Knowledge discovery in databases from a perspective of intelligent information visualization

Juan C. Alvarado-Pérez

Faculty of Engineering
Universidad de Salamanca
Universidad Mariana
jcalvarado@usal.edu.co

Harold Bolaños-Ramírez, Diego H. Peluffo-Ordóñez

Faculty of Engineering
Universidad Cooperativa de Colombia,
Sede Pasto
harold.bolanos@campusucc.edu.co
diego.peluffo@campusucc.edu.co

S. Murillo

Grupo de investigación en Ingeniería de Software
Universidad Autónoma de Manizales
smurillo@autonoma.edu.co

Abstract

This paper reviews some recent and classical, relevant works on information visualization with a special focus on those applied to big data. The central idea dealt in this work relies on how to perform data mining tasks in a visual fashion; that is, using graphical correlation and interaction techniques. The scope of this review encompasses visualization techniques, formal visualization systems, and smart information visualization models. As well, newest approaches consisting of visualization and data mining integration process are explained.

1. Introduction

The exponential growth of the amount of information coming from all the fields of knowledge has naturally caused a high demand for automatic data processing systems. Typically, such systems extract meaningful knowledge from Big Data through rules and models. In this sense, data mining (DM) has shown to be a collection of powerful techniques. Nonetheless, when dealing with Big Data the resultant set of rules is usually represented in a plain fashion, requiring then data analyst to have special skills and expertise to understand such rules and patterns to make the extracted knowledge more intelligible. Because of this, data visualization and visual data exploration play a relevant role in the knowledge discovery in databases (KDD) [1], [2]. Data visualization and smart information scanning tools offer a wide-spread spectrum of possibilities enabling a better understanding of data rules and models through multiple interactive views of the information. Indeed, visualization stages might be incorporated into the DM frameworks. For instance, visual mapping techniques are currently used for both making DM outcomes more intelligible and facilitating users to understand how the methods work. This visual interaction allows analysts to carry out improved and more controlled KDD process, extracting then knowledge more properly. Likewise, researchers have found that visual information plays an important role within the DM algorithms themselves. For instance, visual mapping techniques have been used in two ways: To export DM algorithm results in a more expressive way for the final user, and to help the understanding about the algorithm performance [3]. Also, analysts can make an improved control in the KDD process as well as a knowledge extraction with better quality. In this paper, a review about the relevant works in information visualization (InfoVis) of Big Data is presented. Mainly, discussed works are about the ways to do graphics correlation- and interaction-based techniques by means of visual DM. The present review is focused on visualization techniques, formal visualization systems and models for smart InfoVis.

This paper is organized as follows: In the first section, basics and concepts on datasets, DM and visualization are gathered. Section 2 explains different visualization techniques and special considerations thereof. Next, several visualization systems are described in Section 3. Finally, Section 4 explores techniques for both improving InfoVis using DM techniques, as well as generating intelligible views of the DM results by importing some InfoVis properties. Also, the newest approaches to integrate visualization and KDD processes are studied, which are aiming at improving the user interaction and understanding.

2. Information representation

Information visualization, in short InfoVis, means the use of interactive interfaces with the purpose of representing datasets in a low-entropy way, being more usable for the final user [4]–[7]. In other words, it makes graphics models and visual representations supporting the interaction with the user to explore and extract the information that underlies the data [8]. Smart visualization refers to graphical models built from the representation of simulated objects (normally related to concepts associated to phenomena of the physical world). For instance, the data representation can be embedded into a one- (1D), two- (2D) or three- (3D) dimensional model. Additionally, in some cases, models includes a temporal dimension. Then, time-space visual representation takes place [8]. It is important to remark that data visualization does not always represent physical variables from the real world regarding the time effect. Mostly, visualization represents abstract concepts, for instance the logging information of an internet site, or the number of possible

978-1-4673-9461-1/15/$31.00 ©2015 IEEE

buyers regarding their preferences about a car brand selection. Given this, each data unit or datum usually describes many related features (four or more), which have no a defined time-space nature. Raw data may be represented in many formats, nonetheless a convenient form to representing them is via structured arrays, so-called tuples. The study presented in [9] defines the data table modes as a format of data structure organized by means of rows and columns that holds relationships, moreover corresponding metadata establish such relationships and labels for rows and columns. Normally, these arrays are organized so that the rows represent variables, and the columns represent the cases (Then, a set of values is arranged for each variable). The disposition where the columns represent the variables and the rows represent the cases is more common. The row and column order in a data table should not be relevant in a general rules definition out of data structure [10].

