

FIT5147 - Data Visualization and Exploration

Assignment: Data Visualization Project.

**Subject: Exploring Bike Sharing in Austin, Texas, U.S.A from 2014 to 2016 by
Using Shiny for Interactive Visualisation**



MONASH University

Student name: Trung Kien Nguyen

Student ID: 29057957

Contents

List of figures	2
1. Introduction	3
2. Target Audience.	3
3. Five Sheets Design	4
4. Implementation.	9
5. User guide.	13
6. Conclusion.	18
7. Appendix	19
References	20

List of figures

Figure 1: Five sheets Design - IDEAS	4
Figure 2: Five sheets design - sheet 2	5
Figure 3: Five sheets design - sheet 3	6
Figure 4: Five sheets design - sheet 4	7
Figure 5: Five sheets design - sheet 5	8
Figure 6: Project Structure.....	10
Figure 7: Overview of the project in RStudio	11
Figure 8: Running project on RStudio	12
Figure 9: Running project on website	12
Figure 10: The first page 1	13
Figure 11: The second page 2.....	13
Figure 12: The first page 3	14
Figure 13: The first page 4	14
Figure 14: The daily analysis with night theme.....	15
Figure 15: The daily analysis with daylight theme	16
Figure 16: The daily analysis in dark theme.....	16
Figure 17: The station analysis	17
Figure 18: The station analysis with origin station animated line	17
Figure 19: The station analysis with destination station animated line	18

1. Introduction

Bike share in the U.S has continued its brisk growth, with 35 million trips taken in 2017, 25% more than 2016, according to the report published by National Association of City Transportation Officials (2017). This growth is attributable to increasing ridership in existing systems as well as the launch of several major new bike share systems across the country.

This report is focusing on providing public bike sharing operators in Austin, the U.S. There are three datasets have been taken into consideration. The first dataset is bike share dataset which contains information on bike trips from 2013 to 2017, the second dataset is train station which is the place located the docks in the city. Both datasets are from [Kaggle](#) which was published by the City of Austin. The last dataset is also from [Kaggle](#) which contain weather information of each day in the City of Austin during the period of 3 years between 2013 and 2017. Three datasets will be combined in on single file in order to visualize.

The first report has done with comprehensive visualization and reveals many useful and interesting about the bikeshare system of Austin as well as the correlation between weather condition and customers behaviors on using the system.

This report will focus on designing an interactive visualization in order to convey the insight of the data to the target audience. This report will firstly state the intended audience and define the narrative questions in order to support the audience to keep track of the report as well as understandable of data visualization.

2. Target Audience.

This visualization project aims to support top-level manager, system management and people who operate the entire bikeshare system. The Austin Bikeshare system has operated since 2012, for more than 7 years, the number of customers using the system increase year by year. With the dataset range from 2014 to 2016, this visualization project will focus on analyzing the traffic usage, the number of customers, traffic of each station as well as the daily activity of customers in different type. For more specific, the entire report aims to answer the following question.

- Overview of the entire system (how many customers, how many trips by year).
- How is the traffic of the system by year, month, day, and hours, weekdays and weekend?
- Does the average temperature affect the customer's usage?
- What is the busiest station, and what are popular routes?
- The daily traffic of each station (including origin station and destination station).
- The summary traffic of each station (including origin and destination station).

By answering questions above, the visualization support to better management, the top-level manager can take the advantages of animated visualization in term of busiest hours of the day to adjust the number of docks or number of the bike to decrease or increase in each station at each specific time.

3. Five Sheets Design

Sheet 1: IDEAS

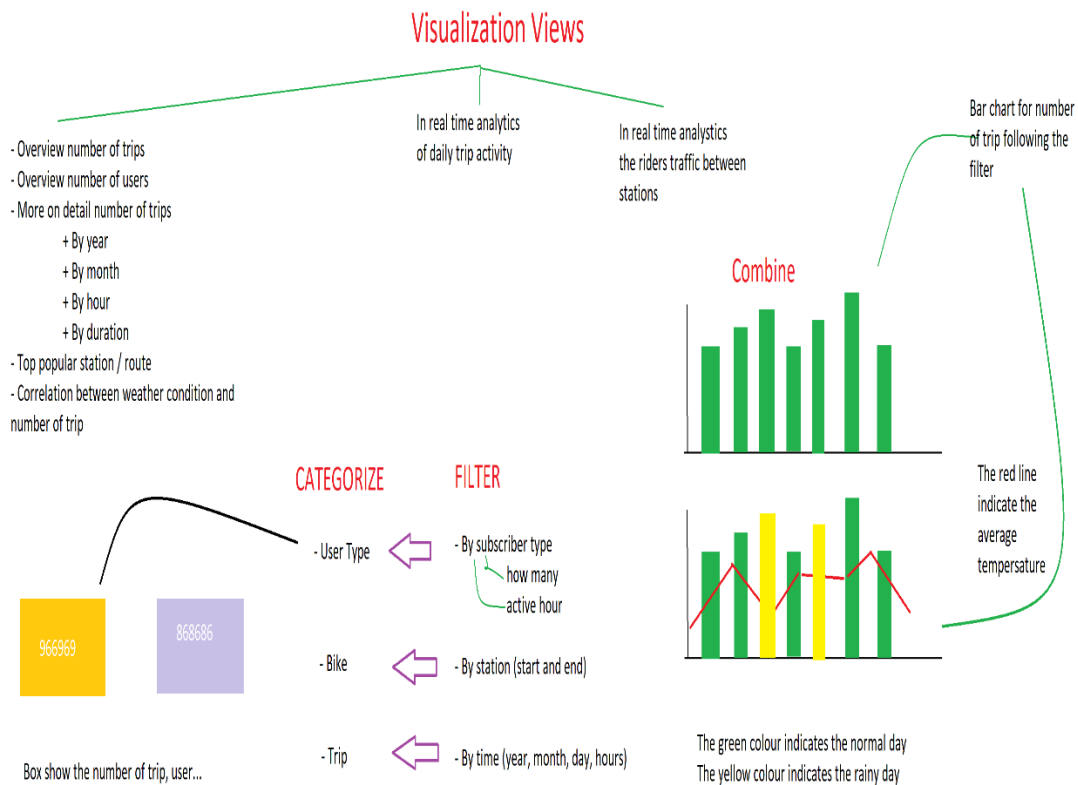


Figure 1: Five sheets Design - IDEAS

The first design sheet is brainstorming for what the visualization project is going to do.

At this sheet, three main views have been listed under the visualization view. This indicates for the narrative visualization as each main view will response to answers some specific questions which stated in the first section.

There is some category that visualization needs to be taken into consideration which is user type or in another word, subscriber type. The second categorize is bike that corresponding to stations and finally is number of trips by riders.

It is clear to see the user type to see how many users the system currently had, how many for each type of users, what is the most active user types. In term of bike category, it will be filtered by the origin station as well as the destination station. Finally, the trip in more intuitive way will be filtered by time which can consider few criteria such as number of trip per year, number of the trip by month of a specific year, by hours or by its duration.

