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: Anish Gade
UID: 2021700022
Batch: BE 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

Example

https://app.powerbigov.us/view?r=eyJrIjoiYmU4MDhiYWItYjEwOS00ZDg5LTk1OTUtNzNlMmU0MDFjNTk5IiwidCI6IjI5MzU3MDllLWMxMGMtNDgwOS1hMzAyLTg1MmQzNjlmODcwMCJ9

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)
```

2. **Load the Dataset:**

```
> data <- read.csv("D:/ADV_EXPS/ADV-LAB-EXPS/Exp4/10_Property_stolen_and_recovered.csv",
+                  sep = ",",            # specify separator
+                  header = TRUE,        # if first row contains column names
+                  stringsAsFactors = FALSE)  # prevent strings from converting to factors
```

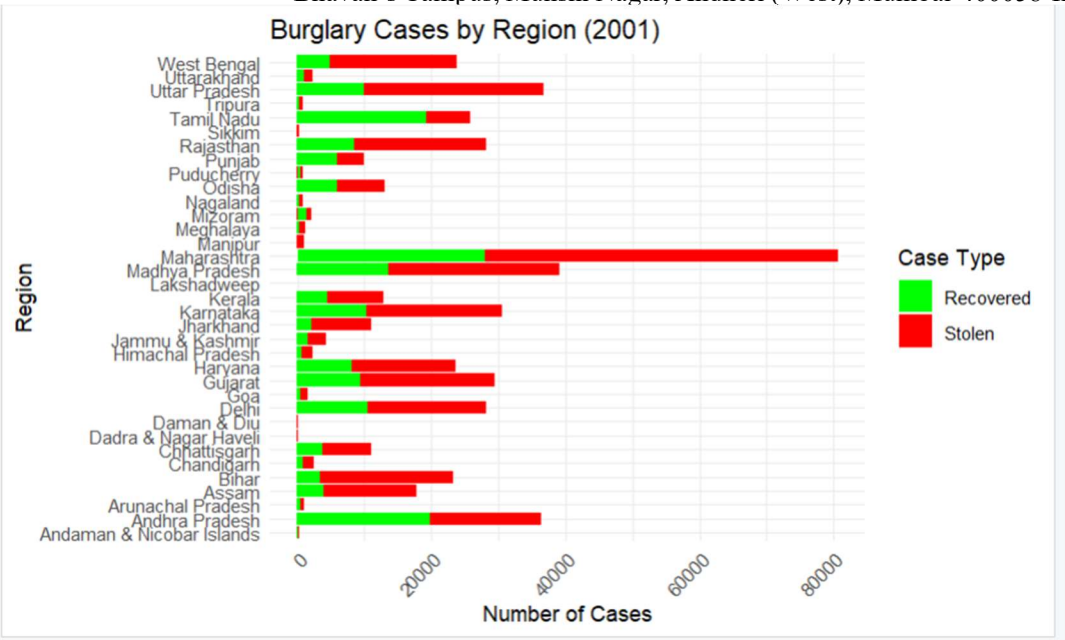
3. **Create Visualizations:**

Bar Chart:

R

