NAME:	Shubham Vishwakarma
UID No.	2021700071
BRANCH:	B.E CSE-DS
BATCH:	D
SUBJECT	Advanced Data Visualization
EXPERIMENT No.	4

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

DATASET: https://www.kaggle.com/datasets/balajivaraprasad/crimes-against-women-in-india-2001-2021?select=CrimesOnWomenData.csv

The dataset contains crime statistics across various Indian states for the year 2001, focusing on different categories of crimes against women, including Women Trafficking (WT), Assault on Women (AoW), Dowry Deaths (DD), and Domestic Violence (DV). It provides a comprehensive view of crime patterns and severity, allowing for trend analysis and regional comparisons.

ANALYSIS:

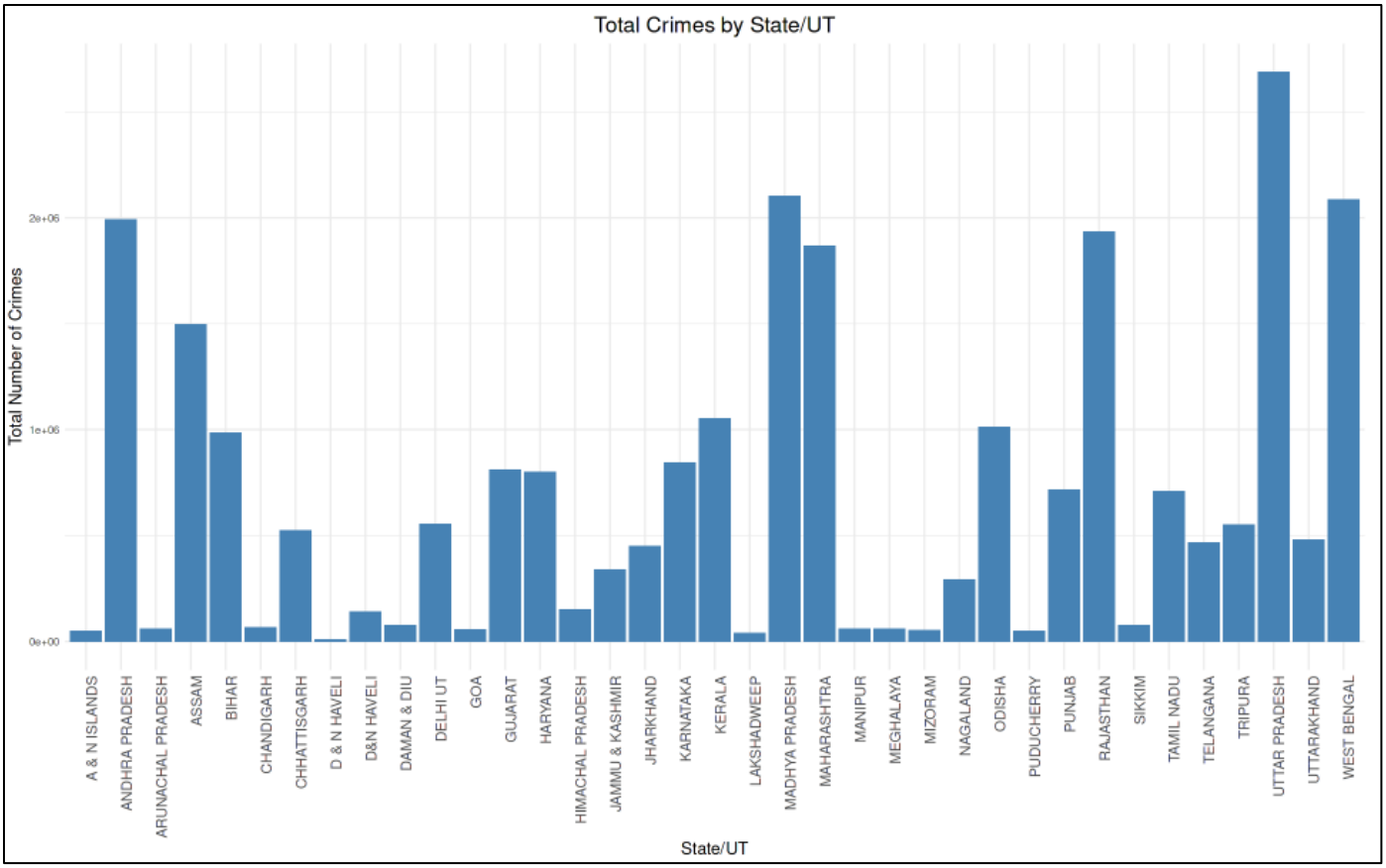
```
crime_data <- read.csv("/kaggle/input/crimes-against-women-in-india-2001-2021/CrimesOnWomenData.csv")
```

