BART Ridership Visualization

Kayu Chen, Shubham Goel, Robyn Perry

I. Project Goals

We set out to create a visualization that would enable users to explore a couple of questions, and we intended to narrow down to a single, clear question that it would answer from the following two:

1. **Which stations have seen dramatic changes in ridership over time?**
2. **Who takes BART?**

Primarily, we wanted our visualization to show ridership flows, and hopefully highlight changes over time, particularly between any source and any destination (question 1 above). We hoped that we'd be able to illuminate bigger trends in transbay travel as well between the East Bay and the Peninsula.

As a secondary goal, we wanted to incorporate at least some of the interesting data from the 2008 Station Profile Survey that we had access to to understand who rides BART (question 2) and whether there are patterns in the demographics of people flowing between sets of stations; this includes data such as gender, income, method of transportation to BART.

We also started out with the goal of doing an accompanying infographic that would tell the story of some of our insights about BART and its riders; this would include the history, possibly incorporate interesting trivia, and a more detailed graphic of the map/BART lines.

We wanted our visualization to allow most to all of these specific tasks:

- view changes over time in total riders using particular stations
- compare passenger flows from any two sets of two stations (or more?) stations on the BART system
- compare BART usage in 2006 to 2013 and years between (and drill down to a particular station for the same time period)
- see changes in usage (if any) between weekday and weekend travel
- learn about BART through an infographic, which would give an overview of just how many people take BART, and the extent of its network (routes, ridership, stations)
- potential future projections of how BART experience would improve with more trains running

During our project process, we also developed some additional goals that drove our design. We wanted to emphasize content over controls, and we aimed to make interaction as intuitive as possible rather than providing instruction.

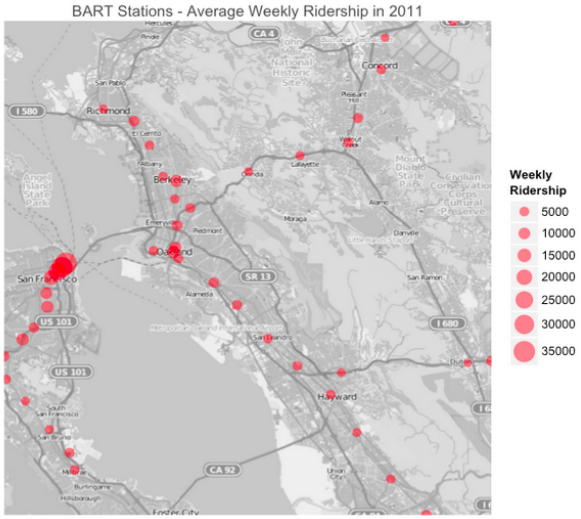
BART Ridership Visualization

Kayu Chen, Shubham Goel, Robyn Perry

We also wanted to enable brushing and/or linking so that users' exploration of the data could take place on multiple dimensions simultaneously. In particular, we imagined users alternating between exploring the dimension of time and comparing different source and destination stations.

II. Related Work

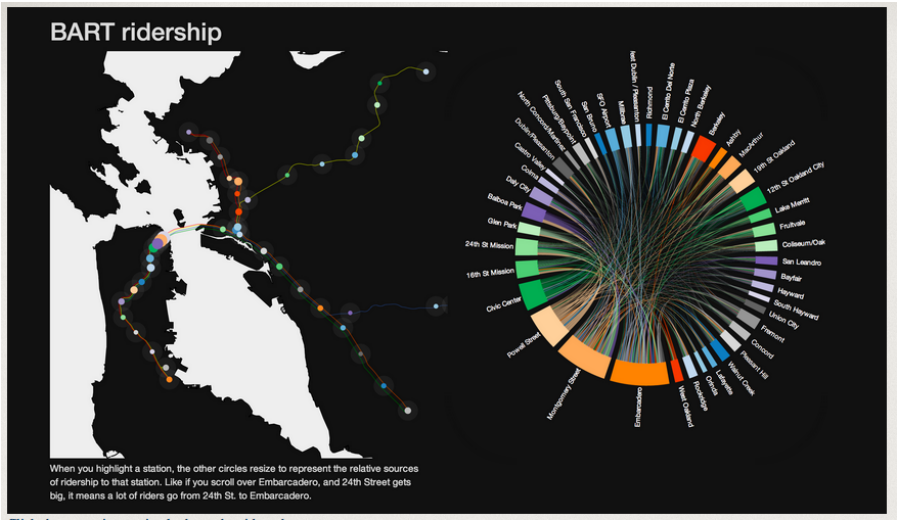
The two visualizations shown below, and this work [http://www.danielledai.com/bartaccess/] about BART access is the extent of the related work in this area that we've come across so far.



