

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. Officers who conduct an SQF event are supposed to complete a form recording details about the event. These forms are then entered into the electronic database, which the NYPD releases annually. 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. We then classified the listed crimes into violent and nonviolent categories for a new variable called “SUSPECTED_VIOLENCE”, which will act as our response variable. “SUSPECTED_VIOLENCE” has a value of 1 if the suspected crime is violent in nature, and 0 if the suspected crime is nonviolent.

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, age, sex, body build type, age, sex, and how the stop was initiated. 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. After an EDA about how each of these factors affects police suspicion of violent crime, we determined height, age, sex, and how the stop was initiated to be the most significant factors, so we will use these factors in our model.

On preliminary examination of the data, it would seem that some racial disparity does exist. When comparing the ratio of stops for suspected violent and nonviolent crimes for each race, black and Hispanic residents are stopped for suspected violent crimes at higher rates than white residents.

Ratio Between Suspected Violent and Nonviolent Crimes by Race



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 create the “SUSPECTED_VIOLENCE” field, which has a value of 1 if the suspected crime is violent in nature, and 0 if nonviolent, we used the data in “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 3130 observations due to missing or impossible data. If a suspect was missing height or age data, that data point was removed. There were also data points removed that were missing the “STOP_WAS_INITIATED” field. There were also impossible data points, such as a suspect who was 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.

In regards to the sex of the suspect, there were some null values present in the data. We decided to label these values as “UNSPECIFIED” since the suspect could have any gender, including non-binary genders that the data cannot capture.

To allow for meaningful interpretation of the intercept, we centered the variables “SUSPECT_HEIGHT” and “SUSPECT_REPORTED_AGE” about their means.

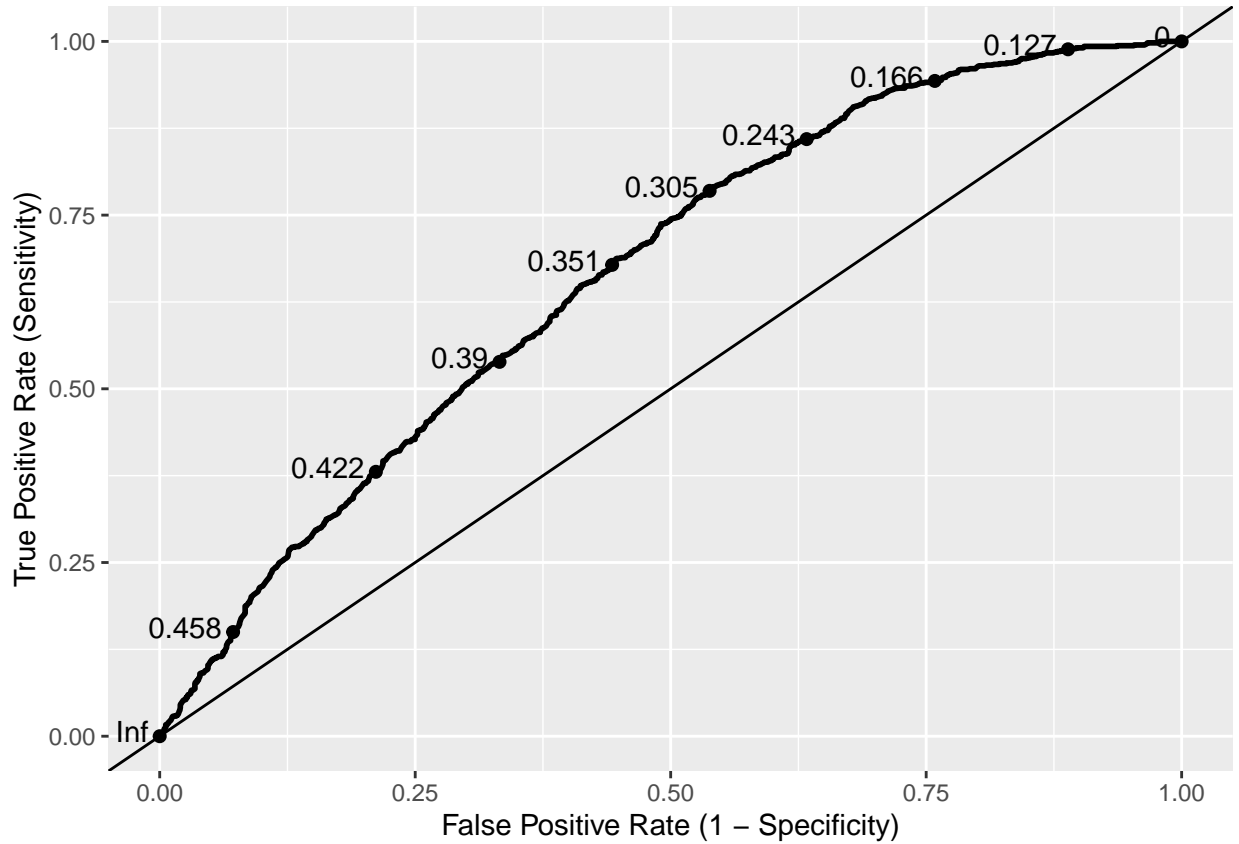
We proceeded to take a simple random sample of 3000 observations out of the remaining 32539 observations to serve as the basis for our model. This also satisfies our model assumption that the data is random.

