A Multidimensional Data Visualization Method Based On Parallel Coordinates and Enhanced Ring

Yi Chen, Xianchen Cheng, Hongqian Chen
College of Computer and Information Engineering,
Beijing Technology and Business University, Beijing, P.R.China
esteechenn@gmail.com

***Abstract*—When Analyzing multidimensional, quantitative data set, statistics of value distribution in each dimension, and comparison of two or more dimensions are common tasks in many domains, such as students' scores, college admissions and stocks. In this paper, we present a multidimensional data visualization method based on parallel coordinates and enhanced ring (PCER). The interactive scheme allows users to do a preliminary statistics on data to get an overview of data's distribution. Users can change the display proportion of one single dimension or all dimensions according to the preliminary statistical results and relative statistical requirement. The enhanced ring will show the distribution of the clutter area in parallel coordinates according to statistical results to help users getting more detailed information about data distribution and relationships among data. Data are filtered by the interaction between parallel coordinates and enhanced ring according to user's statistical requirement to reduce the size of dataset and optimize the visualization effect. The experiment on students' scores shows that PCER is efficient in presenting and quantizing the distribution of students' scores.**

Keywords-parallel coordinates; enhanced ring chart; multidimensional data visualization

I. INTRODUCTION

Information visualization can help people to understand multidimensional data better. Many techniques have been proposed for information visualization. But most of them only focus on one of the data characteristics, such as the hierarchical relationship, but can't quantize the relations. Actually, how to show more details about the distribution and relations of data during visualization process is very important. The trend of a dataset can be observed easily by many visualization techniques, but further more cognize calls for the acknowledgement of quantity. The work of how to quantize a trend or relation is facing the following challenges. 1) Functionally, while dealing with mass data, it's hard to quantize a relation because of the limitation from the method itself and the size of screen. 2) The visualization of mass data will lead to clutter.

We present a visualization method based on parallel coordinates and enhanced ring (PCER), which is an effective method to visualize mass data both qualitatively and quantificationally. Users can get better perception of the represented data by PCER for it both describes the trend of a dataset and quantizes the trend. After testing PCER on the students' scores dataset, it comes to the conclusion that PCER is efficient in both revealing the trend or relation of a dataset and quantizing it.

II. RELATED WORKS

A. Parallel Coordinates

Parallel coordinates uses the 2-D graphic to represent the high dimensional data. Inserberg proposed the multi-dimensional use of parallel coordinates in 1985 and it is used as a computing tool for geometry. He then put forward the use of parallel coordinates as a visualization tool for multidimensional geometry [1].Parallel coordinates has been improved constantly during recent years. A methodology for efficiently generating density plots of uncertain multivariate data sets that draws viewers to identify values of high certainty while not calling attention to uncertain values is proposed [2]. It demonstrates how to augment parallel coordinate plots to incorporate statistically modeled uncertainty. To discern relative frequencies of subsets of points for larger datasets easier, parallel dot plots is used and a mild perspective view and narrow dot towers is arranged to facilitate length-based magnitude comparisons and to minimize occlusion[3]. Pargnostics which is based on the screen space metrics is one of the improvements, too [4]. The metrics include the number of line crossings, crossing angles, convergence, over-plotting, etc. It is used to quantify different visual structures. This method provides a way of optimizing the display automatically, according to user's preferences. In this method, user can pick from a ranked display to construct a visualization view. Pargnostics is limited by whether the metrics set is exhaustive. The metrics are decided by the screen and display size. Parallel coordinates with interactive scheme allows users to interact directly with the parallel coordinate plots and provides great flexibility in exploring and revealing underlying patterns [5]. Users can also dynamically adjust the clustering parameters to reach an optimum.

The method mentioned above improved the visualization effects. But users still can't quantize data's trend or relations of a dataset.

B. Radial Visualizations

Radial, space-filling visualizations can be useful for depicting information hierarchies. It displays data in a circular or elliptical pattern. FanLens is a toolkit that enhances the conventional radial, space-filling visualization (e.g.Sunburst) mainly with incremental Layout and fisheye distortion based selecting [6]. It can address problems such as thin slices, flexibility and large sized data support. Fan charts are a popular method for displaying family trees in a compact way. Interactive Fan Charts can transform what was a static display medium into an interactive tool for browsing and editing genealogical data [7]. It's a radial graph in which nodes can be selectively expanded or collapsed so that a greater proportion of the available space is dynamically allocated to nodes of current interest. DocuBurst is well known as a technique for radial, space-filling visualization [8]. It is a layout of hyponymy (the IS-A relation), over laid with occurrence counts of words in a document of interest to provide visual summaries at varying levels of granularity.

