Report

On

Crime Data Visualization Using R

[CS 534 - R Programming Lab]

Prepared By:

Aditya Kumar Pandey

M.Tech (AI & DS), 240201001

Dept. Of CSE, IIIT Bhagalpur

Submitted To:

Dr. Om Prakash Singh

Assistant Professor

Dept. Of CSE, IIIT Bhagalpur

INDEX

1. Introduction

This project is designed to analyze crime data from Greater Manchester using the statistical programming language R. The objective is not just to view data but to extract insights from crime patterns by employing data visualization techniques.

Key Goals:

- Understand where and when crimes occur most frequently.
- Use graphs, heatmaps, and trend lines to detect patterns.
- Help authorities with decision-making using easy-to-understand visual insights.

2. Problem Statement

Challenges with Raw Crime Data:

- Thousands of records are difficult to interpret manually.
- Trends and hotspots are buried in text and numbers.
- No straightforward way to see which boroughs or categories are more dangerous or increasing in risk.

Solution:

- Use data visualization tools in R to represent the data visually.
- Graphs, charts, and maps help reveal patterns that would otherwise be hidden.
- Makes trend analysis and policy decisions easier and faster.

3. Dataset Information

Source:

- Downloaded from Kaggle, which hosts publicly available crime datasets.

Format:

- CSV file containing 41,450 rows and 7 columns.

Feature	Description
date	Date when the crime was reported
borough	Administrative area (e.g., Manchester, Bolton)
location	Descriptive location (e.g., On or near Supermarket)
lsoa	Local geographic unit used for UK statistics

category	Type of crime (e.g., Burglary, Drugs)
lat	Latitude coordinate of the incident
long	Longitude coordinate of the incident

4. Libraries and Tools Used

- ggplot2: For creating rich visualizations like bar plots, heatmaps, and line charts.
- dplyr: Used for filtering, summarizing, and transforming data.
- readr: Fast and user-friendly CSV reader.
- lubridate: Makes handling dates (months, years) much simpler.
- reshape2: Helps reshape data for heatmaps and other matrix-style visuals.

5. Data Preprocessing

Read Data:

```
crime_data <- read.csv("crime_data.csv"
```

Date Conversion:

```
crime_data$date <- as.Date(crime_data$date, "%Y-%m-%d")
```
Converts strings like "2015-01-01" into actual date objects.

Categorical Conversion:

```
crime_data$borough <- as.factor(crime_data$borough)
crime_data$category <- as.factor(crime_data$category)
```

Treats borough and category as *categories* not plain text (improves efficiency in plots & summaries).

Missing Values:

```
colSums(is.na(crime_data))
```

Identifies any columns with missing data.

Feature Extraction:

```
crime_data$month <- format(crime_data$date, "%m")
crime_data$year <- format(crime_data$date, "%Y")
```

Allows grouping and analysis by month or year.

6. Analysis and Visualizations

a. Crime Count by Category

- Highlights that Violence & Sexual Offences, Anti-Social Behaviour, and Vehicle Crime are the top 3 crimes.
- Useful to prioritize law enforcement focus.

