Scalable Visualization Analytics

of

San Francisco Bicycle Sharing System

Master of Data Science from the **University of Sydney**

by

Xintong Chen

School of Information Technologies

November, 2017

Supervisors**: Prof. Seokhee Hong,**

Dr. Quan Hoang Nguyen

ABSTRACT

The project visualizes the performance and operation status of bike sharing system "Ford Gobike" in San Francisco. Data processing and analysis are conducted prior to visualization to explore the feature of the data and help to determine what visualization methods to choose. Line chart, chord diagram, heatmap set and map-based stacked bar chart are built interactively to visualize the system in both overviewed and detailed perspectives. A novel animation is created to dynamically track the traffic evolvement of the bike sharing system in a day and view the traffic pattern in any time period.

ACKNOWLEDGEMENTS

I would like to thank Professor Seokhee Hong for her guidance and constant inspiration on me to complete this interesting visualization project.

Statement of Originality

I, Xintong Chen, declare that this thesis, submitted as part of the requirements for the award of Master of Data Science, in the faculty of Engineering and Information Technologies, the University of Sydney, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications or assessment at any other academic institution.

Print Name: Xintong Chen

Student ID Number: 470138674

Date: 17th, November, 2017

Table of Contents

ABSTRACT 2

1. Introduction 6

    1.1 Motivation 6

    1.2 Aims 6

    1.3 Contributions 6

    1.4 Organisation: 6

2. Related Works 7

3. Design 11

    3.1 Visualization Framework 11

    3.2 Analysis 13

    3.3 Visualization 18

4. Implementation 21

    4.1 System Architecture 21

    4.2 Data Processing 22

    4.3 GUI 23

5. Evaluation 24

    5.1 Task 1 24

    5.2 Task 2 26

    5.3 Task 3 27

    5.4 Task 4 30

    5.5 Task 5 31

6. Conclusion 32

    6.1 Summary 32

    6.2 Limitations 33

    6.3 Future work 33

References 34

Appendix 34

1. Introduction

1.1 Motivation

Cycling is a fun way to get around, making short trips much more convenient and faster than car or tube. Bicycle share is an effective solution to encourage people to ride in urban area. Bicycle share is gradually becoming a life style among most large cities in the world which is efficient for short distance and friendly for environment. Like most form of public transportation, bicycle sharing needs a system for management and resources allocation. The bicycle sharing system for study in this project is Ford GoBike [1] in the San Francisco Bay Area, California. The system records the real-time information on weathers, availability, and operations like locking and unlocking of bikes at each station. It also keeps track of every rides. In order to improve the efficiency and functionality of the system and to be a better riding guide for riders, all historical data from the system need to be explored and analyzed. However, even cleaned and analyzed data may look unsystematic and uninterpretable, in which case visualization is required.

1.2 Aims

In this project, a set of interactively integrated visualization schemes will be developed to visualize the bike sharing system in both overviewed and detailed perspectives, interactively exploring the system data to view its performance and operation status. Furthermore, the project aims to apply visualization methods proposed in this project to practical problems on route selection, station selection, resource allocation and policy-making, etc. The visualization schemes are driven by a collection of one-year data (Sept 2015 – Aug 2016) from Ford GoBike system [1].

1.3 Contributions

This project introduces different visualization schemes to present different aspects of the system data. Line chart, chord diagram, heatmap set and map-based stacked bar chart are built interactively to visualize overviewed trips and routes network, detailed trips distribution and station availability information. A novel animation is created to dynamically track the traffic evolvement of the bike sharing system in a day and view the traffic pattern in any time period.

1.4 Organisation:

Section 2 summarizes visualization methods for bike sharing system from related works.

Section 3 presents the design of the proposed visualization schemes. Section 4 demonstrates how the visualization methods are realized. Section 5 evaluates the performance of the schemes by putting them into use to solve some defined tasks. Section 6 will make a general conclusion for the project and specify the works to be done in the future.

2. **Related Works**

Though bicycle sharing is a relative new form of transportation, the analysis and visualization of it are in great need due to its increasing popularity around world. In [2], a collection of half-year (2013-2014) data from Bay Area bike share system are used to analyzed and presented. It used different method to visualize processed data from different aspects in an overview scale, for instance, pie charts were used to summarize the amount of rides, docks, stations in different region and rides proportion of customer types, line chart with multiple filters was applied to describe the overview of rides in six-month period under different filter condition, a heatmap set was created to illustrate routes count for each pair of stations (fig 1.1), and an interactive stem-leaf plot (fig 1.2) was shown to address the popularity of every station as both start point and destination.

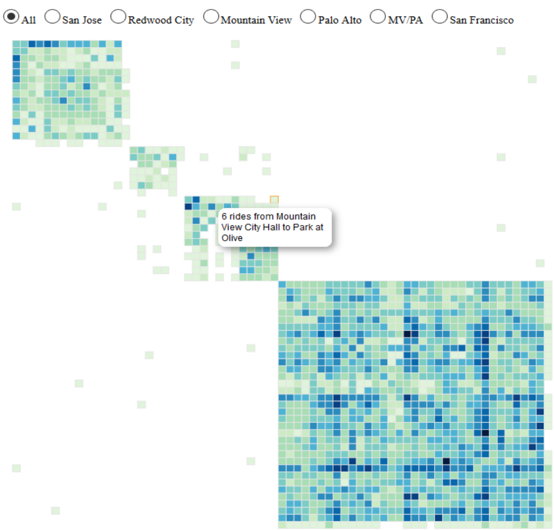

Fig 1.1 Route Heatmap Set

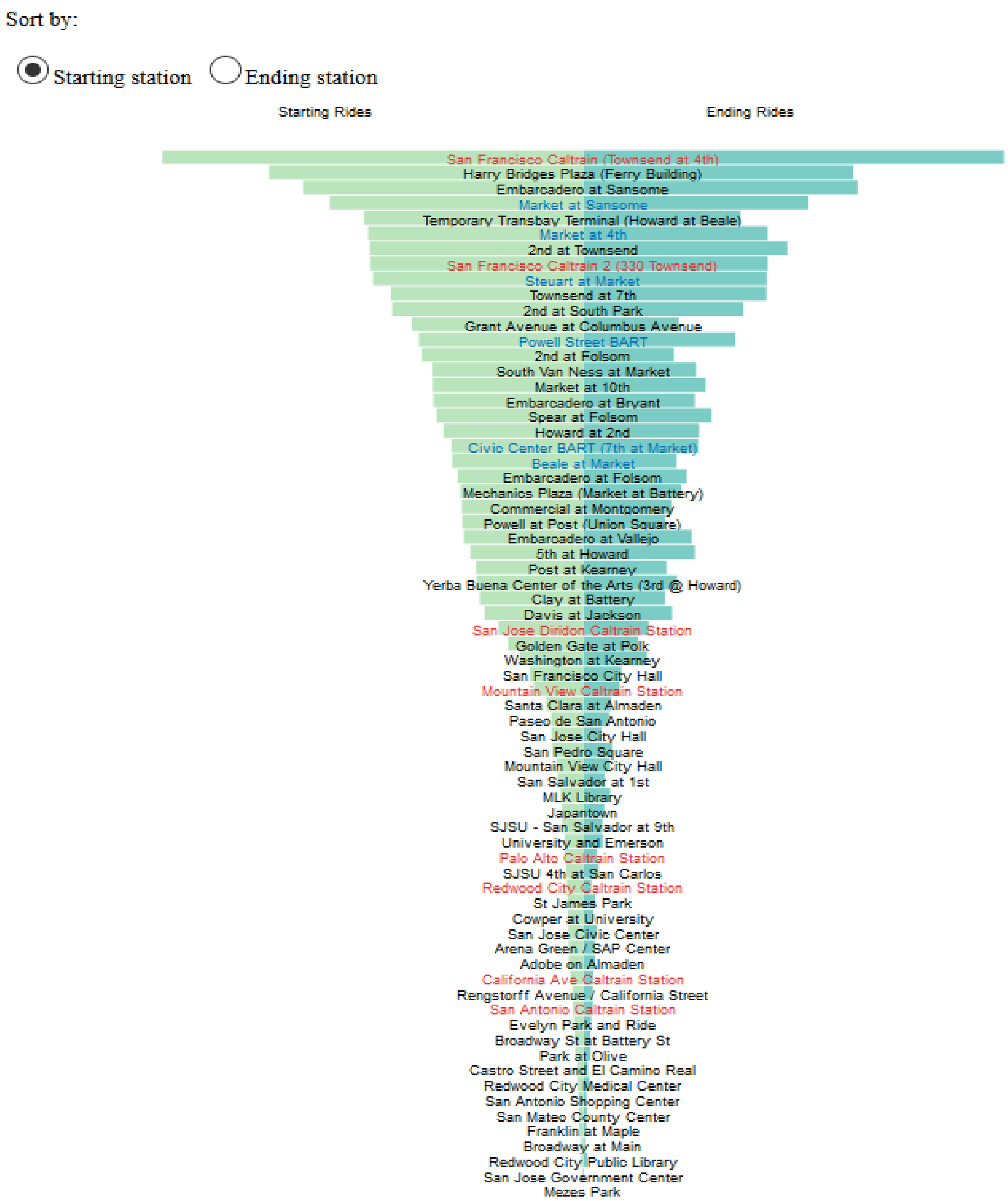

Fig 1.2 Stem-Leaf plot for station popularity

The author of [3] created a small map-based interactive visualization system (fig 1.3) over the same dataset as in [2], it focused on details presenting such as average bike and dock availability throughout whole day at each station, route details between each pair of stations (fig 1.4) and trip distributions against time. Elevation fluctuation is visualized for each route after matching with attitude information.

Fig 1.3 Interface of the map-based visualization system

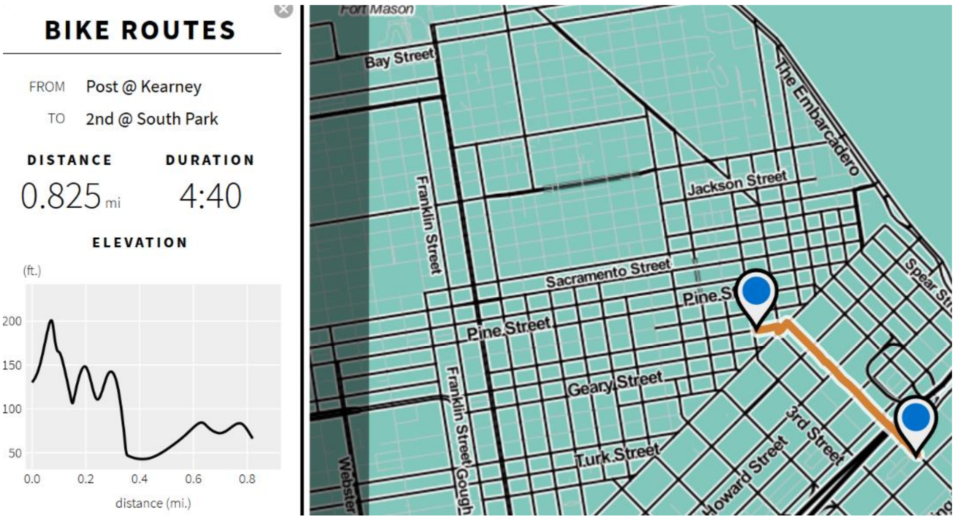

Fig 1.4 Route detail between “Post at Kearney” and 2nd at South Park

The major traffic patterns and trends become apparent in the evolution of trip activity through time in [4] as they are presented by connected constellation maps, all-day trip counts through time are visualized by close encounters of the bird kind which is a wonderful rendering to reveal the detailed evolution of the station’s trip activity, as shown in fig 1.5, neat circles make way for intriguingly patterned blobs with intricate shapes that evoke butteries, birds of prey, or tree leaves.

