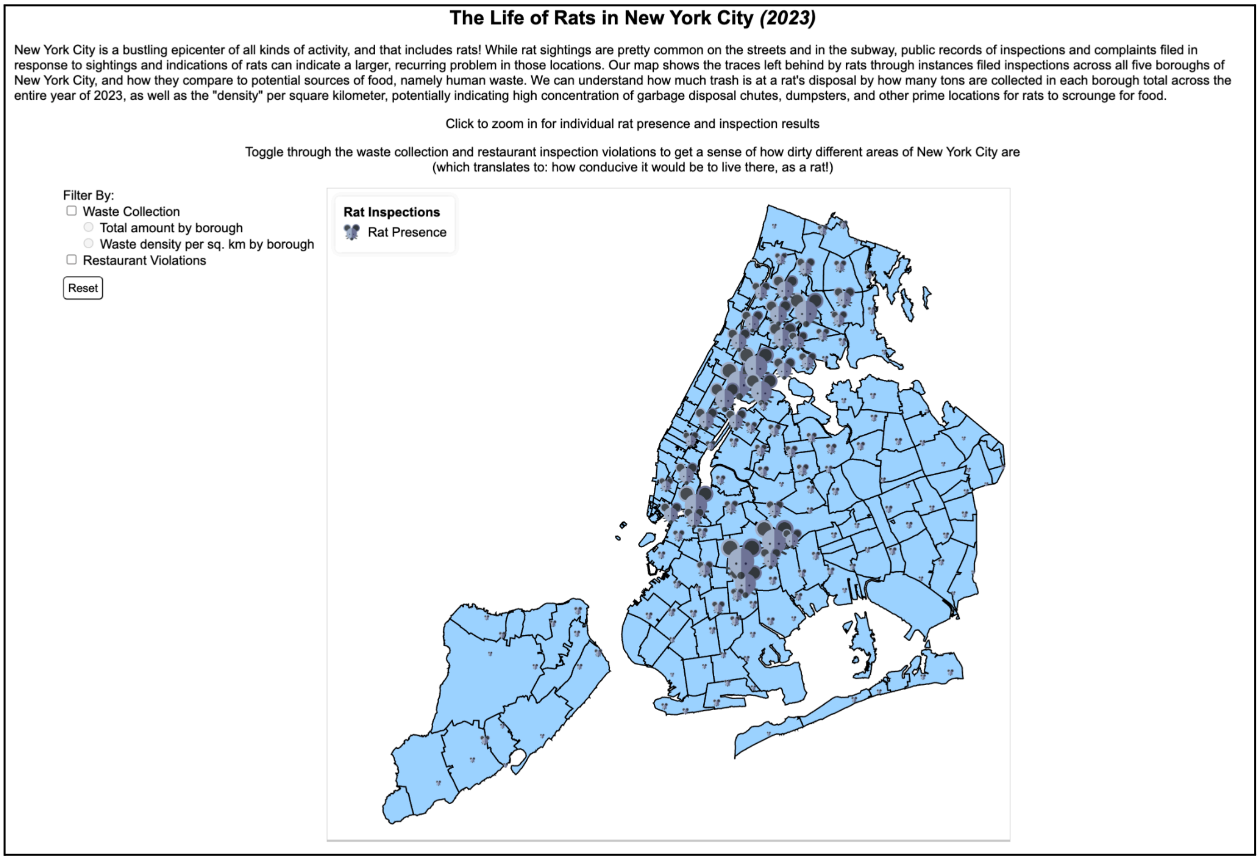
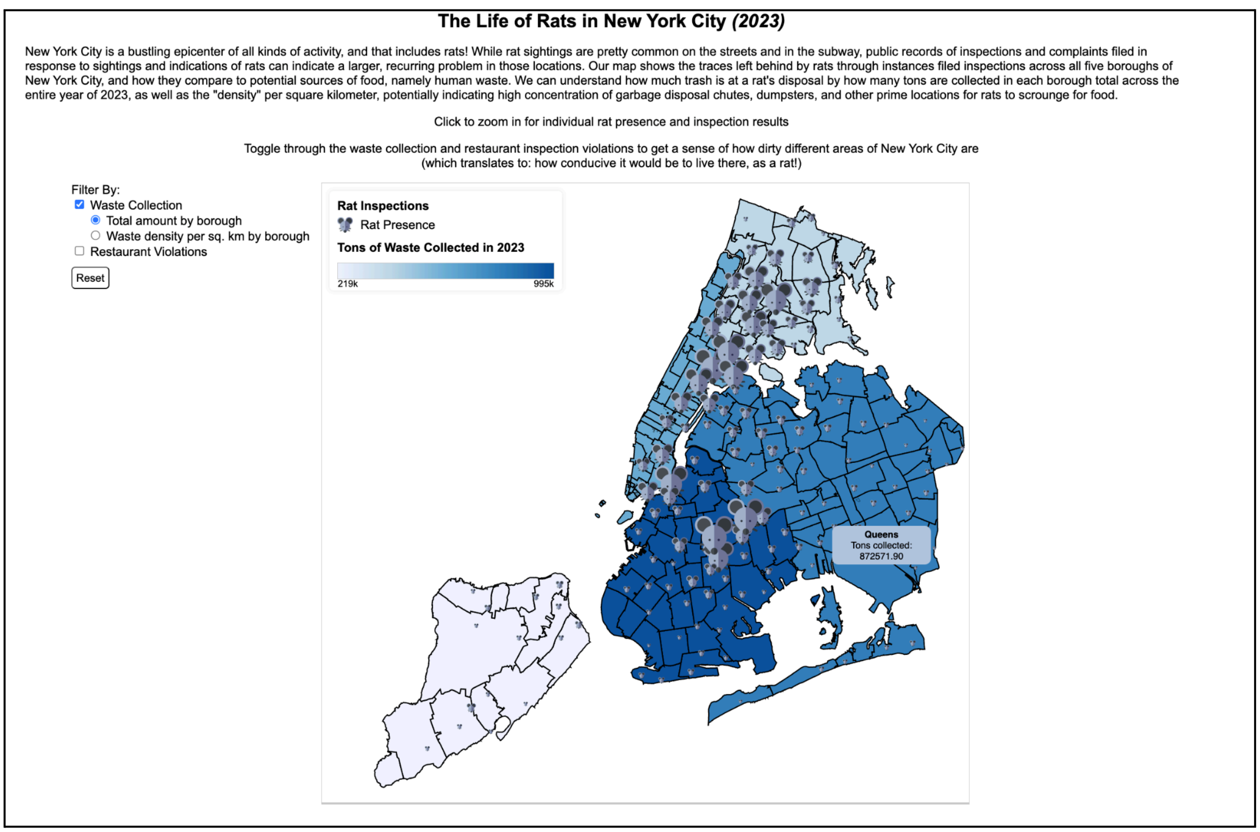
INFO 3300 Project 2 Final Report





