ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

Survey of Star Glyph-Based Visualization Technique for Multivariate Data

Izyan Izzati Kamsani, Norharyati Md Ariff, Noor Elaiza Abd Khalid

Universiti Teknologi Mara (Shah Alam), Department of Computer Science, Faculty of Computer and Mathematical Science, 40000, Shah Alam, Selangor, Malaysia.

ARTICLE INFO

Article history:
Received 13 June 2015
Accepted 5 August 2015
Available online 12 August 2015

Keywords:
Data visualization, Glyph-based, Multivariate data, Star glyph

ABSTRACT

Background: The advancement of modern and scientific technology has led to datasets growing in both size and complexity, exposing the need for more efficient and effective way of visualizing and analyzing data. Current software architecture, management and analysis approaches are unable to cope with the flood of data**. Problem:** Despites, the amount of progress in visualization methods especially multivariate data still poses a number of significant challenges in term of understand and depict with multivariate data involving more than two attribute dimensions. **Objective:** A technique using glyph-based visualization focusing in star glyph visualization is presented to visual and interpret multivariate data. **Results:** As a result, this paper investigates the general statistical of glyph-based application and details of star glyph-based applications based on interaction features. **Conclusion:** This paper reviews the existing guidelines, implementation techniques and surveying the use of star glyph-based visualization which uncommonly used for multivariate data visualization.

To Cite This Article: Izyan Izzati Kamsani, Norharyati Md Ariff, Noor Elaiza Abd Khalid., Survey of Star Glyph-Based Visualization Technique for Multivariate Data. ***Aust. J. Basic & Appl. Sci.,*** *9(26): 61-66, 2015*

INTRODUCTION

The analysis of growing datasets can be extremely difficult to perform if the data is not presented graphically as observed by Spence (Spence, 2001) especially for multivariate data. These data are hard to imagine where it is cannot be easily observed by viewing large tables of data and sometime face difficulty to visualize due to large number of dimensions. One of techniques that are commonly used to visualize multivariate data is iconographic display (Matthew O Ward, 2002). This technique is also known as glyph-based technique (Dzemyda, Kurasova, & Zilinskas, 2013). Fanea *et al* (Fanea, Carpendale, & Isenberg, 2005) pointed that iconographic technique is able to map data in various geometric and different colors for each attributes of glyph. Visualizing pattern between multiple dimensions data is one of the advantages using glyph-based. It is also an effective method which allows a quick comparison of data records and attributes. These dimensions are known as multivariate or multidimensional data. The definition of multivariate is more broadly used to describe the property of such datasets with high dimensionality. While, for attribute less than three, the term univariate and bivariate is frequently used to describe datasets that contain only one or two dimensions respectively. Visualizing multivariate data have been widely used in various field such medical (Ropinski, Oeltze, & Preim, 2011) (Kandogan, Road, & Jose, 2000), business (Marghescu, 2007) (Tekušová & Knuth, 2008), engineering (Haber, 1990) and many more (Fanea *et al*., 2005).

Glyph-based technique is divided into five main branches such Chernoff faces, stick figure, shape coding, color icon and star glyph. Star glyph-based technique is some example of multivariate glyphs. It suited for displaying multivariate and complex datasets. Additionally, star glyph have been widely used in the visualization of data and information and. The objective of this survey is to deliver a better understanding of the application of the star glyph-based technique applied which applied in multivariate data. This paper overviews the existing guideline, implementation technique and applied interaction technique. Discussion includes statistical survey on glyph-based and star glyph publication and analysis on features for star glyph technique.

In this paper, we reviewed star glyph-based techniques to drive and facilitate of visualizing multivariate data and some other interactive technique applied in order to make this technique become more interactive to user. Major gaps in the literature, reviewing the existing guidelines and implementation techniques are also been studied. The

Corresponding Author: Izyan Izzati Kamsani, Universiti Teknologi Mara (Shah Alam), Department of Computer Science, Faculty of Computer and Mathematical Science, 40000, Shah Alam, Selangor, Malaysia.
Ph: +6013-7493793 E-mail: izyanizzati@gmail.com

survey is organized as follows: Section 2 examines the studies of star glyph background and number of application areas where star glyph-based visualization have been deployed. Section 3 surveys for features, geometric channel and interactive technique that have been used for practice. Section 4 summarizes the finding that has emerged during the compilation of this survey.

