Analysis of Crime in the City of Los Angeles

Sta 523 - Final Project

AUTHORS
teammate 1 Binqian Chai
teammate 2 Bohao Yang

Introduction

The city of Los Angeles, a popular place for both domestic travelers and international tourists, offers endless attractions but also faces complex safety challenges. With safety being a crucial priority for both visitors and residents, crime analysis becomes the topic of interest. This project aims to be informative and is divided into two distinct parts: Exploratory Data Analysis (EDA) and an interactive Shiny app. We will use the Los Angeles Crime Data from Kaggle, focusing on the most relevant period from January to mid-April 2024. The found dataset is saved and uploaded as “crime2024.xlsx”. The library “readxl” is used to load the data.

The first part focuses on Exploratory Data Analysis (EDA), where we analyze patterns in the found dataset to uncover meaningful insights about crime trends. By examining temporal, geographic, and demographic aspects of the data, the EDA reveals key patterns, such as which months and weekdays are associated with higher crime rates, the types of crimes most frequently reported, and how crime varies across different areas in Los Angeles. This analysis not only reveals the current trends but also provides actionable insights into potential contributing factors, such as seasonal activity, socioeconomic disparities, and urban characteristics.

The second part involves the creation of a Shiny app, aimed at providing users with an interactive platform to explore crime data in Los Angeles. The app is designed to help users make informed decisions about their travel and movement within the city by offering accessible, data-driven insights. Users can interact with a city map, which uses color-coded markers to represent crime density by area. Filters allow users to explore crime trends based on their selected date range, while bar charts display the top areas with the highest crime rates for comparison. A detailed table further enables users to investigate individual crime incidents, including the date, time, location, and type of crime. Together, these features provide a comprehensive and user-friendly tool for exploring crime patterns.

By combining EDA and the Shiny app, this project achieves a dual purpose: to analyze and interpret crime data for meaningful insights, and to provide an interactive platform that provides users with a deeper understanding of safety trends in Los Angeles. Through these efforts, we aim to enhance public awareness, support better decision-making, and contribute to a safer community.

Variable Explanation

The following are the definitions or brief explanations of variables (column names in the dateset) that will be used inside the project. 1) DATE OCC: the date crime occurred in YYYY-MM-DD 2) TIME OCC:

the time (in 24 hour military time) crime occurred 3) AREA NAME: the geographic area name in the city of Los Angeles that references a landmark or the surrounding community that the crime occurred 4) Part 1-2: crime severity with 1 referring to “serious” and 2 referring to “less serious” 5) Crm Cd Desc: a short phrase describing the crime (type) 6) Vict Descent: the descent of victims (The descent code will be explicitly definined in the Implementation section.) 7) Premis Desc: the type of location where the crime occurred 8) LAT: the specific latitude where the crime actually occurred 9) LON: the specific longitude where the crime actually occurred

Implementation

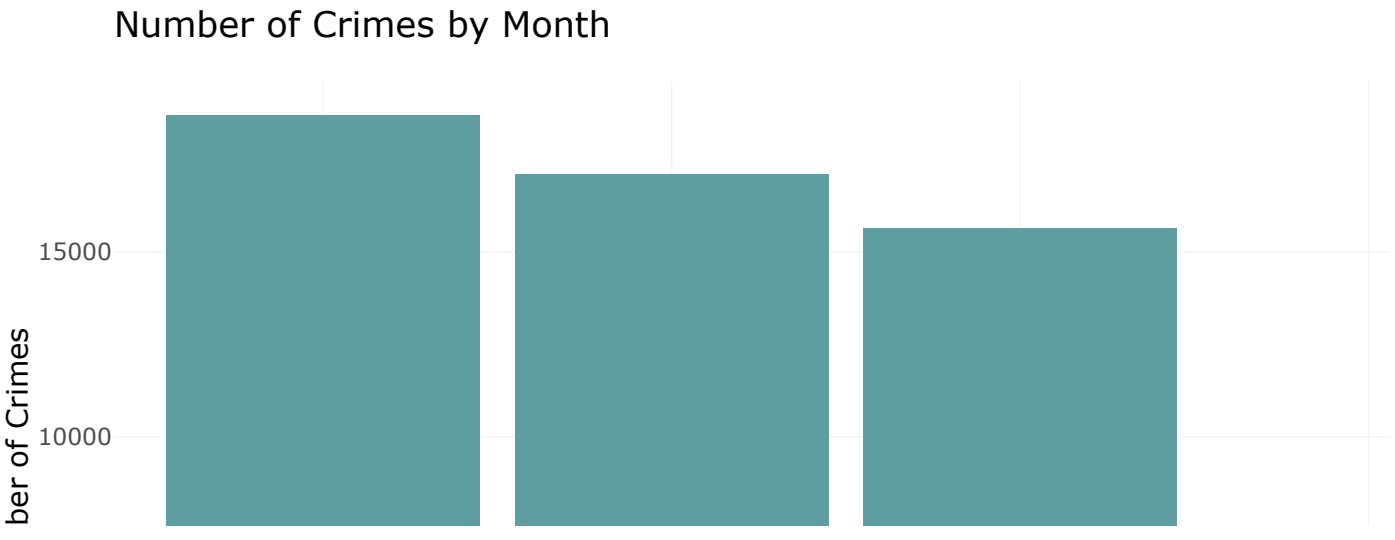
Part 1: Exploratory data analysis with visualization

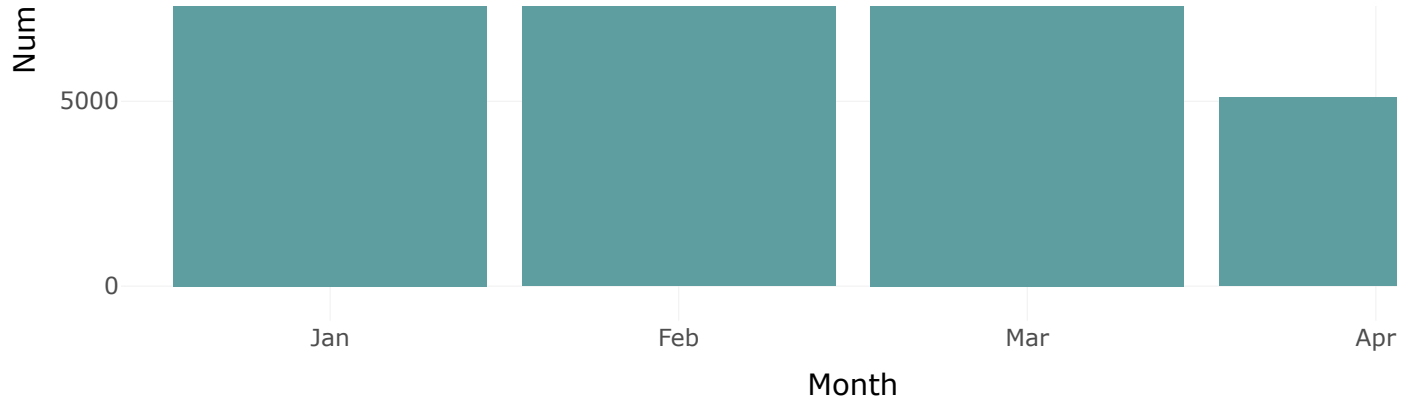
Data Pre-processing

```
# Load the crime data
data <- as.data.frame(read_excel("crime2024.xlsx"))
# Change the format of the date occurred
data$`DATE OCC` <- as.Date(data$`DATE OCC`, format = "%m/%d/%Y %H:%M:%S")
```

1. Compare the number of crimes by month

```
# Filter the months
data$month <- month(data$`DATE OCC`, label=TRUE)

# Plot the number of crimes by month
ggplotly(
  data |>
    count(month) |>
    ggplot(aes(x = month, y = n)) +
    geom_bar(stat = 'identity', fill = 'cadetblue') +
    labs(x = 'Month', y = 'Number of Crimes', title = 'Number of Crimes by Month') +
    theme_minimal()) |>
  config(displayModeBar = FALSE)
```

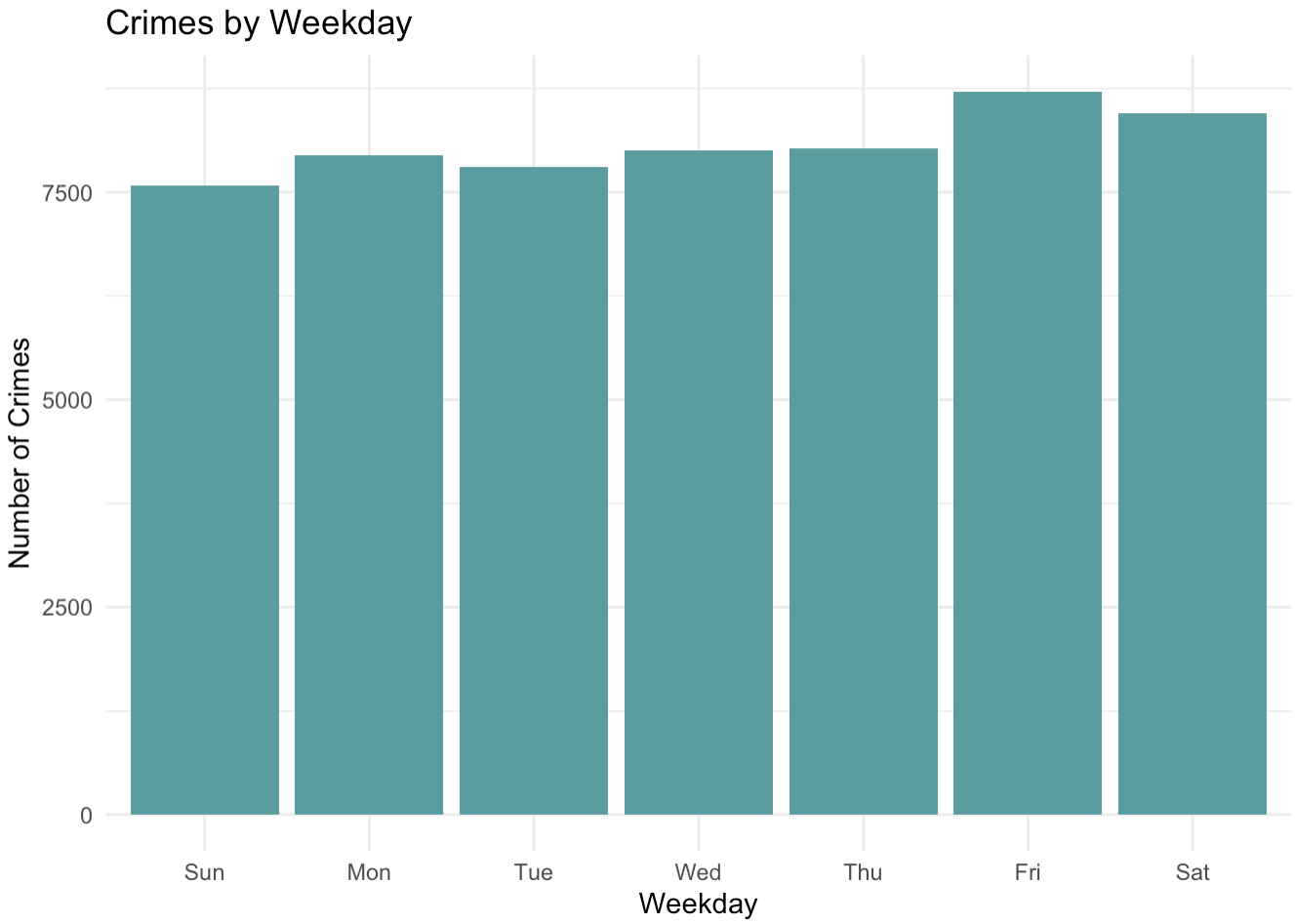




The plot indicates that the number of crimes in January 2024 is the highest. This could possibly be attributed to factors related to the holiday season. 1) The increased shopping, travel, and gatherings during the holidays create more opportunities for theft, burglary, and other property crimes. 2) Financial strain from holiday spending may push some individuals toward opportunistic crimes like shoplifting or fraud. 3) Additionally, celebrations such as New Year's Eve often involve higher alcohol consumption, potentially contributing to crimes or disorderly behavior. Los Angeles, as a tourist-heavy city, may have more crimes targeting travelers, such as pickpocketing or scams.

