A Workplace Study of the Adoption of Information Visualization Systems

Victor González
School of Information and Computer Science
University of California, Irvine
vmgyg@ics.uci.edu

Alfred Kobsa
School of Information and Computer Science
University of California, Irvine
kobsa@ics.uci.edu

Abstract: This paper reports an ongoing longitudinal study of the adoption of information visualization systems by administrative data analysts. Participants were initially excited about the anticipated potential of visual data analysis for their work, but gradually discovered difficulties that eventually precluded a true integration of the visualization system into their daily work practices. These difficulties are unrelated to the specific visualization system used. We conclude that data analysts can take much better advantage of the benefits of information visualization systems when these systems are redesigned to be complementary products of current data analysis and workflow systems, rather than being stand-alone products as is currently the case. Our study offers some insights about how this complementarity can be achieved.

Keywords: Workplace studies, Information Visualization.
Category: H.1.2, H.5.2, J.1, K.4.3, K.6

1 Introduction

Most empirical studies that investigate the merits of information visualizations systems (see [Chen and Yu 2000] for an overview) have been designed as lab experiments. Lab approaches allow one to tightly control the experimental conditions and tasks. They are therefore well suited for the evaluation of individual characteristics of visualization software. However, they are also severely limited since the types of tasks being studied and the short duration of such experiments generally do not allow one to draw conclusions about how visualization systems would assist people in their daily work. Workplace studies [Luff et al. 2000] are needed to reveal the ways in which users would integrate information visualization into their current software infrastructures and their work routines for data analysis and reporting. Studies *in situ* also help determine the critical factors that lead to the adoption of information visualization.

A number of commercial information visualization systems have been on the market for several years, but their dissemination is still limited. There is anecdotic evidence that when first being introduced to such systems, people show great interest and feel thrilled about ways in which they could support their work. After a while though, people would stop using them or use them very rarely. Many reasons may explain this fact, but currently we do not well understand it. We decided to conduct a small longitudinal workplace study to clarify the dynamics in the adoption process, and the significance of factors contributing to the gradual loss of interest in such systems.

In this paper, we report initial results of a study with administrative data analysts in a large corporation and an academic institution. We found that one of the central issues for the adoption of such systems is that people see the role of visual information in the data analysis process as complementary rather than as central. We argue that the adoption of information visualization systems by data analysts therefore deeply hinges on their ability to find ways in which these systems can complement their data analysis practices in a meaningful manner.

2 Data Analysts Participating in our Study

Our subjects were five office workers who routinely use large amounts of numerical data on their jobs and were likely to benefit from information visualization systems. Four subjects came from different administrative units of the University of California, Irvine; the fifth subject works for a major U.S. aerospace company. The job description of each subject is briefly outlined below.

Subject A is a statistical analyst in a Human Resources Department. He is in charge of distributing information to a large number of people in his own unit and in external organizations (e.g. unions). He routinely accesses central databases, and feeds report generators that he had created after joining the department a year ago.

Subject B works as a senior assistant director at a Planning and Analytic Studies unit. She performs high-level analytic work and is proficient in software tools like statistics packages, spreadsheets and graphics programs.

Subject C manages grants for more than twenty faculty who frequently request status reports for their funds. She also produces regular reports for the Dean and other units on campus. She uses mainly spreadsheets, but only exploits a small part of their functionality (e.g., she does not program or use macros).

Subject D works as a director at the Research Administration unit. She supervises more than twenty staff members who provide support to all researchers on campus. She exchanges information (spreadsheets, reports, etc.) with other units on campus, and with federal and state research entities.

Subject E works as Senior Finance Analyst and Project Supervisor in the aerospace industry and monitors the financial situation of more than a hundred projects. He is an expert in programming spreadsheet macros and web forms.

