Evaluating Multidimensional Visualization Techniques in Data Mining Tasks

Dorina Marghescu

TURKU CENTRE *for* COMPUTER SCIENCE

TUCS Dissertations
No 107, September 2008

Evaluating Multidimensional Visualization Techniques in Data Mining Tasks

Dorina Marghescu

To be presented, with the permission of the Faculty of Technology at Åbo Akademi University, for public criticism in Auditorium Gamma, ICT-house, on November 7th, 2008, at 12:00.

Department of Information Technologies
Åbo Akademi University
Turku Centre for Computer Science
Joukahaisenkatu 3-5 B, FIN-20520 Turku, Finland

2008

Supervised by

Professor Barbro Back
Department of Information Technologies
Åbo Akademi University
Turku, Finland

Reviewed by

Professor David L. Olson
Department of Management
University of Nebraska
Lincoln, Nebraska, USA

Professor Waclaw Gregory Wojtkowski
Department of Information Technology and Supply Chain Management
Boise State University
Boise, Idaho, USA

Opponent

Professor Waclaw Gregory Wojtkowski
Department of Information Technology and Supply Chain Management
Boise State University
Boise, Idaho, USA

ISBN 978-952-12-2152-1
ISSN 1239-1883
Painosalama Oy – Turku, Finland, 2008

Abstract

Visual data mining (VDM) tools employ information visualization techniques in order to represent large amounts of high-dimensional data graphically and to involve the user in exploring data at different levels of detail. The users are looking for outliers, patterns and models – in the form of clusters, classes, trends, and relationships – in different categories of data, i.e., financial, business information, etc.

The focus of this thesis is the evaluation of multidimensional visualization techniques, especially from the business user's perspective. We address three research problems. The first problem is the evaluation of projection-based visualizations with respect to their effectiveness in preserving the original distances between data points and the clustering structure of the data. In this respect, we propose the use of existing clustering validity measures. We illustrate their usefulness in evaluating five visualization techniques: Principal Components Analysis (PCA), Sammon's Mapping, Self-Organizing Map (SOM), Radial Coordinate Visualization and Star Coordinates. The second problem is concerned with evaluating different visualization techniques as to their effectiveness in visual data mining of business data. For this purpose, we propose an inquiry evaluation technique and conduct the evaluation of nine visualization techniques. The visualizations under evaluation are Multiple Line Graphs, Permutation Matrix, Survey Plot, Scatter Plot Matrix, Parallel Coordinates, Treemap, PCA, Sammon's Mapping and the SOM. The third problem is the evaluation of quality of use of VDM tools. We provide a conceptual framework for evaluating the quality of use of VDM tools and apply it to the evaluation of the SOM. In the evaluation, we use an inquiry technique for which we developed a questionnaire based on the proposed framework.

The contributions of the thesis consist of three new evaluation techniques and the results obtained by applying these evaluation techniques. The thesis provides a systematic approach to evaluation of various visualization techniques. In this respect, first, we performed and described the evaluations in a systematic way, highlighting the evaluation activities, and their inputs and outputs. Secondly, we integrated the evaluation studies in the broad framework of usability evaluation.

The results of the evaluations are intended to help developers and researchers of visualization systems to select appropriate visualization techniques in specific situations. The results of the evaluations also contribute to the understanding of the strengths and limitations of the visualization techniques evaluated and further to the improvement of these techniques.

Keywords: multidimensional visualization techniques, visual data mining, evaluation of visualization techniques

Abstrakt (Abstract in Swedish)

Visuella datautvinningsverktyg (Visual Data Mining methods, VDM) utnyttjar visualiseringstekniker för att representera stora mängder mångdimensionella data grafiskt. Användaren får då stöd i analysering av data på olika detaljnivåer. Användaren letar ofta efter extremvärden, mönster och modeller – i form av kluster, klasser, trender och relationer – i olika kategorier av data, t.ex. ekonomiska data, affärsinformation, osv.

Denna avhandling fokuserar på att evaluera mångdimensionella visualiseringstekniker, särskilt sett från behov som näringslivets användare har. Avhandlingen behandlar tre olika forskningsproblem. Det första problemet är att evaluera visualiseringstekniker som bygger på projektion, med avseende på teknikernas förmåga att bevara de ursprungliga avstånden mellan datapunkter. Teknikernas förmåga att bevara strukturen i data i samband med klustrering bedöms också genom att använda existerande mått på klustervaliditet. Jag illustrerar verktygens användbarhet genom att evaluera fem visualiseringstekniker: huvudkomponentanalystekniken (PCA), Sammons Mapping-teknik (Sammon’s Mapping), visualiseringstekniken med den självorganiserande kartan (SOM), visualisering med radiella koordinater (Radial Coordinate Visualization) och visualiseringstekniken Star Coordinates. Det andra forskningsproblemet berör evaluering av olika visualiseringstekniker med hänsyn till deras effektivitet i visuell datautvinning av affärsdata. För detta ändamål har jag utvecklat en enkätbaserad evalueringsteknik och använder den för att evaluera nio visualiseringstekniker. De evaluerade visualiseringsteknikerna är: Multipla linjediagram (Multiple Line Graphs), permutationsmatriser (Permutation Matrix), Survey Plot-tekniken, Scatter Plot Matrix-tekniken, tekniken med parallella koordinater (Parallel Coordinates), Treemap-tekniken, huvudkomponentanalystekniken, Sammons Mapping-teknik och SOM-tekniken. Det tredje problemet är att evaluera användbarheten av VDM-verktyg. Jag konstruerar en begreppsram för användbarhetsevaluering av VDM- verktyg och använder den för att evaluera SOM. Evalueringen är enkätbaserad och baserar sig på den konstruerade begreppsramen.

Avhandlingens kontribution består av tre nya evalueringstekniker samt evalueringsresultat som erhållits genom tillämpningen av teknikerna. Avhandlingen presenterar ett systematisk angreppssätt för evaluering av olika visualiseringstekniker. Jag har utfört och beskrivit evalueringen, illustrerat och analyserat evalueringsstegen deras input och output. Jag har sammanställt evalueringsresultaten till en bred begreppsram för användbarhetsevaluering.

Resultaten av evalueringarna är avsedda att hjälpa utvecklare och forskare inom visualisering att välja ändamålsenliga visualiseringstekniker i specifika

situationer. Resultaten av evalueringarna bidrar också till ökad kunskap gällande styrkorna och svagheterna hos de evaluerade visualiseringsteknikerna och vidare till att förbättra dessa tekniker.

Nyckelord: mångdimensionella visualiseringstekniker, visuell datautvinning, evaluering av visualiseringstekniker

Acknowledgements

I thank my supervisor, Professor Barbro Back, for encouraging and guiding me in conducting research in the field of visual data mining. I am grateful for the constructive advice and continuous support that she has given me during my doctoral studies.

I thank Professor David L. Olson from University of Nebraska, USA, and Professor Waclaw Gregory Wojtkowski from Boise State University, USA, for reviewing and evaluating my thesis. I am grateful to Professor Waclaw Gregory Wojtkowski for accepting the invitation to be the opponent of my dissertation.

Turku Centre for Computer Science (TUCS) and Department of Information Technologies at Åbo Akademi University are gratefully acknowledged for the excellent research environment and for the financial support provided for my studies. I gratefully acknowledge also the financial support received from Academy of Finland (grant number 104639), Åbo Akademi University, ESF (Dagmar och Ferdinand Jacobssons fond), Nokia Foundation, Stiftelsen för Åbo Akademi Forskningsinstitut, and TEKES (grant number 40435/05 and 40020/07).

I am grateful to Tomas Eklund for allowing me to use his financial benchmarking dataset in my research. Tomas has also read earlier drafts of my thesis and suggested various improvements. I also thank him for correcting the language of the manuscript and for the professional help that he has given me on many occasions.

I thank Mikko Rajanen for his cooperation in writing our joint papers and for introducing me to the field of usability and human-computer interaction.

Many thanks are due to Professor Christer Carlsson – director of the Institute for Advanced Management Systems Research (IAMSR), Professor Pirkko Walden, and all colleagues at IAMSR for the feedback that they have given me during the postgraduate seminars. In particular, I thank Franck Tétard for the useful comments regarding usability evaluation that he has given me.

I would like to express my gratitude to Professor Barbro Back, Professor Eija Helena Karsten, Professor Sanna Salanterä, Professor Tapio Salakoski and all members of the Louhi research group for positively influencing my research work by providing me with the opportunity to gain new knowledge and experience from the project seminars and discussions. I thank Marketta Hiissa, Shuhua Liu, Hanna Suominen, Filip Ginter, Tapio Pahikkala, and Sampo Pyysalo for the insight given into the evaluation of text mining methods during our cooperation.

Special thanks are due to all members of the Data Mining and Knowledge Management Laboratory for contributing to our seminars with stimulating discussions and ideas. In particular, I thank my colleagues Marketta Hiissa, Piia Hirkman and Minna Kallio for giving me valuable feedback on an earlier version of the thesis.

I am grateful to Cristina Seceleanu for carefully reading and providing me with valuable suggestions on an earlier version of Chapter 1. I thank Dragoş Truşcan for providing me with helpful feedback on a draft of the first chapters of the thesis. I thank also Evgeni Tsivtsivadze for reading an earlier version of Chapter 5 of the thesis and providing me with valuable comments and ideas.

I am also grateful to the administrative personnel from TUCS, Åbo Akademi University, Department of Information Technologies, and IAMSR for always being receptive to my requests.

I gratefully acknowledge the support given by the students at Åbo Akademi University and University of Oulu who have accepted to participate in the empirical studies by answering the questionnaires.

I am grateful to my friends in Turku for the nice times that we spent together and their generosity in offering me help whenever I needed: Alina, Cristina, Diana, Elena, Irina, Luigia, Satu, Dragoş, Eugen, Ion, Tiberiu, Viorel and Vladimir.

I would not have finished this thesis without the moral support given by my family. I thank my fiancé – Mikko Rajanen, Mikko's family, my mother and my sister for their encouragements, trust and friendship. I dedicate this thesis to my family and in the memory of my father.

Turku, September 23, 2008 *Dorina Marghescu*

Contents

List of original publications

Paper 1. Marghescu, D. (2008). Usability Evaluation of Information Systems: A Review of Five International Standards, in Barry, C., Lang, M., Wojtkowski, W., Wojtkowski, G., Wrycza, S., & Zupancic, J. (eds) *The Inter-Networked World: ISD Theory, Practice, and Education,* Springer-Verlag: New York, (*Proceedings of the 16th International Conference on Information Systems Development (ISD 2007)*, Galway, Ireland, August 2007). In Press.

Paper 2. Marghescu, D. (2006). Evaluating the Effectiveness of Projection Techniques in Visual Data Mining. In *Proceedings of the 6th IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP 2006),* Palma de Mallorca, Spain, August 2006.

Paper 3. Marghescu, D. (2007). Evaluation of Projection Techniques Using Hubert's Γ Statistics. In *Proceedings of the IADIS 1st European Conference on Data Mining*, Lisbon, Portugal, July 2007.

Paper 4. Marghescu, D. (2007). Multidimensional Data Visualization Techniques for Exploring Financial Performance Data. In *Proceedings of the 13th Americas' Conference on Information Systems (AMCIS 2007)*, Keystone, Colorado, USA, August 2007.

Paper 5. Marghescu, D. (2007). User Evaluation of Multidimensional Data Visualization Techniques for Financial Benchmarking. In *Proceedings of the European Conference on Information Management and Evaluation (ECIME 2007),* Montpellier, France, September 2007.

Paper 6. Marghescu, D., Rajanen, M., and Back, B. (2004). Evaluating the Quality of Use of Visual Data-Mining Tools. In *Proceedings of the 11th European Conference on IT Evaluation (ECITE 2004),* Amsterdam, The Netherlands, November 2004.

List of abbreviations and acronyms

2D = 2-dimensional
3D = 3-dimensional
DM = Data mining
EC = Equity to capital
EUCS = End-User Computer Satisfaction
IC = Interest coverage
IS = Information systems
ISO = International Organization for Standardization
ISO/IEC = International Organization for Standardization /International Electrotechnical Commission
IT = Information technology
KDD = Knowledge discovery in databases
PC = Principal component
PCA = Principal Components Analysis
QR = Quick ratio
ROTA = Return on total assets
ROE = Return on equity
RQ1 = Research question 1
RQ2 = Research question 2
RQ3 = Research question 3
RT = Receivables turnover
Radviz = Radial Coordinate Visualization
SOM = Self-Organizing Map
SUMI = Software Usability Measurement Inventory
UE = Usability evaluation
VDM = Visual data mining

List of figures

List of tables

Part I. Research Summary

1. Introduction

1.1. Background and motivation

Today, companies easily store large volumes of high-dimensional data with the intention of using them for gaining competitive advantage. To be useful, the collected data have to be analyzed with appropriate tools. Data mining techniques represent one category of such tools. They are used to extract knowledge from data by means of data mining algorithms. The discovered knowledge is referred to as patterns found in data, patterns that must be interesting (novel, valid, potentially useful and understandable) to the user (Fayyad et al. 1996). These patterns are typically represented in the form of clusters, classes, trends, relationships, and summaries of the original data. The totality of patterns of the same nature – e.g., all clusters, all classes, or all relationships – found in the data represents a model or structure of the data (Ferreira de Oliveira and Levkowitz 2003). However, in most cases, these data mining results are communicated to the business user in a format that is difficult to understand and/or interpret (Kohavi et al. 2002).

To address the problem of representing business data and data mining results in an accessible format, researchers investigate the possibilities of using *information visualization techniques*. Information visualization is concerned with representing abstract information (e.g., business data) on computer-supported and interactive graphical displays. The tools that use information visualization technology for *knowledge discovery* are referred to as *visual knowledge tools* (Card et al. 1999). *Visual data mining tools* (Keim 2002) represent a category of visual knowledge tools.

Visual data mining (VDM) tools handle multidimensional data, that is, data with more than three dimensions (variables). Moreover, the volume of the data handled is usually large, that is, from hundreds to millions or more data items (objects or cases).

On the one hand, the VDM tools employ information visualization techniques in order to represent the *data* graphically and to involve the user in exploring data at different levels of detail (Keim 2002). The users examine the graphical representation of the data in order to find outliers (anomalies), and detect patterns and models (in the form of clusters, classes, trends, and relationships) in different categories of data (e.g., financial, business information). On the other hand, the VDM tools employ information visualization techniques in order to represent the *data models* graphically. The models are obtained automatically, by employing a data mining technique (e.g., neural networks, machine-learning, clustering analysis techniques). The models are then presented to the user in a graphical format that should be easy to interpret (Ferreira de Oliveira and Levkowitz 2003).

Most of the research in the fields of information visualization and VDM focuses on *developing* new information visualization techniques, and on exploring their capabilities for providing insights into financial or business datasets. The research literature concerning the *actual use* of visual data mining tools to get insight into financial data is relatively sparse, despite the fact that this technological approach is advocated as being suitable for both financial data and business users. Financial data are very complex due to their high dimensionality, large volume and diversity of data types. Business users are demanding straightforward visualizations and task-relevant outputs, due to the time and performance constraints under which they work (Kohavi et al. 2002).

There is not much evidence that business users use advanced visualization techniques frequently in their work. A survey conducted in 2003 among managers of HEX-listed Finnish companies (Eklund et al. 2004) revealed that spreadsheet tools are employed *daily* by 72.97% of the respondents, while database systems by 44.65%. Decision-support systems, which were defined as being “information systems that present data from databases in a form suitable for decision making”, were mentioned only by 29.73% of the respondents as being used daily.

One of the most popular commercial tools used by business users for visualizing financial data is Microsoft Excel, which provides only the standard visualization tools (bar, column, line, scatter plot, pie, stacked bars, etc.). Excel lacks the capabilities of interacting with and linking different visualizations and does not implement some of the recently developed visualization techniques for multidimensional data. To overcome some of these limitations, one can use the available commercial applications that can be integrated with Excel and provide more sophisticated visualization tools (e.g., Excel Dashboards, Crystal Xcelsius).

In order to increase the popularity of information visualization techniques among business users, such methods need to be thoroughly *evaluated*. The techniques' strengths and limitations can be highlighted only by systematic approaches to evaluating them (Chen and Czerwinski 2000; Plaisant 2004).

In brief, the rationale of thesis is built upon the following facts:

- There are financial and business data available that are difficult to structure in a meaningful manner, due to their large volume and high dimensionality;
- The traditional data analysis tools (e.g., data mining techniques) provide results that are difficult to understand/interpret by business users at certain time and performance parameters;
- The alternative analysis tools supported by the advances in information visualization and VDM are not adopted or are adopted on a small scale by business users;
- The evaluation of the visualization techniques and tools is not extensively discussed and researchers emphasize the need of systematic evaluations and appropriate measures and procedures.

1.2. Aim of the thesis and research questions

1.2.1. Evaluation of visualization techniques

In this thesis, our goal is to *evaluate* multidimensional visualization techniques in data mining tasks, especially from the business user's perspective. The evaluation of the information visualization techniques received little attention until recently, when an avalanche of new visualization techniques has become available to researchers and practitioners. However, the well-established models and metrics for evaluating traditional information systems (IS) may not be entirely adequate or sufficient for the evaluation of information visualization techniques.

Many authors highlight the research potential of the problem of evaluation of visualizations (Keim et al. 1994; Keim and Kriegel 1996; Chen and Czerwinski 2000; Pickett and Grinstein 2002; Plaisant 2004). They all point out the need for test datasets, new procedures and approaches for conducting evaluation of effectiveness of visualization techniques, and the necessity to "get beyond the current demonstrational stage" that characterizes many of the present evaluation approaches.

We concentrate our attention on those visualization techniques that are suitable for graphically representing *table data*. Hoffman and Grinstein (2002) use the term "table visualizations" to refer to this category of visualization techniques. Table data are datasets that are expressed as tables in which, typically, the rows represent cases (records, objects, or data items) and the columns represent variables (attributes, characteristics, or dimensions) of data[1]. Table data do not typically have inherent hierarchical structure.

Moreover, we focus on table visualization techniques that are capable of displaying multidimensional or multivariate data[2]. These techniques are referred to in the literature as *multidimensional data visualizations* (Hoffman and Grinstein 2002) or *multidimensional visualizations* (Card et al. 1999; Soukup and Davidson 2002).

An example of multidimensional visualization is the Self-Organizing Map (SOM). The SOM algorithm is a special type of neural network, developed by Kohonen (2001) in 1980s. The SOM algorithm produces a two-dimensional layout in which high-dimensional data are represented so that similar data items are placed close together. The SOM technique can be used in visualizing large high-dimensional datasets (by projecting data into a lower-dimensional space) and in clustering. The capabilities of the SOM for exploring financial or business data have been extensively investigated in research settings in different domains. For example, Back et al. (1998, 2000), Eklund et al. (2003), Costea and Eklund (2003), Eklund (2004) explored the use of SOM in financial benchmarking; Kaski and Kohonen (1996) – in macroeconomics; and Alhoniemi (2000) – in industrial processes. Kaski et al. (1998) and Oja et al. (2003) provide a comprehensive literature survey of research regarding the SOM technique. However, there is no evidence that the technique is extensively used by business users, for example, in financial benchmarking. One way to find out why business users do not employ this technique in their work is to evaluate its strengths and limitations. Moreover, it is important to analyze the capabilities of the SOM and other visualization techniques of representing and/or revealing patterns in financial benchmarking data.

1.2.2. Research questions

Based on the above exposition, the following *research questions* arise:

RQ1. How can the SOM and/or other projection-based multidimensional visualization techniques be evaluated as to their effectiveness for preserving

---

[1] Card et al. (1999) use the Data Table concept to refer to data represented in a table. They illustrate the concept by assigning attributes to rows and cases to columns.

[2] Throughout the thesis we will use the term dimensions to refer to the variables in the data. Thus, by multidimensional data we denote data with more than three variables. By multidimensional visualizations we denote visualizations that are capable of displaying multidimensional data. Other authors make a distinction between independent and dependent variables, and refer to independent variables with the term data dimensions and to dependent variables with the term variates (see Hoffman and Grinstein 2002).

the original data structure (e.g., distances between data items and clustering)? What are the results of these evaluations? (i.e., are the techniques effective in the mentioned context?)

RQ2. How can the SOM and/or other multidimensional visualization techniques be evaluated as to their effectiveness in financial benchmarking? What are the results of these evaluations? (i.e., are the techniques effective?)

RQ3. What quality-of-use attributes of VDM tools (e.g., tools based on the SOM and/or other visualization techniques) must be addressed in an evaluation? How can these attributes be measured? What are the results of these evaluations? (i.e., what are the strengths and weaknesses of the SOM?)

The evaluation problems mentioned in questions RQ1 and RQ2 are related to the selection of the most adequate techniques to be implemented in a VDM tool. The evaluation problem mentioned in question RQ3 refers to the description of a tool in terms of its strengths and weaknesses, with the purpose of improving the uncovered weak points.

In the thesis, we evaluate the following multidimensional visualizations in addition to the SOM:

- Multiple Line Graphs (Bertin 1981),
- Permutation Matrix (Bertin 1981),
- Survey Plot (Demsar et al. 2004),
- Scatter Plot Matrix (Cleveland 1993),
- Parallel Coordinates (Inselberg 1985),
- Treemap (Johnson and Shneiderman 1991),
- Principal Components Analysis (Sharma 1995; Duda et al. 2000),
- Sammon's Mapping (Sammon 1969),
- Radial Coordinate Visualization (Radviz) (Hoffman et al. 1997; Hoffman 1999), and
- Star Coordinates (Kandogan 2000).

We develop three new visualization evaluation techniques that address the research questions/problems formulated above. We also apply these evaluation techniques to the evaluation of the above visualization techniques.

In evaluating the visualization techniques, we take into consideration the characteristics of both visualization systems and information systems. We provide a systematic approach to the evaluation of various visualization techniques. As such, we firstly design, perform and describe the evaluations in a systematic way, highlighting the evaluation activities (data collection, analysis) and their inputs (attributes, measures) and outputs (results). Secondly, we integrate the evaluation studies in the broad framework of usability evaluation.

The results of the evaluations are intended to help developers and researchers of visualization systems to select appropriate visualization techniques in specific situations. Moreover, the evaluations' results are intended to contribute to the understanding of the strengths and limitations of the visualization techniques evaluated and further to the improvement of these techniques.

1.3. Related work

1.3.1. Related work on evaluation methods/techniques

Plaisant (2004) acknowledges the existence of the following main types of evaluation studies of visualization systems and techniques: *controlled experiments* comparing design elements *or* comparing two or more tools, *usability evaluation* of tools, and *case studies*. We found, in addition to these, another category, namely *informal evaluation* of new visualization tools, aiming at illustrating or demonstrating their effectiveness for solving particular business problems (or engineering problems and other domain-oriented applications, e.g., in (Keim and Kriegel 1994; Inselberg 1997).

Usability evaluation (UE) of a system, in general, is a complex activity, which has received lots of attention from researchers, especially in the human-computer interaction community. There are many models and measures of usability, as well as UE methods described in the literature (Nielsen 1993; Dix et al. 1998; Ivory and Hearst 2001).

There are two main views of usability (Bevan 1995): one regarding usability as a characteristic of a (software) product (e.g., in ISO/IEC 9126-1 (ISO 2000b)), and the other regarding usability as an objective of the (software) product (e.g., in ISO 9241-11 (ISO 1998)). The later view is called by Bevan *quality of use*, and it is defined as being "the extent to which a product satisfies stated and implied needs when used under stated conditions." Bevan also points out that measuring quality of use implies measuring aspects such as *effectiveness*, *efficiency* and *satisfaction* of the users in achieving specified goals in a specified context of use.

Two aspects of usability that are frequently measured in visualization evaluations are *effectiveness* (or *user performance*) and *user satisfaction* (e.g., Mackinlay 1986; Shneiderman 1994). The user satisfaction is typically measured subjectively (user rating) by employing an inquiry technique. For evaluating effectiveness there are many practices and measures proposed.

In information visualization evaluation, there are many views of effectiveness and the concept is termed in different ways by different communities or researchers (e.g., *visual efficacy* by Bertin (1981); *effectiveness* by Mackinlay (1986) and Card et al. (1999)). Mackinlay (1986) defines effectiveness as being the capability of the visualization to exploit the "output medium and the human visual system". Effectiveness is viewed as enabling the user to read, understand and interpret the display easily, accurately, quickly, etc. Accordingly, effectiveness depends not only on the graphical design but also on the capabilities of the viewer. Card et al. (1999) use a similar definition of effectiveness. They define effectiveness as the capability of the visualization to be perceived well by the human: an effective display "is faster to interpret, can covey more distinctions, or leads to fewer errors than some other mapping". Usually, the effectiveness is measured in terms of time to complete a task or a set of tasks or in terms of quality of the tasks' solutions (e.g., Dull and Tegarden 1999; Risden et al. 2000).

Hoffman (1999) and Grinstein et al. (2002) evaluate several multidimensional visualizations with respect to their effectiveness for revealing outliers and patterns (clusters, rule discovery, etc.) in different real benchmark datasets. Keim and Kriegel (1994; 1996) provide a similar approach, but they highlight the importance of using artificial (synthetic) datasets in benchmarking different visualizations. The limitation of these approaches is that the evaluation is based only on authors' experience and use of the techniques. Hoffman (1999) also proposes different objective measures for evaluating different characteristics of the visualization. Other examples of objective measures of the effectiveness are in (Keim and Kriegel 1996).

Keim et al. (1994) propose a conceptual model of generating artificial benchmark datasets based on criteria such as number of data dimensions, number of data points, type of data, structure of data, etc.

1.3.2. Related work on evaluation of projection techniques

Projection techniques are dimensionality reduction techniques (Kohonen 2001, p. 34). They can be used for visualizing multidimensional data when the number of new dimensions in the data is one, two or three. Examples of projection techniques are the SOM, Sammon's Mapping, PCA, and more recent techniques like Radviz and Star Coordinates. The visualization techniques can be evaluated by using objective or subjective approaches.

Regarding the subjective evaluation of the **SOM**[3], Eklund (2004) reported the results of an expert survey in which business users were asked to assess a SOM-

[3] In this section, we use the bold font to highlight that a technique is also evaluated in our thesis.

based model for financial benchmarking. The aspects under evaluation were content, format, ease of use, timeliness and accuracy (based on the End-User Computing Satisfaction model in (Doll and Torkzadeh 1988)). The results reveal that managers showed a positive attitude toward the SOM-based model. Ståhl et al. (2006) tested the usability of a SOM-based prototype tool for financial benchmarking, by conducting a cooperative observational evaluation study. The participants were asked to solve a set of tasks and answer a questionnaire regarding their user satisfaction with the tool. The results reveal good user performance (completion time and error rates) and largely positive user attitudes toward the SOM tool.

Mandl and Eibl (2001) proposed the objective evaluation of SOM and another technique (based on latent semantic indexing) in the context of information retrieval of documents. Their method uses the Spearman correlation coefficient to measure the correlation between the two types of displays, without pointing out which one is better or how well each technique performs.

Sammon's Mapping and **PCA** are classical statistical techniques used for exploratory data analysis and dimensionality reduction. **Radviz** and **Star Coordinates** are recently developed and their capabilities are not extensively explored. Hoffman (1999) and Pillat et al. (2005) have evaluated Radviz. Pillat et al. compared Radviz with Parallel Coordinates and found that the Radviz technique has a layout more difficult to interpret. Hoffman points out a weakness of the Radviz technique in terms of overlapping of data points. He suggests that this problem can be improved by varying the layout of the dimensions and by using data normalization. The advantages of the technique are expressed in terms of the capability to show clusters and concentrating regions, as well as for showing the general features of a dataset. Kandogan (2000) highlights the usefulness of the Star Coordinates technique in visualizing and exploring data with inherent hierarchical clustering structure.

1.3.3. Related work on evaluation of visualizations addressed in the thesis, other than projection techniques

Hoffman and Grinstein (2002) provided a survey of multidimensional visualization techniques, among which Multiple Line Graphs, Permutation Matrix, Survey Plot, Scatter Plot Matrix, Parallel Coordinates, Treemap, Sammon's Mapping, Radviz, and SOM are described and illustrated. **Scatter Plot Matrix** and **(Multiple) Line Graphs** are classical visualization and exploratory data analysis techniques. The Scatter Plot Matrix is especially used for detecting the relationships among variables. (Multiple) Line Graphs are typically used for depicting time-series data in order to detect trends and cycles in the data. **Permutation Matrix** is a variation of column graphs developed by

Bertin (1981). Bertin illustrates the visual efficacy of this technique by analyzing a multidimensional business dataset. **Survey Plot** is a variation of bar graphs in which different bar graphs are used to depict different variables.

Hoffman (1999) subjectively evaluated the effectiveness of five techniques (i.e., **Radviz, Parallel Coordinates, Survey Plot**, Circle Segment, and **Scatter Plot Matrix**) in revealing outliers, clusters, rule discovery (classification), as well as in supporting different data types and data dimensionality. The evaluation was based on experiments on well-known datasets (i.e., from UCI Machine Learning database (Newman et al. 1998)). The results of these studies show that one visualization technique alone is usually not capable of revealing all interesting patterns in the data and, hence, it is desirable that multiple visualizations be employed when analyzing a dataset. Moreover, the effectiveness of the techniques in revealing certain patterns is dependent on the dataset and, thus, the results cannot easily be generalizable across different datasets. Hoffman (1999) also developed objective measures to characterize a visualization for a particular dataset. He defined the concept of *Display Utilization Grid* and, based on it, defined several measures. However, the correlation of these measures with the effectiveness for showing interesting patterns was not formally demonstrated.

Regarding the evaluation of the **Parallel Coordinates** technique, Inselberg (1997) illustrated the benefits of using Parallel Coordinates in a practical problem. Inselberg points out that augmenting Parallel Coordinates with interaction capabilities is important in knowledge discovery. Keim and Kriegel (1996) compared two novel dense-pixel displays with the Parallel Coordinates and Stick Figures displays in revealing outliers (hot spots), clusters, distributions and functional dependencies. Pillat et al. (2005) compared Parallel Coordinates with Radviz. Parallel Coordinates are useful in the identification of outliers, and when subsets of data are displayed, in showing specific features of the dataset as well as specific values of data items.

The **Treemap** technique has also been evaluated in many studies. Barlow and Neville (2001) compared the Treemap technique with three other visualizations of hierarchical data in the context of decision tree analysis. Treemap has been found inferior in terms of response time and accuracy in given tasks, as well as in user preference. Similarly, Stasko et al. (2000) compared the Treemap technique with another technique for visualizing hierarchical structures in the context of file/directory search. The performance of Treemap was lower in the given tasks, especially on initial use, but improved over time. The user preference was also lower in the case of Treemap. Card et al. (1999, p. 151) point out that the Treemap technique is especially effective for displaying data in which one variable is quantitative and large values are important. They also refer to (Johnson 1993), in which Treemap is evaluated empirically and it is shown that the technique is used with efficiency after 15 minutes of training. The Treemap technique is used by SmartMoney (www.smartmoney.com) to enable the online visualization and exploration of high-dimensional financial data. An example of such visualization tool is the Map of the Market, used to observe the changes in stocks in almost real time.

Ward and Theroux (1997) evaluated **Scatter Plot Matrix**, **Parallel Coordinates** and a variation of **PCA** (augmented with star glyphs). They used both real and artificial datasets. They conducted experiments in which the visualizations were evaluated as to their effectiveness in revealing outliers and clusters. 19 users participated in the evaluation of *static images* obtained from the three visualization techniques on different datasets. The purpose of their evaluation was to determine the strengths and weaknesses of the techniques in revealing outliers and clusters when different characteristics of the data were controlled.

1.4. Contributions of the thesis

Given the fact that the evaluation approaches currently employed to evaluate visualization techniques leave room for improvement of evaluation practice, we regard the evaluation problem of visualization techniques as a practical problem still open to research. The evaluation problem is faced by developers of VDM tools in their task to select a set of techniques to be implemented in a system and to assess the strengths and limitations of techniques in order to improve them. Hence, we adopt the *design science* (March and Smith 1995) in answering our research questions with the aim to find suitable ways to evaluate multidimensional visualizations.

The contributions of the thesis consist of three new visualization evaluation techniques that address each of the three research questions, respectively. We have conducted several evaluation studies of multidimensional visualization techniques based on the evaluation techniques proposed. The results of the evaluations are useful in comparing, selecting and improving the visualization techniques.

We have reviewed five ISO[4] standards regarding UE and derived a framework for UE of information systems, which can also be used to systematically conduct/describe the evaluation of information visualization techniques or VDM tools. The UE framework highlights the activities in the UE process. We have

[4]The International Organization for Standardization (ISO) standards reviewed are: ISO/IEC 9126-1 (ISO 2000b), ISO/IEC 14598-1 (ISO 1999b), ISO 9241-11 (ISO 1998), ISO 13407 (ISO 1999a), and ISO 18529 (ISO 2000a).

used this framework to systematize the implementation, presentation and discussion of the evaluation techniques proposed.

To answer the first research question, RQ1, we have proposed an evaluation technique, aiming at *objective evaluation* of different projection techniques with respect to their *effectiveness* for preserving the original structure of a given dataset. The effectiveness concept has been operationalized in terms of *clustering validity measures* (Theodoridis and Koutroumbas 1999). We have provided procedures for calculating different existing quantitative measures of clustering validity with the purpose of evaluating different projection techniques' capabilities in preserving the original structure of the data (in terms of distances between data points and clustering model). The projection techniques evaluated are PCA, Sammon's Mapping, Radviz, Star Coordinates, and the SOM. The advantage of the objective evaluation is that it uses quantitative measures and does not require user involvement in the evaluation. The evaluation results show that the effectiveness of the techniques for preserving the clustering structure differs with the dataset, but in general, PCA, Sammon's Mapping, and the SOM have provided the best results on our datasets.

Our approach overcomes the limitation of the current approaches to objective evaluation of visualizations, which use quantitative measures that do not show a correlation with the effectiveness of the techniques in data mining tasks (e.g., Hoffman 1999; Keim and Kriegel 1996). However, we have focused only on the clustering task and on preserving the original data structure.

To answer the second research question, RQ2, we have proposed an evaluation technique, aiming at *subjective user evaluation* of different visualization techniques as to their effectiveness in financial benchmarking. First, we have provided an initial evaluation of multiple techniques based on a model described by Soukup and Davidson (2002). The authors recommend deriving data mining tasks from a business problem and then applying visualization techniques that are capable of solving the derived tasks. Secondly, we have proposed a user evaluation approach by employing the questionnaire technique in order to collect data about how the users evaluate/interpret the visualizations. We have operationalized the concept of *effectiveness* (Card et al. 1999) in terms of correctness of interpretation of the patterns depicted by the visualization and the number of distinctions (patterns) that a technique is capable of revealing. Our approach is similar to the one in (Ward and Theroux 1997). However, our purpose is to evaluate a set of techniques in order to obtain a subset of them, consisting of those that are most effective in solving a business problem. Moreover, the data analysis of the collected data is different from (Ward and Theroux 1997). They evaluate the users' answers (patterns identified by the users) by comparing them with the patterns found by automated data mining techniques. We subjectively evaluate the users' answers by analyzing the correctness of the answers in terms of identification of patterns and interpretation of the visualization.

The proposed evaluation technique is especially useful in the early stage of the development of a visualization system. We have evaluated the following visualization techniques: Multiple Line Graphs, Permutation Matrix, Survey Plot, Scatter Plot Matrix, Parallel Coordinates, Treemap, PCA, Sammon's Mapping, and the SOM. The evaluation results have highlighted the strengths and limitations of various visualization techniques in effectively solving a financial benchmarking problem. The evaluation results show that the techniques that prove to be most effective in the given context are the SOM, Survey Plot, Permutation Matrix, PCA, Multiple Line Graphs and Parallel Coordinates. We have evaluated the questionnaire by illustrating its use in three distinct situations, with three different groups of users.

One limitation of the existing approaches to subjectively evaluating effectiveness of visualization in data mining tasks is that the evaluation is based only on expert evaluation (e.g., Hoffman 1999; Keim and Kriegel 1996). Other approaches involve several users, but they focus on evaluating the techniques for their performance on different datasets, rather than in an applied context, such as a business problem (e.g., Ward and Theroux 1997).

To address the third research question, RQ3, we have developed a conceptual framework for evaluating visual data mining tools from the user's perspective. We have used the concept of *quality of use* (Bevan 1995) and applied it to the VDM tools. We have reviewed relevant evaluation literature in order to derive attributes of quality-of-use for three levels of analysis: visualization, interaction and information. Based on the conceptual framework, we have developed a questionnaire in order to subjectively measure (by user rating) the quality-of-use attributes. We have applied the questionnaire to the evaluation of the SOM in a financial benchmarking problem. We have also collected data about the user performance in a set of tasks related to the financial benchmarking problem. The results of the evaluation show that the SOM-based tools under analysis provide interesting and new information for the given tasks. The technique is considered helpful in understanding and analyzing the data. The tools are found easy to use by the respondents and most of the visualization features are found helpful and adequate. The limitations seem to be the time to obtain a good map, preciseness and accuracy of the results and difficulty in interpreting the results. We have evaluated the questionnaire by calculating the reliability (internal consistency) of the scales.

The current approaches to evaluating quality of use (usability) or some aspects of it, such as effectiveness or user satisfaction, have the limitation that the evaluation results do not provide detailed information that could help in the improvement of the visualization technique. Our approach to the evaluation of quality of use of VDM tools overcomes this limitation, by providing characteristics and attributes of VDM tools and an instrument for measuring those attributes subjectively, by user rating.

1.5. Overview of thesis

The remainder of the thesis is organized as follows.

In **Chapter 2**, we describe the research methods used in this thesis. The focus is on presenting the characteristics of the design science. Other research approaches used in the thesis, that is, the descriptive research and the theoretical research, are also highlighted.

In **Chapter 3**, we describe the key concepts in information visualization and visual data mining. Among the concepts defined are information visualization, visualization techniques, visual data mining, and data mining tasks. We refer to the classification of the information visualization techniques provided by Keim (2002) in order to provide an overview of the techniques and to highlight what types of visualizations we evaluate.

In **Chapter 4**, we review related work on usability evaluation (UE) of information systems (IS) and evaluation of information visualization techniques/systems. We describe different classifications of UE methods and the factors that explain the differences between different types of methods. Moreover, we present the state of the art in visualization evaluation by highlighting the quality characteristics of visualization, current practices and examples of studies. Chapter 4 also presents the results of a review of the International Organization for Standardization (ISO) standards regarding usability evaluation (**Paper 1:** Marghescu 2008). The standards reviewed are ISO/IEC 9126-1, ISO/IEC 14598-1, ISO 9241-11, ISO 13407, and ISO 18529. Based on this review, we derive a framework for UE of IS, highlighting the activities of UE process.

In **Chapter 5**, we investigate the use of *quantitative measures* for assessing the effectiveness of different visualization techniques for preserving the original structure of the data. The effectiveness is measured by using existing *cluster validity measures*. We evaluate five projection techniques (SOM, Sammon's Mapping, PCA, Radviz and Star Coordinates). The research in this chapter is based on **Paper 2** (Marghescu 2006) and **Paper 3** (Marghescu 2007a).

In **Chapter 6**, we illustrate the use of nine multidimensional data visualization techniques in a *financial benchmarking* problem and provide an *initial evaluation and comparison* of the techniques. The nine visualization techniques under analysis are: Multiple Line Graphs, Permutation Matrix, Survey Plot, Scatter Plot Matrix, Parallel Coordinates, Treemap, PCA, Sammon's Mapping and the SOM. The work presented in this chapter is based on **Paper 4** (Marghescu 2007b).

In **Chapter 7**, we propose a *user evaluation technique* for the evaluation of multiple visualizations in data mining tasks. We describe the use of this method on the evaluation of the nine visualizations described in Chapter 6. The evaluation is subjective, based on user assessments, and concerns the effectiveness of the techniques in solving a financial benchmarking problem – the same as the one defined in Chapter 6. For illustrating and demonstrating the evaluation method, we conduct three empirical studies. The data collection is based on the questionnaire technique. The ideas and one evaluation study in this chapter are presented also in **Paper 5** (Marghescu 2007c).

In **Chapter 8**, we propose a *conceptual framework* for evaluating the *quality of use* of visual data mining tools. We define quality of use as reflecting the satisfaction of the user with all features of the tool: visualization of the data, information provided and interaction. Based on the framework, we develop a questionnaire and apply it to the empirical evaluation of the SOM. This chapter is based on **Paper 6** (Marghescu, Rajanen, and Back 2004).

In **Chapter 9**, we summarize and conclude our work, enumerate the theoretical and practical implications, the limitations of our approach, and propose ideas for future work. Table 1 summarizes the overview of the thesis.

Table 1. Overview of thesis

Topics	Research questions	Chapters	Papers
Introduction to the thesis		Chapter 1	
Research methods		Chapter 2	
Information visualization techniques and visual data mining: key concepts		Chapter 3	
Evaluation of visualization techniques		Chapter 4	Paper 1
Objective evaluation of projection techniques - evaluation of: - Principal Components Analysis, - Sammon's Mapping, - Radviz, - Star Coordinates, and - Self-Organizing Map	RQ1	Chapter 5	Paper 2 Paper 3
Subjective user evaluation of multiple visualization techniques in data mining tasks - evaluation of: - Multiple Line Graphs, - Permutation Matrix, - Survey Plot, - Scatter Plot Matrix, - Parallel Coordinates, - Treemap, - Principal Components Analysis - Sammon's Mapping, and - Self-Organizing Map - the case of financial benchmarking	RQ2	Chapter 6 Chapter 7	Paper 4 Paper 5
Framework for evaluation of visual data mining tools - quality-of-use evaluation - evaluation of the Self-Organizing Map - the case of financial benchmarking	RQ3	Chapter 8	Paper 6
Conclusions		Chapter 9	

2. Research methods: overview

In this chapter, we describe the research approaches adopted in this thesis for addressing the research questions. We start with presenting the characteristics of design science and constructive research. We continue by briefly presenting the descriptive research approach. Finally, we present in brief the theoretical approach. Within each section of the chapter, we first present an overview of the research approach and then we explain how we conducted the research.

2.1. Design science and constructive research

The research questions posed in Section 1.2.2 are *practical* problems faced by developers of VDM tools in their task to select a set of most adequate visualization techniques to be implemented in a system. Another task of developers of VDM tools is the assessment of the strengths and limitations of a visualization technique in order to improve the weaker points. Both tasks are different facets of the general problem of evaluation of visualization techniques.

Many researchers highlight the research potential of the visualization evaluation problem. Keim et al. (1994), Keim and Kriegel (1996), Chen and Czerwinski (2000), Pickett and Grinstein (2002) and Plaisant (2004) stress the need to **develop** *test datasets, new procedures* and *approaches* for conducting systematic evaluations of the visualization techniques/systems. Hence, our research problems (questions RQ1, RQ2 and RQ3 defined in Section 1.2.2) are both *relevant* and *suitable* to be addressed by developing new techniques for evaluation. We, therefore, adopt the *design science* or *constructive research* approach[5] to answer the research questions.

---

[5] Design science and constructive research are often used to denote the same type of scientific research. March and Smith (1995) and Hevner et al. (2004) define **design science** as being concerned with building and evaluating artifacts. Kasanen et al. (1993) define the **constructive approach** as being "a research procedure for producing constructions". Similarly, Iivari (1991) introduces the **constructive research methods** in the context of information systems development and define them as a category of research methods concerned with the development of conceptual or technical artifacts. The term **approach** is a general term used to denote "similar research methods" (e.g., in Järvinen 2001, p. 14). Järvinen defines a **research method** as being "a set and sequence of steps a researcher carries in a singular study." (p. 14). He makes a distinction between methods and **techniques**, the latter being concerned with providing means to solve specific tasks in a research work, such as data collection (e.g., questionnaire technique), or data analysis (e.g., Factor Analysis techniques).

March and Smith (1995) and Kasanen et al. (1993) highlight the *prescriptive* nature of design science and constructive research, respectively, in the sense that they offer prescriptions and develop artifacts that use those prescriptions. By proposing three visualization evaluation techniques, we provide prescriptions (guidance) on how to evaluate different visualizations in three specific situations.

2.1.1. Overview of design science and constructive research

In the Information Technology (IT)/ Information Systems (IS) discipline, the *design science* approach is described as a type of scientific research aimed at the development of new or better ways to achieve human goals (March and Smith 1995; Hevner et al. 2004). The human goals are related to *relevant practical* problems (or tasks) faced by humans in a specific environment. Thus, design science implies the existence of a relevant practical problem for which it tries to find a solution. The solution is usually referred to with the term "artifact" (March and Smith 1995; Hevner et al. 2004).

March and Smith (1995) distinguish between two activities in design science research: *build* the solution (1) and *evaluate* the solution (2). The **build** activity is a process by which researchers create an innovative and valuable solution (artifact). The researcher constructs a solution starting from existing knowledge of the problem-domain, methodology, and technology (Hevner et al. 2004).

The value (utility) of the artifact is assessed in the **evaluate** activity or process. The evaluation is also a difficult process, due to the fact that "performance is related to intended use, and the intended use of an artifact can cover a range of tasks" (March and Smith 1995). The evaluation criteria and metrics may differ for each intended task and context of use of the artifact. Therefore, the criteria and metrics must be determined (developed and/or specified) for the artifact in each particular environment (context of use) for which the artifact is evaluated. The employed metrics are justified by using natural science approaches (e.g., data collection and analysis). As March and Smith point out, *the evaluation of the constructed artifact is important in judging the research effort of building the artifact.*

There are four types of artifacts that design science research can produce: *constructs, models, methods,* and *instantiations* (March and Smith 1995; Hevner et al. 2004). **Constructs** represent the basic vocabulary of a problem. They are a set of concepts that "describe problems within a domain and specify their solutions" (March and Smith 1995). When the objective of the research is to provide new or better constructs for a given problem, the developed constructs should help in clarifying and defining the problem/solution.

Models are representations of the relationships among constructs. They can be viewed as a step forward toward the resolution of the practical problem under analysis. To be useful, they "may need to capture the structure of reality", but they do not need to be accurate on details (March and Smith 1995).

Methods are sets of steps (e.g., algorithms, procedures, and guidelines) used to perform a task. The order of the steps is not implied (it can be sequential, iterative or not relevant to the task - like in guidelines). Methods are based on the constructs and models defined for the practical problem/task to be solved. March and Smith point out that methods involve the *representation of tasks and results*, though not always in an explicit form. This fact corresponds to the translation from the problem model (representation) to the solution model (representation) during the problem-solving process. March and Smith point out that a difference between design science and natural science is that "natural science uses but does not produce methods." An example is given by the IS development methods.

Finally, **instantiations** are "realizations of an artifact in its environment" or "working artifacts" (March and Smith 1995). They can be information systems, or *tools* that address "various aspects of designing information systems". Examples of instantiations are software programs, implementations of algorithms, information systems in the form of prototypes or fully functional systems, etc. March and Smith point out that instantiations operationalize (i.e., make operational, put to use or into operation) constructs, models, and methods.

Based on the distinction between the activities and outputs in IS/IT research, March and Smith propose a framework describing viable research efforts in IT discipline. The framework is a 4x4 table corresponding to the four types of artifacts and the four types of activities: build, evaluate, theorize, and justify. We have previously presented the four types of artifacts: constructs, models, methods and instantiations, as well as the build and evaluate activities. The other two activities (theorize and justify) are part of the natural science approach and are discussed in detail by March and Smith (1995). The authors point out that each cell of the table-framework has a different *objective* and a different *research method* is suitable to employ in order to achieve the objective. Research endeavors can match more than one cell, and the research should be evaluated accordingly:

- Research concerned with the build activity should be evaluated based on the value or utility to a community of users, novelty, demonstration of effectiveness of the artifact, and/or significant improvement in performance.

- Research concerned with the evaluate activity creates metrics[6] and compares the performances of artifacts for specific tasks. A *model* is evaluated for its fidelity with real world phenomena, completeness, level of detail, robustness, and internal consistency. A *method* is evaluated for its operationality (the ability to perform the intended task or the ability of humans to effectively use the method if it is not algorithmic), efficiency, generality, and ease of use. An *instantiation* is evaluated for efficiency, effectiveness, and its impacts on the environment and its users.

General guidelines of conducting design science or constructive research are provided by Hevner et al. (2004) and Kasanen et al. (1993), respectively. Kasanen et al. (1993) describe constructive research process by dividing it into the following phases, whose order may vary:

1. "Find a practically relevant problem which also has research potential;
2. Obtain a general and comprehensive understanding of the topic.
3. Innovate, i.e., construct a solution idea.
4. Demonstrate that the solution works.
5. Show the theoretical connections and the research contribution of the solution concept.
6. Examine the scope of applicability of the solution."

Hevner et al. (2004) propose seven guidelines that assist researchers in conducting and evaluating design science research. The guidelines are:

I) Design as an artifact: the output of design science research must be an artifact (constructs, models, methods or instantiations).
II) Problem relevance: develop solutions (typically, technology-based) to important and relevant business problems.
III) Design evaluation: the artifact must be evaluated for its utility, quality and efficacy.
IV) Research contribution: the research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodology.
V) Research rigor: the research must rely on the use of rigorous methods in both the construction and evaluation of the artifact.
VI) Design as a search process: the search for an effective artifact should be based on the available means.
VII) Communication of research: the research must be presented to the interested (beneficiary) parties (e.g., technology-oriented, management-oriented audiences).

[6] ISO/IEC 14598-1 (ISO 1999b) defines *metric* as being "the measurement method and the measurement scale" and it includes methods for categorizing qualitative data.

2.1.2. Research method used in the thesis

Our work can be positioned into the *evaluate* activity of the design science framework, because we are concerned with evaluating different visualization artifacts (techniques). However, because we focus on the development of suitable evaluation techniques, the research can also be placed into *build* activity, where the research outputs are new *methods*, in particular *evaluation techniques*. The evaluation techniques are intended to assist practitioners in the evaluation of technological artifacts represented by information visualization techniques.

The development of the new evaluation techniques is based on the *constructive research method* described by Kasanen et al. (1993). The phases in the constructive research process are briefly highlighted in the following. The practical relevance and the research potential of our research questions are highlighted by researchers who point out the need to evaluate visualizations and the lack of appropriate and systematic approaches and techniques. The understanding of the topic is acquired by reviewing the relevant literature and the work related to the evaluation of information systems, in general, and of visualization techniques, in particular (Chapter 4). We develop three new evaluation techniques in order to answer the research questions that are defined in Section 1.2.2. We apply and illustrate the evaluation techniques (Chapters 5 – 8). We discuss their practical and theoretical implications, as well as their limitations in Chapter 9.

Whitefield et al. (1991) provide a description of the *method concept*. According to Whitefield et al., the literature highlights two components that define a method: *notation* and *procedure*. Notation refers to theories, models, and representations, and procedure refers to knowledge of scientific and engineering methodology. In our research frameworks, the notation corresponds to the constructs and models used in evaluation and based on which we have built the evaluation techniques. The procedures refer to the ways in which we collected and analyzed the data in order to calculate the measures used in evaluation.

Tables 2 – 4 present the research frameworks underlying our research, using the March and Smith's (1995) model. The contributions are highlighted with blue text. The knowledge on which we build the *evaluation techniques* is pointed out with normal text.

Tables 2 – 4 summarize how we evaluate different visualization techniques. We highlight in the *evaluate column* the fact that the evaluation is carried out by using established *constructs* (e.g., effectiveness, quality of use), and established *models* (e.g., clustering validity measurement, business problem – data mining tasks – visualization evaluation, dataset/task/visualization evaluation). The type of evaluation methods is highlighted in the *evaluate-method cell*. The artifacts being constructed (i.e., quality-of-use framework, metrics, data analysis procedures, questionnaires) are placed in the *build column*. The quality-of-use framework is placed in the *build-model cell*, because it represents a descriptive model highlighting the characteristics of quality of use for a visual data mining tool. We place the developed metrics, procedures and questionnaires in the *build-method cell*, because they represent methods to operationalize the concept of effectiveness or quality of use (user satisfaction).

Table 2. Research framework underlying the studies in Chapter 5 (RQ1)

Research output	**Build**	**Evaluate ***
Construct		Effectiveness (Mackinlay 1986; Card et al. 1999)
Model		• Clustering validity measurement (Theodoridis and Koutroumbas 1999) • Benchmark development: dataset/task/visualization (Keim and Kriegel 1996; Hoffman 1999; Grinstein et al. 2002)
Method	**Procedures for calculating the effectiveness for preserving the data structure by using clustering validity measures**	Simulation method
Instantiation		

* Five visualization techniques are evaluated: PCA, Sammon's Mapping, Radviz, Star Coordinates and the SOM.

Table 2 summarizes how we evaluate the effectiveness of projection-based visualizations for preserving the original structure in the data. We employ the model of clustering validity measurement (Theodoridis and Koutroumbas 1999) and the simulation method for evaluation. We adapt the calculation of the clustering validity measures to our evaluation problem and develop procedures for calculating different measures of effectiveness for preserving the original structure of the data. The evaluation of our evaluation approach is mainly illustrative, by applying the method on different benchmark datasets and visually analyzing the match between the results and the visual representation of the data.

Table 3. Research framework underlying the studies in Chapters 6 and 7 (RQ2)

Research output	**Build**	**Evaluate ***
Construct		• Effectiveness (Mackinlay 1986; Card et al. 1999) • Visual efficacy (Bertin 1981)
Model		• Business problem – Business questions – Data mining tasks – Answers – Effectiveness or Visual efficacy (Soukup and Davidson 2002; Bertin 1981; e.g.,) • Benchmark development: dataset/task/visualization (Keim and Kriegel 1996; Hoffman 1999; Grinstein et al. 2002)
Method	• **New qualitative metrics of effectiveness** • **Questionnaire** • **Data analysis procedure**	• Inspection method **(Chapter 6)** • Inquiry method **(Chapter 7)**
Instantiation		

* Nine visualization techniques are evaluated: Multiple Line Graphs, Permutation Matrix, Scatter Plot Matrix, Survey Plot, Parallel Coordinates, Treemap, PCA, Sammon's Mapping, and the SOM.

Table 3 summarizes how we evaluate the effectiveness of multiple visualizations for solving data mining tasks related to a business problem. We develop a questionnaire that is used to capture qualitative data about the effectiveness of the techniques in data mining tasks. The evaluation technique is developed based on the *effectiveness* (Mackinlay 1986; Card et al. 1999) and *visual efficiency* (Bertin 1981) concepts used to judge a visual presentation. Moreover, we use the models of evaluating different techniques presented by Soukup and Davidson (2002), Keim and Kriegel (1996), Hoffman (1999) and Grinstein et al. (2002). The assessment of the questionnaire is done by conducting three different empirical evaluation of the visualization with three different groups of users and comparing the results.

Table 4. Research framework underlying the studies in Chapter 8 (RQ3)

Research output	Build	**Evaluate ***
Construct		• Quality of use (Bevan 1995) • **Visualization, Interaction and Information characteristics** and **attributes** derived based on literature review: Tufte (1983), Doll and Torkzadeh (1988), Kirakowski (1994).
Model	**Quality-of-use framework**	
Method	• **New quantitative metrics of the attributes** • **Questionnaire** • **Data analysis procedure**	Inquiry method
Instantiation		

* One visualization technique is evaluated: the SOM.

Table 4 summarizes how we evaluate the quality of use of a visual data mining tool, namely the SOM. We develop a conceptual framework (descriptive model) of quality of use of VDM tools. The quality-of-use framework distinguishes between three levels of analysis: visualization, interaction and information, and proposes quality characteristics and attributes at each of these levels. The characteristics have been derived mainly based upon the EUCS (Doll and Torkzadeh 1988) and SUMI (Kirakowski 1994) models, as well as upon the principles of visualization described by Tufte (1983). Based on the framework, we develop a questionnaire. We evaluate the questionnaire by applying it in the evaluation of SOM-based tools, illustrating the results, as well as by calculating the reliability (internal consistency) of the scales.

Our research approach conforms also with the design research guidelines proposed by Hevner et al. (2004). We present subsequently how we accomplish the requirements stated in the guidelines.

I) Design as an artifact: the main outputs of our research are three novel evaluation techniques (i.e., *build-method cell* in the March and Smith (1995) classification). Moreover, we develop a conceptual framework for the evaluation of quality of use of VDM tools (*build-model cell*).

II) Problem relevance: as we have pointed out previously, the visualization evaluation problem is relevant and important.

III) Design evaluation: we perform evaluations of the developed artifacts. We apply the proposed evaluation techniques (questionnaires and procedures) to the evaluation of several visualizations. We also discuss and analyze different aspects of the proposed evaluation techniques.

However, more studies are necessary to investigate their possibilities and limitations.

IV) Research contribution: Hevner et al. (2004) as well as Kasanen et al. (1993) point out that research contributions are assessed for their implications to practice and their implications to theory. The main contribution to theory is given by the novel evaluation techniques of visualizations, while the main contributions to practice are the evaluations results that show the strengths and weaknesses of the visualizations on the test datasets. In Chapter 9, we discuss in detail both types of implications of our research.

V) Research rigor: the construction of the techniques is based on the phases in the research process described in (Kasanen et al. 1993). Our evaluation techniques are based on the principles of simulation, inspection and inquiry evaluation methods.

VI) Design as a search process: we use appropriate concepts (i.e., effectiveness, quality of use) and models in the evaluation. Moreover, we improve iteratively our evaluation techniques/models in order to accomplish the desired goals.

VII) Communication of research: we have communicated the research by submitting and presenting the results to conferences in the fields of Information Systems, Data Mining, Visualization, and Evaluation of Information Systems.

2.2. Descriptive research

Descriptive research aims “to measure a phenomenon – to find out how widespread it is, or how it varies across a given population” (Buckinngham and Saunders 2004, p. 44). In descriptive studies (or non-experimental) the researcher does not manipulate variables actively, but observes events as they occur, with no deliberate interference (Spata 2003, p. 11). The outputs of these studies are descriptions of phenomena or behaviors, and not analyses of the causes that determined them.

In order to demonstrate the utility of our evaluation techniques, we have performed evaluation studies of various visualization techniques. Our evaluation studies can be characterized as descriptive research as we have focused on applying the evaluation methods in order to describe the effectiveness and quality of use of the VDM tools, without investigating the causes of the evaluation results.

In our case, the “phenomenon” under investigation in an evaluation study is the *effectiveness* of the visualization techniques and/or the *quality of use* of a VDM tool. More specifically, we have *measured* the effectiveness of the visualization techniques for VDM tasks (Chapters 5-7) and have gathered information about the attitudes, opinions and performance of users with the SOM tools (Chapter 8).

First, a *simulation* evaluation method (Ivory and Hearst 2001) is employed to gather data about the effectiveness of projection-based visualization techniques in preserving the original structure of the data. We apply different visualization techniques on different real and artificial benchmark datasets and evaluate their effectiveness in preserving the original structure of the data (Chapter 5).

Second, an *inspection* evaluation method (Ivory and Hearst 2001) is used to gather data about the capabilities of different visualization techniques of solving data mining tasks related to a financial benchmarking problem. We examine whether the visualization techniques are capable of revealing interesting patterns in the data (Chapter 6).

Third, the *survey method* (Buckingham and Saunders 2004) is employed to gather data, based on the *questionnaire* technique. The idea is to select a sample of representative actual or potential users of the tool being investigated and to ask them to rate different aspects of the tool. The equivalent of the survey method in UE literature is the *inquiry* or *query technique* (Dix et al. 1998; Ivory and Hearst 2001) (Chapters 7 and 8).

2.3. Theoretical research

The purpose of *theoretical (conceptual-analytical)* research is to collect, integrate and systematize previous research results (Järvinen 2001, p. 17). The result is a theory, model, framework, or taxonomy. Examples of theoretical approaches in information visualization are the reference model for visualization in (Card et al. 1999) and the unified theory of distortion techniques in (Leung and Apperley 1994).

We integrate and systematize previous research results regarding usability evaluation of IS. In particular, we review five international standards addressing usability: ISO/IEC 9126-1, ISO/IEC 14598-1, ISO 9241-11, ISO 13407, and ISO 18529. The purpose is to create a framework that describes the process of the UE of IS, based on the knowledge acquired from the standards. We analyze comparatively the standards and formulate a comprehensive framework of UE process. The resultant framework highlights the main activities and outputs in the UE, and the extent to which the reviewed standards address these activities. Moreover, it points out to what extent the standards address the UE in different stages of the system's life-cycle. The framework can be applied to any information system and its components, in different phases of its life-cycle, therefore, to VDM tools too (Chapter 4 and Paper 1).

3. Information visualization techniques and visual data mining: key concepts

3.1. Background

Given the amounts of data available, there are two main needs of business users: to access the data and to make sense of them. Traditional approaches to fulfill these needs are data management systems (Elmasri and Navathe 2000) and business analytics, including here techniques from statistics, machine learning and data mining techniques (Kohavi et al. 2002). It is often argued that the use of graphical representation of the information is preferable to the use of tabular or textual reports for conveying numerical data to business users. The reasons are the higher confidence that users have in findings shown via visual representation, the faster time of data exploration, and the more intuitive representations (Keim 1996; Kohavi et al. 2002). Thus, information visualization appears as a promising technology that can be employed to enable the users' access to large amounts of data stored in databases. Moreover, it provides techniques to graphically represent the results obtained by applying data mining techniques.

Information visualization is a rather newly emerged field that is concerned with graphically representing abstract data on a computer-supported and interactive medium, in order to amplify cognition (that is, the acquisition and use of knowledge) (Card et al. 1999). Another definition highlights the fact that information visualization involves constructing graphical interfaces that enable humans to understand complex datasets (Fayyad and Grinstein 2002). Catarci and Cruz (1996) describe information visualization as being concerned with "visual mechanisms to communicate clearly to the user the structure of information and improve on the cost of access to large data repositories."

Visual data mining is a function supported by information visualization. It refers to the use of information visualization techniques in order to make sense of large, complex (for example, multidimensional) datasets. In the following, we describe the key concepts of information visualization and visual data mining that we use further in this thesis.

3.2. Classifications of visualization techniques

Visualization techniques support the graphical data representation and the user interaction with the data and the visualization system. The interaction with the data regards the selection of the data (cases and variables) to be visualized. The interaction with the visualization refers to selecting the ways in which the data are represented (that is, changing visualization parameters) (Card et al. 1999; Grinstein and Ward 2002).

In order to characterize visualizations, people create different taxonomies or classifications. In the following, we present some of the existing classifications, which are relevant to the multidimensional visualizations on which we focus. Keim (2002) classifies the information visualization techniques according to three criteria (Table 5): the *data type to be visualized* (1), the *visualization technique* (2), and the *type of interaction and distortion technique* (3).

Other classifications distinguish between *table visualizations* and other visualizations (Hoffman 1999), or focus on the *tasks* (exploratory, confirmatory, presentation) that the techniques support (Grinstein and Ward 2002).

Table 5. Keim (2002)'s classification of the information visualization techniques

Criteria	*Specific techniques*
Data type to be visualized	One-dimensional data Two-dimensional data Multidimensional data Text and hypertext Hierarchies and graphs Algorithms and software
Visualization technique	Standard 2D/3D displays Geometrically transformed displays Icon-based displays Dense-pixel displays Stacked displays
Interaction and distortion technique	Interactive projection Interactive filtering Interactive zooming Interactive distortion Interactive linking and brushing

3.2.1. Data type to be visualized

The **data type** refers to the complexity of the data to be visualized (Shneiderman 1996). The data may be distinguished by the number of dimensions that it has:

one-dimensional, *two-dimensional*, *three-dimensional* and *multidimensional*. In addition, the data may be of the type *text/hypertext*, *hierarchies/graphs*, or *software/algorithms*.

In this thesis, we are concerned with techniques that support the visualization of **multidimensional data**, that is, data with more than three variables, e.g. tables from relational databases.

3.2.2. Visualization techniques

The **visualization techniques** are concerned with the actual graphical representation of the data. Keim (2002) identifies five categories of techniques: *standard 2D/3D* displays, *geometrically transformed* displays, *icon-based* displays, *dense pixel* displays, and *stacked* displays. The techniques differ with respect to the ways in which they graphically represent the data dimensions and arrange the data on the screen (Keim 2001).

In this thesis, we address the evaluation of the techniques presented in Table 6.

Table 6. Visualization techniques addressed in the thesis

Type of technique	*Name of technique*
Variations of standard 2D displays	Multiple Line Graphs, Permutation Matrix, Survey Plot
Geometrically transformed displays	SOM
	Scatter Plot Matrix, Parallel Coordinates
	Sammon's Mapping, PCA
	Radviz, Star Coordinates
Stacked displays	Treemap

- **Standard 2D/3D displays**: This class comprises the most popular techniques, which are very effective for presenting one-, two- and three-dimensional data on a standard 2D or 3D display. Examples of techniques are line graphs, histograms, pie charts, doughnut charts, box plots, x-y(-z) plots (or scatter plots), bar and column charts, radar charts, area graphs, stacked bar and columns graphs (see Bertin 1981; Soukup and Davidson 2002). *Variations of the standard 2D/3D displays* can be employed for representing multidimensional data, for example, Multiple Line Graphs (Bertin 1981), Survey Plots (Demsar et al. 2004) and Permutation Matrix (Bertin 1981).

- **Geometrically transformed displays**: These techniques aim at finding "interesting" transformations of multidimensional datasets (Keim 2002).

They use geometric transformations and projections to produce useful visualizations (Keim 2001). Included are techniques from exploratory statistics (such as Scatter Plot Matrix (Cleveland 1993), Principal Components Analysis (Duda et al. 2000), Dendrograms (Sharma 1995)). Other techniques are Sammon's Mapping (Sammon 1969), Parallel Coordinates (Inselberg 1985), Radial Coordinate Visualization (Hoffman 1999), Self-Organizing Map (Kohonen 2001), Star Coordinates (Kandogan 2000), etc.

- **Stacked displays**: Stacked displays are representations of data that are partitioned in a hierarchical fashion. When the data are multidimensional, the data dimensions to be used in building the hierarchy have to be selected carefully (Keim 2001; 2002). An example of technique in this category is Treemap (Johnson and Shneiderman 1991; Shneiderman 1992).

3.2.3. Interaction and distortion techniques

Interaction and distortion techniques enable the transformation of the data and visualizations according to the user's exploration goals. Common interaction operations include filtering of the data, zooming, and linking multiple visualizations. In this way, the user makes dynamic changes in a visualization. The user can also relate and combine multiple visualizations. Distortion techniques provide means for focusing on particular data items while preserving an overview of the data (Keim 2002).

3.2.4. Table visualizations

Hoffman (1999) distinguishes between *table visualizations* and other types of visualizations. Table visualizations graphically represent data which are structured in a two-dimensional table (typically the rows represent the cases or objects, and the columns represent the variables or dimensions). The particularity of the data that are represented in table visualizations is that they do not explicitly contain internal hierarchical structure or links. The order of the table can sometimes be considered another dimension. If the table represents points in some other sequence, such as a time series, that information should be represented as another column (Hoffman and Grinstein 2002). Examples of techniques in this category include Parallel Coordinates, Survey Plots, Radial Coordinate Visualization.

3.2.5. Tasks supported by visualization techniques

Visualization techniques can support three categories of user tasks (Grinstein and Ward 2002):

- *Explore data*: the user does not necessarily have *a priori* knowledge about the data, nor precise exploration goals. The user looks for meaningful structure, patterns or trends, and hence for formulating a relevant hypothesis.
- *Confirm a hypothesis*: the user looks for certain patterns or structure in data (the user's goal is to verify a hypothesis). Analytic tools may be needed for confirming or refuting the hypothesis.
- *Produce presentation:* the user has a validated hypothesis and his/her goal is to communicate the knowledge to other parties. The focus is on refining the visualization to optimize the presentation.

To summarize, in this thesis, we are concerned with evaluating *table visualizations of multidimensional data*. The tasks for which the visualizations are employed and evaluated are of exploratory nature. We focus mainly on the static presentations of the data (Chapters 5-7), and we address the interaction with visual data mining tools in Chapter 8. We choose techniques belonging to variations of standard 2D displays, geometrically transformed displays and stacked displays.

3.3. Visual data mining

Data mining is the process of extracting knowledge from very large amounts of data. The discovered knowledge takes the form of *patterns* found within the data, patterns that must be interesting to the user (valid, novel, potentially useful and understandable) (Fayyad et al. 1996; Witten and Frank 2000). This process is also known as the *knowledge discovery in databases* (KDD).

KDD consists of many steps, including problem definition, data selection and preparation, data transformation, data mining, interpretation and communication of the results. In the context of KDD, data mining (DM) is a step in the process, which is responsible for automatically extracting patterns from data (by means of an algorithm). One of the many applications of DM is in the field of business. Business analytics (Kohavi et al. 2002) employ DM techniques to gain insight into the business data the companies have available. Examples of patterns obtained via DM are clusters, trends, classes, outliers, and relationships found in the data. The totality of patterns of the same type found in a dataset by applying a DM technique represents a data-mining *model* (for example, clustering, classification, association rules, etc.).

Keim (2001) points out that, in order to be effective, DM has to have a human in the data exploration process. In this way, the human abilities (flexibility, creativity and general knowledge) are combined with computer performance (storage capacity and computational power). The solution to involve the human directly in the process of data exploration is called by Keim *visual data mining* (VDM) or *visual data exploration*. By this process, the user can detect patterns in large multidimensional datasets. In the classification of levels of use supported by information visualization, VDM tools belong to the class of *visual knowledge tools*. The type of information processed is a dataset and the purpose of using information visualization is to get insight into the data and improve knowledge discovering (Card et al. 1999).

The information visualization techniques used for VDM can support two user needs. The first one refers to the *data exploration* by which the user is involved in direct exploration and visualization of large amounts of high-dimensional data. The idea is to represent the data in a graphical and easy to comprehend format in order to give users insight into the data. In addition, the users are enabled to interact with the data. The visual data exploration is typically performed when the user does not have much knowledge about the data nor established exploration goals. The second user need refers to the direct manipulation and visualization of the *data models* obtained by applying business analytics such as data mining techniques. Users can control the process of obtaining and visualizing the data models generated by a data mining technique. Thus, the generated models are represented in a visual form and the user has the possibility to modify the model's or data mining technique's parameters and see the effects of his/her modifications directly on the visualization.

3.4. Data mining tasks

Soukup and Davidson (2002) point out that in order to apply visualization techniques to business problems, one has to transcribe the business problems into business questions and further into data mining tasks. Bertin (1981) provides similar guidelines in the model of decision-making process using graphical information-processing techniques.

Soukup and Davidson (2002) highlight the following DM tasks: *classification*, *estimation*, *association groupings*, *clustering or segmentation*, and *prediction*. Other common DM tasks are *multivariate outliers' detection* (often associated with clustering tasks, Soukup and Davidson 2002), *dependency analysis or modeling*, *change and deviation detection*, and *summarization* (Fayyad et al. 1996).

In the following, we describe the data mining tasks for which we conduct the evaluation.

- Clustering – the aim is to determine subsets (groups) of objects in a dataset that are similar. It is based on the process of dividing a dataset into mutually exclusive groups, without relying on predefined classes (Soukup and Davidson 2002; Jain et al. 1999).
- Clusters description – the aim is to describe the groups obtained via clustering, in terms of the characteristics in each group (i.e., prototypes or representative patterns, e.g., centroids, mean values). Jain et al. (1999) refer to this activity as data abstraction.
- Outlier detection – the aim is to find extreme values that can reflect anomalies or errors in data. In a distribution, the outliers are those data points that lie outside the normal range of data (i.e., univariate outliers). In a relationship between two variables, the outliers are those data points that are not part of the relationship (Hartwig and Dearing 1990) (i.e., bivariate outliers). Often outliers are associated with clustering tasks, a situation in which the outliers are identified as objects that do not belong strongly to any one cluster (Soukup and Davidson 2002) (i.e., multivariate outliers).
- Dependency analysis or modeling – the aim is to find models that describe significant dependencies or relationships between variables (Fayyad et al. 1996). Fayyad et al. point out the existence of two types of dependency modeling, structural and quantitative. Structural models specify (usually in graphical form) the variables that are locally dependent of each other. The quantitative models specify the strength of the dependencies on a numerical scale.
- Classification – the aim is to map (classify) an object into one of the predefined classes. This mapping is based on constructing a classifier from a data set where the classes are known (Fayyad et al. 1996; Soukup and Davidson 2002).
- Class description – it is similar to cluster description, but here the aim is to describe the objects belonging to one *predefined class* in terms of the values of the variables that define the data.
- Comparison of data items – the aim is to compare two or more objects in terms of the variables that define the data. We consider this task as a sub-type of the Change and Deviation Detection task (Fayyad et al. 1996), where the aim is to detect significant changes in the data from previously measured or normative values.

4. Evaluation of visualization techniques

In this chapter, we present the related work on evaluation of visualization techniques. Firstly, we focus on usability evaluation (UE) of information systems, in general. Nielsen (1993) and Dix et al. (1998) emphasize the importance of ensuring a high level of usability of any interactive system. Information visualization is also an interactive process in which users interact with the visualization system in order to acquire and use knowledge. Thus, in the design of information visualization systems as of any other interactive system, it is essential to ensure that the embedded techniques are usable.

Plaisant (2004) acknowledges that UE represents one way to evaluate the visualization techniques and systems. UE is relevant for our research problems too, because we are interested in evaluating the *effectiveness* of visualization techniques in data mining tasks, effectiveness being one characteristic of usability (ISO 9241-11[7]). We also are interested in evaluating *quality of use* of VDM tools.

Secondly, we look at the studies regarding the evaluation of visualization techniques from the point of view of quality characteristics, types of evaluation approaches, datasets and tasks used in evaluation. In the end of the chapter, we provide a classification of the visualization evaluation studies in terms of the type of evaluation method employed.

4.1. Usability evaluation of information systems

4.1.1. Definition of usability and usability evaluation

Usability of a system is often understood as ease of use or user friendliness. Many definitions that try to capture the essence of usability exist, but two of the most used definitions are found in the ISO/IEC 9126-1[8] and ISO 9241-11 standards. ISO/IEC 9126-1 defines usability as being "*the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions*". ISO 9241-11 defines usability as being

---

[7] ISO 9241-11 (ISO 1998) Ergonomic requirements for office work with visual display terminals. Part 11: Guidance on usability.

[8] ISO/IEC 9126-1 (ISO 2000b) Information Technology – Software product quality. Part 1: Quality model.

"*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*." This latter view of usability is encountered in ISO/IEC 9126-1 under the name of "*quality in use*" (i.e., *quality of use* in (Bevan 1995)).

Usability evaluation (UE) refers to the process of planning and conducting the measuring of usability attributes of the user interface and identifying specific problems (Ivory and Hearst 2001). UE should be done throughout the design life-cycle and planned as providing results that can be used for improving the design (Dix et al. 1998).

As stated in the definition, before proceeding to UE, the evaluator has to specify the usability aspects (characteristics and attributes) that have to be assessed. *Usability models* (e.g., Nielsen 1993; Fenton and Pfleeger 1997; Dix et al. 1998; ISO/IEC 9126-1) typically provide those usability characteristics and attributes. For example, Nielsen (1993) uses the following *characteristics* to define usability: *learnability, efficiency, memorability, error rate* and *satisfaction*. ISO/IEC 9126-1 identifies other usability characteristics, some of them being similar to the ones of Nielsen's model: *understandability, learnability, operability, attractiveness,* and *usability compliance*. Moreover, the ISO/IEC 9126-1 defines a model for quality in use, similar to the one in ISO 9241-11 for usability. The characteristics of quality in use (usability in ISO 9241-11) are *effectiveness, productivity (efficiency), safety,* and *satisfaction*.

Attributes are lower-level characteristics that can be measured directly, and they are related to one of the higher-level characteristics (ISO/IEC 14598-1[9]). Regarding the measuring of the attributes, this step requires the use of usability *metrics* for each attribute selected from the model. An evaluation *method* is then employed, which helps the evaluator to systematically plan and conduct the measurements.

Effectiveness is defined in ISO/IEC 9126-1, ISO 9241-11 and ISO 13407[10] as being "accuracy and completeness with which users achieve specified goals". Accuracy and completeness are attributes of effectiveness, according to this definition. They can be measured by employing different metrics. For example, ISO/IEC 9126-4[11] provides three metrics for these attributes, namely, *task effectiveness*, *task completion*, and *error frequency*. Task effectiveness is used to

---

9 ISO/IEC 14598-1 (ISO 1999b) Information Technology – Software product evaluation. Part 1: General overview.
10 ISO 13407 (ISO 1999a) Human-centred design processes for interactive systems.
11 ISO/IEC 9126-4 (ISO 2004) Software Engineering – Product quality – Part 4: Quality in use metrics.

measure “the proportion of the goals of the task achieved correctly”. Task completion measures the proportion of the completed tasks. Error frequency measures the frequency of errors.

4.1.2. A model of the usability evaluation process

In Paper 1, we have reviewed five international standards in order to analyze the similarities and differences among them in defining usability and addressing the UE of information systems. The standards under review were ISO/IEC 9126-1, ISO/IEC 14598-1, ISO 9241-11, ISO 13407, and ISO 18529[12].

ISO/IEC 14598-1 provides a comprehensive model of the evaluation process, in general, which identifies four phases: *Define requirements, Specify evaluation, Plan evaluation,* and *Execute evaluation*. This model can be particularized for the UE process, so that requirements, attributes and metrics, and the evaluation method regard the assessment of usability.

One aspect of the conceptual framework presented in Paper 1 is the description of the UE process in terms of its activities and the standards that provide guidelines to perform these activities (Table 7).

In the first phase, the outputs of the first three activities are specific to the situation being investigated (the system and its components, the tasks under consideration, the intended users, etc.). The fourth activity, though also dependent on the particular situation being investigated, can be very much assisted by the acquired knowledge in the field of UE. There are many available usability models aiming at characterizing usability of information technology, information systems or any interactive system (e.g., Nielsen 1993; Fenton and Pfleeger 1997; Dix et al. 1998; ISO/IEC 9126-1; ISO 9241-1).

