Database Issues for Data Visualization: Supporting Interactive Database Exploration

Daniel A. Keim[1], John Peter Lee[2], Bhavani Thuraisinghaman[3], Craig Wittenbrink[3]

[1] Institute for Computer Science, University of Munich
[2] University of Massachusetts, Lowell
[3] The MITRE Corporation, Bedford MA
[4] University of California, Santa Cruz

1 Introduction

A recurrent theme from the Second Database Issues for Data Visualization workshop was the importance of interactively exploring databases using numerous tools and techniques. Database exploration is a discovery process where relevant information or knowledge is identified and extracted from data. It is related to the field of Knowledge Discovery in Databases (KDD), and emphasizes the *process* of knowledge discovery: the development of hypotheses about the data, and the validation of those hypotheses. Discovery is not only possible from analytic tools, but also from graphical, textual, numeric, and tabular presentations of data. Flexibility in data processing and output presentation are fundamental requirements of any data exploration environment. A shared sentiment among workshop participants was that database exploration requires the cooperation of database management, data analysis and data visualization facilities, as shown in Figure 1.

Interaction is also central to database exploration. The user must *interact with data* to discover information. User-data interactions must, then, be supported by an integrated exploration system. Because of the potential complexity of such a system, interactions occur at many levels between the data, system and user. These include interactions among software modules and user-data interfaces. Process management will play a larger role in the development of adequate exploration environments, because such environments will be integrations of numerous types of systems that focus on smaller aspects of the overall problem. If we are to realize any benefits from such an integration, the whole must be greater than the sum of the components.

1.1 Goals of the report

To build useful data exploration and analysis systems, we must bring the disciplines of database management, data analysis and data visualization together and examine the importance of interaction in this context. This will pose new problems and conditions resulting from the additional capabilities of the integrated system. The primary goal of this report is to look at interactions in

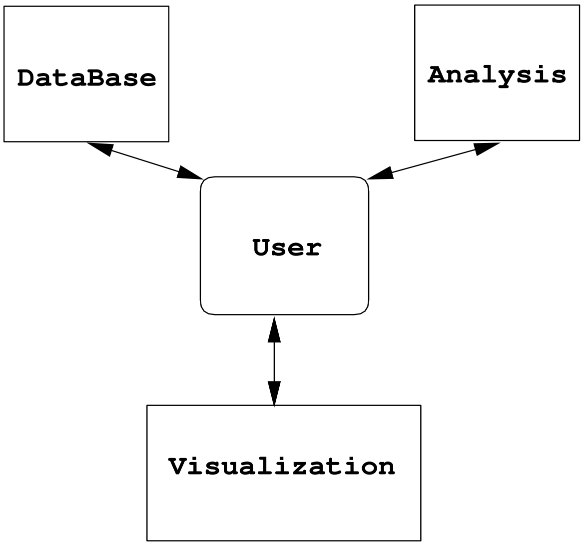

Fig. 1. Interactions between the user, Database, Visualization, and Ana lysis The components of scenarios in user interaction explored in the report

database exploration, from a system and user perspective. The ultimate research goal is to effectively support user-data interactions in both the underlying data models and high-level data interfaces. The two primary focus areas are:

1. **Interactions Between System Components** - What are the system components, and how do they interact/communicate with each other? What are the duties and responsibilities of each component? What constraints must be considered to integrate the components?
2. **Interactions Between User and System** - What does the user interact with and how does the user interact with each system component? What additional constraints does the user impose on the system, or what constraints are placed on the user by the system?

In this report we describe the system components and outline several research issues necessary to make the visualization component the interface to the analysis, database, and visualization. Figure 2 shows the user interacting directly with the visualization, which then acts as an intermediary to the other systems. This changes the user from being surrounded by many tools, to one using and interacting with a single tool, but managing the same sophistication of analysis, data-mining, knowledge discovery, and visualization. Visualization's effective human communication qualities enable it to play this important role.