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

```
crime_long <- pivot_longer(crime_data,
                           cols = Rape:WT,
                           names_to = "Crime_Type",
                           values_to = "Count")
```

1] Bar Graph

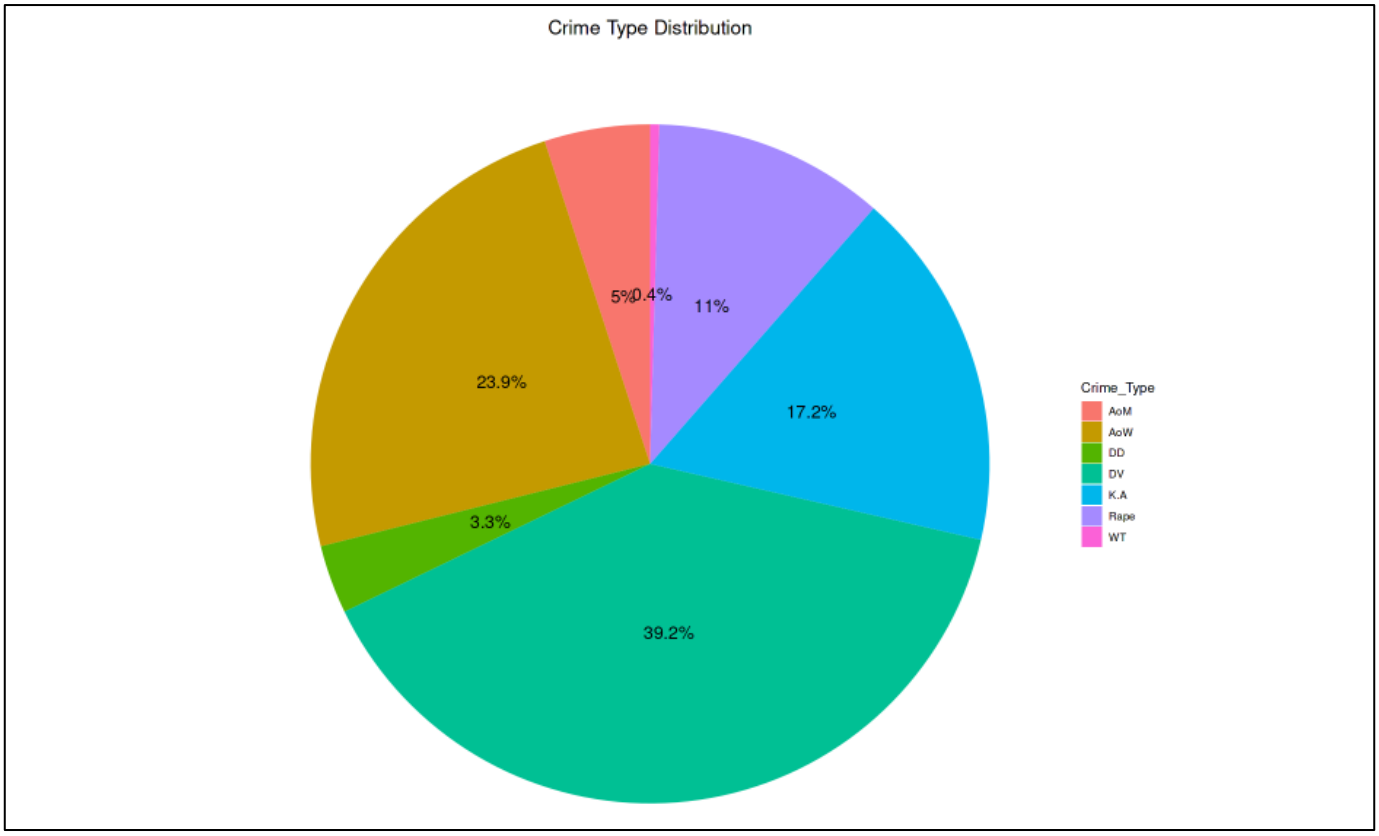
```
crime_data$State <- toupper(trimws(crime_data$State))
crime_summary <- crime_data %>% group_by(State) %>%
  summarise(Crime_Severity = sum(Crime_Severity, na.rm = TRUE))

