

Introduction:

Our data set comes from the fatalencounters.org. As of January 2000, fatalencounters.org has been tracking civilians that have been the victims of fatal police shooting. They have collected over a dozen data points from gender, age, race, to mental illness. The dataset, also, contains a brief description of the circumstance. Fatalencounters.org compiled these attributes by searching local news reports, law enforcement websites and social media and by monitoring independent databases such as Killed by Police and Fatal Encounters.

Problem:

While there is no easy solution for police interdiction fatalities, we believe that analysis will allow for a greater understating of these events. We hope that through this understanding there can be procedural change to how case is handled in situation that have the greatest risk of resulting in police interdiction fatalities. We will, also, identify the cities/states/regions that have a higher risk of police interdiction fatalities. Forecasted time and location data could help lead to more police presence in these areas.

Data Preparation:

Because the data set has been compiled with purpose by fatalencounters.org, the need for data preparation is limited. There is potential to group categories into more streamlined variables. Additionally, we can group incidents for use in modeling

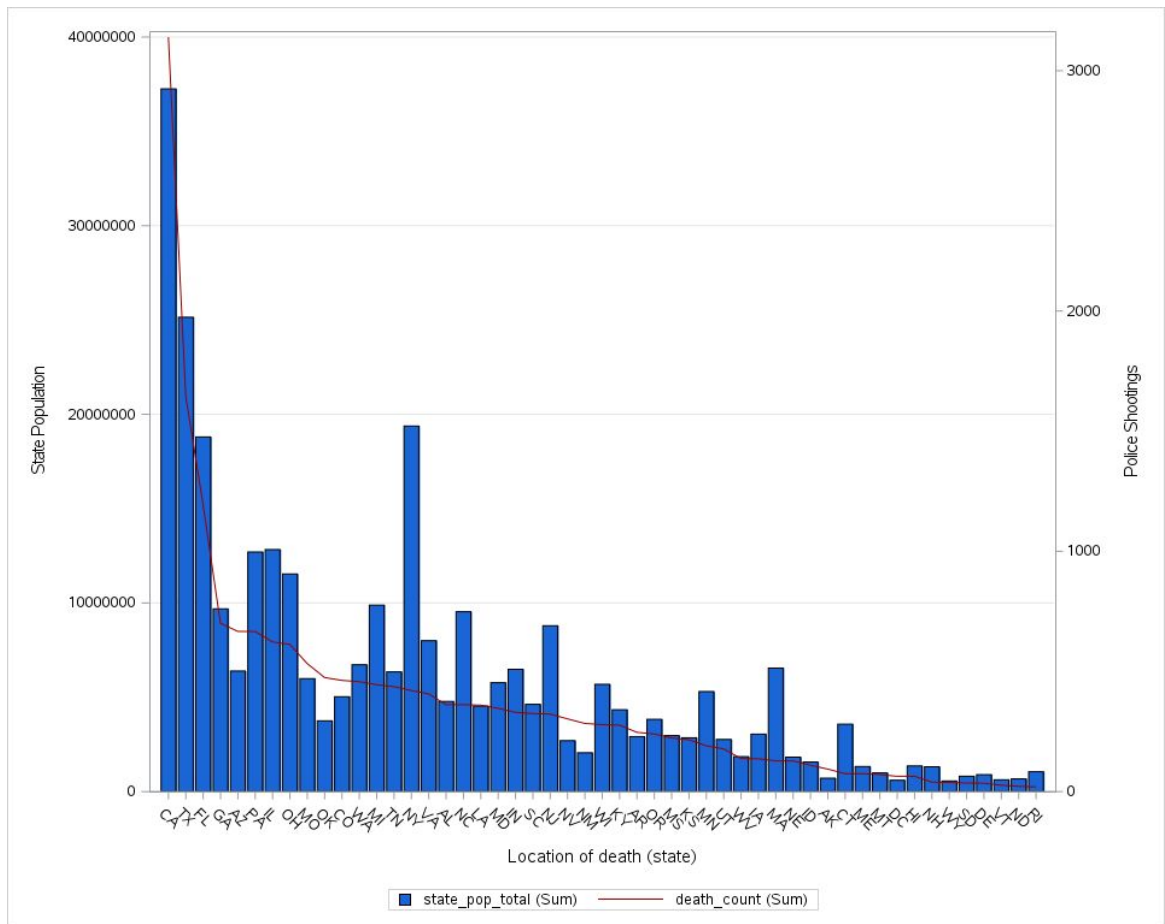
Analysis:

- Variable selection could be used to determine those variables with the most importance to the outcome.
- Overall break-down of the variables through visualizations.
- Time of day and location specific forecasts.
- Age group specific forecast.
- Text analytics of description to find key words and terms with sentiment analysis.
- Suggest future studies with data from specific police departments to collect all data associated with all police dispatches to build predictive and forecasted models for all crimes.

