Building a Browser for Boston's Blue Bike Behavior

Gavin Mooallem Alexander Chang Shreya Mitra

ABSTRACT

In addition to a large public transit system including trains and buses, Boston boasts one of the largest shared bike rental systems in the country, often just called Blue Bikes. With over 400 different bike stations across the Boston Metro Area (and Salem), there is a lot of data to be explored from all of these stations, including ridership, frequency, bike availability, and more. Building out a visualization to view which areas and stations are most or least used can help inform many things, such as where a new station could and should go, which stations need more or less bikes, and which routes are most popular and should have better bike lanes.

1 INTRODUCTION

Bike-sharing systems, such as Blue Bikes in the Boston Metro Area, have become an essential mode of transportation for many urban dwellers, providing a cost-effective, eco-friendly, and flexible option to navigate the city. To ensure the continued success and growth of such systems, it is crucial to understand ridership patterns, bike availability, and route preferences. This paper presents a visualization tool that aims to address these challenges to ultimately enhance the overall efficiency of the Boston-based Blue Bikes system. Our primary goal is for the visualization to provide support in two key areas. The first one is in identifying underserved regions in the Blue Bike network, which can help in planning for the expansion of bike stations or reallocating bikes to areas with higher demand. The second key goal is to understand popular bike transit routes, which can be valuable for urban planners and policymakers to prioritize improvements to bike infrastructure, such as bike lanes, signage, and safety measures. It's important to support these functions as understanding ridership patterns can help ensure bikes are readily available when and where they are needed, improving user satisfaction and encouraging more people to adopt bike-sharing as a mode of transportation. This shift towards bike-sharing can reduce traffic congestion, lower greenhouse gas emissions, and promote an overall healthier lifestyle. Additionally, addressing underserved regions contributes to equitable access to bike-sharing, fostering social inclusivity and promoting sustainable urban mobility. The end users of our visualization tool are diverse, including bike-sharing operators, urban planners, transportation policymakers, and bike-share users. Our visualization can assist operators, for instance, in optimizing their operations and resource allocation, promoting a responsiveness to demand that ensures users are able to find available bikes at their preferred stations. Additionally, urban planners and policymakers can use the insights provided by the visualization to inform the development of bike-friendly policies and infrastructure. Bike-share users can benefit from the tool by identifying bike availability patterns and popular routes, assisting them in planning their trips more effectively. The data used for our visualization is sourced from the Boston Blue Bike official website, specifically trip data from September 2022. This dataset encompasses information on trip duration, start and end times and stations, as well as anonymized user data. Through the visualization of this data, we aim to provide valuable insights into the usage patterns and preferences of Blue Bikes users, ultimately contributing to the optimization and growth of bike-sharing systems in the Boston Metro Area and beyond.

2 RELATED WORK

Due to the publicly available nature of most bike-sharing systems around the world, there is a wealth of available research done on the subject. This ranges from incredibly in depth analyses of the last-mile problem to the optimal way for cities to design their bike systems and more. These papers can sometimes contain data on multiple bike-sharing cities, although the papers discussed in this section each focus on one.

2.1 Visual Analysis of Bike-Sharing Systems [2]

This paper, discussing the Citi Bike system used in New York City, offers multiple complex visualizations that helped inform the early designs of this project. The paper works to aggregate, summarize, and visualize the data provided by the city's bike program, showing metrics like frequency of checkout by station, a matrix illustrating the stations' checkout vs. check in data, and a map of station outages over time. This work, specifically the matrix for each stations data and overlaying the information on a map, was an influence in how we decided to store and represent our data. Our project also keeps a matrix to maintain checkout and check in information, although it's slightly more simplified. We also want to replicate the idea of showing the station locations overlaid on a map of the city they serve, albeit over a simpler, less detailed map for efficiency reasons.

2.2 Travel Patterns of Free-floating E-bike-sharing Users Before and During COVID-19 pandemic [1]

This paper, discussing the bike-sharing system used in Seoul, South Korea, contains an almost overwhelming amount of data and visualizations. The data for bike trips are broken up among many different categories, including rider age, gender, and neighborhood, as well as data about the trips' speed, duration, distance, directness, and more. Overall, most of the visualizations in this paper don't apply to our project, but one visualization that was very intriguing is one where the researchers plot the journeys taken by riders between station on top of the map. While the data given by Boston's Blue Bikes is very different from Seoul's the core idea of plotting the routes between stations is one we hope to implement. Because of the limited nature of our data, such as not having exact user routes, only starting and ending stations, our visualization may look less full than this paper's.

