

# Zpracování dat

Blok je věnován zpracování, základní analýze a prezentaci biologických dat (datové typy, vzorkování, popisná statistika, test normality, klouzavý průměr, typy a tvorba grafů Excel + R). Cílem tohoto bloku je poskytnout obecný informační základ k problematice a praktickou dovednost při prezentaci dat.

# 1. setkání

- cíle kurzu
- obsah kurzu
- požadavky na zápočet
- úvod ke sběru dat

# Fáze výzkumné práce

- 1) Výzkumná otázka, tázání se;
- 2) rešerše stávajícího poznání, návrh/přizpůsobení a testování metodiky;
- 3) sběr dat;
- 4) analýza dat;
- 5) prezentace dat.

# Užijeme-li statistických metod

- *hledání rozdílu* – spočívající v porovnávání paralelních datových řad. Pro jejich prezentaci užíváme sloupcové či krabicové grafy
- *hledání korelace* – spočívající ve zjišťování provázanosti jevů. Zde využíváme zpravidla bodové grafy doplněné o vizualizaci hledané/nalezené korelace
- *hledání asociace* – jež je obdobné předchozímu, až na povahu dat - korelace mají data intervalová, asociace data nominální a ordinální. Vizualizujeme sloupcovými grafy

# Úkol #1 – tabulka pro sběr dat

Sledujeme procentuelní úspěšnost v pretestu a posttestu u tříd několika základních škol (např: ZŠ Mělnická, třída 8.C měla v pretestu 70% a v posttestu 83%)

-->

*navrhněte formát tabulky do níž budete zapisovat zjištěná data.*

# Typy dat

- *intervalová* – jsou to číselná data, se kterými můžeme provádět běžné matematické operace jako je sčítání, odčítání atp. Příkladem je věk žáků, procenta či body z testu, roky školní docházky žáka (nikoli ročník!)
- *ordinální* – hodnoty takovýchto veličin jsou uspořádané, ale i v případě že jsou uváděny jako čísla, nejsou u nich uplatnitelné ony "běžné matematické operace". Můžeme říci že jedna hodnota je vyšší/větší než druhá, ale nemůžeme určit kolikrát přesně je vyšší.
- *nominální* – jako předchozí ale není možné je řadit. Nejlepším příkladem jsou pohlaví, náboženská či státní příslušnost, paralelky ročníku

# Úkol #2 – typy dat

*Rozlište v naší tabulce dílčí typy dat.*

## 2. setkání

- základy R
- popisná statistika a rozložení intervalových dat
- velikost vzorku
- příprava studentů:
  - *Instalace LibreOffice (volitelné)*  
<http://cs.libreoffice.org/download/>
  - *Instalace R* <https://cran.r-project.org/>  
-> výběr dle OS, pro Windows zvolit první, tj "base"  
nebo dočasně vystačíme s online verzí  
<http://www.r-fiddle.org/#/>



# R – první krůčky

- `365*24`
- `dny<-365`
- `hodiny<-24`
- `dny`
- `dny * hodiny -> hodinZaRok`
- `hodinZaRok`

# R – popis

- `age<-c(10,12,13,11,11,12,12,13,12,10,10)`
- `age`
- `summary(age)`
- `stem(age)`
- `boxplot(age)`

# R – popis a normalita

- `hist(age, freq=F); lines(density(age))`
- `plotHistWithNorm<-function(data){h <- hist(data, density = 10, col = "lightgray", xlab = "Proměnná", ylab="Počet výskytů", main = "Přehledovka");xfit <- seq(min(data), max(data), length = 40);yfit <- dnorm(xfit, mean = mean(data), sd = sd(data));yfit <- yfit * diff(h$mids[1:2]) * length(data);lines(xfit, yfit, col = "black", lwd = 2)}`
- `plotHistWithNorm(age)`
- `shapiro.test(age)`
  - `p-value > 0,05` ? normální : neparametrická

