

# Kreativní práce s daty

---

Úvod do ICT, 24. 10. 2022

# Co je nutné umět?

- Za první – data přečíst
- Za druhé – data zpracovat
- Za třetí – data interpretovat
- Za čtvrté (volitelně) – data vizualizovat

# Datové typy proměnných

- Text (řetězec znaků)
- Číslo (celé, reálné, ...)
- Datum (a čas)
- Pravda × nepravda (true × false)
- Kombinovaný typ (např. pole)

# Zpracování dat

- Čištění
- Transformace
- Agregace (seskupování)
- Statistické operace

# Aplikace na práci s daty

- Tabulkový kalkulátor
  - MS Excel, LibreOffice Calc, ...
- Statistický analyzátor
  - SPSS, Statistica, Matlab, PSPP, R, Octave, ...
- Relační databáze
  - Vybrané funkce jazyka SQL

+ spousta  
všelijakých  
utilit

# Regulární výrazy

Vzory (masky) textových řetězců pro:

- Kontrolu vstupních dat
- Vyhledávání v rozsáhlých datech
  - **Co** přesně tam je i **kde** přesně to tam je (pozice)
- Hromadné změny / editace

# Regulární výrazy II

**^** – začátek řádku

**\$** – konec řádku

**.** – libovolný znak

**[ ]** – výčet znaků

**[ ^ ]** – negativní výčet

Opakování znaků:

**?** – žádné nebo jedno

**+** – jedno nebo více

**\*** – libovolné

**{min,max}** – v zadaném rozsahu

# Regulární výrazy III

`^abc` – text **abc** na začátku řádku

`[abc]` – právě jedno z písmen **a**, **b**, nebo **c**

`[^0-9]*` – sekvence znaků s výjimkou číslic

`https?://.+\.com` – webová adresa končící **.com**

`[0-9a-zA-Z]+@[0-9a-zA-Z]+\.[0-9a-zA-Z]{2,3}`



# Regulární výrazy II

- [RE ve Wikipedii](#)
- [regularnivyrazy.info](#)
- [Trénink RE v češtině](#)
- [Online hrátky s RE – regex101.com](#)
- [Tester RE v PHP, Perlu, JavaScriptu, AWK, ...](#)

# Datové typy ve statistice

- Nominální (kategorické)
  - např. pohlaví, studovaná škola
- Ordinální (pořadové)
  - např. ročník studia, dosažené vzdělání
- Intervalové (nejasný počátek)
  - Např. datum, hodnota IQ, stupně Celsia
- Poměrové (jasný počátek)
  - např. věk, rychlost, stupně Kelvina

# Statistické operace s daty

- Popisná statistika (průměry, ...)
- Testování hypotéz
  - test dobré shody, t-test, analýza rozptylu, ...
- Korelace a regresní analýza
- Explorační analýza

[datalab.net](http://datalab.net)

# Kontingenční tabulka

- *Crosstab* v angličtině
- Jednoduchý způsob pro srovnání dvou a více skupin hodnot
- Lze celkem snadno odhalovat odchylky oproti „normálu“

# Rozdíly mezi formami studia

		Forma studia		Celkem
		prezenční	kombinované	prezenční
částečně nebo úplně chyběla praktická cvičení / semináře	prázdné	77,0 %	91,3 %	78,7 %
	souhlas	23,0 %	8,7 %	21,3 %
Celkem		100,0 %	100,0 %	100,0 %

		Forma studia		Celkem
		prezenční	kombinované	prezenční
chybělo vysvětlování učitele k učivu	prázdné	83,9 %	71,7 %	82,5 %
	souhlas	16,1 %	28,3 %	17,5 %
Celkem		100,0 %	100,0 %	100,0 %

# Interpretace dat

- Co znamenají, co z nich vyplývá?
- Interpretační chyby
  - Zdánlivá souvislost, dvojitá příčina, chybějící prvek
- Subjektivní zkreslení
  - První dojem, přehnaná sebedůvěra, přetížení daty
  - Scházející informace, nestejná spolehlivost, zakořeněnost

# Jakou vizualizací zvolit?

- Kdo, co – zobrazíme dotyčný objekt
- Kolik – odpovíme prostým grafem
- Kde – vyjádříme mapou
- Kdy – zachytíme polohu v čase
- Jak – pomocí vývojového diagramu
- Proč – pravá kreativní výzva :-)

# Graf má ...

- Zobrazit údaje
- Vměstnat mnoho čísel do malého prostoru
- Učinit rozsáhlé soubory údajů srozumitelnými
- Podnítit oko ke srovnávání
- Vyvarovat se zkreslení sdělení



