

# Práce s daty – srovnání softwarů

- Excel /open office
- R
- Python

# Excel

- Vhodný k otvírání dat a zorientování se
- Zběžná kontrola (co je v datech, prázdné řádky, formáty, ...)
- Jednoduché operace
  - Řazení
  - Filtrování
  - nahrazování
  - Souhrny
  - Výpočty
- „Grafická úprava“ výstupů
  - Tabulky
  - Grafy
- Rozšiřující makra

# nevýhody

- Proprietární software
- Placená licence
- Omezené množství řádků (cca 1 000 000)
- Při větším počtu případů velmi pomalý
- **OBTÍŽNÝ ZÁZNAM POSTUPU!**
- Relativně slabá komunita

# R

- Open source, nekomerční licence zdarma
- Vhodný zejména ve dvou situacích
  - Opakované provádění podobných postupů
  - Zachycení kompletního postupu
- Vhodný zejména k manipulaci s datovým souborem
  - Spojování datasetů
  - Filtrování
  - Souhrny
  - Rekódování proměnných
- Téměř neomezené možnosti - „packages“
- Analýza
- Silná komunita (stackoverflow, R bloggers,...)

# nevýhody

- Ovládání textem („programování“)
  - Ne úplně intuitivní jazyk
- Bez opakovaného používání rychlá ztráta schopností
- Nepracuje s ním každý
- Kompatibilita různých packages s různými verzemi
- Omezené možnosti nahlížení do dat
- Snadné vyrobiť/přehlédnout chybu
- Rychlost složitějších operací (+ chybí indikátor „hotovosti“)
- Výstupy (možnost integrace s LaTeX)

# Python

- Standardní programovací jazyk
  - Pro osoby se zkušeností s programováním snadnější na naučení se
- Podobné možnosti jako R co se týká práce s daty
  - Širší možnosti v „informatické“ práci – např. aplikace
- Silnější pro analýzu textu
- Machine learning
- AI
- Big Data

# SPSS, SAS, STATA

- Nástroje pro statistickou analýzu
- Pro správu dat nejde o vhodná řešení