SUMMARY – A Tour through the Visualization Zoo

Amount of data we create every minute is electrifying. In order to draw insights from such a huge amount of data we need to analyse the data and find out effective ways to present it in a format which conveys concepts easily. A picture being worth a thousand words, data visualization helps to grasp difficult concepts easily and in an interactive manner.Graphical parameters such as position, shape, color and size are used to map data values.

Author talks about different types of data such as Time Series data which can be represented using index chart (line chart) that shows percentage changes for time series data; stacked graphs for interpreting relative values; small multiples which use small charts for every series for effective visualization; horizon graphs which efficiently use space by splitting graph into bands and layering them.

Some charts are used to understand statistical distribution of data like Stem and Leaf Plots are used for assessing a collection of numbers where a table is used which consists of stem (first digit/digits) and leaf (last digit). Here the content itself is used to represent the plot instead of bars in the histogram. Q-Q Plot compares probability distributions by plotting their quantiles against each other. Scatter Plot Matrix is a collection of multiple scatter plots each representing a pairwise relation between variables. Parallel Axes use multiple axes for variables and a line connecting all the points on these axes represents one row of data.

Maps are another ways of visualizing data which involves flow maps that are used to represent the flow of objects/quantity from one place to another with respect to time; choropleth maps that use color shading to define values of quantities in those areas; graduated symbol maps that does not confound to geographic boundaries and use different symbols with parameters like size, color, position to define data; Cartograms that replace geographical area with some symbol to represent variables like population, obesity etc.

Most of the time data needs to be arranged hierarchically and some of the representations used here are Node-link diagrams that represent a tree-like structure wherein a parent and the child are connected using links; Adjacency list which is a space filling version of node link diagram; enclosure diagrams that encloses child inside parent and is also space filling.

Relationship between data is the most important aspect and some graphs that represent networks are Forced directed layout which represents undirected graphs with nodes and links (forces between nodes); Arc diagram that uses one dimensional layout of nodes and the nodes are linked to each other using circular arcs ; Matrix views that represent information in the form of matrix such that value in row i and column j represents link from i to j.

In this way, the paper describes time series data, statistical distribution, hierarchy, maps and networks and specifies different types of representations for the same. It also states that graphical features like size, position, color and position are important aspects of data representation. The conclusion is that data visualization is a very effective tool for getting insights from the data and there are numerous ways to represent data in a visual form which is based on the type of the data.