

A photograph of the Boston skyline at night, with numerous skyscrapers illuminated and their lights reflecting on the water in the foreground. The sky is dark blue with a faint crescent moon visible.

FINAL PROJECT

ALY6010: Prob Theory and Intro Statistics

BOSTON CRIME ANALYSIS (2022)

By Sruthi Kondra

A decorative graphic in the bottom right corner consisting of several overlapping green triangles of different shades, creating a modern, abstract look.

INTRODUCTION AND OVERVIEW OF THE DATASET

- Selected the "Crime Incident Reports In Boston" dataset from data.boston.gov, focusing on data from the year 2022.
- The dataset comprises 73,852 observations across 16 columns, including offense type, code, street location, exact time, and geographic coordinates.
- Selected to analyze the growing crime rate in Boston and identify patterns for strategic interventions and public safety improvements.



Data Cleaning and Preparation:

1.Key Actions and Improvements:

- 1. Renamed columns for clarity and standardization.**
- 2. Converted data types to appropriate formats (e.g., offense codes to integers).**
- 3. Categorized crimes into distinct groups like 'Violent', 'Moderate', and 'Fraud' for detailed analysis.**

Initial Analysis:

1.Offense Details Subset:

Created a focused subset named 'offense details', isolating specific crime characteristics for deeper examination.

2.Crime Categorization:

- 1. Classified crimes into four main categories: 'Violent', 'Moderate', 'Fraud', and 'Other'.**
- 2. This classification aids in understanding the severity and nature of the crimes reported, setting the stage for targeted analysis.**

3.Foundation for Detailed Investigation:

The categorization forms the basis for exploring crime trends and patterns in subsequent analyses.

Variables of interest:

1.OFFENSE_CODE:

Description: Numerical code representing the type of offense.

2.OFFENSE_DESCRIPTION:

- 1. Description: Detailed description of the offense.**

3.HOUR:

Description: The hour of the day (in 24-hour format) when the incident occurred, used to analyze crime frequency in relation to the time of day.

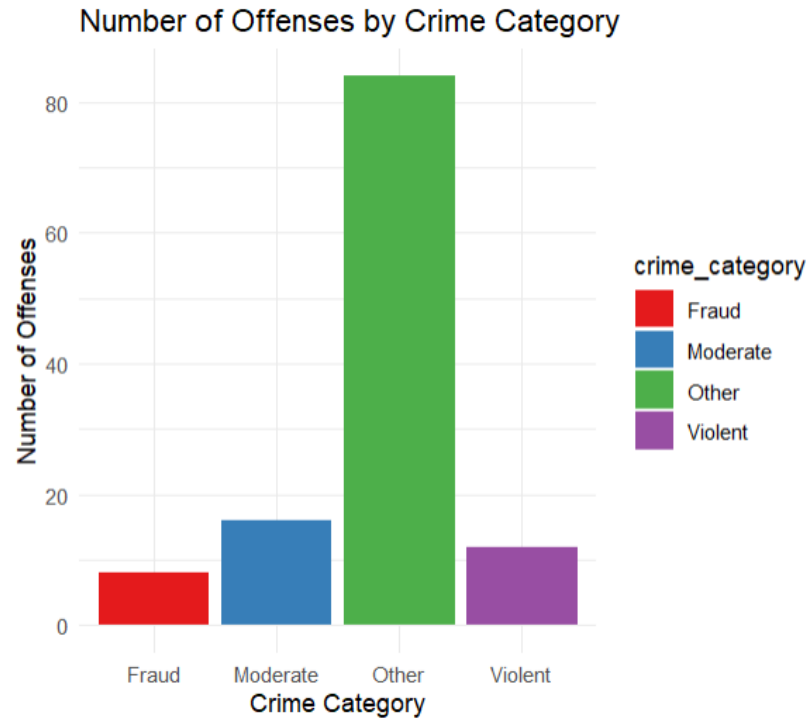
4.Lat (Latitude):

- 1. Description: Geographical latitude coordinate of the incident location, used for analyzing the correlation between crime frequency and geographical location.**

