Directions for Methodological Research in Information Visualization

Brock Craft, Paul Cairns
Institute of Education – London Knowledge Lab, University of York
{b.craft@ioe.ac.uk,pcairns@cs.york.ac.uk }

Abstract

***Abstract**—People within and outside the information visualization community are motivated to create new tools to address their own unique problems of understanding data. However, the techniques which visualizations use to enhance cognition of data are not widely known outside the field. Also, there are currently few resources which comprehensively describe methods for designing novel visualizations. Consequently, people who seek to build new information visualization tools are left to consult design examples, guidelines, and reference models, which do not adequately describe the visualization design process or suggest ways to undertake the process. We identify some of these shortcomings, provide an overview of methodological knowledge in Information Visualization design to date, and suggest that methods from other design-oriented disciplines can bridge these shortcomings.*

Keywords--- **theory development, methodology, reference models, design guidelines**.

1. Introduction

To date, theory development in Information Visualization has focused on defining the boundaries of the domain, identifying crucial concepts, and beginning identification of the relationships among those concepts. More work is needed in all of these areas. Additionally, at this point it would also be useful to identify the area of design methodology as a needed focus for continued theory development. Specifically, there is a lack of descriptive theoretical models regarding the methods of actual design activities. To better delineate Information Visualization as a distinct field, these theories need to be identified and specified leading to a discipline of Information Visualization design as opposed to any other sort of design.

From this perspective, this paper briefly reviews our current understanding of the domain. It is concerned with identifying new directions for developing the theoretical basis for design activities in Information Visualization. It is important to frame this topic by reviewing the boundaries of the domain and description of what models currently exist to explain how visualization systems work. From this basis, an analysis can be made of the shortcomings of visualization reference models. Identifying these allows a case to be made for future research to identify specific design methods.

2. Exposition

There are currently few methodologies that comprehensively describe procedures for creating Information Visualization tools. In particular, techniques for creating interactions and visual representations early in the design process have not been discussed. This leaves people who seek to create new visualizations with few resources to draw on.

Researchers in visualization are currently laying the groundwork to remedy this situation. Current knowledge comes in the form of design Examples, Taxonomies, Guidelines, and Reference Models. These areas of research represent attempts to codify visualization knowledge in a meaningful way, and, implicitly, to assist others in the creation of visualization systems. They do this by giving solutions (Examples), categorizing and listing artefacts (Taxonomies), recommending best practices (Guidelines), and describing how visualization systems work, as a whole (Reference Models). Though each offers an increasing level of depth and robustness, none constitutes a methodology. In terms of creating new visualization systems they describe what and when, rather than how.

Examples comprise the collection of Information Visualization systems that have been reported in the research literature. These are typically the systems that demonstrate new technical or algorithmic solutions [e.g., 27], new visual representations [e.g., 16], and novel interactions [e.g., 1]. Currently, reports about these systems make up the bulk of knowledge in the domain and reported work comes largely from this area. These can serve as a source of design ideas and inspirations for people who seek to design new visualizations, but as they cover very diverse topics, they may not be collected in a useful repository that is pertinent, and may not be relevant to specific design problems.

Taxonomies are attempts to categorize the attributes of visualization systems, applying regular names to visual elements and interactions. They describe ways of classifying information visualization, its characteristics and salient concepts. This body of research involves developing frameworks for organizing and understanding ideas in information visualization or offering categorizations of existing tools. Examples of taxonomies are those proposed by Shneiderman [23], Chi [7], and Tory and Möller [25]. Of these, Shneiderman's taxonomy is particularly important because it has been often cited by others as useful for developing visualizations.

