BIAS IN DATA

TABLE OF CONTENT

01 INTRODUCTION

You can describe the topic of the section here

02 TYPES OF BIASES IN DATA

Some bias can be subconscious

03 IMPACT OF BIASES

You can describe the topic of the section here

04 DATA SCIENCE & ETHICS

You can describe the topic of the section here

01

INTRODUCTION

What is bias? How does it happen? Why is understanding “bias in data” important?

INTRODUCTION

As humans, we have many biases, both implicit and explicit. Biases are systematic errors in thinking influenced by cultural and personal experiences. Biases distort our perception and cause us to make incorrect decisions.

Computers, data, and algorithms are not actually completely objective. It is true that data analysis can help us make better decisions, but it is not immune to bias. Humans create technologies and algorithms. As a result, they often have human biases encoded into them.

02

TYPES OF BIASES IN DATA

Different types of biases that can manifest in training data.

TYPES OF BIAS

CONFIRMATION BIAS

Venus is the second planet from the Sun

HISTORICAL BIAS

AVAILABILITY BIAS

SELECTION BIAS

It’s the closest planet to the Sun and the smallest

SURVIVORSHIP BIAS

OUTLIER BIAS

CONFIRMATION BIAS

Venus is the second planet from the Sun

HISTORICAL BIAS

Venus is the second planet from the Sun

CONFIRMATION BIAS

Venus is the second planet from the Sun

CONFIRMATION BIAS

Venus is the second planet from the Sun

CONFIRMATION BIAS

Venus is the second planet from the Sun

CONFIRMATION BIAS

Venus is the second planet from the Sun

03

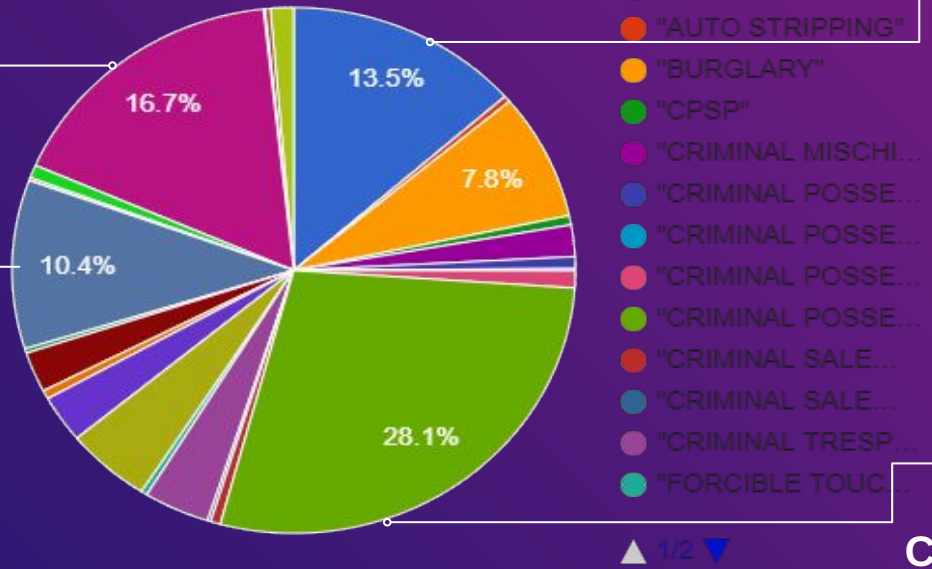
IMPACT OF BIASES IN DATA

Different types of biases that can manifest in training data.

