R Notebook

Code ▾

This is an R Markdown (http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

Hide

```
#start
```

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.

Hide

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

Hide

```
# Load the dataset
crime_data <- read.csv("C:\\Users\\khush\\Downloads\\Crimes_-_2001_to_Present.csv\\Crimes_-_2001
_to_Present.csv")

library(forcats)
# Create a bar chart for Primary Type, sorted by frequency
g <- ggplot(crime_data, aes(x = fct_infreq(Primary.Type))) +
  geom_bar(fill = "skyblue") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(title = "Frequency of Primary Crime Types", x = "Crime Type", y = "Count")

# Print the bar chart
print(g)
```

Frequency of Primary Crime Types
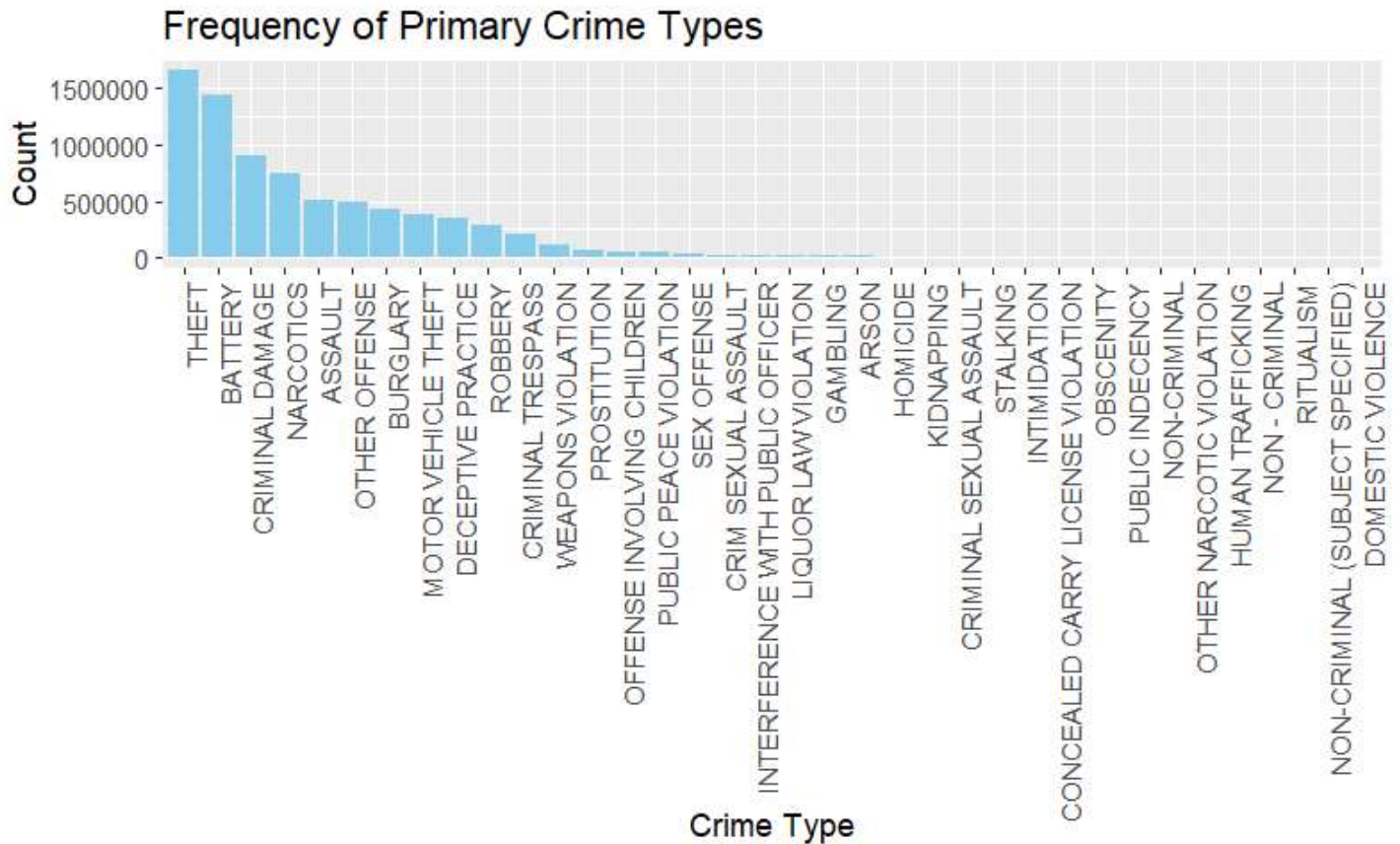
Questions Answered:

- Which crime types are most frequent?

Observations:

- The bar chart shows the frequency of each primary crime type, sorted by the most common types first.

- Theft, battery, criminal damage, narcotics are the most frequent ones.

