



NYPD Arrest Records

2006 - 2019

A Project by Yasmine Hays & Ryland Matthews

Problem Statement

Officers can exercise a high degree of discretion enforcing nonviolent crimes

The NYPD has a history of racially disparate outcomes in their arrest record, but is it so severe as to be predictable?

Background

A long history of data in policing, and racially disparate outcomes

CompStat

- Adopted in 1995
- NYPD one the first in the nation to integrally adopt ‘data-driven-policing’ systems

Stop & Frisk

- Ruled Constitutional in 1968 in Terry v. Ohio
- Widely practiced by NYPD and consciously accelerated under Mayor Bloomberg

Broken Windows Policing

- A policing theory developed in the 1990’s
- Formally adopted by NYPD under Mayor Rudy Giuliani and Police Commissioner William Bratton

Datasets

Census Data

- 8.42 million people
- 27,012 people/sq. mi.
- Demographics:
 - 32.5% White
 - 28.9% Hispanic
 - 22.3% Black
 - 13.5% Asian



Datasets

Arrest Records

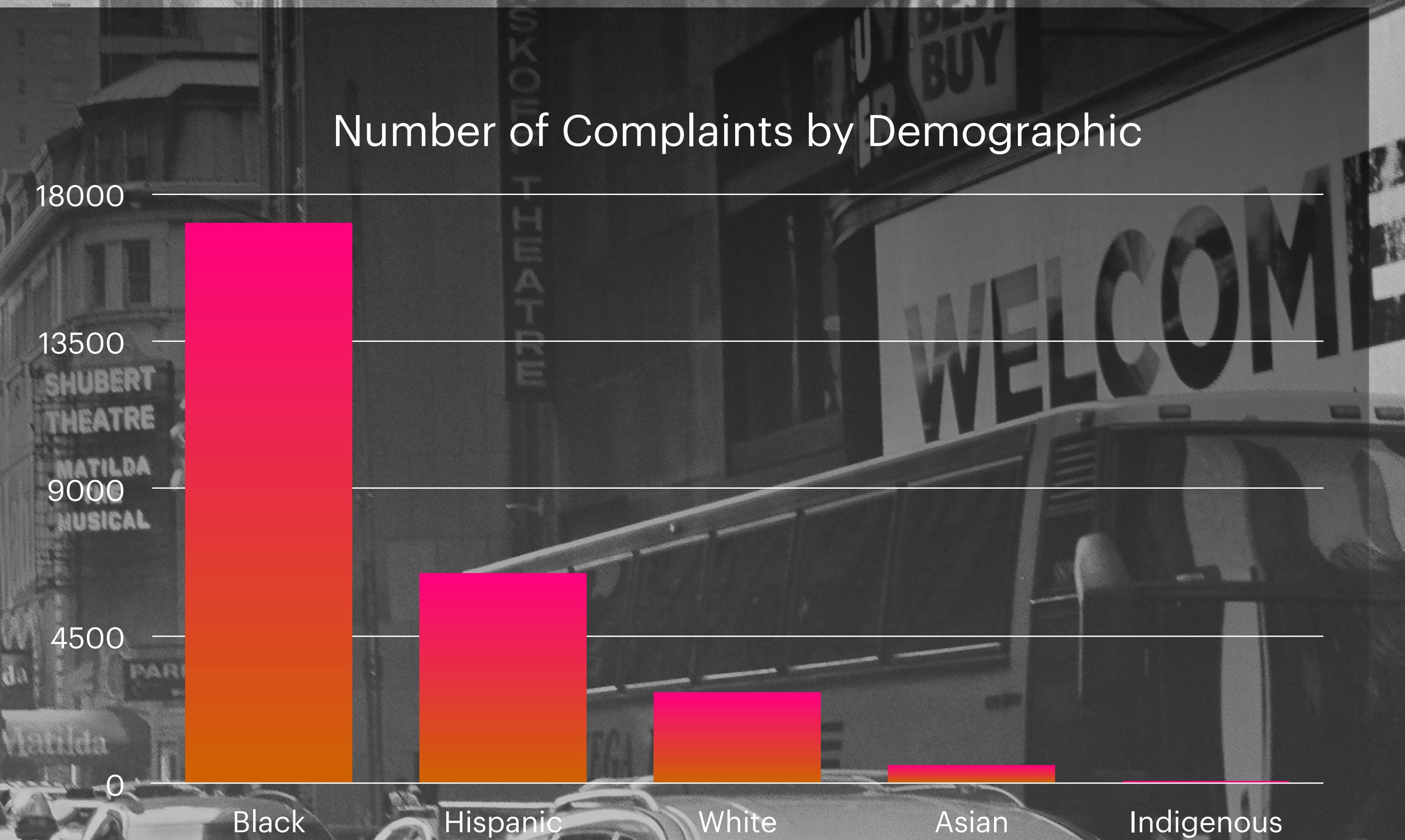
- Covers 2006-2019
- Almost 5 million arrest records
- 77 NYPD Precincts
- 418 types of arrests recorded



Datasets

Civilian complaints against the NYPD

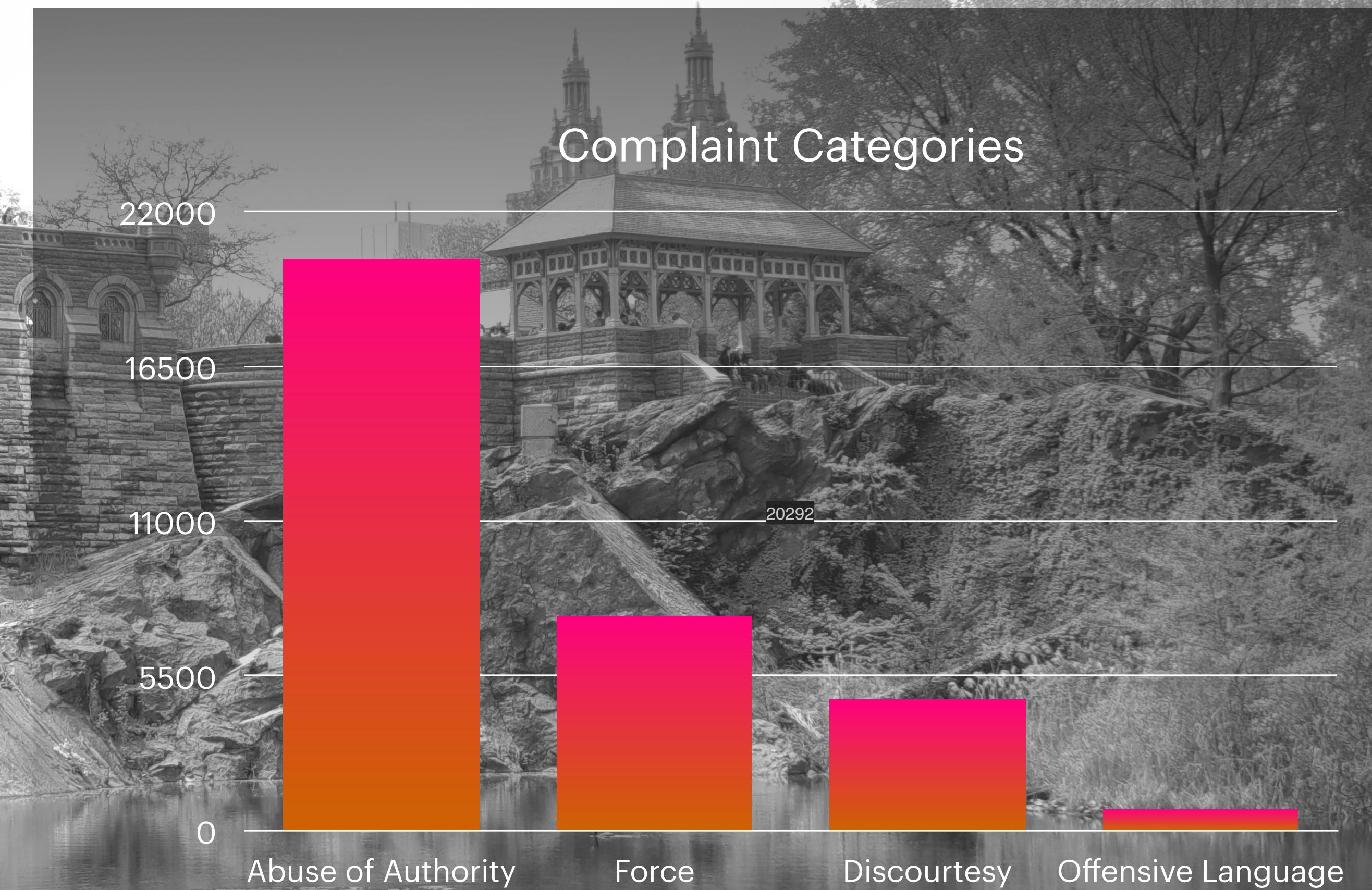
- 33334 Complaints filed
- Spanning 1985 to 2020
- By demographics:
 - 59.2% Black filers
 - 22.2% Hispanic filers
 - 9.6% White filers
 - 1.8% Asian filers