Similarly, the fifth activity in the UE process can be supported by the specialized literature that provides many metrics that measure the usability attributes of interactive systems (e.g., ISO/IEC 9126-2; ISO/IEC 9126-4; Seffah et al. 2006; Hornbæk 2006). However, the activities 6 and 7 depend very much on the evaluation goals and context of use of the systems under evaluation, aspects defined earlier in the process. The eighth activity, planning UE, involves the selection and specification of an evaluation method, method that can be developed for the current purpose or be chosen from the ones available (e.g., in Whitefield et al. 1991; Nielsen 1993; Coutaz 1994; Dix et al. 1998; Ivory and

---

[12] ISO 18529 (ISO 2000a) Ergonomics – Ergonomics of human-system interaction – Human-centred lifecycle process descriptions.

Hearst 2001). Finally, the activities 9, 10 and 11 are highly dependent on the situation under investigation, the steps previously performed and their outputs.

Table 7. Activities in UE process

Activities	*Standards*
Phase I: Define usability requirements	
1. Distinguish between system under development and system in use	ISO 9241-11, ISO 18529
2. Specify purpose of evaluation and evaluation target (user, technology, or system and phase in product/system development life-cycle)	ISO/IEC 9126-1, ISO/IEC 14598-1, ISO 9241-11, ISO 13407, ISO 18529
3. Specify context of use of the information system	ISO 9241-11, ISO 13407, ISO 18529
4. Specify usability and quality in use characteristics, sub-characteristics and attributes	ISO/IEC 9126-1, ISO/IEC 14598-1
Phase II: Specify evaluation	
5. Select or create validated metrics to be used in measurement of the system usability attributes	ISO/IEC 14598-1, ISO 9241-11
6. Specify rating levels for each metric	ISO/IEC 14598-1
7. Specify assessment criteria	ISO/IEC 14598-1
Phase III: Plan usability evaluation	
8. Select and specify an appropriate usability evaluation method	ISO 13407, ISO 18529, ISO/IEC 14598-1
Phase IV: Execute usability evaluation	
9. Measure usability attributes	ISO/IEC 14598-1, ISO 9241-11
10. Map measured values to rating levels	ISO/IEC 14598-1, ISO 9241-11
11. Assess result	ISO/IEC 14598-1, ISO 9241-11

The framework in Table 7 is useful also in the planning and execution of UE of information visualization techniques/systems, and in the description of the evaluation process.

4.1.3. Types of usability evaluation methods

Dix et al. (1998) highlight eight factors that can be used to characterize evaluation *methods* and practices. The first factor is the *phase in the system's life-cycle* at which the evaluation is carried out. According to this factor, the evaluation can take place in the *design* phase (e.g., analytic methods, literature review-based and heuristic methods, model-based methods) and in the *implementation* phase (e.g., experimental methods, observational methods, query methods). Evaluation of implementation requires that a physical artifact exists (any form of artifact ranging from a paper mock-up to a full implementation).

The second factor is the *evaluation style,* which differentiates between *field* and *laboratory* studies. Laboratory studies enable controlled experiments and observations but lose some of the real aspects of the problem. Field studies capture the real interaction with the system, but have limited control over the user's actions.

The third factor is the *objectivity or subjectivity* of the method. Objective evaluation refers to the situation when the results are more or less the same, regardless of who is carrying out the evaluation. Subjective evaluation produces assessments that may vary according to who the evaluators are. The results of subjective evaluation depend, therefore, on the interpretation of the evaluator. The danger of inducing bias into the results can be overcome by involving many evaluators.

The fourth factor is the *type of measures* involved (*quantitative or qualitative*). Quantitative measurements, usually used in objective evaluations, are numeric and can be analyzed with statistical techniques. Qualitative measurements are non-numeric and more difficult to analyze. However, some methods, e.g., in subjective evaluation, capture qualitative data but they map these data into quantitative measures.

The fifth factor is the *information provided* as a result of an evaluation. This can vary from *low-level* information (e.g., assessment of readability of fonts, colors, icons etc.) to *high-level* information (e.g., what is the usability, learnability, or user satisfaction of the system/technique?).

The sixth factor is the *immediacy of the results*. The immediacy of the results regards the time of capturing the usability data (e.g., during interaction, or after interaction).

The seventh factor is the *level of interference* or *intrusiveness* involved by the method. Intrusiveness refers to the extent to which the evaluation method influences the evaluator's interaction with the tool and therefore the assessment process and results.

The eighth factor is represented by the *resources* required for the evaluation. Resources can be classified into equipment, time, money, users, expertise of the evaluator, and context.

Whitefield et al. (1991) provide another framework of evaluation methods, useful in selecting the evaluation method appropriate for a study. Their classification characterizes the methods based on how the *computer* and *user* are present in the evaluation process. Here the computer represents the system or one of its components under evaluation (i.e., the *artifact* under evaluation). The presence of the artifact and of the user can be *real* or *representational*. The result is a four-class framework:

- *Analytic methods* (both user and artifact are representational),
- *User reports* (user is real and the artifact is representational),
- *Specialist reports* (user is representational and the artifact is real),
- *Observational methods* (both user and artifact are real).

Real presence of the computer (artifact under evaluation) means the physical presence of computer (artifact under evaluation) or an "approximation" of it (e.g., implemented systems, prototypes, simulations) in the evaluation process, during assessment. Representational presence of the artifact is given by specifications or notational models (e.g., mental representation of the system under evaluation in questionnaires and survey). In addition, real presence of the user means that actual users or "approximations" are involved (e.g., target users or students). Representational presence of the user means that models or descriptions of users are used.

Ivory and Hearst (2001) classify the UE methods into *testing*, *inspection*, *inquiry*, *analytical modeling* and *simulation*. In usability testing, an evaluator observes users interacting with an interface in completing tasks, in order to identify problems or measure usability attributes such as user performance or time to complete a task. In usability inspection, an evaluator uses a set of criteria or heuristics to identify problems with the interface. In usability inquiry, users assess an interface by answering questions in surveys or interviews. Analytical modeling is the use of user' and interface's models to obtain predictions about usability attributes or problems. Simulation is the use of user' and interface's models that mimic a user interacting with an interface in order to obtain usability data (simulated activities, errors, and other quantitative measures).

In this thesis, we propose three evaluation techniques of information visualizations corresponding to the three research questions (RQ1, RQ2, and RQ3 in Chapter 1). We did not intend to cover all aspects of usability evaluation of a visualization system or technique, but to evaluate the *effectiveness* of different visualization *techniques* in *preserving the original structure of the data* (RQ1) and in different *data mining tasks* (RQ2). Moreover, we develop a framework of *quality of use* and propose an inquiry technique to evaluate different attributes of quality of use of VDM tools (RQ3).

With respect to the UE process model, our research is concerned with assisting developers/evaluators of VDM tools at the following steps of the UE process:

- (4) – Specify usability/quality-of-use characteristics and attributes,
- (5) – Select or create validated metrics for the measurement of the attributes,

- (8) – Select and specify an appropriate UE method,
- (9) – Measure usability attributes.

4.1.4. SUMI and EUCS usability models

The Software Usability Measurement Inventory (SUMI, Kirakowski 1994) and End-User Computing Satisfaction (EUCS, Doll and Torkzadeh 1988) represent two popular models of usability. Both models are rigorously tested. As methods, they demonstrated their effectiveness for measuring usability. We use these models as a basis for the quality-of-use framework of VDM tools (Chapter 8).

SUMI is recognized as a standardized usability measurement instrument (Kirakowski 1994; Fenton and Pfleeger 1997, p. 355). It is used to measure usability (quality of use) of a software product by employing an inquiry UE method based on the questionnaire technique. SUMI model consists of five characteristics of usability: *Efficiency, Affect, Helpfulness, Control, and Learnability*. Each characteristic is measured by a set of 10 questions, whose answers are recorded on a three-point Likert scale (1: Agree; 2: Undecided; and 3: Disagree).

The Efficiency characteristic refers to "the degree to which users feel that the software assists them in their work". The Affect characteristic regards the "user's general emotional reaction to the software", and it could also be called "Likeability". The Helpfulness component refers to "the degree to which the software is self-explanatory", as well as to the quality of documentation and help functions. The Control component regards "the extent to which the user feels in control of the software, as opposed to being controlled by the software, when carrying out the task". Finally, Learnability refers to "the speed and facility with which the user feels that they have been able to master the system", and also to how fast they learn to use new features of the system.

EUCS provides a model of user satisfaction with five characteristics: *Content, Accuracy, Format, Ease of Use, and Timeliness*. The measurement of these characteristics is also carried out by employing the questionnaire technique. The scales of measurement are five-point scales (1: Non-existent; 2: Poor; 3: Fair; 4: Good; and 5: Excellent). The data are recorded as answers to 12 questions, such as Does the system provide the precise information you need? (Content); Is the system accurate? (Accuracy); Is the information clear? (Format); Is the system user friendly? (Ease of use); or Do you get the information you need in time? (Timeliness).

4.2. Evaluation of information visualization techniques

4.2.1. Excellence and integrity of visual displays

The early work in information visualization was concerned with defining quality of graphical displays and developing theories of design of graphical displays (e.g., Bertin 1981; Tufte 1983). Tufte (1983) examines a series of good and bad examples of visual displays and derives the *principles of excellence and integrity* of data graphics. Starting from these principles, we derive some of the attributes of good visualizations, which we include in the quality-of-use framework (Chapter 8, Paper 6).

More precisely, for deriving attributes of the "data display" and "reporting functions" characteristics, we use the requirements specified by Tufte (1983) regarding the excellence (p. 13 & 51), integrity (p. 77), and aesthetics (p. 177) of visual displays. According to Tufte, good displays should:

- *Show the data* with clarity, precision and efficiency.
- Make the viewer *think about the substance of the data*, rather than about the visualization technique (design and computational issues).
- Encourage the eye to *make comparisons* between data.
- Present *multivariate data*.
- Use clear, detailed, and thorough *labeling* in order to avoid graphical distortion and ambiguity.
- Show *data variation*, not design variation.
- Have a properly chosen *format* and *design*.
- Avoid content-free *decoration* (i.e., use of colors and other graphical attributes).

4.2.2. Other quality characteristics of visualizations

Bertin (1981) emphasizes that the aim of visualization is to provide a *higher level of information* from a dataset. The higher the level of information provided by a visualization, the more useful the information is for answering interesting questions. The author distinguishes between three levels of information:

- Elementary: the data points.
- Intermediate: the data points are grouped based on one variable.
- Overall: the data points are grouped based on all variables.

The idea is that the more variables are involved in grouping the data, the more complete view of the data is obtained. Based on these concepts, Bertin defines the *visual efficacy* of a visualization technique as being "the level of question which receives an immediate response".

Mackinlay (1986) defines two characteristics for visualizations, which he uses in developing a system that automatically design visual representations of data. The two characteristics are *expressiveness* and *effectiveness*. Expressiveness refers to the capability of the visual representation to express the desired information (ideally, "all the information and only the information"). Effectiveness refers to the capability of the visualization to exploit the "output medium and the human visual system". Effectiveness is viewed as enabling the user to read, understand and interpret the display easily, accurately, quickly, etc., and thus it depends not only on the graphical design but also on the capabilities of the viewer. These two characteristics of visualizations are adopted also by Card et al. (1999). They define effectiveness as the capability of the visualization to be perceived well by the human: "it is faster to interpret, can covey more distinctions, or leads to fewer errors than some other mapping".

Regarding the evaluation of the interactive capabilities of the visualization, concepts such as *mobility of image* (Bertin 1981) and *design variation* (Mackinlay 1986) can be used. They describe the capabilities of the system to enable the manipulation of the graphical elements, and the use of multiple displays so that all the desired information is revealed. Moreover, Card et al. (1994) define a function, named "the cost-of-knowledge characteristic function" in order to measure the *information access* from dynamic (interactive) displays, in the context of information access systems, such as information retrieval systems.

4.2.3. Classification and examples of evaluation studies of visualization techniques

4.2.3.1. Objective and subjective evaluation studies

One important classification of evaluation methods and studies of visualization techniques is based on the dichotomy between *objective* and *subjective* approaches. Both approaches are important in visualization evaluation, since the user interpretation of the visualization is crucial for knowledge discovery and decision-making (Keim and Kriegel 1996; Keim 1996).

Objective evaluation produces more or less the same results independent on who the evaluator is (Dix et al. 1998). Typically, it involves the computation of quantitative measures of different attributes of visualizations. Among the attributes evaluated in objective approaches are the number of dimensions and data points displayed, data point overlap, but also performance (time, access cost) and effectiveness. Examples of measures and methods used in evaluating such attributes can be found in (Mackinlay 1986; Card et al. 1994; Keim and Kriegel 1996; Hoffman 1999). Other objective studies measure the performance of the visualization in different tasks, for example, in classification (Liu and Salvendy 2007), in prediction (Dull and Tegarden 1999), or in information retrieval (Sutcliffe et al. 2000).

Subjective evaluation produces results that vary according to who the evaluators are (Dix et al. 1998). The results of subjective evaluation depend, therefore, on the interpretation of the evaluator. One example of conducting subjective evaluation by *visually inspecting the output* of the visualizations is in (Hoffman 1999) or in (Grinstein et al. 2002). Hoffman evaluated five visualization techniques (e.g., Radial Coordinate Visualization, Parallel Coordinates, Scatter Plot Matrix, Survey Plot) on 11 datasets, and different tasks such as clustering, outlier detection and rule discovery for classification of the data. The assessment was done by a single experienced user and the author pointed out that many users should be involved, with different levels of experience. A similar approach is used in (Keim and Kriegel 1996), where the authors evaluate six techniques (including Parallel Coordinates) on different datasets (real and artificial) and different tasks (outlier detection, clustering and functional dependencies). Another typical technique used in subjective evaluation is the *questionnaire*, by which data are collected from different users in the form of user ratings (e.g., Risden et al. 2000; Pillat et al. 2005; Liu and Salvendy 2007). Many studies include both objective and subjective approaches (e.g., Liu and Salvendy 2007; Keim and Kriegel 1996).

In user evaluation studies, the *number of users* involved varies from **five** (Pillat et al. 2005; Ståhl et al. 2006), to **15** (Risden et al. 2000), **19** (Ward and Theroux 1997), **20** (Liu and Salvendy 2007; Shneiderman 1996), to **124** (Dull and Tegarden 1999).

4.2.3.2. Datasets used in evaluation

Another practice in evaluation of visualization is to examine the effectiveness of the techniques on different *datasets* (real and/or artificial) and for different *tasks*. Some researchers stress the necessity to evaluate the effectiveness of visualizations on **real and challenging datasets** and problems (e.g., Inselberg 1997). Hoffman (1999) and Grinstein et al. (2002) use **real benchmark datasets** available at UCI Machine Learning Repository (Newman et al. 1998).

On the other hand, Keim and Kriegel (1994, 1996) highlight the importance of **artificial datasets** in evaluation. Keim et al. (1994) developed a conceptual model of generating artificial benchmark datasets based on criteria such as number of data dimensions, number of data points, type of data, structure of data, etc.

4.2.3.3. Tasks evaluated

There are studies that focus on evaluating **problem-domain tasks** such as visual data mining tasks or information retrieval tasks, depending on the functionality of the systems or tools being evaluated. The focus there is on evaluating the *quality of the output* which is provided by the visualization techniques. For example, tasks like clustering, rule discovery, outliers detection, correlation and functional dependencies are commonly evaluated (e.g., in Hoffman 1999; Grinstein et al. 2002; Keim 1996). Other examples of evaluation studies are Sutcliffe et al. 2000 – in information retrieval, Liu and Salvendy (2007) - in classification, Dull and Tegarden (1998) – in prediction tasks.

Other studies focus on evaluating **interaction tasks**, such as selecting the data or saving the results, etc. The focus there is on the *efficiency, ease of learning, or ease of using* the techniques to arrive at the desired output (e.g., Risden et al. 2000).

Regarding the evaluation of **projection techniques' outputs**, Seo (2005) mentions the work of Tukey and Tukey (1985), which highlights the necessity to evaluate "the relative interest of different scatterplots, or the relative importance of showing them and sort out such scatterplots for human analysis". Tukey and Tukey (1985) and Seo (2005) propose the use of metrics based on *correlation coefficients* calculated to the data dimensions, in order to measure and improve the quality of visualization outputs. Seo remarks also on the work of Guo (2003) and Guo et al. (2003), concerned with the evaluation of the projection techniques in clustering tasks based on the *maximum conditional entropy*. The approach in Guo et al. (2003) is different from our approach in the following way. First, their problem is concerned with evaluating a projection with the purpose of selecting the best subset of dimensions that can be then used to determine clusters in the data. This problem is typically found under the name *subspace clustering* (Agrawal et al. 1998; Parsons et al. 2004). Second, they use *cluster tendency measures* (Theodoridis and Koutroumbas 1999; Milligan 1996) in particular, *conditional entropy* (Pyle 1999, p. 417).

Chen and Liu (2004) provide a framework and a visual rendering system "that allows the user to be involved into the clustering process via interactive visualization." They propose that an interactive "visual validation" is better than the use of statistical *clustering validation measures*, such as root-mean-square-standard-deviation, R-squared, and S_Dbw (Sharma 1995; Halkidi et al. 2002 a,b). However, their focus was to provide means to the user to "refine" the clustering solution obtained by applying an algorithm (e.g., the *K-means* technique (MacQueen 1967)).

4.2.3.4. Types of evaluation methods

Based on the classification of evaluation methods in (Ivory and Hearst 2001), we identified the following practices in the visualization evaluation:

- **Usability testing – Observational methods**, for example, Ståhl et al. (2006) evaluate a SOM-based tool with users in order to identify problems of design.
- **Usability testing – Controlled experiments** – These studies aim at comparing two or more design elements of the visual interface, or two or more visualization techniques, with respect to user performance and/or user satisfaction (e.g., Mackinlay 1986; Shneiderman 1994; Ward and Theroux 1997).
- **Inquiry** – these methods are typically employed together with other methods to obtain information otherwise not available (e.g., regarding user satisfaction or preference: Risden et al. 2000; Pillat et al. 20005; Ståhl et al. 2006).
- **Inspection** – for example, Hoffman (1999), Hoffman and Grinstein (2002), Keim and Kriegel (1996). These studies use benchmark datasets (real and artificial) and expert evaluators to examine subjectively the effectiveness of different visualization techniques for different tasks, by visually inspecting the output of the visualization techniques.
- **Simulation** – for example, Hoffman (1999) uses different quantitative measures to objectively evaluate different visualizations of different benchmark datasets. Another example of simulation evaluation is in (Keim 2000) in which the author uses different quantitative measures for evaluating the efficiency and effectiveness of visualization techniques.
- **Analytic modeling** – for example, Pirolli and Rao (1996) use the GOMS analysis (Goals, Operator, Methods, Selection Rules – Card et al. 1983) to measure time estimates for task performance. They analyze a number of exploratory data analysis tasks and compare two visualization tools. Another example is (Card et al. 1994), which uses a new method of analysis, namely "cost-of-knowledge characteristic function".

5. Evaluation of projection techniques using clustering validity measures

5.1. Research problem description

The research problem in this chapter is the evaluation of projection-based visualization techniques as to their *effectiveness* in preserving the original data structure (RQ1). We focus on characterizing the original data structure by *the distances between data points* and *the clustering structure*. We design and conduct an *objective evaluation* of different visualization techniques. Effectiveness is regarded in terms of *accuracy* (i.e., how accurately does a projection preserve the original data structure? and which projection is more effective in preserving the original data structure?). For this purpose, we use *clustering validity measures* (Milligan 1996; Theodoridis and Koutroumbas 1999) as quantitative measures for evaluating the effectiveness (accuracy) of the projection techniques in preserving the original structure of the data. We propose *new procedures* for calculating different clustering validity measures in order to use them in comparing projection techniques.

The evaluation problem is important in the following context. The high-dimensional data need to be mapped onto a lower-dimensional space in order to be visualized by users. A dimension-reduction technique (projection technique) can be used for this purpose. However, different projection techniques can produce very different representations of the data and a user may not know which of these representations reveal the most interesting and accurate facts about the data. In a clustering task, the user is obviously interested in detecting clusters in the data and obtaining an image of the data that closely represents the original data. In this chapter, we explore the use of clustering validity measures for evaluating and comparing different projections. The evaluation approach is based on the *simulation method*. The datasets used in evaluation are real and artificial (synthetic) benchmark datasets.

The following techniques are evaluated: Principal Components Analysis (PCA), Sammon's Mapping, Self-Organizing Map (SOM), Radial Coordinate Visualization (Radviz), and Star Coordinates. The research in this chapter is also presented in Papers 2 and 3.

The clustering validity measures are useful for evaluating the truthfulness or accuracy of a clustering solution which is obtained by applying clustering analysis techniques (Theodoridis and Koutroumbas 1999, Halkidi et al. 2002a,b). The clustering analysis is appropriate for the exploration of interrelationships among the data points and for uncovering the structure of the dataset (Jain et al. 1999). In principle, the relationships among data points that clustering techniques are evaluating are defined in terms of similarity or dissimilarity among data items. Therefore, the structure uncovered by clustering techniques consists of groups of homogenous or similar data items. However, a clustering solution obtained by applying a clustering technique does not always reveal a *real* partition of the data.

In Section 5.2, we present the projection techniques under evaluation. In Section 5.3, we briefly describe each measure. In Section 5.4, we propose the procedures of calculating the measures for evaluating the projection techniques. We apply the procedures on five different real and artificial benchmark datasets and present the results of the evaluations in Section 5.5. The results show that the clustering validity measures are good indicators of the effectiveness of the projection techniques in preserving the structure of a dataset (e.g., distances between data points and clustering).

5.2. Projection techniques

Projection techniques reduce the dimensionality of the data space (Kohonen 2001, p. 34). They transform the original high-dimensional data points into points with a smaller number of dimensions. When the projections are represented in a low-dimensional space such as 2D or 3D space, the projection technique is a possible tool for visualizing the data. The transformation of the data is done so that the information in the data (e.g., structure or variation of the data, etc.) is preserved as well as possible.

In the following we briefly describe and illustrate the five projection techniques evaluated in this chapter (PCA, Sammon's Mapping, Radviz, Star Coordinates and the SOM).

5.2.1. Illustrative dataset: Iris

We illustrate the visualization techniques on the Iris data (Newman et al. 1998). The data concern three species of flowers characterized by four attributes: petal length and width, and sepal length and width. The class variable is the type of flower: Iris-Setosa, Iris-Versicolor, and Iris-Virginica. Each class contains 50 flowers. The class Iris-Setosa is linearly separable from the other two classes, but Iris-Versicolor and Iris-Virginica classes are not linearly separable.

5.2.2. Principal Components Analysis (PCA)

PCA (Sharma 1995; Duda et al. 2000) is a classical statistical technique employing linear transformation of data in order to represent the high-dimensional data onto a lower-dimensional space. The transformation tries to preserve the variance of the original data as well as possible. The PCA technique creates new variables (called principal components), which are linear composites of the original variables and are uncorrelated amongst themselves. The maximum number of new variables that can be formed is equal to the number of original variables. The PCA representation is judged in terms of how well the lower-dimensional space represents the information contained in data, or, geometrically, how well this new low-dimensional data space can capture the original configuration of the data.

Figure 1 represents the Iris dataset plotted on the 2D space formed by the first two principal components. For obtaining the PCs we have used the *standardized*[13] data (i.e., data were *normalized* using the *variance* method (Vesanto et al. 2000)).

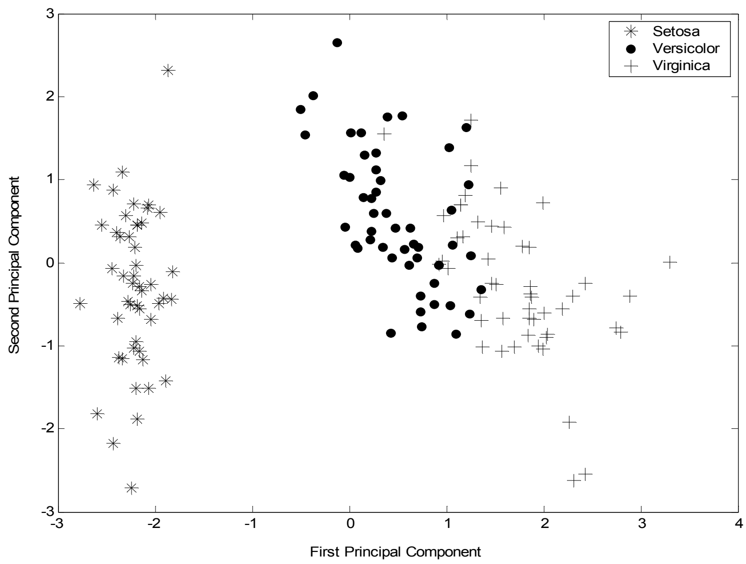


Figure 1. PCA of the Iris data

[13] Standardized data are recommended when it is not wanted that the relative variances of the variables affect the PCA result (Sharma 1995). Thus, by standardizing the data, the variance of each variable is the same, i.e., one. Standardized data are obtained by transforming each variable so that from each data value is subtracted the variable mean and the result is divided by the standard deviation of that variable (Vesanto et al. 2000).

5.2.3. Sammon's Mapping

The Sammon's Mapping (Sammon 1969) is a non-linear projection technique belonging to the class of multidimensional scaling techniques. It tries to match the pairwise distances of the lower-dimensional representations of the data items, with their original distances. Sammon's Mapping is useful for visualizing class distributions, especially the degree of their overlap (Kohonen 2001, p. 37).

Figure 2 illustrates the Sammon's Mapping of the Iris dataset. The data were *normalized* using the *variance* method. The Euclidean distance was used in the Sammon's Mapping algorithm.

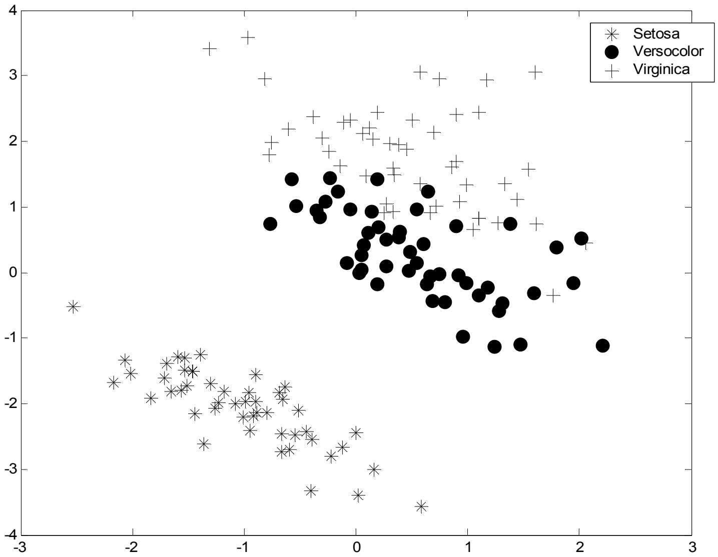


Figure 2. Sammon's Mapping of the Iris data

5.2.4. Radial Coordinate Visualization (Radviz)

The Radviz technique (Hoffman et al. 1997; Hoffman 1999) represents each n-dimensional data item as a point in a two-dimensional space. The points are located within a circle whose perimeter is divided into n equal arcs. The equally spaced points on the perimeter are called *anchorpoints* or *dimensional anchors* (Hoffman 1999; Hoffman et al. 1999). Each data dimension is associated with one anchorpoint. When the data are n-dimensional, each data point will be connected to *n* anchorpoints through *n* different springs. Each data point is then displayed at the position that produces a spring force sum of zero.

The values of each data dimension are normalized to the range [0, 1] [14]. If the data are left in the original range, then the variable with higher values than others will dominate the spring visualization. If all *n* coordinates have the same value (regardless of whether they are low or high), the data point lies exactly in the centre of the circle. If the point is a unit vector point, it lies exactly at the fixed point on the edge of the circle, where the spring for that dimension is fixed (Hoffman 1999).

Figure 3 shows the Radviz projection of the Iris dataset. Each variable was *normalized to the range [0, 1]*.

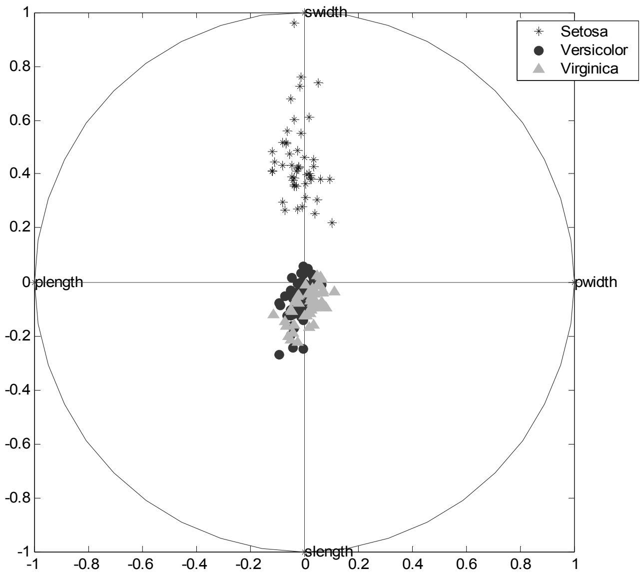


Figure 3. Radviz of the Iris data

5.2.5. Star Coordinates

The Star Coordinates technique (Kandogan 2000; 2001) maps *n*-dimensional data onto a two-dimensional space. The idea of Star Coordinates is to arrange the *n* coordinate axes on a two-dimensional plane, such that all axes share the

[14] This type of normalization transforms each variable by subtracting the minimum value and then dividing the result by the difference between maximum and minimum values of that variable (Vesanto et al. 2000).

same origin point, but they are not necessarily orthogonal to each other. The minimum data value on each dimension is mapped to the origin, and the maximum value is mapped to the other end of the coordinate axis. It is recommended that each variable is normalized to the range [0, 1]. Each image point corresponding to a data point has a location on the two-dimensional plane determined by the sum vector of all unit vectors at each coordinate, multiplied by the value of the data element for that coordinate.

Figure 4 displays the Star Coordinates projection of the Iris data. Each variable was *normalized to the range [0, 1]*.

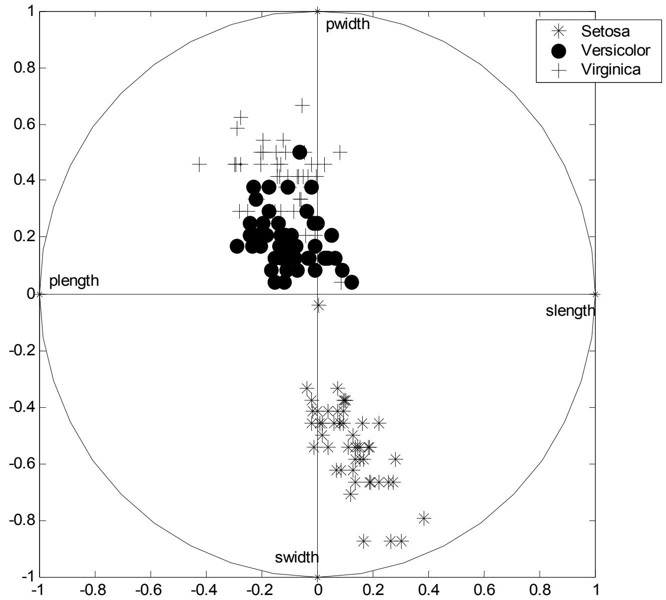


Figure 4. Star Coordinates of the Iris data

5.2.6. Self-Organizing Map (SOM)

The SOM technique (Kohonen 2001) is a special type of neural network based on unsupervised learning. The SOM algorithm is similar to the K-Means clustering algorithm (MacQueen 1967), but the output of a SOM is topological and neighboring clusters are similar. The SOM represents the data items on a two-dimensional grid, where each item is assigned to a node of the grid in an orderly way so that similar data items are mapped to the same node or neighboring nodes. The grid consists of units (nodes) that have assigned reference vectors with the same dimensionality as the original data. After learning is complete, the reference vectors are updated such that they resemble most of the data items, as much as possible. Each data item is then mapped to the map unit whose reference vector is most similar to the data vector. Multiple data items mapped onto the same unit are similar (i.e., a cluster).

Figure 5 represents the Iris dataset on a SOM grid of 12x9 nodes. The technique of *jittering*[15] was used to slightly change the position of each data item so that the items mapped to the same node will not overlap. The data were *normalized* using the *variance* method. Other parameters of the SOM were initialized as follows: Gaussian neighborhood, radius [12, 1], batch training, and linear initialization.

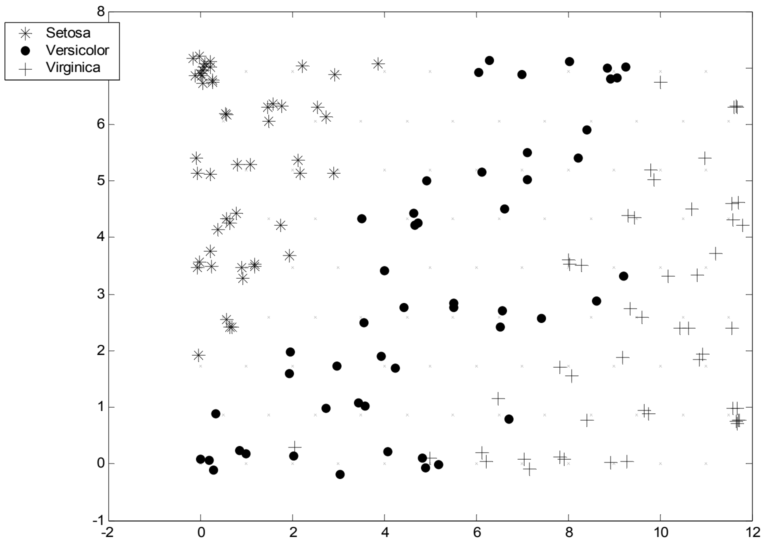


Figure 5. Self-Organizing Map of the Iris data

5.3. Objective measures for clustering evaluation

Clustering structures can be assessed by using quantitative measures called *clustering validity measures* (Milligan 1996; Theodoridis and Koutroumbas 1999; Halkidi et al. 2002a,b). There are three types of measures used in the evaluation of a clustering structure: *external*, *internal* and *relative* measures (Theodoridis and Koutroumbas 1999).

External measures are used to evaluate to what extent a clustering solution, obtained after applying a clustering technique, matches a known or assumed

[15] Jittering is a technique used for changing with a small value the position of each data point (its coordinates on the map) in order to solve the overlap problem (Cleveland 1993, p. 120; Hoffman and Grinstein 2002). Without employing this technique, the data points mapped to the same map unit would overlap.

structure of the dataset. Internal measures are used to assess a clustering solution in terms of the internal relationships among the data items, for example, the distances among them. External and internal measures are used in conjunction with hypothesis testing. A detailed discussion of the use of hypothesis testing is presented by Theodoridis and Koutroumbas (1999, p. 544-48).

Relative measures are used to compare different clustering solutions obtained by using different clustering techniques or different parameters of the same technique. Relative measures do not require hypothesis testing, and therefore are less time-consuming to calculate.

We have explored the use of the following measures:

- Rand statistic, Jaccard coefficient, Fowlkes and Mallows index,
- Hubert's Γ or normalized ($\hat{\Gamma}$) statistic, and
- The modified Hubert's Γ or $\hat{\Gamma}$ statistic.

These measures are described in detail by Theodoridis and Koutroumbas (1999), and we will briefly describe them in the following. Before we describe the measures, we define the following mathematical notations. The dataset is denoted by X and the data points are denoted by $x_i, i = 1,...,N$, N being the total number of the data items in X. The clustering structure produced by applying a clustering technique to the dataset X is denoted by C. A predetermined or a priori known partition of the data is denoted by P. The *proximity matrix* D of a dataset X is a matrix whose elements $D(i, j)$ are equal to the distances between the vectors x_i and x_j of X, $i, j = 1,..., N, i \neq j$.

5.3.1. Rand statistic (R), Jaccard coefficient (J) and Fowlkes and Mallows index (FM)

Rand statistic (*R*), Jaccard coefficient (*J*), Fowlkes and Mallows index (*FM*) belong to the category of external validity measures and are typically used to compare a clustering structure C, produced by a clustering technique, with a known partition P of the dataset X.

The main idea underlying the calculation of these measures is the evaluation of whether pairs of data points belong to the same cluster or not in both the known partition P and the obtained partition C. Their calculation is presented in more detail in Theodoridis and Koutroumbas (1999) and Paper 2.

R, *J*, and *FM* have values in [0, 1]. The larger the values of *R*, *J* and/or *FM* are, the higher the agreement between C and P is. In order to ensure that these values are significant and not achieved merely by chance, Theodoridis and Koutroumbas (1999) recommend the statistical testing of these values using Monte Carlo techniques for estimating the probability density function of the measures under the null hypothesis.

5.3.2. Hubert's Γ and normalized ($\hat{\Gamma}$) statistics

Hubert's Γ statistic (Hubert and Schultz 1976) is an index that measures the correlation between two matrices, A and B, of dimensions $N \times N$, drawn independently of each other (Theodoridis and Koutroumbas 1999). The normalized Γ statistic, denoted by $\hat{\Gamma}$, can also be calculated and it has values in [-1, 1]. Large absolute values of $\hat{\Gamma}$ indicate agreement between the matrices A and B. Theodoridis and Koutroumbas discuss the use of this type of statistics as external and internal measures for evaluating clustering validity. In the following, we focus only on the description of the normalized Γ statistic, $\hat{\Gamma}$.

In **external validity assessment**, the Hubert's $\hat{\Gamma}$ statistic is used in comparing the known partition P with the proximity matrix D of the data. It measures the degree to which the proximity matrix D of X matches a predefined partition P. Thus, *it is not necessary to apply a clustering technique to the data. But it is required to know a priori the structure of the data (the number of clusters and the composition of each cluster).*

In **internal validity assessment**, the Hubert's $\hat{\Gamma}$ statistic is used to determine if the resulted clustering structure matches the information inherent in the data. The information inherent in the data is usually represented by the distances between data points (i.e., proximity matrix). The calculation of the $\hat{\Gamma}$ statistic is based on the information given by the resulting clustering, C, and the proximity matrix of the data, D. The value of $\hat{\Gamma}$ is a measure of the degree of correspondence between D and C.

In order to test the statistical significance of the calculated $\hat{\Gamma}$, hypothesis testing is used. The detailed calculation of the measure can be found in (Theodoridis and Koutroumbas 1999; Halkidi et al. 2002a; Paper 2).

5.3.3. The modified Hubert's Γ and $\hat{\Gamma}$ statistics

The *modified* Hubert's Γ and $\hat{\Gamma}$ statistics are **relative measures**, employed to compare different clustering solutions obtained by using different clustering techniques. The normalized modified $\hat{\Gamma}$ statistic is calculated for the proximity matrix D of the data and a matrix obtained from the information given by C. The detailed calculation can be found in (Theodoridis and Koutroumbas 1999; Paper 3).

The modified $\hat{\Gamma}$ has values in [-1,1]. Large absolute values of $\hat{\Gamma}$ indicate agreement between the matrices D and C. The difference between the measure used in the internal validity assessment and the modified $\hat{\Gamma}$ is the type of information used to characterize the obtained clustering C. In the former case, the membership of data points to different clusters in C is taken into account. In the latter case, the distance between the cluster representatives is used in the calculation of the measure.

5.4. Proposed approaches to evaluate projections

In evaluating projections, we propose the use of the above measures for comparing the information contained in the *original data* (e.g., the known clustering structure or the distances between data points) with the information contained in the *projection*. The latter can be represented by the distances between data points after a projection technique has been applied or by the information obtained from clustering the projection.

Table 8 shows a summary of the measures, highlighting the information that is compared.

Table 8. Summary of clustering validity measures used for the evaluation of projections

Measures	*Information under comparison*	
	Original data	*Projected data*
R, J, FM	Known partition of the original data	Obtained clustering from the projection
$\hat{\Gamma}$*_ext*	Known partition of the original data	Proximity matrix of the projection
$\hat{\Gamma}$*_int*	Proximity matrix of the original data	Obtained clustering from the projection
$\hat{\Gamma}$*_dist*	Proximity matrix of the original data	Proximity matrix of the projection
$\hat{\Gamma}$*_m*	Proximity matrix of the original data	Proximity matrix of the projection, taking into account the distances between cluster representatives, when clustering is applied to the projection

The normalized Hubert's statistic, calculated for external and internal evaluation of projections, is denoted in the following by $\hat{\Gamma}$ *_ext* and $\hat{\Gamma}$ *_int,* respectively. Moreover, the index that measures the match between proximity matrices of original and projected data is denoted by $\hat{\Gamma}$ *_dist*. The normalized modified Hubert's $\hat{\Gamma}$ is denoted by $\hat{\Gamma}$*_m.*

5.4.1. The use of R, J and FM measures to evaluate a projection

The R, J and FM measures can be used only if one knows a priori the clustering structure of the data. The goal of using these indices is to evaluate the extent to which the projections preserve a known clustering structure of the data. The procedure of calculating these indices for this purpose is presented in the following. We present the procedure for calculating only the *R* measure, because the calculation of *J* and *FM* is similar. When evaluating/comparing multiple projections, the same clustering technique and distance measure must be used for the computation of the index.

1. We cluster the projected data using a clustering technique. The number of obtained clusters should be equal to the number of known classes (groups) in the data.
2. We calculate the *R* measure using the formulas in (Theodoridis and Koutroumbas 1999), also described in Paper 2. The *R* measure is computed for the known partition of the data and the clustering structure obtained at Step 1;
3. We calculate the estimate of *R* under the null hypothesis of no clustering structure at a chosen significance level.

Interpretation of R's value and assessment of the projection: If the value of *R* is higher than the estimate of *R* under the null hypothesis at the chosen significance level, then the null hypothesis of random structure is rejected. This result signifies that the clustering solution is meaningful and not achieved merely by chance. The value of *R* represents the extent to which the clustering solution resembles the known structure in the data. A high value of *R* shows a good match between the clustering solution on the projected data and the known structure. We can say that in this case the *projection preserves the clustering structure of the data. Therefore, the projected data can be used instead of the original data in clustering tasks and in representing graphically the known or derived clusters.*

5.4.2. *The use of Hubert's $\hat{\Gamma}$ statistic to evaluate a projection*

Hubert's $\hat{\Gamma}$ statistic can be used in three different situations. First, it can be used as an external validity measure in order to compare a known partition of the data to the proximity matrix of the projected data (Paper 2). Second, Hubert's $\hat{\Gamma}$ statistic can be used as an internal validity measure in order to evaluate the extent to which the clustering obtained on the projection is internally consistent with the information inherent in the original data (Paper 2). Moreover, Hubert's $\hat{\Gamma}$ statistic can be used as a measure of association between the proximity matrices of the original data and projected data (Paper 3). In the following we describe the procedures for calculating these indices in each of the three situations mentioned above.

The computation of the Hubert's $\hat{\Gamma}$ statistic involves the use of proximity matrices. In order to use the indices in comparisons of the projections, the proximity matrices should be comparable. Thus, to cope with eventual large differences between distances calculated on different projections, *we normalize all proximity matrices using the global histogram equalization method.* We use this normalization method because it does not take into account the actual values of the distances, but their ranks. This normalization method works in two steps. First, the distances are replaced by their ranks. Then, the ranks are normalized to the range [0, 1] (Vesanto et al. 2000).

5.4.2.1. *The use of Hubert's $\hat{\Gamma}$ statistic in comparing a known partition with the proximity matrix of a projection ($\hat{\Gamma}$ _ext)*

The goal of using the Hubert's $\hat{\Gamma}$ statistic in this case is to evaluate the extent to which the proximity matrix of a projection matches a known clustering structure in the data. When evaluating/comparing multiple projections, the same distance measure must be used for the computation of the proximity matrix of each projection. *A prerequisite of conducting this measurement is to know a priori the clustering structure of the data.*

1. We calculate the proximity matrix of the projection;
2. We normalize the proximity matrix obtained at Step 1 using the global histogram equalization method;
3. We calculate the Hubert's $\hat{\Gamma}$ statistic for a matrix defined based on the membership of the data points to the clusters in P and the proximity matrix obtained at Step 2;
4. We calculate the estimate of the Hubert's $\hat{\Gamma}$ under the null hypothesis of no clustering structure of the data, at a chosen significance level.

Interpretation of the value of $\hat{\Gamma}$*_ext and assessment of the projection*: If the value of $\hat{\Gamma}$*_ext* is large enough, the null hypothesis is rejected, signifying that the known structure matches the projected data. A high value of the $\hat{\Gamma}$*_ext* (close to 1) shows a good match between the projection and the known clustering. In this case, the projection can be used to represent the known clustering graphically.

5.4.2.2. The use of Hubert's $\hat{\Gamma}$ *statistic in evaluating the internal consistency of the clusters obtained on the projected data (*$\hat{\Gamma}$*_int)*

The goal of using Hubert's $\hat{\Gamma}$ statistic in this case is to evaluate the extent to which the clusters obtained on projected data match *the information inherent in the original data*. In our case, this information is represented by the proximity matrix of the original data. Thus, we do not compare the resulting clustering against the proximity matrix of the projected data, but against original data. This comparison against original data will also be involved at the hypothesis testing step. When evaluating/comparing multiple projections, the same clustering technique and distance measure must be used for the computation of the statistic.

In the following we describe the procedure for calculating the $\hat{\Gamma}$*_int* for projections.

1. We calculate the proximity matrix of the original data;
2. We normalize the proximity matrix obtained at Step 1 using the global histogram equalization method;
3. We cluster the projected data by applying a clustering technique to partition the projected data;
4. We calculate the Hubert's $\hat{\Gamma}$ statistic for the proximity matrix obtained at Step 2 and a matrix defined for the clustering solution obtained at Step 3;
5. We calculate the estimate of Hubert's $\hat{\Gamma}$ under the null hypothesis of no clustering structure of the data and at a chosen significance level;

Interpretation of the value of $\hat{\Gamma}$*_int and assessment of the projection*: If the value of $\hat{\Gamma}$*_int* is higher than its estimate under the null hypothesis for a certain projection and it is close to 1, then the resulting clustering matches the information inherent in the original data. Thus, the use of the projected data in clustering tasks determines the obtaining of meaningful clusters.

5.4.2.3. The use of Hubert's $\hat{\Gamma}$ statistic in evaluating the association between the proximity matrices of original data and projected data ($\hat{\Gamma}$_dist)

The goal of using the Hubert's $\hat{\Gamma}$ statistic in this case is to evaluate the extent to which the projected data preserve the distances between data points. We compare the proximity matrix of projected data against proximity matrix of original data.

The procedure of calculating the $\hat{\Gamma}$*_dist* is given below.

1. We calculate the proximity matrices of the original standardized data and of the projected data;
2. We normalize the proximity matrices of the original and projected data using the global histogram equalization method.
3. We calculate the Hubert's $\hat{\Gamma}$ statistic for the normalized proximity matrices obtained at Step 2.

Interpretation of the value of $\hat{\Gamma}$*_dist and assessment of the projection*: If $\hat{\Gamma}$*_dist* has a high value, the association between the proximity matrices of original and projected data is high. This result signifies that the projection preserve the original data relationships in terms of distances.

5.4.3. The use of the modified Hubert's $\hat{\Gamma}$ statistic in evaluating projections ($\hat{\Gamma}$_m)

The goal of using the modified Hubert's $\hat{\Gamma}$ statistic is to evaluate the extent to which the clustering produced by a projection reflects the information inherent in the original data, represented by its proximity matrix. The idea of the measure is similar to the one corresponding to $\hat{\Gamma}$*_int*, but different matrices are involved in the calculation of these measures (see Table 8).

For comparing different projections, the same clustering technique and the same distance measure must be used in the calculation of the $\hat{\Gamma}$*_m*. The procedure for calculating $\hat{\Gamma}$*_m* is given below.

1. We calculate the proximity matrix of the original *standardized* data;
2. We cluster the projected data using a clustering technique;

3. We calculate a proximity matrix for the projection, based on the distances between the clusters representatives in the clustering solution obtained at Step 2.
4. We normalize the proximity matrices obtained at Steps 1 and 3 using the global histogram equalization method;
5. We calculate the $\hat{\Gamma}\_m$ for the normalized proximity matrices obtained at Step 4;

Interpretation of the value of $\hat{\Gamma}\_m$ *and assessment of the projection:* If the value of $\hat{\Gamma}\_m$ is close to 1, then the resulting clustering on the projected data reflect the original relationships between data points, relationships represented in terms of distances. Thus, the projected data can be used in clustering tasks and in representing the corresponding clustering solutions.

5.5. Empirical results

5.5.1. Benchmark datasets

We have applied the procedures described above on five datasets. Three of them are from UCI Machine Learning Repository (Newman et al. 1998): Iris dataset, Voting records database and Wine recognition data. Iris dataset and Wine recognition data contain three a priori known classes (groups), while Voting records contain two classes (groups). We have also used two artificial datasets: Artificial 1 – containing three classes and Artificial 2 – with no clustering structure. The datasets are described in Paper 2 and 3.

When clustering of the data was required, we have used the *K-means technique* (MacQueen 1967). The K-means is an efficient algorithm for large datasets, and it is particularly suitable for detecting hyperspherical clusters (compact and well-separated) (Jain et al. 1999; Guo et al. 2003). To Iris dataset, Wine recognition, Artificial 1 and Artificial 2 we have applied K-means for obtaining three clusters. The Voting records dataset was clustered in two clusters. In all calculations of the proximity matrices, we have used the Euclidean distance for calculating the distance between two data points.

In the following, we report the results of the evaluations, first, as summaries by each measure, and, then, separately for each dataset. We report also on the values of the above measures calculated on the *original data*, before any projection technique was applied.

5.5.2. Summary of results for R, J and FM indices

The values of the *R*, *J*, and *FM* indicate whether the clusters obtained by applying K-means to the projected data resemble the known classes (groups) in the data. If the values of the indices are high (close to 1), then the clusters evaluated are very similar to the ones *a priori* known. We can say in this case that the projection preserves the known structure of the data and it can be used in visualizing both the known clusters and the derived clusters.

Figure 6 shows the values of *R* for the Iris, Artificial 1, Wine recognition, and Voting records datasets. Similar results are obtained for the *J* and *FM* indices. In the case of Artificial 2 dataset we did not calculate the *R*, *J* and *FM* statistics because the dataset does not have clustering structure.

The results show that the effectiveness for preserving the clustering structure varies according to the dataset under analysis. For example, for Artificial 1 data, Radviz and Star Coordinates provide the poorest results, while for the Voting data, they are the best techniques.

There are some differences between the results reported in this chapter and the ones in Paper 2 in the case of the Voting and Wine datasets at the Radviz and Star Coordinates. The Radviz and Star Coordinates techniques can provide different mappings based on the order of the dimensions. If the dataset has 13 or 16 dimensions, then the total number of possibilities to arrange the dimensions is 13! or 16!, respectively. Therefore, we have recalculated the indices *R*, *J* and *FM* and chose the maximum that we obtained by making 10,000 random arrangements. The differences are due to the fact that the values presented here are calculated based on the best projections obtained out of 10,000 random arrangements of the dimensions.

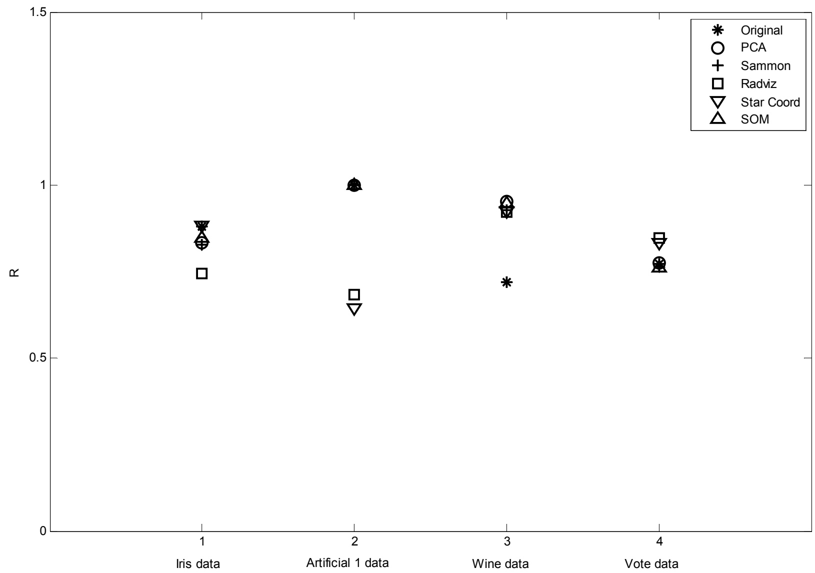


Figure 6. Plot of *R* statistic for different projection and different datasets

5.5.3.Summary of results for Hubert's $\hat{\Gamma}$ external ($\hat{\Gamma}$_ext)

The value of $\hat{\Gamma}$*_ext* indicates the extent to which the proximity matrix of a projection matches the known clustering of the data. We have also calculated this measure for the original data to examine whether a projection can resemble better the clustering structure in the data (Figure 7).

In the case of Iris data, the original data and the SOM provided the best match between the known clustering and their proximity matrices. Radviz provides the poorest match. For Artificial 1 data, all projections but Radviz and Star Coordinates provide good match between the known clustering and the proximity matrices of the projections. In the case of Wine and Vote datasets, the original data does not match the known clustering, but the projections reflect more (e.g., SOM in the case of Wine data) or less (e.g., Sammon's Mapping) this clustering.

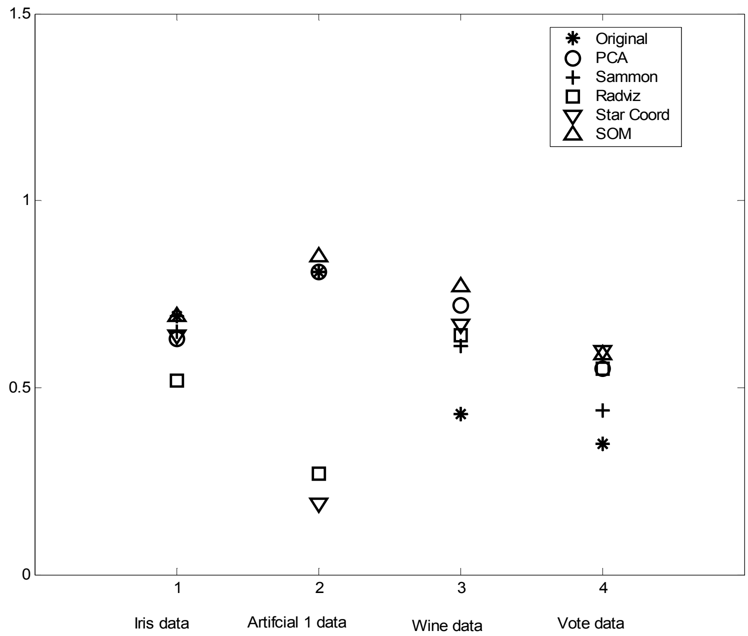


Figure 7. Plot of $\hat{\Gamma}$ *_ext* in external evaluation of the projections

5.5.4. Summary of results for Hubert's $\hat{\Gamma}$ internal ($\hat{\Gamma}$ _int)

The value of $\hat{\Gamma}$ *_int* indicates the extent to which a clustering obtained from projected data reflects the original distances between the data points. In the case of Iris data, there are no large differences between the clusterings obtained from different projections. In the case of Artificial 1 data, the Radviz and Star Coordinates projections provide the poorest results (Figure 8).

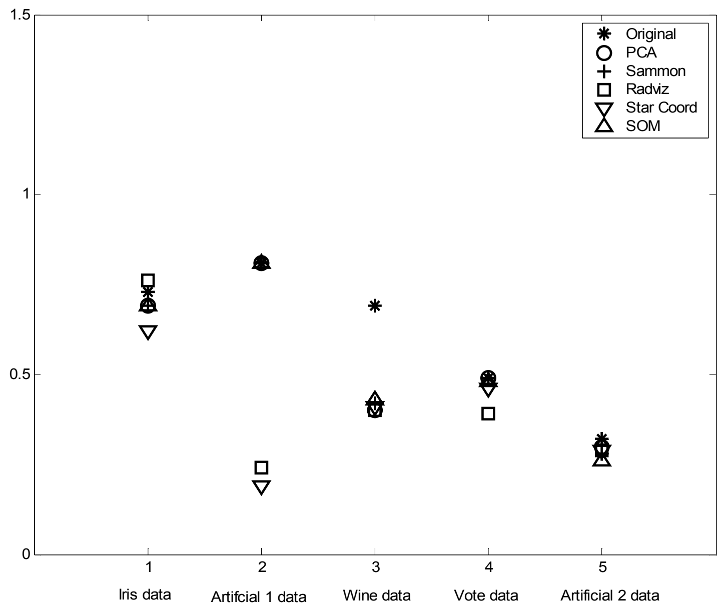


Figure 8. Plot of $\hat{\Gamma}$ *_int* in internal evaluation of the clusters obtained from projections

5.5.5. Summary of results for Hubert's $\hat{\Gamma}$ for proximity matrices ($\hat{\Gamma}$_dist)

The value of $\hat{\Gamma}$ *_dist* indicates the extent to which the proximity matrices of the original data and the projection are similar (i.e., the projection preserves the distances between data points). In almost all cases, the Radviz and Star Coordinates projections have less similar proximity matrices with the original data (Figure 9). The best results are obtained instead at Sammon's Mapping – a technique that tries indeed to maximize the fit with the original distances between data. PCA and SOM provide also projections which preserve the original distances between data points.

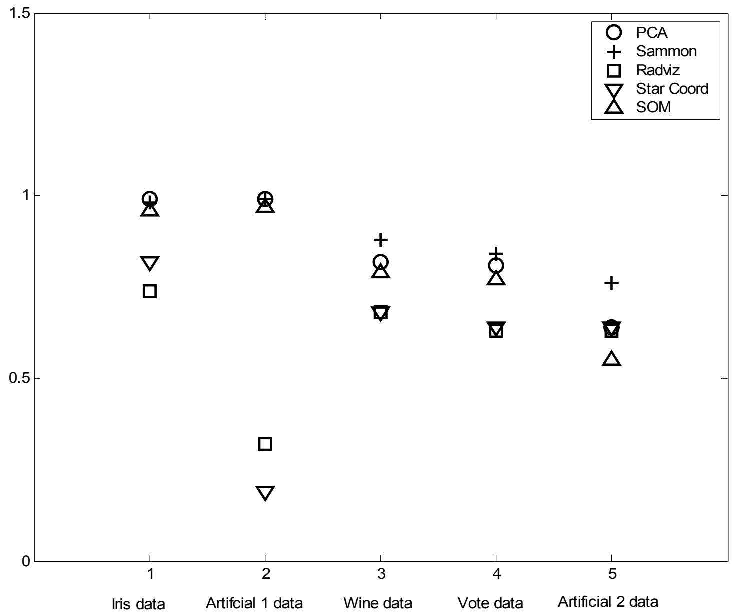


Figure 9. Plot of $\hat{\Gamma}$ _*dist* comparing the proximity matrices of the projection and of the original data

5.5.6. Summary of results for modified Hubert's $\hat{\Gamma}$ ($\hat{\Gamma}$_m)

The value of $\hat{\Gamma}$_*m* indicates the extent to which the clustering obtained by applying a clustering technique (K-means) on the projection data matches the information contained in the original data (e.g., distances between data points). Figure 10 shows that Sammon's Mapping, PCA, and SOM determine clusterings that best reflect the distances in the Iris data. PCA, Sammon's Mapping and SOM provide the best mapping of the Artificial 1 data, while the Radviz and Star Coordinates provide very poor mappings. In the case of Wine data, all techniques provide similar clusterings. However, the resulting clusterings do not reflect very well the distances among original data points. Similar results are obtained for the Voting data, where the Radviz technique provides the poorest results. The clustering in Artificial 2 data is not very meaningful in terms of distances among data points, result which is expected given the fact that this dataset does not have clustering structure.

In the following, we will discuss each dataset and projection separately, given the results obtained for each measure. The presentation of the results for each dataset follows the same outline. We first show the values of the measures for the original data. Then we present the values of the measures for each projection in a table, as well as the graphical representation of the data based on the respective projection. We present the known structure in one image and the clustering structure obtained on the projected data in another image.

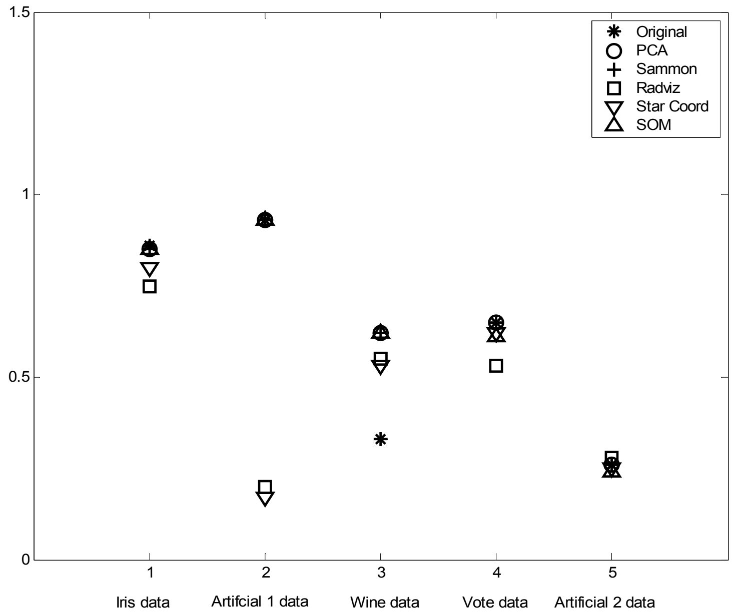


Figure 10. Plot of the modified Hubert's $\hat{\Gamma}$ statistic ($\hat{\Gamma}$*_m*)

5.5.7. Iris data

When the *original data* is used in clustering, the values of the indices are:

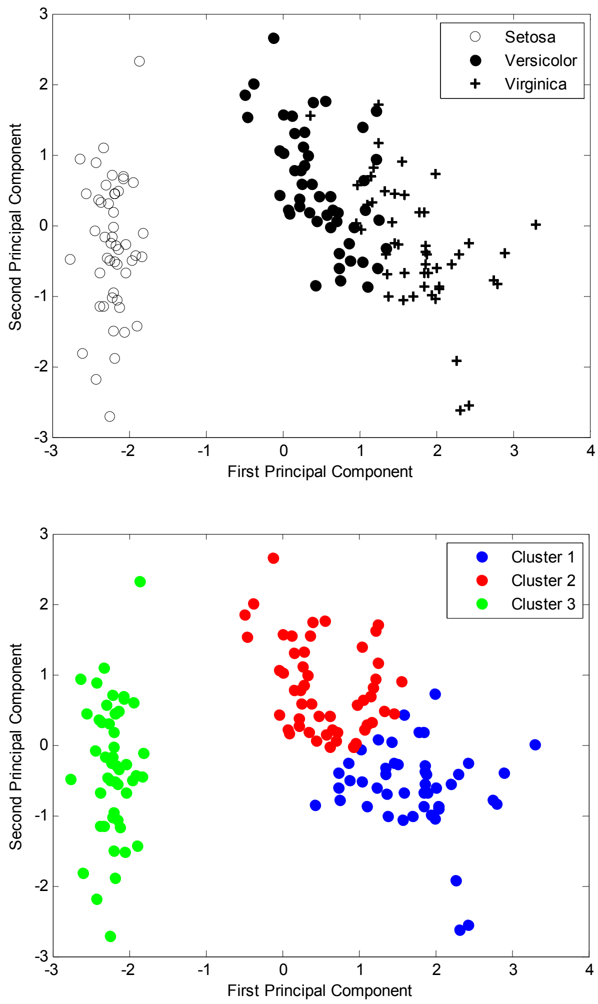
R	*J*	*FM*	$\hat{\Gamma}$*_ext*	$\hat{\Gamma}$*_int*	$\hat{\Gamma}$*_dist*	$\hat{\Gamma}$*_m*
0.880	0.696	0.821	0.69	0.73	N/A	0.86

The values of *R*, *J* and *FM* are high, signifying that the resulting clustering resembles the known structure. $\hat{\Gamma}$*_ext* = 0.69 shows that the original data is quite good in reflecting the clustering structure. Moreover, the clusters obtained are quite consistent internally in terms of distances between data points. The value of $\hat{\Gamma}$*_m* is high, which reflects again a good match between the derived clustering and the distances in the data.

5.5.7.1. PCA

Figure 11 shows the PCA projection of the Iris data. The upper image highlights the known clusters (the three Iris classes). The lower image highlights the obtained clusters. It is observed that the Iris-Setosa is well recognized by the K-means (green dots in lower image correspond to the white circles in upper image). The other two classes, being not linearly separable, were not correctly recognized (red and blue dots in lower image do not entirely correspond to the black circles and +'s in the upper image, respectively). The failure of recovering the complete clustering structure is captured in the *R*, *J* and *FM* measures.

In the left-side of the figure, there are presented the values of all indices. One can also compare the values of these measures with the ones obtained on the original data. In this case, the original data appear to be more appropriate to be used in clustering tasks.



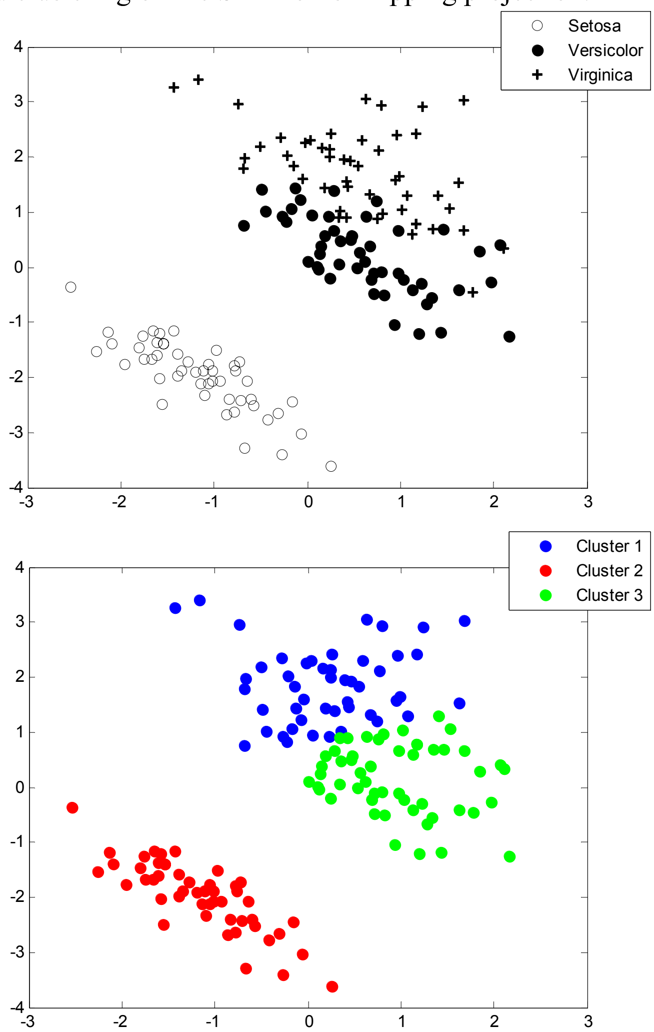
R	0.832
J	0.594
FM	0.745
$\hat{\Gamma}_{ext}$	0.63
$\hat{\Gamma}_{int}$	0.69
$\hat{\Gamma}_{dist}$	0.99
$\hat{\Gamma}_{m}$	0.85

Figure 11. PCA (Iris data). Up: known classes; Down: obtained clusters

$\hat{\Gamma}\_dist$ has a very high value, which shows that the original distances between data points are well represented in the projection. Hence, the PCA projection is effective in preserving the distances between data points.

5.5.7.2. Sammon's Mapping

Similarly with the previous figure, Figure 12 shows the known clustering structure and the obtained clustering on the Sammon's Mapping projection.



R	0.828
J	0.585
FM	0.738
$\hat{\Gamma}\_ext$	0.65
$\hat{\Gamma}\_int$	0.69
$\hat{\Gamma}\_dist$	0.98
$\hat{\Gamma}\_m$	0.85

Figure 12. Sammon's Mapping (Iris). Up: known classes; Down: obtained clusters

The values of the indices are slightly smaller than the ones obtained on the original data. They are also very similar to the ones obtained at PCA. The class

Setosa was fully recognized, but the other two classes are not completely recognized, because they are not linearly separable.

5.5.7.3. Radviz

Figure 13 shows the known clustering structure and the obtained clustering on the Radviz projection.

R	0.743
J	0.535
FM	0.716
$\hat{\Gamma}\_ext$	0.52
$\hat{\Gamma}\_int$	0.76
$\hat{\Gamma}\_dist$	0.74
$\hat{\Gamma}\_m$	0.75

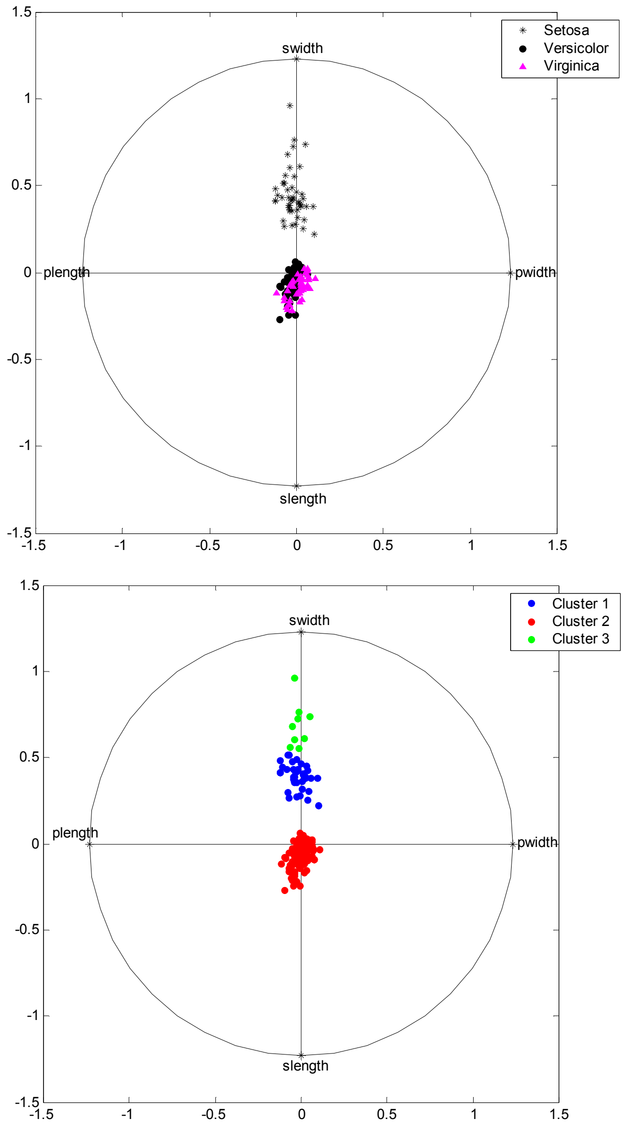


Figure 13. Radviz (Iris). Up: known classes; Down: obtained clusters

The values of the indices are smaller than the ones obtained on the original data, PCA and Sammon's Mapping, except $\hat{\Gamma}$ *_int*. On the representation produced by Radviz, Iris-Versicolor and Iris-Virginica are very much overlapping each other; therefore, the clustering of the data using K-means on this projection is not very successful in revealing the true classes in the data. The class Setosa is divided into two clusters (green and blue dots in the lower image), while the classes Versicolor and Virginica are grouped into a single cluster (red dots in the lower image). Thus, the small values of the indices reflect that the projection does not preserve the clustering structure existent in the data.

5.5.7.4. Star Coordinates

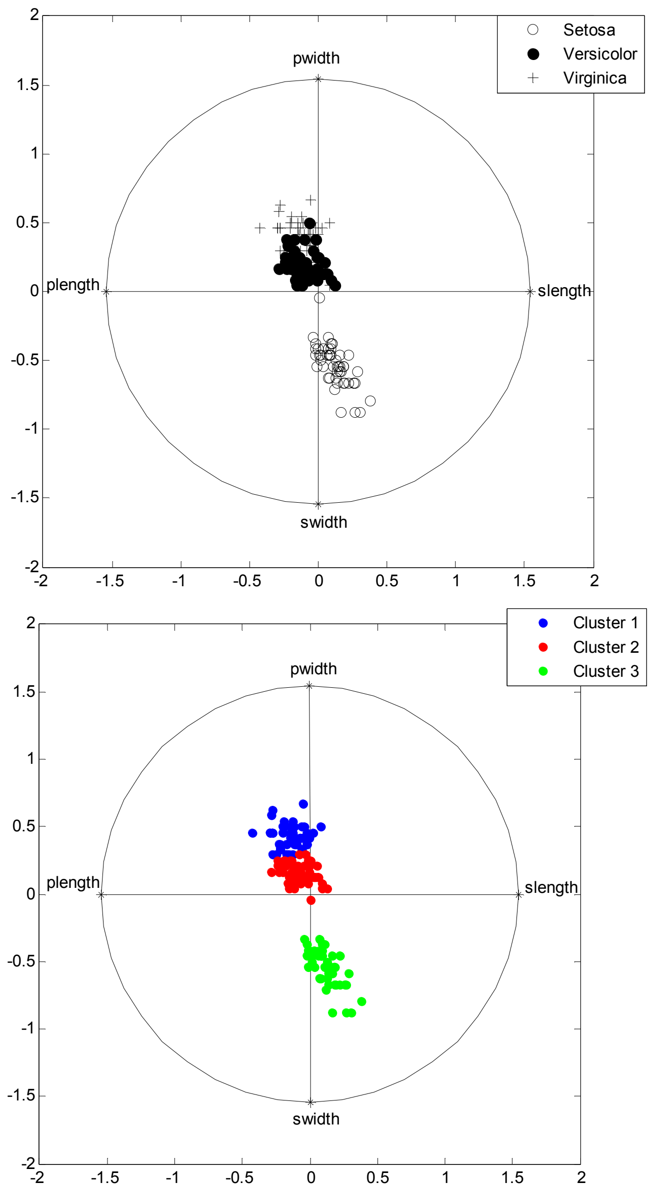
Figure 14 shows the known clustering structure and the obtained clustering based on the Star Coordinates representation. *R*, *J* and *FM*' values are higher than the ones obtained on the original data and the previous projections.

The values of $\hat{\Gamma}$ *_int*, $\hat{\Gamma}$ *_dist* and $\hat{\Gamma}$ *_m* are lower than at previous projections. This fact may indicate that the known clustering is based not only on the original Euclidean distances between data points, but also on other information in the data, which the Star Coordinates projection is capable of representing. The class Setosa is well represented and recognized and fewer errors than in previous projections were made by clustering the Star Coordinates projection.

5.5.7.5. SOM

Figure 15 shows the known clustering structure and the obtained clustering on the SOM projection. *R*, *J* and *FM*'s values are smaller than the ones obtained on the original data and Star Coordinates projection, but they are slightly better than the ones obtained on the PCA, Sammon's Mapping, and Radviz. The internal validation of the clusters is similar to the ones obtained for PCA and Sammon's Mapping. The original distances are well preserved by the projection and the derived clustering reflects the original distances in the data.

The class Setosa is well recognized, but in the other two classes some errors are seen in the upper and lower parts of the map.