1.2 Definitions of Important Terms

This domain has an ever increasing number of terms that often mean the same thing, and the fundamental definitions in this domain are fairly broad. We distil

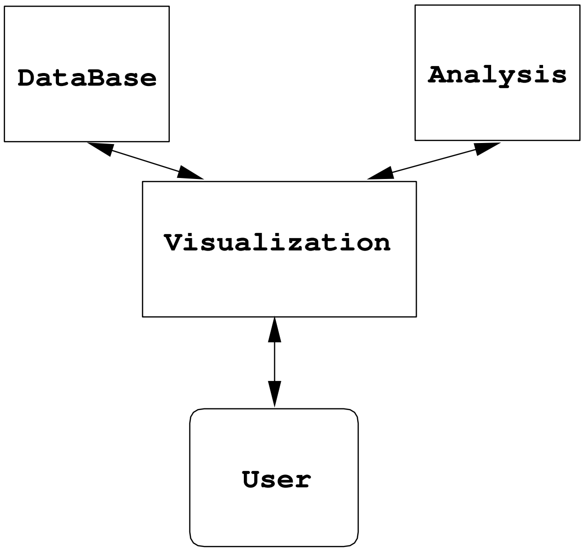

Fig. 2. User primarily interacting with the visualization to access the other system elements.

the following from the universe of definitions to suit our discussions:

data analysis - The computation of quantitative deductions from data. This is the broadest defintion, and often the ultimate goal of any database exploration process. Data analysis can have numeric, symbolic or graphical, etc., components.

data visualization - The graphical presentation of data, whether the data is base data, summary data, configuration data, or knowledge extracted from data. This is a type of visual data analysis, where the analytic component is offloaded to human perception. Often, qualitative deductions are performed from the visualization, and must be confirmed by quantitative data analysis.

knowledge discovery / data mining - The nontrivial, algorithmic, and predominantly autonomous extraction of previously unknown information (patterns, clusters, rules, etc.) from data. The primary focus here is on autonomous calculations and algorithms to extract structure from data. Data mining often specifically relates to the algorithms, and knowledge discovery often relates to the overall processes.

database exploration - The process of finding previously unknown information from a database, using database management, data analysis, data visualization, and data mining facilities. This places greater focus on the human user component of the knowledge discovery process. Since visualization is a critical component, user perception and guidance are also critical. More formally, database exploration is the process of finding subsets D' of a database D and hypotheses $H(D', U, C)$ which a user U considers "useful" in a context C. Database exploration is inherently interactive, focusing on refining hypotheses based on results from interacting with the data (visualizations, data analysis, database queries, etc.). Note that what we have called database exploration is also referred to by some as data mining. We have assumed that database exploration includes what we have referred to as data mining together with other functions such as data analysis, data visualization etc. As stated earlier, there are no standard terms. We have defined the terms to clarify the ideas presented in this paper.

1.3 Overview of Report

Section 2 describes a system model for interactive database exploration, and system interaction scenarios. Section 3 discusses research issues that must be addressed. Section 4 concludes the paper.

2 Overview of an Interactive Data Exploration System

Database exploration is a complex process that cannot adequately be described in terms of database, analysis, or visualization domains alone. An empirical study [Spri92] developed a taxonomy of data analysis tasks, most of which are not adequately supported by current software environments. These tasks encompass all three domains we are considering, and exist at a very low level. A higher level description of the knowledge discovery process [Brac96] emphasized the human component and highlighted the process as iterative, protracted over time, and requiring a great deal of bookeeping. Sequences of task iterations have been modeled in various forms for some time [Nich83, Carr86, Oldf88, Vell90, Youn91], usually to help manage the data analysis process, but not to analyze or exploit it. Exploitation of interaction histories are outlined in [Lee90], and provide the impetus for analyzing database explorations for improvements in system performance and user modeling. There has been little, if any, research done, however, to model and exploit the database exploration process.

In this section, we describe the primary components of a prototypical interactive database exploration system, the inter-component interactions, and the user-component interactions.

