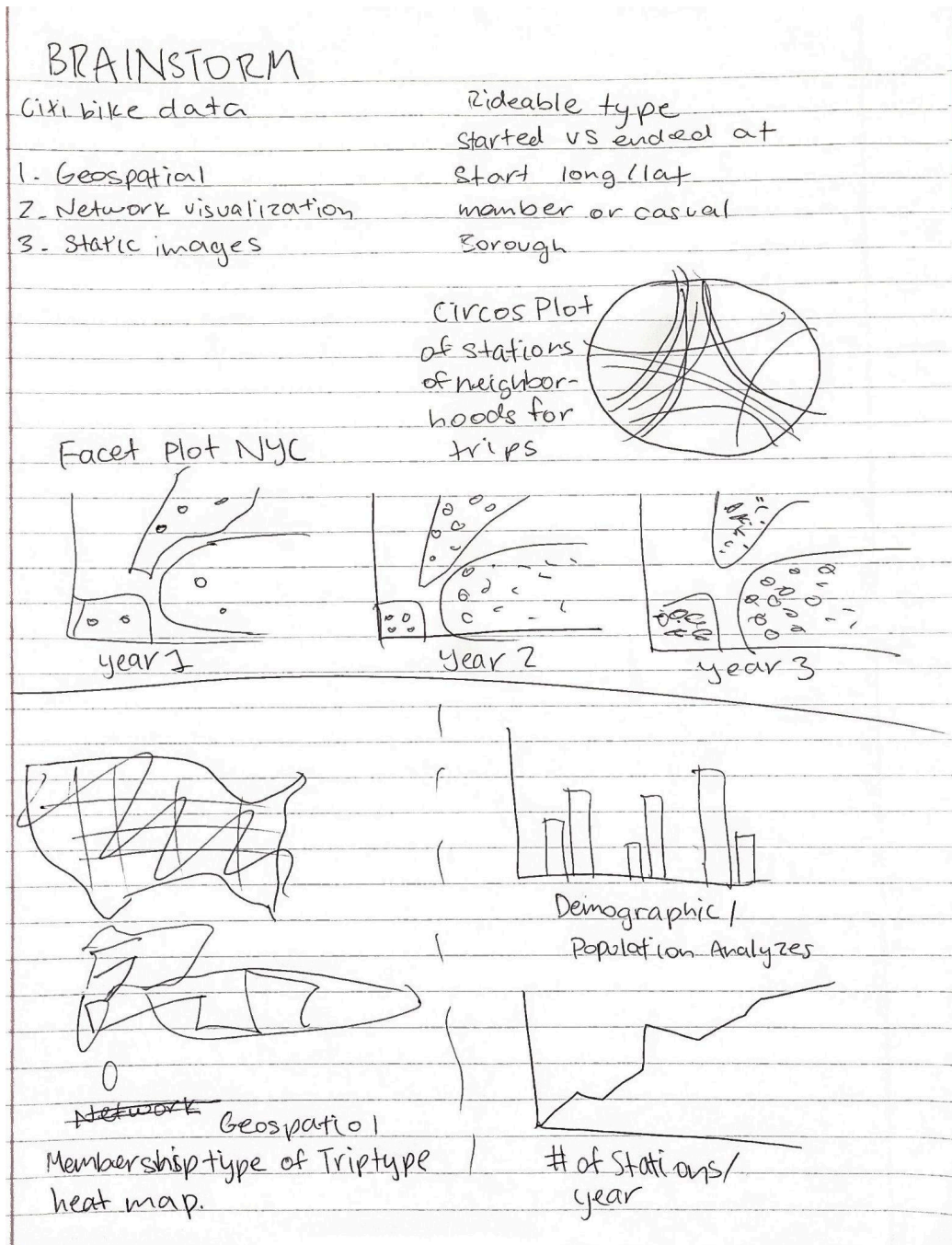


Project Book | Group A Citi Bike

The project book begins with hand-drawn designs from our initial brainstorming sessions. It then outlines the challenges we encountered, how our design choices evolved, and the reasoning behind the current layout and presentation of the plots on the website.

Original Hand Drawn Designs



Evolution of Manhattan Bike Stations (2014–2024):

For this plot, we were interested in looking at the change or progression of bike stations over time. Originally we wanted to plot stations for all the 10 years before 2024, but we soon realized the plot would become too crowded. Moreover, the change between consecutive years was not that significant, therefore to best demonstrate the spread and trend of change for bike stations over time, we chose a 5 year gap and only presented data from 2014, 2019, and 2024. Similarly, as stations did not vary much throughout a single year, we chose the month July for easier data processing and comparison for visualization.

