

A Task-Oriented View of Information Visualization

Stacie L. Hibino

Bell Labs, Lucent Technologies
263 Shuman Boulevard
Naperville, IL 60566 USA
<http://www.bell-labs.com/~hibino/>
hibino@research.bell-labs.com

ABSTRACT

Much of the research in information visualization has primarily focused on providing new views and frameworks to aid users in exploring or accessing data. Very little work has been done to support users through their *full analysis process*—from transforming their raw data into a set of polished final results. In this pilot study, we conducted a task analysis on five experts' use of an existing information visualization system when analyzing a complex data set. Our preliminary results indicate that users conduct several tasks *outside* of data exploration—tasks such as preparing the data, collecting results, and gathering evidence for a presentation. In addition, they give these other tasks high importance ratings with respect to the analysis process.

Keywords

Data analysis process, information visualization, task analysis, exploratory data visualizer.

INTRODUCTION

Information visualization (infoVis) research has primarily focused on enhancing data access and exploration either through general approaches [1, 9] or for particular data domains such as software or temporal data [7, 6]. Little, if any work has examined the larger problem of supporting users' data analysis *process*—the processes they use for transforming raw data into a set of key results based on using infoVis as their primary analysis tool.

What tasks *do* users perform when analyzing data using an infoVis environment? Do they conduct other tasks besides data exploration? If so, how important are these other tasks to the data analysis process? Several infoVis taxonomies have been proposed (e.g., [8, 2]), but these have focused on categorizing aspects of infoVis limited to accessing and exploring data—aspects such as data and visualization types and exploration tasks. The goal of this study is to take a *process-oriented* view of infoVis by assessing users' tasks within the context of real-life data analysis sessions.

METHOD

We interviewed, observed and surveyed five infoVis experts using an existing infoVis environment (EDV: the Exploratory Data Visualizer [9]) to analyze a national disease data set on tuberculosis (TB) [5]. The users' goal was to analyze the TB data and create a presentation of key results. They were given two 1-hour sessions, separated in time by at least one week, to accomplish this goal. They used EDV as their primary tool for analyzing the data, and were told that they could use any other tools that they typically use in conjunction with EDV to accomplish the analysis task. Output of the users' screen was captured on video, along with audio of their verbal protocols.

RESULTS AND DISCUSSION

Although users took different approaches to their analyses, their core set of tasks and findings were very similar. However, due to time constraints and the amount of effort required, none of the users created an actual presentation of results. Instead, they either captured key findings on paper or articulated them aloud during their analysis. Information about presentation tasks was gathered through informal post-interviews and the post-questionnaire.

Types of Data Analysis Tasks

Through observational data, previous infoVis taxonomies [8] and informal interviews, we identified 44 separate data analysis tasks conducted by the users. Grouping similar tasks together revealed seven categories of high-level tasks: prepare, plan, explore, present, overlay, re-orient, and other (see Table 1). When asked, users did not present an alternative list or additional categories of high-level tasks.

Task Importance

Users rated the importance of each of the low-level tasks to the analysis process based on a scale of 1 to 5 (1=unimportant, 5=very important). Column 1 of Table 1 shows average importance ratings for each high-level task, based on an average of corresponding low-level task ratings. While the ratings are fairly high, taking individual users into account, an ANOVA indicates that differences in importance ratings between categories is significant, leading to a rank order of importance: plan > explore > prepare > present > statistics > overlay > re-orient tasks.

Table 1. List of high-level tasks and sample low level tasks. Average user ratings (\pm standard deviation) of high-level task importance are indicated in column 1.