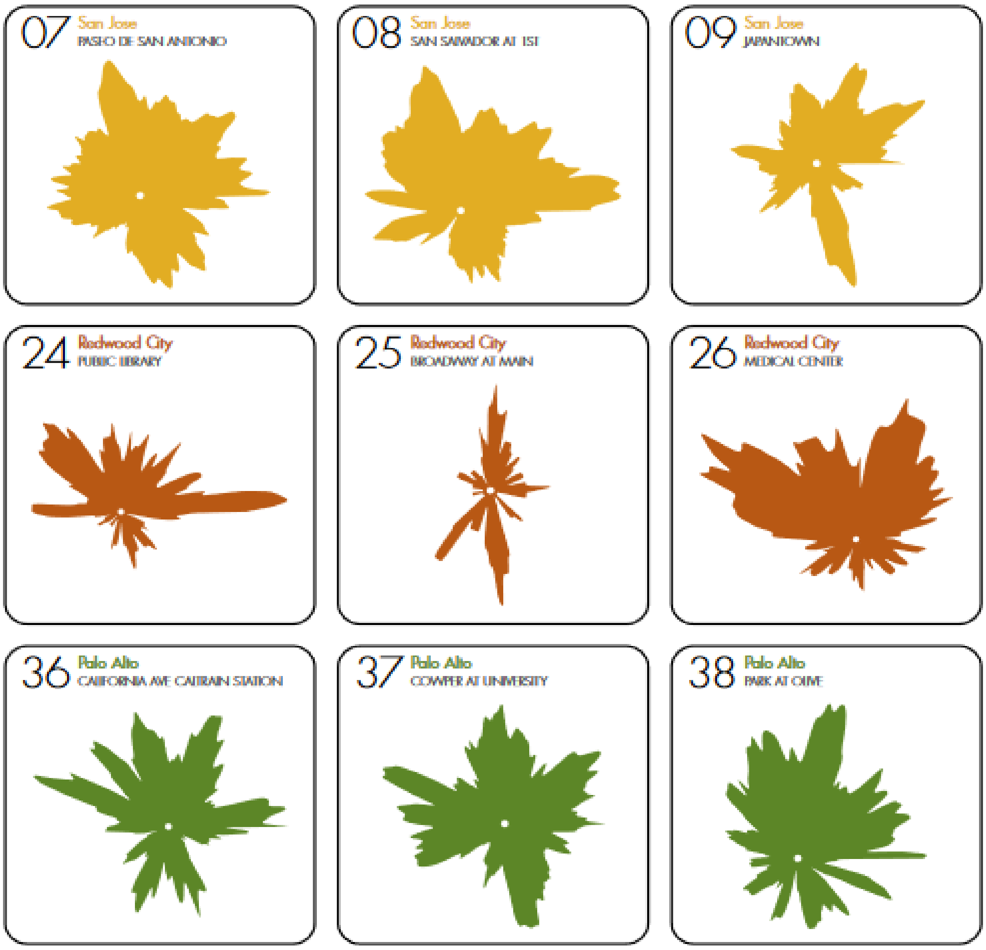

Fig 1.5 Traffic Pattern in bird-kind

[5] studies bicycle system in Washington which uses a colored mask matrix with 60 days data to reveal the masked bike demand of a single station, together with true demand plots they can achieve planning for rush-hour usage.

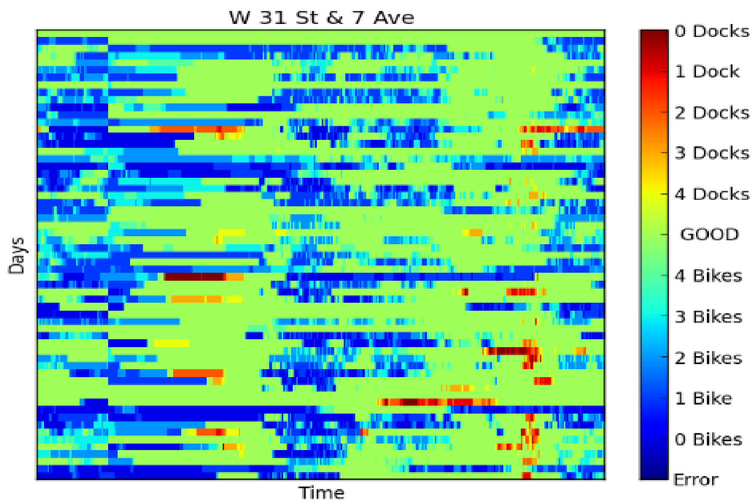

Fig 1.6 Plot of Masked Bike Demand

3. Design

3.1 Visualization Framework

The design of the project is straightforward, as demonstrated in fig 3.1, the original data need to be well processed to be explored and visualized. Thus, counting rides in daily and hourly perspectives, counting routes for different pair of stations, availability information for each station and geographic matching for all stations will be produced from raw data to feed the section of visualization and analysis. Analysis of processed data such as rides distribution throughout a day, popularity of stations and routes are necessary to get insight of the performance and operation of the bike sharing system, and it can also help to suggest what kind of visualization approaches to use to better present the data. Visualization methods are provided for both overviewed and detailed perspectives of the data.

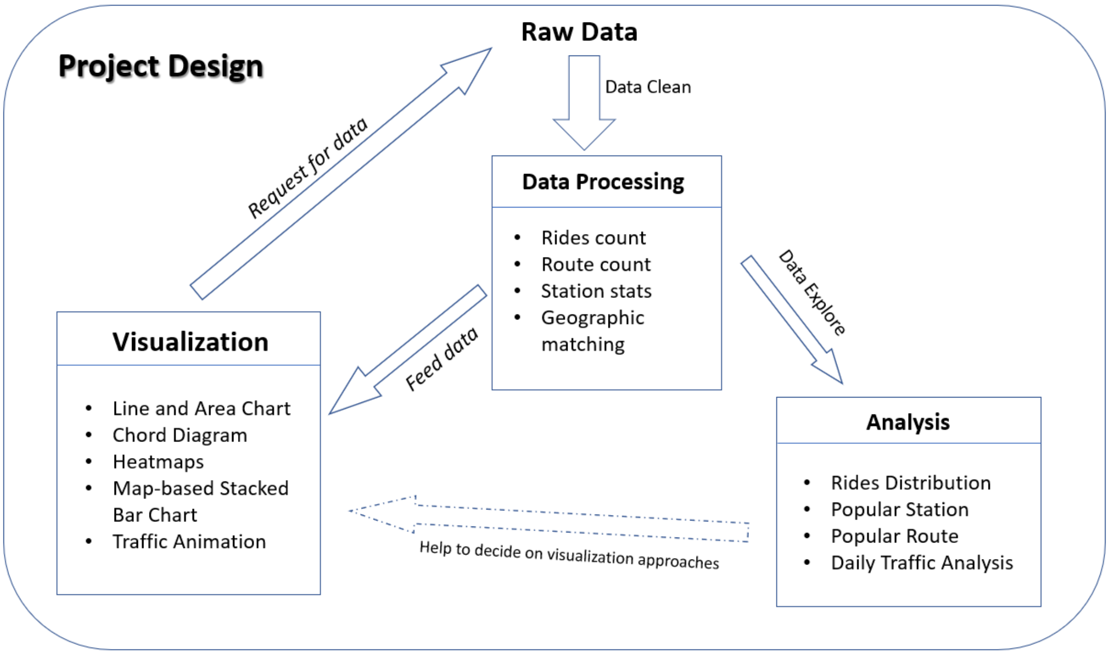

Fig 3.1 Project Structure

In overviewed perspective, a line chart is created to present total trips of all stations counted for each day, multiple lines (temperatures and moving average) are integrated into the chart while only trips line is filled with colored area for stressing. Filters are added to highlight trips during certain period of time such as weekends, holidays and bad weathers. The chart is interactive so that any lines or filters can be optional visible, in this case, the relations between total trips and different combination of conditions can be explored. Moreover, an annulus-based chord diagram is proposed in fig 3.1 to view route network between stations. An arc on the annulus represents a station whose radian size denotes number of out-going trips the station has. A ribbon connecting two stations represents a route between them, the width of an end of the ribbon indicates the outgoing trips from the station at that end to the station at the other end. The chord diagram will interactively hide routes of all other stations or all other routes when a specific station or a certain route is chosen (detailed usage is shown in following sections).

Fig 3.1 Chord Diagram for Visualizing Route network

In detailed perspective, two interactive heatmaps (for out-going and in-coming trips respectively) signifying trips distribution among different day of week for different stations are introduced. The scheme also treats different months separately. Moreover, a map-based interactive graph is used to integrate availability information with stations on the map graph. The “availability” above stands for available bikes and docks at different time of a day. In the graph, stations are marked with its popularity and average availability information (weekdays are separated with weekends) will be viewed when clicking a specific station. Last but not least, an animation is programmed to show daily traffic evolvement and traffic patterns. It selects top-6 popular stations as demonstration and uses bubbles as bikes travelling between those stations. A clock is running to reveal times that trips happen when a bubble travels from a staring station to a terminal station.

3.2 Analysis

Non-programming tools like Tableau and Plotly are used for data exploration and analysis.

In a general view, total number of daily trips of all stations appear to be periodically against time in fig 3.2, and the weekends are highlighted by grey shadows the total trips always drop down to the lower points among adjoining days, besides, holidays such as Thanksgiving Day, Christmas, and New Year are emphasized by light green shadow where total trips are the least. It seems that people in San Francisco have less preference to ride sharing bikes at weekends or in holiday when they are off duties. In fig 3.3, bad weather such as raining, fogging and winding are used as filter to select daily trips under such conditions. Bad weather conditions may happen at both weekdays and weekends, at weekends or holidays, the effect of weather may not be realized for the number of rides has already reduced a lot, however at weekdays, numbers of totals trips are large while trips in the days with bad weather seem to be relative lower. This means that bad weather condition did affect people’s choice on riding bikes, but not as much as weekends or holidays. Furthermore, temperature and moving average lines are integrated with “trips area” in fig 3.4. The moving average shows the rough trend of total trips across the whole year while temperature indicates a strong correlation with the trend. This suggests that in long term, temperature may have a big impact on people’s riding in San Francisco, September makes the most average daily trips with the relatively high temperature and January owns the least trips when the temperature is the lowest.