2.1 System Components

An effective data exploration system has three principal components: database management, data analysis, and data visualization. There is also a user component that must be supported by the system as shown in Figure 2. This is an incremental, evolutionary approach towards an effective database exploration environment. We envision these three separate system components to be tightly integrated in some way in the future, but currently can only develop interfaces between them, due to their complexities. Knowledge gained from research into this integration will help drive the design of next-generation database exploration, analysis, and visualization environments.

The database management component provides persistent data storage, data integrity and concurrency. It possesses a formal data model with well-defined access functions. A data manipulation language (DML) is avaliable to specify the selection of data. The data model is most often relational, though object-oriented and object-relational models are gaining popularity. Database management systems rely heavily on the user to pose queries and manipulate query results. Database management systems may manage centralized databases or distributed databases. The databases may also be heterogeneous in nature. Furthermore, heterogeneity could be with respect to data models, query processing, query languages, and transaction management algorithms. Most of the commercial database management systems have been geared toward OLTP (online transaction processing) applications. However, due to the large quantities of heterogeneous information, there is a need for many applications to process complex queries. As a result, data warehouses are not being developed. A data warehouse extracts data from multiple databases depending on the queries to be posed and stores the data in a single logical location so that a user need to query only the warehouse. Data warehouses are developed for OLAP (online applications processing) and decision support. Data exploration techniques can be used to extract information from the warehouse. Queries that update data warehouses are extremely application dependent, so an application analysis is necessary in order to abstract out the relevant queries for further analyses.

The data analysis component performs statistical operations over data distributions, transformations into new data, and provides discovery functions such as clustering, regressions and pattern matching. Some have used data mining for data analysis. However, we have defined data analysis as computing some derived data from the data in the database. Application-specific computations such as fluid dynamics and finite element methods are also supported by this component. The data model is often based on matrices and arrays. Typical analysis systems are most often file-oriented systems with some means for accessing data stored in relational databases, or data warehouses. Automatic discovery algorithms determine interesting and significant pattern classifications from this data with minimal user intervention. A more human-centered approach is exemplified by On-Line-Analytic-Processing (OLAP) systems, where data is organized into hierarchies of different resolutions, and a "drill-down" capability is used to access lower-level data from retrieved summary data. Other data navigation aids are used such as cross-tabulations of numeric and categoric data. Aside from comparing graphs, a popular method for analyzing data graphically is to link data displays together, and brush one display to see the effects on a different display [Buja91, Buja96].

The visualization component is primarily concerned with data presentation in the form of graphical displays. A secondary concern is providing an adequate user interface to map data and operations to intuitive forms for the user to interact with data. It is both an output and input component, providing results and facilitating graphical interactions. The data model is often a structural model having connectivity and topologic constraints. The visualization field has yielded many techniques for portraying data in 2D and 3D spaces, with animations where time is a variable. They place the discovery burden upon the user, who can apply graphical probes or alter the display parameters (such as opacity, thresholds, data ranges and viewing orientation). They encourage more "intimate" data interactions because data is mapped into a comprehensible (graphical) format for presentation. Visualization systems, however, are file-oriented, and lack robust data selection or data analysis capabilities.

The user component drives the discovery process by formulating hypotheses, testing hypotheses (issue query, transform and analyze result, compute new data, construct presentations, etc.) and drawing conclusions. The user usually has some knowledge of the application domain, and perhaps some knowledge of the exploration system and its internal data representations. The user performs sequences of linked interactions that serve to isolate relevant portions of the database, to generate and validate hypotheses about the data. There exist elements of navigation through data spaces, comparisons and annotations of important results and processes that must be supported by the exploration system components.

2.2 System Component Interactions

Data Analysis - Database Management: Most database systems are not used for data analysis, only data selection, so the analysis component must operate over retrieved data subsets. The analysis might produce new data, that might need to be inserted back into the database. Also, having a drill-down feature at the analysis component would necessitate queries to be issued to the database from the analysis component. It might be easy to transform between data models, because a matrix is rather similar to a relational table. Some analyses might need special data structures, so a mapping is required to a possibly more complex format. If the database system supports analysis, it would probably serve some needs well, as autonomous agents can continually scour the base data for interesting information, or extremely large databases can be analyzed in-place, without the need to transfer data to and from another tool.