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.

```
## # A tibble: 12 x 5
##   term                                estimate std.error statistic  p.value
##   <chr>                                <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)                        -2.28      0.177     -12.9  4.45e-38
## 2 RACEASIAN/PAC.ISL                   0.870     0.310      2.80  5.07e- 3
## 3 RACEW. HISP.                        0.488     0.170      2.87  4.09e- 3
## 4 RACEB. HISP.                        0.445     0.201      2.21  2.68e- 2
## 5 RACEBLACK                           0.517     0.155      3.33  8.56e- 4
## 6 RACEAMER IND                       -12.7     308.      -0.0412 9.67e- 1
## 7 SUSPECT_HEIGHT                      0.0364    0.108      0.337  7.36e- 1
## 8 SUSPECT_REPORTED_AGE                -0.0223   0.00356   -6.25  3.98e-10
## 9 SUSPECT_SEXFEMALE                  -0.0596   0.147     -0.405  6.86e- 1
## 10 SUSPECT_SEXUNSPECIFIED             -0.247    0.845     -0.292  7.70e- 1
## 11 STOP_WAS_INITIATEDBased on Radio Run  1.27     0.114     11.1   9.23e-29
## 12 STOP_WAS_INITIATEDBased on C/W on Scene 1.66     0.142     11.7   7.71e-32
```

The model has the form $\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_0 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_3 + \beta_5 x_4 + \beta_6 x_5 + \beta_7 x_6 + \beta_8 x_7 + \beta_9 x_8 + \beta_{10} x_9 + \beta_{11} x_{10}$, where β_0 is the intercept, β_1 through β_{11} are the coefficients for the model parameters, and x_1 through x_{10} are different model parameters.



