

Stat 240 Final Project, Group 2: Analysis of Crime in New York City

By: Samuel Pekofsky, Kevin Park, Hazel Mittal, and Joe Doyle

Introduction

In the bustling metropolis of New York City, crime and its demographics are constantly under scrutiny. New York City is the largest and most influential American metropolis in the country. Crime rates are a complex phenomenon that vary significantly based on geographic location, demographic factors such as age, and the severity of the offenses committed. Understanding the intricacies of criminal behavior across New York's boroughs is paramount for effective law enforcement strategies and social policy formulation. Consequently, we felt that the social significance of this topic made it worth investigating. In addition to affecting public safety, crime is often a reflection of larger socio-economic issues and urban dynamics.

In this project we aim to answer the questions: Is there a difference in the proportions of crime types and the age groups that commit those crimes in the 5 boroughs of New York city? We specifically want to answer the question: How do crime rates and proportions of felony offenses differ between boroughs, if at all? With this in mind, we chose to analyze how arrests for varying crimes differ across the boroughs, especially with respect to age. Towards the end of our analysis, we put a special focus on the age group of 18-24 year olds and felonies simultaneously. We believe that this age group can be very receptive to rehabilitation while also being young enough that if they are not rehabilitated, incidences of recidivism may occur many more times within their lifetime. While the same points can be made about the <18 age group, people under the age of 18 are not legally adults and are subject to different rules, so we felt it would be less useful to focus on this age group. Also, we believe that minors are, on average, far more likely to experience confounding variables such as their home lives and the knowledge that they may be treated more lightly under the law. Additionally, we chose to emphasize felonies because they are commonly perceived as the most threatening categorization of crime to society.

In order to get a sense of varying arrests and crime throughout the boroughs, we make use of an arrest data set from the NYPD and population data from the US Census Bureau. Through analyzing this data, we found wide variances in arrests per borough, even when accounting for population differences.

Background

The data that will be used is all contained within the following csv file: "NYPD_Arrest_Data_Year_to_Date.csv". This file was provided by the City of New York¹ This data was collected by the NYPD by recording arrests as they occurred. It has been manually extracted and publicly released on a quarterly basis. The file we will be using was created on November 10, 2020 and last updated on January 19, 2024. This data is entirely from the year of 2023 so moving through this analysis all the data we will be using will come from the same year only, 2023. As for the data, each instance, or row, represents an arrest in NYC performed by the NYPD.

- We will be using three key variables:
 - **AGE_GROUP**: this is the age group that an arrested individual falls into. Groupings are <18, 18-24, 25-44, 45-64, and 65+.

¹<https://catalog.data.gov/dataset/nypd-arrest-data-year-to-date>

- **ARREST_BORO**: This is the borough in which the arrest took place. This column has with single letters representing the boroughs as follows: **B**: The Bronx, **S**: Staten Island, **K**: Brooklyn, **M**: Manhattan, **Q**: Queens.
- **LAW_CAT_CD**: This is the level of offense, which is a severity ranking of the crime(s) committed, ranging from **V**: violation (least severe), **M**: misdemeanor, and **F**: felony (most severe).

Data outside the data set was also used to find the populations of the 5 boroughs. This data was taken from the City Population Website.² The population of the boroughs of New York City were found according to census results. It should be noted that the data only records the borough of the arrest, not the borough where the crime took place, potentially affecting data interpretation in cases where the arrest occurs in a different borough. With this being said, the size of the boroughs, and the physical barriers that separate them call into question the likelihood that there is a mismatch between the borough of the crime and the borough of the arrest. Still, this is worth considering as a potential source of noise.

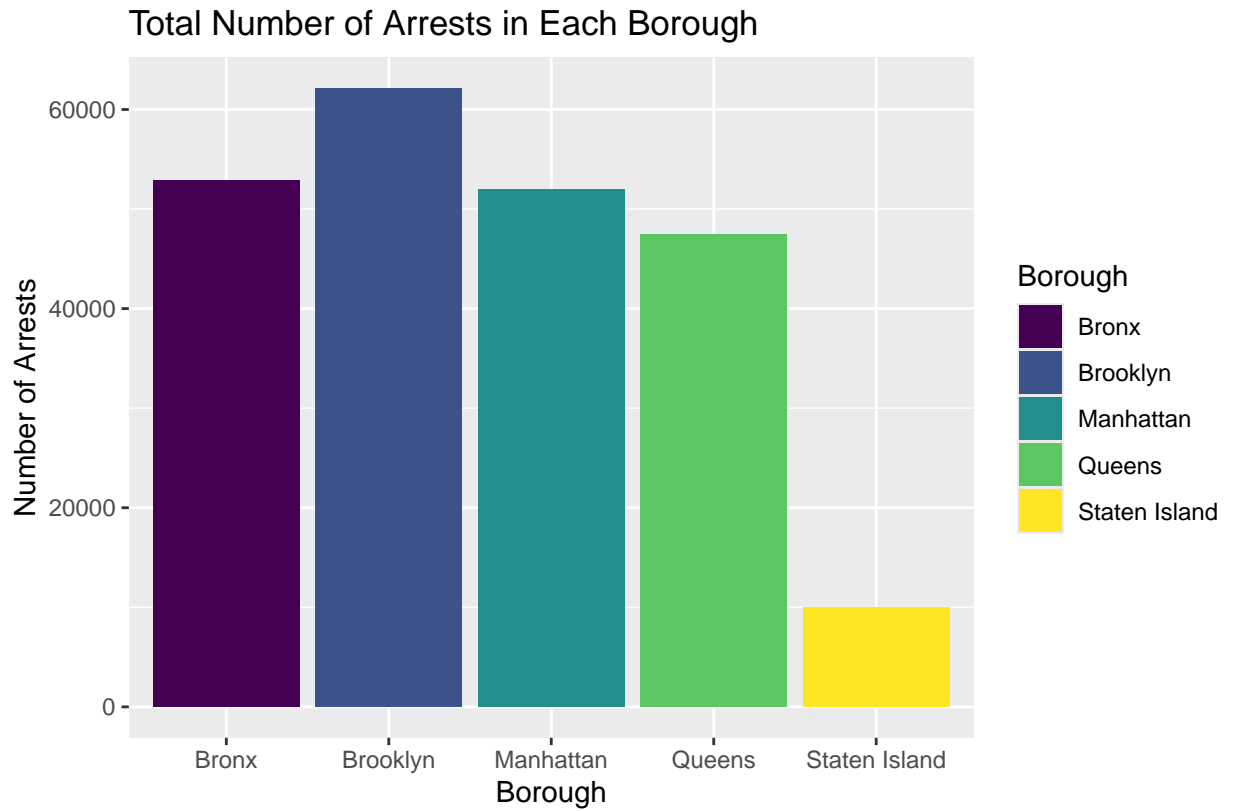
In the rest of this report, we will first clean up and manipulate the data to remove instances with missing relevant data, and then we will remove unneeded fields. With this accomplished, we will then manipulate the data by creating different summaries to understand the distribution of arrests across age groups, severity levels, and boroughs. Following this, we will calculate the percentage of crimes by severity and age group, normalized by borough population, to allow for comparison between boroughs of different sizes. We will also perform t-tests comparing Manhattan to the other four boroughs. By calculating the p-values associated with these tests we will assess the probability of observing the observed differences if the null hypothesis were true, i.e. if there were no true differences between the boroughs. We will look at the confidence intervals so that we can quantify the uncertainty associated with our estimates of the true differences in crime percentages between boroughs.

