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

```
Installing package into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)

Installing package into ‘/usr/local/lib/R/site-library’
(as ‘lib’ is unspecified)


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
```

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

```
colSums(is.na(crime_data))
```

```
                   Area_Name                         Year 
                           0                            0 
                  Group_Name               Sub_Group_Name 
                           0                            0 
   Cases_Property_Recovered       Cases_Property_Stolen 
                           0                            0 
Value_of_Property_Recovered    Value_of_Property_Stolen 
                           0                            0 
```

```
data<- crime_data
```

```
head(data)
```

```
  Area_Name                   Year Group_Name          Sub_Group_Name
1 Andaman & Nicobar Islands 2001 Burglary - Property 3. Burglary
2 Andhra Pradesh            2001 Burglary - Property 3. Burglary
3 Arunachal Pradesh         2001 Burglary - Property 3. Burglary
4 Assam                     2001 Burglary - Property 3. Burglary
5 Bihar                     2001 Burglary - Property 3. Burglary
6 Chandigarh                2001 Burglary - Property 3. Burglary
  Cases_Property_Recovered Cases_Property_Stolen
Value_of_Property_Recovered
```

```
1   27                           64                         755858

2 3321                         7134                       51483437

3   66                          248                         825115

4  539                         2423                        3722850

5  367                         3231                        2327135

6  119                          364                        1804823

  Value_of_Property_Stolen
1   1321961
2 147019348
3   4931904
4  21466955
5  17023937
6  10217378
```

```
head(data$Area_Name)
```

```
[1] "Andaman & Nicobar Islands" "Andhra Pradesh"
[3] "Arunachal Pradesh"         "Assam"
[5] "Bihar"                     "Chandigarh"
```

```
data$Area_Name <- as.factor(data$Area_Name)
data$Group_Name <- as.factor(data$Group_Name)
data$Sub_Group_Name <- as.factor(data$Sub_Group_Name)

head(data)
```

```
  Area_Name                 Year Group_Name          Sub_Group_Name
1 Andaman & Nicobar Islands 2001 Burglary - Property 3. Burglary
2 Andhra Pradesh            2001 Burglary - Property 3. Burglary
3 Arunachal Pradesh         2001 Burglary - Property 3. Burglary
4 Assam                     2001 Burglary - Property 3. Burglary
5 Bihar                     2001 Burglary - Property 3. Burglary
6 Chandigarh                2001 Burglary - Property 3. Burglary
  Cases_Property_Recovered Cases_Property_Stolen
Value_of_Property_Recovered
1   27                           64                         755858

2 3321                         7134                       51483437

3   66                          248                         825115

4  539                         2423                        3722850

5  367                         3231                        2327135
```

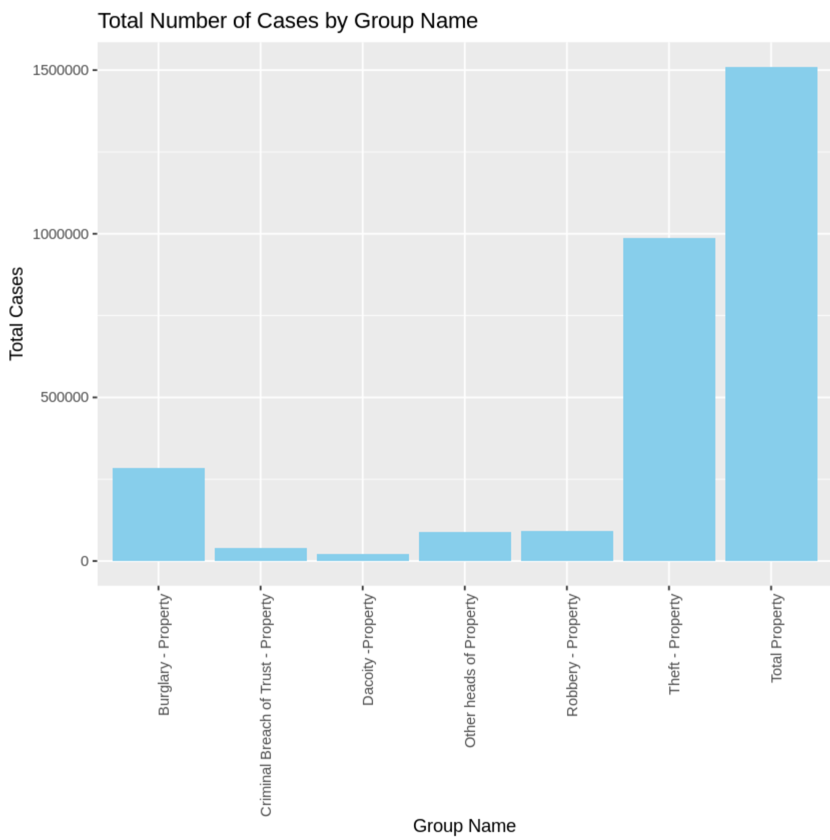
```
6  119                        364                     1804823

  Value_of_Property_Stolen
1    1321961
2 147019348
3    4931904
4   21466955
5   17023937
6   10217378

str(data$"Total Cases")

 NULL

# Bar chart
ggplot(data, aes(x = Group_Name)) +
       geom_bar(aes(weight = Cases_Property_Recovered), fill =
"skyblue") +
       theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
       labs(title = "Total Number of Cases by Group Name", x = "Group
Name", y = "Total Cases"
)
```

