Manrattan - A Look into NYC's Rats

Alex Cheng, Liz Chu, Arthur Jakobsson, Kevin Ren

December 12, 2022

Rat Population Data Analysis - New York City

Descriptions of Datasets

Rat Sightings Dataset

https://data.cityofnewyork.us/Social-Services/Rat-Sightings/3q43-55fe

Our analysis primarily revolves around this dataset, with several supplementary datasets appended to this one for further in-depth analysis. This dataset contains 208,000 different rat sightings in the City of New York between 2010 to the present day, reported by citizens to the City of New York and accessed from NYC Open Data. 38 different variables are recorded for each sighting; notably, geographic data such as latitude, longitude, and borough data, and the date of opening and closing of the complaint.

Supplementary Datasets

We join various auxiliary datasets (described below) to our rat sightings dataset in order to better examine how rat sightings correlate to other demographic and geographic factors.

Subway Dataset

https://data.cityofnewyork.us/Transportation/Subway-Entrances/drex-xx56

This dataset, also sourced from NYC Open Data, contains the names, line numbers, and geographic coordinates of 1928 subways in New York City to date.

Tax Return Dataset

https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi

This is an 2019 IRS-sourced dataset which contains tax return information for each of the 178 zip codes in NYC; namely, the number of returns and total amounts requested by eligible citizens of each of the zip codes for their individual tax returns.

Restaurant Inspection Dataset

https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j

This is an NYC Open Data dataset, most recently updated on December 10, 2022, containing 231,000 data, each corresponding to a health violation citation given to a restaurant in NYC by the City of New York's Health Department. We are given 27 different variables that most importantly provide the location and zip code of each restaurant which was issued a citation.

Research Questions

Going into this project, our group had several questions we wanted to answer regarding the distribution of rats in the city. Namely:

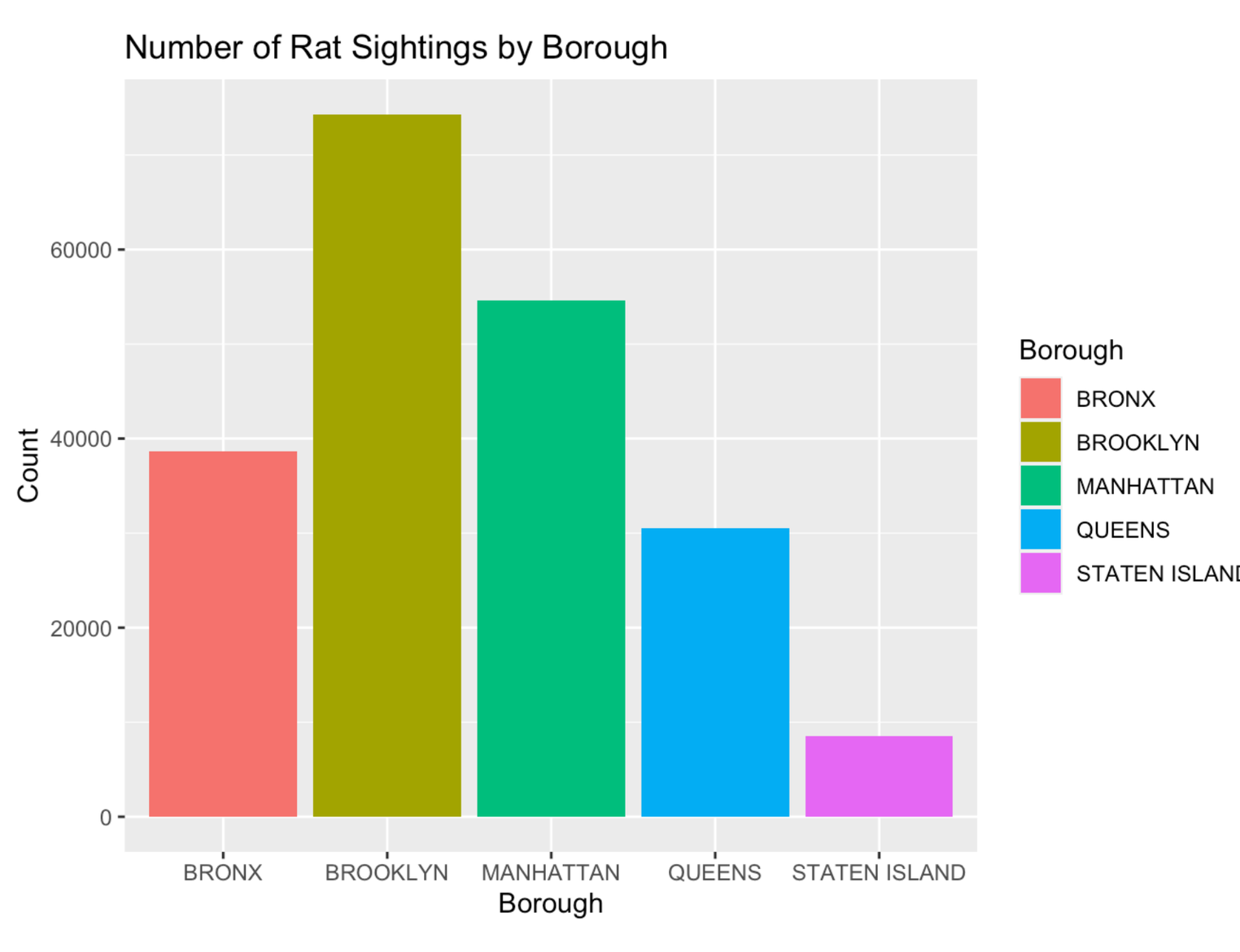
1. How do rat sightings differ geographically and by borough?
2. How has the number of rats reported changed over time?
3. How does well do wealth and geographic data combined correlate with rat sightings?
4. How do rat sightings correlate with candidate features such as subways and restaurants?

