The Citi of Bikes

Group A

Group members

Adit Anand ara2205@cumc.columbia.edu
Yanwei Li yl5403@cumc.columbia.edu
Varsha Varkhedi vv2392@cumc.columbia.edu

Proposed Visualizations and Analyses:

- Map of how the bike stations evolved over time. This can be a facet plot or an interactive plot, where users can choose what year to see.
    - Growth in number of stations over the years (NYC)
- Electric vs. normal bike distribution across stations + paths
- Member vs casual membership based on location (demographic). We can do a heatmap on a map to display this information.
- Trip type: We can do a heatmap on a map to display this information.
- Population analysis: the gender and age of people who usually use bikes.
- Analysis on the distance of rides and duration of rides.
- Network analysis of the bikes, show examples of the radius of how far they are often ridden. We can also plot a Circos graph to show which paths are most commonly taken.
- Socioeconomic analysis of borough type overlaid with station distribution
- Neighborhood type (residential, working, public spaces) vs stations
- The alignment between subway lines and Citi Bike paths.
- Membership type vs trip type
- Seasonal effects on number of rides and length of rides

Links to data sources/API:

The Citi Bike dataset schema can be found on the following information page, and the dataset files can be found on the following page.
We also aim to incorporate the following external sources of geospatial into our plots:
- NYC zoning district data
- Subway data
- Population demographics

Types of visualizations we will use:

1. Static images/charts

2. Network visualization
3. Maps

Hurdles and Challenges

One challenge we foresee is the change of data format and data types that were recorded within the database. There may be data of interest we would like to use across time, but does not have the data recorded before or after some time frame. Moreover, due to the differences in data collection and cleaning methods, there may exist challenges during data harmonization between historical and current data.

We are also worried that there may be missing data and erroneous data that we need to address. Moreover, we are unsure how many records will have all the data we are interested in. There is a chance that we will only end up with a very small subset of the data.
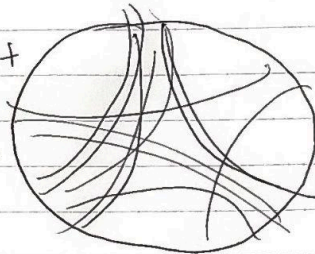
Brainstorm Session
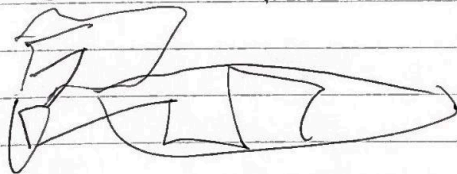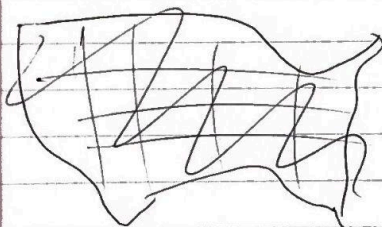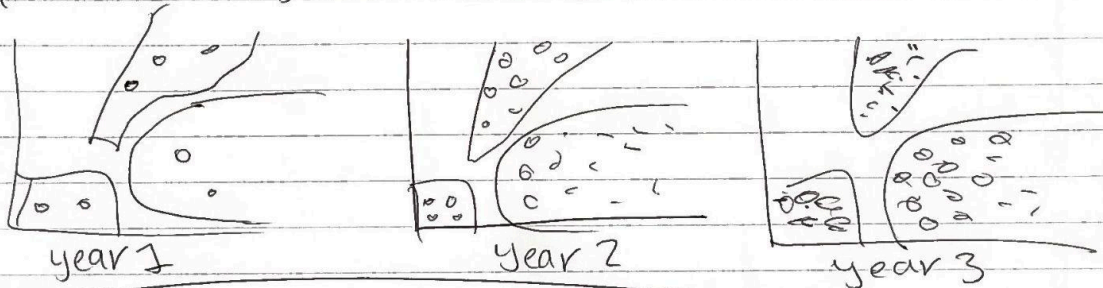
BRAINSTORM

Citi bike data

1. Geospatial
2. Network visualization
3. Static images

Rideable type
started vs ended at
Start long/lat
member or casual
Borough

Circos Plot
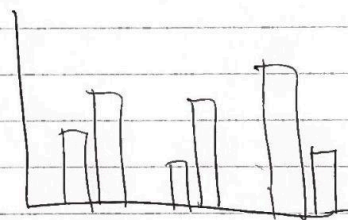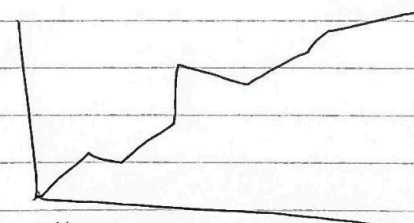of stations
of neighbor-
hoods for
trips

Facet Plot NYC

year 1

Year 2

year 3

Demographic/
Population Analyzes

~~Network~~ Geospatial
Membership type of Triptype
heat map.

# of Stations/
year