

Data Visualization Synced Graphs

PEREIRA CARDOSO Luc

26/11/2024

1 Introduction

This report outlines the relationship between the two graphs and the justification for the specific choices. The scatter plot was already decided for the assignment, and I decided to add a parallel coordinate plot to obtain a more detailed view of the Data and better see relations between different variables. In addition, it is possible to add any numerical value to both graphs to allow even more visualization of the data.

2 Data Set

The dataset consists of **8,760 data points**, representing hourly observations across various attributes. The dataset includes the following headings:

- **Date:** The specific date of the observation.
- **RentedBikeCount:** The number of bikes rented during the corresponding hour.
- **Hour:** The hour of the day (0–23).
- **Temperature:** The temperature.
- **Humidity:** The relative humidity.
- **WindSpeed:** The wind speed.
- **Visibility:** The visibility.
- **DewPointTemperature:** The dew point temperature.
- **SolarRadiation:** The solar radiation.
- **Rainfall:** The amount of rainfall.
- **Snowfall:** The amount of snowfall.
- **Seasons:** The season during which the observation was recorded (e.g., Spring, Summer, Autumn, Winter).
- **Holiday:** Whether the day is a holiday (Yes/No).
- **FunctioningDay:** Whether the day is a functioning day (Yes/No).

Some of these fields are not available mainly because they require a lot of modification to the code to show any type of data. That's why I decided only to use numerical values; they are the majority and can already demonstrate many different relations.

3 Task Bar

To simplify the program's use, A taskbar was added with a button to reset to the defaults.

4 Scatter Plot

The Scatter Plot can dynamically show any numerical combination of x and y. Depending on the trend we want to see, having the option to change the axes freely can be beneficial. In addition, a colour scale is used from blue for low values to red for high values. This is especially beneficial when using temperatures, but it can also be used for other variables since it is pretty simple to understand.

The decision was made not to encode any more data visually because, given the large number of data points, it would make it less readable and harder to change the axis dynamically. In addition, a Parallel Coordinates plot allows us to show the relation between many different data axes, which does a better job than adding more visual encoding to the Scatter Plot.

5 Parallel Coordinates Plot

A Parallel Coordinate Plot allows us to display relations between many different variables. It is a good fit for this case given that the first graph is a Scatter plot, which can be a bit difficult to encode many variables into. Having a Parallel Coordinate Plot can make it simpler to use and understand the relation between variables. In addition, having the option to choose which axes we want makes it even better to detect the trends we want. This plot also uses the same colors as the Scatter plot and helps identify trends.

6 Interaction

6.1 Scatter Plot

For the scatter plot a 2D brushing is used when selecting some of the points the other graph will re-render while only showing the selected subset and updating the axis.

6.2 Parallel Coordinates Plot

For the Parallel Coordinates Plot it uses a 1d brush for each axis the selected point are then showed in the Scatter plot only the subset is shown but it retains the original scale this helps to visualize the position in the data. In addition the colors update to represent the lowest and highest value in that subset and not in the global subset.

7 Improvements

A few improvements could be made. Adding a hover box to show additional information on the points could be exciting. Also, adding multiple color scales and other visual options can be beneficial. The code seems stable and doesn't seem to have too many bugs, but the structure is imperfect. It works and is stable and performance is also pretty good,

which can be problematic when having larger data sets. In this case, the set is not big, but it can already lead to some performance issues.

8 Conclusion

In conclusion I am very happy with the result and personally find it looks good and offers greater ways to show trends between data points. Given that we got templates for the cases and components, I decided to use them even though I personally don't think that classes in React offer any benefit. At least in normal React, I do not know if there are performance benefits to using it for d3.