CS-5630 / CS-6630 (Fall 2022)

Final Project

Traffic Routes in The State of Utah

(With emphasis on Salt Lake City)

Authors: Milena Belianovich, Tark Patel, Xiaoya Tang.

(Emails: u1419504@umail.utah.edu, tpat@sci.utah.edu, u1368791@umail.utah.edu.)

Basic Information.

Title: “Traffic routes in the state of Utah” (may change at a later date).
Contributors: Milena Belianovich, Tark Patel, Xiaoya Tang.
Emails: u1419504@umail.utah.edu, tpat@sci.utah.edu, u1368791@umail.utah.edu.
UIDs: u1419504, u0893472, u1368791.
Project Repository Link: https://github.com/xiaoyatang/Final-Project-VIS-2022.git (Since our repository is private, we have added the TAs accounts to the collaborators.)
Project Website Link:
https://xiaoyatang.github.io/Final-Project-VIS-2022_Tark_Xiaoya_Milena/
Video Link:

Background and Motivation.

Background:
Background work that was done on this subject is the reading of “Topological Analysis of the Evolution of Public Transport Networks”. This article analyzed the topology of Stockholm's tram lines, calculated multiple metrics like the Pearson coefficient and closeness centrality, and visualized node charts with those metrics for coloring. This project showed that topological analysis describes the significance of important stations through the amount of their connections and age of creation, as the backbone of the network tends to be made first.

Motivation:
Topological Analysis is a powerful tool for understanding the relationships within complex data. Applying this to public transportation networks is useful for route planners and riders. An example is that route planners can analyze the connections of a traffic route with respect to how close it is to a core of the network. A dashboard for visualizing these properties could prove to be a powerful tool for better planning.

Milena:
What motivated me to do this project was the possibility of creating dynamic visualizations as well as the project's close connection to topology. I am currently working under Professor Bei Wang-Philips, who is known for her love for topology, so after learning many new things as her laboratory rotation student, I plan to implement them in this project. Creating a project using a programming language I have not used for projects before is very challenging but exciting at the same time, so I am looking forward to seeing how we all do our best to achieve set goals.

Tark:
I think this project will apply everything we've learned in this class so far while providing more challenges due to the preprocessing work. New techniques we will have to learn are how to zoom on maps, combine state and city views, and display routes onto maps.

Xiaoya:
I am thrilled to apply what I learn from this class to a truly meaningful project. The first idea came to us is to make a route map of the public transportation in Utah, especially in Salt Lake City. Traffic routes on maps are often not intuitive. I think a nice-illustrated public transportation network can be truly helpful in route planning and time schedule, especially for newcomers like us. What we also want to do is to see the changes of traffic routes in this city during recent years, and to explore its connection with human and economic factors like population distribution or annual tourism revenue in a city. There are many fascinating visual analyses that can be done if we want to go further.

Project Objectives.

The most important entities for our project will be bus stops, railway stations and their connection to different routes in a large range of years (depending on how many datasets we can find). We hope users can choose the correct paths using our visualization and we can also see the obvious change of traffic routes in the state of Utah over the years from our visualization.
If we further find some dataset of on-board people during different time slots in one day, we will give a better and detailed transportation map for people who see our work. It will allow people to better schedule their time and to plan the routes. Besides, we may acquire some additional data like population distributions and apply it to the changes of traffic routes(to be determined). We plan to find some interesting connections between the public transportations and some human and economic factors in the city. Also the traffic routes may have some connections with some special facilities like schools, hospitals, supermarkets. We may give them highlights in the map, and we hope to do more fantastic interactions and explorations in our visualizations, besides just including the knowledge acquired from this class.
Some questions that our visualization would provide answers to:

1) How were the routes of different buses changing over the years, and why?
2) Which route (in a selected city) has the best availability? Which has the worst availability?
3) Is there a relationship between usage and number of stops?
4) Does usage of routes correspond to the number of other route connections?
5) What are the connections between most used traffic routes and the nearby facilities?
6) The busiest time slots of the public transportation(potentially).

7) Is there any chance of optimizing public transportations according to our findings?
8) What are the connections between the traffic network and the population distributions in the city, and is there any drawback of the route's design considering the distribution(potentially)?
9) Which city has the best bus transportation availability? (potentially)

Data.

One of the main resources our group is using to collect data is the UTA website: https://data-rideuta.opendata.arcgis.com/.
Routes Data URL: https://data-rideuta.opendata.arcgis.com/datasets/rideuta::uta-routes-and-most-recent-ridership/about
Bus Stops Data URL: https://data-rideuta.opendata.arcgis.com/datasets/rideuta::uta-stops-and-most-recent-ridership/about
Onboard Data (requested and acquired from UTA official website).
Stop Boarding Data (requested and acquired from UTA official website).

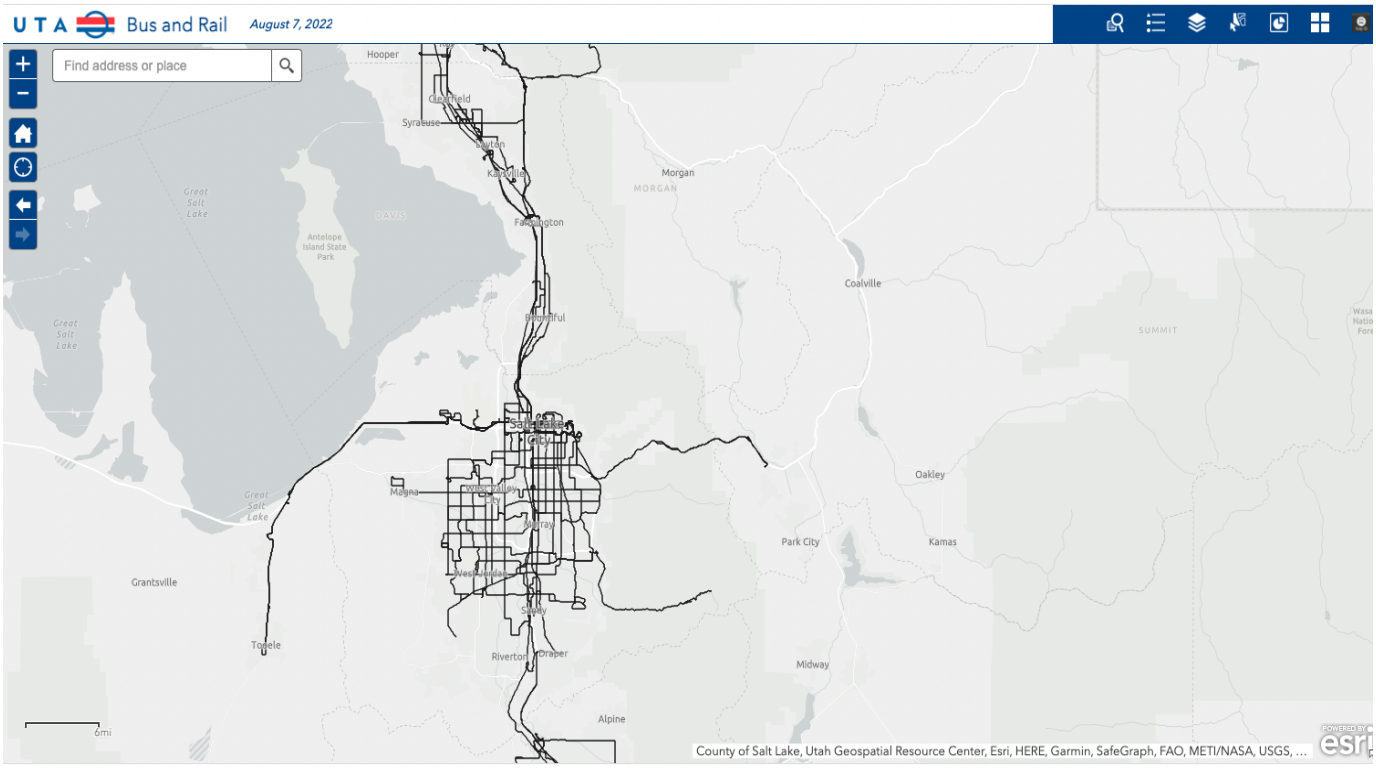
Data Processing.

