Interactive Hybrid Visualization of Seoul Bike Sharing Data

Nicola De March

November 2024

1 Introduction

This project explores the Seoul Bike Sharing Demand dataset through an interactive hybrid visualization designed using the D3.js library. The dataset provides insights into factors affecting bike-sharing demand in Seoul, including weather conditions, seasonal trends, and socio-economic variables.

The visualization integrates two complementary techniques: a scatterplot and parallel coordinates. The scatterplot allows users to observe correlations and distributions between two selected variables, while the parallel coordinates enable multidimensional analysis by visualizing relationships across multiple attributes simultaneously.

The design facilitates seamless navigation through quantitative data while maintaining a clear perspective on categorical variables. This approach proves particularly effective, as the dataset contains a large number of quantitative attributes compared to categorical ones. By allowing users to dynamically select, filter, and highlight data points across both visualizations, the dashboard enhances data exploration and supports the discovery of patterns within the dataset.

2 User Experience and Instructions

The dashboard consists of two main parts: the left panel and the right panel, see Figure 1.

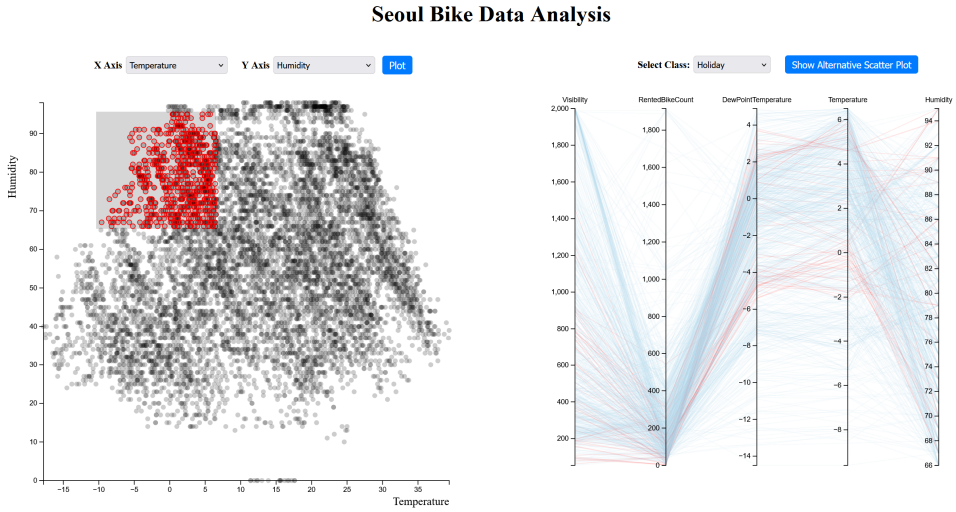

Figure 1: General view of the dashboard.

The left panel serves as a tool for maintaining a general overview of the dataset. It features a high-level scatterplot that allows users to dynamically modify the axis variables through a dropdown menu (see Figure 3). This flexibility enables users to explore different relationships within the data and identify key patterns or trends at a glance. To see the changes, press the button **Plot**. You can click on a datapoint to view all the details of it, as shown in Figure 2.

The right panel, on the other hand, allows users to choose between two different visualizations: a zoomed scatterplot or a parallel coordinates visualization. Both visualizations require that you select data on the main scatterplot using a **2D brush**; otherwise, the right panel will remain empty. The

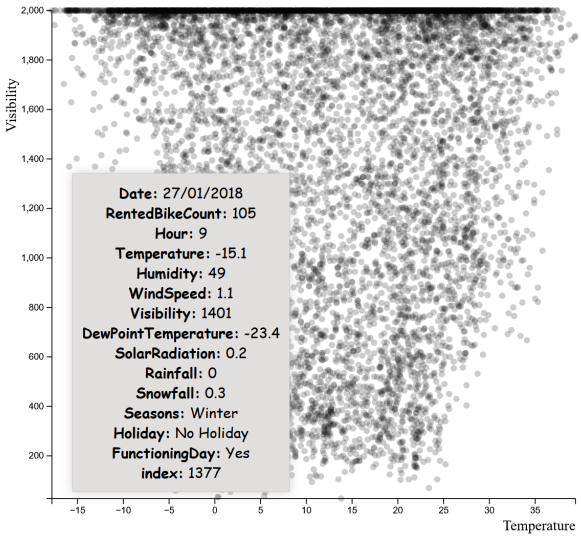

Figure 2: Datapoint details.

