Divvy Bike Information Visualisation

Zheran Xu
University of Nottingham
xuzheran@hotmail.com

INTRODUCTION

Divvy is a bicycle sharing system in the City of Chicago and two adjacent suburbs operated by Motivate for the Chicago Department of Transportation.

Divvy's trip data set is available for public use, which is anonymized and includes trip records and station information.

The motivation of this project is to create a framework system to analyse from various aspects, explore potential relations, and visualise the result, which could the encounter the difficulty of perceiving metadata.

RELATED WORK

Most work related to this project is to draw lessons from existing system, which is visualising transportation information.

1. Divvybike System

- https://member.divvybikes.com/stations
- An official visualisation with station info with the interactive icon. This visualisation allows users to view the station status on map.

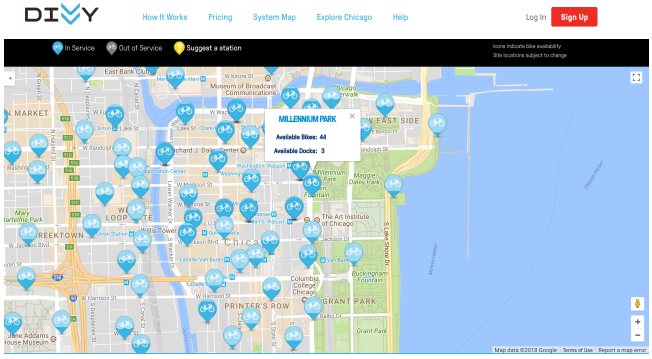

Figure 1. Divvy Bike system

2. Bike Visualisation

- https://www.visualization.bike/
- The most complete bike sharing visualization and analysis websites, which is adapt to dataset from 22 cities bike sharing system. The analysis with visual encodings contain various components from system, station and activity perspectives.

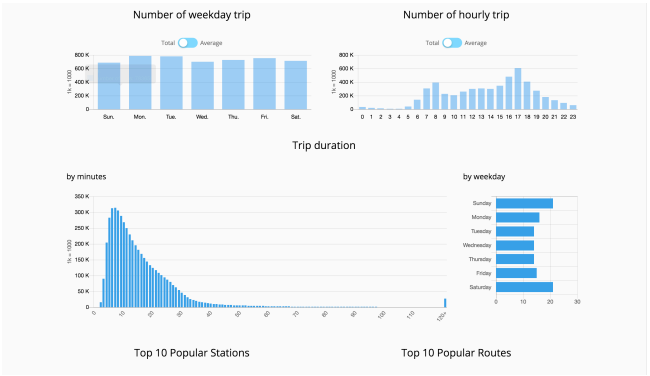

Figure 2. Bike Visualisation Websites

3. NYC Citi Bike Rebalancing Study

- https://urbica.co/citibike/
- A detailed analysis and visualization of trip routes (See Figure 3) and stations (See Figure 4). The information is displayed in a fancy method with in-depth data analysis.

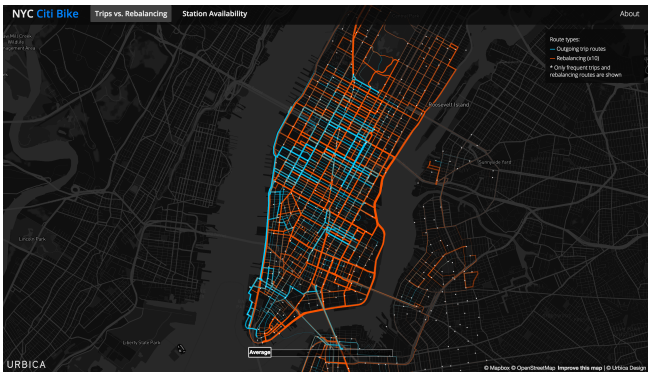

Figure 3. NYC City Bike, Visualisation of Routes by time.

