

Übungsblatt 3: Grundlagen von R

Wissenschaftliches Schreiben und Arbeiten, WiSe 2023/'24

Aufgabe A: Installation von R und RStudio

1. R: Besuche zunächst die Webseite cran.r-project.org und lies dir dort die Anweisungen durch, wie du auf deinem jeweiligen Betriebssystem (Windows / Mac OS / Linux) R installieren kannst.
 - Auf Windows oder Mac OS müsstest du dir erst ein Installationsprogramm herunterladen (*.exe- bzw. *.pkg-Datei) und dieses dann ausführen, um die Installation durchzuführen.
 - Auf Linux kannst du die Installation am einfachsten über deinen Paketmanager durchführen.
2. RStudio: Erst sobald die Installation von R (s. Schritt 1) erledigt ist, solltest du auch RStudio installieren. Besuche dafür die Webseite rstudio.com/products/rstudio/download und mache dich mit den Installationsanweisungen vertraut (für Windows / Mac OS / Linux).
 - Auf Windows oder Mac OS müsstest du dir erst ein Installationsprogramm herunterladen (*.exe- bzw. *.dmg-Datei) und dieses dann ausführen, um die Installation durchzuführen.
 - Auf Linux kannst du die Installation am einfachsten über deinen Paketmanager durchführen.
3. Starte nun RStudio. Siehst du jetzt ein Fenster, das so aussieht wie der [Screenshot auf Folie 16 der Vorlesung vom 17.11.23](#)? Dann hast du alles richtig gemacht!

Aufgabe B: „Hallo Welt!“

1. Erstelle ein neues R-Skript mit dem Dateinamen `abgabe.R` und speichere es irgendwo auf deinem PC.
2. Deine Lösungen zu allen Aufgaben B–F, die im Folgenden im Rahmen dieses Übungsblatts bearbeitet werden müssen, sollen jetzt einfach im selben R-Skript (`abgabe.R`) erledigt werden!
3. Schreibe im R-Skript `abgabe.R` nun einen Befehl auf, der zum Ziel hat, die Zeichenfolge „Hallo Welt!“ auszuwerfen.
4. Führe die Zeile mit diesem Befehl nun aus (z. B. durch Markieren und anschließendes Betätigen der Tastenkombination STRG + ENTER). Vergewissere dich, dass der Befehl tatsächlich das tut, was gewünscht ist, und nicht etwa stattdessen irgendeine Fehlermeldung ausgegeben wird.

Aufgabe C: Zahlenrätsel

1. Erweitere dein R-Skript `abgabe.R` nun folgendermaßen: Definiere eine Variable namens `x`, indem du ihr zunächst irgendeinen Zahlenwert deiner Wahl zuweist. Führe die Code-Zeile aus, in der diese Zuweisung vorgenommen wird.
2. Definiere anschließend darunter eine Variable namens `y`, indem du ihr den Zahlenwert $\frac{1}{2} \cdot x$ zuweist (also die Hälfte von `x`). Führe diese Zeile ebenfalls aus.
3. Nun sollst du darunter neun verschiedene Variablen `a`, `b`, ..., `h`, `i` definieren, denen Wahrheitswerte zugewiesen werden sollen. Die Wahrheitswerte sollen jeweils folgende Aussagen repräsentieren:
 - `a`: y^2 ist gleich `y`.
 - `b`: $y \cdot 9$ ist ungleich 0.
 - `c`: $y + 9$ ist ungleich 10.
 - `d`: `y` ist (ohne Rest) durch 7 teilbar.

- **e:** y ist größer als 20.
 - **f:** y ist kleiner als oder gleich 25.
 - **g:** Die Aussagen **a**, **b** und **c** sind wahr.
 - **h:** Die Aussagen **d**, **e** und **f** sind wahr.
 - **i:** Die Aussage **g** ist wahr oder die Aussage **h** ist wahr.
4. Führe alle oben in Schritt 3 genannten Variablenzuweisungen aus. Schau nun in der Umgebungsübersicht (rechts oben im RStudio-Fenster) nach, welchen Wahrheitswert die Variable **i** schließlich hat. Ist es **TRUE** oder **FALSE**?

Wenn es **FALSE** ist, dann ändere bitte den in Schritt 1 der Variable **x** zugewiesenen Zahlenwert zu einer anderen Zahl und führe den Code ab dieser Zeile bis zum Ende erneut aus. Kannst du einen Zahlenwert für **x** finden, der den Wahrheitswert von **i** wahr (d. h. **TRUE**) macht?

Hinweis: Es gibt nur eine einzige mögliche Zahl für **x**, die **i** wahr machen kann.

