**Chapter 2 – Examining Relationships Assignment #2**

This assignment is to be undertaken individually and is due at the start of class on **Tuesday September 26th 2017.** Any late submissions will only be accepted up until Tuesday October 3rd with a 50% grade reduction, no assignments will be accepted after this date.

Your submission should be a hardcopy report containing the answers to all questions presented in the same order they appear in the assignment description. **Your answers must be clear, concise and supported by attached output from *R* (include code, outputs and clearly labelled plots)*.***

The following questions refer to data on the fuel efficiency of vehicles sold in Canada in 2014, the dataset is provided on Blackboard in csv format (CanFuel.csv). The data is sourced from a study undertaken by Natural Resources Canada (http://www.nrcan.gc.ca/energy/efficiency/11938).

1. For each of the following 3 variables: City Fuel Efficiency (CityMPG), Highway Fuel Efficiency (HwyMPG) and Carbon Dioxide Emissions (CO2):

* construct a histogram
* calculate descriptive statistics (use the **describe** command from the **psych** package)
* construct a boxplot
* calculate the five number summary (min, 1st quartile, median, 3rd quartile, max)

Write a paragraph about each of the 3 variables, for each variable discuss its distribution including: shape, measures of spread, measures of center and outliers or unusual values.

2. Next you will explore the relationships between the 3 variables you looked at in Question 1:

* CityMPG and HwyMPG
* CityMPG and CO2
* HwyMPG and CO2

For each of the 3 pairs of variables listed above:

* construct a scatterplot
* calculate the correlation coefficient “r”

Write a paragraph about each of the 3 pairs of variables, for each pair discuss the relationship including: form, direction and strength. Also discuss any outliers or unusual values.

3. Use least squares regression to fit a linear model that has HwyMPG as the Response Variable and CityMPG as the Explanatory Variable. For your linear model:

* report the slope and the y-intercept
* report the coefficient of determination “*r2*” and describe what it tells you
* construct a scatterplot that includes the regression line

Then repeat what you did above but this time fit a linear model that has CO2 as the Response Variable and CityMPG as the Explanatory Variable.

4. To explore the distribution of the residuals for your linear model of HwyMPG vs CityMPG:

* construct a residual plot
* construct a Normal quantile plot for residuals

Discuss whether you believe a linear (straight line) model is a good fit for this data, if not what form of model might fit better? In answering this refer to the two plots above and the scatterplot with the regression line from Question 3.

Then repeat what you did above for your linear model of CO2 vs CityMPG.

5. Use your first linear model to predict the HwyMPG for a vehicle that has:

* CityMPG = 25
* CityMPG = 75

Use your second linear model to predict the CO2 for a vehicle that has:

* CityMPG = 25
* CityMPG = 75

As well as reporting your four predictions, discuss how confident you are in each prediction and why, in answering this refer to the: scatterplot with the regression line, residual plot and Normal quantile plot for residuals.