For References, see bottom of page or click on the numbers leading to footnotes.

Analysis

In this section, you will find graphs with detailed **descriptions** below them.

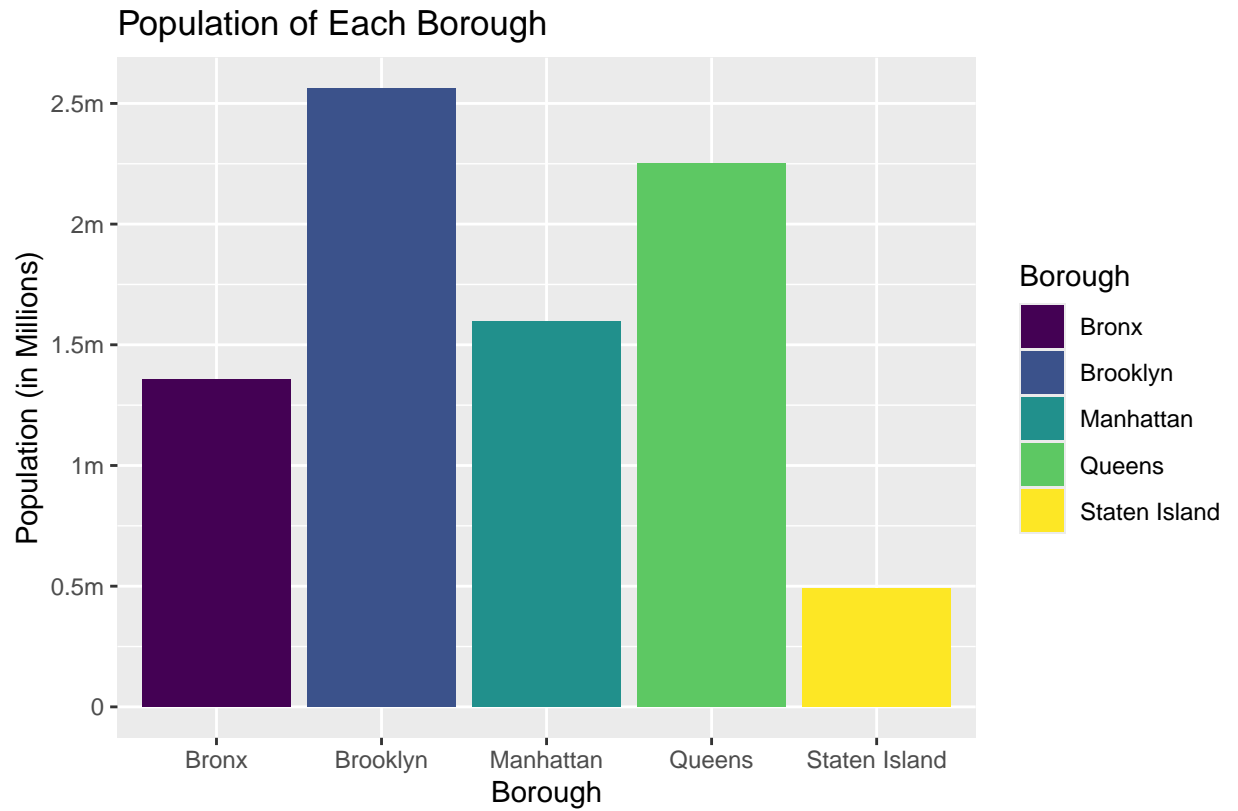
²<https://www.citypopulation.de/en/usa/newyorkcity/>



Data from the NYPD

Graph 1:

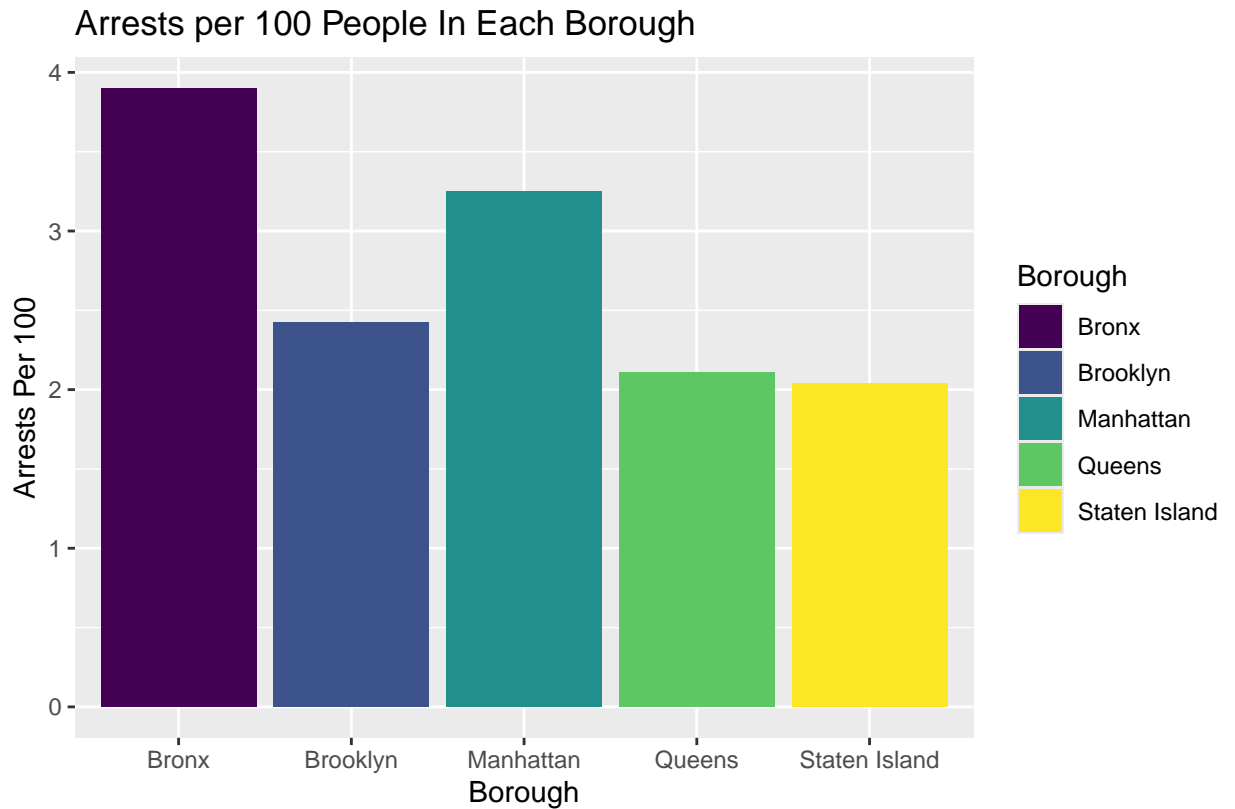
Graph 1, pictured above, displays the total number of arrests for each borough, which is helpful for contextualizing later graphs and analysis. An important thing to consider is that while arrests in each borough vary widely, so does the population, layout, and wealth distribution.



Data from the US Census Bureau

Graph 2:

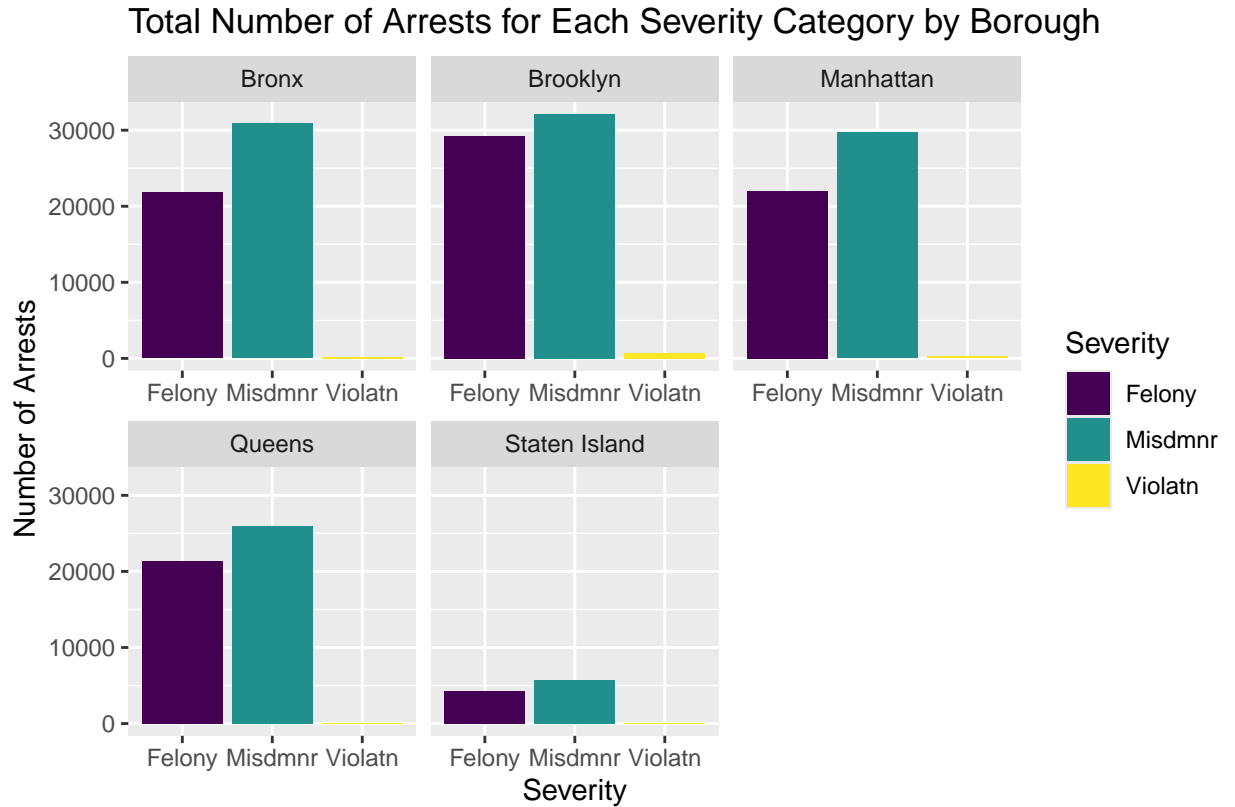
Graph 2, pictured above, displays the total population of each borough, which helps to contextualize the first graph, since it gives an idea of how the number of arrests compares to the overall population of the borough.



