Project 1

Group Member: Peter Fu Chen, Mike Liu

07 March, 2023

Part 1: Visualizing New York City's Rat Problem

Background and Motivation

Source of data: NYC Rat Sightings

It is widely known that rats can pose significant risks to both human health and property. They are carriers of deadly viruses and have the potential to cause fires by gnawing on electrical wires. If you are a resident of New York, it's natural to be concerned about the presence of rats near your house.

The New York City rat dataset is collected and maintained by the City of New York via their 311 portal. It provides information on the location, type, and date of rat sightings, as well as the actions taken to address the issue. The dataset is useful for us to analyze, including mapping rat populations, identifying hotspots, and studying the effectiveness of pest control efforts.

By analyzing the New York City's database of rat sightings, we hope to gain insights into the distribution of rats across different boroughs and types of buildings. This information can then be used to inform the development of effective pest control measures to address the rat infestation problem in the city.

The motivation for this project comes from the need to address the negative impact that rats can have on public health and the environment. By using data analytics techniques to analyze the New York City's database of rat sightings, we hope to provide valuable insights into the distribution of rats in the city and to support the development of effective pest control measures. Through this project, we aim to make a meaningful contribution to the efforts to manage the rat population in New York City and to improve public health and well-being.

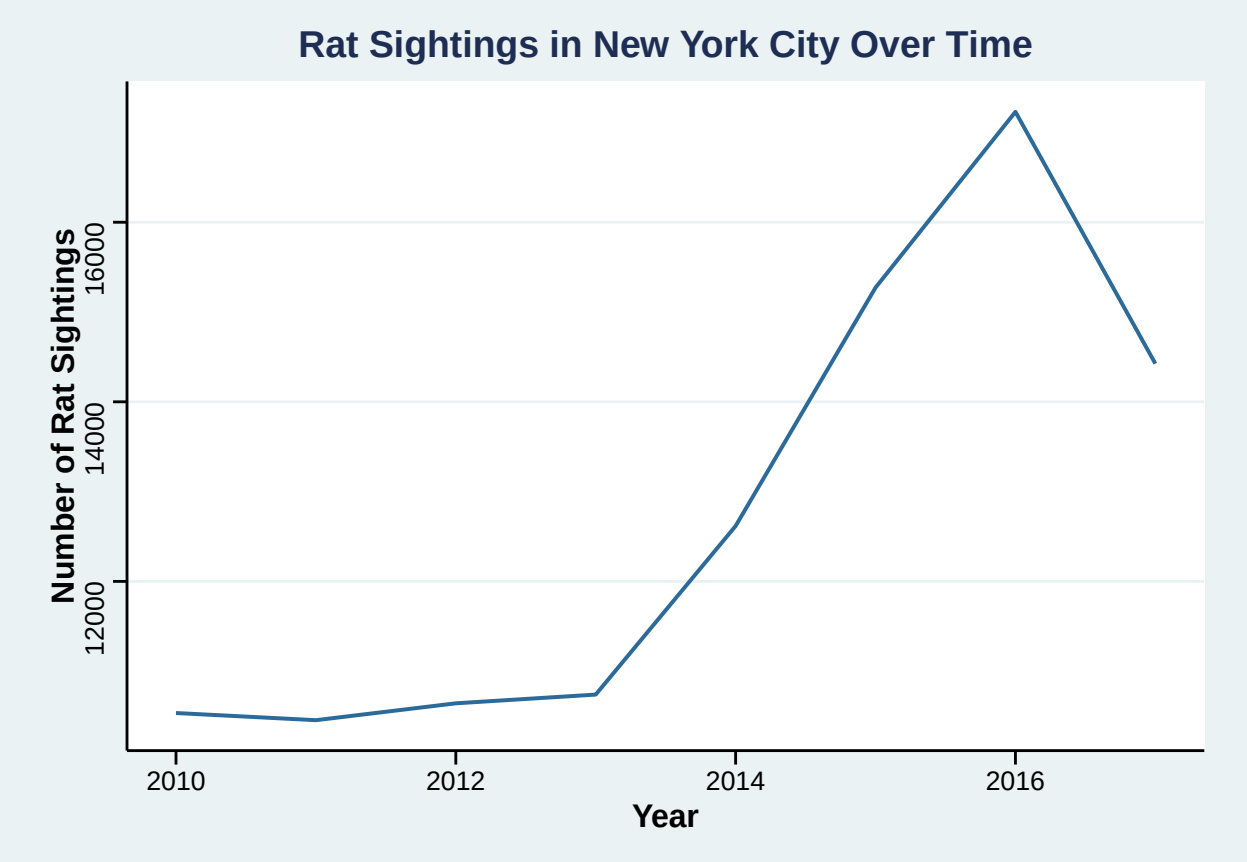
But the question remains, how can we determine where these rats are hiding?

