Table of content

1. Introduction

The city of Portland is organizing an annual bicycle count program. The program has collected the data of the number of bicycles by volunteers at various locations throughout the city in paper form. Also, this program collects data describing trip purpose and cycling habits through in-person surveys. The target of this project is to design a database and data entry interfaces that will be used by program staff to record, view, and summarize the count results in R by Shiny application. Also, this program collects data describing the trip purpose and cycling habits through in-person surveys.

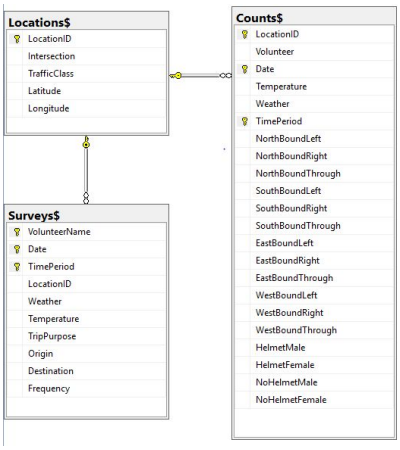
2. Database

2.1. Demand and Assumptions

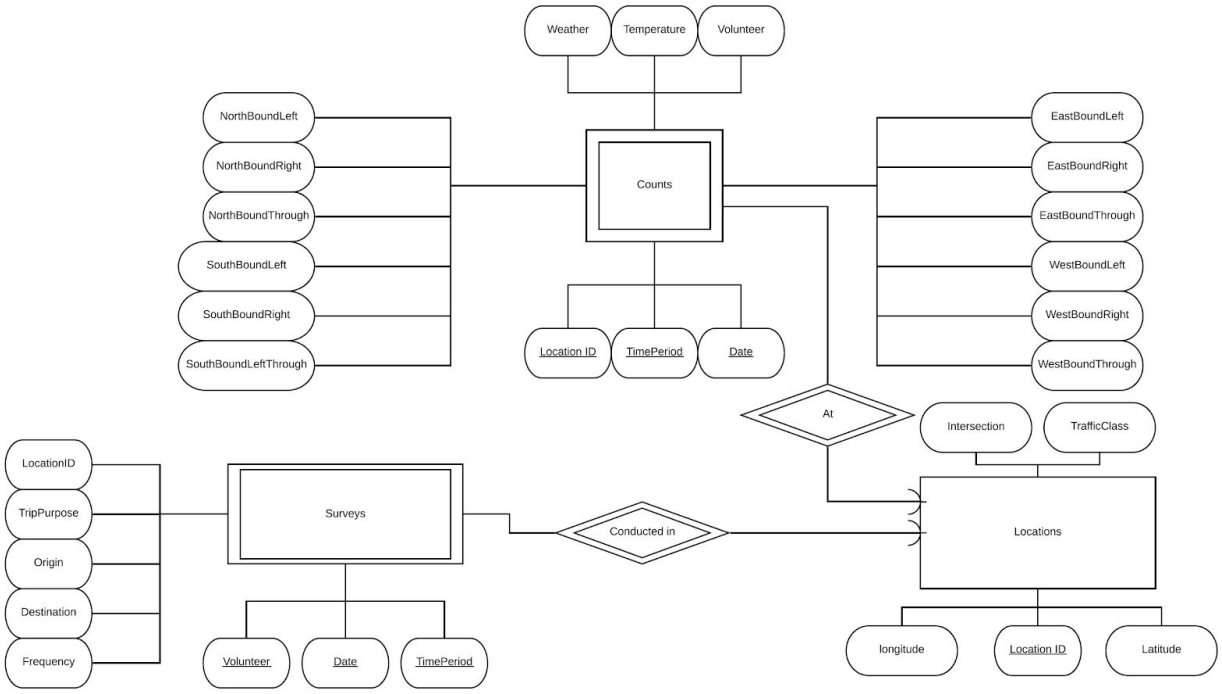
- The volunteer can just choose to submit a Rider survey or a Count survey at a time or submit both.
- The volunteer can take Count surveys or Rider surveys several times in different dates.
- Input data for both surveys will be recorded in respective tables.
- If the location is deleted from the dataset, the corresponding surveys would be removed.
- Multiple Rider surveys can be conducted at one location.
- Multiple Count surveys can be conducted at one location.