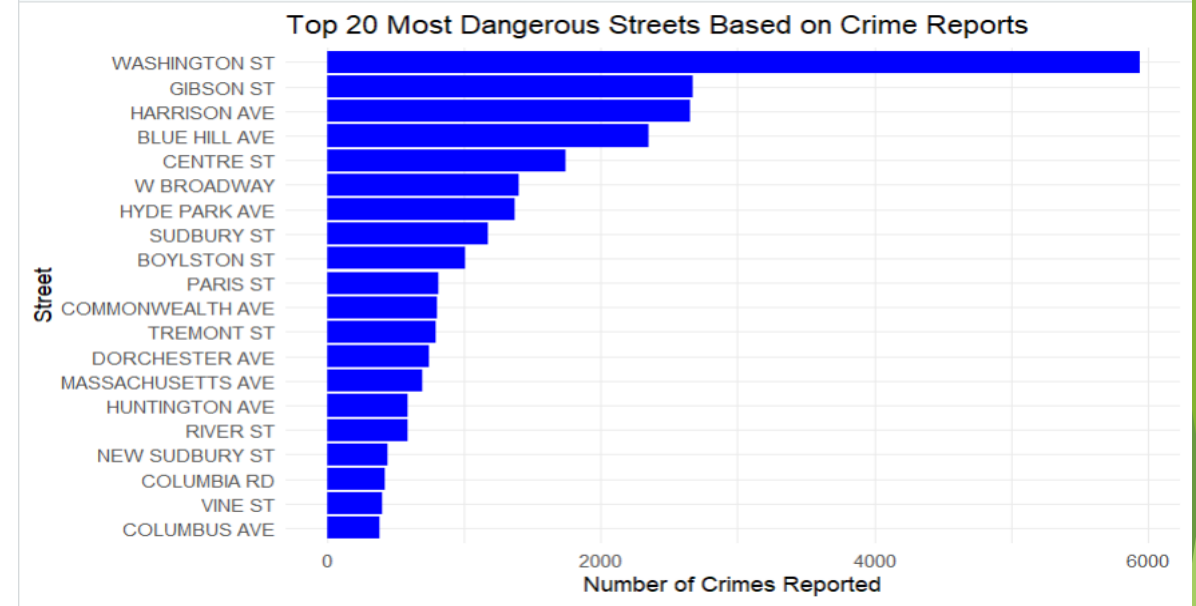
5.MONTH:

- 1. Description: The month when the incident was reported, used in analyzing crime trends across different months of the year.**

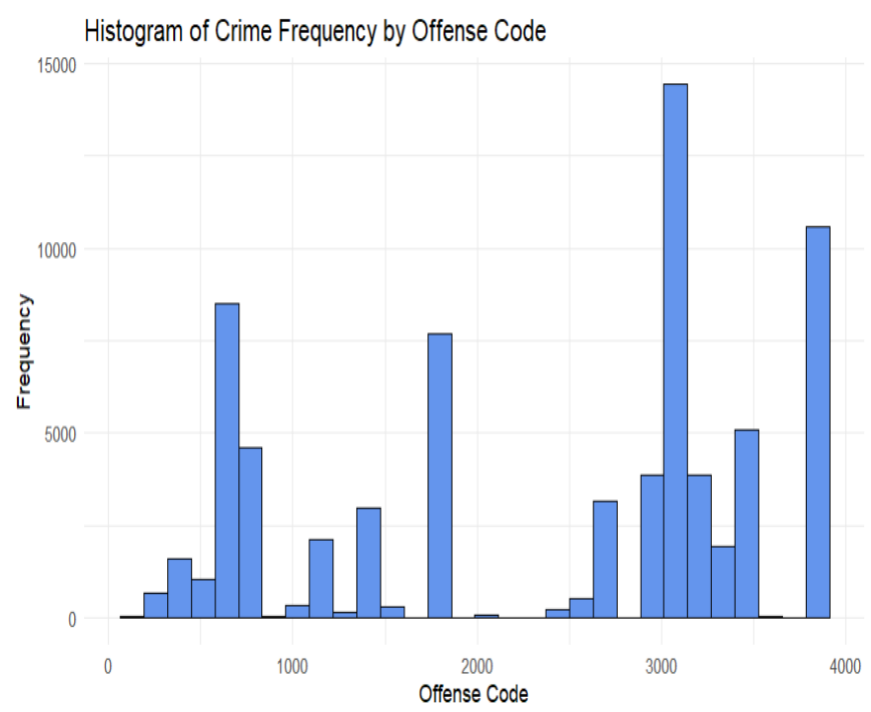
Visualizations from Initial Data Analysis