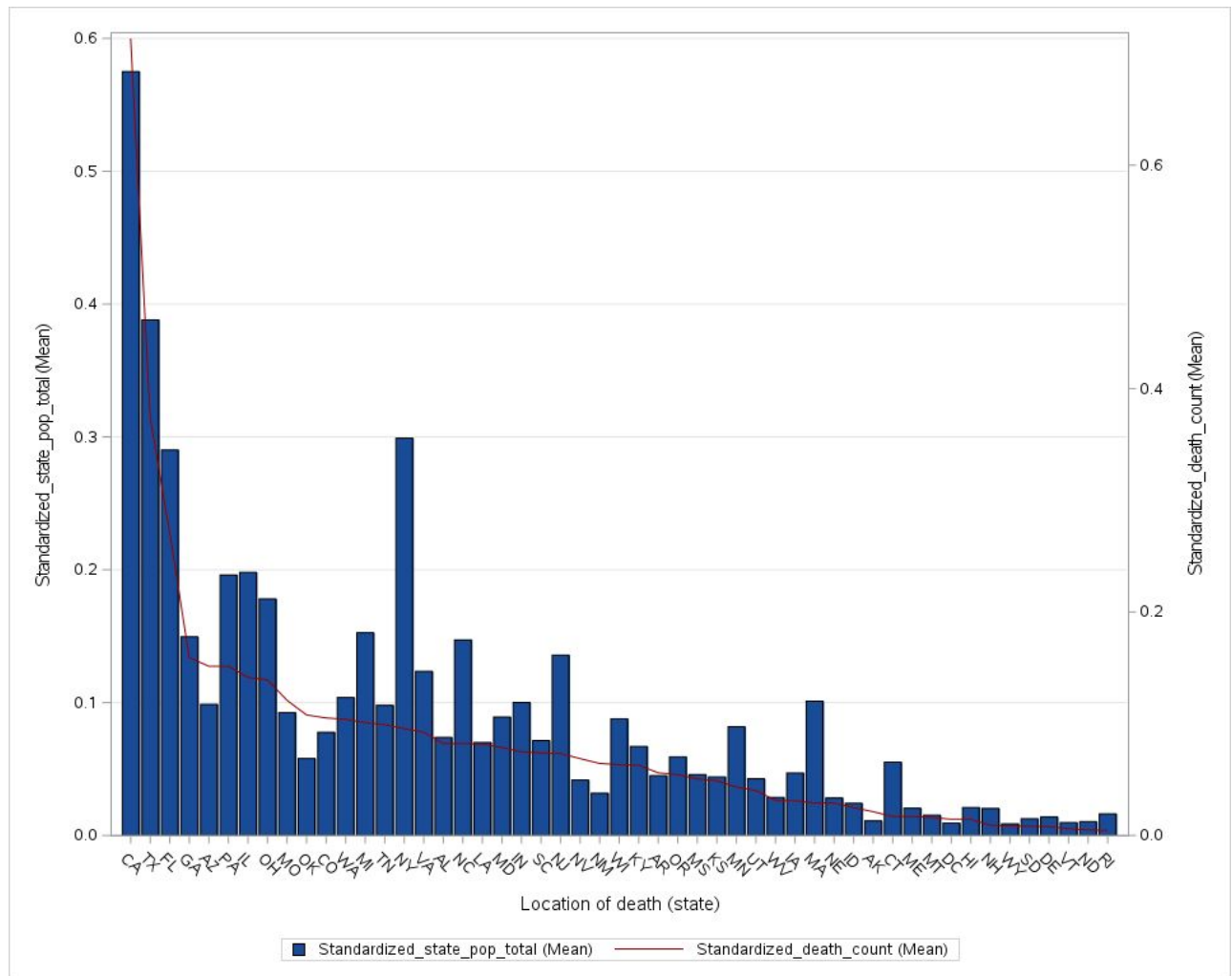
Visualizations

- Trend lines to show changes over time
- Bar Charts to show the various counts and break down
- Colored maps highlighting any location factors