R	0.883
J	0.699
FM	0.823
$\hat{\Gamma}\_ext$	0.64
$\hat{\Gamma}\_int$	0.62
$\hat{\Gamma}\_dist$	0.82
$\hat{\Gamma}\_m$	0.8

Figure 14. Star Coordinates (Iris). Up: known classes; Down: obtained clusters

R	0.846
J	0.622
FM	0.767
$\hat{\Gamma}\_ext$	0.69
$\hat{\Gamma}\_int$	0.69
$\hat{\Gamma}\_dist$	0.96
$\hat{\Gamma}\_m$	0.85

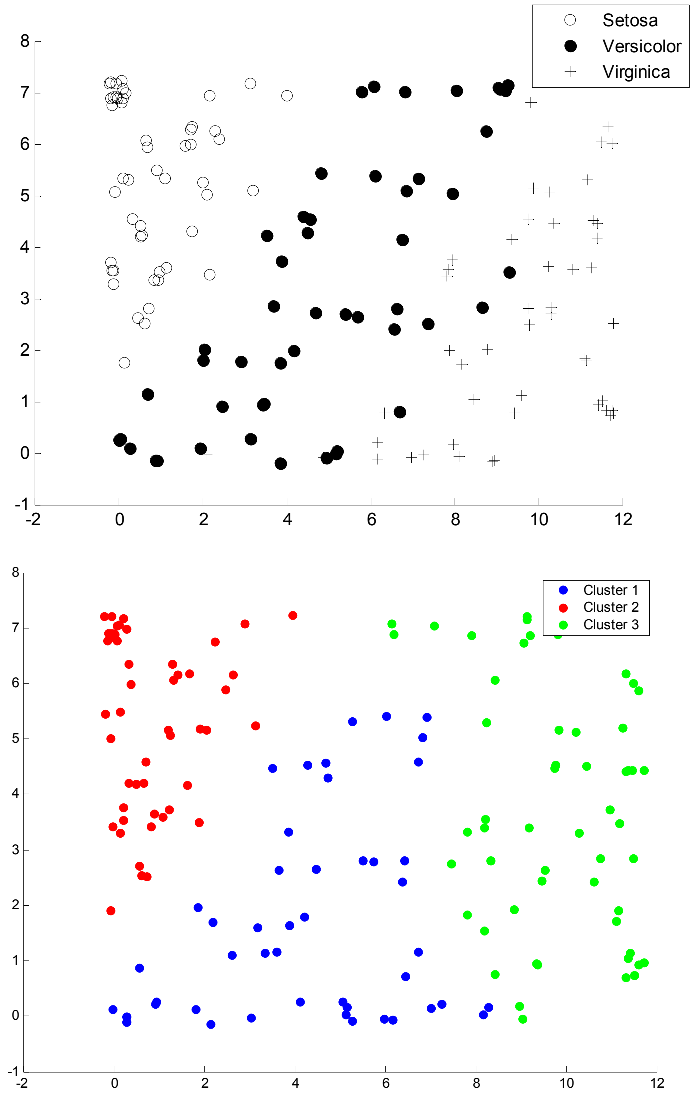


Figure 15. SOM (Iris). Up: known classes; Down: obtained clusters

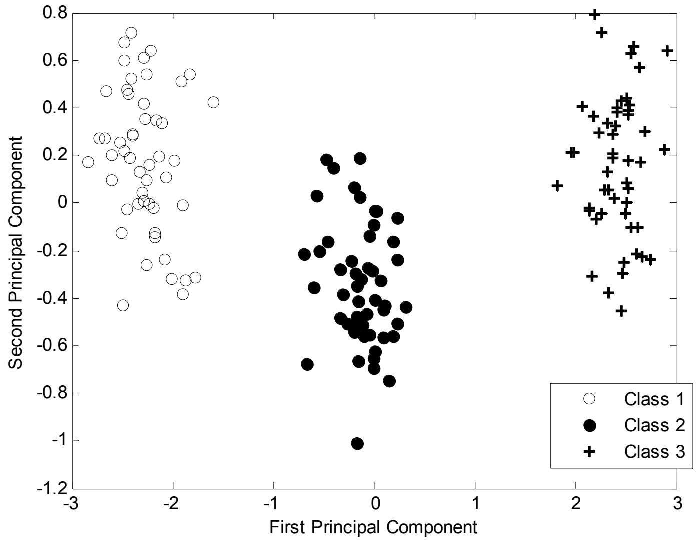
5.5.8. Artificial 1 data

This dataset was randomly generated from three different uniform distributions. It has 150 data points, four attributes and the three classes are linearly separable. The values of *R*, *J* and *FM* after applying K-means to the *original data* are:

R	*J*	*FM*	$\hat{\Gamma}\_ext$	$\hat{\Gamma}\_int$	$\hat{\Gamma}\_dist$	$\hat{\Gamma}\_m$
1	1	1	0.81	0.81	N/A	0.93

5.5.8.1. PCA

Figure 16 shows the known clustering structure and the obtained clustering on the PCA projection. The values of the indices are very high and the same with the ones obtained for original data. Figure 16 shows that the PCA represents well the data so that all three classes are fully uncovered by the K-means.



R	1
J	1
FM	1
$\hat{\Gamma}\_ext$	0.81
$\hat{\Gamma}\_int$	0.81
$\hat{\Gamma}\_dist$	0.99
$\hat{\Gamma}\_m$	0.93

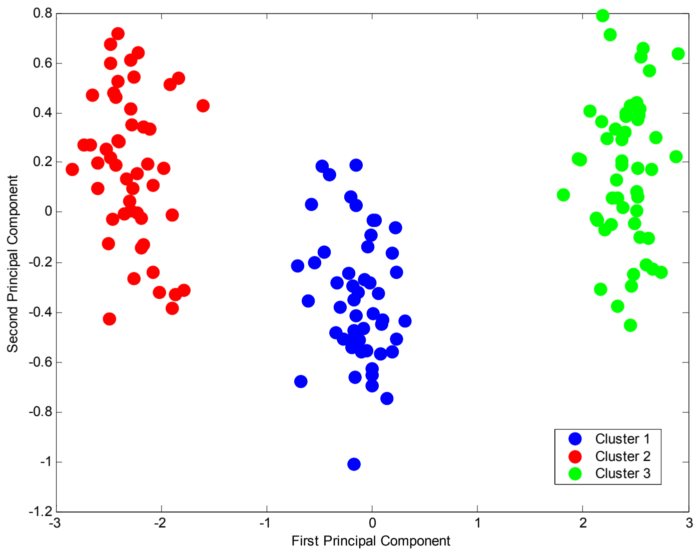
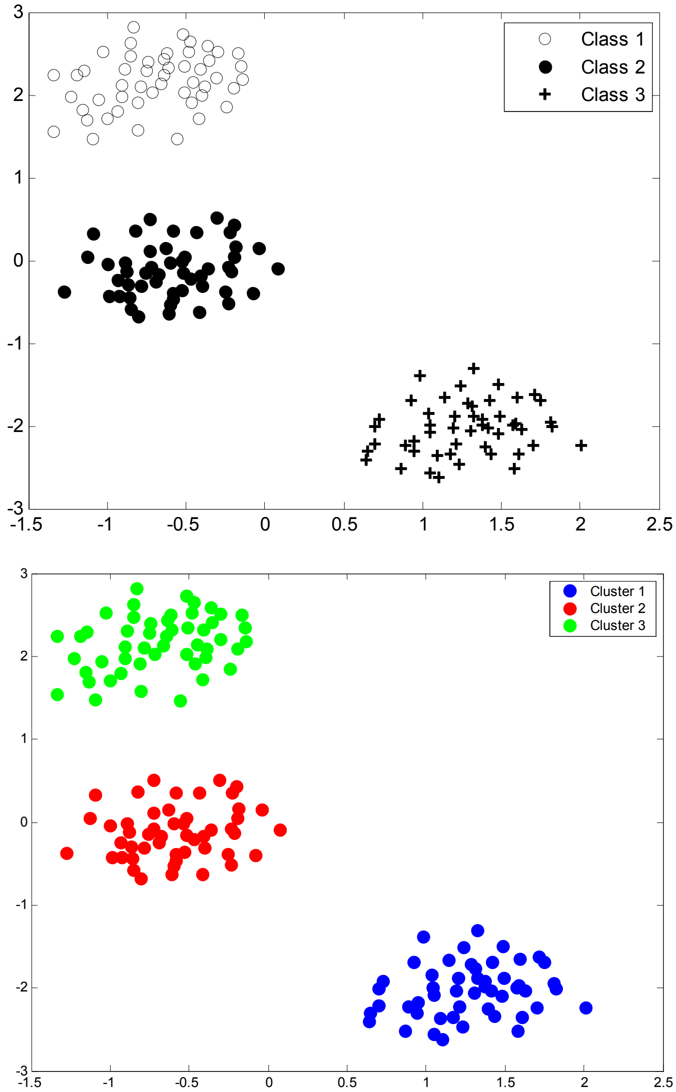


Figure 16. PCA (Artificial 1 data). Up: known classes; Down: obtained clusters

5.5.8.2. Sammon's Mapping

Figure 17 shows the known clustering structure and the obtained clustering on the Sammon's Mapping projection. All indices have high values, which are equal to the ones obtained at original data and PCA. Figure 17 shows that also Sammon's Mapping represents the data so that all three classes are fully uncovered by the K-means.

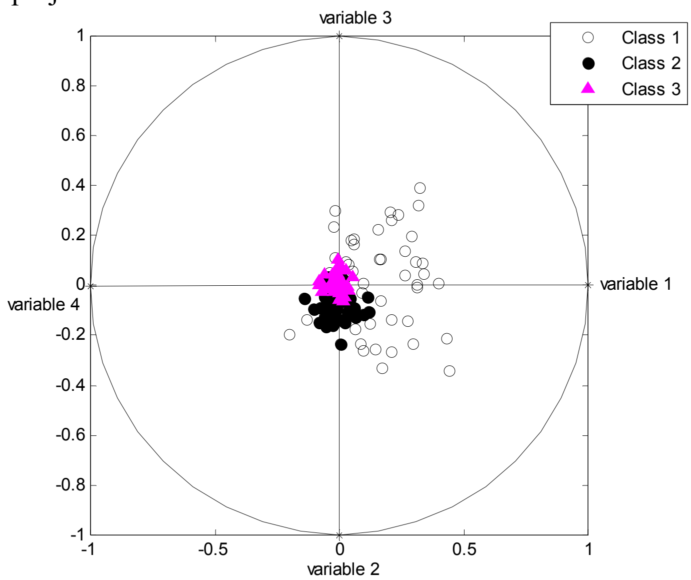


R	1
J	1
FM	1
$\hat{\Gamma}$_*ext*	0.81
$\hat{\Gamma}$_*int*	0.81
$\hat{\Gamma}$_*dist*	0.99
$\hat{\Gamma}$_*m*	0.93

Figure 17. Sammon's Mapping (Artificial 1 data). Up: known classes; Down: obtained clusters

5.5.8.3. Radviz

Figure 18 shows the known clustering structure and the obtained clustering on the Radviz projection. The values of the *R*, *J* and *FM* indices are significant but lower than at previous projections.



R	0.683
J	0.392
FM	0.566
$\hat{\Gamma}$_*ext*	0.27
$\hat{\Gamma}$_*int*	0.24
$\hat{\Gamma}$_*dist*	0.32
$\hat{\Gamma}$_*m*	0.2

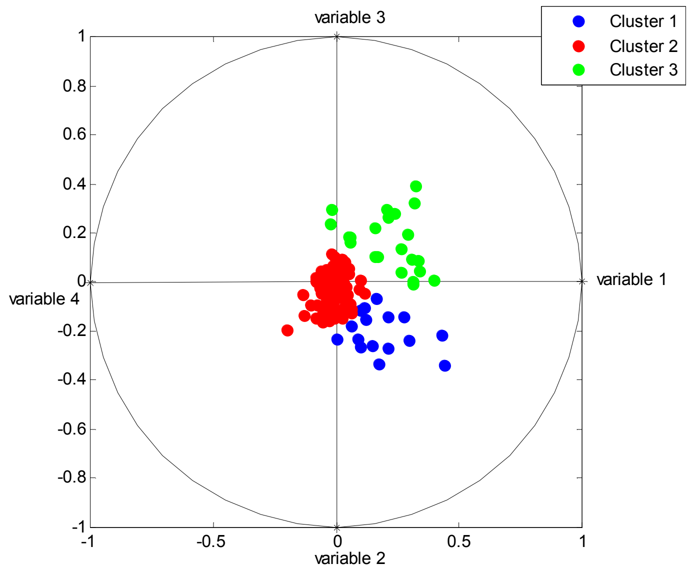


Figure 18. Radviz (Artificial 1 data). Up: known classes; Down: obtained clusters

The Radviz technique represents Class 2 and Class 3 overlapped and K-means does not distinguish between them, but divide Class 1 into two clusters. Thus, as the values of *R*, *J* and *FM* show, this technique is not effective for preserving the clustering structure of the original data. Moreover, the values of the other indices are very low, showing that this projection is not very effective for representing clusters and the original distances in this dataset.

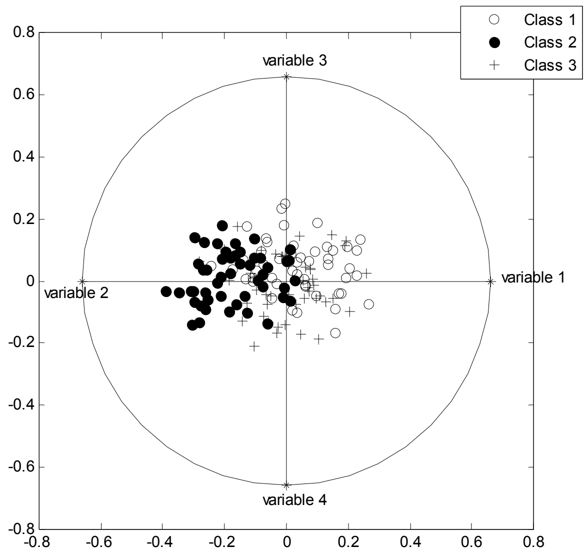
5.5.8.4. Star Coordinates

Figure 19 shows the known clustering structure and the obtained clustering on the Star Coordinates projection. The values of the *R*, *J* and *FM* indices are significant but very low. The Star Coordinates technique represents all three classes overlapping and K-means does not distinguish between them. The clustering result is very different from the true clustering structure. The values of *R*, *J* and *FM* are very low indicating also that the clustering obtained on this projection is not similar to the one *a priori* known. Thus, the Star Coordinates applied to this dataset does not preserve the clustering structure of the data.

Moreover, the values of all Hubert's $\hat{\Gamma}$ statistics are very low, indicating that, similarly as in the case of Radviz, the Star Coordinates projection is not appropriate for representing the distances between data points in this dataset.

5.5.8.5. SOM

Figure 20 shows that the projected data represent well the three classes. All indices have very high values, which show that the SOM is very effective for preserving the distances and the clustering structure of the data.



R	0.646
J	0.309
FM	0.472
$\hat{\Gamma}\_ext$	0.19
$\hat{\Gamma}\_int$	0.19
$\hat{\Gamma}\_dist$	0.19
$\hat{\Gamma}\_m$	0.17

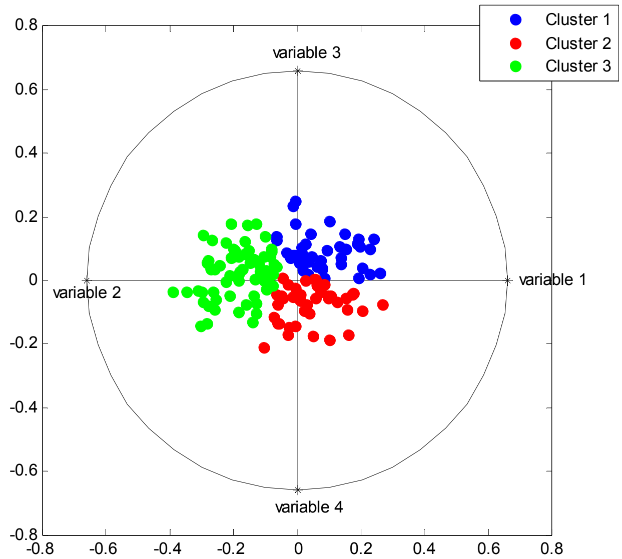


Figure 19. Star Coordinates (Artificial 1 data). Up: known classes; Down: obtained clusters

R	1
J	1
FM	1
$\hat{\Gamma}\_ext$	0.85
$\hat{\Gamma}\_int$	0.81
$\hat{\Gamma}\_dist$	0.97
$\hat{\Gamma}\_m$	0.93

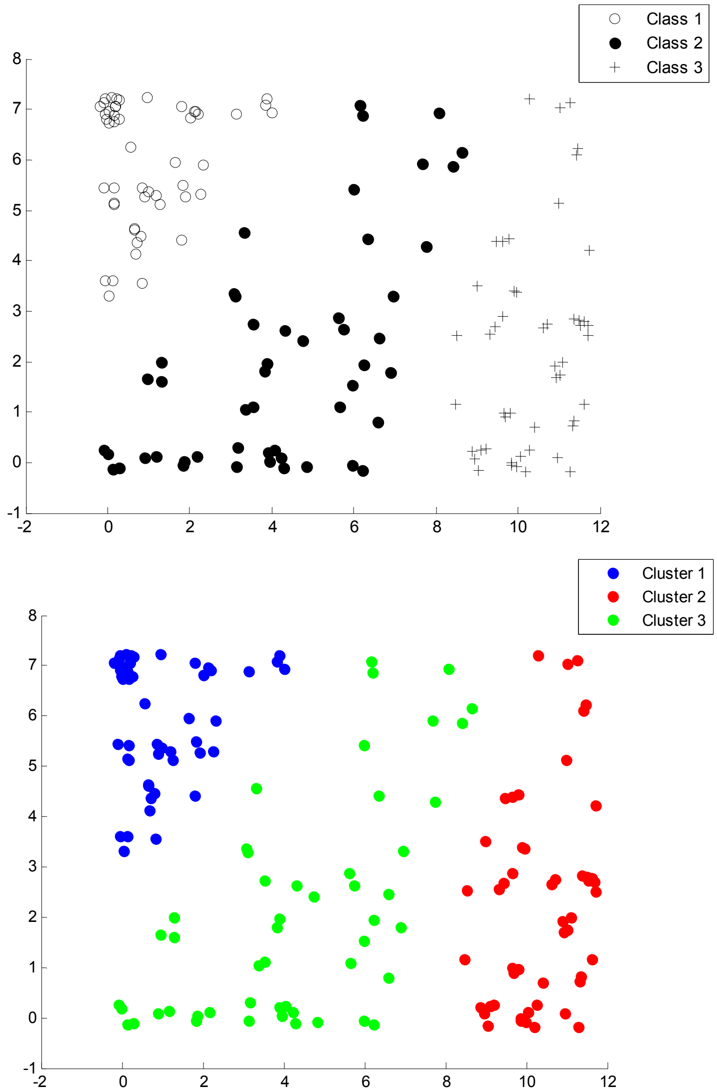


Figure 20. SOM (Artificial 1 data). Up: known classes; Down: obtained clusters

5.5.9. Wine recognition data

This dataset consists of 178 data points, 13 attributes and 3 classes. Appendix 1 presents the plots of each projection, highlighting the classes and resulting clusters. The values of the indices are described in Table 9.

Table 9. Evaluation results for Wine dataset

Projection technique	*R*	*J*	*FM*	$\hat{\Gamma}\_ext$	$\hat{\Gamma}\_int$	$\hat{\Gamma}\_dist$	$\hat{\Gamma}\_m$
Original	0.719	0.412	0.584	0.43	0.69	N/A	0.33
PCA	0.953	0.870	0.930	0.72	0.40	0.82	0.62
Sammon's Mapping	0.927	0.802	0.891	0.61	0.42	0.88	0.62
Radviz	0.922	0.791	0.883	0.64	0.40	0.68	0.55
Star Coordinates	0.926	0.801	0.889	0.67	0.41	0.68	0.53
SOM	0.947	0.854	0.922	0.77	0.43	0.79	0.62

The highest values of *R*, *J* and *FM* respectively are recorded for the PCA projection. The SOM projection yields also very good values of *R*, *J* and *FM*. However, all values are close to one, which means that K-means uncovered closely the true clustering structure of the data by using different projections.

The graphical representations based on the five type of projection techniques show that, indeed, the three resulting clusters obtained by applying K-means are almost the same as the true clusters in the data. Thus, the values of *R*, *J* and *FM* indicate us that all the projections are good for representing the Wine recognition data and the clusters inherent in these data.

5.5.10. Voting records data

This dataset consists of 435 data points, 16 attributes and two classes. The plots of each projection, highlighting the classes and resulting clusters are presented in Appendix 2. The values of the indices are presented in Table 10.

Table 10. Evaluation results for Voting dataset

Projection technique	*R*	*J*	*FM*	$\hat{\Gamma}\_ext$	$\hat{\Gamma}\_int$	$\hat{\Gamma}\_dist$	$\hat{\Gamma}\_m$
Original	0.772	0.636	0.778	0.35	0.49	N/A	0.65
PCA	0.775	0.640	0.781	0.55	0.49	0.81	0.65
Sammon's Mapping	0.765	0.627	0.771	0.44	0.49	0.84	0.65
Radviz	0.848	0.745	0.854	0.55	0.39	0.63	0.53
Star Coordinates	0.833	0.720	0.837	0.60	0.46	0.64	0.62
SOM	0.762	0.622	0.767	0.59	0.48	0.77	0.61

The highest values of *R*, *J* and *FM* respectively are recorded for the Radviz and Star Coordinates projections. The PCA, Sammon's Mapping and SOM projections have more points from one class that are represented among the points from the other class. However, in terms of internal consistency of the clusters produced by K-means on the projections, the Radviz provides the poorest result, but the other projections are not very much different than Radviz.

5.5.11. Artificial 2 data

This dataset consists of 200 data points and 4 attributes. The data were randomly generated so that they contain only one class. The plots of each projection, highlighting the resulting clusters are presented in Appendix 3. We calculated for this dataset only the $\hat{\Gamma}$*_int*, $\hat{\Gamma}$*_dist* and $\hat{\Gamma}$*_m* indices (Table 11).

Table 11. Evaluation results for Artificial 2 data

Projection technique	$\hat{\Gamma}$*_int*	$\hat{\Gamma}$*_dist*	$\hat{\Gamma}$*_m*
Original	0.32	N/A	0.26
PCA	0.30	0.64	0.26
Sammon's Mapping	0.28	0.76	0.26
Radviz	0.29	0.63	0.28
Star Coordinates	0.29	0.64	0.25
SOM	0.26	0.55	0.24

The values of $\hat{\Gamma}$*_int* and $\hat{\Gamma}$*_m* are very low indicating that the three obtained clusters by applying K-means are not meaningful in terms of distances between data points. The values of $\hat{\Gamma}$*_dist* indicate the degree of match between the proximity matrices of original data and different projections. The best match is found at Sammon's Mapping.

5.6. Concluding remarks

In this chapter, we proposed a method of evaluating objectively different projections techniques by employing measures from clustering validity assessment. We provided procedures for adapting the existing measures to the evaluation and comparison of projection techniques. We illustrated the application of this method on three known datasets and two randomly generated datasets.

The methods proposed are applicable in two situations. The first situation is when we want to evaluate the extent to which the projected data preserves the clustering structure of the data. In this case, the following measures can be used: *R*, *J*, *FM*, and $\hat{\Gamma}$*_ext* (when the clustering structure is known); and $\hat{\Gamma}$*_int* and $\hat{\Gamma}$*_m* (when there is no information about the clustering structure of the data). The second situation is when we want to evaluate the degree of match between the proximity matrices of the original and projected data. In this case, the measure used is $\hat{\Gamma}$*_dist*.

The results of the evaluations show that the performances of different projection techniques depend on the original datasets but, generally, the SOM, Sammon's Mapping and PCA provide the best results on the datasets under analysis.

Regarding the UE process model (Table 7), our research was concerned with steps 4, 5, 8 and 9 in that model. We selected the effectiveness characteristic, and the accuracy attribute of a visualization technique (4). We selected a number of clustering validity measures to be used in the measurement and adapted those measures to the evaluation/comparison of projection techniques (5). We selected the simulation method and designed our experiments accordingly (8). We measured the clustering validity of the projection techniques' outputs (9).

Regarding the resources needed in evaluation, the evaluator needs benchmark datasets (real or artificial). The time to obtain the results is very low when no hypothesis testing is required.

A limitation of the research is that the calculation of the indices implies a series of decisions about the parameters of the projection techniques and clustering technique. These decisions may influence the results of the evaluation, issue that we did not address in our papers and in this chapter. This aspect can be studied further.

6. Multiple multidimensional visualization techniques for financial benchmarking

6.1. Research problem description

The research problem in this chapter is to provide an initial evaluation and comparison of nine visualization techniques as to their effectiveness in financial benchmarking (RQ2). The focus is on evaluating *effectiveness* (what patterns can be revealed by the visualization, which a user can perceive?) and *expressiveness* (can all the available data be displayed?). The evaluation approach is based on the *inspection method* (i.e., expert inspection of the visualization output). The evaluation is carried out using a real dataset concerning the financial performance of companies in pulp and paper industry worldwide. The financial benchmarking domain is described in brief in Section 6.2.

In approaching the research problem, we follow the Soukup and Davidson (2002) model of translating a business problem into a visual representation. The authors recommend first translating the business problem into business questions, and the business questions into data mining tasks. Then different visualization techniques are employed and those that are suitable for the data mining tasks are selected. Keim and Kriegel (1994) and Hoffman (1999) follow a similar model of evaluating visualizations: they examine the effectiveness (capabilities) of visualizations for answering different data mining tasks for different datasets.

Bertin (1981) uses the concept of *information level* in order to define the *visual efficacy* (i.e., effectiveness) of a visualization technique. The visual efficacy of a visualization technique is given by the level of information provided (i.e., the level of question that receives an answer).

Thus, based on the Soukup and Davidson (2002) model and the Bertin (1981)'s definition of visual efficacy, we can say that the more data mining tasks (questions) a visualization can solve (answer), the more visually effective it is. Therefore, in this chapter, we evaluate and compare the capabilities of the techniques for solving the data mining tasks which are derived for the financial benchmarking problem. The evaluation is *subjective*, based on the author's assessment of the capabilities of the techniques (Paper 4). We also analyze the visualization techniques in terms of *expressiveness*, i.e., to what extent the visualizations represent graphically all the variables.

6.2. Financial benchmarking

Financial benchmarking is concerned with comparing the financial performance of competing companies (Bendell et. al 1998; Back et al. 1998, 2000; Eklund et al. 2003; Eklund 2004). A manager can be interested in benchmarking for various reasons, especially to improve the performance of his/her company. For solving this problem it is important to decide on the performance measures used for comparison. The performance measures typically used are quantitative financial ratios. Another decision regards the selection of the companies to be compared.

We are not concerned here with selecting variables to be used in a comparison or companies to be analyzed. In this study, we use the data model proposed by Eklund (2004). He constructed a model for financial competitor benchmarking in the pulp and paper industry, with seven financial ratios as a basis for companies' performance comparison.

Eklund uses the Self-Organizing Map (SOM) as the method for data analysis and visualization. We extend the mentioned research and select for data analysis the SOM and other eight data visualization tools. We aim to evaluate the strengths and limitations of these visualization techniques in solving the financial benchmarking problem.

6.2.1. The dataset

The data refer to 80 companies that function in the pulp and paper industry worldwide. The data were collected from the companies' financial reports published on the Internet. The data collection process is described in detail by Eklund (2004). In our study, we use only a subset of the entire dataset, namely the financial ratios of these companies observed during 1997 and 1998. A total of 160 companies are analyzed.

The dataset contains seven numerical financial ratios that characterize the financial performance of companies in the pulp and paper industry. The ratios are grouped in four categories: *profitability* (Operating Margin, Return on Equity, and Return on Total Assets), *solvency* (Interest Coverage, Equity to Capital), *liquidity* (Quick Ratio), and *efficiency* (Receivables Turnover). In the following, we use acronyms when referring to any of the financial ratios (that is, OM, ROE, ROTA, IC, EC, QR, and RT respectively). The dataset contains also three categorical variables: Companies' name, Region (Europe, Northern Europe, USA, Canada and Japan), and Year (1997 and 1998).

Table 12 shows the elements of the data table for the financial benchmarking problem. In this data table, the companies are sorted by Year, then by Region, and within each region they are sorted in alphabetical order.

Table 12. The data table corresponding to the financial benchmarking dataset

Company name	Cartiere Burgo	…	Ahlstrom	…	Tokai Pulp and Paper
Year	1997	…	1997	…	1998
Region	Europe	…	N. Europe	…	Japan
ROE	9.48	…	12.88	…	1.32
ROTA	9.41	…	7.74	…	3.04
OM	5.28	…	6.63	…	4.17
IC	4.10	…	2.55	…	2.02
EC	53.12	…	28.97	…	22.99
QR	1.38	…	0.73	…	0.69
RT	3.42	…	6.03	…	4.07

6.2.2. Business questions and data mining tasks

To be manageable with visualization or VDM tools, any business problem needs to be translated into business questions (Soukup and Davidson 2002). These business questions provide answers that are useful for decision-making. They are not trivial questions that can be answered by using query processing tools. Moreover, Soukup and Davidson (2002) recommend that the derived business questions should be translated into visualization or data mining tasks.

For the problem of financial benchmarking we have derived the business questions and data mining tasks as follows:

a) **Outlier detection**: Do the data contain outliers or anomalies? Are there any companies that show unusual values of financial ratios? This question is important because extreme values of the financial ratios can show some useful information for the decision-maker, but also they can be the result of errors in data collection. In the second case, the outliers identified must be removed from the dataset.
b) **Dependency analysis**: Are there any relationships between variables? Are there financial ratios that are correlated? This question is important because if two or more financial ratios are correlated, a subset of these could be used to explain the remaining ones.
c) **Data clustering**: Are there clusters (groups of companies with similar financial performance) in the data? How many clusters exist? This question is important in order to identify groups of companies with similar financial performance, for example, a cluster with the best performing companies.

d) **Cluster description**: What are the characteristics of each cluster? This question is important because by having the possibility to describe each cluster in terms of financial performance, a manager can compare the companies of interest by comparing the clusters in which they belong.
e) **Class description**: Are there any relationships (common features) among companies located in one region or another? What are these common features? This question can be useful when companies in one region or another are distinguished by certain characteristics specific to their region.
f) **Comparison of data items**: Compare two or more companies with respect to their financial performance. This question is important for finding out the similarities and differences between two or more companies.

For the task f), we have chosen three companies from the dataset to be compared according to their financial performance in 1998: Reno de Medici, Buckeye Technologies, and Donohue. For Reno de Medici we look also at its evolution from 1997 to 1998. These companies are identified on the graphs using the letters A, B, C, and D, respectively. Table 13 presents the financial ratios of these companies.

Table 13. Financial ratios of the companies chosen for comparison

Company	**Reno de Medici**	**Reno de Medici**	**Buckeye technologies**	**Donohue**
Id.	A	B	C	D
Year	1997	1998	1998	1998
Region	Europe	Europe	USA	Canada
OM	4.02	6.7	19.42	21.24
ROE	-15.38	5.34	38.96	17.96
ROTA	0.64	5.27	16.21	15.92
EC	27.94	28.19	20.91	46.35
QR	1.29	1.03	1.36	0.91
IC	0.15	1.68	3.28	5.15
RT	3.3	2.63	7.79	7.96

In the following, in tables and charts, the data mining tasks (a – f) are referred to with shorter terms such as Outliers, Relationships, Clusters, Cluster description, Classes (that is, distinction of classes), Class description, and Comparison, respectively. The task e) is divided into Classes (distinction of classes) and Class description.

6.3. Multidimensional visualization techniques for financial benchmarking

The selection of the visualization techniques depends on the number and types of the variables and cases in the data table (Bertin 1981; Card et al. 1999). Our data is tabular and not hierarchical, and consists of seven variables of numerical or quantitative type. Therefore, *multidimensional data visualizations* are suitable for processing the data. In addition, because the dataset is not very large (i.e., 160 cases), we do not employ the icon-based and dense-pixel display techniques.

We have selected for evaluation the following multidimensional data visualization techniques (Table 6):

- Variations of standard 2D displays: Multiple Line Graphs, Permutation Matrix, Survey Plot,
- Geometrically transformed displays: Scatter Plot Matrix, Parallel Coordinates, Principal Components Analysis (PCA), Sammon's Mapping, and the Self-Organizing Map (SOM).
- Stacked displays: Treemap.

All these techniques, except PCA, are also illustrated in (Hoffman 1999; Hoffman and Grinstein 2002) on different datasets. The authors provide a subjective evaluation, based on their experience and assessment, of the capabilities of the techniques for revealing outliers and clusters, and in rule discovery. Hoffman (1999) augments the evaluation with objective measures for characterizing a visualization, but the correlation of the measures with the effectiveness for data mining is not formally demonstrated.

Our aim is to apply these visualization techniques on the financial performance data and examine their capabilities for answering the business questions and data mining tasks formulated in the previous section. Based on the insight given by visualizing the data using different techniques, an initial evaluation of each technique was performed in terms of their suitability for certain tasks. This is summarized in Section 6.4, Table 14. In the following, we briefly describe each technique and the patterns it highlights. In Paper 4, we only discussed two to three ratios for some techniques, due to page limitations. In this section, we provide a complete discussion of the techniques.

6.3.1. Multiple Line Graphs

Line graphs are used for one dimensional data. On the horizontal axis the values are ordered (e.g., time or the ordering of the table) (Bertin 1981). The vertical axis shows the values of the variable of interest. Multiple Line Graphs can be used to show more than two variables or dimensions (x, y1, y2, y3, etc.) (Bertin 1981; Hoffman 1999). Typically, line graphs are used with time-series data for revealing trends and cycles, or with data in which the variable mapped onto the horizontal axis is of numerical type (i.e., real numbers).

However, we map onto the horizontal axis the companies, in the order of their appearance in the data table (as in the example given in Hoffman and Grinstein 2002). On the vertical axis, we represent the financial ratios. Figure 21 shows line graphs for four ratios (OM, ROE, ROTA and EC), observed in 1997 and 1998.

The graph presents companies from different regions (Europe, Northern Europe, USA, Canada and Japan) in different colors. All the data from the data table can be depicted in the graph, but for saving space on the page, we did not represent the IC, QR and RT.

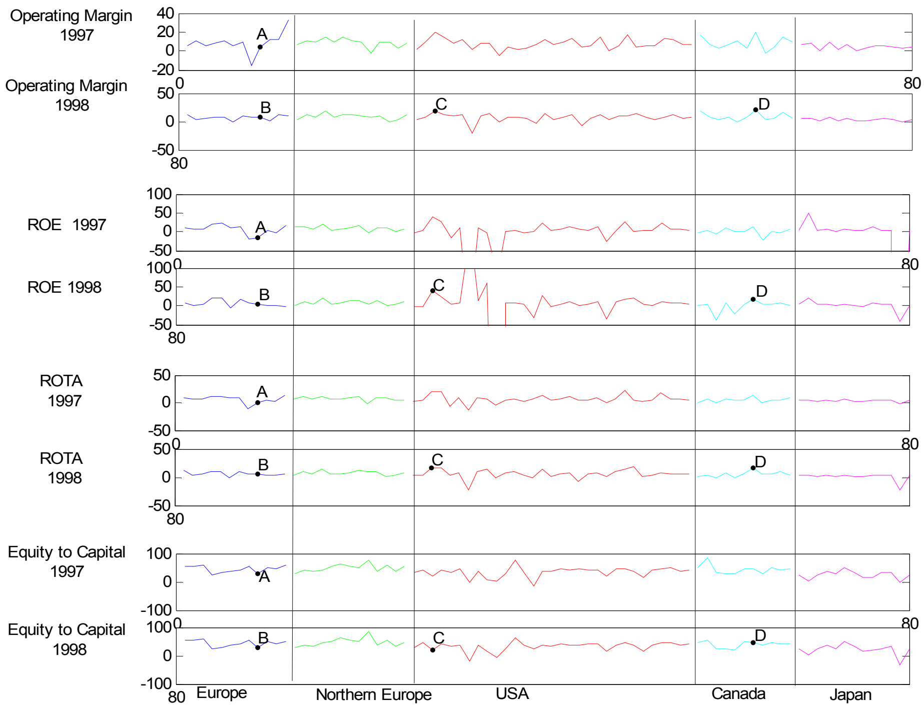


Figure 21. Multiple Line Graphs

This visualization facilitates the detection of outliers in the data, for example, the very low and very high values of ROE for three of the companies, which were further removed from the dataset. It is possible to compare the companies of interest by highlighting them. For example, Donohue 1998 (D) and Buckeye T. 1998 (C) are similar because they have very high ROTA and ROE. In addition, by positioning the two years one under the other, it is possible in principle to follow the evolution of a company's financial ratios. For example, in Europe, one company has in 1997 the lowest OM, but OM of this company increased in 1998. It is also possible to distinguish between companies belonging to different classes (regions), as well as characterizing the companies in one region. For example, companies from Japan have similar financial performance in terms of OM, without having very high or very low values of this ratio. A relationship between ROE 1997 and ROTA 1997 can be observed, especially for companies in Europe.

In summary, Multiple Line Graphs are capable of revealing outliers and relationships, and of distinguishing between classes (regions), characterizing the regions (class description) and comparisons.

6.3.2. Permutation Matrix

The Permutation Matrix technique is a special type of bar graph described by Bertin (1981). In a Permutation Matrix, each data dimension is represented by a vertical bar graph, in which the height of a bar is proportional to the data value for that dimension. The horizontal axes of all bar graphs have the same information (e.g., the time or ordering of the data table). The below-average data values are colored black, and the above-average data values are colored white. A green dashed line plotted over the data represents the average value of each dimension. Implementations of the Permutation Matrix technique allow the interactive changing of the order of the records for observing interesting patterns.

Figure 22 displays a Permutation Matrix created with Visulab (Hinterberger and Schmid 1993). On the horizontal axes the companies are arranged in descending order of ROTA. The companies of interest are highlighted with yellow. The categorical variable Region is not represented in this graph.

This visualization facilitates the detection of outliers (e.g., a very large value of QR). It also reveals relationships between ratios (e.g., ROTA appears to be correlated with ROE, OM and IC). It is also possible to compare the companies of interest (e.g., C and D have higher profitability than B and A).

In summary, Permutation Matrix is capable of revealing outliers, relationships, and comparisons between companies.

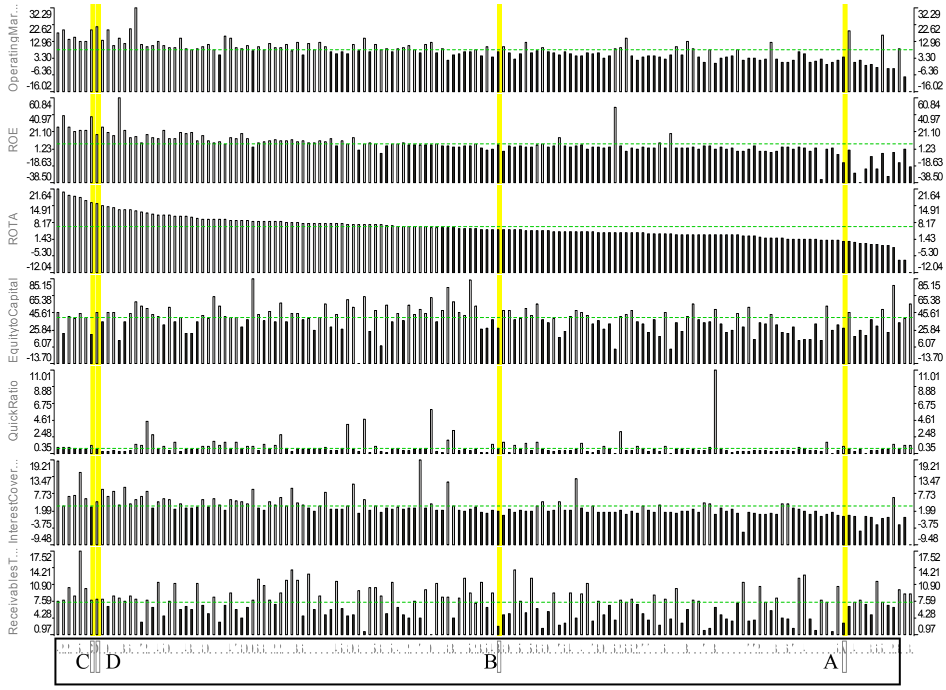


Figure 22. Permutation Matrix created with Visulab

6.3.3. Survey Plot

The Survey Plot technique is a variation of Permutation Matrix (Demsar et al. 2004; Hoffman 1999). Each data dimension is represented by a horizontal bar graph. The length of the bars is proportional to the data values. The bars are centered and there are no spaces separating the bars. One can use colors to distinguish between different classes in the data (if a class variable is present).

Figure 23 displays a Survey Plot created with Orange (Demsar et al. 2004). The data are sorted in the descending order of ROTA. Companies from different regions are displayed with different colors: Europe: blue, Northern Europe: red, USA: green, Canada: violet, and Japan: orange. All the data from the data table are depicted in the graph.

The visualization facilitates the detection of outliers (e.g., a very small value for ROE). It can also reveal relationships between ROTA and OM, ROE and IC. The visualization shows that the companies from Japan are not among the most profitable companies, but have rather medium and low profitability. Instead, many American and European companies display the highest profitability. It is also possible to compare the companies of interest.

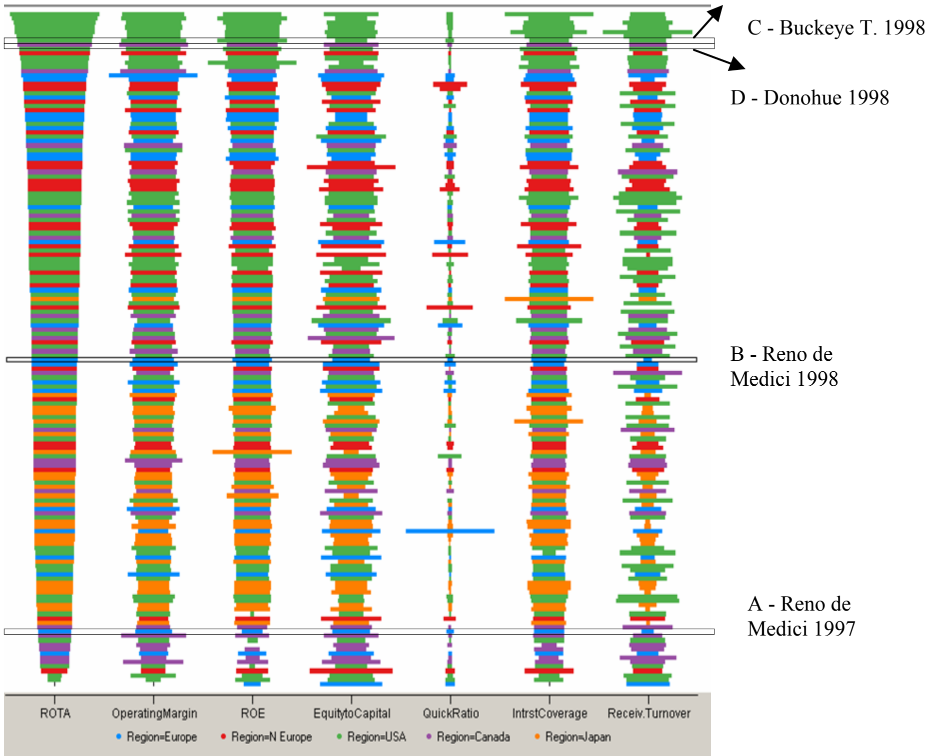


Figure 23. Survey Plot created with Orange

In summary, Survey Plot is capable of showing outliers and relationships, distinguishing between companies in different regions (distinguish classes) and characterizing the companies from one region or another (class description). Survey Plot is also capable of facilitating the comparison of companies.

6.3.4. Scatter Plot Matrix

The Scatter Plot technique is used to plot two-dimensional data so that the horizontal axis shows the values of one variable and the vertical axis shows the values of another variable (Hartwig and Dearing 1990). The Scatter Plot Matrix technique is useful for looking at all possible pairs of variables in a multidimensional dataset (Cleveland 1993).

Figure 24 displays a Scatter Plot Matrix for the financial ratios of the companies. The variable Region is not graphically represented.

The visualization reveals outliers (e.g., a very low value of OM). The visualization reveals that there is no linear relationship between OM and RT, or between OM and QR. There is a linear relationship between ROE and OM, ROTA and OM, etc. The display also facilitates the comparison of companies. For example, Donohue 1998 has the highest OM and EC compared to the other selected companies, but its QR is similar to the others. The positive evolution of Reno de Medici from ’97 to ’98 in terms of profitability (ROE, ROTA and OM) and IC can also be observed.

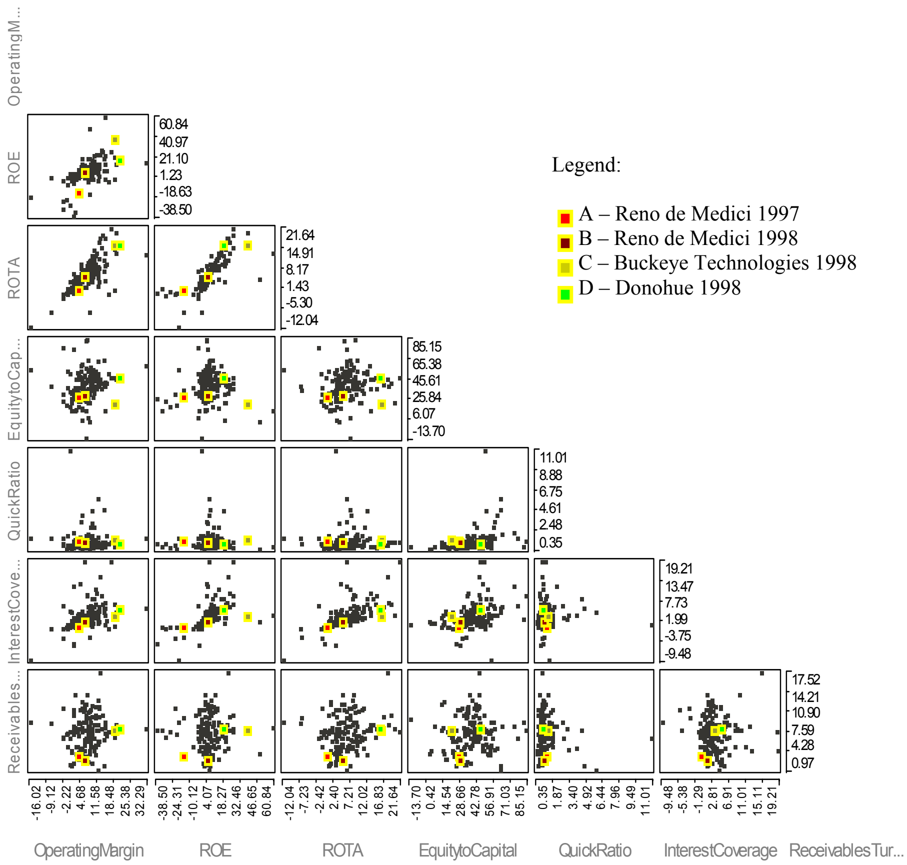


Figure 24. Scatter Plot Matrix created with Visulab

In summary, Scatter Plot Matrix is capable of revealing outliers and relationships, and of facilitating the comparison of companies.

6.3.5. Parallel Coordinates

The Parallel Coordinates technique is introduced by Inselberg (1985; 1997). It is a method to represent multidimensional data using lines. The data dimensions are represented as parallel axes (coordinates). The maximum and minimum values of each dimension are scaled to the upper and lower points on a vertical axis. An *n*-dimensional data point is displayed as a polyline that intersects each axis at a position proportional to the value of the data point for that dimension.

Figure 25 illustrates the Parallel Coordinates technique for our financial data. Each company is represented as a polyline that crosses each axis at a point proportional to the value of the ratio for the corresponding company. The companies of interest are highlighted using different colors. The variable Region is not graphically represented.

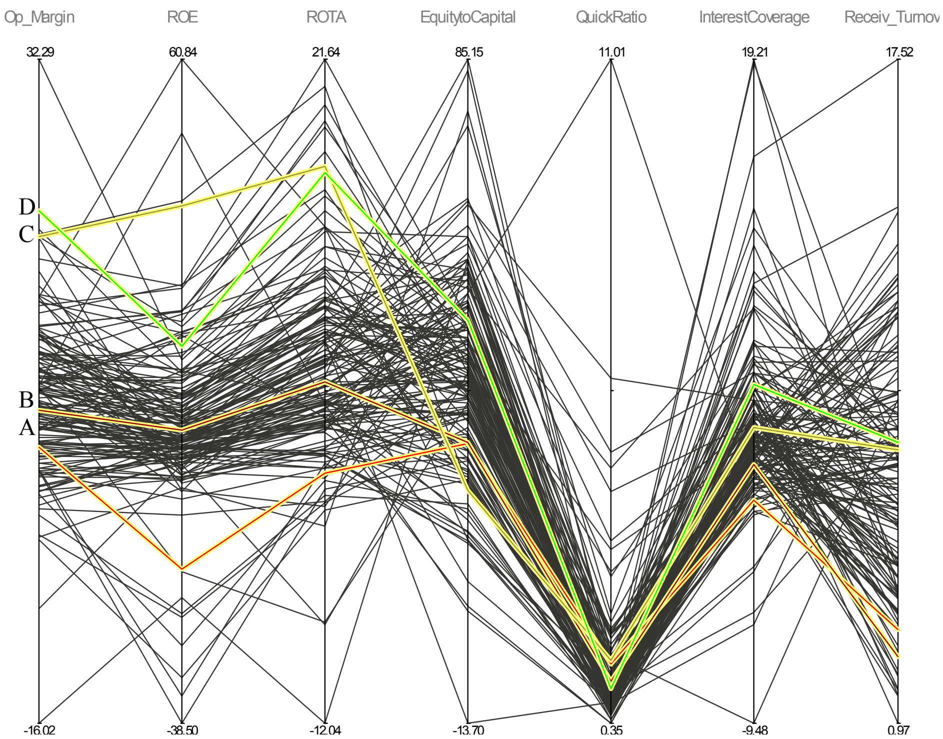


Figure 25. Parallel Coordinates created with Visulab

The Parallel Coordinates technique facilitates the detection of outliers (e.g., an extremely large value of QR, a very low value of OM). It is also possible to compare the companies of interest. Donohue 1998 (D) has the largest OM in the group of companies analyzed. Similar performance is achieved by Buckeye T.

1998 (C), but ROE is larger in the case of Buckeye T. In addition, it is observed that Reno de Medici improved its performance, except QR and RT. All four companies have similar QR. The relationships between two or more variables can be detected if the correlated variables are arranged consecutively (for example, ROE and ROTA, OM and ROE), but in this graph the relationships are not very clear.

In summary, Parallel Coordinates are capable of showing outliers and, to some extent, relationships. They also facilitate the comparisons of companies.

6.3.6. Treemap

The Treemap technique (Johnson and Shneiderman 1991; Shneiderman 1992) is a hierarchical visualization (stacked display) of multidimensional data. Data dimensions are mapped to the size, position, color, and label of nested rectangles.

Figure 26 displays our dataset using the Treemap technique. The figure was created with Treemap 4.1 (2004). Each company is represented by a rectangle. The size of the rectangle indicates the value of RT. The color of the rectangle indicates the value of the ROTA as follows: light green indicates high values of ROTA; light red indicates small values of ROTA; dark red and dark green show values of ROTA close to 14 (see the “color binning” panel in the visualization below). In this visualization, the dataset is organized into categories such as year and region. All the variables can be represented graphically. However, only three variables can be visualized at once (e.g., RT, ROTA and Company’s name), besides the categorical variables Year and Region. A solution can be to transform a numerical variable into a categorical variable and split the existing classes according to the newly formed variable.

The visualization shows in what region the most profitable companies in terms of ROTA are located (the light green rectangles). In addition, one can identify common features or patterns in the industry. Japanese companies have the lowest values of the efficiency ratio (rectangles’ size). One can also compare the financial performance of different companies. Buckeye T. 1998 and Donohue 1998 have higher ROTA and RT than Reno de Medici 1998. Reno de Medici has a higher ROTA in 1998 than in 1997, but a lower RT. Moreover, it is possible to identify outliers (e.g., extremely low and high values of ROTA in USA in 1997).

In summary, the Treemap technique is capable of revealing outliers, distinguishing between classes, describing classes, and facilitating the comparison of companies.

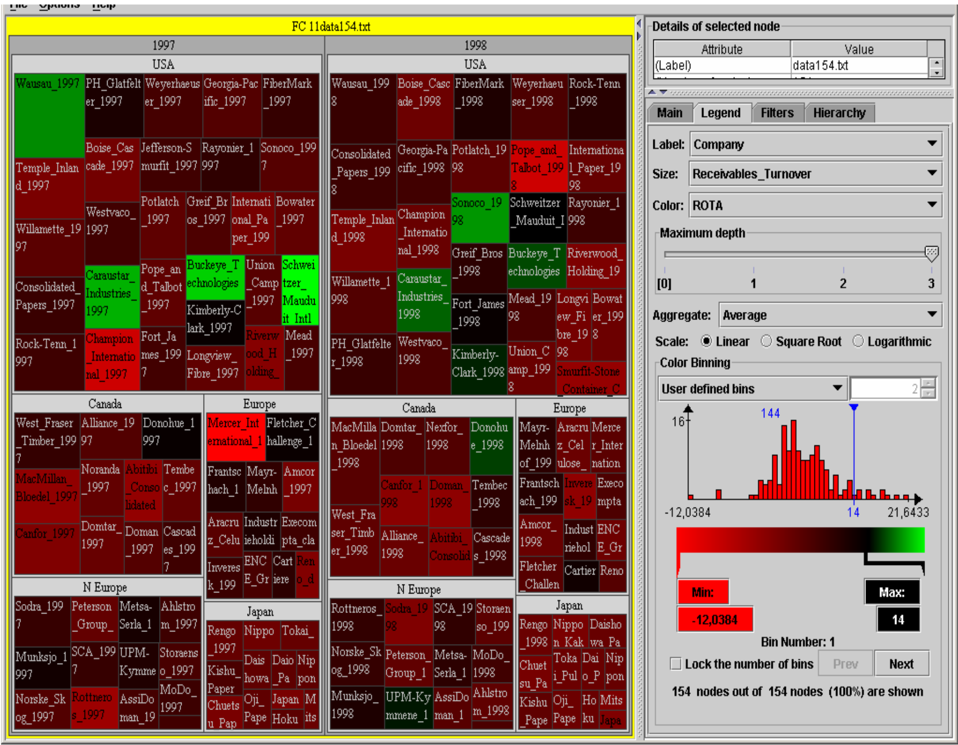


Figure 26. Treemap created with Treemap 4.1

6.3.7. Principal Components Analysis (PCA)

PCA is a classical dimensionality-reducing technique employing linear transformation of data (Sharma 1995; Duda et al. 2000). We have presented the PCA technique in Section 5.2.2.

Figure 27 shows a PCA plot that was constructed from the standardized dataset (the data were normalized using the variance method). The red dot shows the observation closest to the centre of the dataset. The companies of interest are marked with a yellow star and labeled on the graph. The Region variable is not graphically represented.

One can interpret the principal components by inspecting the loadings of each original variable to the PCs. The higher the loading of a variable, the more influence it has in forming the PC score and vice versa. In our case, the first PC (horizontal axis) is highly correlated with the profitability ratios and the IC ratio. Therefore, companies placed towards the right of the horizontal axis have high values of profitability and IC. The second PC (vertical axis) is highly correlated with QR and EC. Companies located on the upper part of the graph have a high liquidity (QR) and high solvency (EC). The amount of variation explained by the two PCs is 40.926% + 19.455%= 60.38% of the total variance. This amount of variance mostly accounts for the variation of six of the ratios; it does not include the variation of efficiency (RT) among the companies.

The visualization reveals outliers (e.g., a very high value of the second PC). It is also possible to compare the companies of interest. Buckeye Technologies 1998 and Donohue 1998 have higher profitability and IC than Reno de Medici 1998. Reno de Medici has improved its profitability compared to 1997, but has lower EC and/or QR. Moreover, the relationships between ratios can be identified by interpreting the principal components. The high correlation of the first PC with all profitability ratios and with the IC ratio indicates that there exists a linear relationship between the profitability ratios and IC. Similarly, the high correlation of the second PC with EC and QR indicates that EC and QR are also correlated.

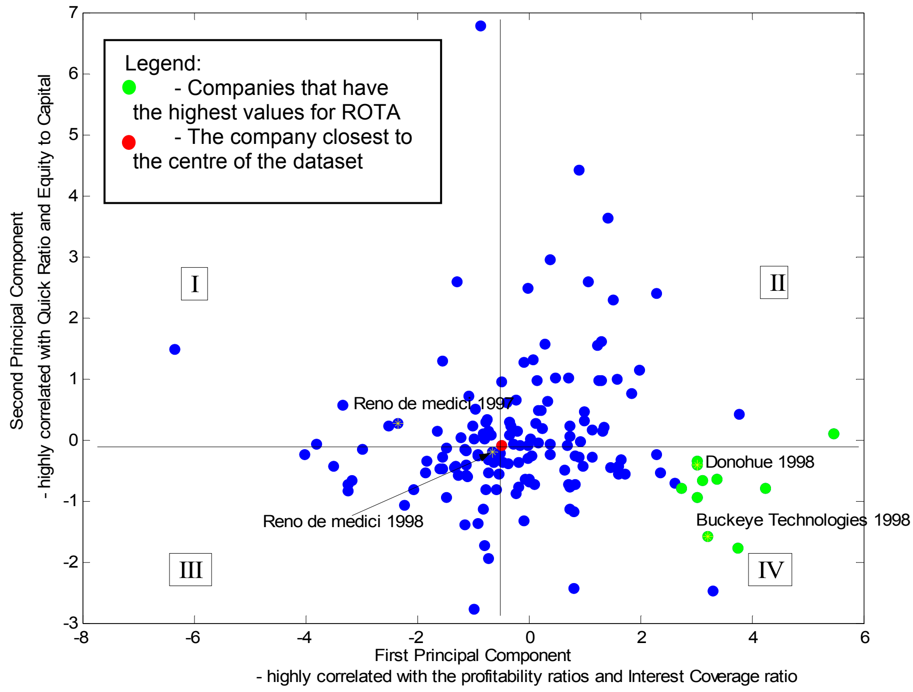


Figure 27. PCA. Data projected on the first two PCs. In area I: medium-high liquidity, low-medium profitability; II: medium-high liquidity, solvency and profitability; III: low-medium liquidity, solvency and profitability; IV: low-medium liquidity, medium-high profitability

By splitting the visual representation in four areas by two orthogonal lines that intersect in the centre of the dataset, one can divide the dataset into four groups of similar observations as shown and described in Figure 27. Based on the meaning of the first two PCs, one can conclude that in area I there are companies with medium-high liquidity and low-medium profitability; in area II, companies with medium-high liquidity, solvency and profitability; in area III, companies with low-medium liquidity, solvency and profitability; and in area IV, companies with low-medium liquidity but medium-high profitability.

In summary, the PCA visualization is capable of revealing outliers, relationships, and clusters. It also enables the description of clusters and the comparison of companies.

6.3.8. Sammon's Mapping

A short description of the Sammon's Mapping (Sammon 1969) has been presented in Section 5.2.3. Before applying the Sammon's Mapping technique to the financial dataset, we have normalized the data using the discrete histogram equalization method.

Figure 28 illustrates the Sammon's Mapping technique for the financial data. Companies from different regions are displayed using different colors. The companies of interest are marked with yellow and labeled on the graph. All the data from the data table are represented. Due to the non-linear transformation of the numerical variables, it is not possible to describe the companies in terms of their financial performance.

The visualization shows that the companies from Canada and USA overlap to some extent and map to the same area of the graph. The companies from Japan, Europe and Northern Europe form three separate groups. However, the degree of overlapping between all these groups is quite high; especially Europe and Northern Europe do not separate well from the other groups. The companies can be compared as to their location on the map, but it is not possible to interpret the similarities/differences between them in terms of financial performance. Thus, Sammon's Mapping is only capable of distinguishing classes.

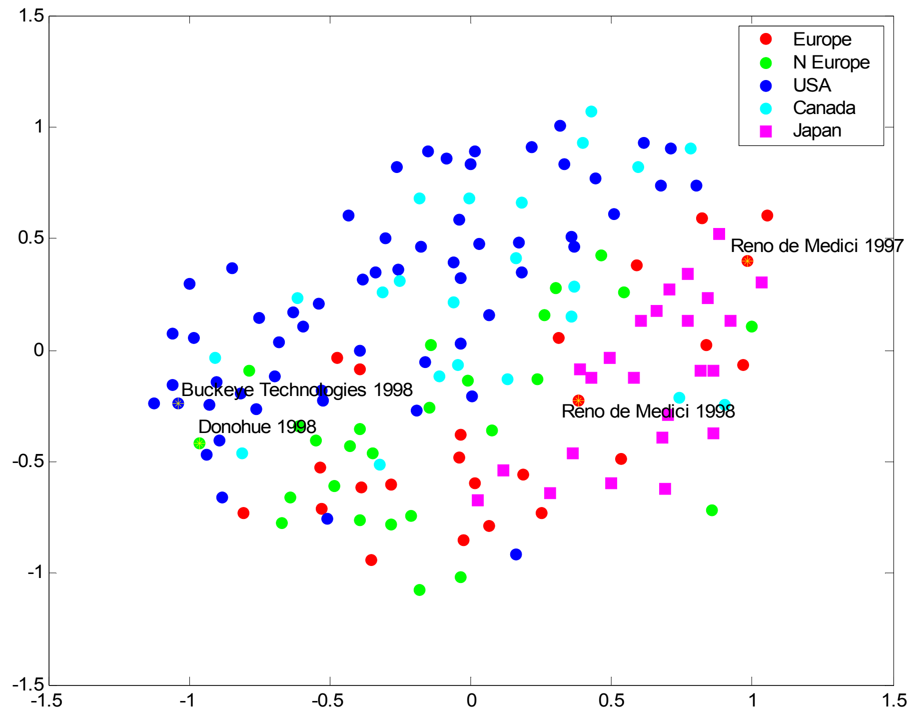


Figure 28. Sammon's Mapping created with SOM Toolbox 2.0 (2005)

6.3.9. Self-Organizing Map (SOM)

A short description of the SOM technique (Kohonen 2001) has been presented in Section 5.2.6. Prior to applying the SOM technique, we have normalized data by using the discrete histogram equalization method. There are many ways to represent the SOM output: SOM–Scatter Plot, SOM–U-matrix, SOM–clustering, SOM–Feature Planes.

The *SOM-Scatter Plot* uses the horizontal and vertical axes produced by the SOM coordinates (i.e., the map size) (Hoffman and Grinstein 2002). The jittering technique is employed to slightly change the positions of the data points.

Figure 29 illustrates the SOM–Scatter Plot view for the financial data. The SOM parameters have been initialized as follows: map size [5, 6], linear initialization, bubble neighborhood, radius [6, 1] and sequential training. The companies from different regions are highlighted using different colors. The entire data table is represented in this graph, but the non-linear mapping of data into the SOM grid makes impossible the description of the companies in terms of their financial performance.

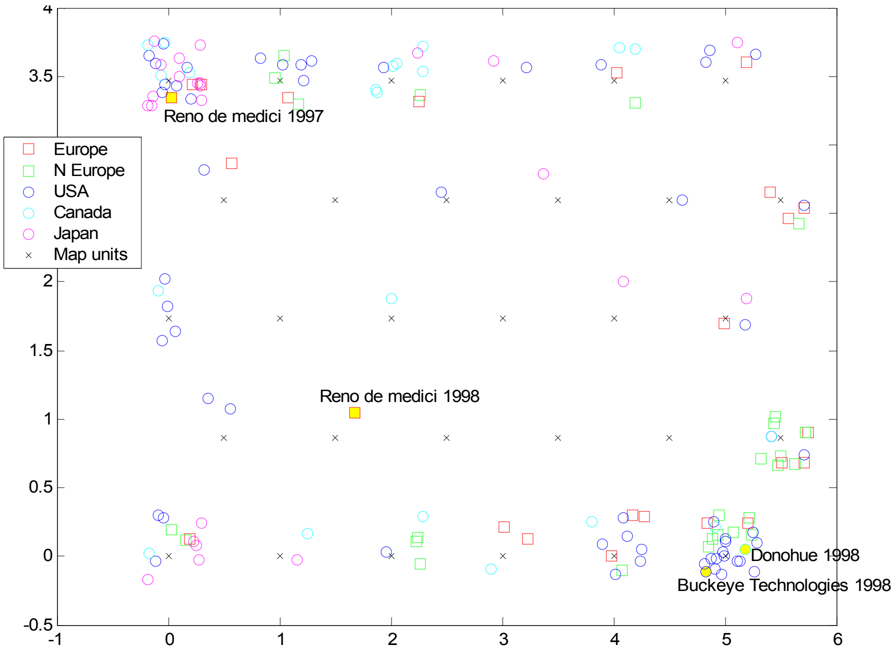


Figure 29. SOM–Scatter Plot view created with SOM Toolbox 2.0

The SOM–Scatter plot view shows many small clusters in the data (the companies that are mapped to the same unit can be interpreted as forming a cluster). One can also observe some isolated companies, which do not belong to a group of companies, and therefore can be interpreted as being outliers. It is possible to distinguish the companies belonging to the different regions, or to identify the position of the companies from one region on the map. However, it is not possible to interpret these classes (regions) and the clusters.

Ultsch and Siemon (1989) developed the *U-matrix graphic display* to illustrate the *clustering of the reference vectors*, by graphically representing the distances between the reference vectors. In the SOM–U-matrix view, each map unit is typically represented by a hexagon. The line or border between two neighboring map-units (hexagons) has a distinguishable color that signifies the distance between the two corresponding reference vectors. Dark green signifies large distances, and light green signifies similarities between the vectors, as indicated by the color bar in the right side of the map (Figure 30). The Region variable is not graphically represented in this map.

By looking at the borders' colors in Figure 30, one can distinguish the main clusters that exist in the data (i.e., the areas delimited by dark lines). However, the SOM–U-matrix does not enable the description of the companies or clusters in terms of financial performance.

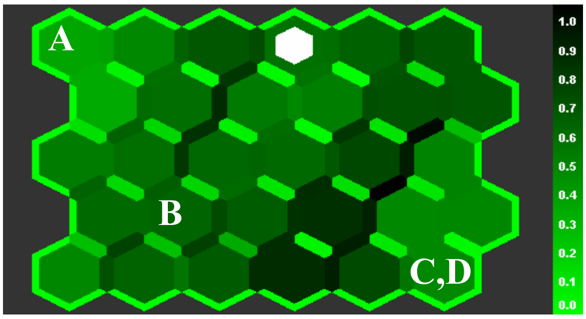


Figure 30. SOM–U-matrix view created with Nenet 1.1 (1999)

A clustering algorithm (e.g., K-means) can be used to automatically partition the map into clusters. This procedure creates the *SOM–Clustering* view (Figure 31). The Region variable is not graphically represented on this visualization.

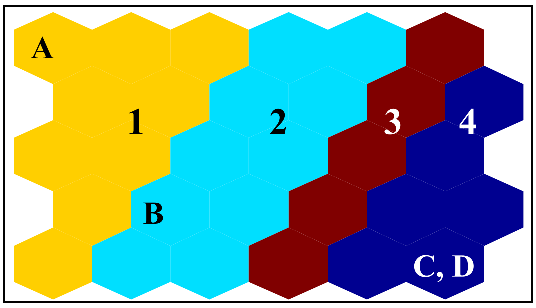


Figure 31. SOM–Clustering view created with SOM Toolbox 2.0

The SOM–Clustering view shows that the dataset contains four main clusters. The description of the clusters in terms of financial performance is not possible.

It is possible to visualize each data dimension using *SOM–Feature Planes*. The Feature Planes graphically display the levels of the variables corresponding to each map unit. The color red signifies high values of the variables, and blue and black correspond to low values of the variables (as indicated by the color bars in Figure 32). The Feature Planes do not graphically represent the Region variable.

The Feature Planes facilitate the comparison of the companies of interest. For example, it can be seen that Buckeye T. 1998 (C) and Donohue 1998 (D) have better financial performance with respect to all ratios, compared to Reno de

Medici 1998 (B). The SOM–Feature Planes also reveal relationships between variables (e.g., OM, ROE and ROTA have the same pattern). Based on the colors of the map units, one can identify clusters and describe the clusters in terms of financial performance (e.g., the blue units in the top-left corner represent a group of companies with very low financial performance).

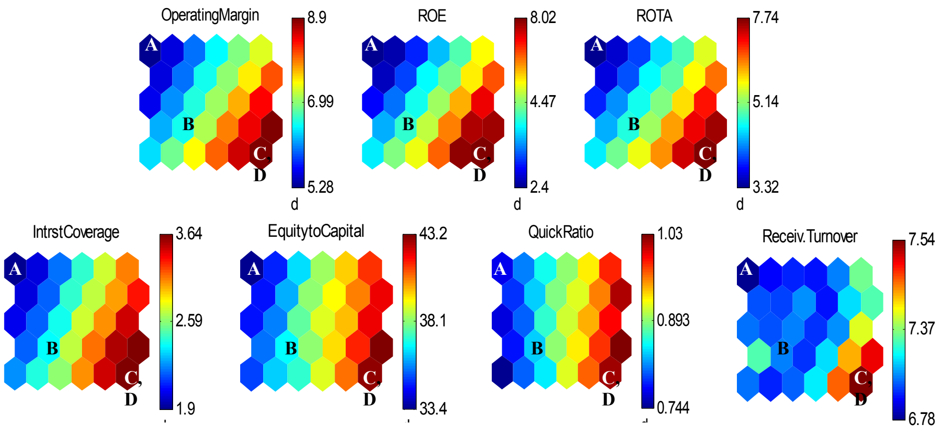


Figure 32. Feature planes created with SOM Toolbox 2.0

In summary, the SOM-based visualizations are capable of showing different patterns in data. SOM–Scatter Plot can convey outliers and clusters, as well as distinguish between classes (regions). The SOM–U-matrix and SOM–Clustering are capable only of showing clusters. The SOM–Feature Planes are capable of showing relationships, clusters, describing the clusters and comparing the companies of interest.

The limitations of a single SOM-based visualization can be overcome by examining in parallel multiple views of the SOM. Thus, by combining the capabilities of all SOM views, it is possible to uncover all interesting patterns in the data. Based on Figure 29, Figure 30, and Figure 31, one can compare the companies of interest with respect to their membership in the identified clusters. Moreover, one can see the composition of each cluster with respect to the variable Region (e.g., Cluster 4 contains mostly American, Northern European and European companies). By examining the SOM–Features Planes in parallel with the SOM–Clustering, one obtains the description of the four clusters as follows. Cluster 1 shows very low profitability, liquidity, solvency and efficiency. It contains the companies with the poorest financial performance. Reno de Medici 1997 is situated in this cluster (A). Cluster 2 shows low to medium profitability, solvency, and liquidity, but low efficiency. Reno de Medici 1998 belongs to this cluster (B). Cluster 3 shows good profitability, liquidity and solvency. Efficiency is medium to low. Cluster 4 shows very high profitability, solvency, liquidity and efficiency. It contains the companies with the best financial performance, among which Buckeye Technologies 1998 and Donohue 1998 are situated (C and D).

6.4. Initial evaluation of visualization techniques

In the previous section, we have highlighted the capabilities of each technique for answering the business questions and data mining tasks related to the financial benchmarking problem. All nine techniques provide an overview of the dataset. The obtained visualizations can convey simultaneously all the companies and most of the variables of the data table. They also give us information about the data structure and particularities. For example, we could observe whether there are clusters in the data, whether the variables are correlated, whether there are outliers, and whether there are similarities and differences between the companies of interest. However, not all techniques offer the same information about the data. Different techniques uncover different patterns in the data.

Below, we evaluate and compare the visualization techniques with respect to four criteria:
1) The capability of the techniques to answer the questions and data mining tasks formulated for the financial benchmarking problem (visual efficacy);
2) The capability of the techniques to show data items (low-level information) or data models (high-level information); and
3) The type of data used as input for the visualization technique (i.e., original data or normalized data).
4) The expressiveness of the techniques (i.e., the extent to which all variables are graphically represented).

Table 14 provides a subjective evaluation of techniques with respect to their capabilities for solving the data mining tasks related to financial benchmarking. The evaluation is based on the previous analysis of the information provided by each graph.

All techniques, except Sammon's Mapping and SOM– (U-matrix, Clustering and Feature Planes) are capable of revealing outliers in this dataset. Scatter Plot Matrix, Survey Plot, Permutation Matrix, PCA and the SOM–Feature Planes are capable of showing relationships between ratios. The SOM and PCA are capable of showing and describing clusters. Treemap, Multiple Line Graphs, and Survey Plot are capable of describing class characteristics. Sammon's Mapping is capable of displaying a class distribution, but does not provide means to describe the characteristics of the classes. Almost all visualization techniques can facilitate the comparison among companies.

Table 14. The capabilities of the visualization techniques on the financial dataset

Technique	*Outliers*	*Relationships*	*Clusters*	*Cluster description*	*Classes*	*Class description*	*Compari -son*
Multiple Line Graphs	✓	✓			✓	✓	✓
Permutation Matrix	✓	✓					✓
Survey Plot	✓	✓			✓	✓	✓
Scatter Plot Matrix	✓	✓					✓
Parallel Coordinates	✓	✓					✓
Treemap	✓				✓	✓	✓
PCA	✓	✓	✓	✓			✓
Sammon's Mapping					✓	*	
SOM Scatter Plot	✓		✓		✓	**	
SOM U-Matrix			✓				
SOM Clustering			✓				
SOM Feature Planes		✓	✓	✓			✓
SOM All Views	✓	✓	✓	✓	✓	✓	✓

* Sammon's Mapping is capable of organizing the dataset so that different classes are distinguishable but does not provide a means to describe the classes.

** SOM–Scatter Plot view is capable of showing where the companies from different classes (regions) are mapped but does not provide a means to describe the classes.

Table 14 also shows that the techniques that uncover most of the patterns in this dataset are Survey Plot, Multiple Line Graphs, PCA, and SOM–All Views. In addition to these, Treemap and SOM–Feature Planes reveal four types of patterns. If we assess separately each SOM-based visualization technique, the results show that each SOM view shows the clustering of the data, but the other patterns are uncovered to a different extent by each SOM view. Multiple Line Graphs and Survey Plot reveal the same types of patterns. The PCA and SOM–Feature Planes reveal almost the same types of patterns; however the information that each of these techniques uncovers can be different. First of all, PCA does not provide information about the RT ratio. Secondly, PCA and SOM–Feature Planes reveal two different clustering structures of the dataset. Moreover, PCA can reveal outliers. Treemap is especially capable of revealing outliers, comparisons between companies and description of classes. However, these tasks can also be answered by the Survey Plot technique. Therefore, we consider that for the financial benchmarking dataset, the SOM–All Views, Survey Plot and PCA are the most suitable visualization techniques.

Secondly, we evaluate and compare the visualization techniques with respect to their capability for showing *data items* or *data models*. All techniques display the data items. The SOM also displays a data mining model (e.g., the clustering of the data). When only the data items are represented, the user has to use his/her perceptual abilities to distinguish the patterns of interest. When a data model is represented, this is automatically generated and explicitly displayed by the visualization technique.

Thirdly, we evaluate and compare the visualization techniques with respect to the type of data processed. The following visualization techniques represent the original data: Multiple Line Graphs, Permutation Matrix, Survey Plot, Scatter Plot, Parallel Coordinates, and Treemap. The other techniques represent normalized data: PCA, Sammon's Mapping, and SOM. The visualizations obtained using normalized data can be more difficult to interpret.

Fourthly, the techniques can be evaluated and compared as to their *expressiveness* (i.e., the extent to which all variables are involved in the construction of the visual representation of the data) (Table 15).

Table 15. The utilization and display of the variables

Technique	*ROTA*	*OM*	*ROE*	*IC*	*EC*	*QR*	*RT*	*Name*	*Year*	*Region*
Multiple Line Graphs *	++	++	++	++	++	++	++	+	++	++
Permutation Matrix	++	++	++	++	++	++	++	+	+	+
Survey Plot	++	++	++	++	++	++	++	+	+	++
Scatter Plot Matrix	++	++	++	++	++	++	++	+	+	+
Parallel Coordinates	++	++	++	++	++	++	++	+	+	+
*Treemap***	++	-	-	-	-	-	++	++	++	++
PCA ***	++	++	++	++	++	++	+	+	+	+
Sammon's Mapping	+	+	+	+	+	+	+	+	+	++
SOM Scatter Plot	+	+	+	+	+	+	+	+	+	++
SOM U-Matrix	+	+	+	+	+	+	+	+	+	+
SOM Clustering	+	+	+	+	+	+	+	+	+	+
SOM Feature Planes	++	++	++	++	++	++	++	+	+	+
SOM All Views	++	++	++	++	++	++	++	+	+	++

Legend:
++ Variables are utilized to construct the graph and are also displayed explicitly.
+ Variables are utilized to construct the graph but are not displayed explicitly. In an interactive situation, they can be displayed at the user's request.
- Variables are not utilized to construct the graph.

Notes:
* For Multiple Line Graphs, we have used all the quantitative variables, but, to save space, we did not display them all.
** With Treemap, the user can interactively change the variables used to construct the graph. However, only five variables can be displayed at a time, unless numerical variables are transformed into categorical variables and used for constructing the hierarchy.
*** With PCA, all numerical variables in the dataset are used to obtain the principal components. However, the RT ratio was mainly represented by the third PC, which was not displayed graphically, because it would have required a 3D graphic display.

6.5. Discussion

The evaluation and comparison of the techniques in Section 6.4 have highlighted the strengths and weaknesses of the visualization techniques. First, we have identified the capabilities of each technique for solving the DM tasks interesting for the financial benchmarking problem (Table 14). The results of the evaluation are useful for selecting a subset of techniques that can together address all of the DM tasks. We have identified this subset as consisting of the SOM–All views, Survey Plot and PCA.

The initial evaluation and comparison of the techniques enable also to identify aspects of the existing visualizations that can be improved. For example, Table 15 shows that Region variable is not explicitly represented in most of the visualizations, therefore it was not possible to distinguish between classes on those displays. This weakness could be addressed by providing color coding for the Region variable. Moreover, there are visualizations (e.g., Sammon's Mapping, SOM–Scatter Plot) in which the numerical variables are not explicitly displayed, therefore it was impossible to interpret the visualizations in terms of financial ratios. This weakness could be addressed by linking the visualization to other visualizations (e.g., Feature Planes, in the case of SOM).

In addition, there are visualizations that are not capable of displaying clusters (e.g., Permutation Matrix, Survey Plot). This weakness could be addressed by augmenting the visualization technique with analytical tools (e.g., clustering algorithms) and then arrange the data items according to the clustering structure found in the data or using color coding to distinguish between clusters. Similar approach can be used for revealing relationships in the data. For example, the Parallel Coordinates technique could be augmented with analytical tools for discovering the relationships between ratios and the possibility to change the order of the axes so that correlated ratios are placed one near the other.

There are some techniques that require the normalization of the data (PCA, Sammon's Mapping, and SOM). This fact makes the interpretation of the visualization to be more difficult. In these cases it is important to have the possibility to see the real values of the financial ratios of the companies of interest (for example, the SOM-Feature Planes use color bars that indicate the values of the financial ratios corresponding to different colors).

The normalization of the data is recommended when applying the PCA, SOM and Sammon's Mapping techniques, otherwise the variables with higher values have more weight than the other variables. The histogram equalization method has the particularity that it also smoothes out the univariate outliers. However, after trying other normalization techniques in the case of SOM and Sammon's Mapping, we have chosen the histogram equalization method because it determined the best representations of the data.

