

Information Visualization Evaluation in Large Companies: Challenges, Experiences and Recommendations

Michael Sedlmair[†] Petra Isenberg* Dominikus Baur[‡] Andreas Butz[‡]

[†]University of British Columbia

*INRIA

[‡]Ludwig-Maximilians University Munich

msedl@cs.ubc.ca

petra.isenberg@inria.fr

(dominikus.baur | andreas.butz)@ifi.lmu.de

Abstract

We examine the implications of evaluating data analysis processes and information visualization tools in a large company setting. While several researchers have addressed the difficulties of evaluating information visualizations with regards to changing data, tasks, and visual encodings, considerably less work has been published on the difficulties of evaluation within specific work contexts. We specifically focus on the challenges which arise in the context of large companies with several thousand employees. Based on our own experience from a 3.5-year collaboration within a large automotive company, we first present a collection of nine information visualization evaluation challenges. We then discuss these challenges by means of two concrete visualization case studies from our own work. We finally derive a set of sixteen recommendations for planning and conducting evaluations in large company settings. The set of challenges and recommendations and the discussion of our experience are meant to provide practical guidance to other researchers and practitioners who plan to study information visualization in large company settings.

1 Introduction

Evaluation in the context of specific data and task sets is a fundamental part of information visualization research.^{1,2} Systems and techniques developed by researchers are often intended to support everyday work activities for domain-specific tasks and data and, thus, evaluation in the context of these activities is particularly important. In order to more clearly understand and assess “real world” data analysis problems and the use of our tools within a specific work context, a close collaboration with domain experts is instrumental.³⁻⁶ Researchers have to learn about their data analysis habits, requirements, goals, and tool use within the respective work context, or “field”,⁷ where existing types of research strategies can roughly be categorized as field studies and field experiments.⁸ Field studies are described by McGrath⁸ as direct observations with minimal possible intrusion within “natural” work environments, whereas field experiments are

a compromise strategy where features of the system are manipulated in order to gain more precise results.

In this paper, we focus on the challenges of applying field strategies within one specific type of field—that of large industrial companies of several thousand employees. We derive our findings from our 3.5-year experience designing and evaluating information visualization and visual analytics tools within a large automotive company (BMW Group). From our own experience we know that applying and evaluating visualization systems directly within a large company context is a fruitful endeavor and can produce valuable insights for the field of information visualization in general. Within such an environment a wide range of real data analysis problems, tasks, and datasets are available. Large companies are also often highly interested in applied research and will even fund it.

Evaluation of information visualization solutions within this context, however, has its own unique set of requirements and challenges due to the structural differences in contrast to small companies, such as a higher degree of organizational complexity, more specialization, formalization and decentralization.⁹ In this article, we categorize field-specific challenges that may arise in evaluating data analysis habits and information visualization systems in such an industrial environment based on our experience in building and evaluating these for automotive engineers. We specifically focus on the evaluation of information visualization “tools” which are meant to solve specific analysis problems for company employees as part of their everyday work. This means that our analysis may not necessarily apply to other types of none work-related information visualization scenarios such as casual information visualization or social data analysis. In this article, we use the term “evaluation” to refer to the validation of data analysis processes and/or information visualization tools using methods with user involvement, including both field and lab studies. We do not address validation methods without user involvement such as formal theory⁸, measuring system time/memory or qualitative discussion of resulting images¹⁰ which are also used in the area of information visualization. We also do not address evaluations that try to assess specific interaction techniques for information visualization (e. g., usability evaluation of a specific visualization techniques) but focus instead on evaluations for information

visualization design that are more holistic in nature. Furthermore, we use the term “evaluation” to refer to both studies of already developed visualization tools (tool-centric) as well as attempts to evaluate current work practices informing the design process of a visualization tool (work-centric). We included work-centric evaluation for two reasons. First, retrospectively we learned that studying current practices of end-users was very important for the success of our visualization tools. Second, we followed the call of information visualization researchers who underlined the importance of such studies for visualization tool design.^{10–12}

We discuss a set of nine challenges we encountered, our own experiences with these challenges in using different evaluation methods, and present a set of sixteen recommendations for evaluation within a large company context. Challenges and recommendations include both aspects specific for information visualization evaluation but also more generic considerations which are no less important for our research. We hope that this article will help to serve as a reference for others who are planning information visualization evaluations within a large company context and provide them with practical guidance. While our article focuses on evaluation within a large company setting, we believe that our discussion will also be informative for visualization researchers collaborating with all kinds of industrial partners—including small or medium-sized companies—and that some of our challenges, experiences and recommendations might hold true for such settings as well. However, as the work presented in this article is solely based on our collaboration with a large company, we deliberately refrain from making assumptions on generalizability to other contexts.

Finally, this article is based on the considerable extension of a previous workshop paper on the same topic.¹³

2 Related Work

In this section, we discuss previous field strategies that were conducted with information visualization tools and go into more detail on obstacles of field research as discussed in the general HCI literature.

2.1 Information Visualization Evaluation in the Field

Despite the well known drawbacks of artificial scenarios and hypothetical tasks, most evaluations for information visualization tools are still conducted in lab settings.³ Some researchers, however, have called for more real world applications of research^{1,10–12} and a growing number of researchers are beginning to invite their target audience to participate in user studies: Perer et al.,¹⁴ for example, studied their social network tool with several experts from different fields of data analysis. Ethnographic studies have also been used within a user-centered design process with domain experts and have

been shown valuable as a formative part of the design process: Tory et al.,¹⁵ for example, documented the results of a qualitative analysis in the building design field and concluded that their structured analysis of qualitative study data provided deep insight on the work processes with visualization. Long-term studies¹⁶ are another type of field strategy that offers the chance for deep insight and learning of the workings of a field and possible merits of visualization use. Unfortunately, they are laborious and only few have been reported on in the literature.^{17–19} The work by González and Kobsa, for example, describes the adoption of an information visualization tool by data analysts over a longer period of time.¹⁷ In a follow-up paper, they describe further observations on the merits of such tools in the workplace.²⁰ While these examples are promising steps towards more evaluation in close contact to domain experts, more insight is needed on the challenges of conducting information visualization evaluation within specific work contexts. Our article is a step in this direction and lists a first set of challenges, experiences, and recommendations for deploying and evaluating information visualization within a large industrial company.

2.2 Organizational obstacles known from HCI

In the area of HCI, more precisely in Participatory and Contextual User-Centered Design (UCD), a considerable amount of previous work exists on how to meet usability evaluation and user needs by actively involving all stakeholders (e.g., end-users, management, decision makers). Much of this research has been conducted in industry settings.^{21,22} Grudin²³ explicitly discusses obstacles encountered in large companies such as finding “representative” participants and crossing organizational barriers during a UCD process. Poltrock and Grudin²⁴ conducted two observational studies in large companies and reported how several organizational factors (e.g., missing commitment, unsatisfying training) can block UCD. Jeffries et al.²⁵ provided a comparison of four formative usability studies in real world environments and recommended heuristic evaluation and usability testing methods for evaluation when considering the number of found problems, UI expertise, and costs.

The main difference of our work to most of these approaches is that we do not examine business-to-customer situations: While much of the previous work was concerned with employing UCD to develop tools for expert users on the outside, we are interested in designing information visualization tools for use within a large company to improve the work processes of its employees. While novel requirements and challenges applying UCD for in-house tools, such as platform and application buying concerns, change management, or the IT life cycle, have been previously discussed^{26,27} related work in this specific area is still rare. In particular, the challenges of information visualization evaluation—as opposed to general usability evaluation—have not received much attention in this context. We contribute a first collection of challenges and recommendations for applying different evaluation methods within a large company context and hope that

this collection will be expanded and modified as more evaluations of information visualizations will be conducted in this work context.

3 Problems and Challenges

While designing and evaluating information visualizations within a large company, we have experienced several field characteristics that pose particular challenges to evaluation. These challenges arise due to the large company setting where workflow, bureaucracy, or hierarchical structures may be quite differently defined compared to smaller companies.^{9,28} For instance, large industrial companies are often characterized by a high degree of collaboration and specification. A single employee often is highly specialized, responsible for a small subset of a highly specific collaborative task set and usually not able to understand all facets of the entire domain.⁹ Therefore, the know-how in a company is widely distributed often resulting in high coordination costs (telephone calls, meetings, etc.).²⁹ In a small company or research lab, on the other hand, a problem domain may be very specific and employees may be able to maintain a comprehensive understanding of their work context and may even be able to deal with many tasks personally without coordination overhead.

When attempting to evaluate information visualization within a large company context, it is imperative to understand the characteristics of this specific evaluation field in order to be prepared for the challenges that may arise in planning and conducting a study and finally analyzing and disseminating the results. In the following section, we describe nine specific challenges to information visualization evaluation in large companies. We categorize them along the typical flow of a user study: study/application design, participant gathering, data collection, and result generation.³⁰ We ground our collection of challenges in our own experience with different techniques, including qualitative and quantitative as well as tool-centric and work-centric (see Section 1), and on general lessons learned from both the HCI and sociology literature and do not focus on a specific evaluation methodology, data collection or analysis methods. We also include challenges related to tool deployment within the company setting which has been both extremely valuable, as a prerequisite to longer-term studies, as well as challenging for us. We describe challenges of deployment that are information visualization-specific but also other challenges in order to give a more complete picture of our experiences. Table 1 gives an overview of the challenges and summarizes in which evaluation phase a particular challenge is most important to consider.

