

# CS171 Final Project Process Book

Team Blue Riders

## **Team Leader:**

Shiyun Qiu (HUID: 21304484) shiyunqiu@g.harvard.edu (424) 666-5878

## **Team Members:**

Wei-Te Mark Ting (HUID: 30904516) weiteting@college.harvard.edu (617) 575-9565

Xiangru Shu (HUID: 51314832) xiangru\_shu@g.harvard.edu (919) 917-5830

Gloria Chang (HUID: 11240348) gchang@gsd.harvard.edu 713 997-9495

## **Project Link:**

<https://github.com/mat3049/cs171-2018-final>

# Table of Contents

<b>Table of Contents</b>	<b>2</b>
<b>Changelog</b>	<b>4</b>
<b>Week 1: Dataset/Argument</b>	<b>5</b>
<b>Week 2: Team Abstract</b>	<b>6</b>
<b>Week 3: Final Project Plan</b>	<b>8</b>
Topic: The Life of a Blue Bike	8
Goals	8
Datasets	10
Blue Bikes Data	10
Carbon/Fuel Costs	10
Travel Time Comparison	10
Team Roles	12
Sketches of Visualization Ideas	13
Interaction Storyboard Sketch	15
Sketch of Webpage Layout/Storytelling	16
Team Expectations Agreement	17
<b>Week 4</b>	<b>18</b>
Meeting with TF	18
Questions from other team (Boston Crime Has Not Departed)	18
Prototype V1	18
Implementation of Visualizations	19
<b>Week 5</b>	<b>26</b>
Prototype V2	26
<b>Week 6</b>	<b>31</b>
Final Website	31
Screencast	31
Github Repository	31
Final Updates	31
Data Sourcing	31
Visualizations	32
Instructions for Interactions	32



# Changelog

2018-12-02

- Final deliverables completed
- Added links to the final website, screencast video, and Github Repository
- Logged design evolutions happened in the last week
- Added overall evaluation of the project

2018-11-25

- Prototype V2 finished
- Created database for trip data
- Logged the problem of multi-station interactions
- Added screenshots of Prototype V2

2018-11-17

- Prototype V1 finished
- Added screenshots of implemented visualizations

2018-11-10

- Milestone 2 finished
- Added visualization sketches
- Added team expectations agreement

2018-11-08

- Discussed topic questions
- Presented visualization ideas

2018-11-02

- Milestone 1 finished
- Decided on Hubway dataset
- Finalized group objectives
- Added team abstract and motivations

2018-10-29

- Created process book

# Week 1: Dataset/Argument

## Public Datasets

<https://github.com/awesomedata/awesome-public-datasets>

## PILAC Database

Collaboration with HLS' PILAC (Program on International Law and Armed Conflict) for ongoing research on the use of Article 51 on the UN Charter

- Article 51. “Nothing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security.
- Penultimate Database (link removed)
- This dataset is the result of ongoing research and will be finalized within a few weeks (prior to final submission). It contains data on complaints filed from 1945 to the present day, which countries submitted the complaint, and which are involved, as well as additional information on the type of complaint, nature of the state, and the implicated parties.
- The dataset has been assembled through the Official Document System of the United Nations.

## Plane Crash Database, since 1920

<http://www.planecrashinfo.com/database.htm>

Notes:

- Difficult to scrape
- Data inconsistencies

## Hubway Trip History Data

<http://hubwaydatachallenge.org/trip-history-data/>

<https://s3.amazonaws.com/hubway-data/index.html>

## Medical Research Payment Data

<https://openpaymentsdata.cms.gov/browse>

# Week 2: Team Abstract

**Topic Consensus:** Blue Bikes trip history data

**Project Title:** Pedaling Our Way to a Better Environment

**Existing Site:** <http://bostonography.com/hubwaymap/>

## **Motivation:**

To promote increased utilization of public bike services within Boston for improved environmental sustainability and economic efficiency, we look at which style of usage and demographics of users have increased usage of bikes throughout 2011-2018. This information should help us target groups for higher bike use, as well as encourage them to use bikes for specific purposes.

## **Dataset Description:**

Blue Bikes publishes downloadable files of blue bikes trip data every quarter. The datasets can be found on <https://s3.amazonaws.com/hubway-data/index.html>. Columns available in the datasets are trip duration, start/end date and time, start/end station, user type, user gender, etc. We also have access to the longitude and latitude of all the blue bikes stations, and the number of docks in each station.

<http://hubwaydatachallenge.org/trip-history-data/>

<https://s3.amazonaws.com/hubway-data/index.html>

**Questions:**

- What's the average trip duration for annual members vs. casual users?
- What are the peak Blue Bikes hours?
- Which days of the week get the most Blue Bikes traffic?
- Which stations are most popular? Which stations make up the most popular origin/destination pairs?
- Which stations are the most asymmetric - more trips start there than end there, or vice versa? Are they all at the top of hills?
- What does a year in the life of a Blue Bikes look like?
- Which stations get the most tourist traffic, and which get the most commuters?
- How far does Blue Bikes really reach? Which community should be the next to get Blue Bikes stations?
- Are all of the Blue Bikes rentals at 2:00am by people under 25?
- Are there different top stations for male vs. female Blue Bikes members?
- Which demographics are more likely to ride bikes? Which ones are rapidly increasing?

# Week 3: Final Project Plan

## **Topic: The Life of a Blue Bike**

- Audience: potential/current blue bikes users, people using public transportations, environmentalists, commuters, tourists
- Client: City of Boston

## Goals

Blue Bikes is a bike sharing system in Boston metropolitan area. It now has around 2200 bikes at more than 210 stations throughout Boston, Brookline, Cambridge and Somerville. Blue bikes now has become a popular public transportation in Boston metropolitan area, as it is convenient, fast and also affordable.

The project is aim to provide detailed statistics about Blue Bikes usage in Boston metropolitan area through interactive visualizations, compare bike riding with car driving from environmental perspective and ultimately promote increased utilization of public bike services within Boston for improved environmental sustainability and economic efficiency.

## Tasks

- Data collection & cleaning
  - Blue Bikes Data: combine monthly data into annual data
  - Collect data on carbon/fuel costs
  - Query Google Map API for estimated travel time data
- Background Research
  - History of Blue Bikes: When and how it was founded
  - What is the make of Blue Bikes
  - Carbon/fuel costs of bikes
  - End of the Blue Bike (disuse, disassembly, reuse)
- Storyline
  - Identify important messages to be told
  - Construct narratives for all pages
- Design
  - Web page layout
  - Interactive visualizations
- D3 Implementation
  - Implement visualizations in D3
- Webpage
  - Storytelling(text, images, videos, etc.)
  - Styling(color theme, background image, font, spacing, etc)
- Video
  - Summary of our project, and demo of how to use our website

# Datasets

## Blue Bikes Data

- Blue Bikes publishes downloadable files of blue bikes trip data every quarter. The datasets can be found on <https://s3.amazonaws.com/hubway-data/index.html>.
- Columns available: trip duration, start/end date and time, start/end station, user type, user gender, longitude & latitude of all the blue bikes stations, and the number of docks in each station
- Time frame: 2011-2018
- Frequency: Per trip

## Carbon/Fuel Costs

- We can get an estimate of carbon/fuel costs per mile and calculate the total costs for each trip.

## Travel Time Comparison

- Google Map API: <https://developers.google.com/maps/documentation/directions/intro>
- We can query all possible routes between stations in advance and store the data.

## Possible Visualizations:

- (Where does a Blue Bike go?) Popular stations and routes on map
- (How many users does a Blue Bike have? / How long does a Blue Bike last?) Number of rentals, trips
- (How fast is a Blue Bike?) Temporal comparison of travel times for vehicles vs bikes
- (How does a Blue Bike save the environment?) Compare to carbon cost, fuel cost of vehicle transportation

## Storyline Items:

- Making of a Blue Bike (basic sourcing)
- Possible interviews with the makers or users of Blue Bikes
- End of the Blue Bike (disuse, disassembly, reuse?)