6.6. Concluding remarks

In this chapter, we have illustrated the use of different multidimensional visualization techniques for exploring financial benchmarking data. We provided an initial evaluation and comparison of nine techniques: Multiple Line Graphs, Permutation Matrix, Survey Plot, Scatter Plot Matrix, Parallel Coordinates, Treemap, PCA, Sammon's Mapping, and the SOM.

The evaluation framework was based on the model described by Soukup and Davidson (2002) of mapping a business problem to business questions, data mining tasks, and visualization techniques.

The evaluation shows that no visualization technique is capable alone of answering all DM tasks, and therefore, multiple visualizations are more suitable to be implemented in a benchmarking tool for business. The need of using multiple visualizations due to the limitations of a single technique is also emphasized by other evaluations performed on artificial or benchmark datasets for different types of tasks (e.g., Keim 1996; Hoffman 1999).

Our evaluation has yielded that the SOM–All Views, Survey Plot and PCA are among the best techniques capable of revealing interesting patterns in the dataset under analysis. The results also show that the use of all SOM views (especially the Feature Planes together with one of the other SOM views) is highly effective for financial benchmarking.

A limitation of this initial evaluation is that it is subjective and it is based only on the author's experience of the visualization techniques and dataset. Collecting data from several users would provide more confidence in the results.

Regarding the UE process model (Table 7), this chapter contributes with providing guidance on steps 4, 5, 8 and 9 in that model. We evaluated the visualizations with respect to their *effectiveness* (*visual efficacy)* and *expressiveness* in presenting the information (4). The measures of these attributes were of *qualitative nature*, capturing whether or not the techniques were capable of providing answers to the data mining tasks formulated for the financial benchmarking problem, or whether all variables are represented graphically, respectively (5). We chose to perform a *subjective* evaluation, using an *inspection method* (8). The actual evaluation consisted of examining the visualizations, after interacting with the tools (9).

7. User evaluation of multiple multidimensional visualization techniques for financial benchmarking

7.1. Research problem description

The research problem in this chapter is the evaluation of different visualization techniques as to their effectiveness in financial benchmarking (RQ2). We extend the evaluation described in Chapter 6 by *involving users* in the evaluation. We follow the same model of deriving data mining tasks from the financial benchmarking problem and mapping the data mining tasks to visualization techniques.

The visualization techniques under evaluation are the nine techniques described in the previous chapter: Multiple Line Graphs, Permutation Matrix, Survey Plot, Scatter Plot Matrix, Parallel Coordinates, Treemap, Principal Components Analysis (PCA), Sammon's Mapping, and the Self-Organizing Map (SOM).

The tasks in which these techniques are evaluated are the data mining tasks derived for the financial benchmarking problem. They are described in detail in the previous chapter: outlier detection, dependency analysis, clustering, cluster description, description of classes and comparison between companies.

The evaluation regards the *effectiveness* of the visualization techniques for solving the financial benchmarking problem. Similar to the approach in Chapter 6, we regard effectiveness as *visual efficacy* (Bertin 1981), that is, the level of question that receives an immediate answer (i.e., patterns that can be identified).

In addition, Mackinlay (1986) points out that effectiveness of a visualization can be regarded from different perspectives: enabling the user to read, understand and interpret the display easily, accurately, quickly, etc. In other words, Card et al. (1999) say that a visual mapping is more *effective* than other visual mapping if "it can be perceived better by a human, i.e., it is faster to interpret, can convey more distinctions, or does lead to fewer errors in interpretation". In our evaluation approach, we also regard effectiveness in terms of *correctness* of interpretation (the visualization does lead to fewer errors in interpretation). In the following, we propose an inquiry evaluation technique based on questionnaire data collection. The ideas and one evaluation study in this chapter are also presented in Paper 5.

7.2. Evaluation method

The basic idea of the proposed evaluation technique is to ask users to identify interesting patterns in the data by examining different visual displays. In evaluating the *effectiveness* of the visualization techniques for revealing interesting patterns, we focus on two aspects of the visualization: the correct interpretation of the visualization by the users (i.e., *correctness of interpretation*), and the number of distinctions or patterns that the visualization can convey (i.e., *visual efficacy,* Bertin 1981). We do not address the interactive capabilities of the techniques and therefore, we focus only on the *static display* of the data.

We consider that the visualization is correctly interpreted by the users if they can identify valid patterns (i.e., users correctly understand and interpret the graphical properties of the visualization when describing the identified patterns). We consider that the visualization conveys many distinctions if many patterns of interest can be identified and described based on visual substrate (axes) and the graphical properties of the technique (color, position, etc.).

7.2.1. Data capture

We develop a questionnaire in order to capture data about how effective a technique is for revealing interesting patterns. The patterns are directly related to the data mining tasks (a-f) derived for the financial benchmarking problem (Section 6.2.2).

For each pattern (data mining task), we create two questions. The first question asks the user whether he/she can identify the pattern by examining the visual display. This question is *pre-coded* with YES/NO answers. The second question requests the user to mark on the graph or shortly describe the identified pattern. This question is *open* question and it ensures the validity of the responses to the first question. By analyzing the answers to the second question we assess whether the user has correctly understood/interpreted the visualization. Thus, the questionnaire contains assessment questions that require YES/NO answers, and task-based questions that require “open” answers.

To avoid bias, we repeat the same questions to each visualization technique, though, it is obviously that some visualizations are not capable of reveling all patterns of interest (e.g., Permutation Matrix does not display information about Region). In this way, we try to ensure that the respondent is not inclined to answer positively all the questions posed for one technique. However, this choice of repeating all questions to all techniques increases the size of the questionnaire.

To illustrate the contents of the questionnaire, Figure 33 presents the questions corresponding to the assessment of the Permutation Matrix technique. Similar questions are created for all the other visualization techniques. The numbering of the questions in Figure 33 is different than the one used in the questionnaire. However, for the data analysis and interpretation we renumber the questions to conform to Figure 33.

1. Can you identify any **outliers** in this dataset by examining the above Permutation Matrix? – YES/ NO
   - If yes, please tell in brief for which ratio you identified an outlier or mark it on the graph.

2. Can you identify any **relationships** between the ratios in this dataset by examining the above permutation matrix? – YES/ NO
   - If yes, please name one pair of ratios that you identified as correlated.

3. Can you identify any **clusters** in this dataset by examining the above permutation matrix? – YES/ NO
   - If yes, please tell how many clusters you identified.

4. Can you **describe the clusters** that you have identified, by examining the above permutation matrix? – YES/ NO
   - If yes, please **mark one of the clusters** and describe it briefly in terms of values for financial ratios. Use Low, Medium, and High for indicating the prevailing level of the ratios in that cluster.

Operating margin	
ROE	
ROTA	
Equity to capital	

5. Can you distinguish between the companies from one region or another? – YES /NO
   - If yes, can you describe the characteristics of companies from Japan in comparison with the other regions by examining the above permutation matrix?

6. Can you **compare** the characteristics of the companies A, B, C and D by examining the above permutation matrix? – YES /NO
   - If yes, please tell in brief how Reno de Medici 1998 (B) performs in comparison with Buckeye Technologies 1998 (C) and Donohue 1998 (D). Also, tell how financial performance of Reno de Medici changed from 1997 (A) to 1998 (B).

Figure 33. Fragment of questionnaire

7.2.2. Data analysis

The collected data are of qualitative nature (that is, YES/NO answers accompanied by descriptions of the identified patterns). We analyze the answers *subjectively*, by evaluating the *correctness* of the answers to the open questions. To obtain a *measure of the effectiveness* of the techniques for certain tasks, we look at the percentage of *positive correct answers*. The positive correct answers are recorded when users have answered that they can identify certain patterns (YES answer) and they also have provided correct answers to the open questions.

7.3. Evaluation studies

We have used the above evaluation method in three different studies. The first study is presented in Paper 5. The other two studies have been conducted in order to examine the applicability of the method to other groups of users. All three evaluation studies have consisted of the following parts:

- Introduction of the study to participants,
- Data collection based on the questionnaire technique, and
- Data analysis and interpretation.

7.3.1. Participants

The participants in all three studies are students. In the first and second study, the participants have been recruited from one advanced-level course in Information Systems. The students have been encouraged to participate in the study by designating the study as a voluntary assignment resulting in extra study points. Hereafter, we refer to these users as Group 1 and Group 2. The students in the third study have been recruited from a course in Human-Computer Interaction, and they have participated in the evaluation as part of an exercise for the course, which was compulsory.

In the first study, we have had access to a group of students who already possessed some experience of the business problem, dataset and one of the visualization techniques, namely the SOM. This experience was acquired from the course from which they have been recruited. We have considered this group as resembling the targeted users of the visualization techniques.

We have conducted the evaluation of the techniques on other two groups of users, groups that differ from the first group with respect to the experience levels of the financial benchmarking problem and the SOM technique. The summary of the characteristics of the participants in each study is presented in the following. Appendix 4 presents the background information requested of the participants.

Group 1 consists of 12 users, with international backgrounds and majors in Information Systems, Computer Science, Economics, or Business Administration. These students have experience of analyzing the financial benchmarking dataset and of the SOM. This group can be considered to have more experience of data analysis and more domain knowledge than any of the other groups analyzed. Their answers are presented in detail in Paper 5.

Group 2 consists of 13 students recruited from the same course as that in Group 1, but in a different academic year (2007) and before they started working with the financial benchmarking data and the SOM. These students have international backgrounds and majors in Information Systems, E-Commerce, Computer Science, Information Technology, or Management Science. They have theoretical knowledge of the SOM and the dataset, but they do not have experience of working with them.

Group 3 consists of 27 students with majors in Information/Data Processing, Information Security, Computer Science, Computer Engineering, Digital Media, Information Technology, or Software Production. These students do not have a priori knowledge about the dataset and most of them do not have experience of working with the SOM technique.

Before involving the students in the evaluation, we have presented a short tutorial about information visualization where the visualization techniques and data mining tasks under evaluation have been briefly illustrated. There are participants in Groups 1 and 3 that did not attend the tutorial. We have statistically analyzed whether the participation to the introductory lecture influenced the answers, but the collected data did not support this hypothesis.

7.3.2. Data collection

In the questionnaire, we have briefly presented the financial benchmarking problem, the dataset and the data mining tasks derived for solving the business problem. We have also provided short descriptions of the visualization techniques, which were followed by the corresponding graphical representations of the data (the same images as the ones in Figure 21 - Figure 32). After each graphical representation we have included the set of questions (i.e., Figure 33). In the questionnaire, the visualization techniques are presented to all participants in the same order (the order was identical to the presentation in Section 6.3).

Most of the respondents provided their answers during the class (Groups 2 and 3, and four participants in Group 1). However, some participants in Group 1 have answered by e-mail. We did not record for each participant the duration to answer a question or the entire questionnaire. The users have returned the questionnaires in 30 minutes to 2 hours. For the participants in Group 1 that answered by e-mail, the time it took them to answer the questionnaire is not available.

We have been interested in the number of distinctions (patterns) that can be identified by users and the correctness of the answers provided.

7.3.3. Data analysis

The analysis of the collected data yielded four categories of answers: *positive*, *negative*, *invalid*, and *non-response*:

- **Positive** answers: The user has answered *positively* (i.e., YES), that he/she could identify certain patterns, and the explanation or illustration of the patterns identified were *correct*.
- **Negative** answers: The user has answered *negatively* (i.e., NO), that he/she could not identify certain patterns.
- **Invalid** answers: The user answered *positively* (i.e., YES) that he/she could identify certain patterns, but the explanations or illustrations of the patterns were *incomplete, incorrect* or *not clear*.
- **Non-responses**: The user did not answer specific questions addressing a task or an entire set of questions regarding a visualization technique.

We have interpreted the percentage of positive answers as the extent to which a visualization technique is effective for uncovering certain patterns. Thus, a visualization technique is effective for some task if most of the users correctly identify the patterns corresponding to that task. The percentage of negative answers indicates the extent to which the visualization technique is not effective for highlighting the patterns of interest. The invalid answers might have been determined by a misunderstanding of the question, task, or visualization. Therefore, they are more difficult to interpret. However, only a relatively small number of invalid answers and non-responses have been recorded.

As was previously mentioned, the effectiveness of a visualization technique for revealing patterns of interest has been measured by the percentages of positive answers. The assessment of the answers' correctness was based on evaluating whether or not the responses indicated identification of the patterns and good understanding of the visual representation (the meaning of axes, colors, position, etc.).

Thus, to classify the answers as positive, we have examined whether the users correctly identified patterns (i.e., how the data are organized or structured) and that they correctly understood the means that each visualization technique provides to interpret the patterns (i.e., the visual substrate (axes or other organization) and the graphical attributes such as color, position, size of the marks).

The following criteria have been used to assess whether the answers are correct:
- Outlier detection: correctly identify at least one outlier.
- Dependency analysis: correctly identify relationships between two or more variables.
- Data clustering: specify a number of clusters. Here it was difficult to assess whether the user correctly identified the clusters in the data, because the question did not ask to mark the clusters on the graph, but only to specify the number of clusters detected. However, if from subsequent answers we have observed that the respondents incorrectly identified the clusters, we have considered their answers invalid.
- Cluster description: describe a cluster by using all the variables mentioned in the question.
- Distinguish classes: distinguish the companies from different regions on those graphs in which this distinction was possible to make.
- Class description: describe the companies from Japan in terms of their similarities and their differences to companies from other regions.
- Data comparison: describe the companies of interest in terms of at least one financial ratio. Not all users have described the performance of the companies in terms of all the financial ratios depicted by the visualization, though we have considered their answers as correct.

It is worth mentioning that for the task Distinguish classes, if the user has answered "YES", it was considered positive. If the user has subsequently not answered anything concerning the Class description, we considered it a "NO" answer. This is a limitation of the questionnaire, the same type of problem has been found regarding the Clustering task (when the user is asked to report the number of clusters but not to mark them). This limitation of the questionnaire has made it difficult to interpret the answers to those questions.

7.4. Empirical results

The answers of the participants are summarized as percentages of positive, negative, invalid answers and non-responses. In Paper 5, we presented the results obtained for Group 1. Below, we present comparatively the results for all three groups. Figure 34 - Figure 46 show the percentages of positive answers recorded for each visualization technique in each group of participants.

7.4.1. Multiple Line Graphs

Figure 34 shows the percentage of positive answers recorded in each of the three groups for the evaluation of Multiple Line Graphs. For the tasks outliers, relationships, and class description, Group 1 has provided the largest percentage of positive answers. High percentages of positive answers are also encountered for the distinguish classes task. Thus, for experienced users multiple graphs are quite effective for these tasks. However, in the comparison task the visualization is not very effective and, thus, Multiple Line Graphs are not very suitable for this task. Very many users in Groups 2 and 3 have found outliers. More complex tasks, such as class description and comparison have fewer positive answers in Group 3.

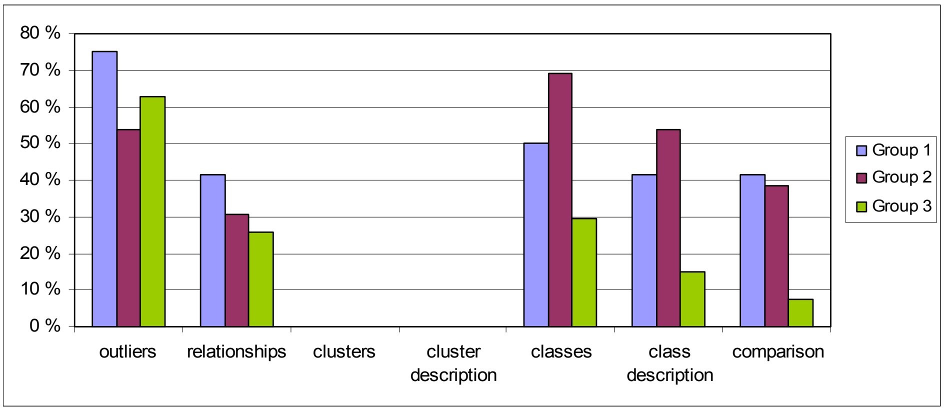


Figure 34. Positive answers (%) for evaluating the Multiple Line Graphs

7.4.2. Permutation Matrix

Figure 35 shows the percentages of positive answers obtained in each group for the evaluation of Permutation Matrix. Group 1 has provided a larger proportion of positive answers for the tasks outliers, relationships and comparisons between companies. One user has identified clusters and has been able to correctly describe one cluster. The proportion of positive answers in Group 3 is almost as high as in Group 2 at identifying outliers and relationships. Some differences between Group 3 and the others groups are noticeable for the task comparison.

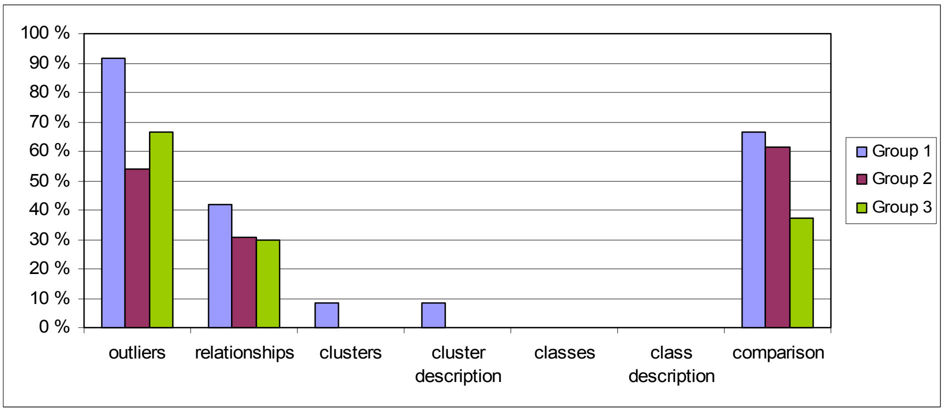


Figure 35. Positive answers (%) for evaluating the Permutation Matrix

7.4.3. Survey Plot

Figure 36 presents the percentages of positive answers obtained for the Survey Plot technique. The results show that Survey Plot is effective for revealing outliers, classes, class descriptions, and comparisons between companies. It also appears that Group 1 has provided a larger number of positive answers for the class distinction, class description and comparison tasks. The Survey Plot technique appears to be effective also when used by novice users.

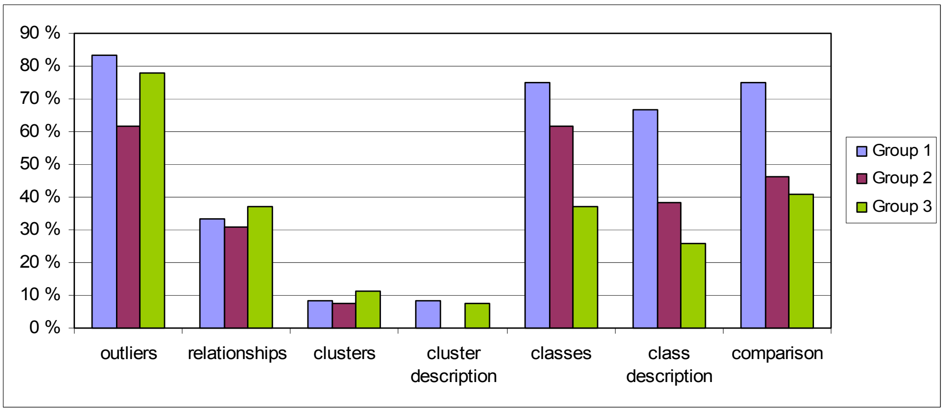


Figure 36. Positive answers (%) for evaluating the Survey Plot

7.4.4. Scatter Plot Matrix

Figure 37 shows that the Scatter Plot Matrix technique is effective for revealing outliers, relationships between variables, and comparisons between companies. The technique's effectiveness varies with the users' experience. For a simple task, such as outlier detection, many users have given correct answers. However, for complex tasks, such as comparison and relationships, the percentages of positive answers in Groups 2 and 3 are smaller.

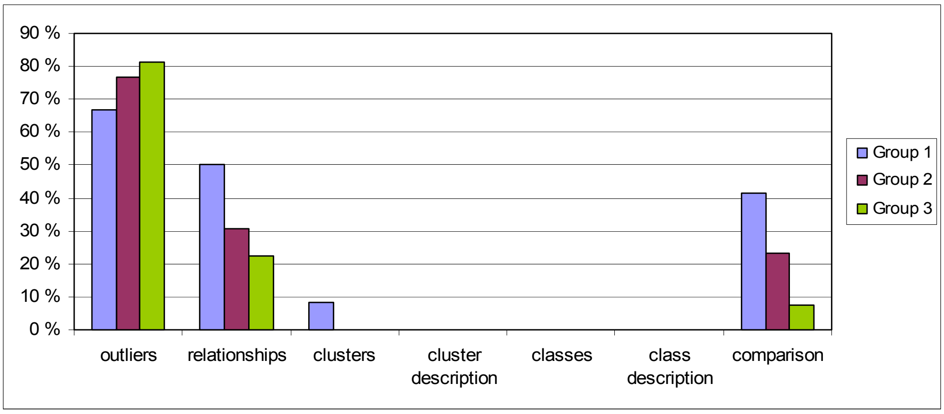


Figure 37. Positive answers (%) for evaluating the Scatter Plot Matrix

7.4.5. Parallel Coordinates

Figure 38 shows that Parallel Coordinates are effective for revealing outliers and comparisons between companies. Few users in Groups 1 and 2 have identified relationships between variables and clusters in the data.

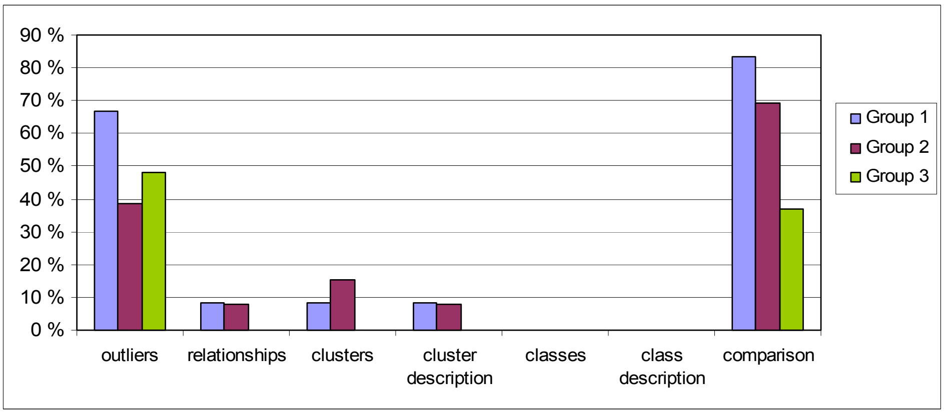


Figure 38. Positive answers (%) for evaluating the Parallel Coordinates

7.4.6. Treemap

The Treemap technique seems effective for uncovering outliers, classes, class descriptions. To a lesser extent, it is also effective for facilitating comparisons between companies (Figure 39). Group 1 has provided the highest percentage of positive answers for these tasks. The lower percentage of positive answers for the comparison task may be partly due to the difficulty of the respondents to identify the companies of interest. Comments such as “too hard to separate” and “very difficult to find companies” support this conclusion. The answers for class description have been considered correct if users have identified similarities among Japan’s companies in terms of RT and ROTA. Other correct answers have expressed that these companies do not have extreme values, but medium or low RT or ROTA in comparison with other regions.

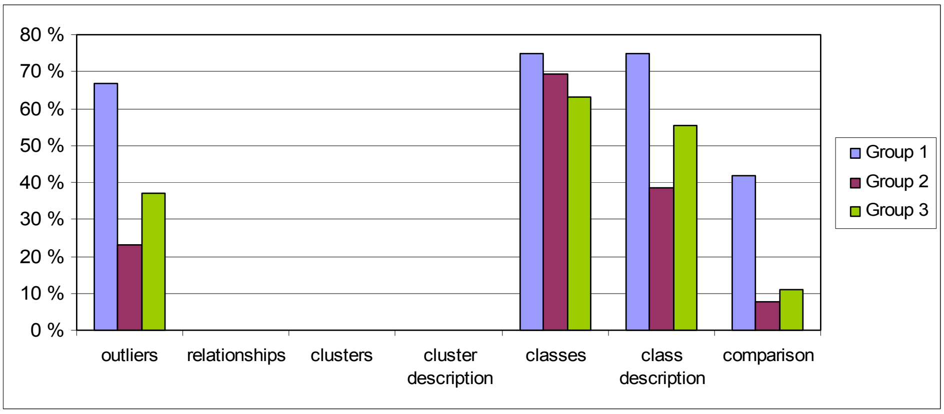


Figure 39. Positive answers (%) for evaluating the Treemap

7.4.7. Principal Component Analysis (PCA)

The PCA technique is effective for revealing outliers and clusters, and for facilitating comparisons between companies (Figure 40). Few respondents have recognized relationships between variables. Group 1 has provided the highest percentage of positive answers for these tasks. For the clustering task, we have considered the answers correct if users have recognized the four areas as clusters, or if they have distinguished between the companies with the highest ROTA and the rest of the companies as representing two different clusters.

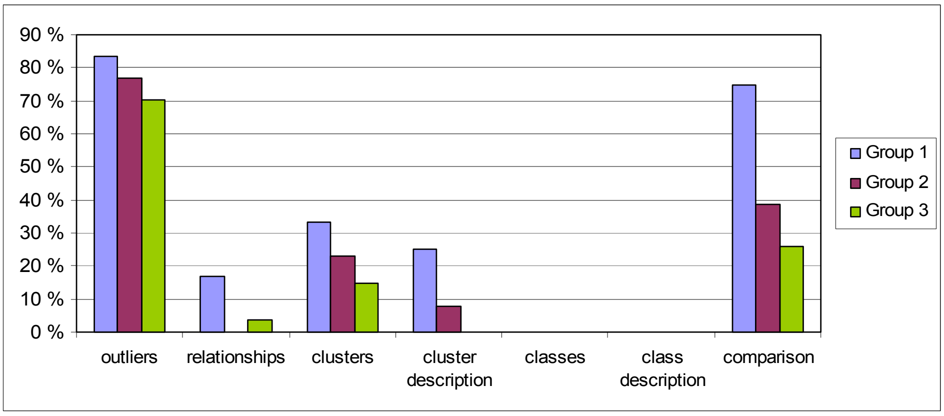


Figure 40. Positive answers (%) for evaluating the PCA

7.4.8. Sammon's Mapping

Figure 41 shows that Sammon's Mapping is effective for revealing classes. Most of the respondents in all groups have answered that they could indeed distinguish the regions. Some respondents in Groups 1 and 3 have also noticed that Japanese companies are more similar to each other than the companies from other regions, and thus forming a "tighter" group than the others regions. We have evaluated this as a positive answer for the class description task. Also, some students have correctly identified outliers based on the color, for example the green and blue marks in the lower part of the map.

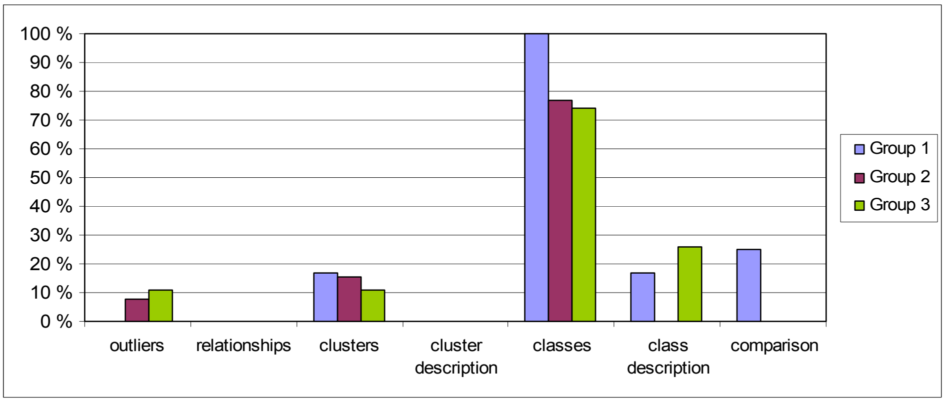


Figure 41. Positive answers (%) for evaluating the Sammon's Mapping

Moreover, some students have identified clusters based on the color (that is, region). We have considered the answer for the clustering task positive if the user has identified two or three clusters (based on the region or color). For the comparison task, answers that have highlighted that the companies of interest have different financial performance have been considered correct.

7.4.9. SOM – Scatter Plot

The SOM–Scatter Plot view is especially effective for revealing clusters (Figure 42). However the number of clusters identified by respondents varies very much, from 1 to 18. We have considered as correct, and thus recorded as positive answers, even those answers which have indicated only one cluster. We took this decision because the question was not very clear, and did not ask to cluster the dataset into disjoint partition but to "identify any clusters". The respondents have indicated the single data points that did not belong to a larger cluster as being outliers. In Group 1, more respondents have distinguished between the companies belonging to different geographical regions. In Group 2 and Group 3, not very many users have positively answered this question. The positive answers for the comparison task correspond to the users' observations according to which the companies of interest belong to different clusters, thus having different financial performance. However, the map does not provide information to assess the companies' performances.

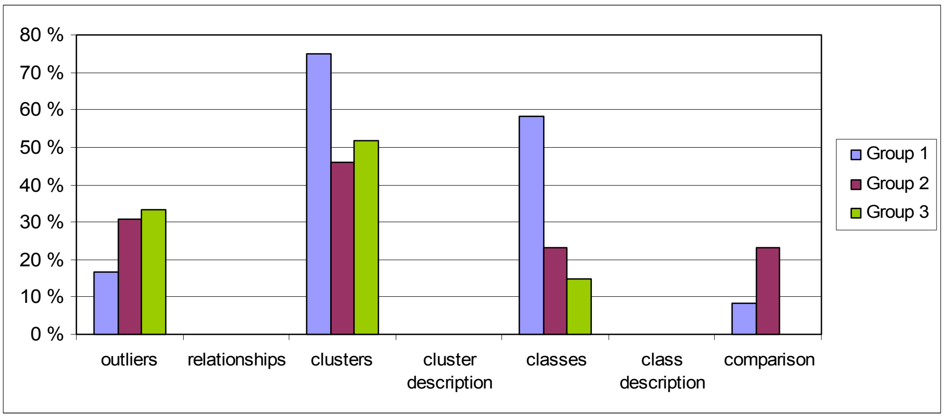


Figure 42. Positive answers (%) for evaluating the SOM–Scatter Plot view

7.4.10. SOM–U-matrix

The SOM–U-matrix view is effective for displaying clusters (Figure 43). For the clusters task, the respondents have provided answers that indicated different numbers of clusters (between one and six clusters have been identified). We have considered correct those answers that have indicated at least one correct cluster on the SOM–U-matrix view. The SOM–U-matrix makes possible the identification of larger clusters in the data. For the comparison task, we have considered correct those answers that have mentioned that the companies belong to certain clusters. However, this representation is not suitable for comparison in terms of financial performance, but only to give a hint about the differences that exist in the data.

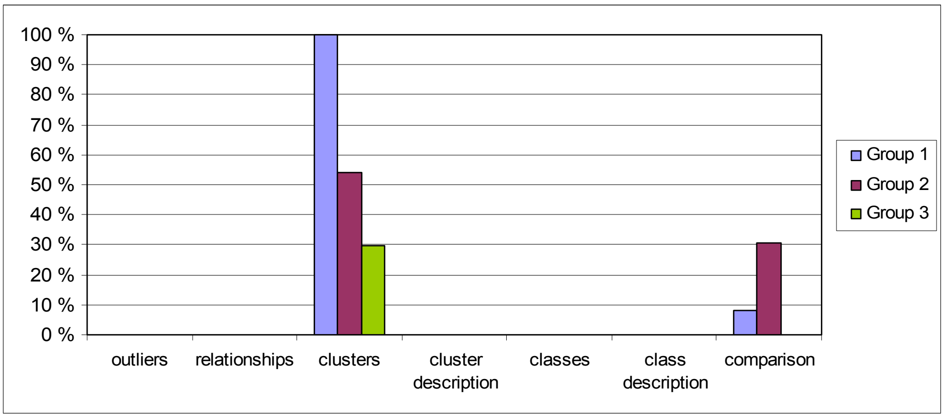


Figure 43. Positive answers (%) for evaluating the SOM–U-matrix view

7.4.11. SOM–Clustering

The SOM–Clustering view is effective for identifying clusters in the dataset (Figure 44). Most of the users have successfully identified four clusters in the data. For the comparison task, the answers have been considered positive if the users have compared the companies based on the membership to the different clusters.

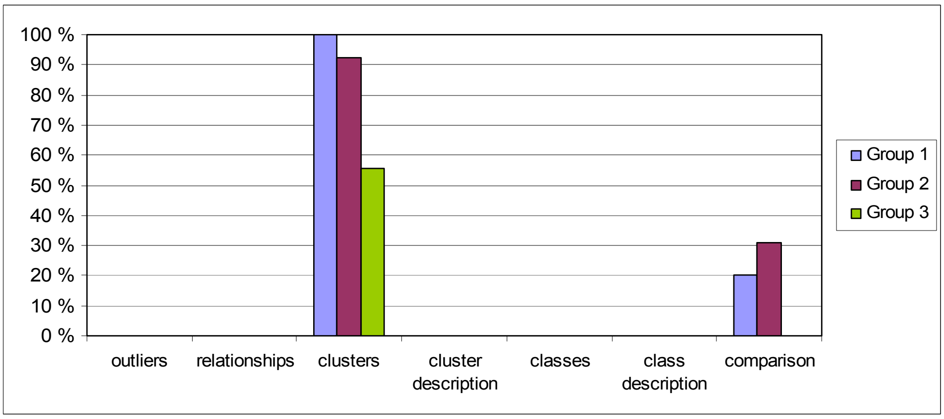


Figure 44. Positive answers (%) for evaluating the SOM–Clustering view

7.4.12. SOM–Feature Planes

Figure 45 shows that the SOM–Feature Planes are effective for revealing relationships between variables, clusters, cluster descriptions, and comparisons different companies. The results show differences in answers between Group 1 and the other groups, especially Group 3. This means that the SOM–Feature Planes view is more effective when users have experience of working with the tool (Group 1) or at least have been taught the basics of this tool (Group 2). The number of clusters identified ranges from one to eight. For the outliers task, answers that have indicated extreme values (dark blue or red) or unusual values in a cluster (e.g., the green color surrounded by blue colors in the RT plane) have been considered positive.

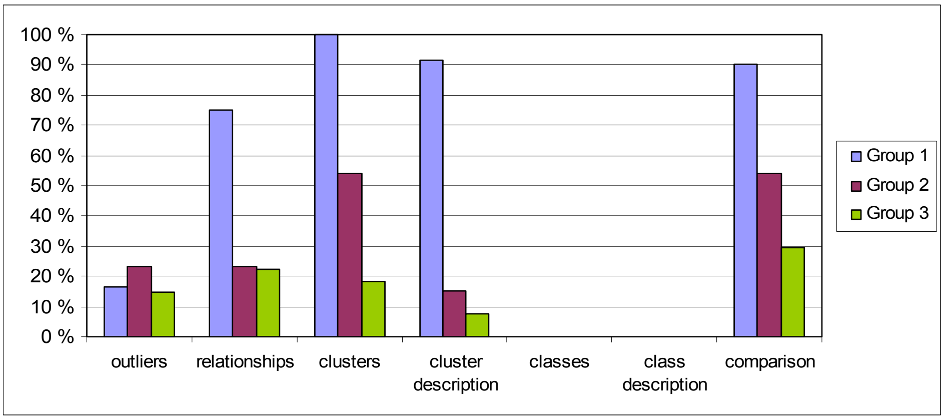


Figure 45. Positive answers (%) for evaluating the SOM–Feature Planes view

7.4.13. SOM–All Views

The SOM–All Views are effective for almost all tasks, but less effective for revealing outliers (Figure 46). There are differences between users who have experience of working with the SOM (Group 1) and users with less or no experience of the SOM (Groups 2 and 3).

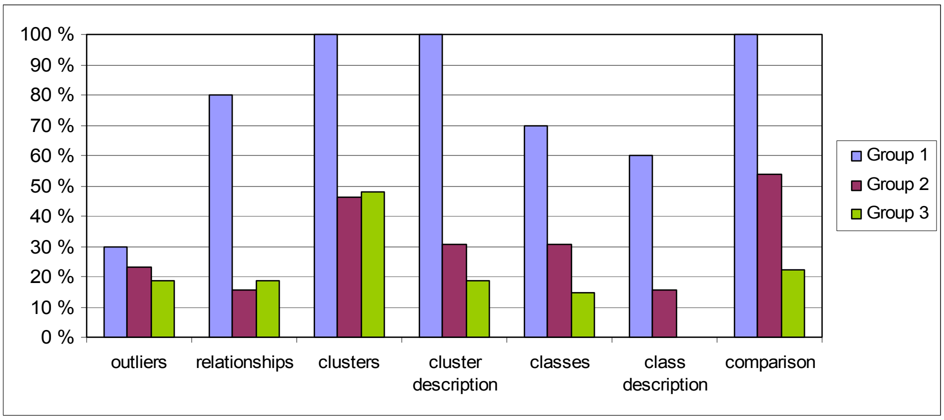


Figure 46. Positive answers (%) for evaluating the SOM–All Views

7.4.14. Summary of empirical results

Table 16 presents a summarization of the evaluations' results, aggregated over the three groups, without taking into account the size of the groups. The separate summarizations for each group are depicted in Table 17 - Table 19.

Table 16. The extent to which the visualization techniques are effective for uncovering the interesting patterns of the financial benchmarking dataset

Technique	*Outliers*	*Relationships*	*Clusters*	*Cluster description*	*Classes*	*Class description*	*Compari -son*
Multiple Line Graphs	++	+	-	-	+	+	+
Permutation Matrix	++	+	+	+	-	-	++
Survey Plot	++	+	+	+	++	+	++
Scatter Plot Matrix	++	+	+	-	-	-	+
Parallel Coordinates	++	+	+	+	-	-	++
Treemap	+	-	-	-	++	++	+
PCA	++	+	+	+	-	-	+
Sammon's Mapping	+	-	+	-	++	+	+
SOM Scatter Plot	+	-	++	-	+	-	+
SOM U-Matrix	-	-	++	-	-	-	+
SOM Clustering	-	-	++	-	-	-	+
SOM Feature Planes	+	+	+	+	-	-	+
SOM All Views	+	+	++	+	+	+	+

Legend:* + + *More than or equal to 50 percent of the participants identified correctly the patterns,
+ *Less than 50 percent of the participants identified correctly the patterns,*
- *No users identified the patterns.*

In Table 16, we have marked the tasks that have received more than or equal to 50% positive answers with ++, those that have received less than 50% with +, and those which did not receive positive answers with -. We have chosen these limits in order to highlight which tasks have been successfully answered by a *majority* of respondents by examining a given graph. In a practical situation, the limits can vary according to the aim of the evaluation and the context of use of the system/techniques under evaluation. For example, one can split the interval (0, 50) into (0, 25) and [25, 50), in order to differentiate the tasks that received positive answers from a very small number of respondents.

Taking into account the majority of positive answers provided for a given question, the centralization of all answers shows the following:

- The most effective techniques for uncovering *outliers* in the data are Multiple Line Graphs, Permutation Matrix, Survey Plot, Scatter Plot Matrix, Parallel Coordinates and PCA. To a lesser extent the other techniques, except SOM–U-matrix and SOM–Clustering, also reveal outliers.
- The techniques effective to some extent (less than 50%) for revealing *relationships* between variables are Multiple Line Graphs, Permutation Matrix, Survey Plot, Scatter Plot Matrix, Parallel Coordinates, PCA, SOM–Feature Planes and SOM–All Views.
- The most effective techniques for revealing *clusters* in the data are all the SOM-based visualizations. To a lesser extent all other techniques, except Treemap and Multiple Line Graphs, can reveal clusters.
- The techniques that are to some extent effective for *cluster description* are Permutation Matrix, Survey Plot, Parallel Coordinates, PCA, SOM–Feature Planes and SOM–All Views.
- The most effective techniques for revealing *classes* are Survey Plot, Treemap, and Sammon's Mapping. To a lesser extent, Multiple Line Graphs, SOM–Scatter Plot and SOM–All Views are effective for revealing classes.
- The most effective technique for facilitating the *class description* is Treemap. To a lesser extent Multiple Line Graphs, Survey Plot, Sammon's Mapping, and SOM–All Views are effective for this task.
- The most effective techniques for facilitating the *comparison* between companies are Permutation Matrix, Survey Plot, and Parallel Coordinates. All other technique are effective for facilitating comparisons but to a lesser extent.

Taking into account the number of tasks that a visualization technique can solve, the most effective techniques are Survey Plot, Multiple Line Graphs, Permutation Matrix, Parallel Coordinates, PCA, SOM–Feature Planes and SOM–All Views.

The initial evaluation in Chapter 6 has yielded similar results, that is, Survey Plot, Multiple Line Graphs, PCA, Treemap, SOM–Feature Planes, and SOM–All Views have provided the most capabilities. By comparing the patterns revealed by each technique, we have selected, in Chapter 6, a subset of the most suitable techniques for financial benchmarking as consisting of Survey Plot, PCA, and SOM–All Views. In a similar manner, based on the results in Table 16, the subset of most effective techniques for financial benchmarking consists of Survey Plot, PCA, and SOM–All Views. The Multiple Line Graphs, Permutation Matrix and Parallel Coordinates have their strengths, especially in revealing outliers and comparisons, but these tasks can also be answered by the other techniques.

The results also show that applying only one technique in financial benchmarking is not enough for getting a complete and accurate understanding of the data under analysis, since no single view is effective for fully uncovering all the patterns in the data. The Survey Plot appears to provide answers to all tasks, but a clustering structure is not fully uncovered by this technique.

The separate analysis of the results for each group shows that the percentages of positive answers in Group 1 are higher than in the other groups in almost all visualization/task evaluations (Table 17). This fact can be explained by the users' experience of working with the dataset, the financial benchmarking problem and the data analysis using the SOM tools.

Table 17. Effectiveness of visualization techniques for uncovering the patterns; Results obtained in *Group 1*

Technique	*Outliers*	*Relationships*	*Clusters*	*Cluster description*	*Classes*	*Class description*	*Compari-son*
Multiple Line Graphs	++	+	-	-	++	+	+
Permutation Matrix	++	+	+	+	-	-	++
Survey Plot	++	+	+	+	++	++	++
Scatter Plot Matrix	++	++	+	-	-	-	+
Parallel Coordinates	++	+	+	+	-	-	++
Treemap	++	-	-	-	++	++	+
PCA	++	+	+	+	-	-	++
Sammon's Mapping	-	-	+	-	++	+	+
SOM Scatter Plot	+	-	++	-	++	-	+
SOM U-Matrix	-	-	++	-	-	-	+
SOM Clustering	-	-	++	-	-	-	+
SOM Feature Planes	+	++	++	++	-	-	++
SOM All Views	+	++	++	++	++	++	++

Legend: + + More than or equal to 50 percent of the participants identified correctly the patterns,
+ Less than 50 percent of the participants identified correctly the patterns,
- No users identified the patterns.

In the case of Group 1, the visualization technique that has proved to be the most effective is the SOM, but only if all views have been analyzed together.

However, even the SOM–All Views have the limitation of not being capable of showing the outliers that other visualizations can uncover. The second best techniques are Survey Plot, SOM–Feature Planes, and Treemap.

The results in Table 17 slightly differ from those reported in Paper 5. First of all, this is due to the fact that we have repeated the analysis of the answers in Group 1 and have corrected some mistakes in the interpretation of the answers. Secondly, the results for the tasks cluster description and class description are calculated as the positive answers in the total answers in a group, while in Paper 5, the percentages are from the positive answers at clusters and classes tasks, respectively.

In Groups 2 and 3, the SOM–All Views and Survey Plot have also been found as providing the answers for almost all tasks. However, the percentages of positive answers are much lower than in Group 1. The respondents in Groups 2 and 3 have provided high percentages of positive answers especially at the task outlier's detection.

Table 18. Effectiveness of visualization techniques for uncovering the patterns; Results obtained in *Group 2*

Technique	*Outliers*	*Relationships*	*Clusters*	*Cluster description*	*Classes*	*Class description*	*Compari-son*
Multiple Line Graphs	++	+	-	-	++	++	+
Permutation Matrix	++	+	-	-	-	-	++
Survey Plot	++	+	+	-	++	+	+
Scatter Plot Matrix	++	+	-	-	-	-	+
Parallel Coordinates	+	+	+	+	-	-	++
Treemap	+	-	-	-	++	+	+
PCA	++	-	+	+	-	-	+
Sammon's Mapping	+	-	+	-	++	-	-
SOM Scatter Plot	+	-	+	-	+	-	+
SOM U-Matrix	-	-	++	-	-	-	+
SOM Clustering	-	-	++	-	-	-	+
SOM Feature Planes	+	+	++	+	-	-	++
SOM All Views	+	+	+	+	+	+	++

Legend: + + More than or equal to 50 percent of the participants identified correctly the patterns,
+ Less than 50 percent of the participants identified correctly the patterns,
- No users identified the patterns.

Table 19. Effectiveness of visualization techniques for uncovering the patterns; Results obtained in *Group 3*

Technique	*Outliers*	*Relationships*	*Clusters*	*Cluster description*	*Classes*	*Class description*	*Compari -son*
Multiple Line Graphs	++	+	-	-	+	+	+
Permutation Matrix	++	+	-	-	-	-	+
Survey Plot	++	+	+	+	+	+	+
Scatter Plot Matrix	++	+	-	-	-	-	+
Parallel Coordinates	+	-	-	-	-	-	+
Treemap	+	-	-	-	++	++	+
PCA	++	+	+	-	-	-	+
Sammon's Mapping	+	-	+	-	++	+	-
SOM Scatter Plot	+	-	++	-	+	-	-
SOM U-Matrix	-	-	+	-	-	-	-
SOM Clustering	-	-	++	-	-	-	-
SOM Feature Planes	+	+	+	+	-	-	+
SOM All Views	+	+	+	+	+	-	+

Legend: + + More than or equal to 50 percent of the participants identified correctly the patterns,
+ Less than 50 percent of the participants identified correctly the patterns,
- No users identified the patterns.

7.5. Method discussion

The resources involved in performing the evaluations are discussed in the following. No technical equipment has been involved in collecting the data. The time to answer the questionnaires has been between 30 to 120 minutes. It has also been possible to answer the questionnaire by e-mail. The number of users participating in each study is medium-sized compared to other evaluation studies. However, we have found that users resembling the target population (e.g., Group 1) tend to provide a larger number of correct answers. Moreover, for the results to be generalizable to the target population, the selection of the users from the target population or from a similar population is important. Regarding the evaluation expertise, there are no requirements for the participants to have experience of evaluation. The interpretation of the answers to open questions is quite time-consuming.

Comparing the results in this chapter (Table 16 – Table 19) with the results in Chapter 6 (Table 14), we have observed that some users have identified more interesting patterns that we initially did. For example, for the comparison task, users have compared the companies based on their membership to a cluster or another in the SOM-based visualizations (Table 17, Table 18). These results show that user evaluation is important because it provides more insight into the capabilities of the techniques and users' interests when they explore data. This fact also demonstrates that formulating open questions is appropriate when the tasks are of exploratory nature.

Based on the evaluation results, we conclude that one advantage of our evaluation technique is that it uses open questions in evaluating the effectiveness of visualization techniques for revealing interesting patterns. This is especially useful, when the tasks are exploratory and the user goals and experience play an important role. However, the answers to open questions are more difficult to assess. Therefore, we have obtained a number of answers that were incomplete, unclear or difficult to interpret, which we have classified as *invalid* answers (Appendix 5). The explanations for invalid answers can be the misunderstanding of the questions and tasks, or the misunderstanding of the visualizations. After conducting the first two empirical studies (Groups 1 and 2), we have tried to improve the questions so that the tasks become clearer to the respondents. The new questions were slightly different from the old ones. The new questions are presented in Appendix 6.

The participants answering the new questionnaire (denoted by Group 4) have been recruited using the same conditions with those in Group 3. They have answered the questionnaire at the same time with those in Group 3. The questionnaires have been given to the participants in random order so that the numbers of students in each group are similar. Group 4 consists of 25 students. They have similar majors to those in Group 3 and they are not familiar with the dataset. One of the participants in Group 4 has experience of working with the SOM technique.

We have analyzed the answers to this new questionnaire separately from the other answers, but using the same criteria of interpreting and evaluating the answers. We have used non-parametric statistical tests (Siegel and Castellan 1988) in order to test whether the differences observed in the answers of Groups 3 and 4 are significant. We have used the two-tailed tests and report the results at two significance levels, alpha = 0.1 and alpha=0.05.

Because the Chi-square test was applicable only in two visualization/task evaluations, we have applied the *Chi corrected* and *Fisher tests*. For this purpose, we have grouped the answers into two categories. First, we have obtained the dichotomy: *clear* (positive and negative) versus *unclear* (missing and invalid) answers. The results of the Chi corrected and Fisher tests are depicted in Table 20. The second dichotomy was *positive* versus *non-positive* (negative, missing and invalid) answers. The results of Chi corrected and Fisher tests for this dichotomy are depicted in Table 21.

Table 20. Significant differences observed between Groups 3 and 4 using the Chi corrected or Fisher tests, when the categories of answers are grouped into clear (positive and negative) and unclear (missing and invalid)

Technique	*Task*	*Level of significance*
Parallel Coordinates	Relationships	0.1
PCA	Cluster description	0.1, 0.05

Only two significant differences have been observed between groups. In both cases, the larger number of “unclear” answers was observed in Group 3.

Table 21. Significant differences observed between Groups 3 and 4 using the Chi corrected or Fisher tests, when the categories of answers are grouped into positive and non-positive (negative, missing and invalid)

Technique	*Task*	*Level of significance*
PCA	Clusters	0.1, 0.05
SOM-All Views	Clusters	0.1

In the case of PCA / clusters evaluation, the larger number of “non-positive” answers has been observed in Group 3. In the case of SOM-All Views / clusters, the larger number of “non-positive” answers has been observed in Group 4.

Given these results, which do not show strong evidence that differences in answers are due to the questions, we conclude that the way in which the questions were formulated did not influence the answer.

7.6. Concluding remarks

7.6.1.Summary of results

In this chapter, we have evaluated nine techniques concerning their effectiveness for revealing interesting patterns in a financial benchmarking dataset. For evaluating the effectiveness of a technique, we have used a new method based on questionnaire data-collection. The effectiveness of a technique in a given task was measured by the percentage of positive answers to question regarding the identification and/or description of the corresponding patterns in the data. The overall effectiveness of a technique for the financial benchmarking problem was given by the number of patterns that the technique can reveal.

We have performed three empirical studies which show the strengths and limitations of each technique (Table 16 – Table 19). Likely due to the differences between the three groups of users, the results of the evaluation in the three groups differ to some extent. Overall, among the most effective visualizations for uncovering different types of patterns are SOM–All Views, Survey Plot, SOM–Feature Planes, PCA, Parallel Coordinates, and Permutation Matrix. The SOM technique is especially effective when used after training and experience of the tool (Group 1) and when different views of SOM are combined.

In the first empirical study, the respondents better resemble the target population, managers working with data analysis tasks, specifically in business intelligence and competitor analysis domains. Their answers are thus more important in the evaluation. We have used the evaluation method with other two groups of respondents, which were not familiar with the business problem and the dataset. The results show that, in the first group, the proportion of positive answers, which reflect correct identification of interesting patterns, was higher than in the groups where users were not familiar with the financial benchmarking and the dataset in almost all tasks and techniques evaluated.

We have summarized in Table 16 the results without making differences between the groups. Moreover, we have emphasized the differences between tasks that received more than 50% positive answers and those which received less than 50%. In a practical situation, the evaluator can select different rating levels.

The results also show that for simpler tasks such as outlier detection and class distinction, novice users in terms of familiarity with the business problem and the dataset are as good as the more experienced users. However, when the tasks require complex interpretation of the visualization (e.g., colors, size, position, meaning of axes), novice users do not provide elaborated or correct answers in the same proportion as experienced users do.

The method based on questionnaire data-collection is relatively easy to implement and yields useful results. Small variations in formulating the questions do not seem to influence the users' answers. The introduction to information visualization and presentation of the techniques as a tutorial lecture before handing out the questionnaires to the participants do not seem to influence the answers either. However, the understanding of the patterns is important, but we did not control this variable.

Regarding the UE process model (Table 7), this chapter contributes with providing guidance on steps 4, 5, 8 and 9 in that model. We evaluated the visualizations with respect to their *effectiveness* (*correctness of interpretation* and *visual efficacy*) in presenting the information (4). The measures were of *qualitative nature*, capturing the extent to which the techniques are effective in providing answers to the data mining tasks formulated for the financial benchmarking problem (5). We chose to perform a *subjective* evaluation, using an *inquiry method* (8). The actual evaluation consisted of asking users to assess whether given visualization techniques reveal interesting patterns in the data (9).

In conclusion, this chapter extends the empirical studies described in Papers 4 and 5. The contribution of these empirical studies is that they propose, illustrate and explore an evaluation method of multidimensional visualization techniques based on questionnaire data-collection. The method is especially useful in the early stage of the development of a visualization system, that is, in the selection of the techniques to be implemented. Moreover, the evaluation results highlight the strengths and limitations of various visualization techniques in effectively providing insight into a financial benchmarking dataset.

7.6.2. Limitations

One limitation of the evaluation studies regards the generalizability of the results to other business problems than financial benchmarking. We did not intend to obtain generalizable assessments of the techniques to other problems than financial benchmarking. Moreover, we cannot claim that the same results would be obtained for other datasets than the one used in these studies. Further empirical research would be necessary to make such claims.

Another limitation of the studies is that users evaluated only static visualizations of the data, while the interaction with the visualizations could have influenced the users' perceptions of the capabilities of the techniques. For example, by changing the order of companies in Multiple Line Graphs or using color-coding in PCA, these techniques could have revealed more patterns.

Another limitation is that the users in the first empirical study were familiar with the SOM, while other techniques may have been unknown to the users. However, each technique was briefly described in the questionnaire.

Another limitation is that the visualization techniques have been presented in the same order to all respondents. This fact may have biased the results towards the last techniques (e.g., the SOM-based techniques) due to the knowledge acquired by answering the preceding questions. However, this bias did not manifest in Groups 2 and 3, and the good results obtained for SOM–All views evaluation in Group 1 are more likely due to the fact the users had experience of the SOM.

A limitation of the method is the subjectivity involved in interpreting the answers to the open questions. Replication of the empirical studies with different users and the use of advanced automated tools in analyzing the answers represent two solutions in making the method more reliable.

8. Evaluating the quality of use of visual data mining tools

8.1. Research problem description

In previous chapters, we have explored the capabilities and effectiveness of various visualization techniques for different data mining tasks. The SOM technique appeared to be one of the most effective techniques for the tasks under evaluation. In this chapter, we are concerned with the evaluation of the *quality of use* of the SOM-based tools.

The research problem is to identify the quality-of-use *characteristics* and *attributes* of visual data mining (VDM) tools (RQ3). We create a *conceptual framework* for evaluating the quality of use of VDM tools, which consists of a working definition of the quality of use and its characteristics and attributes.

Quality of use is defined in (Bevan 1995) as "the extent to which a product satisfies stated and implied needs when used under stated conditions." This corresponds to the view on usability, according to which usability is an objective of a software product or interactive system (ISO 9241-11 (ISO 1998); ISO/IEC 9126-1 (ISO 2000b)). Bevan also points out that measuring quality of use implies measuring aspects such as *effectiveness*, *efficiency* and *satisfaction* of the users in achieving specified goals in a specified context of use.

We are concerned with evaluating (measuring) the quality of use of *VDM* tools. According to the model of UE process (Table 7), we need a model that describes the quality-of-use characteristics and attributes of the system. The attributes are lower-level characteristics of the system that can be directly measured.

Though there exist many general models of usability and/or quality of use (e.g., ISO/IEC 9126-1; ISO 9241-11; Nielsen 1993; Kirakowski 1994), they do not address the particularities of VDM tools. These tools have the particularity that they represent data in a visual form with which users can interact in order to gain information. A high level of information quality is essential for the users, and this can be achieved through a high-quality interaction with the system and a high-quality visualization of the data. Therefore, we propose a framework of quality of use of VDM tools that reflects characteristics from three levels of analysis: *visualization*, *interaction*, and *information*. We identify and select from the literature a number of characteristics that can portray the effectiveness, efficiency and satisfaction of the user at the three levels of analysis specified above. The characteristics, their definitions and relationships are presented in the following section.

To examine the applicability of the framework, we conduct an empirical study in which we evaluate the SOM tools. We identify important attributes of the SOM and measure them by employing a *subjective evaluation,* by asking users' opinions. We also capture the *attitude of users (user satisfaction)* towards the SOM tools and the *performance* of the users in given tasks. We employ the questionnaire technique to gather data about users' attitudes and opinions regarding the SOM tools. For evaluating the performance of the users in given tasks, we analyze multiple case studies, collected in the form of reports on the solutions provided by the users to the tasks. In addition, we evaluate the questionnaire in terms of reliability and internal consistency of the scales. The contents of this chapter are based on Paper 6.

8.2. A framework of evaluating the quality of use of visual data mining tools

For constructing the framework, we have reviewed the relevant usability evaluation literature (Doll and Torkzadeh 1988; Kirakowski 1994; Bevan 1995; Fenton and Pfleeger 1997; Dix et al. 1998; ISO/IEC 9126-1) and visualization evaluation literature (Bertin 1981; Tufte 1983; Shneiderman 1996). The definitions of the concepts involved in the framework are mainly based on the definitions of relevant characteristics found in the above references. In particular, the SUMI (Kirakowski 1994) and EUCS (Doll and Torkzadeh 1988) models have provided us a basis for characterizing the *interaction* and *information* aspects of a VDM tool. Regarding the visualization, the principles of excellence and integrity in visualization (Tufte 1983) served us as a basis for defining the characteristics of the *visualization* aspect of a VDM tool.

8.2.1. Quality of use

In order to take into account the particularities of a VDM tool, we define quality of use of a VDM tool as being *the totality of features and characteristics of the tool that reflect its ability to satisfy the users' needs*. This definition is based on the definition of *quality* in ISO/IEC 14598-1 (ISO 1999b)[16]. Our definition points out that a VDM tool has a high-level of quality of use if its features and characteristics are capable of satisfying the users' needs. We propose that the main and direct features of the VDM tool that influence the user satisfaction are

[16] Quality is defined in ISO/IEC 1459-1 (ISO 1999b) as being "the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs".

the *visualization* of data, the *interaction* with the system, and the obtained *information*.

8.2.2. Quality of visualization

At this level, we are concerned with evaluating the capability of the visualization system to transform the input data and to make them accessible to the user. The issues to be evaluated are presented in Figure 47.

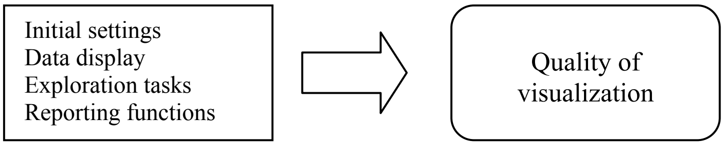


Figure 47. Evaluating the quality of visualization

- *Initial settings* refer to the requirements on input data format, the degree of data abstraction, and the setting of the parameters for visualization.
- *Data display* regards the possibility to visualize the data structure, data variation, data content, and data comparison. Moreover, the description, tabulation and decoration of data are important to evaluate.
- *Exploration tasks* include the five visual tasks identified by Shneiderman (1996), i.e. overview, details of data, filter, details on demand, and relate.
- *Reporting functions* represent those system functions that allow the user to transfer the results outside the application for various purposes. In this part we are concerned with evaluating whether the user is satisfied with how he/she benefits from the visualization. We also ask whether the user is encouraged by the visualization to think of the data, rather than of the graphical design and methodology.

We have chosen the four characteristics of the quality of visualization (Figure 47) so that they cover all “user environments that the software may affect”, which, according to ISO/IEC 9126-1, may include preparation for usage (which we called *initial settings*) and evaluation of results (which we called *reporting functions*). The *data display* (Tufte 1983) and the *data exploration* (Shneiderman 1996) are the main functions of the visualization techniques. Shneiderman also mentions the importance of enabling the use of the results in other applications.

8.2.3. Quality of interaction

At this level, we are concerned with evaluating the extent to which the users consider the system easy to use and learn, accurate (reliable), effective and efficient. We classify the interaction characteristics in five groups (Figure 48).

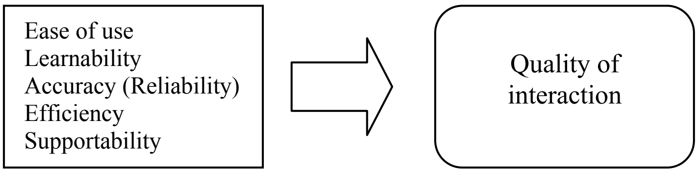


Figure 48. Evaluating the quality of interaction

- *Ease of use* stands for the characteristic of the system to be easy to control by the user and to provide the user with freedom of action (controllability and flexibility).
- *Learnability* affects how easily and quickly the users feel that they master the system enough to perform the desired tasks.
- *Accuracy* (reliability) reflects the frequency and severity of system errors or failures.
- *Efficiency* measures the degree to which users feel that the software helps them in their work (to tailor frequent actions, improve working performance, and receive fast responses to queries).
- *Supportability* regards the users' access to documentation and support, when needed.

We have derived the characteristics of the quality of interaction (Figure 48) mainly from the *SUMI model* (Kirakowski 1994), but also from the *ISO/IEC 9126-1 models of usability and quality of use* (Table 22). However, we did not include here the user satisfaction or affect characteristics. Table 22 presents the similarities and differences between the quality-of-interaction characteristics, the SUMI and ISO/IEC 9126-1 models. We have included the *accuracy* characteristic because it is important in the "operation" of a system, according to the *McCall software quality model* described by Fenton and Pfleeger (1997, p. 339).

Table 22. Comparison of quality of interaction with other models of usability

Quality of interaction	SUMI **Usability**	ISO/IEC 9126-1	
		Usability	**Quality of use**
Ease of use	Control	Operability	
Learnability	Learnability	Learnability	
Accuracy (reliability)			
Efficiency	Efficiency		Effectiveness Productivity
Supportability	Helpfulness	Understandability	
	Affect	Attractiveness	Satisfaction
		Usability compliance	
			Safety

8.2.4. Quality of information

At this level, we are concerned with evaluating the extent to which the users are satisfied with the obtained information. Figure 49 shows the four characteristics of information, which the users might require.

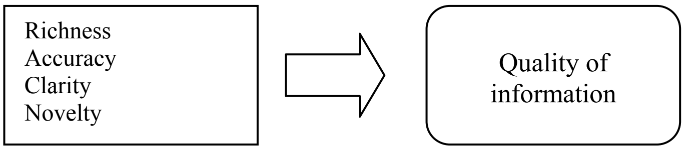


Figure 49. Evaluating the quality of information

- *Richness* of information stands for completeness, usefulness, and interestingness. In addition, it must correspond to users' needs and expectations.
- *Accuracy* of the information regards the degree to which the information is precise, correct, and consistent with users' knowledge.
- *Clarity* of information means that the information is presented in a clear and understandable way, and allows interpretation and inferences.
- *Novelty* of information reflects the characteristic of being new and up-to-date.

We have derived the characteristics of the quality of information (Figure 49) mainly from the *EUCS model* of user satisfaction (Doll and Torkzadeh 1988). Table 23 shows the relationships between the quality-of-information characteristics and the EUCS model.

Table 23. Comparison of quality of information and EUCS model of user satisfaction

Quality of information	EUCS: **User satisfaction**
Richness	Content
Accuracy (correctness)	Accuracy
Clarity	Format
	Ease of use
Novelty	Timeliness

8.2.5. Relationships between characteristics

We hypothesize that the characteristics of quality of use at different levels of analysis are not independent but they influence each other. We depict in Figure 50 the relationships between the characteristics corresponding to the three levels of assessment. The relationship (1) reflects the fact that when a user examines the data display and uses the results, he/she must find the information to be rich, accurate, clear, easy to interpret, novel and up-to-date. The two-way arrow indicates that the perceived quality of visualization influences the perceived quality of information and vice versa. The relationship (2) reflects the fact that whenever a user interacts with the system, he/she wishes the process to be easy, accurate, and effective. The two-way arrow indicates that the perceived quality of visualization influences the perceived quality of interaction and vice versa. The relationship (3) reflects the fact that the interaction with the tool should provide the information needed in the tasks for which the tool has been employed. The two-way arrow signifies that the quality of interaction may influence the quality of information provided and vice versa, the quality of information obtained may have influence on the perceived quality of interaction.

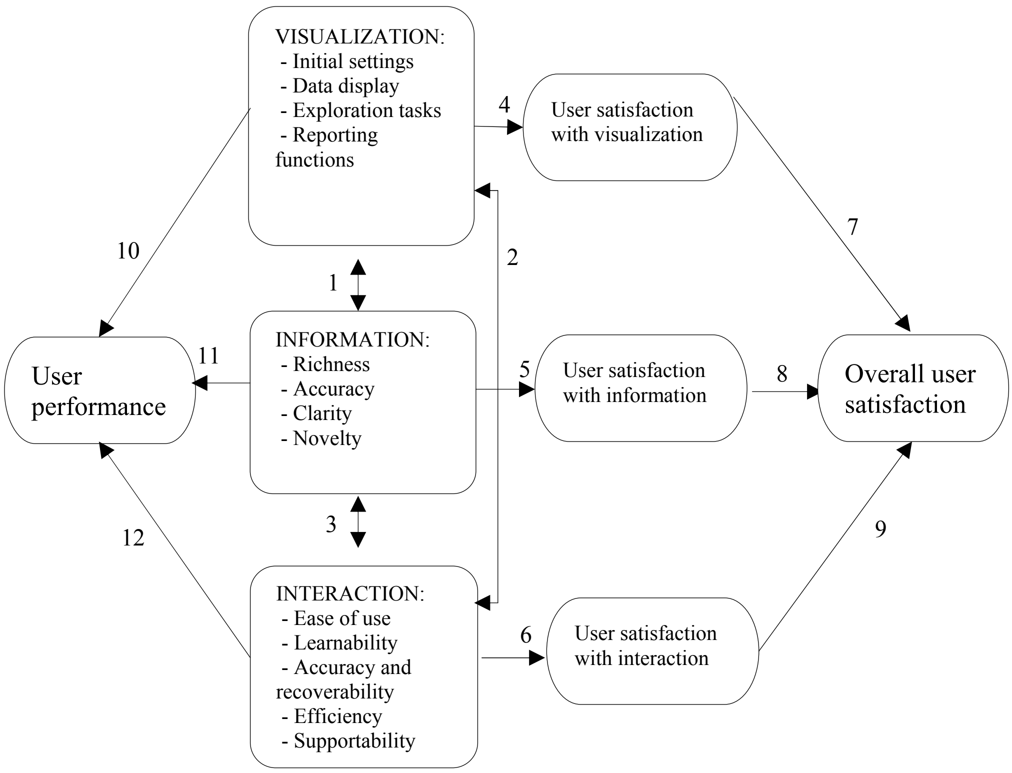


Figure 50. Relationships between attributes

Relationships (4) - (6) show that the quality of visualization, information and interaction influence the user satisfaction with the visualization, information and interaction, respectively. Relationships (7) - (9) indicate that the user satisfaction with the visualization, information and interaction influence the overall user satisfaction with the visual data mining tool. Finally, the relationships (10) - (12) indicate that the quality of visualization, information and interaction all may influence the user performance.

8.3. Evaluation method

8.3.1. Data capture

In order to capture data about the characteristics of quality of use, we first divided each characteristic into sub-characteristics or attributes that are easier to measure (Table 24, Table 25, and Table 26). Then, based on the attributes, we developed a questionnaire. The questions request the users to rate the identified attributes. The questionnaire is presented in Appendix 7.

8.3.2. The quality-of-use attributes

Table 24. Visualization characteristics and attributes

Characteristics	Attributes
Initial settings	- Requirements on input data format - Adequacy of normalized data - Easy to understand parameters - Easy to use parameters
Data display	- Data structure: Data clusters, trends, attribute values, correlations between attributes - Data content: Exploration and description of data - Data variation - Data comparison - Tabulation of data - Decoration of data - Description of data: labeling - Dimensionality and size of the graph

Reporting functions	- Thinking about what is seen:
	    - Substance of the data
	    - Design elements
	    - Computational issues
	- Easy to integrate the resulting maps within other software applications

Table 25. Interaction characteristics and attributes

Characteristics	Attributes
Ease of use	- Too many steps required - Easy to use tool
Learnability	- Easy to learn tool - Satisfaction with learnability
Efficiency	- Time needed to obtain a good map - Provides the information needed
Accuracy	- Satisfaction with the accuracy of the system

Table 26. Information characteristics and attributes

Characteristics	Attributes
Richness	- Reliable - Complete - Interesting - Needed - Useful
Accuracy	- Accurate - Precise - Correct
Clarity	- Clear and understandable - Easy to interpret
Novelty	- New

8.3.3. Scales of measurement

All questions request *qualitative* information. For example, they ask whether the information is accurate or whether the user agrees with some statements. We use a 5-point scale to record the answers, thus obtaining *quantitative measures* of the attributes. The 5-point scales vary according to the way in which we formulate different questions, e.g. “1 very good”; “2 good”; “3 medium”; “4 poor”; “5 very poor”, or “1 fully agree”, “2 partially agree”, “3 neutral”, “4 partially disagree”, and “5 fully disagree”.

8.3.4. Data analysis

The collected data is of qualitative nature, but measured on a quantitative (interval) scale from 1 to 5. For assessing the results, we are interested in the *percentages of positive answers*, that is, answers that reflect high user satisfaction or positive opinions about different attributes of the tools. We map the answers on a 3-point scale as follows: positive answers: “very good” and “good”, neutral answers: “medium”; negative answers: “poor” and “very poor”.

8.4. Evaluation of the SOM tools

In the evaluation study of the SOM tools, we have employed the mixed methods research design. This research design is suitable when the researcher needs to collect both quantitative and qualitative data (Creswell 2003). In the quantitative part of the study, which concerned the evaluation of the quality of use of SOM tools, we have used the questionnaire technique to collect data. We have used the questionnaire described in Section 8.3. In the qualitative part of the study, we have collected and analyzed the participants' solutions to the given tasks in order to determine the users' performance.

The participants in the study are 26 students enrolled for an Information Systems course. The demographics of these participants are presented in Paper 6.

8.4.1. Materials

In the evaluation, we have used three software packages that implement the SOM technique. These packages are SOM_PAK 3.1 (1995), SOM Toolbox 2.0 for MATLAB (2005), and Nenet 1.1 (1999).

The data collection process has comprised the following phases:

1. Training the students to use all three SOM tools,
2. Asking the students to solve an assignment and report their findings,
3. Asking the students to answer a questionnaire evaluating the SOM tools.

When solving the assignment, the students were given the opportunity to choose the tools that they wanted to work with, choosing from SOM_PAK, SOM Toolbox, and Nenet. Nenet has been definitely preferred by all students for visualizing the maps, while students used either SOM_PAK or SOM Toolbox to train the maps. We have used the Binomial and Chi-square tests (Siegel and Castellan 1988) to check whether there are differences in attitudes between users of the SOM_PAK and SOM Toolbox. No significant differences have been found.

8.4.2. Data and tasks

The input data used for the assignment has been the same *financial benchmarking dataset* described in Section 6.2. The assignment required the students to train several different SOMs with the input data provided, until they obtained a map on which to visualize the data and identify the clusters. Students have been requested to answer five questions and document their solutions in a report. The questions were:

1: How many clusters do you identify and what are the characteristics of each cluster?
2: Which is the cluster that contains the best performers in the market?
3: Which is the cluster that contains the worst performers in the market?
4: Discuss the performance of three specific companies based on their positions on the map and compare the results with the real data from the file provided.
5: Benchmark five specific companies one against the other, based on their positions on the map.

8.5. Empirical results

The main empirical results of the evaluation study consist of:

- Evaluation of quality of use of SOM tools in VDM (measuring the quality of visualization, information and interaction)
- Evaluation of the user performance, and
- Identification of problems and limitations of the SOM-tools.

8.5.1. Quality of use of SOM tools

Figure 51 depicts the results regarding the *quality of visualization*. Among the positive aspects, we observe the good visualization of data clusters (92% respondents agree), and the visualization of the comparable data and data trends.

The *initial settings* have not revealed major problems. However, the SOM parameters have been found easy to understand by only 50% of the students. Regarding the *data display* features, relatively low scores have been noticed for tabulation of data, decoration of data, visualization of the correlations between attributes, and visualization of the attributes values. In the *reporting functions* category, we observe that more than 75% of the users have found the results easy to use within other applications. The attention is focused on the substance of data for more than 65% of the users.

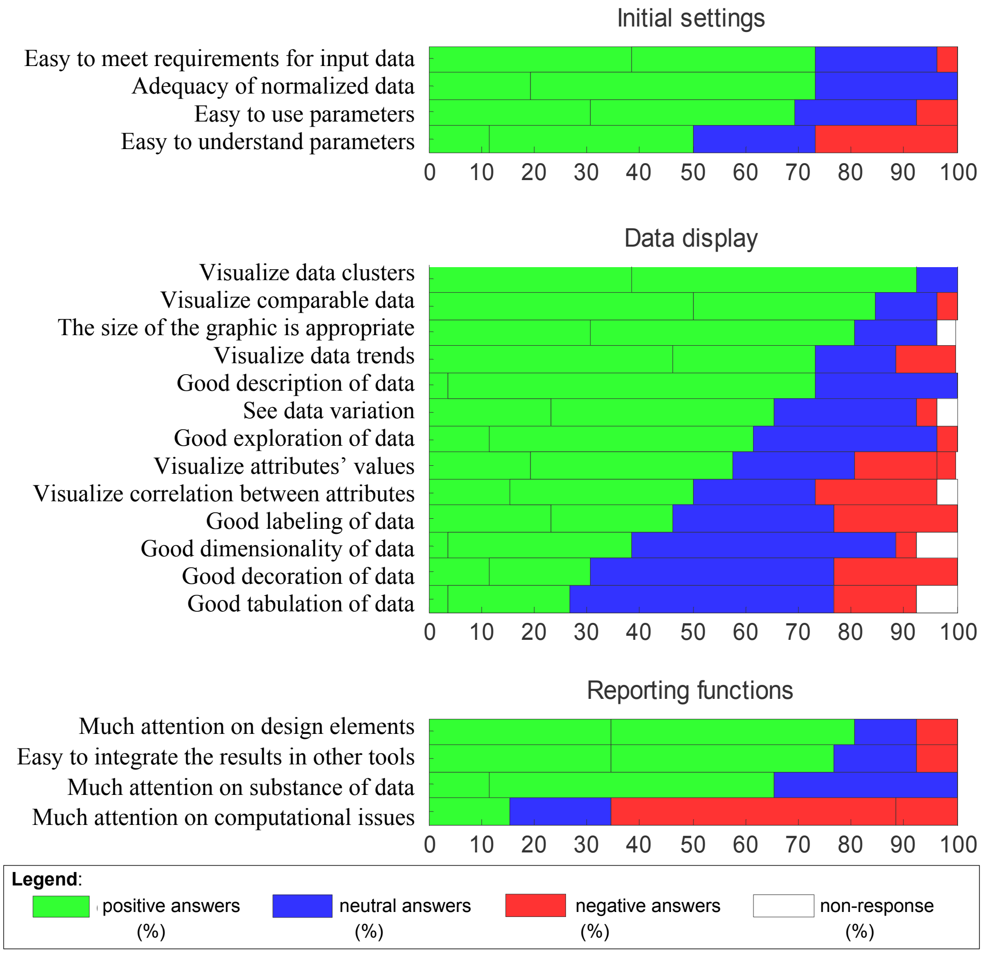


Figure 51. Quality of visualization

We have asked a number of questions about the degree to which different design elements helped in interpreting the visualization (map). Table 27 presents the results.

Table 27. Assessment of the SOM's graphic elements

	Helpful			Adequate		
(%)	Agree	Neutral	Disagree	Good	Medium	Poor
Colors	92	8	0	88	12	0
Scales (color bars)	85	15	0	85	15	0
Grids, neurons, borders	81	19	0	57.5	31	11.5
Attribute values	69	19	8	54	31	15
Data labels	77	15	8	61.6	19	19.4

Figure 52 presents the opinions and attitudes regarding the *quality of interaction*. Among the positive attributes are the ease of use, and ease of learning. In addition, most of the users (82.60%) agreed that the system provided the information needed.

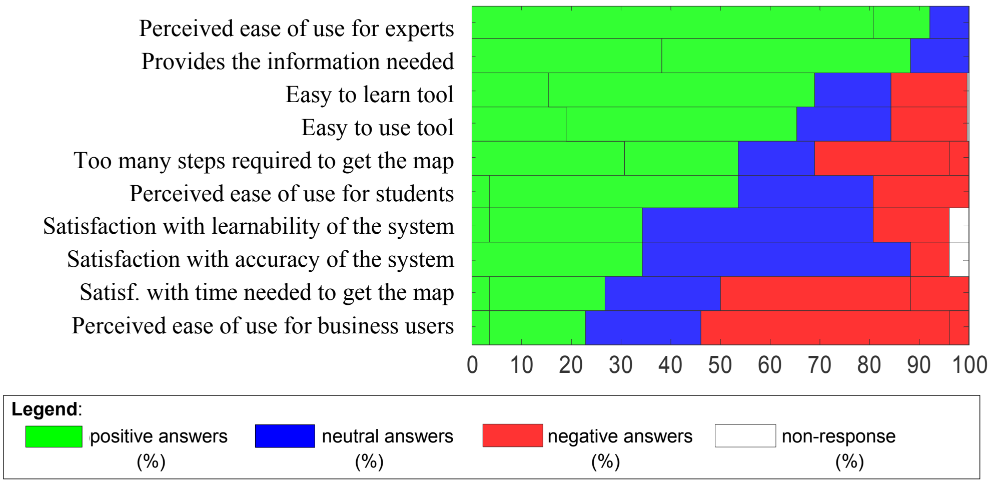


Figure 52. Quality of interaction

The weak points perceived by the students are system flexibility (54% of the respondents have agreed that there are too many steps required to get a good map), and efficiency (only 27% of the respondents have been satisfied with the time needed to get a good map).

