



Bharatiya Vidya Bhavan's

# Sardar Patel Institute of Technology

(Autonomous Institute Affiliated to University of Mumbai)

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India

## Experiment no 4

NAME: Sanimar Singh Manghera

UID: 2021700037

BATCH: CSE-DS

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

### Objectives:

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

### Theory:

Data visualization is an essential skill in data analysis that helps in understanding trends, patterns, and relationships within a dataset. R, a powerful statistical programming language, provides a wide range of tools for creating visually appealing and informative charts. In this experiment, we will use basic chart types to analyze crime data and derive insights.

### Chart Types:

1. **Bar Chart:** A bar chart is used to display categorical data with rectangular bars representing the frequency or count of each category.
2. **Pie Chart:** A pie chart shows the proportion of categories as slices of a pie, useful for comparing parts of a whole.
3. **Histogram:** A histogram is used to represent the distribution of numerical data by grouping it into bins.
4. **Timeline Chart:** A timeline chart visualizes data points in chronological order, often used to show trends over time.
5. **Scatter Plot:** A scatter plot displays the relationship between two numerical variables

using points in a Cartesian plane.

6. **Bubble Plot:** A bubble plot is an extension of a scatter plot where the size of the points (bubbles) represents an additional variable.

## Steps to Perform in R:

### 1. Set Up the Environment:

- Install and load necessary libraries.

R

Copy code

```
install.packages("ggplot2")
install.packages("dplyr")
library(ggplot2)
library(dplyr)
```

```
> library(ggplot2)
> library(dplyr)
```

Attaching package: 'dplyr'

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

filter, lag

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

intersect, setdiff, setequal, union

### 2. Load the Dataset:

- Load the crime dataset (replace `crime_data.csv` with your dataset's file name).

R

Copy code

```
crime_data <- read.csv("crime_data.csv")
```

```
> crime_data <- read.csv("crime_data.csv")
```

### 3. Data Preprocessing:

- Inspect and clean the data if necessary (handle missing values, filter relevant columns, etc.).

R

Copy code

```
crime_data <- crime_data %>% na.omit()
> colnames(crime_data)
[1] "Area_Name"          "Year"
[4] "Sub_Group_Name"     "Cases_Property_Recovered"
[7] "Value_of_Property_Recovered" "Value_of_Property_Stolen"
                                "Group_Name"
                                "Cases_Property_Stolen"
```

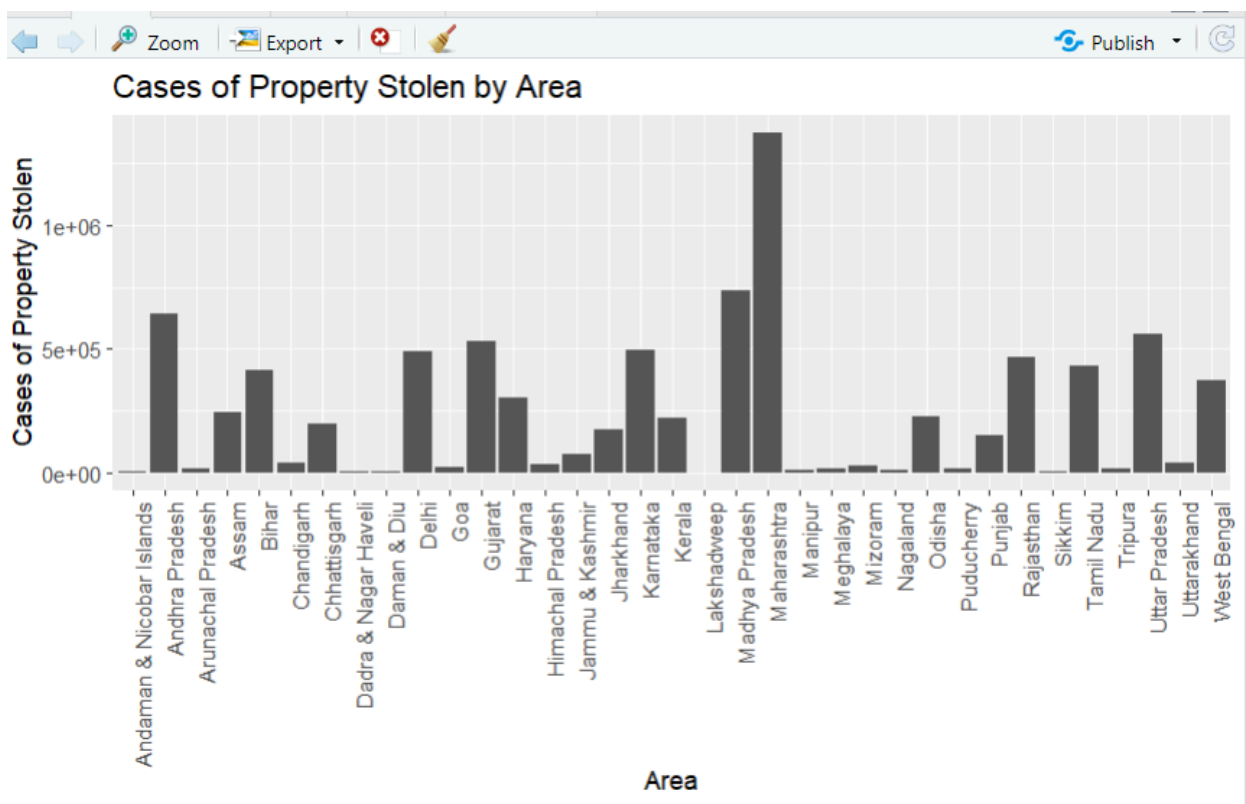
#### 4. Create Visualizations:

**Bar Chart:**

R

Copy code

```
ggplot(crime_data, aes(x = Area_Name, y = Cases_Property_Stolen)) +
  geom_bar(stat = "identity") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) + # Rotate the
x-axis labels if needed
  labs(title = "Cases of Property Stolen by Area",
        x = "Area",
        y = "Cases of Property Stolen")
```



- **Observation:** This shows the Number of cases of Property Stolen in each State of

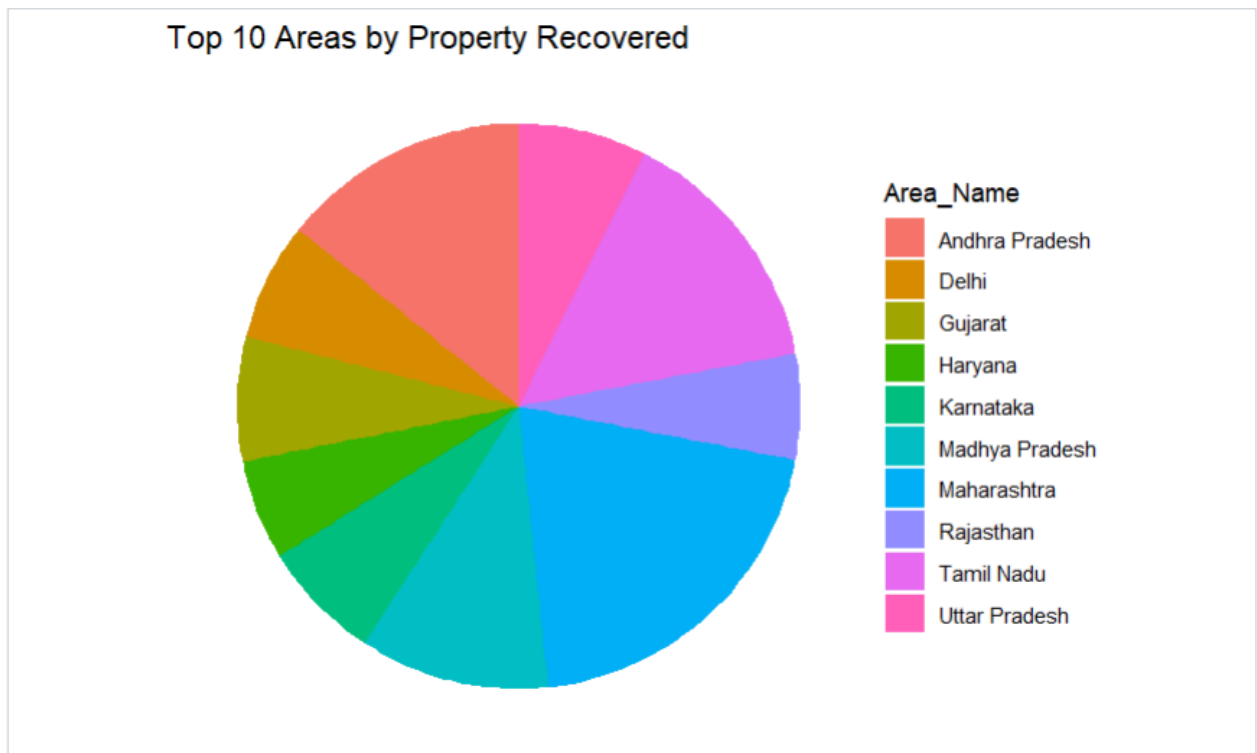
India.

