CS171 Final Project Process Book

Ina Chen and Fox Morone

April 2014

1 Overview and Motivation

Bike sharing is an increasingly popular mode of transportation within large cities such as Boston, New York, Chicago, D.C., and others. Unlike ride shares or car pooling where location and rides are relatively easily adjusted to rider volume and locations, bike sharing requires infrastructure carefully planned to accommodate riders at convenient locations and stocked to ensure availability of bikes. As the number of bike share users increases, it becomes useful for both city planners and bike program management as well as bike share users to look at ride patterns along with the flow of bikes at stations to complement program planning or program participation. For these reasons, a few bike share programs have released data on ride trips and some have even encouraged clear and creative visualization through visualization challenges. Here, we take advantage of these available bike share data in three cities, Boston, Chicago, and Washington, D.C.

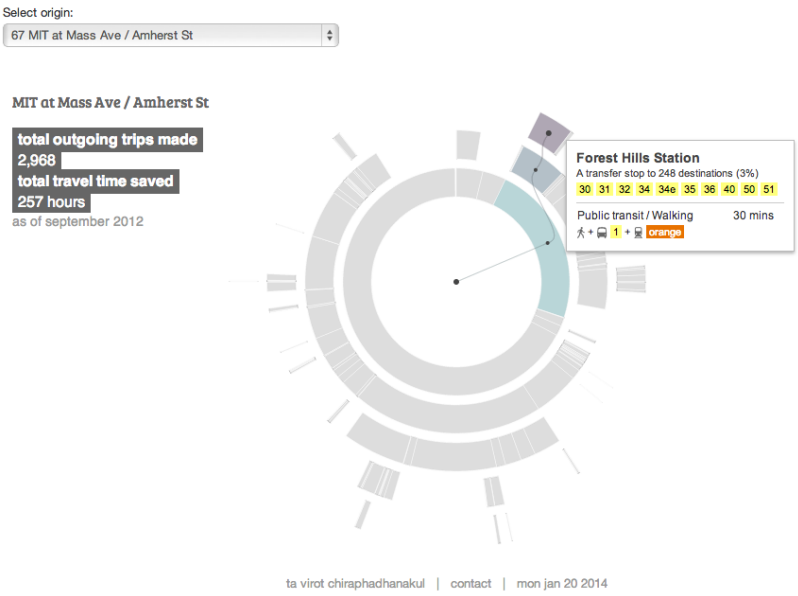
One focus of our visualization is to compare ride patterns between cities. There has been some visualizations of bike share data within individual cities , but we found none that compared bike share patterns between cities. Given the recent trend for programs release bike share data, we thought it would be interesting to compare the ride patterns across different cities. Since our data comes from bike share in large cities, we expect to find similar trends such as commute riders on weekends or tourist riders in summers but also differences that arise due to different geographies or climates of the cities. Comparisons between cities thus could provide both interesting patterns as well as open the way to looking at how bike share programs could improved in different cities.