Methodology:

The survey on glyph-based technique is basically refer from Google Scholar, online journal databases and from books related. These sources were chosen where it is a platform for us to search our literature, gaining plenty of information broadly and find relevant work across the world in research area. 30 papers between the periods 1990 to 2015 were reviewed. 20 papers related to glyph-based technique in various applications were found. However, there are 10 papers only which associated to star glyph technique was reviewed. There were not many papers discussed on star glyph technique since this technique is rarely applied in multivariate data. Glyph-based technique can be divided into five main branches that are Chernoff faces, stick figure, shape coding, color icon and star glyph.

Briefly explain, the most famous in iconography (Chernoff, 1973) is Chernoff face visualization. This technique work by mapping two attributes into the 2D position of a face. While the remaining attributes are mapped to its properties, such the shape of nose, mouth, eyes and the face itself. It also stated that Chernoff faces can only visualize a limited amount of data items (D. a. Keim & Kriegel, 1996). There is common issue involve for multivariate icons where the significant impact on the perceptive effectiveness regarding on semantic relation (Spence, 2001).

Stick figure is another member of glyph-based technique. This technique display two attributes and the remaining attributes are mapped to the angles and / or limb length to the stick of figure icon (D. A. Keim, 2002). If the data items are quite compact with respect to the two dimensionally display, the resulting visualization of texture patterns which differ according to the characteristic of data and determined by pre-attentive perception.

The other technique in glyph-based technique is shape coding. This technique visualizes data using small array of pixels. Pixels in the array are placed in the form of square or rectangle. Meanwhile, the arrays are arranged in a line-by-line pattern. In details, each data items are represented by one such arrays and the pixels are encoded with grayscale according to the attributes values (Beddow, 1990). In other hands, it is highly compressed visualization without any clutter or overlap.

A combination of pixel-based spiral axes and icon-based shape coding techniques perform color icon technique (Levkowitz, 1991). Pixels are replaced by array of color fields that represent the attribute values similar to shape coding. Color, shape, size, orientation, boundaries and area sub-dividers can be used to map the multivariate data.

Star glyph is one of the most widely used glyphs (Fanea *et al*., 2005) for multivariate data. It represent the attributes as equal angular axes release from the center of a circle (Chernoff, 1973) with an outer line connecting the data value points on each axis. Each data item or record is presented by a single star glyph.

Glyph-based Table Review:

There are five (5) techniques that are categorized under glyph-based technique. These techniques contain Chernoff faces, stick figure, shape coding, color icon and star glyph. Figure 1 summarized specifically five (5) techniques in the glyph-based techniques including references.

Glyphs represent the data value using visual features or attributes such as shape, size, colour and position. Normally, the number of graphical object is equal to the cardinality of the high dimensional dataset and they are arranged on the display in such way as to reveal visual pattern. Since a large number of data variations can be incorporated into properties of a single glyph, this makes it a highly suitable technique for communicating and supporting multivariate analysis. Glyph can be placed and viewed either independently from others. But in some cases, glyph can be spatially connected to convey the topological relationships between data points of the underlying data specification. It is expected that these pattern represent interesting behaviour of the data. Glyphs are effective method to visualize pattern between multiple dimensions and allow a quick comparison of data records and attributes.

As a summary, glyph-based technique can handle from small to medium datasets with a few thousand data items. But, glyph-based technique can also been applied to dataset which are multivariate data and these data attributes are treated differently as some visual attributes of the icons may attract more attention than the others. In other way, the way data attributes are mapped to icon are greatly determines the expressiveness of the resulting visualization and what can perceived. There is a difficulty regarding on glyph-based technique during mapping the data when come to complex multivariate data. Additionally, it requires training to interpret the icons since it is not a straight forward interpretation. There is also occurs an overlap records if data attributes are mapped to the icon's display positions. As illustrated from Figure 1, Chernof faces and star glyph are mostly discussed and reviewed by other researchers. This is due to unique characteristics between these two techniques. However, this paper only focused on star glyph technique.

