Crime Analysis of Los Angeles

CMP5352 Data Visualization

Sudeep Fullel
Student Id: 23140750
Bachelor of Science with Honours Computer and Data Science
(Faculty of Computing, Engineering, and the Built Environment)
Birmingham City University (BCU)
Sunway college
Kathmandu, Nepal
Sudeep.Fullel@mail.bcu.ac.uk

Word count:

Abstract

This report looks into crime patterns and trends in Los Angeles using data from the Los Angeles Police Department (LAPD) between 2020 and 2023. Through data visualization techniques, the study aims to find important insights about when and where crimes happen, who the victims are, what weapons are used, and how often cases are solved. The analysis uncovers significant trends like more crimes happening in certain areas, different types of crimes occurring at different times of day, and differences in victims' backgrounds. The study also explores how different crimes relate to the ages of victims in terms of weapon use, giving a comprehensive view of crime in Los Angeles.

Contents

Introduction

Data visualization is an important technique for extracting insights from large datasets. By converting data into helpful visuals, we can discover previously unseen patterns and trends. This enables us to explore stories and make more informed data-driven decisions.

This report aims to investigate crime patterns and trends in Los Angeles utilizing data given by the Los Angeles Police Department (LAPD). The dataset contains information about different crimes in Los Angeles since 2020.

The study begins by outlining why the specific dataset was chosen and what important non-trivial questions may be answered with it. The data is then further examined with the R programming language. This covers tasks like data wrangling and exploratory data analysis (EDA) to identify insights using simple visualizations. Following that, the study moves on to the most interesting section: visualizing non-trivial topics and leveraging the visuals to identify answers. Finally, a summary of the findings is presented, together with the conclusions made.

Motivation and Objectives

Los Angeles crime data is a open data provided by the Los Angeles Police Department (LAPD). It's a useful tool for analyzing crime patterns in Los Angeles. The dataset covers a variety of criminal activity records from 2020 to 2023. The reasons for choosing this dataset were:

- **Relevance:** Crime data is important because it helps improve public safety, allocate legal resources, and find ways to reduce crime.
- **Granularity:** I am personally interested in finding crime patterns and developing solutions through data.
- **Accessibility:** The data is accessible to all, promoting transparency and allowing the study of crime patterns.

2.1 Non-Trivial Research Questions:

From the crime data, following research questions were generated which delve into the complexities of crime in Los Angeles, seeking patterns, relationships, and potential contributing factors that go beyond simple descriptive analysis:

1. **Temporal and Spatial Trends in High-Crime Area:**

   Question: Does the trend of crime concentration in certain areas differ for different crime types??

   Non-Trivial: This question explores the relationship between crime type and geographic concentration over time, It's not immediately clear if the areas with high crime concentrations in certain crime types also expe-rience significant increases in those crime types over time.

2. **Crime Disparities by Demographic and Time: Question:** What crimes are dominant in nighttime and daytime and do the victim demographics in terms of age and gender differ between day and night. **Non-Trivial:** This question connects three different dots: victim demographics, crime types by area, and time of day. It explores whether patterns in victim characteristics can be used to predict the distribution of crime and which crime is dominant in daytime and night.

3. **Consistency and Variations in Crime Patterns: Question:** Do crime patterns in Los Angeles exhibit consistent trends across different time scales (monthly, seasonal, daily) throughout the years, or are there noticeable variations and inconsistencies that suggest the influence of external factors? **Non-Trivial:** This question examines the consistency of crime patterns across different time scales, potentially uncovering the influence of seasonal factors, holiday periods and time that might contribute to variations in crime

rates.

4. **Co-occurrence of Crime Types: Question:** Do certain crime types in Los Angeles frequently occur together at the same locations and time, and if so, what crimes are more likely to occur if one happens? **Non-Trivial:** This analysis investigates the potential relationships between different crime types, examining whether certain crimes tend to occur in close proximity, which could suggest a shared context or potential links between criminal activities.

5. **Weapon Use and Victim Age Relationships: Question:** In the most crime occurring area in Los Angeles, is there a relationship between the age of the victim and the type of weapon used? **Non-Trivial:** This analysis examines the potential association between victim age and weapon type, seeking to understand whether specific weapons are more likely to involve with vitims of specific ages.

6. **Case Clearance Rates and Variations: Question:** Do the differences in case clearance rates across different crime types suggest variations in investigative resources, police priorities, or the inherent difficulty of solving specific crimes?? **Non-Trivial:** This question investigates the success rate of police investigations for different crime types, examining potential factors that contribute to variations in case clearance rates, such as the nature of the crime, the availability of evidence, or resource allocation.

7. **Spatial Clustering and Geographic Features: Question:** Does the spatial distribution of crime in Los Angeles, visualized by area's crime and most common crime types, reveal any clustering patterns or relationships with geographic features? **Non-Trivial:** This question goes beyond simply observing the location of crimes. It explores potential correlations between the spatial arrangement of crimes and factors related to the physical environment.

Experimental Results

This section presents the findings derived from analyzing the LAPD crime dataset using data visualization techniques. We will explore the data through a series of steps, starting with initial data exploration and preparation, then moving towards descriptive analysis and pattern identification.

3.1 Understand the Data

- **Data Source:** The dataset used for this analysis is sourced from the LAPD OpenData platform. The dataset can be accessed through the following link: `https://catalog.data.gov/dataset/crime-data-from-2020-to-present`.

- **Data Description:** The dataset covers crime incident records from 2020 to 2023. It contains a total of 948K observations and 28 variables. The variables include information such as crime type, date occurred, location (latitude/longitude), victim age, weapon used, etc.

- **Initial Exploration:**

  The CSV file "Crime_Data_from_2020_to_Present.csv" was read into a data frame called "crime_data". We identified potential data type issues in the "Crm Cd 4" column and manually specified the data type using the "col_types" argument in the "read_csv()" function. Similarly, we explored the dataset using functions like `head()`, `str()`, and `summary()` in R to gain insights into the dataset's structure and variable types.

  Key variables relevant to our research questions include `crime_type`, `date_occurred`, `location`, `victim_age`, `weapon_used`, etc.

```r
#Required libraries
library(readr)
library(plotly)
library(dplyr)
library(lubridate)
library(RColorBrewer)
library(viridis)
library(treemap)
library(arules)
library(arulesViz)
library(ggplot2)
library(extrafont)
```

```r
# Understanding and loading the Data
crime_data <- read_csv("Crime_Data_from_2020_to_Present.csv")
```

```
## Rows:  932140 Columns:  28
## -- Column specification ------------------------------------------------------------
## Delimiter:  ","
## chr (16):  Date Rptd, DATE OCC, TIME OCC, AREA, AREA NAME, Rpt Dist No,
Crm C...
## dbl (11):  DR_NO, Part 1-2, Crm Cd, Vict Age, Premis Cd, Weapon Used Cd,
Crm ...
## lgl  (1):  Crm Cd 4
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this
message.

# finding the part which is causing problem in data
problems(crime_data)

#checking the column name that has problem
head(crime_data[,24])

crime_data <- read_csv("Crime_Data_from_2020_to_Present.csv",
  col_types = cols(
    `Crm Cd 4` = col_character(),
    .default = col_guess()
)) # providing appropriate data type to solve the problem

#checking if any problem occurs again or not
problems(crime_data)

head(crime_data)

str(crime_data)

summary(crime_data)
```

3.2 Data Wrangling

- **Column Renaming:**

  The columns were renamed for better readability and understanding.

```
# changing columns names for better understanding
crime_data <- crime_data %>%
  rename(
    division_number = DR_NO,
```

```
    date_reported = `Date Rptd`,
    date_occurred = `DATE OCC`,
    time_occurred = `TIME OCC`,
    area = AREA,
    area_name = `AREA NAME`,
    reporting_district = `Rpt Dist No`,
    part = `Part 1-2`,
    crime_code = `Crm Cd`,
    crime_description = `Crm Cd Desc`,
    modus_operandi = Mocodes,
    victim_age = `Vict Age`,
    victim_sex = `Vict Sex`,
    victim_descent = `Vict Descent`,
    premise_code = `Premis Cd`,
    premise_description = `Premis Desc`,
    weapon_code = `Weapon Used Cd`,
    weapon_description = `Weapon Desc`,
    status = Status,
    status_description = `Status Desc`,
    crime_code_1 = `Crm Cd 1`,
    crime_code_2 = `Crm Cd 2`,
    crime_code_3 = `Crm Cd 3`,
    crime_code_4 = `Crm Cd 4`,
    location = LOCATION,
    cross_street = `Cross Street`,
    latitude = LAT,
    longitude = LON
  )
```

- **Data Type Conversion:**

  Relevant columns were converted to appropriate data types so that there would not be any problems in analyzing the data in further report.

```
#converting data to appropriate data types
crime_data <- crime_data %>%
  mutate(
    date_occurred = as.POSIXct(date_occurred, format = "%m/%d/%Y %I:%M:%S %p"),
    time_occurred = as.POSIXct(sprintf("%s", time_occurred), format = "%H%M")
  )
```

- **Missing Data Handling:**

  Many columns with missing values were identified. The dataset had incomplete data for 2024, which was then removed. Missing values in the "victim_sex" column were filled with

"X," and in the "weapon_description" column, they were filled with "Unknown." This was done to keep the data consistent and avoid any biases due to missing information.

```
#Missing Data:
colSums(is.na(crime_data))

#Data Cleaning
#removing data of 2024 beacuse it is incomplete
crime_data <- crime_data %>%
  filter(year(date_occurred) != 2024)

#imputing missing values in sex and weapon description column
unique(crime_data$victim_sex)

crime_data <- crime_data %>%
  mutate(
    victim_sex = ifelse(is.na(victim_sex) |
                        victim_sex == "H" |
                        victim_sex == "-",
                        "X", victim_sex),
    weapon_description = ifelse(is.na(weapon_description),
                                "Unknown", weapon_description)
  )
```

3.3 Dimensionality Reduction

Due to the large number of columns in the dataset, processing and handling became challenging. Therefore, we decided to narrow down our analysis by selecting a subset of relevant variables based on our research questions.

```
#Dimensionality Reduction :section
# Feature Selection based on the research questions
crime_data <- crime_data %>%
  select(
    date_occurred,
    time_occurred,
    area_name,
    crime_description,
    victim_age,
    victim_sex,
    location,
    weapon_description,
    status_description,
    latitude,
```

```
    longitude
  )
```

3.4 Feature Engineering

After considering the requirements for addressing our research questions, it was determined that certain features were necessary. To enhance the dataset for analysis, new features were created. These included temporal features like Year, Month, Weekday, and Season, which allowed for an examination of crime trends across various time scales. Additionally, Time of Day features such as the Hour of occurrence and a "Day/Night" indicator were added to facilitate the analysis of crime patterns throughout the day.

```
# Feature Engineering :section

get_season <- function(month) {
  case_when(
    month %in% c(12, 1, 2) ~ "Winter",
    month %in% c(3, 4, 5) ~ "Spring",
    month %in% c(6, 7, 8) ~ "Summer",
    month %in% c(9, 10, 11) ~ "Fall"
  )
}

# Add year, month, weekday, and season columns
# Define night hours
start_hour = 18
end_hour = 6

crime_data <- crime_data %>%
  mutate(
    year = year(date_occurred),
    month = month(date_occurred, label = TRUE, abbr = TRUE),
    weekday = wday(date_occurred, label = TRUE, abbr = TRUE),
    season = get_season(month(date_occurred)),
    hour_occurred = hour(time_occurred),
    part_of_day = ifelse((hour_occurred >= start_hour) |
                           (hour_occurred < end_hour), "Night", "Day")
  )
```

The cleaned dataset now was saved as "cleaned_crimedata.csv" for shiny package use using follwing command:

```
write_csv(crime_data, "cleaned_crimedata.csv")
```

3.5 Descriptive Analysis

This section presents some basic visualizations to get a preliminary understanding of the crime data.

3.5.1 Number of Crimes Over Time

```
ggplot(crime_data, aes(x = year)) +
  geom_line(stat = "count", color = "#7f0c0f",size = 0.8) +
  geom_hline(yintercept = seq(200000, 240000, by = 10000), color = "grey",
             size = 0.4, linetype = "solid") +
  labs(title = "Number of Crimes Over Time", x = "Years",
             y = "Number of Crimes") +
  theme_minimal() +
  theme(text = element_text(family = "sans", colour = "black"),
        plot.title = element_text(size = 22, face = "bold", hjust = 0.5),
        axis.text.x = element_text(angle = 0, vjust = 0.5, size = 13,
                                   colour = "black"),
        axis.text.y = element_text(size = 13),
        axis.title.x = element_text(size = 16, margin = margin(t = 10)),
        axis.title.y = element_text(size=16, margin = margin(t=10,r=10)),
        axis.line.y = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black")) +
  geom_text(stat = "count",
            aes(label = ifelse(x != 2020,
            paste0(round((..count.. - lag(..count..)) /
                          lag(..count..) * 100, 2), "%"), "")),
            vjust = -1,hjust=0.8, size = 4, color = "#7f0c0f")
```

