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 in order 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 the United States of America, the police department keeps an eye on the criminal activity that occurred with the data they had from the past crime history.

After seeing all this, the police department 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 USA. This 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 easily get an idea of the crime data by looking at the visualizations.

BACKGROUND REASEARCH:

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 paper 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 Kernal density estimation 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 those analysis with a dataset so we will implement the hotspot detection with a dataset and will get the 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. By taking this paper as the reference, we will only concentrate on the Viz part and get better outcomes.

OBJECTIVES:

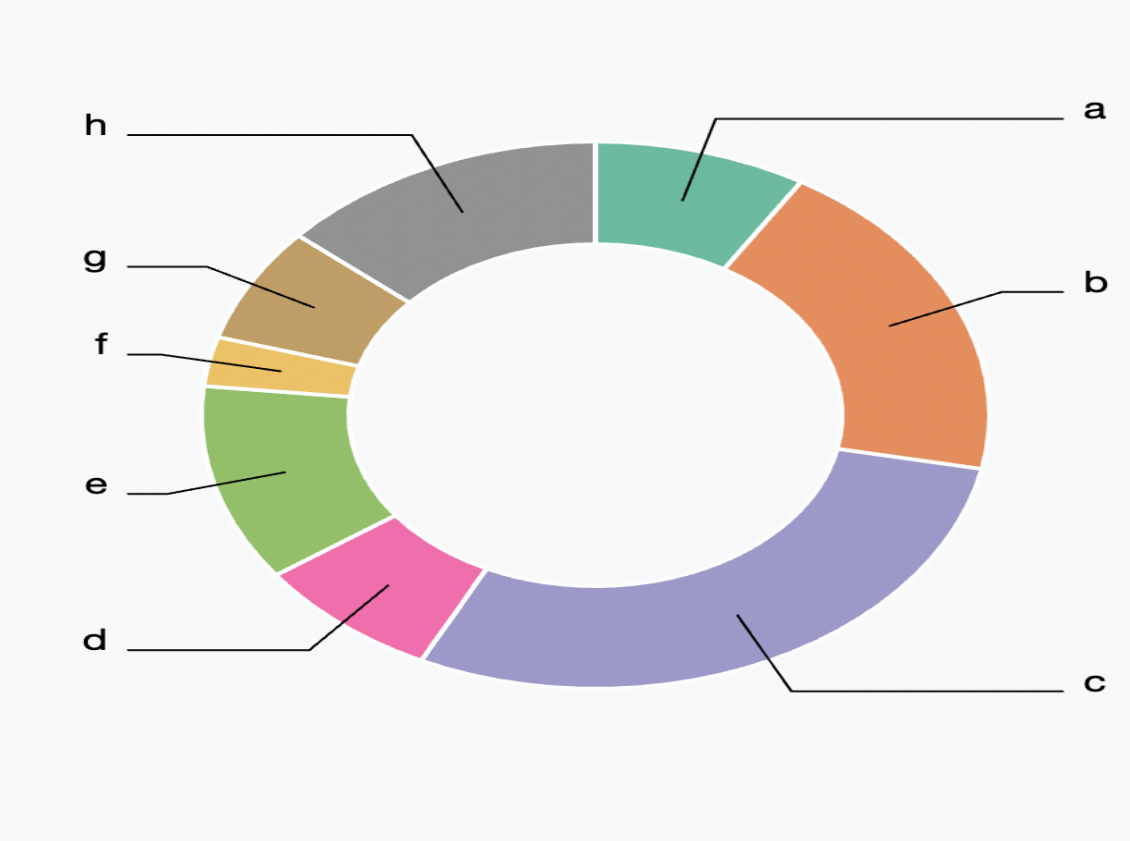
1. Clear explanation of dataset: Using the dataset's information, we will elaborate on the crimes in Baltimore city in the last five years. We will try to figure out what are the different types of crimes that have happened in each year. By doing this, we can have a clear idea about the criminal history of Baltimore city.

2. Finding the hotspots: Using the Geospatial mapping to the Baltimore city will analyze street level and district level hot spots in the city. Also, we will categorize the level of intensity crimes that occurred in the hotspots.