Although researches on radial, space-filling visualization techniques are mainly focused on visualizing hierarchical data, the circular or elliptical pattern still determines that it can be used to describe the proportion of a dataset.

C. Combined Visualization Technique

TripVista is an interactive visual analytics system consisted of ring-style slider, improved themeriver and parallel coordinates [9]. Parallel coordinates is used to visualize the multi-dimensional aspects of the traffic trajectory data. Parallel scatterplots is proposed to analyze the relationships among multidimenshional data [10]. This method improved cue and interactive performance on rendering tens of millions data by using unified shader GPU. Scattering points in parallel coordinates convert two selected neighboring coordinate axes into a scatterplot directly [11]. This scheme is efficient in performing visualizing analysis tasks. PRISMA is a multidimensional information visualization tool using multiple coordinated view [12]. It can provide usability, portability and extensibility.

The combined visualization technique mentioned above shows more details about dataset, but still can't quantize the details. To address this problem, PCER is proposed.

III. VISUALIZATION METHOD BASED ON PCER

PCER is based on parallel coordinates and enhanced ring. It focuses on quantizing the distribution or trend of a dataset.

A. Parallel Coordinates in PCER

1) Transformation from multidimensional data to parallel coordinates.

The average distance of two adjacent axes is calculated by the length of x axis l_x and the amount of variables N as l_x/N. According to the average distance, place N copies of y axis vertically from the beginning of x axis. That constitutes the axes of a N dimensional Euclidean space R^N, labeled as $x_1, x_2, \cdots, x_N$. Record $M\left(c_1, c_2, \cdots, c_N\right)$ is mapped to the coordinate system as a ployline. Its vertex (i, c_i) is on x_i. A 1to1 correspondence is established then between polylines whose vertices are on $x_1, x_2, \cdots, x_N$ and records in R^N.

While dealing with mass data, values intensive focus on a small range leads to clutter. User can't identify the trend of data in detail for that reason. To address this issue, we prefer to change the display scale of parallel coordinates.

2) Change the display scale of parallel coordinates.

Let S be a set of M N-dimensional objects, $S = \{x_m = (x_{m,1}, x_{m,2}, \cdots, x_{m,N})^T \mid 1 \le m \le M\}$ where m is the number of data items, while x_m is object m and N is the dimension of data. For each axis corresponding with dimension j, there is a maximal value $x_{j\max}$ and a minimum value $x_{j\min}$. Let l_j be the length of axis j, C_j be the display scale of dimension j, x_A be the limit value corresponding with C_j, P be the ratio of x_A on axis j and B_n be the limit value of axis n corresponding with C_j while changing all dimensions' display scale according to x_A and C_j, where $0 \le C_j \le 1$ and $x_{j\max} \le x_A \le x_{j\min}$. P can be calculated by $P = \dfrac{x_A}{x_{j\max} - x_{j\min}}$, while $x_{Bn} = P \cdot (x_{n\max} - x_{n\min})$. The display scale changed value $x_{i,j}'$ of $x_{i,j}$ on dimension j is given by formula (1) and redrew as Figure 2.

$$\begin{cases} x_{i,j}' = C_j \cdot l_j \cdot \dfrac{x_{i,j}}{(x_A - x_{j\min})}, & x_{j\min} \le x_{i,j} \le x_A \\ x_{i,j}' = (1 - C_j) \cdot l_j \cdot \dfrac{x_{i,j}}{(x_{j\max} - x_A)}, & x_A < x_{i,j} \le x_{j\max} \end{cases} \qquad (1)$$

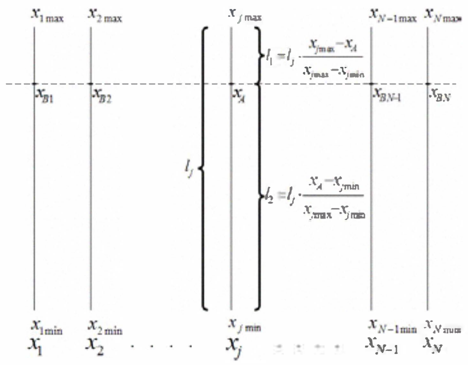

Figure 1. The original dimension of parallel coordinates where $C_j = 1$.

