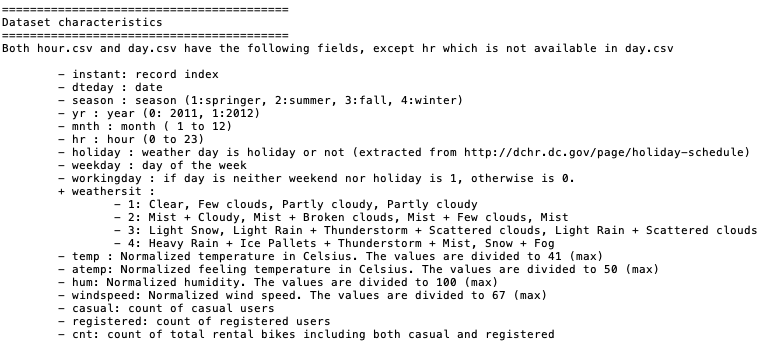
**Introduction**

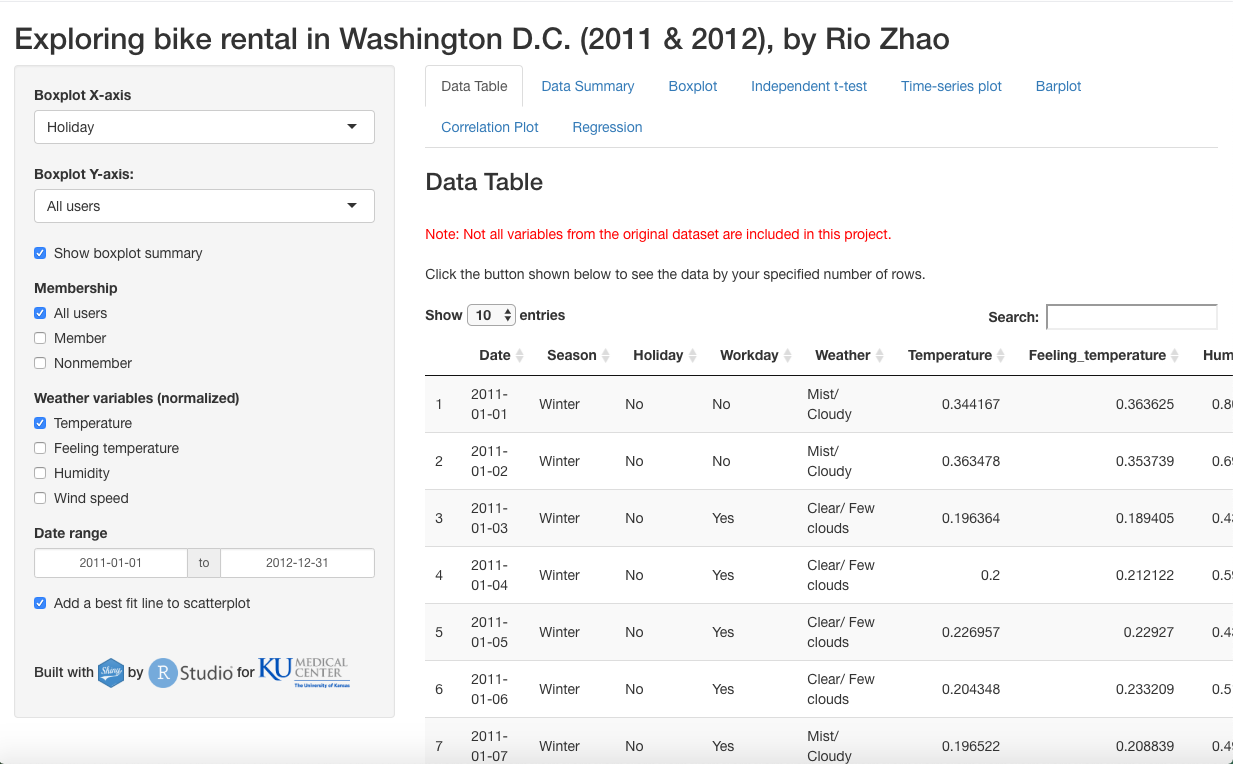
This shiny app was built using dataset obtained from UCI’s Machine Learning Repository. The link to this repository is <https://archive.ics.uci.edu/ml/datasets/bike+sharing+dataset>. In building this shiny app, the day.csv was used. I have included a screenshot of the data dictionary below to provide information about the variables.



The goal of this shiny app is to explore bike rental count in Washington D.C. from 2011 to 2012 using visualization, along with some statistical analyses in eight sections/ tabs. At the end of app exploration, you will start to develop an idea of the factors that affect bike rental among different groups of users. I have provided some descriptions and screenshots of the app to walk you through what it does and explain its output.