However, since the crime data for April may still be incomplete, the above visualization might not provide meaningful insights. To gain a clearer understanding, we compare the number of crimes by weekday instead.

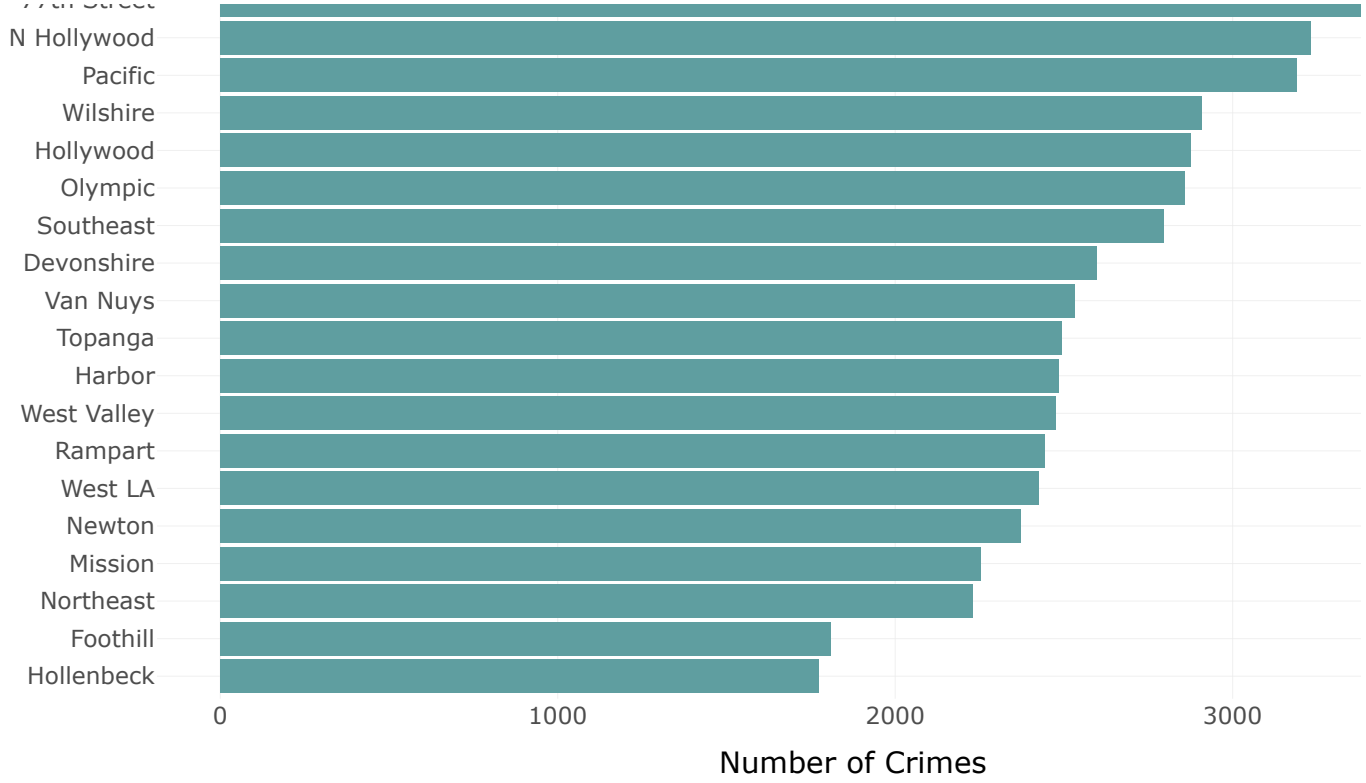
```
# Filter the data by weekday
data$weekday <- wday(data$`DATE OCC`, label=TRUE)

# Plot the number of crimes by weekday
data |>
  count(weekday) |>
  ggplot(aes(x = weekday, y = n)) +
  geom_bar(stat = 'identity', fill = 'cadetblue') +
  labs(x = 'Weekday', y = 'Number of Crimes', title = 'Crimes by Weekday') +
  theme_minimal()
```



This plot shows that Friday has the highest number of crimes compared to other weekdays. Reasons may include: 1) Social Activities: There tends to be more people going out on Fridays, increasing potential targets and reducing supervision. 2) Alcohol Use: People tends to drink more on Friday nights, and drinking can lead to impaired judgment and aggressive behavior. 3) Crime Patterns: Violent crimes like assaults tend to rise on Fridays and weekends.

2. Compare the number of crimes by area in LA

```
# Plot the number of crimes for different areas in LA
ggplotly(
  data |>
    count(`AREA NAME`) |>
    ggplot(aes(x = reorder(`AREA NAME`, n), y = n, fill = n, text = `AREA NAME`)) +
    geom_bar(stat = 'identity', fill = 'cadetblue') +
    coord_flip() +
    labs(x = '', y = 'Number of Crimes') +
    theme_minimal(),
  tooltip = c("text", "y")) |>
  config(displayModeBar = FALSE)
```



From this bar chart, we can see: 1) High-Crime Areas: The top areas with the highest number of crimes are Central, Southwest, and 77th Street, indicating they may require more focused law enforcement or community programs. 2) Moderate-Crime Areas: Areas like Wilshire, Hollywood, and Olympic have moderate crime levels. They may still be hotspots but not as concentrated as the top three. 3) Low-Crime Areas: Foothill, Hollenbeck, and Northeast have the lowest crime counts. These areas might reflect better community safety, fewer opportunities for crime, or underreporting.

There may be a correlation between crime rates and socioeconomic factors, population density, or specific characteristics of these areas. High-crime areas such as Central, Southwest, and 77th Street may be characterized by higher population densities and socioeconomic challenges, including lower median household incomes, higher unemployment rates, and underfunded infrastructure. For instance, areas like Central Los Angeles, with its dense housing and limited access to community resources, may create conditions where property crimes like burglary and vehicle theft are more prevalent. In contrast, low-crime areas like Foothill, Hollenbeck, and Northeast often reflect better community safety, supported by stronger neighborhood cohesion. For example, areas like Northeast Los Angeles, which has a mix of residential and open spaces, may have fewer opportunities for crime and benefit from a smaller transient population, reducing the risk of opportunistic crimes. Furthermore, immigrant-heavy neighborhoods like Foothill may experience underreporting of crimes due to fear of legal repercussions or mistrust in law enforcement, potentially making them appear safer than they truly are.

3. Compare number of crimes by types

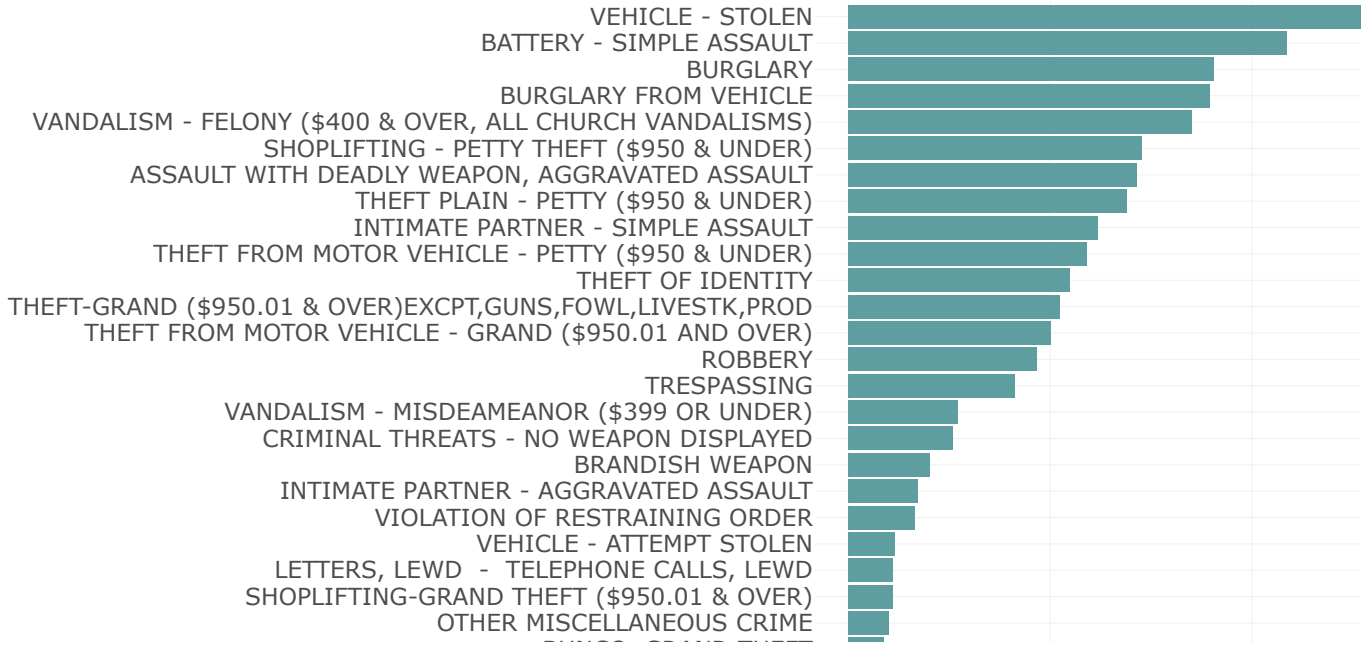
```
# Filter the data
crime_type <- data |>
  group_by(`Crm Cd Desc`) |>
  summarise(count = n()) |>
  arrange(desc(count))
```

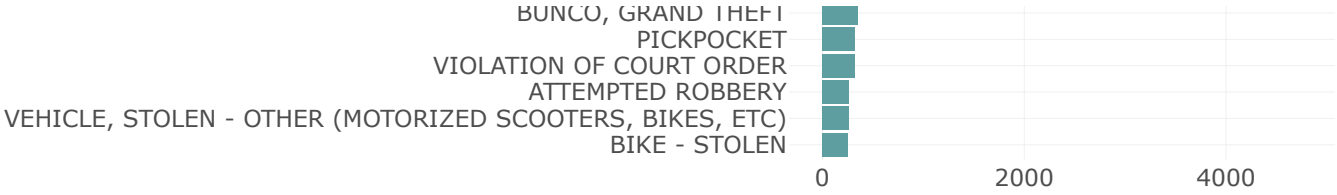
```
# View the whole table
crime_type
```

```
# A tibble: 125 × 2
   `Crm Cd Desc`                                          count
   <chr>                                                  <int>
 1 VEHICLE - STOLEN                                        6611
 2 BATTERY - SIMPLE ASSAULT                                4343
 3 BURGLARY                                                3621
 4 BURGLARY FROM VEHICLE                                   3585
 5 VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS) 3404
 6 SHOPLIFTING - PETTY THEFT ($950 & UNDER)                2912
 7 ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT          2858
 8 THEFT PLAIN - PETTY ($950 & UNDER)                      2757
 9 INTIMATE PARTNER - SIMPLE ASSAULT                       2475
10 THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)         2367
# ℹ 115 more rows
```

From the above table, all different types of crime according with their total number occurred are shown in descending order. Now visualize 30 main types of crimes.

```
# Plot 30 main types of crimes by count
ggplotly(
  crime_type |>
    head(30) |>
    ggplot(aes(x = reorder(`Crm Cd Desc`, count), y = count, text = `Crm Cd Desc`)) +
    geom_bar(stat = 'identity', fill = 'cadetblue') +
    coord_flip() +
    labs(x = '', y = 'Types of Crime') +
    theme_minimal(),
  tooltip = c("text", "y")) |>
  config(displayModeBar = FALSE)
```





The plot shows that: 1) Most Common Crimes: Vehicle Theft is the most frequent type of crime by a significant margin. Other common crimes include Battery (Simple Assault) and Burglary, showcasing issues with personal and property safety. 2) High-Impact Crimes: Assault with Deadly Weapon and Theft (Grand and Petty) appear frequently, reflecting concerns about both violent crimes and property theft. 3) Intimate Partner Crimes: Crimes such as Intimate Partner - Simple Assault and Violation of Restraining Orders highlight ongoing challenges with domestic violence and associated legal compliance. 4) Specialized Thefts: Shoplifting, Theft of Identity, and Bike Theft suggest opportunities for targeted measures to reduce these specific types of crime. 5) Lesser Common Crimes: Pickpocketing, Attempted Robbery, and Bike Theft are less frequent but may still require attention in specific contexts.