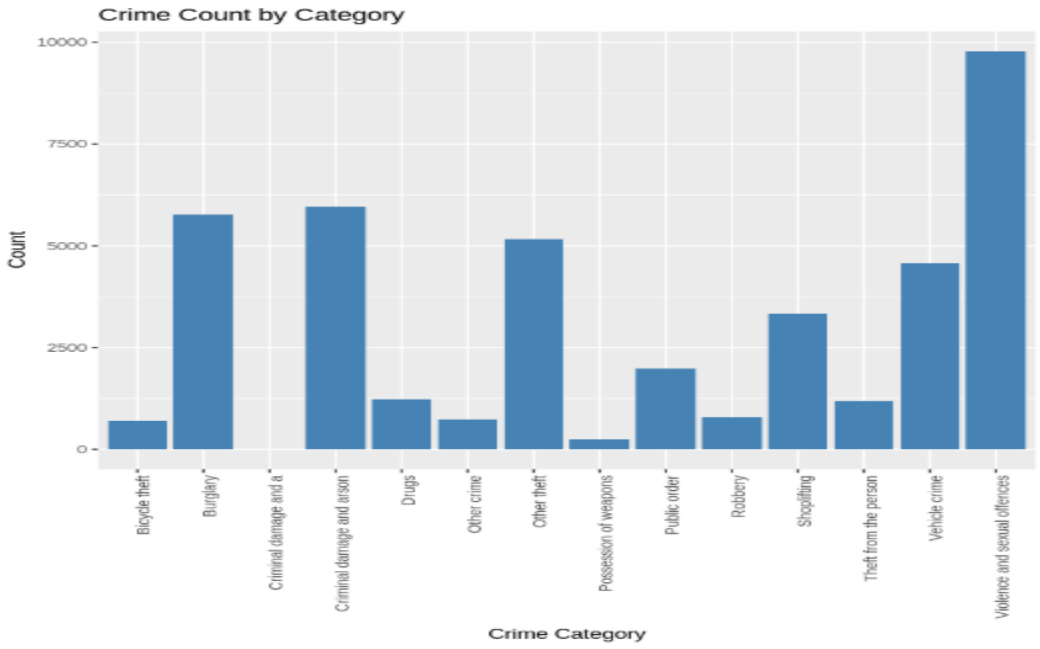

Fig1: *Crime count by Category – Vertical Bar Chart*

b. Top 10 Boroughs with Highest Crimes

- Manchester leads with the highest number of crimes, followed by Bolton.
- Important for resource allocation and patrolling strategy.

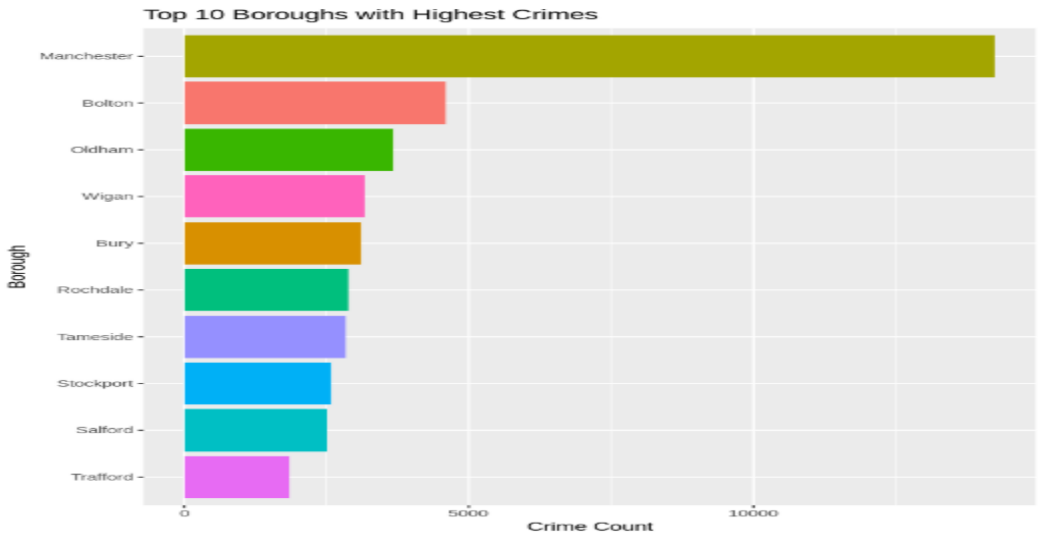

Fig2: *Top 10 Boroughs with Highest Crime – Horizontal Bar Chart*

c. Temporal Crime Trends

- Crimes rose from January to February, then fell sharply in March.
- Possible reasons: policy changes, weather, or data reporting lag.