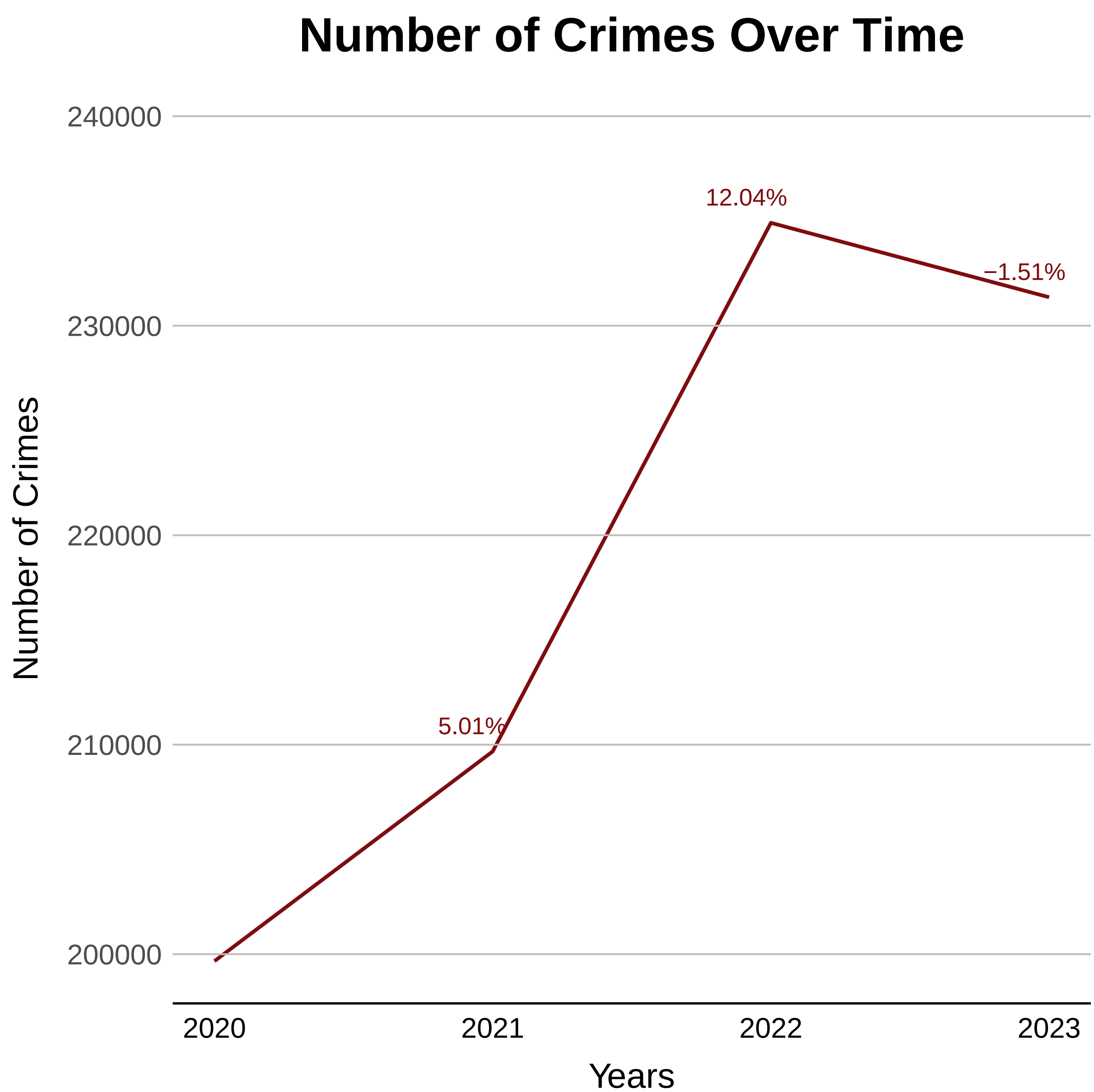

Figure 1: Number of Crimes Over Time

This line graph depicts the total number of crimes reported in Los Angeles from 2020 to 2023. It reveals a consistent upward trend in crime incidents over the years, indicating a potential need for increased crime prevention strategies and resource allocation.

3.5.2 Number of Crimes by Area

```
# Calculating the count of crimes for each area
crime_counts <- crime_data %>%
  group_by(area_name) %>%
```

```r
  summarise(crime_count = n()) %>%
  arrange(desc(crime_count))  # Arranging by descending order of crime count

# Reorder the levels of area_name based on crime_counts
crime_data$area_name <- factor(crime_data$area_name,
                               crime_counts$area_name)

# Plotting the bar chart with ordered levels and similar style to the first plot
ggplot(crime_data, aes(x = area_name)) +
  geom_bar(fill = "#1380A1") +
  labs(title = "Number of Crimes by Area", x = "Area Name",
       y = "Number of Crimes") +
  theme_minimal() +
  theme(text = element_text(family = "sans", colour = "black"),
        plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
        axis.text.x = element_text(angle = 75, hjust = 1, vjust = 1,
                                   size = 12, colour = "black"),
        axis.text.y = element_text(size = 12, colour = "black"),
       axis.title.x = element_text(size = 16, margin = margin(t = 10)),
        axis.title.y = element_text(size=16, margin = margin(t=10,r=10)),
        axis.line.y = element_blank(),
        axis.line.x = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black"))
```

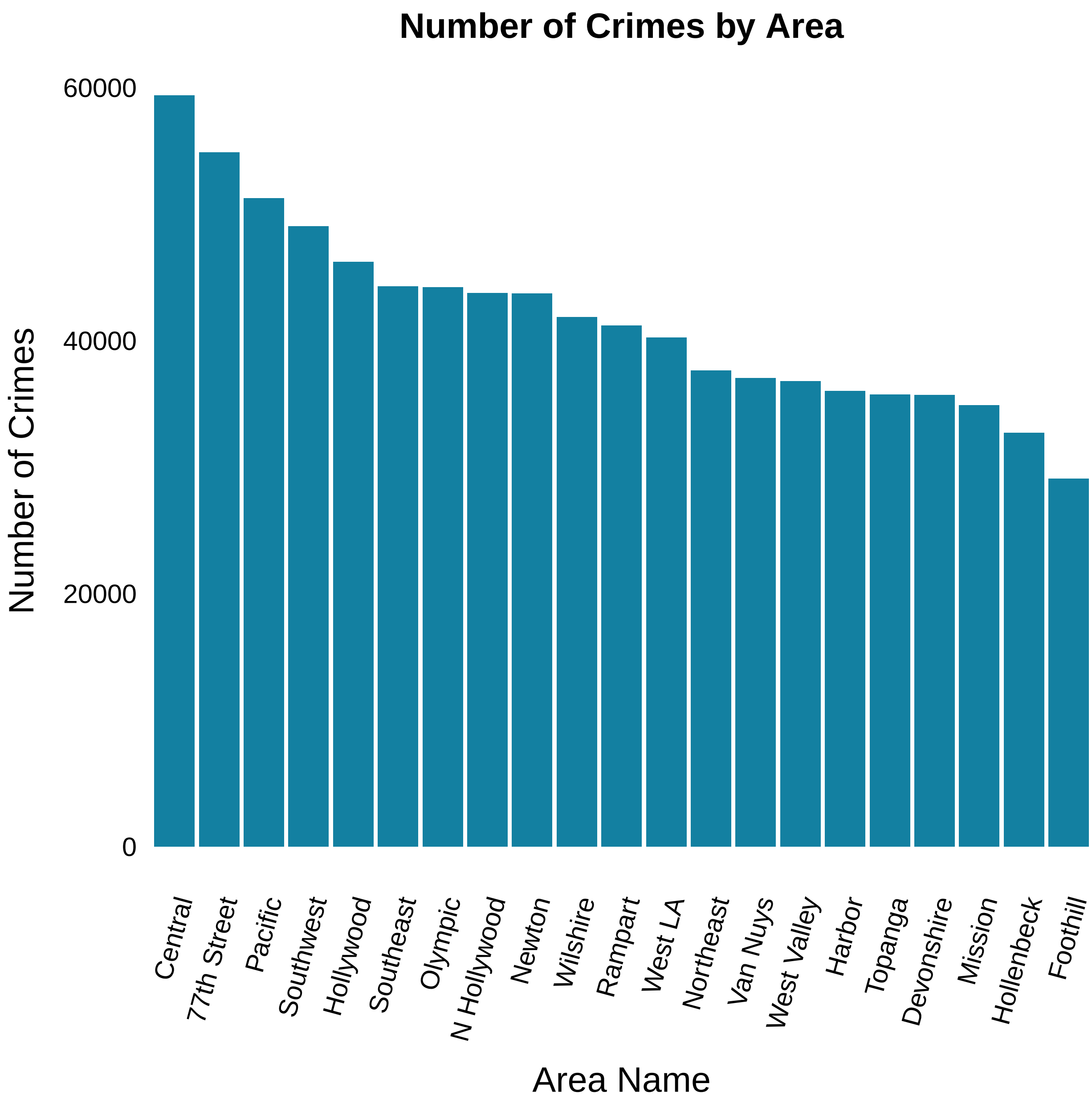

Figure 2: Number of Crimes by Area

This bar chart illustrates the distribution of crime incidents across different policing areas in Los Angeles. The Central Area consistently shows the highest crime count, followed by 77th Street.
newpage

3.5.3 Distribution of Victim Ages

```r
# Box plot of victim ages
ggplot(crime_data, aes(y = victim_age)) +
  geom_boxplot(fill = "orange") +
  labs(title = "Distribution of Victim Ages", y = "Victim Age") +
  theme_minimal() +
  theme(text = element_text(family = "sans", colour = "black"),
        plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
        axis.text.y = element_text(size = 12, colour = "black"),
        axis.title.x = element_text(size = 16, margin = margin(t = 10)),
        axis.title.y = element_text(size=16, margin = margin(t=10,r=10)),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black"))
```

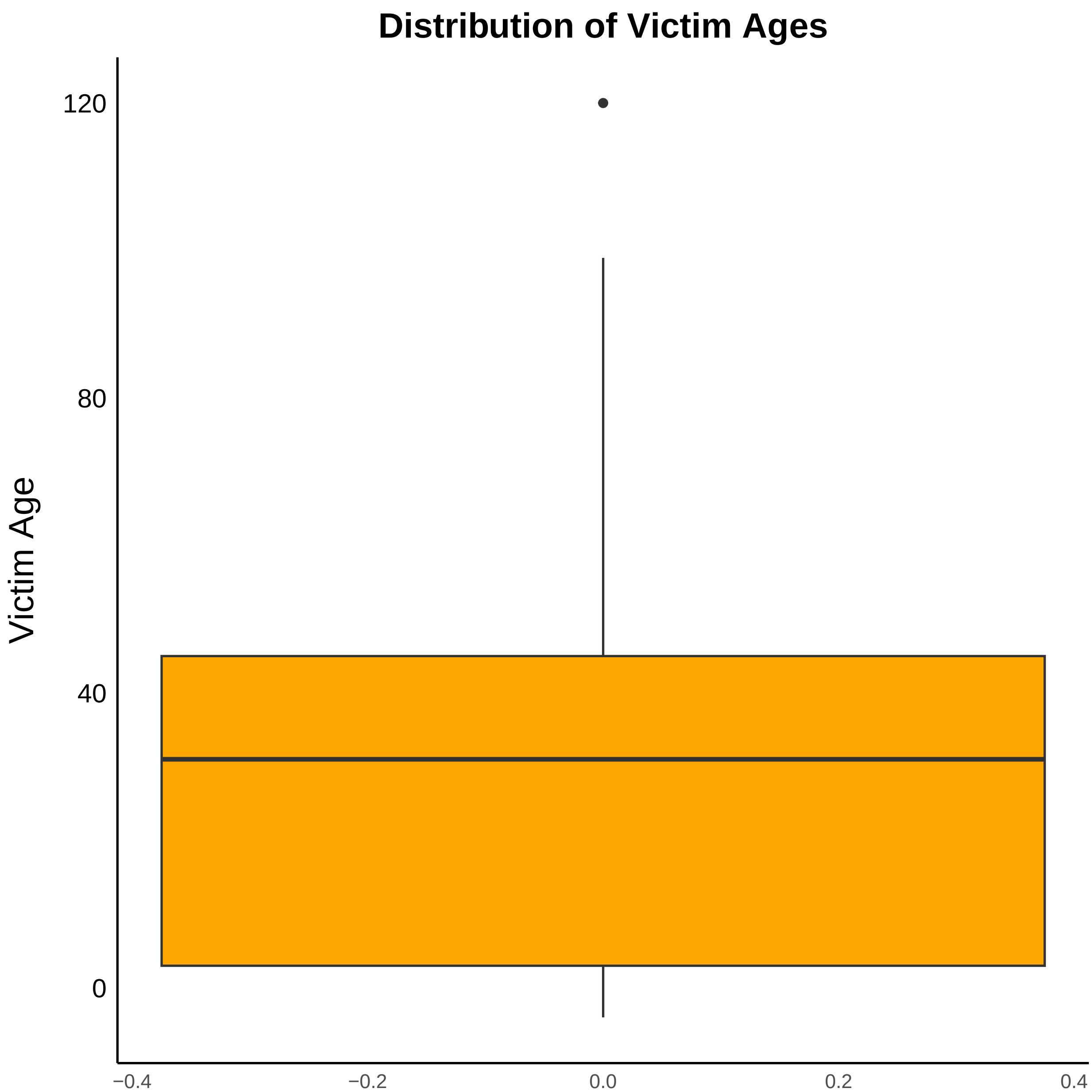

Figure 3: Distribution of Victim Ages

This boxplot showcases the distribution of victim ages across all crime types in the dataset. The majority of victims fall within the 5-45 age range, suggesting a potential focus on prevention strategies targeting this demographic.
newpage

3.5.4 Crimes by Victim Sex

```r
# Calculate the proportions for the pie chart
crime_sex_counts <- crime_data %>%
  filter(!is.na(victim_sex)) %>%
  count(victim_sex) %>%
  mutate(percentage = n / sum(n) * 100)

# Pie chart of crimes by victim sex
ggplot(crime_sex_counts, aes(x = "", y = percentage, fill = victim_sex)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  labs(title = "Crimes by Victim Sex", fill = "Victim Sex", y = "") +
  theme_void() +
  scale_fill_brewer(palette = "Set2") +
  geom_text(aes(label = paste0(round(percentage, 1), "%")),
            position = position_stack(vjust = 0.5),
            size = 5,
            color="white") +
  theme(text = element_text(family = "sans", colour = "black"),
        plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
        legend.position = "right")
```