Based on the insights shown from the plot, possible solutions could be: 1) Resource Allocation: The focus should be on combating vehicle theft, assault, and burglary by deploying targeted law enforcement and public awareness campaigns. 2) Crime Prevention Strategies: Enhanced security measures for vehicles and increased monitoring in high-theft areas can address the most frequent crimes. 3) Social Interventions: Addressing intimate partner violence and restraining order violations may require collaboration between law enforcement and social services.

4. Compare the main crime types by area (top5)

```
# Filter the crime data by area
crime_by_area <- data |>
  group_by(`AREA NAME`, `Crm Cd Desc`) |>
  summarise(count = n(), .groups = 'drop') |>
  arrange(`AREA NAME`, desc(count))

# Find the top 5 areas by total crime count
top_areas <- crime_by_area |>
  group_by(`AREA NAME`) |>
  summarise(total_count = sum(count), .groups = 'drop') |>
  arrange(desc(total_count)) |>
  slice_head(n = 5) |>
  pull(`AREA NAME`)

# View the top 5 areas
top_areas
```

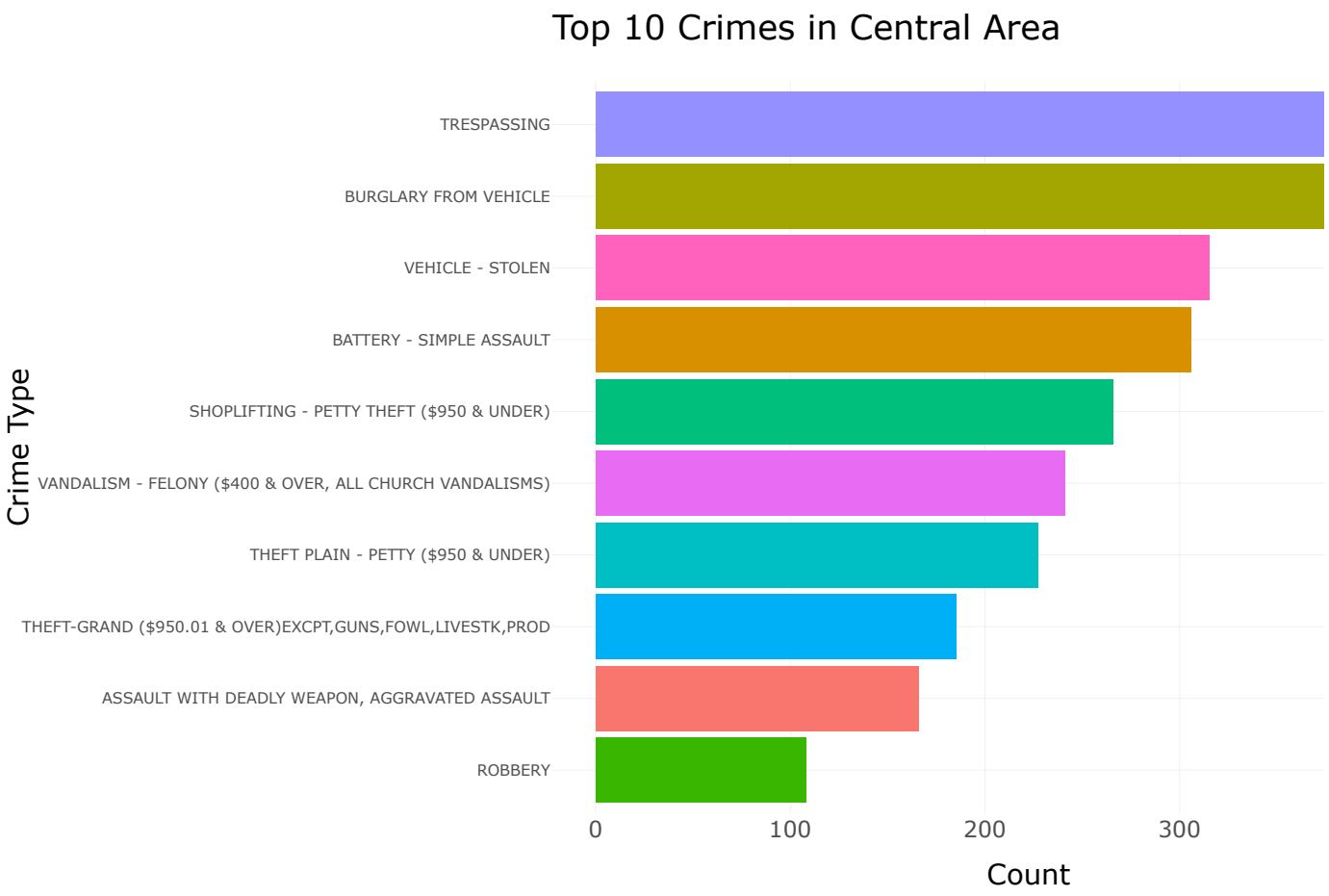
```
[1] "Central"     "Southwest"   "77th Street" "N Hollywood" "Pacific"
```

Top 5 areas with the highest crime numbers are Central, Southwest, 77th Street, N Hollywood and Pacific.

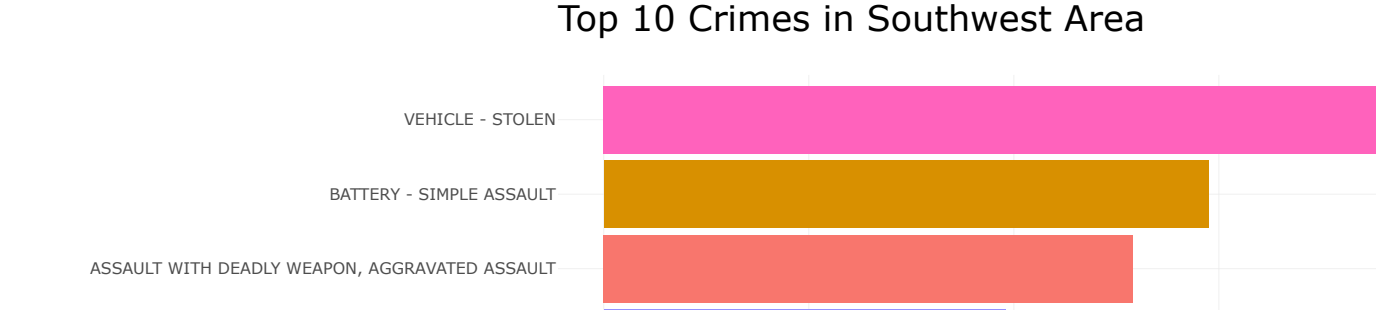
```
# Find the top 10 crime types for each area
top10_Central <- crime_by_area |>
  filter(`AREA NAME` == "Central") |>
  arrange(desc(count)) |>
  head(10)

top10_Southwest <- crime_by_area |>
  filter(`AREA NAME` == "Southwest") |>
  arrange(desc(count)) |>
  head(10)

top10_77st<- crime_by_area |>
  filter(`AREA NAME` == "77th Street") |>
  arrange(desc(count)) |>
  head(10)

top10_nhollywood <- crime_by_area |>
  filter(`AREA NAME` == "N Hollywood") |>
  arrange(desc(count)) |>
  head(10)

top10_pacific <- crime_by_area |>
  filter(`AREA NAME` == "Pacific") |>
  arrange(desc(count)) |>
  head(10)

# List of data for each area
area_data_list <- list(
  "Central" = top10_Central,
  "Southwest" = top10_Southwest,
  "77th Street" = top10_77st,
  "North Hollywood" = top10_nhollywood,
  "Pacific" = top10_pacific)
```

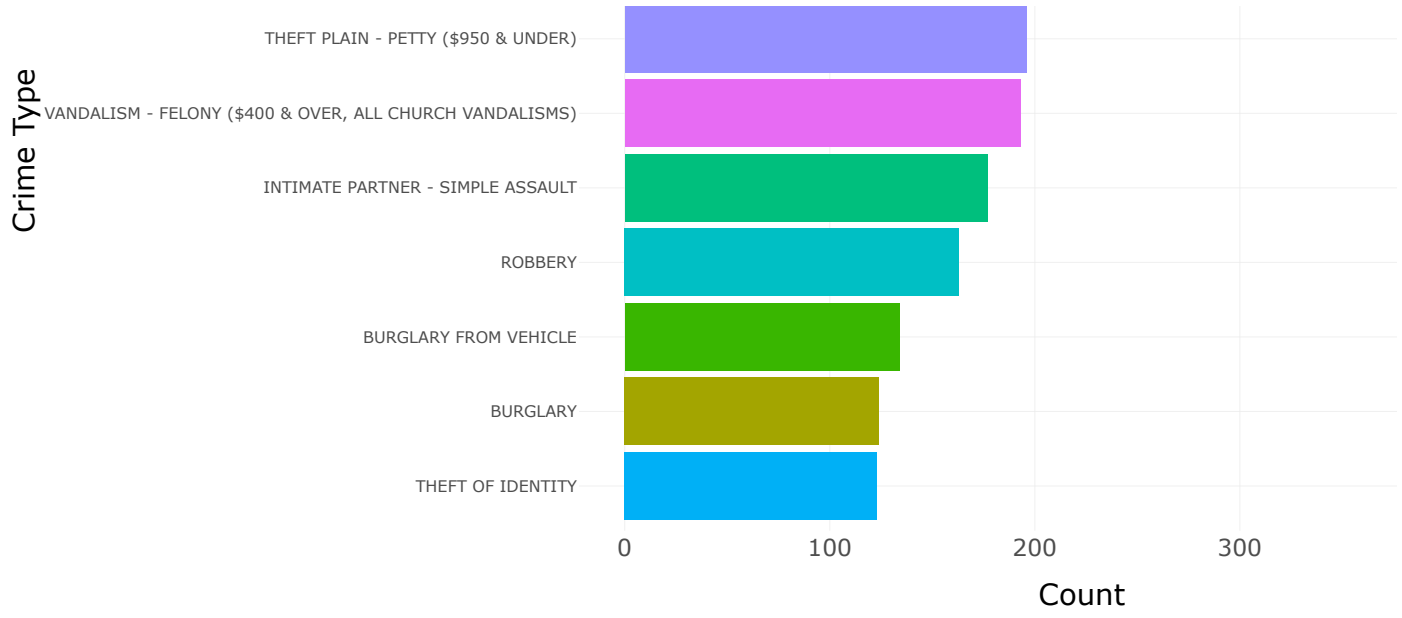
```
# Function to create plot for an area
create_crime_plot <- function(data, area_name) {
  ggplotly(
    ggplot(data, aes(
      x = reorder(`Crm Cd Desc`, count),
      y = count,
      fill = `Crm Cd Desc`,
      text = paste(`Crm Cd Desc`, ":", count))) +
      geom_bar(stat = "identity") +
      coord_flip() +
      labs(
        x = "Crime Type",
        y = "Count",
        title = paste("Top 10 Crimes in", area_name, "Area")) +
      theme_minimal() +
      theme(
```

```
        axis.text.y = element_text(size = 6),
        legend.position = "none"),
    tooltip = c("y")) |>
  config(displayModeBar = FALSE)
}

# Apply the function to all top 5 areas
crime_plots <- lapply(names(area_data_list), function(area_name) {
  create_crime_plot(area_data_list[[area_name]], area_name)
})


# Display the plots
crime_plots[[1]]
```