Data from the NYPD and the US Census Bureau

Graph 3

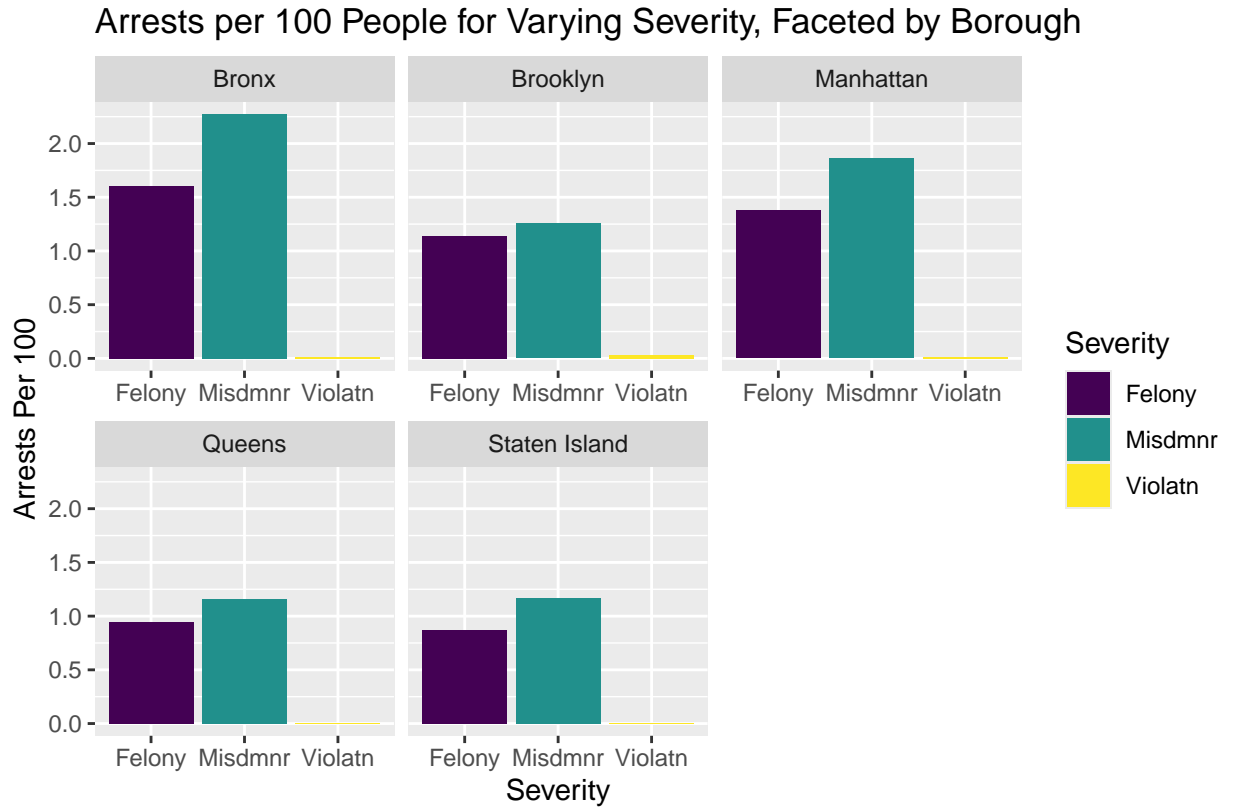
Graph 3, pictured above, combines the data included in the previous graphs, graph 1 and graph 2. Graph 3 shows the number of arrests that occur in each borough per 100 people who live there. For example, the Bronx has a ratio of nearly 4 arrests per 100 people living there, giving it the highest arrest to population ratio. On the other hand it can also be seen that Staten Island has the lowest arrest rate compared to the other boroughs.



Data from the NYPD

Graph 4

Graph 4, pictured above, provides additional insight into the total number of arrests in each borough, breaking crime into three categories ordered from most to least severe: felonies, misdemeanors, and violations. Misdemeanors and Violations are abbreviated as “Misdmnr” and “Violatn”, respectively. These graphs provide insight into the makeup of arrests within each borough. Felonies are the most severe categorization of crime, with prison sentences generally ranging from 1 year to life, if found guilty. Misdemeanors are considered far less serious and often punished with 15 days to 1 year in jail. A violation is considered any non-criminal offense, excluding traffic infractions. As can be observed in graph 4, violation arrests are extremely uncommon and can be punished by a maximum of 15 days in jail and/or a \$250 fine. As it can be seen Brooklyn has the highest amount in number of arrests as it has the most misdemeanor, felony and violation arrests while comparatively Staten Island has the least.