PERCENT OF PERSONS SUSPECTED BY CRIME

16.7%
Robbery

10.4%
Petit Larceny



13.5%
Assault

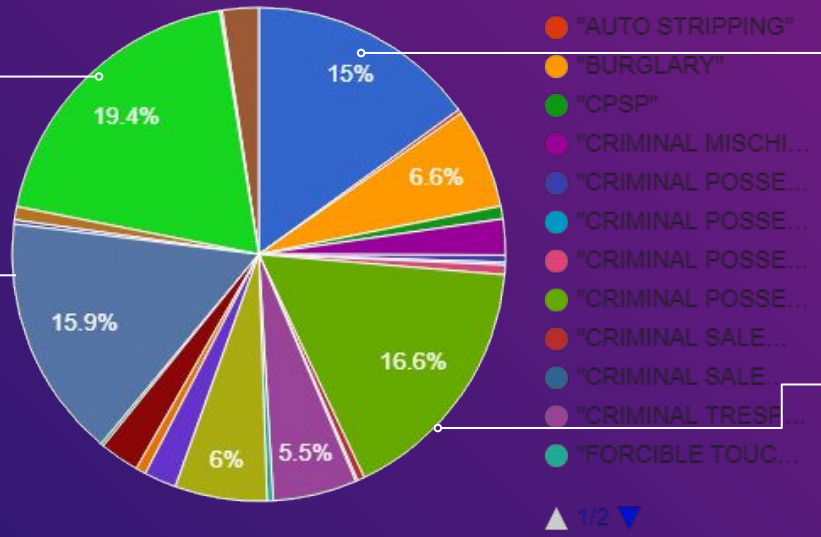
28.1%
Criminal Possession of a weapon

Code: suspected-crimes = pie-chart(nypd-table."suspected-of")

PERCENT OF ARRESTED PERSONS BY CRIMES

19%
Robbery

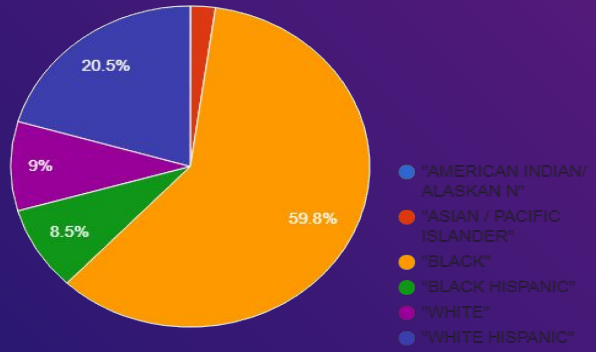
15.9%
Petit Larceny

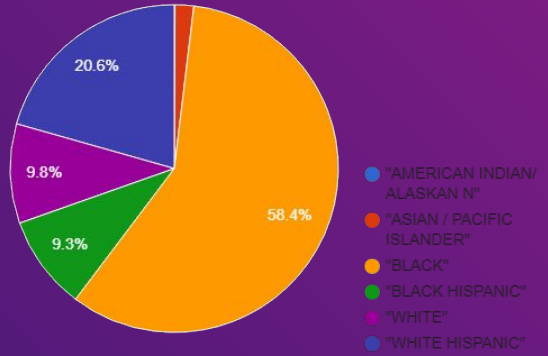


15%
Assault

16.6%
Criminal Possession of a weapon

PIE CHART (ARRESTED VS SUSPECTED BY RACE)