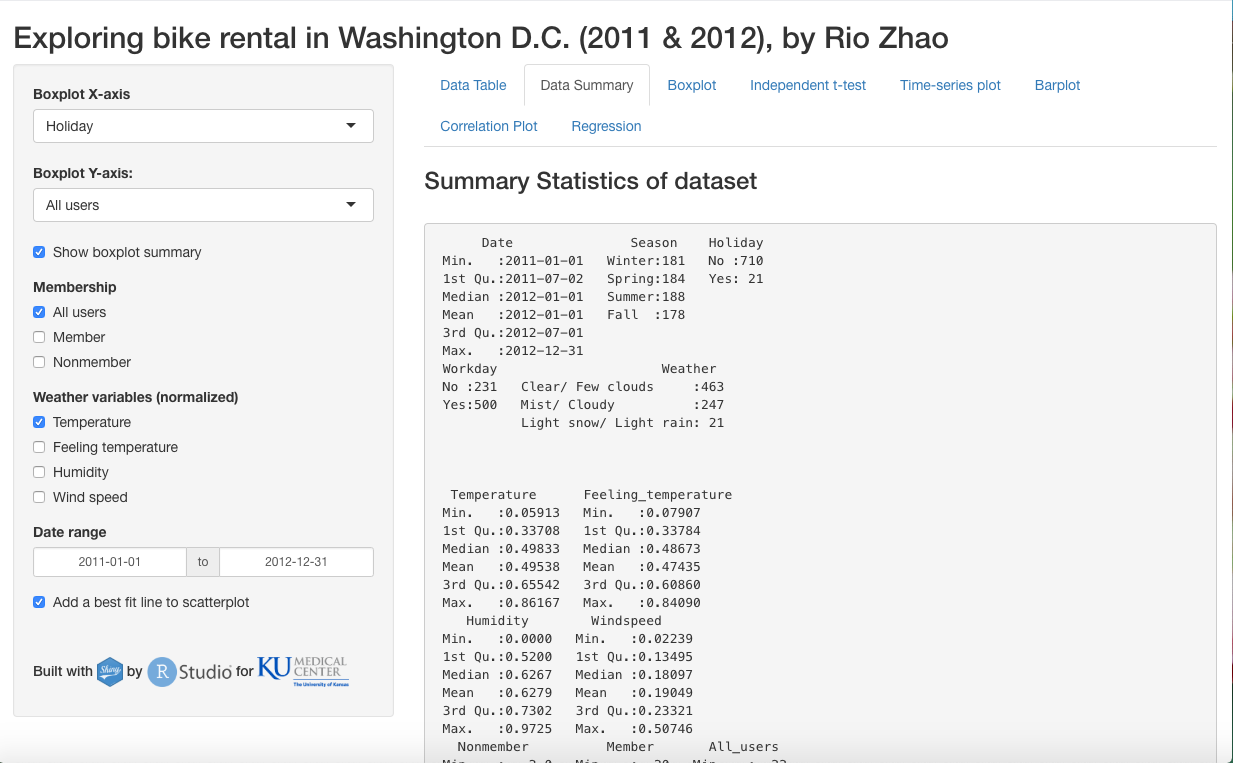
**Tab1 Data Table**

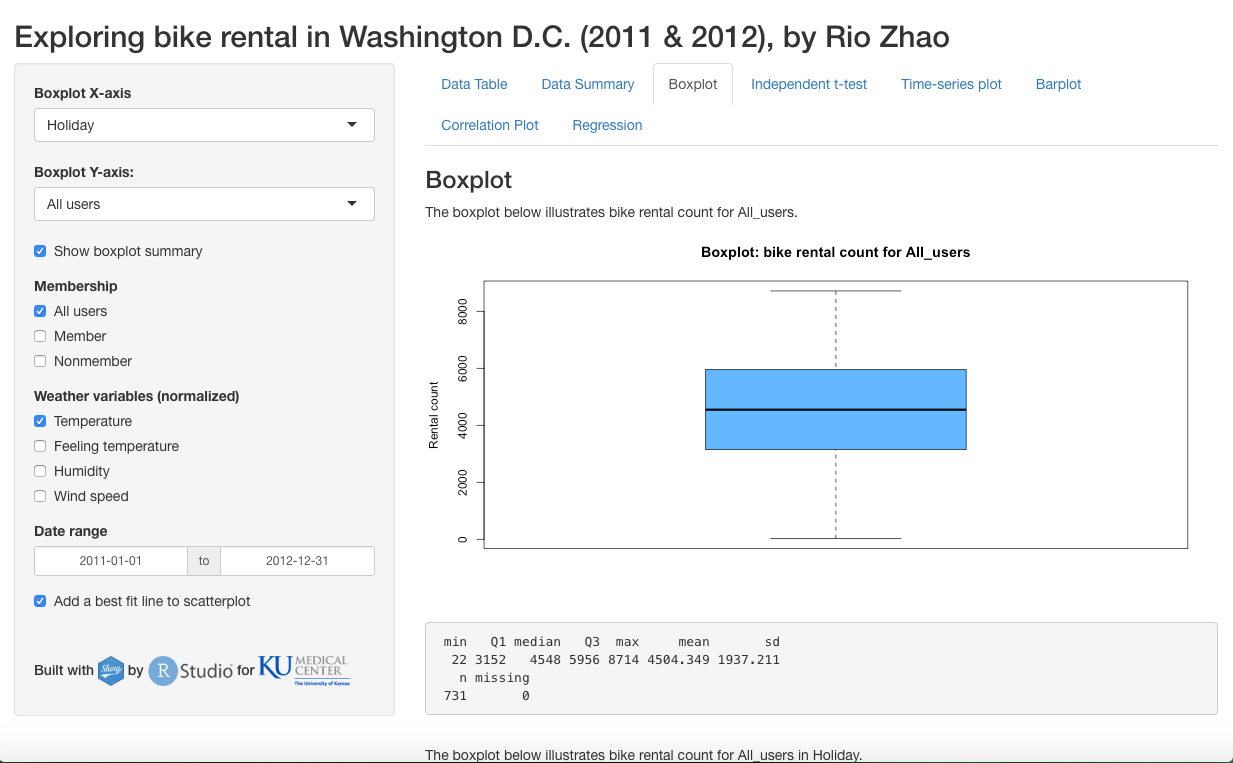
In this section, I imported the modified table into the app, excluding some variables that were not used in this project. With the data table, you can learn about the variables in the dataset, whether they are categorical or continuous. You may also choose to display certain number of entries in the dataset or search for key word in the dataset.



