Bikes on Parade

Background and Motivation.
A lot has been done in the way of analyzing bikeshare data within individual cities, but there is little comparison between cities. This is strange because there is extensive and parallel data on many bikeshare systems, but perhaps because of its short period of existence, not much branching out has been done. Plus this would be interesting because it is the first time that mass bike traffic data has been available for analysis.

About the Team:
Ina is a health nut. She is also REALLY into the environment. She once tied herself to a tree trunk, so that they wouldn't cut the tree down. They cut the tree down anyway, and now she feels like she has to do this project to make up for that shortcoming.

But not as nutty as Fox, who likes bikes, lots of bikes. He mowed down an entire forest once with bikes.

Project Objectives.
Comparing bike share systems across cities, we would like to see what different trends arise, as well as within the city.

1. Comparisons between cities/Overall trends
    - Provide a comparison of bike share usage between cities. Data on trip duration, miles traveled, subscriber vs casual user numbers will be visualized over time and compared between the cities.

2. Bike Usage in each city
    - Dynamics of bike usage in each city: flux of bikes in and out of a station, popular trips from/to stations, difference between casual and subscriber trips (particularly in the summer with tourists in the city)

Data.
From where and how are you collecting your data? If appropriate, provide a link to your data sources.

1. Boston (Hubway)
    - data per trip
    - start/end station, start/end time, gender, zipcode, station long/lat
    - duration and miles traveled can be calculated

2. Chicago (Divvy)
    - data per trip

- start/end station, start/end time, gender, year of birth
- duration and miles traveled can be calculated

3. DC (Capital Bike)
-data per trip
-start/end station, start/end time, gender, year of birth
-duration, miles traveled

Data Processing.

Data can be pulled from website in csv files with the relevant information. Subscribers and one-time users have different user data (gender, birth date) which may need some filtering but the data clean up will be relatively minimal.

We will need to calculate the duration and miles traveled for each trip (if we end up using these quantities). Duration can be calculated from trip start and end timestamps and the miles traveled can be calculated from trip start and end stations (may also need to find station GPS location for DC and Chicago where the station longitude/latitude is not provided).

Visualization.
How will you display your data? Provide some general ideas that you have for the visualization design. Include sketches of your design.

Clustered histograms by month for comparisons across cities normalized by amount of users. (or line graph depending)

Map visualization, possibly a bubble graph to show stats at different stations
- station flux (negative/positive flux in red/blue)
- showing tourists

Chord graphs to visualize trips between stations

Line graph - to show “actual speed” metric

Must-Have Features.
These are features without which you would consider your project to be a failure.

0. A website

1. Overall comparison
- miles traveled, number of trips, duration of trips for each day across the three cities

2. Casual (linked to Tourist) usage trends vs Subscriber usage trends within the same city, and then across cities.

3. Station flux (are any stations consistently empty?) / popular start/end stations

Optional Features.
Those features which you consider would be nice to have, but not critical.

- other parameters to explore in the visualization
    - busiest stations for each city
    - weather associated with each day (precipitation, temperature)

- track a random user

- 24 hour Time-Series Animations

- Measuring Actual speed in different cities. How far can you actually get travelling by bike in different cities?

Project Schedule.
Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

3/30
Data:
- parse/consolidation of data
- process data (functions to calculate geographical distance and duration)

