

# Data Visualization A Tool of Data Mining

<sup>1</sup>Dr. Kishori Lal Bansal, <sup>2</sup>Satish Sood

<sup>1,2</sup>Deptt. of Computer Science, HP University, Summer Hill, Shimla H.P., India

## Abstract

As databases grow ever bigger and as computer hardware grows more powerful in creating and storing the data and presenting it to the end user, there are even greater requirements for tools which provide visualization of data. Visualization technologies graphically display the data in the databases. Much research has been conducted on visualization and the field has advanced a great deal especially with the advent of multimedia computing. For example, the data in the databases could be rows and rows of numerical values. Visualization tools take the data and plot them in some form of a graph.

## Keywords

Data Mining; Data Visualisation.

## I. Introduction

Data visualization is the study of the visual representation of data, meaning “information which has been abstracted in some schematic form, including attributes or variables for the units of information”[1]. According to Friedman (2008) the “main goal of data visualization is to communicate information clearly and effectively through graphical means. It doesn’t mean that data visualization needs to look boring to be functional or extremely sophisticated to look beautiful. To convey ideas effectively, both aesthetic form and functionality need to go hand in hand, providing insights into a rather sparse and complex data set by communicating its key-aspects in a more intuitive way. Yet designers often fail to achieve a balance between design and function, creating gorgeous data visualizations which fail to serve their main purpose - to communicate information”[2]. Data presentation can be beautiful, elegant and descriptive. There is a variety of conventional ways to visualize data – tables, histograms, pie charts and bar graphs are being used every day, in every project and on every possible occasion.

## II. History

In the early 17th century the visualization was seen in the form of Maps and diagrams. In 6200 B.C. the map of Babylon was carved in stone. Similarly the world map’s description was in the book of Herodotus Anaximander of Miletus (c.610BC-546BC). In the year 1600-1699 the new technique was developed in the form of measurements and theories which included time, distance and space for astronomy, navigation and territorial expansion. In 1686 first known meteorological map came in to existence. As the years kept on moving, the year 1700-1799 gave the new graphical form of visualization techniques which included graphs which were easy to study. The year 1800-1900 all forms of static graphs which are used today were developed.

## III. Data Visualization Tools

Data visualization tools are used to create two and three dimensional pictures of business data sets. Some tools even allow us to animate the picture through one or more data dimensions. Simple visualization tools such as line, column, bar, and pie graphs have been used for centuries. However, most businesses still rely on the traditional “green-bar” tabular report for the bulk of the information and communication needs. Recently, with the advance of new visualization techniques, businesses are

finding they can rapidly employ a few visualizations to replace hundreds of pages of tabular reports. Other businesses use these visualizations to augment and summarize their traditional reports. Using visualization tools and techniques can lead to quicker deployment, result in faster business insights, and enable us to easily communicate those insights to others. The data visualization tool used depends on the nature of the business data set and its underlying structure. Data visualization tools can be classified into two main categories:

- Multidimensional visualizations
- Specialized hierarchical and landscape visualizations

**Multidimensional Visualizations:** The most commonly used data visualization tools are those that graph multidimensional data sets. Multidimensional data visualization tools enable users to visually compare data dimensions (column values) with other data dimensions using a spatial coordinate system. Fig. 1.1 shows examples of the most common visualization graph types. Most multidimensional visualizations are used to compare and contrast the values of one column (data dimension) to the values of other columns (data dimensions) in the prepared business data set. They are also used to investigate the relationships between two or more continuous or discrete columns in the business data set.

**Specialized hierarchical and landscape visualizations:** Hierarchical, landscape, and other specialized data visualization tools differ from normal multidimensional tools in that they exploit or enhance the underlining structure of the business data set itself. We are most likely familiar with an organizational chart or a family tree. Some business data sets possess an inherent hierarchical structure. Tree visualizations can be useful for exploring the relationships between the hierarchy levels. Other business data sets have an inherent geographical or spatial structure. For instance, data sets that contain addresses have a geographical structure component. Map visualization can be useful for exploring the geographical relationships in the data set. In other cases, the data set may have a spatial versus geographical structure component. For instance, a data set that contains car part failures inherently has spatial information about the location of the failure within the car. The failures can be “mapped” to a diagram of a car (a car landscape). Another data set may contain where in the factory the failing part was manufactured. The failure can be “mapped” to a diagram of the factory (a factory landscape) to explore whether the failed part has any significance to the location where it was manufactured.



