All You Need to Know about Healthy Ride in Steel City

Yumeng Lu
University of Pittsburgh
135 North Bellefield Avenue
Pittsburgh, PA 15260
yul139@pitt.edu

Fengtao Wu
University of Pittsburgh
135 North Bellefield Avenue
Pittsburgh, PA 15260
few14@pitt.edu

Linghui Liu
University of Pittsburgh
135 North Bellefield Avenue
Pittsburgh, PA 15260
lil103@pitt.edu

Anand Varanasi
University of Pittsburgh
135 North Bellefield Avenue
Pittsburgh, PA 15260
skv9@pitt.edu

ABSTRACT

Pittsburgh's bike sharing system Healthy Ride began the trail operation in summer in 2015 and the ambition of the system is "expanding access to public transit through easy-to-use, affordable active transportation opportunities". The information visualization system we developed visualizes the characteristics of the stations and patterns for the rides from July 1st, 2015 to March 31st, 2016. The information of both the system average level and the specific station level is visualized. Our visualization system will enable both service provider and customers identify the features and dynamics embedded in Healthy Ride on their own interests and in self-defined manners by the means of multiple visualization approaches.

On the system average level, the system visualizes the location of top five stations where the most bikes are rented and the variation of ride counts in the different circumstances within the different time granularities including week, month or quarter. On the specific station level, the system visualizes the location of top five destination stations where the most bikes are returned when a specific station is set to be a departure point and the variation of ride counts in the different circumstances across the different weeks.

Keywords

Information Visualisation; D3.js; Pittsburgh Bike share

1. INTRODUCTION

Renting and riding a bike has become a popular choice of transport for a number of people living and working in Pittsburgh. As stated in the Healthy Ride website, they have received many questions about how the system functions and where the most popular stations are located. Therefore, Healthy Ride releases datasets every quarter in order to provide information regarding how the system is being used.

These datasets includes all the data of bike rides and the information about all the existing stations in Pittsburgh. There are thousands of detailed records and missing values in the datasets, and hence it is almost impossible for both Healthy Ride and the customers to identify how the system is being used by performing basic data analysis using the datasets directly. This information visualization system aims to visualize the characteristics of the stations and patterns of rides from July 1st, 2015 to March 31st, 2016. The information visualization system we implemented will not only visualize the information on the system average level (i.e., to examine how the system is being used), but also visualize the information on the specific station level (i.e., to focus on how each station is being used). Instead of telling the viewers about the customers and the results obtained from the datasets and analysis, the visualization system enables the viewers to identify the features and dynamics on their own interest and in self-defined manners by the means of multiple visualization approaches. The viewers have freedom to choose the time period (i.e., week/month/quarter) and station they are interested in, whether they want to see what happens in the whole system or the specific station and what time granularity is used in the data aggregation.

On the system average level, the visualization system visualizes the location of top five stations where the most bikes are rented and the variation of daily ride counts in the different circumstances which means different weather conditions and different temperature intervals within the different time granularities such as week, month and quarter. The visualization system also compares the hourly ride counts in the course of day between weekday and weekend. On the specific station level, the project visualizes the location of top five destination stations where the most bikes are returned when a specific station is set to be a departure point. Similarly, the variation of ride counts in the different circumstances across the different weeks and the comparison of the hourly ride counts in the course of day between weekday and weekend are visualized for each station.

2. RELATED WORK

Four visualization results on Healthy Ride has been posted in the Healthy Ride website:

DOI: 10.1145/1235

1. Jackson Whitmore[9] published *What's happening with Healthy Ride?* on April 2016. The analysis evaluated data from the first quarter of the system's operations (May 30, 2015 to September 30, 2015), and demonstrated the results of the distribution of the ride count among different types of customers and the distribution of the ride duration.

2. George Lejnine[7] published *A Study of HealthyRide Pittsburgh Data.* on April 2016. The analysis evaluated data from July 1st, 2015 to December 31, 2015, and described how temperature affected the number of ride and the distribution of ride count in the course of week. George also applied the Chord Diagram to visualize the information of station origination and termination.

3. Lauren Renaud[8] published *Tableau: Healthy Ride PGH Bike Share Data.* on February 2016. The analysis evaluated data from the first six months of operation, and compared the hourly ride count in the course of day and the daily ride count in the course of week.

4. Southwestern Pennsylvania Commission (SPC)[1] pubilshed *Healthy Ride Pittsburgh Bike Share Maps.* on September 2015. SPC created the static maps to help visualize the most popular trips and stations throughout the city during the third quarter of 2015.

