Basic Info:

Project Title: How does your city commute?

Team Members:

- Manish Roy - u1145372 -manish.roy@utah.edu
- Rushit Sanghrajka - u1142286 - rush.sanghrajka@utah.edu
- Lama Albarqawi -u6011663 - u6011663@utah.edu

Project Repository: https://github.com/manishroyz/DataVizProject17

Background and Motivation

For most of biking around the neighbourhood was synonymous with growing up and becoming independent. As we grow older much of this relationship with bikes fades away owing to the various challenges like longer commute time, logistical aspects concerning maintenance etc. While many continue biking and also use it to commute for work, this segment largely fluctuates owing to various factors.The prominence of Bike sharing as a service, leverages the interest for biking while diminishing the logistical challenges that usually turn riders away from pedaling away.

Although there exists data and studies, we were quite drawn to notion of augmenting commuting patterns by churning these data sets.

Project Objectives:

With the visualizations we aim to present the macro trends, some of which aren't that abstruse. We intend to correlate the plausible relationships between factors like weather conditions, seasons, and demographics on biking patterns. Also through intuitive representations, we hope to find subtle nuances that shape user preferences. What are the reasons people use bikes for, and to what extent ? Is it for commuting to work, for exercising, for buying groceries at the local store, or simply as part of leisure activity.

There are numerous aspects that are usually not explicitly detailed in reports and statistics about the biking preferences. Many a times such representations can skew people's opinions. We thus want to present a detailed yet unbiased version, that motivates people to bike more

Data

In our project we will use datasets for Bike shares in six cities and their distribution over stations within each city. We are planning to cover New York City, Montreal, Columbus OH, Washington DC, Chattanooga Tennessee and Minneapolis MN. Datasets are accessible in both JSON and csv formats through the link below:

https://github.com/BetaNYC/Bike-Share-Data-Best-Practices/wiki/Bike-Share-Data-Systems

https://montreal.bixi.com/en/open-data

Data Processing

Data cleanup is needed in order to make sure that our study covers the whole domain of interest for each city, which spans from June 2016 until Dec 2016. The attributes that will be derived from these data are as following:

Trip duration, Start Time, Stop Time, Start Station ID, Start Station Name, Start Station Latitude, Start Station Longitude, End Station ID, End Station Name, End Station Latitude, End Station Longitude, User Type, Age (calculated from Birth Year), Gender, as well as counts for the total numbers of cyclists per station.

Since each city has its data divided between several csv sheets, we will start our data processing by organizing sheets per city then we will need to merge them into one sheet also per city, to be used later in the visualizations using d3. Additionally, we need to make all the files columns consistent with each other, for example: User Type in most of the files is categorized to Subscriber or Customer whereas in one of the files is categorized to 1 or 2. This is expected when using multiple sources for the data.

Then we need to do data filtration process. This step is crucial at this point and must be done carefully to take care of the lie factor issue and to guarantee that the resulted story that will be told later from the visualization is faithfully describing data. First step will be to choose bike stations that at least have on average more than 100 cyclists' records per certain day. Then we will filter those stations to ensure gender diversity. That at least we have a ratio 1:5 between both genders in the records per station. Then apply same ratio to the user Type. Then select the stations that have a connection with other station with more than 100 cyclists per day between them and ranking them ascendingly. Then rank the stations according to the number of connection with other stations. Finally searching for where those stations located on the map or what they are called and giving preference for the stations with more familiar locations over others. Finally selecting 7 random stations of the qualified ones.

Visualization Design

Design #1:

When the web page is launched, a map for the US and Canada will be displayed. Cities chosen for our project are the only ones that are going to be displayed on the map (as dots) and user has the option to interact with them. When the user clicks on a specific city (for example New York), it gets highlighted, and below the map area a detailed bike station detailed visualization will be rendered for the selected city. For each selected city 7 small bubbles will represent the bike stations in that city. The dots will be placed in a circular shape. A ribbon connection between 2 stations will be created to show numbers of bikes going from station 1 to station 2. The width of the connection will be dependent on the number they are representing. Upon user hovering over the connection the actual number of bikes will be displayed as a tooltip notification.

Three horizontal bars under the map and above the city stations visualization will be shown. They would have the following details:

- The first one displays the months from June to December,
- The second bar shows the days from 1 to 30 and
- The last one shows the hours from 12am to 11pm.