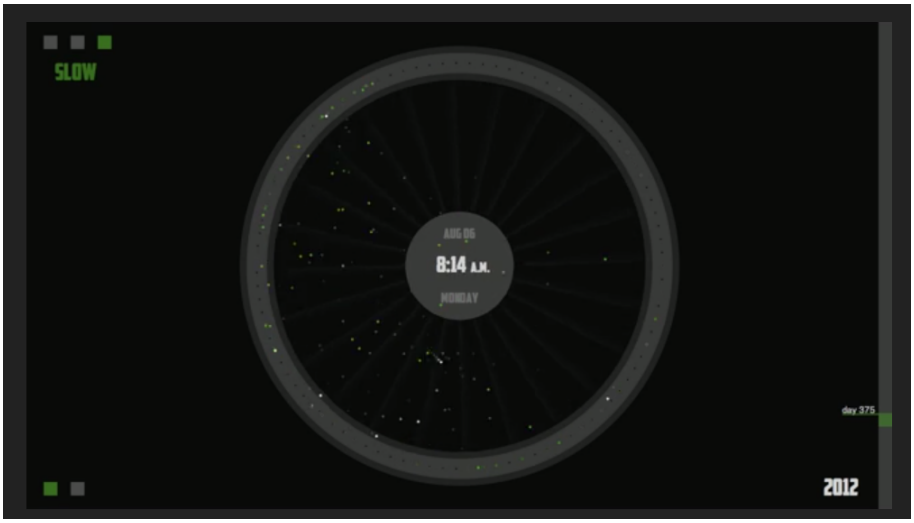
The second focus of our visualization is to look at bike rides data in its native geographical setting. Quite a few bike share data employs chord graphs or summary graphs that, while highlighting certain features, loses the geographical nature of bike rides (see Related Works below). In this focus, we want to complement inter-city comparison visualization with a detailed geographical-based visualization of ride patterns in each city. One particularly interesting feature we wanted to visualize was the flux of bikes at each station and ride routes in and out of each station. In this, we hope to visualize popular stations, trips, and the overall flux of bikes in both space and time.

2 Related Work

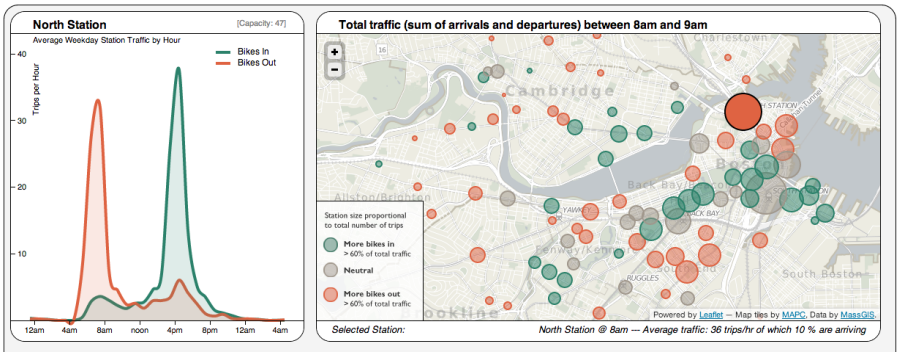
We found and drew inspirations (as well as rejections) from previous bike share visualizations.

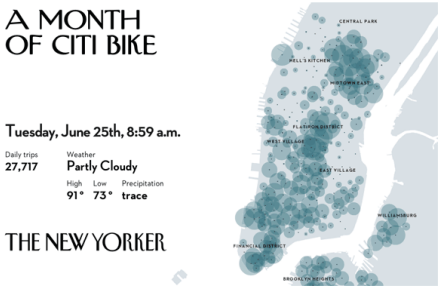
Visualization of rides between stations. We chose not to follow this style.