In summary, all the visualization results mentioned above visualized the following information:

1. The distribution of ride count among different types of customers.
2. The effect of temperature on the daily ride count.
3. The distribution of ride count in the course of day or week.
4. The comparison of ride count between weekday and weekend.
5. The most popular starting or destination stations.

In the information visualization system we implemented, we visualize and display those information in a more user-friendly way, too. Compared to the above four visualization results, our visualization system has the following advantages:

1. The system will not only visualize the information on the system average level to examine how the system is being used, but also visualize the information on the specific station level to focus on how each station is being used.
2. Instead of telling the viewers the static results and statistics already obtained, the system enables the viewers to identify the features and dynamics on their own interest. The viewers has freedom to choose the time period and the station they are interested in, whether they want to see what happens in the whole system or the specific station and what time granularity is used in the data aggregation.
3. The system has better organized visual encoding framework and enriches the information displayed in each chart or diagram. For example, the system enables the viewers to compare the ride count between different hours in the day and between weekday and weekend in one diagram.

3. VISUALIZATION DESIGN

3.1 Design process

We believe that it is the users' requirements that should decide what the functions and components are offered by the information visualization system, instead of the designers themselves. As stated clearly in the Healthy Ride website, identifying how the system functions and where the most popular stations are located are two questions which are addressed and concerned the most. We combine those two questions together with our analysis of data from July 1, 2015 to March 31, 2016, and formulate the following questions to which the viewers will find the answers by the means of the visualization system.

1. locations of the stations where customers can rent bikes and how many bikes does the station offer (rack quantity)?
2. locations of the top 5 stations where most bikes are rented within every week/month/quarter?
3. locations of the top 5 destination stations where the most bikes are returned within each week if the viewer selects the starting station?
4. How does the temperature affect the average daily ride count and what proportion does each type of customers account for on the system level within every week/month/quarter?
5. How does the weather affect the average daily ride count and what proportion does each type of customers account for on the system level within every week/month/quarter?
6. How does the average hourly ride count vary in the course of day and what is the difference of average ride count between weekday and weekend on the system level within every week/month/quarter?
7. How does the temperature affect the average daily ride count and what proportion does each type of customers account for at every station within every week?
8. How does the weather affect the average daily ride count and what proportion does each type of customers account for at every station within every week?
9. How does the average hourly ride count vary in the course of day and what is the difference of average ride count between weekday and weekend at every station within every week?

In summary, the question 1,2,4,5 and 6 belong to the system level, and the question 3,7,8 and 9 belong to the specific station level. For this project, We collect the bike ride record and station data from Healthy Ride, weather data from National Climatic Data Center (NCDC)[6] and calendar data from the City of Pittsburgh Holiday Schedule[2].

3.2 Layout

The layout of the information visualization system is displayed in Figure 1.

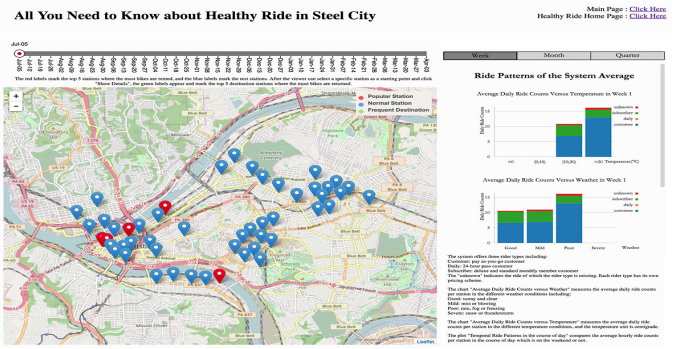

Figure 1: Layout of the Information Visualization System

The time slider on the top left, which is shown in Figure 2, is used to control the movement of time. The range of

the time slider is from July 5, 2015 to April 3, 2016, and each axis point represent the past week. Therefore, the time slider moves by week.

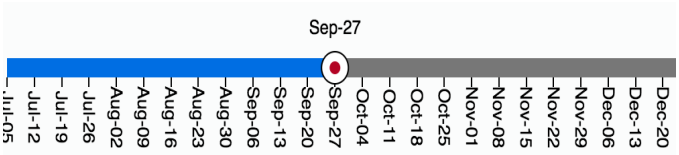

Figure 2: Time Slider Component

The map view is on the bottom left, and all 50 existing stations are labeled in the street map. The map view is shown in Figure 3. The map view component also illustrates where the to five popular stations within the specific time granularity are.

