Benefits of Information Visualization Systems
for Administrative Data Analysts

Victor González, Alfred Kobsa
School of Information and Computer Science
University of California, Irvine
{vmgyg,kobsa}@ics.uci.edu

Abstract[1]

This paper reports results from a study on the adoption of an information visualization system by administrative data analysts. Despite the fact that the system was neither fully integrated with their current software tools nor with their existing data analysis practices, analysts identified a number of key benefits that visualization systems provide to their work. These benefits for the most part occurred when analysts went beyond their habitual and well-mastered data analysis routines and engaged in creative discovery processes. We analyze the conditions under which these benefits arose, to inform the design of visualization systems that can better assist the work of administrative data analysts.

1. Introduction

Empirical studies of the benefits of information visualization systems have generally been designed as laboratory experiments (see [1] for an overview). Lab approaches allow one to maintain tight control over the experimental tasks and conditions, and thereby to identify and measure more precisely the effects of individual characteristics of visualization software. However, lab approaches are limited when we want to study how a system assists people in the context of actual professional work, rather than in isolated experimental tasks. Workplace studies and *in situ* observations are a better means to obtain a more comprehensive picture and to determine the critical factors that lead to the adoption of a system [4]. We decided to conduct such a situated study to reveal the benefits that data analysts obtain from using a visualization system in the context of their current work routines, data, and software infrastructure.

Anecdotic evidence from vendors and developers indicates that data analysts envision clear benefits from information visualization systems as soon as they become introduced to them. They quickly understand that these systems can enhance their work by allowing them to view entire data sets, to flexibly manipulate data, and to easily turn findings into charts and reports. Our study confirms this anecdotic evidence but additionally indicates that these benefits are mainly achieved when analysts are driven to abandon their routine data analysis processes and to come to a new understanding of their data.

In this paper, we present a number of cases in which analysts obtained clear benefits by using an information visualization system to support creative discovery processes. These benefits were achieved when subjects visually explored their data to discover new relationships, trends, and outliers. The administrative data analysts discovered that this kind of analysis was not common practice for them, and that their current tools did not support nor promote such practices. They had to find out how to conduct this kind of analysis, both with respect to doing it visually and the required depth of analysis.

In the remainder of this paper, we will describe the participants of this study (Section 2) as well as the research methods (Section 3). In Section 4, we report that benefits mostly occurred when analysts departed from their routine data analysis, and describe three types of benefits that we observed. In Section 5 we substantiate our claims with a number of case studies. Section 6 finally discusses the results and presents some recommendations for the design of information visualization systems aimed at supporting the work of administrative data analysts.

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 therefore likely to benefit from information visualization systems. Four subjects came from different administrative units of the University of California, Irvine; the fifth works for a major U.S. aerospace company. The role of each subject is briefly described 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 since 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.

All subjects distribute information and generate reports for others on their jobs (e.g., for colleagues, supervisors or external organizations). They have to access, integrate and consolidate data from different repositories (e.g. data warehouses or the Internet). To process, analyze and deliver their data, they routinely use tools like spreadsheets, small database programs (e.g. Microsoft Access), word-processors and presentation software. One subject whose job requires high-level analytical work also uses statistical tools on a daily basis (specifically SPSS).

While the responsibilities and roles of our subjects varied, the common denominator was that their work routinely requires the analysis and interpretation of large amounts of numeric data that are created by themselves or by others. Our subjects worked in different organizational units and at different hierarchical levels, which let us
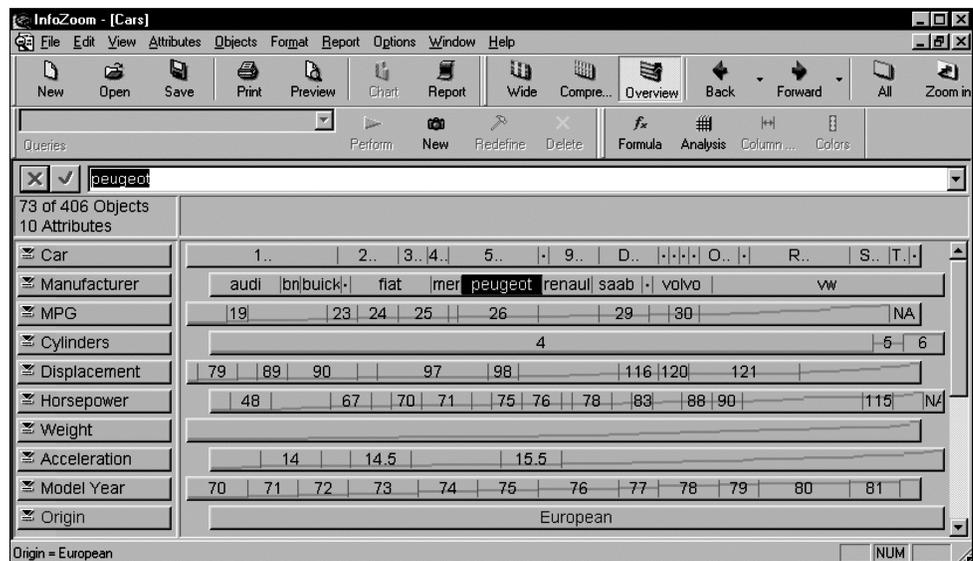

