DS 4200

Group 6: Mental Health Data Report

Names: Lauren Ally, Sean Merkle, Melissa Rejuan, Kedaar Chankankar

Part 1: Additional Data

Lauren's Data:

Articles:

https://www.urbanet.info/urban-sanity/#:~:text=Urban%20conditions%20tend%20to%20increase,be%20reduced%20with%20appropriate%20planning.

- This article could aid us in providing facts to exhibit in our visualization to show the community the effects that certain things could have on people in urban areas. It focuses on "Urban Sanity" and how to create more happiness and sanity in urban areas. This supports not our mission, but the entire class's.

Datasets:

https://boston.maps.arcgis.com/apps/webappviewer/index.html?id=c9de58cb207f448a8212163812d91626

This map is similar to the tree canopy map we were provided when we first started the project. This is a different map since this instead maps out all of the public trees in Boston with a zoom feature and a selection feature incorporating tooltips. This provides us with a bit of inspiration as well as useful data.

https://data.nal.usda.gov/dataset/urban-tree-database

Provides access to a database with urban tree data throughout the country. This could help in comparing other cities to Boston if we would like to make that comparison. It is a bit outdated, but I do like the idea of having data from trees across the country.

Sean's Data:

Article:

https://greatergood.berkeley.edu/article/item/why_trees_can_make_you_happier

I chose this article because it is directly related to our topic. It pinpoints many of the psychological ways that trees can be beneficial to society. It talks about improving physical health and decreasing crime, which are actual things we can study/verify if we want to.

Database for urban trees:

https://www.kaggle.com/datasets/nycparks/tree-census

This dataset is a tree census for NYC, which we thought was relevant because though it is not exactly the same, it does share a lot of similar traits to Boston. This could be good for comparisons, making for interesting subplot based visualizations.

Melissa's Data:

Articles:

1. https://yaleclimateconnections.org/2023/02/the-little-known-physical-and-mental-health-benefits-of-urban-trees/

This Yale article discusses the health benefits of urban forests based on various studies:

- https://www.sciencedirect.com/science/article/pii/S0169204614002941?via%3Dihub

    - Found that patients that have a view of a tree through their window had significantly shorter recovery times post gallbladder surgery.
- https://www.science.org/doi/10.1126/science.6143402
    - Individuals diagnosed with depression in boroughs with higher urban tree density found to have lower antidepressant prescription rates.
- https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0193254
    - Students on campuses with greater tree cover perform better academically.

2. https://www.outsideonline.com/health/wellness/naturequant-app-outdoor-data/

This Outside article discusses Peter James', an assistant professor in Harvard T.H. Chan School of Public Health's Department of Environmental Health, research about the mental health impact of trees. He mentions trees create long-term changes in the occurrence of depression, anxiety, cognitive decline, and chronic diseases.

3. https://canopy.org/blog/impacts-of-trees-on-mental-health/

This Canopy article discusses how trees can combat loneliness by facilitating conversation and providing space for gatherings.

Dataset:

- https://data.boston.gov/dataset/primary-street-trees-public/resource/5edb1d3e-7e16-4c6b-8d02-f7e5be7b653d

    - Contains primary street trees species, longitude and latitude of each tree, and season. This dataset can be useful to distinguish the exact types of trees that are planted in Boston.
- https://www.forestresearch.gov.uk/tools-and-resources/tree-species-database/
    - This database can provide information about specific tree species. Using this information, we will be able to determine what attributes/specific trees are beneficial towards mental health and filter out the Primary Street Trees Dataset.
- https://boston.maps.arcgis.com/apps/webappviewer/index.html?id=c9de58cb207f448a8212163812d91626
    - This is a map where you can locate a specific species of street tree in Boston.

Kedaar's Data:

Dataset:

- https://www.arcgis.com/apps/mapviewer/index.html?webmap=f656506d7ae34f378a5b4855cf44c093
    - This database is a map of the tree inventory in Northeastern. This can allow us to look at locations specifically on campus and can potentially allow us to talk to officials on campus about planting more trees in certain areas on campus.

