Name	Rucha Kulkarni
Class	BE Computer Engineering (Batch F)
UID	2021300067
Exp No.	4

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

- Basic - Bar chart, Pie chart, Histogram, Time line chart, Scatter plot, Bubble plot
- Write observations from each chart

Description:

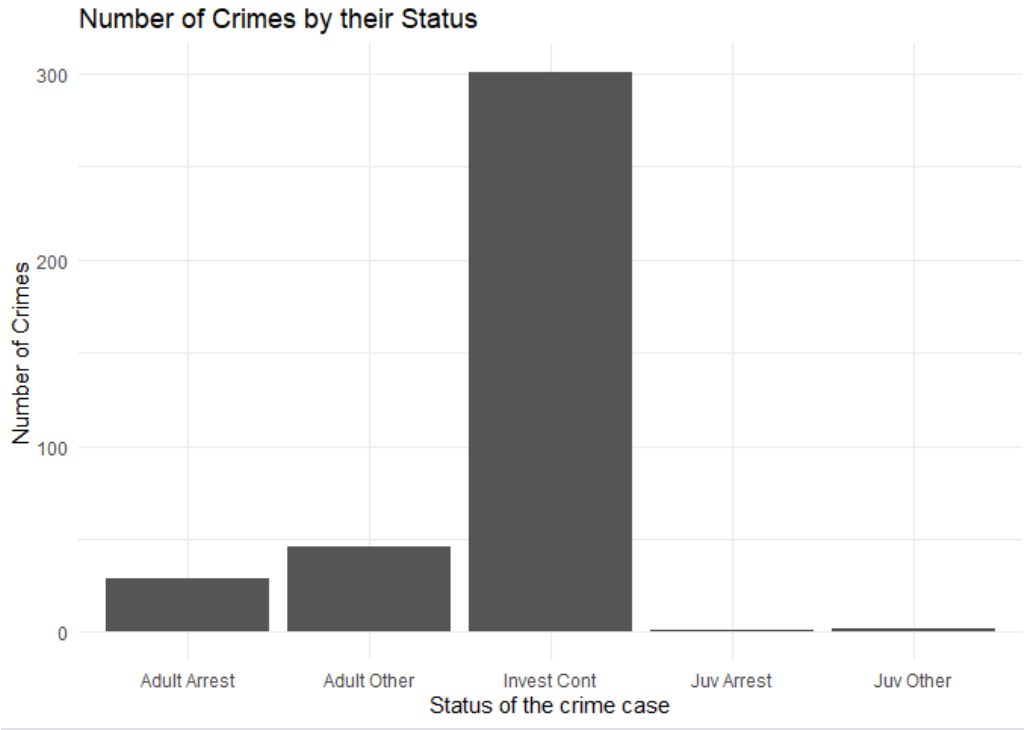
Dataset used is Crime Dataset available at

https://www.kaggle.com/datasets/qnqfbqfqo/crime-data-from-2020-to-present

It contains information about different crime incidents, including details such as the report number, dates and times of occurrence, location, type of crime, involved parties' characteristics, and other relevant information.

Graphs and Observations:

Bar Chart:

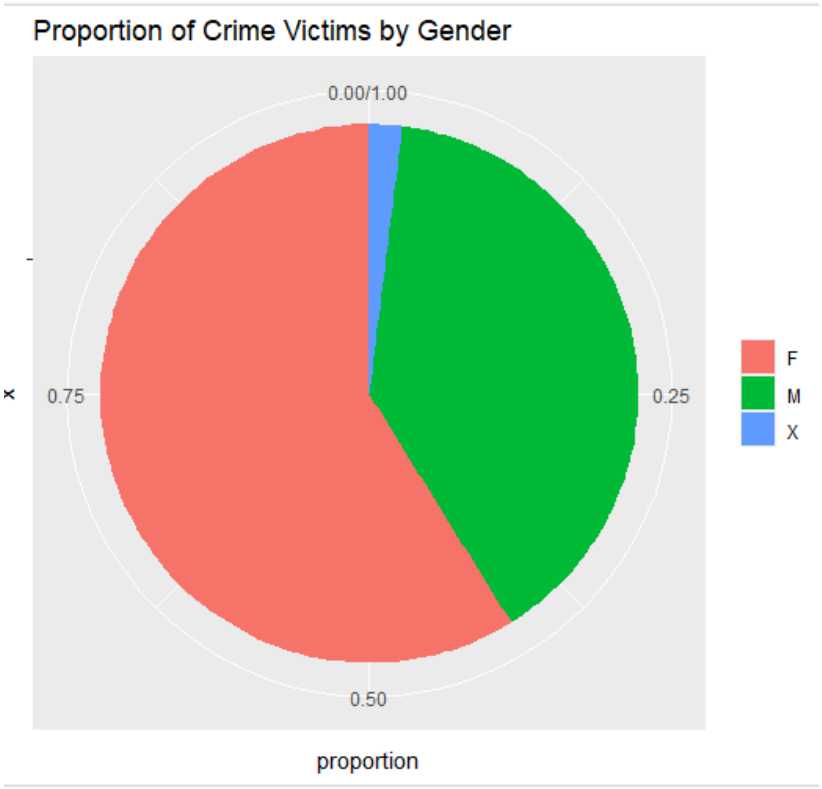


```
> ggplot(crime_data, aes(x = Status.Desc)) +
+     geom_bar() +
+     ggtitle("Number of Crimes by their Status") +
+     xlab("Status of the Case") + ylab("Number of Crime Cases")
>
```

