# Report Data Visualization Assignment

Author: Marzagora Bice

Course: High Performance Data Visualization

## Dataset

The dataset contains bike rental data in Seoul from December 2017 to December 2018. It includes weather-related variables, such as temperature, wind speed, humidity, visibility, snowfall and rainfall, as well as time-based factors, including hours and dates. The aim is to create two synchronized visualizations to explore and analyze this multivariate dataset effectively. My chosen visualizations focus on identifying patterns in bike rentals and uncovering insights to help improve the bike-sharing service.

## Visualizations Used

### Scatterplot and Heatmap

Two main visualizations were used: a scatterplot and a heatmap.

The scatterplot is designed for examining relationships between weather-related variables, allowing users to select attributes such as temperature, humidity, visibility, snowfall, rainfall, and wind speed.

The heatmap, on the other hand, is focused on temporal attributes. It enables users to select both the x-axis and y-axis from options like hour of the day, days of the week, months, seasons, and whether the day is a holiday. Each cell of the matrix of the heatmap represents the total rentals for a specific combination of temporal factors, and its color intensity visually represents rental volumes.

This visualization was chosen because it reveals patterns that might remain hidden in raw data. In fact, certain temporal attributes, such as "days of the week" or "months", are not explicitly available in the dataset. In fact, simply plotting "Date" on the scatterplot does not provide the granularity needed for these attributes. By using a heatmap, I can reveal these patterns indirectly through color intensity, highlighting trends in bike rentals that may be related to specific days or months, even if this information isn't directly present in the dataset.

### 2D Brush Interaction on the Heatmap

When users select a range of data on the heatmap with the 2D brush, the scatterplot updates to display not just the data currently visible in the heatmap but all data corresponding to the attributes selected along the heatmap's axes. I chose this approach with the brush because it allows users to explore relationships between the selected temporal or categorical ranges (like specific months or days of the week) and other variables. in this way, users can analyze relationships across the entire dataset, not just the aggregated values displayed in the heatmap at that moment.

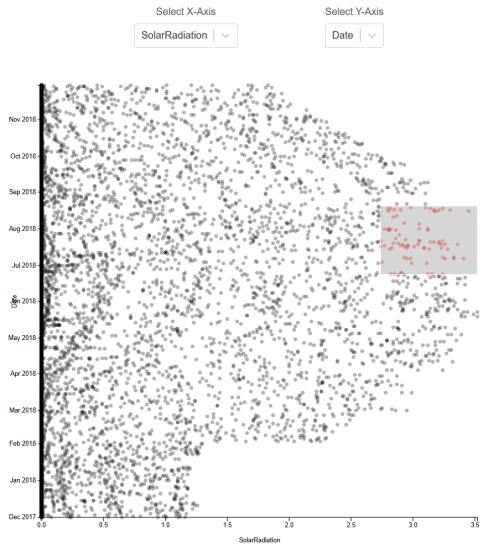
## Key Observations and Insights

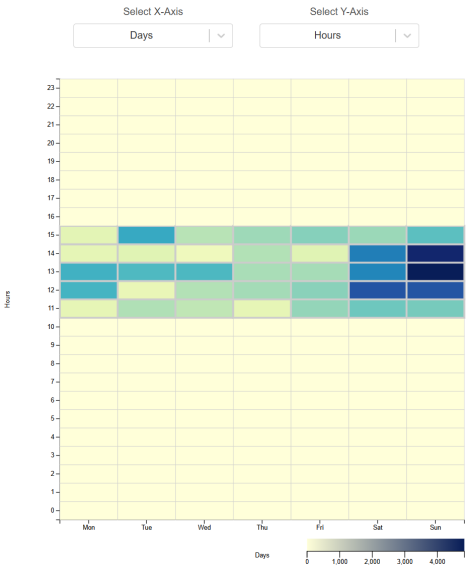
The visual encoding in the scatterplot and heatmap was intentionally designed to align with the properties of the data and the tasks users aim to perform. The scatterplot works well for showing how weather-related data, like temperature and rainfall, affect bike rentals. The heatmap is better for looking at patterns in time-based data, like hours, days, and seasons.

Example of insight gained with this visualization:

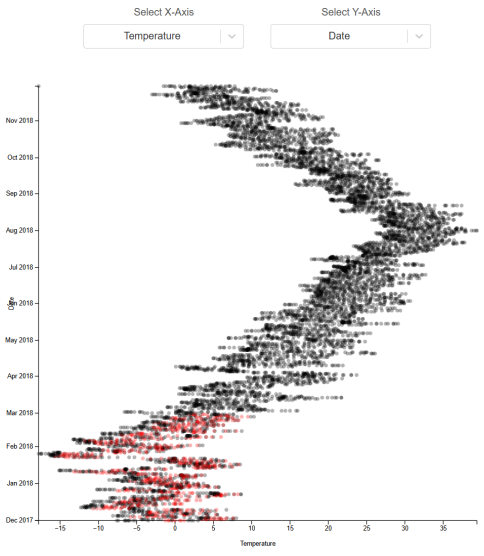
- Weekday rentals peak during commuting hours, with fewer on holidays.
- Rentals increase in spring and drop in winter.

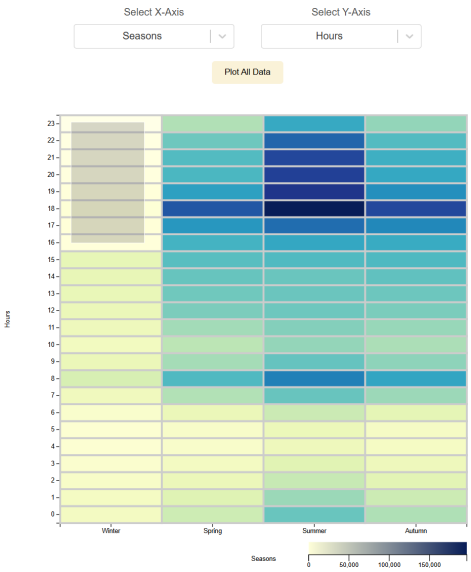
- Rental patterns around holidays and weekends differ significantly from regular weekdays
- Sunny weekends during spring and summer see peak rentals in the mid-afternoon, highlighting their popularity for recreational cycling.




- Colder temperatures in winter correlate with fewer rentals, especially in the evenings.




- Rainy and windy weekday mornings and evenings still have moderate rentals, driven by commuters demand.


## Pro and Cons

**Pro**:

1. The heatmap reveals patterns not explicitly present in the dataset, such as the day of the week or the month.
2. The scatterplot reveals detailed relationships between data points, while the heatmap shows broader trends through color intensity. This combination allows users to analyze both individual details and large-scale patterns effectively.
3. The scatterplot focuses on weather-related factors, while the heatmap captures temporal or categorical patterns, offering a well-rounded view.

**Cons**:

1. The heatmap's aggregation can obscure finer details, especially with high variability.
2. Both plots may struggle with large datasets: dense areas may cause overlapping points, making it hard to interpret specific values, and heatmaps need finer granularity.
3. The heatmap's color scale is relative, meaning the same color can represent different values depending on the 2D brush area in the scatterplot, and can be misleading even if there is a legend.
4. The heatmap's reliance on color may not be effective for users with color vision deficiencies or in low-contrast settings.

# Conclusion

In conclusion, the combination of scatterplots for weather-related data and heatmaps for temporal patterns, enhanced by interactive tools like the 2D brush, provides a useful and intuitive way to explore bike rental data. These insights can help optimize bike-sharing services, such as: increasing bike availability during commuting hours, promoting usage during low-demand periods, and planning for seasonal adjustments.