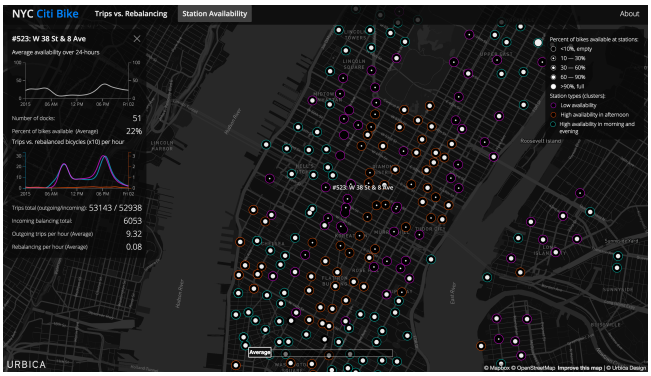

Figure 4. NYC City Bike, Visualisation of Data Analysis by stations and time.

METHODS/DESIGN

1. Data Analysis

Data used in this project is Divvy trips in Quarter 4,2017, which includes 669,239 trip records from 4/1/2017 to 6/30/2017, and 586 station info. Data source is from Divvy system data.

(https://www.divvybikes.com/system-data)

Variables	Descriptions	Example
trip_id	ID attached to each trip taken	17536701
start_time	day and time trip started, in CST	12/31/2017 23:58
stop_time	day and time trip ended, in CST	1/1/2018 0:03
bikeid	ID attached to each bike	3304
tripduration	time of trip in seconds	284
from_station_name	name of station where trip originated	Claremont Ave & Hirsch St
to_station_name	name of station where trip terminated	Damen Ave & Pierce Ave
from_station_id	ID of station where trip originated	159
to_station_id	ID of station where trip terminated	69
usertype	"Customer" is a rider who purchased a 24-Hour Pass; "Subscriber" is a rider who purchased an Annual Membership	Subscriber
gender	gender of rider	Male
birthyear	birth year of rider	1988

* Gender and birthday are only available for Subscribers

Table 1. Metadata for Trips.

Variables	Descriptions	Example
id	ID attached to each station	2
name	station name	Buckingham Fountain
city	Station city	Chicago
latitude	station latitude	41.876393
longitude	station longitude	-87.620328
dpcapacity	number of total docks at each station as of 12/31/2017	27
Online_date	date the station was created in the system	6/10/2013 10:43

Table 2. Metadata for Stations.

- From the variables of trips info and stations info, the data the vision of analysis could be generally separated into three aspects, trips info, station info and user info.
- From the perspective of trips info, for each record, there could be generate routes by from and to stations with time and duration variables, With time info, trips info could also be grouped into time sequence. In this project, the time sequence is floor to 1 hour and use start_time as time point.
- From the perspective of station info, station feature could be analysed at first. With the data from trips, the flow of traffic by station could be calculated. Additionally, it could also merge with time feature. In this project, the calculation of trips frequency that station involved is use both from_station_name and to_station_name.
- From the perspective of user into, general user type could be count, as doing statistics with other features.

2. Visual Encoding Design

For routes and stations with geographic variables, they can be visualised on map, with other features being represented by colour. For statistic results, the figure could be applied into charts such line chart, with category distinguish.

3. Web Design

The visualisation webpage should be user-friendly with consistency. The typeset of charts could be grouped by themes with proper interactive icon.

IMPLEMENTATION

1. Data Manipulation

- R – dplyr/lubridate

  Doing statistics according the visual plan. Most work is filter data by categories and group them together. Then count the figure to get the frequency, which usually is the final visualisation data list.

  Another typical work is to manipulate with time data such as transforming to UNIX timestamps or group them by time sequence.

- Python

Python is used to transform data into the proper data format, such as JSON file or Geojson, which is applied for following visualisation.

2. Data Crawling

- OSRM (Open Street Route Machine)

  OSRM is an open API to generate the shortest route from two locations(See Figure 5). By posting longitude and latitude from two locations, the server would return the shortest route in Geojson line format.

- Python

  A python crawler is developed to collect generated routes data by ONE-DAY(6/1/2017) records of trips. The method of crawling is to keep posting the station location correlated to every single records of trips. Additionally, the return of data have useless properties, which are excluded by scripts.