```
> ggplot(data, aes(x = Area_Name)) +
+     geom_bar(aes(y = Cases_Property_Stolen, fill = "Stolen"), stat = "identity", position = "dodg
e") +
+     geom_bar(aes(y = Cases_Property_Recovered, fill = "Recovered"), stat = "identity", position =
"dodge") +
+     theme_minimal() +
+     theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
+     labs(title = "Burglary Cases by Region (2001)",
+          x = "Region",
+          y = "Number of Cases",
+          fill = "Case Type") +
+     scale_fill_manual(values = c("Stolen" = "red", "Recovered" = "green")) +
+     coord_flip()
```

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



Observation:

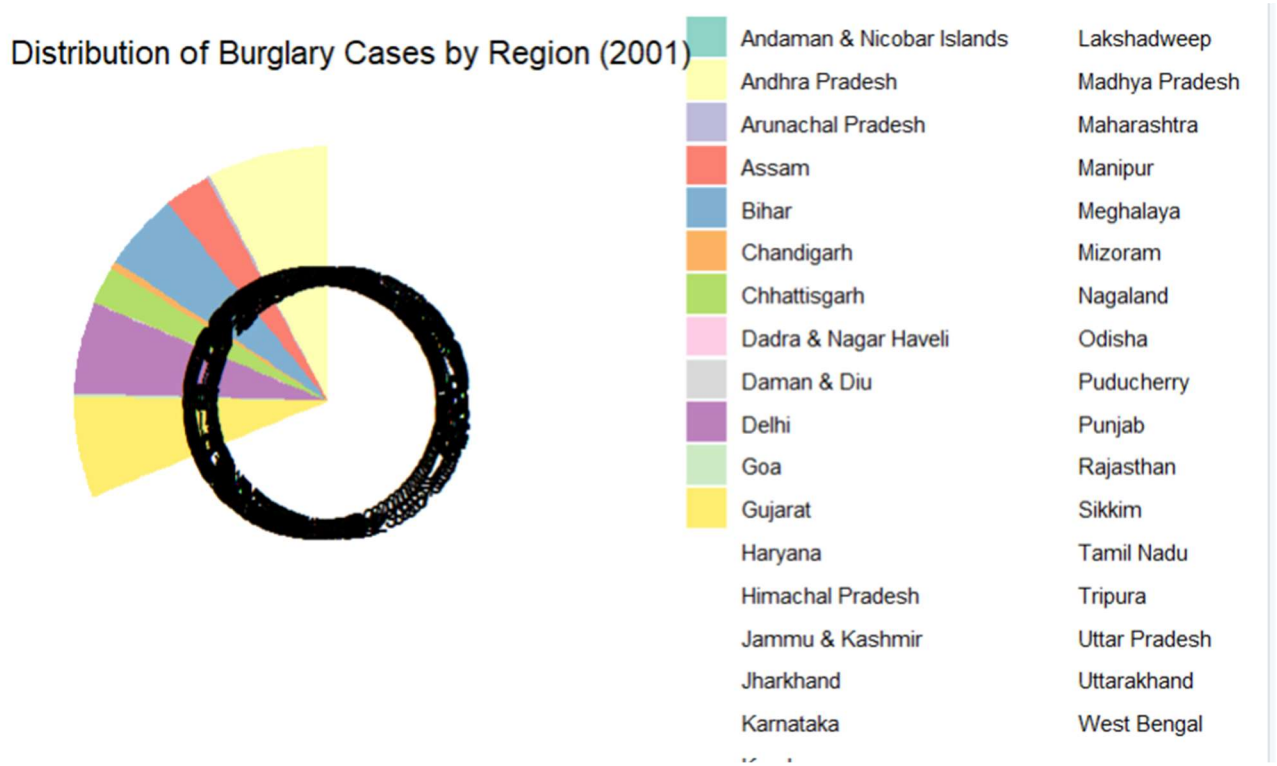
Looking at the 2001 burglary data across Indian regions, there are striking patterns in both theft incidents and recovery effectiveness. Maharashtra emerges as a significant hotspot with approximately 8000 stolen cases, far exceeding other regions, followed by West Bengal, Uttar Pradesh, and Tamil Nadu forming a second tier of high-burglary areas. A consistent pattern across all regions is the notable gap between stolen cases (red bars) and recovered cases (green bars), with recovery rates universally lower than theft rates, suggesting widespread challenges in property recovery efforts. The data reveals a potential correlation between urbanization/industrialization and burglary rates, as more developed states generally show higher incident numbers, while smaller Union Territories and less industrialized regions report fewer cases. This suggests that law enforcement agencies, particularly in larger states, may need to strengthen their recovery mechanisms and potentially learn from smaller regions that show better recovery-to-theft ratios, although no region has achieved a recovery rate that matches or exceeds its theft rate

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

Pie Chart:

R