3.1 Study/Application Design

C1-INTEGRATE: Integrating Tools in Daily Work Processes

Integrating information visualization tools in daily work

practice is a labor-intensive process, not only in large companies. Tools have to be stable, robust to changing datasets and tasks, and—if they replace previous tools—should support the functionalities of the tools being replaced. Besides these common challenges, we describe two critical aspects to consider in large company settings:

(a) *Technical Issues:* Task specialization is common in large industrial companies. Therefore, many specific data analysis tasks exist and most of these will likely already be supported with a variety of different analysis tools. These tools are often well integrated to perform within a chain of other tools so that they together provide more encompassing analysis solutions. Under these circumstances the integration of a new visualization tool may be quite challenging as it may break the chain of analysis processes that are already supported by existing solutions. However, the integration of a specific tool may be a valuable exercise in practice, in particular for studying the use of a tool within a specific established work context.⁵

(b) *Political and Organizational Issues:* Many large companies require the authorization of software or software components upfront for the aforementioned reasons of functionality and security. Initially this may not seem complicated, however, depending on the amount of bureaucracy involved, this process may require highly collaborative synchronization efforts and may become long and exhausting.

C2-DATA: Getting the Data

Not only the tools and techniques but also the domain-specific data itself will likely be distributed across different work groups within large companies. A novel visualization approach, however, might have been designed to improve work with combined and aggregated sources of data. To evaluate a tool with these data sources, issues with interoperability between different data sources on different machines and within different work groups might have to be taken into account. Unavailability, different data versions, different or inappropriate formats, unmaintained sources, and most importantly security restrictions can be additional challenges. However, being able to evaluate visualizations with the data used and created by the participants in their everyday work practices can be critical—not only in evaluating how a tool is used with real-world data characteristics, but also in order to convince the participants or stakeholders that this tool may actually improve everyday work.

3.2 Participants

C3-CHOOSE: Choosing Domain Expert Participants

Employees in large companies have different goals, views, and work habits which represent a variety of personalities and opinions.²⁹ This is particularly important to keep in mind when planning to conduct qualitative work such as interviews, observations, and focus groups with or without information visualization tools. While there might be many teams with (superficially) similar data analysis tasks and data types across a company, the practices and tools of these

Short name	Description	pre	during	post
<i>Study/Application Design</i>				
C1-INTEGRATE	Integrating Tools in Daily Work Processes		x	x
C2-DATA	Getting the Data		x	x
<i>Participants</i>				
C3-CHOOSE	Choosing Domain Expert Participants	x	x	x
C4-TIME	Getting Time From Domain Experts	x	x	x
C5-CONVENTIONS	Attachment to Conventional Techniques			x
<i>Data Collection</i>				
C6-CONFIDENTIALITY	Confidentiality of Information	x	x	x
C7-COMPLEX	Complex Work Processes	x		
<i>Results</i>				
C8-STAKEHOLDERS	Convincing the Stakeholders	x	x	x
C9-PUBLISHING	Publishing	x	x	x

Table 1: Challenges lookup table: The last three columns indicate the phase (pre-design, during-design, post-design) in which a specific challenge may strongly influence information visualization researchers' work in a large company setting (marked with an "x").

groups can vastly differ and therefore studying these groups can result in very different findings. When observing different employee groups, it should, for instance, not surprise that some may have built their own work processes or tools around their work tasks or datasets and that other groups and employees, who may have similar data, may have come up with different solutions while being unaware of solutions from other groups. Besides, experts in large companies often have varying tasks and not all of them may be relevant for an observer and neither do the domain experts want to be observed in every situation. Finding the appropriate balance between unobtrusive observation and intervention when observing work processes requires skill and tact on the side of the evaluator. On the other side, talking to participants in pre-scheduled appointments is also often not sufficient: What people tell you is not always the same as what people do.³¹

C4-TIME: Getting Time From Domain Experts

It is very common that employees in large industrial companies work under heavy time pressure and are bound to strict deadlines. Having to revise a deadline often leads to a considerable loss of revenue. These pressures result in specific challenges for evaluation in general and for evaluation with significant participant involvement in particular: (a) Getting domain experts for studies is generally difficult as time equals money. Every hour participants spend on working with visualization researchers is an extraordinary task without necessarily direct evidence of impact on their actual work tasks; (b) under these circumstances, it becomes difficult to argue for long-term studies (e. g., MILCs¹⁶) without any kind of "pre-evidence" that the required involvement will result in qualitative or quantitative improvements to future work processes.

C5-CONVENTIONS: Attachment to Conventional Techniques

Even if a new system may be designed to improve on conventional tools or techniques, experts are very accustomed to and effective with previous solutions. This effectiveness naturally leads to attachment to the traditional tool and results in

a reluctance to learn a new system. One concrete downside is that by working with their traditional tools over a long period of time experts will have developed skills to estimate the effort and time required for a specific analysis and can factor this knowledge in when planning upcoming deadlines. It is difficult for them to estimate this with a novel tool. In addition, certain domain experts have mastered complex tools and data analysis tasks over the years. If a newly-designed tool significantly simplifies a specific data analysis compared to a previous tool, these experts are stripped from their respected expert status and others are also able to conduct the same tasks.⁷ These issues can complicate both acquiring participants for studies (see also C4-TIME) and conducting and evaluating comparative studies.

3.3 Data Collection

C6-CONFIDENTIALITY: Confidentiality of Information Video-, audio- and screen-recording can be useful data collection tools for information visualization evaluation. Especially for qualitative evaluation such data collection helps to capture participants' actions, conversations and responses and allows systematic coding and analysis of the data in retrospect.³² However, large companies often have confidentiality guidelines and restriction policies (Intellectual Property Rights security requirements, short IPR) that might forbid certain recording techniques. In addition, being discreet about collected data is important. Internal work processes are often secret. This, in particular, means that data cannot be shared (e. g., with a second coder or in online tools), results are only allowed to be discussed anonymously, and that publication restrictions are in place—not only about the results of the study, but also when talking about the data analysis characteristics of the tool (see C9-PUBLISHING).

C7-COMPLEX: Complex Work Processes

One important goal in information visualization is to support people in solving complex tasks. For this purpose, an important first step is to understand current data analysis problems

with pre-design evaluation.¹¹ For us, this type of evaluation has been a very important step in order to focus our work on solving the right real-world problems. Pre-design studies, however, become additionally challenging in large companies where complex problems are often split among several, highly specific sub-problems (see also C3-CHOOSE). Understanding the specifics of both the overarching problem solving process (macro challenges) and the individual (micro) challenges may be difficult for an outsider.

3.4 Results

C8-STAKEHOLDERS: Convincing the Stakeholders

An important evaluation goal in information visualization is to understand how people use visualizations to solve real world problems. This goal does not necessarily align well with the goals of stakeholders whose task it is to maximize profit for the company. Therefore, they are more interested in tools that help to save money and improve the effectiveness and efficiency of their employees (again, time equals money). Another goal of the company is speeding up current work practices (e.g., more insights/time³³) while researchers may be more interested in factors that influence or improve qualitative aspects of the work or the specific factors that may have led to improvements (e.g., how insights were achieved¹⁹).

C9-PUBLISHING: Publishing

To allow information visualization to grow as a field and to share and discuss results with the larger research community they have to be made public. Due to competitive reasons, large companies however often have restrictions on what can be published, in particular if this work leads to a competitive advantage. Researchers may expect a lot of bureaucratic hurdles.

4 Experiences and Methods

In this section, we illustrate some of the listed challenges with examples from our own experience in building visualizations at BMW Group, a large automotive company of roughly 100,000 employees. Over a period of 3.5 years we worked together with engineers who developed and analyzed in-car communication networks. These in-car communication networks connect the electronic components of a car; including up to 100 control units per vehicle, sending and receiving messages; and up to 5 different communication technologies, transporting messages. Over the last years these networks have become highly complex—exchanging roughly 15,000 messages per second with safety critical real-time requirements—and resulted in various and large datasets. Working with and understanding these data during network design and analysis has therefore become a major challenge to automotive engineers.^{34–37} As current techniques are basically text-based, our goal was to find visualization solutions to provide new insights and support engineers in their daily work. We followed this goal and collaborated

with roughly 150 engineers from different subdomains in developing and analyzing in-car communication networks. We designed, implemented and evaluated nine information visualization systems for them resulting in five integrated and adopted systems which are still used by engineers in their daily work practices; Figure 1 gives a temporal overview over all nine projects including the evaluation methods we used; additional details on some of our projects can be found in other publications.^{38–42}

Before, during and after developing these tools, we applied a variety of evaluation techniques in order to better understand problems, practices and requirements (pre-design), to learn how to improve our tool designs (during-design), and to validate the domain value of our own, novel approaches (post-design). In the following, we present two case studies from our own work in order to illustrate the challenges we encountered as well as our approaches to overcome them. After that, we briefly summarize all evaluation methods we used and reflect on our experience applying them.

4.1 Case Study 1: AutobahnVis

AutobahnVis is a visualization tool to support overview, navigation, and pattern recognition for error detection in large network communication logs based on temporal ordered lists of messages (see Figure 2). Technical details on early versions of AutobahnVis can be found in previous publications.^{39,40} Here we discuss our experiences with several types of evaluations within the context of this project.

