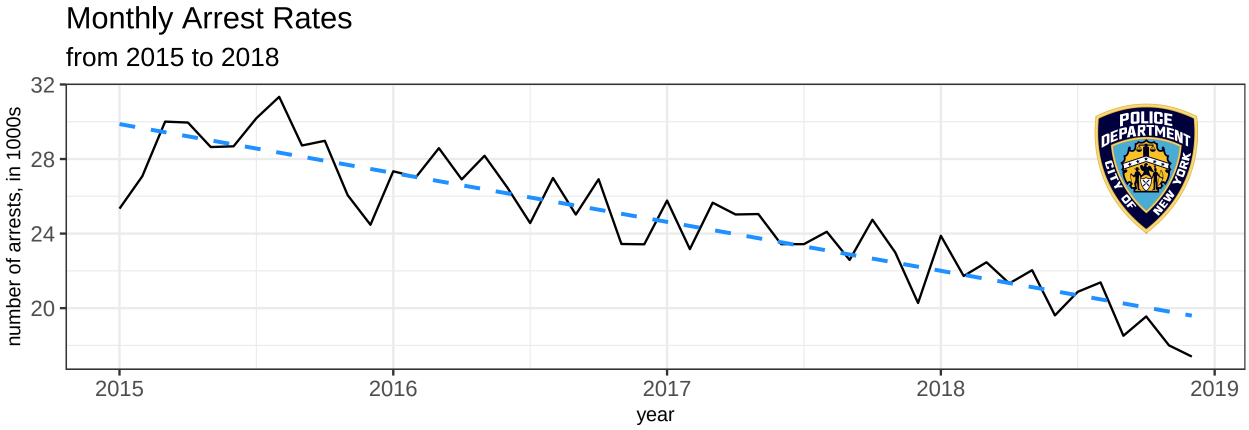
Analysis of NYPD Arrests

Darshan Patel

December 1 2019

Trend of Arrest Rate

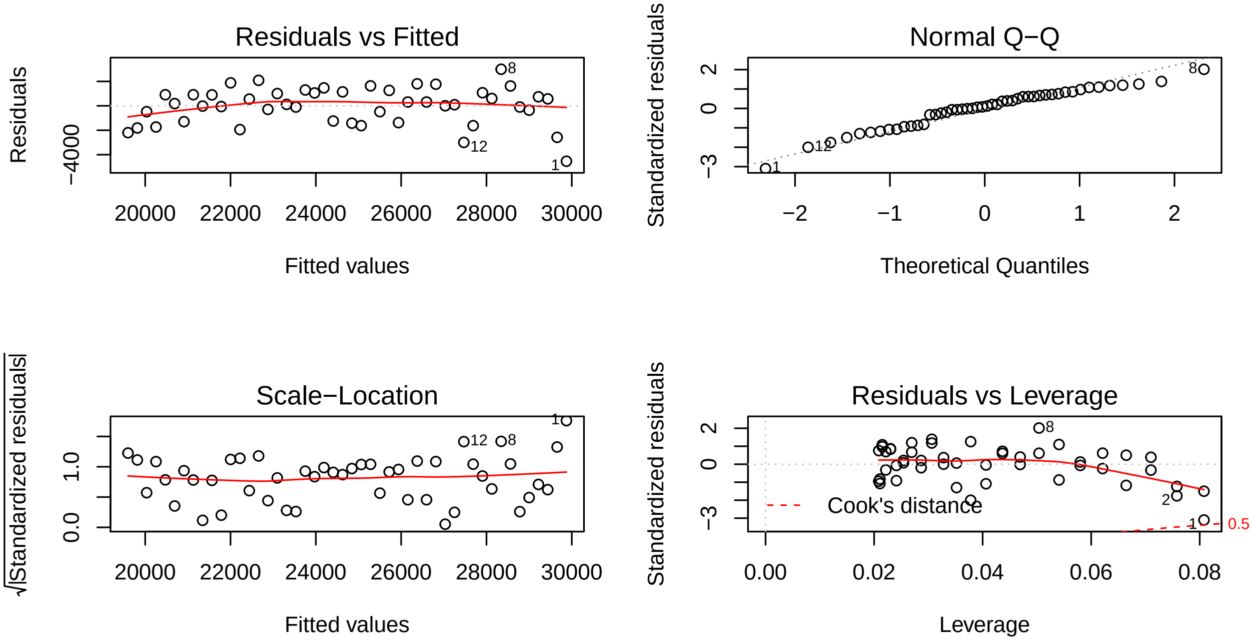
The New York Police Department has made 1187332 arrests from January 2015 to December 2018. A plot of the monthly arrest rates in this time frame has been created and shown below.



