Department of Computer Science and Engineering

Course – Advanced Data Visualization (ADV)

Name	Tejal Subhash Komb
UID	2021600037
Batch	A
Lab no	4

Aim :- Create basic charts using R programming language on dataset Crime or Police / Law and Order.

Objectives :-
1. To understand and apply basic data visualization techniques in R.
2. To create various types of charts (Bar chart, Pie chart, Histogram, Timeline chart, Scatter plot, Bubble plot) using a crime-related dataset.
3. To interpret and analyze the data through visual representations.

Dataset :- Crime in India 2001 to 2013

DISTRICT	YEAR	MURDER	ATTEMPT	CULPABLE	RAPE	OTHER RA	KIDNAPPIN	KIDNAPPIN	KIDNAPPIN	DACOITY	PREPARAT	ROBBERY	BURGLARY	THEFT	AUTO THE	OTHER TH	RIOTS	CRIMINAL	CHEATING	COUNTER	ARSON	HURT/GRE
ADILABAD	2001	101	60	17	50	50	46	30	16	9	0	41	198	199	22	177	78	16	104	1	30	1131
ANANTAPU	2001	151	125	1	23	23	53	30	23	8	0	16	191	366	57	309	168	11	65	8	69	1543
CHITTOOR	2001	101	57	2	27	27	59	34	25	4	0	14	237	723	164	559	156	33	209	9	38	2088
CUDDAPAI	2001	80	53	1	20	20	25	20	5	1	0	4	98	173	36	137	164	12	37	2	23	795
EAST GOD	2001	82	67	1	23	23	49	26	23	4	0	25	437	1021	150	871	70	50	220	3	41	1244
GUNTAKAI	2001	3	1	0	0	0	0	0	0	5	0	2	0	162	0	162	1	0	0	3	0	1
GUNTUR	2001	182	88	2	54	54	82	51	31	16	3	59	338	1122	171	951	244	67	300	8	43	1792
HYDERABA	2001	111	113	7	37	37	80	39	41	13	0	67	1155	2792	1128	1664	65	101	1293	24	0	3137
KARIMNA	2001	162	85	6	56	56	67	49	18	27	1	50	218	392	54	338	220	25	243	5	33	1392
KHAMMA	2001	93	60	1	47	47	41	30	11	1	0	13	172	368	34	334	153	35	130	5	73	1026
KRISHNA	2001	65	51	0	37	37	36	21	15	3	0	15	163	478	27	451	70	24	104	1	62	1985
KURNOOL	2001	133	72	4	29	29	47	47	0	6	0	22	155	297	6	291	84	6	126	2	5	1547
MAHABOC	2001	157	67	26	59	59	42	27	15	8	0	27	249	316	33	283	157	22	84	0	0	867
MEDAK	2001	101	56	12	35	35	26	20	6	27	0	26	219	286	36	250	100	17	87	4	37	1367
NALGOND	2001	122	60	1	35	35	27	19	8	6	0	28	133	318	43	275	220	13	122	8	72	1132
NELLORE	2001	89	69	5	46	46	90	80	10	12	2	16	244	608	72	536	97	20	177	3	65	1119
NIZAMABA	2001	106	49	14	21	21	38	21	17	7	0	22	158	234	48	186	51	61	122	1	30	1383
PRAKASHA	2001	102	82	3	19	19	31	12	19	15	0	14	147	278	33	245	138	16	88	1	43	1266
RANGA RE	2001	214	95	16	72	72	106	83	23	24	3	78	1076	1296	347	949	65	67	527	3	67	2829
SECUNDEF	2001	6	0	0	0	0	0	0	0	0	0	10	2	296	0	296	1	2	4	25	0	17
SRIKAKUL	2001	38	10	4	8	8	12	12	0	1	0	4	118	231	1	230	70	18	53	0	34	679
VIJAYAWA	2001	53	44	5	25	25	70	48	22	3	0	27	491	2057	264	1793	19	34	614	10	17	1578
VIJAYAWA	2001	2	1	0	1	1	0	0	0	0	0	1	0	265	0	265	1	2	3	6	0	9
VISAKHA R	2001	58	29	0	12	12	12	12	0	4	0	3	76	165	0	165	138	19	39	3	31	1476
VISAKHAP	2001	22	10	1	13	13	13	6	7	1	0	5	323	630	172	458	9	37	192	3	15	325
VIZIANAG	2001	33	14	1	8	8	8	2	6	0	0	2	99	144	0	144	36	11	41	2	12	529
WARANGA	2001	158	79	5	53	53	81	25	56	2	0	23	266	418	79	339	330	6	137	2	0	1718
WEST GOD	2001	77	58	1	61	61	41	21	20	7	0	15	257	1116	74	1042	96	63	270	2	32	972
TOTAL	2001	2602	1555	136	871	871	1182	765	417	214	9	629	7220	16751	3051	13700	3001	788	5391	144	872	34947
CHANGLAI	2001	11	2	0	1	1	3	2	1	3	2	4	18	24	0	24	0	4	3	0	1	30
DIBANG V	2001	3	5	0	2	2	4	4	0	2	0	5	18	19	1	18	0	3	0	0	0	27