2D brush highlights the data selected by the user and simultaneously updates the right panel plot in real-time. The brush can be moved freely across the scatterplot, enabling dynamic exploration of the dataset.

You can always choose the class that you want to visualize, thanks to the dropdown menu on the right panel, as shown in Figure 3.

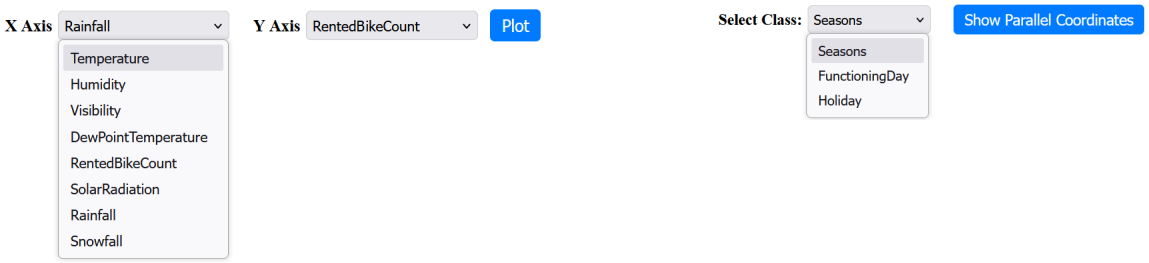

Figure 3: Overview of the dropdown menus and buttons: on the left the menu for the left panel and on the right the menu for the right panel.

2.1 Alternative Scatterplot

On the right panel, you can select a scatterplot that contains all the points chosen with the brush. This feature enables a more detailed analysis, as it allows you to select and visualize a specific class or subset of data points, providing a focused view of the dataset. An example is provided in figure 4.

2.2 Parallel Coordinates

To gain a better understanding of the dataset, a parallel coordinates plot can be used. This type of plot displays multiple variables along parallel axes, making it easier to visualize relationships and trends across multiple dimensions, especially for complex datasets that are difficult to analyze using scatterplots. Additionally, users can interact with the plot by highlighting specific lines. This can be done by positioning the mouse pointer over a line, which allows users to focus on particular data points and observe their relationships across all the axes, as shown in Figure 5.

Moreover, the plot supports the use of 1D brushes on each axis. This feature allows users to select a range of values on any given axis, dynamically filtering the data displayed. The filtered data is updated in real-time, offering a detailed exploration of the dataset's structure and trends. See Figure 5 for an example of how the brush functionality works.