In addition to Taxonomies, many visualization researchers propose Guidelines as a means of describing the important characteristics of visualization systems and when to use them. They are drawn from heuristic impressions of practitioners based on their own experiences and studies. Notable guidelines are offered by Brath [4], Carr [6], Eick [12], Foley and Van Dam [14], Rheingans and Landreth [21], and Shneiderman [23], all of whom have published extensively in the field. However, because guidelines among different authors are not related to one another, they are unstructured and may even be contradictory. For example, in some systems, it may not be appropriate to use a guideline such as, 'provide a simple 3D navigation model' [4], particularly for those having a two-dimensional display. This may seem a rather obvious example, but it is one that highlights the problems brought about by the fact that guidelines rarely describe the context in which they should be used. Other recommendations may be vague or ambiguous. It is difficult to identify precisely what it means to 'emphasize the interesting' [21] or to 'use colour carefully' [4], particularly if a designer has no knowledge of other factors (e.g. colour-blindness) which may bear on a design problem. Still more perplexingly, how is one to know whether 'Visualization is not always the best solution' [6]? As these examples demonstrate, without a structured body of knowledge, it is difficult to build meaningfully on existing knowledge to communicate known solutions in a way that can be shared usefully with other practitioners in the community. In addition, it is hard to engage end-users in the visualization design process using a few scattered examples that are poorly organized or unrelated. Thus, guidelines offer some assistance for visualization design, but have many drawbacks.

Design Patterns [28] go some way to providing the context and integration that guidelines lack but they are currently only proposed for Information Visualization. Their long term value and validity remains to be seen.

Beyond taxonomies and guidelines, Reference Models offer the most thorough descriptions of visualization systems and begin to suggest how designers should go about creating them. These have been drawn from practitioners' experiences in other disciplines, such as software engineering and HCI. In the early days of engineering software systems, ad hoc designs were based upon engineers' intuitive understanding of clients' needs. However, as systems became more complex, it soon became obvious that more considered approaches to software engineering would reduce development costs. This period saw the advent of the Software Development Lifecycle, also known as the 'waterfall' model of systems development, because the results at each stage of development feed into the next [11]. Another approach, the 'Spiral Model' emphasizes an iterative approach to software design, wherein business needs, customer needs, and engineering requirements are continually reassessed as a project progresses [18]. Emerging from these experiences, HCI-centred development models have emphasized user requirements, cognitive and task modelling and interface evaluation. Preece et al. [19] describe a generic model for user-centred design of software systems. This model incorporates an initial phase of user and data requirements gathering, followed by design activities, wherein particulars of the system architecture and user interfaces are created. These can then be used to describe a specification for a prototype, which can be of either a low fidelity to the design concepts (e.g. a paper prototype) or high fidelity, in the form of a functional working prototype. Evaluation of the prototype by informal assessment or usability testing can then suggest design improvements in an iterative and continuous cycle. When the functionality is judged to be adequate, a software release can then be issued. Drawing upon these experiences, a few researchers in the visualization community have begun to propose Reference Models of visualization systems. These Reference Models, which are attempts to codify the components of visualization systems, have differences in emphasis, but their basic components are similar: they all entail an approach which involves manipulation of the visual representation of data by human interaction.

Csinger [10] describes a model for visualization systems drawing upon research in psychophysics, automatic display generation and multi-dimensional data visualization. Csinger's model is a general, high-level abstraction of the major components of a broad range of visualization systems. This model is based upon the work of Ware [26], Roth, and Bertin [3], who articulated the capabilities and limitations of human visual processing, as it relates to abstractions of data. In Csinger's model, real data in the world such as weather-related data are interpreted by a Computational Engine (i.e. a computer) which performs some computation on them. The output is sent to a Display Processor which reduces or alters the dimensions of the data algorithmically in order to match the capacities of human perception and sends them to a Display. This allows a Human User to view and manipulate the Display Processor interactively.

Similarly, Robertson and DeFerrari [20] describe a model which entails the input of data into a visualization system from many possible sources. The data are then transformed according to a set of visual attributes and rendered to the screen. At several points in this process, the human user can intervene to modify either the data or the representations that are encoded and displayed.

Most recently, Card et al., [5] have proposed a basic visualization Reference Model, which attempts to capture the activities involved in Information Visualization design. The model describes the activities that must be completed to create new visualization systems. In this approach, raw data undergo transforms into structured data as data tables, which can be more easily manipulated and so that its features can be identified. This structured data undergoes additional data transforms so that salient derived results can be calculated. These attributes, such as means, frequencies, and other meta-data, describe the data extents and characteristics. These derived attributes then undergo visual mapping transforms wherein the structures inherent in the data can be mapped to abstract visual structures. When graphical views are calculated, the visual structures can be represented by view transformations on the screen, such as changes in shape, colour, size, location, etc. These views can then be altered by human interaction with the system. Such interactions change characteristics of the transformations and mappings so that the visual representation can then be changed to allow exploration of the data.

