

Heuristics for Information Visualization Evaluation

Torre Zuk, Lothar Schlesier, Petra Neumann, Mark S. Hancock, Sheelagh Carpendale

University of Calgary
2500 University Dr. NW
Calgary, Canada T2N 1N4

{zuk,lothars,pneumann,msh,sheelagh}@cpsc.ucalgary.ca

ABSTRACT

Heuristic evaluation is a well known discount evaluation technique in human-computer interaction (HCI) but has not been utilized in information visualization (InfoVis) to the same extent. While several sets of heuristics have been used or proposed for InfoVis, it is not yet known what kind of heuristics are useful for finding general InfoVis problems. We performed a meta-analysis with the goal of exploring the issues of heuristic evaluation for InfoVis. This meta-analysis concentrates on issues pertaining to the selection and organization of heuristics, and the process itself. For this purpose, we used three sets of previously published heuristics to assess a visual decision support system that is used to examine simulation data. The meta-analysis shows that the evaluation process and results have a high dependency on the heuristics and the types of evaluators chosen. We describe issues related to interpretation, redundancy, and conflict in heuristics. We also provide a discussion of generalizability and categorization of these heuristics.

1. INTRODUCTION

Heuristic evaluation is a discount evaluation method commonly used to find usability problems at different development stages of a product. A heuristic evaluation involves a small number of evaluators inspecting a system according to heuristics or guidelines that are relevant for the system. Heuristics exist as shared or general knowledge on design. They often can act as instructional guides for the teaching of novices and can evolve into design patterns for construction such as those that exist for software engineering. They aid in the communication of ideas by providing a common language and promote reuse of proven methods or concepts [10]. Other heuristics can be more general and act as a check on design choices. As heuristic evaluation is a light-weight process that can be cheap, fast, and easy to apply, it has potential for integration within development iterations. It can be used both in design and evaluation phases of develop-

ment and can even be applied to paper-based designs before the first working prototype is created.

While heuristic evaluation has been part of the HCI set of evaluation tools for some time [17], it has not been utilized or examined for evaluating InfoVis to the same extent. Granted usability issues also arise in these systems, but are not the only problems that these systems may have. We discuss issues that call for different or supplemental sets of heuristics for the discount evaluation of InfoVis systems. Utilizing a few sets of previously published heuristics for these systems we analyze *LuMPB Key* (Landscape unit Mountain Pine Beetle Key [18]), a visual decision support system, that is used to examine simulation data, as a case study to demonstrate their application. We assess the value of the used heuristics and suggest implications for further research of the process of heuristic evaluation in InfoVis.

2. HEURISTICS IN INFOVIS

The field of information visualization is influenced by many different research domains including psychology, semiotics, graphic design, and art. The goal of an information visualization is generally defined as providing useful tools and techniques for gaining insight and understanding in a dataset, or more generally to amplify cognition [5]. These are high-level cognitive issues that are hard to measure with quantitative user studies. Tory and Möller in their summary of expert reviews recommend the use of heuristic evaluation for analyzing visualization systems [23]. While usability heuristics, as known from HCI, encompass a wide variety of issues pertaining to visualizations and the interaction with them, we believe that more specific heuristics are of value, in particular since a wide variety of research fields are concerned.

Previous evaluations in InfoVis have proposed heuristics specific to a certain data domain, e.g. for ambient displays [14] or multiple view visualizations [2], for a specific cognitive level based on knowledge and task [1], or based on perception and cognition [26]. Shneiderman's well known "Visual Information-Seeking Mantra" [20] has also been used for heuristic evaluation of information visualizations based more on task and usability (for an overview see [6]). Tory and Möller propose to use heuristics based on both visualization guidelines and usability [23]. At this stage of development of heuristics for information visualization we have reached a similar problem as described by Nielsen in 1994 [17]. It is a difficult problem to assess which list(s) are better for what reasons and under what conditions. This leads to the challenges of developing an optimal list that comprises the most important or common InfoVis problems. Visual rep-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

BELIV 2006 Venice, Italy

Copyright 2006 ACM 1-59593-562-2/06/05 ...\$5.00.

resentation, presentation, and interaction and manipulation of the parameters that build a visualization play a role in the success or failure of the overall high-level goal to amplify cognition. The above mentioned evaluations used different heuristics and methods to evaluate their criteria. They also suggest that data or visualization types and domain specific information processing tasks are a factor for the evaluation of InfoVis systems. Whether it will be possible to find a small set of heuristics that find the most common visualization problems, similar to Nielsen’s [17], is an exciting open problem for the community.