**Tab2 Data Summary**

I pulled summary statistics for the entire dataset being used in this project directly and print them out. With summary statistics of all the variables, you will learn about different levels in the categorical variables, and values of continuous variables such as their mean and median. The purpose of this section is to introduce you the dataset summary.



**Tab3 Boxplot **

In this section, there are three components being displayed. The first one is a boxplot of continuous variables including bike rental count for different groups of users. The second component is the summary for the first boxplot showing statistics such as minimum, 1st quartile, 3rd quartile and standard deviation etc. The last part of this section is another boxplot having both X- and Y-axis. While the X-axis being a categorical variable such as holiday or workday, the Y-axis is continuous variables that show bike rental count for different user groups.

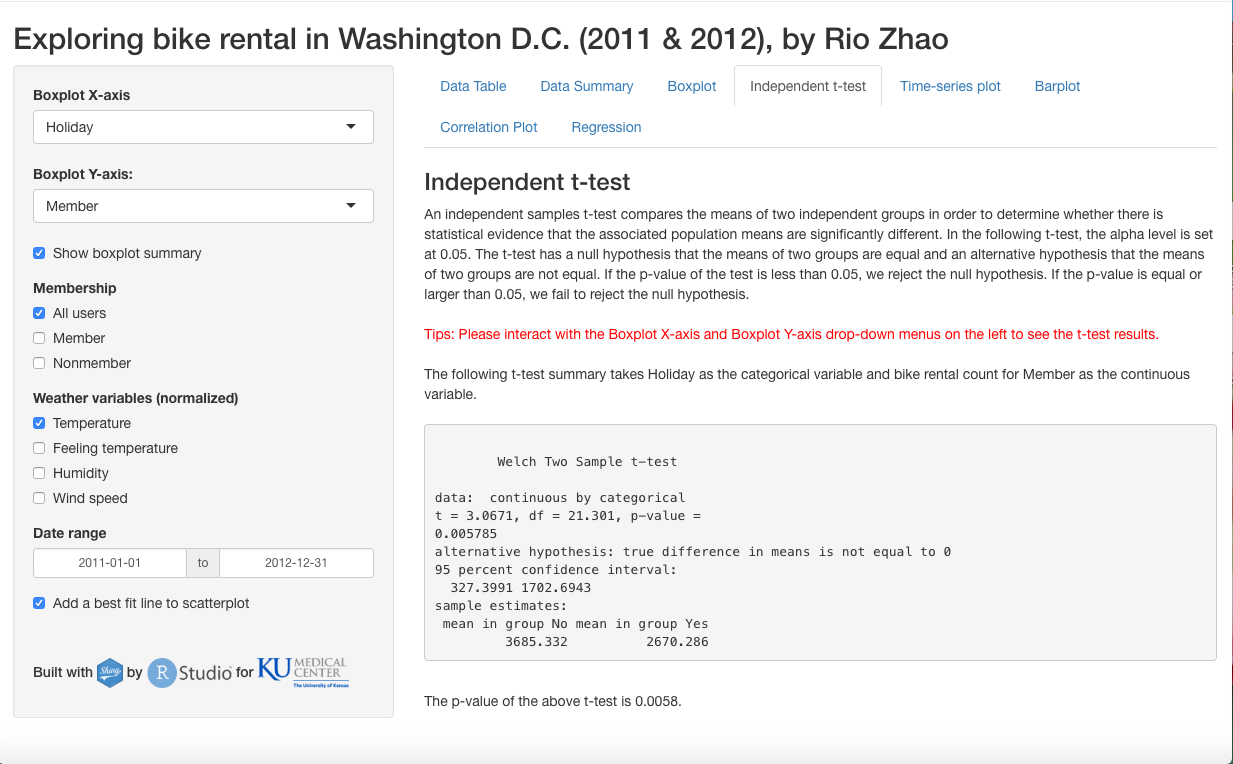
You may interact on the left-side panel to change the X- and Y-axis of the boxplots or choose to include or exclude summary statistics for the first boxplot.