ggplot(crime_summary, aes(x = State, 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 a very high crime count as compared to other states. Here Uttar Pradesh has a very high Crime rate as compared to other states/UT.

2] Pie Chart

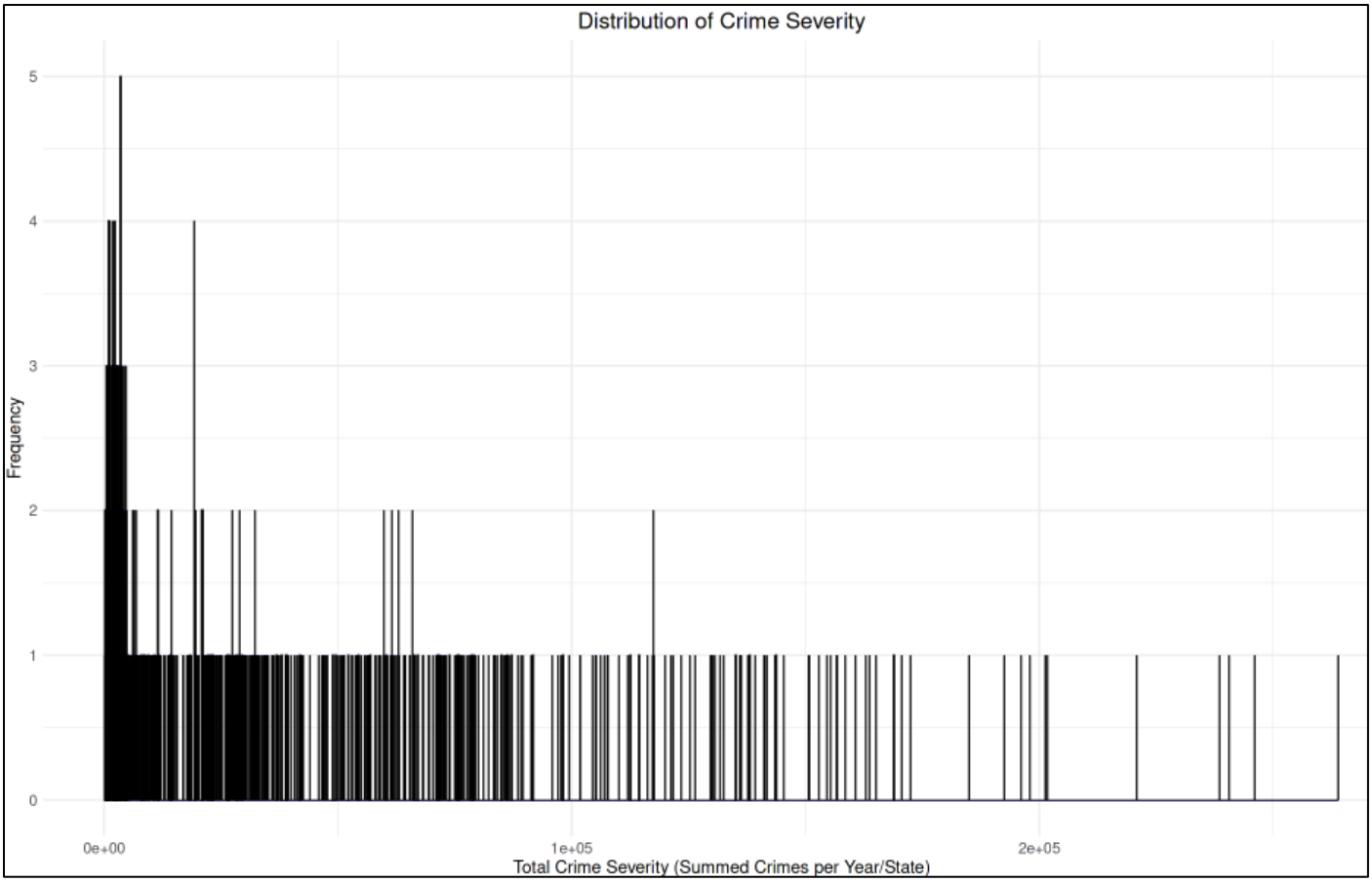
```
crime_long <- crime_data %>% pivot_longer(cols = -c(`State`, Year,`Crime_Severity`, X),
              names_to = "Crime_Type", values_to = "Count")

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 on women that occur in India Dowry Deaths covers about 39.2% and Assault on Women cover about 23.9% making these the highest crime that happen on women.