# Ukázky

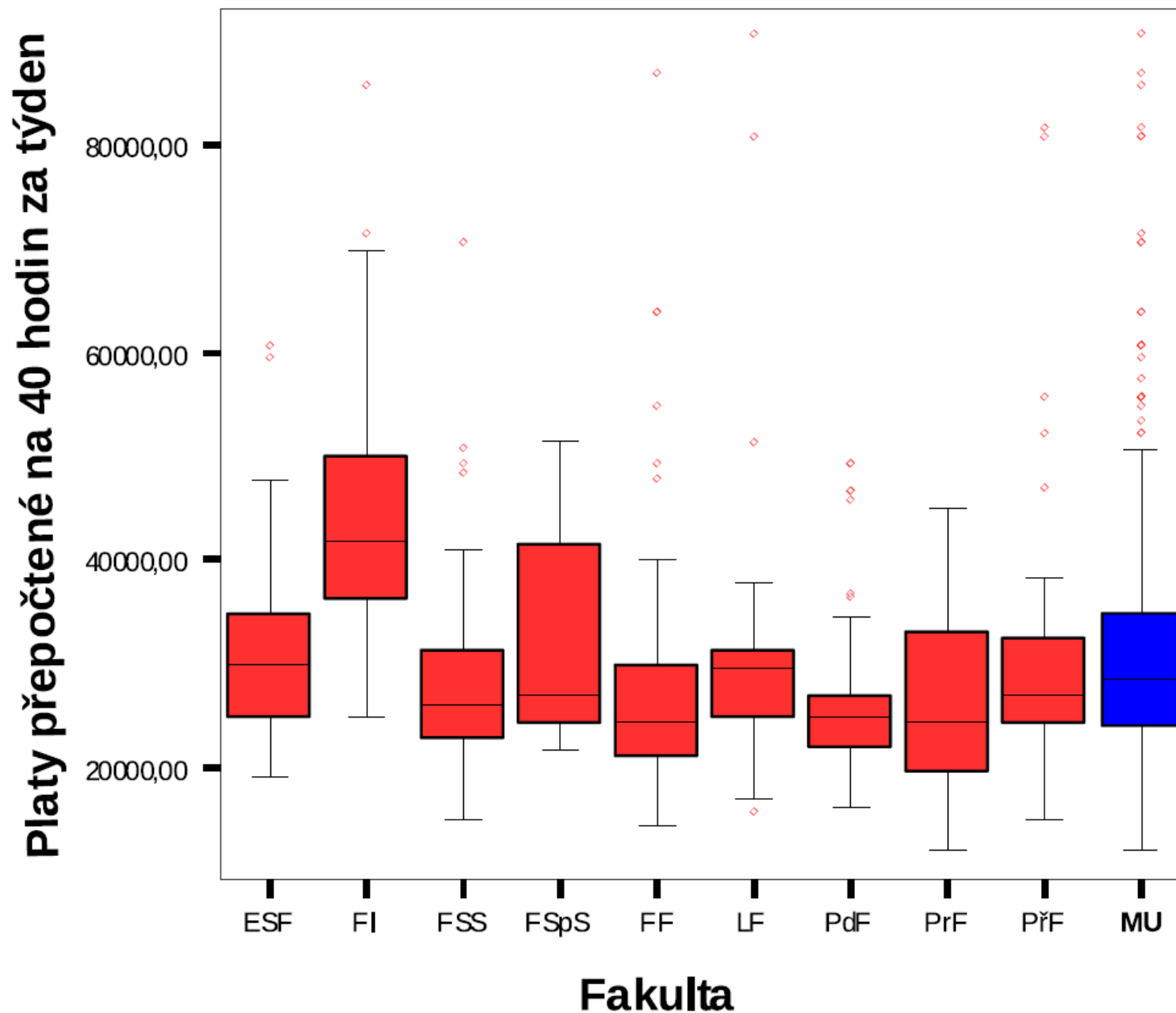
- Základní grafy v MS Forms
- [Mapy](#) Marka Lutonského
- [Rozptylogram](#) z knihy Faktomluva
- Obličejový (Chernoff faces)
- Diagram Ganttův; paprskový