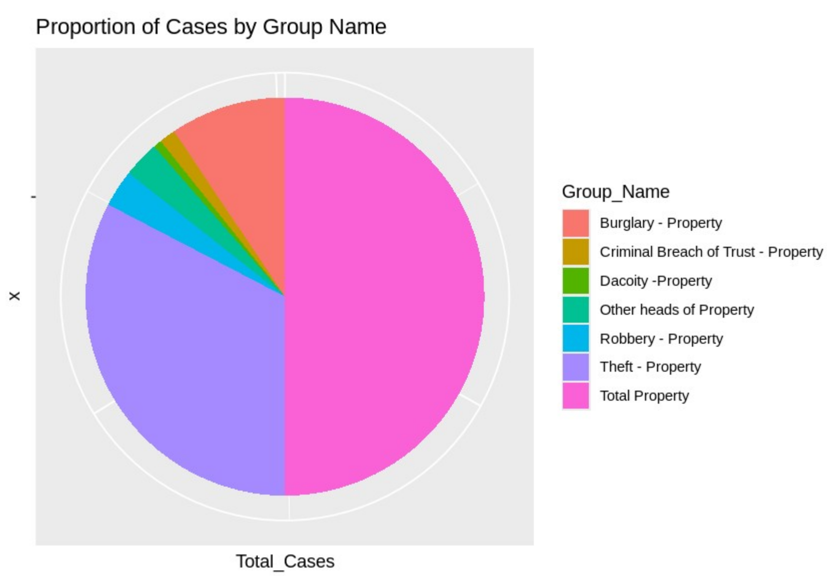


- The bar chart clearly illustrates the distribution of property crimes.
- Theft emerges as the most common crime.

- Burglary ranks second in prevalence, followed by robbery.
- Less frequent property crimes include criminal breach of trust and dacoity.
- This analysis offers important insights into property crime patterns and trends in the region.