In all, we hope to make underlying observations that extend beyond the mere topic of rats, using rat sightings as a proxy for deeper conclusions about socioeconomic and geographic patterns in the City of New York.

Graphs Made

Borough Bar Chart

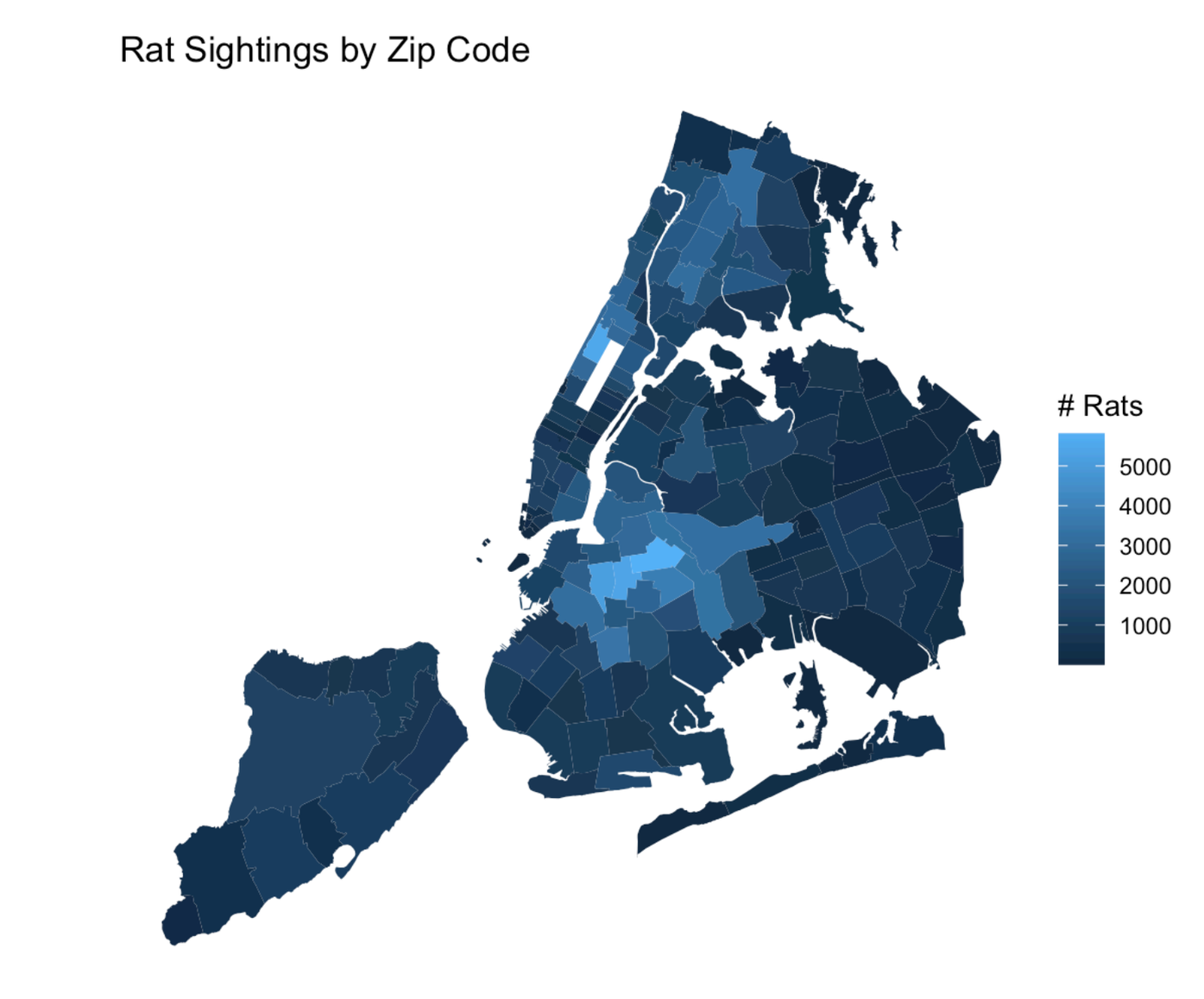
Our first visualization performs some elementary EDA on the distribution of rat sighting counts given the borough of their reporting. We created this graph in order to very directly address our research question of how rat sightings differ by borough.



This bar chart displays the number of rats seen within each borough in New York City. Brooklyn had by far the most rat sightings at 74,302, followed by Manhattan, the Bronx, Queens, and Staten Island, in that order. The low number of rat sightings in Staten Island might reflect its more cut-off nature from the rest of the city, as well as its more suburban feel, which could plausibly explain why Staten Island suffers less from the very urban problem of rats compared to the other boroughs in the city. Similarly, Brooklyn's position in the dead center of the city may explain why it had so many rats. Despite being the smallest borough by land size, Manhattan had the second-most rats, which may reflect the fact that it is one of the main business centers in the city (and the world) which would obviously attract a large number of rats with high concentrations of people and food. Thus, this simple visualization of rat sighting counts allows for greater generalizations about the boroughs in the city.

