Global Journal of Enterprise Information System

Graphs for Research

Sudhansh Sharma[1*]

SOCIS, New Academic Complex, IGNOU, Maidan garhi, New Delhi - 110068; sudhansh@ignou.ac.in

Graph transforms raw data into information, nowadays the electronic environment leads to exponential data growth rate. To analyze such voluminous data, Big Data Analytics is the upcoming field, where Graphs plays dominant role. Further, in any research, this graphical representation of the data strengthens the data analysis to a great extent. Business Data Analytics uses statistical techniques to a great extent; the outcomes of such techniques are generally represented in the form of graphs for better interpretation and analysis. In this paper after studying lots of research work from various disciplines, a brief report of the various types of graphs is prepared along with their area of application and possible alternative techniques. The work performed in this paper is domain independent, and the outcome of the performed work will help to strengthen the research in all disciplines and domain.

Keywords: Application Areas, Big Data Analytics, Data Mining, Types of Graphs, Statistical Techniques

1. Introduction

Graphical analysis and its outcome are very much acceptable in entire research world, as it is a way to represent the research finding in precise and concise way. The graphical representation of research data strengthens the findings and hence the interpretation of data too. This technique of graphical representation is domain independent, be it psychology, computer science, management, physics etc. each and every discipline appreciates this visual data presentation mechanism.

Since it's the time of interdisciplinary studies, it is very much required that one should know about the tools and techniques of other disciplines. To speedup the interdisciplinary understanding graphs may play a vital role in research. Applying the understanding of randomness or probability or other statistical techniques to the data, one may go for data mining, big data analytics, database optimization and many more. The outcomes are very much supported, if they are presented in graphical format. In this paper is consolidated in table-1 below, where various types of graphs are tabulated along with their application areas and alternative techniques.

In the third international workshop on Distributed Statistical computing conducted in Vienna, Austria; Mayer, Zeileis and Hornik et al. presented their work on Visualizing independence using extended association plot. Their work gave insights into various graphical plots which are used to visualize the independence of data; the discussed plots are mentioned in table 1 below. The work performed by Jacoby et al.[1] and Johnson[2] gave detailed description of various graphical techniques for multivariate data analysis. The data plot reference manual designed by Filliben et al.[3,4] cumulatively discussed the graphical techniques for various typed of analysis like cluster analysis, gap analysis, multivariate analysis etc. The graphical structures used to address the stochastic phenomenon are discussed by Medhi et al[5].

The outcome of the literature survey in table 1 given below:

Table 1. Graphs and their application areas along with alternative techniques

S.No.	Graph	Application	Related Techniques
1	Age Pyramid	Genderwise Demographic variation	Histogram, Bi-histogram, Multiple Bar Diagram
2	Andrews Plot	Cluster Analysis , Detection of Multivariate Outlier, Multivariate analysis	Parallel Coordinate Plot, Icon Plot
3	Anom Plot	Quick check of ANOVA results for one way classified data	Jitter Plot, Control Chart
4	Area Chart	Gap Analysis	Line diagram, Stacked Line Chart
5	Association Plot	Contingency Table Data Representation	Mosaic Plot, Sieve Diagram, Chi-Square test for Independence in Contingency Table

