Echoes of Change: Researching Policing Patterns Across Political Shifts

INFO 5371 Final Project Report

Bandar Qadan Doreen Luo Hong (Bee) Chen
Jai K. K. Chandnani Jiaqi Liu Senhuang (Pika) Cai
Wally Chang

Introduction

Research Question

What emerging trends can be identified in the proportion of searches conducted after a stop, including variations by location and reason, among different racial groups before and after the election of President Donald Trump in Nashville TN, San Jose CA, and New Orleans LA?

Motivation

The significance of scrutinizing these trends stems from an urgent need to discern whether political leadership shifts correlate with changes in law enforcement behaviors that disproportionately affect racial groups. This investigation is particularly crucial in understanding if and how the political climate surrounding the Trump election may have influenced law enforcement practices, potentially exacerbating or mitigating racial profiling and its consequent impact on social inequality.

Dataset

The datasets employed are sourced from the Stanford Policing Dataset. This comprehensive resource captures detailed records of police stops across various U.S. cities. This study's focus is narrowed to three distinct locales: Nashville, TN; San Jose, CA; and New Orleans, LA.

These datasets provide extensive details about each police stop, including demographic data about the individuals involved, the reasons for the stops, outcomes, and the exact dates and locations of these encounters.

In our study, we aim to generalize our findings across all 50 states, with a specific focus on New Orleans, LA; Nashville, TN; and San Jose, CA. These cities serve as our sample from which we intend to draw broader conclusions about nationwide policing practices.

To understand the nuances within these cities, we categorize the data into subgroups based on race, White, Hispanic, Asian/Pacific Islander, and Others. This classification allows us to analyze differential policing outcomes more holistically.

Our primary outcomes of interest include the incidence of police searches, arrests, and the reasons for stops, which we will measure and present as proportions of stops for each racial group. Specifically, we will calculate the number of searches and arrests as a proportion of stops, all segmented by race.

Our unit of analysis is the individual police stop, enabling us to analyze and compare the policing trends at a detailed level within and across the different racial subgroups in our chosen cities. This approach will provide insights into whether and how policing disparities exist based on race during different political climates.

We have the following number of observations (traffic stops) for the time-frame from 21 June 2015 to 8 November 2016 (henceforth referred to as 'Pre-Trump' and 9 November to 31 March 2018 (henceforth referred to as 'Post-Trump'). Please note, we have dropped cases that has 'subject_race' coded as either "Unknown" or "NA":

- 149,805 for New Orleans, LA
- 768,770 for Nashville, TN
- 81,201 for San Jose, CA

By dropping the cases that do not have a 'lat' and 'lng' (map coordinates) value (for visualization #2), the number of cases reduces to:

- 92,463 for New Orleans, LA
- 699,063 for Nashville, TN
- 73,406 for San Jose, CA

By dropping the cases that do not have a 'reason_for_stop' value (for visualization #3), the number of cases reduces to:

- 149,805 for New Orleans, LA
- 766,979 for Nashville, TN
- 80,437 for San Jose, CA

Methodology

Initially, we prepared the data segmented into three distinct datasets from Nashville, San Jose, and New Orleans. Given the potential complications of merging these datasets, we analyzed them separately. This preprocessing involved standardizing date formats, categorizing reasons for stops, and distinguishing data into pre- and post-Trump periods using mutate functions. We also implemented filters to remove unknown and NA values and to retain instances where searches were conducted. Our graph utilizes mapping techniques, logarithmic transformations, and the ggplot library to generate comprehensive visual representations.

We employed bar graphs to show the frequency and reasons for stops using a logarithmic scale to highlight variations across racial groups. We also included interactive map visualizations with Leaflet in R, providing geospatial insights into policing practices, complete with pop-ups for additional stop data. This combination of quantitative and qualitative visual tools helps illustrate temporal and racial trends in policing, aiming to uncover disparities and shifts in enforcement practices linked to political changes.

As an example, we explored police search rates in the three cities, comparing racial distributions and search occurrences. We charted cumulative searches to detect significant changes and analyzed search rate percentages over time by race and Trump's election impact, using a moving average for trends and a chi-squared test for statistical significance. The same type of process was applied to the other visualizations as well.

Visualizations

Visualization 1: Search Rate by Race: Urban Policing Trends Across Two Eras

```r
library(tidyverse)
library(lubridate)
library(ggplot2)
library(sf)
library(dplyr)
library(leaflet)
library(ggmap)
library(RColorBrewer)
library(lubridate)
library(htmlwidgets)
library(knitr)
library(kableExtra)
library(forcats)
library(cowplot)
library(patchwork)
```

```r
library(zoo)

#Loading the New Orleans datasets
data_path_new_orleans <- "la_new_orleans_2020_04_01.csv"
nola_raw <- read.csv(data_path_new_orleans)

#Cleaning the New Orleans datasets
nola_trump_dated <- nola_raw |>
  mutate(date = as.Date(date, format = "%Y-%m-%d")) |>
  filter(date >= as.Date("2015-06-21") & date <= as.Date("2018-03-31"),
         !is.na(subject_race) & subject_race %in% c("white",
                                                    "black",
                                                    "asian/pacific islander",
                                                    "hispanic",
                                                    "other")) |>
  mutate(
    reason_for_stop = as.factor(reason_for_stop),
    reason_for_stop = fct_lump_n(reason_for_stop, n = 10),
    period = if_else(date < as.Date("2016-11-09"),
                     "Pre-Trump",
                     "Post-Trump")
  )

#Loading the Nashville datasets
data_path_nashville <- "tn_nashville_2020_04_01.csv"
nashville_raw <- read.csv(data_path_nashville)

#Cleaning the Nashville datasets
nashville_trump_dated <- nashville_raw |>
  mutate(date = as.Date(date, format = "%Y-%m-%d")) |>
  filter(date >= as.Date("2015-06-21") & date <= as.Date("2018-03-31"),
         !is.na(subject_race) & subject_race %in% c("white",
                                                    "black",
                                                    "asian/pacific islander",
                                                    "hispanic",
                                                    "other")) |>
  mutate(
    reason_for_stop = as.factor(reason_for_stop),
    reason_for_stop = fct_lump_n(reason_for_stop, n = 10),
    period = if_else(date < as.Date("2016-11-09"),
                     "Pre-Trump",
```

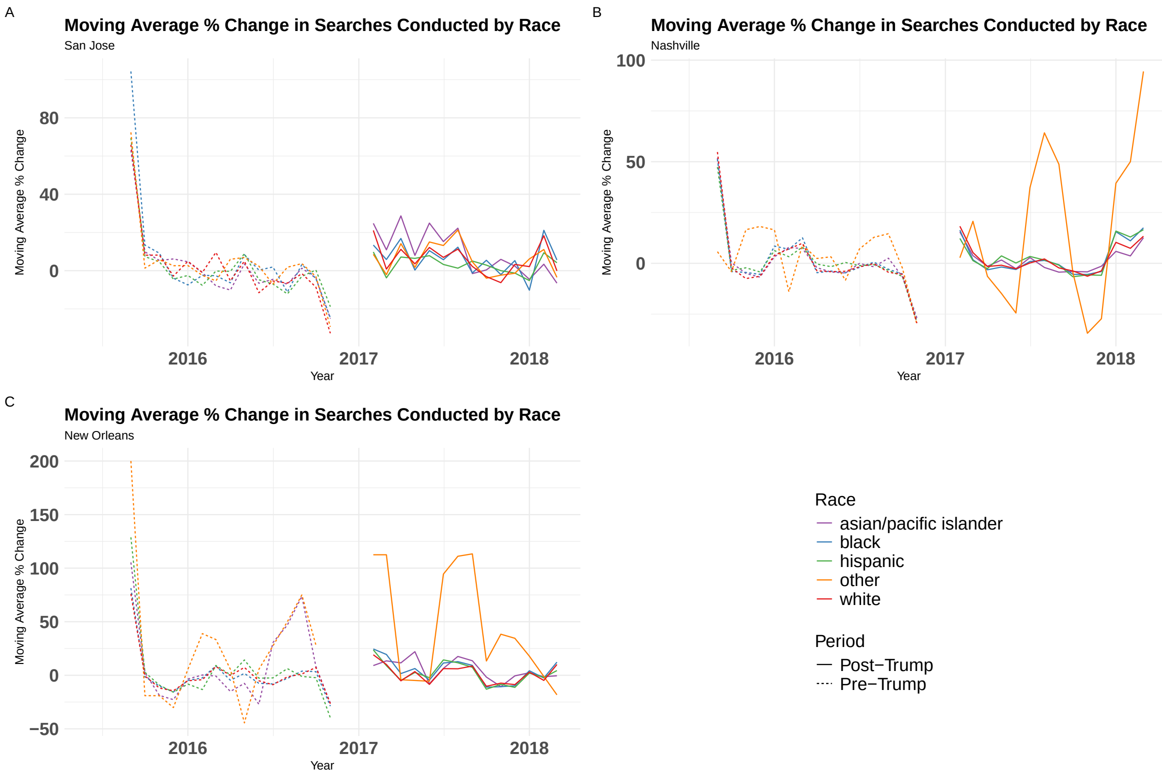
```
                      "Post-Trump")
  )

#Loading the San Jose datasets
data_path_sanjose <- "ca_san_jose_2020_04_01.csv"
san_jose_raw <- read.csv(data_path_sanjose)

#Cleaning the San Jose datasets
san_jose_trump_dated <- san_jose_raw |>
  mutate(date = as.Date(date, format = "%Y-%m-%d")) |>
  filter(date >= as.Date("2015-06-21") & date <= as.Date("2018-03-31"),
         !is.na(subject_race) & subject_race %in% c("white",
                                                    "black",
                                                    "asian/pacific islander",
                                                    "hispanic",
                                                    "other")) |>
  mutate(
    reason_for_stop = as.factor(reason_for_stop),
    reason_for_stop = fct_lump_n(reason_for_stop, n = 10),
    period = if_else(date < as.Date("2016-11-09"),
                      "Pre-Trump",
                      "Post-Trump")
  )

colors <- c("white" = "#E41A1C",
            "black" = "#377EB8",
            "hispanic" = "#4DAF4A",
            "asian/pacific islander" = "#984EA3",
            "other" = "#FF7F00")
```

```
#Calculating monthly frequency of searches conducted by race and period
monthly_frequency <- san_jose_trump_dated |>
  group_by(month = format(date, "%Y-%m"), subject_race, period) |>
  summarise(search_frequency = n()) |>
  ungroup()

#Calculating the percentage change in frequency between consecutive
#months for each race and period
monthly_frequency <- monthly_frequency |>
  group_by(subject_race, period) |>
  mutate(
```

```
    percent_change = (search_frequency - lag(search_frequency))
    / lag(search_frequency) * 100)

#Calculating the moving average of the percentage change for each race and period
window_size <- 3
monthly_frequency <- monthly_frequency |>
  group_by(subject_race, period) |>
  mutate(moving_avg_percent_change =
           rollmean(percent_change, k = window_size, fill = NA, align = "right"))

#Converting month to date format
monthly_frequency$month <- ymd(paste0(monthly_frequency$month, "-01"))

#Defining the cutoff date for filtering
cutoff_date <- as.Date("2015-06-01")

#Filtering the data to include only months after the cutoff date
monthly_frequency_filtered <- monthly_frequency |>
  filter(month >= cutoff_date)

#Plotting the results
figure1a <- ggplot(monthly_frequency_filtered, aes(
  x = month,
  y = moving_avg_percent_change,
  color = subject_race,
  linetype = period)) +
  geom_line() +
  scale_color_manual(values = colors) +
  labs(title = "Moving Average % Change in Searches Conducted by Race",
       subtitle = "San Jose",
       x = "Year", y = "Moving Average % Change",
       color = "Race", linetype = "Period") +
  theme_minimal()

#Running above analysis for Nashville
monthly_frequency <- nashville_trump_dated |>
  group_by(month = format(date, "%Y-%m"), subject_race, period) |>
  summarise(search_frequency = n()) |>
  ungroup()

monthly_frequency <- monthly_frequency |>
```

```r
  group_by(subject_race, period) |>
  mutate(percent_change = (search_frequency - lag(search_frequency))
         / lag(search_frequency) * 100)

window_size <- 3
monthly_frequency <- monthly_frequency |>
  group_by(subject_race, period) |>
  mutate(moving_avg_percent_change =
           rollmean(percent_change, k = window_size, fill = NA, align = "right"))

monthly_frequency$month <- ymd(paste0(monthly_frequency$month, "-01"))

cutoff_date <- as.Date("2015-06-01")

monthly_frequency_filtered <- monthly_frequency |>
  filter(month >= cutoff_date)

figure1b <- ggplot(monthly_frequency_filtered, aes(
  x = month,
  y = moving_avg_percent_change,
  color = subject_race,
  linetype = period)) +
  geom_line() +
  scale_color_manual(values = colors) +
  labs(title = "Moving Average % Change in Searches Conducted by Race",
       subtitle = "Nashville",
       x = "Year", y = "Moving Average % Change",
       color = "Race", linetype = "Period") +
  theme_minimal()

#Running above analysis for New Orleans
monthly_frequency <- nola_trump_dated |>
  group_by(month = format(date, "%Y-%m"), subject_race, period) |>
  summarise(search_frequency = n()) |>
  ungroup()

monthly_frequency <- monthly_frequency |>
  group_by(subject_race, period) |>
  mutate(percent_change = (search_frequency - lag(search_frequency))
         / lag(search_frequency) * 100)
```

```r
window_size <- 3
monthly_frequency <- monthly_frequency |>
  group_by(subject_race, period) |>
  mutate(moving_avg_percent_change =
           rollmean(percent_change,
                    k = window_size,
                    fill = NA,
                    align = "right"))

monthly_frequency$month <- ymd(paste0(monthly_frequency$month, "-01"))

cutoff_date <- as.Date("2015-06-01")

monthly_frequency_filtered <- monthly_frequency |>
  filter(month >= cutoff_date)

figure1c <- ggplot(monthly_frequency_filtered, aes(
  x = month,
  y = moving_avg_percent_change,
  color = subject_race,
  linetype = period)) +
  geom_line() +
  scale_color_manual(values = colors) +
  labs(title = "Moving Average % Change in Searches Conducted by Race",
       subtitle = "New Orleans",
       x = "Year", y = "Moving Average % Change",
       color = "Race", linetype = "Period") +
  theme_minimal()

#Creating a combined plot
figure1a + figure1b + figure1c +
  guide_area() +
  plot_annotation(tag_levels = 'A') +
  plot_layout(guides = 'collect') &
  theme(
    legend.text = element_text(size = 16),
    legend.title = element_text(size = 16),
    axis.text = element_text(size = 16, face = "bold"),
    plot.title = element_text(size = 16, face = "bold")
  )
```



Our visualization of a 3 month moving average, percentage change in searches conducted showed that out of the three cities, San Jose had the most obviously visually significant changes in search rate change before and after Trump's election in November of 2016. New Orleans sees a slight abnormal uptick of "asian/pacific islander" in early/mid 2017, but similarly to the "other" category, this abnormality is likely attributable to a small sample size and sensitivity to small scale changes.