Source: NYPD Arrests Data (Historic)

As the years progressed from 2015, the monthly arrest rate in NYC have decreased. This can be proved using the linear regression model.

Before running the linear regression model, the assumptions of the model are verified. First, assume that the date values are fixed and measured without error. This is a given since time is measured without error and the dates are fixed. Now, assume that the observations are independent of each other. This can be loosely assumed if we let arrests of one person not be dependent on an arrest of another person.



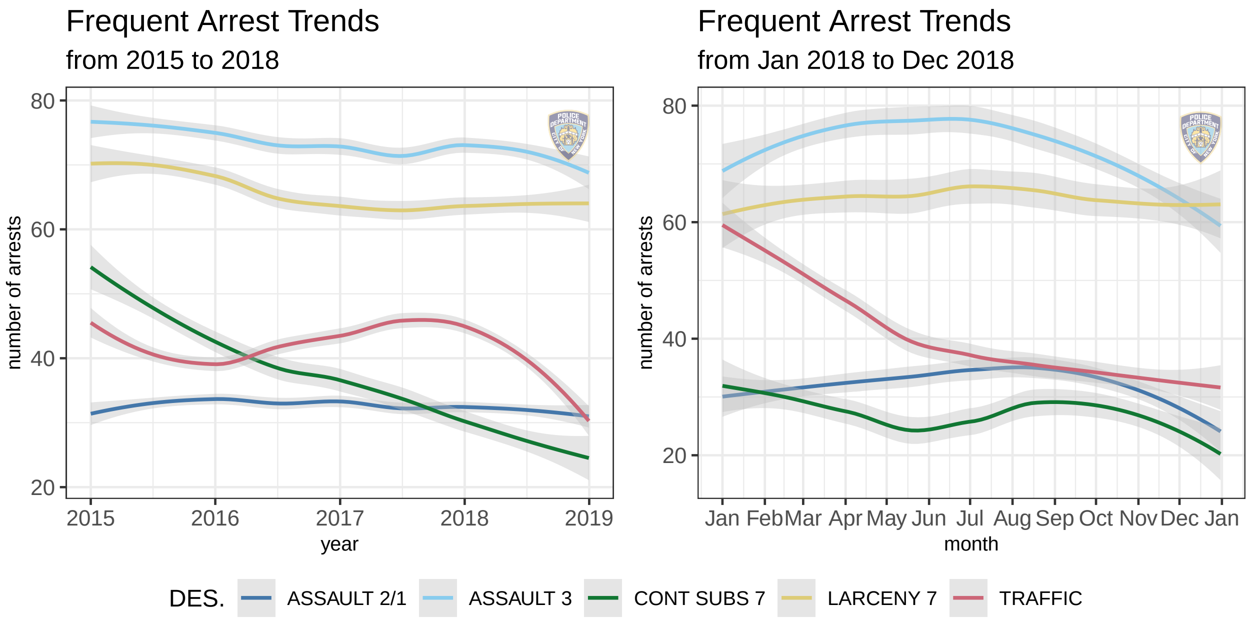
The next set of assumptions say that the relationship between the date and the number of arrests is linearly related. Furthermore, the residuals (errors in calculating number of arrests made) have a normal distribution, with a mean of 0 and a constant variance. The residuals will also be independent of each other. These can be verified by looking at the regression diagnostic plots of the model. Seeing the residuals vs. fitted plot, it is shown that the errors have no pattern, meaning that the relationship between number of arrests made and date is linear, and that the error has mean of 0. A look at the scale-location plot shows that the residuals have a constant variance. However, one point appears to have a high variance in its error. There are nearly equal number of points above and below the red line, indicating a constant variance. Nonetheless, the normal Q-Q plot shows that the residuals are normally distributed. Looking at the residuals vs. leverage plot also indicates that there are 2 points that have high influence on the regression line.