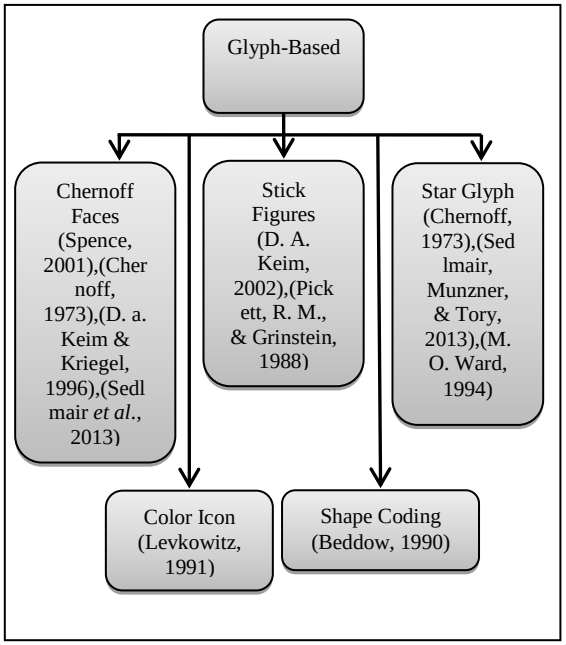

Fig. 1: Categorization of glyph-based technique.

Table 1 shows comparison of five (5) features of the glyph-based techniques. A study regarding this topic have been done on several papers (D. A. Keim, 2002),(Cristina, Oliveira, & Levkowitz, 2003),(Elmqvist & Fekete, n.d.),(Chan, 2006). We categorized each features into specific category such data, type of multi-dimensional data, visualization specification, user's interaction, finding and limitation each techniques in glyph-based technique

Table 1: An analysis of features and requirement of star glyph based technique.

Glyph-based Techniques **Basic idea:** Visualization of the data values as features of icons					
Features / Techniques	**Chernoff [12], [18], [19], [20]**	**Stick Figures [12], [18], [19], [20]**	**Shape Coding [12], [18], [19], [20]**	**Color Icon [12], [18], [19], [20]**	**Star Glyph [12], [18], [19], [20]**
Data					
Types	Multidimensional	Multidimensional	Multidimensional	Multidimensional	Multidimen
Representation	Face	Stick	Rectangular grid	Color	Star
Multivariate Data					
Numerical	Yes	Yes	Yes	Yes	Yes
Categorical	No	No	No	No	No
Visualization					
Dimensionality	Undefined	1000	5000	5000	5000
Cluttered Data	Good	Good	Good	Good	Good
Interactivity					
Brushing	Yes	Yes	No	No	Yes
Zooming	Undefined	Undefined	Undefined	Undefined	Undefined
Linking	Yes	Yes	Yes	Yes	Yes
Finding					
Outliers	Yes	Yes	Yes	Yes	Yes
Gaps	Yes	Yes	Yes	Yes	Yes
Cluster	Yes	Yes	Yes	Yes	Yes
Corelation	Yes	Yes	Yes	Yes	Yes
Pattern	Yes	Yes	Yes	Yes	Yes
Limitation					
Space	Unlimited	Unlimited	Unlimited	Unlimited	Unlimited
Scalability					
- Cases	Unlimited	Unlimited	Unlimited	Unlimited	Unlimited
- Attributes	Unlimited	Unlimited	Unlimited	Unlimited	Unlimited
- Dimensions	Unlimited	Unlimited	Unlimited	Unlimited	Unlimited
Overlap / cluttered	No	Yes	No	Yes	Yes
Labeling	None	None	None	None	None

Data feature is a feature where it is classified into representation and type. Chernoff face, stick figure, shape coding, color icon and star glyph are commonly used to visualize multidimensional or multivariate data. Chernoff face respectively visualize the data into face, stick figure into stick, shape coding into rectangular grid, color icon in color based and star glyph into star representation. Categorical data can also be visualized using glyphs, though only after conversion to numerical data form. Interaction to users is important so that users are free to play around with the data based on their requirement. Brushing enable users to resize the value of the dimension and this stage is crucial to observe the relationship between the attributes in the specific range of the scaling. Labeling is also important and it is preferable to apply into star glyph. Labeling can give details of data as supplementary information to the user by clicking the star glyph right away without making any crowded of data view.