The first is the above BART station map [http://gastonsanchez.com/blog/crunching-data/how-to/visualization/2012/05/25/Bart-Ridership.html] that uses red bubbles of varying sizes to represent weekly ridership at each station. We wanted to go beyond visualization alone and enable exploration based on users' BART riding interests and experience. We also wanted to show interesting relationships between stations with ridership flow data rather than stop at comparing station traffic alone as this one does. The small multiples shown in this report by the same author gave us inspiration for one step of our process: [https://docs.google.com/presentation/d/1B9tFbVbN3LleTd_nVkqq24RUL9oPAeo7w3vQ3PcMZYw/pub?start=false&loop=false&delayms=3000#slide=id.g3f0b065_0_102].

BART Ridership Visualization

Kayu Chen, Shubham Goel, Robyn Perry



The second above includes a chord diagram and has the problems inherent in such a visualization: it's beautiful, but doesn't necessarily allow for precise comparisons of passengers between two different start and endpoints [http://enjalot.github.io/bart/#chapter-05]. Overall, you can see that Embarcadero and Montgomery are heavily trafficked, but not whether there are more trips between Embarcadero and Berkeley as compared to trips Millbrae, for example.
We clarified early on as a group that we wanted to use the BART map as an intuitive site of exploration and interaction, rather than choosing a form unrelated to the system itself. We liked the pairing of two visuals to reinforce the data, but the comparison of areas of small, sometimes tightly-packed circles fails to allow precise comparison.

III. Description

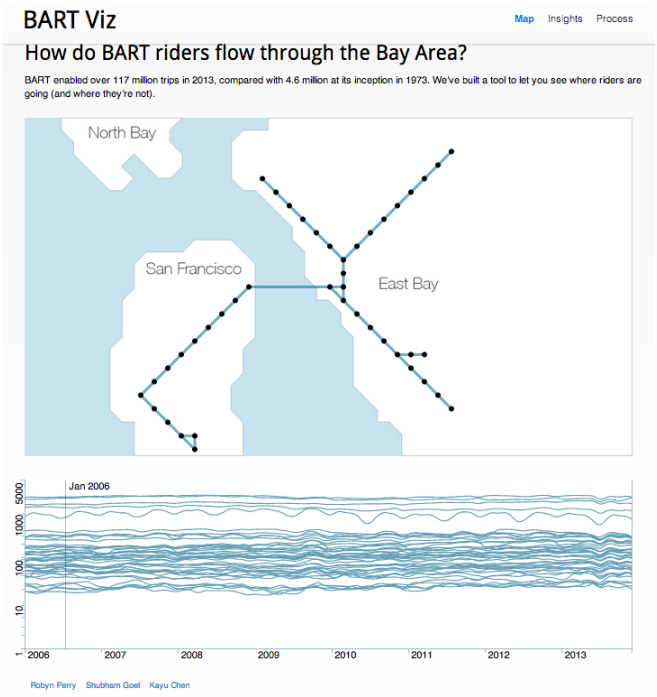
See demo of interactive visualization here [http://youtu.be/CbVhVZPQsdY].

We produced an interactive map accompanied by a dynamically changing time-series that enhances the view of ridership by showing changes over time. Our visualization allows the selection of a source station, which upon clicking will allow users to see using a ridership represented as flows outward from that source station toward all other stations.