When a two sample t-test was performed on these three cities between the two periods, none of them proved to be statistically significantly different between the two periods. Therefore, trends in searches and search percentage that existed before Trump's election stayed relatively similar after Trump's election - there was not a significant difference in its change post-Trump's election.

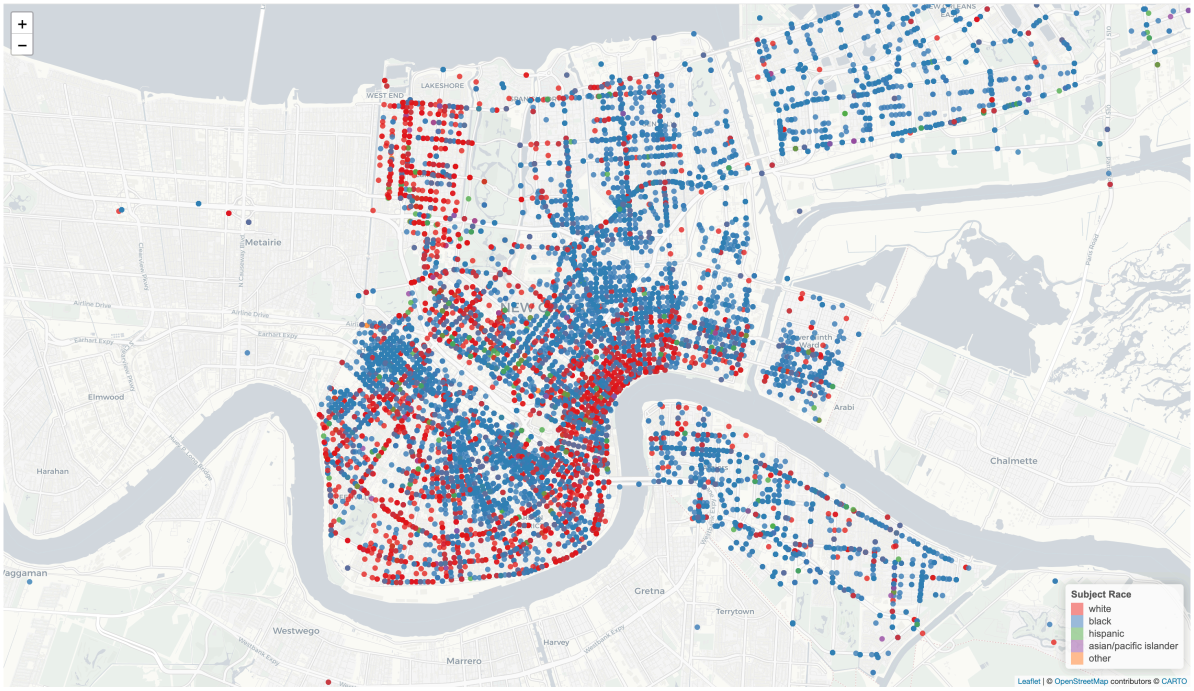
Visualization 2: Arrest Trends: New Orleans, Louisiana

```
#Preparing data for New Orleans map
stops_neworleans_clean <- nola_trump_dated |>
  filter(!is.na(lat), !is.na(lng), !is.na(arrest_made)) |>
  mutate(
    period = factor(period, levels = c("Pre-Trump", "Post-Trump")),
    subject_race = factor(subject_race, levels = names(colors))
  ) |>
  filter(!subject_race %in% c("unknown", "NA"))

#Preparing data for San Jose map
stops_san_jose_clean <- san_jose_trump_dated |>
  filter(!is.na(lat), !is.na(lng), !is.na(arrest_made)) |>
  mutate(
    period = factor(period, levels = c("Pre-Trump", "Post-Trump")),
    subject_race = factor(subject_race, levels = names(colors))
  ) |>
  filter(!subject_race %in% c("unknown", "NA"))

# Preparing data for Nashville map
stops_nashville_clean <- nashville_trump_dated |>
  filter(!is.na(lat), !is.na(lng), !is.na(arrest_made)) |>
  mutate(
    period = factor(period, levels = c("Pre-Trump", "Post-Trump")),
    subject_race = factor(subject_race, levels = names(colors))
  ) |>
  filter(!subject_race %in% c("unknown", "NA"))

#Defining the color mapping function
race_pal <- colorFactor(palette = colors, domain = names(colors))

#Defining map creation function
create_map <- function(data, period_filter, arrest_status = NULL) {
  data_filtered <- data |>
    filter(period == period_filter,
           if (!is.null(arrest_status)) arrest_made == (arrest_status == "Arrest")
           else TRUE)

#Create the leaflet map
  leaflet(data = data_filtered) |>
```

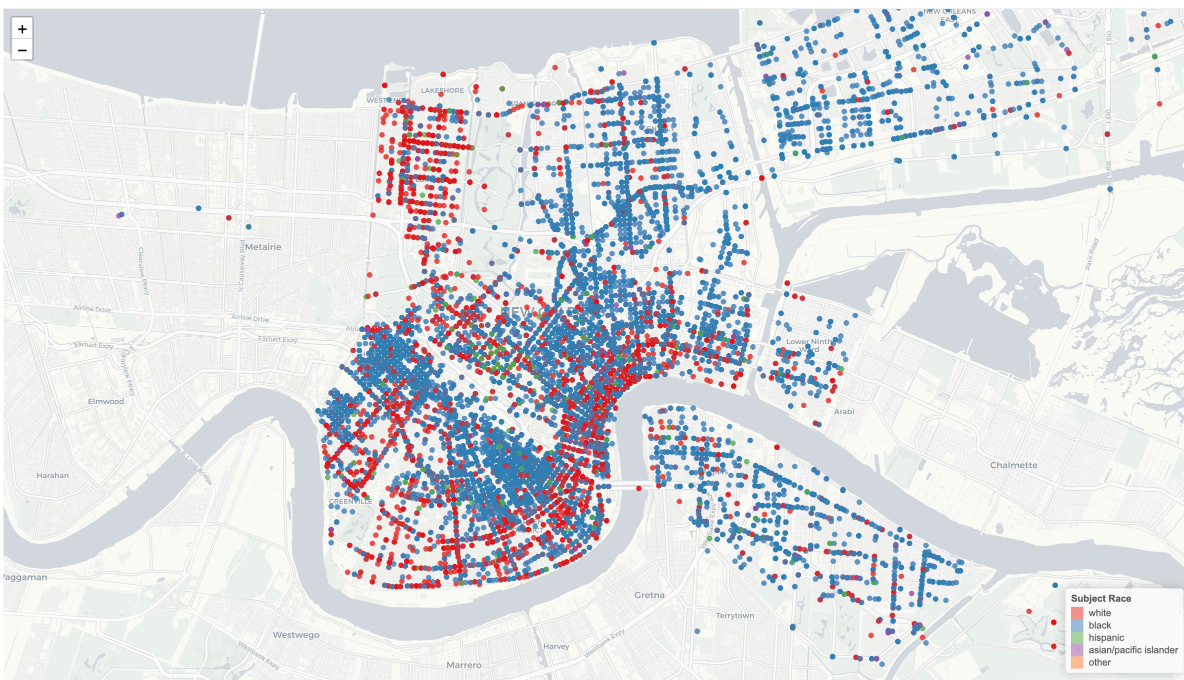
```
    addProviderTiles(providers$CartoDB.Positron) |>
    addCircleMarkers(~lng, ~lat,
                     color = ~case_when(
                       subject_race == "white" ~ "#E41A1C",
                       subject_race == "black" ~ "#377EB8",
                       subject_race == "hispanic" ~ "#4DAF4A",
                       subject_race == "asian/pacific islander" ~ "#984EA3",
                       subject_race == "other" ~ "#FF7F00",
                       TRUE ~ "#FFFFFF"  # Fallback color
                     ),
                     radius = 3.5, fillOpacity = 0.8, stroke = FALSE) |>
    addLegend(colors = colors, labels = names(colors),
              position = "bottomright", title = "Subject Race") |>
    setView(lng = avg_lng, lat = avg_lat, zoom = 13)
}

#Creating and saving the maps for New Orleans
#Calculating average coordinates
avg_lat <- mean(stops_neworleans_clean$lat, na.rm = TRUE)
avg_lng <- mean(stops_neworleans_clean$lng, na.rm = TRUE)

map_pre_trump_stops_neworleans <- create_map(stops_neworleans_clean, "Pre-Trump")
map_post_trump_stops_neworleans <- create_map(stops_neworleans_clean, "Post-Trump")
map_pre_trump_arrests_neworleans <- create_map(stops_neworleans_clean, "Pre-Trump",
                                               arrest_status = "Arrest")
map_post_trump_arrests_neworleans <- create_map(stops_neworleans_clean, "Post-Trump",
                                                arrest_status = "Arrest")

saveWidget(map_pre_trump_stops_neworleans, "Pre_Trump_Stops_Map_Neworleans.html")
saveWidget(map_post_trump_stops_neworleans, "Post_Trump_Stops_Map_Neworleans.html")
saveWidget(map_pre_trump_arrests_neworleans, "Pre_Trump_Arrests_Map_Neworleans.html")
saveWidget(map_post_trump_arrests_neworleans, "Post_Trump_Arrests_Map_Neworleans.html")


#Creating and saving the maps for San Jose
#Calculating average coordinates
avg_lat <- mean(stops_san_jose_clean$lat, na.rm = TRUE)
avg_lng <- mean(stops_san_jose_clean$lng, na.rm = TRUE)

map_pre_trump_stops_sanjose <- create_map(stops_san_jose_clean, "Pre-Trump")
map_post_trump_stops_sanjose <- create_map(stops_san_jose_clean, "Post-Trump")
```

```r
map_pre_trump_arrests_sanjose <- create_map(stops_san_jose_clean, "Pre-Trump",
                                            arrest_status = "Arrest")
map_post_trump_arrests_sanjose <- create_map(stops_san_jose_clean, "Post-Trump",
                                             arrest_status = "Arrest")

saveWidget(map_pre_trump_stops_sanjose, "Pre_Trump_Stops_Map_Sanjose.html")
saveWidget(map_post_trump_stops_sanjose, "Post_Trump_Stops_Map_Sanjose.html")
saveWidget(map_pre_trump_arrests_sanjose, "Pre_Trump_Arrests_Map_Sanjose.html")
saveWidget(map_post_trump_arrests_sanjose, "Post_Trump_Arrests_Map_Sanjose.html")


#Creating and saving the maps for Nashville
#Calculating average coordinates
avg_lat <- mean(stops_nashville_clean$lat, na.rm = TRUE)
avg_lng <- mean(stops_nashville_clean$lng, na.rm = TRUE)

map_pre_trump_stops_nashville <- create_map(stops_nashville_clean, "Pre-Trump")
map_post_trump_stops_nashville <- create_map(stops_nashville_clean, "Post-Trump")
map_pre_trump_arrests_nashville <- create_map(stops_nashville_clean, "Pre-Trump",
                                              arrest_status = "Arrest")
map_post_trump_arrests_nashville <- create_map(stops_nashville_clean, "Post-Trump",
                                               arrest_status = "Arrest")

saveWidget(map_pre_trump_stops_nashville, "Pre_Trump_Stops_Map_Nashville.html")
saveWidget(map_post_trump_stops_nashville, "Post_Trump_Stops_Map_Nashville.html")
saveWidget(map_pre_trump_arrests_nashville, "Pre_Trump_Arrests_Map_Nashville.html")
saveWidget(map_post_trump_arrests_nashville, "Post_Trump_Arrests_Map_Nashville.html")
```

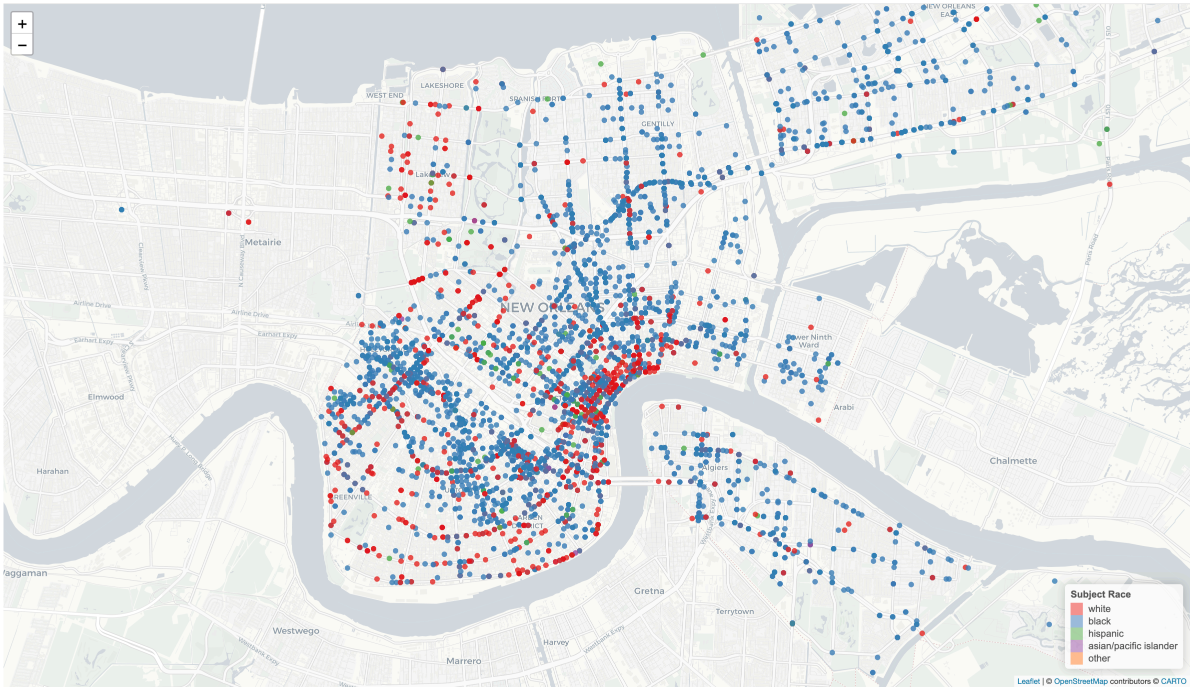
Stops Made in Pre-Trump Era in New Orleans, LA, by Race