RESULT AND DISCUSSION

The result of investigation is discussed in two main section; the general statistical of glyph-based

applications and the details of star glyph-based techniques based on interaction features.

Population of Glyph-based Applications between 1990 to 2015:

Researchers that apply glyph-based technique in visualizing multivariate data began around 1990s. But there is an increasing amount of research applying star glyph technique for multivariate data starting from 2009. This technique is rarely used for visualizing compared to Chernoff face which is quite famous long time ago. However, star glyph technique has become a popular choice of illustrating complex data that involve multivariate data and applied in various applications. Figure 2 depicted the number of publication for glyph-based from 1990 until 2015.

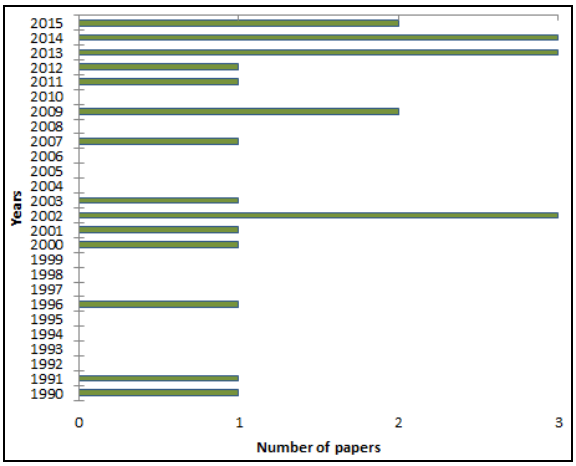

Fig. 2: Glyph-based technique publication by year.

Application of Star Glyph-based Technique based on Multivariate Data:

In this section, survey collections of applications in various fields where star glyph-based visualization has already made an impact. Generally, glyph-based technique explains according to the data's characteristics which make them fundamentally different. Thus, selecting the preferable technique would be depends on the data which are visualized and the task that need to be performed by the user over the visualization (Rzeźniczak, 2013). There are three types of data such event-based, geo-spatial and high-dimensional data. Briefly explain, event-based data is an event that occur at a given time or location. A popular and classic approach that combines the visualization of space and time is the space-time cube concept developed by Hagerstrand in 1989 (Hagerstrand, 1989). While Forlines and Wittenburg (Forlines & Wittenburg, 2010) introduced *Wakame* glyph to depict multi-dimensional sensor reading. Geo-spatial involved analyzing data with a geographical or geo-spatial aspect. MacEachren (MacEachren, Brewer, & Pickle, 1998) present a novel approach to visualize reliability in mortality maps using a bi-variate mapping. Lastly, high dimensional data is a desirable tool due to their applicability to a diverse range of fields. One of the most popular used approaches is star glyph which was first introduced by Siegal *et al*. Star glyph so effective when applied in multivariate data where its ability to compare similarities between multivariate entities based on its geometric properties. It has used to depict a variety of datasets which include myocardial infarction data (Rzeźniczak, 2013), coal data (M.O. Ward & Rundensteiner, 2004) and animal datasets (Lee, Butavicius, & Reilly, 2003) as mention in Figure 2.

Based on Figure 3 it shows that mostly glyph have been applied in multivariate data. However, research on star glyph specifically is still new for this area which uncommonly used for multivariate data visualization compared to glyph-based technique generally. Figure 4 proves the statement according to the research area between star glyph and glyph-based technique. Thus, it is a constraint in order to do a research in star glyph-based technique where the sources are limited and we can summarize that this technique is not widely applied compared to others.

Star Glyph-based Technique based on Interaction Features:

We have implemented our visualization technique using Microsoft Visual Studio (C#) to generate the star glyph visualization. We used benchmark data which is car datasets (Kandogan & Jose, 2001) that contain 400 automobiles from the 1970's until 1980's. The attributes for this dataset includes fuel efficiency (in miles per gallon-MPG), acceleration (time from 0-60MPH in second), engine displacement (in cubic inches), weight and horsepower. We had improved our star glyph visualization such adding legend for future reference and differentiate each attributes applying different colors as shown in Figure 5.