# R – práce na reálných datech

- `getwd()` | `setwd()`
- `data<-read.csv(file.choose(), encoding="UTF-8")`  
`#jen Windows`
- `data<-read.csv("import_data.csv", encoding="UTF-8")`
- `plotHistWithNorm(data$výsledek)`
- `pretest<-subset(data, typ_testu=="pretest",  
select=výsledek)`
- `plotHistWithNorm(pretest$výsledek)`
- `boxplot(pretest$výsledek)`
- `shapiro.test(pretest$výsledek)`

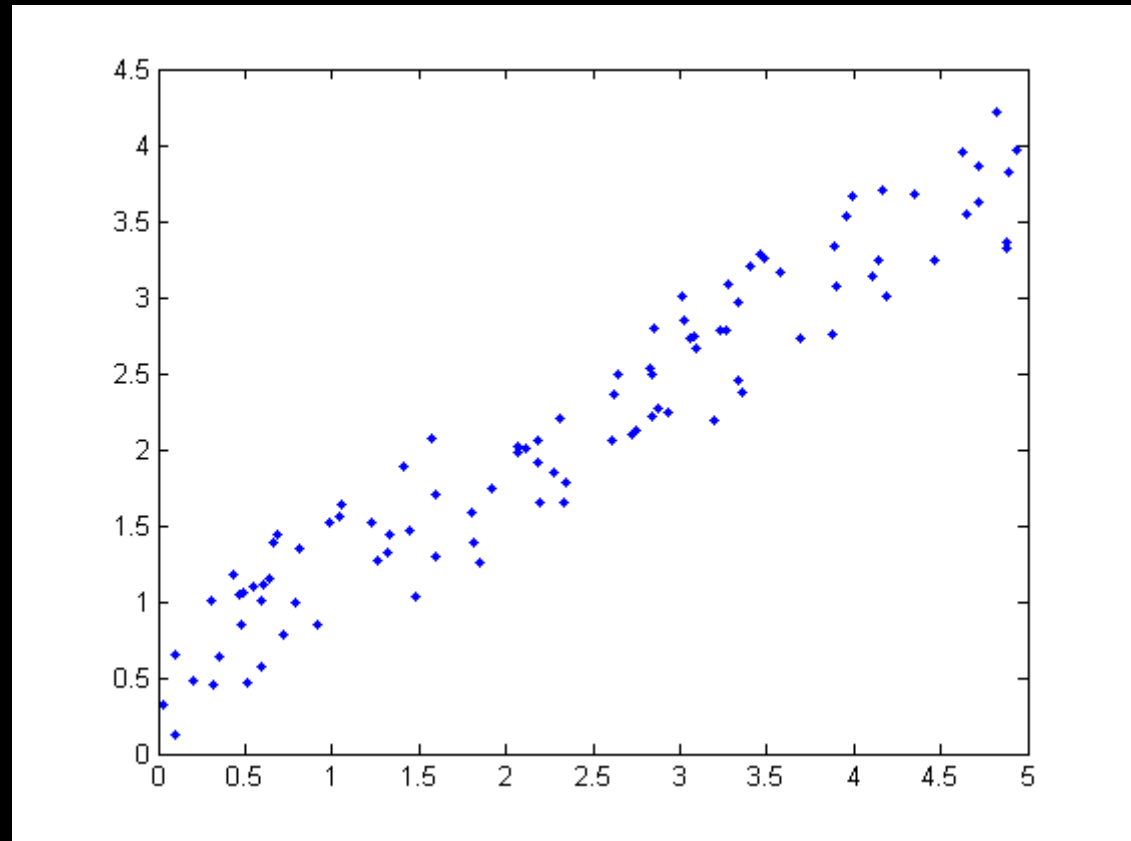
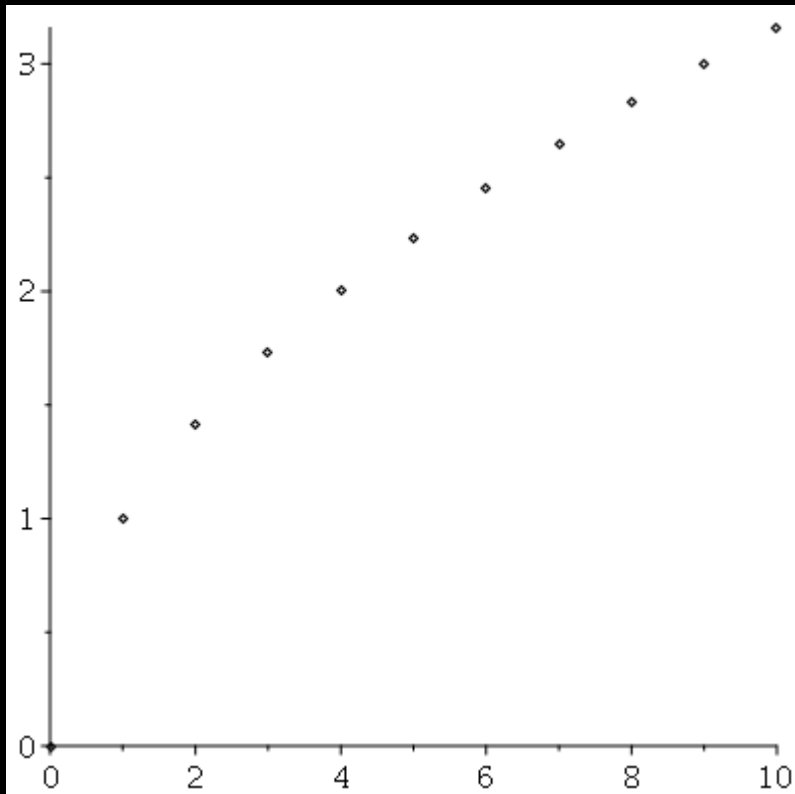
# R – velikost vzorku