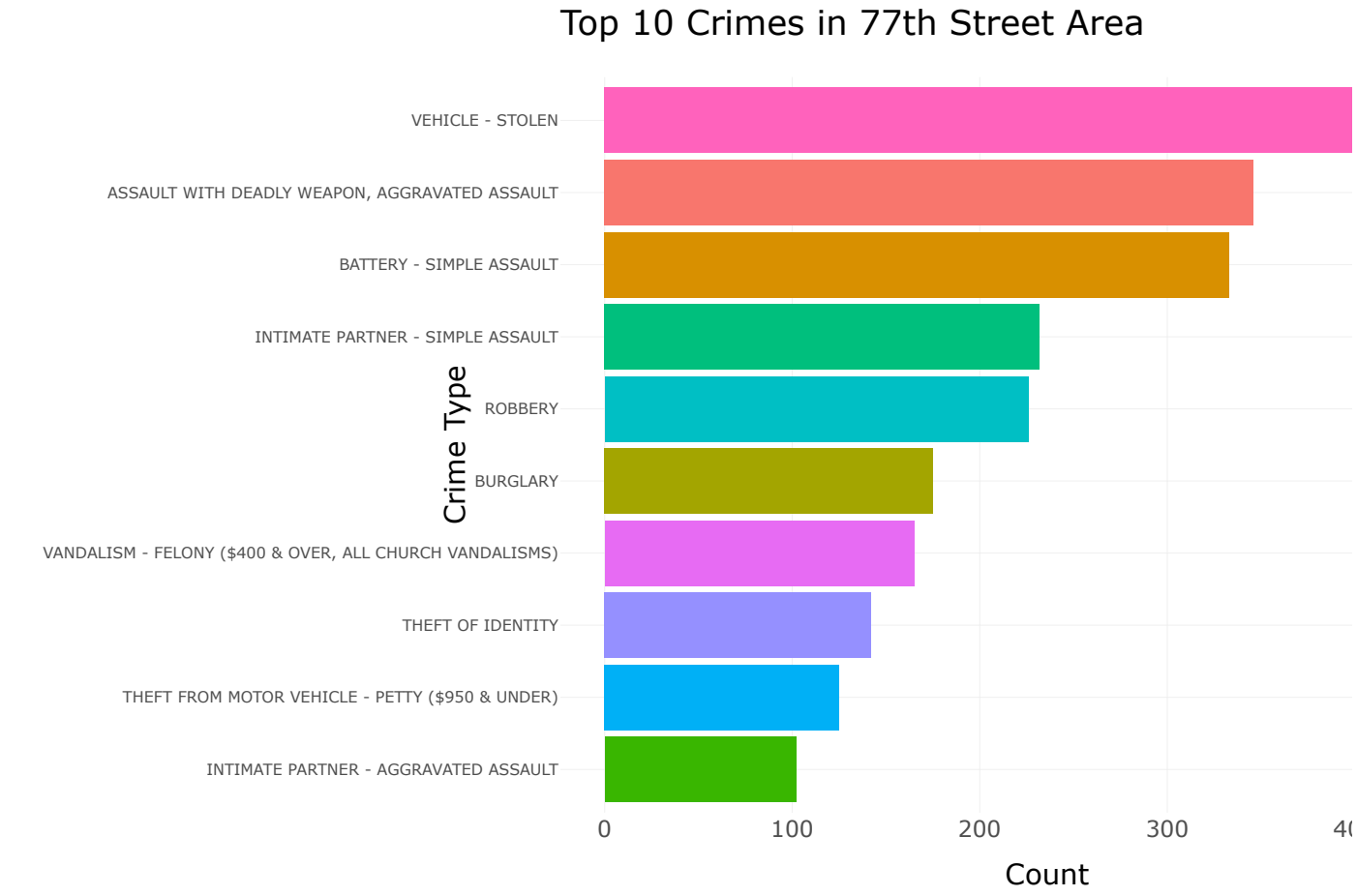



```
crime_plots[[2]]
```

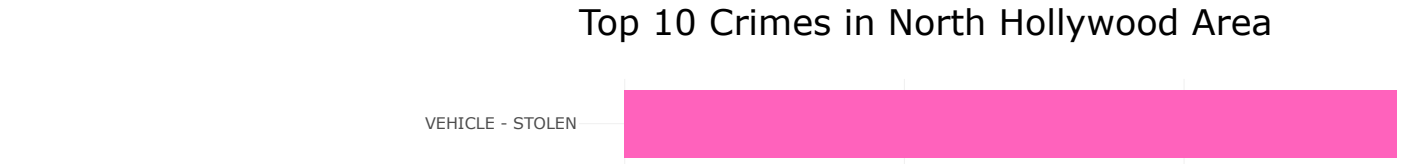


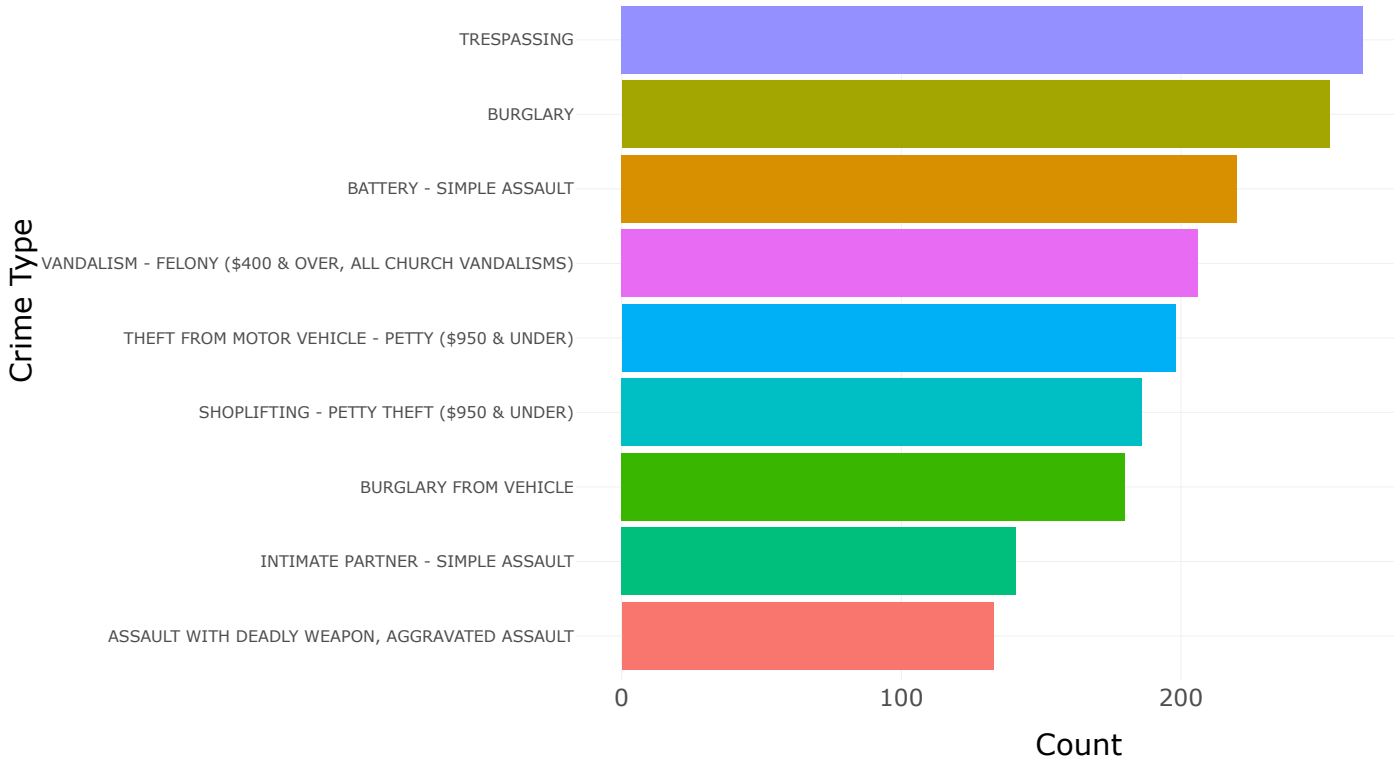



```
crime_plots[[3]]
```

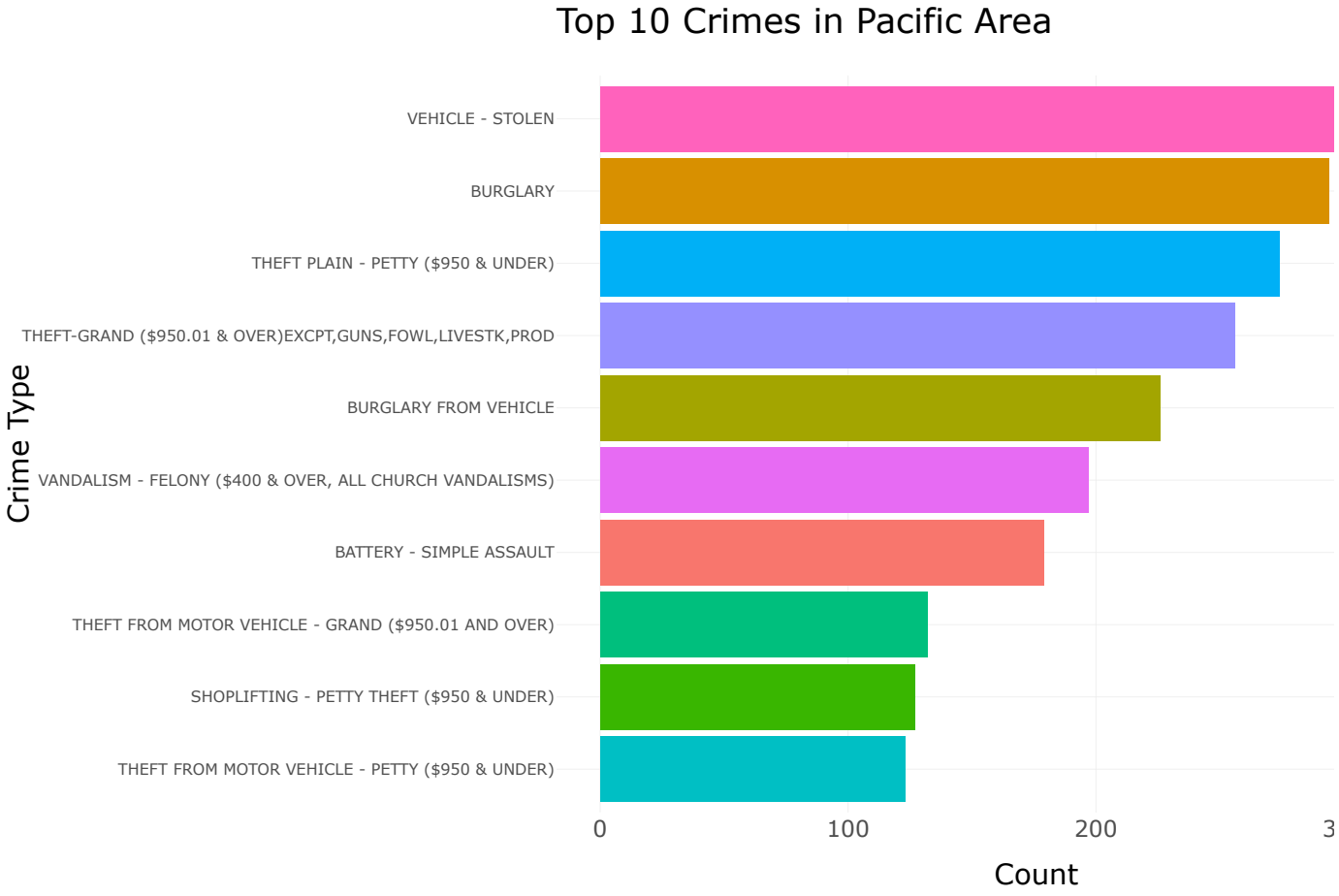



```
crime_plots[[4]]
```






```
crime_plots[[5]]
```



From the plots above, we can see: 1) Central Area: Top crimes: Trespassing, Burglary from Vehicle, and Vehicle Theft. Focus: Improve security for property and vehicles. 2) Southwest Area: Top crimes: Vehicle Theft, Simple Assault, and Assault with Deadly Weapon. Focus: Boost patrolling and address both

violent and property crimes. 3) 77th Street Area: Top crimes: Vehicle Theft, Assaults (Simple and Aggravated), and Domestic Violence. Focus: Prevent domestic violence and reduce thefts. 4) North Hollywood: Top crimes: Vehicle Theft, Trespassing, and Burglary. Focus: Strengthen measures against trespassing and theft. 5) Pacific Area: Top crimes: Vehicle Theft, Burglary, and Grand Theft. Focus: Target high-value property crimes and enhance vehicle security.