Dataset description :- The dataset you provided contains crime statistics for various districts in India for the year 2001. It includes different types of crimes such as murder, attempt to murder, rape, kidnapping, theft, and more. Each row represents a district, and each column represents a specific type of crime, with the last column showing the total number of Indian Penal Code (IPC) crimes reported in that district for the year.

Department of Computer Science and Engineering

Implementation :- Implementation in R programming

```r
library(ggplot2)
library(dplyr)
library(scales)
library(tidyr)

crime_data <- read.csv("D:/Blog/R programming/Crimes_in_india.csv")

# Summarize total IPC crimes by state
top_states <- crime_data %>%
  group_by(STATE.UT) %>%
  summarize(Total_IPC_Crimes = sum(TOTAL.IPC.CRIMES)) %>%
  arrange(desc(Total_IPC_Crimes)) %>%
  top_n(10)


# Bar plot with formatted y-axis
ggplot(top_states, aes(x = reorder(STATE.UT, -Total_IPC_Crimes), y = Total_IPC_Crimes)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  coord_flip() +
  scale_y_continuous(labels = scales::comma) +  # This will format large numbers with commas
  labs(title = "Top 10 States by Total IPC Crimes", x = "State/UT", y = "Total IPC Crimes") +
  theme_minimal()

district_data <- crime_data %>% filter(DISTRICT == "ADILABAD")

# Summarize major crime categories for the pie chart
crime_categories <- district_data %>%
  summarise(Murder = sum(MURDER),
            Rape = sum(RAPE),
            Kidnapping = sum(KIDNAPPING...ABDUCTION),
