

DataHub

Dayne Davis
Niamh Durfee
Alexandra Smith

Background and Motivation

As students, we make use of many forms of alternative transportation to explore and commute around Boston. In particular, bikes are a popular, affordable method of transportation for students and professionals alike. We want to understand who uses bikes and how to get around Boston. One of the most readily available datasets that gives an idea of the prevalence of bikes is that provide by Hubway, a bike sharing/renting program in the Boston and Cambridge.

Project Objectives

We aim to answer the following questions:

- What is the average trip distance?
- What are the most popular start and end destinations in the Boston Area?
- Where are the people using hubway from and where do they go?
- Are trips are for leisure or strictly commuting (based on speed)?

Data

Our data is found at <http://hubwaydatachallenge.org/trip-history-data/>

Data Processing

There will be little data clean up necessary based on our projected uses. We may need to format it more based on GeoJSON requirements. We plan to derive:

- trip length (based on start and end time, and start and end position)
- trip speed (by dividing the quantities above)
- average user age
- gender breakdown
- most popular Hubway stations

This list is by no means exhaustive, and we anticipate more derived statistics as we get deeper in the data.

Visualization

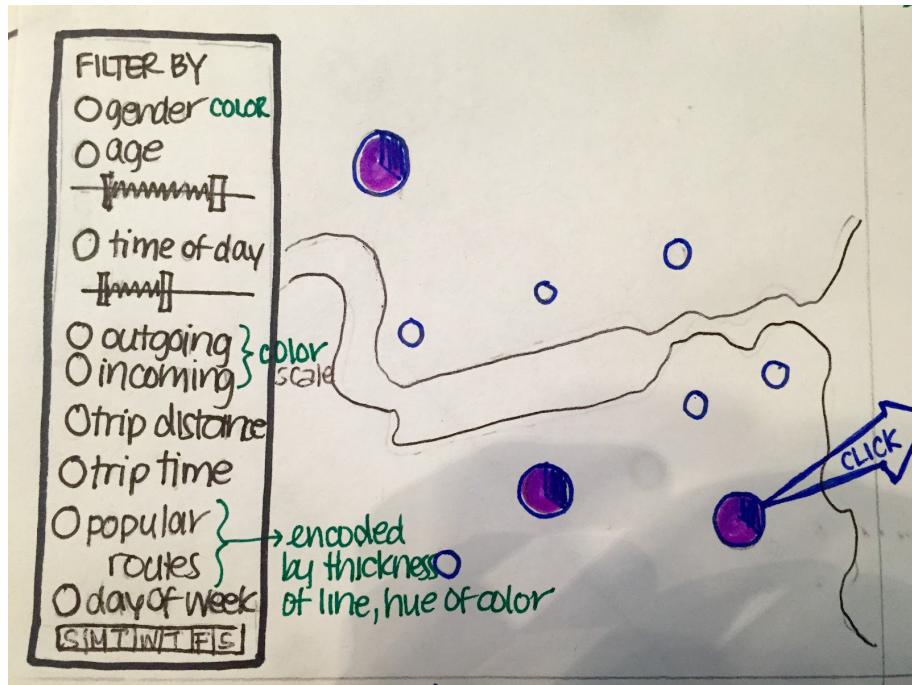


Figure 1: Map visualization by Hubway stations

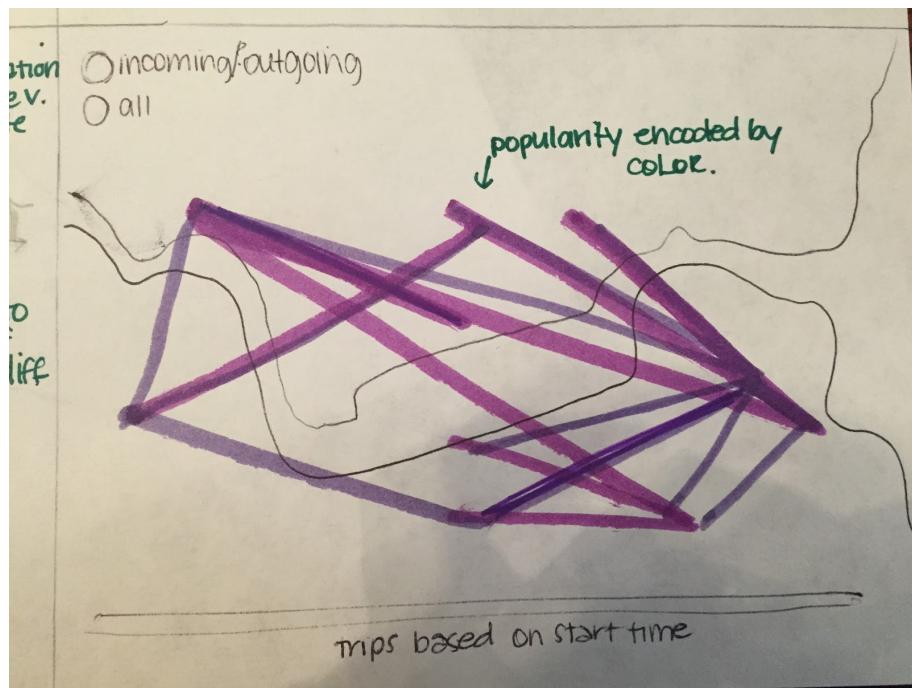


Figure 2: Map visualization of Hubway trips

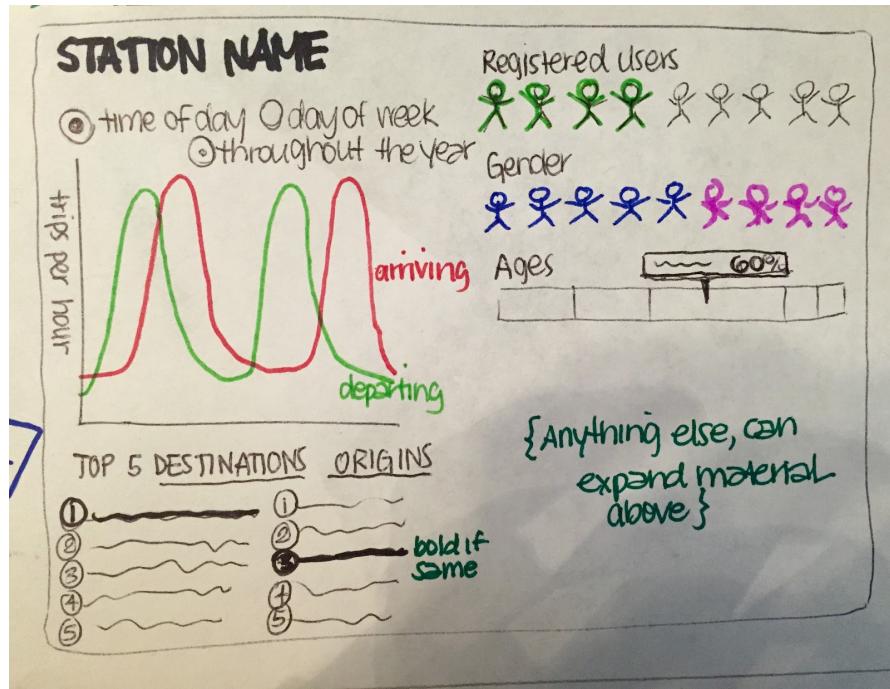


Figure 3: In-depth view of stations

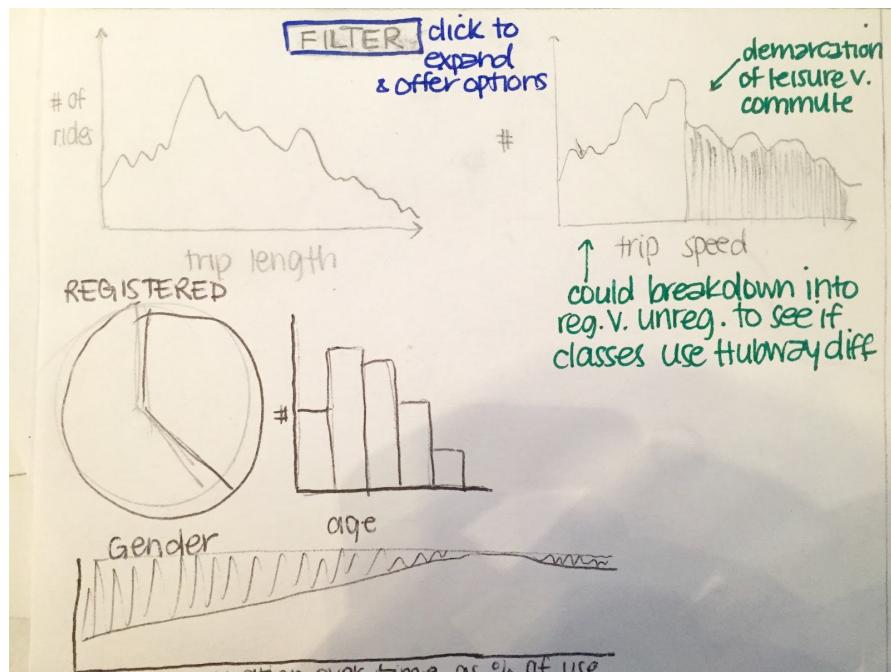


Figure 4: In-depth view of stations + users

