facto standard used by web mapping platforms such as Google Maps, OpenStreetMaps, and Mapbox. It is slightly different from the Mercator projection because it employs spherical formula at all scales, unlike the Mercator maps with the ellipsoidal version of projection at a large scale (Eq. 1, source USNA, 2012[3]).

Mercator Projection formulae

Spherical Forward – earth to map

$$x = R(\lambda - \lambda_0)$$
$$y = R ln\left(\tan\left(\frac{\pi}{4} + \frac{\Phi}{2}\right)\right)$$

Spherical Inverse – map to earth

$$\Phi = \frac{\pi}{2} - 2\arctan\left(e^{-\frac{y}{R}}\right)$$
$$\lambda = \frac{x}{R} + \lambda_0$$

Ellipsoid Forward – earth to map

$$x := a * (\lambda - \lambda_0)$$
$$y := \frac{1}{2} * a * \ln\left(\frac{1+\sin\Phi}{1-\sin\Phi} * \left(\frac{1-e*\sin\Phi}{1+e*\sin\Phi}\right)^{e}\right)$$

Ellipsoid Inverse – map to earth

$$\Phi = \frac{\pi}{2} - 2 * \arctan\left(e^{\frac{-y}{a}} * \left(\frac{1-e*\sin\Phi}{1+e*\sin\Phi}\right)^{\frac{1}{2}e}\right)$$
$$\lambda = \lambda_0 + \frac{x}{a} \qquad (1)$$

[3] https://www.usna.edu/Users/oceano/pguth/md_help/html/mapb0iem.htm

The Web Mercator variation adjusts the world coordinates before applying the zoom. The origin of the coordinate system is on the top left, same as that of a display screen, and hence take tiles and pixels into account (Eq. 2, Wiki, 2021[4]). OpenStreetMap provides comprehensive documentation online on this subject, such as the technique behind the tiling of slippy maps.[5]

Web Mercator Projection Formula

$$x = \left\lfloor \frac{256}{2\pi} * 2^{zoom\ level}(\lambda + \pi) \right\rfloor pixels$$
$$y = \left\lfloor \frac{256}{2\pi} * 2^{zoom\ level}\left(\pi - \ln\left[\tan\left(\frac{\pi}{4} + \frac{\varphi}{2}\right)\right]\right) \right\rfloor pixels \quad (2)$$

The map center coordinates, latitude and longitude are first converted to projected coordinates x and y through a forward spherical Mercator projection but with the Web Mercator variation. The latitude (φ) and the longitude (λ) are converted to radians and multiplied to the constant terms. Since the zoom level is the same for both the latitude and longitude, it was grouped together with the constants and calculated separately. $\frac{256}{2\pi} * 2^{zoomlevel}$ refers to the constants expressing unit distance. After calculating map center x and y through the coordinates and forward projection formula, the distances in x and y direction are calculated as *xDist* and *yDist* columns through the screen distances from the ET screen coordinates of the points. This yields the x and y locations of the points in the projected coordinate system. The final latitude and longitude of these points are computed using the inverse Mercator projection formula as shown above.

Data exports

Once the script was functional for data conversion i.e., the points were converted from screen coordinates to geographic coordinates; the next step was file conversion and export. The main file formats considered for exports in the ET2Spatial tool were Geojson and Shapefile. The reason for selecting these file formats was their widespread usage and popularity amongst the GIS community and their compatibility with most GIS software. Geojson can also be easily visualized with mapping libraries and APIs such as leaflet, Mapbox, and Google Maps.

The script generated a separate file for every participant identified with the participant id that was stored at the beginning of the script. The converted raw gaze points file and the converted fixations points file were provided as an output for each participant. The shapefile function generated a CSV file as a byproduct for an alternative pathway into GIS software.

[4] https://en.wikipedia.org/wiki/Web_Mercator_projection

[5] https://wiki.openstreetmap.org/wiki/Slippy_map_tilenames

When it comes to the structure of the files being exported, only the necessary columns were included, namely; Latitude, Longitude, Zoom level, Time, and Id; with Latitude and Longitude in decimal degrees and Time in milliseconds. The fixations file has an additional duration column, which gives the amount of time the user fixated gaze on that point in milliseconds. The Id is the explicit id from the data frame assigned at the time of data pre-processing and creation of formatted time columns and is in the temporal order of the points. It can be used to re-order points in the output easily.

Graphical user interface

After creating and testing the functionality of all the relevant functions in the python script, the next step was to create a user interface for the script.

Layout design

The conceptual layout of the user interface was kept very simple with a button for each; main imports, function, and exports. The size of the window was kept small because there were no visual aspects to the processing. In addition, basic level function completion notifications were aimed for during the design process. The main components of the layout of ET2Spatial were organized through QT Designer. In addition to the buttons and title of the window, a progress bar was added.

Connecting functions

The python script tested on sample data so far was structured procedurally with line-by-line execution of the program. This code, however, needed to be restructured into functions in a modular manner to be able to connect to buttons and called with single commands. Three import functions were set up for the raw data points, fixation data points, and user interaction data. Figure 5 outlines the function behind each component of the ET2Spatial GUI.

Finally, one button called *Reset* was added to provide convenience to the user. In the current version of the tool, there is no capability for multiple imports and conversion simultaneously. In many cases, including this study, there is more than one participant for which the data needs to be converted. Restarting the application again and again would be very ineffective, and although the user could upload data for the new participant and start over, the configuration of the user interface could be confusing. To simplify this problem, a reset button was added. The only purpose was to delete the existing main variables from memory and reset the labels next to the upload buttons. This would signify a restart of the tool and the progress.

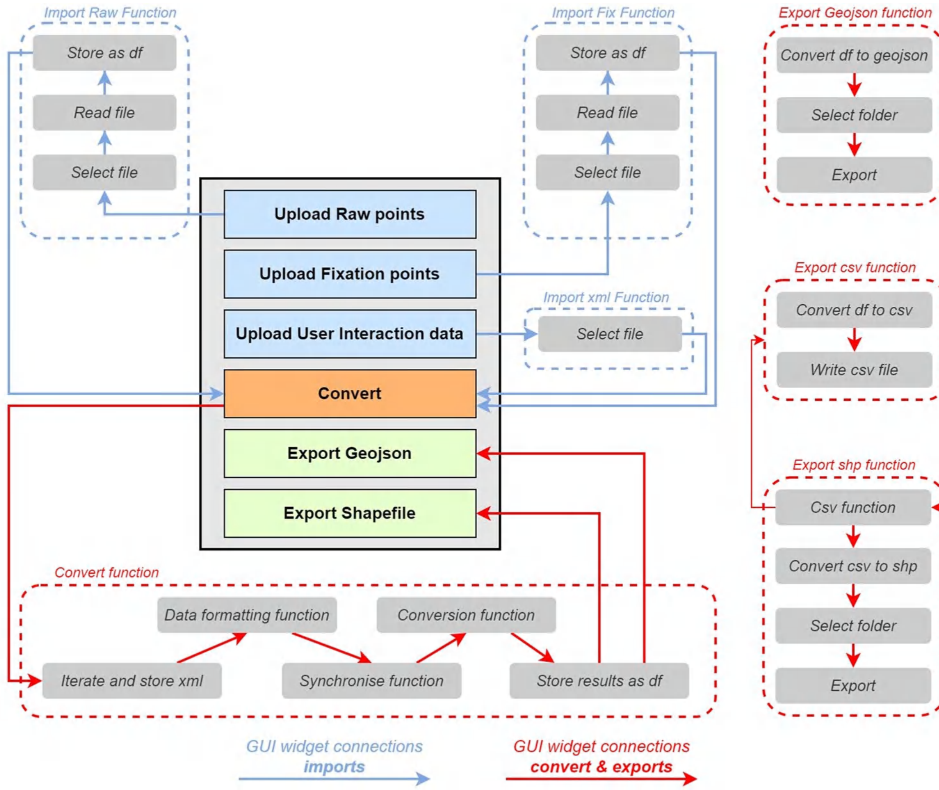


Fig. 5 GUI widget connections

Since one of the objectives of this study was to create a standalone desktop application, an exe file was generated from the python file using PyInstaller module. PyInstaller gives the option to create a folder with all the necessary modules, libraries, and dependencies along with the exe or create a single standalone executable file packaged with all the dependencies implicitly.

One disadvantage of generating a single exe file was the comparatively large size of the executable and the time needed for its initialization. To give notification on the initialization progress, a command line interface was added. This command-line window pops up as soon as the executable file is clicked. It generates notifications on the compilation of the tool and after completion, opens the GUI window of ET2Spatial. Figure 6 shows the final interface window of ET2Spatial.

Tool evaluation

Quality control is a part of the lifecycle in product development of any kind, particularly in software development. For the evaluation of ET2Spatial tool, a similar concept was adopted. The evaluation benchmarks were categorized into output accuracy, tool performance, and visualization of recorded data in GIS environment. The evaluation was an important part of the study as it gave insights into how effective and useful the developed tool was. It also allowed analyzing the strengths and weaknesses of the study in detail.

Output accuracy

Output accuracy refers to the positional accuracy of the exported spatial data from the tool i.e., whether the geographical coordinates of the eye-tracking points were at the exact locations they were supposed to be. The evaluation mechanism was based on the concept of reverse conversion whereby the known geographic coordinates of a point were fed into an algorithm to output screen coordinates as well as on the visual comparison of the point positions on a single frame in the video recording exported through the eye-tracking software.

Tool performance

The second part of the tool assessment was the functional quality i.e., how smooth the tool functions when executed on different systems. While packaging the tool in a single executable file, all the dependencies were included. These packaged python dependencies would allow users to run the tool without installing any modules or language by themselves

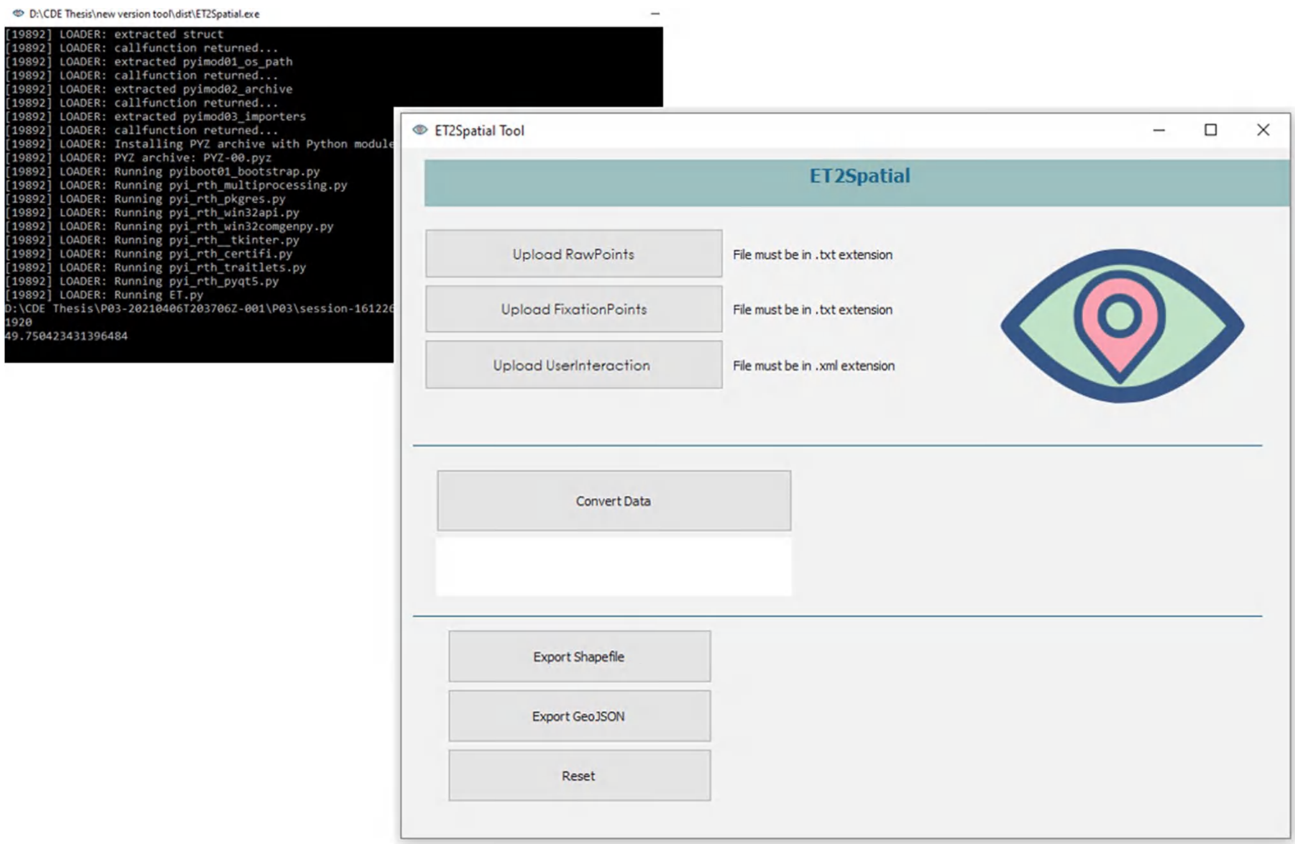


Fig. 6 ET2Spatial GUI and initialization

and would also prevent any system environment clashes. To test this however, the tool was run on four different laptops with a Windows operating system. No noticeable problem was observed during these trials.

One small weakness of the tool was its compilation or initialization time (~30 s). However, a command-line interface was added to portray the initialization process before the final graphical user interface shows up. This was added to lower the inconvenience during usage. To manage unexpected scenarios, a README file was prepared to provide information on necessary columns during export from the eye-tracking software.

Visualization of the results

The final evaluation criteria for the study was the proof of concept. Recorded data were visualized in GIS environment. Different techniques can be applied for displaying the eye-tracking data through manipulation of symbology, labeling, and custom rulesets. Eye movement data can be entered into various types of spatial analyses like any other spatial data.

Concerning the visualizations, precision is only possible if the correct basemaps are loaded that were used at the time of experimentation. Particularly for the usability analyses, the tool can only be meaningful if the exact cartographic renderer is accessed through the GIS software. For map tasks demonstrated in the upcoming sections, mainly Google Maps Roadmap (Czech place labels) and Satellite imagery were used. The ET spatial data was imported into GIS software and projected in the WGS 84 Geographic Coordinate System.

The study environment for collecting the data consisted of an eye-tracker i.e., SMI RED 250 for recording gaze points, a web application MapTrack for the web map interactivity data, SMI Experiment Center for setting up the map tasks and instructions, and SMI BeGaze for exporting the ET data. For post-processing of the results, ArcGIS Pro and QGIS were used. There were eight participants in the experiment and each one was asked to solve a task of identifying the city of Salzburg on a basemap.

Multiple participant data overlay

The first visualization tested was the actualization of the main problem scenario i.e. to overlay eye-tracking data of multiple participants on the same base layer. Figure 7

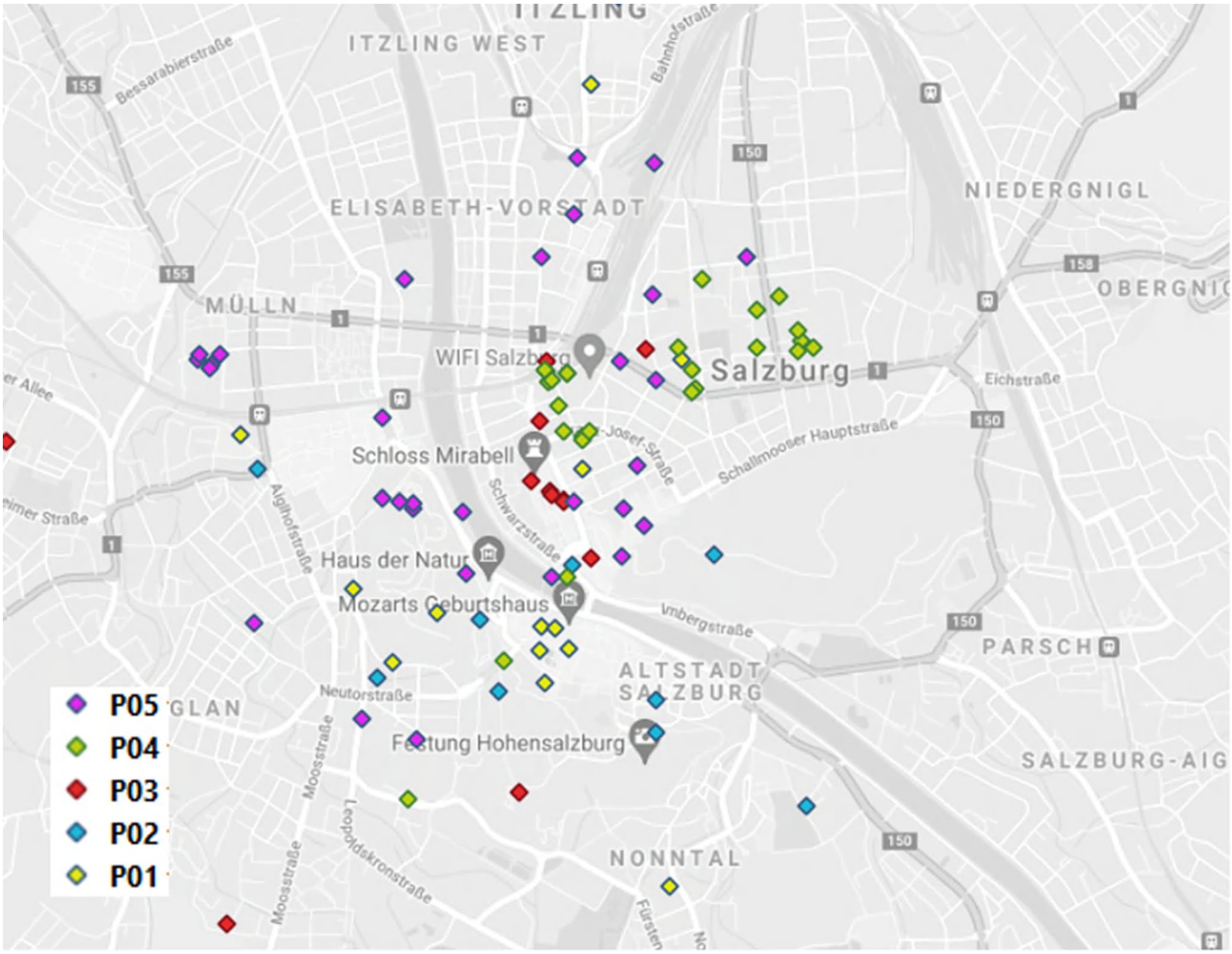


Fig. 7 Multiple participants visualization

shows the fixation points of multiple participants displayed in ArcGIS.

Unlike the traditional eye-tracking software, the symbology for each of the participants can be manipulated as desired. The basemap itself has a few optimization options, such as color toning and transparency. In this case, the basemap was changed to grayscale to better distinguish the point symbology from the map features. A convenient way to have a structured symbology for all the points is by merging the point datasets for all the participants into one layer.

Scale-based rendering

The eye-tracking points were recorded with different variables during the user's interaction with the map i.e., every point was recorded at a certain zoom level and at a certain timestamp. Different zoom levels imply that the underlying content of the map can be different for each value. An accurate position of the ET point is not meaningful unless displayed at the correct scale with the original basemap to analyze the content user was looking at during the experiment. As done with the multiple participant visualization, it is possible to symbolize the ET data based on zoom level and to generate a category for each with different colors and shapes. Labeling can also help in this regard. Figure 8 shows the fixation points of a single user labeled by zoom levels and categorized as well in terms of colors.

