

# Final Project: New York City Stop-Question-Frisk Group Write Up

*Xingyao Chen, Sonia Sehra, Dave Makhervaks, Jenna Kahn Section 5*

*Due December 13, 2016*

## **Background:**

In stop-question-frisk, a police officer is authorized to stop a pedestrian, question them, and then frisk their body searching for contraband items such as weapons or drugs. The motivation for the policy was to prevent crimes from happening in the first place, though recent studies by the New York Civil Liberties Union suggest the policy did not achieve noticeable reductions in crime. For example, NYCLU estimates that guns were found in fewer than 0.2% of stops. Moreover, the policy was found to be discriminatory because Blacks and Latinos were stopped disproportionately more than their participation in crime would suggest. (Since 2014, new restrictions have limited the use of SQF, and the numbers of SQF incidents have decreased precipitously.)

Data from New York City's stop-question-frisk (SQF) program is publicly available online. Whenever a person is stopped under SQF, the officer is required to fill out a form with information about the stop. Each year, the data is compiled and released by the city.

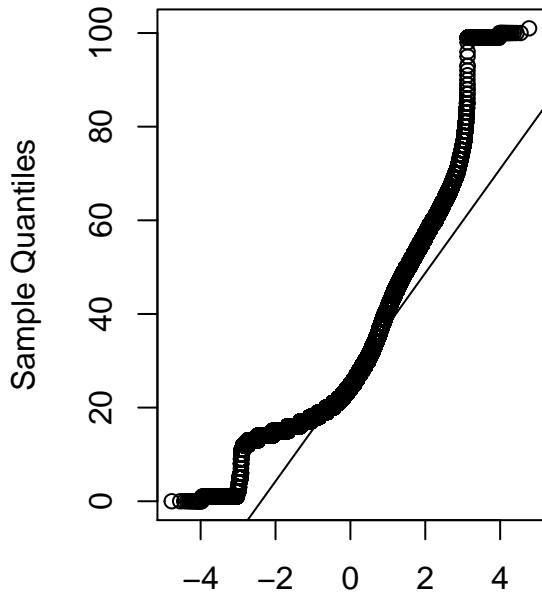
## **Required Questions:**

The probability that someone who is identified by the officer as Hispanic is actually Hispanic is 99.7%. The probability that someone who is identified as Asian/Pacific Islander is actually Asian/Pacific Islander is 0.4964. Refer to R code at *Question 1 Code* in "Group Code and Graphs.rmd"

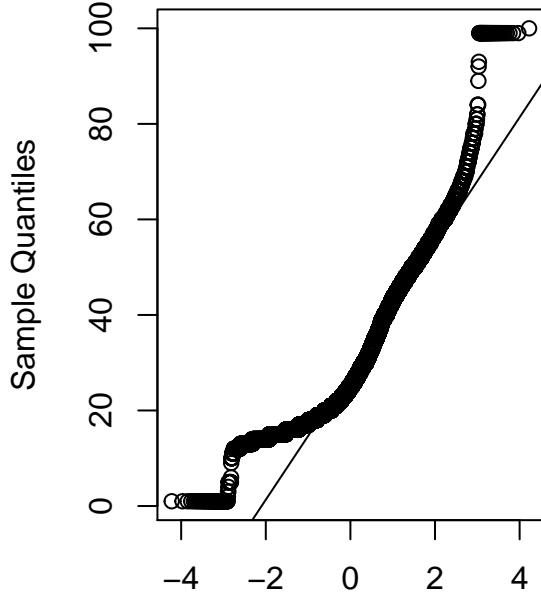
The largest difference in stopped rates, broken down by race is between Bronx and Brooklyn, Bronx and Manhattan, Bronx and Staten Island, Brooklyn and Queens, Manhattan and Queens, and Queens and Staten Island. Refer to R code and table at *Question 2 Chunk 1 Code* in "Group Code and Graphs.rmd"

The smallest difference in stopped rates, broken down by race is between Brooklyn and Manhattan, Unknown and Brooklyn, and Unknown and Staten Island, Unknown and Manhattan Refer to R code and table at *Question 2 Chunk 2 Code* in "Group Code and Graphs.rmd"

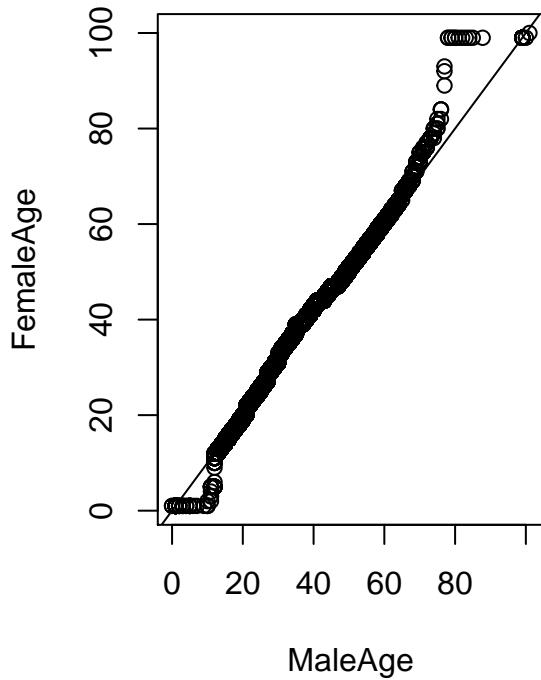
## MaleAge QQ Plot without outlier   FemaleAge QQ Plot without outlier