- `floatMean<-function(data){plot(cumsum(data) / seq(along=data), type='l', xlab= 'Číslo měření', ylab='Plovoucí průměr'); abline(h=mean(data), lty=2, col='blue')}`
- `floatMean(pretest$výsledek)`

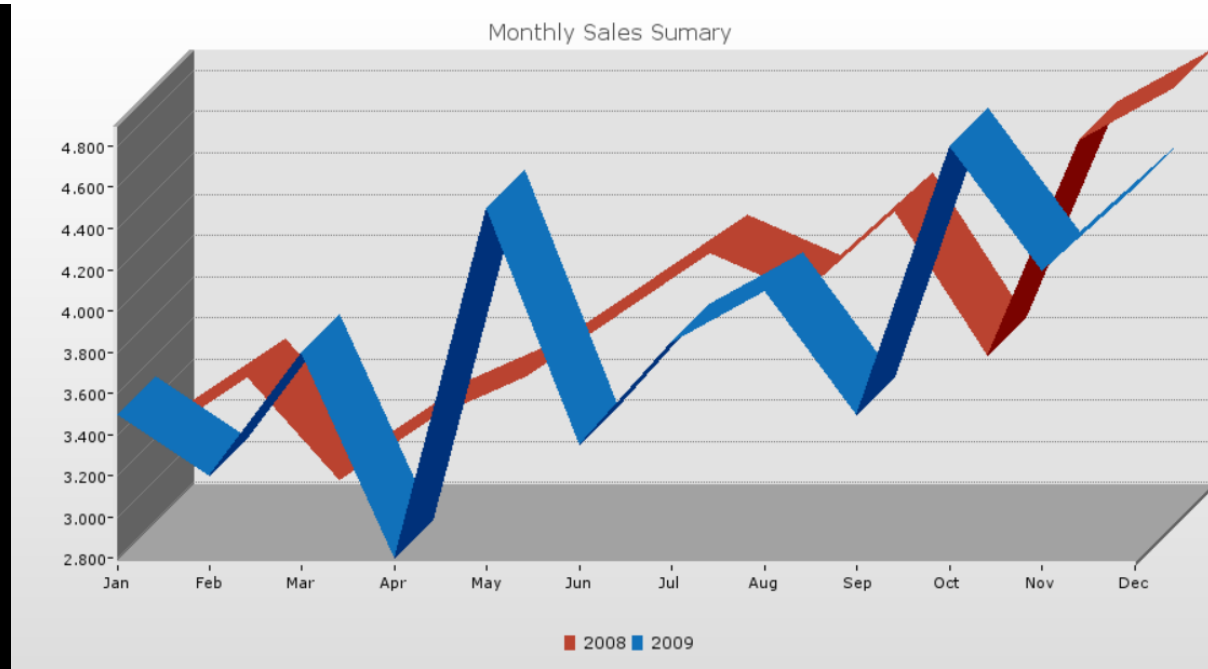
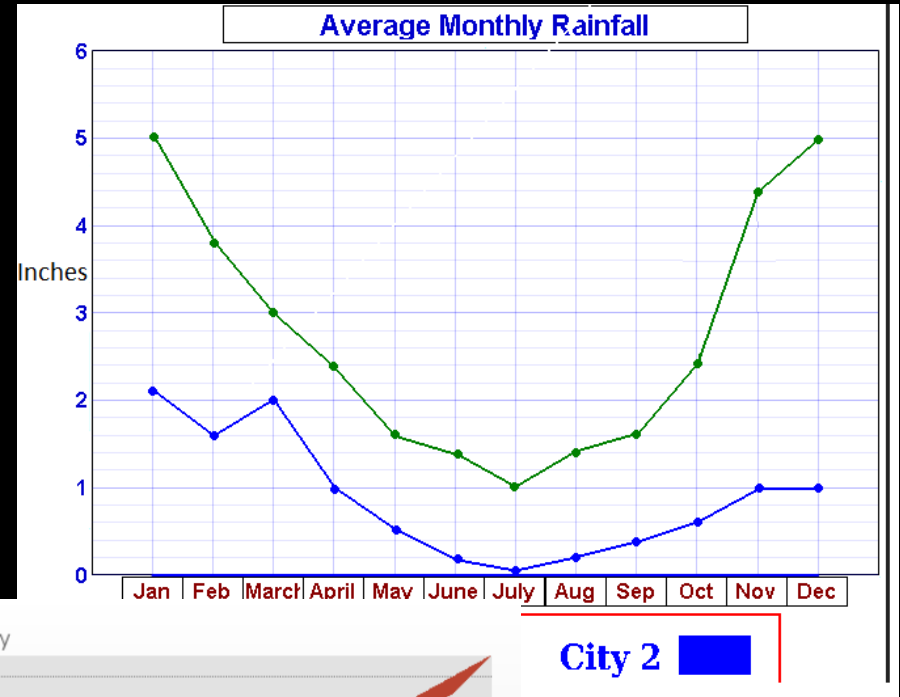
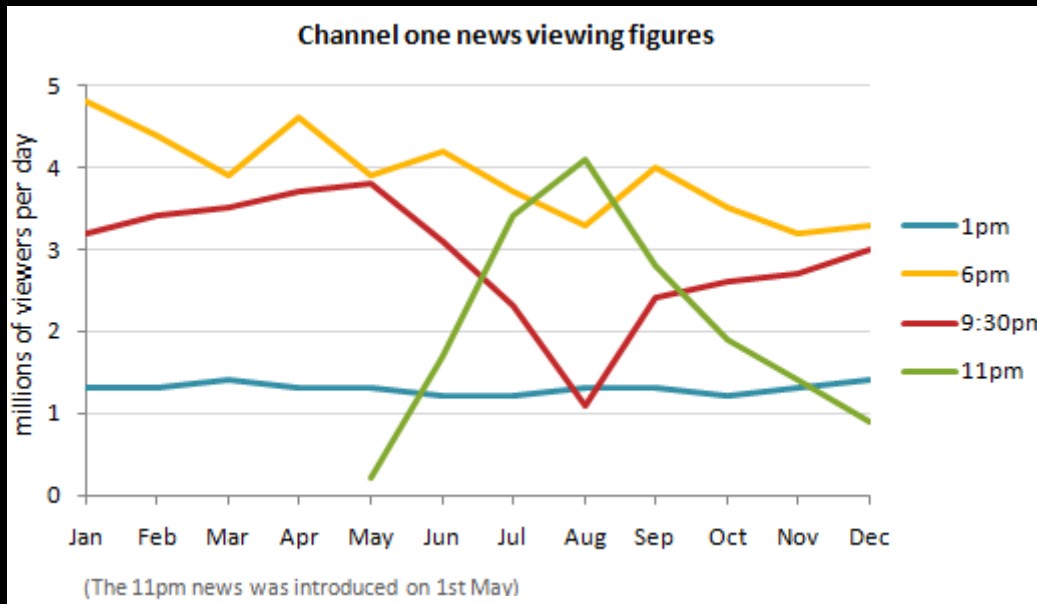
# 3. setkání