4.1.1 Pre-Design Studies

In order to assess situations in which visualization could support the domain expert's work practice and to understand their data analysis problems we first conducted a pre-design study with eight domain experts.³¹ Originally we had planned to use a fly-on-the-wall technique⁴³ for participant observation without any interference, however after a pilot study it became clear that we would not understand their work processes solely through observation. We changed the pure observational design to a variant of a contextual inquiry,⁴⁴ where one analysis expert explained his or her work to 2–3 observers by means of a current work problem. One participant, for instance, explained his work on the raw data to us and made it clear that visualization would not be useful to him as its abstraction was a potential source of error. This participant had learned over the years to read the raw hexadecimal data and had the status of an analysis expert (C7-COMPLEX).

IPR security requirements were in place in the manufacturing areas in which we conducted our study. In order to allow participants to speak openly, we decided not to use any electronic recording devices (C6-CONFIDENTIALITY). Our approach was to counterbalance the potential loss of information by using 2–3 observers instead of just one, all taking

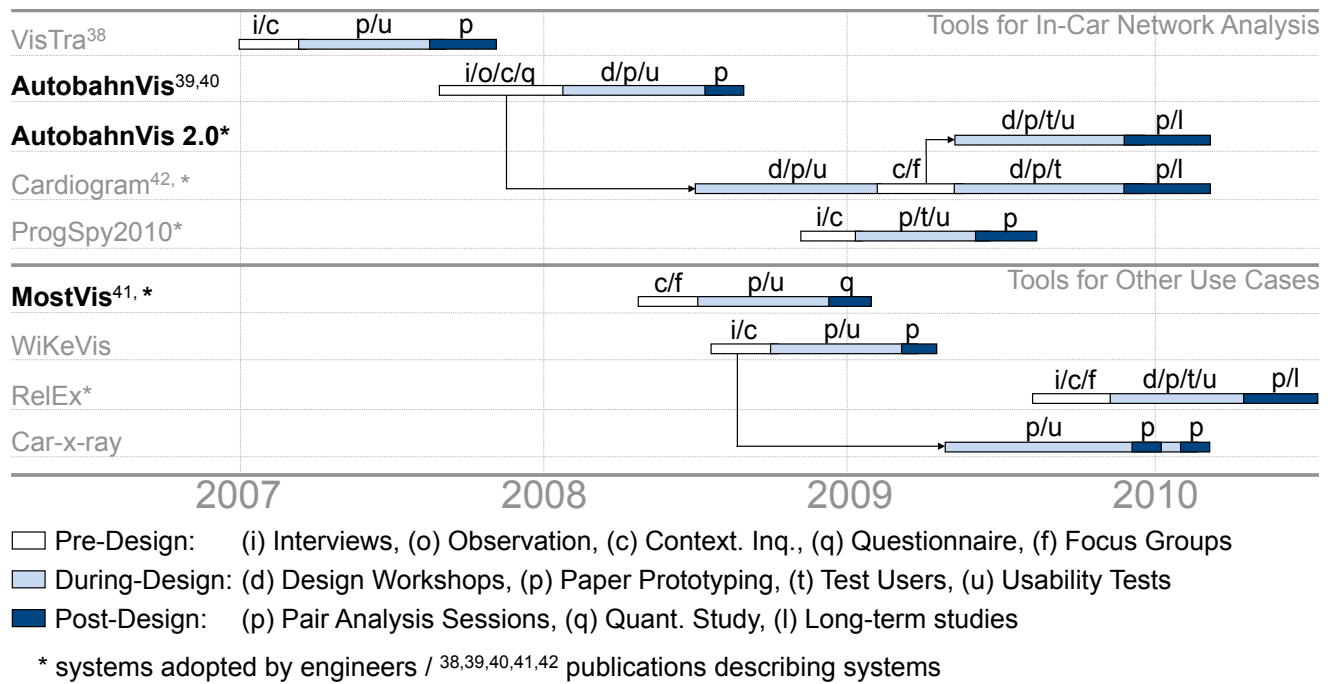


Figure 1: Temporal overview over evaluation and design methods we used in our visualization projects. The horizontal bars roughly show when a project was in which design phase and which evaluation and design methods we used in the particular phases. Naturally, these projects were not independent and strongly influenced each other; requirements from pre-design studies were shared among projects (explicit sharing is indicated via black arrows) or lessons learned derived and used as input for subsequent projects. Bold project names indicate projects that we discuss in this article's case studies.

notes and immediately preparing a summary after the observations. In order to collect more data on our observations, we conducted a second study in form of a brief online questionnaire. We contacted a wider range of analysis experts and asked questions on current work practices as derived from the first study. Our main goal was to reach a wider range of people without strongly interfering with their daily work (C4-TIME) and we eventually recruited 23 additional expert analysts.

4.1.2 Studies During the Design Phase

From the pre-design studies we derived a task list as well as design requirements. Based on these, we designed paper mockups and discussed them with eight domain experts in order to get feedback on our ideas. After we pre-filtered our ideas we developed a concept and validated and fine-tuned it in two design workshops by analyzing how it might be adapted to better support engineers in their daily work practices. A meta-goal of this phase of our studies was to evaluate the degree of interest of each participant. Three participants showed keen interest in our ideas and approaches. For our future collaboration we focused mainly on this sub-group (C3-CHOOSE).

Based on our final design concept, we built a prototype visualizing real data and iteratively evaluated the design. These studies were conducted with students with a usability background in order to save domain experts' time (C4-TIME) and due to challenges of integrating our tool in the domain experts' work environment (C1-INTEGRATE). We also iter-

atively conducted expert reviews with one usability expert with a research background in HCI/ InfoVis.⁴⁵ These two approaches helped us to focus on usability issues alongside the entire development process. For final revisions, we conducted a think aloud study with five students with a usability background plus basic automotive experience (automotive company interns). These studies were used for final usability optimization. The direct integration of our solution in the domain experts' current software environment was extremely difficult due to a variety of different file formats in use (C2-DATA). We therefore opted for an exported (standard) file format (C1-INTEGRATE) and, additionally, had to gather unavailable data (for controlling a semantically coordinated 3D-model view in our tool, cf. Figure 2) manually from textual sources (C2-DATA).

4.1.3 Post-Design Study

After the first study with students, we conducted a qualitative user study in form of a "pair analysis" evaluation to validate the value of our approach with five domain experts. Roughly speaking, one of us worked with one expert analyzing (real, but partly manually translated) test data using our tool and discussing potential benefits and drawbacks. Due to IPR restrictions, we again did not use audio and video recording (C6-CONFIDENTIALITY). During these studies we collected anecdotal evidence¹⁰ that domain experts were able to gain several novel insights from our tool such as the detection of message bursts, better visibility of message cycles and distribution over the network, and a more precise understanding

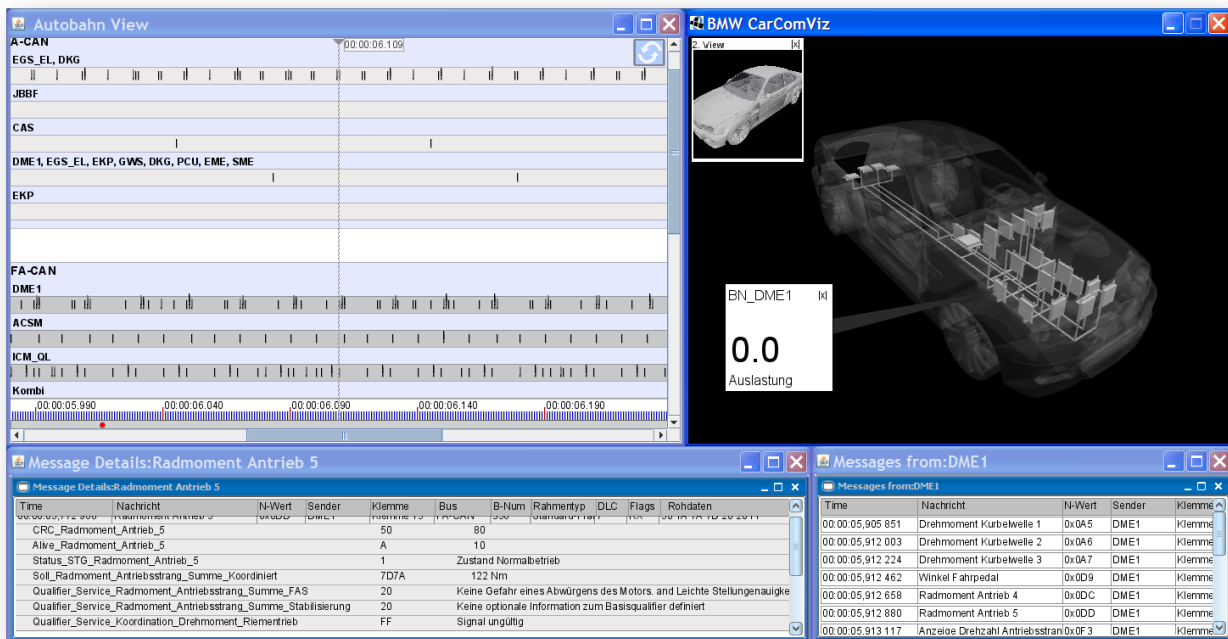


Figure 2: AutobahnVis—Top left: The *MessageView* showing all sent messages ordered by time (horizontally) and Control Unit (vertically); Bottom left and right: *List Views* showing message details, such as row data, signals, time stamps;³⁹ Top right: The *3DModelView* showing a 3D model of the car linked to the other views via semantic linking and linking and brushing.⁴⁰

of connections between electronic and mechanical vehicle behaviour. While these promising results did not provide us with examples from real work settings, they nevertheless (a) showed the validity of our approach for our test datasets and (b) helped us to promote it for a close integration with current analysis software. However, we also learned that only a close integration—without additional time costs (C4-TIME, e. g., exporting and manually translating data)—will allow analysis experts to use our visualization productively on a day-to-day basis with real world, dynamically varying datasets (C2-DATA).