Since most of the assumptions are verified (or loosely verified), it is safe to use the coefficient estimate of the slope of the model. After running the linear regression model on the data, a downward sloping line is created, indicating that the arrest rates is decreasing as time goes on. This is shown on the plot of the arrests.

Another good statistical test to use is the Mann-Kendall trend test. It can give a deeper analysis in trend by looking at arrests by day. To use this, assume that the arrests made per day are independent of each other. Now, let the null hypothesis be that there is no monotonic trend present and that the alternative hypothesis be that there is a decreasing trend. After running the test, it is seen that given the test statistic z = -26.99, and the *p*-value, 9.21e-161, the null hypothesis can be rejected at the α level of 0.01. This means there is sufficient evidence that the trend in the arrest rate is decreasing.

Top Five Most Frequent Arrests in 2018

The top five most frequent arrests in 2018 were: (1) assault of level 2, 1, unclassified, (2) assault of level 3, (3) possession of controlled substance level 7, (4) acts of larceny, petitioned from open areas or unclassified, and (5) unclassified misdemeanor in traffic. Two plots are created to show how the number of arrests change, one by month from 2015 to 2018 and one by day from Jan 2018 to Dec 2018.


Frequent Arrest Trends
from 2015 to 2018

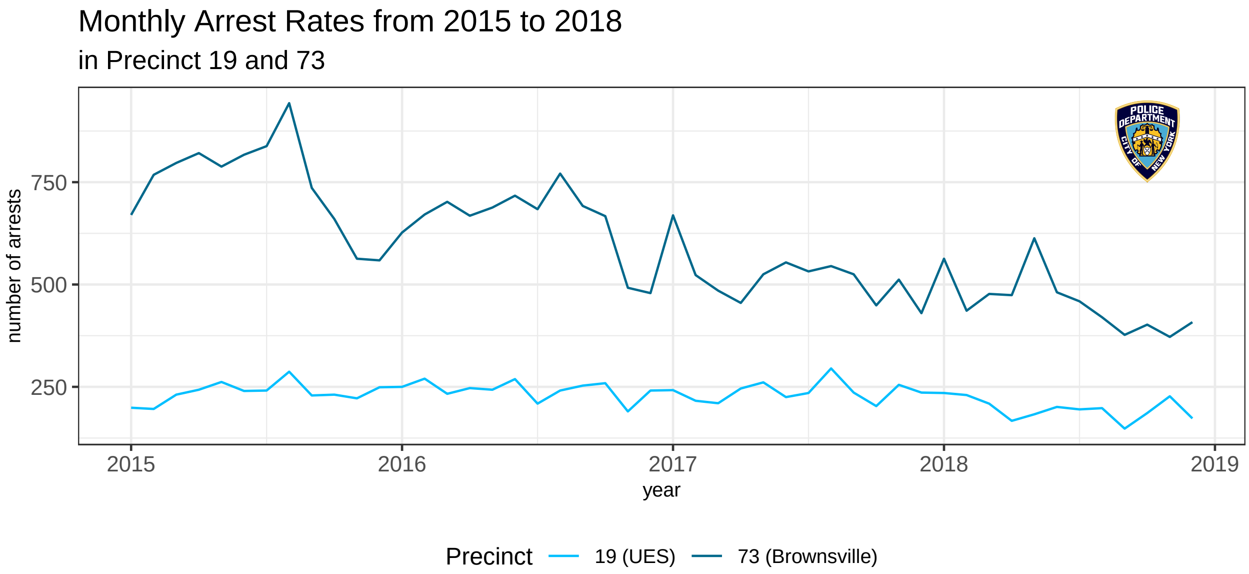
Frequent Arrest Trends
from Jan 2018 to Dec 2018