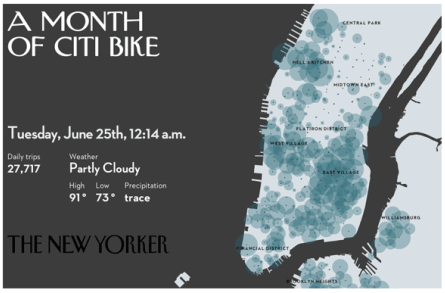


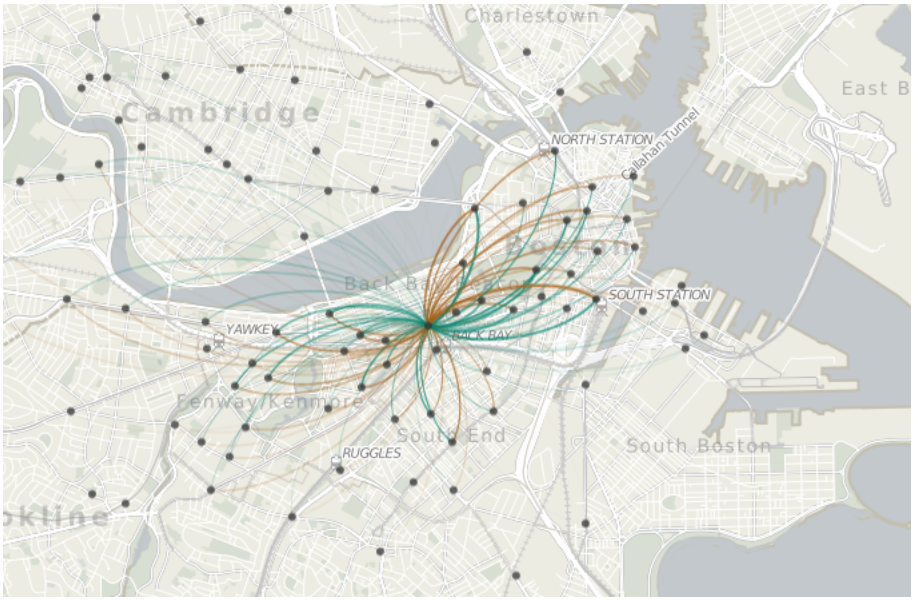


Visualization of station-related on a map. We particularly liked this type of visualization and drew much inspiration from these visualizations

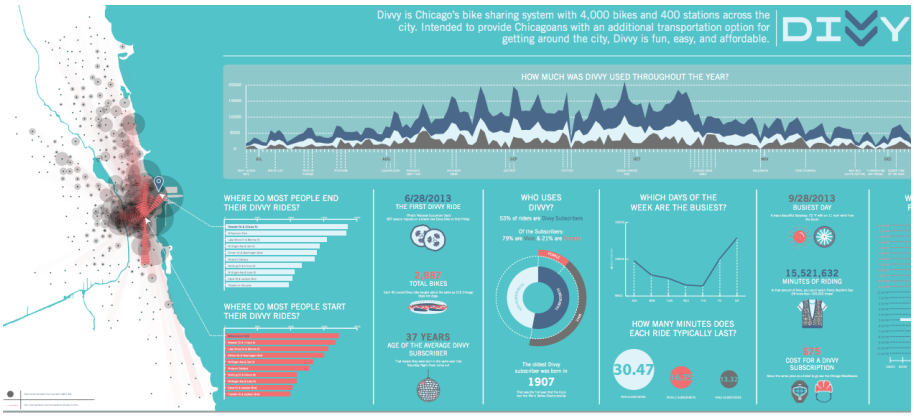


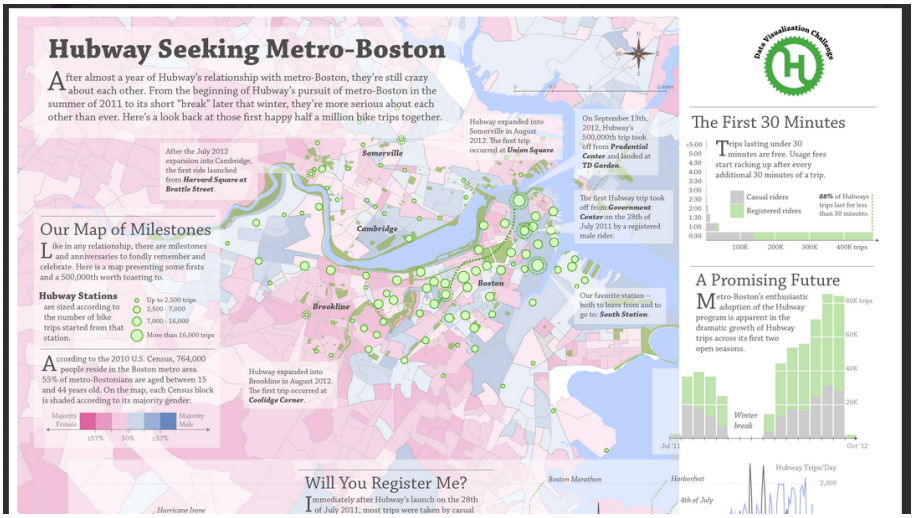






Static Visualizations that told stories. We hope to possibly incorporate some story highlighting feature but will focus on interaction so that the users may explore stories for themselves