Observation:

This bar graph shows the distribution of crimes based on the status of the crime case. Most of the crime cases are under Investigation Control status. The fact that most of the crime cases are in this status suggests that a significant number of cases are still being actively investigated or managed, and have not yet reached resolution or closure.

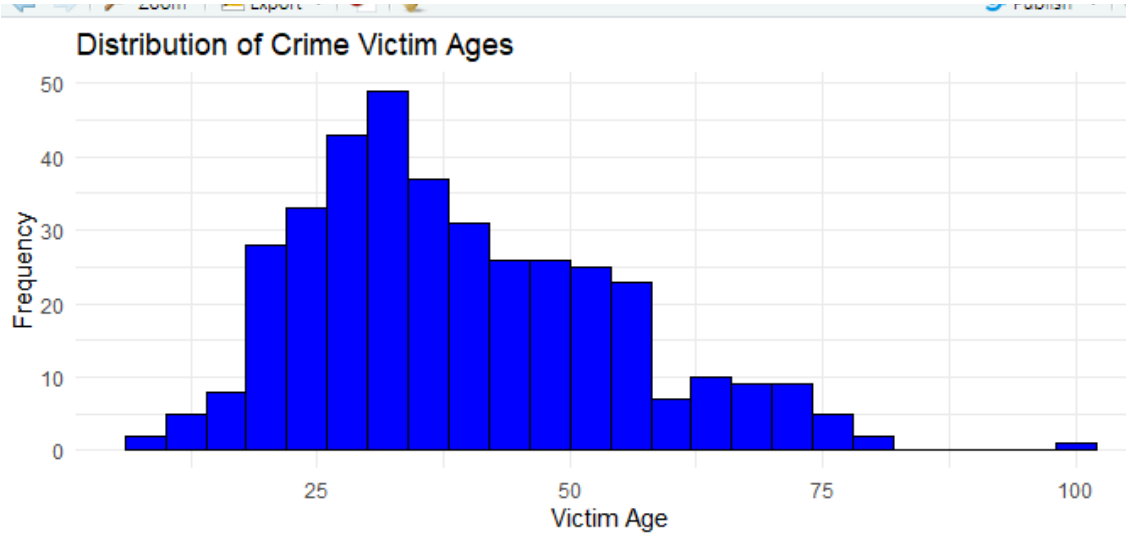
Pie Chart:




```
> crime_victims_gender <- crime_data %>%
+     group_by(Vict.Sex) %>%
+     summarise(count = n()) %>%
+     mutate(proportion = count / sum(count))
> ggplot(crime_victims_gender, aes(x = "", y = proportion, fill = Vict.Sex)) +
+     geom_bar(stat = "identity", width = 1) +
+     coord_polar(theta = "y") +
+     ggtitle("Proportion of Crime Victims by Gender") +
+     theme(legend.title = element_blank())
>
```

Observation: The pie chart displays the proportion of crimes by the victim's gender. It shows that a higher percentage of crimes are committed against female victims compared to male victims, indicating a notable gender disparity in victimization.

