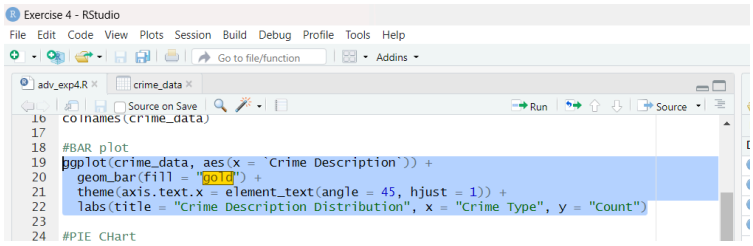
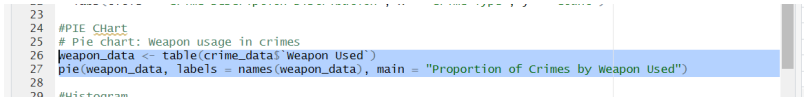
Sardar Patel Institute of Technology

SEM VII:ADVANCE DATA VISUALIZATION.

Name	Smruti Sonekar
UID no.	2021700064
Branch	BE CSE DS (BATCH B)
Experiment no.	5

Topic:	Data Visualization using R programming Language
Aim:	Create basic charts using R programming language on dataset Crime or Police / Law and Order • Basic - Bar chart, Pie chart, Histogram, Timeline chart, Scatter plot, Bubble plot • Write observations from each chart
Theory:	Crime Dataset : https://www.kaggle.com/datasets/sudhanvahg/indian-crimes-dataset This dataset, which spans the years 2020 to 2024, provides a thorough overview of criminal activity in several Indian cities. It contains comprehensive details on the kind of crime, the time and date it happened, the weapon used, the characteristics of the victims, and the extent of police presence. This dataset, which includes crimes ranging from identity theft to homicide, provides insightful information to academics, decision-makers, and law enforcement organizations that want to better understand crime trends and enhance public safety. A clear picture of crime resolution rates in different cities is provided by the data, which also includes information on whether the case was closed. In the R programming language, data visualization is easy as it **provides a range of powerful packages and tools that streamline the process of creating insightful and customizable graphics**. With packages like `ggplot2`, you can create complex and aesthetically pleasing visualizations with relatively simple code. The `dplyr` package simplifies data manipulation, making it straightforward to prepare data for visualization. Additionally, `leaflet` and `ggmap` offer interactive and geographical mapping capabilities, enhancing your ability to present spatial data effectively. The integration of these tools ensures that you can quickly translate data into meaningful visual representations.

	`library(ggplot2)`: Loads the `ggplot2` package, making its functions available for use. `library(dplyr)`: Loads the `dplyr` package, enabling its data manipulation functions. **"Report Number"**: Unique identifier for each crime report. **"Date Reported"**: The date when the crime was reported. **"Date of Occurrence"**: The date when the crime actually occurred. **"Time of Occurrence"**: The time when the crime took place. **"City"**: The city where the crime occurred. **"Crime Code"**: A code that represents the type of crime. **"Crime Description"**: A textual description of the crime. **"Victim Age"**: Age of the victim involved in the crime. **"Victim Gender"**: Gender of the victim. **"Weapon Used"**: The type of weapon used in the crime, if applicable. **"Crime Domain"**: The broader category or domain of the crime. **"Police Deployed"**: Information about the police response or deployment. **"Case Closed"**: Status indicating whether the case has been closed. **"Date Case Closed"**: The date when the case was officially closed.
Program:	Basic Charts 1) Bar Chart: This bar chart displays the distribution of different crime types.Each bar represents a type of crime, with the height showing the count of occurrences for each crime type. The x-axis shows crime descriptions, and the y-axis shows the number of reports.  2) Pie Chart:This pie chart shows the proportion of crimes involving different weapons.Each slice represents a weapon type used in crimes, with the size of the slice corresponding to the percentage of total crimes involving that weapon.  3) Histogram:Bars represent age ranges (bins) of victims, with the height indicating the number of victims in each age range. The x-axis shows age bins, and the y-axis shows the count of victims within each bin.

```
#Histogram
ggplot(crime_data, aes(x = `Victim Age`)) +
  geom_histogram(binwidth = 5, fill = "green", color = "black") +
  labs(title = "Distribution of Victim Ages", x = "Victim Age", y = "Count") +
  theme_minimal()

#Timeline Chart
library(lubridate) #this library helps in date manipulation
crime_data$`Date Reported` <- dmy_hm(crime_data$`Date Reported`)  #here we have converted Date Reported (
```

```
> colnames(crime_data)
 [1] "Report Number"       "Date Reported"       "Date of Occurrence" "Time of Occurrence" "City"
 [6] "Crime Code"          "Crime Description"   "Victim Age"         "Victim Gender"      "Weapon Used"
[11] "Crime Domain"        "Police Deployed"     "Case Closed"        "Date Case Closed"
> ggplot(crime_data, aes(x = `Crime Description`)) +
+   geom_bar(fill = "gold") +
+   theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
+   labs(title = "Crime Description Distribution", x = "Crime Type", y = "Count")
> weapon_data <- table(crime_data$`Weapon Used`)
> pie(weapon_data, labels = names(weapon_data), main = "Proportion of Crimes by Weapon Used")
> #Histogram
> ggplot(crime_data, aes(x = `Victim Age`)) +
+   geom_histogram(binwidth = 5, fill = "green", color = "black") +
+   labs(title = "Distribution of Victim Ages", x = "Victim Age", y = "Count") +
+   theme_minimal()
>
```