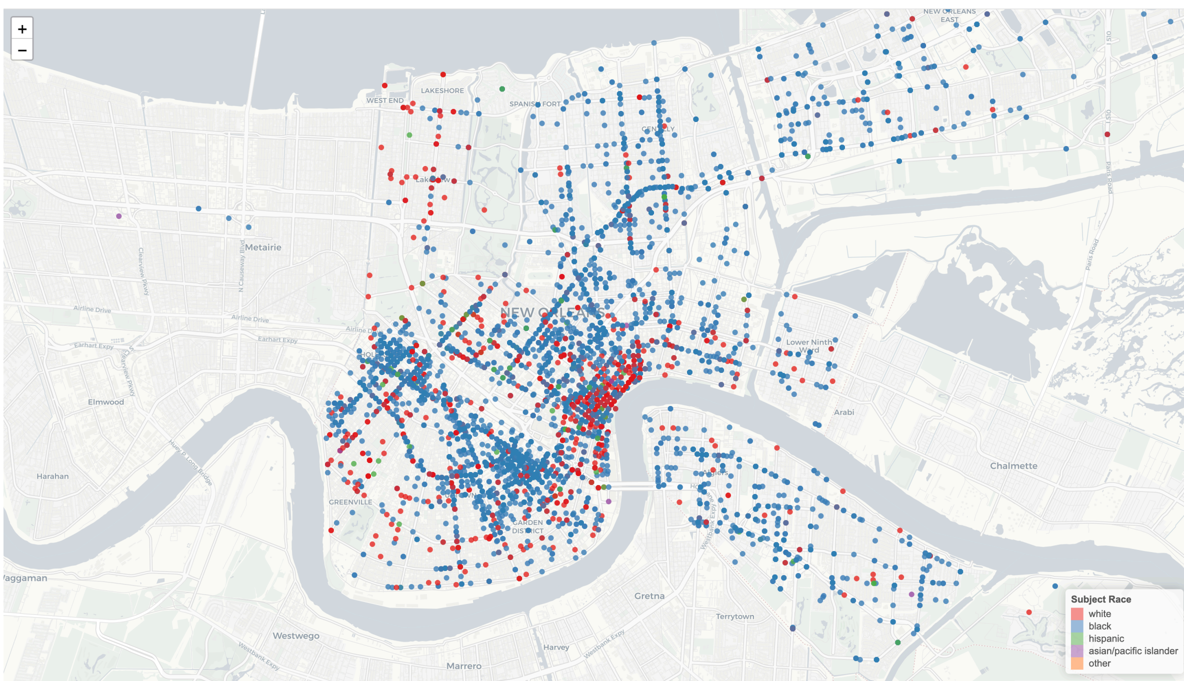
Stops Made in Post-Trump Era in New Orleans, LA, by Race



Arrests Made in Pre-Trump Era in New Orleans, LA, by Race



Arrests Made in Post-Trump Era in New Orleans, LA, by Race



Please note: The leaflet library renders interactive .html pages, allowing for further exploration and analysis. For the purposes of the report, we have attached screenshots of the maps generated.

Further, we have restricted our analysis to New Orleans.

Our visualization of New Orleans demonstrates a detailed analysis of law enforcement data segmented by race. The data reveals a profound discrepancy in stops and arrests in the periods before and after the Trump administration, highlighting the critical intersection of race and policing within this urban landscape. The visual summary exposes a concerning disproportionality in arrest rates, particularly impacting the Black community, whose arrest percentages were markedly higher than those of White, Hispanic, and Asian/Pacific Islander groups across both eras despite a nominal post-Trump decline.

Notably, Black individuals experienced a higher rate of arrests in comparison to their White, Hispanic, and Asian/Pacific Islander counterparts during both periods, with arrest percentages of 16.89% pre-Trump and 18.49% post-Trump. The data indicates a marginal increase for Black individuals in the post-Trump period. It is nominal and continues to highlight a disproportionate impact on this community. In addition, the arrest percentage for White individuals slightly increased from 13.65% to 13.77% when transitioning from the pre- to post-Trump era. Hispanic individuals also slightly increased arrests, moving from 11.02% to 11.38%. The data for Asian/Pacific Islanders suggests a converse trend with an increase in arrest percentage post-Trump. Yet, the numbers remain significantly lower than those for Black and White individuals.

This quantitative evidence suggests a continuity of social inequality within law enforcement practices in New Orleans. Despite shifts in political leadership, these figures indicated that systemic issues persist, affecting racial groups disproportionately.

Visualization 3: Stop Threads - Tied Motives Across Cities

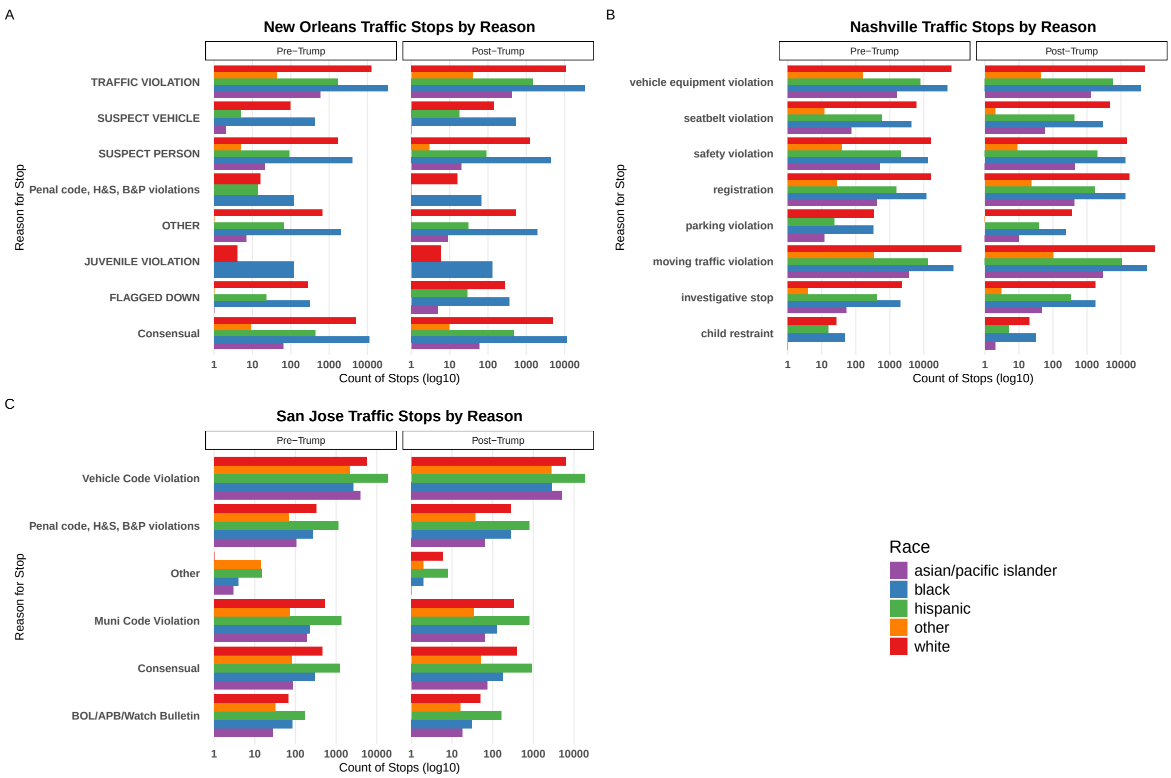
```r
#Cleaning the dataset to remove rows without Reason for Stop variable
nola_trump_dated_reasons <- nola_trump_dated |>
  filter(!is.na(reason_for_stop),
         !is.na(subject_race))

#Coding the reason_for_stop variable
nola_trump_dated_reasons <- nola_trump_dated_reasons |>
  mutate(reason_for_stop = case_when(
  str_detect(reason_for_stop, "^V") ~ "Vehicle Code Violation",
  str_detect(reason_for_stop, "^C") ~ "Consensual",
  str_detect(reason_for_stop, "^P") ~ "Penal code, H&S, B&P violations",
  TRUE ~ reason_for_stop))

#Factoring the period variable
nola_trump_dated_reasons$period <- factor(nola_trump_dated_reasons$period,
```

```r
                              levels = c("Pre-Trump", "Post-Trump"))

#Creating the plot
nola_reasons_plot <- nola_trump_dated_reasons |>
  ggplot(aes(y = reason_for_stop, fill = subject_race)) +
  geom_bar(position = position_dodge()) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    panel.grid.minor = element_blank(),
    panel.grid.major.y = element_blank(),
    strip.background = element_rect(fill="white")
  ) +
  labs(
    title = "New Orleans Traffic Stops by Reason",
    y = "Reason for Stop",
    x = "Count of Stops (log10)",
    fill = "Race"
  ) +
  scale_x_continuous(trans = "sqrt") +
  scale_x_log10(breaks = c(1, 10, 100, 1000, 10000),
                labels = c("1", "10", "100", "1000", "10000")) +
  facet_grid(~period) +
  scale_fill_manual(values = colors)



nashville_trump_dated_reasons <- nashville_trump_dated |>
  filter(!is.na(reason_for_stop),
         !is.na(subject_race))

nashville_trump_dated_reasons <- nashville_trump_dated_reasons |>
  mutate(reason_for_stop = case_when(
  str_detect(reason_for_stop, "^V") ~ "Vehicle Code Violation",
  str_detect(reason_for_stop, "^C") ~ "Consensual",
  str_detect(reason_for_stop, "^P") ~ "Penal code, H&S, B&P violations",
  TRUE ~ reason_for_stop))

nashville_trump_dated_reasons$period <- factor(nashville_trump_dated_reasons$period,
                              levels = c("Pre-Trump", "Post-Trump"))
```

```r
nashville_reasons_plot <- nashville_trump_dated_reasons |>
  ggplot(aes(y = reason_for_stop, fill = subject_race)) +
  geom_bar(position = position_dodge()) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    panel.grid.minor = element_blank(),
    panel.grid.major.y = element_blank(),
    strip.background = element_rect(fill="white")
  ) +
  labs(
    title = "Nashville Traffic Stops by Reason",
    y = "Reason for Stop",
    x = "Count of Stops (log10)",
    fill = "Race"
  ) +
  scale_x_continuous(trans = "sqrt") +
  scale_x_log10(breaks = c(1, 10, 100, 1000, 10000),
              labels = c("1", "10", "100", "1000", "10000")) +
  facet_grid(~period) +
  scale_fill_manual(values = colors)
```

```r
san_jose_trump_dated_reasons <- san_jose_trump_dated |>
  filter(!is.na(reason_for_stop),
         !is.na(subject_race))

san_jose_trump_dated_reasons <- san_jose_trump_dated_reasons |>
  mutate(reason_for_stop = case_when(
  str_detect(reason_for_stop, "^V") ~ "Vehicle Code Violation",
  str_detect(reason_for_stop, "^C") ~ "Consensual",
  str_detect(reason_for_stop, "^P") ~ "Penal code, H&S, B&P violations",
  TRUE ~ reason_for_stop))

san_jose_trump_dated_reasons$period <- factor(san_jose_trump_dated_reasons$period,
                           levels = c("Pre-Trump", "Post-Trump"))

san_jose_reasons_plot <- san_jose_trump_dated_reasons |>
  ggplot(aes(y = reason_for_stop, fill = subject_race)) +
```

```r
  geom_bar(position = position_dodge()) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    panel.grid.minor = element_blank(),
    panel.grid.major.y = element_blank(),
    strip.background = element_rect(fill="white")
  ) +
  labs(
    title = "San Jose Traffic Stops by Reason",
    y = "Reason for Stop",
    x = "Count of Stops (log10)",
    fill = "Race"
  ) +
  scale_x_continuous(trans = "sqrt") +
  scale_x_log10(breaks = c(1, 10, 100, 1000, 10000),
              labels = c("1", "10", "100", "1000", "10000")) +
  facet_grid(~period) +
  scale_fill_manual(values = colors)

#Making the combined plot
nola_reasons_plot +
  nashville_reasons_plot +
  san_jose_reasons_plot +
  guide_area() +
  plot_annotation(tag_levels = 'A') +
  plot_layout(guides = 'collect') &
  theme(
    legend.text = element_text(size = 14),
    legend.title = element_text(size = 16),
    axis.text = element_text(size = 10, face = "bold"),
    plot.title = element_text(size = 14, face = "bold")
  )
```



San Jose, CA

In the visualization of traffic stops in San Jose, the data indicates that ‘Vehicle Code Violation’ consistently ranks as the most frequent reason for traffic stops among all racial groups during both the Pre-Trump and post-Trump periods. Following close behind, ‘Penal Code, H&S, B&P violations’ emerge as the second most common cause, with the Hispanic community experiencing a significant number of these stops throughout both periods. The post-Trump era, in particular, reveals a marked increase in ‘Vehicle Code Violation’ stops for the Asian/Pacific Islander demographic, with slight upticks observed for other racial groups as well. This data is made especially clear thanks to the implementation of a logarithmic scale, which illuminates the presence of less frequent stop reasons, such as ‘Consensual’ or ‘BOL/APB/Watch Bulletin’. These categories, though less numerous, provide essential insight into the broader landscape of traffic stops.

Nashville, TN

Shifting focus to Nashville, the prevalence of ‘Moving Traffic Violation’ and ‘Vehicle Equipment Violation’ as primary stop reasons is evident, with the latter appearing disproportionately among the Black and Hispanic populations in both the Pre-Trump and post-Trump eras.

However, a notable shift occurs post-Trump with a discernible decline in 'Investigative Stop' instances across most racial groups, except for the Asian/Pacific Islander community, where it remains relatively unchanged. Furthermore, an increase in 'Seatbelt Violation' stops particularly affects the White and Black communities, with the latter seeing a rise post-Trump.

New Orleans, LA

Meanwhile, in New Orleans, 'Traffic Violation' stands out as the predominant reason for traffic stops, significantly affecting the Black community during both periods. A decrease in stops related to a 'Suspect Vehicle' is observable among the Hispanic population in the post-Trump phase. Additionally, a high frequency of stops labeled as 'Present at Crime Scene' and 'Other' suggests more frequent police interactions with the Black community for these reasons.

