CS/INFO 3300
Project 1
Final Report
Due Friday, October 7

Correlation Between the Availability of Public Recycling Bins and the Recycling Rate in Five Boroughs of New York City

Description of the Data

In order to make our visualizations, we took data from three different datasets. All three datasets were created by NYC OpenData, a free public data source published by New York City agencies and other partners.

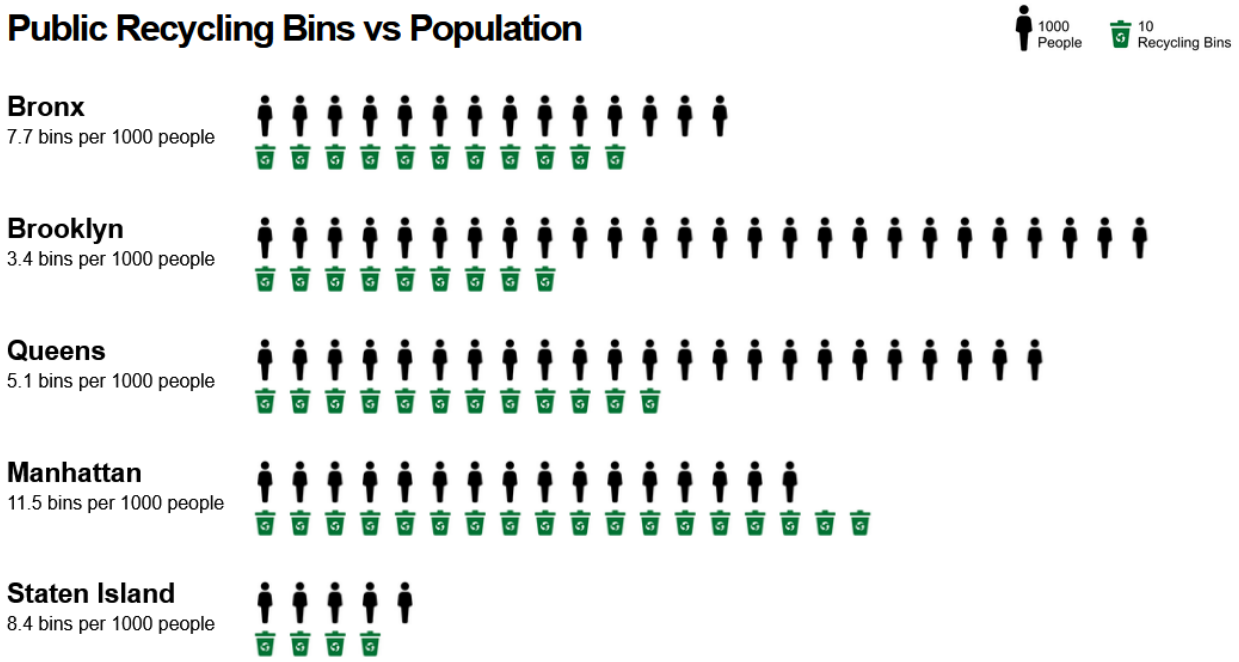
The first dataset gave us the population of each borough, which we eventually used to calculate the number of public recycling bins per capita. Since this dataset also included the information on all the boroughs combined (total NYC), we had to filter out this data as to not skew our scales for the bar charts.

The second dataset gave us information on the different recycling rates by borough. It was last updated February 7th, 2020 by the Department of Sanitation (DSNY). From this dataset, we used the field Recycling Diversion Rate, defined as the percentage of total municipal solid waste collected by the Department of Sanitation that is disposed of by recycling. This was different from the Capture Rate, the percentage of total paper or metal/glass/plastic in the waste stream that is disposed of by recycling. Since this dataset broke up the boroughs into districts, we had to do some filtering and counting. We combined all the districts for each borough to get the average recycling rate across the entire borough, given our visualizations would not be broken down to that level of specification.

