

3.1 Hands-On: ggplot

Marco Wähler

2022-03-25

Aufgabenstellung - Lösung

Aufgabe 1

Importieren Sie die Daten aus dem ALLBUS 2018.

```
library(rio)
```

```
df <- import("data/allbus_2018_gesamt.sav")
```

Aufgabe 2

Folgende Variablen erheben die Mitgliedschaft(en) der Befragten in (zivilgesellschaftlichen) Organisationen

- sm04 - MITGLIEDSSTATUS: KULTURVEREIN
- sm05 - MITGLIEDSSTATUS: SPORTVEREIN
- sm06 - MITGLIEDSSTATUS: SONST. HOBBYVEREIN
- sm07 - MITGLIEDSSTATUS: WOHLTAETIGKEITSVEREIN
- sm08 - MITGLIEDSSTATUS: MENSCHENRECHTSORGAN
- sm09 - MITGLIEDSSTATUS: NATURSCHUTZORGANISATION
- sm10 - MITGLIEDSSTATUS: GESUNDHEITSVEREIN
- sm11 - MITGLIEDSSTATUS: ELTERNORGANISATION
- sm12 - MITGLIEDSSTATUS: SENIORENVEREIN
- sm13 - MITGLIEDSSTATUS: BUERGERINITIATIVE
- sm14 - MITGLIEDSSTATUS: ANDERE VEREINE

Lassen Sie sich zunächst die Eigenschaften der Variablen anzeigen.

```
orga <- c("sm04", "sm05", "sm06", "sm07", "sm08", "sm09", "sm10", "sm11", "sm12", "sm13", "sm14")  
lapply(df[, c(orga)], function(x) attr(x, "label"))
```

```
## $sm04
## [1] "MITGLIEDSSTATUS: KULTURVEREIN"
##
## $sm05
## [1] "MITGLIEDSSTATUS: SPORTVEREIN"
##
## $sm06
## [1] "MITGLIEDSSTATUS: SONST. HOBBYVEREIN"
##
## $sm07
## [1] "MITGLIEDSSTATUS: WOHLTAETIGKEITSVEREIN"
##
## $sm08
## [1] "MITGLIEDSSTATUS: MENSCHENRECHTSORGAN."
##
## $sm09
## [1] "MITGLIEDSSTATUS: NATURSCHUTZORGANISATION"
##
## $sm10
## [1] "MITGLIEDSSTATUS: GESUNDHEITSVEREIN"
##
## $sm11
## [1] "MITGLIEDSSTATUS: ELTERNORGANISATION"
##
## $sm12
## [1] "MITGLIEDSSTATUS: SENIORENVEREIN"
##
## $sm13
## [1] "MITGLIEDSSTATUS: BUERGERINITIATIVE"
##
## $sm14
## [1] "MITGLIEDSSTATUS: ANDERE VEREINE"
```

```
lapply(df[, c(orga)], function(x) attr(x, "labels"))
```

```
## $sm04
##      KEINE ANGABE      KEIN MITGLIED PASSIVES MITGLIED  AKTIVES MITGLIED
##              -9              1              2              3
##      EHRENAMT
##              4
##
## $sm05
##      KEINE ANGABE      KEIN MITGLIED PASSIVES MITGLIED  AKTIVES MITGLIED
##              -9              1              2              3
##      EHRENAMT
##              4
##
## $sm06
##      KEINE ANGABE      KEIN MITGLIED PASSIVES MITGLIED  AKTIVES MITGLIED
##              -9              1              2              3
##      EHRENAMT
##              4
##
## $sm07
```

```

##      KEINE ANGABE      KEIN MITGLIED PASSIVES MITGLIED      AKTIVES MITGLIED
##      -9              1              2              3
##      EHRENAMT
##      4
##
## $sm08
##      KEINE ANGABE      KEIN MITGLIED PASSIVES MITGLIED      AKTIVES MITGLIED
##      -9              1              2              3
##      EHRENAMT
##      4
##
## $sm09
##      KEINE ANGABE      KEIN MITGLIED PASSIVES MITGLIED      AKTIVES MITGLIED
##      -9              1              2              3
##      EHRENAMT
##      4
##
## $sm10
##      KEINE ANGABE      KEIN MITGLIED PASSIVES MITGLIED      AKTIVES MITGLIED
##      -9              1              2              3
##      EHRENAMT
##      4
##
## $sm11
##      KEINE ANGABE      KEIN MITGLIED PASSIVES MITGLIED      AKTIVES MITGLIED
##      -9              1              2              3
##      EHRENAMT
##      4
##
## $sm12
##      KEINE ANGABE      KEIN MITGLIED PASSIVES MITGLIED      AKTIVES MITGLIED
##      -9              1              2              3
##      EHRENAMT
##      4
##
## $sm13
##      KEINE ANGABE      KEIN MITGLIED PASSIVES MITGLIED      AKTIVES MITGLIED
##      -9              1              2              3
##      EHRENAMT
##      4
##
## $sm14
##      KEINE ANGABE      KEIN MITGLIED PASSIVES MITGLIED      AKTIVES MITGLIED
##      -9              1              2              3
##      EHRENAMT
##      4