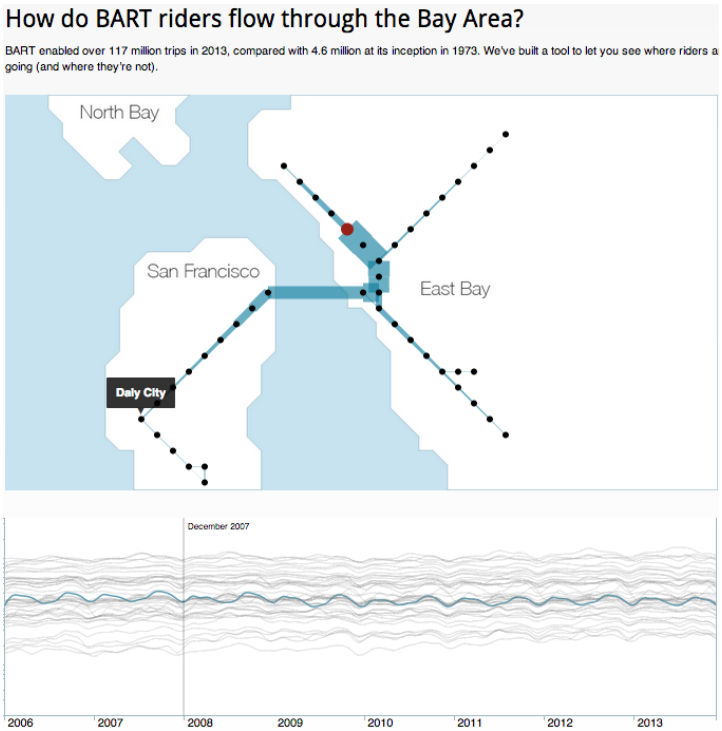
This first image shows the default state of the main Map page with no source station selected.

BART Ridership Visualization

Kayu Chen, Shubham Goel, Robyn Perry



Below, Berkeley has been chosen as the source station as indicated by the red dot that's larger than the other black dots. Daly City is the destination as indicated by the hover text showing "Daly City". The line in the time-series corresponds with monthly trips over time starting at Berkeley and ending at Daly City. Already, it's easy to spot seasonal variation in the line drawn. It appears that there are dips and peaks at regular intervals throughout each year that repeat.



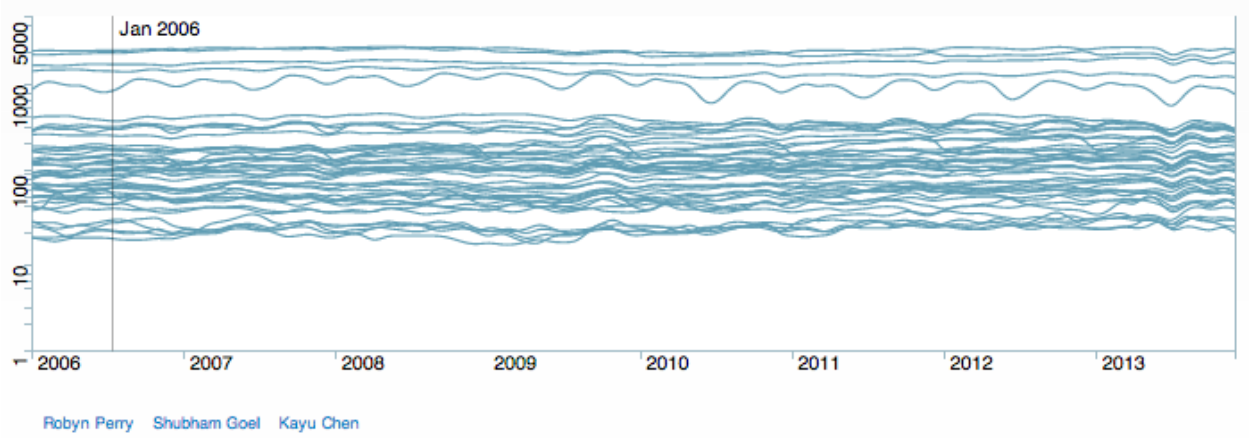
BART Ridership Visualization

Kayu Chen, Shubham Goel, Robyn Perry

The time-series shows the number of people going from the source station to a specific destination, which is chosen by hovering over any station on the map after choosing a source. This line is highlighted in blue and represents the fluctuations of ridership between the selected source station and destination station. Moving the mouse to the left or right over the time series triggers fluctuations in the flows rendered on the map through all the years we have BART data for. The flow map shrinks or grows, matching decreases or increases in ridership over time.