Methodology

Rat Sightings in New York City Over Time

Before answering the above questions, let's take a look at the development history of rat disasters. We analyzed the rat sightings data from 2010 to 2017.

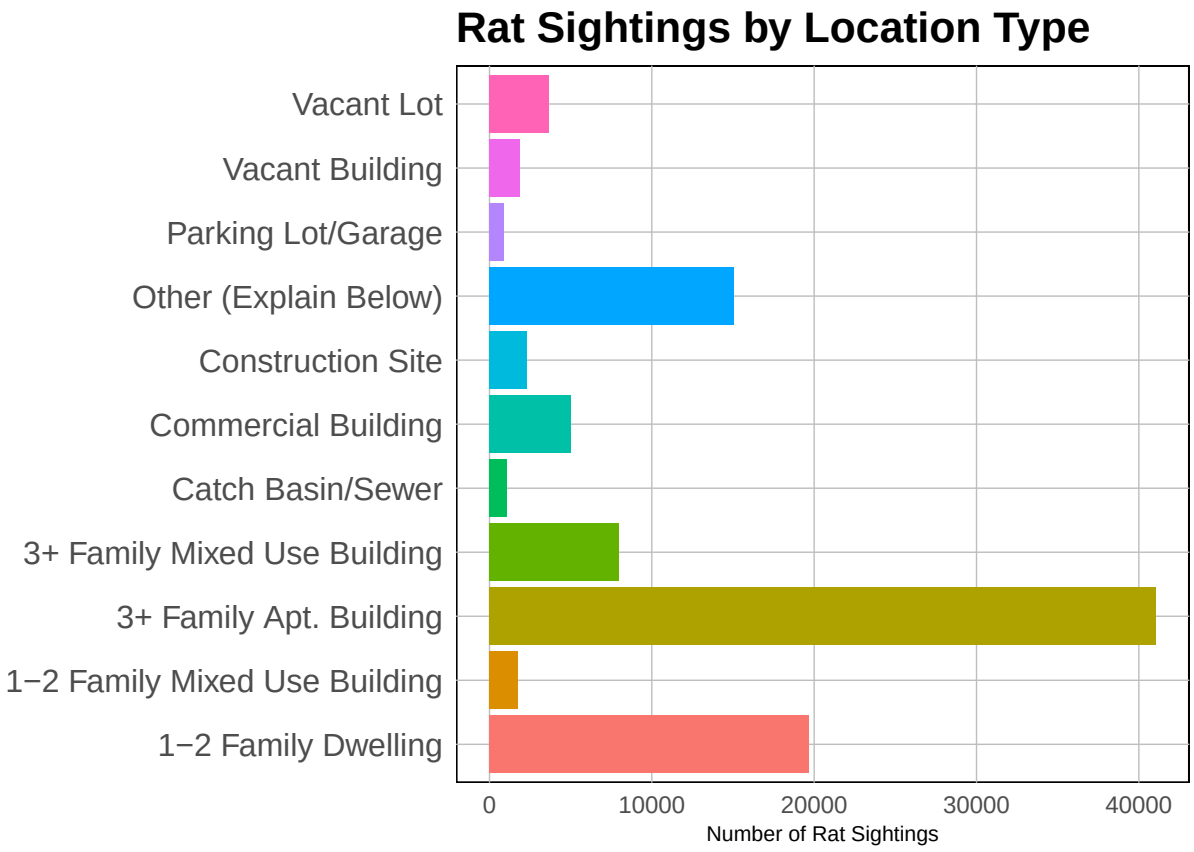
The analysis revealed that there was a significant increase in the number of rat sightings starting from 2013, with the number of sightings reaching a peak in 2016. This trend highlights the need for continued monitoring and implementation of effective pest control measures to address the issue of rat populations in the city.



Rat Sightings by Location Type

We are also concerning what type of buildings had the most sightings?