While subjects had different ranks, positions and responsibilities, their common denominator was that their work requires the analysis and interpretation of large amounts of numeric data that they themselves or others had generated. All analysts must routinely distribute information and generate reports for others (e.g. colleagues, supervisors and external units). They access, verify, consolidate and format data from shared repositories (e.g. data warehouses), or local databases that they maintain. They commonly use spreadsheets (e.g. Excel), small databases (Access), and presentation software to process, analyze and deliver their data. One subject routinely uses SPSS.

3 Methods

3.1 Visualization System Used

Subjects in this study used InfoZoom Professional 3.62 EN from humanIT [Spenke et al. 1996]. We opted for Infozoom since it represents a good example of how the paradigm of information visualization can be implemented in a software tool, and since we found in previous research that users generally had few problems learning and using the system [Kobsa, 2001; Mark et. al., 2002]. InfoZoom presents data in three different views. Fig.❑ shows a small database in the “overview” mode, in which the value distribution of each database attribute is displayed as a horizontal bar and/or graph.

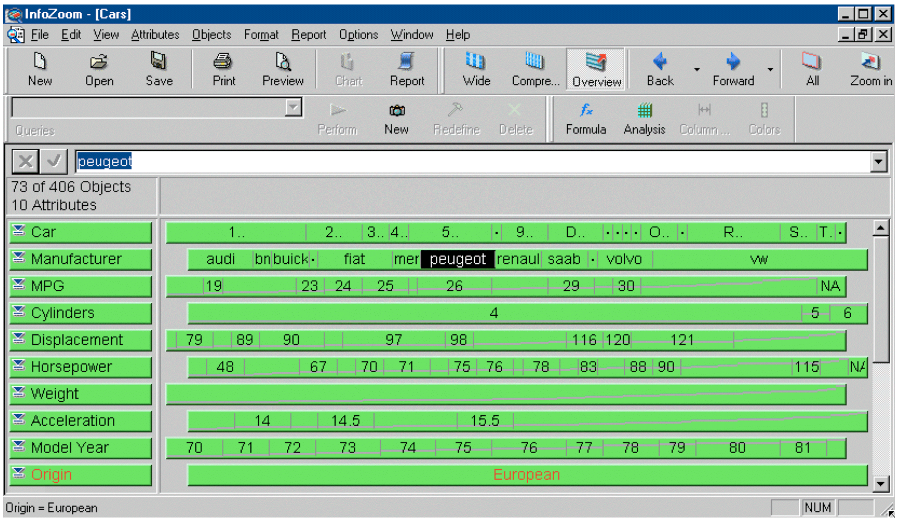

Figure 1: InfoZoom's user interface

InfoZoom's central operation is “zooming” into information subspaces by double-clicking on attribute values, or sets/ranges of values. InfoZoom thereupon shows records only that contain the specific attribute value(s). Slow-motion animation makes it easier to monitor the changes in the other attributes. InfoZoom also allows one to define new variables in dependence of one or two existing variables, to highlight extreme values, and to create a variety of charts (mostly for reporting purposes). InfoZoom can read a number of file formats for tabular data, and also supports ODBC access to database servers. InfoZoom data files can be saved as spreadsheets in which further manipulations can be performed.

3.2 Procedures

We conducted an initial interview with every participant, to familiarize ourselves with their job descriptions and work practices. We asked them about their main responsibilities, their training, the tools they use, processes in which they are involved, and people with whom they interact. We also asked them how they currently perform data analysis, and the software tools they use to support it. The interviews were semi-structured and lasted about 50 minutes on average. Soon after this initial interview, participants received a 90-minute one-on-one tutorial and training on the usage of InfoZoom. Six interviews were scheduled with each participant thereafter, with at least one week in between. Subjects reported their experiences using the system and received help and advice. These interviews lasted between 25 and 45 minutes and were audiotaped and transcribed. In a final questionnaire and interview, we solicited subjects' overall assessment of the usefulness and ease of use of the system.

4 Results

This section presents preliminary findings from our study. We first explain what data analysis means for our subjects, how it is performed as part of their jobs, and which tools and procedures they thereby employ. Then we describe their experiences in integrating the information visualization system into their routines for data analysis.