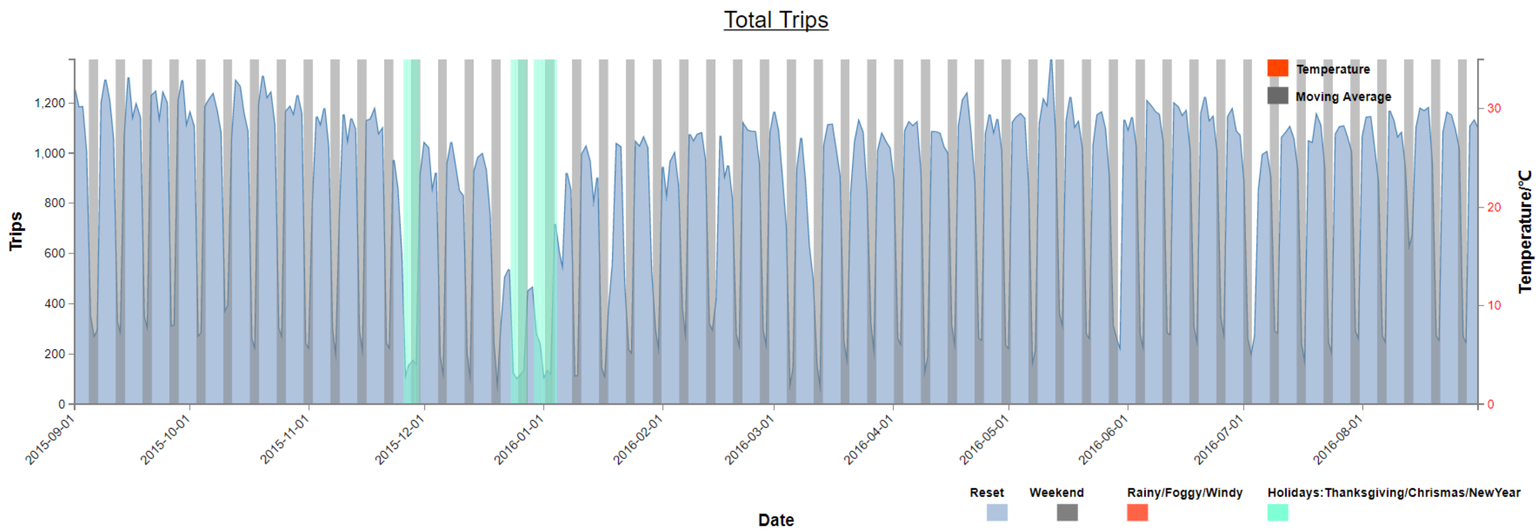

Fig 3.2 Total trips with highlighted weekends and holidays

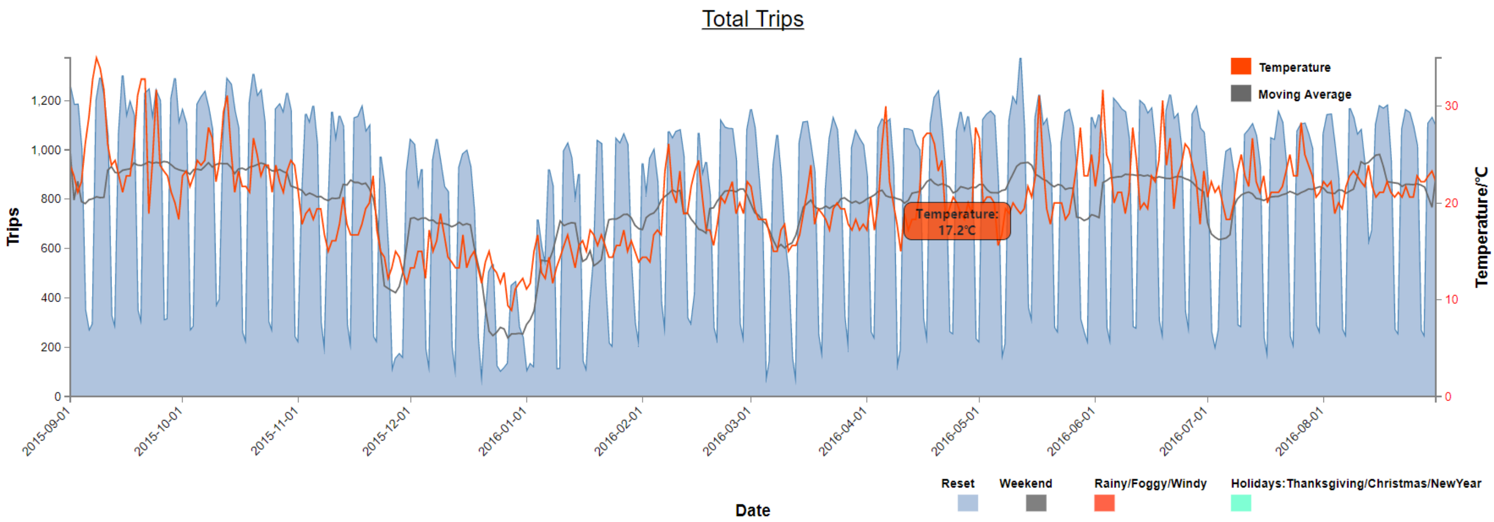

Fig 3.3 Total trips with highlighted weather conditions

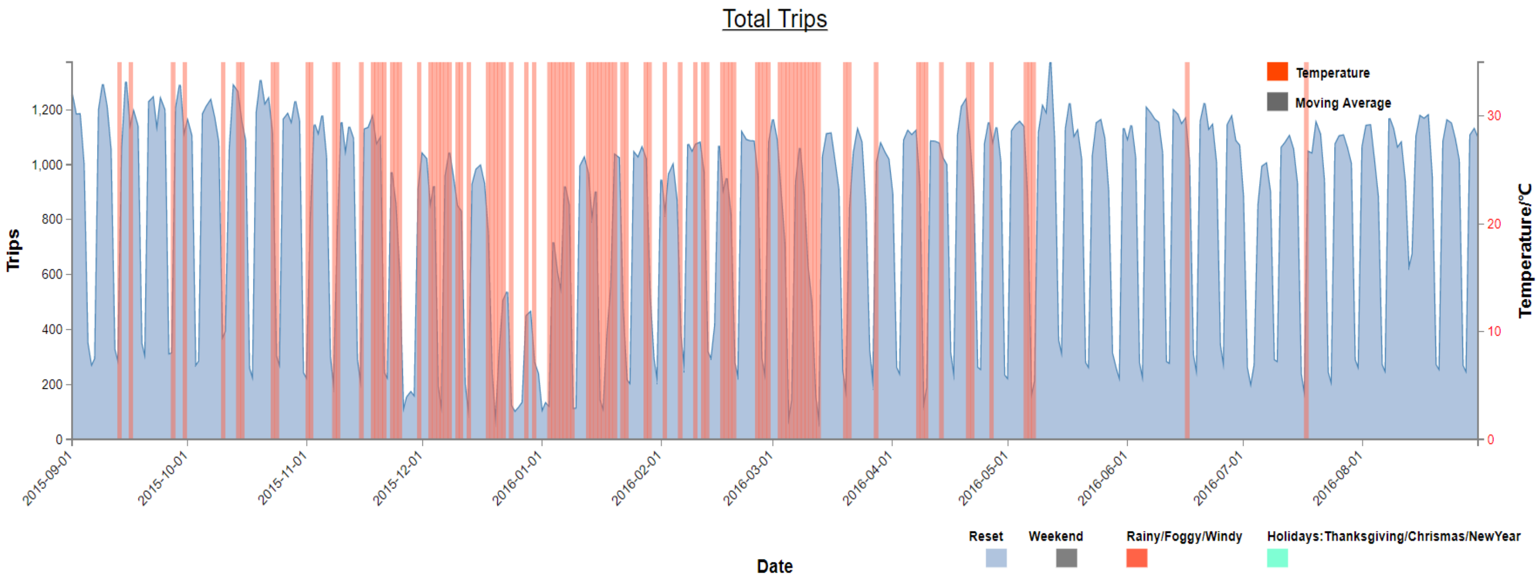

Fig 3.4 Total trips with temperature and moving average lines

When daily trips break down into hourly view, distributions of average trips that happen during a day's period for both weekday and weekend are illustrated in fig 3.5. Most rides at weekdays happen from 7am to 9am and 4pm to 6pm which are times of on-and-off duties, while people at weekends tend to ride sharing bikes between 10am to 5pm when people usually go out for leisure activities. Together with previous analysis, it is obvious that going for work is the main motivation for people to ride sharing bicycles.

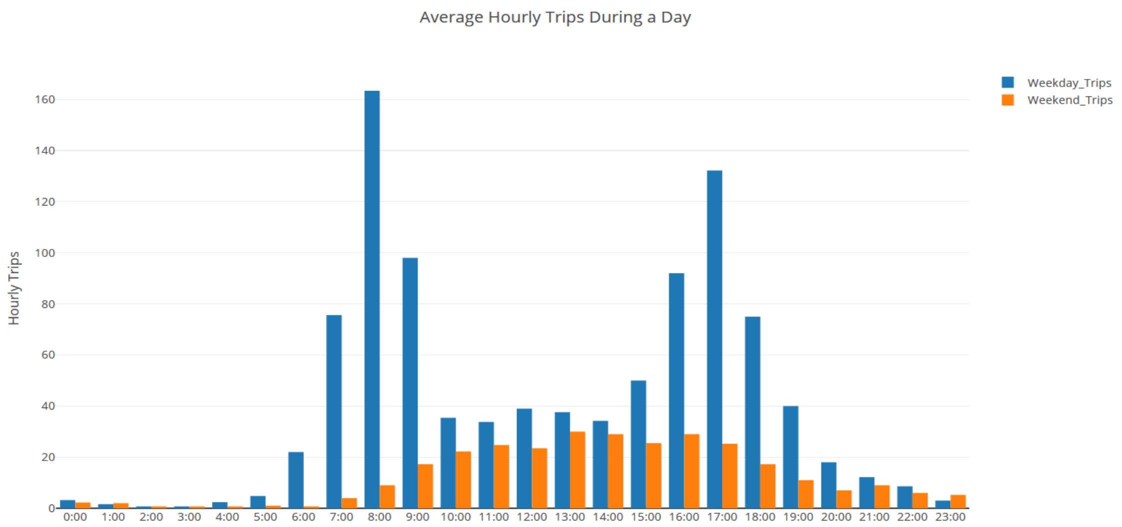

Fig 3.5 Distribution of Average Hourly Trips for all stations throughout a day

The popularity of stations in terms of out-going and in-coming traffic are shown in fig 3.6 and fig 3.7 respectively. The most popular stations as staring points and terminals are almost the same. “SF Caltrain”, “SF Caltrain 2” and “Embarcadero at Sansome” are stations near sea and wharf while “Harry Bridges Plaza” and “Steuart at Market” are close to the ferry building which is not just a terminal for ferries traveling across the Bay, but also a great spot for local eats and shopping.

