

Aufgabe 2: Daten aufbereiten

FOM R-Training

HINWEIS: Für die Beantwortung der Aufgaben und für weitere Informationen kann das Cheat Sheet *C2-Data-Transformation-with-dplyr* zu Hilfe genommen werden.

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")
```

Aufgabe a)

Filtere alle Beobachtungen aus dem Datensatz `Alcohol`, die als Land *Germany* 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.

Filtern und auswählen

Aufgabe b)

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`, `year` und `alcohol` aus.

Ordnen

Aufgabe c)

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

Neue Spalten erzeugen

Aufgabe d)

Verwende die `mutate`-Funktion, um eine neue Spalte `rang` zu erzeugen. Ränge erzeugst du mit Hilfe der Funktion `rank()`. Diese Spalte soll den Rang des Alkoholkonsums in absteigender Reihenfolge enthalten. Speichere diese Daten in einem neuen Dataframe mit Namen `Alcohol.2008.rang`. 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?