Sheet 2: Main Dashboard

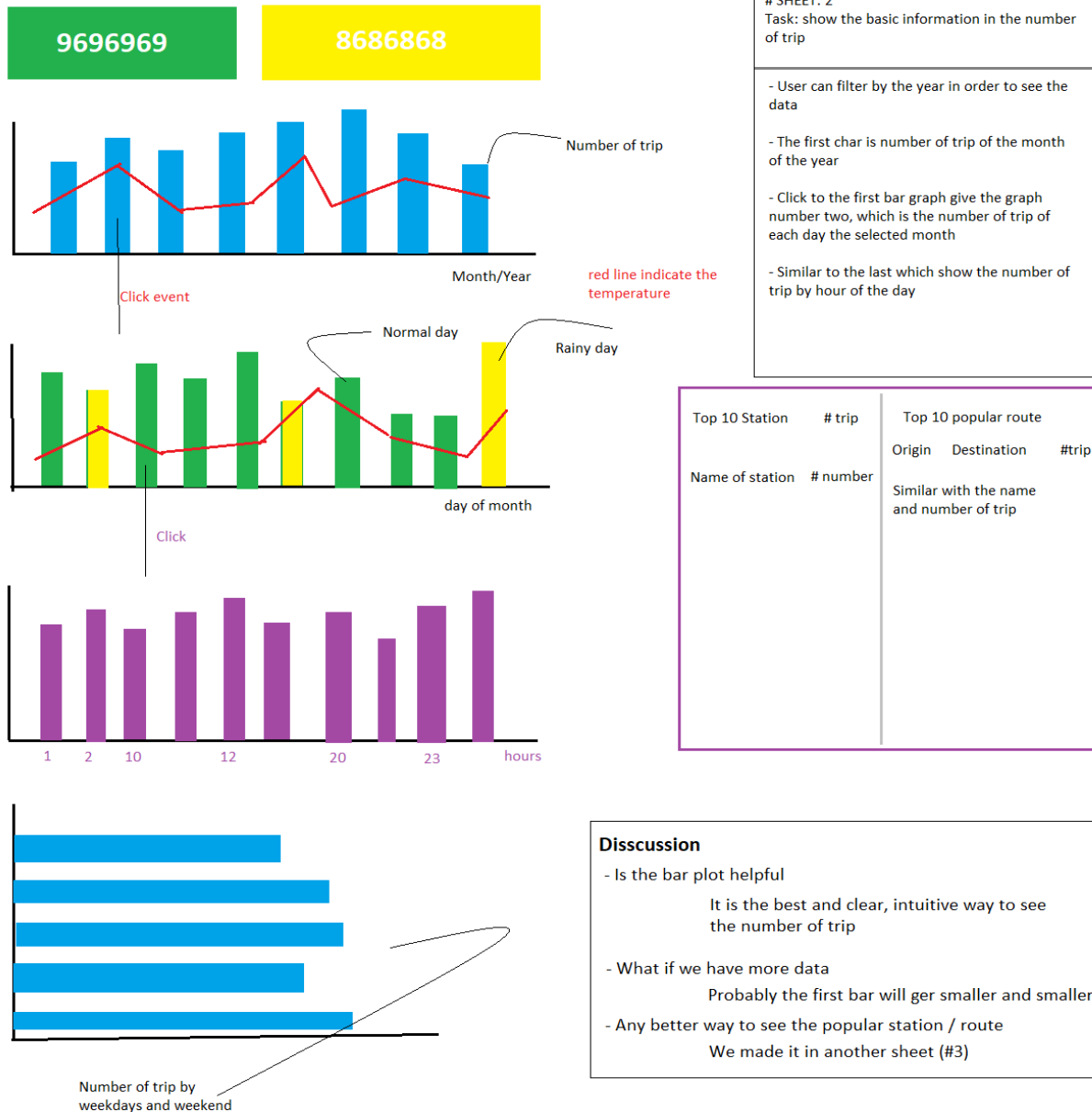


Figure 2: Five sheets design - sheet 2

Going into the second design sheet, which is the show the main dashboard of the system, this sheet shows the overview of the entire bikeshare system, from the number of riders to number of trips. Users can easily select the year to see more in detail the number of trips for each year. Another option is to show the correlation between the weather condition with the number of trips.

Finally, the table shows the top 10 popular stations and the top 10 route that riders normally commute.

SHEET 3: Daily Analysis

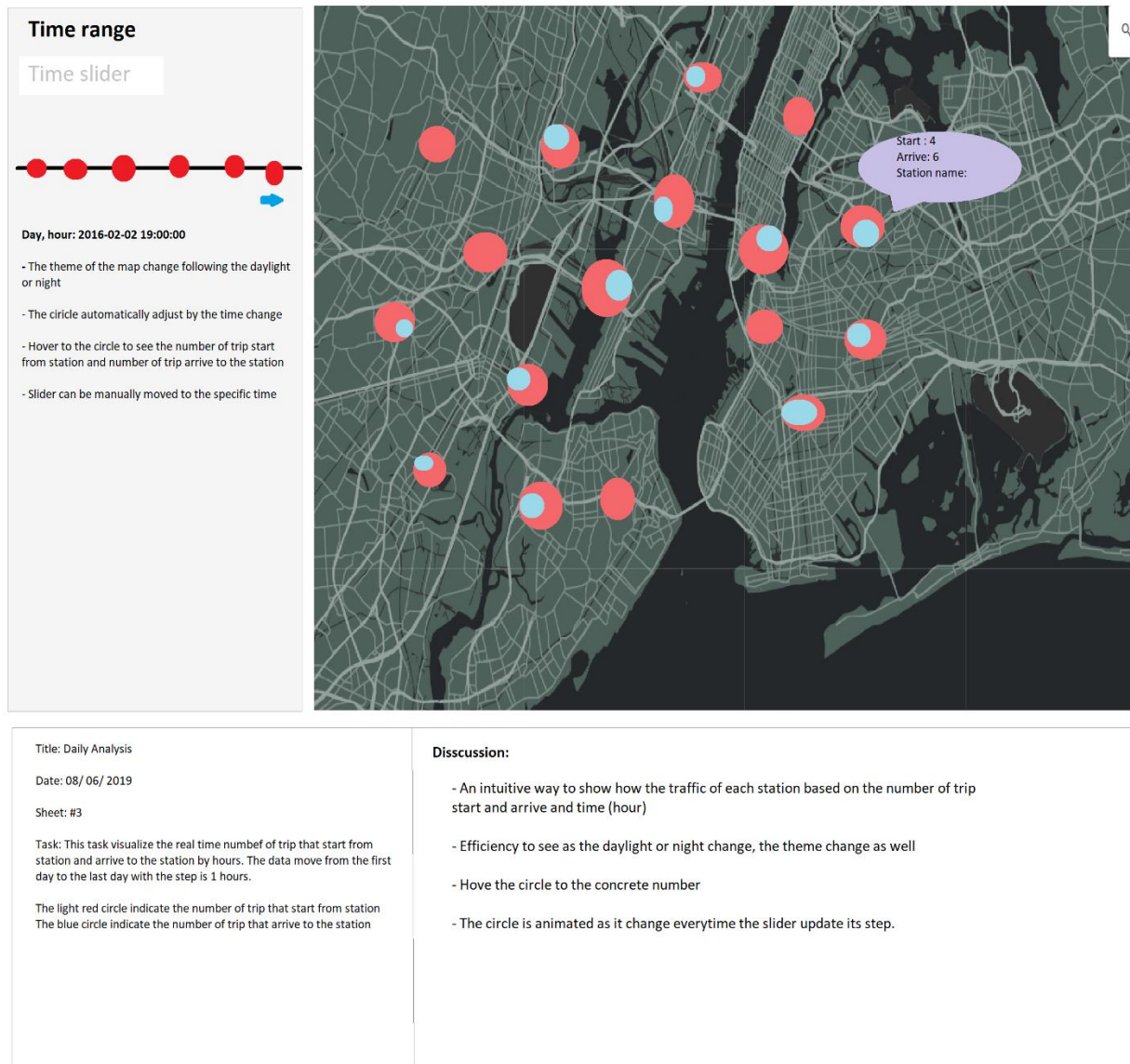


Figure 3:Five sheets design - sheet 3

In this design sheet, the interactive map will be incorporated with the animated map in order to illustrate the daily traffic of Austin's bikeshare system. The time slider will automatically run when the user hit the play button, or the user can easily drag the slider to the specific time.