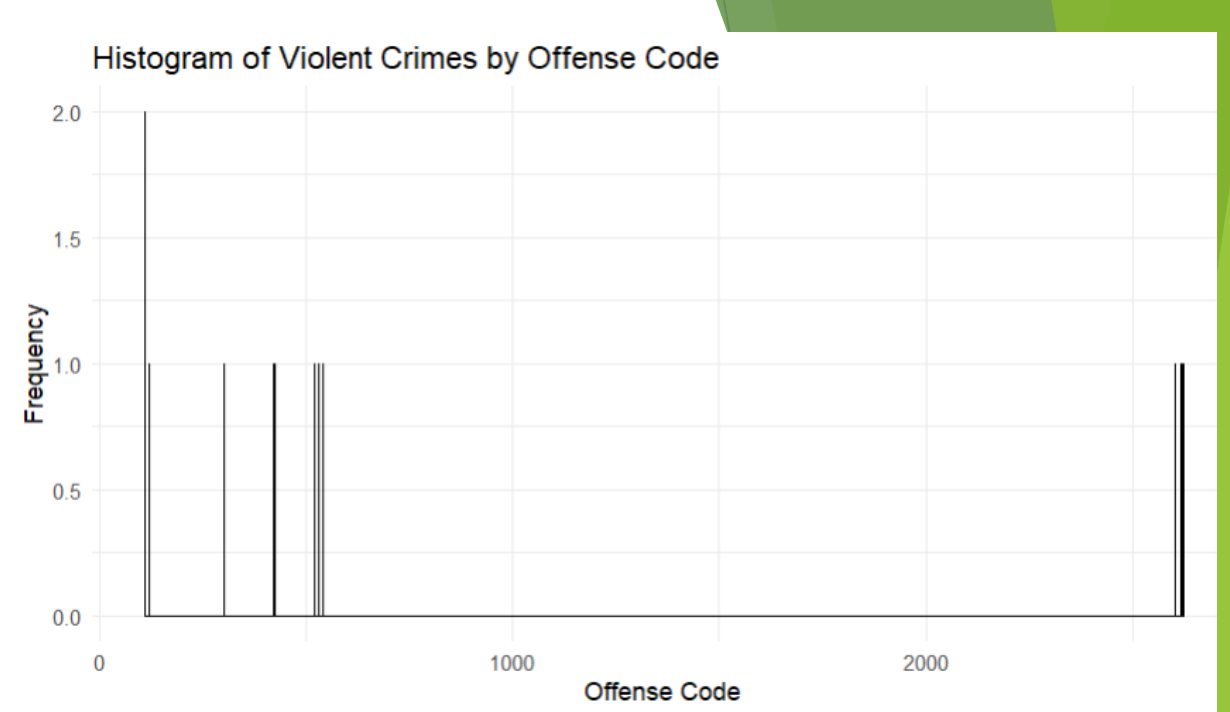
The 'Other' category of crimes significantly surpasses 'Violent', 'Moderate', and 'Fraud' in frequency, suggesting a broad range of criminal activities beyond major categories that could require additional categorization or analysis.



Washington Street stands out as the most reported location for crime, indicating a hotspot that may require targeted law enforcement and community safety interventions.



A few offense codes show a particularly high frequency of crimes, pointing towards specific types of criminal activities that are more prevalent and may need focused prevention strategies.