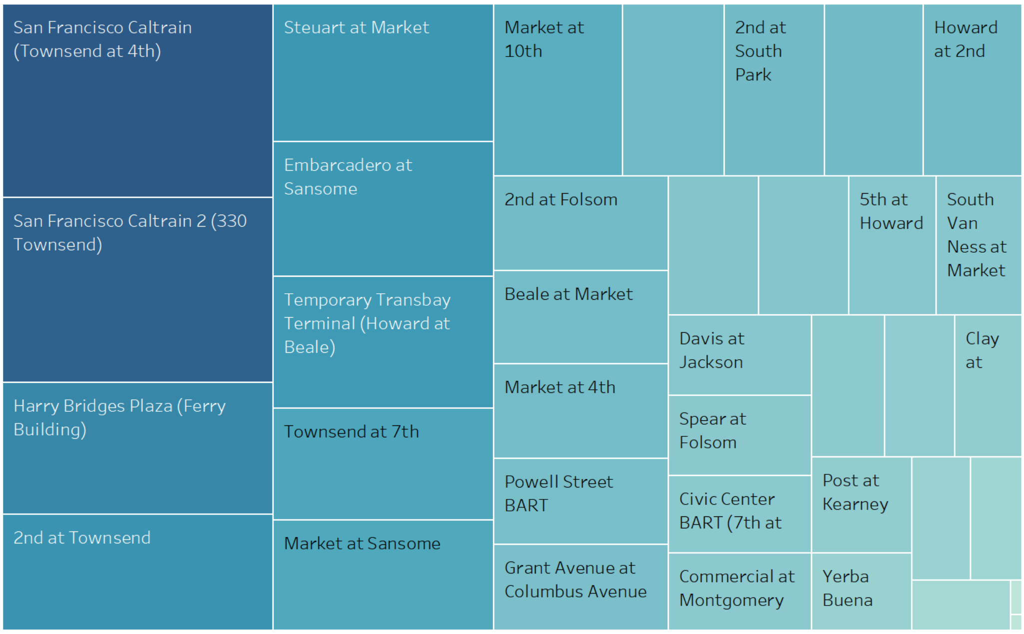

Fig 3.6 Station Popularity of Out-going Traffic

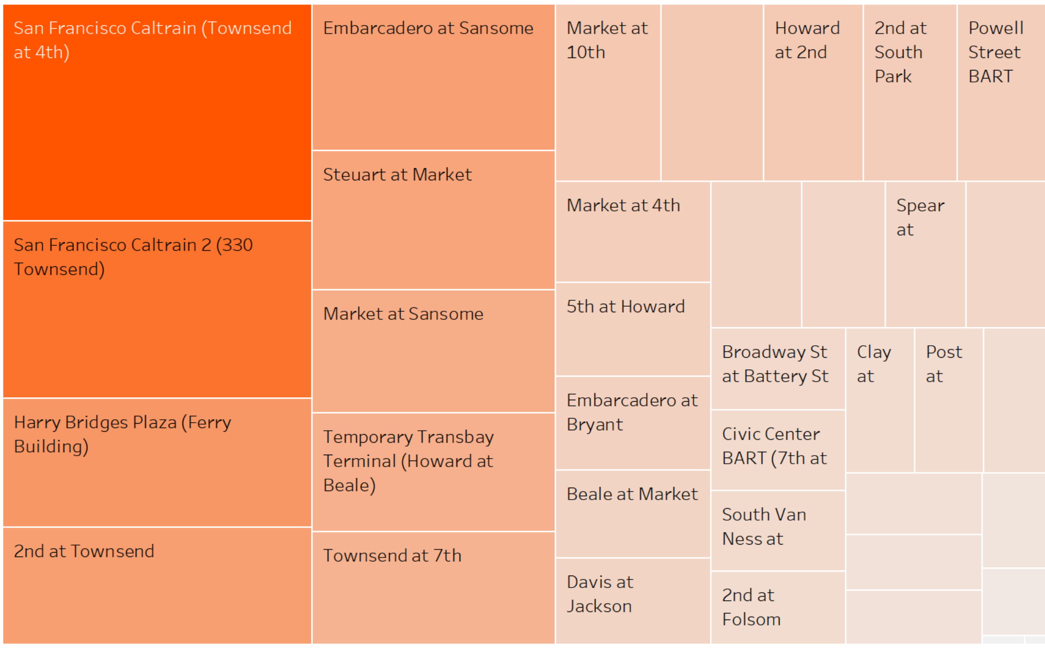

Fig 3.7 Station Popularity of In-coming Traffic

Besides, the top 10 popular routes are listed in table 3.1.

From	To	Number of Trips
Townsend at 7th	SF Caltrain2	3341
Market at 10th	SF Caltrain2	3035
Harry Bridges Plaza	Embarcadero at Sansome	2986
2nd at Townsend	Harry Bridges Plaza	2781
Embarcadero at Sansome	Steuart at Market	2605
SF Caltrain	Harry Bridges Plaza	2593
Harry Bridges Plaza	2nd at Townsend	2531
Steuart at Market	SF Caltrain	2482
Embarcadero at Folsom	SF Caltrain	2384
Harry Bridges Plaza	SF Caltrain	2362

Table 3.1 Top 10 Most Traveled Routes

In fig 3.8, 95% of the total rides last under 30 minutes during which one will only need to pay 3$ to ride, if one keeps a bike out for longer than 30 minutes at a time, it's an extra $3 per additional 15 minutes. The policy of the “Ford Gobike” is quite clear that a single ride is supposed to be kept in 30 to 45 minutes so that the whole system will run effectively and the city traffic will not be affected.

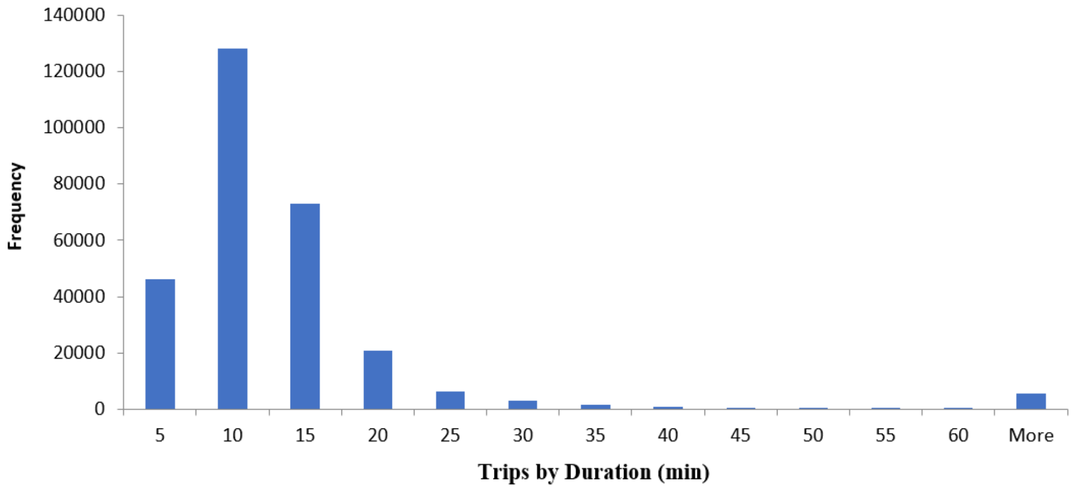

Fig 3.8 Distribution of Trip Duration

3.3 Visualization

On the route network in fig 3.9 of systemwide rides we can note that riders leaving from the most popular station, SF Caltrain, disperse throughout the system. Riders heading to SF Caltrain similarly tend to come from throughout the system.

The third-most ridden route, Harry Bridges Plaza (Ferry Building) to Embarcadero at Sansome in fig 3.10, does not have a return route with as great numbers, indicating that riders tended to ride to Embarcadero at Sansome and continue their journey elsewhere rather than return to the Ferry Building. Embarcadero at Sansome is the northernmost station along the Embarcadero, closest to tourism-heavy Pier 39 and Fisherman's Wharf. The bike path heading north along the Embarcadero is also much more bicycle-friendly than the southern route.

