

# Case Study 0

Austin Kao

1/31/2021

## Introduction

The New York City police department has a Stop-Question-Frisk (SQF) policy where individuals can be stopped by police to be questioned and, in some cases, frisked based on the reasonable suspicion that the individual will commit a crime or is committing a crime. However, opponents of the policy protest that black and Hispanic residents are stopped by police at a significantly higher rate than white and Asian residents, making the practice unconstitutional. Now that a class-action lawsuit has led the NYPD to require justification for any SQF event, we seek to examine any racial disparities that remain in the police rationale to stop residents for violent crimes. We define violent crime to be any crime that involves force or the threat of force, which includes crimes like murder, robbery, and assault.

The data used for this case study comes directly from the NYPD Stop, Question, and Frisk database from the years of 2017 to 2019. We also use NYC police precinct data adapted from data compiled by John Keefe, an editor for the New York Times. To determine what justification police officers are providing for executing an SQF event, we used the “SUSPECTED\_CRIME\_DESCRIPTION” field in the Stop, Question, and Frisk database, which lists the exact crime being investigated. Since there are a variety of different crimes that can be investigated, we classify the crimes into violent and nonviolent categories.

Because we are examining why police officers decide to conduct an SQF event, we want to consider any potential details officers will have access to or will have considered before commencing a stop, which are details like height, weight, the presence of a weapon, body build type, age, and sex. We will also consider race as a factor and compare its influence against these other factors to establish how great an impact race has in these decisions and also to examine what racial disparities exist.