With second boxplot, you will visualize whether the means between two groups are about the same or different. While visualization provides a direct illustration of this idea, its output is only a rough estimation. A statistical test was performed in later section to deliver precise comparison of means between two groups.



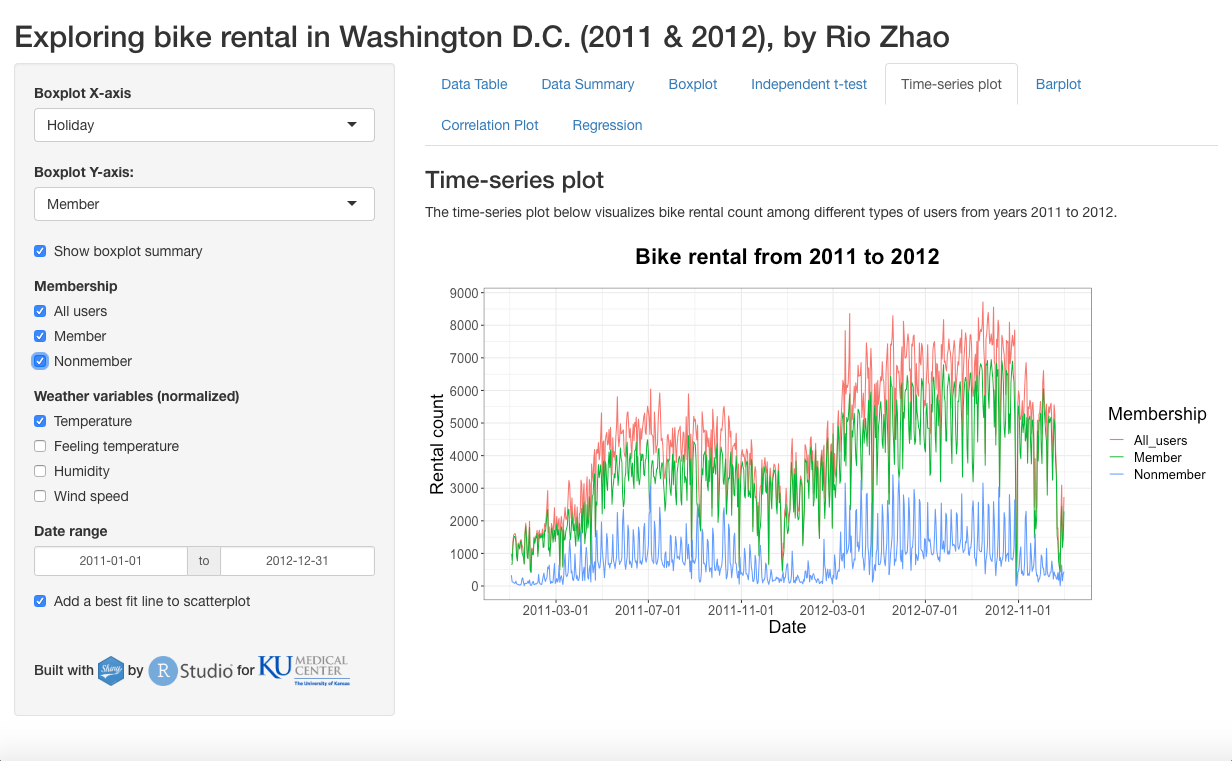
**Tab4 Independent t-test**

In this section, you will have the opportunity to explore an independent t-test that takes a continuous variable, which is bike rental count for various user groups, and a categorical variable, which is either holiday or workday. By choosing different input combinations on the left panel, summary for the corresponding t-test will be displayed and you can see its p-value to draw a conclusion whether to conclude or reject the null hypothesis.