However, there is still need some enhancement for better visualization and presentation. To make

firm basic on star glyph, we had studied on nine (9) papers between 2002 until 2015 which focused on star glyph research area only. We categorized into three (3) categories such features, geometric channels (Chen & Floridi, 2013) and user interaction. We believed that applying all these categories will overcome any weaknesses and enhance star glyph features in future.

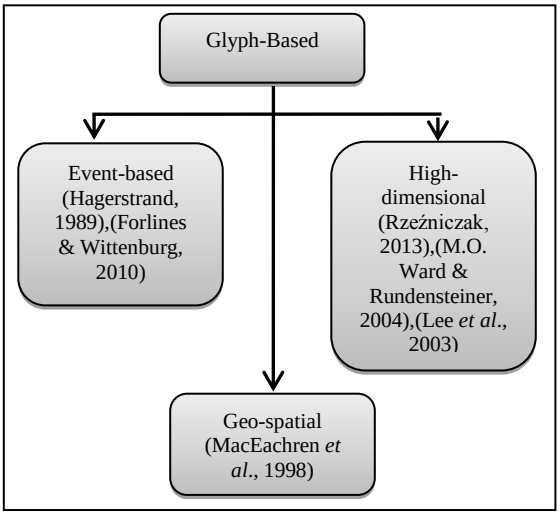

Fig. 3: Different data characteristic used in star glyph applications.

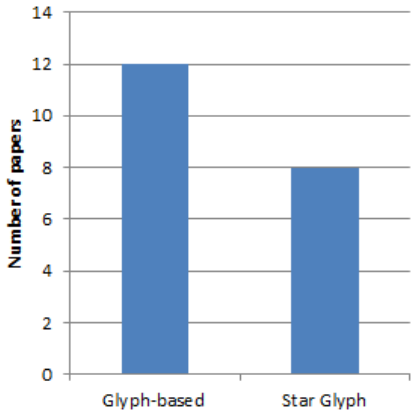

Fig. 4: Comparison between glyph-based and star glyph according to visualization for multivariate data.

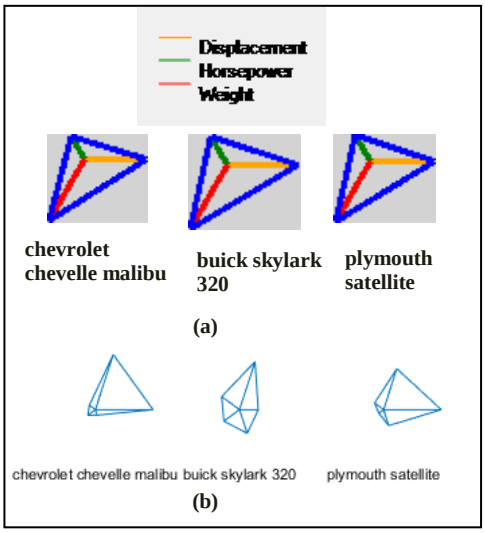

Fig. 5: Examples star glyph that are generated using our visualization technique (a) and image generated using MATLAB (b).

A star glyph is a multivariate graphing technique in which each attributes represent a ray or 'spoke'. Each ray performs an equal angular distance from an origin which extends out via a connecting line. The length of each line is proportional to the magnitude of the variable compared to the maximum value of all the variables. Star glyph visualization represent a set of datasets each has one or more set of features. Each star (records) represents a separate data record and each branch of the star represent a different attributes of datasets. Similar records share a particular features will share the same corresponding branch of the star. Additionally, star glyph based technique allows manipulation of attributes dimensions (Chan, 2006). An interactive view allows users to view data in different angles, axis and attributes manipulation (Nguyen, Nelmes, Huang, Simoff, & Catchpoole, 2014). Applying colours are also allow instant recognition of similarities or differences of the large data items and expressed attributes relationship (D. A. Keim, 2002).

For a detailed overview of research on data glyph, we refer the interested paper for summary. We have been studied the perception of glyph in the context of similarity tasks between year 2000 until 2015 publication. In this section, we were inspired by two main bodies of work: a) related studies investigating the performance of star glyph in information visualization for multivariate data and b) number of application which implement this technique.