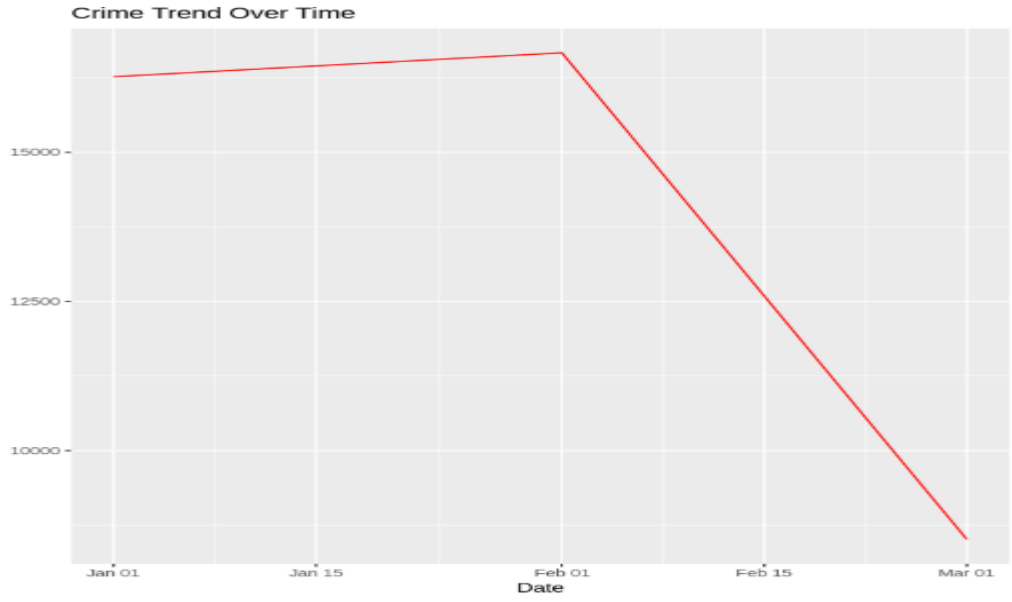

Fig3: *Crime Trend over Time – Line Chart*

d. Crime by Location

- Top locations: Shopping Areas, Supermarkets, Parking Areas.

- Indicates urban and commercial zones are more targeted.

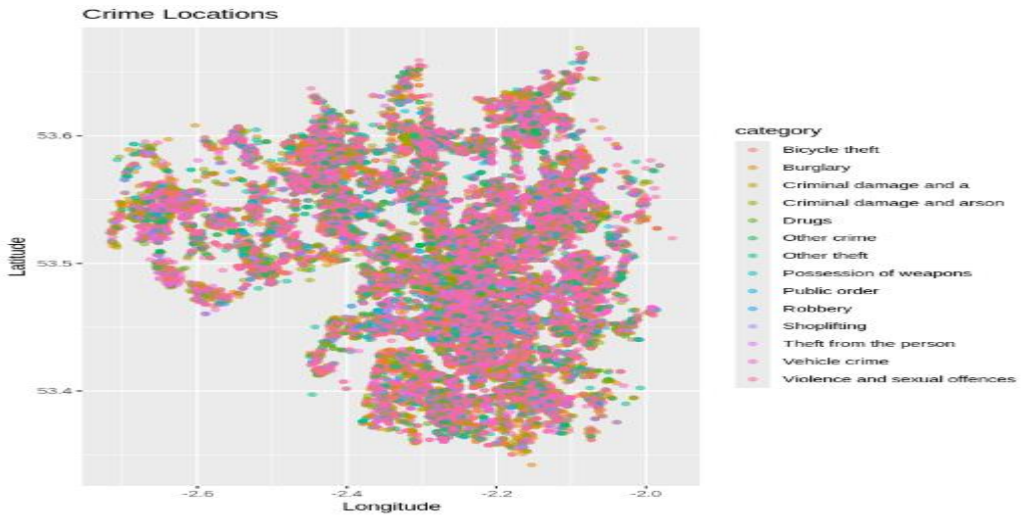

Fig4: *Crime Count by Location– Geospatial Chart*

e. Monthly Crime Trends

- February peaks, March shows a sharp drop.
- May be used for seasonal crime prevention programs.

