Final Project

Coco Donovan

2023-05-08

Abstract:

The California Racial and Identity Profiling Act of 2015 (AB 953) mandated, among other items: law enforcement is prohibited from profiling based on race/ ethnicity or identity and that law enforcement be required to submit data on vehicle/pedestrian stops and citizen complaints alleging racial and identity profiling.

The primary dataset I analyzed comes from the reported Police Stop data from the Long Beach Police Department between Jan 1st, 2019, and Dec 31st, 2021. In this analysis, I intend to focus specifically on the differences in policing people of different ethnicities as measured by stop duration and the number of actions taken by the Long Beach Police Department during the stop.

Loading Necessary Packages

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2
## --

## v ggplot2 3.4.0     v purrr   1.0.1
## v tibble  3.1.8     v stringr 1.5.0
## v tidyr   1.3.0     v forcats 0.5.2
## v readr   2.1.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(openintro)
```

```
## Loading required package: airports
## Loading required package: cherryblossom
## Loading required package: usdata
```

```
library(infer)
library(ggplot2)
```

Data Preparation

```
lb_police_stops <- read.csv('lbpd-ripa-data-annual.csv')

# head(lb_police_stops)

# colnames(lb_police_stops)
```

Limitations: When comparing police "perceived race" with official data on ethnicity for Long Beach the comparison it is clear that there is less nuance to the police "race" data. I understand identity is far more complex than the way I have simplified it for the purpose of analysis, and for the purpose of this less extensive analysis my simplified classifications will be used to present the data.

Population of Long Beach, CA by Ethnicity

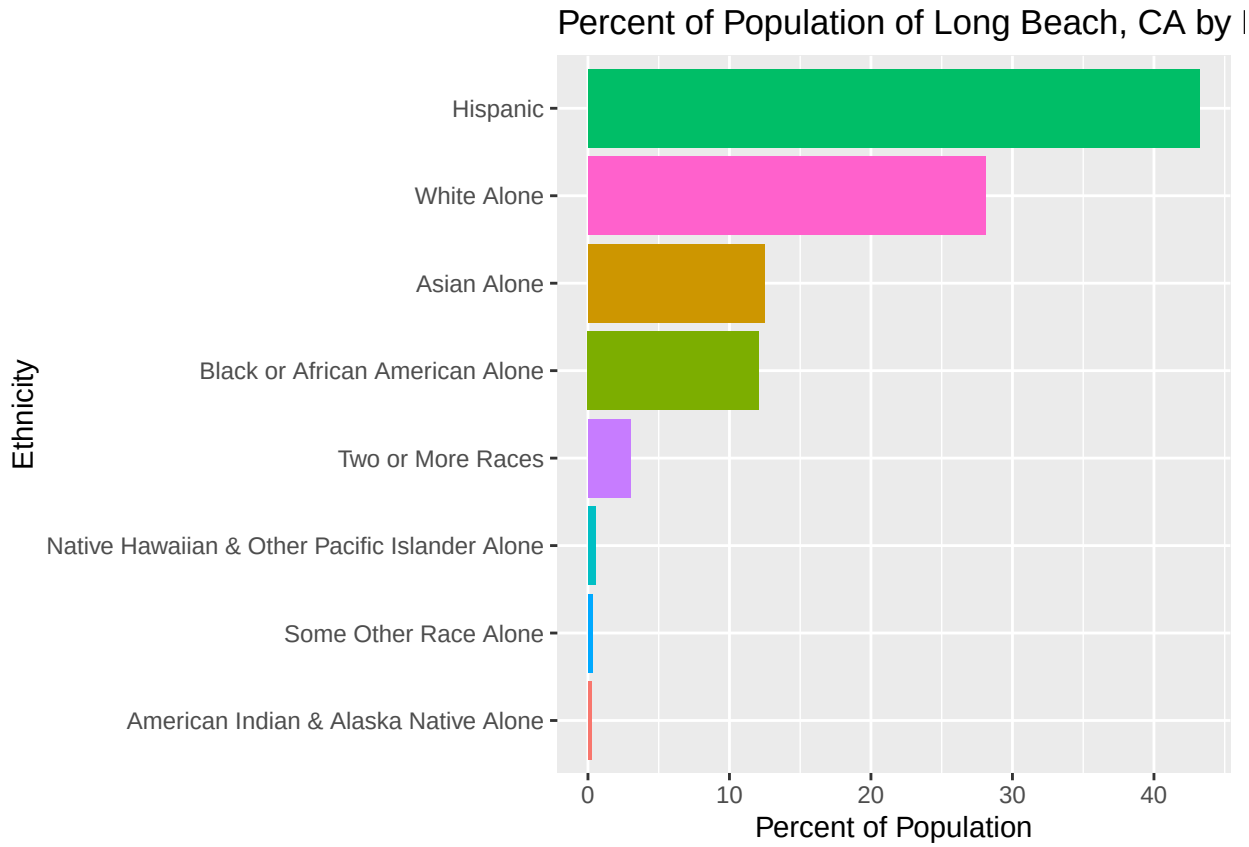
```
lb_population <- read.csv('Race and Ethnicity.csv')