The theme of the map will change from night theme to daylight theme when slider hit the appropriate time. For example, the theme is dark if the current hour is from 7 pm to 6 am of the following day.

Sheet 4: Station Analysis

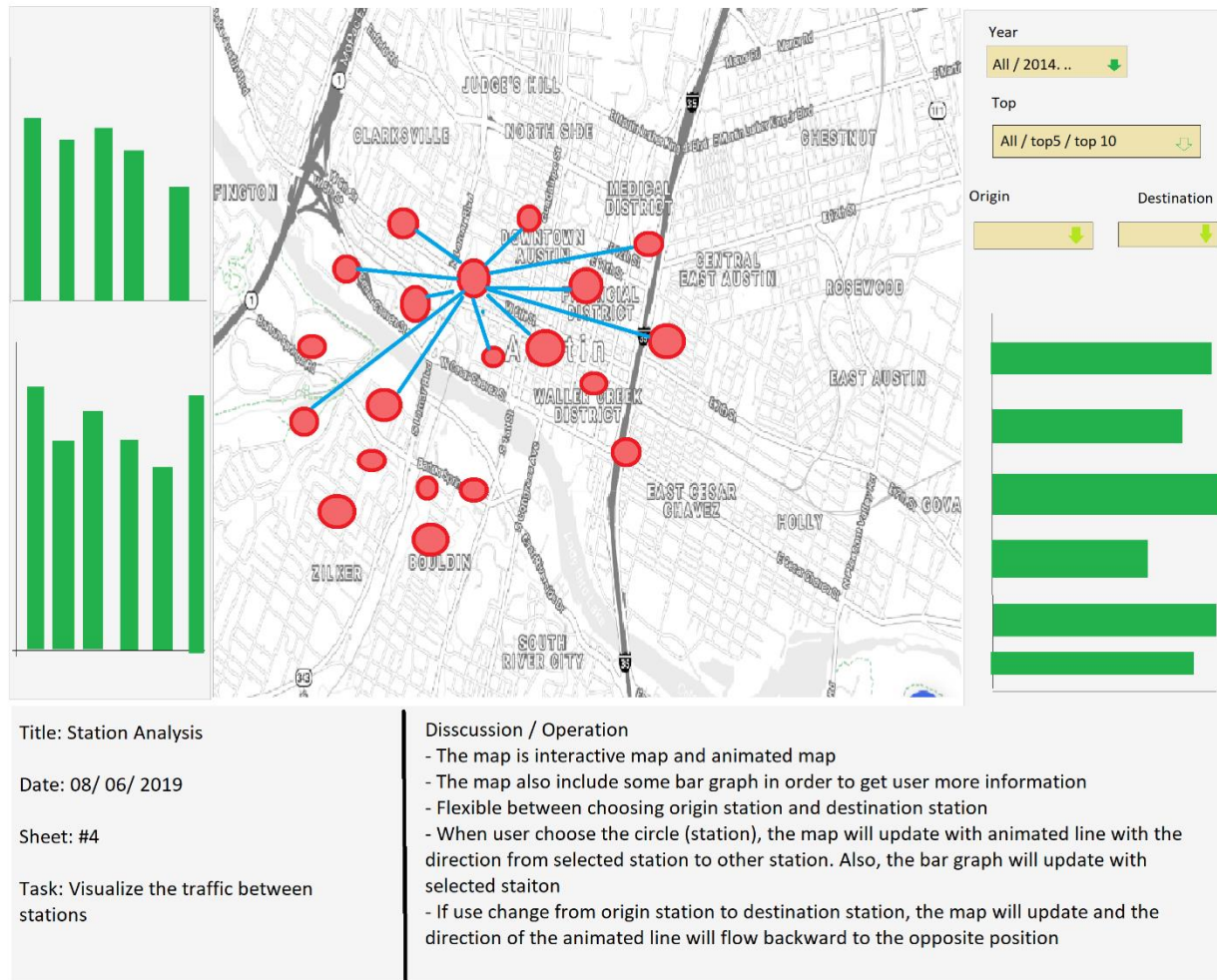


Figure 4: Five sheets design - sheet 4

The last sheet is another interactive and animated map which show the intuitive way of how is the traffic of the system at each station specifically.

In this page, user can select the station that similar to the previous page. However, in this page, instead of show in real time number of trips for each station, it will show the traffic from selected station to the other station by using the animated line to demonstrate the routes. If the user chooses the destination instead of the origin station, the map will be updated and the animated line will go backward from the destination station to origin station.

This page also shows some statistical bar graph that gives the user more information about the station they are currently selected.

The final sheet is realization sheet which shows the project structure and organizes the visualizations. There is 3 stages of visualization that is corresponding to the 3 pages. Therefore, a navbar menu will be the choice in order to move between those pages.

