Manrattan - A Look into NYC Rats

Arthur Jakobsson, Alex Cheng, Liz Chu, Kevin Ren

November 18, 2022

Contents

Rat Population Data Analysis - New York City **1**

Descriptions of Datasets 1

Research Questions 2

Graphs Made 2

Conclusions and Future Work 9

Rat Population Data Analysis - New York City

Descriptions of Datasets

Rat 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:

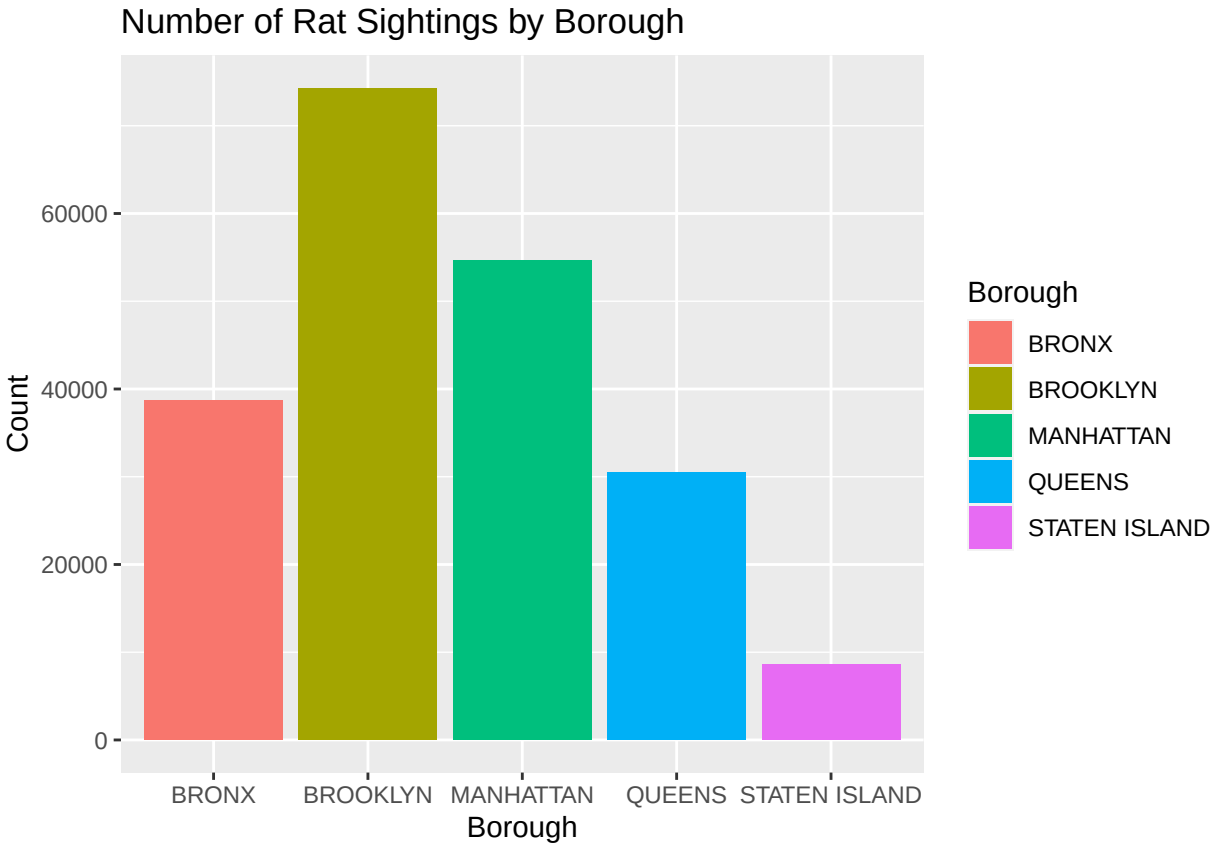
1. How do rat sightings differ geographically and by borough?
2. How has the number of rats reported changed over time?
3. How does well do wealth and geographic data combined correlate with rat sightings?
4. How do rat sightings correlate with candidate features such as subways and restaurants?

