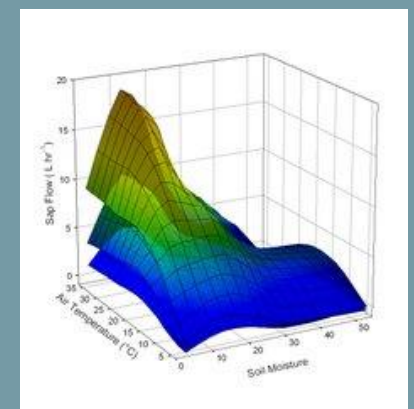
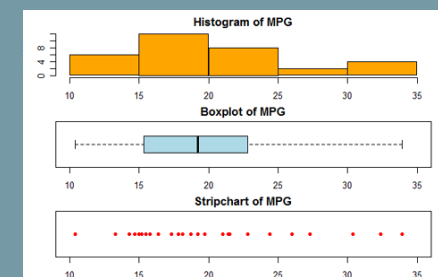
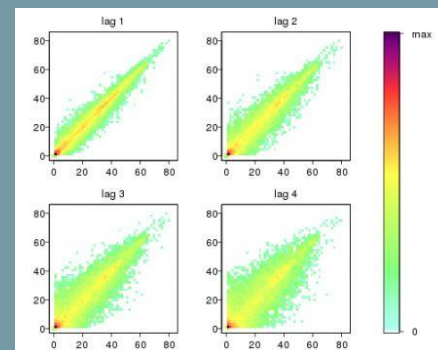


# Základy práce v prostředí R 1



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Co je R

Výpočtové prostředí pro statistické analýzy (programový kód pro zpracování dat)

Vznikl na platformě prostředí S/S+ (Bell Laboratories) na univerzitě Auckland (Ross Ihaka & Robert Gentleman) (1995)

R je neustále kontrolováno (the “R Core Team”)

Funguje pod Windows, Linux, MacOS

Volně stažitelný <http://www.r-project.org/index.html> z archívu CRAN

<http://cran.at.r-project.org/> balíček Base – základní knihovny

Volně šiřitelný pod GNU public licence

Je to open source program

Lehký skriptovací jazyk nabízí objektově orientované programování (možnost využít kód v jiných jazycích C, Fortran)

Obsahuje knihovny (libraries) skládající se z balíčků (packages), které podporují různé vědní obory (bioinformatika, genetika, geografie, sociologie, ekonomie)

Různá rozšíření

Bioconductor (2002)

Statistický nástroj pro bioinformatiky (molekulární biologie, genetika)

Analýza DNA čipů, podpora FASTA formátu

Online připojení na biomedicínské zdroje NCBI

Taxonomie a fylogenetika (ape, ade4), ekologie (vegan), populační ekologie (primer), habitatová preference (adehabitat) ....



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Proč R (výhody/nevýhody ?)

## Komerční software-

- Omezené množství nástrojů, chybí některé specifické metody kvantitativní ekologie (výpočty indexů podobnosti, diverzity, populační modely).
- Dlouhé čekání na nové verze
- Nekompatibilita s jinými programy
- Nemožnost zasahovat do grafiky

## R program-

- Ovládání pomocí příkazového řádku
- Načítání dat (nedostatek operační paměti)
- Problém zápisu a ukládání větších databází
- Úprava tabulek



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Základní syntaxe

R je striktní ve stylu zápisu (case sensitive)

Při nedodržení napíše (syntax error), sémantická chyba

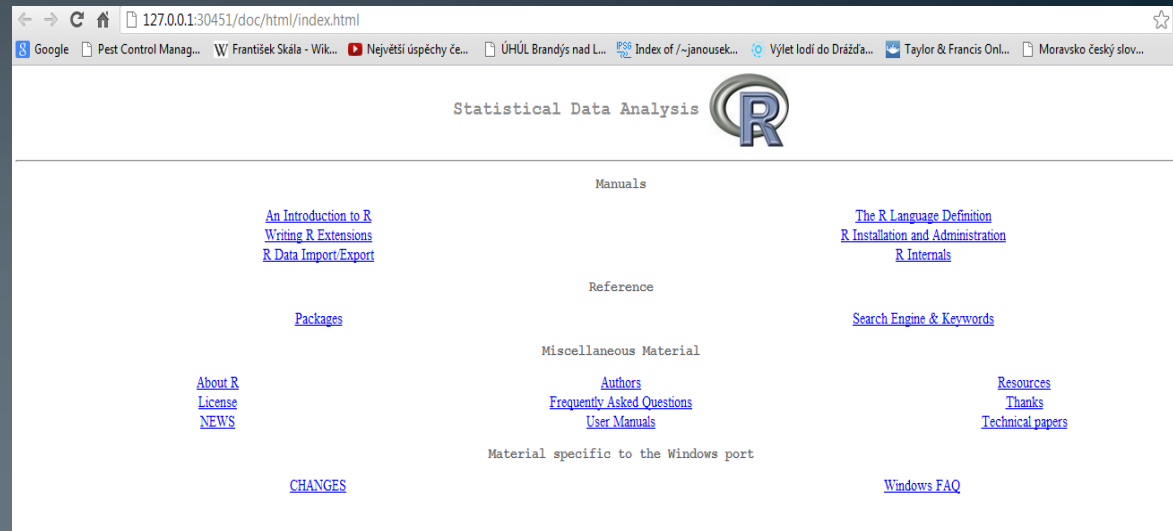
Všechny operace se provádějí pomocí vepisování příkazů (Arial, Courier New)

Parametry funkce a příkazu se

musí zadat do závorek ( )

Více příkazů na jednom řádku oddělíme středníkem ;

Složené funkce a příkazy které se mají spouštět postupně dáváme do složených závorek { }



```
glim (stats) R Documentation

Description
glim is used to fit generalized linear models, specified by giving a symbolic description of the linear predictor and a description of the error distribution.

Usage
glim(formula, family = gaussian, data, weights, subset,
     na.action, start = NULL, etastart, mstart, offset,
     control = list(...), model = TRUE, method = "glm.fit",
     x = FALSE, y = TRUE, contrasts = NULL, ...)

glim.fit(x, y, weights = rep(1, nobs),
        start = NULL, etastart = NULL, mstart = NULL,
        offset = rep(0, nobs), family = gaussian(),
        control = list(), intercept = TRUE)

## S3 method for class 'glm'
weights(object, type = c("prior", "working"), ...)

Arguments
formula an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted. The details of model specification are given under 'Details'.
family a description of the error distribution and link function to be used in the model. This can be a character string naming a family function, a family function or the result of a call to a family function. (See family for details of family functions.)
data an optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which glim is called.
weights an optional vector of 'prior weights' to be used in the fitting process. Should be NULL or a numeric vector.
subset an optional vector specifying a subset of observations to be used in the fitting process.
na.action a function which indicates what should happen when the data contain NA. The default is set by the na.action setting of option, and is na.fail if that is unset. The 'factory-fresh' default is na.omit. Another possible value is NULL, no action. Value na.exclude can be useful.
start starting values for the parameters in the linear predictor.
etastart starting values for the linear predictor.
mstart starting values for the vector of means.
contrasts this can be used to specify in a usual known component
```

## Nápověda

help.search(" ") hledání podtce kontextu  
help ( ) nápověda pro danou funkci  
PDF manuály – Help manuals  
HTML nápověda – Help Html help  
(seznam balíčků a funkcí v jednotlivých knihovnách)



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

```
File Edit View Misc Packages Windows Help
RGui
R Console
ISBN 3-900051-07-0
Platform: i386-pc-mingw32/i386 (32-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Previously saved workspace restored]
> cbind(1:3)
  [,1]
[1,]  1
[2,]  2
[3,]  3
> |
```

# Práce s konzolou

Po instalaci a spuštění se objeví okno s konzolou (okno nápovědy, grafické okno, informační okno)

Statistické analýzy se zadávají pomocí klávesnice do příkazového řádku

Konzole rozlišuje velká malá písmena a diakritiku

V programu lze nalistovat předchozí příkazy použitím šipky nahoru

R je objektově orientovaný jazyk (proměnné, data, výsledky, funkce jsou objekty)

Objekty se vytvářejí přiřazením (přiřazovací operátor se napíše pomocí “<” a “-“: <-)

Import dat pomocí schránky (clipboard), csv formát, GIS.

Vymazání konzoly Edit – clear console

Dlouhý příkaz – rozdělíme ho stiskem klávesy enter (na jednom řádku až 128 znaků)

Ukončení programu R (q)

R používá desetinou tečku ne čárku !!!

Vzhled konzoly lze měnit Edit – GUI preferences

## Příklad tvorby objektů

```
X<- 3+4
```

```
Y<- 5-7
```

```
Z<-X/Y
```

```
M<- sqrt(Z^2) # výpočet příkladu
```

```
ls () vypíše všechny objekty
```

```
rm () vymaže všechny objekty
```

```
str() podrobnosti o objektu
```



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Práce s objekty

Objekty nemohou začínat číslicí ani matematickými operátory nebo funkcemi, nesmí se shodovat s názvem jiného objektu

Rozdělení názvu objektu pomocí tečky

Každý objekt patří do určité třídy podle toho s ním pracujeme (číslo, vektor, faktor, pole, matice, tabulka, funkce, rovnice)

Řada funkcí pracuje pro různé třídy objektů různě

Objekty se ukládají do paměti jsou součástí workspace ukládá se File – Save Workspace

`str( )` podrobnosti o objektu, typ objektu

`class ( )` určí třídu objektu

`objects ( )` – vypíše používané objekty

`attributes()` – vypíše vlastnosti objektu

`search ( )` – zjistí které knihovny a tabulky jsou zpřístupněny

`rm(list=ls())` – vymaže všechny objekty

Generické funkce – vyhodnocení třídy objektu → zpracování funkcí nejvhodnější pro daný objekt (summary, plot)

## Práce s knihovnami a balíčky

Základní balíček neobsahuje některé knihovny

Instalace

`install.packages(" ")`

`Packages - instal Packages/ update Packages`

`library()` přiinstaluje knihovnu

## Základní typy vektorových tříd

Celá čísla (integer);

Reálná čísla (numeric);

Písmena (character);

Logické hodnoty (logical);

Vzorec  $Y \sim x+z+e$  (formula);



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Vkládání Dat

`scan("clipboard")` čtení vektorů, matic ...

Čte data sekvenčně jak jsou uložena v souboru (vhodná pro nepravidelná data)

`read.table("clipboard")`

Převádí soubor do dlouhého řetězce

`read.csv("/Users/artemis/Desktop/rock_lichen_data.csv")`

Čtení dat s tabulkových programů

Kde najdu data `getwd()`

Změna adresáře `setwd("C:/Users/jmeno/Desktop/")`

export dat

`write.table(x,"xx.xls",sep='\t')`

`write.csv(x, file = "foo.csv")`

```
> 1:10
```

```
[1] 1 2 3 4 5 6 7 8 9 10
```

```
> c(1.1, 25, 3.32, 4, 12, 14, 85)
```

```
[1] 1.10 25.00 3.32 4.00 12.00 14.00  
85.00
```

```
> matrix(c(1, 2, 3, 4, 5, 6, 7, 8,  
9), nrow = 3, ncol = 3)
```

```
> array(c(1, 2, 3, 4, 5, 6, 7, 8, 9),  
c(3, 3))
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Ukládání obrázků do souboru

postskript (.ps)

```
x<-postscript(file=„muj.graf.ps“,width=, height=,horizontal=T/F)
```

portable document format (.pdf)

```
x<- pdf(file = "muj.graf.pdf", width =, height = )
```

Bitové mapy (.png, .jpg, .tiff)

```
x<- jpeg(filename = "mujboxplot.jpg", width = , height =, quality =)
```

```
tiff(file="coc2.tiff", width=17, height=17, units="cm", pointsize=9, res=400,  
type="window", antialias="default")
```

```
boxplot(f~factor(fortype2,levels=c("forest.interior","forest.edge","solitaire.tree","me  
adow")),data=x,xlab="Habitats",ylab="Predation  
probability",cex.axis=1.5,cex.lab=1.5)
```

```
dev.off()
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



# Chybějící hodnoty (NA)

Chybějící hodnoty ovlivní způsob fitování modelu, mohou snížit sílu statistické analýzy  
Některé funkce nepracují s chybějícími hodnotami v základním nastavení (průměr)

```
x<-c(4,5,6,7,NA)
```

```
mean(x)
```

```
na.rm= TRUE odstraní chybějící hodnoty
```

```
mean(x, na.rm=T)
```