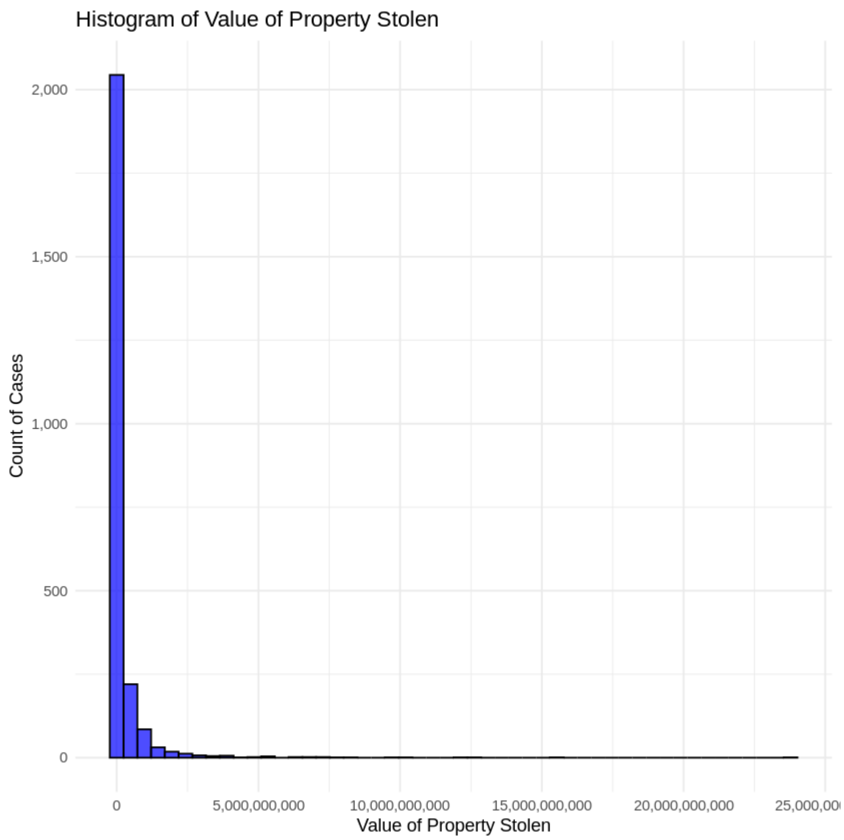
```
# group_summary
library(dplyr)
group_summary <- data %>%
  group_by(Group_Name) %>%
  summarise(Total_Cases = sum(Cases_Property_Recovered))

# Pie chart
ggplot(group_summary, aes(x = "", y = Total_Cases, fill = Group_Name))
+
  geom_bar(width = 1, stat = "identity") +
  coord_polar(theta = "y") +
  labs(title = "Proportion of Cases by Group Name") +
  theme(axis.text.x = element_blank())
```



The pie chart effectively visualizes the distribution of property crimes relative to the total number of cases. It reveals that theft-property is the most prevalent crime, followed by burglary-property and total property. Other property crime categories, such as criminal breach of trust, dacoity, other heads of property, and robbery, have smaller shares of the total cases.

```r
# Histogram
ggplot(data, aes(x = Value_of_Property_Stolen)) +
  geom_histogram(bins = 50, fill = "blue", color = "black", alpha =
0.7) +
  labs(title = "Histogram of Value of Property Stolen", x = "Value of
Property Stolen", y = "Count of Cases") +
  theme_minimal() +
  scale_x_continuous(labels = scales::comma) +
  scale_y_continuous(labels = scales::comma)
```



- The histogram clearly shows the distribution of stolen property values in crime cases.
- Most cases involve minor property losses.
- A few cases involve significant property theft, evident from the long tail on the right.
- This distribution highlights the occasional occurrence of large-scale thefts.

```r
data$Region <- case_when(
  data$Area_Name %in% c("Delhi", "Chandigarh", "Bihar", "Haryana",
"Himachal Pradesh", "Jammu & Kashmir", "Jharkhand", "Punjab",
"Rajasthan", "Uttar Pradesh", "Uttarakhand") ~ "North",
  data$Area_Name %in% c("Assam", "Arunachal Pradesh", "Manipur",
"Meghalaya", "Mizoram", "Nagaland", "Sikkim", "Tripura", "West
Bengal") ~ "East",
  data$Area_Name %in% c("Andhra Pradesh", "Chhattisgarh", "Goa",
"Karnataka", "Kerala", "Madhya Pradesh", "Maharashtra", "Odisha",
```

```
"Telangana", "Tamil Nadu", "Andaman & Nicobar Islands", "Lakshadweep",
"Puducherry") ~ "South",
  data$Area_Name %in% c("Daman & Diu", "Dadra & Nagar Haveli",
"Gujarat") ~ "West",
  TRUE ~ "Other"
)

other_data <- data[data$Region == "Other", ]
print(other_data)
```

```
[1] Area_Name                     Year
[3] Group_Name                    Sub_Group_Name
[5] Cases_Property_Recovered      Cases_Property_Stolen
[7] Value_of_Property_Recovered Value_of_Property_Stolen
[9] Region
<0 rows> (or 0-length row.names)
```

```
library(dplyr)

data_agg <- data %>%
  group_by(Year, Region) %>%
  summarize(Total_Cases = sum(Cases_Property_Stolen))
```