This approach, although aesthetic, can be quite cumbersome while doing analysis or visual inspection. The reason being that the analyst would need to manually zoom to the right scale for every point to see what content lies on the cartographic renderer, and secondly, most GIS software such as ArcGIS or QGIS do not have a 'zoom level' displayed in the windows rather a map scale. This map scale is usually formalized in easy terms as a zoom level for online web

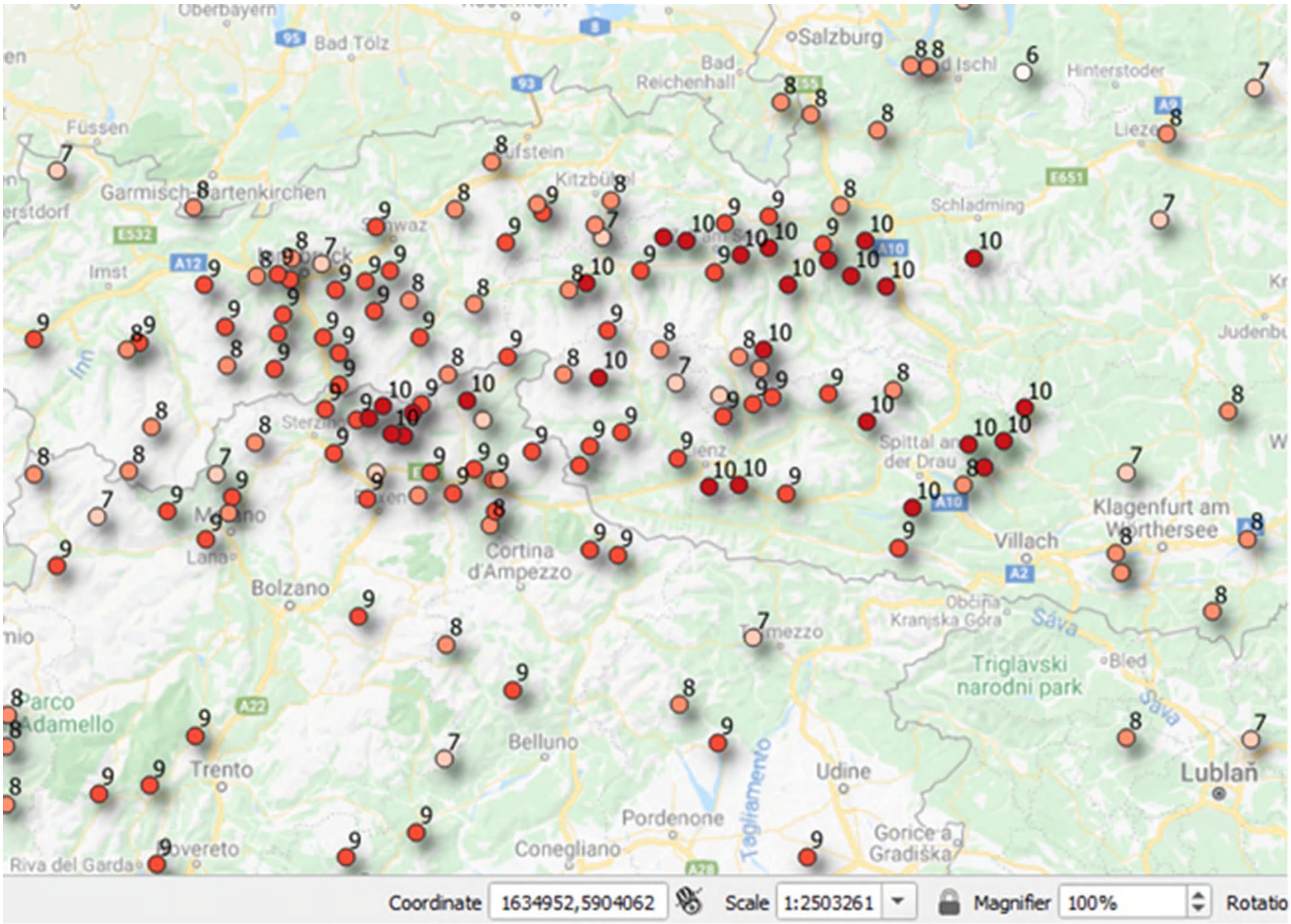


Fig. 8 Categorization by zoom level

maps and can be used in scale-based rendering. Scale-based rendering means that the symbology or labels on the map appear at different scales. This option is usually a part of the most sophisticated GIS software.

Figure 9 shows the example of scale-based rendering which was set up as one of the visualization methods for this study. In QGIS, the points were symbolized using the rule-based symbology, which categorizes points into different groups, in this case, the zoom levels, and then a minimum and maximum scale for each category can be defined. Resulting points in one category, such as zoom level 8 appear only when the user scrolls to a scale of 1:2,311,162 and beyond. The information on the scale value range associated with every zoom level was fetched through documentation online.

Scale-based rendering is a very convenient feature of the GIS software and is even more apt for this study, where the points are dependent on different zoom levels. It rids the analyst of labor-intensive manual scrolling for each point and for each zoom level during analysis and automates the process by only displaying relevant points at every scale.

Attribute based visualization

The symbols can be varied based on their attributes. Similar to the examples shown before where the point symbols were categorized as unique symbols with varying colors, the graduated or proportional symbology for point datasets gives unique visualization capabilities. For the eye-tracking fixation points, the fixation duration can be used as a parameter for the graduated symbology or proportional symbology varying by size. This technique of visualization is somewhat similar to the standard ET software.

In comparison, the GIS software have much more options regarding symbology and labeling. Figure 10 shows the fixations points displayed in ArcGIS Pro as graduated symbols with time duration as labels and a variable. Unlike the ET software where the symbol for the fixations can only be circle or cross, GIS software allow customization of symbols in terms of shapes, colors and combination of both. The maximum and minimum symbol range can also be adjusted to fit the purpose of the analysis.

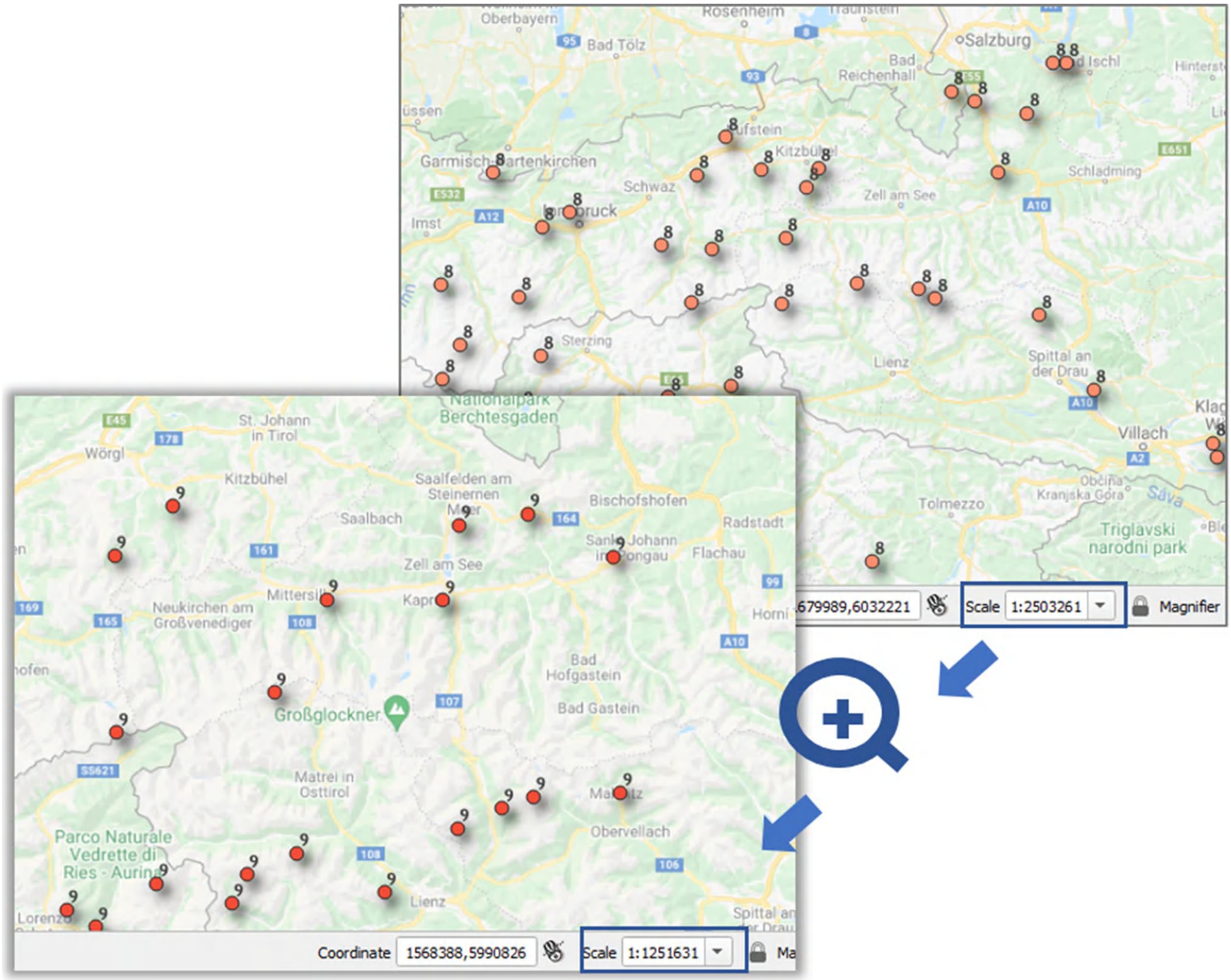


Fig. 9 Scale-based rendering of ET points

Custom scanpaths

Scanpath, in eye-tracking, is the sequence of fixations and saccades i.e., the areas the user viewed on screen one after another. Scanpaths are one of the basic visual analytics mechanisms in eye-tracking and almost every ET software has features for constructing and displaying them.

Scanpaths can be recreated in the GIS software with multiple possibilities. Figure 11 shows the scanpaths created in ArcGIS Pro. Using point to line tool, a new feature dataset was created, this dataset was displayed in addition to the fixation point data layer. The order field takes input of the order of points and the index column was specified here. The symbology for both the point and line dataset can be manipulated separately.

The index column, as exported from ET2Spatial tool, allows sequential temporal ordering of points. The index label is hence used to specify the gaze sequence of the participants as scanpaths. The additional zoom level labels can help in understanding the trend of using zoom levels by the users while doing map tasks. In the referenced Fig. 11, the red labels represent the zoom level associated with each fixation point. The interactivity of scroll and pan in the GIS software also allows the labels to be explored more clearly when zoomed in to a smaller area which otherwise might appear clustered in static images in frame-by-frame analysis through ET software. One thing to note is that in Fig. 11, the scanpaths have been generated for the whole experiment and for all users combined, which is not a straightforward possibility in ET software.

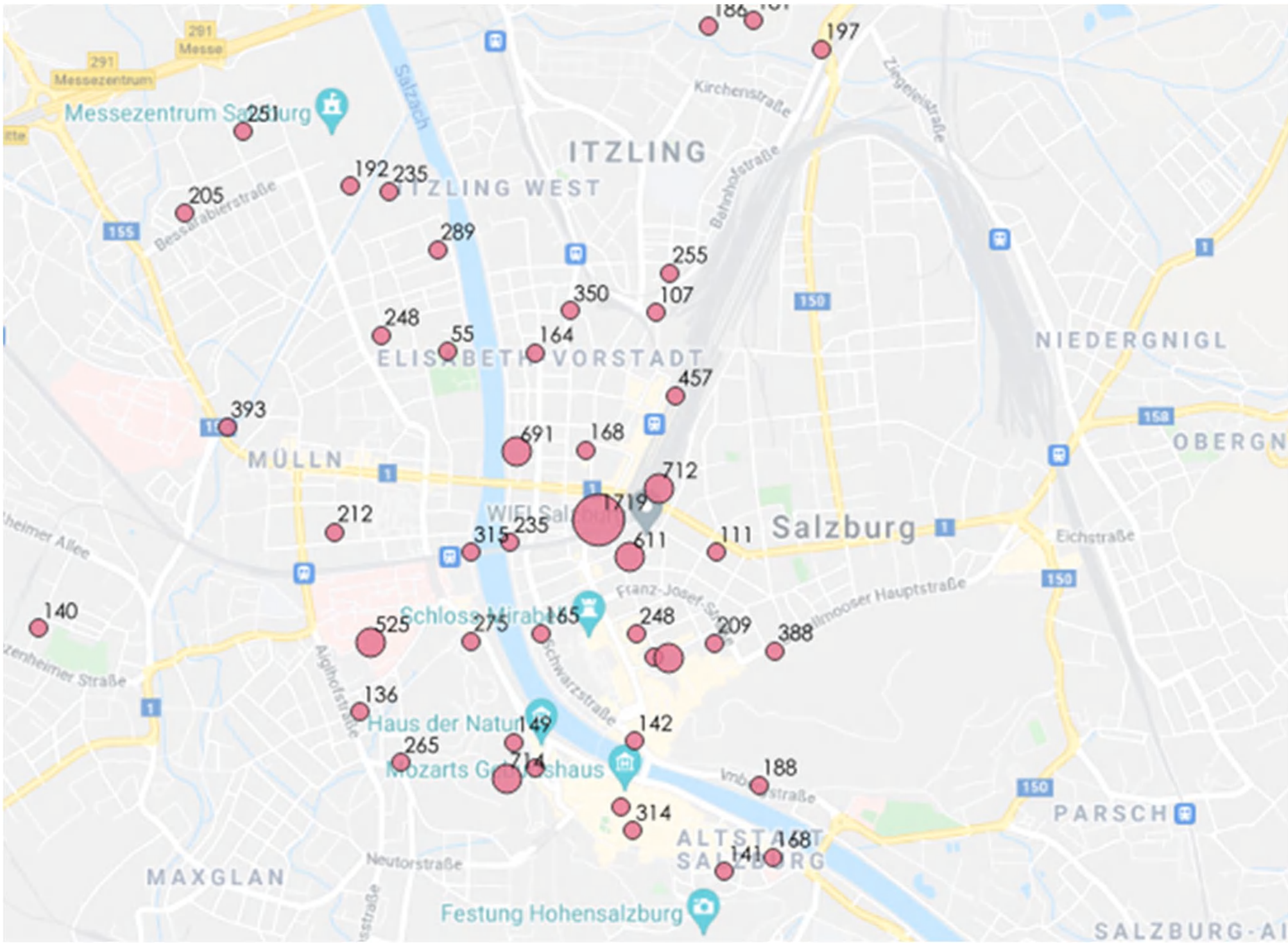


Fig. 10 Fixation points with graduated symbology

Many interesting avenues are opened to explore users' cognitive behavior with interactive maps when the scanpaths are aggregated not just by their temporal sequence but with participant and zoom level individually. To take the multiple labelling ability one step further. In Fig. 12 the left image shows the scanpaths categorized by zoom levels. These lines were aggregated on the zoom column and depict the temporal sequence of fixations at every zoom level. So, for instance, the yellow line shows the sequence of fixations by all users at zoom level 15. A convenient addition to this would be the scale-based rendering for every category. In the right figure, the lines were aggregated for every participant; this scenario is the closest to the one for ET software. For the sequence labels in these cases, the index would need to be reset for each category, something that is done automatically in the ET software for every participant.

Overlay image visualization

All above mentioned visualization methods depict eye movement data recorded over the Google basemap. However, MapTrack (Google Maps API) allows to add vector layers or overlay images to the displayed map. ET2Spatial can thus serve also for analysis of user behavior with these overlay images. As an example, an archaeological predictive map was georeferenced and converted into tiles with the use of MapTiler application. This overlay imagery was then added to the MapTrack environment and the user interaction data was recorded. The dataset along with the eye movement data were processed using ET2Spatial as in the aforementioned examples. The same archaeological map as was used in the MapTrack has been displayed in the ArcGIS environment together with fixation points displayed using graduated points symbology. Each color represents one participant (Fig. 13).

