# PittBike - A visualization exploration for Pittsburgh bike share

Liyun Suo
University of Pittsburgh
LIS86@pitt.edu

Weiyi Tuo
University of Pittsburgh
WET19@pitt.edu

Yuqian Zhanfu
University of Pittsburgh
Yuz98@pitt.edu

Yue Li
University of Pittsburg
YUL134@pitt.ed

## ABSTRACT
A lot of cities started their own city public bike share to encourage an environment-friendly, economic, healthy lifestyle. Similarly, there is an affordable and interesting bike share system waiting for people to explore in Pittsburgh. After observing datasets provided by Healthy Ride Pittsburgh Public bike official website. We planned to build a visualization platform for Pittsburgh bike sharing system to visualize its station, trips and customers information.

## Keywords
Pittsburgh, Public bike share, analysis, visualization platform

## 1. INTRODUCTION
Public bike share is a hot topic among the nation recently. There are a variety of analysis from different cities. Our project is based on Pittsburgh local public bike sharing system--Healthy Ride, which provides affordable public bikes through 50 stations spreading in multiple neighborhood in Pittsburgh. In this project, we are going to dig out more information from Pittsburgh public bike dataset, and present our discovery of the bike share situation in Pittsburgh. This platform is divided into several parts: “Station overview”, ”Station balance visualization”, “Trip period frequency visualization”, “trip duration visualization”. Each part is responsible for a specific visualization result. The “Station overview” section gives an overview including the station geographical location, station details, station traffic statistics; “Station distribution balance visualization” discovers the reasonability of current station arrangement;“Trip period frequency distribution visualization” shows the trip frequency for each station by hour and day; and “ Trip duration visualization” shows the percentage of trip duration for different customer types.

## 2. RELATED WORK
There are a lot of city bike share system researches conducted by researchers, developers, planners. Our design inspiration mainly came from some previous San Francisco bike share researches and local Pittsburgh bike share system researches

### 2.1 San Francisco Bike Sharing System Visualization
San Francisco bike share system can be considered as the most popular public bike share system in the country. A lot of researches regarding to the trip, station network, customers are conducted before. The left graph in Figure 1 shows a visualization design that used dots and size to show the location and poplularity of each station, and the right used lines and arrows to show routes.

Our station overview page was mainly inspired by these two visualization methods

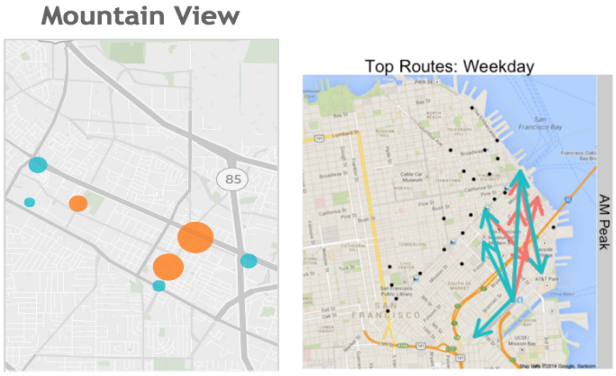


**Figure 1. San Francisco Public Bike Share Map and Route s**

### 2.1 Pittsburgh Bike Sharing System Visualization
The Healthy Ride Pittsburgh Bike Sharing System also published station and trip data from 2015 Q3 to 2016Q3, which contributes to the generation of Pittsburgh bike sharing system research. Figure 2 shows the station map overview of Pittsburgh bike sharing system. Figure2 shows the trip frequency study from previous work. In our project we adopted the same analysis idea with different visualization methods