# Boxplot

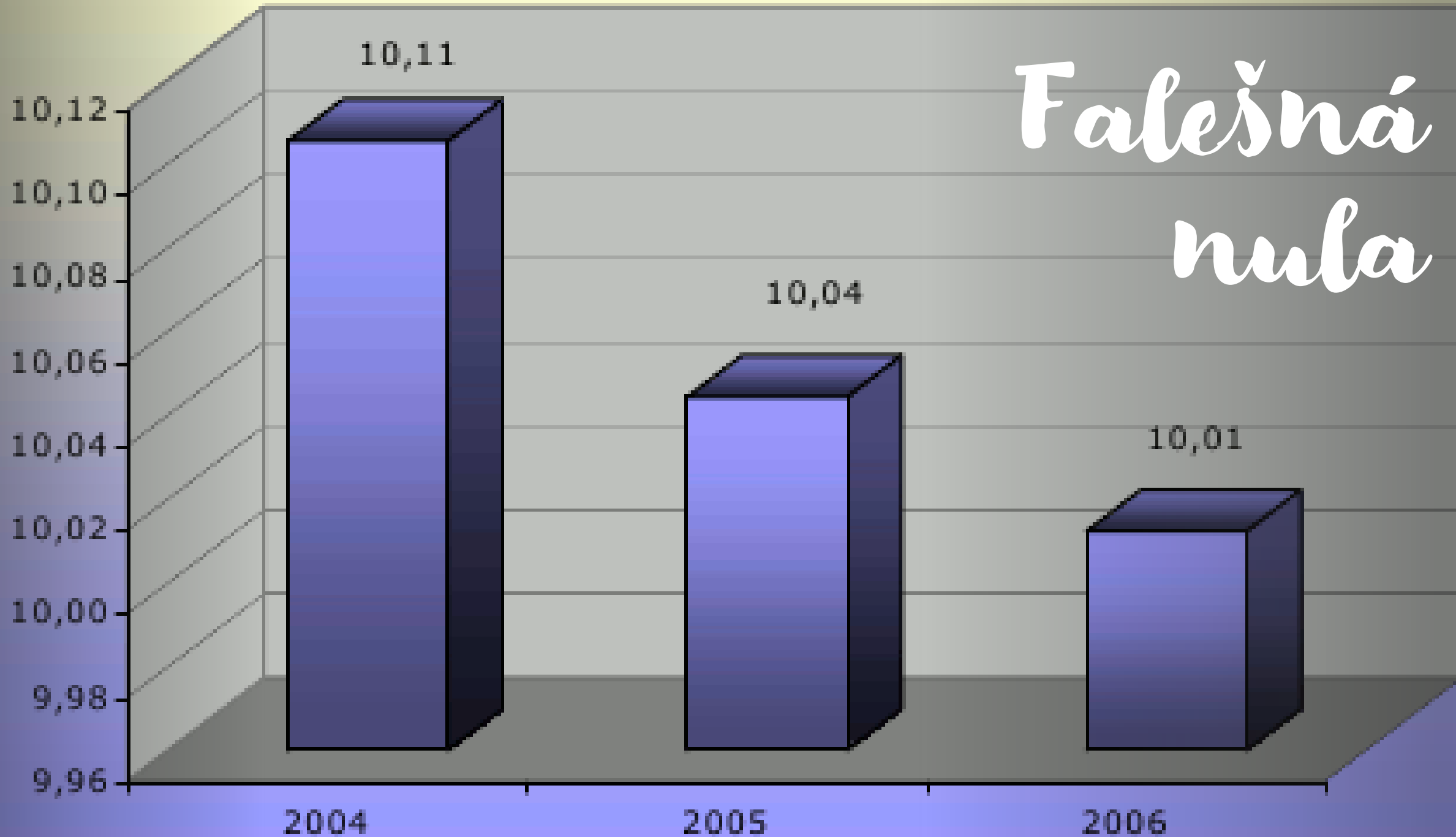
Graficky znázorněné kvartily hodnoty z celého rozsahu

