Fatal Police Shootings (2015-present)

Yousra Shleibik, Elijah Alabi & Suleiman Saka

Introduction

This project examines patterns and factors associated with fatal police shootings in the United States using data analysis and statistical modeling techniques. By exploring geographical distributions, demographic patterns, situational characteristics, and temporal trends, this research aims to identify key factors that influence these tragic events.The analysis begins with thorough data cleaning and preprocessing, followed by an exploratory data analysis that visualizes patterns of our data distribution. Advanced statistical modeling techniques, including hierarchical clustering, Geospatial and survival analysis, are then applied to uncover deeper insights into departmental patterns and temporal dynamics.By combining descriptive statistics with this research contributes to our understanding of the complex factors surrounding police use of lethal force. This work represents an important step toward data-informed approaches to addressing one of the most challenging issues in contemporary law enforcement and public safety.

Data source and Definitions

The dataset documents all fatal encounters involving law enforcement in the United States since 2015. It includes detailed records of thousands of incidents, with variables such as demographic information, incident details, and law enforcement factors.

Source: *https://github.com/washingtonpost/data-police-shootings/blob/master/v2/README.md or https://www.washingtonpost.com/graphics/investigations/police-shootings-database/*

Key Variables*:*

- **ids**: Unique identifier for each incident
- **date**: Date of the incident
- **age**: Age of the individual involved
- **gender**: age of the individual involved
- **race**: race of the individual involved
- **race_source**: Origin of the racial classification information
- **state**: State where the incident occurred
- **city**: City where the incident occurred
- **county**: County where the incident occurred
- **latitude**: Latitude coordinate of the incident location
- **longitude**: Longitude coordinate of the incident location
- **threat_type**: Type of threat perceived by officers
- **armed_with**: Weapon or item the individual was armed with
- **body_camera**: Indicator for body camera usage
- **flee_status**: Indicator for fleeing behavior
- **was_mental_illness_related**: Indicator for mental illness involvement

Data Cleaning and Preprocessing

This section outlines the essential data cleaning procedures applied to the fatal shootings dataset. The process involved handling missing values through multiple approaches: replacing empty cells in character columns with "not_available" and numeric columns with a placeholder value (99999), then using random sampling from existing values for character columns and mean imputation for numeric columns.

```r
df <- read.csv("fatal_shootings.csv")

# Replace empty cells in character columns with "not_available"
df_cleaned <- df %>%
  mutate(across(
    where(is.character),
    ~ ifelse(. == "" | is.na(.), "not_available", .)
  ))

# Replace empty cells in numeric columns with 99999
df_cleaned <- df_cleaned %>%
  mutate(across(
    where(is.numeric),
    ~ replace_na(., 99999)
  ))

# Replace "not_available" with NA in all character columns
df_cleaned <- df_cleaned %>%
  mutate(across(
    where(is.character),
    ~ na_if(., "not_available")
  ))

# Count missing values (NA and 99999) for each column
missing_summary <- df_cleaned %>%
  summarise(across(
    everything(),
    ~ sum(is.na(.) | . == 99999)
  ))
missing_summary_transposed <- t(missing_summary)
colnames(missing_summary_transposed) <- "Missing_Count"
print(missing_summary_transposed)

##                             Missing_Count
## id                                      0
## date                                    0
## threat_type                            68
## flee_status                          1497
## armed_with                            211
## city                                   74
## county                               4691
## state                                   0
## latitude                             1131
```

```
## longitude                              1132
## location_precision                     8194
## name                                    328
## age                                     382
## gender                                   28
## race                                   1208
## race_source                            7378
## was_mental_illness_related                0
## body_camera                               0
## agency_ids                                1
```

```
set.seed(123)

# Function to replace NA with random values from the same column (for
character columns)
replace_na_with_random <- function(column) {
  non_na_values <- column[!is.na(column)]
  if (length(non_na_values) == 0) {
    return(column)
  }
  column[is.na(column)] <- sample(non_na_values, sum(is.na(column)), replace
= TRUE)
  return(column)
}

# Dataframe for the final cleaned data
df_final <- df_cleaned %>%
  mutate(across(
    where(is.character),
    ~ replace_na_with_random(.)
  )) %>%
  mutate(across(
    where(is.numeric),
    ~ ifelse(. == 99999, mean(.[. != 99999], na.rm = TRUE), .)
  )) %>%
  mutate(age = as.integer(age))

# Re-Check for missing values in all columns
missing_summary <- df_final %>%
  summarise(across(
    everything(),
    ~ sum(is.na(.) | . == 99999 | . == "")
  ))

# Confirmation message
if (all(missing_summary == 0)) {
  print("No missing values found in the dataframe.")
} else {
  print("Missing values found in the dataframe:")
```

```
  print(missing_summary)
}

## [1] "No missing values found in the dataframe."
```

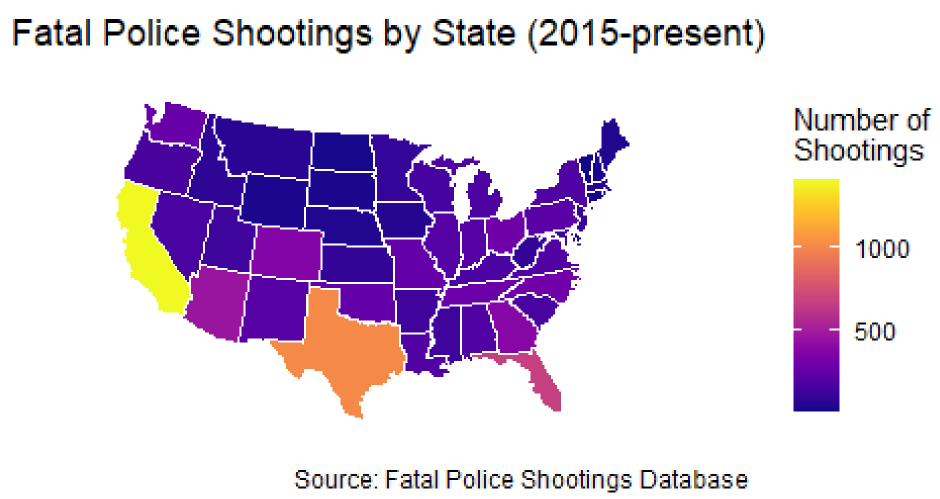
Exploratory Data Analysis (EDA)

This section presents a visual representation of the fatal police shootings dataset. Through a series of visualizations, we examine geographic distribution, demographic patterns, situational characteristics, and temporal trends. This approach enables the identification of patterns and relationships within the data.

1. **Heatmap of shootings by state**: The state-level heatmap reveals significant geographic variation in the frequency of fatal police shootings. States with higher populations such as California, Texas, and Florida show elevated numbers of incidents, but population alone doesn't explain all variations.

```
state_counts <- df_final %>%
  count(state) %>%
  arrange(desc(n))

# State-level map
states_map <- map_data("state")
state_crosswalk <- data.frame(
  state = state.abb,
  state_name = tolower(state.name)
)

# Convert state abbreviations to full names for mapping
state_counts_for_map <- state_counts %>%
  left_join(state_crosswalk, by = "state") %>%
  select(state_name, n)

# map data joined
map_data <- states_map %>%
  left_join(state_counts_for_map, by = c("region" = "state_name"))

# Create the map

ggplot(map_data, aes(long, lat, group = group, fill = n)) +
  geom_polygon(color = "white", linewidth = 0.1) +
  coord_map(projection = "albers", lat0 = 39, lat1 = 45) +
  scale_fill_viridis(option = "plasma", name = "Number of\nShootings",
                     guide = guide_colorbar(title.position = "top")) +
  labs(title = "Fatal Police Shootings by State (2015-present)",
       caption = "Source: Fatal Police Shootings Database") +
  theme_minimal() +
  theme(legend.position = "right",
        axis.text = element_blank(),
```

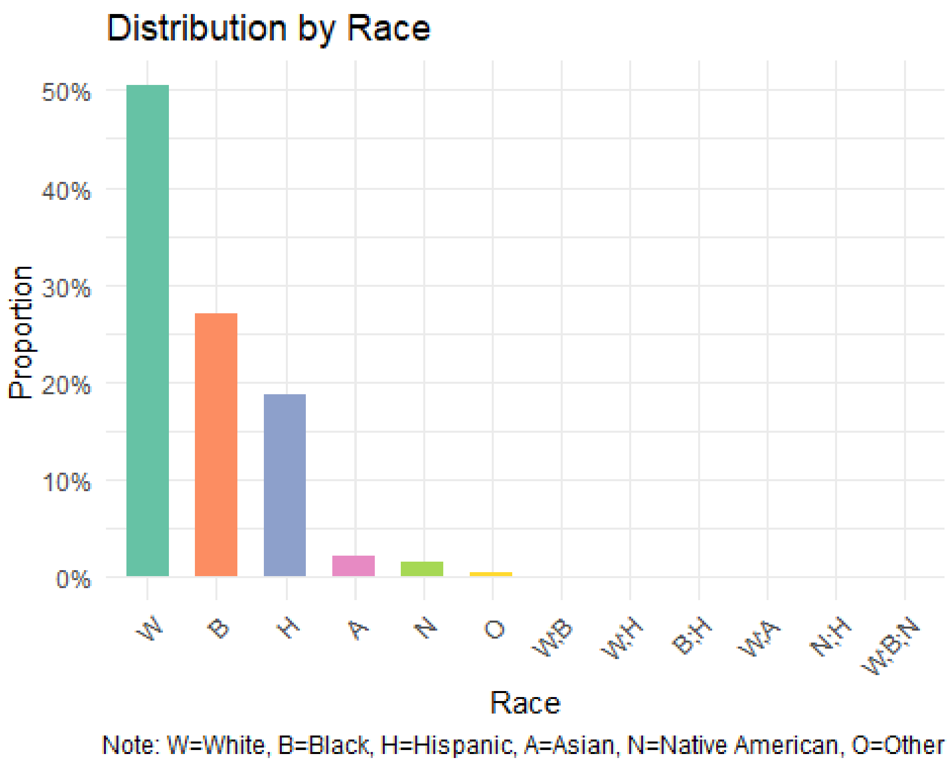
```
        axis.title = element_blank(),
        panel.grid = element_blank())
```



2. **Race Distribution**: This bar chart analyzes the racial distribution of victims in fatal police shootings, showing the proportion of victims by racial category

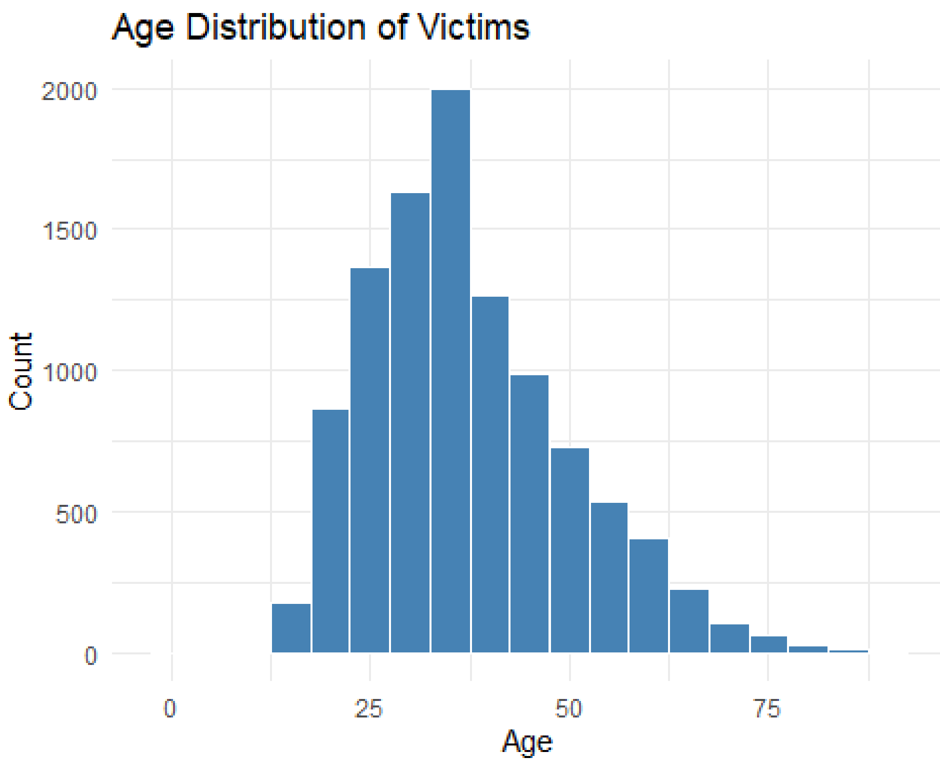
```
race_plot <- df_final %>%
  count(race) %>%
  arrange(desc(n)) %>%
  mutate(proportion = n / sum(n),
         race = factor(race, levels = race[order(n, decreasing = TRUE)])) %>%
  ggplot(aes(x = race, y = proportion, fill = race)) +
  geom_bar(stat = "identity", width = 0.6) +
  scale_fill_brewer(palette = "Set2") +
  labs(title = "Distribution by Race",
       x = "Race",
       y = "Proportion",
       caption = "Note: W=White, B=Black, H=Hispanic, A=Asian, N=Native 
American, O=Other") +
  theme_minimal() +
  theme(
    legend.position = "none",
    axis.text.x = element_text(angle = 45, hjust = 1)
  ) +
  scale_y_continuous(labels = scales::percent)

print(race_plot)
```

```
## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum
for palette Set2 is 8
## Returning the palette you asked for with that many colors
```



3. **Age distribution**: This histogram examines the age distribution of individuals involved in fatal police shootings, highlighting which age groups are most frequently represented among victims and identifying patterns across the age spectrum.

```
age_plot <- ggplot(df_final, aes(x = age)) +
  geom_histogram(binwidth = 5, fill = "steelblue", color = "white") +
  labs(title = "Age Distribution of Victims",
       x = "Age",
       y = "Count") +
  theme_minimal()
print(age_plot)
```

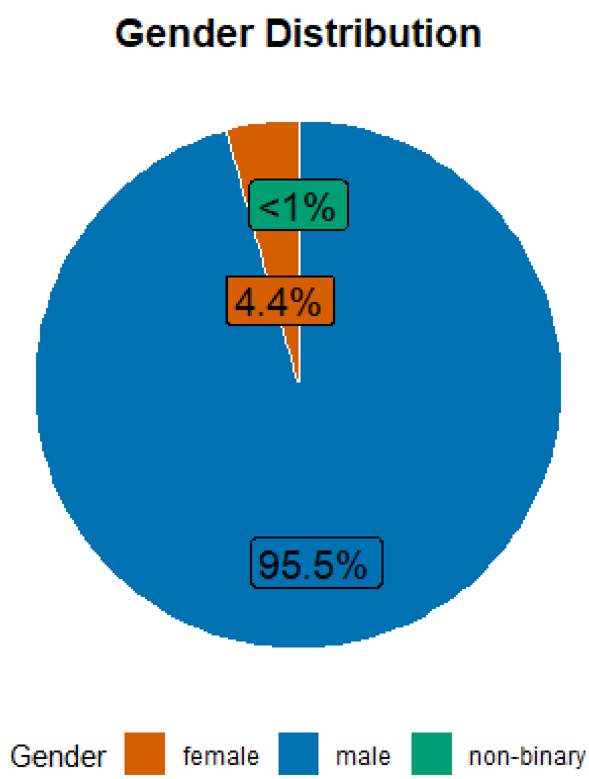



```
age_density <- ggplot(df_final, aes(x = age)) +
  geom_density(fill = "steelblue", alpha = 0.5) +
  labs(title = "Age Distribution Density",
       x = "Age",
       y = "Density") +
  theme_minimal()
```

4. **Gender distribution**: This pie chart illustrates the gender breakdown of victims in fatal police shootings, highlighting the gender disparities in lethal force incidents and providing context for understanding gender-related patterns in policing outcomes.

```
accessible_colors <- c(
  "male" = "#0072B2",       # Blue
  "female" = "#D55E00",     # Orange
  "non-binary" = "#009E73"  # Green
)



gender_plot <- df_final %>%
  count(gender) %>%
  mutate(proportion = n / sum(n),
         label = ifelse(proportion < 0.01, "<1%", scales::percent(proportion,
accuracy = 0.1))) %>%
  ggplot(aes(x = "", y = proportion, fill = gender)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) +
```

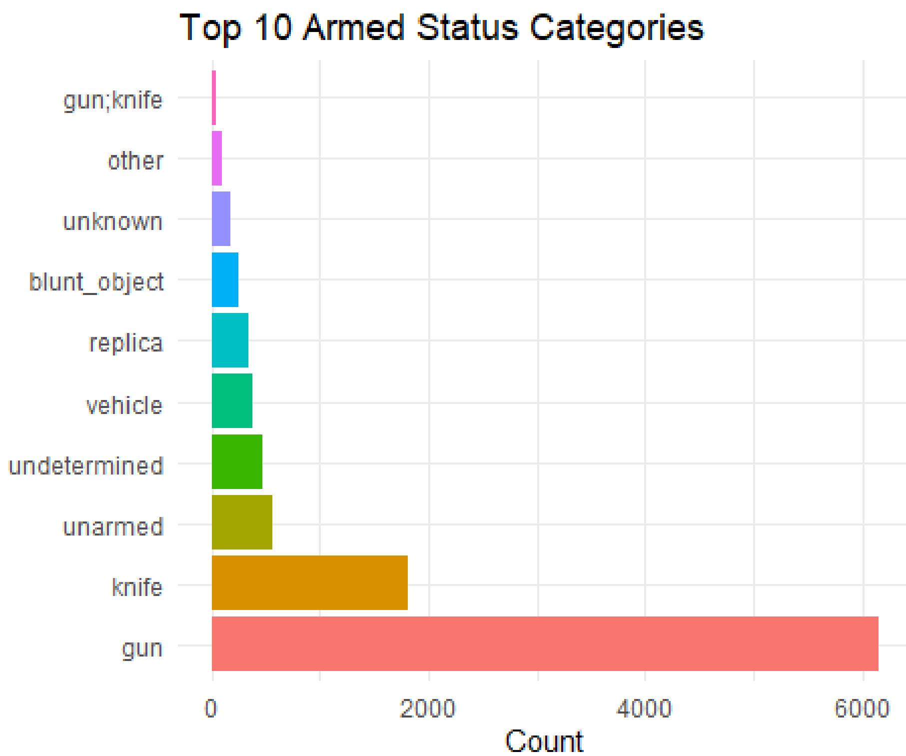
```
  scale_fill_manual(values = accessible_colors) +
  labs(title = "Gender Distribution", fill = "Gender") +
  theme_void() +
  theme(
    legend.position = "bottom",
    plot.title = element_text(hjust = 0.5, face = "bold", size = 14)
  ) +
  geom_label_repel(aes(label = label),
                   position = position_stack(vjust = 0.5),
                   size = 5,
                   color = "black",
                   box.padding = 0.5,
                   direction = "y",
                   show.legend = FALSE)
print(gender_plot)
```