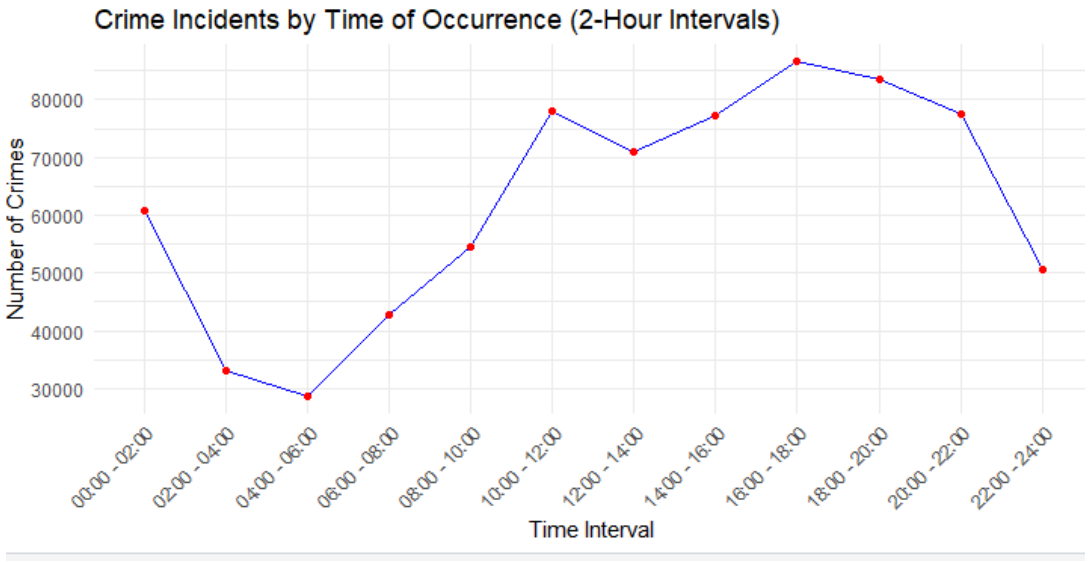
Histogram:




```
> histogram_victim_age <- ggplot(crime_data_filtered, aes(x = Vict.Age)) +
+     geom_histogram(binwidth = 2, fill = "blue", color = "black") +
+     labs(title = "Distribution of Crime Victim Ages", x = "Victim Age", y =
"Frequency") +
+     theme_minimal()
> print(histogram_victim_age)
> histogram_victim_age <- ggplot(crime_data_filtered, aes(x = Vict.Age)) +
+     geom_histogram(binwidth = 4, fill = "blue", color = "black") +
+     labs(title = "Distribution of Crime Victim Ages", x = "Victim Age", y =
"Frequency") +
+     theme_minimal()
> print(histogram_victim_age)
>
```

Observation: This histogram shows the age distribution of crime victims. There is a peak at a certain age range (around 20 to 40 years), indicating that victims within this age group are more common. The histogram tails off at both ends, suggesting fewer crimes involving very young or much older individuals.

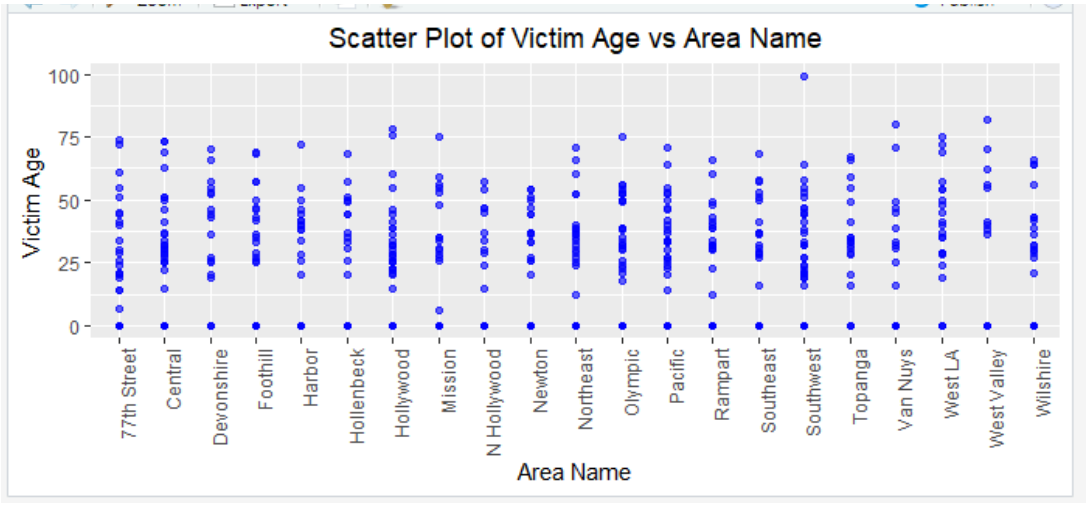
Timeline Chart:




```
> crime_data$TIME.OCC <- sprintf("%04d", crime_data$TIME.OCC)  # Ensure it's 4 digits
> crime_data$TIME.OCC <- as.numeric(crime_data$TIME.OCC)
>
> # Create a new column for 2-hour intervals
> crime_data$Time_Interval <- cut(crime_data$TIME.OCC,
+                                 breaks = seq(0, 2400, by = 200),  # Breaks for 2-hou
r intervals
+                                 labels = sprintf("%02d:00 - %02d:00", seq(0, 22, by
= 2), seq(2, 24, by = 2)),
+                                 include.lowest = TRUE)
>
> # Aggregate data by Time_Interval
> crime_count_time <- crime_data %>% group_by(Time_Interval) %>% summarize(Count = n
())
>
> # Convert Time_Interval to a factor with levels to maintain order
> crime_count_time$Time_Interval <- factor(crime_count_time$Time_Interval,
+                                          levels = sprintf("%02d:00 - %02d:00", seq
(0, 22, by = 2), seq(2, 24, by = 2)))
>
> # Create a time-based plot
> ggplot(crime_count_time, aes(x = Time_Interval, y = Count)) +
+     geom_line(group = 1, color = "blue") +  # Ensure lines are connected
+     geom_point(size = 1.5, color = "red") +
+     labs(title = "Crime Incidents by Time of Occurrence (2-Hour Intervals)", x = "Ti
me Interval", y = "Number of Crimes") +
+     theme_minimal() +
+     theme(axis.text.x = element_text(angle = 45, hjust = 1))  # Rotate x-axis labels
for readability
>
```

