MUNI | RECETOX

Úvod do matematické biologie a biomedicíny

přednáška 02.12.2019

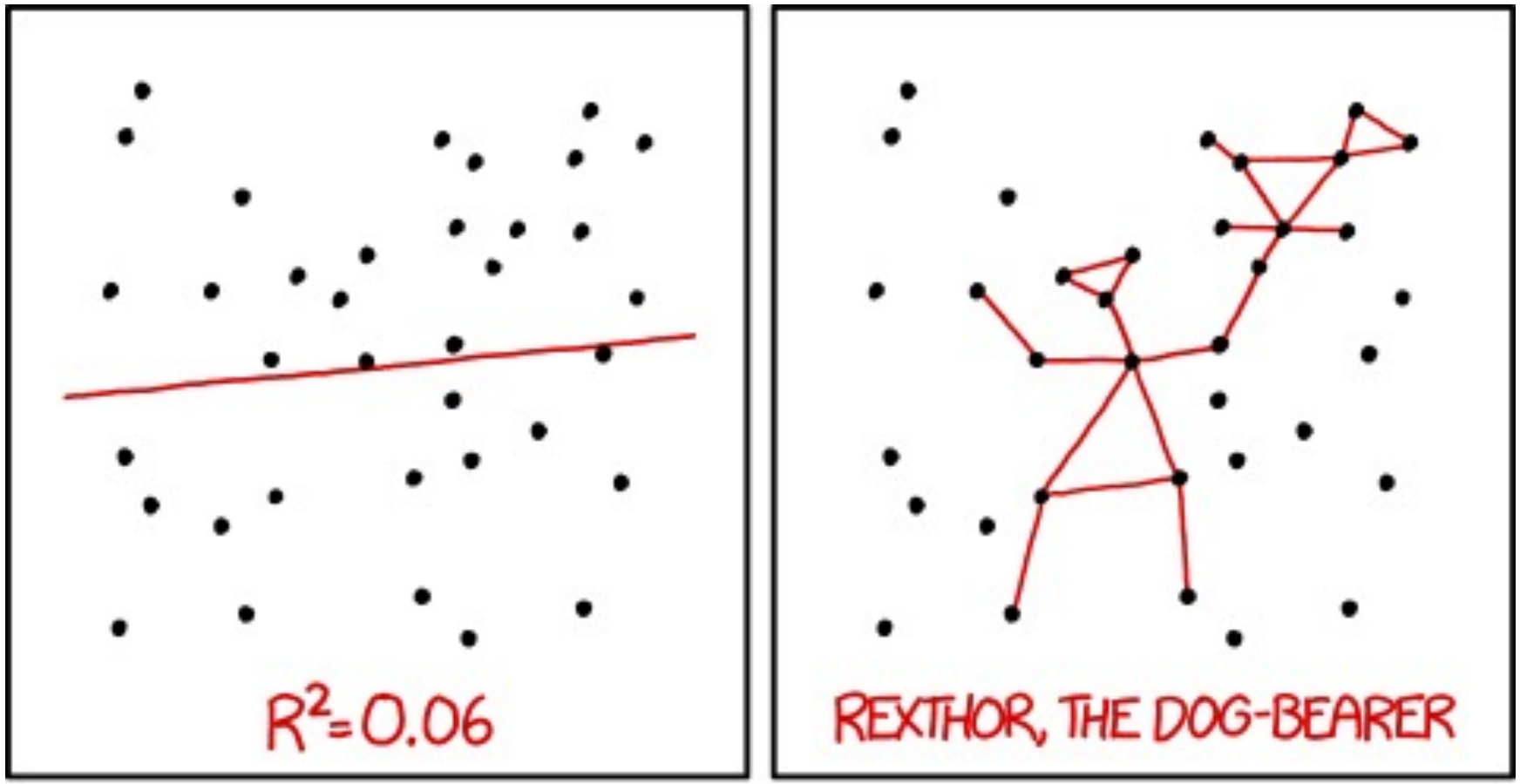
Eva Budinská (budinska@recetox.muni.cz)

MUNI | RECETOX

Co všechno (ne)lze vyčíst z grafů

Síla grafického znázornění dat (nejen) v biologii a medicíně

MUNI | RECETOX

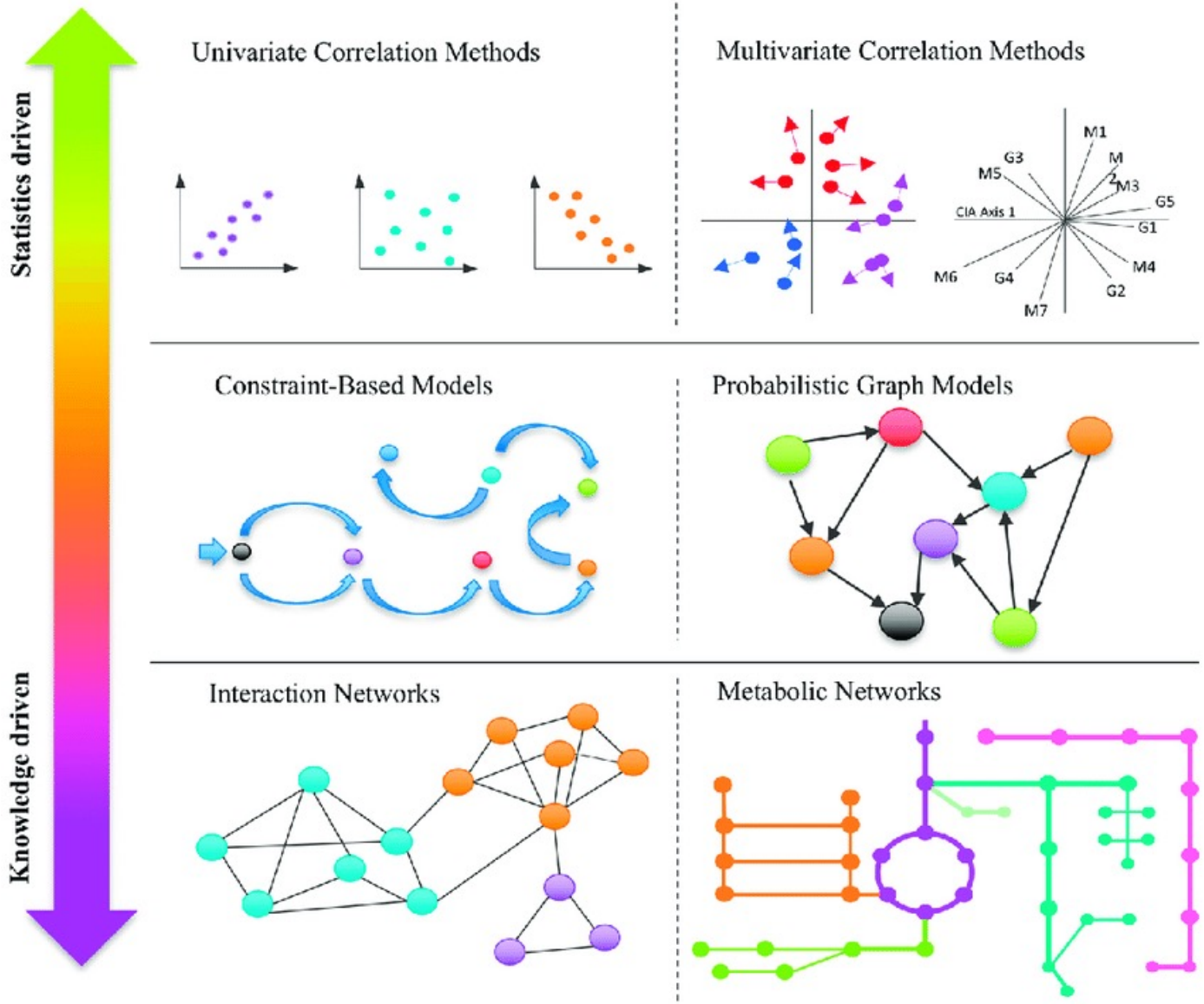


I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

- Zdroj.

Grafické znázornění dat

Nejdůležitější nástroj analýzy a komunikace výsledků!



Základné vlastnosti dobrého grafu

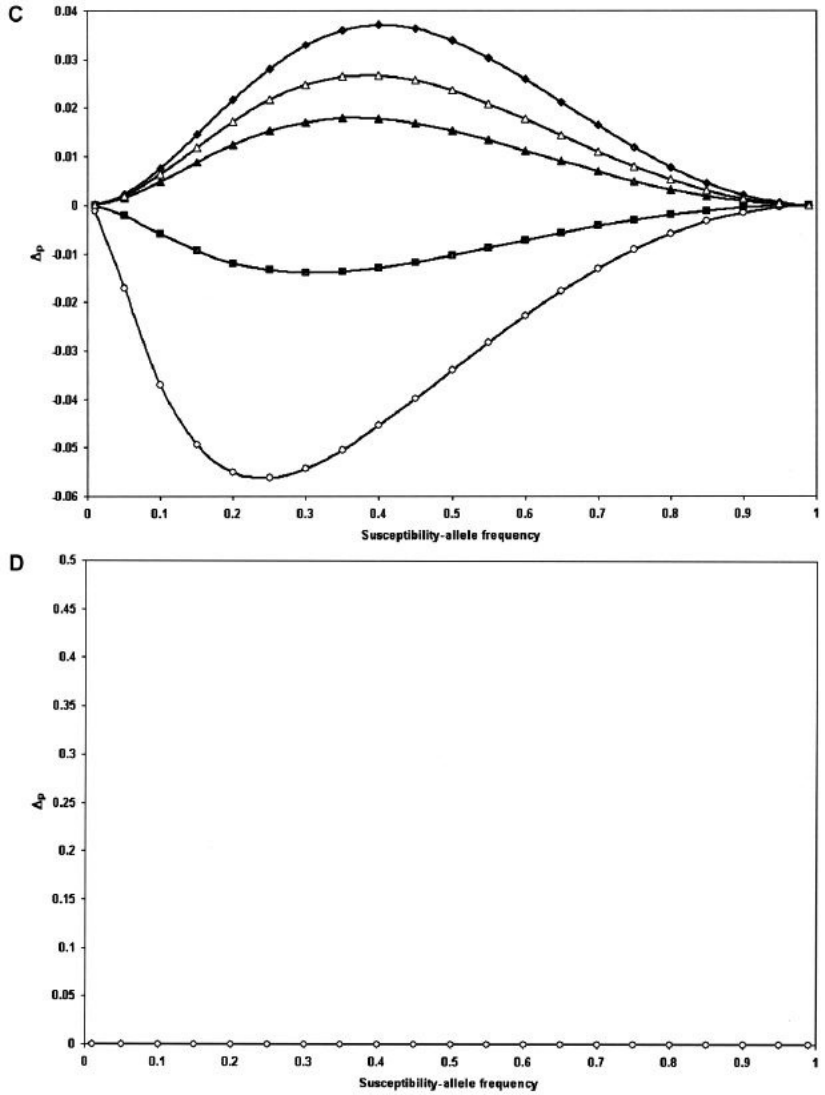
- Obsah!
- Jednoduchost
- Ne(zkreslení)

Obsah

- Graf a jeho legenda musí obsahovat všechny důležité informace

Obsah

- Žádný rozdíl nemusí mít vždy význam zobrazovat



Wittke Thompson, JK,Pluzhnikova,CoxNJ (2005) Rational inferences about departures from Hardy-Weinberg equilibrium. American Journal of Human Genetics 76:967T986,Figure1

Obsah

- Žádný rozdíl nemusí mít vždy význam zobrazovat

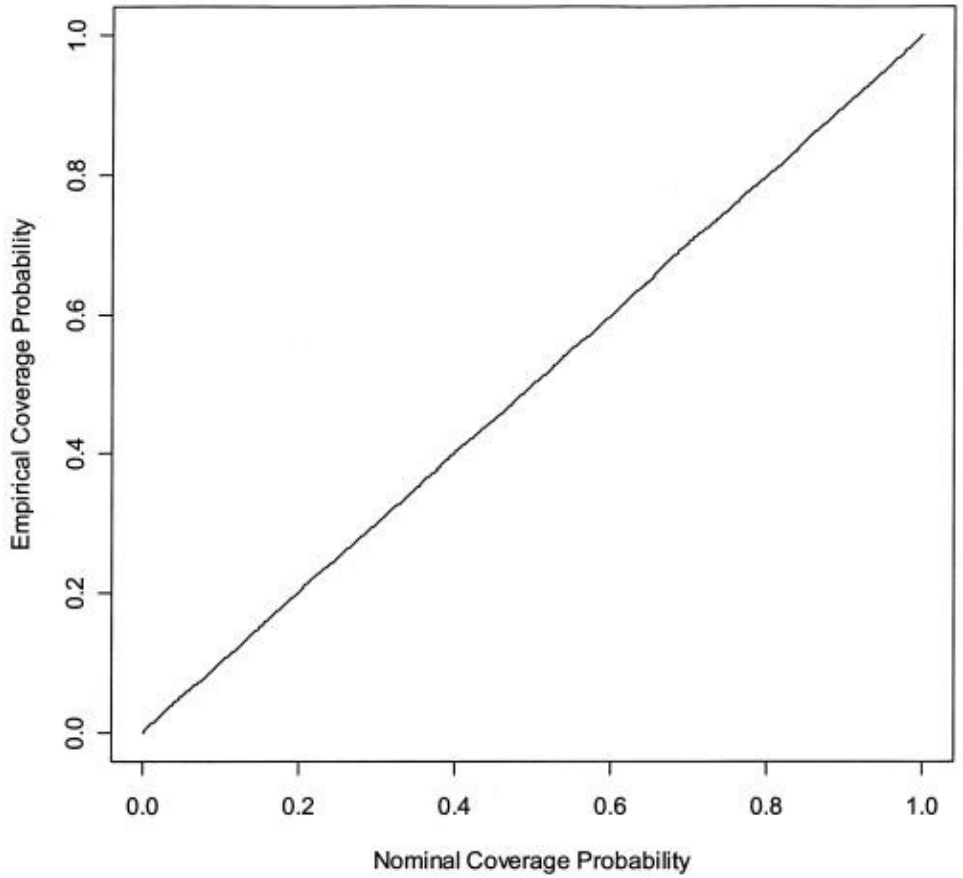


Figure 1. Empirical coverage of CIs for the relative-risk parameter β of haplotype 01100. Results are based on 10,000 simulated data sets with the same haplotype frequencies as the FUSION data. Haplotype 01100 has a multiplicative effect on disease risk, with β=0.35.

EpsteinMP,Satten GA (2003) Inference on haplotype e9ects in case-control studies using unphased genotype data. American Journal of Human Genetics 73:1316T1329, Figure 1

Jednoduchost

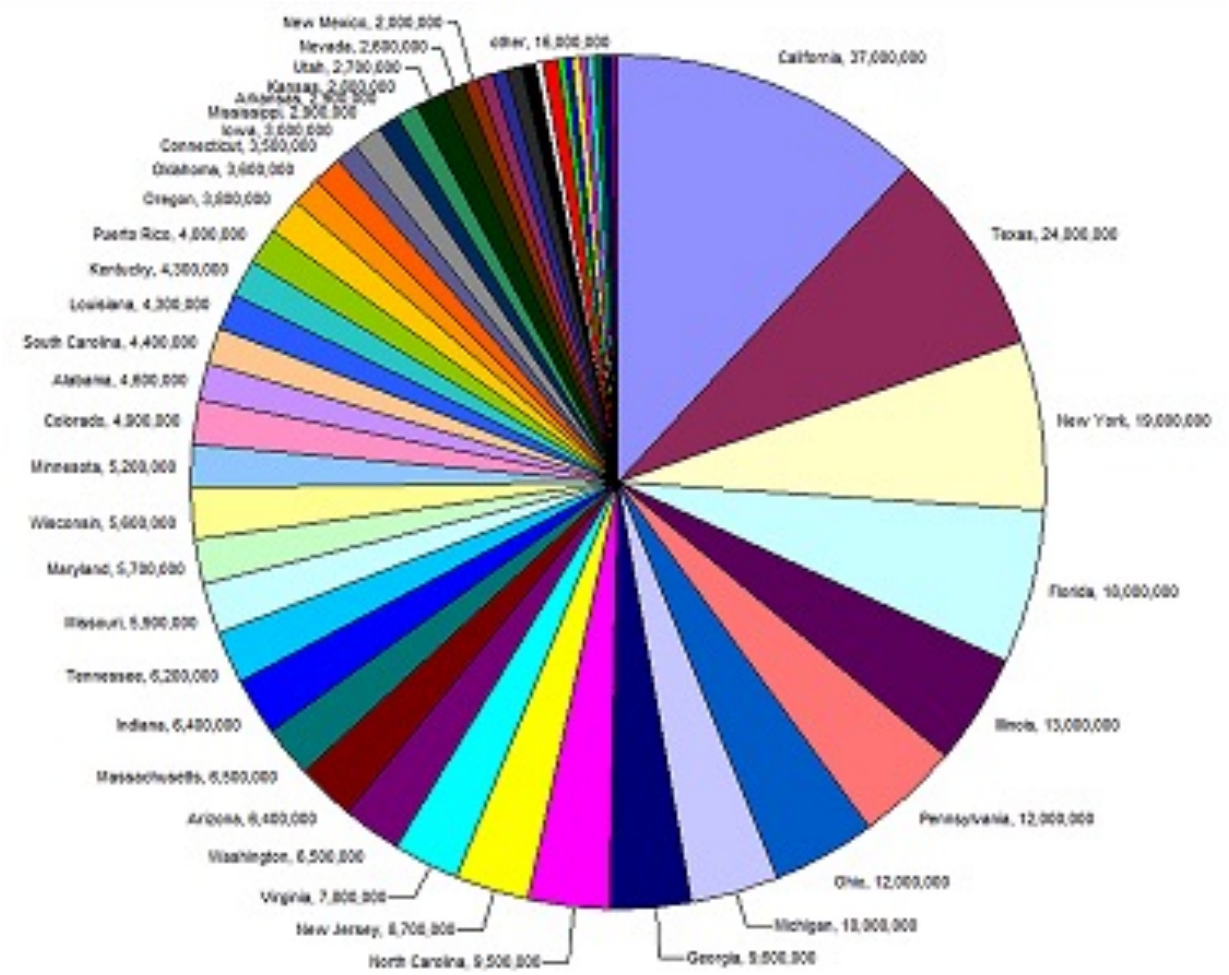
- Dobrý graf není složitější, než informace v něm obsažená
- Graf by měl **mít vysoký poměr** data / inkoust:

$$\frac{\text{Množství inkoustu použitého k zobrazení dat}}{\text{Celkové množství inkoustu použitého k zobrazení grafu}}$$

Jak zkomplikovat graf?

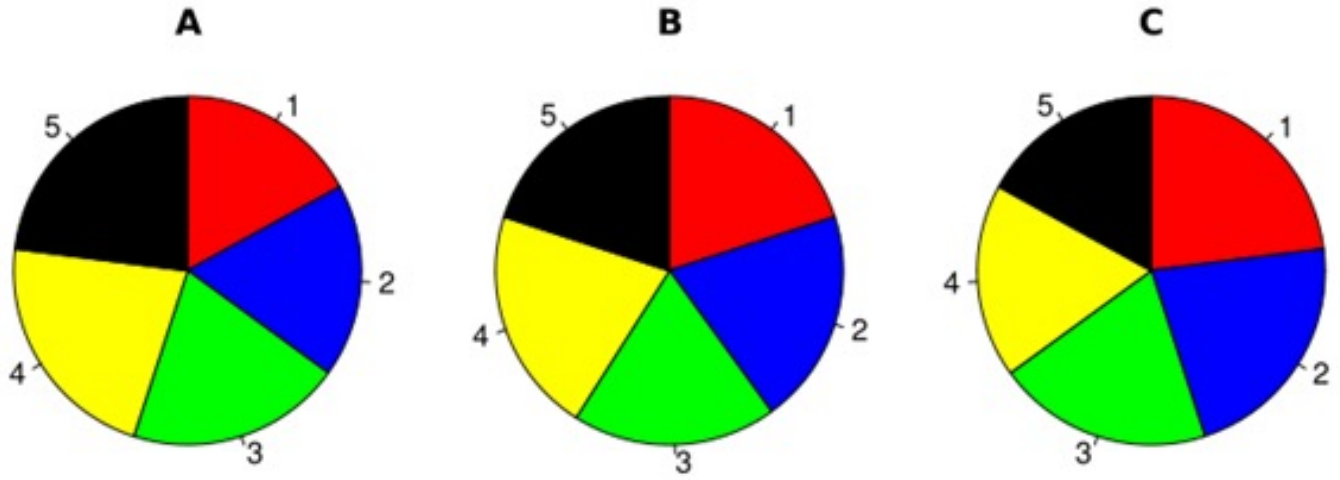
- Výběrem nevhodného zobrazení!
- Ozdobami, které nesouvisí s obsahem
- Nevhodnými a příliš četnými barvami
- Zbytečnými 3D efekty

