**DS 710**

**R Programming Assignment**

**Assignment 3**

1. **Analyzing Used Car Prices**
2. Download Cars 2005.csv, load the data into R, and attach it.

(Dataset: “Car Data," submitted by Shonda Kuiper, Grinnell College. Dataset obtained from the Journal of Statistics Education (http://www.amstat.org/publications/jse). Accessed 3 June 2015. Used by permission of author.)

1. Make a histogram of the prices of cars in the data.  Describe the shape of the distribution.
2. What proportion of cars in the data set cost between $10,000 and $20,000?
3. Find the mean and median price.  Which is larger?  Why does this make sense?
4. Add a vertical line to the histogram to denote the mean price.  Add a legend to the graph.
5. Transform the price to reduce its skew, and make a histogram of the transformed price.  Fit a normal distribution to the transformed price, and graph the normal density curve on the same plot as the histogram.  How well does a normal distribution fit the transformed data?
6. Make a scatterplot of transformed price versus engine size, measured in liters.  Describe the relationship between these two variables.
7. Find the correlation between transformed price and engine size in liters.  Explain what it tells us.
8. Modify your scatterplot in part g to use one color of plotting symbol for cars with leather interiors, and a different color for cars without leather interiors.  Add a legend to your plot.
9. Make a barplot of the types (Sedan, Hatchback, etc.) of cars in the data.
10. Make a barplot of the types of cars and whether they have leather interiors.  Add a legend to your plot.
11. Make a boxplot of (untransformed) price by type of car.  In words, summarize what it shows.
12. Create two different histograms in a vertical stack that allow comparison of (untransformed) price according to whether the car has a leather interior.  Use the same horizontal axis for each to enable comparison, and use informative labels for each graph and the x-axis.
13. Create a single histogram with side-by-side bars to allow the same comparison as in part m.  Add a legend to your plot.
14. **Analyzing the running speed of mammals**
15. In R, type

install.packages("quantreg")

data(Mammals, package="quantreg")

This will load a data set called Mammals, on the maximum land speed of various species of mammal.  Attach the data and look at the first few lines.

(Source:  Garland, T. (1983) The relation between maximal running speed and body mass in terrestrial mammals, *J. Zoology*, 199, 1557-1570.

Metadata:  <http://vincentarelbundock.github.io/Rdatasets/doc/quantreg/Mammals.html>, accessed 7 June 2015.)

b. Decide whether either of the quantitative variables should be transformed.  Justify your decision using appropriate plots and/or descriptive statistics.

c.  Use appropriate graphs and/or descriptive statistics to describe the relationship between maximum land speed and body weight.  Does it matter whether the animal is a “hopper” (such as a kangaroo)?  Explain why you chose the graphs and/or statistics that you chose.

Submit a **single** .docx or .pdf document containing your R code, output and graphs, and interpretations (where requested).  To facilitate a quick turnaround on grading, please keep all parts of a problem together, rather than putting code at the end of the file.

* Please include your full name at the top of your document.
* It is not necessary to include R code where you were testing code or where you made a mistake--just submit the final version.