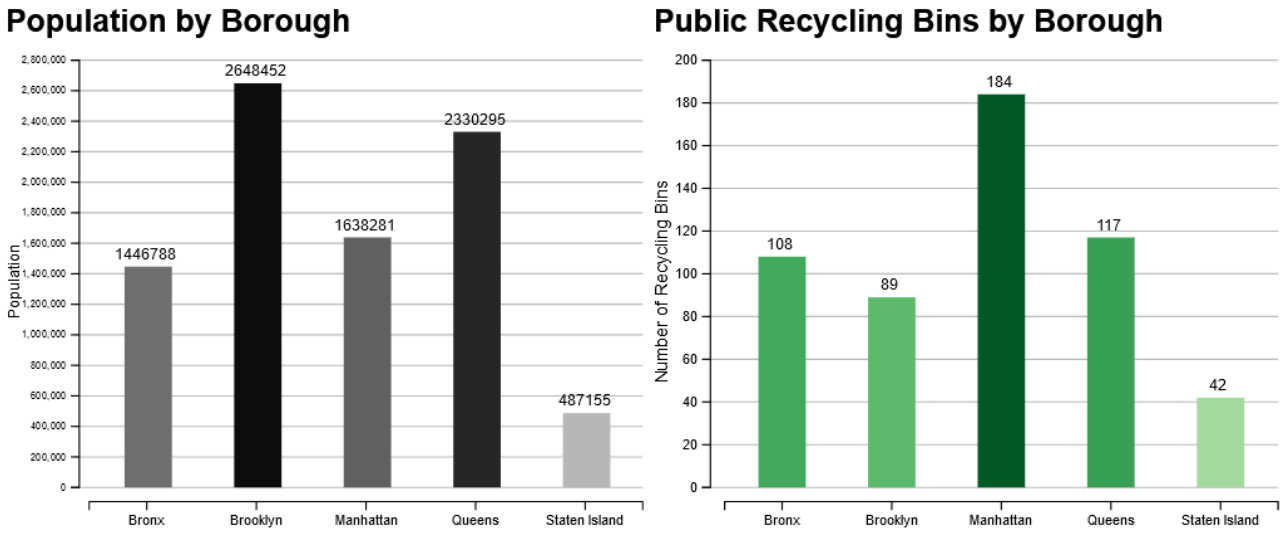
Lastly, the third dataset gave us the exact location of each public recycling bin throughout the entire city of New York. It was last updated October 4th, 2022 by DSNY and originally made public in 2012. Since we needed to calculate the total number of bins per borough, we counted the number of addresses located in each borough in order to get this result.

After completing all the calculations needed from these datasets, we constructed our own list of dictionaries to hold this information, which we defined as recycling_rate. For each borough, we stored its name, population, and average recycling rate across all its districts. Using this list, combined with the rest of the data from the datasets and free open source icons for the human, recycling bin, and borough figures, we were able to construct all visualizations shown below.

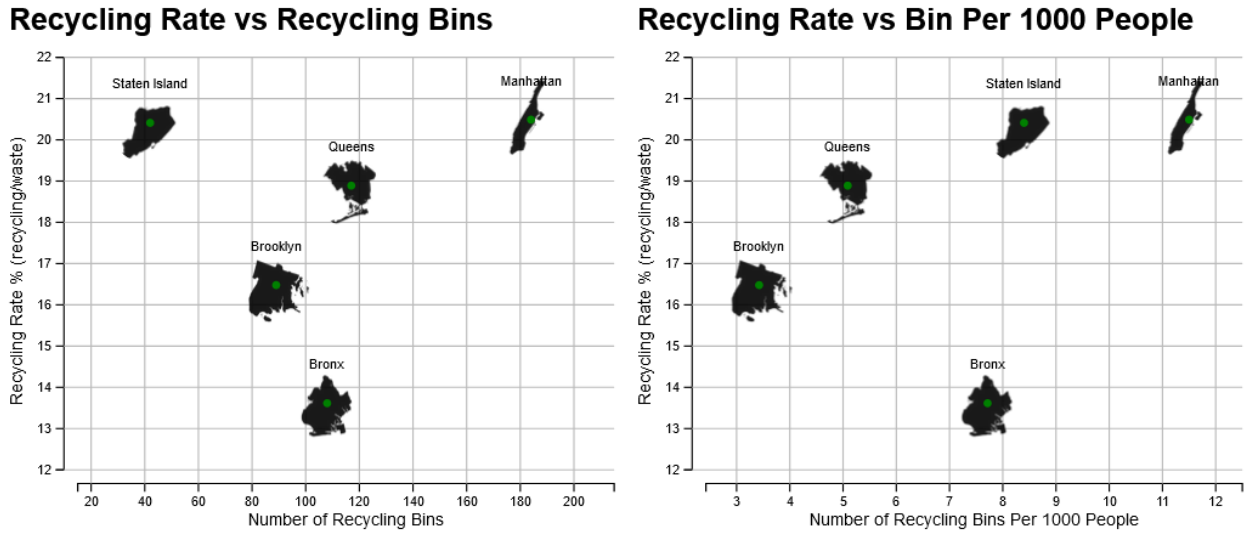
Visualization 1



Visualization 2



Visualization 3



Design Rationale

For the first design visualization, we aim to see how many public recycling bins each borough has for its population. The three variables are the boroughs, the number of recycling bins, and the population. In this visualization, we decided to use icons as marks to represent the recycling bins and population. Each human figure icon in the