4) Timeline Chart:This timeline chart shows the trend of reported crimes over time.A line graph displays the number of crimes reported over time, with the x-axis representing the date and the y-axis representing the count of crimes reported on that date.

```
#Timeline Chart
library(lubridate) #this library helps in date manipulation
crime_data$`Date Reported` <- dmy_hm(crime_data$`Date Reported`)  #here we have converted Date Reported

ggplot(crime_data, aes(x = `Date Reported`)) +
geom_line(stat = "count", color = "orange") +
  labs(title = "Crimes Over Time", x = "Date Reported", y = "Number of Crimes") +
  theme_minimal()

#Scatter Plot
crime_data$`Victim Gender` <- factor(crime_data$`Victim Gender`)
crime_summary <- crime_data %>%
```

```
> library(lubridate) #this library helps in date manipulation

Attaching package: 'lubridate'

The following objects are masked from 'package:base':

    date, intersect, setdiff, union

> crime_data$`Date Reported` <- dmy_hm(crime_data$`Date Reported`)  #here we have converted Date Reported Colu
mn in standard Format
>
> ggplot(crime_data, aes(x = `Date Reported`)) +
+ geom_line(stat = "count", color = "orange") +
+   labs(title = "Crimes Over Time", x = "Date Reported", y = "Number of Crimes") +
+   theme_minimal()
>
```

5) Scatter Plot:This scatter plot visualizes the number of victims for each crime type, differentiated by gender.Points represent crime types with the x-axis showing crime descriptions and the y-axis showing the number of victims. Point color indicates gender, and point size reflects the number of victims.

```
#Scatter Plot
crime_data$`Victim Gender` <- factor(crime_data$`Victim Gender`)
crime_summary <- crime_data %>%
  group_by(`Crime Description`, `Victim Gender`) %>%
  summarise(Victim_Count = n(), .groups = 'drop')

# Scatter plot: Number of victims for each crime type by gender
ggplot(crime_summary, aes(x = `Crime Description`, y = Victim_Count, color = `Victim Gender`, size = Vic
  geom_point(alpha = 0.7) +
  labs(title = "Number of Victims by Crime Type and Gender", x = "Crime Description", y = "Number of Vic
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_size_continuous(range = c(2, 10))
```

```
+   theme_minimal()
> crime_data$`Victim Gender` <- factor(crime_data$`Victim Gender`)
> crime_summary <- crime_data %>%
+   group_by(`Crime Description`, `Victim Gender`) %>%
+   summarise(Victim_Count = n(), .groups = 'drop')
>
> # Scatter plot: Number of victims for each crime type by gender
> ggplot(crime_summary, aes(x = `Crime Description`, y = Victim_Count, color = `Victim Gender`, size = Victim_
Count)) +
+   geom_point(alpha = 0.7) +
+   labs(title = "Number of Victims by Crime Type and Gender", x = "Crime Description", y = "Number of Victim
s") +
+   theme_minimal() +
+   theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
+   scale_size_continuous(range = c(2, 10))
>
```

6) Bubble Plot:This bubble chart displays the relationship between cities and crime domains, with bubble sizes indicating the count of occurrences. The x-axis shows cities, the y-axis shows crime domains, and the size of each bubble represents the number of crimes reported in that city for that crime domain. The color of the bubbles also represents the count, providing additional insight into the distribution of crime domains across cities.