In Figure 1, objects whose values are above x_A are distributed in l_1 and others in l_2. One unit's length in l_1 and l_2 are the same, which is $l_j / (x_{j\max} - x_{jmim})$. The length of l_1' and l_2' in figure 2 covers the same data segment corresponding with l_1 and l_2. But one unit in l_1' changed into $l_j \cdot (1-C_j) / (x_{j\max} - x_A)$ while that of l_2' is $l_j \cdot C_j / (x_A - x_{j\min})$

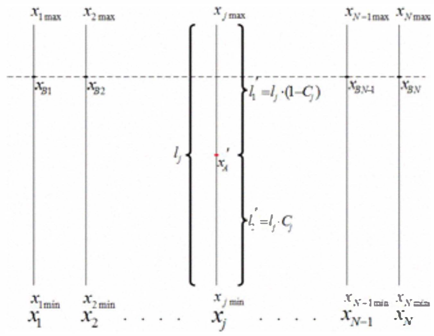

Figure 2. Display scale of dimension j changed with C_j.

Besides changing the display scale of a single dimension, PCER also can change all dimensions' display scale according to x_A and C_j mentioned above. This change will not affect the data trend. $x_{i,n}$ display scale changed value $x_{i,n}'$ on dimension n is given by

$$\begin{cases} x_{i,n}' = C_j \cdot l_n \cdot \dfrac{x_{i,n}}{(x_{Bn} - x_{n\min})}, & x_{n\min} \le x_{i,n} \le x_{Bn} \\ x_{i,n}' = (1-C_j) \cdot l_n \cdot \dfrac{x_{i,n}}{(x_{n\max} - x_{Bn})}, & x_{Bn} < x_{i,n} \le x_{n\max} \end{cases} \quad (2)$$

and redrew as Figure 3.

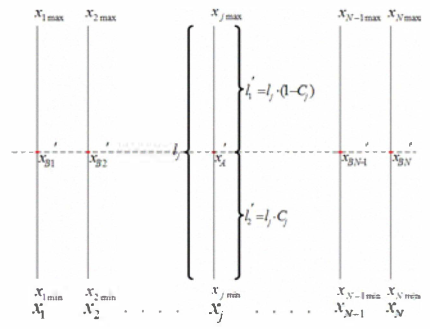

Figure 3. All dimensions' display scale changed according to x_A and C_j.

By changing display scale, users can get more detailed information from the optimized visualization result.

B. Ring Chart and Enhanced Ring Chart in PCER

The dataset of a ring chart is determined by conditions. When users want to get the distribution of students whose physics score is above 75, the condition is 'physics score above 75' and the ring's dataset is then determined. There are several rings in a ring chart. Each single ring is determined by the property of dataset. No relationship is established between different rings. According to the condition mentioned above, if users want to get the students distribution information about scores of Maths, Chinese and English, then the three courses are the properties of the students' scores dataset, and the ring chart is constituted of ring Maths, ring Chinese and ring English. The partition of each single ring is determined by students' amount in every range of the property. The amount and corresponding ratio are what users want to get from a ring chart.

Users need to provide properties and the partition method $S(S = 1,2,3\cdots)$ of each property, which indicates a single ring will be divided into how many parts, to the system. The system will calculate ring i's outside radius r_i by counting the properties' amount N. $r_i = i * \frac{r_{\max}}{N}$, where $r_{\max}$ is the biggest ring's radius the system allowed. The inside radius of the ring is r_{i-1}. The specific part's central angle A_j of a single ring, where $j \in S$,is given by $A_j = 360 * \frac{s_j}{s_{total}}$. s_j indicates part j's record amount, while s_{total} is the record amount of the ring's dataset. A specific part of a ring is then drawn by r_i, r_{i-1} and A_j, and all specific parts constitutes of a ring.

More conditions are added to the enhanced ring chart and the relationship between two adjacent rings is then established to do the data screening. The datasets of different ring in enhanced ring chart are different, and the inner ring's dataset has less data than that of the outer ring. The relationship between two adjacent rings is as follows. $S_i \xrightarrow{P_{i+1}} S_{i+1}$. P_{i+1} indicates condition $i+1$, while S_i indicates the dataset in case of condition i. Ring i is the outer ring while ring $i+1$ is the inner one.