Hide

```
# Summarize the data by block
block_summary <- crime_data %>%
  group_by(Block) %>%
  summarise(Incident_Count = n(), .groups = 'drop') %>%
  arrange(desc(Incident_Count))

# Get the top 15 blocks
top_15_blocks <- head(block_summary, 15)

# Create a bar chart
ggplot(top_15_blocks, aes(x = reorder(Block, Incident_Count), y = Incident_Count)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  coord_flip() +  # Flip coordinates for better readability
  labs(title = "Top 15 Blocks with Most Incidents",
       x = "Block",
       y = "Number of Incidents") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Top 15 Blocks with Most Incidents
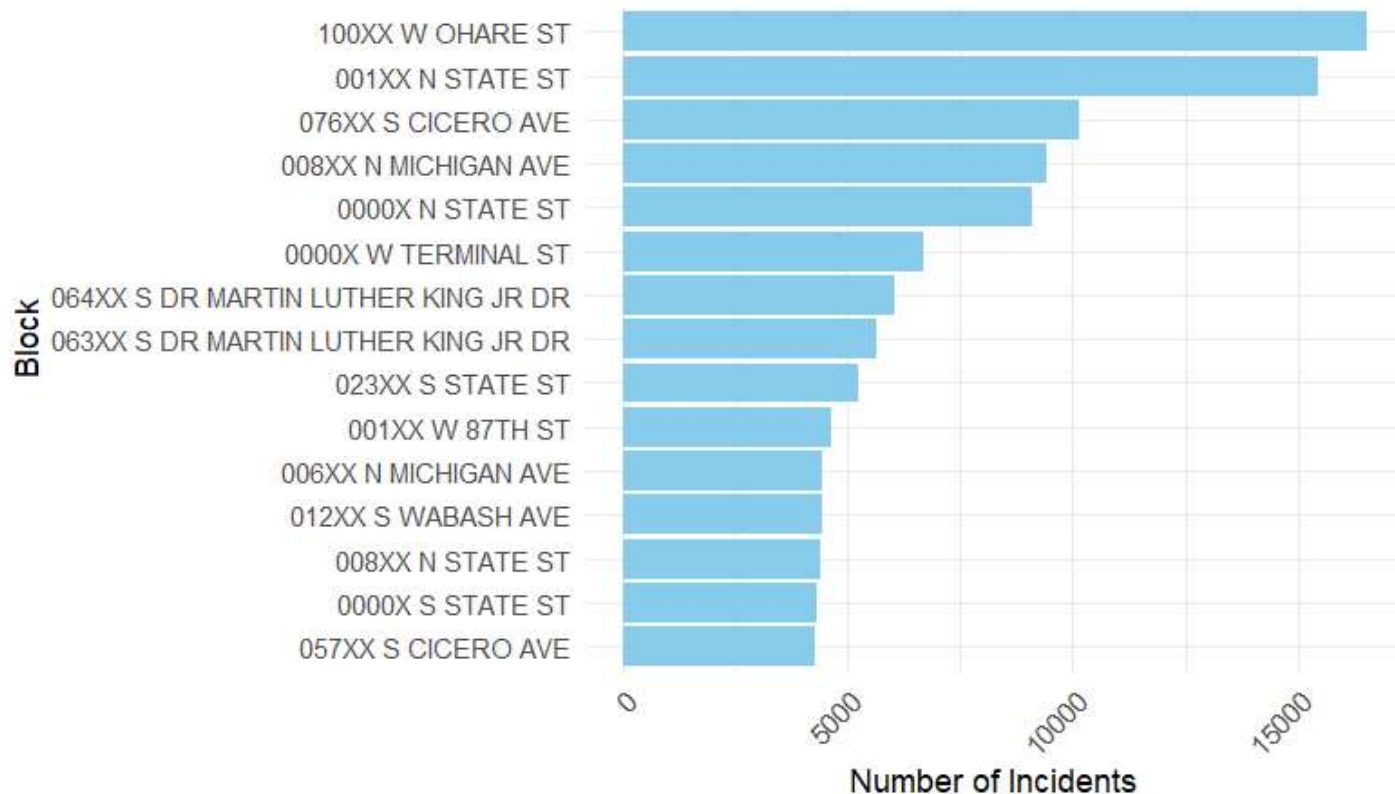
Questions Answered:

- Which blocks have the highest number of incidents?

Observations:

- Incident Distribution: The chart shows the blocks with the most incidents, highlighting where criminal activities are concentrated.

- 100XX W OHARE ST is the block with most number of incidents.

Hide

```
# Summarize the data for Arrest
arrest_data <- table(crime_data$Arrest)

# Convert to a data frame and make sure 'Count' is numeric
arrest_df <- as.data.frame(arrest_data)
names(arrest_df) <- c("Arrest", "Count")
arrest_df$Count <- as.numeric(arrest_df$Count)

