

The Challenge of Information Visualization Evaluation

Catherine Plaisant
Human-Computer Interaction Laboratory
University of Maryland
A.V. Williams Building, College Park, MD 20742, USA
+1(301)405-2768
plaisant@cs.umd.edu

ABSTRACT

As the field of information visualization matures, the tools and ideas described in our research publications are reaching users. The reports of usability studies and controlled experiments are helpful to understand the potential and limitations of our tools, but we need to consider other evaluation approaches that take into account the long exploratory nature of users tasks, the value of potential discoveries or the benefits of overall awareness. We need better metrics and benchmark repositories to compare tools, and we should also seek reports of successful adoption and demonstrated utility.

Categories and Subject Descriptors

H.5.2 [Information interfaces and presentation] User Interfaces - Evaluation/methodology; Graphical user interfaces (GUI)

General Terms

Visualization, Measurement, Performance, Human Factors.

Keywords

Evaluation, user studies, usability, usefulness, return on investment, adoption, technology transfer

1. INTRODUCTION

As the field of information visualization matures, the tools and ideas described in our research publications are reaching users [1]. Information visualization is moving out of research laboratories with a growing number of commercial products (such as those from Spotfire, Inxight, and HumanIT), additions to statistical packages (SPSS/SigmaPlot, SAS/GRAPH, and DataDesk) and commercial development environments (e.g. ILOG JViews). The general public is being exposed to visualizations with SmartMoney financial maps, or the HiveGroup' Peet's Coffee Selector (Fig. 1). Citizens can find health information with interactive maps [2] and view real-time highway traffic information [3].

Visualization is also "infiltrating" main stream applications in subtle ways, for example with the additions of timelines in Adobe Photoshop Album or Microsoft Outlook. Could the modest pace of adoption be increased by more evidence supporting the utility of information visualization? Could we do a better job at guiding potential adopters toward successful applications?



Figure 1. E-commerce users are getting exposed to information visualization. A marketing survey showed that 92% of Peet's Coffee and Tea customers who used the visual treemap interface thought shopping online was easy, opposed to only 12% of those who used the textual lists. (www.peets.com)

The reports of usability studies and controlled experiments are helpful to understand the potential and limitations of our tools. We have been able to demonstrate faster task completion and reduced error rates measured in laboratory settings but how could we further make the case for information visualization to potential adopters? Information visualization is still a novelty for many users who are still struggling to use simple business graphics effectively. However our research community cannot afford to argue that managers are "just not ready to make the jump" as have promoters of many failed innovations in the past. Instead, we need to understand how to improve our methods of evaluation in order to present actionable evidence of measurable benefits that will encourage more widespread adoption.

In this paper we summarize current evaluation practices, review challenges specific to information visualization, and propose initial steps such as the development of benchmarks and repositories, refined evaluation methodologies and toolkits, and

the dissemination of success stories. We then investigate three examples of transformations from prototype to product. These examples highlight components and features needed to increase the utility of information visualization tools, which in turn might increase the likelihood of adoption.

2. CURRENT EVALUATION PRACTICES

In a survey of the literature using about fifty user studies of information visualization systems [4] the authors found four thematic areas of evaluation:

- 1: *Controlled experiments* comparing *design elements*. The studies in this category might compare specific widgets (e.g. comparing alphaslider designs [5]) or compare mappings of information to graphical display [6]
- 2: *Usability evaluation* of a tool. Those studies might provide feedback on the problems users encountered with a tool and show how designers went on to refine the design [7, 8]
- 3: *Controlled experiments* comparing *two or more tools*. This is a common type of study. For example we compared three tree visualization tools: SpaceTree, Hyperbolic and Window Explorer [9]. Those studies usually try to compare a novel technique with the state of the art.
- 4: *Case studies* of tools in realistic settings. This is the least common type of studies, e.g. [10]. The advantage of case studies is that they report on users in their natural environment doing real tasks, demonstrating feasibility and in-context usefulness. The disadvantage is that they are time consuming to conduct, and results may not be replicable and generalizable.

3. CHALLENGES

Geoffrey Moore describes the adoption of new technologies in his books *Crossing the Chasm* (1991) [11] and *Inside the Tornado* (1995) [12]. He describes the process of innovators getting excited about novel technologies and successfully selling the idea to a limited number of early adopters. The chasm is the difficulty in reaching the much larger group, which he calls the early majority. Early adopters are visionaries who enjoy new tools and want to try out all the features. The early majority are pragmatists who want something that is reliable, proven, and solves real problems. One of our roles as information visualization researchers is to provide convincing evidence of utility, which is difficult for any new technology, but presents specific challenges for information visualization.