```

Im Anschluss eine Häufigkeitsauszählung (inklusive Missing Values) jeder Variablen.

```
lapply(df[, c(orga)], function(x) table(x, useNA = "ifany"))
```

```
## $sm04
## x
```

```

##      1      2      3      4 <NA>
## 2995  154  291   35      2
##
## $sm05
## x
##      1      2      3      4 <NA>
## 2347  290  796   42      2
##
## $sm06
## x
##      1      2      3      4 <NA>
## 3163   69  228   15      2
##
## $sm07
## x
##      1      2      3      4 <NA>
## 3109  195  133   38      2
##
## $sm08
## x
##      1      2      3      4 <NA>
## 3389   57   22    7      2
##
## $sm09
## x
##      1      2      3      4 <NA>
## 3199  182   85    9      2
##
## $sm10
## x
##      1      2      3      4 <NA>
## 3304   79   79   13      2
##
## $sm11
## x
##      1      2      3      4 <NA>
## 3312   47   98   18      2
##
## $sm12
## x
##      1      2      3      4 <NA>
## 3418   23   28    5      3
##
## $sm13
## x
##      1      2      3      4 <NA>
## 3401   32   36    6      2
##
## $sm14
## x
##      1      2      3      4 <NA>
## 3056  129  234   56      2

```

Aufgabe 3

Rekodieren Sie die Variablen in Dummy-Variablen mit der Ausprägung 0 (keine Mitgliedschaft) und 1 (passives Mitglied, aktives Mitglied, Ehrenamt)

```
#Rekodierung
```

```
df$sm04_d <- ifelse(df$sm04 == 1, 0, 1)
df$sm05_d <- ifelse(df$sm05 == 1, 0, 1)
df$sm06_d <- ifelse(df$sm06 == 1, 0, 1)
df$sm07_d <- ifelse(df$sm07 == 1, 0, 1)
df$sm08_d <- ifelse(df$sm08 == 1, 0, 1)
df$sm09_d <- ifelse(df$sm09 == 1, 0, 1)
df$sm10_d <- ifelse(df$sm10 == 1, 0, 1)
df$sm11_d <- ifelse(df$sm11 == 1, 0, 1)
df$sm12_d <- ifelse(df$sm12 == 1, 0, 1)
df$sm13_d <- ifelse(df$sm13 == 1, 0, 1)
df$sm14_d <- ifelse(df$sm14 == 1, 0, 1)
```

```
orga_d <- c("sm04_d", "sm05_d", "sm06_d", "sm07_d", "sm08_d", "sm09_d", "sm10_d", "sm11_d", "sm12_d", "sm13_d", "sm14_d")
lapply(df[, c(orga_d)], function(x) table(x, useNA = "ifany"))
```

```
## $sm04_d
## x
##    0    1 <NA>
## 2995  480    2
##
## $sm05_d
## x
##    0    1 <NA>
## 2347 1128    2
##
## $sm06_d
## x
##    0    1 <NA>
## 3163  312    2
##
## $sm07_d
## x
##    0    1 <NA>
## 3109  366    2
##
## $sm08_d
## x
##    0    1 <NA>
## 3389   86    2
##
## $sm09_d
## x
##    0    1 <NA>
## 3199  276    2
##
## $sm10_d
## x
```

```
##      0      1 <NA>
## 3304  171      2
##
## $sm11_d
## x
##      0      1 <NA>
## 3312  163      2
##
## $sm12_d
## x
##      0      1 <NA>
## 3418   56      3
##
## $sm13_d
## x
##      0      1 <NA>
## 3401   74      2
##
## $sm14_d
## x
##      0      1 <NA>
## 3056  419      2
```

Aufgabe 4

Erstellen Sie nun einen additiven Index, der aus zählt, in wie vielen Organisationen die Befragten eine Mitgliedschaft haben. Geben Sie die absolute und prozentuale Häufigkeit aus

```
#index
df$orga_index <- rowSums(df[, c(orga_d)])
```

```
table(df$orga_index)
```

```
##
##      0      1      2      3      4      5      6      7      9     10     11
## 1478 1045   586   227    92    31     7     4     2     1     1
```

```
100*prop.table(table(df$orga_index))
```

