

# Einführung in R für SozialwissenschaftlerInnen

## Übungsblatt Woche 7

### Vorgehensweise

Füge, wenn möglich, nur die Befehle in ihre Abgabe ein, die gefordert sind. Alle anderen Befehle können Dein Ergebnis verfälschen. Achte unbedingt darauf, keinen Plot-Befehl einzufügen!

### Aufgabe 1

#### Ziel: Wiederholung

Du bist am Nordpol und hilfst dem Weihnachtsmann. Zuerst beaufsichtigst du die Auswahl der Rentiere. Im November haben die Rentiere jeden Tag ein Wettrennen gemacht, um sich auf Weihnachten vorzubereiten. Die acht Rentiere, die im Durchschnitt am schnellsten waren, dürfen dieses Jahr den Schlitten des Weihnachtsmanns ziehen. Außerdem darf das Rentier Rudolph mit, um den Weg mit seiner roten Nase zu beleuchten.

Der Datensatz "Rentiere.csv" enthält für jedes Wettrennen die Schnelligkeit pro Rentier in km/h. Lies den Datensatz mit folgendem Befehl ein und speichere ihn in der Variable Rentiere.

```
Rentiere<-read.csv("resources/Rentiere.csv", encoding="UTF-8",row.names=1)
```

Wenn du die Hausaufgaben in R-Studio machst musst du evtl den Pfad ändern und noch zuvor dein Working-Directory setzen. **Bevor du das Skript ins Abgabesystem hochlädst musst du diese Befehle wieder löschen. Im Skript sollen sich nur der Befehl, den du hier siehst und die Befehle für die Hausaufgaben, befinden.**

#### a) Maximale Punktzahl: 2 Punkte

Überprüfe deinen Datensatz Rentiere und füge eventuelle Änderungen durch: Der Datensatz soll keine fehlenden Werte enthalten und alle Spalten sollen den Datentyp numeric haben.

#### b) Maximale Punktzahl: 2 Punkte

Berechne für jedes Rentier die durchschnittliche Schnelligkeit über alle Wettrennen hinweg, gerundet auf 0 Nachkommastellen, und speichere das Ergebnis in der Variable zeiten\_durchschn.

#### c) Maximale Punktzahl: 2 Punkte

Schreibe eine Funktion mit dem Namen top\_acht. Die Funktion erhält einen Vektor x. Der Vektor wird aufsteigend sortiert und die ersten 8 Elemente des Vektors werden ausgegeben. Hier siehst du ein Beispiel:

```
vek<-c(6,7,1,3,2,5,9,4,0,8)
top_acht(vek)
```

```
## [1] 0 1 2 3 4 5 6 7
```

#### d) Maximale Punktzahl: 2 Punkte

Erstelle eine Variable mit dem Namen `santas_gefolge`. Diese Variable soll die **Namen** der 8 Rentiere enthalten, die im Durchschnitt am schnellsten waren (=am wenigsten Zeit benötigt haben). Die Namen sollen in aufsteigender Reihenfolge sortiert sein, das heißt das schnellste Rentier an Stelle 1. Bei Rentieren, die gleich schnell sind, soll alphabetisch vorgegangen werden. Benutze die Funktion `top_acht` und die Variable `zeiten_durchschn`. Tipp: Überlege dir, mit welcher Funktion man die Namen eines Vektors erhält.

Füge anschließend den Namen “Rudolph” am Ende des Vektors hinzu.

## Aufgabe 2

### Ziel: Wiederholung

Nun hilfst du in der Wichtelwerkstatt bei der Verteilung und Verpackung der Geschenke.

#### a) Maximale Punktzahl: 1 Punkt

Der Datensatz “Kinder.csv” enthält alle Kinder mit ihrer Adresse und ihrem Benehmen in diesem Jahr. Speichere den Datensatz in der Variable `kinder`.

```
kinder<-read.csv("resources/Kinder.csv", encoding="UTF-8")
```

Wandle dann die Spalte “Benehmen” in den Datentyp `factor` um.