Zjištění lokalizace chybějících hodnot

```
xx<- c(1:5,NA,NA,8:13)
```

```
which(is.na(xx))
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Tvorba vektorů a jednoduchých objektů

Vektor je proměnná s jedním nebo více hodnotami

Hodnoty mohou být přiřazeny do vektoru mnoha způsoby

```
x<- 1:10
```

```
x<- c (1,2,3,4,5,6,7,8,9,10) řetězíčí funkce
```

```
x<- scan()
```

```
names(x)<-0:9
```

```
as.vector(x)
```

Nejběžnější forma jak přidělit data do vektoru je načíst data z externího zdroje

```
read.table(" ")
```

```
seq (from=, to= , by=, lenght.out=) vytvořím sekvenci
```

```
rep (x, times= , each= , lenght.out=) opakování prvků vektoru
```

Vektorové funkce

```
max, min, sum, sort, order,
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Tvorba matic, polí, seznamů, faktorů & tabulek

Matice

`matrix (data= , nrow=, ncol=, byrow=T/F, dimnames=list(xy=c(),xz=1:5))`

`mymat (2,1)` vypíše prvek v 2 řádku a 1 sloupci

Pole

`array(data=, dim=c(4,2), dimnames=list())`

List

`x<-list(matice=matrix(1:8,2,4),data=c("a","b","c"))`

Tabulka (může obsahovat různé typy dat ale vždy jeden v každém sloupci)

`x[c(2,1)]` vypis druhého objektu a následně prvního

`data.frame(rr=c(), bs=c(),va=c())`

`fix(data.frame)` úprava tabulky

`mytab[ , "název sloupce"]` nebo `[, 2]` vvypis sloupce

`mytab [1:2,]` vypis řádků

`mytab[x$město=="Brno",]` vypis tabulky s vybranými hodnotami

`x$město` vypis sloupce

Faktor (vektor znaku, které jsou ordinální nebo nominální)

`factor(c(1,2,3,4), exclude=1, ordered=T)`



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Úprava tabulek, polí a matic

Nejlehčí editace pomocí funkcí `fix()` & `edit()`

`append(x,c(2,5,4))` # přidá hodnoty k vektoru nebo tabulce

`cbind(x1=c( ),x2=c( )); rbind( )` # sloučí vektory po sloupcích a řádcích

`sort(c( ),decreasing=T)` #seřadit hodnoty

`my.sort<-order(my.tab[,2],my.tab$město)` #řazení podle sloupce 2 a potom 1

`my.tab[my.sort,]` #výpis nově seřazené tabulky

`replace(my.tab$město,list=c(4,5),values=c(7,8))` # nahradí hodnoty v tabulce

`unique(my.tab)` # nezobrazí duplikované řádky

`subset(my.tab,x>3)` # vzbere y tabulky určité hodnoty

`names(my.tab)<- c("aa", "as", "ds")` # změna názvu

`t(x)` transpozice matic

`x %*% t(x)` # násobení matic

`rowsum(x)` # suma v každém řádku

`colsum(x)` # suma v každém sloupci

`apply(x, margin=1, fun=sum)` sečti přes řádky (výpočet funkcí pro každý řádek nebo dimenze)

`apply(x,2,function(x) sum(x+2))`

`factor(x)` vytváří z vektoru objekt typu factor



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Funkce pro práci s vektory

`paste (c("a", "s", "g"), c(1,2,3) )` #slučování vektorů

Hromadné provádění výpočtů na objektech

`rowSums ()`, `colSums ()`, `rowMeans ()`, .....

`table (x)` spočítá četnosti faktorů

`lapply (x,sd)`, `sapply (x,sum)` provádí výpočty funkcí v jednotlivých položkách seznamu

`tapply (x$plat,x$meno,mean)` výpočet pro tříděná data v tabulce



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Programování

Je objektivě orientovaný program

(regrese v SAS, SPSS, NCSS vyhodí velké množství výsledků; R ukáže je to co vypíšeme do příkazového řádku)

Polymorfní příkazy stejná funkce může být použita na různé typy objektů, výsledek záleží na typu objektu

R je více programovací jazyk než soubor diskretních příkazů (možná kombinace více příkazů)

Smyčka (Loop)

```
x <- c(5,12,13)
```

```
for (n in x) print(n^2)
```

```
[1] 25
```

```
[1] 144
```

```
[1] 169
```

Vlastní Funkce

```
pr<-function(a=1,b=3,x=5:10) {y<-
```

```
a*x+b
```

```
list (hodnoty=y, graf=plot(x,y))}
```

```
pr()
```