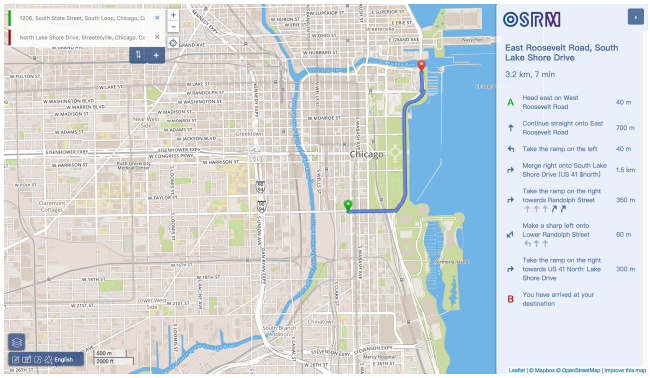

Figure 5. OSRM generating Routes.

Figure 6. Mapbox Studio. Creat map style

3. Geo Visualisation

- Mapbox gl js

  Mapbox is a powerful geographic visualisation tool, which allow the geo-info to be plotted in an interactive map. The first step is to design map style in Mapbox studio(See Figure 6). Then add datasets as tilesets, and create layer on map accordingly. After that add interactive button.

4. Statistic Visualisation

- Highcharts.js

  Highcharts is an interactive JavaScript library for information visualisation.

  The related work is to transform manipulated data into correspond data list. Then visualize by different charts and design theme style. Create HTML pages.

5. Front-end Web Development

- Using Web Template

  In order to create the web site, some work like Analysing the frame of template is been done at first. The following is to transform and integrate all visualisation into one HTML page with using template CSS. And JavaScript is separated in each chart.

RESULTS

Introduction and Header

A brief introduction of projects. The top navigation bar have link to the following section of visualisation parts, which are concluded into four types, Maps, Staions, Trips and User(See Figure 7).

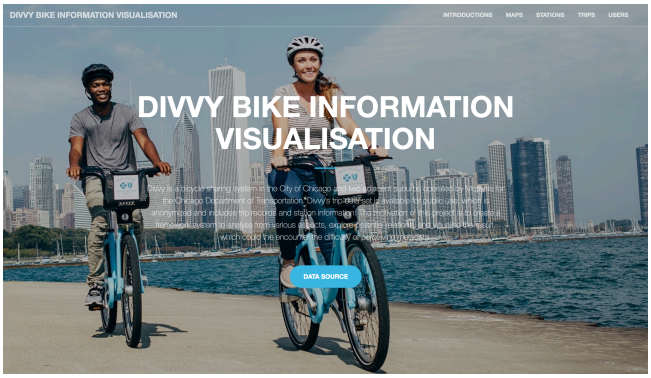

Figure 7. Header.

Maps with visualisation of routes and stations

The first map visualise the data mentioned at crawling. There are interactive buttons to filter data by duration, which is roughly divided into three time sequences, and represented by different colours of line(See Figure 8). The map is zoomable, which allows detailed trip route location with 3D buildings(See Figure 9).

The second map visualise location of stations and display station info on the marker's tooltip when user click it(See Figure 10).

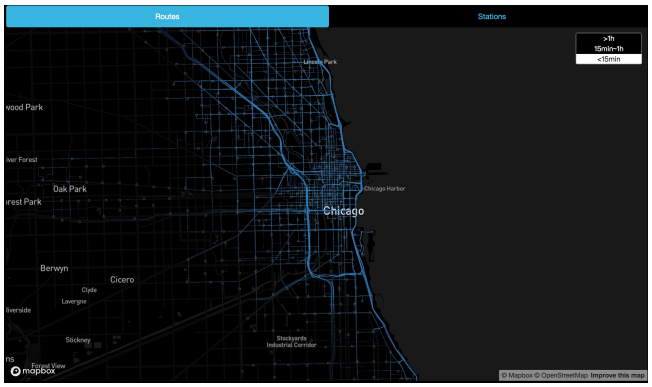

Figure 8. Map. Routes Plot on Map.