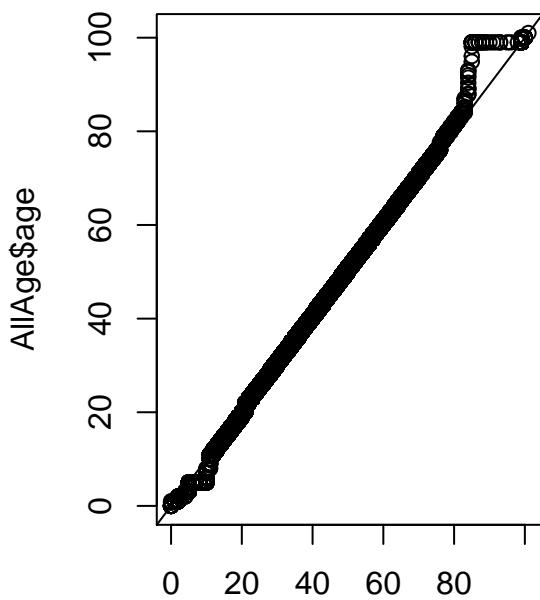
Theoretical Quantiles  
**QQ Plot of Men vs Women Ages**



Theoretical Quantiles  
**QQ Plot of Men vs All Ages**



**MaleAge**



**MaleAge**

The distribution of ages for male suspects with the outliers included is somewhat normal in the middle, but it is pretty messy and is not a very good QQ plot (not a straight line). The same follows for the distribution of ages for female suspects with the outliers included in the data. However, when the outliers are removed, then the QQ-Plot for both the distribution of both male and female ages are reasonably normal. We know this, because the QQ plots without the outliers are reasonably straight linear excluding the sides of the graph. Creating a QQ Plot of Male ages vs Female ages, we get a very linear line, showing that the male

and female distributions are in fact, very similar! Comparing the male distribution of ages against the age distribution of the entire population also yields a linear plot, showing that the distributions are very similar! The distributions between Females and Males are approximately the same, because the plot “QQ Plot of Men vs Women Ages” yielded points that fall closely on the  $y = x$  line. Refer to R code and figures at *Question 3 Code* in “Group Code and Graphs.rmd”

The probability, with a 95% confidence interval, that a suspect was frisked for the entire population in 2015 is  $0.6761955 \pm 0.006105832$ . Refer to R code at *Question 4 Chunk 1 Code* in “Group Code and Graphs.rmd”

The probability, with a 95% confidence interval, that a suspect who refused to provide ID was frisked is  $0.6103286 \pm 0.03784224$ . Refer to R code at *Question 4 Chunk 2 Code* in “Group Code and Graphs.rmd”

A two sample t-test can be used to determine whether these rates significantly differ. The t-test suggests there is a significant difference in the probability a suspect in general will be frisked vs. the probability a suspect who refuses to provide ID will be frisked ( $p < .05$ ). Refer to R code at *Question 4 Chunk 3 Code* in “Group Code and Graphs.rmd”

The  $R^2$  values indicate that the linear models are a poor fit for the data. Because of the low  $R^2$  values, not much variability is explained by the selected variables. The period stopped is not affected by anything, because it completely depends on a case by case scenario. If a suspect is stopped and they comply and everything then the per stop will be short. However, if they dont comply then the perstop could be longer. There are also other factors that come into play. For example, if the officer stops the suspect and then spends a long time trying to find the license plate number in his system, or the suspect has a hard time finding their insurance in their glove box or something. Minor irrelevant things can increase or decrease the perstop, subsequently lowering the goodness-of-fit in the proposed linear models.

Refer to R code and tables at *Question 5 Code* in “Group Code and Graphs.rmd”

### **Team-Chosen Questions:**

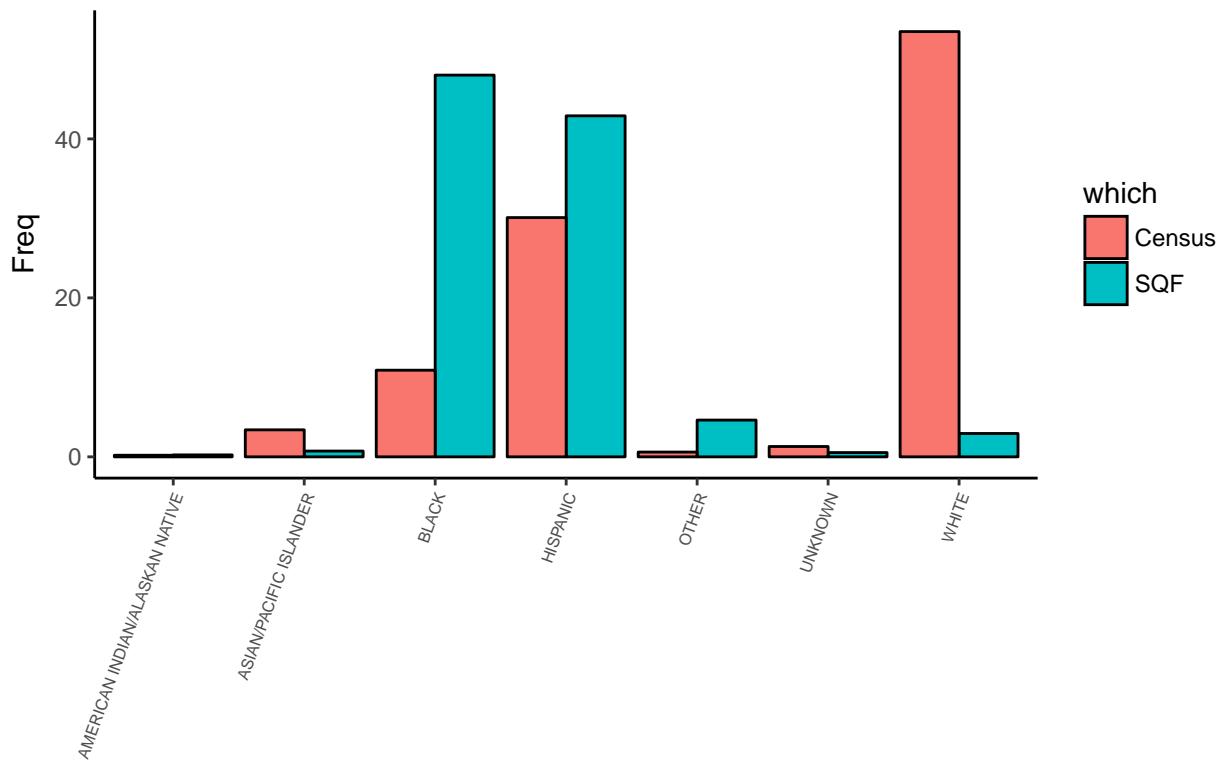
Introduction:

We examined whether people of certain races are more likely to be stopped and/or frisked than would be expected based on the demographic makeup of each borough. This question is important because stop and frisk has faced numerous accusations of racial profiling. If we find that certain racial groups are disproportionately likely to be stopped that, then our findings give support to the claim that stop and frisk unfairly targets those groups. To answer our question, we examined two sets of data. One is the 2010 US government census data for each of the five New York boroughs. We used this to find the racial breakdown of population of each borough. The other data set is a collection of police filings from all of the stop-and-frisk incidents in New York in 2010. Each filing contains information on the race of the stopped individual, where the stop occurred, and whether the individual was frisked during the encounter. We used this data set to the relative frequency with which people of each race were likely to be stopped and frisked. We conducted graphical and statistical analyses to determine whether black, Hispanic, and white people were disproportionately likely to be stopped and frisked.

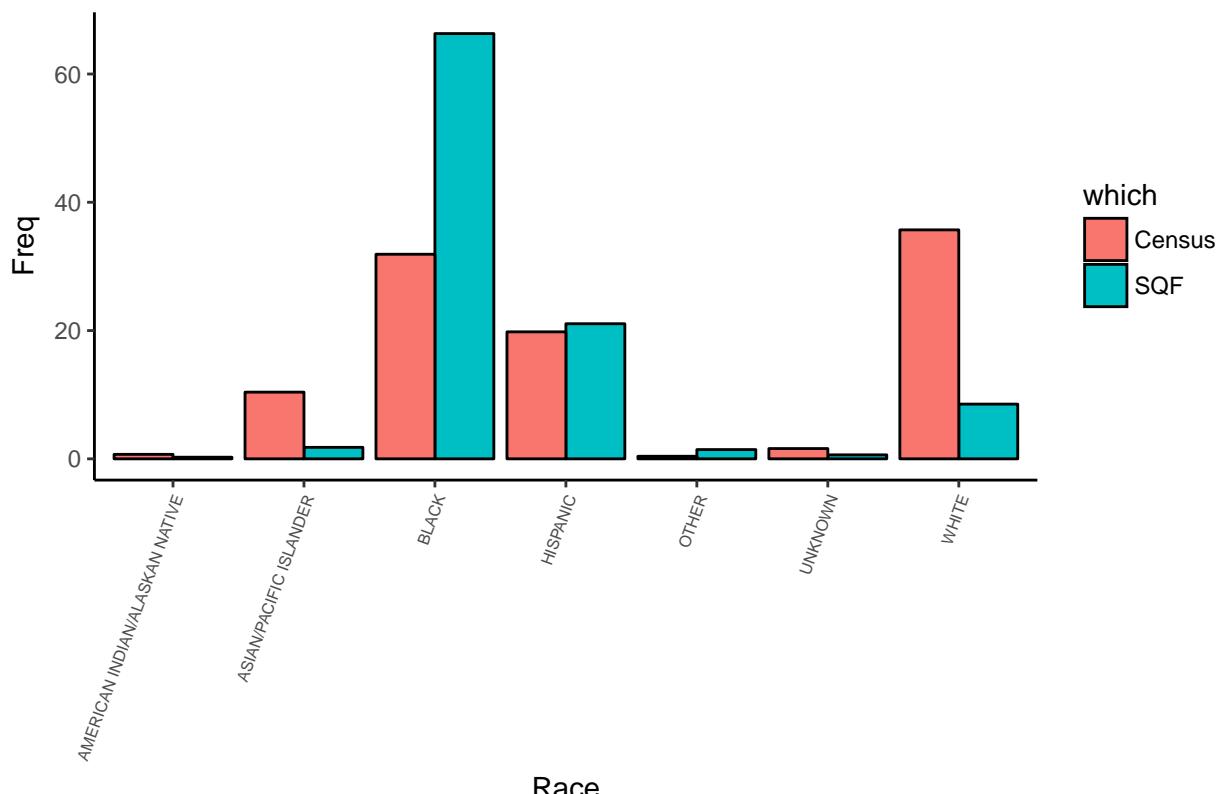
**Question 1: Are people stopped in proportion to the demographic characteristics of the area?**