„Box“ obsahuje polovinu všech případů

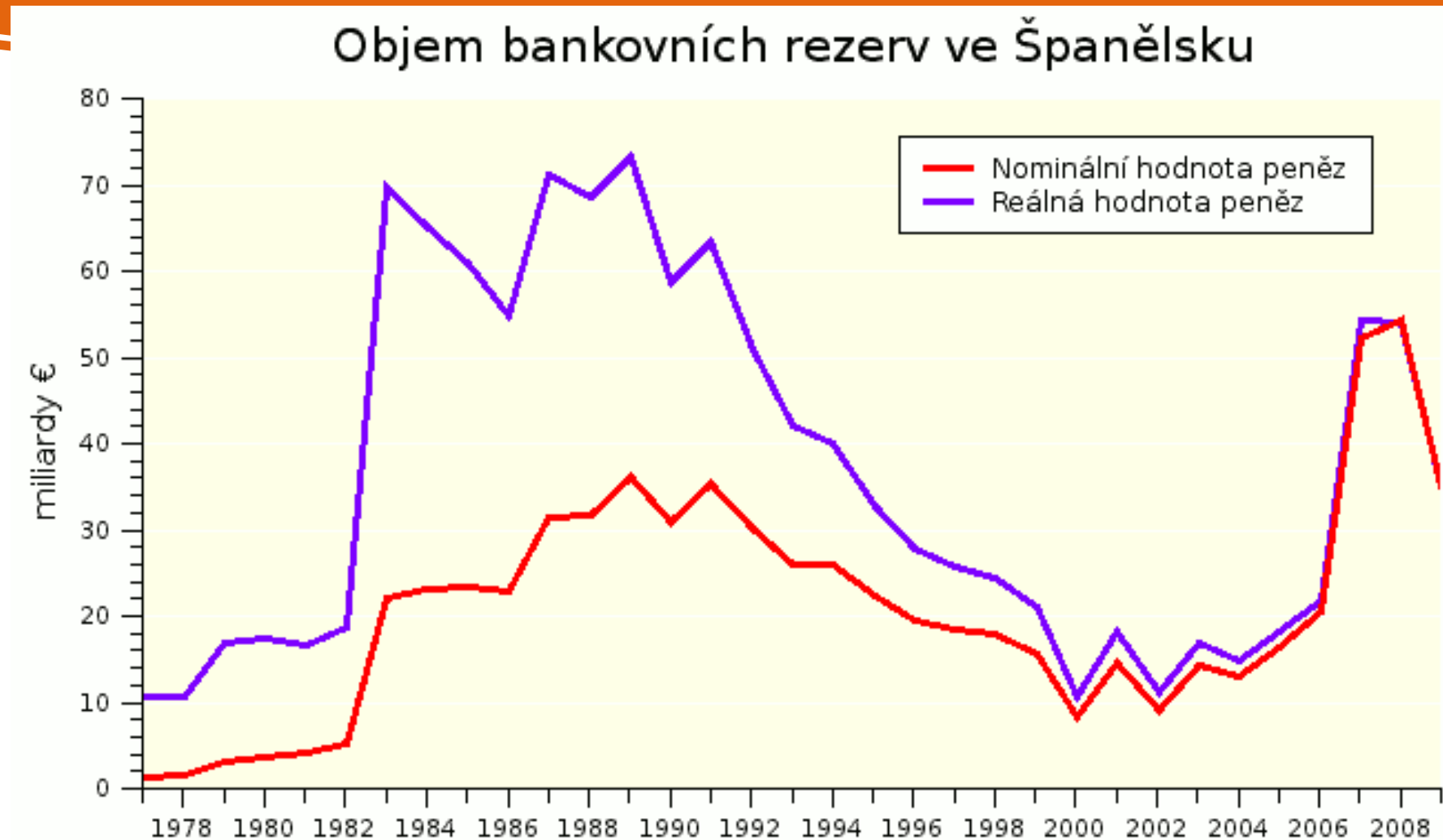
Puntíky = ojedinělé extrémní případy