Violent crimes are concentrated within specific offense codes, which could be pivotal in understanding and addressing the root causes of violence in the community.

crime_category	count	mean_offense_code	median_offense_code
Fraud	8	1,296.2500	1,107.5
Moderate	16	917.0625	618.5
Other	84	2,843.2381	3,023.5
Violent	12	919.3333	525.0

The table is used to obtain descriptive statistics for each crime category, such as mean, median, and count of the offense code and display it in a table

Questions Explored

Question 1: Relationship Between Time of Day and Frequency of Violent Crimes

- Point:** Investigate if there's a significant relationship between the time of day (hour) and the frequency of violent crimes.
- Reason:** To determine if certain hours are more prone to violent crimes, which can inform law enforcement and public safety strategies for targeted interventions during specific times.

```
Call:
lm(formula = frequency ~ hour, data = hourly_crime_frequency)

Residuals:
    Min       1Q   Median       3Q      Max
-74.003 -20.509  -4.788  15.184  159.853

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  105.147    18.527   5.675 0.0000105 ***
hour           3.371     1.380   2.442  0.0231 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 46.81 on 22 degrees of freedom
Multiple R-squared:  0.2133,    Adjusted R-squared:  0.1776
F-statistic: 5.965 on 1 and 22 DF,  p-value: 0.02308
```

The type of testing used in the analysis of the relationship between the time of day and the frequency of violent crimes is linear regression analysis. The linear regression analysis reveals that as each hour passes, violent crimes rise by about **3.371 incidents**

Question 2: Correlation Between Latitude and Crime Frequency

- Point:** Assess if there's a significant correlation between the latitude of a crime incident and the frequency of crimes in that area.
- Reason:** To understand if geographical factors, such as location within the city, influence crime rates. This can help in identifying high-risk areas and in resource allocation for crime prevention.

Pearson's product-moment correlation

```
data: latitude_crime_frequency$lat and latitude_crime_frequency$frequency
t = 0.92448, df = 12540, p-value = 0.3553
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.009247437  0.025752938
sample estimates:
      cor
0.008255279
```

For this hypothesis testing we used a correlation test called the Pearson's correlation coefficient. The analysis suggests that there is no significant correlation between latitude and crime frequency. The high p-value supports maintaining the null hypothesis, indicating that the latitude of a crime incident does not show a measurable impact on the frequency of crimes in that area.