3 Questions

Questions we attempt to answer with this visualization include how do bike ride patterns vary between subscriber and casual riders in different cities. Ride patterns is assessed through various metrics including average speed, number of trips, average duration per trip, average distance per trip, distance traveled which is calculated for each day over the course of a year. What are some interesting stories that the visualization say about deviations in bike rides over weeks, seasons, or for special events? What are patterns in trips within a city? What does the flux of bikes in and out of stations say about the commutes or travels of the riders?

As we worked on the project, questions were to adapted such that we can best visualize the data to maximize features that can be clearly presented while also maximizing smooth visualizations by reducing the data volume. This was particular important for the map visualization. While we initially asked if there are interesting stories from tracking bike data over the course of the year. When it became quickly obvious that this was somewhat unfeasible to tackle even the data size (and 300+ stations per city), we focused our attention on asking questions related to average patterns of bike share rides. On average over a month, how does the trip patterns change through the week? We are also considering further reducing the data to ask how bike patterns varied over a week when aggregated through a season.

4 Data

We pulled available CSVs (1000000 rows each) with ride data from each of the three cities. Generally, we have end and start station, end and start date/time, and casual user/subscriber data. we also had to get separate CSVs with stations' latitudes and longitudes.

The CSVs were several hundred megabytes and therefore unsuitable for a d3 project and also un-pushable to Github.

Cleaning was thorough and can be seen in our iPython notebook. Some of the ride data CSVs have strings of station names within their rows. These had to be removed to consolidate CSV size. Also, not all the cities provided distance data, so longitude and latitude values had to be used in conjunction with a global distance calculator to get that statistic.

As of now, we have cleaned all Washington D.C. data in order to start doing preliminary visualizations, but have not finished cleaning the Boston or Chicago files. This should be trivial after completing the D.C. set, however.

Concerning structure, we originally wished to use a CSV which essentially contained all the data we needed, but quickly saw that would also be too large of a file. We continued to restructure and aggregate data, and ultimately we have begun using a JSON file of average daily values of distance, ride duration, etc, which works well for our exploratory visualizations. On the other hand, our initial map visualizations are driven by CSV data which is separated by weekday, hour of the day, and station. Both of these files are tractable for a d3 project.

We thought that it would be interesting to look at the ride share data in the context of neighborhoods in the city. To generate these maps, we first looked for files of city maps for each of the three cities. GeoJSON files were acquired for all three cities at https://github.com/codeforamerica/click$_t hat_h ood/and we re convertedto TopoJSON fi$

5 Exploratory Data Analysis

As of now we are using rudimentary, simple visualizations to take a better look at our data. This includes looking at the overall data over different metrics such as distance, ride duration, the resulting average speed, and number of rides per day.

We plan to look at (rides$_i n - rides_o ut) to display something of a' flux' from each station. Hope fully depending on the hour and wee$

6 Design Evolution

Our original map visualization idea involved getting to-from data for each station and drawing thicker or thinner chords between station nodes that were more or less 'connected' by bike usage. After much data cleaning, however, we realized two things: the number of rides—even over the entire year—for just about any pair of stations was generally minimal, and that storing so many small values in different JSON objects or CSV rows would result is data files that were far too large.

