

# Project proposal

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

10/09/20

## Setup

```
library(tidyverse)
police <- read.csv("data/nc_durham_2020_04_01.csv")
```

## Section 1. Introduction

In the wake of the election and the Black Lives Matter movement, it is essential that the newest generation of voters educate themselves on the systemic racism within the criminal justice and policing system. The US incarcerates more people than any other country (<https://www.sentencingproject.org/criminal-justice-facts/>), and people of color make up a disproportionate percent of the prison population. Police funding has grown significantly over the past four decades, and overpolicing in communities of color is a serious issue. In Michelle Alexander's book, *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*, she discusses the rise of incarceration rates for black and brown people in the US. Alexander cites the War on Drugs as one of the biggest causes of contemporary mass incarceration, and she uses police pretext stops as an example. In pretext stops, cops can pull over a "suspicious" driver on the pretext of a very minor traffic violation (e.g. turning on red, going over the speed limit) and then do a drug sweep of the car, which may result in an arrest for drug-related charges. According to a Pew Research Center survey (<https://www.pewresearch.org/fact-tank/2020/06/03/10-things-we-know-about-race-and-policing-in-the-u-s/>), "black adults are about five times as likely as whites to say they've been unfairly stopped by police because of their race or ethnicity." Given this information, we wish to investigate traffic stops in Durham from the Stanford Open Policing Project. 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?

Ultimately, in continuation with past literature, 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 (<http://censusviewer.com/city/NC/Durham>).

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.

## Section 2. Data description

The data set is a census of individual police stops in Durham created by the Stanford Open Policing project. It has 29 variables and 326024 observations.

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.

The data in this dataset was collected as part of the Stanford Open Policing Project which “[collects] and [standardizes] data on vehicle and pedestrian stops from law enforcement departments across the country”(https://openpolicing.stanford.edu/) According to the website, the Stanford Open Policing Project has compiled data from “21 state patrol agencies and 29 municipal police departments, comprising nearly 100 million traffic stops.”(https://openpolicing.stanford.edu/findings/) One of the challenges which the website mentioned was the lack of information collected in traffic stops, and how they had to find states which track the data and release it in ways which can then be standardized.(https://openpolicing.stanford.edu/findings/)

### Section 3. Glimpse of data

```
glimpse(police)
```

```
## Rows: 326,024
## Columns: 29
## $ raw_row_number    <int> 1219541, 1221117, 1221124, 1221133, 1221146,...
## $ date              <chr> "2001-12-28", "2002-01-02", "2002-01-02", "2...
## $ time              <chr> NA, "04:00:00", NA, "19:50:00", "20:45:00", ...
## $ location          <chr> "Durham, Durham County", "Durham, Durham Cou...
## $ county_name       <chr> "Durham County", "Durham County", "Durham Co...
## $ subject_age       <int> 22, 28, 21, 28, 36, 68, 24, 30, 38, 56, 28, ...
## $ subject_race      <chr> "black", "hispanic", "black", "black", "blac...
## $ subject_sex       <chr> "female", "male", "male", "female", "female"...
## $ officer_id_hash   <chr> "5b9908c1a5", "0c140e68b5", "0c140e68b5", "4...
## $ department_name   <chr> "Durham Police Department", "Durham Police D...
## $ type              <chr> "vehicular", "vehicular", "vehicular", "vehi...
## $ arrest_made       <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FA...
## $ citation_issued   <lgl> FALSE, TRUE, FALSE, TRUE, FALSE, FALSE, FALS...
## $ warning_issued    <lgl> TRUE, FALSE, TRUE, FALSE, TRUE, TRUE, FALSE,...
## $ outcome           <chr> "warning", "citation", "warning", "citation"...
## $ contraband_found   <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ contraband_drugs   <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ contraband_weapons <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ frisk_performed    <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FA...
## $ search_conducted   <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FA...
## $ search_person      <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FA...
## $ search_vehicle     <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FA...
## $ search_basis       <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ reason_for_frisk   <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ reason_for_search  <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ...
## $ reason_for_stop    <chr> "Investigation", "Stop Light/Sign Violation"...
## $ raw_Ethnicity      <chr> "N", "H", "N", "N", "N", "N", "H", "N", "N",...
```

```
## $ raw_Race          <chr> "B", "W", "B", "B", "B", "B", "B", "W", "B",...  
## $ raw_action_description <chr> "Verbal Warning", "Citation Issued", "Verbal..."
```