2.2. Relational schema

- Counts (LocationID, TimePeriod, Date, Weather, Temperature, Volunteer, NorthBoundLeft, NorthBoundRight, NorthBoundThrough, SouthBoundLeft, SouthBoundRight, SouthBoundThrough, EastBoundLeft, EastBoundRight, EastBoundThrough, WestBoundLeft, WestBoundRight, WestBoundThrough)
- Locations (LocationID, Latitude, Longitude, Intersection, TrafficClass)
- Surveys (Date, TimePeriod, Volunteer, LocationID, Weather, Temperature, TripPurpose, Origin, Destination, Frequency)



2.3. ER diagram



3. User manual

3.1. Attributes in relations

In this user manual, we can see each attribute's name, type, length, description, and constraint. The relations among these three tables are shown below in the constraint.

The relation between table Counts and Locations is Counts.LocationID = Locations.LocationID. The relation between Surveys and Locations is Surveys.LocationID = Locations.LocationID.

Counts entity

Column name	Data Type	Description	Constraint
LocationID	int	Location ID number, a unique value for each location	Key primary, Foreign key constraint (Locations.LocationID)
Volunteer	varchar(225)	Volunteer recorder name	
Date	date	Date	Key primary
Temperature	int	Temperature (F)	
Weather	varchar(225)	Weather condition	

TimePeriod	varchar(225)	Time period of the day(e.g. 7-9 am)	Key primary
NorthBoundLeft	int	Number of northbound turn left bicycle	
NorthBoundRight	int	Number of northbound turn right bicycle	
NorthBoundThrough	int	Number of Northbound go straight bicycle	
SouthBoundLeft	int	Number of southbound turn left bicycle	
SouthBoundRight	int	Number of southbound turn right bicycle	
SouthBoundThrough	int	Number of southbound go straight bicycle	
EastBoundLeft	int	Number of eastbound turn left bicycle	
EastBoundRight	int	Number of eastbound turn right bicycle	
EastBoundThrough	int	Number of eastbound go straight bicycle	
WestBoundLeft	int	Number of westbound turn left bicycle	
WestBoundRight	int	Number of westbound turn right bicycle	
WestBoundThrough	int	Number of westbound go straight bicycle	
HelmetMale	int	Number of male wearing helmet	
HelmetFemale	int	Number of female wearing helmet	
NoHelmetMale	int	Number of male not wearing helmet	
NoHelmetFemale	int	Number of female not wearing helmet	

Locations entity

Column name	Data Type	Description	Constraint
LocationID	int	Location ID number, a unique value for each location	Key primary
Intersection	varchar(225)	Location-specific name	
TrafficClass	varchar(225)	Traffic class(low, medium, high)	
Latitude	int	Latitude	
Longitude	int	Longitude	

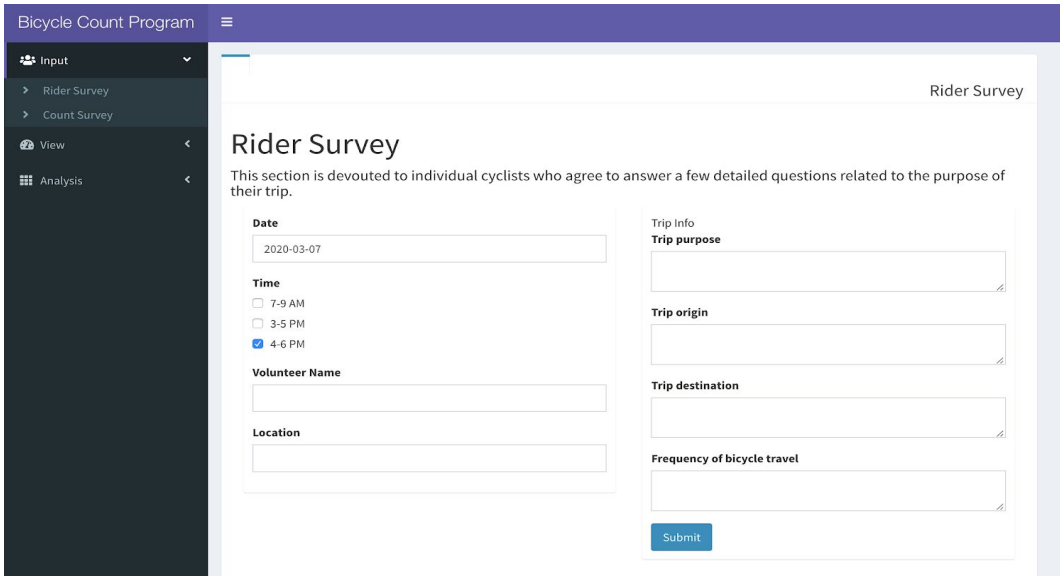
Surveys entity

Column name	Data Type	Description	Constraint
VolunteerName	varchar(225)	Volunteer recorder name	Key primary
Date	date	Date	Key primary
TimePeriod	varchar(225)	Time period of the day(e.g. 7-9am)	Key primary
LocationID	int	Location ID number, a unique value for each location	Foreign key constraint (Locations.LocationID)
TripPurpose	varchar(225)	Trip purpose	
Origin	varchar(225)	Trip origin	
Destination	varchar(225)	Trip destination	
Frequency	varchar(225)	Frequency of bicycle travel	

