Homework 1 (Due: Tuesday, January 12th )

1. This exercise involves \textit{cdc} data set studied in class. Answer the following questions:

   (a) Make a scatterplot of weight versus desired weight. Describe the relationship between these two variables.

   (b) Let’s consider a new variable: the difference between desired weight (wtdesire) and current weight (weight). Create this new variable by subtracting the two columns in the data frame and assigning them to a new object called wdiff.

   (c) What type of data is wdiff? If an observation wdiff is 0, what does this mean about the person’s weight and desired weight. What if wdiff is positive or negative?

   (d) Describe the distribution of wdiff in terms of its center, shape, and spread, including any plots you use. What does this tell us about how people feel about their current weight?

   (e) Using numerical summaries and a side-by-side box plot, determine if men tend to view their weight differently than women.

   (f) Now it’s time to get creative. Find the mean and standard deviation of weight and determine what proportion of the weights are within one standard deviation of the mean.

2. This exercise involves the \textit{Auto} data set available in the course webpage. Make sure that the missing values have been removed from the data.

   (a) Which of the predictors are quantitative, and which are qualitative?

   (b) What is the range of each quantitative predictor? You can answer this using the range() function.

   (c) What is the mean and standard deviation of each quantitative predictor?

   (d) Now remove the 10th through 85th observations. What is the range, mean, and standard deviation of each predictor in the subset of the data that remains?

   (e) Using the full data set, investigate the predictors graphically, using scatterplots or other tools of your choice. Create some plots highlighting the relationships among the predictors. Comment on your findings.

   (f) Suppose that we wish to predict gas mileage (mpg) on the basis of the other variables. Do your plots suggest that any of the other variables might be useful in predicting mpg? Justify your answer.