3. DETERMINING A HEURISTIC SET

Although there are several lists of usability heuristics which do apply to information visualization (not just to the user interface) [22, 17, 12], there are fewer specifically tailored to the visual aspect [1, 26, 19]. How to decide the optimal or even appropriate heuristics is the question.

A *hierarchical* or *taxonomic* way of grouping may aid in selecting an appropriate set of heuristics. A tree-traversal-like approach could be used in which a depth-first search is performed with pruning occurring if the more general heuristics are not appropriate. Morse et al. also pruned an extensive task taxonomy to create a test set (for evaluation-question generation) using the rationale “sample as broadly as possible rather than deeply, and select those which varied significantly” [15]. This organization could lead from a more general heuristic, such as “consider the implications of colour encoding”, to child heuristics such as “color perception varies with size of colored item” [25], or “don’t expect a reading order from color” [3]. The heuristics at the leaf level would likely be “chunked” by experts so that they only need to descend to the more general heuristics to trigger the set of considerations they feel appropriate, but would serve a teaching role to novices. One such possible tree organization is shown in Figure 1.

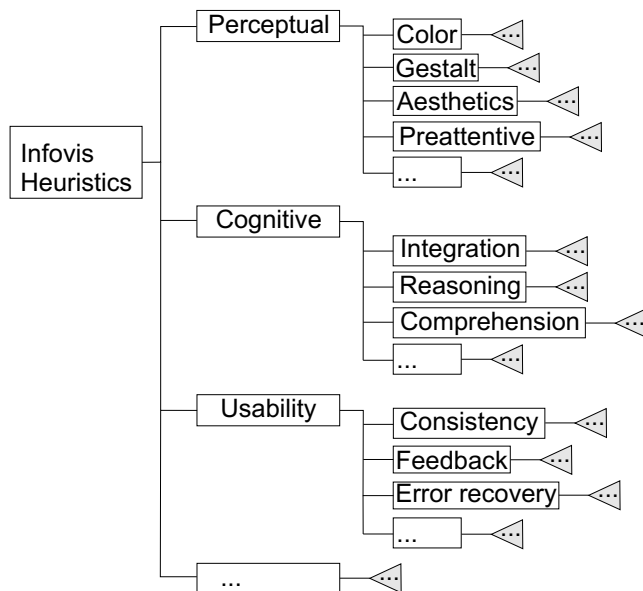


Figure 1: Evaluation Tree.

Another approach is to empirically determine a *minimal* set of heuristics. Nielsen [17] describes a method of refinement of a large set of usability problems into a small set of 10 heuristics that are intended to be general and easily understandable.

4. DETERMINING A PROCESS

The process of heuristic evaluation may evolve just as the heuristics themselves can evolve over time. The original presentation of heuristic evaluation for usability proposed at least two passes of an interface: the first pass to provide a general feel, and the second pass for the application of all heuristics to each interface item [16]. While we initially want to learn from accepted practices we do not want to limit ourselves to that process as the nature of the problem is in some ways fundamentally different.

HCI studies showed that using five evaluators may be enough to find most usability problems, adding more would reduce the benefit to cost ratio, and suggested that three may suffice [16]. More recently Spool and Schroeder [21], and a CHI panel [4] reviewed how many evaluators are required for web site usability analysis. They found for some problems more than five are likely needed to find the majority of problems, and the exact number will likely be product specific. Because the use of heuristics in InfoVis has not yet been fully studied, it is still uncertain if this knowledge will transfer. We can only suspect that for evaluating information visualizations, the required number of evaluators to guarantee finding most problems may also be visualization specific. In heuristic evaluation for usability, as performed in HCI, the evaluators are commonly usability specialists. It still has to be determined, however, what is required of an “InfoVis specialist” when applying a heuristic evaluation. Tory and Möller suggest to use both visualization (data display) and usability experts [23]. What knowledge is required of a “visualization specialist” will have to be discovered.