# Create a pie chart for Arrests
ggplot(arrest_df, aes(x = "", y = Count, fill = Arrest)) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar(theta = "y") +
  scale_fill_manual(values = c("red", "green")) +
  labs(title = "Proportion of Arrests") +
  theme_void() +
  theme(legend.position = "right")
```

Proportion of Arrests
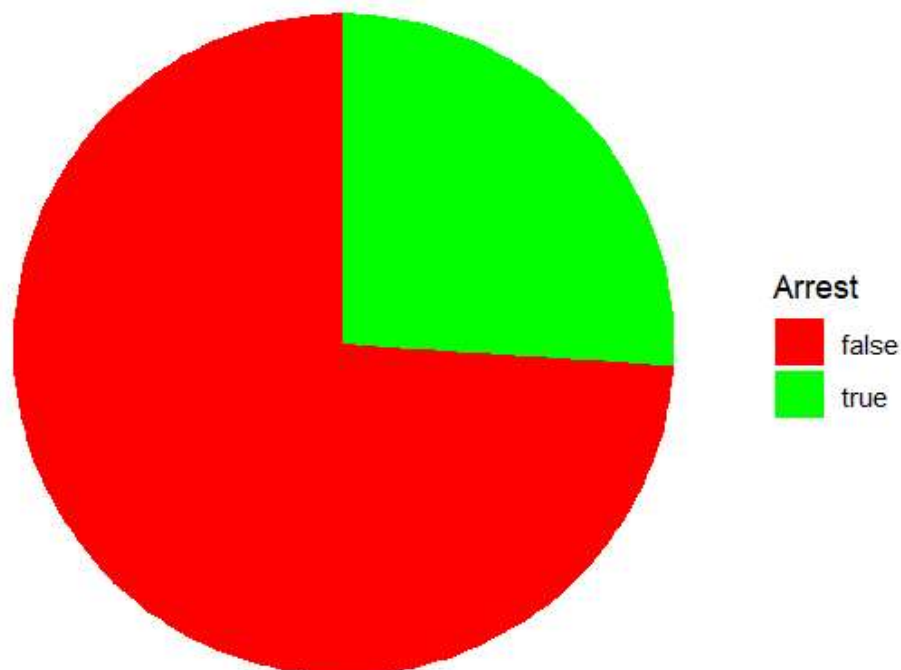
Questions Answered:

- What is the proportion of cases where arrests were made compared to cases where no arrest was made?

Observations:

- This pie chart shows the proportion of incidents with and without arrests.

- It shows that the number of incidents with arrests is less than the number of incidents without arrests, highlighting a larger proportion of cases where no arrest was made.

Hide

```
# Summarize the data for Domestic
domestic_data <- table(crime_data$Domestic)

# Convert to a data frame and make sure 'Count' is numeric
domestic_df <- as.data.frame(domestic_data)
names(domestic_df) <- c("Domestic", "Count")
domestic_df$Count <- as.numeric(domestic_df$Count)

