Homework 3: Making a data plan

Complete this homework as a jupyter notebook titled "LASTNAME_DataPlan.ipynb" posted on your GitHub account. Turn in your assignment by emailing a link to the notebook to timothyv@andrew.cmu.edu by no later than 3pm on Thursday Feb. 22nd, 2018.

Now that you have all identified the data sets that you will be working with for your final project, it is time to come up with a data plan. Make a new Jupyter notebook on your github account and describe the following:

a) The observations (n) and all variables (p) in your data set. What are the response variables and what are the predictor variables? What are quantitative and what are qualitative variables?
b) The data architecture (e.g., data file organization, directory structures), if applicable.
c) Types of anticipated anomalies in your data and approach for data cleansing.
d) How your cleaned data table will be organized.
e) Hypotheses you will test.
f) Approaches to visualizing critical aspects of your data that are necessary for answering your hypotheses.

Later in the semester we will add descriptions of data analysis approaches and predictions. So consider this a "work in progress" plan.