1) Traffic Routes.
Routes JSON file: https://services.arcgis.com/5QAphMT1g51Tw2X4/arcgis/rest/services/UTA_Routes_0419/FeatureServer/0/query?outFields=*&where=1%3D1&f=geojson
Attributes: route abbreviation, route name, frequency(minutes), route type, city, county, most recent average weekday boardings, shape_length.
Data cleanup: We have concluded that substantial data cleanup is not required, except for adding some "NaN" instead of missing data for certain columns, if we see fit.
2) Bus and rail stops.
Stops JSON file: https://services.arcgis.com/5QAphMT1g51Tw2X4/arcgis/rest/services/UTA_Stops_and_Most_Recent_Ridership/FeatureServer/0/query?outFields=*&where=1%3D1&f=geojson
Attributes: stop name, city, zip code, county, routes served, stop number (integer), stop abbreviation (text), mode (bus or rail), average weekday boardings and alightings, latitude, longitude.
Data cleanup: We have concluded that this dataset requires no substantial data cleanup.
3) Onboard Data.
Attributes: ObjectId, Mode, LineAbbr, Month, Year, ServiceType, AvgBoardings, City, County.
Data cleanup: We have concluded that this dataset requires no substantial data cleanup.

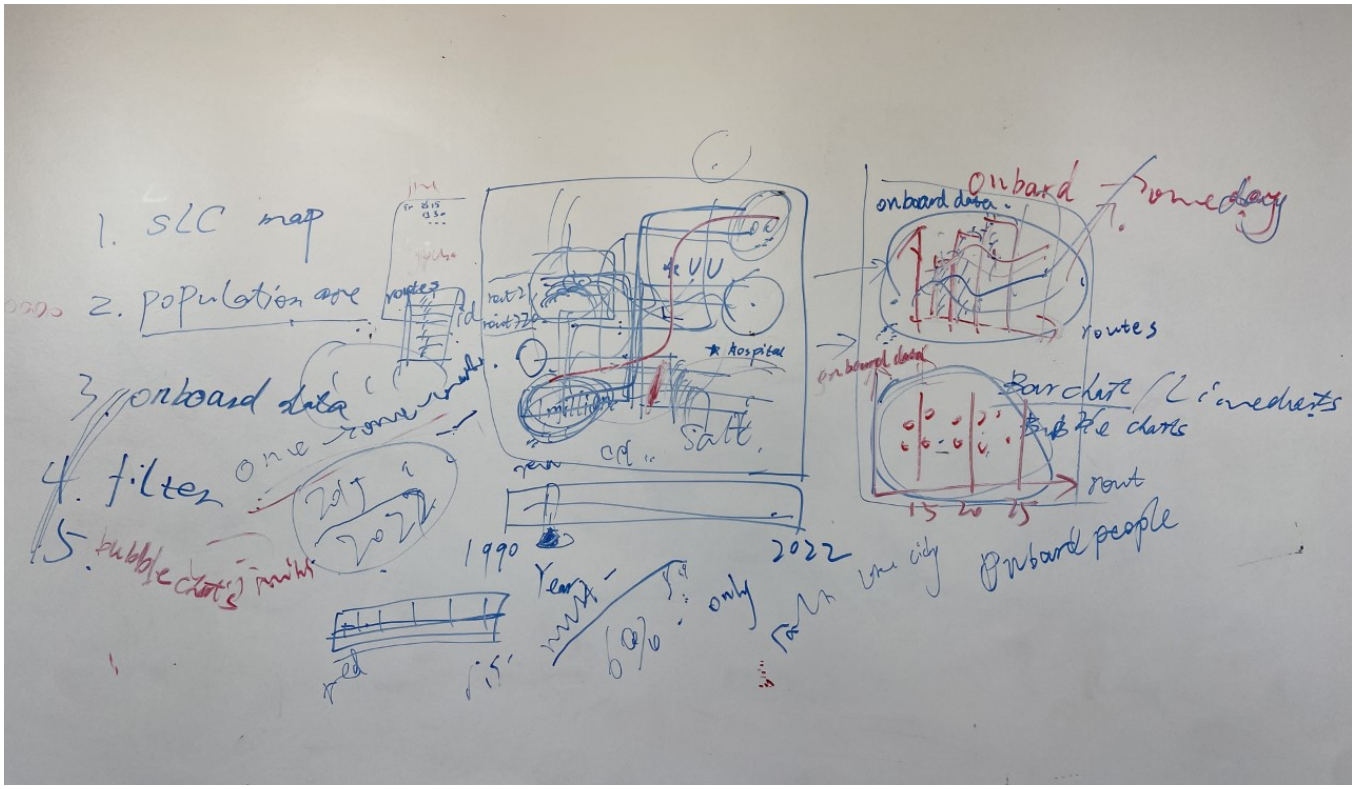
4) Stop Boarding Data.
Attributes: ServiceType, Month, Year, StopAbbr, StopName, City, County, AvgBoard, AvgAlight, Routes, Lat, Lon, ObjectId.
Data cleanup: We have concluded that this dataset requires no substantial data cleanup.

Here begins the **skeleton**.

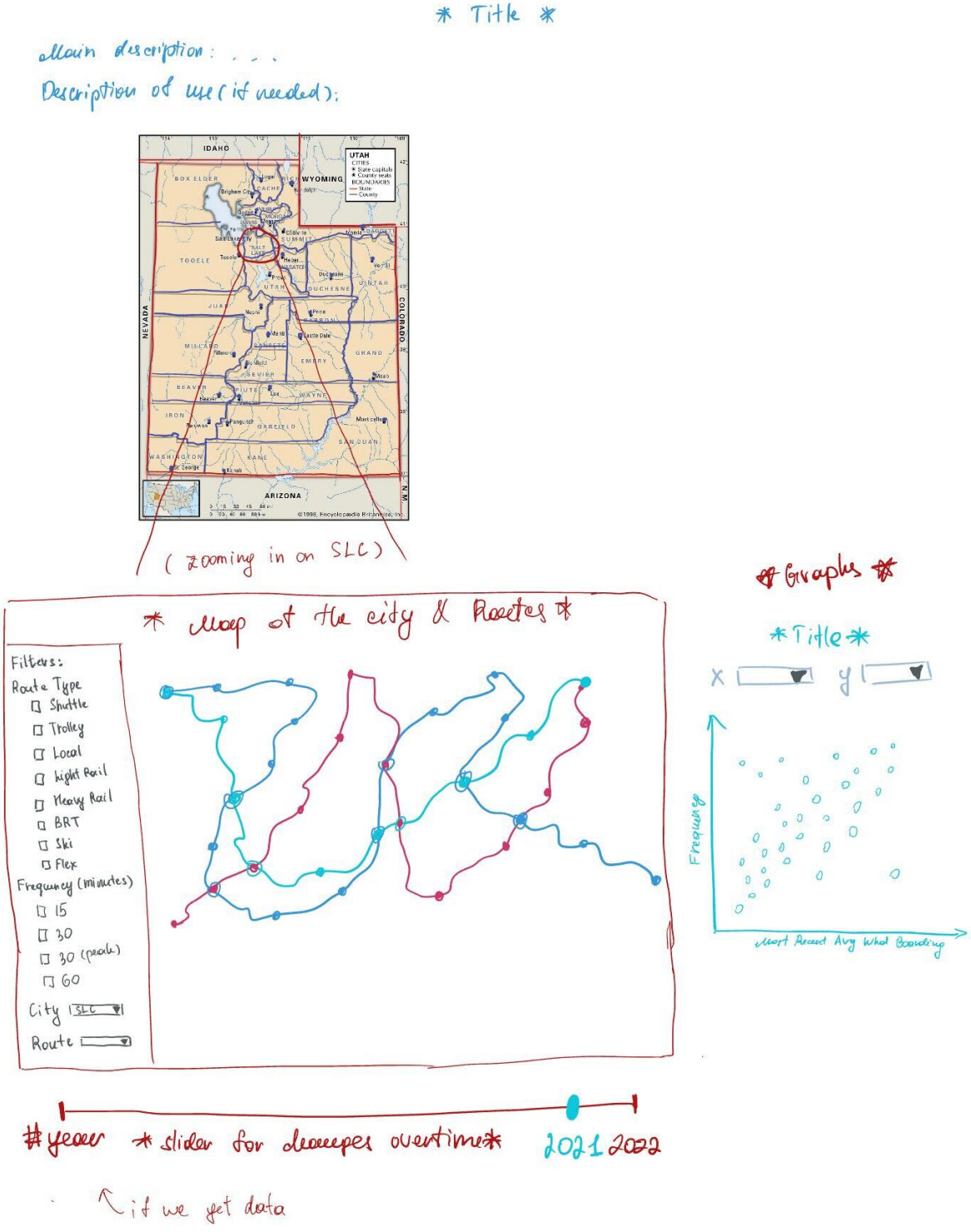
1. Inspiration comes from:



2. Our sketch of the final goal:



3. Previous thoughts and sketches:



Title

Description:

Description of use:

Authors:
Copyright:
Sources:
etc. ...

