Lab 4: Exploratory Data Analysis

```
sqf_2011 <-
  sqf_2011 |>
  #Add a variable for weapon found
  mutate(wpnfound = case_when(pistol == 1 |
                                riflshot == 1 |
                                asltweap == 1 |
                                knifcuti == 1 |
                                machgun == 1 |
                                othrweap == 1 ~ 1,
                              TRUE ~ 0))
sqf_2011 <-
  sqf_2011 |>
  #Add a variable for arrest made or summons issued
  mutate(arrestsumm = case_when(sumissue == 1 |
                                 arstmade == 1 ~ 1,
                               TRUE ~ 0))
```

```
sqf_2011 <-
  sqf_2011 |>
  select(frisked, arrestsumm,arstmade, wpnfound, race_cat, age)
```

Calculating the number of steps

```
total_stops <-
  sqf_2011 |>
  summarize(Count = n()) |>
  pull()

total_stops
```

How many stops were there per race in 2011? What percentage of stops per race in 2011? Arrange by number of stops in descending order.

```
total_stops_race_recorded <-
  sqf_2011 |>
  #Subset to rows where race_cat is not NA or "OTHER"
  filter(race_cat != is.na(race_cat) & race_cat != "OTHER") |>
  summarize(Count = n()) |>
  pull()

sqf_2011_stops <-
  sqf_2011 |>
  #Subset to rows where race_cat is not NA or "OTHER"
  filter(race_cat != is.na(race_cat) & race_cat != "OTHER") |>
  #Group by race
  group_by(race_cat) |>
  #Calculate number of observations
  summarize(stops = n(),
            percent_stops = n() / total_stops_race_recorded * 100) |>
  #Sort by stops in descending order
  arrange(desc(stops))
```

```
total_stops_race_recorded <-
  sqf_2011 |>
  #Subset to rows where race_cat is not NA or "OTHER"
  filter(race_cat != is.na(race_cat) & race_cat != "OTHER") |>
  summarize(Count = n()) |>
  pull()

sqf_2011_analysis <-
  sqf_2011 |>
  #Subset to rows where race_cat is not NA or "OTHER"
  filter(race_cat != is.na(race_cat) & race_cat != "OTHER") |>
  #Group by race
  group_by(race_cat) |>
  #Calculate number of observations
  summarize(stops = n(),
            percent_stops = n() / total_stops_race_recorded * 100,
            percent_frisked = sum(frisked)/stops*100,
            percent_wpnfound = sum(sqf_2011$wpnfound, na.rm = TRUE)/stops*100,
            percent_arrestsumm = sum(arrestsumm)/stops*100) |>
```

```
  #Sort by stops in descending order
  arrange(desc(stops))
```

I learned that drivers in New York City are stopped and frisked by police officers disproportionately based on their race. Based on statistics, it appears that there is discrimination or bias towards certain races. since an excessive amount of Black individuals and White Hispanic individuals in New York are stopped by the police compared to other race categories such as white and American Indian/Alaskan Native.

The races have similar percentage of people in each race arrested or a summoned issued which means that there might be discrimination involved. This implies that in order to prevent police officers from discriminating against certain marginalised groups because of their own biases, the NYC Police Department needs to implement stricter and more uniform policies.

```
ggplot(data = sqf_2011_analysis,
       aes(x = race_cat, y = percent_stops))+
  geom_col(fill = "lightblue")+
  labs(x = "Race",
       y = "Percentage of people stopped per race category",
       title = "The Demographic Breakdown of People Stopped By Police In NYC In 2011",
       caption = "Source: New York City Policy Department, accessed in March, 8 2024")+
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
  theme(axis.text.x = element_text(size = 6))
```

```
ggplot(data = sqf_2011_analysis,
       aes(x = race_cat, y = percent_arrestsumm))+
  geom_col(fill = "lightpink")+
  labs(x = "Race",
       y = "% of people arrested or summons issued per race",
       title = "People Arrested or Summons Issued After Stop By Race in NYC, 2011",
       caption = "Source: New York City Policy Department, accessed in March, 8 2024")+
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
  theme(axis.text.x = element_text(size = 6))
```

I learned that drivers in New York City are stopped and frisked by police officers disproportionately based on their race. Based on statistics, it appears that there is discrimination or bias towards certain races. since an excessive amount of Black individuals and White Hispanic individuals in New York are stopped by the police compared to other race categories such as white and American Indian/Alaskan Native.

The races have similar percentage of people in each race arrested or a summoned issued which means that there might be discrimination involved. This implies that in order to prevent police officers from discriminating against certain marginalised groups because of their own biases, the NYC Police Department needs to implement stricter and more uniform policies.