

Basic Info:

The Project Title is: A Brief History of Crime Group member 1 name: Austin Hickey e-mail: U1041943@utah.edu
UID: U1041943

Group member 2 name: Spencer Sawas e-mail: spencer.sawas@utah.edu UID: U1065866

Group member 3 name: Marko Miholjcic e-mail: u0984549@utah.edu UID: u0984549

Background and Motivation

Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.

Across the nation, the police strive to make the world a safer place for civilians. In Chicago, for example, 310 shootings occurred in 2019. A majority of the shootings were gang related. Police in Chicago are trained specifically for handling gang violence. Similar training programs exist for other cities across the United states, but quite frequently, these are often large cities with similar types of crime as Chicago. The strategies for combating crime and ensuring public safety have extended beyond responding to phone calls for service. Law enforcement agencies are collecting data on crimes, and crime mapping in order to increase crime prevention, responsivity, and intervention. The motivation for our project is to find ways to improve the police academies, and learn more information about if and how gun ownership, and crime influence political affiliation.

Project Objectives

Using crime data from the "FBI Crime Data Explorer API" and the Department of Justice "National Crime Victimization Survey API" we want to assess the most common types of crime in cities across America and identify specialized types of training that might be best suited for city-specific police academies. To go a step further, we intend to use the crime data and, using the "ATF Trace Data Report API", identify if there is any relationship between gun ownership and specific types of crime observed.

It must be noted though, street gangs often use ghost guns or "private sales weapons" which are unregistered and not reported to the ATF. Therefore, data from the ATF API may not yield the most precise results.

Finally, we plan to use Google's "Civic Information API" to identify relationships between a) gun ownership vs political affiliation, b) political affiliation vs most common types of crime, and c) assess whether there are higher rates of gun violence per gun owner-density for those living in "red" or "blue" counties.

Data

Data will be collected from Developer Resources section on the United States Department of Justice, google's civic information api, and the FBI crime data explorer website. The url for these websites are:

<https://www.justice.gov/developer> (<https://www.justice.gov/developer>) , <https://crime-data-explorer.fr.cloud.gov/api> (<https://crime-data-explorer.fr.cloud.gov/api>), <https://developers.google.com/civic-information/docs/v2/divisions#resource> (<https://developers.google.com/civic-information/docs/v2/divisions#resource>)

We plan on collecting data by scraping data from the APIs present on the websites. These APIs include: "ATF Trace Data Report API" "FBI Crime Data Explorer API" "National Crime Victimization Survey API" "Civic Information API"

Ethical considerations

Some problems when working with crime data is that due to misreported information by police (such as made up/exaggerated ages) can result in the data misrepresenting the true story of the crime in a given area. Similarly, by analyzing data relating to crime and reporting locations of high volume, you run the risk of the police or other authority figures using it to oppress or stereotype individuals from specific regions with comparatively high crime rates. This has the potential of increased police brutality or police death, and could lead to decreased attention from the city which could cause more unrest in the area.

This data is not meant to conclude whether one political party has any prowess over the other, and the reporting in this must be unbiased so that no political agendas influence the outcome of the data analysis. Honest reporting will need to take place to ensure there is no conflict of interest.

Political candidates may be incentivised to use these data in pushing a political agenda, regardless of political party. Local political officials may also be motivated to make large claims based on criminal data maps, gun-ownership data, a relationship of the two, or create an argument based on current political leadership and crime. Local elections, or large scale political campaigns can use crime data relationships made with gun-ownership as evidence to push for -or against- increased gun control legislation. Additionally, local government can use this data analysis in making a budget for local police departments. Funding for police departments may be hindered, or assisted depending how the data is interpreted by local governments. This can potentially harm an area in need of increased policing.

Data Processing

Ideally the APIs we will be accessing will provide the data in a well organized fashion, so data cleanup will be basic- such as removing white space, making all numbers integers, cleaning up any values that are NaN (by either removing the row or replacing with a local mean value of the category).

We intend to derive the frequency and types of crimes associated with locations across America. We will derive geo political divisions and locally elected officials political affiliation for identifying the local political-landscape. We will use longitude and latitude data from the ATF api to identify gun-owner density and highlight a political relationship between gun owner density and crime occurrences. Using .describe we will identify outliers in the data and determine how outliers impact the data. The mean values of specific categories will be used to fill in Nan values in a dataframe. Pandas will be used to organize a dataframe. sSaborn will be used to visualize statistical relationships such as the cluster analysis, scatter plot matrices, a geospatial heatmap will be used to visualize correlation coefficients calculated between each of the categories identified. Using scikit-extratreesclassifier, we will identify the most important variables associated with types of crime.

Exploratory Analysis

We will map the most common types of crime to Cities across America by creating a Geospatial plot. We will create scatter plots that display the different types of crime for different cities. We provide correlation matrices for gun ownership and specific types of crime across America. In addition, correlation matrices will be made relating gun ownership to political affiliation, political affiliation to common types of crime, and gun violence per gun owner density to those living in democrat or republican leaning counties. We will also create scatter plot matrices to visualize all of these relationships. We will perform classification on the different cities in order to determine which police strategy is the best for the area. We will perform a cluster analysis to group together the different cities based on types of crime, gun ownership, and political affiliation.

Analysis Methodology

We are planning to look at the descriptive statistics of each variable in each city, as well as each the descriptive statistics of variables across all cities. We plan to look at correlation coefficients to determine relationships between the variables in each city. We will develop a classification algorithm to determine the appropriate police strategy based on the characteristics of a given area. We will cluster together the different cities based on types of crime, gun ownership, and political affiliation. We will try to determine which types of crime are most important in determining the police strategy in each area.

Project Schedule

Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

Week 1: -the data must be imported into the jupyter notebook. -Begin cleaning the data -Participate in peer review

Week 2: -Spring Break

Week 3: -Finish cleaning the data -Begin exploring the data

Week 4: -Finish Exploring the Data -Begin Analyzing the Data -Determine the method of creating the video

Week 5: -Complete the analysis of the Data -Begin creating the video

Week 6: -Finalize the jupyter notebook -Finish the video

Week 7: -Practice the presentation

Week 8: -Present the presentation