Pre-Trump vs. Post-Trump

Over time, the data reflects potential shifts in local law enforcement strategies. In San Jose, the increase in 'Vehicle Code Violation' stops suggests an increased traffic enforcement. For the 'Penal Code, H&S, B&P violations', the slight uptick in stops for Black and Hispanic groups post-Trump could warrant further examination into the specific enforcement of these violations. In Nashville, the reduced number of 'Investigative Stop' and 'Child Restraint' related stops could indicate changes in policing focus or improvements in community education and compliance. In contrast, New Orleans' increase in 'Traffic Violation' stops for White and Asian/Pacific Islander groups and the decrease in 'Suspect Vehicle' stops for the Hispanic community post-Trump might suggest adjustments in profiling or targeted enforcement practices.

These visualizations underscore the dynamic nature of traffic enforcement and its intersection with racial and political factors. They highlight the importance of continuous monitoring and analysis to ensure fair and equitable law enforcement across communities.

Conclusion

Findings and Limitations

Our study highlights notable shifts in policing practices across Nashville, New Orleans, and San Jose during the post-Trump era. In Nashville, arrest rates for White individuals slightly increased, while they decreased for Black and Hispanic groups, suggesting a change in enforcement policies. New Orleans continued to show higher arrest rates for Black individuals, underscoring ongoing racial disparities. San Jose experienced a general reduction in arrest rates across all groups, indicating potentially less aggressive policing. The primary reasons for stops were vehicle and traffic violations, with clear racial disparities in stop frequency and nature.

Despite its insights, the study has limitations due to its reliance on the dataset, which may not fully capture all interactions or accurately represent racial profiling instances. Additionally, focusing on only 3 cities limits the generalizability of the findings across different regions. Local policies and broader societal shifts could also influence observed outcomes, and unaccounted sociopolitical factors like local elections or national movements could skew the results.

Policy Recommendations and Final Thoughts

Based on our findings, we recommend establishing independent oversight bodies to audit police practices and mandating racial data reporting by police departments to enhance transparency. Law enforcement officers should receive extensive anti-bias and cultural sensitivity training, with regular updates on legal and human rights developments.

Our study reveals nuanced changes in enforcement affecting different racial groups. Implementing these recommendations can foster a more equitable justice system. As a stepping stone, we must continue analyzing policing data and adapting our strategies to ensure that law enforcement practices respect all individuals' rights and build trust within communities.

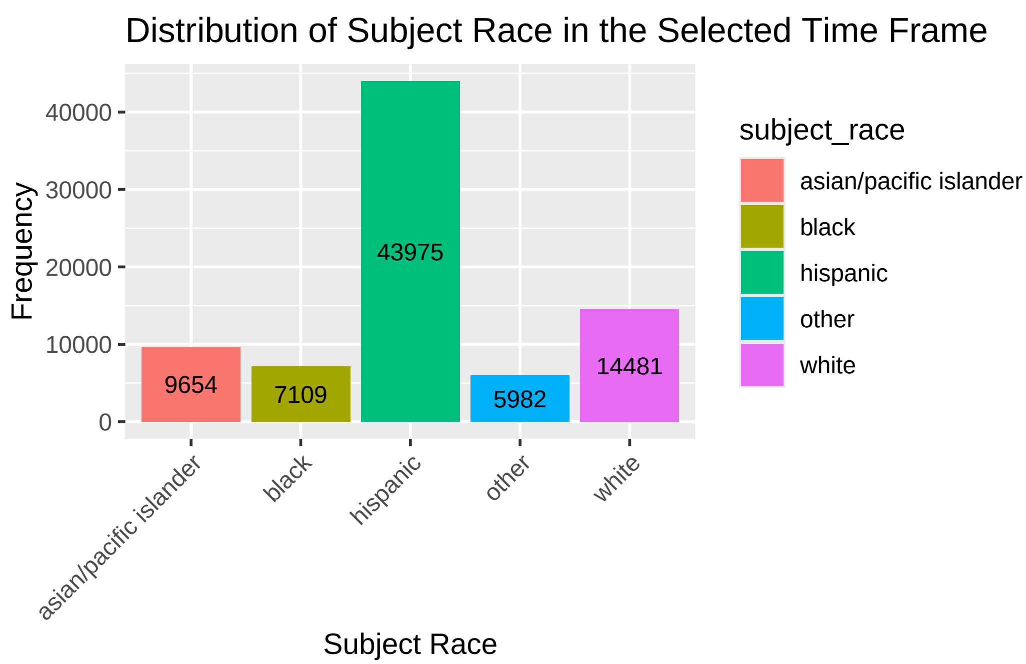
Appendix

```r
knitr::opts_chunk$set(echo = TRUE)
```

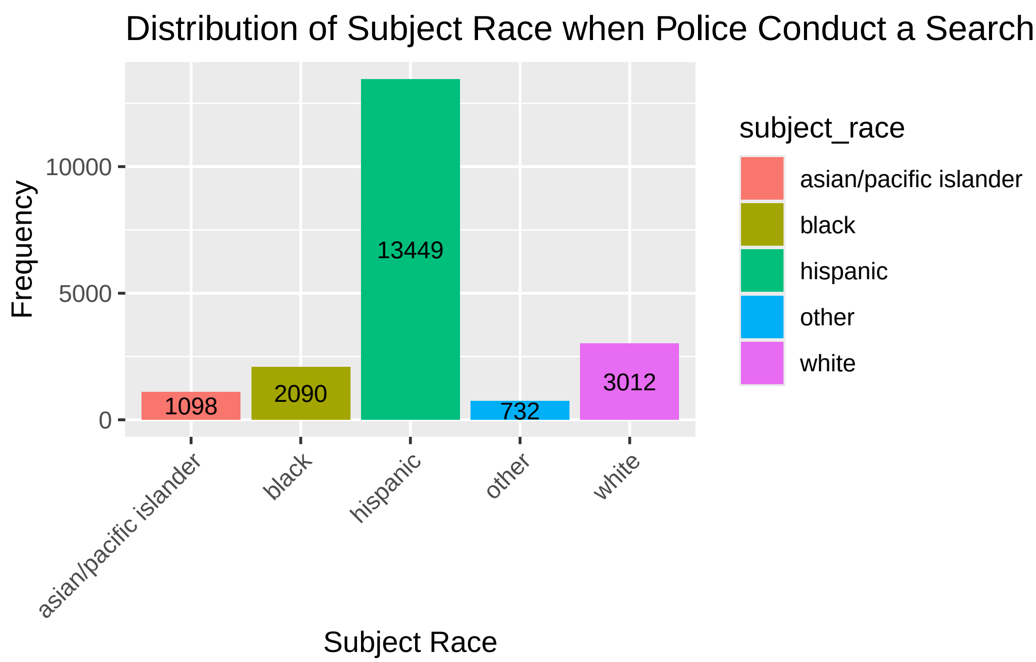
Basic Visualizations of Data and Race Distributions

```r
ggplot(san_jose_trump_dated, aes(x = subject_race,
                                 fill = subject_race)) +
  geom_bar() +
  geom_text(
    stat = "count",
    aes(label = ..count..),
    position = position_stack(vjust = 0.5),
    color = "black",
    size = 3
  ) +
theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
labs(
  title = "Distribution of Subject Race in the Selected Time Frame",
  x = "Subject Race",
  y = "Frequency"
```

```
)
```



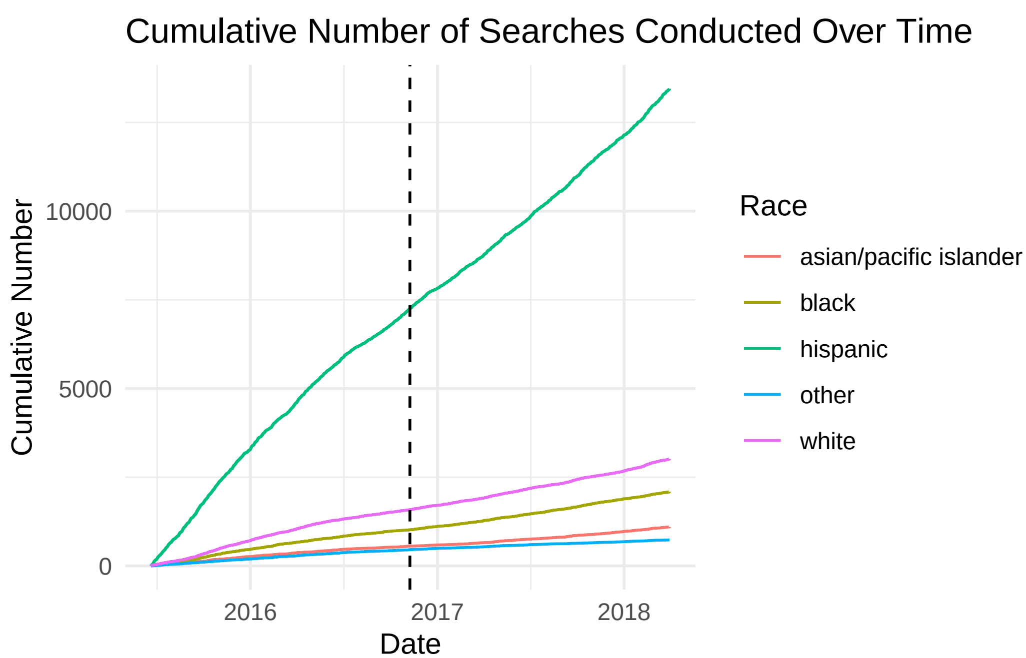

```
san_jose_search_conducted_true <- san_jose_trump_dated |>
  filter(search_conducted == TRUE)

ggplot(san_jose_search_conducted_true, aes(x = subject_race,
                                           fill = subject_race)) +
  geom_bar() +
  geom_text(
    stat = "count",
    aes(label = ..count..),
    position = position_stack(vjust = 0.5),
    color = "black",
    size = 3
  ) +
theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
labs(
  title = "Distribution of Subject Race when Police Conduct a Search",
  x = "Subject Race",
  y = "Frequency"
)
```



Examine Cumulation of Total Searches Conducted Over Time

```
date_range <- seq(min(san_jose_search_conducted_true$date),
                  max(san_jose_search_conducted_true$date),
                  by = "day")

full_data <- expand.grid(date = date_range,
                         subject_race = unique(
                           san_jose_search_conducted_true$subject_race))

cumulative_data <- left_join(full_data,
                             san_jose_search_conducted_true,
                             by = c("date", "subject_race")) |>
  group_by(subject_race) |>
  arrange(date) |>
  mutate(cumulative_number = cumsum(!is.na(search_conducted))) |>
  ungroup()

ggplot(cumulative_data, aes(x = date,
                            y = cumulative_number,
                            color = subject_race,
```

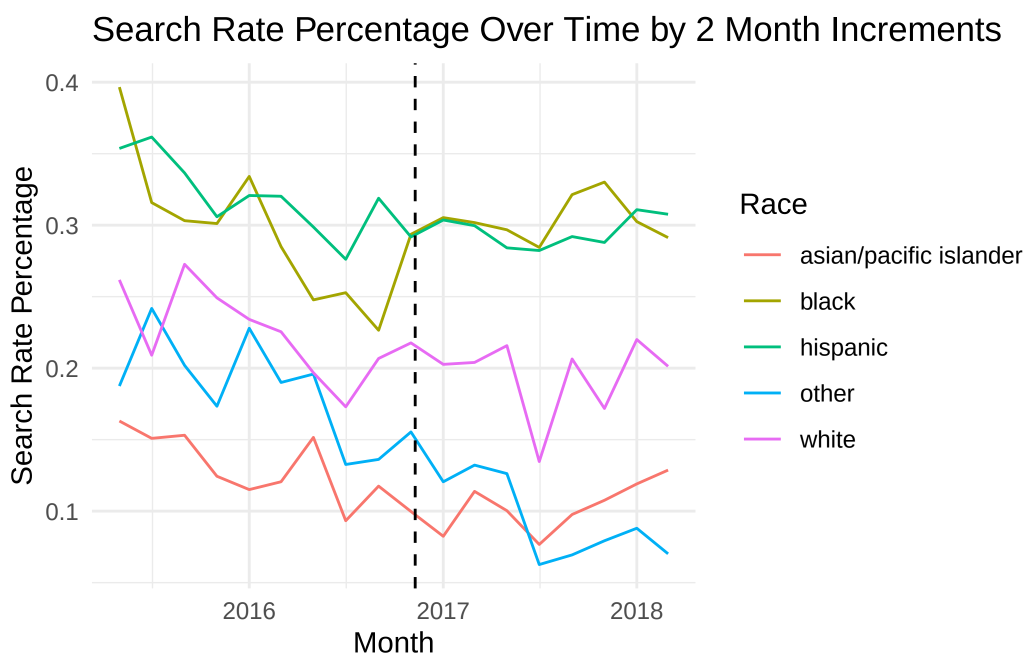
```r
                                       group = subject_race)) +
  geom_line() +
  labs(title = "Cumulative Number of Searches Conducted Over Time",
       x = "Date",
       y = "Cumulative Number",
       color = "Race") +
  theme_minimal() +
  geom_vline(xintercept = as.numeric(as.Date("2016-11-08")),
             linetype = "dashed")
```



Examine Search Rate Percentage Over Time

```r
san_jose_trump_dated_month <- san_jose_trump_dated |>
  mutate(month = floor_date(date, "2 months"))

search_rate_data <- san_jose_trump_dated_month |>
  group_by(month, subject_race) |>
  summarise(search_rate = mean(search_conducted, ,
                               na.rm = TRUE)) |>
  ungroup()
```

```r
ggplot(search_rate_data, aes(x = month,
                             y = search_rate,
                             color = subject_race,
                             group = subject_race)) +
  geom_line() +
  labs(title = "Search Rate Percentage Over Time by 2 Month Increments",
       x = "Month",
       y = "Search Rate Percentage",
       color = "Race") +
  theme_minimal() +
  geom_vline(xintercept = as.numeric(as.Date("2016-11-09")),
             linetype = "dashed")
```



Is there a statistically significant difference in the percent change of searches conducted on a monthly basis between pre-Trump and post-Trump? We will use a t-test.

```r
search_rate_data <- search_rate_data |>
  mutate(period = if_else(month < as.Date("2016-11-09"),
                         "Pre-Trump",
                         "Post-Trump"))
```

```r
average_search_rates <- search_rate_data |>
  group_by(period) |>
  summarise(mean_search_rate = mean(search_rate,
                                    na.rm = TRUE))

percent_change <-
  (average_search_rates$mean_search_rate[
    average_search_rates$period == "Post-Trump"] /
                   average_search_rates$mean_search_rate[
                     average_search_rates$period == "Pre-Trump"] - 1) * 100

monthly_frequency <- search_rate_data |>
  mutate(percent_change = percent_change)

t_test_result <- t.test(search_rate ~ period,
                        data = search_rate_data)
print(t_test_result)
```

```
	Welch Two Sample t-test

data:  search_rate by period
t = -1.7362, df = 75.7, p-value = 0.08659
alternative hypothesis: true difference in means between group Post-Trump and group Pre-Trump
95 percent confidence interval:
 -0.068810106  0.004717053
sample estimates:
mean in group Post-Trump  mean in group Pre-Trump
               0.1983271                0.2303736
```

No.