Nevhodné zobrazení – mnoho kategorií

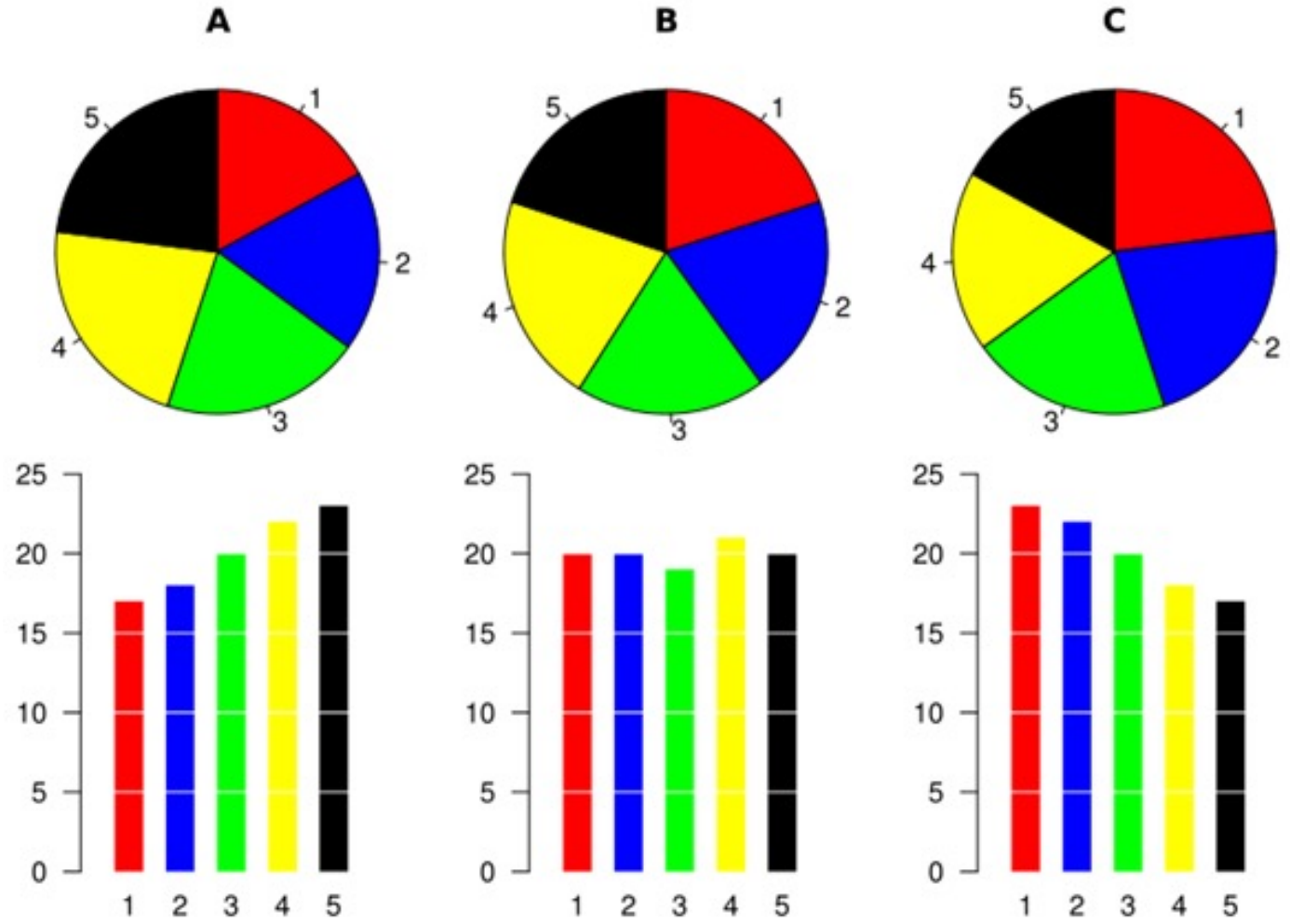


MUNI | RECETOX

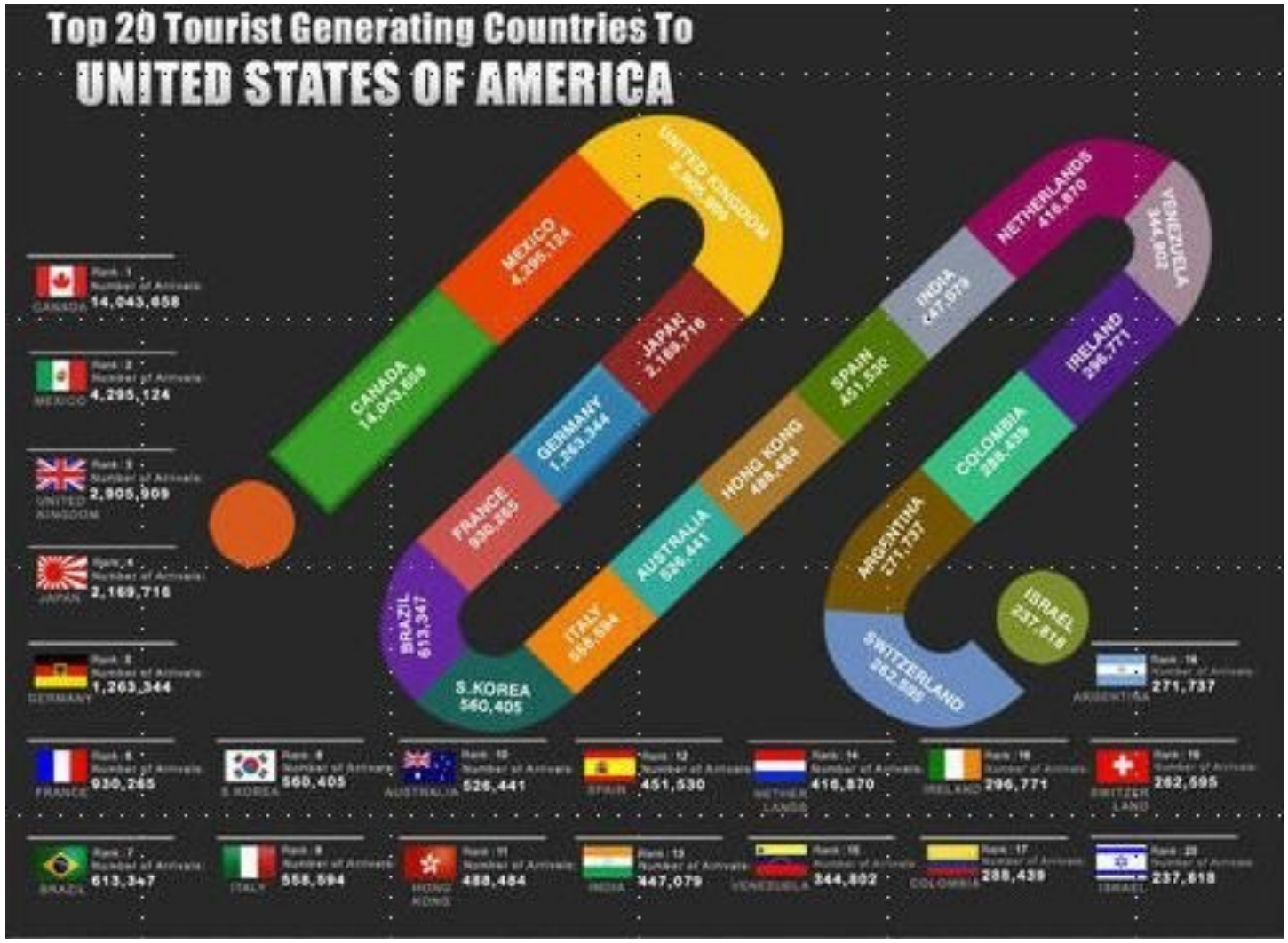
Nevhodné zobrazení – málo kategorií



Nevhodné zobrazení – málo kategorií

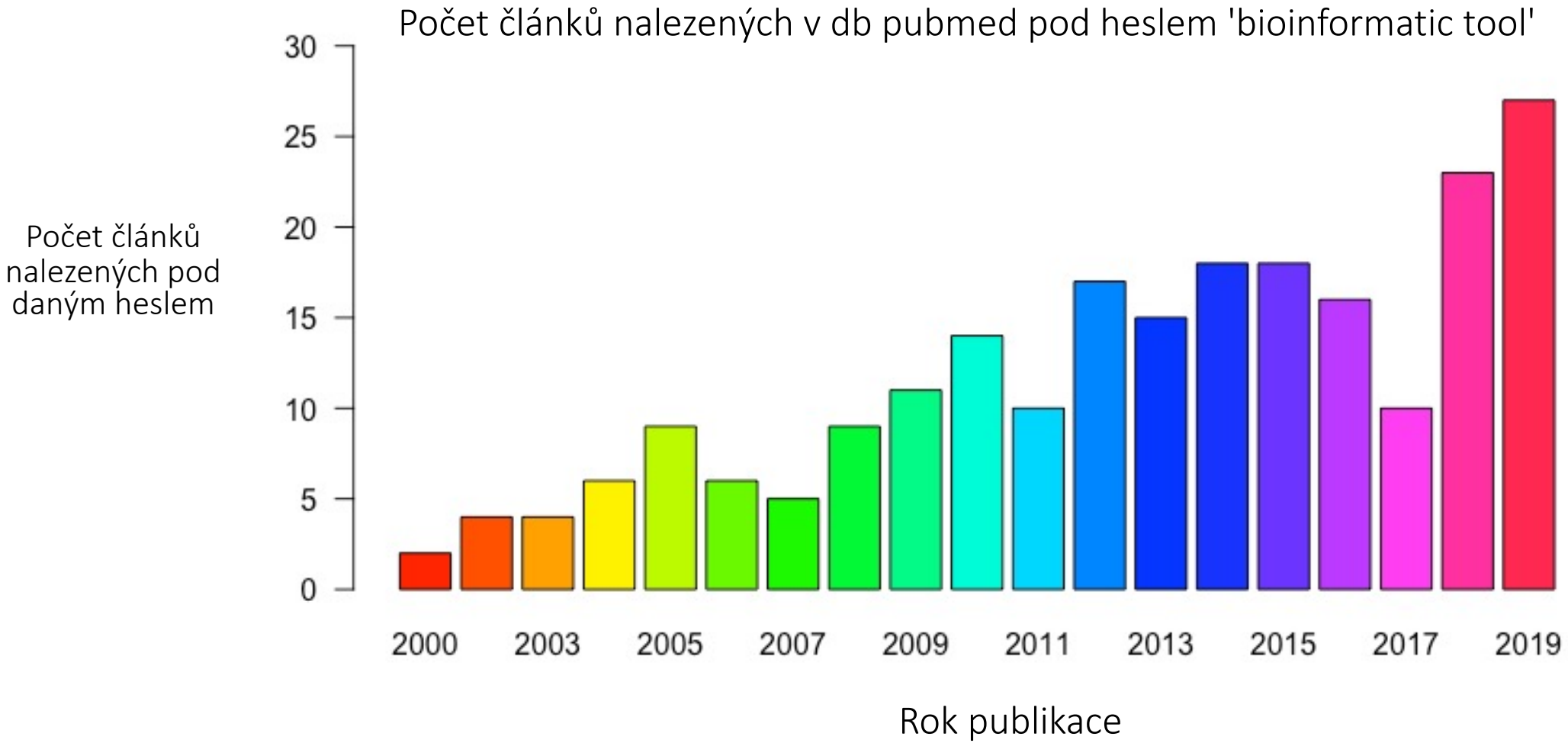


Ozdoby, které nesouvisí s obsahem



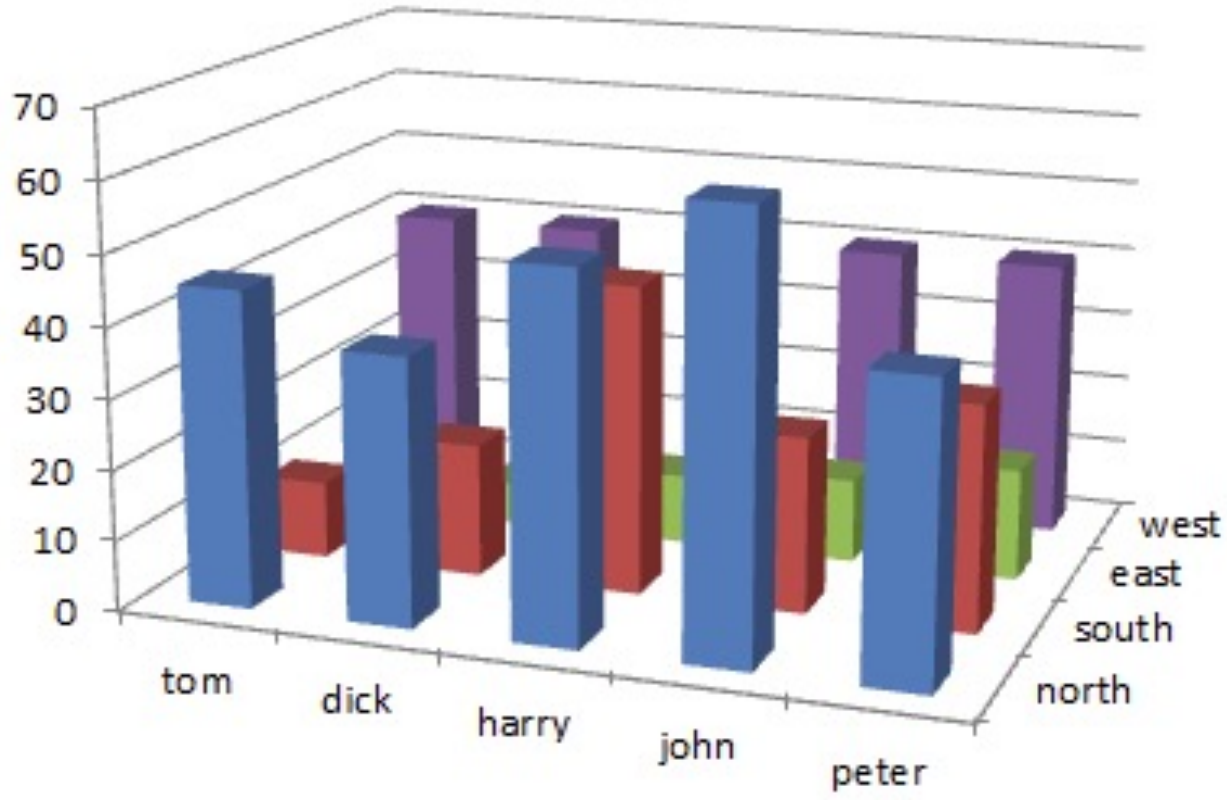
MUNI | RECETOX

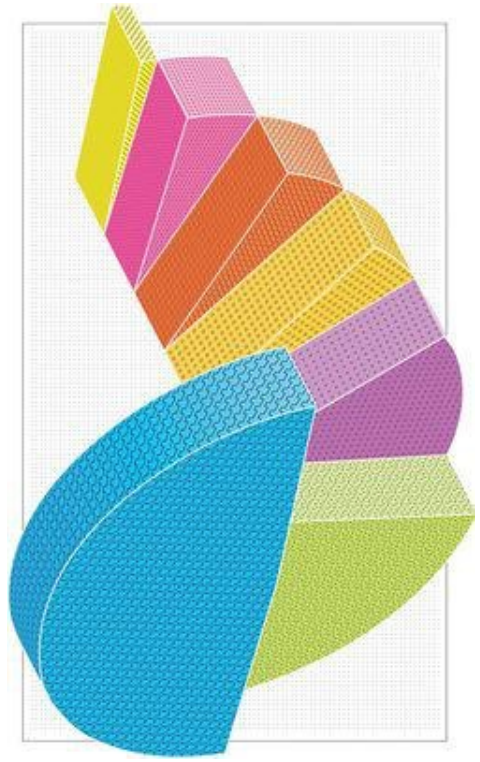
Nevhodné a příliš četné barvy



Zbytečné 3D efekty

Vědecké studie ukazují, že 3D efekty snižují srozumitelnost grafu.

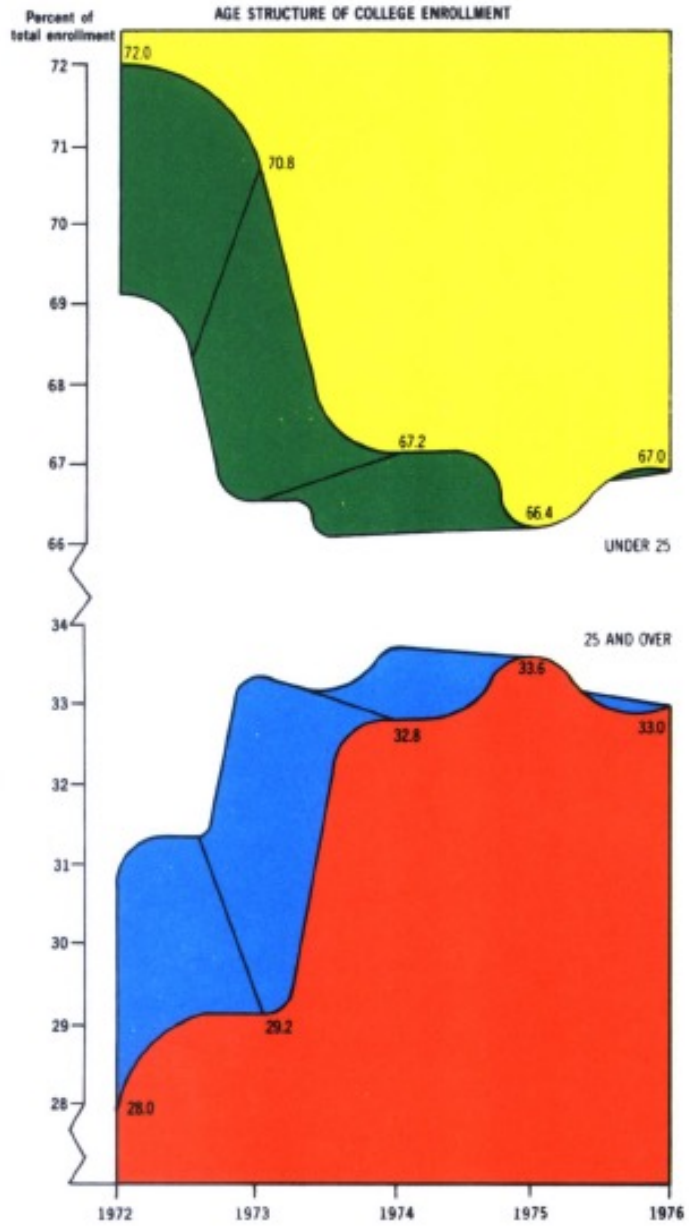




MUNI | RECETOX

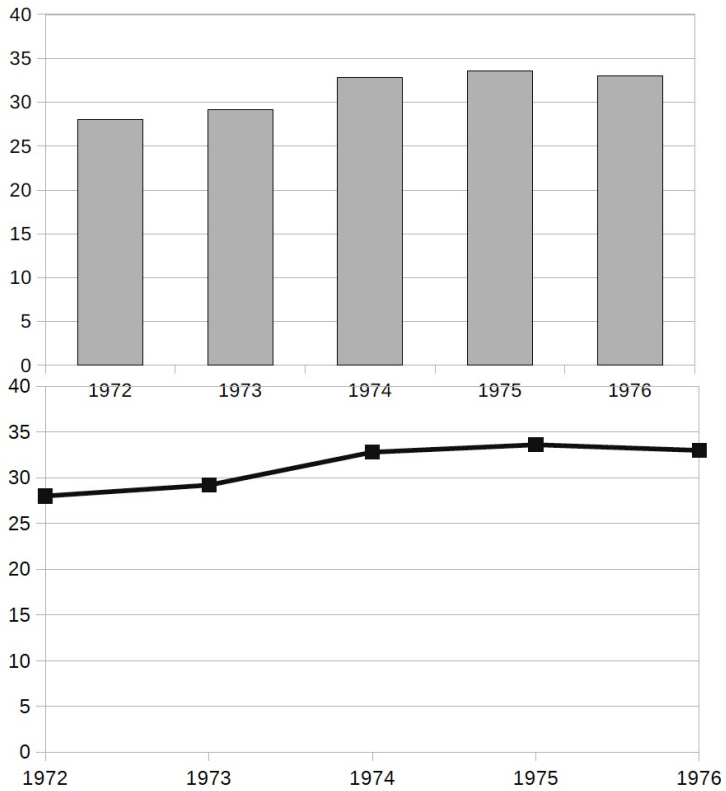
… a kombinací výše uvedeného

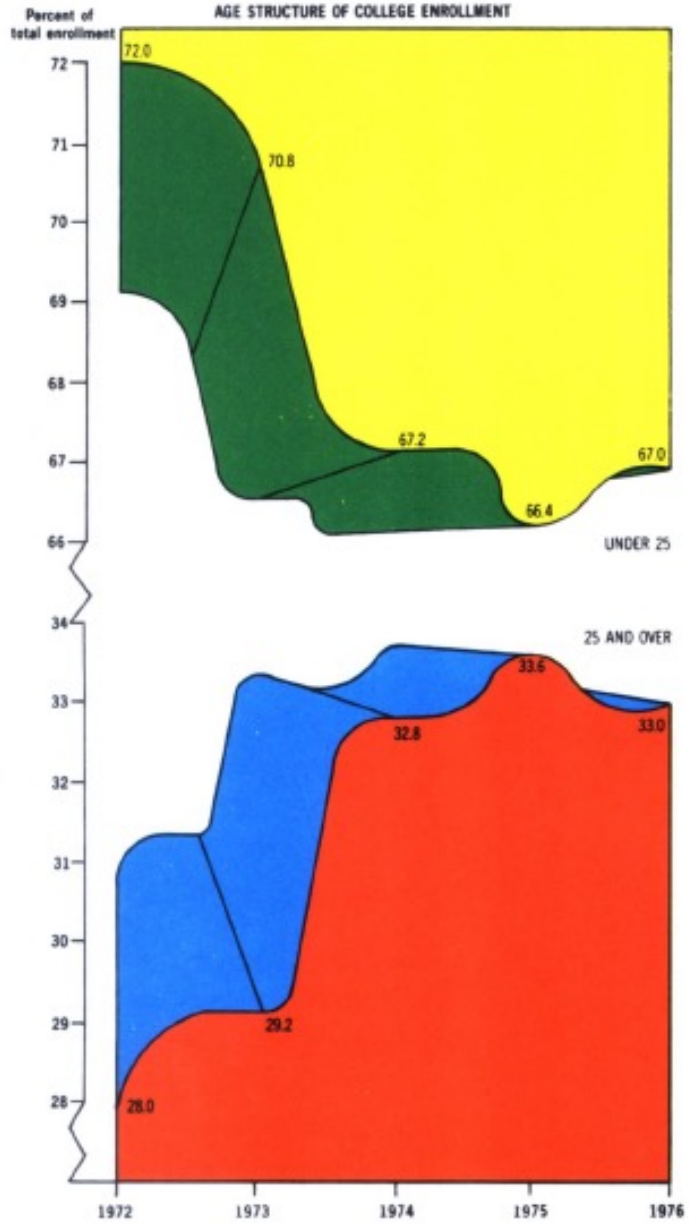
- Graf znázorňuje 5 čísel!
- Podíly vysokoškolských studentů nad a pod 25 let, v letech 1972 až 1976.



… a kombinací výše uvedeného

- Graf znázorňuje 5 čísel!
- Podíly vysokoškolských studentů nad a pod 25 let, v letech 1972 až 1976.





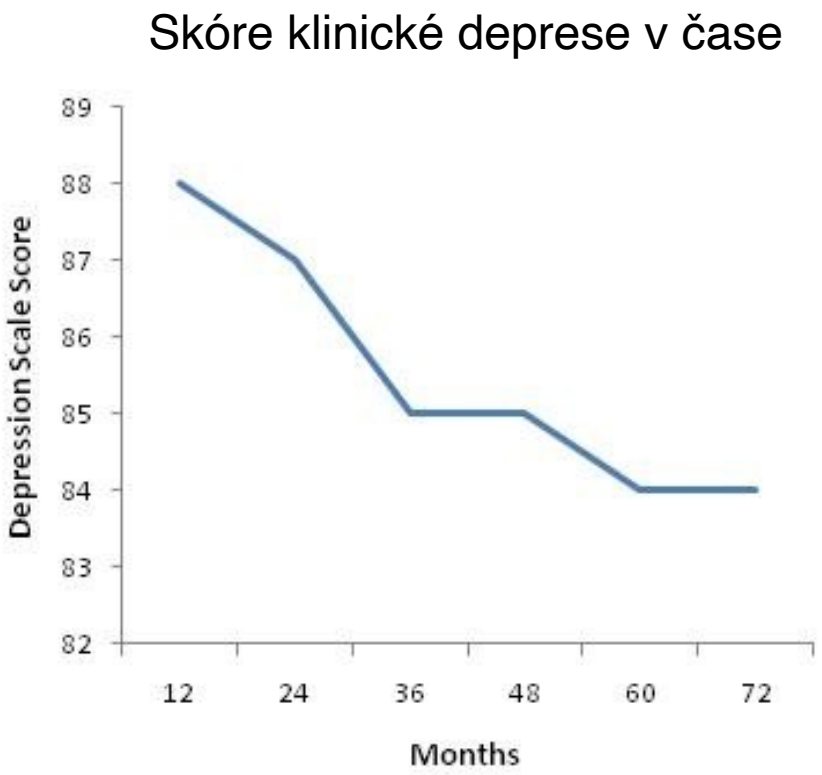
MUNI | RECETOX

Zkreslení

- Graf by neměl zobrazovat zkreslenou skutečnost (ať už účelově nebo náhodou)

Zkreslení škálou I.

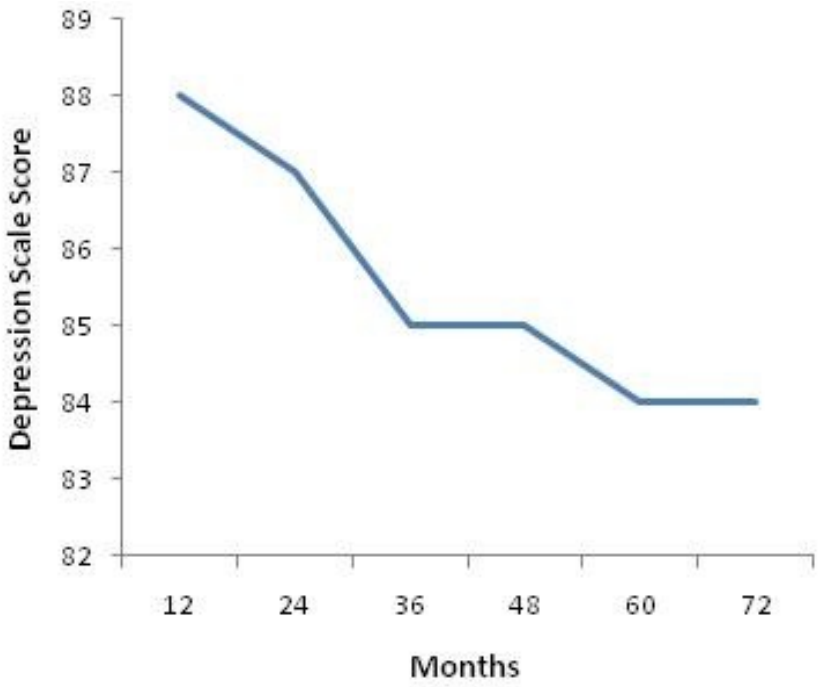
- Každý automaticky předpokládá, že osa Y začíná nulou!



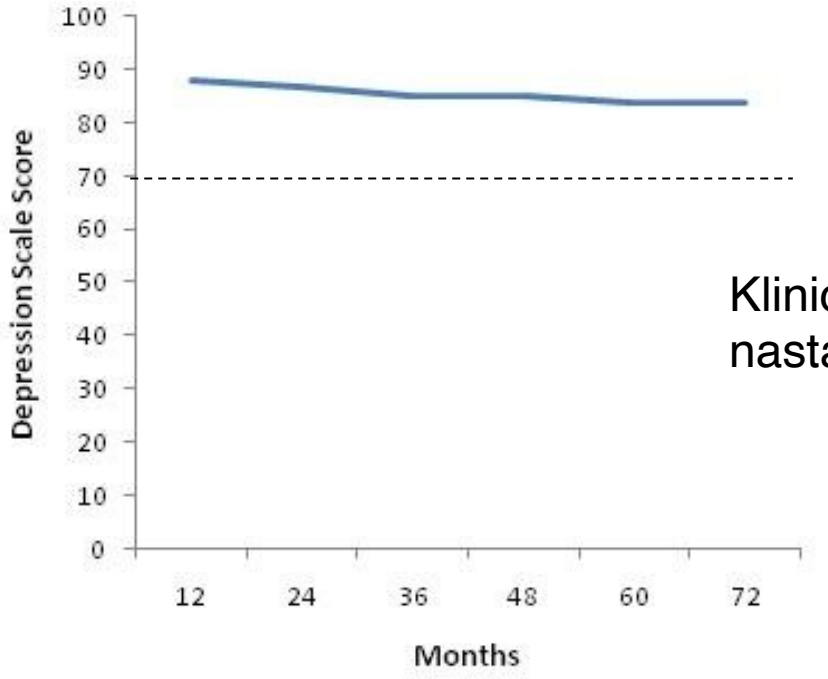
Zkreslení škálou I.

- Každý automaticky předpokládá, že osa Y začíná nulou!

Skóre klinické deprese v čase

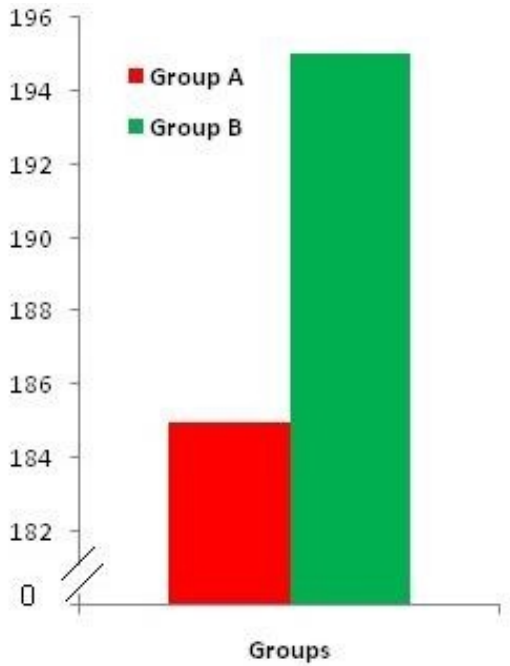


Skóre klinické deprese v čase



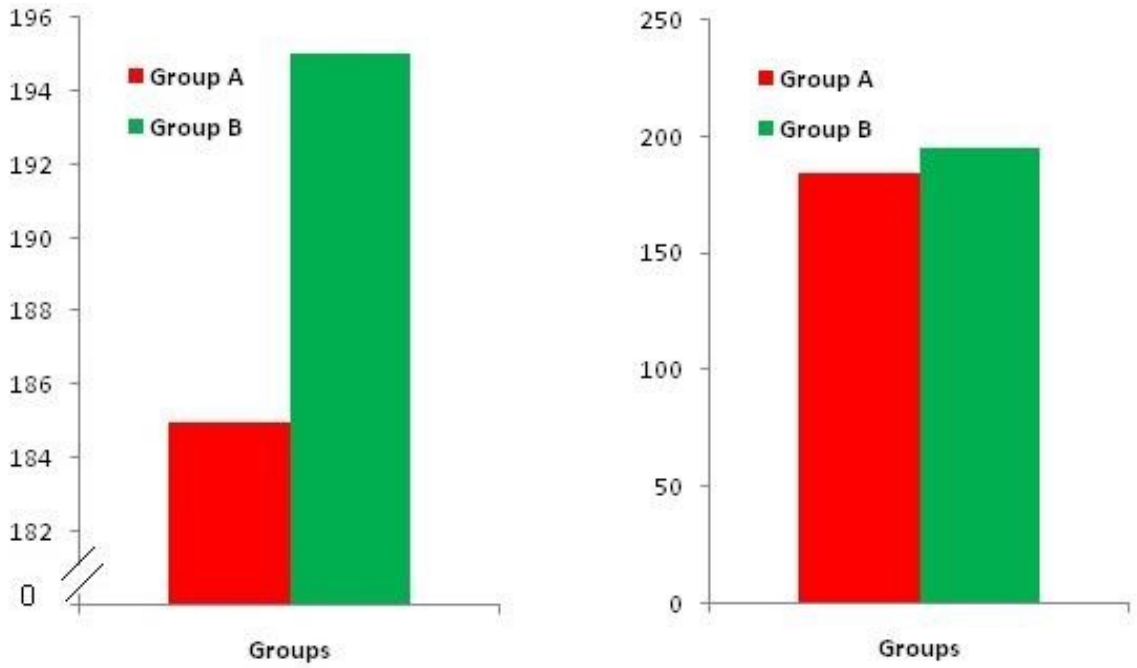
Klinická deprese nastává při skóre >70

Zkreslení škálou II.



Zkreslení škálou II.

