CM50266: Applied Data Science
Weather Visualisation Report

Freud Fan

Nov. 2021

Visual Mappings Used and Why

Statistics for weather data calculated from a dataframe in Python are summarised as below.

	Maximum	Minimum	Mean	Standard Deviation
Humidity	59.0	37.0	48.727231	5.278108
Indoor Temp.	29.21	18.04	21.868713	2.048979
Indoor Temp. (Low)	28.2	14.9	20.608813	2.399339
Indoor Temp. (High)	31.1	19.7	23.559582	1.686869
Outdoor Temp.	26.38	-1.81	11.255206	5.350538
Outdoor Temp. (Low)	18.7	-4.1	7.977727	4.878844
Outdoor Temp. (High)	38.5	1.5	15.668797	7.021080
Baro	1035.6	979.6	1009.991547	9.733374
Rainfall (mm)	23.2	0.0	1.547852	3.265407

Table 1: Statistics of Weather Features Last Year (Full Dataset)

Missing data are detected and inserted with monthly mean values to lower bias to some extent. Also view two infographics for seasonal and monthly data in another attachment, in which **histograms**, **box plots**, **scatter plots** and **line charts** are used to illustrate weather variations across last year, **percentage**, **peak** and **valley values** are presented to highlight certain weather features.

In particular, a histogram is used for seasonable **mean** values in the dataset. One can see at once the distribution of mean baro across four seasons, represented by different brightness levels of colour. In which range does the majority of the data fall for different seasons can also be clearly viewed through this histogram.

A scatter plot and several brief line charts are also produced for seasonable indoor and outdoor temperature to show their **correlation** and **trend**. With the dots encoding relative positions, one can easily conclude that there are positive correlations between indoor and outdoor temperatures, which is strongly expected regardless of which statistic among **minimum**, mean and **maximum** value is considered. On the other hand, the line charts allow readers to see the relative seasonable values via its slope - e.g. the mean temperature is at its highest during last Jun.∼Aug.

In addition, a box plot is here to visualise the **standard deviation** as well as the **trend** of the seasonable mean, maximum and minimum mean humidity, giving readers a vivid picture about the variation of extreme atmospheric conditions. For example, one will find that during last year, the highest humidity recorded lies in season 2, which is also the time when the humidity is most stable compared to that in the rest of the year.

For monthly infographic, line charts are mostly used to illustrate the **trend** from Jan. to Dec. According to Mackinlay's ranking on retinal variables, the line with its start and end point positions and its slope, can very well represent the relation between two variables, here the weather data and the month. Readers can see an overall trend of each weather component in different colours and possibly the difference between them. For instance, the drop between the highest and lowest outdoor temperature is large during season 2 and 3 but small during 1 and 4, in complete contrast to that of indoor temperature. The number of different colours used is kept under or equal to five by convention, which also matches that of the main weather components and the legend, with no redundant colour or line pattern to distract attention.

In both infographics, certain features are represented individually by a series of **numbers**, to highlight their significance. For example, the **percentage** of rainy days - by counting days when rainfall is nonzero - across months, seasons and the whole year; the total rainfall, or the lowest/highest indoor temperature, with the lowest marked blue and highest marked red indicating apparent temperature. They together with resemblant line charts all allow readers to catch the main points immediately on a seasonable or monthly basis, with constant referring from one another.