Figure 1: InfoZoom's overview mode

observe how people with different perspectives use an information visualization system. Since our subjects used different software systems to support their data analysis, we also expected to see how the introduction of a visualization system would affect the usage of their current tools.

3. Methods

3.1 Visualization system used

Subjects in this study used humanIT's InfoZoom Professional 3.62 EN (formerly FOCUS [6][7]). 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 prior research that users generally had few problems learning and using this system [3][5]. InfoZoom presents data in three different views.

In the "overview" mode, the value distribution of each database attribute is displayed as a horizontal bar (in the case of categorical data) or as an ascending or descending horizontal graph (in the case of quantitative data). Fig. 1 shows a small car database in this visualization mode. In the "Manufacturer" line, one can see that most car models in the database are being produced by VW. The "Cylinders" line shows that nearly all cars have four cylinders. The ascending line in the "MPG" attribute (miles per gallon) indicates that the mileage is not evenly distributed, but is skewed towards lower values. The same holds true for "Horsepower", "Weight" and "Acceleration". Attribute lines are completely independent in this mode, i.e. one cannot make any comparisons between them.

In InfoZooms "wide view", the whole database is displayed in table format. This view is useful for comparing a few objects only and for viewing individual data values.

The "compressed view", finally, compacts the visualization horizontally to fit the window width. In this view, a database object is a vertical "column" in the visualization. Through (hierarchical) sorting of attribute lines, correlations and other dependencies between attributes can be revealed. An example for a discovery in InfoZoom's compressed view is given in [8].

In all three views, values of (identical adjacent) attributes become textually, numerically or symbolically displayed wherever space permits this. Making sense of a database thereby becomes facilitated [3].

InfoZoom's central operation is "zooming" into information subspaces by double-clicking on attribute values, or sets/ranges of values. InfoZoom thereupon displays records only that contain the specific attribute value(s). Fig. 1 actually does not show the complete cars database, but rather the result of such a zooming operation into those car models whose "Origin" is European (to indicate this, InfoZoom displays the value "European" in red, which is not visible in Fig. 1). Slow-motion animation makes it easier to monitor the changes in the other attributes. InfoZoom also allows one to exclude certain attribute values or disregard attributes, define new variables in dependence of one or two existing ones, highlight extreme values, and 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

Prior to installing InfoZoom at the workplaces of our participants, we conducted initial interviews to familiarize ourselves with their job descriptions and work practices. We asked about their educational backgrounds, training, main responsibilities, work processes in which they are involved, people with whom they interact, etc. We devoted a significant amount of time talking about the ways in which they currently perform data analysis, their routines, the situations that trigger it, and the software tools they use to support it. The interviews were semi-structured and lasted for about 50 minutes on average.

Soon after the initial interview, participants received a 90-minute one-on-one tutorial and training in the usage of InfoZoom. They had the chance to practice using the system through exercises that covered its main functionality.

After the training, we asked the subjects to use InfoZoom with their own data. They could choose any data set that was related to any project they were carrying out at the moment. The goal was to let them understand the system within the context of their own work.

Six interviews were scheduled with each participant thereafter during the duration of the study, with at least one week in between. Subjects could report their experiences analyzing their own data with InfoZoom and trying to integrate the system into their current work practices. During these sessions, they also received additional help and advice. The interviews lasted between 25 and 45 minutes and were audiotaped and transcribed.

A final interview was scheduled at the end of the study, where we discussed in detail the overall experience of using InfoZoom.

4. Results

This section presents some of the preliminary results of our study. We first describe the circumstances under which most of the observed benefits occurred. We then describe these benefits in more detail and provide examples.