Figure 9. Map. Displaying with 3D Building Model.

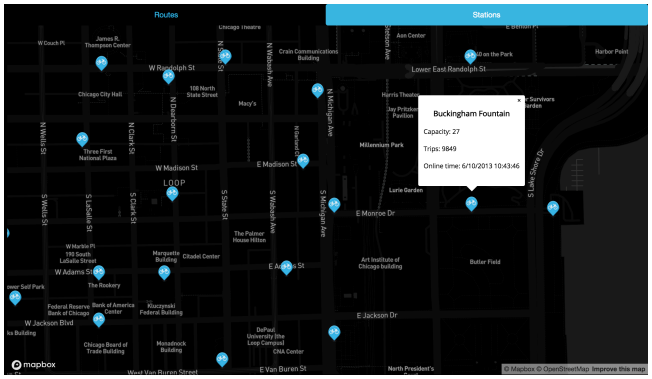

Figure 20. Map. Interaction with Marker.

More detailed visualisation of station info

The Streamgraph (See Figure 11) visualise the data of top 10 popular stations, which means they have leading numbers of trips frequency. The area of categories represents the sum of trips frequency in one quarters by hourly interval.

The Stacked column chart (See Figure 12) visualise the trips frequency of stations and divide trips into four colours by trip duration.

The Parallel coordinates charts (See Figure 13) group all stations by four variables into on charts. And illustrate that the relation between variables such the capacity and trips frequency.

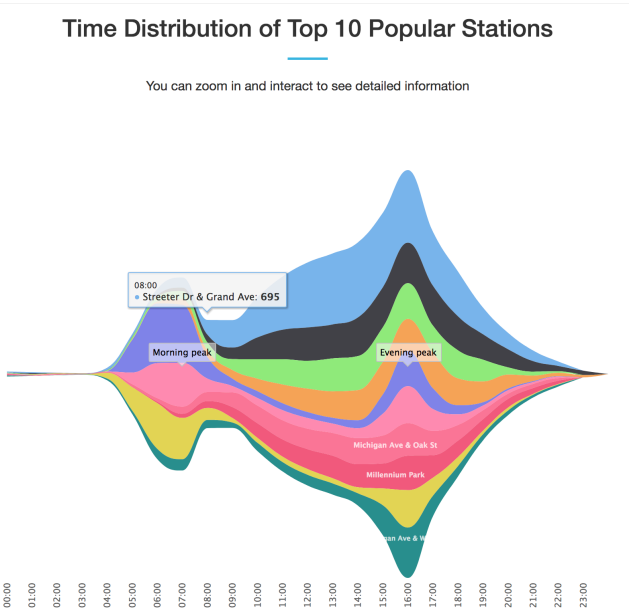

Figure 11. Station. Streamgraph.

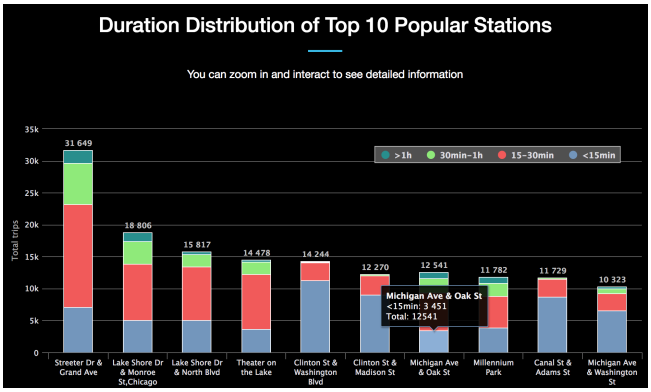

Figure 32. Station. Streamgraph.

