

Lösungen zu den Aufgaben

1. Aufgabe

Laden Sie die folgende Tabellen mit folgendem Befehl aus dem Paket `tidyverse`:

Insgesamt sollten Sie als folgende Tabellen in Ihrem `environment` verfügbar haben:

- `table1`
- `table2`
- `table3`
- `table4`
- `table5`

Welche der Tabellen ist in der Normalform?

- a. `table1`
- b. `table2`
- c. `table3`
- d. `table4`
- e. `table5`

Lösung

Laden wir noch die übrigen Tabellen

Betrachten wir alle fünf Tabellen:

```
## # A tibble: 6 × 4
##   country    year cases population
##   <chr>      <dbl> <dbl>      <dbl>
## 1 Afghanistan 1999     745    19987071
## 2 Afghanistan 2000    2666    20595360
## 3 Brazil      1999   37737    172006362
## 4 Brazil      2000   80488    174504898
## 5 China       1999  212258   1272915272
## 6 China       2000  213766   1280428583

## # A tibble: 12 × 4
##   country    year type      count
##   <chr>      <dbl> <chr>      <dbl>
## 1 Afghanistan 1999 cases         745
## 2 Afghanistan 1999 population 19987071
## 3 Afghanistan 2000 cases         2666
## 4 Afghanistan 2000 population 20595360
## 5 Brazil      1999 cases         37737
## # ... with 7 more rows

## # A tibble: 6 × 3
##   country    year rate
##   <chr>      <dbl> <chr>
## 1 Afghanistan 1999 745/19987071
## 2 Afghanistan 2000 2666/20595360
## 3 Brazil      1999 37737/172006362
## 4 Brazil      2000 80488/174504898
## 5 China       1999 212258/1272915272
## 6 China       2000 213766/1280428583
```

```
## # A tibble: 3 × 3
##   country    `1999` `2000`
##   <chr>      <dbl> <dbl>
## 1 Afghanistan    745   2666
## 2 Brazil        37737  80488
## 3 China         212258 213766

## # A tibble: 6 × 4
##   country    century year    rate
##   <chr>      <dbl> <chr> <chr>
## 1 Afghanistan    19  99    745/19987071
## 2 Afghanistan    20  00    2666/20595360
## 3 Brazil         19  99    37737/172006362
## 4 Brazil         20  00    80488/174504898
## 5 China          19  99    212258/1272915272
## 6 China          20  00    213766/1280428583
```

Man sieht, dass nur Tabelle 1 “tidy” ist.

[Quelle](#)

- a. Wahr
- b. Falsch
- c. Falsch
- d. Falsch
- e. Falsch

2. Aufgabe

Laden Sie den Datensatz `affairs`:

```
affairs_path <- "https://vincentarelbundock.github.io/Rdatasets/csv/AER/Affairs.csv"
affairs <- read_csv(affairs_path)
```

Lesen Sie das *Data Dictionary* [hier](#).

Wir definieren als “Halodrie” eine Person mit mindestens einer Affäre (laut Datensatz).

Bearbeiten Sie folgende Aufgaben:

1. Filtern Sie mal nach Halodries!
2. Sortieren Sie (absteigend) nach Anzahl der Affären!
3. Wählen Sie die Spalten zu Anzahl der Affären, ob es Kinder in der Ehe gibt und die Zufriedenheit mit der Ehe. Dann sortieren Sie dann nach Anzahl der Kinder und *danach* nach der Anzahl der Affären.
4. Berechnen Sie die mittlere Anzahl der Affären!
5. Berechnen Sie die mittlere Anzahl der Affären pro Geschlecht und aufgeteilt auf Partnerschaften mit bzw. ohne Kinder.
6. Geben Sie für jede Person die höhere der zwei Zahlen von Religiosität und Ehezufriedenheit aus!
7. Berechnen Sie jeweils das Heiratsalter!

