STOR 455: Assignment #2 Due: Fri 9/14/18
Value: 20 Points

Correlation & Simple Linear Regression
Applied to a Virtual Used Car Lot

SITUATION:
Suppose that you are interested in purchasing a used car. How much should you expect to pay? Obviously the price will depend on the type of car you get (the model) and how much it's been used. For this project you will investigate how the **price** might depend on the **age** (in years with 2017=1 year old).

DATA SOURCE:
To get a sample of cars, use the UsedCarLot CSV file on Sakai, which contains car prices obtained from *autotrader.com.* Choose a car make and model and then create a new dataset containing only that type of car. The dataset should have columns for **year**, **price** (in $1,000's) and **mileage** (in 1,000's). You should add a variable called **age** which is **`2018-year`**.

REPORT:

1. Calculate the least squares regression line that best fits your data. Interpret (in context) what the slope estimate tells you about prices and ages of your used car model. Explain why the sign (positive/negative) make sense.
2. Produce a scatterplot of the relationship with the regression line drawn on it.
3. Produce appropriate residual plots and comment on how well your data appear to fit the conditions for a simple linear model. Don't worry about doing transformations at this point if there are problems with the conditions.
4. Find the car in your sample with the largest residual (in magnitude – positive or negative). For that car, find its standardized and studentized residual. Would this value be considered unusual?
5. Compute and interpret in context a 90% confidence interval for the slope of your model.
6. Test the strength of the linear relationship between your variables using each of the three methods.
   - $\Rightarrow$ test for correlation
   - $\Rightarrow$ test for slope
   - $\Rightarrow$ ANOVA for regression
7. Suppose that you are interested in purchasing a car of this model that is five years old (**`age=5`**). Calculate each of the following quantities. Write sentences that carefully interpret each of the intervals (in terms of car prices).
   - $\Rightarrow$ predicted value for the **price**
   - $\Rightarrow$ 90% confidence interval for the mean **price** at this **age**.
   - $\Rightarrow$ 90% prediction interval for the **price** of an individual car at this **age**
8. According to your model, is there an age at which the car should be free? If so, find this age and comment on what the "free car" phenomenon says about the appropriateness of your model.
9. Experiment with some transformations to attempt to 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 squares line. Explain why you think that this transformation does or does not improve satisfying the linearity condition.