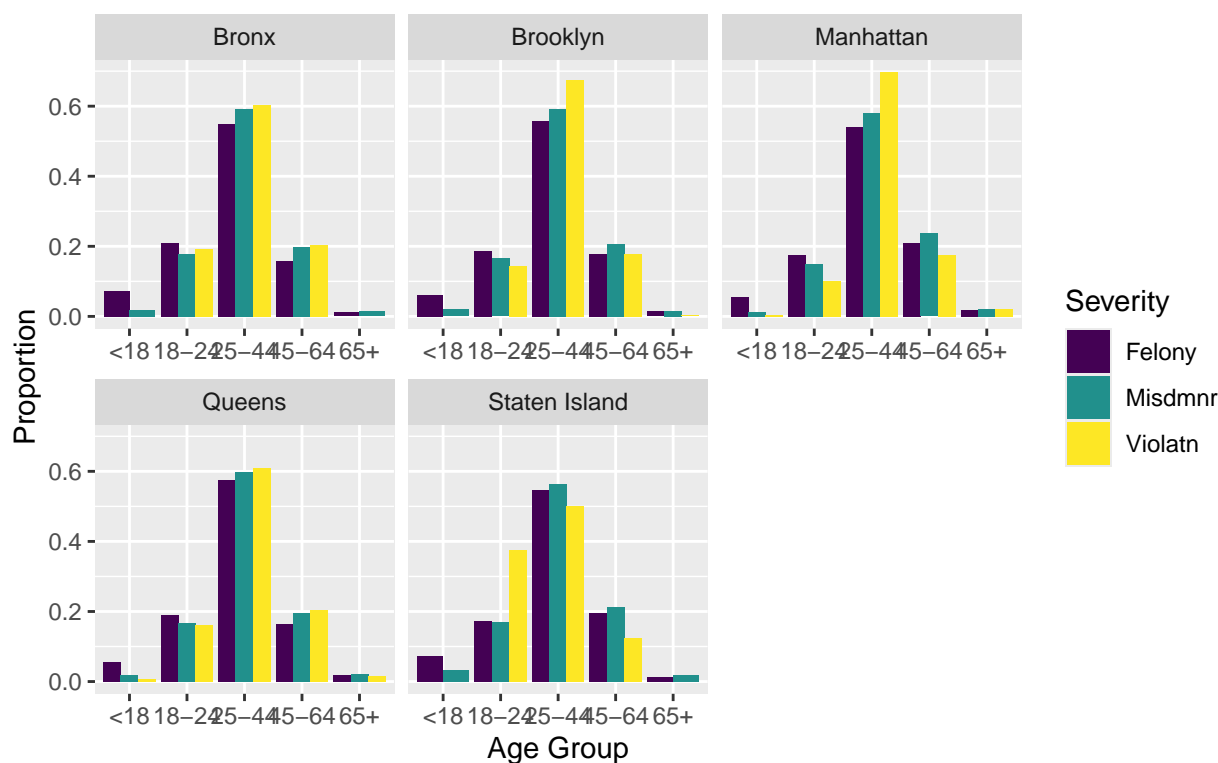


Data from the NYPD and the US Census Bureau

Graph 5

Graph 5, pictured above, combines the data displayed in graph 4 and graph 2, much like the way graph 3 combines data from graph 1 and graph 2. This graph gives insight into the rate at which arrests for each severity categorization are made relative to the total population of each borough. Please note that this graph uses the same abbreviations as graph 4. From this graph we can see that the Bronx has the highest rate of felony and misdemeanor arrests relative to its population, while Brooklyn has the highest rate of violation arrests. This graph contextualizes things better as even though from Graph 4 we saw that Brooklyn had the most amount of arrests in each severity level, in this graph we can see when taking the borough population into consideration the arrest rate in the Bronx is quite worse with Manhattan coming right behind it. This graph also shows while the arrest numbers in Staten Island looked the least in Graph 4, when considering population its arrest rate is very similar to that of Queens.