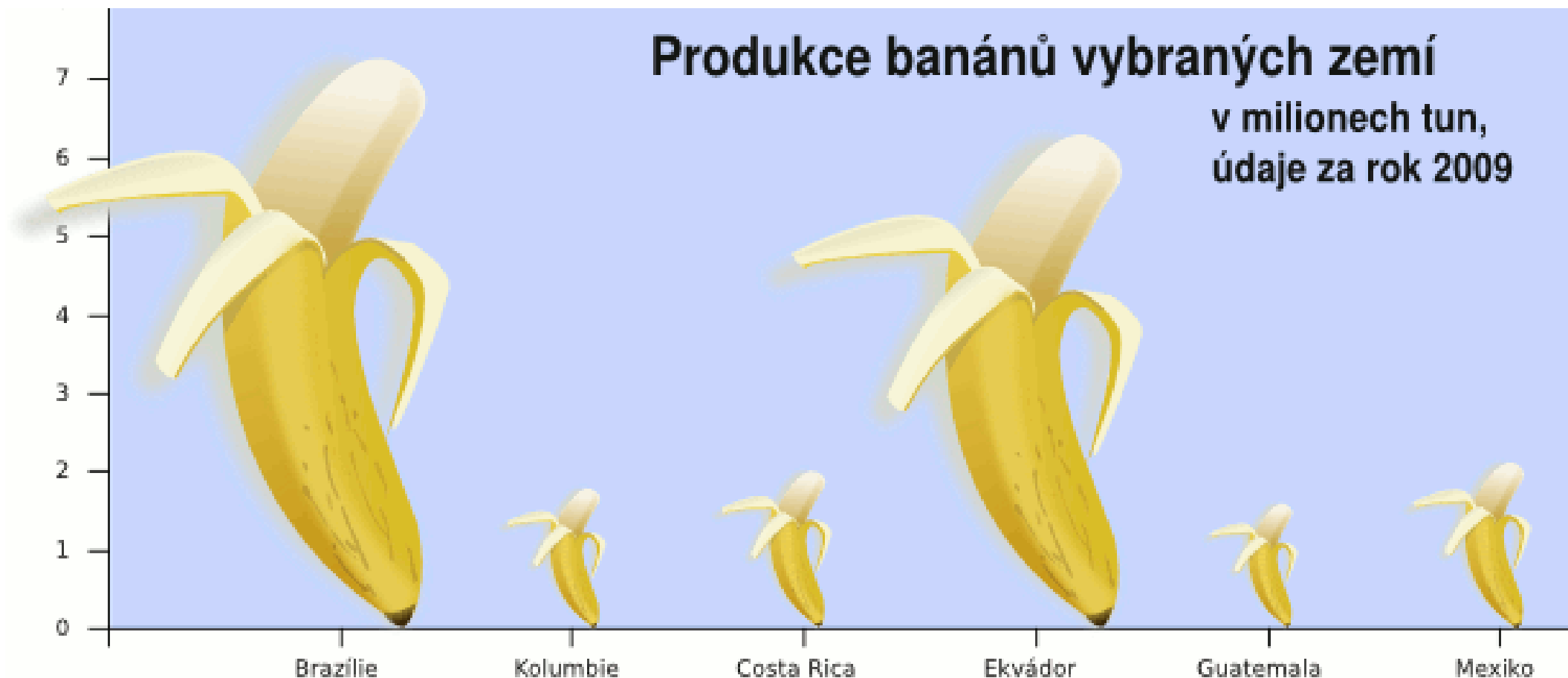
# Falešná nula



# Ne normovaná data

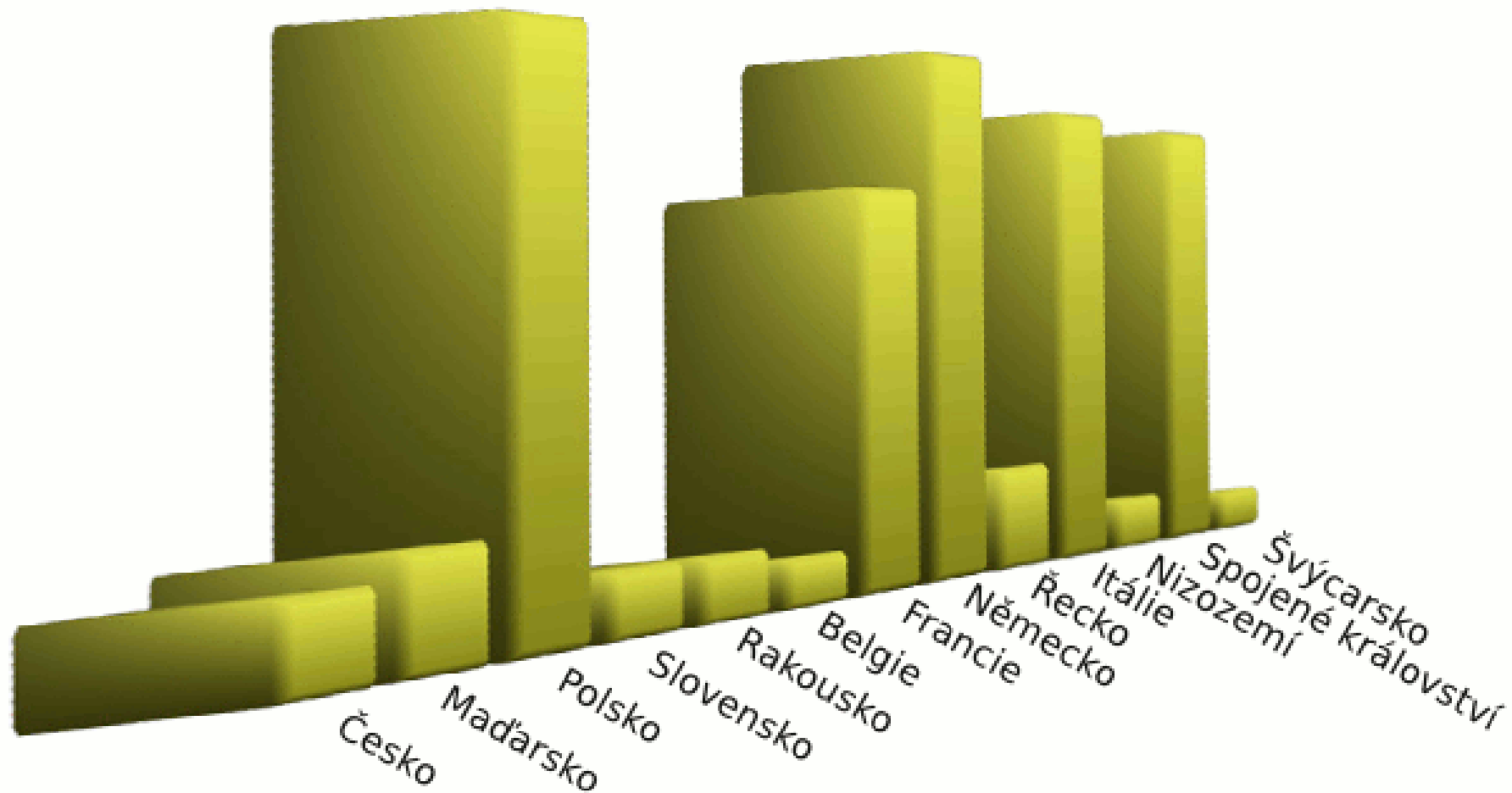


# Plošné znázornění hodnot