

SOC 4930/5050: Lab-02 - Initial Data Cleaning

Christopher Prener, Ph.D.

September 4th, 2017

Directions

Complete all of the following questions using the data from the testDriveR package. To install testDriveR, see instructions in the Week 02 jotter. Your well-formatted R Notebook source (the .Rmd file) and html output should be uploaded to your assignments repository by 4:15PM on Monday, September 18th, 2017.

Part 1: Cleaning Data

Use the auto17 data frame saved in the testDriveR package and make the following changes using “piped” code:

1. Extract observations for German cars (those manufactured by Audi, BMW, Mercedes-Benz, Porsche, and Volkswagen)
2. Keep only the following variables: id, mfrDivision, carLine, combFE, guzzlerStr, displ
3. Rename the mfrDivision and combFE variables¹
4. Create a new logical variable that is TRUE if the vehicle is a guzzler (guzzlerStr == "G") and is FALSE otherwise
5. Re-order the data frame based on your re-named combFE variable from high to low
6. Print the “head” of the data frame - what is the most fuel efficient German car for sale in the United States for model year 2017?
7. How many German cars in total are for sale in the United States for model year 2017?
8. How many German cars are “gas guzzlers”?

¹ Not sure what these variables measure? Type ?auto17 into the console of RStudio and scroll through the help file for the data set.

Part 2: Plotting Data

Use the your cleaned German car data to produce the following plots:.

9. Create a bar plot of the logical “gas guzzler” variable you created

10. Create a histogram of the average fuel efficiency variable
11. Create a scatter plot of the average fuel efficiency and displ variables that (a) highlights “gas guzzler” vehicles and (b) uses the “jitter” positions adjustment