**Must-Have Feature List:**

- Map showing popular stations and trips and filtered by time
- Travel time comparison between bike and car
- Chord diagram showing ride flows
- Diagrams showing aggregated distance of Blue Bikes

**Good-To-Have:**

- Videos about Blue Bikes
- Making of a Blue Bike
- How Blue Bikes are recycled
- Carbon/fuel cost calculation

**Optional Feature List:**

- Self-narrative of a Blue Bike

**Project Schedule:**

- November 18: Prototype V1 Due
  - Finish core visualizations: interactive map, chord diagram, etc.
- November 25: Prototype V2 Due
  - Basic web page structure, storyline complete
  - Integrate the visualizations into the website
- December 2: Final Version Due
  - Final Version
  - Evaluation
  - Video

## Team Roles

Technical Design: Mark, Xiangru, Shiyun

*Primary design and implementation of technical aspects of the website, including visualizations and other interactives.*

Screencast Production: Mark, Gloria

*Screenwriting, recording, and editing of screencast for final presentation.*

Information Coordination: Gloria

*Reach out to Blue Bikes for more information as needed. Proof text and find story relevant information, images, and videos.*

Story/Graphic Design: Gloria

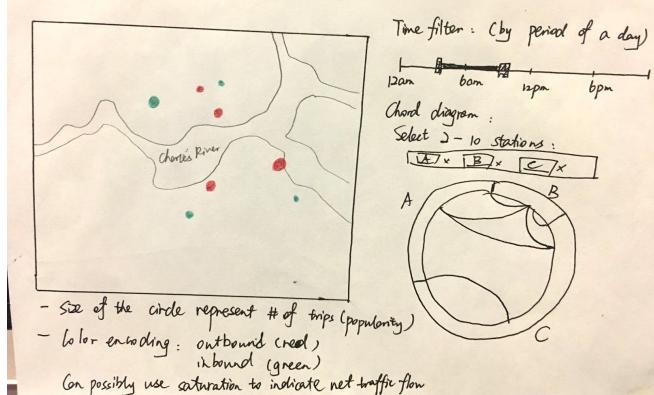
*Establish and ensure consistent use of designated style guidelines, including color schemes and general tone.*

Data Analysis: Shiyun, Xiangru

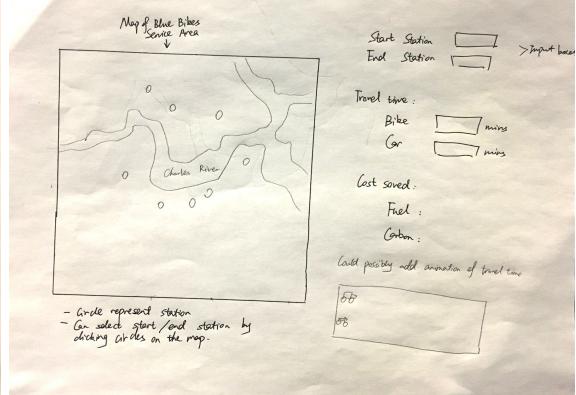
*Collect data and clean up the data.*

# Sketches of Visualization Ideas

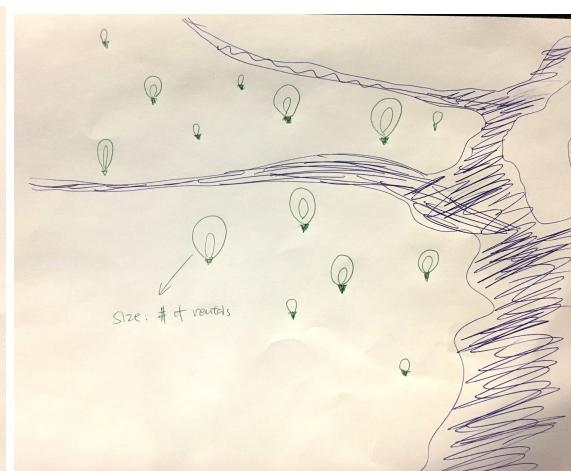
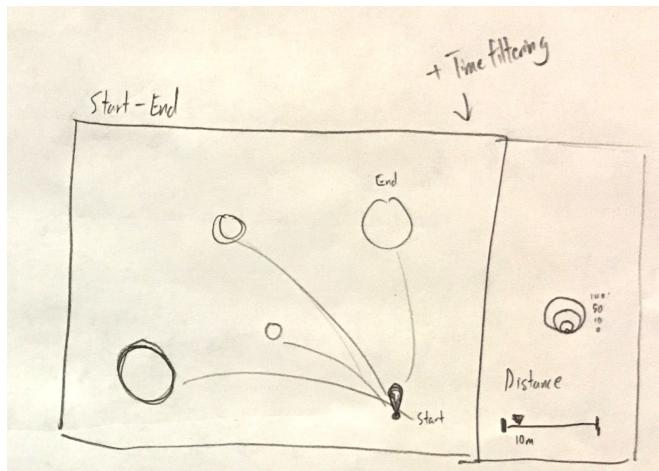
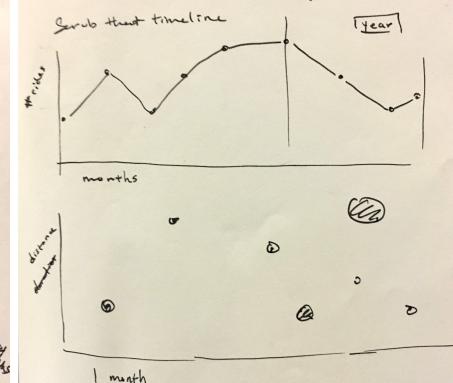
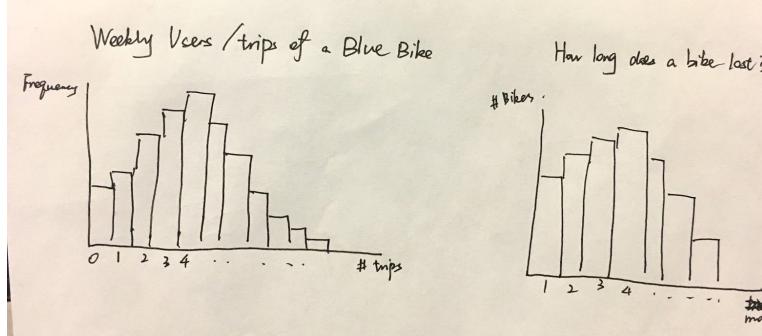
Interactive map: (Popularity of station & Interactions between stations)