Complaints Exploratory Data Analysis

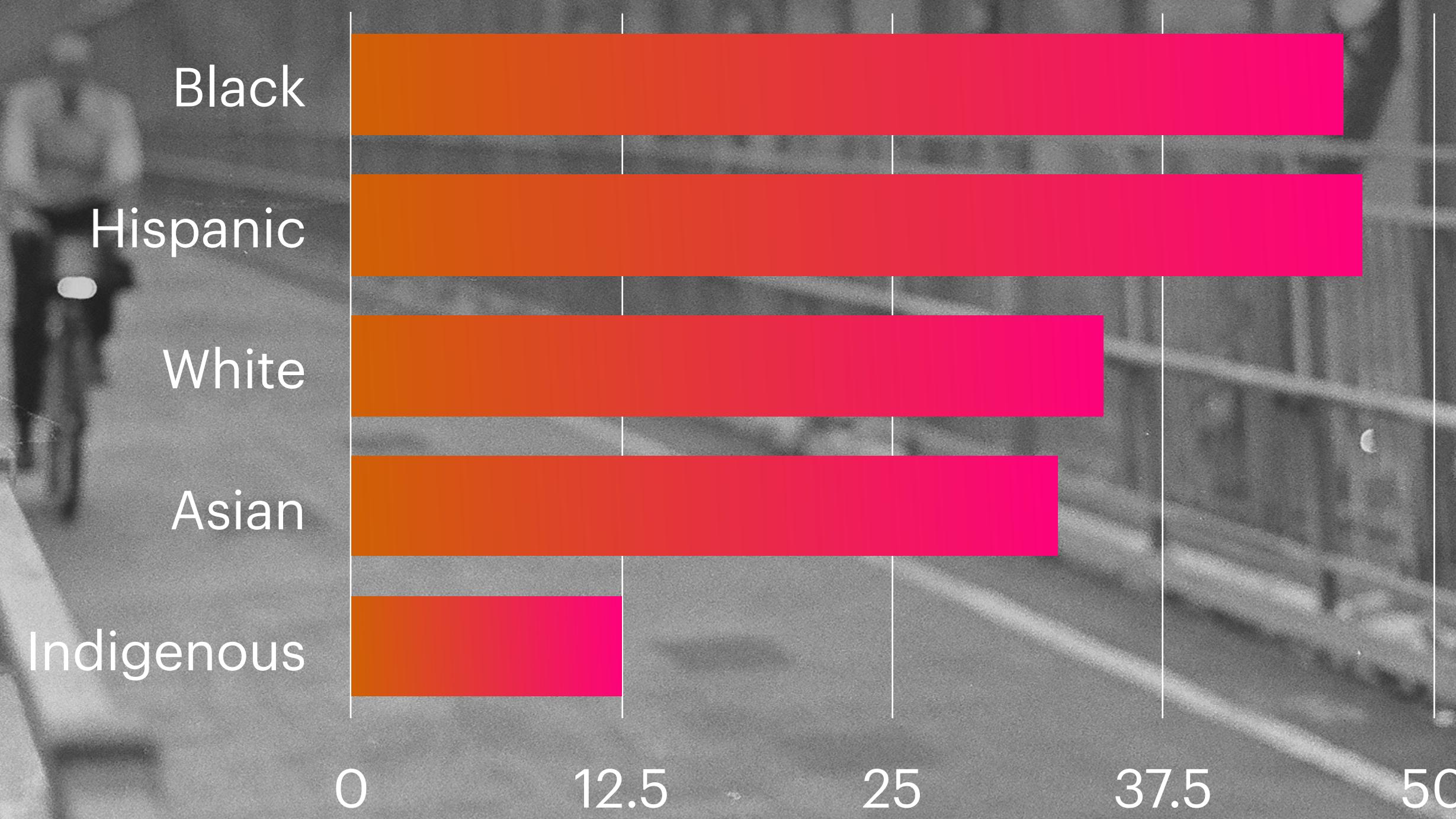
NYPD files complaints under four categories

The top allegations against NYPD police officers are of a physical nature for Black and Hispanic People, but of a verbal nature for white people and Asian people. Allegations made by Indigenous people mostly pertained to being searched.