Part 2: Data Report Rough Draft

Project Summary:

For our project, our group focuses on mental health and the effect that trees and open space could have on this. We have multiple datasets to help us show this through visualizations. Our goal is to take the data we have found or have been given and show what it says through different types of visualizations and different levels of interactivity. Since we already know that

it has been proven that tree canopies are related to better general health and less psychological distress, we definitely want to find areas that do not have a dense tree canopy and visualize that. Also, it has been proven that open space and having the ability to walk through trees could support the community's mental health. The plan is to dive deeper into this through our visualization and get the community more informed.

Mental health awareness is something that is becoming more and more of a public concern that people are becoming more comfortable talking about. So, if there is a way to show the community how something may be affecting them without them knowing, we think they should know.

We want the community to be able to identify which parts of the neighborhood are more likely to experience psychological distress and identify where there is the ability to walk through open space through our visualizations. We plan to look towards creative, eye-catching visualizations to make sure that a general individual would be able to easily understand our visualization but also remember the data being presented.

Overall, with all of our data, we have a pretty solid plan of how we want to inform the community as well as a few set goals to do so.

Data:

For this project, we have lots of different data that coils help us inform the community and identify places that might be more likely to face psychological distress. We have chosen four datasets that would be useful to us—all of which provide data about tree canopies and open space in Boston. Our data shows us more about the open space and recreational space available so that we could further assess the potential improvement for mental and physical health of the community.

In these datasets, there are a lot of different types of data too. With this, it gives us more room to create interesting charts and graphs that simplify these huge datasets for the community to understand. Some include coordinates, sizes, regions, and even park names.

The first dataset we chose has to do with tree canopies in boston and where trees are located. We will refer to this dataset as “Tree Data” and was found on Boston’s Open Data website (https://data.boston.gov/dataset/trees/resource/ca30d1b5-fbef-4420-a0fb-18a642cc4aa7). The data here allows us to see the locations of specific park and street trees, including geographical coordinates and classifications for each tree. Seeing the locations and classifications for each tree can provide valuable insight into how these trees can affect the mental well being of those that reside within these locations. With this data, we could look at making an interactive map, scatter plot, and many other things to visualize where trees tend to reside in Boston. The Tree data has five total columns containing quantitative and categorical data. The data is categorized like so: “Id” is categorical, “X” is quantitative, “Y” is quantitative, “ObjectID” is categorical, and “Type” is categorical.

The second dataset we chose is related to canopy changes and we could refer to it as “Canopy Assessment Data”. It was found on Boston’s Open Data website as well (https://data.boston.gov/dataset/canopy-change-assessment-2019-tree-canopy-polygons/resource/7d6d68ba-d7a9-4d93-8390-99ba7a4f7a3d). This data shows us a map of boston that displays the coverage of trees within certain areas. Looking at the coverage within the entirety of the city can be helpful when looking at mental health related data within areas of high tree canopy coverage, and potentially allow us to see a correlation between the coverage and the overall well being of those within those areas. We will research more into different charts to find the most effective type to use, but for now we are leaning towards something similar to a scatter plot, maybe even a

bullet or radar chart. The Canopy Assessment data has three total columns that are also quantitative and categorical. In the data, “FID” is categorical, “Shape_Length” is quantitative, and “Shape_Area” is quantitative.

The next dataset we chose to use for this project has to do with open space assets in Boston, where they reside, and what they include. We can refer to this data as “Park Asset Data” and this was found on Boston’s Open Data website (https://data.boston.gov/dataset/boston-park-assets/resource/56b3003f-f397-43a1-9321-cd46b4ad9097). This file contains a list of park assets owned by the city of boston. Having this information can allow Speak for the Trees to know where they are able to plant trees with city permission. This can also be utilized alongside the tree canopy map in order to look at which parks could use an influx of planted trees in order to improve coverage, thereby, in theory, aiding in the improvement of the overall well being of the residents. The Park Asset data has four columns that we categorized as all categorical. We said that “Id” is categorical, “Park” is categorical, “Location” is categorical, and “feature(s)” is categorical.

