

Problem Set 1 - Grundlagen

- Laden Sie den Datensatz **ToyotaCorolla.csv** von GitHub herunter und speichern Sie den Datensatz in Ihrem Arbeitsverzeichnis.
- Erstellen Sie ein neues R-Skript mit dem Namen **ProblemSet_1.R**
- Ergänzen Sie in ihrem R-Skript folgende erste Zeile: **rm(list=ls())**
- Lesen Sie den Datensatz **ToyotaCorolla.csv** über das Menü von RStudio ein und vergeben Sie für diesen DataFrame den Namen **cars_df**. Vergessen Sie nicht den generierten Code in Ihr R-Skript zu kopieren.
- Lassen Sie sich die ersten Zeilen des Datensatzes anzeigen.
- Wählen Sie die ersten 2 Spalten der Daten aus.
- Wählen Sie die Zeilen 7-12 der Daten aus.
- Wählen Sie die erste, dritte, siebte und achte Spalte aus und weisen Sie das Ergebnis einem neuen DataFrame mit dem Namen **cars_reduced_df** zu.
- Nutzen Sie die Funktion **dim()** um die Anzahl an Beobachtungen und Variablen des neuen DataFrame **cars_reduced_df** zu bestimmen.
- Entfernen Sie alle Beobachtungen mit fehlenden Werten aus dem neuen Datensatz. Prüfen Sie die Anzahl der Beobachtungen nochmals. Wie viele fehlende Werte waren vorhanden?
- Löschen Sie den originalen Datensatz **cars_df** aus dem Speicher.
- Erstellen Sie eine Häufigkeitstabelle der Variable **Fuel_Type** aus dem DataFrame **cars_reduced_df**.
- Berechnen Sie den Mittelwert und die Varianz der Variablen **Price** und **KM** aus dem **cars_reduced_df**.
- Untersuchen Sie grafisch den Zusammenhang zwischen **Price** und **KM**. Tragen Sie **KM** auf der X-Achse ab und wählen Sie eine korrekte Beschriftung der beiden Achsen. Geben Sie der Abbildung einen sinnvollen Titel.
- Speichern Sie die Abbildung als **ProblemSet1.pdf** im Arbeitsverzeichnis ab.
- Wählen Sie eine geeignete Abbildung, um die Frage zu beantworten, ob sich die Autos je nach Kraftstofftyp im Preis unterscheiden.
- Lassen Sie sich eine numerische Zusammenfassung des Datensatzes ausgeben.
- Speichern Sie das R-Skript ab, prüfen Sie ob Ihre Ergebnisse replizierbar sind.