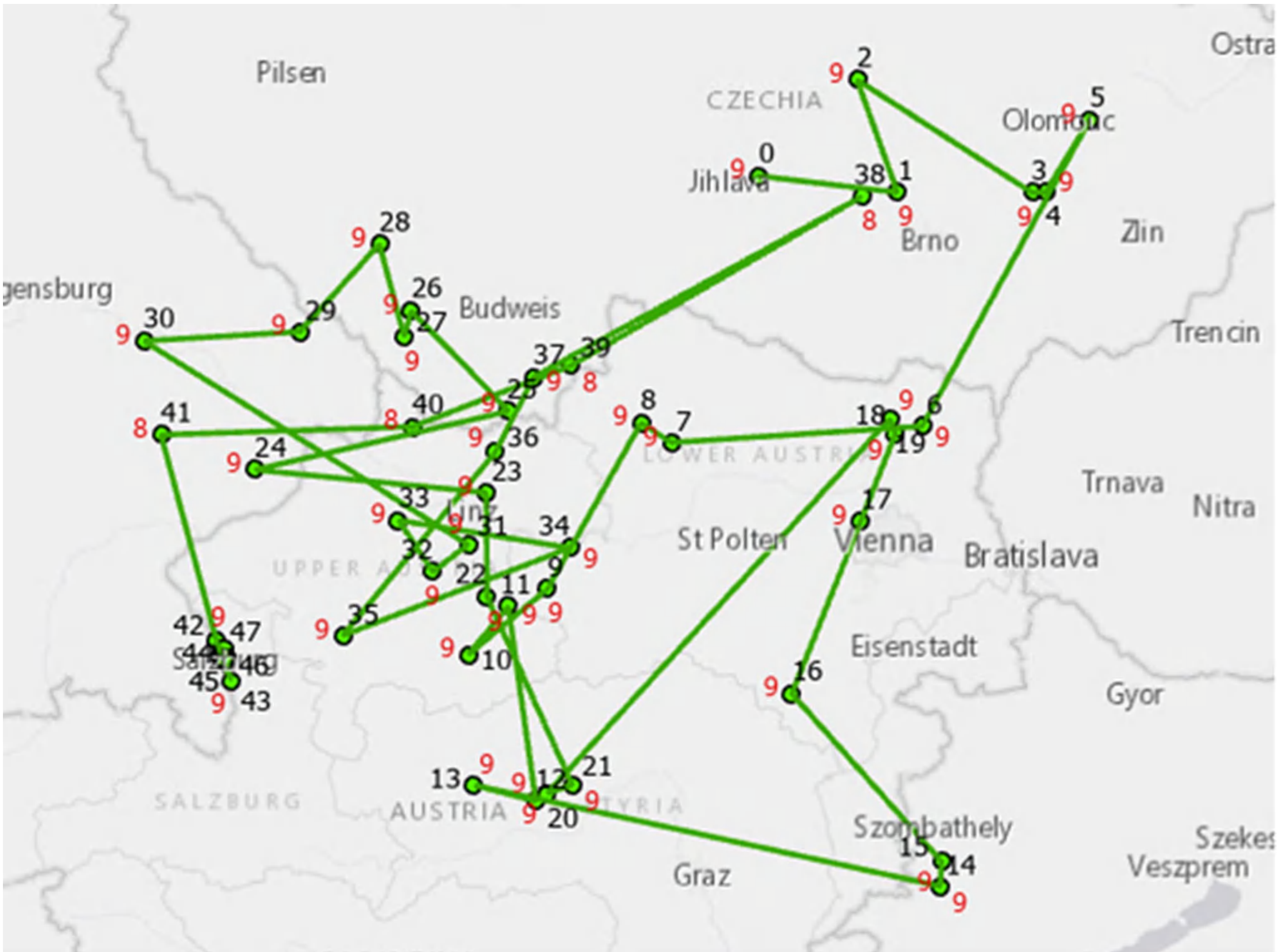


Fig. 11 Scanpaths in ArcGIS Pro with zoom level labels in red

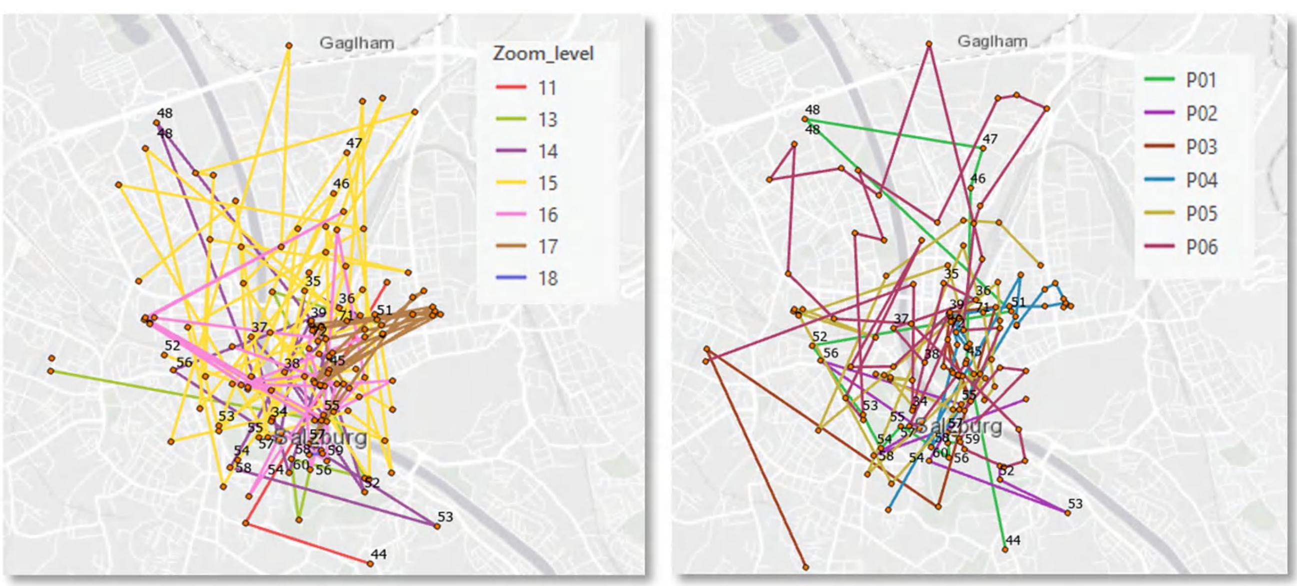


Fig. 12 Scanpaths aggregated by zoom level vs participant

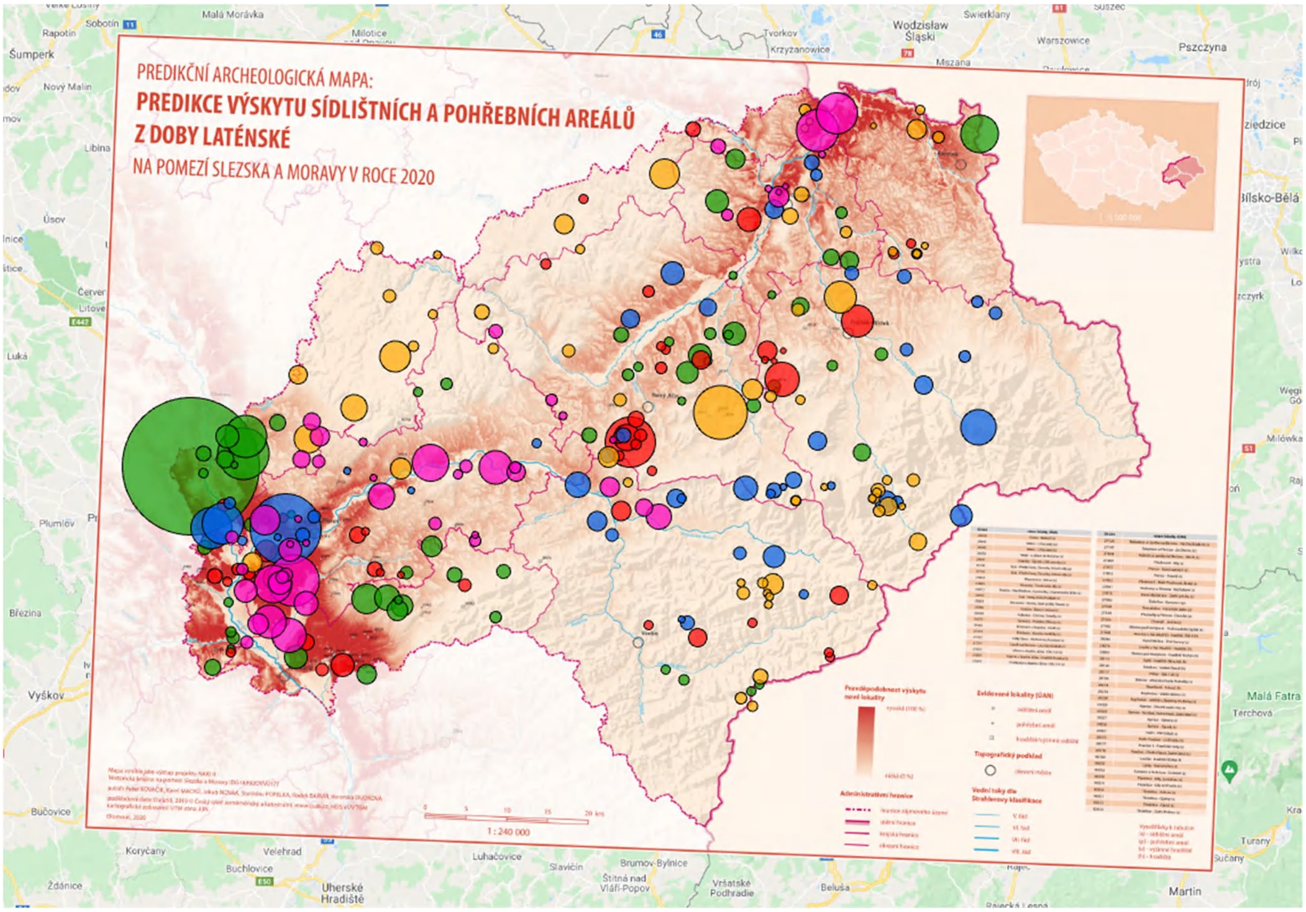


Fig. 13 Fixation points with graduated symbology recorded on image overlay

Discussion

The driving force behind this study were the cumbersome analysis techniques for eye-tracking data captured on interactive stimuli, particularly web maps. The tool developed as a result addressed these issues by georeferencing screen coordinates observed through eye-tracking.

As was mentioned in the introduction, it is not the first attempt to deal with this task. Ooms et al. (2015) proposed a similar approach. However, the proposed methodology differs from ET2Spatial in various ways. Dedicated Map APIs are not taken into account, and the entire referencing procedure is based on detailed user-logging actions such as pan, zoom, scroll, and click. However, some smaller parts of the proposed technique have been adopted in this tool, such as the imposition of tactile user input for easier synchronization of data.

The concept of *ET2Spatial* is also similar to FeatureEyeTrack (Göbel et al. 2019), but the framework, approach, and used technologies are different. Moreover, FeatureEyeTrack is not available under a public license which created room for similar tools to be developed. *ET2Spatial* mainly works in post-processing mode and is independent of database requirements, system settings, and any installations. It builds on free, open-source technologies and components. The tool created works with specific technologies as of now i.e., MapTrack and SMI RED 250 eye-tracking system. Although ET2Sspatial is configured with the output from SMI BeGaze, the tool can theoretically work with output from any ET system if naming conventions for the columns are followed. Tobii2spatial[6] was developed as an auxiliary tool to provide convenience when working with Tobii Pro eye-trackers, by converting the Tobii output files in ET2Spatial readable file structures.

The tool ET2Spatial, takes the input of one participant's files at a time, which in large experiments could be time-consuming. The tool can be scaled to have a bulk file upload feature, but to simplify the continuous conversion process for multiple participants, the reset feature was added.

6 www.eyetracking.upol.cz/tobii2spatial

A fair advantage of transforming the eye-tracking points into the geographic space was the availability of spatial functions that could be applied to the eye-tracking data of several participants. These GIS functions will help in analyzing the ET data from a geospatial perspective. The paper contains a few examples how eye movement data can be viewed and analyzed in GIS software, however, the possibilities of spatial analyses are enormous and will be further investigated.

Conclusions

Standard eye-tracking systems offer good evaluation techniques for static stimuli when it comes to geovisualization products. However, the current practices carried out for analysis with eye-tracking on interactive environments such as web maps are very cumbersome and time-consuming. Research done in this regard has been scarce, and the availability of free, open-source tools addressing the issue is far lesser.

The goal of this study was to create a tool that would solve the problems of eye-tracking analysis on dynamic interactive web maps. The tool developed called ET2Spatial, converts screen coordinates recorded during an eye-tracking experiment to real-world geographic coordinates.

ET2Spatial was developed in Python using a variety of modules. The tool takes three input files; the raw ET points, fixation points, and the user interaction data. These input datasets are pre-processed, synchronized based on timestamps, and stitched together. The main conversion of points relies on the Web Mercator projection formulas. Eventually, the tool offers export in shapefile format along with a CSV and a geojson format.

The tool was evaluated in terms of accuracy of output, performance, and its general usability. Post-conversion, recorded ET datasets were imported in GIS software such as ArcGIS and QGIS to test the visualization capabilities as well as the evaluation of points with spatial techniques. The pilot studies demonstrated good results and multiple options for visualizing eye-tracking data in the GIS environment and the usage of spatial functions to amplify the scope of analysis.

ET2Spatial builds on the existing technology developed at the department, i.e., MapTrack, and aims to provide a harmonized framework for carrying out future research. Despite being slightly technology-specific, the tool is open-source and can be used anywhere provided an eye-tracking system exists. The tool has the potential to be scaled up and can be employed for usability studies of interactive cartographic mediums as well as analysis of human interaction and cognition with web maps.

Software files

https://github.com/minha94/ET2Spatial

https://github.com/ondrejruzicka/maptrack

Availability and Requirements

The tool is standalone and should not need additional installations. However, in case of problems, these are the recommended steps:

- Installation of Python v 3.5 + on the local system (https://www.python.org/downloads/release/python-387/), Make sure that you add python to PATH
- Installation of the packages. In the requirements.txt all the necessary modules are listed, you can install them individually or through 'pip install -r requirements.txt' in shell.

Getting data from MapTrack:

- Maptrack is avalilable at https://github.com/ondrejruzicka/maptrack.
- To use MapTrack data with eye-tracker, it must be selected as a stimulus in presentation settings.
- Enter the participant Id, Matching with the one entered in eye-tracking system. Press F11 to enter fullscreen mode. Press F2 to start.
- After the experiment is finished, Press F2 again to terminate the 'interaction mode'. The results can be accessed at: http://eyetracking.upol.cz/maptrack/results/
- Download the xml file for the relevant participant ID and finished time.

How to use the tool:

- Download the tool https://github.com/minha94/ET2Spatial.
- Double click the exe file. The tool takes maximum 1 min to compile. The icon files should be in the same directory as the exe file
- For Eye-tracking Data following convention should be followed:
- In the Raw gaze points: ['Time of Day [h:m:s:ms]','RecordingTime [ms]','Point of Regard Right X [px]','Point of Regard Right Y [px]','Participant']
- In the Fixations gaze points following elements should exist: ['Event Start Trial Time [ms]','Event Duration [ms]','Fixation Position X [px]','Fixation Position Y [px]''Participant']
- Users that are using Tobii Pro eye-trackers can convert their data using www.eyetracking.upol.cz/tobii2spatial. For that, it is necessary to export Single TSV (timestamp precision = miliseconds) file with these columns: 'Recording timestamp', 'Sensor', 'Participant name', 'Event value', 'Gaze point X', 'Gaze point Y', 'Presented Stimulus name', 'Eye movement type', 'Gaze event dura-

tion', 'Eye movement type index', 'Fixation point X', 'Fixation point Y'

- For the User Interaction file, the tool currently only works with MapTrack export (maptrack/results)
- On the initiation of the tool a terminal window also opens, the notifications for the export buttons are printed in this window.

Exports:

The exports are named after the participant number which was given in input files.

The tool exports

1) shapefile: Which will produce CSV and Shapefiles for both raw points and fixation points. The shapefiles are exported in the EPSG 3857 Projection System.
2) GeoJSON: Which will produce GeoJSON for both raw points and fixation points

Shapefile:

.shp,.shx,.dbf.

The shapefile has the following contents:

- **Raw Points.shp**
  - ind: index or serial number in temporal order
  - latitude: in decimal degrees
  - longitude: in decimal degrees
  - zoom level: Google maps zoom factor at which each point was recorded
  - Format_RT: the time stamp in millisecond for each point
- **Fixation Points.shp**
  - ind: index or serial number in temporal order
  - latitude: in decimal degrees
  - longitude: in decimal degrees
  - zoom level: the Google maps zoom factor at which each point was recorded
  - Sync_time: the time stamp in millisecond for each point
  - Duration: the amount of time of a fixation on screen

Along with the shapefile export there is a:

- PROJECTION FILE: It contains the information on the projection of the exported points in WKT format
- CSV FILES: all the above mentioned attributes of points in csv format.

Geojson:

The geojson file contains a point geometry file with associated properties:

Raw Points.geojson

- Geometry: latitude, longitude
- Properties: zoom level, Format_RT

Fixation Points.geojson

- Geometry: latitude, longitude
- Properties: zoom level, Sync_time, Duration

Funding This study was done as part of the master thesis of first author, supported by the Erasmus + Programme of the European Union. The paper was also supported by the Internal Grant Agency of Palacký University Olomouc (grant number IGA_PrF_2022_027) and by the Ministry of Culture Czech Republic (grant number DG18P02OVV017).

Declarations

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

References

Andrienko G, Andrienko N, Burch M, Weiskopf D (2012) Visual analytics methodology for eye movement studies. IEEE Trans Visual Comput Graphics 18(12):2889–2898. https://doi.org/10.1109/TVCG.2012.276

Beitlova M, Popelka S, Vozenilek V (2020) Differences in thematic map reading by students and their geography teacher. ISPRS Int J Geo Inf 9(9):492

Burch M (2019) Interaction graphs: visual analysis of eye movement data from interactive stimuli. In: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications. Denver, Colorado. https://doi.org/10.1145/3317960.3321617

Burch M, Kurzhals K, Kleinhans N, Weiskopf D (2018) EyeMSA: exploring eye movement data with pairwise and multiple sequence alignment. Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications. Warsaw, Poland. https://doi.org/10.1145/3204493.3204565

Dolezalova J, Popelka S (2016) ScanGraph: a novel scanpath comparison method using visualisation of graph cliques. J Eye Mov Res 9(4):Article 5. https://doi.org/10.16910/jemr.9.4.5

Dong W, Liao H, Liu B, Zhan Z, Liu H, Meng L, Liu Y (2020) Comparing pedestrians' gaze behavior in desktop and in real environments. Cartogr Geogr Inf Sci 47(5):432–451. https://doi.org/10.1080/15230406.2020.1762513

Enoch JM (1959) Effect of the size of a complex display upon visual search. JOSA 49(3):280–285. https://doi.org/10.1364/JOSA.49.000280

Eraslan S, Yaneva V, Yesilada Y, Harper S (2019) Web users with autism: eye tracking evidence for differences. Behav Inf Technol 38(7):678–700

Giannopoulos I, Kiefer P, Raubal M (2012) GeoGazemarks: providing gaze history for the orientation on small display maps. In: Proceedings of the 14th ACM international conference on Multimodal interaction. Santa Monica, California, USA. https://doi.org/10.1145/2388676.2388711

Göbel F, Kiefer P, Raubal M (2019) FeaturEyeTrack: automatic matching of eye tracking data with map features on interactive maps. GeoInformatica 2019(23):663–687. https://doi.org/10.1007/s10707-019-00344-3

Havelková L, Gołębiowska IM (2020) What went wrong for bad solvers during thematic map analysis? Lessons learned from an eye-tracking study. ISPRS Int J Geo Inf 9(1):9

Herman L, Popelka S, Hejlova V (2017) Eye-tracking analysis of interactive 3D geovisualization. J Eye Mov Res 10(3):1–15. https://doi.org/10.16910/jemr.10.3.2. Article 2

Holmqvist K, Nyström M, Andersson R, Dewhurst R, Jarodzka H, Van de Weijer J (2011) Eye tracking: a comprehensive guide to methods and measures. Oxford University Press, Oxford

Jenks GF (1973) Visual integration in thematic mapping: fact or fiction? International Yearbook of Cartography, 13 pp 27–35

Jenks GF (1974) The average map-reader lives. Paper presented at Annual meeting of the Association of American Geographers, Seattle, Washington

Kiefer P, Giannopoulos I (2012) Gaze map matching: mapping eye tracking data to geographic vector features. In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems. Redondo Beach, California. https://doi.org/10.1145/2424321.2424367

Krassanakis V, Cybulski P (2019) A review on eye movement analysis in map reading process: the status of the last decade. Geod Cartogr 68(1):191–209

Krassanakis V, Filippakopoulou V, Nakos B (2014) EyeMMV toolbox: an eye movement post-analysis tool based on a two-step spatial dispersion threshold for fixation identification. J Eye Mov Res 7(1). https://doi.org/10.16910/jemr.7.1.1

Ooms K, De Maeyer P, Fack V, Van Assche E, Witlox F (2012) Interpreting maps through the eyes of expert and novice users. Int J Geogr Inf Sci 26(10):1773–1788. https://doi.org/10.1080/13658816.2011.642801

Ooms K, Coltekin A, De Maeyer P, Dupont L, Fabrikant S, Incoul A, Kuhn M, Slabbinck H, Vansteenkiste P, Van der Haegen L (2015) Combining user logging with eye tracking for interactive and dynamic applications. Behav Res Methods 47(4):977–993. https://doi.org/10.3758/s13428-014-0542-3

Papenmeier F, Huff M (2010) DynAOI: a tool for matching eye-movement data with dynamic areas of interest in animations and movies. Behav Res Methods 42(1):179–187. https://doi.org/10.3758/BRM.42.1.179

Pfeiffer T (2012) Measuring and visualizing attention in space with 3D attention volumes. Symposium on Eye Tracking research & applications (ETRA 2012), Santa Barbara, California, USA

Popelka S, Beitlova M (2022) Scanpath Comparison using ScanGraph for Education and Learning Purposes: Summary of previous educational studies performed with the use of ScanGraph. 2022 Symposium on Eye Tracking Research and Applications, Seattle, WA, USA. https://doi.org/10.1145/3517031.3529243

Popelka S, Burian J, Beitlova M (2022) Swipe versus multiple view: a comprehensive analysis using eye-tracking to evaluate user interaction with web maps. Cartography and Geographic Information Science, pp 1–19. https://doi.org/10.1080/15230406.2021.2015721

Popelka S, Dolezalova J, Beitlova M (2018) New features of scangraph: a tool for revealing participants' strategy from eye-movement data. In: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications. Warsaw, Poland. https://doi.org/10.1145/3204493.3208334

Růžička O (2012) Profil uživatele webových map. Palacký University Olomouc, Olomouc

Šašinka Č, Morong K, Stachoň Z (2017) The hypothesis platform: an online tool for experimental research into work with maps and behavior in electronic environments. ISPRS Int J Geo Inf 6(12):1–22

Šašinka Č, Stachoň Z, Čeněk J, Šašinková A, Popelka S, Ugwitz P, Lacko D (2021) A comparison of the performance on extrinsic and intrinsic cartographic visualizations through correctness, response time and cognitive processing. PLoS ONE 16(4):e0250164. https://doi.org/10.1371/journal.pone.0250164

Skrabankova J, Popelka S, Beitlova M (2020) Students' ability to work with graphs in physics studies related to three typical student groups. J Balt Sci Educ 19(2):298–316

Słomska K (2018) Types of maps used as a stimuli in cartographical empirical research. Misc Geogr 22(3):157–171. https://doi.org/10.2478/mgrsd-2018-0014

Unrau R, Kray C (2019) Usability evaluation for geographic information systems: a systematic literature review. Int J Geogr Inf Sci 33(4):645–665. https://doi.org/10.1080/13658816.2018.1554813

Paper Dashboards

Porti Suarez, A., & **Popelka, S.*** (2023). Design Aspects for COVID-19 Dashboards–Evidence from Eye-Tracking Evaluation. International Journal of Human–Computer Interaction, 1-19. https://doi.org/https://doi.org/10.1080/10447318.2023.2272078

[IF 3.4; Q2]

[0 citations on WoS; 0 citations without autocitations]

[Author's contribution: 50%]

Abstract

The outbreak of the COVID-19 pandemic caused dashboards to become widely used by the public and decision-makers. Nevertheless, dashboard interfaces have been related to business intelligence since their origins, and the search for improvements in their design is not new. This article's objective is to conduct a user evaluation of COVID-19 dashboards that contain geospatial information. This is done through a formative study to identify problematic aspects of user/dashboard interaction. This is enhanced by comparing two self-developed dashboards that, according to previous tests, have functionalities with different appearances. User evaluation is performed through mixed research that combines objective (eye-tracking) and subjective (a questionnaire and an interview) methods. The results generate recommendations for better-designed dashboard interfaces that can transfer information appropriately. The vital elements needed to achieve this are interactivity, the option to choose the metrics, and the distribution of the elements in the layout, all playing a role in a more user-friendly interaction between the user and the dashboard.

