Name : Nathan Jude Cardoso **UID** 2021600009 **CSE(AI/ML)**

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

DATASET: Crime Dataset India: https://www.kaggle.com/datasets/amisha0528/crime-dataset-india/data

The dataset contains crime statistics for various states/UTs in India, detailing incidents of different crime types such as murder, rape, theft, and assault across multiple years. Each row represents crime data for a specific state and year, with columns indicating the number of occurrences for each type of crime.

ANALYSIS:

```
crime_data <- read.csv("/kaggle/input/crime-dataset-india/newtrial - Sheet 1 - 01_District_wise_crim 2.csv")
```

```
crime_data <- crime_data %>% na.omit()
```

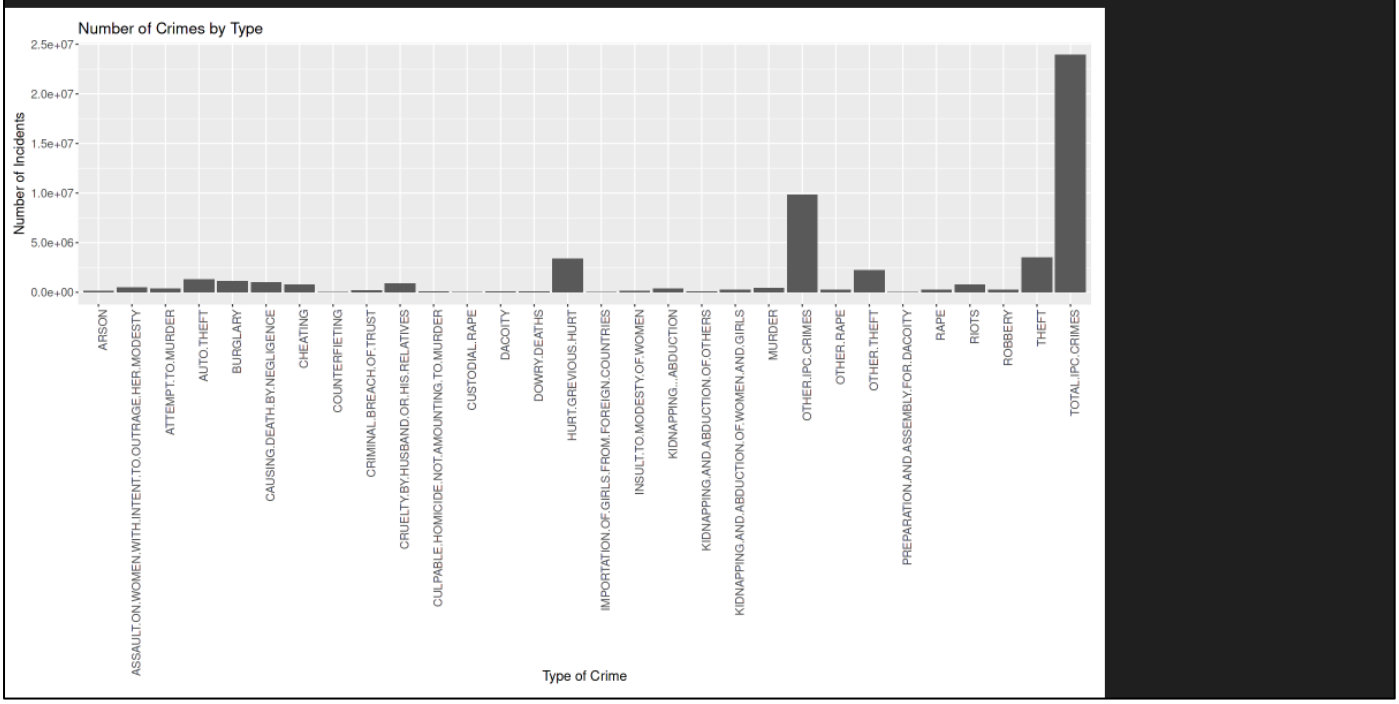
```
crime_long <- pivot_longer(crime_data,
                           cols = MURDER:TOTAL.IPC.CRIMES,
                           names_to = "Crime_Type",
                           values_to = "Count")
head(crime_long)
```

A tibble: 6 × 4

STATE.UT	YEAR	Crime_Type	Count
<chr>	<int>	<chr>	<int>
A & N ISLANDS	2001	MURDER	13
A & N ISLANDS	2001	ATTEMPT.TO.MURDER	0
A & N ISLANDS	2001	CULPABLE.HOMICIDE.NOT.AMOUNTING.TO.MURDER	0
A & N ISLANDS	2001	RAPE	3
A & N ISLANDS	2001	CUSTODIAL.RAPE	0
A & N ISLANDS	2001	OTHER.RAPE	3