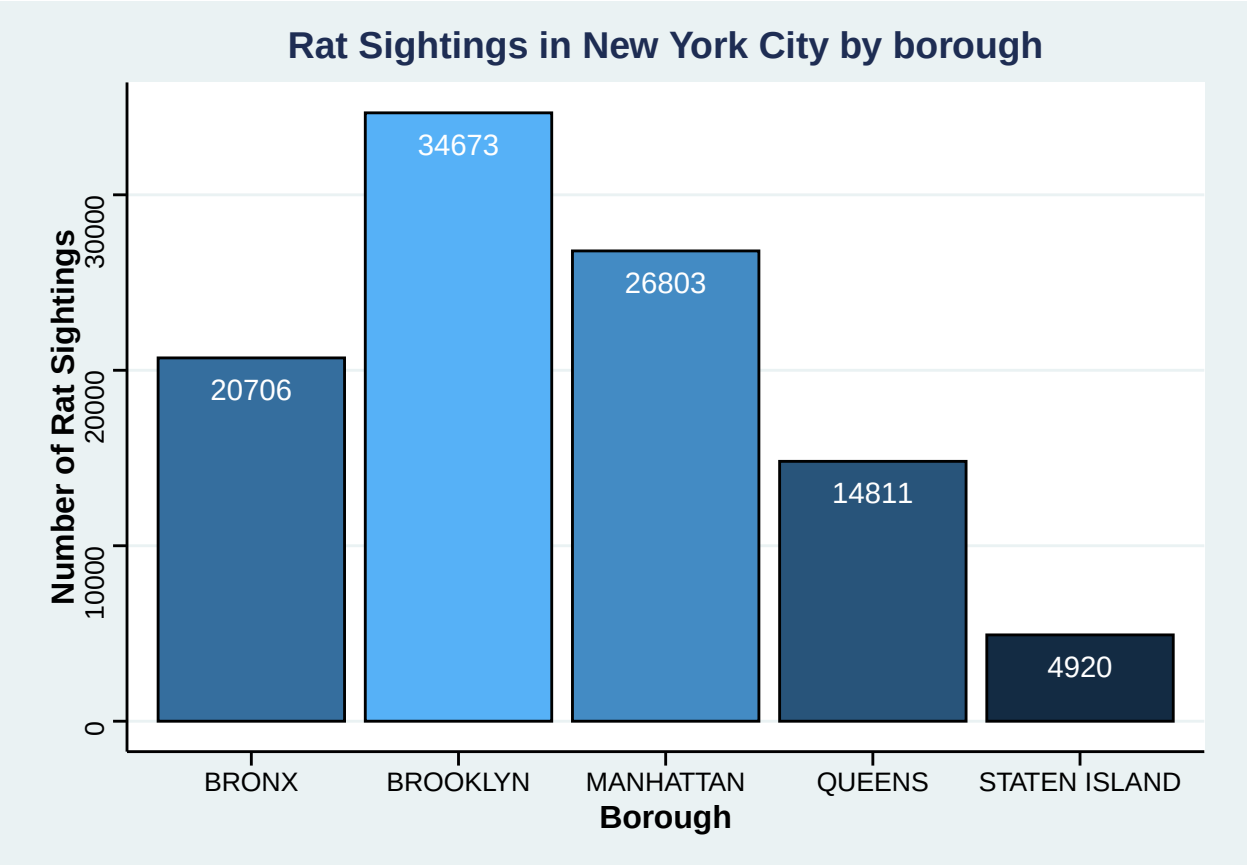
Based on the analysis, it appears that the type of building with the most rat sightings is 3+ family buildings. This conclusion was drawn from the plot shown below that the relationship between the location type and the number of rat sightings. By visualizing the data in this way, we were able to identify which type of building had the highest number of rat sightings and better understand the distribution of rats in the city.



Rat Sightings in New York City by borough

Next, we divide by borough, and look at the growth history of rats in different borough in recent years.

