Racial Disparities in Traffic Stops/Citations

Keohane sQUAD: Chris Liang, Andrew Qin, Bob Qian, and Katie Nash

2020-10-26

Introduction and Data

Our data is a census of individual police stops in Durham created by the Stanford Open Policing project. The Stanford Open Policing Project "[collects] and [standardizes] data on vehicle and pedestrian stops from law enforcement departments across the country"(https://openpolicing.stanford.edu/). We would like to see if that same kind of racial bias is evident in police stops in Durham. In doing so, we also wish to examine if other demographic characteristics (such as sex or age) influence traffic stops. Our general research question is the following: what is the relationship between a subject's demographic attributes (sex, race, or age) and the likelihood of being stopped by police in traffic in Durham?

We hypothesize that race and the likelihood of being stopped by police in traffic in Durham are related, with black people representing disproportionately more of the people being stopped relative to their proportion within the population. We hypothesize that younger people (roughly 18-30) have a disproportionately higher chance of being stopped in traffic (not necessarily due to bias but other lurking variables, such as inexperienced driving). We also hypothesize that sex has no significant relationship with being stopped in traffic. To find the true population proportions of people by race, sex, and age in Durham, we will utilize the 2010 Durham census data (https://www.census.gov/quickfacts/fact/table/durhamcountynorthcarolina/RHI625219#RHI625219).

Additionally, we will examine whether race, sex, or age are related to the outcome of the traffic stop (whether a citation will be issued). We hypothesize that race and the likelihood of receiving a citation are related, with black people more likely to receive a citation upon being stopped. We additionally hypothesize that younger people have a higher chance of receiving a citation upon being stopped and that sex has no significant relationship with being stopped in traffic.

It has 29 variables and 323147 observations, and each observation in the data set is an individual police stop recorded in Durham during 2001 to 2015. A categorical variable in the data set is `subject_race`, which describes the race of the subject involved in the traffic stop. A discrete numerical variable in the data set is `subject_age`, which describes the age of the subject at the time of the traffic stop. A continuous numerical variable in the data set is `time`, which describes the hour, minute, and second that the stop was recorded. Other variables in the data set include `outcome`, which is what resulted from the stop (a warning or a citation, for example); `reason_for_stop`, which describes what the violation leading to the stop was; and `search_conducted`, whether a search of the subject was conducted during the stop.