Default values for each of these 3 attributes will be used to draw connections between stations once a city is clicked, in order to give the user a head start on how he/she can interact with the visualization. Then the user can select the time and date that he/she is interested in seeing stations status during it. Based on the user selection the ribbon connections will reflect the respective data.

Four small labels at the top of the stations visualization will be visible on the web page, contains categories on which the user can select data to be visualized with respect to the selected category. We are planning to work on the following categories; Gender, Customer Type, Age group, Trip duration. User will be allowed to filter data based on one of those categories one at a time. User selection and filtration for the data will split each connection (ribbon) into two connections. Each connection (after category selection) will be presented in a specific color. A color map guide will be displayed for the user under the chart. For example, if the user chose Gender as a category, all the connections between the bike stations will either split into two connections or change the color of the connections. If connection between A and B has only male cyclists that selections will change the connection from black into blue. If connection A and B have both (female and male) cyclists, the connection will split into two connections which their width would be proportional to the count of the cyclists' gender. Each connection will have the gender color (blue for male, pink for female) as well.

Design #2

When the web page is launched, a group of bar charts listed horizontally will be shown. A bar chart for each city will represent the top 7 stations that have the most number of cyclists along the whole year. Each city (bar chart) will be presented with appropriate labeling. When the user clicks on a city, it will become highlighted and a map for that city will appear under the bar charts, as well as a line chart.

The 3 horizontal bars for the months, days and hours will also be available here at this design, directly after the bar charts. In addition to a new bar that holds categories on which the user can later select data to be visualized with respect to the selected category.

When the user clicks on a specific station, a map for the city of the selected station appears, showing the different densities of bike shares within the top 7 stations in that city. We decided to show only the top 7 stations in order to keep our visualization neat and preserve readability. The mark used for showing density will be a small bike shape.

A layered line chart will also appear when the bar is clicked, it will be populated by default with the counts of bike users that are traveling to or from the selected station over specific period of time and according to one of the categories mentioned above. This category as well as times for the visualized trips will have default values on the first load for the station layered line chart.

Design #3

Six tiles will be displayed once the page loads, each tile represents a city as a label in the middle of it. When one of the cities is clicked, the tile flips over to show a new box. The box will contain tabular view to hold four different categories on which data can be represented with respect to them.

Under this tabular view there will be pie charts that depict the distribution of the population on different categories.

Our **final design** that we have chosen to implement is **Design #1**. We found that a good and a rich story can be told from the data and this kind of visualizations, which as we imagine will be interactive and usable.

Must-Have Features

- Having the ability to visualize data for each station according to a specific category (one of the following categories: Gender, Customer Type, Age group, Trip duration).
- Providing the user with the more interactivity with the visualization by changing the time and date that is being visualized.
- Using tooltips to display more details on hovering over a station node or over a connection and using appropriate labeling
- Story being told beside the station visualization based on its data.

Project Schedule

Week 1:
Precursor work for Data munging and designing the detailed layouts for the project and code structure.The visualization techniques and the inter-relationships will also be sorted as a detailed blueprint.

Week 2:
Initial layout as an implementation of the designs and static data representations. Rendering all the required information that needs to be the primer layer. Present the same for the Milestone.

