

Aufgabe 2: Daten aufbereiten

FOM R-Training

HINWEIS: Für die Beantwortung der Aufgaben und für weitere Informationen kann das Cheat Sheet *CS2-Data-Transformation-with-dplyr* zu Hilfe genommen werden.

Aufgabe a) Überblick über die Daten verschaffen

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 bereits in R existierenden Datensatz `Alcohol` ein und wenden Sie `inspect` darauf an. Wie viele Beobachtungen liegen vor? Wie viele Kategoriale Variablen liegen vor?

Aufgabe b) Filtern von Beobachtungen

Mit der `filter`-Funktion lassen sich Zeilen aus einem Dataframe extrahieren, die eine bestimmte Bedingung erfüllen. Das Ergebnis wird in einem neuen Dataframe gespeichert. Der Syntax ist wie folgt

```
dataframe.new <- dataframe.old %>%  
  filter(variablename == "condition")
```

Filtere alle Beobachtungen aus dem Datensatz `Alcohol`, die als Land *Deutschland* haben und speichere das Ergebnis in einem neuen Dataframe `Alcohol.GER`.

Der oben beschriebene Syntax der Funktion ist der gleiche für `filter`, `select`, `arrange`, `mutate`, `group_by` und `summarize`. Deswegen wird auf weitere Illustrationen im Folgenden verzichtet.

Aufgabe c) Filtern und auswählen

Verwende die `filter`- und `select`-Funktionen und speichere in einem neuen Dataframe mit Namen `Alcohol.2008` alle Beobachtungen, die als Jahr *2008* haben und wähle nur die Variablen `country` und `year` aus.

Aufgabe d) Ordnen

Verwende die `arrange`-Funktion und speichere in einem neuen Dataframe mit Namen `Alcohol.2008.sort` alle Beobachtungen, die als Jahr *2008* haben und sortiere nach den Alkoholgehalt. Welches Land hatte im Jahr 2008 den höchsten Alkoholkonsum?

Aufgabe e) Neue Spalten erzeugen

Verwende die `mutate`-Funktion, um eine neue Spalte in einem neuen Dataframe mit Namen `Alcohol.2008.rang` zu erzeugen. Sortiere den Dataframe nach der Spalte `rang`. Auf welchem Platz liegt Deutschland und mit wie viel Alkoholkonsum? Wer ist auf Rang 1 mit wie viel Alkoholkonsum?