3. Clasification of Visualization Techniques

Keim and Kriegel in [11] group multivariate and multidimensional visualization techniques in six classes: geometric projection, icon based, oriented to pixels, hierarchical, graphics based and hybrid. In both works, a general description is given and the representative techniques are compared. In Figure 1, the classification provided by this author is graphically explained.

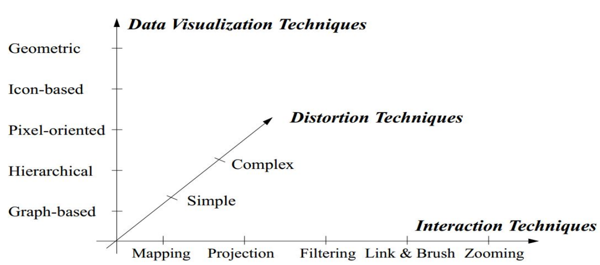

Figure 1. Data visualization techniques. Taken from Daniel A. Keim[12]

3.1 Geometric projection

This technique gives support in works where users need to find informative projections from multidimensional data sets. It includes techniques of exploratory statistics extensively used for processing data as principal component analysis, factor analysis, multidimensional scaling and classical scale graphics, where each couple of features is projected into two dimensions (X and Y-axes).

3.2 Parallel Coordinates

The technique of parallel coordinates is widely known [13]–[15]. It consists of mapping k-dimensional data on a 2D view, where each dimension is associated with a parallel axis, so it should be drawn with k equidistant parallel axes. Also, linear scaling linearly in a standardized way can be applied by associating the corresponding range of attributes with the shaft size, if necessary. Each data item is represented using a polygonal line that intersects each axis corresponding to the items associated with the attribute value of the data point. The technique is effective to reveal a wide range of data characteristics, and their different distributions and functional dependencies. The major limitation of this technique is that it is not as representative a small data set, so if this happens can lead to error in the interpretation of the visualization. [10] adopted a radial arrangement to coordinate axes called Parallel circular Fua et al. [16] presents an extension adapted to cope with huge data sets. His approach was to deploy aggregation of information from the results of a hierarchical clustering procedure. The cluster data can be visualized at different levels of abstraction and using a color code based on proximity cluster approach, as shown in Figure 2. They also introduce an interactive mechanism for dynamic navigation and hierarchical filters, it is denoted as structure based on brushing [17].

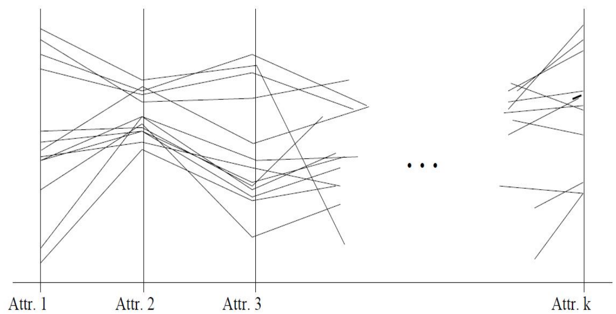

Figure 2. Parallel coordinates. Taken from Daniel A. Keim[12]

3.3 Radial Coordinate Visualization

Other technique based on geometry for high dimension treatment, Rad Viz [10]. The feature visualization is associated to a radial line and its contour intersection, n radial lines represent *n* dimensions, as seen in Figure 3.

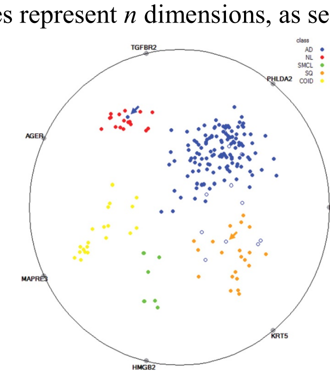

Figure 3. RadViz show a cancer lung dataset using gene expression information.

Dots represent tissue samples, its color is about the respective diagnosis (adenocarcionomas blue, cell carcinomas Brown, carcinoid yellow, lung small cells carcinomas Green and normal lung samples red). Scaly is shown as blue empty circles. Arrows show the representative cases in the work. Taken from Minca Mramor, et al.[18]

3.4 Visualization based on icons or iconographic

This visualization technique maps each multidimensional item in an icon representation, the visual features change as the date values. Chernoff faces is the first approximation [19]. Features are mapped into faces traits building an iconographic face, the shape of nose, mouth and eyes represent the original data. The principal issue of this technique is the subjective appreciation of human faces; usually people focus on eyes expression more than nose or ears disposition.