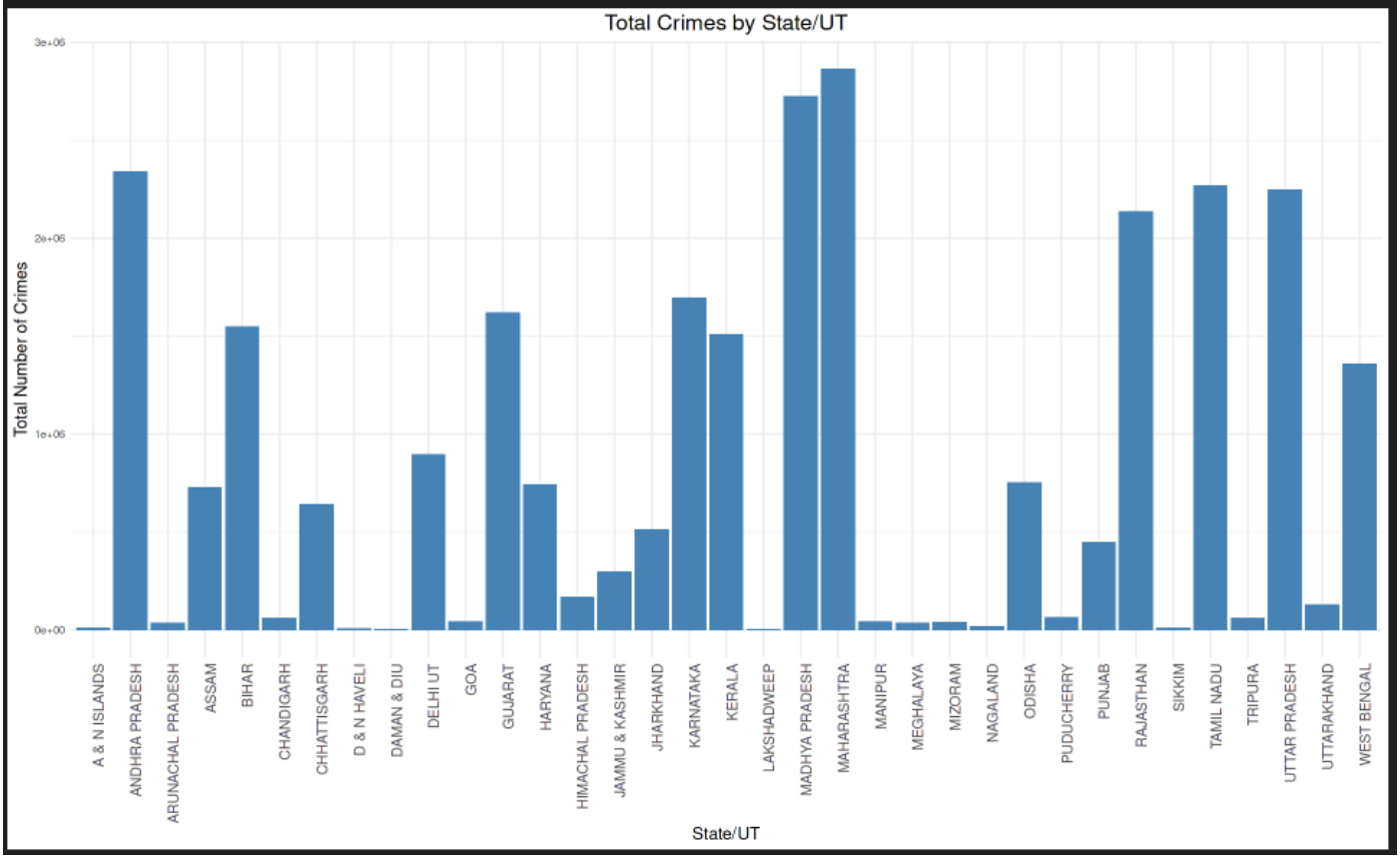
1] Bar Graph

```
options(repr.plot.width=16, repr.plot.height=10)
ggplot(crime_long, aes(x = Crime_Type, y = Count)) + geom_bar(stat = "identity") + ggtitle("Number of Crimes by Type") +
  xlab("Type of Crime") + ylab("Number of Incidents") + theme(axis.text.x = element_text(angle = 90, hjust = 1,size = 12 ),
  axis.text.y = element_text(size = 12), axis.title = element_text(size = 14), plot.title = element_text(size = 16), plot.margin = margin(15, 15, 15, 15))
```



The above bar graph shows us that Other IPC Crimes and Hurt Crime are the most frequent crimes that take place in India.

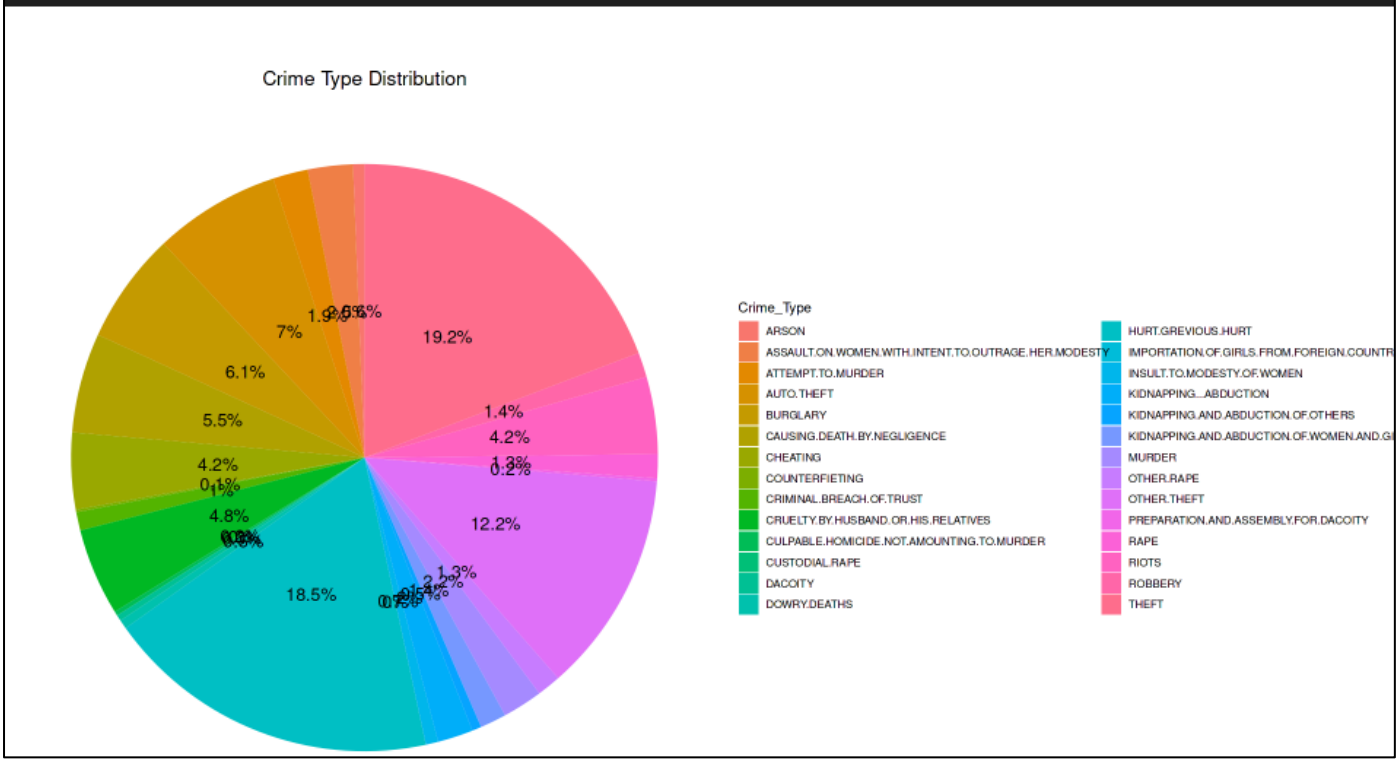
```
ggplot(crime_data, aes(x = `STATE.UT`, y = Crime_Severity)) + geom_bar(stat = "identity", fill = "steelblue") +
  ggtitle("Total Crimes by State/UT") + xlab("State/UT") + ylab("Total Number of Crimes") + theme_minimal() +
  theme( axis.text.x = element_text(angle = 90, hjust = 1, size = 12), axis.title.x = element_text(size = 14),
  axis.title.y = element_text(size = 14), plot.title = element_text(size = 18, hjust = 0.5) )
```