5. **Armed status distribution:** This horizontal bar chart represent the most common types of weapons or items individuals were armed with during fatal police encounters. The visualization provides crucial context about threat levels and situational factors that may influence police decisions to use lethal force.

```
armed_plot <- df_final %>%
  count(armed_with) %>%
  arrange(desc(n)) %>%
  head(10) %>%  # Top 10 categories
  mutate(armed_with = factor(armed_with, levels = armed_with[order(n,
decreasing = TRUE)])) %>%
```

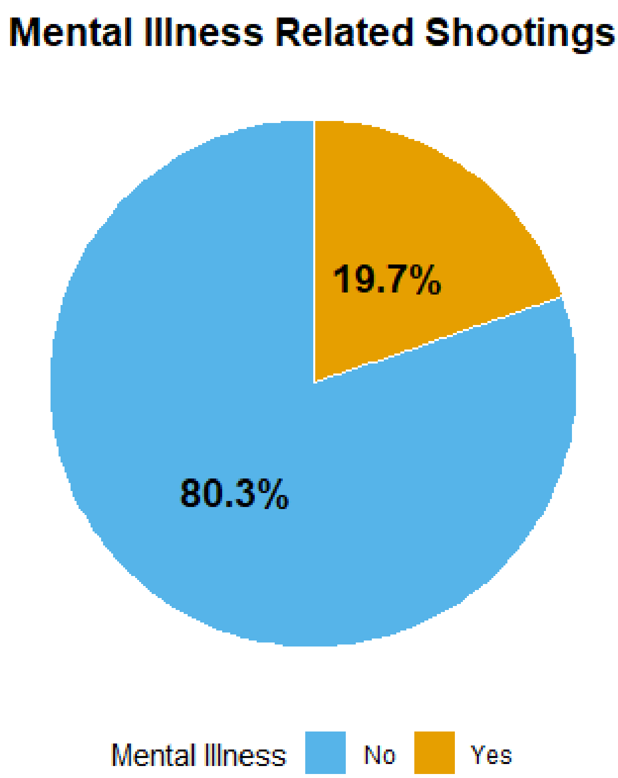
```r
  ggplot(aes(x = armed_with, y = n, fill = armed_with)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(title = "Top 10 Armed Status Categories",
       x = "",
       y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
print(armed_plot)
```



6. **Mental illness related shootings**: This pie chart shows the proportion of fatal police shootings involving individuals with suspected mental illness. The visualization highlights the intersection of mental health issues and lethal police encounters.

```r
accessible_colors <- c("Yes" = "#E69F00",
                       "No" = "#56B4E9")

mental_illness_plot <- df_final %>%
  mutate(was_mental_illness_related = as.logical(was_mental_illness_related))
%>%
  count(was_mental_illness_related) %>%
  mutate(label = ifelse(was_mental_illness_related, "Yes", "No"),
         proportion = n / sum(n),
         display_label = scales::percent(proportion, accuracy = 0.1)) %>%
  ggplot(aes(x = "", y = proportion, fill = label)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) +
```

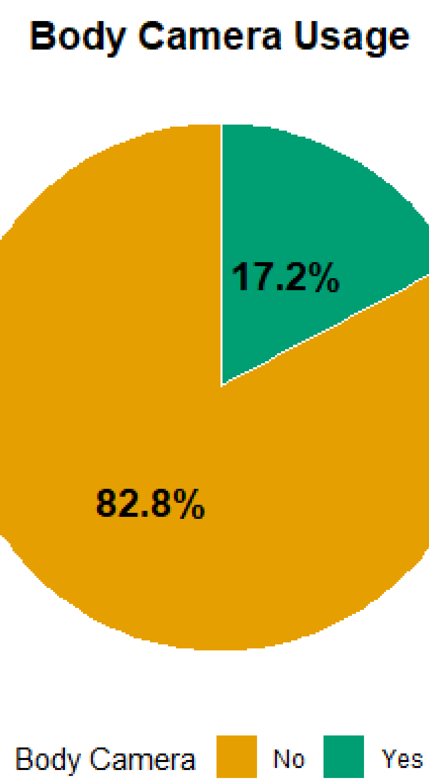
```r
  scale_fill_manual(values = accessible_colors) +
  labs(title = "Mental Illness Related Shootings", fill = "Mental Illness") +
  theme_void() +
  theme(
    legend.position = "bottom",
    plot.title = element_text(hjust = 0.5, face = "bold", size = 14)
  ) +
  geom_text(aes(label = display_label),
            position = position_stack(vjust = 0.5),
            size = 5,
            color = "black",
            fontface = "bold")

print(mental_illness_plot)
```



7. **Body camera usage**: This pie chart illustrates the prevalence of body camera usage during fatal police shootings. The visualization provides insight into police transparency and accountability measures in incidents involving lethal force.

```r
accessible_colors <- c("Yes" = "#009E73",
                       "No" = "#E69F00")

body_camera_plot <- df_final %>%
  mutate(body_camera = as.logical(body_camera)) %>%
  count(body_camera) %>%
  mutate(label = ifelse(body_camera, "Yes", "No"),
         proportion = n / sum(n),
```

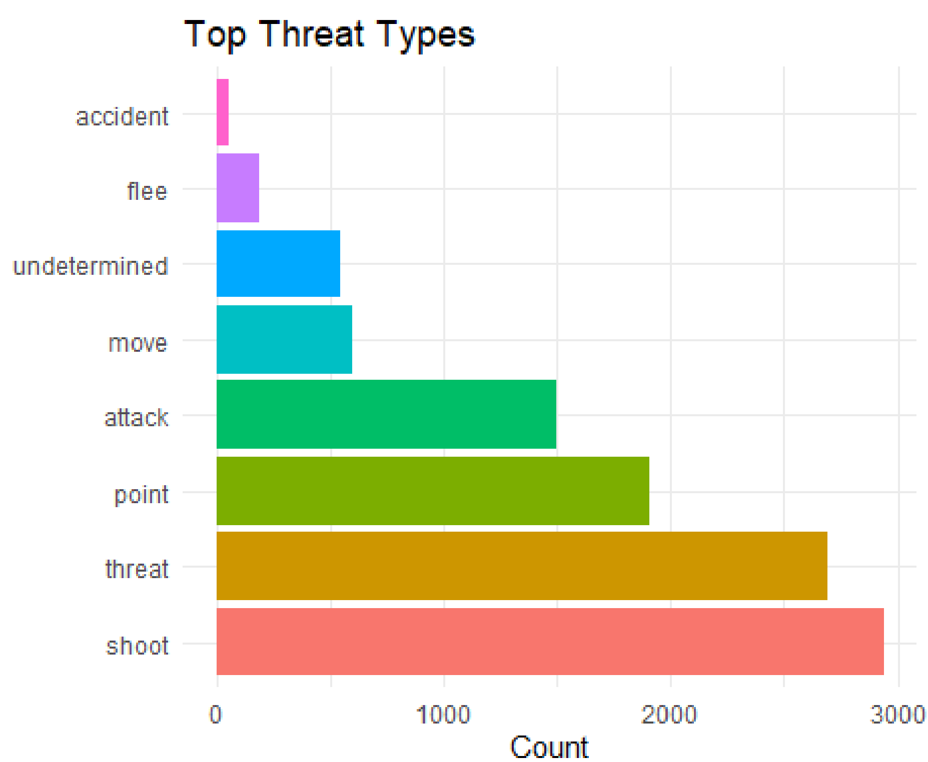
```r
        display_label = scales::percent(proportion, accuracy = 0.1)) %>%
  ggplot(aes(x = "", y = proportion, fill = label)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar("y", start = 0) +
  scale_fill_manual(values = accessible_colors) +
  labs(title = "Body Camera Usage", fill = "Body Camera") +
  theme_void() +
  theme(
    legend.position = "bottom",
    plot.title = element_text(hjust = 0.5, face = "bold", size = 14)
  ) +
  geom_text(aes(label = display_label),
            position = position_stack(vjust = 0.5),
            size = 5,
            color = "black",
            fontface = "bold")

print(body_camera_plot)
```



8. **Threat Type Distribution**: This horizontal bar chart examines the types of threats that officers reported facing during fatal police shooting incidents. The visualization categorizes the perceived threats that preceded the use of lethal force.

```r
threat_plot <- df_final %>%
  count(threat_type) %>%
  arrange(desc(n)) %>%
  head(10) %>%
```

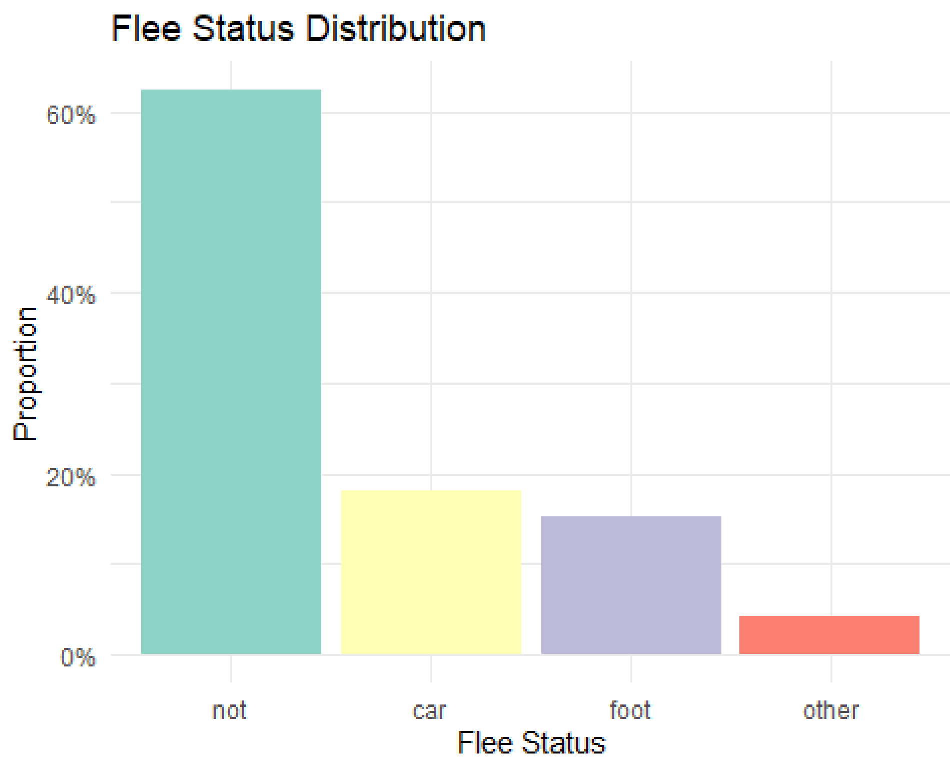
```
  mutate(threat_type = factor(threat_type, levels = threat_type[order(n,
decreasing = TRUE)])) %>%
  ggplot(aes(x = threat_type, y = n, fill = threat_type)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  labs(title = "Top Threat Types",
       x = "",
       y = "Count") +
  theme_minimal() +
  theme(legend.position = "none")
print(threat_plot)
```



9. **Flee status distribution**: The chart shows the distribution of "flee status" among victims of fatal police shootings, indicating whether individuals were attempting to flee from police and by what means when lethal force was used.

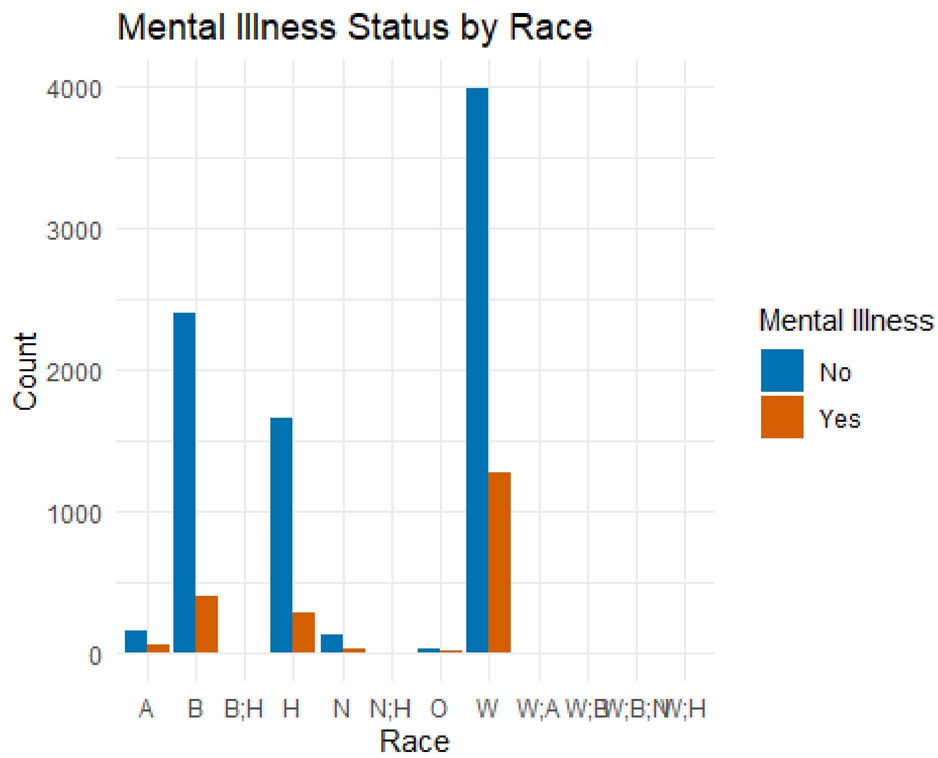
```
flee_plot <- df_final %>%
  count(flee_status) %>%
  mutate(proportion = n / sum(n),
         flee_status = factor(flee_status, levels = flee_status[order(n,
decreasing = TRUE)])) %>%
  ggplot(aes(x = flee_status, y = proportion, fill = flee_status)) +
  geom_bar(stat = "identity") +
  scale_fill_brewer(palette = "Set3") +
  labs(title = "Flee Status Distribution",
       x = "Flee Status",
       y = "Proportion") +
```

```r
  theme_minimal() +
  theme(legend.position = "none") +
  scale_y_continuous(labels = scales::percent)
print(flee_plot)
```



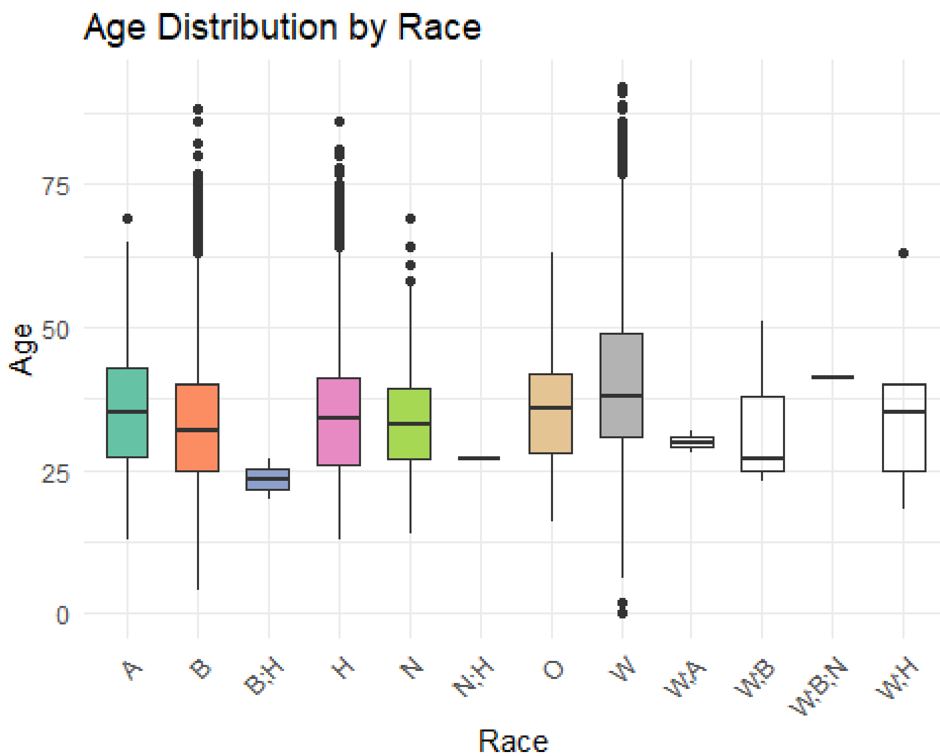
10. **Race and Mental Illness Interaction**: This grouped bar chart examines the intersection of race and mental illness status in fatal police shootings. By showing how mental illness relates to different racial groups.

```r
race_mental_plot <- df_final %>%
  group_by(race, was_mental_illness_related) %>%
  summarise(count = n(), .groups = "drop") %>%
  mutate(mental_illness = ifelse(was_mental_illness_related, "Yes", "No"))
%>%
  group_by(race) %>%
  filter(sum(count) > 0) %>%
  ungroup() %>%
  ggplot(aes(x = race, y = count, fill = mental_illness)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_manual(values = c("Yes" = "#D55E00", "No" = "#0072B2")) +
  labs(title = "Mental Illness Status by Race",
       x = "Race",
       y = "Count",
       fill = "Mental Illness") +
  theme_minimal()

print(race_mental_plot)
```