            Theft = sum(THEFT),
            Riots = sum(RIOTS))

# Reshape the data using pivot_longer instead of gather
crime_categories_melt <- pivot_longer(crime_categories,
                                       cols = everything(),
                                       names_to = "Crime_Type",
                                       values_to = "Count")

# Pie chart
ggplot(crime_categories_melt, aes(x = "", y = Count, fill = Crime_Type)) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start = 0) +
  labs(title = "Crime Proportion in ADILABAD District") +
  theme_void()


# Filter out zero theft values to focus on districts with theft cases
crime_non_zero_theft <- crime_data %>% filter(THEFT > 0)
```

```r
52  optimal_binwidth_theft <- 500   # Larger binwidth for thicker bars
53
54  # Increase the overall plot size
55  options(repr.plot.width = 10, repr.plot.height = 6)  # Adjust the width and height of the plot
56
57  # Histogram with larger binwidth for thicker bars
58  ggplot(crime_non_zero_theft, aes(x = THEFT)) +
59    geom_histogram(binwidth = optimal_binwidth_theft, fill = "dodgerblue", color = "black") +
60    labs(title = "Distribution of Non-Zero Theft Cases Across Districts", x = "Number of Theft Cases", y = "Count of Districts") +
61    theme_minimal() +
62    theme(plot.title = element_text(size = 16),  # Increase title size
63          axis.title.x = element_text(size = 14),  # Increase x-axis label size
64          axis.title.y = element_text(size = 14))  # Increase y-axis label size
65
66