3.5 Pixel-based visualization

One pixel is used to represent values: different features are shown in a set of sub windows, and the variations are represented by a fixed color map [11], [20], [21]. The technique, suitable for large multi-dimensional data sets, is then categorized as "independent of query" or "query-dependent". In the independent query technic’s, the pixels arrays is fixed in the sub windows, independently of the data values. In the query-dependent, the query items are established in advance and distances from the values to the given query values are computed using some metric. The layout of the colors of the pixels is based on the computation of distances for each attribute, and pixels in each sub window are arranged according to their total distances to the item of data from a query.

Figure 4 shows the visualization using two possible pixels arrays, called, spiral and axes array (left and right, respectively), produced from of set of dummy data with normal distribution and five cluster. The set of data has 7000 records with eight attributes. Adopted color code ranges from the range of colors from yellow to green, blue, dark red, and a dark gray color, based on the distance from the item of data to the correct answer to the query: the items of data that best satisfy the query are colored with light yellow and those who are further away are colored with black. In Figure 4, the sub window upper-left corner in each visualization represents the color code and the array of data items in each case (spiral and axes). In general are called result windows, in which the pixels that represent each item are placed according to their overall distance relative to the exact answer that satisfied the query. The rest sub windows show the behavior of the eight attributes regarding the query. The pixels for each item are placed in the same relative position as general appeared in the results window. The regions of different colors allow identifying data clusters with a comparable distance and correlation between different attributes.

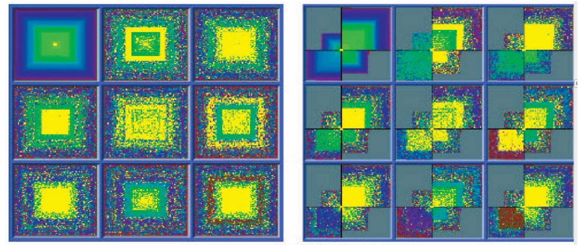

Figure 4. Visualization based on spiral pixels (left) and axes (right) using query dependent pixels array whit a 7000 records of dummy dataset and 8 attributes. Taken from [22]

3.6 Hierarchical

This technique subdivided the k-dimensions of the data space and organizes subspaces in a hierarchical way. This form of representation is known as "worlds within other worlds" [23], [24] and dimensions stacking [25]. That two techniques can be mapped the data table described in a non-hierarchical k-dimensional space on a hierarchical view in to 2D space. Treemaps [26] and Cone Trees [27] also are examples of hierarchical visualization technique, but they take on a hierarchical structure without the data space and thus its objective are not directly the data.

3.7 Graphic-based approaches

This technique allows for visualizing large graphics using specific algorithms of layers, query languages and abstraction techniques [28]–[31] to convey their meaning clearly and quickly. There are many other approaches and systems with specific objectives in the domain field, that appear on the category network, which are described in Section 6.

3.8 Hybrid

This technique integrates multiple visualization techniques, either in one or more windows to enhance the expressiveness of the visualizations. This methodology is a very useful resource and most techniques rely heavily on the dynamics and interaction.

4. Classification of visualization systems

Card et al. [9] adopt an alternative approach to the categorization of visualization of the information, grouping the applications on 4 different levels. At the highest level are visualization tools that allow users visual access to external data collections with intermediate development environments, such as the Internet or online databases on a server. On the second level are the tools which aims to provide support to users in the execution of tasks, allowing quick access and highly interactive visualizations of representation of information in the workspace required by the task. They are the spatial work tools that argue its use in give ability to interact within the workspace information. On the third level are visual knowledge tools that display visual representations of some data and a set of controls to interact with these representations where the users can determine and extract the data relationships. This category covers most of the tools that are used to represent knowledge from data tables. Finally, on the fourth level they are improved visualization objects that focus on providing additional information about an object from the intrinsic visually. A good example is the medical visualization of a human organ using direct visualization of the volume to represent internal structures. The knowledge visualization tools are also categorized based on the types of visual structures that they adopt. The concept of Visual structures involves how the space is used to encode the information, or in other words, the use given to the dimensionality of the represented data. Most common visual structures are:

Physical refers mainly to the representation of the data that has a direct correspondence with the real world objects, essentially this is scientific visualization. These include techniques for construction and visualization of 3D information, of objects of the real world as the human body, buildings, or molecules with the purpose of extracting information.