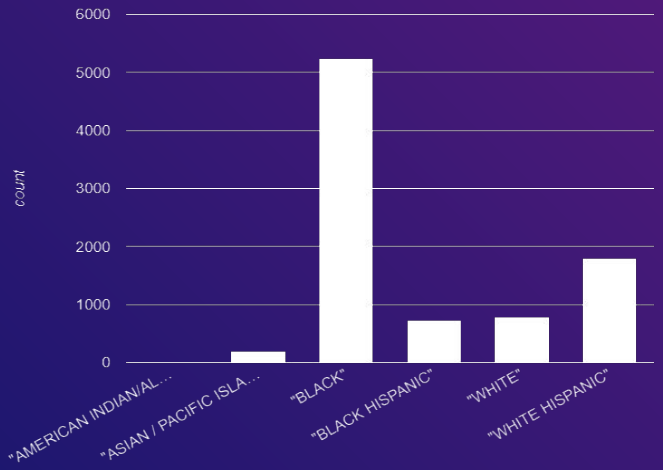


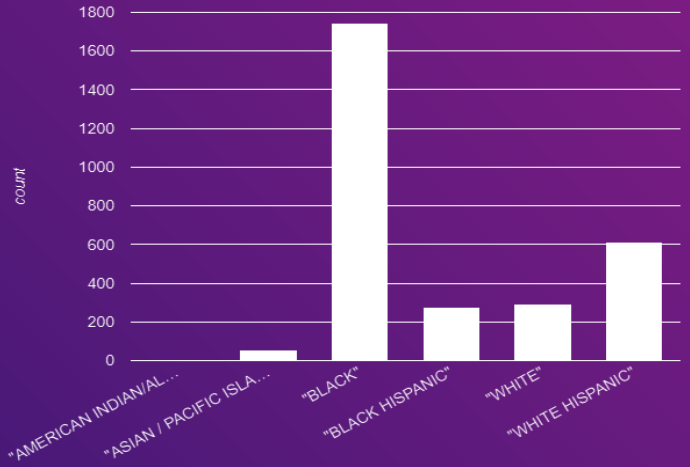


Charts represent the suspects who were stopped and frisked and those who were actually arrested for their suspected crimes. Compared to other races Blacks are twice times more likely to either be suspected or arrested during these stop and frisks. Does this mean that Blacks commit the most crimes? Or are they just twice times more likely to be stopped by a police officer and suspected of a crime. Is that a fair assumption?

Code: pie-chart(nypd-race-table,"race") | pie-chart(nypd-stop-and-frisk-arrests-table,"race")

SUSPECTED OR ARRESTED BASED ON RACE





Charts represent the suspects who were stopped and frisked and those who were actually arrested for their suspected crimes. Compared to other races Blacks are twice times more likely to either be suspected or arrested during these stop and frisks. Does this mean that Blacks commit the most crimes? Or are they just twice times more likely to be stopped by a police officer and suspected of a crime. Is that a fair assumption?

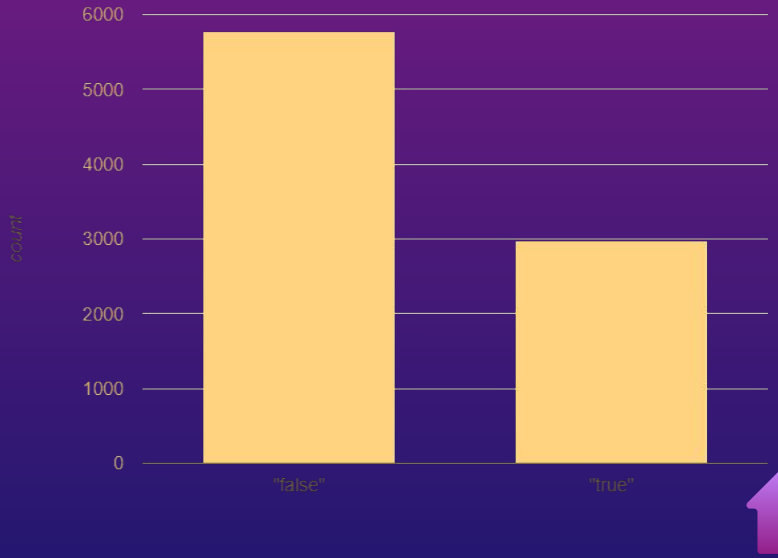
ALL STOPPED - NUMBER ARRESTED

What suspected stop and frisk crimes lead to arrest?

Out of 8764 stop and frisk records for 2019, only 2986 lead to an actual arrest.

This number does not tell whether the number of arrests were actually convicted of the suspected crime.

One can conclude that about 65% of stop and frisk detainments are unnecessary and harassment by the cops. (Would this be accurate based on the chart?)