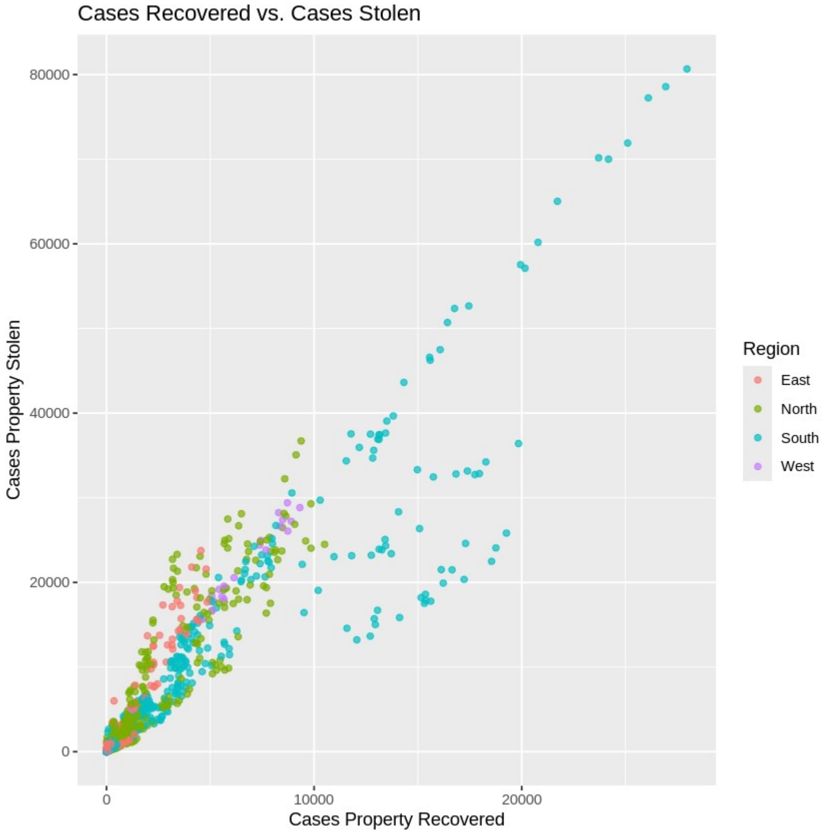
```
`summarise()` has grouped output by 'Year'. You can override using the
`.groups` argument.
```

```
# Timeline chart
ggplot(data_agg, aes(x = Year, y = Total_Cases, color = Region)) +
  geom_line(size = 0.8, alpha = 0.7) +
  geom_point(size = 3) +
  labs(
    title = "Trend of Property Stolen Cases Over Time",
    x = "Year",
    y = "Number of Cases"
  ) +
  theme_minimal() +
  scale_x_continuous(breaks = seq(min(data_agg$Year),
max(data_agg$Year), by = 2)) +
  scale_y_continuous(labels = scales::comma, limits = c(0,
max(data_agg$Total_Cases) * 1.2)) +
  theme(
    legend.position = "bottom",
    legend.title = element_blank()
  )
```

```
Warning message:
“Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
ℹ Please use `linewidth` instead.”
```