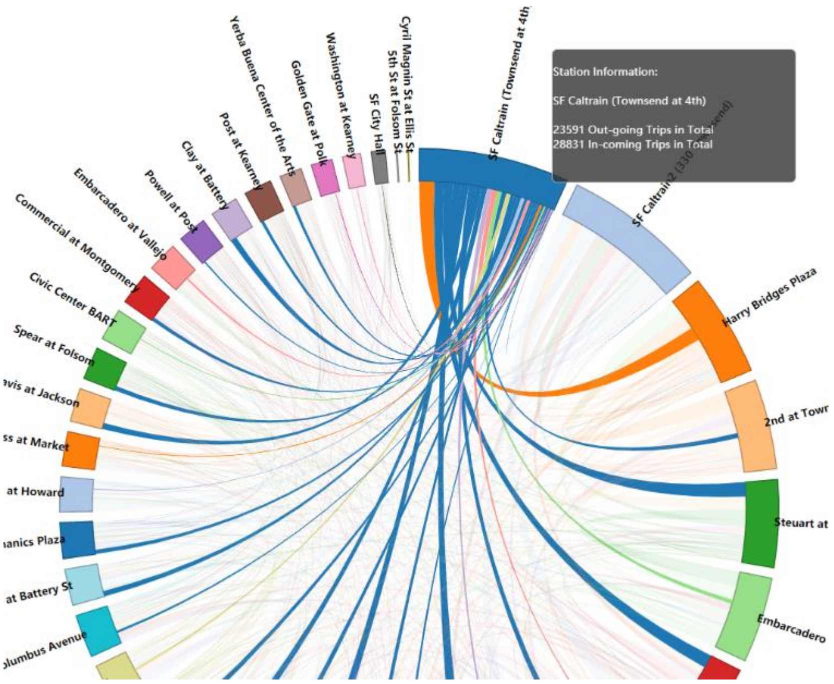

Fig 3.9 Chord diagram with full routes network of “SF Caltrain” Station

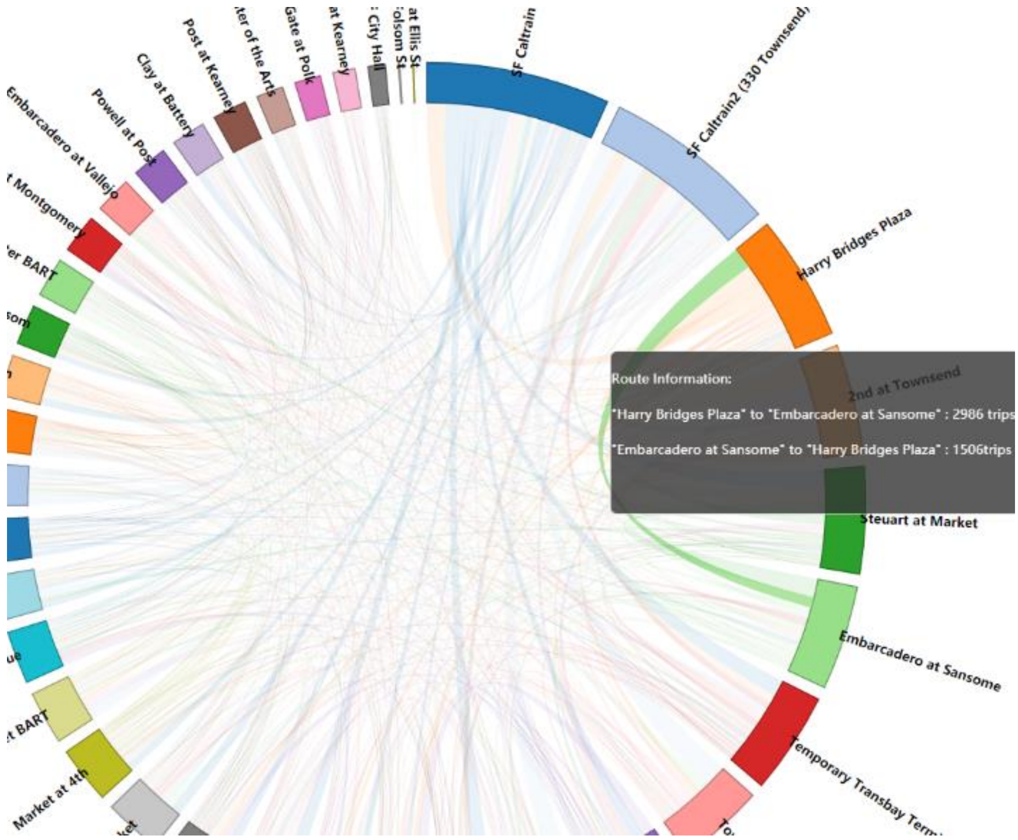

Fig 3.10 Chord Diagram with a single route highlighted

Through the monthly trip distribution, we can find that all stations have least rides at weekends for both out-going and in-coming traffic because people do not need to go to work, while Fridays always have less rides than other weekdays (as shown in fig 3.11 and fig 3.12, taking the data of December as example), so people may get tired and lazy after 5 days' hard working in which case some of them take taxi instead of riding bikes.

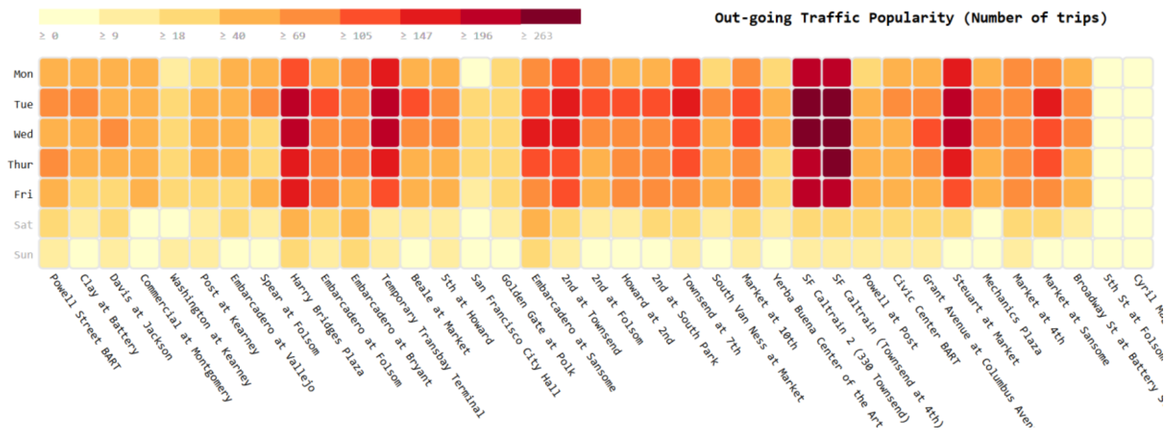

Fig 3.11 Monthly rides distribution of out-going traffic (December)

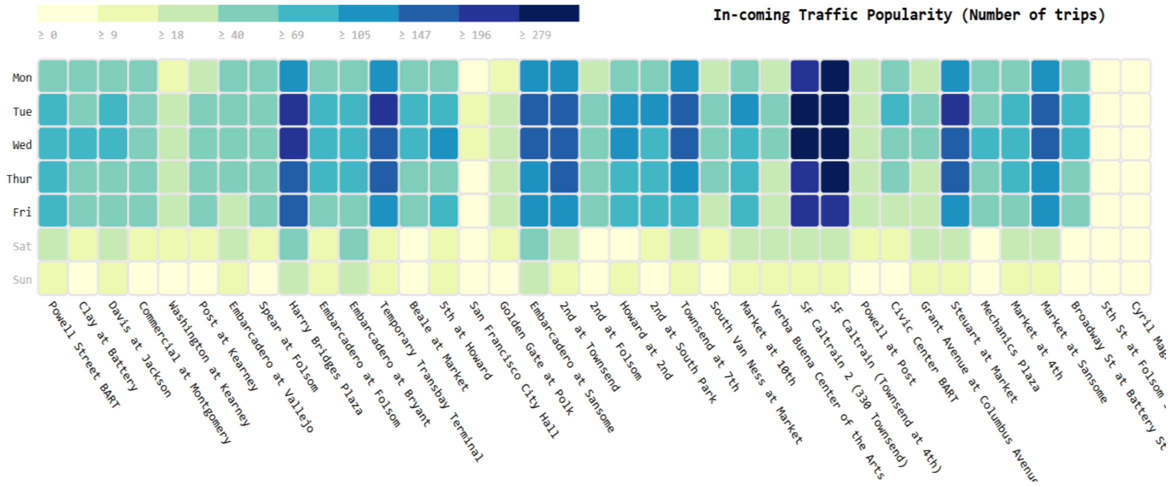

Fig 3.12 Monthly rides distribution of in-coming traffic (December)

As sharing bikes are mostly used for the short trips in the way on and off works, tracking the available bikes at different time for stations can lead to clear conclusion on which popular stations are more likely to be the terminal point near workspaces and verse vice. As shown in fig 3.13, “San Francisco Caltrain” is obviously a popular starting point in the morning when bikes are consumed quickly and are returned in the afternoon when people are off work. “Embarcadero at Vallejo” in fig 3.14, however, is a popular terminal near working space which people flood into in the morning and leave away in the afternoon. Besides, weekends of the stations share the same trend as weekdays with much smaller traffic volume.

Fig 3.13 Availability Information of “San Francisco Caltrain”

Fig 3.14 Availability Information of "Embarcadero at Vallejo"

4. Implementation

This section demonstrates how the visualization methods are realized.

4.1 System Architecture

The project contains 5 main parts of visualization methods that are from either overviewed or detailed perspectives, as shown in fig 4.1, each of the methods requires specific data to feed and specific data format to match. Thus, the they need to be produced from the raw data which are extracted from "Ford GoBike" system.

The raw data includes 4 parts - Station data describes station information such as station name, station id, latitude, longitude and dock number; Status data records available bikes and docks in every minute; Trip data contains details of all trips such as start and terminal stations, starting and ending time and bike id; Weather data records weather status, and temperature for every day. In next stage, they are combinedly processed and formatted according to desired visualization approaches.

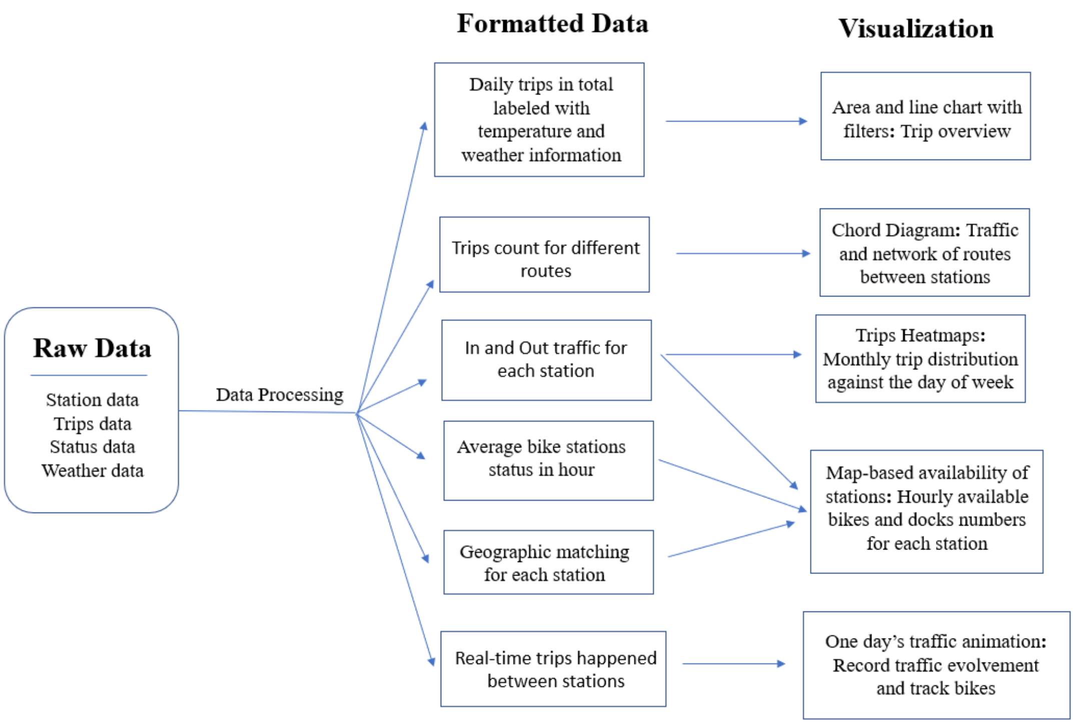

Fig 4.1 Flow Chat for the Whole Process

4.2 Data Processing

Python and R are used to clean and extract data, combine renamed stations, count trips, combine tables, filter condition, transform to specific format of data, the details are listed as following:

The data is in one year's period in which some of the stations may change locations, be renamed or be removed, so station information has been updated, combined and removed; Daily trips of all stations are counted for trip overview of the whole year, hourly trips of all stations are counted for trip view of a whole day; Total yearly trips are counted for each station as a start station which will reveal the popular starting points, and total yearly trips are counted for each station as an end station to reveal the popular ending points; Total yearly trips between each pair of stations are counted to for viewing route popularity, hourly trips between each pair of stations are counted to for demonstrating route information at different times in a day; A matrix-like structure is created to format yearly trips of all routes, namely, "Traffic Matrix" for further visualizations; Hourly and daily availability information which contains the number of available bikes and available docks for each stations are extracted, and in-and-out trips records are further produced by availability information for viewing accessing status for each station; Important weather factors like raining, fogging and winding are bound to daily trip data for further analysis.

4.3 GUI

D3.js and Leaflet.js are used on map-based availability graph, all other measures like area and line chart, trip heatmaps, chord diagram and traffic evolvement animation are built by D3.js.

In line chart of trip overview, information of week of day, weather condition, temperature, moving average and holidays are bound to trip data over time. Check boxes are added to control the visibility of those information while daily trips are indelible with tooltips to show details when mouseover.

In chord diagram, the clockwise order of arcs on the annulus follows the out-going popularity order of the stations while the clockwise order of ribbons connected to a station complies with the out-going popularity order of all routes starting from the station. For clarity, tooltips with station information and route information are bound to each arc (station) and ribbon (route) respectively.