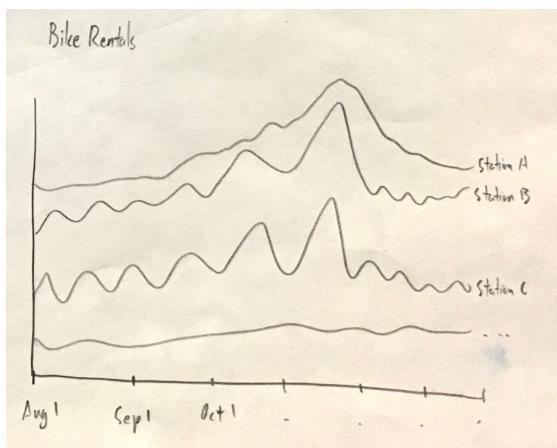
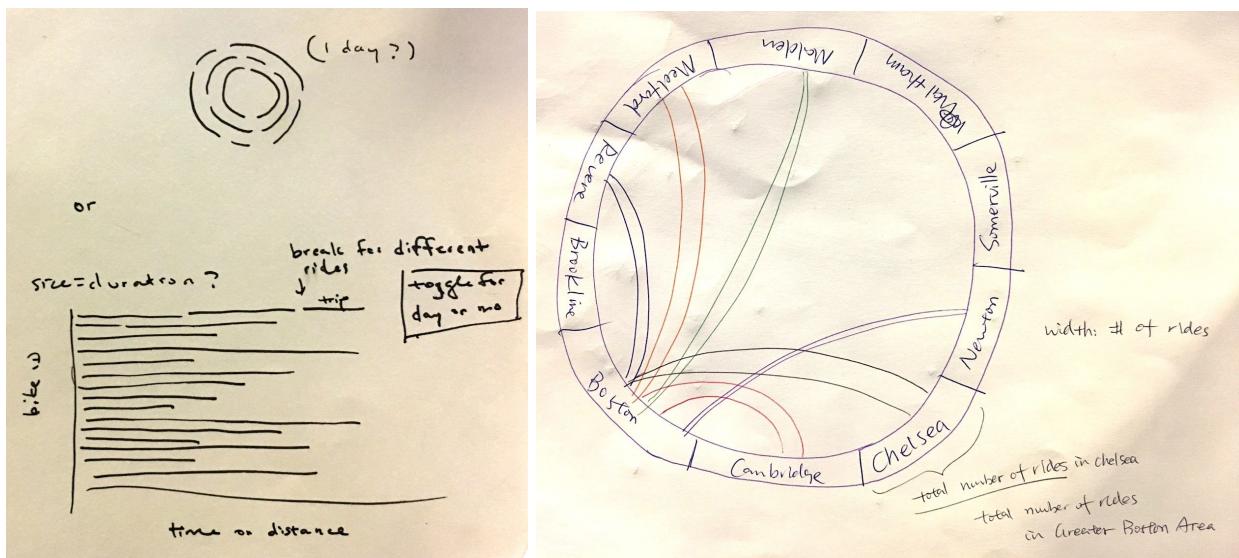
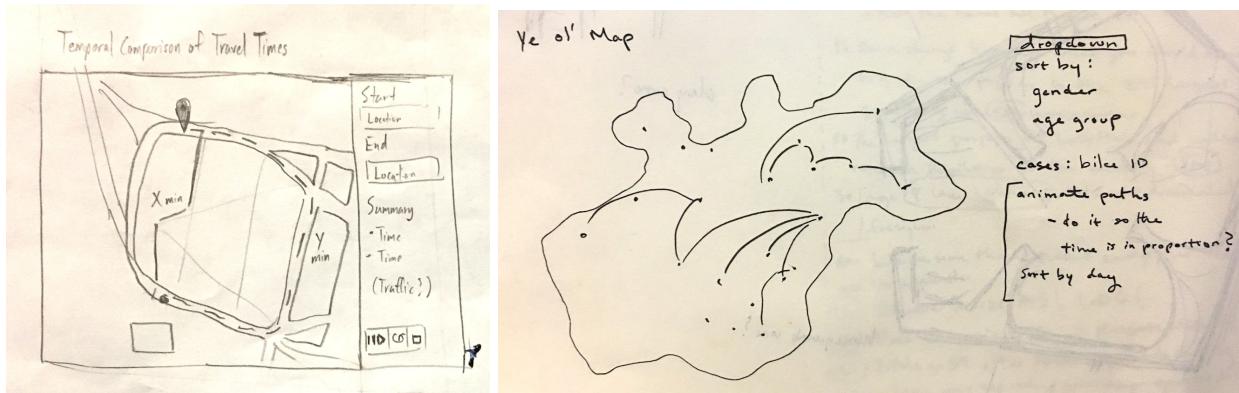


Interactive map: (Travel time Comparison & Carbon/Fuel Cost.)