5. Compare the number of crimes by victim descents

```
# Create a lookup table for Victim Descents
descent_lookup <- c(
  "A" = "Other Asian",
  "B" = "Black",
  "C" = "Chinese",
  "D" = "Cambodian",
  "F" = "Filipino",
  "G" = "Guamanian",
  "H" = "Hispanic/Latin/Mexican",
  "I" = "American Indian/Alaskan Native",
  "J" = "Japanese",
  "K" = "Korean",
  "L" = "Laotian",
  "O" = "Other",
  "P" = "Pacific Islander",
  "S" = "Samoan",
  "U" = "Hawaiian",
  "V" = "Vietnamese",
  "W" = "White",
  "X" = "Unknown",
  "Z" = "Asian Indian")

# Replace Vict Descent codes with descriptions
data2 <- data |>
  mutate(`Vict Descent` = descent_lookup[`Vict Descent`])

# Find the number of victim by descent in LA
vict <- data2 |>
  filter(!is.na(`Vict Descent`)) |>
  group_by(`Vict Descent`) |>
  summarise(count = n(), .groups = 'drop') |>
  arrange(desc(count))

# View the whole table
vict
```

```
# A tibble: 19 × 2
   `Vict Descent`                count
   <chr>                         <int>
 1 Hispanic/Latin/Mexican        15832
 2 White                         10314
```

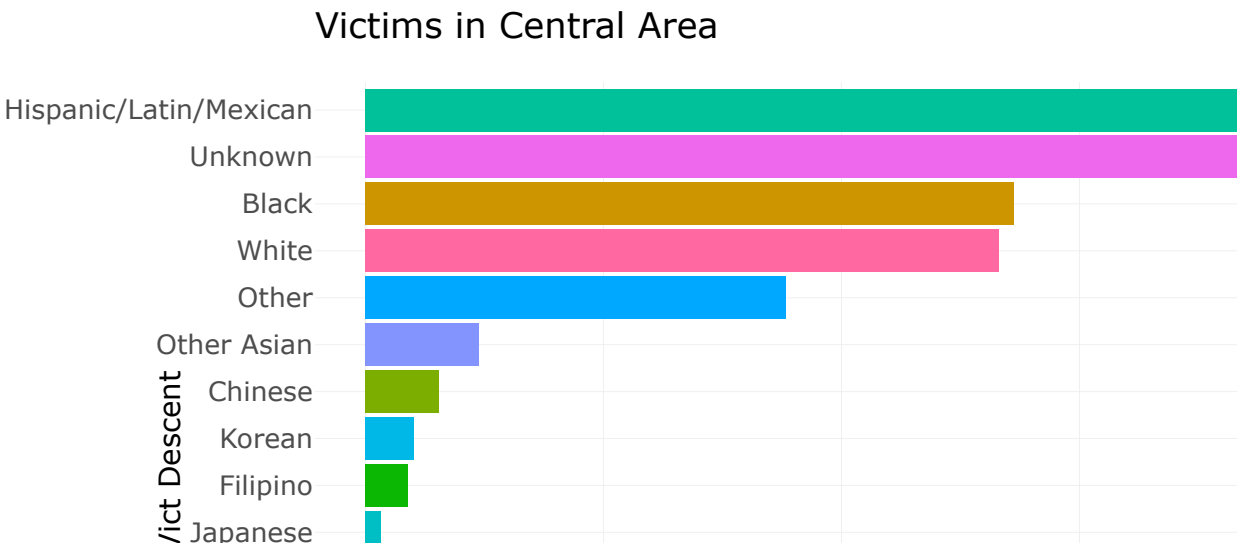
3	Unknown	8148
4	Black	6625
5	Other	4210
6	Other Asian	1208
7	Korean	378
8	Filipino	313
9	Chinese	299
10	Japanese	85
11	Vietnamese	70
12	American Indian/Alaskan Native	60
13	Hawaiian	19
14	Asian Indian	18
15	Pacific Islander	14
16	Laotian	7
17	Cambodian	4
18	Guamanian	3
19	Samoan	2

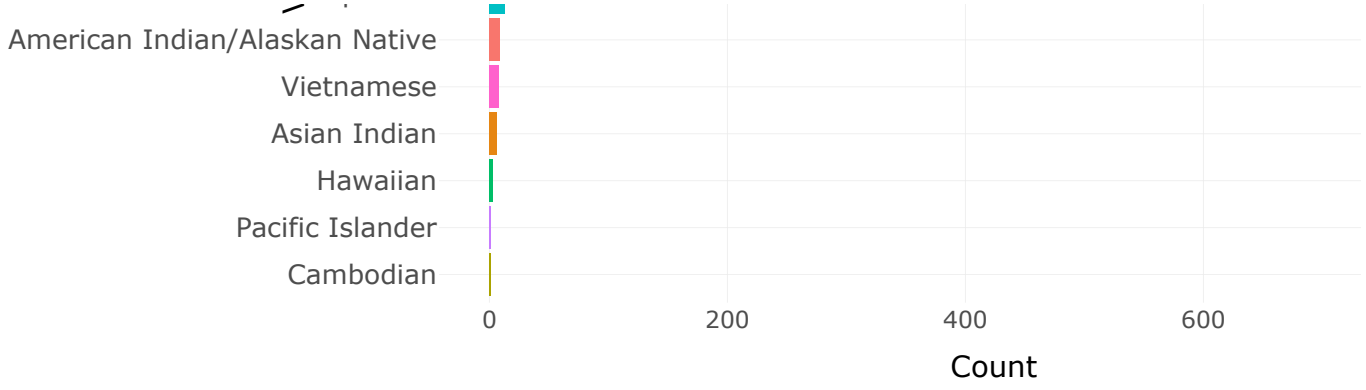
From the table above we can see the most frequent victim descent in Los Angeles (whole area) is Hispanic/Latin/Mexican with a count of 15,832, and the second most frequent descent is White, with 10,314 crimes. Now visualize the frequency of different victim descent for top 5 crime areas.

```
# Filter the number of victim by descent in top 5 areas
vict_by_area <- data2 |>
  filter(!is.na(`Vict Descent`)) |>
  group_by(`AREA NAME`, `Vict Descent`) |>
  summarise(count = n(), .groups = 'drop') |>
  arrange(`AREA NAME`, desc(count))

vict_Central <- vict_by_area |>
  filter(`AREA NAME` == "Central") |>
  arrange(desc(count))

vict_Southwest <- vict_by_area |>
  filter(`AREA NAME` == "Southwest") |>
  arrange(desc(count))

vict_77st<- vict_by_area |>
  filter(`AREA NAME` == "77th Street") |>
  arrange(desc(count))

vict_nhollywood <- vict_by_area |>
  filter(`AREA NAME` == "N Hollywood") |>
  arrange(desc(count))

vict_pacific <- vict_by_area |>
  filter(`AREA NAME` == "Pacific") |>
  arrange(desc(count))

# List of victim descent data for each area
area_vict_data <- list(
```

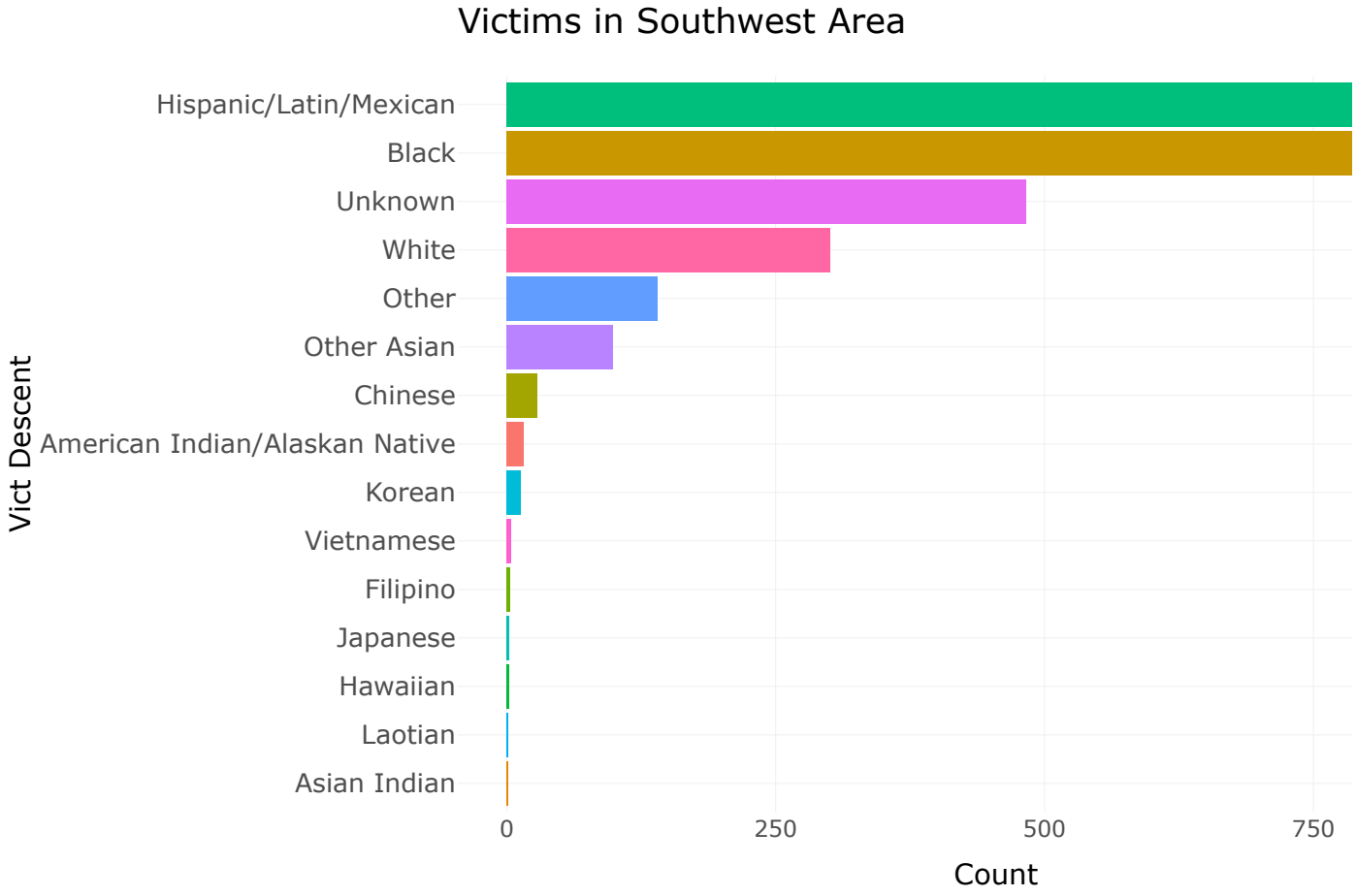
```
  "Central" = vict_Central,
  "Southwest" = vict_Southwest,
  "77th Street" = vict_77st,
  "North Hollywood" = vict_nhollywood,
  "Pacific" = vict_pacific)
```

```
# Function to create ggplotly for a given area
create_victim_plot <- function(data, area_name) {
  ggplotly(
    ggplot(data, aes(
      x = reorder(`Vict Descent`, count),
      y = count,
      fill = `Vict Descent`,
      text = paste(`Vict Descent`, ":", count))) +
      geom_bar(stat = "identity") +
      coord_flip() +
      labs(
        x = "Vict Descent",
        y = "Count",
        title = paste("Victims in", area_name, "Area")) +
      theme_minimal() +
      theme(
        axis.text.y = element_text(size = 10),
        legend.position = "none"),
    tooltip = c("y")) |>
    config(displayModeBar = FALSE)
}

# Apply the function to generate plots for all areas
victim_plots <- lapply(names(area_vict_data), function(area_name) {
  create_victim_plot(area_vict_data[[area_name]], area_name)
})

# Display all plots
victim_plots[[1]]
```