lb_population <- lb_population %>%
  filter(Year == 2020)

# lb_population

lb_population_eth_percentage <- lb_population %>%
  mutate(eth_true = ifelse(Ethnicity == 'Hispanic or Latino', 'Hispanic', Race)) %>%
  group_by(eth_true) %>%
  summarize(percentage = round(100 * sum(share),2)) %>%
  arrange(desc(percentage))

knitr::kable(lb_population_eth_percentage)
```

eth_true	percentage
Hispanic	43.24
White Alone	28.07
Asian Alone	12.49
Black or African American Alone	12.09

eth_true	percentage
Two or More Races	3.01
Native Hawaiian & Other Pacific Islander Alone	0.55
Some Other Race Alone	0.30
American Indian & Alaska Native Alone	0.25

Visualization

```
ggplot(data = lb_population_eth_percentage, mapping = aes(y = reorder(eth_true, percentage), x = percen
  geom_bar(stat = 'identity') +
  labs(y = 'Ethnicity', x = 'Percent of Population', fill = 'Ethnicity', title = "Percent of Population
  theme(legend.position = "none")
```



Percentages by Ethnicity Breakdown of all Stops

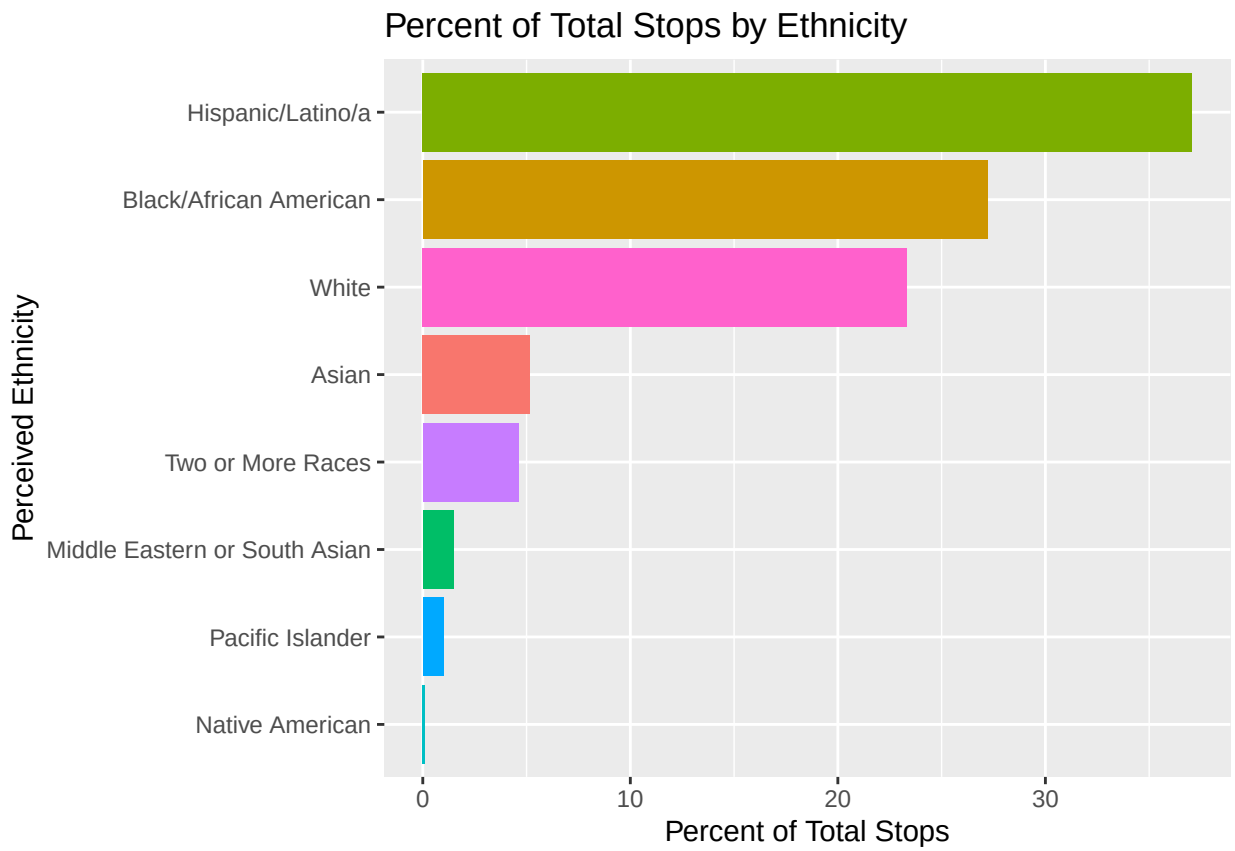
Note: maybe pull the demographics of long beach and compare for proportionality I know it's two data sources, but I'll live