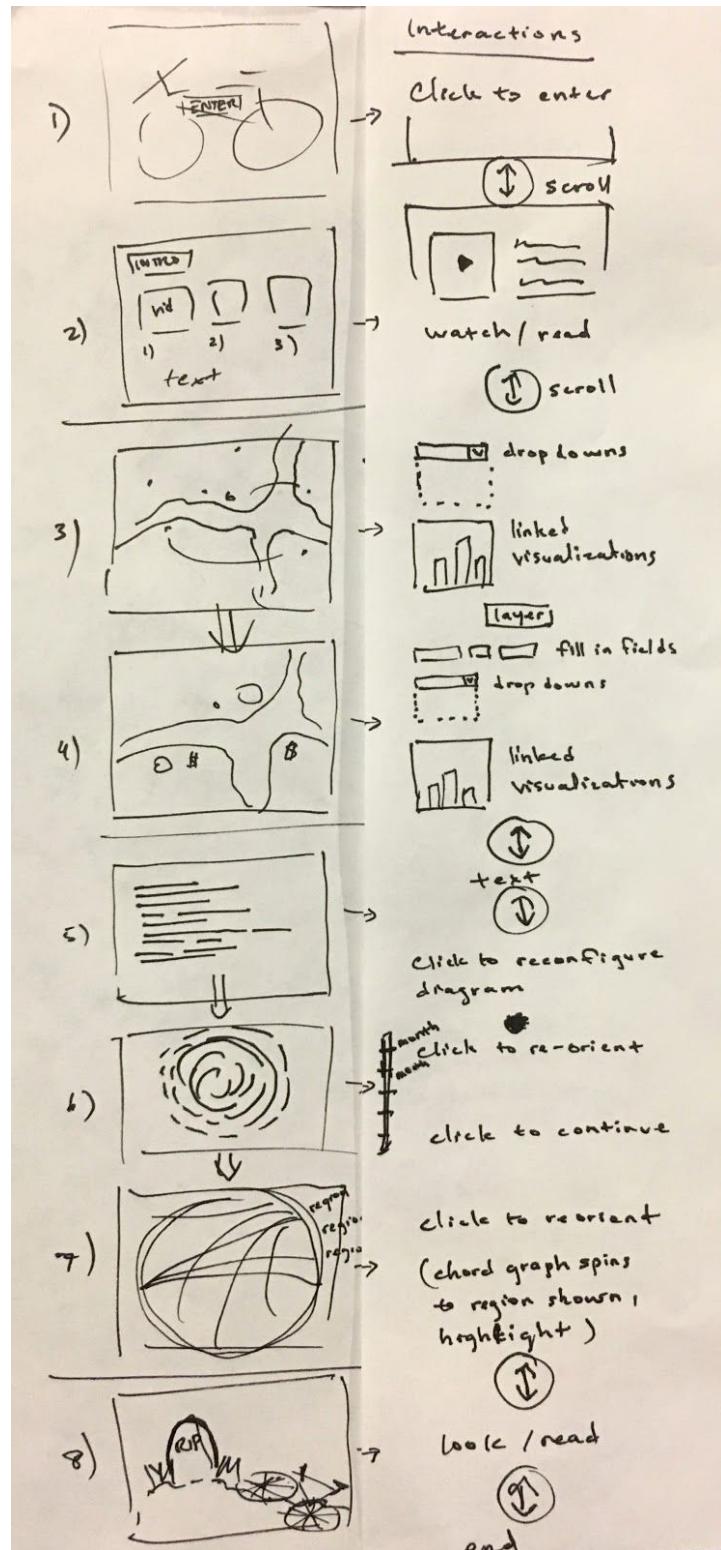


Static plot: Histogram

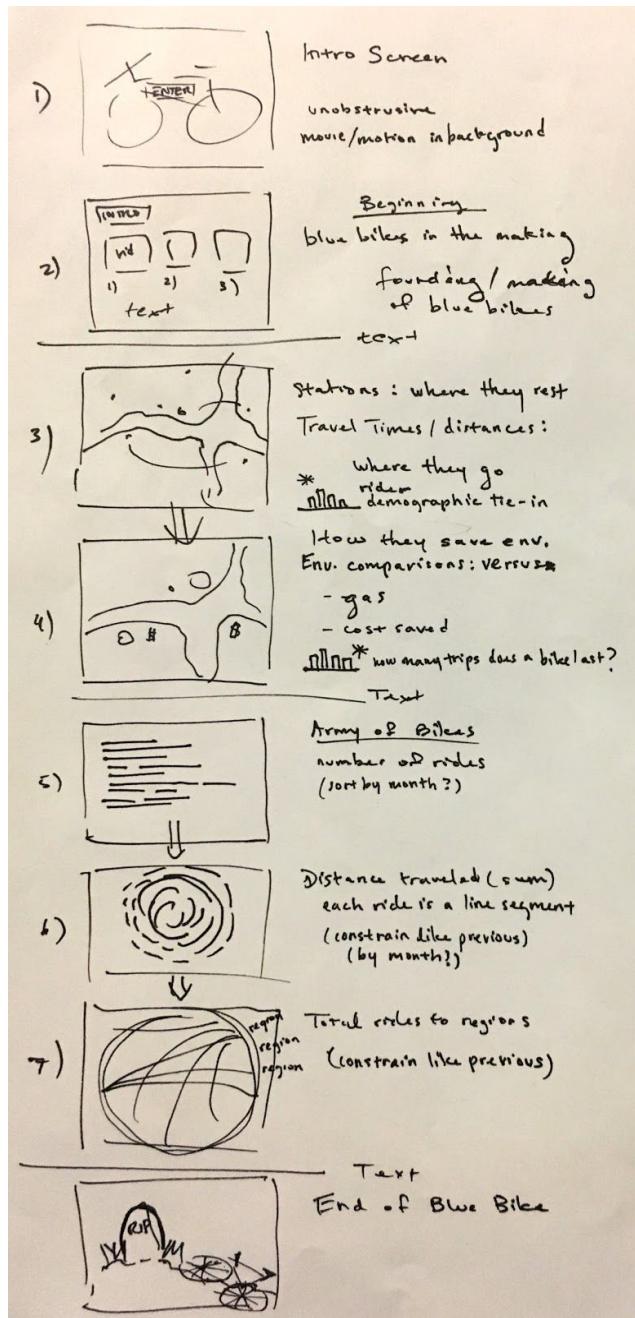




## Interaction Storyboard Sketch



# Sketch of Webpage Layout/Storytelling



- (1) Title/intro page.
- (2) Beginning of the Blue Bikes, how it was founded, how they are made.
- (3-4) Show where the bikes rest, where they go, and how they save the environment - all on a map.
- (5-7) Show what the army of bikes can do - how many rides, what the aggregate distance looks like, and how much they service each region.
- (8) Tell the ending of the Blue Bike - how one retires or is put to rest.

# Team Expectations Agreement

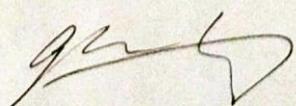
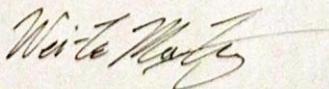
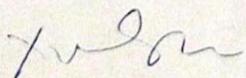
## Team Expectations Agreement

**Team Name:**  
Blue Riders

**Team Members:**  
Shiyun Qiu  
Wei-Te Mark Ting  
Xiangru Shu  
Gloria Chang

### Rules and Expectations:

- Arrive at the weekly group meetings on time
- Prepare for the group meetings and present work at the meetings
- Communicate frankly but with respect when conflicts arise
- Read through email communications and respond in time



# Week 4

## Meeting with TF

Fundamental ideas are all set

- Consider ways to aggregate/filter data
- Consider ways to bundle granular data (not necessarily consolidate, but group)
- Check-in again after Thanksgiving

## Questions from other team (Boston Crime Has Not Departed)

- What are you unsure about? (answer: size of data set might be too big)
- Good storyline, might compete with main map
- Make sure the story aligns / stays focused

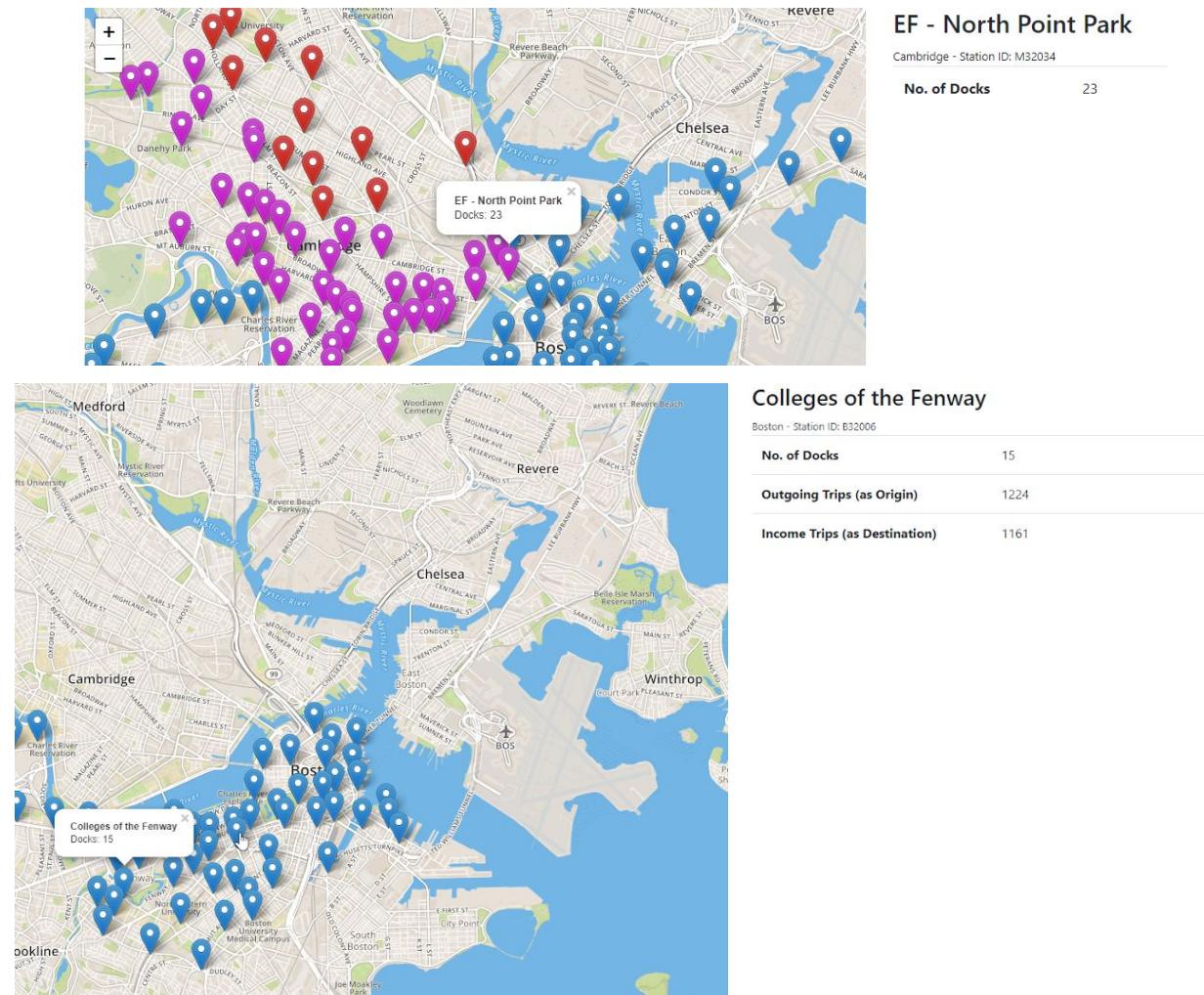
## Prototype V1

### Data Scraping and Cleaning

We've downloaded data from <https://s3.amazonaws.com/hubway-data/index.html>, and combined it into annual data. We have matched station information with the trip information. For this prototype, we only used one year of data. The dataset we used for this prototype is in the data folder of our submitted zip file.

# Implementation of Visualizations

## Visualization 1: Interactive Map

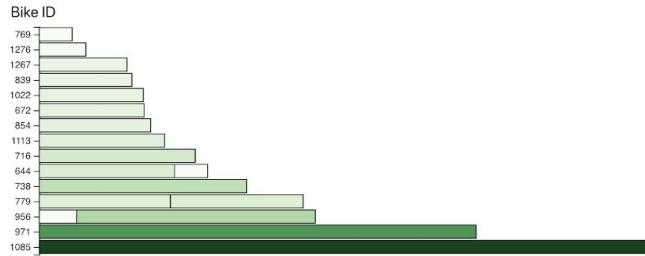


**Description:** Interactive map is zoomable and pannable. User can click on station nodes to get more detailed information on the particular station. Stations from different “regions” are colored differently. Clicking on each station will show information in the sidebar such as the number of inbound/outbound trips. We will be adding time and distance sliders as well as other options to the sidebar to allow detailed filtering of the stations. There will also be additional interactions between stations (lines, highlighting, etc.)