We have worked on several projects that concern related issues of mapping data from the analysis packages to and from the database management systems. In REINAS the Real-Time Environmental Information Network and Analysis System [Long95] we have found that conceptualization of system components affects not only the performance and design, but the way in which users work with the system. Different user views of the data become embedded simply from experience with the tools, that in part coaches their interactions with the data. In other words, the mappings are experiential. And therefore an evolution of coexisting views of the data means that there are coexisting mappings. Even though the views are not mandated, they become important to the user in the training, conceptualizations, and expectations of what the system can and cannot do.

Database Management - Data Visualization: The visualization field is still very immature when considering data models and management. There are a multitude of data models, making translation to the database difficult. Most visualization data models are data format specifications, with little thought given to allowable operations over the data, as in the relational data model. Visualization often works in a batch mode, on huge numbers of records, so a cursor interface might not be appropriate for database-visualization interaction. By nature, the visualization component affords greater interactions with data because the graphical output may uncover relationships and information undetected by automated tools. Thus, there are visual probes and selections occuring that need to be mapped to the database. The querying and drill-down capability from visualization to database is being explored in the Exbase system [Lee94, Lee95, Lee96]. Exbase is a layered system that translates both data and interactions between database and visualization. We are currently exploring means for rapid data selection, such as dynamic queries [Albh92] for retrieved data, and to augment this with actual database queries.

Also, in traditional data base management systems there is no possibility to visualize the result of a query or the relevance of the data items with repsect to the query. New techniques directly presenting huge amounts of data to the user are needed if data visualization shall be used to support exploring large databases. An example for such techniques are the pixel-per-value techniques developed in the VisDB project [Keim94, Keim95, Keim95a]. The basic idea of pixel-per-value techniques is to map each data value to a colored pixel and present the data values belonging to each dimension in a separate window. This permits visualization of up to a million data values. The pixel-oriented techniques use different arrangements of pixels for different purposes. The user may use a query-independent visualization technique which sorts the data according to some dimension(s) and uses a screen-filling pattern to arrange the data values on the display. This is useful for data with a natural ordering according to one variable (e.g., time series data). However, if there is no natural ordering of the data and the main goal is an interactive exploration of the database, the user will be more interested in *feedback* of some query. In this case, the user may turn to the query-dependent visualization techniques which visualize the relevance of the data items with respect to a query. The query-dependent visualization techniques calculate the distances between data and query values, combine the distances for each data item into an overall distance, and visualize the distances for the dimensions and the overall distance, sorted by the overall distance. The arrangement of the data items centers the most relevant data items in the middle of the window, and disperses less relevant data items in a spiral-shape towards the outside of the window.

Data Analysis - Data Visualization: The data visualization component is critical to display the results produced by data analysis. This may be, for example, the visualization of summary data or the curves that represent portions of the data. On the other hand, the data analysis component may carry out further analysis from the visualization data produced by the data visualization component of the data in the database. Data analysis and computation actions are often performed over data to be visualized. Popular visualization dataflow environments often have a suite of analysis and compute modules to apply to data. These modules operate over the native data model of the visualization environment, so there is a tight integration of the two. In fact there are a large number of systems available that have explicit data analysis - data visualization capabilities, many developed from the initial theories of Tukey [Tukey77]. Examples include S Plus / Trellis [Beck88], XGobi [Sway92], and Data Desk [Vell92, Wilh95]. Direct interaction techniques, by not using a strict dataflow approach have also been developed such as used in the modular visualization environment of the *Spray Rendering* approach [Pang95]. Methods for evaluating visualizations have been researched in [Clev85].