Hover text on the map shows station names according to the cursor position. The time series also displays the month and year of data being displayed as you move your mouse through time.

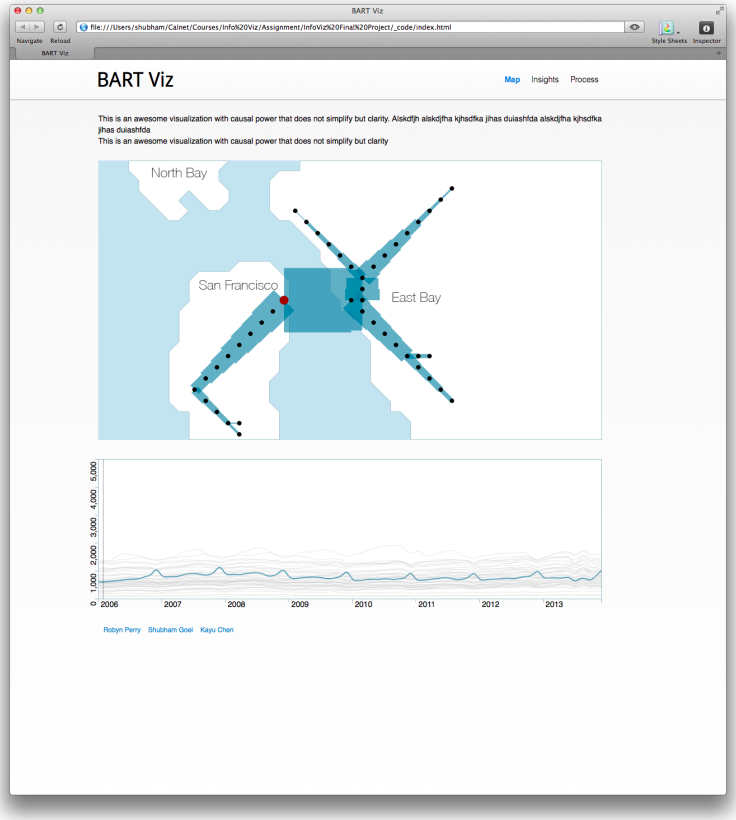
The log scale came in handy to display ridership along the y-axis of the time series portion of the visualization, as shown below:



This helps to make the range of ridership volume more visible. Without the log scale, time series lines representing monthly ridership under 100 all stack on top on each other making it impossible to differentiate individual lines, as shown below.