## Visualization 2: Interactive Bar Chart

### Trip Duration and Distance

Sort



Weekly  data:

Duration  of Blue Bikes starting from station

359 Broadway - Broadway at Fayette Street

in

Week

### Trip Duration and Distance

Sort



Weekly  data:

Duration  of Blue Bikes starting from station

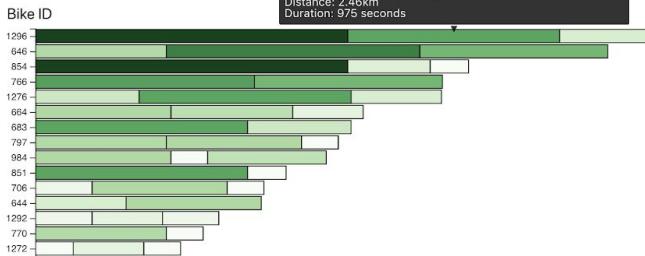
359 Broadway - Broadway at Fayette Street

in

Week

### Trip Duration and Distance

Sort



Monthly  data:

Distance  of Blue Bikes starting from station

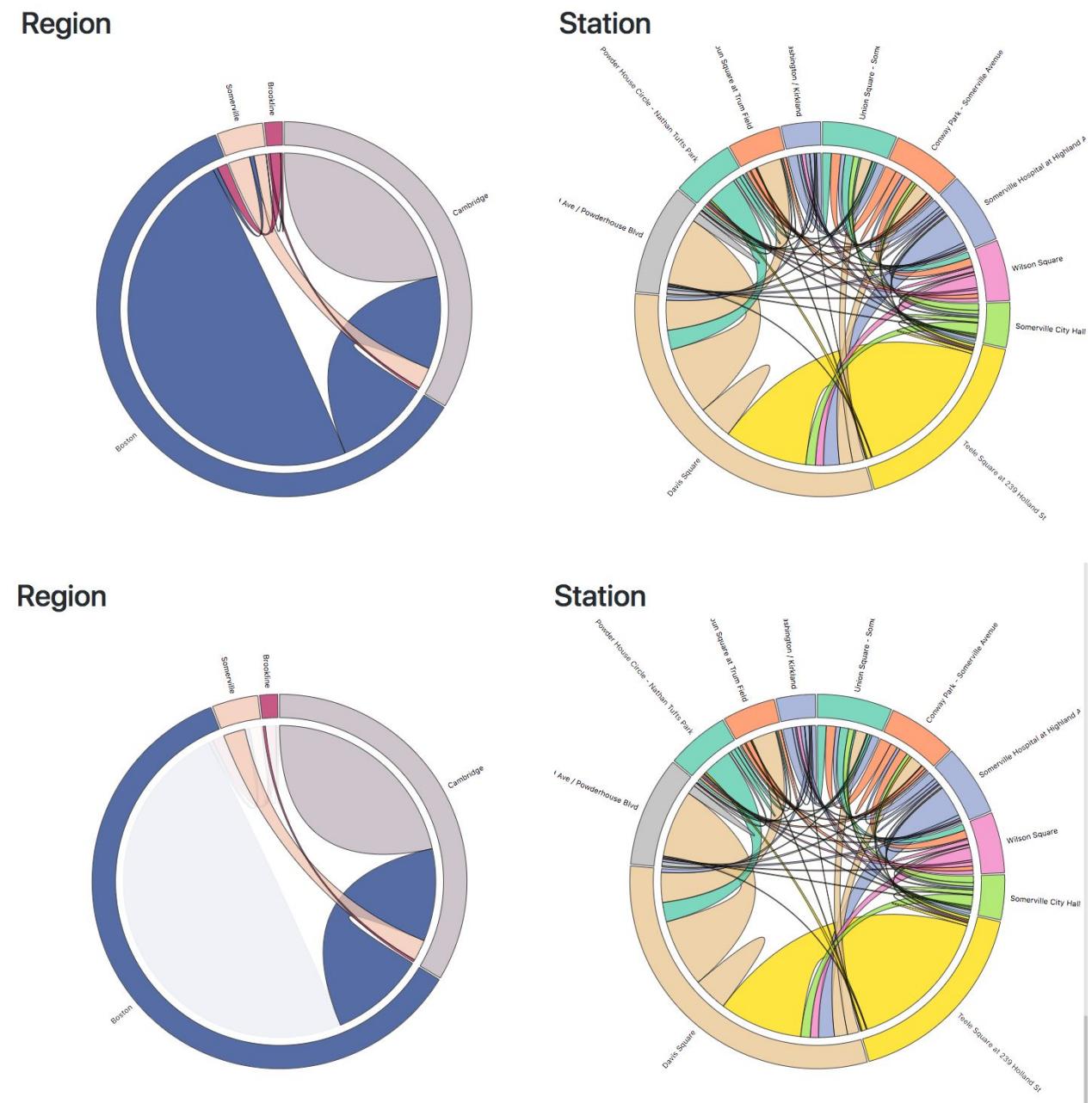
Binney St / Sixth St

in

January

**Description:** The user can choose the aggregation time period (monthly/weekly) they want to visualize. The default mode is weekly data, with a drop down box containing values 1-52, corresponding to each week in a year. If monthly is chosen, a drop down box with selection options January-December will appear. A user can also choose whether he/she wants to visualize the distance or duration of each blue bike trip. To better visualize the data on the screen, a user has to select the start station for bike trips. We will display the top 15 bikes with longest trip duration/distance given user selections. In addition, d3 tool-tip is implemented. When a user hovers the mouse over a bar, detailed trip information will pop up.

### Visualization 3: Chord Diagram



#### Description:

The user can click each region in the first chord diagram, and the second chord diagram will show the stations in that region. A time filter or a drop down box will be added for user to select specific time period. When the user hover on one of the arcs in both chord diagrams, the diagrams will only show the flows to that arc and the other flows will become transparent.

## Rough Website Design / Storytelling

(find this layout in [index 2.html](#))



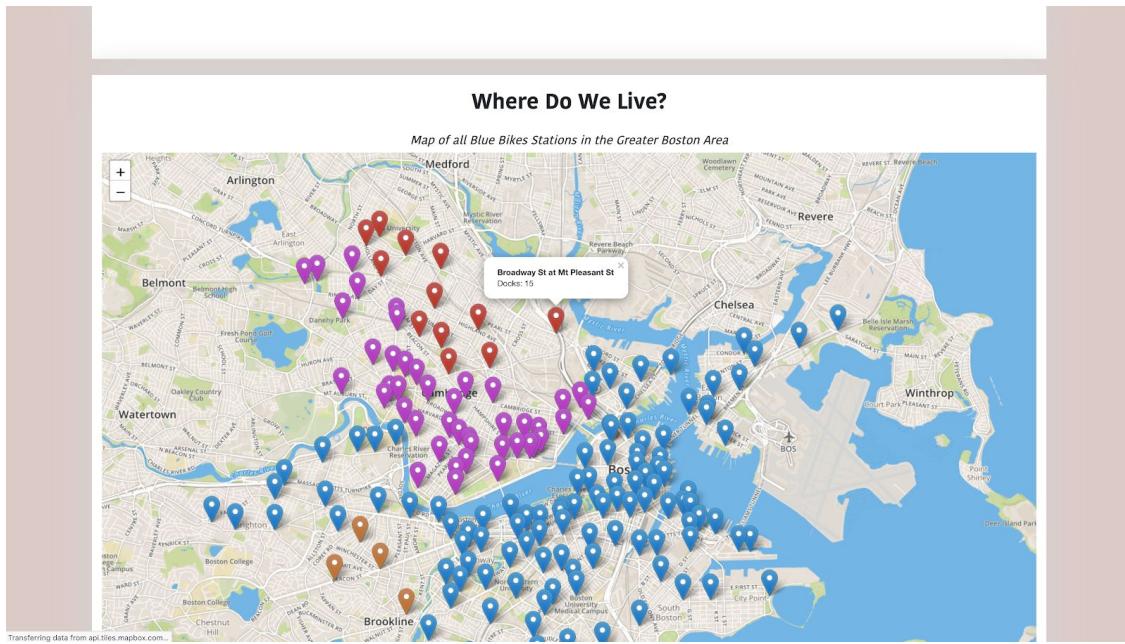
### *Page 1: Intro Screen*

This page will contain our project title. The background will be a moving gif related to Blue Bikes. Video in screenshot is in temporary resolution, video in file is from external site.