**Question 2: Are they frisked in proportion to the demographic characteristics of the area?**

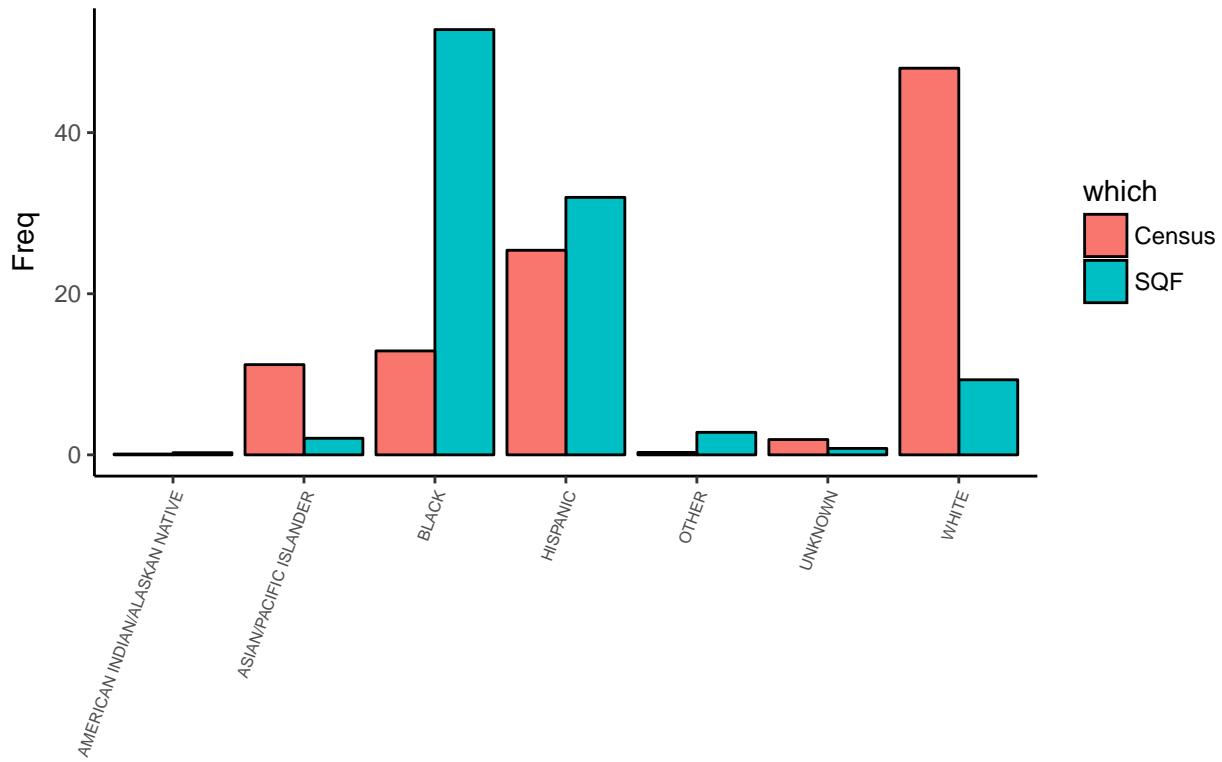
### BRONX Population vs Stopped Population



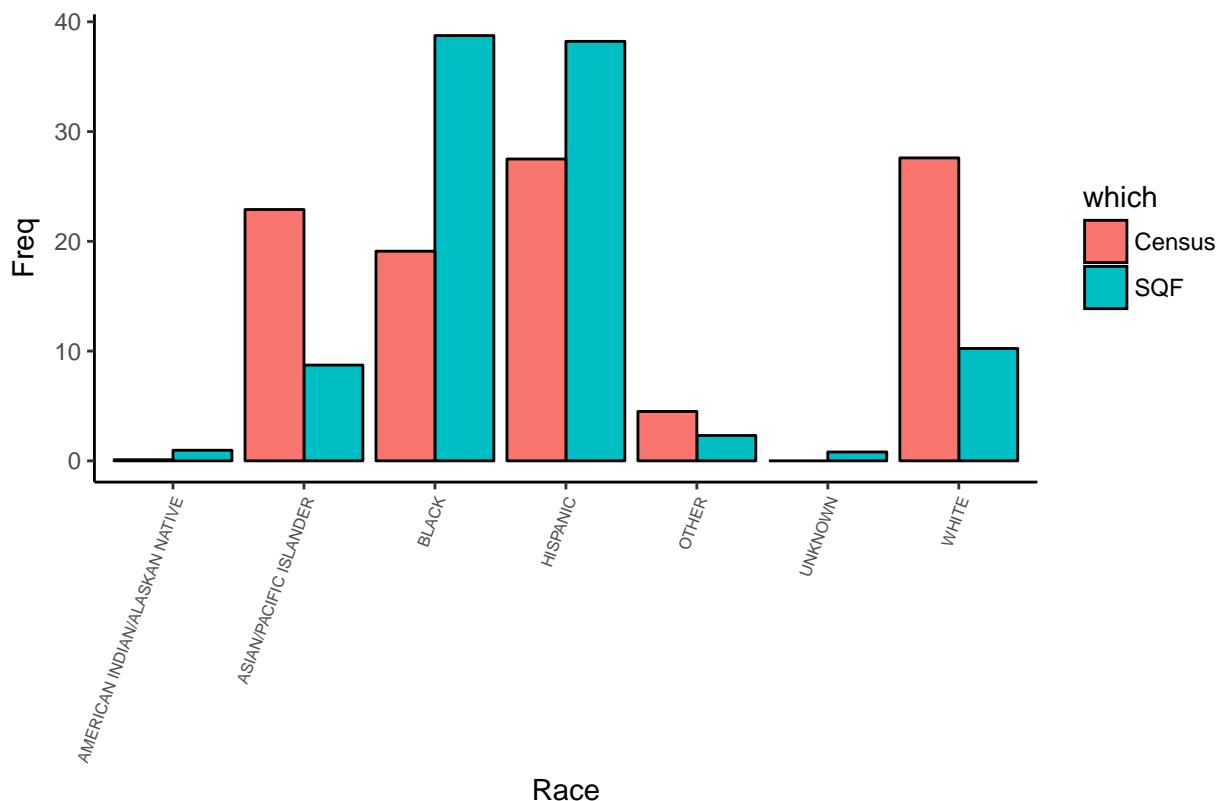
### BROOKLYN Population vs Stopped Population