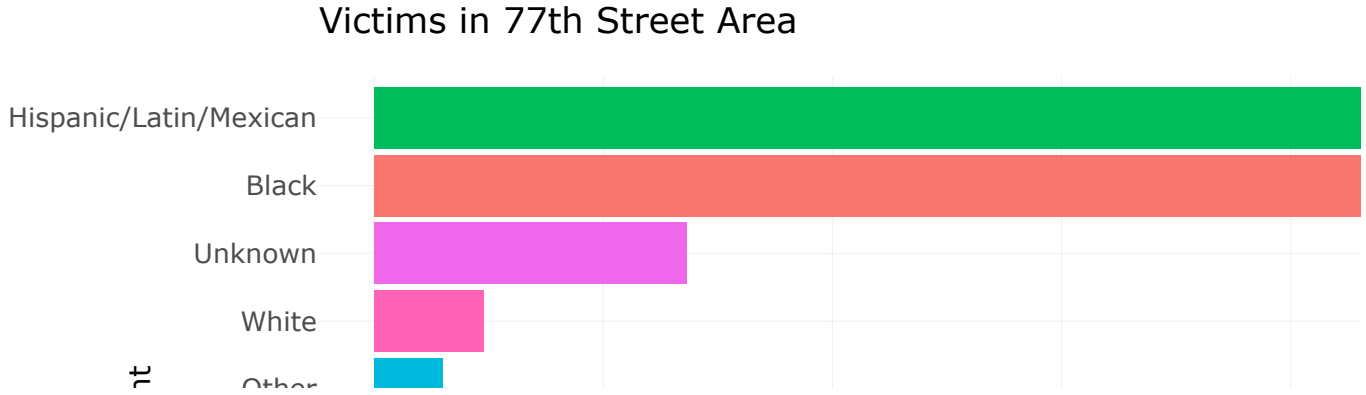


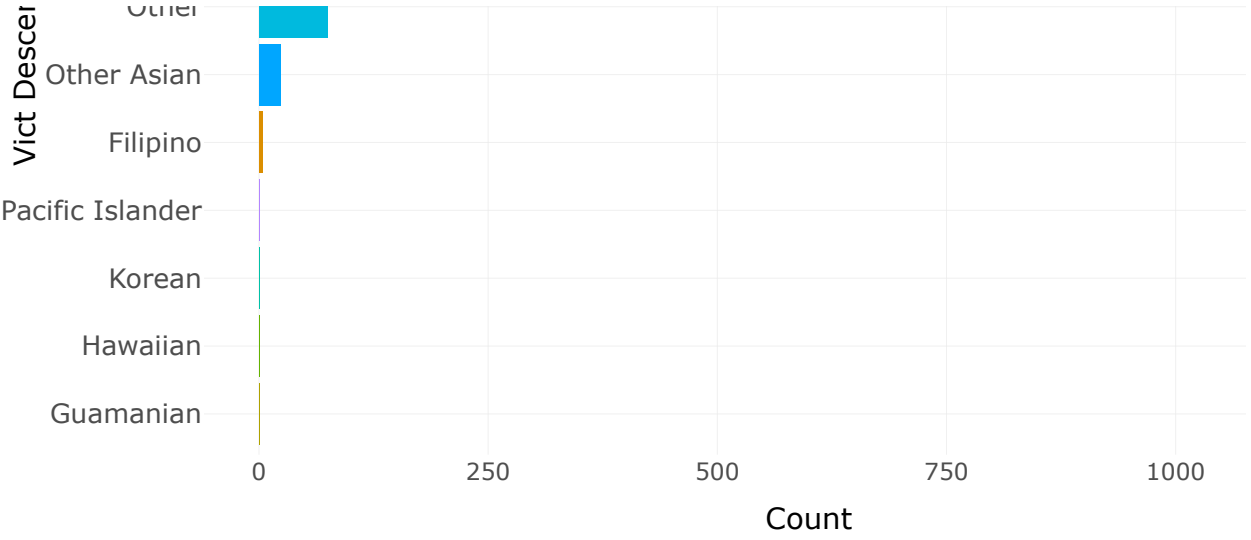



```
victim_plots[[2]]
```

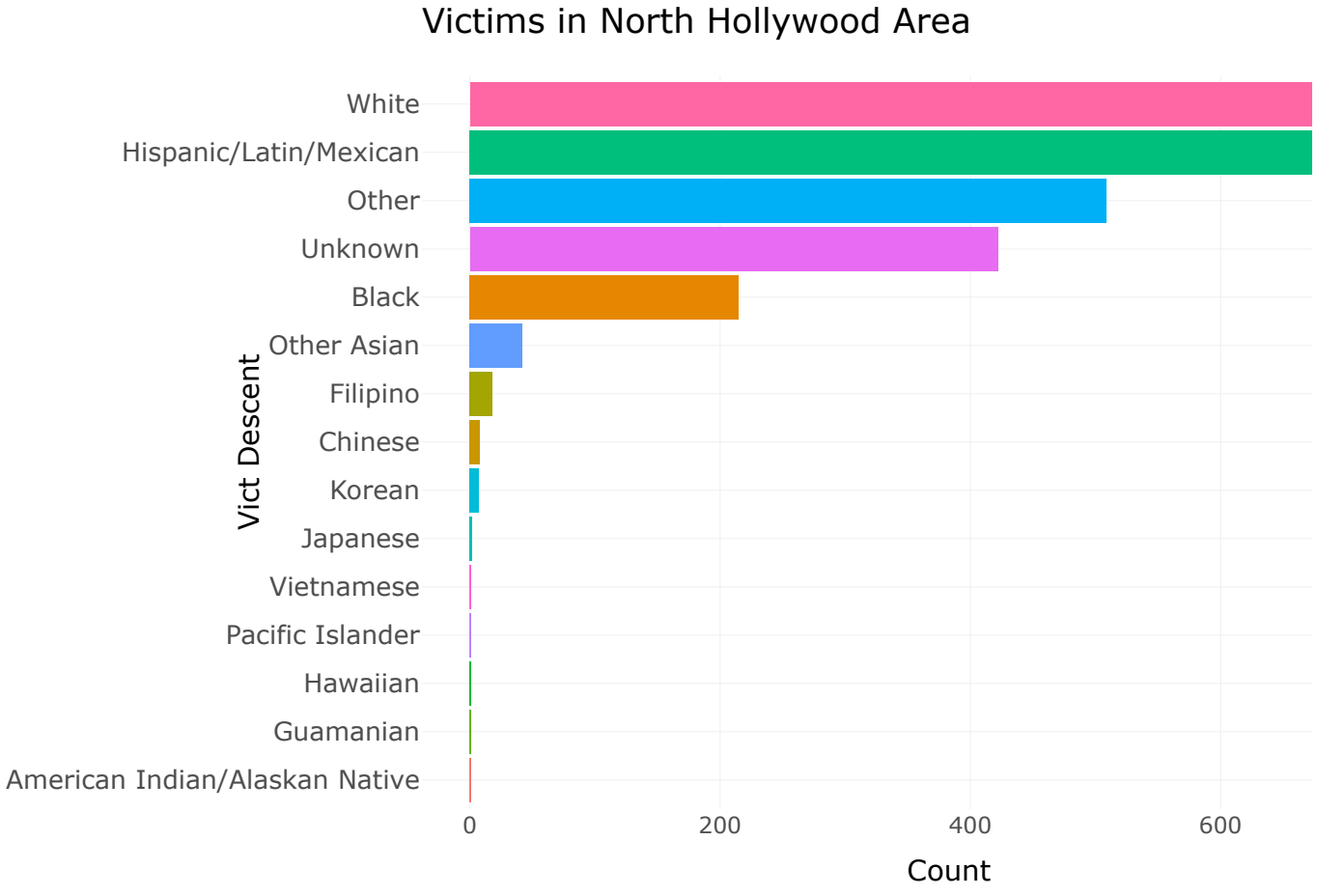



```
victim_plots[[3]]
```

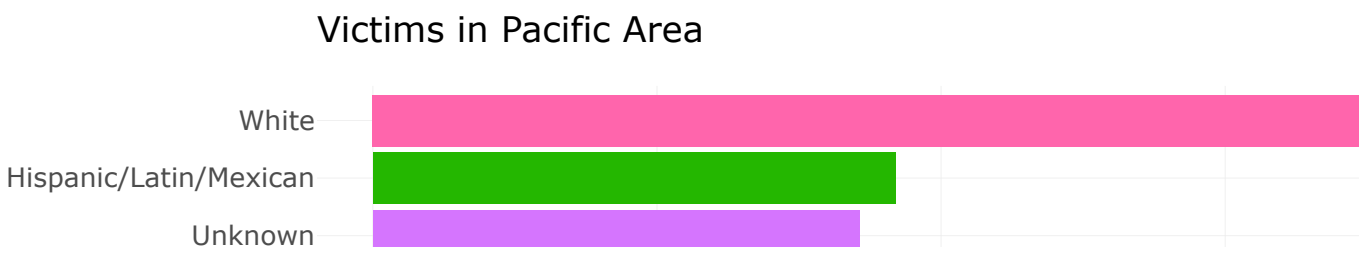


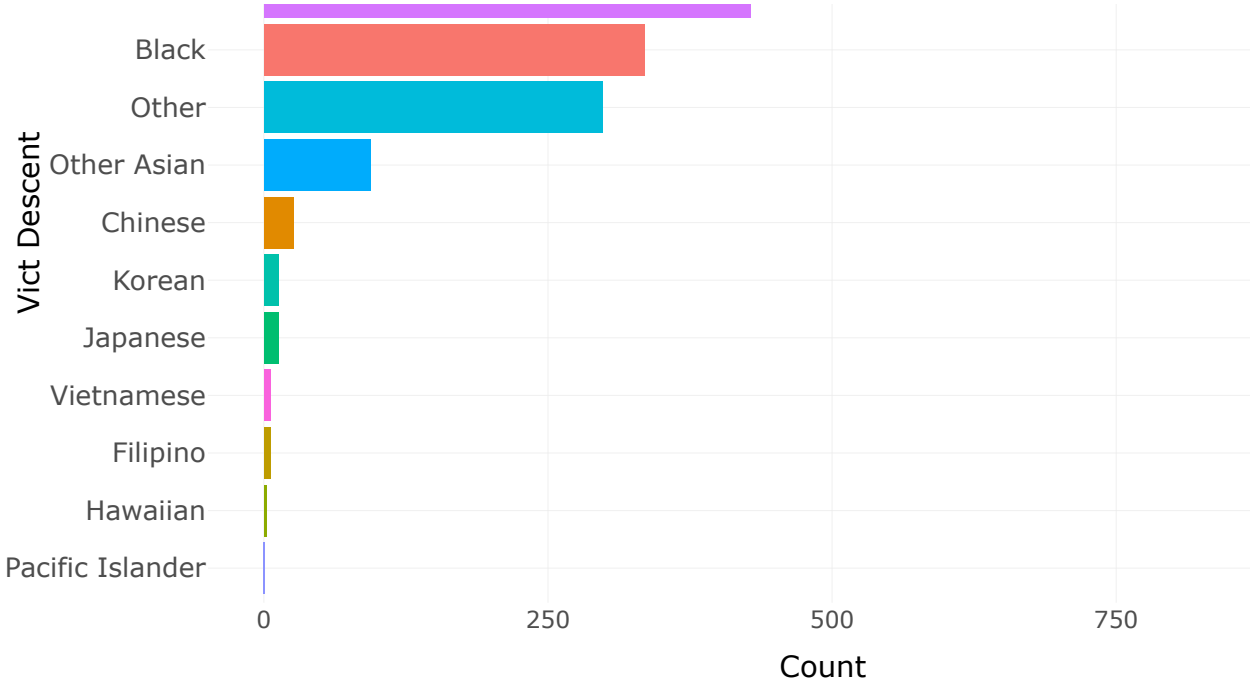



```
victim_plots[[4]]
```




```
victim_plots[[5]]
```





The plots indicate: 1) Central Area: Top Victim Groups: Hispanic/Latin/Mexican, Unknown descent, and Black. Key Insight: The large “Unknown” category indicates a potential gap in data reporting. 2) Southwest Area: Top Victim Groups: Hispanic/Latin/Mexican, Black, and Unknown descent. Key Insight: A significant proportion of victims are from minority communities. 3) 77th Street Area: Top Victim Groups: Hispanic/Latin/Mexican, Black, and Unknown descent. Key Insight: Similar patterns to Southwest, highlighting shared socioeconomic and systemic challenges. 4) North Hollywood: Top Victim Groups: White, Hispanic/Latin/Mexican, and Other. Key Insight: A more diverse spread compared to other areas, with Whites being the most common victim group. 5) Pacific Area: Top Victim Groups: White, Hispanic/Latin/Mexican, and Unknown descent. Key Insight: Similar to North Hollywood, with Whites forming a larger portion of victims.

Therefore, we can conclude that: 1) Hispanic/Latin/Mexican and Black communities are disproportionately affected in Southwest and 77th Street areas. 2) White victims are more prevalent in North Hollywood and Pacific areas. 3) Unknown and NA values account for a large part of the crime data, which need to be handle.