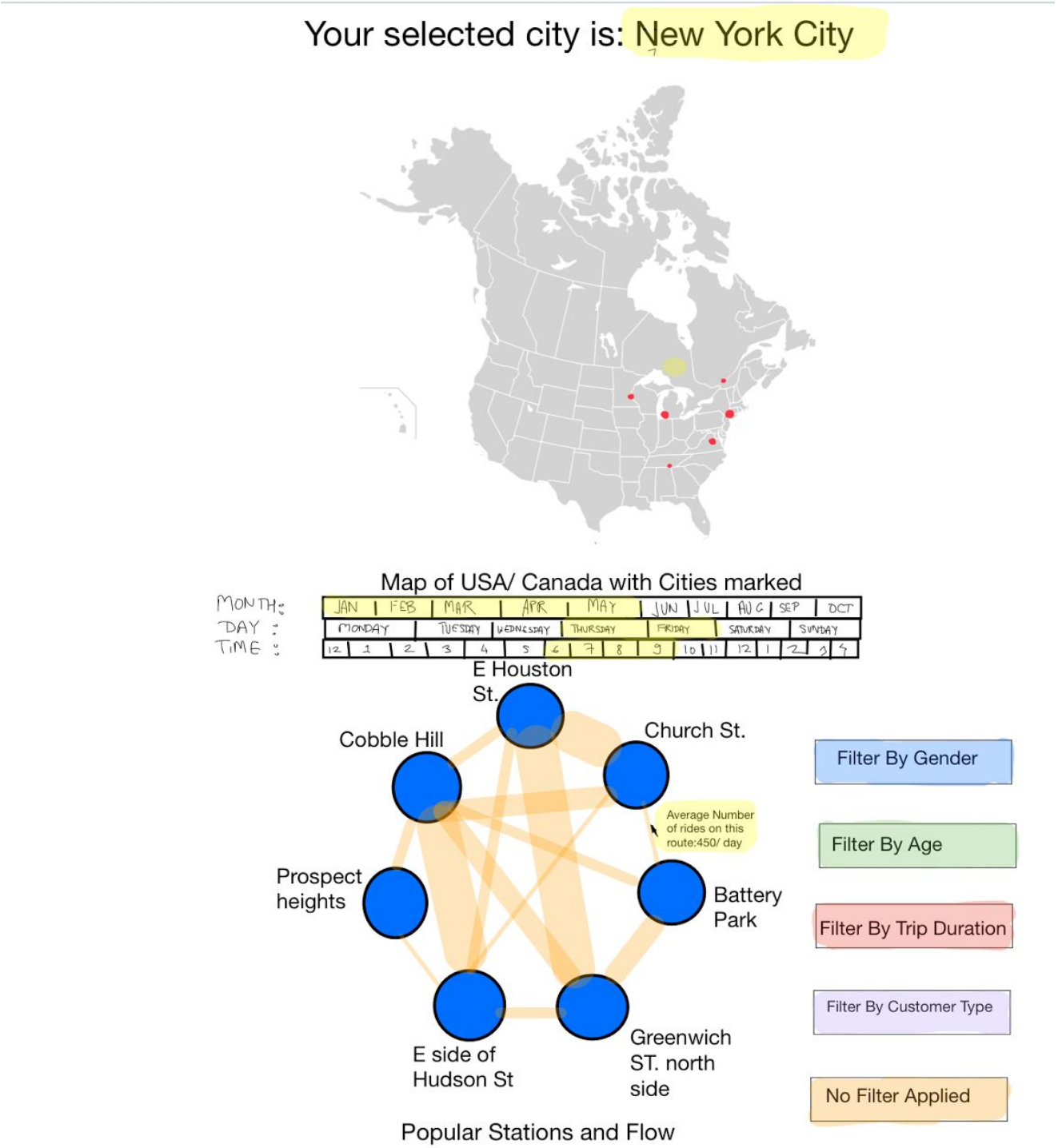
Week 3:

Implementing all the visualization layers and interactivity and exploring the results iteratively to draw insights into the data. Based on these tweak the designs wherever necessary to augment the interface and highlight the most important nuances in the data.

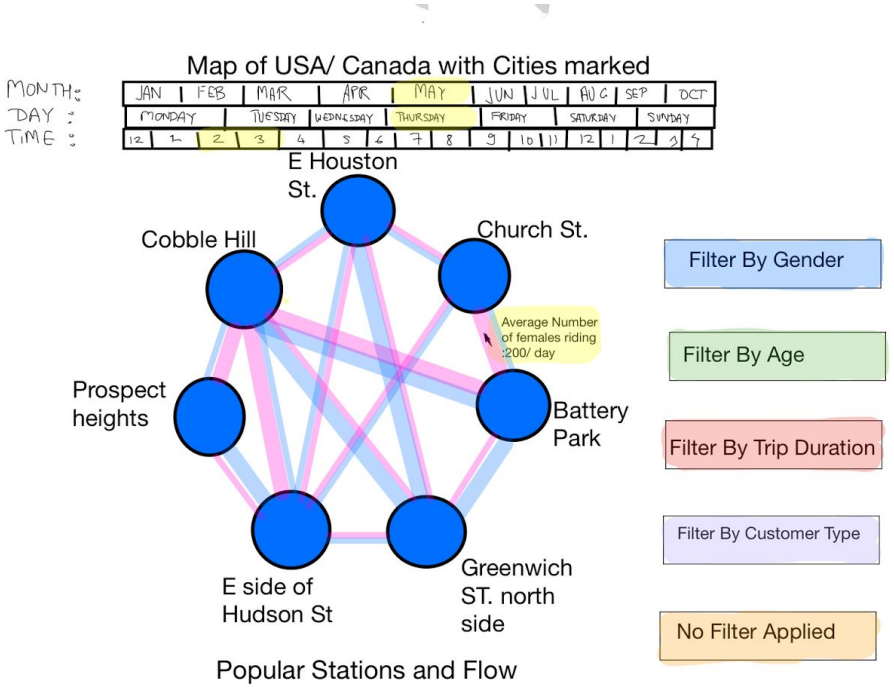
Week 4:
Complete few rounds of peer review to test any User experience aspects that could have been missed out by the team and fix the same .
Collate the project progress and sketches and designs made along the way in the Process book.
Prepare the screencast with narration along with slides for the project narration.