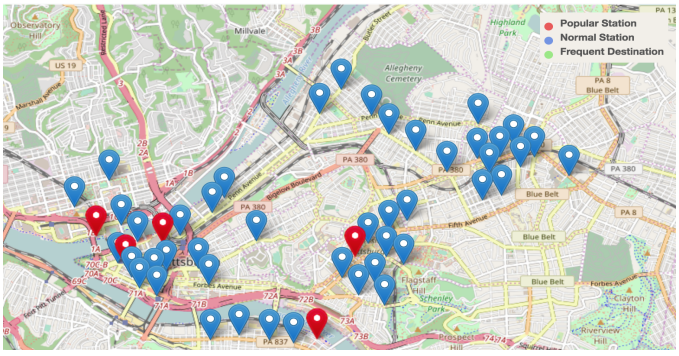

Figure 3: Map View Component Highlight the Top 5 Popular Stations

By clicking any specific station label, the tool-tip will appear to show the detailed information including the station ID, the address and the rack quantity. There is a button indicating clicking means displaying more details in the tool-tip, too. Once the button clicked, the map will highlight top five destination stations in green for chosen station. The effect is shown in Figure 4.

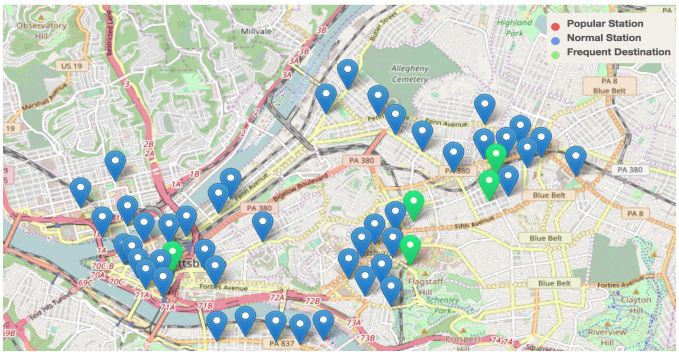

Figure 4: Map View Component Highlight the Top 5 Destination Stations for Each Starting Station

The granularity panel is on the top right, and the user of the system can choose the time granularity used in the data aggregation. The granularity panel is shown in Figure 5. For example, the handle in the time slider stops at September 27, 2015. If the user chooses "Week", the data aggregation is done for the past week and the system displays the result of the past week. If the user chooses "Month", the data aggregation is done for September 2015 and the system displays the result of September 2015. If the user chooses "Quarter", the data aggregation is done for the third quarter in 2015 and the system displays the result of the third quarter in 2015.

Figure 5: Granularity Component

Three charts including two bar charts which describe the effect brought by the temperature and weather to ride count on the system average level and one line chart which compares the hourly ride count in the course of day between weekday and weekend on the system average level are on the middle right. The room is enough to see two charts at the same time when the viewer is moving the time slider, and hence the scrolling bar enables the viewer to watch the line chart and move the time slider at the same time.

Once a specific station is selected by the user and the "Show details" button is clicked, the charts component will display the information for that specific station. There are still three charts including two bar charts and one line chart. The layout of the charts component are shown in Figure 6 and 7. When the mouse is hovered over the proportion of each customer type, the value corresponding to the proportion appears.

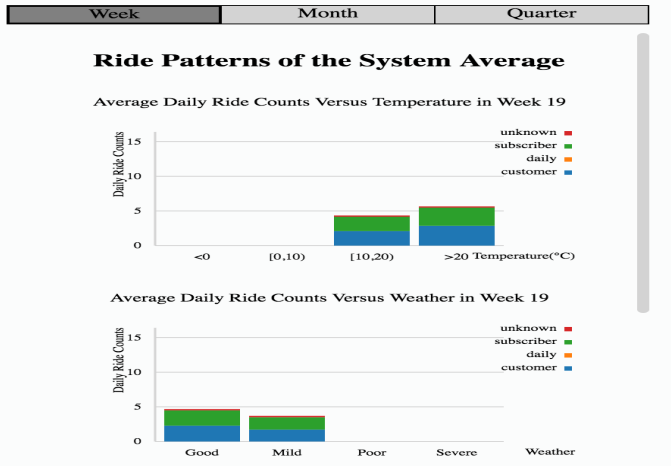

Figure 6: Bar chart of the Information Visualization System

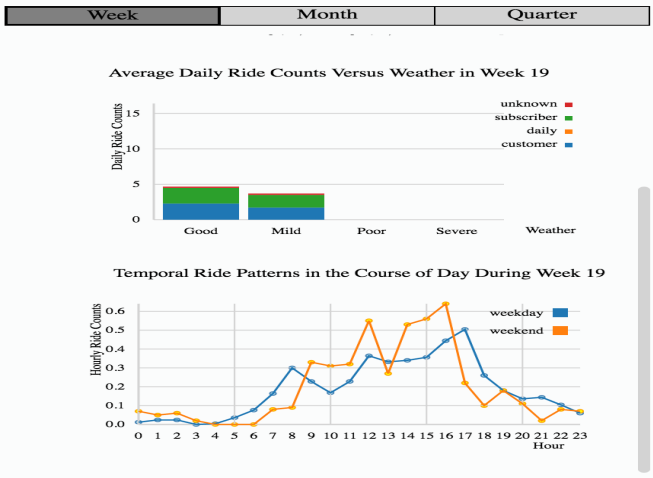

Figure 7: Line chart of the Information Visualization System

The explanation for notations and legends are on the bottom right, and it is shown in Figure 8.

