Group 1: Analysis on NYC Open Parking and Camera Violations

Bingrui Chen, Jinglin Gao, Jiaowei Li, Yaopeng Ma

December 5, 2022

Introduction

New York City is notorious for the traffic and crazy drivers, driving in New York City is a nightmare, but finding a parking lot is even worse, especially in Manhattan. Due to the limitation of parking space and expensive parking garage rates, many people would risk it to park illegally on the street, some people get away from punishment, but most people would receive a ticket. In this project, we are going to look at parking violations in New York City, we collected the data from the five counties(New York, Richmond, Queens, Kings, and Bronx) belonging to New York City. We want to investigate if some hidden factors are playing an important role in the number of violations issued.

Data Description and Data Cleanning

Table 1: Columns in NYC Open Parking and Camera Violations

Feature Name	data type	description
Plate	text	license plate number
State	text	State registered
License Type	text	Type of Vehicle
Issue Date	text	Date of Violation Occur
Violation Time	text	Time of Violation Occur
Violation	text	types of Violation
Judgement.Entry.Date	text	Final Recording of the Decision and Opinion
Fine.Amount	number	The Initial Amount
Penalty.Amount	number	The late fee
Interest.Amount	number	The interests on the unpaid amount
Reduction.Amount	number	Amount Deduct by Court
Payment.Amount	number	The amount of payment made
Amount.Due	number	The outstanding balance
Precinct	text	a district of a city or town as defined for police purpose
County	text	County names
Issuing Agency	text	department names
Violations.Status	text	result of hearing if applicable
Image Link	website link	images of the ticket

The original data set contains over 15 million observations, and the file size is around 20 GB. Due to our computer power, we cannot load the entire data set into R. In order to minimize the data set without losing information beneath the data, we first inspect a small portion of data. There are 19 variables for each observation. First, we noticed that the status of the violations has 95% of values equal to Nan. So we decide to drop this variable. Also, the image of the parking tickets provides the same information as other variables. Thus it is removed from the data set. Other variables such as Judgement entry date and reduction amount have too many Nan values and are unrelated to our interests, so we can safely remove them from

the data. Our first supplementary data is the parking violations by fiscal year which is collected by the department of finance in NYC. This data set has all the information regarding every parking ticket issued in New York City. It has the same summon numbers from our data set. It also has the exact location of the violation, it also has a detailed description of the violation vehicles such as car make and car color. The second supplementary data set is obtained from the Department of Motor Vehicles in New York City, which has information on all the registered vehicles in New York City

Statements of Purpose

In this project, we will investigate the relationship of factors behind the data, and we decided on four goals for this project.

Purpose 1:

In New York City, also many people refer to it as Financial Hub. According to New York University estimates, approximately 3.1 million people in Manhattan during the work day, compared to a residential population of 1.6 million. Most workers commute from nearby regions(Long Island, New Jersey, and Connecticut). In some sense, the number of parking violations reflects the movement of people. Thus we assume that if there is a growing development trend, there will be more population movement, which will increase the total number of violations. On the contrary, when a recession occurs, there will be less population movement, which we think will reduce the total number of violations. We use the weekly economic index of New York City (WEI) to analyze the trend of development, after we are going to compare the trend between WEI and the number of violations given the x-axis to be Dated from 2018 till now. The pandemic that started in 2020, and protests activities during the summer of 2020 would affect both economics and the way parking rules were enforced.

Purpose 2:

We also have a question about whether the parking violation is related to the population in the same county because theoretically more people there are, the more violation would have. So we add secondary data about the population change of the five counties from 2018 to 2021, and then we will observe if they have the same tendency or follow the same distribution by ggplot visualization and some hypothesis tests.

Purpose 3:

From figure 6, we found that there are huge discrepancies between Richmond County and Queens County. For example, for the Violation type [7], there are a lot of violations that happened in Queens County. However, only a few of the type [7] violations happened in Richmond County. What is more, for the Violation type [5], we can see a prominent increasing number of violations in Richmond County compared with the other type of violations. Nevertheless, we found that Queens County did not show much more violations in this type of violation compared with the type[7]. Then we wonder if the proportion of all types of violations has a prominent difference between Queens County and Richmond County, or if it is just an exception.

Purpose 4:

We read an article about a study by researchers from the University of Nevada that suggests drivers of expensive cars are the worst. A similar experiment conducted by the University of Helsinki also noticed a higher percentage of luxury cars would those running through red lights and were unlikely to yield to pedestrians. Therefore, we want to see if this is true for parking violations. Our Parking Violations data set do not have any information regarding car make. Hence we collected detailed violations data through the Department of Finance in New York City. Their violations data is stored by fiscal years and contains the vehicle make and violation number that can use to join both data sets. Also, We obtained the data of all registered vehicles, boats, and snowmobiles from the Department of Motor Vehicles of New York City. We

want to compare the market share of the top 20 car make to the top 20 car make from violations data, to see if there is any anomaly.

Data Visulazition

First, Let us take look at the top 20 most popular cars distributed in the violation data , and in registered vehicles in New York city. If the car brands do not reflect if the owner is more likely to receive parking violations, then the distribution should be the same in both pie charts.