If else

```
If (is.vector (x) mean (x) else apply
```

```
(x,2,mean)
```

While

```
while (exist(name<-paste(x,j,sep="")))
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

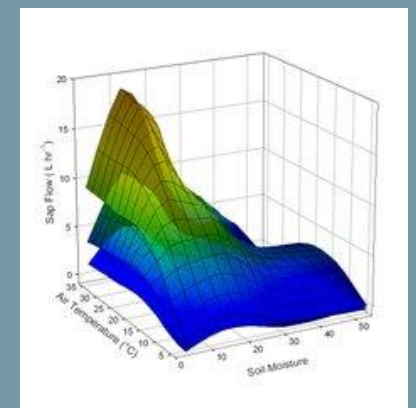
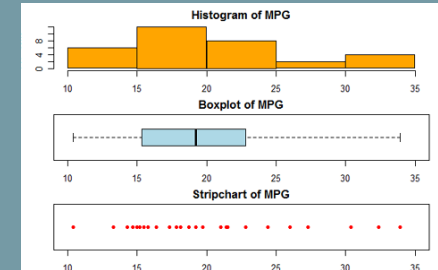
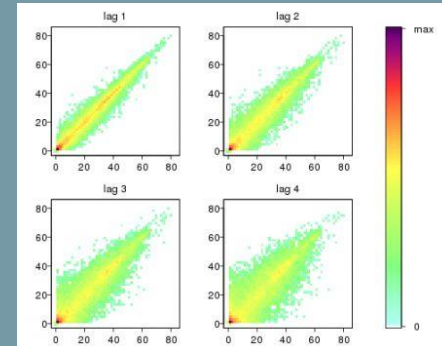


OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



# Základy práce v prostředí R 2



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

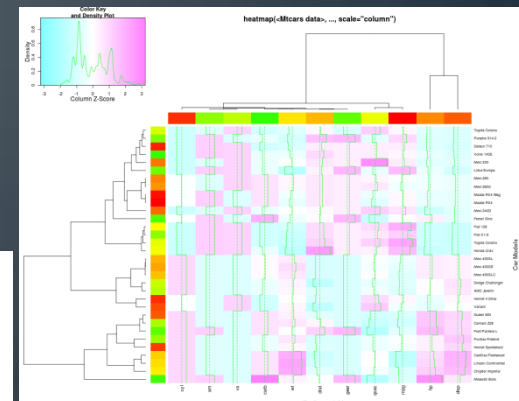
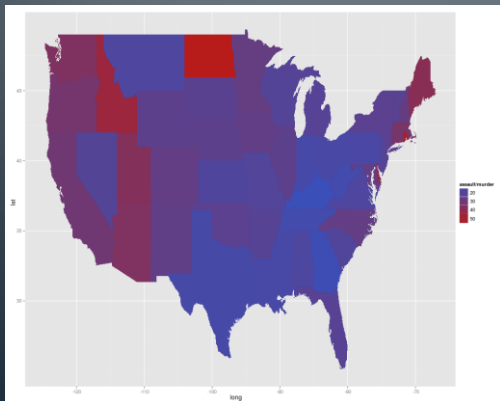
# Úvod do grafického prostředí

Grafické prostředí

Grafické funkce mohou být rozděleny do 3 kategorií

- Grafika High level** (generují grafy v novém grafickém okně, osy, popisky, body) funkce- `plot`, `hist`, `dotplot`
- Grafika Low level** (upravují a zlepšují už existující grafy přidávající více objektů a informací) `lines ()`, `points ()`, `text ()`, `axis ()`
- Interaktivní grafy** (umožňuje odebírání a přidávání dat z grafů)

Propracovanější grafy (lattice, rgl, gplots, ggplot2, IDPmisc)





# Interaktivní grafika

OrdiClust

**- Input (dataframe OR dudi) -**  
 Input dataframe : tab2 Set 271 14  
 Input dudi : Set

**Ordination method :**

**- Ordination method -**  
 cPCA  
 nPCA  
 COA  
 MCA  
 Submit

**- Ordination graph -**  
 X-axis : 1  
 Y-axis : 2  
 Plot rows  
 Plot columns

Number of axes : 2 Set

**Cluster analysis :**

**- Distance -**  
 Euclidean  
 maximum  
 manhattan  
 canberra  
 binary  
 minkowski  
 dudi (ade4)

**- Cluster method -**  
 ward  
 single  
 complete  
 average  
 mcquitty  
 median  
 centroid  
 Submit

**Number of groups**

Number of groups : 4 Cut tree  
 Level height : 7.26292 Cut tree  
 Inertia ratio : 49% Draw curve  
 BGA MCTest p-value : 0.001

Colors  
 Ellipses  
 Conv. hulls

4  
 0 5 10 15 20  
 Dismiss Submit

Box-Cox Demo

**Raw Data**  $\lambda = -1.05$

**Raw Residuals**  $\lambda = -1.05$

lambda: -1.05

Hscale: 2 Vscale: 1.5  
 Refresh Exit

# Přehled typů grafů

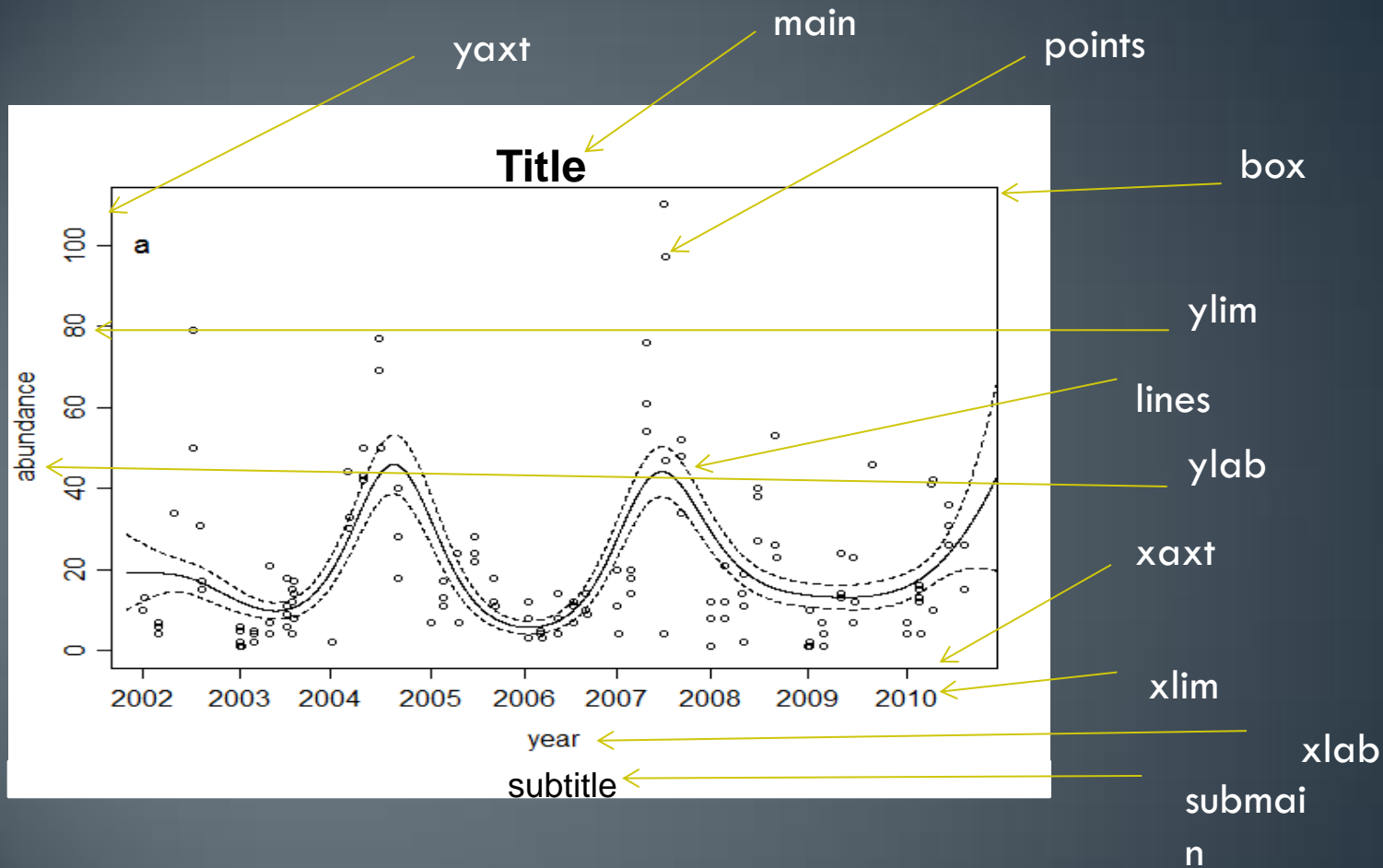
Základní grafy: plot (normální graf), boxplot (krabicový graf), barplot (sloupcový graf), pie (koláčový graf), matplot (graf s více řadami dat), persp (prostorový)

Speciální grafy: pairs (skupina xy grafů), steam (stonek a listy), mosaicplot (mozaikový graf), fourfoldplot (čtyřlístkový graf), assocplot (zobrazení vztahu mezi očekávanou a naměřenou hodnotou)

3D bodové grafy: coplot, pairs, levelplot, xyplot, bwplot

3D grafy: contour, persp, plot3d.ca

# Grafické prvky



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# High level- grafické funkce

- **Bodové grafy** – `plot (x,y)` zobrazení závislosti kvantitativních proměnných
- **Grafy rozdělení** – rozdělení náhodné proměnné `hist ()`, `stem ()`, `qqnorm()`, `qqline()`, `qqplot ()` – rankitový graf srovnává kvantily z předpokládaného normálního rozdělení a empirické kvantily
- **Krabicové grafy** – `boxplot (x,y)` alternativní způsob zobrazení popisných statistik (poloha-průměr, medián, variabilita a tvar- kvantily)
- **Lattice grafy** – `lattice (y~x|A)` vhodné pro vykreslení závislosti spojitých proměnných pro různé úrovně kategoriální proměnné
- **Interakční grafy** – `interaction.plot()` zobrazí průměry závislé proměnné pro všechny úrovně vysvětlující proměnné
- **Sloupcové grafy** – `barplot ()` nejčastěji slouží k porovnání četností
- **Párové grafy**- `pairs ()` vhodný pro zjištění vztahů mezi několika proměnnými



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



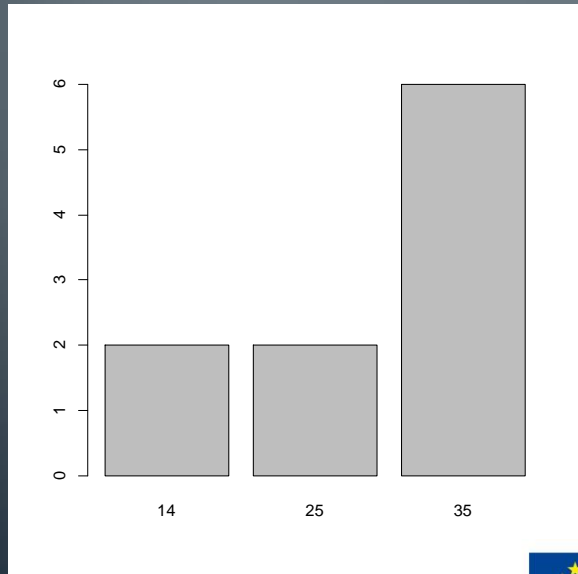
OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

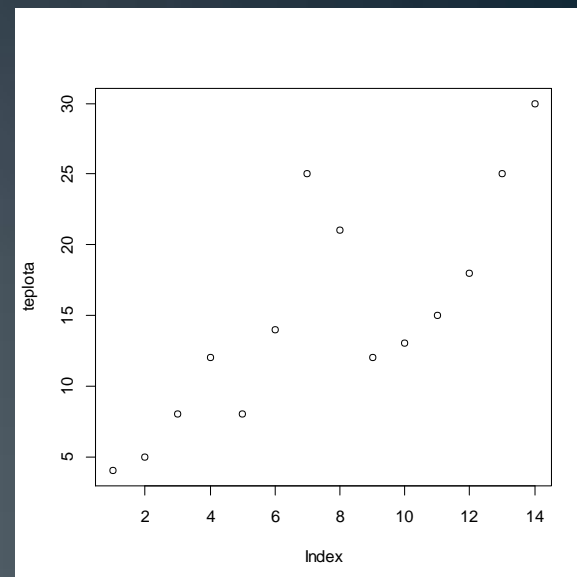
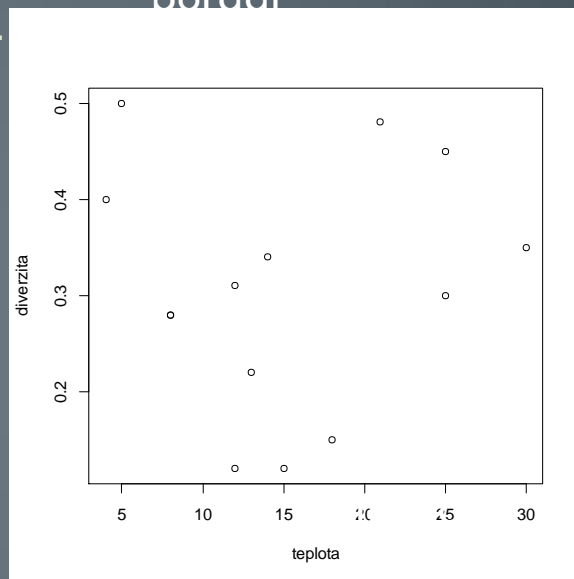
# Funkce plot ()

Funkce plot je generická funkce vyhodnocující třídu objektu a podle toho přiřadí k objektu nejvhodnější funkci.

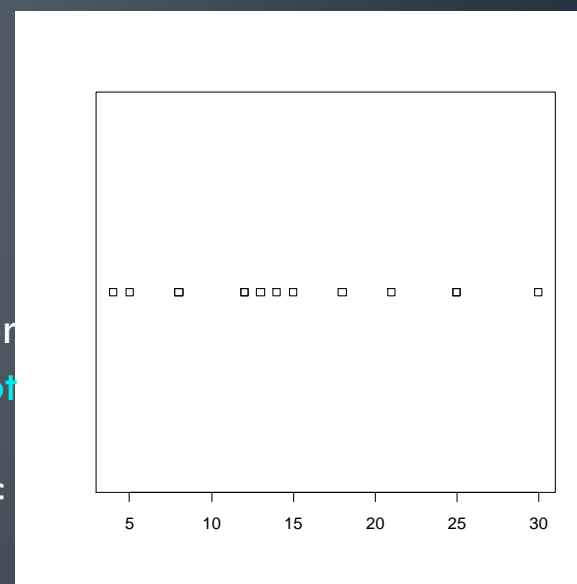
plot(x,y) nebo plot(y~x), funkce plot vytvoří scatterplot



plot(x), z numerického vektoru funkce plot vytvoří časovou řadu, graf teploty v závislosti na jejich pořadí

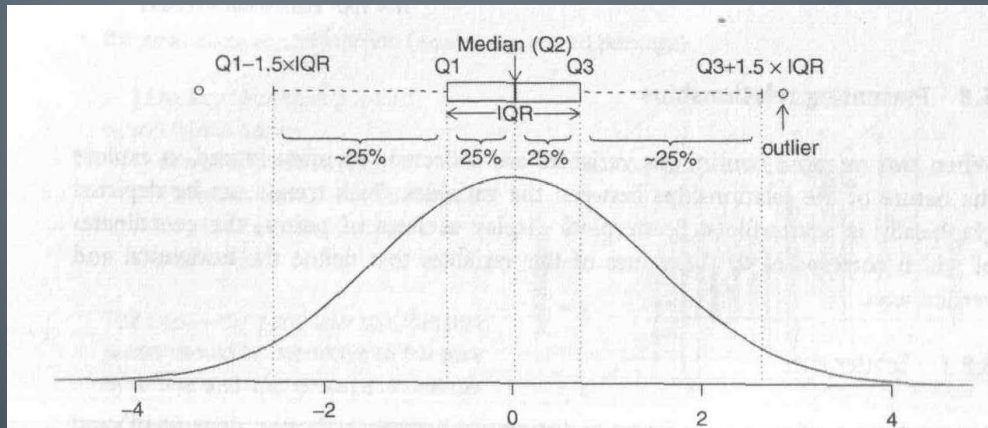
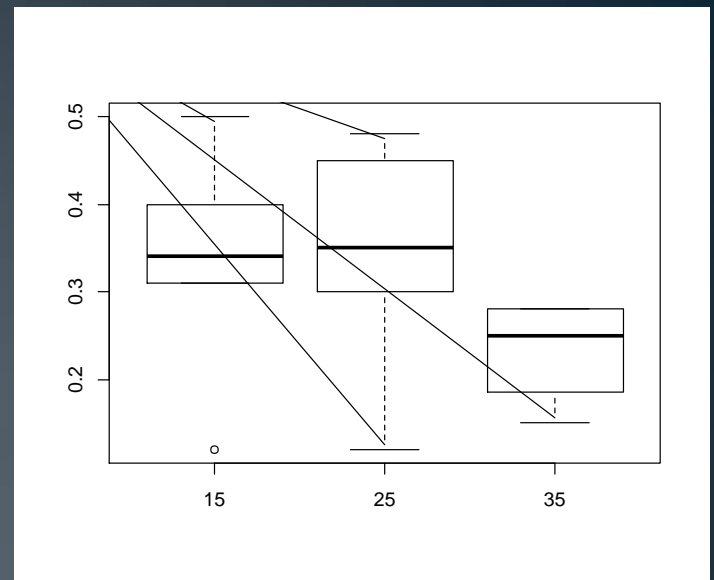


pak funkce plot(y) vytvoří sloupcový graf barplot



plot(~x), tato forma funkce plot vytvoří stripchart

`plot(x,y)` nebo `plot(y~x)`,  
jestliže  $x$  je faktor a  $y$   
vektor, funkce `plot()` vytvoří  
`boxplot()` y pro každou  
hodnotu  $x$ .



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Grafické parametry

Grafické parametry upravují vlastnosti high a low grafických funkcí

Základní nastavení grafu při otevření grafického okna (před vykreslením grafu) !!!

Každý z těchto parametrů může být zadán jako argument do funkce `par ()`

`par ()` globální grafický parametr který se aplikuje na všechny funkce aktivní v grafickém okně

`par (mfrow=c(1,3))` první číslo specifikuje počet řádků druhé počet sloupců

# Parametry zodpovědné za kontrolu rozměrů a rozvržení grafického okna

`din` – šířka grafického okna

`fin` – šířka vykreslovací oblasti

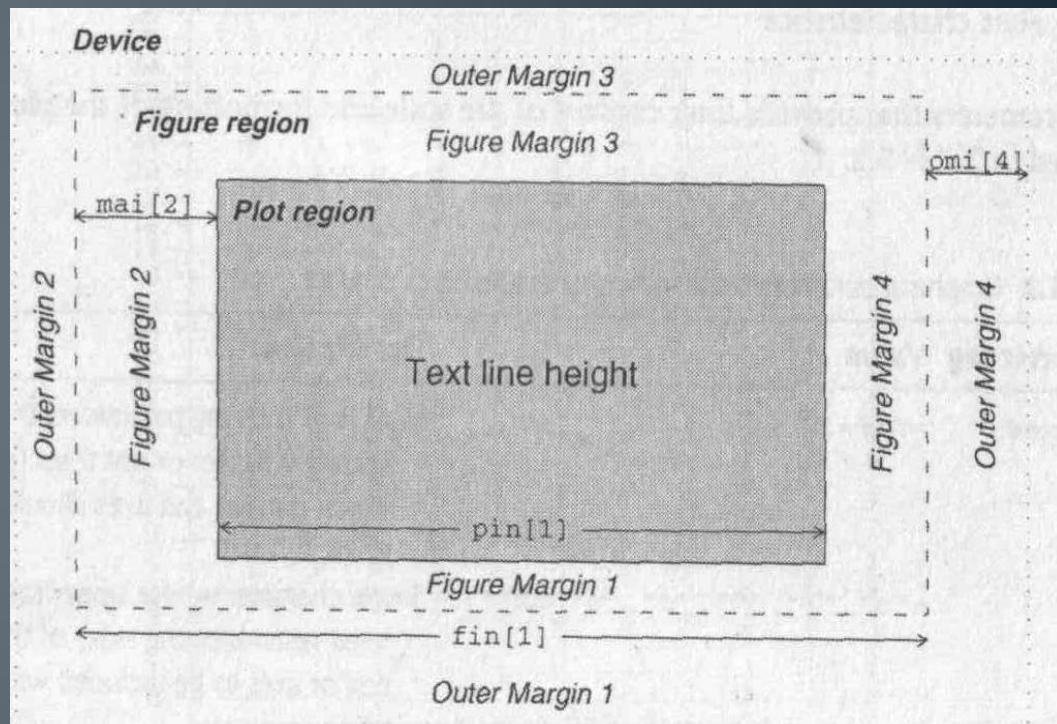
`pin` – šířka grafu

`fig c(left, right, bottom, top)` – souřadnice vykreslovací oblasti

`mai c(bottom, left, top, right)` velikost okrajů vykreslovací oblasti

`mar` počet řádků na okrajích vykreslovací oblasti

`omi c()` velikost vnějších okrajů





# Argumenty Funkce plot ()

Je generická funkce produkující různé typy grafů podle třídy objektu.  
Výsledný graf lze upravovat podle nastavení argumentů grafu.

Základní argumenty:

`type` ("p"-bodový, "l"-liniový, "n"-nevynášet) typ grafu

`axes` (true/false) vykreslení os

`xlab`, `ylab` () pojmenování os

`xlim`, `ylim` () měřítko os

`bty` () typ ohraničení grafu

`cex` () relativní velikost znaků a textu (jak mají být zvětšeny v poměru k základní velikosti) `cex.lab`- velikost znaku názvu os, `cex.axis`- velikost znaku popisků os

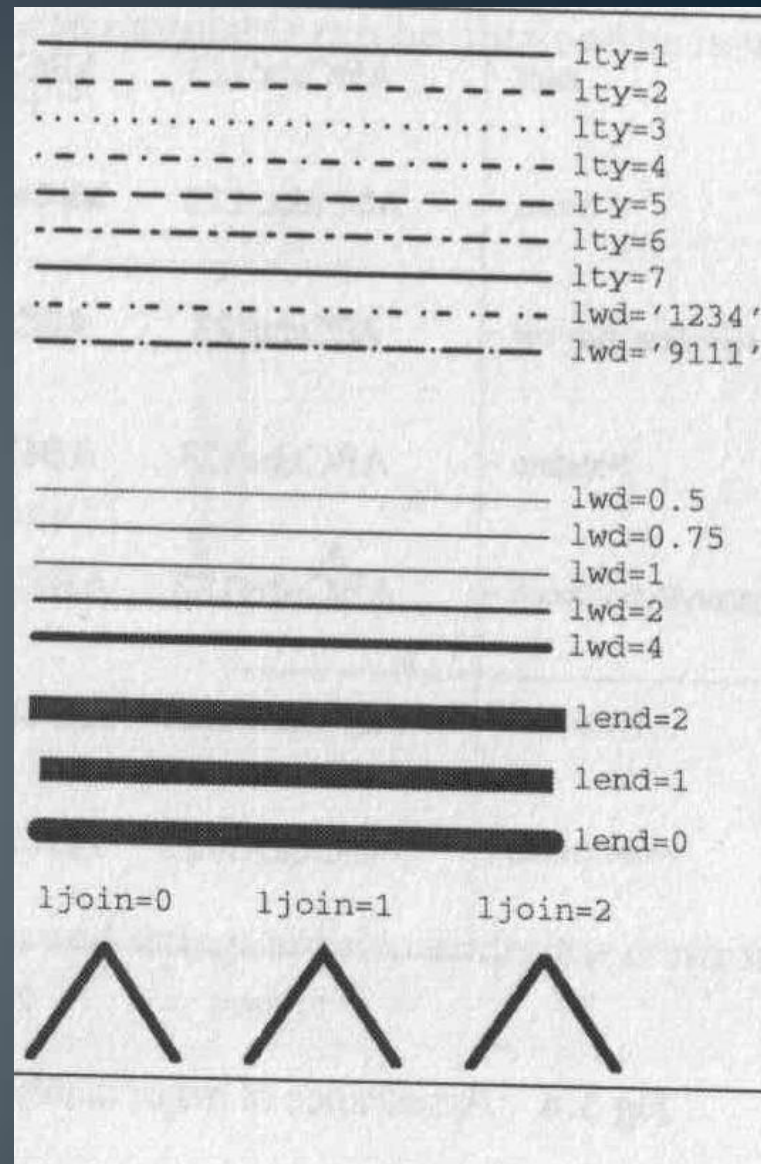
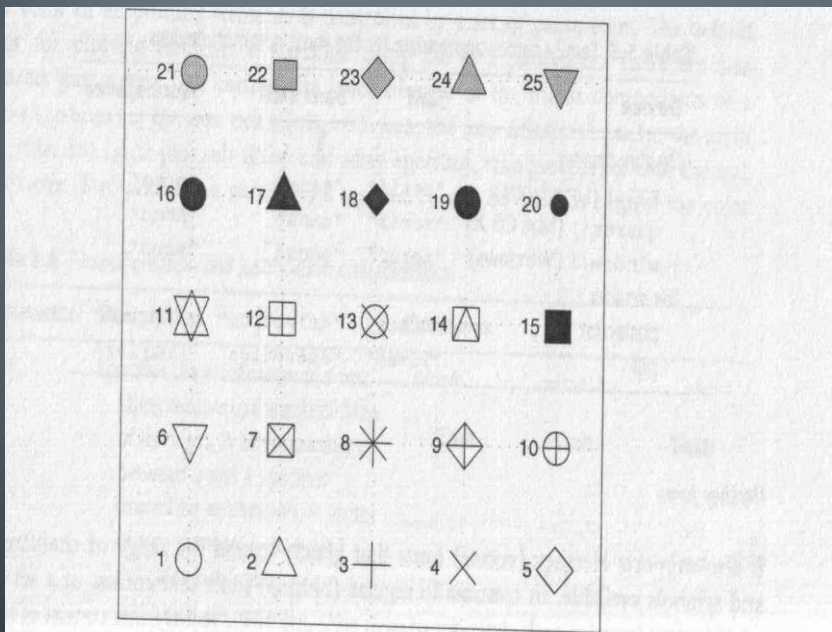
`xaxt`, `yaxt` ="y" , "n" – jestli se mají vykreslit osy

`xlog`, `ylog` =T, F – jestli se mají osy vykreslovat na logaritmické škále

`las` c(1-xy vodorovně, 2- x svisle, y vodorovně) – otočení popisků

OS

# Typ, šířka čáry a typ vykreslovaných bodů



# Font a vzhled písma