```
Percent_by_Eth <- lb_police_stops %>%
  select("Perceived.Race.Simplified") %>%
  group_by(Perceived.Race.Simplified) %>%
  summarise(Percent_of_Stops = 100 * n()/nrow(.), total_stops = n()) %>%
  arrange(desc(Percent_of_Stops))

knitr::kable(Percent_by_Eth)
```

Perceived.Race.Simplified	Percent_of_Stops	total_stops
Hispanic/Latino/a	37.0601483	25835
Black/African American	27.2166516	18973
White	23.3349113	16267
Asian	5.1727848	3606
Two or More Races	4.6248081	3224
Middle Eastern or South Asian	1.4918736	1040
Pacific Islander	0.9854973	687
Native American	0.1133250	79

Visualization

```
ggplot(data = Percent_by_Eth, mapping = aes(y = reorder(Perceived.Race.Simplified, Percent_of_Stops), x
  geom_bar(stat = 'identity') +
  labs(y = 'Perceived Ethnicity', x = 'Percent of Total Stops', fill = 'Perceived Ethnicity', title = '
  theme(legend.position = "none")
```



Ethnicity vs Duration of Stop

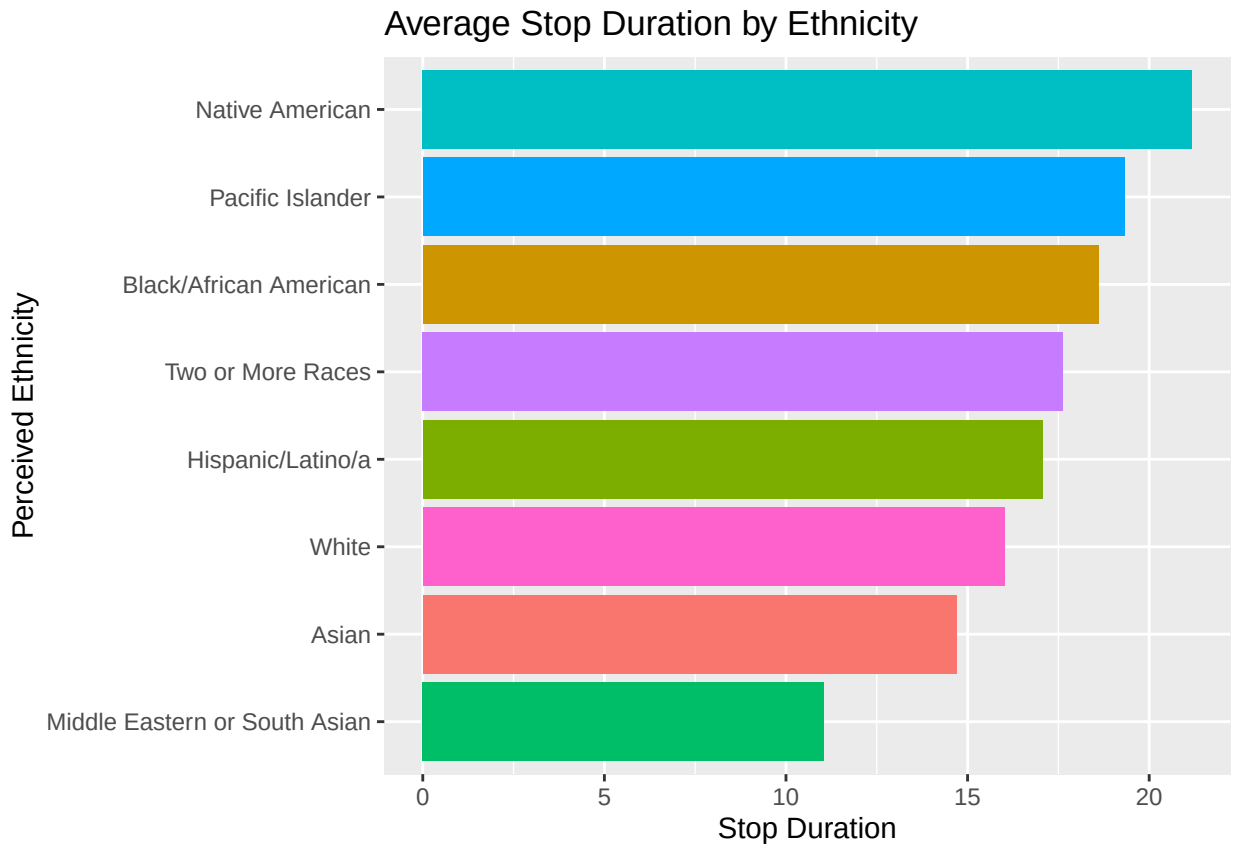
```
# colnames(lb_police_stops)

Stop_duration_by_Eth <- lb_police_stops %>%
  select("Perceived.Race.Simplified", "Stop.Duration") %>%
  group_by(Perceived.Race.Simplified) %>%
  summarise(avg_duration = mean(Stop.Duration)) %>%
  arrange(desc(avg_duration))

knitr::kable(Stop_duration_by_Eth)
```

Perceived.Race.Simplified	avg_duration
Native American	21.17722
Pacific Islander	19.31441
Black/African American	18.60280
Two or More Races	17.62624
Hispanic/Latino/a	17.06108
White	16.02090
Asian	14.69717

Perceived.Race.Simplified	avg_duration
Middle Eastern or South Asian	11.04808

Visualization