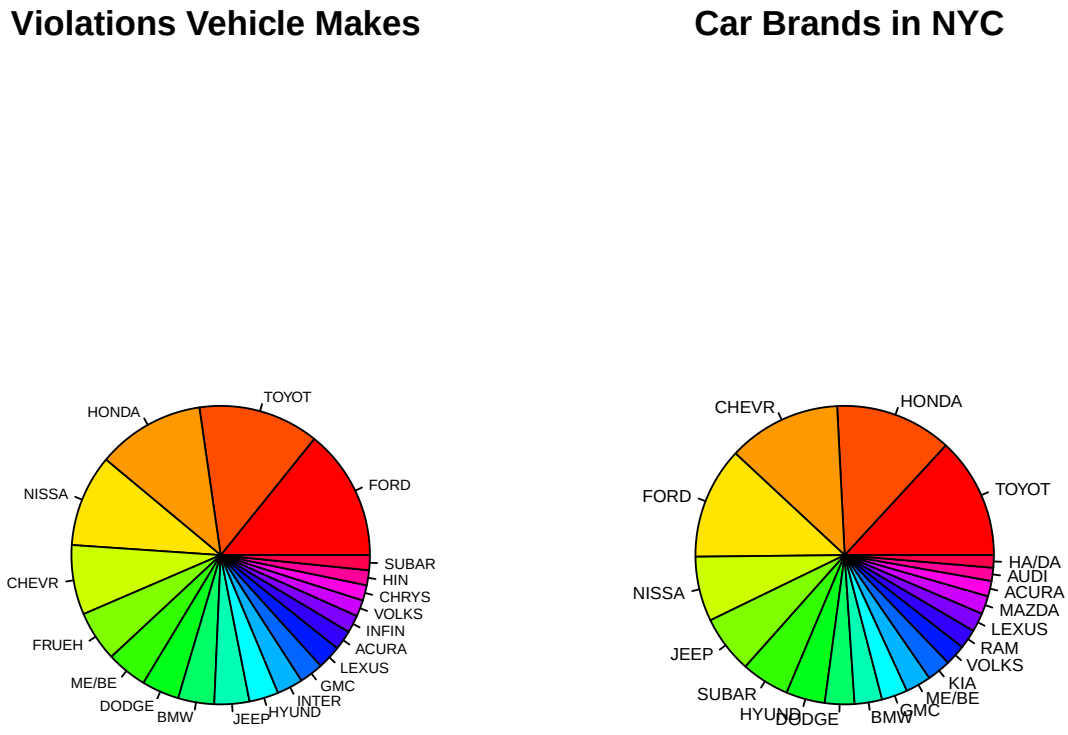
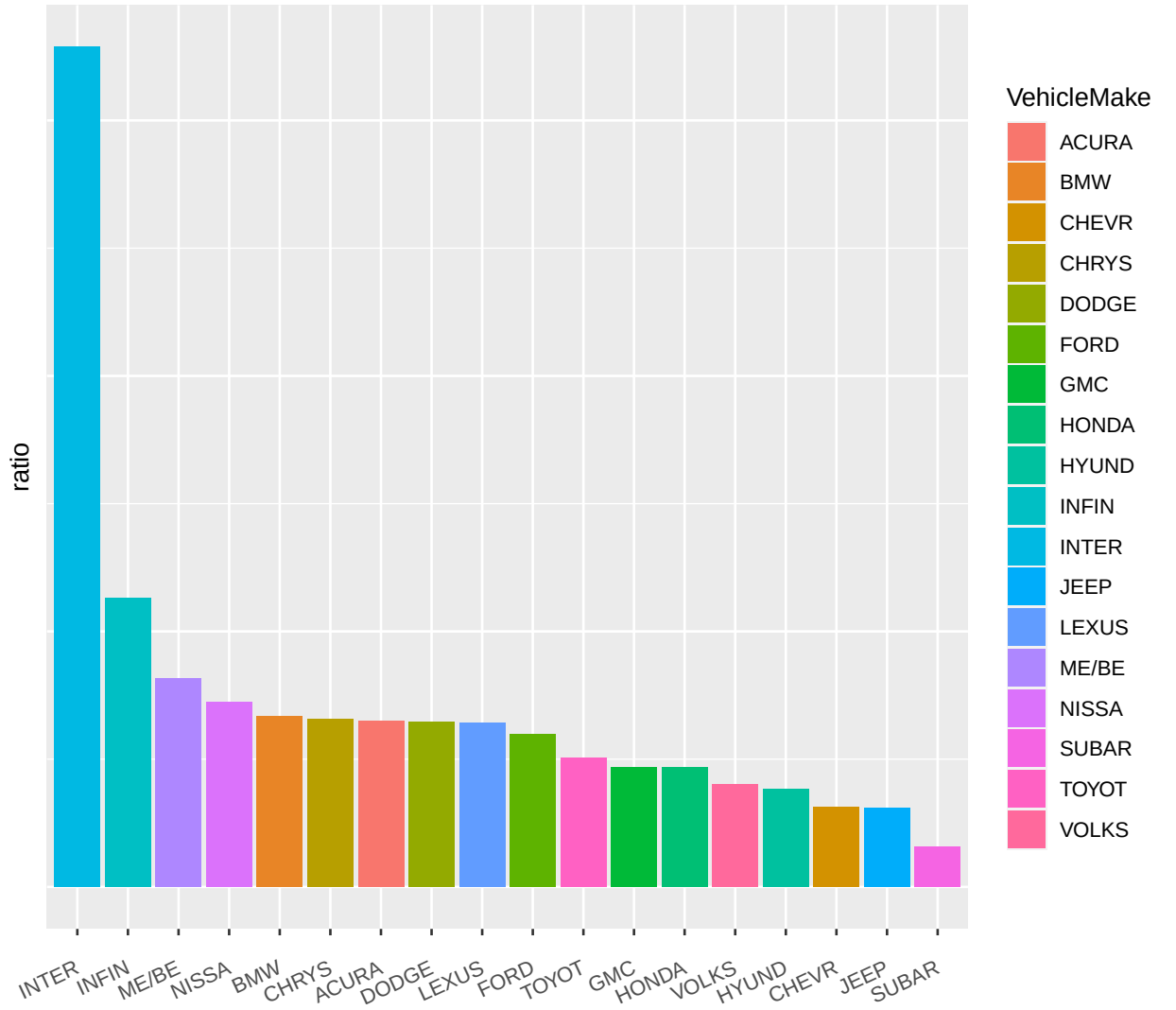


Figure 1: Comparison between the popularity of car brands and the total number of violations in NY city

We noticed that there is some difference in the two pie charts, but we still cannot make a conclusion based on the difference in the pie charts. After that, we calculate the ratio of the number of violations of each brand to the number of registered cars in New York City. We use the ratio to determine which car brand is more likely to receive a parking ticket. If the car has a really high ratio, then it means that people who drove this brand are more likely to violate parking rules.



We see that the first one has some outstanding ratios compared to the other. The INTER stands for international truck, which is a truck brand company. After observing the data, and from multiple online research. We conclude that most trucks are registered in other states since NYC is the transportation hub for the New England area. Therefore trucks from all over the country would travel through here. Since this is uncorrelated to our purpose, so we can disregard it. Now by looking top 10 brands of cars, we see that the Infinity, Mercedes Benz, BMW, Acura, and Lexus, are all considered luxury brands. This graph supports our assumption. The explanation for luxury car brands being more likely to receive parking tickets is that people who drove luxury cars would value their time more. Thus they would rather pay for the ticker than find an open parking space or go to the nearest parking garage.