Map Title

Filters

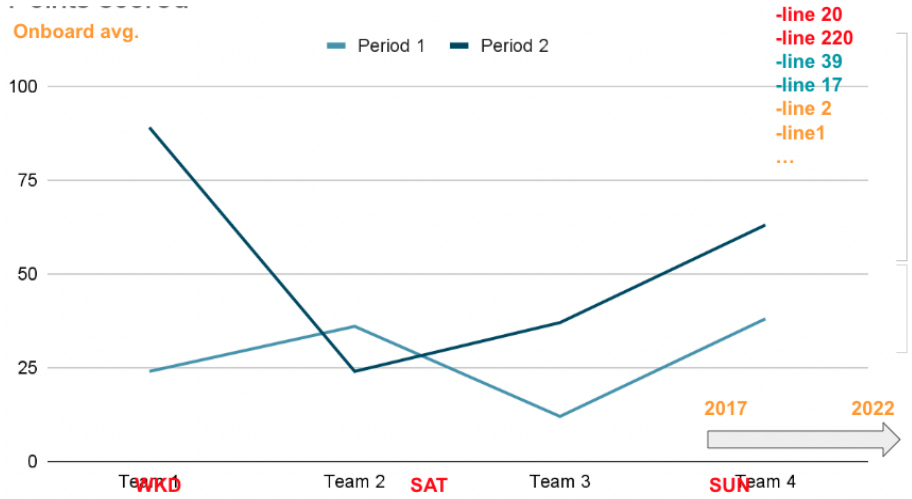
Map

(zoom in/out)

Graphs

Extra space for other implementations, legends, etc.

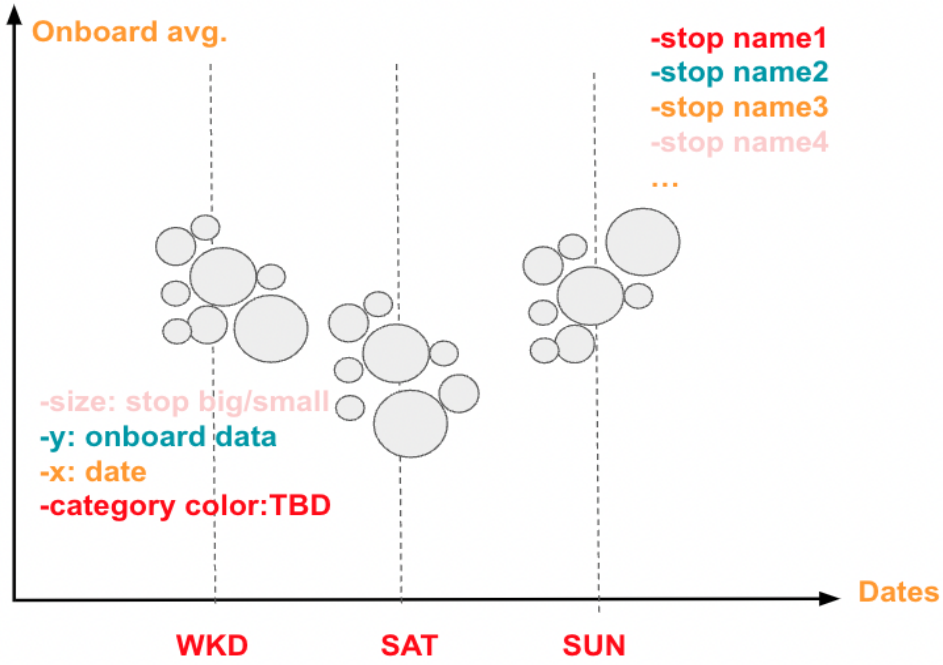
Line Chart representing individual routes as selected on the map.



Bubble chart for interactions with map:

x-axis: WKD, SAT,SUN, y-axis: Stop boarding data(there are 2000 bus stops in SLC); Size of circles: stop big or small? I.e., how many routes serve the stop? Color of circles: TBD.

Interactions: choose boarding or alighting; choose 2020, 2021 or 2022.



Skeleton includes:

Data:

map.json(Aug. 2022)✅;

stops.json(Nov. 2022)✅;

onboard data(WKD,SAT,and SUN from 2017-2022).csv✅;

Stop Boardings - Bus(WKD,SAT,and SUN from 2020-2022).csv✅

Population(2017).csv/json✅(optional)

highlight positions locations❎; annual data of routes❎(requested)

In the UTA website it only shows the data above in 2022. For acquiring the historical data, we contacted the recording staff of UTA. The data of stops are given by .shapefiles. We use libraries of d3 and arcgis in our implementations of all data.

Graphs:

a. **Map** graph of traffic route in SLC:

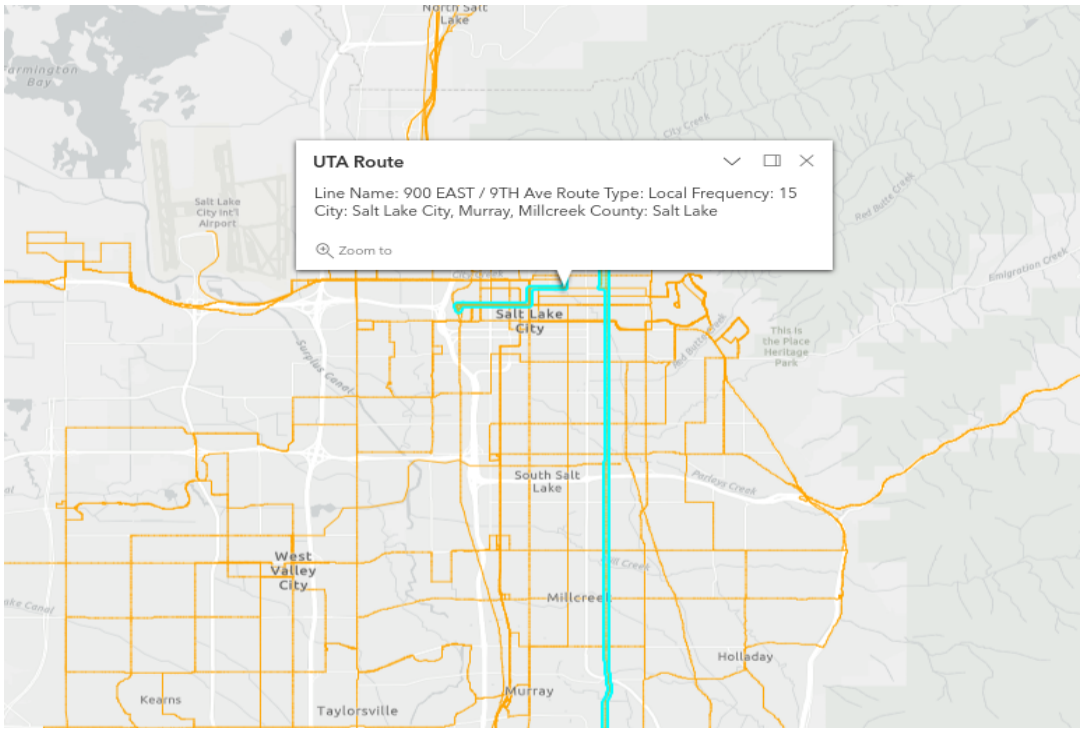
*__zooming__ in and out of the map to choose the desired view of the map routes/stops

*__tooltips__ for routes and stops on the map that show additional information;

*__brushing__ implemented for selecting/deselecting routes and stops in certain areas such that the graphs on the side update accordingly;

Progress: We used the ArcGIS API to visualize transit routes as lines. The map view starts at the northern part of Salt Lake City, but panning and scrolling are supported. The second image shows transit stops as points. Clicking either routes or stops highlights the selection and shows a tooltip with more information.

The UTA's official map website uses the ArcGIS API too. We chose this over D3's map because it allows better interaction. Another benefit we found is that clicking selects multiple elements and there are arrow buttons in the tooltip to switch between them. This is particularly useful when the zoom is far and the number of possible selections is large.