2.3 User Interactions

A user should be free to interact with any of the three system components, to control their configuration and behavior. A user also may control the processes between components. For example, with respect to the data analysis and data visualization component interaction, a user may observe the results produced by the visualization component and guide the data analysis component as to the queries to be posed to the database management system. When the results produced by the database management system are visualized, the user again observes the results, and then guides the data analysis component again. In other words, a user could be a participant for every possible interaction. A worthwhile goal should be to effectively map the interactions at the user interface to the appropriate system action, so that the user is not burdened with unnecessary system details.

Data Visualization Component - User: The data visualization component enables a user to visualize, among other things, the database structure, base data in the database tables, database queries, query results and data analyses. Thus, it is the crucial component to support navigations through data spaces and the system components. Visualization allows the user to configure displays and control selection, analysis and display parameters by providing consistent and intuitive visual metaphors. It guides the exploration, and serves as the glue that holds the environment together. Visualizations of the data may also be used to generate new (qualitative) hypotheses and for confirming hypotheses, so there is a definite tight coupling between the visualization and the user.

Database Management - User: The data management system supports the user by providing persistent storage for both the base data under investigation and the metadata that is generated by the investigation. The metadata summarizes discoveries, and can be used for prediction and planning purposes. It also supports advanced users in directly posing complex queries using the database query language. The primary use of the DBMS is to satisfy queries that select data based on some criteria of interest. Other tasks include granting access priveledges, configuring access and storage parameters, and ensuring data integrity. Database interaction through a DML is often extremely difficult, so this must be shielded from the user to enable efficient querying with minimal effort.

Data Analysis - User: Data analysis aids the user in making deductions, performing data mining, and in decision support. This is primarily for quantitative data analysis to confirm a hypothesis, or to isolate data based on some rule that is not specifiable in the databases' DML. It may, however, also be useful for qualitative data analysis as a means of confirming stages of an exploration. Data must be supplied from either the database or visualization component, as the user directs some database subset or visualization portion to be analyzed. In many cases, the data analysis is visual, so means of comparing and computing over visualizations is necessary.

3 Database Exploration Research Issues

In practice, we usually do not find a database exploration system as described in Figures 1 or 2. In most cases, only a limited portion of the model is realized, usually being restricted to one system component and the user. As a consequence, the interactions between the system components is also restricted. Depending on the system component which is used in the model, three different classes of users may be distinguished:

1. **Statistics-oriented User** Statistics-oriented users perform a statistical analysis of the data for obtaining usually quantitative information about the data.
2. **Database-oriented User** Database-oriented users focus on exploring their databases by using the querying facilities of the database management system. The results may be of a qualitative or quantitative nature.
3. **Visualization-oriented User** Visualization-oriented users try to obtain new insight into the data by using different visualizations of the data. The results are usually qualitative in nature.

It is clear that all three interaction scenarios have their place in an integrated database exploration environment. For practical database exploration, the techniques employed in each of the domains have to be smoothly integrated. Also, an effective interaction between the system components as well as between the system components and the user has to be developed.

If data exploration tools from the different areas are used at all, it is usually the user who has to manually combine the different tools. The user might get some hypotheses about a subset of the database by using a visualization tool, s/he might retrieve the subset by using the database query lanuage, and finally s/he might try to confirm the hypotheses by using statistical analyses. In some other case, the user might get some initial hypotheses by using a statistical analysis tool and then turns to the database querying or visualization tool to confirm the hypotheses. Any sequence of using the different tools would be possible since there are many different techniques in either of the domains. A problem of the current situation, however, is that the interaction between the tools from the different domains is usually handled manually by the user. The user first obtains results by using one of the techniques and then decides which other techniques from the same or a different domain are suitable to continue the data exploration process. In our model, all three domains (statistical analysis, database querying, and visualization) are assumed to be fully integrated which, in our opinion, is necessary for an efficient and effective database exploration. A smooth integration however also raises many research issues which have to be solved to make the data exploration environment work.