Complaints Exploratory Data Analysis

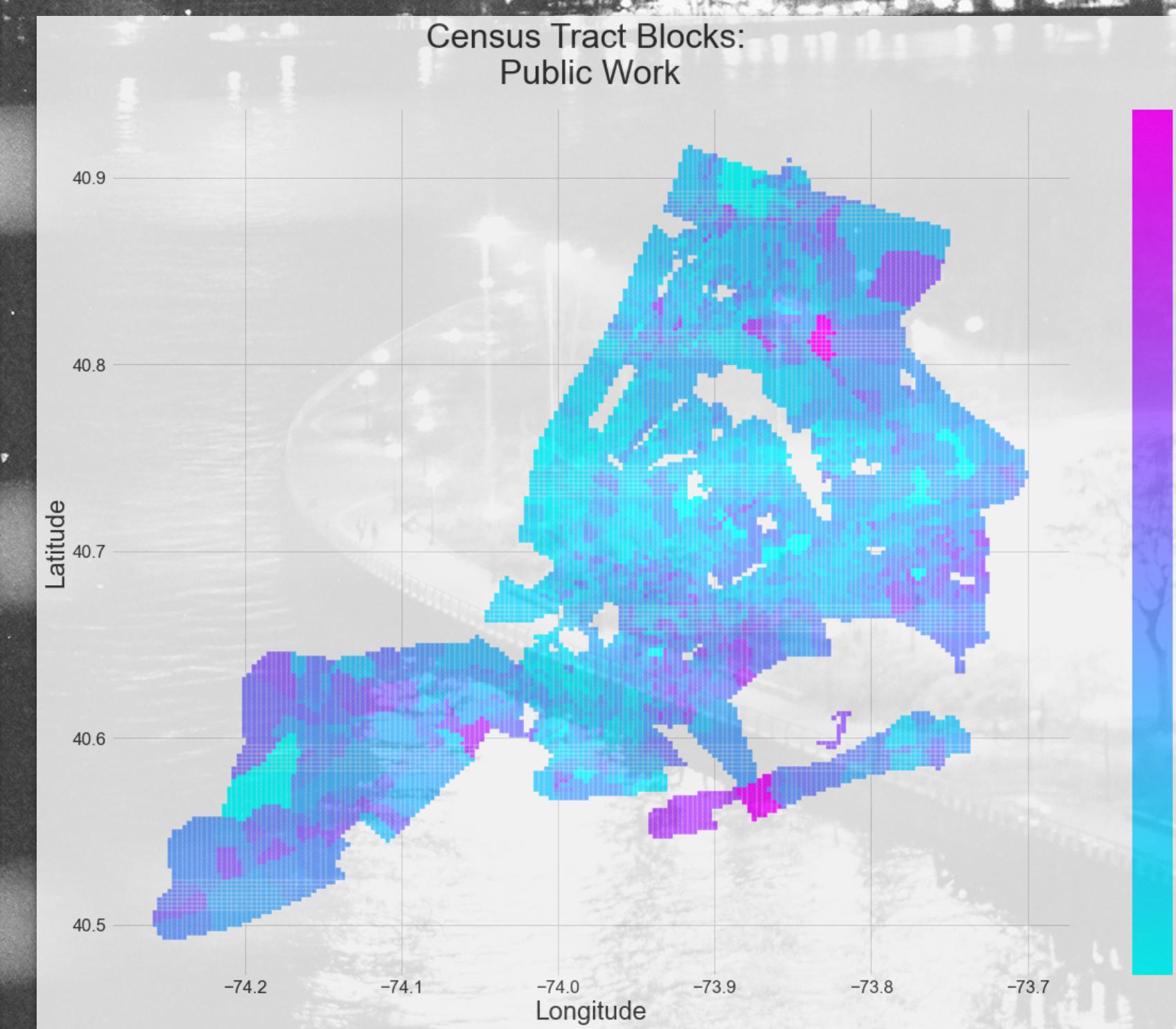
Complaints filed by arrestees



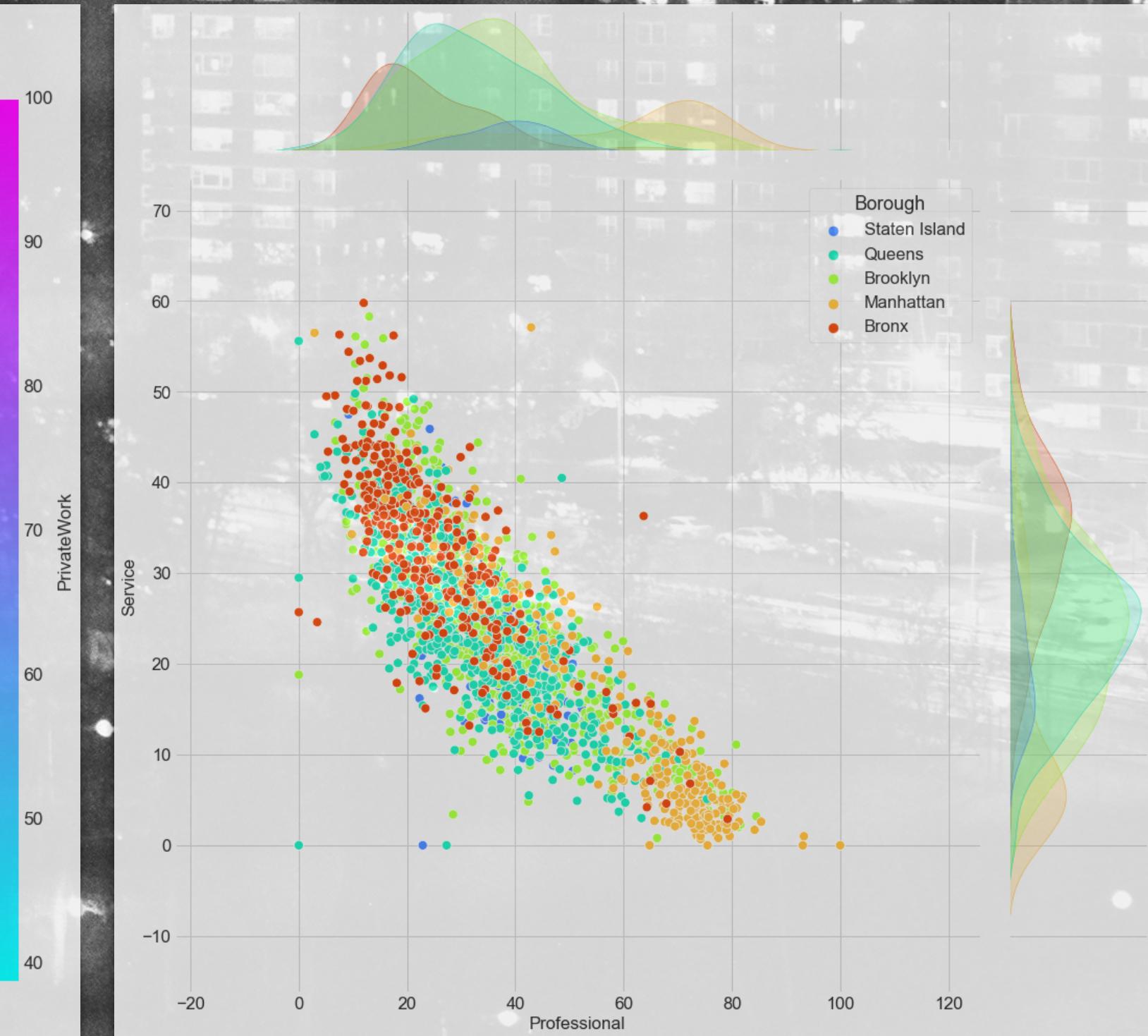
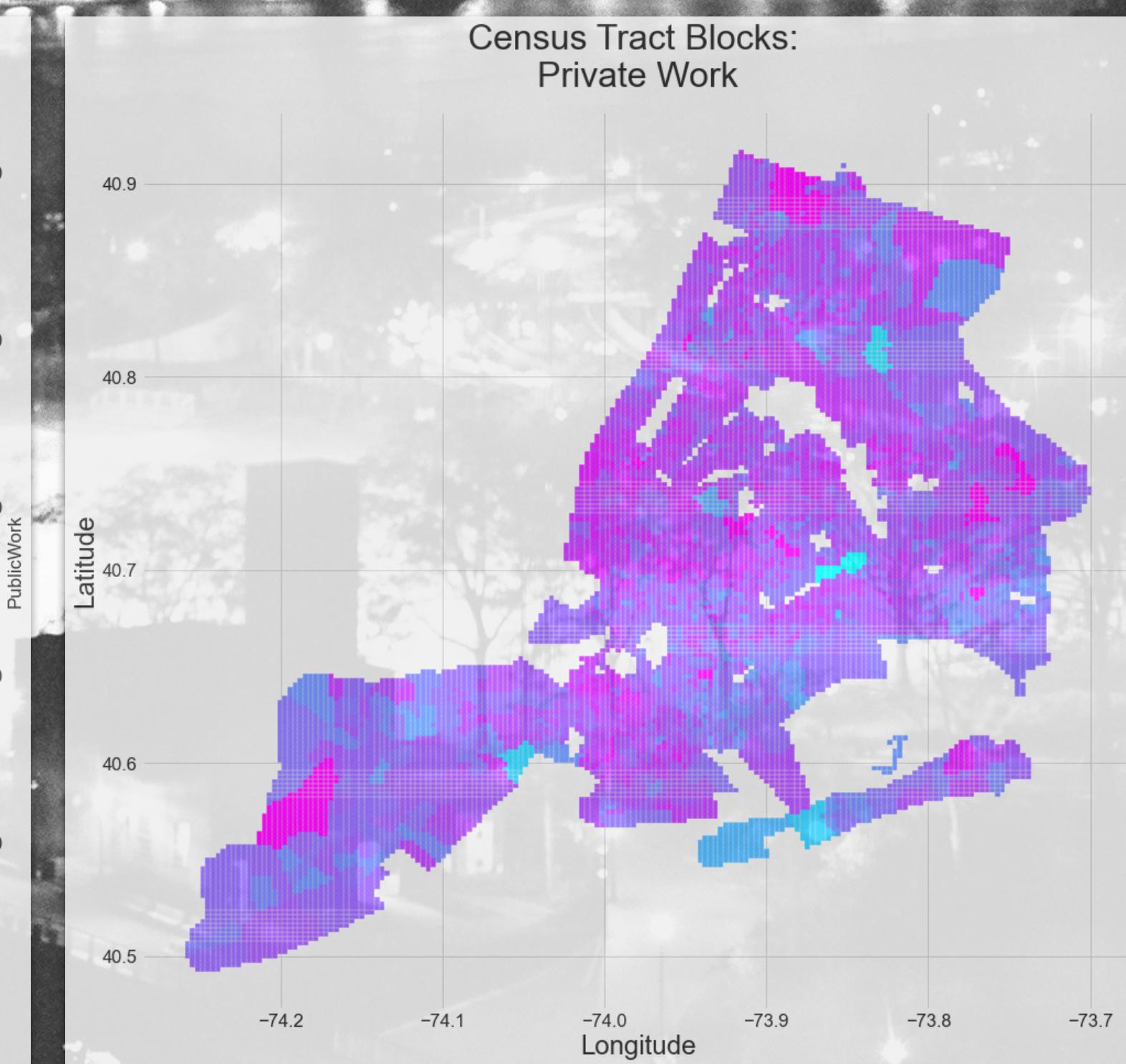
Census Data

