Visualization of crime data in Baltimore

Group No: 14
Ravi Teja Kodali
Email: rkodali@g.clemson.edu
Clemson id: C10966668

Venkata Revanth Naidu Danala
Email: vdanala@g.clemson.edu
Clemson id: C18459834
Repository link:
VenkataRevanthNaidu/DatavizG14

Motivation towards selecting this field:

Now a days, all the countries and their governments in the world are stepping towards public security to control the crimes that are being occurred. Billions of dollars are being spent each year to stop criminal activity. We can see that the crime rate in many countries is reducing gradually year by year. In countries like the United States of America, the police department keeps an eye on the criminal activity that occurred with their data from past crime history.

After seeing all this, the police departments in the USA are speedy and tactical in finding out the crimes, solving the crimes. This is one of the main reasons to decrease criminal activity in the USA. This is our main motive behind selecting this field. By studying the crime data, we can create many Visualizations, which will be helpful to understand the data quickly, and we will implement Viz techniques and their tools to a particular crime dataset to get better outcomes. Using the results of our visualization, the police department can quickly get an idea of the crime data by looking at the visualizations.

BACKGROUND RESEARCH:

By studying many research papers, we found that many are implementing visualizations to get the predictions as part of their research. We found very few who are conducting research where their state of the art is Viz.
Also, few of the researchers are creating new tools to make the study easy. Those tools make life easier for the developers.

One of the research papers we took as the reference is "CriPAV: Street-Level Crime Patterns Analysis and Visualization." They created a new tool called CriPAV, used to get the street-level view of crime patterns. This paper used probability and crime intensity to identify high-risk crimes, time visualization, and analytical tasks used in hotspot detection. We found this research interesting because the detection of hotspots is a unique move in Viz. But they have not implemented that analysis with a dataset, so we will implement the hotspot detection with a dataset and get a good outcome.

Another research paper we took as the reference is "Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data." The work in this paper is mainly focused on EDA (exploratory data analysis). EDA is conducted for trend prediction and visualization. Though they had done the visualization, they didn't show any perfect results or outcomes. The authors just compared the crimes between the three cities. This paper will only concentrate on the Viz part and get better outcomes by taking this paper as the reference.

OBJECTIVES:

1. Scatter plot: Data processing is done by using python. We computed probability and intensity using the crime information in the dataset. We divided the dataset by year wise, calculated the probability and intensity in the neighborhoods, and took the crime with more probability. Intensity is nothing but the total crimes that have occurred in that neighborhood. Lastly, we created a scatter plot using intensity and probability.

2. Finding the hotspots: Using the Geospatial mapping to Baltimore city we analyzed street-level hot spots. Also, we categorized the level of intensity crimes that occurred in the hotspots. This is done by using a javascript open-source library 'leaflet'.

3. Sankey graph: With the help of the Sankey graph, we will compare the occurrence of crimes with all the districts we had in the dataset. This helps us in which district the most crimes are happening.

Visualization 1: Scatter plot is visualized for finding the hotspots based on probability and intensity. These hotspots have been presented in the following visualization (geospatial map).

Visualization 2: We implemented Geospatial mapping to find out the hotspots in Baltimore city for the second visualization. Also, we will classify the crime intensity of every hotspot. By implementing this visualization, we can easily understand the area with the highest possibility of crime.

Visualization 3: The third visualization will be the Sankey graph, where the crime counts will be shown in a flow and match the respective year from 2012 to 2017. Lastly, compare them year-wise.

By implementing these visualizations, we can have a better and simple understanding of the crimes in Baltimore.

QUESTIONS THAT WILL BE ANSWERED:

Q1. What are the hotspots in Baltimore city based on the probability of the crimes? Scatter plot
Q2. What is the crime that has the highest probability of occurring in a neighborhood? Geospatial graph
Q3. What is the individual crime count for all five years? Sankey graph.

DATA PROCESSING:

Data processing is required for the dataset. For computing the probability and intensity we have divided the dataset according to years and by using python we have calculated the probability and intensity. The time attribute must be split into the hours and minutes to implement the Time series visualization for hourly crime rate and, we must do the same process for date attribute to split into the day, month, year for year wise crime rate.

DATASET LINK: http://www.fbi.gov/about-us/cjis/ucr/ucr

MUST HAVE FEATURES:

The vital feature we considered in our project is longitude and latitude, which play a significant role in visualization. Without these two attributes, we would think the whole project would be a failure.

Future Work:

In the coming weeks, we will add some styles for the visualizations and complete the other visualizations. This is one-year data for the above visualizations and should do for the remaining five years. Lastly, the Sankey graph will be done along with these visualizations.

Paper References:

10.1109/TVCG.2021.3111146

10.1109/ACCESS.2019.2930410