lineárních dat



# Zkreslení vinou pseudo3D

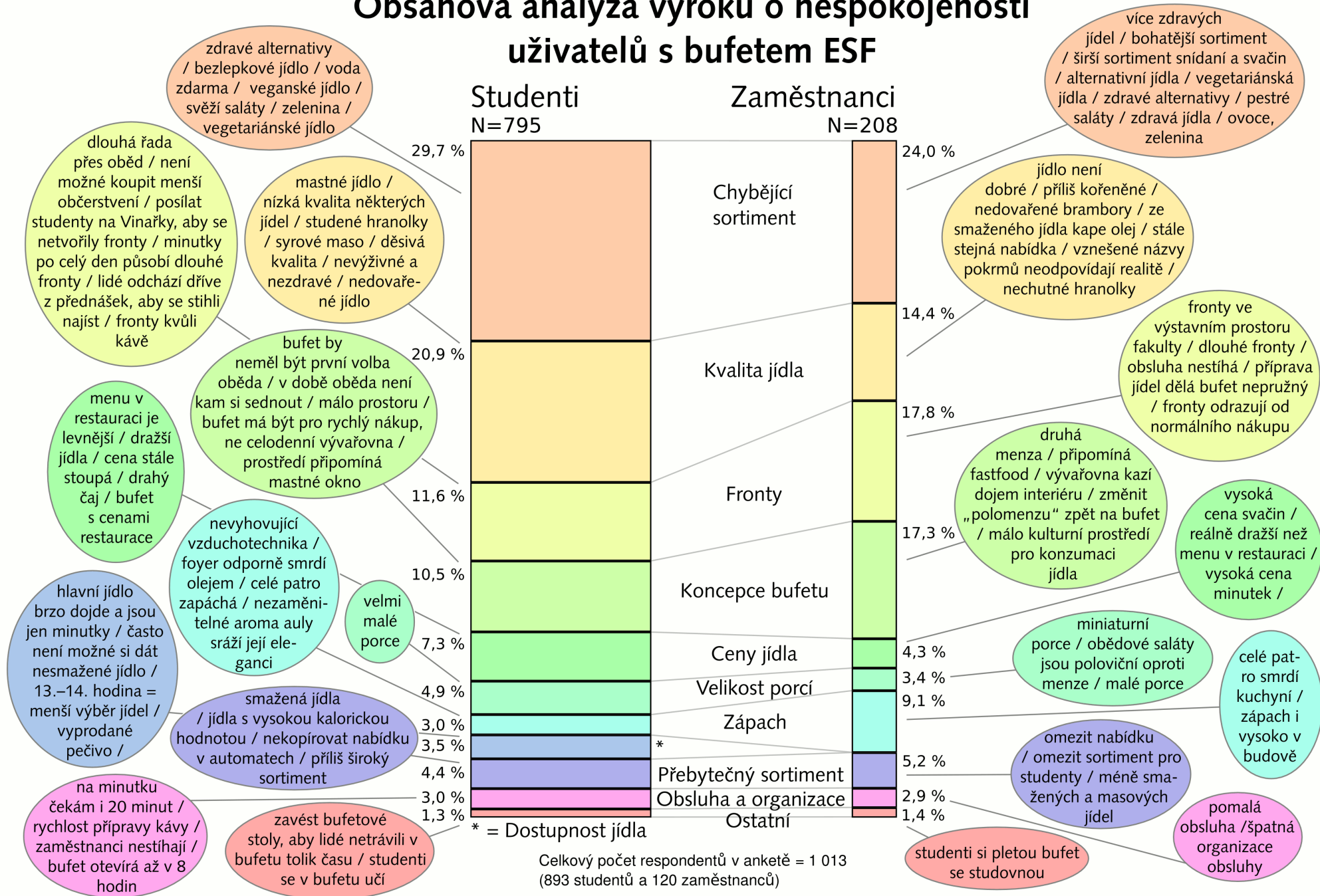
Počty univerzitních studentů ve vybraných zemích