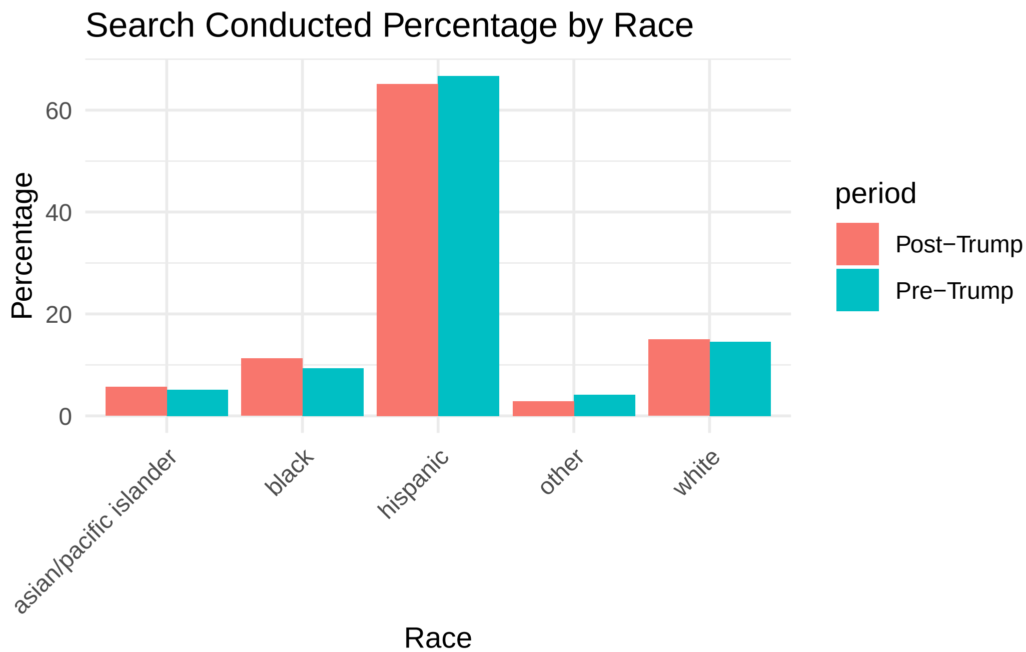
Plot Percentage of Total Searches Conducted Pre and Post Trump

```r
pre_search_percentage <- san_jose_trump_dated |>
  filter(period == "Pre-Trump", search_conducted == TRUE) |>
  group_by(period, subject_race) |>
  summarise(arrest_count = n(), .groups = 'drop') |>
  mutate(percentage = (arrest_count / sum(arrest_count)) * 100)
```

```r
post_search_percentage <- san_jose_trump_dated |>
  filter(period == "Post-Trump", search_conducted == TRUE) |>
  group_by(period, subject_race) |>
  summarise(arrest_count = n(), .groups = 'drop') |>
  mutate(percentage = (arrest_count / sum(arrest_count)) * 100)

post_plot <- ggplot(post_search_percentage, aes(x = subject_race,
                                                y = percentage,
                                                fill = "Post-Trump")) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Search Conducted Percentage by Race (Post-2017)",
       x = "Race",
       y = "Percentage") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

pre_plot <- ggplot(pre_search_percentage, aes(x = subject_race,
                                              y = percentage,
                                              fill = "Pre-Trump")) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Search Conducted Percentage by Race (Pre-2017)",
       x = "Race",
       y = "Percentage") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```r
combined_data <- rbind(
  transform(pre_search_percentage, period = "Pre-Trump"),
  transform(post_search_percentage, period = "Post-Trump")
)

overlay_plot <- ggplot(combined_data, aes(x = subject_race,
                                          y = percentage,
                                          fill = period)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Search Conducted Percentage by Race",
       x = "Race",
       y = "Percentage") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```r
print(overlay_plot)
```



Test for Significance in Association of Arrest Counts and Subject Race, Pre and Post Trump's Election

```r
if (is.character(combined_data$subject_race)) {
  combined_data$subject_race <- factor(combined_data$subject_race)
}

combined_data$subject_race <- droplevels(combined_data$subject_race)

contingency_table <- xtabs(arrest_count ~ period + subject_race,
                           data = combined_data)

print(contingency_table)
```

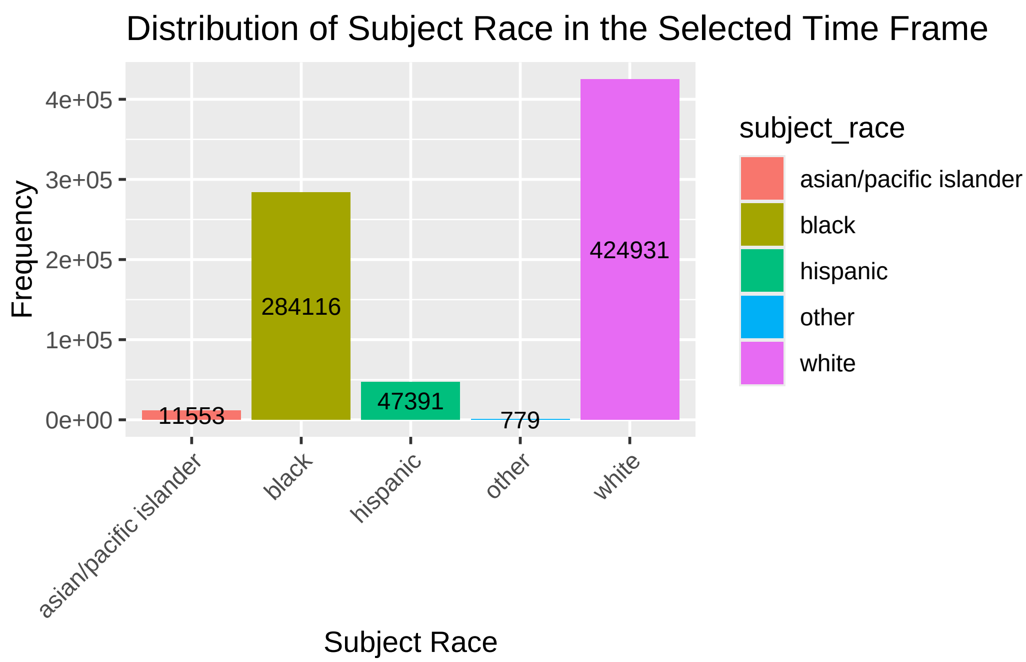
```
            subject_race
period        asian/pacific islander black hispanic other white
  Post-Trump                     540  1071     6197   277  1426
  Pre-Trump                      558  1019     7252   455  1586
```

```r
chi_square_result <- chisq.test(contingency_table)

print(chi_square_result)
```

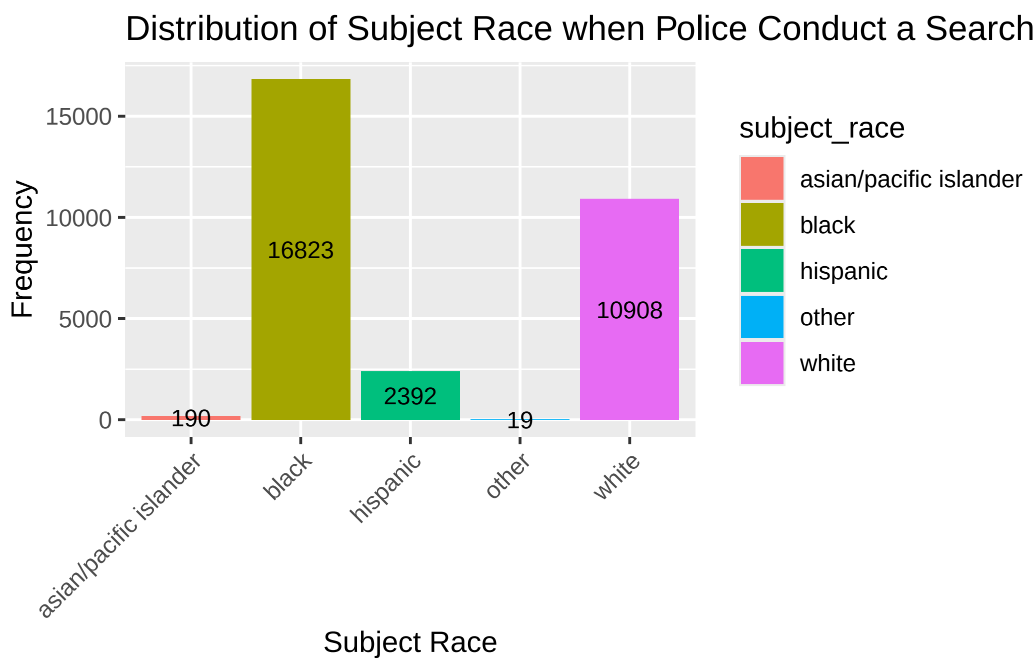
```

	Pearson's Chi-squared test

data:  contingency_table
X-squared = 45.717, df = 4, p-value = 2.821e-09
```

Using a chi-square test, we test for significant association between period and a subject's race and arrest count. With a significance level of 0.05, we reject the null hypothesis and say there is a statistically significant relationship between the period and the subject race number of arrests.

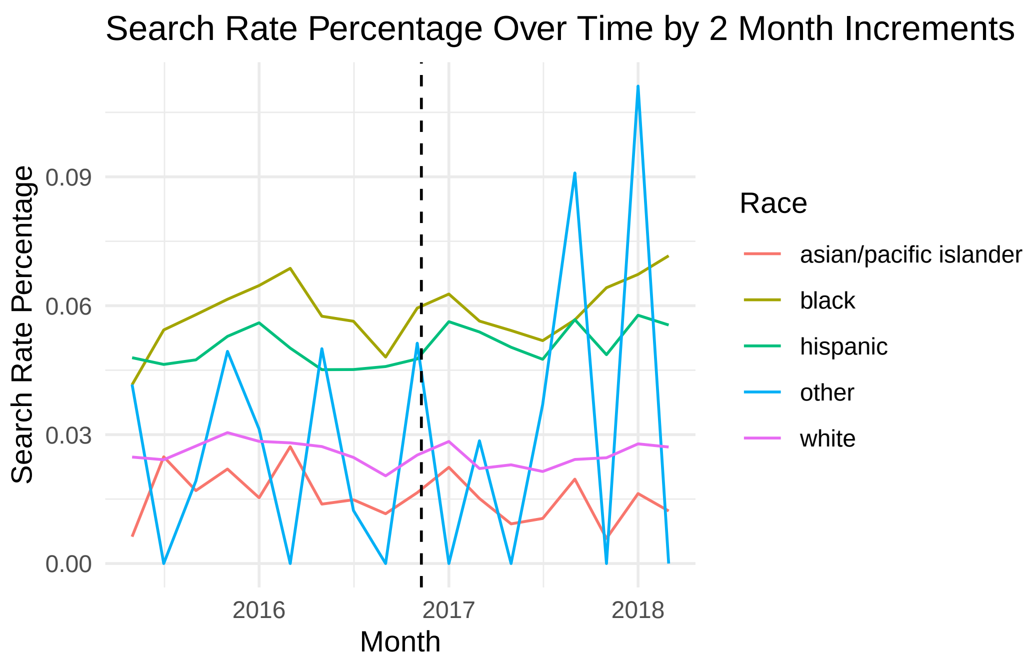
```r
ggplot(nashville_trump_dated, aes(x = subject_race,
                                  fill = subject_race)) +
  geom_bar() +
  geom_text(stat = "count", aes(label = ..count..),
            position = position_stack(vjust = 0.5),
            color = "black",
            size = 3) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Distribution of Subject Race in the Selected Time Frame",
       x = "Subject Race",
       y = "Frequency")
```



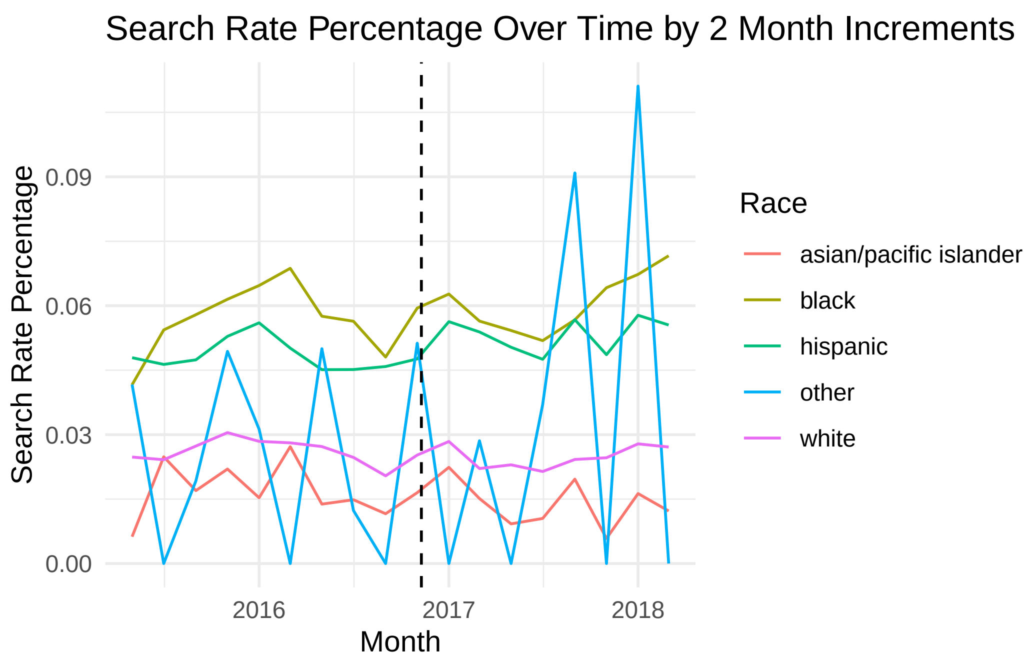

```
nashville_search_conducted_true <- nashville_trump_dated |>
  filter(search_conducted == TRUE)

ggplot(nashville_search_conducted_true, aes(x = subject_race,
                                            fill = subject_race)) +
  geom_bar() +
  geom_text(stat = "count", aes(label = ..count..),
            position = position_stack(vjust = 0.5),
            color = "black",
            size = 3) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Distribution of Subject Race when Police Conduct a Search",
       x = "Subject Race",
       y = "Frequency")
```




```r
nashville_trump_dated_month <- nashville_trump_dated |>
  mutate(month = floor_date(date, unit = "2 months"))

search_rate_data <- nashville_trump_dated_month |>
  group_by(month, subject_race) |>
  summarise(search_rate = mean(as.numeric(search_conducted), na.rm = TRUE)) |>
  ungroup()

ggplot(search_rate_data, aes(x = month,
                             y = search_rate,
                             color = subject_race,
                             group = subject_race)) +
  geom_line() +
  geom_vline(xintercept = as.numeric(as.Date("2016-11-09")),
             linetype = "dashed") +
  labs(title = "Search Rate Percentage Over Time by 2 Month Increments",
       x = "Month",
       y = "Search Rate Percentage",
       color = "Race") +
  theme_minimal()
```