UTA Transit Routes

Authors: Milena Belianovich, Tark Patel, Xiaoya Tang

b. **Bar chart** for interactions with map:

***Basic settings:**

X-axis: months;

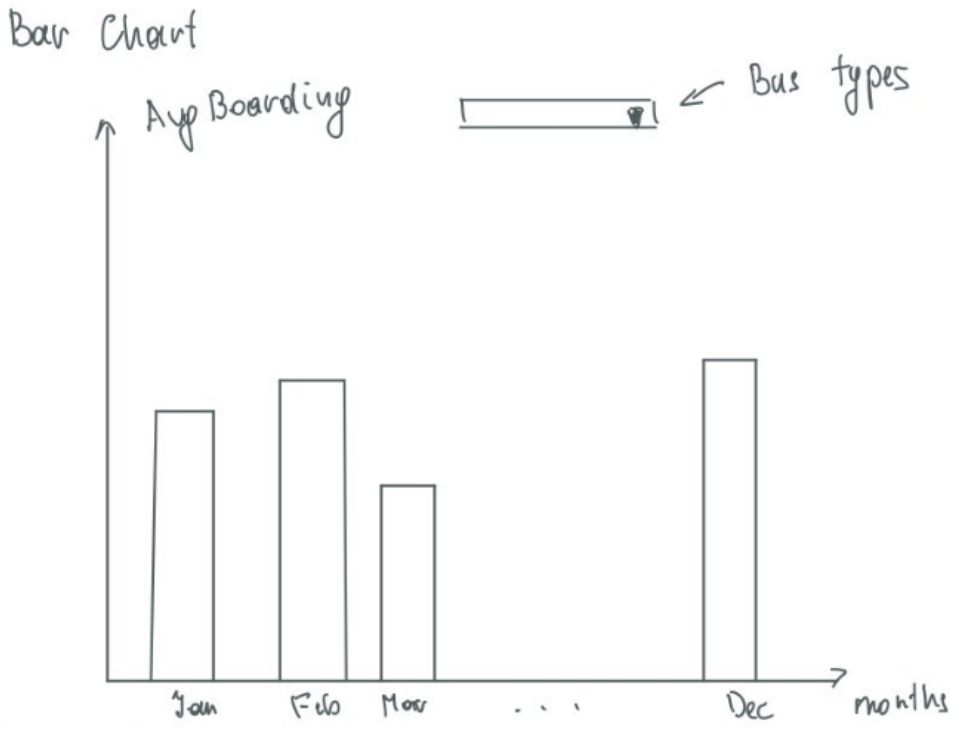
Y-axis: average boarding data(AvgBoarding);

Color: regular on mouse out, gray on mouse on;

Label: sort by bus types, same as in map (drop-down);

Timeline: 2017-2022 (drop-down);

*Interactions: brushing over routes on the map allows viewing the data for the routes in the bar chart, picking different bus types filters the data in the bar chart.



c. **Scatter plot** for interactions with map:

***Basic settings:**

X-axis: average alighting (AvgAlighting);

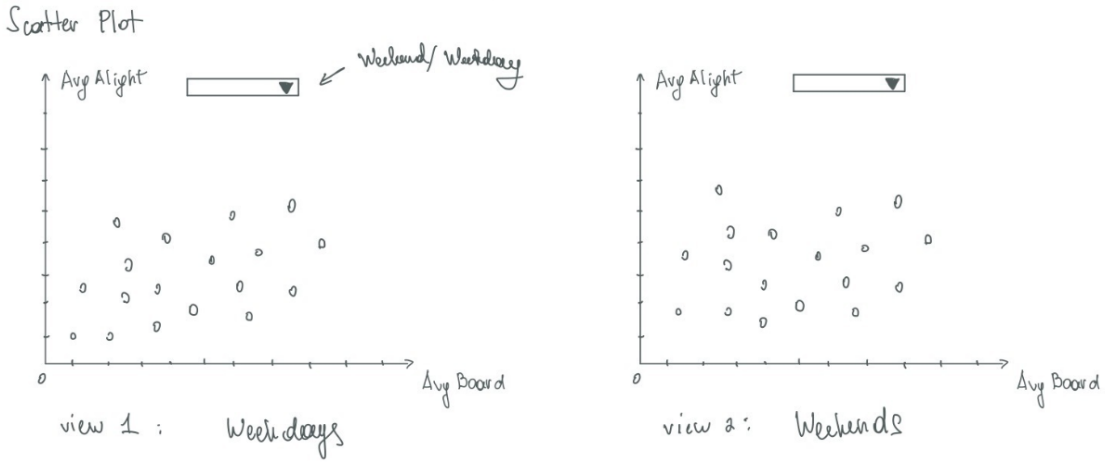
Y-axis: average boarding (AvgBoarding);

Color of circles: months encoded (+ legend for the colors).

Label: sort by types of days: week day/weekend (drop-down);

Timeline: 2017-2022 (drop-down).

*Interactions: brushing over routes on the map allows viewing the data for the routes in the bar chart, picking different day types filters the data in the bar chart.



Must-have Features.
Correct data application and processing, correct map about traffic routes, correct add-on map (layer) about bus stops, drop-down selectors/sliders for year switch (if we acquire data), map zoom from the state to city view and back, brushing over the map with routes/stop selection, charts with interaction (connected to the map).

Optional Features.
Highlighted special locations, district borders, population distributions in the state, **tooltips for each route/stop** (done), interactive popout when you choose some “connection” buttons (besides the tooltips), search bar for routes/stops.

4. Timetables:

Overall timetable:

Week	Milena's part	Tark's part	Xiaoya's part	End result for the week
Week 1 (fall break)	Send the request for data acquiring, document the progress. Work on the proposal, upload sketches.	~	Discussing the ideas. Work on the proposal.	Acquired data for traffic routes, proposal draft.
Week 2	Acquire data for bus stops, add files to the data folder on GitHub, upload the proposal.	Help with data preprocessing	Collaborating with the dataset preprocessing. Updating the repository.	Data processing, submit the proposal.
Week 3	Division of working space through different divs, data binding (routes).	Set up basic map view of Utah and cities	Collaborating with the map view.	Initial space division for the map and its extras.
Week 4 (Peer Review)	Data binding (routes) finish, get the feedback for the project.	Add bus and rail stops	Getting additional data for the graphs.	Transit data added to views. Acquire the peer review.
Week 5	Graph ideas & some implementation (if	Add selecting cities from map vs	Collaborating with the interactivity for routes map and graphs	Add interactivity for map routes and stops, add filters,

	time allows)		(suggestions).	add graphs.
Week 6 (Project Milestone)	Peer feedback, debugging (if needed)	Add info on route selection		Finish graphs and filters, debugging.
Week 7	Start graph implementing.	Map with brushes and clearing them	Optimizing the vis. Add potential analysis.	Potential extra work for the visualization as described above.
Week 8	Finish graph implementing, styling the webpage, choosing color scales. Partially updating process book.	Map with brushes and clearing them, interactions between map and graphs, polishing the graph visualizations.	Collecting the required files, updating process book.	Finalize the changes made, make the presentation video, combine all the required files for submission.

Milestone 1 timetable(Nov. 7th - Nov. 11th)

Date	Milena's part	Tark's part	Xiaoya's part	End result for the week
Nov. 7th				1. Hand in our code and our process book in its current state. * Completed data acquisition, data structures are in place. * Working visualization prototype. * GitHub repository updated and opened (or closed with link for instructors). 2.Review with the each other 3. Upload feedback review
Nov. 8th			Write the skeleton of process book, clarify graph details; Specify timetable and tasks.	
Nov. 9th			Data acquisition listed above. Data processing to the right structures.	
Nov. 10th	Write the feedback review file.	Map graph draft, with background and routes on it.		
Nov. 11th	Upload the feedback review file on Github. Edits on the process book.	Create a release on github.	Update the process book. Commit changes to github.	

Final timetable(Nov. 12th - Dec. 4th)

Date	Milena's part	Tark's part	Xiaoya's part	End result for the week
Nov. 12th - Dec. 4th	Drawing the scatter plot. Styling the website.	Drawing the map and box plot. Make interactions between the three plots by filters and brush.	Drawing the bar chart. Announce the website and write the process book.	Finished the must-have features from our process book. Implementing an informative and beautiful website.

Visual Design Evolution:

1. Line chart to bar chart.

   In milestone 1 we have a map of transit routes in Salt Lake City, which is straightforward. And we intend to use a line chart to select the busiest bus lines to draw the distributions of average onboarding data in Weekday, Saturday and Sunday. The sketch is shown in Fig 2-1.