# Create a pie chart for Domestic
ggplot(domestic_df, aes(x = "", y = Count, fill = Domestic)) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar(theta = "y") +
  labs(title = "Proportion of Domestic vs Non-Domestic Incidents") +
  theme_void() +
  theme(legend.position = "right")
```

Proportion of Domestic vs Non-Domestic Incidents
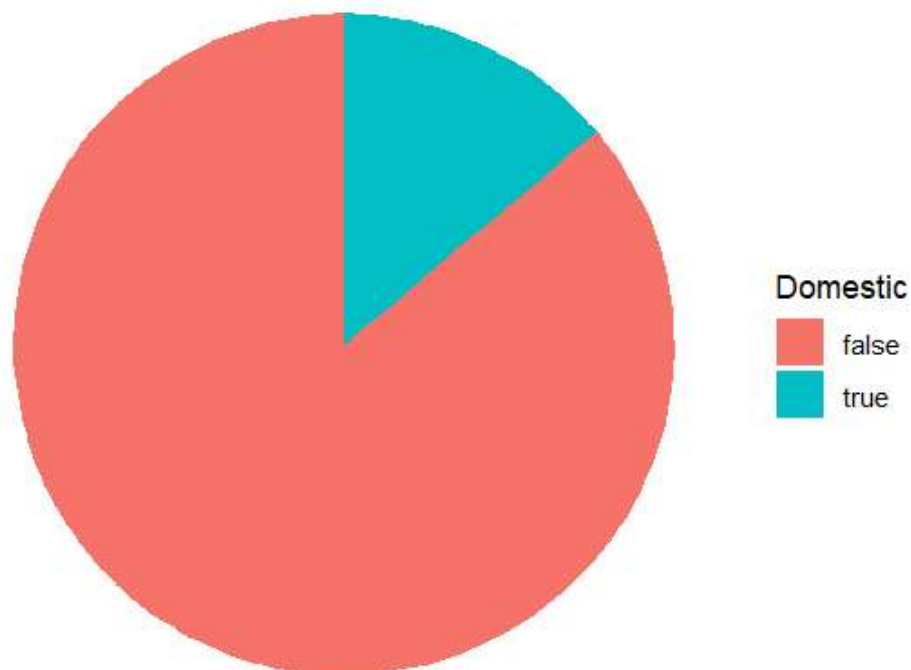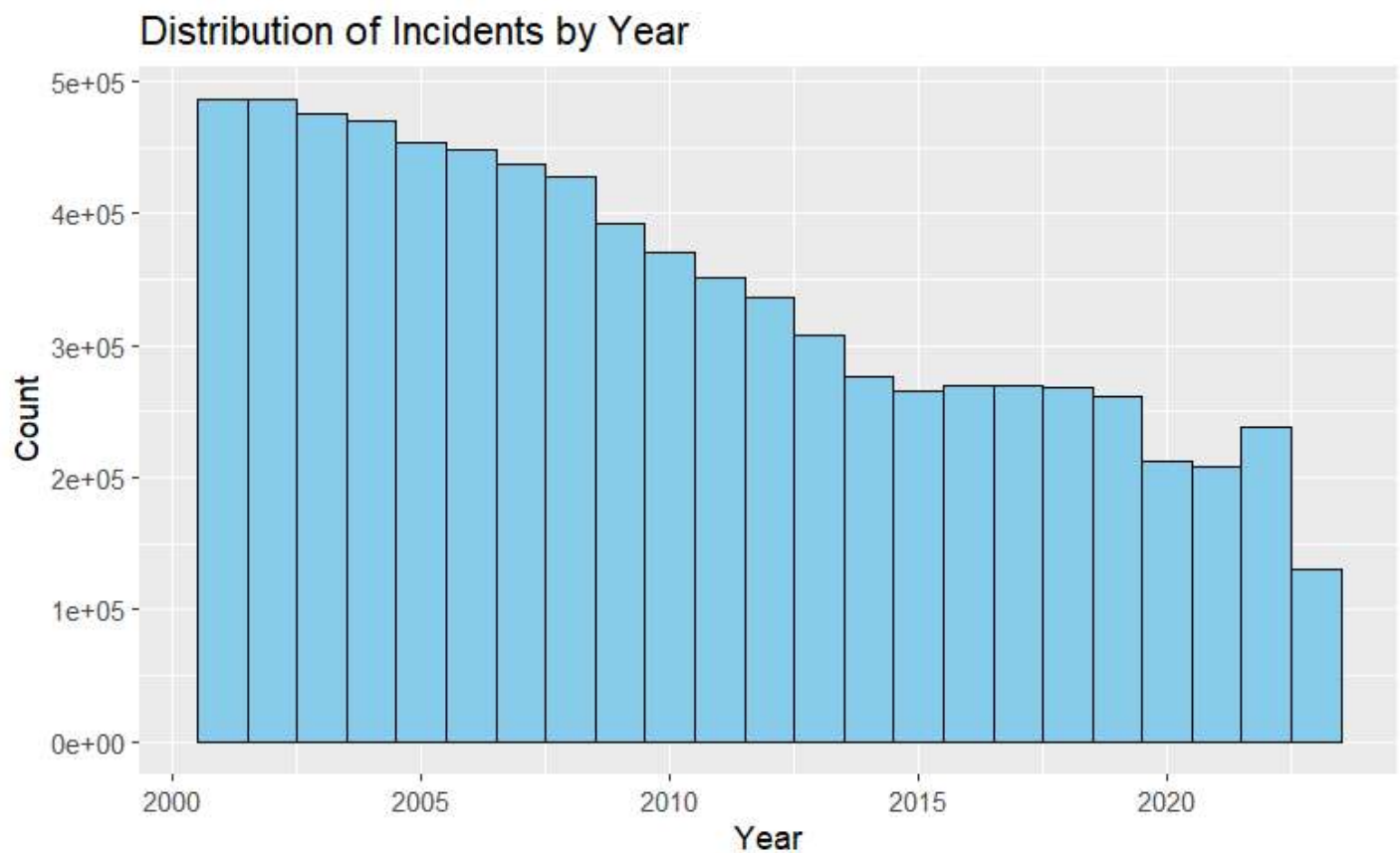

Questions Answered:

- What is the proportion of domestic incidents compared to non-domestic incidents?

Observations:

- The pie chart provides a visual representation of the proportion of domestic versus non-domestic incidents.

- Domestic incidents are less frequent compared to non-domestic incidents, indicating a larger proportion of non-domestic cases in the dataset.

Hide