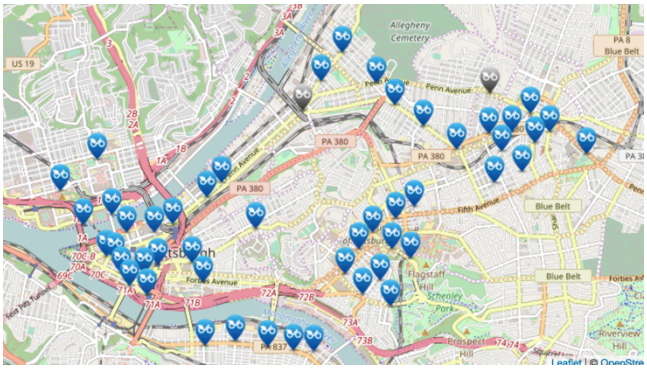


**Figure 2.Pittsburgh Public bike share station map**

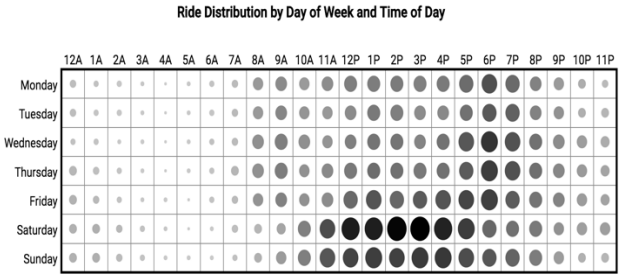


**Figure 3.Pittsburgh Ride Distribution by day of week and Time of Day**

# 3. VISUALIZATION DESIGN

## 3.1 Design Process

### *3.1.1 Data Process*

Firstly, we downloaded historical trip data from the Healthy Ride which includes from 2015 Q3 to 2016 Q3. The valuable information was stored in two files for each quarter. Rental fields include: Trip ID, Bike ID, Trip start and end time, Trip duration (in seconds), Trip start and end station name and station ID, Rider type—Member (pay as-you-go customer); Subscriber (deluxe and standard monthly member customer); Daily (24-hour pass customer). The Stations file includes: Station ID and name, coordinates, number of locker at each station.

The whole dataset contains comprehensive data details, also able to reflect the overall trend well.In the Station overview page. We first obtained the Pittsburgh city geographical dataset from Western Pennsylvania Regional Data Cetenter(WPRDC) to create Pittsburgh map base layer. Then we used simple SQL scripts to aggregate the total trip numbers for each station in the past 15 months. As for the Trip period frequency visualization, we grouped the trip counts from each station by month with SQL scripts and sliced up trips by time of day with Excel. To get correct dataset for Trip distribution part , all records were sorted by its user type and trip duration, we are able to know which kind of user is the most loyal user. In detail, the data is subdivided by station, this naturally leads the comparison among station. Also, it is easy to figure out which station is the top popular station through count “in” and “from” times.

### *3.1.2 Visualization Methods*

We used various methods for the four parts in our visualization system. To simplify the understanding process, we use map to show the overview of station connection detail as well as its lockers’ number. Additionally, sortable small multiple with line charts visualization technique, which is used in Station balance visualization part provides users a clean and simple visualization for balance situation in 50 stations. Slicing up the big plot avoids overplotting, and the sortable function enables users to view the information in multiple ways. Moreover, the circular heat chart used in Trip period frequency visualization, is a novel visualization design to display quantitative data as an array of circular segments, colored according to trip frequency. Finally, in the Trip distribution part, pie chart is the way to see the trip duration preference between different user types.

## 3.2 Visual Encoding

### *3.2.1 Map*

Our first visualization is an interactive map. Since our data involves location information, the best encoding for a geographical attribute would be the map.Dot, size and lines are also used in this map. The dots on the map represents the location.The larger the dot is; more popular the station is. Besides, we chose the yellow color for the stations which is contrast obviously on the dark blue map. All encodings are summarized in Table 1.

**Table 1. Encoding table for Map**

| Variable | Type | Encoding |
|---|---|---|
| Station location | Latitude & Longitude | 2D position - dot |
| Station frequency | Quantitative | Size |

### *3.3.2 Sortable Small Multiples*

We also implemented small multiples visualization technique in our design to visualize the balance situation of each station by months. The general idea of small multiples is to slice up the big plot into multiple small plots by condition. The result is a grid of charts that all follow the same visual formats, and show different pieces of data. With sortable interactive function, the sequence of the multiple small plots can changed by the initial balance gap or the station name. We visualized the balance situation based on each station, which means we have 50 small graphs. In each chart, X – axis stands for time; Y – axis stands for bike balance. We use yellow lines to show the trend in each station.

**Table 2.Encoding table for Sortable Small Multiples**

| Variable | Type | Encoding |
|---|---|---|
| Bike balance | Number | Y-axis |
| Time | Ordinal | X-axis |

### *3.3.3 Circular Heat Chart*

Circular Heat chart is the best way for displaying frequency. In this section, we designed a visualization to display one-week rental trips frequency. The darker the color is, the higher frequency of the trip in this time period is.