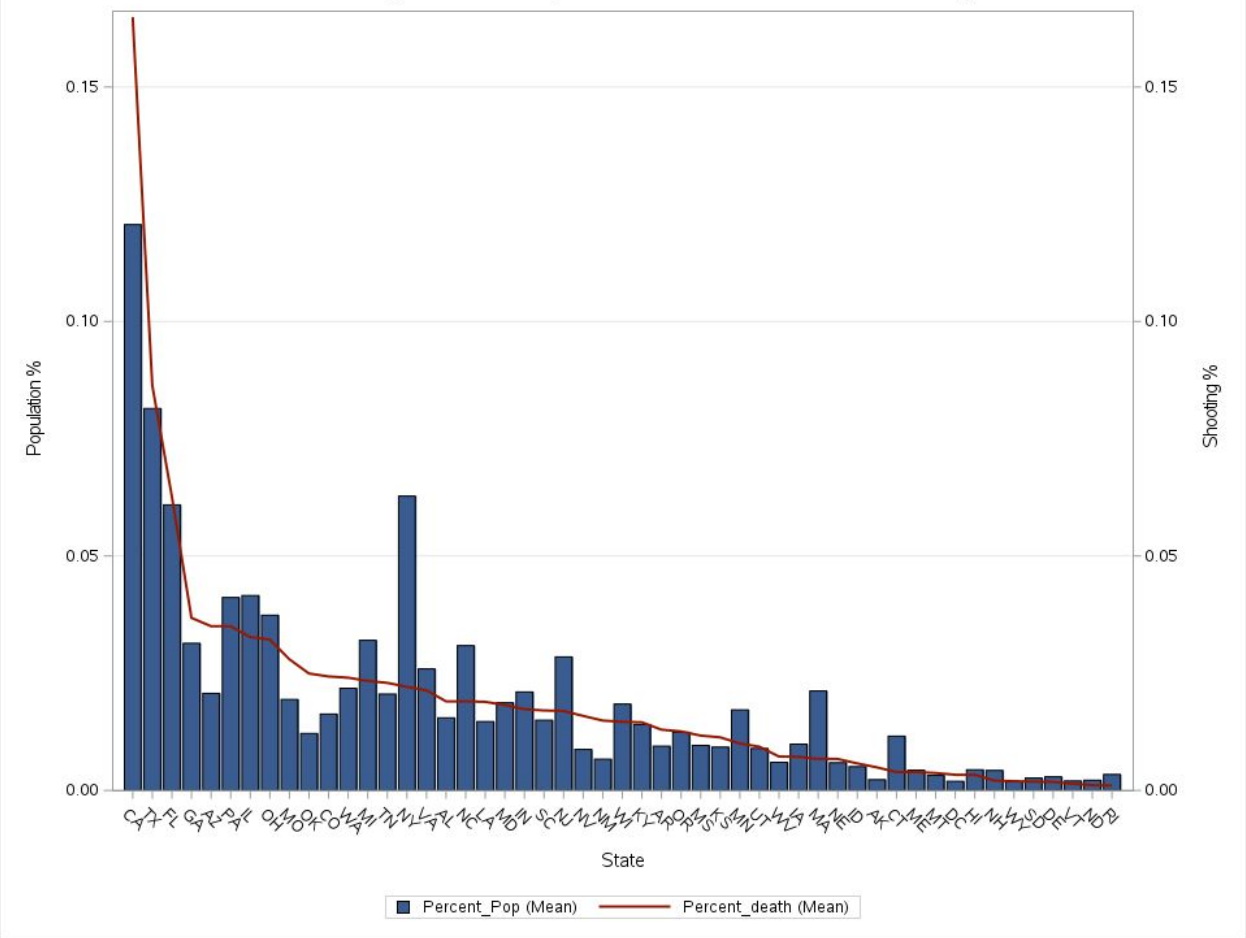
Police Shootings and population non-standardized