Rat Population Choropleth Map

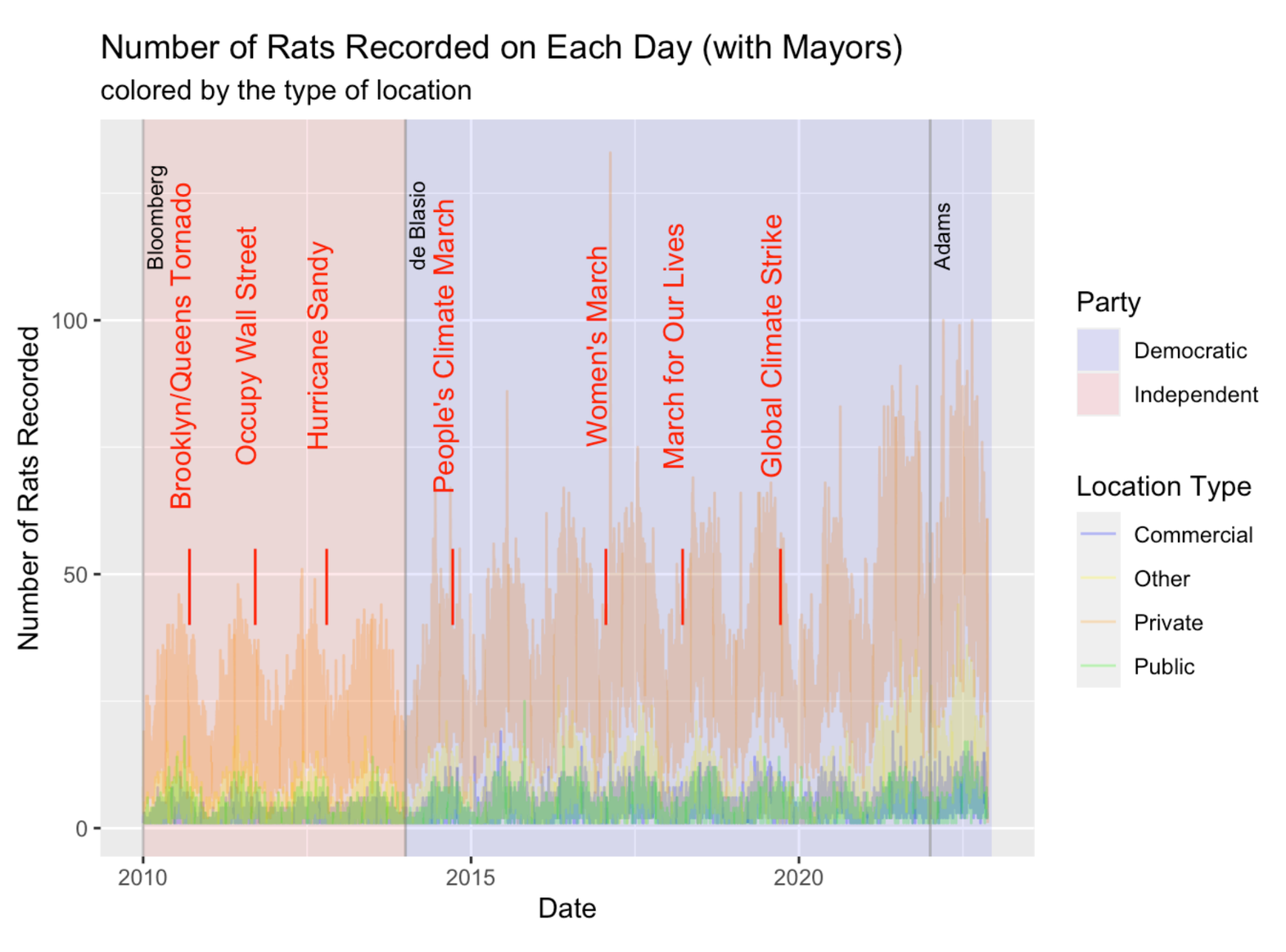
To get a better sense of how the rats in New York City are distributed geographically, we decided to make a choropleth map that showcases the densities of rats by more specific subsections of the city. Upon looking at our data, we realized that each rat sighting was tagged with a zip code, so we created the following choropleth map that shows how many rats in our rat dataset were spotted in a given zip code region.



This map tells us some key information about New York City's rats — we see that a lot of rats are found in the center of the city (specifically, in the middle of Brooklyn) as well as in the northern part of Manhattan. Again, we notice that Staten Island is relatively dark. An interesting section of the map is upper Manhattan, as we notice that a few zip code regions to on the upper west side of Central Park (which is the rectangular cutout in Manhattan) have significantly higher numbers of rat sightings compared to lower Manhattan. The Upper West Side is known for its affluence, which may seem surprising considering the number of rat sightings, but a possible explanation is that official measures to mitigate rat problems in Manhattan have been more focused on the lower side of the borough, where most of the activity is.