4.1.4 AutobahnVis 2.0: Redesign and Integration

Based on our findings with the first version of AutobahnVis, we engaged in a re-design phase with the overall goal to study the final tool in long-term studies to gain insight into the lengthy, ill-defined and exploratory nature of data analysis tasks.¹⁶ The most important step to enable these kind of studies was to integrate our tool into the engineers' working environment (C1-INTEGRATE) in order to overcome additional time costs (C4-TIME) and to guarantee the seamless connection to real data (C2-DATA).

As a first step, we used our study results to convince stakeholders of the potentials of our approach (C8-STAKEHOLDERS). We presented our tool in a group meeting with stakeholders and provided examples of anecdotal evidence from the pair analysis sessions. This meeting was also attended by some of our study participants—i. e., end users—who liked our novel approach. These previous par-

ticipants supported us in arguing for a close integration of AutobahnVis into their daily practices. As a result, we received the permission to directly integrate our approach into an in-house analysis software. Integrating AutobahnVis into this platform meant a lot of additional software engineering work and technical barriers for us. We also had to dispense some of our initial ideas as they did not match processes and data-interfaces provided by the in-house tool—for instance, we excluded the coordinated 3D-model view as this kind of data was not supported with the in-house analysis tool. Finally, however, the integration of our tool allowed the engineers to use AutobahnVis with their own data, in their daily work practices and us to study our solution under real conditions (C1-INTEGRATE).

To validate this second version of AutobahnVis—AutobahnVis 2.0 (see Figure 3)—we conducted a short MILC study.¹⁶ We recruited five leading automotive analysts to use AutobahnVis 2.0 for eight weeks and gathered qualitative feedback weekly via telephone or in personal meetings. Our participants used our tool during their daily work, for their daily problems, and with their own data of interest (C2-DATA), and provided us with examples where AutobahnVis 2.0 helped them to analyze their data as well as with suggestions to improve it. The fact that AutobahnVis 2.0 was incorporated with their practices also allowed our participants to smoothly use and combine Autobahn 2.0 with conventional techniques from the in-house tool (C5-CONVENTIONS). This helped us to get a richer and more holistic understanding of the entire analysis process (C7-COMPLEX), but also in convincing engineers to participate in our study as they could immediately draw benefits

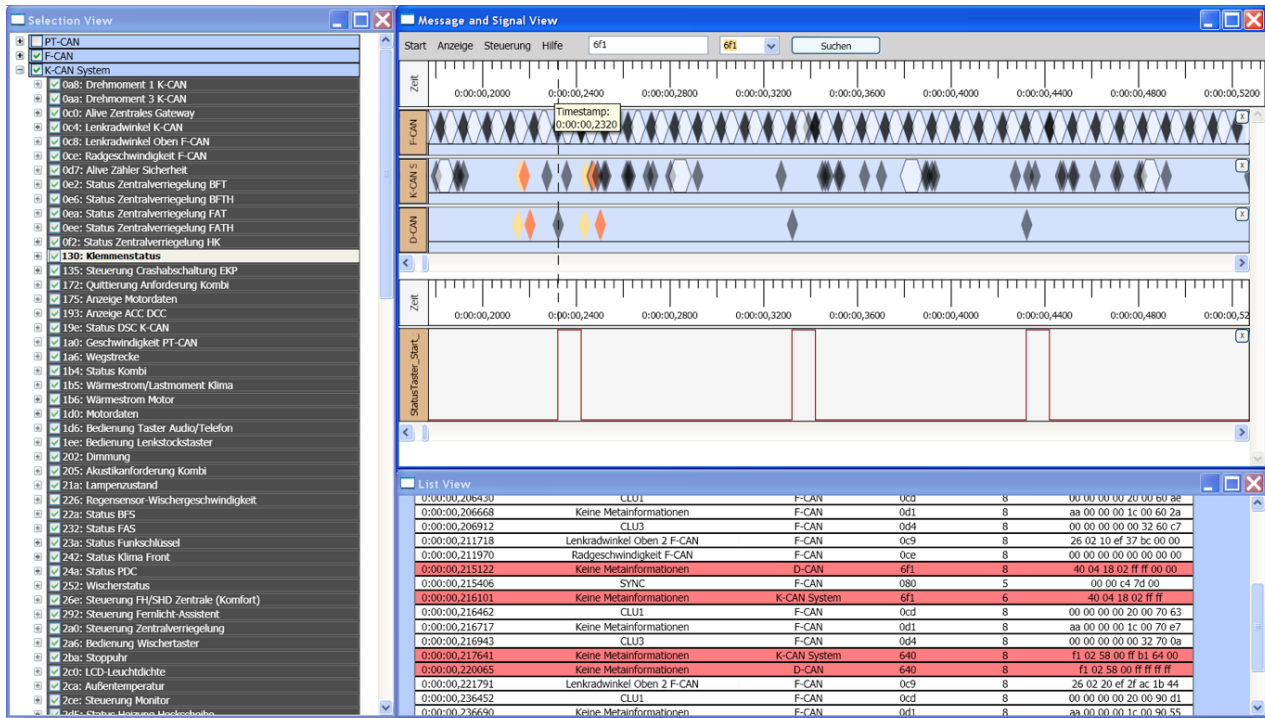


Figure 3: AutobahnVis 2.0—The version of AutobahnVis that we closely integrated with an in-house analysis tool: The original 3DModelView was removed due to missing support for this kind of data by the in-house analysis tool. Based on further studies with domain experts, we also fine-tuned our initial approach (e.g., MessageView: diamonds instead of rectangles indicating exact timing; introduction of semitransparent message clusters) and enriched it with other views (FilterView on the left; SignalView right/center).

from our approach without unnecessary extra time costs (C4-TIME, C8-STAKEHOLDERS).

After the study, we provided our stakeholders with anecdotal evidence from engineers' daily practices underlining our approach's domain value. We showed concrete examples how AutobahnVis 2.0 enriched the engineers' practices by (a) providing insights into aspects of the data unknown so far, or (b) speeding up their work. Our approach also led to the adoption of AutobahnVis 2.0: Our participants still use the tool and we also have recently transferred the code to the tool developers of the in-house tool who are now working on making AutobahnVis 2.0 available as a core-component and to make it accessible for all analysts in the company.

4.2 Case Study 2: MostVis

Our second example, MostVis,⁴¹ was not meant as direct support for error analysis, but as an alternative visual access to auxiliary information. It visually represents large, hierarchically organized information catalogs of one or more bus systems of a car model. Browsing these catalogs has become infeasible on paper as well as with currently used textual database interfaces, so we took a visualization approach to re-enable the analysts' access. In the development of MostVis we followed a similar approach to AutobahnVis including pre-design studies, studies during the design phase, as well as a post-design study. In this section, we solely focus on

experiences that differ from the ones we described above instead of repeating parts of our process description.

4.2.1 Pre-Design Studies

We began our evaluations again with contextual inquiries, based on our positive experience in the AutobahnVis project. After that—instead of distributing a questionnaire—we proceeded by setting up a focus group (see, for instance, Morgan⁴⁶) discussion with five domain experts in order to validate and generalize the findings from contextual inquiries. While the number of participants we reached with this approach was lower than with the questionnaire, we found the qualitative results of the focus group very informative for our design. The results of the focus groups provided us with a more holistic and comprehensive understanding of our application area than the questionnaire did. The focus group also helped us to discuss diverse and even opposed practices and statements we found during individual interviews and helped us to form a common understanding, not just between the experts and us but also among the experts themselves.

4.2.2 Post-Design Study

After designing and implementing MostVis in a user-centered design process (similar to the one described in the AutobahnVis case study), we conducted a quantitative user