Part 2: Shiny App

Instructions about the App features:

1. Map Section: It highlights the city of Los Angeles with a clearly marked boundary to help users focus on the scope of the project and enhance readability. The number of crimes grouped by areas in the city of Los Angeles will be displayed based on the user-selected date range. Colored circle markers represent these areas, with the color transitioning from yellow (representing the least number of crimes) to red (representing the most). Users can click on a circle marker to view the area name and the total number of crimes in that area during the selected period. The map also allows zooming in and out for more detailed exploration of address across the city.

2. Bar Chart Section: It provides a visual representation of the top N areas with the highest number of crimes during the user-selected date range. This date range is based on the one user selected in

the map section and is free to change inside the map section. Users can select the number of areas to display using a drop-down menu, dynamically adjusting the chart. The bars are arranged horizontally and sorted in descending order, with the area with the most crimes appearing at the top. Hovering over each bar displays the total number of crimes for that area. This section complements the map section by focusing on areas with significant crime activity and enabling easy comparison across different regions.

3. Table section: It displays detailed crime data for the city of Los Angeles during the user-selected date range. Users can filter the table by selecting a specific area from the drop-down menu or view all areas collectively. The table includes columns for the date and time of the crime (formatted in 24-hour military time), a brief description of the crime, the type of location where it occurred, and the severity of the crime, categorized as “Serious” or “Less Serious.” The user can also search keywords, such as specific type of crime (e.g. burglary), to quickly locate the crime information he/she wants. This section allows for an in-depth exploration of individual crime records, complementing the broader visual insights provided in the Map and Bar Chart sections.

Together, the Map, Bar Chart, and Table sections will provide a comprehensive and interactive exploration of crime data in the city of Los Angeles for 2024 Janurary to April. The Map offers a spatial overview, visually highlighting crime distribution across areas and allowing users to focus on specific regions by interacting with color-coded markers. The Bar Chart complements this with a ranked comparison of the top N areas with the most crimes, enabling users to identify high-crime regions at a glance. The Table provides details for each crime, including its time, location, description, and severity, allowing for in-depth insights.

Writeup

1. Exploratory Data Analysis: The “cadetblue” color was selected for the bars in the chart as it is a relatively neutral color. It helps to make the graph to be visually clean and helps to focus on the crime data. The color also aligns well with the theme of crime. It conveys a sense of calmness and objectivity, which is appropriate for presenting sensitive topics like crime data. The function of ggplotly() is used so that we can hover over elements to see detailed information, such as exact crime counts. But its toolbar is disabled for simplicity in visualization.

2. Shiny App: To avoid complexity and ensure the app always reflects the latest complete dataset, we chose to reload the data here. Other than the familiar libraries, “new” libraries are included to achieve our goals for the app. “Leaflet” is used to create the crime map, and “plotly” helps turn the static bar chart into an interactive visualization. A GeoJSON file (i.e., City_Boundary.geojson) is loaded to help define the boundaries of the city of Los Angeles. We asked ChatGPT how to mark the specific area of the city of Los Angeles. Inside the server, the data are first filtered for later use, as all three sections (map, bar chart, and table) generate visualization results based on the same user-selected date. In the map section, the city of Los Angeles is marked in blue with the crime data marked in circles. The total number of crime data is calculated, grouped by area name. We failed to find the complete spatial information (e.g., longitudes, latitudes) for the council districts of Los Angeles, so we decided to use the mean longitudes and latitudes of the same area name to mark the crime circles. In this way, the crime circle markers are representative as they mark the

centers where the crime occurred in a specific area. A yellow-to-red color palette is used so that the dangerous levels of areas in Los Angeles can be viewed in the most straightforward way. In the bar chart section, the drop-down menu for the top N is set to 3 or the total number of available areas (whichever is smaller) by default. A horizontal bar chart is created using ggplot(), and the ggplotly() function makes the bar chart interactive such that users can hover over the bars to view statistics. In the table section, the area names in the drop-down menu are in alphabetical order. The columns are renamed, the time is reformatted, and original data are transformed for better readability. We asked ChatGPT for help when customizing the numbers for font sizes, widths, and opacities. Instructions about how users can interact within the map and bar chart sections are included and colored in gray.

Discussion & Conclusion

This project demonstrates how crime data can be effectively visualized and explored to uncover meaningful insights, while also enabling users to interact with the data in a way that highlights key patterns and disparities. The analysis, complemented by the Shiny app, reveals several important findings based on temporal, geographic, demographic, and crime-type trends in Los Angeles.

1. Temporal Patterns: The higher crime rate in January aligns with the holiday season's impact, where increased public activity, financial strain, and alcohol consumption play a role. For example, post-holiday shopping and gatherings provide more opportunities for crimes such as theft and burglary, while celebrations like New Year's Eve may contribute to increased violent crimes or disorderly behavior. The analysis further shows that Fridays experience the highest number of crimes compared to other weekdays, likely due to increased social activities and alcohol consumption on weekends. These temporal patterns suggest a need for targeted law enforcement and public awareness campaigns during high-risk periods.

2. Geographic Disparities: Certain areas, such as Central, Southwest, and 77th Street, consistently show higher crime rates, likely influenced by factors such as socioeconomic disparities, population density, and urban infrastructure. Central Los Angeles, for instance, sees elevated property crimes like trespassing and vehicle theft, while Southwest and 77th Street experience significant violent crimes, including assaults and domestic violence. Conversely, areas like Foothill, Hollenbeck, and Northeast report lower crime rates, which may indicate better community engagement or differing local conditions. These findings highlight the importance of allocating resources strategically, focusing law enforcement and community programs in high-crime areas.

3. Types of Crimes: The analysis reveals that vehicle theft is the most common crime across all areas, emphasizing vulnerabilities in both public and private spaces. Property crimes, including burglary and trespassing, are also frequent, particularly in areas like North Hollywood and the Pacific region. High-impact crimes such as assault with a deadly weapon and domestic violence remain significant concerns, reflecting ongoing challenges in personal safety and community welfare. Specialized thefts, such as shoplifting and identity theft, highlight the need for targeted prevention strategies, including public awareness campaigns and improved security measures.

4. Demographic Disparities: The analysis of victim demographics shows that Hispanic/Latin/Mexican and Black communities are disproportionately affected in areas like Southwest and 77th Street, underscoring the influence of systemic inequalities and socioeconomic challenges. Areas like North Hollywood and Pacific have a more diverse demographic profile, with White victims forming a larger proportion. Notably, the high percentage of victims with “Unknown” descent across all areas indicates potential gaps in data reporting, which could limit the accuracy of demographic analyses.

Recommendations: - Community Programs: Initiatives focusing on education, job training, and youth engagement could address the root causes of crime. For example, introducing mentorship programs in Southwest Los Angeles may help reduce exposure to criminal activities. - Targeted Policing: Deploying additional resources, such as patrols and surveillance, in high-crime areas like the 77th Street region may control crime effectively. - Enhance Public Awareness: Educating residents about crime trends and prevention techniques, such as securing vehicles and identifying early warning signs of domestic violence, can empower communities to contribute to public safety.

Conclusion: By integrating the EDA findings with the Shiny app, our project can serve as a platform for users to explore crime data and understanding its dynamics. The Map, Bar Chart, and Table sections work together to provide a comprehensive and interactive analysis of crime in Los Angeles, bridging spatial, statistical, and contextual details. Users can identify trends, compare regions, and explore specific incidents, making the app a valuable tool for residents, travelers, and policymakers.

Implications

While this project focuses on Los Angeles, the methods and tools developed can be adapted for other cities and datasets. Expanding the temporal range, incorporating additional demographic and socioeconomic data, or integrating predictive analytics could further enhance crime analysis capabilities. Policymakers and law enforcement agencies could use these insights to design data-driven strategies, optimize resource allocation, and engage communities in crime prevention efforts.

Ultimately, our project shows the value of making crime data accessible and interactive. By providing users with intuitive tool and detailed insights, we aim to advocate a safer community and enhance the understanding of crime dynamics.

```
library(shiny)
library(leaflet)
library(readxl)
library(tidyverse)
library(sf)
library(DT)
library(plotly)

# Load and pre-process the crime data
crime_data <- read_excel("crime2024.xlsx") |>
  mutate(
    `DATE OCC` = as.Date(`DATE OCC`, format = "%m/%d/%Y %I:%M:%S %p")
  )

# Load LA boundary GeoJSON
la_boundary <- st_read("City_Boundary.geojson")


# User Interface
ui <- fluidPage(
  # Title
  div(
    style = "text-align: center; font-weight: bold; font-size: 24px; margin-bottom:
20px;",
    "2024 JAN to APR Crime Data in the city of Los Angeles"
  ),
  # Tab bar with Map, Bar Chart, and Table options
  tabsetPanel(
    tabPanel("Map",
             fluidRow(
               column(
                 width = 6,
                 dateRangeInput(
                   "dateRange",
                   "Select Date Range:",
                   start = min(crime_data$`DATE OCC`),
                   end = max(crime_data$`DATE OCC`),
                   min = min(crime_data$`DATE OCC`),
                   max = max(crime_data$`DATE OCC`),
                   format = "yyyy-mm-dd",
                   width = "100%"
                 )
               )
             ),
             fluidRow(
               column(
                 width = 12,
                 div(
                   style = "margin-top: 30px; color: gray;",
                   p("Click on the colored circle marker to view the area name and the
total number of crimes
                     in that area within the selected period. The color of the circle
markers transitions from yellow to red
                     is based on the total number of crimes where red represents the most
and yellow represents the least.
                     You can zoom in or out on the map for more detailed map
exploration.")
                 )
               )
             ),
             fluidRow(
               column(
                 width = 12,
```

```
                leafletOutput("crimeMap", height = "600px")
              )
            )
    ),
    tabPanel("Bar Chart",
             fluidRow(
               column(
                 width = 6,
                 uiOutput("topNSelector")
               )
             ),
             fluidRow(
               column(
                 width = 12,
                 div(
                   style = "margin-top: 30px; color: gray;",
                   p("Hover over the bars to view the total number of crimes in specific
area within
                     the selected period in Map section.")
                 )
               )
             ),
             fluidRow(
               column(
                 width = 12,
                 plotlyOutput("crimeBarChart", height = "500px")
               )
             )
    ),
    tabPanel("Table",
             fluidRow(
               column(
                 width = 6,
                 selectInput(
                   "areaFilter",
                   "Select Area:",
                   choices = c("All", sort(unique(crime_data$`AREA NAME`))),
                   selected = "All"
                 )
               )
             ),
             fluidRow(
               column(
                 width = 12,
                 dataTableOutput("crimeTable")
               )
             )
    )
  )
)


# Server Logic
server <- function(input, output, session) {
  # Filter data based on the selected date
  filteredData <- reactive({
    crime_data |>
      filter(`DATE OCC` >= input$dateRange[1], `DATE OCC` <= input$dateRange[2])
  })

  # Top N selector setup
  output$topNSelector <- renderUI({
    unique_areas <- filteredData() |>
      distinct(`AREA NAME`) |>
```

```
      nrow()
    selectInput(
      "topN",
      "Select top n areas:",
      choices = seq(1, unique_areas),
      selected = min(3, unique_areas) # Default to 3 or total number of areas, whichever
is smaller
    )
  })

  # Map section generating
  output$crimeMap <- renderLeaflet({
    data <- filteredData() |>
      group_by(`AREA NAME`) |>
      summarize(
        Total_Crimes = n(),
        LAT = mean(LAT, na.rm = TRUE),
        LON = mean(LON, na.rm = TRUE)
      )
    # Define a yellow-to-red color palette for crime counts
    crime_palette <- colorNumeric(
      palette = c("yellow", "red"),
      domain = data$Total_Crimes
    )
    leaflet(options = leafletOptions(zoomControl = FALSE)) |> # Exclude the zoom-in zoom-
out tab
      addTiles() |>
      # Mark the area space of the city of Los Angeles
      addPolygons(
        data = la_boundary,
        color = "blue",
        weight = 2,
        fillColor = "blue",
        fillOpacity = 0.1,
        group = "LA Boundary"
      ) |>
      # Mark the central place of each crime area with red dot
      addCircleMarkers(
        data = data,
        lat = ~LAT,
        lng = ~LON,
        radius = 10,
        color = ~crime_palette(Total_Crimes), # Apply the defined yellow-to-red color
palette
        fillOpacity = 0.8,
        popup = ~paste0(
          "<b>Area:</b> ", `AREA NAME`, "<br>",
          "<b>Total Crimes:</b> ", Total_Crimes
        ),
        group = "Crimes"
      )
  })

  # Bar chart section generating
  output$crimeBarChart <- renderPlotly({
    req(input$topN)  # Ensure user input for N is available
    data <- filteredData() |>
      group_by(`AREA NAME`) |>
      summarize(Total_Crimes = n()) |>
      arrange(desc(Total_Crimes)) |>
      head(as.numeric(input$topN))
    plot1 <- ggplot(data, aes(x = reorder(`AREA NAME`, Total_Crimes), y = Total_Crimes)) +
      geom_bar(stat = "identity", fill = "cadetblue") +
      coord_flip() +
```

```
      labs(x = "Area Name", y = "Total Number of Crimes", title = paste("Top", input$topN,
"Areas with Most Crimes")) +
      theme_minimal()
    # Show total number of crimes when hovering over
    ggplotly(plot1, tooltip = "y") |>
      config(displayModeBar = FALSE)  # Disable the toolbar
  })

  # Table section generating
  output$crimeTable <- renderDataTable({
    data <- filteredData()
    # Filter data based on selected area
    if (input$areaFilter != "All") {
      data <- data |> filter(`AREA NAME` == input$areaFilter)
    }
    # Process how the data and columns are displayed for better readability
    data |>
      mutate(
        `TIME OCC` = paste0(substr(`TIME OCC`, 1, 2), ":", substr(`TIME OCC`, 3, 4)),
        `Crime Severity` = case_when(
          `Part 1-2` == 1 ~ "Serious",
          `Part 1-2` == 2 ~ "Less Serious",
          TRUE ~ NA_character_
        )
      ) |>
      select(
        `Date Occurred` = `DATE OCC`,
        `Time Occurred (in 24 hour military time)` = `TIME OCC`,
        `Crime Brief Description` = `Crm Cd Desc`,
        `Type of Location Occurred` = `Premis Desc`,
        `Crime Severity`
      )
  })
}

shinyApp(ui = ui, server = server)
```