```
# Create a histogram for the 'Year' column
ggplot(crime_data, aes(x = Year)) +
  geom_histogram(binwidth = 1, fill = "skyblue", color = "black") +
  labs(title = "Distribution of Incidents by Year", x = "Year", y = "Count")
```

Distribution of Incidents by Year

Questions Answered:

- How are incidents distributed over the years?

- Are there any trends or patterns in the number of incidents per year?
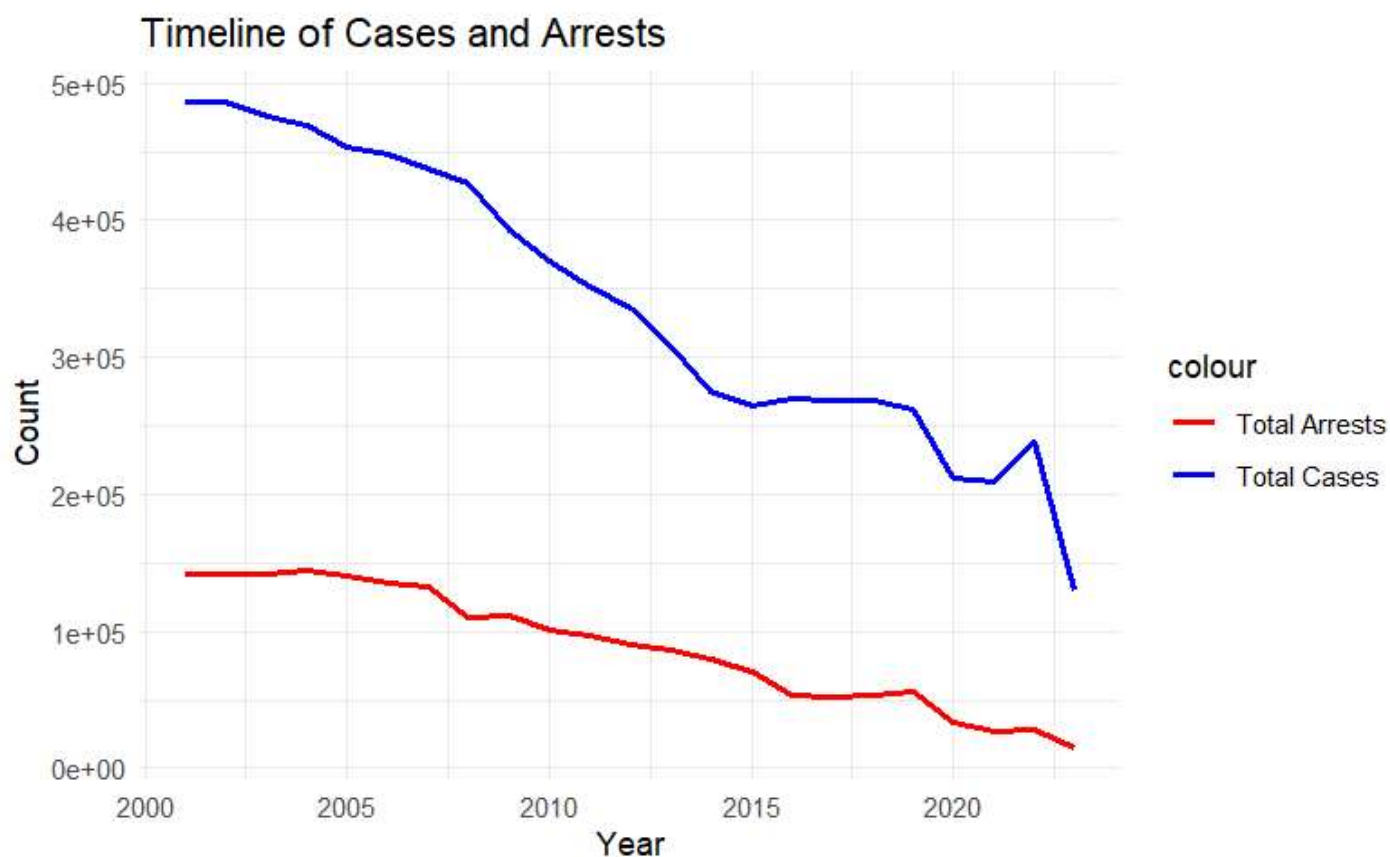
Observations:

- The histogram shows the distribution of incidents by year.

- It highlights how the number of incidents has varied from year to year.

- There is a noticeable decrease in the number of incidents in recent years, suggesting a decline in overall incidents over time.

Hide

```
# Convert 'Date' to Date type
crime_data$Date <- as.Date(crime_data$Date, format = "%m/%d/%Y %I:%M:%S %p")

# Group by year and count total cases and arrests per year
crime_summary <- crime_data %>%
  group_by(Year) %>%
  summarise(Total_Cases = n(),
            Total_Arrests = sum(Arrest == "true", na.rm = TRUE))

# Create a timeline plot for Cases vs. Arrests
ggplot(crime_summary, aes(x = Year)) +
  geom_line(aes(y = Total_Cases, color = "Total Cases"), size = 1) +
  geom_line(aes(y = Total_Arrests, color = "Total Arrests"), size = 1) +
  labs(title = "Timeline of Cases and Arrests", x = "Year", y = "Count") +
  scale_color_manual(values = c("Total Cases" = "blue", "Total Arrests" = "red")) +
  theme_minimal()
```


Questions Answered:

- How have the number of total cases and total arrests changed over the years?

- Are there any noticeable trends or patterns in the total number of cases and arrests?

Observations:

- There is a significant decline in the total number of cases over time.

- The number of arrests has shown a more consistent decrease, indicating that while fewer cases are being reported, the rate of arrests is not declining as sharply.