3] Histogram

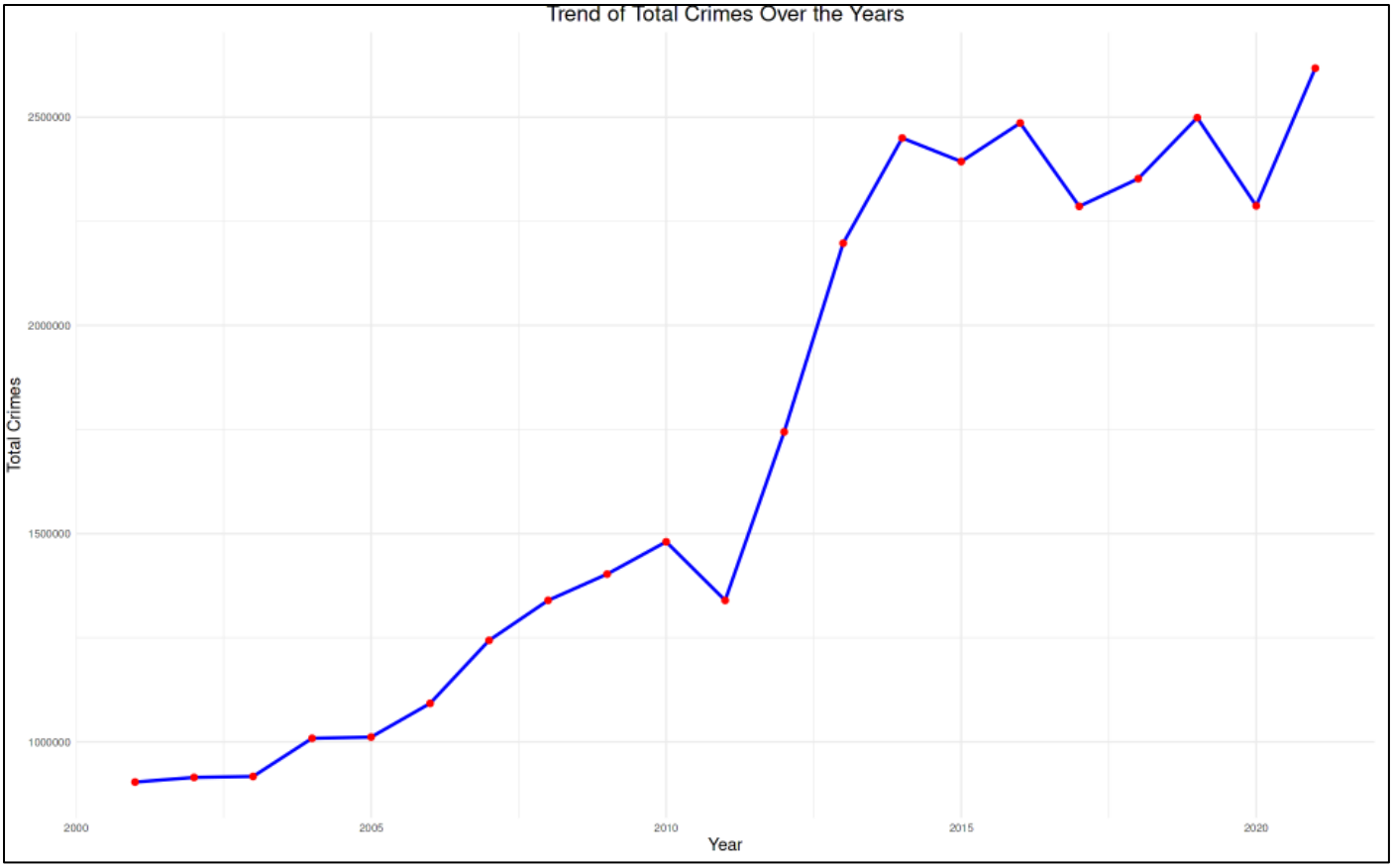
```
ggplot(crime_data, aes(x = Crime_Severity)) +
  geom_histogram(binwidth = 20, fill = "blue", color = "black") +
  ggtitle("Distribution of Crime Severity") +
  xlab("Total Crime Severity (Summed Crimes per Year/State)") +