11. **Age distribution by race**: This boxplot visualization compares age distributions across different racial groups in fatal police shootings. By examining how age patterns vary by race, the chart helps identify potential demographic disparities in police use of lethal force.

```
age_race_plot <- ggplot(df_final, aes(x = race, y = age, fill = race)) +
  geom_boxplot(width = 0.6) +
  scale_fill_brewer(palette = "Set2") +
  labs(title = "Age Distribution by Race",
       x = "Race",
       y = "Age") +
  theme_minimal() +
  theme(
    legend.position = "none",
    axis.text.x = element_text(angle = 45, hjust = 1)
  )

print(age_race_plot)

## Warning in RColorBrewer::brewer.pal(n, pal): n too large, allowed maximum
for palette Set2 is 8
## Returning the palette you asked for with that many colors
```

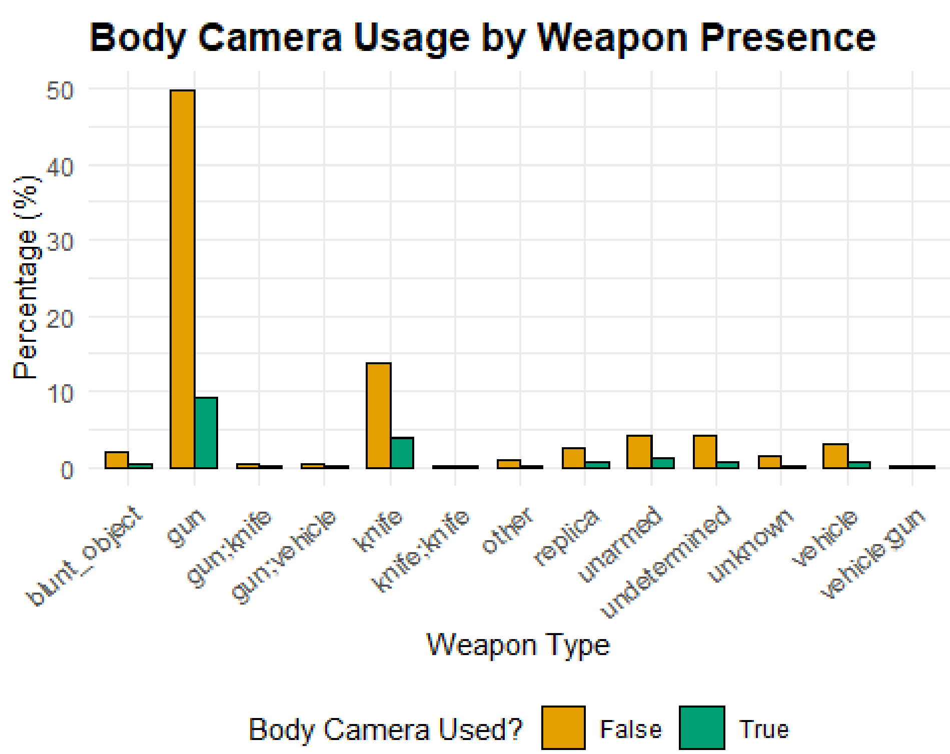


Summary statistics (e.g., mean age, racial proportions) and correlations

```r
# Summary statistics for numerical variables (e.g., age)
summary_stats <- df_final %>%
  summarise(
    Mean_Age = mean(age, na.rm = TRUE),
    Median_Age = median(age, na.rm = TRUE),
    SD_Age = sd(age, na.rm = TRUE),
    Min_Age = min(age, na.rm = TRUE),
    Max_Age = max(age, na.rm = TRUE)
  )
print(summary_stats)

##   Mean_Age Median_Age   SD_Age Min_Age Max_Age
## 1 37.45349         36 12.69518       0      92

# race label mapping
race_mapping <- c(
  "W" = "White",
  "B" = "Black",
  "H" = "Hispanic",
  "A" = "Asian",
  "N" = "Native American",
  "O" = "Other",
  "W;B" = "White; Black",
  "W;H" = "White; Hispanic",
  "B;H" = "Black; Hispanic",
```

```r
  "N;H" = "Native American; Hispanic"
)

# Calculate racial proportions
race_proportions <- df_final %>%
  mutate(
    race = str_replace_all(race, pattern = race_mapping)
  ) %>%
  count(race) %>%
  mutate(Percentage = n / sum(n) * 100) %>%
  arrange(desc(Percentage))

print(race_proportions)

##                          race    n   Percentage
## 1                       White 5261 50.450709628
## 2                       Black 2809 26.937092443
## 3                    Hispanic 1943 18.632527810
## 4                       Asian  206  1.975450710
## 5             Native American  152  1.457614116
## 6                       Other   41  0.393172229
## 7                 White;Black    5  0.047947833
## 8              White;Hispanic    5  0.047947833
## 9              Black;Hispanic    2  0.019179133
## 10                White;Asian    2  0.019179133
## 11   Native American;Hispanic    1  0.009589567
## 12 White;Black;Native American    1  0.009589567

# Gender proportions
gender_proportions <- df_final %>%
  count(gender) %>%
  mutate(Percentage = n / sum(n) * 100)
print(gender_proportions)

##       gender    n  Percentage
## 1     female  462  4.43037975
## 2       male 9961 95.52167242
## 3 non-binary    5  0.04794783

# Calculate proportions

bodycam_weapon_prop <- df_final %>%
  group_by(armed_with, body_camera) %>%
  summarise(n = n(), .groups = "drop") %>%
  mutate(percentage = n / sum(n) * 100)

bodycam_weapon_prop <- bodycam_weapon_prop %>%
  group_by(armed_with) %>%
  filter(sum(n) > 5) %>%
  ungroup()
```

```
colors <- c("True" = "#009E73", "False" = "#E69F00")

ggplot(bodycam_weapon_prop, aes(x = armed_with, y = percentage, fill =
as.factor(body_camera))) +
  geom_bar(stat = "identity", position = "dodge", color = "black", width =
0.7) +
  labs(
    title = "Body Camera Usage by Weapon Presence",
    x = "Weapon Type",
    y = "Percentage (%)",
    fill = "Body Camera Used?"
  ) +
  scale_fill_manual(values = colors) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 40, hjust = 1, size = 10),
    plot.title = element_text(face = "bold", size = 14),
    legend.position = "bottom"
  )
```

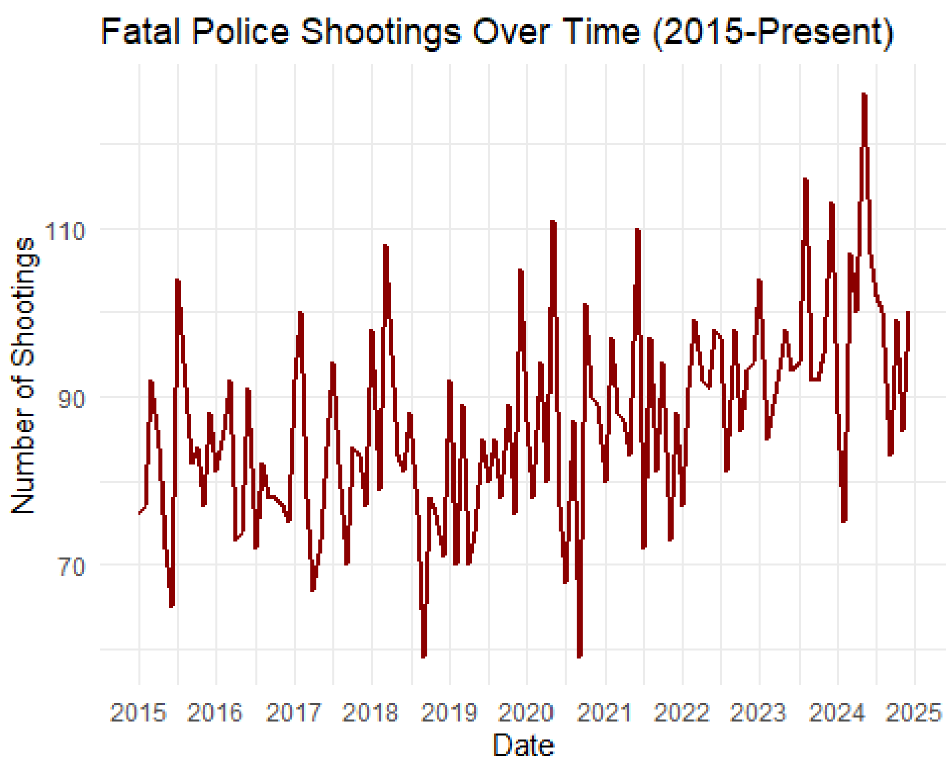


Temporal Trends & Outliers

This section explores how fatal police shootings have changed over time, examining patterns, seasonal variations, and demographic trends. These temporal analyses help identify underlying patterns, outliers, and potential shifts in policing practices or reporting over the study period.

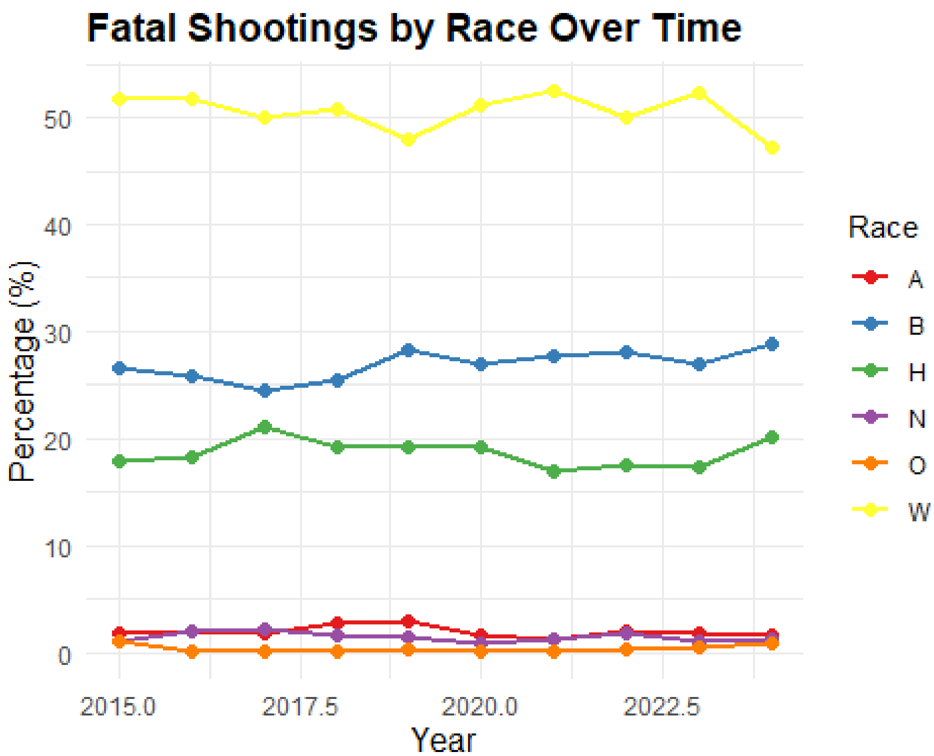
Fatal Police Shootings Over Time: Some seasonal patterns are visible, with slight increases during summer months in certain years.

```r
df_final <- df_final %>%
  mutate(date = as.Date(date)) %>%
  mutate(month_year = floor_date(date, "month"))

# Plot shootings over time
ggplot(df_final, aes(x = month_year)) +
  geom_line(stat = "count", color = "darkred", linewidth = 1) +
  labs(
    title = "Fatal Police Shootings Over Time (2015-Present)",
    x = "Date",
    y = "Number of Shootings"
  ) +
  theme_minimal() +
  scale_x_date(date_breaks = "1 year", date_labels = "%Y")
```



Fatal Shootings by Race Over Time: The visualization of fatal shootings by race over time shows relatively stable proportional representation of different racial groups, with some fluctuations year to year

```
significant_races <- df_final %>%
  group_by(race) %>%
  summarise(total_incidents = n()) %>%
  filter(total_incidents >= 10) %>%
  pull(race)


race_trend <- df_final %>%
  mutate(year = year(date)) %>%
  filter(race %in% significant_races) %>%
  count(year, race) %>%
  group_by(year) %>%
  mutate(percentage = n / sum(n) * 100)


ggplot(race_trend, aes(x = year, y = percentage, color = race)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  labs(
    title = "Fatal Shootings by Race Over Time",
    x = "Year",
    y = "Percentage (%)",
    color = "Race"
  ) +
  scale_color_brewer(palette = "Set1") +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 14),
    axis.title = element_text(size = 12)
  )

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

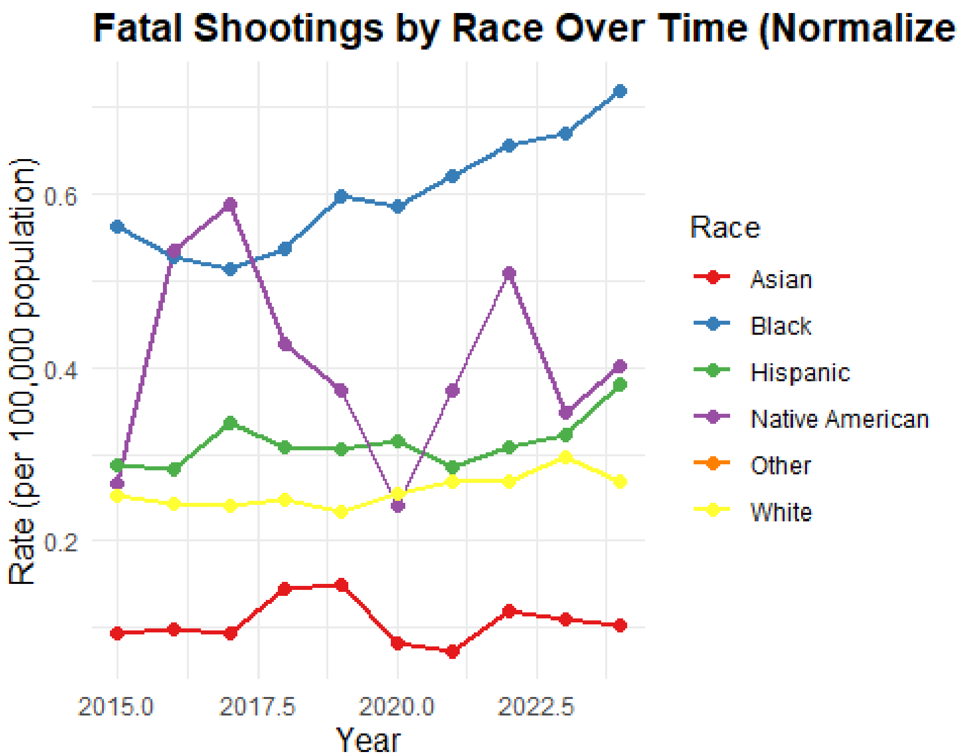


Fatal Shootings by Race Over Time (Normalized by Population): When normalized by population, the disparities become more apparent, with Black and Native American individuals consistently experiencing higher per capita rates of fatal police shootings compared to other racial groups.

```
df_final <- df_final %>%
  mutate(race = case_when(
    race == "W" ~ "White",
    race == "B" ~ "Black",
    race == "H" ~ "Hispanic",
    race == "A" ~ "Asian",
    race == "N" ~ "Native American",
    TRUE        ~ "Other"
  ))


significant_races <- df_final %>%
  group_by(race) %>%
  summarize(total_incidents = n()) %>%
  filter(total_incidents >= 10) %>%
  pull(race)

# Aggregate by race and year
race_trend <- df_final %>%
  mutate(year = year(date)) %>%
  filter(race %in% significant_races) %>%
```

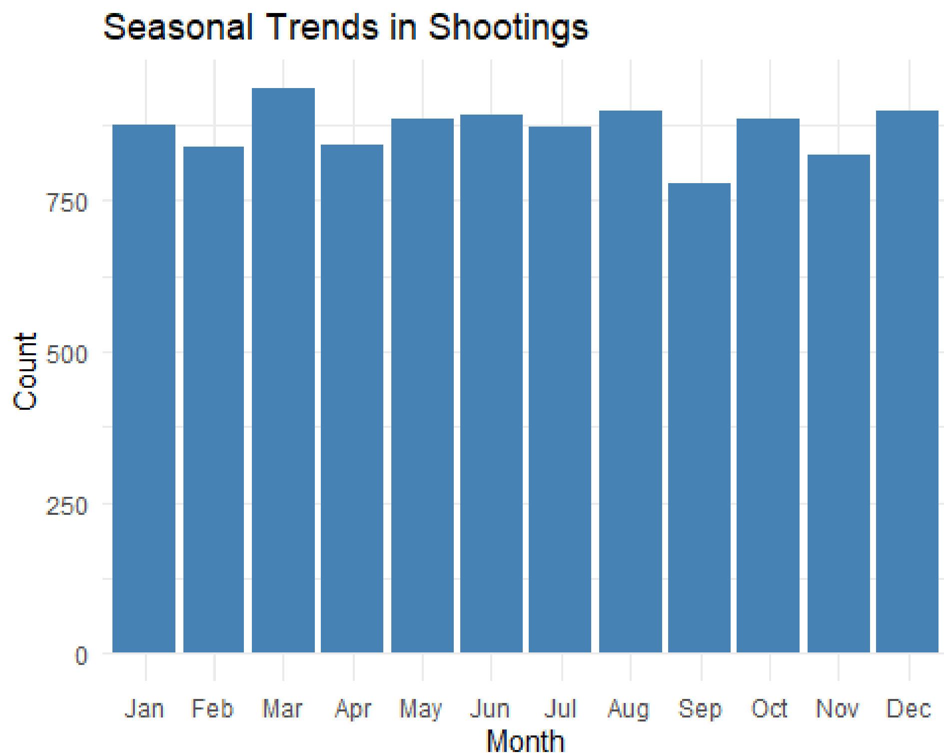
```r
  count(year, race)

# Create a population data frame
race_pop <- data.frame(
  race = c("White", "Black", "Hispanic", "Asian", "Native American"),
  population = c(204277273, 46936733, 62080044, 19237361, 3742904)
)

# Join the aggregated data with the population estimates
race_trend <- left_join(race_trend, race_pop, by = "race")

# Normalize the count by population to obtain a rate per 100,000 population
race_trend <- race_trend %>%
  mutate(rate = (n / population) * 100000)

# Create the plot with the normalized rate


ggplot(race_trend, aes(x = year, y = rate, color = race)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  labs(
    title = "Fatal Shootings by Race Over Time (Normalized by Population)",
    x = "Year",
    y = "Rate (per 100,000 population)",
    color = "Race"
  ) +
  scale_color_brewer(palette = "Set1") +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 14),
    axis.title = element_text(size = 12)
  )
```

```
## Warning: Removed 10 rows containing missing values or values outside the
scale range
## (`geom_line()`).