4.1 What does data analysis mean for our subjects?

People had similar conceptions about what is involved in data analysis. However, they emphasized different aspects of it, depending on what their jobs are. Those subjects who routinely allocate resources (money, time, staff, etc.) conceive data analysis as the process of concurrently verifying the resource allocation until it satisfies a criterion. Other subjects define data analysis as the process of checking trends of data, and comparing them across different periods of time. When the subjects were not so much interested in drawing conclusions based the data but only in furnishing the data to other people for closer analysis, they define data analysis as the process of formatting the data to make it meaningful. In other words, data analysis means for them to achieve a richer presentation of data, so that others (e.g. supervisors) could easily read and understand the data and draw their own conclusions.

We noticed that our subjects always have a purpose in mind when performing data analysis. They either already knew beforehand which trends they wanted to check or which relationships are relevant in their data. They rarely explored data in a completely unrestrained manner, so as to find new and unexpected relationships among data. They also did not define data analysis in that way. However, it was interesting to notice that one of the most appealing characteristics of Infozoom in their view was precisely that it allows this kind of free data discovery.

4.2 Current routines and practices for data analysis

Our subjects analyze data when preparing reports for themselves or for other people. Colleagues and supervisors often prescribe the general outline of the reports to be produced. In most cases, our subjects rely on templates or automated routines to generate reports, to save time. They develop these routines as they learn what is required in their jobs. However, when new requirements are imposed or when their results do not meet their expectations, they perform a deeper analysis of what is required in a report, what variables are involved, and how they should be arranged.

In general, analysts first gather the data that is to be analyzed from the central data warehouse. Then they use different techniques to ascertain their validity: they check for duplicates, empty fields and unusual values. Cleaning can also involve restricting data to those variables that are relevant for the current report. These processes are automated at different levels (depending on the abilities of our subjects), ranging from a manual cleanup to the use of spreadsheet macros or small programs. Once data is clean, subjects format or arrange it in such a way that they can clearly see the trends and relationships they are looking for. The output is either tables or graphs. If trends are not clear or if further filtering is required, they manipulate the data, verify its cleanness again, and rearrange it. After arranging trends, values and results, subjects either integrate the results into documents or distribute their data as spreadsheets or text documents. Most output from data analysis is kept in a tabular format. When required, subjects complement tables with line graphs or pie charts.

During most part of their current data analyses, subjects only sporadically represent data visually. Visual depictions are only used to present the data to others and to enrich the reports. They were not mentioned as a way to clean up or filter data (this is rather performed by Excel formulas, macros and other automated procedures). Only in a few circumstances do subjects use visual representations of data to check and compare trends. However, they do support their observations with hard data, specifically with statistics.

4.3 Integration of an information visualization system into current practices.

Contrary to our initial assumptions, the adoption of Infozoom by data analysts did not mean that they used it predominantly or at least extensively in their current data analysis routines, but rather that they found ways in which it can meaningfully complement these routines. This can be explained in part by the fact that our subjects already have robust software tools at hand to perform data analysis. They use commercial tools (e.g., Excel and SPSS) and created their own (e.g. Excel pivot tables or Access databases). These tools had been extensively tested in the past and are now tightly integrated into the data analysis routines of some of the subjects. Participants quickly understood that InfoZoom’s functionality is not comprehensive enough to replace their current tools. All along the study, they therefore explored how InfoZoom could fit into some stages of their current data analysis process, and ways in which InfoZoom's functionality could be integrated with the tools they already had at hand.

For instance, two of our subjects were keen to use InfoZoom for data cleaning. They liked how easy it was to filter out whole chunks of data, but also mentioned that the time saved on cleaning was lost again when they had to transfer the data to another system for further analysis. Some subjects liked using InfoZoom for checking trends (which was one of the stages of data analysis in which InfoZoom's functionality surpassed any other system available to them). However, they mention that once data is explored they have to make an additional step to support their observations, namely find out the level of statistical significance of their findings. They were dissatisfied by the fact that transferring InfoZoom results to SPSS required some efforts.