Big Events Time Series

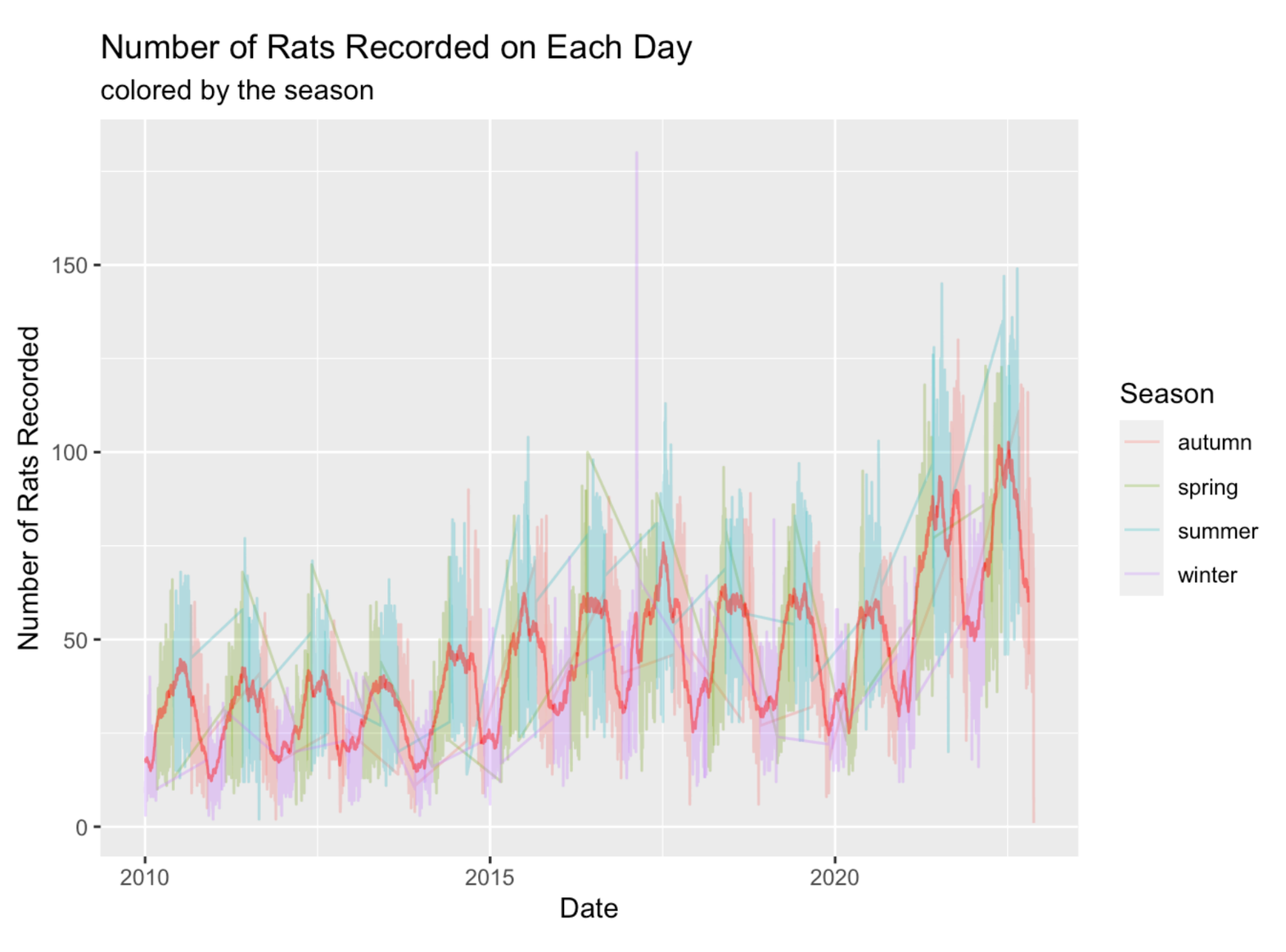
We hoped to garner a further understanding of what causes upticks in rat population, or at least rat spotting. We plotted a time series map of the progression of total rat population in our dataset from 2010, when the first rat was recorded, to present. We also added information which could be correlated potentially with changes in trend of the rat population or a sudden uptick in rats. This includes major events in New York that could be connected to sanitary conditions, number of people outside, or policy changes such as storms, protests, and changes in mayor. We also thought to explore the relative locations for which the rats were spotted to understand the locations in which the rats lived as well as to know if some events caused increases in some domains of the rats.


Number of Rats Recorded on Each Day (with Mayors)
colored by the type of location
Number of Rats Recorded
100
50
0
2010
2015
2020
Date
Bloomberg
Brooklyn/Queens Tornado
Occupy Wall Street
Hurricane Sandy
de Blasio
People's Climate March
Women's March
March for Our Lives
Global Climate Strike
Adams
Party
Democratic
Independent
Location Type
Commercial
Other
Private
Public

We notice a clear increase over time of the rat population. The most evident trend can be correlated with the strong seasonal changes of the New York area which we dissect further in the next section. With regards to the relative location, we observe large number of rats spotted in private areas followed by public and then other, with the fewest rats being spotted in commercial areas. We infer that this is likely due to the fact that many rat spotting is likely connected to building complaints as well as calling of exterminators which is probably most common for private residences to do. Another large reason that this trend may be observed is that a large part of New York is residential, considerably more than any of the other sectors. Among significant events we observed very few major deviations in rat population after events. Several of the protests (Occupy Wall Street, People's Climate March and the Global Climate Strike) as well as the Brooklyn/Queens Tornado seem to be related to slight upticks in rat population, but without further analysis we cannot state whether this is statistically significant. The most noticeable, sudden increase in rat population follows the Women's March protest in 2017, but we again cannot draw any conclusions and must follow this up with statistical tests to draw any conclusions.