Take students' scores as an example. The ring chart is drawn as Figure 4 by settings according to Table 1, and the enhanced ring chart is drawn as Figure 5 by settings according to Table 2. The 2nd ring shows the student distribution on English in both ring chart and enhanced ring chart, but the dataset are different. In the ring chart, the dataset of its three single ring are the same, which is the students whose physics score is above 79. The enhanced ring chart's three single ring's datasets are differ from each other, while the dataset of 2nd ring is students whose physics score is above 79, Chinese score is above 83 and English score is above 89. The score's ranges of the 2nd ring are different between ring chart and the enhanced one, which ranges from 80-100 in ring chart while that of the enhanced one is 90-100. In addition, the enhanced ring chart's 2nd ring's dataset is smaller than that of the ring chart for dataset of enhanced ring chart is filtered based on the ring chart's dataset.

TABLE I. RING CHART'S CONDITION

course	condition	value	role	partition
physics	>	79	condition	Null
Chinese	Null	Null	property	4
English	Null	Null	property	4
Maths	Null	Null	property	4

TABLE II. ENHANCING RING CHART'S CONDITION

course	condition	value	role	partition
physics	>	79	condition	Null
Chinese	>	83	property	4
English	>	89	property	4
Maths	>	79	property	4

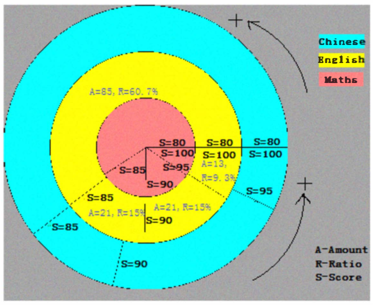

Figure 4. Ring chart in case of settings in table 1.

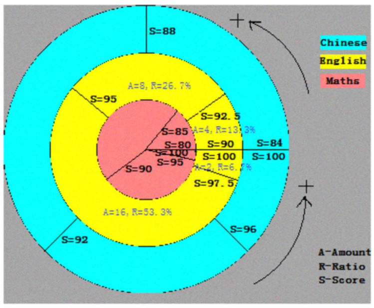

Figure 5. Enhanced ring chart in case of settings in table 2

C. Interact in PCER

PCER provides preliminary statistics to help users to discover the relations among data. Users can select the dimension and data range where the clutter occurs to overview data's proportion, and set the display scale according to their preferences to get a more detailed parallel coordinates visualization result.

The specific part in the ring will present more details by clicking the specific part. Let $d_{currently}$ be the distance from mouse-click location (x_{mouse}, y_{mouse}) to circle center's location (x_{center}, y_{center}). The central angle $A_{currently}$ is given by

$$A_{currently} = \arcsin \frac{d_{currently}}{|y_{mouse} - y_{center}|}. \quad (5)$$

Then compare $d_{currently}$ with all rings' radius to confirm which ring the selected part belongs to. Assisted with $A_{currently}$, the system can identify the specific ring part's dataset. According to the division, the specific part will show more details about the proportion.

While a dimension is selected in parallel coordinates, this dimension will be added to the ring chart. If the ring chart's condition has already been chosen, it'll be added as a property, or it'll be added as the condition. Parallel coordinates will redraw according to the selected part's dataset of a ring.

IV. CASE STUDY

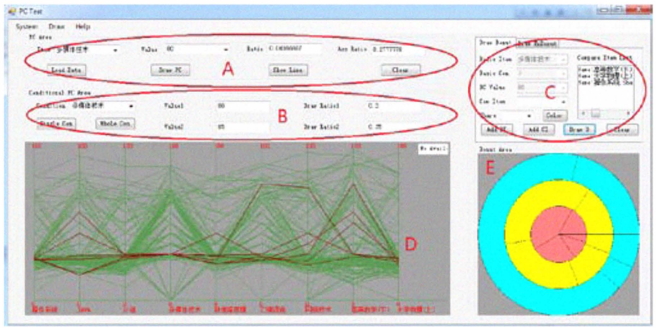

Figure 6. Students' score dataset visualization result with PCER

In this section, we demonstrate the use of PCER with student score dataset. There're five areas in PCER as Figure 6. Area A provides preliminary statistics while the display scale conditions can be set in area B. Users set the condition, property and proportion for ring chart in area C. Area D and E are used to do the visualization.