Even though Infozoom can import and export data in several file formats, analysts found they needed a higher level of integration at the software level. They want a quick way to move data forward and backward between Infozoom and their current tools. Subjects also repeatedly mentioned a desire to complement their reports with "live" data. They wanted to distribute memos that include InfoZoom files, so that others could confirm the memos, and possibly explore the data in different ways.

5 Conclusion

"Some innovations restructure the way people think and work," suggested [Shneiderman 1994] in reference to dynamic query mechanisms for information visualization systems. In our study, administrative data analysts likewise experience changes in the way they *think* about their data, but not fundamental changes in the way they *work* with their data. Most end up finding a *complementary* use of the information visualization system, in symbiosis with the tools that they already employ (they were well aware though that this symbiosis comes for a price, due to lack of integration).

Complementary use does not rule out frequent use. However, we also noticed that our subjects performed free data discovery with InfoZoom only when their current data analysis practices broke down (namely when reports were requested that they could not generate with existing routines [González and Kobsa 2003]). Even in these situations they already have some expectations about likely trends and relevant variables, and can therefore often search for answers with their current tools. From these observations we conclude that information visualization systems are likely to only be infrequently used by administrative data analysts. This was indeed also what our subjects noticed at the end of their trial periods: they anticipated a sporadic use of the system only.

We believe that information visualization systems have to be redesigned to be complementary to existing data analysis practices, and to be highly integrated with currently used systems. Our analysts were continuously exploring how useful InfoZoom could be for different phases in the data analysis process, and from what we found it is possible that it is useful all along the process. One helpful kind of integration that we envisage would be to enhance export and import functions of data analysis tools including visualization tools in such a way that not only the data but also the current findings are saved and loaded. Users then would not have to reproduce their current discoveries any more when switching to a different system. Tighter forms of integration could be the inclusion of information visualization systems as add-ons to current data analysis and workflow systems, or as resident systems that pop up and visualize data in any application whenever they are required.

Acknowledgements

This research was supported by a grant from the NSF Center for Research on Information Technology and Organizations (CRITO) at UC Irvine. We would like to thank Jinwoo Kim and Warren Liang for their contributions to this project.

References

[Chen and Yu 2000] Chen, C., Yu, Y. (2000): "Empirical Studies of Information Visualization: A Meta-Analysis"; Int. J. Human-Computer Studies 53: 851-866.

[González and Kobsa 2003] González, V., Kobsa, A. (2003): “Benefits of Information Visualization Systems for Administrative Data Analysts”; Proc. 7^{th} International Conference on Information Visualisation, London, England, IEEE Press.

[Kobsa 2001] Kobsa, A. (2001): “An Empirical Comparison of Three Commercial Information Visualization Systems”; IEEE InfoVis’03, San Diego, CA, 123-130.

[Luff et al. 2000] Luff, P., Hindmarsh, ,J., Heath, C., eds. (2000): "Workplace Studies: Recovering Work Practice and Informing Design"; Cambridge University Press.

[Mark et al. 2002] Mark, G., A. Kobsa, V. Gonzales. (2002): “Do Four Eyes See Better than Two? Collaborative versus Individual Discovery in Data Visualization Systems”; Proc. 6^{th} International Conference on Information Visualisation, London, England, IEEE Press, 249-255.

[Spenke et al. 1996] Spenke, M., Beilken, C. and Berlage, T. (1996): *The Interactive Table for Product Comparison and Selection*, UIST 96 Ninth Annual Symposium on User Interface Software and Technology, Seattle, 1996, pp. 41-50.

[Shneiderman 1994] Shneiderman, B. (1994): “Dynamic Queries for Visual Information Seeking”; IEEE Software 11(6), 70-77.