Visualizing the State of MBTA Rapid Transit Ridership

Nop Lertsumitkul* Kai Goerg† Clayton Yan‡ Mikaela Gordon §

ABSTRACT

The public transportation system in a city such as Boston is vital for its citizens. By understanding the movement of people in Boston, specifically ridership on the T, we can better design our transportation systems and allocate funding to the lines that need it most. We hope to create a visualization to help understand the current state of the MBTA's train system (the T), how it serves the Boston community, and what we can do to make it more efficient and reliable for those it serves.
`https://github.com/DS4200-S22/final-project-visualize-the-mbta`

1 INTRODUCTION

Our motivation behind this project is to better understand the T system and its shortcomings by looking at data over time across each of the different train lines. This would help to pinpoint and find trends that could be addressed by allocating resources more efficiently. Correctly allocated resources could be used to provide a better transportation experience, as the MBTA would be adapting to the needs of riders everyday. This is meant to start the conversation on how we are able to leverage data to optimize the transportation experience for the riders, operators, stakeholders, and the entire Boston community.

2 RELATED WORK

Visualizing Public Transport Systems: State-of-the-Art and Future Challenges [3]:
This project takes from literature about public transportation as well as interviewing experts in the field. The paper is more theory-based and discusses ways of improving public transport systems. It brings up issues such as improving the visual encoding of landmarks and hhttps://www.overleaf.com/project/62156df498aa56060b9d38f7as one of the visualisations being a map showing time taken to travel from a starting point. Our group can take ideas from this to come up with more theory-based visualisations.

Visualizing Mobility of Public Transportation System [5]:
This project visualises the movement of public transport passengers and uses the help of transportation researchers to perform analytical tasks on the findings. This project has 3 main visualisations: isochrome maps, isotime flow maps, and OD-pair journey views. In this project, they created a mobility model that analyses passenger mobility to estimate aspects such as waiting time, riding time, and transfer time. The idea of creating a model using the data is a nice inspiration that we could use.

Visualizing MBTA Data [1]:
This project uses a line graph with time on the y axis and current position in the system of the trains to show a large overview of the trips trains take in a day. The interactive aspect of this graph which is linked to a map of the train system is something our group can take inspiration from. This article is also similar to what we are trying to achieve, being of similar focus and on the MBTA.

Identifying passenger flow characteristics and evaluating travel time reliability by visualizing AFC data: a case study of Shanghai Metro [4]:
The paper's main objectives were to analyze passenger flow characteristics and evaluate travel time reliability for the Shanghai Metro network by visualizing the automatic fare collection (AFC) data. The paper showed several visualization that we can take inspiration from. Since our data set is large, we can draw many conclusions from the data in a wide variety of ways, just like this paper does. The paper uses a mix of graphs, maps and charts, using color, shape, size and transparency as ways to show data.

A survey of Traffic Data Visualization [2]:
This paper introduces the basic concept and pipeline of traffic data visualization, provides an overview of related data processing techniques, and summarizes existing methods for depicting the temporal, spatial, numerical, and categorical properties of traffic data. This is particularly useful to us because this is exactly the type of data we will be using for our visualization, and we can use this article to learn methods on how to display our data. There even is a section on the Boston subway system which is our focus and so we can draw inspiration from that.

3 USE CASE

This visualization would be able to help those that are working in the transportation agencies to understand how, or which lines are effective or why people may or may not take that line. It is something that agencies are moving towards to leverage their data. In the agency's data/finance division, we can integrate this visualization to help aid in making decisions on how the agency will upgrade and maintain their vehicles based on ridership and popularity as well as needs. The visualization will show the lines as well as stops with details on the usage of them. The MBTA already collects this data, and so it would be useful for both the Boston community and transportation planners to be able to make sense of it in an intuitive and interactive way.

4 DATA

(https://mbta-massdot.opendata.arcgis.com/datasets/MassDOT::mbta-rail-ridership-by-time-period-season-route-line-and-stop/about)

The dataset above contains detailed ridership data for heavy rail and light rail and comes from the MBTA's open data portal. Attributes in the dataset include the average ridership for a typical day during the rating by route/line, direction, stop, day type, and time period. Data is available for Fall 2017, Fall 2018, and Fall 2019. This is reliable data that is collected by the MBTA themselves and can be used to find the flow of people at various stations and lines.

*e-mail: lertsumitkul.n@northeastern.edu
†e-mail:goerg.k@northeastern.edu
‡e-mail:yan.cl@northeastern.edu
§e-mail:gordon.mik@northeastern.edu

5 Task Analysis

Task ID #	Domain Task	Analyze Task	Search Task	Analytic Task
1	Needs visualization to show ridership based on different metrics, to find out which stations are being used the most and when that is	Discover	Explore	Compare
2	Visualization will allow MBTA to figure out where to send resources and the best time to do maintenance during the day based on when it is less busy	Discover	Locate	Identify
3	Wants stats for total ridership, flow, average ridership over a week for a given line, and ways to compare the data by time and day of the week, and more largely comparing different years	Present	Explore	Compare
4	Dashboard showing data based on given criteria such as line, station, or time of day	Present	Lookup	Identify

Final Review

Our primary consumer for our visualization will be MBTA employees and managers. The primary purposes of the visualization would be to discover patterns in data and present them in a way that can be easily used for decision making by higher management. The first two tasks are to do with discovery by gathering data. That data is then used to create meaningful comparisons and identifications that are presented to management to be used in decision making.

6 Implementation Plan

For our implementation plan, we intend to use four visual encodings with data from the CSV file. We do not plan to use an API. These visualizations include a map, a bar chart, a table, and a parallel coordinates graph. The map would be the main visualization showing all the stations and being connected to each of the three other encoding.