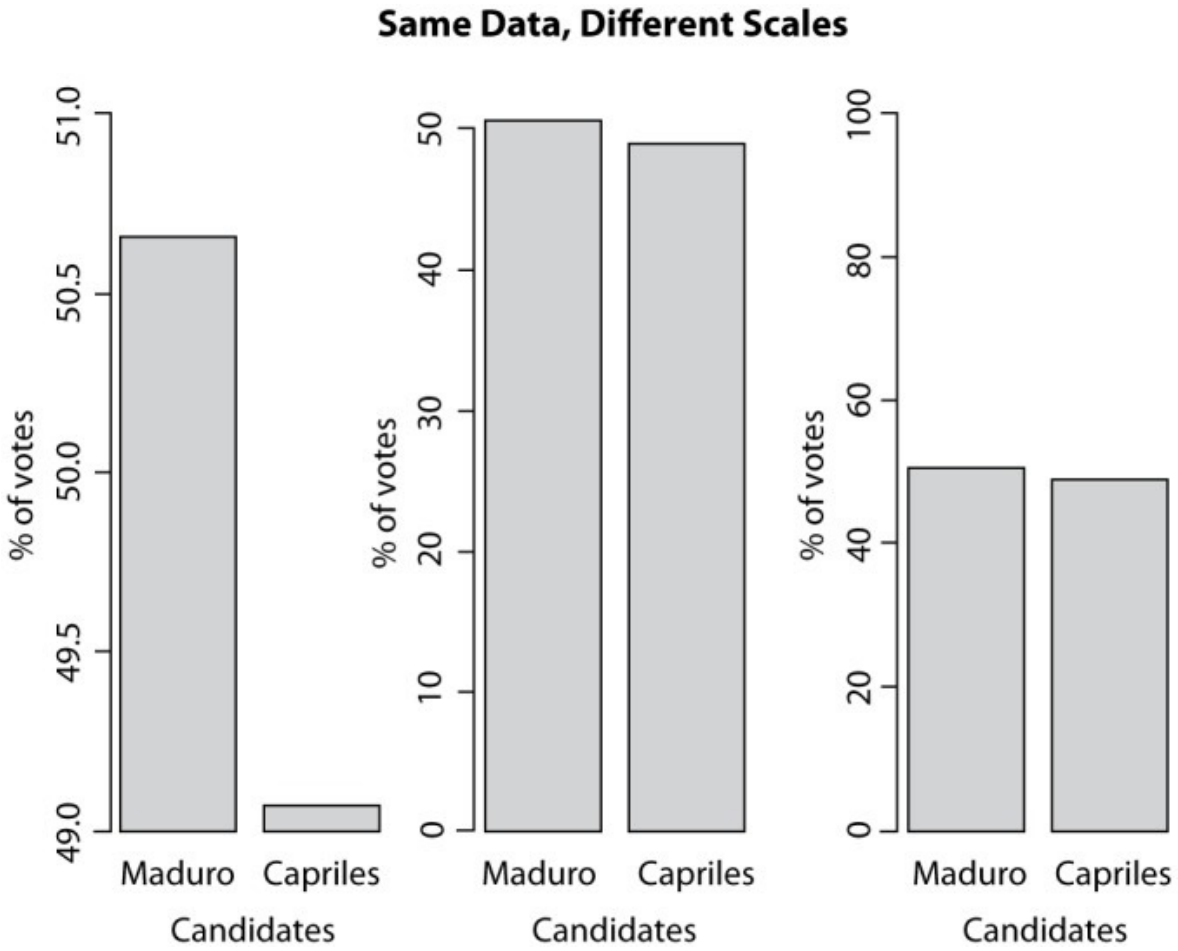
- Zkrácení osy y vyvolává dojem velkého rozdílu



http://www.exercisebiology.com/index.php/site/articles/how_graphs_can_fool_you/

Zkreslení škálou II

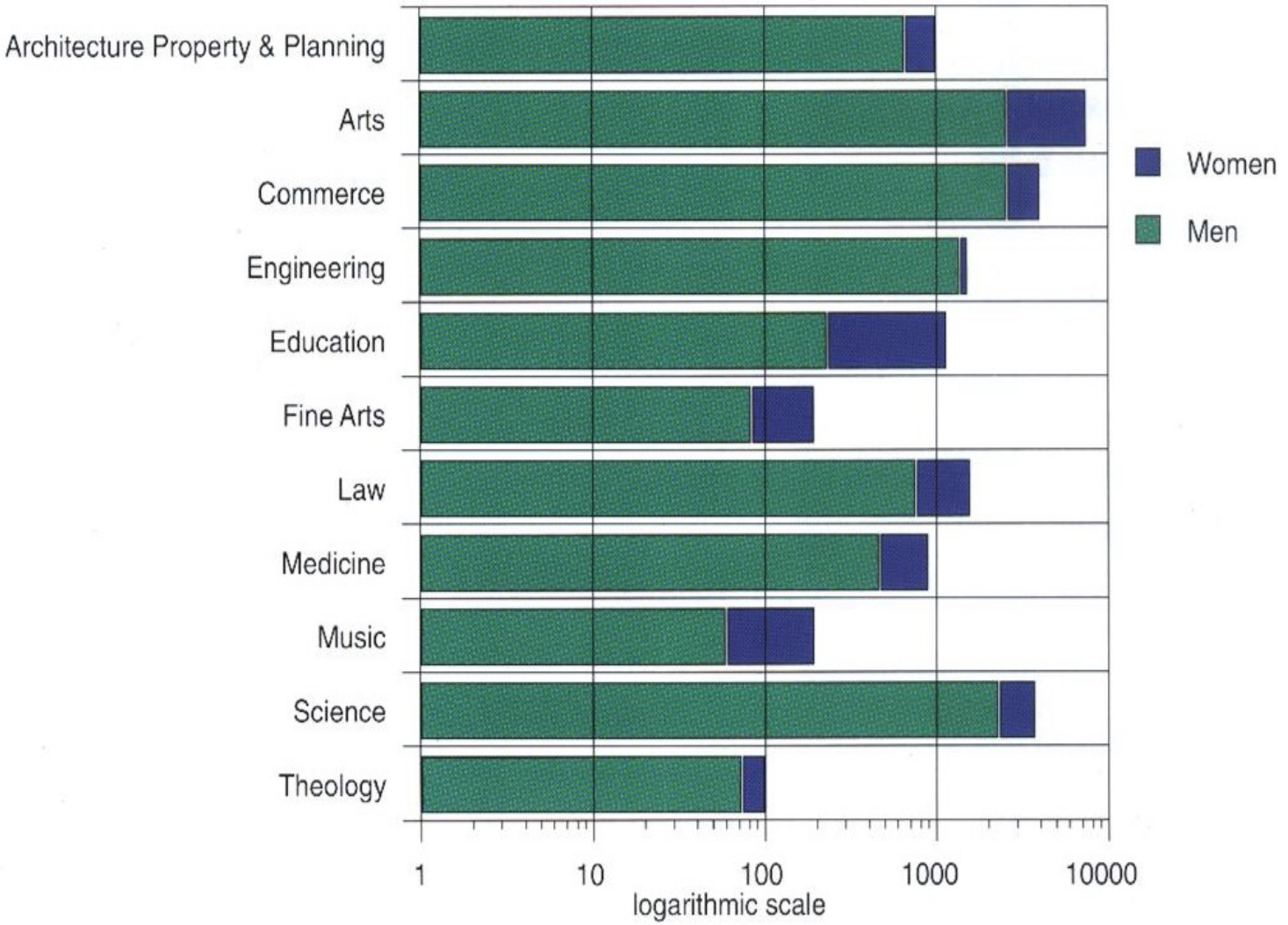
- Prezidentské volby ve Venezuele



Takto rozdíl prezentován v novinách

MUNI | RECETOX

Počty studentů na různých fakultách a podíl mužů a žen

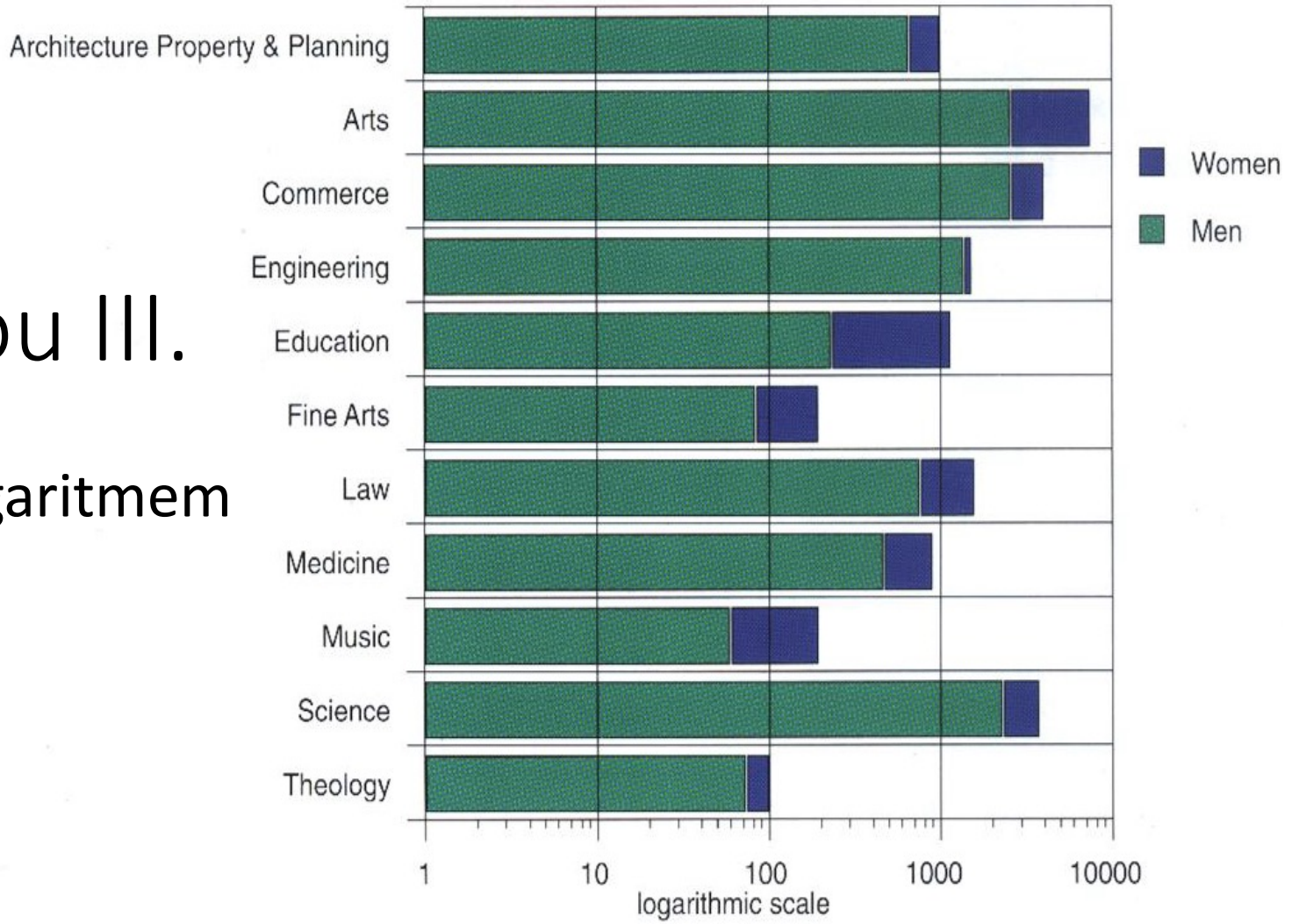


http://www.stat.auckland.ac.nz/~ihaka/120/Lectures/lecture03.pdf

Počty studentů na různých fakultách a podíl mužů a žen

Zkreslení škálou III.

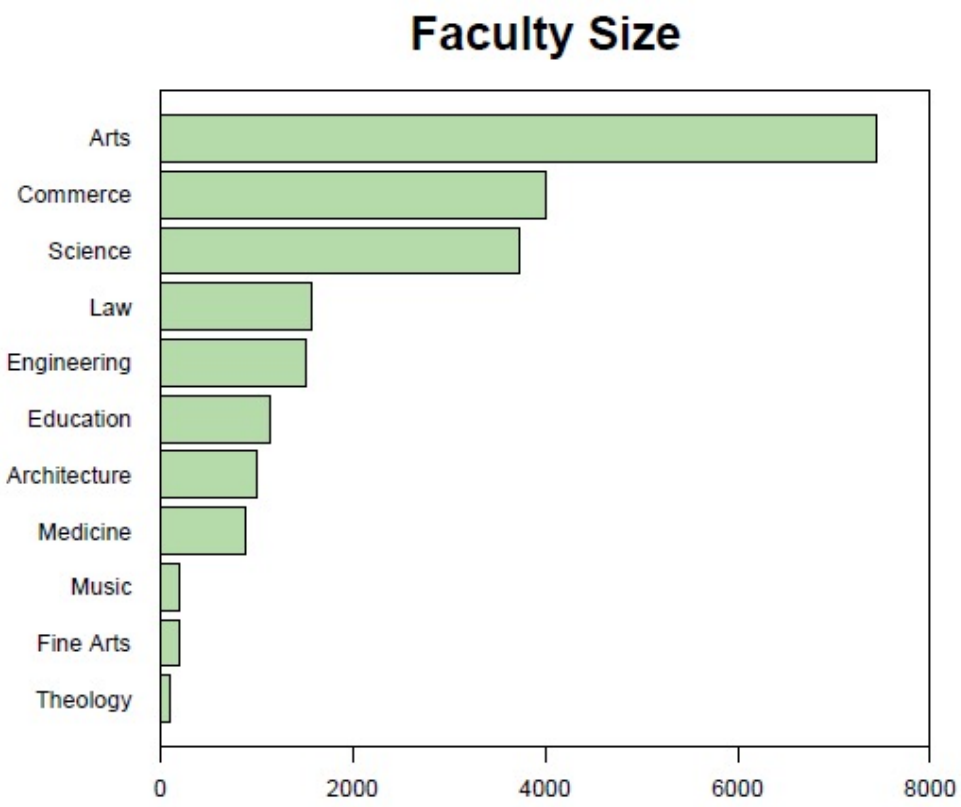
- Transformace osy logaritmem

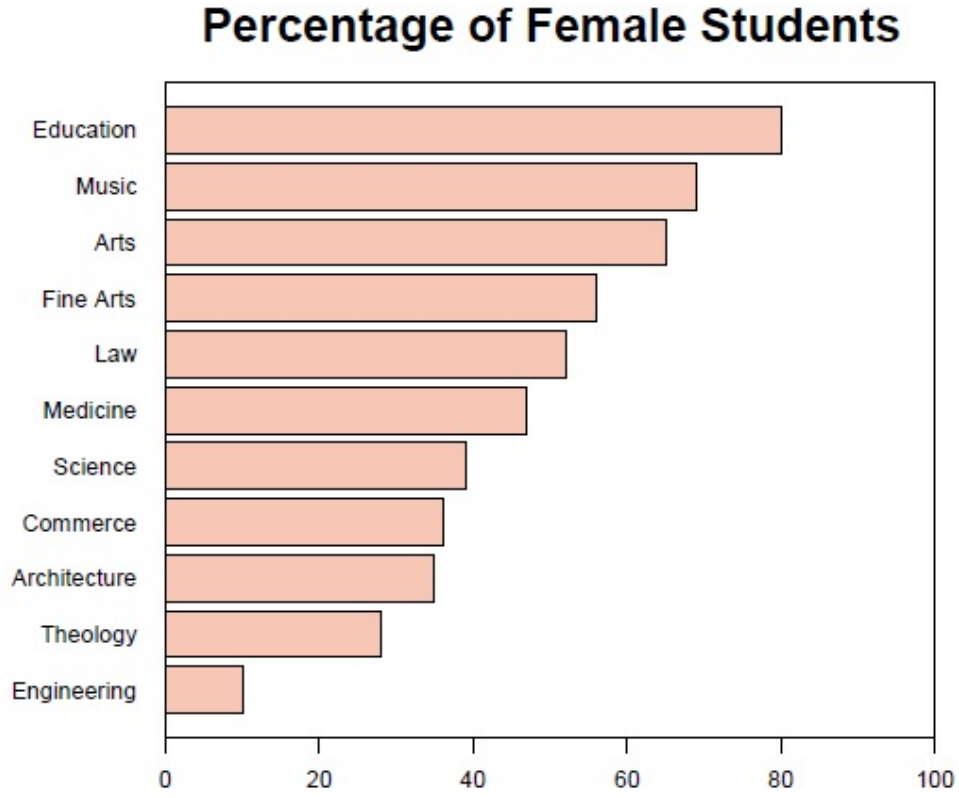


http://www.stat.auckland.ac.nz/~ihaka/120/Lectures/lecture03.pdf

Zkreslení škálou III.

- … to samé bez transformace

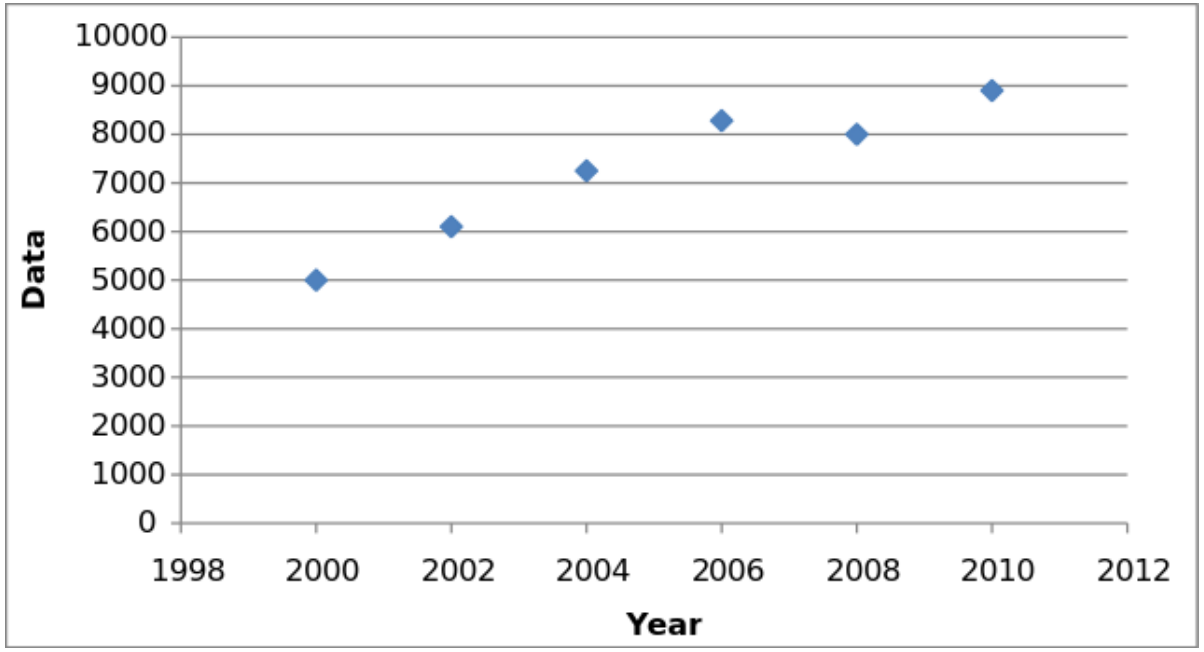




http://www.stat.auckland.ac.nz/~ihaka/120/Lectures/lecture03.pdf

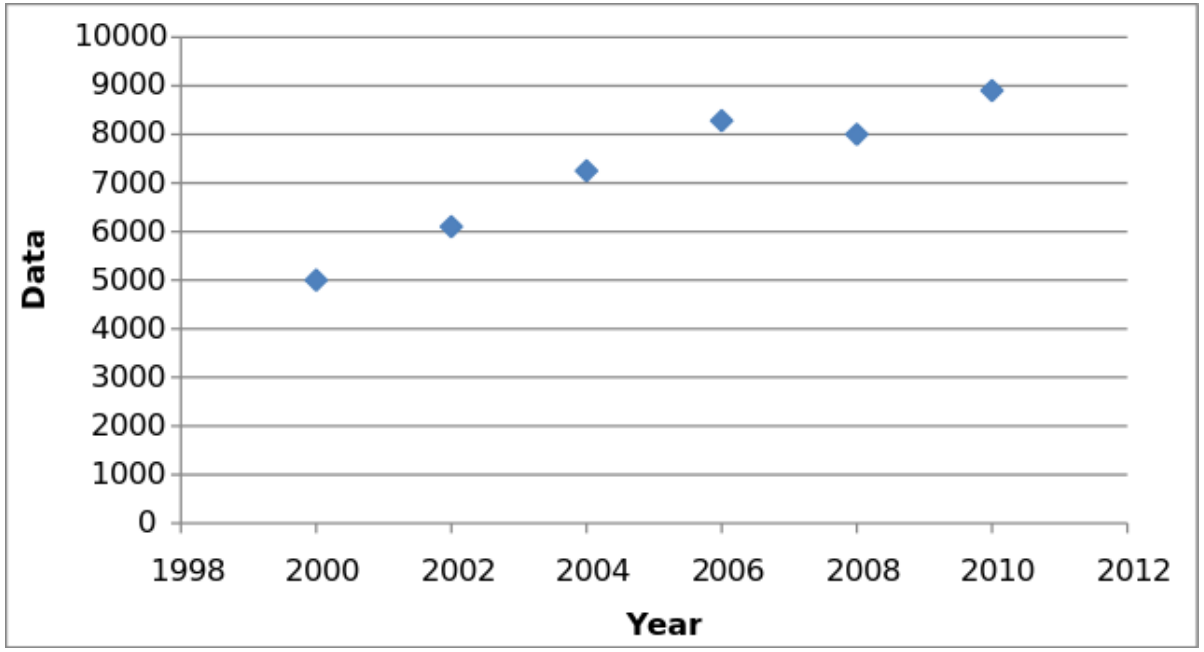
Zkreslení škálou IV.

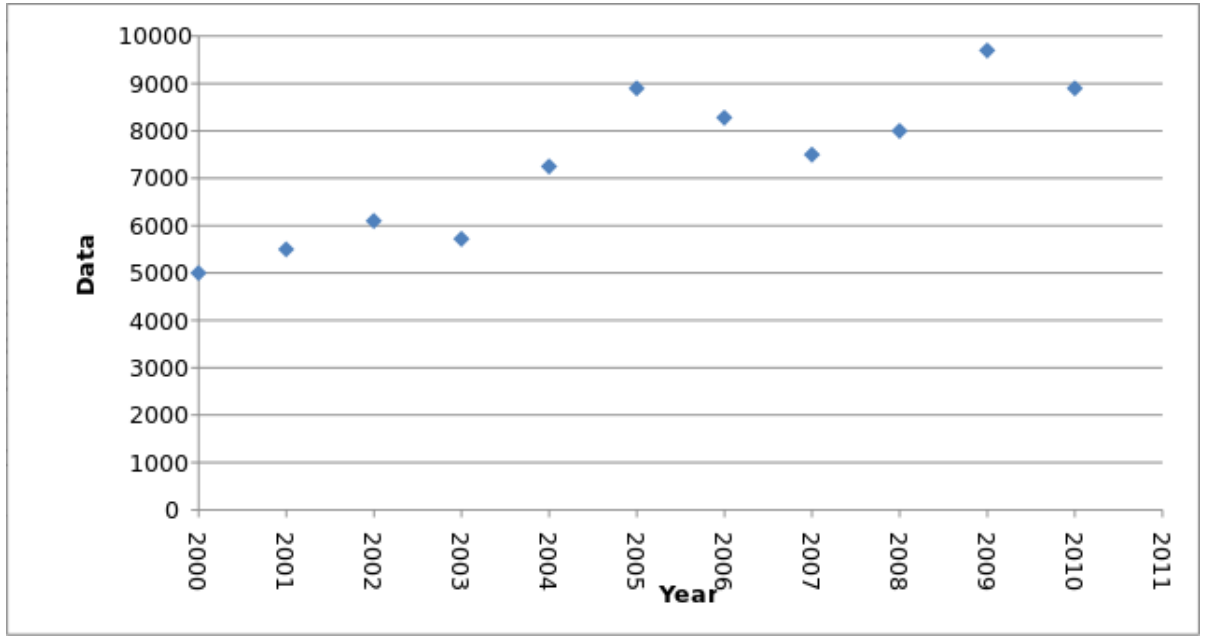
Chybějící body na ose x – zkreslení linearity



Zkreslení škálou IV.

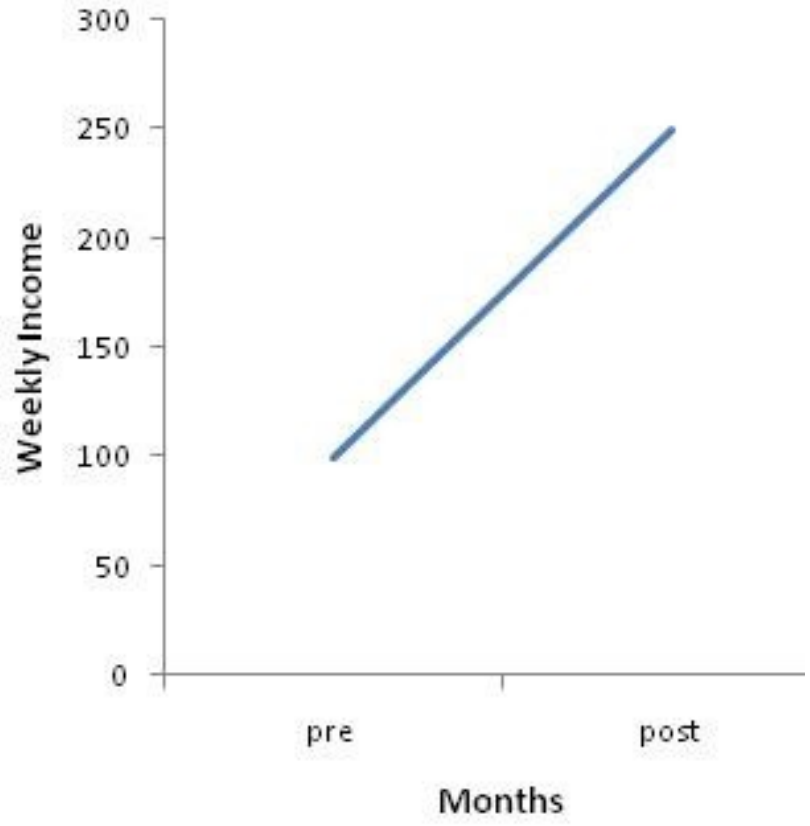
Chybějící body na ose x – zkreslení linearity