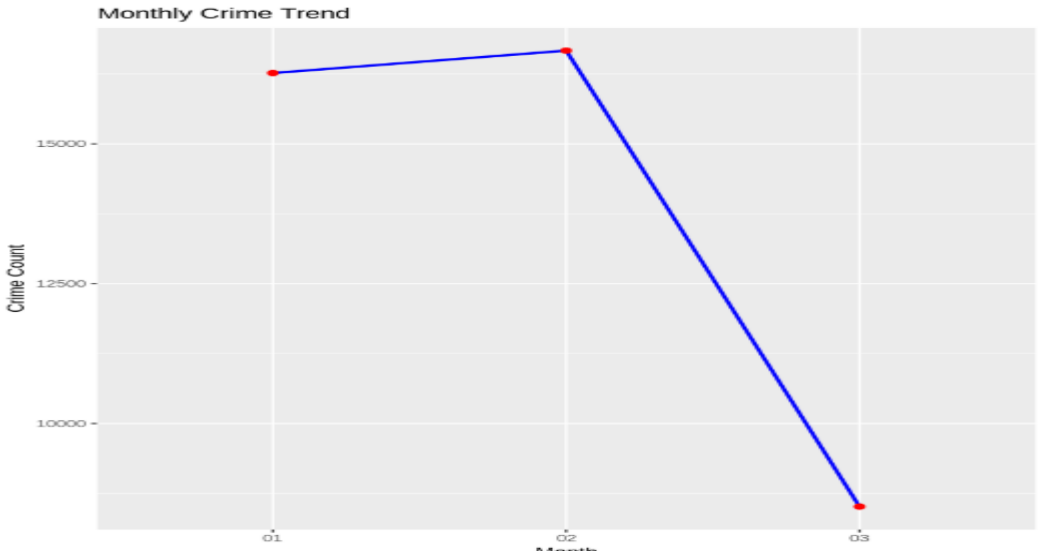

Fig5: *Monthly Crime Trend– Line Chart*

f. Crime Trends in Manchester, Bolton & Wigan

- Manchester has a clear upward trend.
- Helps in comparing boroughs' performance and adjusting plans.

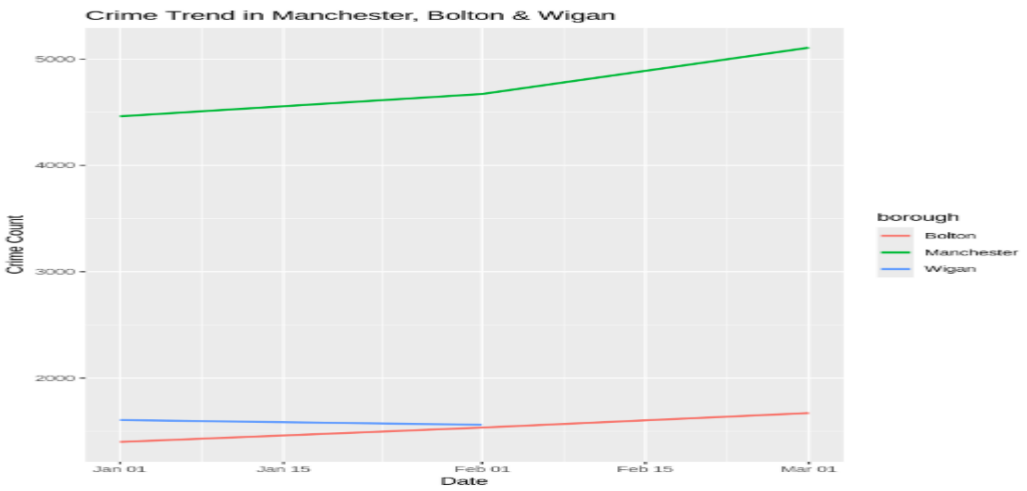

Fig6: *Crime Trends in Manchester, Bolton & Wigan – Multi - Line Chart*

g. Heatmap: Borough vs Month

- Visualizes which borough had more crimes in which months.
- Manchester always peaks—sign of a chronic hotspot.