3.1 Matching tools with users, tasks and real problems

Is a Segway a better vehicle? Questions race to your mind: Better than what? For whom? To go where? In which state of mind? How important is it to get there quickly, or to get there at all? What if you don't really know where you are going, what is the best vehicle then? Should I use my savings to buy a Segway when I have other needs? Usability studies and formal comparison of speed characteristics and incident data might help worried potential drivers but it is their judgment of utility that will likely trigger adoption.

Usability of information visualization tools can be measured in a laboratory however, to be convincing, utility needs to be demonstrated in a real setting, that is a given application domain and set of users. For researchers, choosing and preparing convincing examples goes a long way in attracting potential adopters. Using real datasets with more than a few items, and demonstrating realistic tasks is important.

When utility has been demonstrated in one setting, potential adopters often still need to jump across application domains and envision how the tool could be used. What could we do to encourage this leap of faith? Paradoxically our quest for generality can become an impairment. Researchers are encouraged to – and rewarded for – designing techniques that are generic in nature, can be used with a wide variety of data, and used in many application domains. However, potential adopters might be turned off if they perceive that the tool they are evaluating has not been designed specifically for their particular needs: biologists are looking for biology tools and petroleum engineers will be attracted to tools developed for petroleum engineering. Researchers should consider investing some resources to produce tailored implementations, and establish partnerships that allow them to publish in journals and magazines of the chosen application domains.

3.2 Improving user testing

Usability testing and controlled experiments remain the backbone of evaluation [13]. The reports of such laboratory studies help potential adopters understand the potential and limitations of the tools. They also guide researchers trying to improve their tools. In fact, laboratory user studies often are the only time researchers and developers spend observing users using their tools. Statistically significant differences in time, error rate or satisfaction are obviously a plus, but observations recorded during the tests become the basis for refinement or redesigns, leading to better implementations, guidelines for designers and the refinement of theories. These benefits are very appealing to researchers but much less for potential adopters who are left wondering what performance would be with the improved interface. Most importantly potential adopters are often bothered by the constraints imposed by practical user testing. They are left wondering what would have happened if subjects had been trained longer (or not trained at all), had been given more time to explore the data at their own pace and consult with colleagues, had been using their own data and been able to choose the tasks themselves? Including practical summaries for practitioners in research papers is helpful. Reporting informally on usage and performance by developers can shed some light on the potential utility of the tool and performance of trained users.

Specific aspects of empirical studies could be improved. For example empirical evaluations generally include only simple tasks. The literature survey mentioned above [4] confirms this fact by stating that experiments usually include locate and identify tasks, but that tasks requiring users to compare, associate, distinguish, rank, cluster, correlate or categorize are rarely covered. Another problem lies in the fact that comparative studies often report overall performance for a combined set of tasks while our experience suggests that reporting results per task is preferable. In all our visualization experiments, tools performed differently for different tasks, e.g. [9]. The composition of a set of tasks can favor one tool or

another when measuring overall performance, therefore introducing a bias. Reporting on individual tasks can help potential adopters match tools with their own particular tasks. Finally, data and task selection remains an ad-hoc process which would be aided by the development of task taxonomies and benchmark repositories of datasets and tasks (see 4.1). Even small steps toward reuse will contribute to making our user studies more replicable, and facilitate potential adopters attempts at combining results from multiple studies.

Some characteristics of information visualization render its evaluation particularly challenging. Users often need to look at the same data from different perspectives and over a long time. They also may be able to formulate and answer questions they didn't anticipate having before looking at the visualization. Finally, discoveries can have a huge impact but they occur very rarely, or not at all.

3.2.1 *Looking at the same data from different perspectives, over a long time*

Discovery is seldom an instantaneous event, but requires studying and manipulating the data repetitively from multiple perspectives and possibly using multiple tools. For example biologists might study the same dataset over several weeks or months looking for patterns, interacting with scattergrams and hierarchical clusters of micro-array data, inspecting molecules with custom viewers, consulting gene databases and re-examining the original views from a different perspective. Facilitating the transfer of data between heterogeneous tools and keeping the history of the investigation might well be just as important for discovery as the functionalities of individual tools. Studies typically observe users for a short period of time, and focus on a particular tool or visualization component. They use novice users who are still learning the low-level syntax of the interface and may not be able to focus on the meaning of what they see. Longitudinal studies may be more helpful but they are more difficult to conduct. Using domain experts will lead to more realistic results but individual differences between subjects should be controlled. Measuring the impact of specific visualization components that require users to manipulate visual as well textual representations, use the web to find complementary information, integrate data mining and spend hours brainstorming with colleagues remains a challenge.