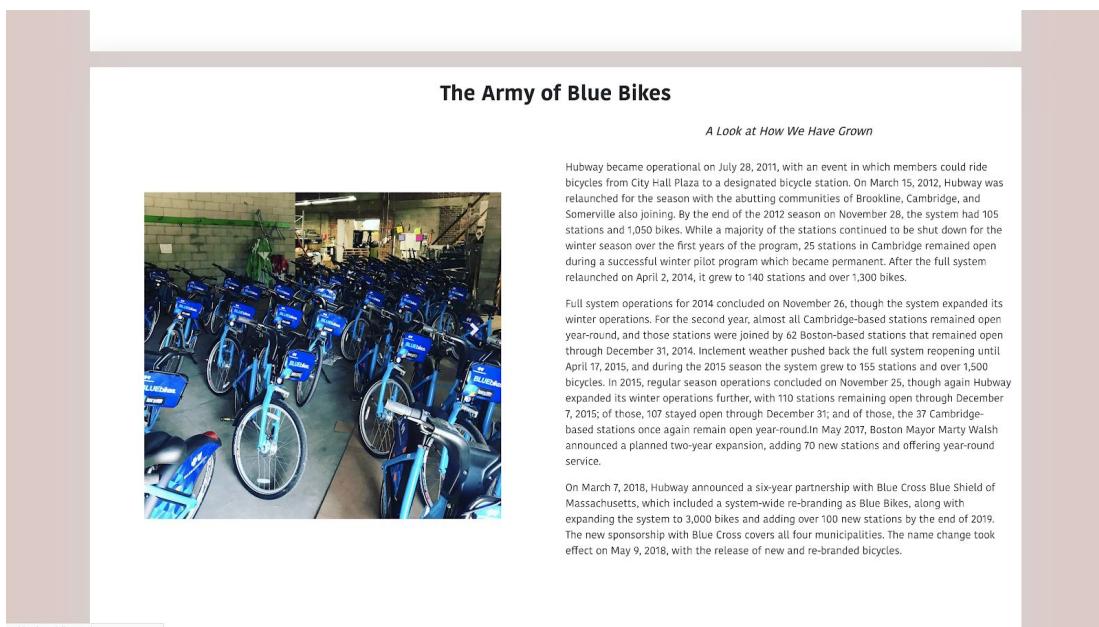
### *Page 2: Birth of Blue Bikes*

We will present background information of Blue Bikes. There will be both textual content and videos introducing Blue Bikes, including information about how the Blue Bikes Program was made possible. Diagram is in progress.



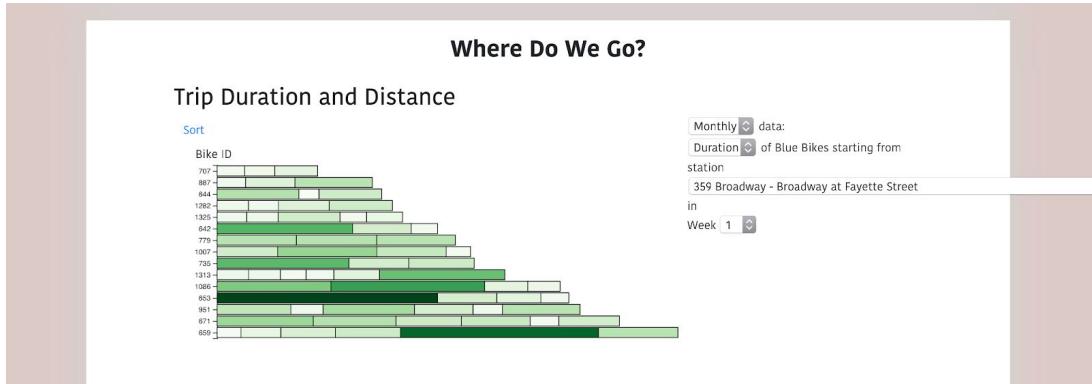
### *Page 3: Interactive Map*

We will present blue bike stations on the map, showing where a blue bike rests in its life. On this page, we will also tell the story of where a blue bike goes (travel times/distance), how it saves energy (carbon/fuel cost) and whom it serves (rider demographic information).



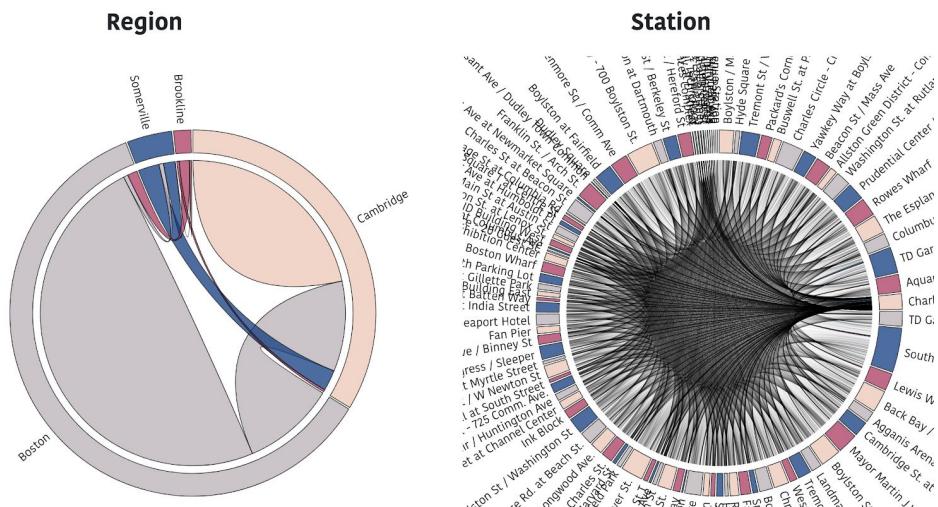
### *Page 4: Army of Bikes*

A look at the dedication and expansion of the Blue Bike system. Yet another step in the life of a Blue Bike, between birth and death: proliferation.



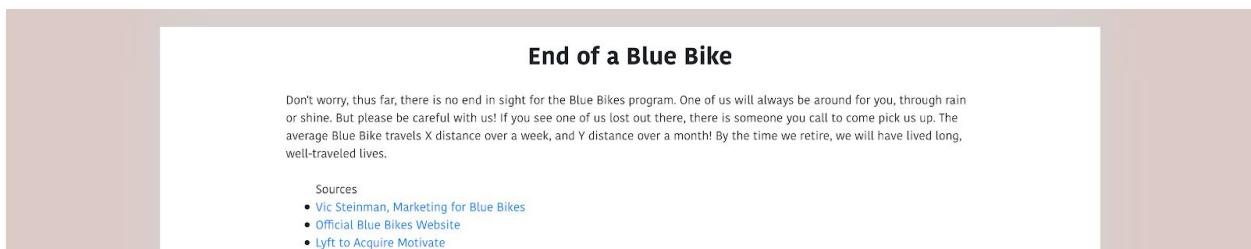
### Page 5: A Week (or Month) in the Life of a Blue Bike

This page will focus on the trip duration/distance of a blue bike. Users can choose which blue bikes they want to analyze through multiple filters including time, starting station, etc. The bar charts will show detailed information about all trips given the filters.



### Page 5: Trip Flow Chord Diagrams

There will be two chord diagrams showing the total number of rides and traffic flow directions between different regions and stations. A time filter will also be integrated on this page. We will provide detailed information about how the total number of rides differs between regions and stations and show the busiest region and station based on different time period.



*Page 6: End of Blue Bike*

The ending page will focus on the summary of a blue bike's life. For example, we will provide statistics about trip duration and distance (total/average), and analytics about riders demographic information. In addition, we may interview some blue bike users and share their stories about how blue bikes affect their lives. If the information is available, we may also present how a blue bike is decomposed and recycled.