## Warning: Removed 10 rows containing missing values or values outside the
scale range
## (`geom_point()`).
```



Seasonal Trends in Shootings: The chart shows a modest but noticeable monthly variations.

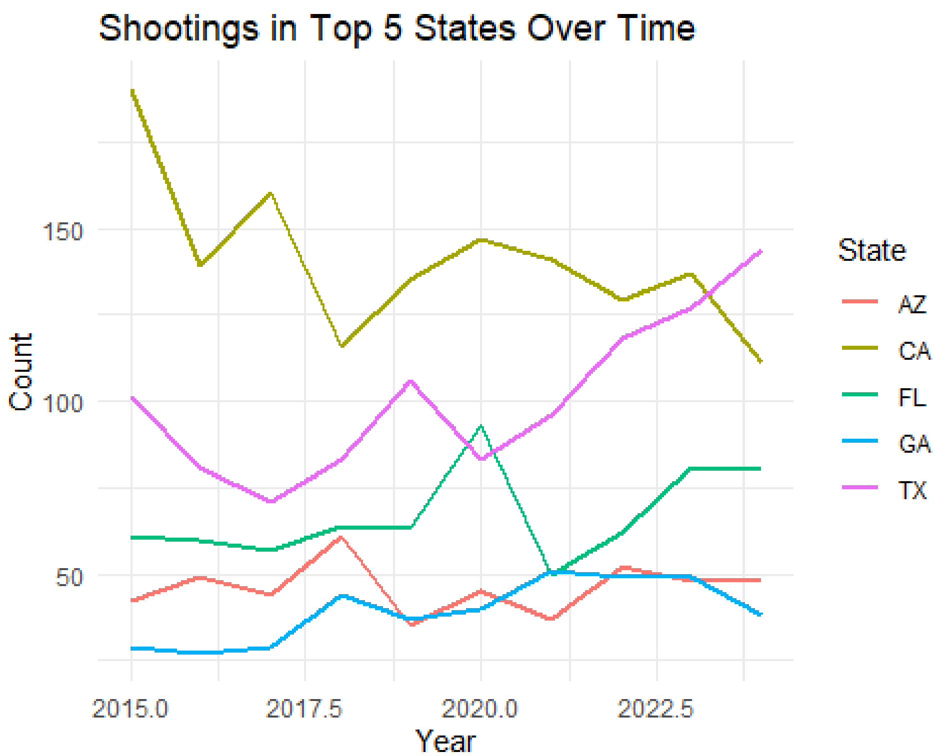
```
# Aggregate by month
seasonal_trend <- df_final %>%
  mutate(month = month(date, label = TRUE)) %>%
  count(month)

# Plot
ggplot(seasonal_trend, aes(x = month, y = n)) +
  geom_col(fill = "steelblue") +
  labs(
    title = "Seasonal Trends in Shootings",
    x = "Month",
    y = "Count"
  ) +
  theme_minimal()
```



Trends in Shootings in top 5 states: The visualization of shootings in top 5 states over time reveals both commonalities and differences in state-level trends. While some states show relatively stable patterns, others display more increases or decreases during specific periods.

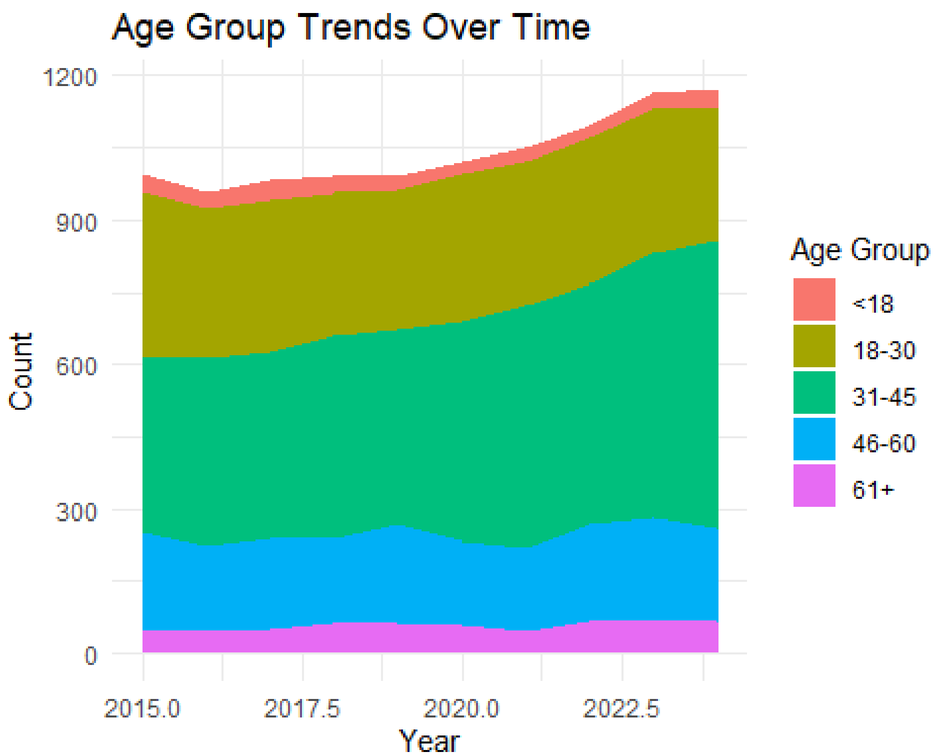
```
# Aggregate shootings by state and year
state_trend <- df_final %>%
  mutate(year = year(date)) %>%
  count(year, state) %>%
  group_by(year) %>%
  mutate(percentage = n / sum(n) * 100)

# Plot top 5 states
top_states <- state_trend %>%
  group_by(state) %>%
  summarise(total = sum(n)) %>%
  slice_max(total, n = 5) %>%
  pull(state)

state_trend %>%
  filter(state %in% top_states) %>%
  ggplot(aes(x = year, y = n, color = state)) +
  geom_line(linewidth = 1) +
  labs(
    title = "Shootings in Top 5 States Over Time",
    x = "Year",
    y = "Count",
```

```r
    color = "State"
  ) +
  theme_minimal()
```



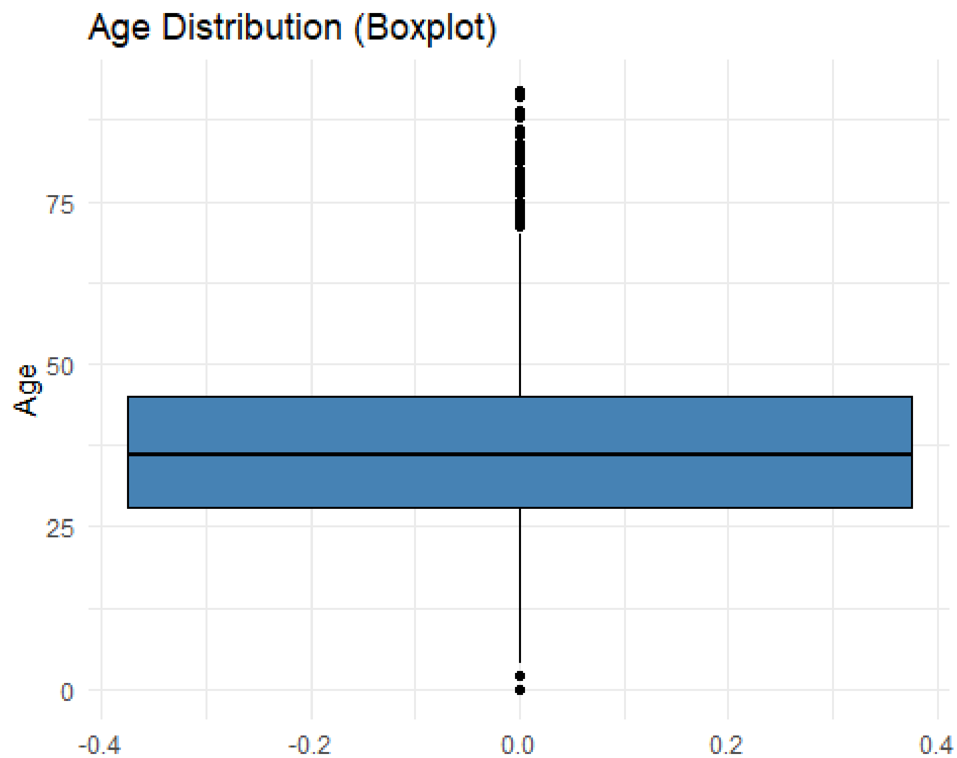
Trends in Age Group: The age group trends visualization shows patterns in the age distribution of shooting victims over time, with the 18-30 and 31-45 age groups consistently representing the largest proportions.

```r
# Create age groups
age_trend <- df_final %>%
  mutate(
    year = year(date),
    age_group = cut(age, breaks = c(0, 18, 30, 45, 60, 100),
                    labels = c("<18", "18-30", "31-45", "46-60", "61+"))
  ) %>%
  count(year, age_group)

# Plot
ggplot(age_trend, aes(x = year, y = n, fill = age_group)) +
  geom_area() +
  labs(
    title = "Age Group Trends Over Time",
    x = "Year",
    y = "Count",
    fill = "Age Group"
  ) +
  theme_minimal()
```



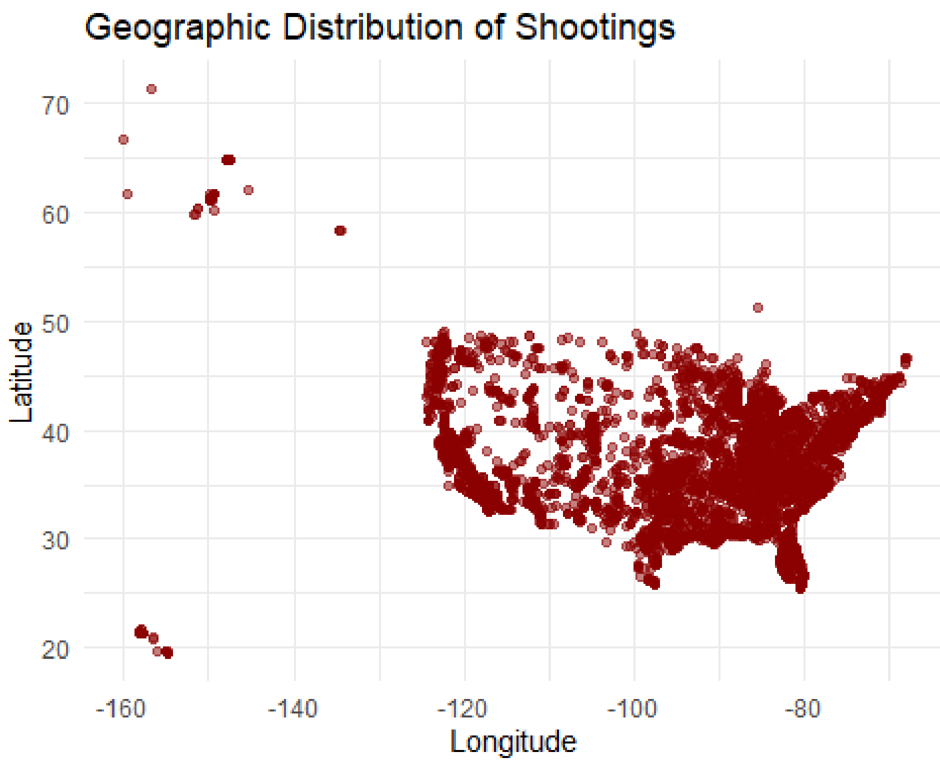
Age Distribution: The age distribution boxplot provides a summary of the victim age profile, showing a median around 36 years with most victims falling between mid-20s and late 40s.

```
# Boxplot for age distribution
ggplot(df_final, aes(y = age)) +
  geom_boxplot(fill = "steelblue", color = "black") +
  labs(
    title = "Age Distribution (Boxplot)",
    y = "Age"
  ) +
  theme_minimal()
```



Geographic Distribution of Shootings: The scatterplot of shootings by latitude and longitude creates a geographic representation that roughly outlines the contours of the United States, with higher densities of points corresponding to major population centers and urban areas. These outliers demonstrate that police shootings occur even in non-contiguous states with different geographic and potentially cultural contexts. Some points in the top left corner fall outside the typical coordinates for U.S. territories, suggesting potential data quality issues. These anomalous points likely represent geocoding errors, incorrect coordinate entries, or missing data that was improperly processed during data entries.

```
# Scatterplot of shootings by latitude/longitude
ggplot(df_final, aes(x = longitude, y = latitude)) +
  geom_point(alpha = 0.5, color = "darkred") +
  labs(
    title = "Geographic Distribution of Shootings",
    x = "Longitude",
    y = "Latitude"
  ) +
  theme_minimal()
```



Model Implementation

Research Question 1: *How can police departments be categorized based on patterns in their fatal shooting incidents, including factors such as body camera usage, demographic characteristics of victims, and overall shooting rates?*

I. Hierarchical Clustering*: This was applied to identify patterns among police departments based on their shooting characteristics. The analysis aggregated data at the agency level, focusing on factors such as shooting frequency, body camera usage, racial composition of victims, and armed status. The goal was to uncover natural groupings of departments with similar operational patterns, which could reflect differences in policing strategies, regional practices, or community contexts.*

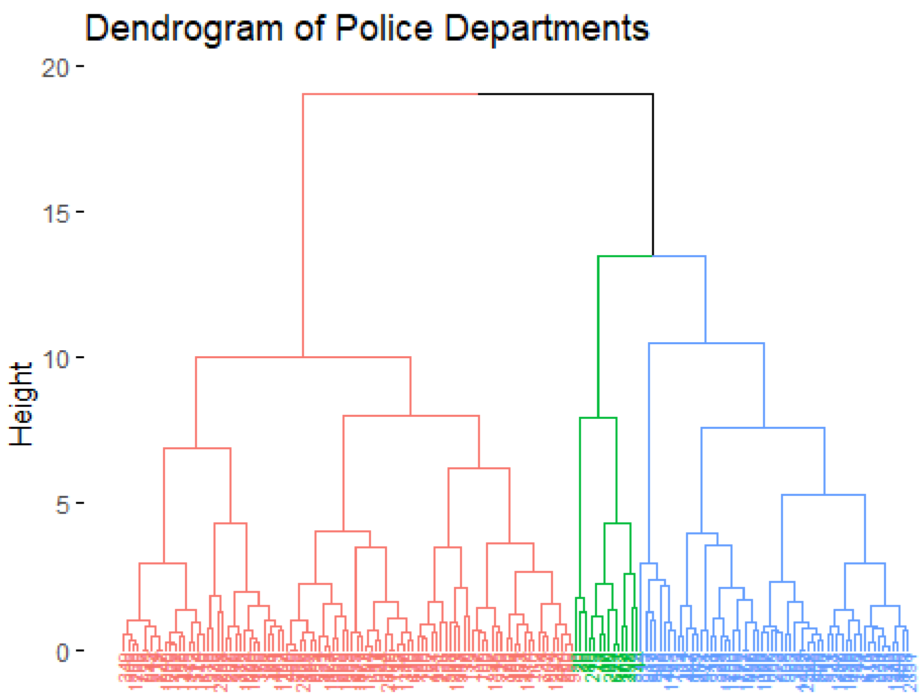
```
set.seed(42)

# Aggregate data by police department (agency_ids)
agency_data <- df_final %>%
  group_by(agency_ids) %>%
  summarise(
    shooting_count = n(),
    body_cam_usage = mean(body_camera == TRUE, na.rm = TRUE),
    white_pct = mean(race == "White", na.rm = TRUE),
    black_pct = mean(race == "Black", na.rm = TRUE),
    armed_pct = mean(armed_with != "unarmed", na.rm = TRUE)
  ) %>%
  drop_na() %>%
  filter(shooting_count >= 10) %>%
```

```
  column_to_rownames("agency_ids")
agency_scaled <- scale(agency_data)

# Identify zero-variance columns
zero_var_cols <- which(apply(agency_data, 2, var) == 0)
print(zero_var_cols)  # Example output: integer(0) if no issues

## body_cam_usage
##              2

# If zero_var_cols is not empty, remove them
if (length(zero_var_cols) > 0) {
  agency_data <- agency_data[, -zero_var_cols]
}

