Police Data Challenge Deadlines
10/20 - Declaration of Intent Form Due
11/03 - Presentation Submitted

Put your statistical and data visualization skills to work to help create safer communities. Using data available through the Police Data Initiative, you will analyze complex data sets from the Baltimore, Cincinnati and Seattle Police Departments, and recommend innovative solutions to enhance public safety. Although not required, teams may identify and utilize external data sets.

This Is Statistics: http://thisisstatistics.org/policedatachallenge/

Judging Rubric
http://thisisstatistics.org/wp-content/uploads/2017/09/PoliceDataChallengeJudgingRubric.pdf
Contest Rules & Requirements
http://thisisstatistics.org/wp-content/uploads/2017/09/Police-Data-Challenge-Rules.pdf

Data - Baltimore Calls for Service dataset (for Eastern region)
https://www.policedatainitiative.org/datasets/calls-for-service/
Data Description
http://thisisstatistics.org/wp-content/uploads/2017/09/Calls-for-Service-Data-101.pdf

STAT 4310 (VIZ WIZ) Deadlines
10/04 - Groups
10/11 - Dataset & Questions
10/25 - EDA Report Due

Datasets
Calls for Service
---External---
~~Arrests: https://data.world/baltimore/baltimore-pd-arrests~~
Employee File: https://data.world/baltimore/baltimore-police-employees
~~Use of Force: https://data.world/baltimore/baltimore-police-use-of-force~~
Police Stations: https://data.world/baltimore/baltimore-police-stations
~~Victim Based Crime: https://data.world/baltimore/baltimore-crime-data~~
Gun Offenders: https://data.baltimorecity.gov/Public-Safety/Gun-Offenders/aivj-4x23
---Includes 2017---
BPDvictims:
https://data.baltimorecity.gov/Public-Safety/BPD-Part-1-Victim-Based-Crime-Data/wsfq-mvij
BPDarrests: https://data.baltimorecity.gov/Public-Safety/BPD-Arrests/3i3v-ibrt

Questions of Interest
• Indicating the dataset (or set of datasets) you plan to use

• Indicating your question of interest (what do you hope to ask of your data)
• Indicating why you think the dataset may contain the answer
    • What are potential variables help explain your story
    • What variables are you interested in possibly conditioning on

1. ***How can Baltimore police more efficiently and effectively deal with crime?***
    a. **Are there specific neighborhoods that are more "dangerous"? That regularly have calls for service that require arrests and use of force?**
        i. Variables from Baltimore Calls Dataset: We will be able to see the multitudes of calls made to police about different instances and where they came from. This will give us locations and/or areas where we can create visualization maps that show where the most arrests are happening, the use of force is happening, and the type of incidents happening as well.
            1. *district* - to look at which districts are considered more dangerous
            2. *incidentLocation* - possible area groupings of the locations of the incidents to see locations where crime is happening, specifically more "dangerous" crimes. (heat map)
            3. *location* - same as above but with precise coordinates, possible merge on variable
            4. *priority* - see where all the high priority emergencies are happening, possibly filtering out high priority emergencies that do not fall under what we are looking at/considered "dangerous" (e.g. medical emergency)
            5. *description* - to filter on types of calls that would be considered "dangerous"
        ii. Variables from Arrests Dataset:
            1. *arresttime* - longer arrest times could signal more troublesome arrests or ones that use force
        iii. Variables from Use of Force Dataset:
            1. *location* - could be used as a merge on variable to the main Baltimore Calls dataset (or variables district; coordinates, but will involve string manipulation)
    b. **Are police stations used effectively to shorten wait time?**
        i. Variables from Police Station Dataset:
            1. *Name* - identifiers for the police stations in Baltimore

2. ***What innovative approaches can be used to increase public safety for both police officers and citizens?***
    a. **Are there false alarms due to racial profiling?**
        i. Variables from the Baltimore Calls Dataset:
            1. *incidentLocation* - could be used to merge on with arrests dataset (if it falls in a reasonable time period and descriptions match)

ii. Variables from Arrests Dataset:
    1. *arrest* - identifiers for the arrests made
    2. *race* - look at which races are getting arrested
    3. *incidentoffense* - to be able to see what offenses different races are being arrested for

b. Will stricter gun control laws decrease shootings and homicides?

i. Will have to address who is getting the guns (perhaps age, race, gender?) and where they are getting it from.

IMPT INFO FROM DATA CLEANING

Datetime

- Check that all dates are from 2015-2017

District, Neighborhood, Address

- Perhaps make all districts CAPSLOCK uppercase for consistency

Location (long=x, lat=y)

- Had to decrease number of decimals ⇒ may lead to too broad locations (more decimals = more precise location)
    - Should be corrected for by datetime matching and address matching
- *Does decreasing number of decimals decrease uniqueness of rows?*
    - Yes but not enough to make significant difference b/c decreasing decimals lead to more matching