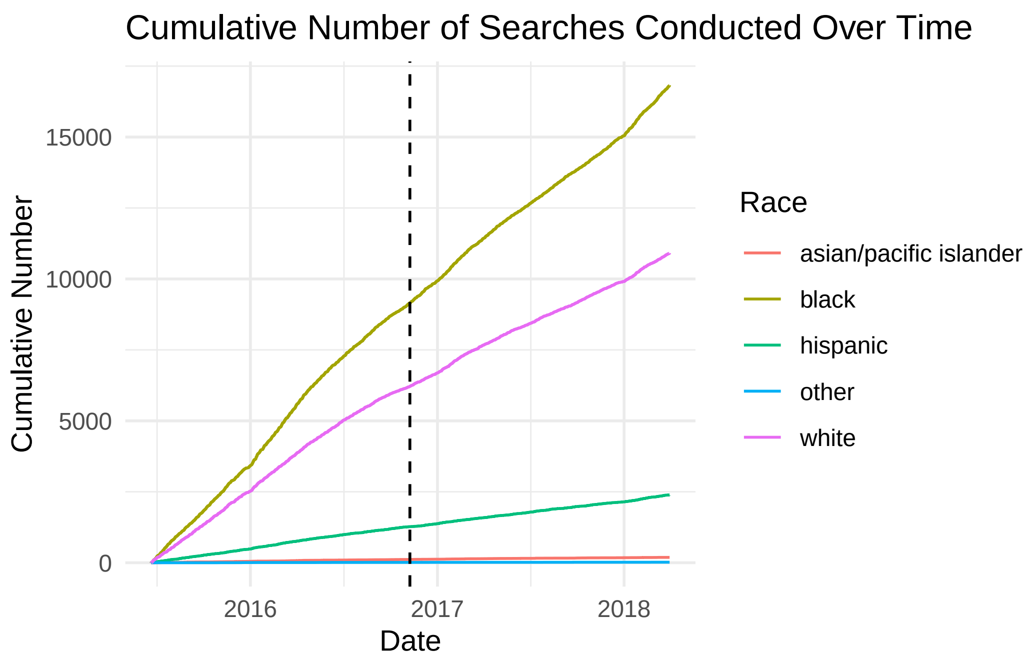



```
nashville_trump_dated_month <- nashville_trump_dated |>
  mutate(month = floor_date(date, "2 months"))

search_rate_data <- nashville_trump_dated_month |>
  group_by(month, subject_race) |>
  summarise(search_rate = mean(search_conducted, , na.rm = TRUE)) |>
  ungroup()

ggplot(search_rate_data, aes(x = month,
                             y = search_rate,
                             color = subject_race,
                             group = subject_race)) +
  geom_line() +
  labs(title = "Search Rate Percentage Over Time by 2 Month Increments",
       x = "Month",
       y = "Search Rate Percentage",
       color = "Race") +
  theme_minimal() +
  geom_vline(xintercept = as.numeric(as.Date("2016-11-09")),
             linetype = "dashed")
```




```
date_range <- seq(from = min(nashville_search_conducted_true$date),
                  to = max(nashville_search_conducted_true$date),
                  by = "day")
full_data <- expand.grid(date = date_range,
                         subject_race = unique(
                           nashville_search_conducted_true$subject_race))

cumulative_data <- left_join(full_data,
                             nashville_search_conducted_true,
                             by = c("date", "subject_race")) |>
  group_by(subject_race) |>
  arrange(date) |>
  mutate(cumulative_number = cumsum(!is.na(search_conducted))) |>
  ungroup()

ggplot(cumulative_data, aes(x = date,
                            y = cumulative_number,
                            color = subject_race,
                            group = subject_race)) +
  geom_line() +
  geom_vline(xintercept = as.numeric(as.Date("2016-11-08")),
```

```
             linetype = "dashed") +
  labs(title = "Cumulative Number of Searches Conducted Over Time",
       x = "Date",
       y = "Cumulative Number", color = "Race") +
  theme_minimal()
```

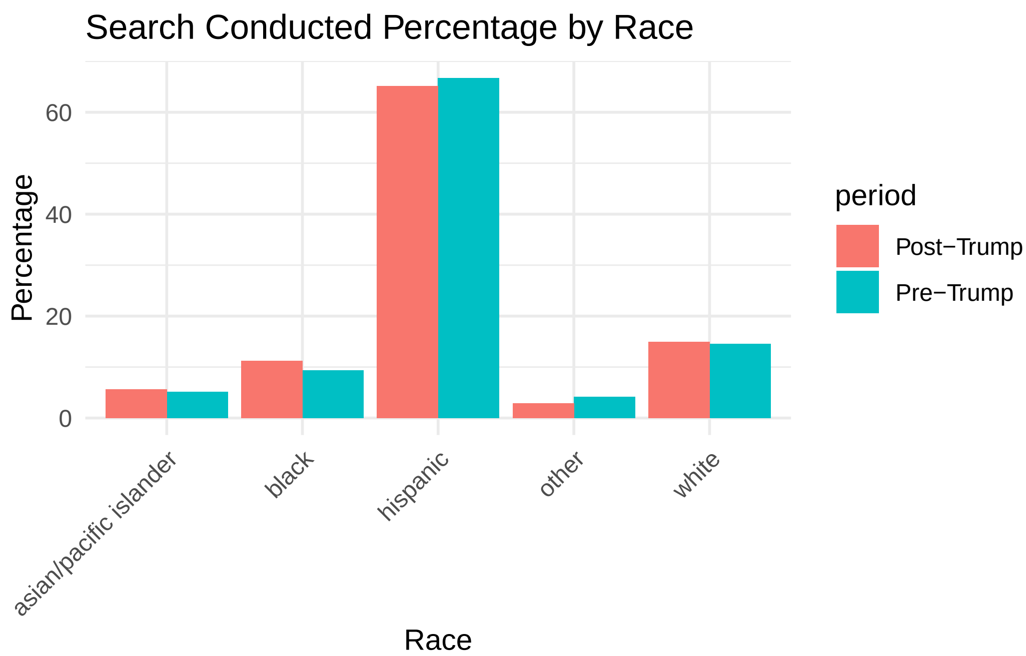



```
nashville_trump_dated <- nashville_raw |>
  mutate(
    date = as.Date(date, format = "%Y-%m-%d"),
    reason_for_stop = as.factor(reason_for_stop),
    reason_for_stop = fct_lump_n(reason_for_stop, n = 10),
    period = if_else(date <= as.Date("2016-11-08"),
                     "Pre-Trump",
                     "Post-Trump"),
    search_rate = as.numeric(search_conducted)
  ) |>
  filter(date >= as.Date("2015-06-21") & date <= as.Date("2018-03-31"),
         !is.na(subject_race), subject_race %in% c("white",
                                                   "black",
                                                   "asian/pacific islander",
                                                   "hispanic",
```

```r
                                                   "other"))

period_search_rates <- nashville_trump_dated |>
  group_by(period) |>
  summarise(mean_search_rate = mean(search_rate, na.rm = TRUE)) |>
  ungroup()

percent_change_pre <-
  period_search_rates$mean_search_rate[
    period_search_rates$period == "Pre-Trump"]
percent_change_post <-
  period_search_rates$mean_search_rate[
    period_search_rates$period == "Post-Trump"]
percent_change <-
  (percent_change_post - percent_change_pre) / percent_change_pre * 100
```

```r
pre_search_percentage <- nashville_trump_dated |>
  filter(period == "Pre-Trump", search_conducted == TRUE) |>
  group_by(period, subject_race) |>
  summarise(arrest_count = n(), .groups = 'drop') |>
  mutate(percentage = (arrest_count / sum(arrest_count)) * 100)

post_search_percentage <- nashville_trump_dated |>
  filter(period == "Post-Trump", search_conducted == TRUE) |>
  group_by(period, subject_race) |>
  summarise(arrest_count = n(), .groups = 'drop') |>
  mutate(percentage = (arrest_count / sum(arrest_count)) * 100)

combined_data$subject_race <- factor(combined_data$subject_race)

combined_data$subject_race <- droplevels(combined_data$subject_race)

overlay_plot <- ggplot(combined_data,
                       aes(x = subject_race,
                           y = percentage,
                           fill = period)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Search Conducted Percentage by Race",
       x = "Race",
       y = "Percentage") +
  theme_minimal() +
```

```
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
print(overlay_plot)
```



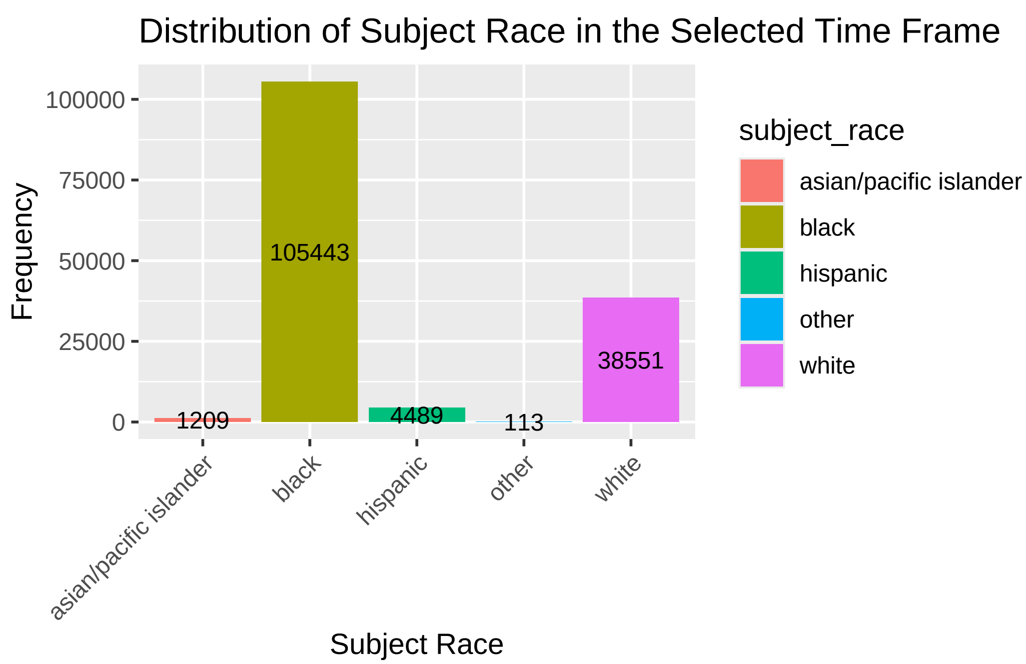

```
contingency_table <- xtabs(arrest_count ~ period +
                             subject_race,
                           data = combined_data)

print(contingency_table)
```

```
            subject_race
period       asian/pacific islander black hispanic other white
  Post-Trump                    540  1071     6197   277  1426
  Pre-Trump                     558  1019     7252   455  1586
```

```
chi_square_result <- chisq.test(contingency_table)

print(chi_square_result)
```

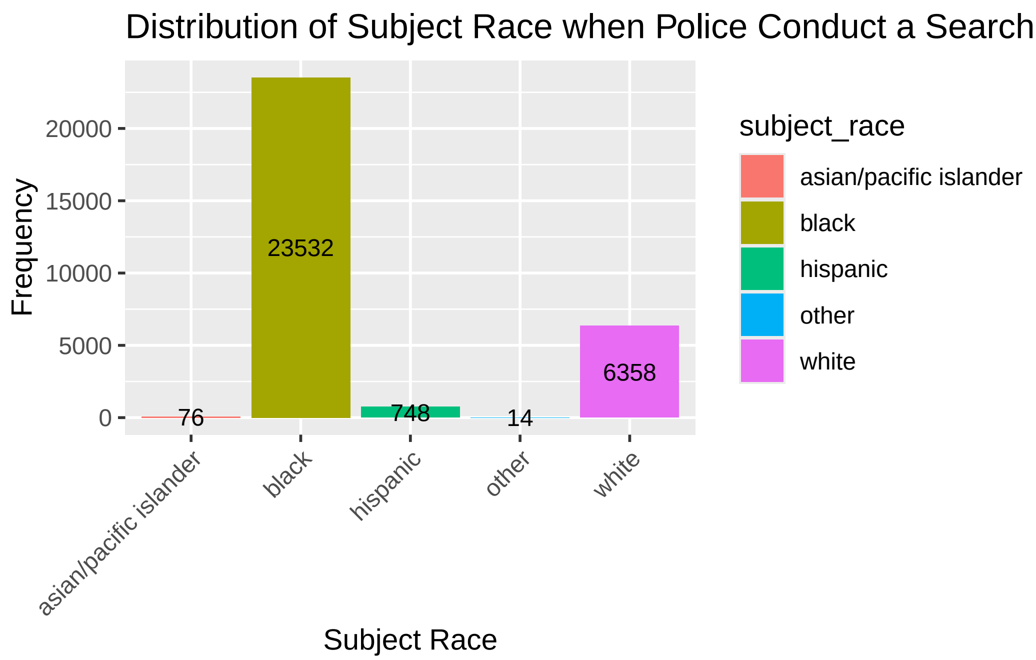
```
	Pearson's Chi-squared test

data:  contingency_table
X-squared = 45.717, df = 4, p-value = 2.821e-09
```

```
ggplot(nola_trump_dated, aes(x = subject_race,
                             fill = subject_race)) +
  geom_bar() +
  geom_text(stat = "count", aes(label = ..count..),
            position = position_stack(vjust = 0.5), size = 3) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Distribution of Subject Race in the Selected Time Frame",
       x = "Subject Race",
       y = "Frequency")
```