3.2. Functions

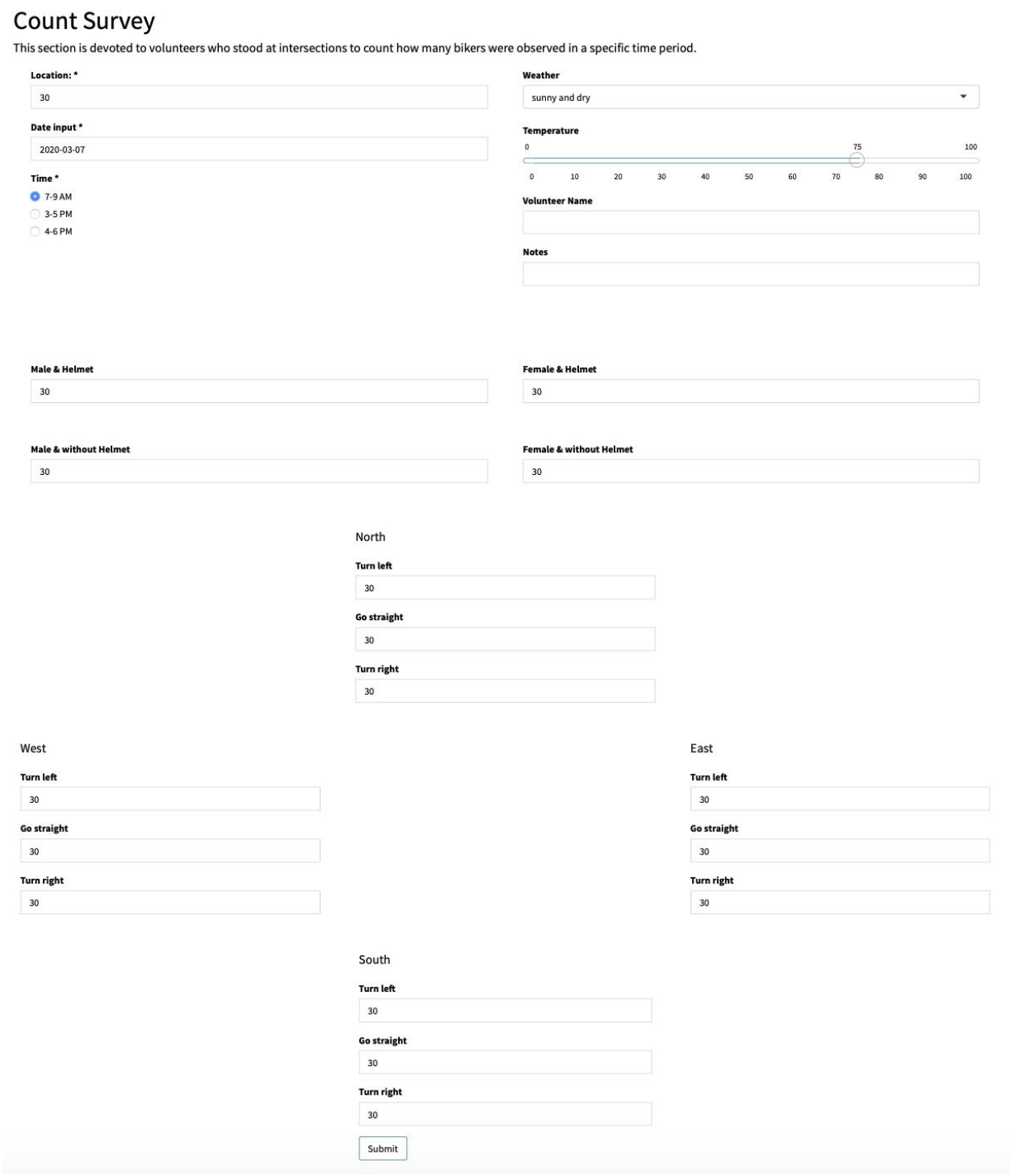
3.2.1. Rider survey input

Rider survey Input page permits users to input data including date and time of the survey, volunteer name, location, trip purpose, trip origin, trip destination, frequency of bicycle travel. Once the user clicks the submit button, the input data will synchronize to the Table Survey in the SQL.