Question 3: Variation of Crime Frequency Across Different Months

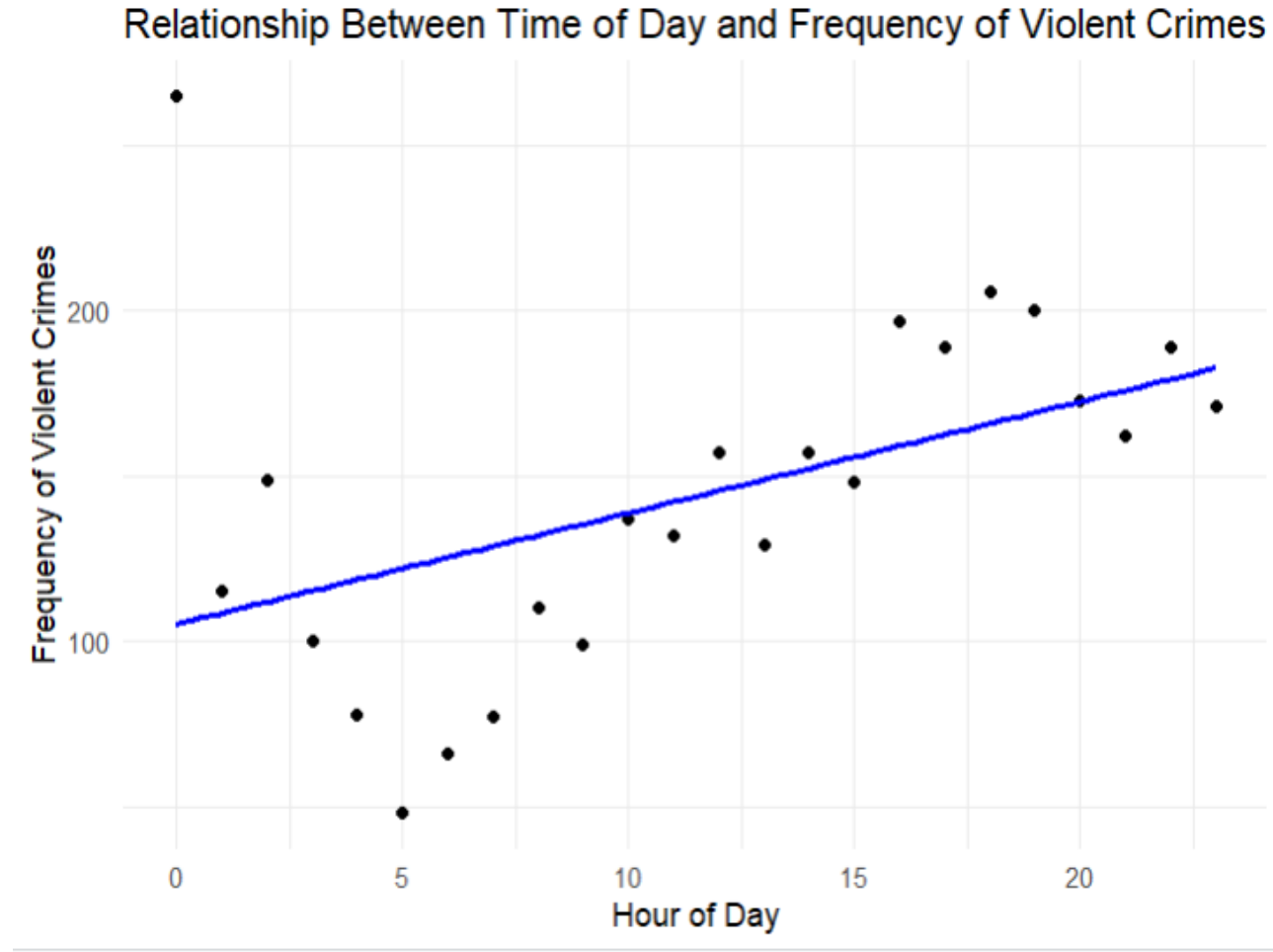
- Point:** Determine if the frequency of total crimes significantly varies across different months of the year.
- Reason:** To identify potential seasonal trends in crime rates, which can guide public safety measures and resource planning on a seasonal basis.