The final dataset chosen by our group also has to do with open space but has different data than the dataset mentioned previously. We could refer to this dataset as “Open Space Data” and was found on Boston’s Open Data website as well (https://data.boston.gov/dataset/open-space/resource/9d5d9470-ab6b-42dc-b4e2-59a2c5ce3cf5). This file contains ownership of numerous parks within Boston. The speak for the trees organization can utilize this to figure out where they are able to plant trees as well as who they need to ask for permission in order to do so. Furthermore, this can also allow us to look at mental health data within these parks to determine where trees can be planted in order to benefit this. The bolded column names are the columns that we think will be most useful to us in our

visualizations, and will be our main focus when assessing the public, open space in Boston. The Open Space data has nineteen total columns, but we have identified seven columns that we plan to use in our project. The entire dataset is a combination of ordinal, categorical, and quantitative data. Of the columns we plan to focus on, they are categorized like so: “ObjectID” is categorical, “SITE_NAME” is categorical, “Ownership” is categorical, “District” is categorical, and “TypeLong” is categorical.

The Data Clean-up Process and Challenges:

There are a few steps our group is planning to take when cleaning our data to use our multiple dataset most efficiently. Since we have so much information from each dataset, we plan on not using some columns from some sets. So, we plan to delete those columns once loading in to avoid clutter.

Another one of our first steps is to make ourselves a data dictionary. This way, if we ever get confused, we could always look back to see exactly what each column means in the dataset. It does take a bit to completely understand each dataset and what it contains due to the fact that there is not much information on this very clearly, so this is something we would like to accomplish ourselves.

A big challenge we will face is the sheer amount of data we plan to look at. To combat this, we plan to take things one step at a time so that we will slowly work towards our bigger goal.

Task Analysis Table:

Shown below is the task analysis table that our group made together at the beginning of the project. It is reordered according to priority and it is what we based our project goals off of.

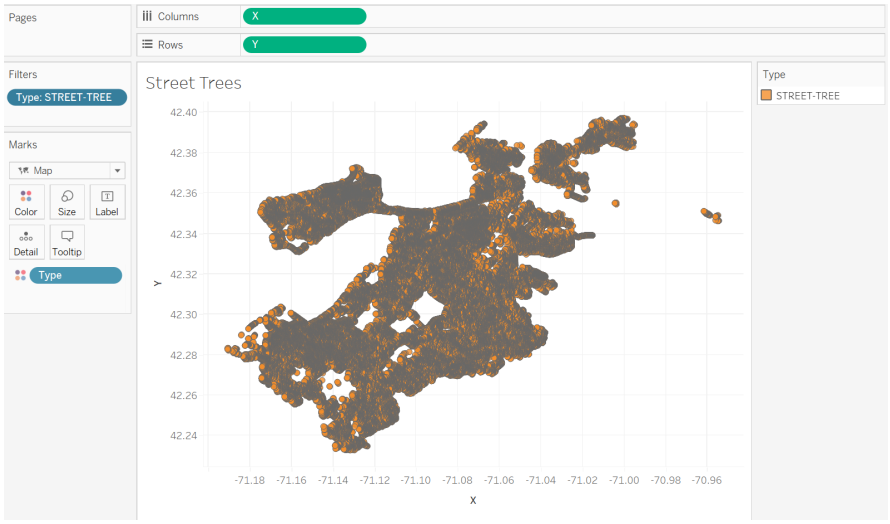
Index (ID #)	“Domain” Task	Analytic Task (Low-level, “Query”)	Search Task (Mid-level)	Analyze Task (High-level)