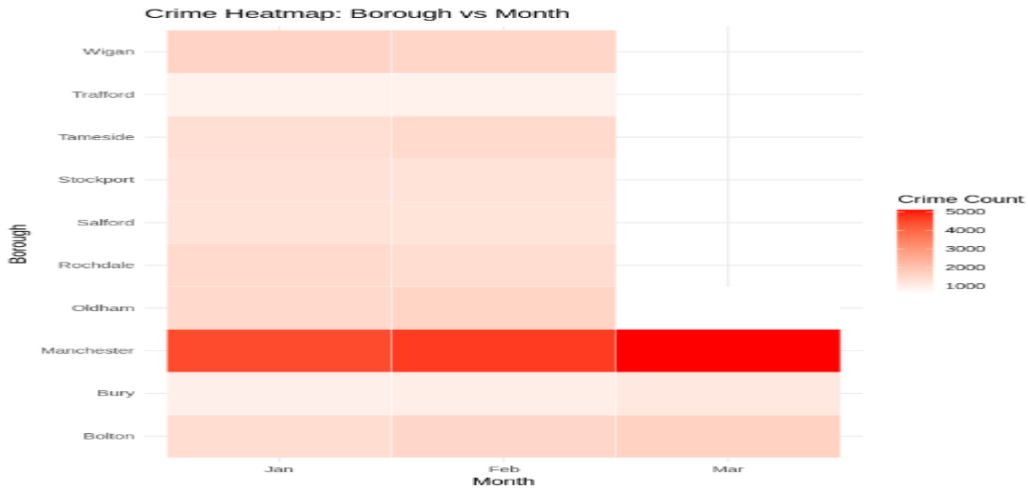

Fig7: *Crime Heatmap: Borough vs. Month– Heatmap*

h. LSOA-Level Crime

- Focuses on micro-areas within boroughs.
- Useful for hyperlocal interventions, neighborhood watch, or CCTV installations.

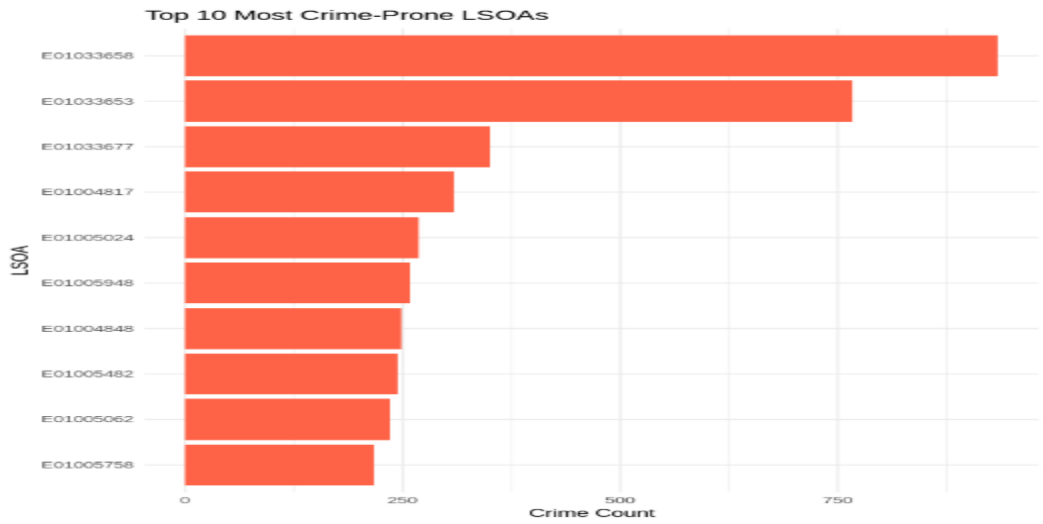

Fig8: *LSOA Analysis – Horizontal Bar Chart*

i. Crime Categories per Borough Heatmap

- This heatmap highlights the intensity of various crime categories across boroughs.
- Manchester leads significantly in violent and sexual offences.
- Helps detect borough-specific crime patterns for focused interventions.