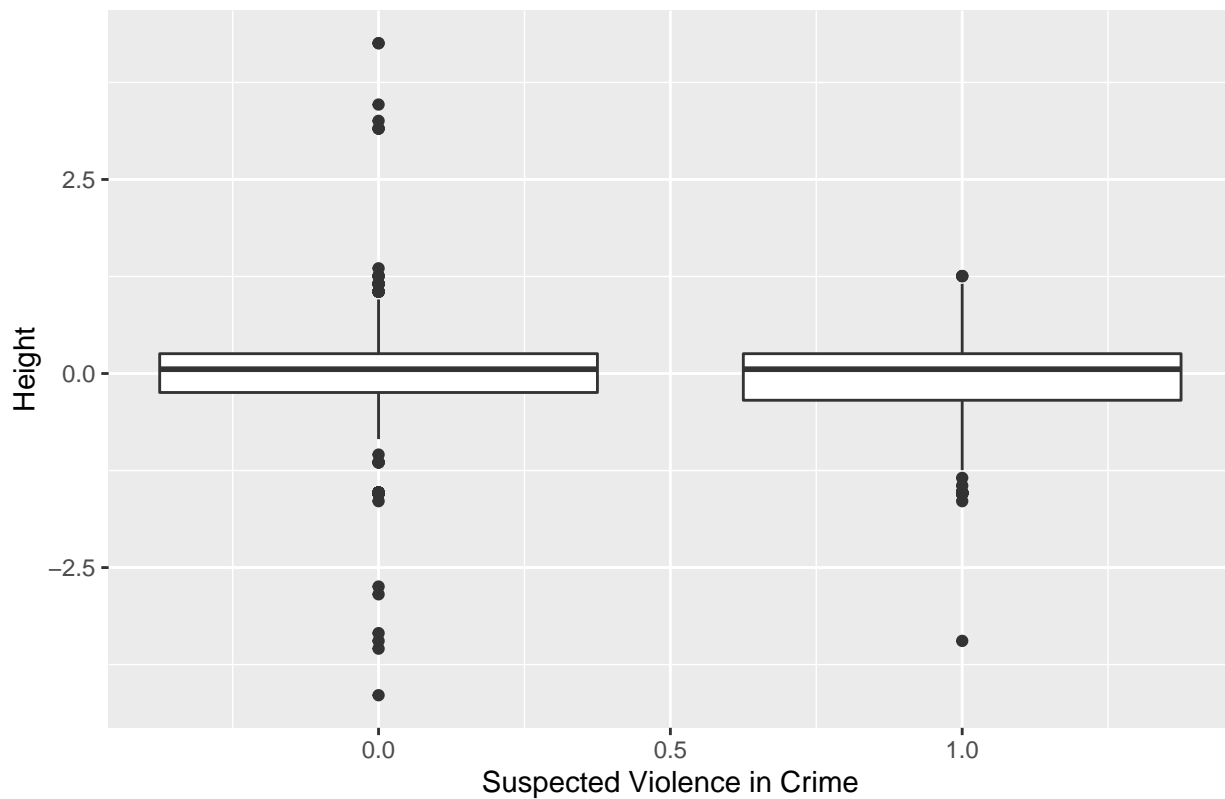
Discussion

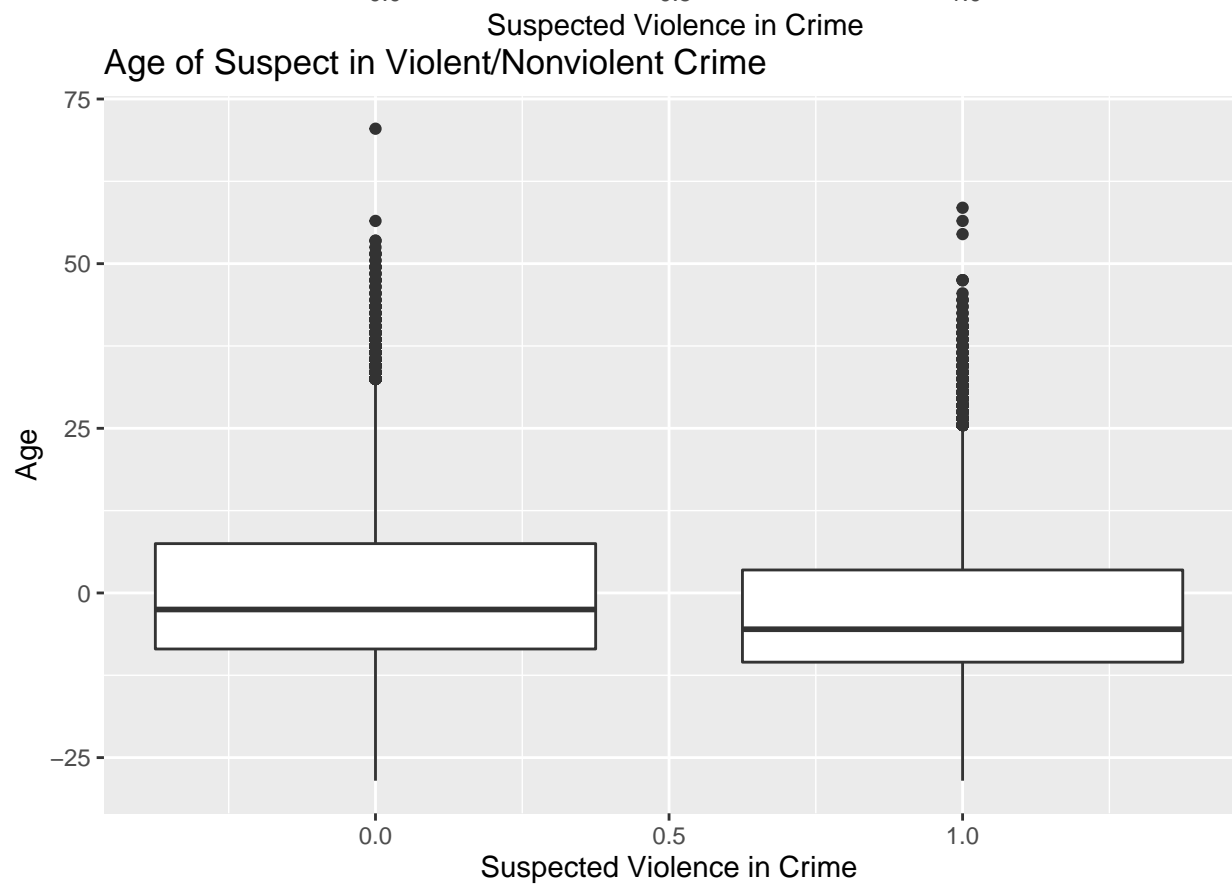
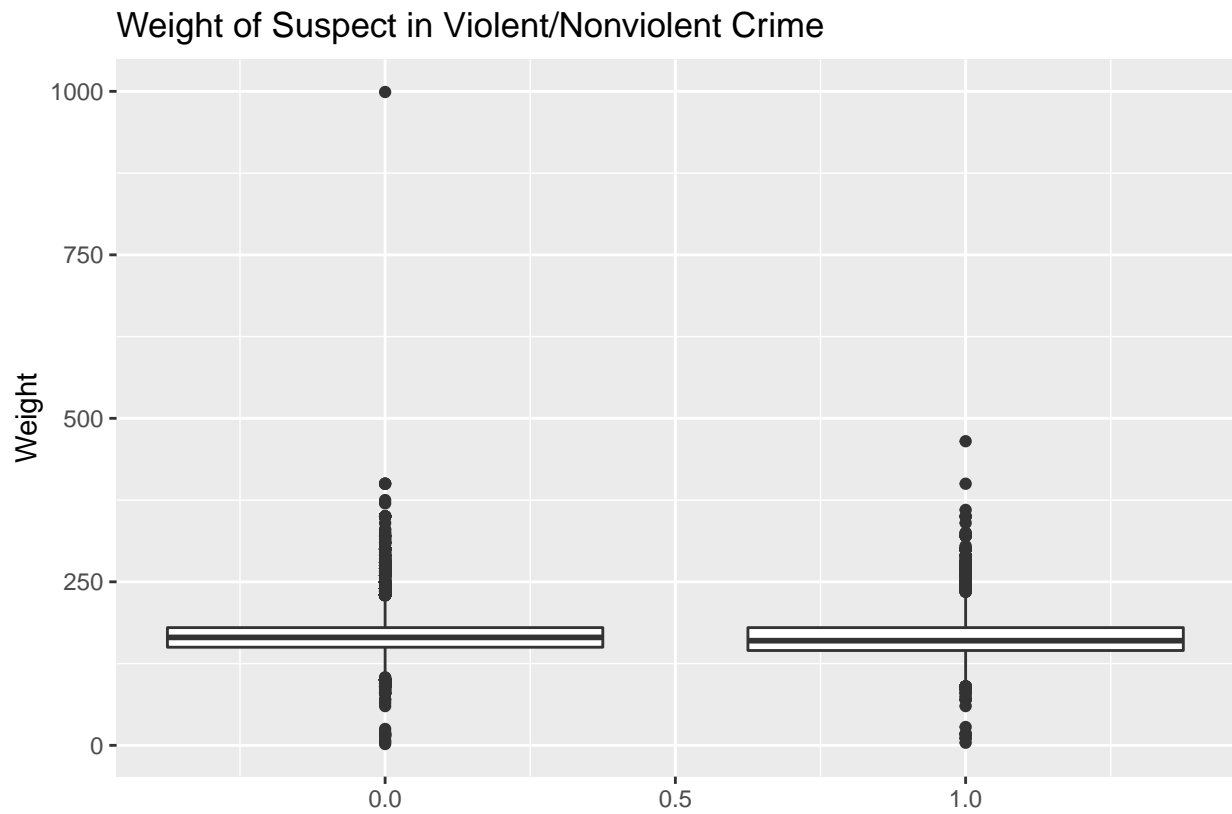
From these results, we can see how race has a significant impact on police decisions. For our baseline, we used a white male with mean height and age who was stopped based on the officer's own initiative. The log odds of police suspicion of a violent crime increase by 7.26 for a white Hispanic than for a white person. This means that the police consistently use race to make them suspicious enough for an SQF event.