visualization is a mark representing 1000 people in the dataset. We tried to pick a gender-neutral-looking icon. Each recycling bin icon in the visualization represents ten recycling bins in the dataset. The rationale is that if we have 2648452 human icons on the screen for everyone in Brooklyn, the icons will be too small to be legible. However, the trade of using one icon to represent 1000 people is that if the user didn't see the legend, they might not know the amount each icon represents. The visualization relies on the user noticing the legend. This trade-off is something that we are willing to accept since the primary goal of visualization 1 is to preface the proportion of people and recycling bins before showing them the actual numbers. Since the goal is to show the proportions, we decided to include a text with each borough indicating the number of bins per 1000 people. We chose to calculate and display it in our visualization so that the user doesn't have to do it for themselves. We showed this relationship per 1000 people because of the icons, and we want consistency throughout our project. Establishing the relationship per 1000 people also gives the user the number of bins in single digits. For example, if we show this relationship per 10 people, we would have to write 0.084 bins per 10 people for Staten island. We are afraid that users will not be able to see what a 0.084 bin looks like. One can imagine this visualization as five scatterplots for each borough, and circles are used as the marks instead of circles.

Another factor we considered in this visualization is color. We picked black for humans since it provided contrast with the white background. We chose green for recycling bins. Since recycling bins relate to environmentally friendly concepts, we decided green would be a standard association that the user might already have. The

order of boroughs is in alphabetical order. Ordering the boroughs by the number of population or bins would be beneficial since repetition would provide a clearer understanding. However, we accepted this trade-off mainly because we do not want to have a different order of boroughs between visualization 1 and 2. The first visualization is horizontal instead of vertical because it feels more natural for recycling bins and humans to be next to each other like they are on the ground instead of stacked on top. Another reason this visualization is not vertical is that when the user first opens the website, they would see the icons at the top but not the borough labels. Presenting this visualization horizontally, we can give the idea of a diagram at a glance for at least the first two boroughs.

The second visualization is actually made before the first visualization. We decided to come up with the first visualization after completing the second visualization because we wanted to tell a story first before presenting the numbers. The first visualization provides the story and the data, but it doesn't offer the user the actual numbers. Therefore we decided to keep the second visualization to show the exact number population and borough without rounding it to the thousand or the ten. Visualization 2 has two sub-visualization. For this report, we will name the visualization on the left a visualization 2a and the one on the right 2b. For visualization 2a, the boroughs are on the x-axis, and the population is on the y-axis. For visualization 2b, the boroughs are on the x-axis, and the number of bins is on the y-axis. The marks in this visualization are rectangular bars. The channels in both 2a and 2b are the height of the rectangles on a uniform scale representing the number of people or bins. The second

channel in these visualizations is color. The colors go from gray/light green to black to dark green representing the number of bins. This repetition in the channels is intentional. Our goal is to use repetition to reinforce the users association of which boroughs have more population vs. bins. In addition to providing the numbers scale on the axis with grid lines, we provide the exact value right on top of the rectangle.

Visualization 1 shows the population and recycling bins side by side, allowing the user to understand the ratio between these two variables. Visualization 2 presents these two variables separately. While visualization 1 tells the story using visual icons to represent rounded values, visualization 2 shows the raw number without the visual elements. The tradeoff is that there are more visualizations. We are okay with this tradeoff, considering that there's less clutter in both visualizations, and each visualization has a different purpose.

The last visualization is a scatter plot. Visualization 3 has two sub-Visualization. In this report, the visualization on the left will be called visualization 3a, and the visualization on the right will be called visualization 3b. The **Our story** section will further clarify why we have two visualizations showing recycling rate vs. recycling bins and recycling rate vs. bin per 1000 people. The x-axis in this visualization is the number of recycling bins for 3a and the number of recycling bins per 1000 people for 3b. The y-axis in this visualization is the recycling rate. The marks in this visualization are circle dots.

Additionally, we decided to add text and shape labels to the circles based on the borough the data is from. The shape of each borough is used to help the user identify each

borough. Easy identification of the boroughs would help when the user is going back and forth between visualization 3a and visualization 3b. We added the map of the five boroughs of NYC at the beginning of the web page to foreshadow this part of the visualization. The tradeoff in using the shapes is that users would have difficulty guessing the shape's area. However, our visualization does not currently rely on the area of each borough as a variable, so we accepted this tradeoff. The circle marks' position depends on the number of recycling bins and the recycling rate on a uniform scale. We showed visualization 3 after we introduced visualization 1 and 2. We want to introduce the concept of recycling bins per 1000 people in visualization 1 and the population variation in each borough in visualization two before showing visualization 3.

Our Story

Intuitively, one could assume that the amount of recycling in an area would increase if the availability of recycling bins is also increased. Since NYC Open Data does a great job compiling recycling data, we decided to focus on New York City to see if this relationship is true. Our visualization aims to display the availability of public recycling bins vs the recycling rate in each borough in New York City in relation to the population.