Plot for Violations in Different Time Frame

Figure 2: In this part, we are interested in the difference in the total amount of violations among different counties. We want to see if the plot would tell us something about how the time frame would affect the number of violations given out.

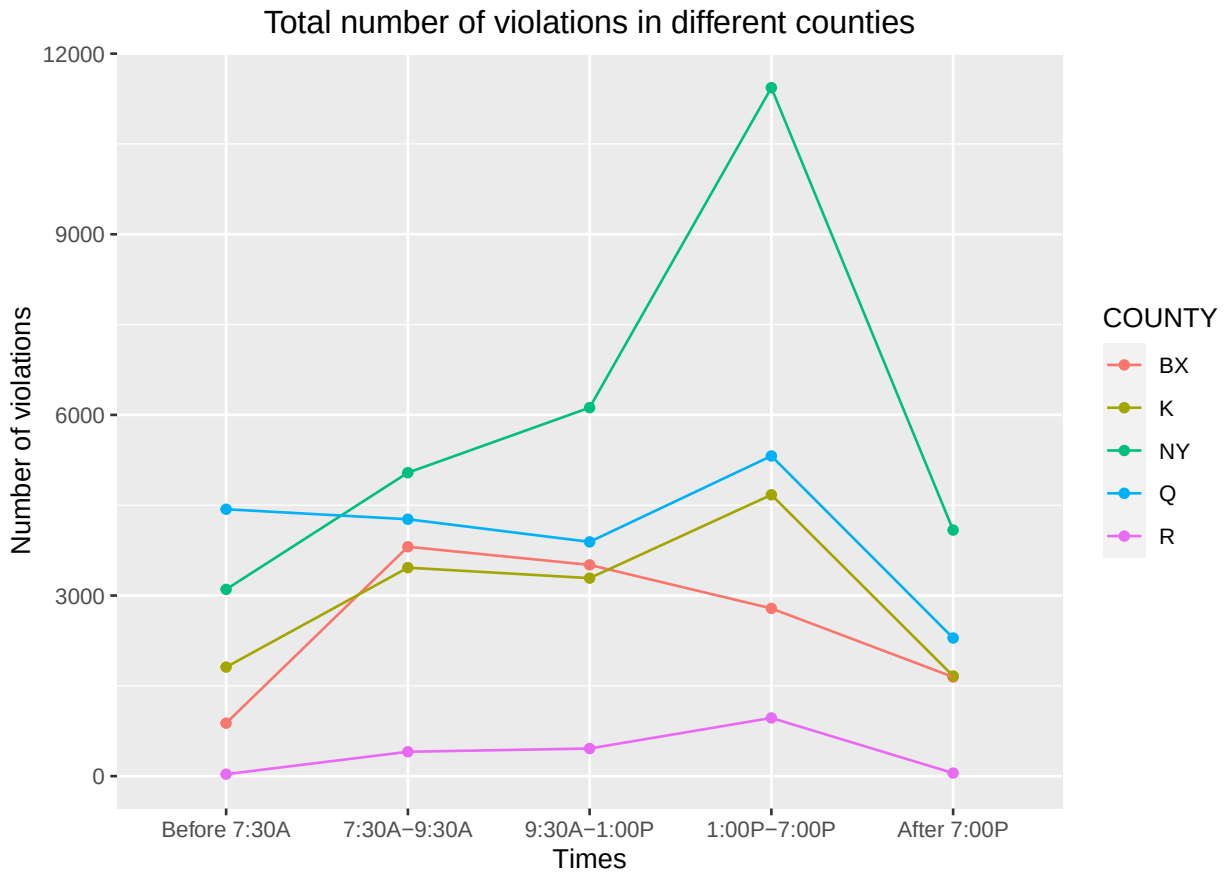


Figure 2: Violations in Different Time Frame

At rush hour between 1:00 pm to 7:00 pm, the number of violations reaches its maximum, and the daytime violations generally over the nighttime violations. But we can also see the difference between different counties. We can see that the line represents NY county is much higher than others, especially during rush hours. We think this is because most of the people work in NY and the incremental of vehicles on the road will increase the possibility of violations

Plot for Amount of Payment by Years

Figure 3: Here we plot the amount of payment of violations by every county in NYC. We want to see if the trend can provide us with any useful information regarding people traveling in and out of NYC from the beginning of 2017 to the end of 2021.