Zkreslení škálou V

Chybějící body na ose x – zkreslení linearity

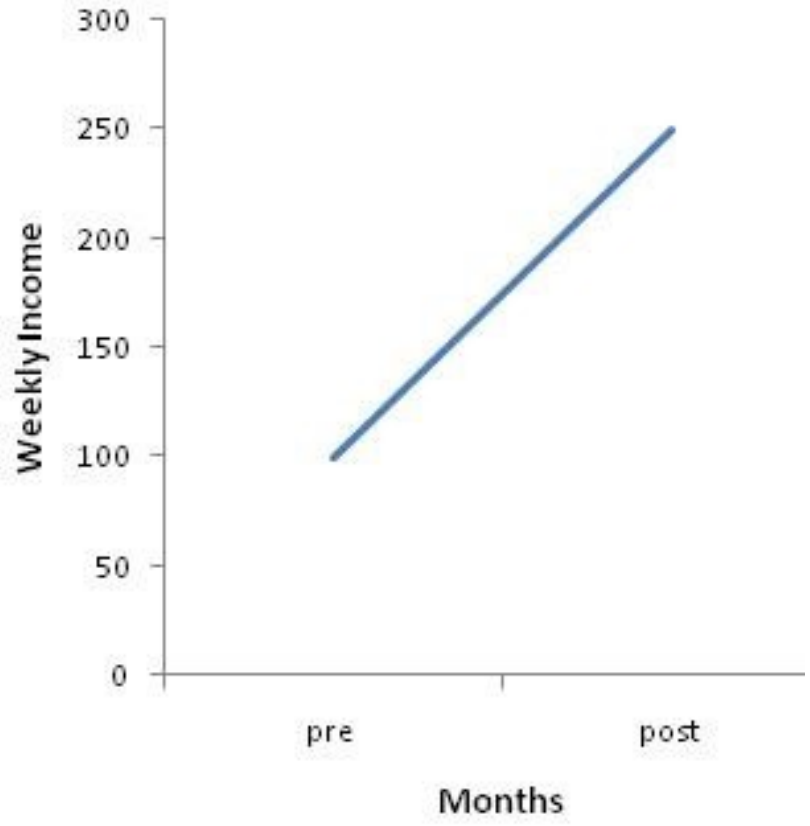


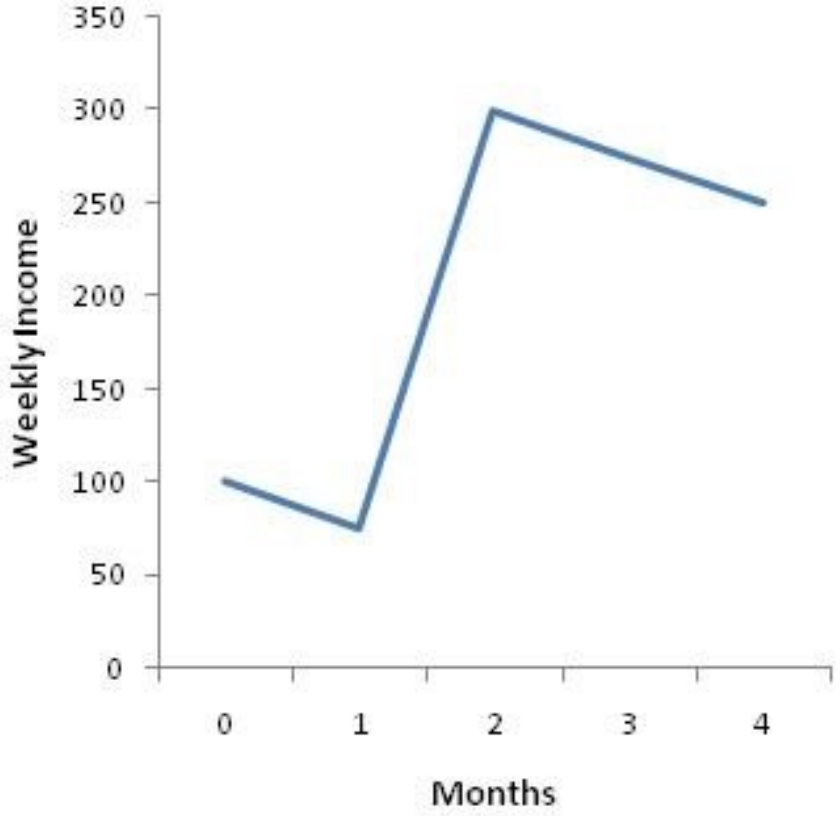
Změny týdenního příjmu

http://www.exercisebiology.com/index.php/site/articles/how_graphs_can_fool_you/

Zkreslení škálou V

Chybějící body na ose x – zkreslení linearity



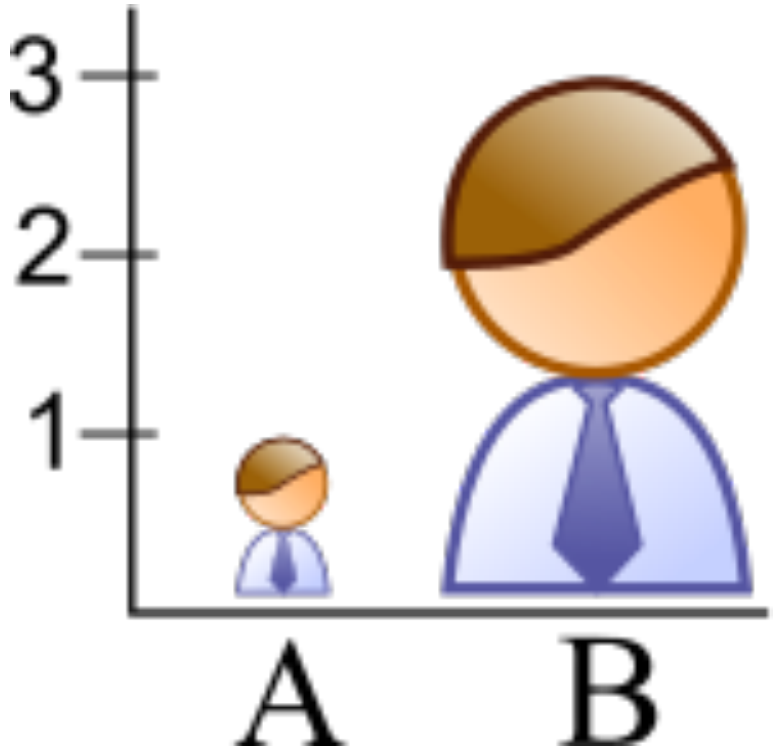


Změny týdenního příjmu

http://www.exercisebiology.com/index.php/site/articles/how_graphs_can_fool_you/

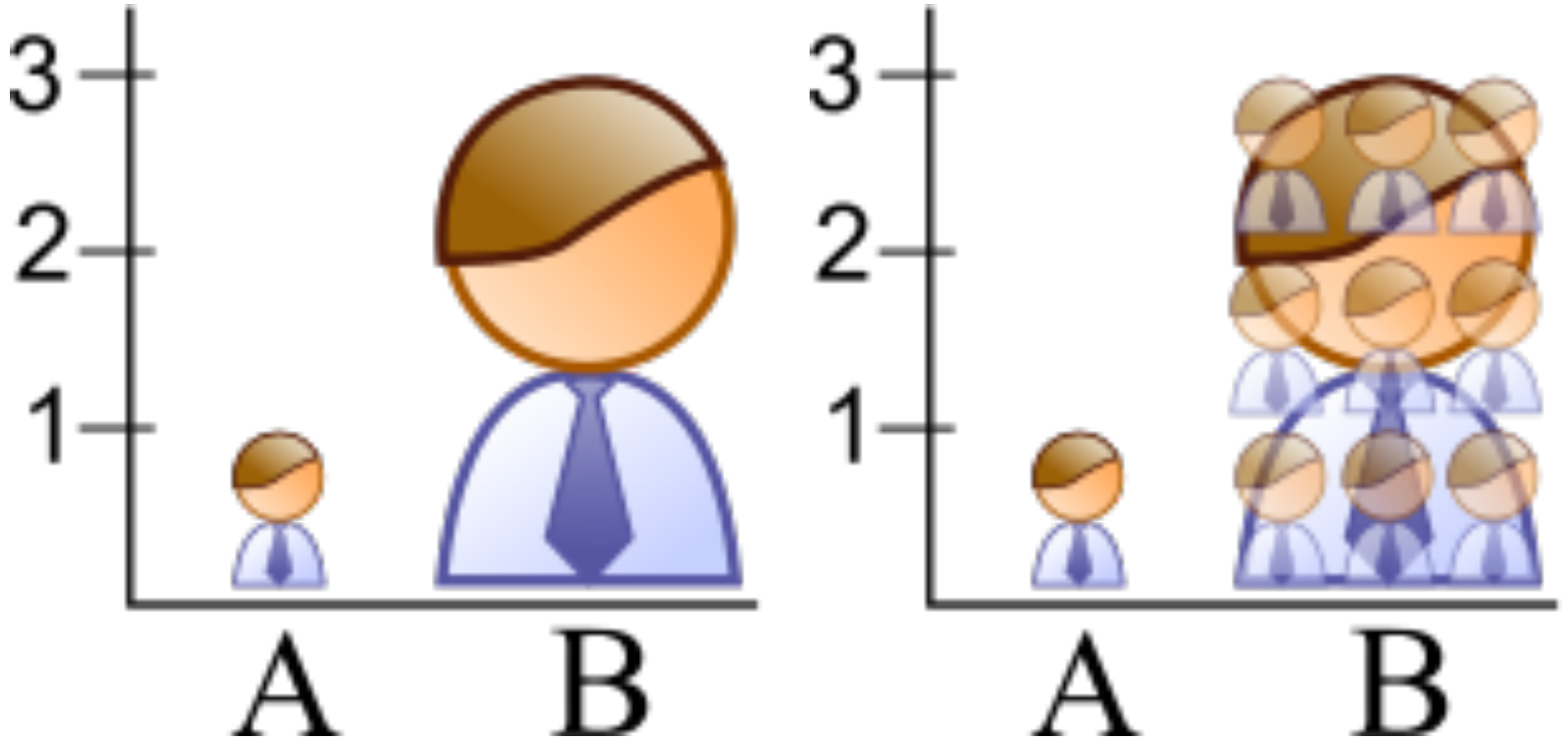
Zkreslení škálou VI

Nesprávné zobrazení jednorozměrného násobného rozdílu pomocí plochy



Zkreslení škálou VI

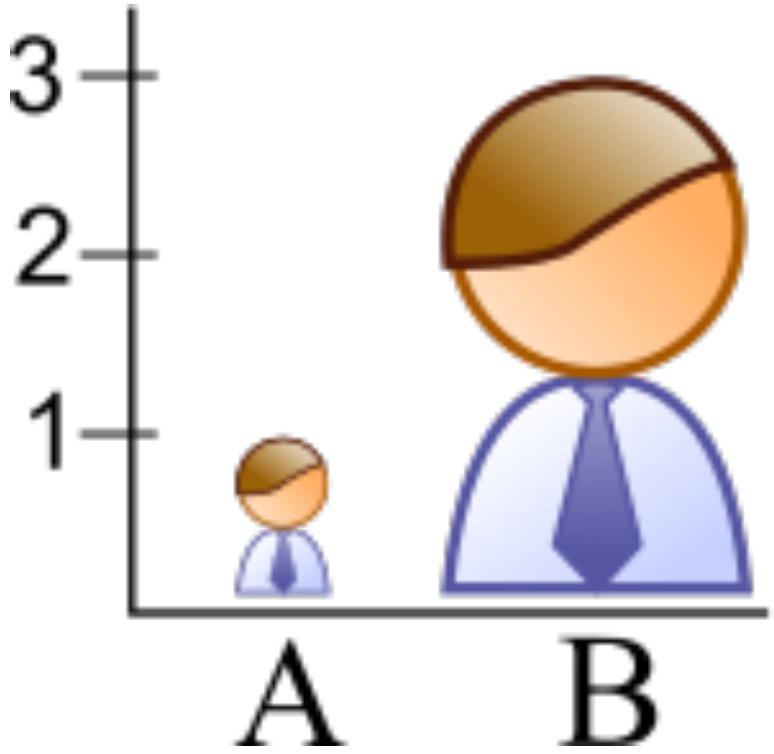
Nesprávné zobrazení jednorozměrného násobného rozdílu pomocí plochy

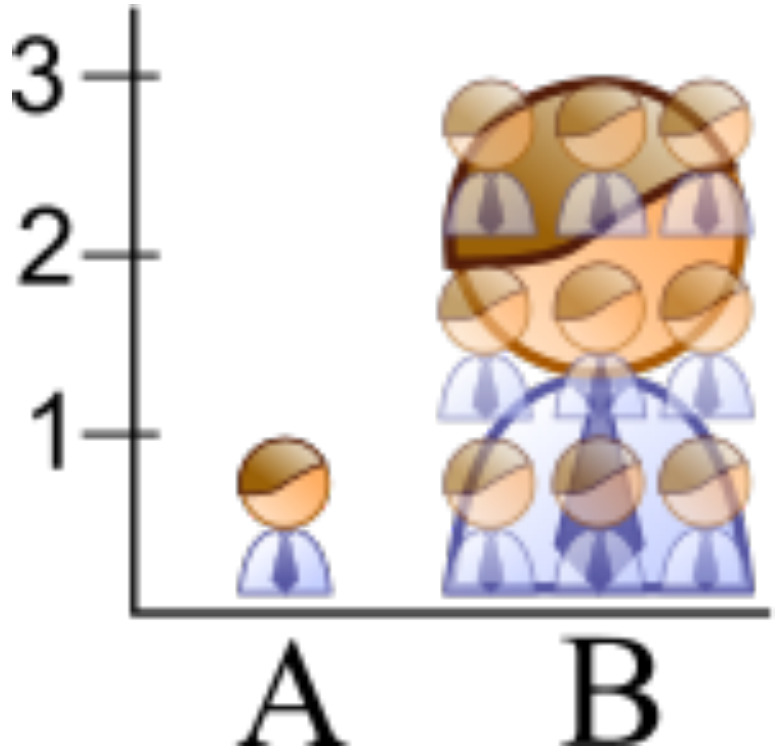


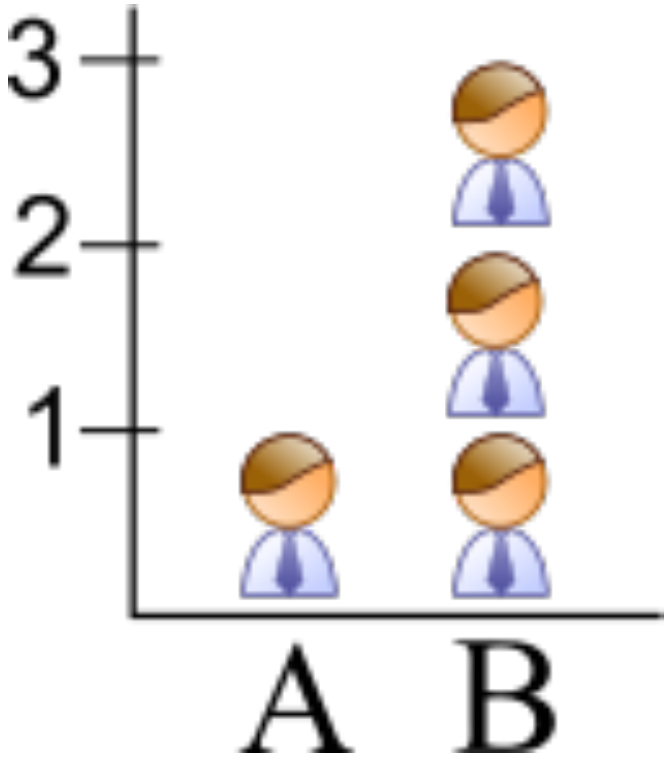
Trojnásobný rozdíl působí jako devítinásobný

Zkreslení škálou VI

Nesprávné zobrazení jednorozměrného násobného rozdílu pomocí plochy





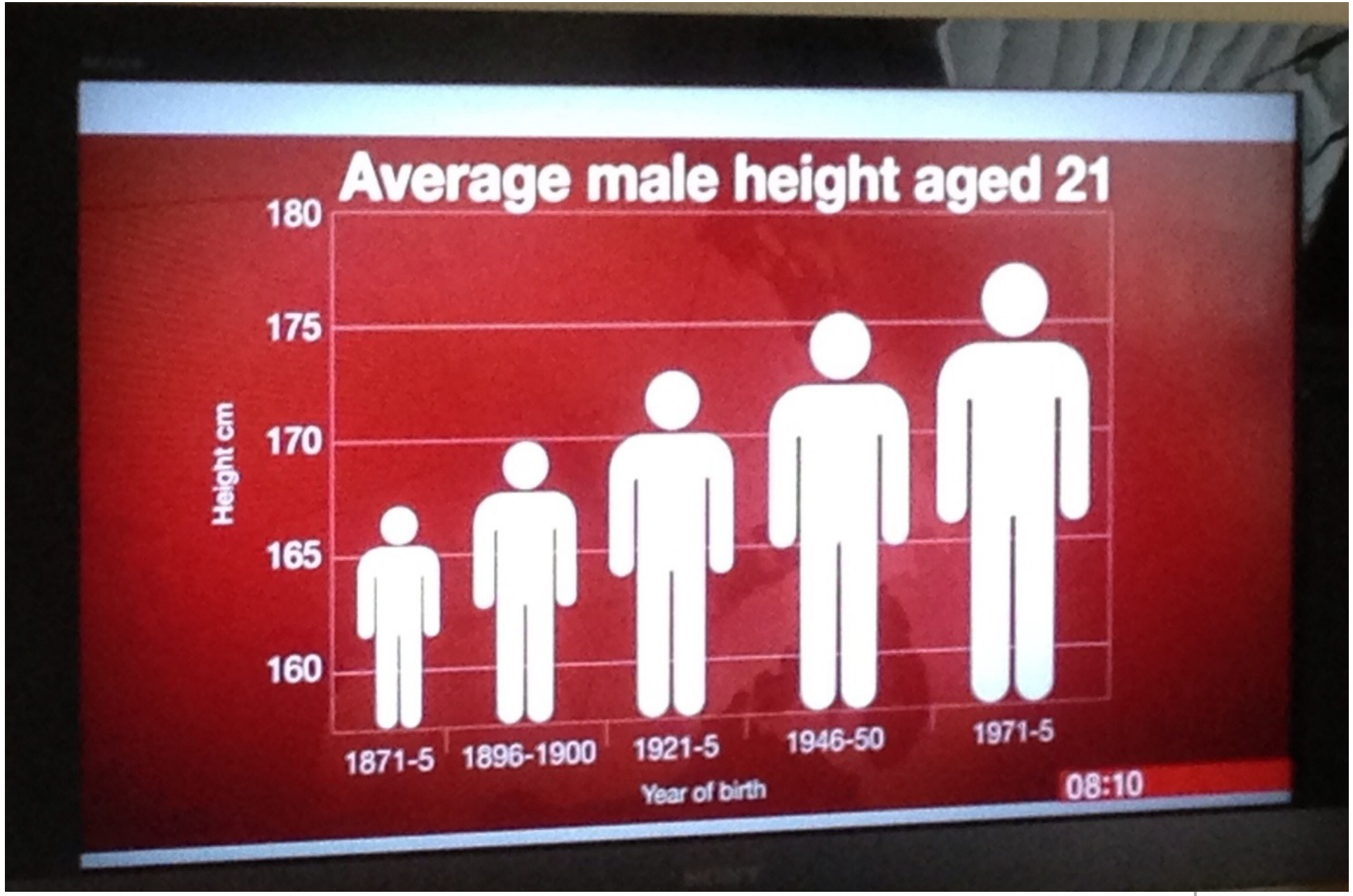


Trojnásobný rozdíl působí jako devítinásobný

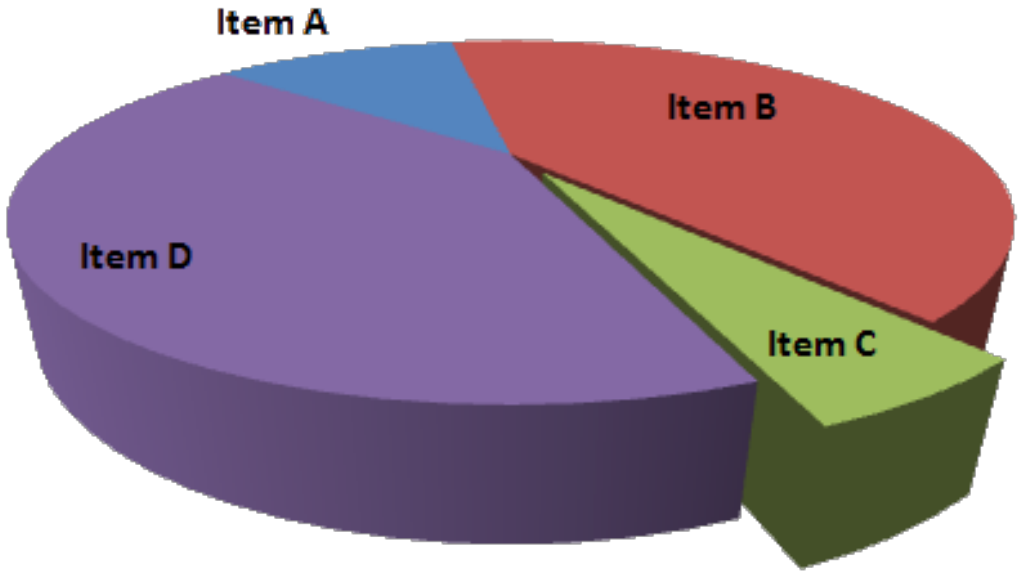
Správné zobrazení

MUNI | RECETOX

Zkreslení škálou VI.

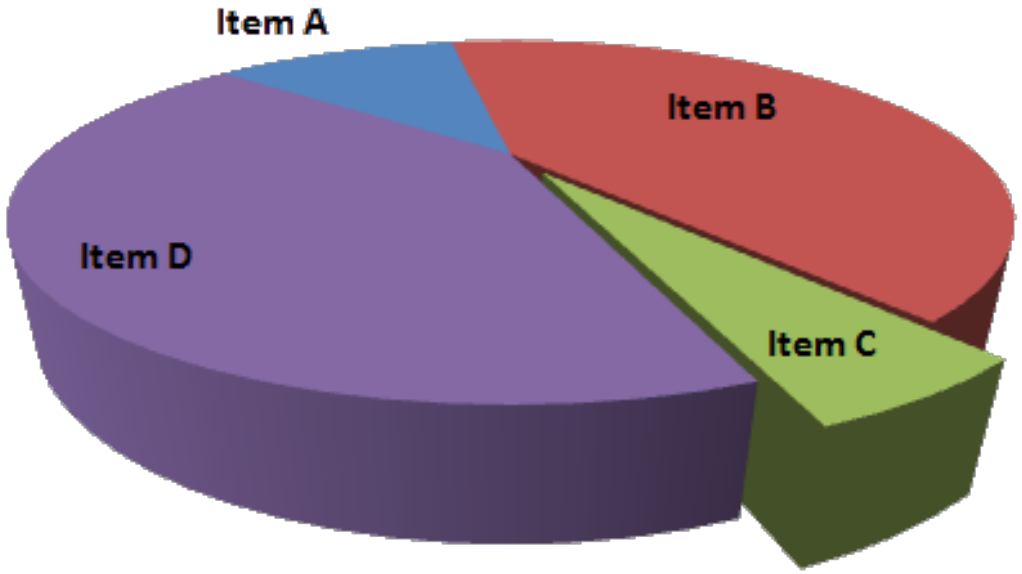

Average male height aged 21
180
175
170
165
160
Height cm
1871-5
1896-1900
1921-5
1946-50
1971-5
Year of birth
08:10

Zkreslení úhlem pohledu a 3D efekty



Item C > Item A ?

Zkreslení úhlem pohledu a 3D efekty



Item C > Item A ?

Item C < Item A !

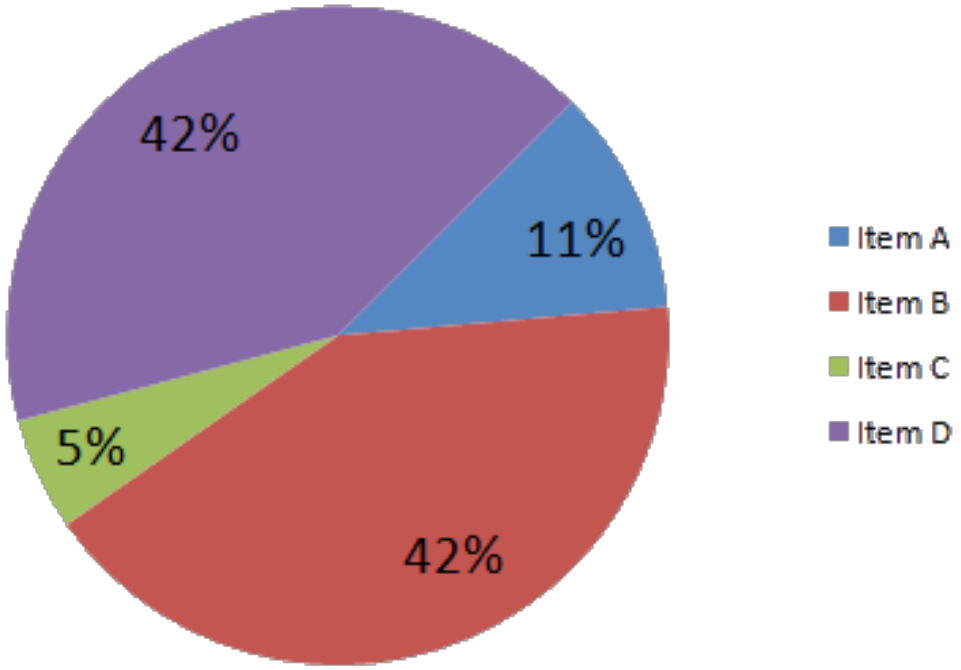


Figure 1. Source: Erickson Times

https://www.forbes.com/sites/naomirobbins/2012/02/16/misleading-graphs-figures-not-drawn-to-scale/#7ea98afa15ef

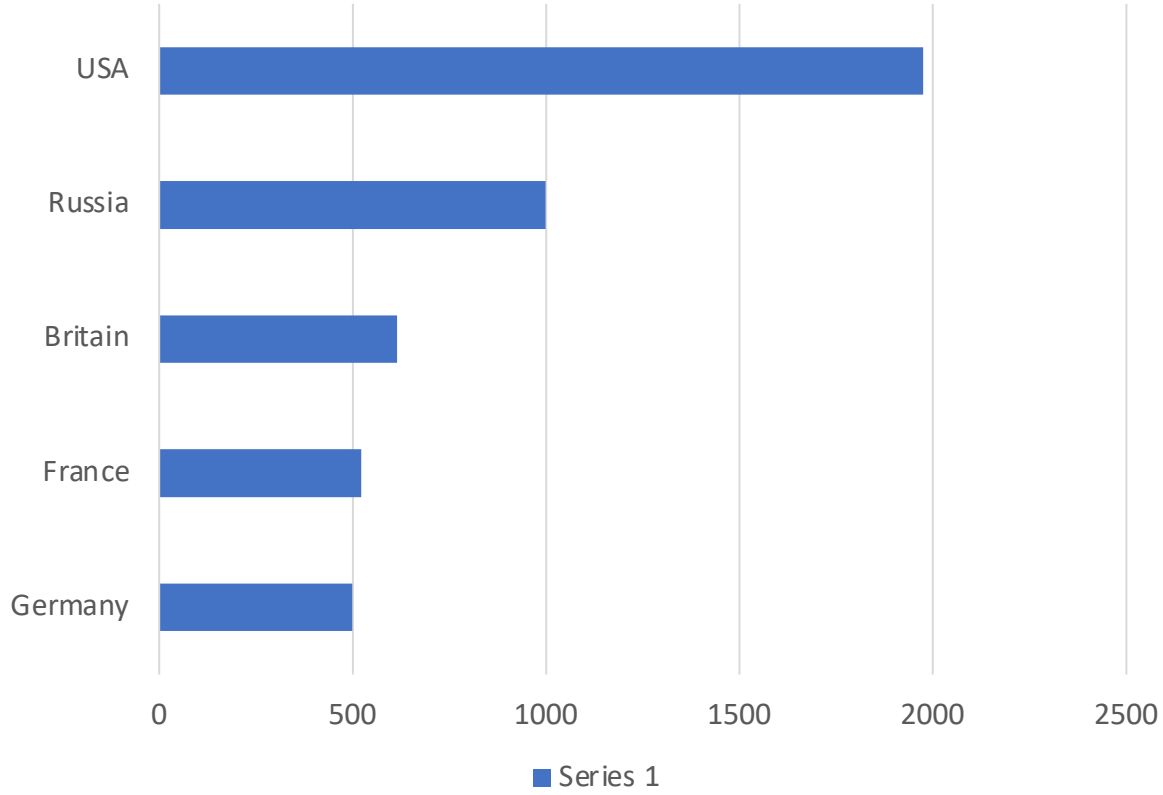


Figure 1. Source: Erickson Times

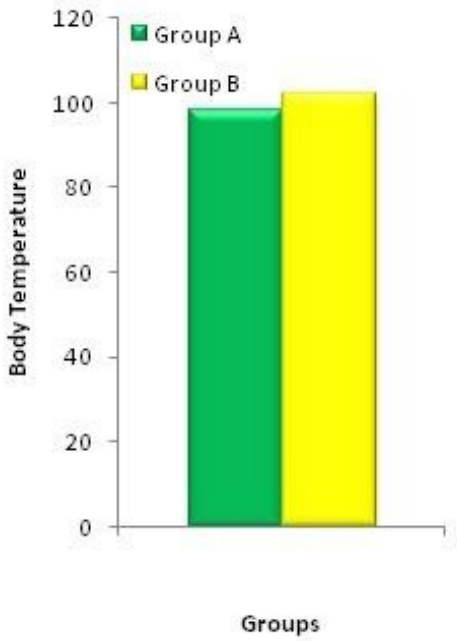
https://www.forbes.com/sites/naomirobbins/2012/02/16/misleading-graphs-figures-not-drawn-to-scale/#7ea98afa15ef

Faktor klamu

$$Faktor\ klamu = \frac{Velikost\ efektu\ v\ grafu}{Velikost\ efektu\ v\ datech}$$

Faktor klamu >1 => změna v grafu je přehnaná
Faktor klamu mezi 0 a 1 => změna v grafu není dostatečně viditelná
Faktor klamu = 1 => perfektní reprezentace skutečného rozdílu

Nesprávný graf – zkreslení významu

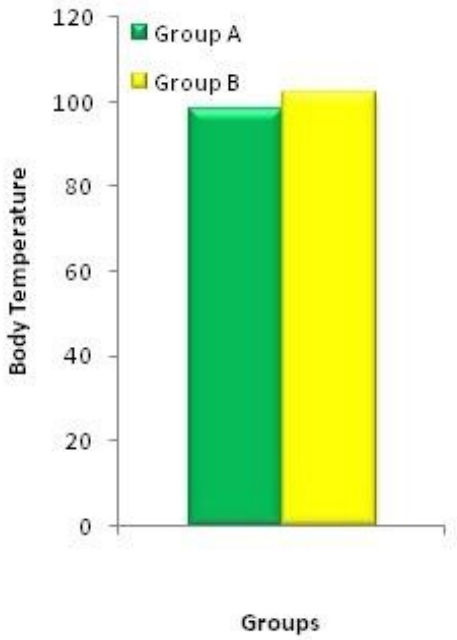


Průměrná tělesná teplota ve dvou skupinách

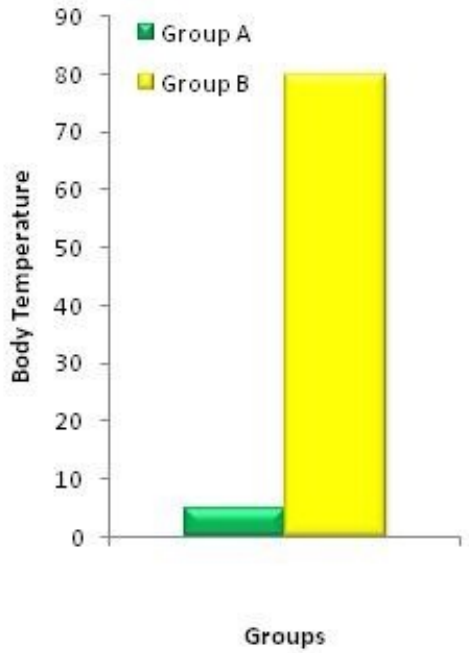
http://www.exercisebiology.com/index.php/site/articles/how_graphs_can_fool_you/