```
# Calculate total counts for each crime type
crime_counts <- crime_data %>%
  group_by(Primary.Type) %>%
  summarise(Total = n()) %>%
  arrange(desc(Total))

# Get the top 5 most frequent crimes
top_5_crimes <- head(crime_counts, 5)
print(top_5_crimes)
```

| Primary.Type	Total
<chr>	<int>
THEFT	1654618
BATTERY	1433129
CRIMINAL DAMAGE	894068
NARCOTICS	748690
ASSAULT	512441
5 rows	

```
# Filter dataset to include only the top 5 crimes
top_5_crimes_list <- top_5_crimes$Primary.Type
filtered_data <- crime_data %>%
  filter(Primary.Type %in% top_5_crimes_list)

# Summarize the data by year and crime type
timeline_data <- filtered_data %>%
  group_by(Year, Primary.Type) %>%
  summarise(Count = n(), .groups = 'drop')

# Create a timeline plot for the top 5 crimes
ggplot(timeline_data, aes(x = Year, y = Count, color = Primary.Type)) +
  geom_line() +
  labs(title = "Number of Crimes by Year for the Top 5 Most Common Types",
       x = "Year",
       y = "Count of Crimes",
       color = "Crime Type") +
  theme_minimal()
```

Number of Crimes by Year for the Top 5 Most Common Types
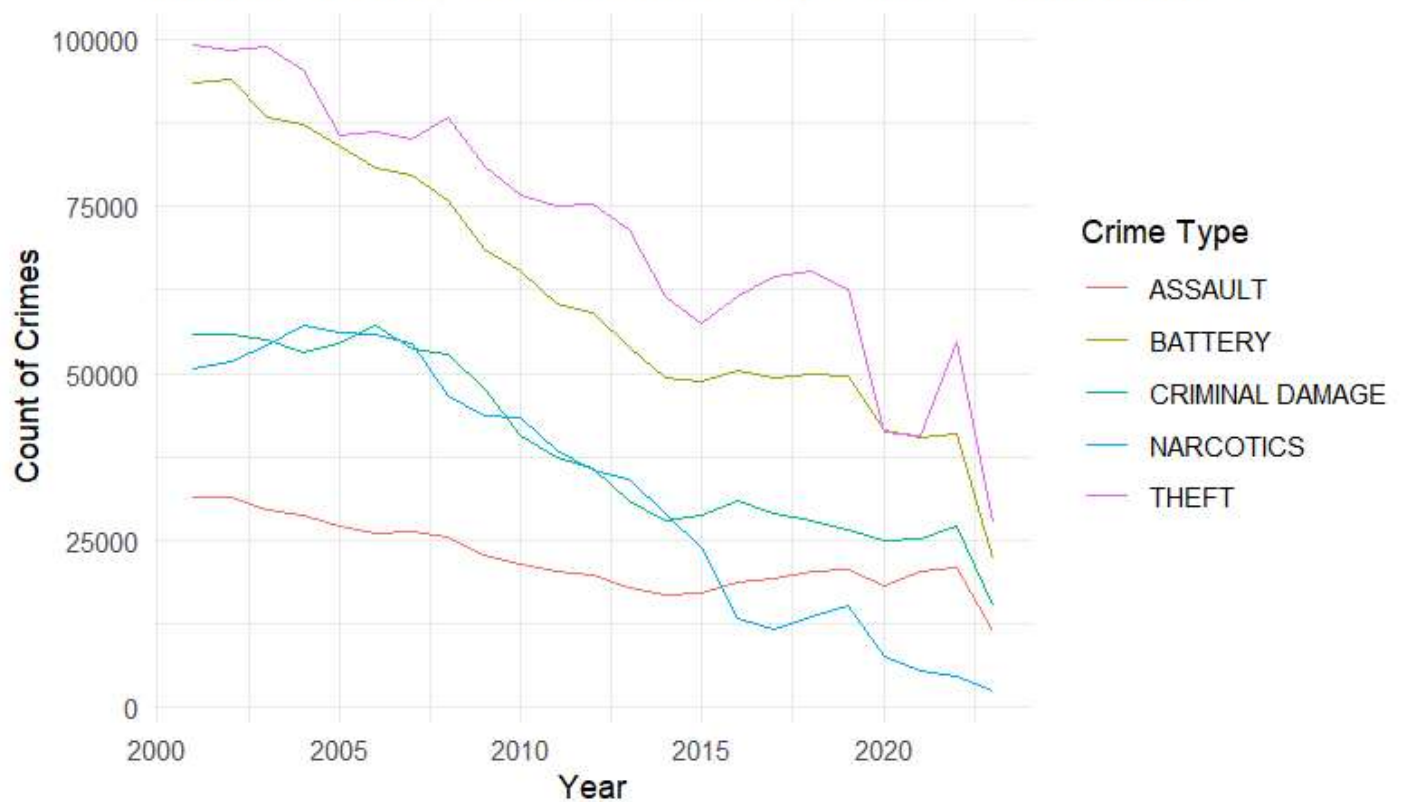
Questions Answered:

- What are the trends in the number of cases for the top 5 most common crime types over the years?

- How do the frequencies of these top 5 crimes change annually?

Observations:

- Narcotics, Theft, and Battery show a significant decline in the number of reported cases over the years.

- Assault has shown relative stability, with little change in its frequency over time compared to the other crime types.

Hide