We present several important research issues which we believe to be interesting, challenging and worthwhile. As mentioned previously, the integration of all three areas is important for effective data analysis. For example, because of the massive amounts of data being collected and archived into many DBMS's, any pruning, preprocessing, or discovery of important information by an analysis module as data is collected would be a great aid. Data mining alone will not determine interesting patterns and correlations that were unanticipated in its algorithms, but a visualization might portray such information easily. The synthesis of visualization with exploration of data bases will provide capabilities that would otherwise not be found, because of the capabilities of the people who use the tools we may develop.

An integrated system model: It is one issue to assemble the components such that they work together. This involves providing adequate data models and translation functions. It is another issue to provide an environment that exploits each component and the user exploration tasks at hand. This involves user-data interaction support, database subset manipulations and discovery representations and annotations. Having some intelligent retrieval mechanisms to access and analyze relevant discoveries from a knowledge base is an important aspect of this goal. An associated goal would be the ability to build rules from a visualization discovery, and insert these rules back into the knowledge base. We also need an interaction model, and a transaction model at the database. A complete model will not be realized until numerous system experiments are conducted, so this is probably the largest open research problem stated here.

Query formulation and refinement: Database exploration has two main goals: (1) to find important information within data, and (2) to find the important questions that (1) answers. The user often doesn't know what to look for, only that s/he must find *something* interesting. This implies that we not only need robust tools to explore and analyze the data, but also the means to express the results to others and to the integrated system. We need to have powerful knowledge representations and management, as well as the requisite user interfaces to enable efficient query specifications and interactions. Some promising research is described in [Brac93]. As the presentation of knowledge is refined, so is its specification. Improving query formulation and refinement requires research in defining an effective query, evaluation of users interacting with visualization and DBMS systems, and developing new interactive techniques and systems. For example, how can queries be specified from an immersive virtual environment? How can analyses or visualizations be specified in queries, or should they be left out?

Visualization for knowledge discovery: Visualization is inherently qualitative, but can be used in a quantitative manner as well. We need to know when it is most effective, for example, as a guidance mechanism to identify interesting regions of the data, for data reduction and exploration focus, or as a useful mechanism for in-depth comparative analysis. Again, the importance of experiments to determine when visualization is appropriate and can be relied upon for analysis is crucial. Little has been done on quantifying the effectiveness of visualization, as opposed to the underlying performance and system issues [Rush95].

Additional visualization discovery tools: The visualization component has incorporated into it aspects of database selection and data analysis, so these tasks must be supported by the visualization user interface. There is, then, a need for more visualization discovery tools, and visualization techniques for database exploration. For example, we have simple data probes, and can brush linked displays. But, what else can be used? Related work is represented by some developments in modular visualization environments, such as Spray Rendering [Pang95]. Direct manipulation interfaces will serve a useful purpose here. Dynamic query interfaces have been shown to be useful for data selection, but are limited to queries expressed as ranges over attributes. Visual programming is also being used to express queries to databases, but query networks are often more complicated than the SQL they support. Navigation through multidimensional data spaces is still an important issue deserving attention.

Cognitive models of data: The user has mental models of the data and the system. These need to be supported with a minimum of cognitive effort. Does the user need to interact with all components, or just one or two? It might seem natural for all user interaction to take place between the user and visualization system, and map data and interactions from the visualization to the other components. Though other workgroups investigated the modelling of data, the effects ripple through the interactions. The efficiency of any user task depends upon their knowledge of the system, the data, and the goals (tasks), and therefore chaining the model of the data and system may have a direct correlation to the task efficiency.

Data Exploration on the World Wide Web(WWW): There is now much work on interfacing database management systems with the WWW. SQL queries are now embedded into HTML home pages. This way users could access the multiple databases on WWW. Database exploration research should also include extracting the information from the WWW databases. What are the additional features that need to be added to HTML home pages? What is the impact on languages such as JAVA on data exploration? What are the appropriate data exploration techniques for digital libraries? Much research is needed.