The Life of Rats in New York City *(2023)*

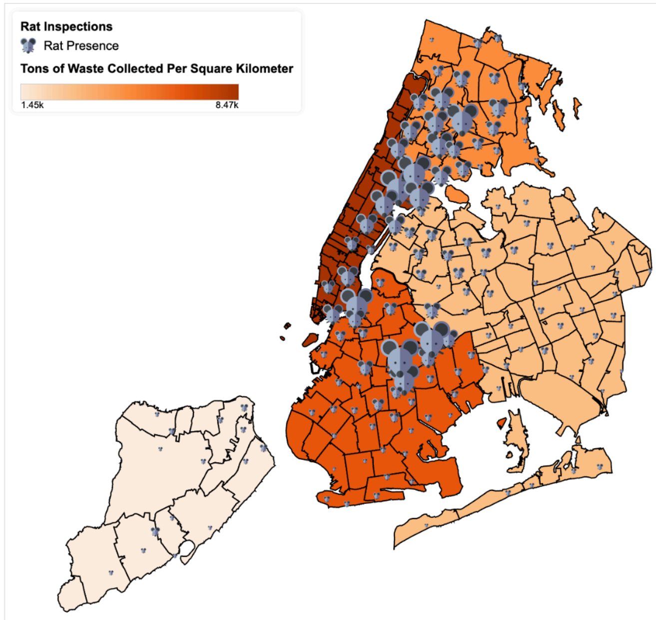
New York City is a bustling epicenter of all kinds of activity, and that includes rats! While rat sightings are pretty common on the streets and in the subway, public records of inspections and complaints filed in response to sightings and indications of rats can indicate a larger, recurring problem in those locations. Our map shows the traces left behind by rats through instances filed inspections across all five boroughs of New York City, and how they compare to potential sources of food, namely human waste. We can understand how much trash is at a rat's disposal by how many tons are collected in each borough total across the entire year of 2023, as well as the "density" per square kilometer, potentially indicating high concentration of garbage disposal chutes, dumpsters, and other prime locations for rats to scrounge for food.