3.2.2 *Answering questions you didn't know you had*

Information visualization is sometimes described as a way to answer questions you didn't know you had. This paradox presents another challenge for visualization evaluation. Tasks used during laboratory user studies need to be simple enough to be accomplished in a short –or at least predictable– amount of time, and specific enough for performance to be measured. One way to address in part this challenge is to allow users – before or after the controlled tasks – to explore the data freely on their own and report on what they are able to see and understand. This technique can lead to informative results when subjects are motivated, but is likely to fail with the subjects of typical studies using subject pools. Finding and navigating tasks can be achieved without much difficulty by every subject, but effective exploration and discovery requires active intellectual engagement which is difficult to trigger and control. Letting

subjects use their own data can be helpful but also difficult to arrange.

3.2.3 *Factoring in the chances of discovery and the benefits of awareness*

User studies usually report on the time to complete tasks and error rates. Measures of subjective satisfaction are becoming common, and learnability is sometimes addressed as well. However, potential adopters have to estimate the risks associated with errors (a common problem for any interface) but for information visualization they also need to estimate the chances of discovering a new trend or phenomena in the data. When a big payoff cannot be anticipated, potential adopters will look the ability to streamline repetitive tasks (so that the tool becomes cost effective immediately) in addition to the potential for the large benefits linked to occasional discoveries (that justify widespread use as an investment). Another challenge is that success may not be traceable back to the visualization. For example an effective visualization used on a daily basis by managers may heighten their awareness of business operations by allowing them to absorb and remember large amount of information effortlessly. However it might be difficult or impossible to link their decisions to a particular tool as awareness is difficult to identify and measure, and decision-making uses information from diverse sources. In fact, the introduction of visualization might even trigger changes in work practices, exacerbating the problem of identifying cause and effect.

3.4 Addressing universal usability

The discovery of complex phenomena has been a strong motivation for the use of visualization, but more techniques and tools are now aimed at the general public. Making visualization tools accessible to diverse users regardless of their backgrounds, technical disadvantages, or personal disabilities is necessary when the tools are to be used by the public, but it remains a challenge for designers [14]. Researchers work with high-end computers but they need to address the range of devices and network speeds available in people's homes and businesses. Special algorithms are needed to guarantee rapid downloads and adequate interaction [15]. Visually impaired users may need to use text-based alternatives to the visual display; an example is provided by the National Cancer Institutes cancer maps [2] which provides table views in addition to maps. Encouraging results have been found with the sonification of graphs, scattergrams, and tables [16]. Spatial sound might help sonify more complex data representations [17]. Users with color deficiencies can be provided with alternative palettes or tools to customize the colors. For example, SmartMoney Map of the Market provides two choices of color: red/green and blue/yellow. ColorBrewer [18] and VisCheck [19] offer guidelines on color schemes that work for those with color vision impairment. Help to get novice users started may be needed, and multi-layer interfaces [20] may assist users find the level of complexity which is right for them.

4. POSSIBLE NEXT STEPS

We discuss three possible first steps to improve information visualization evaluation and facilitate adoption: the development of repositories of data and tasks, the gathering of case

studies and success stories, and the strengthening of the role of toolkits.

4.1 Repositories of data and tasks

One way to improve evaluation is to create benchmark datasets and tasks. Sample datasets have been made available (e.g. for graphs and time series) but user testing requires benchmark tasks as well. To promote this approach we organized the InfoVis contest [21]. Our goal was to initiate the development of the benchmarks, establish a forum to promote evaluation methods, and also create a new interesting event at the conference. The first contest took place in 2003 and the 2004 contest is now underway [22]. We invited submissions of case studies of the use of information visualization for the analysis of tree structured data, in particular to look at differences between pairs of similar trees. Three pairs of datasets were provided in a standard format, along with a taxonomy of general tasks (about 40 tasks in 11 categories). For each dataset the application domain of the dataset was described (phylogenies, taxonomies and file systems, with about 60, 200,000 and 70,000 nodes respectively), and open-ended domain specific tasks were provided to guide the analysis. After five months we received eight entries (a small number, but satisfactory for a first year). The main finding was that it was difficult to compare systems even with specific datasets and tasks. We had hoped to focus the attention of submitters on tasks and results (insights), but the majority of the materials we received focused on descriptions of system features. Little information was provided on how users could accomplish the tasks and what the results meant, making it very difficult for the judge to compare. The systems presented were extremely diverse, each using different approaches to visualize the data. Each tool addressed only a subset of the tasks, for a subset of the datasets. The phylogeny, which consisted of a small binary tree was not used, probably because the tasks were complex and required working with biologists (i.e. chosen to be realistic).