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

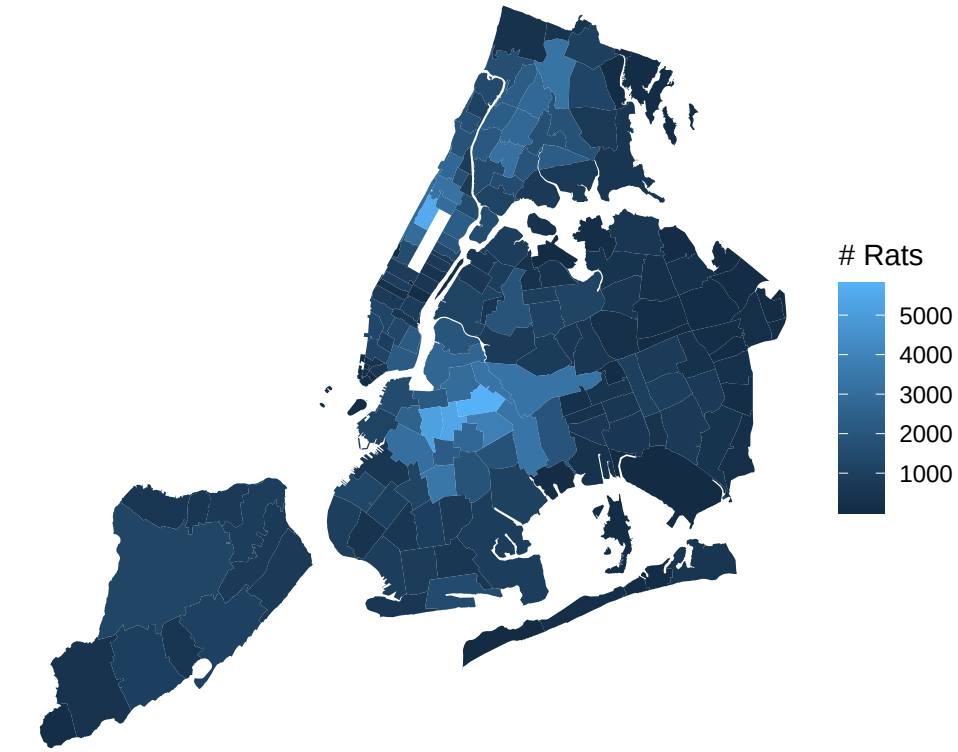
borough bar chart

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.