These reference models attempt to capture the salient features of Information Visualization systems. Each description varies slightly in both terminology and approach, but all of them describe some means of altering the visual representation of a collection of data to be more easily manipulated by the user. Yet, seen from a methodological perspective, the description of how to use the reference model is essentially data-oriented, as can be seen from this excerpt: "The reference model of Information Visualization developed in this chapter approximates the basic steps for visualizing information. The first step is to translate Raw Data to a Data Table, which can then be mapped fairly directly to a Visual Structure. View transformations are used to increase the amount of information that can be visualized. Human interaction with these Visual Structures and the parameters of the mappings create an information workspace for visual sense making." [5]. On the basis of this description alone, it would be easy at design time to place more weight on the nature of the data and to downplay the importance of human interaction.

Although reference models make visualization design methods easier to understand and undertake by breaking up the process into constituent parts, they do not provide guidance in how to create the new visual transformations or interaction activities that comprise a novel visualization. Moreover, they do not address the creativity and problem-solving challenges which occur in the design process.

In addition to these reference models, there are a number of systems which have attempted to automate the creation of visualizations and thus, to make concrete some of the procedural steps of visualization design. Notably Mackinlay [17], Zhou and Feiner [29], and Salisbury [22] have described systems to create visualizations automatically. These systems vary in implementation but all involve automated parsing of a data set, assignment of visual attributes to the data, and provision for manipulation of the representation of these attributes. For example, based on a system of ranking, the software described by Mackinlay could map the position of a quantitative variable to its most effective representation on the screen. Other variables might be represented by colour, texture, shape, etc., depending on their importance. The layout of these elements would be determined algorithmically. Other systems of this type or which offer libraries of techniques are Prefuse [15], The InfoViz Toolkit [13], and the XML visualization toolkit [2].

Because they have a limited repertoire of visual attributes and presentations, automatic visualization systems have a limited set of possible visual representations. For example, the system described by Salisbury in the domain of urban planning offers a specific set of visual abstractions in the form of charts, plots, tables, 2D maps, 3D maps, surface maps and others. When no visualization can be mapped to a single representation, the system presents multiple views. Regardless of the presentation method, however, the visualization system must draw upon a pre-determined set of visual representations. This rules out or tends to limit opportunities for creative explorations of novel and abstract representations.

It is important to emphasize here that one must not confuse specific *techniques* and *design methods*. Techniques are tools in the designer's repertoire but used in isolation they are not design methods. Most of the literature describes new techniques for representation, interaction or both, but authors typically do not describe how they created their visualizations, or what implications this may have for how others should approach visualization design problems.

The situation can be understood by an analogy to architecture. Imagine the task of trying to build a new house. Though we may not pause to examine them closely, we see examples of houses every day. If we were to try to build a new house, these examples would provide a useful source of knowledge about what might be needed. By studying examples we can even learn of new techniques of housing systems that we may want to incorporate into our own design. For example, circulated hot water heating systems have advantages over forced-air heating. Many systems are available for managing sewerage and waste, depending on local requirements. Similarly, visualization designs can be improved by reviewing examples from the many sources in the literature. Our new house will be composed of its constituent parts: a foundation, nails, doors, windows, walls, rooms, etc. A taxonomy describing these parts and can aid us as a reminder of things that might otherwise have been forgotten. Guidelines can also be useful. For example, a room with windows on two sides is often considered to be more pleasant than a room with a single window or none at all. Kitchen and dining areas are typically placed away from sleeping quarters. A house should have at least two exits, in case of fire. Such guidelines, developed from expert practice can suggest effective approaches for addressing design problems and which will lead to a more functional living space. Finally, diagrammatic reference models of homebuilding can describe the functional elements of housing systems, how they work in unison and what elements must exist. A roofing system protects from the elements but also provides structural stability. The façade serves both as an exterior wall and as a public indication of the social function of the building. Similarly for the domain of visualization, the reference models describe the important parts of visualization systems and their relationships.

