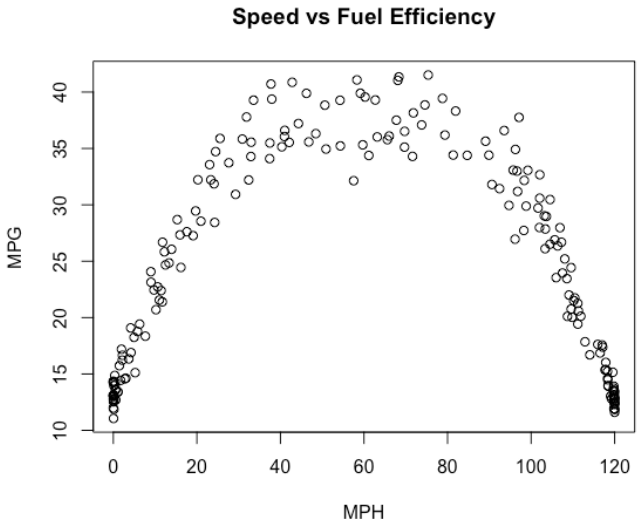
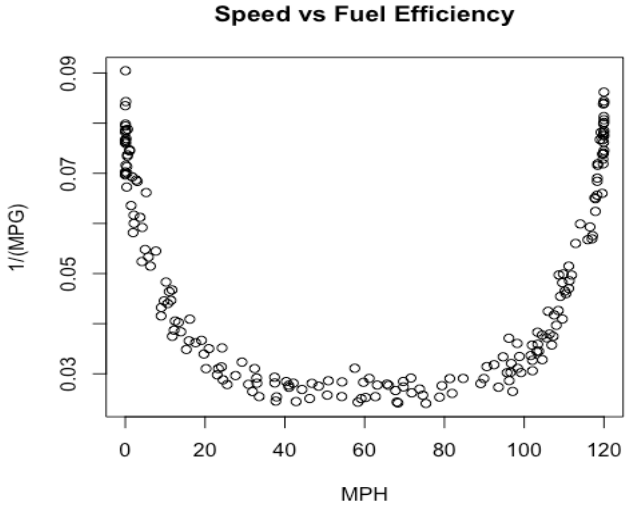
Luke Beebe
STAT 463 – Assignment 4 [code is in R file]

To start the project, I loaded the mpg_data txt file as a table and analyzed its plot.



1) There appears to be a strong nonlinear relationship between the variables.
2) It looks like the variance is not constant throughout the data. At the center of the plot, the height of the curve, the variance increases. It is heteroscedastic. The transformation I applied was making the MPG.t=MPG^(-1). This produced the next plot on the right. I tinkered with it being MPG.t=MPG^(-2) which produced a more homoscedastic plot, but it squished the variables so much towards zero that when I transformed it back, the confidence bands seemed to overcompensate the deviations in the data. That's why I chose MPG.t=MPG^(-1).

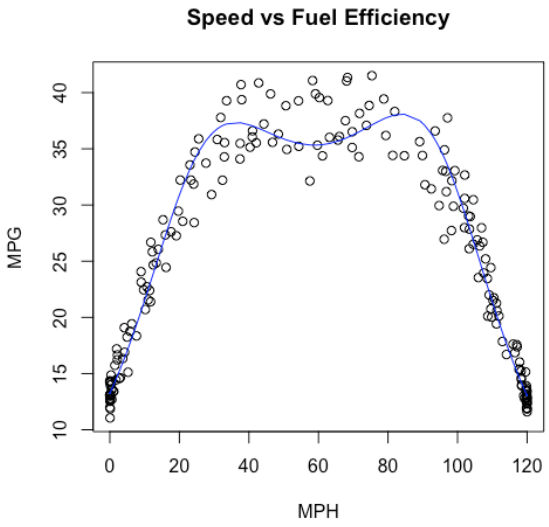


3) While fitting a polynomial model to our data, I went from a second order equation to a fifth order equation. Each model had a p-value of 2.2e-16, which meant there was a significant relationship between the variables given the model. However, two of the models had a R^2 value of 96.54% which was more accurate than the other two. Between these two models, the fourth order equation used all of the variables, while the fifth order equation's fifth order variable was not statistically significant. So, I chose the fourth order equation model to describe the data. This was the equation I came up with:

MPG = 1/MPG.t = -3.872(10^-3)*(MPH)+1.094(10^-4)*(MPH^2)-1.292(10^-6)*(MPH^3)
+5.412(10^-9)*(MPH^4)+7.551(10^-2)

There are two transformations presented above, MPH→MPG.t→MPG

4) I then overlaid the regression model on the plot to the right.



5) From the model constructed in part 3, one can conclude there is a statistically significant relationship between MPG and speed. As I stated before, the p-value of all the models was the same at 2.2e-16, which is statistically significant.
6) The correlation coefficient I got after transforming the data back is 0.9765004, which is very high.
7) According to the model, the engine is most fuel efficient at 84.7 MPG. I arrived at this by two lines of code, which found the index of the max MPG value along the regression line. Then I used that index to derive the MPH value from the regression line.
8) Below is the plot with the regression line and 90% confidence bands.