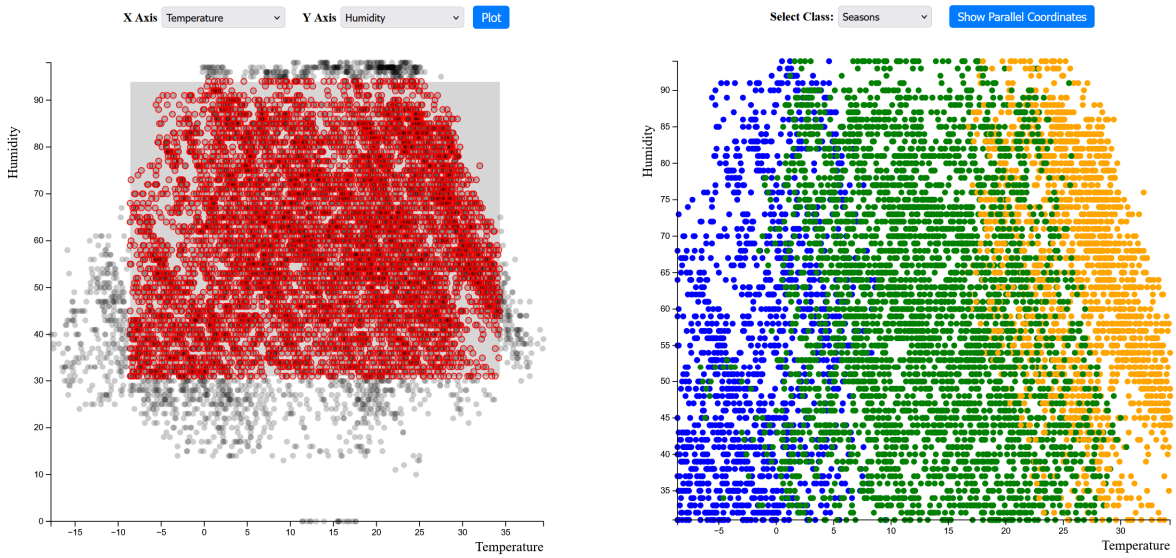

Figure 4: Alternative scatterplot visualization.

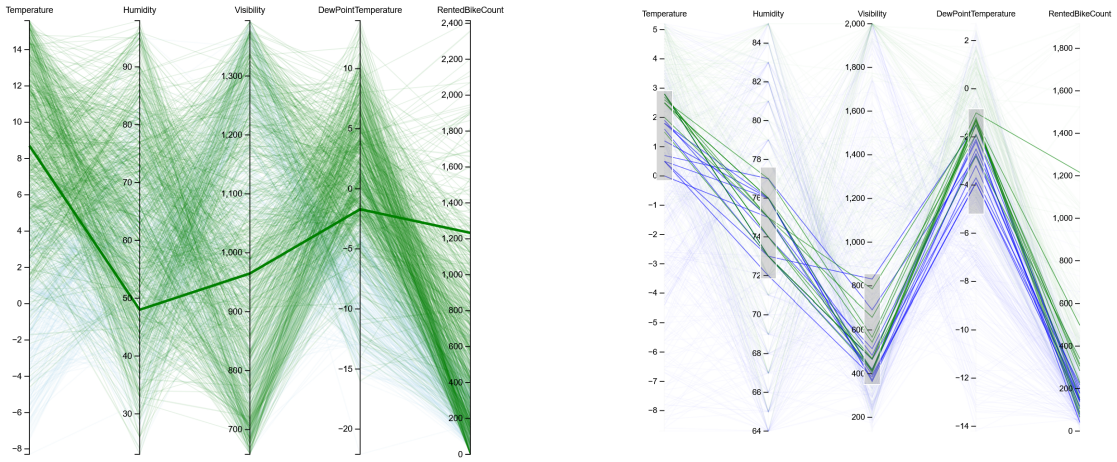

Figure 5: On the left parallel coordinates with a datapoint highlighted, on the right data filtered through 1D brushes.

Moreover, it is possible to change the order of the axes by clicking on a label. The clicked axis will be moved to the first position, starting from the right.

3 Pros and Cons

This project allows you to navigate through the data with ease, offering great flexibility. It enables you to zoom into the data and filter it while maintaining an overall view, thanks to the left panel. Additionally, the system is very responsive and fast, providing a smooth user experience.

On the downside, the left panel does not update based on the selections made in the right panel. It would be a useful feature if filtering the data on the right could automatically highlight the corresponding entries on the left panel.

Another potential drawback is that not all variables are available in the parallel coordinates plot. Only the most important variables, which present more complex patterns, are included. However, this approach allows for more space to plot other variables, making the graph cleaner and easier to read.

One of the strengths of the parallel coordinates visualization is the ability to reorder the axes dynamically. This feature allows users to organize the visualization in a way that makes patterns and relationships clearer, providing a more structured and ordered view of the dataset.