3.2.2. Count survey input

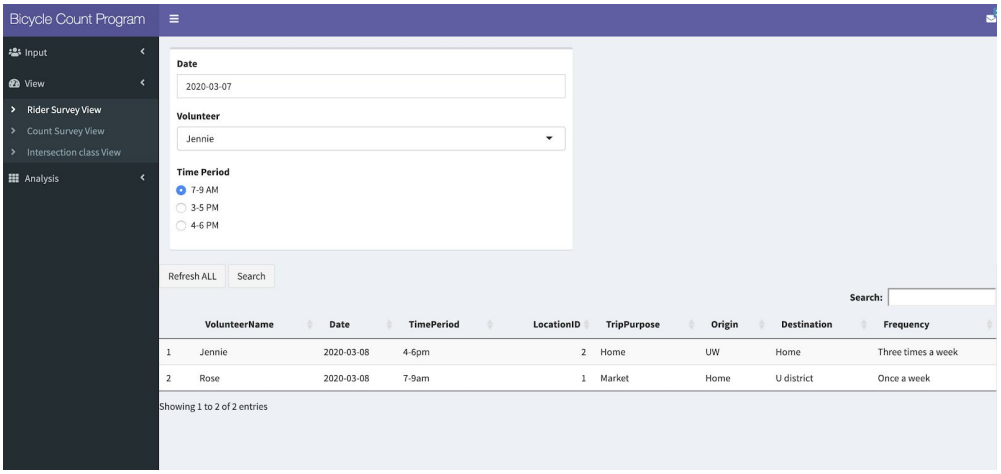
The volume of bicycles of different movements in each bound of the intersection and helmet wearing ratio are mainly input in this part and separately distributed on different panels, Basic Info, Helmet Count and Bicycle Count. Besides, some supplement information, such as the name of the intersection, date, time period, weather, and the name of the recorder, are also input on this page.