The system offers three rider types including:
Customer: pay as-you-go customer
Daily: 24-hour pass customer
Subscriber: deluxe and standard monthly member customer
The "unknown" indicates the ride of which the rider type is missing. Each rider type has its own pricing scheme.

The chart "Average Daily Ride Counts versus Weather" measures the average daily ride counts per station in the different weather conditions including:
Good: sunny and clear
Mild: mist or blowing
Poor: rain, fog or freezing
Severe: snow or thunderstorm

The chart "Average Daily Ride Counts versus Temperature" measures the average daily ride counts per station in the different temperature conditions, and the temperature unit is centigrade.

The plot "Temporal Ride Patterns in the course of day" compares the average hourly ride counts per station in the course of day which is on the weekend or not.

Figure 8: Annotation of the Information Visualization System

3.3 Visual Encoding

We applied different visual encoding rules to design the system. The encoding schemes used in different components of our system are illustrated in Table 1,2 and 3.

Table 1: Visual Encoding Scheme in the Time Slider and Map View Component

Variable	Type	Encoding
Station Address	Nominal	Text
Station Location	Geographical	Point
Rack Quantity	Quantitative	Text
Station Type	Categorical	Color
Time	Quantitative	Slider

Table 2: Visual Encoding Scheme in the Bar Chart Component

Variable	Type	Encoding
Temperature/Weather	Nominal	X
Rides Count	Quantitative	Bar Length
Customer Type	Nominal	Color

Table 3: Visual Encoding Scheme in the Line Chart Component

Variable	Type	Encoding
Time	Quantitative	X
Ride Count	Quantitative	Point Position
Day Type	Nominal	Color

3.4 Interactions

Our system provides great freedom to the users, and the users can choose the time period and the station they are interested in, whether they want to see what happens in the whole system or the specific station and what time granularity is used in the data aggregation. In a other word, the interaction comes from three components: time slider, granularity panel and map view.

3.4.1 Interactions from Time Slider

Time slider is an important component in our system, and the users can use slider bar to control the time movement. The map view will automatically update the top five popular stations when the user moves the time slider. For example, when the slider handle is moved from July 5, 2015 to November 29, 2015, the top five popular stations are updated, which is depicted in Figure 9 and 10.