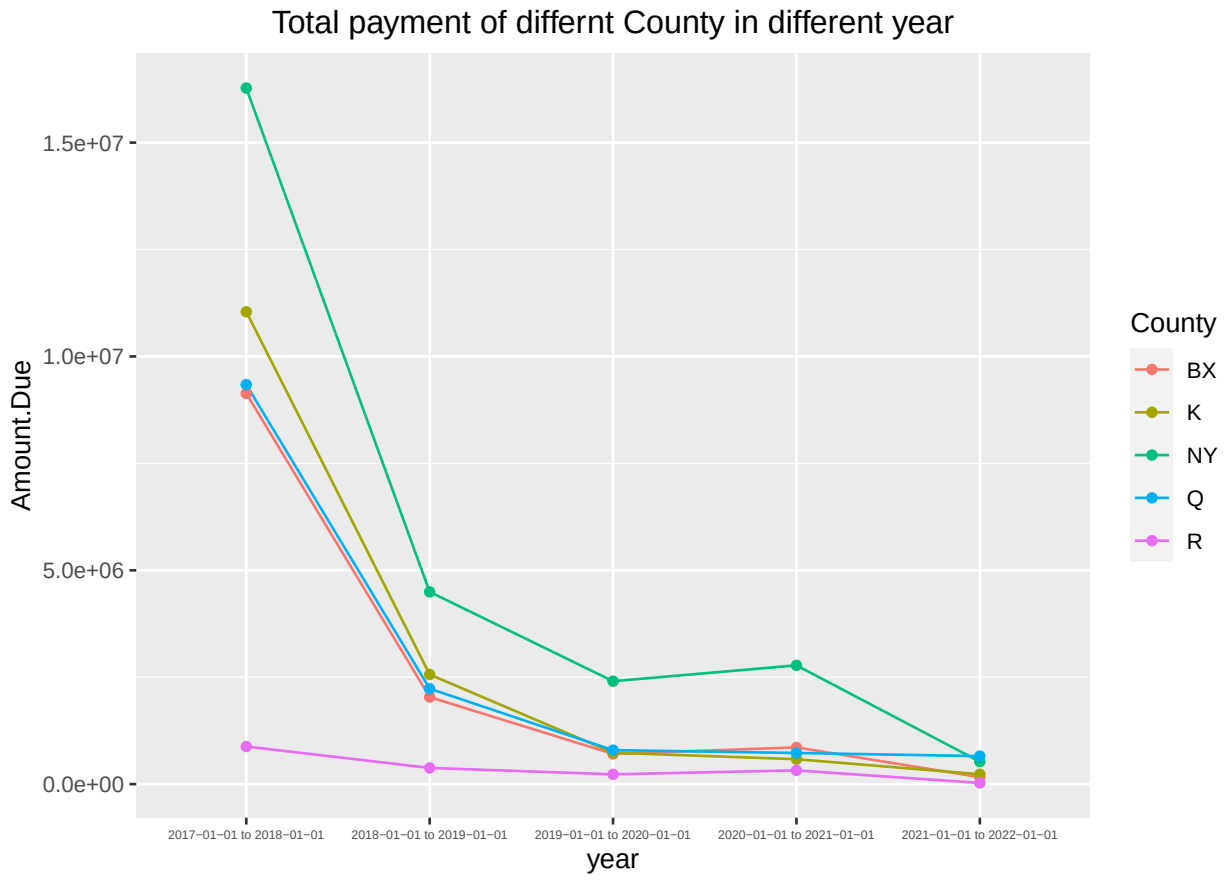
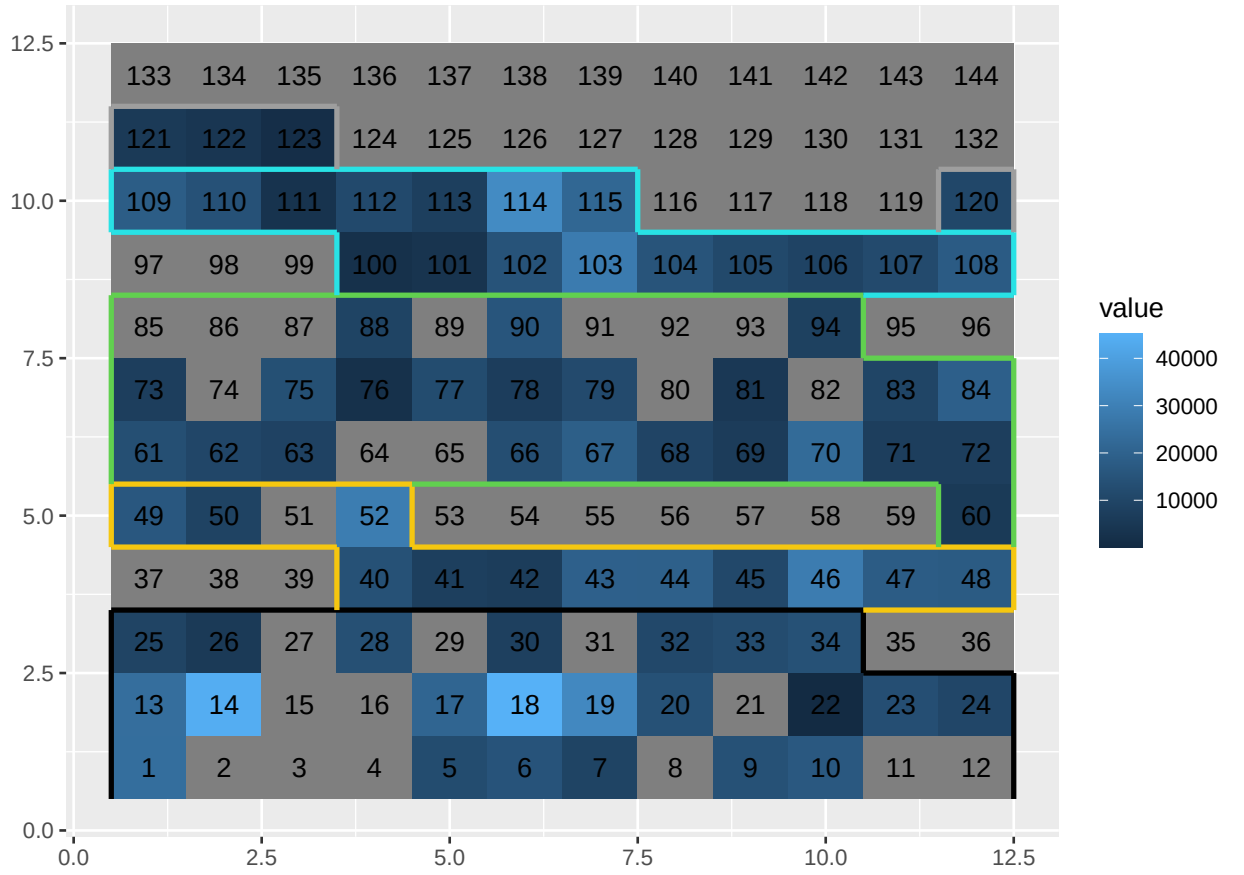


Figure 3: Amount of Payments by Years

The graph shows that the total amount of Payment for violations in every County is decreasing every year. The huge decrease in 2018 may be because of some unknown reasons in traffic policy that effect people's travel. And County NY always has the most payment amount with the largest range of fluctuations by year, County R has the least payment amount and hardly changed over time, which reflects that County NY is busier and maybe has more large traffic. In County NY, there is a continuing decrease in the intensity of traffic in 2021 and five Counties all stayed in a low traffic activity, maybe affected by the epidemics, so people could not go outside.

Heatmap of Number of violations by Precinct

In this section, we want to investigate which precinct has the highest violation numbers, and also which county it belongs to. Here each block represents a precinct, and the lighter the blue means a higher number of violations.