1	What are the benefits for trees and parks along the Mass Ave Corridor to provide space for improved mental health through green space and recreation?	Filter	Browse	Present
4	How have trees proven to have biological mental health benefits?	Retrieve Value	Browse	Present
3	Are there specific urban planning designs/ideas that are most conducive to supporting mental health (and if so how could the Mass Ave Corridor be improved)?	Cluster	Lookup	Present
2	Are there specific types of trees that are most conducive to supporting mental health (and if so how could the Mass Ave Corridor be improved)?	Cluster	Lookup	Present

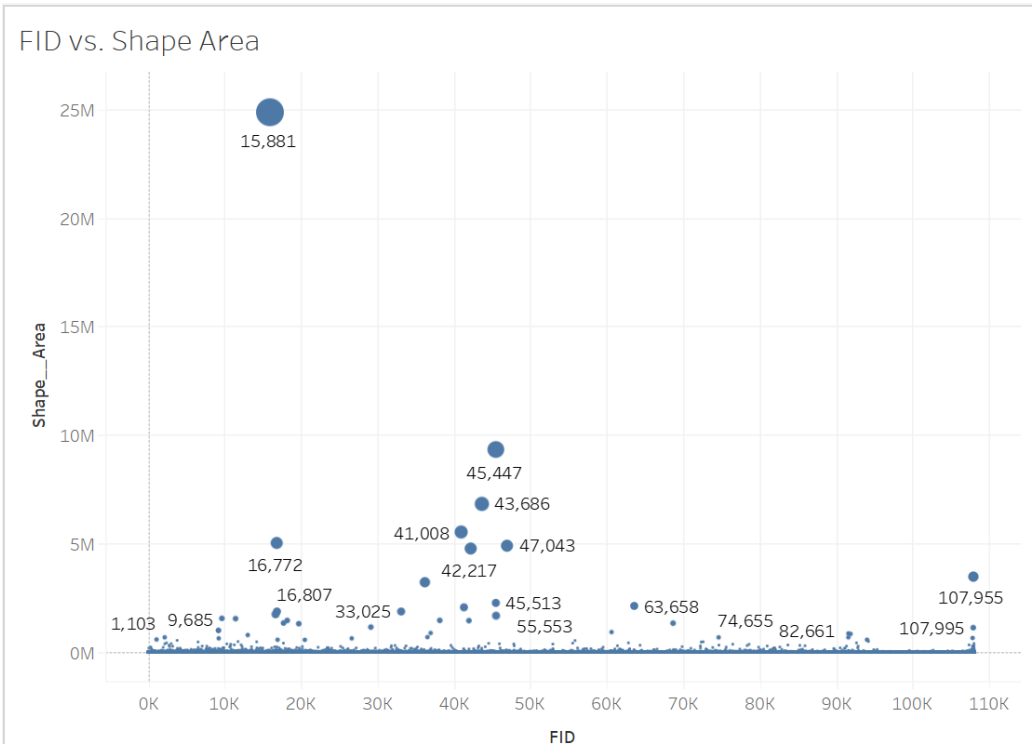
Observations and Insights:

1) 

In this visualization, we explored the amount of park trees in Boston. We looked at the 'Type', 'X,' and 'Y' columns in our dataset, which shows the type of trees (park trees or street trees) and the latitude and longitude of each tree. The visual encodings we used are points, horizontal and vertical positions, and shape. The points indicate exactly where the park trees are. The x-axis shows the latitude of each tree and the y-axis shows the longitude of each tree. We decided to create a map as a channel because it would visualize the locations of the park trees optimally. We found that park trees are scattered all over Boston, as Boston has an abundance of parks. However, in the next visualization, you can see that there are less park trees than street trees.