Ave. Rating	High-Level Task Category and Sample Low-Level Tasks
4.3 \pm .9	Prepare: data background and preparation tasks <ul style="list-style-type: none"> • reformat data for suitable input, • check data for potential data errors, • transform the data (e.g., split variables, extract subset, rollup/aggregate data).
4.5 \pm .8	Plan: analysis planning and strategizing tasks <ul style="list-style-type: none"> • hypothesize, • make a strategy or plan for all or a part of your analysis (e.g., decide what, how, and how much to investigate or explore).
4.3 \pm .9	Explore: data exploration tasks <ul style="list-style-type: none"> • get an overview of the data, • "query" or filter the database, • identify curiosities to investigate further.
4.1 \pm .8	Present: presentation-related tasks <ul style="list-style-type: none"> • record or keep track of trends and results tested and found, • articulate importance of a result (rank it or identify it as "interesting").
3.7 \pm 1.1	Overlay: overlay and assessment tasks <ul style="list-style-type: none"> • take notes, • window management (move & resize windows), • assess your observations (e.g., does this observation/conclusion make sense?).
3.4 \pm .9	Re-Orient: re-orientation tasks <ul style="list-style-type: none"> • review goal(s), • review progress, • identify starting point for current session.
3.8 \pm .8	Other Tasks <ul style="list-style-type: none"> • statistics

IMPLICATIONS FOR INFOVIS ENVIRONMENTS

In EDV, as in many infoVis systems, little or no system support is provided for tasks outside of data exploration. Users are thus forced to turn to alternative mechanisms such as writing a Perl script to aggregate or split data, taking notes with pen and paper, or printing out static results to compare and rank them. Users are then also burdened with additional bookkeeping tasks of organizing these independent pieces (scripts, notes, multiple data files, etc.) and keeping track of how the pieces fit into their analysis.

Previous work in information workspaces [3, 5] offers glimpses of an organizing framework, but much of this work focuses on organizing *data and objects* (e.g., documents) rather than a user's *process*. One can, however, imagine using a rooms [5] or book metaphor [3] for organizing an infoVis analysis. For example, rooms or books could be used to logically separate the analysis along themes or threads (e.g., separate rooms could be dedicated to investigations of different hypotheses). The challenge in using these metaphors, however, is in understanding how, if, and when process support can be bridged across rooms.

We are currently performing a detailed analysis of the video data and verbal protocols to identify how often users performed each task, how much time they spent on each task, and if there is a pattern in their transitions between high-level tasks. In the mean time, we note that users estimate that they typically spend, on average, about 25% of their analysis time on data exploration and at most 40% of their time; thereby spending over half of their analysis time on tasks *other than data exploration*. They also stress that this is increasingly true for larger and more complex data sets. Further investigations must be conducted to determine how and when system support can be provided to better aid users through the full data analysis process.

ACKNOWLEDGMENTS

Special thanks to expert users who participated in the study, to Graham Wills for EDV, and to Beki Grinter and Ken Cox for reviewing earlier drafts of this paper.

REFERENCES

- Ahlberg, C., & Shneiderman, B. (1994). Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. *CHI'94 Conference Proceedings*. NY:ACM Press, 313-317.
- Card, S. and J. Mackinlay. (1997). The Structure of the Information Visualization Design Space. *IEEE Proceedings of Information Visualization '97*, 92-99.
- Card, S., Robertson, G., and W. York. (1996). The WebBook and the Web Forager: an information workspace for the World-Wide Web. *CHI '96. Conference Proceedings*, 111-119.
- Disease Data from the 1991 ASA Data Exposition. Available at <http://www.stat.cmu.edu/disease/>.
- Henderson, J. and S. Card. (1986). Rooms: The use of multiple virtual workspaces to reduce space contention in window-based graphical user interfaces. *ACM Transactions on Graphics*, 5(3), 211-241.
- Hibino, S. and Rundensteiner, E. (1996a). MMVIS: Design and Implementation of a Multimedia Visual Information Seeking Environment. *ACM Multimedia'96 Conference Proceedings*. NY:ACM Press, 75-86.
- Jerding, D.F., Stasko, J.T. and Ball, T. (1997). Visualizing interactions in program executions. *1997 International Conference on Software Engineering (ICSE'97) Proceedings*. NY:ACM Press, 360-370.
- Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. *IEEE Proceedings of Visual Languages 1996*, 336-343.
- Wills, G. (1995). Visual Exploration of Large Structured Datasets. *New Techniques and Trends in Statistics*. IOS Press, 237-246.