Úkoly

Jan Caha

28 března 2018

## Úkol ze cvičení 1

Vyberte si buď z dat prezentovaných na prvním cvičení, nebo z [webu](https://vincentarelbundock.github.io/Rdatasets/datasets.html) některý z datasetů připravených pro R. Nastudujte si z helpu strukturu dat. Z dané datové sady udělejte libovolnou selekci zájmových dat pomocí funkcí v balíku **dplyr** - kompletní dokumentace je dostupná [zde](https://dplyr.tidyverse.org/). Selekce je pouze pro zopakování práce s balíkem dplyr, záleží tak zcela na vás, stejně jako výběr datové sady. Z výsledných dat vytvořte dva grafy, u nichž se pokuste je co nejvíce “vyladit”, aby byl výsledek co “nejprezentovatelnější”, pomocí balíku **ggplot2** - dokumentace [zde](http://ggplot2.tidyverse.org/). Grafy budou vaším výstupem z tohoto úkolu, nicméně uložte si i kompletní R skript s načtením balíků, selekcí dat a následně kódem pro tvorbu vizualizací (viz ukázky na [GitHubu](https://github.com/JanCaha/Vyuka-R-VSB) z prvního cvičení).

### Zdroje

K oběma balíkům doporučuji tzv. cheatsheaty - [dplyr](https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf) a [ggplot2](https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf).

U balíku **ggplot2** zkuste prozkoumat funkce **geom\_**, ať máte přehled, jaké druhy grafů lze vytvářet. Poměrně pěkný přehled a ukázky jsou např. [tady](http://r4stats.com/examples/graphics-ggplot2/) či [tady](http://r-statistics.co/ggplot2-Tutorial-With-R.html).

Případně je užitečné mít po ruce i knihu [R for Data Science](http://r4ds.had.co.nz/workflow-basics.html), nicméně pokročilejší položky z této knihy na cvičení budeme teprve probírat.