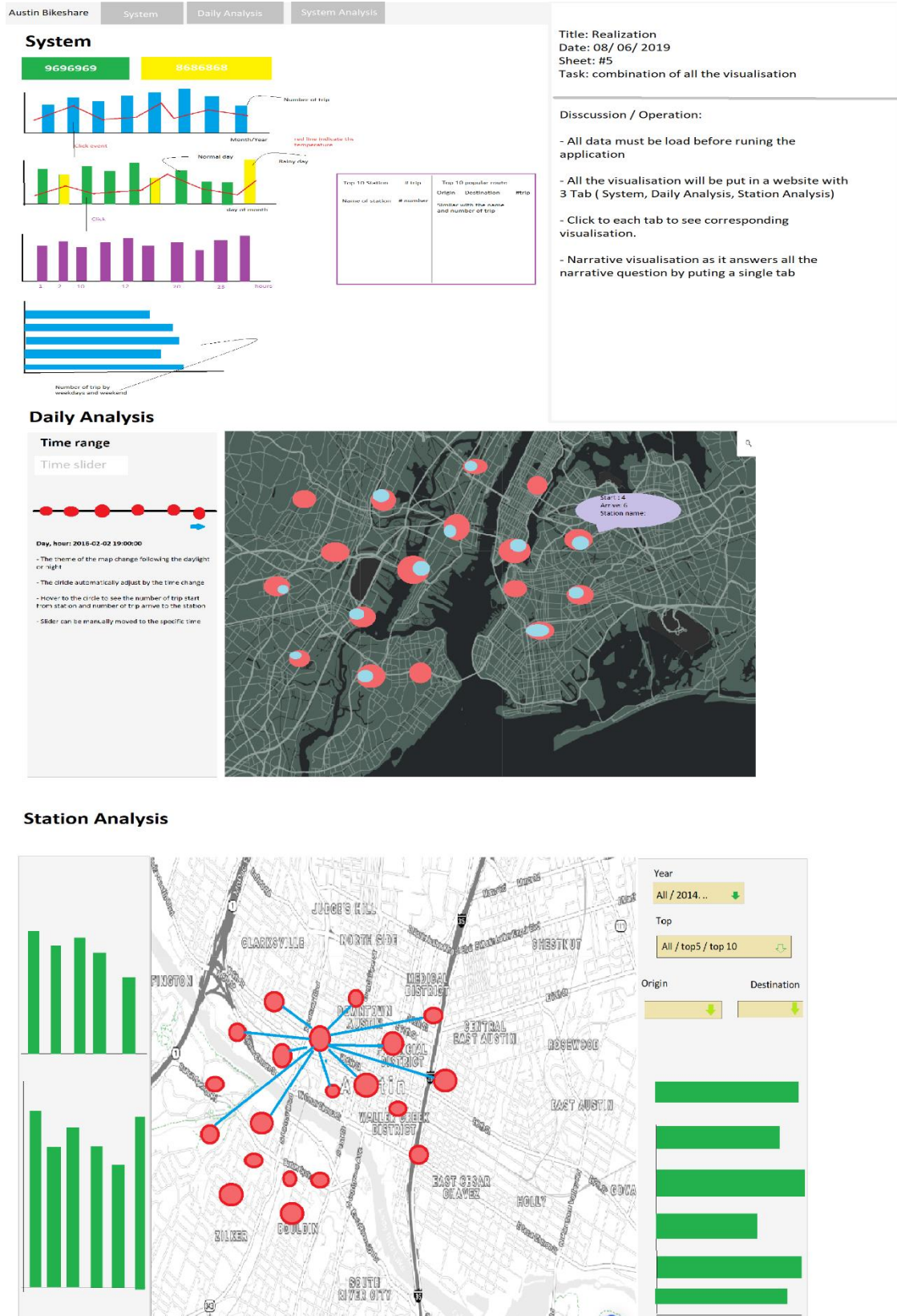


Figure 5: Five sheets design - sheet 5

4. Implementation.

This section will walk through the step to step in order to set up as well as run the visualization project. This visualization has been done by using intensively and an extensively larger number of libraries and preprocessing step. Therefore, it is good to get to know and familiar with all the libraries as well as the development environment of the project.

Environment Development:

Programming Language	Version
R	3.5.3
Javascript	ES6

Operating System: Window 10

Development Tool: RStudio Desktop (Open Source License)

Libraries install in RStudio

Title	Version	Description
shiny	1.2.0	A web framework for building web application using R
shinydashboard	0.7.1	Take advantage from shiny to build a dashboard.
ggplot2	3.1.1	A package for creating graphic visualization
dplyr	0.8.0.1	Use for data wrangling.
plotly	4.9.0	An iterative visualization library.
stringr	1.4.0	For string manipulation
leaflet	2.0.2	A library for creating an interactive web map
shinyWidgets	0.4.8	A small utility addon for shiny library
lubridate	1.7.4	Make easy to work with data and time
data.table	1.12.2	fast and memory efficient in aggregations and merging as the data are more than 500k row, the merging algorithm just does not work when there is not enough memory.
htmltools	0.3.6	A library for HTML generation and output
htmlwidgets	1.3	A framework for easily creating R binding to Javascript libraries.
leaflet.minicharts	0.5.4	add and update small charts on an interactive maps created with the package leaflet.

Library as a javascript plugin: Leaflet-Ant-Path

Though the leaflet library in shiny is powerful, however, it still lacks several plugins written in javascript that only support leaflet in web development (using javascript). Hence, a binding a plugin written by javascript to leaflet library in R needs in order to take immediate advantage of visualization. For more detail, this visualization uses “Leaflet Ant Path” library which is an animated visualization between points in the map.

More information on the link: <https://rubenspgcavalcante.github.io/leaflet-ant-path/>

In the following part of this section will instruct how to install the libraries as well as the environment development. This guide is based on Window 10 operating system, however, for Linux user and MacOS user, the steps are the same.

Step-by-step guides:

- Download and install R by following link: <https://cran.r-project.org/mirrors.html>
- Download and install RStudio choose the free version (open source license), link: <https://www.rstudio.com/products/rstudio/download/>
- The Leaflet Ant Path” library is already included the project under the folder with the name “ant_path”. For more detail and the newest version, refer to the link: <https://github.com/rubenspgcavalcante/leaflet-ant-path>
- After install R and RStudio on computer, unzip the file **bikeshare_app.zip** and go to the unzip folder. The files inside the folder should contain exactly in the figure below.

ant_path	6/6/2019 8:26 PM	File folder	
five_sheet_design	6/8/2019 5:07 PM	File folder	
.Rhistory	6/6/2019 11:52 PM	RHISTORY File	10 KB
AdminLTE	5/9/2019 4:52 PM	Cascading Style Sh...	110 KB
app	6/9/2019 2:41 PM	R File	48 KB
austin_bike_share_with_hour_3	5/29/2019 8:57 AM	Microsoft Excel Co...	108,962 KB
leaflet.polylineDecorator	6/2/2019 1:39 PM	JavaScript File	17 KB
shinydashboard	5/9/2019 4:52 PM	Cascading Style Sh...	3 KB
styles	6/5/2019 3:10 PM	Cascading Style Sh...	2 KB

Figure 6: Project Structure

The AdminLTE.css and styles.css are CSS file use for styling the visualization. The leaflet.poylineDecorator is javascript use for draw interactive the line in the map.