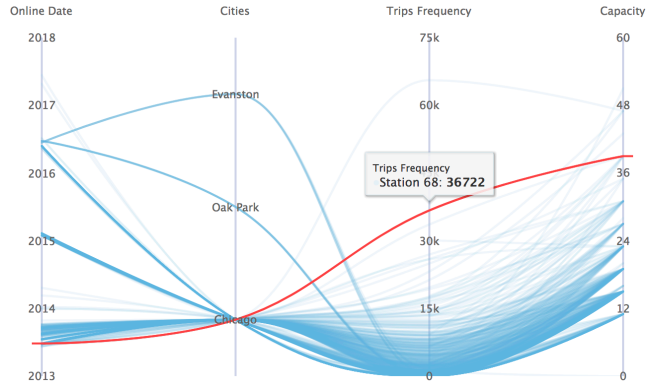

Figure 43. Station. Streamgraph.

Trips with visualisation of time info and bike

The Line chart (See Figure 14) visualise the data of the statistic results by timestamp, which include three months data. And it is zoomable to see the info by hours (See Figure 15).

The zoomable Scatter chart (See Figure 16,17) visualise the sharing times frequency of bikes. It shows a normal distribution trend.

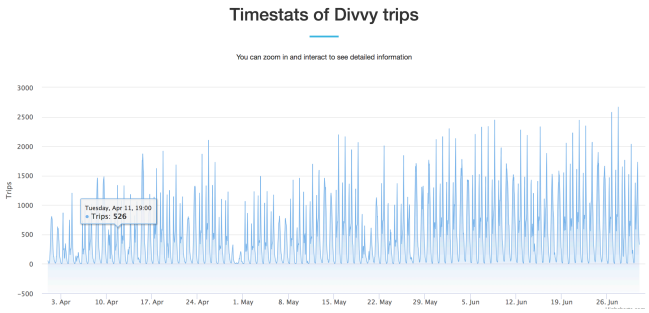

Figure 54. Trips. Line chart.

Figure 65. Trips. Line chart zoom in.

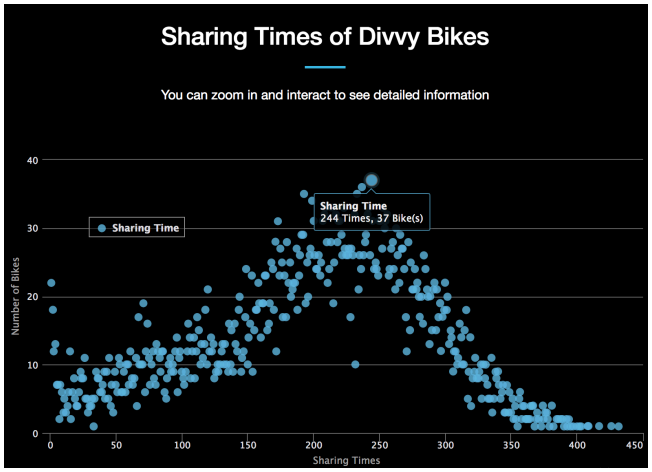

Figure 76. NYC City Bike, Visualisation of Data Analysis by stations and time.

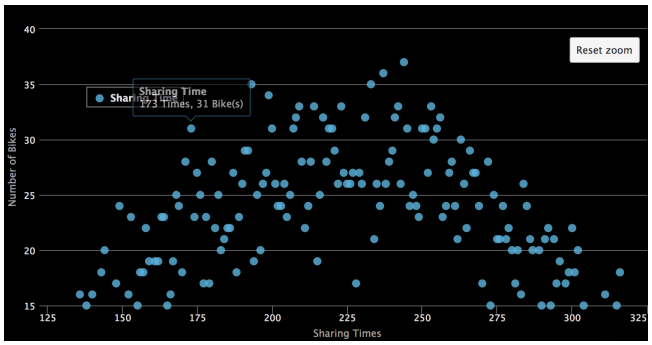

Figure 87. NYC City Bike, Visualisation of Data Analysis by stations and time.

User info

The Pie chart (See Figure 18) visualise the data user info by dimension of variables. Parts of pie have drilldown with another level of variables (See Figure 19).

The Bar chart (See Figure 20) contrast the trips duration of Subscribers and Customers, as they have different feature.

