Dashboard Prototype Presentation

Interactive Data Visualisation

SET09123

Andrew Taison
Matriculation: 40538519
Word | Page count: 1598 | 8.5
(excl. references and appendices)
Submission date: 27th October 2023

Table of Contents

List of Figures

List of Tables

Appendices

Introduction and Description

This report documents the first part of the project in which the role of a Data Scientist is undertaken to analyse a weather dataset for marketing purposes. The aim is to create an interactive dashboard in which to explore the provided dataset and discover patterns.

V. Martínez-de-Albéniz and A. Belkaid (2021) [1], identify the weather as an important factor which can determine product demand especially in goods for which usage is affected by weather conditions. They find that rain largely effects foot traffic increasing it in shopping centres and decreasing it in high street stores. They suggest rain as a first-order factor for channel choice. They also say temperature effects foot traffic less, however it has a large effect on seasonal related sales; more summer items sold under positive temperature shocks and more winter items sold under negative temperature shocks.

The dataset provided by the Met Office contains 3420 records across 15 years from 2007 to 2021 (inclusive). Table 1. (below), shows the dataset attributes along with their levels.

Table 1. Dataset Attributes and Levels

Data attribute:	name (station)	lat	lon	region	year	month	max_temp	min_temp	af_days (air frost)	rain	sun
Attribute Level:	NC	NI	NI	NC	NO	NO	NI	NI	NR	NR	NR

NC = Nominal Categorical
NO = Nominal Ordinal
NI = Numerical Interval
NR = Numerical Ratio

Taking into account V. Martínez-de-Albéniz and A. Belkaid (2021) [1] study, the focus will be on rainfall and temperature. However other visualisations will be created to see if other patterns in the data can be identified.

Development

To begin with, classes were developed for various visualisations. The classes built were Chart, BarChart, BubbleChart, Pienut (pie / donut), LineChart (single and multi lines) and Scatter.

Chart deals with setting up the chart area, appending an svg element inside the selection container, as well as defining common functions among chart types. All other chart types extend Chart. Uncommon functions specific to each visualisation are dealt with in their own class.

The classes have been designed to be easy and quick to use based on an object orientated design. This has however come with some caveats, as the code did contain some bugs / known issues, most of which have been rectified. However due to attempting to set the code up based on OO design, shared highlighting is proving to be difficult to implement. The test code has been provided alongside the main code where shared highlighting was attempted. It is suspected that due to each class using different keys for the x-axis along with the line charts being set up differently to the rest lies at the heart of this issue. Unfortunately, time has not allowed for reworking of the code setup.

Variations of data groups were explored, and it was found that there isn't enough data to get a good insight into trends across the years (with the exception of an extreme low average daily temperature of -6.8°c in 2010 (chilly!), and a peak high average daily temperature of 28.3°c in 2018). Appendix A shows a trial visualisation to show the temperature trends across the years. Due to this, the focus of the visualisations was decided to be month averages.

Some features at the time of writing are not yet included due to time restraints. These include: Visualisation titles (however tool tips provide most of this information), shared highlighting and responsive layout. A horizontal line for the multi line chart showing 0°c would be a nice addition.

Visualisations

Each visualisation for the dashboard prototype is shown and discussed below. For these, a new attribute ('avg_temp') was created when the csv file was read in. This was calculated by adding 'max_temp' and 'min_temp' and dividing by 2 for each record.
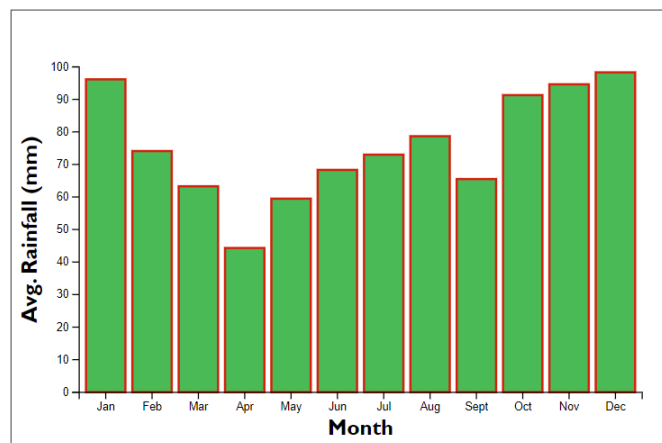
Figure 1. Avg. Monthly Precipitation Bar Chart

Figure 1 (above), shows the average monthly precipitation bar chart. The data has been grouped by month and the average rainfall has been calculated from each group to create the bars. This allows seasonal rainfall patterns to be understood, enables month to month comparisons, and allows identification of months with higher or lower rainfall.