- typy grafů
- hromadné generování grafů v R
- závěrečné shrnutí

# Bodový graf [point plot]

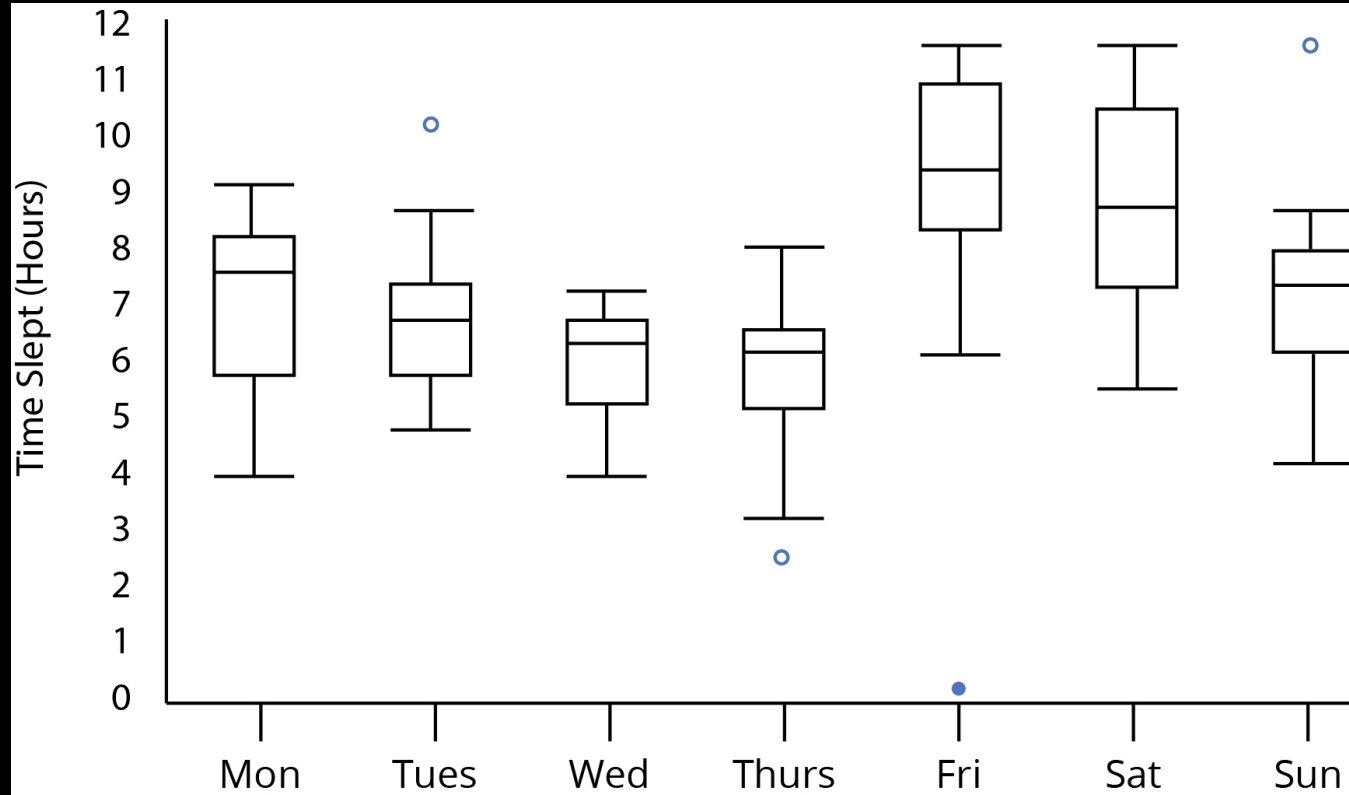
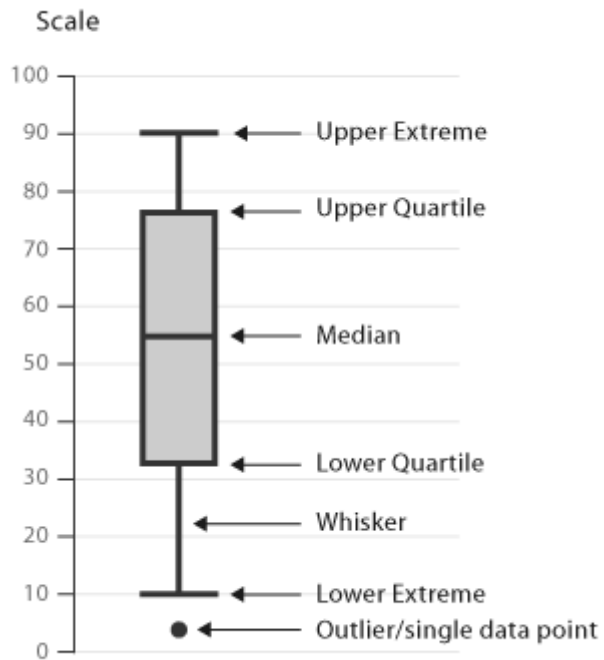