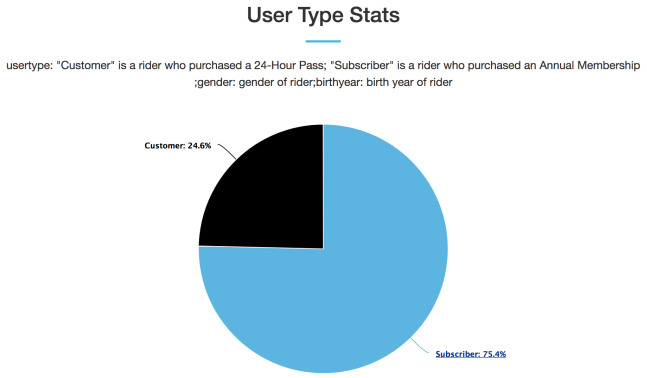

Figure 98. NYC City Bike, Visualisation of Data Analysis by stations and time.

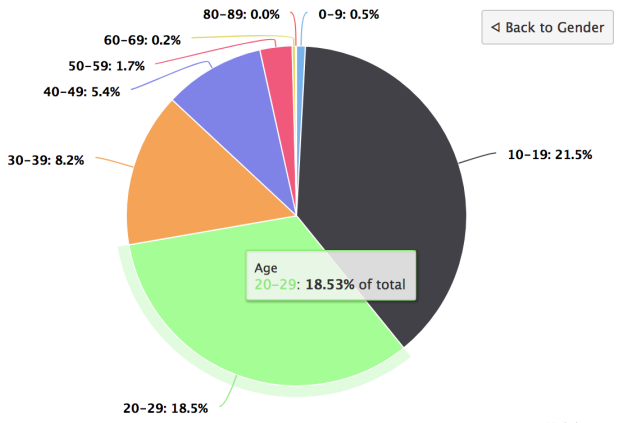

Figure 109. NYC City Bike, Visualisation of Data Analysis by stations and time.

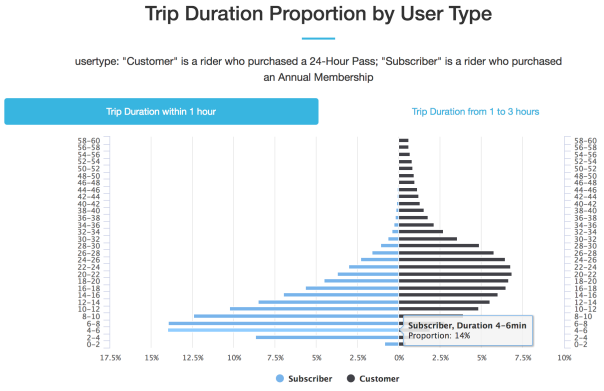

Figure 20. NYC City Bike, Visualisation of Data Analysis by stations and time.

EVALUATION

Most evaluation work is done by user trials. After completing chart visualisation and webpage integration, initial version of project is sent to people who is able to establish localhost. Some charts as JavaScript work well with quick respond time, while other charts have lags in loading zooming and dragging. The initial plan is to plot bubble scatters at highmap.js. But the performance is poor when calling JSON by Jquery (See Figure 21,22).

An issue raise by testing is that some charts take times to understand or have unrelated theme. Therefore, charts are removed or modified in visual encodings.

Another problem is the limitation of statistical analysis, as user wish to see more relations and results from the dataset, such as a visualisation by each station, bikes travel route or even user routes in one month.

