Citi Bike Information Visualization

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

University of Southern California, Los Angeles CA 900007, USA

Abstract. We used Angular, Bootstrap, d3, Mapbox to show various types of charts and maps. We analyzed potential factors that may influence bike-sharing orders through information visualization.

Keywords: Bike-Sharing · Visualization · D3

1 Introduction

1.1 Background and Motivation

Sharing economy is popular in recent years. 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. 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.

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.

1.2 Target User

Citi Bike company, Bike user and Competitor of Citi bike company

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. 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 is offered by NYC Capital Planning Platform [2]. The data is in GeoJSON format. The historical sites, education and infrastructure surrounding top 13 popular stations are selected.

2.3 Precipitation Data

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

3 Approach

3.1 Design

Component and Functionality The project contains two main components. Stations Map 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 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.

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

1). Framework: Angular 6, Bootstrap 4; 2). Language: HTML5, CSS3, JavaScript; 3). Libraries: d3 version 5, Ant Design, MapBox-GL API; 4). Package Manager: npm; 5). Version control: Git

3.3 Development

Angular Framework The file "app.component.html" sets the navigation head bar in all of the pages. The "nav-home-view" component set the layout of home page. The "nav-map-view" component includes 2 sub components: "map-box" and "select-attr". The "nav-analysis-view" component includes 3 sub components: "analysis-age-view", "analysis-infras-view" and "analysis-weather-view".

The development details of each component will be described in README.md.

Angular Techniques Angular router: Bind the router to links on a page and it will navigate to the appropriate application view. Composition interaction: A data service is created to pass values between sibling components.

Maps and Charts Mapbox and d3.map is used to generate station map and top 13 popular station analysis map. d3.scale, d3 data enter, d3 shape, d3 transition and d3 events are used to show various charts.

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.

3.4 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.

Station Variation This page show the average activities for all stations per hour in one day and the variation of stations' count in different years. Users can select multiple years' data to see the line chart and bar chart. Hover on the bars to see the detailed count and neighborhood will be highlighted on the map.

4.2 Stations Analysis

Infrastructure Analysis The map in this page displays the top thirteen popular stations by default. By hovering mouse on different objects, the district's name and stations' address will displayed respectively. By clicking on the station, the map will be zoomed into a bounding box of that station and important infrastructures will be shown as icons.

Age Analysis Top 6 popular stations with maximum orders are used for comparison. The age is divided into four groups and users can click on year buttons to see the age composition of each stations.

Weather Analysis Precipitation data is selected to represent weather and preprocessed by multiply 40000 for intuitive visualization. By concreting period and zoom out the line chart, users can easily find the strong relationship between number of orders and precipitation.

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

From the project, it is obvious stations' activities reached the highest value during rush hours in general. The activities increased with each passing year. From 2013 to 2018, the new installed stations mainly distributed in the eastern and northern districts.

For the three factors, surrounding the top popular stations, there are many important infrastructures such as subway stations, schools and hotels, which means the new stations can be installed around these infrastructure to share the customers. For the age, there is no direct relationship. However each station has special age composition. For example, stations near NYU has more young users and station near park has more old users. It might infer there is a strong relationship between age composition and nearby infrastructure. As for precipitation, the number of orders decreases as precipitation increases. In addition, orders in summer, spring and autumn are higher than that in winter.

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