To implement this, we would extract data from the CSV file. With this data in d3, we will create the map then the three other visualizations which will be linked to the map using listeners, brushing, and linking. For our map, we will implement zoom and brushing over the stations on the map. A way to implement this would be using a scatter plot in D3 to represent the stations and add a brush for them. When brushing over stations on the map, the other three encodings would adjust to the data of the brushed stations. For the parallel coordinates graph, hovering or brushing over each line would make it light up to make it easy for the user to track each line. The bar chart would also have a hover tooltip that shows the user more details about the segment such as the exact rider count.

7 Visualization Design

Our final visualization is a dashboard consisting of an interactive map of MBTA Rapid Transit (T) stations and connected tables and graphs. The main map includes panning, zooming, and the ability to select stations. Based on the selection, there is a table that updates with the total ons, offs, and average flow for the selected station. In addition, there is a pie chart showing the distribution of ridership at the selected stations over the different time periods. Lastly, we have a vertical line chart with all of our data points, plotting them on vertical axes for the time of day, ridership, and day of the week. All of the visualizations will have tooltips and will update to reflect Each of these visualizations is meant to give the MTBA employee relevant and comparable numbers from the larger data set. The dashboard format with an easy-to-use map selection and readable output allows the employee to get the relevant statistics and comparisons quickly. As a whole, this design allows for the data to be reported in a convenient and usable way to answer a variety of transportation questions the MBTA could ask.

8 Appendix B: Data Exploration

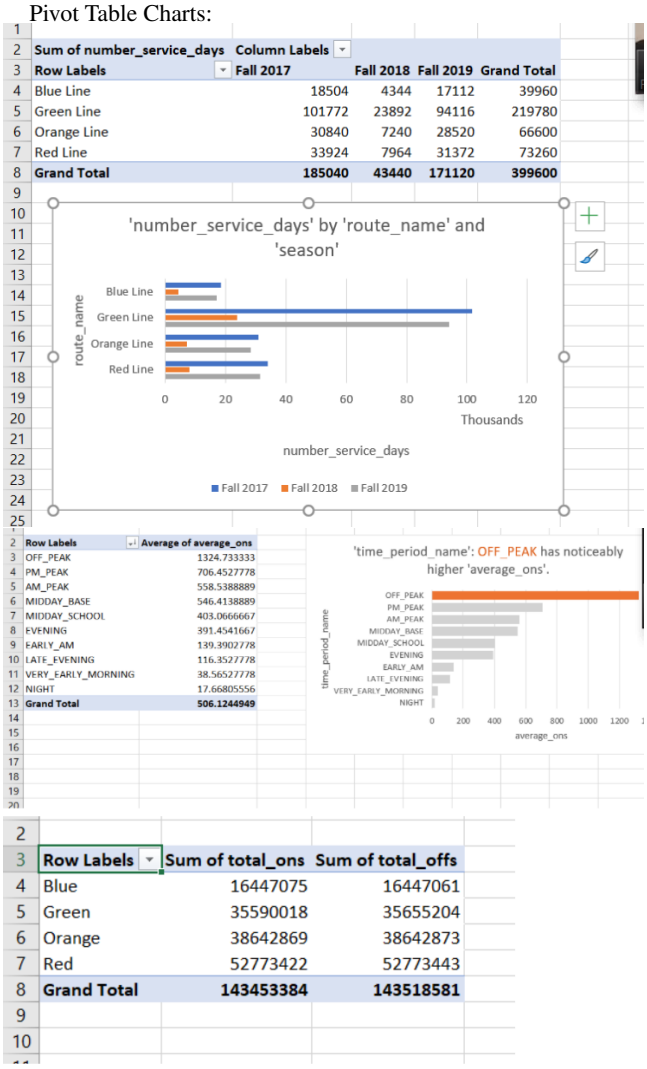
Data Review

The data types present in our data set are as follows, with the first column showing the column title, the second showing the data ranges and/or levels, and the third showing the data type. This information can be found in Figure 1.

FID	1.. Unique id for specific instance	Categorical
mode	0 or 1	Categorical
season	Fall 2017, Fall 2018, Fall 2019	Quantitative
route_id	Blue, Green, Orange, Red	Categorical
direction_id	0 or 1	Categorical
day_type_id	Day_type_01, day_type_2, day_type_3	Categorical
day_type_name	Saturday, sunday, weekday	Categorical
time_period_id	Time_period_01 to 11	Categorical
time_period_name	AM_PEAK, EARLY_AM, EVENING, LATE_EVENING, MIDDAY_BASE, MIDDAY_SCHOOL, NIGHT, OFF_PEAK, PM_PEAK, VERY_EARLY_MORNING	Ordinal
stop_name	Names of station	Categorical
stop_id	place-....	Categorical
total_ons	1-485837	Continuous
total_offs	1-517722	Continuous
number_service_days	5-82	Continuous
average_ons	0-17685	Continuous
average_offs	0-15675	Continuous
average_flow	0-36360	Continuous

Figure 1: The helpful values we have seen are season (which tells us the timing in the year), rout_id (which shows us the route by line), day_type_name (which shows us the timing of the day separated into human readable slots). Total on and offs show how many people board and leave the train which would be useful to show usage. Number of service days shows how many days that particular id is active.