The above Bar graph shows that only a few states have high crime count as compared to other states. Here Maharashtra and Madhya Pradesh have the highest crime count.

2] Pie Chart

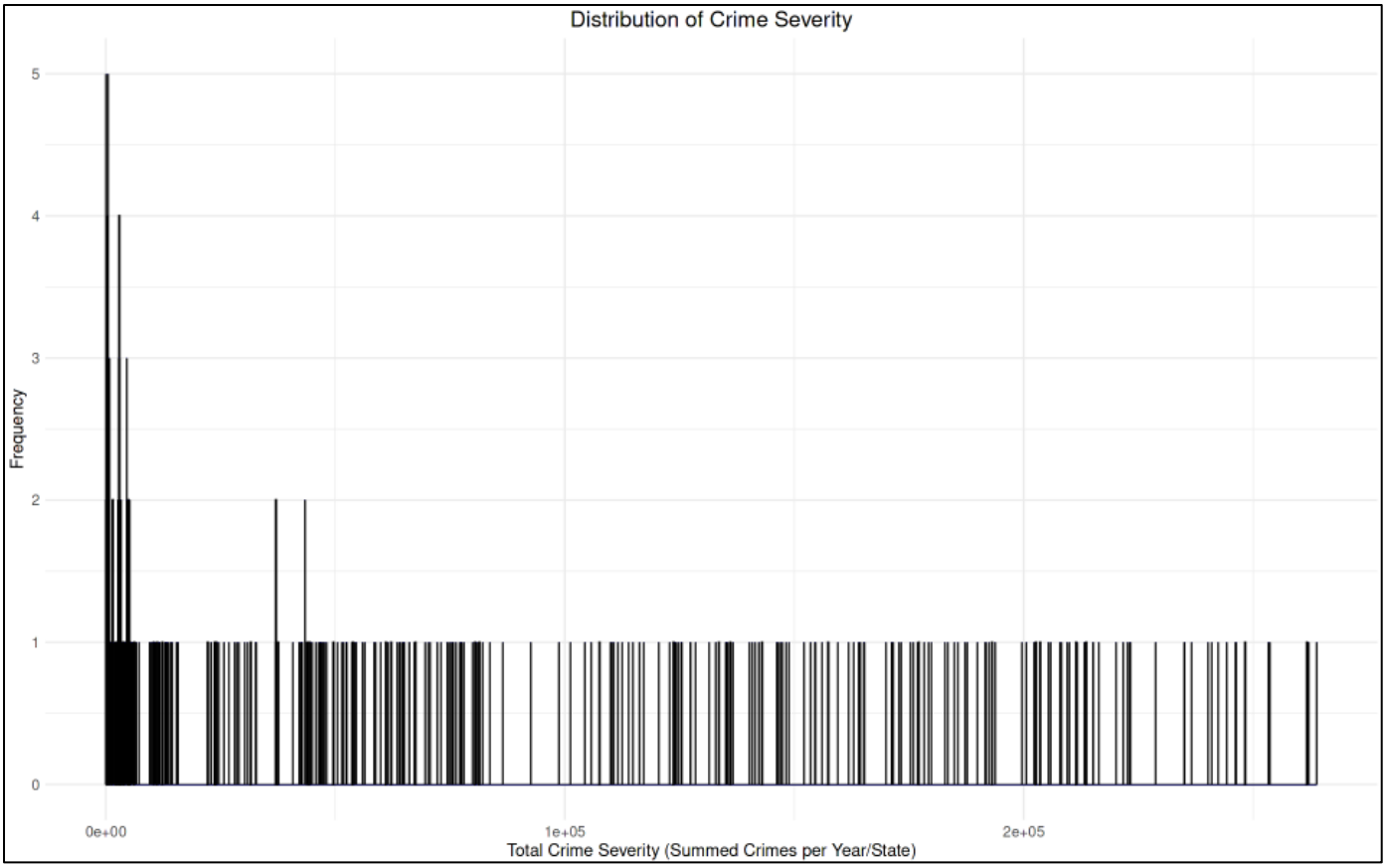
```
crime_long <- crime_data %>% pivot_longer(cols = -c(`STATE.UT`, YEAR, `TOTAL.IPC.CRIMES`,`OTHER.IPC.CRIMES`),
                                          names_to = "Crime_Type", values_to = "Count")
crime_summary <- crime_long %>% group_by(Crime_Type) %>%
  summarise(Count = sum(Count, na.rm = TRUE)) %>% mutate(Percentage = Count / sum(Count) * 100)
ggplot(crime_summary, aes(x = "", y = Count, fill = Crime_Type)) +
  geom_bar(width = 1, stat = "identity") + coord_polar("y") + ggtitle("Crime Type Distribution") +
  theme_void() + geom_text(aes(label = paste0(round(Percentage, 1), "%")),
  position = position_stack(vjust = 0.5), size = 5) + theme(plot.title = element_text(hjust = 0.5, size = 16))
```