Design Visualization Sketches

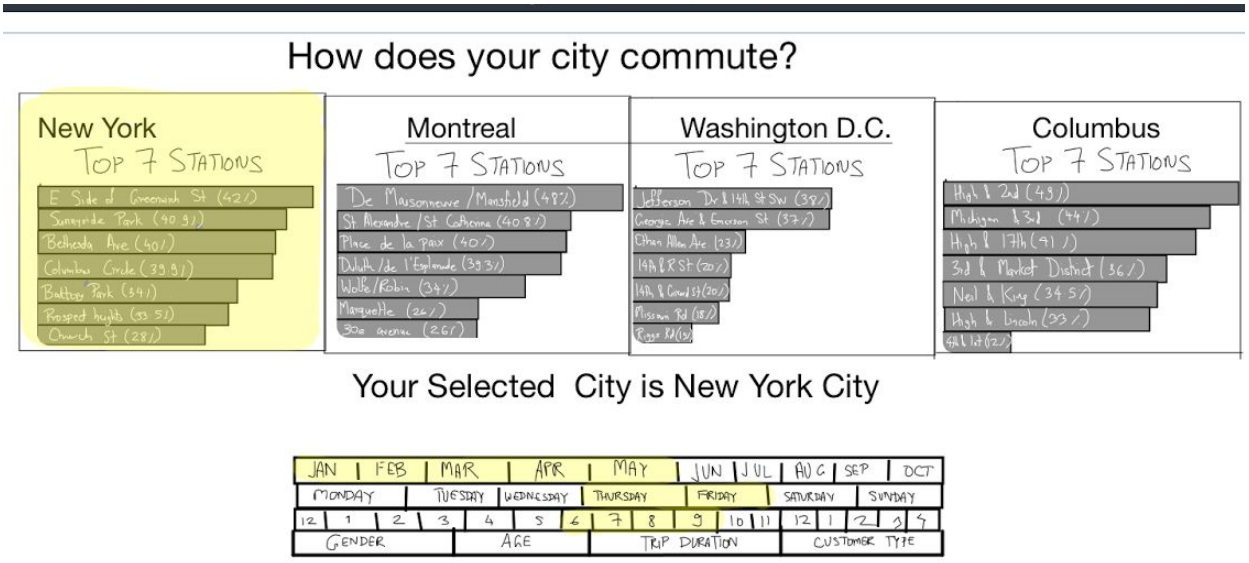
-DESIGN 1

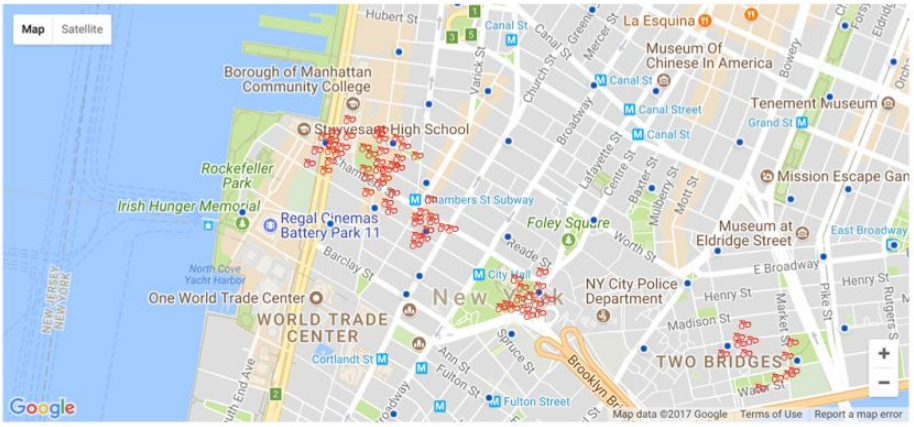


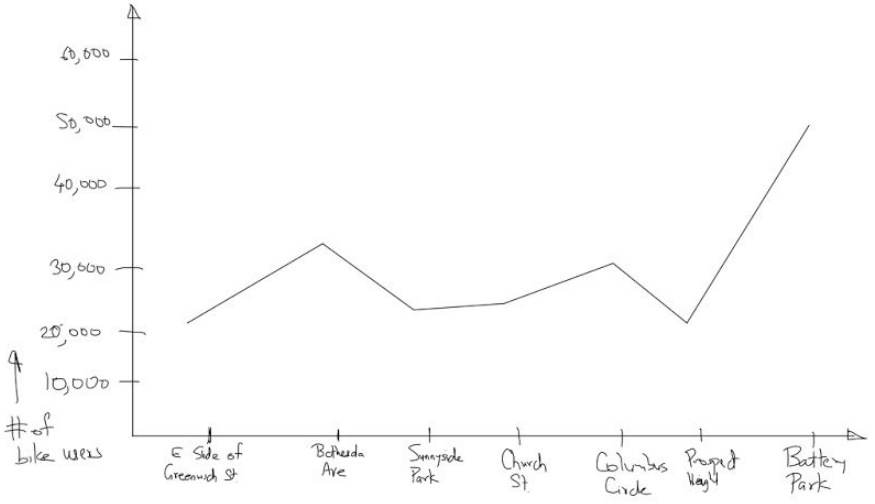
- Filtering by Gender



DESIGN 2







