Metro Bike Usage Visualization

Yueqin Yang[1], Yiran Li[1], and Xiner Ning[1]

University of Southern California, Los Angeles CA 90007, USA

1 Introduction

The growth of bike sharing in the United States has had a trans formative impact on urban transportation. Over the next few years, Metro intends to blanket the city with bike share. By looking at our website, we want to show 1) there might be a good idea to add new stations at some area; 2) some extra work might be needed between some stations. In our website, we have combined the bike stations data with Metro bus lines, rail lines and big business regions in the Los Angeles. By looking at the combination map, we can easily see which area might need some new bike stations. Moreover, most stations usage increased during the summer time, and deceased during the winter. We have also used couple of charts to show the usages for each station duration the daytime and throughout the month. The difference amount between borrow and return for each station shows some stations have a large borrow amount but a very low return amount, then an extra work might be involved in order to keep the station work properly.

2 Approach

2.1 Design Considerations

Before we process the data, we have used Python and R to try to draw the charts first, to test if the graph is making any sense. We also looked some related works on bike sharing, to see what data and charts they used in their researches. Ranjit used lots of line charts and bar charts in his "Evaluation Study of the Bike Share Program in Fargo, North Dakota[3]". In this paper, s/he used bar charts to show the station data, and line charts to show the time data, which gives us a huge inspiration about what kind of data will be relevant for our project. After we have tried a lot of different charts with different axis, we found out some charts fits our story the most: line chart, bar chart and pie chart.

2.2 Technical Considerations

In order to build a nice website, choosing a nice framework is important. We decided to use Vue [1] as our framework. The nice thing about Vue is that it allows us to use JavaScript directly and maintains isolation between components as well.

[1] Vue.js is an open-source JavaScript framework for building user interfaces and single-page applications. https://vuejs.org

We start built the frame by following the guide on the Vue.js office website. The front end is based on Bootstrap which is a responsive grid based framework. Equipping with Bootstrap, our layout is automatically responsive although we still have to do tons of work to adjust the graph based on the size of the window.

2.3 Development

In order to work in parallel, we have divided our website into several components, e.g. Dot map in D3, mapbox map, line charts etc. After having those components defined, each team member takes one or more components. Technically speaking, each component are in MVC structure. For view part, we use bootstrap to make the basic layout responsive. However, it is not sufficient since we also want to change the size of the chart correspondingly. To achieve that, we need to use our controller to redraw the graph, which will be explained later. In order to make our layout clearly, we put the general information on the top of the page and the details are below the general information. The colors of the graphs are also well considered. We try to make the color as clear as possible.

Our team made a very clear plan for this project and follow along. This list gives a summary of what we did every week throughout the semester.

- Week 5: topic choosing / requirement analysis
- Week 6: framework / prototype on whiteboard / proposal presentation
- Week 7,8: Excel prototype / feature engineering
- Week 9: workload distribution / implementation
- Week 10,11: implementation
- Week 12: integration
- Week 13: paper working / summarizing
- Week 14: final webpage / video / final presentation

2.4 Evaluation

Our project is a good one for showing the station usage and optimize the bike usage. By using our website, the company can encourage more people riding bike, which is a very healthy hobby and it can also postpone the global warming.

Visualization Wheel

3 Data Processing

3.1 Data Source

We get our data most from the Metro Bike website [2]. The data is from July 2017 to June 2018. Each .csv file contains data for one quarter of the year. We

[2] https://bikeshare.metro.net/about/data/

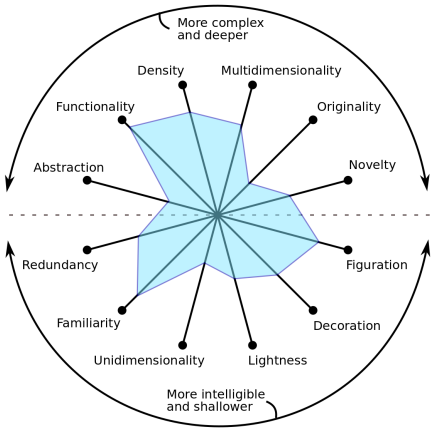

Fig. 1. Visualization wheel for our project.

also got our geojson there, which includes all the station location, makes easier for us to point out the location on the map.

Besides of the bike data, we also fetch the metro bus line and rail line's geojson data of LA country from github as well as the information of shopping malls from LA County open data.

3.2 Data Formatting

After download all the csv files, we used Python Pandas [3] to merge the files together. The data from Metro bike is inconsistent because they change some field names last year. Besides, the data format for date is also inconsistent, which leads to panic in our program. As the result, our first job is to make the data consistent. In particular, it offers data structures and operations for manipulating numerical tables and time series.

3.3 Data Processing using MySql Server

After deciding what kind of charts and data we need. We started process the data using MySql Server.