# Čárový graf [line plot]

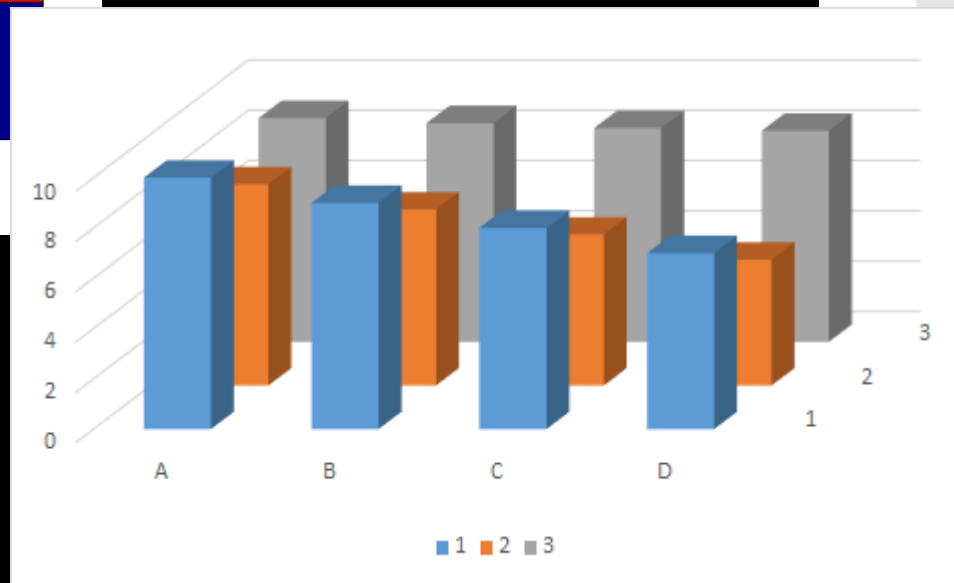
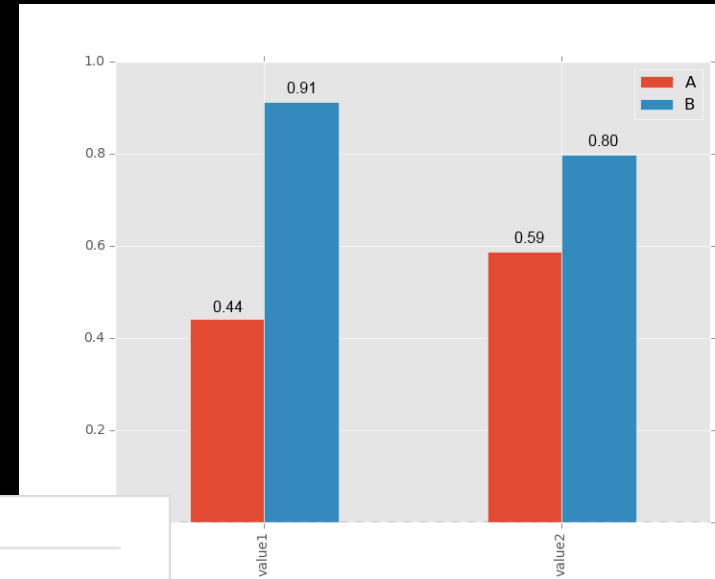
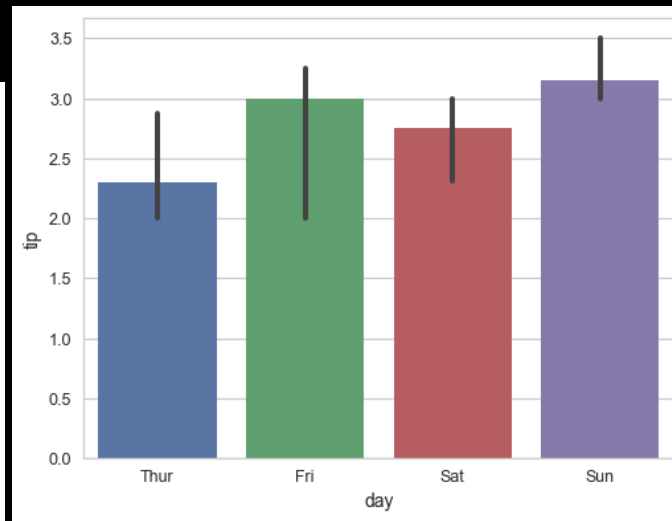
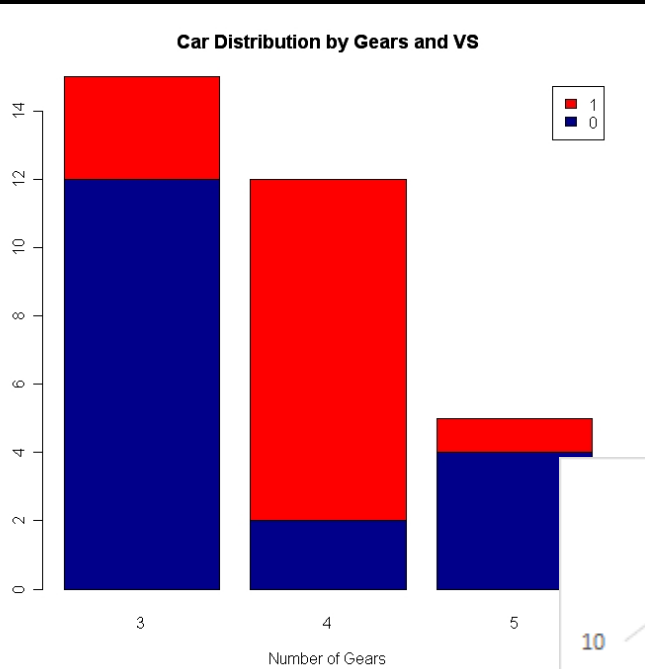




