

Aufgabe 3: Daten explorieren

FOM R-Training

HINWEIS: Für die Beantwortung der Aufgaben und für weitere Informationen können die Cheat Sheets *CS3-Kennzahlen*, *CS4-Visualisierungen-mit-Mosaic* und *CS5-Visualisierungen-generell* zu Hilfe genommen werden.

Aufgabe a) 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)
```

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?

Aufgabe b) Visualisierung eindimensionaler kategorialer Variablen

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

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

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( ~ unabhängige_diskrete_variable | bedingende_variable, data = mydata)
```

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?

Aufgabe c) 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.

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?

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

Aufgabe d) Visualisierung zweidimensionaler quantitativer Variablen

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

```
xyplot( abhängige_variable ~ unabhängige_variable, data = mydata)
```

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?