Since New York's precinct numbers are not continuous, in the plot, we use the gray squares to present that these numbers of the precinct do not exist. Each blue square represents a different precinct with the formal precinct number on it. The heat map shows the sum of violations in each precinct, and the lighter the blue, the more violations in this precinct. From the heat map, we can find that most of the violations happened in the black area and the blue area, which is Manhattan county and Queens county. What is more, in these counties, the 14th, 18th, and 114th precincts had more violations than the other precinct, and the precincts which are around these three precincts also had more violations than the other precinct which are away from the 14th and the 18th precincts. By searching the New York map, we found that the 14th and 18th precincts are the midtown south and north precincts which contain commercial offices, hotels, Times Square, Grand Central Terminal, the Theater District, and Restaurant Row, and have the highest footfall in New York.

Data Visulazition

Plot one: Histrogram of Top 10 Violation Types and County

Figure 4: Here we determined the top 10 violation types from the data set and plotted the bar graph to see any correlation between the number of violations and the counties. New York City has five Counties, Richmond County(R), Kings County(K), Queens County(Q), Bronx County(BX), and New York County(NY). And the violations code with the corresponding name is as the following:

Violation Code	Violation Name
1	"DOUBLE PARKING"
2	"EXPIRED MUNI METER"
3	"FAIL TO DSPLY MUNI METER RECPT"
4	"FIRE HYDRANT"
5	"INSP. STICKER-EXPIRED/MISSING"
6	"NO PARKING-DAY/TIME LIMITS"
7	"NO PARKING-STREET CLEANING"
8	"NO STANDING-BUS STOP"
9	"NO STANDING-DAY/TIME LIMITS"
10	"REG. STICKER-EXPIRED/MISSING"

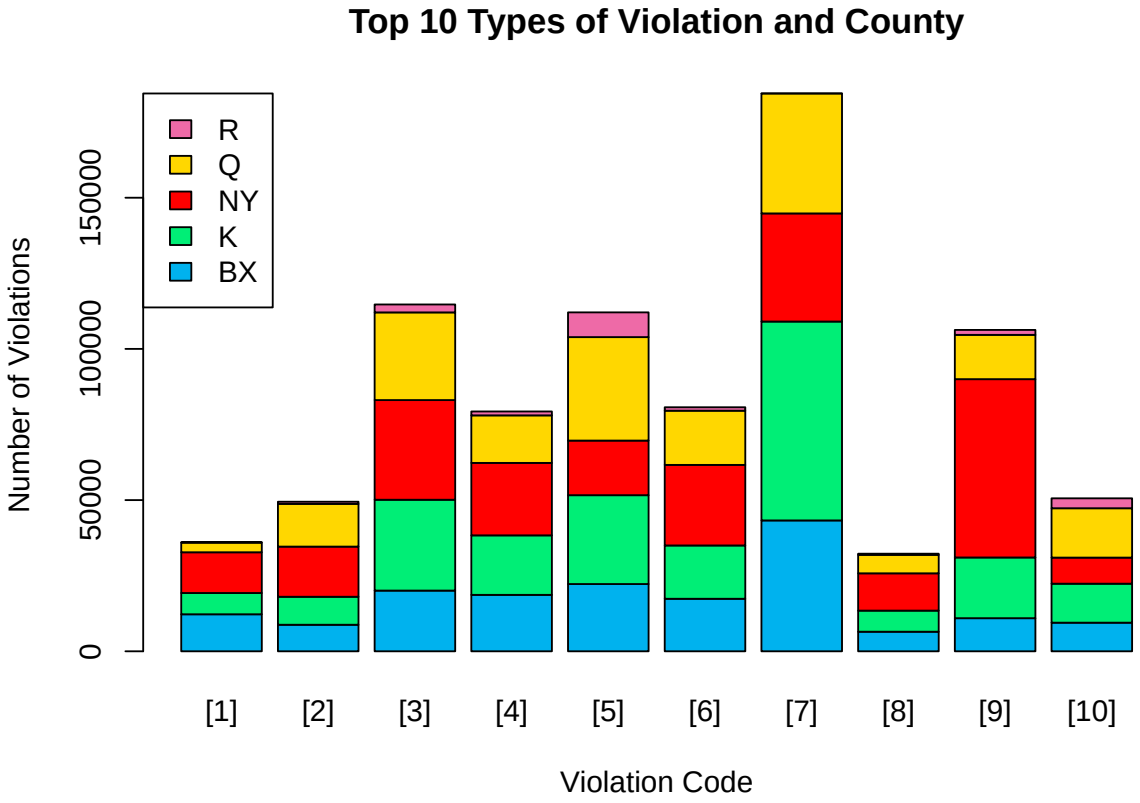


Figure 4: Top 10 Types of Violation and County

The graph shows some unexpected facts about these counties in the top 10 violation types. We see Richmond county(R) barely has any violations of these types. On the other hand, Queens County(Q), Kings County(K), and New York County(NY) make up a majority of these violations number. Here it may be helpful if we look at the locations of these counties. Richmond county is far away from the busy area of NYC, and the other three counties are at the heart of NYC, especially New York County, also known as Manhattan. It is the home of 1.7 million people, and only limited parking spots are available on this small land. Hence, the number of parking violations is extremely higher compared to the others.

Plot two: Comparison Between Richmond County and Queens County

Figure 5 The following graphs compare Richmond county and Queens county by examining their top violation types. The top violations types selected by the number of violations that exceed the county average.