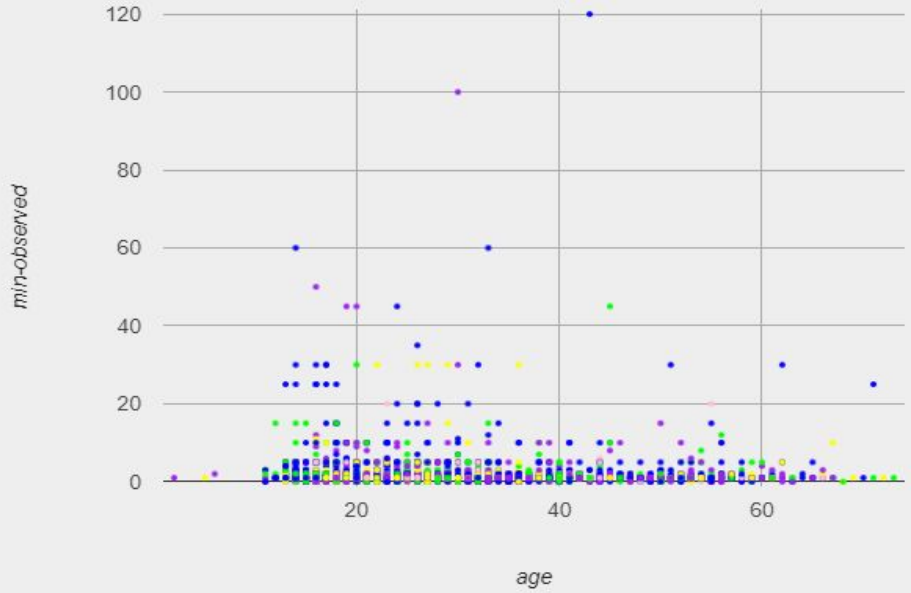


BOROUGHS OF ARRESTED SUSPECTS

- Scatter plot was made to show the different boroughs arrests were made in the stop and frisks, based on age and minutes observed for the suspected crimes.

Surprisingly, the main Borough that arrests were made in was Manhattan.

Is it a fair assumption to make, that if you are a black teenager or young adult, and you visit the Manhattan area that you are likely to be suspected of and arrested for a crime?



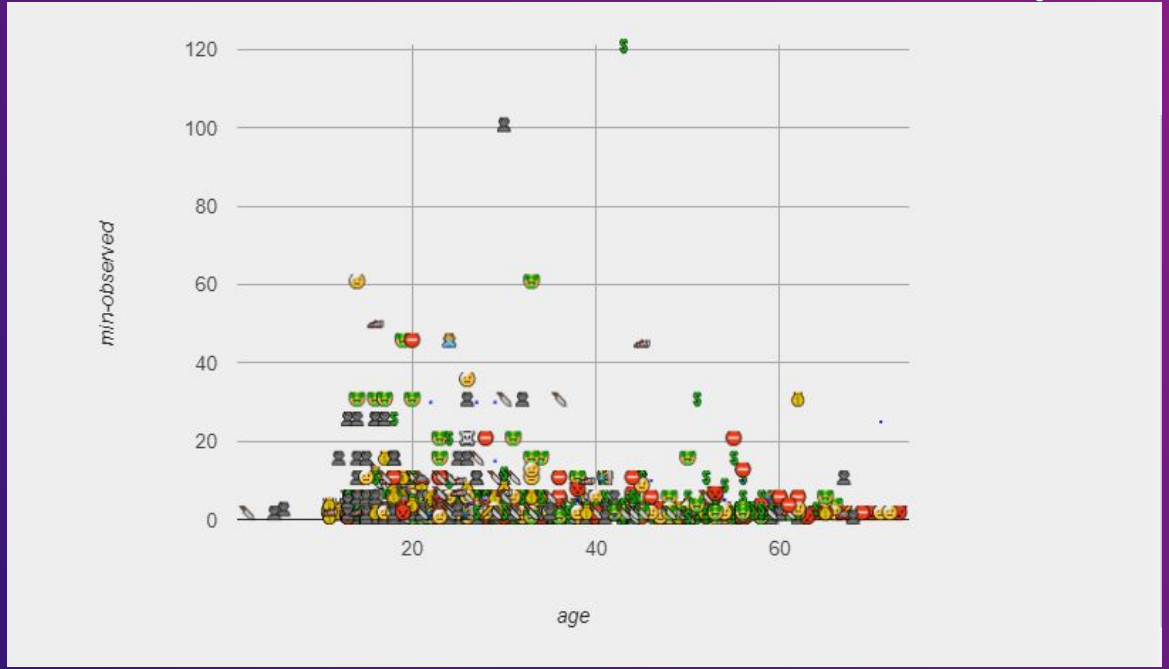
Code:

ARRESTED BY CRIME BASED ON AGE AND TIME OBSERVED

We analyzed whether the types of crimes had a correlation to age and whether that in itself had an impact on the time taken to observe the crime. We filtered to only show the individuals who were actually arrested.