The above pie chart shows that out of the total crimes that occur in India Theft covers about 19.2% and Dowry Deaths cover about 18.5%.

3] Histogram

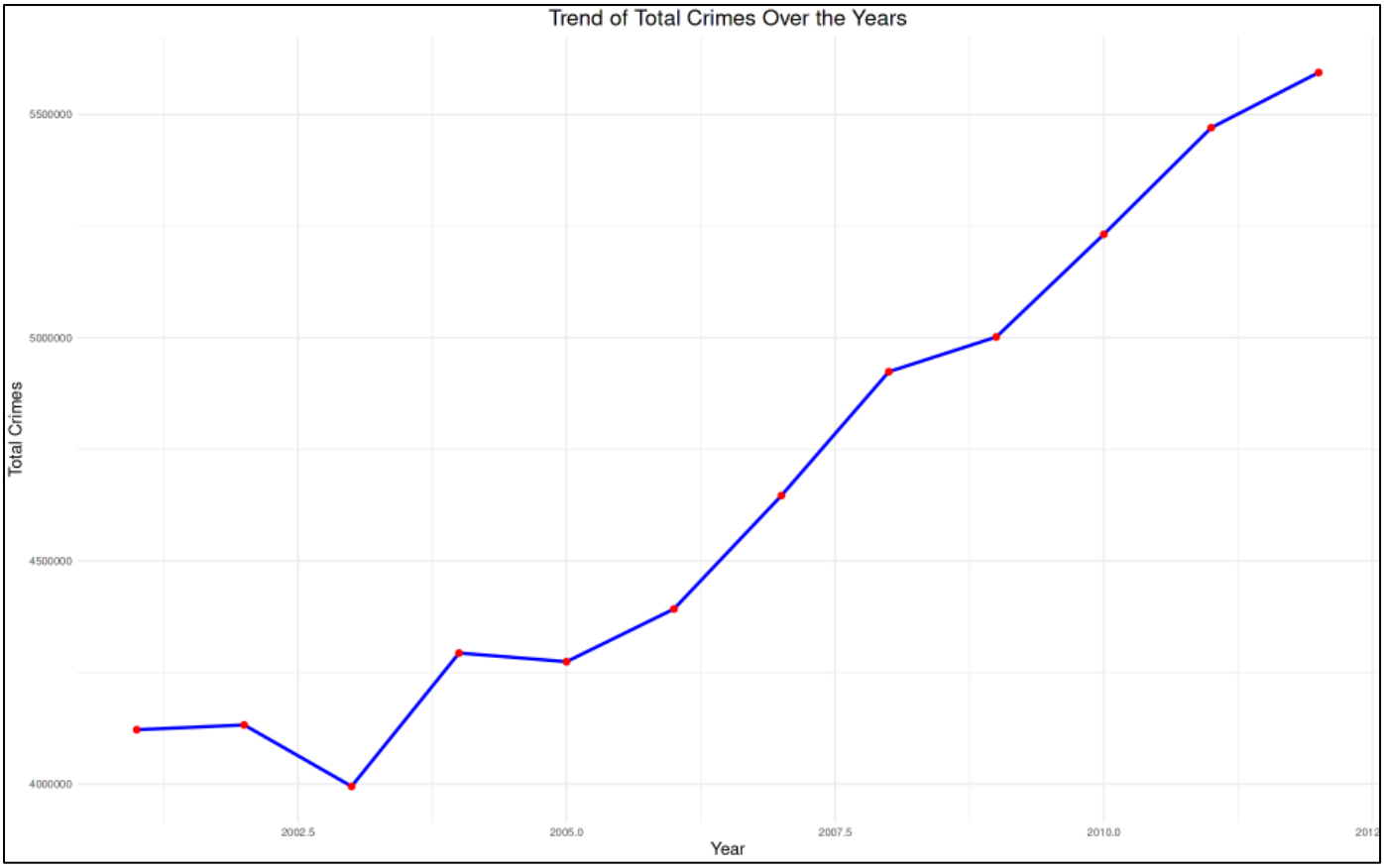
```
crime_data <- crime_data %>% rowwise() %>%
  mutate(Crime_Severity = sum(c_across(-c(`STATE.UT`, YEAR, `TOTAL.IPC.CRIMES`)), na.rm = TRUE))
ggplot(crime_data, aes(x = Crime_Severity)) +
  geom_histogram(binwidth = 20, fill = "blue", color = "black") +
  ggtitle("Distribution of Crime Severity") +
  xlab("Total Crime Severity (Summed Crimes per Year/State)") +
  ylab("Frequency") +
  theme_minimal() +
  theme(
    axis.title.x = element_text(size = 14),
    axis.title.y = element_text(size = 14),
    axis.text.x = element_text(size = 12),
    axis.text.y = element_text(size = 12),
    plot.title = element_text(size = 18, hjust = 0.5)
  )
```



The above histogram states that there are only a few states/year that have the highest crime count as compared to other regions/years.