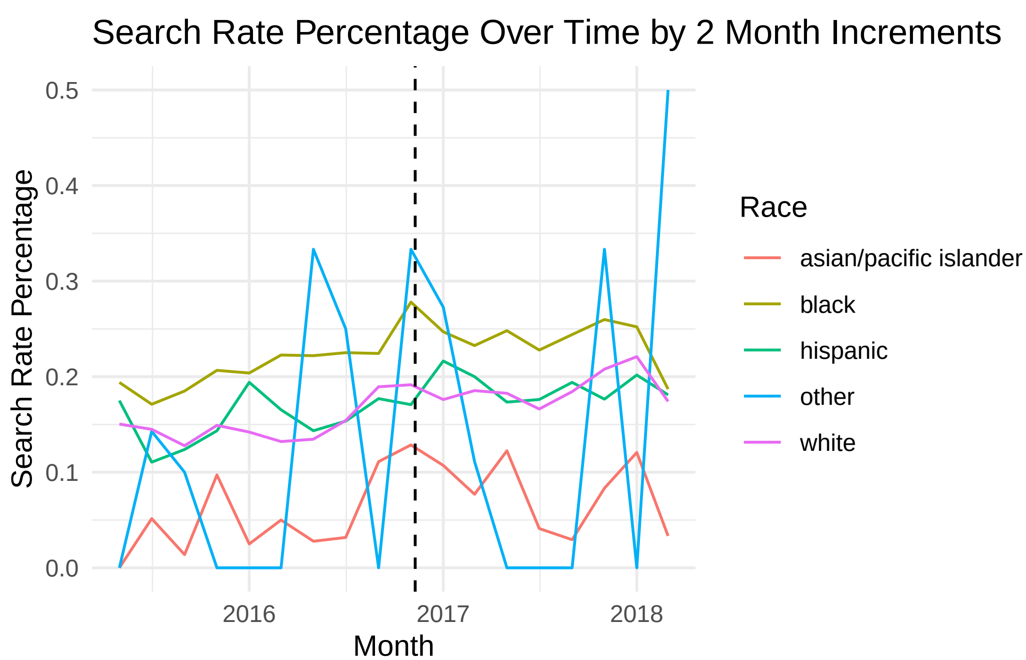

```
nola_search_conducted_true <- nola_trump_dated |>
  filter(search_conducted == TRUE)

ggplot(nola_search_conducted_true, aes(x = subject_race,
```

```
                                        fill = subject_race)) +
  geom_bar() +
  geom_text(stat = "count", aes(label = ..count..),
            position = position_stack(vjust = 0.5),
            size = 3) +
  theme(axis.text.x = element_text(angle = 45,
                                    hjust = 1)) +
  labs(title = "Distribution of Subject Race when Police Conduct a Search",
       x = "Subject Race",
       y = "Frequency")
```



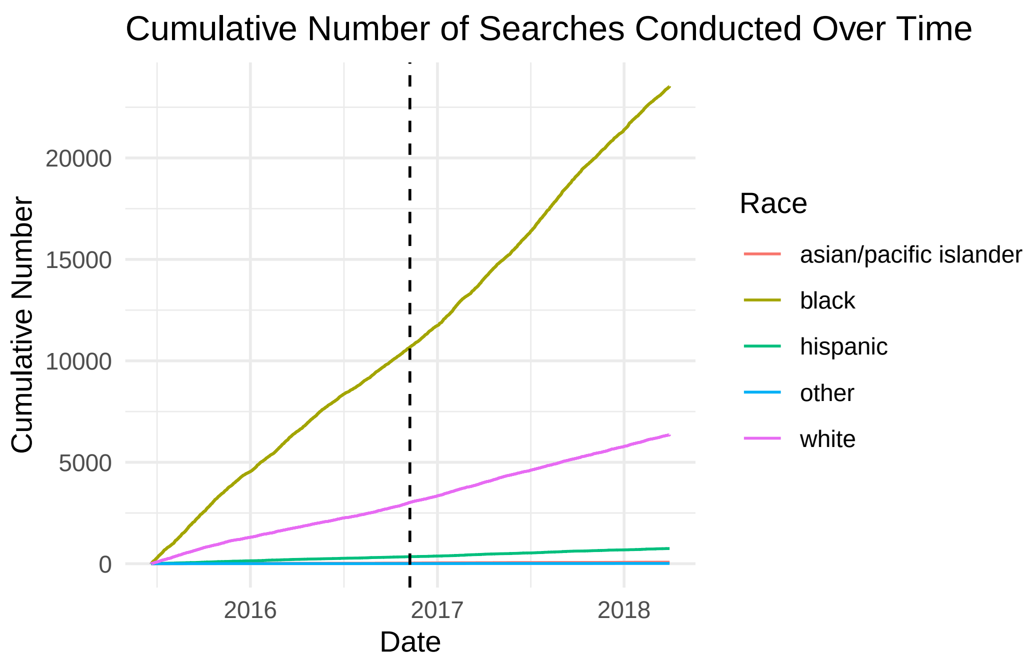

```
nola_trump_dated_month <- nola_trump_dated |>
  mutate(month = floor_date(date, "2 months"))

search_rate_data <- nola_trump_dated_month |>
  group_by(month, subject_race) |>
  summarise(search_rate = mean(as.numeric(search_conducted),
                               na.rm = TRUE)) |>
  ungroup()

ggplot(search_rate_data, aes(x = month,
```

```
                                        y = search_rate,
                                        color = subject_race,
                                        group = subject_race)) +
  geom_line() +
  geom_vline(xintercept = as.numeric(as.Date("2016-11-09")),
             linetype = "dashed") +
  labs(title = "Search Rate Percentage Over Time by 2 Month Increments",
       x = "Month",
       y = "Search Rate Percentage",
       color = "Race") +
  theme_minimal()
```

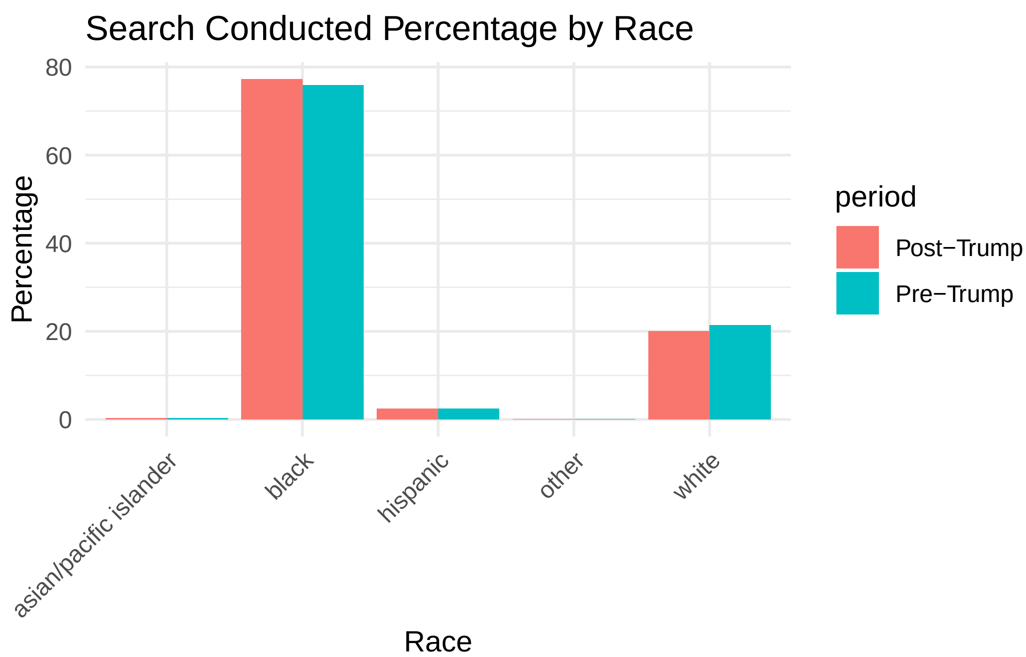



```
date_range <- seq(min(nola_search_conducted_true$date),
                  max(nola_search_conducted_true$date),
                  by = "day")
full_data <- expand.grid(date = date_range,
                         subject_race =
                           unique(nola_search_conducted_true$subject_race))

cumulative_data <- left_join(full_data,
                             nola_search_conducted_true,
```

```r
                          by = c("date", "subject_race")) |>
  group_by(subject_race) |>
  arrange(date) |>
  mutate(cumulative_number = cumsum(!is.na(search_conducted))) |>
  ungroup()

ggplot(cumulative_data, aes(x = date,
                            y = cumulative_number,
                            color = subject_race,
                            group = subject_race)) +
  geom_line() +
  geom_vline(xintercept = as.numeric(as.Date("2016-11-08")),
             linetype = "dashed") +
  labs(title = "Cumulative Number of Searches Conducted Over Time",
       x = "Date",
       y = "Cumulative Number",
       color = "Race") +
  theme_minimal()
```



```r
pre_search_percentage <- nola_trump_dated |>
  filter(period == "Pre-Trump", search_conducted == TRUE) |>
  group_by(period, subject_race) |>
  summarise(arrest_count = n(), .groups = 'drop') |>
  mutate(percentage = (arrest_count / sum(arrest_count)) * 100)

post_search_percentage <- nola_trump_dated |>
  filter(period == "Post-Trump", search_conducted == TRUE) |>
  group_by(period, subject_race) |>
  summarise(arrest_count = n(), .groups = 'drop') |>
  mutate(percentage = (arrest_count / sum(arrest_count)) * 100)

combined_data <- rbind(
  transform(pre_search_percentage, period = "Pre-Trump"),
  transform(post_search_percentage, period = "Post-Trump")
)

overlay_plot <- ggplot(combined_data, aes(x = subject_race,
                                          y = percentage,
                                          fill = period)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Search Conducted Percentage by Race",
       x = "Race",
       y = "Percentage") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
print(overlay_plot)
```



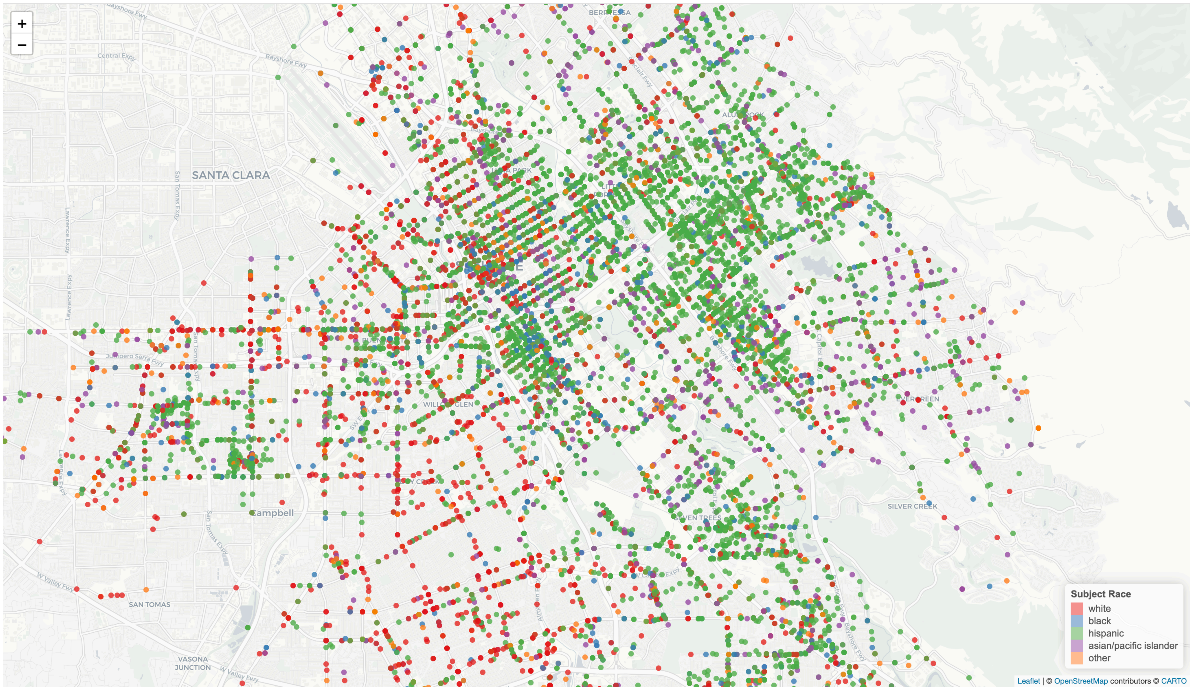

```
contingency_table <- xtabs(arrest_count ~ period +
                            subject_race,
                          data = combined_data)
print(contingency_table)
```

```
            subject_race
period        asian/pacific islander black hispanic other white
  Post-Trump                      43 12849      401     9  3337
  Pre-Trump                       33 10683      347     5  3021
```

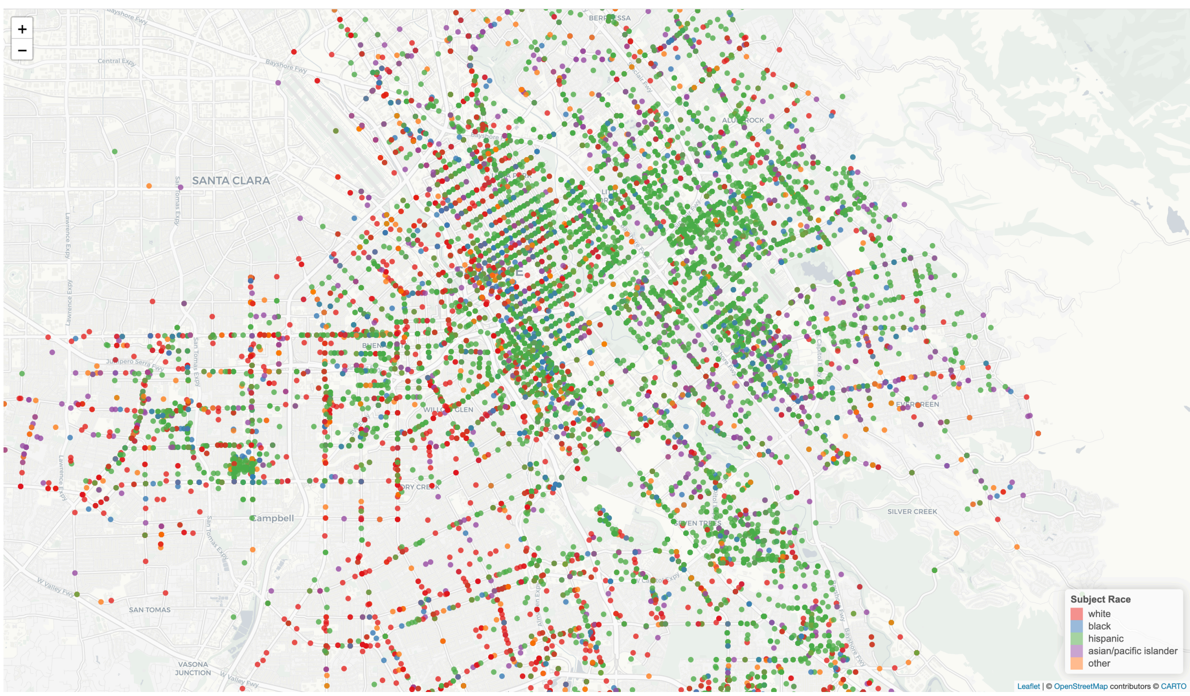
```
chi_square_result <- chisq.test(contingency_table)
print(chi_square_result)
```

```
	Pearson's Chi-squared test

data:  contingency_table
X-squared = 9.8851, df = 4, p-value = 0.04241
```