**Table 3.Encoding for Circular Heat Chart**

| Variable | Type | Encoding |
|---|---|---|
| Frequency | Numbers | Shade of color |
| Weekday | Categorical | Label |

### *3.3.4 Interactive Pie Chart*

We designed the pie chart to present the percentage of users’ trip duration based on different user types. User can play with the pie chart to choose certain user types to see the percentage. The encoding methods are shown in Table 4.

**Table 4.Encoding For Pie Chart**

| Variable | Type | Encoding |
|---|---|---|
| Customer type | Categorical | Color |
| Trip time | Number | Pie |

## 3.3 Layout

The whole visualization platform was divide into four part, by pressing the content button on the left side, the related visualization will display.

### *3.2.1 Station overview*


**Figure 4.Station Overview Page Snippet**

The station overview part uses the Pittsburgh real map to express the general information for different station within 5 quarter. The size of circle represents the locker number, therefore, the bigger the circle the more lockers it has. Also, we use red line to connect two circle, intense red lines represent the station maintains high using frequency. In addition, user can move mouse to point each station to get more detail.

### *3.2.2 Station balance visualization*

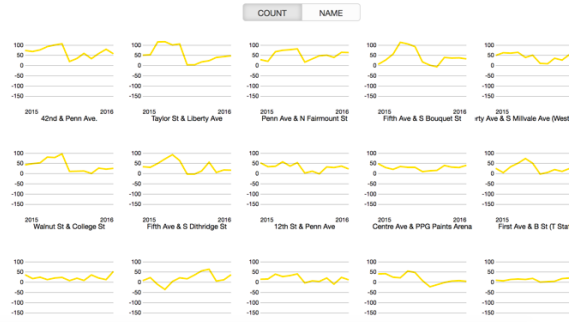


**Figure 5.Station Balance Visualization page Snippet**

This part sliced up the data and uses a separate plot to visualize each slice. Every single line chart indicates the difference of the bike "in" and "out" at each station. The X – axis means time series, the Y – axis means Bike balance, our team classify all of 5 quarters data based on its station name, calculates the monthly average difference between "in" and "out". The default page is ranked by counts, however, the sortable interaction enables review information by station name.

### *3.2.3 Trip period frequency visualization*

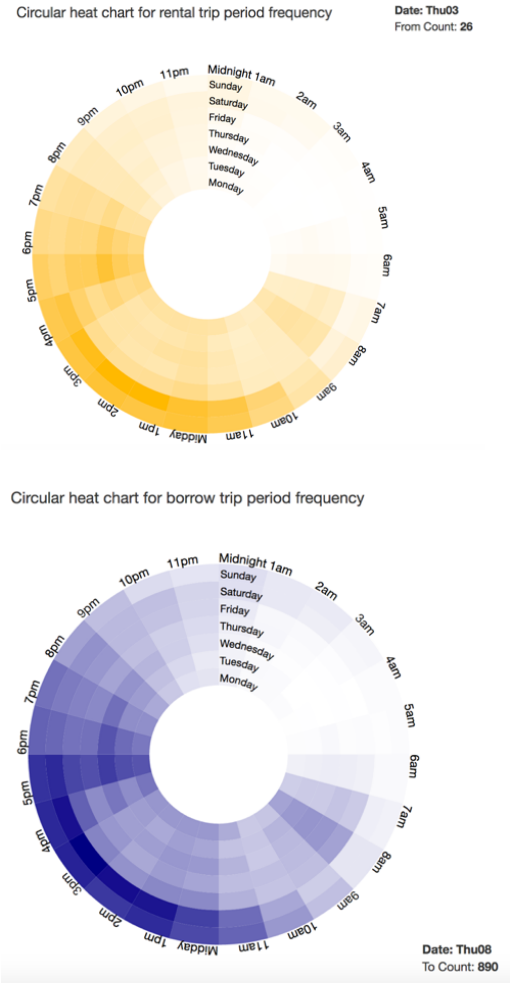


**Figure 6.Trip Period Frequency Page Snippet**

The circular heat chart can be regard as the most creative part of our platform, it displays quantitative data as an array of circular segments, color depth in on behalf of numerical size. Day and hour time were sequentially placed in donut circle, and using hue to show the trip frequency amount. The darker the color is, the higher frequency of the trip in this time-period is. Mouse over the single cell can show the specific date time and trip number

### *3.2.2 Trip duration visualization*

All data were divided into 9 types based on its user type and trip duration length, every color indicates one type. The platform provides two ways with user to make operation. User can interact with check box to compare/see the trip duration time by customer types or duration length. Also, viewers can see average trip duration and percentage, therefore if there are two types maintain close number, user can get the exact value.