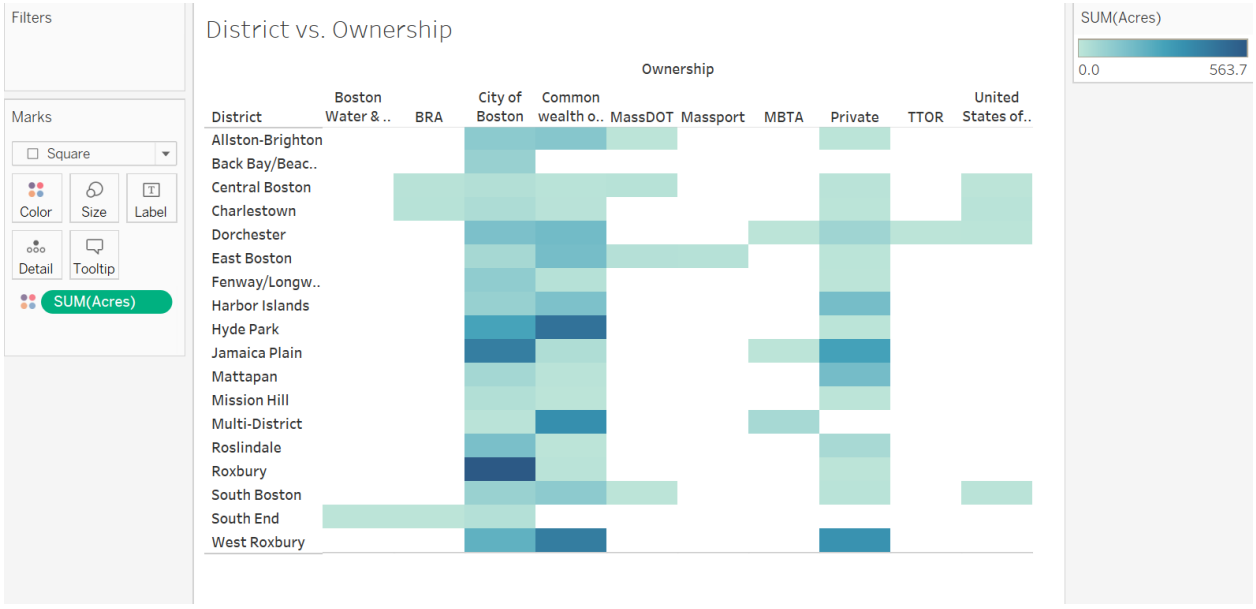
2) 

This visualization is derived from the same dataset of the last visualization and we used the same data from the same columns. We specifically looked at ‘Street Trees’ in the ‘Type’ column. In addition, we used the same visual encodings in the last visualization. We found that there is a greater abundance of street trees than park trees scattered throughout Boston.

3) 