Pivot Table Charts:



Sum of number_service_days	Column Labels			
Row Labels	Fall 2017	Fall 2018	Fall 2019	Grand Total
Blue Line	18504	4344	17112	39960
Green Line	101772	23892	94116	219780
Orange Line	30840	7240	28520	66600
Red Line	33924	7964	31372	73260
Grand Total	**185040**	**43440**	**171120**	**399600**

Row Labels	Average of average_ons
OFF_PEAK	1324.733333
PM_PEAK	706.4527778
AM_PEAK	558.5388889
MIDDAY_BASE	546.4138889
MIDDAY_SCHOOL	403.0666667
EVENING	391.4541667
EARLY_AM	139.3902778
LATE_EVENING	116.3527778
VERY_EARLY_MORNING	38.56527778
NIGHT	17.66805556
Grand Total	**506.1244949**

Row Labels	Sum of total_ons	Sum of total_offs
Blue	16447075	16447061
Green	35590018	35655204
Orange	38642869	38642873
Red	52773422	52773443
Grand Total	**143453384**	**143518581**

Although our data was very clean, we noticed several data quality issues from our initial review of the data. The problems and their solution are as follows:

- The field Direction ID only shows a 0 or a 1. For a viewer of the data or a visualization, this may be hard to understand. It should be changed to a string rather than a number e.g. inbound vs outbound so readers can understand the line of the direction.
- Mode is a field which has a 1 or 0 based on General Transit Feed Specification and does not help us as data that we can use for our project. It can be removed to make our data cleaner.
- Day-type_id is a field which is redundant as there is a corresponding field day_type_name which gives the name based on the id e.g. day_type_03 = Sunday. This field can be removed because we can just use the day_type name field as it is much easier to understand and means the same thing.

Insights

One of the trends we noticed was the on-and-off count for the different lines. The red line had the most ons and offs. Possible reasons could include the fact that it runs through spots such as Harvard and MIT. We also noticed the noticeable drop in ridership count on weekends, with Sunday being less than Saturday. This makes sense since people do not commute as much on weekends, and even less on Sunday since people typically plan outings for Saturday. A big surprise to us came with the data for service days over the years. The number of service days in 2018 was far below 2017 and 2019, which were both similar to each other. One issue we had was that we continued to use time period id's and time period names such as "time_period_01" and "MIDDAY_BASE". It would be more helpful to have a solid time such as 1:00pm but our data did not offer that precision. Our dataset also only has data for ons and offs for the MBTA which is likely not enough to do analysis at a deeper level. To solve this, we plan to merge another dataset with this one to gain a more complex dataset to provide better insight.

Screenshots

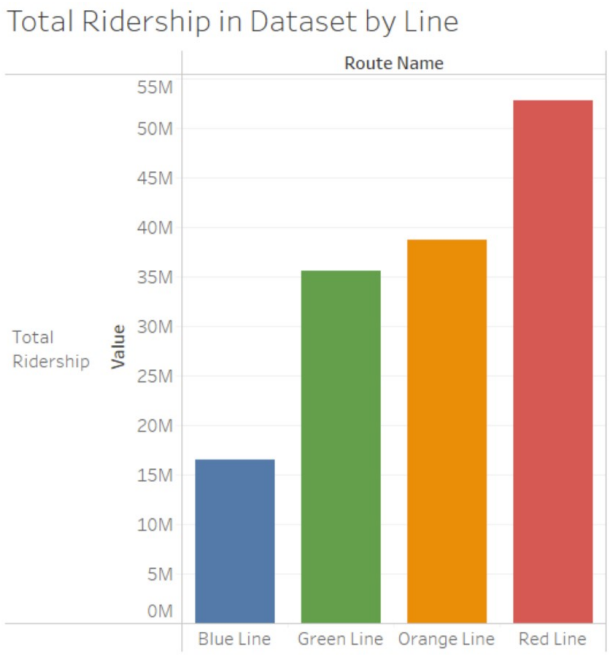


Figure 2: The visualization above illustrates the total ridership, summing both the Ons and the Offs for each route in our dataset. The marks used were areas, in the form of rectangles/bars. The channels used included colors to show the different lines and varying heights/area sizes to convey the total ridership values (as the value is reflected by the height of the bar's top side). It seems that in our dataset there are a lot of Ons and Offs recorded for the Red Line, with much fewer for the Blue Line. The Blue Line's ridership in this dataset is less than half of what is recorded for the next lowest line, which is the Green Line.

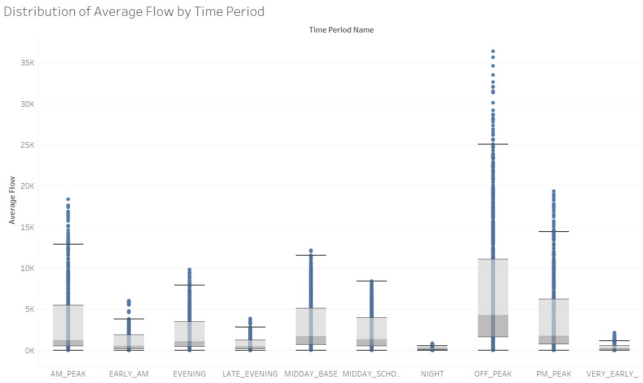


Figure 3: These boxplots show the distribution of average flow for different time periods. The marks used include points to show the individual average flow values and box/whiskers showing the different quartiles and ranges. The channels used include the position of the points on the Y axis, and the box and whisker sizes/heights, demonstrating the Y axis value for each quartile/range value. Based on this exploration, it seems that there is a lot more variability for the Off Peak average flows than for any other time. The night and very early time periods in contrast have very little variability in average flow, shown by the very short (almost flat) box plots).

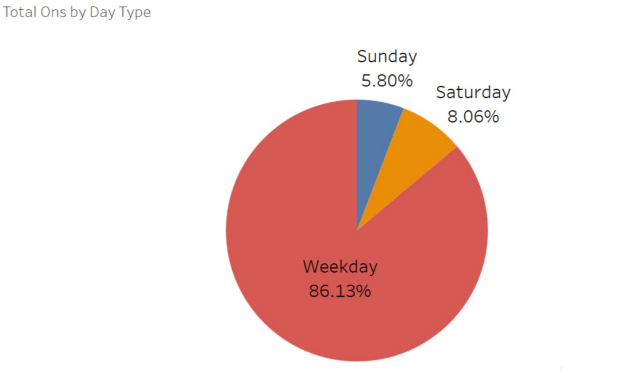


Figure 4: This pie chart shows the Total Ons in our dataset organized by Day Type. The marks used are the pie shaped areas, and the channels are the angles the pie shapes are formed by (which changed this size relative to the percentage being represented) and the colors which represent the different day types marked by the labels. Even though weekdays only make up about 71% of the week (5/7 days), it accounts for 86.13 % of the Total Ons recorded in our dataset. This makes sense, as the T is used by many for commuting to and from work, which takes place during weekdays.