# Krabicový graf [boxplot]

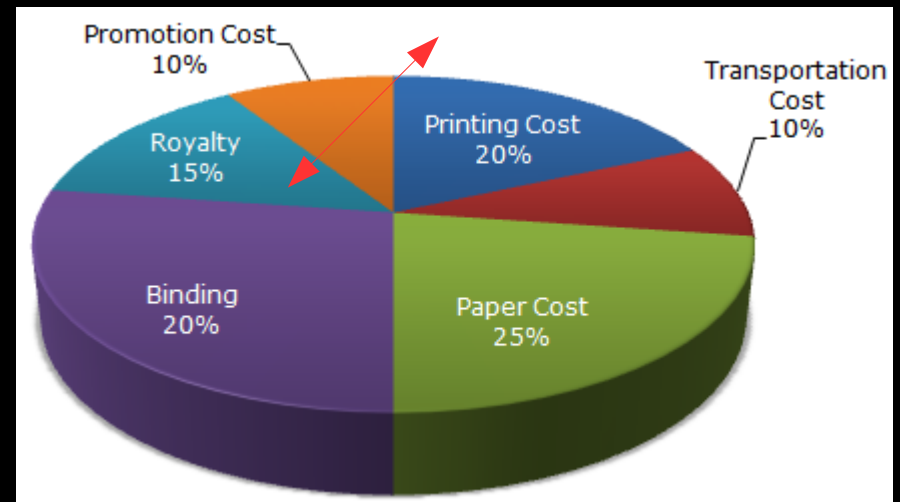
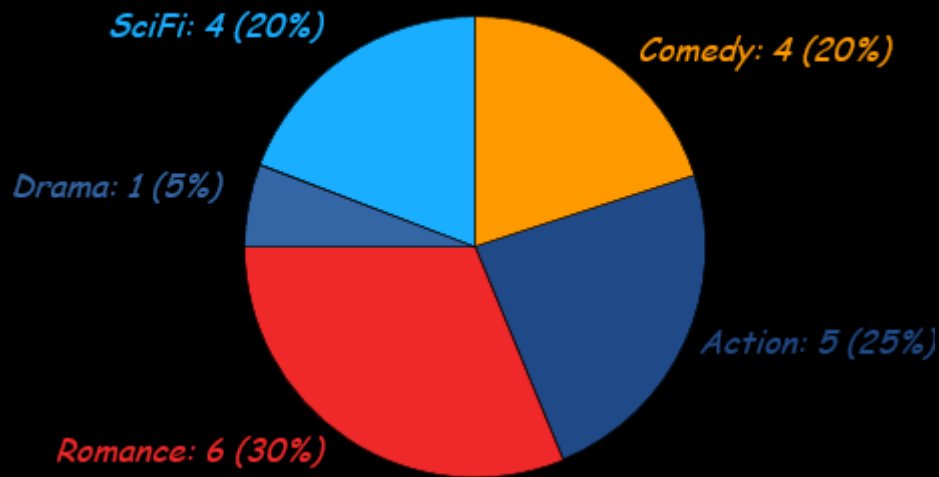


# Sloupcový graf [barplot]



# Koláčový graf [pie chart/plot]

*Favorite Type of Movie*



# Barvy v R

- `age<-c(10,12,13,11,11,12,12,13,12,10,10)`
- `barvy<-  
c("#FFF840", "#FFDD00", "#FFA840", "#FF7400",  
"#FF5D40", "#FF0D00")`
- `barvy<-heat.colors(6)`  
`barvy <- rainbow(6, s = 0.5)`
- `hist(age, col=barvy)`
- `data<-read.csv(file.choose(), encoding="UTF-8")`

# Hromadný export grafů

- `getwd()`
- `outputDir <- "seminarICT";  
dir.create(file.path(getwd(), outputDir),  
showWarnings = FALSE); outputPath <-  
file.path(getwd(), outputDir)`
- `png(paste(outputPath, "graf1.png", sep = "/"),  
width=190,  
height=220,  
units="mm",  
res=300) ;  
boxplot(age, col=barvy);  
dev.off()`

# Závěrečný test

- projděte zadaný kód a pokuste se odhadnout jeho základní funkci
- úspěšně ho spusťte na svém počítači

# 4. setkání

- GoogleMaps, Mapy.cz
- principy WMS/WFS služeb