Methodology

Observed vs Expected Proportions of People Stopped by Race
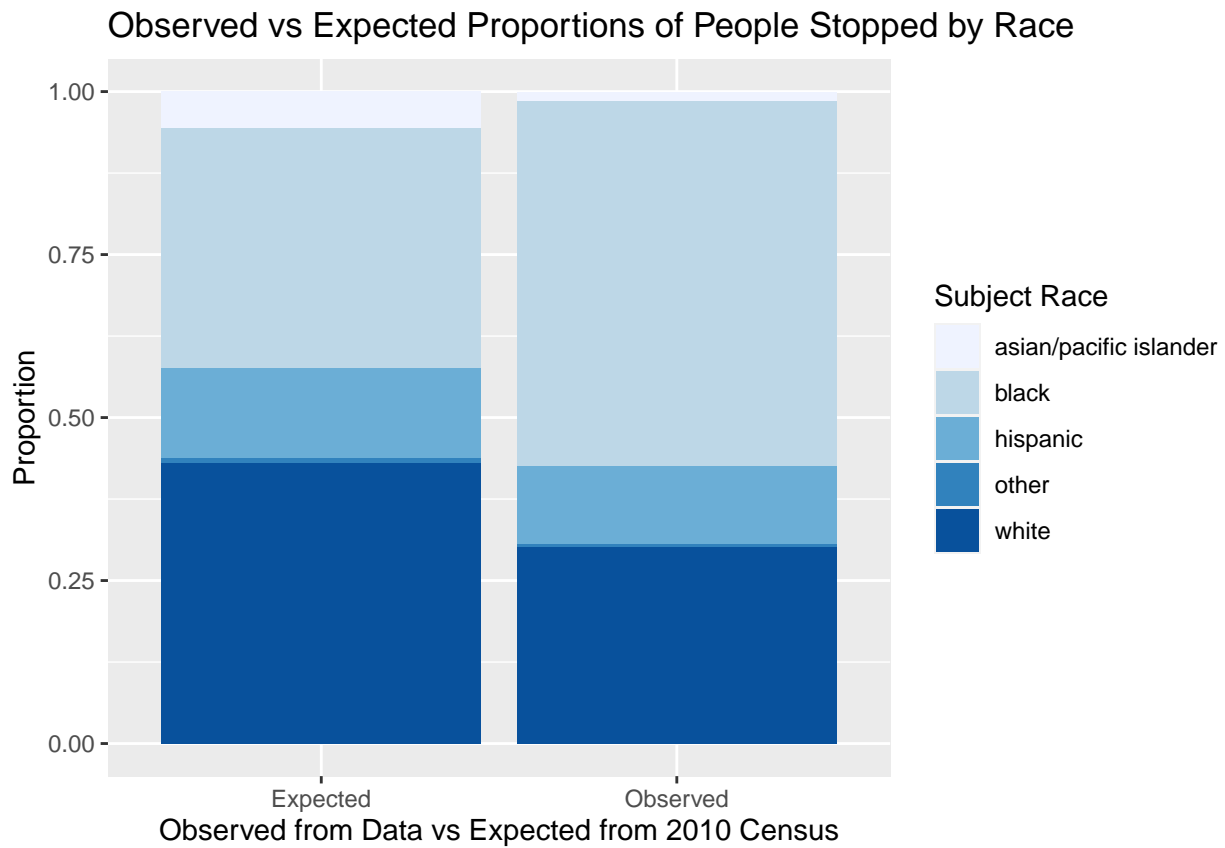

The variables we use to address the research question are `subject_race`, `subject_sex`, and `citation_issued`. For these variables we filter out any unknown and NA values. We also mutate a new categorical variable `age_category` based on values of `subject_age` with the age categories "10-25", "25-40", "40-64", and "65 or older."

To begin with, we visualize a segmented bar graph with the probability of citation based on race below.