Figure 53 shows that the *information* obtained is helpful and useful in data analysis. It is also interesting, easy to understand, and complete for most of the students. However, they are not very satisfied with the correctness of the information and even less with its preciseness. Users still find the SOM content reliable, and overall the satisfaction with the content is high.

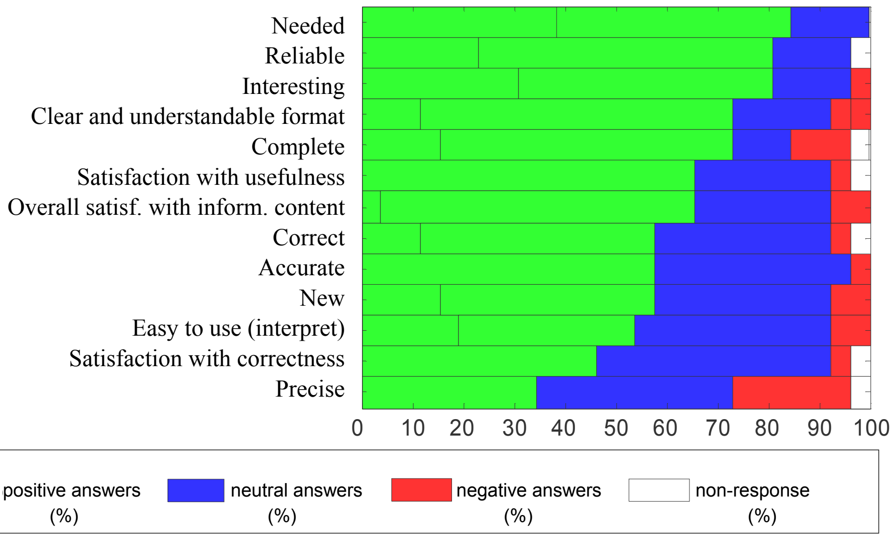
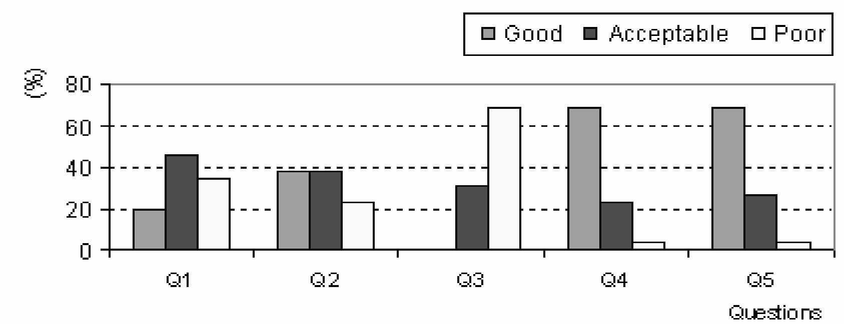


Figure 53. Quality of information

8.5.2. User performance

Participants in the experiment have been asked to solve a complex task with SOM tools, namely to train the SOM until they obtain a map, and with its help to answer five questions. For evaluating the user performance, we have analyzed the students' reports describing the solutions found.



Q1: How many clusters do you identify and what are the characteristics of each cluster?
Q2: Which is the cluster that contains the best performers in the market?
Q3: Which is the cluster that contains the worst performers in the market?
Q4: Discuss the performance of three specific companies based on their positions on the map and compare the results with the real data from the file provided.
Q5: Benchmark five specific companies one against the other, based on their positions on the map.

Figure 54. Quality of solutions reported by participants

Figure 54 shows that the most difficult tasks for students have been to obtain an appropriate map on which to identify correct clusters (Q1, Q2, and Q3). Thus, the first three questions, concerning the number of clusters and their definitions, have received the most varied answers and these have not been very well argued. Students themselves have been aware that their map might not be the correct one, and have noted that an inappropriate map could lead to misinterpretations and mistakes in the decision making process. The last two questions have been much better answered.

Among the explanations the users have given for their imperfect solutions were inexperience of working with SOM tools, unfamiliarity with financial ratios, and the highly subjective criteria for separating the clusters (for some managers, some ratios are more important in a certain time, etc.). Overall, the participants found it very interesting and useful to work with the SOM technique.

8.5.3. SOM tool limitations

Table 28 shows the main limitations of the SOM tools pointed out by our study. For each identified problem, we propose possible solutions and suggestions to improve the software that implements SOM.

Table 28. Problems found and suggestions for improvement

Problem	Suggestion for improvement
Level 1: quality of visualization - Not very easy to understand input parameters - Poor tabulation of data - Poor decoration of data - Medium data labeling	- Automation of parameters selection according to the input data characteristics and the desired results, - Enhance the "Details on demand" feature to display properly the input data and their statistics in tabular reports.
Level 2: quality of interaction - Low perceived ease of use for business users - Medium satisfaction with the time needed to get a good map (visualization), too many steps required - Medium satisfaction with the accuracy of the system - Medium satisfaction with the learnability of the system	- Provide automatic delineation of the clusters. - Due to the fact that SOM reduces the dimensions of the input space, the loss of accuracy is inevitable, but new learning algorithms could be tested for implementation.

Level 3: quality of information - Not very precise - Not high satisfaction with correctness - Not very easy to use (interpret) - Not very accurate	- Add explanations to the information displayed when these are requested.

8.6. Evaluation of the questionnaire

In order to examine the reliability of the scales that we used in evaluation of the SOM tools, we compute the Cronbach's alpha coefficient. A rule of thumb states that the internal consistency of the scales is acceptable when alpha is greater than 0.7 in confirmatory studies and greater than 0.6 in exploratory studies (Straub et al. 2004). Table 29 presents the Cronbach's alpha values for our data. At the Visualization level, there are lower values of alpha for Initial settings and Reporting functions. This is due to the fact that the questions in this section of the questionnaire were focused on distinct issues, thus, no significant similarities between answers have been found. Moreover, the six satisfaction questions are not highly related and, hence, the corresponding Cronbach's alpha is relatively low. These low values can also be explained by the small number of items used, because the value of alpha increases directly with the number of items of the construct, as well as with the correlation between the items.

Table 29. The Cronbach's alpha computed for each level of assessment

Level	alpha	Notice	Alpha
Visualization quality	**0.7724**	when graphical aspects are included:	**0.8704**
Initial settings	0.3971		
Data display	0.7273	when graphical aspects are included:	0.8704
Reporting functions	0.5659		
Interaction quality	**0.6739**	including visualization items:	**0.7046**
Ease of use and learning	0.6143	including visualization items:	0.6774
Accuracy	not computed, only one item used		
Efficiency	not computed, only one item used		
Information quality	**0.7467**	including visualization items:	**0.8748**
Richness	0.5443	including visualization items:	0.7732
Accuracy	0.6075		
Clarity	0.6110		
Novelty	not computed, only one item used		
Satisfaction questions	**0.6291**		

In Table 30, we present the results of Cronbach's alpha obtained for the quality-of-use scale (i.e., all quality questions taken into account). Moreover, the reliability of the user-performance scale is calculated, as well as the overall quality-of-use by combining the user rating and the performance scores.

Table 30. The Cronbach's alpha computed for the overall quality-of-use

Level	alpha	Notice
All quality questions	0.8872	Calculated using the three-point scale, derived from the original five-point scale
User Performance	0.7044	Calculated for the scores assigned to the students' solutions
Overall	0.8845	Calculated for user performance and quality questions

The evaluation of the questionnaire and implicitly of the evaluation method in terms of reliability of the scales shows that that at the overall level the questionnaire is a good instrument of measuring quality of use. Moreover, the measuring of user performance by evaluating the solutions to the chosen tasks is also adequate. Table 29 shows that some scales for measuring quality of use at specific levels and characteristics can be improved (e.g., initial setting, reporting functions, and richness of information).

8.7. Concluding remarks

In this chapter, we have presented a framework for evaluating quality of use of visual data mining (VDM) tools. Quality of use has been defined as being the satisfaction with all of the features of the VDM tool, namely visualization of data, interaction with the system, and information obtained. The framework consists of characteristics and sub-characteristics (attributes) of quality of use at three levels of evaluation of the tools: visualization, information, and interaction. These levels are not completely separated, but interdependent.

Based on the framework, we have developed a questionnaire. To examine the applicability of the framework, we have conducted an exploratory study to evaluate the quality of use of the SOM tools. The results show that the users are satisfied working with the SOM tools. Most of the visual features are considered helpful and adequate. Users are helped by the SOM technique to understand and analyze relatively large amounts of data and to obtain interesting and new information. Regarding the interaction with the tools, users find the tools easy to use and learn. Nevertheless, the SOM tools appear to also have weak points. These are identified in terms of "too long time needed to obtain a good map", relatively low accuracy, preciseness, and correctness of the information, and difficulty in interpreting the results. All these shortcomings, especially the lack of efficiency and preciseness, might explain why business users do not frequently use the SOM tools in financial data analysis.

We have also examined the user performance in the given tasks. The user performance was relatively good, especially in describing and benchmarking certain companies as to their financial performances. Perhaps by enhancing the usability and functionality of the SOM tools, the level of performance and usage of this VDM tool would increase.

We have evaluated the reliability of the questionnaire in measuring the overall quality of use and its characteristics. The results show that, at the overall level, the questionnaire is a good instrument of measuring quality of use. Moreover, the measuring of user performance by evaluating the solutions to the chosen tasks is also adequate. However, some scales used for measuring quality of use at specific levels and characteristics can be improved (e.g., initial setting, reporting functions, etc.).

The resources involved in the evaluation are described in the following. The users' sample is medium-sized, compared to other evaluation studies. The respondents are students resembling business users because they have analyzed a real problem and dataset. It is not necessary that users have experience of evaluation. The time required to collect the questionnaire data is relatively low.

The significance of the study is twofold. Firstly, we provided a comprehensive framework for assessing the VDM tools from the user perspective. The advantage of the framework is that it enables evaluators to derive and focus on important attributes of quality of use. The evaluation of the attributes provides insight into the strengths and limitations of the tools under evaluation. Secondly, the study offers insights into the use of the SOM tools, from data collected through a survey questionnaire and multiple case studies. These insights into how people effectively use and think about the SOM tools can help developers of complex commercial applications in VDM to gather new and interesting information about the tool, its users and their needs.

Regarding the UE process model (Table 7), this chapter contributes with providing guidance on steps 4, 5, 8 and 9 in that model. We developed a quality-of-use framework, which provides characteristics and attributes for the quality of visualization, interaction, and information of a VDM tool (4). We have developed a questionnaire for measuring the identified attributes by using *quantitative measures*. Moreover, we have designed a series of tasks for which we measured the user performance by employing *qualitative measures* (5). We chose to conduct a *subjective* evaluation, using an *inquiry method* (8). The evaluation was conducted by involving users in answering the questionnaire, in order to rate the quality-of-use attributes, and in performing the given tasks (9).

A limitation of the study is that the sample used in the exploratory study does not represent exactly the target population (business users). This drawback might be compensated by the fact that the participants have been asked to solve a real business problem with real data. The sample size is relatively small for being suitable for advanced statistical analysis. Another limitation is that we did not fully explore the relationships between evaluation levels, characteristics and attributes.

9. Conclusions

9.1. Summary

In this thesis, we have addressed three research problems. The first problem is the evaluation of projection-based visualizations with respect to their effectiveness in preserving the original distances between data points and the clustering structure of the data. The second problem is concerned with evaluating different visualization techniques as to their effectiveness for visual data mining of business data. The third problem is the evaluation of quality of use of VDM tools.

We have answered the research questions by developing and applying three different evaluation techniques in the evaluation of 11 visualization techniques. Table 31 presents the visualization techniques under investigation.

Table 31. Visualization techniques under evaluation

Visualization techniques	*Chapters in thesis*	*Type of evaluation method*
Principal Components Analysis (PCA)	5,6,7	Simulation; Inspection; Inquiry
Sammon's Mapping	5,6,7	Simulation; Inspection; Inquiry
Self-Organizing Map (SOM)	5,6,7,8	Simulation; Inspection; Inquiry; Inquiry
Radial Coordinate Visualization (Radviz)	5	Simulation
Star Coordinates	5	Simulation
Multiple Line Graphs	6,7	Inspection; Inquiry
Permutation Matrix	6,7	Inspection; Inquiry
Survey Plot	6,7	Inspection; Inquiry
Scatter Plot Matrix	6,7	Inspection; Inquiry
Parallel Coordinates	6,7	Inspection; Inquiry
Treemap	6,7	Inspection; Inquiry

In the development of the evaluation techniques, we have used the design science approach. The empirical evaluation of the visualization techniques has been based on the descriptive research approach.

We have provided a systematic approach to the evaluation problems, by designing, conducting and describing the evaluations in a systematic way. We have highlighted the evaluation activities (data collection, analysis), and their inputs (attributes, measures) and outputs (results). We have integrated the evaluation in the usability evaluation framework. The proposed evaluation techniques belong to the types of methods presented in Table 31 (i.e., Simulation, Inspection and Inquiry methods).

In Chapter 5, we have evaluated five projection techniques in terms of their effectiveness for preserving the clustering structure and the distances between data points. We have used known clustering validity measures in the evaluation. The evaluation has been conducted based on simulations on artificial and real benchmark datasets. The results show that the effectiveness depends on the dataset, but generally, the SOM, Sammon's Mapping and PCA are the most effective.

Our approach overcomes the limitation of the current approaches to objective evaluation of visualizations, which use quantitative measures that do not show a correlation with the effectiveness of the techniques in data mining tasks (e.g., Hoffman 1999; Keim and Kriegel 1996). However, we have focused only on the clustering task and on preserving the original data structure.

In Chapter 6, we have conducted an initial evaluation and comparison of nine visualization techniques for visual data mining (VDM) tasks related to a financial benchmarking problem. The results show that no single technique is capable of providing answers to all VDM tasks identified (outlier detection, dependency analysis, clustering, cluster description, class description, and comparison). The most effective techniques (answering the highest number of VDM tasks) are PCA, SOM–Feature Planes and SOM–All Views, Survey Plot and Treemap.

In Chapter 7, we have extended the study in Chapter 6 by involving users in the evaluation process. We have provided an evaluation technique based on the questionnaire data collection. The experiment has been limited to the evaluation of static graphical representations of the data. The results show that the most effective techniques for answering the VDM tasks are the SOM–All Views and SOM-Feature Planes, Survey Plot, Permutation Matrix, PCA and Parallel Coordinates.

The approach in Chapter 7, overcomes the limitation of the existing approaches that use a single user for subjectively evaluating the effectiveness of visualization in data mining tasks (e.g., Hoffman 1999; Keim and Kriegel 1996). Other approaches involve several users, but they focus on evaluating the techniques for their performance on different datasets, rather than in an applied context, such as a business problem (e.g., Ward and Theroux 1997). Our evaluation is similar to the one in (Ward and Theroux 1997) because in both cases, static visualizations are analyzed and different patterns are sought by the users. However, we analyze the collected data in a different manner, by analyzing the correctness of the answers in terms of identification of patterns and interpretation of the visualization. Ward and Theroux compare the users' answers with the patterns identified automatically by data mining techniques. Our approach may be more suitable in the early stage of development and evaluation of a system, for selecting appropriate techniques suitable for exploratory tasks.

In Chapter 8, we have provided a framework and an inquiry technique for evaluating the quality of use of VDM tools. The framework provides characteristics and attributes of quality of use at three levels of analysis: visualization, interaction and information. We have applied the framework to the evaluation of SOM-based tools, by developing and implementing a questionnaire. The evaluation results show that the SOM-based tools provide interesting and new information for the tasks given. The SOM technique was considered helpful in understanding and analyzing the data. Regarding the interaction with the tool, respondents found the tools easy to use and learn. Regarding the visualization, most of the features were found helpful and adequate and most attributes of data display and initial setting were positively appreciated. However, the evaluation also revealed weaknesses or limitations of the tools. These are identified in terms of "too long time needed to obtain a good map", relatively low accuracy, preciseness, correctness of the information, and difficulty in interpreting the results. These shortcomings might explain why business users do not frequently use the SOM tools for financial data analysis.

The current approaches to evaluating quality of use (usability) or some aspects of it, such as effectiveness or user satisfaction, have the limitation that the evaluation results do not provide detailed information that could help improving the visualization techniques. Moreover, the usability models suitable for evaluating information systems, in general, do not include attributes for evaluating the quality of visualization (e.g., Doll and Torkzadeh 1988; Kirakowski 1994). Our approach to the evaluation of quality of use of VDM tools overcomes these limitations, by providing characteristics and attributes of VDM tools at three levels of analysis (visualization, interaction and information) and an instrument for measuring those attributes subjectively, by user rating.

Table 32 presents the characteristics of each evaluation study taking into account the factors of categorizing evaluation methods (Section 4.1.3). We follow the approach illustrated in (Dix et al. 1998) for the characterization of different methods, but here we also use the factors presented by Whitefield et al. (1991) and the classification provided by Ivory and Hearst (2001). We have discussed the resources in the corresponding chapters.

Table 32. Characteristics of the evaluation studies in thesis

	Study 1	*Study 2*	*Study 3*
Chapter in thesis	Chapter 5	Chapters 6 and 7	Chapter 8
Technique's life-cycle stage	Implementation	Implementation	Implementation
Evaluation style	Laboratory	Laboratory	Laboratory
Subjective or objective evaluation	Objective	Subjective	Subjective
Qualitative or quantitative measures	Quantitative	Qualitative	Quantitative
Information provided	Low-level	High-level	Medium and High-level
Immediacy of results	Yes	Yes	No
Level of interference	No	No	No
Artifact presence	Representational	Real	Representational
User presence	-	Representational (Chapter 6) Real (Chapter 7)	Real
Method type (Ivory and Hearst 2001)	Simulation	Inspection (Chapter 6) Inquiry (Chapter 7)	Inquiry

9.2. Theoretical implications

The theoretical contributions of the thesis can be divided into two categories: methods and findings. Regarding the methods, we developed three evaluation techniques: two of them belonging to the class of inquiry techniques (Chapter 7 and Chapter 8) and one of them being a simulation technique (Chapter 5). In Chapter 5, we have used existing clustering validity measures in order to evaluate different projection techniques. We provided procedures for calculating the indices for the purpose of evaluating the projections. In Chapter 7, we have developed a questionnaire that can be easily modified according to the tasks and visualization techniques under evaluation. In Chapter 8, we have developed a framework that describes the characteristics and attributes of visual data mining tools. Based on this framework, we created a questionnaire that can be modified according to the tools and tasks evaluated.

Regarding the theoretical findings, our evaluation studies have highlighted a number of results that need to be further researched. In Chapter 5, we calculated different quantitative measures of projections effectiveness. The results show that the effectiveness of the projections depends on the datasets under evaluation. In addition, the results show that, on some datasets, a projection can improve the values of indices, when compared to their values when original data is used for clustering tasks.

In Chapter 7, we have observed that there are differences between the evaluation results among the three groups of users. However, we did not control the variables characterizing the groups, and we can only hypothesize that these observed differences are due to the differences between the groups.

9.3. Practical implications

The evaluation results reported at the end of Chapters 5 – 8 can be useful for practitioners (system developers and users) in selecting appropriate techniques for data visualization. Moreover, when new evaluations are needed, the questionnaires can be modified, if necessary, and used for the evaluation of other techniques and/or tasks. The results can also be used for improving certain visualization techniques or for developing complex visualization systems by integrating multiple techniques. For example, Chapter 7 provides summaries of the visualizations' capabilities for answering different tasks related to the financial benchmarking problem. These summaries can be used for selecting a set of techniques with the purpose to integrate them in a system that fully cover all tasks and data required for the financial benchmarking problem. Moreover, Chapter 8 provides a series of recommendations for improving the quality of use of the SOM tools.

Regarding the evaluation of the SOM in financial benchmarking, our results show that this technique is effective in revealing interesting patterns in the financial data, especially when all types of representation are used (SOM–Scatter Plot, SOM–U-matrix, SOM–Clustering, and SOM–Feature Planes). However, some weaker points in terms of quality-of-use attributes appear to be the time to obtain a good map and the accuracy of the output. In addition to these results, our evaluations show that the use of the SOM technique together with other techniques, such as Survey Plot and PCA, can provide a better and complete understanding of the data. Therefore, if the observed weaknesses are improved, and SOM is integrated with other visualization techniques to overcome some of its weak points, we believe that business users could successfully use the SOM in financial benchmarking.

9.4. Limitations and future work

We have highlighted the limitations of each study at the end of the corresponding chapters. Here, we highlight some of the limitations that need to be further addressed.

In Chapter 5, the limitation of the study is that we did not explore the effect of changing different parameters of the techniques on the measures. Moreover, the clustering of the datasets has been based on K-means technique and we did not explore the effects of using different clustering techniques.

In Chapter 7, the main limitation is that the users have evaluated static presentations, while interaction with the techniques could have provided different insights into the data. Another limitation is that some users have had previous experience of working with one technique (the Self-Organizing Map), while other techniques might have been new to the participants.

In Chapter 8, the main limitation is the relatively small sample size. This limits the use of the data for more advanced statistical analysis.

There are many research ideas that emerged during the research process of this thesis. Some of them are related directly to the limitations stated in each chapter and above. Other ideas come from the necessity to confirm the findings on other datasets and tasks, and to validate the proposed evaluation methods by applying them to different users and comparing the results. Moreover, other research questions can be formulated in order to explore the evaluation of other techniques or other settings.

The approach in Chapter 5 can be extended to other tasks (e.g., classification, information retrieval) by using measures suitable for those tasks. The approach in Chapter 7 can be improved by developing automated tools to evaluate the answers to the open questions used in the questionnaire. The work presented in Chapter 8 can be extended to exploring/confirming the relationships between the quality-of-use characteristics and attributes.

References

Agrawal, R., Gehrke, J., Gunopulos, D., and Raghavan, P. (1998) Automatic subspace clustering of high dimensional data for data mining applications. In *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data* (Seattle, Washington, United States, June 01 - 04, 1998). A. Tiwary and M. Franklin, Eds. SIGMOD '98. ACM, New York, NY, 94-105.

Alhoniemi, E. (2000) Analysis of pulping data using the self-organizing map. In *Tappi Journal, 83*(7): 66.

Back, B., Irjala, M., Sere, K., and Vanharanta, H. (1998) Competitive financial benchmarking using Self-Organizing Maps. In M-A. Vasarhelyi and A. Kogan (eds.), *Artificial Intelligence in Accounting and Auditing. Towards New Paradigms*. Princeton, NJ, Vol. IV:69-81.

Back, B., Öström, K., Sere, K., and Vanharanta, H. (2000) Analyzing company performance using internet data. In *Proceedings of the 11th Meeting of the Euro Working Group on DSS*, Ed. by Zaraté, Toulouse, France, 52-56.

Barlow, T. and Neville, P. (2001) A comparison of 2-D visualizations of hierarchies. In *Proc. of the IEEE Symposium on Information Visualization 2001 (Infovis'01)* (October 22 - 23, 2001). INFOVIS. IEEE Computer Society, Washington, DC, 131.

Bendell, T., Boulter, L., and Kelly, J. (1998) *Benchmarking for Competitive Advantage*. Pitman Publishing, London.

Bertin, J., (1981) *Graphics and Graphics Information Processing*, Berlin: De Gruyter.

Bevan, N. (1995) Measuring usability as quality of use. In *Software Quality Journal*, 4(2):115-130.

Buckingham, A. and Saunders, P. (2004) *The Survey Methods Workbook: From Design to Analysis*, Cambridge: Polity Press.

Card S.K., Moran T.P., and Newell A. (1983) *The Psychology of Human-Computer Interaction*, Hillsdale, Lawrence Erlbaum Associates, N.J.

Card, S. K., Pirolli, P., and Mackinlay, J. D. (1994) The cost-of-knowledge characteristic function: Display evaluation for direct walk information visualizations. In *Proc. of CHI'94*, 238-244.

Card, S.K., Mackinlay, J.D, and Shneiderman, B. (1999) *Readings in Information Visualization - Using Vision to Think*, Morgan Kaufmann Publishers.

Catarci, T. and Cruz, I. F. (1996) Information visualization: Guest editors' foreword. In *ACM Sigmod Record* 25 (4).

Chen, C. and Czerwinski, M. P. (2000) Empirical evaluation of information visualization: An introduction. In *Int. J. Human-Computer Studies*, 53:631-635.

Chen, K. and Liu, L. (2004) VISTA: Validating and refining clusters via visualization. In *Information Visualization* (2004) 3(4): 257-270.

Cleveland, W. S. (1993) *Visualizing Data*. Hobart Press, Summit, New Jersey.

Costea, A. and Eklund, T. (2003) A two-level approach to classifying countries/companies economic/financial performance. In *Proc. 36th Annual Hawai'i International Conf. on System Sciences (HICSS*'36), Decision Technologies for Management track.

Coutaz, J. (1994) Evaluation techniques: Exploring the intersection of HCI and software engineering. In *Proc. of the ICSE Workshop on software Engineering and HCI 1994*, LNCS, Springer-Verlag, 896:35-48.

Creswell, J. W. (2003) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, Thousands Oaks, CA: SAGE.

Demsar, J., Zupan, B., and Leban, G. (2004) *Orange: From Experimental Machine Learning to Interactive Data Mining.* White Paper (www.ailab.si/orange), Faculty of Computer and Information Science, University of Ljubljana.

Dix, A., Finlay, J., Abowd, G., and Beale, R. (1998) *Human-Computer Interaction*, Second Edition, Prentice Hall.

Doll, W.J. and Torkzadeh, G. (1988) The measurement of end-user computing satisfaction. In *MIS Quarterly*, 12 (2): 259-274.

Duda, R. O., Hart, P. E., and Stork, D. G. (2000) *Pattern Classification* (2nd Edition). Wiley-Interscience.

Dull, R. B. and Tegarden, D. P. (1999) A comparison of three visual representations of complex multidimensional accounting information. In *Journal of Information Systems*, 13(2): 117-131.

Eklund, T. (2004) *The Self-Organizing Map in Financial Benchmarking*, Doctoral Dissertation, Åbo Akademi University, Turku Centre for Computer Science, Turku, Finland.

Eklund, T., Back, B., Vanharanta, H., and Visa, A. (2003) Using the Self-Organizing Map as a visualization tool in financial benchmarking. In *Information Visualization*, 2(3):171-181.

Eklund, T., Back, B., Vanharanta, H., and Visa, A. (2004) *Financial Benchmarking Tools in Finnish Companies – A State of the Art Survey*. Technical Report 618, TUCS, Aug 2004.

Elmasri, R. and Navathe, S. (2000) *Fundamentals of Database Systems*, Third ed., Addison–Wesley.

Fayyad, U. and Grinstein, G. G. (2002) Introduction. In U. Fayyad, G.G. Grinstein, and A. Wierse (eds.), *Information Visualization for Data Mining and Knowledge Discovery*, San Francisco: Morgan Kaufmann, 1-17.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996) From data mining to knowledge discovery: An overview. In U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (eds.), *Advances in Knowledge Discovery and Data Mining*, Chapter 1, AAAI Press, Menlo Park, CA, and MIT Press, Cambridge, AM, 1-34.

Fenton, N. E. and Pfleeger, S. L. (1997) *Software Metrics*, Second ed., PWS Publishing Company, Boston, MA.

Ferreira de Oliveira, M. C. and Levkowitz, H. (2003) From visual data exploration to visual data mining: A survey. In *IEEE Transactions on Visualization and Computer Graphics*, 9(3):378-394.

Grinstein, G. G., and Ward, M. O., (2002) Introduction to data visualization. In U. Fayyad, G.G. Grinstein, and A. Wierse (eds.), *Information Visualization for Data Mining and Knowledge Discovery*, San Francisco: Morgan Kaufmann.

Grinstein, G. G., Hoffman, P. E., Pickett, R. M., and Laskowski, S. J. (2002) Benchmark development for the evaluation of visualization for data mining. In U. Fayyad, G.G. Grinstein, and A. Wierse (eds.), *Information Visualization for Data Mining and Knowledge Discovery*, San Francisco: Morgan Kaufmann.

Guo, D. (2003) Coordinating computational and visual approaches for interactive feature selection and multivariate clustering. In *Information Visualization*, 2: 232-246.

Guo, D., Gahegan, M., Peuquet, D., and MacEachren, A. (2003) Breaking down dimensionality: An effective feature selection method for high-dimensional clustering. In *Proceedings of the third SIAM International Conference on Data Mining, Workshop on Clustering High Dimensional Data and its Applications*. San Francisco, CA, USA, 29-42.

Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2002a) Cluster validity methods: part I. In *SIGMOD Rec*., 31(2):40-45.

Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2002b) Cluster validity methods: part II. In *SIGMOD Rec*., 31(3):19-27.

Hartwig, F. and Dearing, B. E. (1990) *Exploratory data analysis* (19^{th} printing), in Quantitative Applications in the Social Sciences, 16, Sage Publications.

Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004) Design science in information systems research. In *MIS Quarterly* 28 (1):75-105.

Hinterberger, H. H., and Schmid, C. (1993) Reducing the influence of biased graphical perception with automatic permutation matrices. In *Proc. of the Seventh Conference of the Scientific Use of Statistic-Software, SoftStat '93*, March 14–18.

Hoffman, P. E. (1999) *Table Visualizations: A Formal Model and Its Applications*. PhD Thesis, University of Massachusetts Lowell.

Hoffman, P. E. and Grinstein, G. G. (2002) A survey of visualizations for high-dimensional data mining. In U. Fayyad, G.G. Grinstein, and A. Wierse (eds.), *Information Visualization for Data Mining and Knowledge Discovery*, San Francisco: Morgan Kaufmann.

Hoffman, P., Grinstein, G., and Pinkney, D. (1999) Dimensional anchors: a graphic primitive for multidimensional multivariate information visualizations. In *Proc. 1999 Workshop on New Paradigms in Information Visualization and Manipulation, in Conjunction with the 8h ACM Int'l. Conf. Information and Knowledge Management (CIKM '99),* 9-16.

Hoffman, P. E., Grinstein, G. G., Marx, K., Grosse, I., and Stanley, E. (1997) DNA visual and analytic data mining. In *Proceedings of IEEE Visualization '97*, Phoenix, Az, 437-441.

Hornbæk, K. (2006) Current practice in measuring usability: Challenges to usability studies and research. In *Int. J. Human-Computer Studies*, 64:79-102.

Hubert, L. J. and Schultz, J., (1976) Quadratic assignment as a general data-analysis strategy. In *British Journal of Mathematical and Statistical Psychology*, 29:190-241.

Iivari, J. (1991) A paradigmatic analysis of contemporary schools of IS development. In *Eur. J. Inf. Systs*. 1(4): 249-272.

Inselberg, A. (1985) The plane with parallel coordinates. In Special Issue on Computational Geometry, *The Visual Computer,* 1: 69-91.

Inselberg, A. (1997) Multidimensional detective. In *Proceedings of the 1997 IEEE Symposium on information Visualization (Infovis '97)* (October 18 - 25, 1997). INFOVIS. IEEE Computer Society, Washington, DC.

Inselberg, A. and Dimsdale, B. (1990) Parallel Coordinates: A Tool for visualizing multi-dimensional geometry. In *Proceedings of IEEE Visualization '90 Conference*, Los Alamitos, CA, 361-375.

ISO (1998) ISO 9241-11 Ergonomic requirements for office work with visual display terminals. Part 11: Guidance on usability.

ISO (1999a) ISO 13407 Human-centred design processes for interactive systems.

ISO (1999b) ISO/IEC 14598-1 Information Technology – Software product evaluation – Part 1: General overview.

ISO (2000a) ISO 18529 Ergonomics – Ergonomics of human-system interaction – Human-centred lifecycle process descriptions.

ISO (2000b) ISO/IEC 9126-1 Information Technology – Software product quality – Part 1: Quality model.

ISO (2004) ISO/IEC 9126-4 Software Engineering – Product quality – Part 4: Quality in use metrics.

Ivory, M. Y. and Hearst, M. A. (2001) The state of the art in automating usability evaluation of user interfaces. In *ACM Computing Surveys*, 33(4):470-516.

Jain, A. K., Murty, M. N., and Flynn, P. J. (1999) Data clustering: A review. In *ACM Comput. Surv.,* vol. 31: 264-323.

Johnson, B. (1993) *Treemaps: Visualizing Hierarchical and Categorical Data*. Unpublished Ph.D. Thesis, University of Maryland.

Johnson, B. and Shneiderman, B. (1991) Treemaps: A space-filling approach to the visualization of hierarchical information structures. In *Proc. of IEEE Visualization '91 Conference*, San Diego, 284-291.

Järvinen, P. (2001) *On Research Methods*. Opinpajan Kirja, Tampere.

Kandogan, E. (2000) Star Coordinates: A multidimensional visualization technique with uniform treatment of dimensions. In *Proc. of the IEEE Information Visualization Symposium,* Late Breaking Hot Topics.

Kandogan, E. (2001) Visualizing multi-dimensional clusters, trends, and outliers using star coordinates. In *Proc. of the Seventh ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining,* 107-116.

Kasanen, E., Lukka, K., and Siitonen, A. (1993) The constructive approach in management accounting research. In *Journal of Management Accounting Research*, 5:243-264.

Kaski, S. and Kohonen, T. (1996) Exploratory data analysis by the Self-Organizing Map: Structures of welfare and poverty in the world. In *Proc. 3rd International Conf. on Neural Networks in the Capital Markets*, London, England, 11-13 October, 1995, Publisher World Scientific, 498-507.

Kaski, S., Kangas, J., and Kohonen, T. (1998) Bibliography of Self-Organizing Map (SOM) papers 1981-1997. In *Neural Computing Surveys*, *1*(3&4): 1-176.

Keim, D. A. (1996) Pixel-oriented visualization techniques for exploring very large databases. In *Journal of Computational and Graphical Statistics*, 5(1), pp. 58-77.

Keim, D. (2000) Designing pixel-oriented visualization techniques: Theory and applications. In *IEEE Trans. Visualization and Computer Graphics*, 6 (1): 59–78.

Keim, D. A. (2001) Visual exploration of large data sets. In *Commun. ACM* , 44 (8) (Aug. 2001): 38-44.

Keim, D.A. (2002) Information visualization and visual data mining. In *IEEE Transactions on Visualization and Computer Graphics*, 8 (1).

Keim, D. A. and Kriegel, H.-P. (1994) Visdb: Database exploration using multidimensional visualization. In *Computer Graphics & Applications*, 6 (Sept.): 40–49.

Keim, D. A. and Kriegel, H.-P. (1996) Visualization techniques for mining large databases: A comparison. In *IEEE Transactions on Knowledge and Data Engineering*, 8 (6): 923–938.

Keim, D., Bergeron, R. D., Pickett, R. (1994) Test data sets for evaluating data visualization techniques. In *Database Issues for Data Visualization*, *IEEE Visualization '93 Workshop*, Springer-Verlag, Berlin.

Kirakowski, J. (1994) *The Use of Questionnaire Methods for Usability Assessment* [online], http://www.ucc.ie/hfrg/questionnaires/sumi/index.html (accessed September 12, 2008).

Kohavi, R., Rothleder, N.J., and Simoudis, E. (2002) Emerging Trends in Business Analytics. In *Communications of the ACM*, 45 (8), Aug., 45–48.

Kohonen, T. (2001) *Self-Organizing Maps*, Third edition, Berlin: Springer-Verlag.

Leung, Y. K. and Apperley, M. D. (1994) A review and taxonomy of distortion-oriented presentation techniques. In *ACM Trans. Comput.-Hum. Interact.* 1, 2 (Jun. 1994), 126-160.

Liu, Y. and Salvendy, G. (2007) Design and evaluation of visualization support to facilitate decision trees classification. In *Int. J. Human-Computer Studies*, 65:95-110.

Mackinlay, J.D. (1986) Automating the design of graphical presentations of relational information. In *ACM transactions on Graphics*, 5 (2): 111-141.

MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations. In *Proc. Fifth Berkeley Symposium on Math. Stat. and Prob.,*1:281-297.

Mandl, T. and Eibl, M. (2001) Evaluating visualizations: A method for comparing 2D maps. In *Proceedings of the HCI International 2001* (*9th International Conference on Human-Computer Interaction*), New Orleans, August 2001. Mahwah, NJ; London: Lawrence Erlbaum Associates, 1145-1149.

March, S. T. and Smith, G. F. (1995) Design and natural science research on information technology, in *Decision Support Systems*, 15(4): 251-266.

Marghescu, D. (2006) Evaluating the effectiveness of projection techniques in visual data mining. In *Proceedings of the 6th IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP 2006),* Palma de Mallorca, Spain, August 2006, ACTA PRESS, 94-103.

Marghescu, D. (2007a) Evaluation of projection techniques using Hubert's Γ statistics. In *Proceedings of the IADIS 1st European Conference on Data Mining*, Lisbon, Portugal, July 2007, IADIS Press, 192-197.

Marghescu, D. (2007b) Multidimensional data visualization techniques for exploring financial performance data. In *Proceedings of the 13th Americas' Conference on Information Systems (AMCIS 2007)*, Keystone, Colorado, USA, August 2007.

Marghescu, D. (2007c) User evaluation of multidimensional data visualization techniques for financial benchmarking. In *Proceedings of the European Conference on Information Management and Evaluation (ECIME 2007),* Montpellier, France, September 2007.

Marghescu, D. (2008) Usability evaluation of information systems: A review of five international standards, in Barry, C., Lang, M., Wojtkowski, W., Wojtkowski, G., Wrycza, S., & Zupancic, J. (eds) *The Inter-Networked World: ISD Theory, Practice, and Education,* Springer-Verlag: New York, (*Proceedings of the 16th International Conference on Information Systems Development (ISD 2007)*, Galway, Ireland, August 2007). In Press.

Marghescu, D., Rajanen, M., and Back, B. (2004) Evaluating the quality of use of visual data-mining tools. In *Proceedings of the 11th European Conference on IT Evaluation (ECITE 2004),* Amsterdam, The Netherlands, November 2004.

Milligan, G. W. (1996) Clustering validation: results and implications for applied analyses. In P. Arabie, L. J. Hubert, & G. De Soete (Eds.), *Clustering and Classification* (pp. 341-375), River Edge, NJ: World Scientific Publishing.

Nenet 1.1 (1999) http://koti.mbnet.fi/~phodju/nenet/Nenet/General.html (accessed March 5th 2008).

Newman, D. J., Hettich, S., Blake, C. L., and Merz, C. J. (1998) *UCI Repository of Machine Learning Databases*, http://www.ics.uci.edu/~mlearn/MLRepository.html.

Nielsen, J. (1993) *Usability Engineering*, Morgan Kaufmann Publishers Inc.

Oja, E., Kaski, S. and Kohonen, T. (2003) Bibliography of Self-Organizing Map (SOM) papers: 1998-2001 Addendum. In *Neural Computing Surveys,* 3:1-156.

Parsons, L., Haque, E., and Liu, H. (2004) Subspace clustering for high dimensional data: A review. In *SIGKDD Explor. Newsl*. 6, 1 (Jun. 2004), 90-105.

Pickett, R. M. and Grinstein, G. G. (2002) Evaluation of visualization systems. Chapter 3 in Fayyad, U., Grinstein, G. G., and Wierse, A. (eds.), *Information Visualization for Data Mining and Knowledge Discovery*, Morgan Kaufmann, San Francisco.

Pillat, R. M., Valiati, E. R., and Freitas, C. M. (2005) Experimental study on evaluation of multidimensional information visualization techniques. In *Proc. of the 2005 Latin American Conference on Human-Computer Interaction* (Cuernavaca, Mexico, October 23 - 26, 2005). CLIHC '05, vol. 124. ACM, New York, NY, 20-30.

Pirolli, P. and Rao, R. (1996) Table Lens as a tool for making sense of data. In *Proc. of AVI'96, Workshop on Advanced Visual Interfaces,* Gubbio, Italy. 67-80.

Plaisant, C. (2004) The challenge of information visualization evaluation. In *Proc. of the Working Conf. on Advanced Visual Interfaces (AVI 2004)*, ACM Press, 109-116.

Pyle, D. (1999) *Data Preparation for Data Mining*, San Francisco, CA: Morgan Kaufmann.

Risden, K., Czerwinski, M.P., Munzner, T., and Cook, D.B. (2000) An initial examination of ease of use for 2D and 3D information visualizations of Web content. In *International Journal of Human-Computer Studies*, 53: 695-714.

Sammon, J. W. (1969) A nonlinear mapping for data structure analysis. In *IEEE Transactions on Computers* C-18 (5): 401–409.

Seffah, A., Donyaee, M., Kline, R. B., and Padda, H. K. (2006) Usability measurement and metrics: A consolidated model. In *Software Quality J*., 14(2): 159-178.

Seo, J. (2005) *Information Visualization Design for Multidimensional Data: Integrating the Rank-By-Feature Framework with Hierarchical Clustering*, Ph.D. Dissertation, Dept. of Computer Science, University of Maryland, College Park, MD, USA.

Sharma, S. (1995) *Applied Multivariate Techniques*, New York, NY: John Wiley & Sons.

Shneiderman, B. (1992) Tree visualization with Treemaps: A 2-dimensional space filling approach. In *ACM Transactions on Graphics* 11(1): 92–99.

Schneiderman, B. (1994) Dynamic queries for visual information seeking. In *IEEE Software,* 11(6): 70-77.

Shneiderman, B. (1996) The eye have it: A task by data type taxonomy for information visualization. In *Proc. of IEEE Workshop on Visual Languages'96*, 336–343.

Siegel, S. and Castellan, N. J. Jr. (1988) *Nonparametric Statistics for the Behavioral Sciences*, second edition, McGraw-Hill Book Company.

SOM_PAK 3.1 (1995) http://www.cis.hut.fi/research/som_lvq_pak.shtml (accessed March 5th 2008).

SOM Toolbox 2.0 for Matlab (2005) http://www.cis.hut.fi/projects/somtoolbox/ (accessed March 5th 2008).

Soukup, T. and Davidson, I. (2002) *Visual Data Mining: Techniques and Tools for Data Visualization and Mining*, John Wiley & Sons.

Spata, A. V. (2003) *Research Methods: Science and Diversity*, John Wiley & Sons, New York, NY.

Stasko, J., Catrambone, R., Guzdial, M., and McDonald, K. (2000) An evaluation of space-filling information visualizations for depicting hierarchical structures. In *Int. J. Human-Computer Studies*, 53: 663-694.

Straub, D., Boudreau, M-C., and Gefen, D. (2004) Validation guidelines for IS positivist research, in *Communications of the Association for Information Systems*, Vol. 13, Article 24: 380-427.

Ståhl, P., Eklund, T., Tétard, F., and Back, B. (2006) *A cooperative usability evaluation of the Domino software*. Technical report Number 775, Turku Centre for Computer Science, Turku, Finland.

Sutcliffe, A. G., Ennis, M., and Hu, J. (2000) Evaluating the effectiveness of visual user interfaces for information retrieval. In *Int. J. Human-Computer Studies*, 53: 741-763.

Theodoridis, S. and Koutroumbas, K. (1999) *Pattern Recognition,* Academic Press, San Diego.

Treemap 4.1 (2004) http://www.cs.umd.edu/hcil/treemap/ (accessed Sept. 12, 2008).

Tukey, J. W. and Tukey, P. A. (1985) Computer graphics and exploratory data analysis: An introduction. In *Proceedings of Annual Conference and Exposition: Computer Graphics*, 3: 773-785.

Tufte, E.R. (1983) *The Visual Display of Quantitative Information*, Cheshire, CT: Graphics Press.

Ultsch, A. and Siemon, H. (1989) *Technical Report 329*. Univ. of Dortmund, Dortmund, Germany.

Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J. (2000) *SOM Toolbox for Matlab 5*, Report A57, Helsinki University of Technology, Libella Oy, Espoo, http://www.cis.hut.fi/projects/somtoolbox/package/papers/techrep.pdf

Ward, M. O. and Theroux, K. J. (1997) Perceptual benchmarking for multivariate data visualization. In *Proceedings of the Conference on Scientific Visualization* (June 09 - 13, 1997). DAGSTUHL. IEEE Computer Society, Washington, DC, 314-321.

Whitefield, A., Wilson, F., and Dowell, J. (1991) A framework for human factors evaluation. In *Behaviour & Information Technology*, 10 (1): 65-79.

Witten, I.H. and Frank, E. (2000) *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufmann, San Francisco, CA.

APPENDIX 1. The projections of the Wine recognition data (Chapter 5)

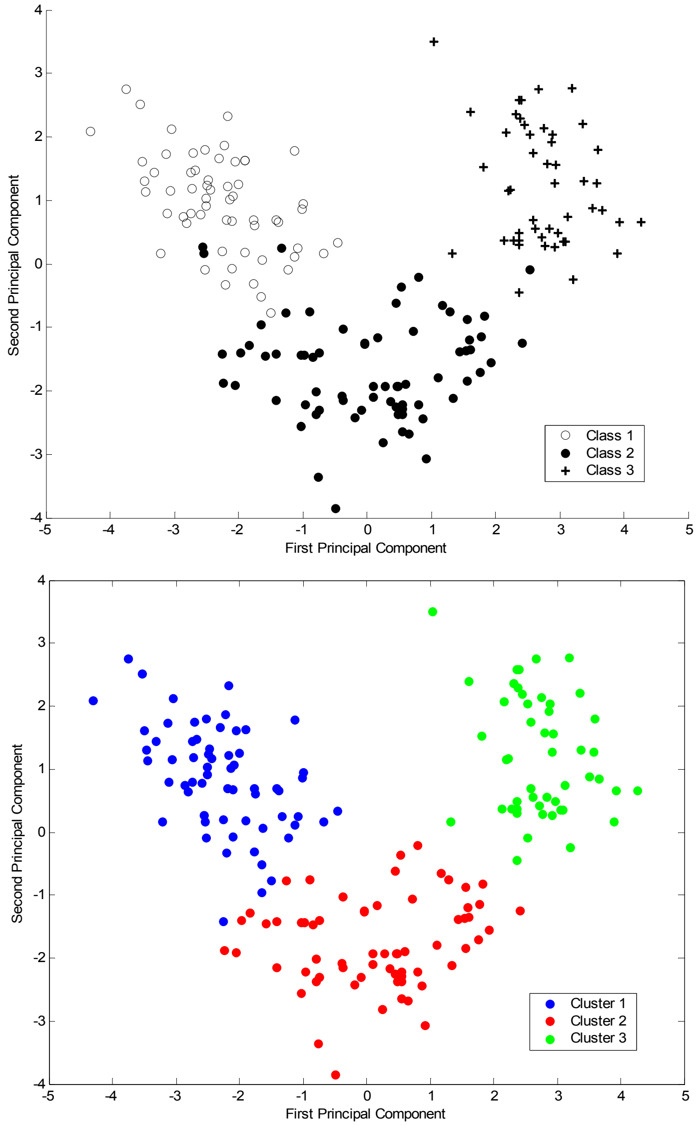


Figure 55. PCA (Wine data). Up: known classes; Down: obtained clusters

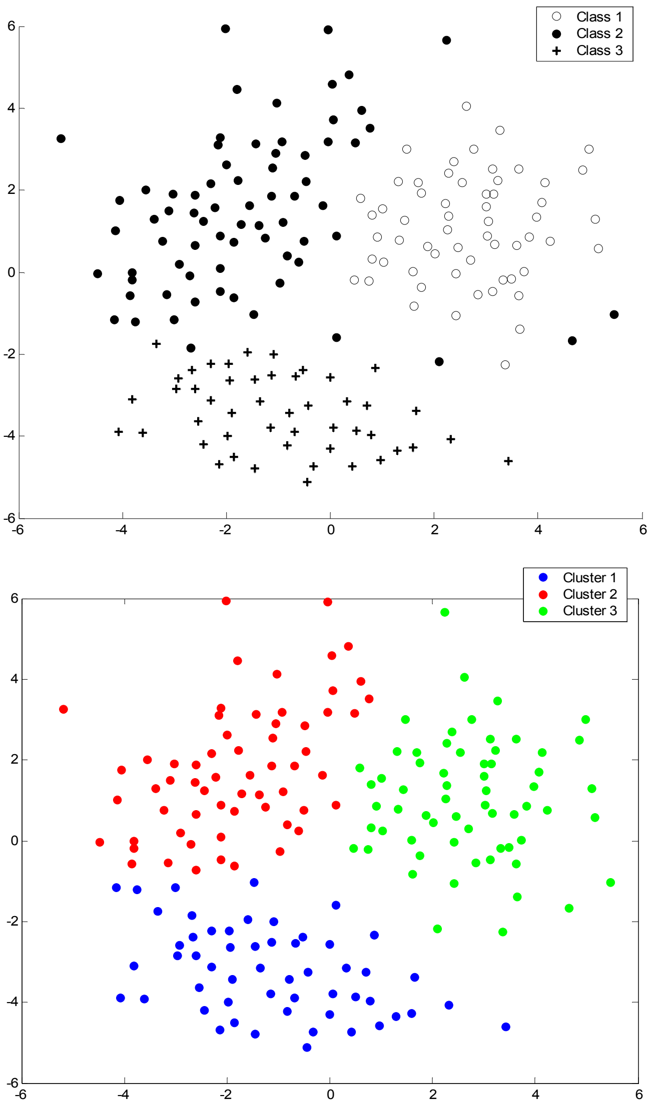


Figure 56. Sammon's Mapping (Wine data). Up: known classes; Down: obtained clusters

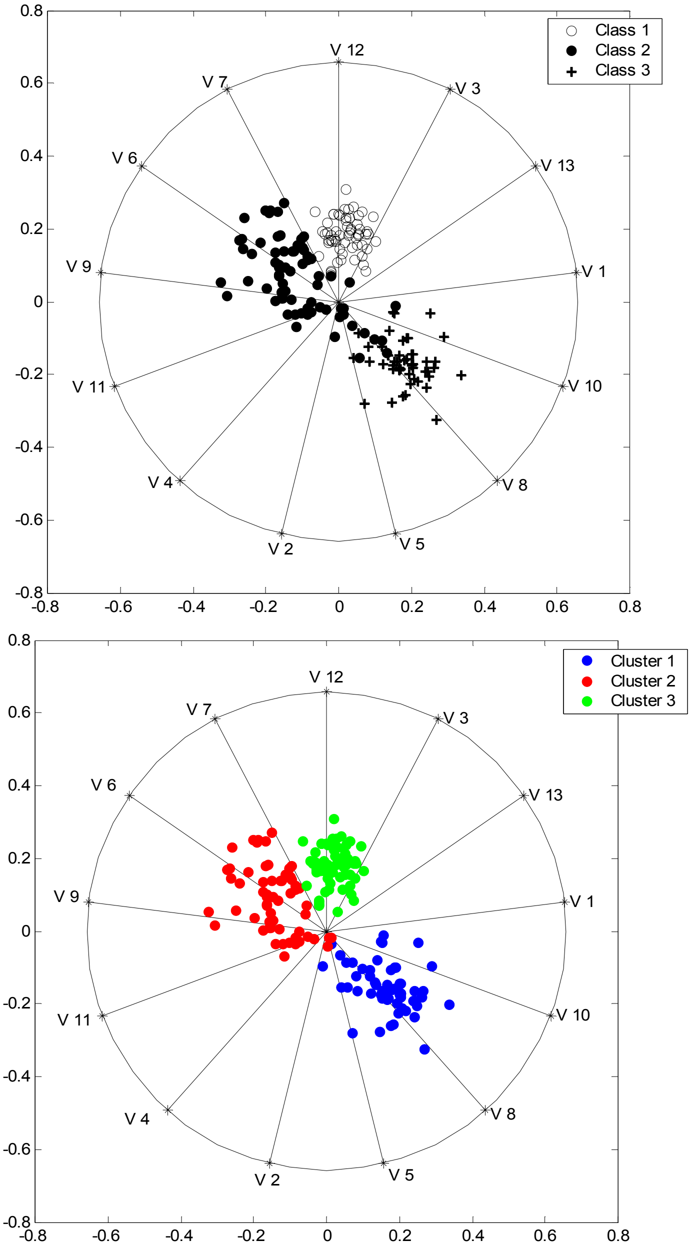


Figure 57. Radviz (Wine data). Up: known classes; Down: obtained clusters

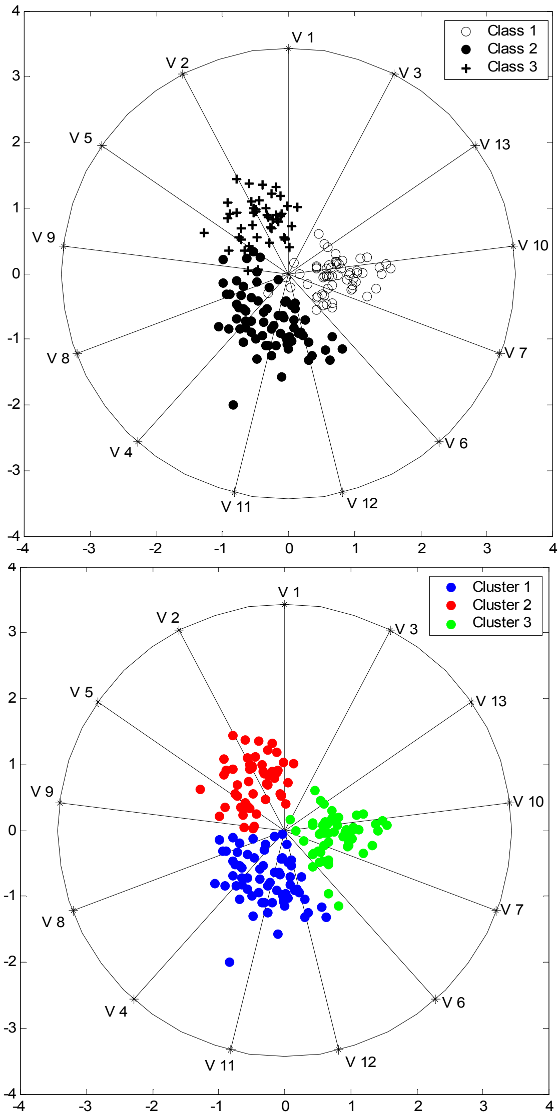


Figure 58. Star Coordinates (Wine data). Up: known classes; Down: obtained clusters

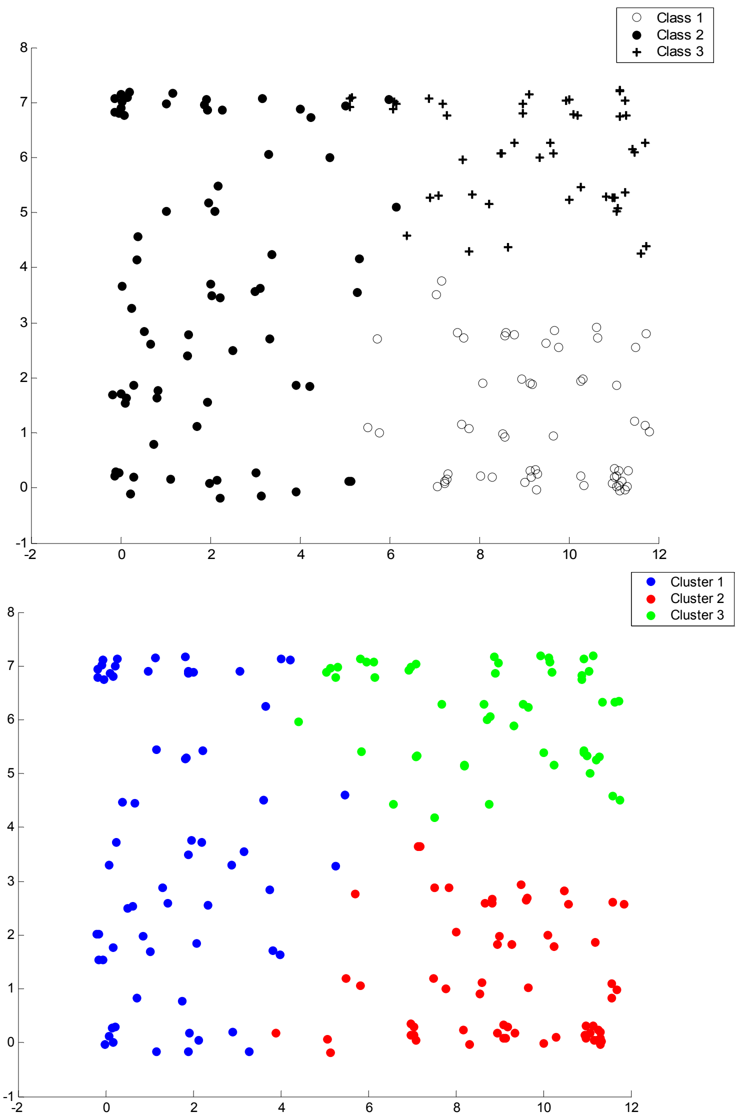
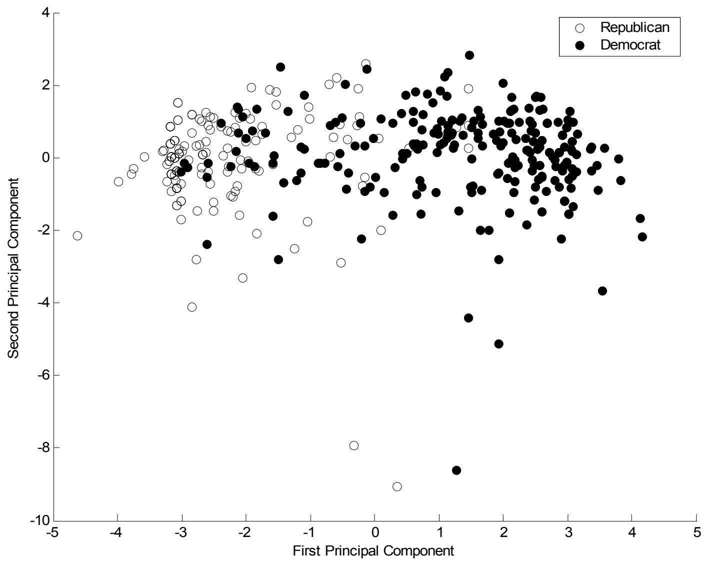


Figure 59. SOM (Wine data). Up: known classes; Down: obtained clusters

APPENDIX 2. The projections of the Voting records data (Chapter 5)



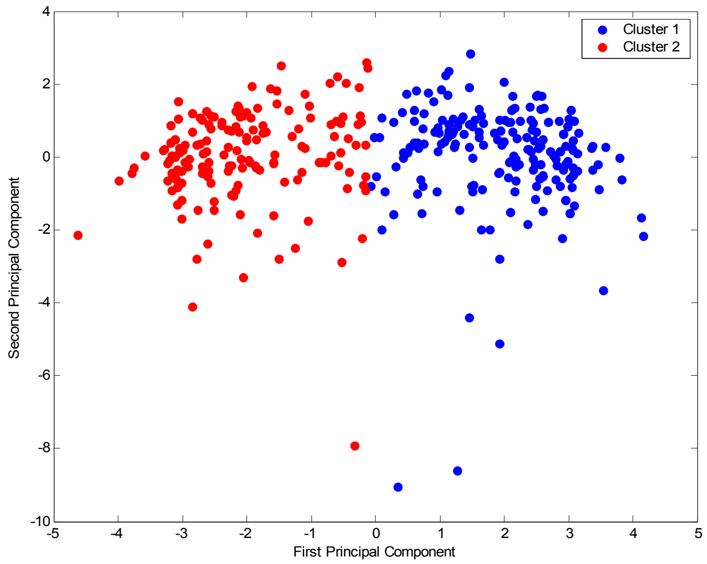
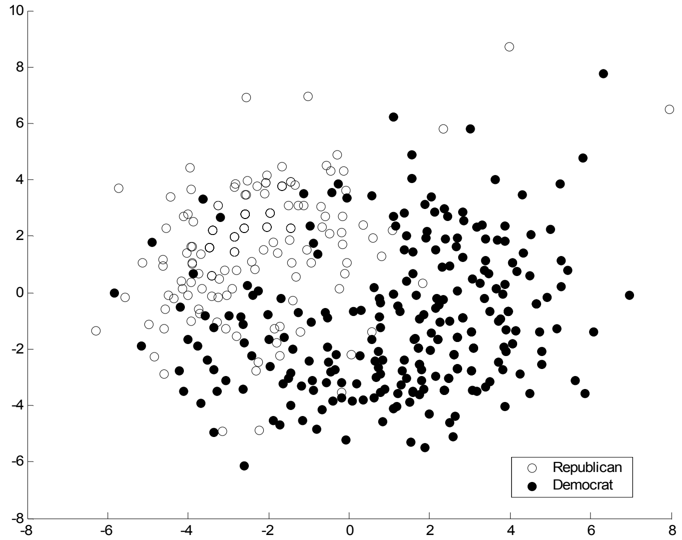


Figure 60. PCA (Voting data). Up: known classes; Down: obtained clusters



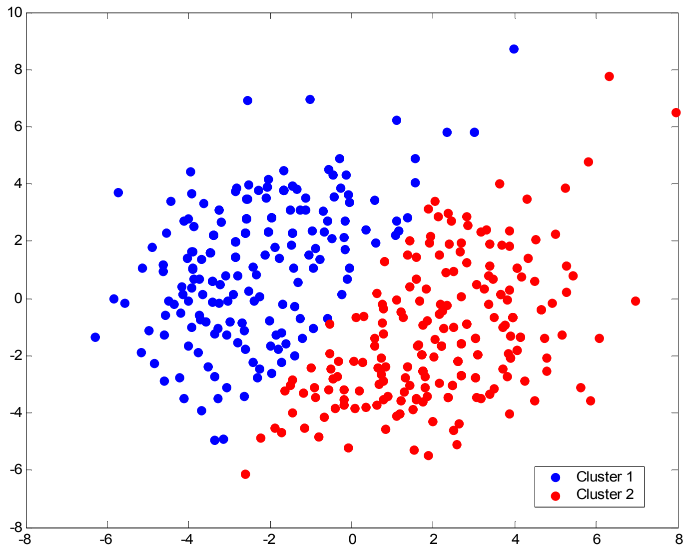


Figure 61. Sammon's Mapping (Voting data). Up: known classes; Down: obtained clusters

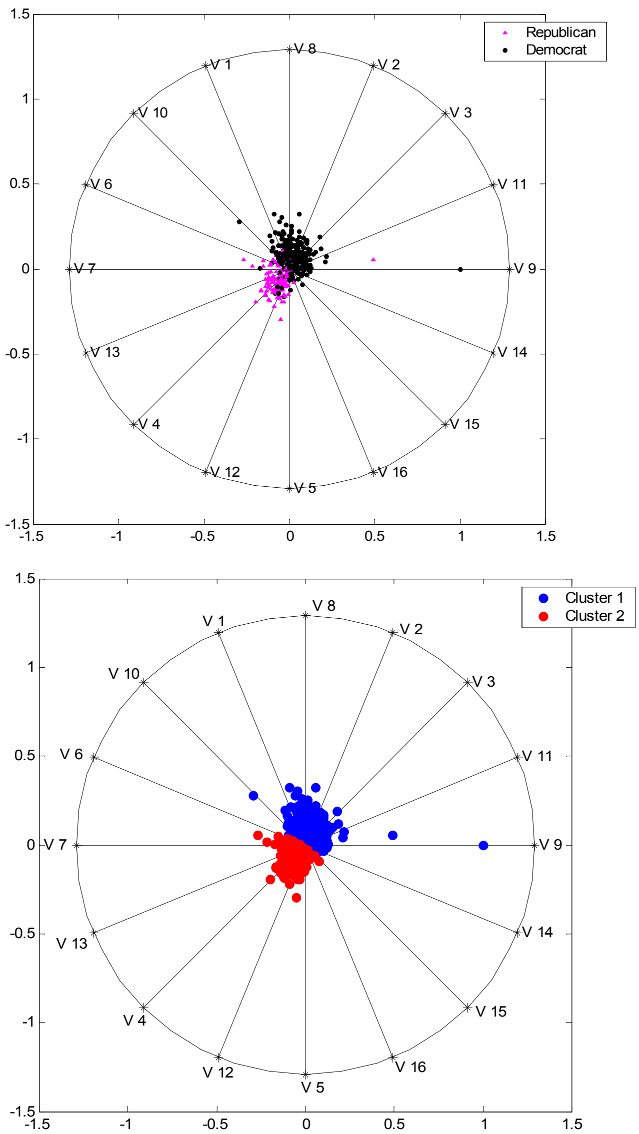


Figure 62. Radviz (Voting data). Up: known classes; Down: obtained clusters

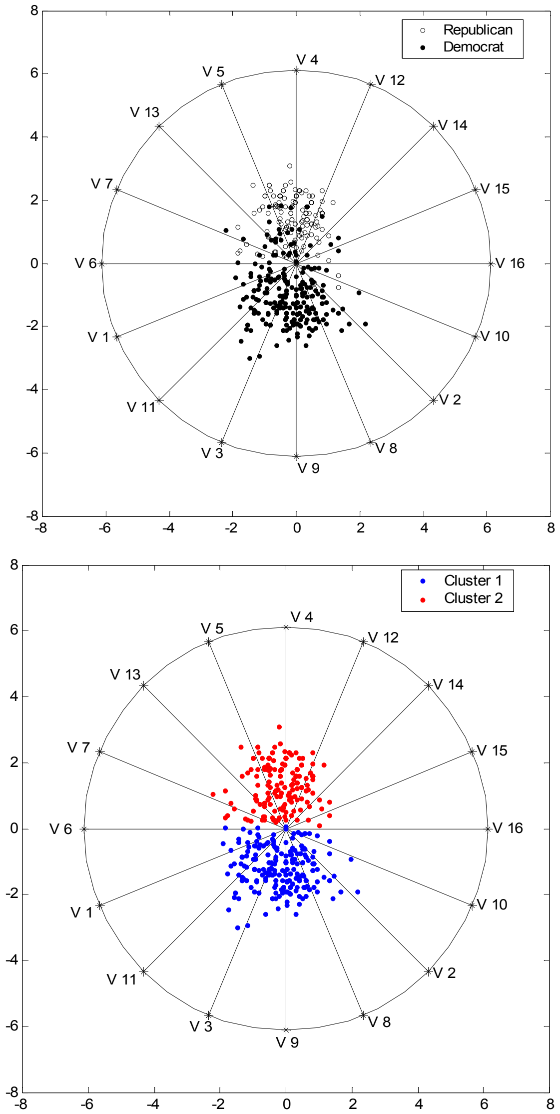


Figure 63. Star Coordinates (Voting data). Up: known classes; Down: obtained clusters

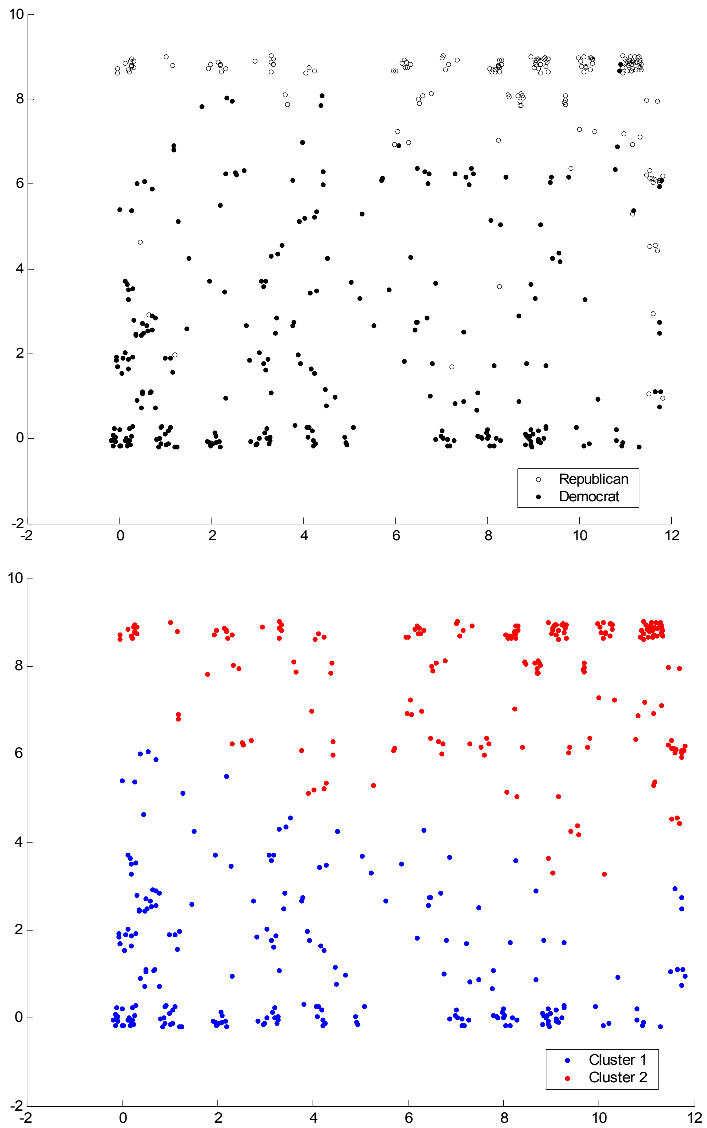


Figure 64. SOM (Voting data). Up: known classes; Down: obtained clusters

APPENDIX 3. The projections of the Artificial 2 data (Chapter 5)

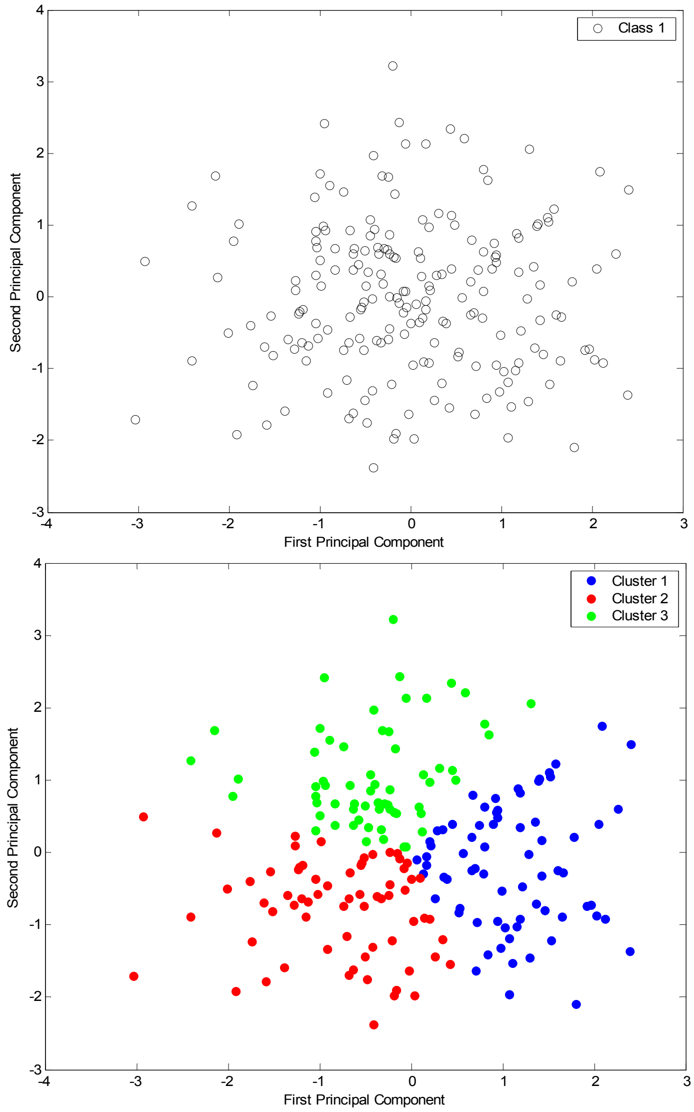


Figure 65. PCA (Artificial 2 data). Up: known classes; Down: obtained clusters

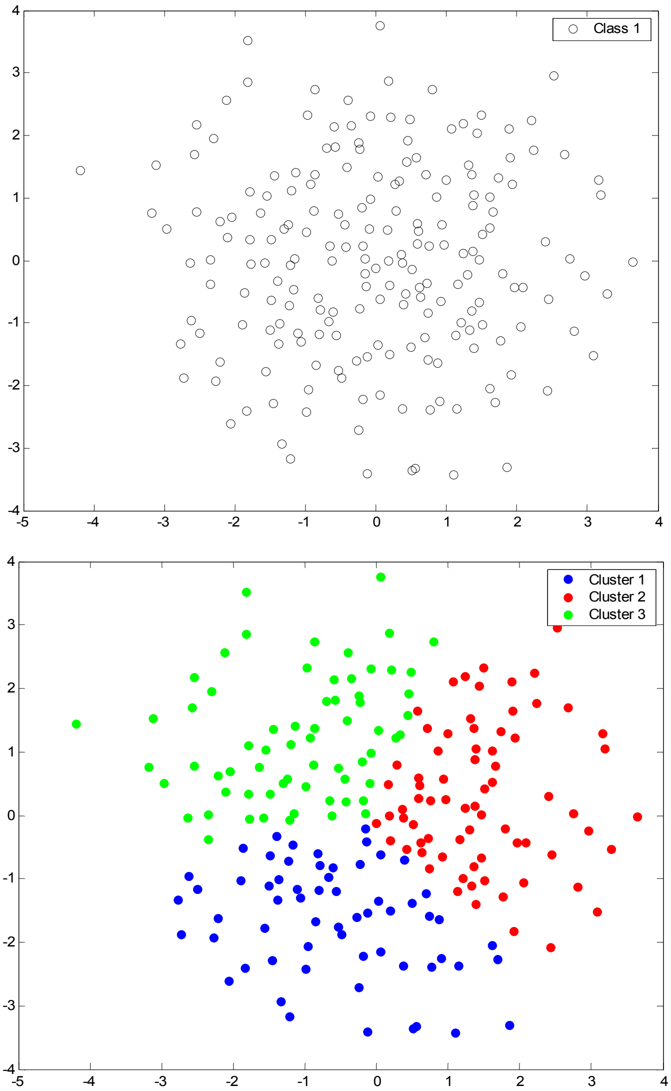


Figure 66. Sammon's Mapping (Artificial 2 data). Up: known classes; Down: obtained clusters

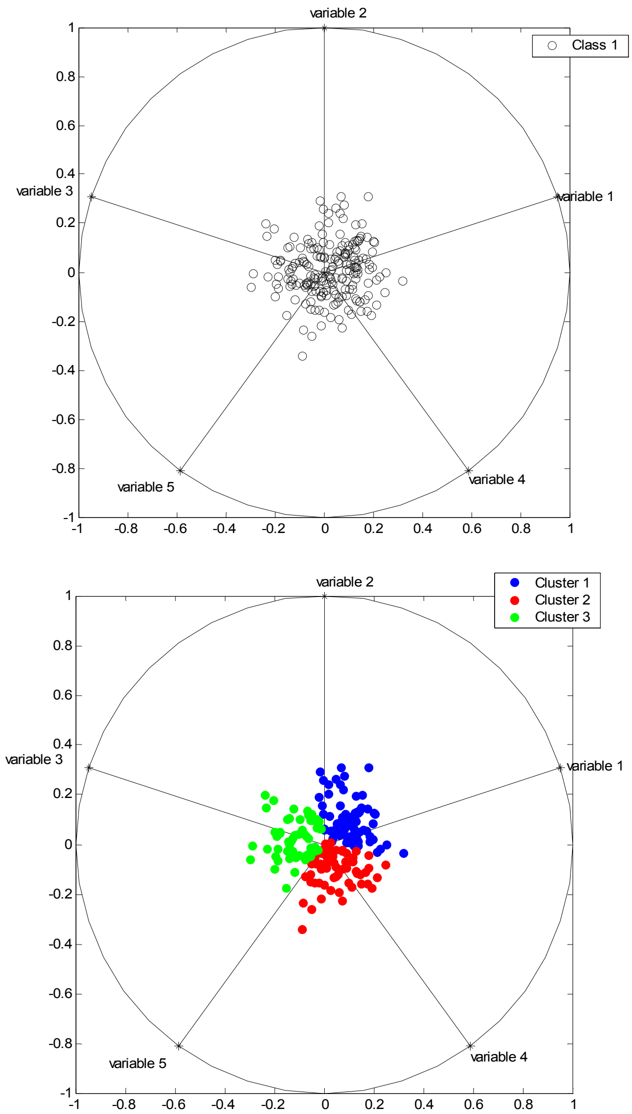


Figure 67. Radviz (Artificial 2 data). Up: known classes; Down: obtained clusters

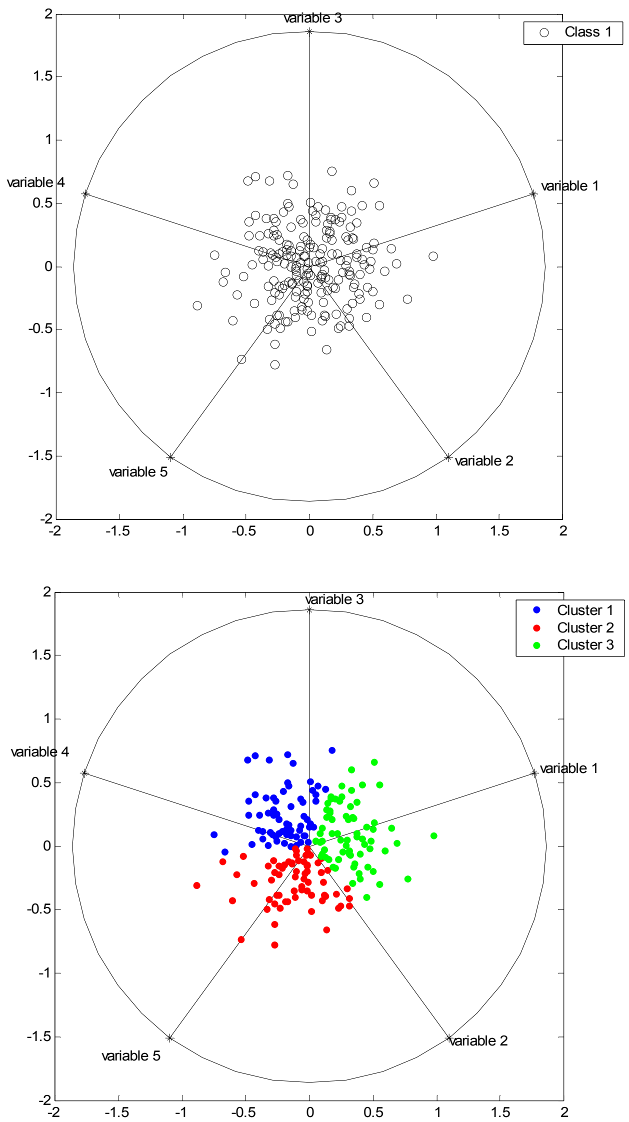


Figure 68. Star Coordinates (Artificial 2 data). Up: known classes; Down: obtained clusters

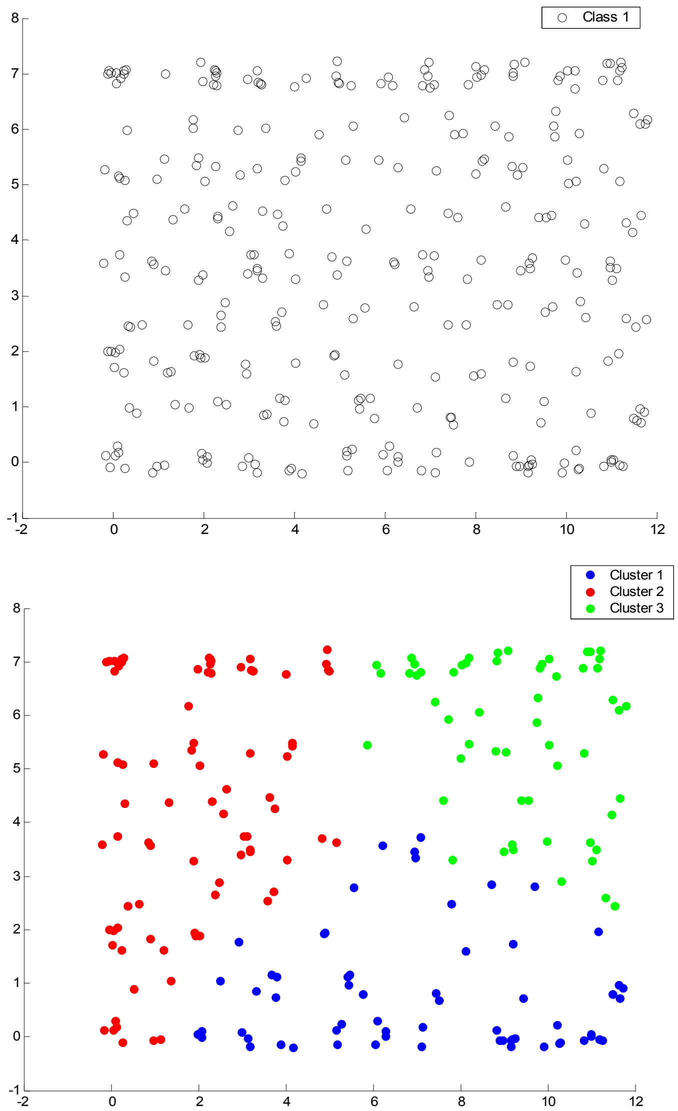


Figure 69. SOM (Artificial 2 data). Up: known classes; Down: obtained clusters

APPENDIX 4. Background information requested in the survey for the evaluation of multiple visualizations (Chapter 7)

Name
University
Major subject
Years at university
Experience with data analysis: Yes/ No

- If yes, what data analysis tools:

Experience with data visualization: Yes/ No

- If yes, what data visualization tools:

Have you used before the following tools for data visualization:[17]

Multiple line graphs …………………...….	Yes/ No
Permutation matrix …………………….......	Yes/ No
Survey plot …………………………...…	Yes/ No
Scatter plot matrix ………………………	Yes/ No
Parallel coordinates ……………………..	Yes/ No
Treemaps ……………………………..	Yes/ No
Principal Components Analysis plot ……	Yes/ No
Sammon's mapping ……………………	Yes/ No
Self-Organizing Maps ..…………………	Yes/ No

Have you attended the *Introduction to data visualization* lecture?[18] Yes/ No

[17] Question asked to participants in Groups 2 and 3.
[18] Question asked only to participants in Group 3.