But just knowing about these key building elements alone will not show us how to build the new house. A description of the parts and systems is not the same as a principled, descriptive approach of how to begin the building process and see it through to fruition. This is somewhat the equivalent of having a kit home delivered to a building site without a set of instructions to do the work.

3. Discussion

The four identified categories of research, Examples, Taxonomies, Guidelines, and Reference Models represent the current state of knowledge in the design of Information Visualization systems. They are useful, in that they are beginning to define the boundaries of the discipline. However, in practical terms, when designers seek to create new visualizations, there are shortcomings. For people who are not intimately familiar with the visualization knowledge domain, the diversity and discontinuity of design knowledge presents a high barrier to understanding how to create new tools. None of these research areas offers a comprehensive and thorough description of how to approach novel visualization design problems. People can use Examples as an inspiration for their own design solutions, adopting a ‘case-based’ design approach, and often do. Spence [24] refers to these as ‘point solutions’. Reports of new visualization systems regularly cite visualization tools which have inspired the reported work. However, examples offer limited help in surmounting the challenges of new designs for visual representation and interaction. They merely present the ‘old favourites’. Taxonomies are only useful in terms of describing the attributes of systems already extant and provide little assistance for designers. Guidelines offer useful suggestions and recommendations of best practice, but they are not unified and they often offer conflicting recommendations. Reference models, which offer the most robust methodological guidance, describe the components of a visualization system which should exist and how those parts should relate to each other. Much of the guidance offered by reference models in terms of actually designing a system is implicit rather than explicit. For example, it is obvious from the visualization reference model of Card et al. [5] that visual mappings between data and on-screen visual structures need to be made during visualization design. But, as with most design examples (such as [15]), little methodological guidance about how to do this is offered in the accompanying text. It is also clear that human interaction needs to figure, in some prominent way, in the design process. Implicitly, the process of knowledge crystallization should be supported. But although their reference model accounts for human tasks, it is left to the designer to interpret which tasks they might perform, how they might perform them, and what components should be made available so that users can achieve their goals. It is up to the designer to fill in the gaps in this high-level model, and indeed, in all of the reference models, even if that designer is inexperienced or unfamiliar with the visualization domain.

Finally, and crucially, the experience and expertise of visualization designers is unaccounted for in all of these areas of research. It is implicit, rather than explicit. Visualization knowledge is captured by successful examples and point solutions. Such knowledge was necessary to generate taxonomies, guidelines and reference models. It is also identified as important by the experts themselves. Experienced designers know the properties of visual representations. Spence notes: “…in the great majority of situations the design of a new visualization tool is a craft activity, the success of which depends upon the designer’s understanding of the task for which the tool is intended, as well as the designer’s possession of many and varied skills ranging from visual design to algorithm design.” Yet the expertise which is apparently necessary to address design problems and generate creative solutions is little described in accounts of the visualization domain. Is this knowledge essential, merely preferable or indeed necessary at all? Moreover, as the domain of Information Visualization practitioners is relatively small, how can this expertise be shared with a larger community, particularly with non-experts, so that visualizations can be beneficial to more people? Rather than relying upon a confusing array of disparate knowledge sources, is there a useful, principled approach which they can use to create successful visualizations?

Understanding how to create visualizations with novel representations and interactions remains a problem. How can this activity be described and supported? What steps are necessary? In addition to a need for a knowledge base, which the current literature provides, there is a need to use design techniques for creativity. Spence presents ‘point solutions’ as a palette of useful techniques which can be extended in new visualization designs. Information Visualization literature serves as a collection of examples which may provide inspiration, but which do not act as a comprehensive guide to solving problems associated with novel visual presentations or user interactions which will enhance knowledge crystallization.

REFER to BELIV here?