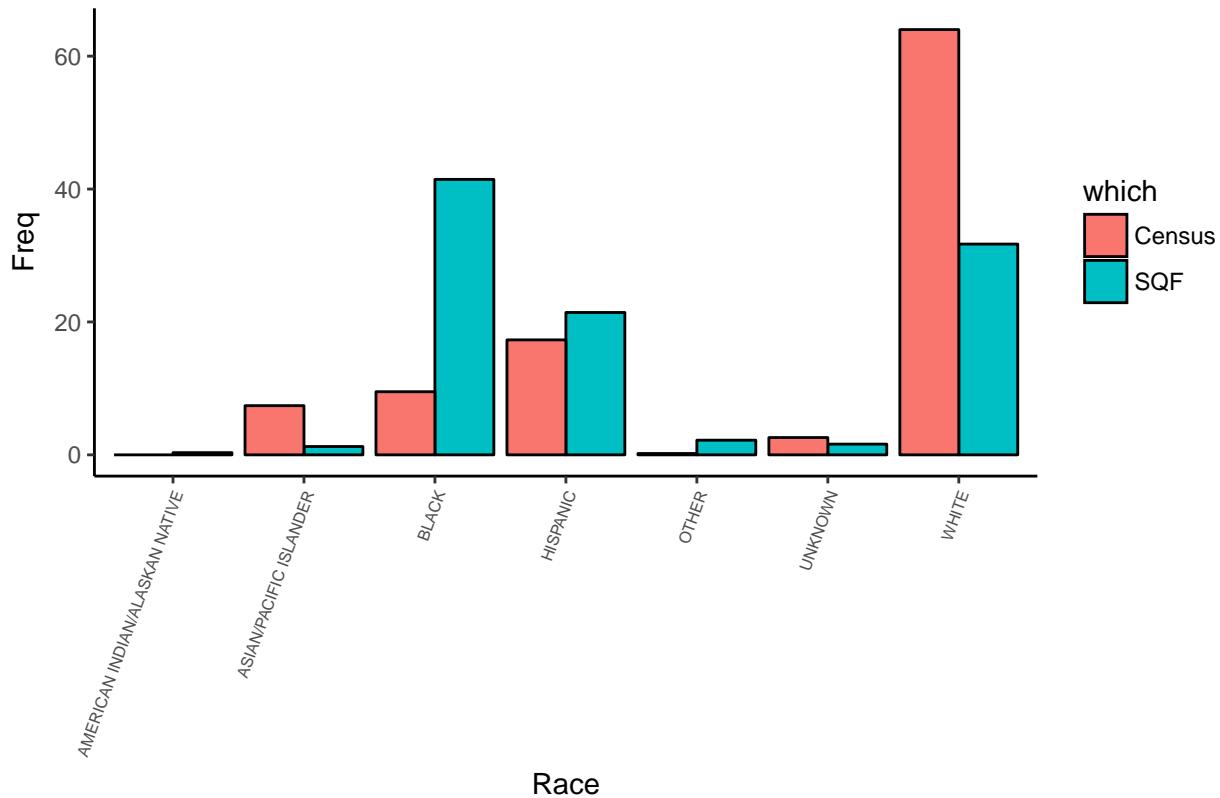
### MANHATTAN Population vs Stopped Population