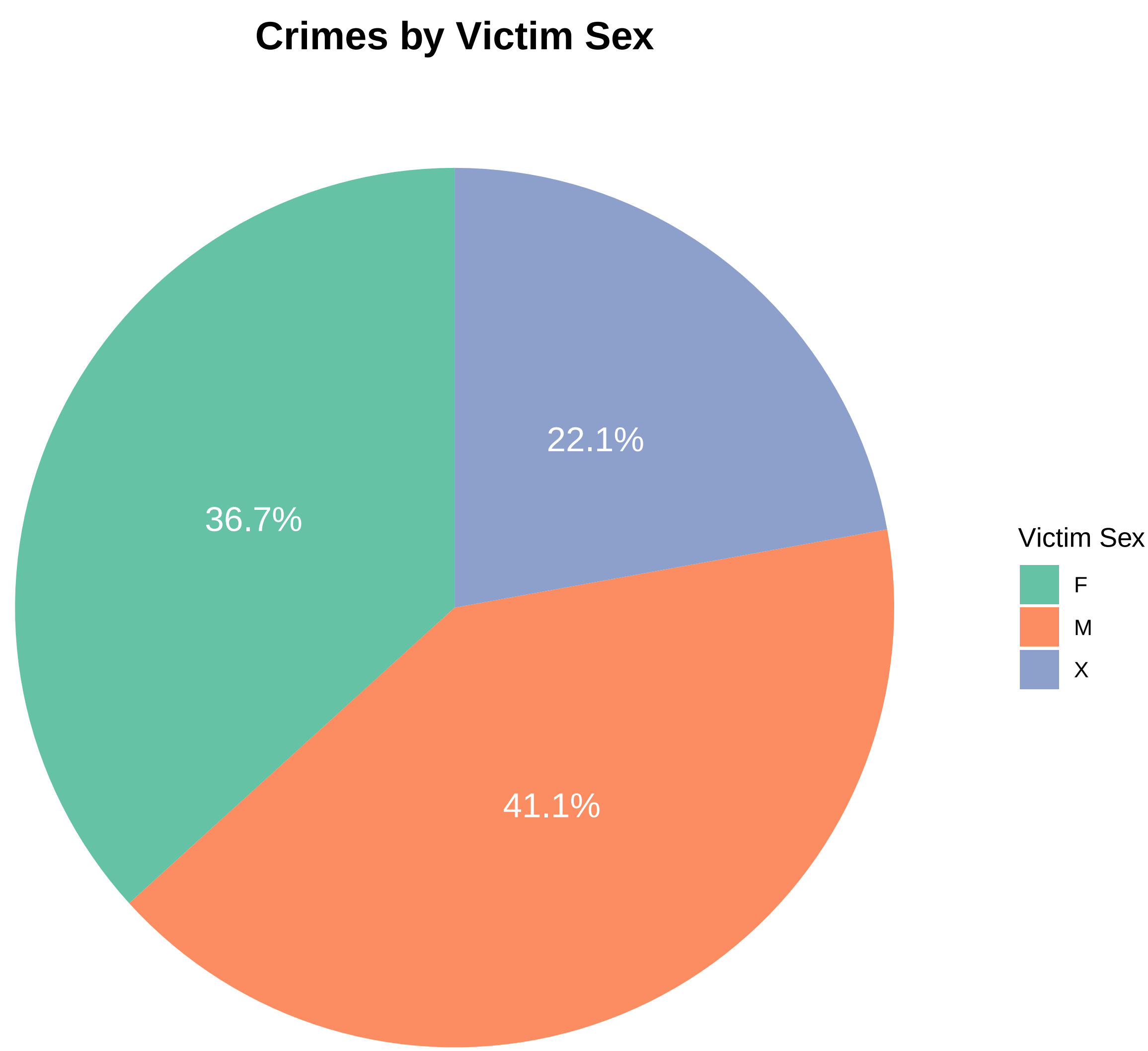

Figure 4: Crimes by Victim Sex

This pie chart reveals the distribution of crimes by victim sex. While males constitute the majority of victims, the "X" category representing unknown or unrecorded sex represents a significant portion.
newpage

3.5.5 Distribution of Victim Ages by Area

```r
# Violin plot of victim age by area
ggplot(crime_data, aes(x = area_name, y = victim_age)) +
  geom_violin(fill = "lightgreen") +
  labs(title = "Distribution of Victim Ages by Area", x = "Area Name",
       y = "Victim Age") +
  theme_minimal() +
  theme(text = element_text(family = "sans", colour = "black"),
        plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
        axis.text.x = element_text(angle = 45, hjust = 1),
        axis.text.y = element_text(size = 12, colour = "black"),
        axis.title.y = element_text(size = 14),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black"))
```

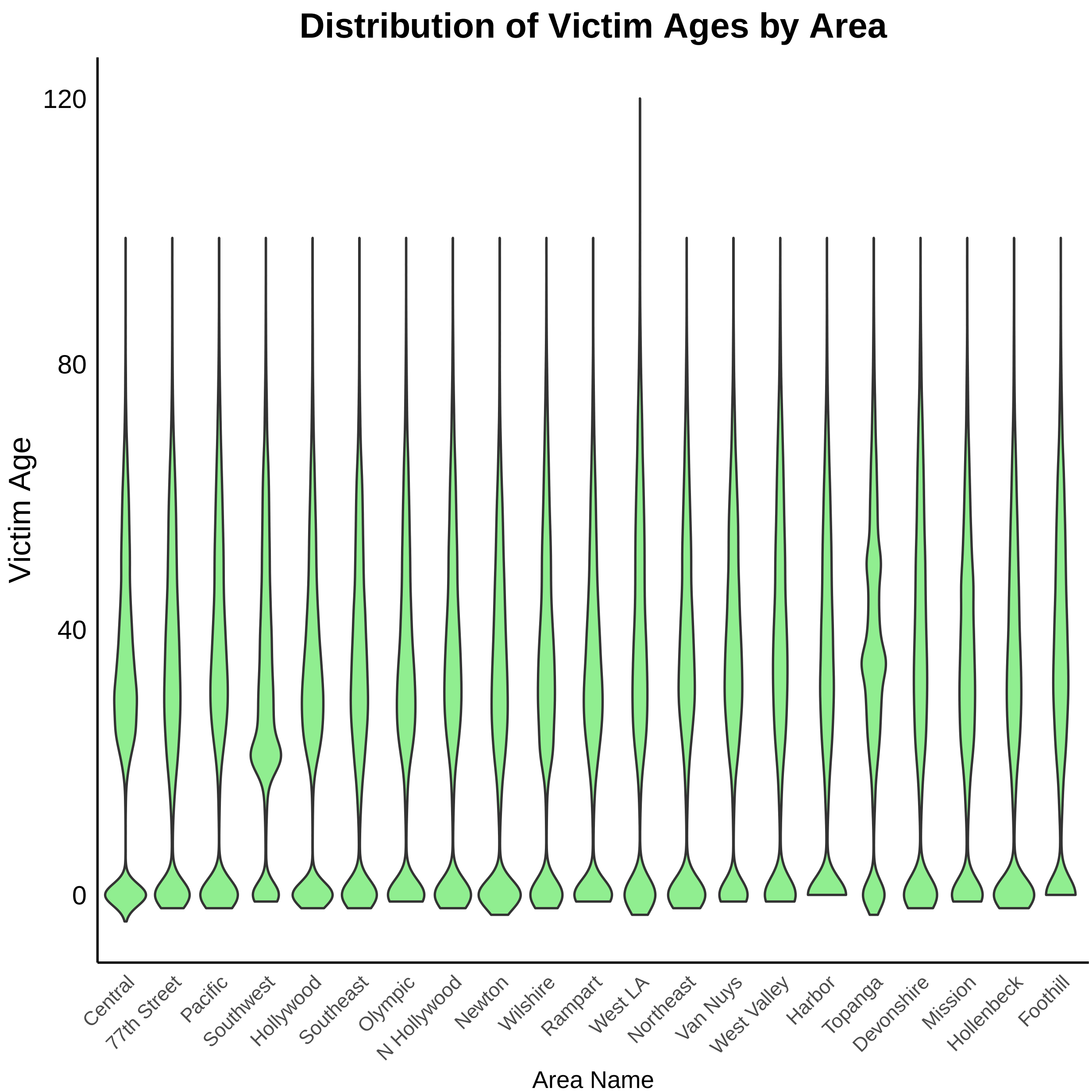

Figure 5: Distribution of Victim Ages by Area

This violin plot displays the distribution of victim ages across different areas in Los Angeles. It suggest a similar pattern crimes in almost all areas.
newpage

3.5.6 Top 10 Most Used Weapons (Excluding Hand or Fist)

```
# Filter out "Unknown" values in weapon_description
crime_data_weapon <- crime_data %>%
  filter(weapon_description != "Unknown")
```

```r
# Calculating the each weapon description and arranging in descending order
weapon_counts <- crime_data_weapon %>%
  count(weapon_description) %>%
  arrange(desc(n)) %>%
  slice(2:10)  # Selected top 10 most used weapons

# Reordering the levels of weapon_description in descending order
weapon_counts$weapon_description <- factor(weapon_counts$weapon_description,
            levels= weapon_counts$weapon_description[order(weapon_counts$n,
                                                    decreasing = TRUE)])

# Ploing the bar chart
ggplot(weapon_counts, aes(x = weapon_description, y = n)) +
  geom_bar(stat = "identity", fill = "skyblue") +
    geom_text(aes(label = paste0(n, " m")),
            vjust = 1.6,
            hjust = 0.5,
            color = "black",
            fontface = 'bold',
            size = 3.5) +
  labs(title = "Top 10 Most Used Weapons(Excluding Hand or Fist)",
       x = "Weapon Description", y = "Number of Crimes") +
  theme_minimal() +
  theme(text = element_text(family = "sans", colour = "black"),
        plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
        axis.text.x = element_text(angle = 75, hjust = 1),
        axis.text.y = element_blank(),
        axis.title.y = element_text(size = 14),
        axis.title.x = element_text(size = 14),
        axis.line.y = element_blank(),
        axis.line.x = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black"))
```

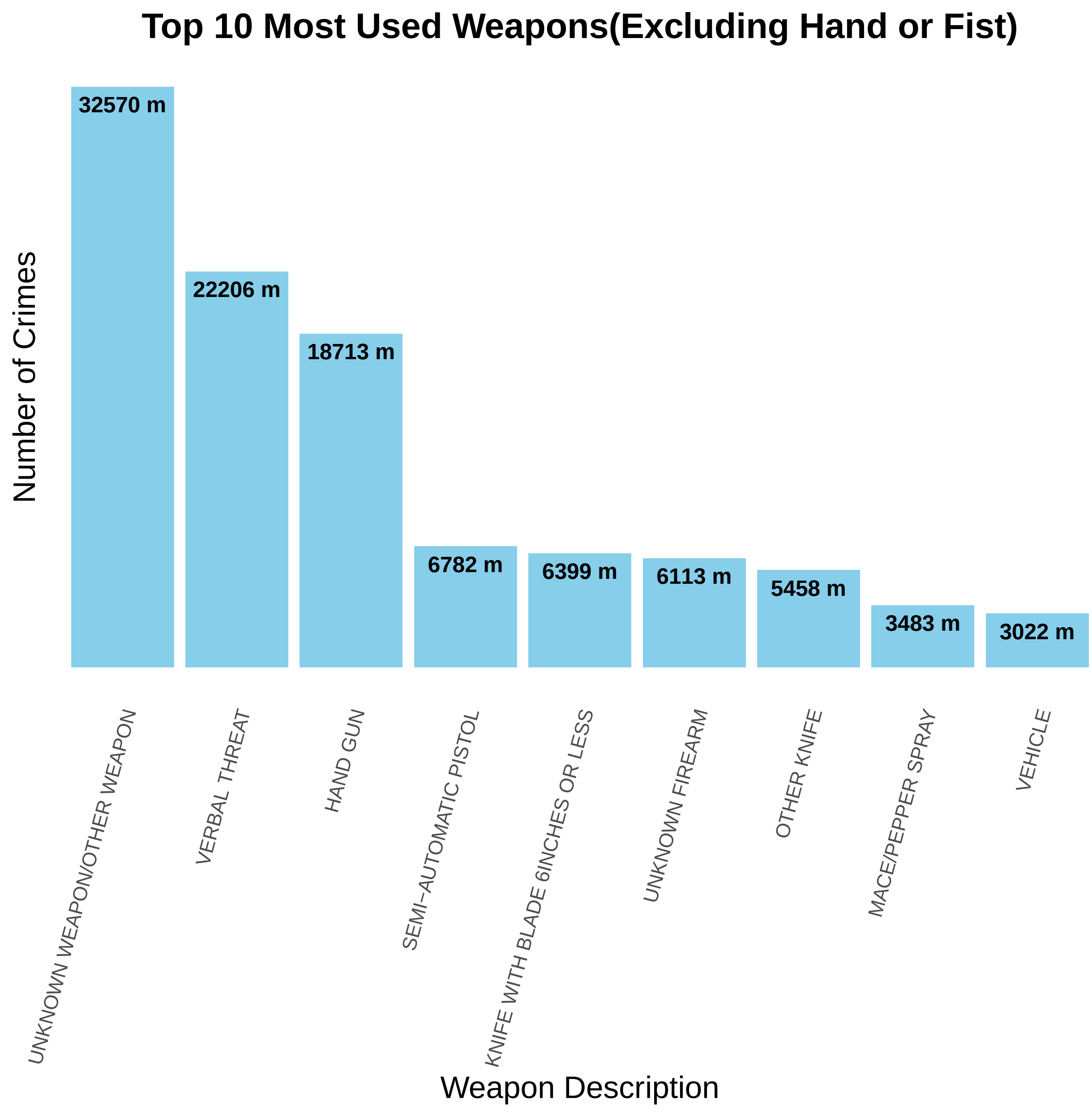

Figure 6: Top 10 Most Used Weapons (Excluding Hand or Fist)

This bar chart showcases the top 10 most frequently used weapons in Los Angeles, excluding hand or fist,which is the most used weapon. beside it, verbal threat, including handguns and semi-automatic pistols, are the most prevalent weapons, telling the pattern of easy hid able weapons used most.
newpage

3.5.7 Distribution of Case Statuses

```r
# Calculate the count of each status description
status_counts <- crime_data %>%
  count(status_description)

# Pie chart for distribution of case statuses
ggplot(status_counts, aes(x = "", y = n, fill = reorder(status_description, n))) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  labs(title = "Distribution of Case Status", fill = "Status Description", y = "") +
  scale_fill_manual(values = brewer.pal(6, "Set2")) +
  theme_void() +
  theme(text = element_text(family = "sans", colour = "black"),
        plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
        legend.position = "right")
```

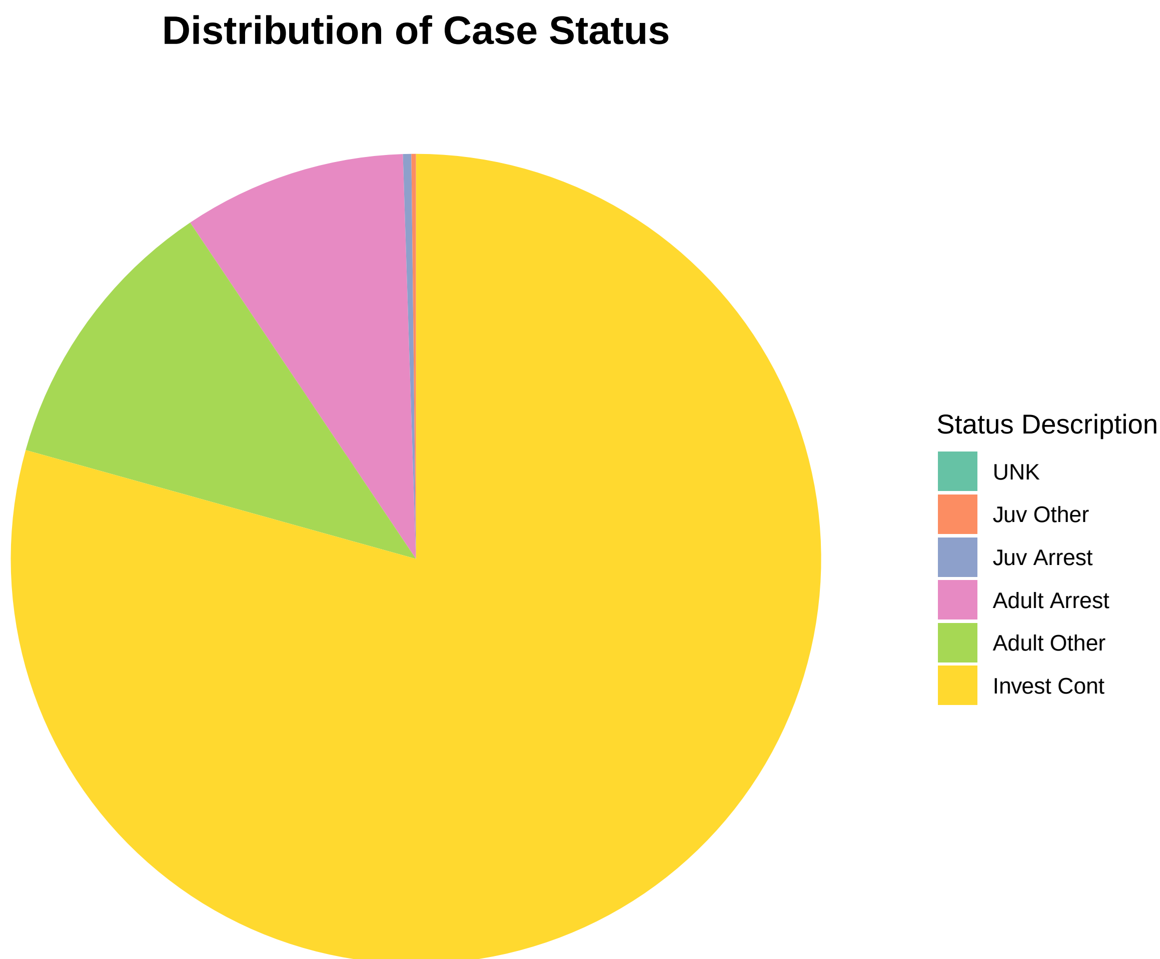

Figure 7: Distribution of Case Status

This pie chart illustrates the distribution of case statuses for reported crimes. The majority of cases are categorized as ”Invest Cont” (investigations continuing) or ”Adult Other” (adult suspects investigated but not arrested), suggesting that many cases are still in progress.

