STOR 455: Assignment #1 Due: Wednesday, 9/5/18
McLean: 8/29/2018 Value: 20 points

Collaboration rules: For this assignment you may collaborate with (or seek assistance from) other students. If you do so you must identify them on the work you turn in. Also, note that it is important for you to get comfortable using the technology options for this course, so don't rely on others to do all your computing for you. You should produce the final solutions you turn in.

Format: Use an R Notebook for all calculations and plots. You will submit the html file to Sakai. Graphical output should always be interpreted with at least a sentence or two.

Data: You should find the dataset for this assignment in the Stat2Data package

1. The text dataset **HighPeaks** contains information on hikes for the 46 highest mountains in the Adirondacks (You can become a 46er by climbing them all!). Consider a model to predict the *Time* for each hike (in hours) based on the *Ascent* (vertical distance in feet from the starting point).

    a. Find the equation of the least squares line for this model (include some R output and write down the prediction equation).

    b. Produce a scatterplot of this relationship (and include the least squares line on the plot).

    c. What time does the model predict for climbing Mt. Marcy? What is the residual for that case?

    d. Which mountain has the largest residual (in absolute value)? What is the value of that residual?

    e. Comment how the conditions for a simple linear model look for this model. Include at least two plots (in addition to the plot in part 2) – with commentary on what each plot tells you specifically about the appropriateness of conditions.

2. The text dataset **Perch** contains information on the *Weight* (in grams), *Length* (in cm.), and *Width* (in cm.) for a sample of 56 fish (perch) caught in a lake in Finland. We are interested in using the *Length* of the fish to predict *Weight*.

    a. Use a plot to show why the linearity condition is clearly not met and a transformation is needed.

    b. Experiment with some transformations until you find one that seems to do a better job of satisfying the linearity condition. Include the summary output for fitting that model and a scatterplot of the transformed variable(s) with the least square line.

    c. While the model in (b) should show better linearity than the one in (a), you might still have concerns about linearity or some other model conditions for the transformed variables. Do you? Include appropriate plots to support your findings.

    d. Suppose that I went fishing in this lake while on his sabbatical and caught a perch that was 30 cm. long but didn't have anything to weigh the fish with. How much do you think it weighed? Show a calculation to support your answer.