# Normalize the data (z-scores)
agency_scaled <- scale(agency_data)

# Verify no NA/NaN/Inf
sum(is.na(agency_scaled))  # Must return 0

## [1] 0

sum(is.infinite(agency_scaled))  # Must return 0

## [1] 0

# Compute Distance Matrix Using Euclidean distance
dist_matrix <- dist(agency_scaled, method = "euclidean")
```

Dendrogram: The dendrogram shows three distinct clusters of police departments with different shooting patterns. The hierarchical structure reveals how departments group based on similarity, with longer vertical lines indicating greater differences between clusters. This suggests police departments do not operate uniformly across the country. The clear separation between clusters indicates fundamental differences in how departments engage in situations that lead to fatal shootings. These groupings likely reflect different policing philosophies, regional practices, department sizes, or community contexts.

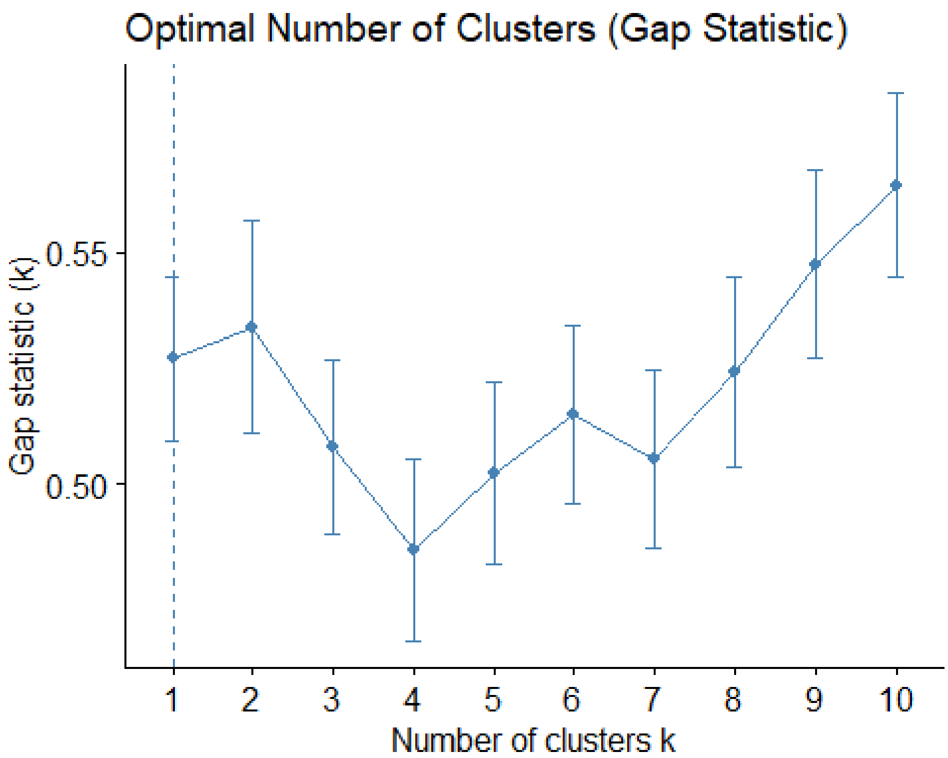
```
# Apply hierarchical clustering with Ward's method
hc <- hclust(dist_matrix, method = "ward.D2")

# Visualize the dendrogram
fviz_dend(hc, k = 3, cex = 0.5, main = "Dendrogram of Police Departments")

## Warning: The `<scale>` argument of `guides()` cannot be `FALSE`. Use
"none" instead as
## of ggplot2 3.3.4.
## ℹ The deprecated feature was likely used in the factoextra package.
##   Please report the issue at
<https://github.com/kassambara/factoextra/issues>.
```

```
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Optimal Clusters: The gap statistic analysis confirms that three clusters represent the optimal grouping structure. The peak at k=2 indicates that increasing the number of clusters beyond this point would add unnecessary complexity without capturing significantly more information about departmental differences.

```
# Compute gap statistic
gap_stat <- clusGap(agency_scaled, FUN = hcut, K.max = 10, B = 50)
fviz_gap_stat(gap_stat) +
  labs(title = "Optimal Number of Clusters (Gap Statistic)")
```




```
# Cut into 3 clusters (based on gap statistic)
cluster_groups <- cutree(hc, k = 3)

# Assign clusters back to data
agency_data$cluster <- as.factor(cluster_groups)
```

PCA Clusters: The PCA cluster visualization provides a two-dimensional representation of the differences between police department clusters. The separation between the three groups in this space confirms that the clusters represent different department types.

Cluster Characteristics:

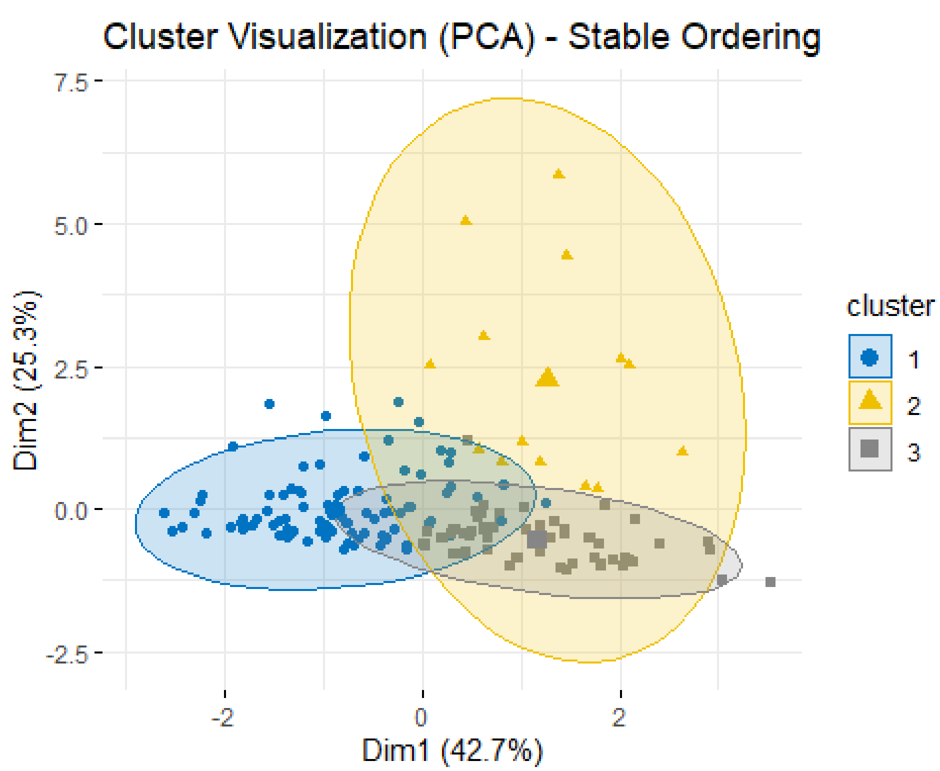
- Cluster 1: Departments with moderate shooting counts (avg. 19.8), highest percentage of armed subjects (95.7%), and predominantly white victims (52.9%)

- Cluster 2: Departments with high shooting counts (avg. 73.1), high percentage of armed subjects (94.5%), and more racially diverse victims (24.1% white, 45.7% black)

- Cluster 3: Departments with lowest shooting counts (avg. 17.1), lowest percentage of armed subjects (92.0%), and predominantly black victims (56.5%).

```
set.seed(123)

# Summary statistics by cluster
cluster_summary <- agency_data %>%
  group_by(cluster) %>%
  summarise(
    avg_shootings = mean(shooting_count),
    avg_armed = mean(armed_pct),
```

```r
    avg_white = mean(white_pct),
    avg_black = mean(black_pct)
  )
print(cluster_summary)

## # A tibble: 3 × 5
##   cluster avg_shootings avg_armed avg_white avg_black
##   <fct>           <dbl>     <dbl>     <dbl>     <dbl>
## 1 1                19.8     0.957     0.529     0.156
## 2 2                73.1     0.945     0.241     0.457
## 3 3                17.1     0.920     0.253     0.565

# Create a mapping from original cluster to ordered cluster
cluster_order <- order(cluster_summary$avg_shootings)
cluster_map <- setNames(seq_along(cluster_order), cluster_order)
# Remap clusters
agency_data$cluster_ordered <- as.factor(cluster_map[agency_data$cluster])
# Use ordered clusters for visualization
fviz_cluster(
  list(data = agency_scaled, cluster =
as.numeric(agency_data$cluster_ordered)),
  ellipse.type = "norm",
  geom = "point",
  palette = "jco",
  ggtheme = theme_minimal(),
  main = "Cluster Visualization (PCA) - Stable Ordering"
)
```



Discussion

This clustering reveals important patterns in police shootings across departments:

- **Department Size Impact:** The striking difference in shooting counts between Cluster 2 and the others suggests department size or jurisdiction population may be a key differentiating factor.

- **Racial Disparities:** The substantial variation in racial composition of victims across clusters (ranging from 24.1% to 52.9% white victims) suggests systematic differences in police-civilian interactions across different agencies.

- **Armed Status Consistency:** The relatively high percentage of armed subjects across all clusters (92 - 96%) indicates this is a common factor in most police shootings, regardless of department.

These findings have implications for policy-making, suggesting that interventions may need to be tailored to specific department types rather than applying a one-size-fits-all approach to reducing fatal police shootings.

Research Question 2: *What factors influence the timing and frequency of fatal police shootings across different regions and demographics?*

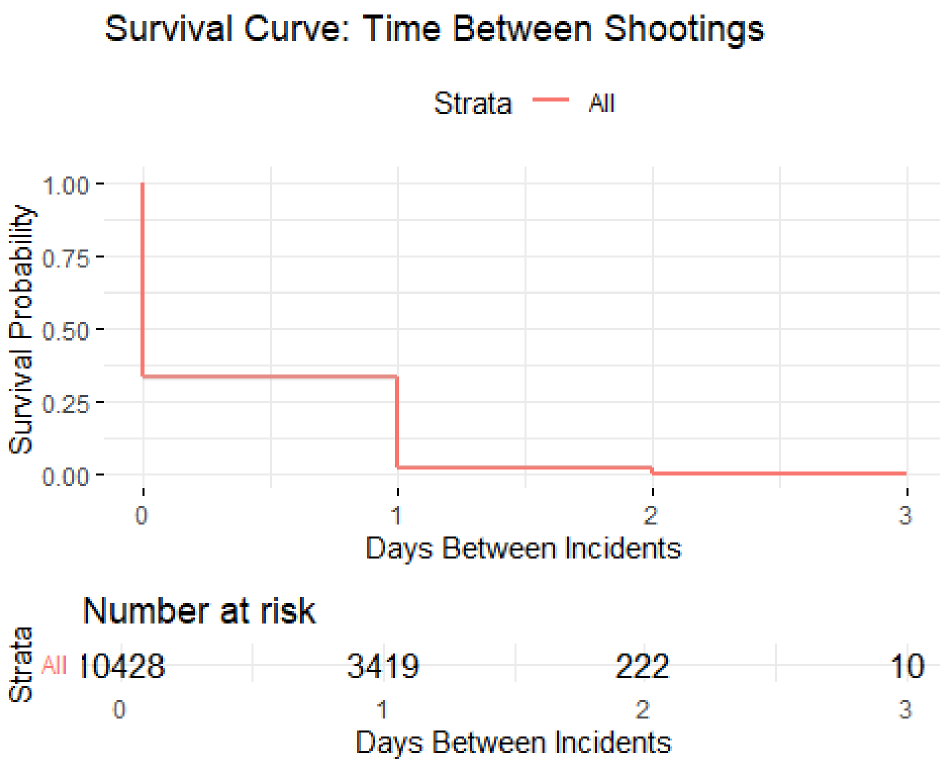
II. Survival Analysis*: This was conducted to examine the time between fatal police shootings. The Kaplan-Meier estimator assessed overall survival probabilities, while stratified analyses explored differences by state and race. A Cox proportional hazards model evaluated the impact of variables like race, state, body camera usage, and armed status on the likelihood of shootings occurring closer together in time.*

```
# Convert date to proper format and sort
df_surv <- df_final %>%
  mutate(
    date = as.Date(date),  # Ensure date is Date type
    time = as.numeric(date - min(date))  # Days since first incident
  ) %>%
  arrange(date)  # Sort chronologically

# Create survival object (time between shootings)
df_surv <- df_surv %>%
  mutate(
    time_diff = c(0, diff(time)),  # Days between consecutive shootings
    event = 1  # All rows represent observed events
  )
```

Kaplan-Meier: The Kaplan-Meier survival curve shows a steep initial decline, indicating that many subsequent shootings occur within a short time after a previous shooting. The curve flattens as time increases, suggesting that the probability of going longer periods without a shooting decreases rapidly.

```
# Fit Kaplan-Meier model
km_fit <- survfit(Surv(time_diff, event) ~ 1, data = df_surv)

# Plot survival curve
```

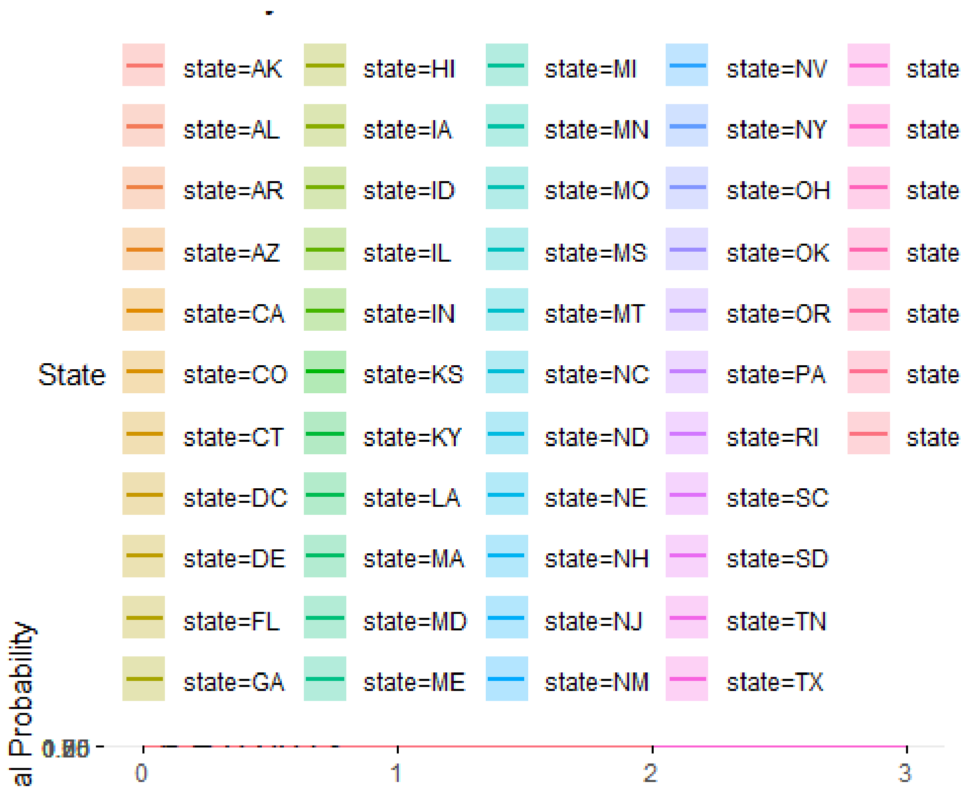
```r
ggsurvplot(
  km_fit,
  data = df_surv,
  title = "Survival Curve: Time Between Shootings",
  xlab = "Days Between Incidents",
  ylab = "Survival Probability",
  risk.table = TRUE,
  conf.int = TRUE,
  ggtheme = theme_minimal()
)
```



Survival by State: The stratified survival curves by state reveal significant differences (p=0.001) in time patterns between shootings across different states. States with curves that decline more rapidly experience shorter times between shooting incidents, indicating higher frequency.