3.6 Non-Trivial Questions Analysis (Identifying Patterns and Trends)

This section will address the non-trivial research questions using visualization techniques, highlighting key observations and insights gleaned from the data.

3.6.1 Temporal and Spatial Trends in High-Crime Areas

We used time-series plots and maps to visualize the trends of the top two most common crimes in Los Angeles' highest-crime areas across the years 2020 to 2023.

```r
# Filter data for the specified areas
selected_areas <- head(crime_counts$area_name, 5)

# Prepare the data
clustered_data1 <- crime_data %>%
  filter(area_name %in% selected_areas)

crime_summary <- clustered_data1 %>%
  group_by(area_name, year, crime_description) %>%
  summarise(crime_count = n()) %>%
  arrange(desc(crime_count))

## `summarise()` has grouped output by 'area_name', 'year'.  You can override
using
## the `.groups` argument.

top_crimes <- crime_summary %>%
  group_by(crime_description) %>%
  summarise(total_crime_count = sum(crime_count)) %>%
  top_n(2, total_crime_count) %>%
  pull(crime_description)

crime_summary_top <- crime_summary %>%
  filter(crime_description %in% top_crimes)

# Heatmap
ggplot(crime_summary_top, aes(x = year, y = reorder(area_name, crime_count), fill = crim
  geom_tile(color = "white", size = 0.2) +
  facet_wrap(~ crime_description, scales = "free_y", ncol = 1) +
  scale_fill_viridis(option = "magma", direction = -1) +
  labs(
    title = "Crime Concentration in Areas along Time",
    x = "Years",
    y = "Area",
    fill = "Crime Count"
  ) +
```

```
theme_minimal() +
theme(
  text = element_text(family = "sans", color = "black"),
  plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
  axis.text.x = element_text(angle = 0, vjust = 0.5, size = 13, margin = margin(t = 10
  axis.text.y = element_text(size = 11),
  axis.title.x = element_text(size = 15, margin = margin(t = 20)),
  axis.title.y = element_text(size = 15),
  panel.grid.minor = element_blank(),
  panel.background = element_blank(),
  axis.line = element_blank(),
  strip.text = element_text(size = 12, face = "bold"),
  legend.title = element_text(size = 13),
  legend.text = element_text(size = 11)
)
```

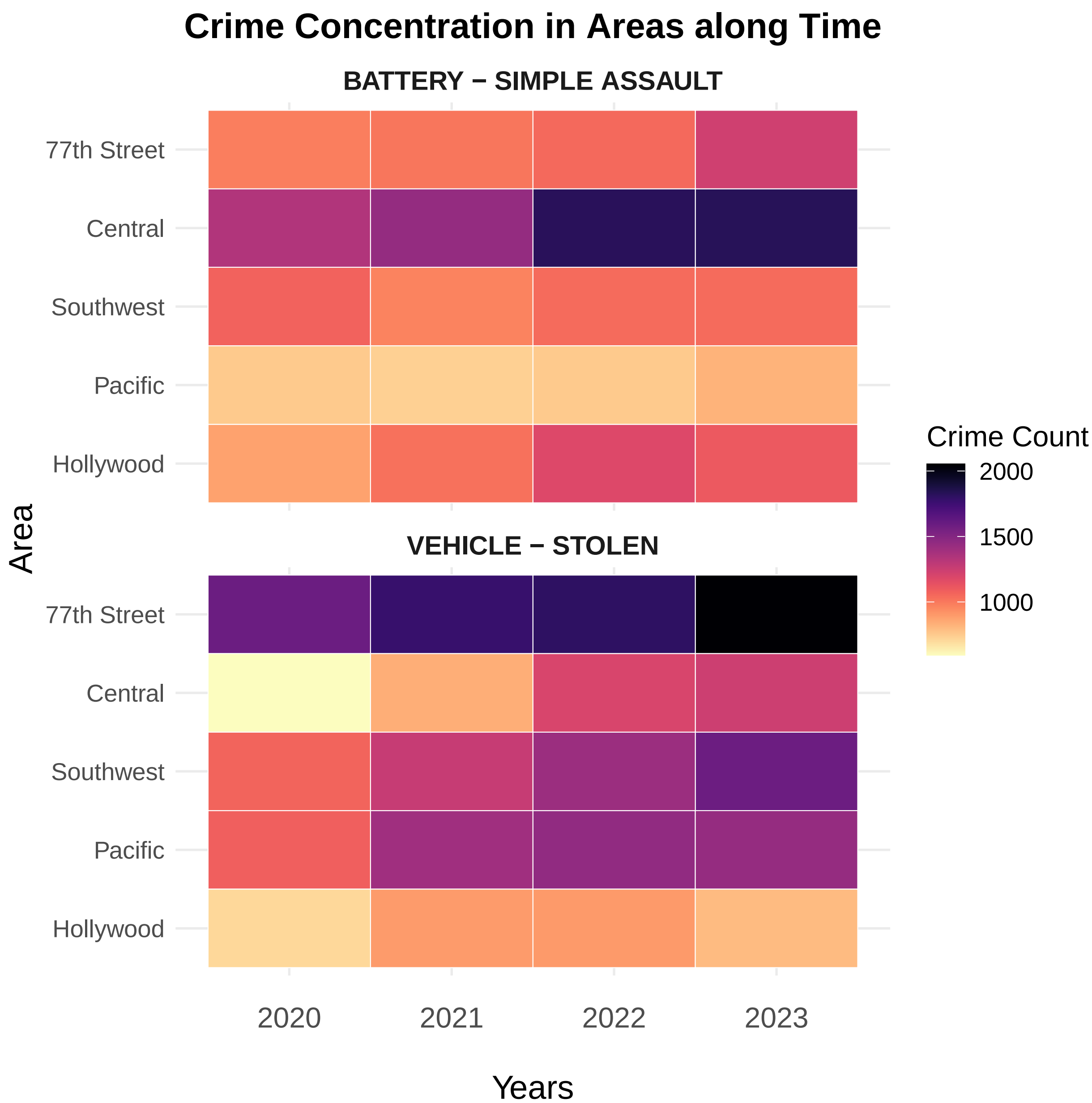

Figure 8: Spatial Trends in High-Crime Areas

This heatmap visualizes the frequency of the top two most common crimes in selected high-crime areas across the years 2020-2023. The colors represent the crime count, highlighting areas with a higher concentration of incidents.

The Central area shows a significant increase in simple assault and battery cases in 2022, continuing into 2023. This suggests a change in crime patterns or contributing factors during that period. In contrast, the 77th Street area consistently has a high number of vehicle thefts each year, peaking in 2023. This indicates an ongoing issue with vehicle theft in that area.

```r
# Line plot
ggplot(crime_summary_top, aes(x = year, y = crime_count, color = area_name)) +
  geom_line(size = 1) +
  facet_wrap(~ crime_description, scales = "free_y", ncol = 1) +  # Adjust ncol as neede
  labs(
    title = "Crime Trends in Areas along Time",
    x = "Years",
    y = "Number of Crimes",
    color = "Area"
  ) +
  theme_minimal() +
  theme(
    text = element_text(family = "sans", color = "black"),
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
    axis.text.x = element_text(angle = 0, vjust = 0.5, size = 11),
    axis.text.y = element_text(size = 11),
    axis.title.x = element_text(size = 13, margin = margin(b = 10)),
    axis.title.y = element_text(size = 13, margin = margin(t = 10)),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    panel.background = element_blank(),
    axis.line = element_line(color = "black"),
    strip.text = element_text(size = 11, face = "bold"),
    legend.position = "none"
    ) +
      geom_text(
        data = crime_summary_top %>% filter(year == 2023),
        aes(label = area_name, x = year, y = crime_count, color = area_name),
        hjust = 1.1, vjust = -0.2, size = 3
      )
```

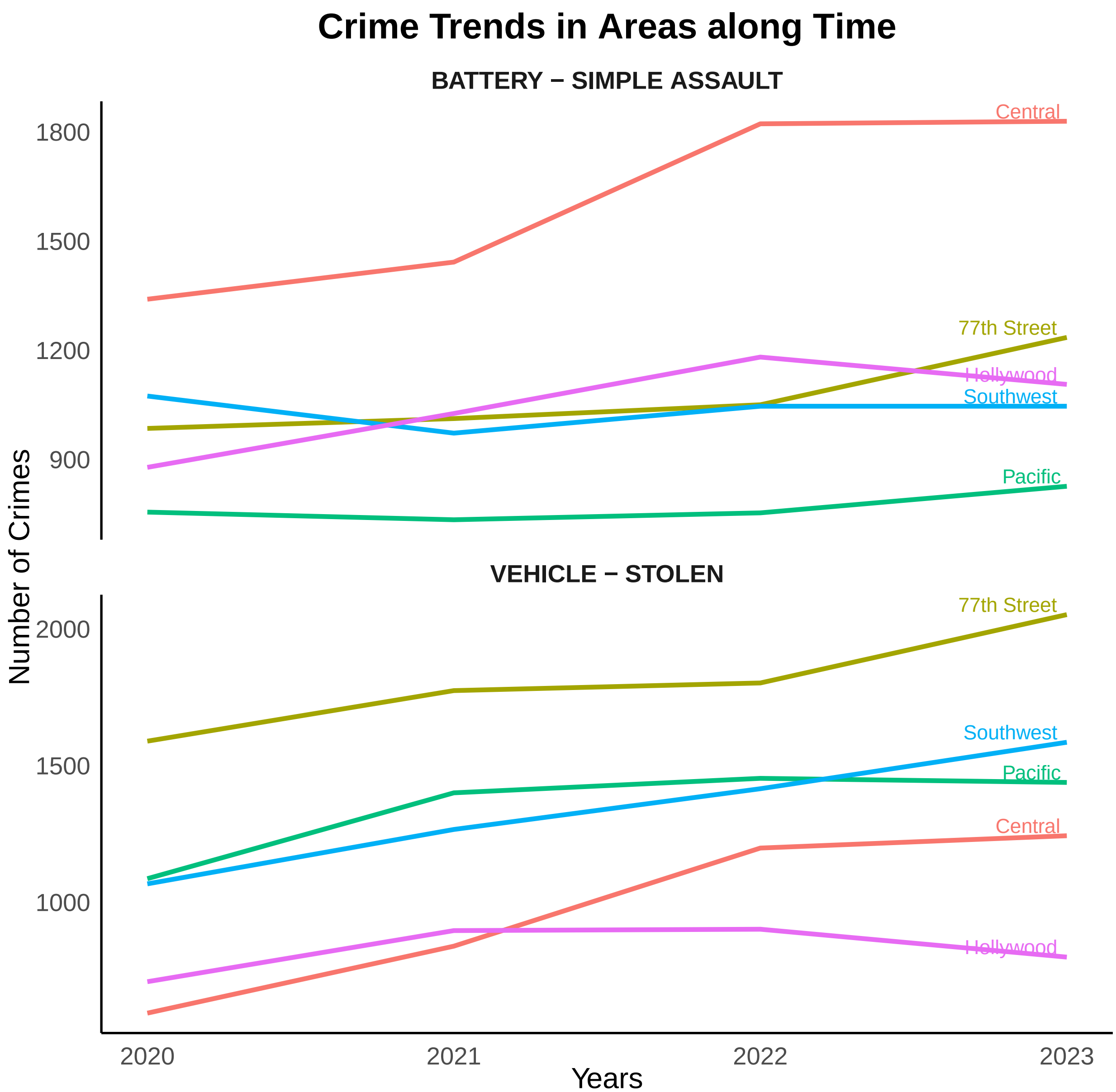

Figure 9: Temporal Trends in High-Crime Areas

This line graph depicts the trends of the top two crimes within selected high-crime areas over the years 2020-2023. The varying trends for different crimes and areas highlight the need for targeted and location-specific interventions.

The line plots reveal that the Central area saw a sharp rise in simple assault and battery cases from 2020 to 2023, reinforcing the heatmap findings and highlighting the seriousness of the issue. Meanwhile, the 77th Street area displays a steady increase in vehicle thefts, confirming the consistently high numbers seen in the heatmap. This suggests a need for targeted preventative measures in the area.