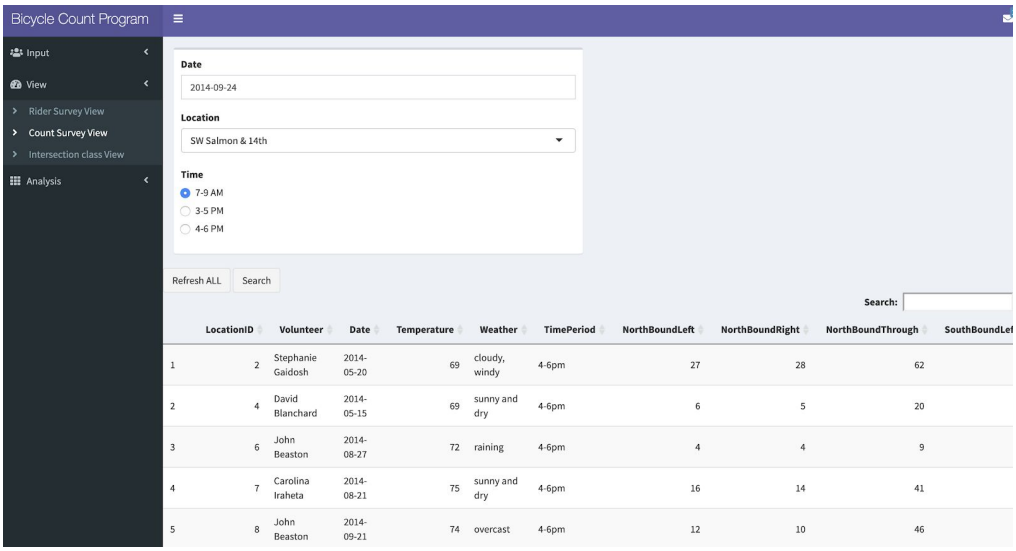


3.2.3. Data management

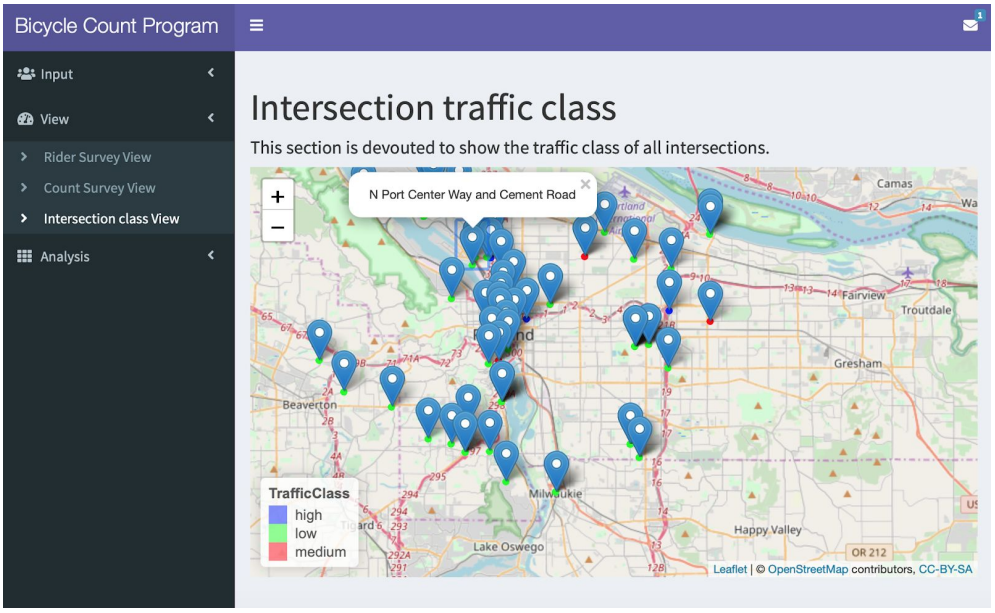
- View and delete
  Users can input data to the database and filter the data based on the date and time period to view. Two types of data are available for users to access, which are Rider Survey View, Count Survey View. The button “Refresh ALL” can refresh data and show

the whole dataset. While the button “search” can show the result table of specified conditions users entered above.





- The View of the Intersections’ Class
  In the third panel of View, we used the Leaflet plot to show the traffic class of intersections, which recorded longitude and latitude in Portland. The green points mean low traffic volume intersections, the blue points mean high volume intersections, while the red points mean medium volume intersections. However, it should be mentioned that the number of volumes may be counted at different times and dates. Thus, it cannot be regarded as the volume condition in a specific time period.

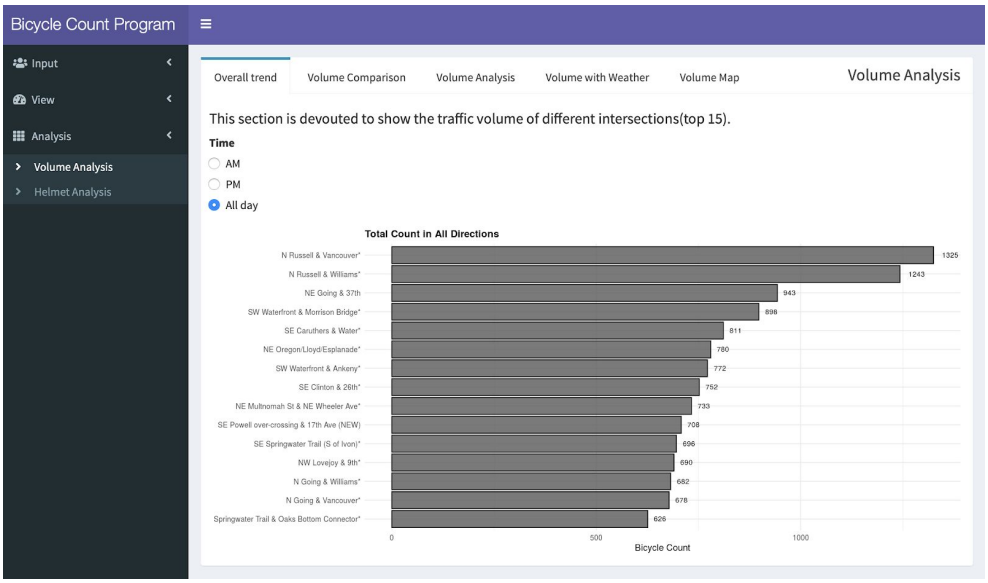


3.2.4. Analysis and Visualization

3.2.4.1. Volume analysis

- Overall trend

  The first panel of volume analysis is showing the traffic volume of different intersections (top 15 intersections). The three histograms respectively represent the volume in 4:00pm - 6:00pm, 7:00am - 9:00am, and total volume in these two periods, namely the plots show the morning peak, evening peak volume, and traffic volume in four hours.