Architecture has used design by example for hundreds of years, but software is more changeable and changes faster. Moreover, design activity draws substantially upon knowledge of previous solutions and these are only likely to be known by people who are already experts in the field. Those who are reporting in the literature tend to focus on a description of the new visualization they have produced and not to describe the

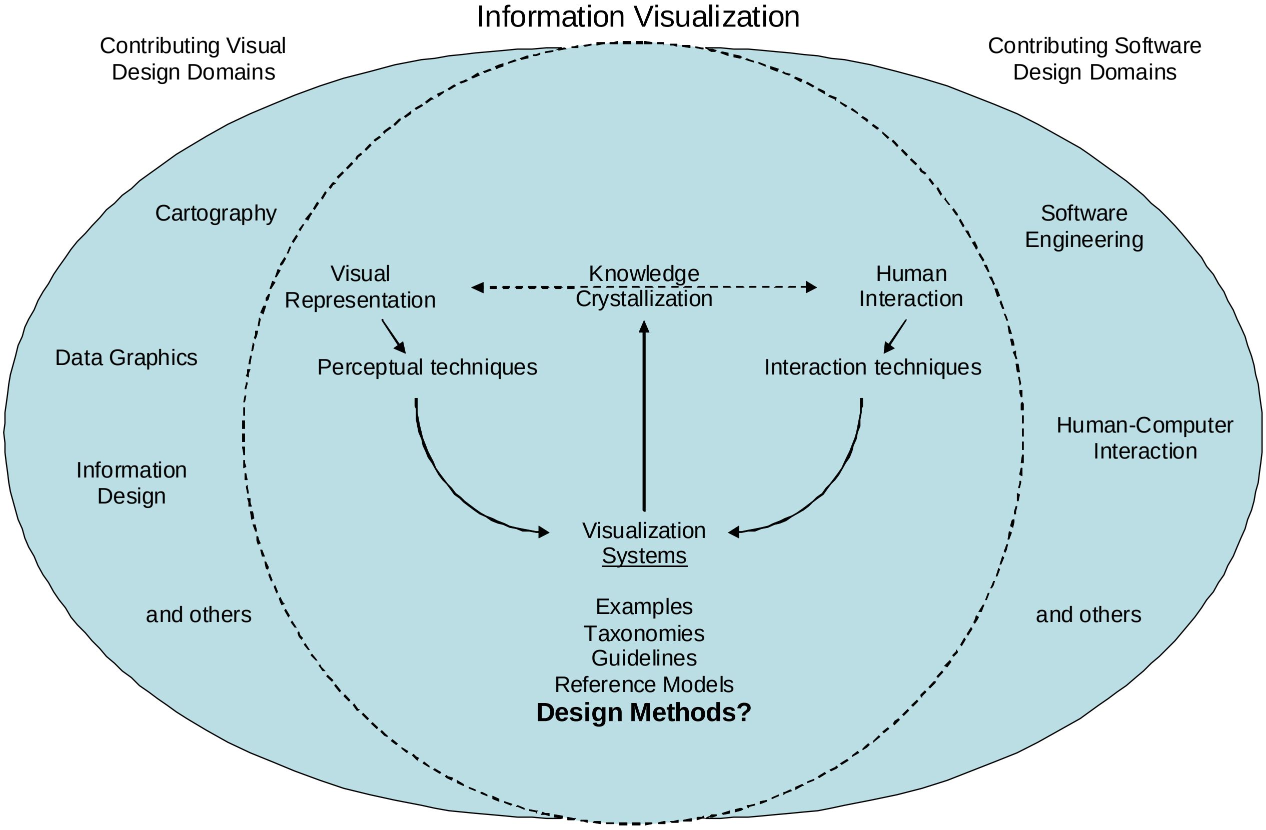

Fig. 1. Theory Development in Information Visualization

entire design process used to generate the new visualization.

This survey has begun to identify some of the shortcomings of existing methodologies of Information Visualization design. In part, the answer is that there is poor representation of the user, and that existing design knowledge is disparate and conflicting. Specifically, principled methods, which lead to designing for knowledge crystallization have not been identified. It could be argued that the design of such systems is so complex that no single methodology could adequately capture all of the necessary design tasks. Indeed, the research until now has concentrated separately on the four separate areas of inquiry discussed above.