Nesprávný graf – zkreslení významu

Nesprávné zobrazení výsledků dává pocit, že jde o nevýznamný rozdíl



Průměrná tělesná teplota ve dvou skupinách



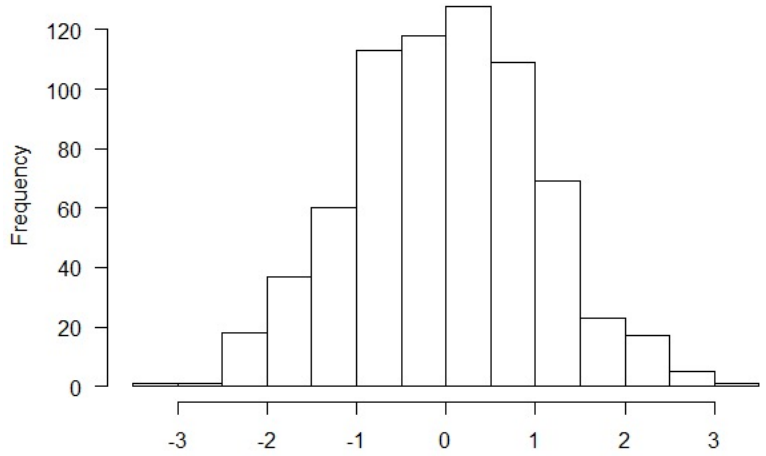
Podíl pacientů se zvýšenou teplotou (>37C)

http://www.exercisebiology.com/index.php/site/articles/how_graphs_can_fool_you/

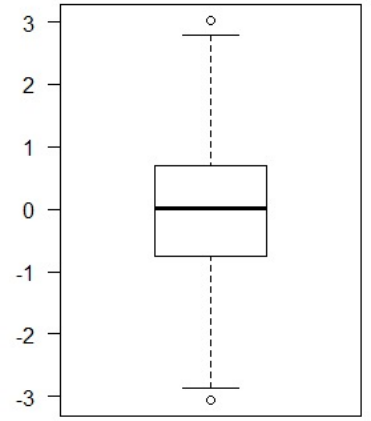
Výběr správného grafu

Grafy zobrazující **rozložení spojitých** proměnných

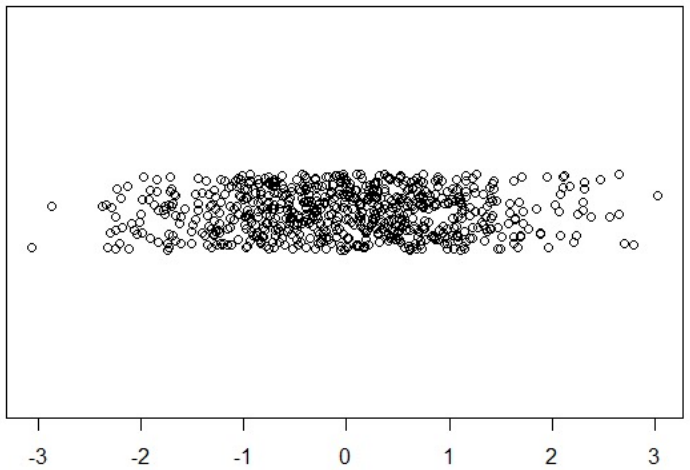
- Histogram



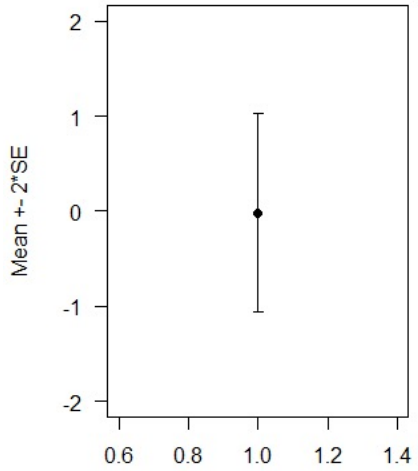
- Krabicový graf



- Jednorozměrný bodový graf

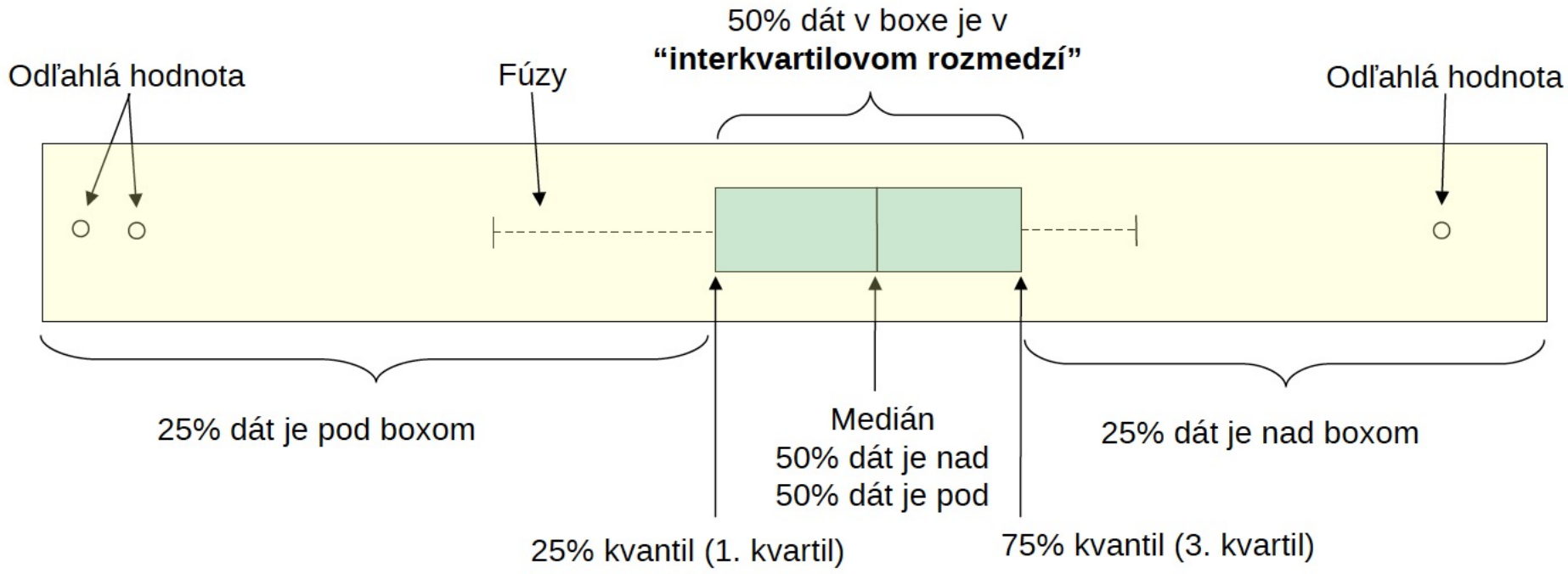


- Průměr a chyba



MUNI | RECETOX

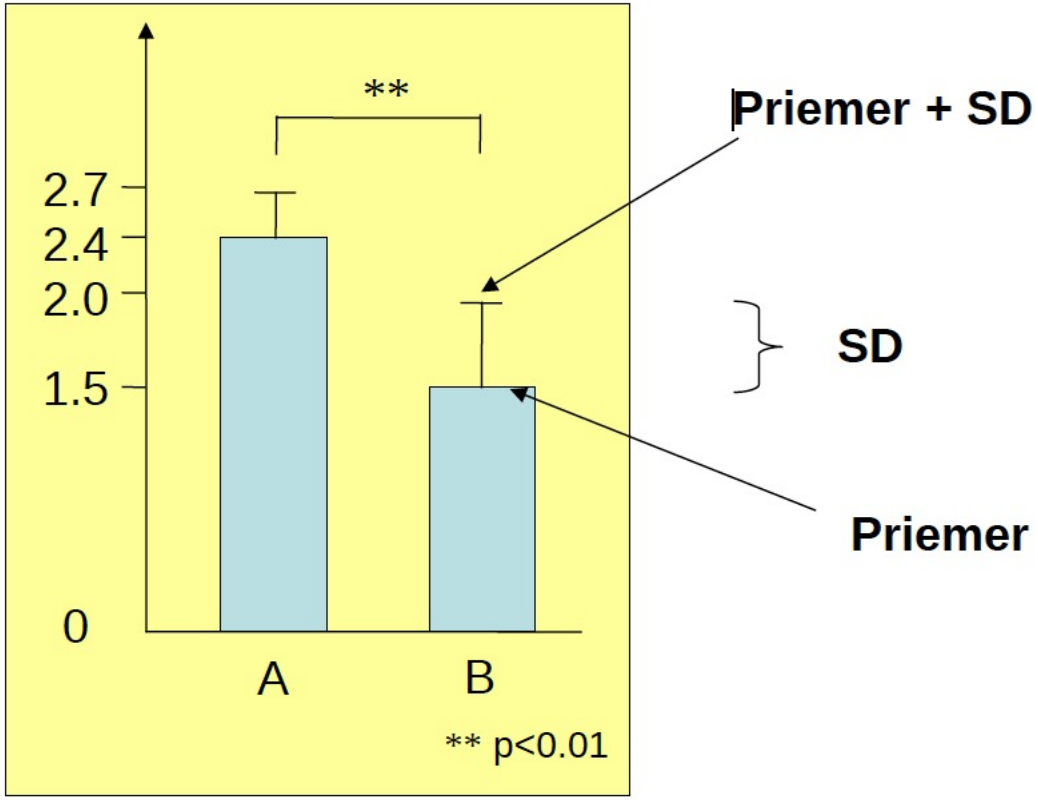
Krabicový graf (box and whisker plot)



- **Odľahlé hodnoty** (neobvyklé hodnoty) – tie dátové body, ktorých vzdialenosť od boxu je väčšia ako 1.5 krát interkvartilové rozmedzie.
- **Fúzy** (whiskers) sa rozpínajú k poslednému neodľahlému bodu
- Boxplot je grafická reprezentácia **5-číselného zhrnutia**:

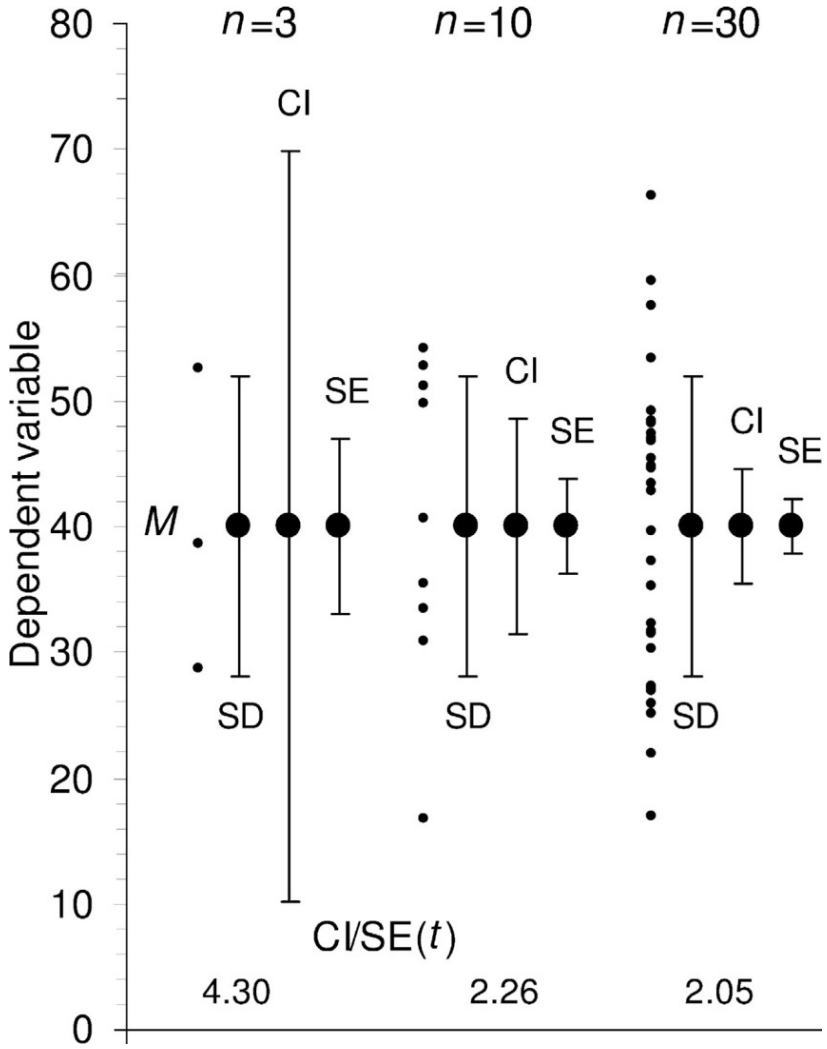
Minimum, Prvý kvartil (25%), Medián, Tretí kvartil (75%), Maximum

Speciální případ – sloupcový graf s chybou



Průměr měření skupiny A (25 vzorků) a B (18 vzorků) s indikací směrodatné odchylky každé skupiny; byl aplikován oboustranný dvouvýběrový T-test.

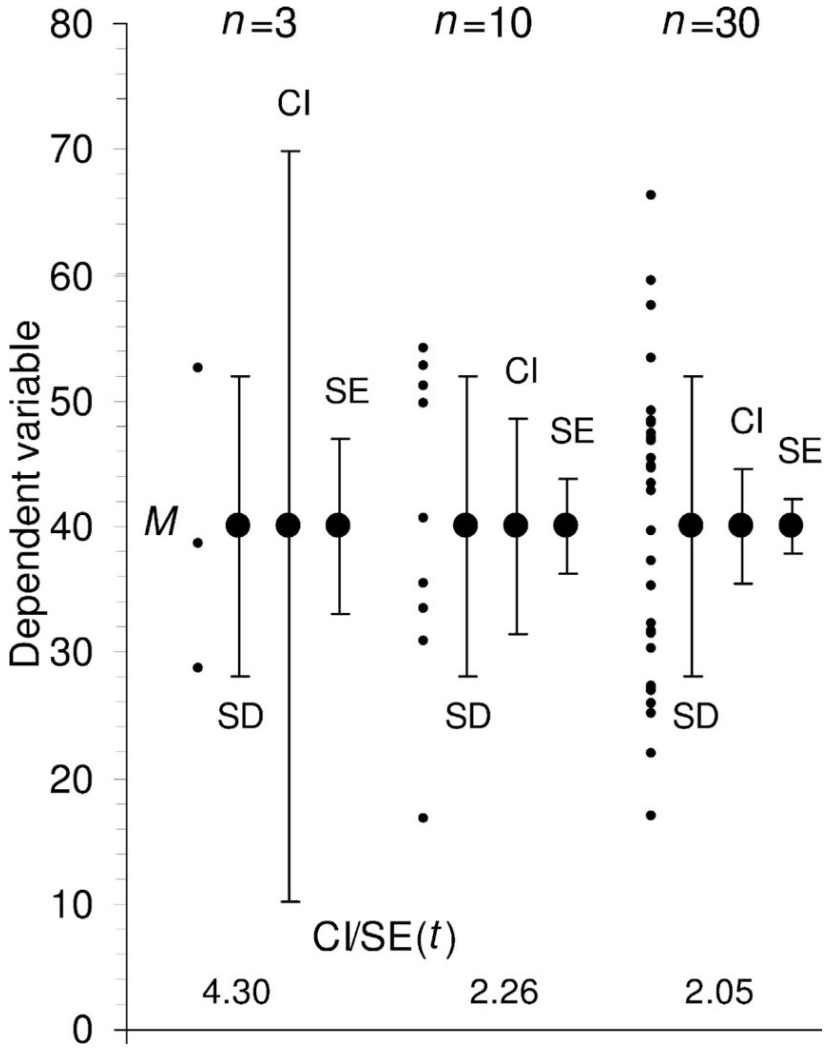
Sloupcový graf s chybou – ale kterou?



Popisná chyba:
SD - směrodatná odchylka

Inferenční (odvozovací) chyba:
SE - standardní chyba
CI - interval spolehlivosti

Sloupcový graf s chybou – ale kterou?



$$SD = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}}$$

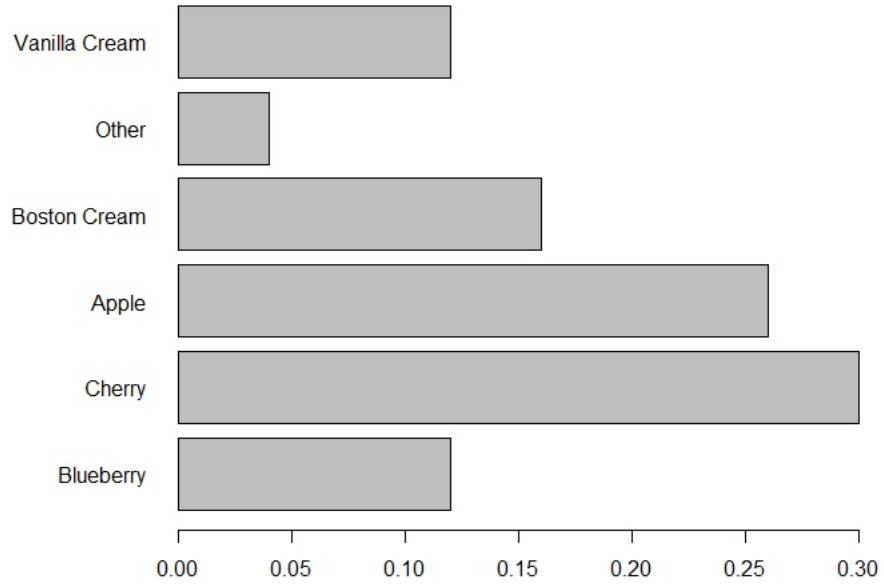
$$SE_{\bar{x}} = \frac{SD}{\sqrt{n}}$$

$$\bar{x} \pm 1.96 \times SE_{\bar{x}}$$

Grafy zobrazující **frekvenci kategoriálních** proměnných

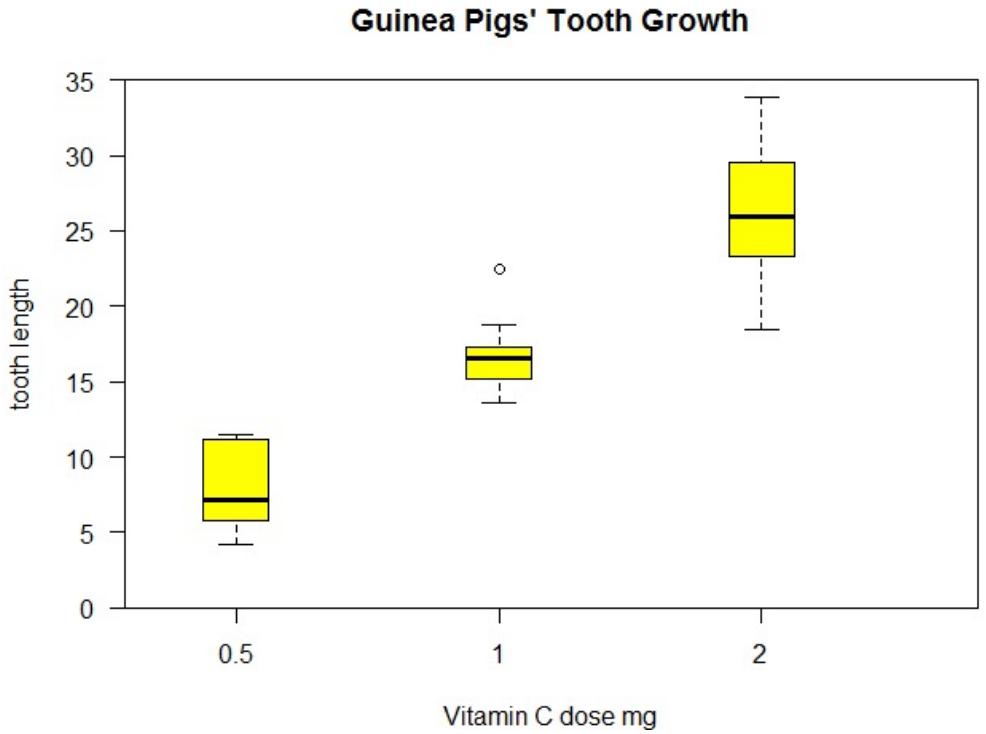
- Koláčový graf

- Sloupcový graf

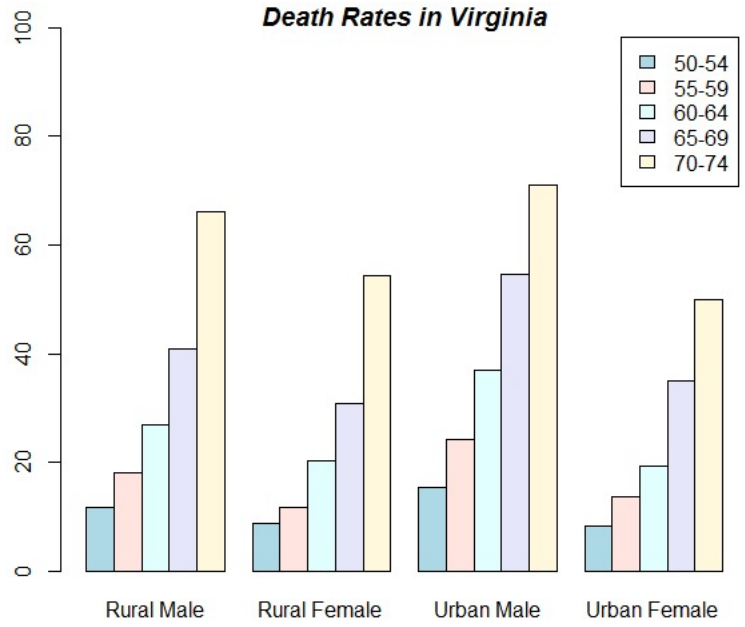


Grafy zobrazující asociaci **kategoriální** a **spojité** proměnné

- Krabicové grafy v kategoriích



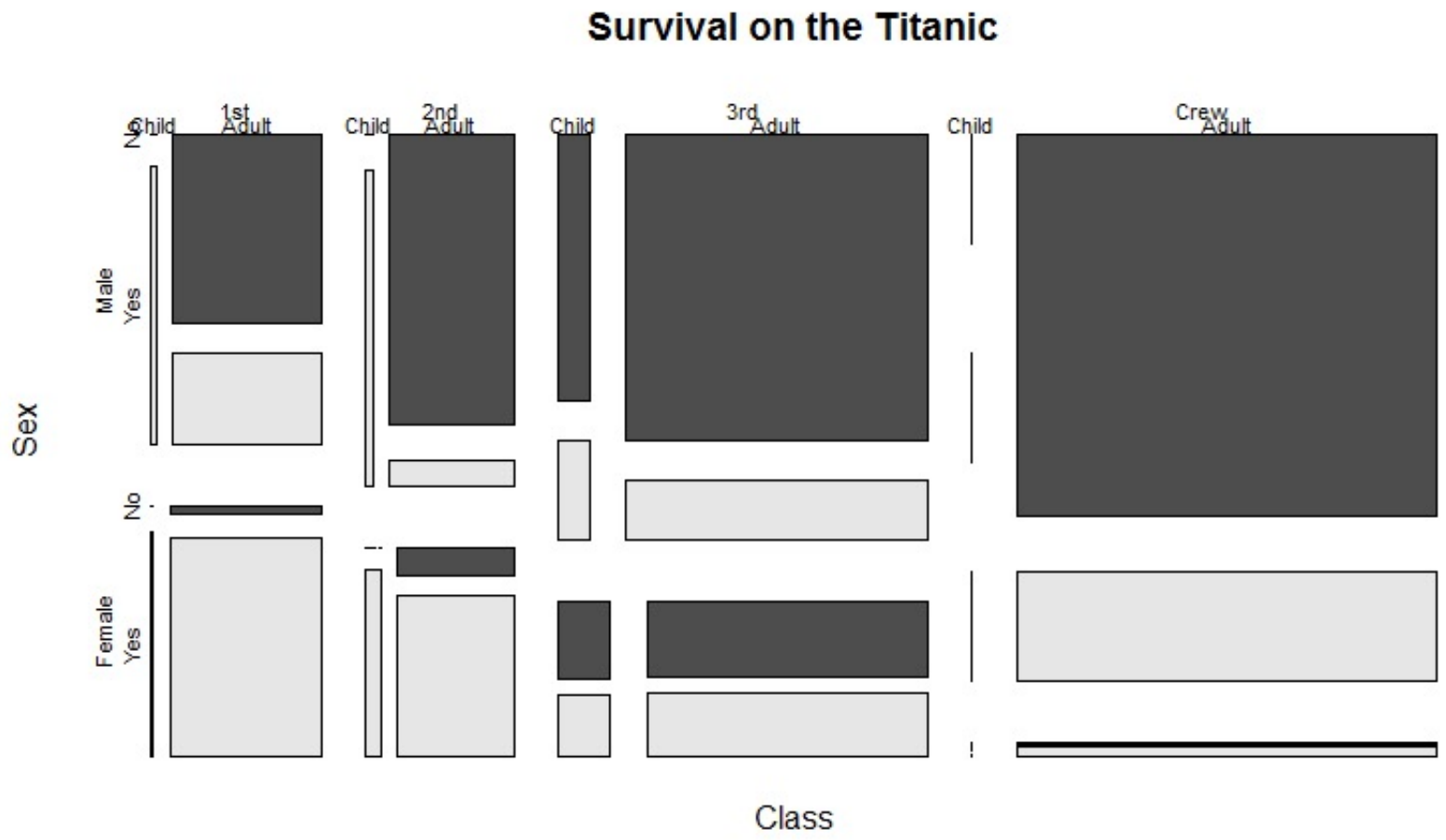
- Kategorizovaný sloupcový graf



MUNI | RECETOX

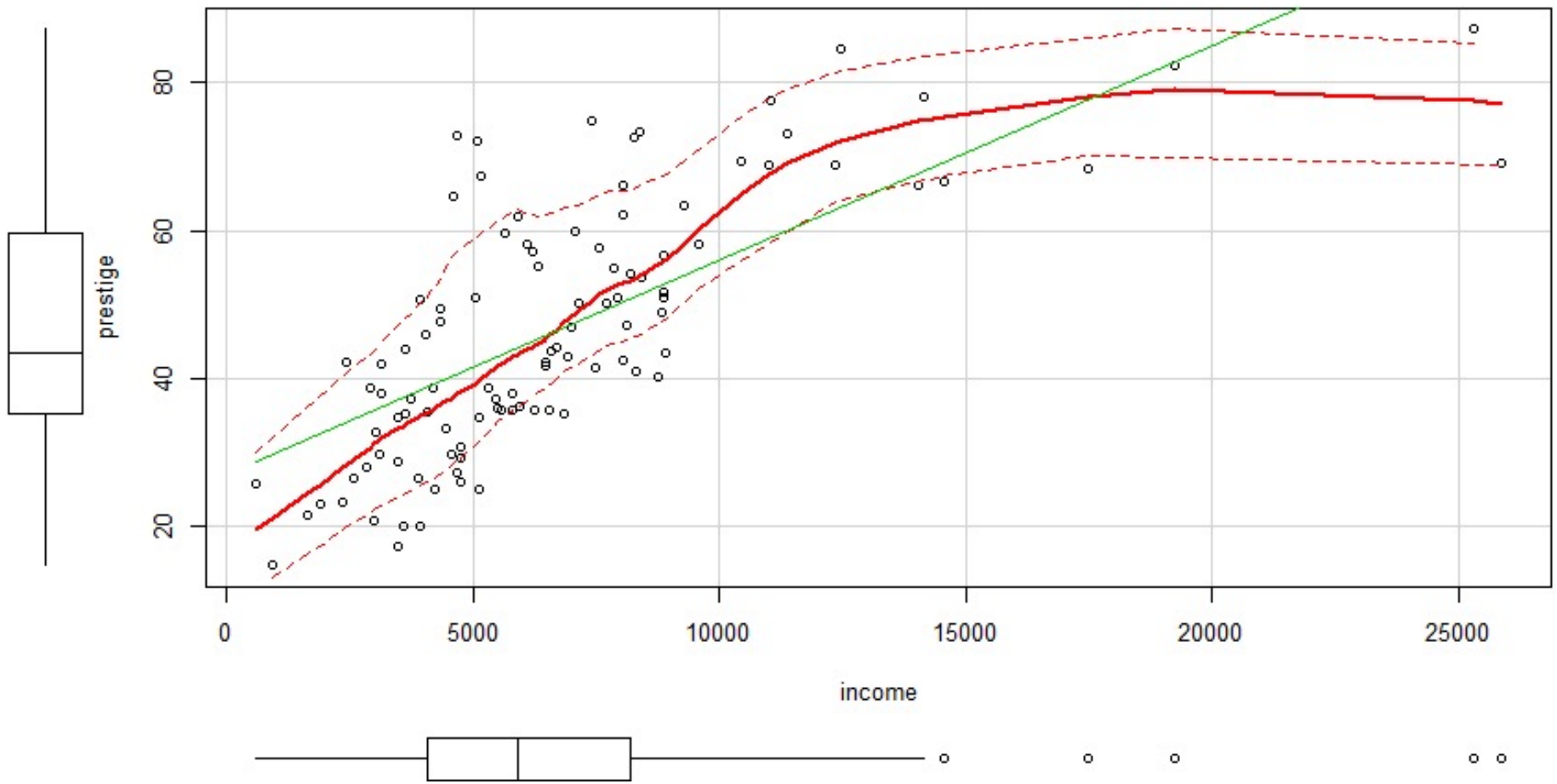
Grafy zobrazující asociaci **dvou kategoriálních** proměnných

