Capstone Proposal

1. PROBLEM: As fossil fuels disappear every more quickly, it becomes incumbent upon responsible citizens to care about finding greener options for all that they do, including the everyday task of driving their automobiles. It is therefore vital to discover the cars that have high mpg ratings. But with so many options available in terms of manufacturers and engine types, it is unclear how to begin the search for energy-efficient cars.
2. CLIENT: Any and all persons who drive could potentially be interested in the results of this research. If there are any patterns to be found that separate the fuel-efficient cars from the fuel-inefficient cars, surely most drivers would be interested in knowing about them.
3. DATA: UC Irvine has a dataset with information about several hundred automobiles available at this website: <http://archive.ics.uci.edu/ml/datasets/Auto+MPG>. Since it’s a .data file, it can be loaded into R with the ‘read.table()’ function.
4. APPROACH: The main idea will be to examine the data in hopes of finding some correlations between the values of some variable(s) and the MPG rankings. That is, I shall treat the MPG variable (column) as my dependent variable, and then look for ways of predicting the value of this variable given other (independent) variables (columns). The best candidates for relevant independent variables (which I can check by using the ‘lm()’ function from the “stats” package as well as plotting tools from the “ggplot2” package) will be things like the number of engine cylinders and the weight of the car. I expect to have to clean up my data a bit first, which will likely call for some tools from the “dplyr” or the “tidyr” packages. By constructing linear models and then looking at their summaries, I can get a sense of which variables have the most significance in predicting a car’s MPG rating.
5. DELIVERABLES: I expect to use an R Markdown document that can be “knitted” into a .pdf. That way I can combine coding information with methodological information as well as with plots and other visual depictions of data and results. I also expect to have some .ppt slides for the highlights of the experimental method and/or results.