#### b) Maximale Punktzahl: 2 Punkte

Erstelle einen Bericht für den Weihnachtsmann und speichere ihn in der Variable `bericht`. Der Bericht soll exakt folgendes Format haben, wobei die Punkte durch die richtige Anzahl an artigen/unartigen Kindern ersetzt werden sollen. Füge keinen zusätzlichen (Leer)Zeichen.

```
## [1] "Dieses Jahr waren ... Kinder artig und ... Kinder unartig."
```

#### c) Maximale Punktzahl: 7 Punkte

Der Weihnachtsmann hat viele Wunschzettel von Kindern erhalten. Lade den Datensatz “Wunschliste.csv” herunter und speichere ihn in der Variable `wuensche`.

```
wuensche<-read.csv("resources/Wunschliste.csv", encoding="UTF-8")
```

Führe dann folgende Änderungen **in genau dieser Reihenfolge** durch:

- Manche Kinder schreiben dem Weihnachtsmann mehrmals. Entferne alle Duplikate. Beziehe dich dabei nur auf die ID der Kinder.

- Manche Briefe waren unleserlich. Entferne alle Zeilen, die einen fehlenden Wert enthalten.
- Füge an das Ende des Datensatzes die Spalte Geschenk hinzu. Kopiere die Werte aus der Spalte Wunsch in die Spalte Geschenk.
- Nicht alle Wünsche sind erfüllbar. Ersetze bei allen Kindern, die sich ein “Geschwisterchen” wünschen, das Geschenk durch eine “Puppe”. Ersetze bei allen Kindern, die sich ein “Haustier” wünschen, das Geschenk durch ein “Kuscheltier”.
- Wunschzettel, die nach dem Nikolaustag eintreffen, können nicht mehr bearbeitet werden. Ersetze bei allen Kindern, deren Wunschzettel nach dem 06.12.20 eingegangen ist, das Geschenk durch eine “Uhr”.

#### d) Maximale Punktzahl: 4 Punkte

Nur die Kinder, die dieses Jahr brav waren, erhalten Geschenke. Führe die folgenden Anweisungen **in genau dieser Reihenfolge** durch:

- Füge den Datensatz kinder mit der Spalte ID und Geschenk aus dem Datensatz wuensche zusammen und speichere ihn in der Variable gesamt. Auch Kinder, die keinen Wunschzettel geschrieben haben, sollen in dem Datensatz vorkommen.
- Gib bei allen Kindern, die keinen Wunschzettel geschrieben haben, das Geschenk “Stifte” an.
- Ersetze bei allen Kindern, die unartig waren, das Geschenk durch “Kohle”.
- Lösche die Spalte Benehmen aus dem Datensatz gesamt.

#### e) Maximale Punktzahl: 7 Punkte

Nun, da die Geschenke für alle Kinder feststehen, müssen die Geschenkesäcke genäht werden. Für jeden Kontinent gibt es einen Geschenkesack. Damit diese genug Gewicht aushalten, muss das Gesamtgewicht aller Geschenke für diesen Kontinent berechnet werden.

- Schreibe eine Funktion gewicht\_ermitteln, die das Gewicht eines Geschenks in Kilogramm berechnet. Die Funktion bekommt einen Wert x. Wenn x eine “Ritterburg” oder ein “Roller” ist, gibt die Funktion 4.5 zurück. Wenn x eine “Puppe” oder ein “Buch” ist, gibt die Funktion 1.0 zurück. Sonst gibt die Funktion 0.5 zurück.
- Erstelle einen Datensatz verpackung, der die Spalten Kontinent und Geschenk aus dem Datensatz gesamt enthält. Verändere nicht den Datensatz gesamt.
- Füge zum Datensatz verpackung eine Spalte Gewicht hinzu. Diese Spalte soll für jedes Geschenk in der Zeile das entsprechende Gewicht enthalten. Nutze deine Funktion gewicht\_ermitteln. Tipp: Die Funktion apply, die sehr ähnlich zur Funktion apply ist, könnte hilfreich sein. Recherchiere wie die Funktion apply funktioniert.
- Erstelle einen Dataframe gewicht\_uebersicht, der für jeden Kontinent das Gesamtgewicht der Geschenke enthält.

#### f) Maximale Punktzahl: 1 Punkt

Speicher die Anzahl an Kindern, die im Datensatz gesamt ein Puzzle als Geschenk erhalten, in der Variable puzzle.

Wenn du die Variable richtig belegt hast und in deinem Skript diesen Code ausführst (lade diesen Teil nicht ins Abgabesystem hoch), erhältst du eine Überraschung:

```
letters<-c("X","C","E","S","H","O","N","F","C","R","I"," ",""),":")
paste(letters[puzzle+c(1,6,2,3,0,4,0,9,5,0,7,8,0,4,11,10)], collapse="")
```