In traffic popularity heatmaps, processed data need to be further extracted, combined and formatted to make a "day of week-against-stations" matrix for each month. After data are feed in the model, single color systems with varied depth (cool color and warm color system for in-coming and out-going traffic heatmaps respectively) are used for color scheme of heatmaps, and bin sizes of the color ranges are adjusted automatically to make obvious contrast for heatmaps.

In map-based availability graph, leaflet.js is used to integrate stations by their geographical coordinates into street map of San Francisco, each station is assigned a marker whose size is defined by popularity of corresponding station, stacked bar charts are created to illustrate average available bikes and docks at a same time in a day.

To track the real-time traffic of a day, forced bubble animation is driven by the real trips data and availability status data of 1st of September, 2015. As the animation only focus on the traffic among six stations, traffic with all other stations are excluded, so the bikes that are supposed to go to stations other than the six ones remain stayed at original stations, they will be assigned with a new destination among the six stations because they are the only bike resource to use without in-coming bikes from outside of the small traffic system in the animation. Also, tracking lines are created pointing to destinations when trips happen, they will be partially faded after certain time period.

5. **Evaluation**

This section will assess the visualization performance by solving some tasks which are defined to solve practical problems on route selection, station selection, resource allocation and policy-making.

5.1 Task 1

What are the top 3 routes for the most popular station as a start terminal and what are the proportion of trips the routes take up among all routes of the station? How about incoming trips from the destinations of the 3 routes?

From the Chord Diagram (fig 5.1), the most popular station as starting point which has the largest arc is "SF Caltrain (Townsend at 4th)", the 3 top routes are represented by the first 3 highlighted ribbons from left side; The destination of the routes are "Harry Bridge Plaza", "Temporary Transbay Terminal" and "Steuart at Market" respectively.

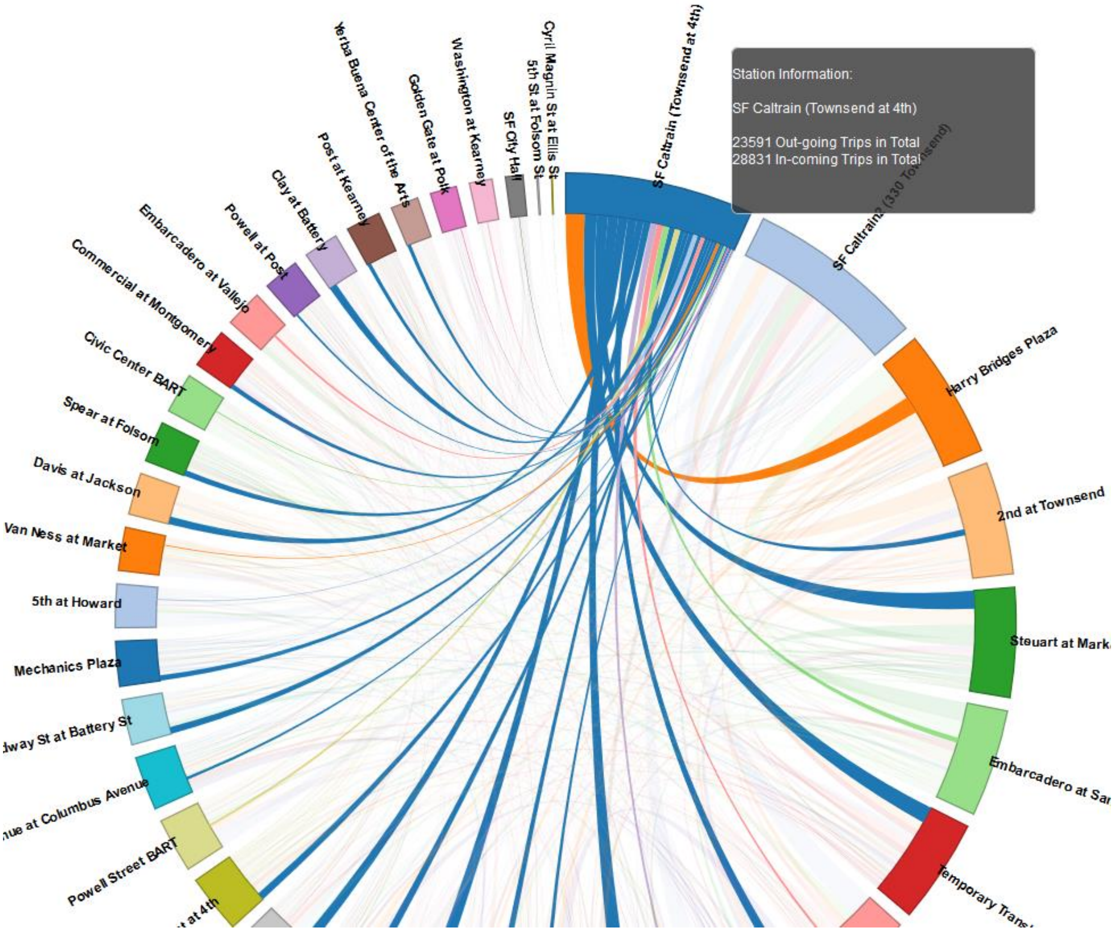

Fig 5.1 Route Pattern for "SF Caltrain" Station

Take the first route for example, as illustrated in fig 5.2, the number of out-going trips is 2593 which takes up 11% of total trips of "SF Caltrain" station. In-coming trips from "Harry Bridges Plaza" is 2362 that makes up 8.2% of the total in-coming traffics. Sharing the same concept, the second and the third route take up 7.1% and 6.4% of out-going traffic respectively while in-coming trips from "Temporary Transbay Terminal" and "Steuart at Market" make up for 7.8% and 8.6% of the total in-coming traffic respectively.

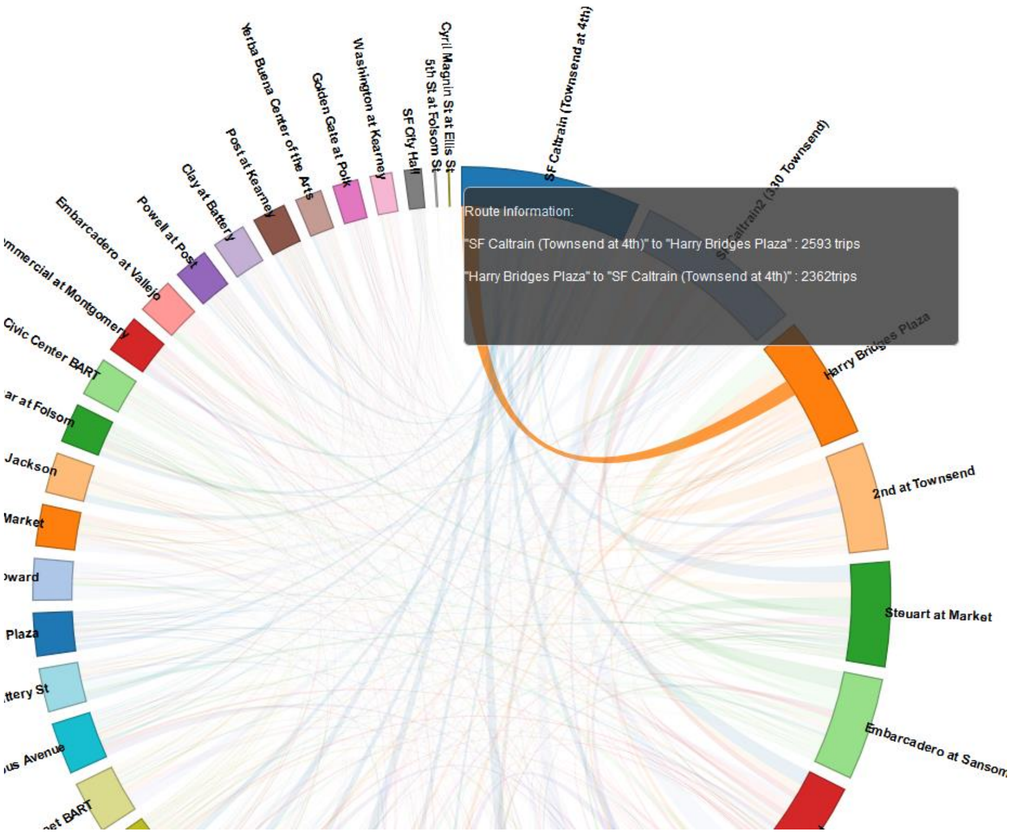

Fig 5.2 Top 1 route for “SF Caltrain” Station

5.2 Task 2

<u>Low temperature may lead to less rides, how it affects cycling traffic for different stations in different days of week?</u>

From the fig 5.3 we can find that there is a strong correlation between temperature line and moving average line of trips, so natural trend of temperature is the main reason for the long-term fluctuation of total rides. Specifically, September and January have the most and least daily trips respectively, the heatmaps for the two months are loaded to view the changes that temperature bring to the traffic. The station that are affected more are circled in fig 5.4 where the reduce of trips are reflected by color depth, besides, trip decrement usually happens at weekdays where working dominate the rides contribution.