APPENDIX 5. The percentages of invalid answers in each group (Chapter 7)

	G1	G2	G3	G4
Multiple Line Graphs				
outliers	0%	0%	4%	0%
relationships	0%	8%	11%	28%
clusters	0%	23%	11%	36%
cluster description	0%	0%	0%	0%
classes	0%	0%	0%	0%
class description	0%	15%	0%	4%
comparison	0%	15%	7%	12%
Permutation Matrix				
outliers	8%	8%	19%	0%
relationships	0%	15%	11%	4%
clusters	17%	0%	7%	20%
cluster description	0%	0%	0%	0%
classes	0%	8%	7%	8%
class description	0%	0%	*0%*	0%
comparison	0%	15%	7%	0%
Survey Plot				
outliers	0%	8%	4%	12%
relationships	0%	15%	15%	0%
clusters	8%	15%	7%	12%
cluster description	0%	0%	0%	0%
classes	0%	0%	4%	0%
class description	8%	8%	4%	4%
comparison	0%	8%	4%	0%
Scatter Plot Matrix				
outliers	0%	8%	0%	0%
relationships	0%	0%	15%	4%
clusters	8%	54%	52%	44%
cluster description	8%	8%	0%	0%
classes	0%	8%	4%	4%
class description	0%	0%	0%	0%
comparison	0%	8%	15%	4%

Parallel Coordinates				
outliers	0%	8%	0%	0%
relationships	25%	8%	33%	4%
clusters	8%	15%	30%	24%
cluster description	0%	0%	0%	0%
classes	0%	23%	4%	4%
class description	0%	0%	0%	0%
comparison	0%	8%	7%	0%
Treemap				
outliers	0%	0%	0%	0%
relationships	0%	0%	11%	0%
clusters	0%	38%	15%	16%
cluster description	0%	0%	0%	0%
classes	0%	0%	0%	0%
class description	0%	23%	4%	4%
comparison	0%	15%	4%	0%
PCA				
outliers	0%	8%	4%	4%
relationships	8%	38%	7%	4%
clusters	17%	31%	30%	8%
cluster description	0%	0%	11%	0%
classes	0%	8%	4%	0%
class description	0%	0%	0%	0%
comparison	8%	15%	22%	8%
Sammon's Mapping				
outliers	0%	8%	4%	4%
relationships	0%	8%	11%	0%
clusters	8%	31%	22%	12%
cluster description	0%	0%	0%	0%
classes	0%	0%	0%	0%
class description	58%	62%	*19%*	8%
comparison	0%	15%	15%	12%
SOM - Scatter Plot				
outliers	17%	0%	4%	4%
relationships	0%	0%	0%	0%
clusters	17%	23%	4%	4%
cluster description	0%	0%	0%	0%

classes	0%	0%	0%	0%
class description	0%	15%	0%	0%
comparison	8%	23%	15%	8%
SOM - U-matrix				
outliers	0%	15%	22%	4%
relationships	8%	8%	7%	0%
clusters	0%	15%	11%	4%
cluster description	8%	8%	0%	0%
classes	0%	0%	4%	0%
class description	0%	0%	0%	0%
comparison	17%	31%	22%	8%
SOM - Clustering				
outliers	0%	0%	7%	4%
relationships	0%	8%	7%	4%
clusters	0%	8%	4%	8%
cluster description	8%	8%	4%	0%
classes	0%	23%	0%	4%
class description	0%	0%	0%	0%
comparison	8%	23%	19%	12%
SOM - Feature Planes				
outliers	8%	0%	11%	16%
relationships	0%	31%	15%	0%
clusters	0%	8%	11%	16%
cluster description	8%	8%	4%	0%
classes	0%	8%	4%	0%
class description	0%	0%	0%	0%
comparison	0%	15%	11%	4%
SOM - All Views				
outliers	10%	8%	4%	12%
relationships	0%	31%	7%	4%
clusters	0%	23%	11%	28%
cluster description	0%	0%	0%	0%
classes	0%	8%	0%	0%
class description	0%	8%	4%	0%
comparison	0%	15%	15%	4%

APPENDIX 6. The modified version of the questionnaire (Chapter 7)

Questions - Line graphs:

1. Can you identify any **outliers** in this dataset by examining the above line graphs? – YES/ NO
    - If YES, please mark the outliers on the graphic.
2. Can you identify any **relationships** (correlations) between the financial ratios in this dataset by examining the above line graphs? – YES/ NO
    - If YES, please name one pair of financial ratios that you identified as correlated.
3. Can you identify **clusters** of companies with similar financial performance in this dataset by examining the above line graphs? – YES/ NO
    - If YES, please tell how many clusters you identified and circle each cluster on the graphic.
4. Can you **describe the clusters** that you have identified, by examining the above line graphs? – YES/ NO
    - If YES, please **specify one of the clusters** and **describe** it briefly in terms of values of the financial ratios in the table. Use Low, Medium, and High for indicating the prevailing level of the ratios in that cluster.

Operating margin	
ROE	
ROTA	
Equity to capital	

5. Can you distinguish between the companies from one region or another? – YES /NO
    - If YES, can you describe the financial characteristics of companies from Japan in comparison with the other regions by examining the above line graphs? Please describe them below. If you cannot describe, write “no”.
6. Can you **compare** the financial characteristics of the companies Reno de Medici 1997 (A), Reno de Medici 1998 (B), Buckeye Technologies 1998 (C) and Donohue 1998 (D) by examining the above line graphs? – YES /NO
    - If YES, please tell in brief how Reno de Medici 1998 (B) performs in comparison with Buckeye Technologies 1998 (C) and Donohue 1998 (D). Also tell how financial performance of Reno de Medici changed from 1997 (A) to 1998 (B).

APPENDIX 7. SOM evaluation questionnaire (Chapter 8)

BACKGROUND INFORMATION:
University:
Faculty/Department:
Major of your studies:
Years at University:
Programming experience:
    Yes/No
    Which programming languages:
Data analysis/Data visualization experience:
    Yes/No:
    Which tools:
For the assignment you have used mostly:
    SOM_PAK and Nenet ❒
    Matlab and Nenet ❒
    Matlab ❒

1. The requirements on the input data format (text files**) when applying SOM tools are easy to meet.**

1 Fully agree | 2 Partially agree | 3 Neutral | 4 Partially disagree | 5 Fully disagree

2. The normalized input data used by the SOM are adequate for clustering and visualization.

1 Fully agree | 2 Partially agree | 3 Neutral | 4 Partially disagree | 5 Fully disagree

3. The parameters of the SOM (*e.g. learning rate, size of the map, etc***) are:**

a) Easy to understand.

1 Fully agree | 2 Partially agree | 3 Neutral | 4 Partially disagree | 5 Fully disagree

b) Easy to use.

1 Fully agree | 2 Partially agree | 3 Neutral | 4 Partially disagree | 5 Fully disagree

4. When examining the resulting maps, your attention is focused on:

a) Attribute values, data samples, similarities, dissimilarities, and trends.

1 Very much | 2 Much | 3 Medium | 4 Little | 5 Very little

b) Colours, grids, borders, map size

1 Very much | 2 Much | 3 Medium | 4 Little | 5 Very little

c) Computational issues (statistics/ methodology/ formula used to create the maps)

1 Very much | 2 Much | 3 Medium | 4 Little | 5 Very little

5. The size of the graphic makes the reading of the information displayed easy.

1 Fully agree | 2 Partially agree | 3 Neutral | 4 Partially disagree | 5 Fully disagree

6. The maps obtained help you to see and analyse comparable data (*e.g. different objects/samples/companies on the map***).**

1 Fully agree | 2 Partially agree | 3 Neutral | 4 Partially disagree | 5 Fully disagree

7. The maps obtained help you to see and analyse <u>data trends</u> (*e.g. the same objects in different periods of time***).**

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

8. The maps obtained help you to see and analyse <u>correlations between attributes.</u>

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

9. The maps obtained help you to see and analyse <u>data clusters</u> (*e.g. groups of similar entities***)**

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

10. The maps obtained help you to see and analyse <u>attribute values</u>

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

11. When you look at the map you see <u>variation in the data</u> (*i.e. differences in the data are displayed differently***).**

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

12. The maps you obtain using SOM tools support your needs for analysis of the data.

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

13. The way the information is presented by the SOM is clear and understandable.

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

14. There are too many steps required to get a good map.

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

15. How <u>easy</u> is it to integrate the obtained results within other applications for presentation/communication purposes (*e.g. Power Point, MS Word, Paint, etc***)?**

1 Very easy	2 Easy	3 Medium	4 Difficult	5 Very difficult

16. <u>From your experience, given the reliability, completeness and correctness of the input information used, how would you characterize the *output information* provided by SOM?</u>

a) Reliable

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

b) Complete

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

c) Correct

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

d) Precise

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

17. <u>Please rate the following *aspects of visualization* with Self Organizing Maps</u>

a) Description of data (*i.e. how the data is displayed***)**

1 Very good	2 Good	3 Medium	4 Poor	5 Very poor

b) Exploration of data (*i.e. the possibility to see data at different levels of detail, e.g. feature planes, zoom in, zoom out, focus+context***)**

1 Very good	2 Good	3 Medium	4 Poor	5 Very poor

c) Tabulation of data (*i.e. how the data is presented in tables***)**

1 Very good	2 Good	3 Medium	4 Poor	5 Very poor

d) Decoration of data (*i.e. how the data is made attractive, decorated*)

1 Very good	2 Good	3 Medium	4 Poor	5 Very poor

e) Labelling of data (*i.e. words and explanations assigned to the data to make the reading easier*)

1 Very good	2 Good	3 Medium	4 Poor	5 Very poor

f) Dimensionality of data (*i.e. the number of variables used*)

1 Very good	2 Good	3 Medium	4 Poor	5 Very poor

18. In your opinion, applying SOM in data analysis is *easy* to do for

a) experts (*they have good domain knowledge and high experience of data analysis*)

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

b) students (*in class assignments*)

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

c) end/business users (*they have domain knowledge but not high experience in data analysis*)

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

19. The following elements of graphical design help you to interpret the map:

a) Colours

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

b) Scales (colour bars)

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

c) Grids, map units (neurons), borders

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

d) Attribute values

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

e) Data labels

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

20. How adequately are the following graphic elements represented on the maps:

a) Colours

1 Very good	2 Good	3 Medium	4 Poor	5 Very poor

b) Scales (colour bars)

1 Very good	2 Good	3 Medium	4 Poor	5 Very poor

c) Grids, map units (neurons), borders

1 Very good	2 Good	3 Medium	4 Poor	5 Very poor

d) Attribute values

1 Very good	2 Good	3 Medium	4 Poor	5 Very poor

e) Data labels

1 Very good	2 Good	3 Medium	4 Poor	5 Very poor

21. **When using SOM tools for data analysis and visualization, <u>to what extent are you satisfied with the content of the information displayed</u>?**

1 Very high	2 High	3 Neutral	4 Low	5 Very low

22. **When using SOM tools for data analysis and visualization, <u>to what extent are you satisfied with the correctness of the *information* displayed</u>?**

1 Very high	2 High	3 Neutral	4 Low	5 Very low

23. **When using SOM tools for data analysis and visualization, <u>to what extent are you satisfied with the usefulness of the *information* displayed</u>?**

1 Very high	2 High	3 Neutral	4 Low	5 Very low

24. **When using SOM tools for data analysis and visualization, <u>to what extent are you satisfied with the accuracy of the *system*</u>?**

1 Very high	2 High	3 Neutral	4 Low	5 Very low

25. **When using SOM tools for data analysis and visualization, <u>to what extent are you satisfied with the *time* needed to obtain good maps</u>?**

1 Very high	2 High	3 Neutral	4 Low	5 Very low

26. **When using SOM tools for data analysis and visualization, <u>to what extent are you satisfied with the learnability of the system</u>?**

1 Very high	2 High	3 Neutral	4 Low	5 Very low

27. **<u>From your experience, SOM as a tool for data analysis and data visualization:</u>**

a) **Is easy <u>to learn</u>.**

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

b) **Is easy <u>to use</u>.**

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

c) **Provides <u>accurate information</u>.**

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

d) **Provides <u>easy to interpret</u> visualizations (maps).**

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

e) **Provides <u>interesting information</u>.**

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

f) **Provides <u>new information</u>.**

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

28. **Did the SOM provide the information you needed for the assignment?**

1 Fully agree	2 Partially agree	3 Neutral	4 Partially disagree	5 Fully disagree

Part II. Original Publications

Paper 1

Marghescu, D. (2008). Usability Evaluation of Information Systems: A Review of Five International Standards, in Barry, C., Lang, M., Wojtkowski, W., Wojtkowski, G., Wrycza, S., & Zupancic, J. (eds) *The Inter-Networked World: ISD Theory, Practice, and Education,* Springer-Verlag: New York, (*Proceedings of the 16th International Conference on Information Systems Development (ISD 2007)*, Galway, Ireland, August 2007). In Press.

Usability Evaluation of Information Systems: A Review of Five International Standards

Dorina Marghescu

Turku Centre for Computer Science, Åbo Akademi University, Department of IT
dmarghes@abo.fi

Abstract. In this paper, we review five international standards that are concerned with defining and evaluating usability of information technology and interactive systems. The aim is to investigate the extent to which the international standards provide guidelines for planning and conducting usability evaluation of information systems. We first compare the standards in order to uncover the differences and relationships between the guidelines provided. Then, based on the guidelines, we provide a framework that highlights the activities required for usability evaluation of information systems. In the end, we discuss the way in which the standards cover the usability evaluation of information systems from the perspective of users, technology, system, and phase in system's life cycle.

Keywords: usability evaluation, information systems, international standards, ISO/IEC 9126-1, ISO/IEC 14598-1, ISO 9241-11, ISO 13407, ISO 18529.

1 Introduction

Evaluation of information systems (IS) represents an important topic among practitioners and researchers of information systems development (ISD) field. The evaluation of an IS may regard different aspects of the system, for example, performance, cost-benefit analysis, user acceptability, usability, reliability, functionality, efficiency, job satisfaction, etc. In this paper, we will focus on usability evaluation (UE) of information systems.

UE is concerned with planning and conducting the measuring of the usability attributes of the user interface and identifying specific problems (Ivory and Hearst 2001). Dix, Finlay, Abowd, and Beale (1998) point out that UE should be done throughout the design life-cycle and planned as providing results that can be used for improving the design. There are many models of usability that define the usability attributes that have to be measured. For example, Nielsen (1993) highlights the following usability attributes: learnability, efficiency, memorability, error rate, and satisfaction. However, the literature provides several other attributes as well as methods, techniques and metrics that can be used for measurement (see, for example, Hornbæk 2006).

Though UE is extensively discussed in literature, especially from the perspective of usability models, methods, techniques, and metrics used in evaluation, in the ISD methodologies, evaluation, in general, and usability evaluation, in particular, are not well addressed. Comparing different ISD methodologies, Avison and Fitzgerald (2003) identified only a small number of the existing methodologies that address in detail the system evaluation at the post implementation stage. The other methodologies mention the post-implementation evaluation as being important or do not address evaluation at all. An exception is Merise methodology that considers the usability and acceptance criteria in the decision cycle, when users together with system developers and senior managers are involved in discussing various options for the technical part (i.e., software and hardware) of the information system. However, Merise does not address a post-implementation evaluation.

Motivated by this lack of details that Avison and Fitzgerald (2003) pointed out regarding the evaluation of IS in the ISD methodologies, we address in this paper one important aspect of IS evaluation, namely usability evaluation. We review five standards, developed by the International Organisation for Standardization (ISO), that address usability of information technology (IT) and interactive systems. The standards that we review are: ISO/IEC 9126 – 1, ISO/IEC 14598 – 1, ISO 9241 – 11, ISO 13407, and ISO 18529. We study these standards because they are intended to provide guidelines and general principles for planning and conducting evaluation during product/system development life-cycle. We do not analyse the standards that guide the assessment of the UE process capability (e.g., ISO/IEC 15504 – 1).

Other studies concerned with analysing the international standards that address usability are, for example, Jokela, Iivari, Matero and Karukka (2003) and Seffah, Donyaee, Kline and Padda (2006). In an interpretative study, Jokela et al. analysed ISO 13407 against ISO 9241-11 with the aim to find whether the two standards are consistent. Seffah et al. reviewed three ISO standards (ISO 9241-11, ISO/IEC 9126 – Parts 1 and 4, and ISO/IEC 14598-1) and other usability models from literature. Their aim was to highlight the need for a unified theory of usability measurement and propose a model for usability measurement. Nielsen (1993) describe and compare the advantages of using interface standards (national, international, industry or in-house built standards) when designing interactive systems.

Our aim is to identify how the ISO standards address the UE *process* of IS. We do not focus on the measurement component of the evaluation process (that is, on methods, usability models and metrics used in evaluation), but on the phases and activities that are required in planning and conducting UE. The research question is to what extent the standards provide guidance for planning and conducting UE of information systems. We study the guidelines provided by the standards with respect to planning and conducting UE of information systems. We first compare the standards to uncover differences and relationships between the guidelines provided. Then, based on the guidelines, we provide a framework that highlights the activities required for UE of IS.

The definition of an information system in Avison and Fitzgerald (2003) emphasizes that people and technology are equally important for the success of an information system: users and IT interact in order to accomplish a function, goal or task in a specified environment. Therefore the evaluation of the information system has to take into account the users, the technology (software), and the systems as whole. Having in mind this definition of an information system, we try also to examine the way in which the standards cover the UE from the perspective of the users, technology, and the system as a whole. Moreover, the system has to be evaluated throughout the system's life-cycle, and we will analyse whether the standards provide guidelines for UE in different stages of system's life-cycle (requirements, design, implementation, use, and maintenance).

The paper is organised as follows. Section 2 introduces the five ISO standards under analysis. Section 3 compares the scopes of the standards. Section 4 presents the definitions of usability encountered in international standards. Section 5 compares the usability evaluation approaches provided by the international standards. Section 6 discusses the comparative analysis of international standards and provides a framework for usability evaluation of information systems based on the standards' guidelines. It also discusses the way in which the standards cover the usability evaluation from the perspective of the users, technology, the system as a whole, and the system's life cycle. Section 7 concludes the paper.

2 International Standards that Address Usability

International standards provide practitioners with a common technical language necessary in development, acquisition, supply and evaluation of products and services and in communicating to other parties. They are also a means to ensure that the final product attains the desired quality. There are five ISO standards that address usability of information technology and interactive systems:

- ISO/IEC 9126 – Part 1 (2000) - Information Technology – Software product quality – Part 1: Quality model,
- ISO/IEC 14598 – Part 1 (1999) - Information Technology – Software product evaluation – Part 1: General overview,
- ISO 9241 – Part 11 (1998) - Ergonomic requirements for office work with visual display terminals. Part 11: Guidance on usability,
- ISO 13407 (1999) - Human-centred design processes for interactive systems,
- ISO 18529 (2000) - Ergonomics – Ergonomics of human-system interaction – Human-centred life-cycle process descriptions.

ISO/IEC 9126 defines a software product quality model, characteristics and metrics and Part 1 of this standard focuses on defining the quality model. ISO/IEC 14598 describes the process of evaluation of software product quality and Part 1 of this standard provides a general overview of this process. A new series, namely ISO/IEC 25000 (2005) – Software Engineering – Software product quality requirements and evaluation (SQuaRE) – Guide to SQuaRE, is intended to replace and integrate the multipart standards ISO/IEC 9126 and 14598. However, only ISO/IEC 25000 from the new series has been released and published while other parts that detail the quality model and process of evaluation are currently under development.

ISO/IEC 25000 introduces the other parts of SQuaRE and provides an overview of software quality modelling and evaluation as well as the relationships between current standards (ISO/IEC 9126 and ISO/IEC 14598) and the new SQuaRE series of standards. Until the other parts of the new series SQuaRE are published, the multipart ISO/IEC 9126 and ISO/IEC 14598 are applicable and therefore they are appropriate for analysis.

ISO 9241 – Part 11 addresses the definition and evaluation of usability in the context of visual display terminals or other products with which a user interacts in order to achieve a goal. ISO 13407 addresses the process of human-centred design for interactive systems and uses the definition of usability from ISO 9241 – 11. Finally, ISO 18529 defines a model of the human-centred processes described in ISO 13407. It is intended to be used in the specification, assessment and improvement of the human-centred processes in system development and operation.

To uncover the relationships and differences between the above standards with respect to usability evaluation, we analyse comparatively the standards focusing on:

- Scope of the standard,
- Definition of usability adopted by a standard, and
- The evaluation approach proposed by the standard.

3 Scopes of International Standards

The analysis of the scope of the standards aims at identifying three elements: 1) the entity under evaluation (e.g., software product, hardware, and system), 2) the main stakeholders to whom the standard is addressed, and 3) the phase in life-cycle at which the product or system is evaluated (e.g., during development, use, acquisition, etc.). Table 1 presents the scope of each standard.

Table 1. Scopes of ISO/IEC 9126-1, ISO/IEC 14598-1, ISO 9241-11, ISO 13407, ISO 18529

Standard	*Entity*	*Stakeholders*	*Phase in life-cycle*
ISO/IEC 9126-1 ISO/IEC 14598-1	Software product	Designers, developers, evaluators, maintainers, acquirers	Requirements, development, use, evaluation, support, maintenance, quality assurance, audit of software, acquisition
ISO 9241-11	Software, hardware or service product in interactive systems	Designer, developer , evaluator, acquirer	Design, development, evaluation, procurement
ISO 13407	Computer-based interactive system	Project managers, All parties involved in human-centred system development	Throughout the system development life-cycle
ISO 18529	Life-cycle process of computer-based interactive system, software and hardware	Those involved in design, use and assessment of life-cycle processes	Design, development, use and assessment of life-cycle process of system, software and hardware

By comparing the standards with respect to their scopes, the following observations can be made:

- ISO/IEC 9126-1 and ISO/IEC 14598-1 (ISO/IEC 25000, respectively) focus on defining and evaluating *quality* of any kind of *software products* (including computer programs, data contained in firmware). They are, on the one hand, narrower in scope than the other standards, because they focus on the software product; while the others take into account both hardware and software of an interactive system. On the other hand, they are broader in scope than the other standards because they cover all quality characteristics of the software product, not only usability.
- ISO 9241-11 focuses on defining and evaluating *usability* of any *product* that is part of an *interactive system* and can be of nature software, hardware or service (e.g., Visual display terminals).
- ISO/IEC 9126-1, ISO/IEC 14598-1 and ISO 9241-11 are addressed to *stakeholders* involved in all phases of product/system life-cycle: development, procurement or evaluation of software products and interactive systems, respectively. ISO/IEC 9126-1 and ISO/IEC 14598-1 are addressed to maintainers too.
- ISO 13407 focuses on *designing* computer-based interactive systems and evaluating different *design solutions throughout the system development life-cycle* and is addressed mainly to *project managers*, but to other stakeholders involved in system's development life-cycle too.
- ISO 18529, unlike the other standards, addresses the system's life-cycle *process* (how this process should be designed, used and assessed). The others standards address the software, hardware or systems (how these should be designed, implemented, used, and assessed). Therefore, it focuses not on developing a system, hardware or software, but on modelling their *life-cycle processes*.

4 Definition of Usability in International Standards

ISO/IEC 9126-1 (ISO/IEC 25000, respectively) represents the Software Engineering (SE) perspective on usability. In SE, usability is defined as being "the capability of the software product to be *understood*, *learned*, *used* and *attractive* to the user, when used under specified conditions."

ISO 9241-11 represents the Human Computer Interaction (HCI) perspective on usability that is defined as being "the extent to which a product can be used by specified users to achieve specified goals with *effectiveness*, *efficiency* and *satisfaction* in a specified context of use."

The differences in defining the usability can be explained by the focus that each community has. While SE is concerned with providing high quality intermediate or final *software products* that conform to specified requirements, HCI is concerned with developing usable *interactive systems*. Therefore, in SE usability is just one component of software product quality, while in HCI reaching a high level of usability is the ultimate goal of the system development. Another difference is that SE focuses on *software product* development and evaluation, while HCI focuses on *system* (interaction between software, hardware and users) development and evaluation.

To include the latter perspective, ISO/IEC 9126-1 (ISO/IEC 25000) added a new concept, namely, *quality in use,* which is defined as being "the capability of the software product to enable specified users to achieve specified goals with effectiveness, productivity, safety and satisfaction in specified contexts of use." In this definition, the attribute safety is added and efficiency is named productivity.

According to these definitions, usability is not an intrinsic quality of a product, but its measured level depends on many factors, which are generically described as the *context of use*. The context of use may include the users and their goals, the tasks, the equipment, and the environment. Both ISO 13407 and ISO 18529 use the definition of usability given in ISO 9241-11.

Table 2 summarizes the definitions of usability adopted by each standard. Regardless which definition is used, assessing the level of usability achieved by an information technology or information system requires careful planning of the measurement. Usability evaluation represents the methodology used for planning and conducting usability measurement. In the next section, we present different approaches to usability evaluation as provided by the international standards.

Table 2. Definitions of usability in international standards

Standard	*Definition of usability*
ISO/IEC 9126-1 (ISO/IEC 25000)	**Usability** is the capability of the software product to be *understood*, *learned*, *used* and *attractive* to the user, when used under specified conditions
	Quality in use is the capability of the software product to enable specified users to achieve specified goals with *effectiveness, productivity, safety* and *satisfaction* in specified contexts of use.
ISO 9241-11	**Usability** is the extent to which a product can be used by specified users to achieve specified goals with *effectiveness*, *efficiency* and *satisfaction* in a specified context of use.

5 Usability Evaluation Approaches in International Standards

Here we compare the approaches to UE provided by different standards and examine the level of detail of the guidelines. There are four standards that address UE of information technology and interactive systems. These standards are: ISO/IEC 14598-1, ISO 9241-11, ISO 13407 and ISO 18529. Table 3 presents the approaches to UE provided by the four standards.

By comparing the four approaches, the following observations can be made:

- ISO/IEC 14598-1 provides a general process of software product quality evaluation. The phases and activities of the process are well defined. The guidance allows identification of inputs and outputs of each activity. The evaluation is based on the quality model specified in the first phase of the process. The evaluation process can be employed for any intermediate or final product throughout its lifecycle.

- ISO 9241-11 distinguishes between evaluation of a product during design and evaluation of a system during its use, but the evaluation process is not discussed in detail.
- ISO 13407 addresses mainly evaluation of design, but also the long-term monitoring of the system in use in the context of human-centred design process. The guidance provided is a framework that identifies different situations and objectives that usability evaluation can address.
- ISO 18529 describes the evaluation of the design against requirements as an iterative sub-process within the general human-centred design process. The evaluation is carried out in all phases of the human-centred design process from requirements to use. The standard provides guidance with regard to the activities to be performed in different stages of the evaluation process. However, the standard does not detail how these activities have to be conducted.

Table 3. Evaluation approaches in ISO/IEC 14598-1, ISO 9241-11, ISO 13407 & ISO 18529

Standard	*Evaluation approach*
ISO/IEC 14598-1 (ISO/IEC 25000)	Four phases evaluation process, applicable throughout the software product life-cycle 1. Quality requirements definition 2. Specify the evaluation 3. Plan the evaluation 4. Execute the evaluation
ISO 9241-11	Distinction between specification and evaluation of usability during design and during use of system: 1. During design: usability input to a quality plan. 2. During use: evaluate how changes in components of context of use affect efficiency, effectiveness, and/or satisfaction of the user
ISO13407	- Three phases prior to evaluation (human-centred design approach) 1. Identification of the context of use and its components represents a basis for evaluation. 2. Specification of user and organizational requirements should be done in terms that permit subsequent testing and should be confirmed or updated during the life of the project. (Reference to ISO/IEC 14598-1 for specifying software) 3. Produce design solutions 4. Evaluate designs against requirements - Identifies and addresses three goals of evaluation (providing feedback to improve the design, assess whether the objectives have been met, and monitor long term use of the product). - Discusses evaluation plan and different evaluation situations
ISO 18529	- Distinction between formative evaluation (identification of problems) and summative evaluation (assessment of whether objectives are met). - Discusses the following evaluation activities: 1. Specify context of evaluation 2. Evaluate for requirements 3. Evaluate to improve design 4. Evaluate against system requirements 5. Evaluate against required practice 6. Evaluate in use.

The standard that provides the most extensive guidance on how to proceed with the quality evaluation of information technology is ISO/IEC 14598-1. We summarize in Table 4 the phases, activities, and outputs of the UE process as derived from the general model of quality evaluation given in ISO/IEC 14598-1. Accordingly, the UE process usability consists of four phases: (1) quality requirements definition, (2) specify the evaluation, (3) plan the evaluation, and (4) execute the evaluation.

Table 4. Process model for usability evaluation (derived from ISO/IEC 14598-1)

Phase	*Activity*	*Output/documents*
1. Define usability requirements	1.1 Establish purpose of evaluation	Purpose of evaluation document
	1.2 Identify type of product	Specification of product and context of use
	1.3 Specify usability model	Usability characteristics, sub-characteristics, attributes
2. Specify the evaluation	2.1 Select metrics for each attribute	Metrics
	2.2 Establish rating levels for each metric	Rating levels
	2.3 Establish assessment criteria	- Assessment criteria - Procedure to summarize evaluation results
3. Plan the evaluation	3.1 Produce evaluation plan	Specification of evaluation method, schedule and evaluator actions
4. Execute the evaluation	4.1 Take measures	Measured values for each metric
	4.2 Compare with criteria	Rated level for each metric
	4.3 Assess results	- Statement of usability - Managerial decision (acceptance or rejection of product)

In the first phase, quality requirements definition, the context of use and the usability attributes are identified. In the second phase, suitable metrics for each attribute are identified. Moreover, rating levels for the metrics as well as assessment criteria are defined. In the third phase, an evaluation plan is produced in order to describe the evaluation method to be used, the schedule of the evaluation and the evaluator actions. The evaluation phase consists of a series of activities such as *measurement*, *rating*, and *assessment* of the usability attributes. The *measurement* means applying usability or quality in use metrics to the system or component under evaluation. For conducting this activity, a usability evaluation method is defined in order to properly collect, analyse and summarise the measurement data. The *rating* means that the measured value is mapped to a rating level, a priori established. Finally, the *assessment* requires applying assessment criteria to the system or component under analysis for determining the acceptance of the system in terms of usability.

The other standards, ISO 9241-11, ISO 13407, and ISO 18529 do not address in detail the steps of UE. ISO 9241-11 focuses on specifying the context of use. ISO 13407 highlights the phases needed for UE and mentions the use of appropriate evaluation methods (ISO 13407). ISO 18529 provides general guidelines applicable to different evaluation purposes.

6 A Framework for UE of IS

The comparative review of the international standards has yielded that these standards complement each other as follows. First, ISO/IEC 9126-1 and ISO/IEC 14598-1 focus on software products, while the other standards take into account also the hardware and interactive system as a whole. Moreover, ISO/IEC 9126-1 and ISO/IEC 14598-1 are addressed to the stakeholders responsible with all phases in software product development life-cycle, including maintenance. The other standards are not addressed to maintainers. ISO 18529 focus on modelling the life-cycle process, and not on software, hardware or interactive system. In addition to these differences and relationships observed by analysing the scopes of the standards, other differences between the standards regard the way they define usability and approach usability evaluation (Table 2 and Table 3, respectively).

Regarding the approach to usability evaluation recommended by a standard, we observed that the standard that provides the most extensive guidance on how to proceed with the evaluation is ISO/IEC 14598-1. This standard provides the phases and activities required in the process of quality evaluation of a software product, but the guidelines can be extended to usability evaluation of information systems (Table 4).

Taking into account also the guidelines provided by the other standards, we propose a framework for UE of IS. The framework lists the principal activities in the UE process (Table 5), as derived from the guidelines provided by all standards (Tables 3 and 4). The framework also provides references to the standards that mention and/or guide the activities.

Table 5. Activities in UE process

Activities	*Standards*
1. Distinguish between system under development and system in use	ISO 9241-11, ISO 18529
2. Specify purpose of evaluation and evaluation target *(users, technology, or system and phase in product/system development life-cycle)*	ISO/IEC 9126-1, ISO/IEC 14598-1, ISO 9241-11, ISO 13407 and ISO 18529
3. Specify context of use of the information system	ISO 9241-11, ISO 13407 and ISO 18529
4. Specify usability and quality in use characteristics, sub-characteristics, attributes	ISO/IEC 9126-1, ISO/IEC 14598-1
5. Select or create validated metrics to be used in measurement of the system usability attributes	ISO/IEC 14598-1, ISO 9241-11
6. Specify rating levels for each metric	ISO/IEC 14598-1
7. Specify assessment criteria	ISO/IEC 14598-1
8. Select and specify an appropriate usability evaluation method	ISO 13407, ISO 18529, ISO/IEC 14598-1
9. Measure usability attributes	ISO/IEC 14598-1, ISO 9241-11
10. Map measured values to rating levels	ISO/IEC 14598-1, ISO 9241-11
11. Assess result	ISO/IEC 14598-1, ISO 9241-11

Activities 1, 2, 3 and 4 belong to the Define usability requirements phase. Activities 5, 6 and 7 represent the Specify evaluation phase. Activity 8 represents the Plan evaluation phase. Activities 9, 10 and 11 belong to the Execute evaluation phase.

Table 5 points out that at activity 2 the target of the evaluation has to be specified and that all five standards mention or provide guidelines for this activity. However, the extent to which the standards guide the evaluation from each of these perspectives (i.e., users, technology, systems, and phase in life-cycle) differs as follows.

ISO/IEC 9126-1 focuses on software product evaluation, but addresses the evaluation from all three perspectives (users, technology and system). It focuses on providing a detailed model of usability and quality in use characteristics and sub-characteristics. ISO/IEC 14598-1 focuses on a general evaluation process and activities that can be applied to evaluate attributes that are relevant to users, technology, and system. ISO 9241-11 and ISO 13407 focus on interactive system evaluation. They especially guide the specification of the context of use. In addition, ISO 13407 and ISO 18529 discuss the involvement of the user during the development life-cycle process. ISO 18529 does not address the user or technology evaluation but lists the activities required in the system evaluation in the context of human-centred design approach (Table 3).

Table 6 summarizes the aspects for which guidelines are provided in each standard with respect to user, technology or system.

Table 6. Guidance on usability evaluation with respect to user, technology or system

	User	*Technology (software product)*	*System*
ISO/IEC 9126-1	- User requirements - User satisfaction	- Usability attributes	- Quality in use attributes
ISO/IEC 14598-1	- Activities for UE of attributes relevant to user	- Activities for UE of attributes relevant to technology	- Activities for UE of attributes relevant to system
ISO 9241-11	- Context of use (identification of users)	- Context of use (identification of equipment and tasks)	- Context of use (identification of users, equipment, tasks, and environment) - Activities for UE
ISO 13407	- User involvement in design process, including evaluation - User requirements specification	- Activities for UE	- Context of use (users, tasks, environment) - Activities for UE
ISO 18529	- Focus on user throughout the system development life-cycle process, including system strategy and evaluation	-	- Activities in the evaluation process during system development life-cycle process

Table 7 provides an assessment of the level of detail at which each standard addresses the evaluation process at different phases of systems' life-cycle. The difference between "extensive guidance" and "guidance at some detail" is that the former includes guidelines on how to perform different activities in the evaluation process, while the latter limits to enumerating the activities and distinguishing between different evaluation situations. ISO/IEC 9126-1 provides a model of usability and quality in use characteristics that can be used in any phase of life-cycle. However, it does not address the overall UE process and, therefore, we did not include it in the table.

Table 7. Guidance on usability evaluation during the system life-cycle

	Requirements	*Design*	*Implementation*	*Use*	*Maintenance*
ISO/IEC 14598-1	+++	+++	+++	+++	+
ISO 9241-11	+++	++	++	++	-
ISO13407	++	++	-	++	-
ISO 18529	++	++	++	++	-
Legend: +++: extensive guidance, ++: guidance at some detail, +: mentioning the situation , -: no guidance					

7 Conclusion

In this paper, we analysed comparatively five international standards that address the definition and evaluation of usability of information technology (software) and interactive systems: ISO/IEC 9126-1, ISO/IEC 14598-1, ISO 9241-11, ISO 13407 and ISO 18529. The purpose of the study was to investigate the extent to which the standards provide guidance for planning and conducting usability evaluation of information systems.

We first analysed the standards to uncover relationships and differences between the guidelines. The standards provide different approaches on defining and evaluating usability (Table 2 and Table 3, respectively) and provide different levels of detail and coverage of the aspects considered (Table 1, Table 6, and Table 7). Based on the standards' approaches to usability evaluation, we provided a framework for usability evaluation of information systems that highlights the main activities of usability evaluation process (Table 5). We also discussed the extent to which the guidelines for conducting usability evaluation take into account the user, technology and system (Table 6), and different phases in system life-cycle (Table 7).

The analysis of the standards showed that the guidelines complement each other as follows. ISO/IEC 9126-1 and ISO/IEC 14598-1 focus on software products, while the other standards address also the hardware and interactive system as a whole. Moreover, ISO/IEC 9126-1 and ISO/IEC 14598-1 are addressed to the stakeholders responsible with all phases in software product development life-cycle, including maintenance. The other standards are not addressed to maintainers. ISO 18529 focus on modelling the life-cycle process, and not on software, hardware or interactive system. Regarding the usability evaluation approaches provided by the standards, ISO/IEC 14598-1 provides the most detailed guidelines to usability evaluation process in almost all phases of product and system life-cycle (Tables 3, 4 and 7), but does not discuss in detail the identification of context of use of the product/system. ISO 9241-11 focuses on describing the context of use (Table 6), but does not discuss in detail the phases, activities, inputs and outputs required in the usability evaluation process. ISO 13407 and ISO 18529 focus on highlighting the importance of user involvement throughout the system development life-cycle process in the context of human-centred design approach, and limit themselves to enumerating the activities of this process, but do not discuss them in detail. ISO/IEC 9126-1 provides extensive models of usability and quality in use.

Acknowledgements

I gratefully acknowledge the contribution of Professor Barbro Back and the financial support of Tekes (grant number 40435/05) and the Academy of Finland (grant number 104639).

References

Avison, D. E. and Fitzgerald, G. (2003) *Information Systems Development: Methodologies, Techniques and Tools*. 3rd Edition. McGraw-Hill, London.
Dix. A., Finlay, J., Abowd, G., and Beale, R. (1998) *Human Computer Interaction*, 2nd Edition, Prentice Hall, Inc.
Hornbæk, K. (2006) Current practice in measuring usability: Challenges to usability studies and research. *Int. J. Human-Computer Studies*. 64 (2), Feb., 79-102.
ISO 9241-11 (1998) Ergonomic requirements for office work with visual display terminals. Part 11: Guidance on usability.
ISO 13407 (1999) Human-centred design processes for interactive systems.
ISO/IEC 14598-1 (1999) Information Technology – Software product evaluation – Part 1: General overview.
ISO/IEC 9126-1 (2000) Information Technology – Software product quality – Part 1: Quality model.
ISO/IEC 9126-4 (2004) Software Engineering – Product quality – Part 4: Quality in use metrics.
ISO/IEC 25000 (2005) Software Engineering – Software product quality requirements and evaluation (SQuaRE) – Guide to SQuaRE.
ISO 18529 (2000) Ergonomics – Ergonomics of human-system interaction – Human-centred lifecycle process descriptions.
ISO/IEC 15504 -1 (2004) Information technology - Process assessment - Part 1: Concepts and vocabulary.
Ivory, M. Y. & Hearst, M. A. (2001) The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys*. 33 (4), 470-516.
Jokela, T., Iivari, N., Matero, J., and Karukka, M. (2003) The standard of user-centered design and the standard definition of usability: analyzing ISO 13407 against ISO 9241-11. In *Proc. of the Latin American Conference on Human-Computer interaction* (Rio de Janeiro, Brazil, August 17 - 20, 2003), CLIHC '03, vol. 46. ACM Press, New York, NY, 53-60.
Nielsen, J. (1993) *Usability Engineering,* Morgan Kaufmann Publishers Inc.
Seffah, A., Donyaee, M., Kline, R. B., and Padda, H. K. (2006) Usability measurement and metrics: A consolidated model. *Software Quality Control* 14, 2 (Jun.), 159-178.

Paper 2

Marghescu, D. (2006). Evaluating the Effectiveness of Projection Techniques in Visual Data Mining. In *Proceedings of the 6th IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP 2006),* Palma de Mallorca, Spain, August 2006.

Proceedings of the Sixth IASTED International Conference
Visualization, Imaging, and Image Processing
August 28–30, 2006, Palma de Mallorca, Spain
ISBN Hardcopy: 0-88986-598-1 / CD: 0-88986-600-7

EVALUATING THE EFFECTIVENESS OF PROJECTION TECHNIQUES IN VISUAL DATA MINING

Dorina Marghescu
Turku Centre for Computer Science
Åbo Akademi University
Lemminkäisenkatu 14 A, 5th floor
FIN-20520 Turku, Finland
dorina.marghescu@abo.fi

ABSTRACT
When dealing with large amounts of high-dimensional data, one approach is to reduce the number of data dimensions by applying a projection technique. The later reduces the data dimensionality by combining the original variables into a smaller number of new dimensions, in a linear or nonlinear manner. The projection methods are particularly useful because they lend themselves to visual representations of data, when the number of new dimensions is one, two or three. In this paper, we compare different visualization techniques based on projection techniques with respect to their effectiveness for solving a data-mining task such as clustering. We investigate the use of cluster validity measures in order to judge the effectiveness of the projection techniques in visual data mining. The results show that cluster validity is a successful approach to evaluate objectively the visualization techniques.

KEY WORDS
Data treatment and visualization, evaluation, visual data mining, cluster validity

1. Introduction

Nowadays, business people, but also any knowledge workers face problems caused by the difficulty or even the impossibility to deal efficiently with the data available. Encountered in the research literature as the problem of *information overload* [1] and the problem of *information complexity* [2], these problems become even more acute when the users need to rely less on other people in order to get insight into the data, and they are working under time and performance constraints. One solution to these problems appears to be provided by the *visual data mining tools* or *visual analytics*.

According to Thomas and Cook [3], visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces. Knowledge workers (e.g., business users) use visual analytics (or visual data mining [4]) tools and techniques, in order to synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data. These tools are also effective in detecting the expected but also in discovering the unexpected patterns in the data.

One category of visualization techniques that are effective in visualizing high-dimensional datasets is based on *projection techniques* [5]. The high-dimensional data is mapped onto a lower-dimensional space, usually of two dimensions. The resulting dataset is plotted on a Cartesian coordinate system. In pattern recognition, this process of reducing the dimensionality of the dataset is referred to as *feature extraction* [6-8]. Feature extraction is usually performed in the context of pattern classification, and its goal is to find or derive the most important features of the dataset that are used further by a classifier. In information visualization, the transformation of a multidimensional dataset and its representation on a 2D space is referred to as *geometrically transformed displays* [4].

In this paper, we analyze the effectiveness of five projection techniques for visual data mining or visual analysis. In particular, we evaluate two classical projection techniques, that is, PCA [7, 9] and Sammon's mapping [10], and three more recently developed techniques, namely SOM [5], Radviz [11], and Star Coordinates [12]. One can use both qualitative and quantitative criteria, for evaluation.

The qualitative criteria are subjective (i.e., the user interprets the effectiveness of the visual representation, based on his own experience and goals). The quantitative criteria are objective measures of how well the projections preserve the original structure of the datasets. Both types of evaluation are important [13]. The evaluation of visual representation is an increasingly important research area, many researchers emphasizing the need to develop methodologies for evaluation, but relatively little has been done in this area [14]. Keim [13] and Grinstein [2] have illustrated the use of objective as well as subjective measures to evaluate the capabilities of a number of visualization techniques. In this paper, we propose and illustrate the use of *clustering validity measures* in order to assess *objectively* the effectiveness of the projection techniques in visual data mining.

The paper is organized as follows. In Section 2, we define the research problem and formulate the research questions. Section 3 describes briefly the projection techniques under analysis. Section 4 presents the method used for evaluation and the datasets used in our study. Section 5 presents the evaluation results. We conclude with final remarks and future work ideas in Section 6.

2. Research problem and questions

In this paper, we intend to compare different data representations (derived by applying the PCA, Sammon's mapping, SOM, Radviz, and Star Coordinates techniques) with respect to their effectiveness in preserving the structure of an original dataset. For this purpose, we study the possibility of using *clustering methodology* to evaluate different projection techniques. By employing the clustering methodology, we first perform partitioning of both the original and transformed data, and then we evaluate the obtained clustering solutions.

The clustering of data is an appropriate procedure for discovering internal relationships among data points, especially when these are not labelled. In principle, the relationships among data points that clustering techniques are evaluating are defined in terms of similarity or dissimilarity among data items. Therefore, clustering techniques are used in order to partition the dataset into groups of homogenous or similar data items. However, a clustering solution obtained by applying a clustering algorithm does not always reveals a *real* partition of the data. One can evaluate the truthfulness or accuracy of a clustering solution by using *cluster validity measures* [15-17]. Measuring cluster validity can be done via *external validity* and *internal validity*, but also by *relative criteria* [17].

External validity criteria are used to evaluate to what extent a clustering solution, obtained after applying a clustering algorithm on a dataset, matches an a priori known or assumed structure of the dataset. *Internal validity criteria* are used to assess a clustering solution, obtained by applying a clustering algorithm, in terms of internal relationships among the quantities represented by the data under study, for example, the proximity matrix. Finally, *relative criteria* are used to compare different clustering solutions obtained by using different clustering algorithms or different parameters of the same algorithm.

In our study, we intend to use only the ***external*** and ***internal validity measures***, and consider for future work the use of relative criteria. By using the external validity criteria, we want to determine the extent to which a clustering solution obtained on a "transformed" dataset confirms the known or assumed structure of the original dataset. By using the internal validity measures, we aim to determine whether the clustering solutions resulted are indeed real. It is a known fact that most clustering algorithms impose a clustering structure on a dataset, even though the dataset may not possess such a structure. Therefore, we need to assess the degree to which the clustering solution is recovered in the original data. In most of the cases, the structure of the data is not known a priori. In such situations, we cannot perform external validity assessment, but we can evaluate the internal validity of the obtained clustering solution.

The research questions that have guided our study are presented in the following: to what extent the new representation of the data obtained through projection techniques reveals the structure (clusters) of the original dataset? Can we use clustering validity criteria to evaluate the suitability of using a particular projection technique to represent our datasets?

To answer these questions, we performed a number of experiments on known datasets. All the plots and required computations were realized in Matlab [18]. For obtaining the SOM and Sammon projections, we also used the SOM Toolbox for Matlab [19].

3. Projection Techniques

Projection methods are used to reduce the variable space (i.e., the original variables are combined into a smaller number of new dimensions). The projection methods are particularly useful because they facilitate visual representations of data. In this section, we look at two classical techniques: Principal Components Analysis (PCA) and Sammon's mapping, and three more recently developed techniques: Self Organizing Maps (SOMs), Radviz and Star Coordinates. We briefly describe each technique and give to the interested readers the references where they can find the mathematical details behind each technique.

3.1. Illustrative dataset: Iris dataset

For illustrating the capabilities of each projection technique, we use the *Iris* dataset [20]. This dataset is a well known reference in pattern recognition community, due to its public availability and suitability for classification and clustering tasks. The data concerns three species of flowers characterised by four attributes: petal length and width, and sepal length and width. The class variable is the type of flower: Iris-Setosa, Iris-Versicolor, and Iris-Virginica. The dataset contains 150 observations, each class containing 50 flowers. The class Iris-Setosa is linearly separable from the other two classes, but Iris-Versicolor and Iris-Virginica classes are not linearly separable.

3.2. PCA

PCA is a classical statistical technique [9] that maps high-dimensional data items onto a lower-dimensional space. The PCA technique is known in the signal processing community as Karhunen-Loéve expansion [7]. The transformation tries to preserve the variance of the original data as well as possible. The PCA technique simply creates new variables (called principal components), which are linear composites of the original variables. The maximum number of new variables that can be formed is equal to the number of original variables, and the new variables are uncorrelated among themselves.

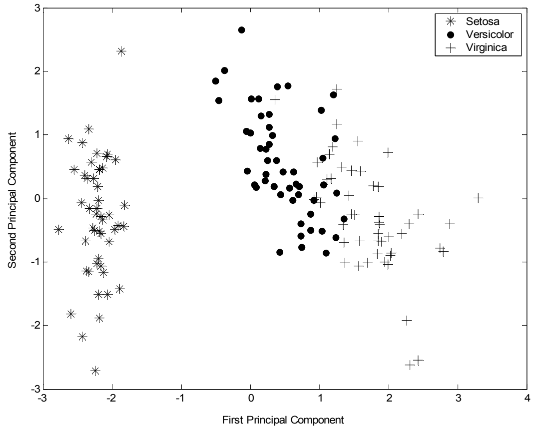


Figure 1. PCA of the Iris data

The solution to PCA is obtained by computing the eigenvalues and eigenvectors of the covariance matrix. The eigenvectors give the weights that are used for computing the new variables, whereas the eigenvalues give the variances of the new variables. The solution depends on the relative variances of the variables. For obtaining the PCs we have used the standardized data. Figure 1 represents the Iris dataset plotted on the 2D space formed by the first two principal components.

By choosing only the two first PCs to represent the data, we can determine the loss of information resulted from data reduction. For the Iris dataset, the percentage of variance explained by the first two PCs is 95.80%, which means that the part of the variance remained unexplained is 4.20%.

3.3. Sammon's Mapping

Multidimensional scaling is a class of methods used widely in behavioural, econometric, and social sciences. The data is projected down to two dimensions, while trying to preserve the distances between the data points [5].

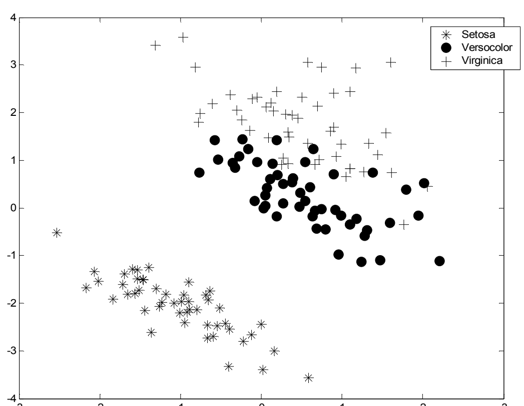


Figure 2. Sammon's mapping of the Iris data

The Sammon's mapping [10] is a multidimensional scaling technique that tries to match the pairwise distances of the lower-dimensional representations of the data items, with their original distances. Sammon's mapping is useful in visualizing class distributions, especially the degree of their overlap.

Figure 2 illustrates the Sammon's mapping applied to the Iris dataset.

3.4. SOM

Developed by Kohonen [5] in 1982, SOM is a technique widely used for clustering and abstraction based on unsupervised learning neural network. The SOM method lends itself to visually displaying the clusters formed by the data. It is similar to K-Means clustering algorithm, but the output of a SOM is topological and neighbouring clusters are similar. As a projection technique of multidimensional data onto a two-dimensional grid, the SOM method is similar to multidimensional scaling techniques, such as Sammon plot. One way to represent the data is to plot it (usually with jittering) as a scatter plot, using the horizontal and vertical axes produced by the Kohonen network (i.e., the map size).

Figure 3 represents the Iris dataset on a SOM grid of 12x9 nodes.

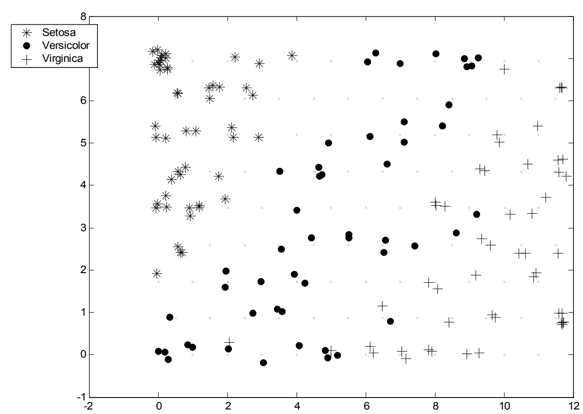


Figure 3. Self-Organizing Map of the Iris dataset

3.5. Radviz

The technique was developed at the Institute for Visualizations and Perception Research at the University of Massachusetts, Lowell, and is described in [11]. The idea is to represent each n-dimensional data item as a point, in a two-dimensional space (like in a scatter plot). The points are usually located within a circle, and more than two variables' values determine the actual position of each point. The perimeter of the circle is divided in n equal arcs. The equally spaced points on the perimeter are called anchorpoints or dimensional anchors [21]. Each data dimension is being associated with one anchorpoint. A data point is connected to a particular anchorpoint through a spring. When the data is n-dimensional, each data point will be connected to n anchorpoints through n different springs. The values of each data dimension are normalized to range within [0…1]. Each data point is then displayed at the position that produces a spring force sum of zero. If the data remains in the original range, then the variable with higher values than others will dominate the spring visualization. If all n coordinates have the same value (regardless whether they are low or high), the data point lies exactly in the centre of the circle. If the point is a unit vector point, it lies exactly at the fixed point on the edge of the circle, where the spring for that dimension is fixed.

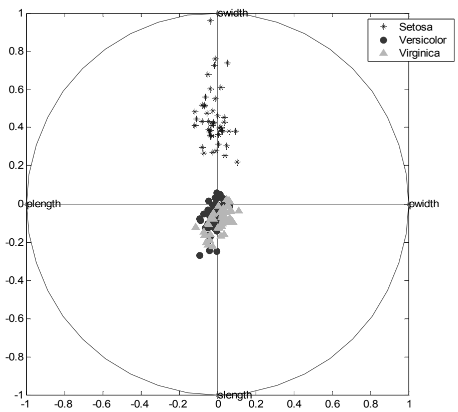


Figure 4. Radviz of the Iris data – local normalization

Figure 4 shows the Iris dataset on a two-dimensional space using the radviz technique. The ordering of the dimensions appears to be very important for the effectiveness of the visualization, different arrangements producing different visualizations. The data values were locally normalized so that each variable have mean zero and unit variance.

The difficulty of interpreting this type of visualization consists in the fact that many different points can map to the same position (for example, in Figure 4, the Versicolor and Virginica classes overlap and lie close to the centre of the circle).

3.6. Star Coordinates

Star Coordinates [12] is a multi-dimensional visualization technique, which is capable of mapping n-dimensional data onto a two-dimensional space. The idea of Star Coordinates is to arrange the n coordinate axes on a two-dimensional plane, such that all axes share the same origin point, but they are not necessarily orthogonal to each other. The minimum data value on each dimension is mapped to the origin, and the maximum value is mapped to the other end of the coordinate axis. It is recommended that each variable is normalized to range within [0…1]. Each image point corresponding to a data point has a location on the two-dimensional plane determined by the sum vector of all unit vectors on each coordinate, multiplied by the value of the data element for that coordinate.

Figure 5 displays the Iris dataset using the Star Coordinates technique.

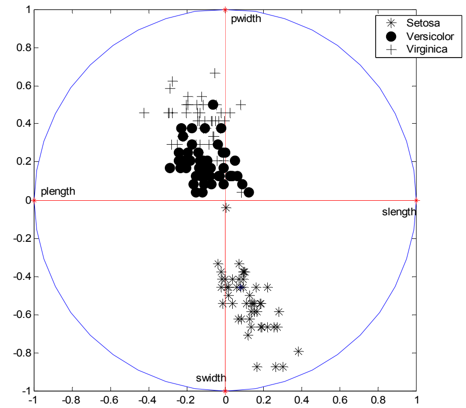


Figure 5. Star coordinates of Iris dataset – local normalization

Other projection methods are, for example, non-metric multidimensional scaling, principal curves and curvilinear component analysis [5].

4. Proposed approach to evaluation

4.1. Methods

According to Jain et al. [8], clustering methodology is particularly appropriate for the exploration of interrelationships among the data points, to make an assessment of their structure. Clustering techniques attempt to group points, so that the classes thereby obtained reflect the different data generation process represented in the dataset. Duda and Hart [6] point out that the clustering is a useful method in an early stage of an investigation, when it is valuable to gain insight into the nature or structure of the data. Consequently, the discovery of distinct subclasses or major departures from expected characteristics may significantly contribute to the understanding of the data for further processing (e.g., in classification tasks).

We investigate the use of clustering methodology to evaluate the effectiveness of projection techniques in preserving the structure of the original dataset. We first partition the data in clusters, and then evaluate the clustering solution obtained. Out of a very large number of available clustering algorithms [8, 17], we chose the *K-means* algorithm [22] for clustering the original and the transformed data. K-means is a partitional clustering technique, which performs well even on large datasets. Then, we calculate the external and internal clustering validity measures of the clustering solutions obtained for each dataset (original and transformed).

4.2. Clustering Validity Assessment

Cluster validity analysis is the assessment of a clustering procedure's output. Validity assessments are *objective* [8], and are performed to determine whether the output is meaningful.

A clustering structure is valid if it could not reasonably have occurred by chance or as an artefact of a clustering algorithm. When statistical approaches to clustering are used, validation is accomplished by carefully applying statistical methods and testing hypothesis.

There are three main types of validation studies: external validity, internal validity, and relative assessment. In this study, we will use only the external and internal validity criteria.

4.2.1. External validity

The ex*ternal validity assessment* compares the resulting structure to an a priori known structure. According to [17], external validity criteria are used for two purposes. The first is to compare a clustering structure *C*, produced by a clustering algorithm, with a partition *P* of a dataset *X*, which is known a priori. The second is to measure the degree of agreement between a predetermined partition *P* and the proximity matrix of *X*. The proximity matrix of *X* is a symmetric matrix in which the element (i, j) is defined as the distance between the x_i and x_j data points, from the dataset *X*, for *i, j=1…N*. Different distance measures can be used for calculating the proximity matrix, and in this study we have used Euclidean distance.

Comparison of *P* with a clustering *C*

Let $C = \{C_1,...,C_m\}$ and $P = \{P_1,...,P_s\}$ represent the clustering *C* and the predetermined partition *P* of a dataset *X*. The number of clusters in *C* is not necessarily the same as the number of groups in *P*. For a pair of vectors (x_v, x_u), we have the following cases:

a) SS – both vectors belong to the same cluster in *C* and to the same group in *P*.
b) SD – the vectors belong to the same cluster in *C*, but to different groups in *P*.
c) DS – the vectors belong to different clusters in *C*, but to the same group in *P*.
d) DD – both vectors belong to different clusters in *C* and to different groups in *P*.

Let *a*, *b*, *c*, and *d* be the number of SS, SD, DS, and DD pairs of vectors of *X*, respectively. Then, $a+b+c+d = M$, where M is the total number of possible pairs in *X*, that is, $M = N*(N-1)/2$, *N* being the number of data points.

Using the above definitions, the following statistical indices (statistics) are derived, in order to measure the degree to which *C* matches *P* [17]:

- *Rand statistic*: $R = (a+d)/M$
- *Jaccard coefficient*: $J = a/(a+b+c)$
- *Fowlkes and Mallows index*: $FM = \sqrt{\frac{a}{a+b} * \frac{a}{a+c}}$

For all the above defined indices, the larger are their values, the higher is the agreement between C and P. The corresponding statistical tests of these indices are *right tailed*. The statistical tests are needed to be carried out in order to ensure that the results (the value of the indices) are not achieved merely by chance.

We illustrate in Section 4.3.1 in Table 1, the calculation of the R, J, and FM statistics for the Iris dataset and the projected data, after applying K-means.

Theodoridis and Koutroumbas [17] provide an effective procedure of testing the null hypothesis of random structure, by using the Monte Carlo simulation technique. According to this procedure, in order to test the null hypothesis of random structure (or no structure), one should perform the following operations:

- For $i = 1:r$ (usually, $r = 100$):
  - Generate a dataset X_i of *N* vectors in the area of interest X, so that vectors are uniformly distributed in it;
  - Assign each vector $y^i{}_j \in X_i$ to the group where $x_j \in X$ belongs, according to the structure imposed by *P*;
  - Run the same clustering algorithm, used for obtaining *C* on X_i, and let C_i be the resulting clustering;
  - Compute the value $q(C_i)$ of the corresponding statistical index *q* for *P* and C_i;
- End {For};
- Create the histogram of $q(C_i)$'s.

By this procedure, one can generate *r* random samples of uniformly distributed data points with the same dimensionality as the original dataset *X*. For each random sample X_i, the clustering of the data is performed. The obtained structure C_i is compared to the predefined partition *P*. The total of *r* indices $q(C_i)$ represents an estimation of the theoretical distribution (probability density function) of the population index *q*, under the null hypothesis. A right-tailed test implies that for a significance level α, we reject the null hypothesis if the calculated value of the index is larger than the $q(C_i)$ corresponding to the significance level α.

For example, if we generate r = 100 random samples, we obtain for each random sample the R, J, and FM indices. We will have 100 indices R, 100 indices J, and 100 indices FM, which will be sorted in ascending order. For a chosen significance level α = 5%, one can reject the null hypothesis (that states that the clustering solution does not match the a priori known data structure), if the calculated statistics R, J and FM, respectively, for our dataset are greater than the highest 5 (i.e., 5%) indices R, J, and FM, respectively, obtained on the random samples. The lower is the value of the significance level, the more confidence one has that the result of the test is accurate.

Comparison of *P* with the proximity matrix *D*

For this purpose, the Hubert's Γ or normalized $\hat{\Gamma}$ statistic can be used. These statistics measure the degree to which the proximity matrix *D* of *X* matches a predefined partition *P*.

Let us consider the matrix *Y*, whose (i,j) element, $Y(i,j)$, is defined as follows (for $i,j=1,...,N$):

$$Y(i,j)=\begin{cases}1, \text{if } x_i \text{ and } x_j \text{ belong to different groups}\\ 0, \text{otherwise}\end{cases}.$$

Then, the Γ (or $\hat{\Gamma}$) statistic is applied to the *Y* and proximity matrix *D*. Its value is a measure of the degree to which *Y* matches *D*. For symmetric matrices *X* and *Y*, the Γ and $\hat{\Gamma}$ statistics are defined as follows:

- *Hubert's* Γ *statistic:*

$\Gamma=(1/M)\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}X(i,j)Y(i,j)$, where $X(i,j)$ and $Y(i,j)$ are the (i,j) elements of the matrices X and Y. High values of Γ indicate close agreement between X and Y.

- *Normalized* Γ *statistic:*

$$\hat{\Gamma}=\frac{(1/M)\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}(X(i,j)-\mu_X)(Y(i,j)-\mu_Y)}{\sigma_X\sigma_Y}.$$

$\mu_X,\mu_Y,\sigma_X,\sigma_Y$ represent the means and variances of matrices *X* and *Y*, respectively. $\hat{\Gamma}$ has values between -1 and 1. Large absolute values of $\hat{\Gamma}$ suggest agreement between the matrices *X* and *Y*.

We can also estimate the pdf (probability density function) of the Γ (or $\hat{\Gamma}$) under the null hypothesis of random structure. We generate *r* random *mappings* (i.e., cluster labels) g_i, for $i=1,...,r$. For each *mapping*, we form the corresponding Y_i matrix and we apply Γ (or $\hat{\Gamma}$) statistic to D and Y_i, for $i=1,...,r$. Then, we proceed as usually for the acceptance or rejection of the null hypothesis of random structure, at a given significance level α.

4.2.2. Internal validity

The *internal validity assessment* tries to determine if the resulted structure matches information that is inherent in the dataset *X*. The method of assessment is based on the information contained in the proximity matrix *D* of the original data. The Γ (or $\hat{\Gamma}$) statistic can be used for assessment. The (i,j) element of the matrix *Y* is defined as follows (for $i,j=1,...,N$):

$$Y(i,j)=\begin{cases}1, \text{if } x_i \text{ and } x_j \text{ belong to different clusters}\\ 0, \text{otherwise}\end{cases}.$$

Y is computed for the resulting clustering solutions. Then, the Γ (or $\hat{\Gamma}$) statistic is applied to the *Y* and proximity matrix *D*. Its value is a measure of the degree of correspondence between *Y* and *D*. We can also estimate the pdf of the Γ (or $\hat{\Gamma}$) under the null hypothesis of random structure. We generate *r* random *samples* X_i, for $i=1,...,r$. For each *sample*, we compute the proximity matrix D_i. Then, we apply to each sample the same clustering algorithm in order to obtain a structure C_i. We compute Y_i for each clustering obtained, and apply the Γ (or $\hat{\Gamma}$) statistic. Finally, we decide the acceptance or rejection of the null hypothesis at a given significance level α.

4.3. Datasets

For our study, we have selected three datasets that are well known for their suitability for clustering and classification tasks. We have chosen these datasets based on the framework proposed in [2] for benchmarking visualization techniques. The evaluation approach described in [2] was based on defining a number of tasks suitable for each dataset and evaluating subjectively (by visual inspection) the extent to which one visualization technique is effective in solving the tasks. In our study, we investigate the possibility of using objective measures (cluster validity measures) in order to evaluate the effectiveness of a projection technique in preserving the original data structure (clusters). The datasets are available at the UCI Machine Learning Repository [20]. We have also used a fourth dataset, which was artificially created so that it contains three well separated groups.

4.3.1. Iris dataset

This dataset was described in Section 3.1. For illustrating the computation of the external cluster validity indices, we exemplify on this dataset. The number of data points is *N=150*, the total number of pairs of data points is *M=N*(N-1)/2=11175*.

Table 1. Illustration of the calculation of R, J, and FM

Dataset	a	b	c	d	R	J	FM
Original	3075	600	744	6756	0.880	0.696	0.821
PCA	2742	933	942	6558	0.832	0.594	0.745
Sammon	2715	960	964	6536	0.828	0.585	0.738
Radviz	3306	369	2500	5000	0.743	0.535	0.716
Star C.	3024	651	652	6848	0.883	0.699	0.823
SOM	3030	645	766	6734	0.874	0.682	0.811

4.3.2. Voting records database

This dataset consists of votes for each of the US House of Representatives Congressmen, on the 16 key issues. There are nine different types of votes, which are simplified to three types: yes, no, and unknown. There are 435 instances (267 democrats, and 168 republicans). There are 16 variables and one class variable with two values to be predicted (democrat or republican).

4.3.3. Wine recognition data

These data are the results of a chemical analysis of wines grown in a region of Italy, but derived from three different cultivars. The analysis determined the quantities of 13 constituents found in each of the three types of wines. There are 178 instances (59 in the first class, 71 in the second class and 48 in the third class). The three classes are separable.

4.3.4. Artificially created data

This dataset contains 150 data points of dimensionality 4. The data was randomly generated from three normal distributions with known means and variances. Therefore, the dataset consists of three distinct groups, each group having 50 data points. The means and variances characterising each group in this dataset are given by matrix *M* and *V*. *M* represents the means of the 4 variables (columns) in each group (rows), and *V* represents the respective variances.

$$M = \begin{bmatrix} 1 & 2 & 3 & 3 \\ 4 & 7 & 8 & 8 \\ 10 & 10 & 15 & 15 \end{bmatrix}, \text{ and } V = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

5. Results

5.1. External validity

Comparison of *P* with a clustering *C*

In the following, we present the results of comparing the known partitions of the datasets with the clusterings obtained using K-means. For comparison we have used the indices R, J and FM. We have also tested the null hypothesis according to which the data has no structure, and therefore the assignment of the data points to different clusters is random. The following three tables (Table 2-Table 4) show the results of the external validation of the clusterings obtained, based on different datasets.

The rows of each table correspond to different transformations of the dataset. The corresponding indices are calculated based on comparison between the clustering obtained using K-means algorithm and the a priori known structure of the dataset. The *original* data represents the data without any transformation. The results obtained on the original data are useful for assessing the extent to which the structure (clusters) of the data can be revealed by the clustering algorithm used.

Each table shows also the estimates for the R, J and FM statistics, respectively, under the null hypothesis H_0, at the *significance level of 1%* (R/H_0, J/H_0, and FM/H_0). By comparing the estimate with the actual value of the statistics, one can reject or accept the null hypothesis as follows. If the actual statistic is larger than the estimate, we reject H_0, otherwise we accept H_0. The rejection of H_0 signifies that the resulting clustering matches the a priori known structure of the data.

Table 2. The external validation of clustering obtained on Iris datasets. The higher the R, J and FM, the more effective the projection

Projection	R	R/H_0	J	J/H_0	FM	FM/H_0
Original	0.880	0.569	0.696	0.212	0.821	0.350
PCA	0.832	0.569	0.594	0.215	0.745	0.355
Sammon	0.828	0.568	0.585	0.215	0.738	0.354
Radviz	0.743	0.573	0.535	0.213	0.716	0.351
StarCoord	0.883	0.573	0.699	0.215	0.823	0.354
SOM	0.846	0.568	0.622	0.212	0.767	0.350

Table 2 show high values of R, J and FM obtained on the Iris dataset. For this dataset, Star coordinates projection leads to the best representation of data, and the K-means recovers very well the clustering structure of the data. The projection which does not exactly represent the known structure of the data is given by the Radviz technique. One can compare these results with the plots corresponding to each projection technique (Section 3), and confirm the findings. Table 2 shows that the hypothesis of random structure has been rejected in all cases.

Table 3 shows that, in case of the Voting dataset, the Star Coordinates technique gives the most different representation of data from its original structure. The projection technique that best preserves the structure of the data is PCA, followed by Sammon's mapping and SOM. The hypothesis of random structure has been rejected in all six cases.

Table 3. The external validation of clustering obtained on Voting datasets. The higher the R, J and FM, the more effective the projection

Projection	R	R/H_0	J	J/H_0	FM	FM/H_0
Original	0.772	0.508	0.636	0.361	0.778	0.530
PCA	0.775	0.508	0.640	0.351	0.781	0.520
Sammon	0.762	0.504	0.622	0.349	0.767	0.517
Radviz	0.720	0.505	0.572	0.348	0.728	0.517
StarCoord	0.682	0.512	0.527	0.357	0.690	0.527
SOM	0.762	0.508	0.622	0.352	0.767	0.521

Table 4 reveals another interesting result, namely that a projection technique can be used in conjunction with a clustering technique to recover the structure of the dataset even with better results than using the original dataset. The PCA, Sammon's mapping and SOM are well suited techniques for mapping this dataset onto a low-dimensional space and the K-means technique reveals the true structure of the data after the projection. In all cases, the null hypothesis has been rejected.

Table 4. The external validation of clustering obtained on Wine datasets. The higher the R, J and FM, the more effective the projection

Projection	R	R/H_0	J	J/H_0	FM	FM/H_0
Original	0.719	0.567	0.412	0.214	0.584	0.352
PCA	0.953	0.570	0.870	0.217	0.930	0.357
Sammon	0.925	0.564	0.799	0.217	0.888	0.356
Radviz	0.753	0.566	0.471	0.214	0.641	0.353
StarCoord	0.801	0.577	0.540	0.225	0.702	0.367
SOM	0.947	0.564	0.855	0.210	0.922	0.347

Table 5 presents the R, J, and FM obtained on the artificial dataset after applying K-means. For this dataset, PCA, Sammon's mapping and SOM have yielded the best projections of the data, the R, J and FM values showing exact match between the resulted clustering and the actual partitioning. The other two projection techniques did not perform well with respect to preserving the original structure in this dataset.

Table 5. The external validation of clustering obtained on Artificial datasets. The higher the R, J and FM, the more effective the projection

Projection	R	R/H_0	J	J/H_0	FM	FM/H_0
Original	1	0.575	1	0.215	1	0.354
PCA	1	0.566	1	0.215	1	0.355
Sammon	1	0.568	1	0.222	1	0.364
Radviz	0.571	0.582	0.362	0.224	0.554	0.366
StarCoord	0.627	0.568	0.282	0.213	0.440	0.352
SOM	1	0.572	1	0.218	1	0.358

Comparison of *P* with the proximity matrix *D*

In the following, we present the results of comparing the proximity matrices of each transformed dataset (including also the original dataset) with the a priori known partition of the dataset. For comparison we have used the $\hat{\Gamma}$ statistic, calculated based on the procedure described in Section 4.2.1. The proximity matrices of each dataset (original or transformed) were normalized using he histogram equalization method. The histogram equalization method works in two steps. Firstly, the elements of the proximity matrix are sorted and replaced by their rank. Secondly, the newly data elements of the proximity matrix are normalized to range in [0…1].

We have also tested the null hypothesis according to which the dataset does not have structure and the assignment of the data points to the a priori known classes is random.

Table 6 shows the Hubert's normalized statistic, $\hat{\Gamma}$, obtained for our datasets. The higher is the value of $\hat{\Gamma}$, the higher is the degree of agreement between the proximity matrix of the original data and the clustering solution obtained. The data obtained using SOM, PCA and Sammon's mapping techniques appear to preserve best the original data structure. The null hypothesis has been rejected in all cases at significance level of 1%.