Proportions of Severity of Crime Committed by Age Group In Each Borough

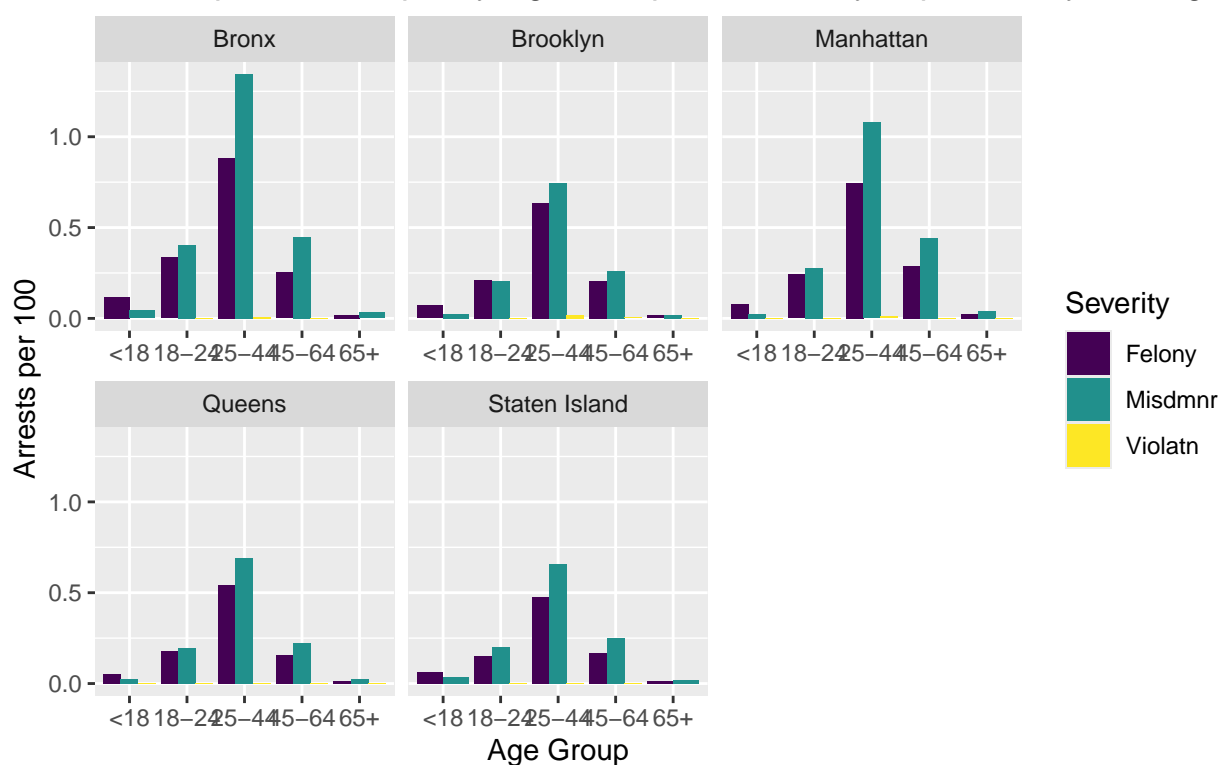


Graph 6

Data from the NYPD

Graph 6, pictured above, displays the proportions of arrests for a given severity level of crime for each age group, separated by boroughs. For example, the Brooklyn graph shows that roughly 60% of misdemeanors are committed by people who are between 25 years old and 44 years old, inclusive. Additionally, in Queens, 18 to 24 year olds make up nearly 20% of felony arrests.

Arrests per 100 People by Age Group and Severity, Separated by Borough



Data from the NYPD and the US Census Bureau

Graph 7

Graph 7, pictured above, shows us the rate at which each age group is arrested for each crime severity relative to the entire population of the borough. For example, we can see that in Brooklyn, for every 100 people living there, there were roughly 0.75 felony arrests on individuals who were 25-44 years old. From this graph we can infer many things. Notably, there is an exception in the “<18” age group, where felony arrests are more frequent than misdemeanor arrests, suggesting a concerning trend warranting further investigation. The age group between 25-44 emerges as the most active cohort in criminal activities across all boroughs. Interestingly, misdemeanors prevail as the predominant type of arrest within this age bracket, indicating a consistent pattern of lesser severity offenses. Misdemeanors appear to be the most common type of arrest across all age groups, except for those below 18 and a singular instance in Brooklyn, where the rate of felony arrests slightly surpass misdemeanor arrests in the 18-24 age group. The rate of violation arrests remains extremely low across all age groups and all boroughs, suggesting minimal prevalence in comparison to other categories. An overarching observation is the perception of the Bronx as the least safe neighborhood in New York City, evidenced by higher crime rates across various age groups and severity levels.

Next, the Welch Two Sample t-test was conducted to compare the proportions of felony arrests for the 18-24 age group between Manhattan (often seen as New York City’s center, or New York City itself) and the other four boroughs. The test evaluates whether there is a statistically significant difference in the proportions of felony arrests for this age group between Manhattan and each of the other boroughs.

Focusing on t-tests for felony arrests within the 18-24 age group serves multiple purposes. The 18-24 age group represents a critical transitional period when individuals may be more susceptible to involvement in criminal activities. By focusing on felony arrests within this age range, the

analysis can identify potential areas for targeted intervention and rehabilitation programs aimed at reducing recidivism rates and promoting positive behavior change. Felony offenses are typically more serious in nature and can have profound implications for public safety. Focusing on felony arrests within the 18-24 age group allows for a targeted examination of crimes that pose the greatest threat to community well-being, thereby informing strategies to enhance public safety and reduce the overall burden of criminal activity on society. Also by conducting t-tests comparing felony arrest proportions between Manhattan and other boroughs specifically within the 18-24 age group, the analysis can provide insights into geographical variations in crime patterns and law enforcement practices. This comparative approach helps to contextualize the findings and identify potential disparities or areas of concern that may require further investigation or intervention.