3 USE CASE

For the ecologically conscious commuter, oftentimes the most affordable and least environmentally damaging option for transportation is the bicycle. Boston's network of bike lanes and low-traffic side streets creates the ideal environment for such an individual. Despite this, bike ownership is not always the optimal option in the city of Boston. It can be difficult to find places to store bikes, and if they are left out in the open there is an ever-present risk of theft or damage. For some, the one time investment cost of a bike can be difficult to justify. The solution: Blue Bike, Boston's publicly owned and operated bike share program. The perfect option for those who want to transport themselves via bike, but do not want to own one for any number of reasons. Its low cost and wide-spread availability makes it the perfect option for many, with one glaring caveat: the empty

Blue Bike station. Unfortunately, it can be all too common to find your local Blue Bike station completely empty of bikes, hanging Blue Bikers out to dry.

By visualizing Blue Bike ridership data it is possible to improve bike availability for riders. Indeed, Blue Bike employees can utilize such a visualization to understand where the demand for bikes is highest. This tool will allow for the visualization of demand across different stations over the course of the month of September (the busiest month of the year for Blue Bike ridership). Additionally, it will provide insight into the connections between different stations and how they interact with each other. This can aid in decision making with regards to where new bike stations should be added or expanded.

3.1 Understand underserved regions of Boston

Blue Bike stations are located throughout Boston and it's neighboring regions, stretching as far as Salem, Massachusetts. The extensive network of bikes and stations aims to provide adequate coverage and access to the different neighborhoods. While the existing system contains over 400 stations and several thousand bikes, there still exists a possibility of lack of access in certain communities.

3.2 Understand popular bike transit routes

It is to be expected that certain stations tend to be popular for departures and arrivals. It is likely that commuters who regularly take Blue Bikes will typically leave from stations in residential zones and arrive at commercial / business areas during the work week. It is also likely that tourists who take Blue Bikes will bike from housing hubs to areas of interest throughout the city.

4 Data

The data for this visualization comes from the Boston Blue Bike official website, specifically trip data from September 2022. Boston Blue Bike publishes data regarding trips and stations on a monthly basis, and the original pre-cleaned data set can be found at bluebikes.com/system-data. This data includes information about the duration of the trip, the start time and station, the end time and station, as well as some simple data about the user who took the trip. All of this data is anonymized, so any potentially identifying information is removed from the data set before it is published.

There are very few ethical considerations that need to be made when utilizing this data. Since the data is already anonymized and does not contain personal information regarding the user who took a particular trip it would nearly impossible to use this dataset to identify a particular individual. There is a small possibility that a determined individual could track down a particular rider if they know some combination of information about a rider's exact start / end time or station. However, our visualization tool will not allow for exploration of data regarding individual trips so it will not be useful for such a nefarious goal.

Data provided from the Blue Bike system data does not contain any missing values, every row had complete information (except for the postal code column which was irrelevant to this data visualization). Additionally, data is consistent across categorical attributes (ie Station Name) and requires no cleaning to that end.

The data was processed to better understand activity across the different stations and dates. The original data set was modified such that individual trips were removed. Instead, each station was given its own row. The original data set of raw trip data was then parsed to get the number of trips to and from each station for each day of the month. The region / neighborhood of each station was also added to the data set.

Additionally, multiple new data sets were created. These data sets contain a matrix where each station has a row and a column. The data in the new data sets represents the number of trips between two particular stations, where the row represents that start station and the column represents the end station.