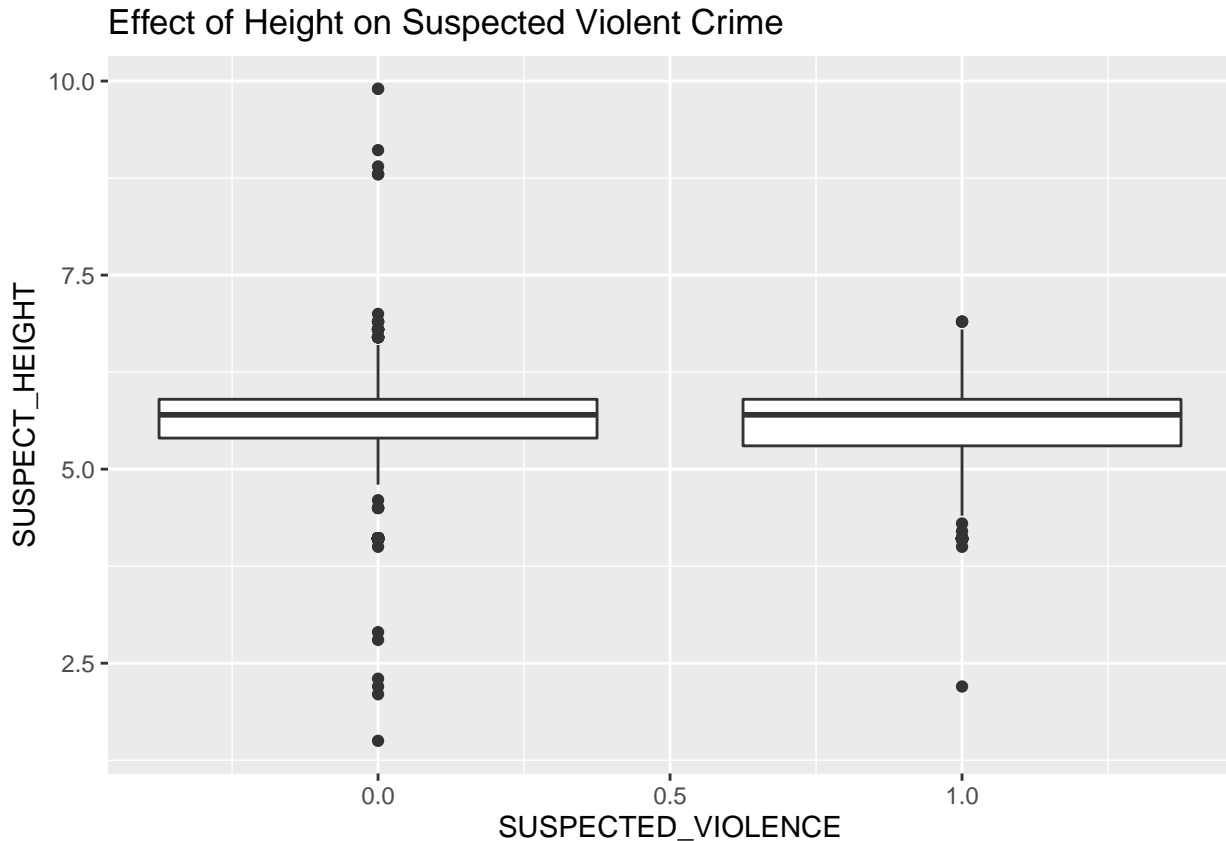
## Methodology

Because we are primarily concerned with the race of the suspects at SQF events, we filtered out 427 observations where the race was listed as null or male. Also, to examine SQF events across the years between 2017 and 2019, we combined the data sets from 2017 to 2019 into one data set. Any fields that are present in one data set but absent in another were excluded from the overall data set. The fields are named “SUPERVISING\_ACTION\_CORRESPONDING\_ACTIVITY\_LOG\_ENTRY\_REVIEWED” and “STOP\_LOCATION\_PREMISES\_NAME” are both present in the 2017 and 2018 data sets but absent in the 2019 data, while the fields named “ASK\_FOR\_CONSENT\_FLG” and “CONSENT\_GIVEN\_FLG” are both present in the 2019 data set but absent in both the 2017 and 2018 data. None of the fields being excluded affect police decisions to execute an SQF event except for perhaps the premises, but other stop location fields are more comprehensive and can act as sufficient substitutes.

To classify a crime as violent, we created the “SUSPECTED\_VIOLENCE” column, which has a value of 1 if the suspected crime is violent in nature, and 0 if the suspected crime is nonviolent. We based the “SUSPECTED\_VIOLENCE” column on the information given for “SUSPECTED\_CRIME\_DESCRIPTION”. The crimes we decided were violent are assault, menacing, forcible touching, murder, rape, terrorism, reckless endangerment, and robbery, while all other crimes were classified as nonviolent.

We also removed an additional 1749 observations due to missing or impossible height and weight data. There was a suspect recorded as being 60.11 feet tall. This data point was also removed since it is unclear if the officer was using inches or added an extra zero. Some suspects had a recorded weight of 0, while one suspect had a weight of 1, which are both impossible, so these data points were also removed.

```
ggplot(data = filtered_1719, aes(group = SUSPECTED_VIOLENCE, x = SUSPECTED_VIOLENCE, y = SUSPECT_HEIGHT)) +
  geom_boxplot() +
  labs(title = "Effect of Height on Suspected Violent Crime")
```



## Results

Based off the predictor variables we chose, we generated a model for the log odds that a suspect will be stopped by the police for a violent crime.

```
violence_model <- glm(SUSPECTED_VIOLENCE ~ RACE + SUSPECT_HEIGHT + SUSPECT_WEIGHT
  + SUSPECT_REPORTED_AGE + SUSPECT_SEX + WEAPON_FOUND, data = filtered_1719)
tidy(violence_model)
```

```
## # A tibble: 12 x 5
##   term                estimate std.error statistic  p.value
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)        0.403     0.0594     6.79 1.12e-11
## 2 RACEASIAN/PAC.ISL  0.0866     0.0196     4.41 1.01e- 5
## 3 RACEW. HISP.       0.0726     0.0100     7.24 4.45e-13
## 4 RACEB. HISP.       0.0841     0.0121     6.95 3.76e-12
## 5 RACEBLACK          0.0788     0.00907    8.68 4.06e-18
## 6 RACEAMER IND      -0.0795     0.0855    -0.931 3.52e- 1
```

##	7	SUSPECT_HEIGHT	-0.00789	0.00689	-1.14	2.52e- 1
##	8	SUSPECT_WEIGHT	-0.0000781	0.0000843	-0.926	3.54e- 1
##	9	SUSPECT_REPORTED_AGE	-0.00374	0.000222	-16.9	1.33e-63
##	10	SUSPECT_SEXFEMALE	0.0173	0.0452	0.383	7.02e- 1
##	11	SUSPECT_SEXMALE	0.00818	0.0444	0.184	8.54e- 1
##	12	WEAPON_FOUNDY	-0.101	0.00872	-11.6	3.63e-31

The model is

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_0 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_3 + \beta_5 x_4 + \beta_6 x_5 + \beta_7 x_6 + \beta_8 x_7 + \beta_9 x_8 + \beta_{10} x_9 + \beta_{11} x_{10}$$

## Discussion

From these results, we can see how race has a significant impact on police decisions. The log odds of police suspicion of a violent crime increase by 7.26 for a white hispanic than for a white person. This means that the police consistently use race to make them suspicious enough for an SQF event.

## Appendix

Bibliography