

Aufgabe 1:

Laden Sie die Datei *starwars.RData* in Ihren Workspace und betrachten Sie die einzelnen Datensätze. Im Folgenden sollen Sie die Datensätze *tidy* machen.

- (a) Der Datensatz *starwars2* ist nicht *tidy*. Räumen Sie den Datensatz so auf, dass sich alle Werte in einer Spalte befinden und alle Variablen die Spaltennamen sind.
- (b) Der Datensatz *starwars3* ist ebenfalls nicht *tidy*. Hier stehen teilweise Variablennamen in den Zellen. Beheben Sie dieses Problem.
- (c) Im Datensatz *starwars4* ist ebenfalls nicht *tidy*. Hier stehen zwei unterschiedliche Variablen in einer Spalte. Trennen Sie die beiden Variablen, um den Datensatz *tidy* zu machen.
- (d) Der Datensatz *starwars5* ist wie die Datensätze zuvor nicht *tidy*. Hier ist eine Variable auf zwei Spalten getrennt. Führen Sie die beiden Spalten zusammen.

Hinweis: Vergeben Sie für die jeweiligen Objekte, die Sie erzeugen, sinnvolle Objektnamen. Orientieren Sie sich hierbei an den Vorschlägen aus der ersten Lerneinheit.

Aufgabe 2:

Der Datensatz *starwars* existiert im Paket *dplyr*. Rufen Sie den ursprünglichen Datensatz auf und speichern Sie diesen unter einem geeigneten Objektnamen. Wie Sie sehen, enthält der Datensatz verschiedene Informationen wie Name, Gewicht, Körpergröße der Charaktere.

- (a) Nutzen Sie den vollständigen Datensatz *starwars* und lassen Sie sich mithilfe der Funktion *filter* aus dem Paket *dplyr* auszählen, wie viele Personen mit einem Gewicht größer als 75 kg und weniger als 100 kg und einer Körpergröße größer als 180 cm und kleiner als 190 cm welche Augenfarbe haben.
- (b) Filtern Sie als nächstes alle Beobachtungen aus dem Datensatz, die fehlende Werte bei den Variablen *mass* und *height* haben.
- (c) Sortieren Sie Ihren Datensatz einmal aufsteigend nach dem Gewicht und einmal absteigend nach der Körpergröße. Wie groß ist Darth Vader?
- (d) Berechnen Sie den Body-Maß-Index mit folgender vereinfachten Formel: $\frac{\text{Gewicht in kg}}{(\text{height in cm})^2}$. Filtern Sie zuvor alle Droiden und fehlenden Werte beim Gewicht und der Körpergröße aus dem Datensatz. Im Anschluss reduzieren Sie die Anzahl der Spalten auf die Variablen *name*, *height* und *mass* und berechnen dann den Body-Maß-Index.
- (e) Benennen Sie die Variablen *height*, *mass*, *hair_color* und *skin_color* des Datensatzes *starwars* mit ihren deutschen Begriffen.
- (f) Filtern Sie wiederum aus dem Datensatz alle Beobachtungen von Droiden und fehlenden Werten von Gewicht und Körpergröße. Wählen Sie dann die Spalten *name*, *height* und *mass* aus und berechnen Sie den Body-Maß-Index. Im Anschluss lassen Sie sich für die Variablen *height*, *mass* und *bmi* das arithmetische Mittel, Median und Standardabweichung in einem Schritt ausgeben.

- (g) Äquivalent zur Aufgabe 2 (f) führen Sie die gleichen Schritte durch, berechnen allerdings das arithmetische Mittel, Median und Standardabweichung der Variablen *height*, *mass* und *bmi* getrennt nach der Spezies.
- (h) Äquivalent zur Aufgabe 2 (f) führen Sie die gleichen Schritte durch, berechnen allerdings das arithmetische Mittel, Median und Standardabweichung der Variablen *height*, *mass* und *bmi* getrennt nach dem sozialen Geschlecht.
- (i) Äquivalent zur Aufgabe 2 (f) führen Sie die gleichen Schritte durch, berechnen allerdings das arithmetische Mittel, Median und Standardabweichung der Variablen *height*, *mass* und *bmi* getrennt nach dem sozialen Geschlecht und der Spezies. Das bedeutet, z. B. das arithmetische Mittel des Gewichts eines weiblichen Menschen.
- (j) Lassen Sie sich die absoluten Häufigkeiten des sozialen Geschlechts ausgeben.
- (k) Fügen Sie die Datensätze *starwars6* und *starwars7* zu einem Datensatz zusammen.

Hinweis: Speichern Sie Ihr Skript und Ihren Workspace in einem geeignetem Ordner.