Lösung

Ad 1.

```

affairs %>%
  filter(affairs > 0)

## # A tibble: 150 × 10
##   ...1 affairs gender age yearsmarried children religiousness education
##   <dbl>   <dbl> <chr>  <dbl>         <dbl> <chr>         <dbl>   <dbl>
## 1      6      3 male    27          1.5  no            3      18
## 2     12      3 female  27          4    yes            3      17
## 3     43      7 male    37          15   yes            5      18
## 4     53     12 female  32          10   yes            3      17
## 5     67      1 male    22          0.125 no            4      16
## # ... with 145 more rows, and 2 more variables: occupation <dbl>, rating <dbl>

```

Ad 2.

```

affairs %>%
  arrange(-affairs)

## # A tibble: 601 × 10
##   ...1 affairs gender age yearsmarried children religiousness education
##   <dbl>   <dbl> <chr>  <dbl>         <dbl> <chr>         <dbl>   <dbl>
## 1     53     12 female  32          10   yes            3      17
## 2    122     12 male    37          15   yes            4      14
## 3    174     12 female  42          15   yes            5       9
## 4    176     12 male    37          10   yes            2     20
## 5    181     12 female  32          15   yes            3      14
## # ... with 596 more rows, and 2 more variables: occupation <dbl>, rating <dbl>

```

Ad 3.

```

affairs %>%
  select(affairs, rating, children) %>%
  arrange(children, affairs)

## # A tibble: 601 × 3
##   affairs rating children
##   <dbl>   <dbl> <chr>
## 1      0      4 no
## 2      0      4 no
## 3      0      3 no
## 4      0      5 no
## 5      0      3 no
## # ... with 596 more rows

```

Ad 4.

```

affairs %>%
  summarise(affairs_mean = mean(affairs))

## # A tibble: 1 × 1
##   affairs_mean
##   <dbl>
## 1      1.46

```

Ad 5.

```

affairs %>%
  group_by(gender, children) %>%
  summarise(affairs_mean = mean(affairs))

## # A tibble: 4 × 3
## # Groups:   gender [2]
##   gender children affairs_mean
##   <chr>   <chr>         <dbl>
## 1 female no            0.838
## 2 female yes           1.69

```

```
## 3 male    no          1.01
## 4 male    yes         1.66
```

Ad 6.

```
affairs %>%
  group_by(...1) %>%
  summarise(max(c(religiousness, rating)))

## # A tibble: 601 × 2
##   ...1 `max(c(religiousness, rating))`
##   <dbl>                                <dbl>
## 1      4                                4
## 2      5                                4
## 3      6                                4
## 4     11                                4
## 5     12                                5
## # ... with 596 more rows
```

Ad 7.

```
affairs %>%
  mutate(heiratsalter = age - yearsmarried)

## # A tibble: 601 × 11
##   ...1 affairs gender age yearsmarried children religiousness education
##   <dbl>   <dbl> <chr>  <dbl>         <dbl> <chr>          <dbl>   <dbl>
## 1      4      0 male    37          10    no            3       18
## 2      5      0 female  27           4    no            4       14
## 3     11      0 female  32          15    yes           1       12
## 4     16      0 male    57          15    yes           5       18
## 5     23      0 male    22          0.75 no            2       17
## # ... with 596 more rows, and 3 more variables: occupation <dbl>, rating <dbl>,
## #   heiratsalter <dbl>
```

3. Aufgabe

Importieren Sie den folgenden Datensatz in R:

```
mtcars <- read_csv("https://vincentarelbundock.github.io/Rdatasets/csv/datasets/mtcars.csv")
```

Übersetzen Sie dann die folgende R-Sequenz ins Deutsche:

```
mtcars %>%
  drop_na() %>%
  select(mpg, hp, cyl) %>%
  filter(hp > 100, cyl >= 6) %>%
  group_by(cyl) %>%
  summarise(mpg_mean = mean(mpg))

## # A tibble: 2 × 2
##   cyl mpg_mean
##   <dbl>   <dbl>
## 1      6    19.7
## 2      8    15.1
```

Lösung

Hey R:

1. Nimm den Datensatz `mtcars` UND DANN
2. hau alle Zeilen raus, in denen es fehlende Werte gibt UND DANN
3. wähle (selektiere) die folgenden Spalten: Spritverbrauch, PS, Zylinder UND DANN
4. filter Autos mit mehr als 100 PS und mit mindestens 6 Zylindern UND DANN
5. gruppieren nach der Zahl der Zylinder UND DANN
6. fasse den Verbrauch zum Mittelwert zusammen.