- Open the file app.R

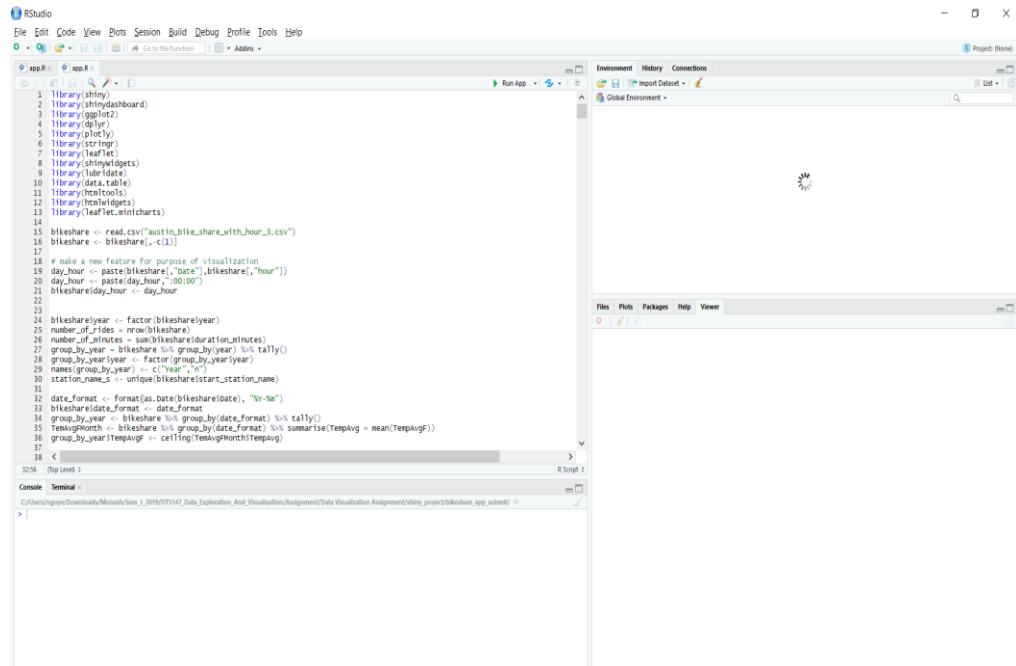


Figure 7: Overview of the project in RStudio

- Now, go to console and type the following command to install necessary libraries:
 - **install.packages("shiny")**
 - **install.packages("shinydashboard")**
 - **install.packages("ggplot2")**
 - **install.packages("dplyr")**
 - **install.packages("plotly")**
 - **install.packages("stringr")**
 - **install.packages("leaflet")**
 - **install.packages("shinyWidgets")**
 - **install.packages("lubridate")**
 - **install.packages("data.table")**
 - **install.packages("htmltools")**
 - **install.packages("htmlwidgets")**
 - **install.packages("leaflet.minicharts")**
- Click to the play button "Run App" to run the project.



- If all the steps have been successful, the running application will be shown.

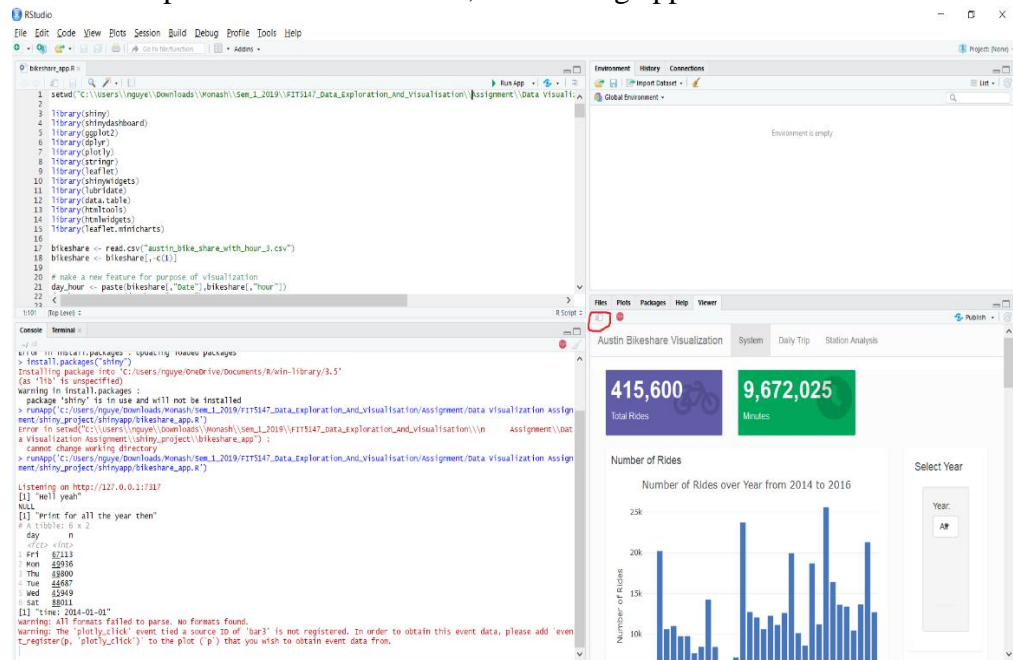


Figure 8: Running project on RStudio

- Click to the button which is a circle to show the entire project on the website as look like this figure:

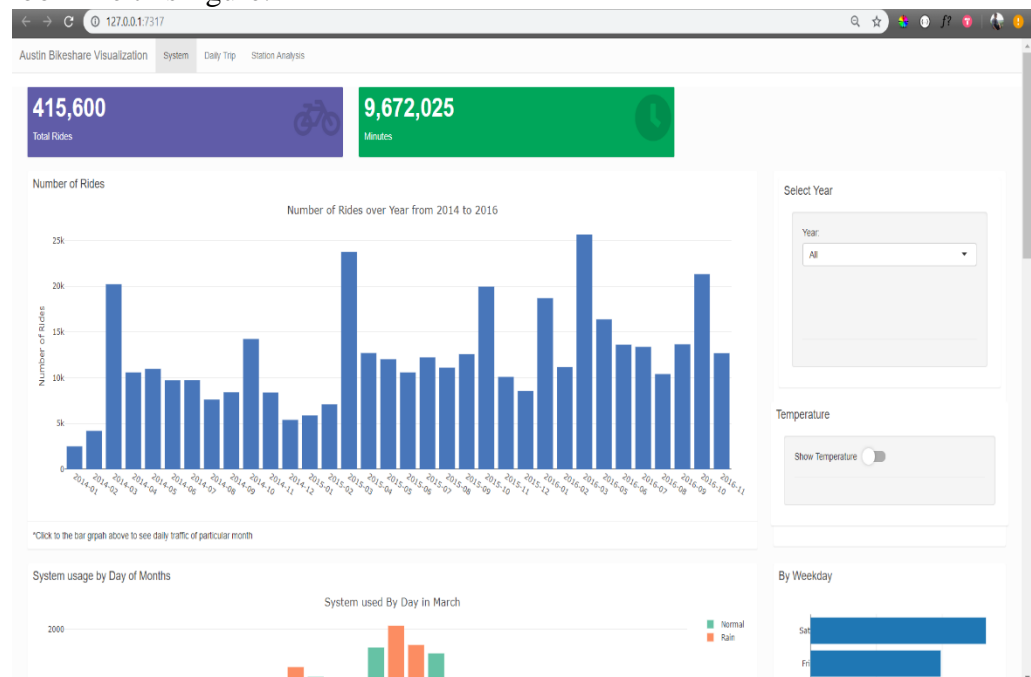


Figure 9: Running project on website

5. User guide.

This section is guiding on how to interact with the visualization project. When user run the web application, the first page is the overview dashboard, which contains information regarding number of trips, how much people use the system.

The first graph is the number of rides by month over the year from 2014 to 2016. There is also a filter to select a specific year rather than view the entire timeline. Another option use can use is a temperature toggle, by default, it will be turned off, if toggle on, a red line will be added into the graph indicate the average temperature.



Figure 10: The first page 1

Click to the blue bar plot, the next graph will show up which indicate the number of the trip on a particular month of the year.



Figure 11: The second page 2

The graph above illustrates the number of trips by a specific month, which is, in this case, in March 2015. The red line indicates the average temperature of each day of the month while the organ color show if there was a raining day, the green indicates the normal day.

The next plot shows the total number of trips between weekdays and the weekend of the year and month.

Click on the bar plot of a specific day, another graph will be shown which is tell the traffic hour of that day, following the number of the trip by its duration.