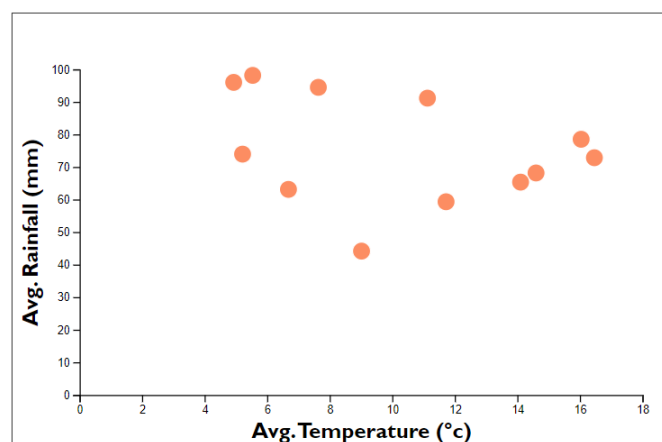*Figure 2.* Avg. Temperature vs Avg. Rainfall Scatter Plot

Figure 2 (above), shows the average temperature vs average rainfall scatter plot. The data has been grouped by month, then the temperature and rainfall averages have been calculated and displayed on a scatter plot. This can reveal the relationship, if one exists, between temperature and rainfall to see if there is a correlation between the two.
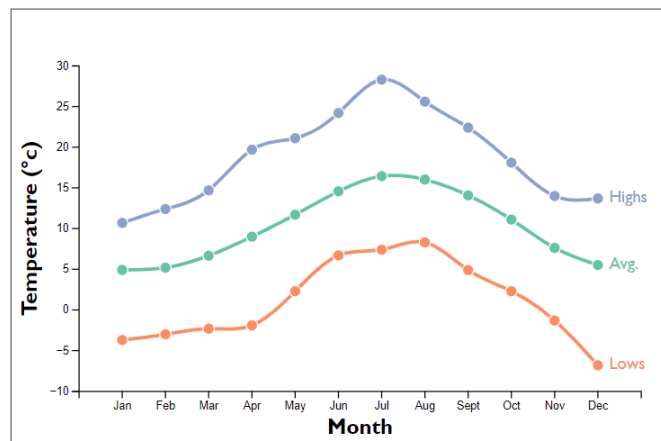
Figure 3. Month vs Temperature Multi Line Chart

Figure 3 (above), shows a multi line chart. The 'Lows' and 'Highs' lines show the lowest and highest temperature on record for each month respectfully. The 'Avg.' line shows the mean average temperature for each month over recorded years. This can visualise temperature trends throughout the year and compare how temperature varies by month. It also shows how the highest and lowest temperatures are correlated with the average temperature.

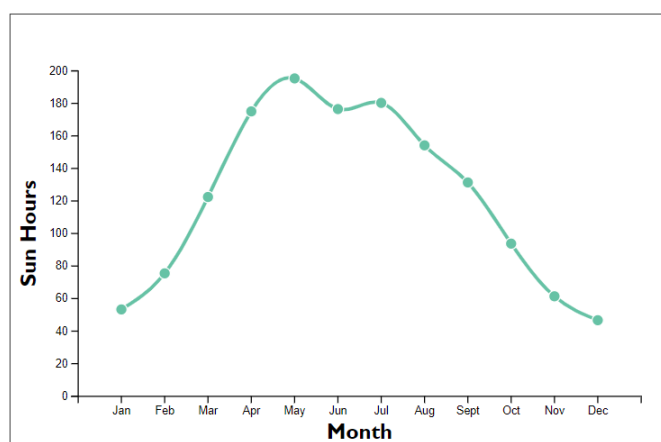*Figure 4. Avg. Monthly Sunshine Hours Line Chart*