```
Plot  
(family="serif"  
, font=2)
```

	Common fonts			
sans	ABCabc123	<b>ABCabc123</b>	<i>ABCabc123</i>	<b><i>ABCabc123</i></b>
serif	ABCabc123	<b>ABCabc123</b>	<i>ABCabc123</i>	<b><i>ABCabc123</i></b>
mono	ABCabc123	<b>ABCabc123</b>	<i>ABCabc123</i>	<b><i>ABCabc123</i></b>
	Postscript/PDF fonts			
Helvetica-Narrow	ABCabc123	<b>ABCabc123</b>	<i>ABCabc123</i>	<b><i>ABCabc123</i></b>
Palatino	ABCabc123	<b>ABCabc123</b>	<i>ABCabc123</i>	<b><i>ABCabc123</i></b>
NewCenturySchoolbook	ABCabc123	<b>ABCabc123</b>	<i>ABCabc123</i>	<b><i>ABCabc123</i></b>
Bookman	ABCabc123	<b>ABCabc123</b>	<i>ABCabc123</i>	<b><i>ABCabc123</i></b>
AvantGarde	ABCabc123	<b>ABCabc123</b>	<i>ABCabc123</i>	<b><i>ABCabc123</i></b>
	1 (Plain)	2 (Bold)	3 (Italic)	4 (Bold + Italic)



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

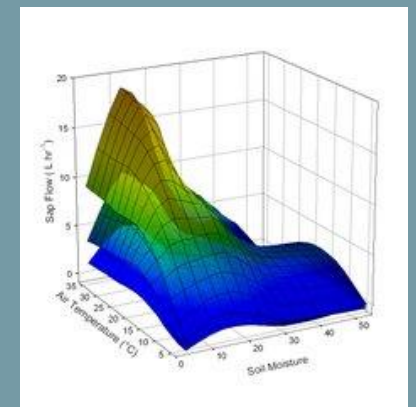
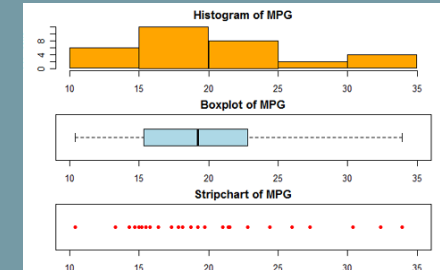
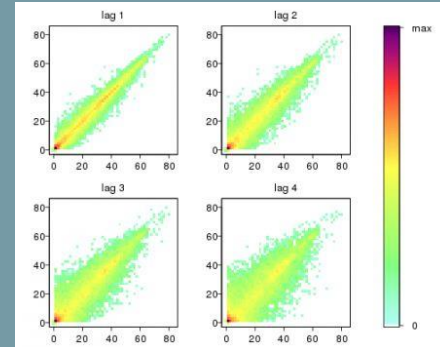


OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



# Základy práce v prostředí R 3



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Tvorba a vkládání legendy do grafu

`legend (x,y)` souřadnice legendy v grafu  
`title = "teplota"` pojmenování legendy  
`box.lty, box.lwd` typ a tloušťka čáry ohraničující legendu  
`bg="grey"` barva pozadí,  
`text.col("red","black","yellow")` barva písma  
`horiz= T` horizontální poloha  
`pch =c(1,2,3)` typ vykreslovaného bodu  
`fill=c("gray")` barva výplně bodu  
`pt.cex =1:3` relativní velikost bodu  
`pt.lwd=2` šířka ohraničení bodu  
`lty=c(1,2,3)` typ čáry  
`lwd =c(1.5)` šířka čáry



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



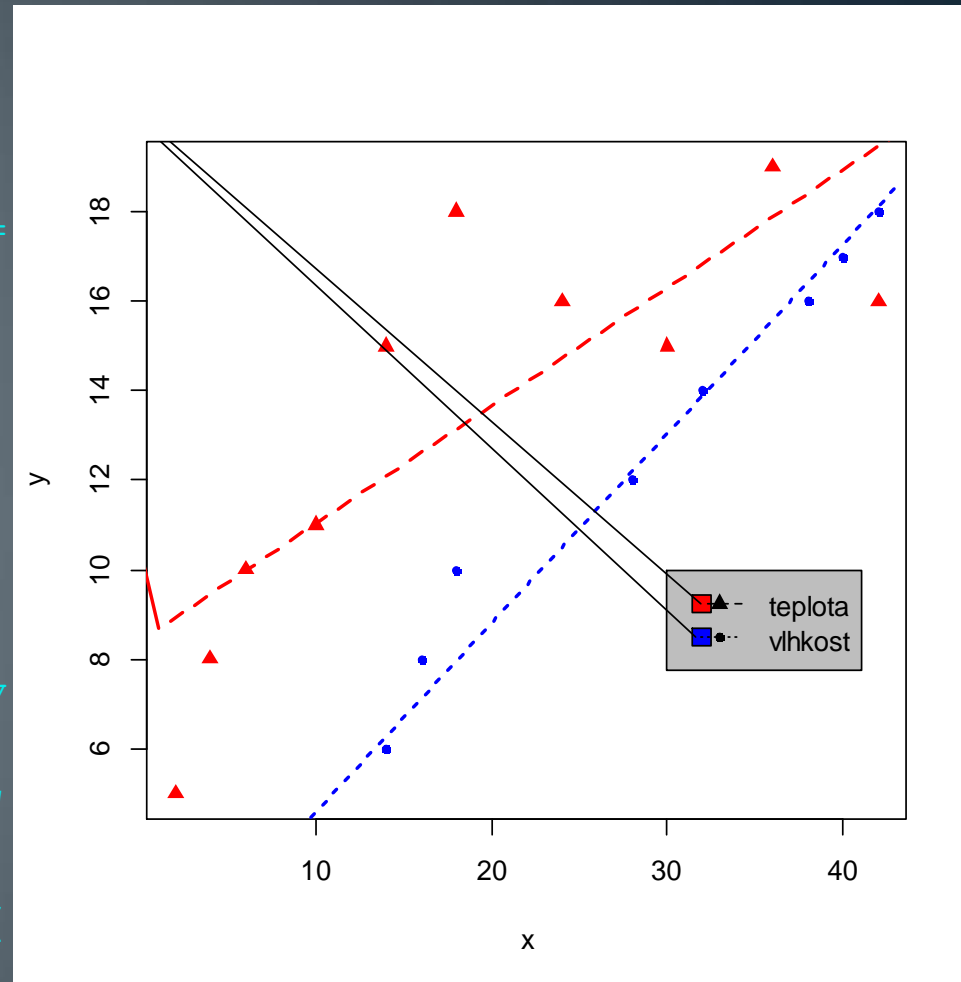
OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