While evidence has shown for usability that a small set of heuristics can find a majority of problems, we as yet have no evidence for a similar potential from visualization heuristics. Craft and Cairns [6] recently undertook the process of analyzing the heuristics of the “Visual Information-Seeking Mantra”. They reviewed others’ use of the “Mantra” and found a lack of empirical evidence validating the heuristics. They noted that even though the heuristics were presented as descriptive in nature they have been used as prescriptive [6]. They conclude by calling for a more rigorous design methodology that: takes into account the useful techniques that guidelines and patterns suggest, has measurable validity, is based upon a user-centred development framework, provides step-by-step approach, and is useful for both novices and experts.

Kahn and Prail have provided a set of design heuristics [12] to help design the evaluation process itself. These are: minimize time cost to engineers who are on the critical path, maximize involvement of engineers who will implement changes, create a method that is an “event” in the usability life-cycle, team-based approach, adapt existing method (i.e. help do what is done better), leverage the language and structure of well-established methods solving similar problems, task orientation, and clear potential integration with other parts of the usability engineering life-cycle. There is a danger in assuming too much in reusing the process of heuristic evaluation from usability, therefore we should consider using these

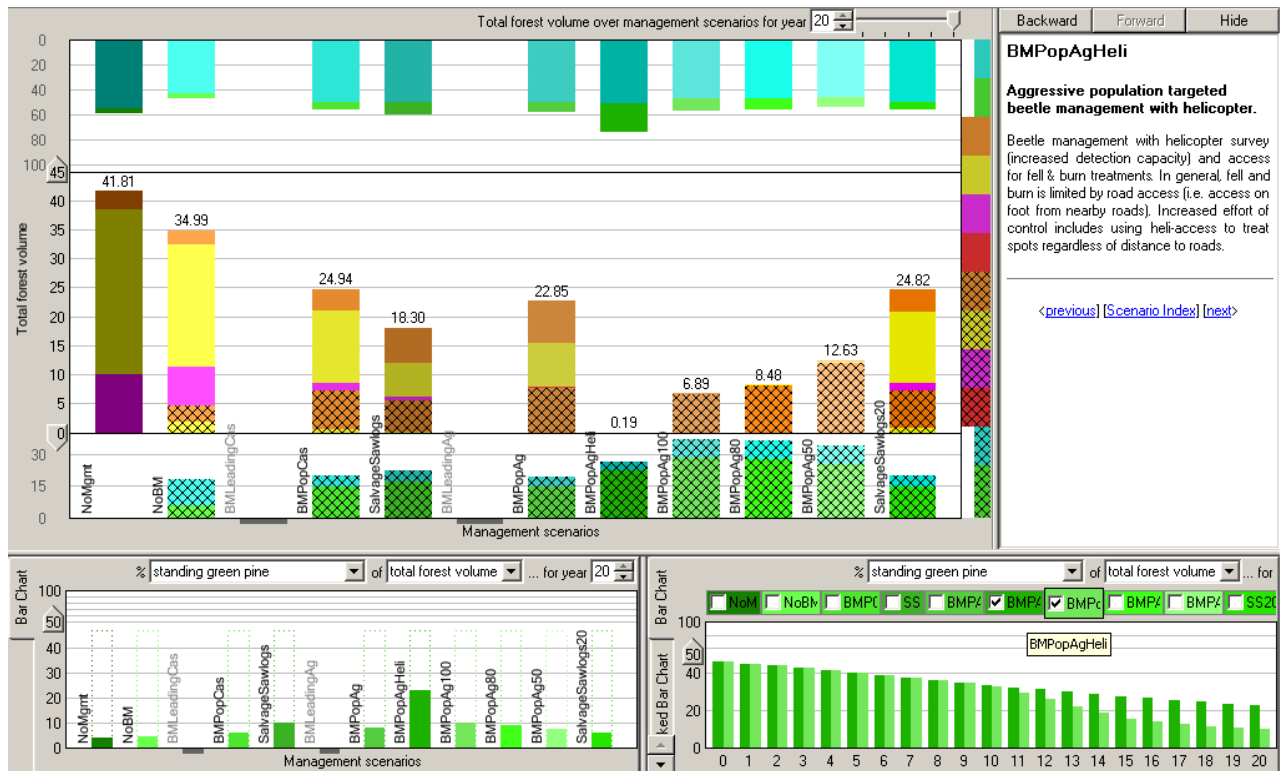


Figure 2: Screen shot of the *LuMPB Key* tool: showing different views on a data set and a text providing context information about a management scenario.

heuristics to re-evaluate the process in its application to information visualizations.

5. CASE STUDY

In order to study the understandability and applicability of a set of heuristics and explore a methodology, we performed a heuristic evaluation of a visualization of simulation data for measuring the impact of mountain pine beetles (MPB) on forests.

5.1 Method

Our method involved applying three different and distinct sets of heuristics to a single visualization, then analyzing the evaluation results individually, followed by a discussion between all evaluators. The discussion included both an analysis of the individual findings and a meta-analysis of the heuristics and process. The discussion was based on the specific findings, but actively considered the ability to generalize. Rather than considerations for pursuing a high-quality evaluation (high percentage of all problems found), our methodology was chosen to support the meta-analysis.

5.1.1 Evaluators

Four computer science graduate students in the Interactions Lab at the University of Calgary each independently performed a heuristic evaluation of a single visualization in the *LuMPB Key* tool. One student was the lead programmer of the visualization tool, two were Ph.D. students in Information Visualization and the fourth was a Ph.D. student in human-computer interaction. One of the Ph.D. students was

