DS-510: Project 1
Assigned Nov. 8, 2016
Due Nov. 22, 2016

In this project you will investigate the impact of a number of automobile engine factors on the vehicle's mpg. The dataset *auto-mpg.data* contains information for 398 different automobile models. Information regarding the number of cylinders, displacement, horsepower, weight, acceleration, model year, origin, and car name as well as mpg are contained in the file.

1. Using the first 300 samples in the *auto-mpg.data*, run a simple linear regression to determine the relationship between mpg and a single, appropriate dependent variable of your choice. Report all appropriate information regarding your regression and include explanatory graphs.
   In addition, include plots for the following:

   (a) residuals vs. the predictor variable

   (b) absolute value of the residuals vs. the predictor variable

   (c) histogram of the residuals

   Examine the above plots and discuss how you might interpret these plots to determine the appropriateness of your linear model.
   For the remaining 98 samples in the dataset, use your linear model to predict each automobile's mpg and report how your predictions compare to the car's actual reported mgp.

2. Repeat 1., this time using more than one independent variable. Try to construct a methodology for determining which independent factors should be included to obtain the "best" linear model for mpg. Note that you will have to modify your approach to the examination of the residuals as there should now be more than one independent variable in your model.