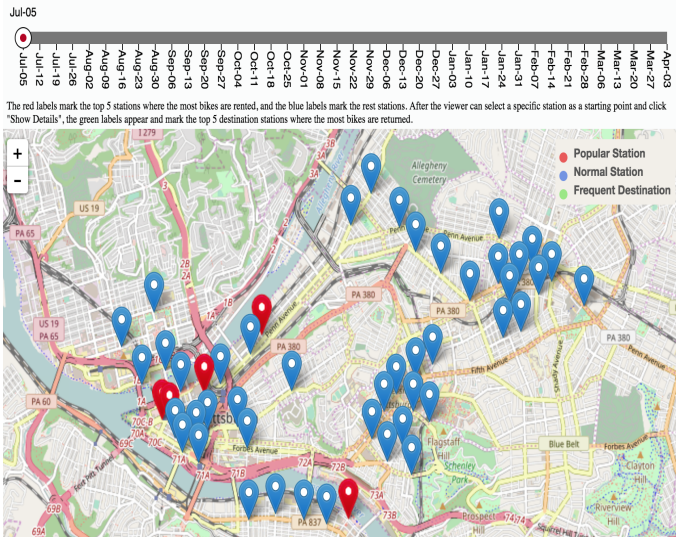

Figure 9: The Top 5 Popular Stations in July 5, 2015

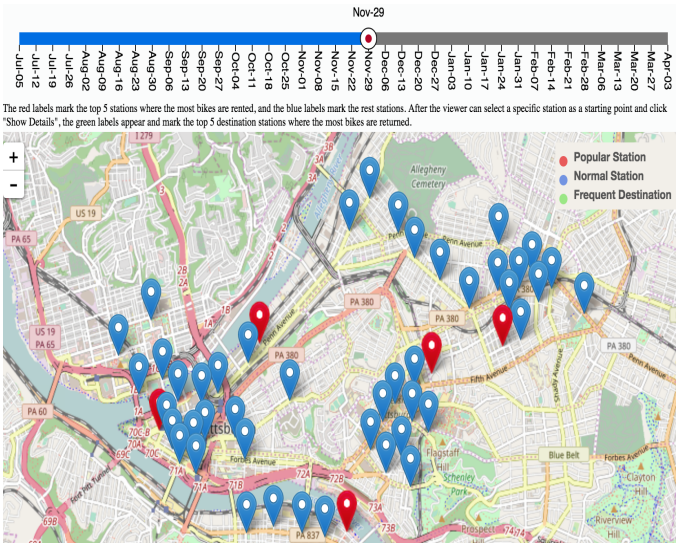

Figure 10: The Top 5 Popular Stations in November 29, 2015

Time slider also will update what is displayed in the chart component. For example, when the slider handle is moved from July 5, 2015 to November 29, 2015, the bar chart describing the relationship between weather and ride count is updated, which is depicted in Figure 11 and 12.

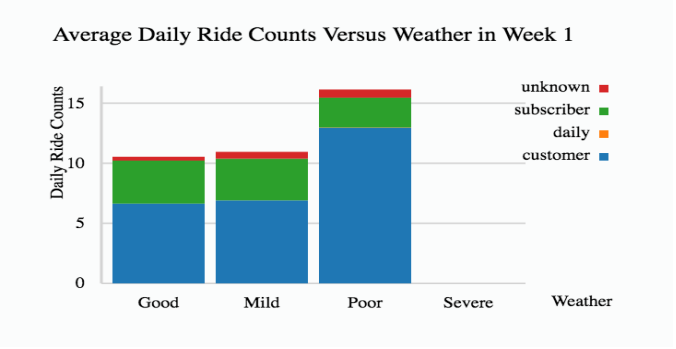

Figure 11: Average Daily Ride Counts Versus Weather in July 5, 2015

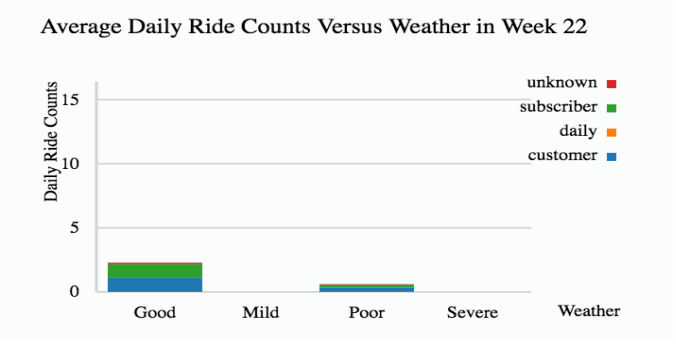

Figure 12: Average Daily Ride Counts Versus Weather in November 29, 2015

3.4.2 Interactions from Granularity Panel

The granularity panel is another important component affecting the interaction function. It will influence the time granularity used in the data aggregation, which is critical in viewing the result of the system. Both the map view component and the chart component will update the display according to the change of time granularity. For example, Figure 13 shows bar chart describing the relationship between temperature and ride count in week 22 (Time slider stops at November 29, 2015), when the granularity is set to be week. If the granularity is changed to month, the bar chart updates automatically and shows relationship between temperature and ride count in November. The update is depicted in Figure 14.

