Lab 8: RStudio Review

Introduction

In this lab you will review the different techniques we have discussed for analyzing data in RStudio. The dataset contains information about various 2015 car models.

Start Project and Get Data

- Download the file at http://www.lock5stat.com/datasets/Cars2015.xlsx
- Start a new project in RStudio (Lab_8).
- Upload the Excel file to the Lab_8 project (**Files** tab at lower right, **upload** button)
- Import the data from Excel to RStudio (click the file in the Files tab, or go to the **Environment** tab at upper-right and click **import dataset** button).
- Start a new RMarkdown document for your lab report.
- Add **two chunks** to your RMarkdown, one to load the hanoverbase library and one which includes the two lines needed to import the data from Excel.
- Make sure you can view the Cars2015 data in the data viewer.
- Save and knit. Remember to save and knit often as you work on your report.

Understanding the Variables in the Cars Dataset

If you find yourself unsure of, for example, the units used for the weight of a car, you should do a Google search on one of the car makes and models, like "Hyundai Accent", to find its weight. That information should help you determine proper units for the weight variable in the dataset.

Statistical Investigations

Provide **relevant graphs**, **numerical summaries**, and **discussions** to answer the following questions. A graph cannot speak for itself; you need to provide a discussion of what you see in the graph.

As you investigate the data, you should create **at least one of each** of the following types of graphs (you can also use other kinds of graphs, of course):

- piechart
- simple barchart
- labeled dotplot (optional)
- histogram
- boxplot
- stacked barchart
- simple scatterplot with smooth fit line
- scatterplot with panels

Make sure each graph has a main title and suitable axis labels.

For help with writing your R code chunks, refer to the R and RStudio Cheatsheet and to the previous lab assignments: https://hanoverstatslabs.github.io/resources/

Questions

When looking at a new dataset, the first thing we must do is understand each of the variables **individually**. We will only look at a subset of the variables here.

1. Investigate *individually* the distributions of the car type and size for this dataset.
2. Investigate *individually* the distributions of city mileage, highway mileage, acceleration time from 0 to 60, and weight. Make sure to identify the outliers in the data (e.g. by make and model).

Once we have looked at individual variables, we would typically proceed to look at relationships between two or possibly three variables. The next set of questions requires this kind of analysis.

3. Investigate the relationship between type and size.
4. Investigate how highway mileage and city mileage are related, and also how they are related to type and size.
5. What effect, if any, does the car type and weight have on the car's time to accelerate from 0 to 60 mph? Does the effect of weight on acceleration differ across the different car types? Which car type tends to have the fastest acceleration?

Submission

- Make sure to knit one last time, then download the Lab8Report.Rmd file: in the **Files pane** (lower right), click the checkbox for the RMD file, then choose More > Export ... > Download. *Do the same for Lab8Report.html.*
- **Submit both files via Moodle**.