9 APPENDIX C: INTERVIEW

End User

I am an employee at the MBTA (transportation agency) working as an analyst in the budget and planning department. My motivation is to better understand the T system and help allocate resources to help management make better decisions on how to fund projects. This is critical because we want funding to go to projects that help the community in Boston be more mobile and make for a better rider experience.

I have used excel to help visualize data that we collect, however new visualizations that express relationships in the data would be very helpful. A specific task I encounter in my work includes finding which projects would need additional funding to report up to the project manager, along with how these projects are serving riders. I need to manually gather data ad-hoc in excel for my manager when they ask for a ridership performance. This task is very redundant and we are not able to build or remember insights after I send the report to them. I envision a data visualization that may help me gather quick insights on the performance of a line and also see how ridership may have changed over time. This would allow me to come up with quicker response times that help managers move their projects quickly and efficiently. I would like some interactivity to select different lines in the system and it would be helpful to have averages of key categories displayed when I hover over them. It would also be great if there is an ability to select different times/seasons to see where we are looking at data from. It would also help to see which stations are associated with which lines and the time_period_name column used to see when the busiest times are in the day.

Interview Script

1. What is the reason you're looking to have a visualization made?
2. How have you been deciding how to allocate resources in the past? How do you feel about the current experience?
3. Can you explain two pain points you have for the process currently?
4. What would you like the visualization to show and convey overall and what role would it play in your resource allocation process?
5. What kinds of statistics and comparisons would be the most helpful for you? Could you elaborate on how you would like to measure performance?

Interview Notes

- The MBTA allocates resources based on stations that see the most riders.
- Resources: maintenance workers, money to upgrade station facilities.
- Right now, to get information they put the data into excel and make calculations and graphs.
- Small scale insights used on demand, but nothing bigger.
- Pain points: Manual calculations and redundant tasks.
- Needs visualization to show ridership based on different metrics, to find out which stations are being used the most and when that is.
- Dashboard showing data based on given criteria such as line, station, or time of day.
- Visualization will allow MBTA to figure out where to send resources and the best time to do maintenance during the day based on when it is less busy.
- Wants stats for total ridership, flow, average ridership over a week for a given line, and ways to compare the data by time and day of the week, and more largely comparing different years.

Interview Results

1. Why is the MBTA looking for a visualization currently?

   The MBTA is looking for a visualization because they want to better allocate resources such as maintenance funding to the stations that see the most people. In addition, they want to be able to figure out the best time to send maintenance and do work during the day.

2. What has this process looked like in the past?

   Currently, to get relevant data they have taken their data into excel on a small scale and made insights on demand. They have never made a visualization/dashboard of all of their data like this on a large scale. As a result, managers reach out for manual visualizations often.

3. What pain points exist in the current process?

   Two pain points that exist in the process currently are the fact that the calculations are being done manually in excel. In addition, the tasks in excel become redundant over time as their calculations are done on a small scale each time and are repeated every couple months.

4. What would you like the visualization to show and what role would it play?

   The visualization should include a dashboard of statistics based on the given criteria such as Line or time of day, and be able to include graphs and charts comparing the data from different stations and lines. This visualization will allow the MBTA to figure out where to send resources and the best time to do maintenance during the day based on when it is less busy.

5. What statistics or comparisons would be the most helpful?

   They would like statistics showing total ridership for different lines and stations, the distribution of flow throughout the day and average ridership over a week for a given line, among others. They want to be able to compare the data and statistics for different stations, times and days of the week, and more largely the ridership over different years.

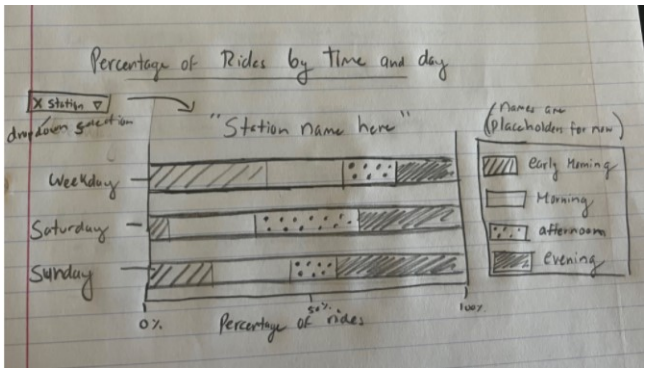


Figure 5: Mikaela Sketch 1 - I chose to use area marks and then color and size/position for the channels. Using this combination of marks and channels allows the user to visualize how the percentage of riders in a day is distributed over the different time periods. This also allows them to compare the ride time distribution for different days of the week, highlighting when the busiest and slowest times are for each. This addresses task 2, as it allows the user to see within a given day of the week which times would be the slowest and therefore the best for maintenance.

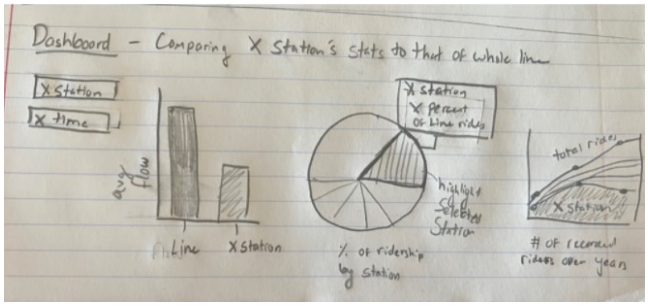


Figure 6: Mikaela Sketch 2 - This visualization uses multiple marks and channels over the three dashboard visualizations. The bar chart uses the bar area as a mark and the position of the top line of the bar as the channel, along with color to highlight the chosen station's bar. The pie chart uses areas as marks and then color, tilt, and area as channels. Lastly, the line chart uses areas (bound on top by lines) as marks and then color and size as channels. This addresses task 4, as it shows some possible dashboard visualizations to include. Certain averages and values can be displayed in pop ups as well but are not shown here.