Click to zoom in for individual rat presence and inspection results

Toggle through the waste collection and restaurant inspection violations to get a sense of how dirty different areas of New York City are
(which translates to: how conducive it would be to live there, as a rat!)

Filter By:
- [x] Waste Collection
  - ○ Total amount by borough
  - ◉ Waste density per sq. km by borough
- [ ] Restaurant Violations

Reset



The Life of Rats in New York City *(2023)*

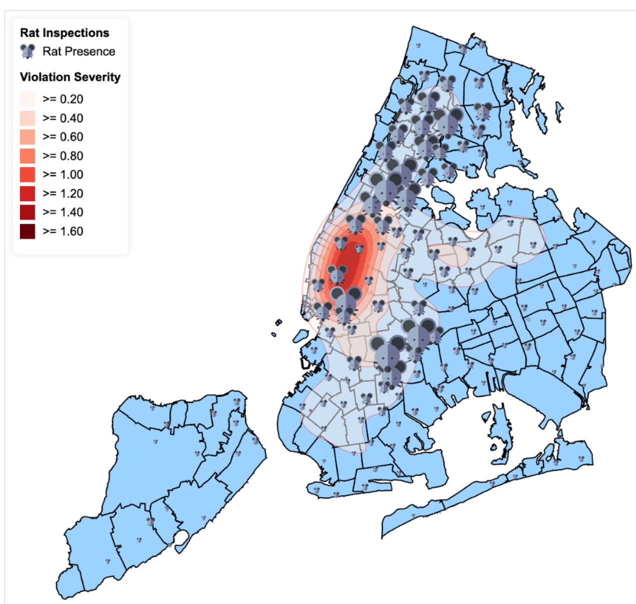
New York City is a bustling epicenter of all kinds of activity, and that includes rats! While rat sightings are pretty common on the streets and in the subway, public records of inspections and complaints filed in response to sightings and indications of rats can indicate a larger, recurring problem in those locations. Our map shows the traces left behind by rats through instances filed inspections across all five boroughs of New York City, and how they compare to potential sources of food, namely human waste. We can understand how much trash is at a rat's disposal by how many tons are collected in each borough total across the entire year of 2023, as well as the "density" per square kilometer, potentially indicating high concentration of garbage disposal chutes, dumpsters, and other prime locations for rats to scrounge for food.

Click to zoom in for individual rat presence and inspection results

Toggle through the waste collection and restaurant inspection violations to get a sense of how dirty different areas of New York City are
(which translates to: how conducive it would be to live there, as a rat!)

Filter By:
- [ ] Waste Collection
  - ○ Total amount by borough
  - ○ Waste density per sq. km by borough
- [x] Restaurant Violations

Reset



Data Description

For our project, we obtained our datasets from NYC Open Data and Kaggle.

Our first dataset, [Rodent Inspection] contains the information on rat inspections sourced from the NYC Department of Health and Mental Hygiene. This dataset contains detailed information on rodent-related inspections conducted across the five boroughs of New York City and includes geographic and inspection-related attributes.

We filtered out unnecessary variables from the Rodent Inspection dataset, such as JOB_TICKET_OR_WORK_ORDER_ID, HOUSE_NUMBER, COMMUNITY BOARD, COUNCIL DISTRICT, and CENSUS TRACT, to focus on relevant attributes and reduce file size for easier processing. Additionally, we randomly sampled 10,000 records from the original 100,000 to address loading issues and prevent visual clutter caused by overlapping data points in the visualization.