Economic factors

Census Tract Blocks:
Public Work

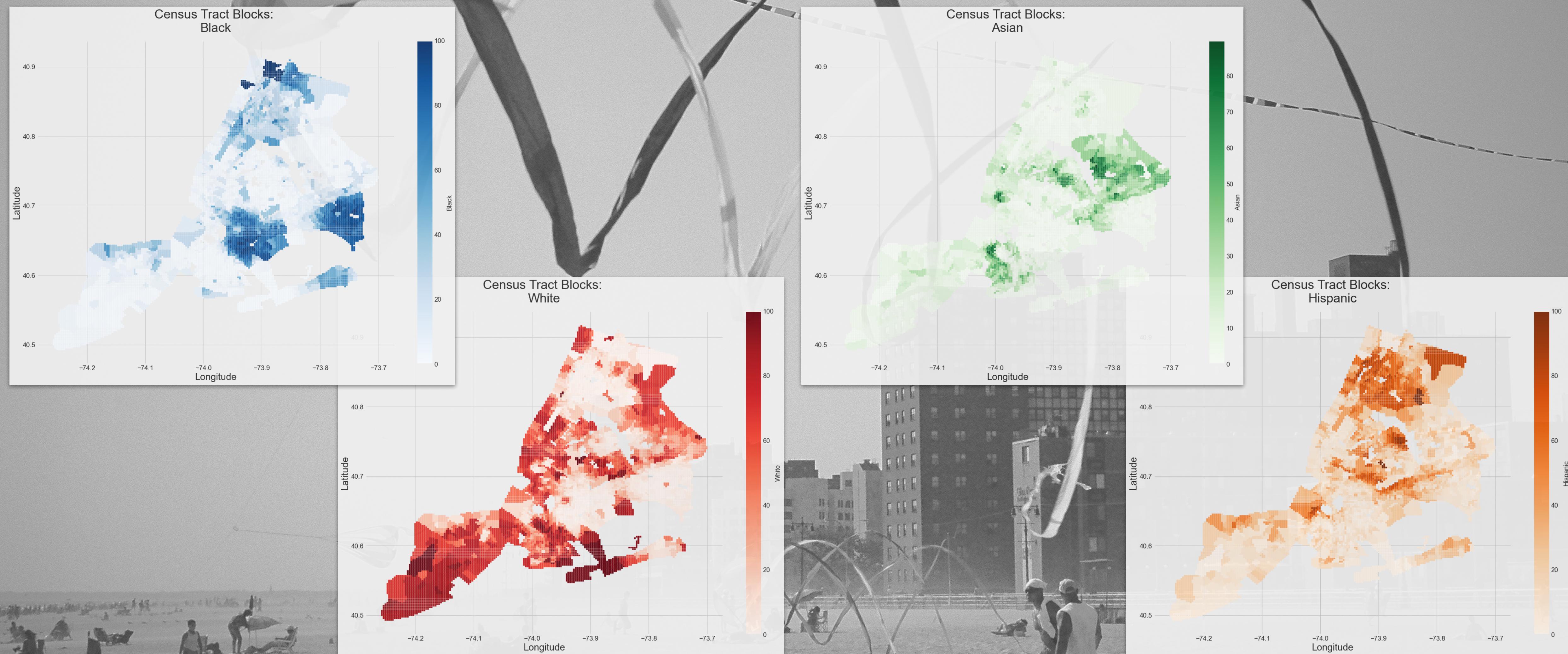


Census Tract Blocks:
Private Work



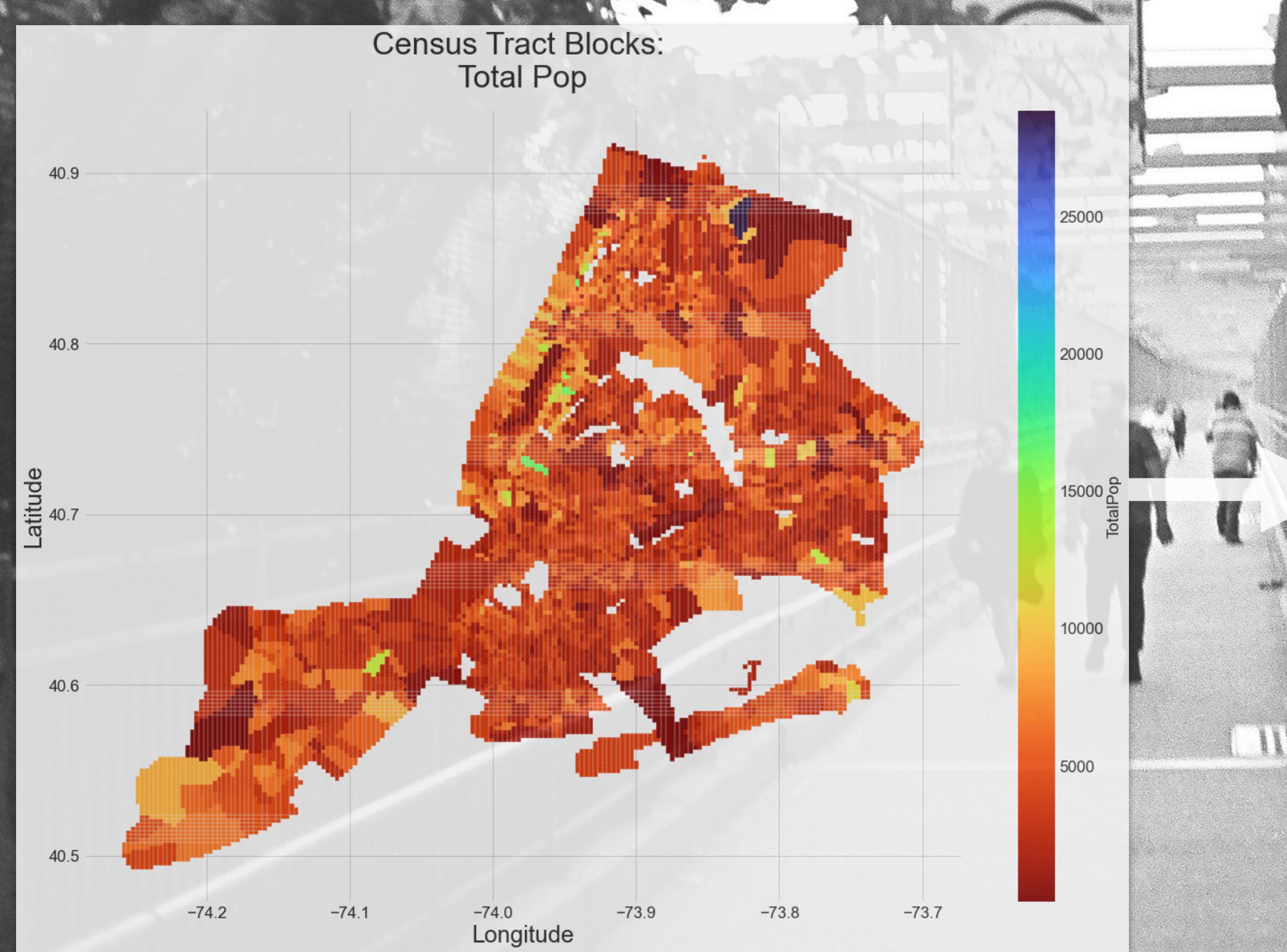
Census Data

Demographics

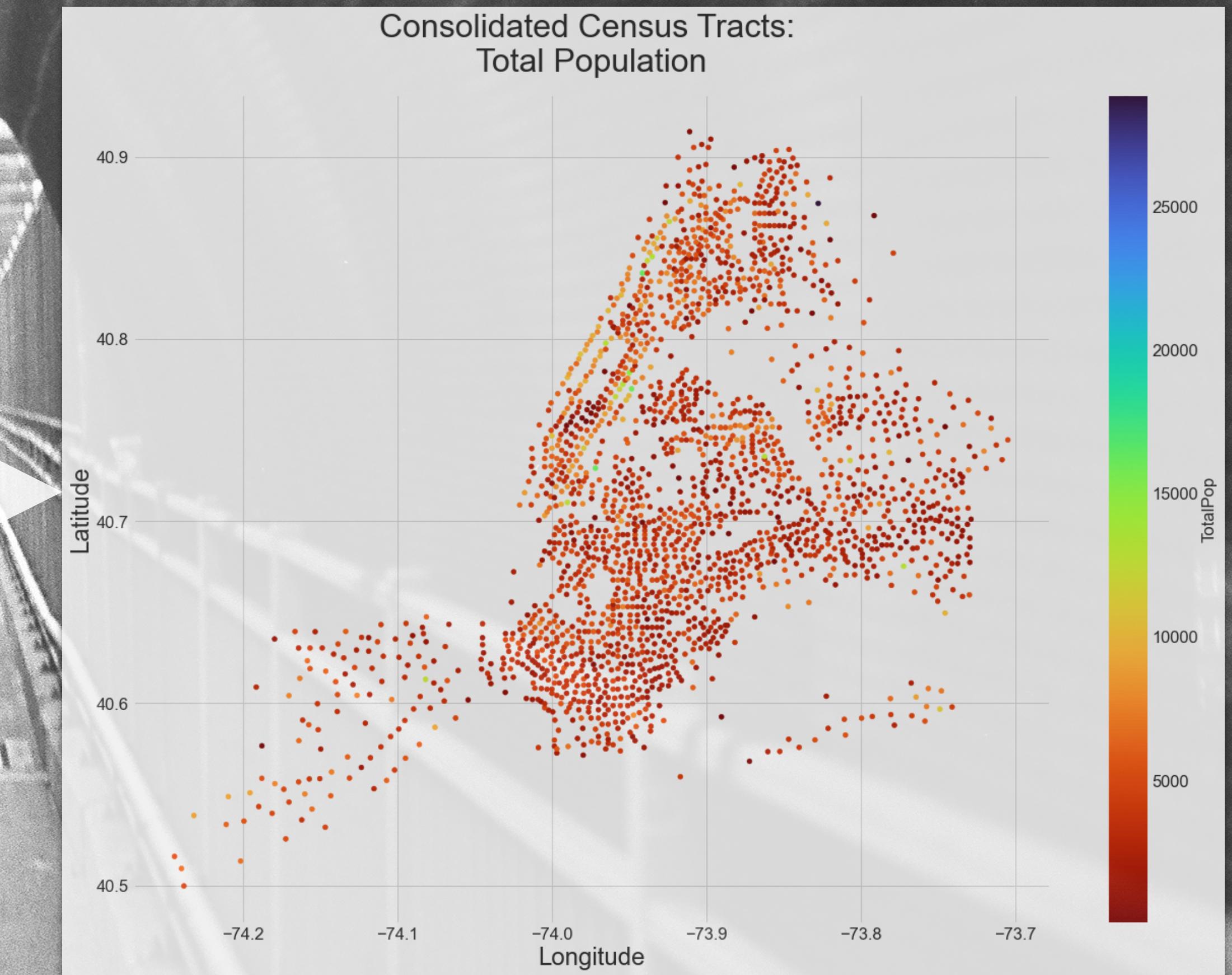