The student score dataset consists of 144 records and nine attributes. Each attribute is a course and each value is one student's scores. Firstly, the dataset is visualized by original parallel coordinates as Figure 7. Users can just get an overview on the distribution of data. The 4th dimension's data are concentrated and leads to clutter as shown in area A of Figure 7. Users can hardly get any further information or detailed information from the parallel coordinates.

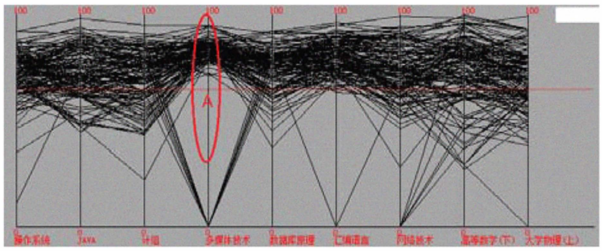

Figure 7. Students' score visualization with parallel coordinates.

Users can use area A in Figure 6 to do a preliminary statistics about data on the 4th dimension according to the overview from Figure 7. The 4th dimension is then selected as the reference for the enhancing donut. The parallel coordinates display scale can be changed by setting value and ratio in Figure 6 area B. The clutter in Figure 7 area A is caused by scores concentrate between 80 and 85. Then we can set area B in Figure 6 as following: $value_1 = 80$, $ratio_1 = 0.3$, $value_2 = 85$, $ratio_2 = 0.35$. The display scale of scores on 4th dimension is than changed as: 30% length of 4th dimension for scores under 80 as shown in Figure 8 area C, 35% length of 4th dimension for scores between 80-85 as shown in Figure 8 area B and the others for scores above 85 as shown in Figure 8 area A. Figure 9 shows all dimensions changing result according to the settings in Figure 6.

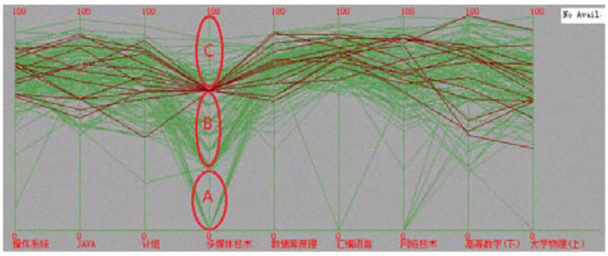

Figure 8. Single dimension display scale changed parallel coordinates.

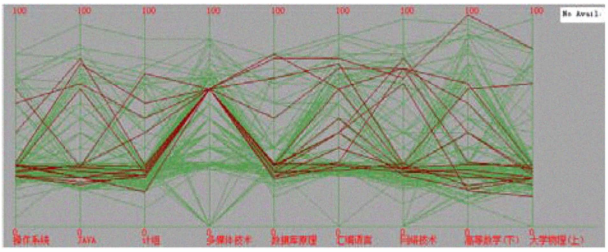

Figure 9. All dimensions display scale changed parallel coordinates.

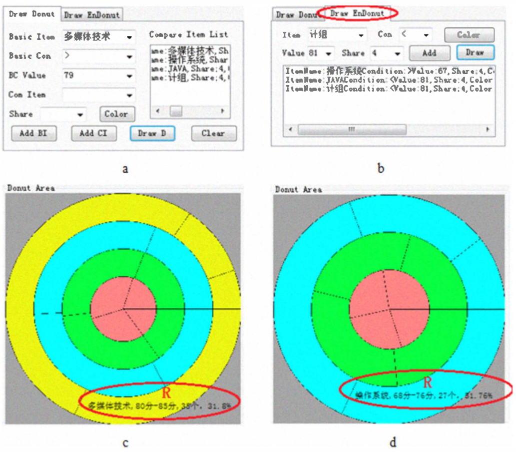

Figure 10. Settings and results of ring and enhanced ring.