BART Ridership Visualization

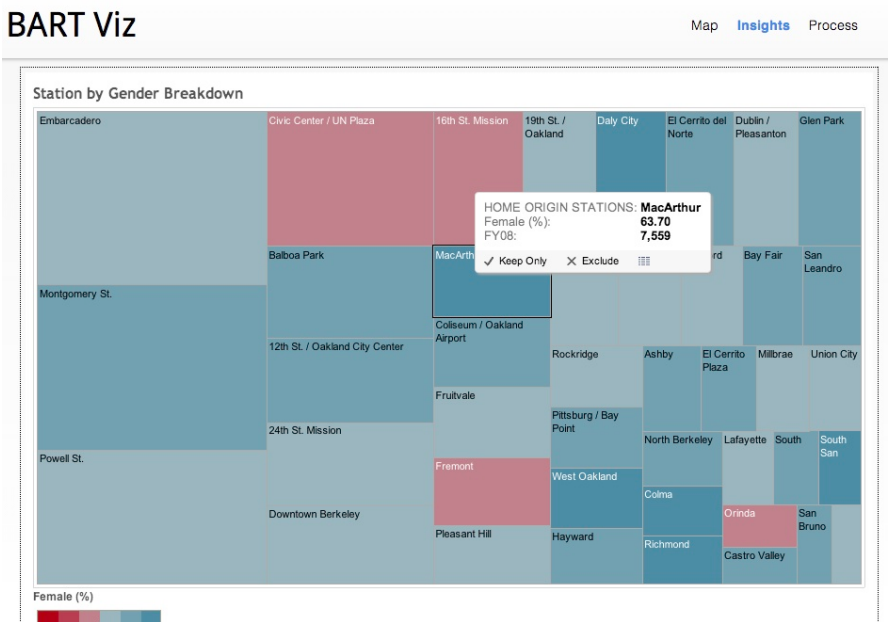
Kayu Chen, Shubham Goel, Robyn Perry



The Insights tab has a Tableau tree map visualization that shows the gender breakdown by BART station, and the percentage of female to male riders. The station profile data showed that more females use BART than males, so we wanted to highlight this. On the map, the size of the area represents the weekly ridership average of that station for 2008 when the survey was taken. The colors represent the ratio of female to male riders: the more blue the rectangle, the more females take that station, while red represents more males. We decided to use two colors rather than one, in order to emphasize that the majority of BART stations are used by women (so the few that have slightly more men are obvious), and if you hover over each area, you can see the specific percentage breakdown.

BART Ridership Visualization

Kayu Chen, Shubham Goel, Robyn Perry



In summary, our visualization allows exploration of a number of variables including source station, destination station and their corresponding ridership, and change over time.

IV. Data Sources

- BART monthly ridership data from 2006-present [http://www.bart.gov/about/reports/ridership]
- Station Profile Report [http://www.bart.gov/sites/default/files/docs/2008StationProfileReport_web.pdf] (we obtained the Excel data behind this publicly available PDF thanks to a contact that Raymon Sutedjo-The connected us with).
- Weekly Average Ridership [http://www.bart.gov/about/reports]

V. Tools

This is a summary of the tools we used and for what purpose we used them:

	Purpose:
Excel	Calculations, some data manipulation and aggregation
Tableau	Exploratory data analysis of Station Profile Survey data; rendering of gender treemap

BART Ridership Visualization

Kayu Chen, Shubham Goel, Robyn Perry

Python	Wrote program to read monthly files and generate json file for each source station for entire time period
D3/Javascript	Development of map with lines and dots for each station; creation of small-multiples and later, the time-series portion of the
Photoshop	Development of layout mockups of whole visualization, colors, and smaller components of visualization

VI. Steps

Design Process

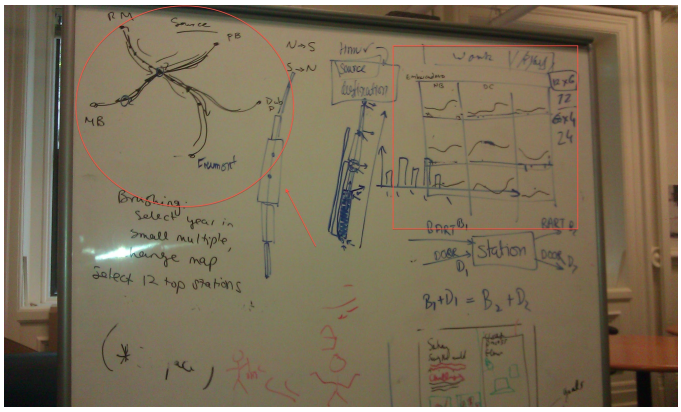
1. We discussed initial ideas, first deciding on a mixed project with interactive and static narrative elements. We envisioned an interactive map showing ridership with a time component and a static infographic in a separate section or pane.

2. Our process followed an iterative cycle of brainstorming ideas, designing mockups according to the brainstorm, coding example versions with test data, and then massaging the actual data to finish implementation. We solicited feedback at various times in the process to refine our product and validate our end goals.

3. We sketched out ideas on paper first during our meetings and then made finalized, color mockups using Photoshop. We tested how certain features would look, such as alternative visual layouts of the map, and the interplay between the two interactive visual components. Shown below is an early version of our thinking which included the idea of small multiple time series (in the red square) to accompany the map (circled). We also intended to create a Minard-style segmented line on the map, as shown by the red arrow in the middle.