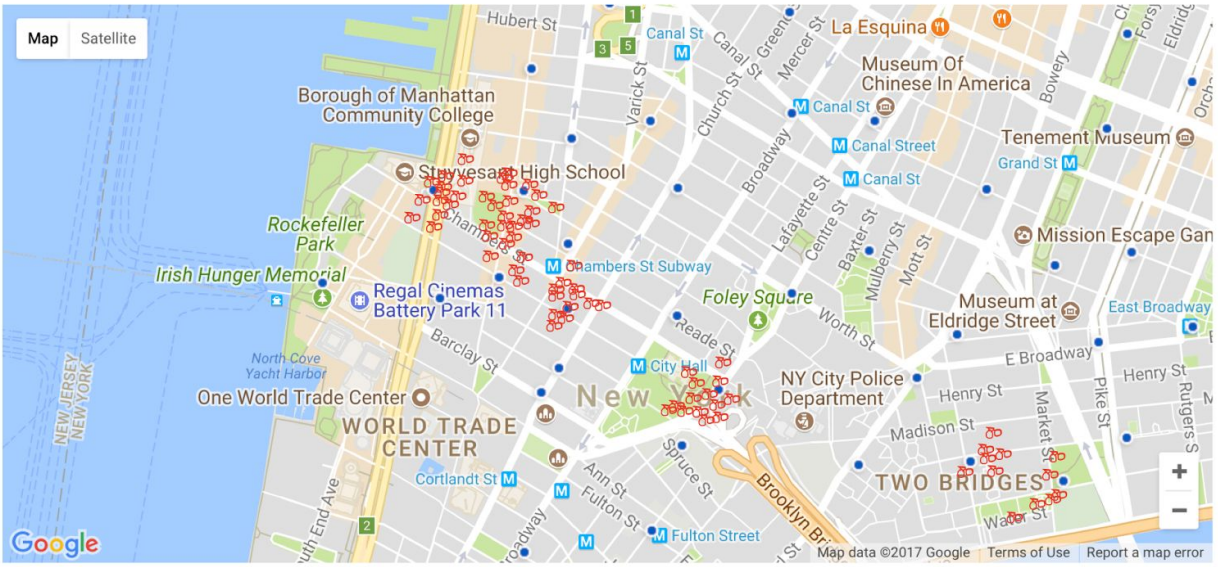
- Zoomed in view of map

JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT

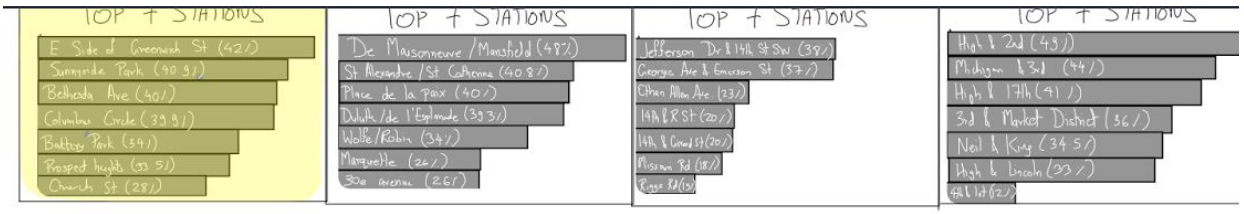
MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY	SUNDAY