To ensure consistency across datasets, we filtered the data to include only inspections from 2023, aligning with the timeframes of the other datasets used in the project. Furthermore, we filtered the RESULT column to retain only records marked with "Rat Activity" since our analysis focuses specifically on rat activity, reducing unnecessary information and allowing us to highlight relevant insights more clearly without overwhelming the map with extraneous data.

The following variables are included in the final dataset:

- **INSPECTION_TYPE**:Describes the type of inspection performed
    - **Initial Inspection**: Conducted in response to a 311 complaint or through proactive neighborhood indexing
    - **Compliance Inspection**: Follow-up inspection performed after a failed initial inspection.
- **BORO_CODE**: A numerical code assigned to each NYC borough. This code also serves as the first digit in the BBL tax lot code.
- **STREET_NAME**: The street name of the tax lot that was inspected.
- **ZIP_CODE**: The postal ZIP code of the inspected tax lot.
- **LATITUDE**: Geographic latitude (in decimal degrees, WGS 1984) of the inspected tax lot.
- **LONGITUDE**: Geographic longitude (in decimal degrees, WGS 1984) of the inspected tax lot.
- **BOROUGH**: The name of the NYC borough where the inspection occurred (e.g., Manhattan, Brooklyn).
- **INSPECTION_DATE**: The date when the inspection was conducted.
- **RESULT**: Outcome of the inspection.
    - **Active Rat Signs (ARS)**: Evidence of rats, such as tracks, droppings, burrows, runways, rub marks, gnawing marks, or live rats.
    - **Problem Conditions**: Issues like garbage (poor food waste containment leading to rat feeding), harborage (clutter or vegetation promoting nesting), or evidence of mice.

Our second dataset, [DSNY Monthly Tonnage Data], utilizes data from the New York City Department of Sanitation (DSNY), which provides monthly collection tonnages of refuse, recyclables, and organics collected from NYC residences and institutions. The dataset contains information about borough-level waste collection across the city.

We removed variables such as SCHOOLORGANICTONS, LEAVESORGANICTONS, XMASTREETONS, and BOROUGH_ID because these fields often contained null or incomplete data, which would have introduced inaccuracies in our calculations. Including these could have skewed the overall tonnage results and did not align with our focus on borough-level analysis. Similar to the Rodent Inspection dataset, we filtered the data to include only records from 2023 to allow for comparisons between waste tonnage and rodent activity during the same time period. We focused on tonnage variables (REFUSETONSCOLLECTED, PAPERTONSCOLLECTED, MGPTONSCOLLECTED, RESORGANICSTONS) because they provide a comprehensive view of waste collection that potentially correlates with rodent activity in NYC.

The final variables are as follows:

- **MONTH**: Year and month of data collection.
- **BOROUGH**: One of the five boroughs of NYC (e.g., Manhattan, Queens).
- **COMMUNITYDISTRICT**: Corresponds to one of NYC's 59 community districts, which align with sanitation districts.
- **REFUSETONSCOLLECTED**: Total tons of general refuse (trash) collected.
- **PAPERTONSCOLLECTED**: Tons of recyclable paper collected, separated at the source.
- **MGPTONSCOLLECTED**: Tons of recyclable metal, glass, plastic, and beverage cartons collected, separated at the source.
- **RESORGANICSTONS**: Tons of residential organics (e.g., food scraps, yard waste) collected, separated at the source.

Our third dataset, [NYC Restaurant Inspection Results 2023] provides comprehensive information on restaurant health inspections across New York City, including attributes related to location, inspection dates, results, and violations.