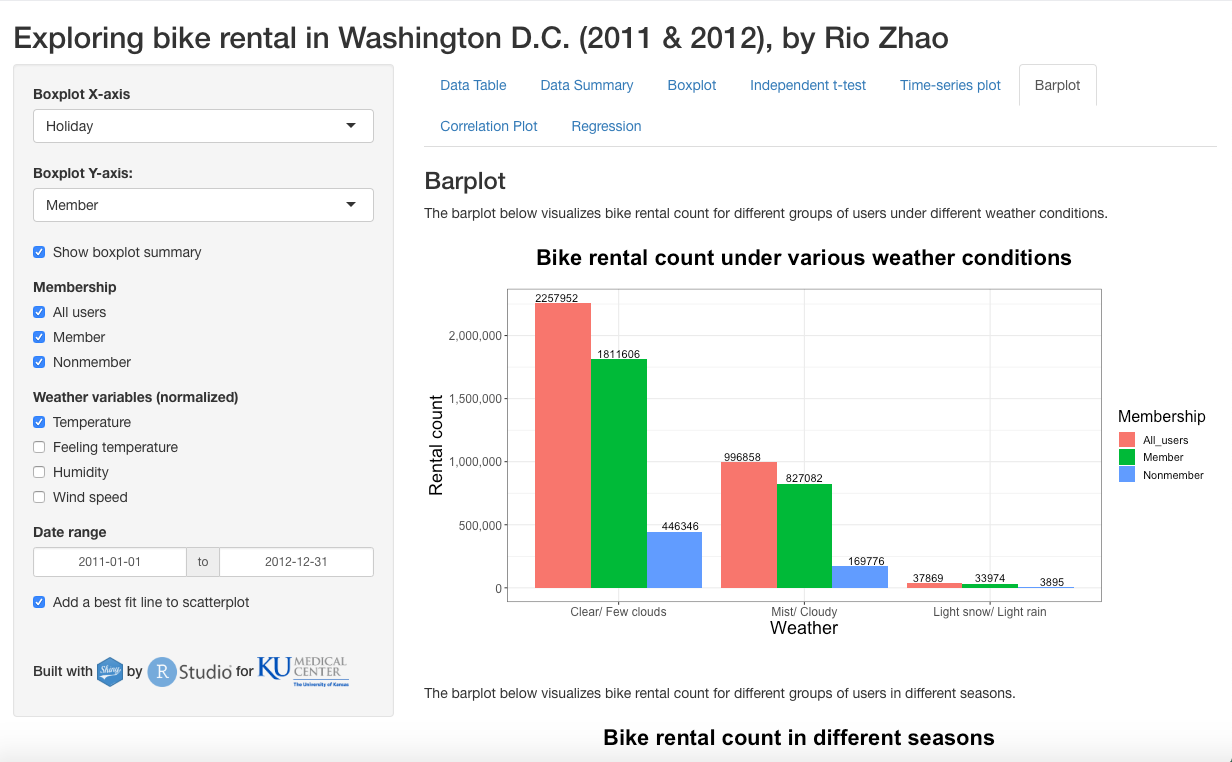


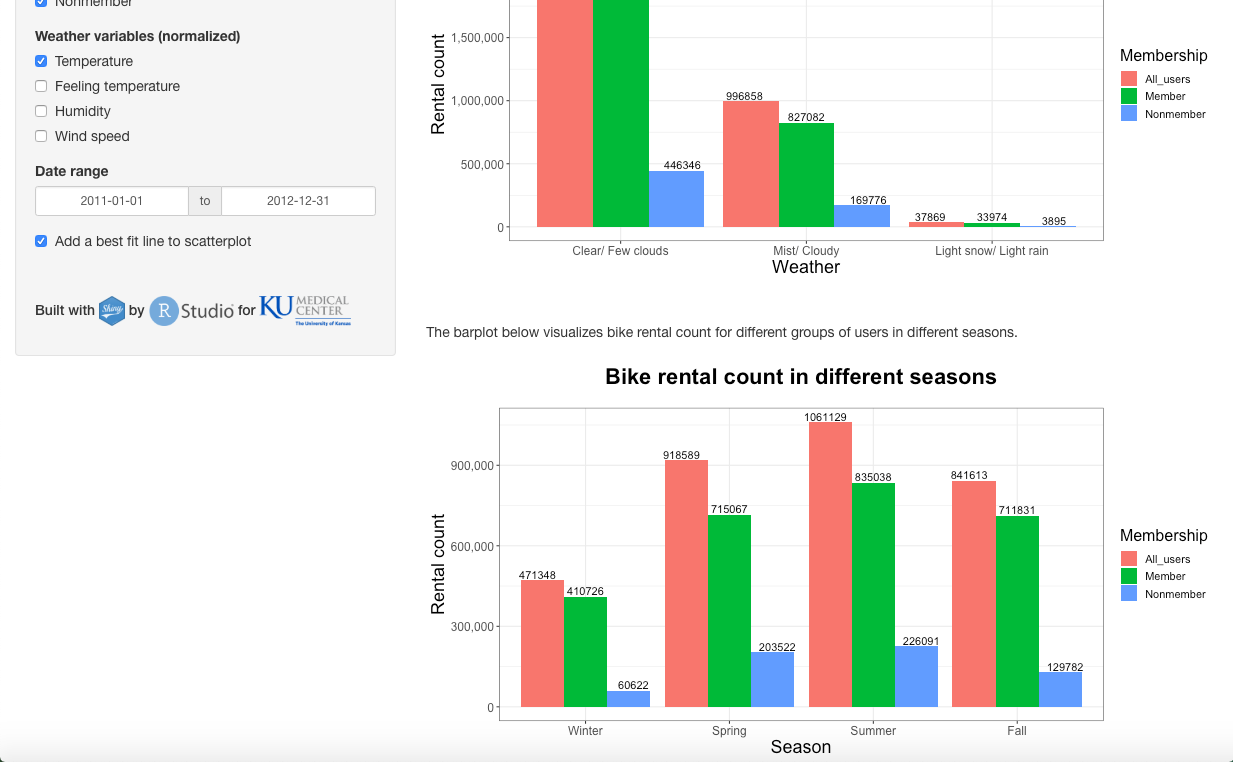
**Tab5 Time-series plot**

I have included a time-series plot that takes date as the X-axis and bike rental count for different user groups as the Y-axis. You may choose to display bike rental count for one to all user groups in the plot by interacting with the Membership check group on the left panel. You may also choose your own date range by entering desired start and end date on the left panel. Please note that the valid date range for this project is between 2011-01-01 and 2012-12-31. Specifying any dates outside this window will not result in a valid plot.