Figure 4 (above), shows the average monthly sunshine hours as a line chart. The data was again grouped by month, and then the average sunshine hours was calculated. Similar to the last chart, this can be used to compare trends over time and to access daylight patterns.
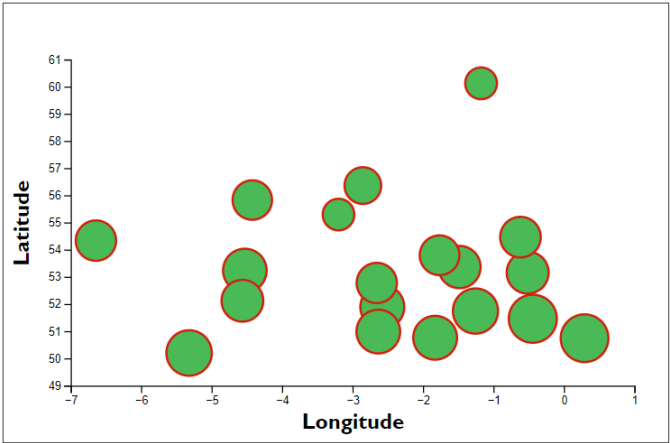

Figure 5. *Avg. Temperature by Location Bubble Chart*

Figure 5 (above), displays average temperature by location. Data was grouped first by longitude, then by latitude and the mean temperature was calculated. Longitude is plotted against latitude (in a way plotting a map), each bubble then ends up representing a weather station. The size of each bubble is determined by average temperature. Larger bubbles are plotted towards the back, smaller ones on top. This can provide geographical insights into variations across different locations and regions, and how longitude, latitude, and temperature are correlated.

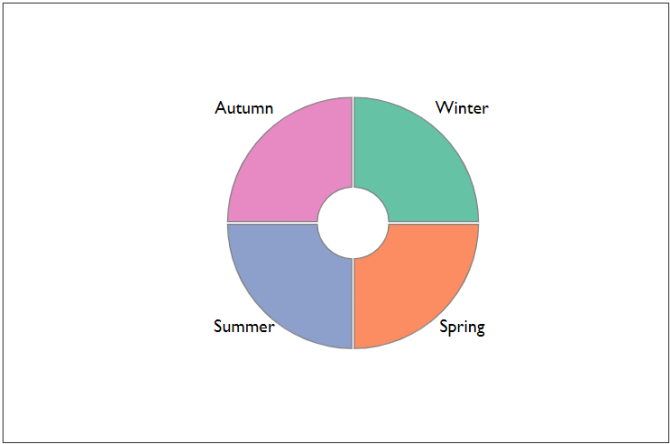*Figure 6.* *Seasonal Distribution of Dataset Donut Chart*

Figure 6 (above), shows the seasonal distribution of the dataset. The data has been grouped into segments by month determining the season, and the length (number of records) of each of these has been used to determine the segment size. This makes it clear how the dataset is composed by seasonal data.

Interactions

There were a fair few bugs which (frustratingly so) teased out, but this has seen a lot of scheduled time for the project eaten into. All tooltips are functioning as intended, as is individual chart highlighting. Shared highlighting ran into issues as mentioned in development (above).

Tooltips are provided for each chart allowing users to obtain detailed information about each data point. Highlighting of each data point when a user mouses over provides feedback to the user.

Figures 7-14 (below) show each of the visualisations with an element clicked and a datapoint / line being highlighted

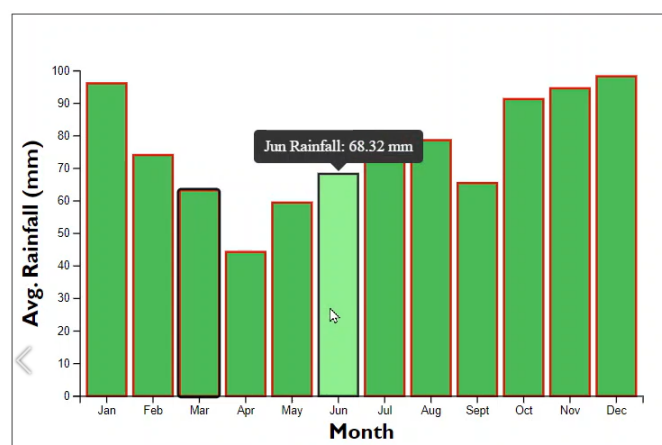*Figure 7. Highlighting and Tooltips, Bar Chart*