There were three first-place entries (see [21, 23] for more information on all entries). TreeJuxtaposer [24] submitted the most convincing description of how the tasks could be conducted and results interpreted. Zoomology [25] demonstrated how a custom design for a single dataset could lead to a useful tool that addressed many of the tasks satisfactorily. InfoZoom [26] was the most surprising entry. This tool was designed for manipulating tables and not trees. However the authors impressed the judges by showing that they could perform most of the tasks, find errors in the data and provide insights in the data. The three second-place entries showed promises but provided less information to the judges on how the tasks were conducted and what the results meant. EVAT [27] demonstrated that powerful analytical tools complementing the visualization could assist users to accomplish their tasks. Taxonote [28] demonstrated that labeling is an important issue that makes textual displays attractive. The submission from Indiana University [29] illustrated the benefits of toolkits by quickly preparing an entry combining several tools, each accomplishing different tasks. All entries were given a chance to revise their materials after the contest. We required participants to fill a structured form with screenshots and explanations for each task. That information is now archived in the Information Visualization Benchmark Repository [23].

With the InfoVis2003 contest we attempted to provide real data and tasks while trying to narrow the problem to one data type (trees) and three representative tree types. The contest taught us that the problem was still too large for a contest and that the vague nature of the tasks made it impossible to compare answers effectively. Our next step - the 2004 contest - will only have one dataset, much fewer tasks and a more structured reporting format. Nevertheless, we anticipate that the open-ended nature of realistic tasks and the diversity of approaches will still make judging a challenge. The contest also illustrated the difficulty of presenting convincing evidence. Demonstrating the power of a tool is difficult. Researchers are trained to describe their tools' novel features more than illustrating them with convincing examples using real data.

Contests are an artificial testing situation where the opinion of judges reflects the quality of the submitted materials, opposed to the actual merits exhibited when tools are tested interactively and discussed with designers. The impact of contests may be limited but the datasets and tasks remain available after the contests. They can be used by developers to exercise their tools and identify missing features, and by evaluators to enrich their testing procedures with complex tasks. We hope that more specific lists of tasks can be added to the repository and used in controlled experiments.

4.2 Case studies and success stories

Case studies report on users in their natural environment doing real tasks. They can describe the entire discovery process, collaborations among users, the frustrations of data cleansing and the excitement of seeing the first overview of the data. They can report on frequency of use and benefits gained. The disadvantage is that results may not be replicable and generalizable. Case studies have the potential to convince users and managers that information visualization is a legitimate investment. They can be extremely convincing for potential adopters working in the same application domain as the one addressed by the study. Others will need to be able to extrapolate the results to their domains and imagine how the tool could be useful for them.

Case studies may use entirely different measures of success than the ones traditionally used in user studies. E-commerce criteria might be the percentage of completed sales, not time or error. In the example of Figure 1 Peets Coffee goals are to sell coffee and tea. What matters is that users complete the order and come again. User satisfaction remains a likely indicator of success, hopefully correlated with order completion. What makes Peets coffee's interface more satisfying remains unverified, even though we may have some hypothesis... Users might feel more in control because they can review all the choices at once? The novelty of the interface might entice expert computer users who are also more able to work thru the interface? Longitudinal studies will be helpful.

The recording of usage data with appropriate privacy protections is a powerful method to glean information about user behavior. It can inform designers about the frequency of feature use, guide screen layouts to speed interaction, pinpoint possible sources of dissatisfaction and inform the revision of help materials. It can also provide evidence of success when users

choose an interface over another one, and allocate resources where they are most needed.

One first step would be for our community to create a directory of success stories (similar to a “return on investment” website [30] and complementary to informational websites [31]). Another approach is to encourage submission of stories from industry and government agencies in information visualization journals or magazine. Success stories set examples for others to follow.

4.3 The role of toolkits and development tools

If adoption is considered a strong criteria for evaluating the value of an information visualization techniques then making that technique a part of a toolkit greatly improves its chance of success. For example ILOG JViews now makes it much easier for designers to use dynamic queries, treemaps, coupled overviews and details, and so on. Many resources are available to designers and developers such as XmdvTool, Xgobi, Common GIS, GeoVISTA studio, the Indiana University InfoVis Repository, Jean-Daniel Fekete’s InfoVis Toolkit and the University of Maryland Piccolo toolkit for zooming user interfaces.

It is likely that information visualization will spread more rapidly into mainstream applications when components such as double-sided sliders, fisheye tables and treemap viewers become part of the mainstream user interface development tools. With wide use, the usability testing of information visualization components or task-based comparisons of components can have a much more significant impact, universal usability aspects might be addressed more easily, and the “early majority” of pragmatists might look at information visualization in a different light.

5. LEARNING FROM EXAMPLES OF TECHNOLOGY TRANSFER

A common evaluation measure for any technology is adoption by others, and the move into commercial products. It is clear that the quality of a product alone does not guarantee commercial success, and then many aspects such as financial alliances, marketing strategies, personal networking and sometime plain luck often play a role as well. However some interesting lessons may be gleaned from examples of transformations from prototypes to products and applications. We look at examples based on three techniques originally developed at the University of Maryland (because we know them well.) The first two examples are of successful products, the third one still has an uncertain future.