  ylab("Frequency") + theme_minimal() + theme(
    axis.title.x = element_text(size = 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 means that regions are not safe for women.

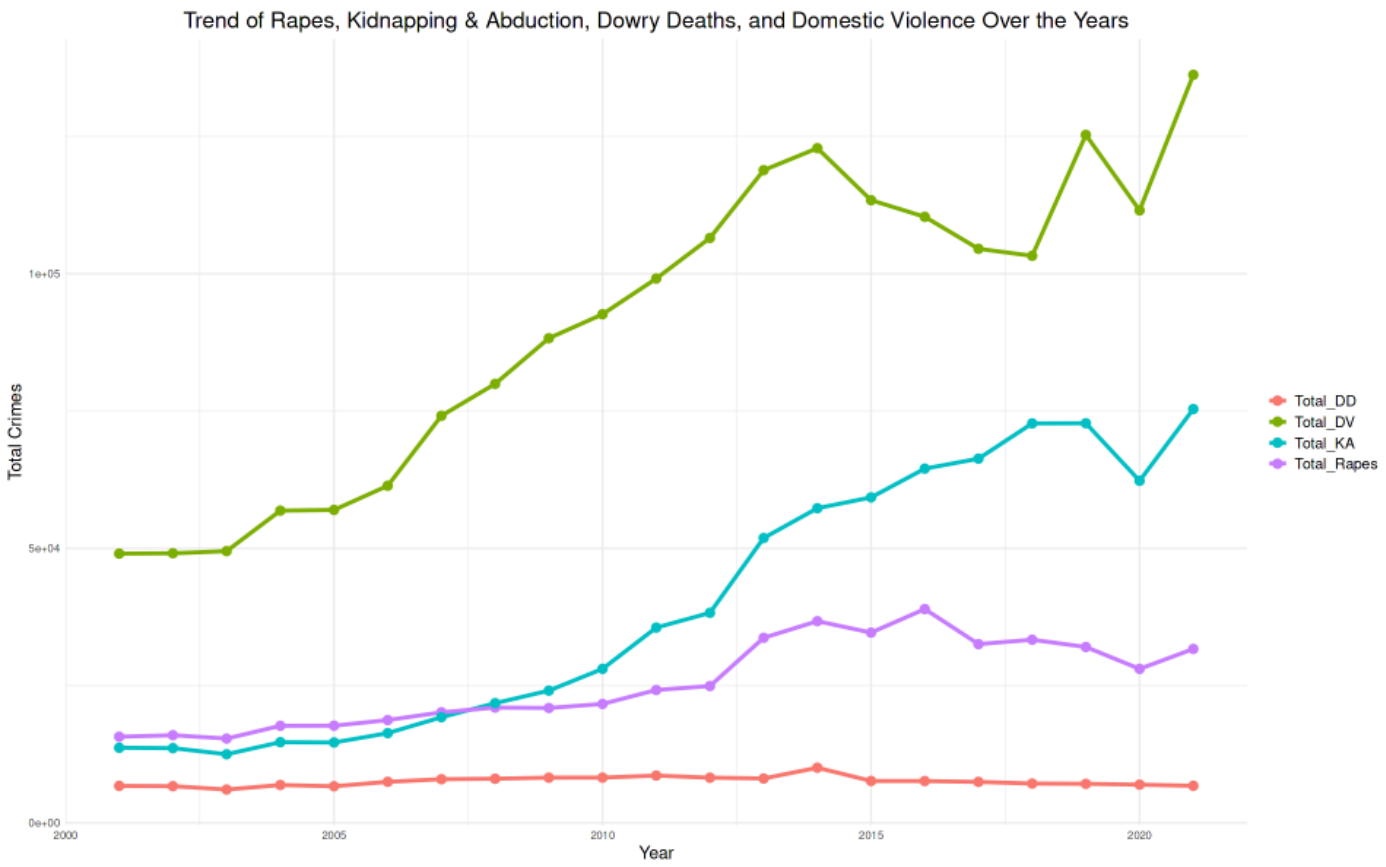
4] Timeline Chart