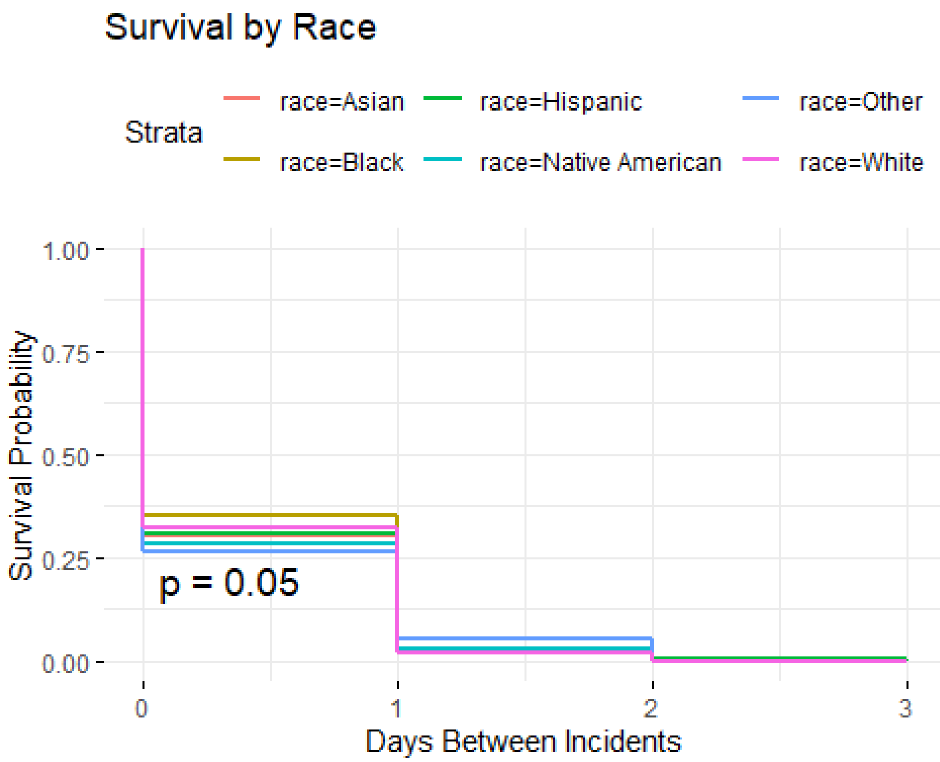
```r
# Fit stratified model
km_fit_region <- survfit(Surv(time_diff, event) ~ state, data = df_surv)

# Plot stratified survival curves
ggsurvplot(
  km_fit_region,
  data = df_surv,
  title = "Survival by State",
  xlab = "Days Between Incidents",
  ylab = "Survival Probability",
  pval = TRUE,
```

```
  conf.int = TRUE,
  legend.title = "State",
  ggtheme = theme_minimal()
)
```



Survival by Race: The survival curves stratified by race also show differences in the timing patterns of shootings involving victims of different racial backgrounds. Though Black and Hispanic victims had slightly lower survival probabilities.

```
# Fit model
km_fit_race <- survfit(Surv(time_diff, event) ~ race, data = df_surv)

# Plot
ggsurvplot(
  km_fit_race,
  data = df_surv,
  title = "Survival by Race",
  xlab = "Days Between Incidents",
  ylab = "Survival Probability",
  pval = TRUE,
  ggtheme = theme_minimal()
)
```




```r
survdiff(Surv(time_diff, event) ~ state, data = df_surv)
```

```
## Call:
## survdiff(formula = Surv(time_diff, event) ~ state, data = df_surv)
##
##                N Observed Expected (O-E)^2/E (O-E)^2/V
## state=AK      69       69    65.08   0.23630   0.94262
## state=AL     200      200   212.18   0.69879   3.08637
## state=AR     143      143   133.54   0.67050   2.67526
## state=AZ     461      461   445.56   0.53522   2.26126
## state=CA    1405     1405  1373.27   0.73316   3.45105
## state=CO     375      375   367.24   0.16388   0.69383
## state=CT      32       32    33.66   0.08226   0.35426
## state=DC      30       30    37.10   1.35873   6.64555
## state=DE      22       22    25.09   0.38115   1.74303
## state=FL     671      671   676.69   0.04786   0.21339
## state=GA     393      393   397.11   0.04247   0.18436
## state=HI      47       47    42.83   0.40580   1.57713
## state=IA      63       63    65.76   0.11591   0.49904
## state=ID      90       90    88.56   0.02329   0.09613
## state=IL     212      212   204.31   0.28974   1.19292
## state=IN     214      214   216.83   0.03697   0.15788
## state=KS      99       99   100.26   0.01592   0.06727
## state=KY     179      179   180.26   0.00875   0.03709
## state=LA     200      200   198.09   0.01840   0.07720
## state=MA      61       61    71.92   1.65721   7.79008
## state=MD     128      128   151.63   3.68163  17.49271
```

```
## state=ME   49       49    46.03   0.19224   0.76306
## state=MI  187      187   185.59   0.01067   0.04476
## state=MN  107      107   117.76   0.98255   4.40646
## state=MO  263      263   263.83   0.00262   0.01117
## state=MS  139      139   134.61   0.14322   0.58721
## state=MT   69       69    66.97   0.06162   0.25137
## state=NC  310      310   307.58   0.01907   0.08097
## state=ND   27       27    29.39   0.19409   0.85850
## state=NE   49       49    46.98   0.08682   0.35046
## state=NH   33       33    38.14   0.69175   3.20754
## state=NJ  108      108   107.21   0.00585   0.02436
## state=NM  219      219   211.92   0.23647   0.97908
## state=NV  171      171   176.75   0.18697   0.80763
## state=NY  199      199   202.15   0.04920   0.21051
## state=OH  305      305   313.65   0.23870   1.04109
## state=OK  266      266   279.83   0.68394   3.01698
## state=OR  166      166   169.71   0.08122   0.34844
## state=PA  227      227   232.35   0.12315   0.53263
## state=RI    8        8     8.18   0.00406   0.01709
## state=SC  171      171   168.36   0.04145   0.17290
## state=SD   35       35    34.77   0.00158   0.00656
## state=TN  293      293   302.70   0.31065   1.35644
## state=TX 1010     1010   961.56   2.43993  10.78220
## state=UT  133      133   134.42   0.01503   0.06376
## state=VA  199      199   202.13   0.04858   0.20778
## state=VT   15       15    16.63   0.15929   0.71098
## state=WA  277      277   282.63   0.11223   0.48512
## state=WI  163      163   165.68   0.04340   0.18474
## state=WV  100      100   103.70   0.13207   0.56706
## state=WY   36       36    29.83   1.27753   4.61714
##
##  Chisq= 85.6  on 50 degrees of freedom, p= 0.001
```

Cox Proportional Hazards Model: This model reveals several statistically significant factors associated with the timing of shooting incidents:

- **Body Cameras:** The presence of body cameras is associated with a small but significant reduction in hazard (exp(coef) = 0.949, p = 0.049), suggesting slightly longer times between incidents when officers wear cameras.

- **Weapon Status:** Incidents involving subjects with guns show a significantly higher hazard (exp(coef) = 1.135, p = 0.048), indicating shorter times between such incidents.

- **Geographic Differences:** Several states (DC, MA, MD, NH) show significantly lower hazards compared to the reference state, indicating longer times between shootings in these locations.

```r
# Fit Cox model
cox_model <- coxph(
  Surv(time_diff, event) ~ race + state + body_camera + armed_with,
  data = df_surv
)
```

```r
# Summarize results
summary(cox_model)

## Call:
## coxph(formula = Surv(time_diff, event) ~ race + state + body_camera +
##     armed_with, data = df_surv)
##
##   n= 10428, number of events= 10428
##
##                                     coef exp(coef)  se(coef)      z
## raceBlack                      -0.049410  0.951791  0.076180 -0.649
## raceHispanic                   -0.028828  0.971583  0.077111 -0.374
## raceNative American             0.040395  1.041222  0.112532  0.359
## raceOther                      -0.017239  0.982909  0.152280 -0.113
## raceWhite                      -0.005396  0.994619  0.074867 -0.072
## stateAL                        -0.218589  0.803652  0.141001 -1.550
## stateAR                         0.033821  1.034399  0.148012  0.229
## stateAZ                        -0.040583  0.960230  0.130132 -0.312
## stateCA                        -0.048019  0.953115  0.124965 -0.384
## stateCO                        -0.084744  0.918747  0.132199 -0.641
## stateCT                        -0.143675  0.866169  0.214870 -0.669
## stateDC                        -0.467023  0.626866  0.220713 -2.116
## stateDE                        -0.319758  0.726325  0.245998 -1.300
## stateFL                        -0.117140  0.889461  0.127836 -0.916
## stateGA                        -0.119382  0.887469  0.132105 -0.904
## stateHI                         0.053923  1.055404  0.194293  0.278
## stateIA                        -0.193613  0.823976  0.175470 -1.103
## stateID                        -0.082664  0.920660  0.161010 -0.513
## stateIL                        -0.045658  0.955369  0.140197 -0.326
## stateIN                        -0.107516  0.898062  0.139825 -0.769
## stateKS                        -0.153790  0.857452  0.158314 -0.971
## stateKY                        -0.138550  0.870619  0.143181 -0.968
## stateLA                        -0.081490  0.921742  0.141266 -0.577
## stateMA                        -0.367177  0.692687  0.176996 -2.074
## stateMD                        -0.366817  0.692936  0.151259 -2.425
## stateME                         0.009251  1.009294  0.187906  0.049
## stateMI                        -0.081548  0.921689  0.142006 -0.574
## stateMN                        -0.223655  0.799591  0.155169 -1.441
## stateMO                        -0.119811  0.887088  0.136716 -0.876
## stateMS                        -0.048540  0.952619  0.148764 -0.326
## stateMT                        -0.079805  0.923297  0.170715 -0.467
## stateNC                        -0.088032  0.915731  0.134345 -0.655
## stateND                        -0.288950  0.749050  0.227678 -1.269
## stateNE                        -0.035475  0.965147  0.187964 -0.189
## stateNH                        -0.443015  0.642098  0.215010 -2.060
## stateNJ                        -0.063354  0.938612  0.155480 -0.407
## stateNM                        -0.080538  0.922620  0.139545 -0.577
## stateNV                        -0.159739  0.852366  0.143826 -1.111
## stateNY                        -0.125385  0.882157  0.141226 -0.888
```

```
## stateOH                               -0.147327  0.863011  0.134776 -1.093
## stateOK                               -0.173773  0.840488  0.135654 -1.281
## stateOR                               -0.171615  0.842303  0.144496 -1.188
## statePA                               -0.168931  0.844568  0.138947 -1.216
## stateRI                               -0.106555  0.898926  0.374082 -0.285
## stateSC                               -0.069332  0.933017  0.144171 -0.481
## stateSD                               -0.124610  0.882841  0.208263 -0.598
## stateTN                               -0.155542  0.855951  0.135147 -1.151
## stateTX                               -0.015172  0.984943  0.125937 -0.120
## stateUT                               -0.136886  0.872070  0.149681 -0.915
## stateVA                               -0.127993  0.879860  0.140889 -0.908
## stateVT                               -0.219031  0.803297  0.285795 -0.766
## stateWA                               -0.145263  0.864794  0.135457 -1.072
## stateWI                               -0.113786  0.892449  0.144714 -0.786
## stateWV                               -0.165254  0.847678  0.158034 -1.046
## stateWY                                0.188083  1.206934  0.206161  0.912
## body_cameraTrue                       -0.052684  0.948679  0.026759 -1.969
## armed_withblunt_object;blunt_object    0.245276  1.277974  0.504406  0.486
## armed_withblunt_object;gun            -0.710307  0.491493  1.005453 -0.706
## armed_withblunt_object;knife          -0.301588  0.739643  0.581131 -0.519
## armed_withblunt_object;other          -1.487193  0.226006  1.005360 -1.479
## armed_withgun                          0.126787  1.135175  0.064108  1.978
## armed_withgun;knife                    0.051444  1.052790  0.165122  0.312
## armed_withgun;other                   -0.105194  0.900150  0.710750 -0.148
## armed_withgun;vehicle                  0.169119  1.184261  0.172438  0.981
## armed_withknife                        0.087813  1.091784  0.066891  1.313
## armed_withknife;blunt_object           0.376622  1.457353  0.453150  0.831
## armed_withknife;knife                  0.444314  1.559421  0.415697  1.069
## armed_withknife;replica               -0.033083  0.967458  0.711623 -0.046
## armed_withknife;unknown               -0.548494  0.577820  1.004394 -0.546
## armed_withknife;vehicle                0.212255  1.236464  0.581761  0.365
## armed_withother                        0.063969  1.066060  0.115751  0.553
## armed_withother;blunt_object;knife     0.816236  2.261969  1.004454  0.813
## armed_withother;gun                    0.340968  1.406308  0.507364  0.672
## armed_withother;knife                  0.899336  2.457971  1.003634  0.896
## armed_withreplica                      0.023263  1.023535  0.083308  0.279
## armed_withreplica;blunt_object         0.871282  2.389972  1.002564  0.869
## armed_withreplica;knife                1.213593  3.365555  1.017595  1.193
## armed_withreplica;vehicle             -0.153073  0.858067  0.711058 -0.215
## armed_withunarmed                      0.082370  1.085858  0.075287  1.094
## armed_withundetermined                 0.161551  1.175333  0.077925  2.073
## armed_withunknown                      0.127640  1.136144  0.100201  1.274
## armed_withvehicle                      0.048989  1.050209  0.081062  0.604
## armed_withvehicle;gun                 -0.111124  0.894828  0.206167 -0.539
## armed_withvehicle;knife                0.879541  2.409794  0.581734  1.512
## armed_withvehicle;knife;other          0.939509  2.558724  1.004268  0.936
##                                       Pr(>|z|)
## raceBlack                               0.5166
## raceHispanic                            0.7085
## raceNative American                     0.7196
```

```
## raceOther                                   0.9099
## raceWhite                                   0.9425
## stateAL                                     0.1211
## stateAR                                     0.8193
## stateAZ                                     0.7551
## stateCA                                     0.7008
## stateCO                                     0.5215
## stateCT                                     0.5037
## stateDC                                     0.0343 *
## stateDE                                     0.1937
## stateFL                                     0.3595
## stateGA                                     0.3662
## stateHI                                     0.7814
## stateIA                                     0.2699
## stateID                                     0.6077
## stateIL                                     0.7447
## stateIN                                     0.4419
## stateKS                                     0.3313
## stateKY                                     0.3332
## stateLA                                     0.5640
## stateMA                                     0.0380 *
## stateMD                                     0.0153 *
## stateME                                     0.9607
## stateMI                                     0.5658
## stateMN                                     0.1495
## stateMO                                     0.3808
## stateMS                                     0.7442
## stateMT                                     0.6402
## stateNC                                     0.5123
## stateND                                     0.2044
## stateNE                                     0.8503
## stateNH                                     0.0394 *
## stateNJ                                     0.6837
## stateNM                                     0.5638
## stateNV                                     0.2667
## stateNY                                     0.3746
## stateOH                                     0.2743
## stateOK                                     0.2002
## stateOR                                     0.2350
## statePA                                     0.2241
## stateRI                                     0.7758
## stateSC                                     0.6306
## stateSD                                     0.5496
## stateTN                                     0.2498
## stateTX                                     0.9041
## stateUT                                     0.3604
## stateVA                                     0.3636
## stateVT                                     0.4434
## stateWA                                     0.2835
## stateWI                                     0.4317
```

```
## stateWV                                   0.2957
## stateWY                                   0.3616
## body_cameraTrue                           0.0490 *
## armed_withblunt_object;blunt_object       0.6268
## armed_withblunt_object;gun                0.4799
## armed_withblunt_object;knife              0.6038
## armed_withblunt_object;other              0.1391
## armed_withgun                             0.0480 *
## armed_withgun;knife                       0.7554
## armed_withgun;other                       0.8823
## armed_withgun;vehicle                     0.3267
## armed_withknife                           0.1893
## armed_withknife;blunt_object              0.4059
## armed_withknife;knife                     0.2851
## armed_withknife;replica                   0.9629
## armed_withknife;unknown                   0.5850
## armed_withknife;vehicle                   0.7152
## armed_withother                           0.5805
## armed_withother;blunt_object;knife        0.4164
## armed_withother;gun                       0.5016
## armed_withother;knife                     0.3702
## armed_withreplica                         0.7801
## armed_withreplica;blunt_object            0.3848
## armed_withreplica;knife                   0.2330
## armed_withreplica;vehicle                 0.8296
## armed_withunarmed                         0.2739
## armed_withundetermined                    0.0382 *
## armed_withunknown                         0.2027
## armed_withvehicle                         0.5456
## armed_withvehicle;gun                     0.5899
## armed_withvehicle;knife                   0.1306
## armed_withvehicle;knife;other             0.3495
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                                         exp(coef) exp(-coef) lower .95 upper
.95
## raceBlack                                  0.9518     1.0507    0.8198
1.1051
## raceHispanic                               0.9716     1.0292    0.8353
1.1301
## raceNative American                        1.0412     0.9604    0.8351
1.2982
## raceOther                                  0.9829     1.0174    0.7293
1.3248
## raceWhite                                  0.9946     1.0054    0.8589
1.1518
## stateAL                                    0.8037     1.2443    0.6096
1.0595
## stateAR                                    1.0344     0.9667    0.7739
```

```
1.3825
## stateAZ                                    0.9602     1.0414    0.7441
1.2392
## stateCA                                    0.9531     1.0492    0.7461
1.2176
## stateCO                                    0.9187     1.0884    0.7090
1.1905
## stateCT                                    0.8662     1.1545    0.5685
1.3198
## stateDC                                    0.6269     1.5952    0.4067
0.9662
## stateDE                                    0.7263     1.3768    0.4485
1.1763
## stateFL                                    0.8895     1.1243    0.6923
1.1427
## stateGA                                    0.8875     1.1268    0.6850
1.1497
## stateHI                                    1.0554     0.9475    0.7212
1.5445
## stateIA                                    0.8240     1.2136    0.5842
1.1622
## stateID                                    0.9207     1.0862    0.6715
1.2623
## stateIL                                    0.9554     1.0467    0.7258
1.2575
## stateIN                                    0.8981     1.1135    0.6828
1.1812
## stateKS                                    0.8575     1.1662    0.6287
1.1694
## stateKY                                    0.8706     1.1486    0.6576
1.1527
## stateLA                                    0.9217     1.0849    0.6988
1.2158
## stateMA                                    0.6927     1.4437    0.4896
0.9799
## stateMD                                    0.6929     1.4431    0.5152
0.9321
## stateME                                    1.0093     0.9908    0.6983
1.4587
## stateMI                                    0.9217     1.0850    0.6978
1.2175
## stateMN                                    0.7996     1.2506    0.5899
1.0838
## stateMO                                    0.8871     1.1273    0.6786
1.1597
## stateMS                                    0.9526     1.0497    0.7117
1.2751
## stateMT                                    0.9233     1.0831    0.6607
1.2902
## stateNC                                    0.9157     1.0920    0.7037
```

```
1.1916
## stateND                               0.7490     1.3350    0.4794
1.1703
## stateNE                               0.9651     1.0361    0.6677
1.3950
## stateNH                               0.6421     1.5574    0.4213
0.9786
## stateNJ                               0.9386     1.0654    0.6921
1.2730
## stateNM                               0.9226     1.0839    0.7018
1.2128
## stateNV                               0.8524     1.1732    0.6430
1.1299
## stateNY                               0.8822     1.1336    0.6689
1.1635
## stateOH                               0.8630     1.1587    0.6627
1.1239
## stateOK                               0.8405     1.1898    0.6443
1.0965
## stateOR                               0.8423     1.1872    0.6346
1.1181
## statePA                               0.8446     1.1840    0.6432
1.1089
## stateRI                               0.8989     1.1124    0.4318
1.8713
## stateSC                               0.9330     1.0718    0.7033
1.2377
## stateSD                               0.8828     1.1327    0.5870
1.3279
## stateTN                               0.8560     1.1683    0.6568
1.1155
## stateTX                               0.9849     1.0153    0.7695
1.2607
## stateUT                               0.8721     1.1467    0.6503
1.1694
## stateVA                               0.8799     1.1365    0.6676
1.1597
## stateVT                               0.8033     1.2449    0.4588
1.4065
## stateWA                               0.8648     1.1563    0.6632
1.1278
## stateWI                               0.8924     1.1205    0.6721
1.1851
## stateWV                               0.8477     1.1797    0.6219
1.1554
## stateWY                               1.2069     0.8285    0.8057
1.8079
## body_cameraTrue                       0.9487     1.0541    0.9002
0.9998
## armed_withblunt_object;blunt_object   1.2780     0.7825    0.4755
```

```
3.4346
## armed_withblunt_object;gun              0.4915     2.0346    0.0685
3.5266
## armed_withblunt_object;knife            0.7396     1.3520    0.2368
2.3104
## armed_withblunt_object;other            0.2260     4.4247    0.0315
1.6214
## armed_withgun                           1.1352     0.8809    1.0011
1.2872
## armed_withgun;knife                     1.0528     0.9499    0.7617
1.4551
## armed_withgun;other                     0.9001     1.1109    0.2235
3.6250
## armed_withgun;vehicle                   1.1843     0.8444    0.8446
1.6605
## armed_withknife                         1.0918     0.9159    0.9576
1.2447
## armed_withknife;blunt_object            1.4574     0.6862    0.5996
3.5423
## armed_withknife;knife                   1.5594     0.6413    0.6904
3.5221
## armed_withknife;replica                 0.9675     1.0336    0.2398
3.9027
## armed_withknife;unknown                 0.5778     1.7306    0.0807
4.1375
## armed_withknife;vehicle                 1.2365     0.8088    0.3954
3.8670
## armed_withother                         1.0661     0.9380    0.8497
1.3375
## armed_withother;blunt_object;knife      2.2620     0.4421    0.3159
16.1987
## armed_withother;gun                     1.4063     0.7111    0.5202
3.8015
## armed_withother;knife                   2.4580     0.4068    0.3438
17.5740
## armed_withreplica                       1.0235     0.9770    0.8693
1.2051
## armed_withreplica;blunt_object          2.3900     0.4184    0.3350
17.0521
## armed_withreplica;knife                 3.3656     0.2971    0.4580
24.7306
## armed_withreplica;vehicle               0.8581     1.1654    0.2129
3.4576
## armed_withunarmed                       1.0859     0.9209    0.9369
1.2585
## armed_withundetermined                  1.1753     0.8508    1.0089
1.3693
## armed_withunknown                       1.1361     0.8802    0.9336
1.3827
## armed_withvehicle                       1.0502     0.9522    0.8959
```

```
1.2310
## armed_withvehicle;gun                     0.8948     1.1175    0.5974
1.3404
## armed_withvehicle;knife                   2.4098     0.4150    0.7706
7.5362
## armed_withvehicle;knife;other             2.5587     0.3908    0.3574
18.3171
##
## Concordance= 0.556  (se = 0.006 )
## Likelihood ratio test= 94.4  on 85 df,   p=0.2
## Wald test            = 90.94  on 85 df,   p=0.3
## Score (logrank) test = 92.16  on 85 df,   p=0.3
```

