

# Aufgabe 3: Daten explorieren

## *FOM R-Training*

HINWEIS: Für die Beantwortung der Aufgaben und für weitere Informationen können die Cheat Sheets *C3-Kennzahlen*, *C3-Visualisierungen-basics* und *C3-Visualisierungen-advanced* zu Hilfe genommen werden.

### **Kennzahlen-Summary des Datensatzes**

Die Funktion `inspect` aus dem Paket `mosaic` zeigt auf einen Blick alle notwendigen Informationen über einen Datensatz. Der Befehl lautet

```
inspect(mydata)
```

### **Aufgabe a)**

Laden Sie den Datensatz `tips.csv` ein und wenden Sie `inspect` darauf an. Wie viele kategoriale Beobachtungen liegen vor? Wie hoch ist das mittlere Trinkgeld?

---

### **Visualisierung eindimensionaler kategorialer Variablen**

Die Funktion `bargraph` aus dem Paket `mosaic` erstellt ein Balkendiagramm nach folgendem Syntax:

```
bargraph( ~ variable_of_interest, data = mydata)
```

### **Aufgabe b)**

Wenden Sie diesen Befehl auf den `tips`-Datensatz an, um den Anteil von Frauen und Männern (Variable `sex`) grafisch darzustellen. Welches Geschlecht zahlt häufiger die Rechnung?

Die gleiche Funktion kann verwendet werden, um den Anteil von Frauen und Männer je nach Zeitpunkt des Essens (Variable `time`) zu unterteilen. Der Syntax ist

```
bargraph( ~ variable_of_interest | grouping_variable, data = mydata)
```

### **Aufgabe c)**

Wenden Sie diesen Befehl auf den `tips`-Datensatz an, um den Anteil von Frauen und Männern bedingt auf den Zeit des Essens grafisch darzustellen. Welches Geschlecht zahlt zu welcher Zeit häufiger die Rechnung?

---

---

## Visualisierung eindimensionaler quantitativer Variablen

Die Funktion `histogram` aus dem Paket `mosaic` erstellt ein Histogramm nach dem gleichen Syntax wie die Funktion `bargraph`. Als Input wird lediglich eine quantitative (auch metrische) Variable verlangt.

### Aufgabe d)

Wenden Sie `histogram` auf den `tips`-Datensatz an, um die Verteilung der gezahlten Rechnungen (Variable `total_bill`) abzubilden. Wie hoch ist der Rechnungsbetrag der am häufigsten gezahlt wird?

### Aufgabe e)

Unterteilen Sie auch das Histogramm der Rechnungshöhe nach Frauen und Männern. Was fällt auf, wenn Sie die beiden Verteilungen miteinander vergleichen?

---

## Visualisierung zweidimensionaler quantitativer Variablen

Die Funktion `xyplot` aus dem Paket `mosaic` erstellt ein XY-Streuungsdiagramm nach folgendem Syntax:

```
xyplot( yvariable ~ xvariable, data = mydata)
```

### Aufgabe f)

Wenden Sie `xyplot` auf den `tips`-Datensatz an, um den Zusammenhang zwischen der Trinkgeldhöhe (Variable `tip`) und der Rechnungshöhe (Variable `total_bill`) abzubilden. Gibt es einen Zusammenhang zwischen diesen beiden Variablen?