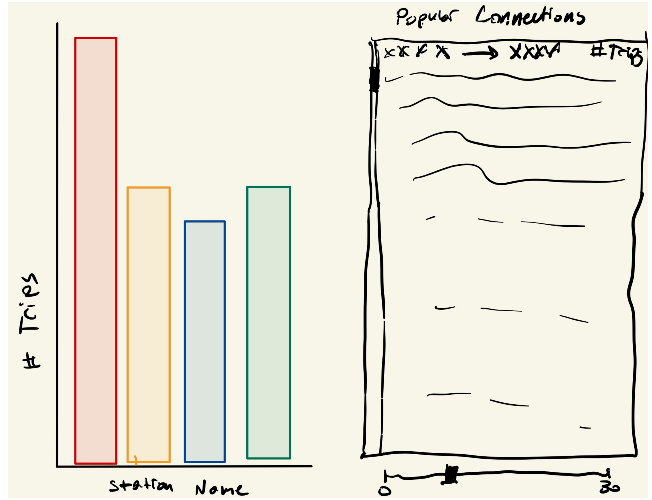

Figure 1: The initial draft.

5 Design Process

Figure one shows the initial draft of the project. This illustration has two views to it. The left side shows a bar chart that would display the total number of rides at each station and show the number of rides between different stations. There is also a slider at the bottom of the page that allows the user to adjust the date for the which the data is being shown. Ultimately, this design proved to be far too simplistic to convey the information necessary to draw useful conclusions. However it includes key components that will be used in later designs, such as a view for viewing trips by station, a way to select particular dates to view data for, and a view for data between different stations.

The second draft of the project is shown in figure two. Here the design has evolved to include a full map of all regions where there are Blue Bike stations present. Each station is marked on the map using a red dot. The date selection tool has evolved to be a series of boxes that each represent a day of the month. The boxes are colored based on the number of trips taken on each particular day. The same station to station view was carried over. Overall, there is significant improvements from the first design to this one. However, we identified improvements that could be made to the station to station connection view.

Figure 3 is the third draft of the visualization. The map and date selection tools have carried over from the previous iteration. However, stations on the map are now color coded to represent the number of trips at a station for a particular day. Additionally, dates are selected using a brush tool to make date selection easier. The largest change comes from the station to station view. This view is now portrayed as a network diagram to show connections between different stations.

Figure 4 shows the final implementation of the visualization. The most significant change from the third draft of the visualization is the replacement of the network diagram with the bar chart. It was decided that the network diagram was redundant as the stations are already visualized on the map. Additionally, it was found that when displaying a large amount of data the network diagram was nearly useless. We found instead that a bar chart was much more useful for displaying information. The bar chart is a very effective way to compare numerical data, in our case trip data across stations. Users can easily see which stations have more trips compared to others.

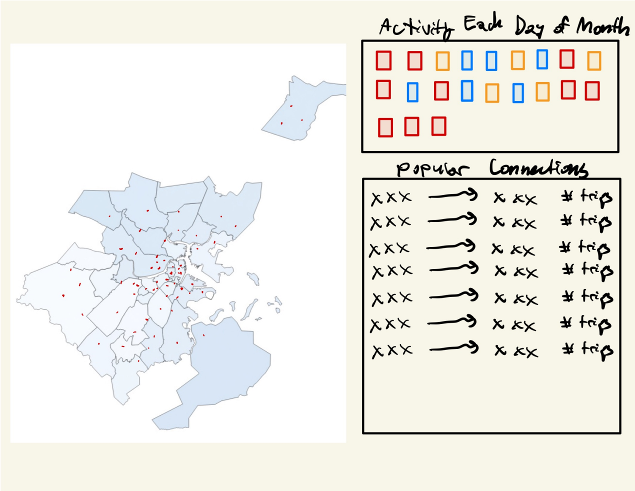

Figure 2: The second draft.

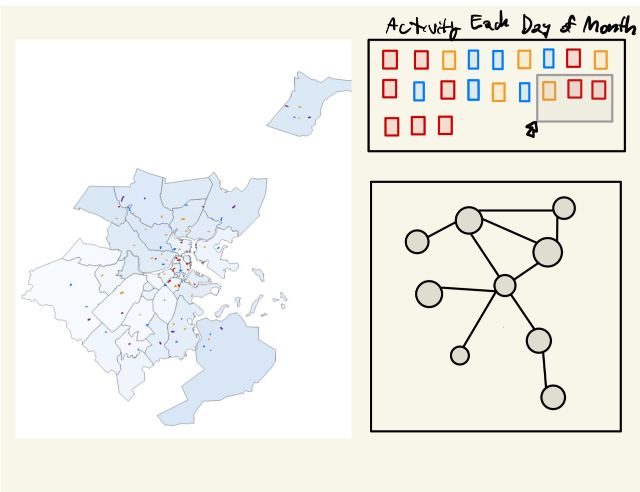

Figure 3: The third draft.