Aufgabe D: Pakete installieren und laden

1. Installiere das Paket **dplyr** durch Anwendung des `install.packages(...)`-Befehls.
2. Installiere das Paket **ggplot2** durch Anwendung des `install.packages(...)`-Befehls.
3. Lade nun die Funktionalitäten des Pakets **dplyr** mithilfe des `library(...)`-Befehls.
4. Lade ebenfalls die Funktionalitäten des Pakets **ggplot2** mithilfe des `library(...)`-Befehls.

Aufgabe E: 11 Städte der EU

1. Lade dir die Tabellendatei **11-staedte.csv** aus dem Moodle herunter und speichere sie auf deinem PC in irgendeinem Verzeichnis (d. h. Ordner) deiner Wahl ab.
2. (Weiterhin in **abgabe.R**.) Nutze den `setwd(...)`-Befehl, um das Verzeichnis, in dem du die Datei **11-staedte.csv** abgespeichert hast, als aktuelles Arbeitsverzeichnis von R festzulegen.
3. Vergewissere dich nun, indem du die Befehle `getwd()` und anschließend `dir()` ausführst und dir den Output in der Konsole anschaust, dass du tatsächlich das gewünschte Arbeitsverzeichnis festgelegt hast und sich darin auch wirklich die Datei **11-staedte.csv** befindet.
4. Definiere eine Variable namens **data**, indem du ihr einen Datenrahmen zuweist, der die Daten aus der Tabellendatei **11-staedte.csv** enthält.
Hinweis: Du wirst `read.csv(...)` brauchen, um die Daten aus der Tabellendatei zu importieren.
5. Lass dir den Inhalt des Datenrahmens **data** anzeigen, indem du den `View(...)`-Befehl anwendest.
6. Berechne nun zunächst komplett manuell, also unter Verwendung der dir bekannten arithmetischen Operatoren in R, den Durchschnitt der Spalte „**Fläche..Quadratkilometer.**“, indem du erst alle 11 gegebenen Werte abschreibst und aufaddierst und ihre Summe dann durch 11 teilst.
Weise das Ergebnis einer neuen Variable namens **durchschnitt_km2** zu.
7. Stelle durch Nachdenken fest, was der Median der Spalte „**Fläche..Quadratkilometer.**“ ist.
Weise deinen ermittelten Medianwert dann einer neuen Variable namens **median_km2** zu.
Hinweis: Falls du dir nicht sicher bist, was der Unterschied zwischen den Konzepten Durchschnitt und Median ist, kannst du es hier noch einmal nachlesen: prlbr.de/2016/durchschnitt-median-modus
8. Lass dir durch Anwendung des `summary(...)`-Befehls eine Zusammenfassung der Daten aus dem Datenrahmen **data** ausgeben. Vergewissere dich nun, dass die in der Zusammenfassung angezeigten Werte „**Mean** :“ (Durchschnitt) und „**Median** :“ für die Spalte „**Fläche..Quadratkilometer.**“ mit den von dir manuell bestimmten Werten aus den obigen Aufgabenschritten 6 und 7 übereinstimmen.

9. Führe folgenden Befehl aus:

```
data <- data %>% rename(city = Name,  
                        residents = Einwohner..in.Millionen.,  
                        km2 = Fläche..Quadratkilometer.)
```

Dies hat zur Folge, dass im Datenrahmen `data` die Spalten „Name“, „Einwohner..in.Millionen.“ und „Fläche..Quadratkilometer.“ umbenannt werden zu jeweils „city“, „residents“ und „km2“.

Hinweis: Die Ausführung dieses Befehls wird nur dann gelingen, wenn du zuvor (s. Aufgabe D) bereits erfolgreich das Paket `dplyr` installiert sowie geladen hast.

10. Führe zuletzt noch folgenden Befehl aus:

```
ggplot(data, aes(x = km2, y = residents)) + geom_point()
```

Dadurch wird eine Grafik erstellt, die die Beziehung der beiden Wertespalten „km2“ und „residents“ in einem Streudiagramm darstellt. Diese sollte dir dann standardmäßig unten rechts in RStudio, nämlich im Grafik-Ausgabebereich, angezeigt werden.

Hinweis: Die Ausführung dieses Befehls wird nur dann gelingen, wenn du zuvor (s. Aufgabe D) bereits erfolgreich das Paket `ggplot2` installiert sowie geladen hast.

Abgabe:

Lade bis zum 30.11.23 um 23:59 Uhr folgendes R-Skript in der Abgabemaske im Kurs-Moodle hoch:

- `abgabe.R` (nachdem du alle Aufgaben B–E erkennbar darin erledigt hast)