**Tab6 Barplot**

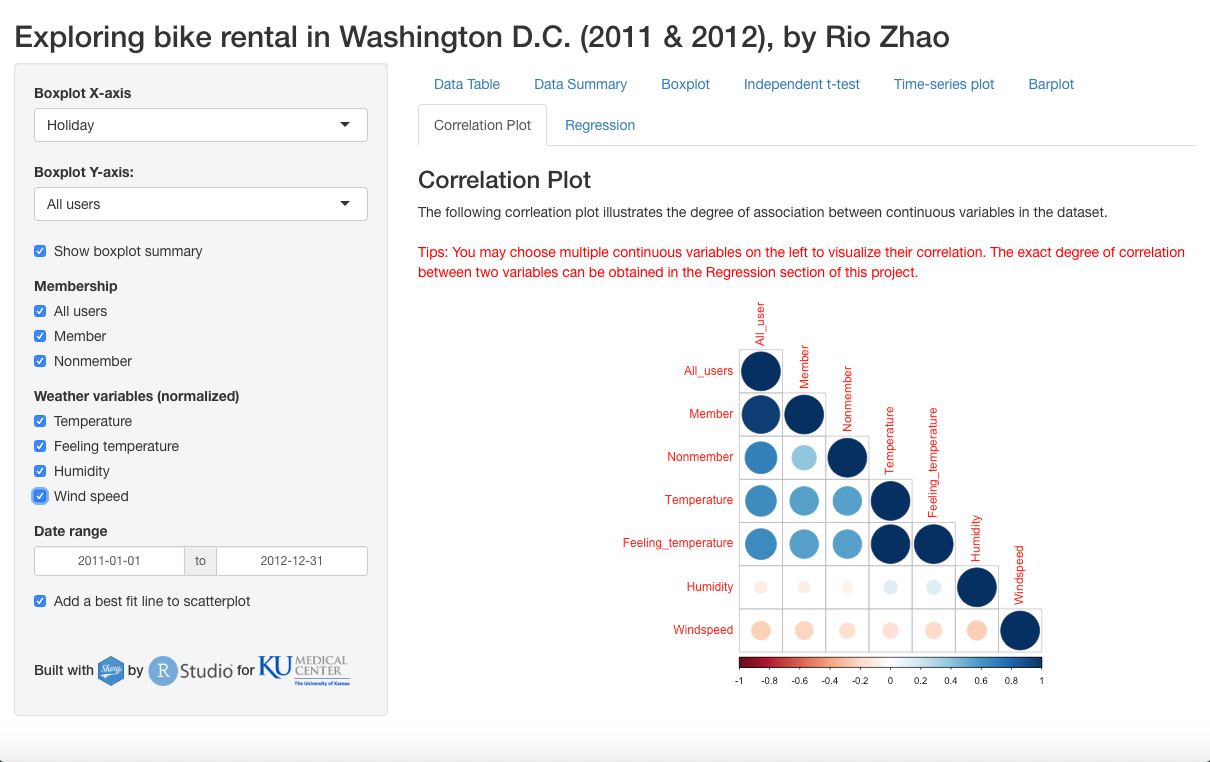




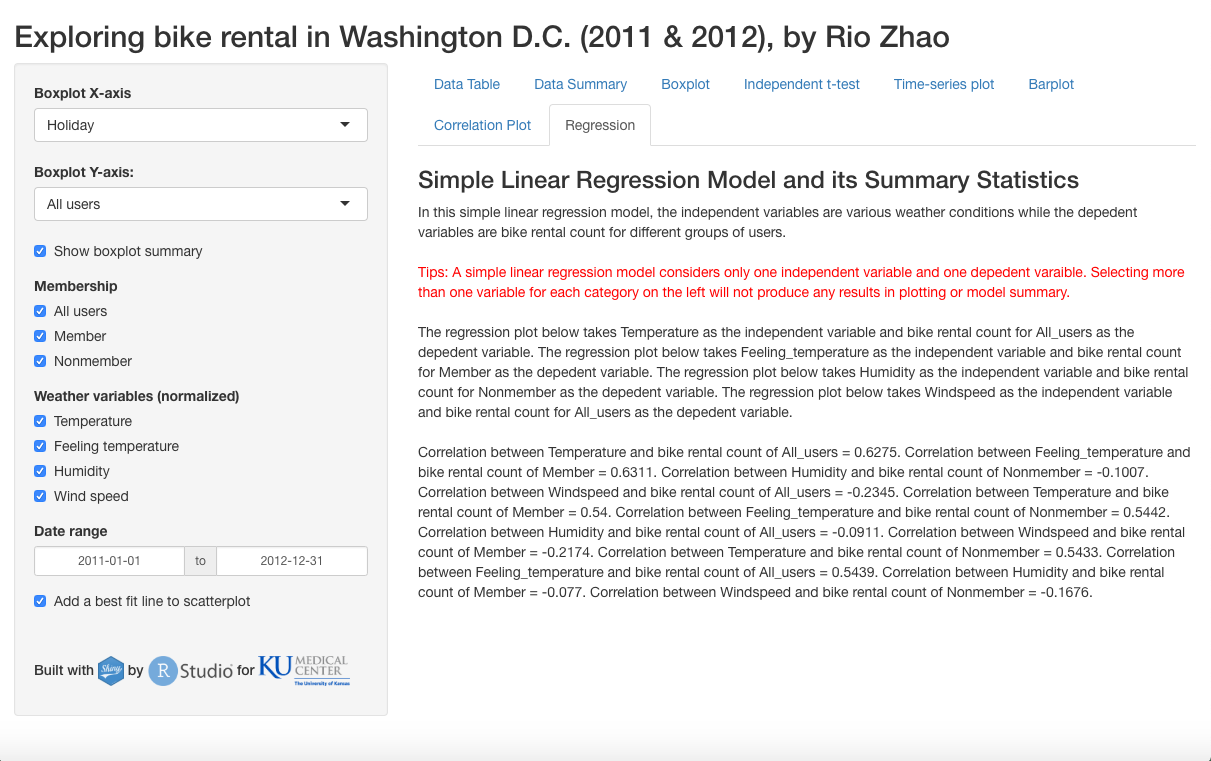
In this section, you will find two barplots. The first one illustrates the bike rental count for various user groups under various weather conditions in a day. The second one illustrates the same thing in different seasons. By interacting with the Membership check group on the left panel, you may be able to visualize comparison of bike rentals count for different user groups under different conditions. On each bar, there is also an exact count for bike rental to give you precise count.