Table 6. The $\hat{\Gamma}$ statistic for comparing the proximity matrix of each transformed dataset with the actual partition of the original dataset.

	Iris data		Voting data		Wine data		Artificial	
Projection	$\hat{\Gamma}$	$\hat{\Gamma}$ / H_0	$\hat{\Gamma}$	$\hat{\Gamma}$ / H_0	$\hat{\Gamma}$	$\hat{\Gamma}$ / H_0	$\hat{\Gamma}$	$\hat{\Gamma}$ / H_0
Original	0.69	0.04	0.35	0.01	0.43	0.08	0.81	0.04
PCA	0.63	0.05	0.55	0.01	0.72	0.03	0.81	0.04
Sammon	0.65	0.04	0.46	0.01	0.63	0.03	0.81	0.04
Radviz	0.52	0.05	0.31	0.01	0.51	0.04	0.27	0.11
Star C	0.64	0.05	0.35	0.02	0.43	0.05	0.19	0.08
SOM	0.69	0.03	0.59	0.02	0.75	0.03	0.85	0.03

5.2. Internal validity

In this subsection, we present the results of assessing the internal validity of the clustering solutions obtained on original and transformed datasets. We have used K-means for partitioning each dataset (original or transformed) in three clusters. For evaluation, we have used $\hat{\Gamma}$ statistic, calculated based on the proximity matrix of the original dataset and the clustering solutions obtained on the original or transformed datasets. The proximity matrices were normalized using histogram equalization method.

We have also tested the null hypothesis according to which the clustering solution is obtained merely by chance (i.e., the data has no structure). Rejection of this hypothesis signifies that the obtained clustering are meaningful, and the data points in each cluster are similar to each other, while data points from different clusters are different. High values of $\hat{\Gamma}$ show good internal validity of obtained the clusters. Table 7 shows the results.

Table 7. The $\hat{\Gamma}$ statistic for assessing internal validity of the partitions obtained using K-means on the original and transformed data

	Iris data		Voting data		Wine data		Artificial	
Projection	$\hat{\Gamma}$	$\hat{\Gamma}$ / H_0	$\hat{\Gamma}$	$\hat{\Gamma}$ / H_0	$\hat{\Gamma}$	$\hat{\Gamma}$ / H_0	$\hat{\Gamma}$	$\hat{\Gamma}$ / H_0
Original	0.73	0.04	0.49	0.01	0.69	0.04	0.81	0.03
PCA	0.69	0.06	0.49	0.01	0.40	0.04	0.81	0.02
Sammon	0.69	0.06	0.49	0.01	0.41	0.04	0.81	0.02
Radviz	0.76	0.06	0.32	0.01	0.24	0.04	0.24	0.02
Star C	0.62	0.06	0.44	0.01	0.36	0.04	0.12	0.02
SOM	0.69	0.04	0.48	0.01	0.43	0.04	0.81	0.03

Table 7 reveals for the Iris data that, in terms of internal validity, the clusters obtained on the data projected using Radviz technique are the most compact clusters. However, the clusters evaluated by this procedure are not necessary the same clusters as the known classes in the dataset. By looking at Table 2, one can see that the clustering obtained on Radviz data is the least fitted to the known structure of the dataset. The result has the following message: even if the projected data does not fit best an a priori known structure, it may reveal other interesting patterns or structure. For the other three datasets, the comparison among techniques shows that PCA, Sammon's mapping and SOM are better than Radviz and Star Coordinates to be used in clustering. The null hypothesis has been rejected in all cases at significance level of 1%.

5.3. Discussion

There are few remarks that need to be made. The results obtained for all the cluster validity measures used previously are dependent on the selection of a number of parameters. These parameters regard aspects such as: the selection of the clustering algorithm, the parameters of the clustering algorithm, and, finally, the parameters of the projection techniques.

Regarding the clustering algorithm, we have chosen to use the K-means. Depending on the structure of the original dataset (e.g., the shape of clusters, etc.) other clustering algorithms can be used.

Regarding the parameters of the clustering algorithm (e.g., K-means), there are several choices that can be made. One decision regards the similarity measure used by the algorithm (Euclidean, Squared Euclidean, Mahalanobis, Cityblock, etc.). In our experiments, we have used Squared Euclidean distance. The use of other distance measure can result in other values of the cluster validity measures.

Finally, regarding the parameters of the projection techniques, they concern each technique in part. In the case of PCA, one can use standardized or mean corrected data to obtain the PCA solution. We have used standardized data. The Sammon's mapping was computed using the Euclidean distance, but other distance measures can be used. For the Radviz and Star Coordinates projection we used local normalization of the data. Each of these two techniques creates a different representation of the data depending on the arrangement of the dimensions in the 2D space. Different arrangements will yield different mappings in 2D, and therefore different cluster validity measures after applying the clustering. SOM, as any neural network algorithm, has several parameters to be set. To name just a few, we mention the map size and the type of learning.

Therefore, the decisions made by the researcher in designing the study and experiments can influence the results of the experiments. However, the hypothesis testing approach represents a means to check that the results are not obtained merely by chance.

The results summarised in Tables 2-7 show that the external validity measures are good indicators of how well the structure of the new data resembles the actual structure of the data. All four statistics used to measure the external validity confirmed the less capability of Radviz technique to reveal the three clusters the Iris data form, as Figure 4 also showed. The external validity measures can also be used in comparing different projection techniques with respect to their capability of preserving the original structure. A limitation of this approach is that in real applications the structure of the data is rarely known.

Figure 6 shows how the external validity measure R assesses the clustering obtained on different datasets and their projections. Similar plots were obtained for J, FM and $\hat{\Gamma}$ statistics. The plots illustrate that the external validity measures are good indicators of the effectiveness of a projection technique in preserving and displaying the clusters in the data. They also show that effectiveness of the projection techniques is highly depended on the original datasets.

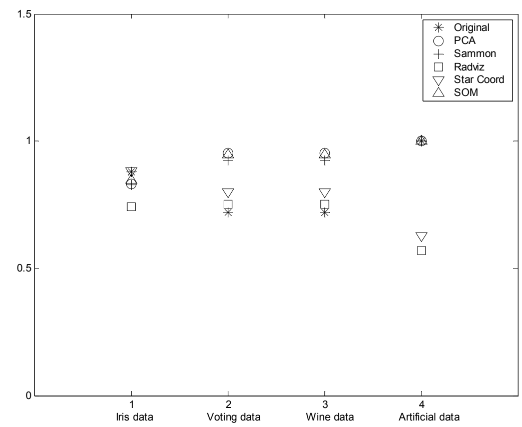


Figure 6. External validity: Plot of the R statistic for different projections and different datasets

The internal validity measure is also a good measure to assess the effectiveness of a projection technique in displaying clusters in data. This can be used even when the structure of the data is not known, because it measures the degree to which the obtained clusters are internally consistent, and not the degree to which they match an a priori known structure.

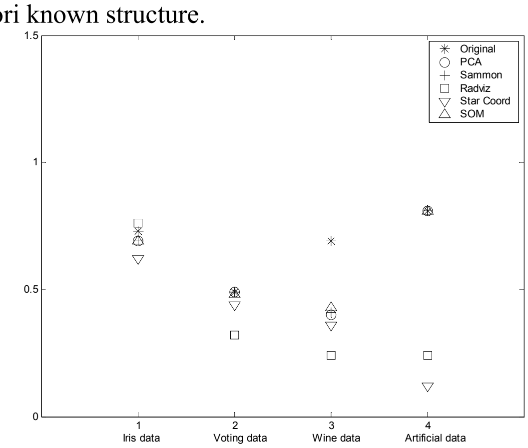


Figure 7. Internal validity: Plot of the $\hat{\Gamma}$ statistic for different projections and different datasets

6. Conclusion

In this paper, we investigated the effectiveness of five projection techniques (PCA, Sammon's Mapping, Radviz, Star Coordinates, and SOM) in visual data mining. We illustrated the use of these techniques on the well known Iris dataset, in order to facilitate visual comparison of the capabilities of each projection technique with respect to their effectiveness for solving a data-mining task such as clustering. Because this visual inspection is highly subjective, we investigated the use of cluster validity measures in order to judge the effectiveness of the projection techniques in visual data mining. The results showed that cluster validity is a successful approach to evaluate objectively the visualization techniques.

The results also showed that all the five projection techniques are capable of representing the structure of the dataset, and that their effectiveness depends on the dataset under analysis. We illustrated that external and internal cluster validity measures are suitable for evaluating the effectiveness of different projection technique in visual data mining, and can be used to choose the right data representation technique for a particular task, before further processing the data.

For future work, we intend to investigate the use of relative criteria for assessing the effectiveness of applying a clustering technique on projected data. We aim also to study the capabilities of other projection techniques to represent the high-dimensional data down to two dimensions.

References

[1] A. Edmunds and A. Morris, The problem of information overload in business organisations: a review of the literature, *Int. J. Inf. Management, vol. 20*, 2000, 17-28.

[2] G. G. Grinstein, P. E. Hoffman, R. M. Pickett and S. J. Laskowski, Benchmark development for the evaluation of visualization for data mining, in Fayyad, U., Grinstein, G. G., and Wierse, A. (eds.), *Information visualization for data mining and knowledge discovery* (San Francisco: Morgan Kaufmann, 2002), 129-176.

[3] J. J. Thomas and K. A. Cook, *Illuminating the Path: The Research and Development Agenda for Visual Analytics* (National Visualization and Analytics Center, 2005).

[4] D. A. Keim, Information visualization and visual data mining, *IEEE Transactions on Visualization and Computer Graphics, vol. 8*, 2002, 1-8.

[5] T. Kohonen, *Self-Organizing Maps, 3rd* ed. (Berlin; New York: Springer, 2001).

[6] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis* (New York: John Wiley and Sons, Inc., NY, 1973).

[7] K. Fukunaga, *Introduction to Statistical Pattern Recognition* (San Diego, CA: Academic Press, 1990).

[8] A. K. Jain, M. N. Murty and P. J. Flynn, Data clustering: a review, *ACM Comput. Surv., vol. 31*, 1999, 264-323.

[9] S. Sharma, *Applied Multivariate Techniques* (New York, NY: John Wiley & Sons, Inc., 1995).

[10] J. W. Sammon, A nonlinear mapping for data structure analysis, *IEEE Trans. Comput., vol. C-18*, 1969, 401-409.

[11] P. E. Hoffman, G. G. Grinstein, K. Marx, I. Grosse, and E. Stanley, DNA visual and analytic data mining, in *Proceedings of IEEE Visualization '97*, Phoenix, Az, 1997, 437–441.

[12] E. Kandogan, Visualizing multi-dimensional clusters, trends, and outliers using star coordinates, *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2001,* 107-116.

[13] D. A. Keim, Designing pixel-oriented visualization techniques: theory and applications, *IEEE Trans. Visual. Comput. Graphics, vol. 6*, 2000, 59-78.

[14] C. Chen and M. P. Czerwinski, Empirical evaluation of information visualizations: an introduction, *Int. J. Human-Computer Studies, vol. 53*, 2000, 631-635.

[15] M. Halkidi, Y. Batistakis and M. Vazirgiannis, Cluster validity methods: part I, *SIGMOD Rec., vol. 31*, 2002, 40-45.

[16] M. Halkidi, Y. Batistakis and M. Vazirgiannis, Cluster validity methods: part II, S*IGMOD Rec*., vol. 31, 2002, 19-27.

[17] S. Theodoridis and K. Koutroumbas, *Pattern Recognition* (San Diego: Academic Press, 1999).

[18] The MathWorks, *MATLAB-The Language of Technical Computing* (Natick, MA: The MathWorks, Inc, 2000).

[19] J. Vesanto, J. Himberg, E. Alhoniemi and J. Parhankangas, Self-organizing map in Matlab: the SOM toolbox, *Proceedings of the Matlab DSP Conference,* 1999, 35-40.

[20] D. J. Newman, S. Hettich, C. L. Blake and C. J. Merz, UCI Repository of machine learning databases, http://www.ics.uci.edu/~mlearn/MLRepository.html, 1998.

[21] P. Hoffman, G. Grinstein and D. Pinkney, Dimensional anchors: a graphic primitive for multidimensional multivariate information visualizations, *Proc. 1999 Workshop on New Paradigms in Information Visualization and Manipulation, in Conjunction with the 8h ACM Int'l. Conf. Information and Knowledge Management (CIKM '99),* 1999, 9-16.

[22] J. MacQueen, Some methods for classification and analysis of multivariate observations, *Proc. Fifth Berkeley Symposium on Math. Stat. and Prob., vol. 1*, 1967, 281-297.

Paper 3

Marghescu, D. (2007). Evaluation of Projection Techniques Using Hubert's Γ Statistics. In *Proceedings of the IADIS 1st European Conference on Data Mining (DM 2007)*, Lisbon, Portugal, July 2007.

ISBN: 978-972-8924-40-9 ©2007 IADIS

EVALUATION OF PROJECTION TECHNIQUES USING HUBERT'S Γ STATISTICS

Dorina Marghescu
Turku Centre for Computer Science, Åbo Akademi University, Dept. of IT

ABSTRACT

Projection techniques reduce the data dimensionality by combining the original variables into a smaller number of new dimensions, in a linear or nonlinear manner. The projection methods are particularly useful because they lend themselves to visual representations of data, when the number of new dimensions is one, two or three. In this paper, the aim is to evaluate different visualization techniques based on projection techniques with respect to their effectiveness in preserving the inherent relationships and structure of the dataset. For this purpose, we investigate the use of the Hubert's Γ statistics for evaluating the fit between the distance matrices of original data and projected data. Moreover, we investigate the use of the modified Hubert's Γ statistics for evaluating the effectiveness of projection techniques in preserving the clustering structure inherent in the dataset, if such structure is present.

KEYWORDS

Projection techniques, visualization, evaluation, Hubert's Γ statistic

1. INTRODUCTION

In this paper, we illustrate the use of projection techniques for visualizing high-dimensional data and evaluate their effectiveness in preserving the inherent relationships and structure that exist in the dataset. We study the use of *Hubert's Γ statistics* in order to *objectively* assess the effectiveness of the projection techniques in preserving the relationships and structure in the data. More specifically, we investigate the use of the Hubert's Γ statistics for evaluating the *fit between the distance matrices* of original data and projected data. We also examine the use of the modified Hubert Γ statistics for evaluating the effectiveness of projection techniques in preserving the *clustering structure* inherent in the dataset, if such structure is present.

Section 2 illustrates the projection techniques under analysis. Section 3 presents the definitions of Hubert's Γ statistics and its modified version. Section 4 proposes the procedures for calculating Hubert's Γ statistics and its modified version for evaluating the effectiveness of the projection techniques. Section 5 presents the evaluation results. We conclude with final remarks and future work ideas in Section 6.

2. PROJECTION TECHNIQUES

Projection methods are used to reduce the variable space (i.e., the original variables are combined in a linear or nonlinear manner into a smaller number of new data dimensions). The projection methods are particularly useful because they facilitate visual representations of data. In this section, we look at two classical projection techniques: Principal Components Analysis (PCA) and Sammon's mapping, and three more recently developed techniques: Self Organizing Maps (SOMs), Radviz and Star Coordinates.

For illustrating the capabilities of each projection technique, we use the Iris dataset (Newman et al. 1998), due to its suitability for classification and clustering tasks. The data concerns three species of flowers characterised by four attributes: petal length and width, and sepal length and width. The class variable is the type of flower: Iris-Setosa, Iris-Versicolor, and Iris-Virginica. The dataset contains 150 observations, each class containing 50 flowers. The class Iris-Setosa is linearly separable from the other two classes, but Iris-Versicolor and Iris-Virginica classes are not linearly separable.

2.1 PCA

PCA is a classical statistical technique that maps high-dimensional data items onto a lower-dimensional space (Sharma 1995). The transformation tries to preserve the variance of the original data as well as possible. The PCA technique creates new variables (called principal components), which are linear composites of the original variables. The maximum number of new variables that can be formed is equal to the number of original variables, and the new variables are uncorrelated among themselves. Figure 1 represents the PCA projection of the standardized data using the first two principal components.

2.2 Sammon's Mapping

The Sammon's mapping (Sammon 1969) is a metric multidimensional scaling technique that tries to match the pairwise distances of the lower-dimensional representations of the data items with their original distances. Figure 2 illustrates the Sammon's mapping of the Iris dataset, after applying variance normalization to the original data and using the Euclidean distance in the Sammon's mapping algorithm.

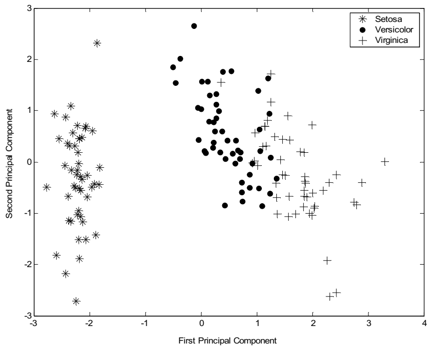


Figure 1. PCA

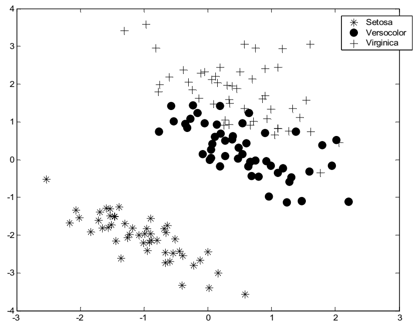


Figure 2. Sammon's mapping

2.3 SOM

Developed by Kohonen in 1982, SOM is a technique widely used for clustering and abstraction based on unsupervised learning neural network (Kohonen 2001). The SOM represents the data items on a two-dimensional grid, where each item is assigned to a node of the grid in an orderly way so that similar data items are mapped to the same node or neighbouring nodes. Figure 3 represents the Iris dataset on a SOM grid of 12x9 nodes. The technique of jittering was used to change with a small value the position of each data item so that the items mapped to the same node will not overlap. The data was first normalized using the variance method. Other parameters of the SOM were initialised as follows: Gaussian neighbourhood, radius [12, 1], batch training, and linear initialization.

2.4 Radviz

The technique was developed by Hoffman et al. (1997). The *n*-dimensional data items are represented as points in a two-dimensional space. The points are located within a circle whose perimeter is divided in *n* equal arcs. The equally spaced points on the perimeter are called *anchorpoints* or *dimensional anchors* (Hoffman et al. 1999), they being associated with each data dimension. The values of each data dimension must be normalized to range within [0…1]. The data item is connected to the *n* anchorpoints through *n* different *springs*. Each data point is then displayed at the position that produces a spring force sum of zero. If all *n* coordinates have the same value (regardless whether they are low or high), the data point lies exactly in the centre of the circle. If the point is a unit vector point, it lies exactly at the fixed point on the edge of the circle, where the spring for that dimension is fixed. Figure 4 shows the Radviz projection of the Iris dataset, after local normalization of the data in the range [0 1].

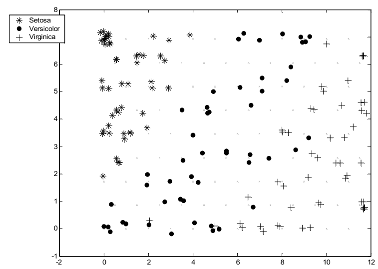


Figure 3. Self-Organizing Map

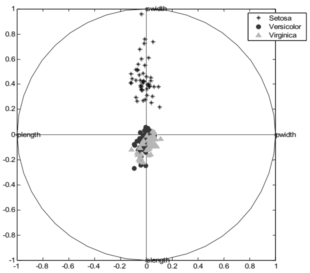


Figure 4. Radviz

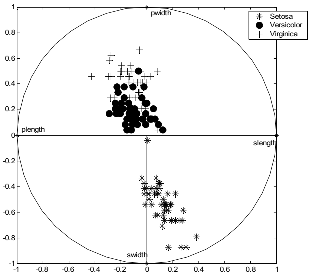


Figure 5. Star Coordinates

2.5 Star Coordinates

Star Coordinates (Kandogan 2001) maps *n*-dimensional data onto a two-dimensional space. The *n* coordinate axes are arranged on a two-dimensional plane, such that all axes share the same origin point, but they are not necessarily orthogonal to each other. The minimum value on each dimension is mapped to the origin, and the maximum value is mapped to the other end of the coordinate axis. Each variable is normalized to range within [0…1]. Each image point corresponding to a data item has a location on the two-dimensional plane determined by the sum vector of all unit vectors on each coordinate, multiplied by the value of the data item for that coordinate. Figure 5 displays the Star Coordinates projection of the Iris data, after normalization.

3. HUBERT'S Γ STATISTICS AND ITS MODIFIED VERSION

Hubert's Γ statistic (Hubert and Schultz 1976) is an index that measures the correlation between two matrices, *A* and *B*, of dimensions *N x N*, drawn independently of each other (Theodoridis and Koutroumbas 1999). Theodoridis and Koutroumbas discuss the use of this type of statistics as external and internal criteria for assessing clustering validity. For two symmetric matrices, *A* and *B*, this statistic is defined as follows:

$$Hubert's\ \Gamma = (1/M)\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} A(i,j)B(i,j) \qquad (1),$$

where $A(i,j)$ and $B(i,j)$ are the (i,j) elements of matrices *A* and *B,* and $M = N(N-1)/2$. High values of Γ indicate close agreement between *A* and *B.* The normalized Γ statistic, denoted $\hat{\Gamma}$, can also be used.

$$\hat{\Gamma} = \frac{(1/M)\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}(A(i,j)-\mu_A)(B(i,j)-\mu_B)}{\sigma_A\sigma_B} \qquad (2),$$

where $\mu_A, \mu_B, \sigma_A, \sigma_B$ represent the means and squared variances of *A* and *B*, respectively. $\hat{\Gamma}$ has values between -1 and 1. Large absolute values of $\hat{\Gamma}$ suggest agreement between the matrices *A* and *B*.

Theodoridis and Koutroumbas (1999) discuss also the **modified Hubert's Γ statistic** as a *relative* measure to compare different clustering solutions obtained by using different clustering algorithms. Let *X* be the original dataset containing *N* data items and *P*, its *proximity matrix*, where the element $P(i,j)$ is the distance between two vectors x_i and x_j of *X*. By applying a clustering algorithm to the dataset *X*, we obtain the partition *C*, in which individual clusters are denoted $C_1, C_2, \ldots, C_m$, *m* being the total number of clusters. Let *Q* be the *N x N* matrix whose (i,j) element, $Q(i,j)$, is equal to the distance $d(w_{c_i}, w_{c_j})$ between the representatives w_{c_i}, w_{c_j} of the clusters where x_i and x_j belong. The same distance measure must be used for both *P* and *Q*. The modified Hubert's Γ statistic and its normalized version are defined using the Eq. (1) and (2), respectively, where the matrix *A* is the proximity matrix *P* of the dataset and the matrix *B* is the matrix *Q*. The normalized modified $\hat{\Gamma}$ has values between -1 and 1. Large absolute values of $\hat{\Gamma}$ suggest agreement between the matrices *P* and *Q*.

4. PROPOSED APPROACH TO EVALUATION OF PROJECTIONS

In this section, we propose two procedures for evaluating the projection techniques based on *Hubert's Γ and modified Hubert's Γ statistics*. First, we use the Hubert's Γ to evaluate the extent to which the obtained projected data preserved the inherent relationships between the data points, measured in terms of *proximity matrix*. The **proximity matrix** of a dataset *X* is a symmetric matrix consisting of the pair-wise distances of elements of *X*. Second, we use the modified index to evaluate the extent to which the obtained projected data preserves the clustering structure inherent in the data, if such structure exists.

Both procedures require that the original data dimensions are standardized (by subtracting the mean and dividing by standard deviation), before calculating the proximity matrix of original data. Moreover, the procedures include the normalization of the proximity matrices using the *global histogram equalization* method. This method works in two steps: first, the data values are replaced by the order index, and then these values are normalized to be in the range [0, 1], by applying a linear transformation.

4.1 Procedure for Calculating Hubert's Γ and $\hat{\Gamma}$ Statistics for Projected Data

1. Calculate the proximity matrix *Dx* of the original *standardized* data.
2. Calculate the proximity matrix *Dp* for projected data.
3. Normalize *Dx* and *Dp* using *global histogram equalization* method so that the distances are comparable.
4. Calculate the Hubert's Γ statistic by applying Eq. (1) to normalized proximity matrices *Dx* and *Dp*. The value of this statistic will indicate the extent to which the two proximity matrices (of original data and projected data) reflect the same inherent relationships of the data points.
5. Similarly, we can calculate the normalized Hubert's Γ statistic given in Eq. (2) using the normalized proximity matrices *Dx* and *Dp*.
6. Repeat steps 1-5 for each obtained projected dataset.
7. The higher the Hubert Γ (or the absolute values of $\hat{\Gamma}$) statistics, the better is the preservation of the data relationships after projection.

4.2 Procedure for Calculating Modified Hubert's Γ And $\hat{\Gamma}$ Statistics for Projected Data after Clustering

1. Apply a clustering algorithm (e.g., K-means algorithm (MacQueen 1967)) on projected data.
2. Calculate the proximity matrix *Dx* of the original *standardized* data.
3. Calculate a *N x N* matrix *Q*, whose (i, j) element $Q(i, j)$ is equal to the distance $d(w_{c_i}, w_{c_j})$ between the representatives of the clusters where x_i and x_j belong. The same distance measure must be used for calculating both *Dx* and *Q*.
4. Normalize *Dx* and *Q* using *global histogram equalization* method so that the distances are comparable.
5. Calculate the modified Hubert's Γ statistic by applying Eq. (1) to the normalized proximity matrices *Dx* and *Q*. This statistics will indicate the extent to which the clustering of the projected data reflects the inherent structure of the original data (given by the normalized proximity matrix *Dx*).
6. Similarly, we can calculate the normalized modified Hubert's Γ statistic by applying Eq. (2) to the normalized proximity matrices *Dx* and *Q*.
7. Repeat steps 1-6 for each obtained projected dataset.
8. The higher the modified Hubert Γ (or the absolute value of $\hat{\Gamma}$) statistics, the better is the preservation of the data structure after projection.

To investigate the use of these indices for the evaluation of projection techniques, we performed a number of experiments on known datasets. The computations were realized in Matlab (The MathWorks 2000). For obtaining the SOM and Sammon's mapping, we used the SOM Toolbox (Vesanto et al. 1999).

5. RESULTS

Three datasets with clustering structure available at UCI Machine Learning Repository (Newman et al. 1998) were used for analysis in this study: *Iris dataset*, *Voting records database* and *Wine recognition data*. We also used two artificially created datasets, one which contains three well separated clusters (*Artificial 1*) and another that does not have a clustering structure (*Artificial 2*). **Artificial 1** dataset contains 150 data points of dimensionality 4. The data was randomly generated from three normal distributions with known means and variances. Therefore, the dataset consists of three distinct groups, each group having 50 data points. The means and variances characterising each group in this dataset are given by the matrices M_1 and V_1. M_1 represents the means of the 4 variables (columns) in each group (rows), and V_1 represents the respective variances:

$$M_1 = \begin{bmatrix} 1 & 2 & 3 & 3 \\ 4 & 7 & 8 & 8 \\ 10 & 10 & 15 & 15 \end{bmatrix} \text{ and } V_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Artificial 2 dataset contains 200 data points of dimensionality 5. The data was randomly generated from one normal distribution with means $M_2 = [1\ 2\ 3\ 3\ 5]$ and variances $V_2 = [1\ 1\ 1\ 1\ 1]$.

5.1 Results of Normalized Hubert's Γ Statistics for All Datasets

Table 1 shows the values of the normalized Hubert's Γ statistics calculated for the proximity matrices of the projected datasets. We used for calculating these indices the procedure 4.1. It is observed that for these datasets, the PCA, Sammon's mapping and SOM are the most effective techniques in preserving the inherent relationships between the data points calculated in terms of Euclidean distance.

Table 1. Normalized Hubert's Γ statistics of projected data (the higher the index, the better the projection)

	Iris	*Vote*	*Wine*	*Artif. 1*	*Artif. 2*
PCA	**0.98**	**0.81**	**0.82**	**0.99**	0.64
Sammon's mapping	**0.97**	**0.84**	**0.89**	**0.99**	**0.76**
SOM	**0.96**	**0.77**	**0.78**	**0.97**	0.55
Radviz	0.74	0.49	0.68	0.32	0.63
Star Coordinates	0.83	0.59	0.61	0.20	0.64

Table 2. Normalized modified Hubert's Γ statistics after clustering (the higher the index, the better the obtained clustering)

	Iris	*Vote*	*Wine*	*Artif 1*	*Artif 2*
Original	**0.88**	**0.65**	0.33	**0.93**	0.26
PCA	**0.87**	**0.65**	**0.62**	**0.93**	0.26
Sammon's mapping	**0.86**	**0.65**	**0.63**	**0.93**	0.26
SOM	**0.87**	**0.64**	**0.62**	**0.93**	0.24
Radviz	0.75	0.42	**0.61**	0.21	0.28
Star Coordinates	**0.80**	0.56	0.51	0.11	0.25

The results also reveal that the poorest projection for Iris dataset is given by Radviz technique (see also Figure 4). By examining in parallel Figures 1-5 with the results showed in Table 1, we observe that the values of the indices are good indicators of the effectiveness of projection techniques in preserving the relationships between data points. It is interesting to observe that in the case of Artificial 2 data (which does not have a clustering structure) the best projection technique that preserves the original distances between the data points is Sammon's mapping. This result confirms the capability of Sammon's mapping of preserving the pair-wise distances of the data points. In this case, SOM technique has the lowest performance.

5.2 Results of Modified Hubert's Γ Statistics for All Datasets

Table 2 shows the values of the normalized modified Hubert's Γ statistic obtained after clustering the data using the K-means algorithm. We used for calculating these indices the procedure 4.2, in which Euclidean distance was chosen for calculating the proximity matrices. The column *Original* contains the values of the indices after applying the clustering on the original dataset. It is observed that for these datasets, the PCA, Sammon's mapping and SOM are the most effective techniques in preserving the inherent clustering structure that exists in the datasets. Because Artificial 2 dataset does not have a clustering structure, the indices have low values. In the case of Wine data, the clustering of the original data does not reveal the true clustering structure of the dataset, and the use of any projection technique may be used to visualize and detect the clusters inherent in the dataset.

The results obtained for all these indices (Tables 1 and 2) are dependent on the selection of a number of parameters. These parameters may regard the projection techniques, the distance metric used for calculating proximity matrices, and the clustering algorithm.

6. CONCLUSION

In this paper, we investigated the use of Hubert's Γ statistic and its modified version for evaluating the effectiveness of five projection techniques (PCA, Sammon's Mapping, SOM, Radviz, and Star Coordinates) in data relationships and structure preservation. We illustrated the use of these projection techniques on the well known Iris dataset, in order to facilitate visual comparison of the capabilities of each projection technique with respect to their effectiveness for solving a data-mining task such as clustering. Because this visual inspection is highly subjective, we investigated the use of objective measures such as Hubert's Γ statistics for the evaluation.

We proposed two procedures for calculating the Hubert's Γ statistics and its modified version in order to assess the extent to which the projected data preserves, first, the inherent relationships and, second, the structure that exists in the dataset. The results showed that our approach can be used to evaluate objectively the data visualizations based on projection techniques. The performance of the projection techniques depends on the dataset under analysis, but generally the PCA, Sammon's mapping and SOM were found the most effective projections for our datasets. The evaluation approach proposed in this paper can be used to assess whether a given projection is good enough in preserving the data relationships and structure, before using that projection for further processing the data. For future work, we intend to use this approach in investigating the capabilities of other projection techniques to represent different high-dimensional datasets.

REFERENCES

Hoffman, P. E., Grinstein, G. G., Marx, K., Grosse, I., and Stanley, E., 1997. DNA Visual and Analytic Data Mining, *Proceedings of IEEE Visualization '97*, Phoenix, Az, 437–441.

Hoffman, P., Grinstein, G., and Pinkney, D., 1999. Dimensional Anchors: a Graphic Primitive for Multidimensional Multivariate Information Visualizations, *Proc. 1999 Workshop on New Paradigms in Information Visualization and Manipulation, in Conjunction with the 8h ACM Int'l. Conf. Information and Knowledge Management (CIKM '99),*pp. 9-16.

Hubert, L. J., and Schultz, J., 1976. Quadratic Assignment as a General Data-Analysis Strategy, *British Journal of Mathematical and Statistical Psychology*, Vol. 29, pp. 190-241.

Kandogan, E., 2001. Visualizing multi-dimensional clusters, trends, and outliers using star coordinates, *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* pp. 107-116.

Kohonen, T., 2001. *Self-Organizing Maps, 3rd* ed. (Berlin; New York: Springer).

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations, *Proc. Fifth Berkeley Symposium on Math. Stat. and Prob., vol. 1*, pp. 281-297.

The MathWorks, 2000. *MATLAB-The Language of Technical Computing* (Natick, MA: The MathWorks, Inc).

Newman, D. J., Hettich, S., Blake, C. L. and Merz, C. J., 1998. UCI Repository of machine learning databases, http://www.ics.uci.edu/~mlearn/MLRepository.html.

Sammon, J. W., 1969. A nonlinear mapping for data structure analysis, *IEEE Trans. Comput., vol. C-18*, pp. 401-409.

Sharma, S., 1995. *Applied Multivariate Techniques* (New York, NY: John Wiley & Sons, Inc.).

Theodoridis, S. and Koutroumbas, K., 1999. *Pattern recognition* (Academic Press, San Diego, CA).

Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J., 1999. Self-organizing map in Matlab: the SOM toolbox, *Proceedings of the Matlab DSP Conference,* pp. 35-40.

Paper 4

Marghescu, D. (2007). Multidimensional Data Visualization Techniques for Exploring Financial Performance Data. In *Proceedings of the 13th Americas' Conference on Information Systems (AMCIS 2007)*, Keystone, Colorado, USA, August 2007.

MULTIDIMENSIONAL DATA VISUALIZATION TECHNIQUES FOR EXPLORING FINANCIAL PERFORMANCE DATA

Dorina Marghescu
Turku Centre for Computer Science, Finland
Åbo Akademi University, Dept. of IT, Finland
dorina.marghescu@abo.fi

Abstract

In this paper, we review nine visualization techniques that can be used for visual exploration of multidimensional financial data. We illustrate the use of these techniques by studying the financial performance of companies from the pulp and paper industry. We also illustrate the use of visualization techniques for detecting multivariate outliers, and other patterns in financial performance data in the form of clusters, relationships, and trends. We provide a subjective comparison between different visualization techniques as to their capabilities for providing insight into financial performance data. The strengths of each technique and the potential benefits of using multiple visualization techniques for gaining insight into financial performance data are highlighted.

Keywords: multidimensional data visualization techniques; financial performance; financial data visualization

Introduction

Many novel visualization techniques have been developed in the fields of information visualization (Card et al. 1999) and visual data mining (Keim 2002). However, the research literature concerning the use of visual data mining for gaining insight into *financial data* is relatively sparse, despite the fact that this technological approach is suitable for both financial data and business users. Financial data are very complex due to their high dimensionality, large volume and diversity of data types. Business users are demanding straightforward visualizations and task-relevant outputs, due to the time and performance constraints under which they work (Kohavi et al. 2002).

In this paper, we review nine visualization techniques that are suitable for representing multidimensional data. The aim is to examine the extent to which they are capable of providing insight into *financial performance data*. In particular, we focus on the problem of financial benchmarking, which is concerned with comparing the financial performance of companies.

The approach consists of the following steps. First, we formulate the financial benchmarking problem in terms of business questions and associated data mining tasks. Second, we investigate the capabilities of each visualization technique in solving the derived data mining tasks and uncovering interesting patterns in data. Third, we compare the visualization techniques from three different perspectives such as: 1) the capability of the techniques to uncover interesting patterns in the data (task fitness); 2) the capability to visualize data items or data models; and 3) the type of data processed (i.e., original data or normalized data).

The analysis highlights the strengths of each technique and the potential benefits of using multiple techniques for exploring financial data. In this paper, we do not address the interactive capabilities of the visualization techniques.

The paper is organised as follows. In the next section, we outline the problem of financial benchmarking, describe the dataset to which we applied the visualization techniques, and derive the business questions and data mining tasks. In Section three, we describe nine multidimensional data visualization techniques and highlight their capabilities for solving the derived data mining tasks. Section four provides a subjective comparison of the techniques and discusses the results. We conclude with final remarks and future work ideas.

The problem of financial benchmarking

One of the problems that business intelligence people are confronted with nowadays is performing comparisons of companies' financial performance. This problem of comparing financial performance of companies is known as *financial competitor benchmarking* (Eklund 2004). The problem is non-trivial since many variables (financial ratios) must be considered. One part of the problem is choosing the ratios to be used when describing the financial performance of a company. Eklund (2004) proposed a model for financial competitor benchmarking in the pulp and paper industry, with seven financial ratios as a basis for companies' performance comparison, and the Self-Organizing Map (SOM) as the method for data analysis. In this paper, we build on the mentioned research to explore the use of other visualization techniques for gaining insight into financial data.

Illustrative Dataset

The dataset analysed in this paper is a subset of a dataset whose collection process including variable and company selection are described by Eklund (2004). The data values are entirely based on the information obtained from companies' financial reports available on the Internet.

The data refer to 80 companies that function in the pulp and paper industry worldwide, observed during 1997 and 1998. A total of 160 observations are analysed. The dataset contains seven numerical variables, namely seven ratios that characterize the financial performance of companies in the pulp and paper industry. The ratios are grouped in four categories: *profitability* (Operating Margin, Return on Equity, and Return on Total Assets), *solvency* (Interest Coverage, Equity to Capital), *liquidity* (Quick Ratio), and *efficiency* (Receivables Turnover). In the following, we use acronyms when referring to any of the financial ratios (that is, OM, ROE, ROTA, IC, EC, QR, and RT respectively). The dataset contains three categorical variables: companies' name, region (Europe, Northern Europe, USA, Canada and Japan), and year (1997 or 1998). The choice of this particular dataset was due to the availability of the dataset, and to its suitability for data mining (e.g., cluster detection, cluster characterization, class characterization, outlier detection, and dependency analysis).

Business Questions and Data Mining Tasks

According to Soukup and Davidson (2002), in order to use information visualization for solving a business problem, the problem should be translated in terms of business questions and further into visualization or data mining tasks. For the problem of financial benchmarking we have derived the business questions and data mining tasks as follows:

a) Outlier detection: Do the data contain outliers or anomalies? Are there any companies that show unusual values of financial ratios?

b) Dependency analysis: Are there any relationships between variables?

c) Data clustering: Are there clusters (groups of companies with similar financial performance) in the data? How many clusters exist?

d) Cluster description: What are the characteristics of each cluster?

e) Class description: Are there any relationships (common features) among companies located in one region or another? What are these common features?

f) Comparison of data items: Compare two or more companies with respect to their financial performance.

For the task f), we have chosen three companies to be compared according to their financial performance in 1998: Reno de Medici, Buckeye Technologies, and Donohue. For Reno de Medici we look also at its evolution from 1997 to 1998. These companies are identified on the graphs using the letters A, B, C, and D, respectively. Table 1 presents the financial ratios of these companies.

Table 1. Financial ratios of the companies chosen for comparison

Company	*Id.*	*Year*	*Region*	*OM*	*ROE*	*ROTA*	*EC*	*QR*	*IC*	*RT*
Reno de Medici 1997	A	1997	Europe	4.02	-15.38	0.64	27.94	1.29	0.15	3.3
Reno de Medici 1998	B	1998	Europe	6.7	5.34	5.27	28.19	1.03	1.68	2.63
Buckeye technologies 1998	C	1998	USA	19.42	38.96	16.21	20.91	1.36	3.28	7.79
Donohue 1998	D	1998	Canada	21.24	17.96	15.92	46.35	0.91	5.15	7.96

Multidimensional data visualization techniques

Because our dataset is tabular data, that is, the rows represent records and the columns represent attributes or dimensions of data, and the data has more than two dimensions, we selected *multidimensional data visualizations* for analysis (Hoffman and Grinstein 2002). The multidimensional data visualization techniques that are reviewed in our paper are multiple line graphs, permutation matrix, survey plot, scatter plot matrix, parallel coordinates, treemap, Principal Components Analysis, Sammon's mapping, and the Self-Organizing Maps. In the following, we apply these visualization techniques on the financial performance data and highlight their capabilities for answering the business questions and data mining tasks formulated in the previous section. Due to page limitations, we are only discussing two to three ratios for each technique. A complete discussion can be found in Marghescu (2007).

Multiple line graphs

Line graphs are used for one dimensional data. On the horizontal axis (Ox) the values are not repeated (e.g., time or the ordering of the table). The vertical axis (Oy) shows the values of the variable of interest. Multiple line graphs can be used to show more than two variables or dimensions (x, y1, y2, y3, etc.).

Figure 1 shows line graphs for two ratios (OM and ROE), observed in 1997 and 1998. The companies are mapped to the horizontal axis, in the order of appearance in the data table. The graph presents companies from different regions (Europe, Northern Europe, USA, Canada and Japan) in different colours, facilitating the characterization of companies from one region or another. By positioning the two years of data one under the other, one can follow the evolution of some company's financial ratios, and make comparisons between companies' financial states.

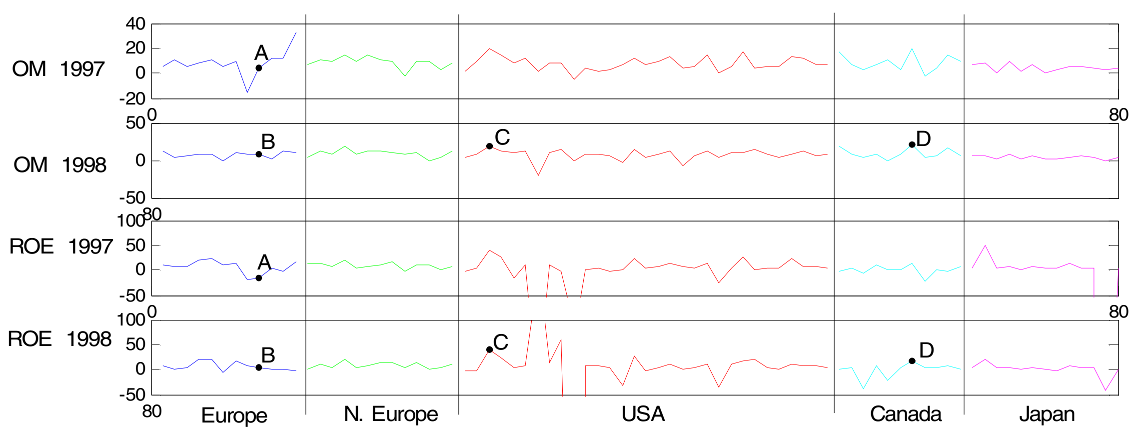


Figure 1 Multiple line graphs

This graph also facilitates the detection of outliers or anomalies in the data, for example, the very low and very high values of ROE for three of the companies, which were further removed from the dataset. By highlighting the companies to be compared, one can see the differences and similarities among them.

Permutation matrix

The permutation matrix is a special type of bar graph described by Bertin (1983). In a permutation matrix, each data dimension is represented by a bar graph in which the heights of the bars represent the data values. The horizontal axes of all bar graphs have the same information (e.g., the time or ordering of the data table). The below average data values are coloured black, and the above average data values are coloured white. A green dashed line plotted over the data represents the average value of each dimension. Implementations of permutation matrixes allow the interactive changing of the order of the records for observing interesting patterns.

Figure 2 displays a permutation matrix created with Visulab (Hinterberger and Schmid 1993). On the horizontal axes the companies are arranged in descending order of ROTA. The companies of interest are highlighted. This graph facilitates the detection of relationships between ratios and the comparison of companies. It also reveals anomalies in the data.

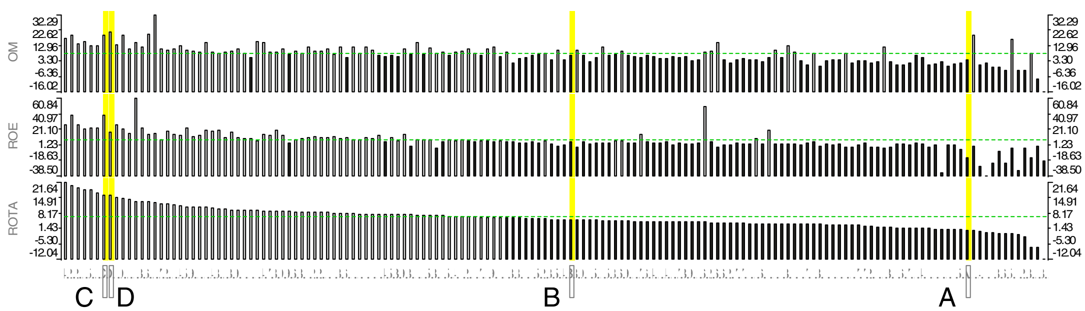


Figure 2 Permutation matrix created with Visulab

Survey plot

The survey plot is a variation of the permutation matrix. The values of each data dimension are represented as horizontal bars. The width of the bars is proportional to the data values. The bars are centred and there are no spaces separating the bars. One can use colours to distinguish between different classes in the data (if a class variable is present).

Figure 3 displays a survey plot, in which the data are sorted according to ROTA. This facilitates the detection of relationships between ROTA and other ratios, for example OM, ROE and IC.

Companies from different regions are displayed with different colours. The graph shows that the Japanese companies are not among the most profitable ones, while the American and European companies display the highest profitability. The technique facilitates the detection of outliers and comparison between two or more companies.

Scatter-plot matrix

A scatter plot is used to plot two dimensional data so that the horizontal axis shows the values of one variable and the vertical axis shows the values of another variable. The scatter-plot matrix is useful for looking at all possible pairs of variables in the dataset.

Figure 4 displays a scatter plot matrix for three financial ratios (OM, ROE and RT). The plots reveal relationships between ROE and OM. The visualization also reveals outliers, and facilitates the comparison of companies.

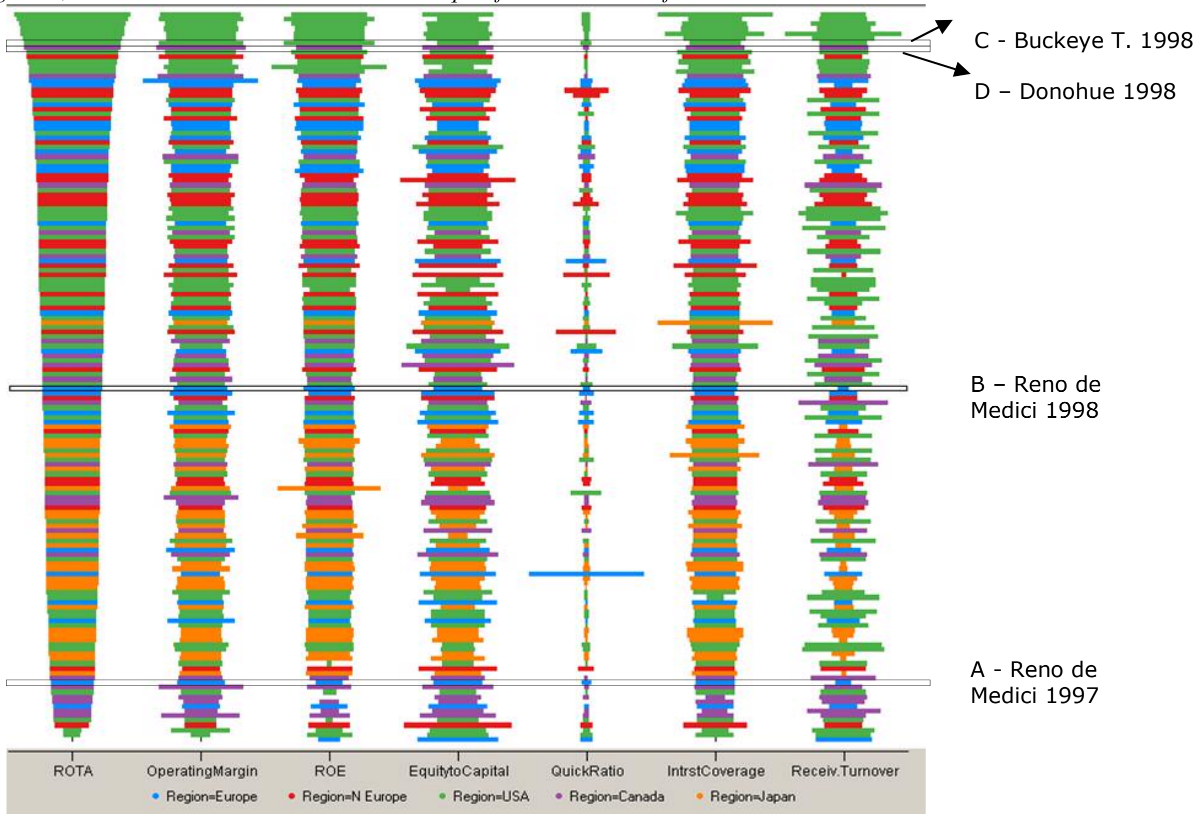


Figure 3 Survey plot created with Orange (Demsar 2004)

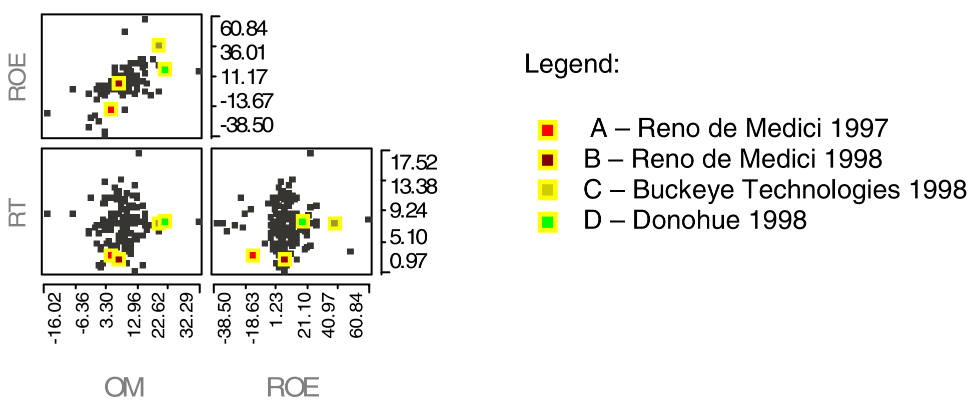


Figure 4 Scatter-plot matrix created with Visulab

Parallel coordinates

Introduced by Inselberg (1985), parallel coordinates represent multidimensional data using lines. The data dimensions are represented as parallel axes (coordinates). The maximum and minimum values of each dimension are scaled to the upper and lower points on a vertical axis. An n-dimensional data point is displayed as a *polyline* that crosses each axis at a position proportional to its value for that dimension.

Figure 5 represents the financial ratios as parallel axes and each company as a polyline that crosses each axis at a point proportional to the value of the ratio for the corresponding company. The companies of interest are highlighted using different colours.

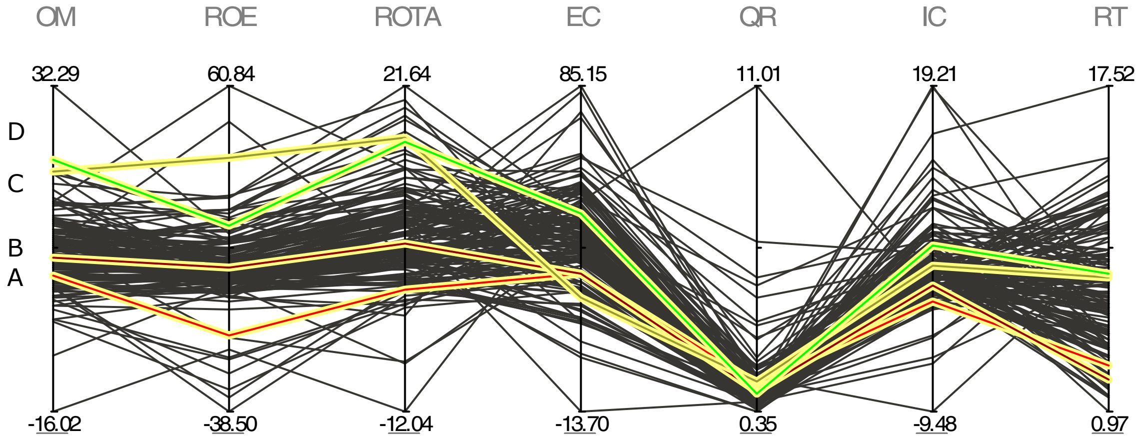


Figure 5 Parallel coordinates created with Visulab

The display facilitates detection and characterization of outliers. One can compare the financial performance of different companies. The relationships between two or more variables can be detected if the correlated variables are arranged consecutively (for example, ROE and ROTA).

Treemaps

The treemaps (Johnson and Shneiderman 1991) are hierarchical visualizations of multidimensional data. Data dimensions are mapped to the size, position, colour, and label of nested rectangles.

Figure 6 displays the dataset using the treemaps technique. The figure was created with Treemap 4.1 (2004). Each company is represented by a rectangle. The size of the rectangle indicates the value of RT. The colour of the rectangle indicates the value of the ROTA ratio as follows: light green indicates high values of ROTA; light red indicates small values of ROTA; dark red and dark green shows values of ROTA close to 14 (see the “colour binning” panel in the visualization below). In this visualization, the dataset is organised into categories such as year and region.

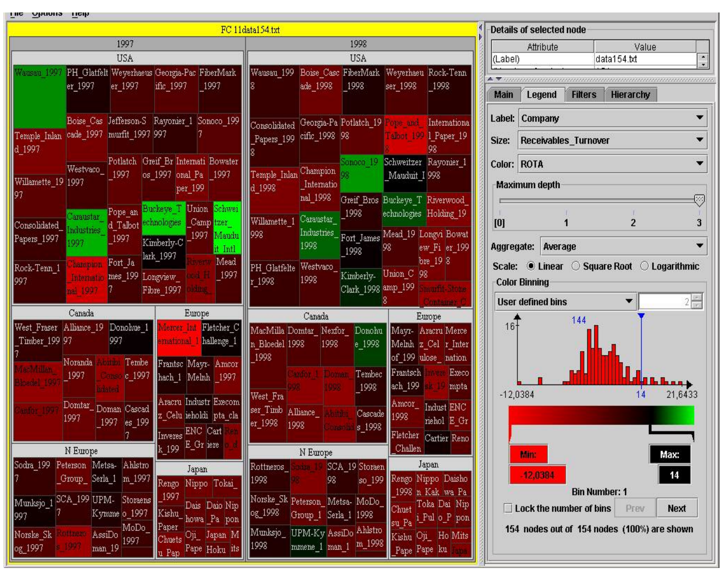


Figure 6 Treemap created with Treemap 4.1

This treemap representation shows where the most profitable companies in terms of ROTA are located, and how the companies of interest have evolved over time. In addition, one can identify common features or patterns in the industry, for example, that Japanese companies have the lowest values of the efficiency ratio. One can also compare the financial performance of different companies.

Principal component analysis (PCA)

PCA is a dimensionality-reducing technique employing *linear transformation of data* (Sharma 1995). The projection of high-dimensional data onto a lower-dimensional space tries to preserve the variance of the original data as well as possible. The PCA technique creates new variables (called principal components), which are linear composites of the original variables and are uncorrelated amongst themselves. The maximum number of new variables that can be formed is equal to the number of original variables. The PCA output is judged in terms of how well the new variables represent the information contained in data, or, geometrically, how well the new dimensions can capture the original configuration of the data.

Figure 7 shows PCA plot that was constructed from the standardized dataset. The red dot shows the observation closest to the centre of the dataset. The companies of interest are marked with a yellow star and labelled on the graph.

One can interpret the principal components by inspecting the loadings of each original variable to the PCs. The higher the loading of a variable, the more influence it has in forming the PC score and vice versa. In our case, the first PC (horizontal axis) is highly correlated with the profitability ratios and the IC ratio. Therefore, companies placed towards the right of the horizontal axis, have high values in profitability and IC. The second PC (vertical axis) is highly correlated with QR and EC. Companies located on the upper part of the graph have a high liquidity and high solvency. The amount of variation explained by the two PCs is 40.926% + 19.455%= 60.38% of the total variance. While this amount of variance accounts for the variation of six of the ratios, it does not consider the variation of efficiency (RT) among the companies.

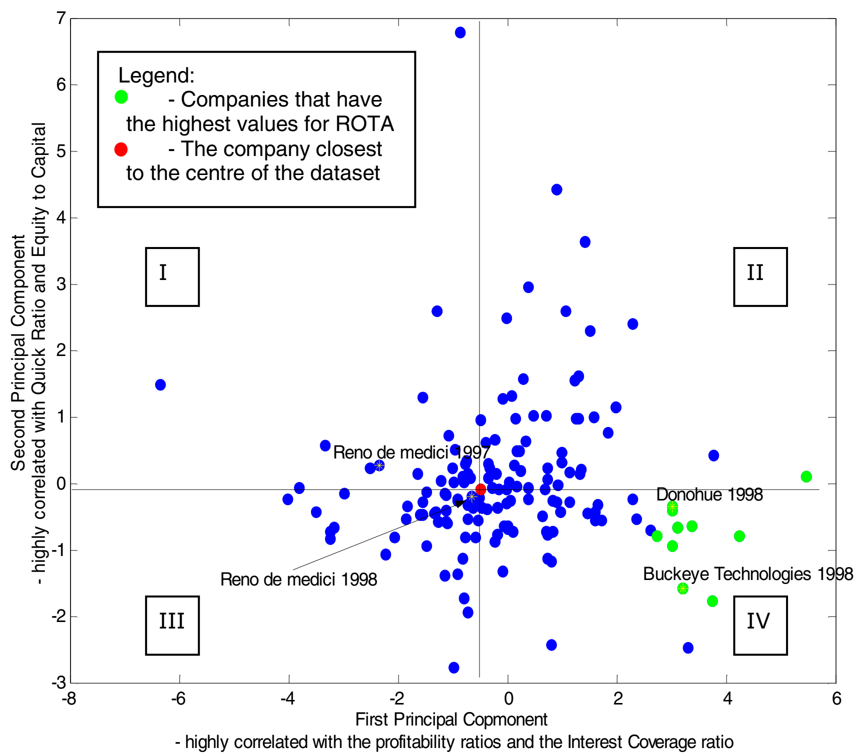


Figure 7 Data projected on the first two PCs created with the Statistics Toolbox (The MathWorks 2002). In area I: medium-high liquidity, low-medium profitability; II: medium-high liquidity, solvency and profitability; III: low-medium liquidity, solvency and profitability; IV: low-medium liquidity, medium-high profitability

In addition to its usefulness as a data reduction method, PCA is also useful in finding numerous patterns in data (Figure 7). The graph shows the high profitability of Buckeye Technologies 1998 and Donohue 1998, and the increase in profitability for Reno and Medici in 1998. The high correlation of the first PC with all profitability ratios and with the IC ratio indicates that there also exists a relationship between the profitability ratios and IC. Similarly, the high correlation of the second PC with EC and QR indicates that EC and QR are also correlated.

By splitting the visual representation in four areas by two orthogonal lines that intersect in the centre of the dataset, one can divide the dataset into four groups of similar observations as shown and described in Figure 7. Based on the meaning of the first two PCs, one can conclude that in area I there are companies with medium-high liquidity and low-medium profitability; in area II, companies with medium-high liquidity, solvency and profitability; in area III, companies with low-medium liquidity, solvency and profitability; and in area IV, companies with low-medium liquidity but medium-high profitability. Based on this evaluation, one can compare the financial performance of the companies of interest.

Sammon's mapping

Sammon's mapping is a *nonlinear projection* of the multidimensional data down to two dimensions so that the distances between data points are preserved (Kohonen 2001). It belongs to multidimensional scaling techniques.

Figure 8 illustrates our financial dataset using Sammon's mapping. The data values were normalized using the discrete histogram equalization method. The normalization method works in two steps: first, the data values of each variable are replaced by the order index, and then these values are normalized to be in the range [0, 1], by applying a linear transformation. Companies from different regions are displayed using different colours. The companies of interest are marked with yellow stars and labelled on the graph.

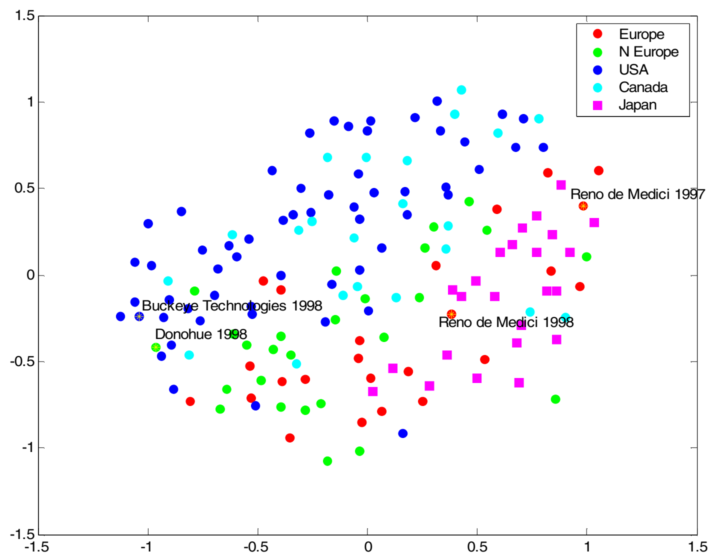


Figure 8 Sammon's mapping created with SOM Toolbox 2.0 (2005)

The technique is useful in visualizing class distributions, especially the degree of their overlap. One can see that companies from Canada and USA overlap and map to the same area of the graph, whereas Japan, Europe and Northern Europe form three separate groups. However, the degree of overlapping between all these classes is quite high; especially Europe and Northern Europe do not separate well from the other groups. The differences and similarities between the companies are easy to distinguish, but not easy to interpret.

Self-Organizing Maps (SOM)

The SOM technique, developed by Kohonen (2001) is a special type of neural network based on unsupervised learning. The SOM algorithm is similar to the K-Means clustering algorithm, but the output of a SOM is topological and neighbouring clusters are similar. As a projection technique of multidimensional data onto a two-dimensional grid, the SOM method is similar to multidimensional scaling techniques, such as Sammon's mapping. The grid consists of units that have assigned reference vectors with the same dimensionality as the original data. After learning is complete, the reference vectors are updated such that they resemble most of the data items, as much as possible. Each data item is then mapped to the unit where the highest similarity between the reference vector and the data item is calculated. Multiple data items mapped onto the same unit are similar and form a cluster.

We have used the SOM technique on normalized data obtained byapplying the d iscrete histogram equalization method. There are many ways to represent the SOM output. One way to represent the data is to use the *scatter plot* technique (usually with jittering), in which the horizontal and vertical axes are produced by the Kohonen network (i.e., the map size, in our case 6x5 units).

Figure 9 is a scatter plot of the dataset based on the SOM coordinates. Companies from different regions are highlighted using different colours. The technique of jittering was used in order to change with a small value the position of each company; otherwise the companies mapped to the same unit would have overlapped. Figure 9 shows many clusters in the data (if more companies are mapped to the same map unit, they may be interpreted as forming a cluster). One can also observe some isolated companies. However, the interpretability of this map is not easy. One can distinguish among the companies belonging to the same region, or identify the placement of these companies on the map but cannot interpret these classes or the clusters formed.

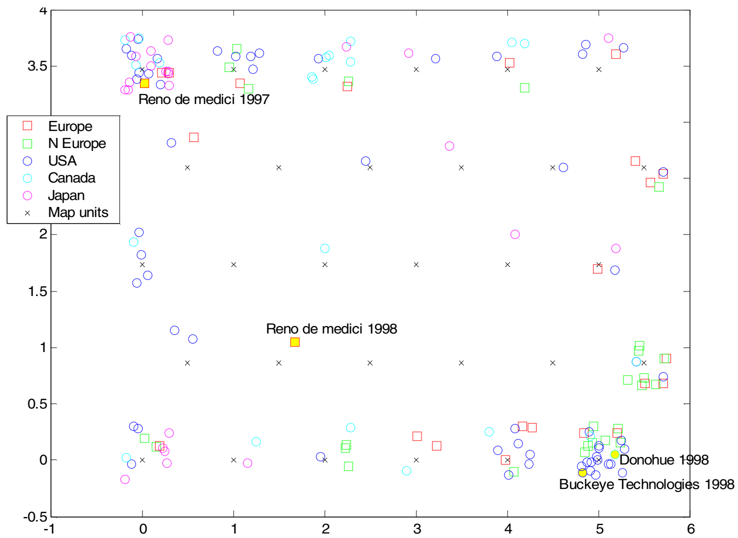
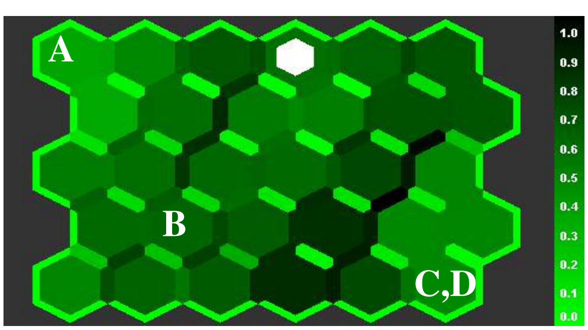


Figure 9 Self-Organizing Map – scatter plot view created with SOM Toolbox 2.0

Ultsch and Siemon (1989) developed the *U-matrix* graphic display to illustrate the clustering of the reference vectors, by representing graphically the distances between map units. In this visual representation, each map unit is typically represented by a hexagon. The line or border between two neighbouring map-units (hexagons) has a distinguishable colour that signifies the distance between the two corresponding reference vectors. Dark green signifies for large distances, and light green signifies similarities between the vectors, as indicated by the colour bar (Figure 10-Left).



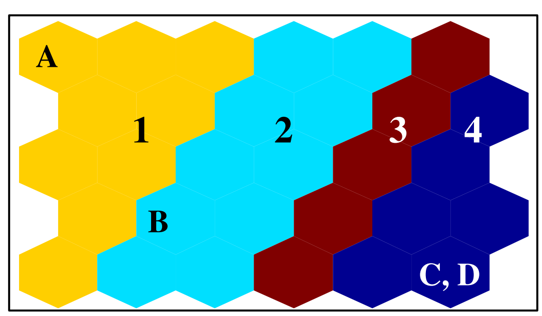


Figure 10 Left: Self-Organizing Map - U-matrix view created with Nenet 1.1 (1999); Right: Clustering of SOM view created with SOM Toolbox 2.0

By looking at the borders' colours in Figure 10-Left, one can distinguish the main clusters that exist in the data. A clustering algorithm (e.g., K-means) can be used to automatically partition the map into similar clusters (Figure 10-Right), creating the *clustering of SOM* view. The dataset appears to contain four main clusters. Based on Figure 9 and Figure 10, one can compare the companies of interest with respect to their membership in the identified clusters. Moreover, one can see the composition of each cluster with respect to the variable Region (e.g., Cluster 4 contains mostly American, Northern European and European companies).

It is also possible to visualize each data dimension using *feature planes*. These represent graphically the levels of the variables in each map unit. The colour red signifies high values of the variables, and blue and black correspond to low values of the variables (as indicated by the colour bars, Figure 11).

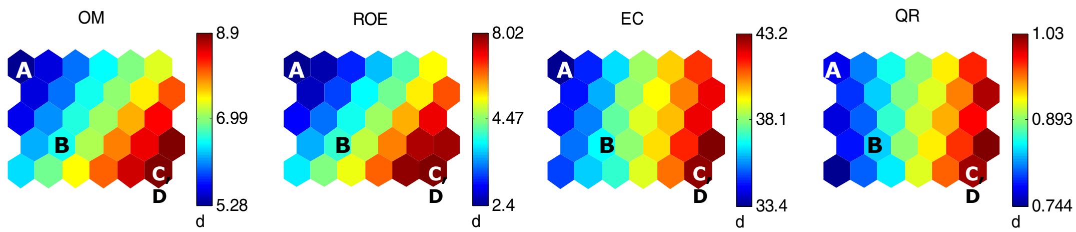


Figure 11 Feature planes created with SOM Toolbox 2.0

The feature planes facilitate the comparison of the companies of interest. The feature planes also help to identify the relationships between variables (e.g., OM is correlated with ROE; EC is correlated with QR).

By examining the features planes in parallel with the clustering of the SOM, one obtains the description of the four clusters identified previously as follows. Cluster 1 shows very low profitability, liquidity, solvency and efficiency. It contains the companies with the poorest financial performance. Reno de Medici 1997 is situated in this cluster (A). Cluster 2 shows medium profitability, solvency, and liquidity, but low efficiency. Reno de Medici 1998 belongs to this cluster (B). Cluster 3 shows good profitability, liquidity and solvency. Efficiency is medium to low. Cluster 4 shows very high profitability, solvency, liquidity and efficiency. It contains the companies with the best financial performance, among which Buckeye Technologies 1998 and Donohue 1998 (C and D) are situated.

Comparison of visualization techniques

In the previous section, we illustrated the use of multidimensional data visualization techniques for exploring financial performance data. All visualization techniques used are capable of providing an overview of the dataset under analysis, and different techniques uncover different patterns in the data. We highlighted the capabilities of each technique for answering the business questions and data mining tasks related to the financial benchmarking problem. In this section, we compare the techniques with respect to three criteria: 1) their capabilities to answer the questions and data mining tasks formulated for the financial benchmarking problem; 2) their capabilities to show data items or data models; and 3) the type of data used as input for the visualization technique (i.e., original data or normalized data).

First, we provide in Table 2 a *subjective comparison* of the techniques with respect to their capabilities for solving the data mining tasks related to financial benchmarking. The assessment concerns only this business problem and the associated dataset. We do not intend to generalize the results to other datasets, because for a different dataset (with different types of data, number of variables, number of observations, underlying structure) the results of the evaluation could be different. Table 2 can be used as a means to map the data mining tasks to different visualization techniques for this dataset. This table can, therefore, be used in the process of selection of visualization techniques suitable for representing and exploring the data in the financial benchmarking problem.

Table 2 shows that there are data mining tasks for which more than one visualization technique can be used. On the other hand, one data mining task may be addressed using different visualization techniques but with a different outcome (e.g., clustering solutions produced by the SOM and PCA). Almost all visualization techniques can facilitate the comparison among companies. Moreover, all techniques, except Sammon's mapping, are effective for finding outliers or anomalies in this dataset. The scatter plots, survey plot, permutation matrix, PCA and the SOM are good in showing relationships between ratios. The SOM and PCA are capable of showing and describing clusters. The treemap, line graphs, and survey plot are capable of describing class characteristics. Treemap is typically effective in displaying hierarchical data, and in our example proved to be very effective in making comparisons between companies and highlighting the characteristics of companies from one region or another with respect to the values of financial ratios. Sammon's mapping is effective in displaying class distribution, but does not provide means to describe the characteristics of the classes.

Table 2 also shows that the most effective techniques in uncovering patterns in this specific dataset are the SOM (when all views are analysed together) and PCA. If we assess separately each SOM-based visualization technique, the results show that all SOM views show the clustering of the data, but the other patterns are uncovered to a different extent by each SOM view.

Table 2. The capabilities of the visualization techniques on the dataset under analysis

Visualization technique	*Outliers detection*	*Dependency analysis*	*Clustering*	*Cluster description*	*Class description*	*Comparison*
Line graphs	✓	✓			✓	✓
Permutation matrix	✓	✓				✓
Survey plot	✓	✓			✓	✓
Scatter plot matrix	✓	✓				✓
Parallel coordinates	✓	✓				✓
Treemaps	✓				✓	✓
PCA	✓	✓	✓	✓		✓
Sammon's mapping					✓*	
SOM – scatter plot	✓		✓		✓**	
SOM – U-matrix			✓			
SOM – clustering			✓			
SOM – feature planes		✓	✓	✓		✓
SOM – all views combined	✓	✓	✓	✓	✓	✓

* Sammon's mapping is capable of organizing the dataset so that different classes are distinguishable but does not provide a means to interpret and describe the classes.
** SOM – scatter plot view is capable of showing where the companies from different classes (regions) are mapped but does not provide a means to interpret and describe the classes.

Second, the visualization techniques are compared with respect to their capability for showing data items or data models. All techniques display the data items. The SOM displays also a data mining model (e.g., the clustering of the data). In the former case, the user has to use his/her perceptual abilities to distinguish the patterns of interest. In the later case, the data model is automatically generated and displayed by the visualization.

Third, the visualization techniques are compared with respect to the type of data processed. The following visualization techniques represent the original data: multiple line graphs, permutation matrix, survey plot, scatter plot, parallel coordinates, and treemap. The other techniques represent standardized or normalized data: PCA, Sammon's mapping, and SOM. The visualizations obtained using standardized or normalized data are more difficult to interpret.

In summary, the techniques reviewed in this paper complement each other in uncovering all patterns in the financial benchmarking dataset. Using a single technique for data exploration may result in a limited understanding of the data. Therefore, the use of multiple techniques could be beneficial for the user. For example, combining different visualizations that are based on the SOM, we obtain a very good understanding of the data, while if we consider only one view, we understand very little about the data. One benefit of using multiple visualizations is that one can see different facets of the data and problem under investigation by using visualizations that uncover distinct patterns. Another benefit is that the analyst has the possibility to confirm that the patterns or outliers highlighted by one visualization technique are indeed real, and not an artefact, thereby increasing confidence in the findings. A third benefit is given by the descriptive power of some techniques over the others.

Conclusion

In this paper, we reviewed nine multidimensional data visualization techniques for representing financial performance data. We investigated the capabilities of different visualization techniques for uncovering interesting patterns in financial performance data, patterns described in terms of outliers, clusters, classes, relationships and trends.

By deriving business questions and data mining tasks from the financial benchmarking problem as recommended by Soukup and Davidson (2002), and mapping these tasks to appropriate visualization techniques, we provided a means to subjectively compare and assess the capabilities of different visualization techniques to solve the financial benchmarking problem. This approach can serve visual data mining systems' developers in assessing the strength of various techniques in the early stage of system development and accordingly select the most appropriate techniques. The approach can be extended by involving users in further evaluation studies of the selected visualization techniques.

We highlighted the potential benefits of using multiple visualization techniques for solving a business problem such as the financial benchmarking problem and uncovering all interesting patterns in data. Empirical studies of users facing multiple visualizations are needed to quantify these benefits. The study can also be extended by analysing other financial datasets and/or other multidimensional data visualization techniques.

Acknowledgements

I gratefully acknowledge the contributions of Professor Barbro Back and Dr. Tomas Eklund, and the financial support of Tekes (grant number 40435/05) and the Academy of Finland (grant number 104639).

References

Bertin, J., Semiology of Graphics: Diagrams, Networks, Maps, (W.J. Berg, Trans.), Madison, WI: University of Wisconsin Press, 1967/1983.

Card, S.K., Mackinlay, J.D., and Shneiderman, B., Readings in Information Visualization - Using Vision to Think, San Francisco, CA: Morgan Kaufmann Publishers, 1999.

Demsar J., Zupan B., and Leban G., "Orange: From Experimental Machine Learning to Interactive Data Mining", White Paper (www.ailab.si/orange), Faculty of Computer and Information Science, University of Ljubljana, 2004.

Eklund, T., The Self-Organizing Map in Financial Benchmarking, TUCS Doctoral Dissertation, Åbo Akademi University, Turku, 2004.

Hinterberger, H. H., and Schmid, C., "Reducing the Influence of Biased Graphical Perception with Automatic Permutation Matrices", SoftStat '93, Proceedings of the Seventh Conference of the Scientific Use of Statistic-Software, March 1993.

Hoffman, P. E., and Grinstein, G. G., "A Survey of Visualizations for High-Dimensional Data Mining", in Fayyad, U., Grinstein, G. G., and Wierse, A. (eds.), Information Visualization for Data Mining and Knowledge Discovery, San Francisco: Morgan Kaufmann, 2002.

Inselberg, A., "The Plane with Parallel Coordinates", Special Issue on Computational Geometry, The Visual Computer 1, 1985, pp. 69–91.

Johnson, B., and Shneiderman, B., "Treemaps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures", Proc. of IEEE Visualization '91 Conf., San Diego, 1991, pp. 284 291.

Keim, D.A., "Information Visualization and Visual Data Mining", IEEE Transactions on Visualization and Computer Graphics, Vol. 8, No. 1, January-March 2002.

Kohavi, R., Rothleder, N. J., and Simoudis, E., "Emerging Trends in Business Analytics", Communications of the ACM, 45(8), 2002, pp. 45-48.

Kohonen, T., Self-Organizing Maps, Third edition, Berlin: Springer-Verlag, 2001.

Marghescu, D., Multidimensional Data Visualization Techniques for Financial Performance Data: A Review, TUCS Technical Report no. 810, Turku Centre for Computer Science, Turku, Finland, February 2007.

Nenet 1.1, http://koti.mbnet.fi/~phodju/nenet/Nenet/General.html, 1999.

Sharma, S., Applied Multivariate Techniques, New York, NY: John Wiley & Sons, 1995.

SOM Toolbox 2.0 for MATLAB, http://www.cis.hut.fi/projects/somtoolbox/, 2005.

Soukup, T., and Davidson, J., Visual Data Mining: Techniques and Tools for Data Visualization and Mining, New York, NY: John Wiley & Sons, 2002.

The MathWorks, Statistics Toolbox for Use with MATLAB - User's Guide, Version 4, 2002.

Treemap 4.1, http://www.cs.umd.edu/hcil/treemap/, 2004.

Ultsch, A., and Siemon, H., Technical Report 329, Univ. of Dortmund, Dortmund, Germany, 1989.

Paper 5

Marghescu, D. (2007). User Evaluation of Multidimensional Data Visualization Techniques for Financial Benchmarking. In *Proceedings of the European Conference on Information Management and Evaluation (ECIME 2007),* Montpellier, France, September 2007.