- Mozaikový graf



Grafy zobrazující asociaci **dvou spojitých** proměnných

- Dvouroměrný x-y graf



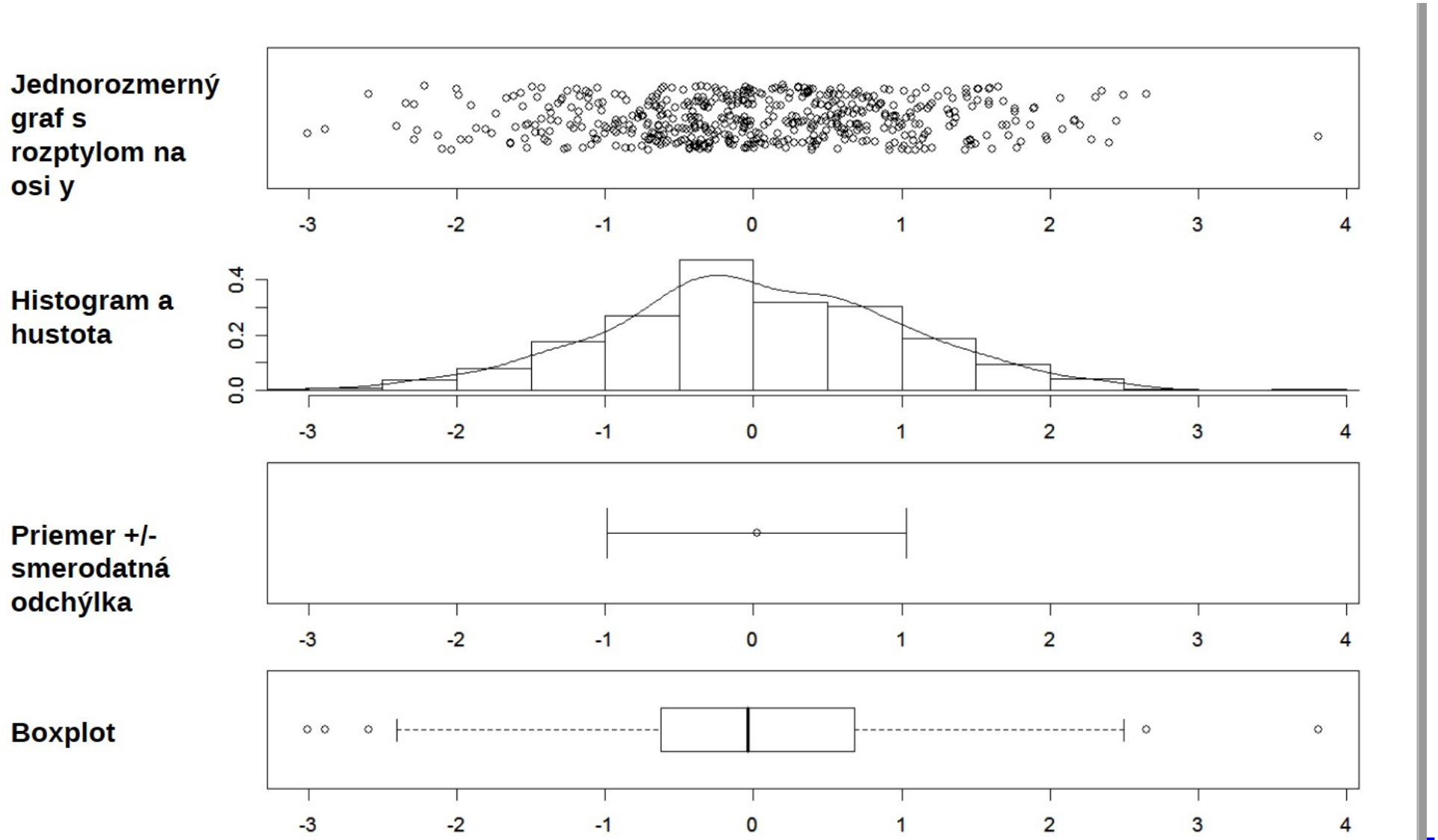
Cvičení

- Různé datové soubory vs. 4 metody zobrazení:

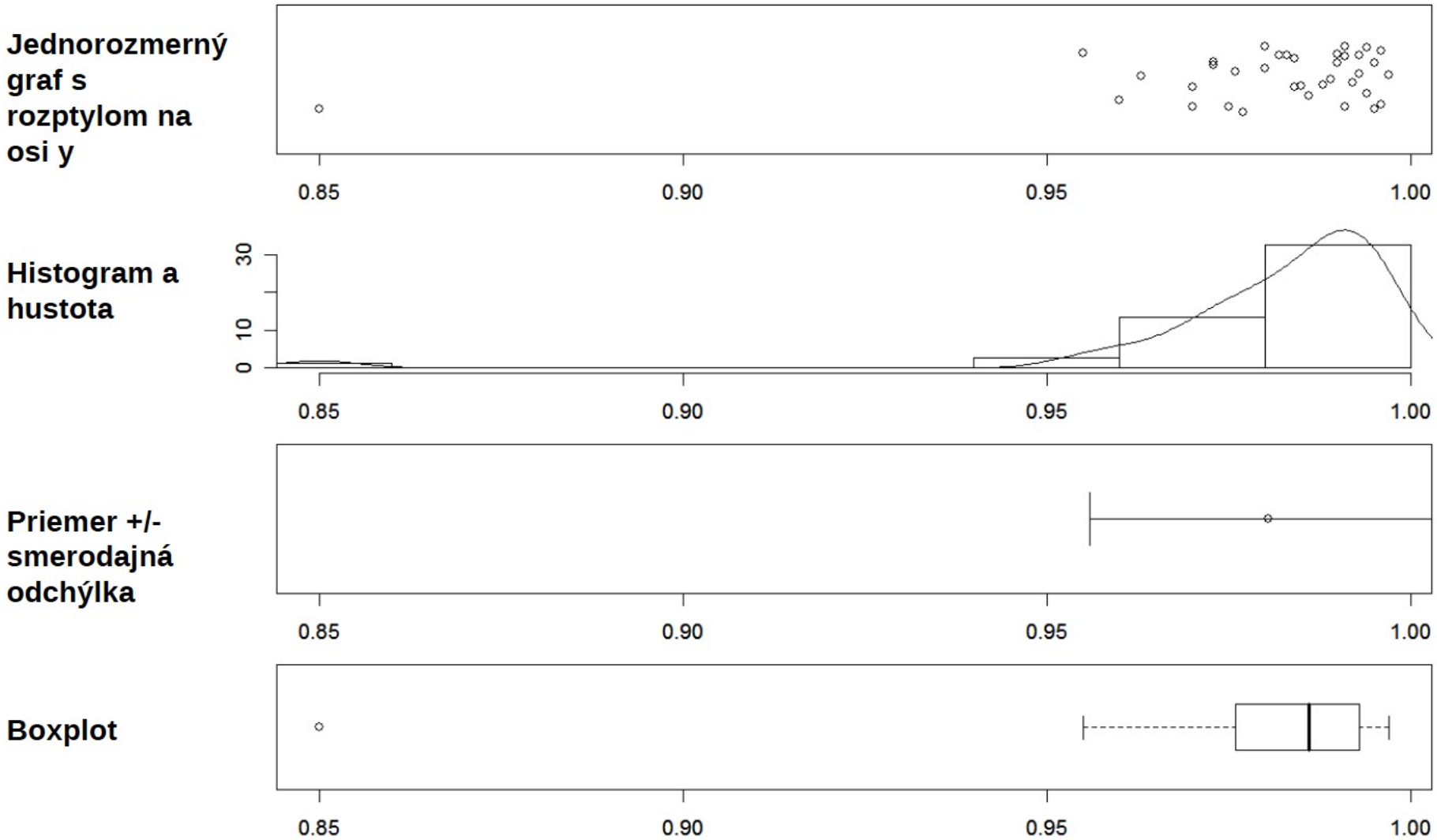
- Jednorozměrné individuální body (s rozptylem na ose y)
- Histogram s hustotou
- Průměr +/- směrodatná odchylka
- Boxplot

Vyberte **nejvíce** a **nejméně informativní** graf co se týká **rozložení hodnot**, u každého datového souboru.

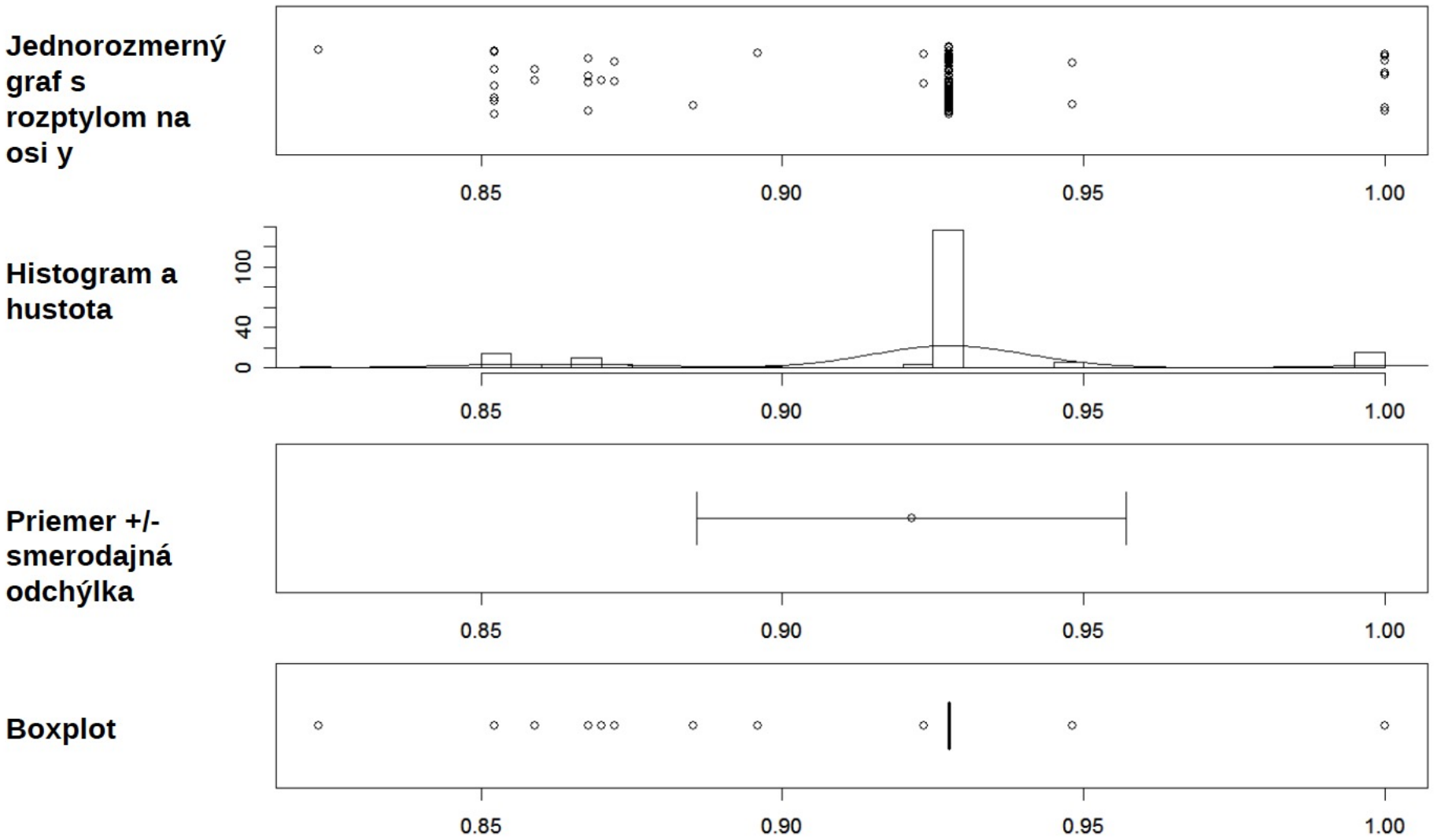
Příklad 1. Náhodné rozložení, N=400



Príklad 2. N=37



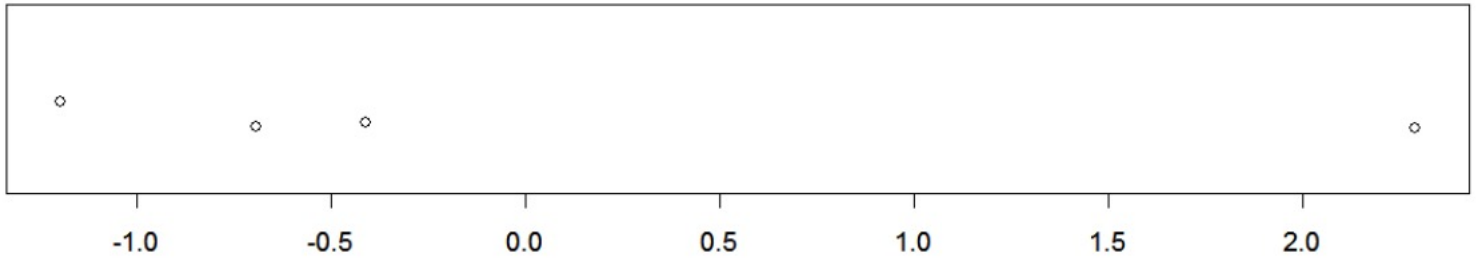
Príklad 3. N=100



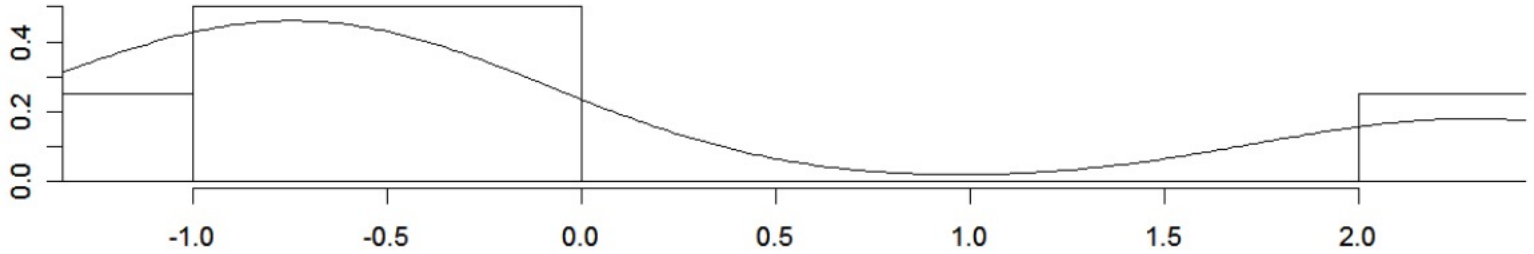
MUNI | RECETOX

Príklad 4. N=4

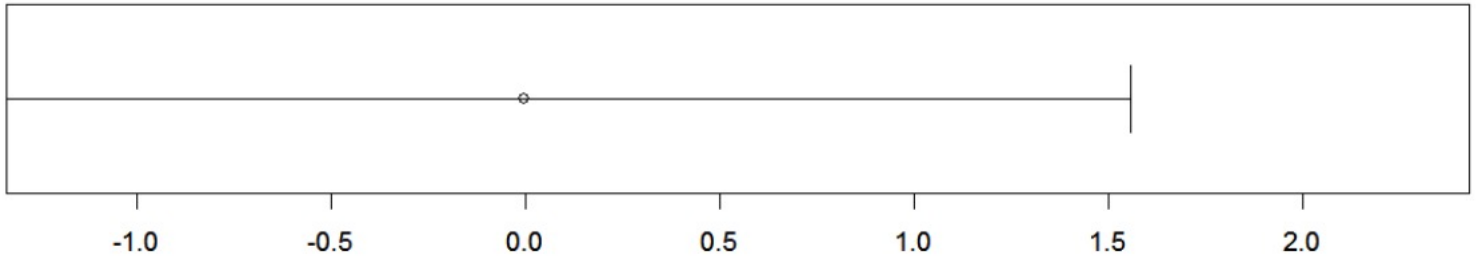
Jednorozmerný graf s rozptylom na osi y



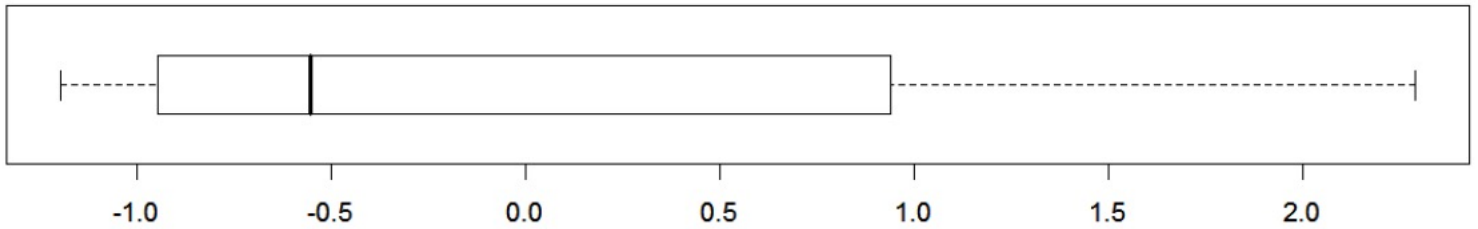
Histogram a hustota



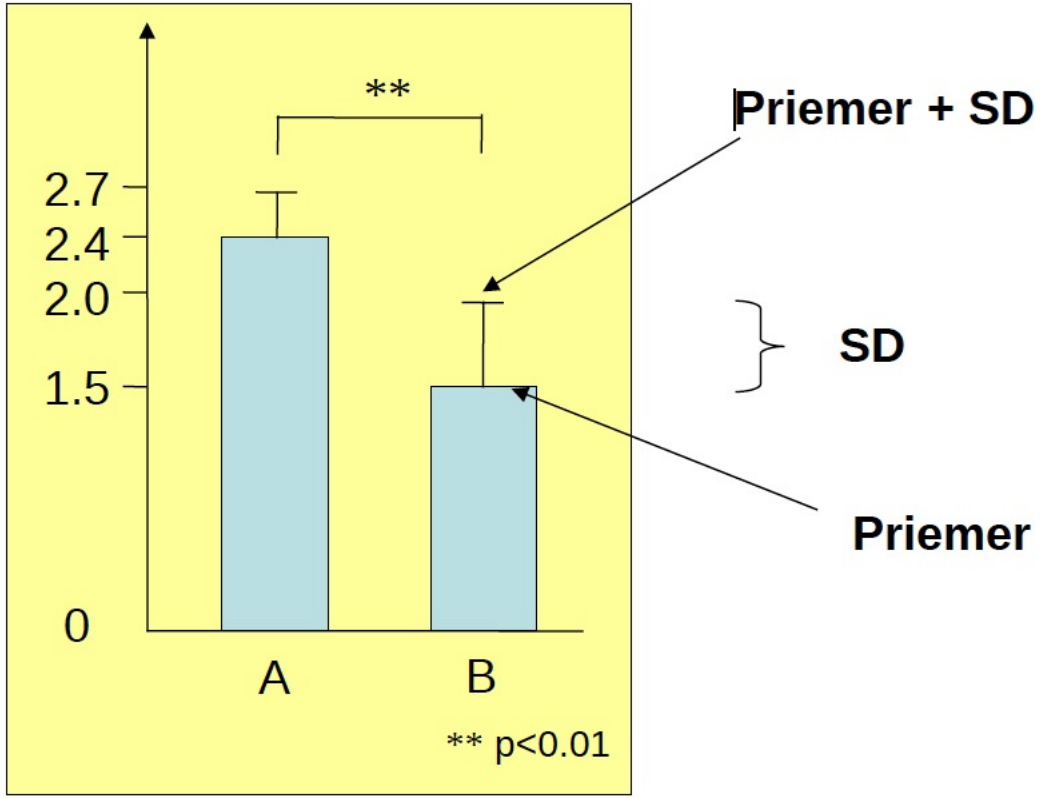
Priemer +/- smerodajná odchýlka



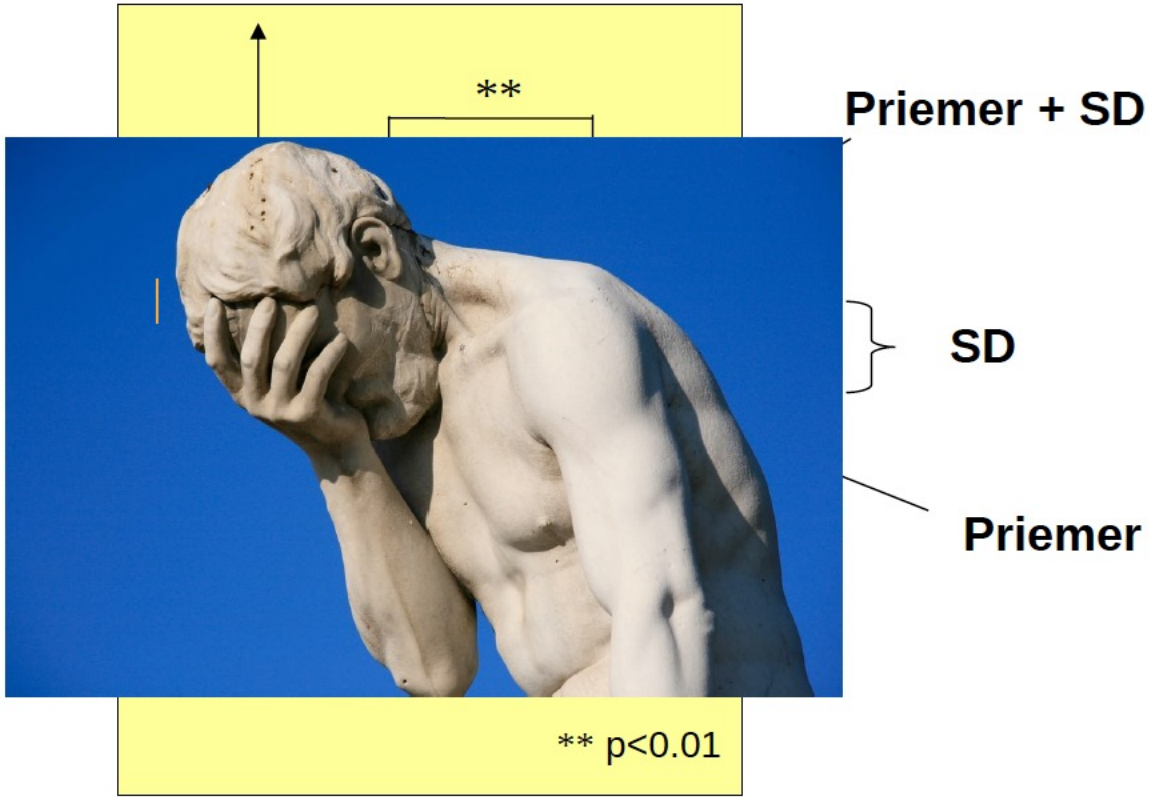
Boxplot



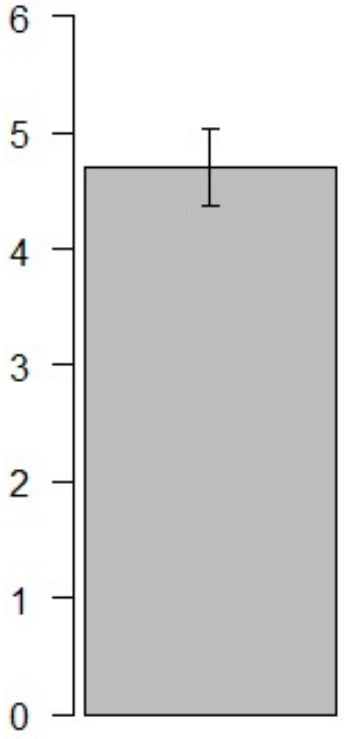
Proč se vyhýbat sloupcovým grafům s chybou



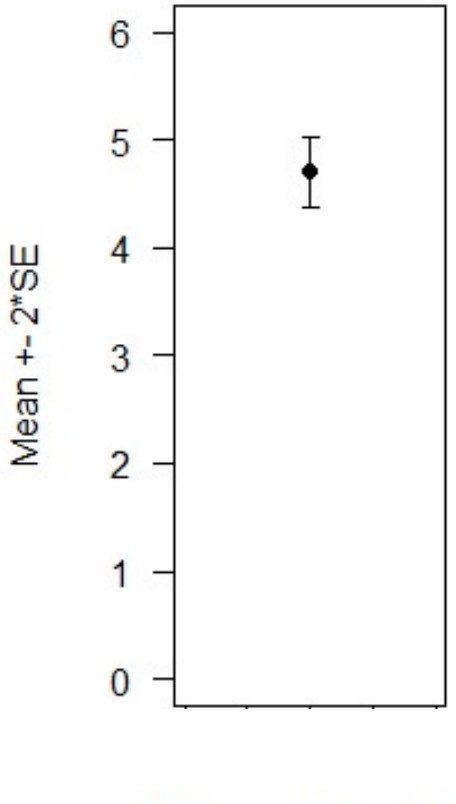
Proč se vyhýbat sloupcovým grafům s chybou



Proč se vyhýbat sloupcovým grafům s chybou



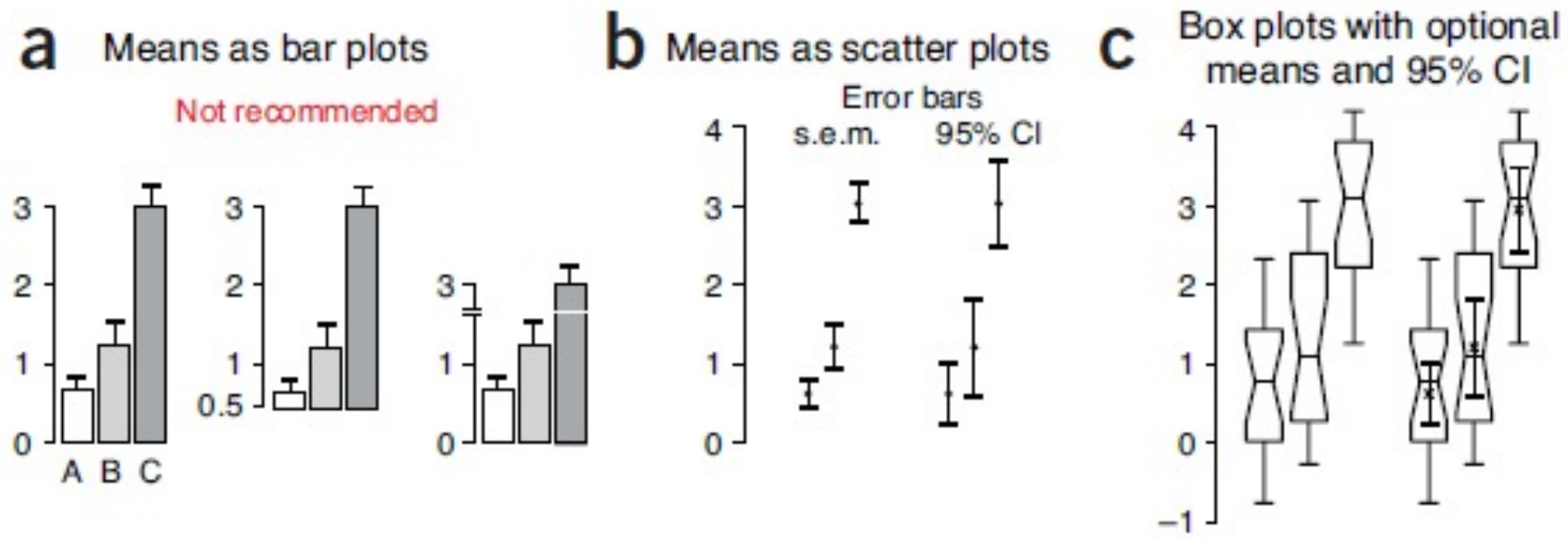
Mají příliš nízký poměr data/atrament



Alternativa bez sloupce

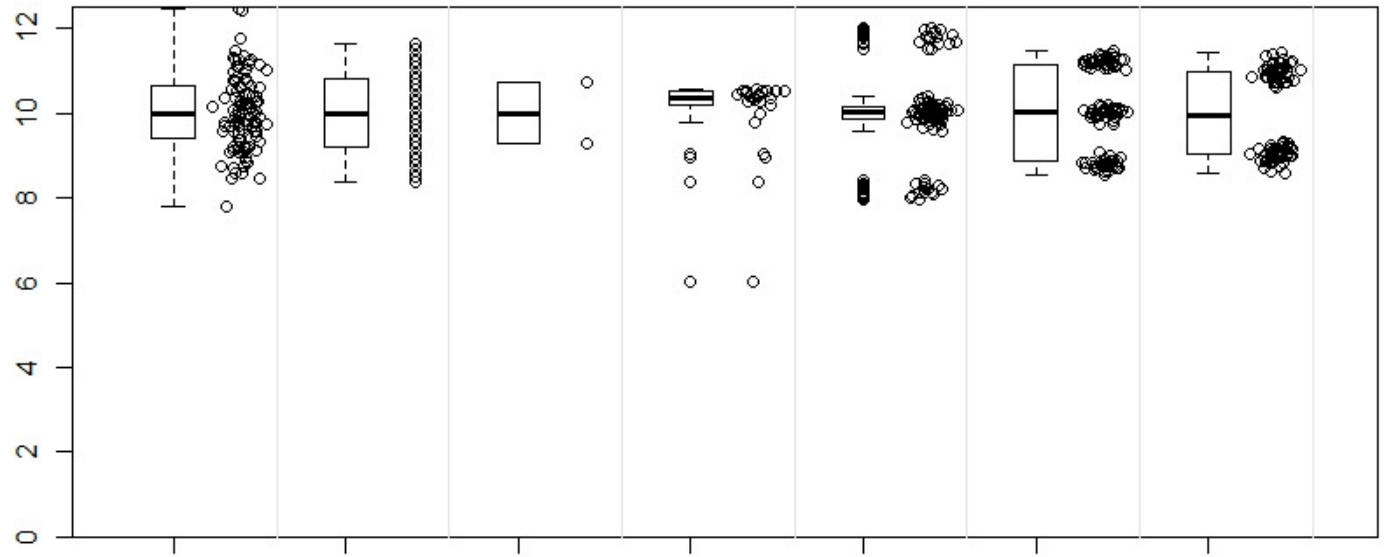
Proč se vyhýbat sloupcovým grafům s chybou

