To submit Homework 14, please select your dataset and turn in a draft of these tables and graphs by Monday, December 6. The grade will be completion-based, but you have to submit something!!

You may pick any dataset for your final project, provided that the dataset has:

- At least fifty cases/rows;
- A numerical response–if discrete, there should be a large range of values (20 or more).
- At least one numerical explanatory variable (you can have more if you like).
- At least one categorical explanatory variable (you can have more if you like).

I recommend looking at the sites in the "Project Datasets" handout included in the Final Project folder on Moodle–there are many datasets that are suitable for this class! You are welcome to find a dataset from a different site, or even use your own data. I recommend that you check with me before moving forward with the rest of the project so that you don't have to start over if something turns out wrong–this will be one of the questions in our regular progress checks.

Analysis of a dataset often begins with summarizing each variable in a dataset. A discussion of graphs and summary statistics for each variable is a good way to introduce your audience to the data. Your project should include:

1. Graphical displays for each variable. Points will be taken off if you use an inappropriate graph to summarize the distribution of a variable (e.g., a bar chart for a numerical variable). Graphs should have titles and axes labeled, plus legends if you include different colors.
2. Summary statistics for each variable. You may choose which summary statistics to report, but you should have a table of the summary statistics for the numerical variables and a table of the summary statistics for the categorical variables.
3. One written paragraph per variable, referencing the summary statistics and commenting on the graphs.

Note: If you have one and only one categorical variable in your dataset, you may choose to include it in various graphical displays rather than creating a separate graphical display. For example, you may choose to color code your graphs for numerical variables by male versus female observations rather than create a bar chart specifically displaying the distribution of sex. If you have questions, please ask.