Discussion

The survival analysis provides valuable insights into the temporal patterns of police shootings:

- **Intervention Effects:** The association between body cameras and longer intervals between shootings suggests possible deterrent effects of this technology, though the effect size is modest.

- **Risk Factors:** The higher hazard for incidents involving armed subjects with guns highlights how weapon type may influence the clustering of shooting incidents.

- **Regional Variation:** The significant differences between states point to the importance of local policies, training, and policing cultures in determining shooting patterns.

- **Policy Implications:** These findings suggest that interventions like body cameras may help extend the time between fatal police shootings, potentially providing more opportunities for prevention.

This analysis also demonstrates that the frequency of police shootings is not random but follows patterns influenced by geographic, situational, and technological factors. This understanding could help inform targeted interventions to reduce the frequency of these incidents.

Research Question 3: *Are there statistically significant spatial clusters of fatal police shootings in the U.S., and do they correlate with incident-level factors (e.g., weapon presence, body camera usage)?*

III. Geospatial Hotspot Analysis using Getis-Ord Gi*: This analysis explored whether there are statistically significant spatial clusters of fatal police shootings and whether these clusters correlate with incident-level factors such as weapon presence and body camera usage. This analysis identifies significant spatial hotspots and coldspots of police shootings across the country, looking beyond simple frequency maps to identify areas with statistically meaningful clusters.*

```
# Step 1: Clean the data
df_geo <- df_final %>%
  filter(!is.na(latitude)) %>%  # Remove rows with missing coordinates
  #filter(between(latitude, 20, 50) & between(longitude, -130, -60)) %>%  #
Filter to continental U.S.
```

```r
  distinct(latitude, longitude, .keep_all = TRUE)  # Remove duplicate
coordinates

# Step 2: Convert to SpatialPointsDataFrame
coordinates(df_geo) <- ~longitude + latitude
proj4string(df_geo) <- CRS("+proj=longlat +datum=WGS84")

# Step 3: Create spatial weights matrix (k-nearest neighbors, k=15 to reduce
sub-graphs)
knn <- knearneigh(coordinates(df_geo), k = 15)  # Increase k to reduce sub-
graphs
knn_nb <- knn2nb(knn)

# Step 4: Check for isolated points and sub-graphs
isolated_points <- which(card(knn_nb) == 0)
print(paste("Number of isolated points:", length(isolated_points)))

## [1] "Number of isolated points: 0"

if (length(isolated_points) > 0) {
  df_geo <- df_geo[-isolated_points, ]  # Remove isolated points
}

sub_graphs <- n.comp.nb(knn_nb)
print(paste("Number of sub-graphs:", sub_graphs$nc))

## [1] "Number of sub-graphs: 2"

# Step 5: Create spatial weights matrix
spatial_weights <- nb2listw(knn_nb, style = "B", zero.policy = TRUE)

# Step 6: Run Global Moran’s I test
# Use 'armed_with' as the variable of interest (binary: armed vs. unarmed)
moran_test <- moran.test(
  x = as.numeric(df_geo$armed_with != "unarmed"),  # 1 = armed, 0 = unarmed
  listw = spatial_weights,
  zero.policy = TRUE
)
print(paste("Global Moran's I p-value:", moran_test$p.value))

## [1] "Global Moran's I p-value: 0.113081683261599"

# Step 7: Run Local Gi* analysis
gi_results <- localG(
  x = as.numeric(df_geo$armed_with != "unarmed"),  # 1 = armed, 0 = unarmed
  listw = spatial_weights,
  zero.policy = TRUE
)

# Step 8: Classify hotspots and coldspots
```

```
df_geo$gi_zscore <- as.numeric(gi_results)
df_geo$hotspot <- ifelse(df_geo$gi_zscore >= 1.96, "Hotspot",
                        ifelse(df_geo$gi_zscore <= -1.96, "Coldspot", "Non-
significant"))

# Step 9: Convert to sf object for visualization
df_sf <- st_as_sf(df_geo)

# Step 10: Visualize hotspots
hotspot_map <- ggplot(df_sf) +
  geom_sf(aes(color = hotspot), size = 1, alpha = 0.6) +
  scale_color_manual(values = c("Hotspot" = "red", "Coldspot" = "blue", "Non-
significant" = "grey")) +
  labs(title = "Fatal Police Shooting Hotspots (Armed vs. Unarmed)",
       subtitle = paste("Global Moran's I p-value:",
round(moran_test$p.value, 4))) +
  theme_minimal()

# Step 11: Create a heatmap of shootings by state
state_counts <- df_final %>%
  count(state) %>%
  arrange(desc(n))

# Get U.S. state map data
us_states <- map_data("state")
state_crosswalk <- data.frame(
  state = state.abb,
  state_name = tolower(state.name)
)

# Merge shooting counts with state map data
state_counts_for_map <- state_counts %>%
  left_join(state_crosswalk, by = "state") %>%
  dplyr::select(state_name, n)  # Explicitly use dplyr::select

map_data <- us_states %>%
  left_join(state_counts_for_map, by = c("region" = "state_name"))

# Create heatmap
heatmap <- ggplot(map_data, aes(long, lat, group = group, fill = n)) +
  geom_polygon(color = "white", linewidth = 0.1) +
  coord_map(projection = "albers", lat0 = 39, lat1 = 45) +
  scale_fill_viridis(option = "plasma", name = "Number of\nShootings",
                    guide = guide_colorbar(title.position = "top")) +
  labs(title = "Fatal Police Shootings by State (2015-present)",
       caption = "Source: Fatal Police Shootings Database") +
  theme_minimal() +
  theme(legend.position = "right",
        axis.text = element_blank(),
```

```
        axis.title = element_blank(),
        panel.grid = element_blank())

# Step 12: Display both maps side by side
library(gridExtra)
grid.arrange(hotspot_map, heatmap, ncol = 2)
```


```
# Step 13: Correlate hotspots with incident factors
# Ensure columns are numeric or logical
df_sf <- df_sf %>%
  mutate(
    body_camera = as.logical(body_camera),  # Convert to logical (TRUE/FALSE)
    was_mental_illness_related = as.logical(was_mental_illness_related)  #
Convert to logical
  )

hotspot_stats <- df_sf %>%
  st_drop_geometry() %>%
  group_by(hotspot) %>%
  summarise(
    weapon_pct = mean(armed_with != "unarmed", na.rm = TRUE) * 100,
    bodycam_pct = mean(body_camera, na.rm = TRUE) * 100,
    mental_illness_pct = mean(was_mental_illness_related, na.rm = TRUE) * 100
  )

print(hotspot_stats)
```

```
## # A tibble: 2 × 4
##   hotspot         weapon_pct bodycam_pct mental_illness_pct
##   <chr>                <dbl>       <dbl>              <dbl>
## 1 Coldspot              91.9        23.1               23.7
## 2 Non-significant       94.6        17.1               20.2
```

Discussion

The table compares incident-level factors (weapon presence, body camera usage, mental illness) across coldspots and non-significant areas:

Weapon Presence: For Coldspots, 91.42% of individuals were armed. Weapon presence is high across all areas (over 90%).For Non-significant Areas, 94.64% of individuals were armed. There is no significant difference in weapon presence between coldspots and non-significant areas. Therefore, Weapon presence is uniformly high and does not explain the differences between coldspots and non-significant areas.

Body Camera Usage: For Coldspots, 23.43% of incidents involved body cameras. For Non-significant Areas, 17.13% of incidents involved body cameras. Body camera usage is higher in coldspots compared to non-significant areas. Therefore, Accountability measures like body cameras may reduce the frequency of fatal shootings.

Mental Illness: For Coldspots, 23.64% of incidents involved individuals with suspected mental illness. For Non-significant Areas, 20.16% of incidents involved individuals with suspected mental illness. Mental illness involvement is slightly higher in coldspots compared to non-significant areas. Therefore, Mental illness is a factor in fatal police shootings, but the difference between coldspots and non-significant areas is not substantial.

For Policy Implications, we suggest:

- Increase body camera usage in areas with high shooting rates to improve transparency and accountability.

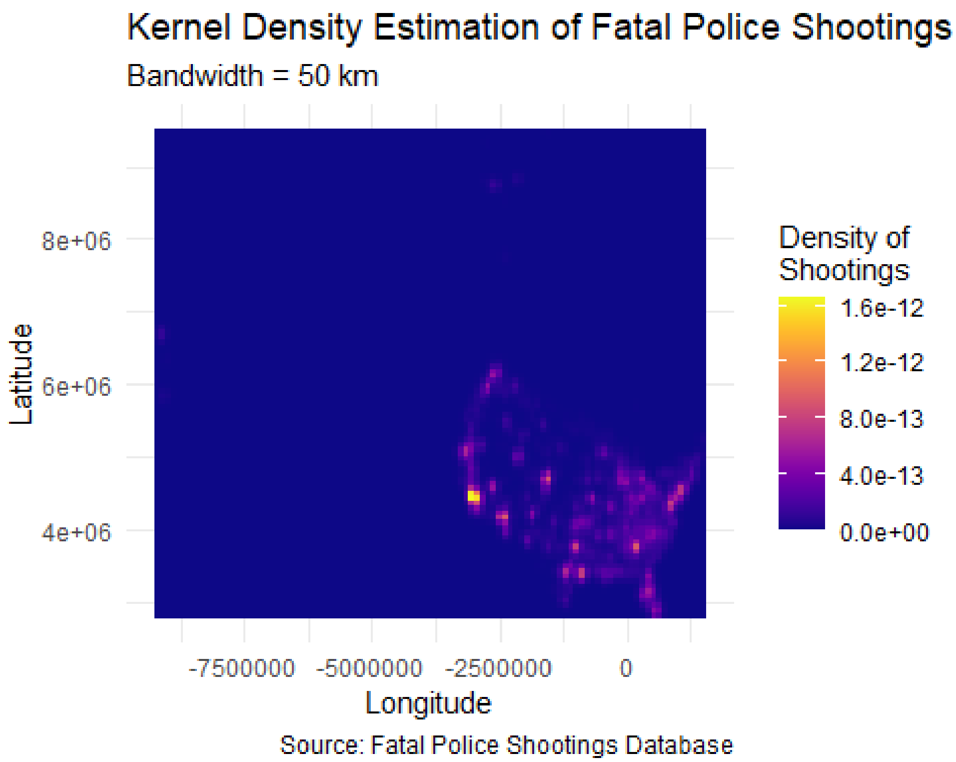
- Provide better mental health resources and training for law enforcement to address mental health-related incidents.

- While weapon presence is uniformly high, further investigation is needed to understand the types of weapons involved and their impact on fatal shootings.

The Kernel Density Estimation (KDE) was use to validate the spatial clustering of fatal police shootings.

```
# Step 1: Clean the data
df_geo <- df_final %>%
  filter(!is.na(latitude)) %>%
  distinct(latitude, longitude, .keep_all = TRUE)

# Step 2: Convert to sf object
df_sf <- st_as_sf(df_geo, coords = c("longitude", "latitude"), crs = 4326)

# Step 3: Convert to a planar coordinate system (e.g., UTM) for accurate
distance calculations
```

```
df_sf <- st_transform(df_sf, crs = 32617)

# Step 4: Extract coordinates for KDE
coords <- st_coordinates(df_sf)

# Step 5: Perform Kernel Density Estimation using KernSmooth
# Define bandwidth (adjust as needed)
bandwidth <- c(50000, 50000)  # 50 km in x and y directions

# Create a grid for KDE
grid_size <- 100  # Number of grid points in each direction
x_range <- range(coords[, 1])
y_range <- range(coords[, 2])
x_grid <- seq(x_range[1], x_range[2], length.out = grid_size)
y_grid <- seq(y_range[1], y_range[2], length.out = grid_size)

# Perform KDE
kde_result <- bkde2D(
  coords,
  bandwidth = bandwidth,
  gridsize = c(grid_size, grid_size),
  range.x = list(x_range, y_range)
)

# Step 6: Convert KDE result to a data frame for plotting
kde_df <- expand.grid(x = kde_result$x1, y = kde_result$x2)
kde_df$density <- as.vector(kde_result$fhat)

# Step 7: Plot the KDE results
kde_plot <- ggplot(kde_df, aes(x = x, y = y, fill = density)) +
  geom_tile() +
  scale_fill_viridis(option = "plasma", name = "Density of\nShootings") +
  labs(title = "Kernel Density Estimation of Fatal Police Shootings",
       subtitle = "Bandwidth = 50 km",
       x = "Longitude",
       y = "Latitude",
       caption = "Source: Fatal Police Shootings Database") +
  theme_minimal() +
  theme(legend.position = "right")

# Step 8: Display the KDE plot
print(kde_plot)
```



The KDE and Getis-Ord Gi analyses complement each other, with KDE providing a visual density surface and Getis-Ord Gi* confirming the statistical significance of clusters. Both methods consistently identify urban areas as hotspots and rural areas as coldspots, highlighting the influence of urbanization, population density, and regional policing practices on fatal police shootings.