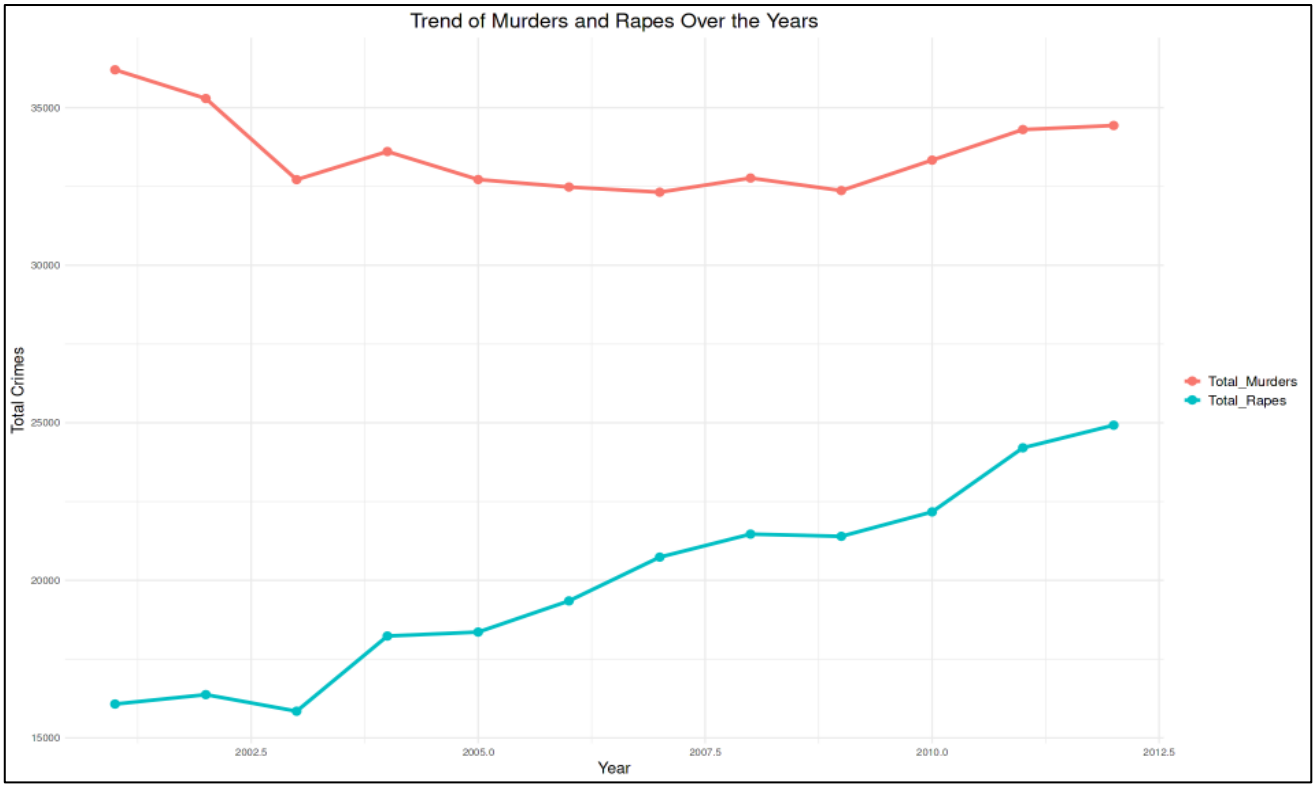
4] Timeline Chart

```
crime_data <- crime_data %>% rowwise() %>%
  mutate(Crime_Severity = sum(c_across(-c(`STATE.UT`, YEAR, `TOTAL.IPC.CRIMES`)), na.rm = TRUE))

crime_trend <- crime_data %>% group_by(YEAR) %>% summarise(Total_Crimes = sum(Crime_Severity))

ggplot(crime_trend, aes(x = YEAR, y = Total_Crimes)) +
  geom_line(color = "blue", size = 1) +
  geom_point(color = "red", size = 2) +
  ggtitle("Trend of Total Crimes Over the Years") +
  xlab("Year") +
  ylab("Total Crimes") +
  theme_minimal() +
  theme(
    axis.title.x = element_text(size = 14),
    axis.title.y = element_text(size = 14),
    plot.title = element_text(size = 18, hjust = 0.5)
  )
```



The above graph shows that the crime rates has been increasing over the years. The only time the count was low was in the year 2003.