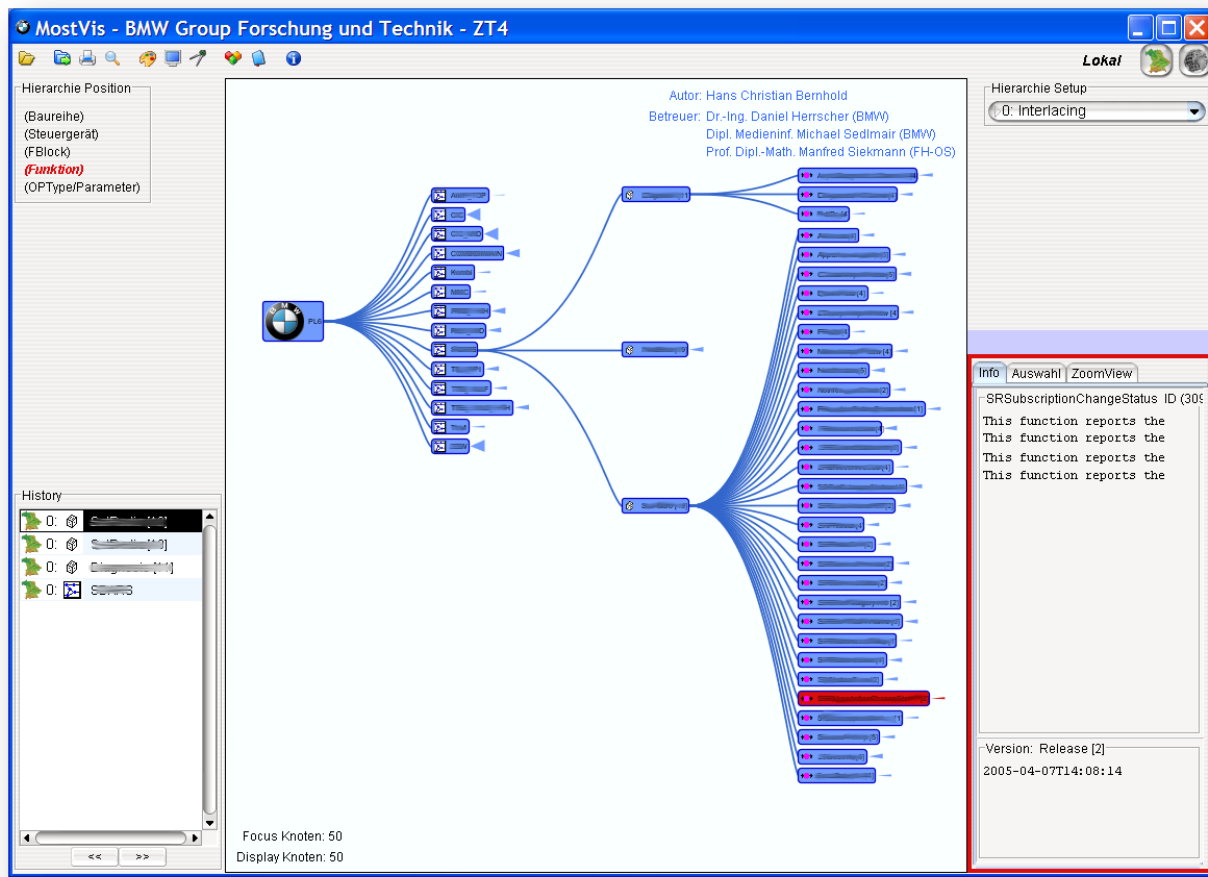


Figure 4: MostVis with an interactive, left-to-right presentation of a MOST function catalog; details, zoom, interaction history and grouping functionality in additional views.⁴¹ Labels have been blurred due to IPR restrictions.

study, instead of a qualitative user study as in the case of AutobahnVis. Our quantitative study was conducted with 14 domain experts from our target group of engineers to assess performance, understandability, as well as to collect qualitative feedback. Finding participants for this study was easier compared to AutobahnVis for two reasons: First, MostVis was a pull-solution (i. e., directly requested by employees). Second, our study was less time-intensive and took under one hour per participant (C4-TIME). We tested a set of domain-specific tasks and for each task measured time and errors with MostVis and with current tools. We conducted a statistical comparative analysis and showed that MostVis was significantly faster and less error-prone for experts (C5-CONVENTIONS) and non-expert users.⁴¹

This study approach lead to different results compared to the post-design studies of AutobahnVis. Unfortunately, we did not learn as much about insight generation and the analysis process due to our controlled study setup. As MostVis used a completely different analysis approach than the tool we compared it to, we could not specify exactly which parameter/variable was the reason for the improvements but could only give a general statement about increased speed in the tasks we tested in our study.⁴⁷ Despite these limita-

tions, we saw some important benefits of this study for the large company setting. The statistically significant results which showed that MostVis sped up work processes were very convincing to the stakeholders and opened up new opportunities for integrating our tool in the current software environment (C8-STAKEHOLDERS). After seeing our results, the stakeholders decided to additionally fund and tightly integrate MostVis with a large database that was subject to strict access regulations. Without their agreement we would probably have missed such an opportunity (C1-INTEGRATE). Conducting a quantitative user study was therefore our ticket to reach a large number of end-users (C4-TIME) with manageable effort and real data (C2-DATA) in real environments (C1-INTEGRATE) and opened new possibilities for future long-term and more in-depth studies.*

4.3 Reflections on Evaluation Methods

In this section we reflect on the breadth of evaluation techniques we used during our work in industry—including methods from the case studies above but also such used in our

*At the time of writing this article, the integration process of MostVis is underway but has not yet been completed.

other projects. We strive for completeness in order to provide a holistic overview of all evaluation methods we used and a summary of our experience we gained in using them. Table 2 provides an overview over all evaluation methods we employed. For a temporal overview over all our projects and the particular evaluation methods we used during each project see also Figure 1.

4.3.1 Pre-Design Studies

One important factor for success in our projects turned out to be to carefully study end-users' current practices, problems, and requirements before starting the actual design of our own tools. We therefore want to echo the call by other researchers for the importance of such pre-design studies^{10–12} and especially emphasize its value in a large company context where it helps to better understand the variety of tools and practices in use (C7-COMPLEX). In our projects, we used the following pre-design study methods:

- *Interviews:* After a review of technical documentations and literature, we usually started our projects with semi-structured interviews to get first-hand information about practices, challenges and problems of our end-users. While we found the “pre-filtered” opinions about problems and challenges helpful for understanding current practices, our experience, however, also underlined that interviewing alone could not provide insights into all important facets of current analysis techniques (see AutobahnVis, Section 4.1).³¹
- *User Observation and Contextual Inquiries:* To get a richer and more realistic understanding of our participants' practices and needs, we often enriched the information gathered from interviews with observational studies and contextual inquiries. The high specialization and complexity of our end-users' tasks (C7-COMPLEX) hindered us in using pure observational studies without any interruptions of work processes (see AutobahnVis).
- *Questionnaire:* In our AutobahnVis project, we also used a questionnaire to evaluate and generalize our findings from the qualitative studies. This questionnaire particularly helped us to reach more and especially time-restricted participants (C4-TIME), and to validate (simple) questions with a broader audience of potential end-users (e. g., to find out about current analysis tools used by our experts). Compared to our other pre-design methods, however, the questionnaire did not provide any deep insights into data analysis processes and therefore was only helpful as an extension to other techniques (C7-COMPLEX).
- *Focus Groups:* Another technique that we frequently used for evaluating findings from individual interviews and observations was focus groups (see MostVis, Section 4.2). Because of their interactive and group-oriented structure we found focus groups very helpful

in terms of clarifying inconsistent findings from individual studies and for getting a more holistic and coherent understanding. In our case it also turned out to be helpful, not only to invite end-users to our focus groups but also designers of current analysis tools and stakeholders to let them participate early in the process (C8-STAKEHOLDERS).

4.3.2 Studies during the Design Phases

Similar to our AutobahnVis example, we designed all our tools in close cooperation with end-users applying a participatory design approach⁴⁸ with several steps of prototyping. Our general approach is in spirit similar to McLachlan et al.'s staged design and development process¹⁸ and turned out to be a valid approach in our domain, finally leading to the adoption of five of our tools into current practices. During our design phases we especially used the following methods:

- *Design Workshops:* Based on our positive experience with focus groups for pre-design studies, we similarly set up design workshops with end-users, tool designers and stakeholders in order to let them participate in our design decisions (e. g., AutobahnVis). In these workshops we introduced information visualization techniques to the participants, discussed our and their ideas and fine-tuned possible solutions (features, system designs, etc.). We also used design workshops to validate basic concepts of our tools before we started implementing them. Group discussions helped us to rule out concerns and problems, and to merge our ideas with the ideas of our experts (C1-INTEGRATE).
- *Paper Prototyping and Mockups:* For validating and refining our ideas and designs, we also used paper prototyping sessions with single experts (see also AutobahnVis). In such sessions we provided the experts with paper mockups of our systems, discussed benefits and drawbacks, and used this information to adapt our final concept.
- *Test Users:* Especially when we developed our later, closely integrated tools (e. g., Autobahn 2.0), we got the chance to improve our designs based on feedback from engineers using our tools under realistic work conditions. Towards the end of the development process of these projects, we provided a group of carefully chosen test engineers with frequent iterative releases of our tools and elicited feedback in personal meetings or via telephone. Feedback in this phase helped us to adopt our tools to the specific needs of our test users. To save engineers time (C4-TIME), we usually combined this phase with post-design long-term studies (see below) where our test users provided us with both, suggestions for improvement as well as anecdotal evidence of the usefulness of the tool.
- *Dedicated Usability Tests:* In order to fix usability problems that were not domain-specific we applied traditional usability methods, most importantly heuristic

<i>Evaluation method</i>	<i>Number of times performed</i>	<i>Cost estimation</i>	<i>Helpfulness</i>
interviews	~50	low	helpful
pure user observations	2 pilot studies	low	not helpful
contextual inquiries	~25	low	very helpful
questionnaire	1 with 23 participants	medium	helpful
focus groups	~25 with 3–10 participants each	low	very helpful
design workshops	~20	low	very helpful
paper prototyping	~50 different participants	low	very helpful
test users	5 tools with 2–6 test users each	medium (integration helpful)	very helpful
usability tests	~50 sessions in 11 studies	low	helpful
pair analysis sessions	8 tools with 1–12 participants each	low (no integration necessary)	helpful
quantitative study with domain experts	1 with 14 participants (see MostVis, Section 4.2)	low (no integration necessary)	helpful
long-term studies	3 tools with 2–7 participants each; 5–8 weeks in length*	high (integration necessary)	very helpful
informal collaboration	>200 casual meetings with ~150 domain experts	low	very helpful

* We also partly participated in the evaluation of an automatic data analysis approach with 15 engineers over the period of 1 year.⁴²

Table 2: Overview of the evaluation methods we used in our projects: The table shows how often we applied the particular techniques. We also include our personal estimation of costs for the researcher in terms of time for preparing and running these studies. Finally, we added our estimation of how helpful we found a particular study type for gaining new insights and developing successful information visualization tools in our domain.

evaluations,⁴⁹ expert reviews and think-aloud studies. As mentioned in the example of AutobahnVis, we invited outside testers for these studies in order to save domain experts' time (C4-TIME).