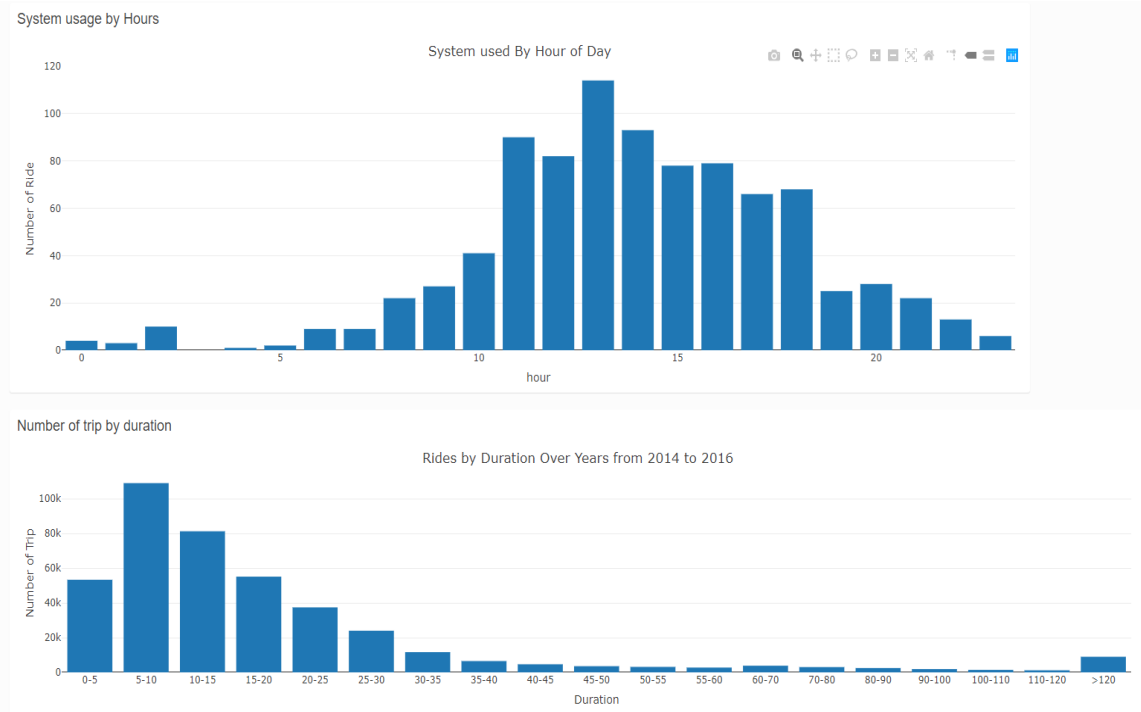


Figure 12: The first page 3

Finally, in the dashboard, there is a table showing the top 10 popular stations as well as the top 10 routes that riders normally commute.

Top 10 Popular Stations		Top 10 Popular Routes		
	# of trip	Origin	Destination	# of trip
5th & Bowie	19942	5th & Bowie	4th & Congress	2079
City Hall / Lavaca & 2nd	19592	4th & Congress	5th & Bowie	1896
4th & Congress	18292	5th & Bowie	City Hall / Lavaca & 2nd	1835
Convention Center / 4th St. @ MetroRail	17421	Republic Square @ Guadalupe & 4th St.	5th & Bowie	1664
2nd & Congress	17309	3rd & West	City Hall / Lavaca & 2nd	1503
Riverside @ S. Lamar	15045	City Hall / Lavaca & 2nd	5th & Bowie	1471
Davis at Rainey Street	14732	Convention Center / 4th St. @ MetroRail	City Hall / Lavaca & 2nd	1466
Capitol Station / Congress & 11th	13352	South Congress & Academy	2nd & Congress	1434
Rainey St @ Cummings	13258	5th & Bowie	Convention Center / 4th St. @ MetroRail	1377
Pflugger Bridge @ W 2nd Street	12845	Rainey St @ Cummings	City Hall / Lavaca & 2nd	1306

Figure 13: The first page 4

In this dashboard page, the visualization is already answered the first forth question which state in the first section:

- Overview of the entire system (how many customers, how many trips by year).
- How is the traffic of the system by year, month, day, and hours, weekdays and weekend?
- Does average temperature affect the customer's usage?
- What is the busiest station, and what are popular routes?

Now, in order to answer the last two questions, two tabs have been given which is Daily Trip and Station Analysis.

Moving to Daily Trip tab, there will be a map of Austin city and a panel to show more information in detail.

The panel consists of two elements, the first element is a timeline slider which shows the time from the first day of 2014 to the last day of 2016. User can easily move the slider to see the change in the map. When the user hits the play button below the slider, the time slider will automatically move and the map will automatically change to show the data by hour of the day.

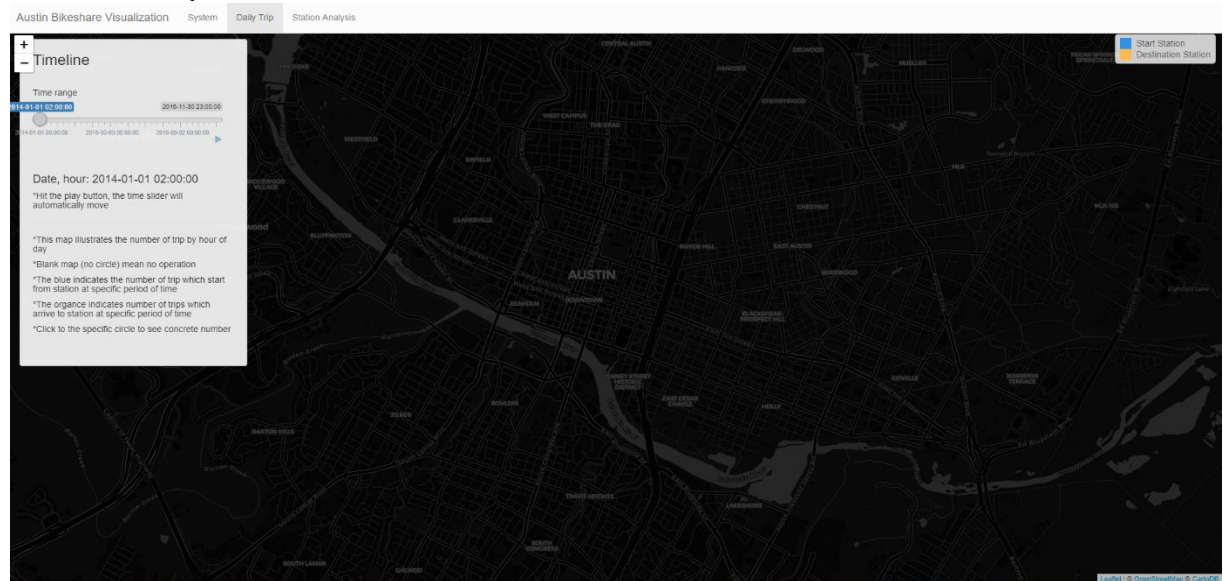


Figure 14: The daily analysis with night theme

The second element is shown what is the day and hour of that day that slider is currently in along with the user guide.

The time slider moves day by day with step is 1 hour. If the hour is on the daylight, the map will change to light theme. The circle indicates number of trips start from the station (blue color) and number of trips arrive at the station (orange color). The circle automatically changes when the slider moves.