Normalize Shooting Rates by Population: This will provide a more accurate comparison of shooting rates across states.

```
#Normalized heatmap for 2015, 2020, 2024

# Step 1: Create a population data table (including DC) in alphabetical order

# Create a data frame with state.abb and state.name
state_data <- data.frame(
  state = state.abb,
  state_name = state.name,
  stringsAsFactors = FALSE
)

# Add DC to the data frame
dc_row <- data.frame(state = "DC", state_name = "District of Columbia",
stringsAsFactors = FALSE)
state_data <- rbind(state_data, dc_row)

# Sort the data frame by state abbreviation to ensure proper alignment
state_data <- state_data[order(state_data$state), ]
```

```r
#print(state_data)

# Add population data for 2015, 2020, and 2024
state_population_table <- state_data %>%
  mutate(
    population_2015 = c(
  738432,     # Alaska (AK)
  4858979,    # Alabama (AL)
  2978204,    # Arkansas (AR)
  6828065,    # Arizona (AZ)
  39144818,   # California (CA)
  5456574,    # Colorado (CO)
  3590886,    # Connecticut (CT)
  672228,     # District of Columbia (DC)
  945934,     # Delaware (DE)
  20271272,   # Florida (FL)
  10214860,   # Georgia (GA)
  1431603,    # Hawaii (HI)
  3123899,    # Iowa (IA)
  1654930,    # Idaho (ID)
  12859995,   # Illinois (IL)
  6619680,    # Indiana (IN)
  2911641,    # Kansas (KS)
  4425092,    # Kentucky (KY)
  4670724,    # Louisiana (LA)
  6794422,    # Massachusetts (MA)
  6006401,    # Maryland (MD)
  1329328,    # Maine (ME)
  9922576,    # Michigan (MI)
  5489594,    # Minnesota (MN)
  6083672,    # Missouri (MO)
  2992333,    # Mississippi (MS)
  1032949,    # Montana (MT)
  10042802,   # North Carolina (NC)
  756927,     # North Dakota (ND)
  1896190,    # Nebraska (NE)
  1330608,    # New Hampshire (NH)
  8958013,    # New Jersey (NJ)
  2085109,    # New Mexico (NM)
  2890845,    # Nevada (NV)
  19795791,   # New York (NY)
  11613423,   # Ohio (OH)
  3911338,    # Oklahoma (OK)
  4028977,    # Oregon (OR)
  12802503,   # Pennsylvania (PA)
  1056298,    # Rhode Island (RI)
  4896146,    # South Carolina (SC)
  858469,     # South Dakota (SD)
  6600299,    # Tennessee (TN)
  27469114,   # Texas (TX)
```

```r
  2995919,   # Utah (UT)
  8382993,   # Virginia (VA)
  626042,    # Vermont (VT)
  7170351,   # Washington (WA)
  5771337,   # Wisconsin (WI)
  1844128,   # West Virginia (WV)
  586107     # Wyoming (WY)
),
    population_2020 = c(
  731545,    # Alaska (AK)
  4921532,   # Alabama (AL)
  3013756,   # Arkansas (AR)
  7276316,   # Arizona (AZ)
  39538223,  # California (CA)
  5773714,   # Colorado (CO)
  3605944,   # Connecticut (CT)
  689545,    # District of Columbia (DC)
  989948,    # Delaware (DE)
  21538187,  # Florida (FL)
  10711908,  # Georgia (GA)
  1455271,   # Hawaii (HI)
  3190369,   # Iowa (IA)
  1839106,   # Idaho (ID)
  12812508,  # Illinois (IL)
  6785528,   # Indiana (IN)
  2937880,   # Kansas (KS)
  4505836,   # Kentucky (KY)
  4648794,   # Louisiana (LA)
  7029917,   # Massachusetts (MA)
  6177224,   # Maryland (MD)
  1362359,   # Maine (ME)
  10077331,  # Michigan (MI)
  5706494,   # Minnesota (MN)
  6154913,   # Missouri (MO)
  2961279,   # Mississippi (MS)
  1084225,   # Montana (MT)
  10439388,  # North Carolina (NC)
  779094,    # North Dakota (ND)
  1961504,   # Nebraska (NE)
  1371246,   # New Hampshire (NH)
  9288994,   # New Jersey (NJ)
  2117522,   # New Mexico (NM)
  3104614,   # Nevada (NV)
  20201249,  # New York (NY)
  11799448,  # Ohio (OH)
  3986639,   # Oklahoma (OK)
  4237256,   # Oregon (OR)
  13002700,  # Pennsylvania (PA)
  1097379,   # Rhode Island (RI)
  5118425,   # South Carolina (SC)
```

```
  886667,    # South Dakota (SD)
  6916897,   # Tennessee (TN)
  29145505,  # Texas (TX)
  3271616,   # Utah (UT)
  8654542,   # Virginia (VA)
  643077,    # Vermont (VT)
  7705281,   # Washington (WA)
  5893718,   # Wisconsin (WI)
  1792147,   # West Virginia (WV)
  576851     # Wyoming (WY)
),
    # Extract 2024 population data
population_2024 = c(
  740133,    # Alaska (AK)
  5157699,   # Alabama (AL)
  3088354,   # Arkansas (AR)
  7582384,   # Arizona (AZ)
  39431263,  # California (CA)
  5957493,   # Colorado (CO)
  3675069,   # Connecticut (CT)
  702250,    # District of Columbia (DC)
  1051917,   # Delaware (DE)
  23372215,  # Florida (FL)
  11180878,  # Georgia (GA)
  1446146,   # Hawaii (HI)
  3241488,   # Iowa (IA)
  2001619,   # Idaho (ID)
  12710158,  # Illinois (IL)
  6924275,   # Indiana (IN)
  2970606,   # Kansas (KS)
  4588372,   # Kentucky (KY)
  4597740,   # Louisiana (LA)
  7136171,   # Massachusetts (MA)
  6263220,   # Maryland (MD)
  1405012,   # Maine (ME)
  10140459,  # Michigan (MI)
  5793151,   # Minnesota (MN)
  2943045,   # Mississippi (MS)
  6245466,   # Missouri (MO)
  1137233,   # Montana (MT)
  11046024,  # North Carolina (NC)
  796568,    # North Dakota (ND)
  2005465,   # Nebraska (NE)
  1409032,   # New Hampshire (NH)
  9500851,   # New Jersey (NJ)
  2130256,   # New Mexico (NM)
  3267467,   # Nevada (NV)
  19867248,  # New York (NY)
  11883304,  # Ohio (OH)
  4095393,   # Oklahoma (OK)
```

```r
  4272371,   # Oregon (OR)
  13078751,  # Pennsylvania (PA)
  1112308,   # Rhode Island (RI)
  5478831,   # South Carolina (SC)
  924669,    # South Dakota (SD)
  7227750,   # Tennessee (TN)
  31290831,  # Texas (TX)
  3503613,   # Utah (UT)
  8811195,   # Virginia (VA)
  648493,    # Vermont (VT)
  7958180,   # Washington (WA)
  1769979,   # West Virginia (WV)
  5960975,   # Wisconsin (WI)
  587618     # Wyoming (WY)
)


  )
# Display the population table
kable(state_population_table, caption = "State Populations for 2015, 2020, and 2024")
```

State Populations for 2015, 2020, and 2024

	state	state_name	population_2015	population_2020	population_2024
2	AK	Alaska	738432	731545	740133
1	AL	Alabama	4858979	4921532	5157699
4	AR	Arkansas	2978204	3013756	3088354
3	AZ	Arizona	6828065	7276316	7582384
5	CA	California	39144818	39538223	39431263
6	CO	Colorado	5456574	5773714	5957493
7	CT	Connecticut	3590886	3605944	3675069
51	DC	District of Columbia	672228	689545	702250
8	DE	Delaware	945934	989948	1051917
9	FL	Florida	20271272	21538187	23372215
10	GA	Georgia	10214860	10711908	11180878
11	HI	Hawaii	1431603	1455271	1446146
15	IA	Iowa	3123899	3190369	3241488

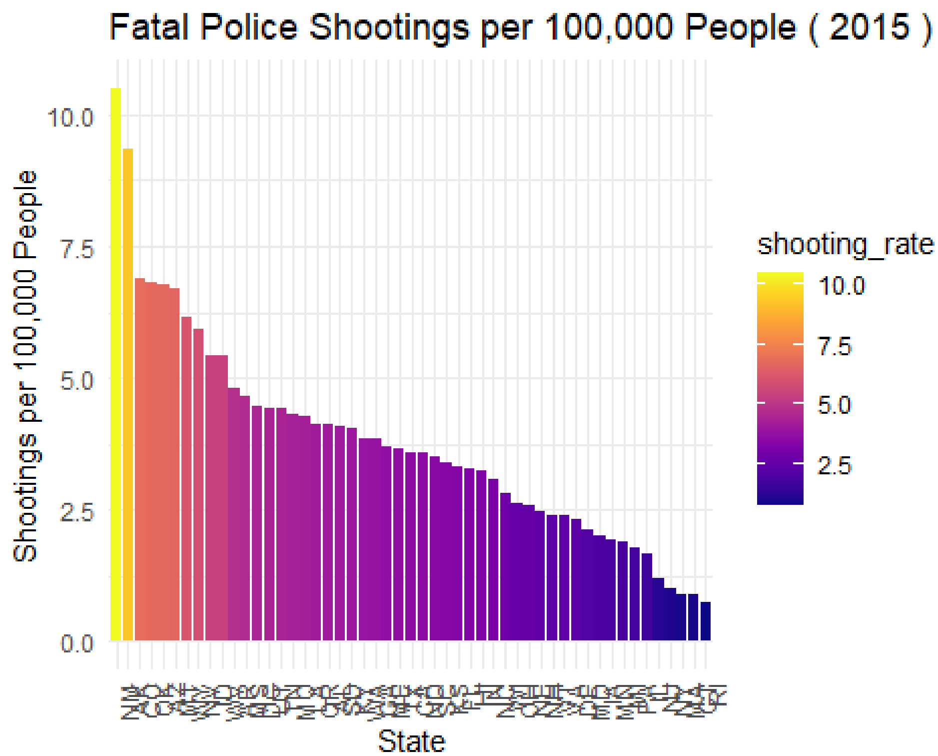
	state	state_name	population_2015	population_2020	population_2024
12	ID	Idaho	1654930	1839106	2001619
13	IL	Illinois	12859995	12812508	12710158
14	IN	Indiana	6619680	6785528	6924275
16	KS	Kansas	2911641	2937880	2970606
17	KY	Kentucky	4425092	4505836	4588372
18	LA	Louisiana	4670724	4648794	4597740
21	MA	Massachusetts	6794422	7029917	7136171
20	MD	Maryland	6006401	6177224	6263220
19	ME	Maine	1329328	1362359	1405012
22	MI	Michigan	9922576	10077331	10140459
23	MN	Minnesota	5489594	5706494	5793151
25	MO	Missouri	6083672	6154913	2943045
24	MS	Mississippi	2992333	2961279	6245466
26	MT	Montana	1032949	1084225	1137233
33	NC	North Carolina	10042802	10439388	11046024
34	ND	North Dakota	756927	779094	796568
27	NE	Nebraska	1896190	1961504	2005465
29	NH	New Hampshire	1330608	1371246	1409032
3	NJ	New Jersey	8958013	9288994	9500851

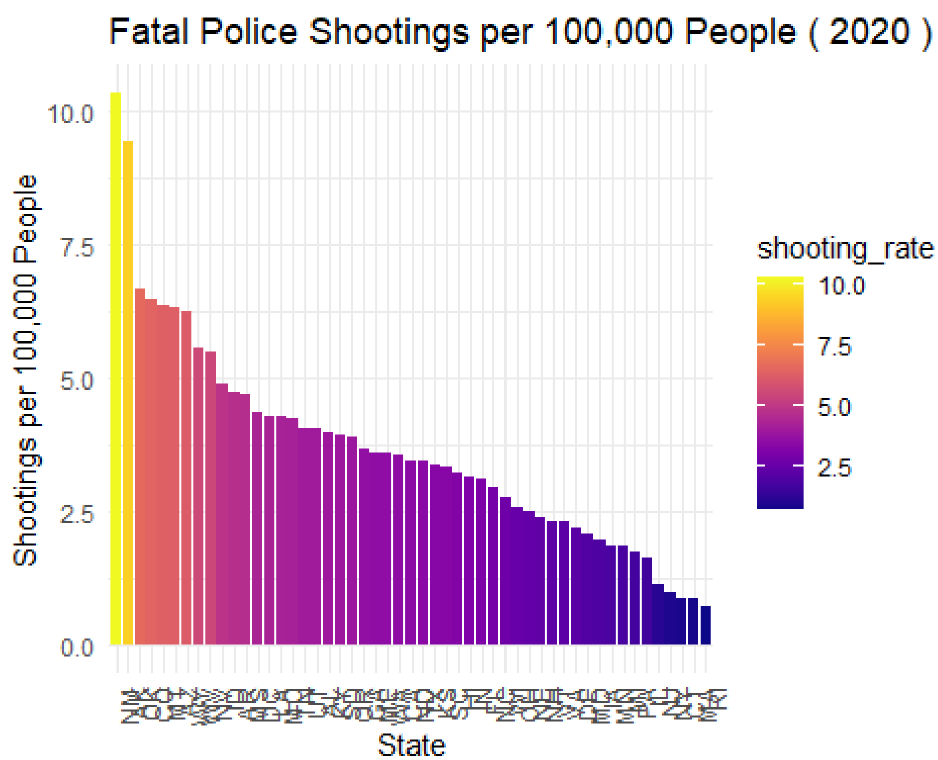
	state	state_name	population_2015	population_2020	population_2024
0					
31	NM	New Mexico	2085109	2117522	2130256
28	NV	Nevada	2890845	3104614	3267467
32	NY	New York	19795791	20201249	19867248
35	OH	Ohio	11613423	11799448	11883304
36	OK	Oklahoma	3911338	3986639	4095393
37	OR	Oregon	4028977	4237256	4272371
38	PA	Pennsylvania	12802503	13002700	13078751
39	RI	Rhode Island	1056298	1097379	1112308
40	SC	South Carolina	4896146	5118425	5478831
41	SD	South Dakota	858469	886667	924669
42	TN	Tennessee	6600299	6916897	7227750
43	TX	Texas	27469114	29145505	31290831
44	UT	Utah	2995919	3271616	3503613
46	VA	Virginia	8382993	8654542	8811195
45	VT	Vermont	626042	643077	648493
47	WA	Washington	7170351	7705281	7958180
49	WI	Wisconsin	5771337	5893718	1769979
48	WV	West Virginia	1844128	1792147	5960975

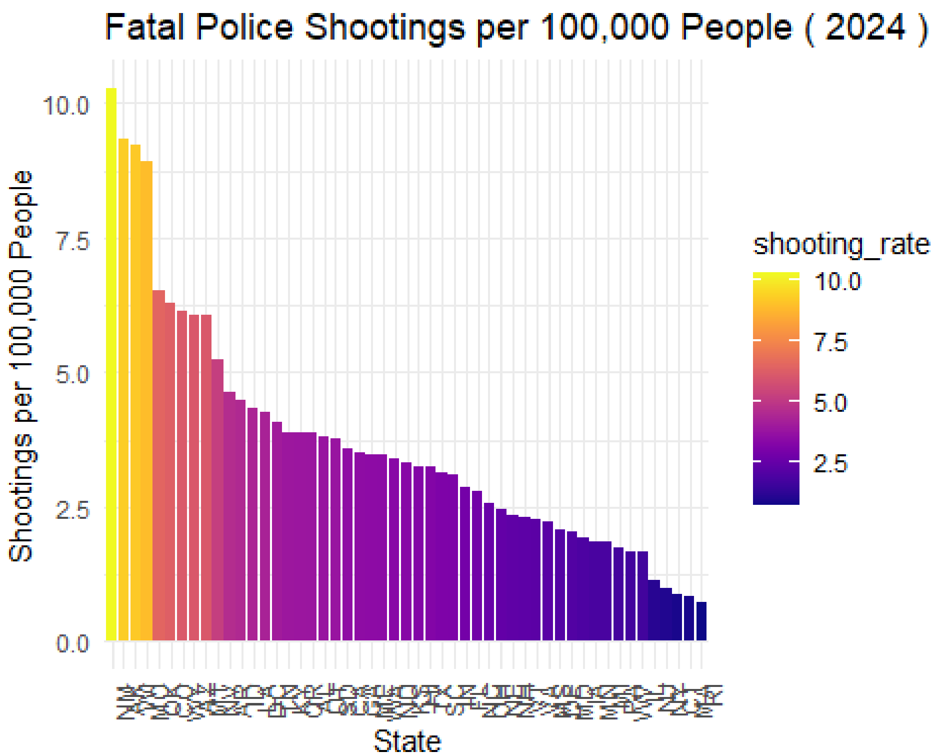
	state	state_name	population_2015	population_2020	population_2024
50	WY	Wyoming	586107	576851	587618

```
# Step 2: Aggregate shooting counts by state (including DC)
# Assuming df_final contains the shooting data with a 'state' column
state_counts <- df_final %>%
  count(state) %>%
  #print(count(state)) %>%
  rename(shooting_count = n)

# Step 3: Generate plots for each year
years <- c("2015", "2020", "2024")

for (year in years) {
  # Select the population column for the current year
  population_column <- paste0("population_", year)

  # Merge shooting counts with population data
  state_data <- state_counts %>%
    left_join(
      state_population_table[, c("state", population_column)],  # Base R
subsetting
      by = "state"
    ) %>%
    mutate(
      shooting_rate = (shooting_count / .data[[population_column]]) * 100000,
      #print(shooting_rate)
    )

  # Create a bar plot
  bar_plot <- ggplot(state_data, aes(x = reorder(state, -shooting_rate), y =
shooting_rate, fill = shooting_rate)) +
    geom_bar(stat = "identity") +
    scale_fill_viridis(option = "plasma") +
    labs(title = paste("Fatal Police Shootings per 100,000 People (", year,
")"),
         x = "State", y = "Shootings per 100,000 People") +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 90, hjust = 1))

  # Save the plot as a file (PNG)
  ggsave(
    filename = paste0("shooting_rate_", year, ".png"),
    plot = bar_plot,
    width = 10,
    height = 6,
    dpi = 300,
```

```
    bg   = "white"
  )
  print(bar_plot)
}
```


Fatal Police Shootings per 100,000 People (2015)
Shootings per 100,000 People
10.0
7.5
5.0
2.5
0.0
State
shooting_rate
10.0
7.5
5.0
2.5


Fatal Police Shootings per 100,000 People (2020)
Shootings per 100,000 People
10.0
7.5
5.0
2.5
0.0
State
shooting_rate
10.0
7.5
5.0
2.5



While California has the highest raw number of fatal police shootings, New Mexico has a higher rate per capita, and Rhode Island maintains a low rate despite its small population. Across 2015, 2020, and 2024, New Mexico consistently leads in shootings per 100,000 people, followed by Alaska and Arizona. Normalizing by population accounts for state size and density, ensuring fair comparisons and identifying areas with disproportionately high rates for targeted interventions.