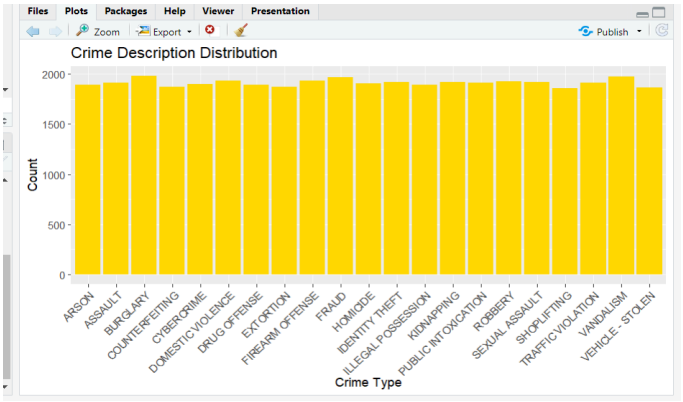
```r
#BUBBLE CHART
bubble_data <- crime_data %>%
  group_by(City, `Crime Domain`) %>%
  summarise(Count = n(), .groups = 'drop')

# Create the bubble chart
ggplot(bubble_data, aes(x = City, y = `Crime Domain`, size = Count, color = Count)) +
  geom_point(alpha = 0.7) +
  labs(title = "Bubble Chart of City and Crime Domain", x = "City", y = "Crime Domain") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +  # Rotate x-axis labels for reada
  scale_size_continuous(range = c(2, 20))  # Adjust range for bubble sizes
```

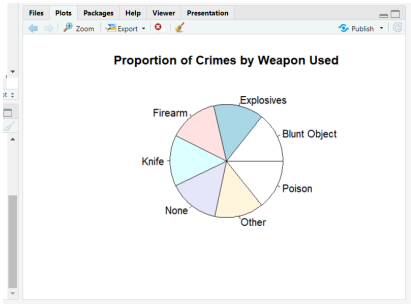
```
+   theme_minimal() +
+   theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
+   scale_size_continuous(range = c(2, 10))
> bubble_data <- crime_data %>%
+   group_by(City, `Crime Domain`) %>%
+   summarise(Count = n(), .groups = 'drop')
>
> # Create the bubble chart
> ggplot(bubble_data, aes(x = City, y = `Crime Domain`, size = Count, color = Count)) +
+   geom_point(alpha = 0.7) +
+   labs(title = "Bubble Chart of City and Crime Domain", x = "City", y = "Crime Domain") +
+   theme_minimal() +
+   theme(axis.text.x = element_text(angle = 45, hjust = 1)) +  # Rotate x-axis labels for readabili
ty
+   scale_size_continuous(range = c(2, 20))
>
```

Result:

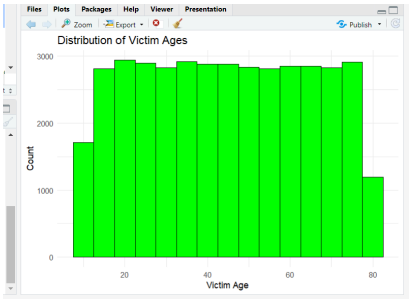
Bar Chart:



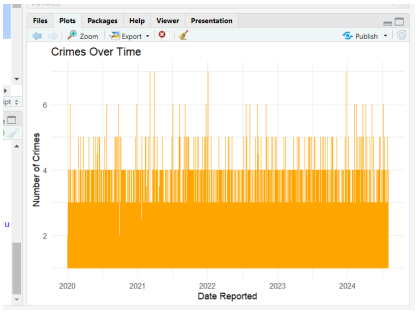
Pie Chart:



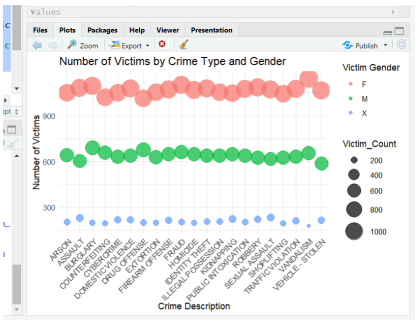
Histogram:



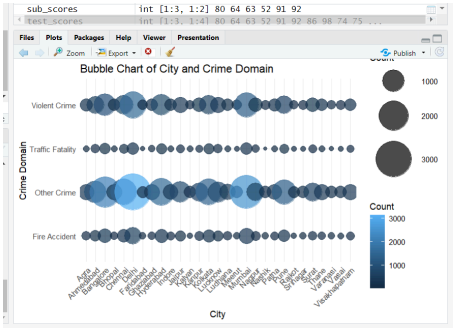
Timeline Chart:



Scatter Plot:



Bubble Plot:



Conclusion:	The crime dataset features various visualizations, including bar charts, pie charts, histograms, timeline charts, scatter plots, and bubble charts using R programming language. These provide a comprehensive view of crime patterns, victim demographics, and geographic distribution. They help identify prevalent crime types, weapon usage patterns, age groups, trends over time, gender impacts, and hotspots in different locations.