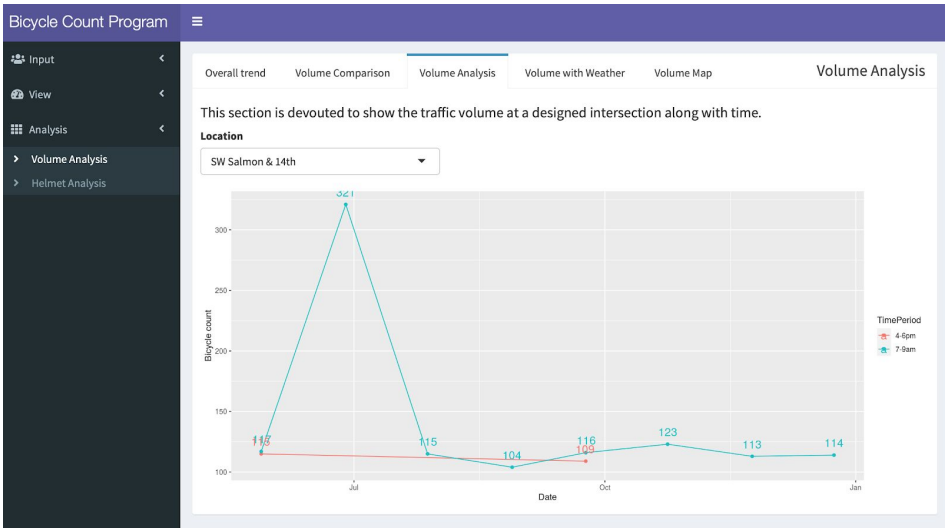


- Volume Comparison

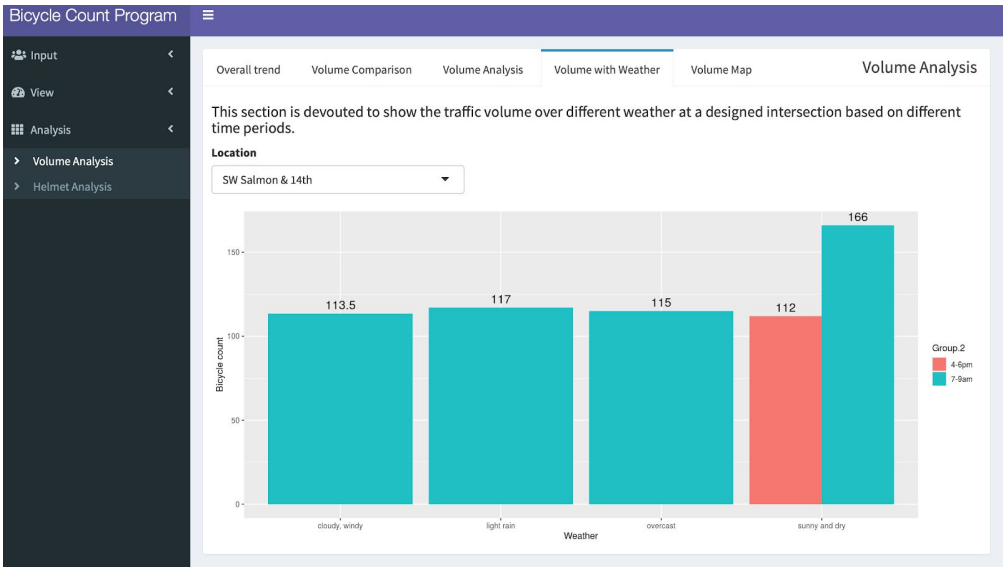
  The second panel of volume analysis is showing the traffic volume of different intersections at a specific date based on different time periods. First of all, we need to

select the date. Then, we can get the histogram showing that the maximum volume at 4:00pm - 6:00pm is in the NE Going & 37th intersection up to 943 bicycles. we also add a table to display the detailed information of the intersection. Besides, we also design the search box, which allows us to search the specific crossing.

