

03 Einführung R - Übung

Kurs: 'Einführung in die quantitativen Forschungsmethoden', Prof. Dr. Theresa Gessler SoSe 2023

Teil 1

Führen Sie die folgende Zeile aus

```
?print
```

Beginnen Sie (langsam) einen R-Befehle zu tippen - z.B. library(). Was passiert?

Fragen Sie nach Hilfe zum Befehl sum()

Führen Sie die nächsten beiden Zeilen aus. Welche Unterschiede sehen Sie?

```
1+3
```

```
[1] 4
```

```
summe <- 1+3
```

Erzeugen Sie einen Vektor mit mehreren Werten.

```
myvector <- c(5,3,4)
```

Fragen Sie die Klasse dieses Vektors ab

```
class(myvector)
```

```
[1] "numeric"
```

Tabellieren Sie die Häufigkeit der Inhalte des Vektors

```
table(myvector)
```

```
myvector  
3 4 5  
1 1 1
```

Drucken Sie den Vektor mit `print()`

```
print(myvector)
```

```
[1] 5 3 4
```

Teil 2

Installieren und laden Sie das tidyverse Paket.

```
install.packages("tidyverse")
```

```
library(tidyverse)
```

Der Befehl erzeugt einen Dataframe. Führen Sie ihn aus.

```
df <- data.frame(semester=c("SoSe","WiSe"),  
                    jahr=c(2023,2024),  
                    kurs=c("Quantitative Sozialforschung","Ihr nächster Kurs"))
```

Rufen Sie eine Variable aus dem Datensatz ab.

```
df$jahr
```

```
[1] 2023 2024
```

Sehen Sie den Data Frame und die Variablen in R an. Welche Variablen Typen sehen Sie mit `class()`?

```
class(df)
```

```
[1] "data.frame"
```

```
class(df$semester)
```

```
[1] "character"
```

```
class(df$jahr)
```

```
[1] "numeric"
```

```
class(df$kurs)
```

```
[1] "character"
```

Erstellen Sie eine neue Variable mit ECTS Punkten.

```
df$ects <- c(3,9)
```

Der folgende Code erzeugt einen Datensatz mit einigen der Ergebnisse unserer Umfrage aus der ersten Sitzung. (Normalerweise würden wir die Daten aus einem Datensatz laden). Führen Sie den Befehl aus.

```
umfrage <- structure(list(geburtsjahr = c(2004,
1998, 1997, 1992, 2002, 2003, 2002, 1997, 1996, 2002, 1993, 2002,
1998, 2003, 2001, 2001, 1998, 2005, 2003, 2003, 1999),
geschlecht = c("männlich",
"weiblich", "männlich", "männlich", "männlich", "weiblich",
"weiblich", "weiblich", "männlich", "weiblich", "weiblich",
"weiblich", "weiblich", "weiblich", "männlich", "männlich",
"männlich", "weiblich", "männlich", "weiblich",
"weiblich"),
gesellig = c("Ungefähr gleich oft", "Ungefähr gleich oft",
"Seltener als die meisten", "Viel seltener als die meisten",
"Ungefähr gleich oft", "Seltener als die meisten", "Ungefähr gleich oft",
"Häufiger als die meisten", "Ungefähr gleich oft", "Viel seltener als die meisten",
"Ungefähr gleich oft", "Ungefähr gleich oft", "Seltener als die meisten",
"Häufiger als die meisten", "Seltener als die meisten", "Seltener als die meisten",
"Ungefähr gleich oft", "Viel häufiger als die meisten", "Ungefähr gleich oft",
"Häufiger als die meisten", "Häufiger als die meisten"
)),
```

```
row.names = c(NA, -21L), class = c("spec_tbl_df",  
"tbl_df", "tbl", "data.frame"))
```

Sehen Sie sich den Datensatz an. Sehen Sie die ersten Observationen an und überprüfen Sie die Variablentypen.

```
head(umfrage)
```

```
# A tibble: 6 x 3  
  geburtsjahr geschlecht gesellig  
    <dbl> <chr>      <chr>  
1     2004 männlich  Ungefähr gleich oft  
2     1998 weiblich  Ungefähr gleich oft  
3     1997 männlich  Seltener als die meisten  
4     1992 männlich  Viel seltener als die meisten  
5     2002 männlich  Ungefähr gleich oft  
6     2003 weiblich  Seltener als die meisten
```

```
class(umfrage$geburtsjahr)
```

```
[1] "numeric"
```

```
class(umfrage$geschlecht)
```

```
[1] "character"
```

```
class(umfrage$gesellig)
```

```
[1] "character"
```

Erstellen Sie eine Tabelle der Geschlechter

```
table(umfrage$geschlecht)
```

```
männlich weiblich  
      9      12
```

Erstellen Sie eine Tabelle der Geselligkeit (wie oft nehmen Sie an geselligen Treffen teil)

```
table(umfrage$geburtsjahr)
```

```
1992 1993 1996 1997 1998 1999 2001 2002 2003 2004 2005
  1     1     1     2     3     1     2     4     4     1     1
```

Berechnen Sie mit `mean()` das durchschnittliche Geburtsjahr.

Tipp: Fügen Sie `na.rm=TRUE` hinzu, um fehlende Werte zu überspringen.

```
mean(umfrage$geburtsjahr,na.rm=T)
```

```
[1] 1999.952
```

Erstellen Sie eine neue Variable die angibt, ob das Geburtsjahr nach 2000 ist.

```
umfrage$post2000 <- ifelse(umfrage$geburtsjahr>2000,"nach","vor")
```

Bonus: Berechnen Sie das durchschnittliche Geburtsjahr pro Geschlecht

```
umfrage_gruppiert <- group_by(umfrage,geschlecht)
summarize(umfrage_gruppiert,mean=mean(geburtsjahr))
```

```
# A tibble: 2 x 2
  geschlecht mean
  <chr>      <dbl>
1 männlich  1999.
2 weiblich  2000.
```

Bonus: Berechnen Sie das ungefähre Alter der Kursteilnehmenden.

Tipp: Sie können R beim Erstellen neuer Variablen einfach als Taschenrechner benutzen und Variablen von Zahlen abziehen.

```
umfrage$alter <- 2023-umfrage$geburtsjahr
```