Based on Table 3, features give dimensionality, representation, data similarity, attributes, mapping, and labelling, numerical, categorical, overlap and closed versus open contour. Below are the analyses during the study on multivariate data in star glyph visualization:

- **Dimensionality** –star glyph was selected as the one with the highest dimensionality where it is still not limitless. Authors Dias M.M *et al*. (Dias, Yamaguchi, Rabelo, & Franco, 2012) had mention in his paper that more than 80 attributes may contribute to blur visualization.
- **Attribute's domain type** –star glyph is the suitable technique for representing numerical value based on the data type perspective. However it is not suitable for nominal qualitative data but it can be done by assign them into '0' and '1' numerical value for example to ease during plotting time. Hence, it is easy to notice that the visual representation is built of symptoms with '0' and '1' which contribute to the picture indirectly.
- **Scalability** –star glyph can present only limited number of object simultaneously because of the space occupied by the graphical entity. However, according to the task type, the limited scalability of star glyph is not an obstacle here.
- **Labelling** – labelling gives user the ability to display any important information or either hide the label if necessary. Each spokes may have its own label or may have more details to display in the visualization. For instance, user may be interested in the exact of the spoke represents. By clicking in that particular star glyph, detailed information about that category will be displayed. Using this feature, the user can view the specific details for focal point to compare and analysis.

Interaction visualization looks at the ability to navigate and interrogate datasets through interaction to improve understanding. However, like in many multivariate approaches, the interaction of star glyph-based visualization is an important aspect for the visual exploration of complex datasets. There are also challenges for viewing data due to screen space, resolution and displaying many glyphs which can cause data overlaps. Besides, it is difficult to interpret the relationship between these data in order to make fast decision and summarization.

Conclusion:

We do not state this opinion are complete or describe a theory, but still the observation we have made during our literature reviewed. We believe it can be a helpful survey for the interested reader. Literature review based on star glyph is limited since 2012. In future will be more experiment in visualizing multivariate and complex data using star glyph visualization.

ACKNOWLEDGEMENT

Izyan Izzati Kamsani is a PhD student in Computer Science. Norharyati Md Ariff is a master student in Computer Science and Dr. Noor Elaiza A. Khalid is a senior lecturer in Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia. Author' email addresses: izyanizzati@gmail.com

REFERENCES

Beddow, J., 1990. Shape coding of multidimensional data on a microcomputer display. *Proceedings of the First IEEE Conference on Visualization: Visualization `90*. http://doi.org/10.1109/VISUAL.1990.146387

Chan, W.W., 2006. *A Survey on Multivariate Data Visualization*.

Chen, M., L. Floridi, 2013. An analysis of information visualization. *Syntheses*, *190*(Synthese), 3421–3438.

Chernoff, H. (1973). The Use of Faces to Represent Points in K-Dimensional Space Graphically. *Journal of the American Statistical Association*, *68*(342): 361-368.

Cristina, M., Oliveira, F. De, H. Levkowitz, 2003. From Visual Data Exploration to Visual Data Mining : A Survey. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, *9*(3): 378-394.

Dias, M.M., J.K. Yamaguchi, E. Rabelo, C. Franco, 2012. Visualization Techniques : Which is the Most Appropriate in the Process of Knowledge Discovery in Data Base ? *Advances in Data Mining Knowledge Discovery and Applications*, 155-180. http://doi.org/10.5772/3349.

Dzemyda, G., O. Kurasova, J. Zilinskas, 2013. *Multidimensional Data Visualization*. (G. Dzemyda, O. Kurasova, & J. Zilinskas, Eds.). Springer Optimization and Its Applications.

Elmqvist, N., J.D. Fekete, (n.d.)., Hierarchical aggregation for information visualization: overview, techniques, and design guidelines. *IEEE Transactions on Visualization and Computer Graphics*, *16*(3): 439-54. http://doi.org/10.1109/TVCG.2009.84.

Fanea, E., S. Carpendale, T. Isenberg, 2005. An interactive 3D integration of Parallel Coordinates and Star Glyphs. *Proceedings - IEEE Symposium on Information Visualization, INFO VIS*, 149-156. http://doi.org/10.1109/INFVIS.2005.1532141.