3.6.2 Crime Disparities by Demographic and Time

We used stacked bar charts to compare the frequency of crimes across demographic groups during nighttime hours and daytime hours. similarly, treemap to look for dominant crime type during day and night in different areas of Los angels.

```
# Define nighttime hours (e.g., 10 PM to 6 AM)
top_crimes_combined <- crime_data %>%
  group_by(area_name, crime_description, part_of_day) %>%
  summarise(crime_count = n()) %>%
  arrange(area_name, desc(crime_count)) %>%
  group_by(area_name, part_of_day) %>%
  top_n(1, crime_count) %>%
  ungroup()

# Treemap Visualization (Facets for Day & Night)
treemap(
  top_crimes_combined,
  index = c("area_name", "part_of_day", "crime_description"),
  vSize = "crime_count",
  title = "Trend of Top Crimes in Areas by Day/Night",
  palette = "Set3",
  fontsize.labels = c(8, 8, 7),  # Increase font size for all levels
  fontfamily.labels = "sans",
  fontface.labels = c(2, 1, 1),  # Bold area names (level 1)
  bg.labels = "transparent",
  align.labels = list(
    c("left", "top"),
    c("center", "bottom"),
    c("left", "center")
  ),
  border.col = c("white", "grey", "grey"),
  border.lwds = c(1, 0.5, 0.5),
  overlap.labels = 0.5,
)
```

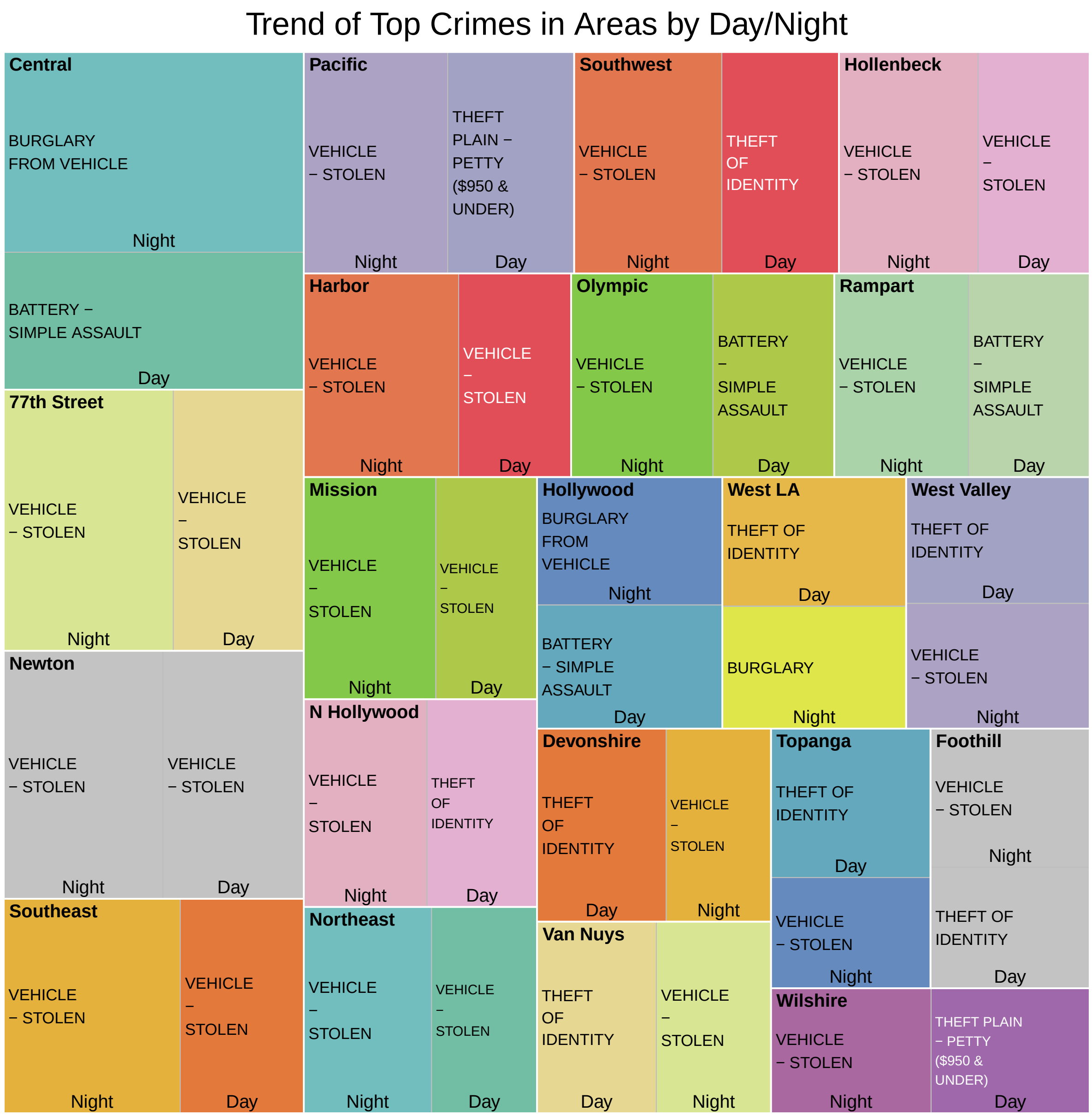

Figure 10: Crime Disparities by Demographic and Time

The treemap shows the most common crime types in different areas during the day and night. Vehicle-related crimes are frequent in many areas at both times. Certain areas, like Rampart and olympic, have more battery and simple assault crimes, mainly at day. Areas like Hollywood and West LA have more burglaries and thefts from vehicles at night, indicating patterns in criminal activity.

```
# Stacked Bar Plot (Day & Night)
victim_demographics <- crime_data %>%
  mutate(age_group = case_when(
```

```r
    victim_age < 18 ~ "<18",
    victim_age >= 18 & victim_age < 30 ~ "18-29",
    victim_age >= 30 & victim_age < 40 ~ "30-39",
    victim_age >= 40 & victim_age < 50 ~ "40-49",
    victim_age >= 50 & victim_age < 60 ~ "50-59",
    victim_age >= 60 ~ "60+"
  )) %>%
  group_by(age_group, victim_sex, part_of_day) %>%
  summarise(victim_count = n()) %>%
  arrange(part_of_day, age_group, desc(victim_count))

ggplot(victim_demographics, aes(x = age_group, y = victim_count,
                                fill = victim_sex)) +
  geom_bar(stat = "identity", color = "black", width = 0.7, position = "stack") +
  facet_wrap(~ part_of_day) +
  labs(
    title = "Victim Demographics by Age and Sex (Day & Night)",
    x = "Age Group",
    y = "Count",
    fill = "Gender"
  ) +
  scale_fill_manual(
    values = viridis_pal(option = "D")(3),
    labels = c("F" = "Female", "M" = "Male", "X" = "Unknown")
  ) +
  theme_minimal() +
  theme(
    text = element_text(family = "sans", color = "black"),
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
    axis.text = element_text(size = 12),
    axis.title = element_text(size = 14, margin = margin(t = 20)),
    legend.title = element_text(size = 13),
    legend.text = element_text(size = 12),
    legend.position = "bottom",
    panel.grid.major = element_blank()
  )
```

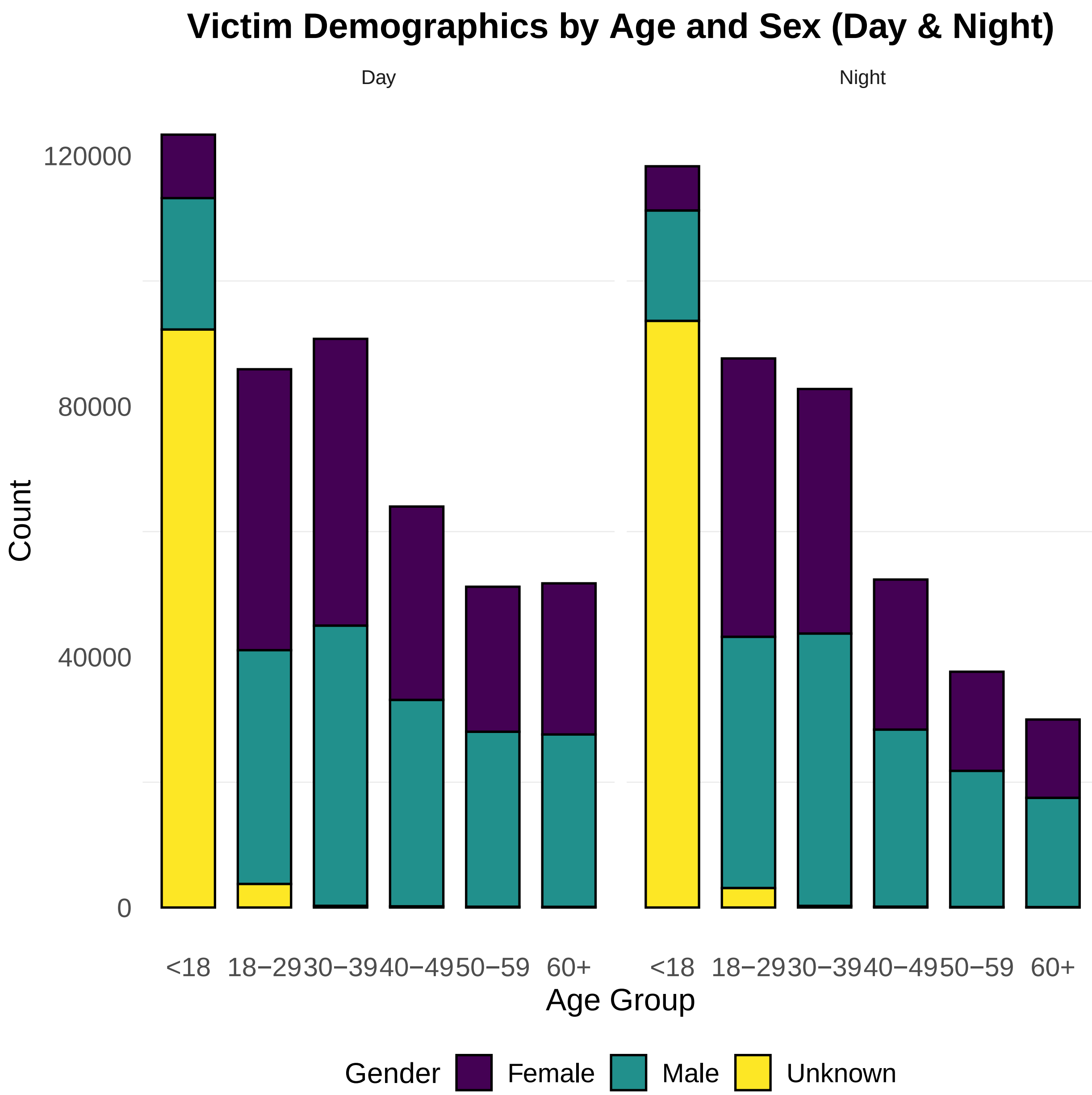

Figure 11: Crime Disparities by Demographic and Time

The stacked bar chart shows that most victims are female, with a significant number of unknown gender. The highest number of victims is in the 18-39 age group. There is a slight difference in the number of victims by gender and age group between day and night. For example, there are slightly more victims in the 18-29 age group during the night, while there are slightly more victims in the 30-39 age group at day.

3.6.3 Consistency and Variations in Crime Patterns

We created time-series plots to visualize crime patterns across different time scales (monthly, seasonal, daily).

3.6.4 Consistency and Variations in Crime Patterns

We created time-series plots to visualize crime patterns across different time scales (monthly, seasonal, daily and hourly).

```
crime_counts_month <- crime_data %>%
  group_by(year, month) %>%
  summarize(Total_Crimes = n(), .groups = 'drop')

# Plot the monthly crime trend
ggplot(crime_counts_month, aes(x = month, y = Total_Crimes,
                               color = factor(year), group = year)) +
  geom_line(size = 1) +
  labs(
    title = "Crime Trends Over Time (Monthly)",
    x = "Month",
    y = "Number of Crimes"
  ) +
  scale_x_discrete(labels = month.abb) +
  theme_minimal() +
  theme(
    text = element_text(family = "sans", color = "black"),
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
    axis.text = element_text(size = 12),
    axis.title = element_text(size = 14),
    panel.grid.major.x = element_blank(),
    legend.position = "none"
  ) +
  geom_text(
    data = crime_counts_month %>% filter(month == "Dec"),
    aes(label = year, x = month, y = Total_Crimes, color = factor(year)),
    hjust = 0.1, vjust = -0.9, size = 4
  )
```

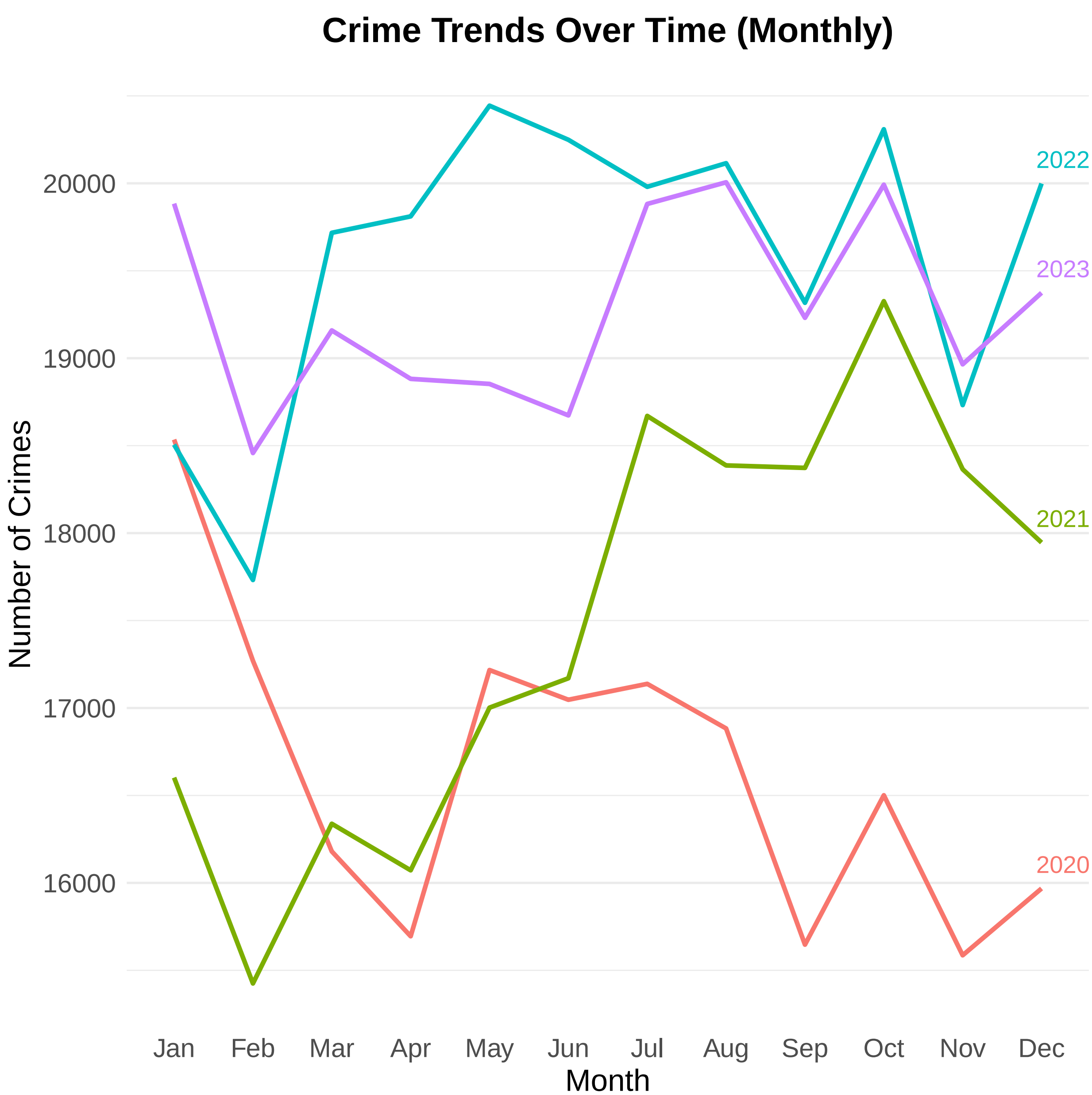

Figure 12: Crime Trends Over Time (Monthly)

