

Analýza a vizualizácia dát

11. marca 2024

- ▶ možné termíny:
 - utorok 14:50
 - štvrtok 16:30

grafy

- ▶ rýchly prenos informácie
- ▶ pozor aj na vizuálnu stránku

bar / column plot

- ▶ najčastejšie pre kategoriálne dáta
- ▶ grafy početností (percent)
- ▶ spolupracujú s pivotkami
- ▶ *Department (HRdata)*
- ▶ úprava textu

X Y scatter plot

- ▶ dve číselné premenné
- ▶ sledovanie závislosti
- ▶ *Monthly Income vs Total Working Years*
- ▶ zmeniť popisy osí
- ▶ trendline

line chart

- ▶ časové rady - vykreslenie trendu v čase
- ▶ spurious correlations

pie chart

- ▶ často pre kategoriálne dáta s malým počtom kategórií
- ▶ pri kategoriálnych dátach niekedy treba spraviť najskôr pivotku
- ▶ 3D vie zavádzať
- ▶ *Marital Status*

histogram

- ▶ histogram = graf, ktorý vykresľuje intervalovú frekvenčnú tabuľku
- ▶ column chart je podobný
- ▶ jeden z najpoužívanejších grafov (vizuálne počiatkové zhodnotenie dát)
- ▶ porovnanie s teoretickými rozdeleniami
- ▶ *Hourly Rate*

boxplot

- ▶ zobrazuje veľa číselných charakteristík v jednom grafe
- ▶ minimum, maximum, kvartily, priemer, outlierov ($q_{0,25} - 1,5IQR; q_{0,75} + 1,5IQR$), niekedy aj rozptyl
- ▶ pozor, *QUARTILE.EXC()*
- ▶ porovnanie mužov a žien *Hourly Rate*

iné

- ▶ area, surface, ...
- ▶ špeciálne typy dát