1D, 2D, or 3D refers to visualization that encodes the information into marks pointing out geometric coordinates on orthogonal axes. 1D visualization structure is used frequently to represent time-lines and documents, usually as part of a large structure of visualization. They also have the ability to allow control by themselves, using navigation and scroll bars indicate the range of values for a particular parameter. Examples of 2D visualization structures are scatter plots and matrices of scatter plots. 3D visualization structures are commonly for physical data, but also often used for compositions of 2D and 3D visualizations in abstract representations.

Multi –d are development that manage information visualization of abstract objects that have many attributes to be encoded directly in 1D, 2D or 3D; often, multiple attributes are not of the nature of the primary space, and do not have explicit structures or relationships. Scientific visualization also offers the possibility of Multi-D representations, but many of the scientific dataset have spatial attributes which are decisive for creation of visualizations. The most common tasks that can be supported by this kind of development involves obtaining knowledge of the data, how to find patterns, relationships, clusters and outliers, or find specific items using interaction as approaches actions, filters, and selection.

Trees and networks denote how visual structures used connections and attachments to encode data relationships. These correspond, with some exceptions, the groups of hierarchical techniques based graphics and classification of Keim. The hierarchy naturally occurs when described, for example, taxonomies, organized structures, and managing information on the disk. The purpose of visualization techniques is the simultaneous control of many nodes, not the whole tree, while providing mechanisms for browsing and searching that allow users to retain the overall structure of the tree and reduce disorientation. Hierarchies are similar to Multi-d data in the sense that its nodes usually contain a number of favorable attributes. Visualization of network structures often describe data consisting of nodes that also show the relationships between different points, additional information related to the items of data or connections. Much of the work in this field has already developed, however, the complexity of relationships and tasks the user especially in large networks, still leaves much to be done. Applications of this type of structures seen in areas such as networking and traffic management, digital libraries, and the visualization structure of World Wide Web (WWW).

Interaction gives the user the possibility to be able to use their perception of information when visually explore data sets [32] and virtually all visualization techniques are used in combination dynamics and interactivity. The ability to interact with Visual representations can greatly reduce the inconveniences of techniques, particularly those related to the disorder and the overlapping objects, providing the user with mechanisms to handle the complexity of large data sets. [12] identified two categories for the interaction techniques. The first group includes those operating on visual representation to allow to visualize large data sets. Mainly, many of these techniques have been designed as a component integrated into the tool, whose objective is in the specific domain and can currently be seen as visualization technique on itself, as view of FishEye for graphic visualization and hyperbolic trees for visualization of hierarchies. The second category includes those techniques that support more effective exploration of data because it allows a dynamic and interactive mapping of attributes to display parameters or direct interaction with visualization models, the best known examples are Linking-and-Brushing [33], dynamic queries [34], details on request [35]. Note in [36] a brief description of these and other interaction techniques and references.

Many of the visual exploration systems work on interaction techniques, some of these are available in academic tools (XmdvTool [37] and XGobi), other are available in commercial distributions (IVEE [38], now SpotFire). There are many web-based resources for exploration and data visualization.

5. Formal visualization models

The creation of visualization so far has been considered as a complementary process, without any formal method design, engineering, or evaluation, however, some studies have attempted to formalize the process of visualization. The usefulness of these formal models is very relevant. Firstly, because they offer a guide consistent to the user on how to address the process of visualizations from the data. Secondly, many of them can help to partially automate the process of creating visualizations. Thirdly, these models can provide an objective based on the comparison of the effectiveness of different views on the same data to select an effective task and thus be able to give new insights into the creation of new techniques. Much of the previous work on visualization models have aimed to both graphic presentation [39], [40], and scientific visualization [24], [41]–[45]. Many research efforts have been carried out with the purpose of automating the visualization process influenced by the presentation tools that have been developed [40]. Although it is important to mention that, these do not offer much help in analyzing visualizations in high-dimensional data tables such as scientific data, since these differ from the natural data. A notable exception is the reference model for state data proposed by Chi and Riedl [46], using a framework for characterization operations of the different techniques of visualization. Chi [47] argues that, although the taxonomy of visualization techniques help end users to dominate data, they do not help to implement an understanding of the design options and the potential applicability of these techniques. State of data model (see Figure 5) is used as a basis for visualization taxonomy.