4.1. Benefits occurred when routine data analysis broke down

Our subjects perform data analysis when they have to produce reports for themselves or for other people. As data analysts learn what is required on their jobs, they develop routines and use templates to facilitate the creation of such reports and to save time. Report templates or previous reports are reused periodically:

"*We have a table format, and a lot of what we do we've been doing for some time historically. So we have various formats archived, and it's often easier to fall back on what's already written and then just make some minor changes.*!!
Data Analyst B - 01-17-03

In the case of such recurring reports, analysts already know beforehand what the important relationships are. Data analysis in this case only consists of checking that the current data is consistent with these expectations. (If it is inconsistent, analysts will first assume that there are errors in the data rather than changing their expectations).

Whenever data analysts need to create new reports for which previous reports are irrelevant, they have to perform a much deeper analysis. Analysts thereby lack a point of reference for a comparison and have to spend considerable time ascertaining that their understanding of the data is correct. Our study showed that this second type of analysis is neither a common nor a predictable part of the their daily work. Such an analysis may be needed, e.g., when an external entity requests a report for the first time, or when supervisors or colleagues encounter problems and want relevant data to be analyzed. Data

analysts develop routines and templates to accommodate such extraordinary situations, but they cannot go very far making provisions for every possible request.

The need for a deep analysis arises whenever the creation of reports requires from data analyst to go beyond what is considered in current routines. Such a *breakdown* of current processes forces people to abandon well-mastered routines and faces them with the challenge of recreating them. This breakdown can only be overcome by trying to make sense of the data from a new perspective.

In such a situation, software tools are required that allow one to quickly manipulate data and to look at it in different ways. If the tools are adequate, they will help overcome the breakdowns. If this is complicated or tedious, or if the effort required is too high compared with the achievable benefits, it is very likely that these breakdowns remain unsolved. When first introduced to InfoZoom, our subjects mentioned many such unsolved questions and expressed the expectation that the system could help them to finally obtain an answer.

4.2. Identified benefits for data analysis

In general, the administrative data analysts in our study felt that visual tools benefited them in three different ways, which will be described below.

4.2.1 Flexibility in handling data. InfoZoom gave data analysts more flexibility in handling information. Data analysts appreciated being able to quickly include or exclude data for visualization, so they could concentrate on certain portions of data:

"...[InfoZoom] allowed me to focus and compare... by comparing each committee, maybe I'm able to identify, 'well there's a training need for this persons' or 'maybe we've been disproportionately assigning workload to this third committee'. Whereas if I had a table, it wouldn't be nearly as easy to say, 'OK, let's separate out all the C and let's just look at A and B because their work is more similar to each other'. [InfoZoom] allowed me to look at problems from different vantage points more easily and quickly"   *Data Analyst C - 12-10-02*

4.2.2 Full picture of complete data. InfoZoom allows data analysts to obtain a clear picture of their complete data, and to be able to see the wealth of information that lies in this data. Data analysts quickly discovered properties of their data that were not clearly visible or straightforwardly obtainable with other tools.

"InfoZoom helped me to understand my data.... just like that one example of how we could find how many years the average PhD student took to complete the program under each faculty. I didn't know it could come

up with that information, or I don't know if Excel could have. Maybe it could have, but it would have taken more of my time, but it surprised me some of the information that it could come up with."   *Data Analyst D - 02-21-03*

4.2.3 Easy extraction of charts and reports. Finally, InfoZoom offered to analysts easier ways for extracting reports from data than their current tools. They found that querying the data with InfoZoom was easier since the system provides immediate visual feedback. After evaluating the results, operations can also be undone very easily.

"In Excel, it's not easy... you have to select attributes of the people you want to go in there. I don't know, but I use it, it's easier because I can tell [InfoZoom] what I want by clicking instead of trying to figure out where I want it on my graph..."   *Data Analyst D - 02-21-03*

5. Case studies of benefits and lack of benefits

5.1 Quest for complementarity vs. usage in breakdown situations

The potential benefits of the visualization system were detected early in the adoption process, in fact mostly already the first time that subjects had the chance to visually interact with their own data. While using InfoZoom for work-related problems was a requirement for participating in the study, subjects turned out to actually be very keen to try out the system on some of their current problems and data.