The current state of knowledge can be expressed as a knowledge diagram, see Figure 1. The central component of this conceptualization is that, as described by Card, et al., knowledge crystallization is at the core of information visualization software and it is this characteristic which distinguishes information visualisation as a distinct discipline. Knowledge crystallization is supported by the dynamic interaction of users and visual representations. Domains such as Cartography, Data Graphics, Information Design have contributed to our understanding of how data can be visually represented. Traditional Software Engineering and HCI disciplines have contributed to our knowledge of interaction design.

To date, the four knowledge areas we have identified summarise our design knowledge. It has been assumed that the design examples, taxonomies, guidelines, and reference models, which describe the major components of information visualizations, are sufficient as design methods. But we argue that this is not the case. These knowledge areas, while providing a very thorough description of that new house described above, do not set us on the path of beginning its construction. The descriptions are insufficient, particularly for people who are not frequently engaged in visualization design.

Merely grafting techniques from software engineering and HCI onto the reference models also will not work. This is because such grafting does not take into account the particular attributes of information visualization design, which makes it different than other kinds of software. The methods of visual representation and human interaction, which lead to knowledge crystallization, are not accounted for by this approach.

We believe that using methods from other design domains is a promising route to take and will lead to a more general methodology for information visualisation design. We have begun work in this area by combining user-centered design with techniques with those from architecture, graphic design, and traditional engineering disciplines. Our ongoing research [8, 9] has examined the efficacy of participatory design, visualization design patterns proposed by Wilkins [28], and design sketching as useful tools for developing a design methodology for visualization.

A roadmap for developing further methodologies in visualization would be to study the efficacy of design methods used in other design disciplines as applied to the particular problem of knowledge crystallization. We do not suggest that there will be a holy grail, a grand "unified theory" visualization methodology. Instead, in searching for methods, we must learn from other design fields that can contribute meaningfully to the information visualization. Without the prior knowledge areas we have described, which have begun to identify the boundaries of visualisation as a discipline, this would not have been possible.

4. Conclusions

We have presented an overview of the Information Visualization literature, which has constituted the main thrust of research to date. This literature has comprised four main areas: examples, taxonomies, guidelines, and reference models. This has proved a very fruitful vein of knowledge which continues to produce successful systems, as is demonstrated by the increasing number and quality of systems reported in conferences and journals. It is now time to build on this knowledge base to describe clear and detailed methods, which set the domain of visualization apart from other software engineering (and indeed visualization) domains. We believe that other design-oriented disciplines provide the best source of knowledge for defining clear and specific design methods for information visualisation which will be useful for both expert and novice practitioners.

References

[1] Ahlberg, C., Williamson, C., and Shneiderman, B., (1992). "Dynamic queries for information exploration: An implementation and evaluation", Proceedings of ACM SIGCHI'92, pages 619--626.

[2] Baumgartner, J. and Börner, K. (2002). "Towards an XML Toolkit for a Software Repository Supporting Information Visualization Education." In Proceedings of IEEE Information Visualization Conference, (Boston, MA, 2002).

[3] Bertin, J., (1967). Semiologie Graphique. Paris: Gauthier-Villars.

[4] Brath, R., (1999). "Effective Information Visualization: Guidelines and Metrics for 3D Interactive Representations of Business Data", Masters of Computer Science Thesis, Graduate Department of Computer Science, University of Toronto, Canada.

[5] Card, S., Mackinlay, J., and Shneiderman, B., (1999). Readings in Visualization: Using Vision toThink. Morgan Kaufman, San Francisco, California.

[6] Carr, D.A., (1999). "Guidelines for Designing Information Visualization Applications", Ericsson Conference on Usability Engineering '99. Stockholm, Sweden, 1-3 December 1999.Chi, E., (2000). "A taxonomy of visualization techniques using the data state reference model", Proc. of the Symposium on Information Visualization (InfoVis '00), pages 69--75. IEEE Press, 2000. Salt Lake City, Utah.

[7] Chi, E., (2000). "A taxonomy of visualization techniques using the data state reference model", Proc. of the Symposium on Information Visualization (InfoVis '00), pages 69--75. IEEE Press, 2000. Salt Lake City, Utah.

[8] Craft, B., and Cairns, P., "Using Sketching to Aid the Collaborative Design of Information Visualisation Software - A Case Study" , HWID '06 Human Work Interaction Design: Designing for Human Work, Madeira, Portugal 13-15 Feb. 2006.