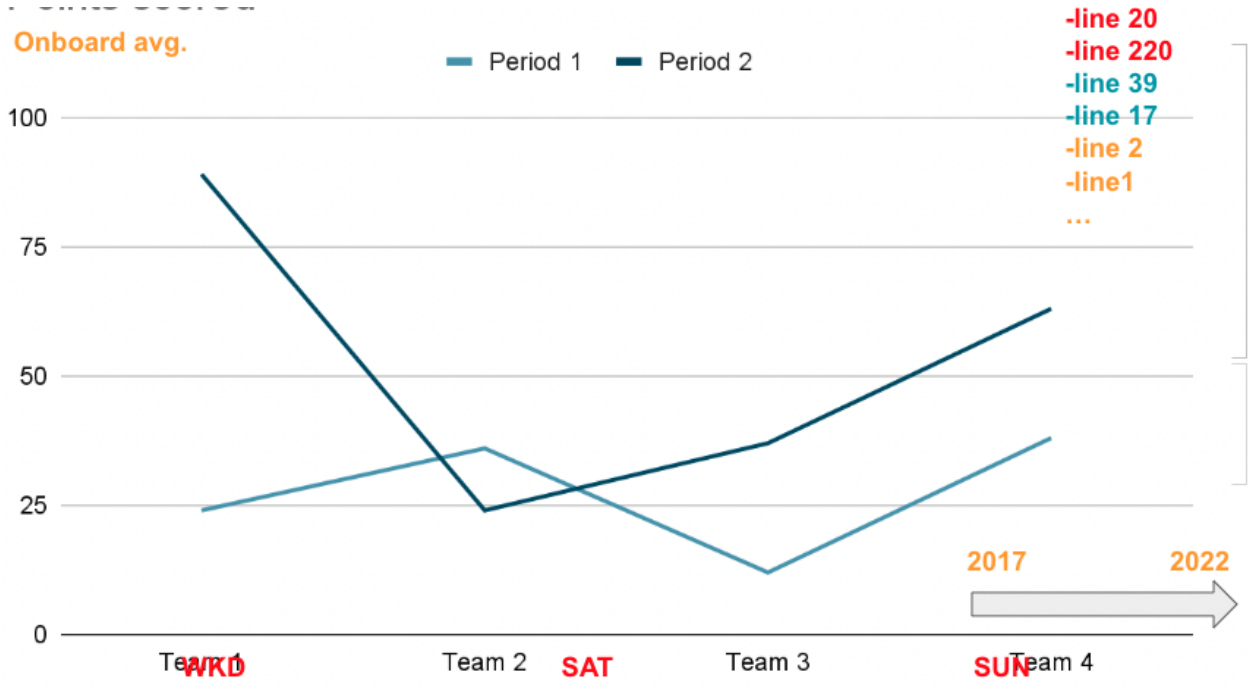


Fig 2-1. The line chart design

We then found the lines are three-fold, which is determined by our data of average aboardings. The line chart is not the best way to show the onboarding numbers of bus lines. We changed the design to a bar chart to illustrate the same data, shown in Fig 2-2.

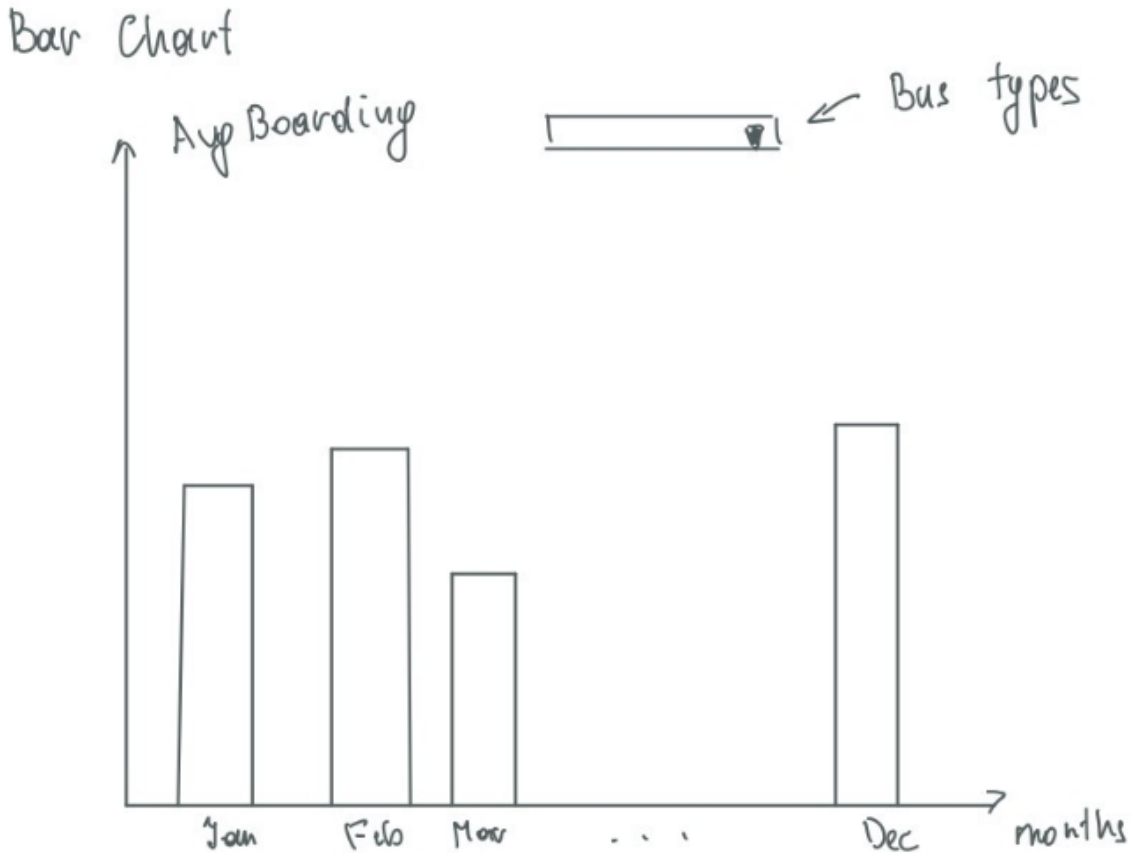


Fig 2-2. The barchart design

A bar chart filtered by months is better for illustrating the distribution of onboardings over time. A selector of bus types is used to separate the different types of the bus lines in the data. We can then analyze the use of different types of buses over the change in months. Also there is a change over weekdays and weekends, which are encoded in another selector of day types.

Final implementation:

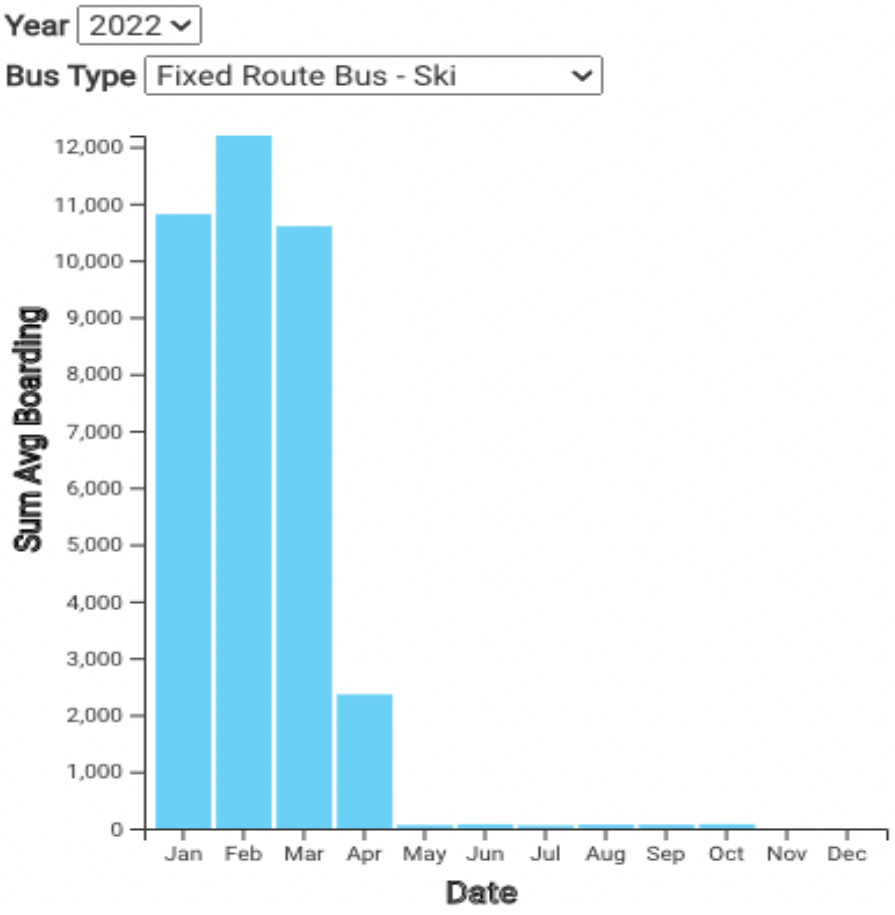


Fig 2-3. The barchart implementation

In the implemented bar chart we show the summation of average onboarding people over months in all eight types of bus lines in SLC. If you simply choose the “Bus Type”, it will show the distribution over months. Due to the large scale in our data, 2020-2022 are chosen in all of our plots. There is only data before November in 2022, when the data has been acquired.