Seasons Time Series

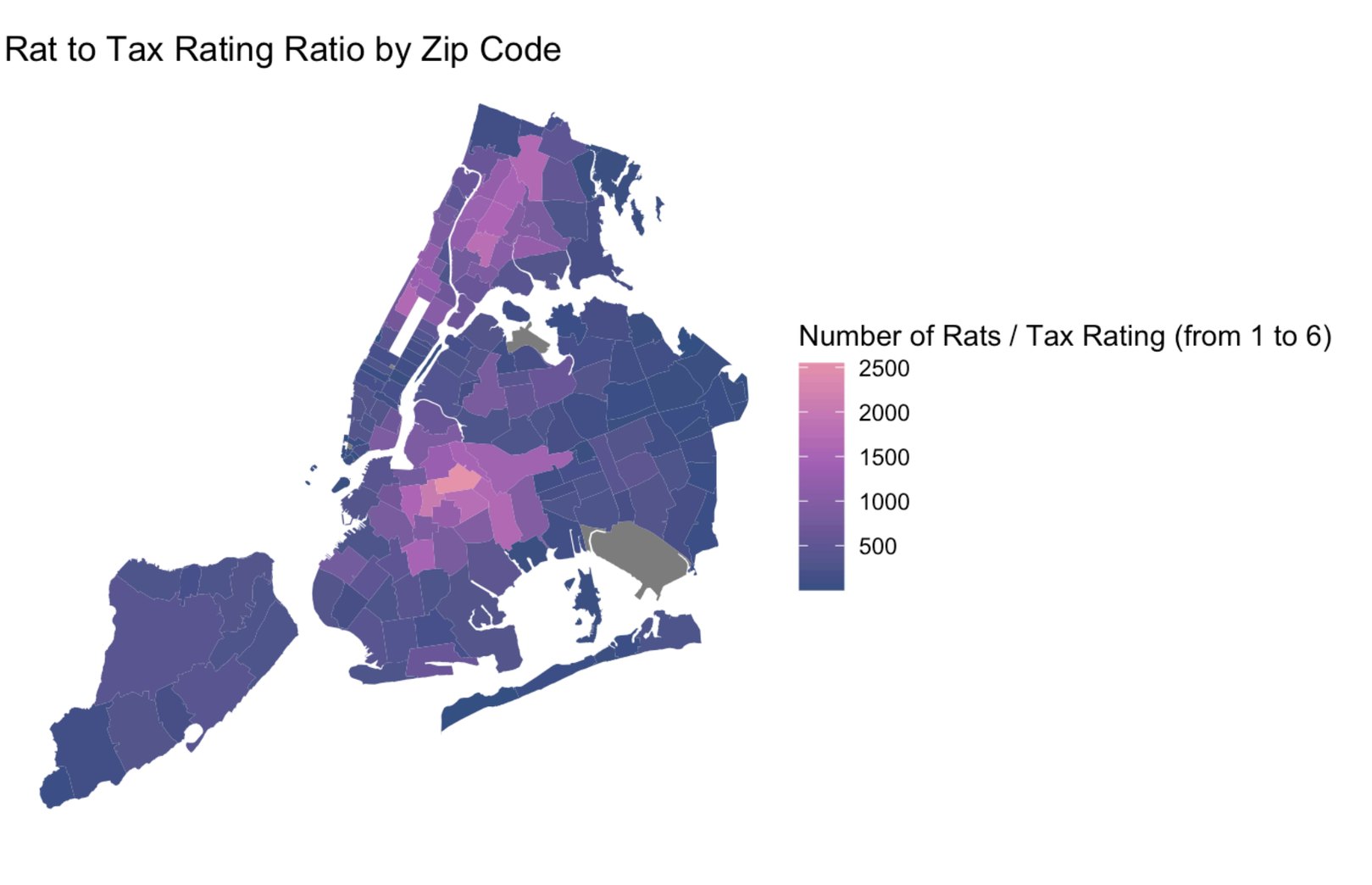
Building on this time series analysis, we now turn to a seasonal approach to modeling rat sightings over time, hoping to further address our research question of how temporal factors impact rat sighting counts.



The above graph plots the moving average for the number of rats seen each month in the red line in order to track the trends, as well as the actual observed number of rats per day. Furthermore, we colored the observed rats by the season in which it was observed, and found a harmonic pattern - there would always be a lot of rats observed in the summer, and not many rats observed in the winter (except for one fateful day in 2017!). This could reflect a few things - rats don't like the cold and tend to stay inside, so they are less likely to be seen. However, humans don't like cold either, so they are less likely to go outside and observe rats in New York. Overall, it is interesting to note the changes in rat observations each season.