Visualization
- Setup map framework

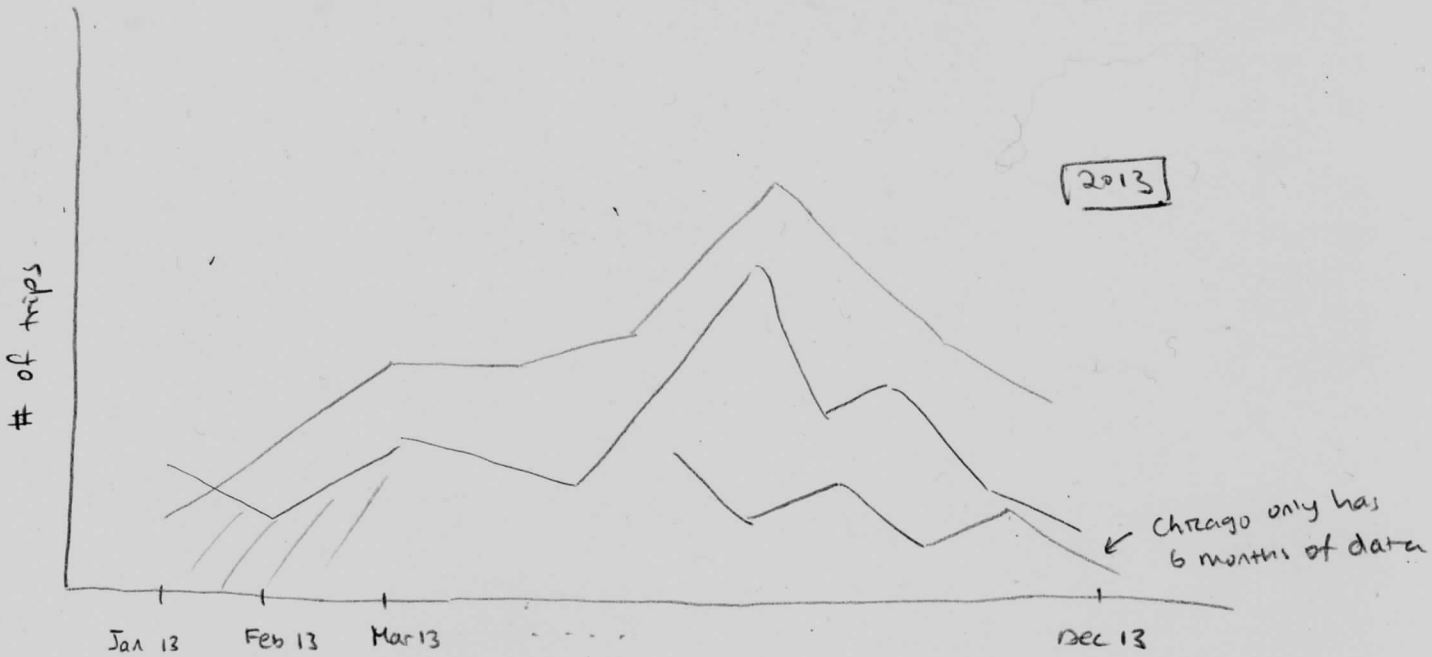
4/6
Visualization
- histogram (overall comparison)
- progress on visualizing within city

4/10
Functional draft!

4/20
Aim to finish must-have features

Until 5/1
Depending on progress, either doing Optional features or tweaking must-haves

Cities Overview

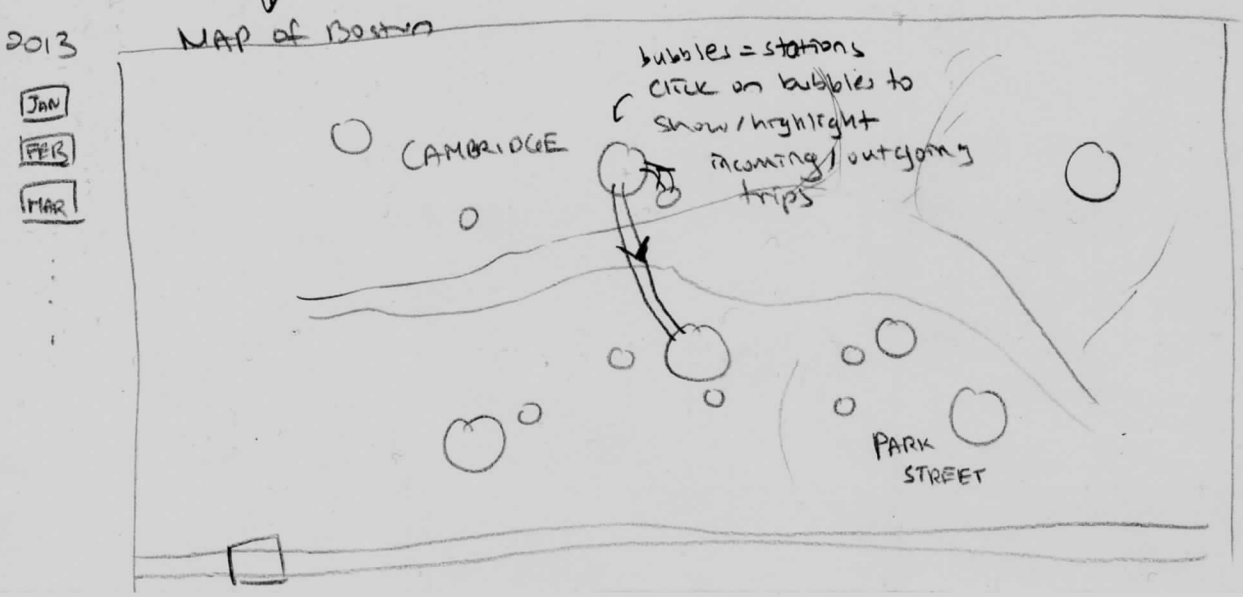


- ○ # trips / day
- ○ miles / trip
- ○ total miles / day
- ○ average speed

(option #1)

- ○ casual
- ○ subscriber

(option #2)



- ○ Visualize all Trips
- ○ Clear all Trips