Appendix

Note: To reproduce all the results in this case study, simply knit the Analysis.rmd document.

Height of Suspect in Violent/Nonviolent Crime



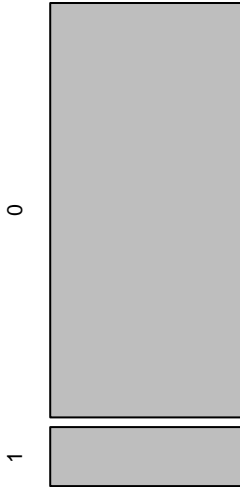


Ratio of Violent and Nonviolent Crime by Sex

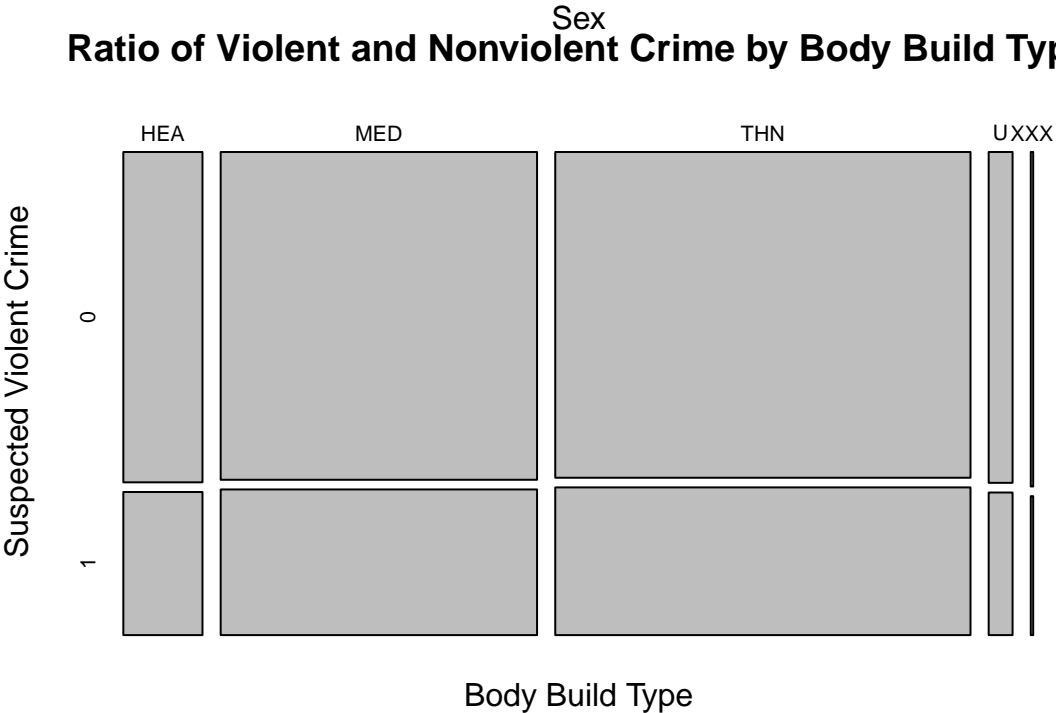


Non

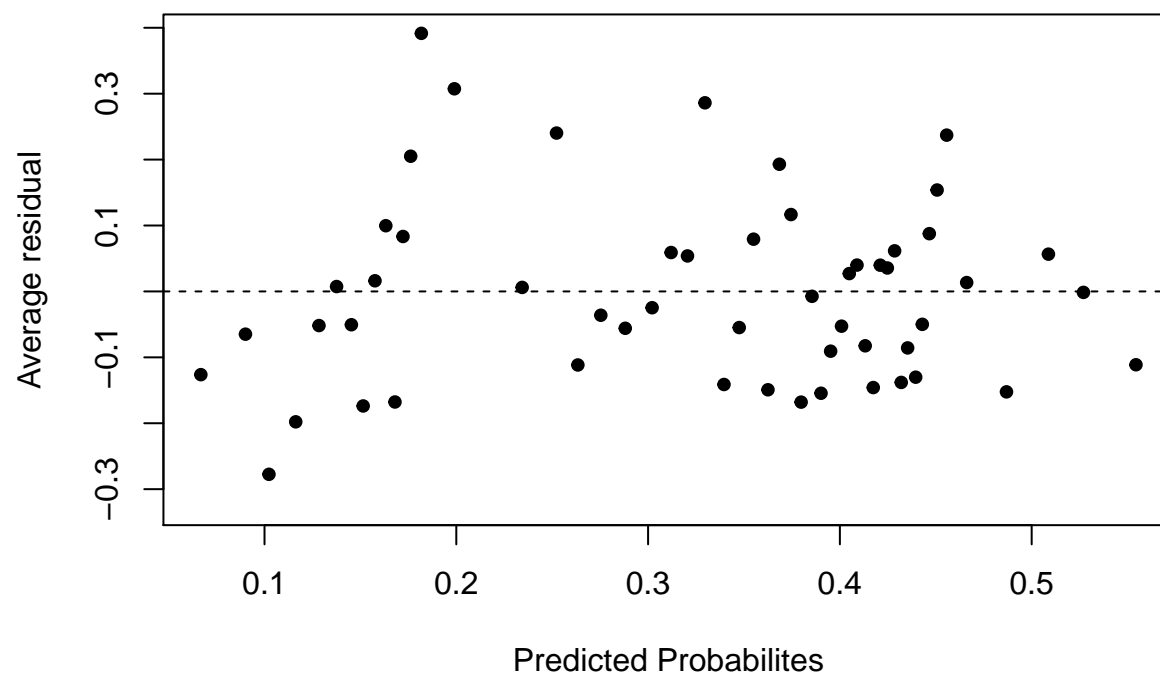
Based on Self Initia



Ratio of Violent and Nonviolent Crime by Body Build Type



Binned Residuals vs. Probabilities



b

Bibliography

<https://www.nyclu.org/en/stop-and-frisk-data>