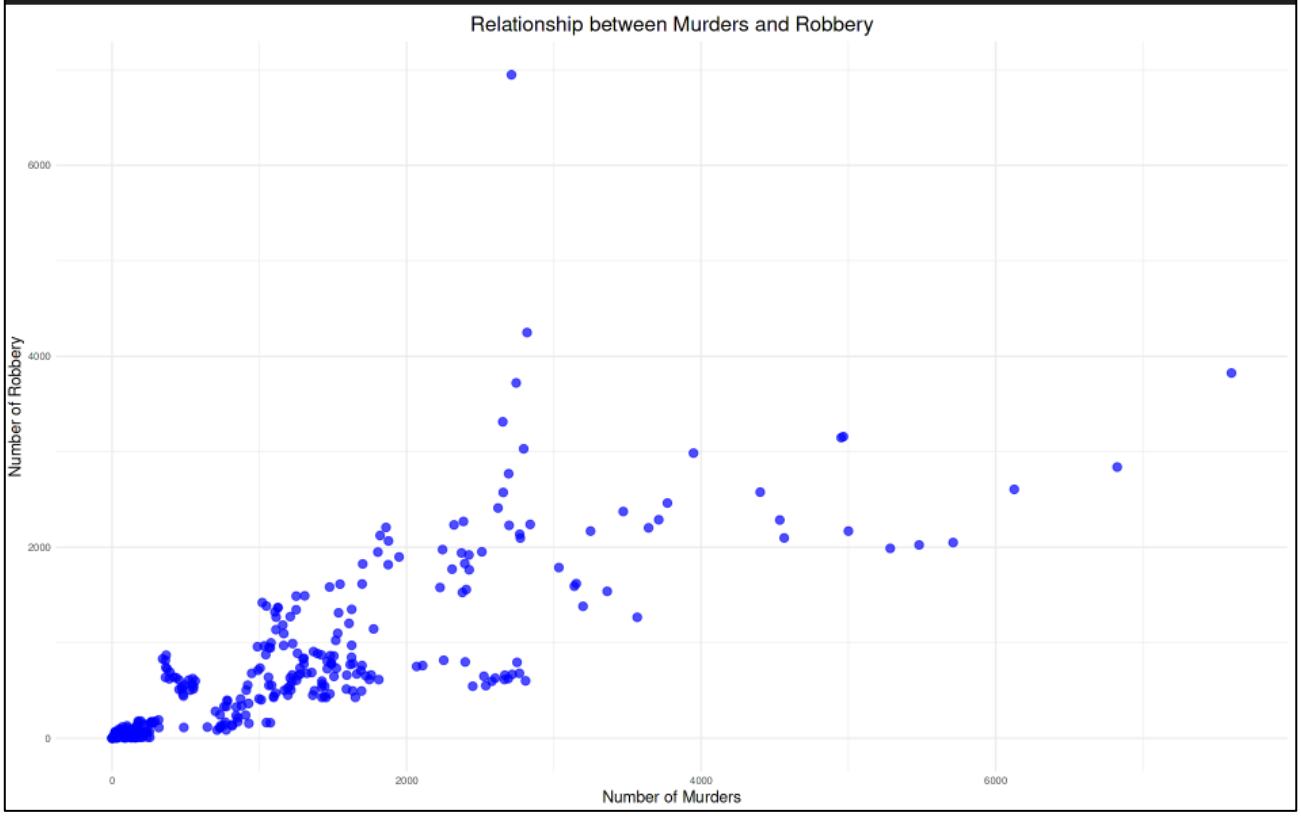
```r
crime_comparison <- crime_data %>%
  select(`STATE.UT`, YEAR, MURDER, RAPE) %>%
  group_by(YEAR) %>%
  summarise(
    Total_Murders = sum(MURDER, na.rm = TRUE),
    Total_Rapes = sum(RAPE, na.rm = TRUE)
  )

crime_comparison_long <- crime_comparison %>%
  pivot_longer(cols = c(Total_Murders, Total_Rapes),
               names_to = "Crime_Type",
               values_to = "Total_Crimes")

ggplot(crime_comparison_long, aes(x = YEAR, y = Total_Crimes, color = Crime_Type)) +
  geom_line(size = 1.2) +
  geom_point(size = 3) +
  ggtitle("Trend of Murders and Rapes Over the Years") +
  xlab("Year") +
  ylab("Total Crimes") +
  theme_minimal() +
  theme(
    axis.title.x = element_text(size = 14),
    axis.title.y = element_text(size = 14),
    plot.title = element_text(size = 18, hjust = 0.5),
    legend.title = element_blank(),
    legend.text = element_text(size = 12),
    plot.margin = margin(10, 10, 10, 10)
  )
```



The above timeline graphs shows that the count of murders over the years have always been higher than the rapes in India but the rate of growth of rapes are higher than murders.