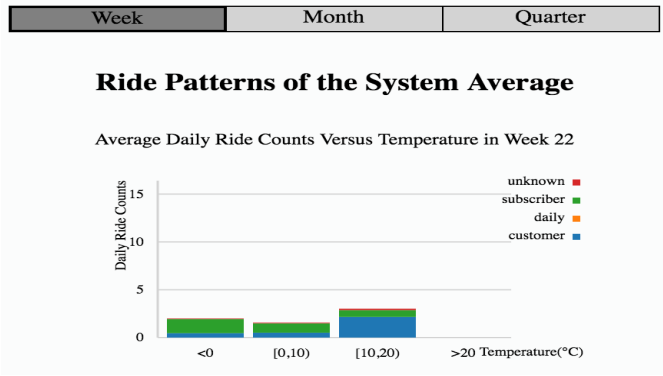

Figure 13: Bar Chart When Time Granularity is Week

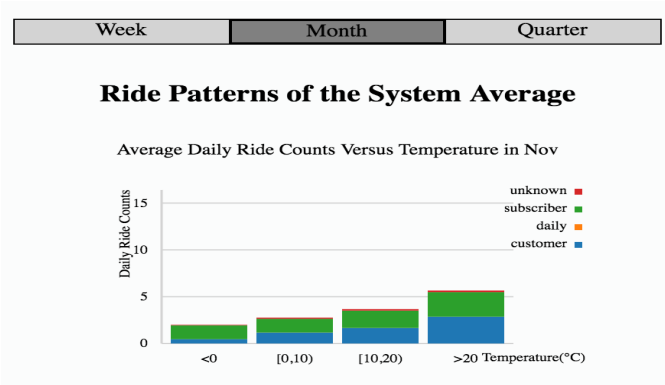

Figure 14: Bar Chart When Time Granularity is Month

3.4.3 Interactions from Map View

The map view is the component having the most mouse interaction functions in our system. First the user can zoom in and zoom out to see the detail and bird view in the map. Second, the top 5 popular stations are automatically labeled in red, which is depicted in Figure 16. all the stations are marked in the map, and they are all interactive. The user can choose any station he or she is interested in by clicking, the tool-tip appears to show the detailed information including the station ID, the address and the rack quantity. There is a button indicating clicking means displaying more

details in the tool-tip, too. Once the button clicked, the map will highlight top five destination stations in green for chosen station. The effect is shown in Figure 17.

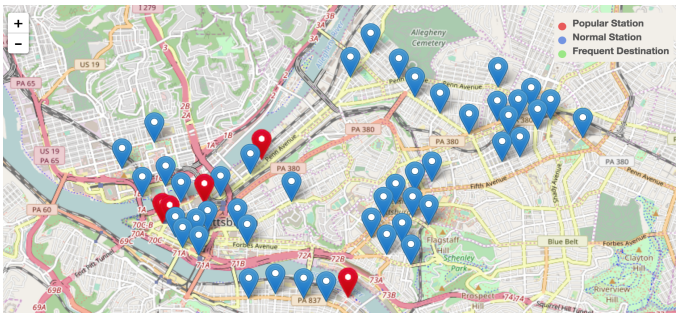

Figure 15: Five highlighting red station with chosen station

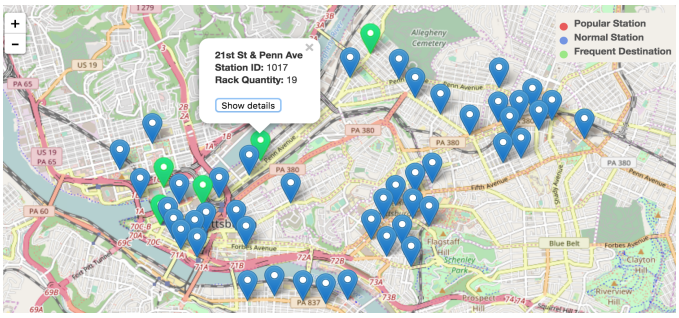

Figure 16: Five highlighting red station without chosen station

Once a specific station is selected by the user and the "Show details" button is clicked, the charts component will display the information for that specific station. There are still three charts including two bar charts and one line chart. Figure 17 and 18 display the temporal ride patterns in the course of day during week 22 for the stations whose ID are 1017 and 1035. The station 1017 is one of the popular stations in week 22 and the station 1035 is one of the normal stations.

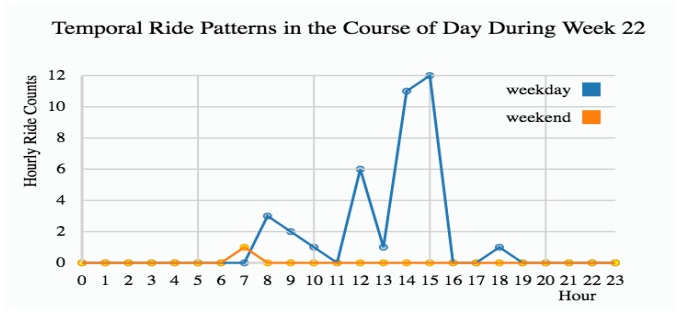

Figure 17: Temporal Ride Patterns in the Course of Day During Week 22 in station 1017

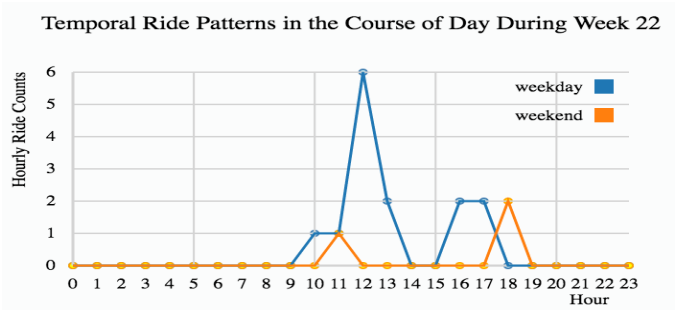

Figure 18: Temporal Ride Patterns in the Course of Day During Week 22 in station 1035

4. IMPLEMENTING NOTES

The ambition of our system is to visualize the data in a dynamic way by the means of integrating the interaction functions from time slider, granularity panel and map view.