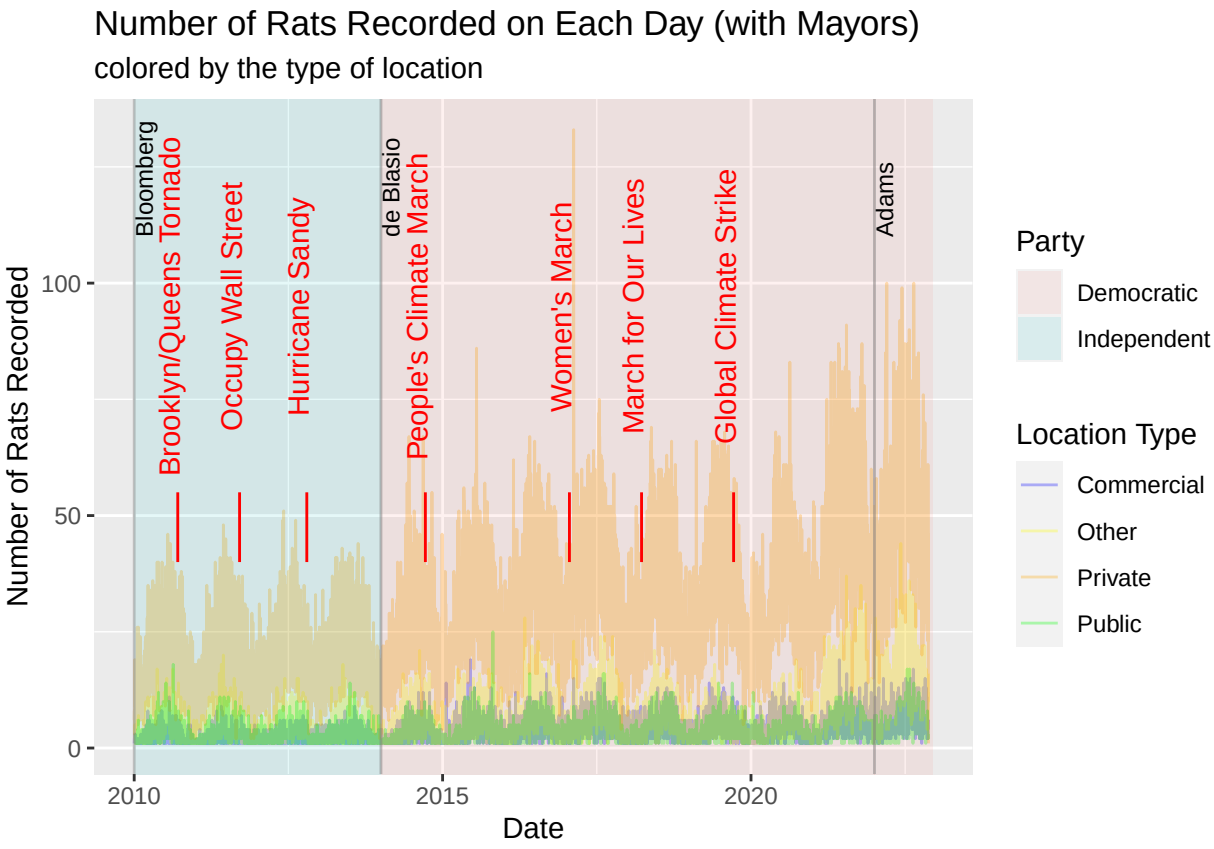


This bar chart displays the number of rats seen within each borough in New York City. Brooklyn had by far the most rat sightings at 74,302, followed by Manhattan, the Bronx, Queens, and Staten Island, in that order. The low number of rat sightings in Staten Island might reflect its more cut-off nature 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 map