Forlines, C., K. Wittenburg, 2010. Wakame: Sense making of multi-dimensional spatial-temporal data. *Proceedings of the Workshop on Advanced Visual Interfaces AVI*, 33-40. http://doi.org/10.1145/1842993.1843000.

Haber, R.B., 1990. Visualization techniques for engineering mechanics. *Computing Systems in Engineering*, *1*(1): 37-50. http://doi.org/10.1016/0956-0521(90)90046-N.

Hagerstrand, 1989. What About People in Regional Science.Pdf. Ninth European Congress of The Regional Science Association.

Kandogan, E., S. Jose, 2001. Visualizing Multi-dimensional Clusters , Trends , and Outliers using Star Coordinates. *In Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 107-116.

Kandogan, E., H. Road, S. Jose, 2000. Star Coordinates : A Multi-dimensional Visualization Technique with Uniform Treatment of Dimensions.

Keim, D.A., 2002. Information Visualization and Visual Data Mining. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, *8*(1): 1-8.

Keim, D.A., H.P. Kriegel, 1996. Visualization techniques for mining large databases: a comparison. *IEEE Transactions on Knowledge and Data Engineering*, *8*(6): 923-938. http://doi.org/10.1109/69.553159.

Lee, M.D., M.A. Butavicius, R.E. Reilly, 2003. Visualizations of binary data: A comparative evaluation. *International Journal of Human Computer Studies*, *59*(5): 569-602. http://doi.org/10.1016/S1071-5819(03)00082-X.

Levkowitz, H., 1991. Color icons-merging color and texture perception for integrated visualization of multiple parameters. *Proceeding Visualization '*91: 164-170. http://doi.org/10.1109/VISUAL.1991.175795.

MacEachren, A.M., C.A. Brewer, L.W. Pickle, 1998. Visualizing georeferenced data: representing reliability of health statistics. *Environment and Planning A*, *30*(9): 1547-1561. http://doi.org/10.1068/a301547.

Marghescu, D., 2007. Multidimensional Data Visualization Techniques for Financial Performance Data: A Review. *TUCS Technical Report*, *810*(July). Retrieved from http://users.abo.fi/dmarghes/publications_files/TR810.pdf.

Nguyen, Q. V., G. Nelmes, M.L. Huang, S. Simoff, D. Catchpoole, 2014. Interactive Visualization for Patient-to-Patient Comparison. *Genomics & Informatics*, *12*(1): 21-34. http://doi.org/10.5808/GI.2014.12.1.21.

Pickett, R.M., G.G. Grinstein, 1988. Iconographic Displays For Visualizing Multidimensional Data. *In Systems, Man, and Cybernetics, 1988. Proceedings of the 1988 IEEE International Conference*, 1: 514-519.

Ropinski, T., S. Oeltze, B. Preim, 2011. Survey of glyph-based visualization techniques for spatial multivariate medical data. *Computers and Graphics (Pergamon)*, *35*(2): 392-401. http://doi.org/10.1016/j.cag.2011.01.011.

Rzeźniczak, T., 2013. Evaluation of multidimensional visualization techniques for medical patterns representation. *Journal of Theoretical and Applied Computer Science*, *7*(4): 70-85.

Sedlmair, M., T. Munzner, M. Tory, 2013. Empirical guidance on scatterplot and dimension reduction technique choices. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12): 2634-43. http://doi.org/10.1109/TVCG.2013.153.

Spence, R., 2001. *Information Visualization*. ACM Press Book.

Tekušová, T., M. Knuth, 2008. Data quality visualization for multivariate hierarchic data. *InfoVis Demo. Http://www. …*, 1–2. Retrieved from http://bib.dbvis.de/uploadedFiles/infovis08-paper.pdf.

Ward, M.O., 1994. XmdvTool: integrating multiple methods for visualizing multivariate data. *Proceedings Visualization '94*, *7*. http://doi.org/10.1109/VISUAL.1994.346302.

Ward, M.O., 2002. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1(3-4): 194-210. http://doi.org/10.1057/palgrave.ivs.9500025.

Ward, M.O., E.A. Rundensteiner, 2004. Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering. *IEEE Symposium on Information Visualization*, 89-96. http://doi.org/10.1109/INFVIS.2004.15.