Example 1: From the Film Finder to Spotfire

The transformation from the FilmFinder [32] to today’s Spotfire is a 11-year-long voyage [33]. The visualization technique based on dynamic scattergrams (Fig 2) remained basically unchanged but the tool has seen dramatic transformations to increase its utility, which has now been demonstrated in multiple application domains (Fig. 3). Other visualizations such as parallel coordinates, table viewers or standard business graphics

were combined with the scattergrams. Data can be easily imported and results published and shared with others. Analytical tools were integrated to facilitate quantitative analysis.

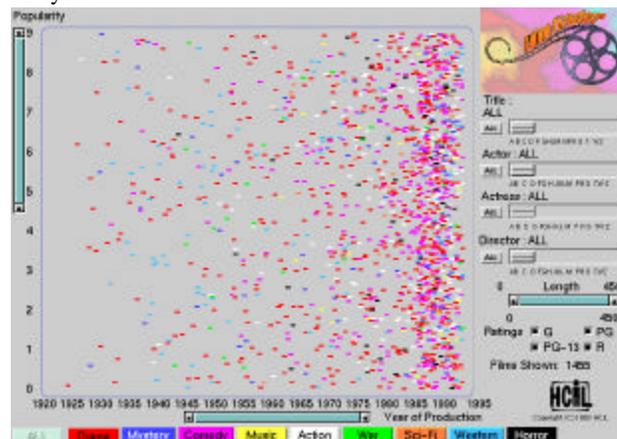


Figure 2. The 1993 version of the Film Finder, with dynamic query sliders to filter a scattergram.

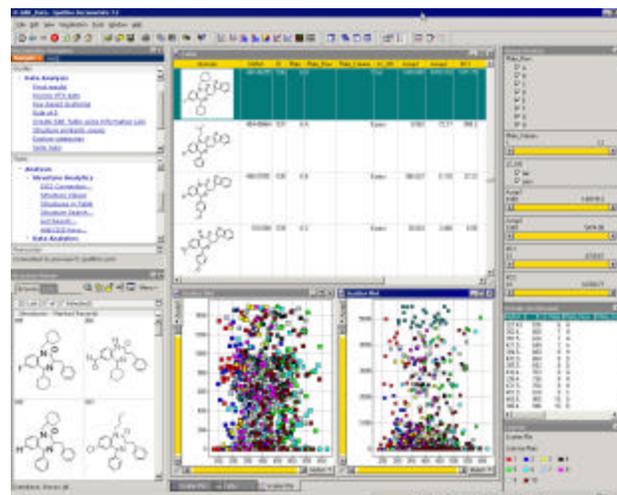


Figure 3. Biochemists explore compound properties with Spotfire (www.spotfire.com)

Separate products were designed to fit the needs of specific application domains such as drug discovery or petroleum industry with a substantial effort going into the development or integration of domain specific tools such as molecule viewers for drug discovery or GIS tools for the petroleum industry. Building a versatile tool while focusing on specific applications seemed important ingredients of success.

Example 2: From treemap to the Map of the Market and other treemap tools

Treemap was first prototyped in 1990, fourteen years ago [34] [35]. Empirical studies demonstrated benefits but acknowledged that treemap required training before users could use it effectively (Fig. 4). The visualization spread slowly, mostly to other research projects until the SmartMoney Map of the Market became the first widely known commercial application of

treemaps (Fig. 5). It provided a powerful success story that early adopters could use to convince managers of the feasibility of treemaps, and a playground for designers who were able to envision how to put treemaps to use in other applications. The success of the Map of the Market illustrates that techniques initially described as challenging to use can become accessible and highly successful interfaces when the right ingredients are present. The popularity of this application with the general public might be attributed in part to the fact that the data is familiar, and important to users who visit a financial website. The careful simplification and refinement of the initial treemap interface also played a large part. The names and groupings of stocks are familiar to users of the SmartMoney website. The hierarchy is simple, being shallow and of fixed depth, and the color mapping is natural to most users (using green for gains or increased values of the stocks, red for losses). An elegant new algorithm had been devised by Martin Wattenberg which resulted in a better aspect ratio for all the rectangles making the display more readable and pleasing, and leading to further refinements [36]. Labeling and links to details were optimized for that specific application. Finally the application was carefully written in lightweight Java so that it could actually run on most users' machines without having to download special plugins or worry about Java versioning. This combination of factors has been effective in making the map usable by the general public, as indicated by the large number of return users the website enjoys, industry awards and suite of imitators.