Electric vs. Classic Bike Usage by Neighborhood in 2024:

The second question we were interested to learn about was usage patterns of electric vs classic bike usage across the different NYC neighborhoods. At first we wanted to create pins or clusters of pins for bike usage within each neighborhood, but soon we realized as the count for rides were in magnitude of millions, it is impractical to plot each individual ride and thus we ended up creating a heatmap to convey the message in the data.

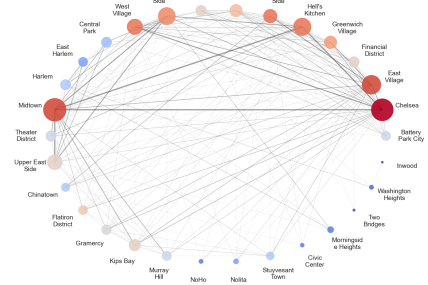
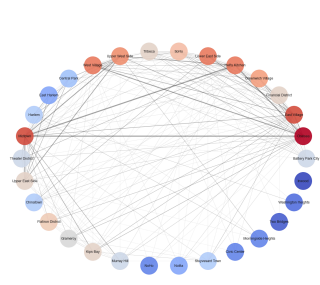
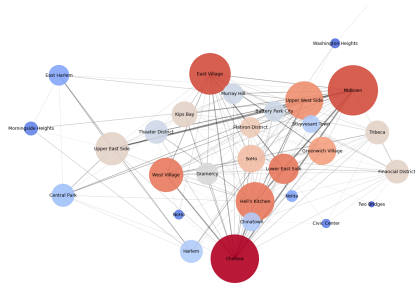
Furthermore, because we wanted to demonstrate the different usage patterns of electric vs classic bikes per neighborhood, we tried creating layers that show difference of ride counts for electric vs classic bikes. However, electric bikes always had significantly more rides compared to classic bikes, the ride count difference made little sense for direct interpretation. Therefore, we decided to show the percentage of electric bike rides over the total number of rides within that neighborhood instead. We also created the heatmap layer using the normalized percentages for better visualization to show contrast/comparison between the neighborhoods.

Member vs. Casual Bike Usage by Neighborhood in 2024:

For the comparison of member vs casual bike usage per neighborhood, we did not want to create a heatmap again like the plot before. We then decided to use overlapping bar plots to display the data. We first plotted the data with the neighborhoods on the x-axis, but soon realized the plot would then be too wide and make the neighborhood names hard to read. For cleaner visualization, we also ordered the bars in descending order of the total number of member bike rides. This would allow us to also easily identify the major neighborhoods with more bike rides in general.

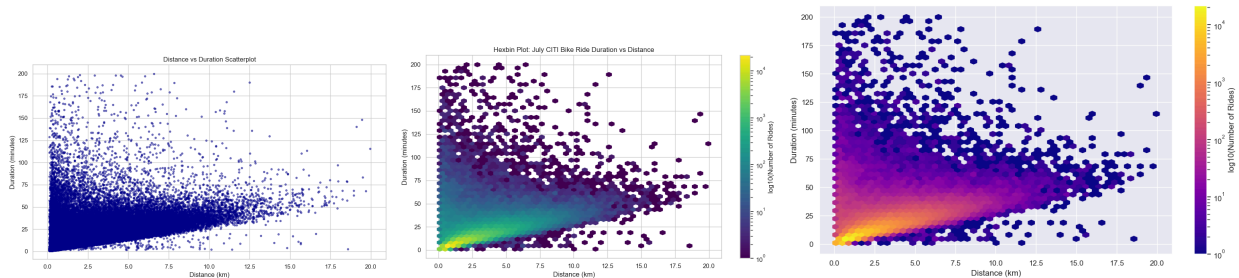
Bike Trip Network Across Manhattan Neighborhoods:

This network visualization was a big challenge and took many iterations. It was important that it told a different story than the neighborhood visualizations that were performed earlier. So I brought in multiple metrics such as degree centrality and trip number. Starting with a kawaii layout, it was visually confusing as the neighborhoods did not layout how they were in real life, so I switched to a circular image to replicate a Circos-like graph. Then I added node size and color to show which neighborhoods were hubs of public transport. We were able to show which neighborhoods had the most trips in and out.