# Infografika

- Spojení grafu a umění
- Oblíbené u médií (novinové články)
- Obvykle kombinace grafu a kontextu
  - [www.easel.ly](http://www.easel.ly)    [www.infogram.com](http://www.infogram.com)
  - [www.piktochart.com](http://www.piktochart.com)

# Obsahová analýza výroků o nespokojenosti uživatelů s bufetem ESF



zdravé alternativy / bezlepkové jídlo / voda zdarma / veganské jídlo / svěží saláty / zelenina / vegetariánské jídlo

více zdravých jídel / bohatější sortiment / širší sortiment snídaní a svačin / alternativní jídla / vegetariánská jídla / zdravé alternativy / pestré saláty / zdravá jídla / ovoce, zelenina

dlouhá řada přes oběd / není možné koupit menší občerstvení / posílat studenty na Vinařky, aby se netvořily fronty / minutky po celý den působí dlouhé fronty / lidé odchází dříve z přednášek, aby se stihli najíst / fronty kvůli kávě

mastné jídlo / nízká kvalita některých jídel / studené hranolky / syrové maso / děsivá kvalita / nevyživné a nezdravé / nedovařené jídlo

jídlo není dobré / příliš kořeněné / nedovařené brambory / ze smaženého jídla kape olej / stále stejná nabídka / vznešené názvy pokrmů neodpovídají realitě / nechutné hranolky

bufet by neměl být první volba oběda / v době oběda není kam si sednout / málo prostoru / bufet má být pro rychlý nákup, ne celodenní vývařovna / prostředí připomíná mastné okno

fronty ve výstavním prostoru fakulty / dlouhé fronty / obsluha nestíhá / příprava jídel dělá bufet nepružný / fronty odrazují od normálního nákupu

menu v restauraci je levnější / dražší jídla / cena stále stoupá / drahý čaj / bufet s cenami restaurace

druhá menza / připomíná fastfood / vývařovna kazí dojem interiéru / změnit „polomenzu“ zpět na bufet / málo kulturní prostředí pro konzumaci jídla

nevyhovující vzduchotechnika / foyer odporně smrdí olejem / celé patro zapáchá / nezaměnitelné aroma auly sráží její eleganci

vysoká cena svačin / reálně dražší než menu v restauraci / vysoká cena minutek /

hlavní jídlo brzo dojde a jsou jen minutky / často není možné si dát nesmažené jídlo / 13.-14. hodina = menší výběr jídel / vyprodané pečivo /

velmi malé porce

miniaturní porce / obědové saláty jsou poloviční oproti menze / malé porce

smažená jídla / jídla s vysokou kalorickou hodnotou / nekopírovat nabídku v automatech / příliš široký sortiment

celé patro smrdí kuchyní / zápach i vysoko v budově

na minutku čekám i 20 minut / rychlost přípravy kávy / zaměstnanci nestíhají / bufet otevírá až v 8 hodin

zavést bufetové stoly, aby lidé netrávili v bufetu tolik času / studenti se v bufetu učí

omezit nabídku / omezit sortiment pro studenty / méně smažených a masových jídel

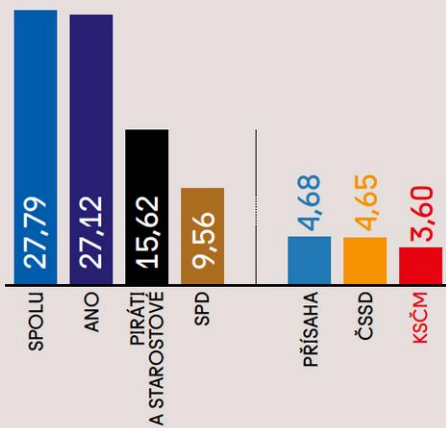
pomalá obsluha / špatná organizace obsluhy