```
ggplot(data = Stop_duration_by_Eth, mapping = aes(y = reorder(Perceived.Race.Simplified, avg_duration),
  geom_bar(stat = 'identity') +
  labs(y = 'Perceived Ethnicity', x = 'Stop Duration', fill = 'Perceived Ethnicity', title = "Average S
  theme(legend.position = "none")
```



95% Confidence Intervals of Stop Duration by Ethnicity

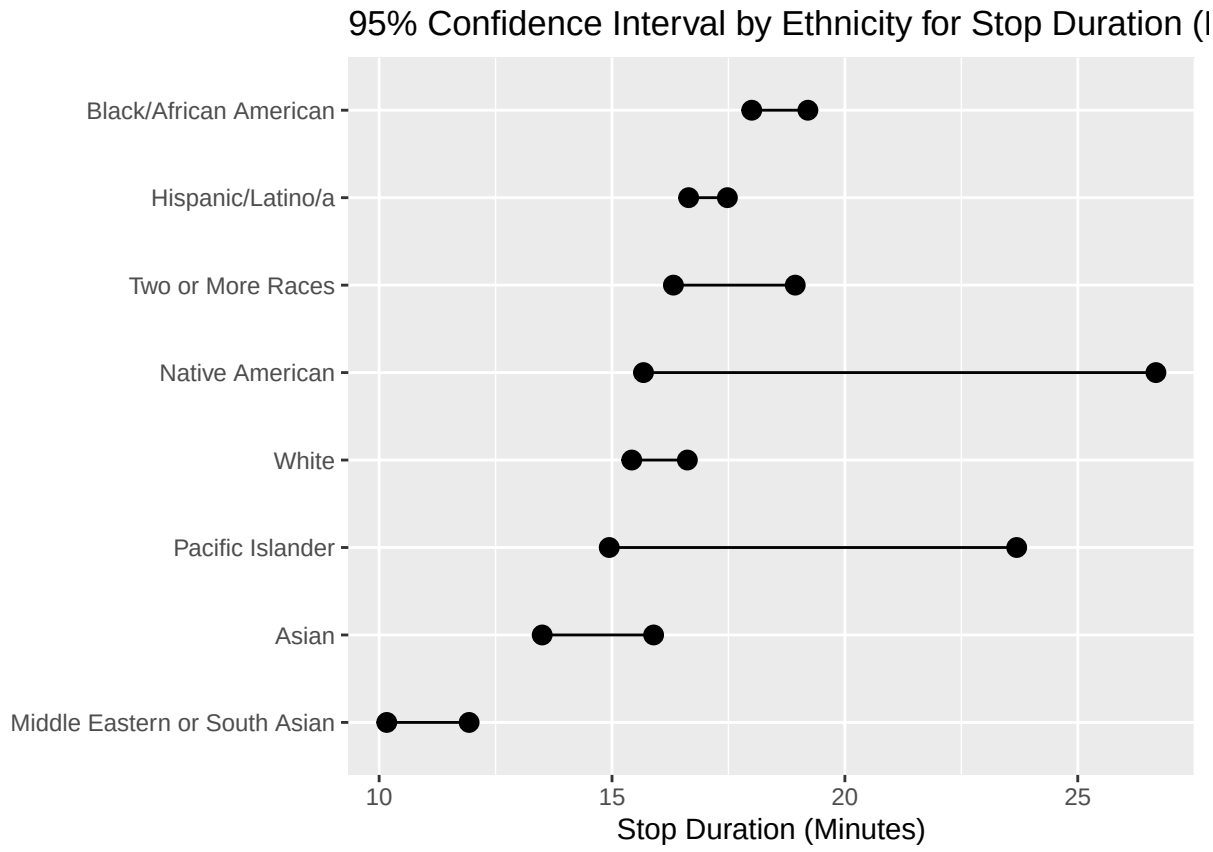
```
summary_stops_by_eth <- lb_police_stops %>%
  group_by(Perceived.Race.Simplified) %>%
  summarise(sd_time = sd(Stop.Duration, na.rm = TRUE), mean_time = mean(Stop.Duration, na.rm = TRUE), c

summary_stops_by_eth <- summary_stops_by_eth %>%
  mutate(lower_ci = mean_time - 1.96 * (sd_time / sqrt(count_stops)), upper_ci = mean_time + 1.96 * (sd

arranged_summary_stops <- summary_stops_by_eth %>%
  select(Perceived.Race.Simplified, lower_ci, upper_ci, count_stops) %>%
```

```
  arrange(desc(lower_ci))

knitr::kable(arranged_summary_stops)
```

Perceived.Race.Simplified	lower_ci	upper_ci	count_stops
Black/African American	18.00018	19.20542	18973
Hispanic/Latino/a	16.64524	17.47692	25835
Two or More Races	16.31854	18.93394	3224
Native American	15.67620	26.67823	79
White	15.42385	16.61795	16267
Pacific Islander	14.94027	23.68855	687
Asian	13.50078	15.89356	3606
Middle Eastern or South Asian	10.16427	11.93188	1040

Visualization

```
ggplot(arranged_summary_stops) +
  geom_segment( aes(x=reorder(Perceived.Race.Simplified, lower_ci), xend=Perceived.Race.Simplified, y=lo
  geom_point( aes(x=Perceived.Race.Simplified, y=lower_ci), size=3 ) +
  geom_point( aes(x=Perceived.Race.Simplified, y=upper_ci), size=3 ) +
  coord_flip()+
  theme(
    legend.position = "none",
  ) +
  xlab("") +
  ylab("Stop Duration (Minutes)") +
  labs(title= "95% Confidence Interval by Ethnicity for Stop Duration (Minutes)")
```



Ethnicty vs Action Taken During Stop