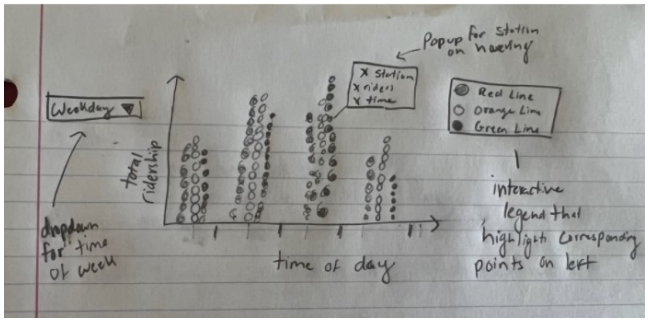


Figure 7: Mikaela Sketch 3 - This visualization uses points as marks, and then color and position as channels. It addresses task 3 as it allows the user to compare the total ridership of the different lines based on time of day. Each point will represent a different station's data, and so the user will be able to see the distribution to total ridership over each line and over each time of day.

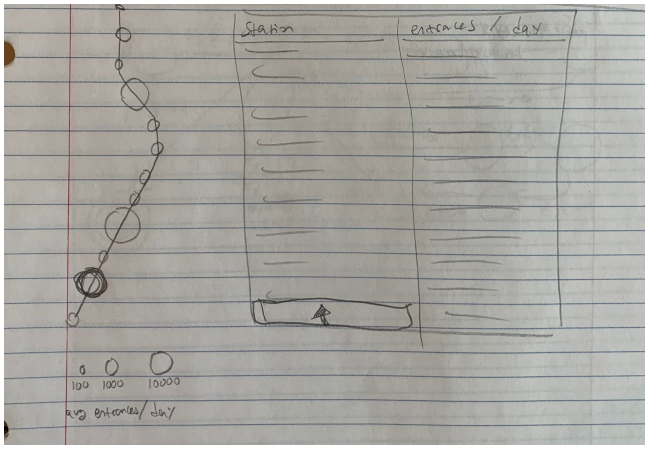


Figure 8: Kai Sketch 1: This visualization will use the map for each line and will use the stations as marks. The size of the station will represent how great the number of entrances there are for that station per day on average. There will also be a table on the side of it which will be linked to the map. The row that is hovered on will be linked to the map and both will be highlighted in bold. That way the user can tell the exact number of entrances. This can be done for each line, or we could even combine all the lines together. We could also do the same visualization for exits per day. This would be for task 1.

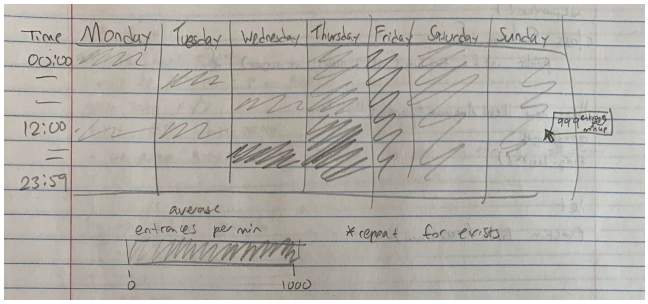


Figure 9: Kai Sketch 2: This visualization will use color to communicate the number of entrances per day or minute for all stations based on time. The darker the color, the greater the number of entrances. This could also be done for exits. When the user hovers over a section of the grid, it will show the exact number. This will help will task 2.

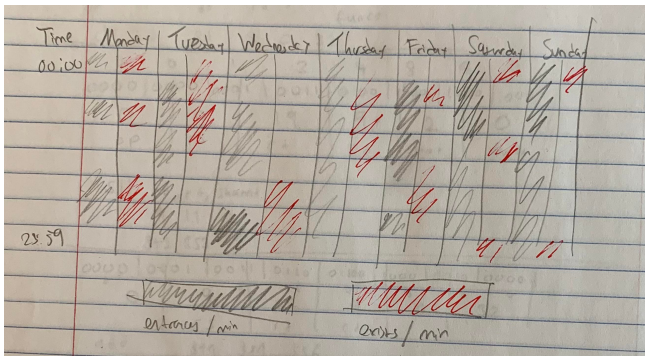


Figure 10: Kai Sketch 3: This visualization is an extension of drawing 2 and uses color to communicate the number of entrances per minute as well as the number of exits per minute, giving the user a better idea of which times do people exit more at in addition to which times people enter at. Color will be used to show the amount of entrances/exits. Different colors will also be used for exits and entrances so the user can differentiate between the two. This will help for task 2.

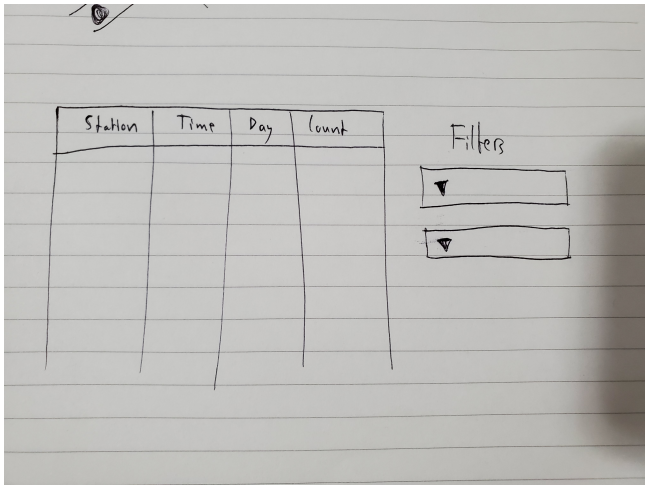


Figure 12: Nop Sketch 2 - This second idea targets task ID 4 and is essentially a filterable interactive table. The user can filter out depending on station, time of day, sort by count, etc. This allows us to get very specific data. The marks for this are the boxes in the table, and the channels are the values in the boxes.