67  # Filter data for a specific district (e.g., "ADILABAD")
68  district_trend <- crime_data %>%
69    filter(DISTRICT == "ADILABAD") %>%
70    group_by(YEAR) %>%
71    summarize(Total_Crimes = sum(TOTAL.IPC.CRIMES))
72
73  # Line plot (Timeline)
74  ggplot(district_trend, aes(x = YEAR, y = Total_Crimes)) +
75    geom_line(color = "blue") +
76    geom_point() +
77    labs(title = "Crime Trend in ADILABAD District Over Years", x = "Year", y = "Total Crimes") +
78    theme_minimal()
79
80  # Scatter plot of murder vs kidnapping cases
81  ggplot(crime_data, aes(x = MURDER, y = KIDNAPPING...ABDUCTION)) +
82    geom_point(color = "darkgreen", alpha = 0.6) +
83    labs(title = "Murder vs Kidnapping Cases", x = "Murder Cases", y = "Kidnapping & Abduction Cases") +
84    theme_minimal()
85
86  # Prepare data for bubble plot (Top 10 states)
87  bubble_data <- crime_data %>%
88    group_by(STATE.UT) %>%
89    summarize(Murder = sum(MURDER),
90              Rape = sum(RAPE),
91              Theft = sum(THEFT),
92              Total_Crimes = sum(TOTAL.IPC.CRIMES)) %>%
93    top_n(10, Total_Crimes)
94
95  # Bubble plot (Size of points represents total crimes)
96  ggplot(bubble_data, aes(x = Murder, y = Rape, size = Theft, color = STATE.UT)) +
97    geom_point(alpha = 0.7) +
98    labs(title = "Bubble Plot of Major Crimes in Top 10 States", x = "Murder Cases", y = "Rape Cases", size = "Theft Cases") +
99    theme_minimal()
```
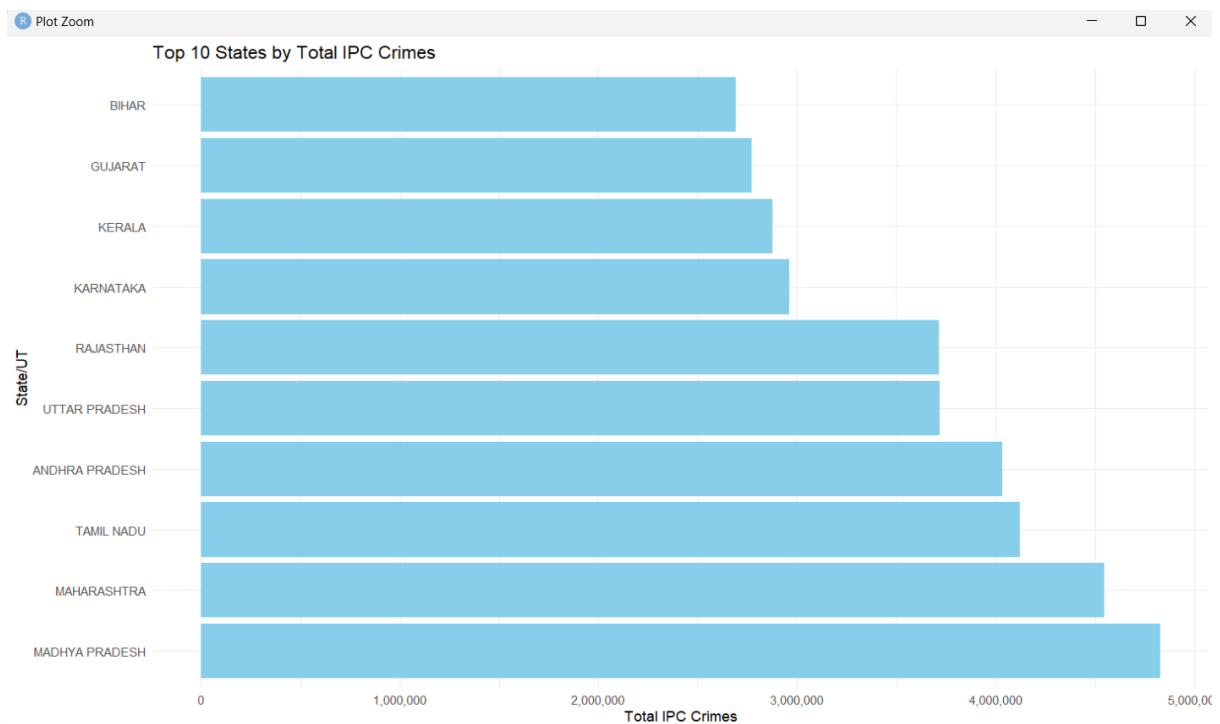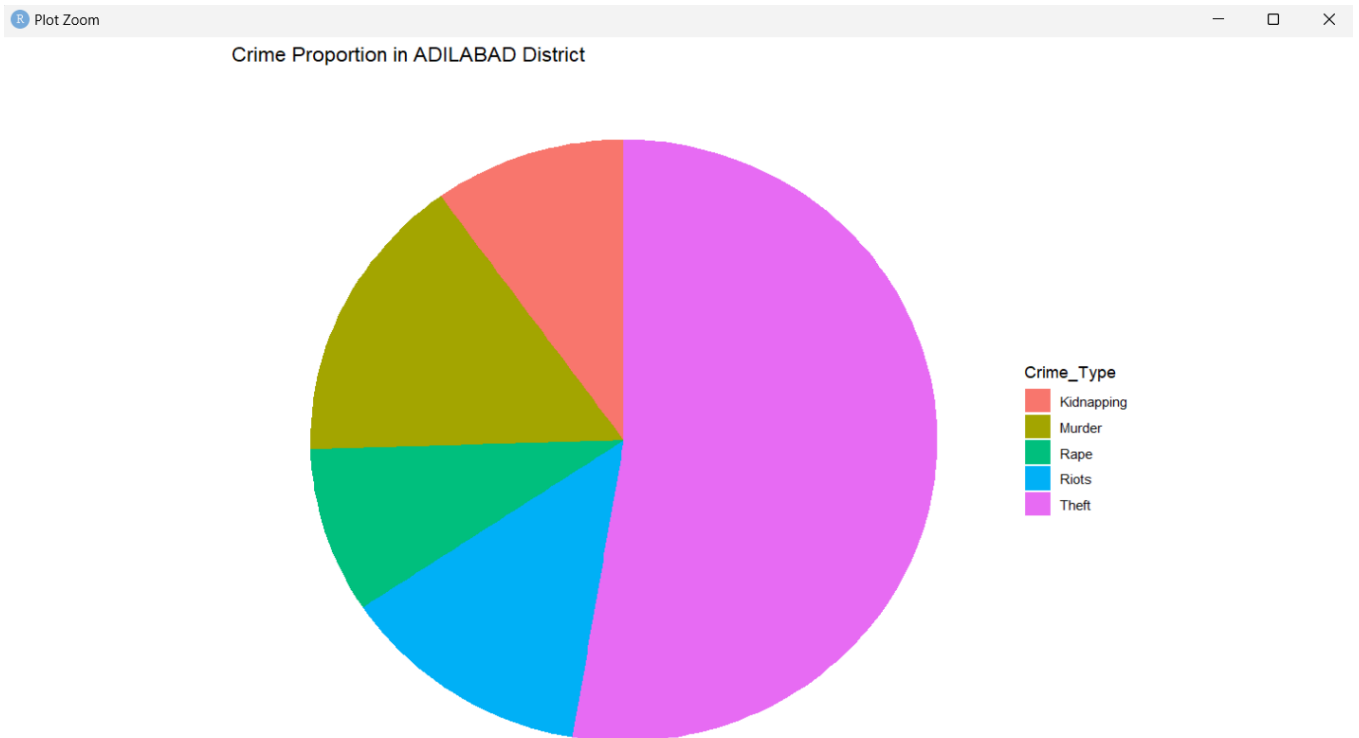
99:18   (Top Level) ‡