Most crimes committed by teenagers to young adults appeared to be robbery. Most of the suspects were arrested within minutes of being suspected.

Is it fair to say that law enforcement is targeting this population or that this population prefers to rob others and does not like to work for the things they want?

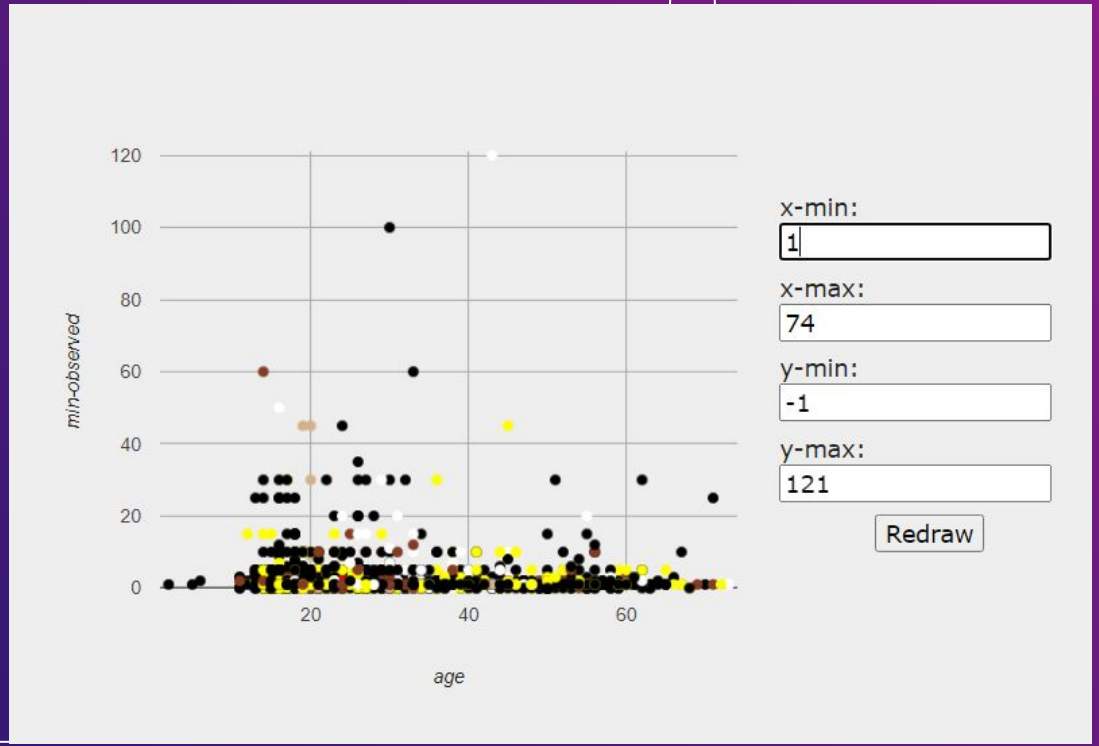


RACE OF ARRESTED SUSPECTS

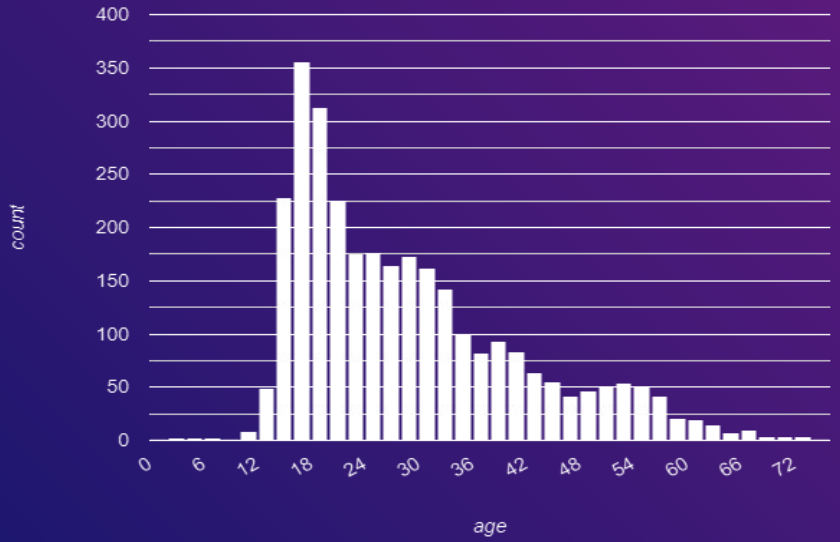
Scatter plot was made to show the different races of suspects arrested.