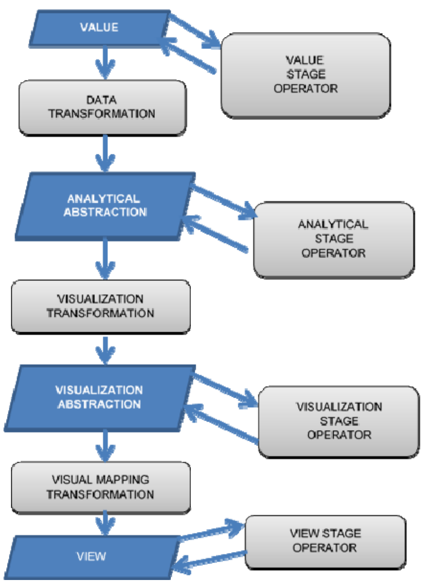

Figure 5. Reference model of States of data for visualization of information. Taken from Chi [46].

This model breaks the source of information inside of four data states and puts a Data Transformation operator. The steps reflect the nature of the data while they are transformed. In the initial stage -called value data-, the operations are applied to the original data. Likewise, in the Abstraction Analytical Stage, the operations are applied on metadata or information extracted from the data, meanwhile in the Abstraction Visualization Stage, the operations are applied to the visual information displayed on the screen. Finally, in the View Stage, the operations are applied on the visualization as a whole. The data processing operations are of three types, and transform data from one stage to another to produce manipulated data abstraction at every stage along the information visualization. The operators of the framework created by Chis provide a conceptual model that extracts all critical visualization operations (significant interactions of users) throughout the process information visualization, enabling the identification of important design artifacts. At the same time, this allows to end users to get better results while carrying out the visualization process, creating a simple understanding of the operators that are being applied, when it can be applied and what it can do. Final users can predict the results of many of the actions of the interactions and choose the appropriate operators to get a desired result. Designers can classify and understand the relationships between operators and the composition of the interactions and identify similarities between the different techniques, having the potential to strengthen the system modularization and standardization.

6. New approaches to intelligent information visualization

Much of the commercial software systems marketed as Visual Data Mining (VDM) has a high component of visual interaction for data exploration, with many levels of support for the pre-processing data, external database connections, and possibly, data mining algorithms. The biggest challenge with this type of tool is that many have not standardized the visual techniques that are scalable respect to large data. Grunenwald and Goebel [47] [48] conducted a survey of several data mining tools and knowledge discovery both commercial and academic, and the revised total of 42 products, 23 said include resources to observe the analysis conveniently (Visualization Models) and 10 also included to facilities for explore the data analysis. However, the visualization have not a important role that supports data mining (DM), this is not widely integrated into commercial software. Many data analysis using visualization as part of a process of integration strategy of data mining and visualization to get a goal, an approach commonly found in many research applications and Visual DM techniques. As noted by Wong (Wong, 1999), usually analytical techniques extraction does not rely on visualization. Many of the works that describes the approaches and applications Visual DM techniques are classified into two categories. The use of visual exploration systems or techniques that support a target of extracting knowledge or specific data mining task, or the use of visualization to display the results of a mining algorithm, as a cluster or classification process, and this mode provides greater understanding of the results to final users. Below are described some cases.

6.1 Visual data exploration mining

The tasks of data mining usually require techniques capable of handling large multidimensional data sets, often in the form of data tables or relational databases. Parallel coordinates and scatter plots are most often used in this context. Also, interaction mechanisms for filtering, consultations and selection of data are usually required to handle large data sets. This point is strongly emphasized by Inselberg [49] in a paper that illustrating the strong integration of parallel coordinates with effective and interactive consultation mechanisms with visual signals in the discovery process. Hoffman et al. [50] describes a case study showing how the technique of visual exploration of higher dimension as RadViz, parallel coordinates and graphics Sammon [51], have used a combination with classifiers based rules and neural networks to classify the DNA sequences. Cvek [52] used analytical and visualization techniques to mine the data set of functional genome of yeast. They compared exhaustively classification and clustering techniques on the same data set, showing the application of parallel coordinates, circle segments and RadViz, helped to get the expected data, and visually compared and contrasted analytical techniques. These and other writings and websites describe the mining and visualization tools applied to Bioinformatics [53] clearly shows this domain as one of the possible challenges of research in mining and visualization.