Console

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

Output :-

1) **Bar chart - Top 10 states by total IPC crimes :-**

Observation: The bar plot shows the top 10 states/UTs with the highest total IPC crimes. States like Maharashtra, Uttar Pradesh, and Madhya Pradesh are among the top contributors to the total IPC crimes. This indicates higher crime rates in these states compared to others.
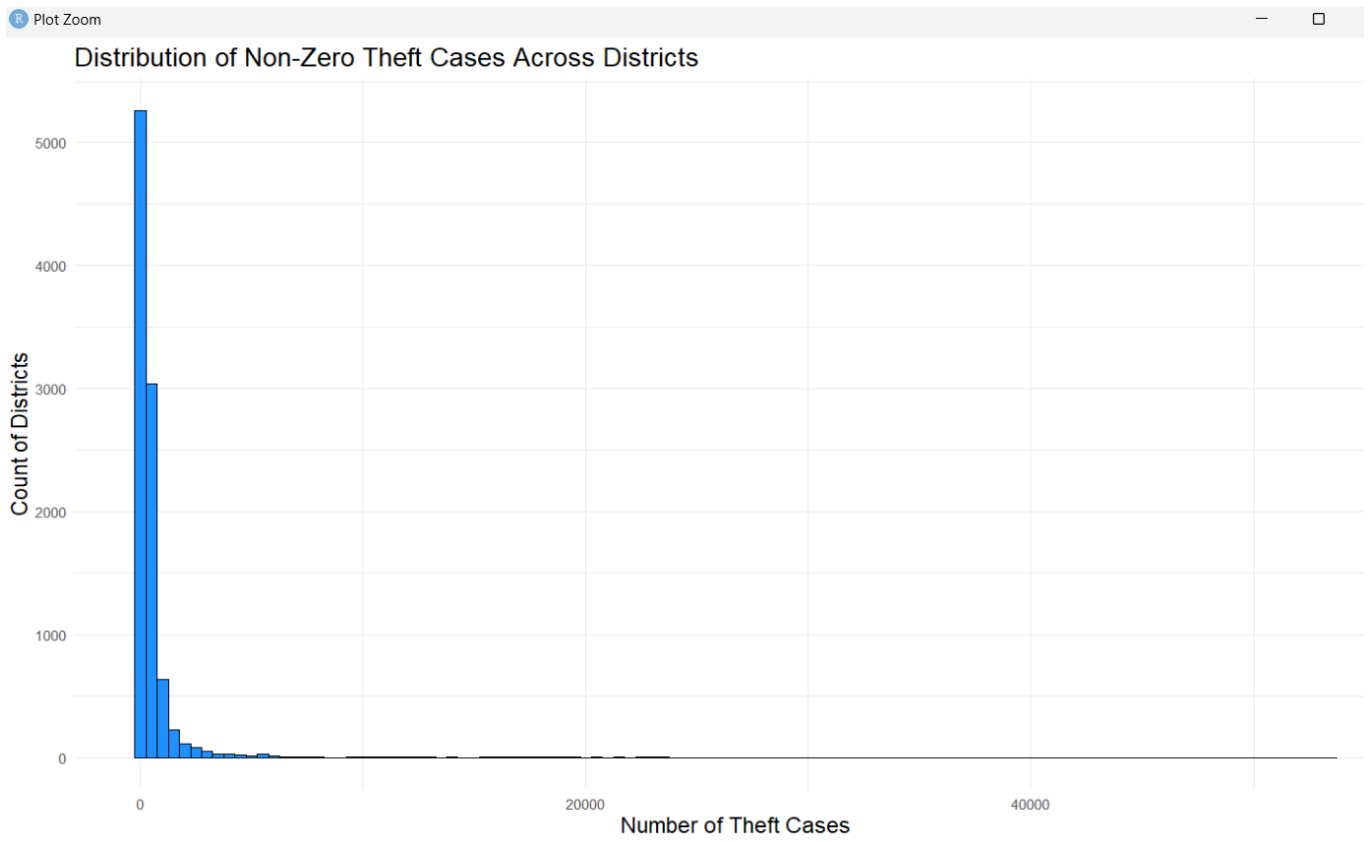
BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

2) Pie chart - Proportion of different crimes in a district :-

Observation: The pie chart illustrates the proportion of different crime categories in the Adilabad district. Theft and murder are significant contributors, indicating these are the most prevalent crimes in this district.