```
> library(dplyr)
> data$percentage <- round((data$Cases_Property_Stolen / sum(data$Cases_Property_Stolen)) * 100, 1)
> data$label <- paste0(data$Area_Name, "\n(", data$percentage, "%)")
> ggplot(data, aes(x = "", y = Cases_Property_Stolen, fill = Area_Name)) +
+     geom_bar(stat = "identity", width = 1) +
+     coord_polar("y", start = 0) +
+     theme_minimal() +
+     theme(
+         axis.title.x = element_blank(),
+         axis.title.y = element_blank(),
+         panel.border = element_blank(),
+         panel.grid = element_blank(),
+         axis.ticks = element_blank(),
+         axis.text = element_blank(),
+         legend.title = element_blank(),
+         legend.position = "right"
+     ) +
+     labs(title = "Distribution of Burglary Cases by Region (2001)") +
+     scale_fill_brewer(palette = "Set3") +
+     geom_text(aes(label = percentage),
+               position = position_stack(vjust = 0.5))
```



Observation:

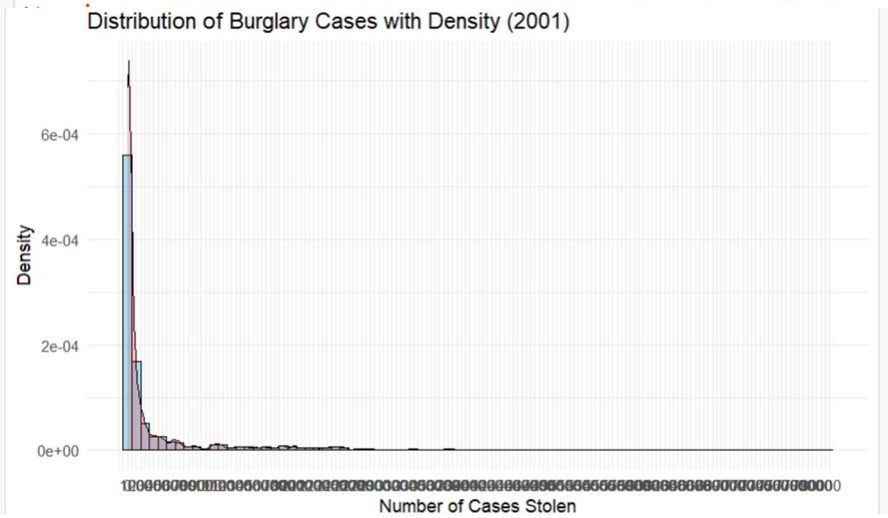
The pie chart visualization of 2001 burglary cases across Indian regions reveals significant disparities in crime distribution, with larger and more developed states like Andhra Pradesh commanding substantial portions of total cases, while Union Territories and smaller states show minimal representation. This uneven distribution suggests a correlation between region size, urbanization, and burglary incidents, though it's worth noting that larger slices might also reflect better crime reporting systems in more developed areas rather than just higher crime rates. The stark contrast between the largest and smallest slices emphasizes the need for

region-specific approaches to crime prevention and management.

Histogram:

R