6.2 Viewing Mining Model

Another typical use of visualization on data mining consists in to make converge the results of mining tasks, such as clustering or classification, to facilitate the interpretation of the user. As example, is presented the tool BLOB and BLOB H-clustering algorithms [54][55] tool, that using implicit interfaces for display data cluster.

6.3 Visual data mining

Wong [56] argues that, instead of using tools of visual data exploration and data mining algorithms separately, strong strategy of data mining would be to merge processes of data mining and information visualization in one data analysis tool. Many mining techniques involve different phases of math that require the user intervention. Some of these can be complex and visualization can withstand processes involving decision making. From this perspective, a technique of visual data mining is not just a visualization technique being applied to exploit the data in some phases of a data mining process, but data mining algorithm with visualization can have better performance. Although current data exploration systems will certainly play an important role in the future and is likely to visualization systems with data mining must change to adapt to the new paradigm.

6. Final remarks

This papers presents an overview of works on the incursion of smart visualization techniques for representing information as a key stage for designing Big Data analysis systems. As discussed, visualization techniques has enhanced the data mining tasks performance by introducing visual tools into the knowledge discovery process. One of the most important benefits is that better understanding of the rules and patterns is brought visually closer to the user.

Still, since the use of InfoVis within the Big Data and data mining context is relatively new, there are many open issues and challenges to deal with. Develops in this emerging topic may significantly contribute to the field of data exploratory analysis.

References

[1]U. M. Fayyad, “Mining Databases: Towards Algorithms for Knowledge Discovery,” *Data Eng. Bull*, vol. 21, no. 1, pp. 39–48, 1998.

[2]U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and others, “Knowledge discovery and data mining: Towards a unifying framework,” *Knowl. Discov. Data Min.*, pp. 82–88, 1996.

[3]C. Ahlberg and E. Wistrand, “An Information Visualization and Exploration Environment,” in *Int’,l Symp. Information Visualization*, 1995, pp. 66–73.

[4]J.-D. Fekete, J. J. Wilk, J. T. Stasko, and C. North, “The Value of Information Visualization. Theoretical Foundations of Information Visualization,” in *Information Visualization: Human Centered Issues and Perspectives*, Lecture Notes In Computer Science, Ed. 2008, pp. 1–18.

[5]D. Keim, F. Mansmann, J. Schneidewind, and H. Ziegler, “Challenges in Visual Data Analysis,” in *Proceedings of Information Visualization*, 2006, pp. 9–16.

[6]R. Mazza, *Introduction to Information Visualization*. London, 2009.

[7]H. Purchase, N. Andrienko, T. J. Jankun-Kelly, and M. Ward, “Theoretical Foundations of Information Visualization,” in *Information Visualization: Human Centered Issues and Perspectives*, Lecture Notes In Computer Science, Ed. 2008, pp. 46–64.

[8]E. H. Chi, “A Taxonomy of Visualization Techniques Using the Data State Reference Model,” in *Symp. Information Visualization (InfoVis 2000)*, 2000, pp. 69–75.

[9]S. K. Card, J. D. Mackinlay, and B. Shneiderman, *Readings in Information Visualization—Using Vision to Think*. San Francisco, {CA}, {USA}: Morgan Kaufmann, 1999.

[10] P. E. Hoffman, “Table Visualizations: A Formal Model and Its Applications,” University of of Massachusetts, 1999.

[11] D. A. Keim and H. P. Kriegel, “Visualization Techniques for Mining Large Databases: A Comparison,” *IEEE Trans. Knowl. Data Eng*, vol. 8, no. 6, pp. 923–936, 1996.

[12] D. A. Keim, “Visual Database Exploration Techniques,” in *Tutorial KDD ’97 Int’l Conf. Knowledge Discovery and Data Mining*, 1997.

[13] A. Inselberg and B. Dimsdale, “Parallel Coordinates: A Tool for Visualizing Multidimensional Geometry,” in *Proc. IEEE Visualization ’90*, 1990, pp. 361–375.

[14] A. Inselberg, “The Plane with Parallel Coordinates,” *Vis. Comput.*, vol. 1, pp. 69–91, 1985.