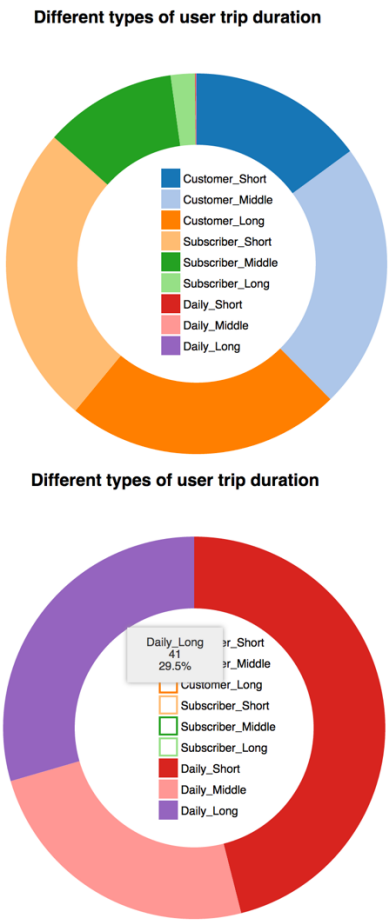


**Figure 7.Trip Duration Visualization Snippet**

## 3.4 Interaction

### *3.4.1 Station overview*

In this part, hovering on one circle, users can learn about which other circles connect with it. And in the left-top of page, the name of station will appear.

### *3.4.2 Station balance visualization*

In this part, all single line chart are connected with each other, when user choose any time point in any line chart, rest of all will show the number. Also, the sorted methods enable users to ranked the line charts by name and counts.

### *3.4.3 Trip period frequency visualization*

In this part, when hovering on the heat circle, both of two circle will specific number at that time point.

### *3.4.4 Trip duration visualization*

In this part, hovering on one part, users can get specific number of one type and the percentage in the current situation.

# 4. IMPLEMENTATION NOTES

Our most important steps in this visualization design is to identify data and determine information presented. Based on the dataset and the information we want to communicate, we start the coding, implementation, testing part.

To clean data sets, we used some simple SQL scripts and Excel to aggregate the original data set, and transferred file formats respectively into .tsv , .json, . csv depending on visualization graph type. The entire data visualization part are built with Data Driven Documents JavaScript library(D3.js) . Bootstrap, a HTML,CSS and JS framework is implemented into the main page design. As for the interaction part, we used some JavaScript library such JQuery, CoffeeScript to realize functions

# 5. EVALUATION

The usability evaluation step plays essential role in the success of a visualization platform. This step helps evaluate the quality of communication between a visualization platform and users. Our questionare is divided into two part, which included 8 questions. The first 3 questions are aimed to evaluate the general information efficiency, function and interaction of the platform. The rest four problems are designed to evaluate if users can obtain information we want to provide with our visualization methods. The measure scale is based on System Usability Scale(SUS), which is consist of five response option for respondents. Scores 1 and 5 in the scale respectively represents “Strongly disagree” and “Strongly agree”. The bigger the score is, the better the performance is in this evaluation item.