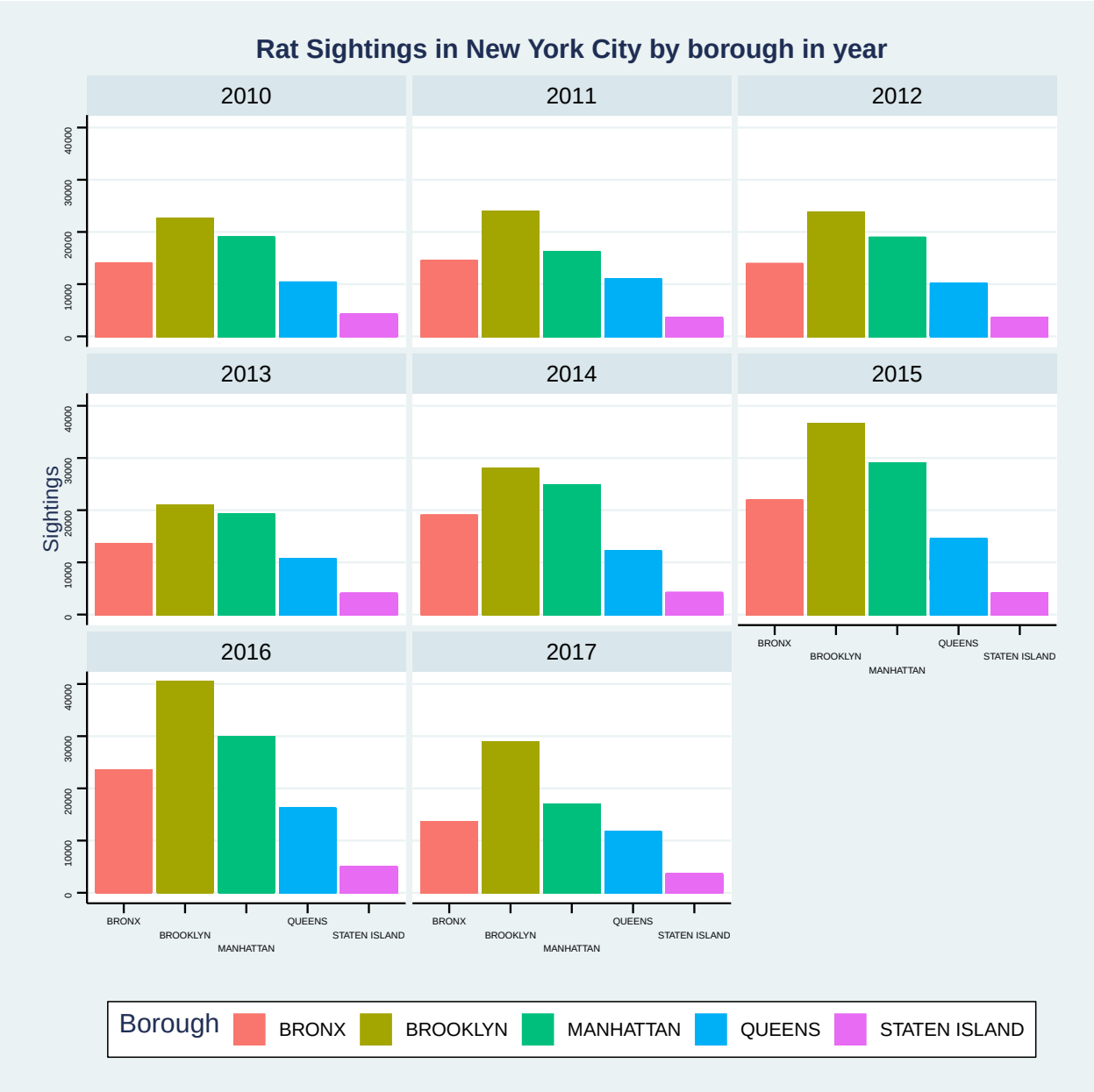
The analysis of the rat sightings by borough in New York City reveals that Brooklyn has consistently been the borough with the highest number of rat sightings. With a total of 34,673 sightings, Brooklyn outpaces the other boroughs. Manhattan ranks second in terms of rat sightings, according to the bar plot. These findings highlight the need for targeted pest control efforts in these areas in order to effectively manage the rat population.



The analysis of the rat sightings in New York City can also be conducted at the borough level and over time. As seen in the first plot, the number of rat sightings in the city increased exponentially after 2013. To understand which boroughs were contributing to this trend, we analyzed the data by borough and year in the below plot.

The results of this analysis, as shown in the earlier plot, indicate that the growth in rat sightings in New York City was driven by an increase in sightings in Brooklyn. The number of rat sightings in Brooklyn began to rise after 2013 and reached a peak in 2017 before declining. This trend is consistent with the overall trend seen in the first plot, which highlights the relationship between the growth in the rat population and specific boroughs.

These findings demonstrate the importance of understanding the distribution of the rat problem in New York City and the need for targeted pest control efforts in the affected areas. By analyzing the data at both the city-wide and borough levels, we can gain a more complete understanding of the problem and develop more effective strategies for addressing it.