[15] A. Inselberg, “Multidimensional Detective,”,” in *IEEE Symp. Information Visualization (InfoVis ’97)*, 1997, pp. 100–107.

[16] Y. H. Fua, M. O. Ward, and A. Rundensteiner, “Hierarchical Parallel Coordinates for Exploration of Large Datasets,” in *Visualization ’99*, 1999.

[17] M. Goebel and L. Gruenwald, “A Survey of Data Mining and Knowledge Discovery Software Tools,” *ACM SIGKDD Explor.*, vol. 1, no. 1, pp. 20–33, 1999.

[18] M. Mramor, G. Leban, J. Demsar, and B. Zupan, “Visualization-based cancer microarray data classification analysis.,” *Bioinformatics*, vol. 23, no. 16, pp. 2147–54, Aug. 2007.

[19] H. Chernoff, “The Use of Faces to Represent Points in kDimensional Space Graphically,” *Am. Stat. Assoc.*, vol. 68, pp. 361–368, 1973.

[20] D. A. Keim, R. D. Bergeron, and R. M. Pickett, “Test Data Sets for Evaluating Data Visualization Techniques,” in *Perceptual Issues in Visualization*, G. G. G. and H. Levkowitz, Ed. Springer-Verlag, 1995, pp. 9–22.

[21] D. A. Keim, “Designing Pixel-Oriented Visualization Techniques: Theory and Applications,” *Vis. Comput. Graph.*, vol. 6, no. 1, pp. 59–78, 2000.

[22] D. A. Keim and H. P. Kriegel, “VisDB: Database Exploration Using Multidimensional Visualization,” *Comput. Graph. Appl.*, vol. 14, no. 5, pp. 40–49, 1994.

[23] C. G. Beshers and S. K. Feiner, “Visualizing n-Dimensional Virtual Worlds within n-Vision,” *Comput. Graph. (ACM).*, vol. 24, no. 2, pp. 37–38, 1990.

[24] C. G. Beshers and S. K. Feiner, “Automated Design of Data Visualizations,” *Scientific Visualization—Advances and Challenges*, pp. 88–102, 1994.

[25] J. LeBlanc, M. O. Ward, and N. Wittels, “Exploring N-Dimensional Databases,” in *IEEE Visualization ’90*, 1994, pp. 230–237.

[26] G. Robertson, S. Card, and J. Mackinlay, “Cone Trees: Animated 3D Visualizations of Hierarchical Information,” in *Int’l Conf. Human Factors in Computing*, 1991, pp. 189–194.

[27] B. Shneiderman, “Tree Visualization with Treemaps: A 2D SpaceFilling Approach,” *Trans. Graph.*, vol. 11, no. 1, pp. 92–99, 1992.

[28] G. D. Battista, P. Eades, R. Tamassia, and I. G. Tollis, “Annotated Bibliography on Graph Drawing,” *Comput. Geom. Theory Appl.*, vol. 4, no. 5, pp. 235–282, 1994.

[29] B. Becker, R. Kohavi, and D. Sommerfield, “Visualizing the Simple Bayesian Classifier,” in *SIGKDD ’97 Workshop Issues on the Integration of Data Mining and Data Visualization*, 1997.

[30] S. G. Eick and G. J. Wills, “Navigating Large Networks with Hierarchies,” in *IEEE Visualization ’93*, 1993, pp. 204–210.

[31] R. J. Hendley, N. S. Drew, A. M. Wood, and R. Beale, “Narcissus: Visualizing Information,” in *Int’l Symp. Information Visualization (InfoViz ’95)*, 1995, pp. 90–96.

[32] W. Hibbard, H. Levkowitz, J. Haswell, P. Rheingans, and F. Schroeder, *Interaction in Perceptually-Based Visualization,” Perceptual Issues in Visualization, G.G.* 1995.

[33] W. S. Cleveland, “Visualizing Data,” *Horbart Press*, New York, 1993.

[34] B. Shneiderman, “Dynamic Queries for Visual Information Seeking,”,” *IEEE Softw.*, vol. 11, no. 6, pp. 70–77, 1994.

[35] C. B. Kreitzberg, “Details on Demand: Hypertext Models for Coping with Information Overload,” in *Interfaces for Information Retrieval and Online Systems*, New York, 1991, pp. 169–176.

[36] M. C. F. De Oliveira and H. Levkowitz, “On-Line Resource on Visual Data Mining,” *http://www.cs.uml.edu/~mcristin/ vdm-resource.htm*, 2001. .