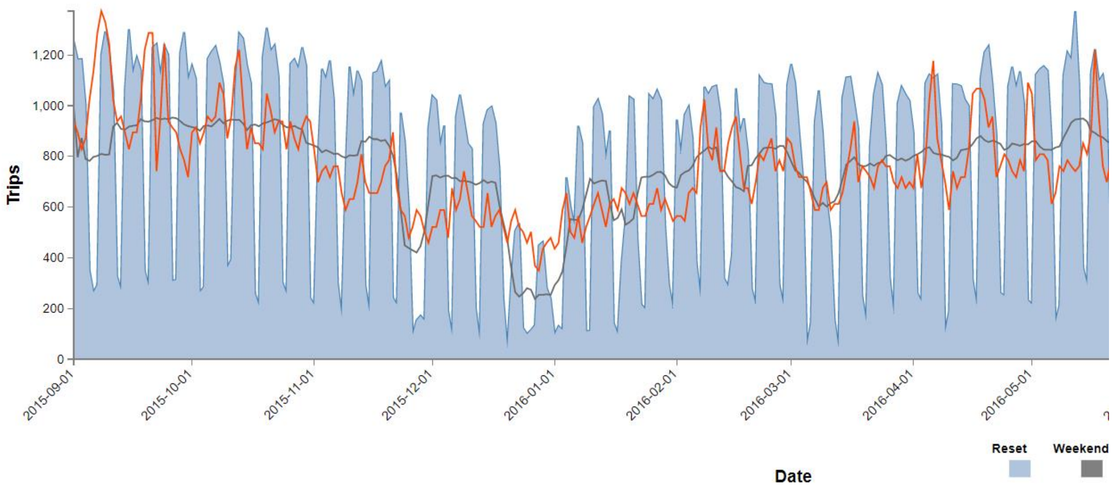

Fig 5.3 Correlations between Temperature and Total Trips

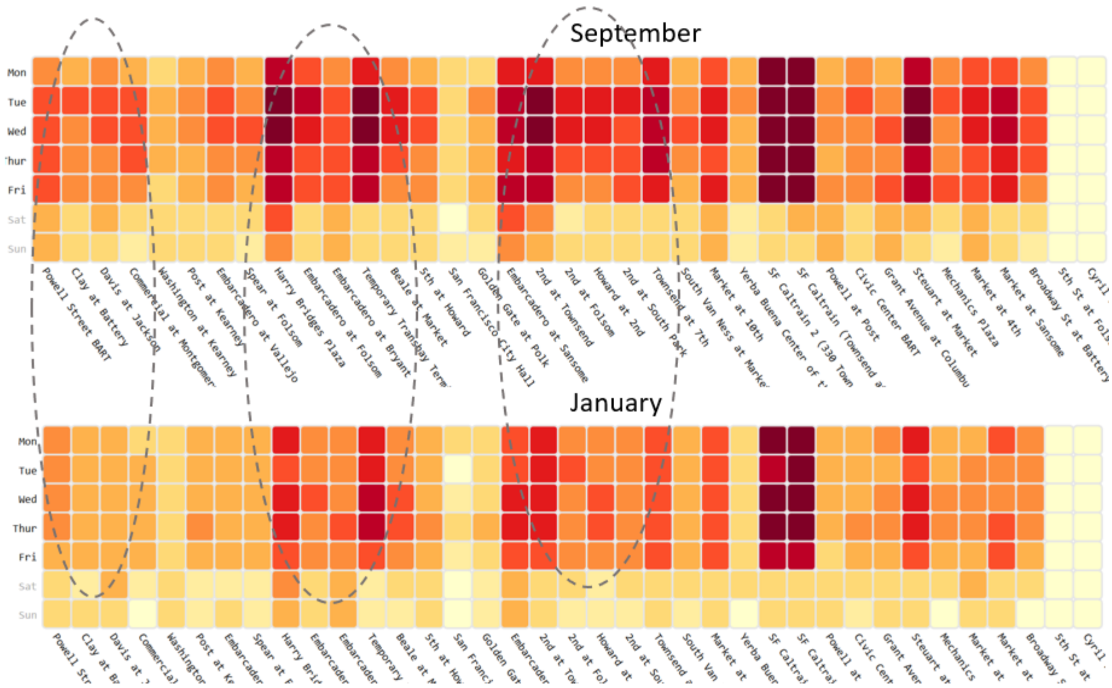

Fig 5.4 Heatmaps Comparison between September and January

5.3 Task 3

<u>Which stations need extra bikes supply and which stations can be bikes resources after one day’s traffic? How to plan the route of supply in terms of loading capacity of transfer vehicles and shortest route?</u>

A thorough search on availability information for each station is needed to find out which stations need bikes or need to unload bikes. 5am and 10pm are set to the start and the end time of a day for counting respectively, weekdays and weekends are treated separately. An index is introduced to describe the bikes imbalance of stations, as shown in the formula below, the difference of available bikes between 10pm and 5am of a day (demonstrated in fig 5.5) is counted for both weekdays and weekends of each station, and the difference is divided by the half of the sum of the available bikes of 10pm and 5am, where the AB_{10pm} and AB_{5am} stands for available bikes at 10pm and 5am respectively.

$$\text{Index} = (AB_{10pm} - AB_{5am})/[0.5*(AB_{10pm} + AB_{5am})]$$

We define the index above 0.05 to be a sign of bike shortage, and the index below -0.05 to be a sign of bikes excess.

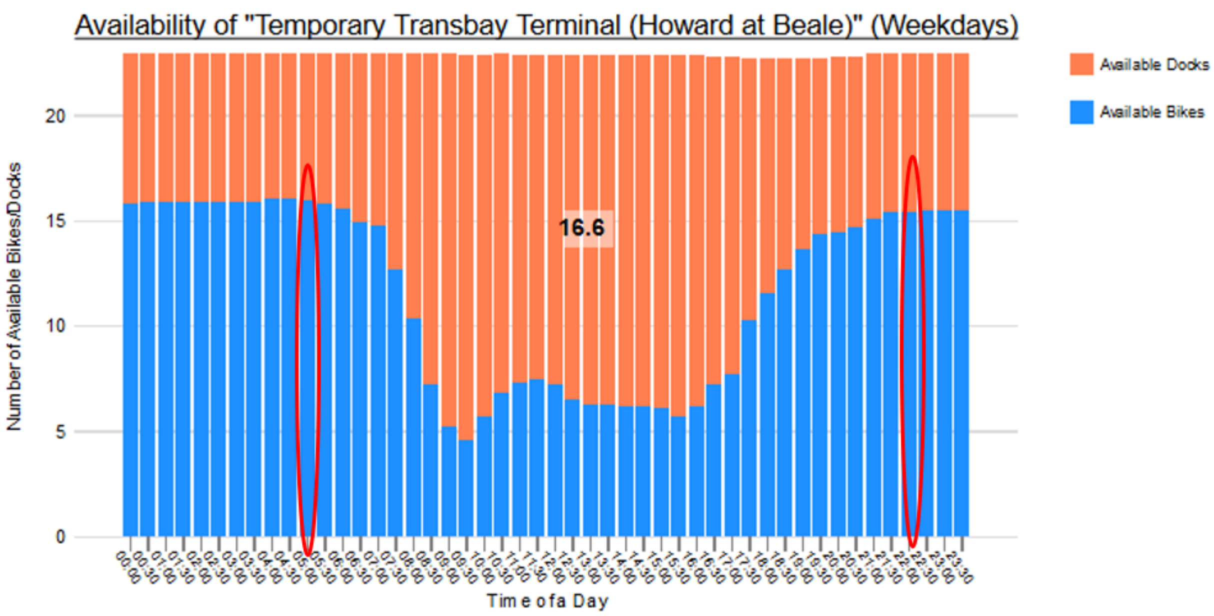

Fig 5.5

At weekdays in fig 5.6, “2nd at Folsom”, “Clay at Battery”, “5th at Howard”, “Davis at Jackson”, “Embarcadero at Sansome” and “Embarcadero at Bryant” are marked as stations with bikes shortage by red circle, while “Golden Gate at Polk”, “Temporary Transbay Terminal (Howard at Beale)”, “Steuart at Market”, “Commercial at Montgomery”, “Spear at Folsom” and “Grant Avenue at Columbus Avenue” are marked as stations with bikes excess by black rectangular.

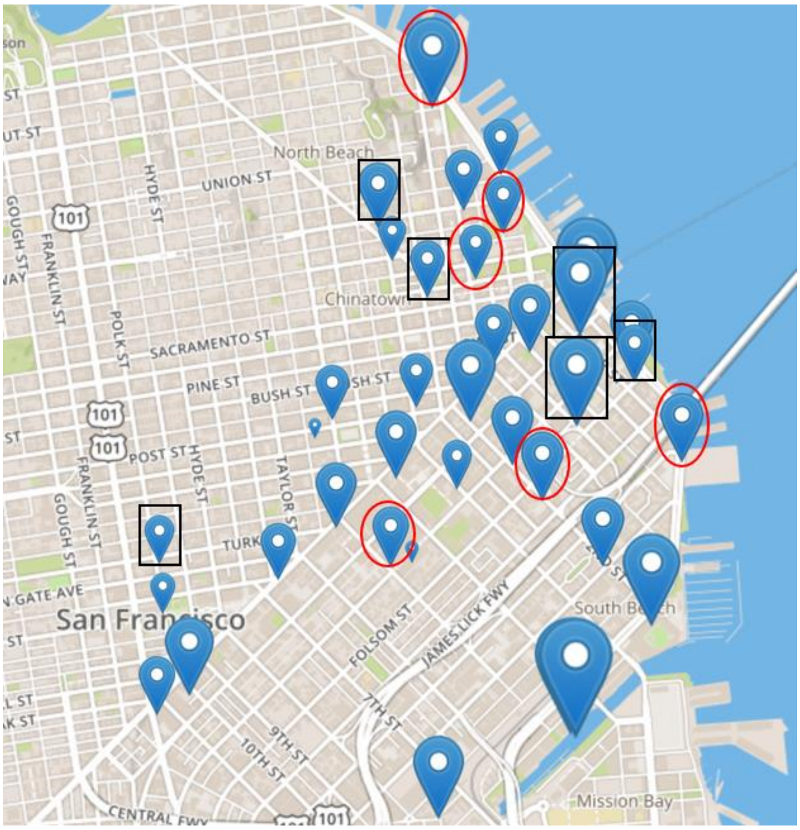

Fig 5.6 Weekdays

At weekends in fig 5.7, “Beale at Market”, “South Van Ness at Market”, “Harry Bridges Plaza”, “Market at 4th”, “Golden Gate at Polk”, “Embarcadero at Bryant”, “Grant Avenue at Columbus Avenue” are marked stations with bikes shortage by red circle, while “2nd at Townsend”, “Mechanics Plaza (Market at Battery)”, “Davis at Jackson”, “Washington at Kearney”, “2nd at Folsom”, “5th at Howard”, “Embarcadero at Sansome” and “Townsend at 7th” are marked as stations with bikes excess by black rectangular.

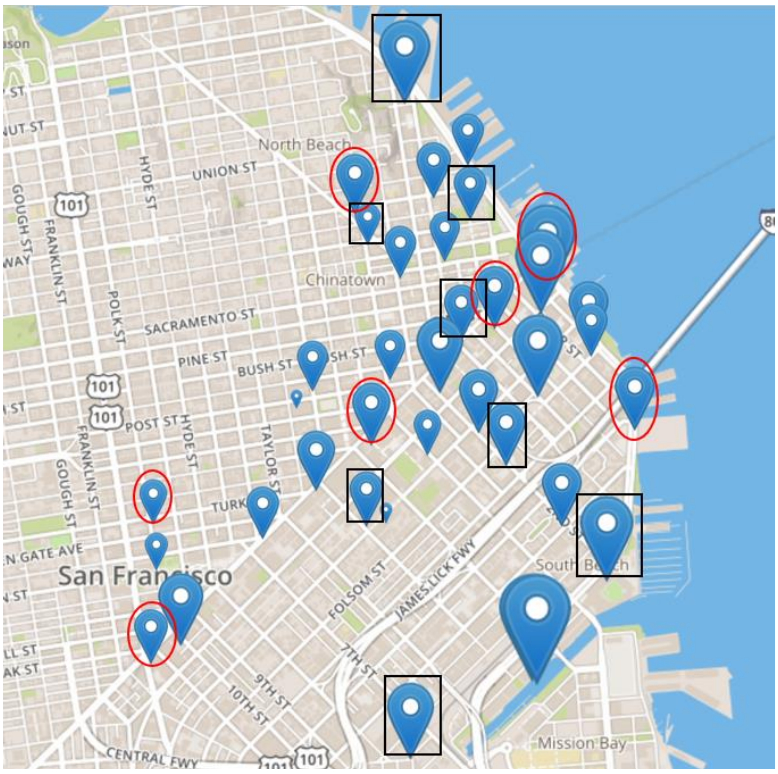

Fig 5.7 Weekends

5.4 Task 4

<u>At what time the bikes from other stations can make up around 1/3 of number of total bikes in "Steuart at Market" station?</u>

In animation fig 5.8, this task can be easily solved by watching out the real-time proportion of different colors of bubbles in "Steuart at Market" station.