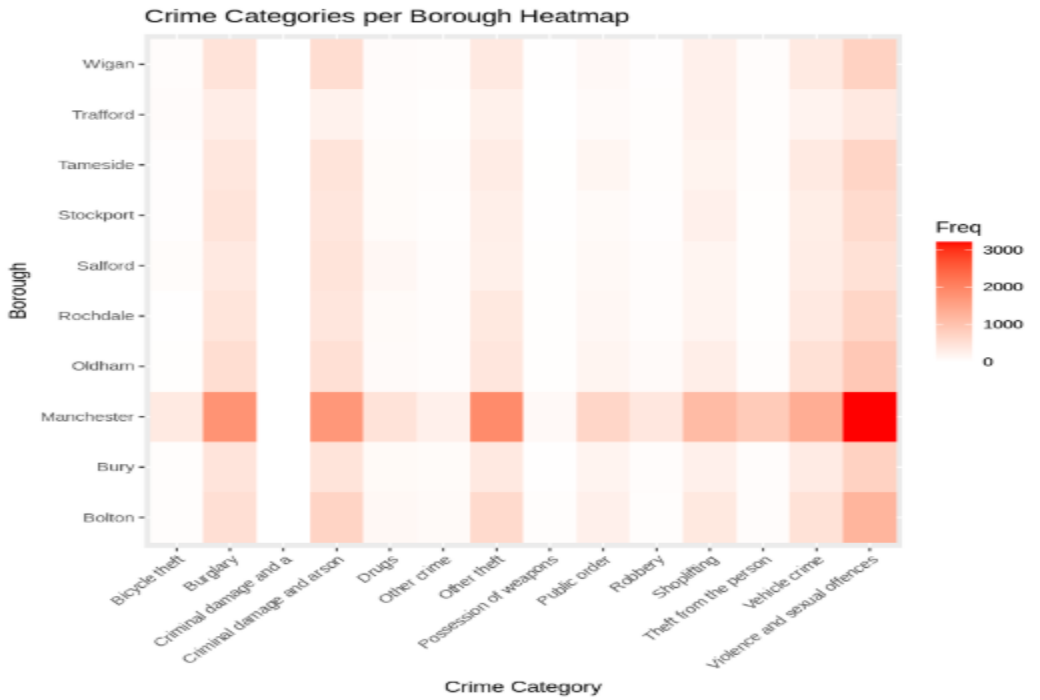

Fig8: *Crime Categories per Borough Heatmap – Heatmap*

7. Recommendations

Based on visual insights:

1. Install CCTV in shopping areas, supermarkets, and parking lots.
2. Increase police presence in Manchester and Bolton.
3. Use month-wise trends to adjust patrols and resource planning.
4. Focus LSOA neighborhoods with high crime for safety programs.
5. Launch public awareness campaigns for common crimes in each borough.

8. Conclusion

- Visualization has transformed complex crime data into clear, actionable knowledge.
- Enables law enforcement to detect patterns, optimize resource use, and engage communities.
- This approach can be repeated for other cities and scaled using automation.

9. References

- Kaggle Dataset: Greater Manchester Police Crime Data
- R Documentation: ggplot2, dplyr, lubridate

10. Important Code Snippets

#Load necessary libraries

library(ggplot2) library(dplyr) library(readr) library(lubridate) library(reshape2)

#**Convert columns**

crime_data$date <- as.Date(crime_data$date, format="%Y-%m-%d") crime_data$borough <- as.factor(crime_data$borough) crime_data$category <- as.factor(crime_data$category)

#Extract time-based features

crime_data$month <- format(crime_data$date, "%m") crime_data$year <- format(crime_data$date, "%Y")

#Check for missing values

colSums(is.na(crime_data))

#Top 10 boroughs with highest crimes

top_boroughs <- crime_data %>% count(borough, sort=TRUE) %>% top_n(10)

#Crime trend over time

```
crime_trend <- crime_data %>% group_by(date) %>% summarise(crime_count = n())
```

#Heatmap: Borough vs Crime Category

```
Crime_heatmap <- table(crime_data$borough, crime_data$category) crime_heatmap_df <-
as.data.frame(crime_heatmap)

ggplot(crime_heatmap_df, aes(Var2, Var1, fill=Freq)) + geom_tile() + labs(title="Crime Categories per Borough
Heatmap", x="Crime Category", y="Borough") + scale_fill_gradient(low="white", high="red") + theme(axis.text.x
= element_text(angle = 45, hjust = 1))
```