### Pie Chart:

R

Copy code

```
ggplot(top_10_areas, aes(x = "", y = Total_Recovered, fill = Area_Name)) +  
+   geom_bar(stat = "identity", width = 1) +  
+   coord_polar(theta = "y") +  
+   labs(title = "Top 10 Areas by Property Recovered",  
+         x = NULL,  
+         y = NULL) +  
+   theme_void()
```



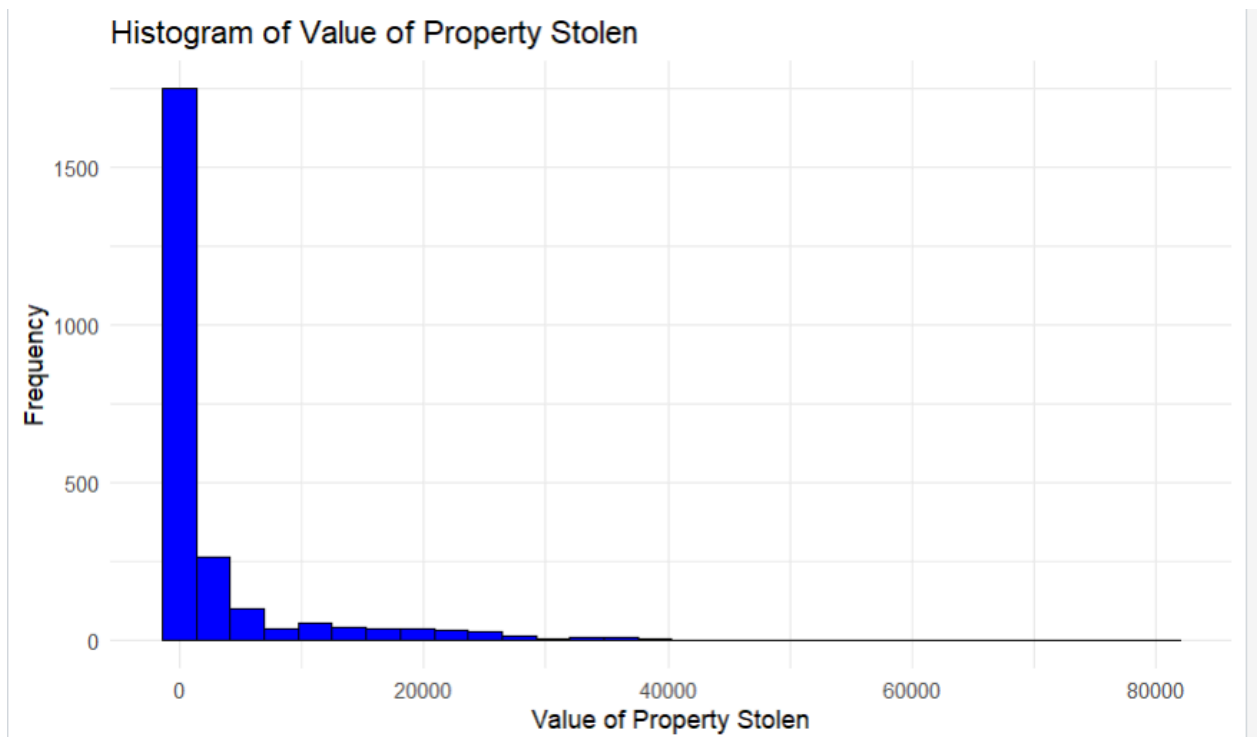
- **Observation:** The pie chart illustrates the top States for which the Cases of Stealing were Recovered.