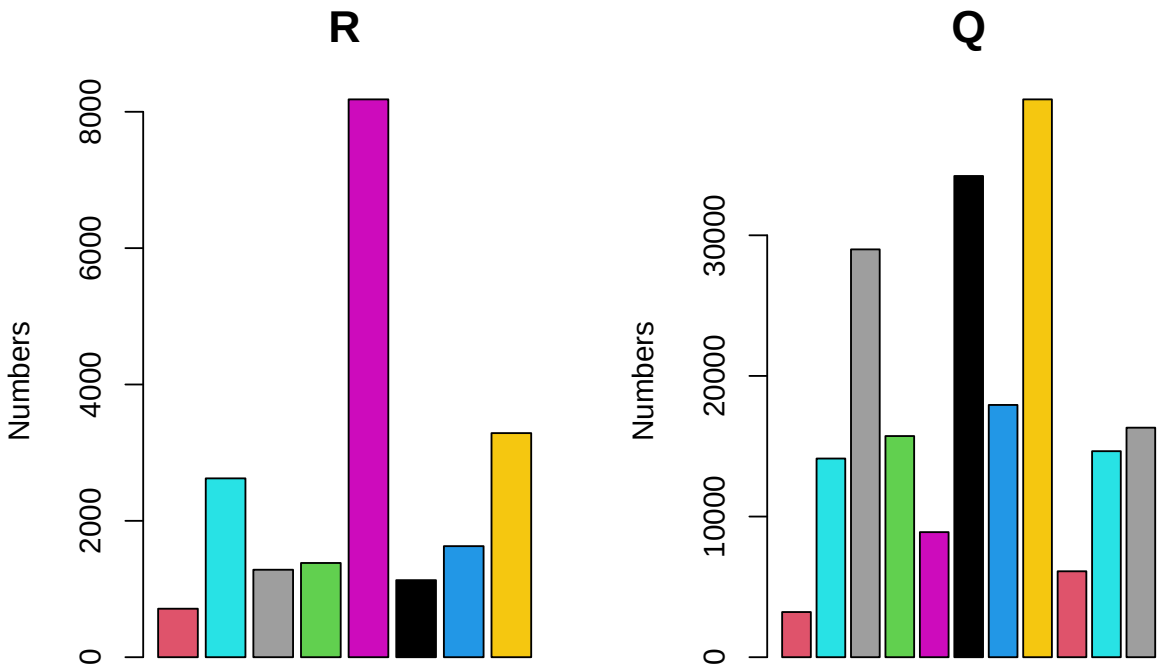


Figure 5: Comparison Between Richmond County and Queens County

From the top 10 types of violation and county's barplot, we find that the proportion of some types of violations has a prominent difference between Queens County and Richmond County. Then we want to analysis more about this phenomenon, so we abstract all types of violations which numbers above the average in Richmond County and Queens County. Then we found that some differences. For example, for the Double parking violation, there are a lot of violations that happened in Queens County. However, only a few of the Double parking violations happened in Richmond County. What is more, for the sticker-expired/missing violation, we can see a prominent increasing number of violations in Richmond County compared with the other type of violations. Nevertheless, we found that Queens County did not show much more violations in this type of violation compared with the Double parking violation.

Time Frame and Number of Violations

Figure 6 Here, we are going to look at the data for several years, and for each year, we calculate the number of violations given out at each time frame during one day. The time frame is separated by the night time, morning rush hours, lunch hours and afternoon rush hours. We want to see if there is any trends in different years.