the co-author of Zuk and Carpendale’s [26] heuristic evaluation paper and so had experience in applying their heuristics. All four evaluators are authors of this paper. Note that these evaluators were chosen for the purpose of generating valuable discussion in the meta-evaluation and *not* to appropriately evaluate this specific system.

5.1.2 System

The *LuMPB Key* simulation tool can be used to visualize complex simulation data created with the Spatially Explicit Landscape Event Simulator (SELES) [8, 9]. In these simulations mountain pine beetle impact on forest is observed for various conditions. One goal of the simulations is to see which forest management strategy is best to protect pine trees. Our heuristic evaluation focused on a particular pair of visualizations within the *LuMPB Key* system.

A stacked bar chart is used to display the relative proportions of tree types (e.g. amount of cumulative logged pine trees) in the forest over different management scenarios for a given year (upper left part of Figure 2). Bar charts are used to display a single tree type over management scenarios for a given year (lower left area of Figure 2), or to show a time series for a tree type for one or more scenarios (lower right area of Figure 2). Furthermore, text describing management scenarios or tree types can be brought on to the screen.

5.1.3 Procedure

Each evaluator was asked to identify both positive and negative aspects of two specific visualizations (views) in the *LuMPB Key* system, based on three sets of heuristics. The

Table 1: Heuristics applied in evaluation.

Set	Heuristics
Zuk and Carpendale’s <i>Selection of perceptual and cognitive heuristics</i> [26]	Ensure visual variable has sufficient length [3][25][26] Don’t expect a reading order from color [3][25][26] Color perception varies with size of colored item [25][3][26] Local contrast affects color & gray perception [25][26] Consider people with color blindness [25][26][22] Preattentive benefits increase with field of view [3][25][26][11] Quantitative assessment requires position or size variation [3][26] Preserve data to graphic dimensionality [24][3][26] Put the most data in the least space [24][26] Remove the extraneous (ink) [24][26] Consider Gestalt Laws [25][26] Provide multiple levels of detail [24][25][26] Integrate text wherever relevant [24][25][26]
Shneiderman’s <i>“Visual Information-Seeking Mantra”</i> [20]	Overview first [20] Zoom and filter [20] Details on demand [20] Relate [20] Extract [20] History [20]
Amar and Stasko’s <i>Knowledge and task-based framework</i> [1]	Expose uncertainty [1] Concretize relationships [1] Determination of Domain Parameters [1] Multivariate Explanation [1] Formulate cause & effect [1] Confirm Hypotheses [1]

visualizations analyzed are the two views on the left side in Figure 2. The user has the ability to swap the positions to bring one into focus (in order to get more details).