12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4

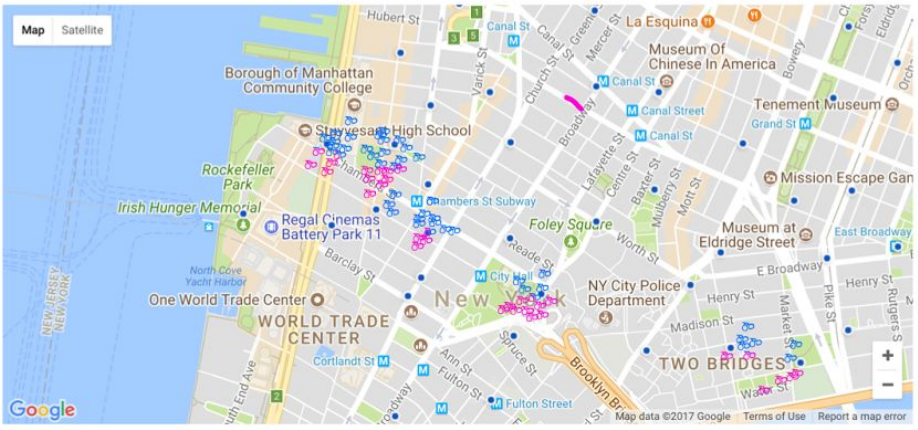
GENDER	AGE	TRIP DURATION	CUSTOMER TYPE

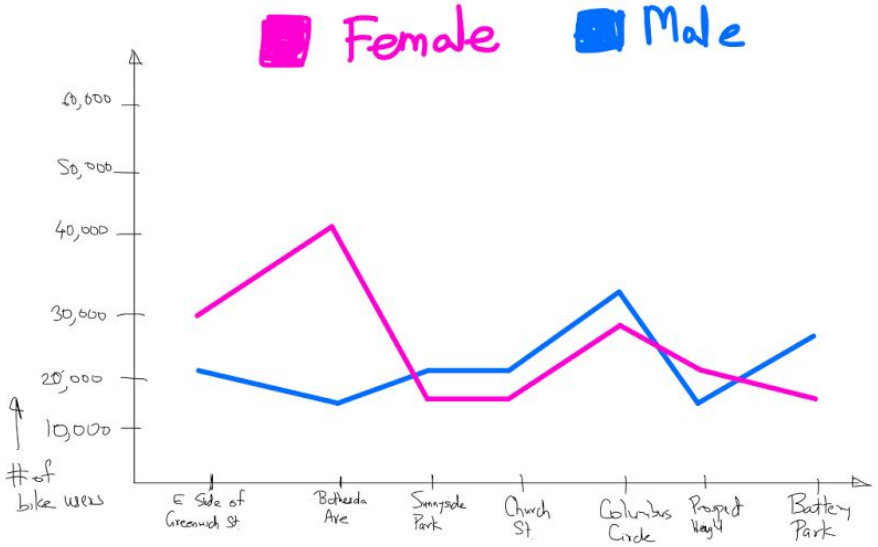


- Filtering by Gender

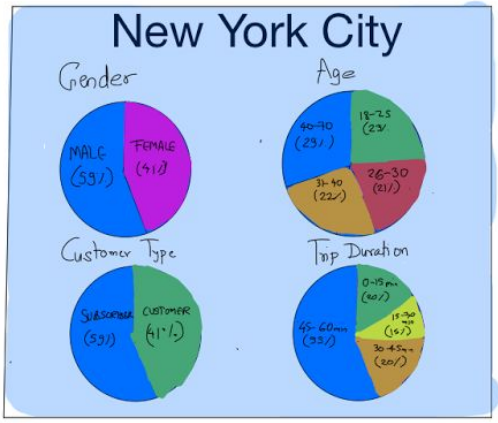


Your Selected City is New York City





DESIGN 3



Washington DC

Montreal

Minneapolis

Chittagong

Chicago

- Zoomed in view of tile