## Likelihood of Receiving Citation Based on Race
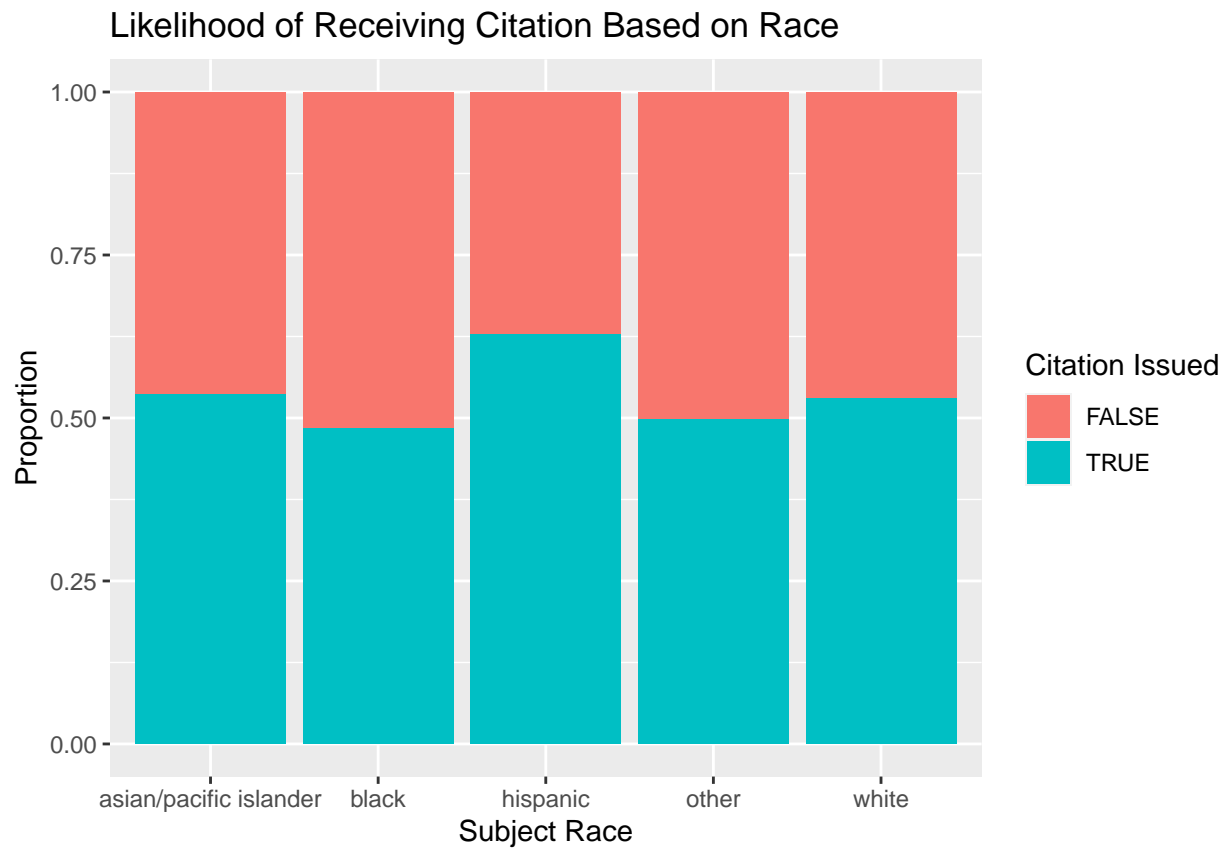
According to the chart, it appears that Hispanics are the race with the highest proportion of citations issued. We also visualize a segmented bar graph with the probability of citation based on age below.

