

# Základy spracovania a vizualizácie dát

26. februára 2024

- ▶ rozcvičky na konci hodiny:
  - 11.3.2024
  - 8.4.2024
  - 29.4.2024
  - 13.5.2024 - opravná / náhradná

# dáta

- ▶ rôzne delenia, napríklad:
  - ▶ kvalitatívne (číselné)
    - diskrétné vs spojité
    - intervalové / rozdielové (rok, °C) vs podielové / pomerové (vek,  $K$ )
  - ▶ kvantitatívne (kategorióálne)
    - ordinálne
    - nominálne

# načítanie dát

- ▶ cesta DATA → GetData
- ▶ rôzne možnosti: .txt, .csv, ...
- ▶ pozor na formát
  - desatinné čísla
  - čísla, text, dátum, ...
- ▶ základná práca s tabuľkou:
  - filtrovanie
  - usporadúvanie
  - rozširovanie riadkov/stĺpcov
  - úprava vizuálu
  - slicers

# cvičenie

- ▶ načítajte dáta *Infant\_mortality\_data\_1965*
- ▶ zmeňte meno tabuľky, vizuál
- ▶ usporiadajte tabuľku, aby boli štáty zoradené abecedne zostupne
- ▶ vytvorte nový stĺpec v tabuľke s názvom *HS in %*, ktorý bude zobrazovať hodnoty *HS* v percentách

# logické operátory

- ▶ fungujú intuitívne
- ▶ *AND()*, *IF()*, *OR()*, *XOR()*, *IFS()*, ...
- ▶ *TRUE*, *FALSE*, *<*, *>*, *<=*, *>=*, *=*, *NA()*

# základné funkcie

- ▶ *ROUND()*, *SQRT()*, *POWER()*, *LN()*, *LOG10()*, *LOG()*, *FLOORMATH()*, ...
- ▶ *VLOOKUP()*, *HLOOKUP()*, ...
- ▶ cesta FORMULAS → MoreFunctions → Statistical

## cvičenie

- ▶ vytvorte v tabuľke nový stĺpec s názvom *pophys* < 10000, ktorý bude obsahovať
  - TRUE, ak je *pophys* < 10000
  - FALSE, ak je *pophys* >= 10000
- ▶ vytvorte v tabuľke nový stĺpec s názvom *Calorie group*, ktorý bude obsahovať
  - 1 pre *Calorie* ∈ [1000, 2000)
  - 2 pre *Calorie* ∈ [2000, 3000)
  - 3 pre *Calorie* ∈ [3000, 4000)
- ▶ využite slider ako filter
- ▶ vytvorte v tabuľke nový stĺpec, ktorý bude avizovať, že nebolo v danej krajine dostatočne veľa zdravotníkov, teda *pophys* aj *popnurs* obsahuje hodnotu väčšiu ako 1000
- ▶ pomocou funkcie *VLOOKUP* nájdite hodnotu novorodeneckej úmrtnosti pre Československo



# samostatné cvičenie

- ▶ načítajte dáta *HR-Employee-Attrition*
- ▶ zmeňte meno tabuľky, vizuál
- ▶ využite sliders na vytriedenie zamestnancov, pre ktorých platí:
  - *Attrition* = "Yes"
  - *Department* = "Human Resources"
  - *Gender* = "Male"
  - *EducationField* = "Life Sciences "
- ▶ vytvorte v tabuľke nový stĺpec s názvom *Att*, ktorý bude prekóduje hodnoty *Attrition*
  - 0 = "No"
  - 1 = "Yes"
- ▶ pomocou funkcie *VLOOKUP* nájdite pre zamestnanca s číslom *EmployeeNumber* = 895
  - *EnvironmentSatisfaction*
  - *Department*