On one hand, we focused on the data processing, and we prepared the datasets which answer the nine questions we purposed. Therefore, we need to prepare the datasets on both the system average level and the specific station level.

On the system average level, we prepared the following datasets:

1. The top 5 stations' information within every week/month/quarter;
2. The daily ride count aggregated by week/month/quarter and weather/temperature conditions;
3. The hourly ride count aggregated by week/month/quarter and the course of day.

On the specific station level, we prepared the following datasets:

1. The top 5 destination stations' information within every week/month/quarter for each station;
2. The daily ride count aggregated by week and weather/temperature conditions for each station;
3. The hourly ride count aggregated by week, the course of day, the weekend indicator (indicating whether the day is on the weekend or not) for each station.

On the other hand, we implemented different components in the system separately. The time slider, granularity panel and the chart components were implemented in D3.js[3], and we deployed the map-creating from a mature map generation tool-Leaflet[5]. Then We integrated map view, time slider, granularity panel and charts components together. The whole web system has two different architecture creating two kind of outcomes.

Architecture one: Time slider transmits the time value selected in the time slider to the charts component. Without choosing any specific station in the map, the charts component displays the result of the bike sharing system in the current week/month/quarter.

Architecture two: Time slider transmits the time value selected in the time slider to the charts component. With choosing the specific station in the map, the charts component displays the result of the specific station in the current week.

Figure 19 demonstrates how the different components in the system are connected. Both time slider and granularity panel will affect the map view. Time slider, granularity panel and map view will all influence the display of charts component.

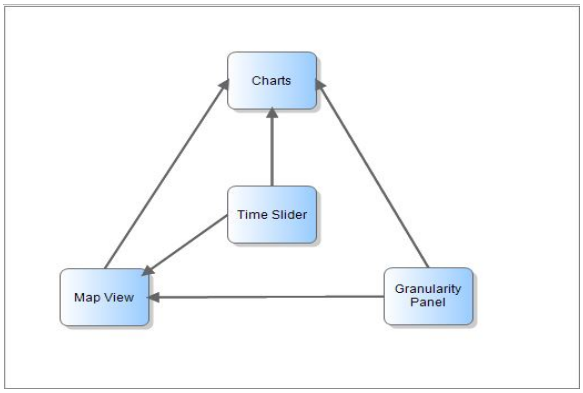

Figure 19: Flow chart of components

5. USAGE EVALUATION

After implementing the system, we would like to evaluate the quality of our system by listening to the users' feedback. We interviewed 10 candidates, who are graduate students, with eight questions. The eight questions include five evaluation type questions, two text questions and one self-evaluation question. These questions are based on the evaluation criteria from online sources[4].

Before introducing our system to the users, we asked the candidates what they knew about Healthy Ride. This process is necessary because it serves as the control group compared with the evaluation result of our system. Our evaluation question includes:

1. Does our system contain the information you want?
2. Does our system inspire you thinking?
3. Do you think it is efficient for you to learn more about Healthy Ride in our system?
4. Is it easy to interact with our system?
5. Do you think our system is helpful?

After interviewing the candidates, we got a preliminary assessment of our system. Table 4 shows the result of evaluation. The six questions are related to five evaluation criteria, which are functionality, effectiveness, efficiency, usability and usefulness. Our system scores over the average scores for different parts. The scores give us a brief overview. Based on the users' feedback, we did well in interaction functions, and they helped users understand the characteristics and patterns embedded in Healthy Ride lively. On the other hand, our system will do better to inspire users' thinking if there are more ways to interpret the data.

Table 4: result of evaluation

Question	Criteria	Score
Does the system contain the info you want?	Efficiency	4.3
Does the system inspire you thinking?	Efficiency	4.3
Is it easy for you to learn from our system?	Usefulness	4.5
Do you think it is helpful?	Usefulness	4.5
Self-Evaluation	Functionality	3.5
Do you think it is efficient for you to learn from our system?	Usefulness	4.5

6. DISCUSSION

6.1 Advantage and Inspiration

Considering the functionality in reality, our system has several advantages:

1. Our system fits the time range perfectly. Time slider moves by week from the third quarter in 2015 to the first quarter in 2016.
2. Our system is designed in a feasible framework. All the interactive functions cooperate well and there is no conflicts among them.
3. Our system offers an user-friendly interactivity. All the interactive functions are easy to use, and the visualization is provided with detailed explanations and annotations.
4. Our system uses an appropriate way to interpret different data in order for the easy and concise comparison of data through time.
5. Our system helps the service provider discover a number of patterns and characteristics embedded in Healthy Ride. For example, subscribers prefer to ride in business day for commuting from work to home and pay-as-you-go customers tend to take ride much more in weekend rather than weekday. These insights will help them offering better service in the future.