Does policy works?

Borough	Control	Sightings
BRONX	Before	3497
BRONX	After	2791
BROOKLYN	Before	5979
BROOKLYN	After	5493
MANHATTAN	Before	4610
MANHATTAN	After	3326
QUEENS	Before	2410
QUEENS	After	2199
STATEN ISLAND	Before	734
STATEN ISLAND	After	616

We take the report,NYC's Newest Weapon Against the Rats? Sterilization, as reference.

The analysis of the effectiveness of pest control measures is critical for determining the success of efforts to control the rat population in New York City. By comparing the number of rat sightings before and after control measures have been implemented, we can evaluate the impact of these efforts on the rat population.

One measure that has been implemented in New York City is the sterilization of rats, as mentioned in the article "NYC's Newest Weapon Against the Rats? Sterilization". This measure was proposed by Mayor Bill De Blasio and is aimed at reducing the rat population by preventing them from reproducing.

At the start of our analysis, we were concerned about the distribution of rats across different boroughs in New York City. After analyzing the data, we found that Brooklyn has consistently been the borough with the most rat sightings. Additionally, we learned that after 2016, the number of rat sightings decreased significantly.

To quantify the effectiveness of the sterilization measure, we plotted the number of rat sightings before and after the implementation of the measure. The results showed a significant decrease in the number of rat sightings across different boroughs. This indicates that the sterilization measure may have been effective in reducing the rat population in New York City.

Overall, our analysis provides valuable insights into the effectiveness of pest control measures in controlling the rat population in New York City. The results of this analysis can be used to inform future pest control efforts and ensure the continued success in reducing the rat population in the city.

Part 2: Data Hunting

- Data Source: FiveThirtyEight
- The Economic Guide To Picking A College Major

FiveThirtyEight is a data-focused news and analysis website that is known for its use of data and statistical analysis in its reporting. Its data sets are particularly interesting for data analytics because they cover a wide range of topics and are often based on extensive and diverse sources of data. The data sets are also well-organized and well-documented, making it easy for data analysts to work with the data and generate insights.

We have selected the College Majors dataset from FiveThirtyEight for our potential research project. Our objective is to determine which major or category has the most favorable prospects in terms of employment and earnings. Despite the common perception that a college degree is a ticket to financial stability, we believe that this is not necessarily the case. There are numerous factors that influence a person's financial success, and choosing the right major is just one of them. By analyzing the data from the College Majors dataset, we hope to gain insights into which majors are more likely to result in lower unemployment rates and higher earning potential.

- Data Source: BuzzFeed News
- Here And Now: These Maps Show How Climate Change Has Already Transformed The Earth

The BuzzFeed News data sets available on GitHub are particularly interesting for data analytics because they cover a wide range of topics and are often based on unique and under-explored sources of data. For example, BuzzFeed News has published data sets on topics such as political campaign donations, police shootings, and consumer product safety.

For our potential research project, we have chosen the 2019 BuzzFeed News post on the current reality of climate change as our source of data. This dataset contains the foundational information for creating a map that displays temperature trends across the globe. Our objective is to conduct a comprehensive analysis of the evolution of climate change on our planet and make predictions about its future trajectory. The ultimate goal of our project is to provide policy makers with an informed understanding of the urgency of the situation and the need for swift action to implement new policies. We believe that our research will be a valuable resource in the fight against climate change, helping to guide decision-making and prioritize resources in a way that will have the greatest impact.

- Data Source: DataHub
- CIRP Freshmen Survey

Datahub's data collections include data on climate change, economic indicators, and health and wellness. Datahub provides a wide range of tools and resources for working with the data, including data cleaning and preprocessing tools, visualization tools, and data analysis libraries. Overall, the quality and diversity of Datahub's data sets make them an excellent resource for data analytics projects.

For our potential research project, we have selected the CIRP Freshmen Survey from DataHub as our data source. The CIRP Freshmen Survey is a comprehensive study that gathers information from first-year college students from institutions across the United States. The survey covers a wide range of topics, including financial aid, academic major, religious affiliation, intended career path, and much more. This dataset is incredibly rich and offers a wealth of information for analysis. Our aim is to use the data from this survey to gain a deeper understanding of the characteristics and experiences of college students in the United States. With this information, we hope to provide insights that could inform policy decisions at the institutional level, particularly with regards to which groups of students may require additional support and resources. We believe that our research has the potential to make a significant contribution to the field and have a positive impact on the lives of students.