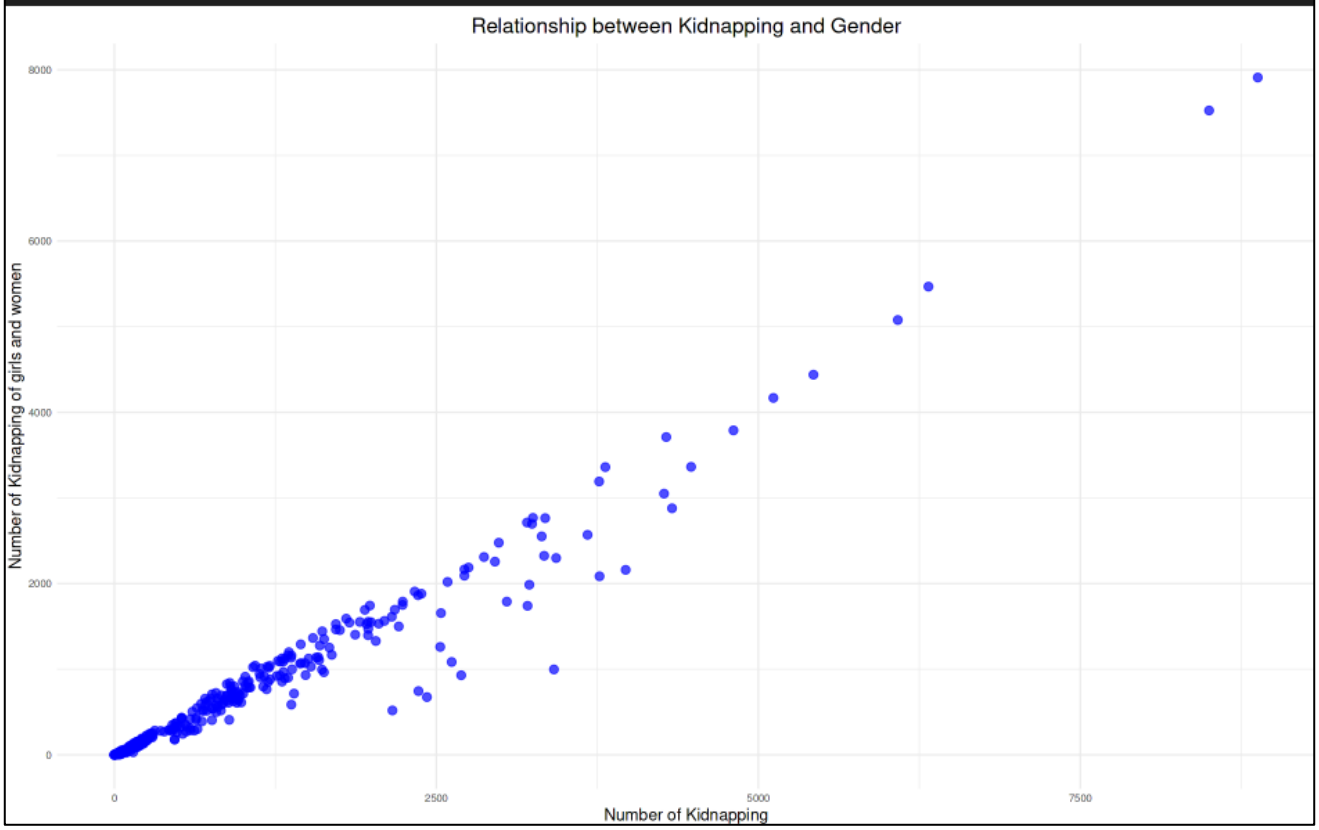
5] Scatterplot

```
ggplot(crime_data, aes(x = MURDER, y = ROBBERY)) + geom_point(color = "blue", size = 3, alpha = 0.7) +
  ggtitle("Relationship between Murders and Robbery") + xlab("Number of Murders") + ylab("Number of Robbery") +
  theme_minimal() + theme( axis.title.x = element_text(size = 14), axis.title.y = element_text(size = 14),
  plot.title = element_text(size = 18, hjust = 0.5), plot.margin = margin(10, 10, 10, 10) )
```



The above scatterplot shows that there is a relationship between Robbery and Murders and most Robberies results into Murders.

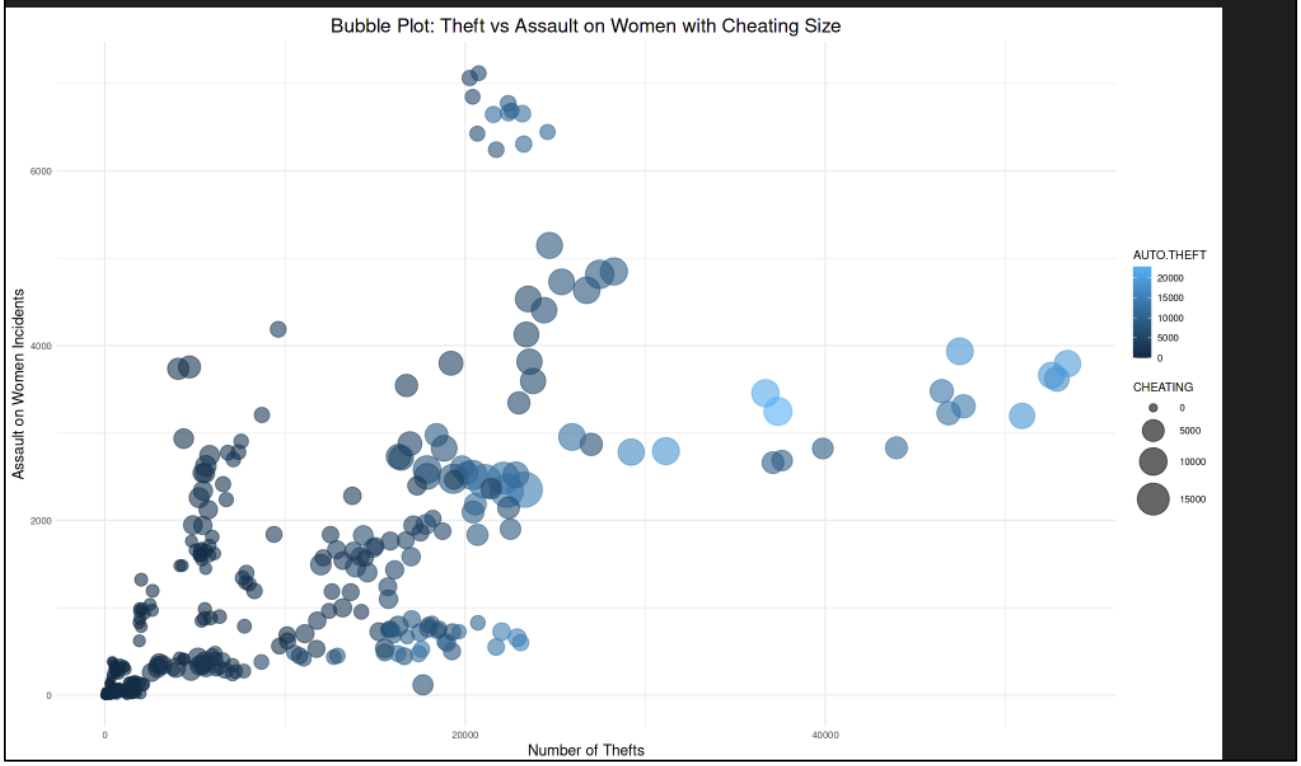
```
ggplot(crime_data, aes(x = KIDNAPPING...ABDUCTION, y = KIDNAPPING.AND.ABDUCTION.OF.WOMEN.AND.GIRLS)) +
  geom_point(color = "blue", size = 3, alpha = 0.7) + ggtitle("Relationship between Kidnapping and Gender") +
  xlab("Number of Kidnapping") + ylab("Number of Kidnapping of girls and women") +
  theme_minimal() + theme( axis.title.x = element_text(size = 14), axis.title.y = element_text(size = 14)
```



The above graph shows that there is a linear relationship between the Kidnapping and Kidnapping of girls and women which indicates that most kidnapping incidents occur with girls and women.