The success of the Map of the Market is now spawning more interest in treemaps and several companies are doing business with their own versions of treemap. For example the Hive Group recently deployed an application used by the US marines for inventory management. The success of this new generation of products seems to rely on 1) the availability of tools for data preparation and publishing, allowing expert users to prepare automatically updated views for large numbers of end-users. 2) a simplified interface that allows end-users to view the data and perform limited filtering and grouping operations.

Example 3: DataMap

The third example reports on the long voyage from a prototype to an interface developed for disseminating Census data. An interactive choropleth map interface called Dynamap (Fig. 7) was prototyped in 1993 for the National Center for Health Statistics [37]. Many versions were later prototyped by others and us. Unfortunately our prototypes seemed to never be available in the environment potential adopters needed to be able to reach their users. Early SVGA graphic libraries became obsolete too quickly, proprietary GIS software a problem, Java was deemed too inaccessible, then became required, etc. Implementation issues are powerful agents in the process of stalling adoption. In the recent years, a new University of Maryland version was developed, named DataMap by the Census Bureau. Users can click on a map of the US to display facts about states or counties, select multiple areas for comparison, zoom on the map or filter the map with dynamic queries. A scattergram, tightly coupled to the map and table, shows relationships between two attributes. DataMap was identified by the Census Bureau as a candidate interface to release their data on CDs. A Java version may follow.

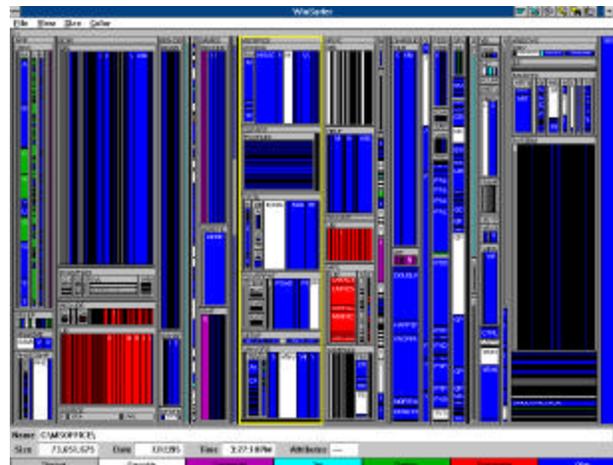


Figure 4. An early version of Treemap showing a file directory with the slice and dice algorithm.



Figure 5. SmartMoney Map of the Market treemap (www.smartmoney.com)



Figure 6. The HiveGroup Honeycomb treemap is being used by thousands of US Marine Corps Logistics Command users for equipment readiness.

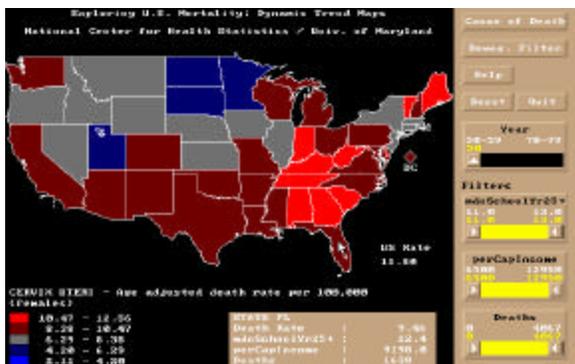


Figure 7. The 1993 Dynamap

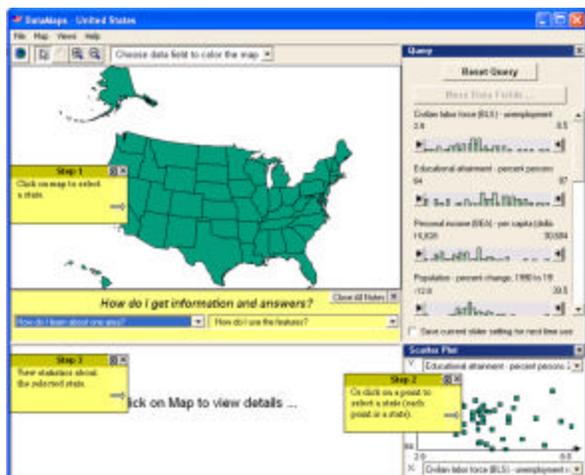


Figure 8. DataMap might be used by the Census Bureau to release data on CDs.