# Week 5

## Prototype V2

### Data Sourcing

During our implementation, we found that loading the entirety of the dataset (2011-2018) was not feasible due to the sheer amount of data, so we focused on the 2015 data. We specifically chose this year as prior years (2011-2014) had different column names and referred to older stations.

To further reduce dependency on bundled data (and overhead), we then uploaded the ride and station data into MySQL and implemented our own API for stations and trips in 2015. The various API endpoints are written in PHP and can be found here:

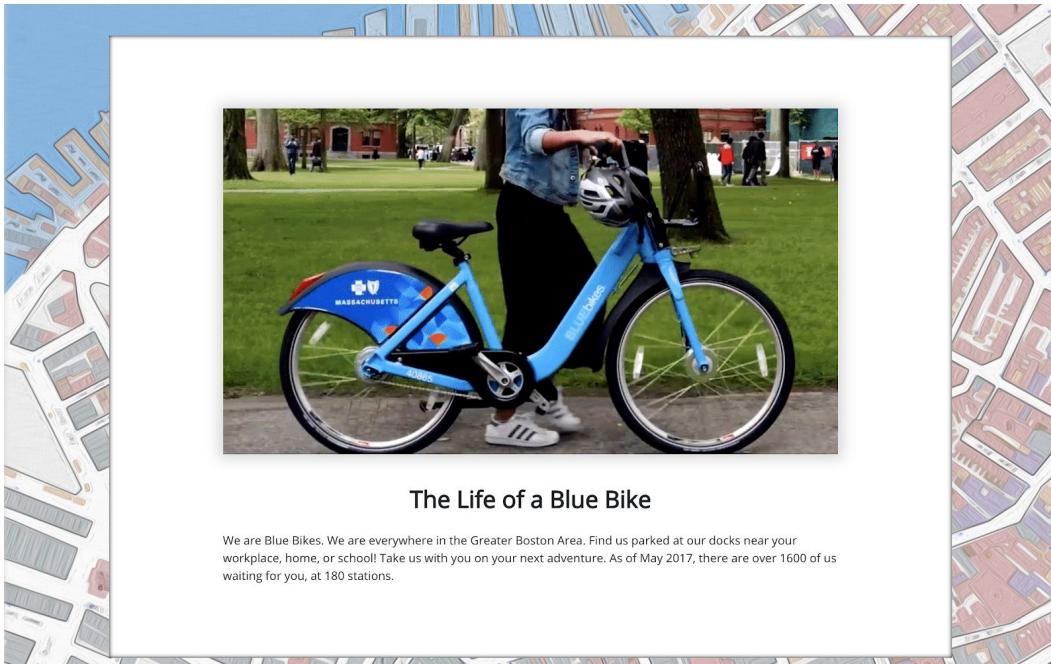
<http://vps77598.vps.ovh.ca/bluebikes/>

To facilitate usage, a data management layer in the form of the [DataService](#) class was written and integrated into the StationMap.

### Problems to be Solved

Apart from the map, the other visualizations rely on the full dataset, so we have not yet migrated those ones to API calls. We first explored using Leaflet markers for station interaction, but found it insufficient for our needs. D3.js was much more versatile, and so we integrated D3 into the marker layer within a Leaflet map. However, due to continuing difficulty with multi-station interactions, we are currently exploring additional interaction methods and solutions.

## Current Version of Website Design and Storytelling



### Page 1: Intro Screen

This page contains our project title. The first image is a moving gif related to Blue Bikes.



### Page 2: Birth of Blue Bikes

We present background information on Blue Bikes. The diagram shows information about how the Blue Bikes Program was made possible. Each icon links to the official website.

## Where Do We Live?

We stay docked at stations around Boston, Brookline, Cambridge and Somerville. Take us for a quick one-way trip or a leisurely ride around town. Commute to work or school, run errands and explore the city. At the end of your journey, return your Blue Bike to any station. Slide the us firmly into an empty dock and wait for the green light to make sure the lock is fully engaged. See the map below to find out where we are!

No. of Docks 19

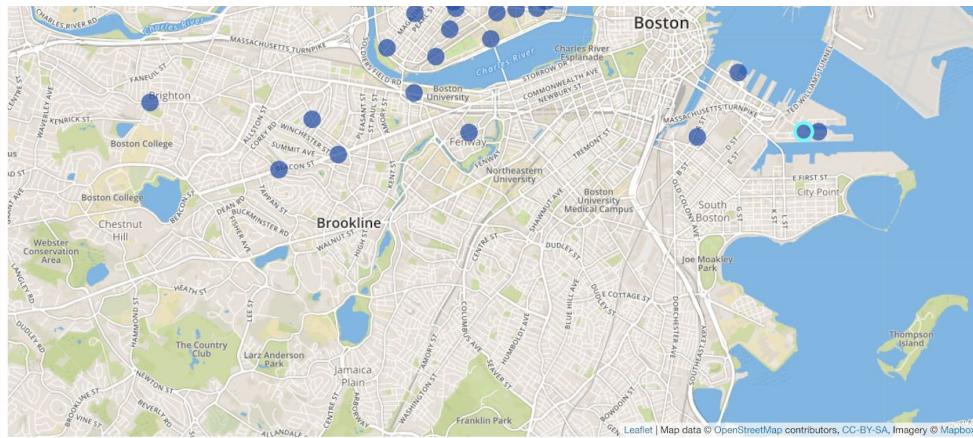
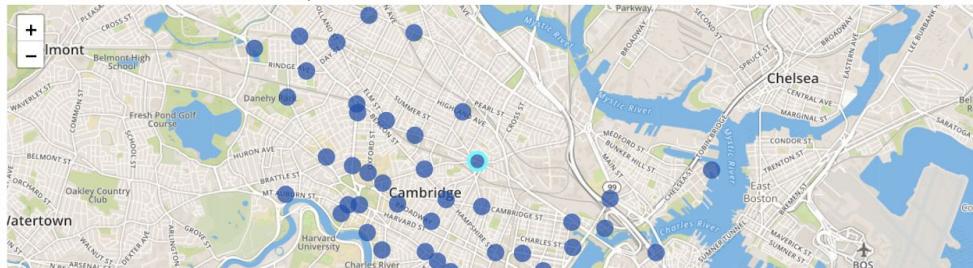
Outgoing Trips (as Origin) -

Incoming Trips (as Destination) -

Union Square - Somerville

Somerville (ID: S32002)

Map of all Blue Bikes Stations in the Greater Boston Area



## Who Are Our Riders?

While we are here for everyone, some people like to take us out more. See how this varies by area on the map above. Change the drop down options below to see who our most frequent riders are, by gender or age (upcoming).

Rider Demographics

Female 79

Male 379

Unknown 0

## Page 3: Interactive Map

We present blue bike stations on the map, showing where a blue bike rests in its life. On this page, we tell the story of where a blue bike resides (station information) and whom it serves (rider demographic information). There will be more information of outbound and inbound flow after we solved the problem of multi-station interactions.



## The Army of Blue Bikes

### *Our Growth and Identity*



Hubway became operational on July 28, 2011, launching in the communities of Brookline, Cambridge, and Somerville on March 15, 2012. By the end of the 2012 season on November 28, the system had 105 stations and 1,050 bikes. After the full system relaunched on April 2, 2014, it grew to 140 stations and over 1,300 bikes.

Inclement weather pushed back another full system reopening in 2015 until April 17, and during the 2015 season, the system grew to 155 stations and over 1,500 bicycles. Hubway expanded its winter operations further, with 110 stations remaining open through December 7, 2015. In May 2017, Boston Mayor Marty Walsh announced a planned two-year expansion, adding 70 new stations and offering year-round service.

On March 7, 2018, Hubway announced a six-year partnership with Blue Cross Blue Shield of Massachusetts, which included a system-wide re-branding as Blue Bikes, along with expanding the system to 3,000 bikes and adding over 100 new stations by the end of 2019. The new sponsorship with Blue Cross covers all four municipalities. The name change took effect on May 9, 2018, with the release of new and re-branded bicycles.



## *Page 4: Army of Bikes*

A look at the dedication and expansion of the Blue Bike system, leading up to the aggregated effects of the bikes.