This line graph depicts the monthly crime trends from 2020 to 2023. An interesting insight is that crime rates significantly increase from September to October and then drop in November, with this pattern repeating every year. There are noticeable peaks and dips in crime rates throughout the year, suggesting that factors like weather, holidays, and school schedules could play a role. Each year shows a unique monthly crime pattern, hinting at influences such as economic changes or policy updates.

```
# Aggregate the crime counts by season and year
crime_counts_season <- crime_data %>%
```

```
  group_by(year, season) %>%
  summarize(Total_Crimes = n(), .groups = 'drop')

# Plot the seasonal crime trend
ggplot(crime_counts_season, aes(x = year, y = Total_Crimes, color = season, group = seas
  geom_line(size = 1) +
  geom_point(size = 1) +
  labs(
    title = "Seasonal Crime Trends Over Time",
    x = "Year",
    y = "Total Crimes",
    color = "Season"
  ) +
  theme_minimal() +
  theme(
    text = element_text(family = "sans", color = "black"),
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
    axis.text = element_text(size = 12),
    axis.title = element_text(size = 14),
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.title = element_text(size = 13),
    legend.text = element_text(size = 11),
    panel.grid.major.x = element_blank(),
    panel.grid.minor = element_blank(),
    legend.position = "none"
  ) +
  geom_text(
    data = crime_counts_season %>% filter(year == 2021 & season %in%
                                          c("Fall", "Summer")),
    aes(label = season, x = year, y = Total_Crimes, color = season),
    hjust = 0, vjust = -0.5, size = 4
  ) +
  geom_text(
    data = crime_counts_season %>% filter(year == 2020 & season %in%
                                          c("Winter", "Spring")),
    aes(label = season, x = year, y = Total_Crimes, color = season),
    hjust = 0, vjust = -0.5, size = 4
  )
```

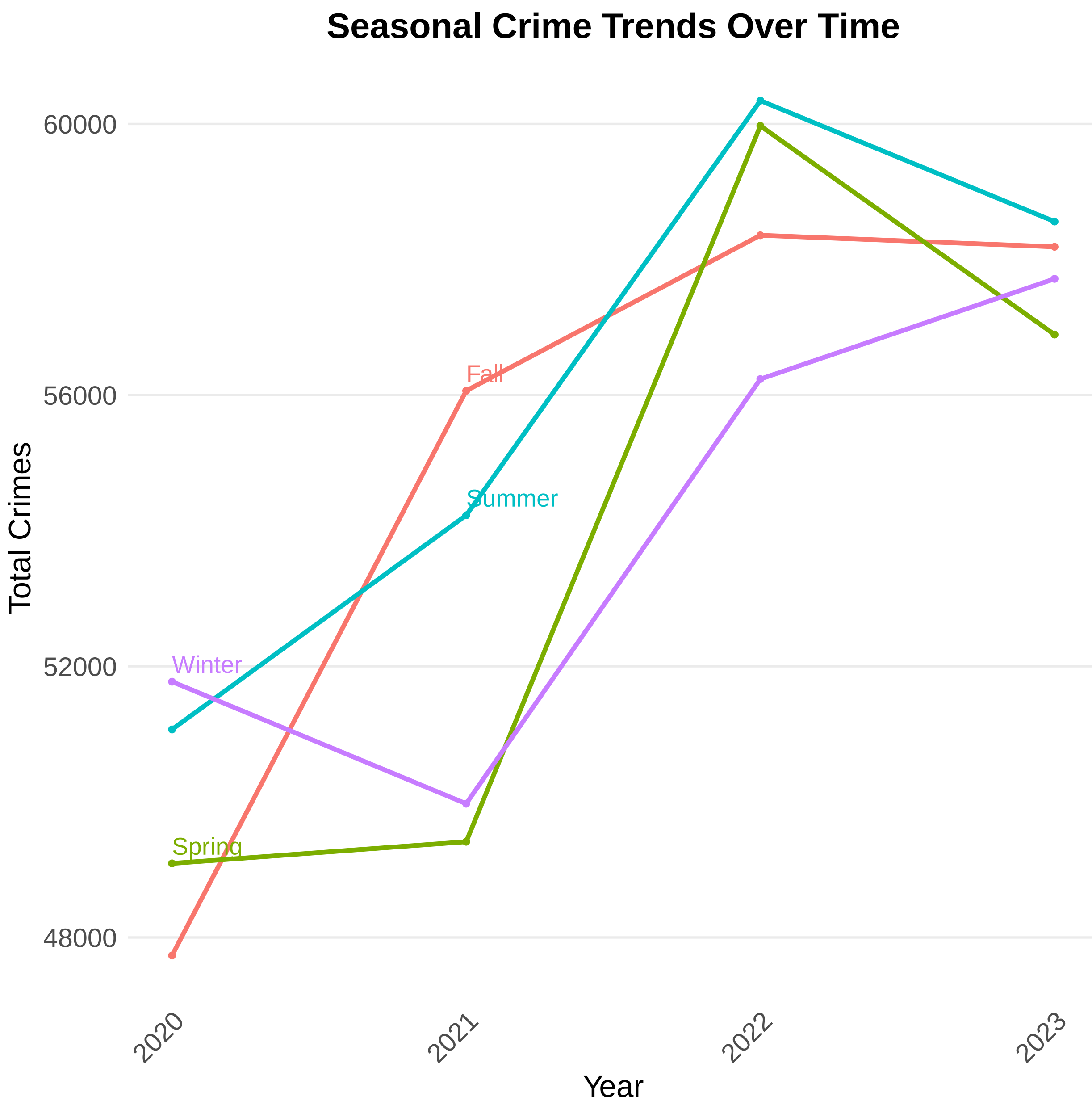

Figure 13: Seasonal Crime Trends Over Time

This line graph shows the seasonal crime trends from 2020 to 2023.Crime rates peak during summer months, likely due to increased outdoor activities, heat-related tensions, and longer daylight hours. There is also a noticeable rise in crime during spring, possibly linked to warmer weather and increased outdoor activity.

```
# Aggregate crime counts by year and weekday
crime_counts_weekday <- crime_data %>%
  group_by(year, weekday) %>%
  summarize(Crime_Count = n(), .groups = 'drop')
```

```r
# Plot crime distribution by day of week divided by year
ggplot(crime_counts_weekday, aes(x = weekday, y = Crime_Count)) +
  geom_col(fill = "#3182bd", color = "black", size = 0.5, width = 0.7) +
  facet_wrap(~ year, ncol = 2, scales = "free_y") +
  labs(
    title = "Crime Trends Over Time(Weekdays)",
    x = "Day of Week",
    y = "Number of Crimes"
  ) +
  theme_minimal() +
  theme(
    text = element_text(family = "sans", color = "black"),
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
    axis.text = element_text(size = 12),
    axis.text.x = element_text(angle = 45, hjust = 1),
    axis.title = element_text(size = 14),
    strip.text = element_text(size = 12, face = "bold"),
    panel.grid.major.x = element_blank(),
    panel.spacing = unit(1.2, "lines") # Add spacing between facets
  )
```

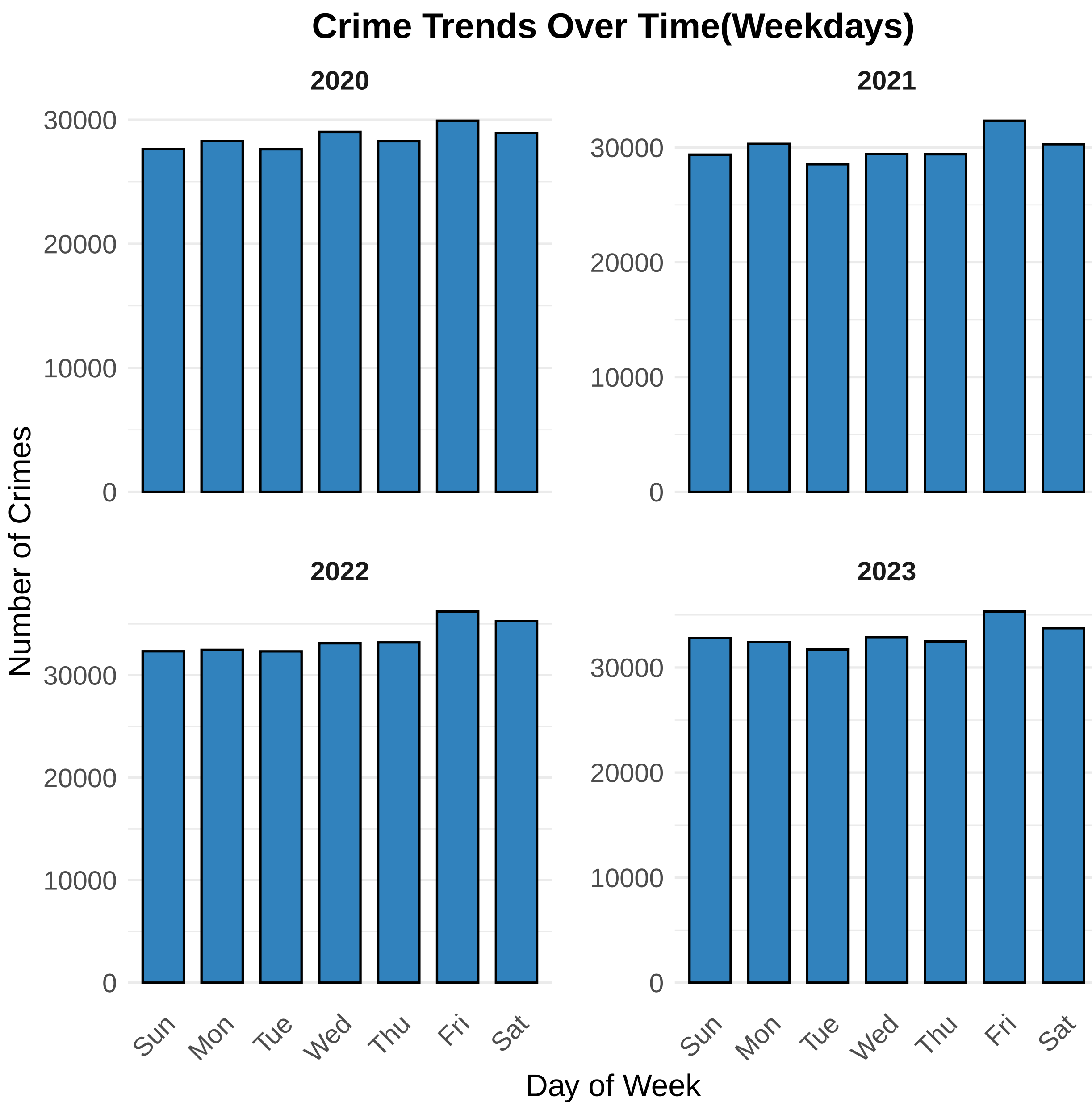

Figure 14: Crime Trends Over Time(Weekdays)

This bar chart showcases the distribution of crime by day of the week for the years 2020 to 2023. Crime rates remain relatively consistent across weekdays, indicating that routine activities and public spaces may play major roles. However, there is a slight increase in crime on weekends, suggesting increased social activity or reduced security presence.

```
# Aggregate crime counts by hour and year
crime_counts_hour <- crime_data %>%
  group_by(year, hour_occurred) %>%
  summarize(Crime_Count = n(), .groups = 'drop')
```

```r
# Plot crime distribution by hour divided by year
ggplot(crime_counts_hour, aes(x = hour_occurred, y = Crime_Count)) +
  geom_col(fill = "#3182bd", color = "black", size = 0.5, width = 0.7) +
  facet_wrap(~ year, ncol = 2, scales = "free_y") +
  labs(
    title = "Crime Distribution by Hour of Day",
    x = "Hour of Day",
    y = "Number of Crimes"
  ) +
  scale_x_continuous(breaks = seq(0, 23, by = 3)) +
  theme_minimal() +
  theme(
    text = element_text(family = "sans", color = "black"),
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
    axis.text = element_text(size = 12),
    axis.text.x = element_text(angle = 0, hjust = 1, vjust = 1),
    axis.title = element_text(size = 14),
    strip.text = element_text(size = 12, face = "bold"),
    panel.grid = element_blank(),
    panel.spacing = unit(1.2, "lines")
  )
```