The first set of heuristics were Zuk and Carpendale’s [26] selection of perceptual and cognitive heuristics. These heuristics were chosen because they were designed to be used as heuristics and have been in practice. Shneiderman’s “Visual Information-Seeking Mantra” [20] were chosen as the second because they too have been used to evaluate information visualizations, even though they were not designed this way. Amar and Stasko’s knowledge and task-based framework [1] was chosen as the third set because they were designed to be used to evaluate (and design) information visualizations, but (to our knowledge) evidence for their use in evaluation has not been published. The heuristics are listed in Table 1; detailed descriptions are available in the original papers. Each set of heuristics was to be considered separately in the order shown in Table 1.

5.2 Discussion

Our analysis (meta-analysis) was performed by reviewing as a group all of the individual evaluation results. We proceeded through the heuristics in the order that they were applied looking for commonality, discussing problems found, problem solutions, and to a lesser degree positive findings. At a higher level we also discussed problems and generalizations and what could be improved in the heuristics and the evaluation process.

5.2.1 Heuristics for Communicating Patterns

One aspect of heuristics as design patterns is the communication of ideas. However, we found there existed a variety of interpretations of the heuristics across the four evaluators.

Placing Bertin’s definitions in the perceptual-based heuristics was particularly problematic, as the strict separation of perception from cognition and/or symbolism was not usually maintained. While the heuristics were described in more detail in the original papers, only the summary heuristic was provided as a cue for the evaluation. As the heuristics will likely evolve along with the considerations they evoke, tying a concise description to a heuristic will be helpful. Creating consistency of definitions across the community of practice would help in general usefulness and in the possibility of meta-comparisons. This will also aid in the communication and transfer of knowledge from the findings.

The generally high specificity of Zuk and Carpendale’s heuristics was also called into question. Loosely defined terms and more general wording in a heuristic may allow the flexibility in interpretation needed to catch a broader range of related problems. For example, the “preattentive benefits increase with field of view” heuristic was considered too narrow, with a potential replacement being “use preattentive visual variables wisely”.

5.2.2 Redundancy

The three different groups of heuristics did at times find the same problem from different perspectives. If the main goal of the heuristics is to identify problems then redundant coverage goes against the goal of a minimal set of heuristics. However, if the intention is to also indicate possible solutions to the problems, then finding the same problem via different heuristics can suggest different solutions. Instead of redundancy we can consider that heuristics may support each other by revealing the same problem from different standpoints. In our case study, *details on demand* and *integrate text where relevant* are an example where two

heuristics pointed out the same problem and the same solution. Both revealed that tool tips could be used to display the mean values and standard deviations in the stacked bar chart.

5.2.3 *Conflicting Heuristics*

Heuristics, especially from different sets, may also in some ways contradict each other. This leads to the consideration of trade-offs in the design and it needs to be determined which heuristic has a higher priority. Stakeholders (commonly the domain experts) may also have the right to override heuristics based on domain knowledge or other constraints. For example, colours for the stacked bar chart in the evaluated system were chosen by the domain experts to reflect common usage and could, therefore, not be changed to account for colour-blindness. This domain-dependent weighting of heuristics also creates the variability which adds difficulty in producing a minimal set.

5.2.4 *Heuristic Taxonomy*

Our case study was a preliminary exploration of how we might develop a set of appropriate heuristics for evaluation of information visualization. We are not yet at the stage of producing a taxonomy, but our combined evaluations led to a discussion of how best to organize the heuristics to provide experts with a better “chunked” representation (and thus improved recall) of potential problems to look for. One suggested categorization was to organize the heuristics according to their applicability to *perception*, *usability*, and *discovery process*. In particular, we found it useful to think of the *LuMPB Key* system by separating our criticism into these three aspects. Specifically, Zuk and Carpendale’s [26] were most useful for evaluating perception, Schneiderman’s heuristics were most useful for evaluating usability [20], and Amar and Stasko’s heuristics were most useful for evaluating the discovery process [1]. However, there was significant overlap between these sets in terms of this categorization.