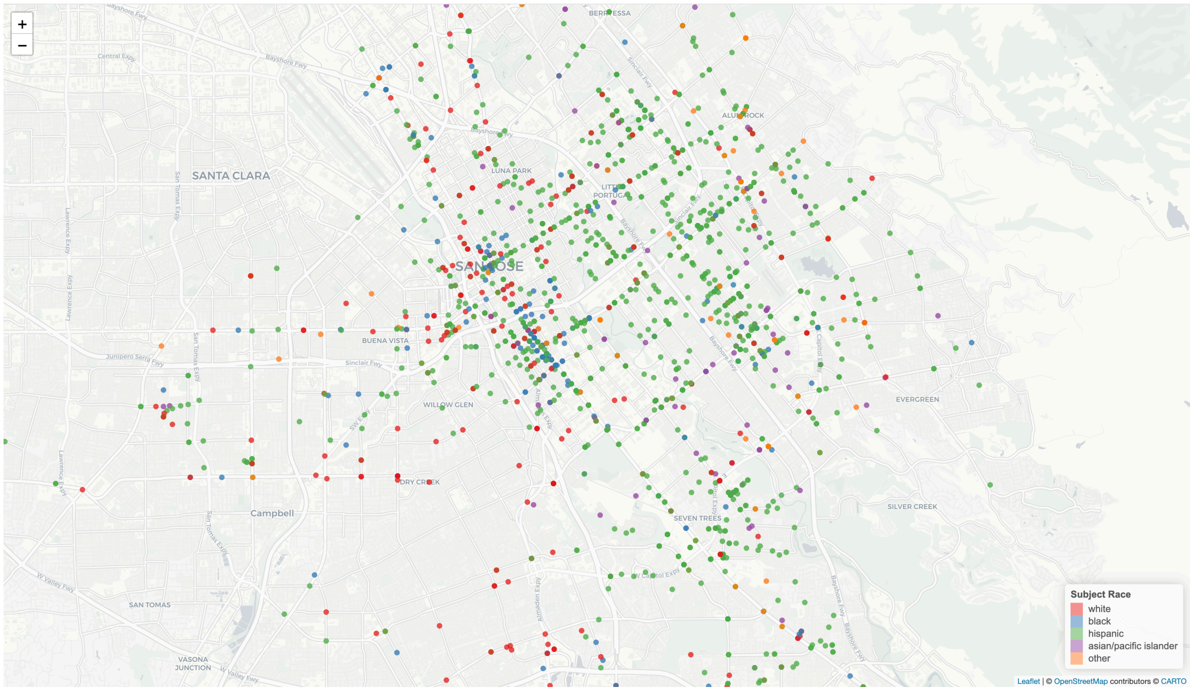
Stops Made in Pre-Trump Era in San Jose, CA, by Race



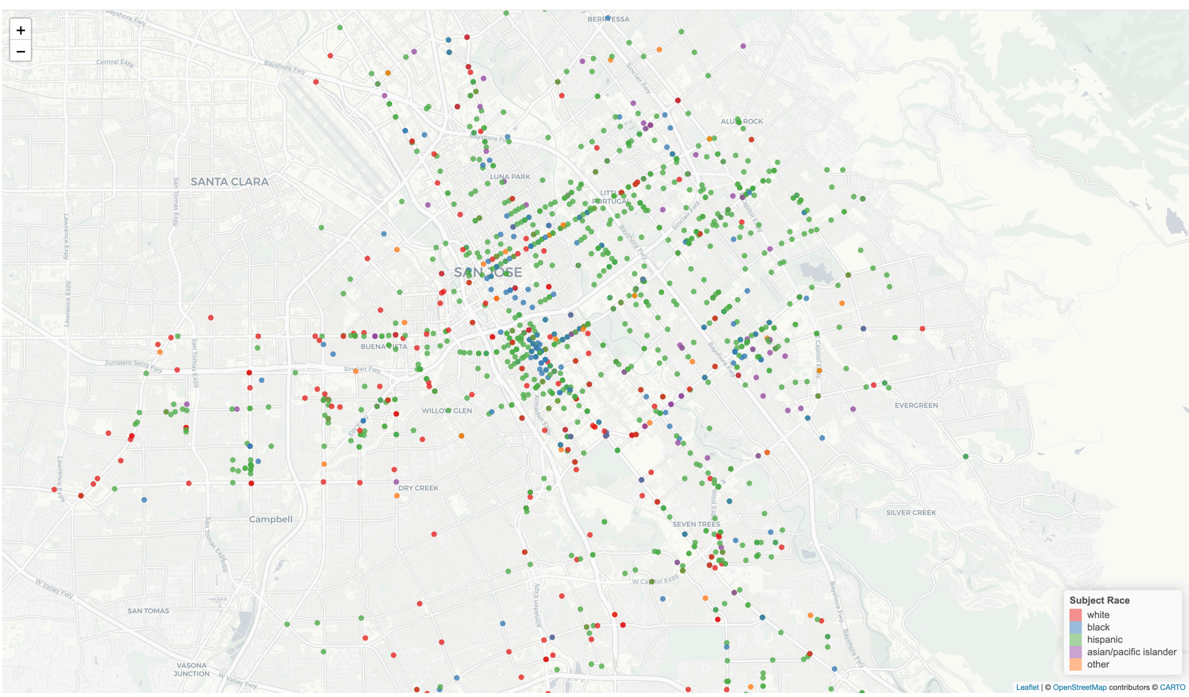
Stops Made in Post-Trump Era in San Jose, CA, by Race



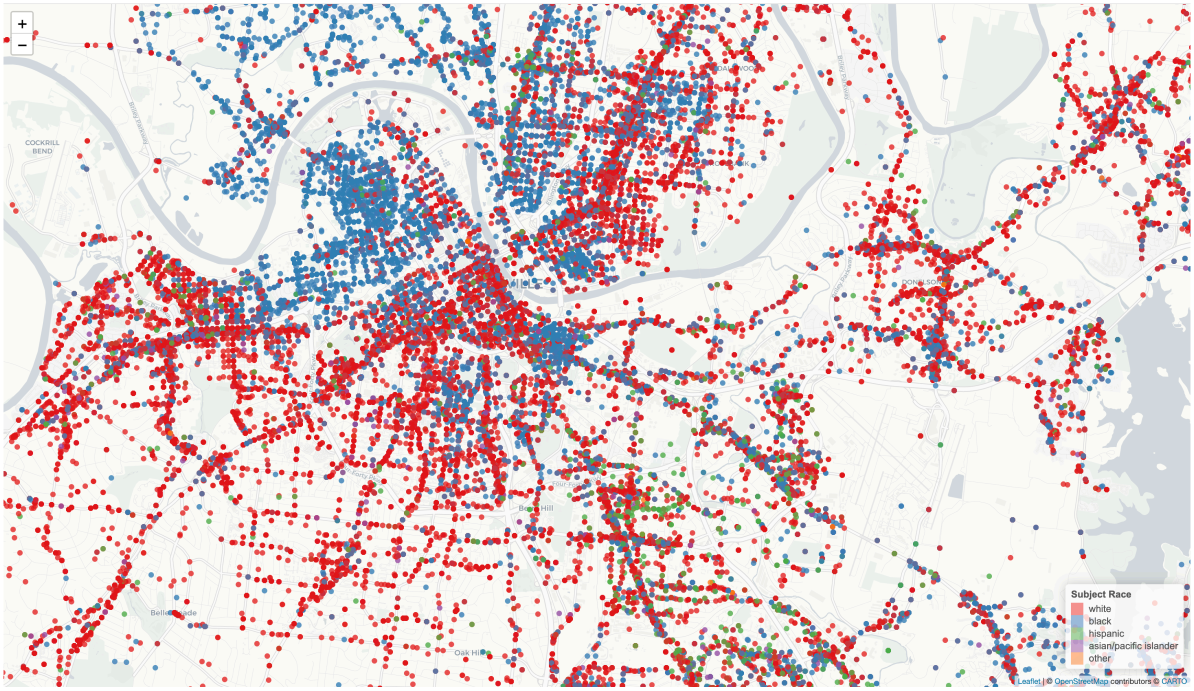
Arrests Made in Pre-Trump Era in San Jose, CA, by Race



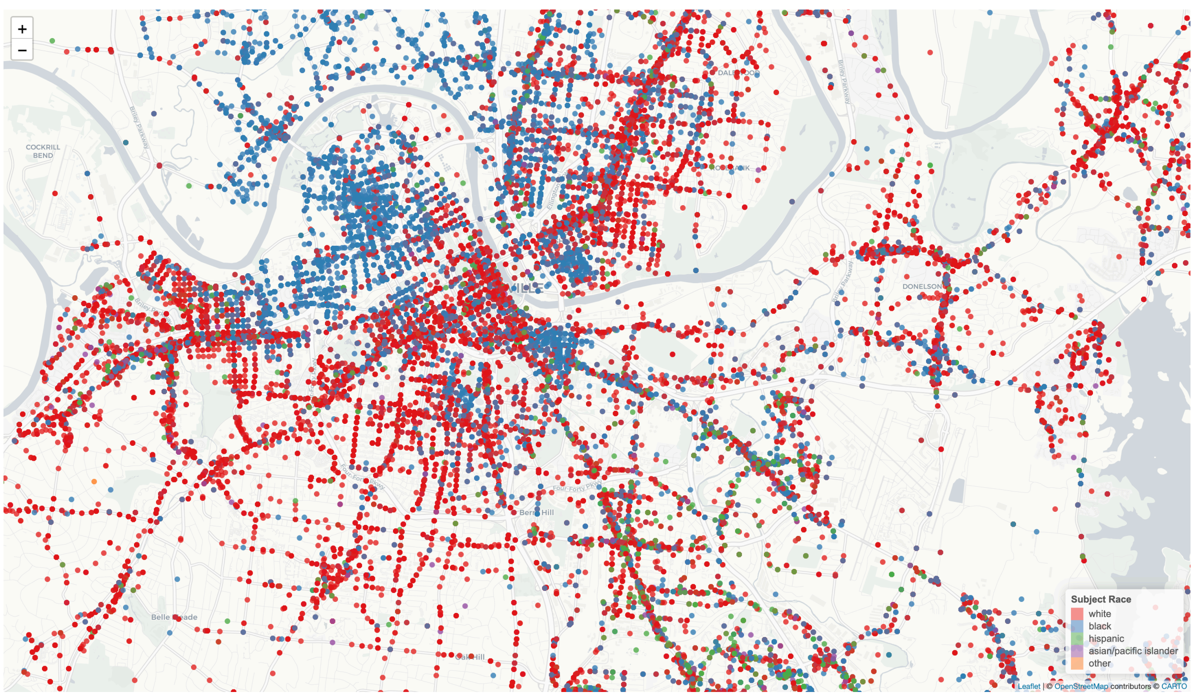
Arrests Made in Post-Trump Era in San Jose, CA, by Race



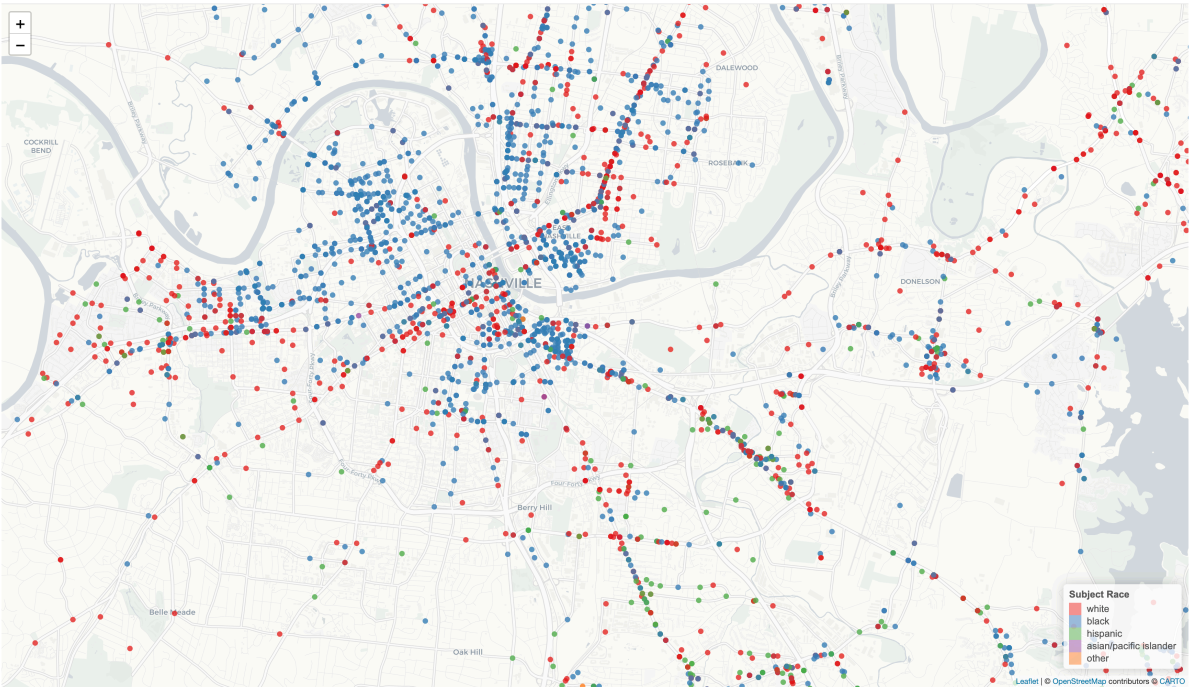
Stops Made in Pre-Trump Era in Nashville, TN, by Race



Stops Made in Post-Trump Era in Nashville, TN by Race



Arrests Made in Pre-Trump Era in Nashville, TN, by Race



Arrests Made in Post-Trump Era in Nashville, TN by Race