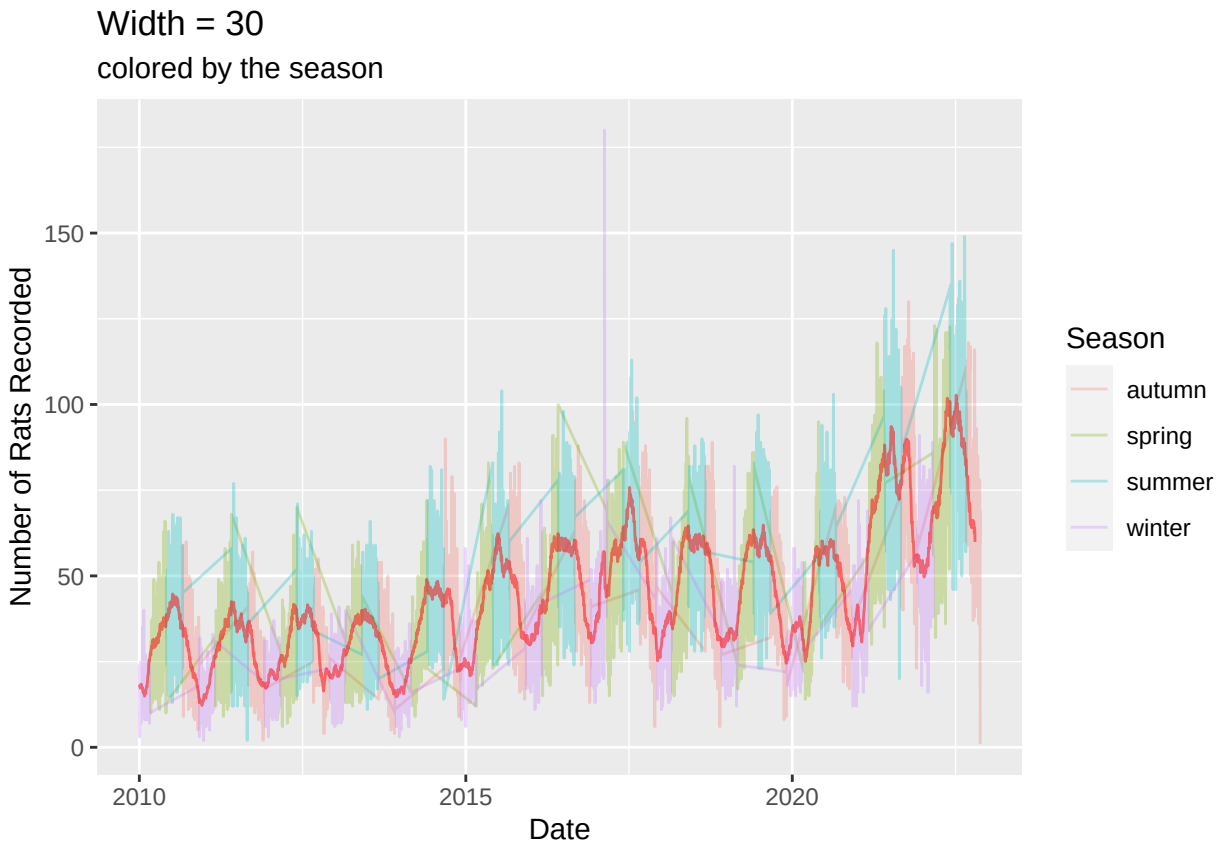


big events time series



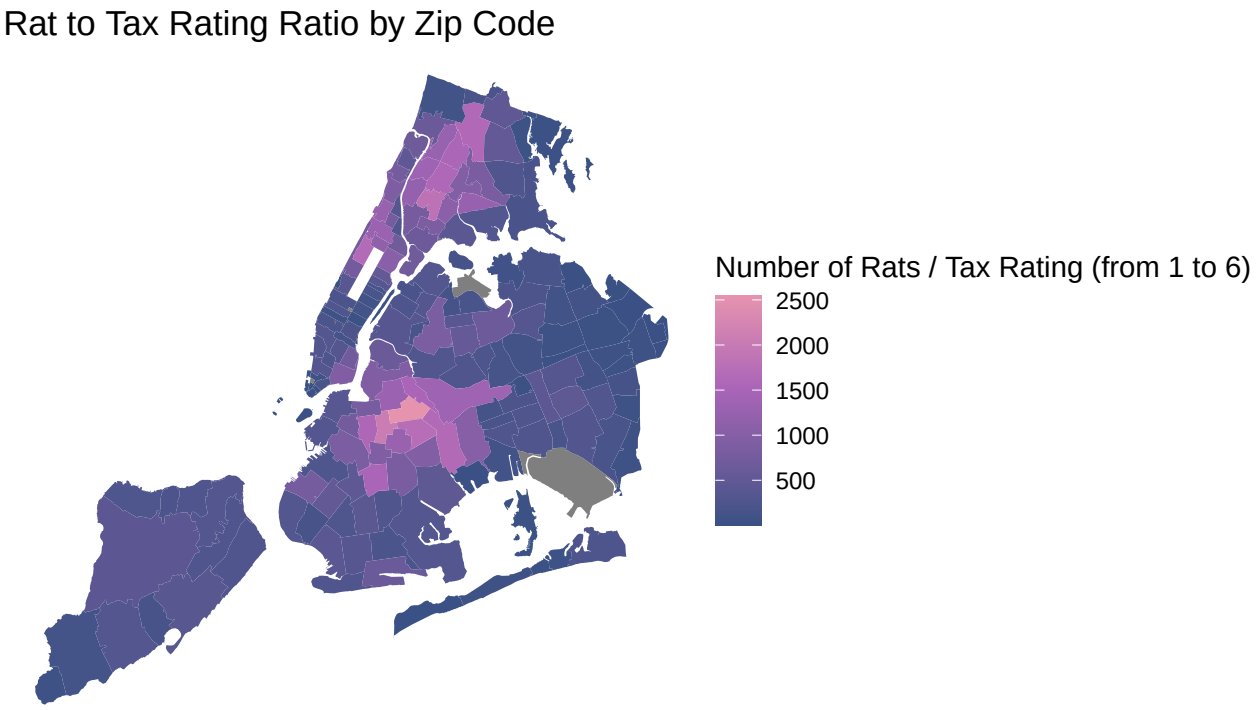
seasons time series

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.



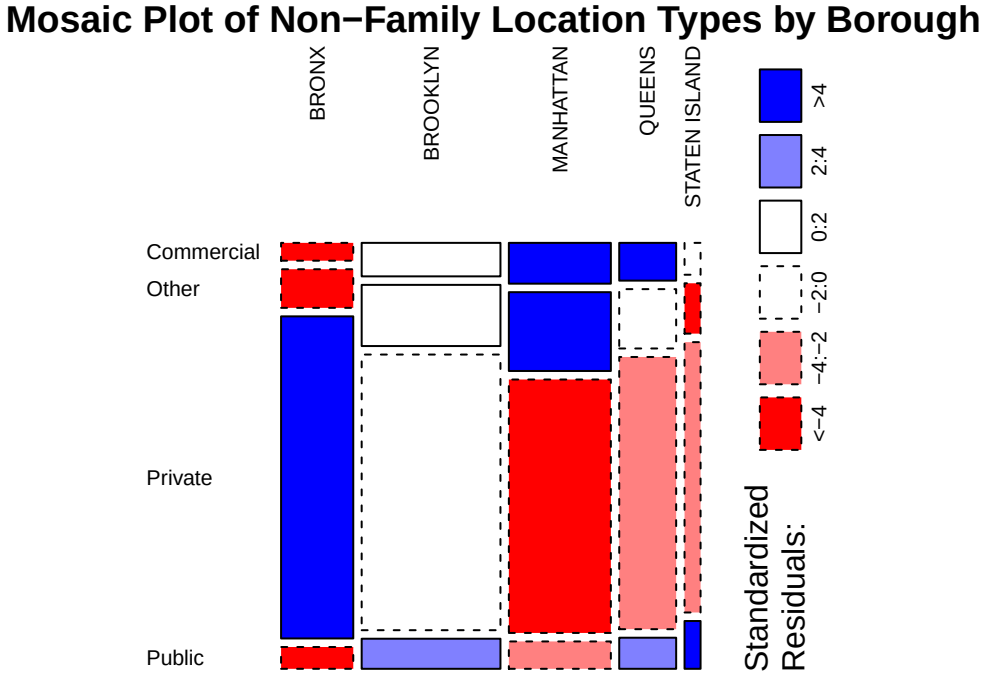
The above graph plots the moving average for the number of rats seen each month in the red line in order to track the trends, as well as the actual observed number of rats per day. Furthermore, we colored the observed rats by the season in which it was observed, and found a harmonic pattern - there would 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 map



