

Uvod u R/R studio

Opis kursa:

Nova era informacija stvara nove prilike na tržištu rada. Pojavljuju se nova radna mjesta kao što su Data Mining, Data Science, Manager Savvy, Data citizen, dok „stara“ radna mjesta kao što su „Data Analyst“ zahtjevaju poznavanje novih alata.

R je besplatan, ima preko 2 miliona korisnika (do 2014.g.), broj korisnika je u stalnom porast, preko 7.800 paketa koje su nadogradnja na basic R (do 2016.g.). R prvenstveno služi za statističke analize i vizualizaciju podataka. Veliki broj paketa omogućava njegovu primjenu u svakoj industriji. Pored statističara koriste ga data scientists, naučnici, analitičari.

Ako ste istraživač, ili radite analize podataka, onda je R pravi izbor. Kako R nije point-and-click software, spašavanje kodova omogućava reprodukciju analize bilo kad i bilo gdje. Na kursu će se raditi u R Studiju, koji je prilagođen korisnicima („user friendly“). Pored osnovne manipulacije bazama podataka i osnovnih analiza, polaznici će kraju kursa moći da urade analizu baze podataka u R markdownu.

Kurs je namjenjen osobama bez ikakvog prethodnog znanja u R-u. Očekivanja su da će polaznici biti raznih profila i profesija: studenti, ekonomisti, biolozi, farmaceuti, statističari, IT stručnjaci, osobe koje se bave naučnoistraživačkim radom, osobe koje se bave medicinskim istraživanjima itd.

Neophodni materijali/oprema:

- Na kurs je potrebno ponijeti vlastiti laptop

Plan i program rada:

| Broj sati | Tema | Opis teme | Paketi |
|-----------|---------------------------------------|--|--|
| 1 | Uvod u R i R studio | upoznavanje sa R i R studiom, osnovne matematičke i statističke operacije, importovanje file-ova (.xls, .csv...), kreiranje vektora, baza podataka, indeksiranje | Basic R paket |
| 3 | Manipulacija bazama i "ciscenje baza" | Filtriranje i selektovanje varijabli, dodavanje i brisanje varijabli, dodavanje i brisanje obzervacija, spajanje baza podataka, formatiranje varijabli (numericke, kategoricke i datumi), sortiranje baze, grupisanje, Ujednačavanje velikih ili malih slova, ciscenje baze. Upotreba opratora %in% | Basic, dplyr, lubridate, tidyr, stringr, exess, reshape2 |

| | | | |
|---|---|---|---|
| 2 | Analiza podataka | Upoznavanje sa bazom podataka, Sumarni pregledi, deklinativna statistika, osnovni grafikoni (density, box-plot, bargraf, linijski, rukovanje sa nedostajucim podacima) | Basic, stargazer |
| 2 | Deskriptivna statistika + statistički testovi | sum, sd, mean, median, t test, chi-square test, | |
| 3 | Linearna regresija | Uvod u jednostavnu linearnu tzv.univariate regresiju i uvod u linearnu multivariate regresiju: pretpostavke i dijagnostika | Koristeni paket basic + stargazer za prikazivanje outputa |
| 4 | Vizualizacija podataka | Graficki prikaz najpopularnijim R paketom: linijski, barplot, boxplot., prikaz regresione linije, barplot sa error bar, vizualizacija podataka na mapi, prikaz vise od jednog grafa | Paketi: ggplot2, GridExtra, corrplot, ggmap |
| 1 | R Markdown | Upoznavanje sa RMarkdown i rad u njemu koji omogucava ubacivanje kodova u tzv. Chunks i onda uz pomoc paketa knitr da se podaci "ispletu" u finalni dokument. Finalni dokument koji moze biti u html, word ili pdf verziji ujedno prikazuje i kodove i output kodova po zelji | Paket Rmarkdown i knitr |