### Histogram:

R

Copy code

```
ggplot(crime_data, aes(x = Cases_Property_Stolen)) +  
+   geom_histogram(fill = "blue", color = "black") +  
+   labs(title = "Histogram of Value of Property Stolen",  
+         x = "Value of Property Stolen",  
+         y = "Frequency") +  
+   theme_minimal()
```



- **Observation:** The histogram shows the distribution of value of Property stolen giving us the idea that too many expensive things were not stolen. But things of minimal value were stolen a lot/ reported.

### Timeline Chart:

R

Copy code

```
ggplot(crime_data, aes(x = Year, y = Value_of_Property_Stolen)) +  
+   geom_line(color = "blue", linewidth = 1) + # Use linewidth instead of  
size  
+   geom_point(color = "red") + # Optional: Add points to mark the values  
at each year  
+   labs(title = "Timeline of Property Stolen by Year",  
+         x = "Year",  
+         y = "Value of Property Stolen") +  
+   theme_minimal()
```



- **Observation:** The timeline chart highlights trends in crime frequency over time, revealing periods of high and low crime rates, lately it seems that the Crime frequency has gone to an all time high.

#### Scatter Plot:

R

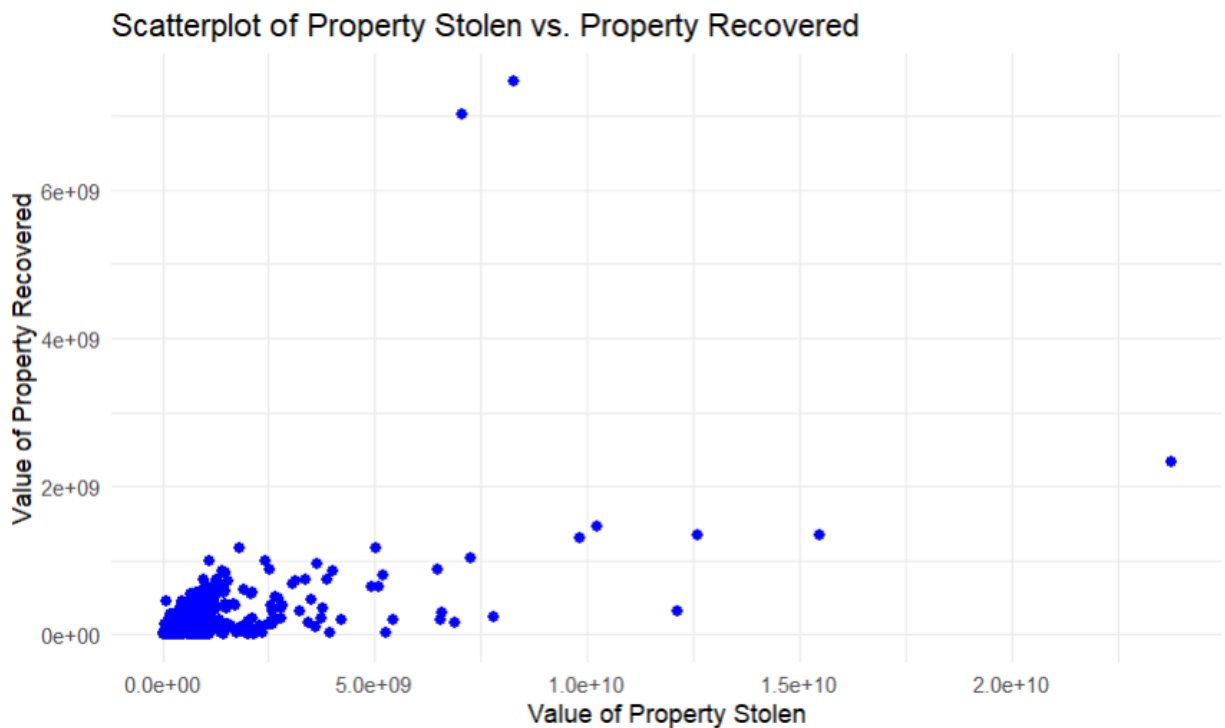
Copy code

```
ggplot(crime_data, aes(x = Value_of_Property_Stolen, y =
  Value_of_Property_Recovered)) +

+   geom_point(color = "blue", size = 2) + # Use color and size to style
  the points

+   labs(title = "Scatterplot of Property Stolen vs. Property Recovered",
+         x = "Value of Property Stolen",
+         y = "Value of Property Recovered") +

+   theme_minimal()
```