mosaic plot

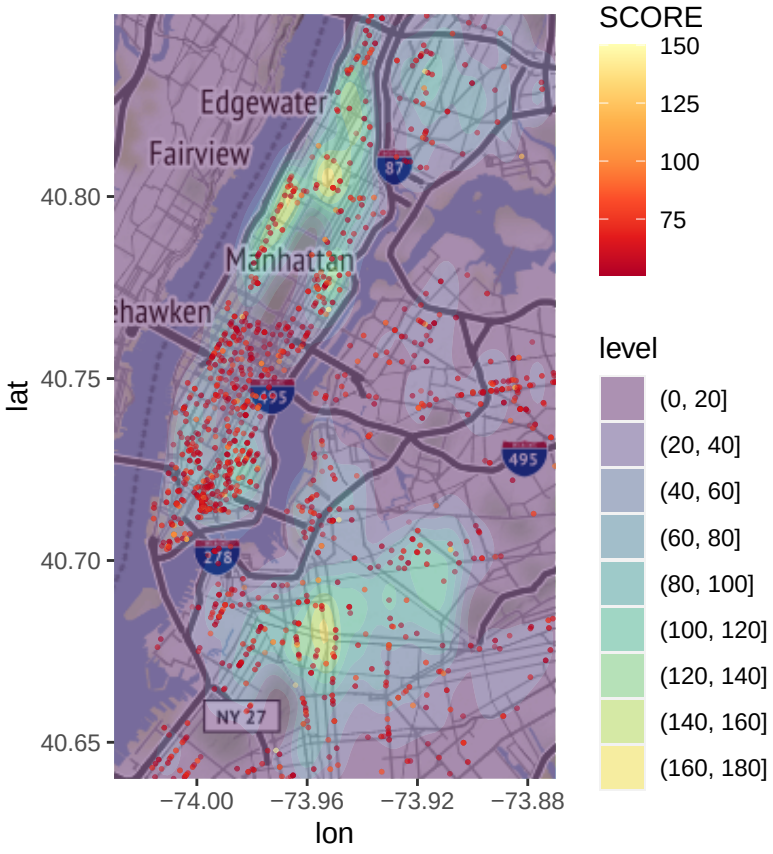
With some temporal and time series analysis of rat sightings done, we now turn to analyzing the conditional distribution of rat sighting locations given borough in an effort to address the degree to which this property of a given rat sighting differs between boroughs.



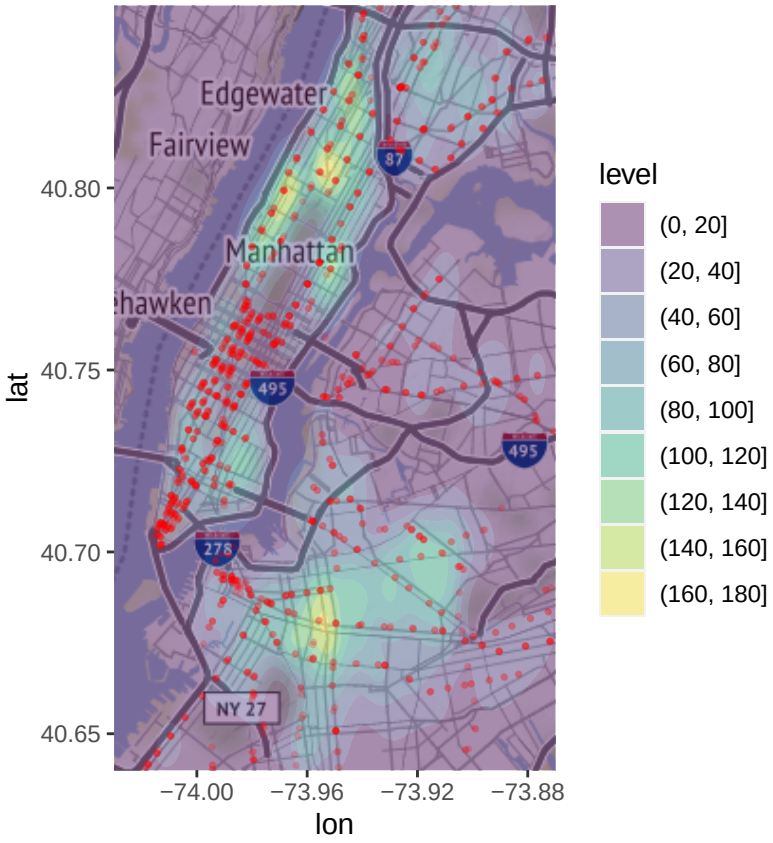
This mosaic plot visualizes the conditional distribution of reporting sites of rats given borough. Based on this

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 reflect's 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 map



subway map



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 quanitity of rat sightings differ greatly between boroughs, zip codes within boroughs, and even specific types of locations within different boroughs. Futhermore, 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.