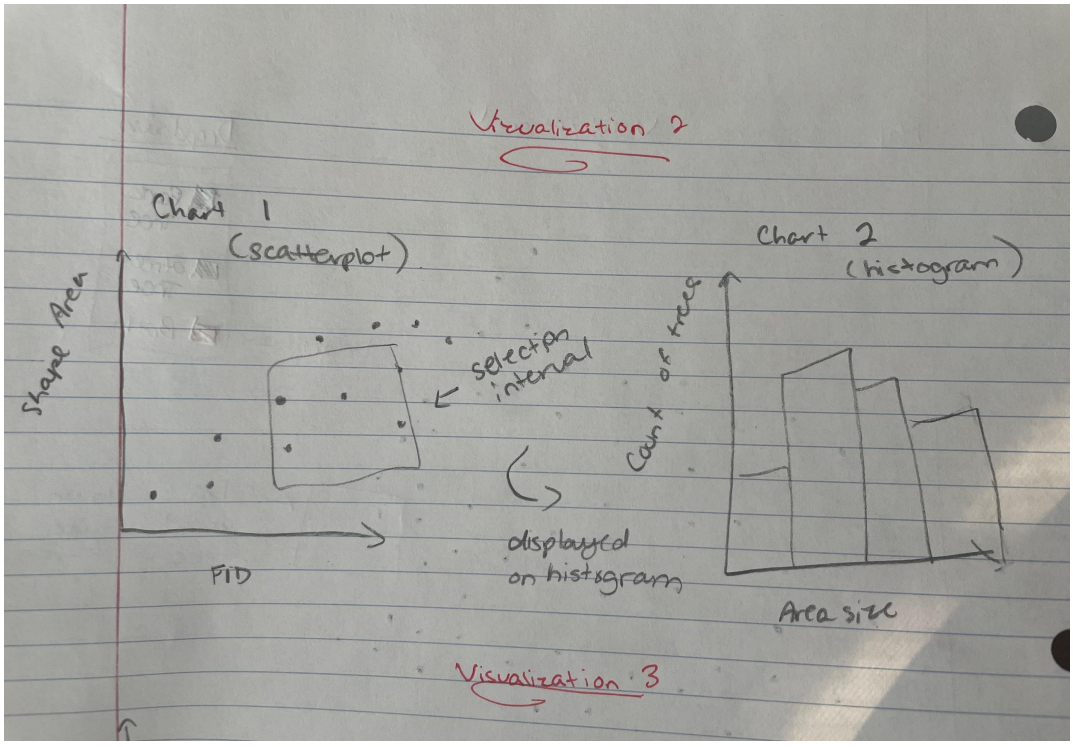
In this visualization, we explored the amount of area tree canopies take up. We looked at the ‘FID’ and ‘Shape_Area’ columns in our dataset, which indicates the tree canopy number and tree canopy shape area. The visual encodings we used are points, horizontal and vertical positions, and area size. Each point indicates the area of the tree canopy. The x-axis shows FID and the y-axis shows the shape area. The size of each point indicates how high the area is of the tree canopy. Therefore, greater the size of the point, higher the area. We found the tree canopy that has a FID of around 15,000 has the greatest shape area. Overall, we learned tree canopies cover a vast amount of area.

4)



In this visualization, we explored who owns land per district in Boston. We looked at the ‘District’ and ‘Ownership’ columns in the dataset. The visual encodings we used are shape, horizontal and vertical positions, and color. We decided that using squares as a marker is better than scatter points to distinguish between colors more easily. The x-axis shows different owners for districts. The y-axis shows each district. The intensity of color indicates the amount of acres

of land each district owns. Therefore, darker the color, the more land the owner has. We found that the City of Boston and the Commonwealth of Massachusetts owns the most land in Boston.

5) 

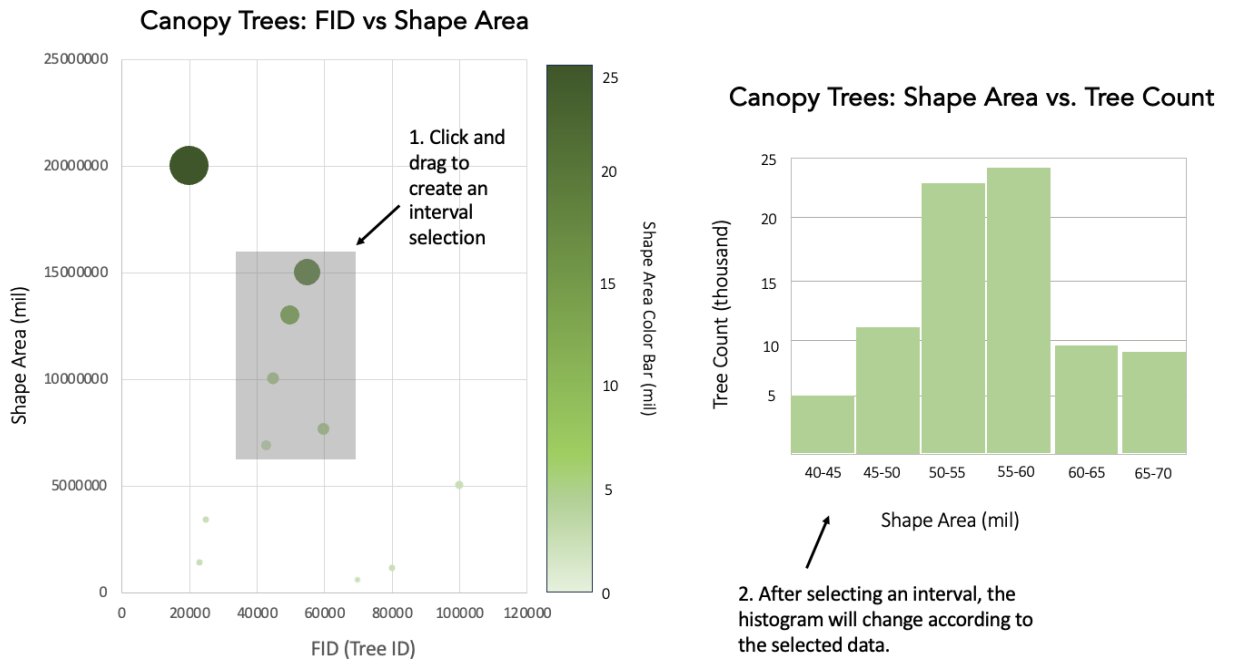
In this visualization, we will explore the shape areas of canopy trees. Although we have not visualized it yet through Python or Tableau, we expect to find greater counts of canopy trees in larger area sizes.