```

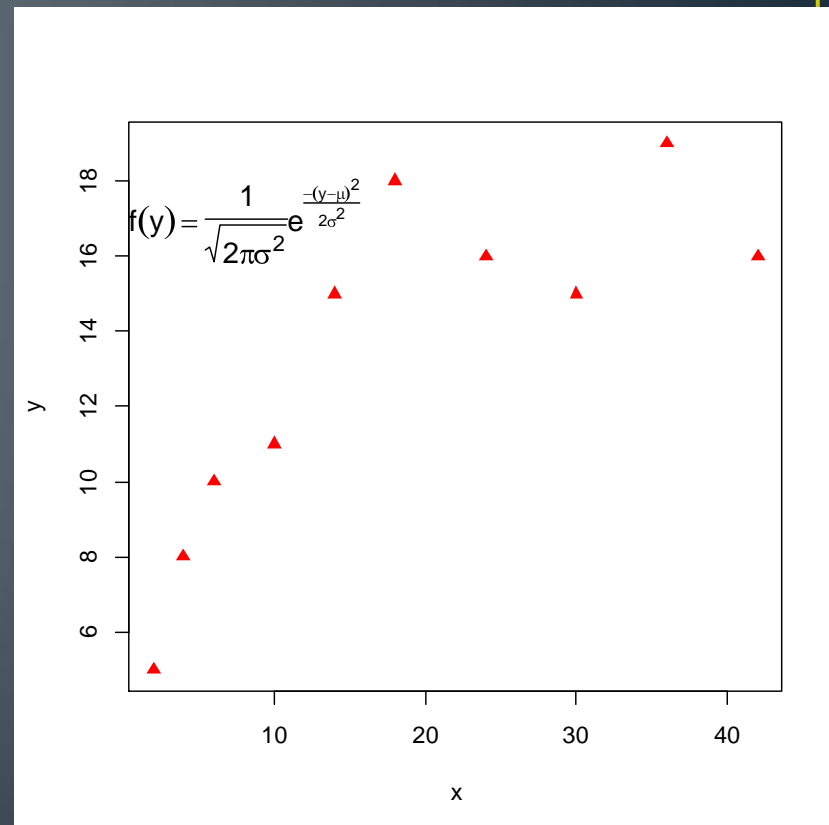
x<-
c(2,4,6,10,14,18,24,30,36,42)
y<-
c(5,8,10,11,15,18,16,15,19,16)
plot(y~x,pch=17,col="red")
lm1<-lm(y~x)
abline(lm1,lwd=2,col="red",lty=
2)
x2<-
c(7,10,14,16,18,28,32,38,40,42)
y2<-
c(2,4,6,8,10,12,14,16,17,18)
points(y2~x2,pch=20,col="blue")
lm2<-lm(y2~x2)
abline(lm2,lwd=2,col="blue",lty
=3)
legend(30,10,legend=c("teplota"
,"vlhkost"),bg="grey",pch=c(17,
20),fill=c("red","blue"),lty=c(
2,3))

```



# Vkládání vzorců a textu

- `text(8,17,`  
`expression(f(y)==frac(1,sqrt(2*pi*sigma^2)`  
`)*e^frac(-(y-mu)^2,2*sigma^2)),cex=1.25)`



# Vykreslování os

`side` – výběr osy

`at` – na jakém místě se zobrazí body na ose

`labels` – zobrazí popisky k jednotlivým bodům

`line` – specifikuje počet textových řádků od kraje

`lwd` – šířka čáry

`lty` – typ čáry

`col` – barva čáry

`padj` – specifikuje vzdálenost popisků od osy

`tick` – specifikuje jestli se má vypsát osa

```
axis(side=1,at=c(10,
20,30),labels=c("low",
"medium","high"),l
ine=-
1,lwd=2,lty=3,col="b
lue",padj=0,tick=F)
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

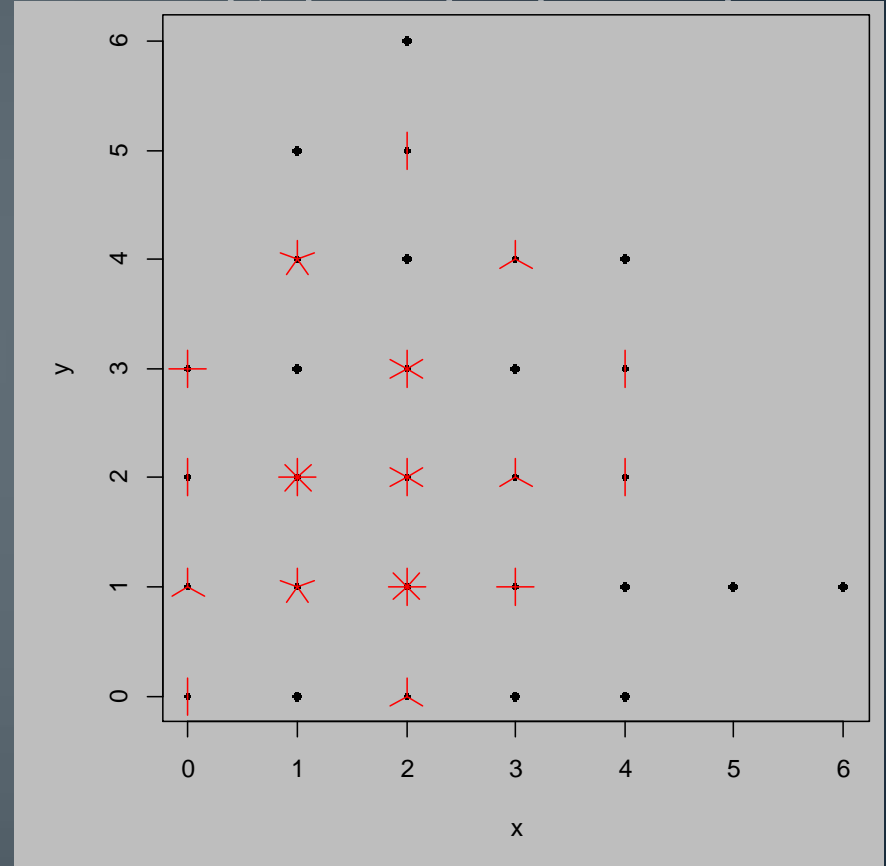
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



# Pokročilejší grafy

`sunflowerplot` – xy graf (slunečnicový), překrývající body se zobrazují jako lístečky

```
x<-rpois(80,2)
y<-rpois(80,2)
sunflowerplot(x,y)
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

`Fourfoldplot` (čtyřlístkový graf), znázorňuje vztahy v čtyřpolní tabulce. V grafu jsou znázorněny také konfidenční intervaly.

`x` – matice dat

`color` – barevné rozlišení

`conf.level` – vykreslí konfidenční intervaly

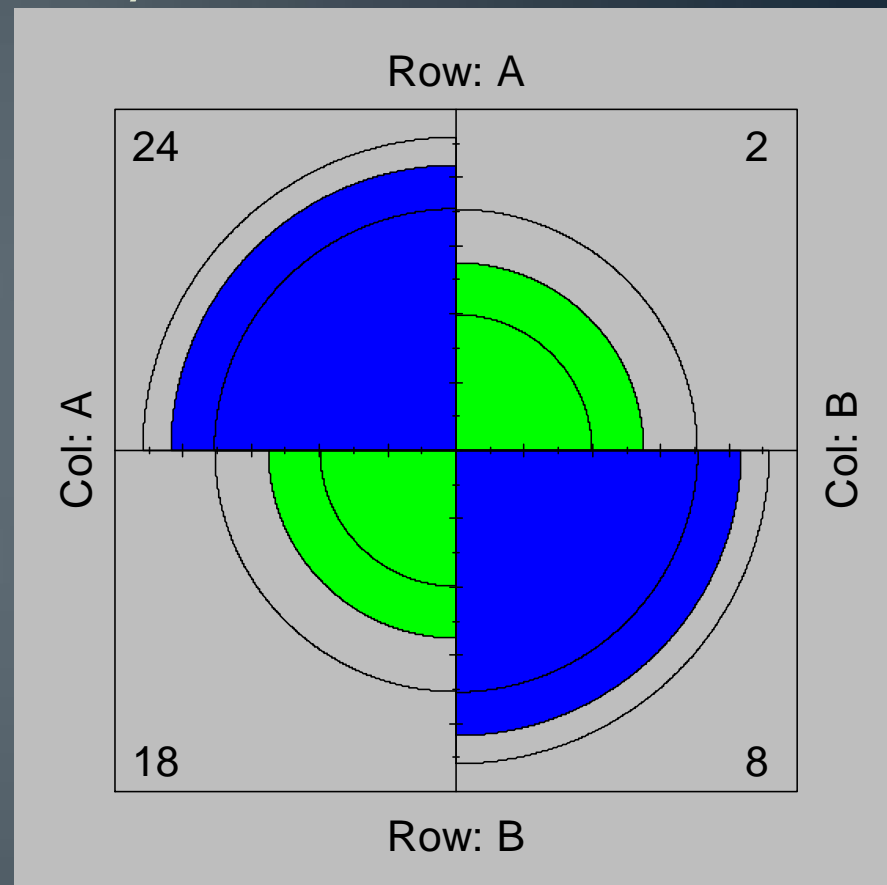
`std` – standardizace dat („margin“- vyvážení po uhlopříčkách)

`margins = 2` standardizace po sloupcích, 1 po řádcích, `c(1,2)` obojí

```
x<-
```

```
array(c(24,18,2,8),di  
m=c(2,2))
```

```
fourfoldplot(x,color=  
c("green","blue"))
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



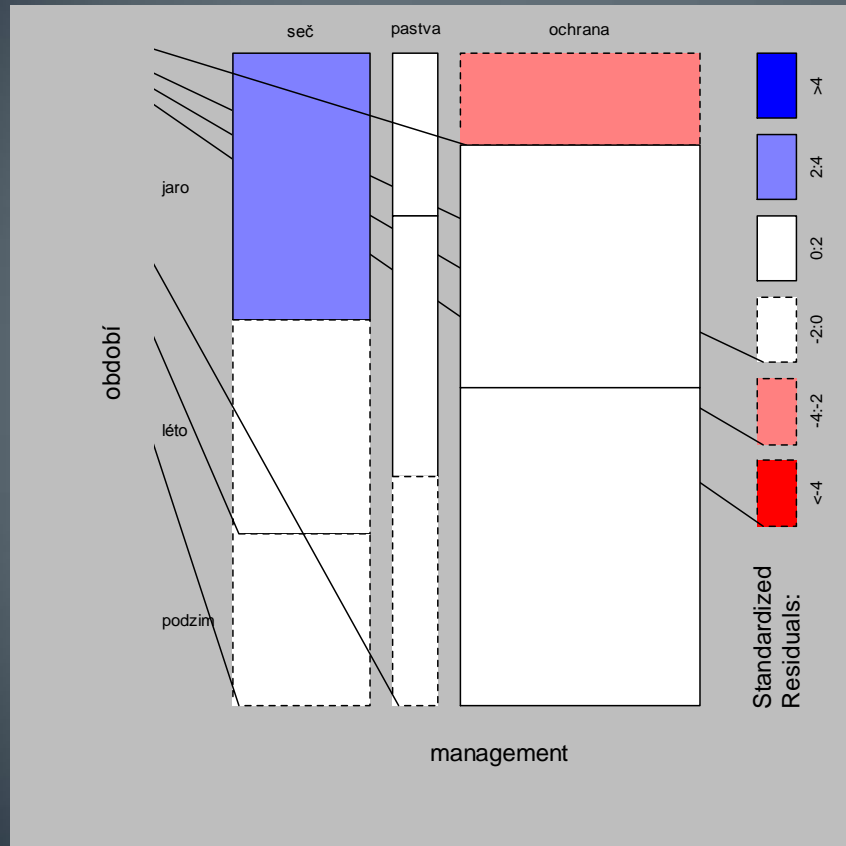
MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Mosaic plot – vytváří graf s obdélníky s délkou stran podle velikosti vzorku. Využívá se pro znázornění dat v kontingenčních tabulkách