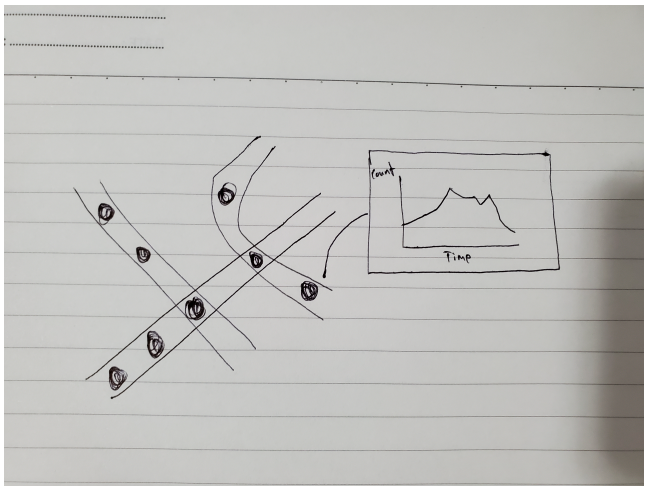


Figure 11: Favorite - Nop Sketch 1 - This one has the base form of a map of the lines. The marks are the stops and the channels are the color of the stop indicator. The warmer the color, the more active the stop over a selected period of time. Hovering over a stop will show the user a number-by-time activity line graph for a time period (eg. one day). This targets all the task IDs that we identified but it may be an overloaded idea.

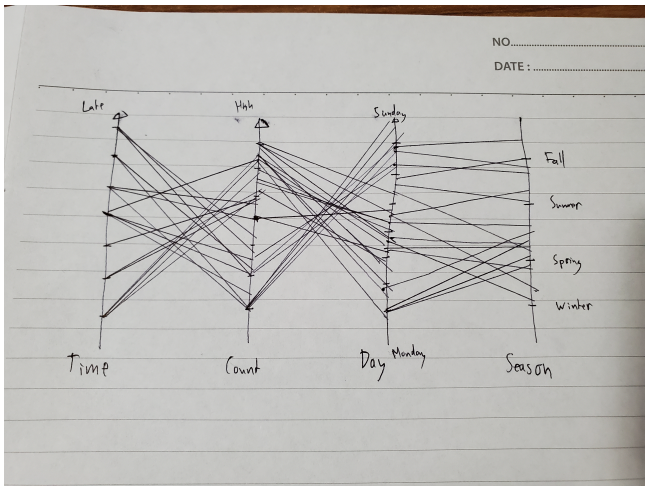


Figure 13: Favorite - Nop Sketch 3 - This is a parallel coordinates graph which allows us to link the different variables together to see trends. For example, we predict that during early and late hours of the day, there will be higher activity. So we should see more lines going to the High part of the count section coming from those early/late parts of the time section. If things are strongly correlated, we would see a denser concentration of lines and more random lines if things are weakly correlated. This mainly assists task IDs 1 and 2 as it is mainly for identifying trends and correlation rather than specific data. The marks are the lines and the channels are where the lines point to.

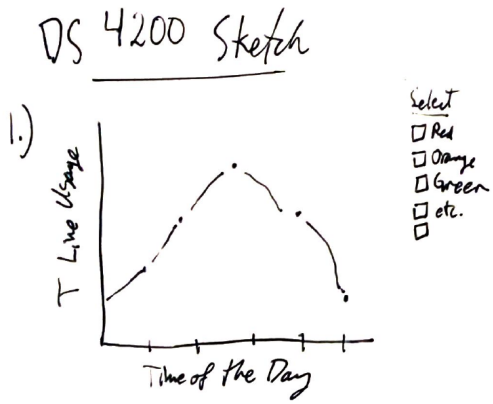


Figure 14: Clayton - Drawing 1 communicates the importance of seeking the peak hours in the T system. We allow the users to select which line they would want information on and it populates the aggregate usage based on time period name. Interaction would include selecting line and also the ability to see specific numbers at each time period. We would be able to compare if two lines or more are selected and their colors would be corresponding. Corresponds with task 2

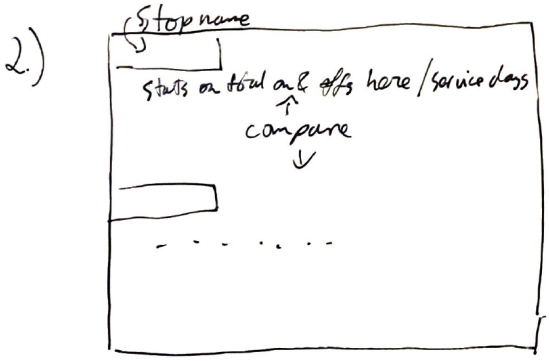


Figure 15: Clayton - Drawing 2 is a simple interface to get quick information on comparing two stops. We would let users search for a stop based on their name and it would return the stats on total on and offs or even service days. The use case here is the agency wants to compare two stops to see which is under more strain, this would help. Additional elements to add are relation to which line it is apart of. Corresponding to task 3 for showing stats for stops in the system.

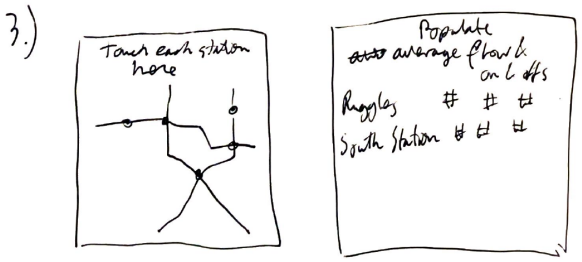


Figure 16: Favorite - Clayton - Drawing 3 is a 2 view visualization to show each stations flow. We would try to incorporate each station in a map on the left. When users hover or select a station, we would populate the average flow information on the right. This incorporates brushing and allows users to make the connection and compare flows because flows are generally related for each line. Corresponds to task 1 to discover data for each station based on flow.