Our numerous demonstrations and tests of DataMap in the lab had made us confident that average users could understand and use it after a minute of two of demonstration. However, usability tests conducted by the Census Bureau with "off-the-streets" novice users receiving no training revealed that many users had difficulties getting started. Some would simply dismiss the interface saying, "It's just too complicated". Others had problems with zooming, selecting multiple regions or using the sliders. Several users were puzzled by the interactive scattergram, some not guessing that dots represented areas and some not knowing how to read a scattergram

The first step was to revise and retest the interface following standard usability practices. This was done by a team of colleagues at Virginia Tech led by Chris North (Fig. 8) [38]. Most problems were handled by simplifying the initial interface (e.g. reducing the starting number of attribute sliders with the possibility of adding more with a control panel) and avoiding situations confusing to users (e.g. by controlling the zoom to avoid zooming in to empty areas of the map). Some aspects of the interfaces still needed explanations (e.g. demonstrating the use of sliders, or introducing the scattergram, so help needed to be provided to get users started. We explored alternative approaches to providing such help [14]. A "sticky note" technique was implemented into the Census Bureau version, and short narrated recordings of demonstrations may also be used. As the CD version is still being polished there is now more interest in a Java version. We have demonstrated the feasibility of dynamic choropleth maps with a variety of platforms and

network speeds [15], we are now exploring the challenge of providing access to the data for blind users [17]. Simultaneously, other dynamic choropleth maps are being developed and it is not clear which tool will become a standard for dissemination of statistical information data to the public. The prototypes have been created and tested, the technology is available, agencies are actively looking for ways to disseminate their data, but these components seem to have difficulties crystallizing into a decisive product. We are still a long way from understanding how our research tools become products. Gathering struggle and success stories from research and product teams should help our community make more rapid progress.

6. CONCLUSIONS

As the field evolves, research prototypes are finding their ways into commercial products. However, information-visualization research must understand the principles that will help the field cross the chasm to wider success. We must complement our love for novelty by carefully integrating visualization tools into solutions for real problems. This may mean facilitating the importation of data, coping with large volumes of incomplete data, enabling users to integrate with other tools and collaborate with others, but it also means reporting on long term use in natural settings. Information visualization is making steady gains. We should promote field studies, investigate new evaluation procedures, and celebrate successes.

7. ACKNOWLEDGMENTS

I particularly thank Jean-Daniel Fekete for the discussions we had while organizing the InfoVis2003 contest, and Ben Shneiderman for his many suggestions. I also thank Mary Czerwinski, George Robertson and Anita Komlodi for their feedback on early versions; and the AVI conference chairs for giving me the opportunity me to write on this new topic. Partial support was provided by NSF grant EIA 0129978, ChevronTexaco, the Census Bureau, the National Center for Health Statistics, and Microsoft Research.

8. REFERENCES

- [1] Borzo, Janet, Get the picture, *Wall Street Journal*, January 12 (2004)
- [2] National Cancer Institute Cancer Mortality Maps and Graphs www3.cancer.gov/atlasplus.
- [3] Sytadin. www.sytadin.tm.fr
- [4] Komlodi, A., Sears, A., Stanziola, E., Information Visualization Evaluation Review, *ISRC Tech. Report, Dept. of Information Systems, UMBC*. UMBC-ISRC-2004-1 http://www.research.umbc.edu/~komlodi/IV_eval (2004)
- [5] Ahlberg, C., Shneiderman, B.. The alphaslider: a compact and rapid selector, *Proc. of the SIGCHI conference on Human factors in computing systems* (1994) 365 - 371
- [6] Irani, P. , Ware, C., Diagramming information structures using 3D perceptual primitives, *ACM Transactions on Computer-Human Interaction*, 10, 1 (2003), 1-19