aa<-

```
array(c(25,5,15,20,8,40,16,7,52),dim=c(3,3),dimnames=list("management"=c("seč", "pastva", "ochrana"),období=c("jaro", "léto", "podzim")))
```

shade – barevné odlišení residuálů

off – velikost mezer mezi obdélníky

las – popisky os

type – typ residuálů



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



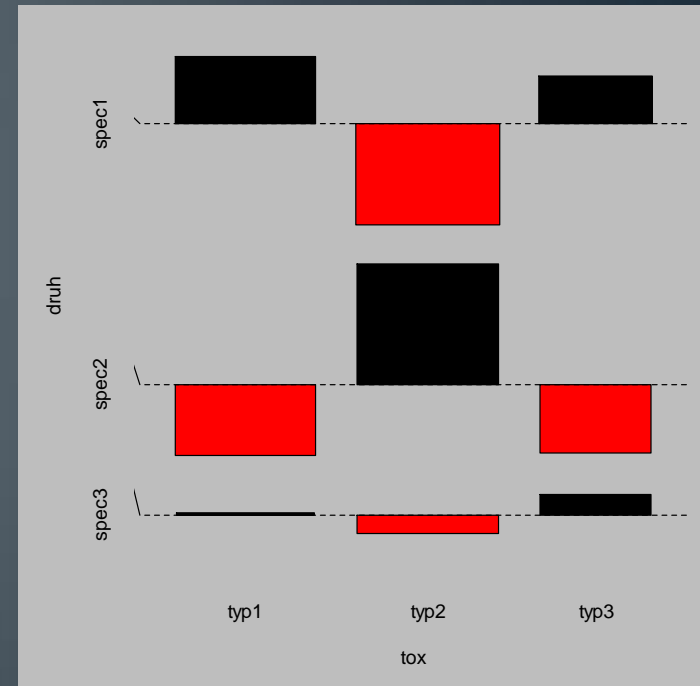
OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

`assocplot` – graficky zobrazuje vztahy mezi očekávanými a naměřenými hodnotami v kontingenční tabulce.

V případě že jsou naměřené hodnoty vyšší než očekávané, obdélník se vykreslí černě, v opačném případě červeně.

```
tabb<-  
as.table(array(c(35,26,22,26  
,39,16,30,30,20),c(3,3),  
+  
list(tox=c("typ1","typ2","ty  
p3"),druh=c("spec1","spec2",  
"spec3"))))  
assocplot(tabb)
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



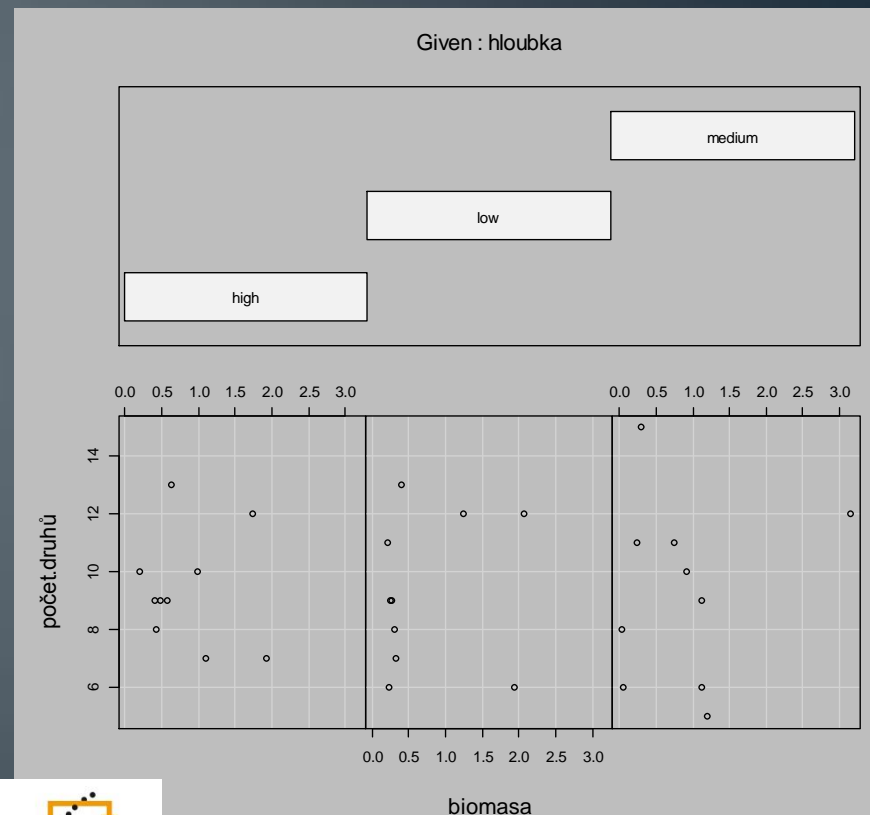
OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Coplot – xy graf znázorňující závislost dvou proměnných, rozdělených podle určité úrovně.

```
tabb<-  
data.frame(hloubka=rep(c("low", "medium", "high"), 10), biomasa=abs(rnorm(30)),  
+ počet.druhů=rpois(30, 10))
```

```
coplot(počet.druhů ~  
biomasa | hloubka, data =  
tabb, rows=1)
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



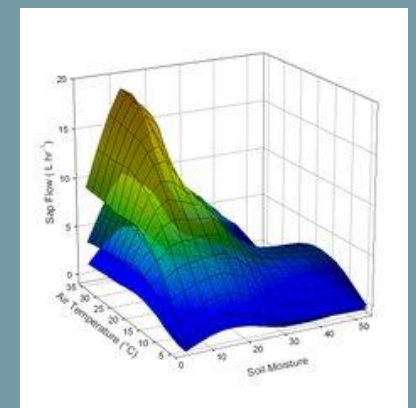
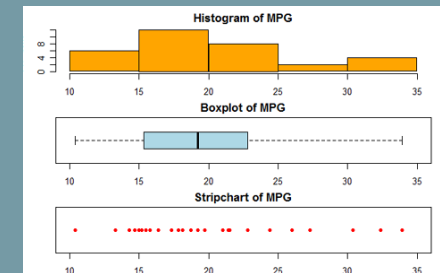
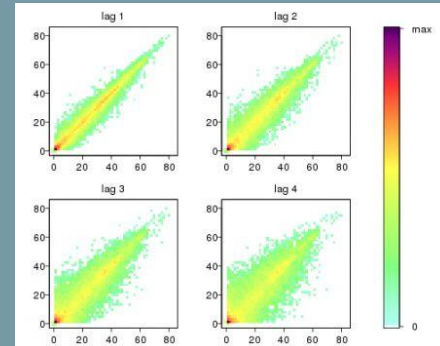
MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Základy práce v prostředí R 4



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Rozdělení četnosti

`table ()` umožňuje rozdělení jednotlivých částí variační řady, funkce vypočítává absolutní četnosti

```
x<-rpois(80,2)
```

```
čet<-table(x)
```

Výpočet relativních četností

```
r.čet<-čet/sum(čet)
```

Výpočet kumulativní četnosti

```
cum.čet<-cumsum(čet)
```

Výpočet relativní kumulativní četnosti

```
cum.r.čet<-cumsum(r.čet)
```

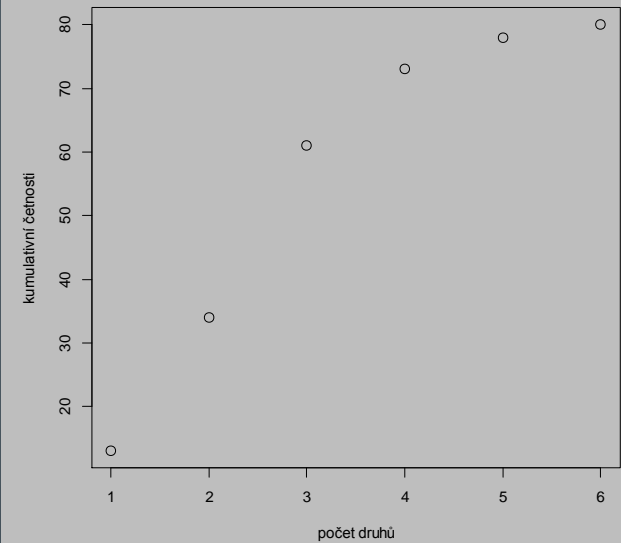
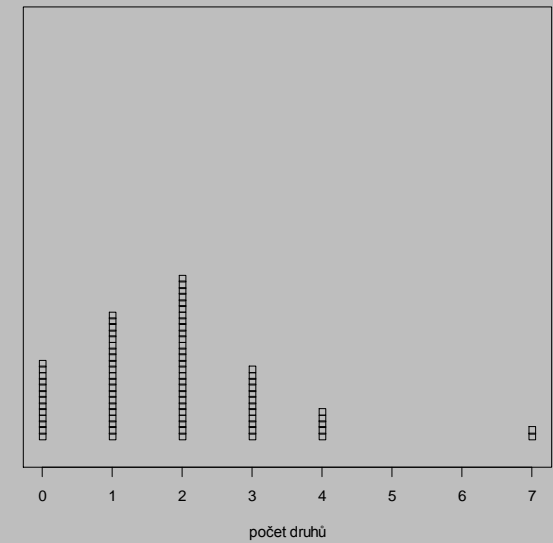
Tabulka četností

```
tabb<- cbind(n=čet, f=r.čet, N=cum.čet, F=cum.r.čet)
```





```
dotchart (sort (x), xlab="počet  
druhů")  
stripchart (x, "stack", xlab="počet  
druhů", ylim=c(0, 30))
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Intervalové rozdělení četnosti

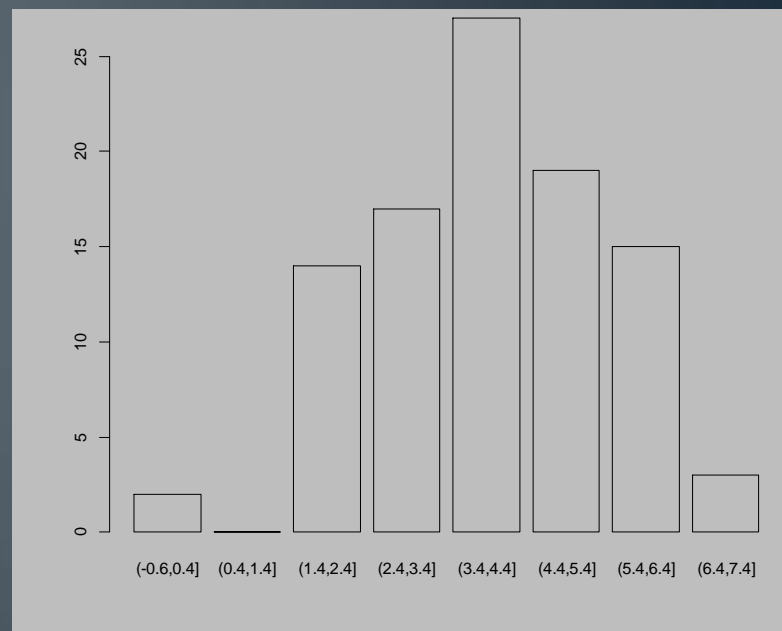
```
xn<-rnorm(100,4,1.5)
range(xn) vybere minimální a
maximální hodnotu
range(x)[2]-range(x)[1]
variační rozpětí
seq(-0.6, 7.9,1) vytvoříme si
vlastní intervaly četnosti
pretty(xn, n=8) automatický se
vytvoří intervaly četnosti
```

```
čet<-cut(xn,breaks=seq(-0.6,
7.9,1), include=T) výpočet
četností
```

```
abs<-table(čet)
```

Výpočet intervalového rozdělení četností pomocí funkce `hist`

```
hist(xn,plot=F)
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Typy rozdělení dat