[9] Craft, B., and Cairns, P., (2006). "Using Sketching to Support Visualisation Design" In British HCI - V&I 2006 (Combining Visualisation and Interaction to Facilitate Scientific Exploration and Discovery). London, UK.

[10] Csinger, A., (1992). The Psychology of Visualisation. University of British Columbia, Department of Computer Science. Technical Report; TR-92-28.

[11] Dix, A., (1998). Human-Computer Interaction (Second edition). Hertforshire, UK: Prentice Hall Europe.

[12] Eick, S.G., (1995). "Engineering Perceptually Effective Visualizations for Abstract Data", In: Nielson, G.M., Hagen, H., Muller, H. (eds.) Scientific Visualization Overviews, Methodologies and Techniques. USA: IEEE Computer Science Press. pp 191-210.

[13] Fekete, J.-D. (2004). "The InfoVis Toolkit". In Proceedings of InfoVis '04, pp. 167-174, 2004.

[14] Foley, J.D., Van Dam, A., (1982). Fundamentals of Interactive Computer Graphics. Boston, Massachusetts: Addison-Wesley.

[15] Heer, J., Card, S. K., and Landay, J. A. (2005). "Prefuse: a toolkit for interactive information visualization." In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Portland, Oregon, USA, April 02 - 07, 2005). CHI '05. ACM Press, New York, NY, 421-430.

[16] Johnson, B., and Shneiderman, B., (1991). "Tree-maps: A Space-Filling Approach to the Visualisation of Hierarchical Information", in IEEE Visualisation 1992, pp 284-291.

[17] Mackinlay, J., (1986). "Automatic the Design of Graphical Presentations of Relational Information", ACM Transactions of Graphics, 5(2), pp 110-141.

[18] Pfleeger, S.L., (1987). Software Engineering: The production of quality software. Indianapolis: Macmillan Publishing Company.

[19] Preece, J., Rogers, Y, Sharp, H., Benyon, D, Holland, S., and Carey, T. (1994). Human-Computer Interaction. Harlow, England: Addison-Wesley.

[20] Robertson, P., and DeFerrari, L., (1994). "Systematic Approaches to Visualisation: Is a Reference Model Needed?" In Rosenblum, R.A. et al., eds. Scientific Visualisation: Advances and Challenges. New York: Academic Press.Roth, S. and Mattis, J., (1990). "Data Characteristization for Intelligent Graphics Presentation", in Proceedings CHI'90, April, pp 193-200. ACM Press.

[21] Rheingans, P., Landreth, C., (1995). Perceptual Principles for Effective Visualizations. In: Grinstein, G., Levkowitz, H., (eds.) Perceptual Issues in Visualization. Berlin: Springer- Verlag. pp 59-73.

[22] Salisbury, L.D.P., (2001). "Automatic Visual Display Design and Creation", PhD Dissertation, University of Washington, Seattle, USA.

[23] Shneiderman, B., (1996). “The eyes have it: A task by data type taxonomy of information visualizations”. Proc. IEEE Visual Languages '96, pp. 336-343.

[24] Spence, R., (2001). Information Visualisation. New York: ACM Press.

[25] Tory, M., Möller, T., (2004). “A Model-based Visualization Taxonomy”, in Proceedings of the 2004 IEEE Symposium on Information Visualization, October 2004.

[26] Ware, C., (2000). Information Visualisation: Perception for design. San Francisco: Morgan Kaufman Publishers.

[27] van Wijk, J. J., and Nuij, W., (2003). “Smooth and Efficient Zooming and Panning”, in Proceedings of the 2003 IEEE Symposium on Information Visualisation, October 2003.

[28] Wilkins, B., (2003). “MELD: A Pattern Supported Methodology for Visualisation Design”, PhD Dissertation, University of Birmingham, UK.

[29] Zhou, M.X. and Feiner, S.K., (1998). “Data Characterisation for Automatically Visualizing Heterogeneous Information”, Conference on Human Factors in Computing Systems. Los Angeles, Ca., USA, 28-29 October 1996. pp 392-399.