DATA 606 Data Project Proposal

Coco Donovan

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
```

Data Preparation

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


head(lb_police_stops)
```

Research question

You should phrase your research question in a way that matches up with the scope of inference your dataset allows for.

Research Question: Is there a meaningful difference in the stop frequency and actions taken by Long Beach Police Officers for civilians of different "Perceived [Ethnicity]?"

Notes: The wording is "Perceived Race," but the correct wording should be "Perceived Ethnicity," hence the brackets around "Ethnicity." Also, it's important to note the word "Perceived," this comes from the Police Officer's Perception of of a civilian and is not an exact science, and it is the data to which we have access.

Cases

What are the cases, and how many are there?

The cases are the individual traffic stops that took place in the City of Long Beach's Police Departments jurisdiction from 1/1/2019 to 12/31/2021. There are 69,711 recorded stops and therefore there are 69,711 cases in this data.

```
min(lb_police_stops$Stop.Date)
```

```
## [1] "2019-01-01T00:00:00-08:00"
```

```
max(lb_police_stops$Stop.Date)
```

```
## [1] "2021-12-31T00:00:00-08:00"
```

```
nrow(lb_police_stops)
```

```
## [1] 69711
```

Data collection

Describe the method of data collection.

This comes directly from the City of Long Beach Data Website, "This dataset captures information about the stops that were conducted by Long Beach Police Department. LBPD officers conducting stops capture information required by The Racial and Identity Profiling Act of 2015 (AB 953). The data captured are submitted to the State of California annually, and is made available to the public via this dataset."

Type of study

What type of study is this (observational/experiment)?

This is an observational study. In theory the police continue along with their normal policing tactics, but just record the information needed in accordance with the law.

Data Source

If you collected the data, state self-collected. If not, provide a citation/link.

Link = "https://data.longbeach.gov/explore/dataset/lbpd-ripa-data-annual/table/?disjunctive.pid&disjunctive.perceivedage&disjunctive.actions_taken&disjunctive.did_consent_for_person_search&disjunctive.did_consent_for_property_search&disjunctive.perceived_races&disjunctive.num_perceived_races&disjunctive.perceived_race_simplified&disjunctive.all_reason_for_stop&disjunctive.all_reason_for_stopdesc&disjunctive.all_reason_for_stopcode_text&disjunctive.all_reason_for_stop_detail&disjunctive.num_reason_for_stop&disjunctive.reason_for_stop_simplified&disjunctive.all_result_of_stop&disjunctive.all_result_of_stop_code&disjunctive.all_result_of_stopcode_text&disjunctive.result_of_stop_simplified&disjunctive.was_detained&disjunctive.num_result_of_stop&disjunctive.stophour&disjunctive.timeofday&disjunctive.was_handcuffed&disjunctive.was_searched&disjunctive.was_removed_from_vehicle&rows=1000&dataChart=eyJxdWVyaWVzIjpbeyJjaGFydHMiOlt7InR5cGUiOiJiYXIiLCJmdW5jIjoiQ09VTlQ

Dependent Variable

What is the response variable? Is it quantitative or qualitative?

I'd like to look at a few response variables, Duration of the Stop, and treatment during the stop. Also, this is not so much a variable, but I think it will be important to look at the stop frequency compared across the specific demographic that I am looking at (Ethnicity, Gender, Able-Bodied/Disabled, etc.)

Independent Variable(s)

The main independent variable I want to look at is "Perceived [Ethnicity]" and given that the data also provides "Perceived Gender," "Perceived LGBT," and "Perceived Disability," there are a few values that I could use as my indepedent variable to see how changing demographics may affect traffic stop frequency and treatment during the stop.

Relevant summary statistics

Provide summary statistics for each the variables. Also include appropriate visualizations related to your research question (e.g. scatter plot, boxplots, etc). This step requires the use of R, hence a code chunk is provided below. Insert more code chunks as needed.

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

knitr::kable(Percent_by_Eth)
```

Overview: Percentages by Ethnicity Breakdown of all Stops

Perceived.Race.Simplified	Percent_of_Stops
Hispanic/Latino/a	37.0601483
Black/African American	27.2166516
White	23.3349113
Asian	5.1727848
Two or More Races	4.6248081
Middle Eastern or South Asian	1.4918736
Pacific Islander	0.9854973
Native American	0.1133250

```
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)
```

Ethnicty vs Action Taken During Stop

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

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

knitr::kable(Stop_time_by_Eth)
```

Ethnicity vs Stop Duration Boxplot

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

```
# Tried a boxplot, but it was too crowded and there were too many extreme outliers

# boxplot(Stop.Duration~Perceived.Race.Simplified, data=lb_police_stops, main="Ethnicity and Traffic St
#   xlab="Ethnicity", ylab="Stop Duration")
```

Because of the outliers, the boxplot may not be the clearest visualization of the dynamic Ethnicity may play in influencing the Traffic Stop Experience.