The main goal for our data processing was to keep only necessary variables for our analysis as well as to reduce the file size. The dataset size was reduced from 4395 GB to 6 MB after filtering and column selection. Our process was as follows:

- Inspection Date: To ensure consistency with the rat inspection data from 2023, only inspections conducted in 2023 were included, focusing on correlations between restaurant violations and rat activity.
- Critical Violations: Only records flagged as CRITICAL were retained to focus on significant health code violations.
- Column Selection and Variable Descriptions: Columns that were irrelevant to the purpose of our analysis were removed (BUILDING, STREET, PHONE, VIOLATION DESCRIPTION, SCORE, GRADE, GRADE DATE, record date, inspection type) and the following key variables were kept in our final dataset:
  - CAMIS: Unique identifier for the restaurant.
  - DBA: The name (doing business as) of the restaurant.
  - BORO: The borough where the restaurant is located.
  - ZIPCODE: Zip code of the restaurant's location.
  - CUISINE: Description of the restaurant's cuisine type.
  - INSPECTION DATE: Date of the inspection.
  - VIOLATION: The violation code is associated with the inspection.
  - CRITICAL FLAG: Indicator for critical violations (Critical, Not Critical, Not Applicable).
  - Latitude and Longitude: Geographical coordinates of the restaurant.

Our fourth dataset **NYC_boroughs** was used for purposes of providing geographical context to outlines the borough boundaries of New York City. The file provides geographic data for each borough; specifically we used the polygon coordinates it contained, which defined the shapes of NYC's boroughs: Manhattan, Brooklyn, Queens, Bronx, and Staten Island, to draw out our map.

For our fifth dataset, the **NYC_zipcodes** is a GeoJSON file provides geographic boundary data for NYC postal zip codes. We used this dataset to overlay zip code boundaries on the borough map, providing finer spatial granularity for our visualizations. This dataset was obtained from a publicly available source and contains the following variables:

- **postalCode**: The postal zip code for the region (e.g., "11372").
- **PO_NAME**: The name of the postal area associated with the zip code (e.g., "Jackson Heights").
- **borough**: The NYC borough to which the zip code belongs (e.g., "Queens").
- **geometry**: A polygon defining the boundary of the zip code using latitude and longitude coordinates.
- **OBJECTID**: A unique identifier for each zip code region.
- **Shape_Leng**: The calculated perimeter of the zip code boundary in meters.
- **Shape_Area**: The calculated area of the zip code boundary in square meters.

Visual Design Rationale

Map: Rodents in New York City

We chose to use a map to visualize trends in rodent activity across New York City. The borough and zip code boundaries are displayed with black outlines, with borough boundaries given bolder strokes to provide a clear spatial hierarchy. The boroughs are filled with a light blue color to establish spatial context while maintaining a neutral, non-distracting background. At a high-level initial view, rodent data points are aggregated into clusters represented by scaled rat icons. This approach reduces visual clutter and prevents overlap, enhancing readability, which would otherwise be compromised by displaying all 10,000 data points at once. A tradeoff is the loss of granularity at this high level, but this is mitigated by interactivity—at more zoomed-in levels, individual inspections are depicted as slightly transparent grey circles, allowing users to inspect specific locations without overwhelming the display. A legend is included in the top-left corner to clarify the meaning of the rat icons, their sizes, and how they scale with inspection counts.

In terms of visual channels, the map employs spatial positioning along vertical and horizontal axes to map both the boroughs and rodent inspections accurately, using a Mercator projection to align NYC's geographic coordinates. The area of the rat cluster icons is proportional to the number of inspections, using square root scaling to emphasize differences in inspection counts while compressing extreme values for better readability. Additionally, saturation is used for individual inspection points, with slightly transparent gray circles that distinguish individual data points while maintaining clarity against the light blue background of the boroughs. The boroughs themselves are colored in light blue to ensure contrast and visibility against the black outlines of boundaries and the gray and black rat icons.

Map: Tonnage and Rodents

We chose to create a choropleth map to visualize rodent activity and waste (refuse, paper, MGPT, and organics) across NYC boroughs. The map retains the earlier layout, with borough and zip code boundaries, but introduces a sequential blue color scale as the fill color to represent either total tonnage or tonnage per km². Radio buttons allow users to toggle between these two metrics: total tonnage highlights absolute waste contributions, while tonnage per km² emphasizes the density of waste per unit area, normalizing borough comparisons. However, since the tonnage data is aggregated at the borough level, we couldn't explore relationships at a finer spatial resolution (e.g., zip code correlations with rodent densities). This trade-off prioritizes visual clarity and avoids clutter from less granular data.