Likelihood of Receiving Citation Based on Age
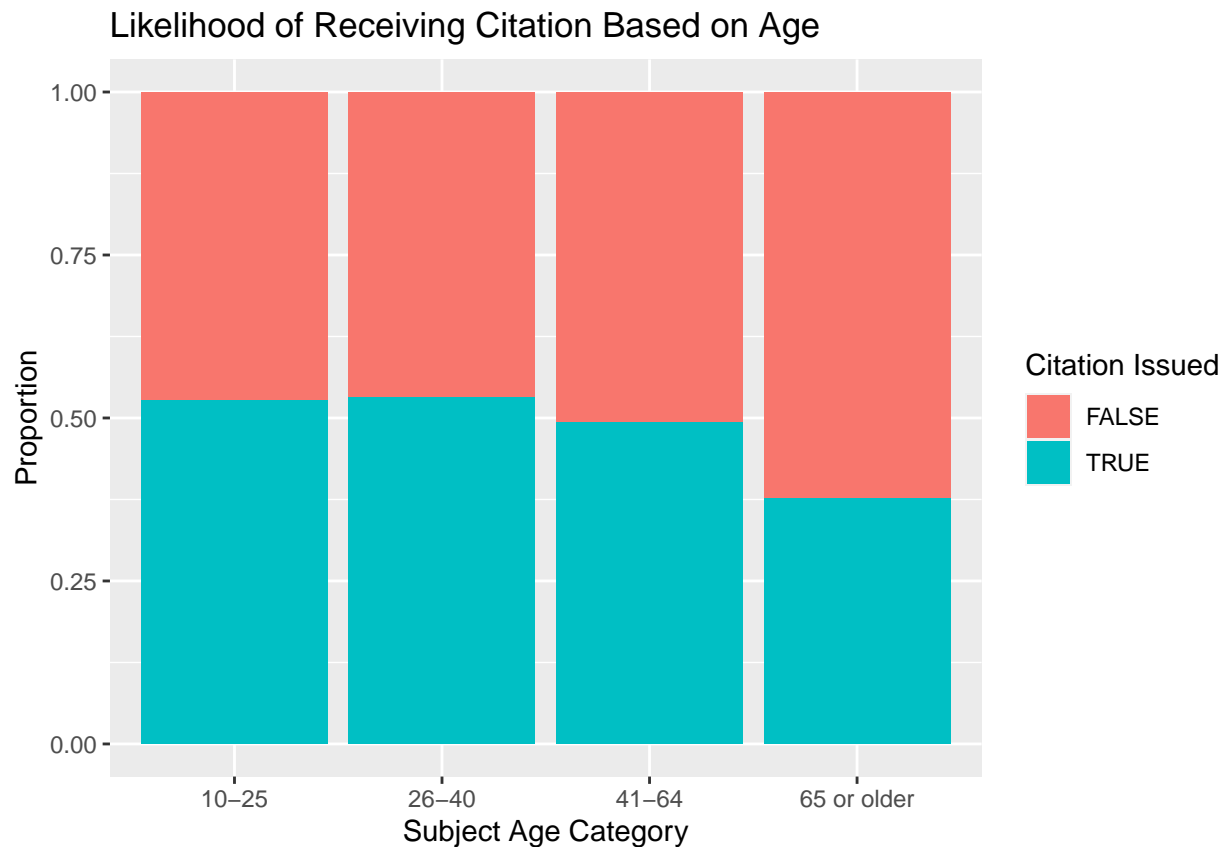
Based on this graph, it appears that there is a roughly equal proportion of citations issued to those in the 10-25 age group and 25-40 age group. The proportion of citations issued then begins to decrease in the next two age groups, with the citation proportion of the 40-64 age group being less than the previous groups' proportions and the 65 or older group's proportion being less than the 40-64 group's citation proportion.

Lastly, we visualize a segmented bar graph with the probability of citation based on sex below.

Likelihood of Receiving Citation Based on Sex
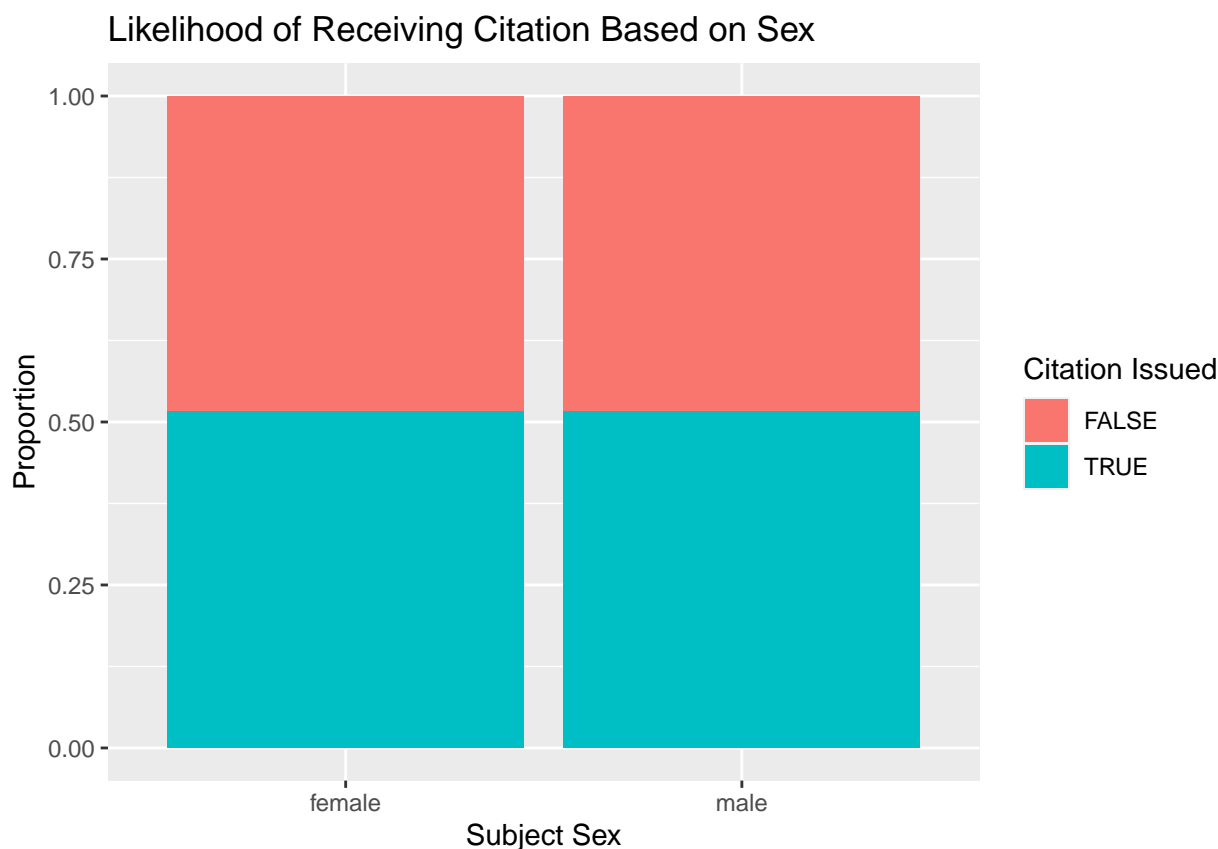
Based on this chart, the proportion of citations issued for females and males appears to be roughly equal.