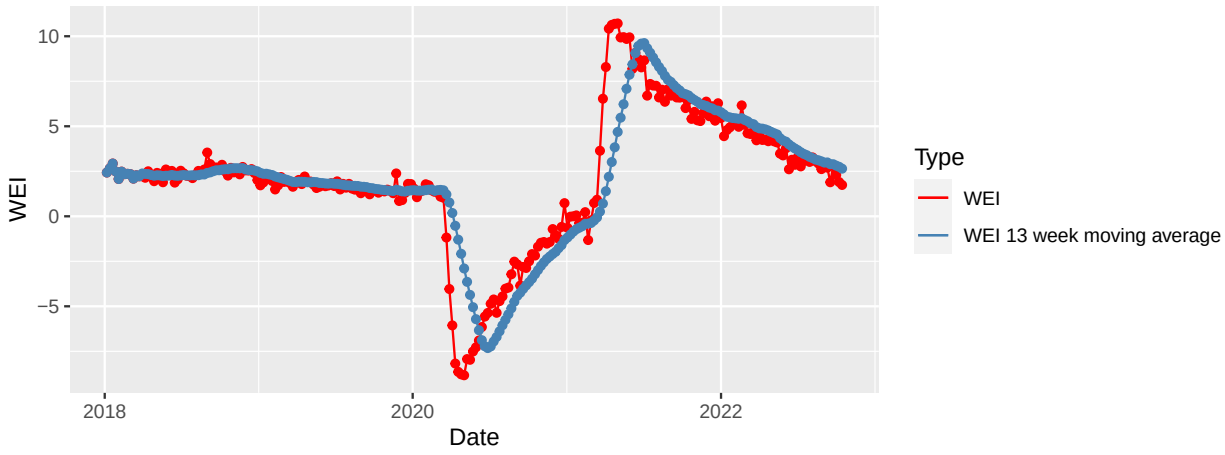


Figure 6: Weekly Economic Index (WEI) of NewYork City

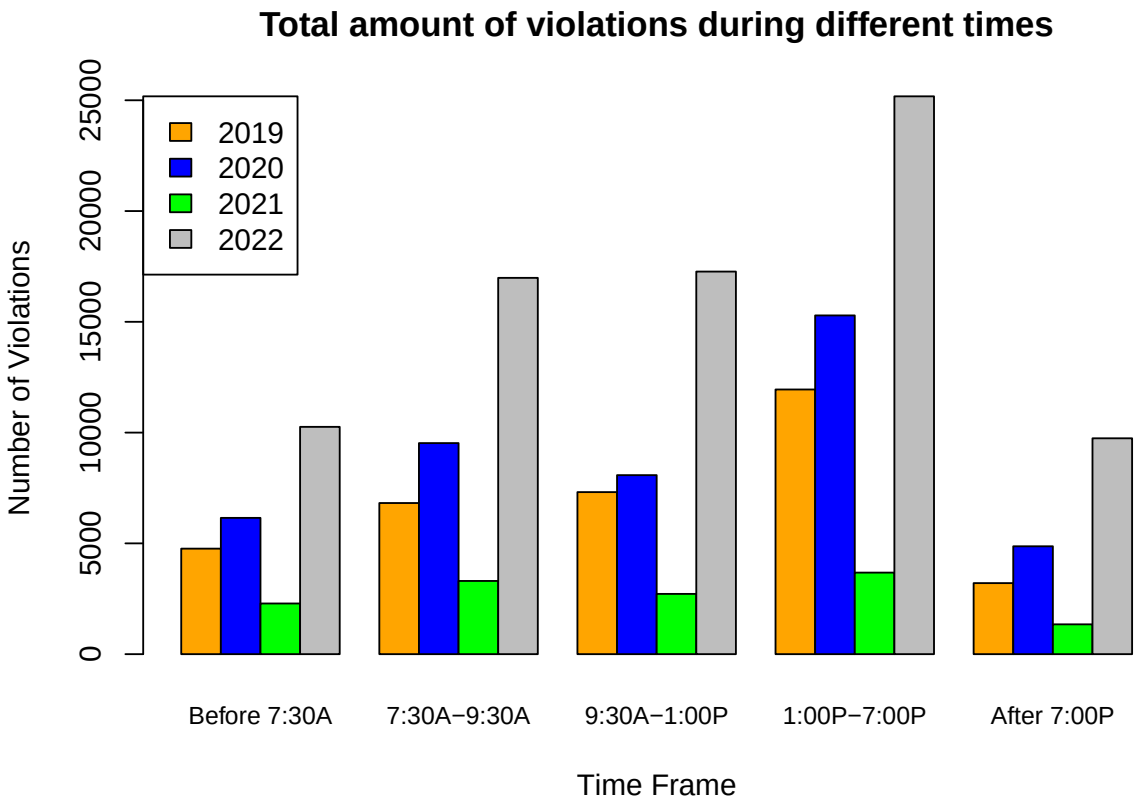


Figure 7: Number of Violations by Timeframe in Different Years

From the graph, we see a steep drop in the number of violations in every time frame starting in the year

2020, this is when the pandemic first started and the quarantine started. Many families sold their cars at the beginning of the pandemic because most people were staying home, less travelling implies less usage of cars, therefore less parking violations were given out. In 2021, there is another drop in the number. The reason for this could be the many companies in NYC offer working from home for their employees, which reduced the number of cars travelling into NYC. Hence, there are more parking spots in the city, and fewer parking violations given out.

Model

We noticed from figure 6 and figure 7 that both of them have an obvious drop during the years 2020 and 2021. Thus, we assume that WEI is highly likely to have some relationship with the number of violations. We first tried to fit the data with a linear model but it turned out to be model under-fitting and poor prediction. Therefore, we modified our data by doing logarithm on both WEI and the number of violations and then fit it with a generalized additive model (non-linear regression model). The result is acceptable with errors in a reasonable range. We conclude that we can estimate the number of violations by checking the economic status, generally speaking, the more prosperous the economy, the higher the number of violations. However, the interaction is not linear.

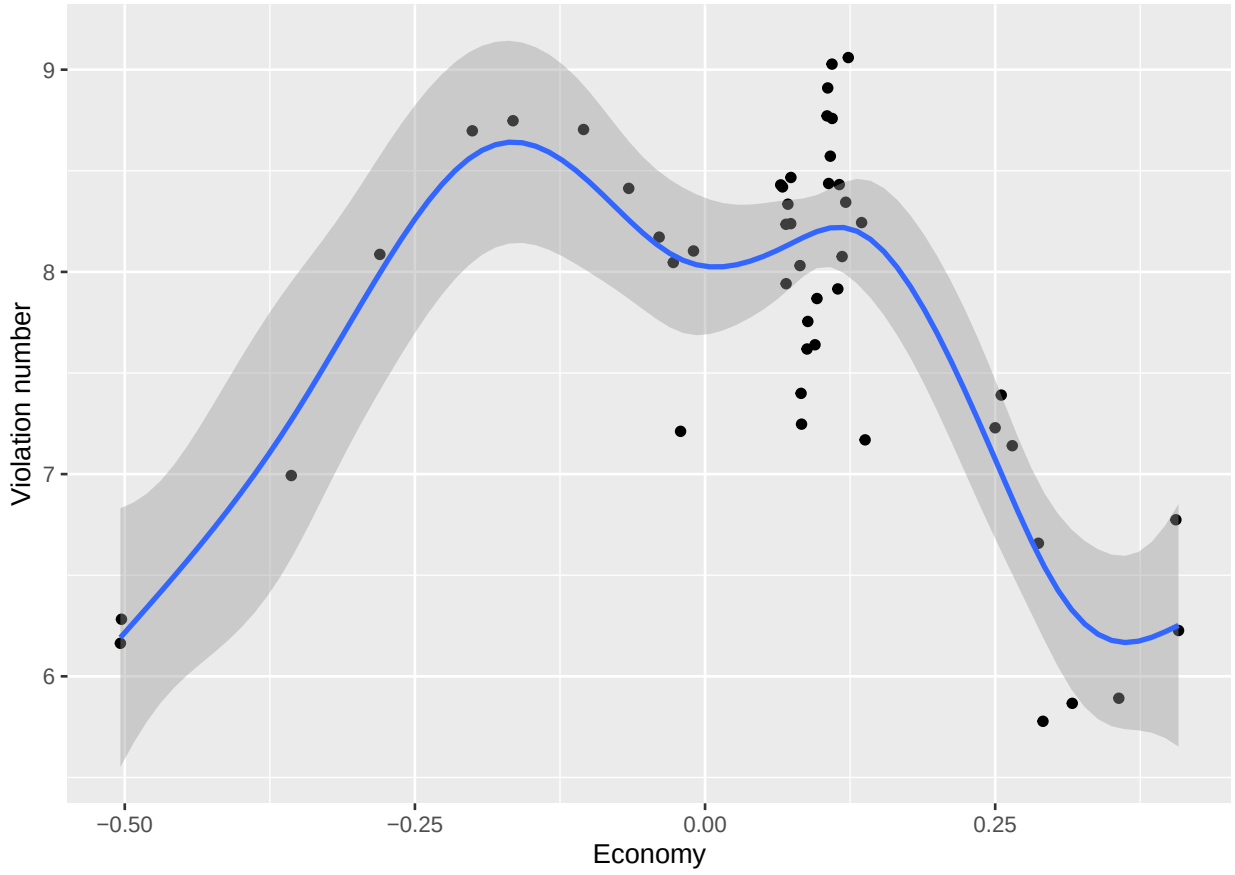
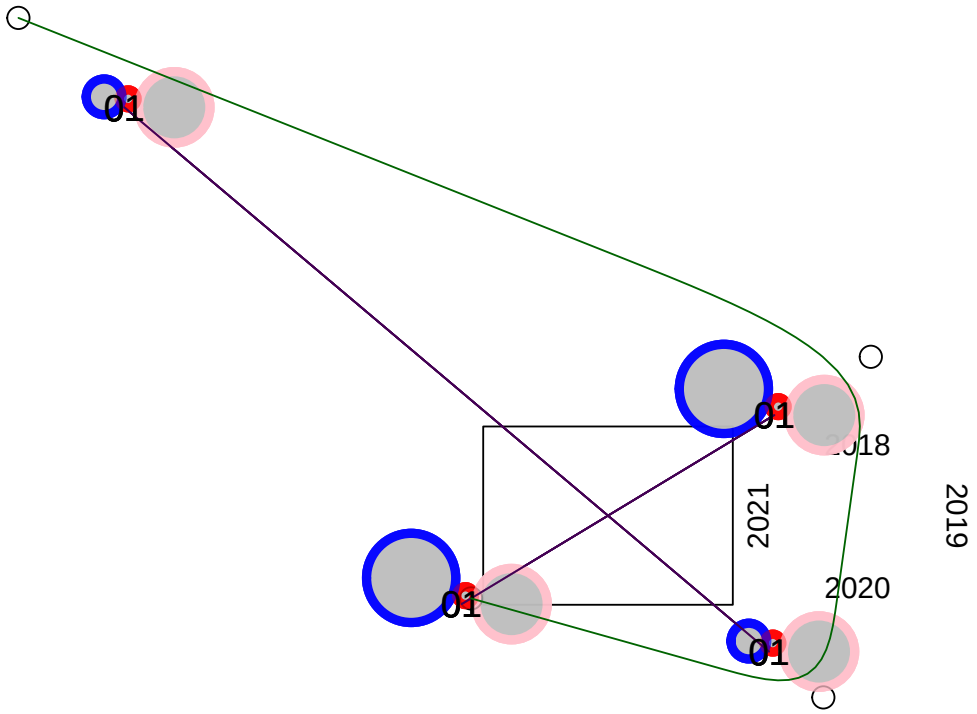


