

R Shiny Application: Exploratory Data Analysis of Crimes in Boston

Project By: Lahari Kuchibhotla, Suchita Negi, Stuti Sanghavi, Tanmayi Varanasi

Data: <https://www.kaggle.com/AnalyzeBoston/crimes-in-boston>

Online hosted app link : https://stuti.shinyapps.io/Crime_in_Boston/

Data Analysis:

Our dataset “Crimes in Boston” from Kaggle, has both categorical and numerical information which makes it a good dataset for analysis. First, we cleaned up the data by removing the NAs. Then we found that a few of the latitudes and longitudes in the dataset do not correspond to the city of Boston so we removed those specific rows as well.

We used several libraries and techniques for this project that we haven’t used in class. One of these was Leaflet which was used for Geographic Information System (GIS). We also wanted to create a time series analysis of the crime data for which we utilized the Forecast library. Overall, the new techniques used were GIS, Time Series Analysis and Forecasting.

Tab 1: Crime Locations: What are the most common crimes? Where are the most common crimes most likely to occur? How does the frequency of crime change for each district?

The first plot shows the top five most common crimes for each year. From this insight, we wanted to know the specific locations that these crimes were occurring. To plot the crimes we used GIS and a new library called Leaflet which plots the offenses based on latitude and longitude. In the Side Panel, there is a drop-down menu that has the top five most common crimes. The user can select one or more offenses to view. The user can further select which year, range of months and day of the week that they want to view information for. Based on all these filters the crimes will be plotted on the map with circle markers. If the user selects multiple offenses to view, they will be plotted on the map in different colors, so that the user can clearly differentiate between them. We also made our side panel fluid so as the user scrolls down the page they can still view the filters they selected. The map is useful in pinpointing the exact locations of the crimes but if the user wanted a broader understanding of crimes, the last plot in this tab shows the overall crime trends for each district in descending order.

TAB 2: Trends: What are the monthly and hourly trends for the top five most common crimes based on user filters?

While working on the first graph in this tab, we noticed that the years 2015 and 2018 didn’t have data for all 12 months. This could throw an error to the user if they selected a month which is not in the dataset. In order to fix this, we created multiple buckets for each year, storing the number of months in that year. These are used as a reference for the slider. Now, the slider displays only the number of months valid for that particular year.

Our goal, with this graph, was to show a breakdown of crime statistics in Boston during the years 2015 to 2018. The first graph shows the frequency of crimes that occurred each month. In order to make this graph more dynamic and interactive, we added a slider in the side panel that gets updated with the user input. With this feature, the user can select the months they wish to see the data for.

For the second graph, we plotted the number of crimes occurring per hour for the top five crimes. We utilized an inner join to match the description of the crimes with the top five crimes. This graph illustrates the hourly trends and gives some insight into what times of the day are unsafe. The user can use the drop-down menu, to select single/multiple crimes and see the frequency of each of these crimes by the hour. The graph becomes interactive, as the user has the option of selecting the year, month and day of the week that they want to focus on the most.

The last graph shows the overall crime trends for each year. As seen the total number of crimes is increasing.

TAB 3 & 4: UCR Trends: What are the monthly and yearly crime trends for categorized crimes?

These two tabs focus on the Uniform Crime Reporting Index which categorizes the crimes based on severity. The first plot in Tab 3 has the UCR monthly crime trend. As seen there is a spike for all types of crime in June, July, and August. The table allows the user to choose a specific filter and just focus on that data. The table option is present in both tab 3 and tab 4. In Tab 4 the plot shows the UCR yearly trend. Part 3 crimes which are the least severe are most common in all the years.

TAB 5: Forecast: What are the crime trends in the future based on time series analysis?

This section focuses on predicting the number of crimes for future years by using the time series analysis of the crime data in the UCR part. We analyzed the categorical crime data and used the forecast library to store the data in a time series object. Using this data, we plotted the time series to forecast crime trends for future years.

The second graph uses the decomposition of additive time series into seasonal, trend and irregular components using moving averages for all the years and predict for the next two years. (2019-2020).

Summary:

Overall, our R Shiny App analyzes crimes in Boston based on different parameters. The locations of crimes can be viewed on the map which shows which neighborhoods are safe or unsafe. The app also shows monthly and yearly trends based on the user's selection to illustrate what times of the day are unsafe. Lastly, the app forecasts crime trends for 2 years.