Domača naloga 4

Alen Kahteran

09. 10. 2020

Pridobivanje podatkov

Preko knjižnice `curl` pridobimo podatke ki jih nato spremenimo v `tibble`, ki se obnaša praktično enako kot `data.frame`, le da je meni osebno lepši prikaz.

```r
library(curl)
df_full <- read.csv(
            curl("https://raw.githubusercontent.com/sledilnik/data/master/csv/stats.csv"))
library(tidyverse)
library(lubridate)
df_full <- tibble(df_full)
```

Obdelava podatkov

Za ustrezen prikaz podatkov potrebujemo naslednje stolpce:

- `cases.confirmed`
- `tests.performed`
- `date`

stolpec `date` je v obliki `character`, zato ga ustrezno pretvorimo v `date`.

Dodamo stolpec z zapisom dneva v tednu (Po, T, Sr, C, P, S, n)

```r
# select only required columns
df_simple <- select(df_full, date, tests.performed, cases.confirmed)
# convert dates from character to datetime and then to date (due to ggplot)
df_simple$date <- parse_date_time(df_simple$date, c("Ymd"))
df_simple$date <- as.Date(df_simple$date)

# add a column day_in_week which converts dates into numbers from 1-7
# 1 is Sunday, 2 is monday, ...
df_simple <- mutate(df_simple, day_in_week = wday(date))
# recode changes numerical values to what we want to have in the plot.
df_simple$day_in_week <- recode(df_simple$day_in_week,
                                `1` = "Nedelja",
                                `2` = "Ponedeljek",
                                `3` = "Torek",
                                `4` = "Sreda",
                                `5` = "Cetrtek",
                                `6` = "Petek",
                                `7` = "Sobota")
# adding a column for correct coloring in ggplot
df_simple <- df_simple %>%
    mutate(relative_conf_perf = cases.confirmed / tests.performed) %>%
```

```
    filter(date > as.Date("2020-03-04", format="%Y-%m-%d")) %>%
    mutate(week = week(date) - min(week(date)))
# converting to factor due to ggplot2 correct ordering
df_simple$day_in_week <- factor(df_simple$day_in_week,
                                    levels=c("Ponedeljek",
                                             "Torek",
                                             "Sreda",
                                             "Cetrtek",
                                             "Petek",
                                             "Sobota",
                                             "Nedelja"),
                                    labels=c("Ponedeljek",
                                             "Torek",
                                             "Sreda",
                                             "Cetrtek",
                                             "Petek",
                                             "Sobota",
                                             "Nedelja")
                              )
```
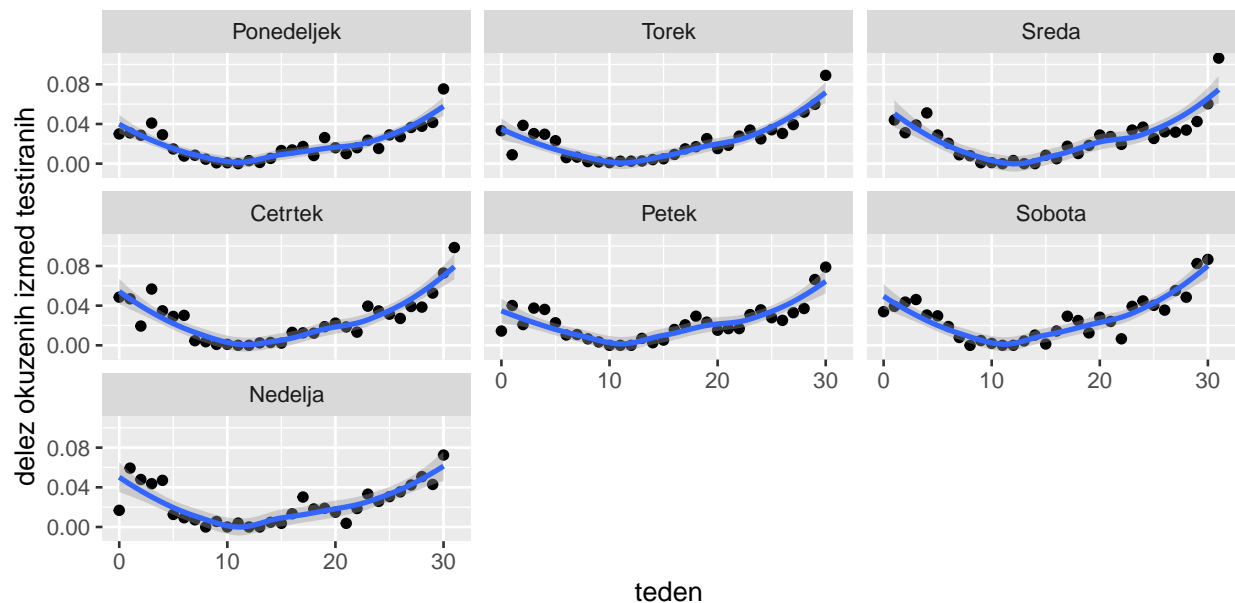
```
# selecting week as x variable and relative_conf_perf as y
fw <- ggplot(df_simple, aes(x=week, y=relative_conf_perf)) +
  # selectin points for plotting
  geom_point() +
  # adding smoothing line
  # using loess as a smoothing line
  geom_smooth()
# creating facet_wrap by day_in_week and adding labels
fw + facet_wrap(vars(day_in_week)) +
  ylab("delez okuzenih izmed testiranih") +
  xlab("teden")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```r
# selecting week as x variable
# relative_conf_perf as y variable
# day_in_week as color
g <- ggplot(df_simple, aes(x=week, y=relative_conf_perf, color=day_in_week)) +
  # selecting points for plotting
  geom_point() +
  # adding smoothing line with alpha 0.1 as you can't see a lot otherwise
  # using loess as a smoothing line
  geom_smooth(alpha=0.1)
# adding labels
g + labs(x="teden",
        y="delez okuzenih izmed testiranih",
        color="dan v tednu")
```

`geom_smooth()` using method = 'loess' and formula 'y ~ x'