## Diskrétní rozdělení náhodné proměnné

### Binomické rozdělení

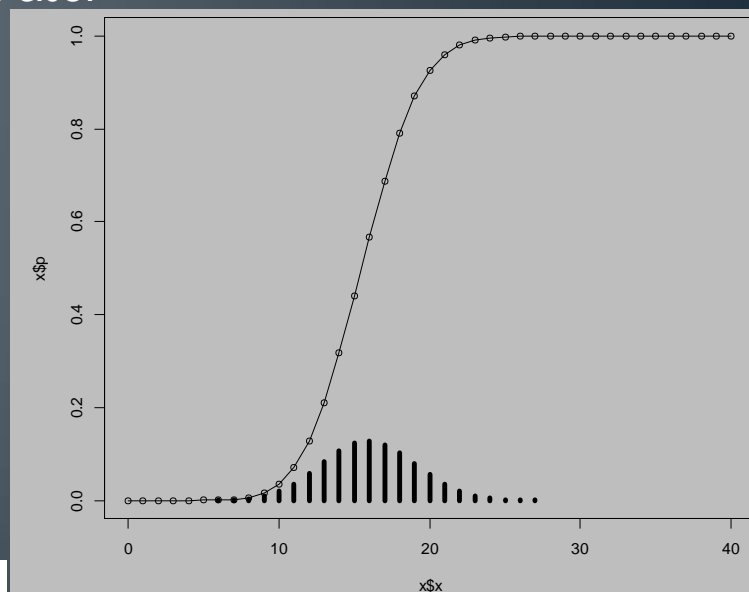
`dbinom (x, size, prob)` – rozdělení pravděpodobnosti

`pbinom (q, size, prob)` – distribuční funkce

`qbinom (p, size, prob)` – kvantily rozdělení

`rbinom (n, size, prob)` – generátor náhodných čísel

```
x<-  
data.frame(x=0:40, p=dbinom(0:40,  
40, 0.4), F=pbinom(0:40, 40, 0.4))  
plot(x$x, x$p, type="h", ylim=c(0, 1),  
lwd=5)  
lines(x$x, x$F, type="o")
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Poissonovo rozdělení

`dpois (x, lambda)` – rozdělení pravděpodobnosti

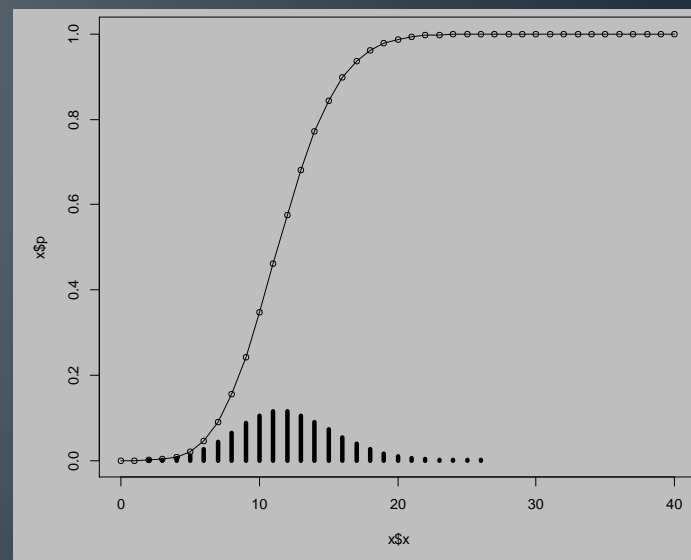
`ppois (q, lambda)` – distribuční funkce

`qpois (q, lambda)` – kvantily rozdělení

`Rpois (n, lambda)` – generátor náhodných čísel

Lambda – průměrný počet jevů v čase nebo prostoru

```
x<-  
data.frame(x=0:40,p=dpois(0:40,12),F=ppois(0:40,12))  
plot(x$x,x$p,type="h",ylim=c(0,1),lwd=5)  
lines(x$x,x$F,type="o")
```



## Negativně binomické rozdělení

`dnbinom (x, size, prob, mu)` – rozdělení  
pravděpodobnosti

`pnbinom (x, size, prob, mu)` – distribuční funkce

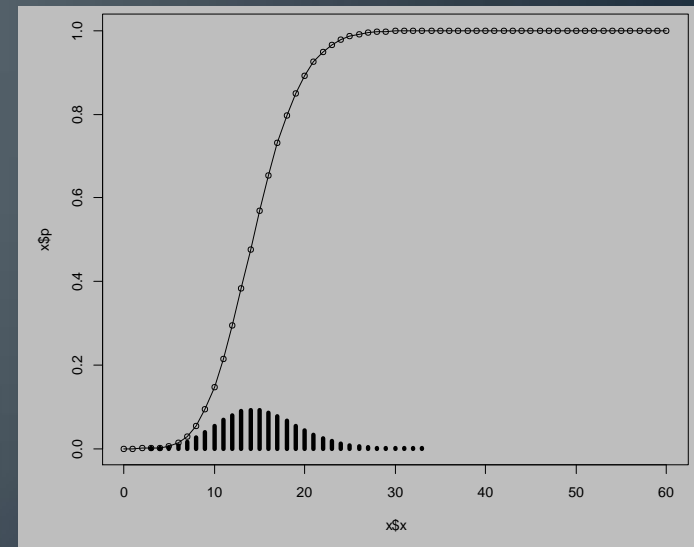
`qnbinom (p, size, prob, mu)` – kvantily rozdělení

`rnbinom (p, size, prob, mu)` – generátor  
náhodných čísel

`prob` – pravděpodobnost úspěchu

`mu` – alternativní zápis pomocí průměru

```
x<-  
data.frame(x=0:60,p=dnbinom(0:60  
,60,0.8),F=pnbinom(0:60,60,0.8))  
plot(x$x,x$p,type="h",ylim=c(0,1  
) ,lwd=5)  
lines(x$x,x$F,type="o")
```



# Spojité rozdělení náhodné proměnné

## Normální rozdělení

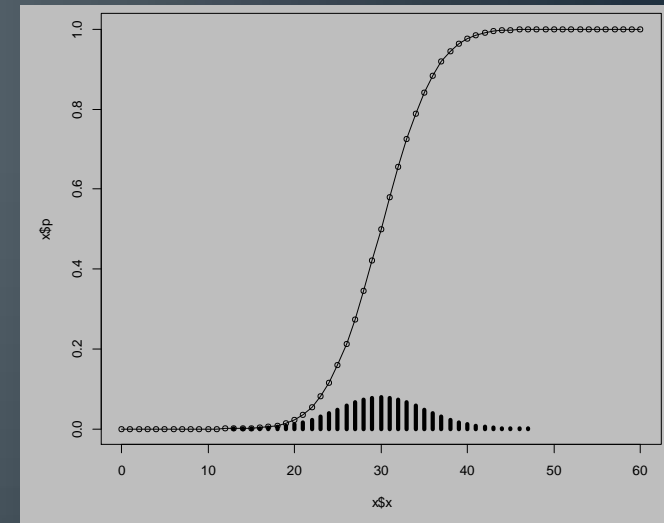
`dnorm (x, mean, sd)` – hustota  
pravděpodobnosti

`pnorm (q, mean, sd)` – distribuční funkce

`qnorm (p, mean, sd)` – kvantil rozdělení  
náhodné proměnné

`rnorm (n, mean, sd)` – generátor  
náhodných čísel

```
x<-  
data.frame(x=0:60, p=dnorm(0:60  
, 30, 5), F=pnorm(0:60, 30, 5))  
plot(x$x, x$p, type="h", ylim=c(0  
, 1), lwd=5)  
lines(x$x, x$F, type="o")
```



# $\chi^2$ rozdělení

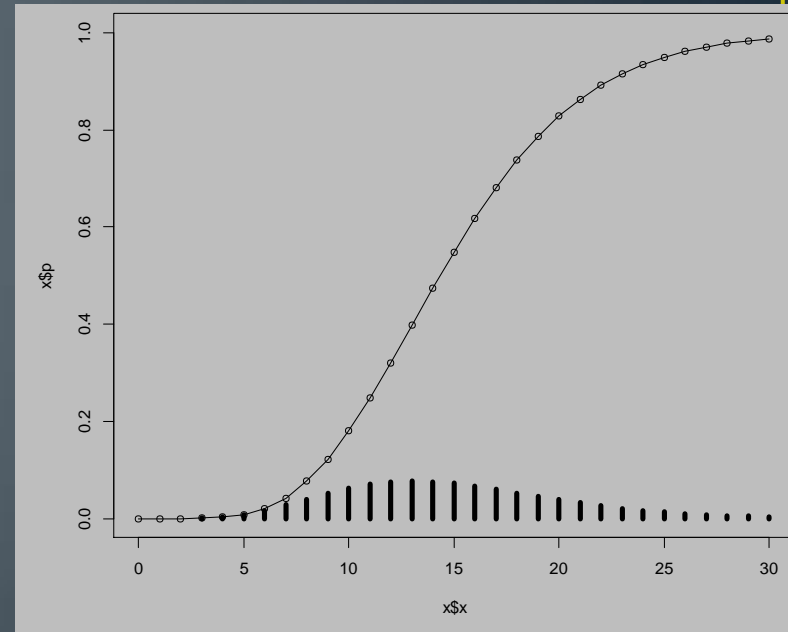
`dchisq (n, df)` – hustota  
pravděpodobnosti

`pchisq (q, df)` – distribuční funkce

`qchisq (p, df)` – kvantil rozdělení náhodné  
proměnné

`rchisq (n, df)` – generátor náhodných čísel

```
x<-  
data.frame(x=0:30,p=dchisq  
(0:30,15),F=pchisq(0:30,15  
) )  
plot(x$x,x$p,type="h",ylim  
=c(0,1),lwd=5)  
lines(x$x,x$F,type="o")
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

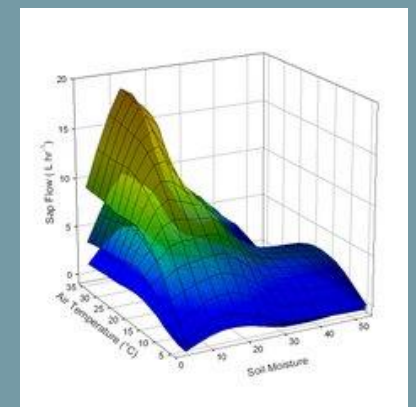
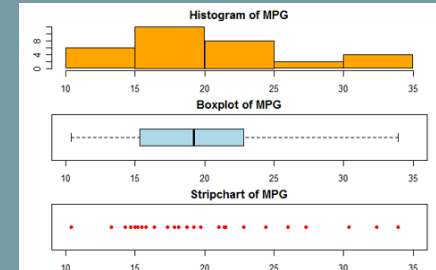
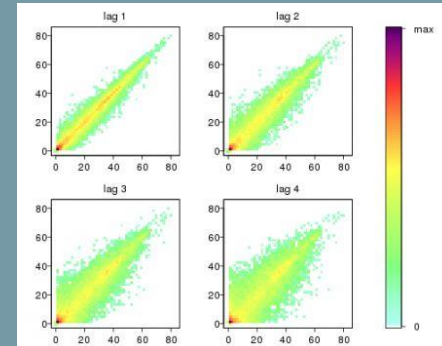


OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



# Základy práce v prostředí R 5



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ



# Základní charakteristiky dat

## exploatační analýza dat

### Charakteristiky polohy

`min ()`

`max ()`

`median ()`

průměr

`mean ()`

Vážený průměr

`weighted.mean (x, w)` x- hodnoty, w- jejich váhy

`x<-rpois (0:2000, 50)`

`quantile (x, prob=c (0.2, 0.5, 0.6, 0.8))`

Harmonický průměr

`harm<-function (x) length(x) /sum (1/x)` - výpočet rychlosti

Geometrický průměr

`geom<-function (x) prod(x)^(1/length(x))` - výpočet průměrného růstu



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

## Charakteristiky variability

`Var ()` rozptyl

`Sd ()` směrodatná odchylka

`fivenum ()` počítá minimum, dolní kvantil, median, horní kvantil, maximum

`boxplot ()` funkce `boxplot` je jednou z nejpoužívanějších v exploatační analýze dat

```
x<-rpois(0:2000,50)
```

```
boxplot.stats(x)
```

Počítá charakteristiky polohy a variability

`$stats` minimální (30) a maximální hodnota (70), dolní kvantil (45) horní kvantil (55), medián je (50)