Tax Rating Choropleth Map

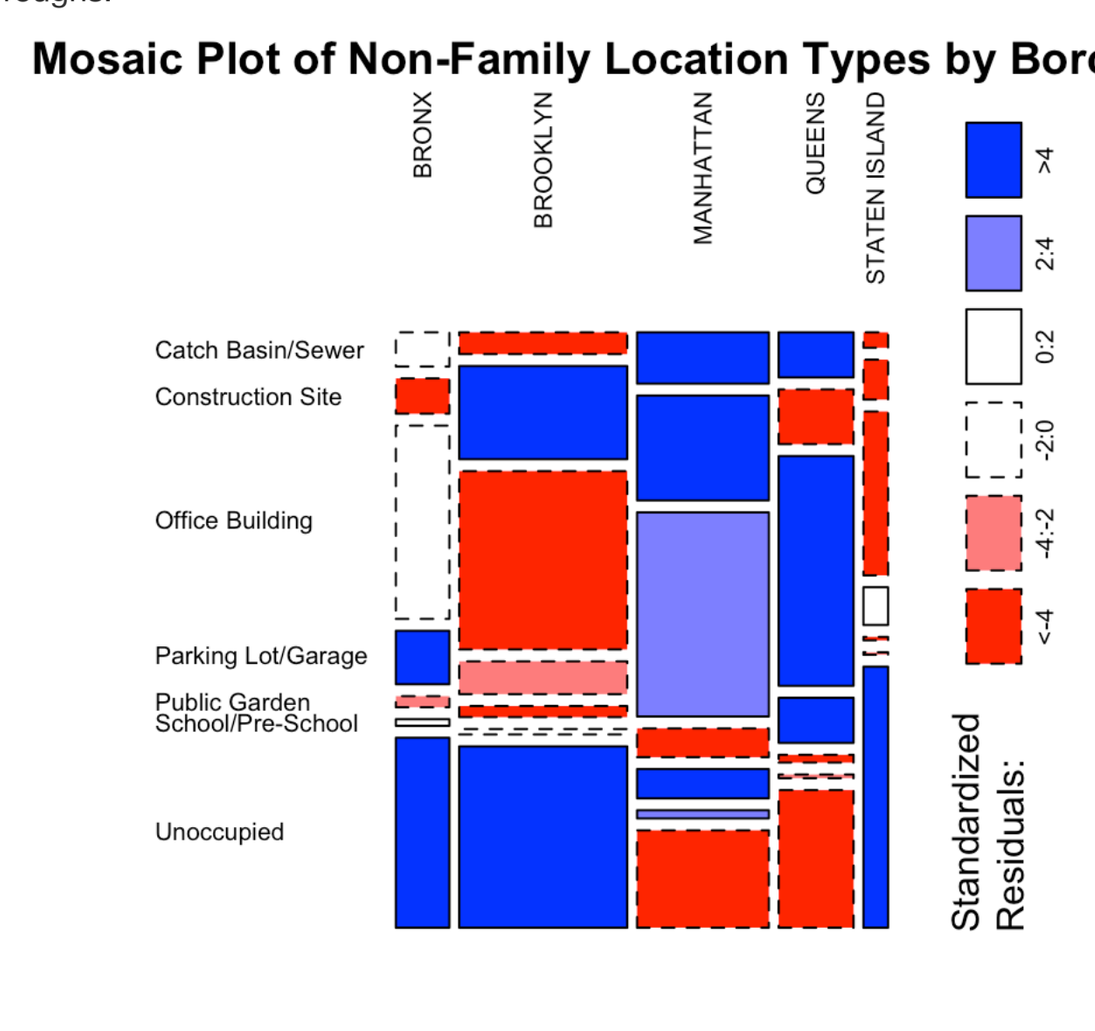
To answer our question about wealth and rat sightings, we turn to the tax return dataset. In order to see the relationship between the affluence of residents and number of rat sightings in each zip code area, we turn to choropleth maps again. This time, we first establish a tax rating system, which is a weighted average of all the tax returns in a given zip code area. The weighted average is calculated by multiplying each individual by their tax "rating" (1 meaning they belong in the lowest bracket and 6 meaning the highest bracket), and then taking the average of all the tax ratings in each zip code area. Thus, this value ranges from 1 to 6, with 1 meaning that everyone from that area is in the lowest tax bracket and 6 meaning the same for the highest tax bracket. To relate this value to rats, we will display the ratio between tax rating and number of rat sightings for each region.



This map has some clear differences when we compare it to the rat sightings choropleth map. First, the Upper West Side is no longer as extreme of a value, as we know that that area is very affluent (which means a smaller rat-to-tax rating ratio). However, we notice that Brooklyn still looks very similar to the original choropleth map, which may suggest a high amount of rats and a lower tax rating in these areas. As we know that Brooklyn is less affluent than Manhattan in general, a plausible explanation for the cluster of higher rat-to-tax rating ratios in the center of the city may be that there are more rats in this area compared to other areas with higher tax ratings (or more affluence).

Location Type Mosaic Plot

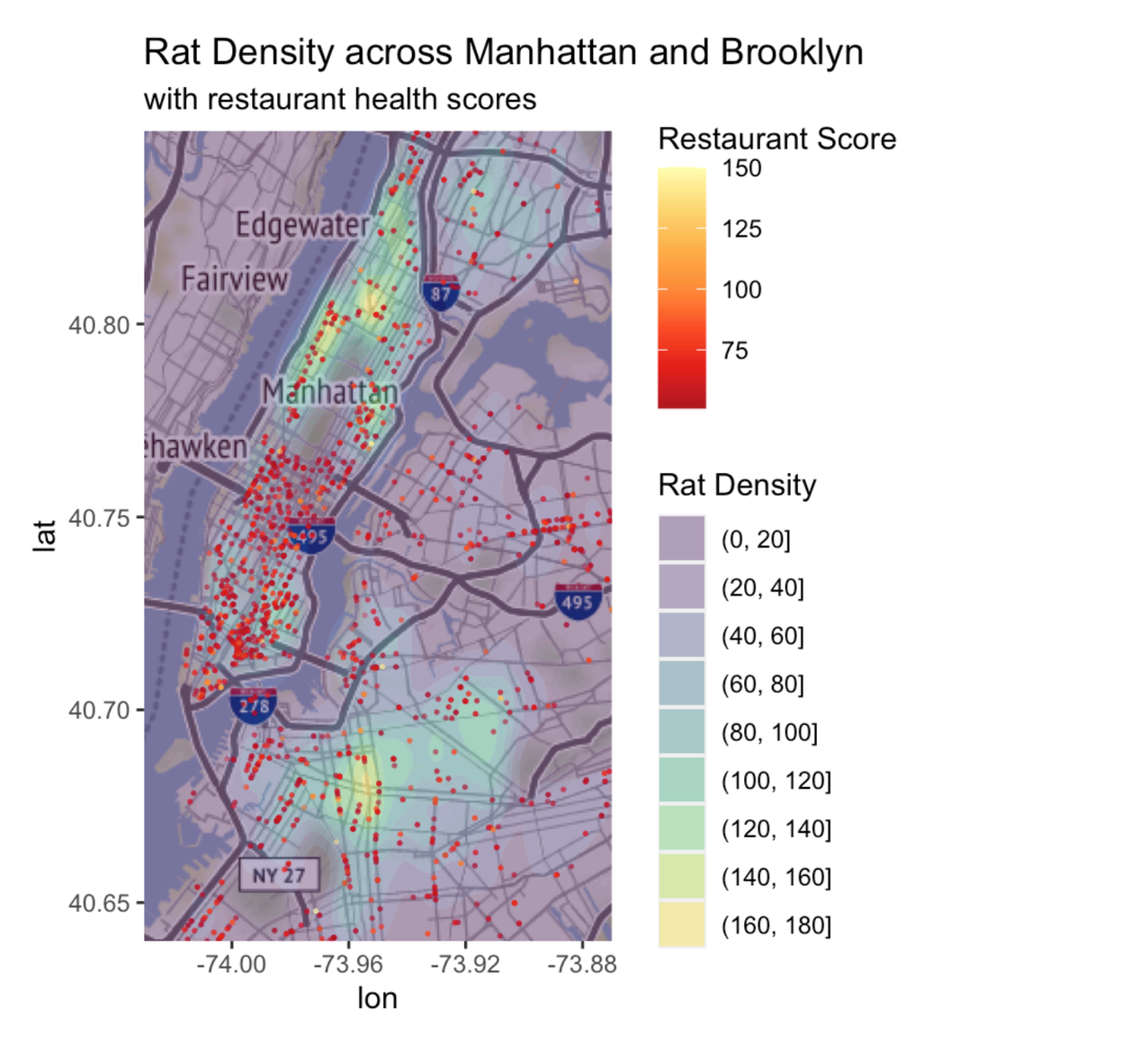
With some temporal and time series analysis of rat sightings done, we now turn to analyzing the conditional distribution of rat sighting locations given borough in an effort to address the degree to which this property of a given rat sighting differs between boroughs.