Collaborative Database Exploration: Another important area is collaborative computing. The idea is for users to collaboratively work together to solve a problem, and the users could be at different locations. Since database exploration is a complex process, it may not be possible for one person to extract useful information in many cases. Therefore multiple individuals possibly at different locations may have to work together to collaboratively explore and extract useful information. Therefore, database exploration needs to be integrated with collaborative computing technology.

/em Performance Techniques: Of course, the demand for greater performance with larger and larger archives of data will continue to drive research. Techniques for parallelization and distributing of query execution, as well as for parallelization of the even more demanding data mining and knowledge discovery algorithms must be developed. Data mining is a batch application now, but if it were possible to make such computer intensive applications interactive, the capabilites for interactive data exploration would be tremendously improved.

4 Conclusions

In order for visualization to be ultimately useful as a knowledge discovery technique, it must be integrated into a system that also supports database and data analysis processes. We envision an evolution from monolithic visualization systems to integrated database exploration and knowledge discovery environments as outlined in this report. Once visulalization systems become more "data-centric", by offering data selection and analysis at the visualization interface, they will attain a more prominent status in knowledge discovery environments. Their greatest utility is in presentation and interaction affordances, and so they must seamlessly integrate with other tools that perform the core knowledge discovery tasks.

The interactions and metaphors for interaction determine in large part the success and usability of any database system. By focusing on the efforts necessary in combining visualization front ends with database engines, we have sought to clarify the research issues in pushing this field forward as quickly as possible. The possibility is there of a user interacting purely with a visualization front end to control the database, data mining, and data analysis investigations of very large datasets. Of course there are complications in connecting each of the modules together directly, and there are added complications in human interface choices. What we have shown is that the assumptions about where the user lies in relation to the system, and the tasks that a performed change the definition of the interaction, and add/change the necessary systems development. Direct graphics manipulation can be easily used for a high percentage of interactions with visualization from databases, but cannot handle all tasks. The open problem of how to map gestures to queries is intriguing from practical and theoretical standpoints. We hope that this group report serves to aid those in investigating this important area.

References

[Albh92] Albherg, C., C. Williamson, B. Schneiderman, "Dynamic Queries for Information Exploration: An Implementation and and Evaluation", *Proceedings of CHI'92*, pages 619-626, 1992.

[Beck88] Becker R., Chambers J. M., Wilks A. R.: 'The New S Language', Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, CA., 1988.

[Brac93] Brachman, R., P. Selfridge, L. Terveen, B. Altman, A. Borgida, F. Halper, T. Kirk, A. Lazar, D. McGuinness, L. Resnick, "Integrated Support for Data Archaeology", *International Journal of Intelligent and Cooperative Information Systems*, vol. 2, no. 2, pages 159-185, 1993.

[Brac96] Brachman, R., T. Anand, "The Process of Knowledge Discovery in Databases", in *Advances in Knowledge Discovery and Data Mining*, Fayyad, Piatetsky-Shapiro, Smyth, Uthurusamy, eds., MIT Press, pages 37-58, 1996.

[Buja91] Buja, A., J. MacDonald, J. Michalak, W. Stuetze, "Interactive Data Visualization Using Focusing and Linking", *Proceedings of Visualization'91*, pages 156-163, 1991.

[Buja96] Buja, A., D. Cook, D. Swayne, "Interactive High-Dimensional Data Visualization", *Journal of Compueational and Graphical Statistics*, vol. 5, number 1, pages 78-99, 1996.

[Carr86] Carr, D., W. Nicholson, P. Cowley, "Data Analysis Management - Goals and Experience", *in Proceedings of the American Statistical Society*, Stat. Comp. Section, pages 25-31, 1986.

[Clev85] Cleveland, William S., "The Elements of Graphing Data", Wadsworth, 1985.

[Keim94] Keim D. A., Kriegel H.-P., "VisDB: Database Exploration using Multidimensional Visualization",*Computer Graphics and Applications*, Sept.1994, pages 40-49.