```
# Filter dataset to include only the top 5 crimes
top_5_crimes_list <- top_5_crimes$Primary.Type
filtered_data <- crime_data %>%
  filter(Primary.Type %in% top_5_crimes_list)

# Summarize the data by year and crime type
summary_data <- filtered_data %>%
  group_by(Year, Primary.Type) %>%
  summarise(Cases = n(),
            Arrests = sum(Arrest == "true"),.groups = 'drop')

# Create a scatter plot
ggplot(summary_data, aes(x = Cases, y = Arrests, color = Primary.Type)) +
  geom_point(alpha = 0.7, size = 3) +
  labs(title = "Number of Cases vs. Number of Arrests for Top 5 Crimes",
       x = "Number of Cases",
       y = "Number of Arrests",
       color = "Primary Type") +
  theme_minimal()
```

Number of Cases vs. Number of Arrests for Top 5 Crimes
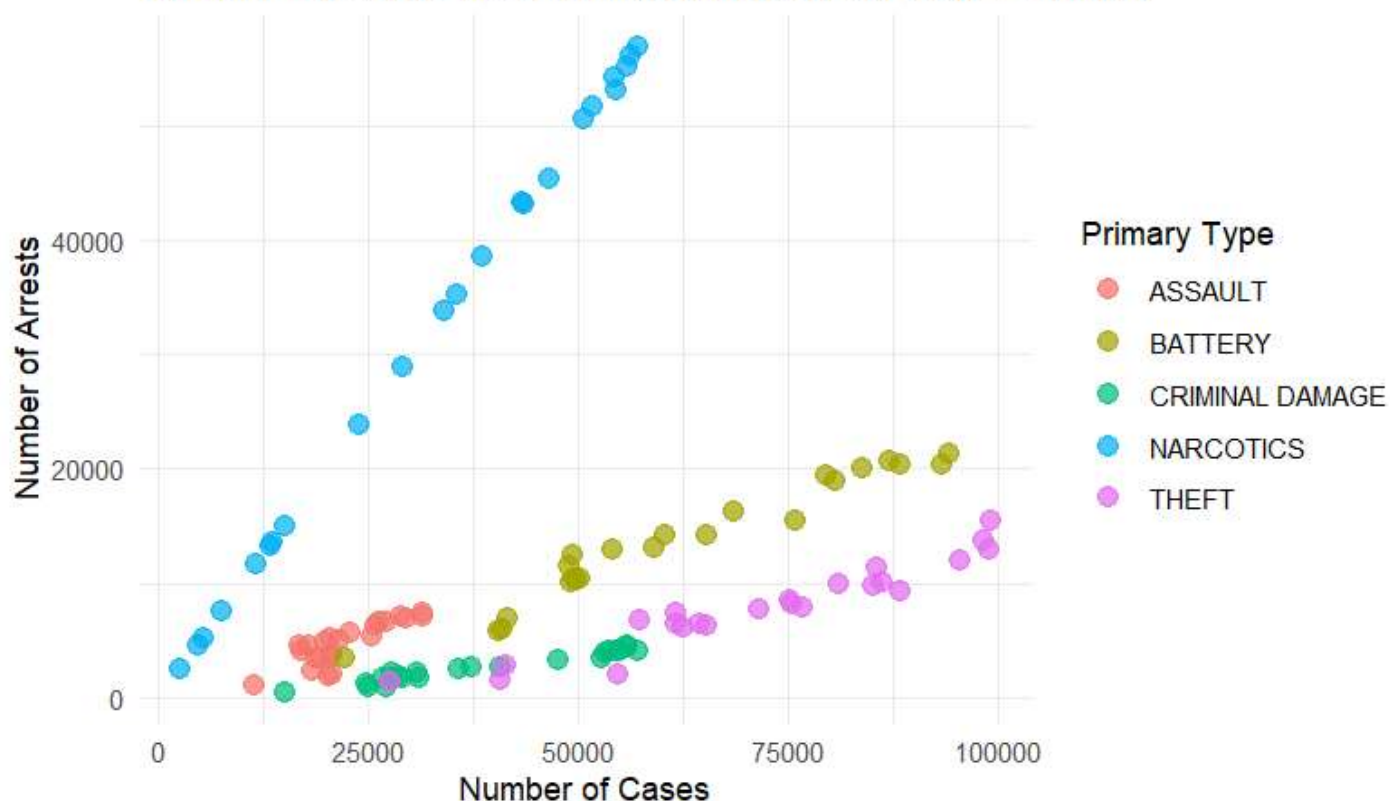
Questions Answered:

- What is the relationship between the number of cases and the number of arrests for the top 5 crimes?

- Are there any patterns in how arrests correspond to the number of cases for different crime types?
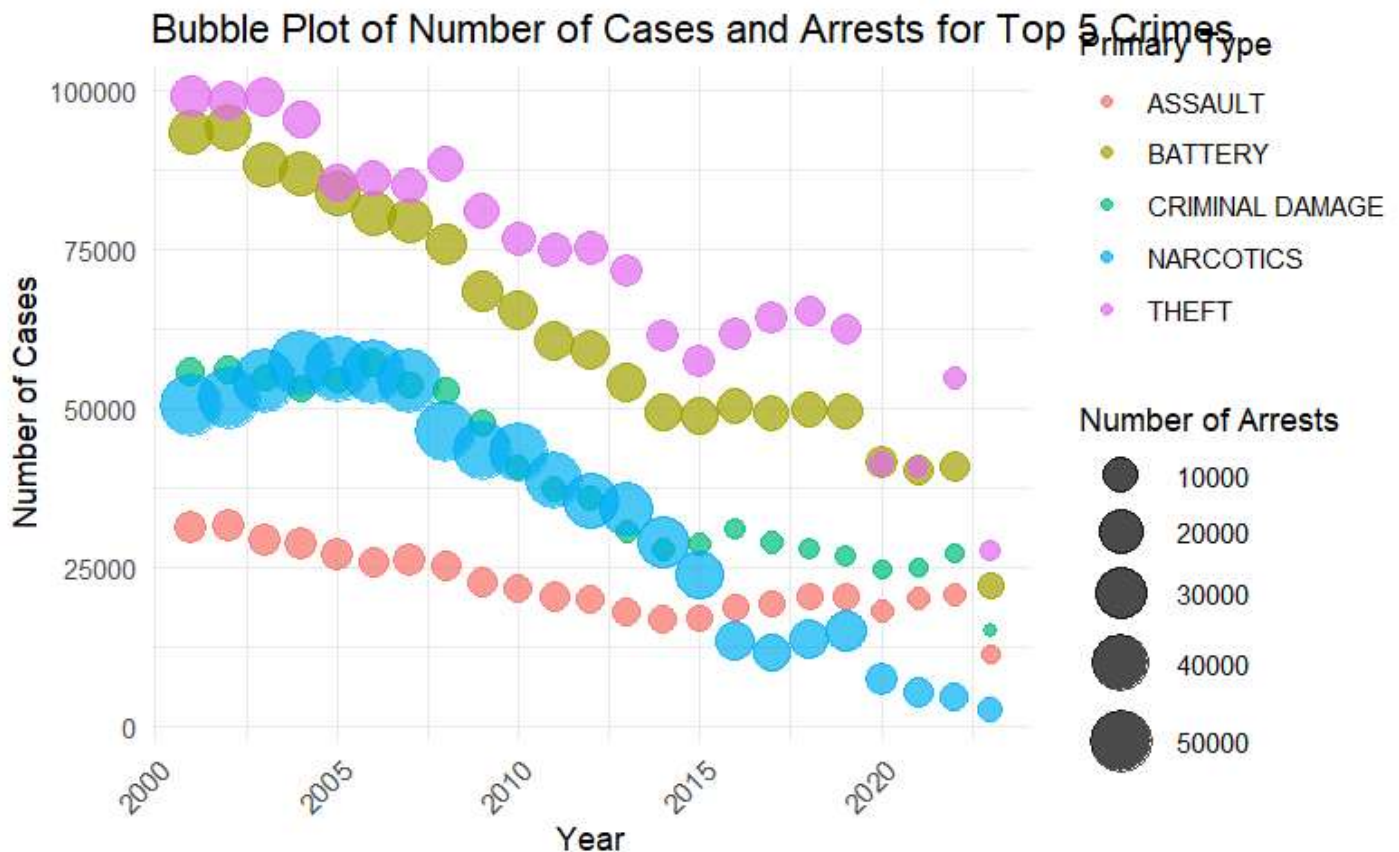
Observations:

- There is a general trend where the number of arrests increases with the number of cases for most crime types.

- Narcotics shows the strongest correlation, with a notable increase in arrests corresponding to a higher number of cases.

- Other crime types also exhibit an increase in arrests with more cases, but the correlation is less pronounced compared to narcotics.

Hide

```
# Summarize the data by year and crime type
summary_data <- filtered_data %>%
  group_by(Year, Primary.Type) %>%
  summarise(Cases = n(),
            Arrests = sum(Arrest == "true"),  # Summarize the number of arrests
            .groups = 'drop')

# Create a bubble plot
ggplot(summary_data, aes(x = Year, y = Cases, size = Arrests, color = Primary.Type)) +
  geom_point(alpha = 0.7) +
  scale_size_continuous(range = c(2, 10), name = "Number of Arrests") +
  labs(title = "Bubble Plot of Number of Cases and Arrests for Top 5 Crimes",
       x = "Year",
       y = "Number of Cases",
       color = "Primary Type") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```


Bubble Plot of Number of Cases and Arrests for Top 5 Crimes

Questions Answered:

- What is the relationship between the number of cases and the number of arrests for different crime types over the years?

Observations:

- Bubble Size reflects the number of arrests. Larger bubbles indicate a higher number of arrests.

- Color Coding allows comparison of different crime types and their trends.

- Narcotics shows consistently larger bubbles and a significant drop in the number of cases over the years.