Their first approach to using InfoZoom was to see how it could complement their existing tools. Initially, many of our subjects tried to duplicate with InfoZoom what other tools already provided to them. Gradually they realized, however, that their current procedures for data analysis were shaped by the characteristics of their current systems. Using InfoZoom would therefore require major changes in these procedures, which was not worthwhile upfront. While InfoZoom offered some potential advantages compared with their current tools, they realized that the actual challenge was that by replacing some of their tools with InfoZoom, they would discard not only these tools but also comprehensive familiar solutions for data analysis. This situation produced some frustration in two of our subjects, since they liked the advantages of information visualization but could not integrate the system in such a way as to complement their current well-mastered practices. They kept looking for ways to achieve complementarity, but their existing data analysis solutions were too refined and self-contained to make changes seem worthwhile.

Other subjects, after trying for a short time to simulate their existing system, opted instead to use InfoZoom to

solve data analysis breakdowns that occurred in their work. They had questions about relationships among data attributes, wanted to know their distributions, or understand dependencies among them. Solving these questions was not perceived as being critical for their daily work, but rather as having a positive impact on future analyses. Furthermore, they considered this to be a good way to explore the integration of InfoZoom into their work processes.

Using InfoZoom to solve breakdowns let our subjects analyze data in a new way. Soon they started to realize that the kind of data exploration that InfoZoom enabled was definitely not part of their typical routines for data analysis, in two senses. First, given that they had been working as data analysts for quite a long time, deeper analysis is not something they do every day; it emerges only once in a while, when uncommon reports are requested from them. Second, they generally do not analyze data visually due to a lack of appropriate tools. They gradually discovered that InfoZoom was empowering them to be creative, to move beyond their present concerns and questions regarding their data, and to discover new information in the data. Experiencing the value of visual representations as part of creative processes was a strong source of motivation for them during our study. In the following section, we describe three cases how subjects engaged in creative processes while using InfoZoom.

5.2 Examples of benefits when analysts went beyond routine data analysis

5.2.1 Debunking wrong assumptions. A project manager who is well versed in the usage of spreadsheets and has good analytical skills was interested in exploring her data to support in-house training efforts of her department. For some time she had noticed an increased delay in some of their internal processes (namely protocol approval). She decided to use InfoZoom and to visually comprehend the relationships between different variables and how they affected the time to complete the process. She was able to quickly determine that her original intuition regarding the reasons for this delay were wrong, discover the real cause, and take further actions to alleviate the problem:

"Well, I would say the most dramatic result was where a trend clearly and obviously debunked a misapprehension that we had.... We had made an assumption that certain things would hold up protocol approval, and it was clear immediately visually that the number one variable that holds up protocol approval is whether there is any off-site component. It was a real revelation for us, and I don't think we would have come to that drawing down tables."   *Data Analyst C - 12-10-02*

After debunking their initial flawed assumptions, she and her team decided to continue monitoring this chief factor for delays, and to restructure their reports accordingly. In other words, InfoZoom let them to instigate a new data analysis process and a new way of looking at their data that will give high weight to this particular variable.

5.2.2 Answering long-time unsolved questions. Another subject, a financial analyst, had kept for a long time a list of unanswered questions about some properties of the data that she works with on a daily basis. The answers to these questions were not critical for her work, but knowing them would enable her to be more responsive to external demands. People from other units often asked her about these properties, and her reaction was to send them the data so they could find out the answer for themselves. She knows how to use her software tools to produce the results, but was reluctant to spend much time doing it herself. When we started the study, she was motivated because she realized that InfoZoom offered a quick way to answer these questions.

"They were always asking us that kind of questions or "how many women in the program vs. men," "how many undergrads we have vs. grads," "how many are in the PhD, in the Master's Program." So we were asked those kind of questions but it was tedious to get the results."

     Data Analyst D - 02-21-03

This analyst determined the answers to these questions, but also came up with further insights that she shared with people from other administrative units. She used the discovered information to advise other administrators of certain previously unknown relationships in their data. After this exercise, she suggested many ways in which their processes for data analysis and reporting could be modified, so as to include references to the valuable information that she had found.

5.2.3 Checking trends before the extra step. A senior data analyst used InfoZoom as a way to pre-explore data. She is very well versed in statistical packages (specifically SPSS). For a while, she tried in vain to duplicate with InfoZoom what she can easily do with her current tools. Nevertheless she decided that visual exploration was something that she was lacking. While conducting research for a major report to extra-mural organizations, she used visual representations to derive new interesting trends from her data. For those trends that looked promising, she took the extra step and applied more formal statistics to confirm its significance. This is another example of a creative use of InfoZoom, where the goal was to look at the data and to explore them. Once candidates for relationships and trends were identified, they could be incorporated into existing work practices.