[37] M. O. Ward, “XmdvTool: Integrating Multiple Methods for Visualizing Multivariate Data,” in *IEEE Visualization ’94*, 1994, pp. 326–336.

[38] C. Ahlberg and E. Wistrand, “IVEE: An Information Visualization and Exploration Environment,” in *Int’l Symp. Information Visualization (InfoVis ’95)*, 1995, pp. 66–73.

[39] J. Bertin, “Semiology of Graphics: Diagrams, Networks, Maps,” *Univ. of Wisconsin Press*, 1983.

[40] J. D. Mackinlay, “Automating the Design of Graphical Presentations of Relational Information,” *Trans. Graph.*, vol. 5, no. 6, pp. 110–141, 1986.

[41] C. G. Beshers and S. K. Feiner, “AutoVisual: Rule-Based Design of Interactive Multivariate Visualizations,” *Comput. Graph. Appl.*, vol. 13, no. 4, pp. 41–49, 1993.

[42] W. H. Hibbard, C. R. Dyer, and B. E. Paul, “A Lattice Model for Data Display,” in *Visualization ’94*, 1994, pp. 310–317.

[43] S. F. Roth and J. Mattis, “Data Characterization for Intelligent Graphics Presentations,” in *Human Factors in Computing Systems Conf.*, 1990, pp. 193–200.

[44] S. F. Roth, P. Lucas, J. A. Senn, C. C. Gomberg, M. B. Burks, P. J. Stroffolino, J. A. Kolojejchick, and C. Dunmire, “Visage: A User Interface Environment for Exploring Information,” in *Symp. Information Visualization (InfoVis ’96)*, 1996, pp. 401–409.

[45] H. Senay and E. A. Ignatius, ““Knowledge-Based System for Visualization Design,” *Comput. Graph. Appl.*, vol. 14, no. 6, pp. 36–47, 1994.

[46] E. H. Chi and J. T. Riedl, “An Operator Interaction Framework for Visualization Systems,” in *Symp. Information Visualization (InfoVis ’98)*, 1998, pp. 63–70.

[47] E. H. Chi, “A Taxonomy of Visualization Techniques Using the Data State Reference Model,” in *Proc. Symp. Information Visualization (InfoVis 2000)*, 2000, pp. 69–75.

[48] M. Goebel y L. Gruenwald, «A Survey of Data Mining and Knowledge Discovery Software Tools», *SIGKDD Explor Newsl*, vol. 1, n.º 1, pp. 20–33, jun. 1999.

[49] A. Inselberg y B. Dimsdale, «Parallel Coordinates: A Tool for Visualizing Multi-dimensional Geometry», en *Proceedings of the 1st Conference on Visualization ’90*, Los Alamitos, CA, USA, 1990, pp. 361–378.

[50] P. Hoffman, G. Grinstein, K. Marx, I. Grosse, y E. Stanley, «DNA visual and analytic data mining», en *Visualization’97., Proceedings*, 1997, pp. 437–441.

[51] J. W. Sammon, «A Nonlinear Mapping for Data Structure Analysis», *IEEE Trans. Comput.*, vol. 18, n.º 5, pp. 401-409, 1969.

[52] G. Leban, B. Zupan, G. Vidmar, y I. Bratko, «VizRank: Data Visualization Guided by Machine Learning», *Data Min. Knowl. Discov.*, vol. 13, n.º 2, pp. 119-136, may 2006.

[53] G. A. Helt, S. Lewis, A. E. Loraine, y G. M. Rubin, «BioViews: Java-Based Tools for Genomic Data□Visualization», *Genome Res.*, vol. 8, n.º 3, pp. 291-305, mar. 1998.

[54] T. C. Sprenger, R. Brunella, y M. H. Gross, «H-BLOB: A Hierarchical Visual Clustering Method Using Implicit Surfaces», en *Proceedings of the Conference on Visualization ’00*, Los Alamitos, CA, USA, 2000, pp. 61–68.

[55] M. H. Gross, T. C. Sprenger, y J. Finger, «Visualizing information on a sphere», en *, IEEE Symposium on Information Visualization, 1997. Proceedings*, 1997, pp. 11-16.

[56] P. C. Wong, «Guest Editor’s Introduction: Visual Data Mining», *IEEE Computer Graphics and Applications*, vol. 19, n.º 5, pp. 20-21, 1999.