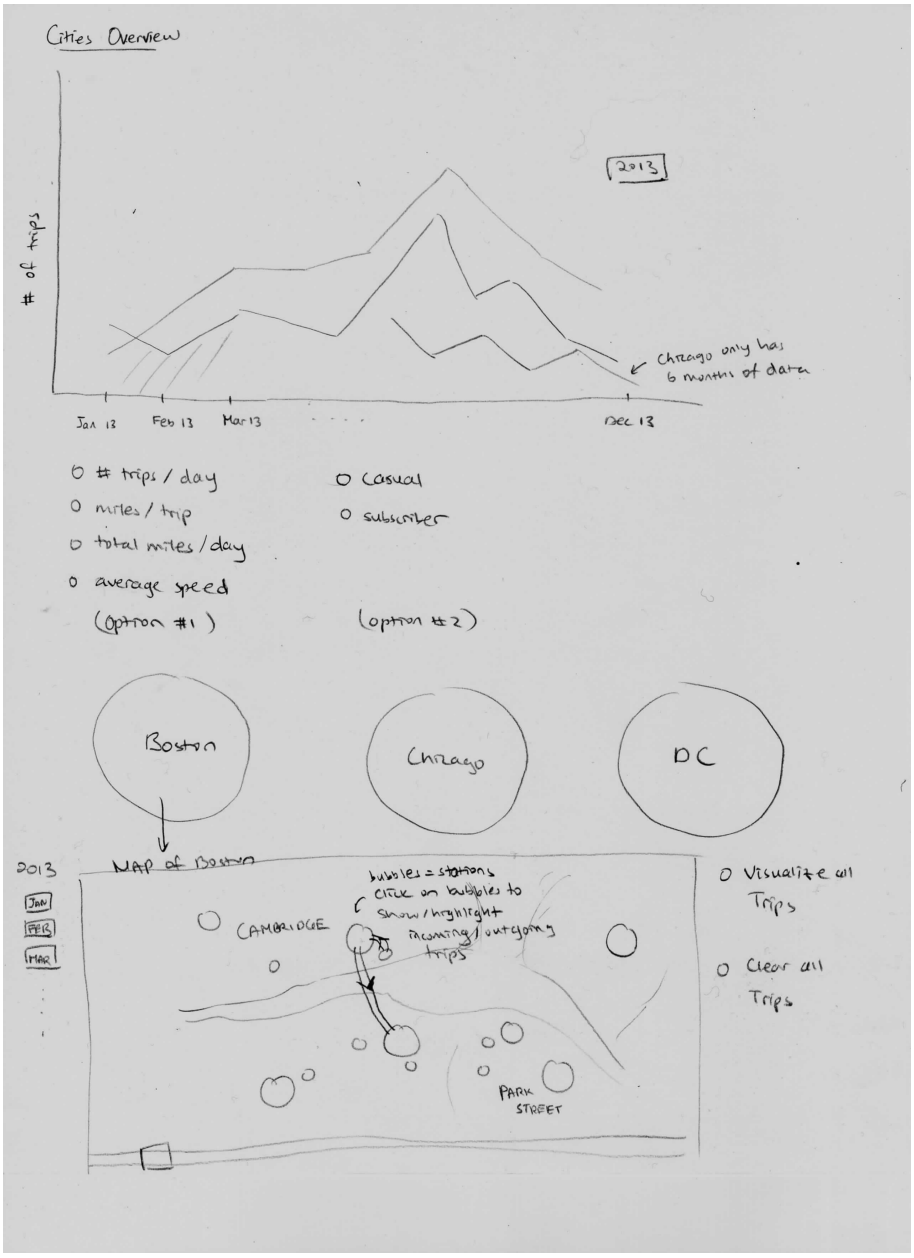
That idea eventually led to data aggregation which is discussed in the 'Data' Section above.

The line graph aspect of our visualization has been there since the beginning and to us seems like a very apt method of comparison between the cities.