```
crime_data <- crime_data %>% rowwise() %>%
  mutate(Crime_Severity = sum(c_across(-c(`State`, Year)), 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 on women has been increasing over the years. The crime rates decreased in the year 2011, 2016 and 2020 there might be several reasons to it.

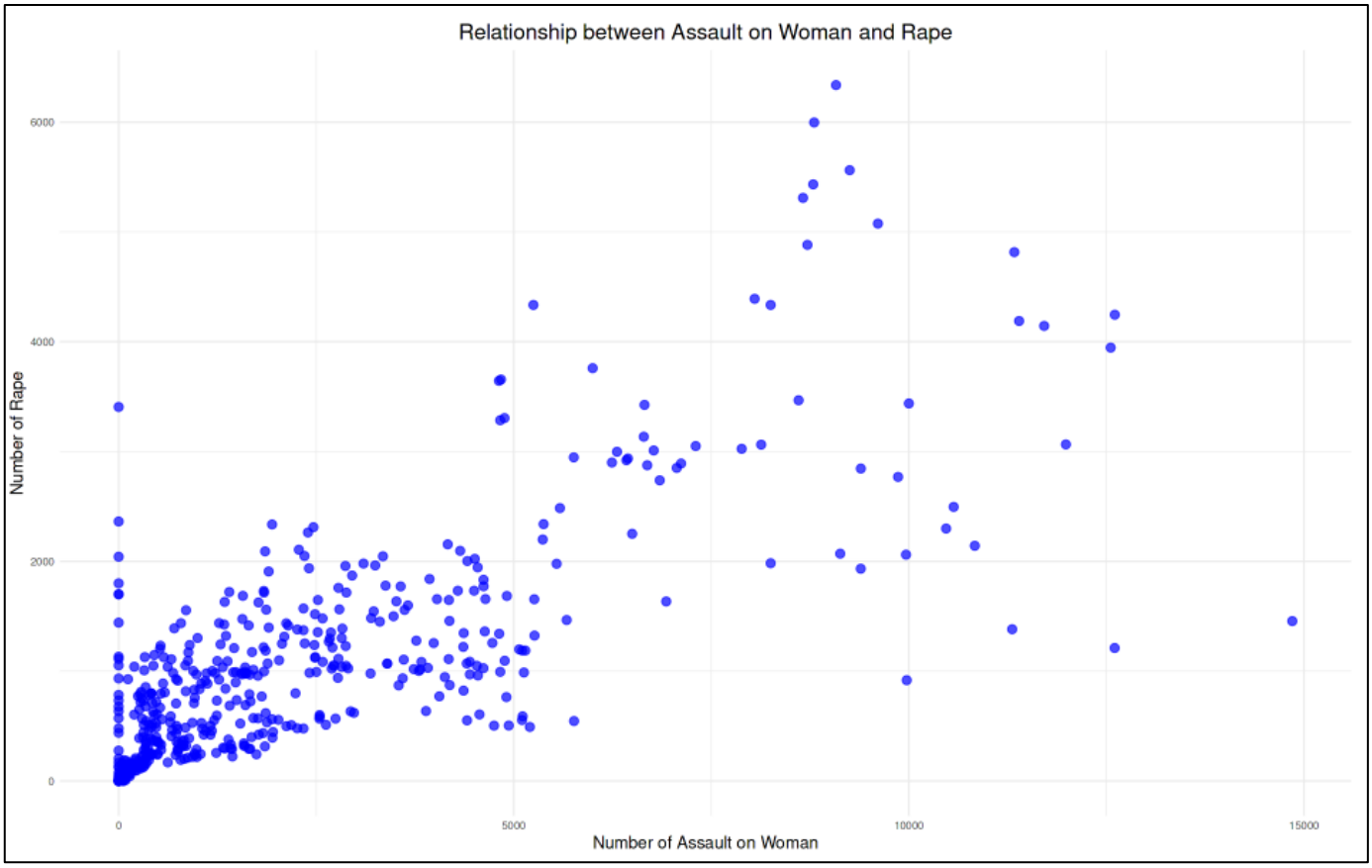
```
crime_comparison <- crime_data %>%
  select(`State`, Year, Rape, K.A, DD, DV) %>% group_by(Year) %>%
  summarise( Total_Rapes = sum(Rape, na.rm = TRUE),
  Total_KA = sum(K.A, na.rm = TRUE),Total_DD = sum(DD, na.rm = TRUE),
    Total_DV = sum(DV, na.rm = TRUE))