Pearson's product-moment correlation

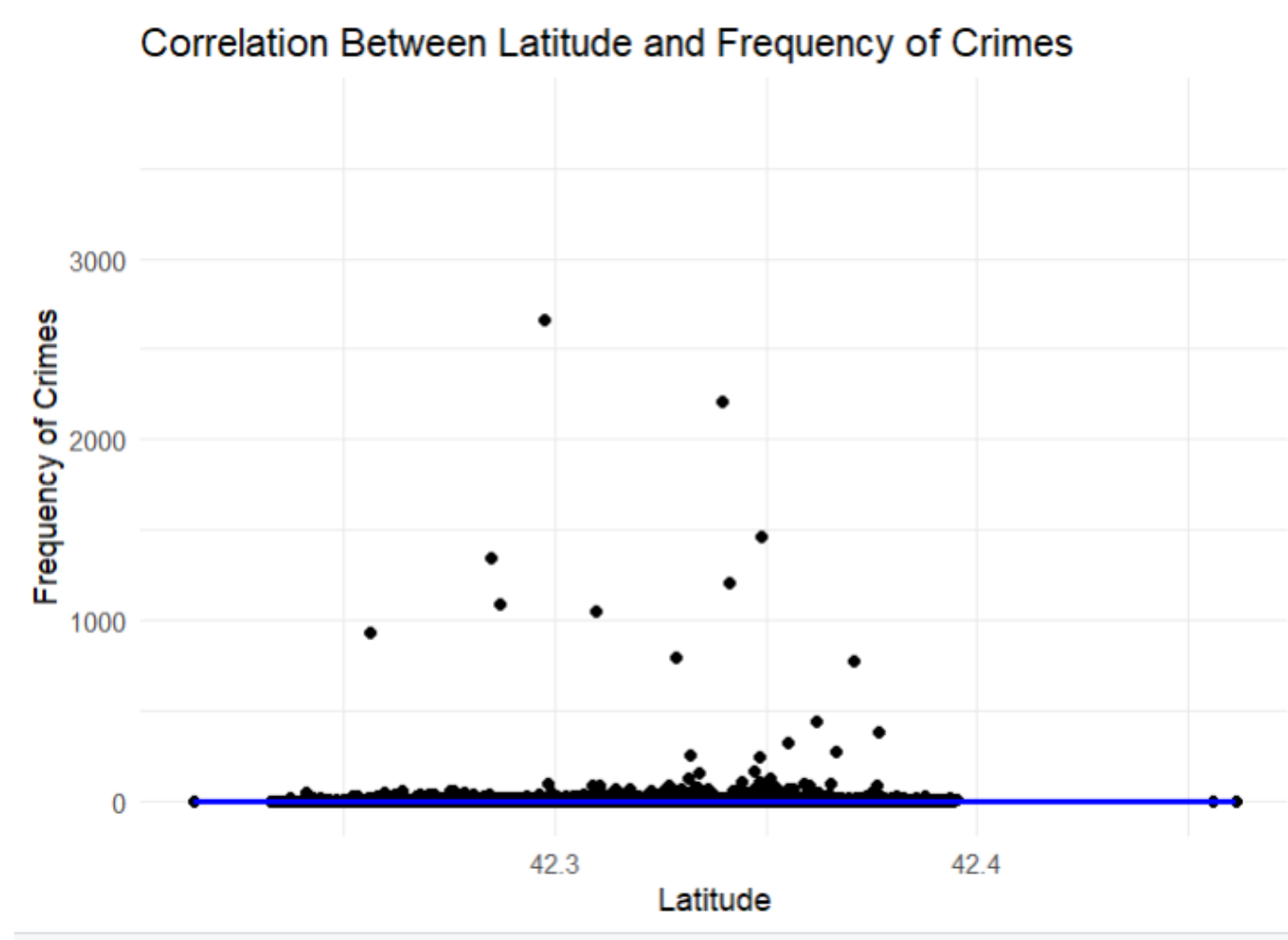
```
data: monthly_crime_frequency$month_num and monthly_crime_frequency$frequency
t = 1.1232, df = 10, p-value = 0.2876
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.2960781  0.7621929
sample estimates:
      cor
0.3346948
```

A correlation test was conducted to analyze the relationship between the month and crime frequency. The higher p-value leads us to retain the null hypothesis, suggesting that the month of the year does not have a measurable influence on the overall frequency of crimes.