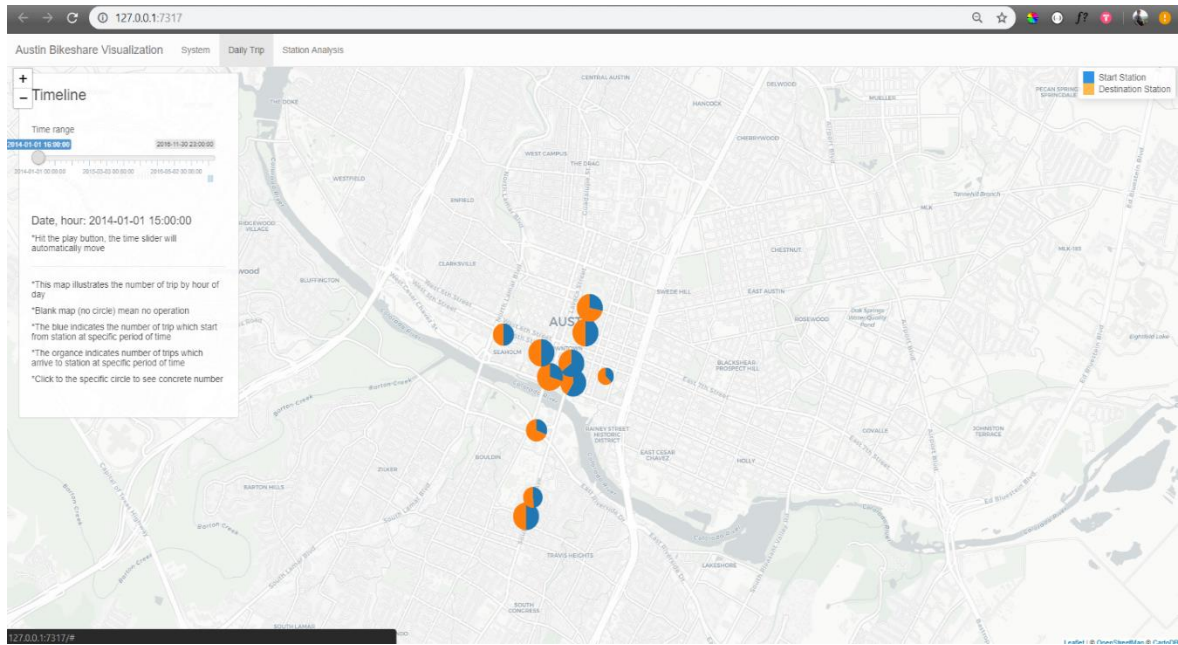


Figure 15: The daily analysis with daylight theme

Another figure illustrates the night traffic of the system.

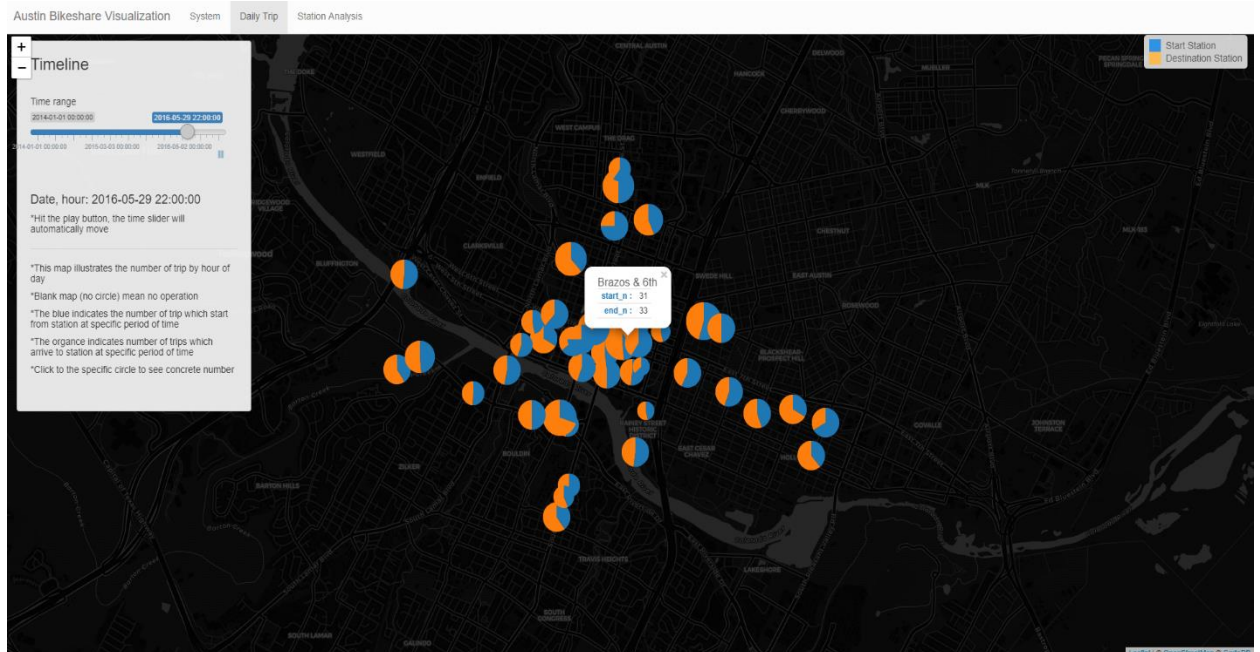


Figure 16: The daily analysis in dark theme

This visualization is already answered for the fifth question, which is the daily traffic of each station (including origin station and destination station). Manager can take advantage of this animated graph to know the busiest time of the station, so they can adjust the number of bikes at each station to increase the efficiency.

The tab Station Analysis will answer the last question

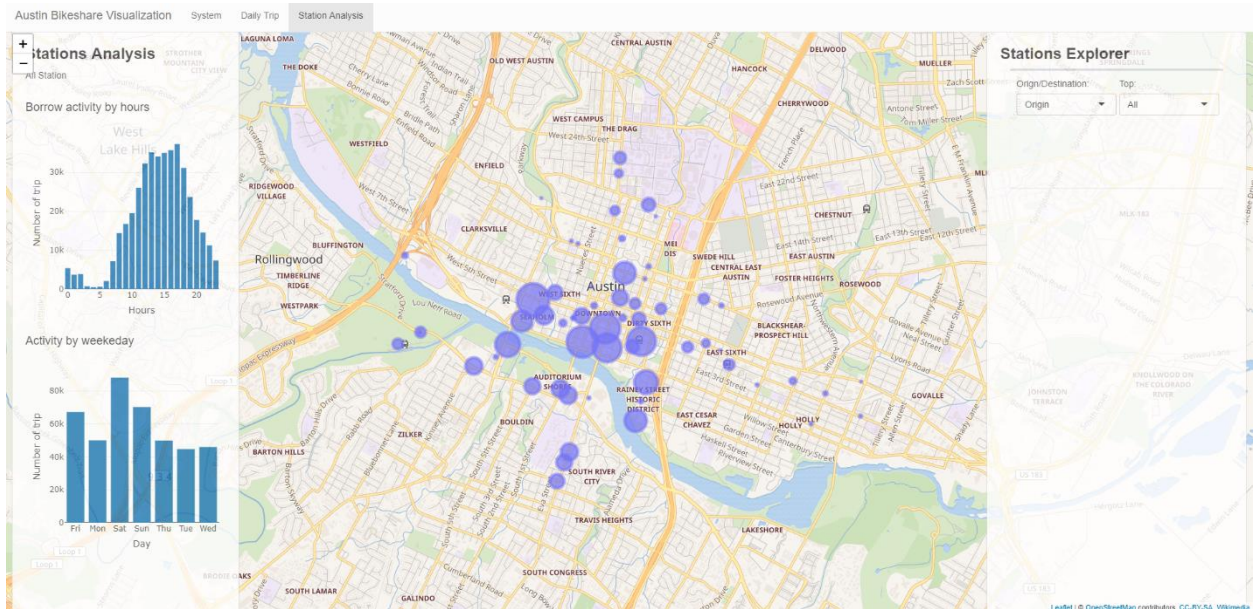


Figure 17: The station analysis

This map consists of three main elements. The first element shows some statistic information includes the borrowing activity by hours and the borrowing activity by weekdays of each station.

