CSE-564: Visualization

PROJECT PROPOSAL

Analysis and Visualization of Crime Data in United States

ANKIT AGRAHARI (110946823)
Tarun Lohani (110921666)

Problem Statement:

The newly elected government wants to eradicate various crimes from US has instructed the federal and state police to take preventive measures. The police department want to appoint their force based on the number of criminal cases as well as based on the severity of the crime. They want to see the recent trends and patterns through visualization in the crime across the country with detailed comparision among states and cities.

Objectives:

We will develop an Interactive web visualization which will be helpful for the government to gain insight about the trends; increase/decrease in number of crimes in recent years. We will visualize the regions in United States where the crime rates are high along with the type of crime so that the appropriate police staff could be appointed accordingly. We will also get a picture of the ratio of solved cases with respect to total number of cases, effect of population on crime rates and severity and try to get interesting correlations in any.

We intent to evaluate the following:

1) Figure out the regions with low, medium and high crime rates densities.
2) Filter different regions according to the type of crime.
3) Draw different plots for different regions with the frequency of different kind of crimes occurred in that region.
4) Draw plots for recent years of crime data to see the increase or decrease in the crime rates for each state.
5) Seek some correlation among different type of crimes in all the states and effect of population and other factors on crime rates and severity.

Literature Review:

The reported U.S. violent crime rate includes:

- Murder
- Rapes and sexual assault
- Robbery
- Aggravated assault

Crime rates are necessarily altered by averaging neighborhood higher or lower local rates over a larger population which includes the entire city. Having small pockets of dense crime may lower a city's average crime rate. US violent crime rates includes:

- Homicide
- Gun Violence
- Property Crime
- Crime Against Children

Crime rates vary in the United States depending on the type of community, within metropolitan statistical areas both violent and property crime rates are higher than the national average; in cities located outside metropolitan areas, violent crime was lower than the national average, while property crime was higher for rural areas, both property and violent crime rates were lower than the national average.

Dataset:

Dataset	**Timeline/Year**	**No. of Entries**	**Major Columns/Features**	**Data Source**
Fbi gov crime data	2013	10000 x 14	Population, Violent Crime, Murder and nonnegligent manslaughter, Rapes and sexual assault, Robbery, Aggravated assault etc	https://ucr.fbi.gov/
Crime Rates	1975-2015	3000 x 16	Year, State, Population, Homicide, Rapes , Assaults, Robberies, Crime per capita etc.	https://www.kaggle.com/marshallproject/crime-rates
Effect of Population against crimes	2012	300 x 15	Population, Violent_crime_total, Murder_and_Manslaughter, Forcible_rape, Robbery, Aggravated_assault, Property_crime_total, Burglary, Larceny_theft etc.	https://www.kaggle.com/mascotinme/population-against-crime
Homicide Reports	1980-2015	638455 x 24	City, States, Year. Month, Incident etc.	https://www.kaggle.com/murderaccountability/homicide-reports

Analysis:

We have a number of datasets having information of crime data. We have a dataset which contains type of crime for all the cities in a period of one year. Another dataset has information about crime rate per capita which gives us the effect of population on crime rate and severity. Other datasets have records of various crimes of past 35 years and ratio of solved cases with respect to the total number of cases. To get a meaningful visualization from these data, first we will perform following operations with the data:

- Club different datasets having different information to form one data set where every information related to crime (i.e years, crime rates, types of crimes, crime per capita etc.) can be found and processed.
- Transform the data and find correlation among the rows and columns.
- Clean the data to remove inconsistency and redundant data.

Once having both the city wise and state wise crime data, we will start our analysis with

- Putting different weights to different crime types which will denote the severity of that crime.
- Figuring out the correlation between different type of crimes, areas and population.
- Compare the crimes among cities and states by using different visualization techniques.

Goals\Deliverables:

Our final goal will be to deliver the following -

- Show the low, medium and high crime density areas on the US map.
- Provide two modes of analysis - city-wise and state-wise.
- Provide user interaction so that they can set the severity of different type of crimes.
- Provide user interaction to increase and decrease a particular type of crime to filter the areas on the map.
- Draw plots and graphs to compare and correlate among interesting features of the crime data.

Methodology:

- We will use python for data cleaning and processing.
- Server: We will run server on localhost through python Flask framework.
- Client: Our Client side implementation will be on javascript, html and d3.js libraries.
- We will use different python libraries for our data analysis and other plottings.

References:

Literature Reviews:

- https://en.wikipedia.org/wiki/Crime_in_the_United_States
- https://en.wikipedia.org/wiki/Crime_in_the_United_States#International_comparison
- http://chicago.cbslocal.com/2015/10/22/violent-crime-statistics-for-every-city-in-america/

Data Set:

- https://crime-data-explorer.fr.cloud.gov/
- https://ucr.fbi.gov/crime-in-the-u.s
- https://www.kaggle.com/datasets