BART Ridership Visualization

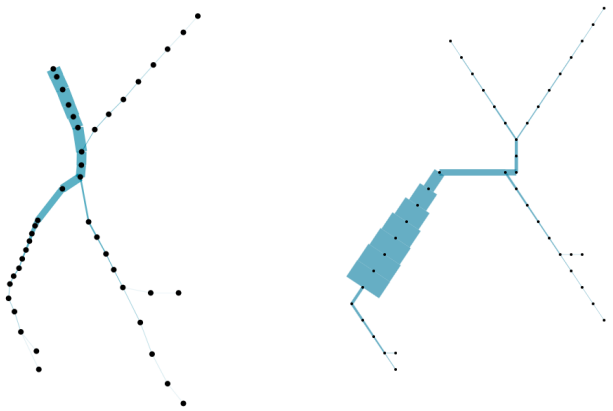
Kayu Chen, Shubham Goel, Robyn Perry



4. We began to focus on the interactive elements and abandoned the infographic idea. We aggressively eliminated ideas that seemed like they wouldn't accomplish a coherent final visualization. For example, an earlier stage of our project included small multiples of time series data, as shown above, for 12-15 destination stations where the single time-series eventually ended up. Feedback from Marti and classmates helped us realize that the small multiples didn't match the data being shown on the map at a given time, and may have caused confusion or distraction.

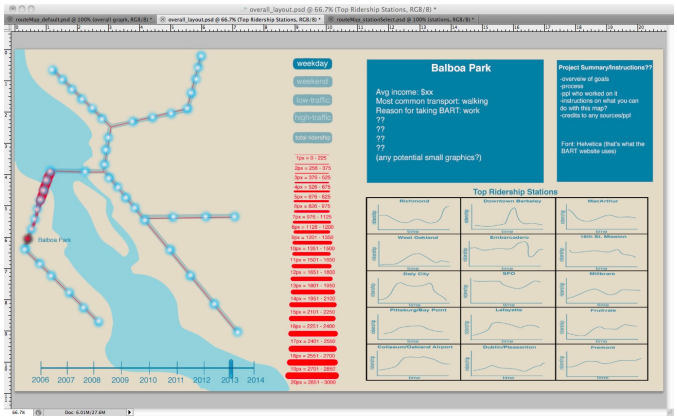
5. Some of the improvements we arrived at included a London tube-style distorted BART map that maintained sharp, simple angles at either 45º, 90º, or 180º; the fade-out of destination lines for all non-selected destinations; muted colors; and an angular representation of the land and Bay.

See the first representation of the map followed by a more angular representation that we stuck with below:



6. We mocked up each refinement in Photoshop to have a clear visual idea of what we wanted to build. This way, we could have a clear vision at each stage, which

made it easier to convey our ideas and get feedback. This also allowed us to actually see what might be feasible, and what may have to be pushed to a lower priority. It also helped to start shaping our list on what would actually have to be built and how.



7. Throughout the entire process, we aimed for a certain thematic aesthetic, based around a minimalistic version of the BART scheme: blue, white, and black. We saved red for the selected source station only. We used clean lines without clutter as to not draw attention away from the main interactive map. A lot of this had to be altered during our final build phase, in order to accommodate for limitations like time and actual deliverables, but the overall theme helped to guide our design process the entire time, and keep a consistent aesthetic in mind.

See the appendix at the end of the report for additional documentation of the process.

Data Manipulation Process

The data required extensive manipulation and massaging.

