

## Hausaufgabe II: Grundlagen ggplot2

1. Laden Sie den Datensatz **ToyotaCorolla.csv** entweder von GitHub oder StudOn herunter (das ist der Datensatz aus Hausaufgabe 1) und speichern Sie den Datensatz in Ihrem Arbeitsverzeichnis.
2. Erstellen Sie ein neues R-Skript mit dem Namen **Hausaufgabe\_2.R**
3. Ergänzen Sie in ihrem R-Skript folgende erste Zeile: `rm(list=ls())`
4. 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.
5. Betrachten Sie den Datensatz mit dem Befehl **View(...)**.
6. Laden Sie die beiden Pakete **tidyverse** und **ggthemes** via des Befehls **library(Paketname)**.
7. Prüfen Sie, ob der Datensatz Beobachtungen mit fehlenden Informationen hat und korrigieren Sie dies bei Bedarf.
8. Nutzen Sie die Funktion **dim( )**, um die Anzahl an Beobachtungen und Variablen des Datensatzes zu bestimmen.
9. Speichern Sie Ihr R-Skript und führen Sie Ihr gesamtes Skript noch einmal aus und prüfen Sie, ob Ihr Vorgehen replizierbar ist.

Sie interessieren sich für den Zusammenhang zwischen dem Preis eines Toyotas (**Price**) und der Anzahl der gefahrenen Kilometer (**KM**). Geben Sie bitte den Code für jeden der folgenden Schritte an.

10. Berechnen Sie den Mittelwert für beide Variablen und speichern Sie die beiden Mittelwerte in zwei R-Objekten mit dem Namen **Variablenname\_mw** ab.
11. Nutzen Sie die Funktion **ggplot( )** um ein Plot Objekt zu definieren, das auf dem Datensatz **cars\_df** basiert und auf der X-Achse **KM** darstellt und **Price** auf der Y-Achse.
12. Ergänzen Sie nun ein layer (**geom**), das ein Streudiagramm der beiden Variablen darstellt.
13. Färben Sie nun die Datenpunkte so ein, dass die unterschiedlichen Arten von Kraftstoff (**Fuel\_Type**) erkennbar sind.

14. Nutzen Sie `geom_smooth()` um eine gemeinsame Trendlinie (für alle Kraftstoffarten) hinzuzufügen. Sie wollen eine lineare Regressionsfunktion haben (nicht den Standard). Nutzen Sie die Hilfefunktion `?geom_smooth()` um herauszufinden, wie Sie diese erhalten können.
15. Nutzen Sie nun das **shape**-Attribut, um die Kraftstoffarten auch in der Form unterscheidbar zu machen.
16. Nutzen Sie **labs()**, um Ihrer Abbildung einen Titel, einen Untertitel sowie sinnvolle Achsenbeschriftungen zu geben. Benennen Sie auch den Titel der Datenlegende korrekt.
17. Nutzen Sie die geoms: `geom_hline()` und `geom_vline()` um eine horizontale und eine vertikale Linie bei den oben berechneten Mittelwerten einzuzeichnen. Färben Sie die Linien **blau** ein und wählen Sie eine **gepunktete Linie**.
18. Ändern Sie die Farbpalette in der Abbildung auf `scale_color_colorblind()` und formatieren Sie Ihre Abbildung mit dem Theme: `theme_bw()`.
19. Erzeugen Sie eine zweite Abbildung, in der der Zusammenhang zwischen Price und KM nach Kraftstoffarten getrennt angezeigt wird.
20. Speichern Sie das R-Skript ab und prüfen Sie, ob Ihre Ergebnisse replizierbar sind.