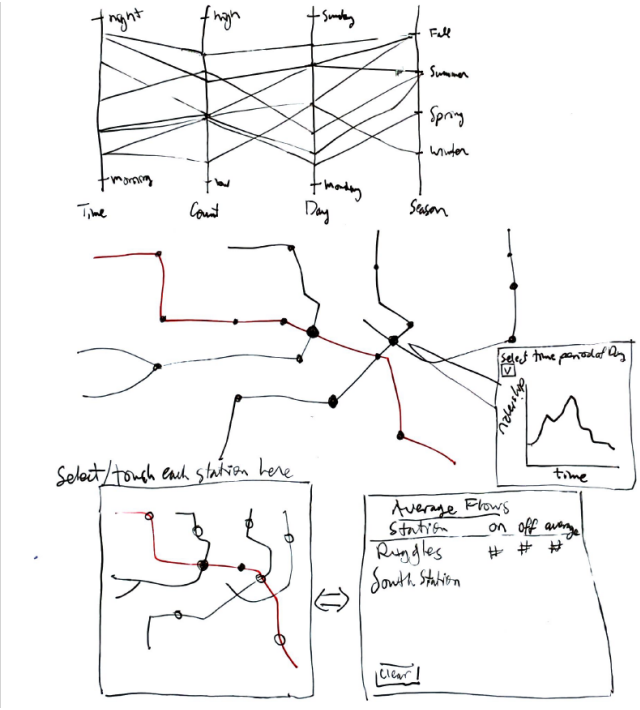


Figure 17: Final Sketch - We chose these three visualizations together because we fulfilled exploratory tasks in Task 1 and 2 to allowing the user to pinpoint information about specific stations and lines further down in the graphs. The user can start by seeing trends and then explore times of the day that they think are busiest. This couples with the ability to see which stations are active at the time and finally allows the user to compare station's flow (task 3 4) We seek to use colors and marks to show the activity and distinction between different stations and lines which make it easier to use. The overall goals helps to address how we will show activity and assist in resource allocation through trends and data specifics.

Why Favorites Were Chosen

We chose these three visualizations because they would provide the user with a great opportunity to explore and understand their data from seeing trends to comparing specific stations. Starting with our first pick, the parallel coordinate graph would show different variables together and allow us to see trends on the groupings of the lines. Denser lines may represent more strong correlations. This supports all data types as we will determine the scale of the categories. It would support tasks 1 and 2 to help us understand the trends in usage for lines and stations.

Secondly, we decided that the map visualization based on station activity would be a great way to see the change over the selected time period. The marks are the stops and the channels are represented by the warmth of the color, signaling more activity. This helps to understand the most active station for each time of the day and may see which lines are also actively working together. This would support tasks 1,2, and 3 because we would also receive statistics on activity on the station we hover over.

Finally, the double view visualization would allow us to brush and link and have interactions for the user to select an area around one or more stations and see corresponding statistics on the other view. We plan to use information on the average flow for this visualization because it allows us to better compare the total number of passengers traveling through the rail system between stations within the aggregated time period. This helps with task 3 and 4 as we get statistics and almost a dashboard view of what we decide to do the flows by station or line.

11 Appendix E: Digital Sketch

We hope that the user would be able to explore the map of the T stations and pick out the areas where data is needed. The MBTA worker could have stations or lines of interest in mind based on a need for maintenance or funding allocation. As they make these selections, the graphs and tables will update to reflect the data for those specific stations or lines. The hope is that the user can quickly and easily pick the stations they need information for, and have all the data they could need at one time. In this way comparisons between stations can be made efficiently, and then the data can be reported to outside agencies quickly as they appear in the table.

This will cover all our domain tasks that we had initially set out as goals for our visualization. The first domain task was to show ridership based on different metrics, which can be viewed when the user selects the stations that want to inspect. The second task was to allow MBTA to figure out the best time to do maintenance. This can be done by looking at the data in our visualization. The third task was to show total, flow and average ridership over time which is also shown. The fourth task was to create a dashboard, which is the overall theme of our visualization.

12 Appendix F: Usability Testing

Preparation

Our visualization tool is intended to help MBTA staff understand the T system and flow through different stations. The bar graph shows how the ridership in our dataset is separated out by line, and the pie chart shows how the ridership is distributed over the different time periods. Our interactive map showcases the entire system and allows users to click on specific stations and get detailed information through the table and pie chart which are connected to the user's interaction. We intend to show flow data along with how time affects flows.

The first thing we would like you to do is tell us what percent of all MBTA rides are taken during the early AM time period. This task is being tested in order to see if our users can read the pie chart. The user should be able to find what percent of rides occur for each of the different time periods. We hope that users will be able to provide us

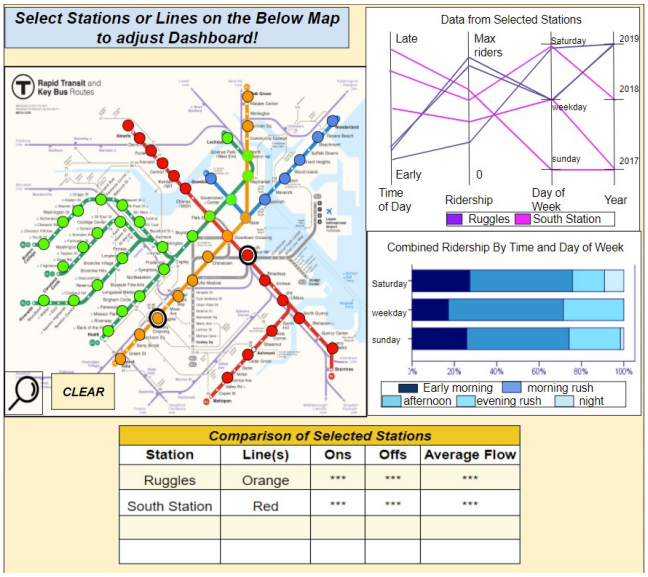


Figure 18: Final Digital Sketch

with the correct value and that it is clear which percentage belongs to which time category

Second, can you tell us which line has the most ridership over the three years in our dataset? This task is being tested to see if our users can read the bar chart and tell which of the four lines in our dataset has the highest ridership. We hope users can successfully compare bars on the chart and give us the correct answer. Users should also easily differentiate and identify which bar corresponds to which line.