The number of arrests for possession of controlled substances (level 7) has been on the downfall from 2015 to the end of 2018. It shows the fastest amount of decline compared to the other four frequent arrests. The rate of arrests made for assault of level 2/1 has been steady since 2015 at around 32 arrests made per month. This contrasts with the arrests for assault of level 3, which has slowly been decreasing in rate since 2015, picked up slightly in the middle of 2017 and then went down again. The arrest rate for larceny (level 7) was at an all time high in 2015 and went down steadily up to the end of 2016. Since then, the arrest rate has been relatively constant around 63 arrests per day. As for traffic misdemeanors, the pattern has been fluctuating, where arrests rate went down til the end of 2015, went up through the fall of 2017 and then started rapidly decreasing til the end of 2018.

When looking at the arrests made daily in 2018, it is apparent that assault of level 3 arrests was at an all time high in the early summer at almost 80 arrests per day and then decreased to around 60 arrests in December. Arrests for assault of level 2/1 were considerably lower at 30 arrests per day in January, increases to 35 arrests in September and then goes down to 25 in December. The larceny arrest rate has been roughly steady throughout 2018 at 62-63 arrests per day. Traffic demeanors occured the most at the beginning of the year at around 50 to 60 arrests per day. The rate continued to decrease rapidly til June where it sat at 40 arrests per day and then slowly decrease til December where it is almost at 30 arrests per day. The arrest rate for controlled substance possession (level 7) has been fluctuating where it had an all time high in January at 32 arrests per day, goes down to 25 arrests per day in the middle of May, shot back up in mid August at 28 arrests per day and then reached an all time low in December at 20 arrests per day. In fact, this is the arrest type that had the lowest daily arrests by the end of 2018. Furthermore, its monthly arrest rate shown the greatest change in the four year span. Arrest rates for assault of level 2/1 have been relatively low but unchanging since 2015 while those of level 3 have been slowly declining. The arrest rate for traffic misdemeanors are most variable and could be investigated further.

Crimes in the Upper East Side vs. Brownsville

A plot of the number of arrests made per month from Jan 2015 to Dec 2018 for precincts 19 (Upper East Side) and 73 (Brownsville) is shown below.