Conclusion

Our visualizations offered insights into the distribution of park and street trees in Boston, highlighting the abundance of park and trees scattered throughout the city's parks (street trees were found to be more prevalent). We also explored the extensive coverage of tree canopies and found that they cover a vast area. Land ownership analysis revealed that the City of Boston and the Commonwealth of Massachusetts owned most of the land. Future work could involve examining further data on tree species, health, and age, as well as general mental health impacts that these trees have in proximity to these areas. These findings lay the groundwork for further

investigations into urban greenery and land dynamics, guiding urban planning, conservation, and equitable access to nature.

Part 3: Polished Visualization Sketch



In this visualization, we chose to prioritize the following tasks: are there specific urban planning designs/ideas that are most conducive to supporting mental health and are there specific types of trees that are most conducive to supporting mental health? By answering these questions, we are able to cluster information about canopy trees, lookup information, and present it. For example, in the visualization above, the user is able to distinguish the shape area of each canopy tree. The size and color of each marker indicates a greater amount of shape area per tree in the scatter plot. The more area a canopy tree can cover, the greater impact it can have on mental health. The scatter plot also contains an interactive feature. Through brushing and linking, the user can select and highlight an interval of points to be presented on a histogram, showing the number of trees

per interval of shape areas. The histogram provides information about the number of trees that cover the greatest amount of area. With this insight, Speak for Trees can increase the number of large canopy trees in Boston if only few trees cover a great amount of area.

We chose the scatter plot and histogram design out of all our original sketches because this visualization allows the user to filter and query data to gain insight about trees, while also providing two different demonstrations of data. In addition to user autonomy and insight, the scatter plot and histogram are easy to follow. In the scatter plot, the user is able to easily distinguish between points based on color and size. In the histogram, the user can easily identify the lowest number of trees and highest number of trees based on area size. We also left small spaces between each bar in the histogram, allowing the user to differentiate each bar optimally.

We used powerpoint to create our final interactive scatter plot and histogram.

Part 4: Group Charter re-visit

After working together for a few weeks now, we have revisited the group charter. We have abided by our group roles and have been working well together for these past weeks. The guidelines we set with each other at the beginning of the assignment are still solid and we still stick to them. For group roles, we are not as stuck to them, and stick to splitting everything equally 4 ways. This is what came naturally to our team and works for us. We are getting more and more comfortable working together as a team and still hope to make a meaningful final project together. Each person, although contributing differently, contributes equally. We all listen to each other and heat out interesting ideas so that we never pass over someone's potentially creative idea.

Our discussion about this was very unanimous as we have always been very transparent with each other and respect each other. We understand that we each bring something different to the table and can work together to combine our strengths.

Part 5: Peer and Self-Evaluation Form:

We all completed the form separately.