

DODATNI PAKETI U R-u

STATISTIČKI PRAKTIKUM 2

- ▶ manipulacija podacima
 - ▶ dplyr
 - ▶ tidyr
 - ▶ lubridate
 - ▶ stringr
 - ▶ data.table
 - ▶ janitor
- ▶ vizualizacija
 - ▶ ggplot2
 - ▶ ggvis
 - ▶ rgl
 - ▶ plotly
 - ▶ corrplot
- ▶ za izvještaj o rezultatima
 - ▶ shiny
 - ▶ R Markdown
 - ▶ xtable
- ▶ za prostorne podatke
 - ▶ ggmaps
 - ▶ sp, maptools

- ▶ modeliranje podataka
 - ▶ tidymodels
 - ▶ car
 - ▶ lme4
 - ▶ randomForest
 - ▶ multcomp
 - ▶ glmnet
 - ▶ vcd
 - ▶ survival
 - ▶ caret
- ▶ za vremenske nizove i
financijske podatke
 - ▶ zoo
 - ▶ xts
 - ▶ quantmod
- ▶ purr
- ▶ recipes
- ▶ simstudy

dplyr

- ▶ 2014.
- ▶ manipulacija podacima
- ▶ rad s `data.frame`, ali i vanjskim bazama podataka
- ▶ funkcije organizirane u 4 skupine, ovisno o tome rade li nad
 - ▶ retcima
 - ▶ stupcima
 - ▶ grupama
 - ▶ tablicama
- ▶ najčešće korištene:
 - ▶ `mutate()`
 - ▶ `select()`
 - ▶ `filter()`
 - ▶ `summarise()`
 - ▶ `arrange()`
 - ▶ `group_by()`
- ▶ "pipes" `%>%` za ulančavanje naredbi

ggplot2

- ▶ 2007.
- ▶ funkcije za grafički prikaz podataka (također radi i s okvirima podataka)
- ▶ funkcija `ggplot()` stvara prazan graf na kojeg dodajemo slojeve spojene oznakom `+`
- ▶ geometrijski elementi grafa:
 - ▶ `geom_point()`
 - ▶ `geom_line()`
 - ▶ `geom_boxplot()`
 - ▶ `geom_bar()`
 - ▶ `geom_histogram()`
- ▶ unutar podnaredbe `aes()` povezujemo varijable s vizualnim svojstvima, a ostale vizualne argumente stavljamo izvan te naredbe
- ▶ ostali elementi koje dodajemo na graf: naslov, naziv osi...,
- ▶ sa `theme` uređujemo ostale elemente grafa (veličina naslova, oznake na osima...)

- ▶ za čišćenje imena stupaca `janitor::clean_names`
- ▶ korisna knjiga sa zadacima "R for Data Science", Wickham et al.
- ▶ "šalabahteri"

Zadatak 1

Učitajte podatke naziva `mtcars` o svojstvima i performansama 32 marke automobila.

- (a) Izdvojite samo podatke o potrošnji (`mpg`), broju konjskih snaga (`hp`) i vrsti prijenosa (`am`).
- (b) Ispišite one marke automobila čiji broj cilindara (`cyl`) je jednak 8, a potrošnja nije veća od 16.
- (c) Spremite u novu varijablu podatke o potrošnji, broju cilindara, težini (`wt`), vremenu za 1/4 milje (`qsec`) i vrsti prijenosa za one automobile čiji oblik motora je ravan (`vs` je jednak 1).
- (d) U tu novu matricu dodajte novi stupac u kojem ćete spremiti podatke od potrošnji po jednom cilindru (`mpg/cyl`).
- (e) Izračunajte prosječnu potrošnju automobila ovisno o broju cilindara, a zatim minimalno i maksimalno vrijeme za 1/4 milje (`qsec`) ovisno o broju cilindara i vrsti motora (`cyl` i `vs`).

Zadatak 2

Za iste podatke napravite sljedeće:

- (a) Grafički prikazite odnos između potrošnje i vremena za 1/4 milje. Neka oblik točke sugerira broj cilindara pojedinog automobila, a boja opisuje težinu. Povećajte veličinu točaka na 4.
- (b) Dodajte naziv grafa i osi. Povećajte naslov grafa.
- (c) Nacrtajte histogram podataka o potrošnji. Promijenite boju grafa i zarotirajte oznake na x-osi za 60 stupnjeva.
- (d) Nacrtajte boxplot podataka o potrošnji ovisno o vrsti motora. Svaku grupu obojite drugom bojom.