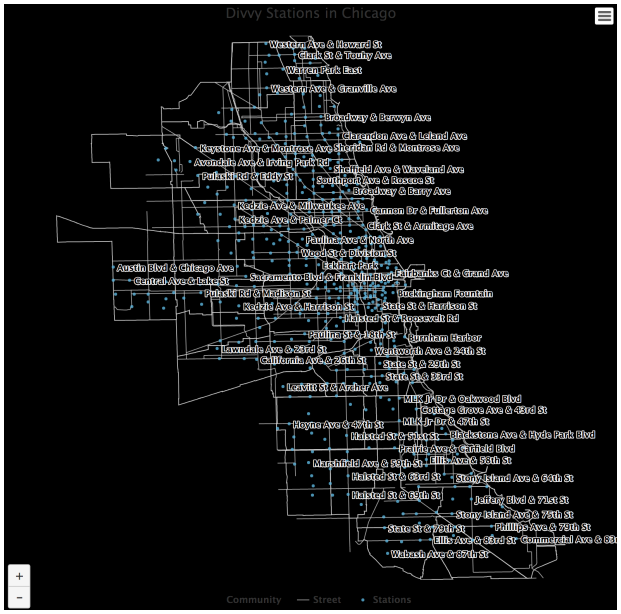

Figure 21. NYC City Bike, Visualisation of Data Analysis by stations and time.

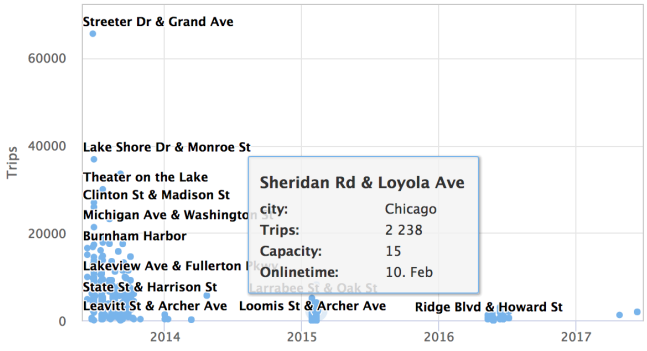

Figure 22. NYC City Bike, Visualisation of Data Analysis by stations and time.

DISCUSSIONS

From the perspective of functionality, the project is a simple frame work for sharing bike system, which encounters the problem that analyzing the trips records from metadata and visualising in an intuitive way, especially for the routes data. The interaction of map is enjoyable. The routes generated by OSRM is an approximate value as in reality there would be much more complicated routes.

The drawback of the projects is evident. For the sake of time limitation, the data analysis is considerably lacking. Another point is that more charts could be integrated into one, such station info integration with map.

FUTURE WORK

- Integration

  Holistic design for visual and put visualisation into more integrative platform based on Mapbox. Using buttons or icon to filter and shift information.

- Data Mining with Statistic Analysis

  More complicated relation of data could be discovered. For example, stations location could be regarded as cluster with network made up by trips, same as user info and time factors.

- Interface with Database

  The size of dataset in the project is considerably small but the during the crawling work it expand to larger size (generate Geojson with multiple points and properties). Database should be applied for storage. The system could also be able to compute the result in time.

- Live Data Visualisation

  Live station info is available from DivvyBikes. By developing API with JSON feed, a real time visualisation could be realised, which is more valuable than statistic one.

BIBIOGRAPHY

1. Mapbox GL JS Documentation
   https://www.mapbox.com/mapbox-gl-js/api/
2. Highcharts Documentation
   https://www.highcharts.com/docs
3. OSRM API Documentation
   http://project-osrm.org/docs/v5.15.2/api/#general-options
4. HTML5 Template
   http://groups.ischool.berkeley.edu/Drugs-Incarceration/
5. Mackinlay, Jock. "Automating the design of graphical presentations of relational information." *Acm Transactions On Graphics (Tog)* 5.2 (1986): 110-141.
6. Viegas, Fernanda B., et al. "Manyeyes: a site for visualization at internet scale." *IEEE transactions on visualization and computer graphics* 13.6 (2007).

7. Inselberg, Alfred. "Multidimensional detective." Information Visualization, 1997. Proceedings., IEEE Symposium on. IEEE, 1997.
8. Rao, Ramana, and Stuart K. Card. "The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information." *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1994.
9. Yi, Ji Soo, Youn ah Kang, and John Stasko. "Toward a deeper understanding of the role of interaction in information visualization." *IEEE transactions on visualization and computer graphics* 13.6 (2007): 1224-1231.
10. Carpendale, Sheelagh. "Evaluating information visualizations." *Information visualization*. Springer, Berlin, Heidelberg, 2008. 19-45.