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2023.2272078

Design Aspects for COVID-19 Dashboards – Evidence from Eye-Tracking Evaluation

Anna Porti Suarez and Stanislav Popelka

Department of Geoinformatics, Palacký University Olomouc, Olomouc, Czechia

ABSTRACT

The outbreak of the COVID-19 pandemic caused dashboards to become widely used by the public and decision-makers. Nevertheless, dashboard interfaces have been related to business intelligence since their origins, and the search for improvements in their design is not new. This article's objective is to conduct a user evaluation of COVID-19 dashboards that contain geospatial information. This is done through a formative study to identify problematic aspects of user/dashboard interaction. This is enhanced by comparing two self-developed dashboards that, according to previous tests, have functionalities with different appearances. User evaluation is performed through mixed research that combines objective (eye-tracking) and subjective (a questionnaire and an interview) methods. The results generate recommendations for better-designed dashboard interfaces that can transfer information appropriately. The vital elements needed to achieve this are interactivity, the option to choose the metrics, and the distribution of the elements in the layout, all playing a role in a more user-friendly interaction between the user and the dashboard.

KEYWORDS

Dashboard; eye-tracking; geospatial information; qualitative methods; usability testing; user interface

1. Introduction

The year 2020 was marked by the presence of the COVID-19 pandemic. The spread of the virus created the need to quickly communicate critical information to a range of stakeholders, including policymakers, healthcare practitioners, and the general public (Kamel Boulos & Geraghty, 2020). Interactive dashboards are an efficient visualization mechanism for presenting information that can be both used by relevant stakeholders to take action and engage in rapid decision-making and by the community to inform themselves. Moreover, they are crucial for guiding modeling approaches and implementing control actions during the initial phases of the outbreak (Dong et al., 2020).

1.1. Dashboards

A dashboard is a type of graphical user interface able to display selected data that regularly updates. Even though there is no exact definition of the term *dashboard*, Few (2006, p. 26) defines it as; "A visual display of the most valuable information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance."

Understanding the relationship between dashboards and the field of business intelligence is essential to contextualizing their histories since their beginnings are significantly related. According to Few (2006), dashboard precursors are executive information systems (EISs). They were limited to executives' offices in terms of accessibility and understanding. Data displayed in the dashboards and integrated with other technologies did not become widespread until the 90s (Few, 2006). Other precursors to dashboards developed in the end of the millennium were key performance indicators (KPIs); tools used by managers to monitor their company's performance, composed of indicators using diverse metrics and highlighting areas of importance (Marr, 2012).

These precursors developed into what a dashboard is today – a display where information is presented visually, usually as a combination of text and graphics (Few, 2006). This highlights five essential points regarding dashboard design and its relation to visual perception:

- Displays the information necessary to achieve one or more specific objectives.
- Fits onto a single computer screen so everything can be seen at once.
- Monitors information at a glance.
- Has small, concise, clear, and intuitive mechanisms.
- Customized to serve the user's purpose.

Dashboards are commonly used to visualize geospatial information by integrating maps with other elements. According to Kitchin et al. (2015), different sections of a dashboard can be clicked on to highlight data points on a graph and the corresponding areas on a map. This facilitates the understanding of the spatial dimension of selected phenomena.

The COVID-19 outbreak in 2020 caused dashboards to be widely used by the public for consulting metrics and,

CONTACT Stanislav Popelka stanislav.popelka@upol.cz Department of Geoinformatics, Palacky University Olomouc, Olomouc, Czechia

according to Google Trends for searching the term *dashboard*, this played a role in increasing the popularity of dashboards. The number of searches for this term quadrupled between January 2020 and March 2021 (GoogleTrend, 2023).

A plethora of web-based GIS dashboards incorporating essential GIS functionality have been developed. These dashboards enable swift data sharing and instant access to information, thereby streamlining the decision-making process (Lan et al., 2021).

Dashboards have surged in popularity recently as instruments for data visualization, yet the data are not always appropriately displayed on all the available platforms. The intricacy of designing dashboards involves abstracting data, managing screen space, grouping components, illustrating connections and incorporating interactive features for purposes such as exploration, drilldown, navigation, and customization (Bach et al., 2022). The art of visual communication encompasses semantics and syntax similar to verbal language. Understanding the rules is crucial for effective communication through graphs, since this process is largely scientific, grounded in established knowledge about visual perception and cognition (Few & Edge, 2007).

1.2. Design aspects of dashboards

Design aspects are crucial to achieving the communication of information displayed. Nevertheless, it is hard to establish rules when designing such interfaces, as they are composed of different elements and are created for different purposes. Sedrakyan et al. (2019) suggest that although dashboard solutions are increasingly popular and numerous, there is still a lack of understanding regarding their design elements, particularly in terms of selecting visualizations during the development of dashboards, which argues that despite a growing number of public examples, case studies, and general guidelines, there is a surprising lack of design guidance for dashboards. In the design of dashboards, particular attention must be paid to usability principles and human factors to deliver interactive and data sharing capabilities (Carayon & Hoonakker, 2019). Dashboards could be a valuable part of public health initiatives and communicate vital information to their users. To achieve this and empower healthcare stakeholders, it is essential that developers adopt a human-centered approach (Monkman et al., 2021).

Ching-Yi et al. (2018) recommend a new design framework inspired by previous works. This includes the principle of classification, which means dividing a larger domain into smaller parts based on specific traits, in order to help users recognize design elements. They also emphasize simplicity, advocating the use of methods to choose, filter, drill down, and request detailed information on demand. Finally, the principle of synthesis is used, combining elements with related but distinct indicators.

Monkman et al. (2021), while examining COVID-19 dashboards, assert that these tools struggle to effectively convey information to the public. They further highlight the ambiguity of whether dashboards are generally successful in conveying pertinent trends and patterns and in meeting the informational needs of users. Following the heuristics for dashboard visualizations by Dowding and Merrill (2018), Monkman et al. (2021) found and listed shortcomings that should be improved in future dashboard design.

- Visibility of the system's status: provide the currency of the data update.
- Match between the system and the real world: label and describe the phenomena they are referencing.
- User control: find a way to restore the dashboard to the default view.
- Recognition rather than recall: display the information on a single screen without the need to scroll.
- Aesthetics and minimalist design: avoid unnecessary words, graphics, and visualizations.
- Orientation: include clear titles and maps.

Therefore, to make dashboards a useful and user-friendly visual tool that provides insights into specific phenomena and, in this case, their spatial dimension, it is necessary to analyze how the information should be integrated into aspects of the design.

1.3. User evaluation

According to MacEachren (1995), creating effective maps requires an understanding of the capabilities and limitations of our visual-cognitive system. Specifically, we need to understand how vision and cognition contribute to our ability to interpret and derive meaning from visual scenes.

Because of this, to achieve a proper dashboard design containing geospatial information, user evaluation needs to be performed through studies into the perception of information by the human brain as it relates to the field of cognitive cartography. Cognitive cartography focuses on how humans perceive information on a map. Eye-tracking is, among others, a method used in cognitive cartography that records eye movement and converts it into data derived and measured to obtain insights into cognitive processes. Combined with other methods, these measurements provide both quantitative and qualitative data for analysis and for gaining insights regarding user experience.

A recent work by Krassanakis and Cybulski (2019) explains the current panorama of existing eye-tracking studies in the field of cartographic research, concluding that eye movement analysis is part of the cartographic field, and that a remarkable number of research studies used eye-tracking technology to analyze map reading processes. Eye movement, according to Coltekin et al. (2009), provides insights that can enhance the understanding of how humans interact with interactive map interfaces.

This also applies to cartography on the web and dashboards, which are compounded by interactive maps. Their interaction with the user is explored in studies such as that by Zuo et al. (2020), proposing a design model for a map-based dashboard with a methodology that combines eye-tracking and interviewing to analyze a user's experience. Their study concludes with details of which specific elements of the layout should be changed or improved, such as the font size or the arrangement itself.

Li et al. (2022) performed an empirical study and made observations with five COVID-19 dashboards. One of the research questions in the study was focused on information visualization. Based on the results of the analysis, they found that tree map, tabular data and donut charts are used for one-dimensional data; line charts, bar charts, and histograms for two-dimensional data; and bubble charts and maps are employed for the visualization of multi-dimensional data types.

Praharaj and Wentz (2022) built a dashboard prototype offering data segregation by population sub-groups that enlighten users about the unequal effects of disasters and pandemics. Then, Praharaj et al. (2023) performed its comprehensive usability assessment using online questionnaire survey filled in by 30 participants. User ratings using a five-point scale were targeted to the appropriateness of the dashboard indicators, the ease of navigation, users' trust in the data sources, and the dashboard's utility for informed decision-making. Their findings suggest that dashboards are perceived differently by various actors based on their specific job roles and information needs.

Fan et al. (2023) used the think-aloud technique to understand how older adults interact with and comprehend online COVID-19 visualizations. They selected five visualizations from the 57 dashboards that they identified. Based on the verbalized thoughts of the participants, the authors confirmed and extended the three-process visualization comprehension theory (Carpenter & Shah, 1998). They identified four types of thought processes: encoding visual information; relating visual information to concepts; associating concepts with existing knowledge; and recovering from errors.

By organizing a dashboard design workshop, Bach et al. (2022) aimed to understand the design patterns of dashboards regarding the information display. During the workshop, they discuss the application of patterns for the dashboard design processes, as well as general design trade-offs and common challenges.

Monkman employed a team of user experience experts to assess visual design of COVID-19 dashboards. The result of this evaluation was that many dashboards violated design heuristics and contained many usability issues that may have created challenges for their users.

1.4. Aim of the article

The aim of the article is to gain insights into design aspects of COVID-19 dashboards, based on user evaluation using eye-tracking. Understanding these design aspects can enhance data visualization and lead to improved understanding. Additionally, it can help rectify design challenges and result in dashboards that are more user-friendly and effective in promoting awareness and informed decision-making across various fields. For this, three goals are established and described as follows.

The *first goal* is to obtain information about four existing dashboards and to acquire insights into the users' interaction with their functionalities and find whether or not the users' goals are met. This consists of a formative study with three steps: the design of the experiment, the recording of the data, and the processing and analysis of the obtained data. The results allow us to identify the problematic elements of these dashboards and formulate recommendations, which are then followed to find which dashboard elements are user-friendly and communicate the information accordingly.

The *second goal* is to create self-developed dashboards according to the insights obtained in goal 1. Once the problematic elements and the recommendations for improving user interaction are identified, two new dashboards are produced to verify the insights obtained in the first goal. The dashboards will be used as the stimuli in the empirical assessment, which is part of the third goal.

The *third goal* is to verify the usability of the two self-developed dashboards considering the insights of goal 1. This goal also consists of a formative study that identifies and compares each dashboard's positive and negative elements to concur with the previous studies. Like the first goal, this involves the design of the experiment, the recording of the data, and its processing and analysis, leading to the results and conclusions.

2. Methods

2.1. Experiment design

To achieve the first goal, experiment I is performed to find problematic aspects of the selected existing dashboards. For the third goal, experiment II is performed in two self-developed dashboards, in which elements and layout are created following the findings obtained in experiment I.

Both experiments are designed using SMI Experiment Center 3.7, and consist of the following steps: Calibration, Introduction, Task solving, Interview, Questionnaire, and Acknowledgement. Experiment II also includes one minute of free exploration for each dashboard. The eye movement is recorded, as well as the participant's voice and image, with a camera and a microphone, the latter of which is particularly important for collecting data during the interview. A summary of the procedure for both experiments is displayed in Figure 1.

2.2. Stimuli and tasks

2.2.1. Experiment I

The four dashboards selected as the stimuli in experiment I meet the characteristics of Few's definition of a dashboard (2006), as they are interfaces displaying information at a glance. Moreover, all selected dashboards have different functionalities and contain geospatial visualizations.

The selected dashboards are shown in Figure 2: COVID-19 Map by the John Hopkins University (D1), Novel Coronavirus Incidence Map by the University of Washington (D2), OCHA COVID-19 Data Explorer (D3), and Health Map (D4).

Three tasks (T1, T2, and T3) that consist of finding an answer to a specific question and interacting with the dashboard interfaces are assigned to each of the dashboards, so

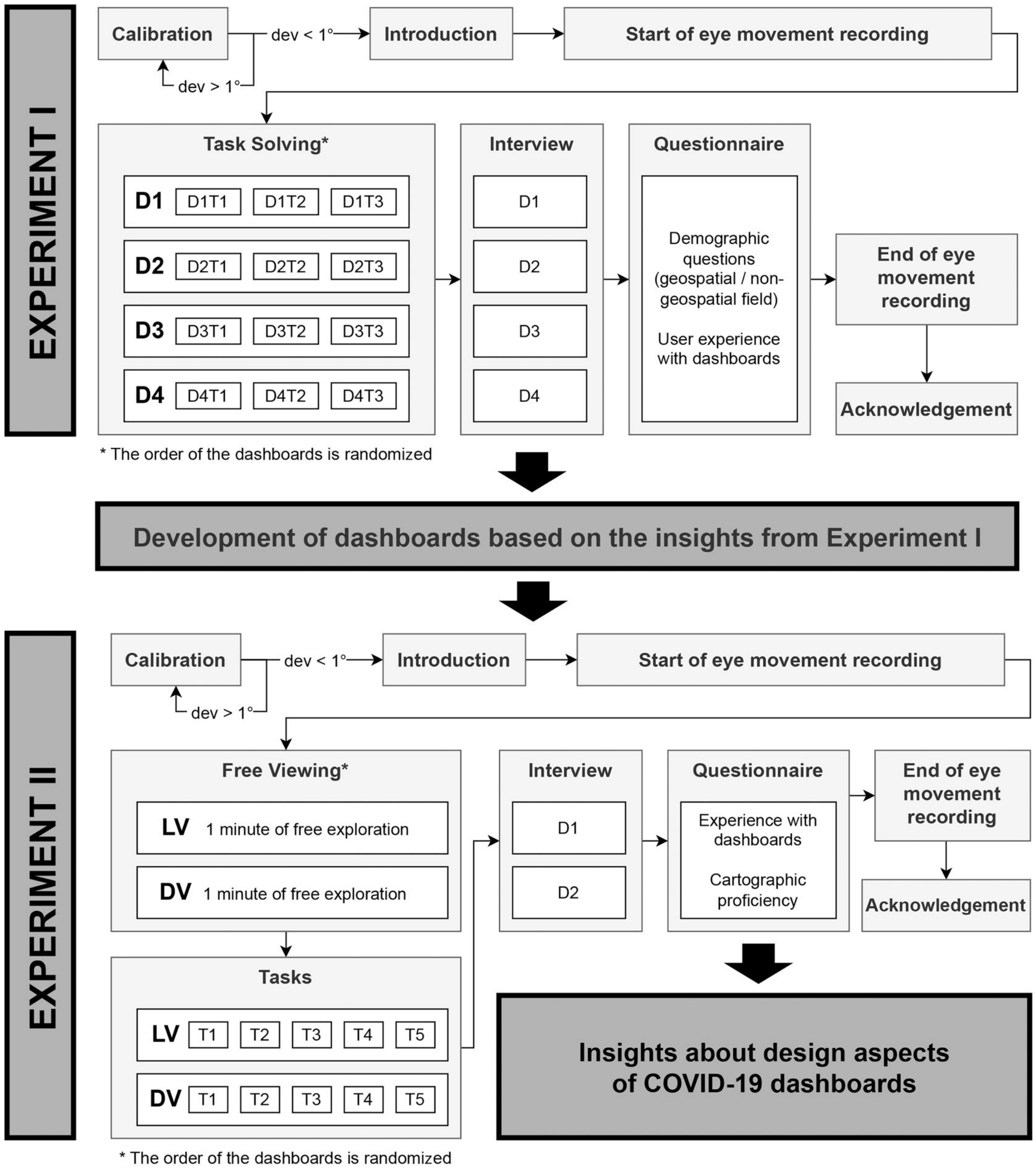


Figure 1. The schema of the study.

there are 12 tasks. The tasks are focused on: identifying the value for a particular area of its most recent update (identify-now), identifying the value for a specific date (identify-date), and comparing or ranking areas on a specific date (compare-date, rank-date). Typology of the tasks was taken from Roth and MacEachren (2016). The level of difficulty increases through tasks 1–3 (Table 1). The selection of the tasks is in line with the work of Fan et al. (2023), who used the think-aloud method to analyze the work of older adults with COVID-19 dashboards who carried out tasks very similar to those in this article.

2.2.2. Dashboard development

The two self-developed dashboards depicting Covid-19 data from Catalunya designed according to insights obtained in experiment I served as stimuli in experiment II.

The methodology of creating these dashboards involved several steps and tools. First, data were collected from the Catalogue of Open Data of the Government of Catalunya and the Statistical Institute of Catalunya. These data included CSV files for daily cases at county and town levels, as well as the geometry of these administrative levels, and population data. Python scripts ran on a Linux server to process and clean these data. PostgreSQL's spatial extension, PostGIS, was used for storing geographic objects. Tableau, a visual analytics platform, was utilized for the front-end design of the dashboards, with middleware facilitating communication between the server and user interface. Tableau directly retrieved the data from the database and joined tables containing the geometry with the COVID-19 metrics from different administrative levels.