### QUEENS Population vs Stopped Population

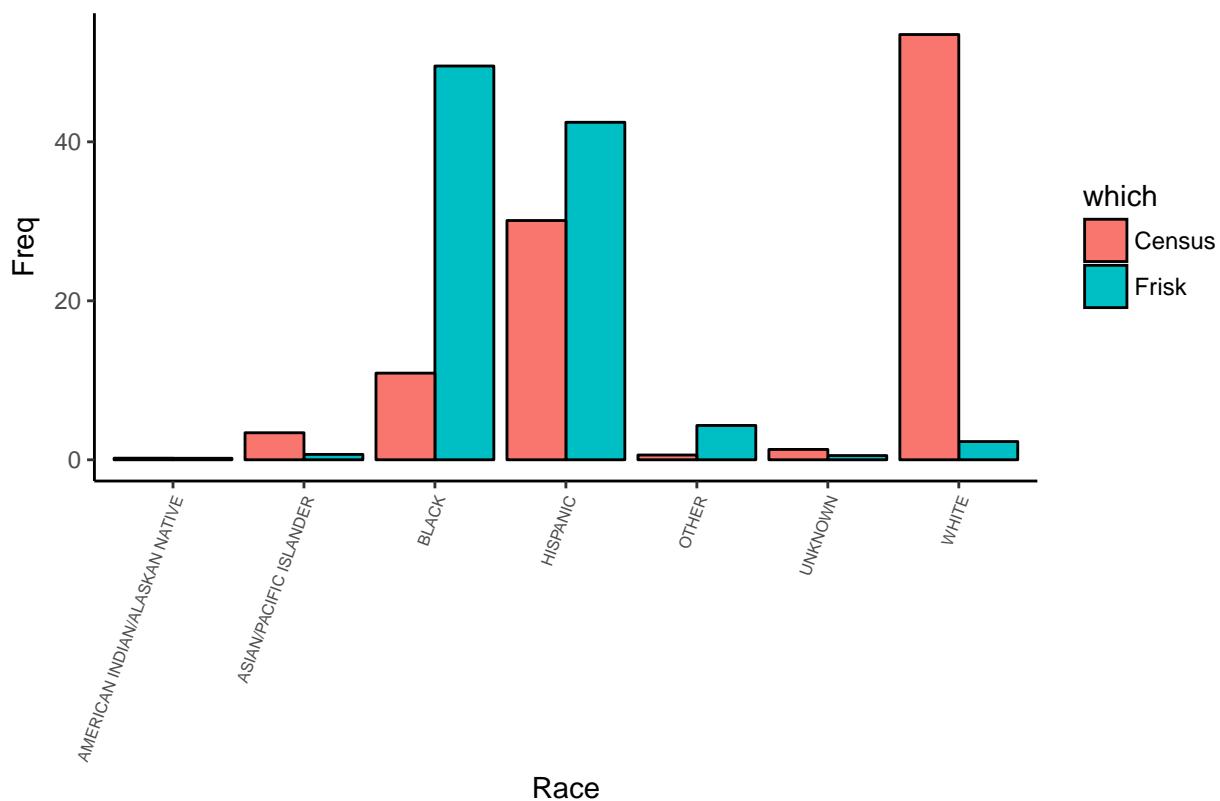


### STATEN IS Population vs Stopped Population



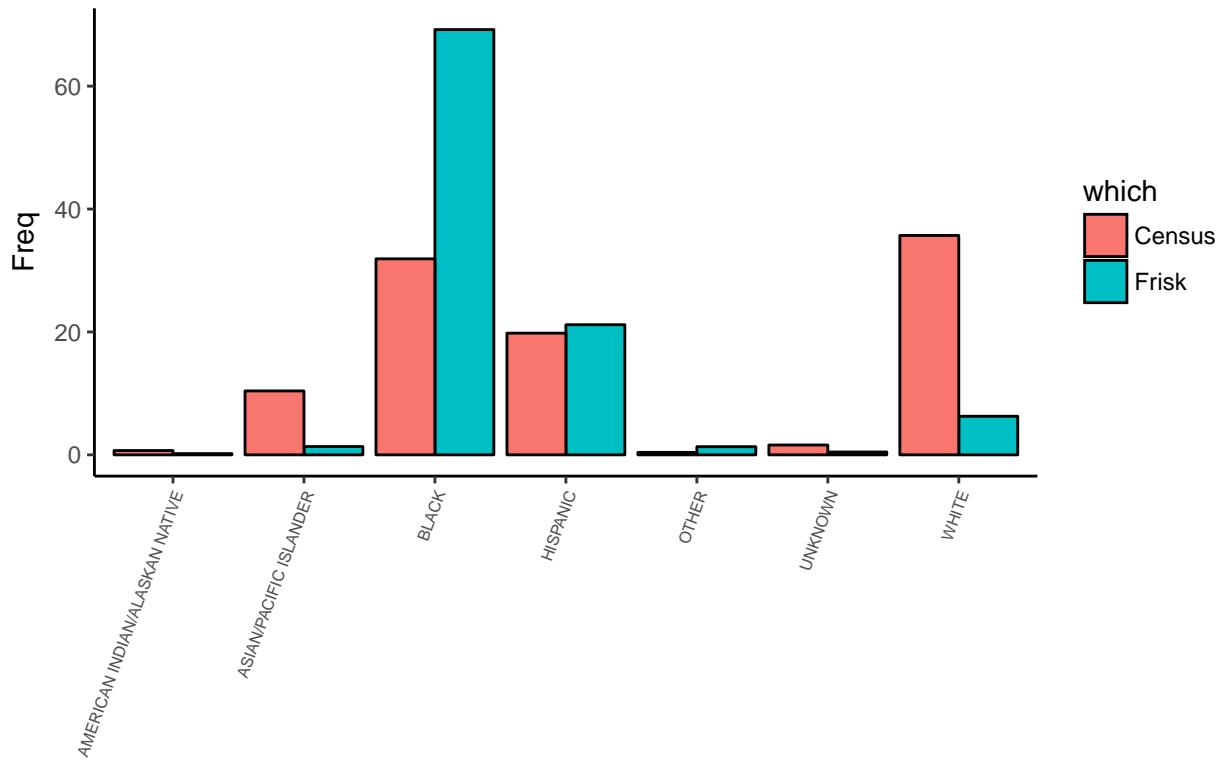
Race