```
> ggplot(data, aes(x = Cases_Property_Stolen)) +
+     geom_histogram(binwidth = 1000, fill = "skyblue", color = "black", alpha = 0.7) +
+     theme_minimal() +
+     labs(title = "Distribution of Burglary Cases (2001)",
+          x = "Number of Cases Stolen",
+          y = "Frequency (Number of Regions)") +
+     geom_vline(aes(xintercept = mean(Cases_Property_Stolen)),
+                color = "red", linetype = "dashed", size = 1) +
+     annotate("text", x = mean(data$Cases_Property_Stolen), y = 5,
+              label = "Mean", color = "red", vjust = -1) +
+     scale_x_continuous(breaks = seq(0, max(data$Cases_Property_Stolen), by = 1000))
```

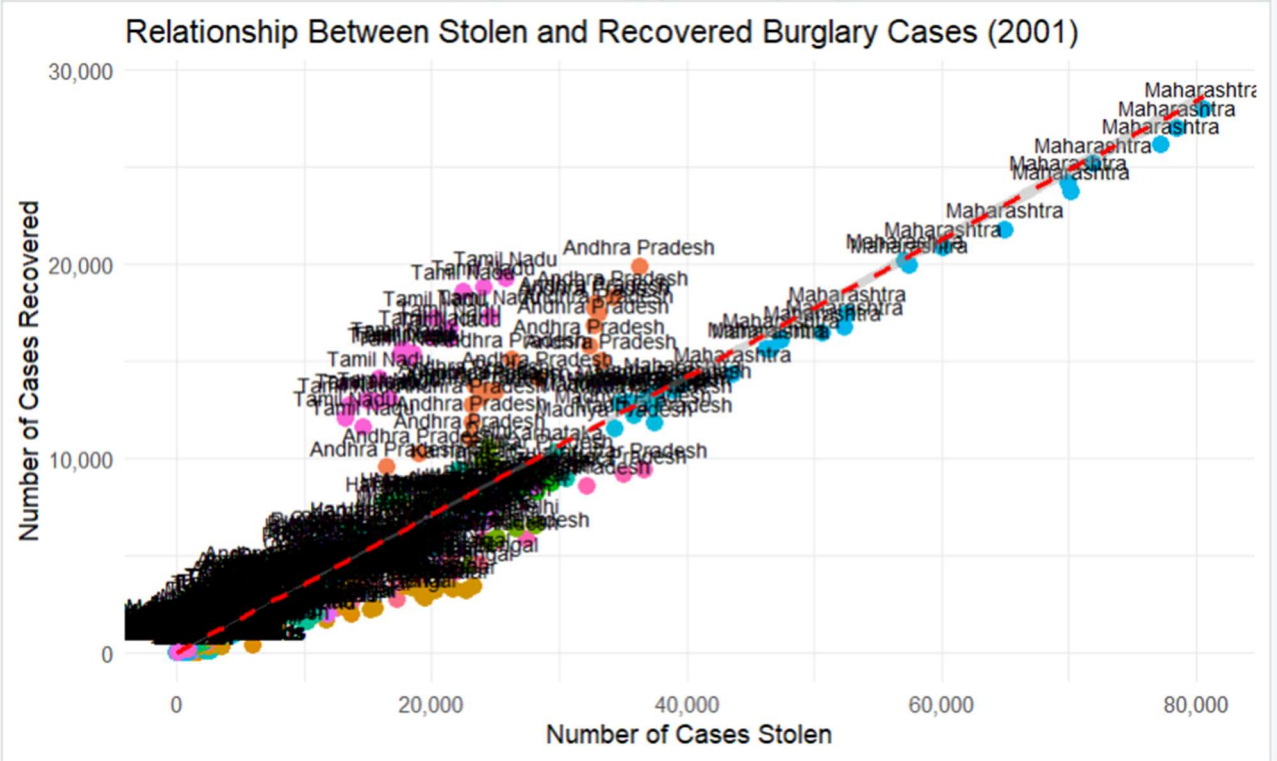


Observation: The histogram shows the distribution of crime severity, identifying common severity levels.

Scatter Plot:

R

```
> ggplot(data, aes(x = Cases_Property_Stolen, y = Cases_Property_Recovered)) +
+     geom_point(aes(color = Area_Name), size = 3) +
+     geom_text(aes(label = Area_Name), vjust = -0.8, size = 3) +
+     geom_smooth(method = "lm", color = "red", se = TRUE, linetype = "dashed") +
+     theme_minimal() +
+     labs(title = "Relationship Between Stolen and Recovered Burglary Cases (2001)",
+          x = "Number of Cases Stolen",
+          y = "Number of Cases Recovered",
+          color = "Region") +
+     theme(legend.position = "none") +
+     scale_x_continuous(labels = scales::comma) +
```



Observation: The scatter plot examines the relationship between crime severity

and police response time, identifying any correlation.

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.

Submission

https://docs.google.com/forms/d/e/1FAIpQLScrs2lYhqrJPz7M9lVvAqZn3M8cM5H1hc58ZpJRCO2jxcWuxg/viewform?usp=sf_link