[Keim95] Keim D. A., Kriegel H.-P., "VisDB: A System for Visualizing Large Databases", System Demonstra tion, *Proc. ACM SIGMOD Int. Conf. on Management of Data*, San Jose, CA, 1995.

[Keim95a] Keim A., Kriegel H.-P., Ankerst M.: 'Recursive Pattern: A Technique for Visualizing Very Large Amounts of Data', Proc. Visualization '95, Atlanta, GA, 1995, pp. 279-286.

[Keim96] Keim D. A., "Enhancing the Visual Clustering of Query-dependent Data Visualization Techniques using Screen-Filling Curves", *Proc. Int. Workshop on Database Issues in Data Visualization*, Atlanta, GA, 1995.

[Lee90] Lee, A., "A Taxonomy of Uses of Interaction History", *Proceedings of Graphics Interface'90*, pages 113-122, 1990.

[Lee94] Lee, J.P., G.G. Grinstein, "Data Exploration Interactions and the Exbase System", *in Database Issues for Data Visualization*, J.P. Lee and G.G. Grinstein, eds., *Lecture Notes in Computer Science*, Springer Verlag, volume 871, pages 118-137, 1994.

[Lee95] Lee, J.P, G.G. Grinstein, "An Architecture for Retaining and Analyzing Visual Explorations of Databases", *Proceedings of Visualization'95*, pages 101-109, 1995.

[Lee96] Lee, J.P., G.G. Grinstein, "Describing visual interactions to the database: closing the loop between user and data", *Proceedings of SPIE: Data Exploration and Analysis III*, volume 2656, pages 93-103, 1996.

[Long95] Long, D.D.E., Mantey, P.E., Wittenbrink,C. M, Haining, T.R., and Montague, B.R., "The real-time environmental information network and analysis system", *Proceedings of COMPCON'95*, pages 482–487.

[Nich83] Nicholson, W., "Analyzing Large Data Sets: A Challenge for Statistical Computing", *in Proceedings of the American Statistical Society*, Stat. Comp. Section, pages 194-199, 1983.

[Oldf88] Oldford, R., S. Peters, "Statistical Analysis Maps", *University of Waterloo Technical Report STAT-88-21*, Department of Statistics and Actuarial Science, 1988.

[Pang95] Pang, A., Wittenbrink, C., "Spray Rendering as a Modular Visualization Environment", *ACM Computer Graphics*, volume 29, number 2, pages 33 – 36, 1995.

[Rush95] Rushmeyer, H., M. Botts, S. Uselton, J. Walton, D. Watson, H. Watkins, "Metrics and Benchmarks for Visualization", *Proceedings of Visualization'95*, Panel Discussion, pages 422-427, 1995.

[Spri92] Springmeyer, R., Blattner, M., and N. Max, "A Characterization of the Scientific Data Analysis Process", *Proceedings of Visualization'92,* pages 235-242, 1992.

[Sway92] Swayne D.F., Cook D., Buja A.: 'User's Manual for XGobi, a Dynamic Graphics Program for Data Analysis', Bellcore Technical Memorandum, 1992.

[Tukey77] Tukey, J.W., "Exploratory Data Analysis", Addison Wesley, 1977.

[Vell90] Velleman, P., "Computing and Modern (Exploratory) Data Analysis", *Proceedings of the American Statistical Society*, Stat. Comp. Section, pages 46-54, 1990.

[Vell92] Velleman P. F: 'Data Desk 4.2: Data Description', Ithaca, NY, 1992.

[Wilh95] Wilhelm A., Unwin A.R., Theus M.: 'Software for Interactive Statistical Graphics - A Review', Proc. Int. Softstat '95 Conf., Heidelberg, Germany, 1995.

[Youn91] Young, F., J. Smith, "Towards a Structured Data Analysis Environment: A Cognition-Based Approach", *in Computing and Graphics in Statistics*, A. Buja and P. Tukey eds., Springer Verlag, 1991.

This article was processed using the LaTeX macro package with LLNCS style