**Tab7 Correlation plot**

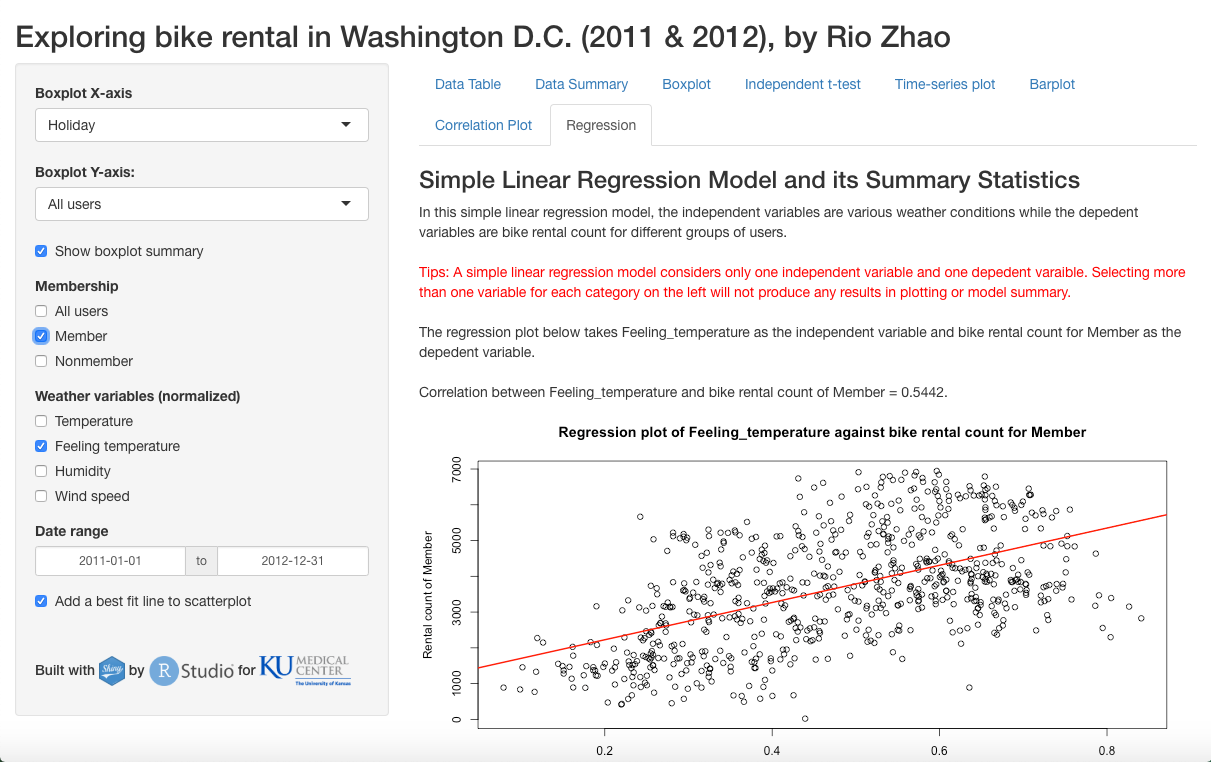
I have created a correlation plot here to show you visual correlation between all continuous variables in the dataset. You may be able to tell which variables are more closely correlate to each other by visual inspection. In other words, you may be able to develop a potential regression model using the variables that are closely correlate to each other. You can also interact with the plot by specifying the number of variables to be included in the plot on the left panel.