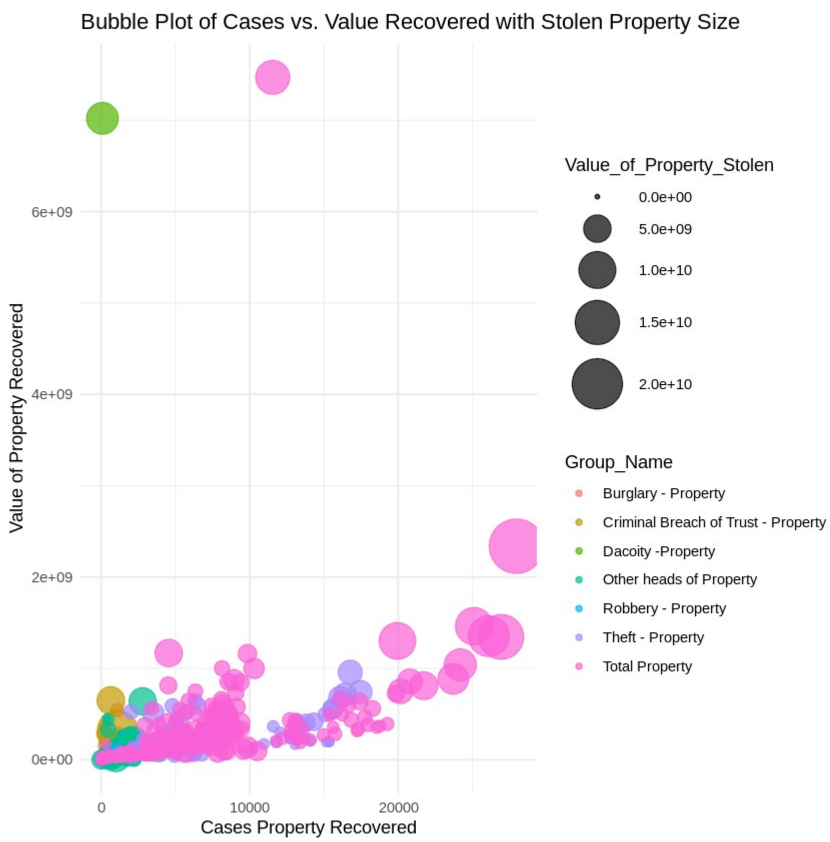
- The line chart illustrates property theft cases from 2001 to 2010 across regions.
- The South (blue) has the highest number of cases, consistently above 400,000, with a gradual increase.
- The North (green) shows steady growth, starting below 200,000 and rising over time.
- The East (pink) and West (purple) have lower and relatively stable trends, with cases remaining below 100,000.
- The data suggests the South faces the most severe and increasing property theft, while the East and West experience lower, stable levels.

```r
# Scatter plot
ggplot(data, aes(x = Cases_Property_Recovered, y =
Cases_Property_Stolen, color = Region)) +
  geom_point(alpha = 0.7) +
  labs(title = "Cases Recovered vs. Cases Stolen", x = "Cases Property
Recovered", y = "Cases Property Stolen")
```



- The data points show a diagonal trend, indicating a positive correlation: more stolen cases are linked to more recovered cases.
- Blue points (South) are more spread out and dominate the higher range on both axes, reflecting a higher volume of stolen and recovered cases.
- Clustering near the origin, especially for green (North) and pink (East) points, suggests lower activity in both stolen and recovered cases in these regions.
- The overall trend highlights regional differences in crime recovery rates, with the South showing more extreme values.

```
# Bubble plot
ggplot(data, aes(x = Cases_Property_Recovered, y =
Value_of_Property_Recovered, size = Value_of_Property_Stolen, color =
Group_Name)) +
  geom_point(alpha = 0.7) +
  scale_size_continuous(range = c(1, 15)) +
  labs(title = "Bubble Plot of Cases vs. Value Recovered with Stolen
Property Size", x = "Cases Property Recovered", y = "Value of Property
Recovered") +
  theme_minimal()
```



- The bubble plot shows the relationship between recovered property cases and the value of recovered property.
- Bubble sizes represent the value of stolen property.
- Most bubbles are clustered at lower values on both axes, indicating frequent recoveries involving low monetary amounts.
- Larger bubbles, representing high stolen property values, are rare and don't necessarily align with higher recovery rates.
- The plot emphasizes that high-value recoveries are uncommon, with crime categories (like Burglary, Theft, etc.) influencing the distribution.

Conclusion

From this experiment, I gained a comprehensive understanding of property crime dynamics through various data visualizations. The charts revealed that theft is the most common property crime, followed by burglary and robbery, while less frequent crimes include criminal breach of trust and dacoity. Regional analysis showed that the South experiences significantly higher rates of both stolen and recovered cases compared to other regions, with an upward trend over time, while the East and West remain relatively stable.

Additionally, the visualizations highlighted economic disparities in crime, showing that most stolen property cases involve minor losses, though some involve substantial thefts. The scatter and bubble plots emphasized that higher stolen property values do not necessarily correlate with higher recovery rates. Overall, these insights offer a clear picture of property crime patterns, regional variations, and the financial impact of such crimes.