4.3.3 Post-Design Studies

Our post-design studies generally focused on evaluating domain utility, i. e., on the question if our approaches provided domain experts with novel insights (see AutobahnVis) or if they were able to speed up current practices (see MostVis). On a meta-level we also used the study results to convince stakeholders (anecdotal evidence for AutobahnVis & significant speedup for MostVis). The evaluation method we used for our post-design studies strongly depended on the degree of integration of our tools (cf. C1-INTEGRATE). We used the following techniques:

- *Pair Analysis Sessions:* This is the evaluation method we usually used when our tools were not properly integrated into the end-users' environment (e. g., first post-design study with AutobahnVis; cf. C1-INTEGRATE, C2-DATA). Testing our tools with prepared datasets in artificial work situations helped us to gain insight into domain potentials of our approaches and into aspects for improvement, but did not tell us about the long-term, exploratory nature of real analysis processes with our tools.
- *Quantitative study with domain experts:* In the MostVis project we conducted a quantitative study with domain experts to learn about engineers' task-performance compared to traditional tools. While the insights gained into the data-analysis process were small, the results were very convincing to stakeholders (C8-STAKEHOLDERS) and finally led to funding and integration.

- *Long-term studies:* We agree with Shneiderman and Plaisant¹⁶ that studies with end-users, with real data, in real environments are extremely important factors for learning about information visualization tools. In our large company setting it turned out that long-term studies require close integration of novel tools with current practices (C1-INTEGRATE) and of all important real data sources (C2-DATA). While the process of integration usually comes with additional technical and organizational hurdles (C1-INTEGRATE), we found that it is worth to overcome them in order to study InfoVis tools in realistic environments.

4.3.4 Informal Collaboration

We found it also invaluable to enrich our formal studies alongside the entire process with informal collaboration, such as meetings in the company's cafeteria, a joint lunch, or casual discussions at the workplace. Informal collaboration with our experts helped us to better understand the various facets of our target domain, iteratively refined our requirements, and to establish novel successful tools. To counterbalance the restricted opportunity of logging data from this kind of informal studies, we always carried notepads in case spontaneous conversations would occur.

5 Recommendations

Based on our experience we finally derive a set of concrete recommendations for other information visualization researchers who are planning to conduct evaluations within a large company setting. The organization of our recommendations reflects the main categories of challenges in Section 3. Some of our recommendations are specific to working with

data and data analysis tasks with information visualizations and some apply to evaluation in this field more generally. Table 3 gives an overview of all recommendations and summarizes in which evaluation phase a particular recommendation may be particularly important to consider.

5.1 Study/Application Design

R1-INTEGRATE: Overcome Technical Obstacles of Tool Integration

To evaluate the full work process of domain experts an information visualization tool should be integrated and coordinated with current domain specific techniques and tools to operate in complete analysis environment. Many of the existing tools in a work environment, however, have often grown and expanded over the years and integration may be a considerable software engineering challenge. In our AutobahnVis project, for instance, we spent more time on overcoming technical restrictions than on designing and implementing the actual visualizations. In the end, however, overcoming the technical hurdles of integrating our tool into the end-users' environment paid off in the opportunity to evaluate AutobahnVis over a longer period of time under realistic circumstances. Of course, instead of an integration one can also consider extending the features of a new visualization tool to unite the capabilities of a previous tool chain. Depending on the amount of previous work this could be a valid solution as well. The costs of either solution should be considered based on the goal of the evaluation. Supplementing existing tools, however, is often the cheaper and more effective way.¹⁷

R2-DATA: Avoid Additional Steps to Work with Domain-specific Data

In line with the previous recommendation, it is similarly important that new tools provide no additional overhead to work with domain-specific data. While our first version of AutobahnVis, for example, was more flexible to adapt to different data and analysis settings, engineers argued that the additional required overhead of file conversion was their reason for not using the tool. For the purpose of evaluating AutobahnVis in more depth we redesigned and tightly integrated AutobahnVis with a conventional data management and analysis tool. In large company settings, tight coupling to the data, even if that is just a subset of all possible data, may be more important than supporting wide applicability. This factor may not be important in research departments (where insight may outweigh time), but the obstacle of additional time requirements is crucial in industrial environments. Having to convert data manually should be a last resort.¹⁶

R3-ENVIRONMENT: Choose your Study Environment with Care

Obstacles for studying solutions or studying work environments often result not only from technical challenges but from political or organizational requirements. To conduct evaluations researchers need permissions and committed collaborators. In order to find those, having dedicated employees who support the project and convince stakeholders is imperative (see R4-PULLPUSH for further recommenda-

tions). Similar data analysis tasks and data types can be found across different groups within a large company. It takes skill as a researcher to generalize from the individual opinions and views encountered to find the right target group and work environment to deploy a planned or built tool and receive the most valuable feedback.

When conducting pre-design studies, connecting with motivated domain experts and starting by identifying and understanding different sub-problems and sub-groups in a problem domain are important first steps. Talking to various people and being open-minded towards existing solutions from other people beyond the target group is equally helpful. When using this knowledge to become an expert in the respective domain, it is important not to try to solve everyone's problems. A more promising course is to find a specific sub-target group with specific and concrete problems and with interest in one's work. After producing specific solutions and validating them, abstracting from the specific lessons learned to a more general approach and giving it back to the research community is the final step.

R4-PULLPUSH: Consider both Employee-pull and Researcher-push Solutions

Generally, we distinguish between two kinds of solutions: (a) *employee-pull solutions* (such as MostVis) where one or several employees request a specific tool, and (b) *researcher-push solutions* (such as AutobahnVis) where the visualization expert advertises a tool. Both approaches may be successful: Employee-pull solutions are often easier because employees can argue that a tool may address a recognized analysis problem and have more trust by the stakeholders than external researchers. However, they are not an inevitable factor for success, as this also depends on their position within the company hierarchy, how important the problem the visualization solves is, and of course on the quality of the solution. Researcher-push solutions, on the other hand, may require very tactful negotiation but are no less important. Specific work practices may have become established over the years and employees may be satisfied with those improvable solutions. In these cases, a push from an outsider can help to provide a new perspective on more advanced data analysis options (cf. AutobahnVis).

R5-DELIGHT: Delight with Usability and Aesthetics, Avoid Eye-candy

The value of usability and aesthetics in visualization should never be underestimated. In-house tools often neglect these aspects,²⁷ so they can become important distinctive features to be used to gain acceptance of novel tools and to convince stakeholders. During the development of MostVis, for example, we intensively focused on usability engineering aided by several usability studies. The higher usability compared to current tools definitely was a major reason for our good results in the comparative study which finally led to the tool being integrated in a larger work context. During our studies with the early 3D view in AutobahnVis our subjects frequently pointed out the aesthetics of the solutions and mentioned that it would be much easier to convince decision makers with such solutions. Still, it takes care to not confuse

Short name	Description	pre	during	post
<i>Study/Application Design</i>				
R1-INTEGRATE	overcome technical obstacles of tool integration		x	x
R2-DATA	avoid additional steps to work with domain-specific data		x	x
R3-ENVIRONMENT	choose your study environment with care	x	x	x
R4-PULLPUSH	consider both employee-pull and researcher-push solutions	x	x	
R5-DELIGHT	delight with usability and aesthetics, avoid eye-candy		x	
R6-SUPPORT	smooth installation and tech support		x	x
<i>Participants</i>				
R7-ONEHOUR	the magic one hour limit	x	x	x
R8-CONVINCE	convincing the target audience		x	x
R9-LEARN	learn from the experts	x		
R10-OUTSIDERS	conduct usability studies with outside testers		x	
R11-REMIND	gentle reminders			x
<i>Data Collection</i>				
R12-LEGAL	try to get an IPR license, do studies in any case	x	x	x
R13-CCC	be in constant, close cooperation	x	x	x
<i>Results</i>				
R14-MONEY	the magic metric: money			x
R15-SKILL	factor in high skill with current techniques			x
R16-PUBLISH	clarify publishing conditions upfront	x	x	x

Table 3: Recommendations Summary: The last three columns indicate the phase (pre-design, during-design, post-design) where a specific recommendation may particularly support information visualization researchers' work in a large company setting (marked with an "x").

"aesthetics" with "eye-candy": Our experiences with trace analysis engineers, for example, showed that rather simple and easily understandable solutions were strongly preferred.

R6-SUPPORT: Smooth Installation and Tech Support

In large company settings, employees' computers often have restricted user accounts, fine-tuned operating systems or specific security policies. Clarifying all relevant technical aspects upfront and guaranteeing a smooth operation and easy installation of visualization tools on these computers is a must when hoping for results that are not biased by software bugs. Easy installations and updates are particularly important when planning for long-term studies such as MILCs¹⁶ where participants may be provided with frequent tool updates.