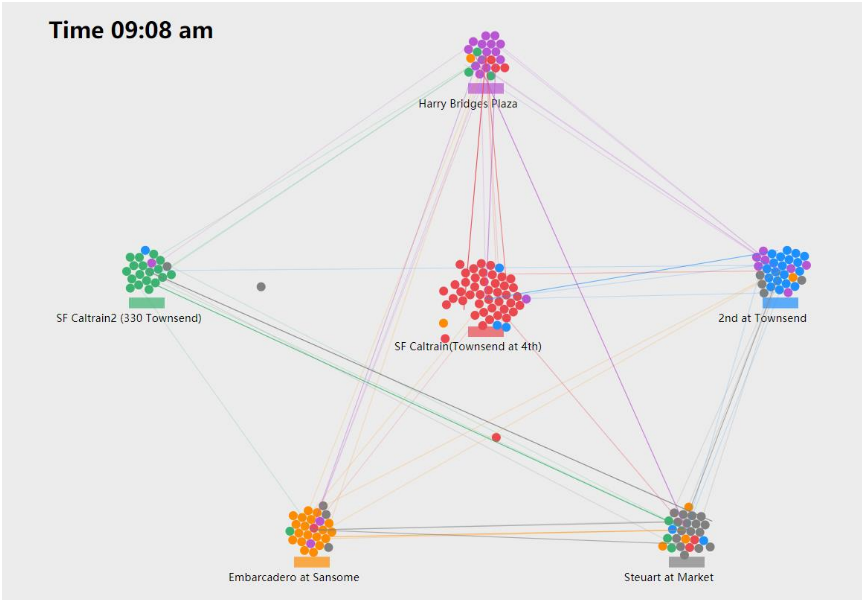

Fig 5.8 Traffic Pattern at 9:08am on 1st Sept 2015

5.5 Task 5

<u>Which pairs of stations have most trips between 7:33am and 8:33am?</u>

The program can be adjusted to let track lines fade completely after 60 minutes in fig 5.9, so lines displayed at 8:33am reveals all trips in the last 60 minutes, among this time period the most popular route is from “Harry Bridges Plazza” to “2nd at Townsend”. More importantly, in this manner, traffic network patterns in any time period can be viewed clearly, and the fig 5.10 shows the full traffic pattern of the day after the animation comes to a close.

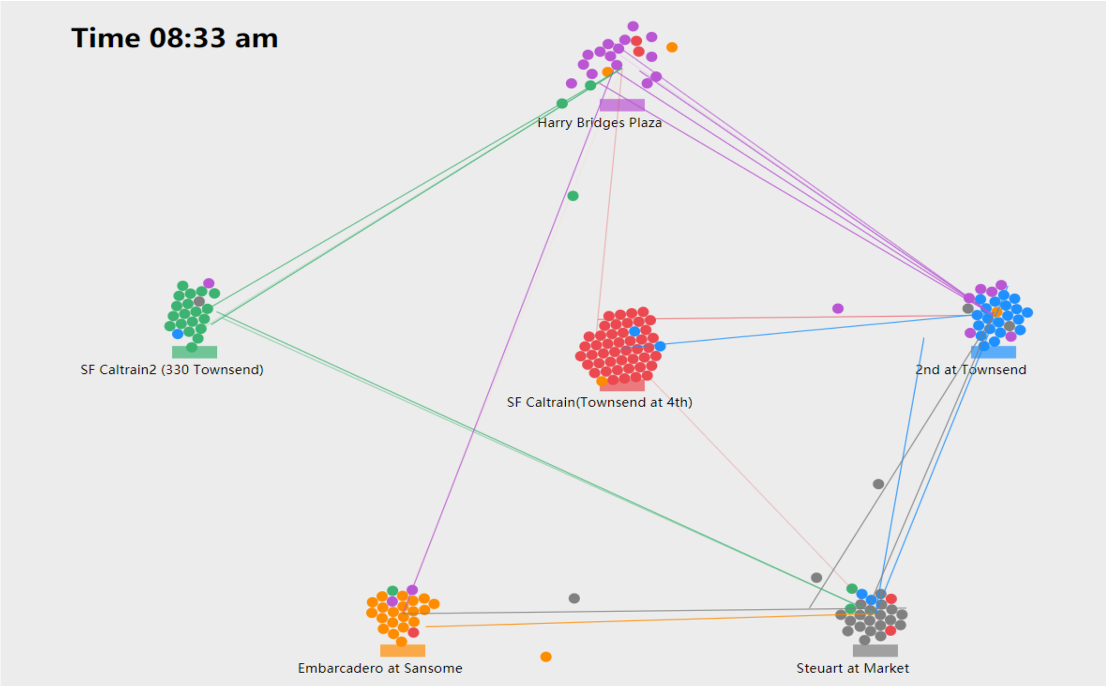

Fig 5.9 Traffic Patterns between 7:33am and 8:33am

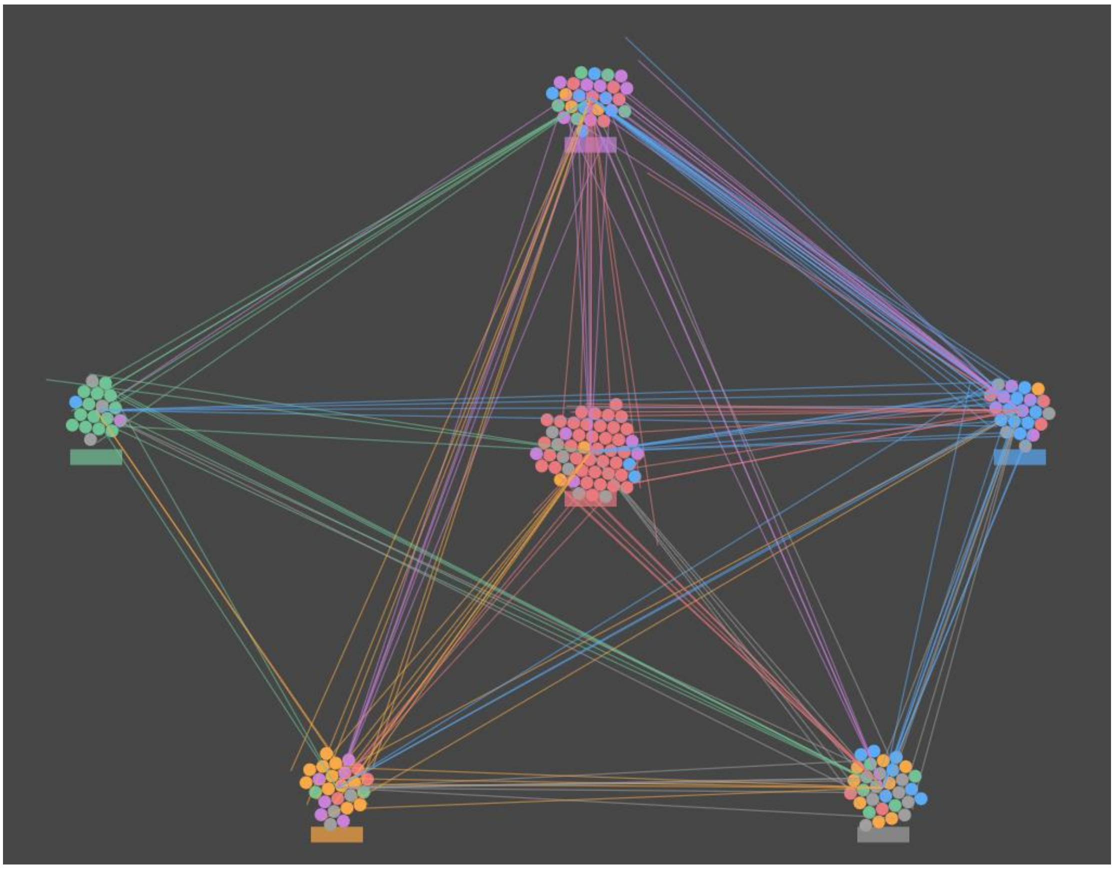

Fig 5.10 Full Traffic Pattern of a Day

6. **Conclusion**

This section includes summaries, limitation and future work of this project

6.1 Summary

The project has shown different measures to visualize both overview and details of the performance and operation of the bike sharing system of San Francisco. In general, people ride sharing bikes mostly for going work and the trips happens more on weekdays than on weekends, the most popular stations are “San Francisco Caltrain (Townsend at 4th)”, “San Francisco Caltrain 2 (330 Townsend)”, “Harry Bridges Plaza (Ferry Building)”, “Steuart at Market” and “Embarcadero at Sansome”. The whole visualization schemes are built interactively in which users can interact with data to better understand the data.

6.2 Limitations

The traffic animation is only a prototype that involves only 6 of the stations, it can be extended to the whole bike system to view evolvement of full traffic. The visualization schemes are only pieces that cannot interact with each other well, so the project cannot be regarded as a system and has limit to analyze complex problems that require different perspectives of data.

6.3 Future work

In task 3, the stations are only marked as shortage or excess, but the route for bike resource reallocation need to be done by optimal research in the future.

The interface for animation is not interactive enough to control the progress of the traffic evolvement in which case one can only adjust program to set fade time of tracking lines, so it is needed to be improved to better control the progress and the clock.

Cross filters should be added to compare data across different aspects of the system.

References

[1] Ford GoBike. (2017). *System Data | Ford GoBike*. [online] Available at: https://www.fordgobike.com/system-data [Accessed 1 Sep. 2017].

[2]Thfield.github.io. (2017). *BABS Data Challenge*. [online] Available at: http://thfield.github.io/babs [Accessed 1 Sep. 2017].

[3] Chiraphadhanakul, T. (2017). *BABS Data Challenge*. [online] Babs.virot.me. Available at: http://babs.virot.me [Accessed 1 Sep. 2017].

[4] Bay_area_bike_share_overview AUG2013-FEB2014. (2017). *Bjorn VERMEERSCH*. [online] Available at: http://www.bayareabikeshare.com/assets/pdf/Bjorn.pdf [Accessed 1 Sep. 2017].

[5] O'Mahony, E. and Shmoys, D.B., 2015, January. Data Analysis and Optimization for (Citi) Bike Sharing. In *AAAI* (pp. 687-694).

[6] Ford GoBike. (2017). *The Bay Area's bike share program | Ford GoBike*. [online] Available at: https://www.fordgobike.com [Accessed 1 Sep. 2017].

[7] Yau, N. (2017). *Make a Moving Bubbles Chart to Show Clustering and Distributions*. [online] FlowingData. Available at: https://flowingdata.com/2016/08/23/make-a-moving-bubbles-chart-to-show-clustering-and-distributions/ [Accessed 1 Sep. 2017].

Appendix

All the data and codes are shared at https://github.com/XintongChen1992/VA-for-SF-bike-sharing-system, instructions on how to interact with the visualization schemes will be added.