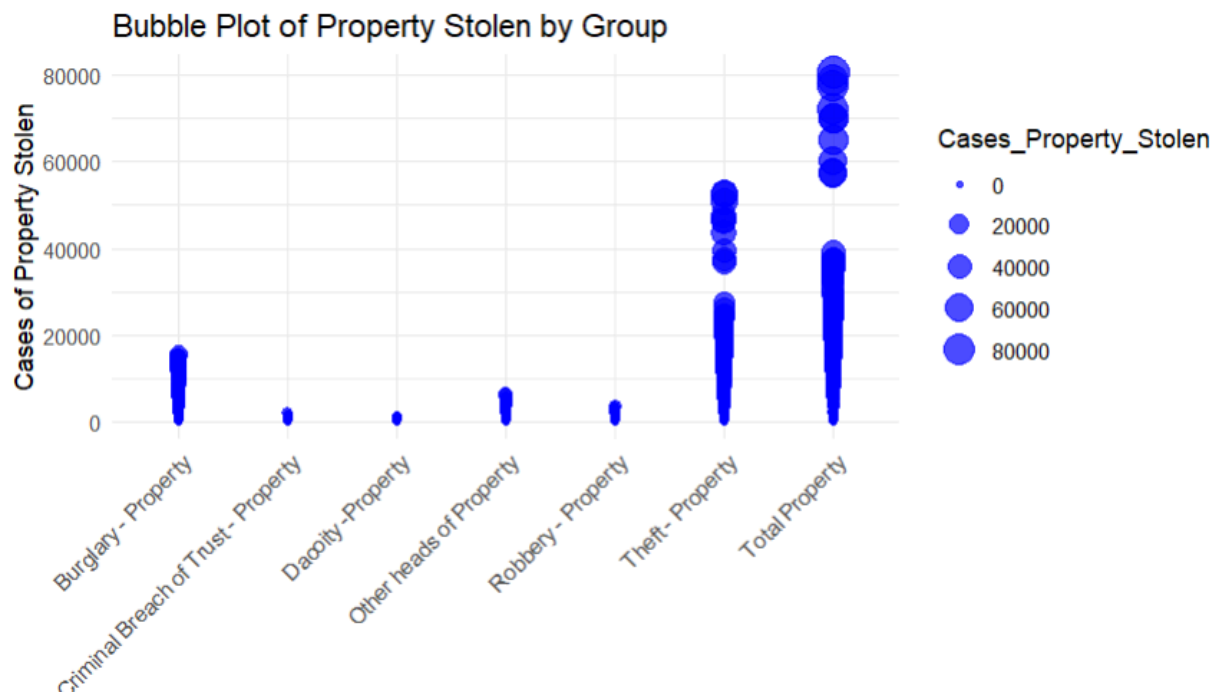
**Observation:** The scatter plot examines the relationship between Value\_of\_Property\_stolen and Value\_of\_Property\_Recovered

### Bubble Plot:

R

Copy code

```
ggplot(crime_data, aes(x = Group_Name, y =  
Cases_Property_Stolen, size = Cases_Property_Stolen)) +  
+   geom_point(color = "blue", alpha = 0.7) + # Set alpha  
for transparency  
+   labs(title = "Bubble Plot of Property Stolen by Group",  
+         x = "Group Name",  
+         y = "Cases of Property Stolen") +  
+   theme_minimal() +  
+   theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



- **Observation:** The bubble plot tells us the kind of theft that mostly happened in India.

#### Outcomes:

- Successfully created multiple types of charts using R to visualize crime data. ● Gained insights into the distribution, frequency, and relationships within the crime dataset.
- Developed an understanding of how different chart types can be used to analyze and present data effectively.

#### Conclusion:

This experiment demonstrated the power of data visualization in uncovering patterns and trends in a crime dataset. By using R, we efficiently created visual representations that allowed us to explore the data from different perspectives, leading to better-informed conclusions.