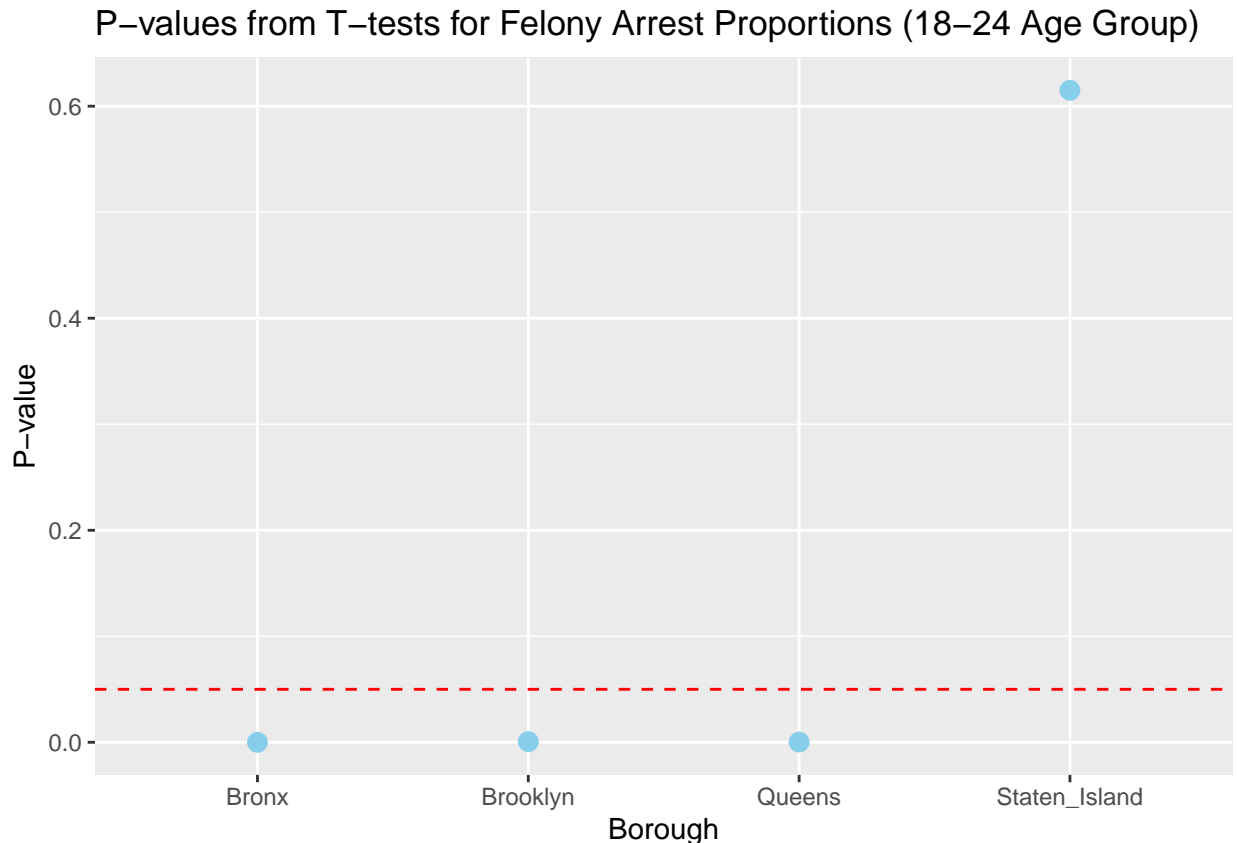
Null Hypothesis (Ho): The proportions of felony arrests for the 18-24 age group in Manhattan are equal to the proportions of felony arrests for the 18-24 age group in each of the other four boroughs.

Alternative Hypothesis (Ha): The proportions of felony arrests for the 18-24 age group in Manhattan are not equal to the proportions of felony arrests for the 18-24 age group in at least one of the other four boroughs.

If the p-value obtained from the test is less than the significance level 0.05, then the null hypothesis will be rejected, indicating that there is a statistically significant difference between the proportions of felony arrests for the 18-24 age group in Manhattan and the other borough being compared. The Alternative Hypothesis (Ha) then being true that there is a significant difference in those proportions. The p-values for these t-tests can be seen below in scientific notation.

##	Borough	p_value
## Bronx	Bronx	9.720353e-19
## Brooklyn	Brooklyn	6.068869e-04
## Queens	Queens	2.222006e-04
## Staten_Island	Staten_Island	6.149448e-01

It can be seen above that the p-values for comparing Manhattan to the Bronx, Brooklyn, and Queens, are all extremely low, and far below 0.05, while the p-value for Staten Island is quite large, as it is over 0.6.



Graph 8

Graph 8, which can be seen above, help to illustrate the magnitude of the p-values for each t-test. The graph shows the significance level 0.05 with a red dashed line to clearly see how close the p-values are to alpha. Therefore it can be seen that the values for the Manhattan-Bronx test, the Manhattan-Brooklyn test, and the Manhattan-Queens test are all extremely close to zero. The graph also shows the significance level 0.05 with a red dashed line to clearly see how close the p-values are to alpha. Further explanations can be seen below:

Bronx: The extremely small p-value (9.720353×10^{-19}) indicates strong evidence against the null hypothesis, leading to the rejection of the null hypothesis. This suggests a significant difference in the proportions of felony arrests for the 18-24 age group between the Bronx and Manhattan.

Brooklyn: With a p-value of 6.068869×10^{-4} , less than the significance value of 0.05, the null hypothesis is rejected. Consequently, there is a significant difference in the proportions of felony arrests for the 18-24 age group between Brooklyn and Manhattan.