Crime Proportion in ADILABAD District

Department of Computer Science and Engineering

3) Histogram - Distribution of murder cases across districts :-

Observation: The histogram shows the distribution of theft cases across districts with non-zero theft values. Most districts have a lower number of theft cases, but there are a few districts with very high theft counts, indicating a skewed distribution.
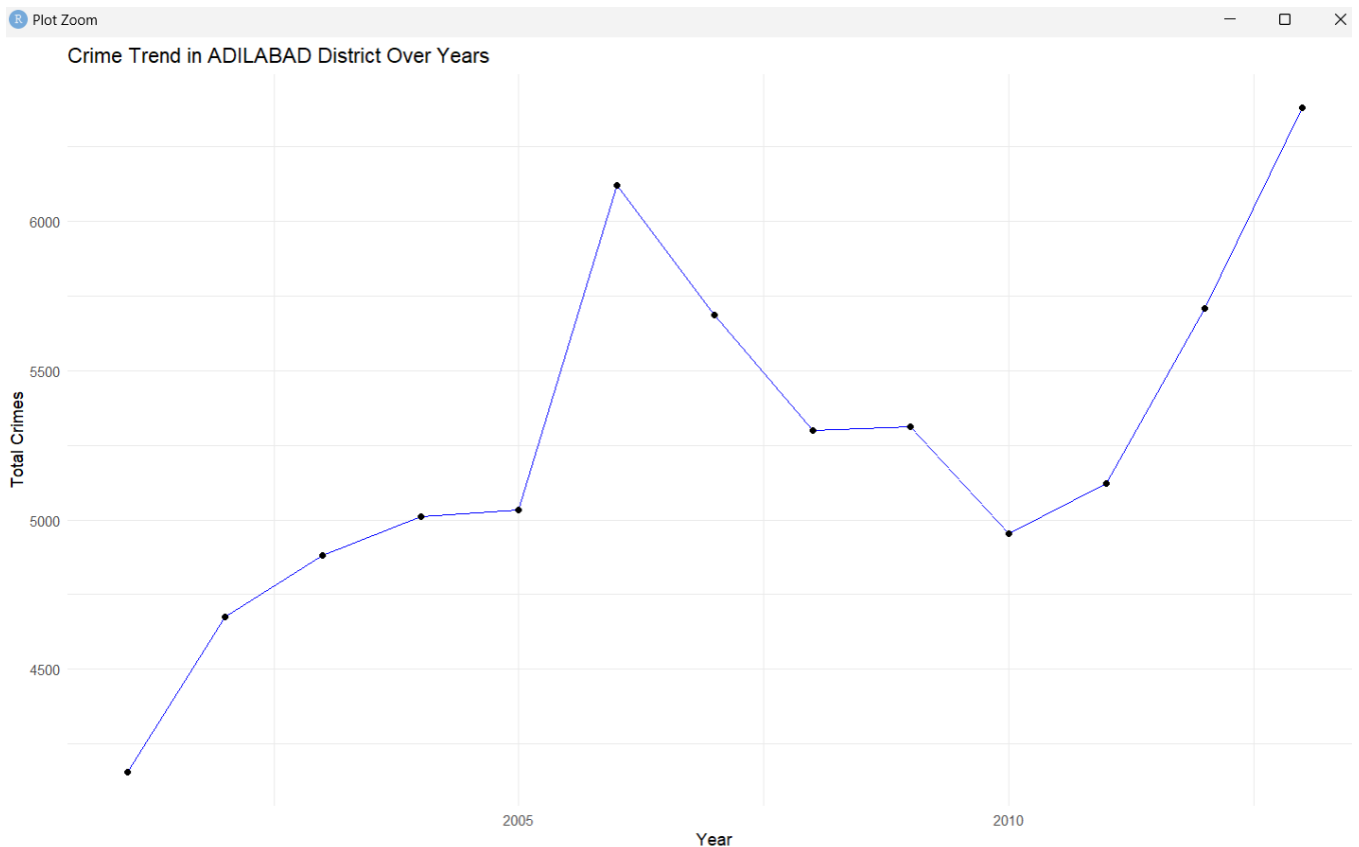