Also, researchers should provide good and fast technical support. Due to the ever-present time pressure (C4-TIME) even small technical barriers can lead to forfeiting all conceptual benefits a tool provides. Good technical support turned also out to be helpful in terms of training with our new software systems. While our collaborators rarely used the written documentation we provided (lack of time, see C4-TIME), they usually relied on the personal introduction we gave them when we first delivered a tool, and then called or emailed us whenever they had further questions.

5.2 Participants

R7-ONEHOUR: The Magic One Hour Limit

Our experience showed that recruiting participants for one hour or less is significantly easier than for longer time periods. Employees are occupied with meetings, appointments, and deadlines and additional involvement in user studies just

adds to this work load. Being prepared and professional in recruiting and conducting the study and sticking to the scheduled time limit are very important.

R8-CONVINCE: Convincing the Target Audience

Even though participants may be very attached and used to their current tools there are some things one can do to convince them of a solution. Trying to solve real problems of a target group even if these first-hand solutions are small and actually not the main focus of the work can help tremendously! People become immediately interested in solutions which they can use with their own data. Participants will be much more motivated to attend studies when they know they will be remunerated by working on solutions to their current problems. One way to achieve this is to integrate some simple but highly desired functions not available with current tools (for a successful example see specific features we added to MostVis⁴¹). Even outlining solutions, e. g., presenting some work of our initial AutobahnVis ideas in the early design phase, was very valuable to convince our participants of the potential value of our work.

R9-LEARN: Learn from the Experts

An important first step is to identify experts in a problem domain. Researchers can learn a lot by interviewing and observing their practices. Often, experts may not be interested in novel solutions because they have mastered problems already using their own approach. Trying to identify why their practices are effective and efficient and thinking about how this knowledge can be applied in a visualization tool can make it available to a wider range of people. During our pre-design studies of AutobahnVis, for example, talking to one specific expert helped us enormously in understanding the variety of potential error sources and the importance of a

hexadecimal representation. Our tool design benefited from his experience.

R10-OUTSIDERS: Conduct Usability Studies with Outside Testers

In all our projects we conducted various heuristic and think-aloud studies alongside the development process. These studies helped us to focus on usability issues and we often conducted them with students and external testers with a usability background.⁴⁵ This did not replace usability studies with target users—as they are the only ones who provided domain expert feedback—but it definitely helped us to save valuable experts' time.

R11-REMIND: Gentle Reminders

Especially in long-term studies it might be valuable to gently remind participants of the existence of a tool. Employees usually have a variety of different, not only analysis-related tasks and a visualization tool may therefore only support a fraction of these tasks. Time-pressure and long periods between analysis tasks may lead experts to revert back to confirmed habits regardless of the tool's quality. Gently reminding participants of the benefits can help both the researcher and the practitioner, in slightly integrating the tool with daily work practices and in studying the utility of it. Informal meetings can be a good opportunity for such reminders.

5.3 Data Collection

R12-LEGAL: Try to get an IPR License, Do Studies in any Case

Checking IPR policies (see C6-CONFIDENTIALITY) and, if required, trying to get permission to video or audio tape participants and equipment are central to keeping research within legal boundaries. We agree with Dix et al.³⁰ that the analysis of recorded video or audio helps to gain a much deeper understanding of the scenario under study. If a permission is received, equipment has to be carefully installed. It is imperative that participants know about recording devices and that privacy concerns are thoroughly discussed. In particular in large companies, employees may be concerned about the company "watching" them.

In some areas IPR restrictions might be very strict and digitally recording study sessions may not be possible. In these cases, qualitative user studies can be done anyway and the loss of documentation can be counterbalanced with more than one observer and immediate notes and a summary (see AutobahnVis). Especially in secure areas this methodology additionally may allow participants to be more open about their work processes, data, and tasks.

R13-CCC: Be in Constant, Close Cooperation

To support specific domain experts with information visualization it is important to get a clear understanding of their problem domain.⁷ We have made good experiences with informal collaborations that helped us to get a very well-grounded and detailed knowledge about our target group: We have talked to almost 150 domain experts, we conducted several types of studies (from pre- to post-design) and we di-

rectly worked together with the domain experts. We refer to such a process as "constant, close cooperation". The ambitious goal was to gain a deep understanding of our problem domain. From our experience, this kind of constant, close cooperation is valuable especially in large industries where problems are often highly diverse and complex (C7-COMPLEX). If possible, being flexible and spontaneous in terms of time can help to counterbalance business and time-restrictions. Especially for investigating daily practices this helped us to observe "real" situations, both in our pre-design studies and in summative long-term studies. We are aware that understanding all facets of a problem domain is time-intensive, however, we found that this approach helps to clearly tailor solutions to the needs of a target group and to develop effective and efficient tools. Being in constant, close cooperation can help to overcome some of the pitfalls of evaluation.¹⁰

5.4 Results

R14-MONEY: The Magic Metric: Money

In industrial settings the benefits of a new tool are often measured in terms of cost savings. These savings are closely related to other metrics used in information visualization evaluation such as insights¹⁹ or errors⁵⁰ (note that there are of course also differences between these metrics, e. g., cost savings and reduction of errors might be easier to measure than insights). In an industry setting, the most related one may be time (again, time equals money) and important quality metrics for stakeholders reflect this, for instance, in terms of decisions per hour.³³ Reporting the results of a study and presenting evidence that a tool can lead to measurable benefits in terms of such metrics may be very important for convincing the stakeholders (see C8-STAKEHOLDERS). While studies that measure these metrics may not always be relevant for more holistic research questions, they could be a ticket for reaching more domain experts and studying solutions in-depth in real working environments. MostVis is a good example for such an "enabler" study: Proving that MostVis was significantly faster for a set of predefined user-tasks was very convincing to our stakeholders and we finally received the opportunity to tightly integrate MostVis into a current software environment that is subject to strict access regulations (C8-STAKEHOLDERS). Conducting a quantitative user study was therefore our ticket to reaching many end-users (C4-TIME) with real data (C2-DATA) in real environments (C1-INTEGRATE) and opened new possibilities for future long-term and more in-depth studies.

R15-SKILL: Factor in High Skill with Current Techniques

When comparing traditional to new tools, one must consider that participants may have become very skilled with current techniques (see C5-CONVENTIONS) and factor in learning time and potential reluctance towards a new tool as these factors can initially distort a comparative evaluation.¹⁶

R16-PUBLISH: Clarify Publishing Conditions Upfront

Often a main part of research is publication. In this case, it

is vital to make concrete agreements with a company upfront and preferably not just verbally. Discussing details beforehand such as what parts of the results can be published, how and if they have to be anonymized, what pictures (if any) are allowed to be included and if the company wants internal reviews for drafts can prevent bad surprises.

6 Discussion: Advantages, Risks and Limitations

In this article, we summarize design challenges, examples, and recommendations for working within a large company to develop, deploy, and evaluate information visualization tools. We thus emphasize the difficulties of conducting research in cooperation with a large company, rather than focusing on the advantages. Nevertheless, we want to encourage research cooperations with large companies for several reasons: As already mentioned, large companies provide a lot of interesting challenges and complex real-world datasets for information visualization research. Based on the tight network between employees in a large company we also found it relatively easy to get in contact with a wide range of potential end-users in order to study their practices and our tools. In addition, although deployment and evaluation might be a long and laborious process, there are good chances that valuable solutions will be approved and integrated into real working environments. Thus, domain experts can benefit from dedicated information visualization solutions and researchers in return can investigate their systems under realistic circumstances.¹⁶

We also have not explicitly focused on risks that might appear in research collaborations with large companies. These risks usually result from the different goals of large companies and researchers (see, for instance, C8-STAKEHOLDERS). Our challenge C9-PUBLISHING goes in this direction. Beyond that, it is also possible that a desired solution for the company might not align well with the goals of the researcher, e.g., the information visualization techniques (s)he wants to investigate. Such mismatches can result in collaboration breakdown.⁵¹

Since a large part of our work is based on our own experiences there is a limitation to be aware of: Experiences in other companies might differ or go beyond the ones we made. While the lessons we learned can serve as a reference for others who are planning information visualization evaluations within a large company context, this work should also encourage information visualization researchers (a) to report their own experiences evaluating within this field to create a broader knowledge base, and (b) to try out alternative evaluation approaches in order to address the challenges outlined.

In this article, we specifically outlined the advantages of integrating tools into industrial workflows (see C1-INTEGRATE). We found this type of applied work very rewarding as it provided considerably more realistic insights than lab studies—even though conducting these studies was often more diffi-

cult than a lab-based study in particular because it takes additional time and effort to integrate tools into industrial workflows. Of course, one might argue that research does not immediately benefit from the additional engineering costs that are required of such undertakings. However, we argue that eventually ‘moving research into practice’ remains one of our grand challenges² and will feed back valuable information to our research.⁵² In those lines, closely collaborating with large companies and integrating/studying our tools in such environments can help to broaden our understanding of the practical value of information visualization.

7 Conclusions

We presented a number of specific challenges arising in the context of evaluating information visualizations in a large company with several thousands of employees. These nine challenges are grouped into those relating to: study/application design, participants, data collection, and results. This collection is based on a 3.5-year body of work involving nine prototypes and their development process. To illustrate these challenges, we discussed our experience in developing several visualization systems for a large automotive company and summarized the different information visualization evaluation techniques we used. Based on our experience, we presented a set of sixteen recommendations for practitioners. With this collection of experiences and insights we hope to help others in preparing and conducting information visualization evaluations in similar settings and to encourage them to add and compare their experiences to our work.