crime_comparison_long <- crime_comparison %>%
  pivot_longer(cols = c(Total_Rapes, Total_KA, Total_DD, Total_DV),
               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 Rapes, Kidnapping & Abduction, Dowry Deaths, and Domestic Violence 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 there has been a rapid growth in the Domestic Violence that takes place on women, not only that but the Kidnapping of girls and women also increased in the 2008 which explains the high rates or rapes and Women Trafficking in other graphs.

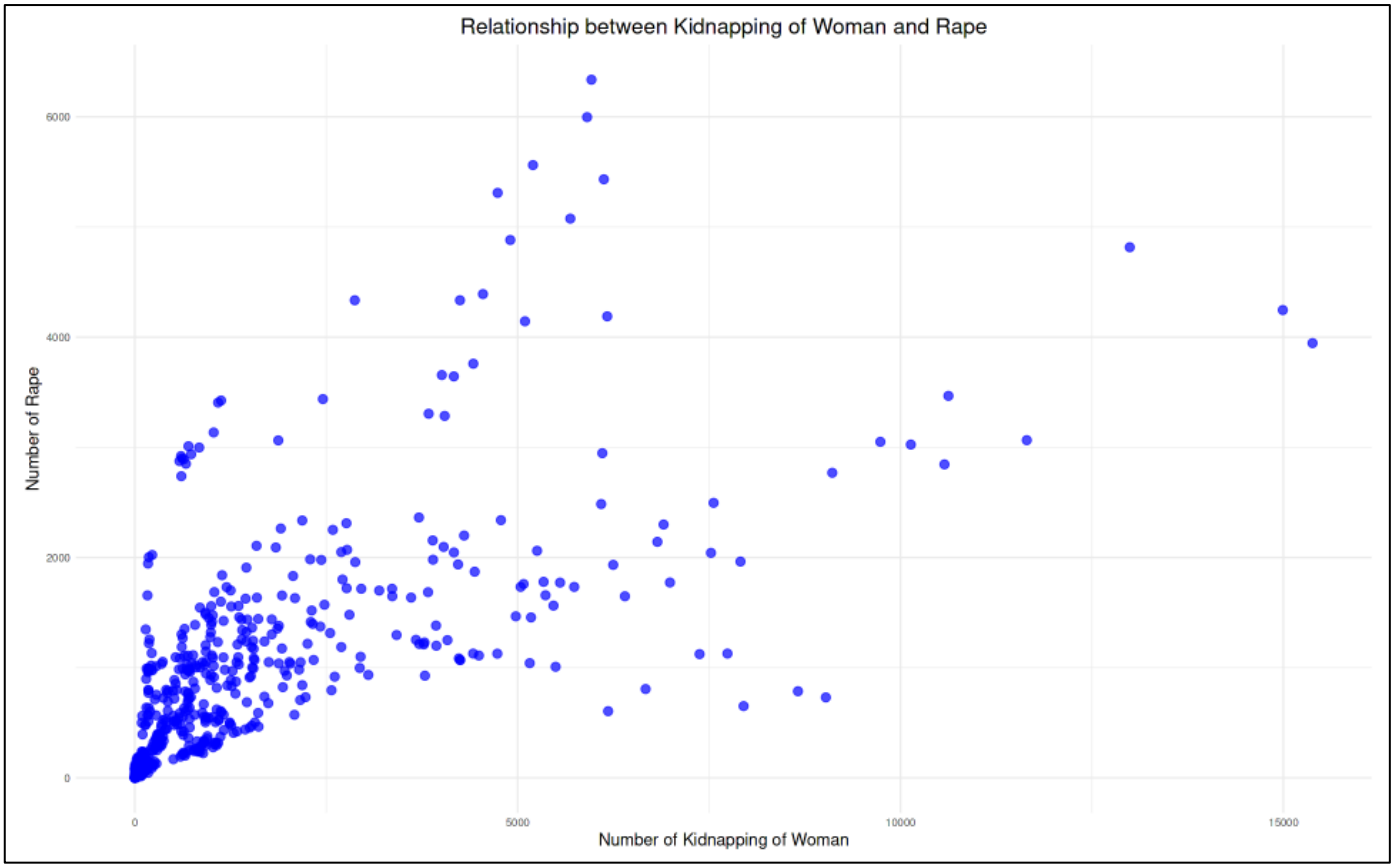
5] Scatterplot