Mosaic Plot of Non-Family Location Types by Borough
BRONX
BROOKLYN
MANHATTAN
QUEENS
STATEN ISLAND
Catch Basin/Sewer
Construction Site
Office Building
Parking Lot/Garage
Public Garden
School/Pre-School
Unoccupied
>4
2:4
0:2
-2:0
-4:-2
<-4
Standardized Residuals:

This mosaic plot visualizes the conditional distribution of reporting sites of rats given borough. Based on this visualization, we see many statistically significantly high and low combinations of borough and reporting site under the assumption of independence between the two variables plotted. It is interesting to note the way rat reporting sites reflect the distinctive landscapes of each borough. For instance, we have significant evidence that Manhattan has higher proportions of rat sightings made at office buildings and construction sites than would be expected under independence, which reflects Manhattan's reputation as a bustling metropolis with many developed and developing commercial construction projects. It is also interesting to note that the statistical significance of the proportion of reports made in Unoccupied sites (which we categorized as reports made in either Vacant Buildings or Vacant Lots) for every single borough; the high proportion of such sightings in Staten Island, Brooklyn, and the Bronx may suggest the presence of pockets of high poverty or low economic development in these boroughs, and the significantly low proportion of sightings in Unoccupied regions may suggest a relatively high degree of property and economic development in these boroughs, where fewer spaces are left unused by homeowners or businesses. In all, reveals that rats are generally found in very different sets of locations in different boroughs.