- [7] Sutcliffe, A. G., Ennis, M., Hu, J., Evaluating the effectiveness of visual user interfaces for information retrieval. *International Journal of Human Computer Studies*, 53, 5 (2000), 741-763.
- [8] Byrd, D. (1999). A scrollbar-based visualization for document navigation. *Proc. of the fourth ACM Conference on Digital Libraries* (1999) 122 - 129
- [9] Plaisant, C., Grosjean, J., and Bederson, B. B. SpaceTree: Supporting exploration in large node-link tree: design evolution and empirical evaluation *IEEE Symposium on Information Visualization* (2002), 57-64.
- [10] Trafton, J., Tsui, T., Miyamoto, R.; Ballas, J., Raymond, P., Turning pictures into numbers: extracting and generating information from complex visualizations. *International Journal of Human Computer Studies*, 53, 5 (2000), 827-850.
- [11] Moore, G., *Crossing the Chasm*, HarpersCollins Publishers, New York, NY (1991)
- [12] Moore, G., *Inside the Tornado*, HarpersCollins Publishers, New York, NY (1995)
- [13] Chen, C., Czerwinski, M. (Eds.) Introduction to the Special Issue on Empirical evaluation of information visualizations, *International Journal of Human-Computer Studies*, 53, 5, (2000), 631-635.
- [14] Plaisant, C., Information visualization and the challenge of universal access, in *Exploring Geovisualization*, Dykes, J., MacEachren, A. and Kraak, M.-J. (Eds.), Elsevier, Oxford (2004) in press.
- [15] Zhao, H., Shneiderman, B., Image-based highly interactive web mapping for geo-referenced data publishing, *Tech Report CS-TR-4431*. (2003) www.cs.umd.edu/hcil
- [16] Brewster, S, A., Non-speech auditory output. In Jacko, J. and Sears, A. (Eds.), *Human Computer Interaction Handbook*, Lawrence Erlbaum Associates, Mahwah, NJ (2002), 220-239.
- [17] Zhao, H., Plaisant, C., Shneiderman, B., Duraiswani, R., Sonification of geo-referenced data for auditory information seeking: Design principle and pilot study, *University of Maryland Tech. Report HCIL 2004-03* (2003) available at www.cs.umd.edu/hcil
- [18] Brewer, Cynthia, ColorBrewer, www.personal.psu.edu/faculty/c/a/cab38/ColorBrewerBeta.html
- [19] Dougherty, R., Wade, A. VisCheck: www.vischeck.com
- [20] Shneiderman, B., Promoting universal usability with multi-layer interface design, *Proc. of ACM Conf. on Universal Usability* (2003) 1-8.
- [21] Fekete, J-D and Plaisant, C., InfoVis 2003 Contest, www.cs.umd.edu/hcil/iv03contest (2003)
- [22] Fekete, J.-D., Greinstein, G. and Plaisant, C., InfoVis 2004 Contest, www.cs.umd.edu/hcil/iv04contest
- [23] Fekete, J-D and Plaisant, C., Information Visualization Benchmark Repository www.cs.umd.edu/hcil/InfovisRepository
- [24] Munzner, T., Guimbretière, F., Tasiran, S., Zhang, L. and Zhou, Y., TreeJuxtaposer: Scalable tree comparison using Focus+Context with guaranteed visibility. *ACM Transactions on Graphics, SIGGRAPH 03* (2003) 453-462
- [25] Hong, J. Y., D'Andries, J., Richman, M., Westfall, M., Zoomology: Comparing Two Large Hierarchical Trees, in *Poster Compendium of IEEE Information Visualization* (2003) (also available at www.cs.umd.edu/hcil/iv03contest)
- [26] Spence, M., Beilken, C., InfoZoom - Analysing Formula One racing results with an interactive data mining and visualization tool, Ebecken, N. *Data mining II*, (2000), 455-464
- [27] Auber, D., Delest, M., Domenger, J-P., Ferraro, P., Strandh, R., EVAT - Environment for Visualization and Analysis of Trees, in *Poster Compendium of IEEE Information Visualization* (2003) (also available at www.cs.umd.edu/hcil/iv03contest)
- [28] Morse, D. R., Ytow, N., Roberts, D. McL., Sato, A., Comparison of Multiple Taxonomic Hierarchies Using TaxoNote, in *Poster Compendium of IEEE Information Visualization* (2003) (also available at www.cs.umd.edu/hcil/iv03contest)
- [29] Sheth, N., Börner, K., Baumgartner, J., Mane, K., Wernert, E., Treemap, Radial Tree, and 3D Tree Visualizations, in *Poster Compendium of IEEE Information Visualization* (2003) (also available at www.cs.umd.edu/hcil/iv03contest)
- [30] Alexander, Dey, Return on investment resources, <http://deyalexander.com/resources/roi.html>
- [31] Dürsteler, Juan Carlos, www.infovis.net
- [32] Ahlberg, Christopher and Shneiderman, Ben, Visual information seeking: Tight coupling of dynamic query filters with starfield displays, *Proc. CHI '94 Conference: Human Factors in Computing Systems*, ACM, New York (1994), 313-321 and color plates.
- [33] Shneiderman, B. Dynamic queries, starfield displays, and the path to Spotfire, www.cs.umd.edu/hcil/spotfire
- [34] Shneiderman, B., Treemaps for space-constrained visualization of hierarchies, www.cs.umd.edu/hcil/treemaps
- [35] Shneiderman, Ben, Tree visualization with treemaps: a 2-d space-filling approach, *ACM Transactions on Graphics*, 11, 1 (1992) 92-99.
- [36] Bederson, B. B., Shneiderman, B., and Wattenberg, M. Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies, *ACM Transactions on Graphics*, 21, 4 (2002) 833-854.
- [37] Plaisant, C., Facilitating Data Exploration: Dynamic Queries on a Health Statistics Map. *Proc. of the Annual Meeting of the American Statistical Association (Government Statistics Section)*, ASA, Alexandria, VA (1993) 18-23
- [38] Li, Q., North, C., Empirical comparison of dynamic queries sliders and brushing histograms, *Proc. IEEE Information Visualization* (2003), 147-15