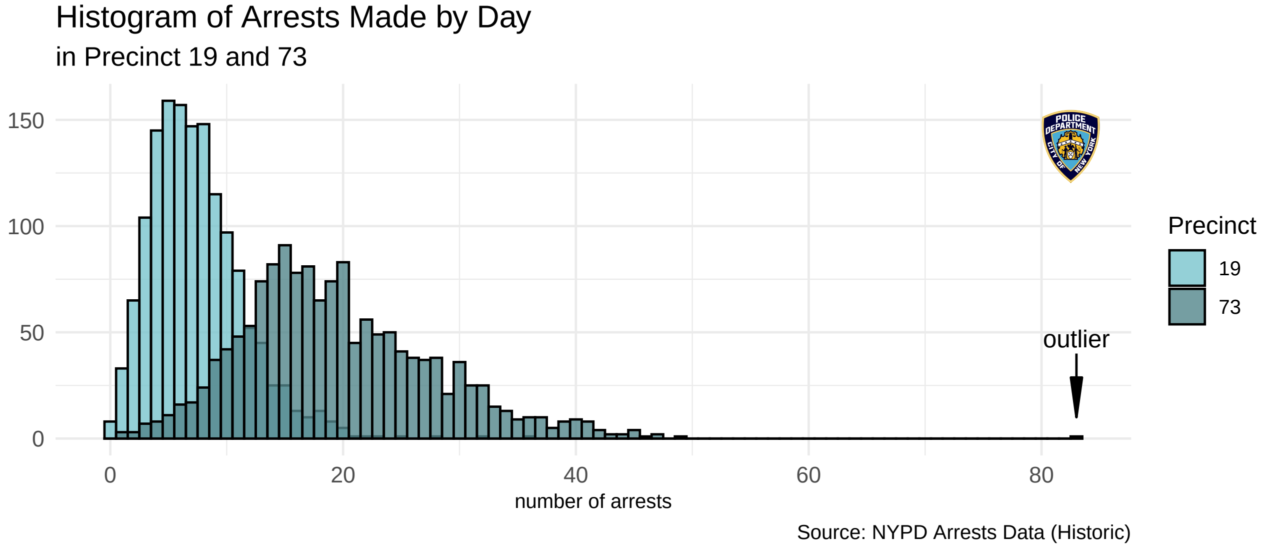
Monthly Arrest Rates from 2015 to 2018

in Precinct 19 and 73

Source: NYPD Arrests Data (Historic)

When looking at the big picture, if arrests are thought of as a sample of total crimes, then there has been less crimes in precinct 19 than precinct 73. But when looking more closely, there are some remarks that are to be made. Since 2015, the number of criminal arrests happening per month in precinct 19 has been in the ballpark of 250 arrests and only started going down slightly after 2018. There does not appear to be a trend present. As for precinct 73, monthly arrest rates have been everchanging between 2015 and 2018, with arrest rates the highest in July 2015 at more than than 875 arrests per month and then shooting down to 600 towards the end of 2015. Monthly arrest rates for this precinct continue to fluctate heavily between mid 700s and high 400s up to mid 2018 and then decreases further down to around 375 arrests per month at the end of 2018. The trend for criminal arrests in precinct 73 is decreasing, which shows that the area is becoming much safer.

A paired sample t test may be able to verify if the differences in number of arrests between both precincts from 2015 to 2018 is equal to zero, or less/greater than zero. To use this test, it would have to be true that the daily arrest rates are normally distributed and do not have any outliers. When plotted, it was found that there was one outlier, an abnormally large number of arrests (83!) made in precinct 19 on August 2 2018). In addition, the distribution of arrest per day in precinct 19 is not normal. This is seen in the histogram shown below.



Instead, a nonparametric method must be used for statistical testing.

The Mann-Whitney U test (or Mann-Whitney-Wilcoxon test) can be used to statistically determine if a randomly selected arrest rate from precinct 19 is less than a randomly selected arrest rate from precinct 73. Assume that all arrests made are independent from the other precinct. This can be loosely justified since the Upper East Side and Brownsville are not close in location and so arrests are from two different populations. In addition, assume that the arrest rates are ordinal, that it is clear which of two arrest rates is greater than the other. Since this is a non-parametric test, it is not needed to assume that arrest rates in both precincts follow a normal distribution. Now, allow the null hypothesis to be that the distributions of arrest rates in precincts 19 and 73 are equal. The alternative hypothesis is that the distributions are not equal and that the distribution of arrest rates for precinct 19 is shifted to the left of the distribution of arrest rates for precinct 31. Under these hypotheses, the test statistic is calculated to be V = 12350.5, with a p-value of 1.3e-225. At the confidence level of $\alpha = 0.01$, the p value is statistically significant and the null hypothesis can be rejected. The distribution of arrest rates for precinct 19 is shifted to the left of the distribution of arrest rates for precinct 31. This is also to say that there are more crimes in precinct 31 than precinct 19.

Predicting Crimes

To better allocate NYPD resources, a Bayesian model that predicts likelihood of different crimes in each precinct could be developed using the criminal history of the area. Variables that would be included in the model include past criminal history of the area, geographical locations of popular crime scenes by crime type and socio-economical information of the area. Included in the past criminal history of the area would be demographic information of the perpetrators. This could lead to racial profiling, a challenge that should be overcome. The model will be evaluated daily by founding out which crimes happened in the area that day and how that compares to the number of different crimes predicted by the model. The model will consistently learn new information on crimes happening in NYC and so the NYPD will be able to better allocate their resources.