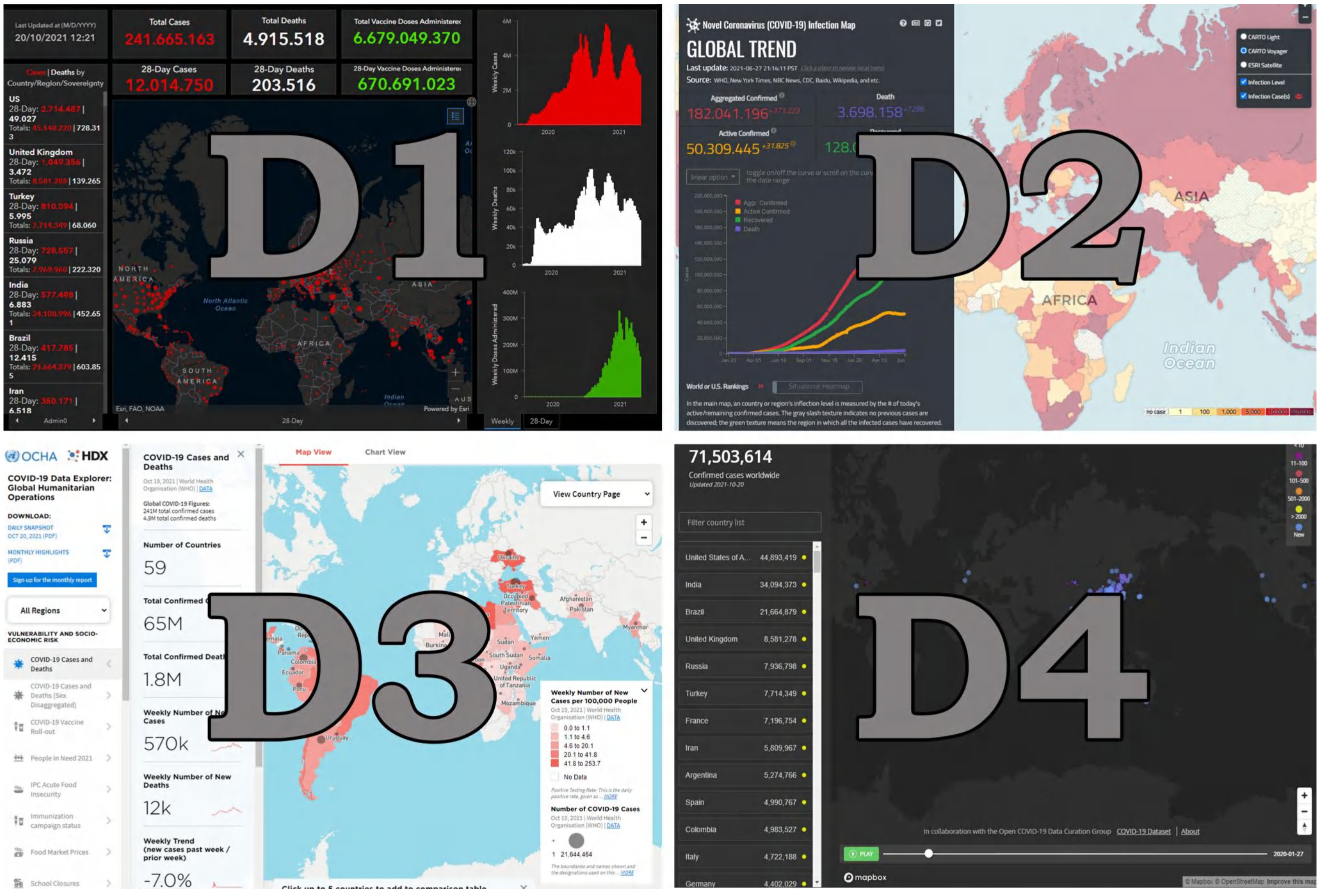


Figure 2. Four exiting dashboards used for the experiment I.

Table 1. Experiment I list of tasks.

ID	Task	Objective
D1T1	Total accumulated cases in Switzerland by the most recent update	Identify-now
D1T2	Daily cases in France on October 25 2021	Identify-date
D1T3	Weekly cases in California (US) during the week of January 3 2021	Identify-date
D2T1	Confirmed cases in Germany by the most recent update	Identify-now
D2T2	Aggregated confirmed cases in Russia on February 25 2021	Identify-date
D2T3	If the situation regarding the evolution of cases in Norway has improved or worsened from June 7 2021 to June 27 2021	Compare-date
D3T1	Total accumulated cases in Cameroon by the most recent update	Identify-now
D3T2	Number of accumulated cases in Ethiopia by July 10 2021	Identify-date
D3T3	A country with more weekly new cases (by the most recent update) between Brazil, Argentina, and Bolivia	Rank-now
D4T1	Accumulated cases in Italy by the most recent update	Identify-now
D4T2	Accumulated cases in Nicaragua by the most recent update	Identify-country
D4T3	Number of cases in Melbourne (Australia) on March 5 2020 and April 5 2020	Compare-date

The resulting dashboards are the Light Version (from now on LV) and the Dark Version (from now on DV). Both dashboards are visible in Figure 3.

2.2.3. Experiment II

Five tasks (T1, T2, T3, T4, and T5) are assigned to each of the self-developed dashboards, so in total there are 10 tasks (Table 2). Again, through tasks 1–5, the level of difficulty increases: T1 is the simplest with no need to change any parameter (identify-now); T2 and T3 require the changing of a parameter, either the region, the date, or both, to find numerical values (identify-now or identify-date); T4 and T5 require the changing of the administrative levels of the regions by selecting a tab, as well as a comparison of the numeric metrics from different dates and regions (compare-date or compare-place). The typology of the tasks is again taken from Roth and MacEachren (2016).

2.3. Questionnaire and interview

Both eye-tracking experiments were supplemented by a simple questionnaire, which contained Likert scale questions focused on the subjective opinions of the participants. First, they rated the level of difficulty in solving the tasks for each dashboard, from 1 (very difficult) to 5 (very easy). The participants were then asked to rate the aesthetics of the dashboards, from 1 (very ugly) to 5 (very attractive). The questionnaire was followed by an interview, where participants were asked to freely speak their minds regarding the dashboards and to talk about their positive and negative experiences during the task-solving processes.

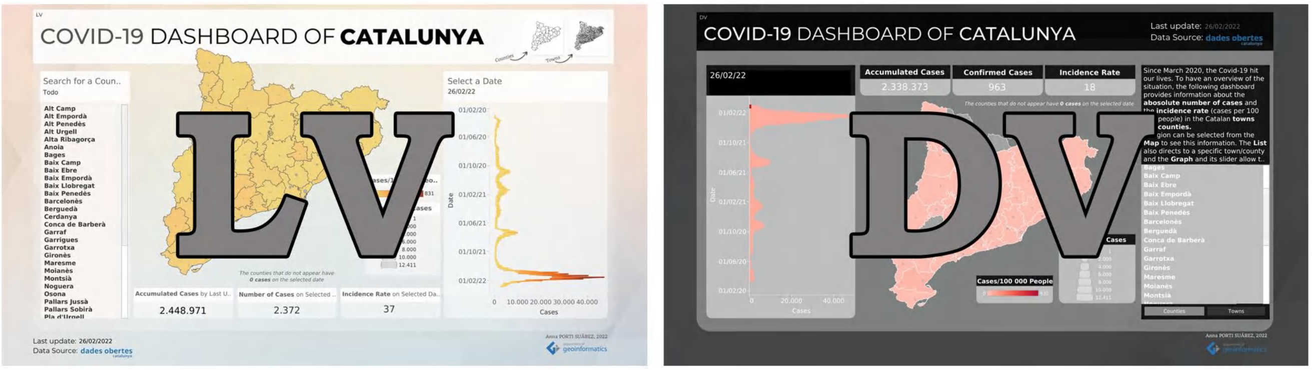


Figure 3. Two self-developed dashboards used in experiment II.

Table 2. Experiment II list of tasks.

ID	Task	Objective
LVT1	Accumulated cases in Catalunya by the most recent update	Identify-now
LVT2	Incidence rate in the county of Anoia by the most recent update	Identify-now
LVT3	Number of cases in the county of Maresme on January 1 2022	Identify-date
LVT4	If the number of cases is higher in the town of Girona, on October 1 2020 or October 1 2021	Compare-date
LVT5	If the incidence rate is higher in the town of Girona or Tarragona on February 2 2022	Compare-place
DVT1	Number of confirmed cases in Catalunya by the most recent update	Identify-now
DVT2	Incidence rate in the town of Barcelona by the most recent update	Identify-now
DVT3	Incidence rate in the county of Bages on December 12 2021	Identify-date
DVT4	If the incidence rate is higher in Barcelona by the most recent update or the exact date one month ago (January 26 2022)	Compare-date
DVT5	If the number of cases is higher in the town of Lleida or Terrassa on January 1 2021	Compare-place

2.4. Apparatus

The apparatus employed to record eye movement is the eye-tracker SMI RED 250, which records at a frequency of 250 Hz, and is located in the eye-tracking laboratory of the Department of Geoinformatics of Palacký University Olomouc, Olomouc, Czechia. A camera with a microphone (Logitech C920) is also used, and SMI Experiment Center 3.7 is the software used to design the experiment, as well as to record it, combined with iView X. The stimuli are displayed on the participants' 24-inch screen in a Google Chrome browser.

2.5. Participants

Experiment I was performed between the December 18 2021 and December 27 2021 with 19 participants, 58% of whom have experience in the geospatial field. Experiment II was conducted with 20 participants between the March 9 2022 and March 18 2022. In this case, their level of expertise in cartography was evaluated as: 25% of the participants had no expertise in the cartographic field; 20% had either little, medium, or good expertise; and 15% were experts (according to their self-reporting). We recruited the participants through social media platforms, word-of-mouth, and snowball sampling. All the participants were undergraduate students, and they were not paid for their involvement in the experiment. Seven individuals participated in both experiments. We focused on the general public since it is an important target group of dashboards.

The Sample Size Calculator for Discovering Problems in a User Interface (Lewis, 2001; Sauro, 2023) was used to identify the ideal number of participants for both studies. From the sample data, Calculator estimates the problem occurrence (p) using the Good-Turing and Normalization procedure devised by Turner et al. (2006). In both studies, an estimation was made regarding how many respondents would be appropriate to detect at least 85% of the problems encountered. For experiment I, considering the 27 problems encountered, the result is 13 participants. For experiment II, the problems encountered are eleven, and the result is 12 participants. In both cases, the number of participants is higher than recommended.

2.6. Data pre-processing and analyses

The I-DT algorithm is used to identify the fixations and the saccades as it is appropriate for data measured at 250 Hz and below (Holmqvist et al., 2011). Two thresholds are applied: one for the minimum fixation duration and one for the maximum dispersion. The optimal fixation detection settings for cartographic research in SMI BeGaze 3.7 are 80 milliseconds for minimum fixation duration, and 50 pixels for maximum dispersion, according to Popelka (2014).

The recorded eye movement and screen monitoring need to be processed in such a way that allows the data to be separated into different individual tasks. SMI BeGaze 3.7 enables the segmentation of video recordings into *custom trials*. These are used to analyze eye-tracking metrics in the general context of individual tasks. For detailed participant behavior while using a dashboard, we establish areas of interest (AOIs).

2.6.1. Experiment I

Several methods are used to analyze the processed data. The correctness rate of the task solving evaluates the effectiveness, whether the participant can reach the targets (answer correctly) or not. The trial duration, or the time needed to solve the tasks, visualized in boxplots, evaluates the participant's efficiency, how quickly they solve the tasks (Rubin & Chisnell, 2008). A longer trial duration means the user interface has problems or the complexity of the task is high (Goldberg & Kotval, 1999). For the statistical analysis, the normal distribution of each value can be analyzed using the Shapiro–Wilk test. Most values do not have a normal distribution. A non-parametric Nemenyi's all-pairs rank comparison test (Pohlert, 2023) is performed.

The AOIs show the fixation time of each participant on different elements of the dashboard per task, enabling us to see their distribution of attention when interacting with the interfaces. Each dashboard has different elements, so the AOIs will be different in each case and will present a different layout (Figure 4).

The experiment also includes an interview that is used to measure each participant's satisfaction and perception of the elements of the interface, and a questionnaire designed to assess user satisfaction regarding their confidence when solving the tasks and testing the dashboards' usability. Participants were also asked whether they found the dashboards to be aesthetically pleasant and whether their academic/professional background was in the geospatial field.

2.6.2. Experiment II

The second experiment uses the same analysis methods as experiment I, adding the time to the first fixation, which indicates the average duration a respondent takes, and how long all respondents take on average, to target the first fixation into an AOI (SensoMotoricInstruments, 2010). In this case, the distribution of attention to the AOIs is visualized with sequence charts for each task; and the two dashboards have the same elements organized in different layouts (Figure 5). The search bar on the top of the LV and the DV's blocks of text are the only two elements unique to each dashboard.

For this experiment, the interview enables a comparison between the same functionalities displayed differently on each dashboard, and the questionnaire identifies the level of expertise in the field of cartography of the users. With these methods, conclusions regarding this experiment can be elaborated on, and the insights from experiment I can be confirmed or rejected in order for us to reach the final conclusions.

3. Results

3.1. Experiment I

3.1.1. Correctness

The results relating to the correctness of the answers show that D1 has the lowest success rate, with no one able to solve T3 (see Figure 6). This relates to the necessity of accessing the tabs to change the administrative levels of the countries and the periods (weekly or daily), which proved not to be intuitive for the participants, who could not find how to assess the task. D2 has the highest percentage of

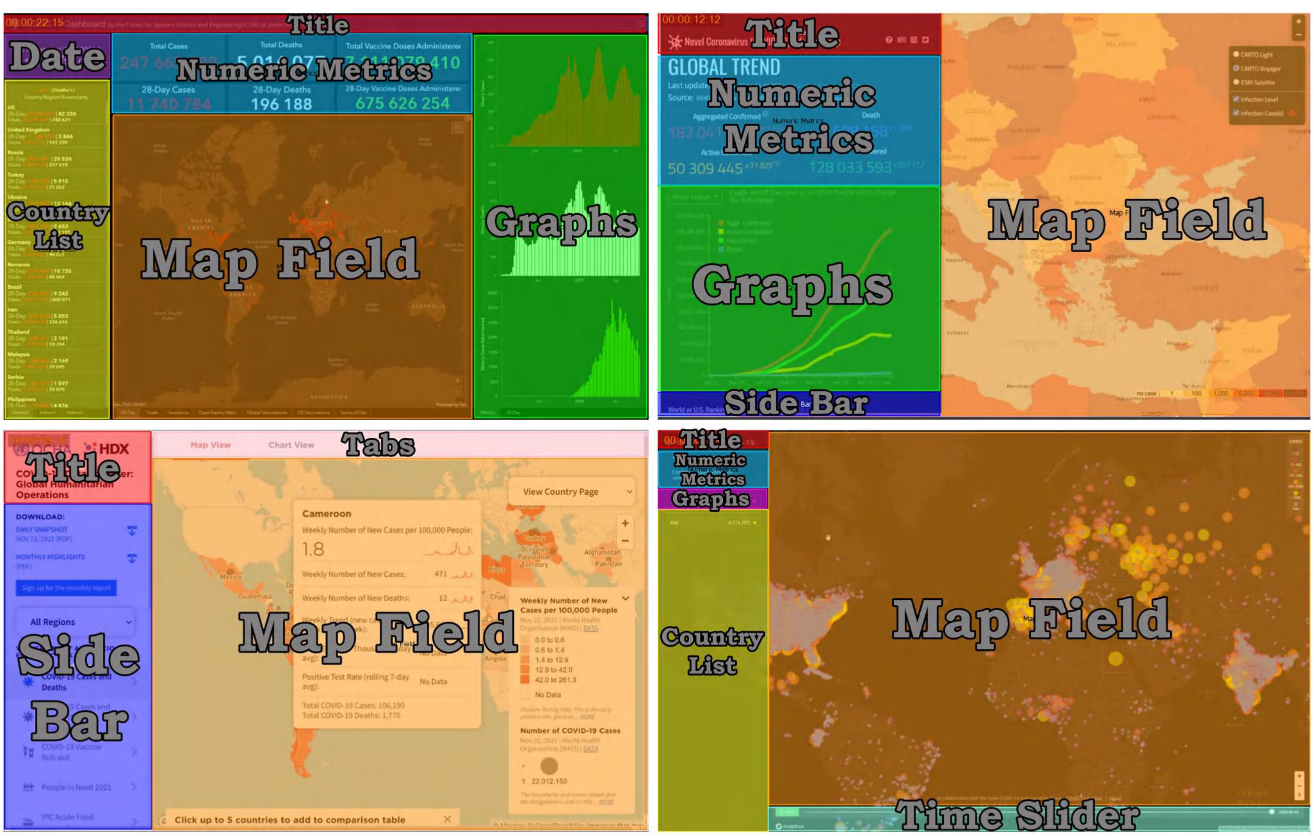


Figure 4. Areas of interest of D1, D2, D3, and D4.

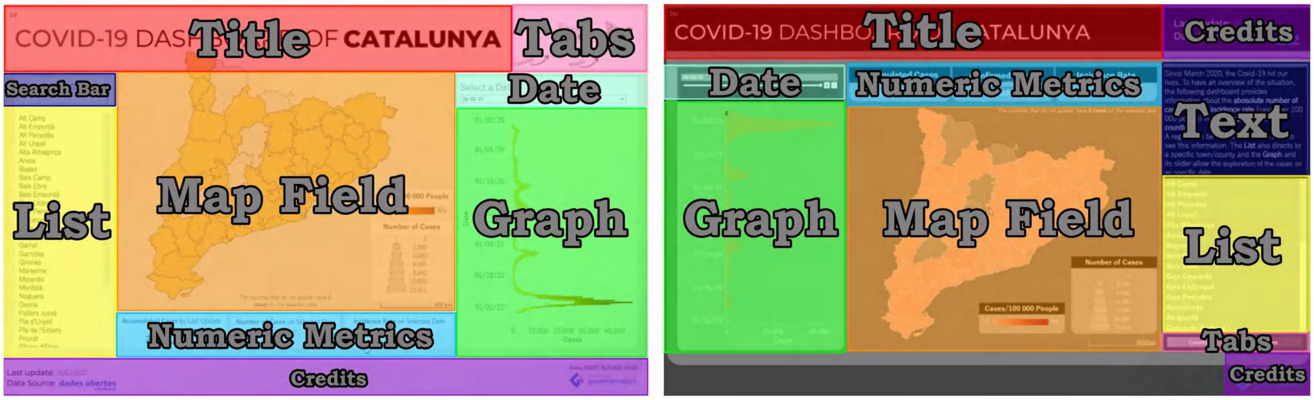


Figure 5. Areas of interest of LV and DV.

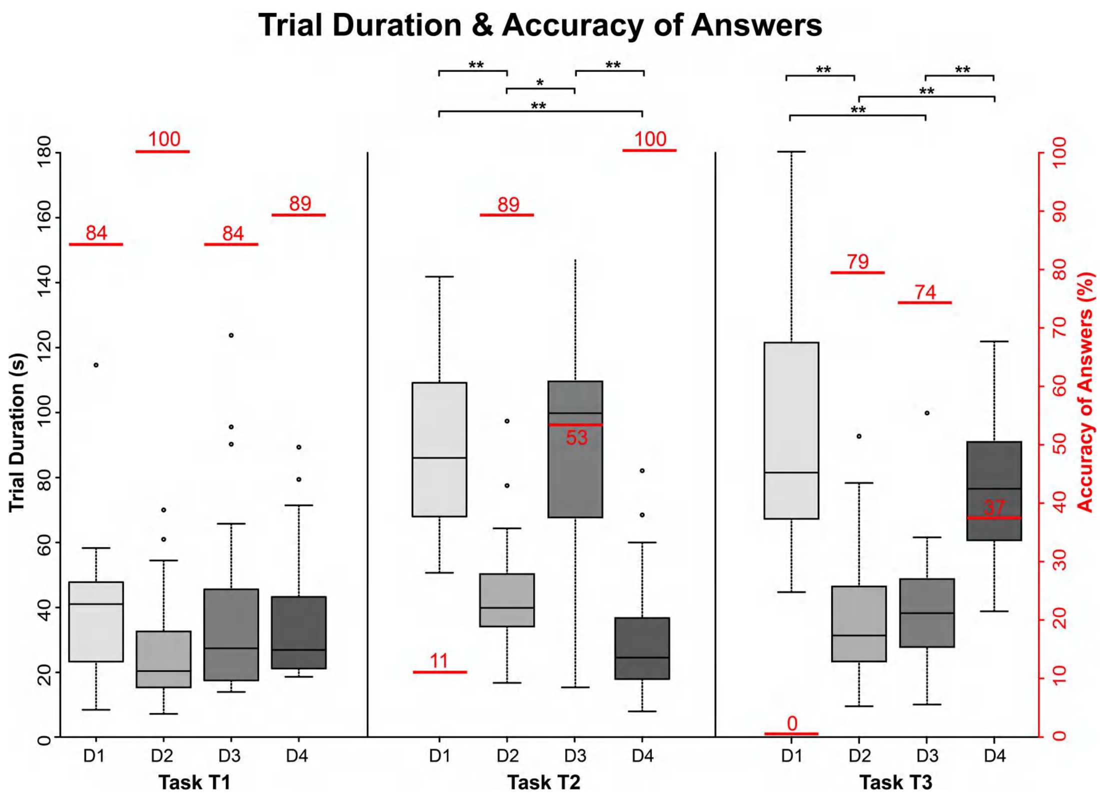


Figure 6. Trial duration and correct answers of each task per dashboard in experiment I.

correct answers, while for D3, more than half the participants correctly answered each question; and D4T1 and D4T2 have an almost perfect success rate (89% and 100%), but D4T3 has only a 37% success rate since it refers to the functionality of the dashboard, which is not intuitive at first sight; the time-slider.