*Author for correspondence

Table 1. Continued

6	Auto Correlation Plot	Quick Check over the Randomness of data	Run Test, Lag Plot
7	Bar Chart	Comparative analysis of one or more variables under different categories	Pie Diagram, Histogram, Multiple Bar Diagram
8	Bihistogram	Simultaneous detection of several parameters like location, dispersion, symmetry/ skewness, outliers	Block Plot, ANOVA, t-Test, F-test
9	Binomialness Plot	Checking goodness of fittness of data to binomial distribution	Probability Plot, Poissonness Plot
10	Cartogram	Geographical comparison of various features	Pictogram
11	C-Chart	To identify the existence of any assignable cause of variation	Process control chart, p chart, s chart
12	Chi Plot	Visual testing of variable dependence, study randomness of data set, Randomness of residual from fitted model	Run Test, Serial Correlation Test, Run Sequence Plot
13	Chi Diagram	basic Study of goodness of fit of a data set	Histogram, Residual Histogram, Residual Root gram
14	Contour Plot	Visualizing 3D plane in 2D Surface	
15	Coplot	Analysis of the effect of third variable over the relationship of two variables	scatter plot matrix, scatter diagram
16	Correlation Plot	Too study the independence of two time series	Auto-Correlation plot, Serial Plot, Lag Plot
17	Dendrogram	Visualize single linkage clustering, Visualize Euclidean distance, division of data into any number of clusters	Hierarchical clustering method, Complete Linkage Clustering
18	Deterended Probability Plot	Checking the normality of the residuals from the model	Probability plot, Normal Probability Plot, Chi-Square Test
19	Deviation Plot	Visualization of the mean of a set of observations, categorical data analysis	Bar Diagram, Residual Bar Plot, Column Plot
20	Double Y-Axis Plot	Analyze variation of two variable with respect to the third variable	Categorical Scatter Plot, Glyph Plot
21	Droughnut Chart	compare various components of the response variables	Bar Diagram, Pie Chart, Pie Icon Plot
22	Empirical Distribution Function (Edf) Plot	Check normality of data	Probability plot, Normal Probability Plot, Chi-Square Test for the goodness of Fit.
23	Error Bar Plot	comparison of the central values of a number of groups	Jitter Plot, ANOVA, t-test, F test
24	Lag Plot	Deciding suitable model for a given time series data	scatter plot, auto correlation, serial correlation
25	Lorenz Curve	Generally used in economics to describe inequality in the distribution of wealth amongst the population	Gini's Coefficient, Lorentz Asymmetry Coefficient
26	Moving Average Plot	To visualize/compare the trend value variation with actual value of data	line diagram, histogram, Run sequence plot
27	Mean Sum Of Squares (Mse) Plot	Compare distributional pattern of several time series data, also used to search best model for given time series data	Tests of Goodness to fit, Tests for Linearity of Data
28	Normality Plot	To quickly check the normality of data, a necessary condition for many statistical tests	EDF Plot, Probability Plot, Chi Square Test for Goodness to fit tests
29	Ogive	To Find Median, Quartiles, Deciles, Percentiles of a frequency distribution, Cumulative frequency for given value of variable.	Frequency Polygon, Histogram, Frequency Distribution

Table 1. Continued

30	Pareto Plot	To study the type of economy in the country	Lorenz curve, Engel's Law
31	Poissonness Plot	To check the goodness of fit of data to Poisson Distribution	Probability plot, Chi square test for goodness of fit
32	Probability Plot	Quick check to the goodness of fit test	EDF Plot, Normal Probability Plot, Chi Square Test for Goodness to fit tests
33	Quantile-Quantile (Q-Q) Plot	Compare Tail behavior of two data sets; check if the two data sets have similar distribution shapes	EDF Plot, Probability Plot, Chi Square Test
34	Residual Plot	To Check the randomness of the residuals from an model.	scatter diagram, serial correlation plot
35	Run Sequence Plot	To identify the shift in the location of data or any shift in the spread of the data	scatter plot, Lag Plot, Auto Correlation Plot

2. Conclusion

The work performed in this paper is expected to strengthen and speedup the research. The young researchers will get a glimpse and idea of various graphical techniques available at their disposal and they can identify which graphical technique is to be used for which purpose. The work performed in this paper will bring conceptual clarity to the researchers about graphs and their usage areas.

3. References

1. Jacoby W. Statistical graphics for visualizing multivariate data. Sage University Papers Series on Quantitative Applications in the Social Sciences. Thousand Oaks, CA: Sage; 1998.
2. Johnson RA, Wichern DW. Applied multivariate statistical analysis. 3rd ed. NJ, USA: Prentice-Hall International; 1992.
3. Filliben JJ. Dataplot reference manual. Statistical Engineering Division. Information Technology Laboratory: NIST; 1997.
4. Filliben JJ, Cetinkunt Y, Dommenz. Explanatory data analysis techniques as applied to a high-precision turning machine. New York: Elsevier; 1993. 199–223.
5. Medhi J. Stochastic process. New Delhi, India: Wiley Eastern Limited; 1994. p. 128.