Homework 1 - Diagnostic
Due April 4, 2017

Problem 1: Data Acquisition and Analysis

This problem deals with data from the 311 system in Chicago. Your task is to download data about 311 requests from the Chciago Open Data Portal and analyze it to better understand what type of requests come from what (type of) neighborhoods.
You should use Python to do the following tasks.

1. You should download and combine data from the open data portal for the past year about the following 311 requests
    a. Graffiti Removal
    b. Vacant and Abandoned Buildings Reported
    c. Potholes
    d. Sanitation Code Complaints

2. Generate summary statistics for these requests including but not limited to number of requests of each type (and subtype within each of the types above) over time, by neighborhood, response time by the city. Please use a combination of tables and graphs to present these summary stats.

3. Based on these summary statistics, tell me 5 interesting things you learned (about Chicago and the different neighborhoods) using the 311 data.

Problem 2: Data Augmentation and APIs

All of the service requests you just analyzed have address and lat/long fields. The task is to augment that data with American Community Survey data. For each address, use one of the census APIs to get some additional data (at least 3-4 useful variables) about the block where the 311 request came from. This could include information about demographics of the block (race, income, family size, etc.). If you run into API rate limit issues, just do it on "Vacant and Abandoned Buildings Reported" and "Sanitation Code Complaints" and for the past 3 or 6 months.

Based on this augmented data, do some descriptive statistics to describe:
1. What types of blocks get "Vacant and Abandoned Buildings Reported"?
2. What types of blocks get "Sanitation Code Complaints"?
3. Does that change over time in the data you collected?
4. What is the difference in blocks that get "Vacant and Abandoned Buildings Reported" vs "Sanitation Code Complaints"?

For all of these questions, submit:
1. a link to your github repository with the code on chalk.
2. A writeup answering the questions intended for a policy audience (either as a document on chalk or a jupyter notebook in the github repo).

Problem 3:
Assume you are running the 311 call center for Chicago. You get a call from 7500 S Wolcott Ave.

A. Of the four types of requests you have data for, which request type is the most likely given the call came from 7500 S Wolcott Ave? What are the probabilities for each type of request?

B. Let's now assume that a call comes in about Graffiti Removal. Which is more likely – that the call came from Lawndale or Uptown? How much more or less likely is it to be from Lawndale versus Uptown?

C. Now assume that you don't have access to all the raw data and you know the following things:

   There are a total of 1000 calls, 600 from Englewood and 400 from Uptown. Of the 600 calls from Englewood, 100 of them are about Graffiti Removal. Of the 400 calls from Uptown, 160 are about Graffiti Removal. If a call comes about Graffiti Removal, how much more/less likely is it that the call came from Englewood versus Uptown?

You should show your work and submit the answers on chalk.