Citi Bike Data Visualization

Jing Zhang, Tianyang Li, Ziwei Yuan, and Tong Lyu

University of Southern California, Los Angeles CA 900007, USA

Abstract. Analyze potential factors that may influence bike-sharing orders through data visualization.

Keywords: Sharing Bike · Visualization · D3

1 Introduction

1.1 Background

Sharing economy is popular in recent year. Bike-sharing, car-sharing and workspace-sharing are all eco-friendly measures that make full use of our resources. Bicycles are popular vehicles in our daily life, shared bicycles perfectly solves the problem of the last one mile in the city for people. It is also a healthy and eco-friendly means of transportation. In the meantime, lots of new companies which focus on sharing-economy are changing our life. Uber, Didi, OfO and Citi bike.

Citi Bike is the nation's largest bike share program, with 12,000 bikes and 750 stations across Manhattan, Brooklyn, Queens and Jersey City. It was designed for quick trips with convenience in mind, and its a fun and affordable way to get around town. If you are a visitor who want to travel NYC or local resident that always need short travel, then Citi bike is a perfect solution to meet your needs.

1.2 Motivation

Last month, we visited NYC and try to use Citi bike. There are more than four bike stations around our hotel within 2 blocks. But none of us can borrow a bike when we start to explore this city. The contradiction is there are more than 300 stations located at Manhattan and thousands of bikes distributed to different stations. User still cannot borrow bikes at anywhere and anytime. That is a pain point for most of their target users. So as their target user, we decided to find potential factors that may influence orders and location of stations.

We assume some potential factors before we start to gather data. That is stations nearby infrastructure, weather, seasons, users portrait and borrow/return hours. Due to data limitation, we just get users birth year and gender. Thus we focused on analyze users age distribution of each station.

Also, there are lots of data dimensions may influence the site of stations, number of orders and other data. We would like to analyze different data to find potential factors. And help bike-sharing company make decisions.

1.3 Target User

Citi Bike company There exists some problems of bike distribution. Our project analyzed potential factors and focus on finding relationship between these factors and Citi company can get uses laws and suggestions from our website. Also, if Citi company want to expand their business on new cities or set more new stations, our websites can give them ideas.

Bike user User need to know when there has available bikes at stations. And our website can tell users when there has available bikes base on history data.

Competitor of Citi bike company Our data from Citi bikes sharing website. So their competitor can make their plan by looking our website.

2 Data

2.1 Citi Bike Trip Histories

Citi Bike Trip History data is offered by Citi Bike website [1]. The dataset is about trip information and is summarized for each month from 2013 to now, including Start Time and Date, End Time and Date, Start Station Name, Start Station Name, Station Lat/Long, Station ID, User Year of Birth, etc. We downloaded trip data of June from 2013 to 2018. We processed the original data and generated Station GeoJSON data of 6 years, total number of borrowed bikes per hour of each station for June of each year, total number of returned bikes per hour of each station for June of each year and user age summary. Here we use the birth year data to calculate the total using in each age group. Fig. 1 gives a sample of data.


```
{
  "type": "FeatureCollection",
  "features": [{
    "type":"Feature",
    "properties":{
      "addr":"Broadway & W 24 St",
      "id":"444",
      "use":"9546"
      },
    "geometry":{
      "type":"Point",
      "coordinates":[-73.98915076,40.7423543]
    }
  }]
}
```


Station Id	0	1	2	...	23
”444”	50	38	28	...	57
”406”	43	25	11	...	73
”474”	25	12	9	...	49

Fig. 1. Sample (Left: Station GeoJSON Data, Right: Borrowed Bikes Data)

2.2 NYC Facilities

The New York City facilities data has two sources.
The first one is NYC Capital Planning Platform [2] including various types of facilities such as core infrastructure, historical sites and parks. The format of the data is GeoJSON. The properties of datasets include Agency ID, Facility Name, Street Name. Address, City, Latitude and Longitude. We used historical sites, education and infrastructure datasets to show the surroundings of top 13 popular stations.
The second data sources is Google Map[]. We filtered the important infrastructures around several popular stations on the map and recorded their coordinates. In the data processing phase, we use the xlrd package in python to read the value of each item directly from the source file. Then we deleted the redundant data, fixed some of the wrong data, and unified the data format.

2.3 Precipitation Data

Precipitation data is offered by Kaggle [3]. The original data records the precipitation amount of 2016 by data. We used the data to show relationship between orders of sharing bikes and precipitation amount in 2016.

3 Approach

The project aims to provide activity statistics and variation visualization of sharing bike stations, and analyze the external factors influencing the popularity of stations.

3.1 Design

Component and Functionality The project contains two main components.

Stations Map The component is a map of all stations accompanied by station information including a summary of borrowed/returned bikes and a summary of annual variation in 6 years.