Additionally, we added the option to order by trip count or station name to make it easier for users to identify a particular stations within the bar chart. Even on large amounts of data, it is easy to compare the relative quantity of trips across different stations. To replace the ”connectivity” aspect that the network diagram provides we modified the map such that it filters only stations that are shown on the bar chart, and it draws a line between stations when a bar is highlighted.

5.1 Usability Testing

As part of the design process, usability testing was conducted on a prototype version of the visualization.

5.1.1 Tasks

We asked participants to complete the following tasks:

1. Find the station with the highest number of trips during the last week of the month
   (a) Follow-up: From this station, which station has the highest number of connections?
2. Find the Seaport Blvd at Sleeper St station
   (a) Follow up: How many station have at least 120 trips between Seaport Blvd and the station?
3. Pick two distinct stations, determine the number of trips between them for the entire month.

We collected data regarding the time to complete each task and the correctness of their responses. We also asked participants to rate their confusion on a scale of 1 to 10, where 1 is not confused at all and 10 is entirely confused.

5.1.2 Results

The results of each task and subtask are shown below. Here DNF means that the participant did not finish the task (given a time limit of one minute per task). Users were provided an explanation on the correct answer to a task before completing follow up tasks if they did not finish. Each participant was asked to complete all the tasks in the order listed below.
Result for task 1:

	Time	Correct	Confusion
P1	41s	Yes	7
P2	DNF	No	10

Result for task 1a:

	Time	Correct	Confusion
P1	15s	Yes	5
P2	40s	Yes	10

Result for task 2:

	Time	Correct	Confusion
P1	30s	Yes	2
P2	18s	Yes	1

Result for task 2a:

	Time	Correct	Confusion
P1	21s	Yes	1
P2	18s	Yes	2

Result for task 3:

	Time	Correct	Confusion
P1	23s	Yes	5
P2	15s	Yes	8

Through our usability testing we found that users were often confused initially on how to interpret the visualization. It was misunderstood what the color of the stations represent. This led to the inclusion of the legend in the top left corner of the final map. Additionally, users did not understand the calendar as it lacked dates. To solve this, our final visualization includes the actual day of the month represented by the box and displays a tooltip with more information when hovered. We also found that users were more comfortable with the visualization as they used it, but were very confused at the start. To help assist with initial confusion we added the ”Select a station” text to the screen to guide users down certain paths of interactivity.

6 Final Design

The design of the final visualization is shown in figures 4, 5, and 6. The final iteration of the design has evolved to feature the same map view on the left side, but now displays a calendar view and a bar chart view on the right side. Like many of the previous iterations, the map on the right side features a point to represent each of the Blue Bike stations. The stations are color-coded to represent the number of trips to and from the given station. The map now features a legend in the top left corner to help the user better understand the meaning of the color. The calendar view in the top right is color coded based on the trips for that particular day. Users are able to click on one or several days in order to view the data for that particular day. The bar chat seen beneath the calendar appears once a user selects a station. This bar chart shows the number of trips between the selected station and stations that meet a set threshold. The map view is also adjusted such that only stations that meet the threshold are rendered on the map. The threshold can be changed using the slider located above the bar chart. When a user highlights a bar in the bar chart, a tool tip appears that says the name of the station as well as the number of trips. Additionally, a line is drawn on the map between the currently selected station and the station the highlighted bar represents. This interaction also works in the opposite direction, where if a user highlights a station on the map then the corresponding bar will be outlined in red.

The mark of the map view is a point. Each point represents a single station. There are a few associated channels with the mark. The first channel is the color hue. The hue represents the number of trips to and from the station. The color scheme is divergent from the mean, going from blue to white to red. This allows for users to quickly and easily distinguish between which stations receive above average activity and which receive below average activity.