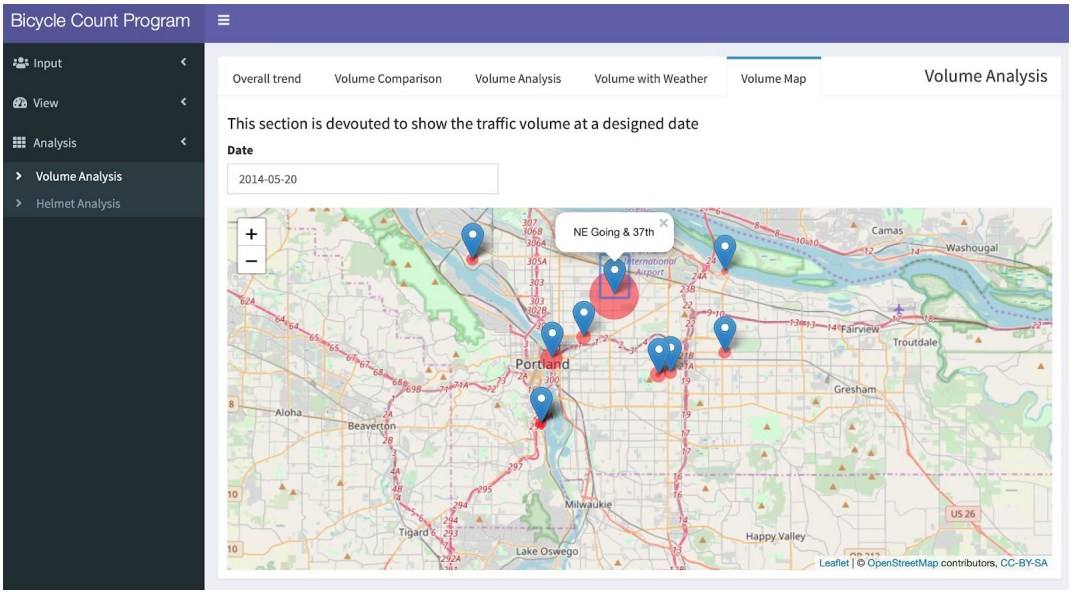
- Volume Analysis
  The third panel of volume analysis is volume analysis at a designed intersection along with time. After choosing the intersection we want to analyze in the select box, there will be a line chart shown below. It clearly reveals the variation of the volume in one intersection over time. Besides, we distinguish the data into two categories, the volume in the morning peak and evening peak. The plot shows little differences between these two periods. However, the volume in July significantly higher than any other month in 2014.



- Relationship between volume and weather
  The fourth panel of volume analysis is the variance of volume over different weather at a designed intersection based on different time periods. As same as the previous panel, we also need to choose a specific intersection in this part. Then the results will feedback on a bar chart showing different volume amounts in different weather over different time periods as the plot shown below.

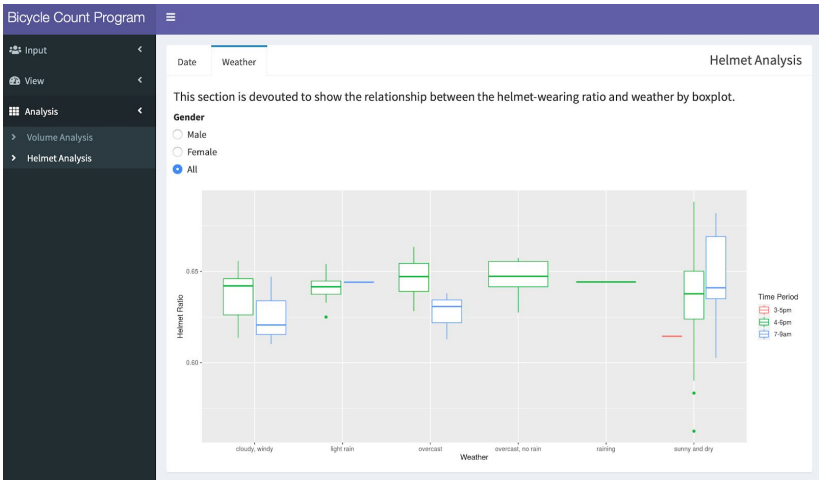


- Volume Map
  The fifth panel of volume analysis is the daily volume map. After setting the date, the volume of each intersection in the research time (i.e. in 4 hours) will be displayed on the map. What's more, when the mouse moves to the intersection, the name of it will float above the mark. The scale of total volumes of different intersections on that day, as the picture is shown below, is represented by the size of the red circle. A bigger circle scale means a larger amount of bikes.