### BRONX Population vs Frisked Population

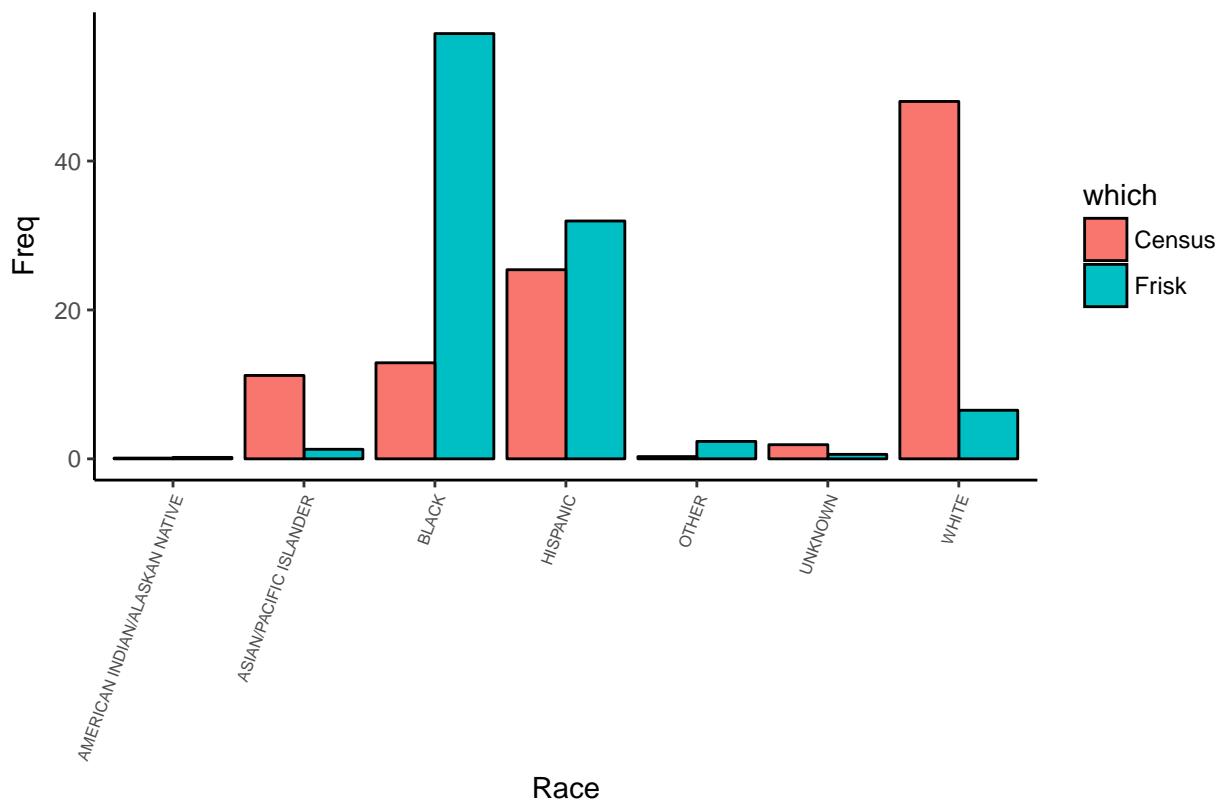


Race

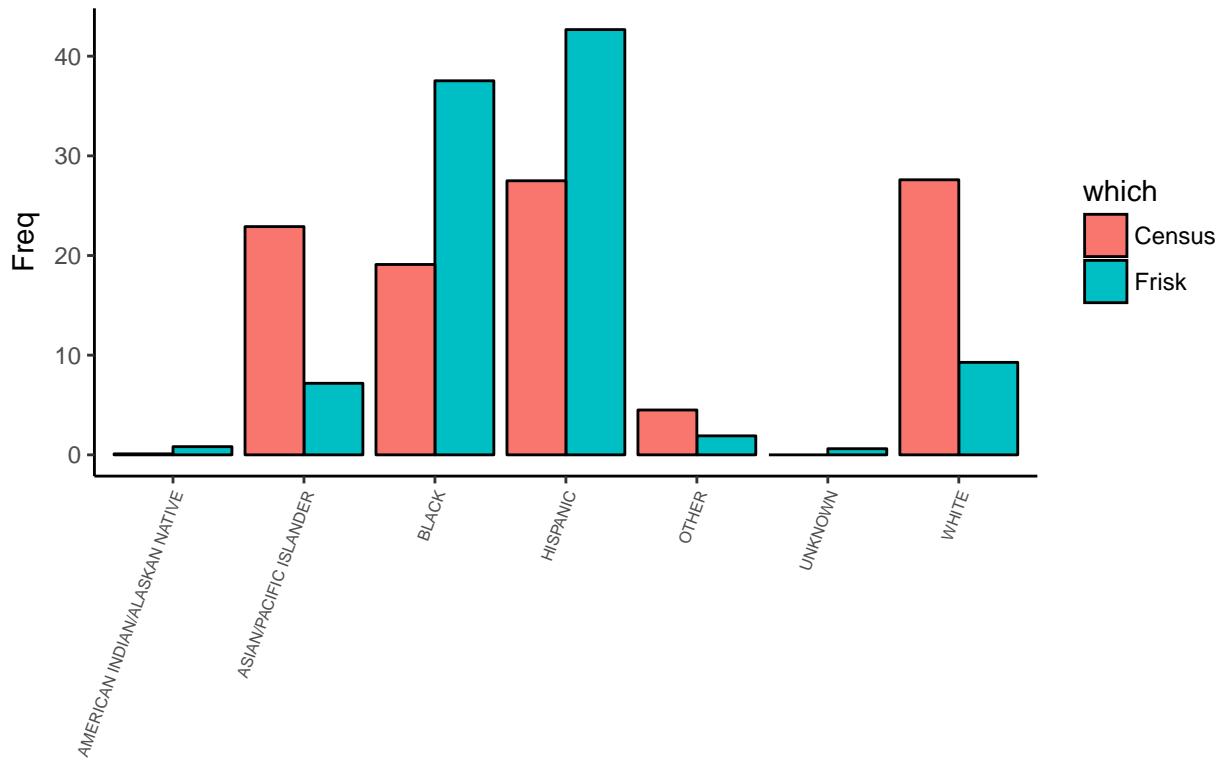
## BROOKLYN Population vs Frisked Population