The marks of the calendar view are the squares that represent a given day. There is just a single channel associated with the mark, which is the hue. The hue is a sequential color scale from white to orange, and the deeper the orange the more trips there is on the day. The marks of the bar chart are the bars themselves which represent the number of trips between a station and the currently selected station. The channels here are size and hue, and both channels relate to the number of trips at a particular station. Here a larger size represents more trips, and a more orange hue also means more trips. In order to accomplish the task of identifying under served regions, users can look to the map. The map shows where every Blue Bike station is located, so it is very easy to quickly see the spread and distribution of stations. Users can extrapolate which geographic regions of the greater Boston area lack Blue Bike stations. In addition to this, it is easy to recognize areas of relatively high demand based on the color of the station. Areas with low bike station density and high station usage can be easily identified as areas that potentially require additional stations.

To understand popular bike transit routes, a user can select a particular station. From there the visualization shows the most popular stations travelled to from the selected station. A user can highlight a particular bar to visualize geographically the travelling done between these two stations. Publicly available data does not allow for the visualization of actual routes, but this tool is still useful in understanding where people are going from a particular location.

7 Discussion

Our visualization tool serves as a means for understanding the demand and patterns of bike-sharing usage. However, it doesn't entirely address the domain problem, as there are a few limitations. While the visualization helps identify underserved regions and popular bike transit routes, it lacks certain elements that could provide a more comprehensive understanding of the system. One limitation is the absence of real-time data, as our visualization is based on historical data from September 2022. Additionally, our tool does not take into account external factors such as weather, special events, or maintenance schedules, which could influence bike availability and

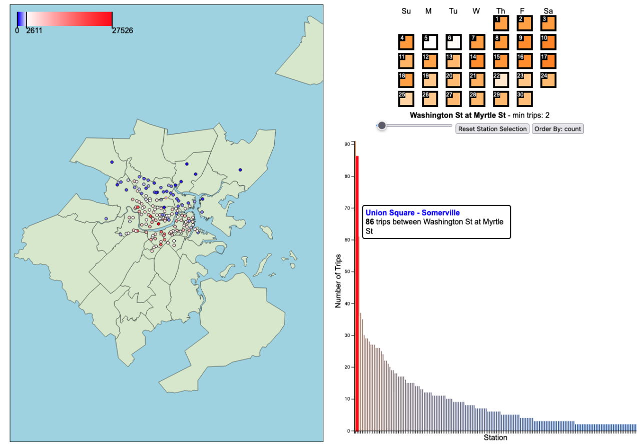

Figure 4: The final design.

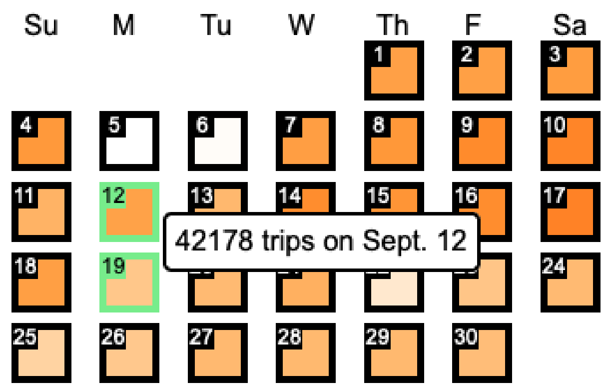

Figure 5: The calendar in the top right. Here Sept 12 and Sept 19 have been selected.

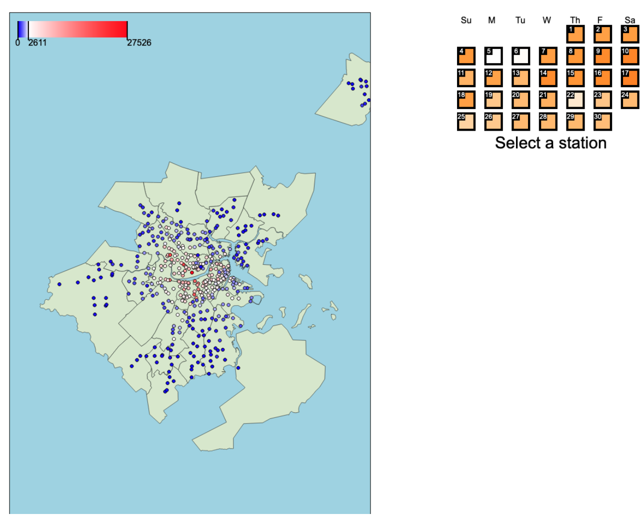

Figure 6: The final design before a user has selected a station.