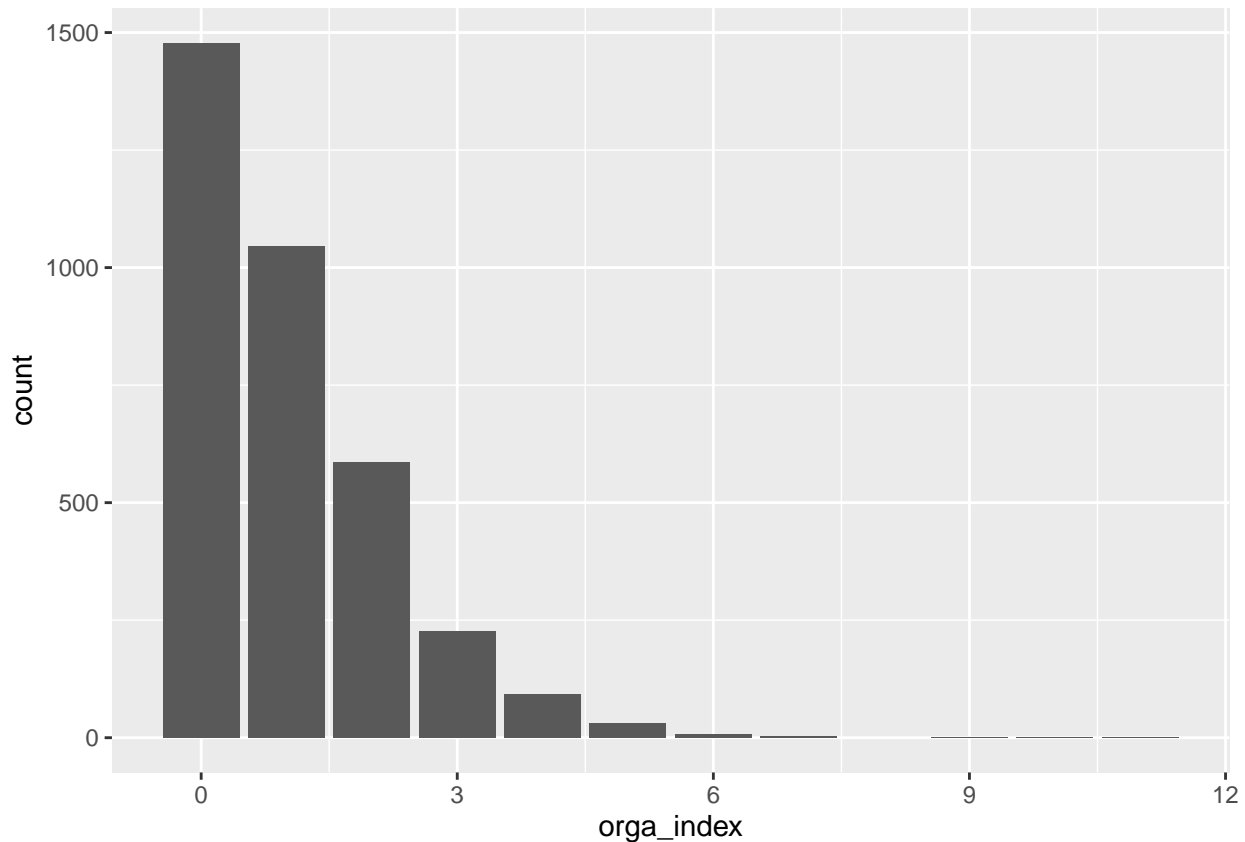
```
##
##           0           1           2           3           4           5
## 42.54461716 30.08059873 16.86816350  6.53425446  2.64824410  0.89234312
##           6           7           9          10          11
##  0.20149683  0.11514105  0.05757052  0.02878526  0.02878526
```

Aufgabe 5

Erstellen Sie nun eine passende Grafik, die einen Überblick über die Häufigkeit der Mitgliedschaften anbietet. Falls es sinnvoll erscheint, fassen Sie die Ausprägungen nochmals zusammen (beispielsweise die Ausprägungen 5 - 12). Dabei kann es hilfreich sein, die neue Variable als **factor** mit entsprechenden **levels** zu speichern. Hinweis: Wenn Sie mit einem **factor** arbeiten, dann kann es sein, dass zusätzlich die Missing Values in die

Graphik aufgenommen werden. Um das zu verhindern, können Sie einfach die Funktion `drop.na(variable)` nutzen.

```
library(tidyverse)
df %>% ggplot(aes(x=orga_index)) +
  geom_bar()
```



Aufgabe 6

Es erscheint nicht sinnvoll alle möglichen Mitgliedschaften zu plotten. Rekodieren Sie die Variable und fassen Sie Befragte mit gleich/mehr als fünf Mitgliedschaften in einer Kategorie zusammen. Auch die neue Variable soll die Klasse `factor` haben. Erstellen Sie anschließend einen Plot mit entsprechender Beschriftung.

```
#5 bis 12 zusammenfassen
df$orga_index[df$orga_index == 0] <- "keine Mitgliedschaft"
df$orga_index[df$orga_index == 1] <- "1"
df$orga_index[df$orga_index == 2] <- "2"
df$orga_index[df$orga_index == 3] <- "3"
df$orga_index[df$orga_index == 4] <- "4"
df$orga_index[df$orga_index %in% c(5:14)] <- ">5 Mitgliedschaften"
```

```
orga_level <- c("keine Mitgliedschaft", "1", "2", "3", "4", ">5 Mitgliedschaften")
df$orga_index <- factor(df$orga_index, levels = orga_level)
```

```
class(df$orga_index)
```

```
## [1] "factor"
```

```
table(df$orga_index)
```

```
##  
## keine Mitgliedschaft      1      2  
##      1478      1045      586  
##      3      4 >5 Mitgliedschaften  
##      227      92      46
```

```
df %>% drop_na(orga_index) %>% ggplot(aes(x=orga_index)) +  
  geom_bar()
```