The ridership data is split into many Excel files with a worksheet for weekday travel and one for weekend travel. Each sheet included a matrix of 43 stations (rows) by 43 (columns) like so:

Exit stations	Entry stations->		ADJUSTED WEEKDAY			Sep 06		
	RM	EN	EP	NB	BK	AS	MA	19
RM	6	151	133	88	389	114	139	129
EN	180	14	130	89	938	182	239	472
EP	142	111	5	57	657	81	115	213
NB	91	76	55	8	222	69	80	182
BK	410	965	677	242	19	425	458	427
AS	112	172	78	61	361	9	106	196
MA	144	228	112	76	429	101	12	143
19	137	484	220	190	386	195	138	8
12	187	630	278	252	509	245	228	75
LM	39	112	97	81	283	97	96	52
FV	97	134	70	82	357	120	163	164
CL	91	245	113	105	345	144	240	231
SL	45	62	25	33	175	64	105	234

BART Ridership Visualization

Kayu Chen, Shubham Goel, Robyn Perry

We did the following:

1. Calculated total ridership entering at each station and subtract drop-offs at each destination to be able to represent each segment of the flow as the total still on the train.
2. Developed a key of all of the segments for each possible source station.
3. Wrote a graph traversal algorithm that includes calculation of riders for each segment from a given source station.
4. Development of a json file for each source station to each destination that included ridership for each month of each year, where the source station was located regionally, and whether the data was for a weekend or weekday.
5. Built a simpler graph with 8 nodes to test our graph traversal algorithm which does the following:
    a. generates directed trees from an undirected graph with cycles for every source station
    b. calculates the weight of all the children for every node of the directed tree
6. Tested it on a data set for one month.
7. This data went into the initial map prototype that we demo-ed for the Mid-project Presentation.
8. Created a python program to read all Excel files and generate JSON files for the aggregated data to power the map visualization and the time-series. This program deals with all the discrepancies in the ridership data. This problem was solved in steps, first for one source destination station and all the data for one year and later for all the stations and all the time periods.

Results

We were fortunate to have multiple rounds of feedback during our project, including some ad-hoc feedback sessions with generous friends and family members.
After our mid-project presentation, we scrapped both the small multiples and the idea of incorporating a simple overlay showing overall changes in ridership since BART's inception as a result of the feedback we received in class.

Due to the positive feedback about the demographic and other interesting data from the 2008 Station Profile Survey, including similar feedback from Marti, we decided to include a page on demographic insights about BART rides, which we had worried might detract from our main goal of displaying ridership flows. Interest countered this concern.

During the project open house, we received some immensely valuable reactions to our work. As a result, we have a checklist of new feature requests, some of which we had intended to include but didn't have time for.

BART Ridership Visualization

Kayu Chen, Shubham Goel, Robyn Perry

Here's what we hope to do in the next iterations, thanks to the generous feedback of those who tested out the BART ridership Viz:

- Though most liked the spare design, we should add more landmarks to orient users better to the BART map so they can quickly locate stations of interest.
- Find a measure in between a regular scale and a log scale to display the ridership segments more smoothly, particularly because the extra wide segments are not visually pleasing.
- Add hover text on segments to show actual numbers of riders (as well as on corresponding line in time series)
- Choose a default for time-series chart on page load
- On selection of a source station, show all lines in time-series by default or a particular one in a transparent way (currently it shows the source station *as* the destination, which is confusing)
- Keep source and destination labeled while exploring a single route (and fade out the other routes and time-series lines so the link between the map and the chart is more obvious)
- Add "presentation modes" to help users zero in on interesting insights right away (by linking to a particular map/time-series state of interest with some narrative)
- Include weekend ridership data
- Consider blending in data related to sports events or weather to show spikes or lows (and indicating as pop-up text on timeline)
- Add dynamically updating legend that contains time and currently selected source and destination
- Consider BART pricing and why ridership doesn't look like it's gone up over time, in spite of the overall ridership data
- Color time-series lines by region on the BART system to highlight possible regional trends from a selected source to stations on the Peninsula, the East Bay, etc.

For the gender map, we will do the following:

- Alter the color scheme so it's not counter-intuitive to US users (pink for males, blue for females currently)
- Change range shown to show actual range or full range (0-100%)
- Group boxes in tree map by location on BART system
- Add bar chart below showing the breakdown of gender ratio by station in order

Our final visualization is the product of much iteration and thoughtful tussle with an interesting problem. We all want to continue working on it.