Our main view will be a map of the Cambridge, Boston area. There will be two visualizations that you can switch between that are overlayed on the map. Figure 1 will be nodes at the station locations that will vary in size depending on how busy they are. You will be able to add filters for the time of day, gender, age, etc. that will modify the nodes. Figure 2 will be lines connecting the stations that will show the popularity of the route based on color. You will be able to apply the same filters to this visualization. Also hovering over a route can show other details of that route such as the average trip length, gender, and age of the rider. We also plan on having two more visualizations that will show other stats of the routes, stations, and the average user of Hubway (Figures 3 & 4). We plan on switching to these views by either clicking on a station or route. Each of these visualizations will also have some filters and a time slider to allow the user to see how they change over time.

Must-Have Features

- trip length distribution
- trip distance distribution (base on direct route between stations)
- trip speed distribution: leisure versus commuting
- popularity of a hub (total volume, outgoing, ingoing) - plotted on a map
- distribution of genders & ages of users
- distribution of time of day/year

Optional Features

- follow the bike (one bike for an extended period of time)
- see common routes
- plot use against weather (interesting to see extreme weather events)
- plot popular routes on a map

Project Schedule

	Dayne	Niamh	Lexi
Week 1 April 5-11	Overall: Data cleaned; Page setup; Map functionality with ability to plot coordinates		
	Setting up map visualization.	Import and clean up data.	Setting up page and ui. Set up website
Week 2 April 12-18	Overall: Both map visualizations done		
	Change nodes by popularity of station	Add lines to map viz and change color by popularity.	Add filters to map visualization.
Week 3 April 19-25	Overall: View of in-depth stations complete		
	Set up visualization.	Line graph and top Destinations	Registered users by stick figures, age stats, etc
Week 4 April 26-May 2	Overall: Last view complete, cleanup		
	Line graphs, Demo video	Pie graph, bar graphs, Gender graph.	Tie all visualizations together, add transitions, and clean up style.