Users can get more detailed information about the 4th dimension's data distribution through Figure 8. According to Figure 9, users can observe the detail information about other dimensions' data distribution in case of setting on 4th dimension. Users can quantize the distribution by using enhanced ring. Users can set according to Figure 10(a) to do a detailed statistics on 1st, 2nd and 3rd dimension when 4th dimension's score is beyond 80. The result is shown in Figure 10(c). More detailed information will be shown as area R in Figure 10(c) when the mouse clicks the specific area. Figure 10(d) shows the enhanced ring visualization result in case of settings like Figure 10(b). Figure 10(c)'s outermost ring shows the distribution of 1st dimension in case of 4th dimension's score is beyond 80, while Figure 10(d)'s dataset is that of 10(c) with additional condition, which is 1st dimension's score is between 68 and 76.

Users can get a qualitative and quantitive analysis on students' score by PCER to help the teachers to get more detailed information about the students and improve the teaching.

V. CONCLUSION AND FUTURE WORK

The visualization approach based on PCER presented in this paper shows a qualitative and quantitive visualization result to the users. The visualization technique ring which is used to visualize the hierarchical data is now used to visualize data which is not hierarchical. PCER can be further used to visualize multidimensional data that contains potential relation, such as students' score dataset, standard data in pesticide residue, and sales information. In addition, users' interaction and priori knowledge is also important to PCER.

In the future, we plan to make this technique covered with more visualization techniques such as bar chart and glyphs to improve the interaction between parallel coordinates and enhanced ring. We will also focus on simplifying and optimizing the user's interaction process to improve PCER.

REFERENCES

[1] Alfred Inselberg, B. Dimsdale: Parallel Coordinates: A Tool for Visualizing Multi-dimensional Geometry. IEEE Visualization 1990: 361-378

[2] David Feng, Lester Kwock, Yueh Lee and Russell M.TaylorII: Matching Visual Saliency to Confidencein Plots of Uncertain Data. IEEE Transactions on Visualization and Computer Graphics, vol. 16, No. 6, November/December 2010: 980-989

[3] Tuan Nhon Dang, Leland Wilkinson and Anushka Anand: Stacking Graphic Elements to Avoid Over-Plotting. IEEE Transactions on Visualization and Computer Graphics, vol. 16, No. 6, November/December 2010: 1044-1052

[4] Aritra Dasgupta and Robert Kosara: Pargnostics: Screen-Space Metrics for Parallel Coordinates. IEEE Transactions on Visualization and Computer Graphics, vol. 16, No. 6, November/December 2010: 1017-1026

[5] Peihong Guo, He Xiao, Zuchao Wang and Xiaoru Yuan: Interactive Local Clustering Operations for High Dimensional Data in Parallel Coordinates. Pacific Visualization Symposium (PacificVis), 2010 IEEE: 97-104

[6] Xinghua Lou, Shixia Liu and Tianshu Wang: FanLens: A Visual Toolkit for Dynamically Exploring the Distribution of Hierarchical Attributes. Visualization Symposium, 2008. PacificVIS '08. IEEE Pacific: 151-158

[7] Geoffrey M. Draper and Richard F. Riesenfeld: Interactive Fan Charts: A Space-saving Technique for Genealogical Graph Exploration. Proc. Eighth Ann. Workshop Technology for Family History and Genealogical Research, 2008.

[8] Christopher Collins, Sheelagh Carpendale andGerald Penn: DocuBurst: Visualizing Document Content using Language Structure. Eurographics/IEEE-VGTC Symposiumon Visualization 2009. vol.28, No.3, 2009

[9] Hanqi Guo, Zuchao Wang, Bowen Yu, Huijing Zhao, Xiaoru Yuan: TripVista: Triple Perspective Visual Trajectory Analytics and Its Application on Microscopic Traffic Data at a Road Intersection. Pacific Visualization Symposium (PacificVis), 2011 IEEE: 163 - 170

[10] Ke Yang, Qiong Luo and Jiaoying Shi: Parallel Scatterplots: Visual Analysis with GPU. Journal Of Computer-Added Design & Computer Graphics. Vol.20, No.9, Sep., 2008: 1219-1228

[11] Xiaoru Yuan, Peihong Guo, He Xiao, Hong Zhou and Huamin Qu. Scattering Points in Parallel Coordinates. Visualization and Computer Graphics, November/December 2009, vol. 15, No. 6: 1001-1008

[12] Godinho, Meiguins., Goncalves Meiguins., Casseb do Carmo, de Brito Garcia, Hernandez Almeida, Lourenco, Belem: PRISMA - A Multidimensional Information Visualization Tool Using Multiple Coordinated Views. Information Visualization, 2007. IV '07. 11th International Conference: 23 – 32.