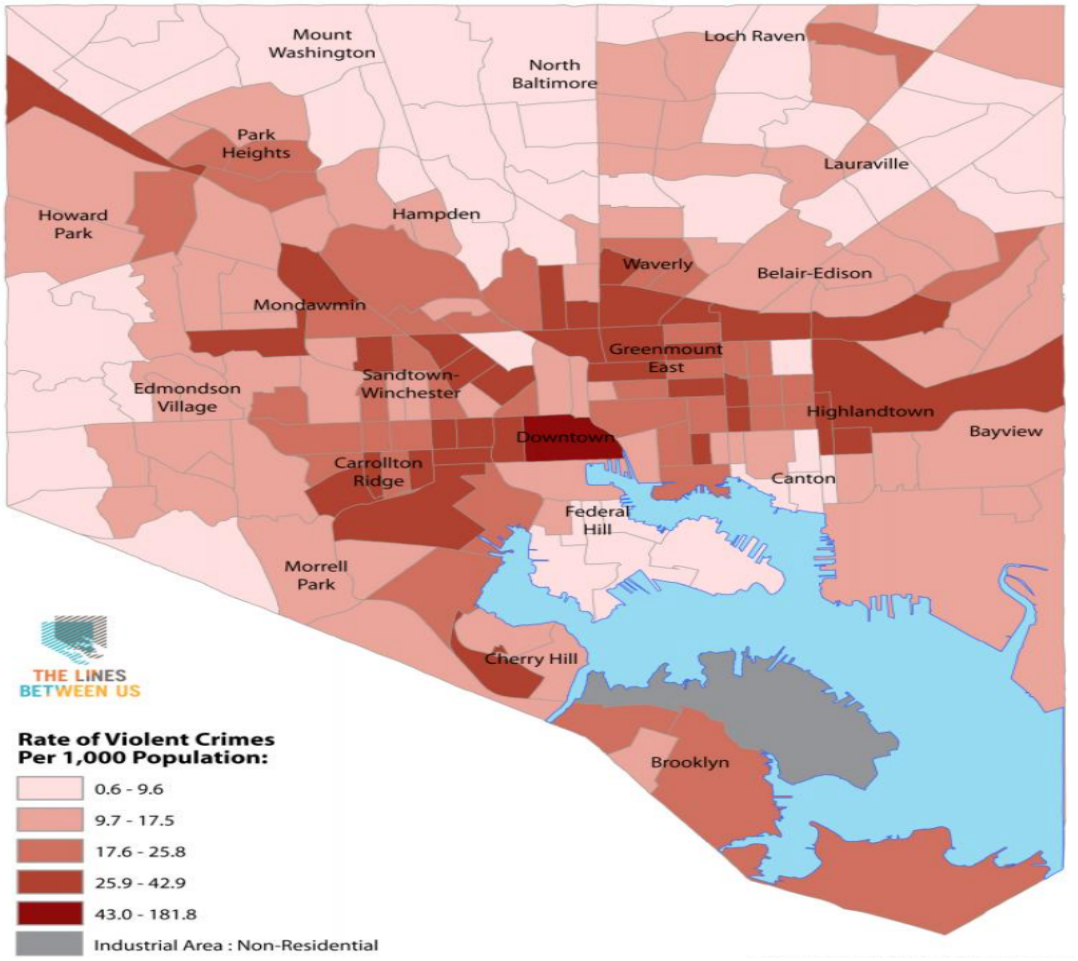
3. Time series visualization: In the dataset, the occurrence of the crime is given, i.e., the time of the crime. By using that, we can create a time-series graph. It will be effortless to understand at what time the occurrence of crimes and the type of crime is happening.

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

Visualization 1: Firstly, to explain the actual crime rate in the span of five years, we will implement the Doughnut chart representing the crime count of all type of crimes yearly. So, we can get the total crime rate for each type of crime in Baltimore city.



Visualization 2: We will implement bubble Geospatial mapping to find out the hotspots in Baltimore city for the second visualization. Also, we will classify the crime intensity of every hotspot with the help of a color hue. By implementing this visualization, we can easily understand the dangerous area and the safe zone.



Visualization 3: In the third visualization, the implementation of the time series visualization will be done. As we have the occurrence of crime time, we can use time series visualization to get an idea at what time more crimes are happening.

Visualization 4: The fourth visualization will be the Sankey graph, where the crimes in each of the four districts will be shown in a flow and match the respective district based on the number of crimes that occurred in each 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 that had occurred in Baltimore.

QUESTIONS THAT WILL BE ANSWERED:

Q1. What are the areas that crimes will occur frequently? Geospatial graph
Q2. What are all the crimes that occurred district wise? Sankey graph
Q3. Which crimes are being occurred very often at a particular time? Time-series visualization
Q4. Total number of crimes that had occur year wise? Doughnut chart

DATA PROCESSING:

A little data processing is required for the dataset. The time attribute must be split into the hours and minutes in order to implement the Time series visualization for hourly crime rate and also, we have to do the same process for date attribute to split into day, month, year for year wise crime rate.

DATASET LINK: https://crime-data-explorer.app.cloud.gov/pages/home

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.

Paper References:

10.1109/TVCG.2021.3111146

10.1109/ACCESS.2019.2930410