`$n` počet hodnot

`$conf` výpočet konfidenčních intervalů

`$out` odlehlé hodnoty



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

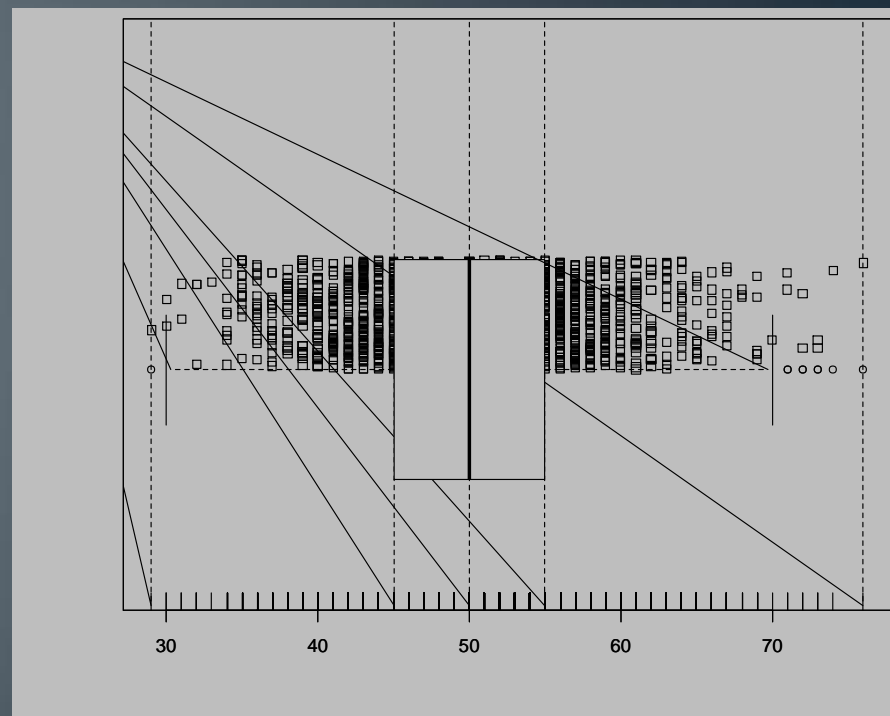


OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Boxplot a stripchart

```
stripchart (x,"jitter")  
abline (v=fivenum (x) ,lty=2)  
boxplot (x,horizontal=T,add=T  
,at=0.9)  
rug (x)
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

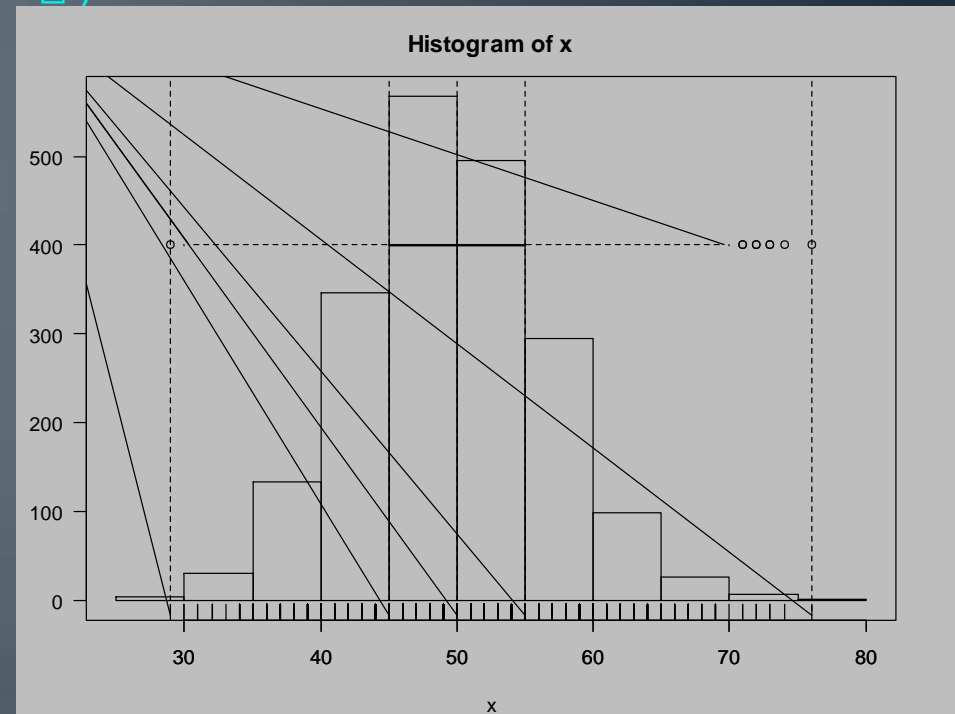


OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

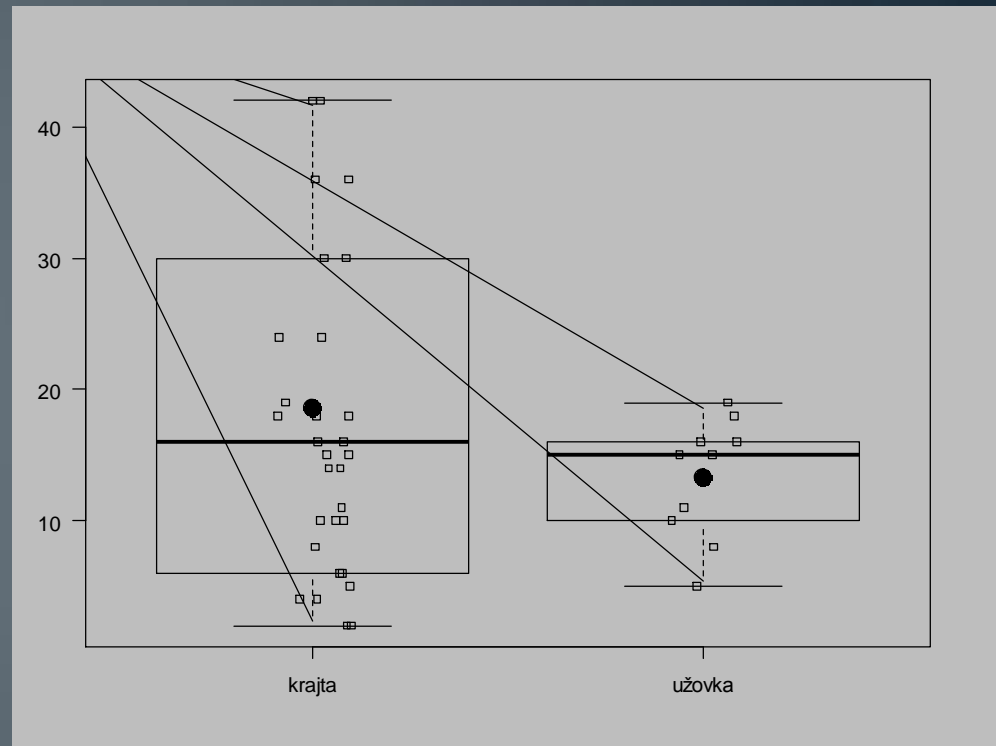
# Histogram

```
hist(x, ylab="četnost")  
boxplot(x, horizontal=T,  
add=T, boxwex=4, at=400)  
abline(v=fivenum(x), lty=2)  
rug(x)
```



# boxplot

```
krajta<-  
c(12,14,16,20,24,28,24  
,30,36,42)  
užovka<-  
c(5,8,10,11,15,18,16,1  
5,19,16)  
boxplot(cbind(krajta,u  
žovka),ylab="hmotnost  
těla")  
stripchart(list(krajta  
,užovka),"jitter",vert  
ical=T,add=T,cex=0.8)  
points(1:2,c(mean(kraj  
ta),mean(užovka)),pch=  
20,cex=3)
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

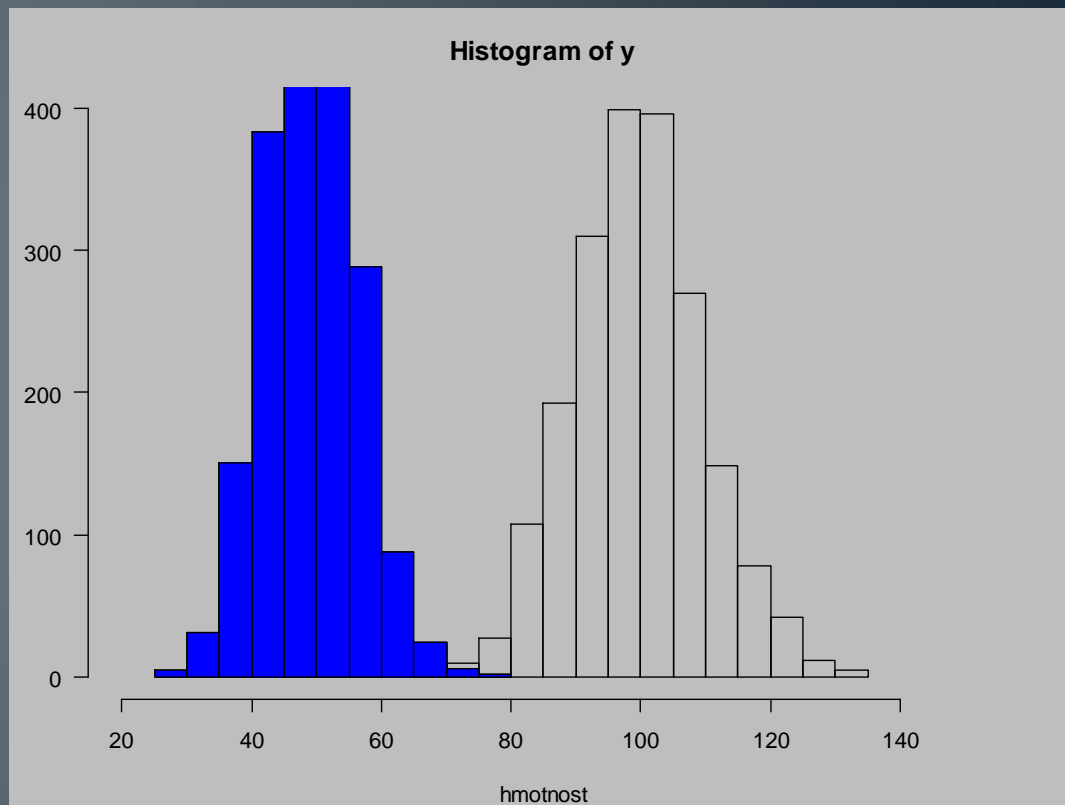


OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

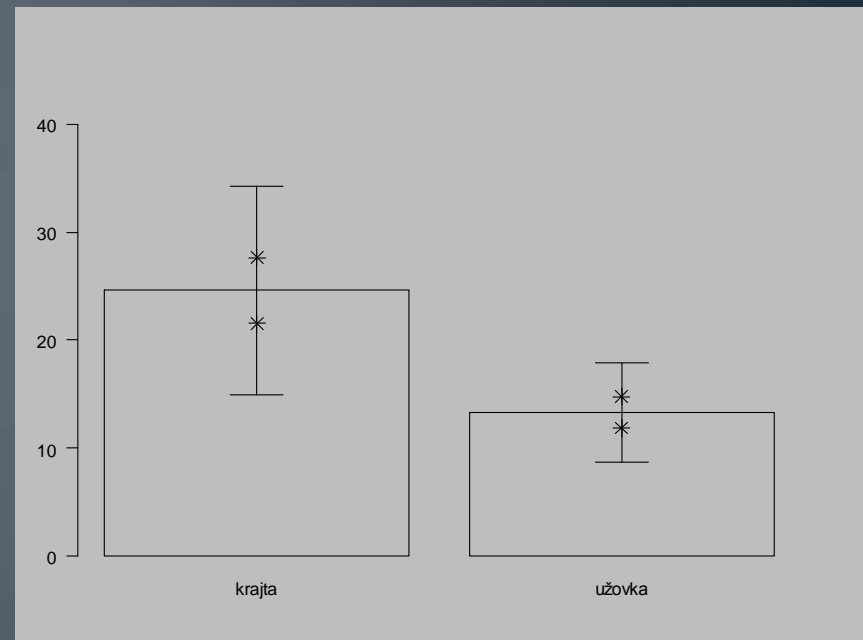
# četnosti

```
x<-rpois(0:2000,50)
y<-rpois(0:2000,100)
hist(y,xlab="hmotnost",ylab="četnosti",x
lim=c(20,150))
hist(x,add=T,col="blue")
```



# Sloupcový graf

```
avg<-  
c(mean(krajta),mean(užovka))  
sdd<-  
c(sd(krajta),sd(užovka))  
střed<-barplot(avg,plot=F)  
barplot(avg,names.arg=c("krajta",  
"užovka"),ylab="hmontnost",ylim=c(0,45))  
arrows(střed,avg-sdd,střed,avg+sdd,angle=90,c  
ode=3)  
my.se<-  
sdd/sqrt(c(length(krajta),length(užovka)))  
points(rep(střed,2),c(avg-my.se,avg+my.se),pch=4,cex=1.5)
```



evropský  
sociální  
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání  
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