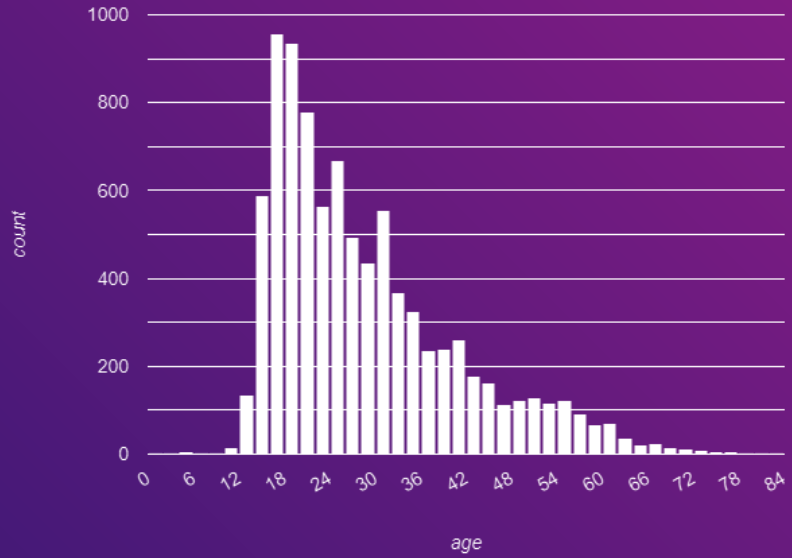
Here we see that blacks no matter the age are predominantly the ones that are arrested for suspected crimes.

Does this mean that Blacks are the main targets of stop and frisk and law enforcement?



HISTOGRAM - AGE RANGE OF ARRESTED VS SUSPECTED





Both charts represent the age range for both arrested and suspected stop and frisks. We can interpret that the general age range for both arrests and suspected cases is between 16 - 24.

04

DATA SCIENCE & ETHICS

Different types of biases that can manifest in training data.

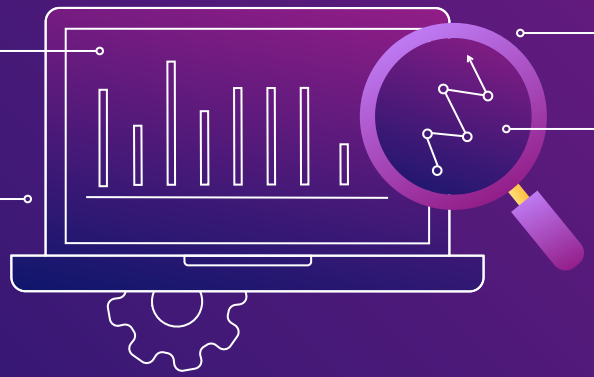
HISTORY OF DATA SCIENCE

MERCURY

It’s the closest planet to the Sun and the smallest one

VENUS

Venus is the second planet from the Sun in the Solar System



MARS

Mars is actually a cold place. It’s full of iron oxide dust

NEPTUNE

It’s the farthest planet from the Sun. It's also the fourth-largest

OUR TEAM

LATOYA BOLAND

Math & Computer Science Teacher
A-Tech High School

ALICIA WADE

Computer Science
East Williamsburg Scholars

RESOURCES

- **Intro**
  - Intro - Resource 1
- **Types of Biases**
  - Types of Biases - Resource 1
  - Types of Biases - Resource 2
- **Data Science - Coding**
  - Bootstrap Data Science
- **Stop and Frisk Spreadsheet**
- **Code**