

Predavanje 00 – Motivacija

Slabosti dela z Excel (v primerjavi z R):

- Veliko priložnosti za napake pri ročnem poseganju v podatke, kopiranju, ipd.
- Delo ni ponovljivo – informacija o korakih se izgubi.
- Manj primeren za večje količine podatkov.
- Manj primeren za zahtevnejše analize.
- Težje pripraviti vizualizacije.

Primer

```
library(xlsx)
```

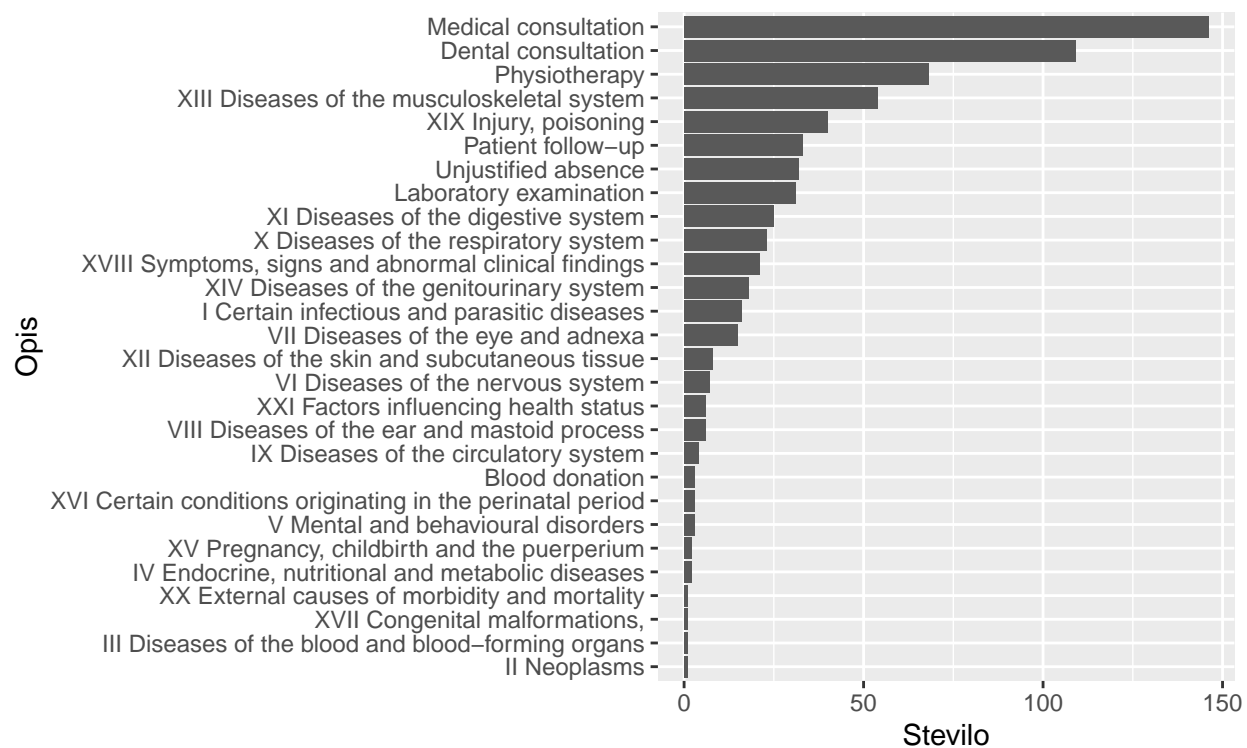
```
## Warning: package 'xlsx' was built under R version 3.5.3
```

```
library(ggplot2)
```

