

Abhiram Kolal

Information Visualization

Professor Veliz

5/10/2020

### Final Project

New York City has been one of the most impactful cities in the world throughout modern history. In a city filled with approximately 8.4 million people, there is bound to be some form of criminal activity within. The City of New York actively releases datasets regarding criminal activity across the city, which I had used to create a proper analysis of New York City crime from 2006-2018.

#### **Data Source**

The data source used in this analysis is provided by the New York City Police Department, which lists every arrest in NYC spanning from 2006 to 2018. According to the footnotes of this dataset, the information is a close approximation of current records, due to the ever-changing nature of arrest revisions and updates. People that have been arrested with multiple charges are listed as a single record, with the top charge being included. Most arrests in this dataset are classified through specific keys that are classified by the NYPD as identifiers, such as the arrest key, which gives continuous numerical data as to a single arrest instance. Other codes given in this dataset are categorical, such as the law code, which refers to the specific law charge corresponding to New York State Penal Law, and arrest precinct, which refer to the specific precinct that arrested the perpetrator. The data also gathers nominal categorical demographic information which includes the perpetrators' age group, gender, and race. To help

map out the geographic location of each arrest, the dataset also provides a coordinate location for each arrest instance.

## **Challenges**

Using this dataset, I wanted to analyze different aspects of the overall nature of crime in New York City. Throughout the given years of 2006 to 2018, I decided to focus on three main aspects: types of crimes committed, locations of criminal activity, and individual perpetrator demographics. Using Tableau, I was able to map out every instance of criminal activity within the city, as shown in Visualization 1. I had separated the data by borough, which proved to be challenging as the codes for each borough were based on NYPD notation, which I had to familiarize myself with. I was able to figure out that B = Bronx, S = Staten Island, K = Brooklyn, M = Manhattan, and Q = Queens by categorizing the data by borough code and matching it by location using the Map function in Tableau. By understanding this notation, I was also able to extract the total number of arrests by borough, which can be seen in Visualization 3. In Visualization 1, I had also provided information as to what type of arrest had occurred, which a full list was provided in Visualization 2. By having access to both the total number of arrests and the approximate date and time of arrest, I was able to visualize arrests over time, as shown in Visualization 4. Throughout Visualization 5-7, I had included information based on perpetrator demographics. I was able to utilize the information given about perpetrator race to create a tree map in Visualization 5, perpetrator age groups to create a pie chart in Visualization 6, and gender distribution in Visualization 7 using packed bubbles. I found that there were many challenges in creating these visualizations, mainly due to cleaning up the data found from the site to create an accurate depiction of each topic. Many of the categorizations had null values which I had to

manually filter out, and some of the offense descriptions had duplicate values, which I had to combine using Excel.