Lastly, can you find where Ruggles is on the map and tell us what line it belongs to? This task is being tested in order to see if users can use the map and understand the coloring of the dots. The hope is that users will be able to zoom and pan around the map and find Ruggles, see that it is orange, and put those together. In our final visualization, users will need to go around the map and select stations, so it is important that the map works and can be understood.

Results

Overall, our usability testing went well. The testers were able to complete each of our tasks successfully, but they did have minor suggestions that would make these tasks easier to complete and would increase overall data understanding. They liked the way that the map could be explored and how it connected to each of the other visualizations. Our users pointed out a few minor improvements that could be made in order to better the visualizations, including adding more descriptive titles and changing the color scheme of one of our plots

Task 1: The users liked the pie chart in that it shows the distribution of ridership over all of the different time periods. Our test was designed to see if the user could successfully read our pie chart and understand what it means. The task was successfully completed, but we did receive some comments. The main feedback we received based on this testing was that the current color scheme was not as readable as it could be. It was suggested that we move from a color map to a discrete color scheme for the different time periods. This is a reasonable critique and should be implemented.

Task 2: The users were able to tell us the line with the most ridership based on the bar chart. They successfully found that the red line had the biggest total ridership over the Fall 2017, Fall 2018, and Fall 2019 time period we have in our dataset. The one piece of

feedback we received was to add a more descriptive title to the plot explaining the years of the data included in the dataset. We plan to implement this in our final visualization by changing the title text.

Task 3: The users were able to successfully find the Ruggles station on the map by dragging and zooming on the map. This means that any user should easily be able to find and select any stop they want to get data from. One positive aspect that they highlighted was that they enjoyed the interactivity of the map and how easy it was to navigate. They did make one suggestion that we can implement, which is to draw the lines connecting each station. We did try to implement this but have not been successful thus far. However, it is not necessary as the user can easily tell the direction of the line from the points and their colors.

Based on our feedback we will be adjusting the color scheme of our pie chart to have discrete colors rather than a color map. This was pointed out as a flaw in our usability testing and will be done to improve readability. In addition, we will be adding more descriptive titles to our plots. This was also noted in our usability testing and will help the user to better understand what kind of data the plots are showing.

REFERENCES

[1] M. Barry and B. Card. Visualizing mbta data. `http://mbtaviz.github.io/`, 2014.

[2] W. Chen, F. Guo, and F.-Y. Wang. A survey of traffic data visualization. *IEEE Transactions on Intelligent Transportation Systems*, 16(6):2970–2984, 2015.

[3] M. De Marchi. Visualizing public transport systems: State-of-the-art and future challenges. *unpublished.[Online]. Available: https://www.evl. uic. edu/creativecoding/cs524/RESEARCH/DeMarch i_Transit. pdf*, 2015.

[4] Y. Sun, J. Shi, and P. M. Schonfeld. Identifying passenger flow characteristics and evaluating travel time reliability by visualizing afc data: a case study of shanghai metro. *Public Transport*, 8(3):341–363, 2016.

[5] W. Zeng, C.-W. Fu, S. M. Arisona, A. Erath, and H. Qu. Visualizing mobility of public transportation system. *IEEE transactions on visualization and computer graphics*, 20(12):1833–1842, 2014.

13 APPENDIX A: GROUP CHARTER

- Group Purpose: We found a common interest in public transportation because we interact and use the T system in Boston and want to better understand it. We are working to use data from the MBTA to provide insights to the agency so that the MBTA may see how different lines operate in relation to stations and ridership.

- Goals: We aim to commit as much time as possible to completing this project and to the best of our ability. We collectively are aiming for the best outcome possible by understanding problems and following design principles learned in class. We aim to be as efficient and as effective as possible with our time, and want to be happy and fulfilled with our final project.

- Membership Roles/Responsibilities:

  - Mikaela - Communication and meeting facilitator
  - Nop - Documentation coordinator
  - Clayton - Project management role/ domain analysis
  - Kai - Information manager/ technical specialist

- Rules: Communication will be done through Facebook Messenger. Meeting times will be decided there and done through Zoom or in person. Group decisions will be made by a vote. Equal contribution by all members is expected through the project. If the group is split on a decision, we will talk it through and make sure we are respectful of the other side's opinions.

- Potential Barriers and Coping Strategies: There may initially be a small communication barrier as members are seeing each other for the first time. However, over time this should not be a problem. Members should try to speak their thoughts and try to get to know each other. If there is conflict of ideas, the best way to go about this is to talk it out and make decisions as a group. Communication is key, and groups who communicate well are less-likely to have problems.

Mid-Project Charter Evaluation
Our initial group charter is providing a successful structure. All the guidelines we had proposed have been effective and were followed by all members. We have all been comfortable with their roles for this group and there are no immediate problems that we need to troubleshoot and address. We are looking forward to the implementation plan and plan to adapt as needed.
Positive comments from each group member:

- Mikaela - Nop is a team player who contributes many ideas, especially with the technological aspects of the project.

- Nop - Clayton is the person who birthed the idea of the project, always guiding the team in the right direction and always gives ideas for new visualisations.

- Clayton - Kai always gives great ideas and is willing to help out others.

- Kai - Mikaela has been great leader, constantly motivating discussion and setting up meetings.