Stations Analysis The component combines station data with other datasets and presents station analysis in three different aspects, that is infrastructure effect for top 13 popular stations, uses age distribution of top 6 popular stations and the relationship between precipitation amount and number of orders.

In the infrastructure effect analysis page, there is a map which shows the top 13 popular stations and some of the important infrastructures around them.

In the weather effect analysis page, there is a interactive line chart which shows the relationship between precipitation and the number of orders.

In the weather effect analysis page, there is a donuts chart which shows the age composition of users of the top 6 popular stations.

Color Ware [4] suggests to use high saturation to code small areas such as symbols and lines, while background uses low saturation to attract attention on targets. We applied strong colors on SVG elements, highlight them and added pop-out effect when we needs to emphasize something.

Responsive We employed multiple methods mentioned by Gardner [5] to implement a responsive web design, such as fluid layout, flexible media and media queries. When the website is displayed on a mobile, information will not be lost too much.

Interactive We optimized the visual queries and user interaction by multiple methods such as pop out effect, coupling effects and appropriate color scheme.

3.2 Technical Considerations

Our technology stack includes:
Framework: Angular 6, Bootstrap 4
Language: HTML5, CSS3, JavaScript
Libraries: d3 version 5, Ant Design, MapBox-GL API
Package Manager: npm
Version control: Git

Maps and Charts Mapbox and d3.map is used to generate station map and top 10 popular station analysis map. d3.scale, d3 data enter, d3 shape, d3 transition and d3 events are used to show various charts.

Angular Router Bind the router to links on a page and it will navigate to the appropriate application view.

Angular Components Interaction A data service is created to pass values between sibling components.

Bootstrap Framework Bootstrap Grid System and components are used to build a responsive framework. Media queries with CSS restrictions like overflow and max-wdith helps to make the web contents responsive.

Event Listener Multiple event listeners are used. For example, a "resize" event listener is added to the window for implementation of responsive charts. To highlight or hover a target, mouse events are bound to relevant DOM elements using d3.selection.on() method and map.on() method.

3.3 Development

Angular Framework Take the following steps to develop the project:
1). Create the Web Application Architecture and Framework
2). Design the Data Structure
3). Develop / Customize the Web Application Module, Libraries and Classes
4.) Complete the Development and Implement all Functionalities - Version 1.0

3.4 Evaluation

Unit Test For each component, ensure the data pipeline goes well to show the result, test whether each chart displays correctly and supports interactive actions with users.

Integration Test Test whether navigation menu items navigate to the corresponding appropriate component view. Inspect the initialization state and rendering the web page in different conditions.

4 System

4.1 Stations Map - Station Activity

The sub-components represents borrowed bikes per hour bar chart and returned bikes per hour bar chart. Users select one year and pick any station on the station map to see two bar charts. Two charts adjust based on the size of the window. Users hover a bar and see the number of bikes in this hour interval for both charts.

4.2 Stations Map - Station Variation

4.3 Infrastructure analysis Page

In the infrastructure analysis page, users can quickly get know about some of the important or famous infrastructures around the popular sharing-bike stations. The map in this page displays the top thirteen popular stations by default. Each brown point represents a popular sharing-bike station. Users could see the district's name by hovering mouse on the map and could see the stations' address by hovering mouse on the station point. By clicking the station point, the whole chart will be zoomed in on this point. In that mode, the important infrastructures which includes park, subway station, shopping mall, hotel, school, sightseeing point and sports center will be shown as icons in different color. Meanwhile, it has a information box on the left side of page which counts the number of each infrastructure. In that case, users can clearly see the reason of that station being popular.

5 Related work

Other websites with the same topics offers a lot of statistics charts and spatial analysis. We referred to some of charts such as returned bikes and borrowed bikes per hour statistics bar charts on a sharing bike visualization website [6]. Most of relevant websites focuses on statistics and analysis. Their target users are tend to data analysts.

Our goal is to use simple statistics charts to tell a story about sharing bikes, showing the temporal and spatial variation of stations and the factors. Besides, our charts are responsive and interactive. We optimized the visual queries and user interaction by multiple methods such as pop out effect, coupling effect and appropriate color scheme.

6 Conclusions

References

1. Citi Bike Trip Histories Data, https://www.citibikenyc.com/system-data. Last accessed 29 Nov 2018
2. NYC Facilities Data, https://capitalplanning.nyc.gov/map/facilities. Last accessed 29 Nov 2018
3. Precipitation Data, https://www.kaggle.com/mathijs/weather-data-in-new-york-city-2016. Last accessed 29 Nov 2018
4. Ware, Colin. Visual thinking: For design. Elsevier, 2010.
5. Gardner, Brett S. "Responsive web design: Enriching the user experience." Sigma Journal: Inside the Digital Ecosystem 11.1 (2011): 13-19.
6. Sharing Bike Visualization Website, https://www.visualization.bike/citibike/system. Last accessed 29 Nov 2018