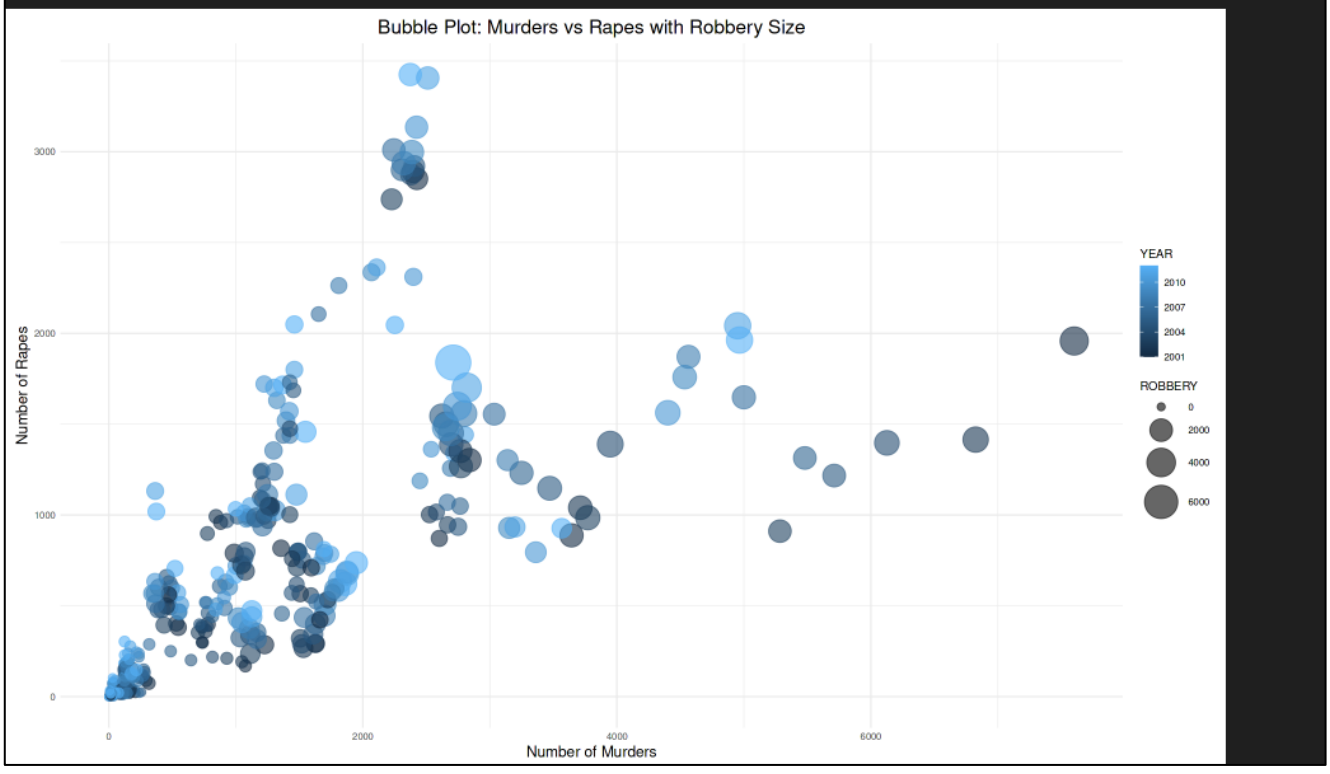
6] Bubble plot

```
ggplot(crime_data, aes(x =THEFT, y =`ASSAULT.ON.WOMEN.WITH.INTENT.TO.OUTRAGE.HER.MODESTY`, size=CHEATING, color =`AUTO.THEFT`))
+ geom_point(alpha = 0.6) + scale_size(range = c(3, 15)) + ggtitle("Bubble Plot: Theft vs Assault on Women with Cheating Size") +
xlab("Number of Thefts") + ylab("Assault on Women Incidents") + theme_minimal() + theme( axis.title.x = element_text(size = 14),
axis.title.y = element_text(size = 14),plot.title = element_text(size = 18, hjust = 0.5), plot.margin = margin(10, 10, 10, 10))
```



The above graph shows that the Number of Thefts and Assault on women are related to Cheating as the graph shows bubbles mostly in the centre.

```
ggplot(crime_data, aes(x = MURDER, y = RAPE, size = ROBBERY, color = YEAR)) + geom_point(alpha = 0.6) + scale_size(range = c(3, 15))
+ ggtitle("Bubble Plot: Murders vs Rapes with Robbery Size") + xlab("Number of Murders") + ylab("Number of Rapes") + theme_minimal()
+ theme( axis.title.x = element_text(size = 14), axis.title.y = element_text(size = 14),
    plot.title = element_text(size = 18, hjust = 0.5), plot.margin = margin(10, 10, 10, 10))
```



The above graph shows that the Number of Murders and Number of Rapes are related for Robbery which states that most Robbery leads to Rapes or Murders.