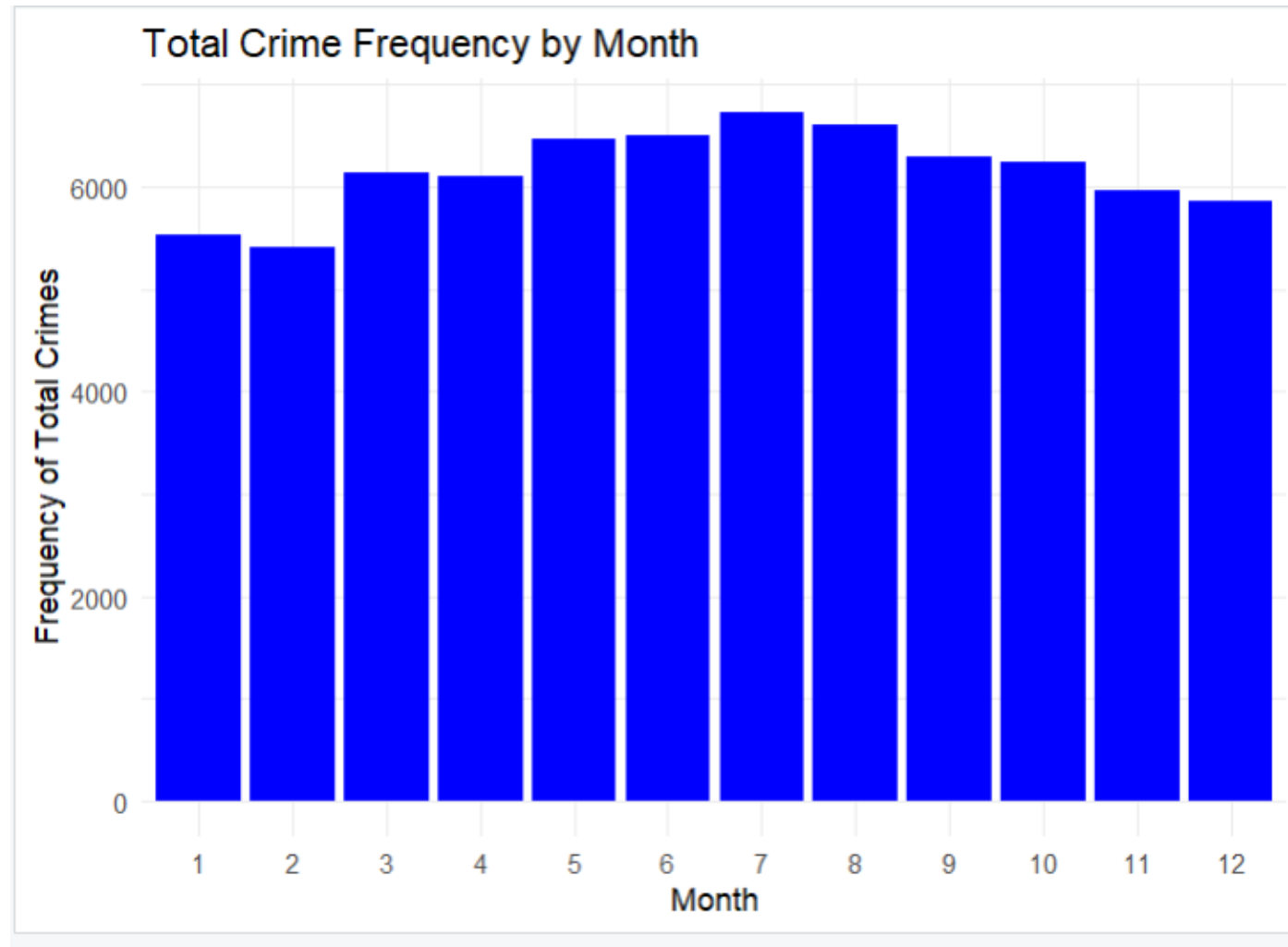
Visualizations:



The scatter plot indicates a clear positive trend, showing that the frequency of violent crimes increases as the day progresses, with a significant peak in the evening hours.



The scatter plot does not show a strong relationship between latitude and crime frequency, indicating that latitude may not be a significant factor in predicting crime occurrences.



The bar chart shows that the total crime frequency remains relatively consistent throughout the year, with no significant peaks or troughs indicating seasonal trends.

Conclusion

Concluding Insights from Crime Data Analysis

1. Time of Day Correlation:

1. The analysis confirmed a statistically significant increase in violent crimes later in the day, highlighting potential times for heightened law enforcement vigilance.

2. Geographical Influence:

1. Latitude showed no substantial impact on crime frequency, suggesting that factors other than geographic location within the city may play a more crucial role in crime occurrence.

3. Monthly Crime Trends:

1. The consistent crime frequency across months indicates the absence of significant seasonal variations, which can assist in steady year-round resource allocation for law enforcement.

4. Strategic Recommendations:

1. Based on the findings, law enforcement strategies can be optimized by focusing on time-specific patrolling, while geographic and monthly allocations can be standardized due to the lack of strong correlations with crime rates.

Thank
you!