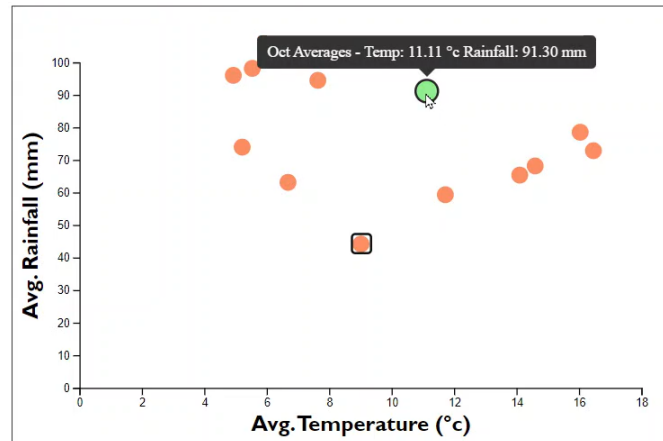*Figure 8.* Highlighting and Tooltips, Scatter Plot

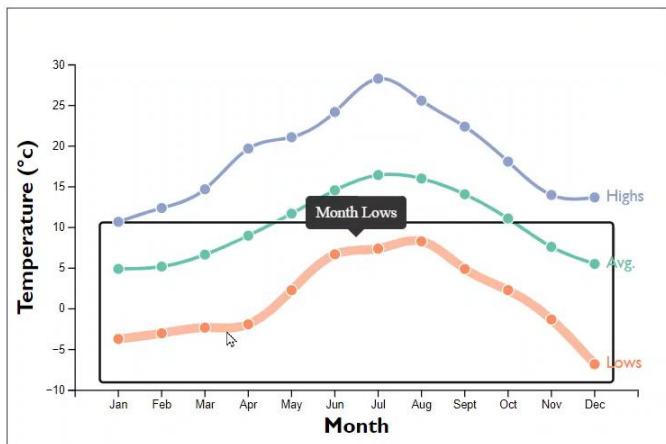*Figure 9.* Multi Line Chart Line Highlight and Tooltip

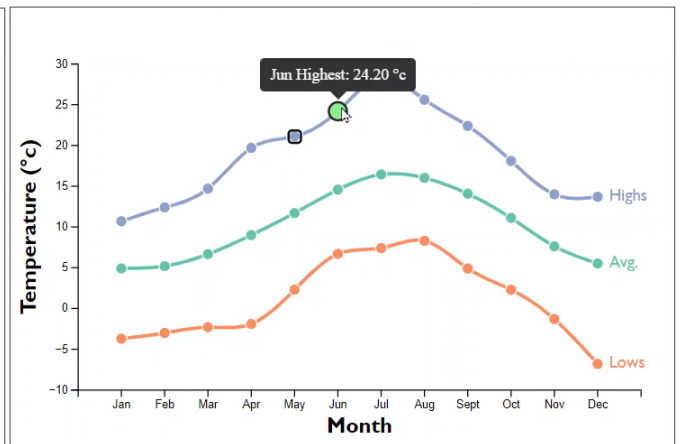*Figure 10.* Multi Line Chart Data Point Highlight and Tooltip

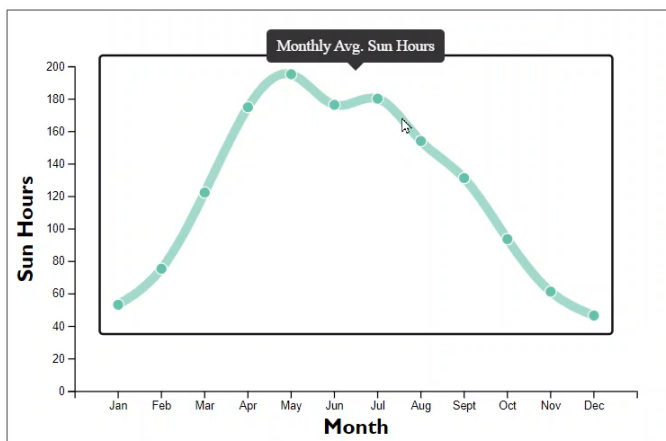*Figure 11.* Single Line Chart Line Highlight and Tooltip