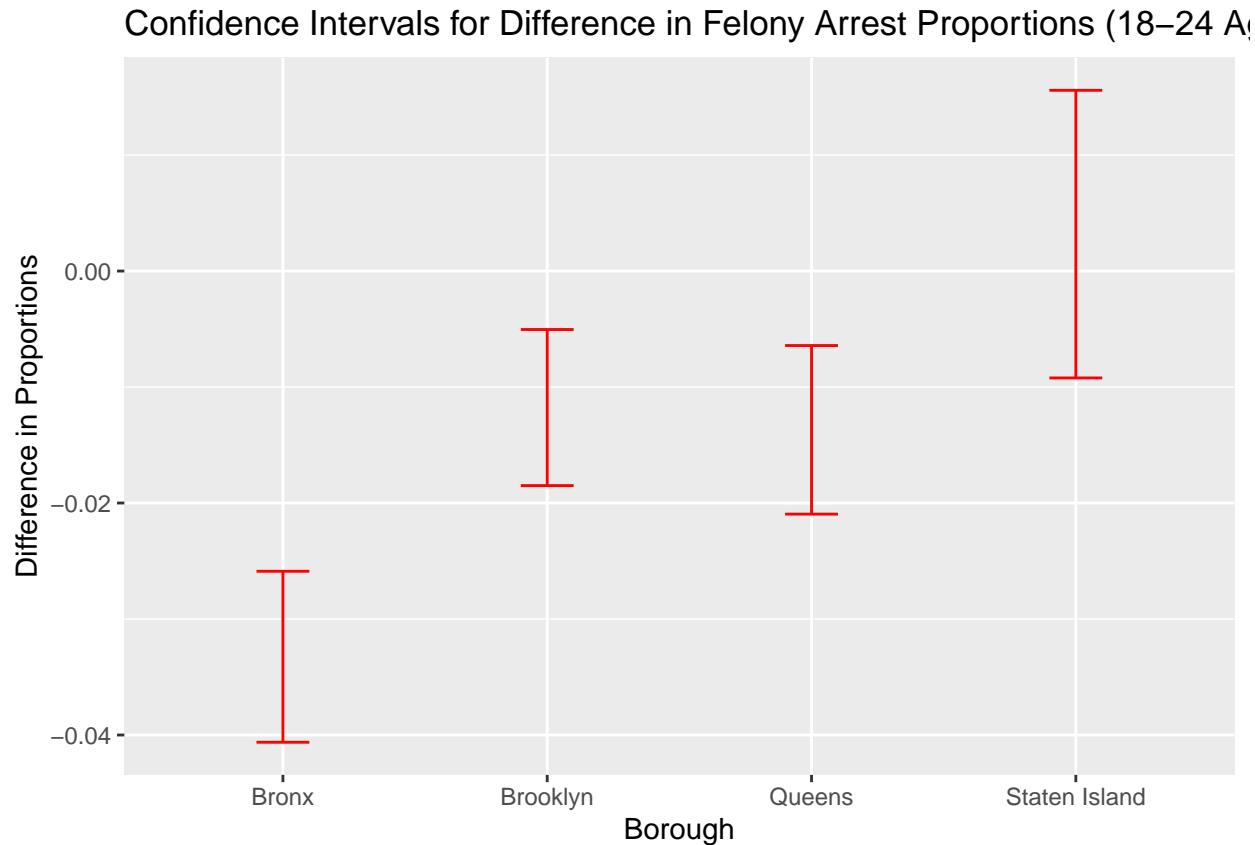
Queens: Similarly to Brooklyn, the p-value (2.222006×10^{-4}) is less than 0.05, indicating a significant difference in the proportions of felony arrests for the 18-24 age group between Queens and Manhattan.

Staten Island: The p-value (6.149448×10^{-1}) is greater than 0.05, suggesting insufficient evidence to reject the null hypothesis. Therefore, there is no significant difference in the proportions of felony arrests for the 18-24 age group between Staten Island and Manhattan.

Using the t-tests, the confidence intervals were also extracted. These confidence intervals represent the range of values within which we are reasonably confident that the true difference in

proportions of felony arrests for the 18-24 age group between Manhattan and each of the other boroughs lies. Each interval consists of two values: the lower bound and the upper bound. The intervals include both positive and negative values, indicating potential variability in the difference in proportions.

##	Borough	Lower_Bound	Upper_Bound
## Bronx	Bronx	-0.040627014	-0.025883932
## Brooklyn	Brooklyn	-0.018498984	-0.005041987
## Queens	Queens	-0.020958917	-0.006424240
## Staten_Island	Staten Island	-0.009217118	0.015580528



Graph 9

These 95 percent confidence intervals show that we are 95 percent confident that the true difference in proportions of felony arrests between Manhattan and the respective borough fall between the upper and lower confidence intervals. Something to note is the difference in the range of Staten-Island. In this case, the interval is wider compared to the others, indicating greater uncertainty about the true difference in proportions between Manhattan and Staten Island. This can be because of the significant difference in the population of Staten-Island and Manhattan which is much greater than between the other boroughs as Staten-Island has the lowest population and the lowest percentage of felonies committed.

Discussion

The analysis reveals significant variations in crime demographics across New York City’s boroughs, particularly concerning age groups and crime severity. By focusing on arrests for felony offenses among 18-24-year-olds, there is light shed on potential areas for intervention and policy formulation. The interpretation

of results suggests that Manhattan, despite its reputation, might not have the highest overall crime rates but does have a higher proportion of felony offenses, especially among young adults. This indicates a need for targeted intervention programs aimed at reducing recidivism rates among this age group.

One potential shortcoming is the reliance on arrest data, which may not fully capture the prevalence of crime due to factors such as under-reporting or differences in policing practices across boroughs. Additionally, the analysis focuses primarily on demographic factors and does not delve deeply into underlying socio-economic or environmental determinants of crime. Furthermore, the use of population data as a normalization factor assumes that crime rates are proportional to population size, which may not always be the case.

Future research could explore the underlying socio-economic factors contributing to crime disparities across boroughs, including poverty rates, unemployment levels, and access to education and social services. Additionally, longitudinal studies could track the trajectories of individuals involved in the criminal justice system to assess the effectiveness of rehabilitation programs. Exploring alternative methodologies, such as predictive modeling or spatial analysis, could offer complementary insights into crime patterns and trends across boroughs. Utilizing advanced statistical techniques like machine learning algorithms could help identify predictive factors associated with criminal behavior and inform targeted intervention strategies. Collecting data on additional variables, such as neighborhood-level socio-economic indicators, policing practices, and access to mental health services, could provide a more comprehensive understanding of crime dynamics. Longitudinal data tracking individuals' interactions with the criminal justice system over time could also offer valuable insights into patterns of recidivism and rehabilitation.

The primary conclusion of the analysis is that crime demographics vary significantly across New York City's boroughs, with age groups and crime severity playing crucial roles. Evidence supporting this conclusion includes the analysis of arrest data, graphical representations illustrating crime distributions, and statistical tests comparing felony arrest proportions between boroughs. These findings underscore the need for targeted interventions tailored to specific demographic groups and crime types to effectively address crime disparities across the city.