ridership. The visualization also relies on data limited to start and end stations for each trip, lacking precise user routes, which would allow for a more granular analysis of bike usage patterns. To enhance the effectiveness of our visualization tool, several improvements could be made in the future. First of all, incorporating real-time data would enable users to make more informed decisions and respond to immediate needs, such as bike redistribution or identifying potential station outages. Secondly, integrating external factors like weather, special events, and maintenance schedules would provide a more holistic understanding of bike-sharing usage patterns and enable better planning and resource allocation. Additionally, integrating GPS data, if available, could help determine the exact routes taken by users, allowing for improved insights into infrastructure needs such as bike lanes, signage, and safety measures. Beyond the data limitations, changes to our visualization tool could include the addition of interactive features, such as dynamic filters and sliders, allowing users to customize their analysis based on specific time frames, stations, or trip duration, to promote more focused and relevant insights. Another enhancement would be the implementation of clustering techniques to identify patterns and trends in bike usage more clearly. Clustering could reveal underlying relationships in the data, such as similarities between stations with high demand or common characteristics among underserved areas. Furthermore, incorporating advanced visual analytics methods, such as heat maps or flow maps, could provide a more intuitive and efficient way for users to interpret the spatial distribution of bike-sharing usage and identify areas that require attention or resources. Ultimately, these improvements would better support the goals of identifying underserved regions, optimizing bike-sharing systems, and promoting sustainable urban mobility in the Boston Metro Area and beyond.

8 Conclusion

While Boston's bike sharing system isn't perfect, tools like this visualization can help make it better by eliminating trips to empty/full stations, identifying underserved regions of the city, and help show route popularity. To make this visualization, publicly available Blue Bike data was cleaned and processed for the month of September 2022 and overlaid on a map of the city. The map was made interactive, with stations showing connections that they have to other stations around the city and the ability to select certain dates to filter. Design of the visualization evolved from having a complex network graph, to a more simple and understandable bar chart, with a variable threshold for adaptive searching. More tools like this are needed in the future to help address gaps in service and other issues within the Blue Bike system, but this visualization is a good start.

9 Appendix

9.1 Data Abstraction

- Data set type - Table
- Row data type - Item
- id - Attribute
- name - Attribute
- latitude - Position
- longitude - Position
- region - Categorical attribute
- X_start_trips - Sequential quantitative ordered attribute
- X_end_trips - Sequential quantitative ordered attribute
- X_total_trips - Sequential quantitative ordered attribute

9.2 Task Abstraction

9.2.1 Domain Task: Understand under-served regions of Boston

- Identify areas with a low number of bike stations and/or availability in order to determine regions in Boston that require expansion or redistribution of bike-sharing resources.
- Compare the distribution of bike stations and usage across different neighborhoods to better understand disparities in access to bike-sharing facilities and inform future decision-making processes in order to address these gaps.
- Discover patterns or trends in bike usage over time in underserved regions to inform the prioritization of future investments in bike-sharing infrastructure and resources.

9.2.2 Domain Task: Understand popular bike transit routes

- Identify the most frequently used routes between bike stations in the dataset to allow users to recognize the most popular routes, as well as learn where additional resources could be allocated.
- Compare the relative popularity of different stations and routes, both within a neighborhood and across the broader network, in order to help users prioritize which stations should receive more or less resources based on their demand.
- Discover patterns, trends, or anomalies in the usage of bike stations and routes across various time frames. Understanding these patterns can help inform future planning related to bike infrastructure, including potential expansion or targeted marketing efforts.
- Provide an aggregated view of the overall popularity of bike transit routes across the network, allowing users to quickly identify areas that experience especially high or low demand.

References

[1] S. E. Choi, J. Kim, and D. Seo. Travel patterns of free-floating e-bike-sharing users before and during COVID-19 pandemic. *Cities*, 132(104065):104065, Jan. 2023.

[2] G. N. Oliveira, J. L. Sotomayor, R. P. Torchelsen, C. T. Silva, and J. L. D. Comba. Visual analysis of bike-sharing systems. *Comput. Graph.*, 60:119–129, Nov. 2016.