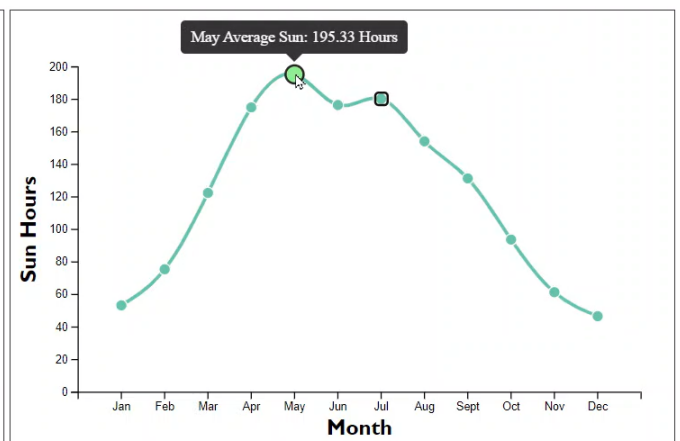*Figure 12.* Single Line Chart Data Point Highlight and Tooltip
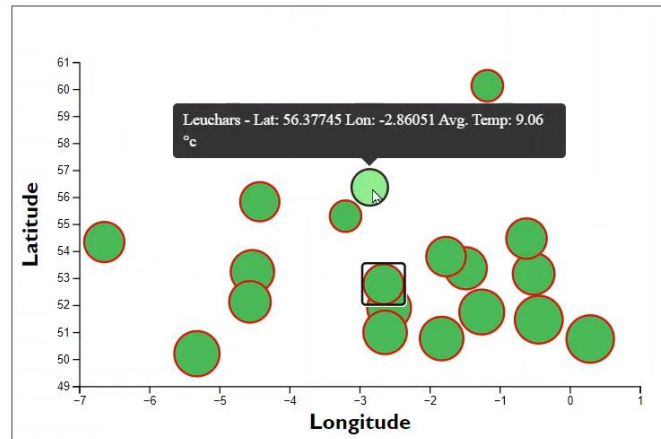
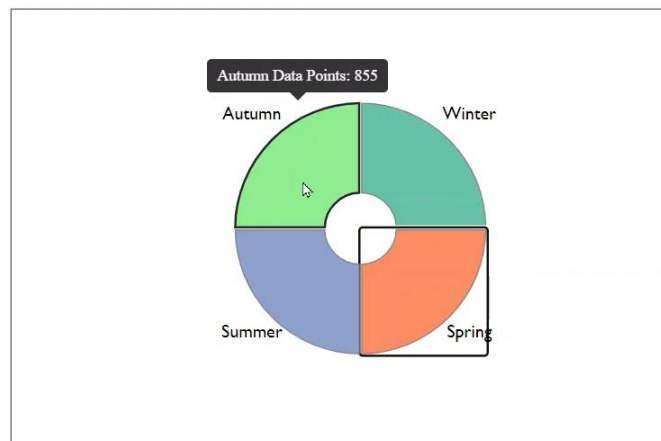*Figure 13. Bubble Chart Highlight and Tooltip*

Figure 14. Donut Chart Highlight and Tooltip

Insights

From the prototype dashboard, it can be seen that the dataset is equally distributed across the seasons and is well balanced. This means that statistical analysis along with any trends or patterns found in the data should be fairly reliable and a good representation without any, or little, bias.

Looking at the scatter plot, it appears that there are possibly two or three groupings of data and suggest that with lower average temperatures implies higher rainfall.

The multi line chart shows that the lowest temperature on record was in December, and the highest in July with the lowest average temperature in January.

It can also be seen with the bar chart that there is a steady increase in rainfall month by month from April to August, followed by a drop in September. The increase then continues until December when average rainfall tends to decline again until April.

Finally looking at the bubble chart, it can be seen that the higher average temperatures can be found in the south east.

References

[1] V. Martínez-de-Albéniz and A. Belkaid, "Here comes the sun: Fashion goods retailing under weather fluctuations" *European Journal of Operational Research*, vol. *294,* no. *3*, pp. 820-830, 2021. [Online]. Available: https://doi.org/10.1016/j.ejor.2020.01.064.

Appendix A – Trial Visualisation

Figure 15. Temperature trends across the years