**Description of the Visualizations**

**Group 32 – Boston Bluebike Analysis**

Link to public website: <https://rxzhang7.github.io/DS-4200/>

1. **Geographic Distribution: Ride Count by Neighborhood (Choropleth Map)**

This visualization uses Folium to present a contour map of Boston, showing the total number of Bluebike rides aggregated by neighborhood. It provides an interactive overview that not only conveys the spatial patterns of bike use but also guides the user through exploring changes in various areas of the city. Users can quickly view ride counts by hovering over any neighborhood, which helps identify hotspots of Bluebike activity.

We utilized GeoPandas, Pandas, Folium, and Shapely to create the Bluebikes choropleth map. Pandas handled data cleaning and aggregation, while GeoPandas and Shapely enabled spatial processing to assign trip points to Boston neighborhoods. Folium was used to render the interactive map, featuring hover tooltips that display ride counts per neighborhood.

A choropleth map was chosen because it effectively shows geographic variations in bike usage. Its color gradient makes it easy to compare neighborhoods at a glance. The core dataset came from Bluebikes trip data, with each ride’s start and end coordinates linked to neighborhood boundaries. Interactivity was achieved through tooltips that appear on hover, providing more detail without cluttering the map. The color scale (yellow to red) visually encodes ride intensity, while styling elements like opacity and simplified borders maintain readability and visual appeal.

The goal of this map is to show which areas in Boston have more or fewer Bluebikes rides. Darker colors mean more rides, while lighter colors mean fewer rides. This visual design makes it clear which neighborhoods are busy and helps guide viewers through the data. The map serves as an easy-to-read introduction that sets up the rest of the project.

1. **Correlation plot (Heatmap)**

The purpose of the heatmap is to show how the different categories of data in our dataset are related to each other, whether that be positively or negatively correlated. This representation was aimed at the public in Boston and potential policymakers that could interpret the data and make changes to the Bluebike system to better account for how consumers utilize these services. The correlation plot is helpful to determine common trends between data, such as determining when each type of user (member vs. casual) uses the services. Our group wanted to make this aspect more visually easy to understand, so we created a color gradient from red to blue, where red means the categories are more closely correlated, and blue represents negative correlation. These colors are bold and color-blindness friendly, and the visualization is not dependent on the colors to be understood. Finally, the data was able to clearly confirm our conclusion that the weather was affecting ridership due to the negative correlations between humidity and rider count and duration.

1. **Hourly Bluebike Usage Trends by Weather Conditions (Line Plot)**

For this visualization, we decided to use a line plot to clearly demonstrate how the ridership numbers changed over time. This design is intuitive to understand that left-to-right shows the total hours in a day, and the line depicts how many people are riding at any given time. We wanted to also incorporate multiple variables into this design, so we created different lines for each of the different weather types. In this way, we were able to draw conclusions not only about the most popular times of the day to ride, but also in which weather conditions were favorable to users. We incorporated color through differentiating which line represented which weather condition. So, for example cloudy is grey, and rain in dark blue.

**4. Extended Analysis of Weather Impact on Ridership (Scatter Plot)**

This scatter plot was built using Altair, a Python-based declarative visualization library that supports interactive features with minimal code complexity. To ensure fair representation across weather categories and to avoid visual clutter, the dataset was filtered to exclude null values and then sampled to 700 rows per weather condition using a grouped sampling approach. This sampling helps create a balanced, interpretable visualization without over-representing any single weather type.

Each weather condition (Clear, Cloudy, Fair, Rain, Snow) is assigned a distinct color using a manual color scale:

* Clear: orange
* Cloudy: gray
* Fair: green
* Rain: blue
* Snow: light blue

An interactive radio button filter lets users toggle between weather categories. The selected category is displayed in full color, while others are faded out in light gray with reduced opacity. This technique draws attention to the chosen subset while still providing context from the rest of the data. The chart includes tooltips for additional details on hover, such as weather type, temperature, ride count, and day of the week. The layout is clean, with clear axis labels and a title that clarifies the sampling approach.

#### **Visual Insights & Interpretation**

The goal of this chart is to explore how riding activity fluctuates with temperature under different weather conditions. A clear trend emerges:

* Between 5°C and 25°C, ridership significantly increases, especially under Fair and Clear weather, indicating this is the optimal temperature range for biking.
* Snowy conditions show a sharp drop in ride counts, with most rides occurring between -10°C and 5°C, and very few above 10°C highlighting a clear deterrent effect of cold and snowy weather.
* Surprisingly, Cloudy weather shows a similar upward trend to Clear weather and even outperforms it slightly at certain temperatures. This suggests that overcast, dry days are still comfortable and may be preferred by some riders.
* Rainy days result in a flatter distribution of ride counts, further reinforcing how adverse weather reduces ridership.

The scatter plot uses color to encode weather, while the density of points reflects ride frequency. With the added interactivity, users can isolate trends by weather type and compare behaviors side by side. This chart clearly demonstrates that ridership is heavily influenced by both temperature and weather, and the best conditions for riding Bluebikes are mild temperatures with dry, fair or clear skies. It provides a compelling visual narrative that complements and extends earlier time-based trend analyses.

**5. Member vs Casual Riders by Day of Week (Bar Plot)**

This chart was created using Altair. The data focuses specifically on daily ride statistics split by user type, Casual and Member.

To prepare the data, the day-of-week values were cleaned and standardized by capitalizing and removing extra spaces. The dataset was reshaped from a wide format into a long format by combining casual and member data into one unified structure and adding a user\_type column for clarity.

A dropdown menu was implemented to allow users to filter the data by specific days of the week. This interactive feature makes it easier to observe how riding behavior varies between weekdays and weekends.

Two bar charts were created and placed side by side. One chart displays the average rider count, while the other shows the average ride duration in seconds. These charts are dynamically filtered based on the selected day. Tooltips offer extra context on hover, and consistent formatting ensures visual clarity across both charts.

#### **Visual Insights & Interpretation**

This visualization is designed to show how riding behavior differs between Member and Casual users throughout the week.

Several key patterns emerge from the chart. During weekdays, Member riders have significantly higher ride counts compared to Casual riders, which suggests that members are primarily using the bikes for commuting or regular travel. On weekends, Casual users tend to have much longer ride durations, indicating their usage is more recreational or leisure focused.

Each metric, rider count and ride duration, is presented in a separate chart for better focus. The use of a day-of-week dropdown filter helps viewers explore variations across the week in a flexible and intuitive way.

Although this chart does not rely on color or size for visual encoding, the side-by-side layout and interactive controls effectively highlight the contrast between user types across key behavioral metrics.

Combined with the first chart that explores the impact of weather and temperature, this visualization helps complete the narrative by focusing on how different user groups behave over time. Together, they provide a comprehensive view of when and under what conditions people are most likely to use Bluebikes.