5.2.5 *Generalizable Problems*

Our preliminary exploration also involved significant discussion of some problems with the *LuMPB Key* system that may be common to other information visualizations. Some of these problems included difficult-to-see visual components due to contrast issues, assignment of colour value resulted in confusion or difficulty to perceive relationships, lack of detailed information in “tooltips”, and many more. In the same way that Nielsen [16] refined a set of usability problems into a small set of heuristics, both to cover all problems found and to cover all *serious* problems found, repeating our process with several other information visualizations could provide this same data set and allow the same form of analysis.

5.2.6 *Process*

Amar and Stasko’s heuristics were found by most evaluators to be difficult to apply without extra domain knowledge. It may generalize that one set of heuristics will benefit most from domain expert involvement, or a particular part of the design life-cycle. Broader heuristics such as Amar and Stasko’s may also lend themselves more toward use in design than evaluation, as they may have major implications for system requirements that need to be addressed earlier in the development process in order to reduce costs.

Higher level heuristics such as Schneiderman’s and Amar and Stasko’s tended to require consideration of additional visualizations the system provided, or the system as a whole, for proper application. Therefore, in our attempt to restrict evaluation to a couple of views, the use of these heuristics led most evaluators to questions about the views not analyzed. One evaluator commonly included another view to aid in the application of the heuristics, while the system developer could not help but consider the entire system. Lower-level heuristics may thus work better when analyzing a decomposed larger system. In order to minimize learning both a complex visualization tool and the related domain knowledge, one could borrow from Extreme Programming and have a domain expert and evaluator work in a pair.

Usability issues were often tied to a detected visualization problem, so a set of usability heuristics would have been a useful addition. With the addition of other sets of potential heuristics some organization may be necessary. This leads to the problem of heuristic selection and whether partitioning a larger set of heuristics is useful, both of which will require further research.

One of the evaluators used supplementary software while applying a heuristic. Screen shots of the charts were automatically recolored to test how a colour blind person would see them [7]. This finding raises the question of if and how tools may support heuristic evaluation. The use of tools for evaluation is related to the automatic design of visualizations based on heuristics, such as Mackinlay’s system for relational information [13] using formal expressiveness and effectiveness criteria.

6. CONCLUSIONS

Our meta-analysis has added to the understanding of using different sets of heuristics for evaluation of information visualizations. The approach of using three different sets of heuristics provided useful results for our case study and revealed some characteristics, such as redundancy and conflict, that may be generally useful when comparing different heuristics. We found value in using visualization-specific heuristics, as problems were found that would not have been discovered by usability heuristics.

Many problems we found crossed theoretical and knowledge boundaries, and therefore the evaluation process would benefit from including experts from visualization, usability, and the domain area. Information Visualization’s focus on amplifying cognition means that heuristics related to higher level cognitive tasks such as Amar and Stasko’s [1] delve into issues that only the domain expert may understand. These higher-level issues also require a holistic evaluation of entire systems and so do not lend themselves to a strategy of divide and conquer.

Both finding an appropriate taxonomy of heuristics and finding a minimal set of heuristics that can find the majority of problems or provide the best guidance will require a large amount of research. This work is the first steps toward these goals and, on this path, we intend to bring a better understanding of the types and frequency of problems. During this research, it may be useful to continually look at different organizations of heuristics and different processes which may be more efficient in finding problems and suggesting solutions.

7. ACKNOWLEDGMENTS

We would like to thank Alberta Ingenuity, Alberta's Informatics Circle of Research Excellence (iCORE), Natural Sciences and Engineering Research Council (NSERC), and Veritas DGC Inc. for funding.

