Data Visualization Notes (02/25/2019)

Sparsh Agarwal

Sparsha2

The LIGO between frequency and time was discussed. The visualization can be found here:

https://gizmodo.com/observatories-across-the-world-announce-groundbreaking-1819500578

The visualization shows the change in waves frequencies over period of time (in seconds). As the energy of the wave with a particular frequency increases, the more yellow in color the wave is represented. It can be seen in the visualization that over the time, increase in energy is resulting in increase in frequency of a particular wave and this the wave that was detected by LIGO scientists. The strength of the visualization is that it represents a wide range of time as well as frequency, which is very important to represent this type of scientific data. The color bar shown as right side is also continuous and gives a precise representation of change in energy. The weakness of the plot is that the color of the wave seems to be getting mixed with the background, thus, it is difficult to follow the path of the wave. Also, as this visualization is made primarily for scientific community, interpreting the information from this plot is difficult for general public.

Pie chart: In the pie chart of zoo animals data, it is difficult to identify the class of animal having the largest size. Thus, pie chart should not be generally used for representing quantitative data, but should be used for representing qualitative data. Even if pie chart is marked with quantitative values, such as percentage numbers, it will still not be good representation, making the pie chart even worse. Instead, the data should be plotted on a bar graph Also, pie chart can be misleading sometimes in terms of quantitative information due to principle of proportional ink. In place of pie chart, the donut chart can be used as it represents quantitative information with help of its arc length. It is still not a very good option as looking at the arc length (the angle involved) is also sometimes difficult.

The tree map is also useless as it has the same problem as pie chart, so it should be avoided too to represent such data. The only advantage of tree map is that it lets you compare data in 1-D. Among all the plots, the horizontal bar chart and the waterfall chart would be the best as they represent all the data in 1-D which makes it easy to understand quantitative data.

For representing hierarchical data, the sunburst or nested box area plots can be used. In sunbursts, as you goes towards outside direction, the more you break into child category. In nested box plot, the area of the box helps in understanding the quantitative information. Stacked bar chart and stacked line chart represents the popularity of animal data very nicely, and can be an alternative representation to pie charts.

Pandas:

BQ plot: It is quite similar to matplotlib, the only difference is that bq plot is more interactive. Two ways to interact: pyplot—set up an environment and update it as you go. Object oriented :takes everything in one go.

Objects: A vehicle object is represented by the number of wheels, color and weight of the vehicle. A phone can't be described with vehicle object as it's traits are entirely different. We define classes which is used to define an object of that particular class. The objects that are made by these classes are used in trailets library. An object can be given specific values to its traits, for instance, declaring a specific person object name with name Weezer. These names can be updated using interactive feature of trailets. Change is akeyword to traitlets. Ipywidgets is used to interactively change traits. More things can be done using ipywidgets like making button objects, or representing data in a tab

Data Visualization Notes (02/25/2019)

Sparsh Agarwal

Sparsha2

format just like we do in chrome or other web browsers. We can also link the objects with ipywidgets so that whenever we make some changes to the object, the changes are reflected into the interactive environment as well.