Census Data

Consolidating Census Tracts for computational efficiency



12999 Census Tract Blocks



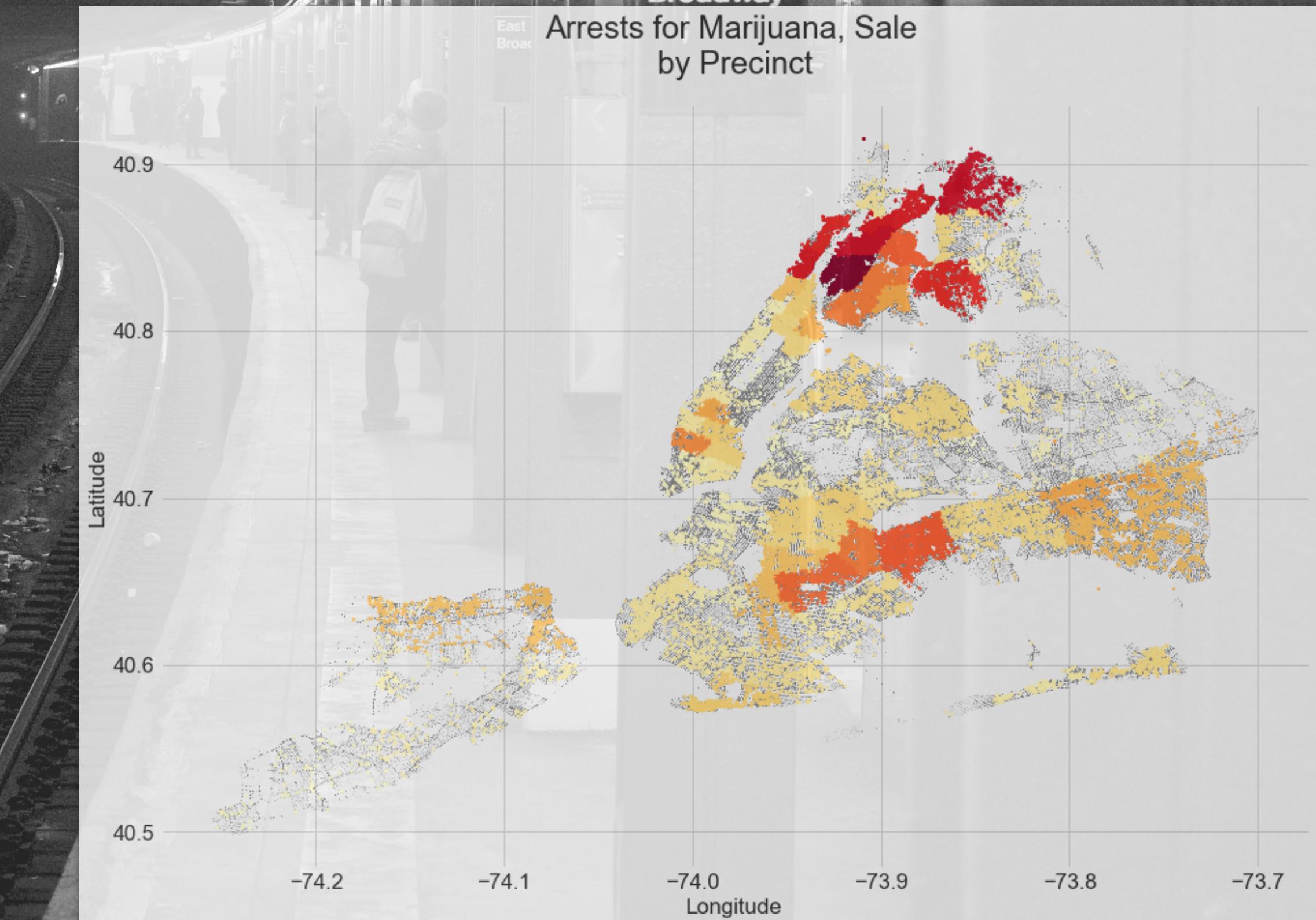
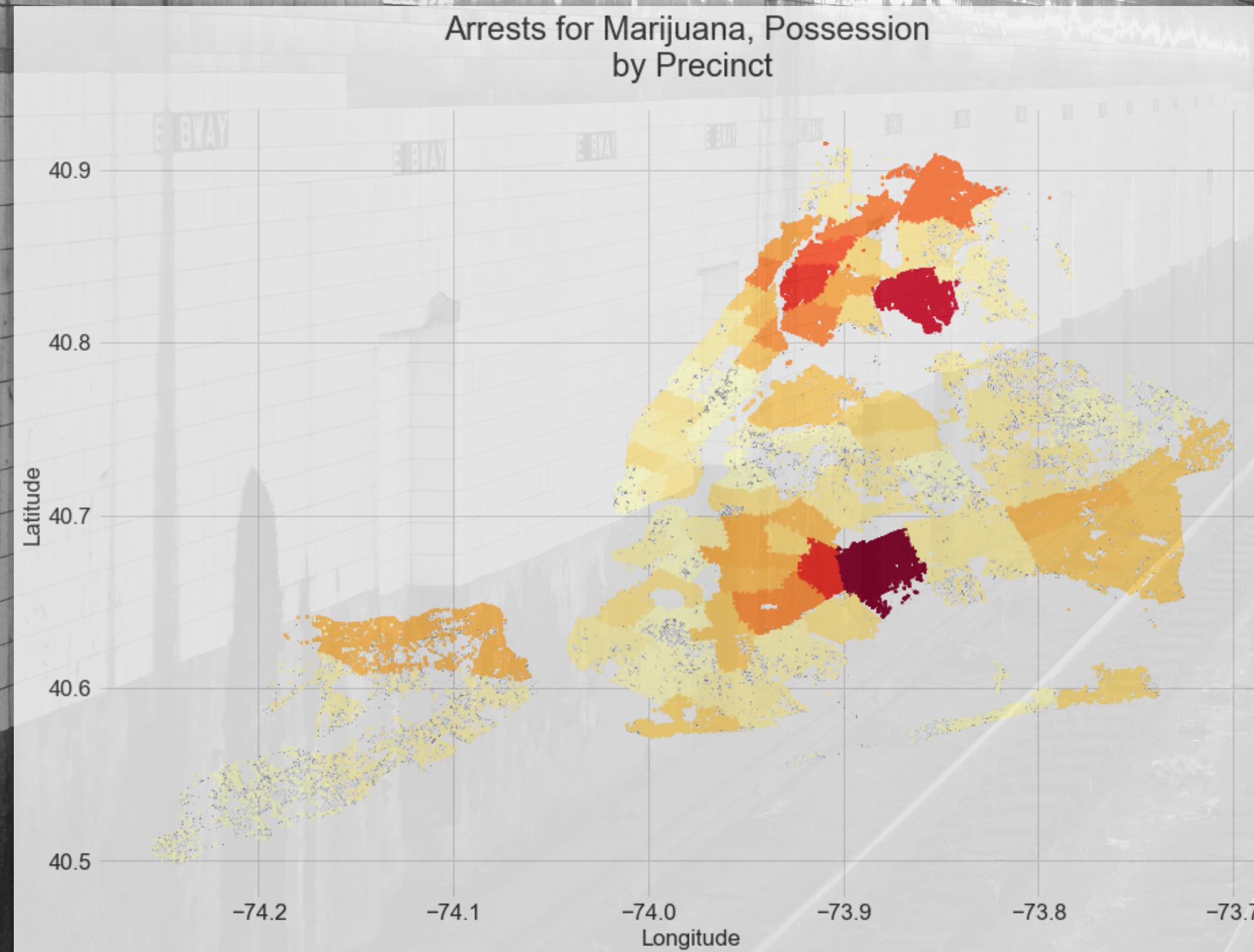
2113 Census Tracts