In terms of visual channels, we utilized spatial positioning to accurately map the boroughs, waste data, and rodent activity based on their geographic coordinates. Additionally, we used the area for the rat clusters, carried over from the earlier map. The sequential blue color scale fills the boroughs, with darker shades representing higher waste metrics and lighter shades representing lower values, providing an intuitive gradient that highlights variation, the visual channel being the saturation of the blue. Transparent gray dots are used to depict individual rodent inspections, balancing visibility and preventing clutter when zoomed in. The color scale leverages quantile scaling, dividing the data into perceptually equal bins to enhance interpretability. This ensures users can compare waste levels meaningfully, even across regions with varying data distributions. A legend at the bottom of the map clarifies the mapping between the shades of blue to the tonnage values, providing a seamless gradient from light blue (representing lower tonnage) to dark blue (indicating higher tonnage) to ensure that users can interpret the data's magnitude across the boroughs.

In terms of transformation, the total tonnage metrics were calculated by summing refuse, paper, MGPT, and organic waste for each borough. For tonnage per km^2, we normalized total tonnage by dividing it by borough area. Borough areas were calculated in spherical radians and converted to square kilometers based on Earth's radius. The resulting values were then scaled using the sequential blue color palette. While quantile scaling ensures equal perceptual groupings, it compresses extreme values, slightly limiting visibility of the highest and lowest data points. The transformation to tonnage per km^2 introduces a density-based view, which highlights the areas where waste production is high relative to borough size.

Map: Restaurant Violations and Rodents

We chose to use a contour map to visualize rodent activity and restaurant violation severity across NYC. The map's layout is likewise retained here as with the previous 2 maps, but utilizes a sequential color red scale that decreases in luminosity (increasing in darkness) as the fill color to represent increasing densities, and thus severities, of restaurant violations. Contours additionally include semi-transparent red strokes and a fill opacity of 0.5 to ensure we avoided obscuring the underlying map features. Laying the contour map underneath the rats' icons (rats inspection results) as well provides users with better visibility to see the correlations between rodent activity and restaurant violation severity.

In terms of visual channels, we utilized spatial positioning in order to map rodent activity and restaurant violation locations and thus their corresponding densities based on their geographic coordinates. Area for rat clusters was, as in the previous map, carried over from the earlier map. Similar to the previous map, the translucent gray dots that show upon zooming in provide visibility and prevent clutter for displaying individual rodent inspections. For the density contour, we used another sequential color scale in which shades of red darken (decreasing in luminosity) with increasing densities of restaurant violations in a way that intuitively associates "concentrations" between the two (color/density of violations). This is additionally denoted through the legend located at the top left corner of the map for the viewer's reference in order to interpret this information in the visualization. The color scale leverages quantile scaling, dividing the data into perceptually equal bins to enhance interpretability. This ensures users can compare waste levels meaningfully, even across regions with varying data distributions.

In terms of transformation, latitude and longitude were converted to pixel positions using a Mercator projection to enable mapping. In order to calculate the density of restaurant violations that allowed us to create a contour map to highlight areas with high concentrations of critical violations, coordinates from the geographic transformation were processed with the d3.contourDensity() function. We then determined the density thresholds (the numerical values displayed on the legend) by segmenting the resulting density values from the density function into 10 intervals to simplify interpretation and emphasize areas of interest. The tradeoff we considered with this latter aspect was that although this limits the intervals of and thus the nuance of density that is displayed, there is greater clarity in terms of a summary view of which parts of NYC have more severe restaurant violations.

Interactive Elements Design Rationale

One of our challenges was displaying all the rodent activity points in a clear manner, since due to the abundance, there was a lot of overlap. To combat this on the NYC level map, we clustered the points by 1 mile radius, making the rat icon size correlate to the amount of points in each cluster. By hovering over the rat icons, the viewers can see how many inspection cases are in that 1 mile radius cluster. However, we still wanted to display all the rodent activity points in some way to not lose visual information. Therefore by adding the zoom interactivity feature that allows the viewer to zoom into sections on the NYC map, it plots all the points on the zoomed section to provide a clearer distribution of all the points in that area. To make this discoverable for viewers, when they hover over zoomable areas, the cursor turns into a magnifying glass. When the map is completely zoomed, the magnifying glass cursor turns into a grabbing cursor, signifying that the user can drag the map to see another area.