Distribution of Non-Zero Theft Cases Across Districts

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

4) Timeline chart - Trend of total crimes over the years in a top district

Observation: The line plot shows the trend of total crimes in Adilabad district over the years. There are fluctuations in the crime rates, with some years showing significant increases or decreases. This can help identify periods of higher crime activity and potential causes.
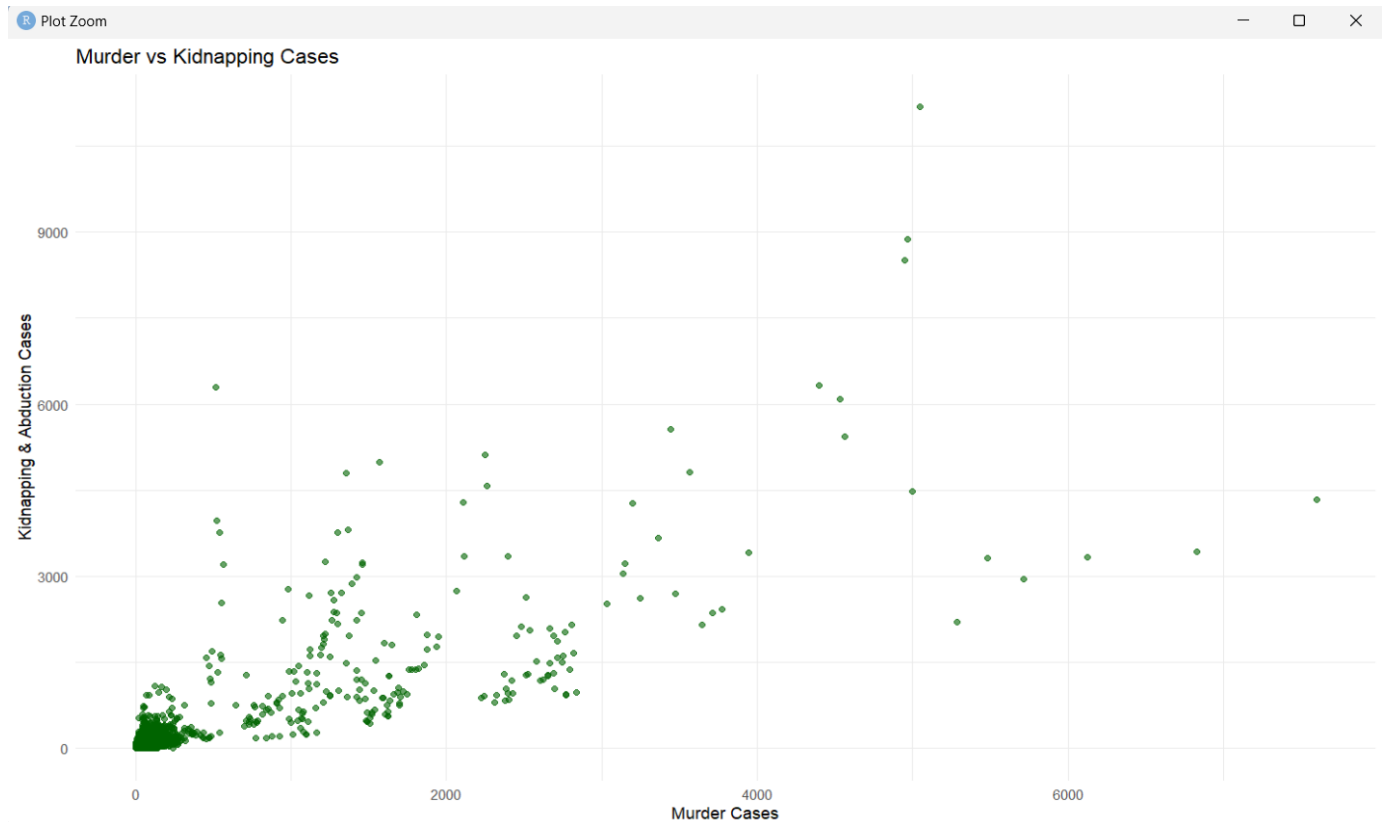
BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

