DS4200 Final Project

Pierre Dang, Melina Lymberopoulos, Shreya Rajeswaran, Rishpiath Satter

**Analysis of Washington D.C. Bike Share Data**

The link to our website is: <https://pidang1.github.io/Bike-Sharing-Project/>

The github repository can be found at: <https://github.com/pidang1/Bike-Sharing-Project>

Line chart - bike share use over time

This is a line graph that represents the overall trend of data over time. More specifically, it displays the number of bike share customers each day from 2011 to 2018 in Washington D.C. There is an interactive component to the plot where if you select a section of data on the line plot, a bar chart will populate to the right that has information about the number of entries in each season. It can be seen that there is a cyclical trend in the data where every year, there is a rise and fall of the data corresponding to the seasons (for example, there is usually a relative peak in the data around July and a relative minimum in the data around January). Additionally, it can be observed that the data has a positive trend overall indicating that as time has gone on, more and more customers have started using the bike share system, especially in the summer months. This suggests that the bike share system has increased in popularity over time and also indicates that more people are likely to use the bike share system during the warmer seasons. The mark for this visualization is a line (for the line plot) and area (for the bar plot). The channels are horizontal and vertical position. The task is to discover trends (high level). We chose this visualization to provide an overview of trends in the data over time, and also added a tooltip interaction to provide additional information of the max temperature and number of casual vs registered customers on any given date, if the viewers are curious.

Bar chart of mean customers based on weather type

This is a visualization that is a bar chart to show the effect of weather types on the total number of customers. Individual bars in this visualization represent the marks for each weather type, while the length of each bar corresponds to the mean total number of customers. The channels employed include the x and y position (length of the bars and what the bar represents). Tooltip is also implemented to inform the viewers of the exact number of mean customers. This visualization is supposed to compare the average number of customers in different weather conditions and help the viewer find the trends and differences in customer behavior depending on the weather conditions. The task is to identify weather types that result in very few or very many customers (low level).

Histogram of casual and registered customer distribution

We also noted that there was a distinction between casual and registered customers in the dataset, and wanted to see if there were any differences that could be seen with these variables. So, we created histograms of the distribution of casual and registered customers, where it can be seen that the casual distribution has a lower mean and is positively skewed, while the registered distribution is fairly symmetrical with a higher mean. This suggests that it's more likely for there to be a high number of registered customers on any given day than it is for there to be casual customers, which might indicate customers who are registered are more consistently using the bike share platform. The mark for this visualization is area, and the channels are horizontal and vertical position. The task is to discover trends (high level).

Scatter plot of customers vs precipitation

This visualization shows the relationship between bike-sharing rentals and amounts of precipitation, showing the casual customers, registered customers and total customers (casual + registered customers) through an interactive dropdown. Tooltip is also used to describe the max temperature and wind speed at each point. This chart reveals how many types of customers respond to weather conditions, with casual customers generally showing a stronger decline in rentals as precipitation increases, while registered customers may be less affected due to regular use for commuting. The total customer plot combines both groups, offering insights into overall bike rental trends in response to precipitation. The marks in this visualization are scatter points, where each point represents a specific observation in the dataset, showing the relationship between precipitation levels (on the x-axis) and the number of customers (on the y-axis). The channels are x and y position; the position of each point along the x-axis reflects the level of precipitation, while the position along the y-axis represents the number of bike-sharing customers. This allows us to visually compare how casual and registered customers behave under different weather conditions, and see how precipitation impacts bike-sharing usage overall. The task is to compare different types of customers (low level), and to discover trends in precipitation (high level).

Scatter plot of temperature vs precipitation

This scatter plot visualizes the relationship between the maximum temperature of the day and the total number of bike-share customers. The x-axis represents the maximum temperature of the day, while the y-axis represents the total number of customers. We also added an interactive slider that filters data points by their max temperature value, allowing users to focus on specific ranges of temperature. From the plot, we saw that there was a positive correlation between temperature and customer count on warmer days. This shows that the bike sharing system is more popular during warmer weather, likely due to more favorable conditions for outdoor activities. We also saw that there was some spread in the data, indicating variability in customer numbers even on days with similar temperatures, which could be influenced by factors such as weather patterns, holidays, or events. The mark for this visualization is a point (representing each data point/circle), and the channels are horizontal and vertical position. The task is to identify patterns and correlations between the variables (high level).