Arrests Exploratory Data Analysis

The arrests categories considered nonviolent for this analysis

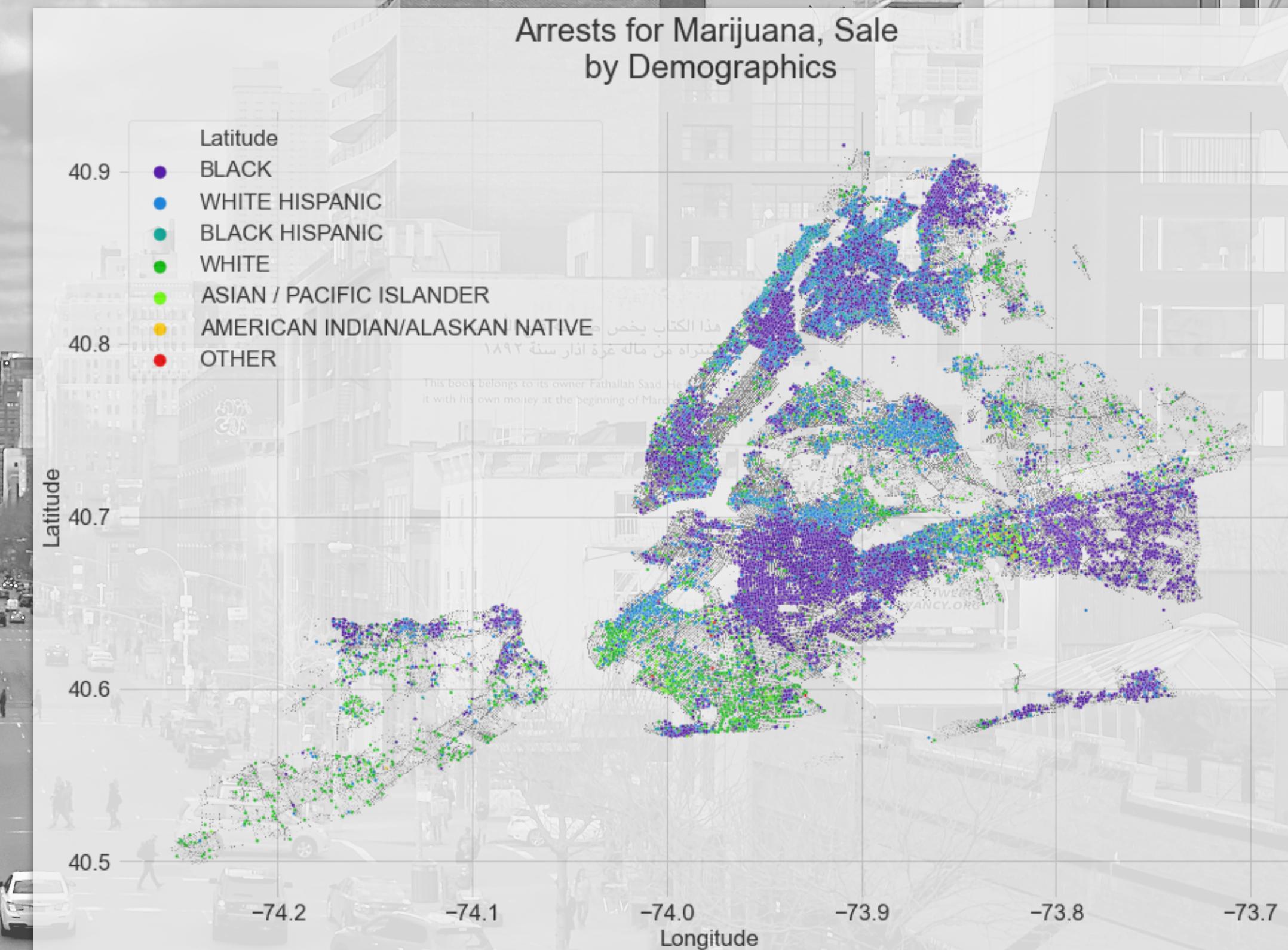
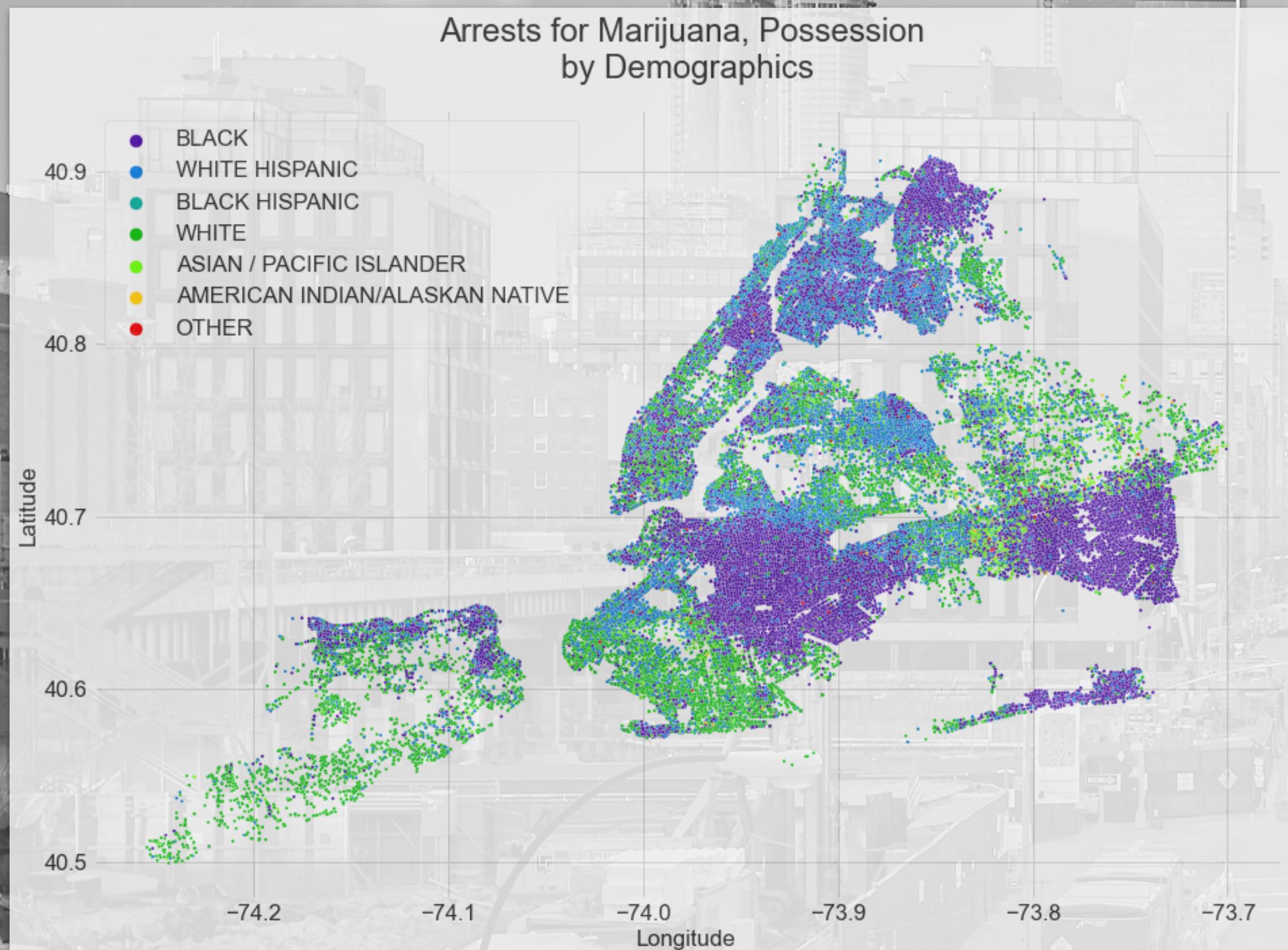
- Marijuana
 - Sale, Possession
- Controlled Substances
 - Sale, Intent to Sell, Possession
- Drug Paraphernalia
- Resisting Arrest
- Misdemeanor Forgery
- Prostitution
 - *special care was taken to not include child or compulsory prostitution which we do consider violent crimes
- Gambling
- Criminal Mischief
- Parks/R&R

