Homework 1 - Diagnostic
Due by 11:59pm April 9, 2019

The goal of this homework is to make sure that everyone is coming in with the pre-requisites for this class and can use Python to do data acquisition (from external APIs), data exploration/analysis, and can communicate the results of the analysis.

For all of these questions for problems 1, 2, and 3 submit:

1. a link to your github repository with the code on canvas.
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 1: Data Acquisition and Analysis [20 pts]

This problem deals with crime reporting data from Chicago. Your task is to download data about Crime Reports from the Chicago Open Data Portal and analyze it to better understand what type of crimes get reported in what (type of) neighborhoods.

You should use Python to do the following tasks.

1. Download reported crime data from the Chicago open data portal for 2017 and 2018.

2. Generate summary statistics for the crime reports data including but not limited to number of crimes of each type, how they change over time, and how they are different by neighborhood. Please use a combination of tables and graphs to present these summary stats.

Please put your code and writeup in a github repository and submit the link on chalk.

Problem 2: Data Augmentation and APIs [40 pts]

All of the crime data you just analyzed have a block address and lat/long fields. The task now is to augment that data with American Community Survey data. For each crime report, use one of the census APIs to get some additional data (at least 3-4 useful variables) about the block or zipcode where the crime report came from. This could include information about demographics of the block or zipcode (race, income, family size, etc.).

Based on this augmented data, provide some descriptive statistics to describe:

1. What types of blocks have reports of “Battery”?
2. What types of blocks get “Homicide”?
3. Does that change over time in the data you collected?
4. What is the difference in blocks that get “Deceptive Practice” vs “Sex Offense”?

Problem 3: Analysis and Communication [25 pts]

Based on the data you have from 1 and 2:

1. Describe how crime has changed in Chicago from 2017 to 2018?

2. One of the alderman candidates from recent elections has some crime statistics on his website: https://www.ringer4results.com/node/8

   Specifically:

   Crime is a major issue in our neighborhoods over the last four years, and the data shows just that.

   Let's break down the Chicago Police Department's report for the month leading up to July 26, 2018, compared to the same week in 2017:

   - Robberies -- up 21 percent over the same time-frame in 2017
   - Aggravated batteries -- up 136 percent
   - Burglaries -- an increase of 50 percent
   - Motor vehicle theft -- up 41 percent.

   All told, crime rose 16 percent in the same 28-day time period in just one year. But take a look at the year-to-date number and you'll see how crime has affected our local neighborhoods in a four-year period:

   - Rose 10 percent since 2017
   - Rose 22 percent since 2016

   A. Are these statistics correct?
   B. Could they be misleading or would you agree with the conclusions he's drawing? Why or why not?

   Please give these answers based on the data you have from 1 and 2 and show your code.

3. As you know, there will be a new mayor in Chicago very soon. Based on these summary statistics, provide 5 key findings to the new mayor's office about crime in Chicago and what they should focus on in order to deal with crime in Chicago.

4. What are some of the key caveats of your recommendations and limitations of the analysis that you just did?

For all of these questions for problems 1, 2, and 3 submit:

- a link to your github repository with the code on canvas.
- 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: [15 pts]
Assume you are running the 911 call center for Chicago. You get a call from someone at 2111 S Michigan Ave

A. Of the types of crimes you have data for, which crime type is the most likely given the call came from 2111 S Michigan Ave? What are the probabilities for each type of request?

B. Let's now assume that a call comes in about Theft. Which is more likely – that the call came from Garfield Park or Uptown? How much more or less likely is it to be from Garfield Park vs 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 Garfield Park and 400 from Uptown. Of the 600 calls from Garfield Park, 100 of them are about Battery. Of the 400 calls from Uptown, 160 are about Battery. If a call comes about Battery, how much more/less likely is it that the call came from Garfield Park versus Uptown?

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