## How Do We Help Save the Environment?

Every week is different for us. Our riders sometimes take us for short rides, or long adventures. Some months are busier than others; check out your favorite holidays, or see if anyone is riding in the cold. Not only do we help people of all ages and genders get around, we can also help save fuel and time. Investigate how many trips we take, how long our riders keep us, and how that reduces carbon emission and fuel use by changing the menus on the right.

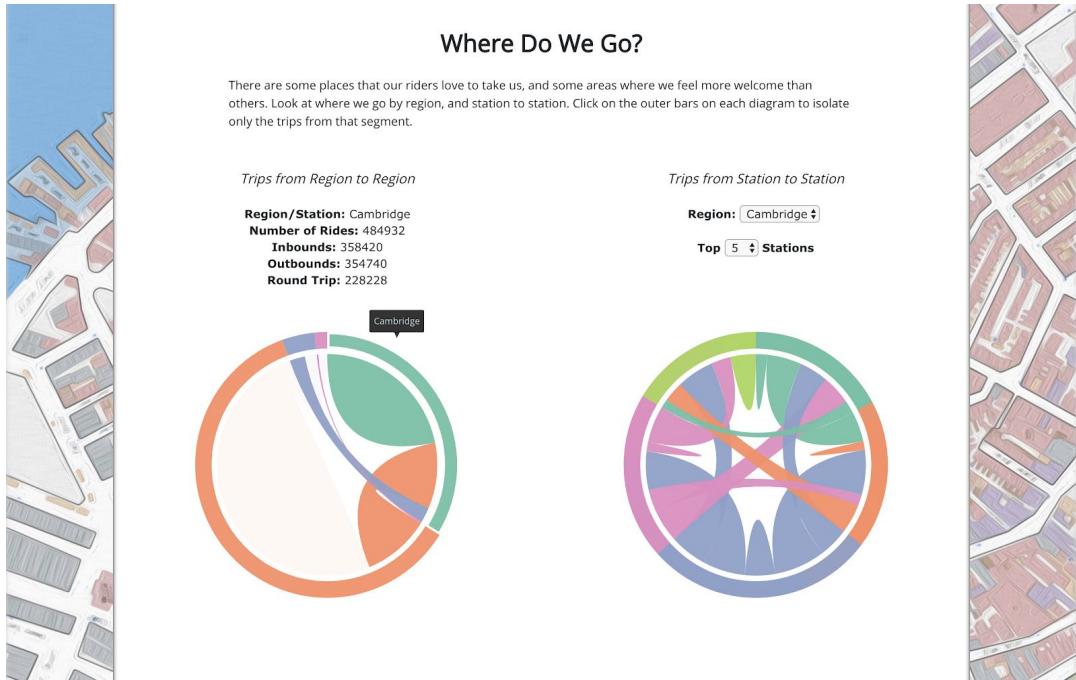


18 bikes started from Porter Square Station in week 1, totaling more than 36.61 km and 241.35 minutes. This saves at least 0.37 liters of fuel and reduces carbon emission by 0.84 kg.



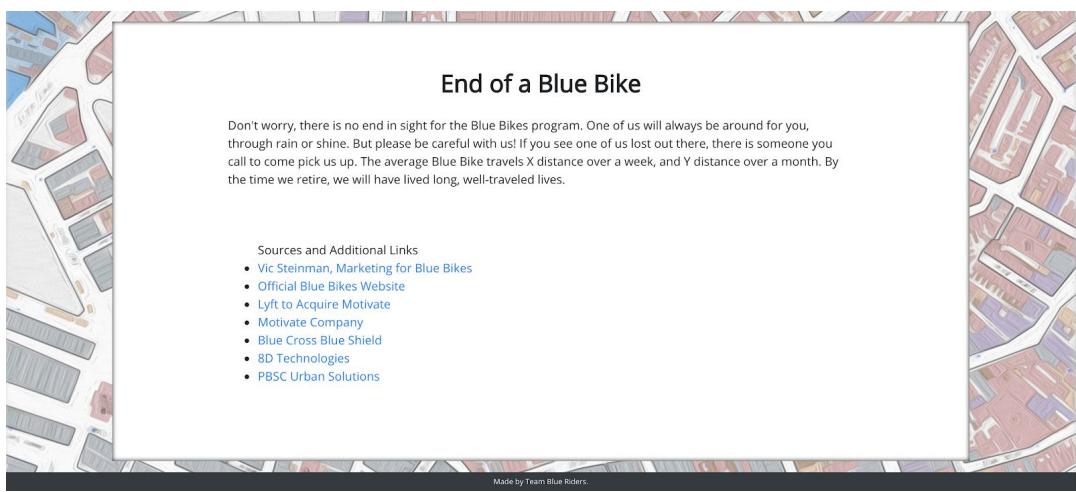
## *Page 5: A Week (or Month) in the Life of a Blue Bike*

This page focuses on the trip duration/distance of a blue bike and the carbon/fuel cost it saves. Users can choose which blue bikes they want to analyze through multiple filters including time, starting station, etc. The bar charts show detailed information about all trips given the filters, and the text block will update automatically showing how much energy those blue bikes save.



### *Page 5: Trip Flow Chord Diagrams*

There are two chord diagrams showing the total number of rides and traffic flow directions between different regions and stations. We provide detailed information about how the total number of rides differs between regions and stations and show the busiest region and station. Clicking on the arc of one region will highlight the region on the left and show the interactions of top 5 or 10 busiest stations within this region on the right.



### *Page 6: End of a Blue Bike*

The ending page focuses on the summary of a blue bike's life. We will provide statistics about average distances traveled and other general conclusions. Not much information is available on the end process after a Blue Bike is put out of commission, so maybe we will change to the spread of bike systems.

# Week 6

## Final Website

The public website of our project can be found on <https://mat3049.github.io/cs171-2018-final/>. If the visualizations do not show up, please make sure to unblock the content and load our scripts. The bar charts and chord diagrams will take around 30 seconds to load as the trip data file is over 150 MB.

## Screencast

We elected to first record the proper narration while going through site interaction, and then “reverse-dub” the screen-recording, following along our narration. The audio and video were then multiplexed into the final product.

The video can be found here: <https://youtu.be/a9ErRUmqRrk>.

## Github Repository

Our codes for the website can be found here:

<https://github.com/mat3049/cs171-2018-final/tree/master>.

## Final Updates

### Data Sourcing

Throughout the course of developing a backing solution for our data, we found that only small queries for ride data were feasible technically. As some of our visualizations rely on summary data, it could be possible to run SQL aggregations over the ride data, but we had limited time. As a result, we maintained usage of the full dataset in CSV format for those visualizations. Due to the large size of total trip data (176 MB) in 2015, we decided to use Git Large File Storage so that we can host our website through Github Pages.

## Visualizations

### 1. Map

We added distance and time comparison, carbon reduction and multi-station interactions to the map. We modified the page layout. Now the map is on the left, and the forms containing station and rider demographic information are on the right. Users no longer need to scroll up and down the page to see the results after clicking a station on the map.

### 2. Bar Charts

The x-axis now shows the value and the type of information displayed by the bars. We changed the color scheme to “PuBu” to match the color used in the other visualizations.

### 3. Chord Diagrams

The chord diagram on the left now displays the names of each area. The color scheme of both diagrams are changed to be more consistent with other colors used on our website.

## Instructions for Interactions

We added more instructional texts to our website, which show the users how they could interact with our visualizations. These texts are all in bold so that the users will notice them easily. In general, we incorporated more texts into our website to give users an overview of the information we want to convey.

## Evaluation

We told the story about the life of a blue bike from the bikes’ perspective, including how Blue Bikes were made possible and how much activity they have had over the course of 2015. We think we did a great job showing where the blue bikes go, how they’ve proliferated, and the effect they have had on their rider’s lives and the environment. We hope our project would motivate more people in Boston to choose bikes over cars when they commute.

Our visualizations convey the message we want to deliver well. If we have more time, we would overlay more barcharts showing station and rider demographics information on the map and we would incorporate color and size encoding to the markers representing stations.