Často trpí neduhem **– zkreslení škálou**

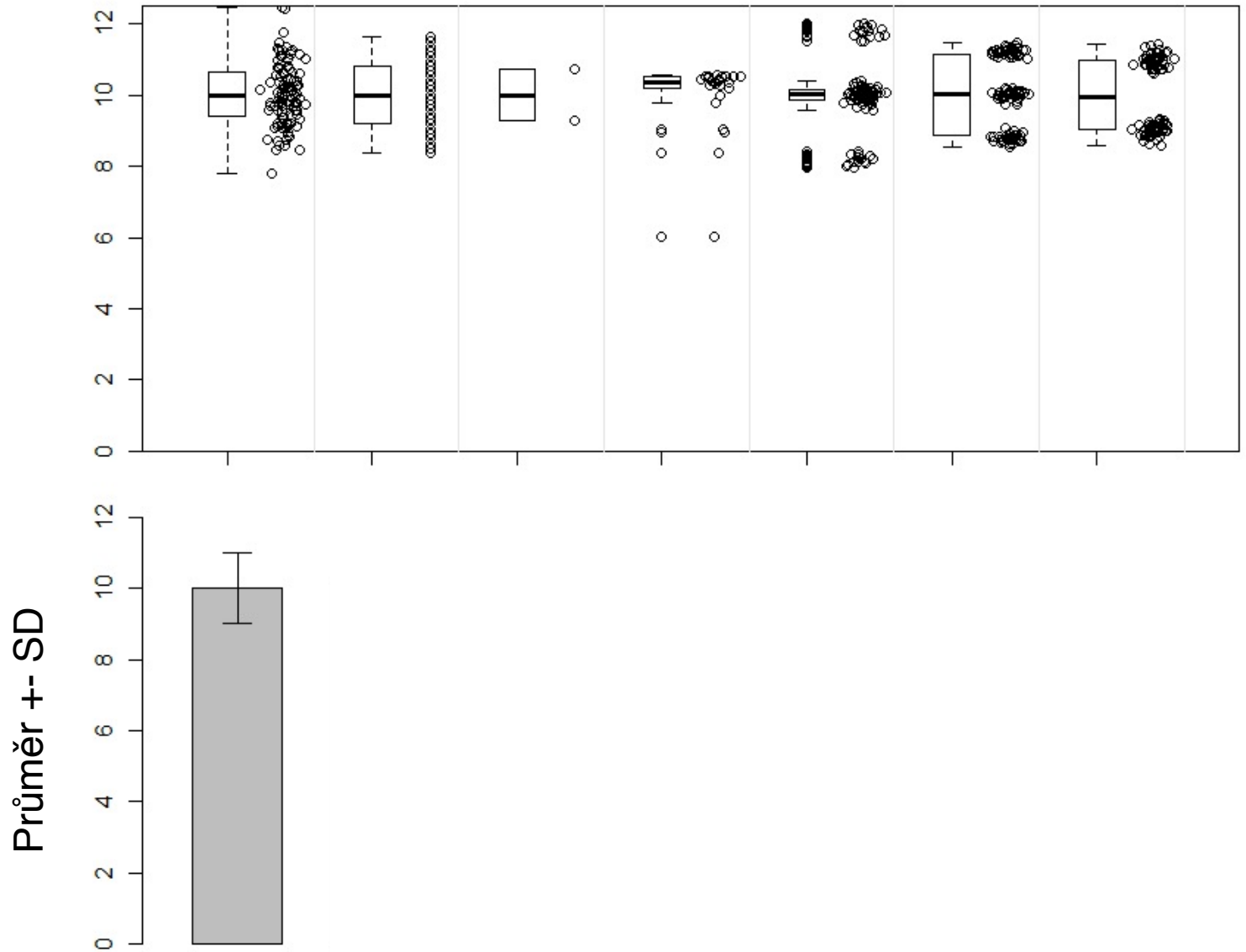


Krzywinski M, Altman N. (2014) Visualizing samples with box plots. Nat Methods. 2014 Feb;11(2):119-20.

Proč se vyhýbat sloupcovým grafům s chybou

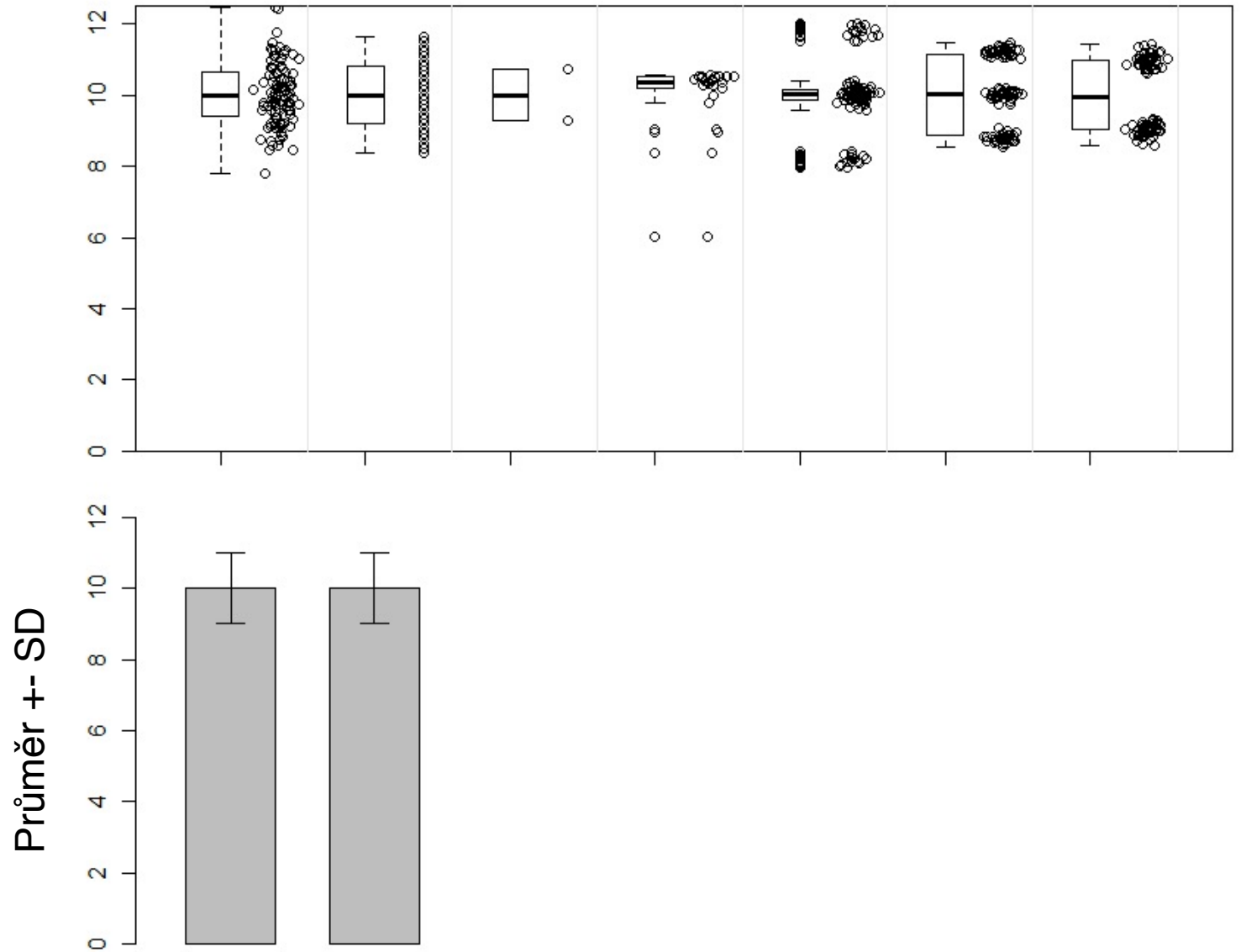


Proč se vyhýbat sloupcovým grafům s chybou



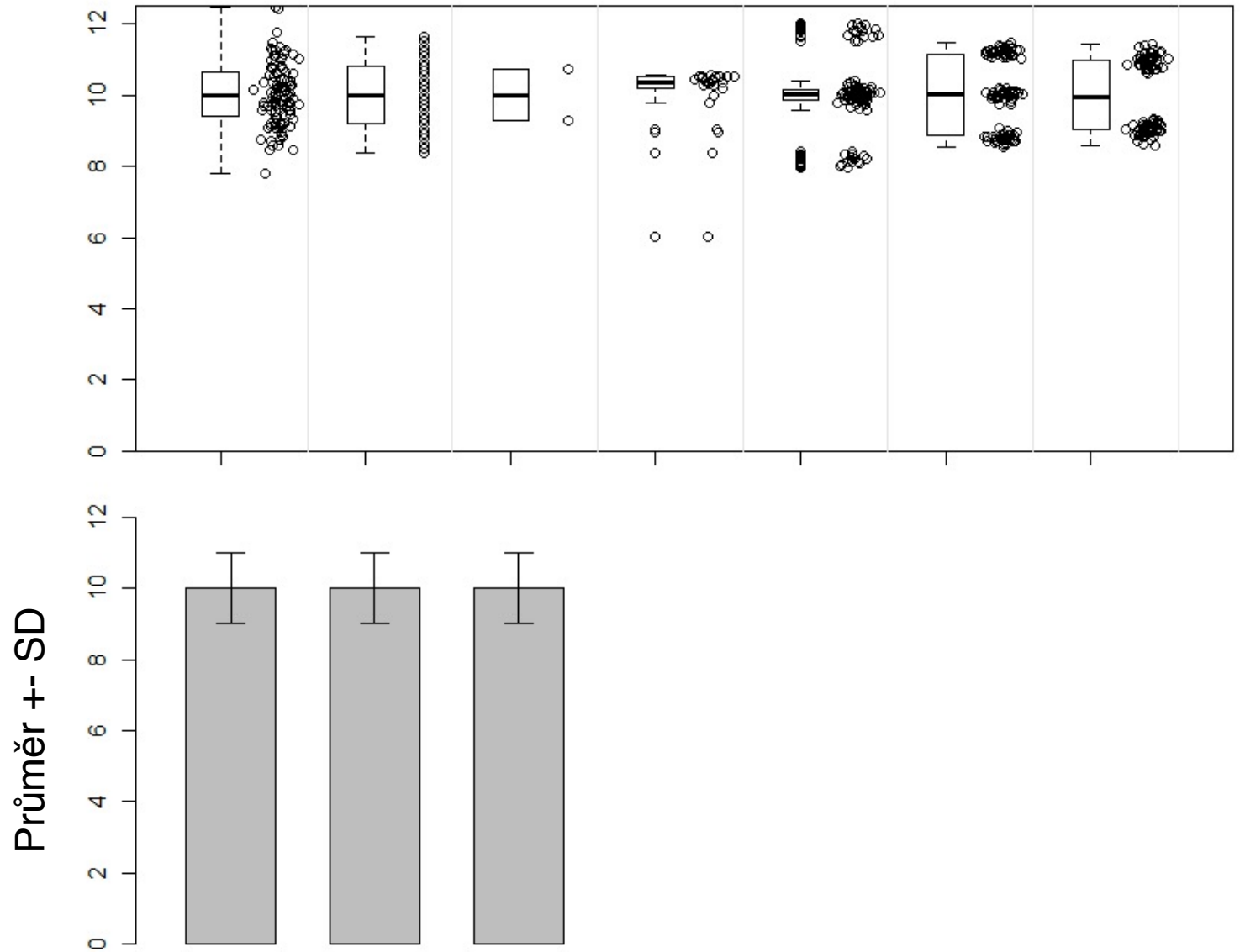
RECETOX

Proč se vyhýbat sloupcovým grafům s chybou



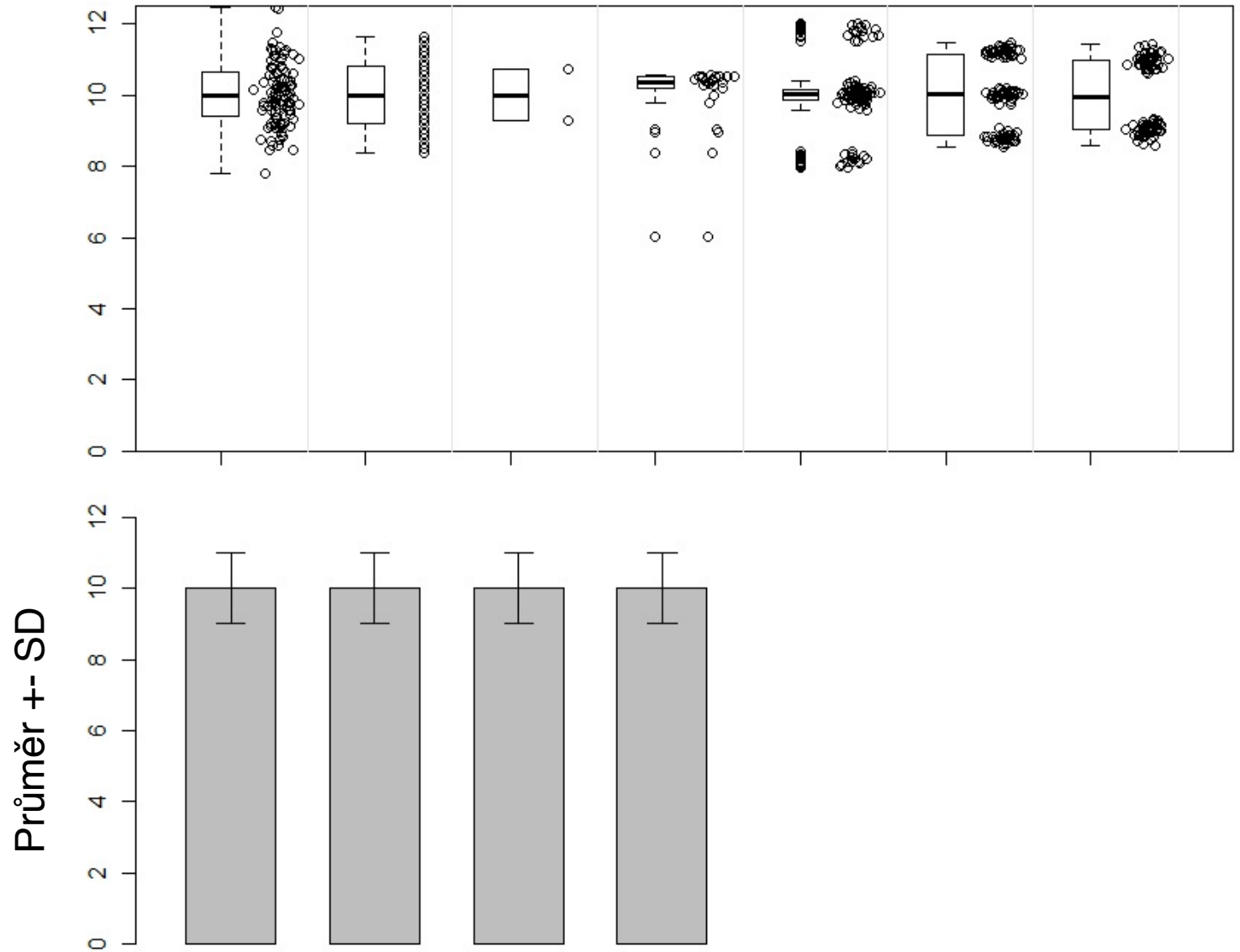
RECETOX

Proč se vyhýbat sloupcovým grafům s chybou



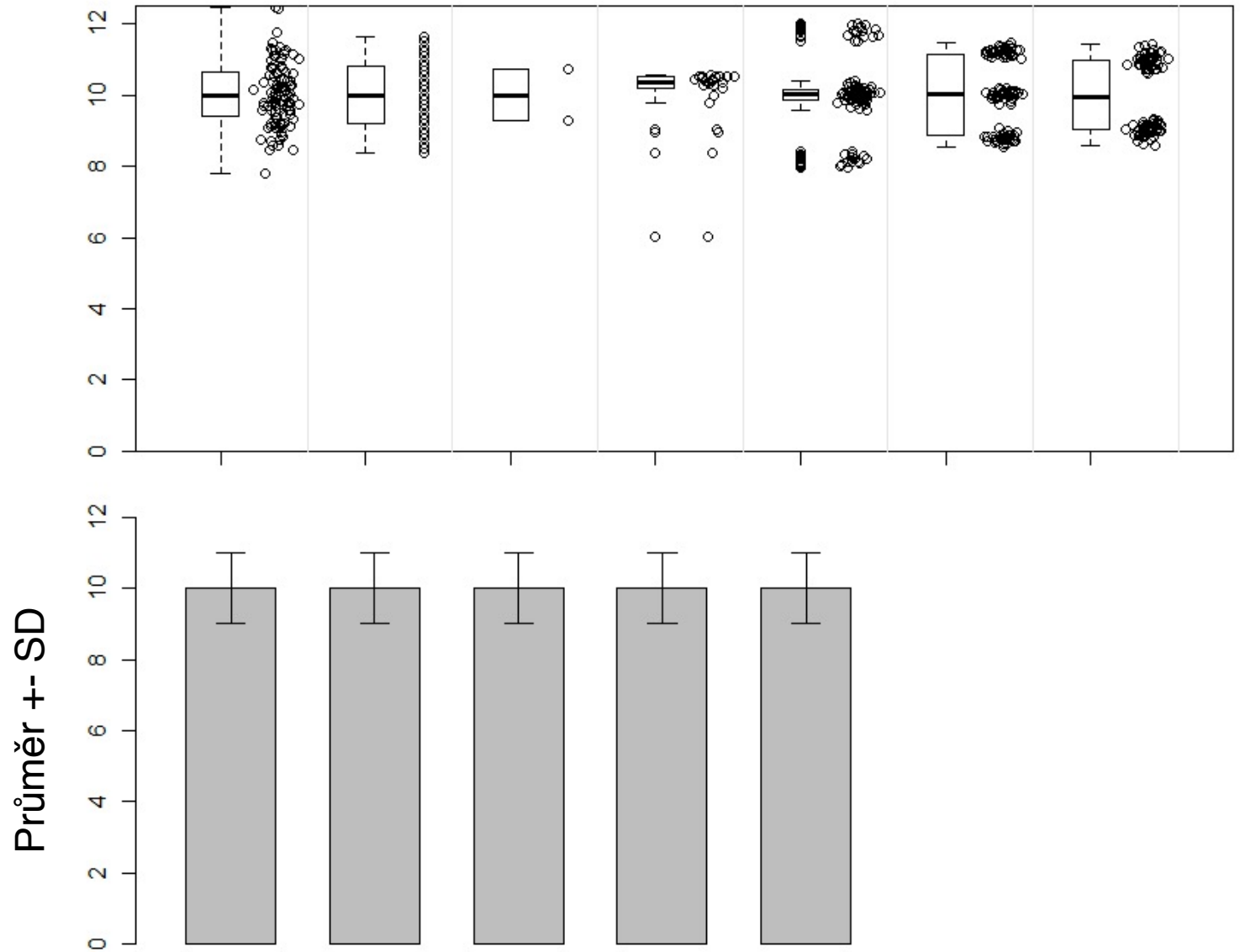
RECETOX

Proč se vyhýbat sloupcovým grafům s chybou



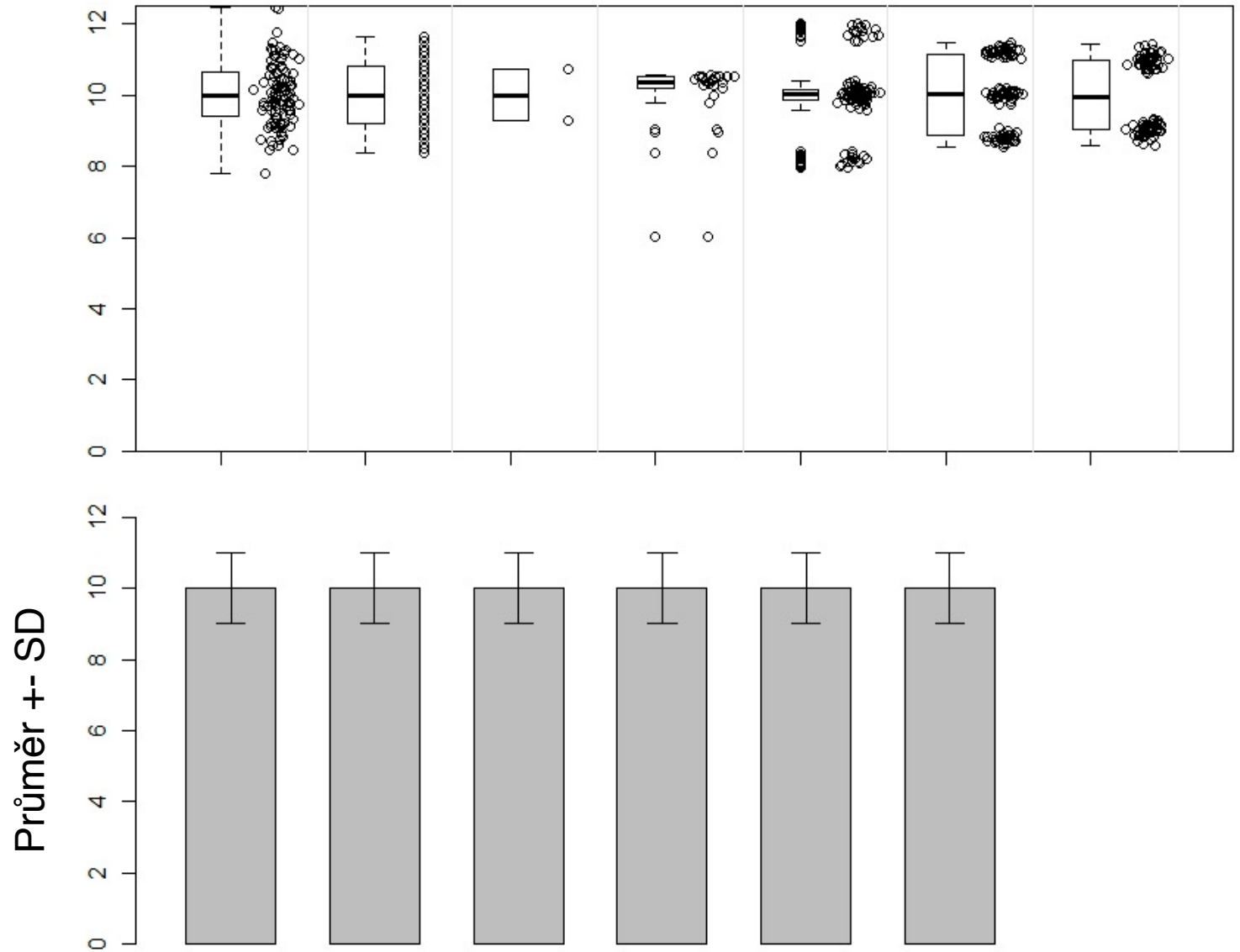
RECETOX

Proč se vyhýbat sloupcovým grafům s chybou



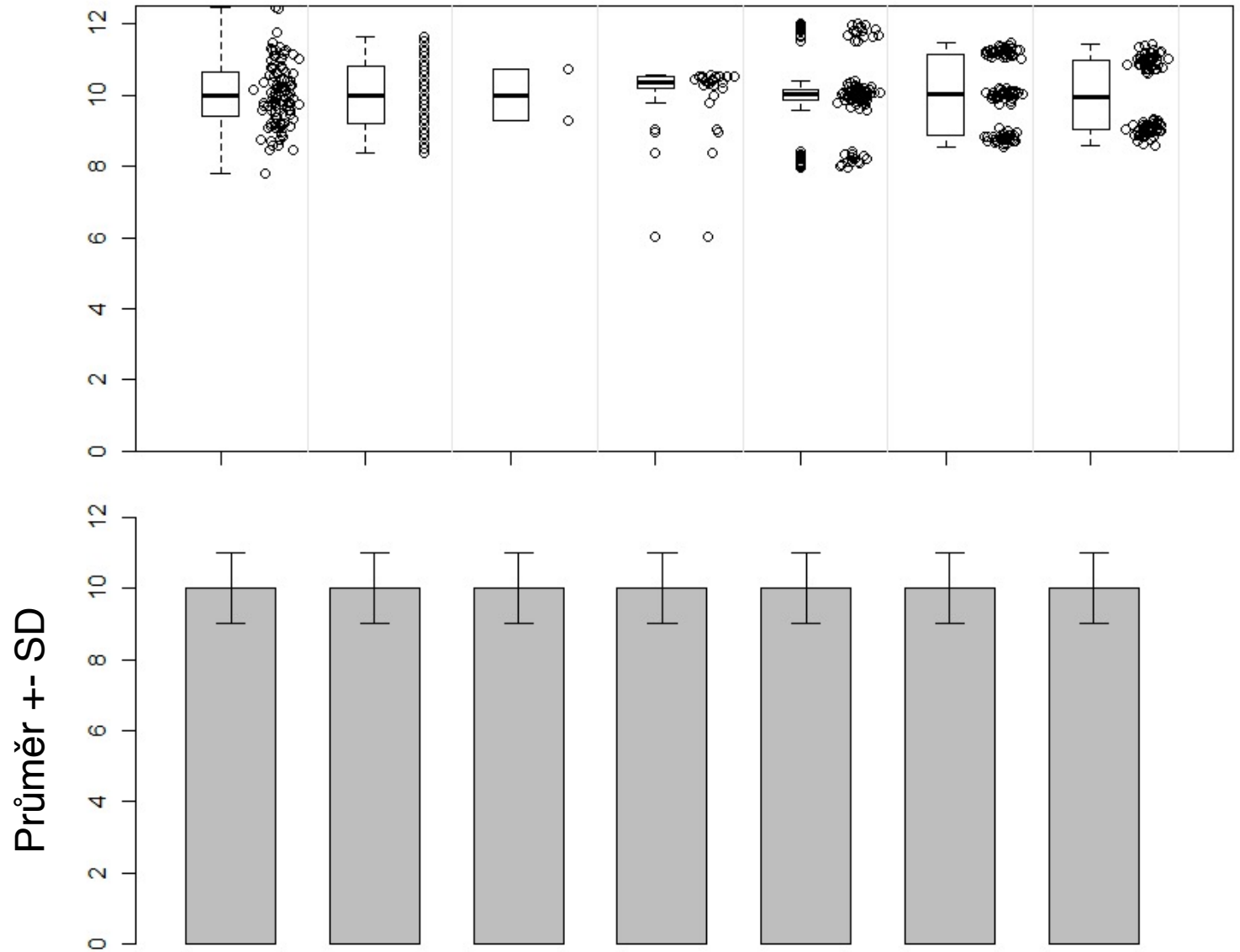
RECETOX

Proč se vyhýbat sloupcovým grafům s chybou



RECETOX

Proč se vyhýbat sloupcovým grafům s chybou



RECETOX

Kdy nejsou vhodné ani krabicové grafy

Když je jen málo bodů na zobrazení, krabicové grafy postrádají význam, jsou degenerované