2. Bubble chart to scatter plot.

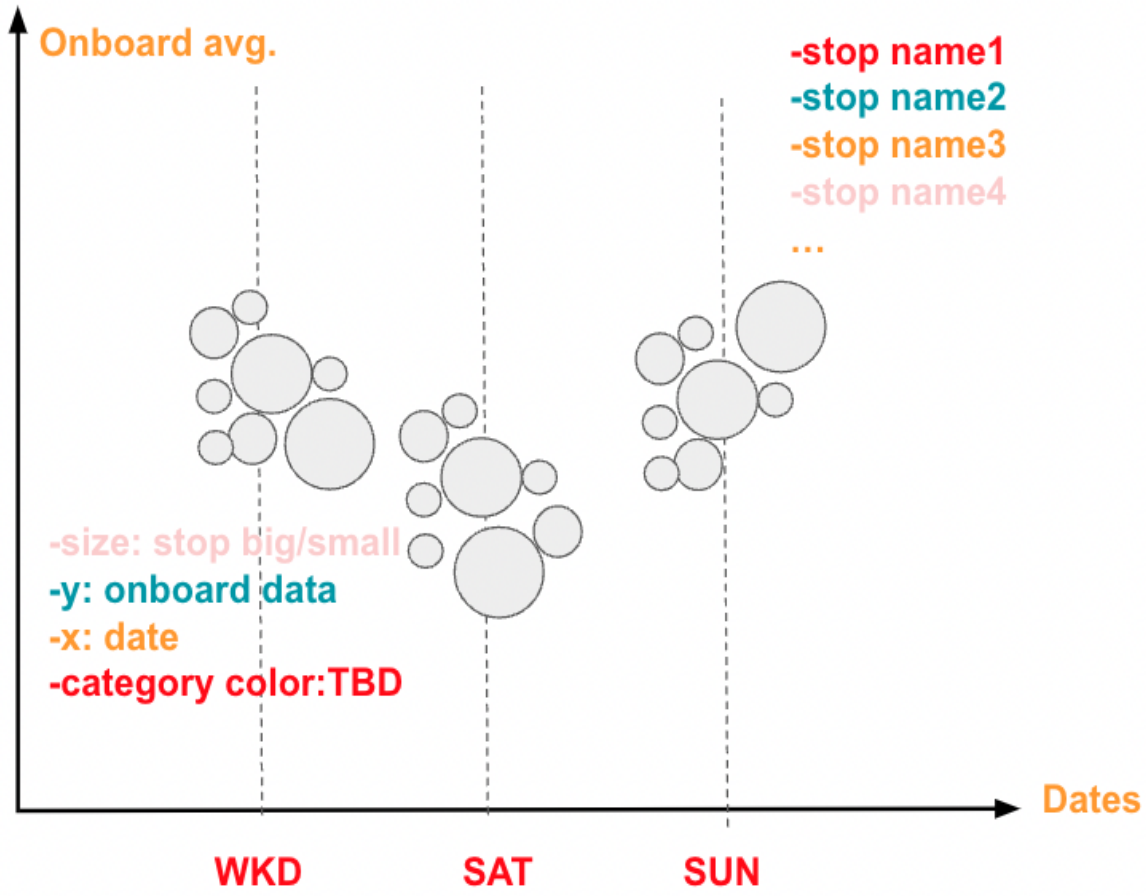


Fig 3-1. The bubble chart design

Our initial thought about the visualization was to use a bubble chart to illustrate the stop-to-level average onboardings. Problems with this visualization are the missing data encoded with the category color of bubbles. Another problem is that there are nearly 2000 stops in our data, which would incur a concern of running time. The last problem is that we design an inappropriate x-axis to divide bubbles. We have only three divisions of dates. Bubbles would aggregate if we use three divisions. We found that a scatter plot was a better way here, shown in Fig 3-2.

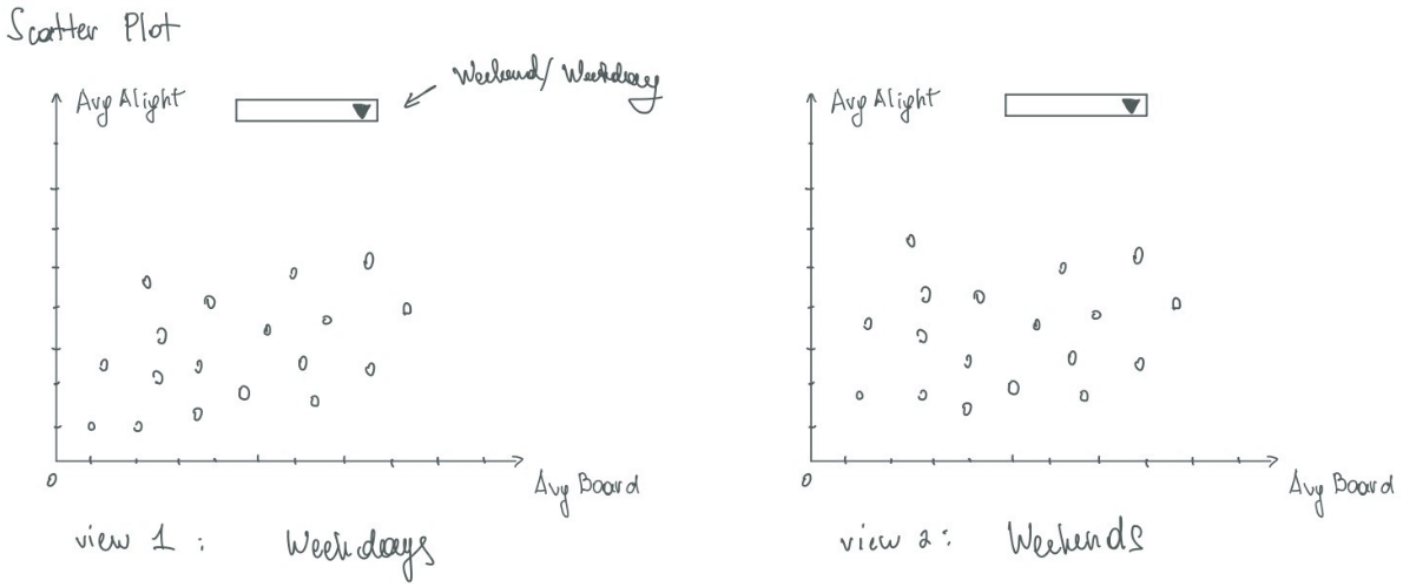


Fig 3-2. The scatter plot design

We planned to do this in two possible plots divided into weekdays and weekends by choosing an appropriate version in the selector. Another data used here is the average amount of people alighting in every bus stop. The user may be interested in how many people are getting on board at this stop and how many people are alighting at the same stop, which this visualization gives the answer to.

Final implementation:

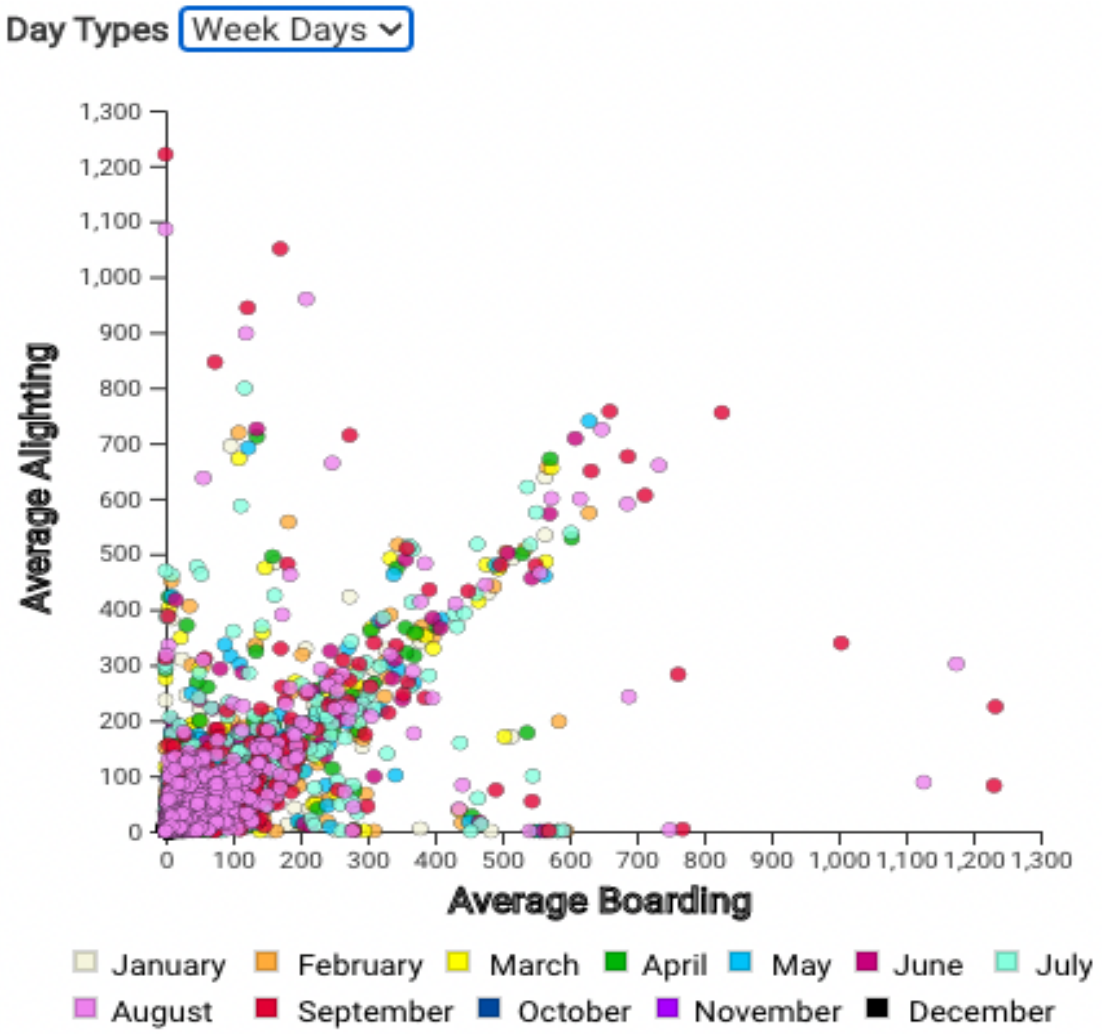


Fig 3-3. The scatter plot implementation

We used a described earlier scatter plot to plot the distributions between average onboardings and average alightings at the stops. We also implemented a selector of day types to see the differences between weekdays and weekends. Different colors are used to categorize the months.

Realization of Project Objectives (partially from proposal):

In our visualization we provide answers to the question below.

1. How were the routes of different buses changing over the years, and why?
To avoid page crashes, we simply choose the nearest three years to draw our visualization. That means the routes did not change obviously. We simply offer the traffic map for public transportation in Salt Lake City. The user can use a zoom tool to see the high-level view of the map and the details of routes in our city. It is easy to see stops and traffic routes and the shared routes between them.

2.Which route (in a selected city) has the best availability? Which has the worst availability?
We use two plots to show the public transportation in Salt Lake City. The user can choose different days and different bus types to see which type of lines has the biggest onboardings and alightings. It can also interactively choose lines from the map to see the detailed information of lines and also the availability shown in the bar chart and scatter plot simultaneously.

3. Is there a relationship between usage and number of stops?
It is hard to see the relationship between usage and the number of stops directly from our visualization. But we can see the usage of stops in different months and dates, which can be meaningful for route planning and go-out schedules.

4. Does usage of routes correspond to the number of other route connections?
We only have the usage data for bus routes. So for the bus routes, the user can simply choose a single line or use a brush to choose multiple lines to see the detailed information. If the user clicks on lines, there would be arrow buttons in the tooltip to switch between the lines who share the same clicked route. At the same time, they can view the usage by the type of bus lines in the bar chart. They may acquire the answer by this interactivity.

5. The busiest time slots of the public transportation(potentially).
The UTA only calculates data by averaging data in Weekdays, Saturdays and Sundays. So we have no answer for this question in our visualization, which is also optional.

6. Is there any chance of optimizing public transportations according to our findings?
We illustrate the usage in different dates by route-level and stop-level. By combining the map, people can get thoughts on optimizing the public transportations from our website.

7. What are the connections between the traffic network and the population distributions in the city, and is there any drawback of the route's design considering the distribution(potentially)?

We only got the population distribution data in 2017. We chose to focus on the usage and availability of traffic routes. So this optional question is discarded.

8. Which city has the best bus transportation availability? (potentially)
Only the SLC is analyzed in our implementation.

Visualization analysis:

We use three visual designs in our website, listed below from Fig 1-3.

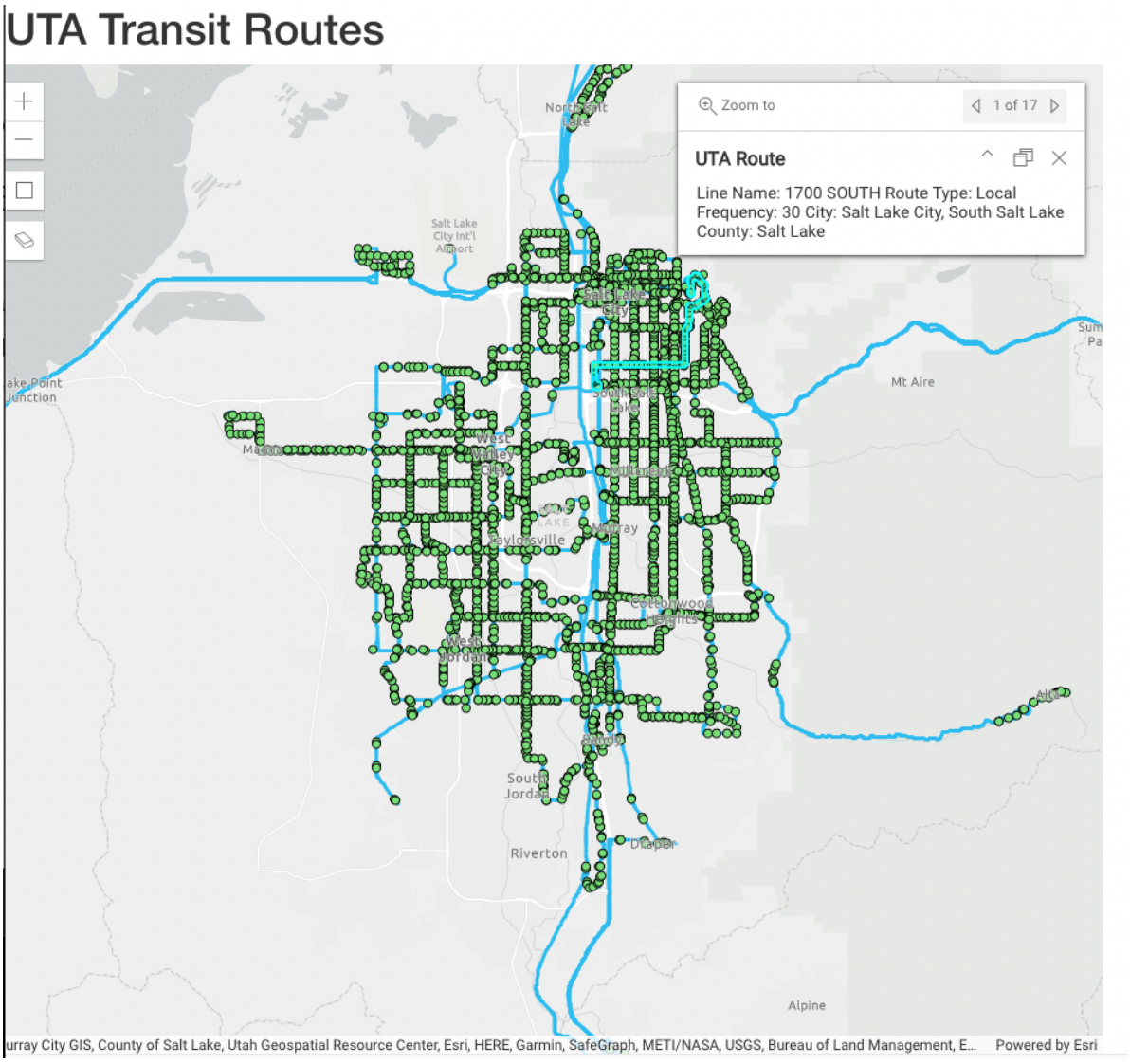


Fig 1. Routes map across Salt Lake City.

Visual Encodings:

As far as the visual encodings, we used markers like points to show stops and lines to show traffic routes in the implemented map. The visibility of the map changes when the user clicks on a route to offer interactivity.

An informative message appears when a user interacts with an element. One more interactivity here is the arrow button to show the linked routes with the clicked ones.

Information we can get:

The user can see the distribution of public traffic routes in Salt Lake City. By choosing the element, they can clearly obtain the route information and the linked routes with it.