**Tab8 Regression**

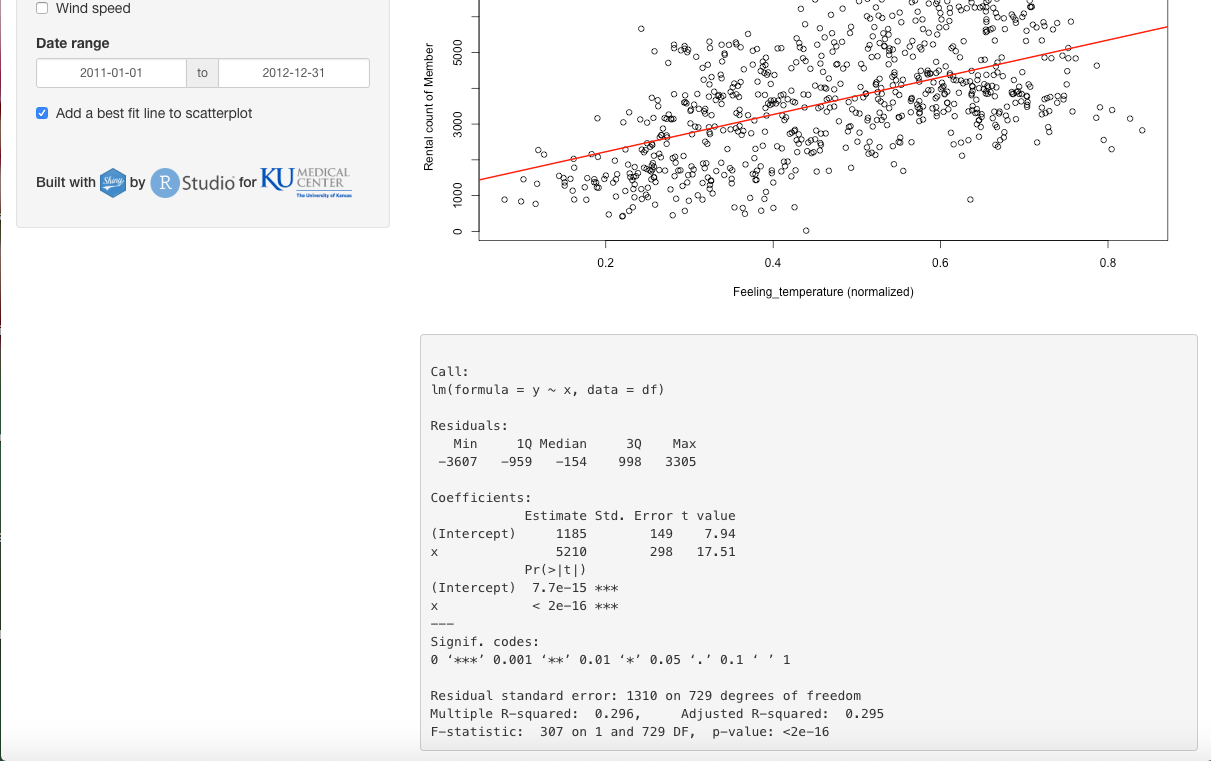


In this section, you will be able to explore some potential simple regression models that predict bike rental count using another continuous variable. Due to the nature of a simple regression model, you may not select more than one variable for the dependent and independent variable, respectively. No model plotting or model summary will be shown if you choose to do so. However, this app is still able to calculate the exact correlation between two continuous variables for any number of variables you check on the left panel.



Selecting only one variable for X and Y in the simple regression model, you will be able to visualize a plotting of all data points in a single plot. This gives you visual understanding of relationship between the dependent and independent variables. You may also choose to include or exclude a best-fit line of the model in the plot.

In addition to visual inspection of the regression model, a model summary will be automatically shown below the plot. It will also be automatically updated to reflect the current regression model when you choose to explore different X and Y variables in the model on the left panel.



**Code for this app**

I have included code for creating this app in two files, ui.R and server.R, along with the original dataset. With appropriate packages in R, you may reproduce this app for an interactive experience.