8 Acknowledgments

We would like to thank the BMW Group Research and Technology for funding this work. We also thank all our reviewers as well as Tamara Munzner, Miriah Meyer, Stephen Ingram and Robert Schuster for their valuable comments and suggestions to further improve this article.

References

1. Plaisant C. The Challenge of Information Visualization Evaluation. In: Proc. Advanced Visual Interfaces (AVI). ACM; 2004. p. 109–116.
2. Thomas JJ, Cook KA. Illuminating the Path. IEEE Computer Society; 2005.
3. Ellis G, Dix A. An explorative analysis of user evaluation studies in information visualisation. In: Proc. of the Workshop Beyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV). ACM; 2006. .

4. Preece J, Rogers Y, Sharp H. *Beyond Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons, Inc.; 2001.
5. Saraiya P, North C, Duca K. An Insight-Based Methodology for Evaluating Bioinformatics Visualizations. *IEEE Transactions on Visualization and Computer Graphics*. 2005;11(4):443–456.
6. Tory M, Möller T. Human Factors in Visualization Research. *IEEE Transactions on Visualization and Computer Graphics*. 2004;10(1):72–84.
7. Paley WB. Interface and Mind - A “Paper Lecture” about a Domain-Specific Design Methodology based on Contemporary Mind Science. *it-Information Technology*. 2009;51(3):131–141.
8. McGrath JE. Methodology Matters: Doing Research in the Behavioral and Social Sciences. In: *Human-Computer Interaction: Toward the Year 2000*. Morgan Kaufmann Publishers; 1995. p. 152–169.
9. Daft RL. *Organization Theory and Design*. West St. Paul, MN; 1995.
10. Munzner T. A Nested Process Model for Visualization Design and Validation. *IEEE Transactions on Visualization and Computer Graphics*. 2009;15(6):921–928.
11. Isenberg P, Zuk T, Collins C, Carpendale S. Grounded Evaluation of Information Visualizations. In: *Proc. of the Workshop Beyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV)*. ACM; 2008. .
12. Carpendale S. Evaluating Information Visualizations. In: Kerren A, Stasko JT, Fekete JD, North C, editors. *Information Visualization: Human-Centered Issues and Perspectives*. vol. 4950. Springer LNCS; 2008. p. 19–45.
13. Sedlmair M, Isenberg P, Baur D, Butz A. Evaluating Information Visualization in Large Companies: Challenges, Experiences and Recommendations. In: *Proc. of the Proceedings of the Workshop Beyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV)*. ACM; 2010. p. 79–86.
14. Perer A, Shneiderman B. Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis. In: *Proc. of the Conference on Human Factors in Computing Systems (CHI)*. ACM; 2008. p. 265–274.
15. Tory M, Staub-French S. Qualitative analysis of visualization: a building design field study. In: *Proc. of the Workshop Beyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV)*. ACM; 2008. .
16. Shneiderman B, Plaisant C. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In: *Proc. of the Workshop Workshop Beyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV)*. ACM; 2006. .
17. González V, Kobsa A. A Workplace Study of the Adoption of Information Visualization Systems. In: *3rd International Conference on Knowledge Management (I-KNOW)*; 2003. p. 92–102.
18. McLachlan P, Munzner T, Koutsofios E, North S. LiveRAC: Interactive visual exploration of system management time-series data. In: *Proc. of the Conference on Human Factors in Computing Systems (CHI)*. ACM; 2008. p. 1483–1492.
19. Saraiya P, North C, Lam V, Duca KA. An insight-based longitudinal study of visual analytics. *IEEE Transactions on Visualization and Computer Graphics*. 2006;12:1511–1522.
20. González V, Kobsa A. Benefits of Information Visualization Systems for Administrative Data Analysts. In: *Proc. of the International Conference on Information Visualization (IV)*. IEEE Computer Society; 2003. p. 331–336.
21. Bak JO, Nguyen K, Risgaard P, Stage J. Obstacles to usability evaluation in practice: A survey of software development organizations. In: *Proc. of the Nordic Conference on Human-Computer Interaction (NordiCHI)*. ACM; 2008. p. 23–32.
22. Mao JY, Vredenburg K, Smith PW, Carey T. The state of user-centered design practice. *Communications of the ACM*. 2005;48(3):105–109.
23. Grudin J. Obstacles to participatory design in large product development organizations. *Participatory Design: Principles and Practices*. 1993;p. 99–119.
24. Poltrock SE, Grudin J. Organizational obstacles to interface design and development: two participant- observer studies. *ACM Transactions on Computer- Human Interaction*. 1994;1(1):52–80.
25. Jeffries R, Miller JR, Wharton C, Uyeda K. User interface evaluation in the real world: a comparison of four techniques. In: *Proc. of the Conference on Human Factors in Computing Systems (CHI)*. ACM; 1991. p. 119–124.
26. Boivie I, Åborg C, Persson J, Loeffberg M. Why usability gets lost or usability in in-house software development. *Interacting with Computers*. 2003;15(4):623–639.
27. Holtzblatt K, Barr J, Holtzblatt L. Driving user centered design into IT organizations: is it possible? In: *Extended Abstracts of the Conference on Human Factors in Computing Systems (CHI)*. ACM; 2009. p. 2727–2730.
28. Pugh DS, Hickson DJ, Hinings CR. An Empirical Taxonomy of Structures of Work Organizations. *Administrative Science Quarterly*. 1969 Mar;14(1):115–126.
29. Drucker PF. The Coming of the New Organization. *Harvard Business Review*. 1988;66(1):45–53.

30. Dix A, Finlay J, Abowd GD, Beale R. *Human-Computer Interaction*. Prentice Hall; 2004.
31. Beyer H, Holtzblatt K. *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann Publishers Inc.; 1998.
32. Bernard HR. *Social Research Methods: Qualitative and Quantitative Approaches*. Sage Publications Inc.; 2000.
33. McNee SM, Arnette B. Productivity as a metric for visual analytics: reflections on e-discovery. In: *Proc. of the Workshop Beyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV)*. ACM; 2008. .
34. Broy M. Challenges in automotive software engineering. In: *Proc. Software Engineering (ICSE)*. ACM; 2006. p. 33–42.
35. Grimm K. Software technology in an automotive company: major challenges. In: *Proc. Software Engineering (ICSE)*. IEEE; 2003. p. 498–503.
36. Heinecke H. Automotive system design-challenges and potential. In: *Proc. Design, Automation and Test (DATE)*. IEEE; 2005. p. 656–657.
37. Pretschner A, Broy M, Kruger IH, Stauner T. Software engineering for automotive systems: A roadmap. In: *International Conference on Software Engineering*. IEEE; 2007. p. 55–71.
38. Sedlmair M, Hintermaier W, Stocker K, Büring T, Butz A. A Dual-View Visualization of In-Car Communication Processes. In: *Proc. of the Conference on Information Visualization (IV)*; 2008. p. 157–162.
39. Sedlmair M, Kunze B, Hintermaier W, Butz A. User-centered Development of a Visual Exploration System for In-Car Communication. In: *Proc. of the Symposium on Smart Graphics (SG)*. Springer; 2009. p. 105–116.
40. Sedlmair M, Ruhland K, Hennecke F, Butz A, Bioletti S, O’Sullivan C. Towards the Big Picture: Enriching 3D Models with Information Visualisation and Vice Versa. In: *Proc. of the Symposium on Smart Graphics (SG)*. Springer; 2009. p. 27–39.
41. Sedlmair M, Bernhold C, Herrscher D, Boring S, Butz A. MostVis: An Interactive Visualization Supporting Automotive Engineers in MOST Catalog Exploration. In: *Proc. of the Conference on Information Visualization (IV)*; 2009. p. 173–182.
42. Sedlmair M, Isenberg P, Baur D, Mauerer M, Pigorsch C, Butz A. Cardiogram: Visual Analytics for Automotive Engineers. In: *Proc. of the Conference on Human Factors in Computing Systems (CHI)*. ACM; 2011. To appear.
43. Moggridge B. *Designing Interactions*. Cambridge, MA, USA: Massachusetts Institute of Technology; 2007.
44. Holtzblatt K, Jones S. Contextual inquiry: A participatory technique for system design. In: *Participatory Design: Principles and Practices*. Lawrence Erlbaum Associates; 1993. p. 177–210.
45. Tory M, Möller T. Evaluating visualizations: do expert reviews work? *IEEE Computer Graphics and Applications*. 2005;25(5):8–11.
46. Morgan DL. *Focus groups as qualitative research*. Sage Publications, Inc; 1997.
47. Lam H, Munzner T. Increasing the utility of quantitative empirical studies for meta-analysis. In: *Proc. of the Workshop Beyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV)*. ACM; 2008. .
48. Kensing F, Simonsen J, Bodker K. MUST: A method for participatory design. *Human-Computer Interaction*. 1998;13(2):167–198.
49. Nielsen J, Mack RL. *Usability Inspection Methods*. John Wiley & Sons, Inc.; 1994.
50. Chen C, Czerwinski MP. Empirical evaluation of information visualizations: an introduction. *International Journal on Human-Computer Studies*. 2000;53(5):631–635.
51. Munzner T, Barsky A, Williams M. Reflections on QuestVis: A Visualization System for an Environmental Sustainability Model. *UBC Computer Science Technical Report TR-2009-24*; 2009.
52. Johnson C, Moorhead R, Munzner T, Pfister H, Rheingans P, Yoo TS. *NIH-NSF Visualization Research Challenges Report*. IEEE Computer Society; 2006.