3.1.2. Trial duration

From T1 to T3 on each dashboard, the difficulty tends to increase, as does the number of incorrect answers in most cases, as previously explained. Considering this, the trial's duration would also be expected to be longer, but that is not always the case.

The tasks are analyzed individually, and the statistically significant differences between T1, T2, and T3 along the dashboards are identified in Figure 6, together with the correctness of the answers. For T1, it is visible that D1 has the highest median, while D2 has the lowest, but there is no statistically significant difference. For T2, solving tasks above D1 and D3 took longer and was harder than D2 and D4. Finally, T3 shows that the most problematic dashboard was D1, where the task-solving took the longest amount of time, and no one answered correctly. In contrast, answers to D2 and D3 were relatively accurate and needed smaller amounts of time.

3.1.3. Distribution of attention to AOIs

Attention paid to each AOI can be described by dwell time value (Table 3). The map field is very predominant in the three D1 tasks. Participants spent in the map on average 40.7% of fixation time. Since T2 and T3 require the use of the graphs, the time spent focusing on this element was also very high. The country list was also visualized frequently since the tasks included looking for a specific region. On the other hand, some participants focused on the numeric metrics during T1, but these were not fixated by the others.

A similar pattern is observed in D2. The map field and the graphs require a lot of fixations, especially in T2 where the questions require participants to check the graphs for answers. They spent 59.0% of fixation time on the graph AOI. Instead of the country list, the sidebar was often checked in T2 and T3 because it contains large text with explanations and numeric metrics and, unlike D1, there are no pop-ups, and the numeric metrics change when different countries are selected to provide information regarding COVID-19 cases. As previously mentioned, it is hard to find a specific date on the graph because it uses a hovering system, which could have an impact on the trial's duration.

D3 has more fixation time spent on the map field (on average 72.7% of fixation time). This is explained by the fact that the graphs are also included in the same area (when switching to viewing the charts) as the pop-ups, the legend and the comparison table of different countries. Therefore, all the tasks involve focusing on this area. The sidebar also received some attention (on average 10.1%) from the participants as it has many tabs for checking different variables and changing parameters. In contrast, the tab for changing

Table 3. Dwell times for individual AOIs in experiment I.

Dwell time (%)		T1	T2	T3	Sum
D1	Country list	13.8	10.2	9.2	11.1
	Date	0.3	1.3	0.3	0.6
	Graphs	2.2	31.0	19.3	17.5
	Map field	54.6	22.0	45.5	40.7
	Numeric metrics	5.6	3.7	1.3	3.5
	Title	0.1	0.1	0.1	0.1
D2	Graphs	8.8	59.0	61.9	43.2
	Map field	30.0	6.4	10.8	15.7
	Numeric metrics	30.1	7.4	9.3	15.6
	Side bar	1.8	7.9	6.8	5.5
	Title	0.1	0.5	0.3	0.3
D3	Map field	74.1	79.2	64.8	72.7
	Side bar	11.2	4.1	15.0	10.1
	Tabs	0.6	0.5	0.8	0.6
	Title	0.2	0.2	0.3	0.2
D4	Country list	28.0	29.4	7.9	21.8
	Map field	37.9	30.9	47.7	38.8
	Numeric metrics	1.2	1.8	0.5	1.2
	Time slider	1.8	1.4	12.8	5.3
	Title	0.2	0.1	0.1	0.1

from the map to the charts did not attract their attention. Most participants did not notice it and could not interact with the graphs in order to answer T2.

Finally, the way users interact with D4 has its particularities. The map field again plays a significant role (on average, 38.8%), as does the countries list (on average, 21.8%). The added feature in this dashboard is the time-slider. T3 involves its usage, and it received many fixations during the search for a solution (12.8% of fixation time). During T1 and T2, the time-slider received only a minimum of fixations. It was similar for the numeric metrics, even though they consist of a static number of cases worldwide. Since there are no informative pop-ups, the users could have expected them to be dynamic.

3.1.4. Questionnaire and interview

The questionnaire and the interview provide information regarding the subjective opinion of the participants. First, they rated how difficult it was to solve the tasks of each dashboard, from 1 (very difficult) to 5 (very easy). The results (Figure 7) show that a higher number of participants rated D1 as the dashboard with the most difficult tasks, contrasting with the situation for D2. D3 tended to be easier, and D4's results are closer to the okay rate since T3 is not intuitive. The participants were also asked to rate the dashboards aesthetically, from 1 (very ugly) to 5 (very attractive). This question is subjective since, for example, some participants appreciate a dark background with colorful symbology (D1 and D4) while others prefer the simplicity of a clear background (D2 and D3).

Nevertheless, D2 shows the highest rates of aesthetic approval; D3 is closer to being considered very attractive, followed by D1 and D4. In general, lighter and simpler dashboards are rated aesthetically higher.

The interviews also show preferences and suggestions worth considering as part of the recommendations. These opinions are visualized in word clouds in Figure 8. Regarding D1, the highlights are in the negative aspects since almost half of the participants mentioned that there was no search option (47%), and the accessibility of the tabs used to change regions or time was not easily identifiable (42%). An excess of information and symbology (graduated symbols) were also mentioned as negative elements by a smaller percentage (21% in both cases). Some positive elements are the graphs, numeric metrics, and

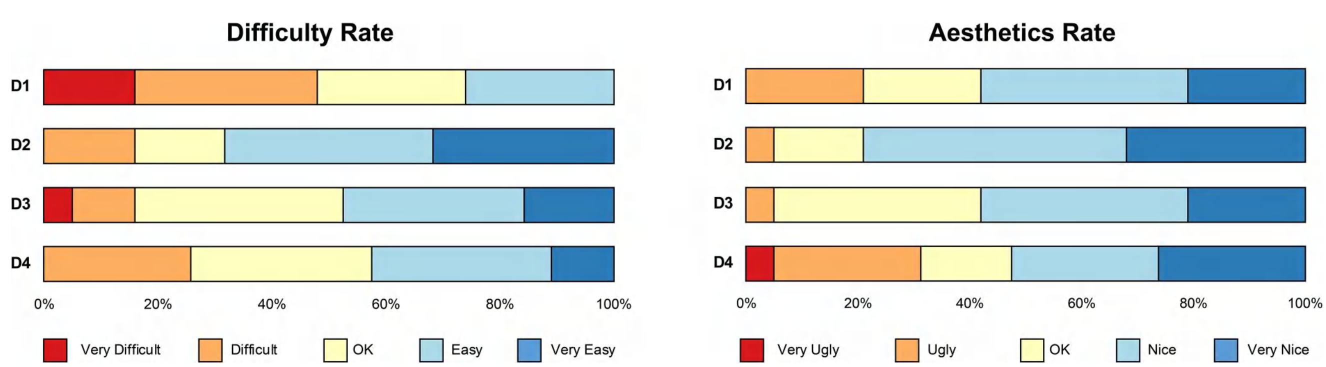


Figure 7. Difficulty and aesthetics rates in experiment I.

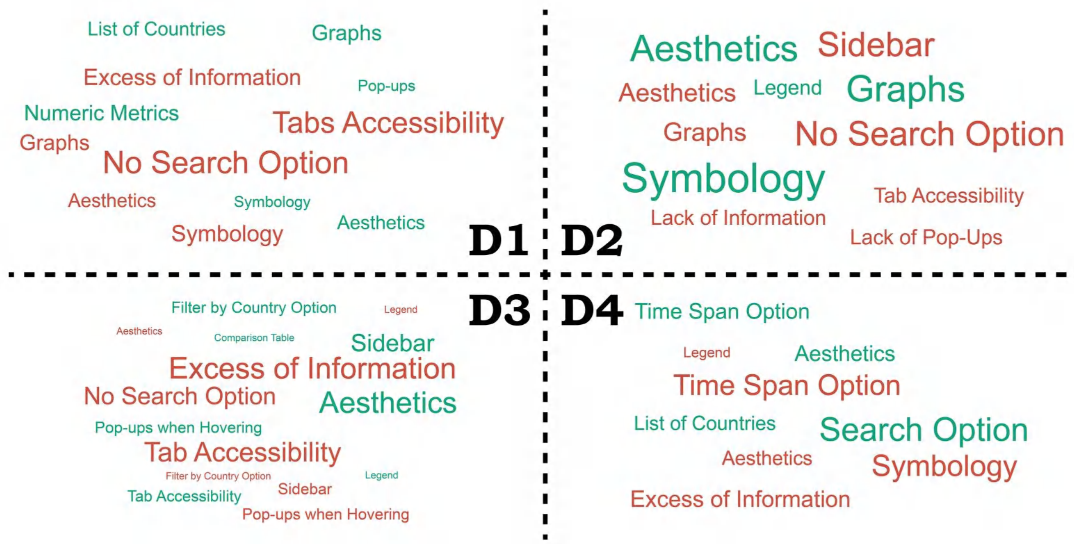


Figure 8. Participants' preferences for dashboards in experiment I visualized via word cloud. Green indicates positive feedback while red indicates negative feedback.

aesthetics, but they were mentioned by a tiny percentage of the participants (lower than 16%).

D2's positive elements were highlighted more. Almost half of the participants mentioned that the symbology (choropleth) was a positive asset (47%), as well as the graph and the aesthetics (32%). These were commented on negatively by only 11% of the participants, together with other assets with lower figures. Only the lack of a search option and the fact that the sidebar occupies too much space in the interface, and contains a lot of text information, were mentioned as negative elements by 26% of the participants.

The most positive aspects of D3 are its aesthetics, pointed out by 32% of the participants, and its sidebar, highlighted by 21%. The most negative aspects were the excess of information, according to 32% of the participants, and the not-so-easy tab accessibility for users to change the parameters, mentioned by 26%. Again 21% pointed to the lack of a search option.

Finally, the search option in D4, unlike the other dashboards, was much appreciated by 68% of the participants. The aesthetics were also positively mentioned in 21% of cases, as well as the time-slider option. Time-slider is tricky since almost half of the participants mentioned it as a negative element (47%), saying it was a good idea but not adequately implemented. The symbology (dots) was also seen as a non-positive element by half of the participants (53%), together with the excess of information and aesthetics by smaller numbers, 26% and 16%, respectively. Sixteen percent also highlighted the list of countries as a positive asset.

3.1.5. Insights from experiment I

The information obtained enables a summary of the results, with which to identify the problematic aspects of the dashboards and to make recommendations for an improved version.

Regarding correct answers, the results show that D1 is the hardest and D2 is the easiest. The results showing the trials' duration are the same, with T1 of each dashboard being done quicker than the last task, except in those cases where the previous questions enabled participants to be familiar with the procedure and answer quicker. Matching quantitative data from eye-tracking and qualitative data from the subjective answers in the interviews also show that D2 is seen as the easiest and D1 as the hardest. The aesthetics of light dashboards are preferred, and while D1 tends to have a higher percentage of negative aspects, D2 is the opposite.

This can be related to the number of elements available to interact with in order to find the information, and also to their accessibility. D1 has many elements, and this makes it hard to find specific information, in contrast to D2. D3's situation is similar to D1, but it appeared to be more intuitive to participants, with a higher correctness rate and a lower trial duration. D4 is similar to D2, but the time-slider is not user-friendly and decreases the correctness rate and increases the trial duration.

The field of study/work of the participants does not have a significant impact in the correctness rate, since the average of correct answers is 61% for participants from this domain and 67% for the ones who are not.

The AOIs show the importance of the map field in all cases and how the nature of the question can change which areas are consulted more. The graphs and the list of countries were often noticed and used to carry out the proposed exercises. The frequency of the numeric metrics varies, depending on whether these are interactive or static, making the static ones a less valuable resource. The specific elements of each dashboard, such as the search option and the time-slider in D4, are present in the participants' fixations because they are required for the tasks. The explanatory text in the sidebars can be a stumbling block rather than a helpful element.

According to the participants' subjective opinions, some negative elements that all dashboards had in common were the lack of a search option, the accessibility of the tabs and the excess of information. Good assets were related to symbology when the choropleth was used in the map section, as well as to the light aesthetics, the graphs, the interactive numeric metrics, and the sidebars that included lists of countries.

Therefore, the following recommendations have been formulated based on the quantitative and qualitative data results. A good dashboard should include light, simple aesthetics with a choropleth map, a country list sidebar with a search option, and numeric metrics that interact with the desired requested information (instead of pop-ups). When graphs and a time-slider showing temporal data are used, they should be user-friendly and clearly visible. A concise, clear title is necessary, and it is important to avoid large blocks of explanatory text. Alternatively, the less appreciated elements were darker colors, graduated symbols or dots, static numeric metrics and hard-to-use/find functionalities.

3.2. Dashboard development

Two dashboards, LV and DV, have been designed according to the results of experiment I, which relate to:

- Good assets: light aesthetics with a choropleth map, a country list sidebar with a search option, numeric metrics that interact with the desired requested information, clearly visible and user-friendly graphs and tabs, an option to choose a date, and a clear title.
- Assets to avoid: darker colors and graduated symbols or dots for the cartographic symbology, static numeric metrics, clickable elements that display pop-ups, hard to use/find functionalities in the case of graphs and tabs, as well as large blocks of explanatory text.

Each dashboard has a title, imprint, a map field, a list of countries, numeric metrics, a graph and tabs to change between administrative regions (counties and towns). To communicate the information in an accurate manner, and following cartographic rules, the map fields display a choropleth map that shows the incidence rate (cases per 100,000 inhabitants) as well as graduated symbols that display the absolute number of cases. The elements that differ from one interface to another are the following:

- The LV design incorporates a light aesthetic, in contrast to the DV's darker theme.
- In the LV, the title is more prominent, while the DV places greater emphasis on the imprint.
- An explanatory block of text is present in the DV, while the LV includes a search option within the list of countries/towns.
- The LV provides interactive numeric metrics that adapt to region/time parameters, whereas the DV shows static values from the most recent update. Users must click on symbols in DV to view more details in a pop-up.
- For displaying temporal time, the LV employs a drop-down menu that also affects the map and the graph (as well as numeric metrics), while the DV uses a time-slider.
- The tabs for switching between administrative regions are larger and situated on the top right of the screen in LV, while in DV, they are smaller and located on the bottom right, under the list.

In general, the elements in the LV are intended to be interactively connected and show temporal and regional data. DV shows the same, but the behavior of the elements is static, and the elements need to be clicked on to display the information, which tends to make it a slighter less user-friendly interface, with other assets considered negative, according to experiment I.

3.3. Experiment II

3.3.1. Entry time

The results regarding time taken to the first fixation metric (in ms) during the free exploration time are shown in Figure 9. They are similar for both dashboards since the map field is the first to be perceived. The title and the list are spotted very quickly with the LV, but it takes much longer with the DV because LV's are bigger, and the list is directly below the title. The opposite situation occurs with the numeric metrics, which can be related to the fact that they are located on the top of the map field for the DV and at the bottom for the LV, thus being more visible in the first case. The graph is one of the first elements perceived and shows similar entry times for both versions. The tabs are a

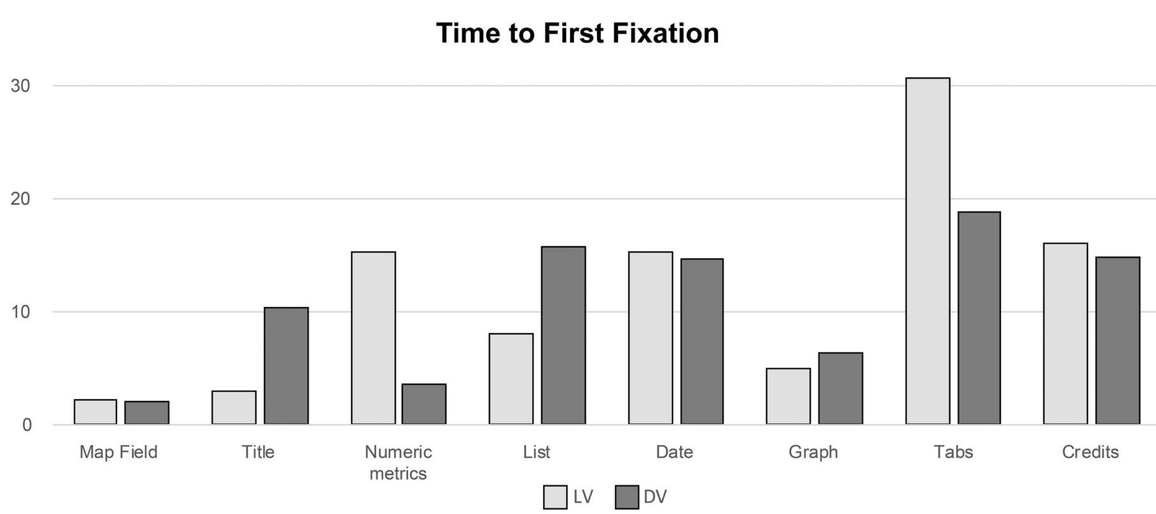


Figure 9. Average time to first fixation in each AOI during the free exploration time in experiment II.

curious case: those from LV are bigger and on the top right, while the DV's are smaller and on the bottom right below the list, yet the DV's are perceived in almost half the time as the LV's.

3.3.2. Correctness

Even with a minute to explore the dashboards freely and get familiar with their functionalities, the users expected similar outcomes in LV and DV and gave wrong answers. The results show big differences in the rates of success in task solving with each dashboard (Figure 10). T1 of each dashboard was usually correctly answered (85% for both versions). Regarding the other questions, the differences between both versions are very significant: T2 and T3 have a 90% and 55% success rate, respectively. In the case of LV, the success rate for T3 is lower because it involves more difficulty, while T2 and T3 success rates for DV are 10% and 5%. This is related to the fact that users expected the numeric metrics to change (as happens in the LV), but these are static, and to find the answer, they need to display a pop-up. Finally, for T4 and T5, the LV's success rate is 75% and 80%, and the DV's is 55% in both cases. For DV, these rates are higher than the previous two, even though they are more difficult, and this is because the participants became familiar with the interface during the experiment. Overall, LV shows a higher success rate, indicating that its interface is more intuitive for its users.

3.3.3. Trial duration

The median time required for T1 of LV is higher than for T2 (Figure 10). This is related to the fact that it is the first question, and the participants were getting adapted to the task-solving process, which is not the same as the DV. Generally, tasks in LV take longer to solve than in DV, but the time difference is minimal. Alternatively, the tasks that take longer to solve in DV take almost double the time they take in LV: from 26 to 47 s in T2 and from 44 s to one minute, 36 seconds in T5.

Regarding the statistically significant differences, it is clear that T1's median of time needed to solve the task in LV is higher. Again, this is related to it being the first question when the participants were adapting to the task-solving process, so it is not necessarily related to the question's difficulty. Their statistical significance also differs. In the case of T2, there was the highest difference in the accuracy of answers. For DV, only one participant answered correctly, while in the case of LV, the accuracy of answers was 90%. For T3, the median of trial duration was significantly higher in the case of DV and the accuracy of answers was low for DV. In T4, no statistically significant differences were observed. For T5, the two dashboards differ statistically. The difference between the median and the time needed to solve the last task was higher for DV due to its difficulty and the lack of search tools that would have made the processes shorter and more straightforward.