BART Ridership Visualization

Kayu Chen, Shubham Goel, Robyn Perry

Accessing Our BART Viz

Currently, view live visualization here: http://whatanapp.com/bartviz/
See demo of interactive visualization here [http://youtu.be/CbVhVZPQsdY].

Distribution of Work

Team Member	Role
Kayu Chen	main graphics designer • overall layout and map development • coordinate determination • mockup generator • color experimentation • mid-project presentation slides • extensive contributions to ideation process
Shubham Goel	main coder, designer of interactive visualization • algorithm development • data massaging with Python script • graph traversal test and implementation of full map • integration of time-series and map • extensive contributions to ideation process
Robyn Perry	designer of interactive visualization, group coordinator • exploratory data analysis • insights page, development of additional demographic charts on income and weekly ridership averages that aren't yet ready for production • coder of small multiples and time-series • main writer for project proposal and final project report • extensive contributions to ideation process

Software Created

Please visit this page for documentation and code related to this project:
https://github.com/nyborrobyn/BART

BART Ridership Visualization

Kayu Chen, Shubham Goel, Robyn Perry

Find the code we wrote in the /_data directory. Additional scripts for data aggregation and munging are in /_docs, and design files are in /_design.

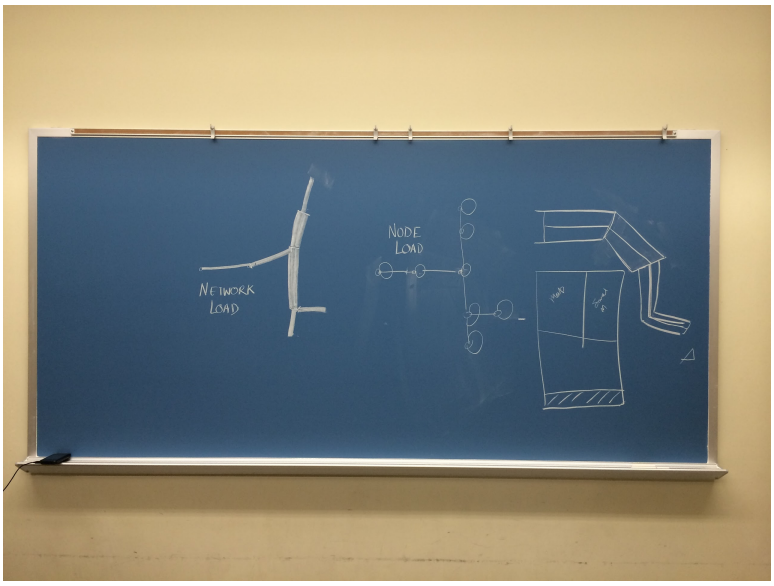
BART Ridership Visualization

Kayu Chen, Shubham Goel, Robyn Perry

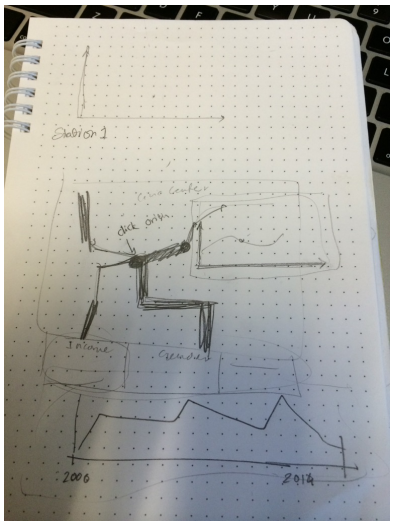
Appendix

Additional Images Documenting Process

The following shows a stage where we considered other ways to represent the load on the BART system, either by representing it at the segments (denoted by “Network Load”) or at the nodes (“Node Load”).



The next image shows a sketch including a time slider. We decided against an explicit slider that *only* performed the function of scrolling through time, remembering the importance of Hans Rosling’s narration in the GapMinder visualization. We thought that using a time slider might mean users could miss important insights in the data as a result of sliding too fast.



BART Ridership Visualization

Kayu Chen, Shubham Goel, Robyn Perry

This final image was one of our final sketches. It displays one of the features we'd like to include, which is coloring of the time-series lines by region to highlight trends in types of BART trips.