The second panel let the user choose two option, either user would like to see the origin station of starting station as well as filter by the number of station user would like to see.

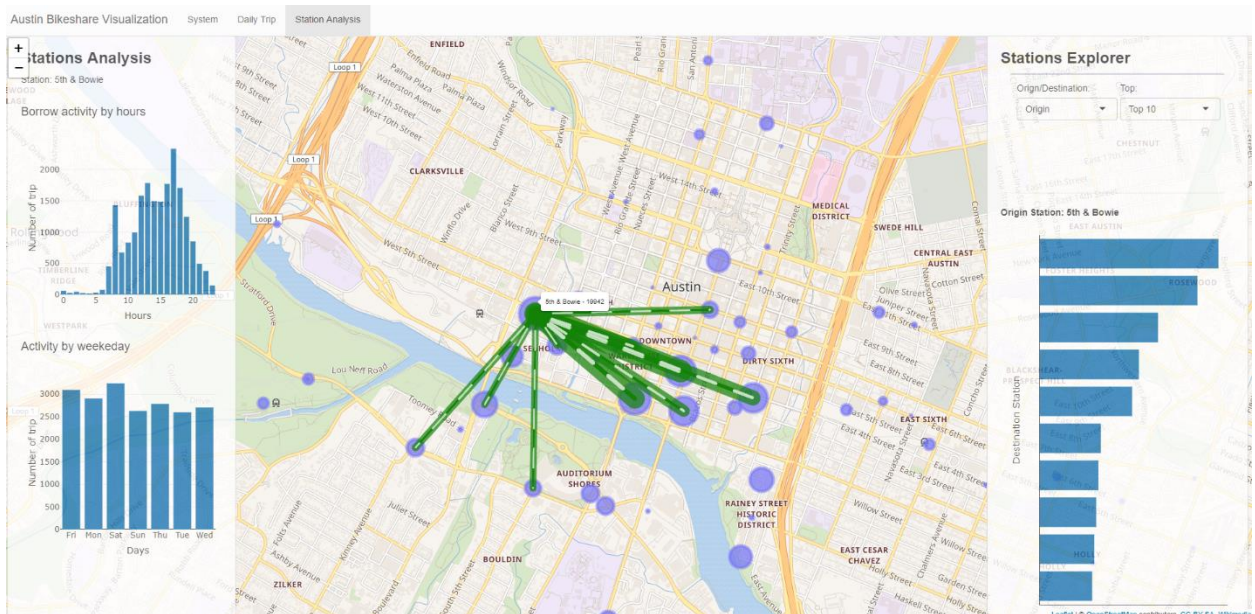


Figure 18: The station analysis with origin station animated line

When hitting to the circle, the animated map will show up to visualize the route traffic of the selected station. If the user selected the Destination Station instead of Origin Station, the map will update like the figure below.

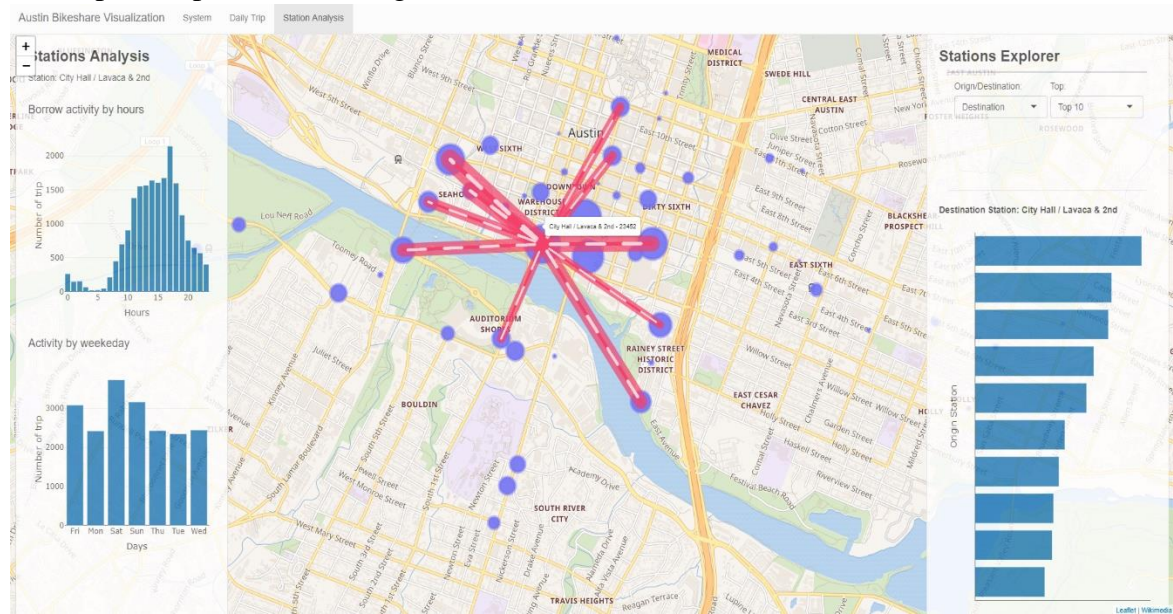


Figure 19: The station analysis with destination station animated line

The traffic now will go backward from all the start station that has destination is selected station, the line also changes the color, and the flow direction also changes to adapt to the filter.

6. Conclusion.

This visualization project has been addressed the questions stated in the first section of the report. Overall, the interactive visualization gives the managers of the bikeshare system an intuitive way of how the Bikeshare system work, hence, improve the efficiency in terms of partition the number of bikes for each station for a specific time of day. In addition, taking advantage of the visualization, it helps to reduce waste of resources by unbalance distribution rather than balance the number of bikes for each dock. This visualization emphasizes the narrative question by going more on the detail of each specific element of the report. As this project has been done by using Shiny with intensively and extensively support libraries, however, it is difficult to get the animated map done. Though the using of Leaflet library gives an easy way to plot the iterative map visualization, however, the lack of support plug-in in Leaflet, especially for making the tab Station Analysis is tough. Some libraries only support Leaflet with Javascript only and there is no support for Leaflet in Shiny, in order to take advantage from such useful plugin, this project has introduced the convenient way to integrate Javascript based plugin into Shiny to improve the graphics visualization.

7. Appendix

The visualization project has been published to cloud.

Checkout the link to see live project

http://nguyenkien1402.shinyapps.io/bikeshare_app_submit

Update: due to some confliction in terms of virtual machine instance, the live project on the link above has occurred some minor error, which could not show the top 10 popular stations and top 10 popular routes.

To archive the full features of visualization project, please run the project locally by following the Implementation section.

References

National Association of City Transportation Officials (NACTO), 2017. NACTO's new reports, Shared Micromobility in the U.S.:2018. Retrieved from <https://nacto.org/sharedmicromobility-2018/>