```
ggplot(crime_data, aes(x = AoW, y = Rape)) +
  geom_point(color = "blue", size = 3, alpha = 0.7) +
  ggtitle("Relationship between Assault on Woman and Rape") +
  xlab("Number of Assault on Woman") + ylab("Number of Rape") +
  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 positive correlation between the Assault on Women and Rapes which would indicate that most Assaults may eventually lead to Rape.

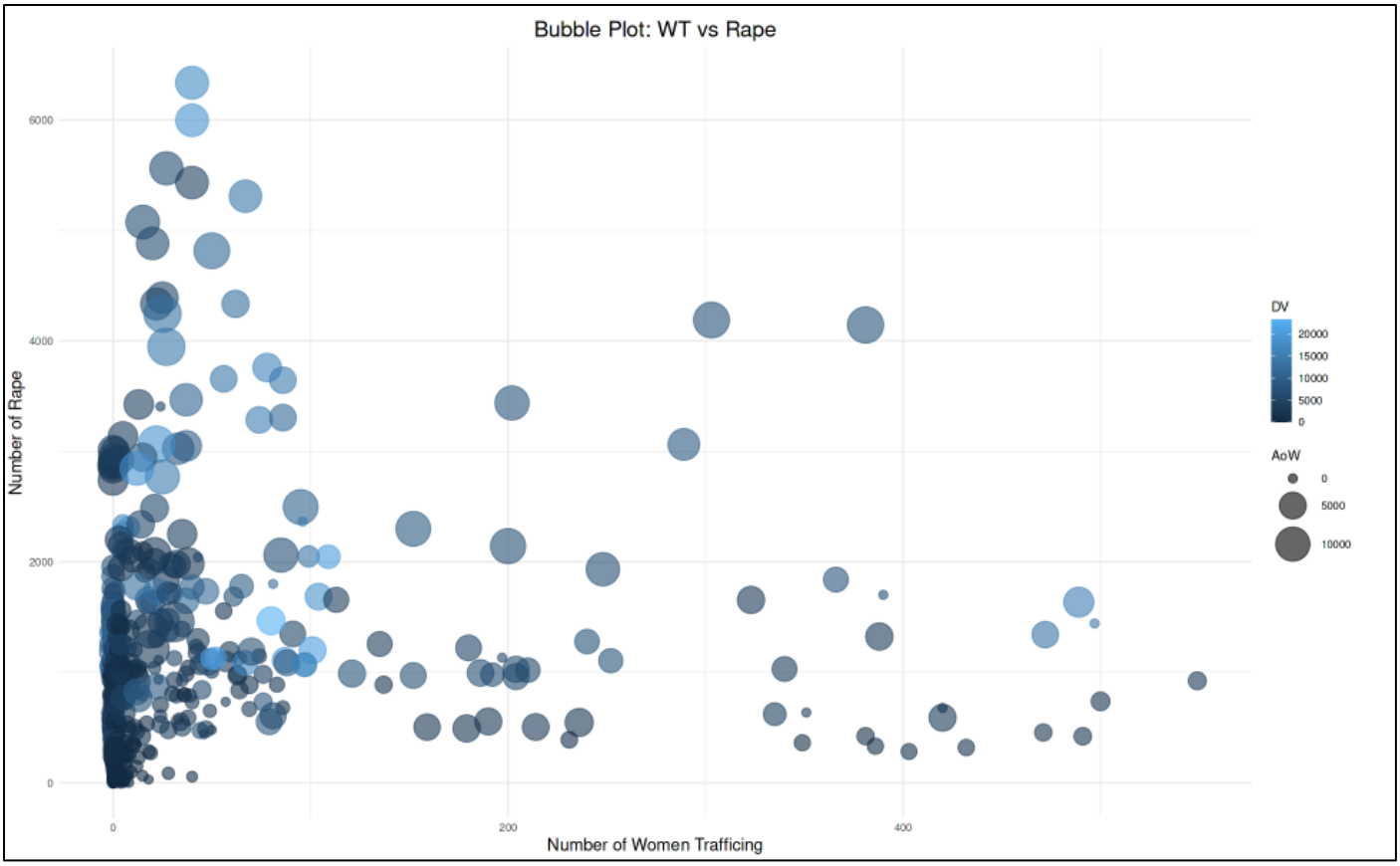
```
ggplot(crime_data, aes(x = K.A, y = Rape)) +
  geom_point(color = "blue", size = 3, alpha = 0.7) +
  ggtitle("Relationship between Kidnapping of Woman and Rape") +
  xlab("Number of Kidnapping of Woman") +
  ylab("Number of Rape") +
  theme_minimal() +
  theme(
    axis.title.x = element_text(size = 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 there is a positive correlation between Kidnapping of girls and women and Rapes which would indicate that most Kidnappings may eventually lead to Women Trafficking and Rapes.

6] Bubble plot

```
ggplot(crime_data, aes(x = WT, y = Rape,size = AoW, color = DV)) +
  geom_point(alpha = 0.6) + scale_size(range = c(3, 15)) +
  ggtitle("Bubble Plot: WT vs Rape") +
  xlab("Number of Women Traffickg") +
  ylab("Number of Rape") + 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 bubble plot shows a positive correlation between the number of women trafficked and the number of rapes. Larger bubbles indicate higher levels of both women trafficking and rape are related to assault on Women, suggesting a potential link between them.