Restaurant Heat Map

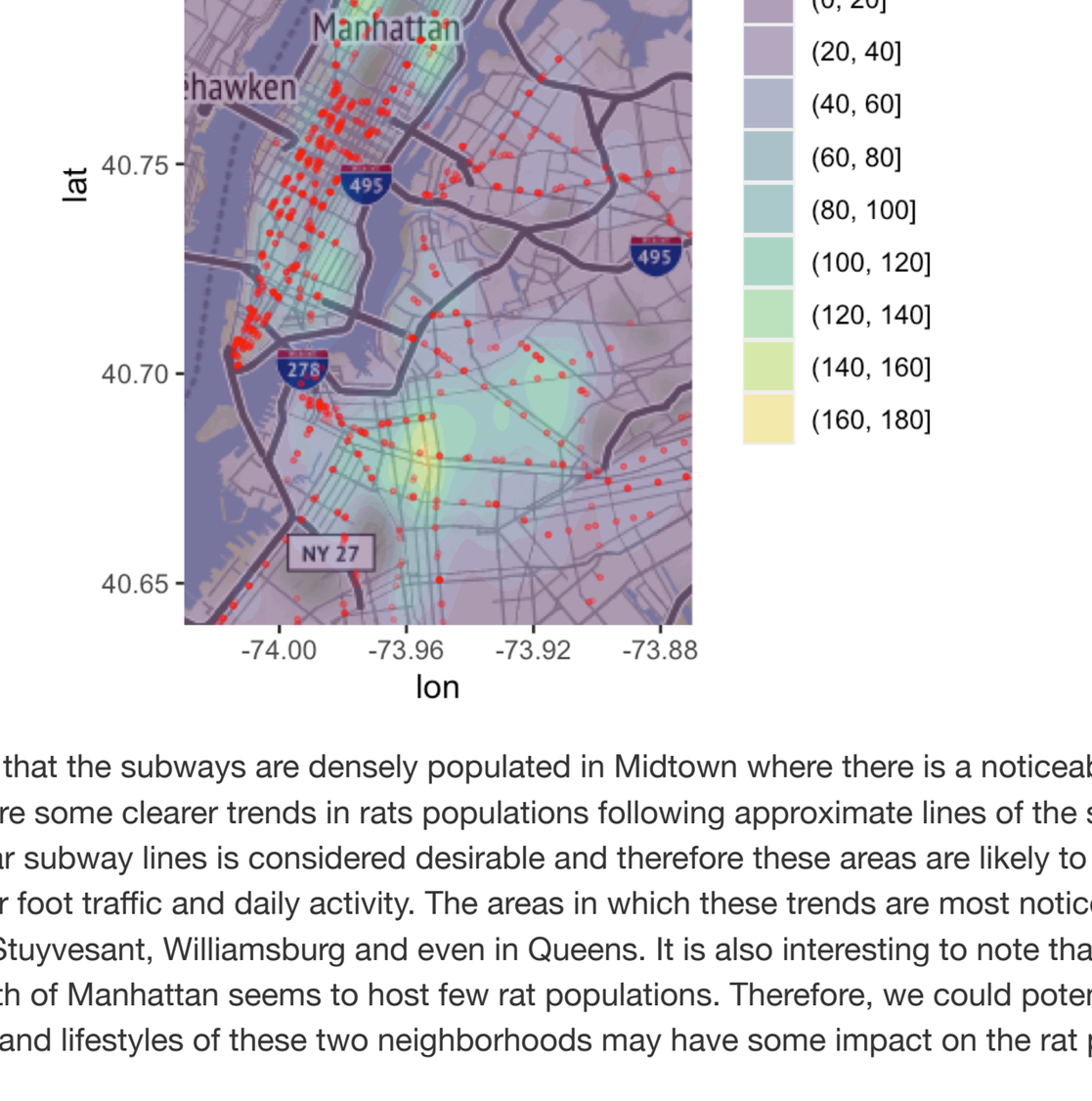
We decided to see if there were any evident relationships between geographic artifacts such as restaurants in New York with the density of the rat population. Our intuition was that an increases in restaurants would increase the number of food sources which would have an immediate impact on rat population. After initial results showing inconsequential results we decided to test whether restaurants that received a poor score on their health examination (arbitrarily 50+ scores, with 0 being a perfect score) had a relationship with rats. We speculated that that worse sanitary conditions of the restaurants may also share some relationship with rat population.


Rat Density across Manhattan and Brooklyn
with restaurant health scores
lat
40.80
40.75
40.70
40.65
-74.00
-73.96
-73.92
-73.88
lon
Edgewater
Fairview
Manhattan
87
495
278
NY 27
Restaurant Score
150
125
100
75
Rat Density
(0, 20]
(20, 40]
(40, 60]
(60, 80]
(80, 100]
(100, 120]
(120, 140]
(140, 160]
(160, 180]

We observe several intersting trends in this graph. We notice that the very dense area of Midtown which hosts most of the restaurants and most of the restaurants with bad health scores also contains very few rats. This in and of itself is worth further investigation. It is worth noting that Combs et. al. *https://www.theatlantic.com/science/archive/2017/11/rats-of-new-york/546959/* observed similar results in their genetic study of rats in New York, with Midtown serving as a natural genetic and physical boundary for the rat populations. Otherwise, the density of the rat populations as well as the restaurants seem to have little in common aside from some small areas. This is counter-intuitive, but supports the idea that rat population may be driven more greatly by other factors such as potentially human movement or human density.

Subway Heat Map

We thought to observe the relationship between human movement and rat population density. We decided to see if we could notice a relationship between subway entrances, a location comically famously associated with rat spotting, and rat density. While subway entrances provide a loose map of where people move and where people live it is important to note that because the New York subway map is designed similarly to most United States rail and public transport systems.The lines take people from the suburbs and the other boroughs into Manhattan to support the idea of a work day commute with people living in the outskirts of the city and traveling into the city for work.



We observe again that the subways are densely populated in Midtown where there is a noticeable absence of rats. However, we do notice that there are some clearer trends in rats populations following approximate lines of the subway. We surmise that this is because living near subway lines is considered desirable and therefore these areas are likely to be more densely populated and also have much heavier foot traffic and daily activity. The areas in which these trends are most noticeable are in the Upper West Side, Harlem, Bedford-Stuyvesant, Williamsburg and even in Queens. It is also interesting to note that, similarly to Midtown, the Financial District in the South of Manhattan seems to host few rat populations. Therefore, we could potentially suggest that the similar industrial designs and lifestyles of these two neighborhoods may have some impact on the rat populations and density in these areas.

Conclusions and Future Work

Through this analysis, we have learned a multitude of interesting things about the conditional distribution of rats in New York City given such variables as geography, temporal events, and physical landmarks. Clearly the distribution of rats in the city correlates highly with many of our tested variables, and displays significant geographic and temporal activity. It seems that the quantity of rat sightings differ greatly between boroughs, zip codes within boroughs, and even specific types of locations within different boroughs. Furthermore, rat sightings display a significant trend and seasonal over time, all the while responding to major events that occur in the city. Future analysis of this topic would do well to analyze a) different datasets that could potentially be compared to rat sighting distributions such as racial or age-related data in order to assess how people of different social groups experience varying levels of rats in their homes, and/or b) dive deeper into the auxiliary variables which we had already selected; for example, correcting for geographic area in our borough and zip code data in order to calculate and visualize how the rats per square mile (and by extension, variables involving rat sighing counts such as rat sighting density to tax rating ratio) changes between geographic regions. In all, this project provided a thoughtful insight into life in New York City from the perspective of its most mainstay citizens - the rats.