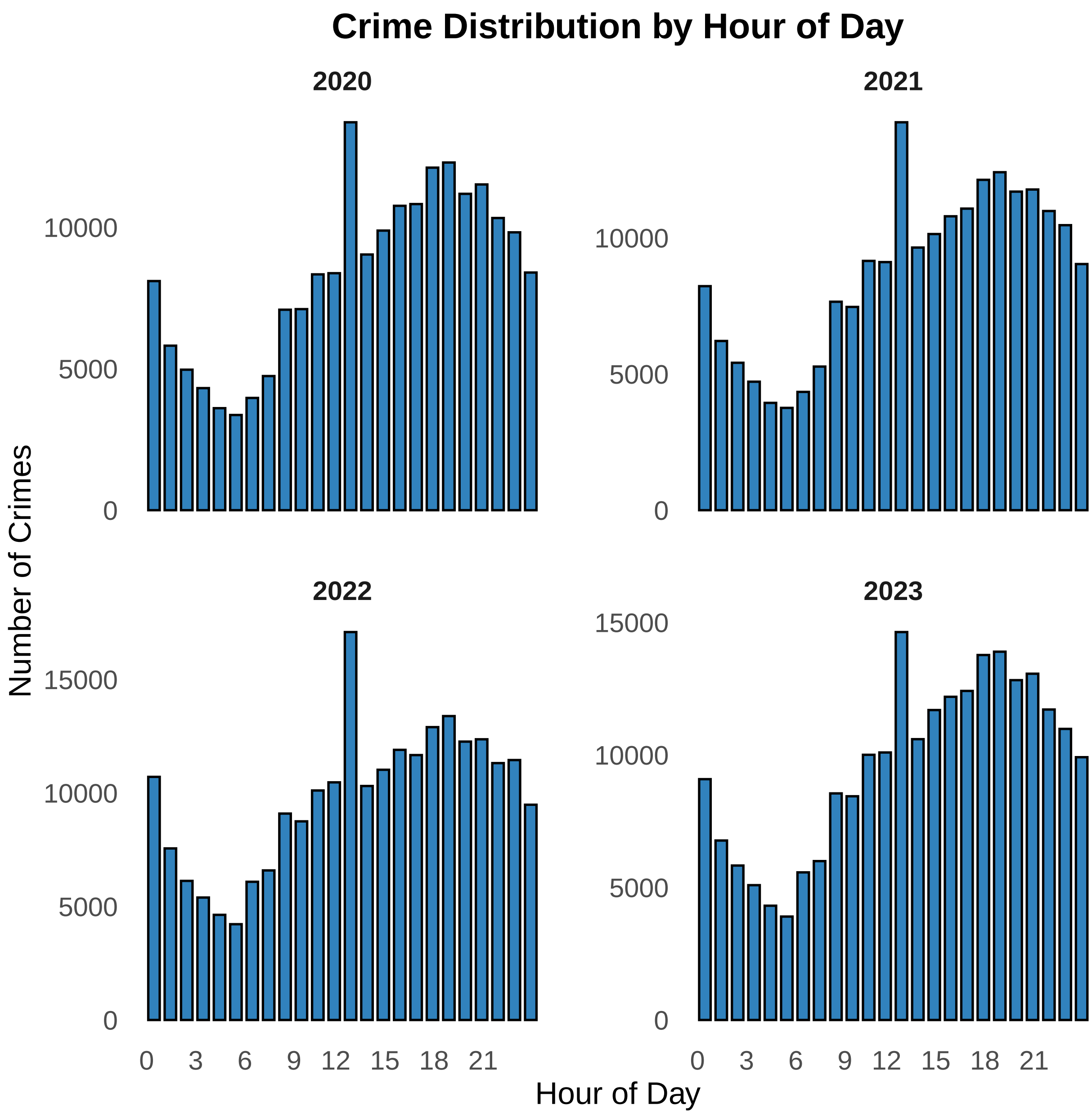

Figure 15: Crime Distribution by Hour of Day

Crime rates consistently peak around 7pm, with another peak observed around 1am, showing a connection with evening and nighttime activities. There are subtle shifts in peak hours and overall distribution across years, possibly reflecting changes in lifestyle or security measures.

3.6.5 Co-occurrence of Crime Types

We used heatmaps to visualize the co-occurrence patterns of crime types at the same locations.This heatmap showcases the co-occurrence patterns of the top five most common crimes in Los Angeles. To identify these patterns, the **Apriori algorithm** was employed. This algorithm is a popular method for association rule mining, which aims to discover relationships between items in a dataset. In this case, the items are crime types, and the algorithm identifies rules that indicate the likelihood of one crime occurring given the presence of another. The heatmap visualizes these relationships as rules, where each rule is formatted as "If Antecedent, then Consequent." Antecedent: This refers to the crime type that is considered the "if" part of the rule. It represents the crime that is already known to have occurred. Consequent: This refers to the crime type that is considered the "then" part of the rule. It represents the crime that is more likely to occur given that the antecedent crime has already happened.

```
# Filter and prepare the data
crime_data_filtered <- crime_data %>%
  select(date_occurred, location, crime_description) %>%
  mutate(date = as.Date(date_occurred)) %>%
  group_by(date, location) %>%
  summarize(crime_types = paste(unique(crime_description), collapse = ","),
            .groups = 'drop')

# Convert the crime types into a list of transactions
transactions_list <- strsplit(crime_data_filtered$crime_types, ",")

# Create a transactions object
transactions <- as(transactions_list, "transactions")

# Apply the Apriori algorithm
rules <- apriori(transactions, parameter = list(supp = 0.001, conf = 0.8))

## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.8    0.1    1 none FALSE            TRUE       5   0.001      1
##  maxlen target  ext
##      10  rules TRUE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 795
##
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[160 item(s), 795281 transaction(s)] done [0.09s].
## sorting and recoding items ... [74 item(s)] done [0.01s].
## creating transaction tree ... done [0.27s].
## checking subsets of size 1 2 3 4 5 done [0.00s].
```

```
## writing ... [149 rule(s)] done [0.00s].
## creating S4 object  ... done [0.04s].
```

```
print(head(transactions))
```

```
## transactions in sparse format with
##  6 transactions (rows) and
##  160 items (columns)
```

```
top_crimes <- sort(table(unlist(transactions_list)), decreasing = TRUE)[1:5]

print(top_crimes)
```

```
##
##         VEHICLE - STOLEN BATTERY - SIMPLE ASSAULT                 BURGLARY
##                    91920                    62027                    55585
##        THEFT OF IDENTITY    BURGLARY FROM VEHICLE
##                    53842                    53376
```

```
# Filter rules to include only the top 5 crimes

rules_top <- subset(rules, lhs %in% names(top_crimes) | rhs %in% names(top_crimes))

rules_df <- data.frame(
  Antecedent = labels(lhs(rules_top)),
  Consequent = labels(rhs(rules_top)),
  Support = rules_top@quality$support,
  Confidence = rules_top@quality$confidence,
  Lift = rules_top@quality$lift
)
# Plot heatmap

ggplot(rules_df, aes(x = Antecedent, y = Consequent, fill = Lift)) +
  geom_tile(color = "white") +
  scale_fill_gradient(low = "white", high = "steelblue") +
  labs(title = "Co-occurrence of Crimes(Top 5))",
       x = "If This Crime Occurs (Antecedent)",
       y = "Then This Crime Is Likely (Consequent)",
       fill = "Lift (Strength of Association)") +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    text = element_text(family = "sans")
  )
```

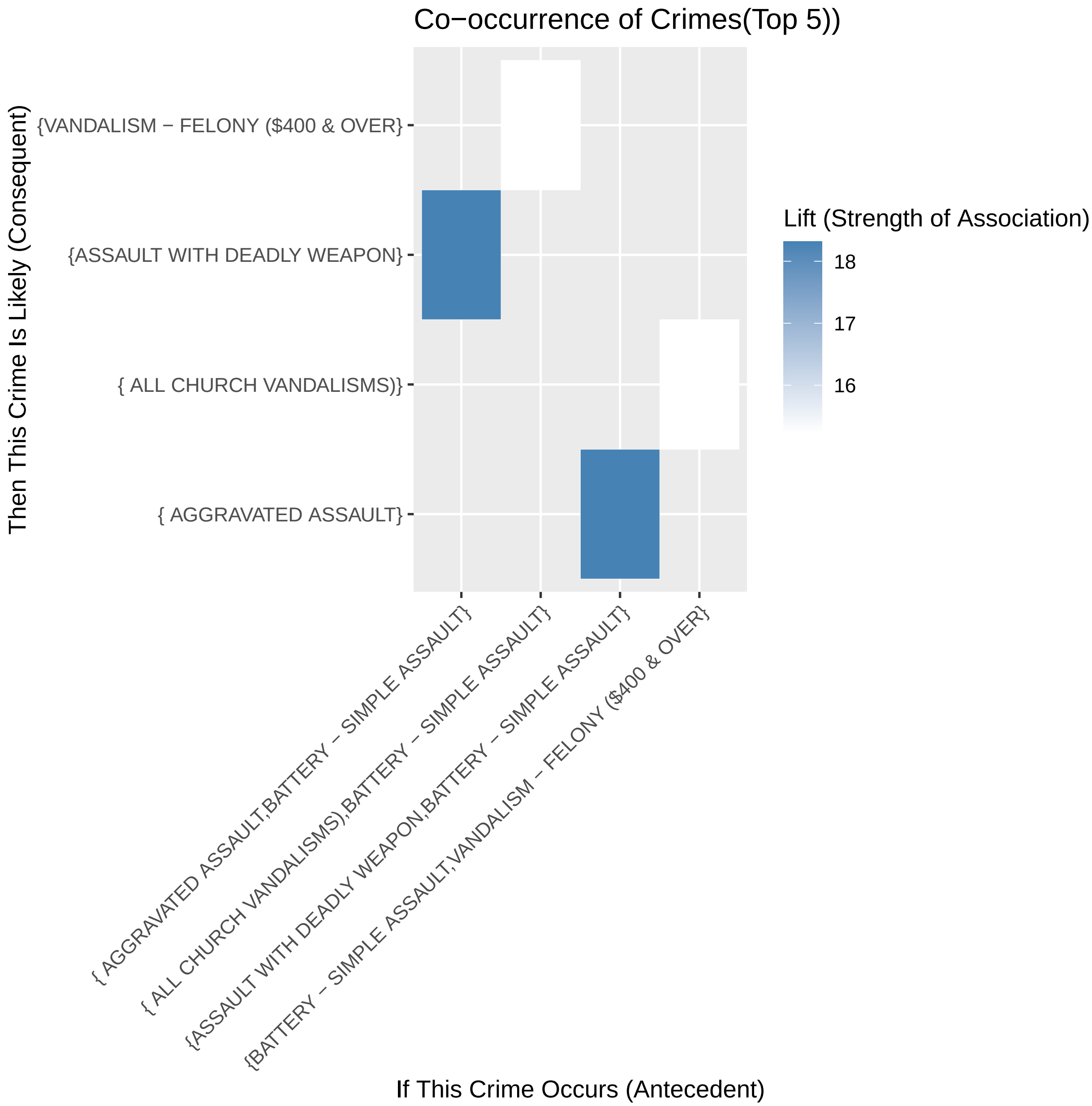

Figure 16: Co-occurrence of Crime Types

Strongest Association: Aggravated assault is most likely to be followed by battery - simple assault, indicating that perpetrators who commit aggravated assaults are more likely to also engage in simple battery offenses.

Moderate Association: Assault with a deadly weapon is followed by battery - simple assault, showing a similar pattern to the aggravated assault case, where a higher level of assault is associated with a simple battery offense.

No Association: The other co-occurrences on the heatmap have no statistically significant association (represented by white squares), meaning that the other crimes don't seem to have a strong relationship with one another.

3.6.6 Spatial Clustering and Geographic Features

open street map from plotly was used to visualize the spatial distribution of crime in Los Angeles.

```
# Grouping data by area and calculating the mean(latitude & longitude) for each area,
district_crime_counts <- crime_data %>%
  group_by(area_name) %>%
  summarise(
    latitude = mean(latitude, na.rm = TRUE),
    longitude = mean(longitude, na.rm = TRUE),
    counts = n(),
    common_crime = names(sort(table(crime_description), decreasing = TRUE)[1])
  ) %>%
  ungroup()

# Verify the data
print(district_crime_counts)

# Map Plotting
fig <- plot_ly(
  data = district_crime_counts,
  lat = ~latitude,
  lon = ~longitude,
  size = ~counts,
  color = ~counts,
  colors = colorRamp(c("blue", "red")),  # Using a custom color ramp
  text = ~paste("Area:", area_name, "<br>Counts:", counts, "<br>Common Crime:", common_c
  hoverinfo = "text",
  type = 'scattermapbox',
  mode = 'markers',
  marker = list(sizeref = 0.1, sizemode = 'area'),
  height = 750,
  width = 1200
) %>%
  layout(
    title = "Spatial Clustering of Crime Counts by District",
    mapbox = list(
      style = "open-street-map",
      zoom = 9,  # Adjust zoom level for better focus
      center = list(lat = 33.922, lon = -117.9437)  # Centering on Los Angeles
    )
```

```
)
```

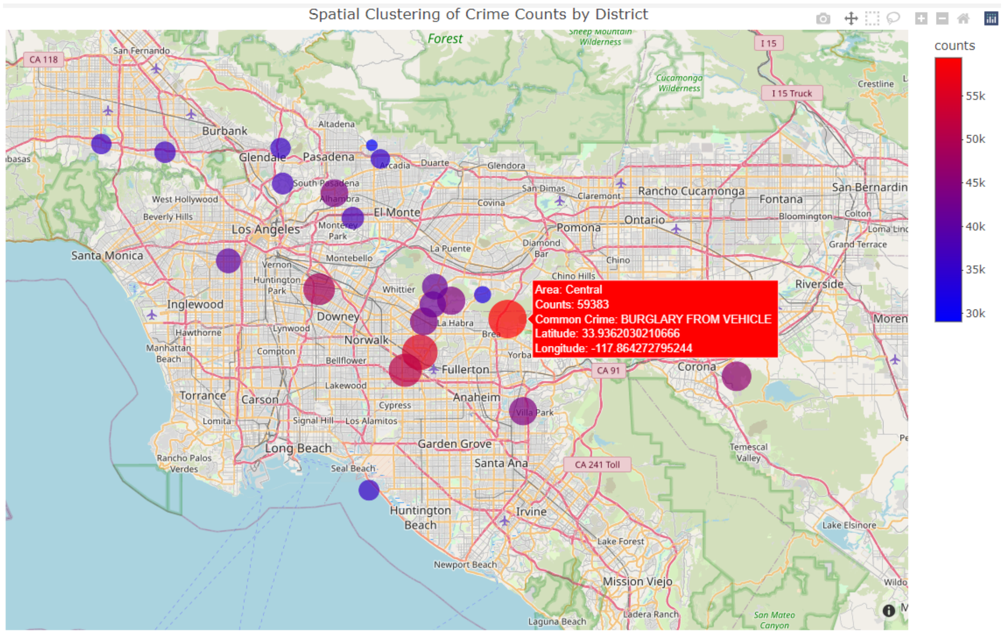

Figure 17: Spatial Clustering of Crime Counts by District (Interactive version available in the accompanying Shiny application)

The central LA area has a higher concentration of crimes, likely because it has more people, more businesses, and is easier to get around. The map shows a clear cluster of vehicle burglary in this area, suggesting that vehicles are more likely to be broken into here. This could be because there are a lot of cars, plenty of parking spaces, and potential targets for thieves.