User Evaluation of Multidimensional Data Visualization Techniques for Financial Benchmarking

Dorina Marghescu
Åbo Akademi University, Turku, Finland
dmarghes@abo.fi

Abstract: In this paper, we investigate the use of information visualization techniques for getting insight into multidimensional financial data. In particular, we focus on exploring different multidimensional data visualization techniques with respect to their effectiveness in solving a financial problem, namely financial competitor benchmarking. Financial competitor benchmarking is concerned with comparing the financial performance of different companies competing in the same market, industry, country or region. We investigate the extent to which different multidimensional visualization techniques are effective in revealing interesting patterns in financial performance data. For this purpose, we conducted a user evaluation study in which nine multidimensional data visualization techniques were assessed. The assessment concerns the extent to which users of these techniques are capable of discovering interesting patterns in multidimensional financial data, patterns associated with the problem of financial benchmarking. These patterns are identified as outlier detection, clustering, cluster description, class description and data comparison. The visualization techniques under analysis are: multiple line graphs, permutation matrix, survey plot, scatter plot matrix, parallel coordinates, treemap, principal components analysis, Sammon's mapping and the Self-Organizing Maps. The evaluation method consists in questionnaire-based data collection and analysis. We obtained answers from 12 students who agreed upon participating in this study. The evaluation we have conducted is useful especially in the early stage of the development of a visualization system, because it helps in the process of selection of most appropriate techniques for solving certain tasks.

Keywords: User evaluation, multidimensional data visualization, financial data visualization

1. Introduction

In this paper, we investigate the use of information visualization techniques for getting insight into multidimensional financial data. In particular, we focus on exploring the effectiveness of multidimensional data visualization techniques in solving a financial problem, namely *financial competitor benchmarking*. Financial competitor benchmarking is concerned with comparing the financial performance of different companies competing in the same market, industry, country or region. Simply referred to as *financial benchmarking*, this problem is complex because many variables (financial ratios) must be considered. One part of the problem is choosing the ratios to be used when describing the financial performance of a company. Eklund (2004) proposed a model for financial benchmarking, in pulp and paper industry, with seven financial ratios as a basis for companies' performance comparison. Moreover, he studied the use of the Self-Organizing Maps (SOM) as the method for data analysis and visualization. Marghescu (2007) explored the use of other multidimensional data visualization techniques for the financial benchmarking problem and provided a subjective comparison of nine techniques, including the SOM.

In this paper, we extend the mentioned research to evaluate the nine multidimensional data visualization techniques investigated in (Marghescu 2007) with respect to their effectiveness in providing insight into financial performance data. To achieve this objective, we conducted a user evaluation study in which 12 participants were involved. The assessment concerns the extent to which users of these techniques are capable of discovering interesting patterns in multidimensional financial data, patterns associated with the problem of financial benchmarking. For the data collection, we used a questionnaire. We focus on those visualization techniques that are suitable for representing graphically *tabular data*, that is, datasets that are expressed as data tables in which the rows represent cases or records and the columns represent attributes or dimensions of data. Hoffman and Grinstein (2002) use the term "table visualizations" to refer to this class of visualization techniques. Moreover, we focus on table visualization techniques that are capable of displaying multidimensional or multivariate data. These are referred to in the literature as *multidimensional data visualizations* (Hoffman and Grinstein 2002) or *multidimensional visualizations* (Soukup and Davidson 2002, p. 208). The paper is organized as follows. Section 2 reviews related work on evaluation of information visualization techniques. Section 3 describes the dataset, tasks and visualization techniques used in evaluation. Section 4 describes the evaluation method, including the participants and the data collection and analysis. Section 5 describes the results of the evaluation. Section 6 discusses the findings and the evaluation approach. We conclude with final remarks and future work ideas in Section 7.

2. Related work

Many researchers (Chen and Czerwinski 2000, Plaisant 2004, Picket and Grinstein 2002) emphasize the need for systematic approaches to evaluation of data and information visualization. The studies about information visualization evaluation differ with respect to, on the one hand, the methods used for data collection and analysis, and, on the other hand, the focus of the studies, that is, what exactly is evaluated. For example, some studies focus on evaluating the *performance* of users in solving business tasks (Dull and Tegarden 1999). Others address the *ease of use* of the system (Risden and Czerwinski 2000), and the *effectiveness* of visualization techniques *in different task-domains* such as information retrieval (Sutcliffe et al. 2000), visual data mining (Marghescu and Rajanen 2005, Liu and Salvendy 2007), or depicting hierarchical structures (Stasko et al. 2000).

Besides the focus of evaluation, these studies can be classified based on the types of measures (qualitative vs. quantitative) or approach (subjective vs. objective) employed. However, there are studies that employ both subjective and objective approaches or use both quantitative and qualitative measures for assessment (e.g., Liu and Salvendy 2007). Objective evaluation is concerned with assessing the effectiveness of a technique in the absence of the human viewer. The criteria used in objective evaluation are directly measurable, such as the number of variables or volume of data that can be displayed. Other measures in objective evaluation reflect the performance or effectiveness in solving specific tasks. In classification tasks, the measures of performance are defined in terms of accuracy of classification (Liu and Salvendy 2007), in clustering tasks the measures can be in terms of internal and external validity of the clusters (Marghescu 2006), and in information retrieval, the performance measures are defined in terms of recall and precision (Sutcliffe et al. 2000). On the other hand, subjective evaluation involves a human viewer (an expert or user) in the process of assessment and is more difficult to quantify (e.g., Marghescu and Rajanen 2005; Liu and Salvendy 2007). According to many authors, both objective and subjective studies are important for evaluating visualization techniques (Keim 1996).

Moreover, according to a review of evaluation practices (Plaisant 2004), an important category of studies are based on controlled experiments. These studies focus on comparing two or more tools or on comparing different design elements or visualization factors and their impact on the user (e.g., Somervell et al. 2002). In the next sections, we present the evaluation approach that we used in this study. The approach is characterized by the following: it is based on user assessment (i.e., subjective approach), and it focuses on assessing the effectiveness of each technique in uncovering interesting patterns in a particular multidimensional financial dataset. The assessment does not consider the interactive capabilities of the techniques, but their *effectiveness* in uncovering interesting patterns in data. Card et al. (1999) say about a visual mapping that is effective if it can be perceived well by a human, i.e., it is fast to interpret, can convey many distinctions, or does not lead to many errors in interpretation. The effectiveness can be judged individually for each visualization, or by comparing different visualizations. In our approach, we used the first alternative, that is, each visualization was assessed individually. We choose not to conduct a comparative study, which would have implied a controlled experiment, due to the relatively small number of participants compared to a rather large number of visualization techniques to be assessed in the study.

3. Materials

3.1 Dataset and tasks

The evaluation was carried out on a dataset concerning 80 companies that function in the pulp and paper industry worldwide, observed during 1997 and 1998. A total of 160 observations are analysed. These data are a subset of a larger dataset whose collection process including variable and company selection is described in (Eklund 2004). The dataset contains seven numerical variables, namely seven ratios that characterize the financial performance of companies in the pulp and paper industry. The ratios are grouped in four categories: *profitability* (Operating Margin, Return on Equity, and Return on Total Assets), *solvency* (Interest Coverage, Equity to Capital), *liquidity* (Quick Ratio), and *efficiency* (Receivables Turnover). Besides the numerical variables, the dataset contains three categorical variables: companies' name, region (Europe, Northern Europe, USA, Canada and Japan), and year (1997 or 1998). According to Soukup and Davidson (2002, p. 49), in order to use information visualization in solving a business problem, this should be translated in terms of business questions and further in visualization or data mining tasks. For the problem of financial benchmarking we have derived the business questions and data mining tasks as follows:

1. Outlier detection: Does the data present outliers or anomalies? Are there any companies that present unusual values of financial ratios?
2. Dependency analysis: Are there any relationships between variables?
3. Data clustering: Are there clusters (groups of companies with similar financial performance) in the data? How many clusters do exist?
4. Cluster description: What are the characteristics of each cluster?
5. Class description: Are there any relationships (common features) among companies located in one region or another? What are these common features?
6. Comparison of data items: Compare two or more companies with respect to their financial performance.

For the task f), we have chosen three companies to be compared as to their financial performance in 1998: Reno de Medici, Buckeye Technologies, and Donohue. For Reno de Medici we look also at its evolution from 1997 to 1998. These companies are identified on the graphs using the letters A, B, C, and D, respectively. Table 1 presents the financial ratios of these companies.

Table 1: Financial ratios of the companies chosen for comparison

Company	Reno de Medici 1997	Reno de Medici 1998	Buckeye technologies 1998	Donohue 1998
Id.	A	B	C	D
Year	1997	1998	1998	1998
Region	Europe	Europe	USA	Canada
OM	4.02	6.7	19.42	21.24
ROE	-15.38	5.34	38.96	17.96
ROTA	0.64	5.27	16.21	15.92
EC	27.94	28.19	20.91	46.35
QR	1.29	1.03	1.36	0.91
IC	0.15	1.68	3.28	5.15
RT	3.3	2.63	7.79	7.96

3.2 Multidimensional data visualization techniques

Because the dataset is tabular data, we selected for evaluation multidimensional data visualization techniques that are suitable for representing this type of data. These techniques and the tools used for their implementation are as follows:

- Multiple line graphs created with Matlab (The MathWorks 2000) – Figure 2,
- Permutation matrix created with Visulab (Hinterberger and Schmid 1993) – Figure 3,
- Survey plot created with Orange (Demsar et al. 2004) – Figure 4,
- Scatter plot matrix created with Visulab – Figure 5,
- Parallel coordinates created with Visulab – Figure 6,
- Treemap created with Treemap 4.1 (2004) – Figure 7,
- Principal Components Analysis created with Statistical Toolbox for Matlab (The MathWorks 2002) – Figure 8,

- Sammon's mapping created with SOM Toolbox 2.0 (2005) – Figure 9,
- SOM created with SOM Toolbox and Nenet (1999) – Figures 10-13.

4. Evaluation method

4.1 Participants

A number of 12 students agreed upon participating in this study. The participants were students with international background at a public university. The respondents have different majors (Information Systems, Computer Science, Economics, and Business Administration). The users were recruited from one advanced-level course in information systems. They were familiar with the dataset and the financial benchmarking problem.

4.2 Data collection

The evaluation study consisted of the following parts:

- Introduction of the study to participants,
- Data collection based on the questionnaire technique, and
- Data analysis.

In the introduction, the participants were given background information about the study, as well as instructions to participate. The background information included the following:

- Short tutorial about information visualization,
- Short description of data mining tasks , and
- Short descriptions of the nine visualization techniques used in the study.

The introductory part of the questionnaire consisted of the descriptions of the data, business problem, and visualization techniques. After introducing a visualization technique, the questionnaire presented the corresponding graphical representation of the dataset, followed by questions asking whether or not certain patterns are identifiable on the visualization. The visualization techniques were presented to all participants in the same order (as given also in Section 3.2). The patterns required to be identified were directly related to data mining tasks (a-f) identified for the financial benchmarking problem (Section 3.1). The purpose of the questions was to assess the extent to which users of the visualization techniques are able to identify and describe interesting patterns in data. To avoid bias, the same questions were asked for each visualization technique. In this way, we ensured that the respondent is not inclined to answer positively to all the questions posed for one technique. To ensure the validity of the responses (i.e., the user has understood the questions, the visualization and the patterns) the questionnaire included open questions asking to illustrate on the graph or to explain the patterns identified. For illustrating the contents of the questionnaire, Figure 1 presents the questions asked for the assessment of the permutation matrix. Similar questions were asked for all the other visualization techniques.

4.3 Data analysis

The analysis of the collected data yielded four categories of answers:

- *Positive* answers: The user has answered positively (i.e., YES), that he can identify certain patterns, and the explanation or illustration of the patterns identified were correct.
- *Negative* answers: The user has answered negatively (i.e., NO), that he could not identify certain patterns.
- *Invalid* answers: We defined as invalid answers those positive answers that were accompanied by incomplete explanations or mistakes.
- *Non-responses*: This category includes the non-responses to specific questions.

We interpreted the number of positive answers of the participants as the extent to which the visualization is effective in uncovering those patterns. The number of negative answers indicates the extent to which the visualization is not effective in highlighting the patterns of interest. The invalid answers might have been determined by the misunderstanding of the question as well as the misunderstanding of the visualization or patterns' definitions. Therefore these are more difficult to interpret. However, a very small number of invalid answers and non-responses was recorded.

- Can you identify any **outliers** in this dataset by examining the above permutation matrix? – YES/ NO
- If yes, please tell in brief for which ratio you identified an outlier or mark it on the graph.
- Can you identify any **relationships** between the ratios in this dataset by examining the above permutation matrix? – YES/ NO
- If yes, please name one pair of ratios that you identified as correlated.
- Can you identify any **clusters** in this dataset by examining the above permutation matrix? – YES/ NO
- If yes, please tell how many clusters you identified.
- Can you **describe the clusters** that you have identified, by examining the above permutation matrix? – YES/ NO
- If yes, please **mark one of the clusters** and describe it briefly in terms of values for financial ratios. Use Low, Medium, and High for indicating the prevailing level of the ratios in that cluster.
- Can you distinguish between the companies from one region or another? – YES /NO
- If yes, can you describe the characteristics of companies from Japan in comparison with the other regions by examining the above permutation matrix?
- Can you **compare** the characteristics of the companies A, B, C and D by examining the above permutation matrix? – YES /NO
- If yes, please tell in brief how Reno de Medici 1998 (B) performs in comparison with Buckeye Technologies 1998 (C) and Donohue 1998 (D). Also, tell how financial performance of Reno de Medici changed from 1997 (A) to 1998 (B).

Figure 1: Fragment from questionnaire

5. Results

In this section, we present the answers of the participants as percentages of positive, negative, invalid answers or non-responses. In Tables 2 – 14, the values in the 'description of clusters' and 'class description' rows (corresponding to the second parts of the questions 4 and 5, respectively) represent the percentages from the positive answers recorded for the 'clusters' and 'distinguish classes' rows (first parts of questions 4 and 5, respectively).

Line graphs are used for depicting one-dimensional data (typically for temporal data, in order to detect changes and fluctuations). Multiple line graphs can be used for depicting multi-dimensional data. Figure 2 represents different ratios, in different years (1997 and 1998). Table 2 shows that 75% of the users identified correctly outliers (question 1), and 42% identified correctly relationships between ratios (question 2). As expected, no respondents identified clusters (question 3). Fifty percent of the respondents distinguished between classes and 83% of these correctly described the Japanese companies (question 5). Finally, 42% of the respondents correctly compared the companies A, B, C and D (question 6).

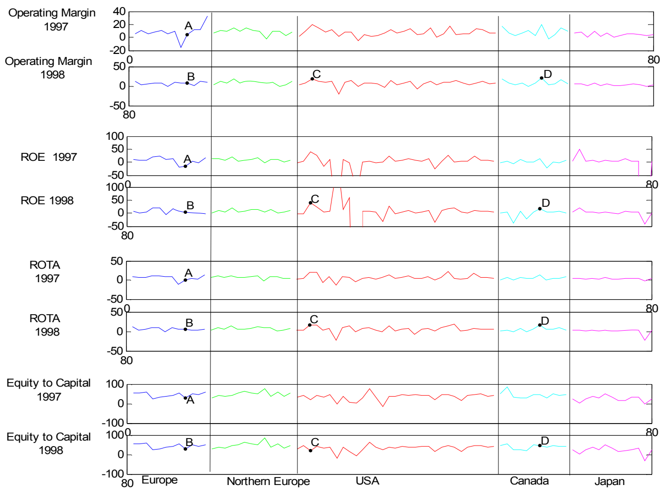


Figure 2: Multiple line graphs

Table 2: Evaluation of multiple line graphs

Answer/ task	Positive	Negative	Invalid
Outliers	75%	25%	0%
Dependency analysis	42%	50%	8%
Clusters	0%	100%	0%
Description of clusters	-	-	-
Distinguish classes	50%	50%	0%
Class description	83%	17%	0%
Comparisons	42%	58%	0%

Permutation matrix represents each dimension by a single bar graph. The name "permutation" is given because the order of the dimensions as well as the order of the observations can be changed interactively so that different patterns can be detected. Figure 3 arranges the observations in descending order according to ROTA. Among the patterns identified correctly by respondents are: outliers, relationships between ratios, and comparisons between companies (Table 3).

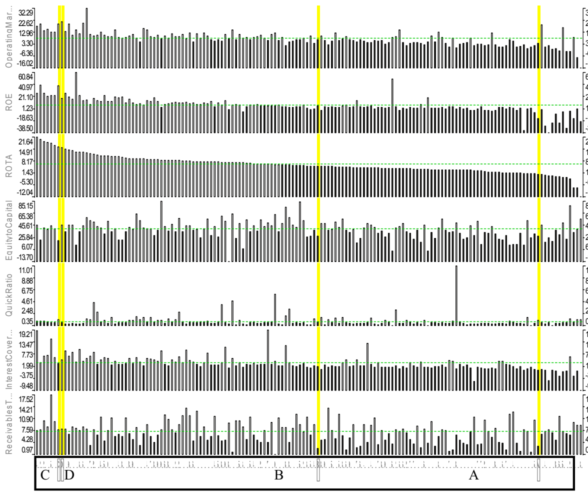


Figure 3: Permutation matrix

Table 3: Evaluation of permutation matrix

Answer/ task	*Positive*	*Negative*	*Invalid*
Outliers	92%	0%	8%
Dependency analysis	42%	58%	0%
Clusters	8%	75%	17%
Description of clusters	100%	-	100%
Distinguish classes	0%	100%	0%
Class description	-	-	-
Comparisons	67%	33%	0%

Survey plot is similar to permutation matrix, but the dimensions are depicted vertically. Figure 4 uses color coding for depicting regions. The companies are arranged in descending order according to ROTA. The patterns identified by most of the users are: outliers, relationships between ratios, class detection and description, and comparisons between companies (Table 4).

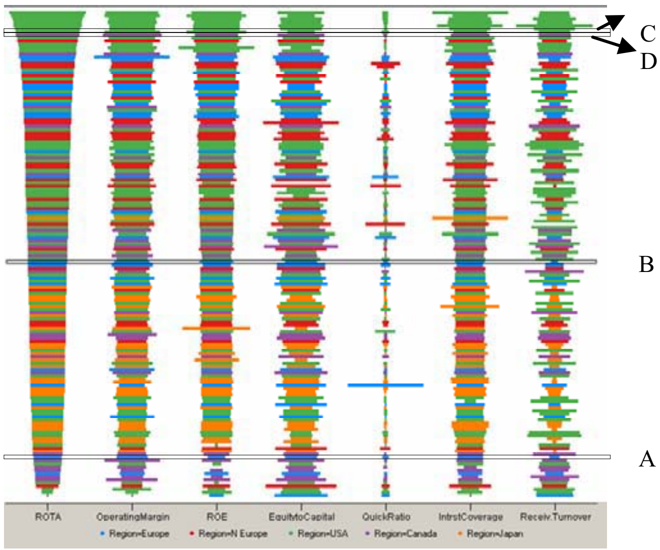


Figure 4: Survey plot

Table 4: Evaluation of survey plot

Answer/ task	*Positive*	*Negative*	*Invalid*
Outliers	83%	17%	0%
Dependency analysis	33%	67%	0%
Clusters	8%	83%	8%
Description of clusters	100%	0%	100%
Distinguish classes	75%	25%	0%
Class description	100%	0%	0%
Comparisons	75%	25%	0%

Scatter plot matrix is a matrix of all possible scatter plots constructed for every two variables in the dataset. A scatter plot usually is used for depicting the relationship between two variables. Table 5 shows that the respondents identified in Figure 5 outliers and relationships and made comparisons between companies.

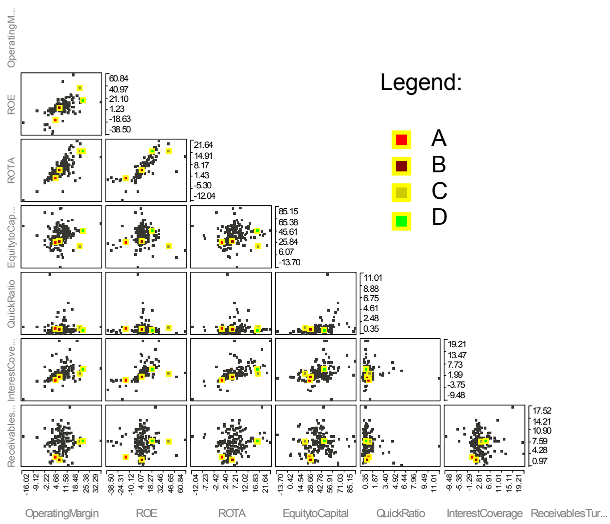


Figure 5: Scatter plot matrix

Table 5: Evaluation of scatter plot matrix

Answer/ task	*Positive*	*Negative*	*Invalid*
Outliers	67%	33%	0%
Dependency analysis	50%	50%	0%
Clusters	8%	83%	8%
Description of clusters	0%	100%	100%
Distinguish classes	0%	100%	0%
Class description	-	-	-
Comparisons	50%	50%	0%

Parallel coordinates depicts the data dimensions as parallel axes. Table 6 shows that the respondents identified in Figure 6 outliers, relationships and made comparisons between companies.

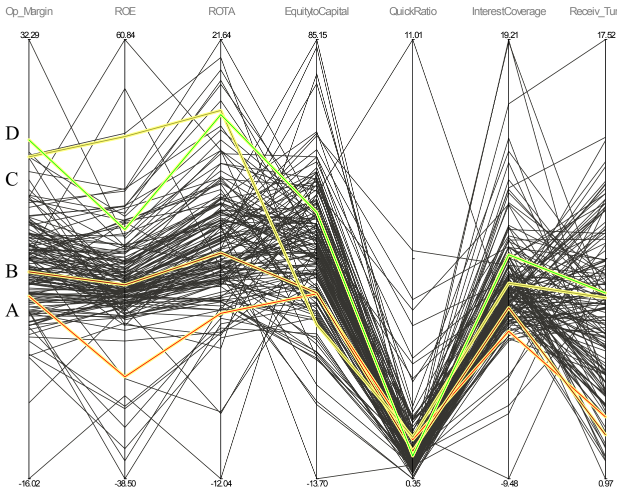


Figure 6: Parallel coordinates

Table 6: Evaluation of parallel coordinates

Answer/ task	*Positive*	*Negative*	*Invalid*	*Non response*
Outliers	67%	33%	0%	0%
Dependency analysis	17%	67%	17%	0%
Clusters	8%	83%	8%	0%
Description of clusters	100%	-	-	-
Distinguish classes	0%	100%	0%	0%
Class description	-	-	-	-
Comparisons	83%	8%	0%	8%

Treemap is used for representing hierarchical data. Figure 7 organizes the data in categories given by the year and region. Table 7 shows that the respondents identified outliers, distinguished and described companies from different regions, and compared the companies' financial performances.

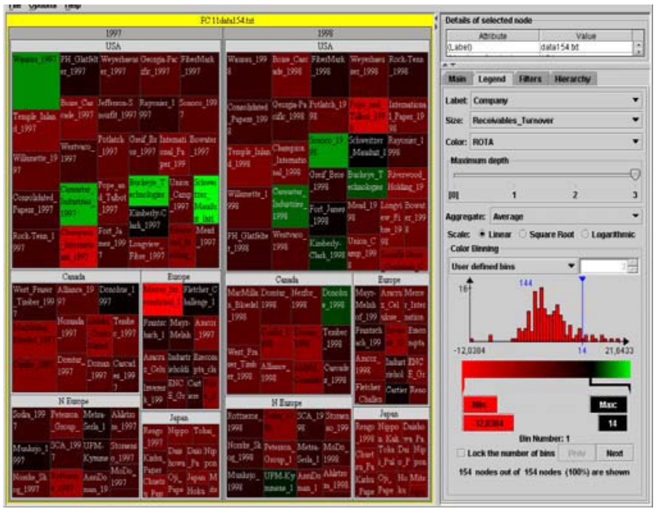


Figure 7: Treemap

Table 7: Evaluation of treemap

Answer/ task	*Positive*	*Negative*	*Non-response*
Outliers	67%	33%	0%
Dependency analysis	0%	100%	0%
Clusters	0%	100%	0%
Description of clusters	-	-	-
Distinguish classes	75%	17%	8%
Class description	100%	0%	100%
Comparisons	42%	58%	0%

Principal components analysis (PCA) is a statistical technique for dimension reduction and visualization of multidimensional data. Figure 8 depicts the dataset using the two first principal components. The respondents identified outliers, relationships, and clusters. They also correctly described the clusters and compared the companies' financial performances.

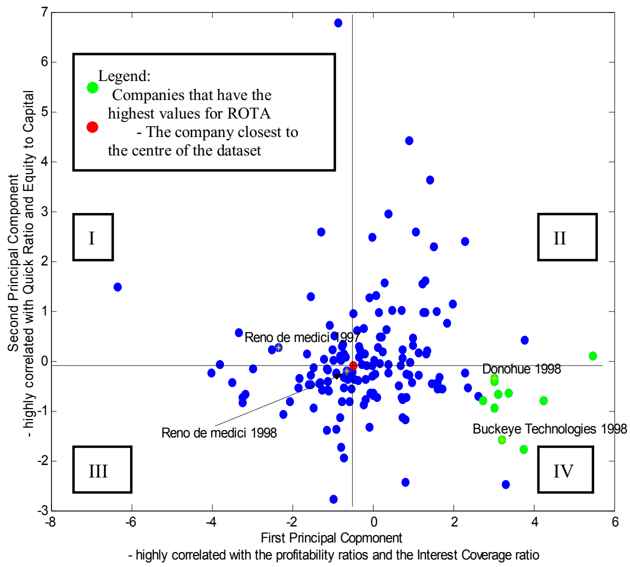


Figure 8: PCA

Table 8: Evaluation of PCA

Answer/ task	*Positive*	*Negative*	*Invalid*
Outliers	83%	17%	0%
Dependency analysis	17%	75%	8%
Clusters	33%	50%	17%
Description of clusters	60%	40%	-
Distinguish classes	0%	100%	0%
Class description	-	-	-
Comparisons	83%	17%	0%

Sammon's mapping is a metric multidimensional scaling technique used for dimension reduction. It is effective in depicting class distributions because it preserves the distances between data points. Figure 9 uses color coding for representing different regions. The users identified clusters and distinguished between different regions but could not describe them (Table 9).

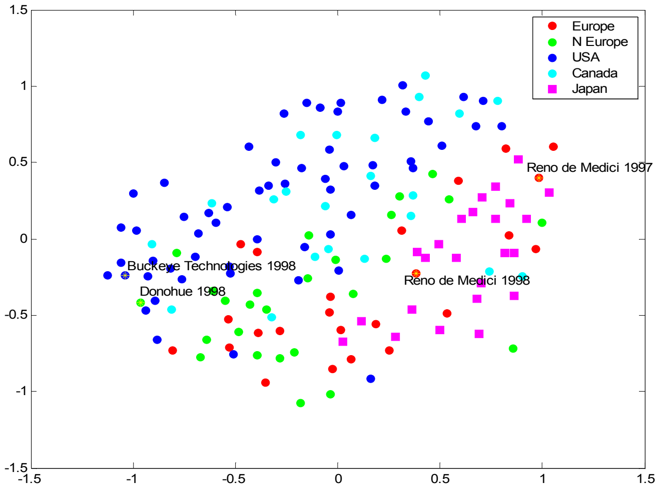


Figure 9: Sammon's mapping

Table 9: Evaluation of Sammon's mapping

Answer/ task	*Positive*	*Negative*	*Non-response*
Outliers	0%	100%	0%
Dependency analysis	0%	100%	0%
Clusters	25%	67%	8%
Description of clusters	0%	100%	0%
Distinguish classes	100%	0%	0%
Class description	0%	100%	0%
Comparisons	0%	100%	0%

SOM is a type of neural network used for clustering and visualization of multidimensional data. SOM output can be represented by using different graphical representations (Figure 10-13). Figure 10 shows the scatter plot view. The companies form different regions are highlighted with different colors. The users identified outliers, clusters and regions, but they could not describe the clusters and classes observed (Table 10). The U-matrix and the clustering views (Figures 11 and 12) reveal to the users only clusters (Table 11 and 12). The feature planes (Figure 13) reveal to most of the users outliers, relationships between ratios, clusters and descriptions of the clusters (Table 13). However, by combining all SOM views, the users identified most of the patterns (Table 14).

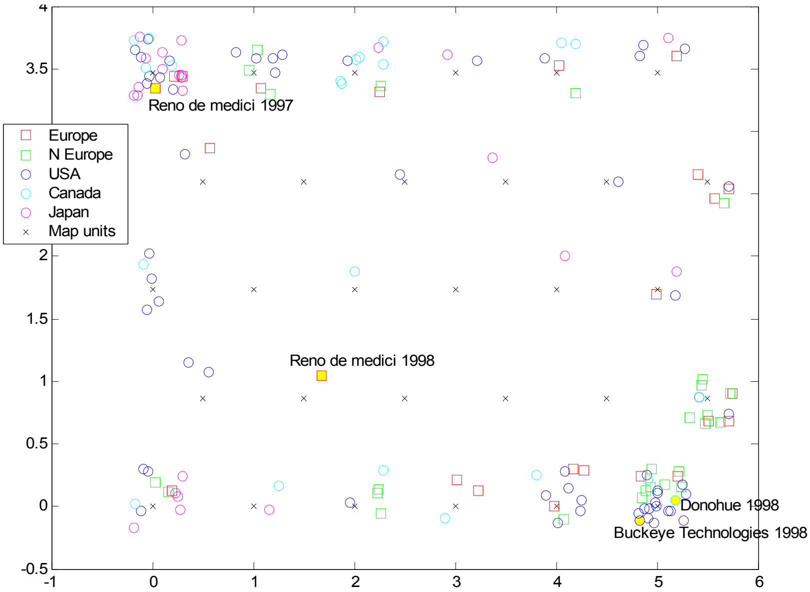


Figure 10: SOM – scatter plot view

Table 10: Evaluation of SOM – scatter plot

Answer/ task	*Positive*	*Negative*	*Invalid*
Outliers	25%	67%	8%
Dependency analysis	0%	100%	0%
Clusters	92%	8%	0%
Description of clusters	0%	91%	9%
Distinguish classes	58%	42%	0%
Class description	0%	86%	14%
Comparisons	0%	92%	8%

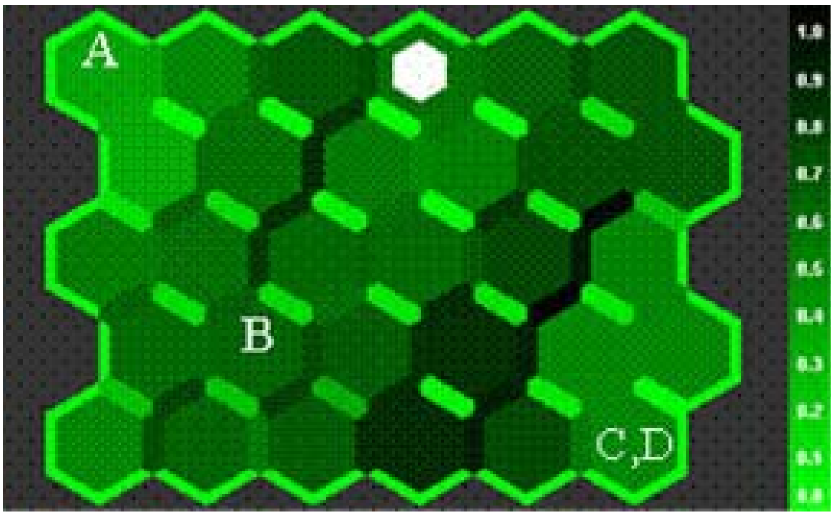


Figure 11: SOM – U-matrix view

Table 11: Evaluation of SOM – U-matrix

Answer/ task	*Positive*	*Negative*	*Invalid*
Outliers	0%	100%	0%
Dependency analysis	0%	92%	8%
Clusters	100%	0%	0%
Description of clusters	0%	92%	8%
Distinguish classes	0%	100%	0%
Class description	-	-	-
Comparisons	0%	83%	17%

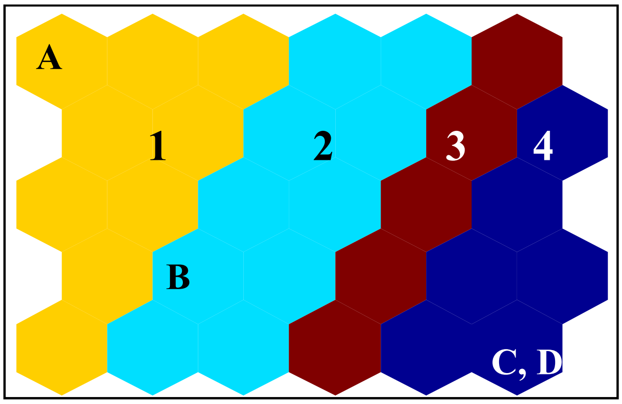


Figure 12: SOM – clustering view

Table 12: Evaluation of SOM – clustering

Answer/ task	*Positive*	*Negative*	*Invalid*
Outliers	0%	100%	0%
Dependency analysis	0%	100%	0%
Clusters	100%	0%	0%
Description of clusters	0%	92%	8%
Distinguish classes	0%	100%	0%
Class description	-	-	-
Comparisons	0%	92%	8%

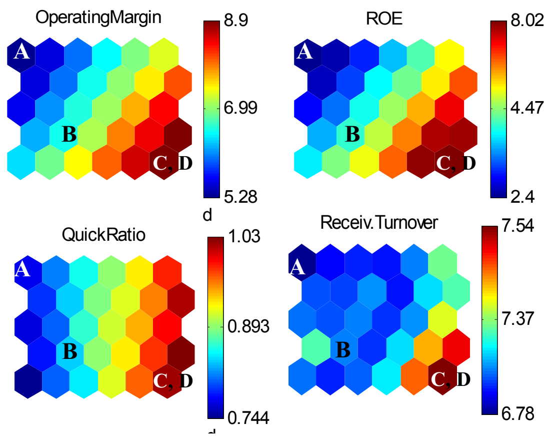


Figure 13: SOM – feature planes

Table 13: Evaluation of SOM - feature planes

Answer/ task	*Positive*	*Negative*	*Invalid*
Outliers	17%	75%	8%
Dependency analysis	75%	25%	0%
Clusters	100%	0%	0%
Description of clusters	92%	0%	8%
Distinguish classes	0%	100%	0%
Class description	-	-	-
Comparisons	90%	10%	0%

Table 14: Evaluation of SOM – all views

Answer/ task	*Positive*	*Negative*	*Invalid*
Outliers	30%	60%	10%
Dependency analysis	80%	20%	0%
Clusters	100%	0%	0%
Description of clusters	100%	0%	0%
Distinguish classes	70%	30%	0%
Class description	86%	14%	0%
Comparisons	100%	0%	0%

6. Discussion

Table 15 summarizes the positive answers obtained from the individual assessments made by participants. As mentioned previously, the number of positive answers indicates the extent to which each visualization technique is effective in revealing interesting patterns to the users. The centralization of all answers shows the following:

- All techniques except Sammon's mapping and SOM were very effective in uncovering *outliers* in the data.
- The SOM-feature planes and the scatter plot were the most effective in revealing *relationships* in the data. To a less extent, the line graphs, permutation matrix, survey plot, parallel coordinates, and PCA were effective in revealing relationships between ratios.
- The SOM visualizations, PCA, and to a less extent Sammon's mapping, permutation matrix, survey plot, scatter plot matrix, and parallel coordinates were effective in revealing *clusters*.
- The SOM-feature planes view was effective for *cluster description*. In addition, a small percentage of the participants that identified clusters on other visualizations than the SOM, were able to describe correctly the clusters found (e.g., permutation matrix, survey plot, parallel coordinates and PCA).
- The line graphs, survey plot, treemap, and all views of the SOM combined were effective in revealing *classes* in the data. In addition, Sammon's mapping and SOM-scatter plot were effective in uncovering the class patterns, but they did not facilitate the description of each class in terms of financial performance (marked with (-) in Table 15).
- The permutation matrix, survey plot, scatter plot, parallel coordinates, PCA and SOM-feature planes were effective in *comparing* companies' financial states.

Table 15 facilitates the comparison of the visualization techniques as to their effectiveness in solving different tasks associated to the financial benchmarking problem and dataset. Therefore, the table shows that among the most effective visualizations in uncovering different types of patterns are the SOM-all views, SOM-feature planes, PCA, survey plot, parallel coordinates, and permutation matrix. In addition, Table 15 shows that, given a particular visualization, some users are capable of detecting patterns in the data while other users are not.

Table 15 shows also that applying only one technique in financial benchmarking is not enough for getting a complete and accurate understanding of the data under analysis, since no single view is effective in uncovering fully all the patterns in the data. The visualization technique, that proved to be the most effective, is the SOM, but only if all combined views were analyzed together. However, even the SOM-all views together have the limitation of not being capable of showing the outliers that other visualizations can uncover. The second best technique is PCA.

Table 15: The extent to which the visualization techniques are effective in uncovering the patterns related to the data mining tasks identified for the financial benchmarking problem

	Tasks					
Technique	Outlier detection	Dependency analysis	Data clustering	Cluster description	Class description	Data comparison
Line graphs	+ +	+	-	-	+ +	+
Permutation matrix	+ +	+	+	+ +	-	+ +
Survey plot	+ +	+	+	+ +	+ +	+ +
Scatter plot matrix	+ +	+ +	+	-	-	+ +
Parallel coordinates	+ +	+	+	+ +	-	+ +
Treemap	+ +	-	-	-	+ +	+
PCA	+ +	+	+	+ +	-	+ +
Sammon's mapping	-	-	+	-	+ + (-)	-
Self Organizing Map – scatter plot	+	-	+ +	-	+ + (-)	-
Self Organizing Map – U-matrix	-	-	+ +	-	-	-
Self Organizing Map – clustering	-	-	+ +	-	-	-
Self Organizing Map – feature planes	+	+ +	+ +	+ +	-	+ +
Self Organizing Map – all views	+	+ +	+ +	+ +	+ +	+ +
Legend: + + more than or equal to 50 percent of the participants identified correctly the patterns, + less than 50 percent of the participants identified correctly the patterns, - no users identified the patterns.						

By using this evaluation method, we did not intend to obtain results generalizable to other datasets or business problems. The aim of the evaluation was to assess subjectively (by asking users) the extent to which they were capable of finding interesting patterns in a certain dataset by using each visualization technique in part. We showed that, by using this approach, we could assess the strengths and weaknesses of each technique in revealing interesting patterns for a particular dataset. In our study, the participants were familiar with the data and the financial benchmarking problem. We consider that the users must be familiar with the dataset and the business problem for which the techniques are assessed in order to obtain meaningful evaluation results.

The limitation of the study is that users evaluated only static visualizations of the data, while the interaction with the visualizations could have influenced the perception of the users as to the capabilities of the techniques. For example, by changing the order of companies in the multiple line graphs or using color-coding in PCA, these techniques could have revealed more patterns. Another limitation is that the users were familiar with the SOM from another university course, while other techniques may have been unknown to the users (e.g., survey plot, treemap, permutation matrix, Sammon's mapping). However, each technique was introduced to the participants before they answered the questions.

7. Conclusion

This paper presented a user evaluation study of nine multidimensional data visualization techniques. The evaluation concerned the use of these techniques for discovering interesting patterns in multidimensional financial data, patterns associated with the problem of financial benchmarking. These patterns are identified as outliers, clustering, cluster description, class description and data comparison. The purpose of the evaluation was to assess the extent to which users of the visualization techniques are able to identify and describe interesting patterns in data, and therefore to prove the effectiveness of the visualization techniques in uncovering these patterns.

The contribution of the study is that it presents empirical evidence of the effectiveness of different multidimensional data visualization techniques for getting insight into a particular financial dataset. Moreover, the evaluation method and the questionnaire were relatively easy to implement and yielded data that were easy to interpret. The evaluation method consisted in questionnaire-based data collection and data analysis of the 12 assessments provided by the participants in the study. The results of evaluation are useful especially in the early stage of the development of a visualization system, because they help in the process of selection of most appropriate techniques for solving certain tasks.

Some future research directions are, for example, to study the effects of color-coding to the effectiveness of techniques, to study the use of multiple techniques for getting insight into data, and to determine the extent to which the fact that users are familiar with the data and problem under analysis affects the evaluation of the techniques. Moreover, it is interesting to investigate whether the results differ when larger datasets are used for analysis, when different categories of users are involved, or when users interact with the visualizations.

Acknowledgments

I gratefully acknowledge the contribution of Professor Barbro Back, and the financial support of Tekes (grant number 40435/05) and the Academy of Finland (grant number 104639).

References

Card, S.K., Mackinlay, J.D., and Shneiderman, B. (1999) Readings in Information Visualization - Using Vision to Think, Morgan Kaufmann Publishers.

Chen, C., and Czerwinski, M. P. (2000) “Empirical evaluation of Information Visualization: an Introduction”, Int. J. Human-Computer Studies, 53, pp. 631-635.

Demsar J., Zupan B., Leban G. (2004) Orange: From Experimental Machine Learning to Interactive Data Mining, White Paper (www.ailab.si/orange), Faculty of Computer and Information Science, University of Ljubljana.

Dull, R. B., and Tegarden, D. P. (1999) “A Comparison of Three Visual Representations of Complex Multidimensional Accounting Information”, Journal of Information Systems, vol. 13, No. 2, pp. 117-131.

Eklund, T. (2004) The Self-Organizing Map in Financial Benchmarking, TUCS Doctoral Dissertation, Åbo Akademi University, Turku.

Hinterberger, H. H., and Schmid, C. (1993) “Reducing the Influence of Biased Graphical Perception with Automatic Permutation Matrices”, SoftStat ’93, Proceedings of the Seventh Conference of the Scientific Use of Statistic-Software, March 14–18.

Hoffman, P. E., and Grinstein, G. G. (2002) “A Survey of Visualizations for High-Dimensional Data Mining”, in Fayyad, U., Grinstein, G. G., and Wierse, A. (eds.), Information Visualization for Data Mining and Knowledge Discovery, Morgan Kaufmann, San Francisco.

Keim, D. A. (1996) “Pixel-Oriented Visualization Techniques for Exploring Very Large Databases”, Journal of Computational and Graphical Statistics, 5(1), pp. 58-77.

Liu, Y., and Salvendy, G. (2007) “Design and Evaluation of visualization support to facilitate decision trees classification”, Int. J. Human-Computer Studies, 65, pp. 95-110.

Marghescu, D. and Rajanen, M. (2005) “Assessing the Use of the SOM Technique in Visual Data Mining”, Proc. of the IASTED Int. Conf. on Databases and Applications (DBA 2005), pp. 181-186.

Marghescu, D. (2006) “Evaluating the Effectiveness of Projection Techniques in Visual Data Mining”, Proc. of the 6th IASTED Int. Conf. on Visualization, Imaging and Image Processing (VIIP 2006), pp. 94-103.

Marghescu, D. (2007) “Multidimensional Data Visualization Techniques for Exploring Financial Performance Data”, Proc. of 2007 Americas Conference on Information Systems (AMCIS 2007), Keystone, Colorado, USA, 9-12 August. To appear.

Nenet 1.1 (1999) http://koti.mbnet.fi/~phodju/nenet/Nenet/General.html

Picket, R. M., and Grinstein, G. G. (2002) “Evaluation of Visualization Systems”, Chapter 3, in Fayyad, U., Grinstein, G. G., and Wierse, A. (eds.), Information Visualization for Data Mining and Knowledge Discovery, Morgan Kaufmann, San Francisco.
Plaisant, C. (2004) “The Challenge of Information Visualization Evaluation”, Proc. of the Working Conference on Advanced Visual Interfaces (AVI 2004), ACM Press, pp. 109-116.
Risden, K. and Czerwinski, M. P. (2000) “An Initial Examination of Ease of Use for 2D and 3D Information Visualizations of Web Content”, Int. J. Human-Computer Studies, 53, pp. 695-714.
SOM Toolbox 2.0 for Matlab (2005) http://www.cis.hut.fi/projects/somtoolbox/
Somervell, J., McCrickard S., North, C., and Shukla, M. (2002) “An Evaluation of Information Visualization in Attention-Limited Environments”, Joint EUROGRAPHICS – IEEE TCVG Symposium on Visualization 2002, pp. 211-216.
Soukup, T., and Davidson, I. (2002) Visual Data Mining: Techniques and Tools for Data Visualization and Mining, John Wiley & Sons.
Stasko, J., Catrambone, R., Guzdial, M., and McDonald, K. (2000) “An evaluation of Space-Filling Information Visualizations for Depicting Hierarchical Structures”, Int. J. Human-Computer Studies, 53, pp. 663-694.
Sutcliffe, A. G., Ennis, M., and Hu, J. (2000) “Evaluating the Effectiveness of Visual User Interfaces for Information Retrieval”, Int. J. Human-Computer Studies, 53, pp. 741-763.
The MathWorks (2000) MATLAB - The Language of Technical Computing, Using Matlab Graphics, Version 6.
The MathWorks (2002) Statistics Toolbox for Use with MATLAB - User’s Guide, Version 4.
Treemap 4.1 (2004) http://www.cs.umd.edu/hcil/treemap/.

Paper 6

Marghescu, D., Rajanen, M., and Back, B. (2004). Evaluating the Quality of Use of Visual Data-Mining Tools. In *Proceedings of the 11th European Conference on IT Evaluation (ECITE 2004),* Amsterdam, The Netherlands, November 2004.

Evaluating the Quality of use of Visual Data-Mining Tools

Dorina Marghescu[1], Mikko Rajanen[2] and Barbro Back[1]
[1]Turku Centre for Computer Science, Åbo Akademi University, Turku, Finland
[2]Dept. of Information Processing Science, University of Oulu, Finland
dorina.marghescu@abo.fi
mikko.rajanen@oulu.fi
barbro.back@abo.fi

Abstract: In this paper we propose a framework for evaluating quality of use of visual data-mining tools. The evaluation framework addresses three levels of analysis: visualisation, interaction, and information. We examine the applicability of the framework to the Self-Organising Maps tools. For this purpose we conducted an exploratory study using the mixed methods research design, and its results are reported in this paper. The conclusion is that our framework can be used for evaluating different visualisations techniques, with small variations from case to case.

Keywords: visual data-mining, usability evaluation, quality of use, Self-Organising Maps, visualisation.

1. Introduction

Data mining is the process of extracting information from large quantities of data by employing advanced computational techniques. Because the data in organisations' databases are rapidly growing, the data-mining activity is not always easy and successful. Users of data-mining tools need fast access to data, real-time interaction with the system, and high-quality information. Whereas traditional algorithmic techniques are analysing the data automatically, information visualisation techniques in data mining involve the human to use his/her capabilities to detect structures and to process patterns in data.

The information visualisation literature reveals a variety of novel and sophisticated visualisation techniques. The problem is that they are not always implemented and/or used to fulfil the real demand of users. One example is Self-Organising Maps (SOM) (Kohonen 2001). The SOM method is a special type of neural network that allows the mapping of high-dimensional data onto a smaller dimensional space, making accessible large amounts of data through a visual model. The capabilities of the SOM technique have been extensively explored in different research areas for more than two decades (Kaski et al. 1998, Oja et al. 2003). Although a large body of research explores the applicability of the SOM method to economic and financial data (Kaski and Kohonen 1996, Back et al. 2000), there is no evidence that business-oriented practitioners use this technique in their work.

This lack of evidence has encouraged us to evaluate the quality of use of the SOM tools. Our approach to evaluating the SOM software consists of three steps: developing a framework of evaluation, selecting the appropriate attributes to measure, identifying the problems and limitations of the SOM tools.

The research problem we intend to tackle in this article is to develop a framework for evaluating the visual data-mining tools from the user perspective (step 1), and to apply it to evaluating the SOM tools (steps 2 and 3). The need of a framework rose because we did not find a suitable model in the literature we reviewed, despite the fact that in the visualisation literature, many authors emphasised the necessity for systematic empirical evaluation of visualisation techniques (Card et al. 1999, Chen and Czerwinski 2000). The framework for evaluating the quality of use of visual data-mining tools that we propose in this study attempts to clarify the following issues:

- How is the quality of use defined?
- What attributes of the visual data-mining system must be assessed?
- How do these attributes relate?

- How could these attributes be assessed?

Based on established theories and empirical studies reported in the literature, we developed the framework for evaluating the quality of use of visual data-mining tools by taking into consideration three levels of analysis: visualisation, interaction and information. For each of the three levels, we identified and described the corresponding attributes.

To examine the applicability of the framework, we conducted an exploratory study on the SOM tools use, and we report the results in this article. The purpose of the study was to examine the attitude of the SOM tools' users, and to shed light on the quality of solutions the SOM users reported. In the quantitative part of the study, we employed the survey technique to collect data about users' attitudes and opinions regarding the SOM tools. The research questions in this part of the study were:

- Determine what attitude the users have regarding the SOM technique,
- Determine the significant relationships between the attributes evaluated,
- Determine the consistency of the measurement.

In the qualitative part of the research, we analysed multiple case studies, collected in the form of reports on the solutions provided by the users to the task given. The research questions for the qualitative part of the study were:

- Determine the quality of the solutions reported by users,
- Determine how the quality of the solutions reflects on the users' attitude on SOM use.

The paper is organised as follows. In section 2, we briefly describe a review of the related literature. In section 3, we propose a framework for evaluating the visual data-mining tools from the user perspective. Section 4 describes the methods and procedures applied for evaluating the quality of use of the SOM tools. In section 5 we report the results obtained. Section 6 contains relevant discussion about our proposed evaluation framework and its generalisability. We conclude in Section 7 with final remarks and future work ideas.

2. Review of related literature

This section highlights few methods from the usability evaluation literature. It also looks into related studies regarding evaluation of the visualisation tools.

2.1 Usability evaluation

Usability is defined in standard ISO/IEC 9126-1 as being the capability of the software product to be *understood*, *learned*, *used* and *attractive* to the user. Bevan (1995) refers to usability with the term *quality of use*. This reflects the extent to which the users can achieve specific goals with *effectiveness*, *efficiency*, and *satisfaction*.

Dix et al. (1998) point out that usability evaluation of the system is conducted in order to ensure that the system behaves in conformity with developers' expectations and users requirements. The evaluation methods are divided into four categories: analytic methods, specialist reports, user reports, and observational reports. The techniques corresponding to user-centric evaluation include experimental methods, observational methods, and surveys.

An example of survey instrument is the End-User Computing Satisfaction (EUCS), developed by Doll and Torkzadeh (1988). It measures the user satisfaction with both information product and ease of use items, using five sub-scales: content, accuracy, ease of use, format, and timeliness.

Another survey instrument is Software Usability Measurement Inventory (SUMI) for assessing user attitudes regarding software tools (Kirakowski 1994).

2.2 Evaluation of the visualisation techniques

Tufte (1997) and Bertin (1981) provide us the bases for defining quality with regard to visualisation. Card et al. (1999) point out the importance attached to the evaluation of visualisation techniques. Moreover, according to Chen and Czerwinski (2000), the proliferation of visualisation techniques also highlights the need for principles and methodologies for empirical evaluation of these techniques. However, relatively little research has been done in this area. Morse et al. (2000) propose a method for evaluation based on a visual taxonomy, intended to test the visualisation in isolation from the rest of the system. Other studies are concerned with the effectiveness and utility of the tools (Stasko et al. 2000), or they are targeted to specific types of visualisation (Risden et al. 2000, Sutcliffe et al. 2000).

In this paper, we are concerned with evaluating the quality of use of visual data-mining tools in order to assess the user satisfaction. We take into consideration all the relevant aspects of the system: visualisation, interaction with the system, and information provided.

3. The framework for evaluating the quality of use of visual data-mining tools

The activities, in which the user is involved during the visual data-mining process, are depicted in Figure 1. To accomplish certain goals and tasks, the user employs the domain knowledge, and the data available in databases. The access to the data is allowed through data-mining systems. In essence, the visualisation represents an interface to the data stored in the databases. For simplicity, we describe the way in which the human uses the system as follows. With a certain goal in mind, the user examines the visualisation, interacts with it, and finally gets some information. The user satisfaction and, therefore, the success of the data-mining process depend on how good the visualisation, the interaction and the information are.

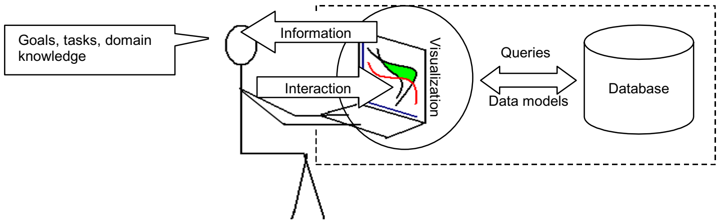


Figure 1: The relations between visualisation, interaction and information in data-mining process

A *good visualisation* properly represents the data of interest. The initial settings should be adequate and practical. The graphical design must convey structures and content of data. The visualisation system should allow a variety of exploration tasks such as overview, details of data, and filter, to facilitate to the user the access to the desired information. Finally, the visualisation should make the user to think about data, and allow the transfer of the results to other applications.

A *good interaction* with the system is ensured when the system is efficient, accurate, and easy to use and learn.

Regarding the *information*, this must be interesting, new, reliable and accurate.

3.1 Definition of terms

3.1.1 Quality of use

Quality of use of a visual data-mining tool is defined as being the totality of features and characteristics of the tool that reflect on its ability to satisfy the users' needs. In other words, quality of use reflects the satisfaction of the user with all features of the tool. As stated above, the main and direct features of the system, that influence the user attitude and behaviour, are: visualisation of data, user-system interaction, and information obtained.

3.1.2 Quality of visualisation

At this level we are concerned with evaluating the capability of the visualisation system to transform the input data and make them accessible to the user. The issues to be evaluated are presented in Figure 2.

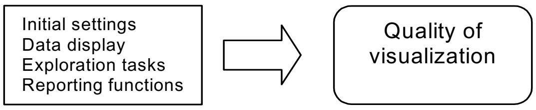


Figure 2: Evaluating the quality of visualisation

- *Initial settings* refer to the requirements on input data format, the degree of data abstraction, and the setting of the parameters for visualisation.
- *Data display* regards the possibility to visualise the data structure, data variation, data content, and data comparison. Moreover, the description, tabulation and decoration of data are important to evaluate.
- *Exploration tasks* include the five visual tasks identified by Shneiderman (1996), i.e. overview, details of data, filter, details on demand, and relate.
- *Reporting functions* represent those system functions that allow the user to transfer the results outside the application for various purposes. In this part we are concerned with evaluating whether the user is satisfied with how s/he benefits from the visualisation. We also ask whether the user is encouraged by the visualisation to think of the data, rather than of the graphical design and methodology.

3.1.3 Quality of interaction

Assessing the quality of interaction is conducted in order to find out whether the users of the system consider the system easy to use and learn, accurate, effective and efficient. We classify the interaction attributes in five groups (Figure 3).

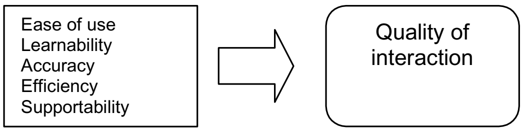


Figure 3: Evaluating the quality of interaction

- *Ease of use* stands for the characteristic of the system to be easy to control by the user and to provide the user with freedom of action (controllability and flexibility).
- *Learnability* affects how easy and fast the users feel that they master the system to perform the desired tasks.
- *Accuracy* (reliability) reflects the frequency and severity of system errors or failures.
- *Efficiency* measures the degree to which users feel that the software helps them in their work (to tailor frequent actions, improve working performance, and receive fast response to queries).
- *Supportability* regards the users' access to documentation and support, when needed.

3.1.4 Quality of information

Assessing the quality of information is meant to answer whether the users are satisfied with the output information provided by the system. Figure 4 shows the four attributes of information, which the user might require.

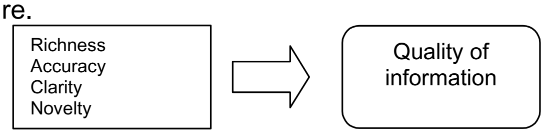


Figure 4: Evaluating the quality of information

- *Richness* of information stands for completeness, usefulness, and interestingness. Also it must correspond to users' needs and expectations.

- *Accuracy* of the information regards the degree to which the information is precise, correct, and consistent with users' knowledge.
- *Clarity* of information means that the information is presented in a clear and understandable way, and allows interpretation and inferences.
- *Novelty* of information reflects the characteristic of being new and up-to-date.

3.2 Relationships between attributes

The relationships between the attributes corresponding to the three levels of assessment are described in Figure 5.

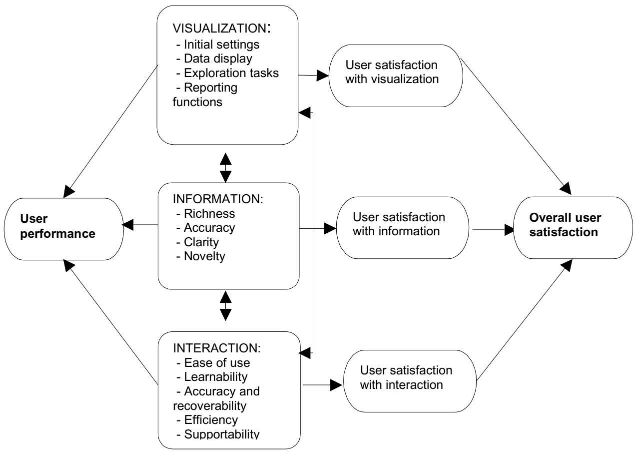


Figure 5: Relationships between attributes

When the user examines the data display, and uses the results, s/he must find the information being rich, accurate, clear, easy to interpret, novel and up-to-date. Moreover, whenever the user interacts with the system, s/he wishes the process to be easy, accurate, and effective.

4. Exploratory study: evaluating quality of use of the SOM tools

We employed the mixed methods research design in order to analyse the quality of the SOM tools, and also to get insight into the quality of the solutions the users found. For the quantitative part of the study, which concerned the quality of use of SOM tools, we used the questionnaire survey technique to collect data. In the qualitative part of the study, we were interested in analysing the participants' solutions to the task they were asked to solve.

4.1 Participants

The participants in our study were 26 students, enrolled for an Information Systems course, in a public university. The research site was the classroom. The demographics of the participants are presented in Table 1.

Table 1: Demographics of the participants in the survey

Category	Values	Percentage
Major	Information systems	61,54
	Computer Science	19,23
	Economics and Computer Science	11,53
	Mathematics	3,85
	Accounting	3,85
Years at university	1, 2 years	26,92
	3, 4 years	34,62
	5 and over	30,77
	Non response	7,69
Programming experience	Yes	80,77
	No	19,23
Data analysis experience	Yes	26,92
	No	73,08

4.2 Materials

In our study, we used three software packages, which implemented the SOM algorithm, all being available online for downloading. These were SOM_PAK, SOM Toolbox for Matlab, and Nenet (Kohonen 2001).

The data collection process consisted of the following phases: 1. the students were trained to use all three SOM tools, 2. they were asked to solve an assignment and report their findings, 3. after returning the solutions, the students were asked to answer the questionnaire.

The students had the possibility to choose the tools they wanted to work with, out of SOM_Pak, SOM Toolbox for Matlab, and Nenet. Nenet was definitely preferred by all students, for visualising the maps, while different students used either SOM_PAK or SOM Toolbox to train the maps. We used the Binomial, and Chi-square tests (Siegel and Castellan 1988) to check whether there are differences in attitudes between users of the SOM_PAK and SOM Toolbox, but no significant differences were found.

4.3 The quality attributes

Based on the framework described in Section 3, we selected the attributes of SOM tools to be evaluated (Figures 6, 7, and 8).

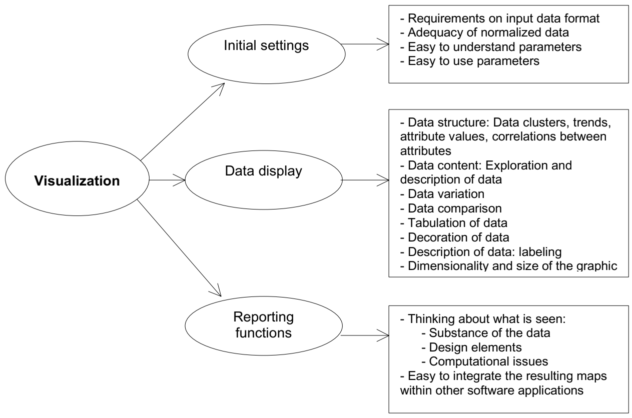


Figure 6: Attributes of visualisation

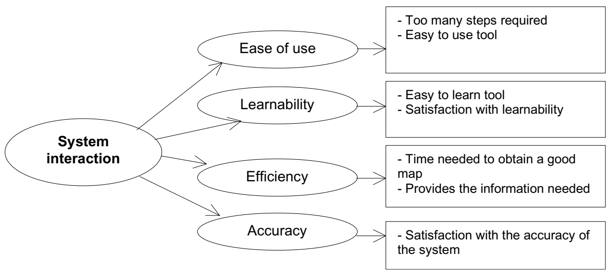


Figure 7: Attributes of interaction with the tool

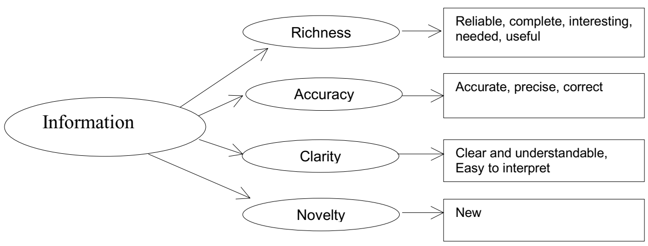


Figure 8: Attributes of information

5. Results

5.1 Quality of use of SOM tools

Figure 9 depicts the opinions regarding the *quality of visualisation*. Among the positive features, we observe the good visualisation of data clusters (92% respondents agree), the visualisation of the comparable data and data trends.

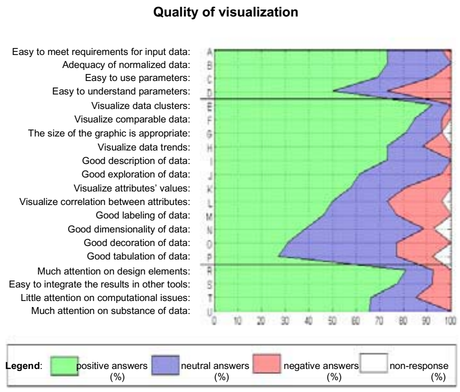


Figure 9: Quality of visualisation. A – D: Initial settings, E – P: Data display, R – U: Reporting functions

The initial settings did not reveal major problems. However, the SOM parameters were found easy to understand only by 50% of students. Regarding the data display features, relatively low scores are noticed for tabulation of data, decoration of data, visualisation of the correlations between attributes, and visualisation of the attributes values. At the reporting functions category, we observe that more than 75% of participants found easy to use the results within other applications, and the attention of the users was focused on the substance of data for more than 65% of participants.

We asked a number of questions about the degree to which different design elements helped in interpreting the visualisation (map). The answers are presented in Table 2.

Table 2: Assessment of the SOM's graphic elements

	Helpful			Adequate		
(%)	Agree	Neutral	Disagree	Good	Medium	Poor
Colors	92	8	0	88	12	0
Scales (color bars)	85	15	0	85	15	0
Grids, neurons, borders	81	19	0	57.5	31	11.5
Attribute values	69	19	8	54	31	15
Data labels	77	15	8	61.6	19	19.4

Figure 10 presents the opinions and attitudes regarding the *quality of interaction*. Among the positive interaction features are the ease of use, and ease of learning. Also, most of the users (82.60%) agreed that the system provided the information needed. The weak points perceived by the students are system flexibility (54% respondents agreed that there are too many steps required to get a good map), and efficiency (only 27% respondents were satisfied with the time needed to get a good map).

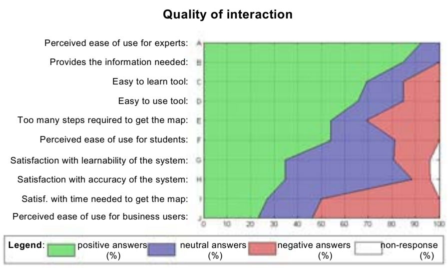


Figure 10: Quality of interaction

Figure 11 shows that the *information* obtained is helpful and useful in data analysis. It is also interesting, easy to understand, and complete for most of the students. However, these are not very satisfied with the correctness of the information and even less with its preciseness. Users still find the SOM content reliable, and overall the satisfaction with the content is high.

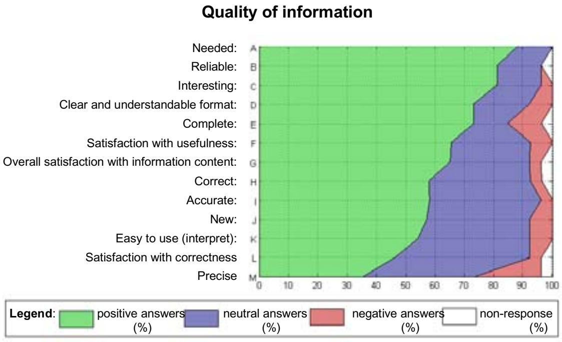


Figure 11: Quality of information

5.2 User performance

Participants in the experiment were asked to solve a complex task with SOM tools, namely to train the SOM until they obtain a map and with its help to answer five questions. For evaluating the user performance we analysed the students' reports describing the solutions found.

Figure 12 shows that the most difficult for students was to obtain an appropriate map on which to identify correct clusters. The first three questions, concerning the number of clusters and their definitions, received the most varied answers and these were not very well argued. Students themselves were aware that their map might not be the correct one, and noticed that an inappropriate map could lead to misinterpretations and mistakes in the decision making process. The last two questions are obviously much better answered.

Among the explanations the users gave to their imperfect solutions were the inexperience of working with SOM tools, the unfamiliarity with financial ratios, and the highly subjective criteria to separate the clusters (for some managers some ratios are more important in a certain time, etc.). Overall, the participants found it very interesting and useful to work with the SOM technique. It must be noticed that even 92% of the students were satisfied with the visualisation of the data clusters, only 62% of the students gave acceptable and good solutions for that task (question Q1).

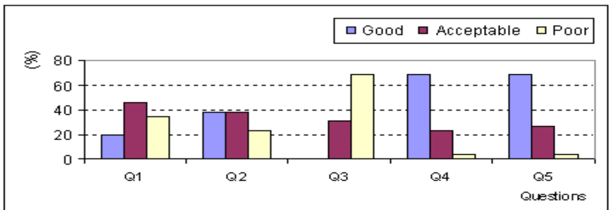


Figure 12: Quality of solutions reported by participants

5.3 SOM tools limitations

Table 3 shows the main limitations of the SOM tools pointed out by our study. For each identified problem we propose possible solutions and suggestions to improve the software that implements SOM, in addition to those stated by Kohonen (2001).

Table 3: Problems found and suggestions for improvement

Problem	Suggestion for improvement
Level 1: quality of visualisation Not very easy to understand input parameters Poor tabulation of data Poor decoration of data Medium data labelling	- Automation of parameters selection according to the input data characteristics and the desired results, - Enhance the "Details on demand" feature to display properly the input data and their statistics in tabular reports.
Level 2: quality of interaction Low perceived ease of use for business users Medium satisfaction with the time needed to get a good map (visualisation), too many steps required Medium satisfaction with the accuracy of the system Medium satisfaction with the learnability of the system	- Provide automatic delineation of the clusters. - Due to the fact that SOM reduces the dimensions of the input space, the loss of accuracy is inevitable, but new learning algorithms could be tested for implementation.
Level 3: quality of information Not very precise Not high satisfaction with correctness Not very easy to use (interpret) Not very accurate	- Add explanations to the information displayed when these are requested.

6. Discussion

6.1 Consistency of the measurements

In order to examine the reliability of the scales that we used in assessment, we have computed the Cronbach's alpha coefficient. A rule of thumb states that the internal consistency of the scales is acceptable when alpha is greater than 0.7. Table 4 presents the Cronbach's alpha values for our data. At the Visualisation level, there are lower values of alpha for Initial settings construct and Reporting functions. This is due to the fact that the questions in this section of the questionnaire were focused on distinct issues, so that no significant similarities in answering were found. Also, the six satisfaction questions that we used were not highly related and the corresponding Cronbach's alpha is relatively low. These low values are justified by the small number of items used, because the value of alpha increases directly with the number of items of the construct and also with the correlation between the items.

Table 4: The Cronbach's alpha computed for each level of assessment

Level	alpha	Notice	alpha
Visualisation quality	**0.7724**	when graphical aspects are included:	**0.8704**
Initial settings	0.3971		
Data display	0.7273	when graphical aspects are included:	0.8704
Reporting functions	0.5659		
Interaction quality	**0.6739**	including visualisation items:	**0.7046**
Ease of use and learning	0.6143	including visualisation items:	0.6774
Accuracy	not computed, only one item used		
Efficiency	not computed, only one item used		
Information quality	**0.7467**	including visualisation items:	**0.8748**
Richness	0.5443	including visualisation items:	0.7732
Accuracy	0.6075		
Clarity	0.6110		
Novelty	not computed, only one item used		
Satisfaction questions	**0.6291**		
All quality questions	**0.8872**	using the three-point scale, derived from the original five-point scale	
User Performance	**0.7044**	for the scores we assigned to the solutions offered by students	
Overall	**0.8845**	user performance and quality questions	

6.2 Interdependencies between attributes

For exploring the interdependencies between variables, we performed an exploratory factor analysis, based on the extraction of the principal components (PC). Applying this technique to the data revealed us that only a selected number of variables were to be retained as significant. Table 5 presents the variables that show a high contribution in the variance of the data corresponding to each level of assessment.

Table 5: The most contributing variables in evaluation

Level	PC	Cumulative variance of rotated components (%)	Most significant variable in the rotated component	Weight in rotated component
Visualisation	1	13.217	Data labels adequacy	0.897
	2	24.957	Colours helpfulness	0.853
	3	33.475	Tabulation of data, Dimensionality of data	0.817 0.820
	4	41.707	Description of data	0.873
	5	49.613	User performance items (Q2), but also Q5	0.695
	6	57.266	Adequacy of normalised data	0.854
	7	63.524	User performance (Q4)	0.759
	8	69.171	Easy to understand parameters	0.823
	9	74.408	Attention on data representation, Data attributes representation	0.661 -0.817
	10	79.504	Data clusters visualisation	0.816
	11	84.546	Requirements on data format	0.813
Interaction	1	18.868	User performance (Q5)	0.861
	2	33.782	Easy to use tool	0.920
	3	47.353	Easy to use for students	0.835
	4	59.111	Satisfaction with accuracy	0.841
	5	70.469	Efficiency	0.872
Information	1	16.978	Easy to interpret	0.809
	2	28.988	User performance (Q5), but also Q1, Q2, Q4	0.777
	3	40.547	Completeness	0.884
	4	51.626	Usefulness	0.815
	5	61.411	Correctness	0.751
	6	70.670	Novelty	0.906
	7	79.109	User performance (Q3)	0.801

We also explored the correlations between the variables derived from the factor analysis. For example, it resulted that the user performance is interdependent with the ease of use of the tool (correlation coefficient = 0.42), preciseness (0.44), clarity (0.401), visualisation of the data attributes correlations (0.488), visualisation of the data variations (0.488), adequacy of the data labels (0.423). Other notable correlations are: attention on data representation is correlated highly with tabulation of data and adequacy of data labels.

The evaluation framework we presented can be generalised by using the approach for generalising from theory to description. According to Lee and Baskerville (2003) this type of generality involves generalising from theoretical statements to empirical (descriptive) statements. The framework can be applied with small variations from the present format in different settings, and with different visualisation techniques or visual data-mining tools.

Regarding the method for assessment, we recommend the user-centric approaches. The user-centric approaches to qualitative evaluation employ a representative number of people out of the actual or potential users of the software tool. One method for collecting the data from these people is the survey. The survey technique is appropriate for our problem because it is designed to assess the relative frequency, distribution, and interrelations of naturally occurring phenomena in the population under study.

7. Conclusions

We developed a framework for evaluating visual data-mining tools, and we examined the satisfaction of the users with SOM tools. The framework consists of three levels of evaluation:

visualisation, interaction, and information. These levels are not completely separated, but interdependent.

To examine the applicability of the framework, we conducted an exploratory study for evaluating the quality of use of the SOM tools. Quality of use was defined as being the satisfaction with all the features of the SOM software, namely visualisation of data, interaction with the system, and information obtained. The results showed that the users were satisfied working with SOM tools. Most of the visual features were considered helpful and adequate. People were helped by the SOM technique to understand and analyse relatively large amount of data and to obtain interesting and new information. Regarding the interaction with the tools, participants in the study found the tools easy to use and learn. Nevertheless, the SOM tools appear to have also weak points. These are identified in terms of “too long time needed to obtain a good map”, relatively low accuracy, preciseness, and correctness of the information, difficulty in interpreting the results. All these shortcomings, especially the lack of efficiency and preciseness might be explanations of why business users do not use frequently the SOM tools in financial data analysis.

The significance of the study is twofold. Firstly, we provided a comprehensive framework for assessing the visual data-mining tools from the user perspective. Secondly, the study offers insights into the use of the SOM tools, from data collected through a survey questionnaire and multiple case studies. These insights into how people effectively use and think about the SOM tools can help developers of complex commercial applications in visual data mining to gather new and interesting information about the tool, its users and their needs.

A limitation of the study is that the sample used in the exploratory study does not represent the target population (business users), but students. This drawback might be compensated by the fact that the students worked on a real life problem and real data. Moreover, the sample size is relatively small.

For future we aim to test thoroughly the applicability of the evaluation framework, by examining other tools. Moreover, the causal relationships that the framework reveals remained unexplored and we intend to conduct formal experiments in order to explore them fully.

References

Back, B., Öström, K., Sere, K., and Vanharanta, H. (2000) “Analyzing Company Performance Using Internet Data”, *Proceedings of the 11th Meeting of the Euro Working Group on DSS*, Ed. by Zaraté, Toulouse, France, pp52-56

Bertin, J. (1981) *Graphics and Graphic Information Processing*, Walter de Gruyter, New York.

Bevan, N. (1995) “Measuring usability as quality of use”, *Software Quality Journal*, Vol. 4, No. 2, pp115-130.

Card, S., Mackinlay, J., and Shneiderman, B. (1999) *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, San Francisco.

Chen, C., and Czerwinski, M. (2000) “Empirical evaluation of Information Visualizations: an introduction”, *International Journal of Human-Computer Studies*, Vol. 53, pp631-635.

Dix, A., Finlay, J., Abowd, G., and Beale, R. (1998) *Human-Computer Interaction*, Second Edition, Prentice Hall.

Doll, W.J. and Torkzadeh, G. (1988) “The measurement of end-user computing satisfaction”, *MIS Quarterly*, Vol. 12, Issue 2, pp259-274.

ISO/IEC 9126-1 (2001) *Software Engineering, Product quality, Part 1: Quality model*, International Standards Organization.

Kaski, S., Kangas, J., and Kohonen, T. (1998) “Bibliography of Self-Organizing Map (SOM) papers 1981-1997”, *Neural Computing Surveys*, pp102-350.

Kaski, S. and Kohonen, T. (1996) “Exploratory Data Analysis by the Self-Organizing Map: Structures of Welfare and Poverty in the World”, *Proceedings of the Third International Conference on Neural Network in the Capital Markets*, World Scientific.

Kirakowski, J., (1994) “The Use of Questionnaire Methods for Usability Assessment” [online], http://www.ucc.ie/hfrg/questionnaires/sumi/index.html, last accessed June 2nd 2004.

Kohonen, T. (2001) *Self-Organizing Maps*, Third Edition, Springer.
Lee, A.S., and Baskerville, R.L. (2003) "Generalizing Generalizability in Information System Research", *Information Systems Research*, Vol.14, No.3, pp221- 243.
Morse, E., Lewis, M., and Olsen, K.A. (2000) "Evaluating visualization: using a taxonomic guide", *International Journal of Human-Computer Studies*, Vol. 53, pp637-662.
Oja, E., Kaski, S., and Kohonen, T. (2003) "Bibliography of Self-Organizing Map (SOM) papers: 1998-2001 Addendum", *Neural Computing Surveys*, 3, pp1-156.
Risden, K., Czerwinski, M.P., Munzner, T., and Cook, D.B. (2000) "An initial examination of ease of use for 2D and 3D information visualizations of Web content", *International Journal of Human-Computer Studies*, Vol. 53, pp695-714.
Shneiderman, B. (1996) "The Eye Have It: A Task by Data Type Taxonomy for Information Visualizations", *Proceedings of Visual Languages* (Boulder, CO, September 3-6). IEEE Computer Science Press, Los Alamitos, CA, pp336-343.
Siegel, S., and Castellan, N. J. (1988) *Non-Parametric Statistics for the Behavioral Sciences*, Second Edition, McGraw-Hill.
Stasko, J., Catrambone, R., Guzdial, M., and McDonald, K. (2000) "An evaluation of space-filling information visualizations for depicting hierarchical structures", *International Journal of Human-Computer Studies*, Vol. 53, pp663-694.
Sutcliffe, A.G., Ennis, M., and Hu, J. (2000) "Evaluating the effectiveness of visual user interfaces for information retrieval", *International Journal of Human-Computer Studies*, Vol. 53, pp741-763.
Tufte, E. R. (1986) *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, Connecticut.

Turku Centre *for* Computer Science

Joukahaisenkatu 3-5 B, 20520 Turku, Finland | www.tucs.fi

University of Turku

- Department of Information Technology
- Department of Mathematics

Åbo Akademi University

- Department of Information Technologies

Turku School of Economics

- Institute of Information Systems Sciences

ISBN 978-952-12-2152-1
ISSN 1239-1883