Project 3 Instructions
SDS348/385
Due Thursday May 5th, 2015 by 11:59 pm on Canvas

For Project 3 you will analyze a data set of your choosing with the specific goal of answering two questions about the data set. You should address each question computationally, using R and/or python, and produce a plot for each question illustrating the trend of interest. You will then answer your questions and interpret the plot, in the context of the questions you proposed.

Each student must turn in his or her own project. If you are working with a partner, both you and your partner must turn in a final project, but these project documents may be the same. Be sure to include both partners' names in the final project.

You should write your project in a single knitted (and converted to PDF) RMarkdown file[1], organized into three main sections: Introduction, Question 1, and Question 2. You should incorporate all R code used into the main RMarkdown document, as done on previous projects. All python code should be provided as a separate iPython Notebook(s), converted to PDF. Be sure to organize your python code properly so that it is clear which code pertains to which question.

You must additionally turn in the raw RMarkdown file as well as all code and data (including your dataset!) used. Please bear in mind that **you will lose points** for any of the following:

- Printing entire dataframes in the RMarkdown file
- Code (either R or python) with no comments
- Code which produces an error message
- Missing code and/or reporting results without corresponding code
- Extraneous code which does not contribute to your final analysis or discussion
- Not turning in your dataset
    - NOTE: if your dataset is built into R or the R packages ggplot2, tidyr, dplyr, or MASS, then you do not need to turn in a file with the data. You must, however, state in your Introduction in which R package the dataset can be found.

The Introduction should be relatively short (3-5 sentences) and should contain the source of and a brief description of your dataset.

The section for each question should begin with the question you are asking, followed by the analysis. If you conducted any aspect of your analysis in python, you should indicate in which file (either an iPython notebook or python script) the code is located. Again, be sure to directly reference the attached iPython Notebook file or

[1] If using Python exclusively for analyses and plots, write your project into a single IPython Notebook. Follow the same format as you would an Rmarkdown document.

Project 3 Instructions
SDS348/385
Due Thursday May 5th, 2015 by 11:59 pm on Canvas

python script. After your analysis has been conducted, you will display your plot(s), and discuss/interpret the results. The discussion **must** accomplish the following:

- Describe your chosen analysis and provide a brief overview of your methods (3-5 sentences)
- Explicitly justify why this analysis makes sense for the question you have asked (1-3 sentences)
- Interpret your plot in the context of the question you asked (2-4 sentences)
- Answer the question you have asked