6. Discussion

The fact that some of our subjects were able to use InfoZoom to support creative problem solving that was not part of their daily routine teaches us important lessons about data analysis. Creative discovery processes are almost never structured, and require lots of interaction with the data before one can make sense of them and discover interesting facts. This explains why InfoZoom became such a valuable tool. InfoZoom is a relatively transparent and intuitively comprehensible system that allows for very flexible interaction [5]. Our users were therefore relatively easily able to identify relationships between variables that they found important.

Once they found such relationships, they aimed at establishing a routine that would let them revisit these relations regularly in the future. They did not want to repeat the discovery process though, but to monitor the discovered trend or relationship over time using their familiar tools and processes:

"...What I would really love to do is to include all the 2002 data because the reports I was running was 2000 and 2001.... it would be helpful, like I said, not to ask new questions, but to go back and ask some of the same questions but including the 2002 data"

Data Analyst C - 12-10-02

Subjects mentioned that once they had discovered a noteworthy property in the data, they would try to use tools other than InfoZoom to either obtain additional information, or create templates or procedures that would facilitate (and possibly automate) the generation of reports about this property. This indicates that InfoZoom, and probably also other visualization tools, support a general discovery phase that *precedes* the routine data analysis process, rather than support the routine analysis processes themselves. For our data analysts, the benefits of visually exploring the data can only be understood in the context of an activity whose final goal is a new data analysis routine supported by templates and automated procedures. The real value of InfoZoom lies in using it in creative ways to discover something, rather than just to confirm something. This implies that tools like InfoZoom have to provide the means not just to visually reveal properties of data, but also to transfer the data and the discovered properties to other applications, so that the findings can be further analyzed and compiled into queries for future reports.

We observed that subjects who wanted to use InfoZoom together with their current tools experienced inconveniences [2]. They wanted to quickly move data forward and backward between InfoZoom and their tools, or even better be able to call InfoZoom functionality within their own tools, but were unable to do either. The system also did not provide any support to export their findings, i.e. the discovered relationships, trends and outliers as well as the steps taken to isolate them. Our subjects would like to have an easy way to transfer these findings to their current applications, where subsequent processing and automation can be performed. Further analysis may provide additional data to define the characteristics of a visualization system that can achieve such an integration.

7. References

[1] C. Chen and Y. Yu, "Empirical Studies of Information Visualization: A Meta-Analysis", *International Journal of Human-Computer Studies* 53, Academic Press, San Diego, CA, 2000, 851-866.

[2] V. González and A. Kobsa, "User Adoption of Information Visualization Systems", *I-KNOW'03 Workshop on Knowledge and Information Visualisation (KIV2003)*, Graz, Austria, Springer-Verlag, 2003, to appear.

[3] A.!Kobsa, "An Empirical Comparison of Three Commercial Information Visualization Systems", *IEEE Symposium on Information Visualization*, San Diego, CA, IEEE Press, Los Alamitos, CA, 2001, 123-130.

[4] Luff, P., J. Hidmarsh, and C. Heath, eds. *Workplace Studies: Recovering Work Practice and Informing Design.* Cambridge University Press, Cambridge, England, 2000.

[5] G. Mark, A. Kobsa, and V. González, "Do Four Eyes See Better than Two? Collaborative versus Individual Discovery in Data Visualization Systems", *Proceedings of the Sixth International Conference on Information Visualisation.* London, England, IEEE Press, Los Alamitos, CA, 2003, to appear.

[6] M. Spenke, C. Beilken, and T. Berlage, "The Interactive Table for Product Comparison and Selection", *UIST-96 Ninth Annual Symposium on User Interface Software and Technology*, Seattle, WA, ACM Press, New York, NY, 1996, 41-50. http://fit.gmd.de/~cici/Focus/Paper/uist96.htm

[7] Spenke, M. and C. Beilken, "Discovery Challenge: Visual, Interactive Data Mining with InfoZoom – the Financial Data Set", *Workshop Notes on "Discovery Challenge", 3rd European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'99)*, 1999, 33-38. http://fit.gmd.de/~cici/InfoZoom/DiscoveryChallenge/Financial.ps

[8] Mark, G., K. Carpenter, and A. Kobsa (2003): "A Model of Synchronous Collaborative Information Visualization", *Proceedings of the Sixth International Conference on Information Visualisation.* London, England, IEEE Press, Los Alamitos, CA, 2003, to appear.