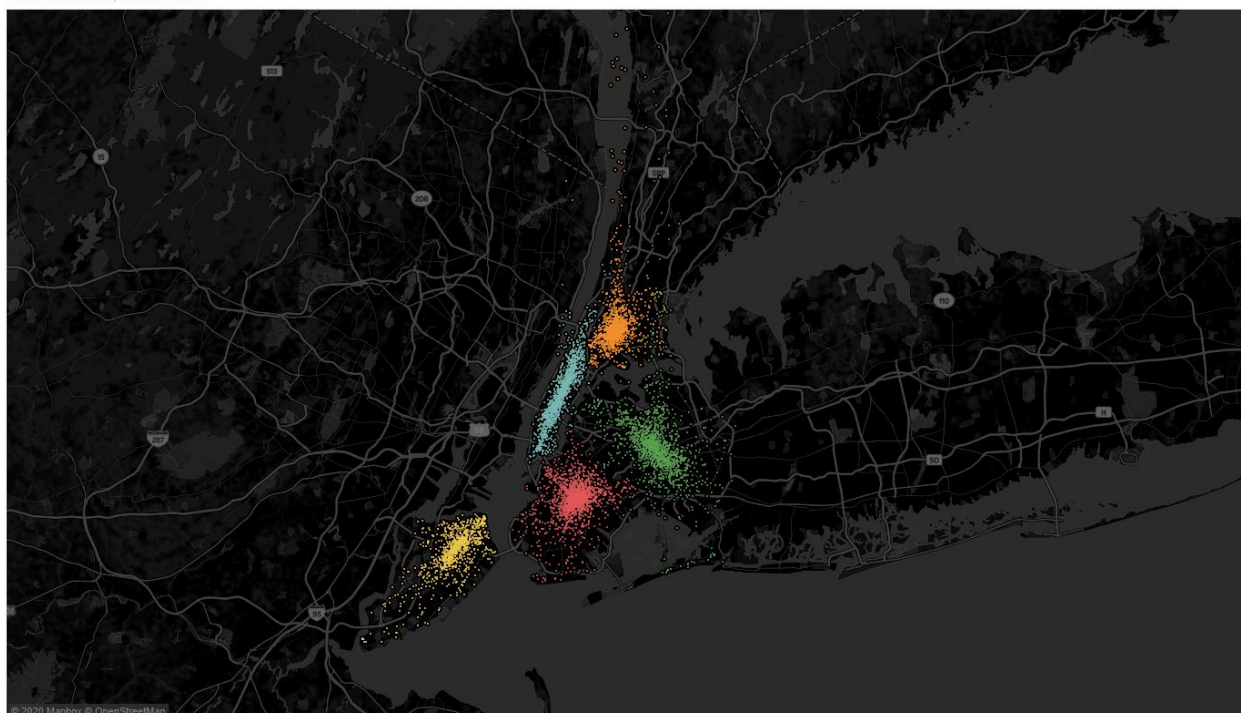
### **Key Findings**

Some key findings that I had derived from this information is that the overall crime rate in New York City has exponentially decreased in the given timeframe of this dataset. As shown in Visualization 4, crime rate has reduced by almost 100,000 cases overall per year since 2006 to 2018 in all five boroughs, which is a huge improvement. The most arrests occurred in Brooklyn, and the least arrests had occurred in Staten Island overall. Drug related offenses were the most common arrest, with almost 500,000 more cases than assault, the second highest offense. The things I found most interesting about this analysis was looking at demographic information, where I found that the most common perpetrator would be an African American male in the age group of 25-44. While it is true that these classifications were the majority in all arrests according to the dataset, this data is insufficient to confirm a prediction on who could be classified as a criminal overall.

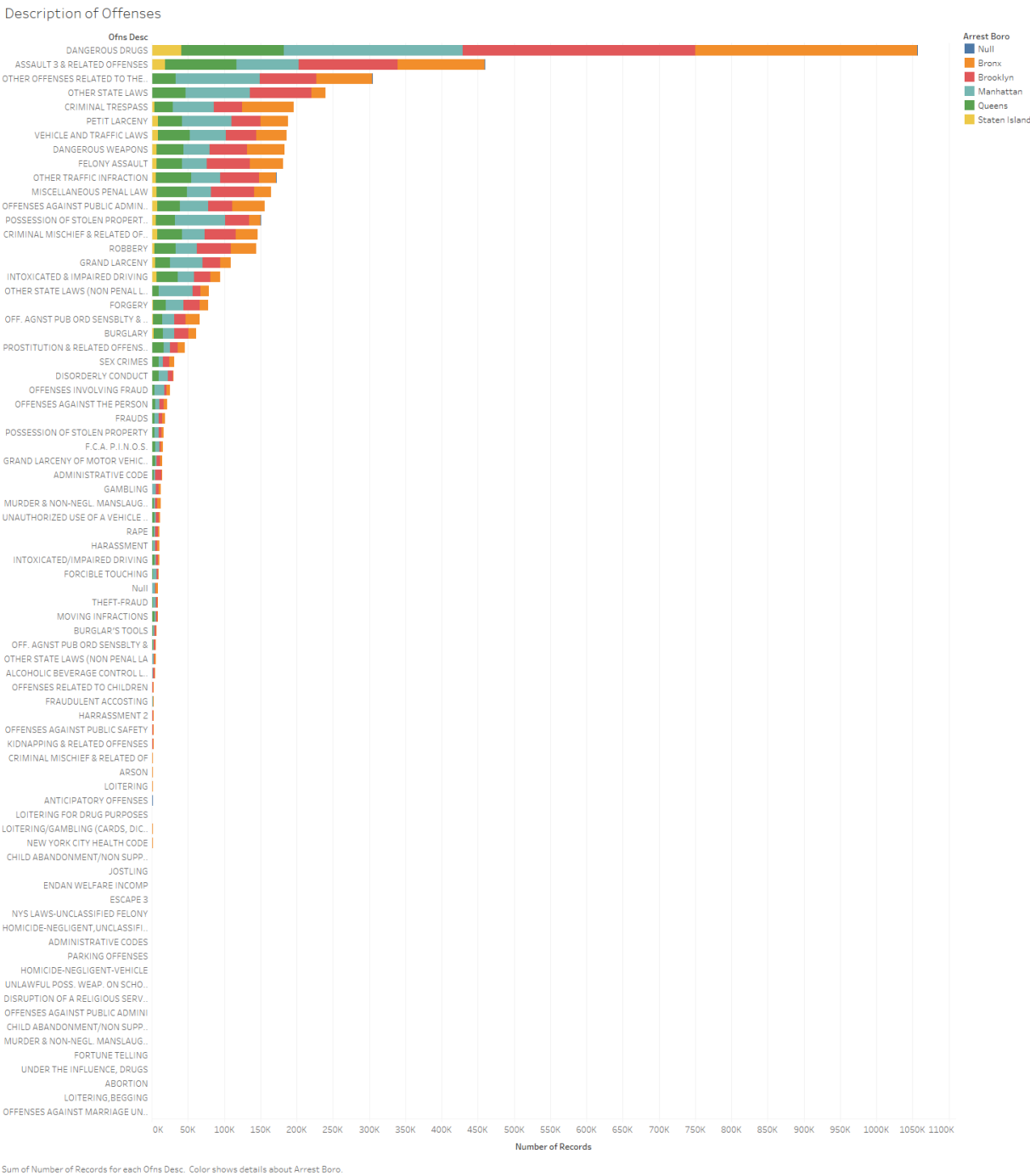
Appendix:

Visualization 1:

NYC Arrests, 2006-2018

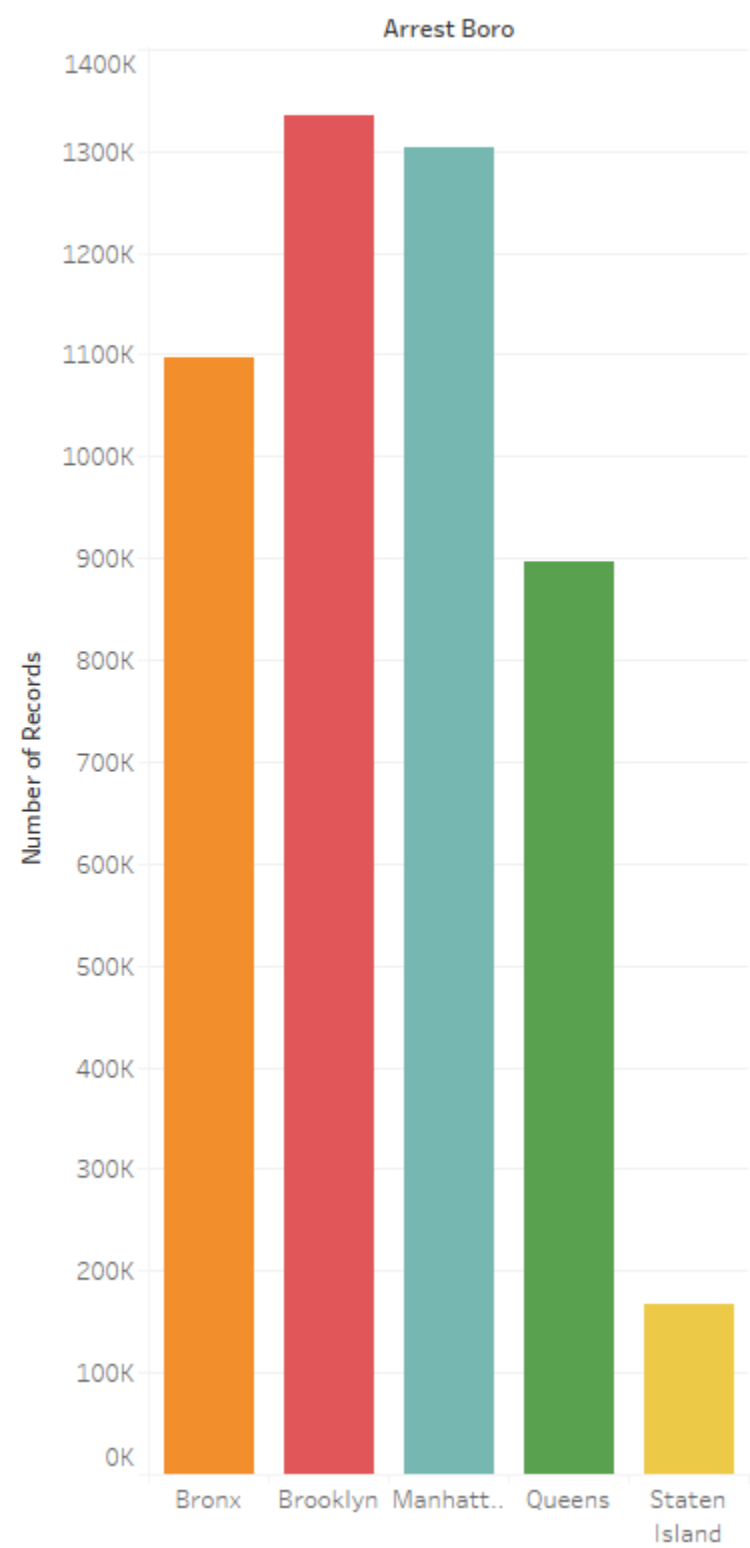