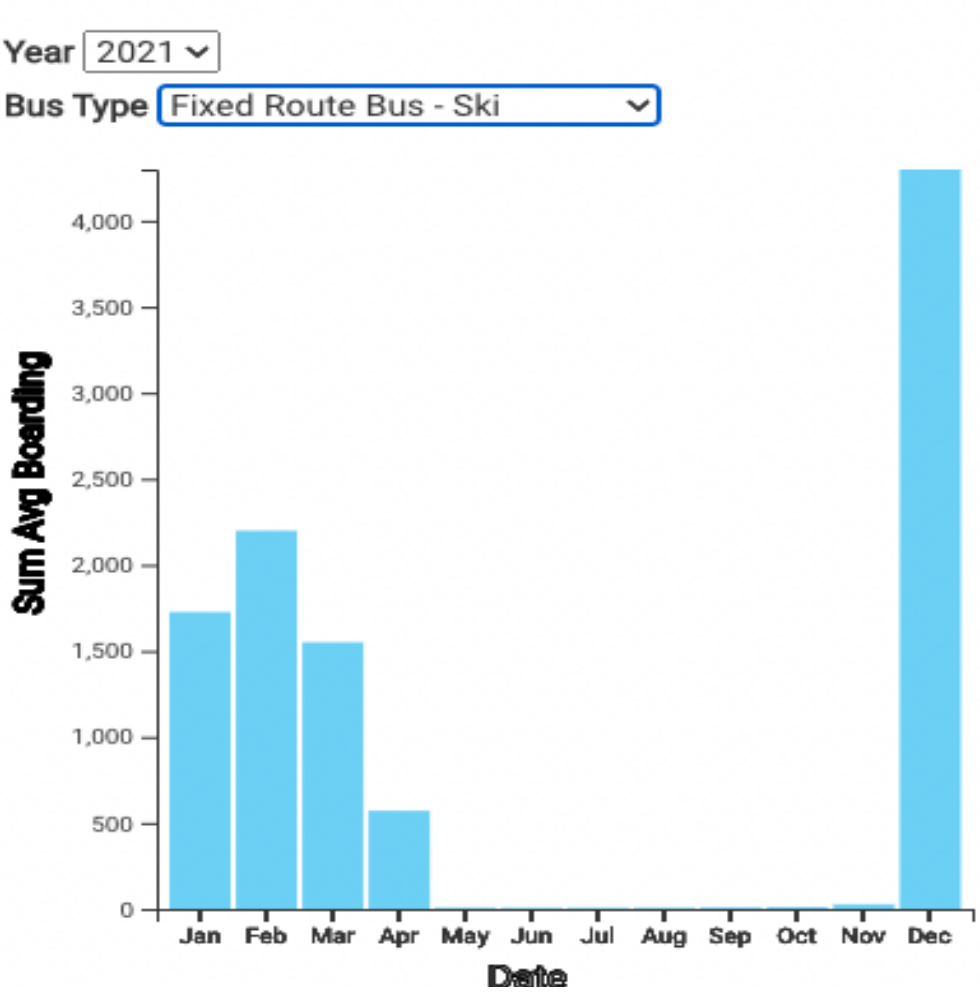


Fig 2. Bar chart for summation of average onboardings from months.

Visual Encodings:

Because our data is both quantitative and categorical. We use marks of lines and channels of vertical lengths and horizontal positions to show our data. Use two selectors of year and bus type for interactivity and story-telling. The user can see the changes between months of different bus types by interacting with the options.

Information we can get:

One interesting observation is that by choosing the “Fixed Bus Route - Ski”, we can see the usage of it reaches the top in December of every year. Salt Lake City contains famous ski resorts, so the ski lines are frequently used during the winter holiday season, especially in December, January, February and sometimes even March.

You can also see the usage of “Fixed Route Bus - Regular”. There is no obvious difference in its usage over months, partly due to the fact that it is one of the busiest lines in the city.

Similar season-connection occurs in “Fixed Route Bus - Express” .

The most frequently used type is “Fixed Route Bus - Regular”. The following ones are “Fixed Bus Route - Ski”, “Fixed Route Bus - UTA Rapid”, “Light Rail” and “Commuter Rail” .

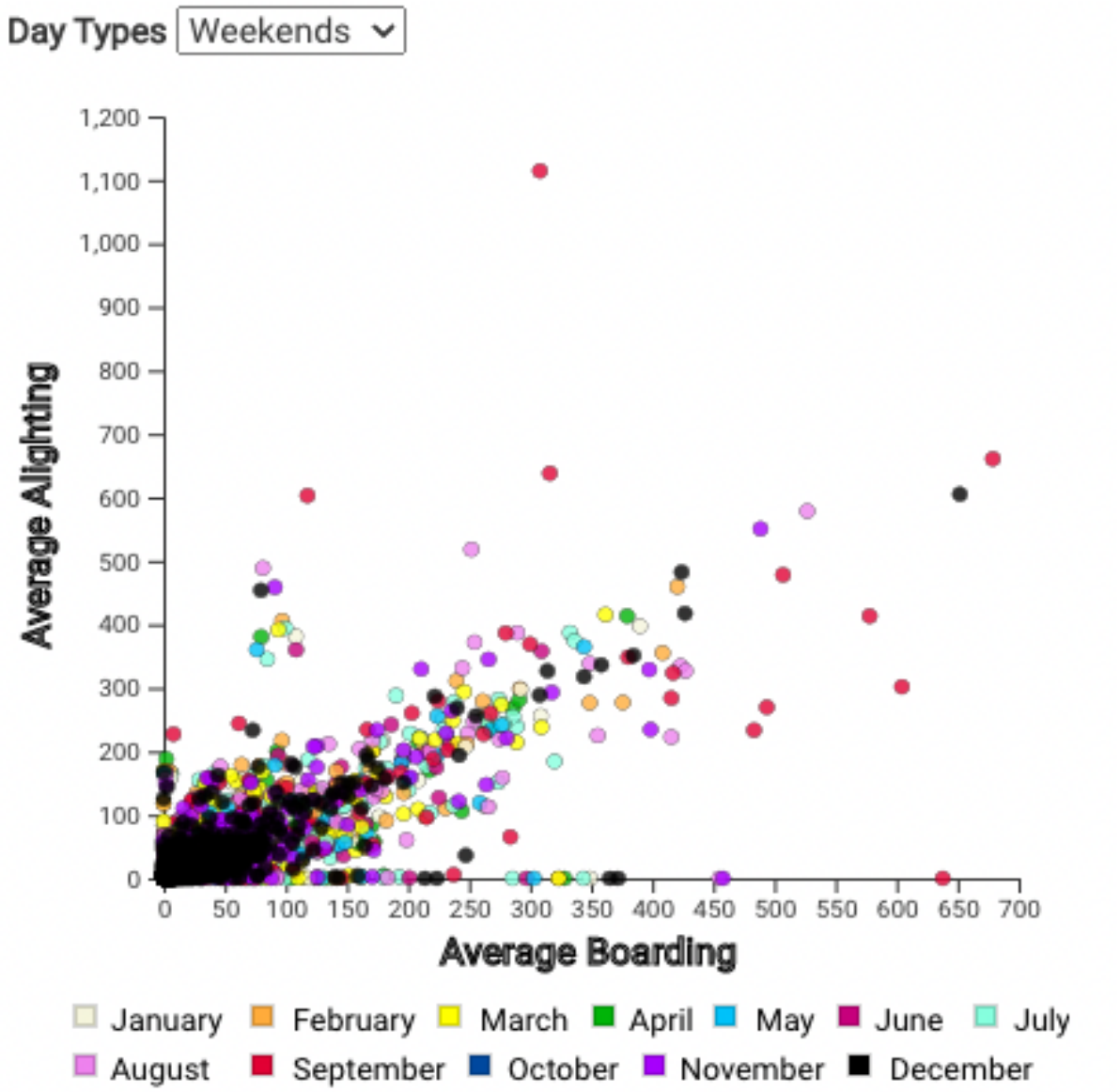


Fig 3. Scatter plot for average onboardings and alightings from months.

Visual Encodings:

We use points as marks and channels of vertical and horizontal positions to show the quantitative properties of our data. We use an identity channel of color hue to show the categorical properties of our data.

Use two selectors of year and bus type for interactivity and story-telling. The user can see the changes of stop usages between months directly from the plot and of different bus types by interacting with the selections.

Information we can get:

Most stops have low average onboardings and alightings in December, while several stops are frequently used at the same time. We can estimate these stops are near the ski places in the city.

Stops are frequently used in September, November and August. The following are July and January. That may be connected to special events in Salt Lake City like the school events.

The usage of stops in 2021 is obviously lower than 2020 and 2022. It may be influenced by the spread of the Coronavirus in this city.

Interactivity and Story-telling:

We offer a public selector of years for the bar chart and scatter plots. The user can see the information in the same year. We also offer separate selectors for these two plots according to the data embedded.

For the interactivity between the three plots, we offer a tool of brush for the user to interactively choose the routes they are interested in and see the changes in bar chart and scatter plot. We hope people can find a lot of fun on our website.

Possible future extensions:

We can advance the map with special positions to see the connections between stops and positions.

The scatter plot used 12 colors to illustrate each of the months. It may be too much for the user to precisely catch the difference in color. Therefore, we could improve the design more by adding another set of dropdowns or some other type of selectors for months representation of the data. There is much information that deserves to be analyzed.

We can also work on implementing the division of the map into neighborhoods to further analyze the possible connections between most populated neighborhoods and route's use and availability.

We hope the audience can acquire the information they are interested in from our website and we are open to discuss our work with the UTA or others.