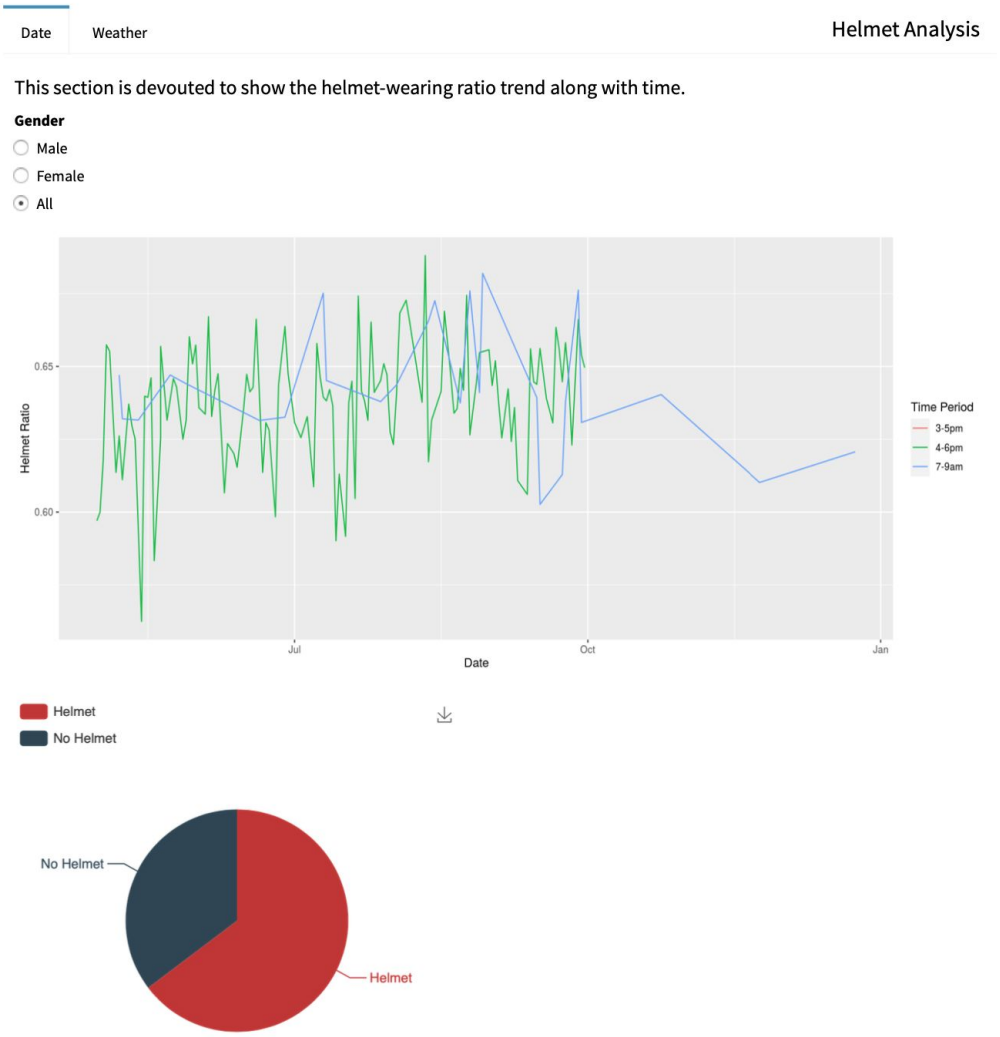
3.2.4.2.Helmet Analysis

- Relationship between helmet wearing ratio and weather
  The second panel in the helmet analysis is studying the relationship between the helmet-wearing ratio and weather. The input is still gender, then, there will be a box plot displayed below.



- Helmet-wearing ratio trend along with time
  The first panel in the helmet analysis is studying the helmet-wearing ratio trend along with time. First of all, we need to select one of three gender categories. After that, the plot will mainly show the change of helmet-wearing ratio in two time periods during different dates in a year, just like the figure. It is easy for us to observe that the ratio fluctuates between 60 percent and 67 percent. Moreover, A pie chart below was divided into sectors that each represents a proportion of the people wearing helmets or not. The entire circle encompasses all of the cyclists who were recorded, and the

sectors represent a percentage of wearing helmets or not, it shows that 64.71% of all cyclists in the research were wearing helmets and the rest were not.



4. Shiny application

https://leahyu.shinyapps.io/Team08/

5. Summary

5.1. Accomplishment

Throughout this project, we learned how to import data into SQL, as well as manipulate the data within SQL to be compatible with linkage to R. We also were able to successfully manipulate and code necessary queries to add Primary keys and Foreign keys to each table as deemed appropriate. The coding for linkage with R and SQL was challenging and complex but it was done through team coordination. Due to certain

conditions, we were able to coordinate and plan out a team presentation by recording our parts and voices into a single format.

5.2. Duties of each team member

- Roberto Gomez - Importing data to SQL, establishing Primary and Foreign Keys, ER diagram, Relational schemas, and making an introduction PowerPoint for the presentation.
- Roy Kim - Importing data to SQL, establishing Primary and Foreign Keys, ER diagram, Relational schemas, and making an introduction PowerPoint for the presentation.
- Lu Liu - Process data in Excel, Shiny part (Visualization), Presentation, Video
- Lu Yu - All code part, Presentation