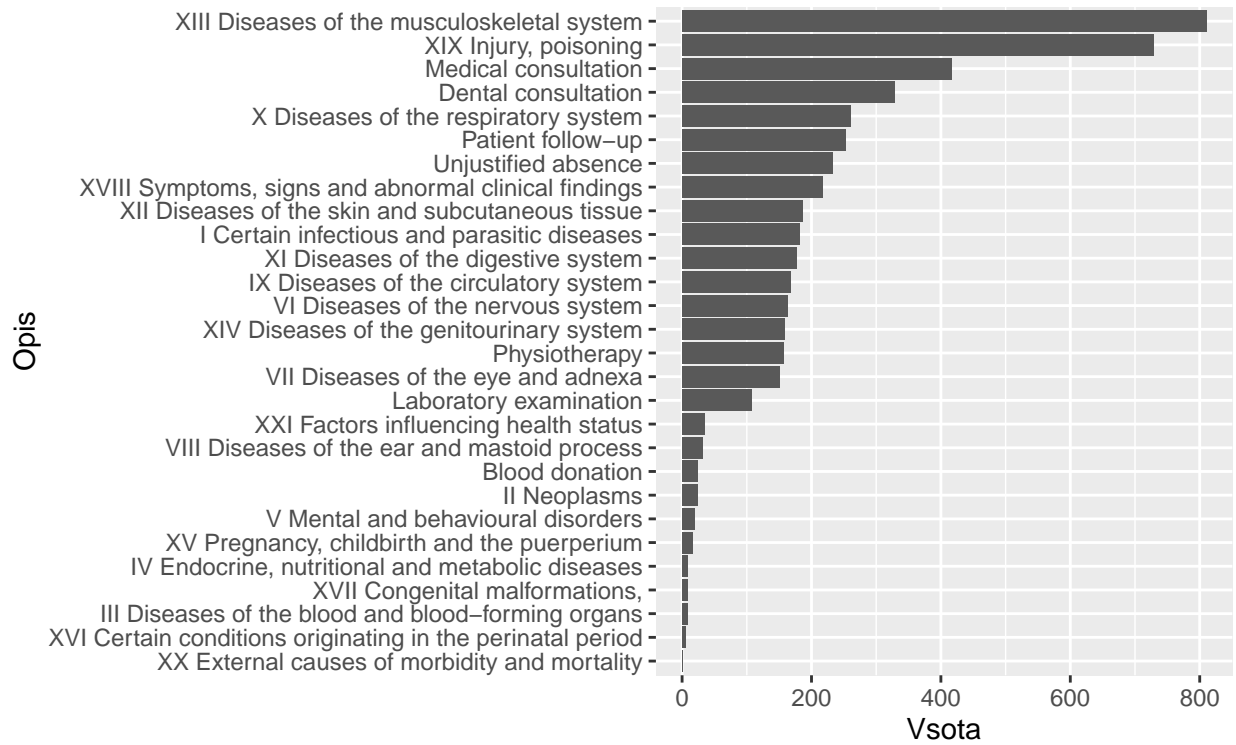
```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
dat <- xlsx::read.xlsx("odsotnosti_R.xls", sheetIndex = 1)
rsn <- xlsx::read.xlsx("sifrant_odsotnosti.xlsx", sheetIndex = 1)
# Podatki so dostopni na https://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work.

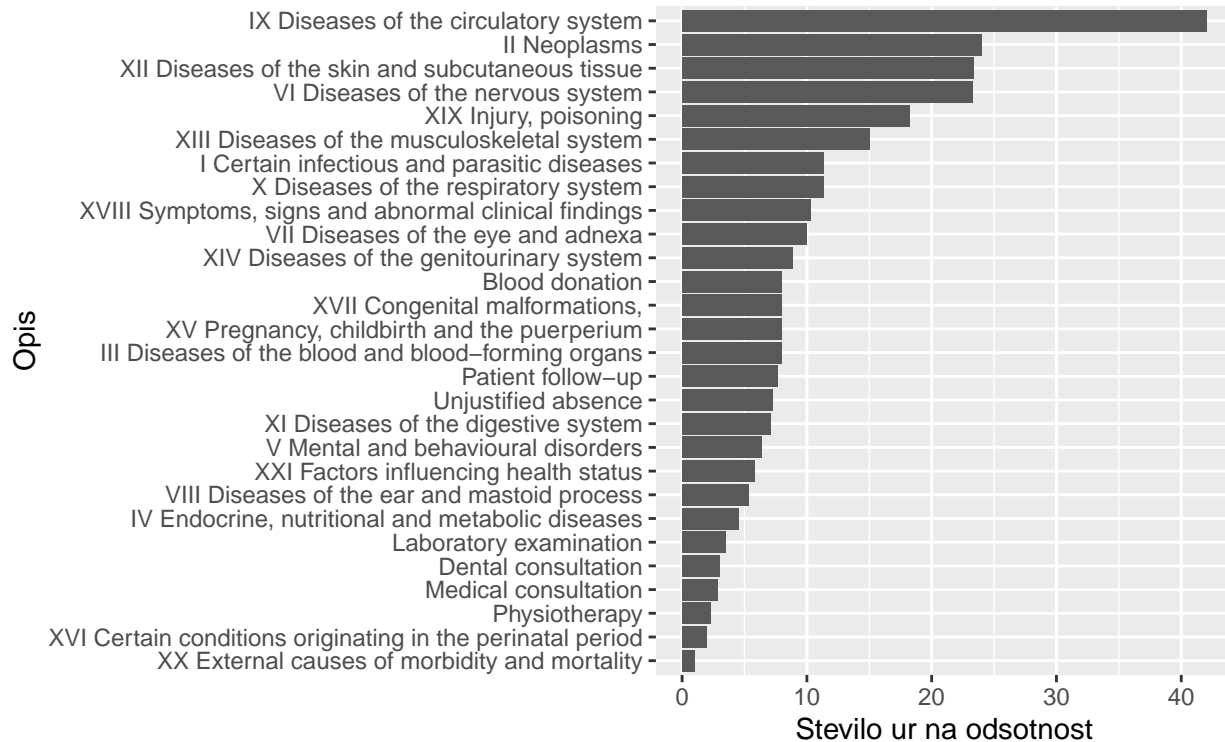
df <- data.frame(rsn,
                 Stevilo = tapply(dat$Razlog, dat$Razlog, length),
                 Vsota = tapply(dat$Ure, dat$Razlog, sum))
df$Opis <- factor(df$Opis, levels = df$Opis[order(df$Stevilo)])
ggplot(df, aes(x = Opis, y = Stevilo)) + geom_bar(stat = "identity") +
  coord_flip()
```



```
df$Opis <- factor(df$Opis, levels = df$Opis[order(df$Vsota)])
ggplot(df, aes(x = Opis, y = Vsota)) + geom_bar(stat = "identity") +
  coord_flip()
```



```
df$Razmerje <- df$Vsota / df$Stevilo
df$Opis <- factor(df$Opis, levels = df$Opis[order(df$Razmerje)])
ggplot(df, aes(x = Opis, y = Razmerje)) + geom_bar(stat = "identity") +
  coord_flip() + ylab("Stevilo ur na odsotnost")
```



Podatki, ki smo jih uporabili v tem poglavju, so del znanstvenega članka Martiniano et al. (2012).

REFERENCES

Martiniano, A, RP Ferreira, RJ Sassi, and C Affonso. 2012. "Application of a neuro fuzzy network in prediction of absenteeism at work." In *7th Iberian Conference on Information Systems and Technologies (Cisti 2012)*, 1–4. IEEE.