Observation: This line graph shows how crime occurrences are distributed over the duration of a day in two hour intervals. There are peaks at certain points, suggesting a pattern in crime timing.Most of the crimes occur during 2 pm to 8pm.

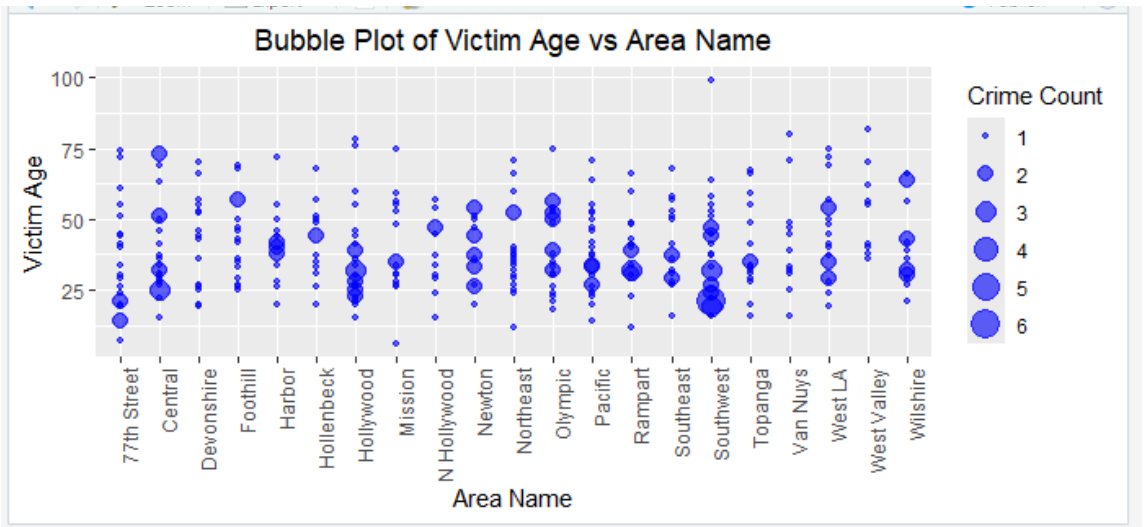
Scatter Plot:




```
> scatter_plot_area_age <- ggplot(crime_data_sample, aes(x = AREA.NAME, y = Vi
ct.Age)) +
+     geom_point(alpha = 0.6, color = "blue") +
+     labs(title = "Scatter Plot of Victim Age vs Area Name", x = "Area Name",
y = "Victim Age") +
+     theme(axis.text.x = element_text(angle = 90, hjust = 1),  # Rotate x-axi
s labels
+           plot.title = element_text(hjust = 0.5))
> print(scatter_plot_area_age)
>
```

Observation: This scatter plot represents victim age across different areas. There are clusters of data points, showing the range of ages per area. Some areas may have a wider age range of victims, while others show a narrower spread. In southwest, the crime victims range between most of the age group(25 to 75 years).

Bubble Plot:



```
> crime_count <- crime_data_filtered %>%
+     group_by(AREA.NAME, Vict.Age) %>%
+     summarize(Count = n(), .groups = "drop")
> bubble_plot_area_age <- ggplot(crime_count, aes(x = AREA.NAME, y = Vict.Age,
size = Count)) +
+     geom_point(alpha = 0.6, color = "blue") +
+     labs(title = "Bubble Plot of Victim Age vs Area Name", x = "Area Name",
y = "Victim Age", size = "Crime Count") +
+     theme(axis.text.x = element_text(angle = 90, hjust = 1),  # Rotate x-axi
s labels
+           plot.title = element_text(hjust = 0.5))
> print(bubble_plot_area_age)
>
```

Observation: The bubble size represents the count of crimes per age group in different areas. Larger bubbles indicate a higher frequency of crimes. This visualization helps understand not only the spread of victim ages across areas but also where crime is most concentrated.In southwest most crime victims are around 25 years.

Conclusion:

Through this experiment, we got to know about R and learned how to use it to create and interpret basic graphs. By applying these techniques to the crime dataset, we were able to uncover trends and insights related to crime distributions and victim demographics.