We split startstation and end station into two different table, and count the usage by stationid for every hour. Then we pivoted the table to stationid and 24 hours columns, which will be a lot easier to use for drawing the d3 chart later on. By grouping the same starting or end date for the same station and then take the different of these two table, we got the data for the positive and negative bar chart. Moreover, we also grouped data by region and got the data used for the pie chart and line chart.

[3] Pandas stands for Python Data Analysis Library. It is free software released under the three-clause BSD license.

4 Story and Solutions

4.1 Story

Our goal is to show does the metro bike station set properly, and does it need more stations in some regions? Since this course is not about machine learning or data mining, we do not go too deep to process the data.

By station is set properly or not, we mean some stations might have a large borrow amount and a very low return, but some other station might have a larger return than borrow amount. So is there a solution to balance these issues?

By setting more stations in some regions, we mean by looking at the station setting Metro Bike has right now, a large amount of area in Los Angeles County has not been covered. So is there a market for Metro Bike or other bike sharing companies to set their bike station in those regions?

4.2 Solutions

d3 Map In article "Optimizing the location of stations in bike-sharing programs: A GIS approach[2]", Juan mentions that set the bike station related to potential demand like population, activities and public transport stations is a key factor for one station to be success.

By using this graph, we want our user have an idea about the distribution of the stations (see Fig. 2).

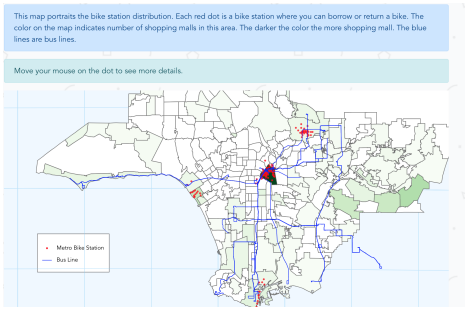

Fig. 2. : This map portraits the bike station distribution.

Pie Chart Bike usage distribution: average bike used of each region in every month. We want to show our user the popularity of each area.

Line Chart Bike usage trend along months: bike usage in summer months is significantly higher than winter months.

Mapbox and Bar Charts We find that stations adjacent to transit hubs receive disproportionate amounts of re-balancing relative to trips and that re-balancing is more often responding to morning and afternoon demand exceeding station dock capacities rather than longer term accumulations of bicycles. Cyrille points out a re-balancing patterns from data extracted for individual stations in "Bike-share re-balancing strategies, patterns, and purpose[1]", which is called bicycle sharing system (BSS).

This graph shows the relationship between time and bike borrow and return amount(see Fig. 3).The bar chart will be charged according to the station user selected on the map.

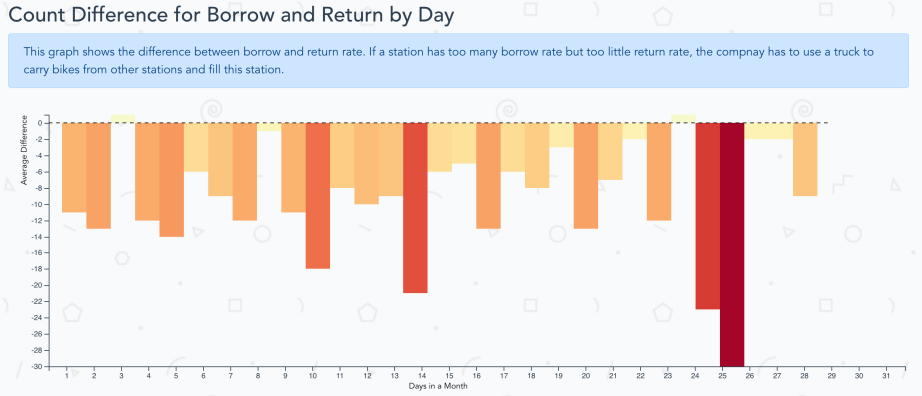

Fig. 3.

5 Conclusion

Our project results showed that the most of the stations are very suitable while some new stations should be built on the eastern part of LA county because there are some shopping malls. Besides, we find some stations have more borrow than return. We have two solutions dealing with the problem. First one is to use more trucks to carry the extra bikes to the insufficient station. The other solution is to change the business model and use different price for one-way and round trip fare.

References

1. de Chardon, C.M., Caruso, G.: Estimating bike-share trips using station level data. Transportation Research Part B: Methodological **78**, 260–279 (2015)
2. García-Palomares, J.C., Gutiérrez, J., Latorre, M.: Optimizing the location of stations in bike-sharing programs: A gis approach. Applied Geography **35**(1-2), 235–246 (2012)
3. Godavarthy, R., Mattson, J., Taleqani, A.R.: Evaluation study of the bike share program in fargo, north dakota. Tech. rep. (2017)