3.3.4. Distribution of attention to AOIs

The findings related to the attention distribution in the LV experiment show that, while not crucial to solving the tasks, the graph (green) played a very important role due to its function as a date search method and a drop-down date selector. In most cases, the list (yellow) received a significant number of fixations, while the search bar (purple) was only perceived by some participants. The map field (orange) and numeric metrics (light blue) were present in most tasks, as

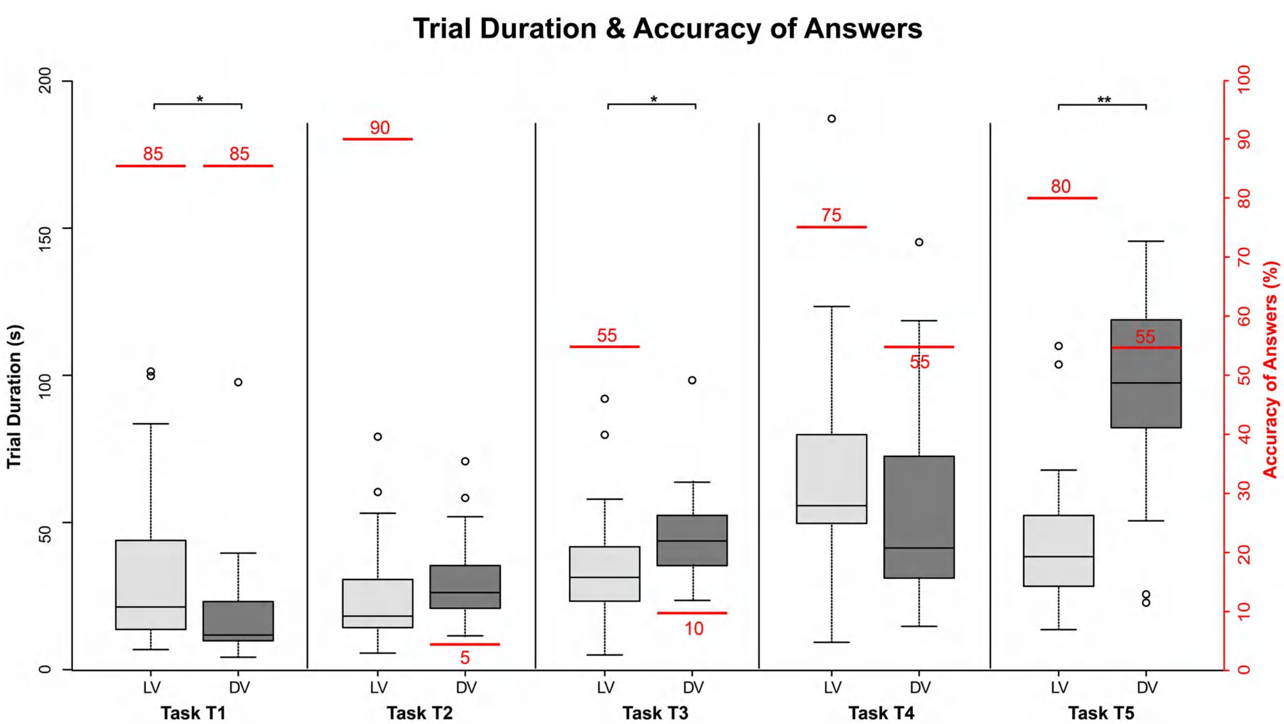


Figure 10. Trial duration and accuracy of answers of each task per dashboard in experiment II.

they dynamically adjusted according to the selected region and date.

Despite their importance, the tabs for changing the administrative level (pink) were overlooked by the participants who, on average, spent only 0.36% of the time on observation. The tabs limited usage led to confusion and a low success rate. Figure 11 gives an example of a sequence chart for LVT4 that shows the relative distribution of attention paid to AOIs during tasks that involved the participants comparing a numeric metric from a specific region over two periods of time. The graph and the list of AOIs were fixated extensively, while the tabs were barely noticed.

In the DV, the average dwell time on the tabs (pink) was more than four times higher (1.48% of observation time) than with LV, despite the tabs being smaller. Surprisingly, the attention paid to the list (yellow) was not as high as expected, considering the absence of a search option in this case, and participants needed to find a specific town or county in the list. Additionally, the text block (black) did not receive considerable attention from participants, indicating a lack of interest in the information it provided, with only 0.1% of observation time in the case of DVT4, which was the task used as an example and depicted in Figure 12. Alternatively, the date selector (time-slider, dark blue) required a lot of fixation time (19.9%) due to the difficulty participants had in selecting specific dates.

In both dashboards, the distribution of attention to AOIs have several similarities, primarily the significant use of the graph and the map field, which users consulted for metrics. As for the numeric metrics, in the DV experiment, the attention time decreased when participants realized that they remained static and did not change when interacting with other elements of the dashboard.

However, one noticeable difference was the higher number of fixations on the list in the LV experiment, which included a search option designed to enhance the search process for users. While its implementation improved the accuracy of the answers, it did not reduce the time required. Another difference, in this case expected, was the increased attention paid to the date selection in the DV experiment, as the time-slider proved to be less efficient than a drop-down menu when choosing a specific date.

3.3.5. Questionnaire and interview

To obtain information regarding the participants' subjective opinions, they were asked again how difficult it was to solve

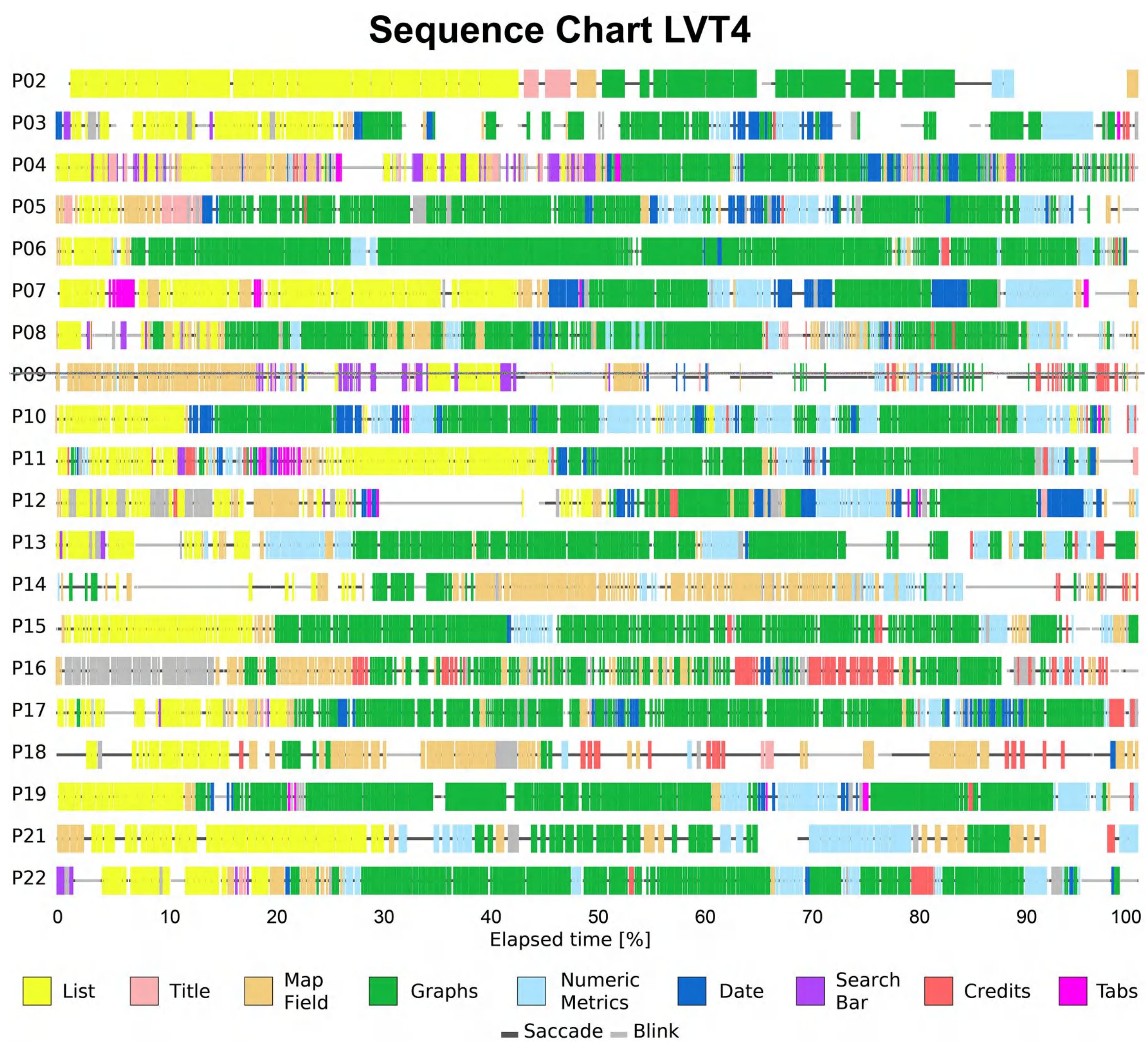


Figure 11. Sequence chart for LVT4.

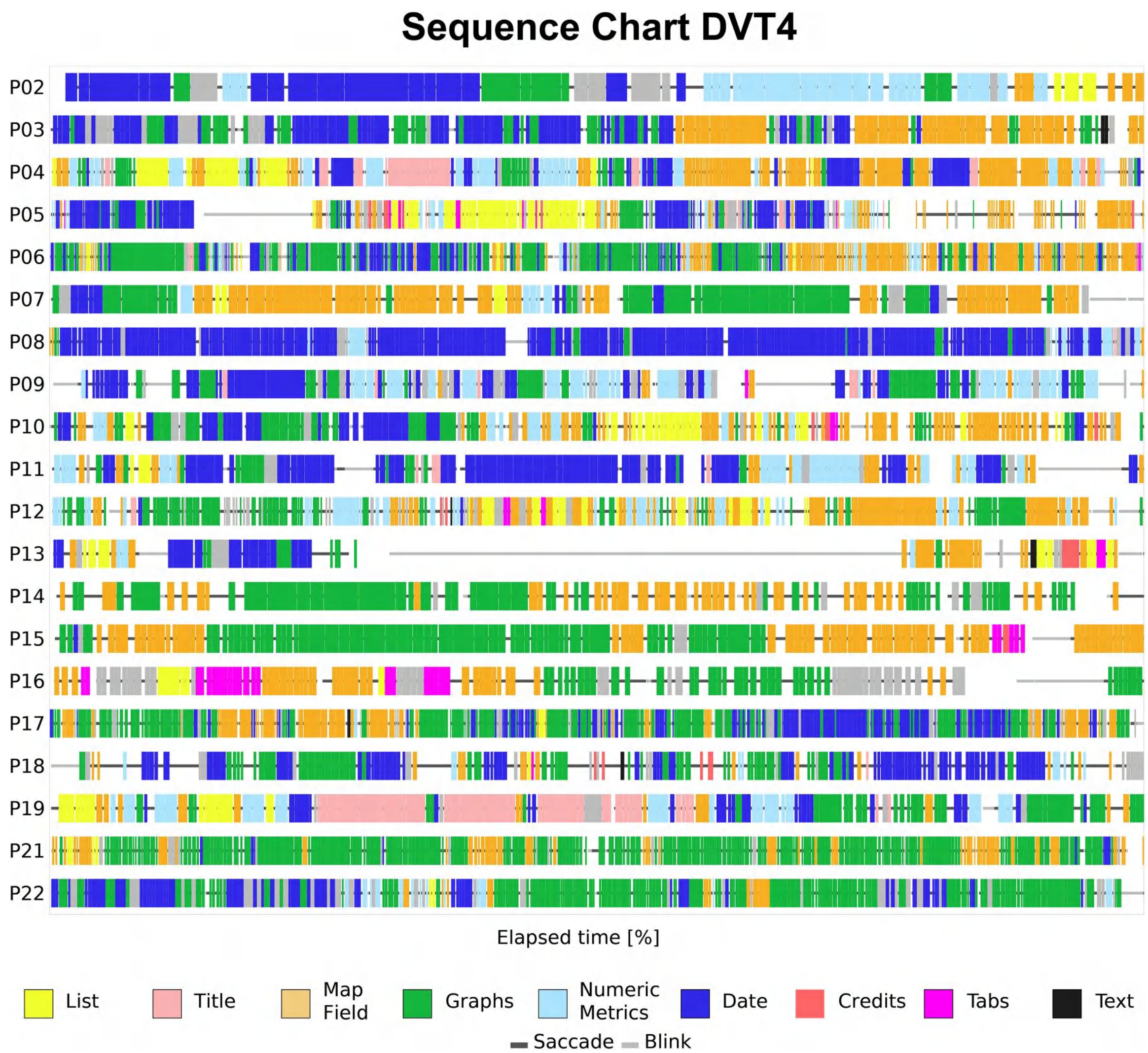


Figure 12. Sequence chart for DVT4.

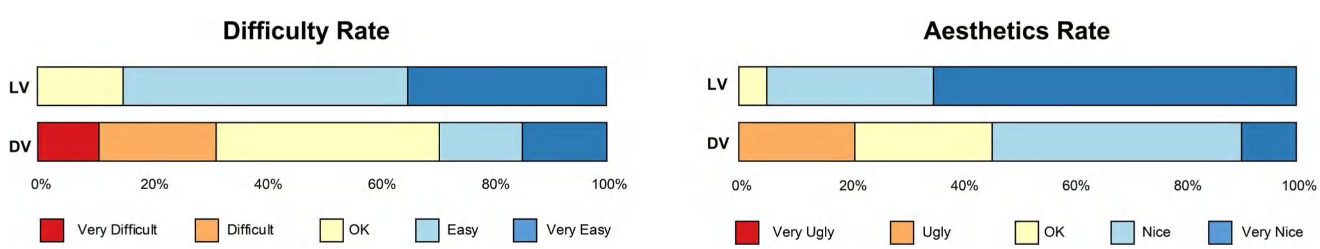


Figure 13. Difficulty and aesthetics rates in experiment II.

the tasks of each dashboard, and their opinions of the aesthetics (Figure 13).

The LV is generally considered easier, whereas the DV has a broader range of opinions. A significant number of the participants considered DV okay (40%). The number of participants that consider DV difficult was higher than the number that considered it easy. Regarding aesthetics, the LV is mostly considered to be attractive, while only 10% thought the DV was attractive, and a fifth of the participants thought it was ugly.

The participants were asked to freely speak their minds regarding their experience during the task-solving process. Considering their opinions, it is possible to compare the same functionality presented differently on both dashboards (Figure 14). Overall, most functionalities of the LV are widely preferred over those of the DV. Rather than the time-slider, the favorite element was the drop-down date option: 85% of the participants said the LV's option was better. The fact that the numeric metrics changed when users interacted with the dashboard and the light aesthetics were preferable meant that the LV was the choice of more than half of the participants (60% for LV and 55% for DV).

Fifteen percent of the users favored dark aesthetics and 10% supported the clicking option to display the values. Regarding the list, 37% mentioned it as a nicer element and more user-friendly in the LV because it was complemented

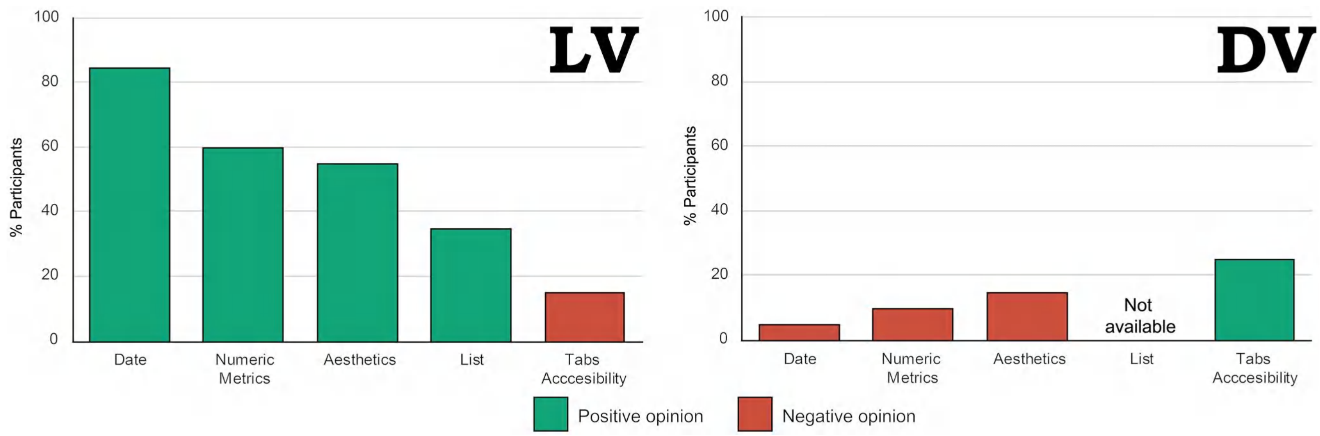


Figure 14. Comparison between LV and DV functionalities.

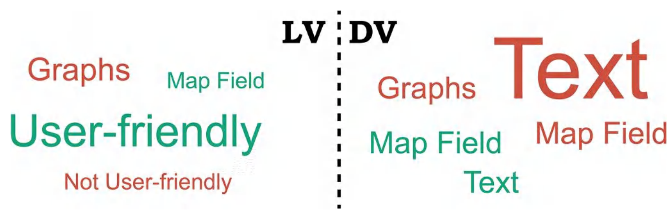


Figure 15. Participants' preferences for dashboards in experiment II visualized via word cloud. Green indicates positive feedback while red indicates negative feedback.

by a search option that the DV did not have. Finally, the DV's tab accessibility was considered better in the opinion of 25% of the participants.

Also, the specific functionalities of each dashboard were mentioned by the participants as good/bad assets (Figure 15). Forty percent of the participants said the LV was user-friendly, while only 5% said the same of the DV. The map field and the graphs are commented on by between 5 and 15% of the participants with the map field mainly seen as a positive asset and the graphs seen as negative assets. Finally, the text in the DV, which is not present in the LV, was considered unnecessary in 20% of cases.

3.3.6. Insights from experiment II

The correctness rate shows significantly better results for the LV, implying that it is easier for the participants to answer correctly using the LV, and the trial duration results are related to this. These quantitative data match the qualitative data in most cases. The subjective opinion of the participants regarding the aesthetics and the difficulty of the dashboards was that for both metrics the LV was more attractive and easier, to a significantly major extent, showing that it is more user-friendly. When the participants freely gave their opinions, comparing the two dashboards, the functionalities of the LV were widely preferred over those of the DV. The only element of the DV that participants preferred over the LV was the tabs' accessibility, proved by the fact that more participants made use of them than with the LV, but the block of text was considered unnecessary in 20% of cases.

In this case, the level of expertise in the field of cartography does not alter the correctness rate either. The participants with an expertise level of 5 (the highest) present an average of 53% correct answers. After them, ordered from superior to lower (4 to 1), the averages are 68%, 53%, 70%, and 54%, respectively, not showing a higher average related to a major level of expertise.

The entry times show that the size and position of the various elements of a dashboard can play a role in attracting a user's attention. For example, the map field is the dominant element of the dashboard, so it is the first element to attract attention, while the credits are the last. Because of their size, prominent elements such as graphs and lists are perceived before others, such as search bars and date selectors. Alternatively, some elements, such as the numeric metrics, are noticed earlier than others, not only when they are dynamic rather than static, but also depending on whether they are at layout. The tabs were more likely to be perceived if they were near the element they referred to (in this case, the list of regions), rather than because of their size or their predominant position in a layout.

The distribution of attention to the AOIs show that, regarding the numeric metrics, the user's attention decreases through the course of the experiment, from the early tasks to the later tasks in the DV as the participants realize it is a static element that does not interact. As expected, the text in DV does not attract a lot of attention but, unexpectedly, the tabs in DV attract more attention than in LV, which leads people to change the administrative level in most cases, giving a higher correctness rate. The graphs, even though they are not essential to solving the tasks, attract many fixations, especially in the case of LV, because they can be used to select a date. Also, the LV allows the date to be entered when searching for it, but not many participants realized that and wasted time scrolling. Alternatively, the DV's date selector (time-slider) required a lot of attention because it was challenging to select a date with precision. This took a lot of time, as with the list of towns/counties where the time to find a solution increased, when users were comparing two regions, because of the lack of a search option.