Fig.1: Multidimensional data visualization graph types.

#### IV. Data Mining

Data mining can be defined as the process of discovering previously unknown patterns and trends in database and applying that information to create predictive modules [3]. It can also be stated as the process of data selection and exploration and creating models by applying enormous data stores to open previously unknown patterns [4]. Data mining is the most powerful new technology with great potential to extract meaningful information from large and mostly unorganized data banks. It is the process of performing automated extraction and generating predictions and forecasting information from large data sources. Data mining tools predict future trends and behaviors allowing businessmen to take proactive measures to gain maximum benefit from the same. "Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets [5]". These models can include statistical models, mathematical algorithms, and machine learning methods (algorithms that improve their performance automatically through experience, such as neural networks or decision trees). Consequently, data mining consists of more than collecting and managing data, it also includes analysis and prediction [6].

#### V. Uses of Data Visualization

Most good data visualization allows the user some key attributes:

- Ability to compare data
- Ability to control scale (look from a high level or drill down to detail)
- Ability to map the visualization back to the detail data that created it
- Ability to filter data to look only at subsets or sub regions of it a given time.

As a result, data visualization is used in a number of places within data mining:

- As a first-pass look at the "data mountain" that provides the user some idea of where to begin mining
- As a way to display the data mining results and predictive model in a way that is understandable to the end user
- As a way of providing confirmation that the data mining was performed the correct way (e.g. to confirm intuitions and common sense at a very high level)
- As a way to perform data mining directly through exploratory analysis, allowing the end user look for and find patterns so efficiently that it can be done in real time by the end users without using automated data mining techniques.

#### VI. Conclusion

Through the visualization technique even the common man can understand what the data is trying to present. So through this visualization technique it is not only easy but simple way to analyse the data and give the results in the smooth and simple way. But this will happen only if we understand it and use it properly. We must embrace what really works and jettison the silly stuff that undermines data visualization today.

#### References

- [1] Michael Friendly (2008). "Milestones in the history of thematic cartography, statistical graphics, and data visualization"
- [2] Vitaly Friedman (2008) "Data Visualization and Infographics" in: Graphics, Monday Inspiration, January 14th, 2008.
- [3] K. Kincade (1998), "Data Mining: digging for healthcare

gold". Insurance & technology, 23(2) IM2-IM7.

- [4] A.Milley (2000), "Healthcare and data mining", Health Management Technology, 21(8), 44-47.
- [5] Pieter Adriaans, DolfZantinge, Two Crows Corporation", Introduction to Data Mining and Knowledge Discovery, Third Edition (Potomac, MD: Two Crows Corporation, 1999); Data Mining (New York: Addison Wesley, 1996).
- [6] Jeffrey W. Seifert, "Data Mining: An Overview", Updated December 16, 2004 Analyst information Science and Technology Policy Resources, Science and Industry Division.
- [7] S. Dai, P. Zhang, "A data mining algorithms in distance learning", Electronic Edition, BibTeX, 2003, pp. 1014-017.
- [8] A. Joshi, R. Krishnapuram, "On mining web access logs". In Proc. Of the 2000 ACM SIGMOD workshop on Research Issues in Data Mining and Knowledge Discovery, pages 63-69, 2000.
- [9] T. Dingsoyr, (1998). "Integration of data mining and case-based reasoning".
- [10] Ming Syan Chen, Jiawei Han, "Data Mining An Overview from a Database Perspective", IEEE Transactions on Knowledge and Data Engineering, Vol. 8, No. 6, December 1996.
- [11] Mattoso & N. F. F. Ebecken COPPE, "Data mining: a database perspective". M. S. Sousa, M. L. Q. Federal University of Rio de Janeiro, RJ, Brazil