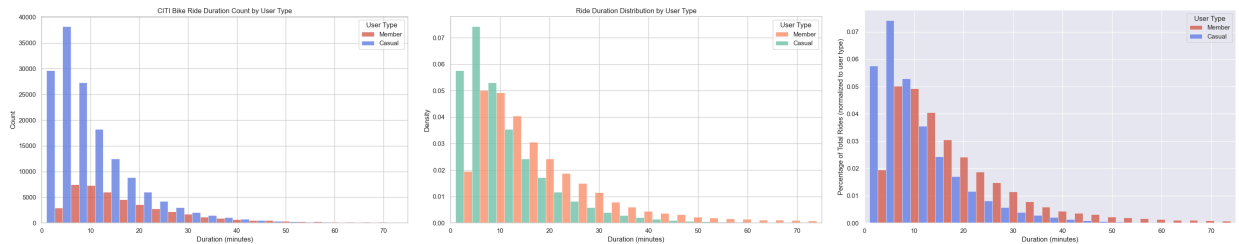
Hexbin CITI Bike Ride Duration vs Distance:

This plot was originally done with the entire 2024 year, but that ended up being too crowded and meaningless. I narrowed down to July to evenly sample commuters and vacationers. The original scatterplot was still too packed and the overlapping points detracted from the meaning of the graph. Therefore a Hexbin plot was used so the higher density and correlation could be seen. The final graph had its coloring changed to better match the websites theming.



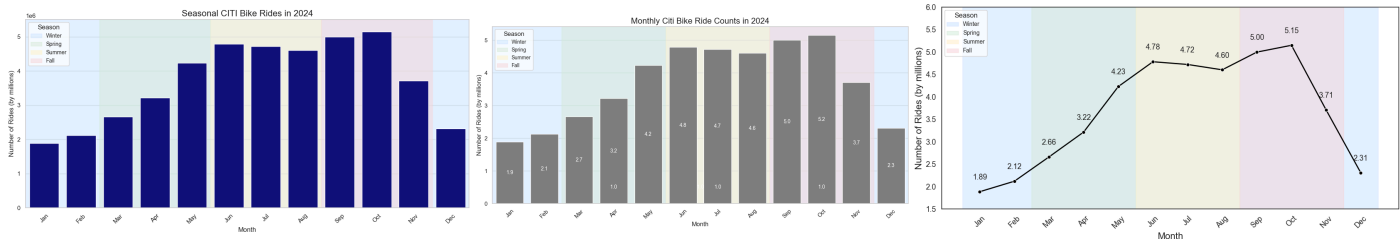
July CITI Bike Ride Duration Density:

The following was an attempt to showcase the most used duration of rides based on user type. The first was a simple diagram of density of each duration of rides (binned). However because of the lack of balance between Member and Casual user type, it ended up detracting from the message, it was confusing that there were more casual users. Therefore I normalized to each user type, this allowed the trend of more members using longer ride types and casual users having shorter durations. The final was to fix coloring to match the website.



Seasonal Trends of CITI Bike Ride Counts in 2024:

We wanted to visualize the changes of CITI bike usage over the seasonal changes. We knew this data was significant just based on the amount of data available for the summer months. Starting with a bar graph, we felt like the message was not being conveyed clearly and the bar graph was taking up too much of the visual field and the seasons were not clear. We changed coloring and added labeling, however it still did not convey the seasonal changes, just mainly the monthly differences. We finally settled on a labeled line graph. The difference in usage was clear and the seasonal markings in the back were clearly displayed.



The alignment between subway stations and Citi Bike stations:

Instead of scoping our focus to strictly bike transportation, we were interested in investigating potential trends in the distribution of Citi Bike stations and MTA subway stations and highlight them in our project. To achieve this, we used a) a dataset with the coordinate information for all Citi Bike stations and b) a dataset with the coordinate information for all MTA subway stations. We performed reverse geocoding on these datasets to obtain neighborhood and borough information. Our initial plan was to generate a map overlaying Manhattan's MTA subway lines with the CITI Bike stations. However, we realized that would make for a convoluted map, so we pivoted to displaying distance from subway stations to their closest Citi Bike station.

Heatmap regarding socioeconomic status (mean income level):

While our project's primary focus is transportation, we wanted to perform an exploratory analysis to identify potential associations between socioeconomic factors and Citi Bike station distribution. We considered performing our analysis based on neighborhood type (i.e., residential vs. commercial) but ultimately decided on median household income per neighborhood for 2023 as reported by the Citizens' Committee for Children of New York. We quickly realized the neighborhood granularity in the income dataset was notably different from that of the dataset used for neighborhood geography, so we had to standardize each dataset's neighborhood information before creating our heatmap.