```
Action_Taken_by_Eth <- lb_police_stops %>%
  select("Perceived.Race.Simplified", "Num.Actions.Taken") %>%
  group_by(Perceived.Race.Simplified) %>%
  summarise(avg_num_of_actions = mean(Num.Actions.Taken)) %>%
  arrange(desc(avg_num_of_actions))

knitr::kable(Action_Taken_by_Eth)
```

Perceived.Race.Simplified	avg_num_of_actions
Native American	1.0126582
Black/African American	0.9612080
Pacific Islander	0.9490539
Hispanic/Latino/a	0.8539191
White	0.7162968
Two or More Races	0.6150744
Asian	0.5008319
Middle Eastern or South Asian	0.3057692

Given a person stopped is black, what is the likelihood action is taken? (Sampling)

```
black_and_action_taken <- lb_police_stops %>%
  filter(Perceived.Race.Simplified == 'Black/African American') %>%
  mutate(action_taken = ifelse(Actions.Taken != 'None', "yes", "no"))

black_and_action_taken %>%
  filter(!is.na(action_taken)) %>%
  specify(response = action_taken, success = "yes") %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = 0.95)
```

```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1    0.377    0.391
```

Given and person is Hispanic or Latino, what is the likelihood a police will take action

```
hispanic_and_action_taken <- lb_police_stops %>%
  filter(Perceived.Race.Simplified == "Hispanic/Latino/a") %>%
  mutate(action_taken = ifelse(Actions.Taken != 'None', "yes", "no"))

hispanic_and_action_taken %>%
  filter(!is.na(action_taken)) %>%
  specify(response = action_taken, success = "yes") %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = 0.95)
```

```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1    0.333    0.344
```

Given a person is white, what is the likelihood of police action?

```
white_and_action_taken <- lb_police_stops %>%
  filter(Perceived.Race.Simplified == 'White') %>%
  mutate(action_taken = ifelse(Actions.Taken != 'None', "yes", "no"))

white_and_action_taken %>%
  filter(!is.na(action_taken)) %>%
  specify(response = action_taken, success = "yes") %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = 0.95)
```

```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1    0.310    0.325
```

Given a person is Asian, what is the likelihood of police action?

```
Asian_and_action_taken <- lb_police_stops %>%
  filter(Perceived.Race.Simplified == 'Asian') %>%
  mutate(action_taken = ifelse(Actions.Taken != 'None', "yes", "no"))

Asian_and_action_taken %>%
  filter(!is.na(action_taken)) %>%
  specify(response = action_taken, success = "yes") %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = 0.95)
```

```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1    0.195    0.222
```

Does more time with the police Mean more actions taken?

```
lb_police_stops %>%
  summarise(cor(Stop.Duration, Num.Actions.Taken, use = "complete.obs"))
```

```
##   cor(Stop.Duration, Num.Actions.Taken, use = "complete.obs")
## 1                                                   0.2704781
```

```
lb_police_stops %>%
  filter(Perceived.Race.Simplified == "Black/African American") %>%
    summarise(cor(Stop.Duration, Num.Actions.Taken, use = "complete.obs"))
```

```
##   cor(Stop.Duration, Num.Actions.Taken, use = "complete.obs")
## 1                                                   0.2492346
```

Conclusion:

In conclusion, we can see from the data that the Black Community in Long Beach experiences disproportionately long Police Stops when compared to people of other races (with exception of the ethnicities that made up smaller portions of the overall population i.e. Native American and Pacific Islander populations).

The Black community also experienced a disproportionate amount of police action, when compared to the ethnicities that made up larger portions of the population.

When I quickly checked to see if there was correlation between the stop duration and the number of actions taken, there was only weak correlation at best.

Though what is clear is that the Black community of long beach faces statistically significant unfair treatment at the hands of the Long Beach Police Department during Police Stops.

Additional Analysis Possibilities:

This is something I did not have time for, but a cool addition to this sort of analysis would be to classify actions by Violent vs Non Violent (Use of Force vs Non-Use-of-Force). I would have to do some reading to see which actions are classified by department definition as use of force and then repeat this sort of analysis with the a binary representing whether or not an encounter involved the use of force and the counted number of uses of force.