http://127.0.0.1:7478 | Open in Browser | Publish

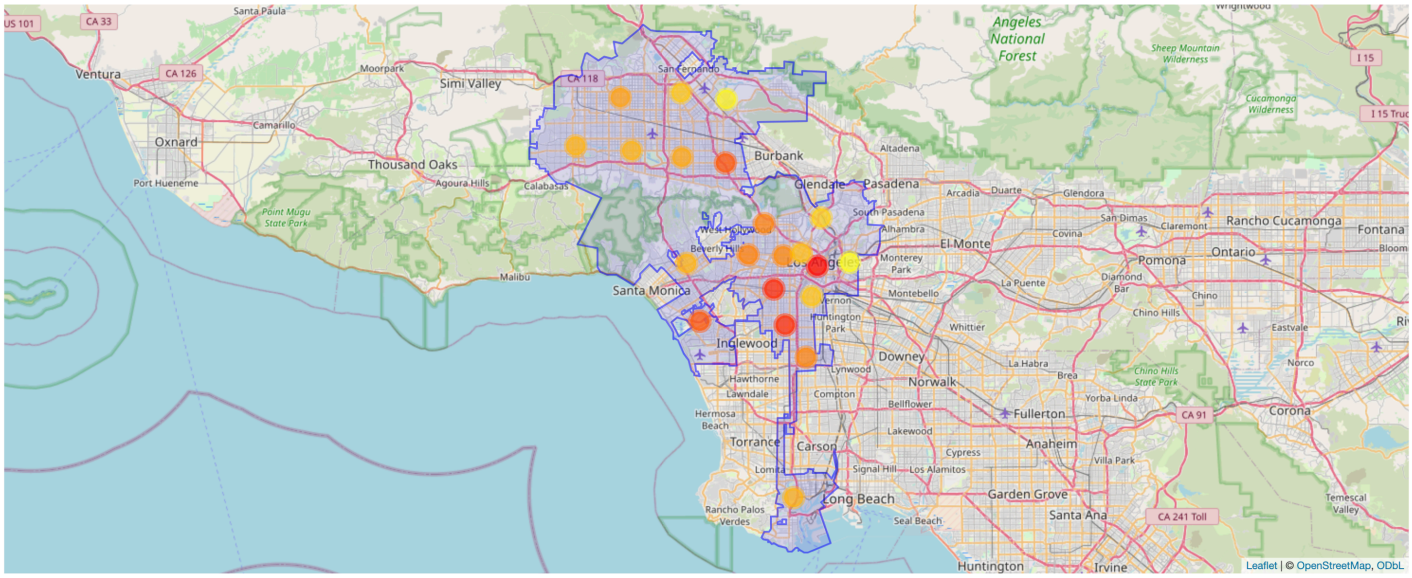
2024 JAN to APR Crime Data in the city of Los Angeles

Map | Bar Chart | Table

Select Date Range:

2024-01-18 to 2024-04-15

Click on the colored circle marker to view the area name and the total number of crimes in that area within the selected period. The color of the circle markers transitions from yellow to red is based on the total number of crimes where red represents the most and yellow represents the least. You can zoom in or out on the map for more detailed map exploration.



http://127.0.0.1:7478 | Open in Browser | Publish

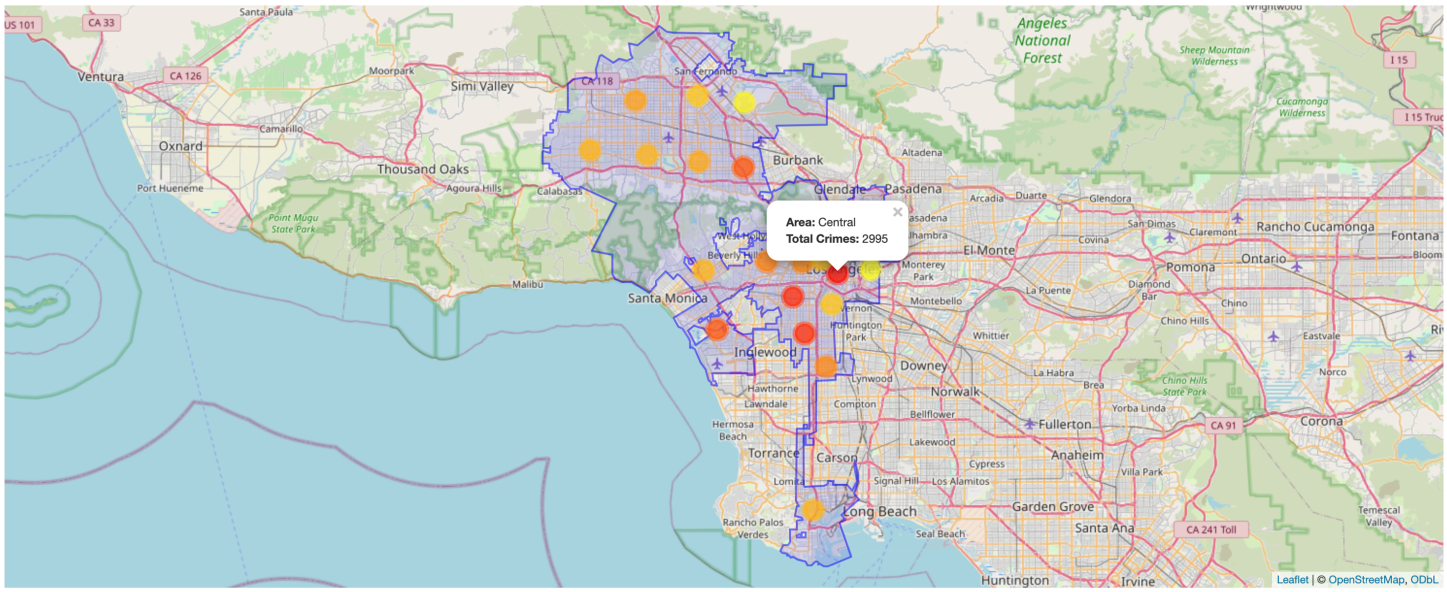
2024 JAN to APR Crime Data in the city of Los Angeles

Map | Bar Chart | Table

Select Date Range:

2024-01-18 to 2024-04-15

Click on the colored circle marker to view the area name and the total number of crimes in that area within the selected period. The color of the circle markers transitions from yellow to red is based on the total number of crimes where red represents the most and yellow represents the least. You can zoom in or out on the map for more detailed map exploration.



http://127.0.0.1:7478 | Open in Browser | Publish

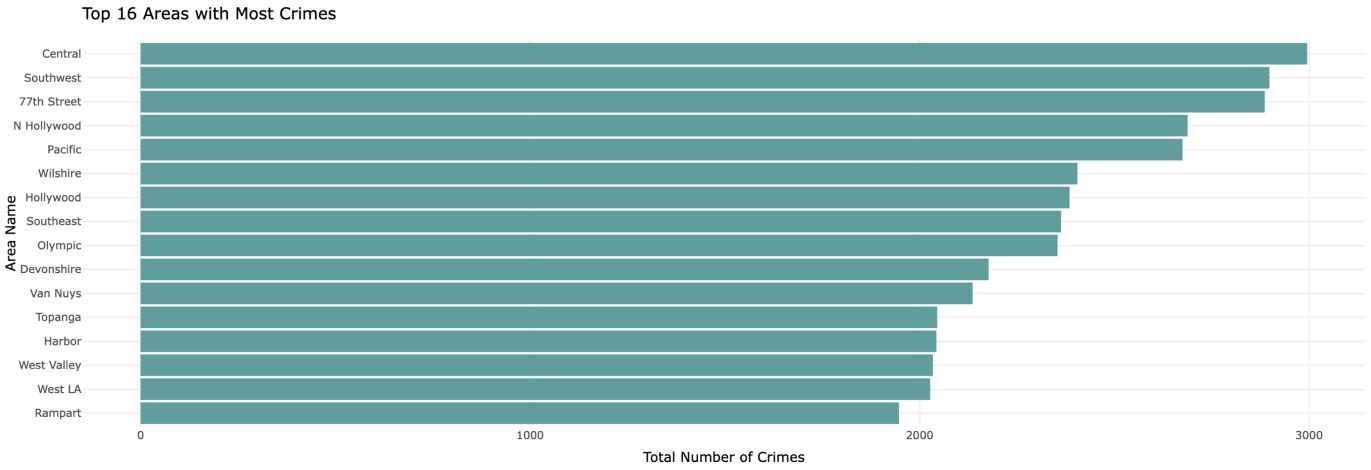
2024 JAN to APR Crime Data in the city of Los Angeles

Map | Bar Chart | Table

Select top n areas:

16

Hover over the bars to view the total number of crimes in specific area within the selected period in Map section.



http://127.0.0.1:7478 | Open in Browser | Publish

2024 JAN to APR Crime Data in the city of Los Angeles

Map | Bar Chart | Table

Select Area:

77th Street

Show 10 **entries** **Search:**

	Date Occurred	Time Occurred (in 24 hour military time)	Crime Brief Description	Type of Location Occurred	Crime Severity
1	2024-02-25	09:00	CHILD NEGLECT (SEE 300 W.I.C.)	ALLEY	Less Serious
2	2024-03-20	19:15	VEHICLE - STOLEN	STREET	Serious
3	2024-02-19	12:00	VEHICLE - STOLEN	STREET	Serious
4	2024-04-09	13:00	VEHICLE - STOLEN	DRIVEWAY	Serious
5	2024-01-26	21:00	VEHICLE - STOLEN	DRIVEWAY	Serious
6	2024-01-23	04:30	THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)	STREET	Serious
7	2024-03-31	10:00	THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)	VEHICLE, PASSENGER/TRUCK	Serious
8	2024-03-31	11:00	CHILD NEGLECT (SEE 300 W.I.C.)	STREET	Less Serious
9	2024-03-30	05:00	BURGLARY	FURNITURE STORE	Serious
10	2024-03-01	13:00	BUNCO, GRAND THEFT	STREET	Less Serious

Showing 1 to 10 of 2,886 entries

Previous 1 2 3 4 5 … 289 Next