Figure 8: Number of Violations by Timeframe in Different Years

Killer Plot

The idea of our killer plot is inspired by the hot balloon. The plot consists of balloons, rope, basket. Each component represents different features. The basket is the square in the middle, with each side representing one year, from 2018 -2021. The side divides into 12 months. The balloons are a circle on the outside, the size of the balloon demonstrates the number of violations, and the color of the balloons represents five counties in New York City. The rope is the line that connects the basket and the circle. The rope shows the total amount of fines collected at the current time. Longer the lines, the higher amount of fines. The green lines outside the balloons are the economy curve. Due to the pandemic, we see that there is a steep drop in the economy, and the trend of the number of violations is also decreasing.

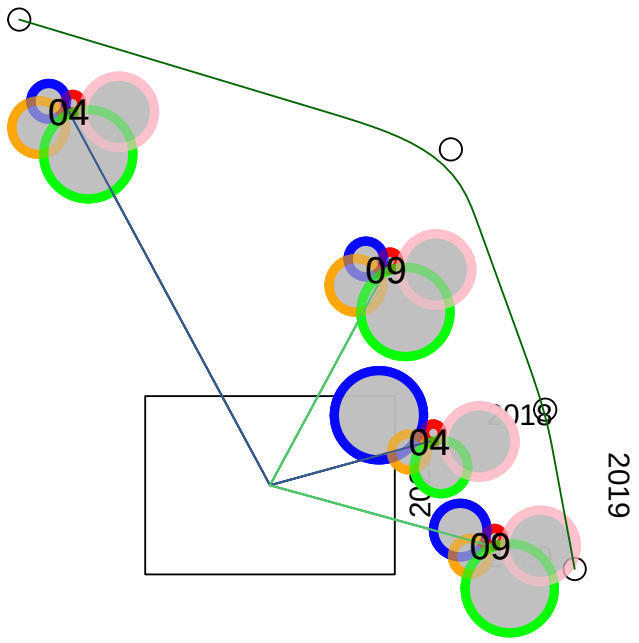


	Issue.Date	Amount.Due	BX	K	R
1	2018-01	1555343.6	1095	2305	155
13	2019-01	452586.8	507	397	38
25	2020-01	517763.6	533	648	435
37	2021-01	398278.4	586	483	80

Figure 9: Killerplot-1

This plot can show us many different relationships between the features, such as which county has the highest number of violations, and which month has the highest number of violations. Also the relationship between the economy and the number of violations issued.

Another View



	Issue.Date	Amount.Due	BX	NY	Q	K	R
4	2018-04	1163436.7	1184	2218	1229	1548	268
9	2018-09	672106.0	642	1517	707	790	148
16	2019-04	354777.2	290	989	464	276	60
21	2019-09	490755.3	498	1922	470	806	76

Figure 10: killerplot-2

Conclusion

By viewing all these relationships between different variables and the number of violations, we have the following conclusions: People who drive luxury cars are more likely to get parking tickets, maybe because of a sense of superiority among others, or because they do not want to waste the time to find the parking lot. Also, the economy plays an important role in the number of people traveling to New York City. According to New York University estimates, during work days, there are 3.1 million people in Manhattan, compare to the 1.7 million residents in Manhattan. Most people go to New York for work or pleasure. If the economy is not very promising, people would most likely reduce traveling, and Companies will lay off workers to reduce costs.

Reference

1. Federal reserve bank of NEWYORK. *(n.d.). Weekly Economic Index (WEI). Retrieved November 2, 2022, from https://www.newyorkfed.org/research/policy/weekly-economic-index#/interactive*
2. Department of Finance. *(2016-2022) "Open Parking and Camera Violations". Available at:https://data.cityofnewyork.us/City-Government/Open-Parking-and-Camera-Violations/nc67-uf89 (Accessed: 22 09 2022)*
3. Department of Motor Vehicles. *(2022 "Vehicles, snowmobiles, boat registration". Available at: https://data.ny.gov/Transportation/Vehicle-Snowmobile-and-Boat-Registrations/w4pv-hbkt (Accessed: 10 10 2022)*
4. "If You Drive an Expensive Car You'Re Probably a bad driver, Scientists Say." *Https://Www.cnn.com/2020/02/26/World/Expensive-Car-Drivers-Study-Scli-Scn-Intl/Index.html, CNN, 26 Feb. 2020,*