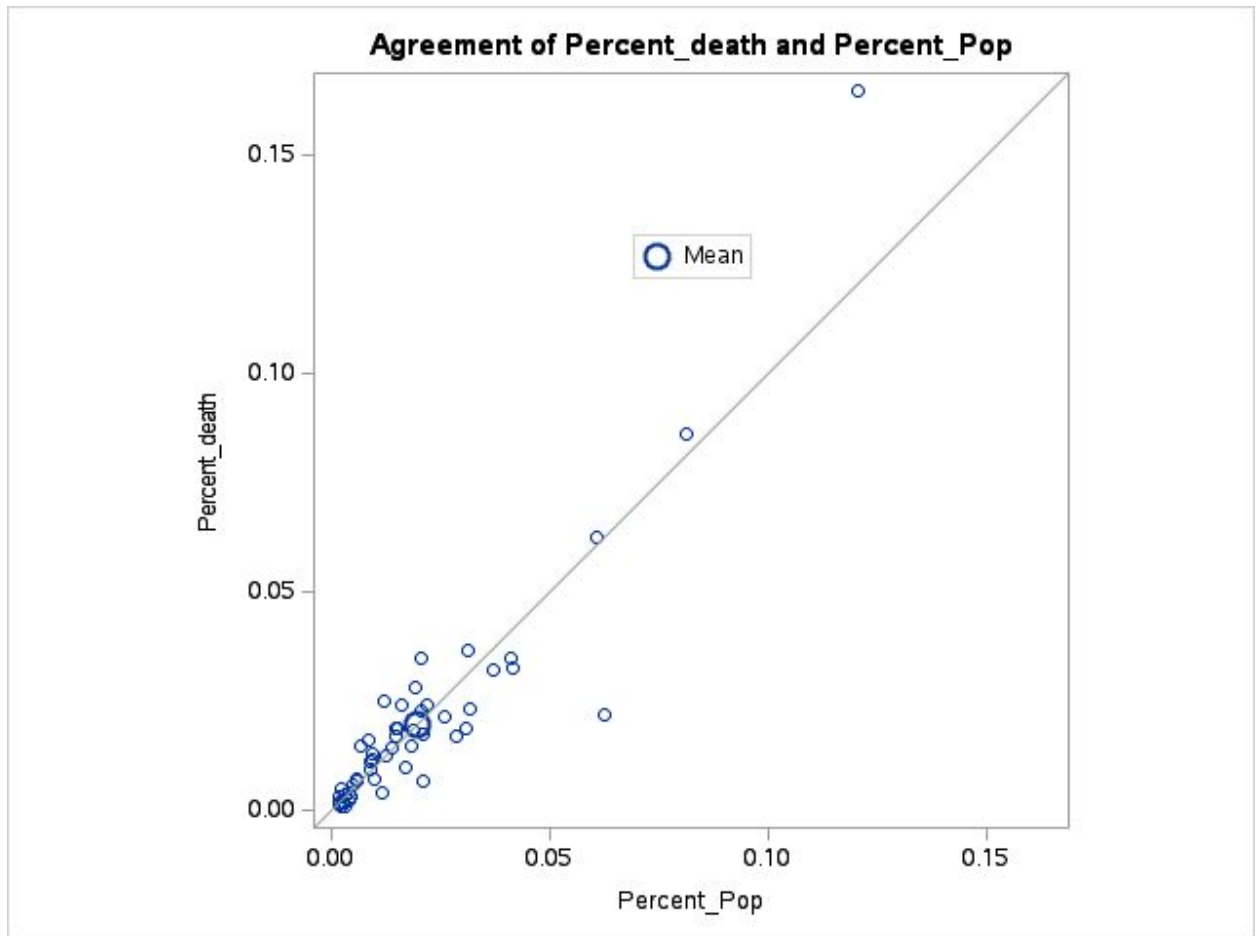


Police shootings and population standardized



Percentage of US Population and Police Shootings





Generalization

We will break down the results of our analysis into easy to understand terms for the non-technical user. Ideally, we will have an outsider review our paper to test their understanding, specifically for this section.

Data set Selection:

- Use case for the analysis and what benefits could be derived from it.
- How well does the set fit the analysis we are trying to perform.
- How accessible is the data and what are the limitation of retrieving it form the source.
- As well we took into consideration the advice we received as we went through the process.

Backup:

National Drug Use Survey

This dataset is very large with over 3000 variables on tobacco/alcohol/drug use ranging from first use to amount of usage in the past month. We could isolate one drug to test the variables from each year to test if the usage is increasing or decreasing from year to year or isolate one drug to test the age and region demographics. We would like to incorporate this data to the police interdiction fatalities data to test if there is a link between drug/alcohol use in region and age demographics that have led to police interdiction fatalities.

FAA Airline Safety

This dataset records all air and ground incidents since the inception of the FAA in 1923. We could test the data to find a connection to certain type of aircraft, manufacturing company, nation, or type of incident that leads to commercial airline mishaps. With this dataset, we would need to conduct web scraping to collect the text writeup on the mishap and more data regarding the aircraft type since there are different models and years produced.

Tools:

- Enterprise miner
- Text and sentiment analysis