6.1.1 limitation and future work

1. Time Slider will be foreseeable crowded when the range of time becomes longer. Because of the limitation of the web page, we need to use different time scale measurement to fix this problem. But we will face the precision problem at the same time.
2. The interactive map view may not work well for the users using the touchpad, and it sometimes causes the inconvenience such as zooming in instead of moving in one direction. In the future, we need to keep the map display the informative area when the viewers use the interactive functions.
3. When the user open charts describing the ride pattern analysis for specific station, some users reply it is hard to find the close button for this function, which means they may have problem in returning to main page. We need to highlight the close or return button in future.
4. In our system, we used bar chart and scatter plot to express our data analysis, because we want to integrate the map view, time slider and data analysis in one web page, which causes the tight layout. We can not show too many functions in one page, so our way of interpreting the data is simple and normal. In future work, we want to make an embedded page in current page to enrich our techniques to interpret data.
5. Scroll bar for the charts component is hard to recognize sometimes. Because we have two bar charts and one line chart in our charts component, we are facing the problem that the space in the page is not enough to put three charts at the same time. We use a scroll bar to fix this problem. Unfortunately, the scroll bar is not visible enough. We may add an embedded page in current page for data display instead of using scroll bar forcing all charts in one page.

7. CONCLUSION

In this paper, we have designed and presented a interactive system to visualize the characteristics of the stations and patterns for the rides from July 1st, 2015 to March 31st,

2016. In order to ensure the system is useful in reality, we integrated several visualization techniques such as map and charts. Our ambition is help people who are interested in our city's bike sharing system and service provider understand our Healthy Ride in a lively manner. In the future, we will improve our system's functionality and provide interface for future add-ones.

8. ACKNOWLEDGMENTS

We would like to thank Dr. Yu-Ru Lin for her instructions and suggestions in each milestone of the project sincerely. We also greatly appreciate the four groups sharing their thinking on the project in the discussion. The data employed in this project is from Healthy Ride, National Climatic Data Center,the government of the city of Pittsburgh and leaflet. We appreciate for people and organization to record, organize and share those data.It is our pleasure to use D3.js in our project, which is a innovative visualization language in nowadays.

9. REFERENCES

[1] Healthy ride pittsburgh bike share maps. http://www.spcregion.org/trans_multi_atp.shtml#health, 2015. [Online; accessed 12-December-2016].

[2] City of pittsburgh holiday schedule. http://apps.pittsburghpa.gov/pcsc/2016_Holiday_Schedule.pdf, 2016. [Online; accessed 12-December-2016].

[3] D3.js. https://d3js.org/, 2016. [Online; accessed 12-December-2016].

[4] Evaluation criteria. https://www.kfw-entwicklungsbank.de/International-financing/KfW-Development-Bank/Evaluations/Evaluation-criteria/, 2016. [Online; accessed 12-December-2016].

[5] Leaflet. http://leafletjs.com/, 2016. [Online; accessed 12-December-2016].

[6] National climatic data center. https://www.ncdc.noaa.gov/, 2016. [Online; accessed 12-December-2016].

[7] A study of healthyride pittsburgh data. http://lejnine.com/bike_analysis.html, 2016. [Online; accessed 12-December-2016].

[8] Tableau: Healthy ride pgh bike share data. http://www.laurenrenaud.com/blog/2016/2/1/healthy-ride-pgh-bike-share-data, 2016. [Online; accessed 12-December-2016].

[9] What's happening with healthy ride. http://suds-cmu.org/2016/04/21/whats-happening-with-healthy-ride/, 2016. [Online; accessed 12-December-2016].

10. AUTHOR CONTRIBUTIONS STATEMENT

During the system implementation process, all four authors contributed the ideas to the goal and design of the system. Fengtao Wu collected and cleaned the data required by the system, and then prepared the datasets used in the system. Yumeng Lu, Linghui Liu and Anand Varanasi were responsible for implementing different web components including charts, map and time slider. Finally, all four authors cooperated together to integrate different components and refine the system layout. All authors reviewed the manuscript.