According to the new outcomes that both the quantitative and the qualitative results have provided, dashboards should contain a map field that follows cartographic rules, as well as light aesthetics, a sidebar list with a search option, numeric metrics that interact with the desired requested

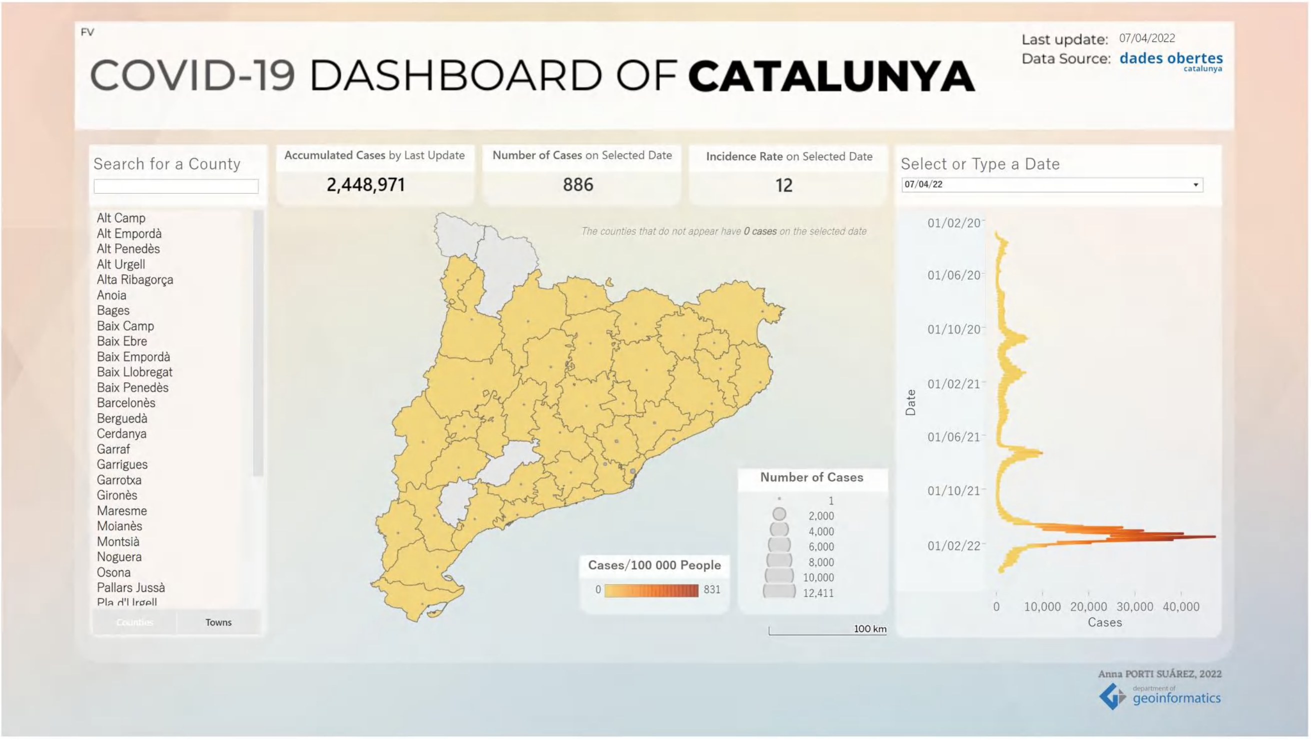


Figure 16. Final version of the COVID-19 dashboard of Catalunya.

information (at the top of the layout), tabs near the elements they refer to, graphs with a date selector as a drop-down menu or typing box, and a clear title. The time-slider as a date selector and blocks of explanatory text are roadblocks.

Figure 16 shows the final version of the dashboard, and it combines the positive elements from the two dashboards used as stimuli in the second experiment, compounding an interface with all the recommended elements.

4. Discussion

Two user experiments were performed to evaluate the design aspects of COVID-19 dashboards. The first evaluated existing solutions, while in the second, self-developed dashboards served as the stimuli. We chose this topic because the public has been accustomed to work with these dashboards in recent years. There exists multiple dashboards on single topic and there is a wealth of freely available data. However, the findings from performed experiments can be applied to other topics with a spatial component.

The selection of dashboards for evaluation in experiment I was not systematic and might not be representative of the population of dashboards. A similar study evaluating interactive visualizations in COVID-19 dashboards was performed by Fan et al. (2023). In that case, the authors investigated different dashboards carrying out a Google search of keywords connected with COVID-19 and visualization. From the 57 identified dashboards, they selected five different visualization methods. They stated that these examples were selected "to cover common visualization types". A similar approach was used in our case, where we aimed to select dashboards containing maps with different functionalities. One drawback is that the results lead to the comparison of questions that were designed with different objectives. Alternatively, this enables the identification of valuable insights regarding user behavior when interacting with interfaces of varying complexity and information load.

In another study focused on the analysis of COVID-19 dashboards, Li et al. (2022) selected five dashboards and analyzed the people's needs on Twitter. The authors stated that external validity might be threatened because of the selection of these dashboards and that could be mitigated by obtaining more dashboards. However, using more than four dashboards could overload the participants.

The participant selection is not methodical either, since COVID-19 dashboard user might range from experts to the general audience. Nevertheless, various fields of study and levels of knowledge related to cartography have been involved. Information needs to be understandable for both decision makers and common users seeking for information. Such study can dig in how developers should enhance the design of user interfaces when the user target is not defined, repeating aspects mentioned in the literature review such as classification and simplicity (Ching-Yi et al., 2018) or the need of avoiding unnecessary elements (Dowding and Merrill 2018; Monkman et al., 2021).

While dashboards are known to be tools often used by policymakers, health officials, and researchers, the onset of the COVID-19 pandemic has broadened their user base to include members of the community. Considering the diversity and limited number of these experts, grouping them into a consistent study sample presents challenges. Therefore, our research involved the general public as participants in our experiments. This study, while offering valuable insights, presents several limitations. First, the limited number of participants may not provide a comprehensive representation of the broader user population. Further, specific user groups, such as health officials, were not included. In experiment 1, the dashboards were not systematically selected, leading to complications in comparisons given the disparities in sizes of AOIs. Recognizing these shortcomings, future improvements would include a more systematic selection of dashboards and recruiting specific user groups for participation.

The overarching nature of this research is formative, focusing not directly on comparing the efficacy or comprehensibility of selected dashboards, but rather on identifying successful and problematic design elements. Nevertheless, we conducted an analysis of the accuracy of responses and trial duration, employing these metrics solely to facilitate the identification of efficacious and problematic design elements. These elements were subsequently incorporated in the construction of our original prototypes (LV and DV). Due to the context of our study, it is crucial to highlight that the numerical data related to accuracy or trial duration are not intended for a comparative analysis between experiment I and experiment II.

One conclusion in both experiments shows that, when interacting with dashboards, people do not spend time reading the explanatory texts that might introduce them to the functionalities of the interface. The question is whether users react in this way in web interfaces or only in these experiments. The short fixation or attention time spent on the text area may not necessarily mean that users did not read the explanatory text. It could be that they were scanning for useful information (Manhartsberger & Zellhofer, 2005). By tracking a gaze plot, Manhartsberger and Zellhofer could identify which text a user read to fulfill specific tasks. In the same study, they concluded that users had problems solving specific tasks when the reading text gave instructions of functionality that were not placed next to the text, causing usability problems, and they recommended that the information a text relates to is placed next to it, such as the previously mentioned issue with the tabs in this study, referring to the law of proximity (Gestalt theory) as applied to interactive web interfaces (Graham, 2008).

A practical example is visible when working with different spatial scales, a potential challenge of this work. The first experiment uses the general country level and some dashboards that allow changing to smaller administrative areas in an unclear way that led to confusion and unsuccessful results. Making this instruction clear to the user should be a key element to consider when designing interfaces. For this reason, one goal of the design of the second experiment dashboards was to check how the participants would be able to find the possibility of changing an administrative level, despite their familiarity with the spatial scales, by placing this option close to the element they refer to.

Cartwright et al. (2001) emphasized that beyond the functionality of the interface, the aesthetic aspect of design plays a pivotal role. They suggested that attractive design elements could potentially boost the early adoption and acceptance of a product. In this work, the preferred light version of the dashboard is also related to more successful task-solving and lower trial duration. This contrasts with Coltekin et al. (2009), whose study concludes that better aesthetics are not related to more accurate answers when performing tasks, meaning a low correctness rate on an interface that is considered aesthetically attractive also translates to usability problems.

In experiment II, the order of appearance of the dashboards was not randomized for the performance of the tasks, with the LV appearing first. Due to the users' lack of familiarity with the process, they made mistakes when answering the first (and easiest) question (LVT1). This shows the importance of randomization when performing user testing experiments.

The qualitative results of this study include suggestions from the participants about how to improve dashboard interfaces. There are two important aspects to highlight for an improvement in the communication of displayed information: (1) add a button to clear the selected information instead of having to manually unselect it. (2) Use a logarithmic scale in the graphs to spread the values around the graphic and provide better visualization when the absolute values are too low and homogeneous and are not visible. However, it is necessary to properly describe the visualization, since users might have issues with understanding (Romano et al., 2020).

Some of the design aspects that follow the suggestions made by Monkman et al. (2021) are essential factors to consider when analyzing the AOIs and the correctness/time trial of the answers, and were an improvement from the existing dashboards to the self-developed ones. The clearest example of this is that the labels describing the phenomena they refer to accurately guide users through the interface, enabling more accurate answers. Alternatively, the indication of the currency of the data did not attract the participants' attention, as they could interact with other functions that explored time variables. Finally, there was no way to restore a dashboard to a default view, and this led to the use of refresh in most cases, which increased the time taken to interact with the dashboards. So, this is a functionality that needs to be taken into account in future implementations.

Together with other user evaluation results, these insights provide recommendations for better interface design and they suggest improvements. Other variables could be investigated that are related to the dashboard's functionalities, such as appearance, size, placement in a layout, and the users' behavior. This work focused on dashboards covering COVID-19 metrics, but it could also apply to assorted topics to observe whether the same patterns are identified within different interfaces.

5. Conclusions

The aim of the article was to perform user evaluation of dashboards containing geospatial information regarding the COVID-19 topic, and to establish three partial goals, using mixed-research methods.

The *first goal* was to obtain insights into user interaction with existing dashboard functionalities, to identify problematic elements and to recommend a more user-friendly interface design that appropriately communicates information. These improvements relate to a map field's symbology as a choropleth, a sidebar with a list of the regions displayed and a search option, numeric metrics that interact with the desired requested information (instead of pop-ups), noticeable tabs for changing parameters, small credits, and a big title. Temporal data should be presented as graphs, including a date selector for choosing a period to visualize. Light aesthetics are preferred to dark, and large blocks of explanatory text and excess information do not appeal to users.

The *second goal* consists of elaborating two self-developed dashboards with the design and placement of the elements according to the insights obtained previously, and to be evaluated in the *third goal*, except for the spatial information, which is displayed following cartographic rules showing a choropleth map with relative values, and graduated symbols for the absolute values, in order to communicate the information accurately. The results of experiment II show that most of the insights are repeated but they also provide new outcomes; the numeric metrics must not only be interactive but should be placed at the top of the interface, the preferred date selector option should be through a drop-down menu or typing, and the use of the tabs should not depend on their size and visibility, but on the fact that they are close to the element they refer to, in this case, administrative levels.

To summarize, this process involved the characterization of analysis methods in cognitive cartography, both quantitative and qualitative, focusing on eye-tracking technologies. We obtained insights into user interaction with COVID-19 dashboards, into the appearance of their functionalities, and into their role in communicating and transferring information correctly. Finally, we formulated recommendations for improving dashboard interface design.

Acknowledgements

The article was supported by the project of the Czech Science Foundation "Identification of Barriers in the Process of Communication of Spatial Socio-Demographic Information" (Grant No. 23-06187S).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The article was supported by the project of the Czech Science Foundation "Identification of barriers in the process of communication of spatial socio-demographic information" (Grant No. 23-06187S).

ORCID

Stanislav Popelka http://orcid.org/0000-0003-2127-2614

References

Bach, B., Freeman, E., Abdul-Rahman, A., Turkay, C., Khan, S., Fan, Y., & Chen, M. (2022). *Dashboard design patterns*. arXiv Preprint arXiv:2205.00757. https://doi.org/10.48550/ARXIV.2205.00757

Carayon, P., & Hoonakker, P. (2019). Human factors and usability for health information technology: Old and new challenges. *Yearbook of Medical Informatics*, *28*(1), 71–77. https://doi.org/10.1055/s-0039-1677907

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, *4*(2), 75–100. https://doi.org/10.1037/1076-898X.4.2.75

Cartwright, W., Crampton, J., Gartner, G., Miller, S., Mitchell, K., Siekierska, E., & Wood, J. (2001). Geospatial information visualization user interface issues. *Cartography and Geographic Information Science*, *28*(1), 45–60. https://doi.org/10.1559/152304001782173961

Ching-Yi, L., Fu-Wen, L., Sheng-Tun, L., & Tsung-Hsueh, L. (2018). 5S dashboard design principles for self-service business intelligence tool users. *Journal of Big Data Research*, *1*(1), 5–19. https://doi.org/10.14302/issn.2768-0207.jbr-18-2175

Coltekin, A., Heil, B., Garlandini, S., & Fabrikant, S. I. (2009). Evaluating the effectiveness of interactive map interface designs: A case study integrating usability metrics with eye-movement analysis. *Cartography and Geographic Information Science*, *36*(1), 5–17. https://doi.org/10.5167/uzh-10125

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet. Infectious Diseases*, *20*(5), 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1

Dowding, D., & Merrill, J. A. (2018). The development of heuristics for evaluation of dashboard visualizations. *Applied Clinical Informatics*, *9*(3), 511–518. https://doi.org/10.1055/s-0038-1666842

Fan, M., Wang, Y., Xie, Y., Li, F. M., & Chen, C. (2023). Understanding how older adults comprehend COVID-19 interactive visualizations via think-aloud protocol. *International Journal of Human–Computer Interaction*, *39*(8), 1626–1642. https://doi.org/10.48550/arXiv.2202.11441

Few, S. (2006). *Information dashboard design: The effective visual communication of data*. O'Reilly Media, Inc.

Few, S., Edge, P. (2007). *Data visualization: Past, present, and future*. IBM Cognos Innovation Center.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, *24*(6), 631–645. https://doi.org/10.1016/S0169-8141(98)00068-7

GoogleTrend. (2023). *Google trends for the term "dashboard"*. https://trends.google.com/trends/explore?date=all&q=dashboard&hl=cs

Graham, L. M. (2008). Gestalt theory in interactive media design. *Journal of Humanities & Social Sciences*, *2*(1), 1–12.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Kamel Boulos, M. N., & Geraghty, E. M. (2020). Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: How 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. *International Journal of Health Geographics*, *19*(1), 8. https://doi.org/10.1186/s12942-020-00202-8

Kitchin, R., Lauriault, T. P., & McArdle, G. (2015). Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. *Regional Studies, Regional Science*, *2*(1), 6–28. https://doi.org/10.1080/21681376.2014.983149

Krassanakis, V., & Cybulski, P. (2019). A review on eye movement analysis in map reading process: The status of the last decade. *Geodesy and Cartography*, *68*(1), 191–209. https://doi.org/10.24425/gac.2019.126088

Lan, Y., Desjardins, M. R., Hohl, A., & Delmelle, E. (2021). Geovisualization of COVID-19: State of the art and opportunities. *Cartographica*, *56*(1), 2–13. https://doi.org/10.3138/cart-2020-0027

Lewis, J. R. (2001). Evaluation of procedures for adjusting problem-discovery rates estimated from small samples. *International Journal of Human–Computer Interaction*, *13*(4), 445–479. https://doi.org/10.1207/S15327590IJHC1304_06

Li, X., Wang, H., Chen, C., & Grundy, J. (2022). An empirical study on how well do COVID-19 information dashboards service users' information needs. *IEEE Transactions on Services Computing*, *15*(3), 1178–1192. https://doi.org/10.1109/TSC.2021.3114673

MacEachren, A. (1995). *How maps work: Representation, visualization and design*. Guilford Press.

Manhartsberger, M., & Zellhofer, N. (2005, November 8). Eye tracking in usability research: What users really see. Empowering software quality: How can usability engineering reach these goals? *1st Usability Symposium* (pp. 141–152). HCI&UE Workgroup.

Marr, B. (2012). *Key performance indicators (KPI): The 75 measures every manager needs to know*. Financial Times Press.

Monkman, H., Martin, S. Z., Minshall, S., Kushniruk, A. W., & Lesselroth, B. J. (2021). Opportunities to improve COVID-19 dashboard designs for the public. In R. Marcilly, L. Dusseljee-Peute, C. E. Kuziemsky, X. Zhu, P. Elkin, & N. Christian (Eds.), *Context sensitive health informatics: The role of informatics in global pandemics* (2021/11/11 ed., Vol. 286, pp. 16–20). IOS Press BV.

Pohlert, T. (2023). *PMCMRplus: Calculate pairwise multiple comparisons of mean rank sums extended*. https://CRAN.R-project.org/package=PMCMRplus

Popelka, S. (2014). *Optimal eye fixation detection settings for cartographic purposes* [Paper presentation]. 14th SGEM GeoConference on Informatics, Geoinformatics and Remote Sensing, Albena, Bulgaria. https://doi.org/10.5593/SGEM2014/B21/S8.091

Praharaj, S., & Wentz, E. A. (2022). Building community resilience through geospatial information dashboards. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *48*(4/W5), 151–157. https://doi.org/10.5194/isprs-archives-XLVIII-4-W5-2022-151-2022

Praharaj, S., Solis, P., & Wentz, E. A. (2023). Deploying geospatial visualization dashboards to combat the socioeconomic impacts of COVID-19. *Environment and Planning B: Urban Analytics and City Science*, *50*(5), 1262–1279. https://doi.org/10.1177/23998083221142863

Romano, A., Sotis, C., Dominioni, G., & Guidi, S. (2020). *Covid-19 data: The logarithmic scale misinforms the public and affects policy preferences*. PsyArXiv Preprints. https://doi.org/10.31234/osf.io/42xfm

Roth, R. E., & MacEachren, A. M. (2016). Geovisual analytics and the science of interaction: An empirical interaction study. *Cartography and Geographic Information Science*, *43*(1), 30–54. https://doi.org/10.1080/15230406.2015.1021714

Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing: How to plan, design and conduct effective tests*. Wiley Publishing Inc.

Sauro, J. (2023). *Sample size calculator for discovering problems in a user interface*. https://measuringu.com/calculators/problem_discovery/

Sedrakyan, G., Mannens, E., & Verbert, K. (2019). Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. *Journal of Computer Languages*, *50*, 19–38. https://doi.org/10.1016/j.jvlc.2018.11.002

SensoMotoricInstruments. (2010). *BeGaze software manual*. SensoMotoric Instruments.

Turner, C. W., Lewis, J. R., & Nielsen, J. (2006). Determining usability test sample size. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors – 3 volume set* (Vol. 3, pp. 3084–3088). CRC Press.

Zuo, C., Ding, L., & Meng, L. (2020). A feasibility study of map-based dashboard for spatiotemporal knowledge acquisition and analysis. *ISPRS International Journal of Geo-Information*, *9*(11), 636. https://www.mdpi.com/2220-9964/9/11/636 https://doi.org/10.3390/ijgi9110636

About the authors

Anna Porti Suarez studied Geography at Universitat de Barcelona and GIS at Universitat Politècnica de Barcelona. Also completed a double master's degree in Geoinformatics and Cartography at Salzburg Universität and Palacký University Olomouc. Currently works developing geoweb applications at Institut Cartogràfic i Geològic de Catalunya.

Stanislav Popelka studied Geography and Geoinformatics (MSc) and then Geoinformatics and Cartography (PhD) at Palacký University Olomouc, Czechia. Now, he works there at the Department of Geoinformatics as the Head of the eye-tracking laboratory. His research is focused on the use of eye movement analysis for evaluation and optimization of maps.