To answer our research question, we utilize the chi-square test. We selected this test because we want to determine whether there is an association between two variables where we have more than two samples.

We perform three chi-square tests. For the first, we ask whether there is an association between someone's race status and whether a citation was issued. For the second, we ask whether there is an association between someone's age category and whether a citation was issued. For the third, we ask whether there is an association between someone's sex and whether a citation was issued.

With each chi-square test, we compare observed versus the expected counts that we would expect if each H_0 were true. If these total differences are "large enough," then we reject the null hypothesis. We will perform each chi-square test at the $\alpha = 0.05$ significance level.

Results

Let ρ equal the true

```
## # A tibble: 1 x 3
##    statistic chisq_df p_value
##        <dbl>    <int>   <dbl>
## 1     2785.        4       0
```

We reject the null... Two things are associated...

Investigate nature of association with logistic model

Conditions (prob in methodology)

```
## # A tibble: 5 x 5
##   term                estimate std.error statistic  p.value
```

```
##   <chr>                      <dbl>      <dbl>      <dbl>     <dbl>
## 1 (Intercept)               0.142     0.0286       4.97  6.66e- 7
## 2 subject_raceblack        -0.206     0.0290      -7.12  1.06e-12
## 3 subject_racehispanic      0.382     0.0304       12.6  3.61e-36
## 4 subject_raceother        -0.153     0.0632      -2.43  1.53e- 2
## 5 subject_racewhite       -0.0240     0.0293     -0.820  4.12e- 1
```

BEing HIspanic associated with increased likelihood... Being black associated with lower likelihood of citation...

Add age to logit model to see if it changes predictions

```
## # A tibble: 1 x 3
##   statistic chisq_df   p_value
##       <dbl>    <int>     <dbl>
## 1      991.        3 2.06e-214

## # A tibble: 1 x 3
##   statistic chisq_df p_value
##       <dbl>    <int>   <dbl>
## 1  0.000593        1   0.981
```

Fail to reject null...

Discussion

Though black people are disproportionately more likely to be stopped for a traffic violation, they were not the most likely to receive a citation upon being stopped–Hispanics were the most likely to receive a citation after being stopped.