5) Scatter plot - Murder vs Kidnapping across districts :-

Observation: The scatter plot shows the relationship between murder and kidnapping cases across districts. There is no clear linear relationship, but some districts with high murder cases also have high kidnapping cases, indicating possible correlations in certain areas.
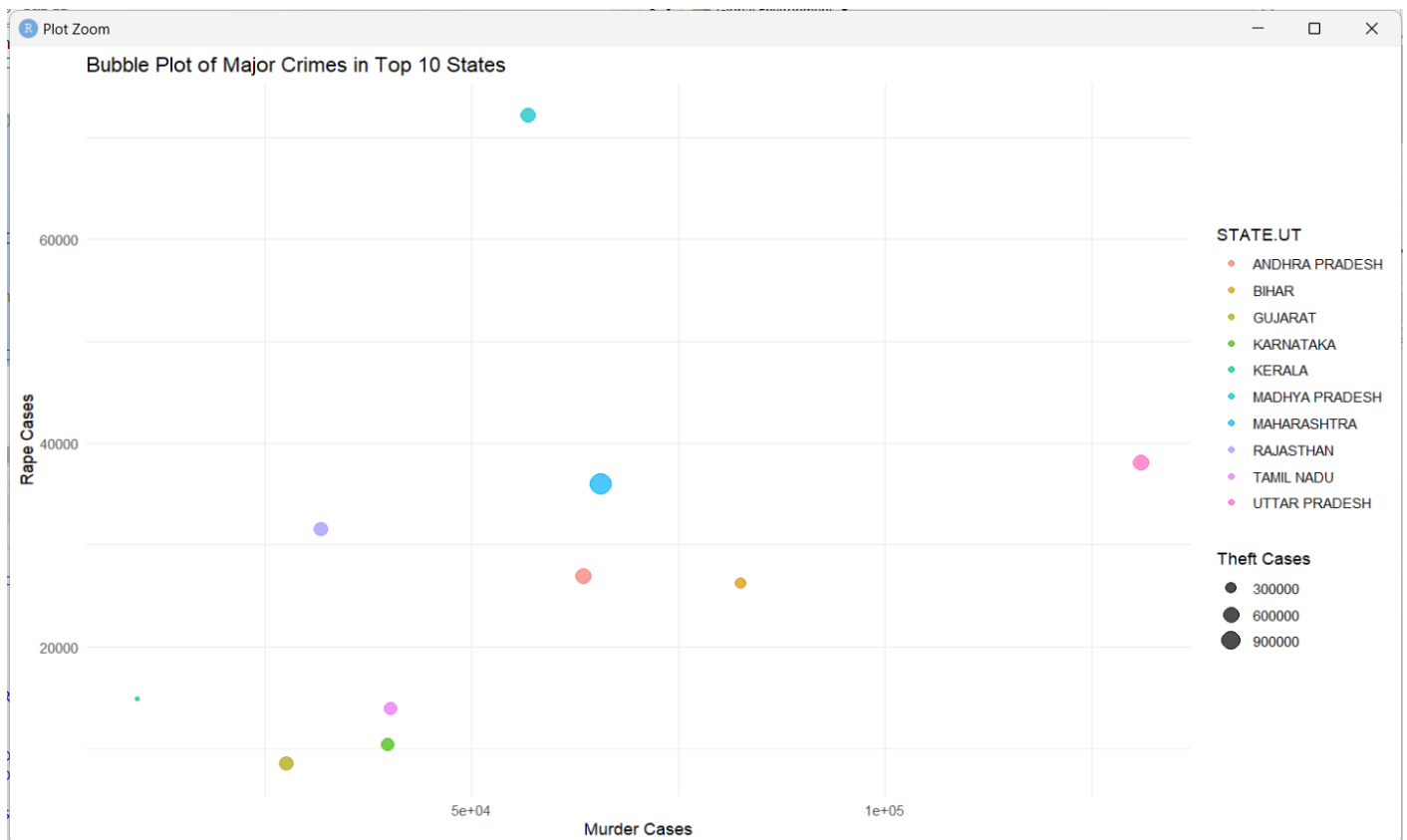
BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

6) Bubble plot - Multiple crimes in top states :-

Observation: The bubble plot shows the relationship between murder, rape, and theft cases in the top 10 states by total crimes. The size of the bubbles represents the number of theft cases. States like Maharashtra and Uttar Pradesh have large bubbles, indicating high theft cases along with significant murder and rape cases.

BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400093-India
(Autonomous College Affiliated to University of Mumbai)

Department of Computer Science and Engineering

Conclusion :- In this experiment we aimed to visualize crime data from various districts in India using R programming. The results showed that states like Maharashtra and Uttar Pradesh have the highest crime rates. In Adilabad district, theft and murder are the most common crimes. The distribution of theft cases is uneven, with some districts having very high numbers. Crime trends in Adilabad fluctuate over the years, indicating periods of higher and lower crime activity. There is no clear relationship between murder and kidnapping cases, but some districts have high numbers of both.