Visualization 2:



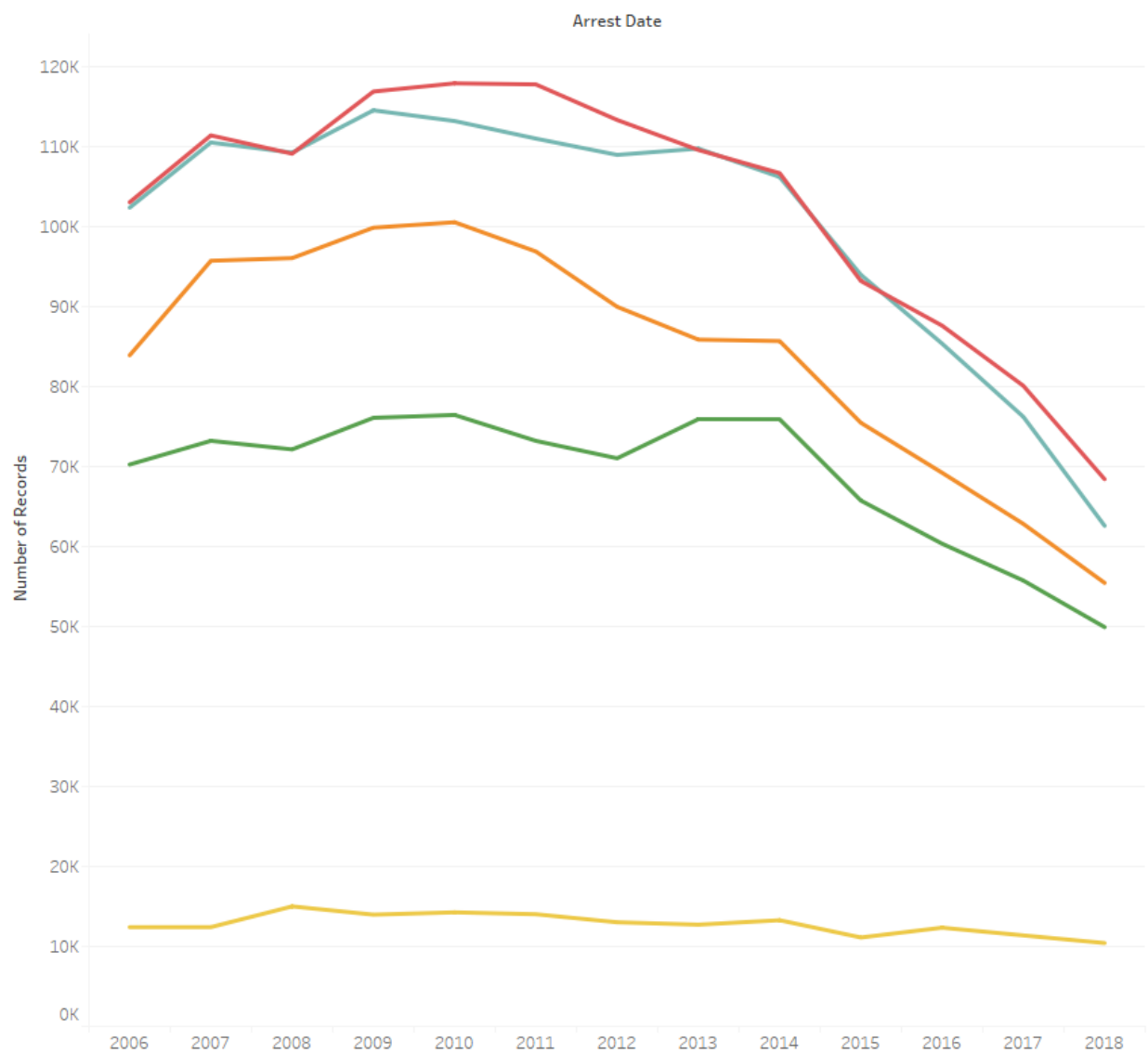
Visualization 3:

### Total Arrests By Borough



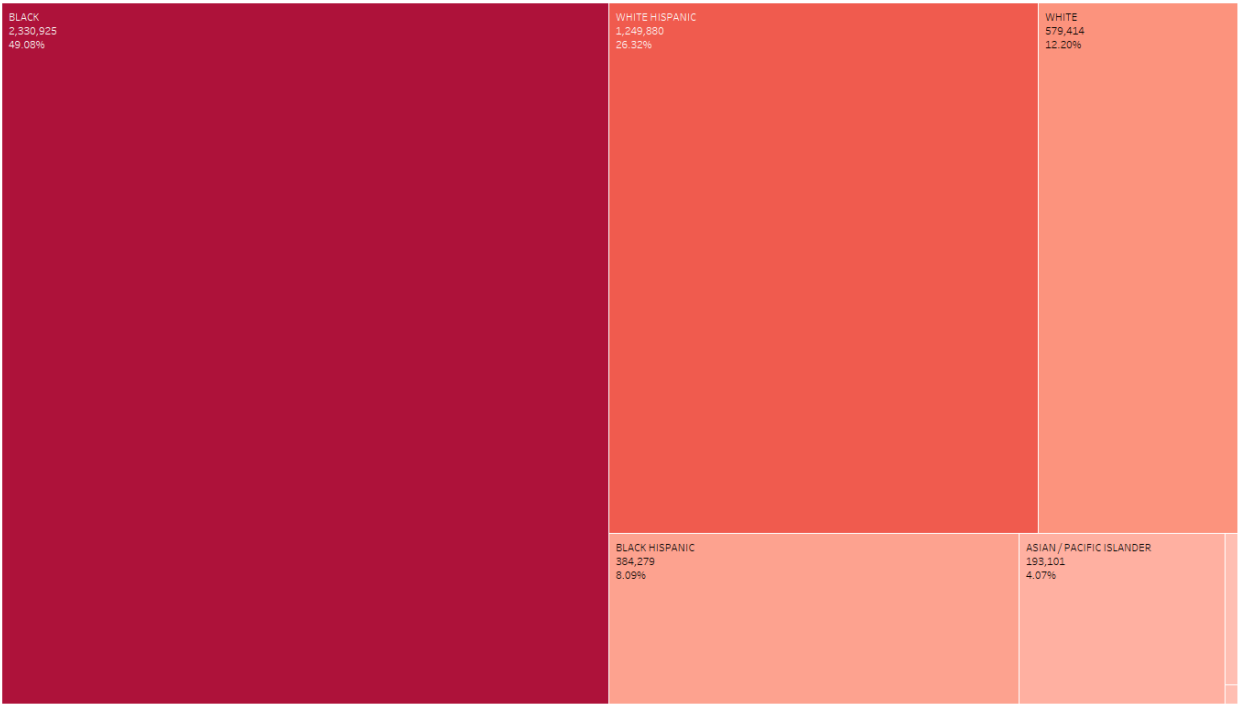
Visualization 4:

Total Arrests Per Year



Visualization 5:

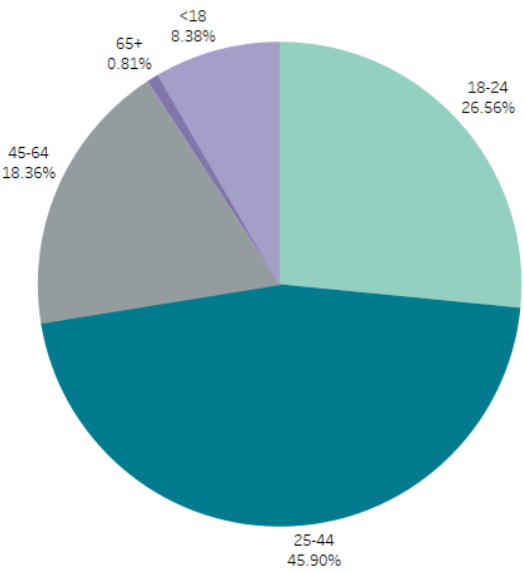
Race of Perpetrators





Visualization 6:

Age Group



Visualization 7:

Gender Distribution

