

Übungseinheit 1 - Grundlagen

Übung 1: Vorbereitendes

- ☐ Machen Sie sich mit der RStudio-Benutzeroberfläche vertraut. Eine recht umfassende Übersicht über die Benutzeroberfläche und die Tastaturkürzel finden Sie hier: <https://rstudio.github.io/cheatsheets/rstudio-ide.pdf>
- ☐ Probieren Sie in der R-Konsole verschiedene Operatoren in R aus (+, -, *, /, :, ^, <, >, ==, >=, <=, !=) sowie die Taschenrechnerfunktionen.
- ☐ Erstellen Sie ein neues Verzeichnis mit dem Namen **GDSR_1** und definieren Sie dieses Verzeichnis als Ihr Arbeitsverzeichnis für heute.
- ☐ Laden Sie den Datensatz **WestRoxbury.csv** von GitHub herunter und speichern Sie den Datensatz in Ihrem Arbeitsverzeichnis.
- ☐ Erstellen Sie ein neues R-Skript mit dem Namen **Aufgabenblatt_1.R** und speichern Sie dieses ebenfalls in Ihrem Arbeitsverzeichnis ab.
- ☐ Ergänzen Sie in Ihrem R-Skript folgende erste Zeile (beachten Sie, dass Sie Kommentare mit # zu Ihrem Skript hinzufügen können): **rm(list=ls()) # Entfernt alle Objekte (Datensätze, Variablen usw.) aus dem Speicher**
- ☐ Lesen Sie den Datensatz **WestRoxbury.csv** über das Menü von RStudio ein und vergeben Sie für diesen DataFrame den Namen **housing_df**. Vergessen Sie nicht den generierten Code in Ihr R-Skript zu kopieren.
- ☐ Betrachten Sie die eingelesenen Daten mit dem Befehl **View()**.
- ☐ Speichern Sie Ihr R-Skript und führen Sie Ihr gesamtes Skript noch einmal aus und prüfen Sie ob Ihr Vorgehen replizierbar ist.

Übung 2: Analyse

- ☐ Wählen Sie die ersten 10 Zeilen der Daten aus. Überprüfen Sie im Variablenviewer ob Sie die richtigen Zeilen ausgewählt haben.
- ☐ Extrahieren Sie den Wert der zweiten Wohnung im Datensatz aus den Daten.
- ☐ Wählen Sie die erste, vierte und letzte Spalte aus und weisen Sie das Ergebnis einem neuen DataFrame mit dem Namen **housing_reduced_df** zu.
- ☐ Suchen Sie mit **?table** die notwendige Syntax für den Befehl und erstellen Sie eine Häufigkeitstabelle der Spalte **REMODEL** aus dem DataFrame **housing_reduced_df**.
- ☐ Räumen Sie Ihren Speicher auf und löschen Sie den DataFrame **housing_reduced_df** aus dem Speicher.
- ☐ Nutzen Sie die Funktion **dim()** um die Anzahl an Beobachtungen und Variablen des DataFrame **housing_df** zu bestimmen. Können Sie mit **length()** die gleichen Informationen erhalten?
- ☐ Berechnen Sie den Mittelwert und die Varianz der Variable **Total.Value** aus dem **housing_df**.
- ☐ Untersuchen Sie grafisch den Zusammenhang zwischen **TOTAL.VALUE** und **LIVING.AREA**. Tragen Sie **LIVING.AREA** auf der X-Achse ab und achten Sie auf eine korrekte Beschriftung der beiden Achsen. Können Sie der Abbildung auch einen Titel geben?
- ☐ Speichern Sie die Abbildung als **Abbildung1.pdf** im Arbeitsverzeichnis ab.
- ☐ Untersuchen Sie grafisch den Zusammenhang zwischen **TOTAL.VALUE** und **REMODEL**. Beachten Sie den Variablentyp von **REMODEL**. Tragen Sie **REMODEL** auf der X-Achse ab und achten Sie auf eine korrekte Beschriftung der beiden Achsen.
- ☐ Lassen Sie sich eine numerische Zusammenfassung des Datensatzes ausgeben.
- ☐ Speichern Sie das R-Skript ab und prüfen Sie ob Ihre Ergebnisse replizierbar sind.