When plotting average recycling rates versus the total number of recycling bins in a borough (in Visualization 3), we found that there wasn’t a clear correlation between recycling rates and recycling bins. For example, the borough with the least recycling bins (Staten Island) and the borough with the most recycling bins (Manhattan) had roughly the same amount of recycling. This was initially surprising since it seems to counter the intuitive relationship that more recycling bins lead to more recycling.

We then decided to take the population of each of the boroughs into account by looking at the number of recycling bins relative to the number of people in the boroughs, which gives a much better metric of the availability of recycling bins in an area. Visualizations 1 and 2 show this by comparing the number of recycling bins to the population in the boroughs. This reason is also behind the rationale on why visualization 1 and 2 are introduced first before visualization 3. After factoring this into our scatter plot so that it plots average recycling rates versus the number of recycling bins per 1000 people, we found a much stronger correlation. It seems to show a positive logarithmic trend where the more recycling bins there are per 1000 people the higher the recycling rate becomes. However, because of the logarithmic shape of the trend, there seem to be diminishing returns where after a certain amount of recycling bin availability the recycling rate starts to plateau.

One thing to note is that The Bronx doesn't follow this trend as it has the lowest average recycling rate despite having a fairly good recycling bin availability. Unfortunately, since we only have 5 data points, we don't have enough data to see if this is just an outlier or if it indicates that the correlation we found is incorrect. Given more time, a next step could've been to add in more data points from other cities to see if the trend still holds.

Team Contributions

Glenn:

- Wrote the description of the data for the final report (~ 45 minutes)
- Filtered the population dataset (~ 30 minutes)
- Constructed counting functions to get the total number of recycling bins / average recycling rate by borough (~ 1 hours)
- Created the two bar charts to compare borough population with the total number of public recycling bins per borough, also gave us the information to calculate recycling bins per

capita (~ 4 hours → had some difficulty manipulating the data from these datasets to fit the scales)

Alex:

- Wrote the description of the design rationale for the final report (~ 2 hours)
- Design Sketches and Figma Mockup - generated multiple iterations for the whole webpage before selecting one (~ 1 hour)
- Data Processing For Visualization 1 (~1.5 hour) the first visualization needs the total population represented with an icon for every 1000 human and total recycling bins represented with an icon for every 10 bin
- Making Visualization 1 (~3.5 hours) - Challenging to make because it is very different from the demo visualizations made in class. Used an ad hoc method using scatter plots
- Finding Icons and Implementing them to visualization 1 & 3 (~2 hours)
- CSS formatting (~2 hours)
- Adding Axis Labels to Visualization 2 & 3 , Rescaling Visualization 2 & 3(~45 minutes)

Ryan:

- Wrote the Story section of the report (~45 minutes)
- Data Processing of datasets (~45 minutes)
- Created two scatter plots: one plotting average recycling rate vs total number of bins and one plotting average recycling rate vs bins per 1000 people (~3 hours)
- Cleaned up code (~1 hour)

Datasets and Sources Cited

(DCP), D. of C. P. (2014, April 29). *New York City population by borough, 1950 - 2040: NYC open data*. New York City Population by Borough, 1950 - 2040 | NYC Open Data. Retrieved October 6, 2022, from https://data.cityofnewyork.us/City-Government/New-York-City-Population-by-Borough-1950-2040/xywu-7bv9

(DSNY), D. of S. (2019, July 24). *Recycling diversion and capture rates: NYC Open Data*. Recycling Diversion and Capture Rates | NYC Open Data. Retrieved October 6, 2022, from https://data.cityofnewyork.us/Environment/Recycling-Diversion-and-Capture-Rates/gaq9-z3hz

(DSNY), D. of S. (2022, October 4). *Public recycling bins: NYC open data*. Public Recycling Bins | NYC Open Data. Retrieved October 6, 2022, from https://data.cityofnewyork.us/Environment/Public-Recycling-Bins/sxx4-xhzg

Human Icon - Human by Adrien Coquet https://thenounproject.com/icon/human-1683012/ by Creative Commons license,

Recycling Trash Can Icon - Recyclable by Alice Design https://thenounproject.com/icon/recyclable-2025392/ Creative Commons license ,

NYC City Icon - New York City by Minna Ninova https://thenounproject.com/icon/new-york-city-160968/ Creative Commons license,

Bronx Icon - The Bronx by Minna Ninova https://thenounproject.com/icon/the-bronx-160969/ Creative Commons license,

Brooklyn Icon - Booklyn by Minna Ninova https://thenounproject.com/icon/brooklyn-160964/ Creative Commons license,

Manhattan Icon - Manhattan by Minna Ninova https://thenounproject.com/icon/manhattan-160965/ Creative Commons license,

Queens Icon - Queens by Minna Ninova https://thenounproject.com/icon/queens-160966/ Creative Commons license,

Staten Island Icon - Staten Island by Minna Ninova https://thenounproject.com/icon/queens-160966/ Creative Commons license,