Arrests Exploratory Data Analysis



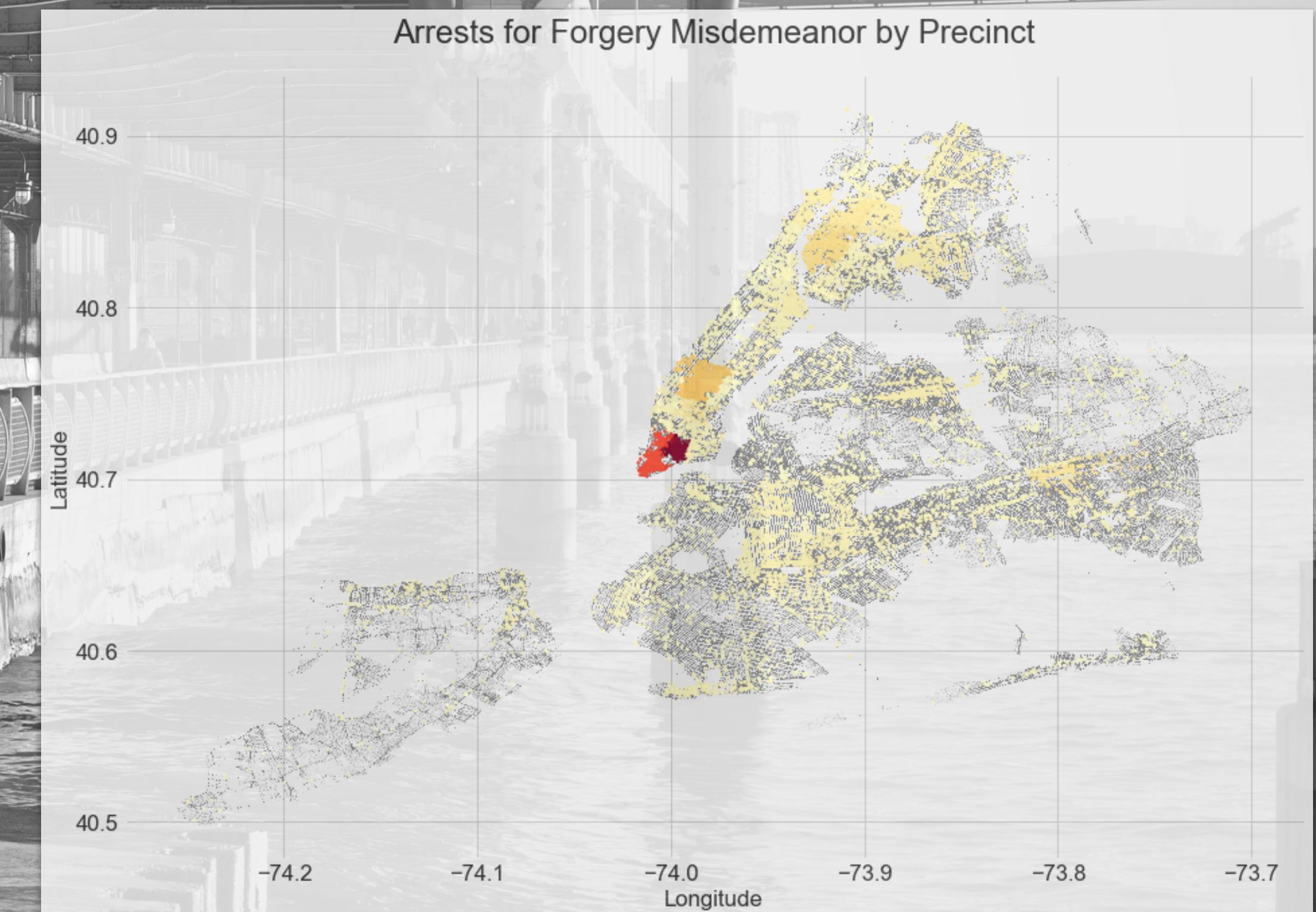
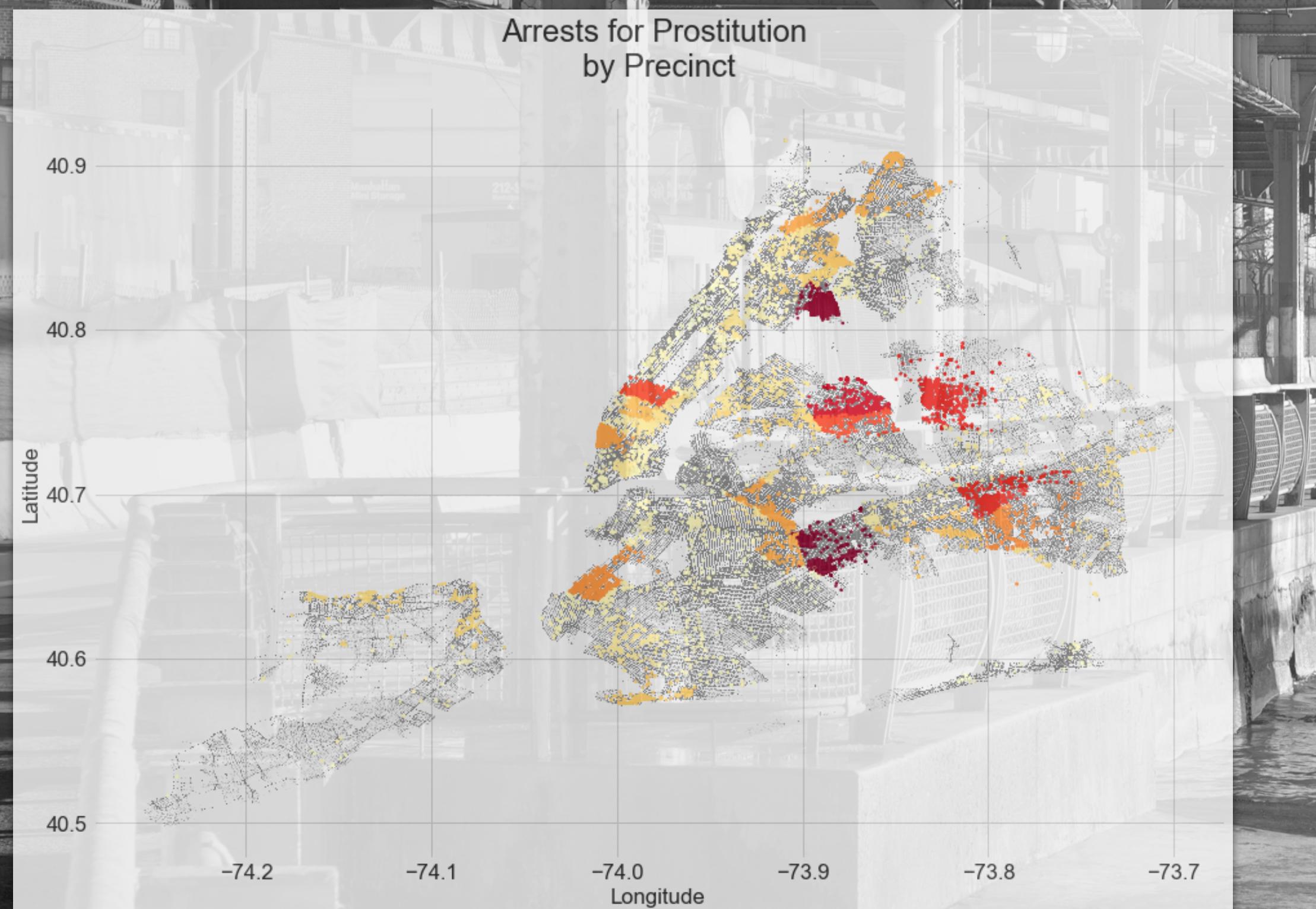
Arrests Exploratory Data Analysis

Most drug arrests are for possession, not sale



Arrests Exploratory Data Analysis

Some arrest categories are highly localized



The Model

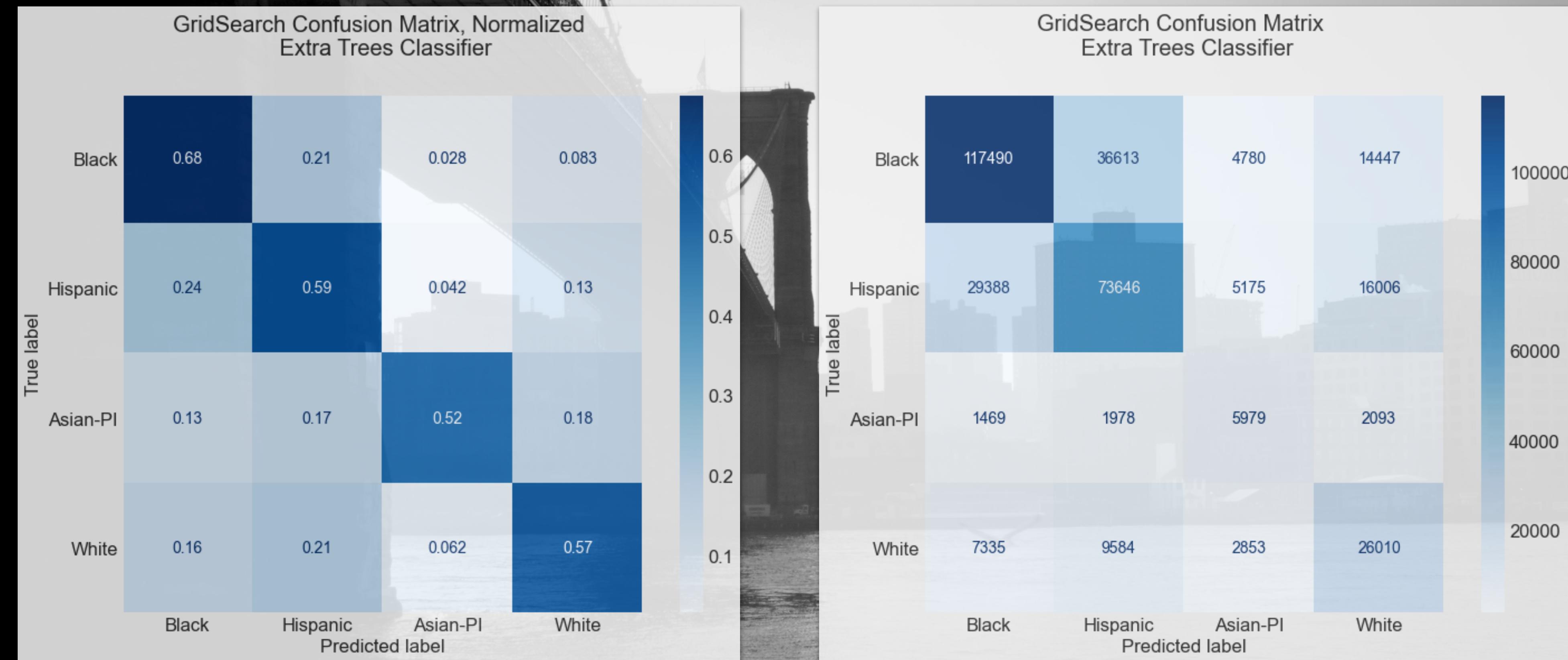


Also considered:

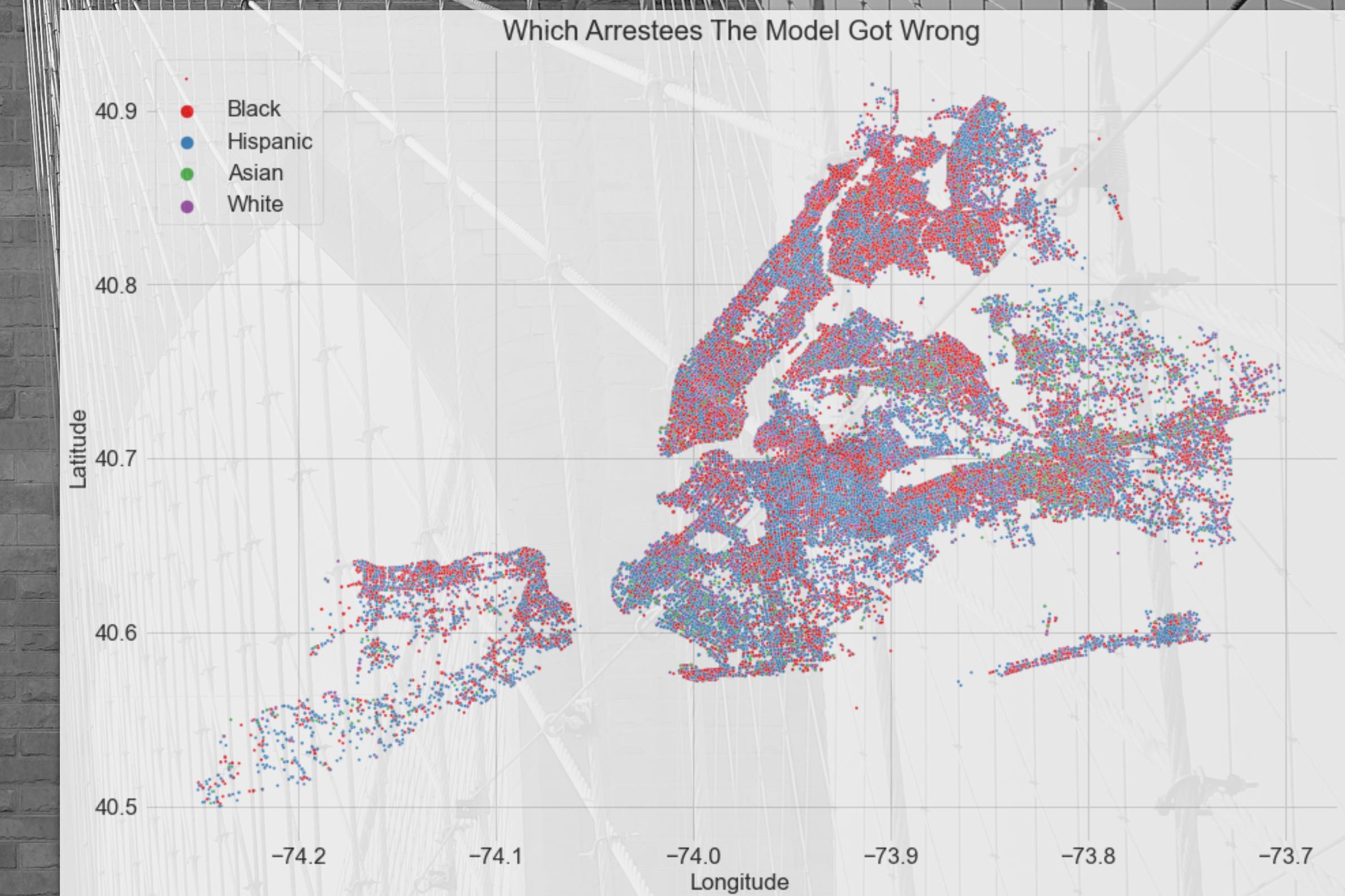
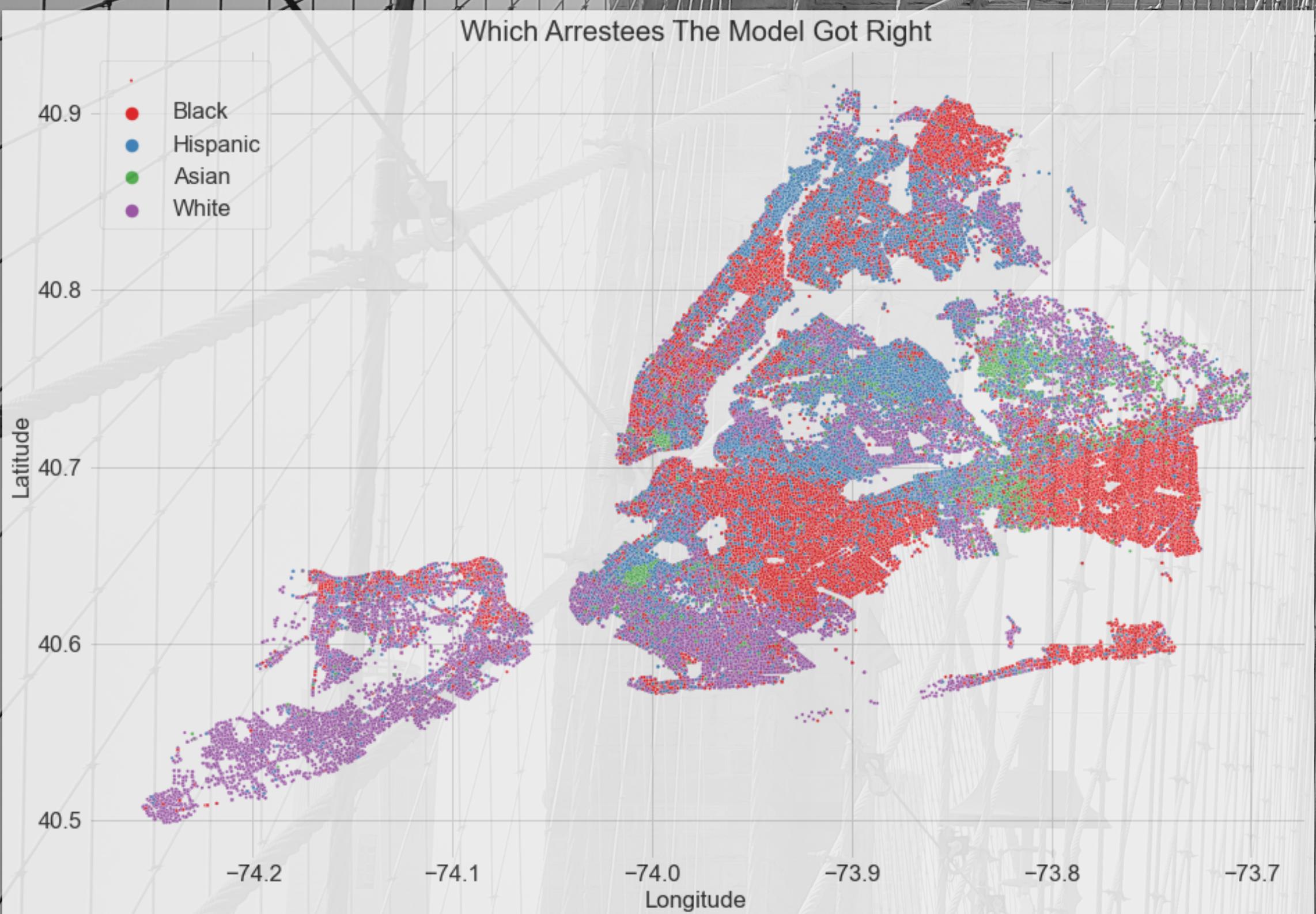
- K-Nearest Neighbors
- Support Vector Classifier

Modeling

Most misclassification was between Black and Hispanic arrestees



Modeling



Conclusions

The NYPD

That the limits on the ability to predict the race of an arrestee mostly falls to conflation between nonviolent Black and Hispanic arrestees illustrates a system that clearly practices greater restraint when it comes to the policing of white people while throwing the book at people of color

Complaints against the NYPD show the same results as the arrests. Black and Hispanic people file complaints against the NYPD much more than white, indigenous or asian people and the complaints filed by people of color are generally based on physical force as opposed to verbal assault

Next steps

Where to go from here?

- Vectorize/multithread the Haversine process for finding nearest Census Tract
- More specific analysis of which precincts have the most disparate outcomes
- Use an unsupervised learning model to determine which precincts have the most disparate outcomes for their locality
- More granular arrest offense classifications
- Use a neural-network model