8. REFERENCES

- [1] R. Amar and J. Stasko. A Knowledge Task-Based Framework for Design and Evaluation of Information Visualizations. In *Proc. of IEEE InfoVis*, pages 143–149, Los Alamitos, USA, 2004. IEEE Press.
- [2] M. Q. W. Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for Using Multiple Views in Information Visualization. In *Proceedings of AVI '00*, pages 110–119, New York, USA, 2000. ACM Press.
- [3] J. Bertin. *Semiology of Graphics*. University of Wisconsin Press, 1983. Translation: William J. Berg.
- [4] N. Bevan, C. Barnum, G. Cockton, J. Nielsen, J. Spool, and D. Wixon. The “Magic Number 5”: is it enough for Web Testing? In *CHI Extended Abstracts*, pages 698–699, New York, USA, 2003. ACM Press.
- [5] S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, San Francisco, CA, USA, 1999.
- [6] B. Craft and P. Cairns. Beyond Guidelines: What can we learn from the Information Seeking Mantra? In *Proceedings of IV*, pages 110–118, Los Alamitos, CA, USA, 2005. IEEE Press.
- [7] R. Dougherty and A. Wade. Vischeck. <http://www.vischeck.com/vischeck/vischeckImage.php>, 2006. visited: March, 2006.
- [8] A. Fall and J. Fall. SELES: A Spatially Explicit Landscape Event Simulator. In *Proc. of GIS and Environmental Modeling*, pages 104–112. National Center for Geographic Information and Analysis, 1996.
- [9] A. Fall and J. Fall. A Domain-Specific Language For Models of Landscape Dynamics. *Ecological Modelling*, 137(1):1–21, February 2001.
- [10] E. Gamma, R. Helma, R. Johnson, and J. Vlissides, editors. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, 1994.
- [11] C. G. Healey. On the Use of Perceptual Cues and Data Mining for Effective Visualization of Scientific Datasets. In *Proc. of GI*, pages 177–184, 1998.
- [12] M. J. Kahn and A. Prail. Formal usability inspections. In *Usability Inspection Methods*, pages 141–171. John Wiley & Sons, Inc., 1994.
- [13] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)*, 5(2):110–141, 1986.
- [14] J. Mankoff, A. K. Dey, G. Hsieh, J. Kientz, S. Lederer, and M. Ames. Heuristic Evaluation of Ambient Displays. In *Proc. of CHI '03*, pages 169–176, New York, NY, USA, 2003. ACM Press.
- [15] E. Morse, M. Lewis, and K. A. Olsen. Evaluating Visualizations: Using a Taxonomic Guide. *Int. J. Human-Computer Studies*, 53:637–662, 2000.
- [16] J. Nielsen. Heuristic Evaluation. In J. Nielsen and R. L. Mack, editors, *Usability Inspection Methods*, pages 25–62. John Wiley & Sons, Inc., 1994.
- [17] J. Nielsen and R. L. Mack. *Usability Inspection Methods*. John Wiley & Sons, Inc., 1994.
- [18] L. Schlesier, J. Hughes, A. Fall, and M. S. T. Carpendale. The LuMPB Key: A Multiple View Interface to Explore High Dimensional Mountain Pine Beetle Simulation Data. In *Proc. of CMV*, pages 31–41. IEEE Computer Society, 2006.
- [19] B. Shneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley, Reading, MA, 1987.
- [20] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proc. of the IEEE Symposium on Visual Languages*, pages 336–343. IEEE Press, 1996.
- [21] J. Spool and W. Schroeder. Testing Web Sites: Five Users is Nowhere Near Enough. In *CHI '01 Extended Abstracts*, pages 285–286, New York, NY, USA, 2001. ACM Press.
- [22] B. Tognazzini. First Principles of Interaction Design. <http://www.asktog.com/basics/firstPrinciples.html>, 2006. visited: March, 2006.
- [23] M. Tory and T. Möller. Evaluating Visualizations: Do Expert Reviews Work. *IEEE Computer Graphics and Applications*, 25(5):8–11, September/October 2005.
- [24] E. R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 2001.
- [25] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, 2nd edition, 2004.
- [26] T. Zuk and M. S. T. Carpendale. Theoretical Analysis of Uncertainty Visualizations. In *Proc. SPIE & IS&T Conf. Electronic Imaging, Vol. 6060: Visualization and Data Analysis 2006*, 2006. 606007.