

3.1 Hands-On: ggplot II

Marco Wähler

2022-03-25

Aufgabenstellung - Lösung

Untersuchen Sie, ob sich im Mittel die Wahlbeteiligungsabsicht zwischen Befragten in Ost- und Westdeutschland unterscheidet. Arbeiten Sie dazu die folgenden Aufgaben durch.

Aufgabe 1

Importieren Sie nun die Daten aus der GLES 2021.

```
library(rio)
```

```
df <- import("data/gles_2021_gesamt.sav")
```

Im Datensatz sind sowohl die Vorwahl- als auch die Nachwahl-Daten. In der Variable “sample” ist die jeweilige Befragung gespeichert. Filtern Sie nun nach der Variable “sample” und speichern Sie die Befragten aus der Vorwahl-Befragung im neuen Dataframe “df_pre” ab. Arbeiten Sie bitte nun mit den Daten der Vorwahl-Befragung.

```
attributes(df$sample)
```

```
## $label
## [1] "Stichprobe"
##
## $format.spss
## [1] "F32.0"
##
## $labels
## GLES Querschnitt 2021, Vorwahl GLES Querschnitt 2021, Nachwahl
##                                7                                8
```

```
#option 1
```

```
df_pre <- df[df$sample == 7, ]
```

```
#option 2
```

```
df_pre <- subset(df, sample == 7)
```

Aufgabe 2

Rekodieren Sie die Variable „ostwest“ in 0 „Ostdeutschland“ und 1 „Westdeutschland“ um. Die neue Variable soll die Klasse `factor` haben.

```
attributes(df$ostwest)

## $label
## [1] "Ost/West (Zuordnung von Berlin zu Ostdeutschland)"
##
## $format.spss
## [1] "F31.0"
##
## $labels
## Ostdeutschland Westdeutschland
##           0           1

df_pre$west <- factor(df_pre$ostwest, labels = c("Ostdeutschland", "Westdeutschland"))

class(df_pre$west)

## [1] "factor"
```

Aufgabe 3

Rekodieren Sie die Variable zur Wahlbeteiligungsabsicht (q6) um, sodass hohe Werte eine hohe Wahlbeteiligungsabsicht signalisieren. Vergeben Sie einen neuen Namen für die rekodierte Variable. * „bestimmt zur Wahl gehen“ erhält den Wert 5 * „wahrscheinlich zur Wahl gehen“ erhält den Wert 4 * „vielleicht zur Wahl gehen“ erhält den Wert 3 * „wahrscheinlich nicht zur Wahl gehen“ erhält den Wert 2 * „bestimmt nicht zur Wahl gehen“ erhält den Wert 1 * alle weiteren Ausprägungen erhalten ein Missing Value

Die neue Variable soll die Klasse „numeric“ haben.

```
df_pre$vote_int[df_pre$q6 == 1] <- 5
df_pre$vote_int[df_pre$q6 == 2] <- 4
df_pre$vote_int[df_pre$q6 == 3] <- 3
df_pre$vote_int[df_pre$q6 == 4] <- 2
df_pre$vote_int[df_pre$q6 == 5] <- 1

class(df_pre$vote_int)

## [1] "numeric"
```

Geben Sie nun eine Häufigkeitsauszählung inklusive der fehlenden Werte aus.

```
table(df_pre$vote_int, useNA = "ifany")

##
##    1    2    3    4    5 <NA>
##   39   55   74  217 3163 1568
```

Aufgabe 4

Berechnen Sie den Mittelwert zur Wahlabsicht getrennt für Ost- und Westdeutschland.

Aufgabe 5

Plotten Sie die Mittelwerte und Standardabweichung zur Wahlabsicht für Ost- und Westdeutschland. Beschriften Sie die X- und Y-Achse sinnvoll. Fügen Sie einen entsprechenden Titel für Ihre Grafik ein. Eine Grafik können Sie mit einem Klick auf „Export“ und der Wahl eines Formats speichern. Können Sie einen Weg finden, um den Plot per Code zu speichern? Suchen Sie in den Skripten oder im Internet nach einer entsprechenden Lösung.

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.6      v dplyr  1.0.8
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
df_sum <- df_pre %>%
  select(west, vote_int) %>%
  group_by(west) %>%
  summarise(mw = mean(vote_int, na.rm=TRUE),
            sd = sd(vote_int, na.rm=TRUE))
```

```
f <- ggplot(df_sum, aes(x=west, y=mw)) +
  geom_errorbar(aes(ymin=mw-sd, ymax=mw+sd), width=.1) +
  geom_line() +
  geom_point() +
  scale_y_continuous(limits = c(1,7), breaks = c(1:7)) + labs(x = "Ost-/Westdeutschland", y = "Wahlabsi
f
```

```
## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?
```