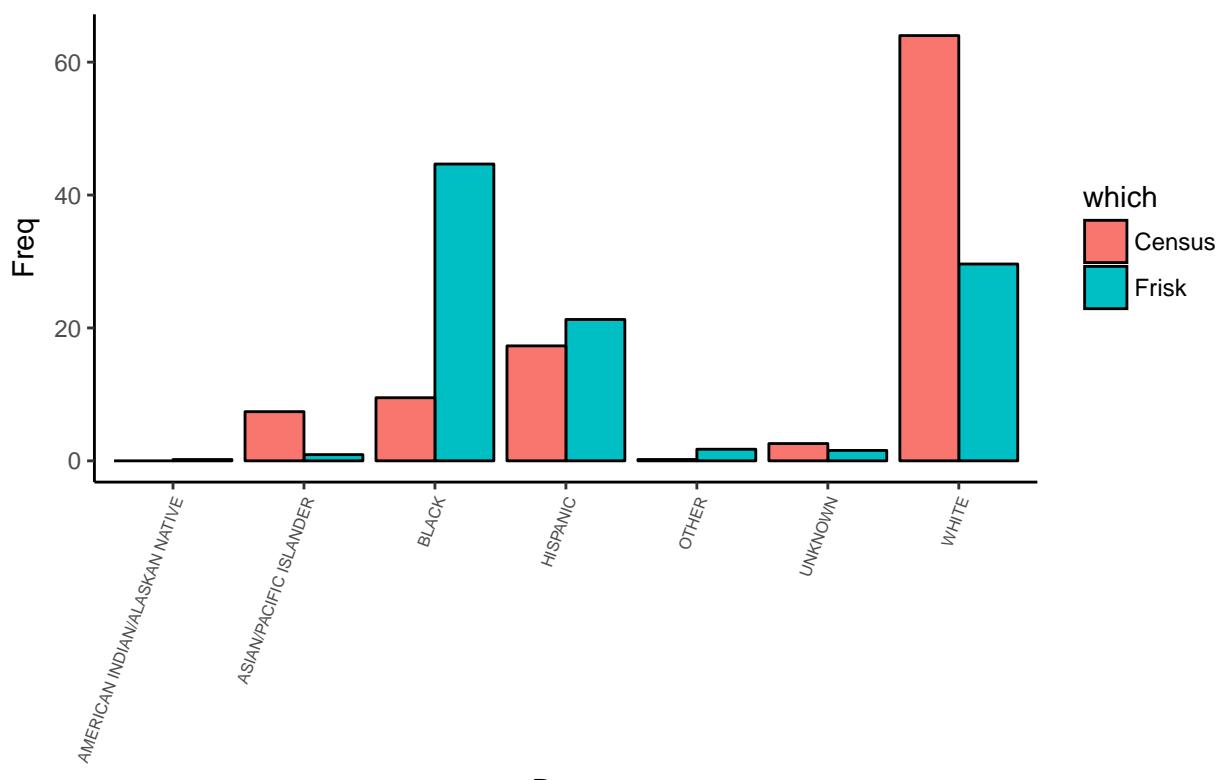
## MANHATTAN Population vs Frisked Population



### QUEENS Population vs Frisked Population



### STATEN IS Population vs Frisked Population



**Analysis** The figures show that there is indeed a difference between the percentage of people of a certain

race stopped or frisked and the percentage of people of a certain race who live in that borough. In order to test if the difference between the percentage of people of a certain race stopped and the percentage of people of a certain race who live in that borough are statistically significant, we used the `prop.test` function in R. For both the black and Hispanic populations we said that our alternative hypothesis was that the percentage of stopped people is greater than the percentage of people living in that borough. We found that the p-value calculated using this function gave us roughly zero for the percentage of stopped people who are black versus the percentage of black people in that borough. Since the p-value is less than 0.05 than we can confidently say that these two percentages are significantly different (which is a problem). The  $H_0$  is rejected. Performing the same `prop.test` for the Hispanic population of each other five boroughs, we get practically identical results, all the p-values are practically 0 (which is less than 0.05) so we can confidently say that these two percentages are significantly different for Hispanics as well. As for the white population, we said the alternative hypothesis was that the percentage of stopped people who are white is less than the percentage of white people in that borough. The p-values we got for each of the five boroughs was approximately 0, rejecting our  $H_0$ . (Refer to R code, and tables at *Team-Chosen Question 1 Chunks 1 and 2 Code* in “Group Code and Graphs.rmd”) This means that we can say that the percentage of white people in the borough is significantly greater than the percentage of stopped people that are white.

We also did the same `prop.test` for the percentage of people frisked in the borough versus percentage of people living in the borough. For the African American population we found that the p-value also approximated to 0 for each of the black boroughs for the some hypothesis we had above. Our results were also identical to the ones above for both Hispanics and Whites. (Refer to R code, and tables at *Team-Chosen Question 2 Chunks 1 and 2 Code* in “Group Code and Graphs.rmd”)

### **Concluding Remarks**

After looking at all this data, we can make a few important conclusions. First, in New York City and most likely the rest of the United States, our police departments do racial profiling and stop and frisk an disproportionate amount of African American and Hispanic people. Second, there must be changes made to this problem. Some proposed changes could be a decrease in the amount of stop and frisk that police departments around the nation and more particularly, in NYC, do. In addition, racial profiling must be addressed in the education and training of police officers around the United States in order to ensure equal treatment of all citizens and residents of this country.

### **Bibliography**

Barone, M., “Stop-and-Frisk Protects Minorities”, <http://www.nationalreview.com/article/356481/stop-and-frisk-protects-minorities-michael-barone>, 2013.

New York Civil Liberties Union website: <http://www.nyclu.org/node/1598>

New York Civil Liberties Union, “Stop and Frisk During the Bloomberg Administration”, [http://www.nyclu.org/files/publications/stopandfrisk\\_briefer\\_2002-2013\\_final.pdf](http://www.nyclu.org/files/publications/stopandfrisk_briefer_2002-2013_final.pdf)

Geller, A., Fagan, J., Tyler, T., Link, B., “Aggressive Policing and the Mental Health of Young Urban Men”, *Am J Public Health*. 2014 December; 104(12): 2321-2327. Published online 2014 December. doi: 10.2105/AJPH.2014.302046