studenti si pletou bufet se studovnou



## MAPA SE PROMĚNILA

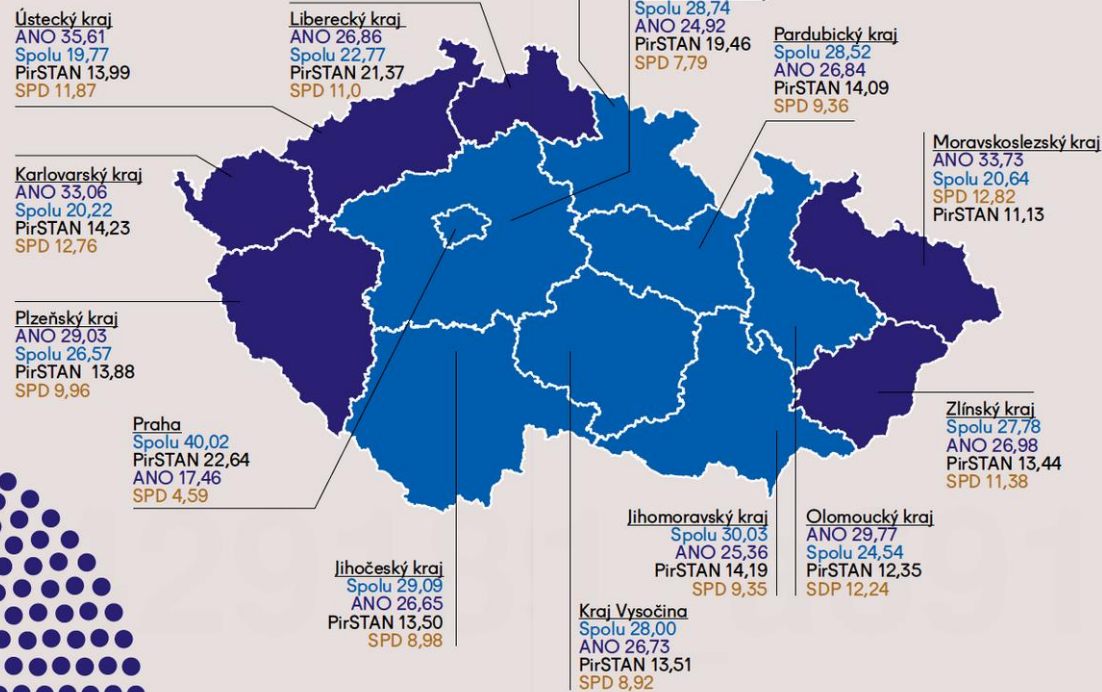
Výsledky v jednotlivých krajích a městech  
(v procentech)



### ROZDĚLENÍ MANDÁTŮ:

- ANO 72 mandátů,
- Spolu 71 mandátů  
(ODS 34, KDU-ČSL 23, TOP 09 14)
- PirSTAN 37 mandátů  
(STAN 33, Piráti 4)
- SPD 20 mandátů

# 2021



### PRAHA



### BRNO



### OSTRAVA



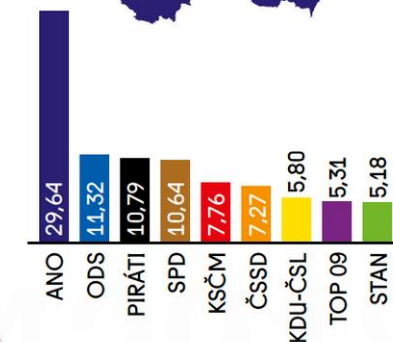
### PLZEŇ



### OLOMOUC



# 2017



## ZASE O TROCHU VÍČ

Počet žen a mužů ve sněmovně letos a v minulém volebním období

150 50

157 44

volební účast **65,43%**

## A NA KROUŽKY VYHRÁL... politici s největším počtem preferenčních hlasů



VÍT RAKUŠAN  
(PirSTAN)

59 792



MARKÉTA PEKAROVÁ  
ADAMOVÁ  
(Spolu)

49 014



PETR FIALA  
(Spolu)

38 515



ANDREJ BABIŠ  
(ANO)

38 271



JANA ČERNOCHOVÁ  
(Spolu)

35 583

# Domácí úlohy

1) Procvičte si tvorbu regulárních výrazů

2) Vytvořte infografiku

- Základ dat:

<https://www.czso.cz/csu/czso/databaze-demografickych-udaju-za-vybrana-mesta-cr>