For the tonnage visualizations, we added a hover effect to enhance interactivity and provide users with contextual information. When users hover over a borough, a tooltip appears, displaying the total tonnage or tonnage per square kilometer for that borough, depending on the selected metric. This interactive feature complements the zoom and hover effects used in the rodent activity visualization, creating a cohesive and engaging user experience across both maps.

We included filters for the purpose of viewing how different hygienic factors could relate to rodent activity. Specifically, we included checkmarks for waste collection and restaurant violations to communicate that both maps could be simultaneously viewed, should users wish to see how both hygienic factors might relate to one another and to rodent activity; there are options to visualize waste collection, allowing the user to choose how they would like to compare waste collection to rat populations across the map (either total tons collected or density by square kilometer). The radio buttons are also initially deactivated, as indicated by the grayed out button, to indicate to users that they must select the checkmark to enable showing of the waste collection map first prior to selecting the specific type of waste collection map they wish to see. We included a reset button that changes color when the mouse hovers over to give users the option to remove all visualizations.

The Story

Our visualization aims to explore the relationship between rodent activity, waste management, and restaurant violations in New York City by leveraging data from multiple sources to uncover potential patterns. The NYC level map highlights spatial trends in boroughs and zip code boundaries, with inspection results represented by scaled rat icons. This visualization allows for easy identification of rodent activity hotspots, while including heat map overlays for additional content illustrating areas of high restaurant violations.

The map reveals that rodent activity is mostly concentrated in the central and northern Manhattan area, with severity levels surrounding areas that exhibit restaurant violations activity. This trend suggests correlations between sanitation challenges, urban density, and rodent activity. Compared to other boroughs like Brooklyn, Queens, and Staten Island, there are also pockets of rodent activity, but their patterns are more dispersed compared to the density of Manhattan's clusters. To further explore this relationship, we examined waste tonnage data. Interestingly, the boroughs with the highest rodent activity—Manhattan and Brooklyn—also have the highest waste tonnages. This observation underscores the potential link between waste management practices and rodent activity across NYC. Waste management insights may include as proximity between trash collection sites and areas of trash that may remain in the same location between trash collection periods, such as

A surprising observation is the close overlap between heatmap areas of restaurant violation severity and the clusters of rodent inspections. All of the major big clusters of rodent inspections all lie within the regions of high violation severity, reinforcing the idea that suggests a correlation between urban activity with rodent activity. In addition, there is relatively low rodent activity within Staten Island compared to the other boroughs which aligns with the 2020 Census data indicating Staten Island is the least populated NYC borough.

Another interesting finding is that when the two maps (waste collection and restaurant violations) are overlaid, there is one part of the map where the two aforementioned patterns overlap; i.e., in lower Manhattan, the borough with the highest indicated waste collection density, there is also a major cluster of rodent inspections, as well as the highest restaurant violation severities. Notably, however, there are also major clusters of rodent inspections around the boundary between Manhattan and Brooklyn, as well as upper Brooklyn. In these particular boroughs, while restaurant violation density is not as high, the waste collection densities are at similarly high levels, yet total waste collected between these locations also differs significantly (i.e., Manhattan/Bronx have high waste collection densities and low total waste collected, whereas Brooklyn has a high waste collection density and a high number for total waste collected). This suggests that there are other possible factors that may contribute to rodent activity.

Ultimately, the primary insight we want to convey to viewers of the “NYC 2023 Rat Inspection Dashboard” is awareness of cleanliness in New York City. The map identifies high-violation and high-activity areas where potential sanitation efforts could have the most impact on the community’s health. In addition to the rodent activity not being uniform across boroughs, this called attention to Manhattan as a priority for intervention. These insights hope to empower the general public of trends and initiate decisions to address this urban challenge.