| | System Level Evaluation | | | Perceptual Evaluation | | | |
|---|---|---|---|---|---|---|---|
| ID | 1.You can easily understand and use the visualization platform functions | 2.This platform provides information that can help you play your bike trip | 3.You can effectively interact with the platform with interactive functions | 4.You can easily see which month has the largest return and borrow imbalance at each station | 5.You can easily see/compare which type of customer tend to use bike longer | 6.You can easily see the connection and trip frequency of stations from map overview page | 7.You can clearly see the busiest trip hour from trip time schedule page |
| 1 | 4 | 3 | 4 | 4 | 4 | 4 | 4 |
| 2 | 5 | 4 | 5 | 4 | 5 | 3 | 5 |
| 3 | 5 | 5 | 4 | 4 | 4 | 3 | 5 |
| 4 | 4 | 4 | 5 | 3 | 5 | 3 | 4 |
| 5 | 4 | 4 | 4 | 3 | 4 | 3 | 5 |
| 6 | 5 | 3 | 3 | 4 | 4 | 3 | 3 |
| 7 | 4 | 4 | 4 | 4 | 5 | 3 | 4 |
| 8 | 5 | 3 | 4 | 3 | 4 | 4 | 4 |
| Average | 4.5 | 3.75 | 4.125 | 3.625 | 4.375 | 3.25 | 4.25 |
| | 4.125 | | | 3.875 | | | |

**Figure 8.User Evaluation Table**

From the evaluation result, the general function score is around 4, which means from an average trend, users are agree that this platform is easy to understand and use. As for the perceptual evaluation part, we need to improve the station balance viz page and the map overview page for a more effective platform.

# 6. DISCUSSION

## 6.1 Advantages and Limitations

keeps a simple clean interface design style. Users cansimply navigate to different visualization sections through the navigation side bar. In addition, compared to previous Pittbike related works, our project creates a consistent visualization platform to share Pittsburgh bike share information from various aspects. There are also some limitations in terms of the dataset and functions:

1. The data set has slightly difference with reality because some bikes are returned at no-recorded stations or stolen, which are removed during the data cleaning process.
2. The map overview page can not be zoomed in and out and the detail information snippet of each station
3. More in depth information can be digged out from time dimension

## 6.2 Improvements

According to the user evaluation and our self evaluation, there are several improvements that we could implement in the future work:

1. The interactive function can be improved to give users more flexibility to communicate with information.
2. Add some analysis and visualization that includes external factors like weathers, holidays, seasons.

3. Improve the loading speed with some efficient data retrieval algorithm.
4. Add some description into visualization page to help users better understand our design

## 7. CONCLUSION

Our project is aimed to explore the information of Pittsburgh public bike share with the data from July 2015 to June 2016. We hope to provide an effective way for users to plan their public bike trip in Pittsburgh , as well as some insightful information for a better public bike system construction. We adopted some novel visualization methods to present the dataset such as sortable small multiple and circle heat chart. For the future work, the function optimization and interface design parts will be worked on.

## 8. CONTRIBUTIONS

**Yue li** was responsible for data cleaning for the trip duration visualization and CSS adjustment on map overview. She also worked on some parts of presentation slides , she and weiyiTuo wrote the interaction part in final paper and uploading the project folder

**Weiyi Tuo** was responsible for the data cleaning, coding,CSS adjustment for trip distribution visualization. He was also responsible for the demo presentation and combination. In addition, he wrote the visual design section in the final paper

**YuQian Zhanfu** was responsible for the abstract, project and progress report, as well as interface structure HTML and CSS design. She and liyun suo worked together on the data cleaning and generation of station distribution visualization . She also worked on the presentation and the final paper compilation

**Liyun Suo** was responsible for most of the coding and D3.js templates parts including station overview, trip frequency visualization and station balance visualization. She modified and adjust a lot of Java Script and CSS code. Besides, she also worked on discussion part in final paper

## 9.REFERENCE


[1] Federici, S., & Borsci, S. (n.d.). Usability evaluation: Models, methods, and applications http://www.academia.edu/11682033/Usability_evaluation_models_methods_and_applications

[2] Allandingham, J. (2016). How to Make Interactive Linked Small Multipleshttps://flowingdata.com/2014/10/15/linked-small-multiples/

[3] Circular heat chart. (n.d.). Retrieved December 13, 2016, from http://prcweb.co.uk/lab/circularheat/

[4] A Study of HealthyRide Pittsburgh Data. (n.d.). http://lejnine.com/bike_analysis.html

[5] What's happening with Healthy Ride? (n.d.). http://suds-cmu.org/2016/04/21/whats-happening-with-healthy-ride/

[6] Bureau, U. C. (n.d.). Geography. http://www.census.gov/geography.html

[7] Bay Area Bike Share Data Challenge. (n.d.). http://thfield.github.io/babs/