3.6.7 Weapon Use and Victim Age Relationships

We used box plots to analyze the relationship between victim age and weapon type for common area.

```
# Filter for the top crime types and top weapons used
# 1. Filter data for the Central area
central_area_data <- crime_data %>%
  filter(area_name == "Central")

# Filter for top weapon types (excluding "Unknown")
top_weapons <- central_area_data %>%
  filter(!is.na(weapon_description) & !(weapon_description %in% "Unknown")) %>%
  count(weapon_description) %>%
  top_n(5, n) %>%
```

```r
  pull(weapon_description)

# Filter data for top weapons and Central area
filtered_data <- central_area_data %>%
  filter(weapon_description %in% top_weapons)


# Box Plot
ggplot(filtered_data, aes(x = weapon_description, y = victim_age)) +
  geom_boxplot(fill = "lightblue", color = "black", outlier.color = "red") +
  labs(
    title = "weapons and Victims(Central Area)",
    subtitle = "Main weapons used against age groups",
    x = "Weapon Used",
    y = "Victim Age"
  ) +
  theme_minimal() +
  theme(
    text = element_text(family = "sans", color = "black"),
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
    axis.text.x = element_text(angle = 45, hjust = 1),
    axis.text = element_text(size = 12),
    axis.title = element_text(size = 14),
    panel.grid.major.x = element_blank()  # Remove vertical grid lines
  )
```

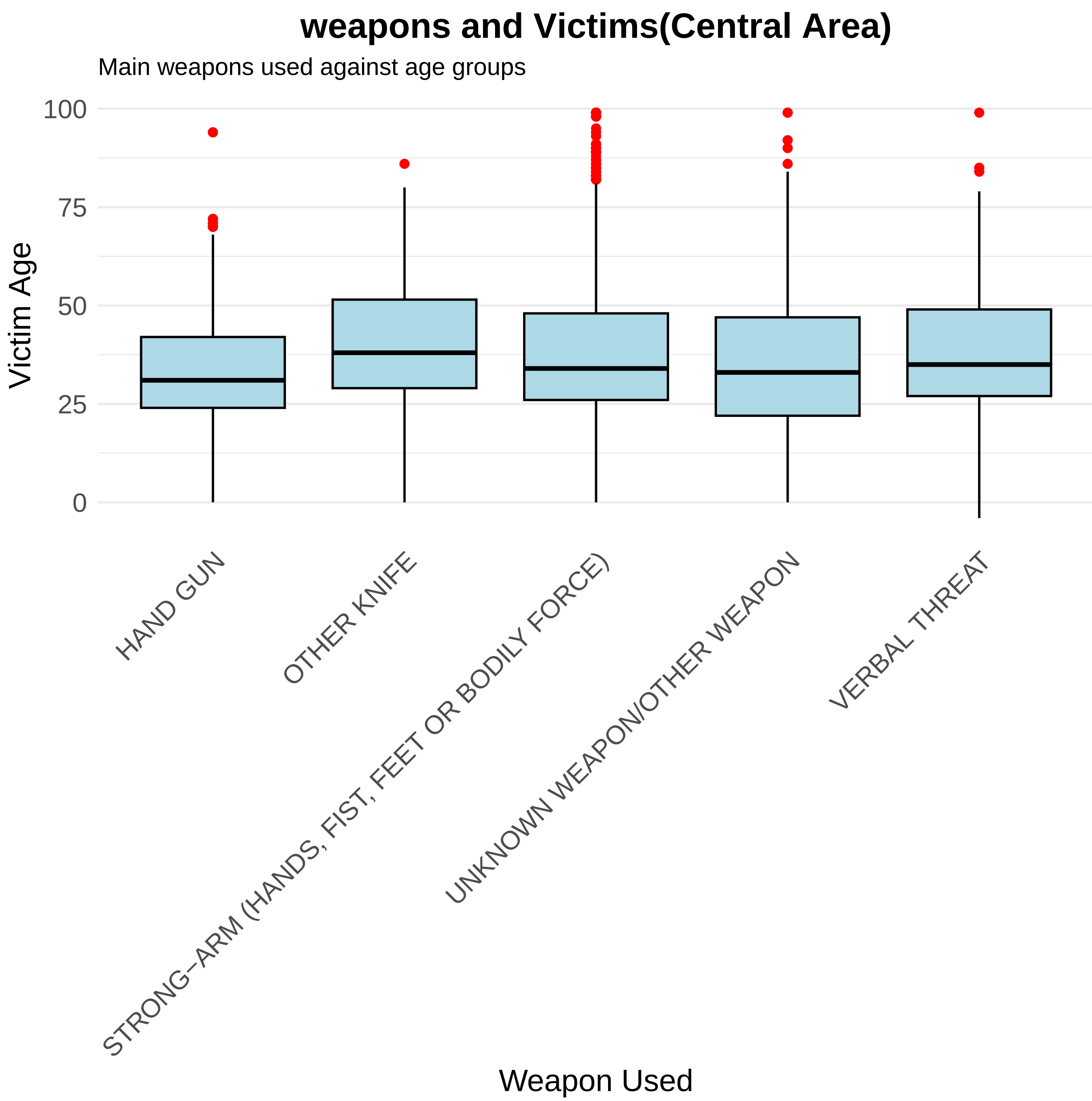

Figure 18: Weapon Use and Victim Age Relationships

The boxplots show a broad range of victim ages across all weapon types, indicating no specific age group is targeted more than others. The median age is typically around 40-50 years across most categories. Outliers, especially for "Hand Gun" and "Unknown Weapon/Other Weapon," suggest older victims may be involved in those incidents. The "Strong-arm (Hands, Fist, Feet, or Bodily Force)" category has a wider age range of victims compared to others, suggesting this method of assault could target a wider variety of victims. However, some categories like "Other Knife" and "Verbal Threat" have limited data points, which may limit our conclusions. Overall, the type of weapon used does not significantly influence the age of victims.

3.6.8 Case Clearance Rates and Variations

We used bar charts and stacked bar charts to visualize the distribution of case statuses for different crime types.

```r
# Filter for the top 10 most frequent crime types
top_crimes <- crime_data %>%
  count(crime_description) %>%
  top_n(6, n) %>%
  pull(crime_description)

# Filter and group the data
crime_status_data <- crime_data %>%
  filter(crime_description %in% top_crimes) %>%
  group_by(crime_description, status_description) %>%
  summarize(count = n(), .groups = 'drop')

# Reorder levels of status_description within each crime_description group
crime_status_data <- crime_status_data %>%
  arrange(crime_description, desc(count)) %>%
  mutate(status_description = factor(status_description, levels = unique(status_descript

# Faceted Bar Plot
ggplot(crime_status_data, aes(x = status_description, y = count, fill = status_descripti
  geom_bar(stat = "identity", color = "black", size = 0.3, width = 0.7) +  # Added borde
  facet_wrap(~ crime_description, scales = "free_y", ncol = 2) +
  labs(
    x = "Status Description",
    y = "Count",
    fill = "Status Description",
    title = "Case Clearance Rates and Variations along Time"
  ) +
  theme_minimal() +
  theme(
    text = element_text(family = "sans", color = "black"),
    plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
    axis.text = element_text(size = 11),
    axis.text.x = element_text(angle = 45, hjust = 1),
    axis.title = element_text(size = 13),
    strip.text = element_text(size = 8, face = "bold"),
    panel.grid.major.x = element_blank(),
    panel.spacing = unit(1.2, "lines") # Add spacing between facets
  )
```

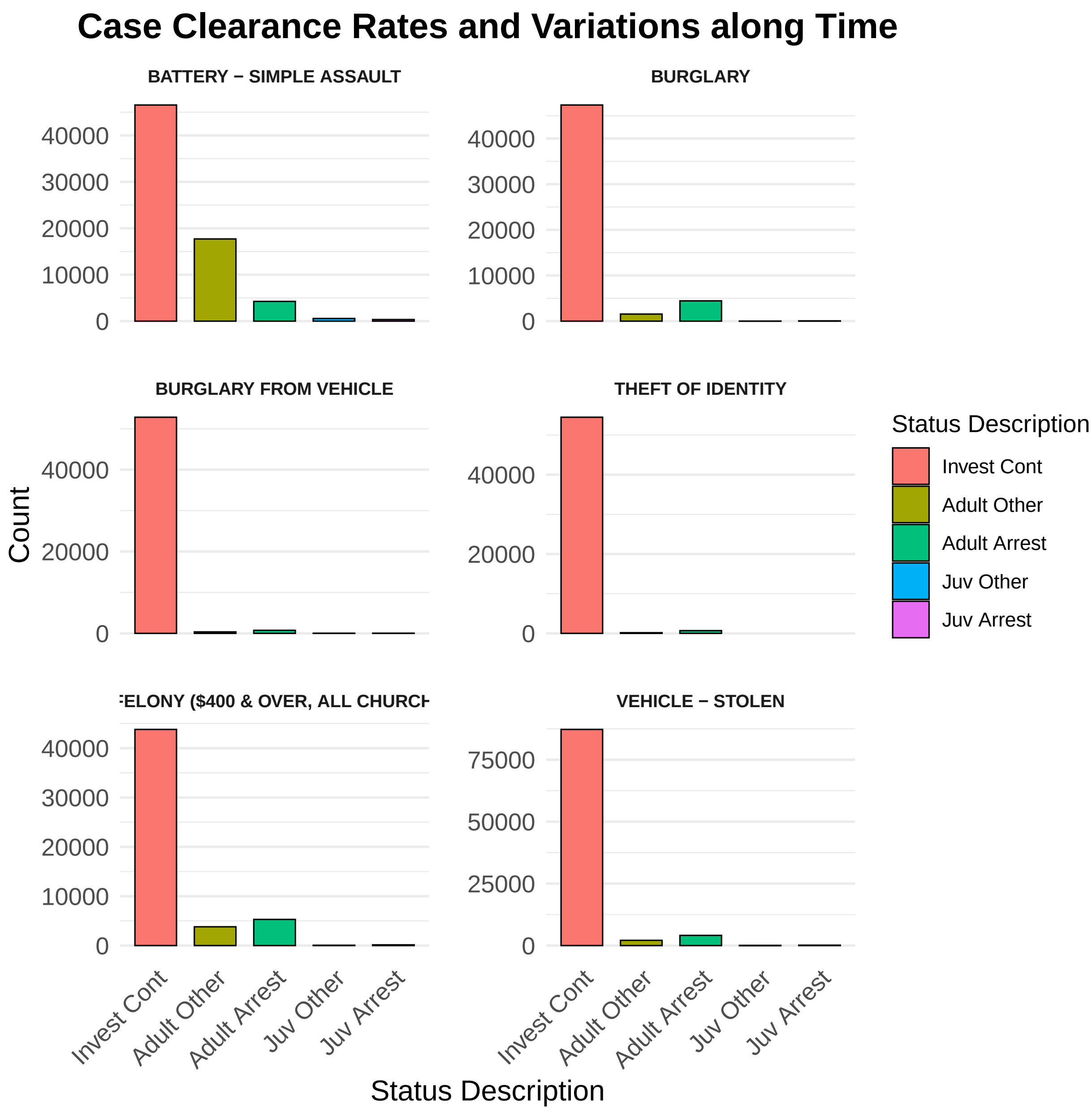

Figure 19: Case Clearance Rates and Variations

Vehicle-related Crimes like "Burglary from Vehicle" and "Vehicle - Stolen" have moderate clearance rates, attributed to factors and challenges in physical evidence, witness reports, and targeted investigations. On the other hand, Violent Crimes like "Battery - Simple Assault" and "Burglary" face higher clearance rates due to evidence, witness accounts, underreporting, and identifying perpetrators.

Summary

1. **Does the trend of crime concentration in certain areas differ for different crime types?**
   - **Expected Answer:** Trends might differ for various crime types over time and by area.
   - **Actual Answer:** Yes, different crime types show varied trends in concentration across different areas over time.
2. **What crimes are dominant in nighttime and daytime, and do the victim demographics in terms of age and gender differ between day and night?**
   - **Expected Answer:** Nighttime and daytime crimes differ, with varying victim demographics by time of day.
   - **Actual Answer:** Nighttime crimes are typically violent, while daytime crimes are more diverse. Victim demographics vary slightly by time of day.
3. **Do crime patterns in Los Angeles exhibit consistent trends across different time scales (monthly, seasonal, daily) throughout the years, or are there noticeable variations and inconsistencies that suggest the influence of external factors?**
   - **Expected Answer:** Crime patterns might exhibit consistent trends across different time scales (monthly, seasonal, daily).
   - **Actual Answer:** Crime patterns show both consistent trends and seasonal variations influenced by external factors.
4. **Do certain crime types in Los Angeles frequently occur together at the same locations and time, and if so, what crimes are more likely to occur if one happens?**
   - **Expected Answer:** Certain crime types in Los Angeles might frequently occur together at the same locations and time, and there might be a connection between these crimes.
   - **Actual Answer:** Yes, certain crime types, such as burglary and vehicle theft, frequently co-occur at the same locations and times.
5. **In the most crime-occurring area in Los Angeles, is there a relationship between the age of the victim and the type of weapon used?**
   - **Expected Answer:** There might be a relationship between the age of the victim and the type of weapon used.
   - **Actual Answer:** No specific age group is disproportionately targeted, indicating that the type of weapon used does not seem to be a major factor in determining the age of victims.

6. **Do the differences in case clearance rates across different crime types suggest variations in investigative resources, police priorities, or the inherent difficulty of solving specific crimes?**

   - **Expected Answer:** Variations in case clearance rates across different crime types might be linked to investigative resources, police priorities, or the inherent difficulty of solving specific crimes.
   - **Actual Answer:** Yes, vehicle-related crimes have higher clearance rates, while violent crimes have lower clearance rates, likely due to varying investigative challenges.

7. **Does the spatial distribution of crime in Los Angeles, visualized by area's crime and most common crime types, reveal any clustering patterns or relationships with geographic features??**

   - **Expected Answer:** Spatial distribution might reveal clustering patterns or relationships with geographic features.
   - **Actual Answer:** Crime distribution reveals clustering patterns related to geographic features like commercial areas and transport hubs.

Conclusions

The LAPD crime data analysis from 2020 to 2023 reveals some important points. Crimes are most common between 7 PM and 11 PM. Assault and buglary related crimes are solved more often than other crimes. Different types of weapons don't seem to target specific age groups more than others. This suggests that the type of weapon used doesn't strongly affect the victim's age. The report emphasizes the usefulness of visualizing data to find hidden patterns and trends. This can offer valuable insights for law enforcement and public safety planning.A lot of insights and stories can be extracted from these dataset in future to make law and policy even powerful.

References

1. Los Angeles Police Department (LAPD). (2020-2023). Crime Data from 2020 to Present. Retrieved from `https://catalog.data.gov/dataset/crime-data-from-2020-to-present`

2. Walker, J.T. and Drawve, G.R. (2018) 'What is crime analysis?', Foundations of Crime Analysis, pp. 18–34. doi:10.4324/9781315716442-2.

3. 'Apriori algorithm' (no date) SpringerReference [Preprint]. doi:10.1007/springerreference_178739.

4. Ariza, J.M. and Solymosi, R. (2023) 'Crime along spatial networks', Crime Mapping and Spatial Data Analysis using R, pp. 205–238. doi:10.1201/9781003154914-8.

5. Mauron, Y. (2020) 'Catmaply: Heatmap for categorical data using "plotly"', CRAN: Contributed Packages [Preprint]. doi:10.32614/cran.package.catmaply.