6.1 Proposed Design

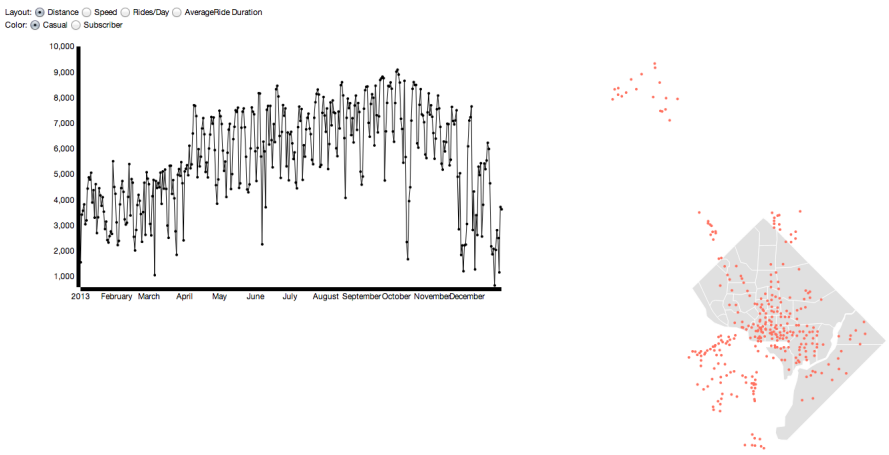
We initially proposed a split presentation with the overview graph on top and the map visualization on the bottom. The top and bottom graphs cannot be show on the screen at the same time. The overall data

visualization was to include interactions to display different features of the graph (two sets of features to choose from as shown in the image). The map visualization was to be an animation that displayed the ride patterns over the course of the year.



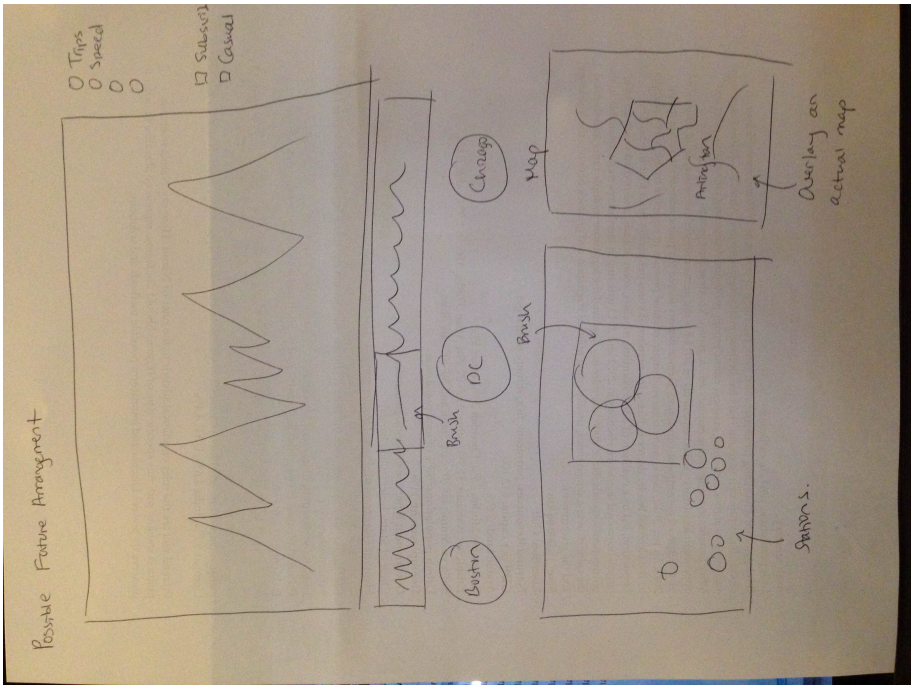
6.2 Work-in-Progress Design

Current layout



6.3 Next Step Design

We are thinking of tweaking the design layout again. We are also going to add a map layer as well as brushing.



7 Implementation

We will expand when our final implementation is more cemented.

8 Evaluation

We will expand when our final implementation is more cemented.

9 Fox's Middle Name

It starts with an "i"

So far eliminated: igloo