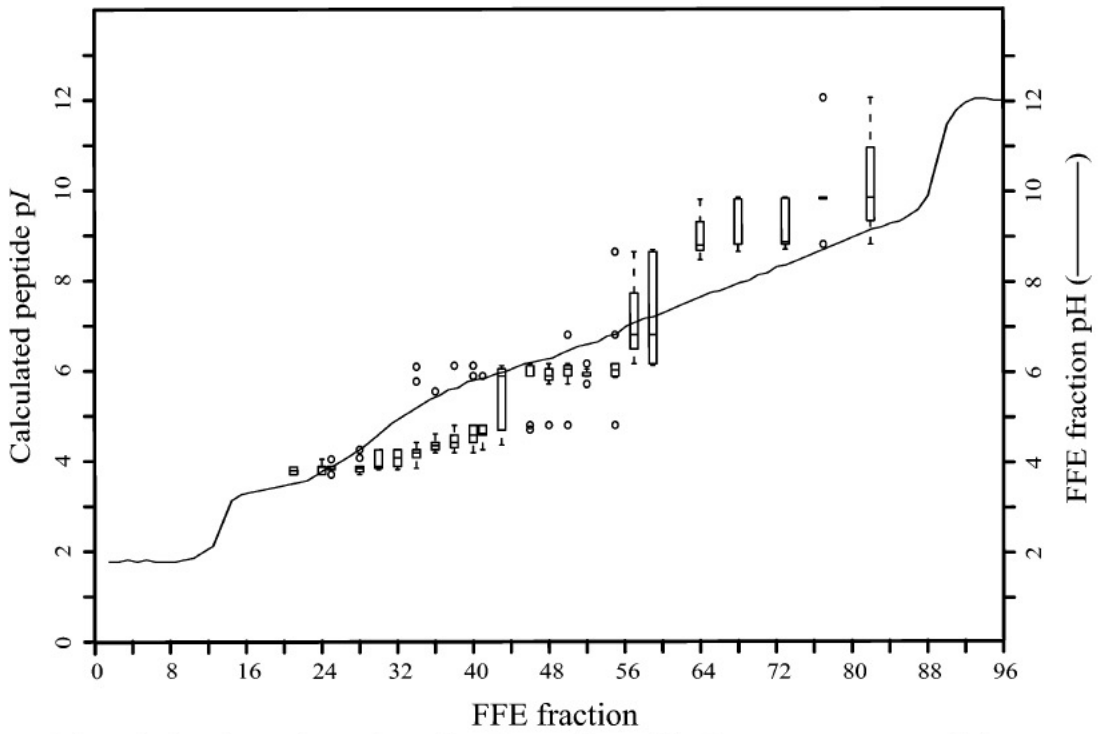
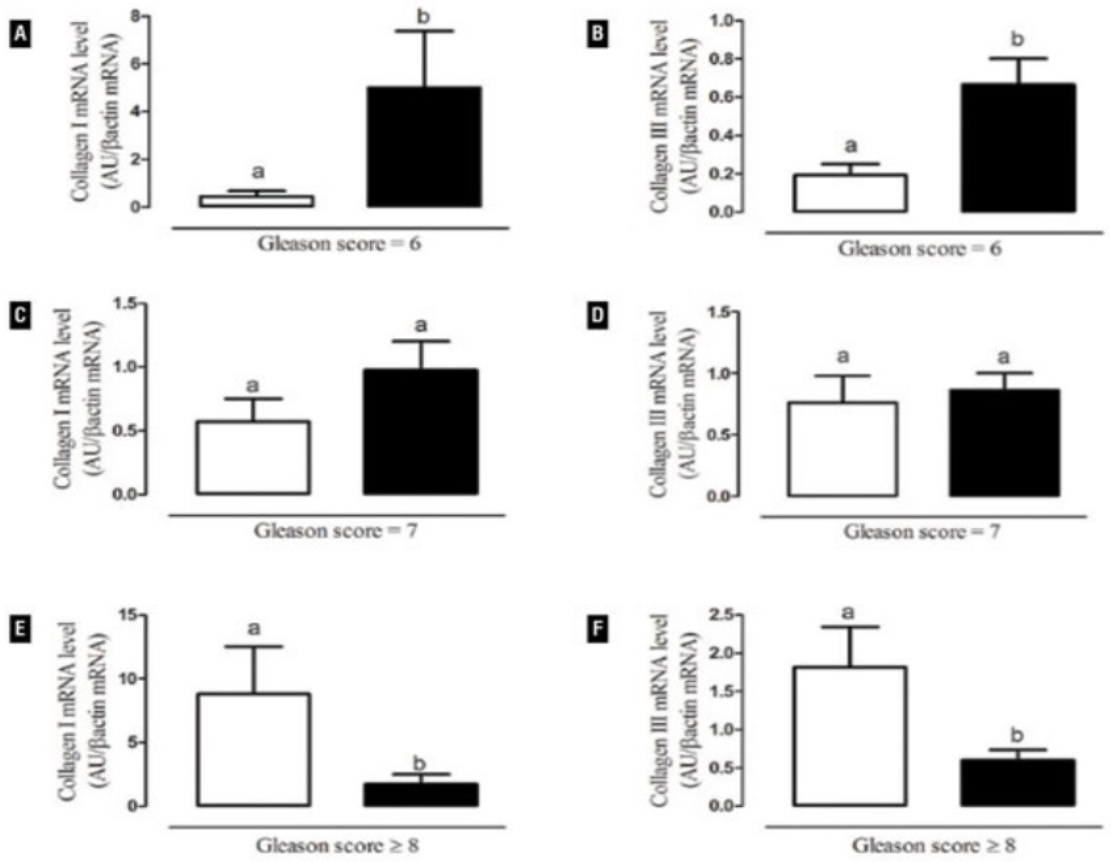


Figure 9. Comparison of theoretical and experimental p/ values for tryptic peptides from LIM 1215 cytosolic lysate separated by 2D FFE-IEF (pH 3−10)/RP-HPLC. Box plots were automatically generated using the statistical package R, version 1.5.0 (http://www.r-project.org) using the default parameters (i.e., the box represents scores between 25 and 75%, with the median at 50%, outliners are scores >1.5 times the interquartile range (75−25%) from the box and are indicated by dots (○), whiskers (+ − +) extend to the highest or lowest score not considered to be an outlier).

From Moritz et al., Anal. Chem. 2004 Aug 15; 76(16):4811-24

Nájdite aspoň 3 chyby

Figure 2 - Gene expression of collagen I and III in the benign (white bar) and tumor (black bar) areas of prostate of Gleason score=6 (A,B), Gleason score=7 (C,D) and Gleason score≥8 (E,F). β actin was used as an internal control. Data are represented as means ± SEM. Sample numbers were 13 for Gleason score=6; 10 for Gleason score=7 and 10 for Gleason score≥8. Different letters mean statistical significance evaluated by Student's-t-test.

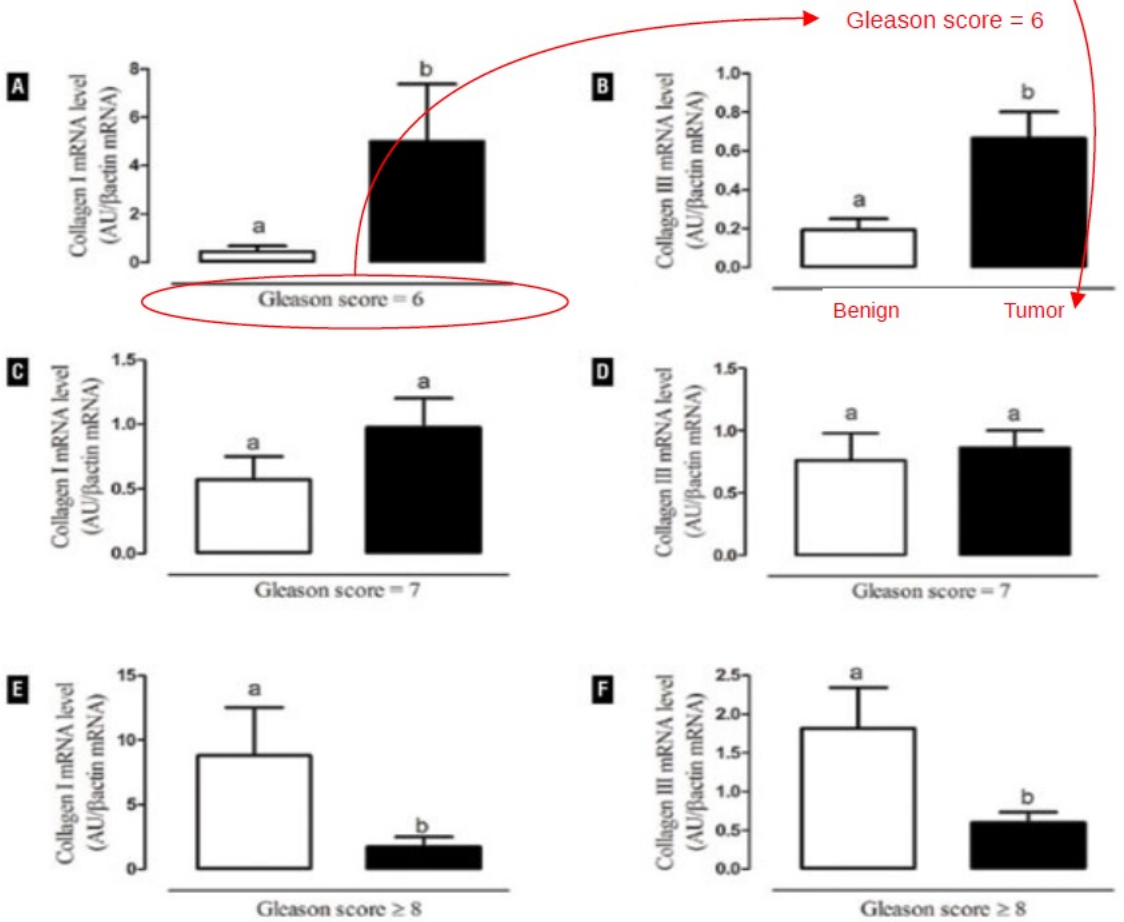


Antonio H. Duarte; Sicilia Colli; Jorge L. Alves-Pereira; Max P. Martins; Francisco J. B. Sampaio; Cristiane F. Ramos. Collagen I and III and metalloproteinase gene and protein expression in prostate cancer in relation to Gleason score. Int. braz j urol. vol.38 no.3 Rio de Janeiro May/June 2012

ETOX

Nájdite 3 chyby

Figure 2 - Gene expression of collagen I and III in the benign (white bar) and tumor (black bar) areas of prostate of Gleason score=6 (A,B), Gleason score=7 (C,D) and Gleason score≥8 (E,F). β actin was used as an internal control. Data are represented as means ± SEM. Sample numbers were 13 for Gleason score=6; 10 for Gleason score=7 and 10 for Gleason score≥8. Different letters mean statistical significance evaluated by Student's-t-test.



Gleason score = 6

1. Chýbajúci popis kategórií boxplotov na osi x

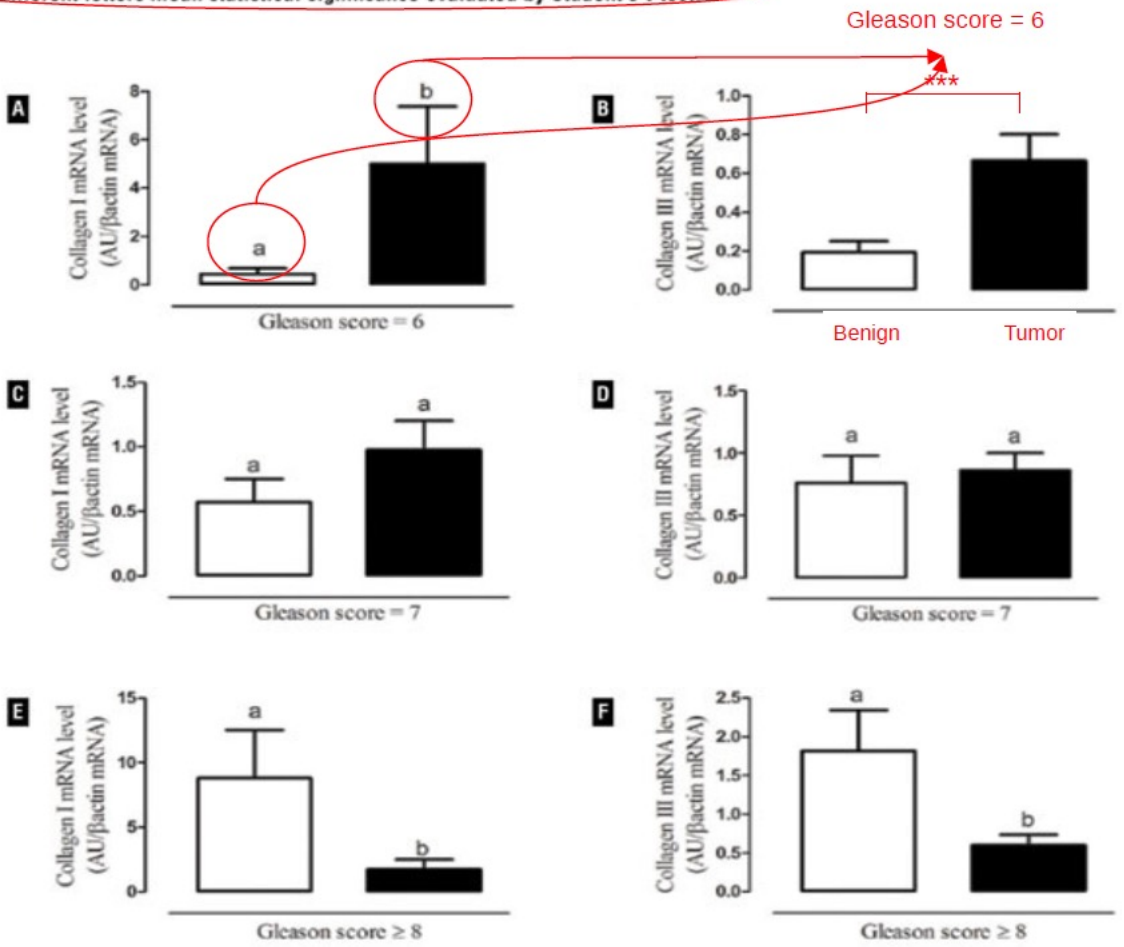
2. Popis skóre uvedený nesprávne na osi x

Antonio H. Duarte; Sicilia Colli; Jorge L. Alves-Pereira; Max P. Martins; Francisco J. B. Sampaio; Cristiane F. Ramos. Collagen I and III and metalloproteinase gene and protein expression in prostate cancer in relation to Gleason score. Int. braz j urol. vol.38 no.3 Rio de Janeiro May/June 2012

ETOX

2. Nevhodné zobrazenie štatistickej významnosti

Figure 2 - Gene expression of collagen I and III in the benign (white bar) and tumor (black bar) areas of prostate of Gleason score=6 (A,B), Gleason score=7 (C,D) and Gleason score≥8 (E,F). β actin was used as an internal control. Data are represented as means ± SEM. Sample numbers were 13 for Gleason score=6; 10 for Gleason score=7 and 10 for Gleason score≥8. Different letters mean statistical significance evaluated by Student's-t-test.



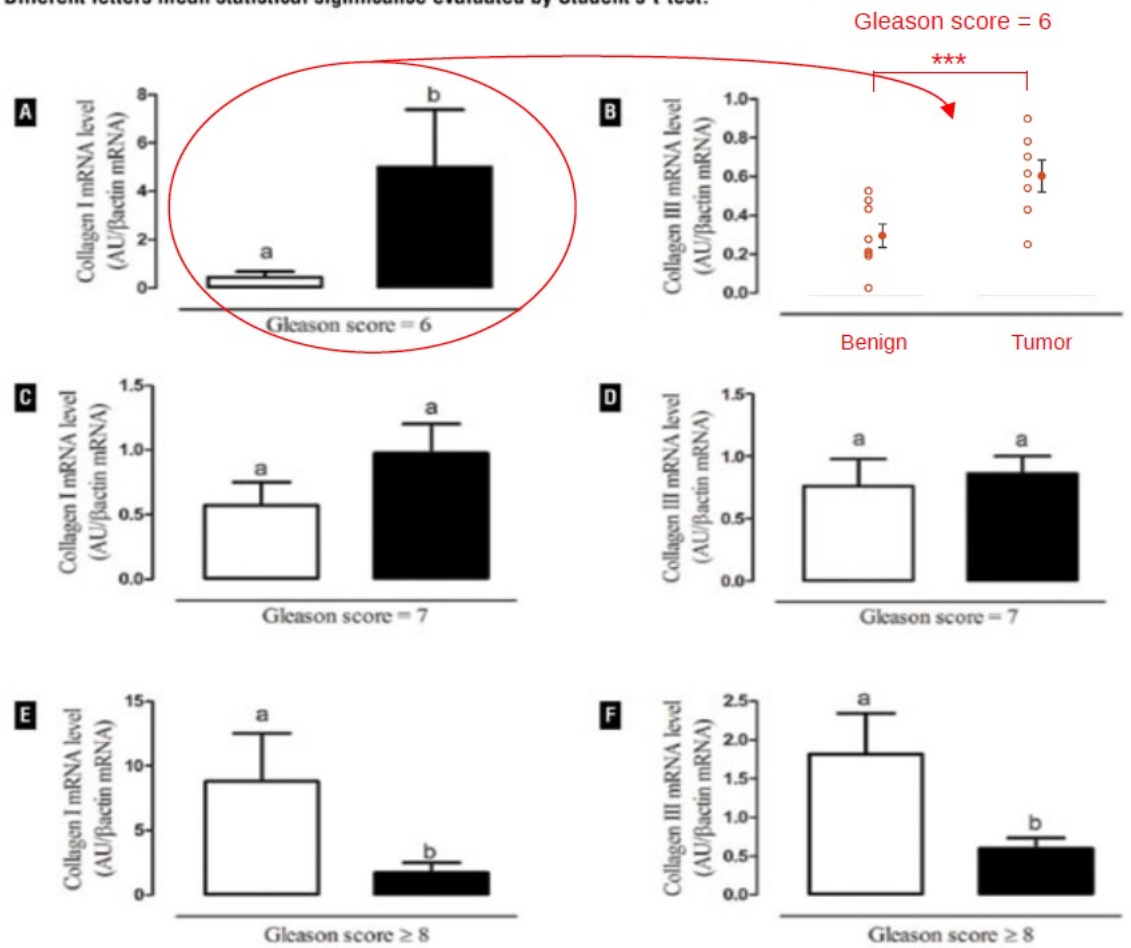
Veľmi neštandardné, navyše písmená a, b vyzerajú ako názvy kategórií/boxplotov

Antonio H. Duarte; Sicilia Colli; Jorge L. Alves-Pereira; Max P. Martins; Francisco J. B. Sampaio; Cristiane F. Ramos. Collagen I and III and metalloproteinase gene and protein expression in prostate cancer in relation to Gleason score. Int. braz j urol. vol.38 no.3 Rio de Janeiro May/June 2012

ETOX

3. Nevhodný graf !

Figure 2 - Gene expression of collagen I and III in the benign (white bar) and tumor (black bar) areas of prostate of Gleason score=6 (A,B), Gleason score=7 (C,D) and Gleason score≥8 (E,F). β actin was used as an internal control. Data are represented as means ± SEM. Sample numbers were 13 for Gleason score=6; 10 for Gleason score=7 and 10 for Gleason score≥8. Different letters mean statistical significance evaluated by Student's-t-test.



Menej atramentu, viac informácie!

N=13

N=10

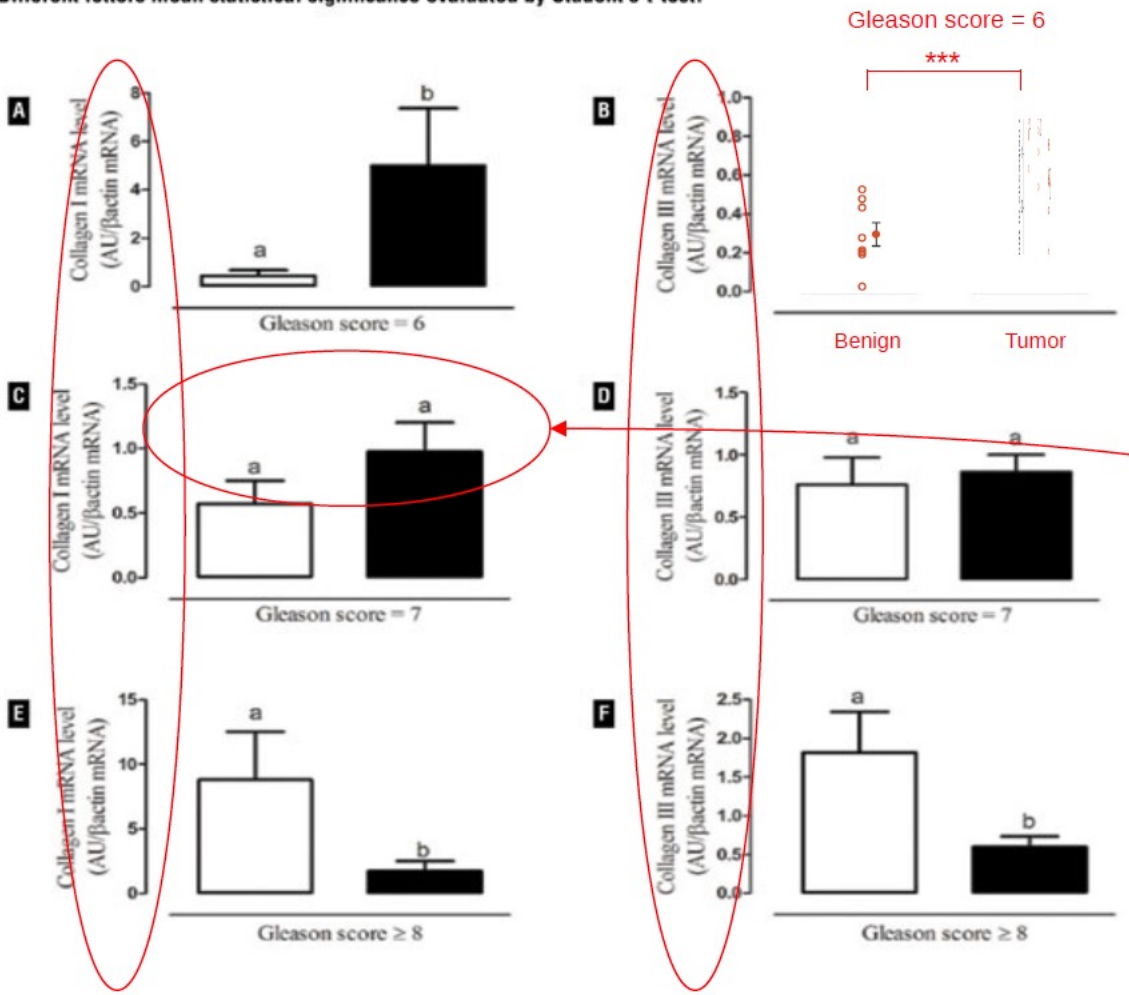
N=10

Antonio H. Duarte; Sicilia Colli; Jorge L. Alves-Pereira; Max P. Martins; Francisco J. B. Sampaio; Cristiane F. Ramos. Collagen I and III and metalloproteinase gene and protein expression in prostate cancer in relation to Gleason score. Int. braz j urol. vol.38 no.3 Rio de Janeiro May/June 2012

ETOX

4. Odlišné škály na osi y

Figure 2 - Gene expression of collagen I and III in the benign (white bar) and tumor (black bar) areas of prostate of Gleason score=6 (A,B), Gleason score=7 (C,D) and Gleason score≥8 (E,F). β actin was used as an internal control. Data are represented as means ± SEM. Sample numbers were 13 for Gleason score=6; 10 for Gleason score=7 and 10 for Gleason score≥8. Different letters mean statistical significance evaluated by Student's-t-test.



Odlišná škála pôsobí dojmom veľkého rozdielu v expresii!

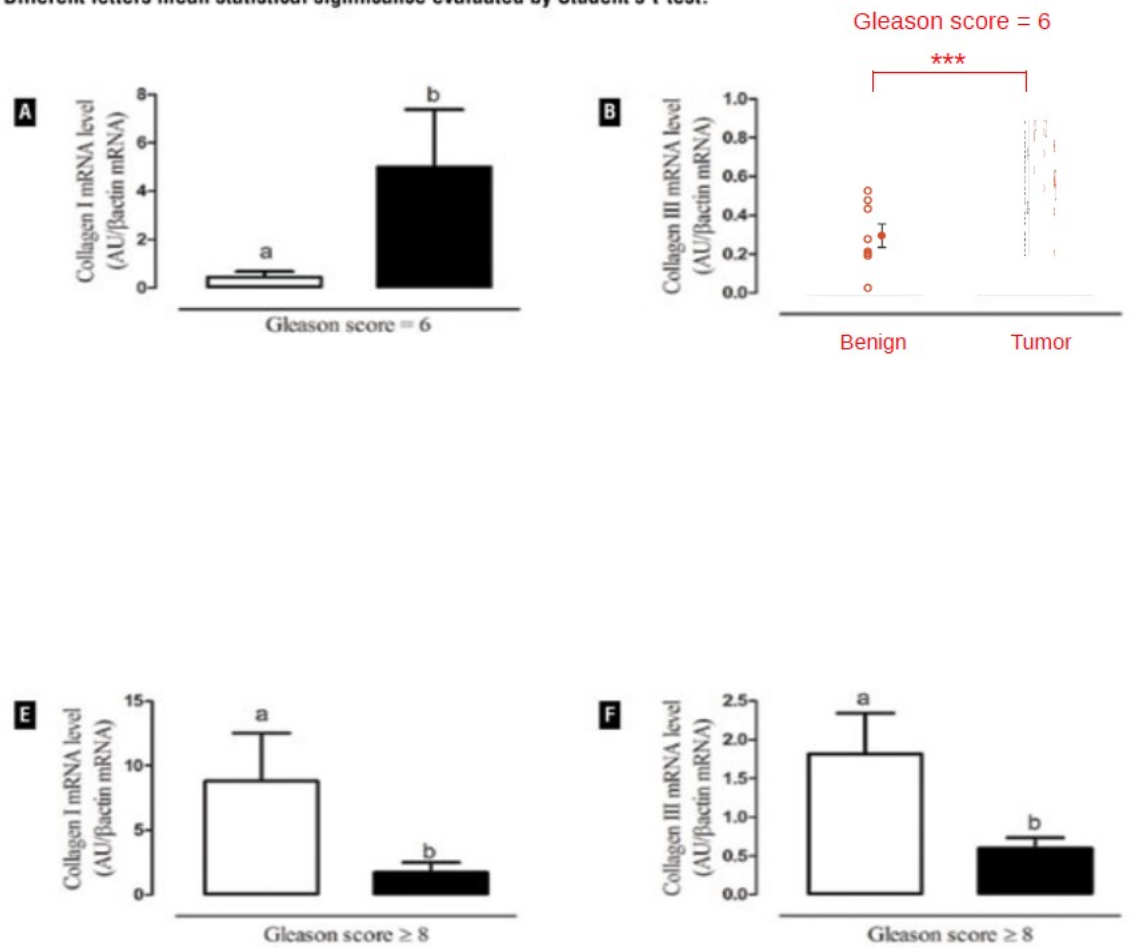
Nemožnosť priameho porovnania medzi grafmi

Antonio H. Duarte; Sicilia Colli; Jorge L. Alves-Pereira; Max P. Martins; Francisco J. B. Sampaio; Cristiane F. Ramos. Collagen I and III and metalloproteinase gene and protein expression in prostate cancer in relation to Gleason score. Int. braz j urol. vol.38 no.3 Rio de Janeiro May/June 2012

ETOX

5. Redundantné grafy C a D

Figure 2 - Gene expression of collagen I and III in the benign (white bar) and tumor (black bar) areas of prostate of Gleason score=6 (A,B), Gleason score=7 (C,D) and Gleason score≥8 (E,F). β actin was used as an internal control. Data are represented as means ± SEM. Sample numbers were 13 for Gleason score=6; 10 for Gleason score=7 and 10 for Gleason score≥8. Different letters mean statistical significance evaluated by Student's-t-test.



Nevýznamný výsledok netreba zobrazovať, prípadne dať do supplemental

Antonio H. Duarte; Sicilia Colli; Jorge L. Alves-Pereira; Max P. Martins; Francisco J. B. Sampaio; Cristiane F. Ramos. Collagen I and III and metalloproteinase gene and protein expression in prostate cancer in relation to Gleason score. Int. braz j urol. vol.38 no.3 Rio de Janeiro May/June 2012

ETOX

Další čtení

- https://eagereyes.org
- http://www.biostat.wisc.edu/~kbroman/topten_worstgraphs/mykland_disc.html
- http://www.exercisebiology.com/index.php/site/articles/how_graphs_can_fool_you/
- https://www.eea.europa.eu/data-and-maps/daviz/learn-more/chart-dos-and-donts#toc-2
- http://www.radford.edu/jkell/statsgraphs.pdf
- http://www.stat.auckland.ac.nz/~ihaka/120/Lectures/lecture03.pdf
- http://people.stat.sfu.ca/~cschwarz/Stat650/Notes/PDF/ChapterBadgraphs.pdf
- http://www.datavis.ca/gallery/index.php
- http://www.doc.govt.nz/documents/science-and-technical/docts32.pdf
- http://www.edwardtufte.com/tufte/

- Geoff Cumming, Fiona Fidler, David L. Vaux (2007). Error bars in experimental biology. JCB Hom. 177 (1): 7e