User Studies in Visualization: A Reflection on Methods

Melanie Tory[1]

Abstract In this chapter I will reflect on many years of running user studies in visualization, examining my experience with how effectively different methodological approaches worked for different goals. I first introduce my own categorization of user studies based on their major goals (*understanding* versus *evaluation*, each with specific subcategories) and common methodological approaches (*quantitative experiment, qualitative observational study, inspection,* and *usability study*), providing examples of each combination. I then use examples from my own experience to reflect upon the strengths and weaknesses of each methodological approach.

1 Introduction

User studies are becoming standard practice in visualization research and design, as a way of both understanding users and evaluating visualization methods and tools. But the "user study" is not one single research method; rather it encompasses a large collection of empirical methods involving human participants. Designing a user study can be a complex and challenging activity involving many difficult choices. Which methods should I consider? How should I choose which method to use? When is each type of study appropriate? What metrics can best answer my question? Answering these questions is often very difficult.

Moreover, the adoption of user study methods within visualization has taken some time because visualization has some unique challenges compared to other fields such as Human Computer Interaction (HCI). Many of these were documented by Plaisant [18]. For example, many real world scenarios involve experts who will use the visualization tools on an extended basis; yet it is hard to get much time with experts, and training novices for a long time period is impractical. Additionally, tasks in visualization studies often need to be over-simplified in order to make them measurable. There is also a trade-off between a desire to build generalizable tools and the need to make sure tools work in specific domains.

The purpose of this chapter is to share lessons learned from running user studies in visualization, and to provide a structure to help designers and researchers navigate the rather large space of methods involving human participants. For readers who are new to the field, the chapter may be used as an overview of empirical study approaches and when to use them, with references to follow for more detail on specific methods. For those who have

[1] Melanie Tory
Department of Computer Science, University of Victoria, 3800 Finnerty Road, Victoria, BC, Canada, V8P 5C2, e-mail: mtory@cs.uvic.ca.

done some work in the area, the chapter provides a conceptual structure to explain the design space of user studies, and some insight on the state of user study methods in visualization and where to go from here.

2 User Study Goals and Types

User studies have been categorized in numerous different ways. In this section, I focus on categorizing studies based on the researcher's primary goals.

Some researchers have classified user studies in visualization by doing surveys of research papers. For example, Komlodi et al. [11] identified four major types: *controlled experiments comparing design elements, usability evaluation of a tool, controlled experiments comparing two or more tools,* and *case studies of tools in realistic settings*. In a more recent review, Lam et al. [12] characterized seven different empirical study scenarios, divided at the highest level into *understanding data analysis* and *understanding visualizations*. Nearly all of these categories are captured within my own description of study types.

Munzner's [16] nested model of the visualization design and validation process identifies four nested layers of design: *domain problem characterization, data / operation abstraction design, encoding / interaction technique design,* and *algorithm design*. While her model does not directly focus on characterizing user study techniques, it makes the important points that evaluation needs to occur within each layer, and the evaluation method chosen needs to match the design contribution of the layer. As an extreme example, evaluating the speed of an algorithm does not verify that the visualization algorithm solves a specific domain problem. The high-level user study goals that I present below have some similarity to the goals of Munzner's design layers.

2.1 Types of Visualization Studies Based on High Level Goals

Over the last several years within the field of visualization the term *evaluation* has become synonymous with *user study*. This is rather unfortunate because there are other reasons to run users studies besides evaluating visualization tools and techniques. In particular, many user studies are designed to better understand human behaviour, in order to provide insight to design. Often these studies take place before a visualization system or technique has even been developed.

I propose a simplified categorization of user study types based loosely on the prior work described at the beginning of section 2, but mostly based on classifying past research studies that I am familiar with. I divide user study goals into two high-level categories, *understanding* and *evaluation*, each with two lower-level categories, as shown in Table 1.

From a practical perspective, a user study designer can use these categories as the first step of study design. Choosing one of the study types can help the designer to articulate their goal, and can narrow down the choice of empirical methods (see section 3 for how the study types relate to empirical methods).

Table 1: Study types based on the high level goals of the researcher. Study types are classified at the highest level into those with understanding-based goals and those with evaluation-based goals.

Study goal		**Description**
Understanding	Understand perceptual and cognitive principles	Understand human perceptual or cognitive characteristics, often by measuring performance at abstract tasks.
	Understand context	Understand the context in which a visualization or future visualization will be used, including user characteristics, tasks, environment, social context, work practices, communication practices, etc.
Evaluation	Compare visualization techniques, tools, or interaction techniques	Directly compare two or more approaches to identify strengths and weaknesses of each or to validate a hypothesized improvement over a baseline design.
	Evaluate one visualization technique, tool, or interaction technique	Identify strengths, weaknesses, and/or limitations of one single approach.

Understanding Perceptual and Cognitive Principles

In this sort of study, the researcher wishes to learn something about human perception or cognition in order to inform design. Typically, the study employs similar methods to applied psychology studies, and aims to provide insights into human behaviour that can be used to formulate generalizable guidelines for visualization design. For example, Healey and Enns [6] described a series of experiments to better understand how humans perceive textural elements such as height, density, regularity and colour. They used the results of these experiments to inform the design of geographic representations of multivariate data.

Understanding Context of Use

In a context of use study, the researcher wishes to better understand the circumstances surrounding the use (or potential use) of a visualization tool, often for a specific domain. This could involve understanding the domain problem, user characteristics, tasks, environment, social context, work practices, collaborative or communication practices, analytical reasoning processes, etc. Typically a study would not investigate all of these, but would focus on the subset that is most relevant to the problem at hand. Most often, studies to understand context are done before initiating design, in order to inform the design process; such studies have been termed *pre-design studies* by Isenberg et al. [8]. However, this does not mean that visualization tools are never used in the studies. For example, people might already make use of visualization tools in their current work practices, and a study might assess how well the existing tools support the users' tasks, attempting to identify ways in which the tools might be modified or enhanced. Kang and Stasko [9] provide a recent example of a study to understand context. Through a longitudinal field study, they observed the work process of intelligence analysts to inform the design of analytics tools. Similarly, Sedlmair et al. [19] conducted extensive studies of automotive engineering prior to designing their visualization tool to support current work practices.

Compare Visualization Techniques, Tools, or Interaction Techniques

In a comparison study, the researcher directly compares two or more visualization techniques, complete visualization tools, or interaction techniques. Here *visualization technique* refers to a visual encoding method, *interaction technique* refers to a software and/or hardware interface for the user to interact with a visualization technique, and *tool* refers to a complete visualization system that would typically include multiple visualization and interaction techniques. The goal of a comparison study may be either to compare a prototype idea to a baseline 'best-of-breed' approach, to validate that the new technique is better, or to compare two or more competing ideas to identify their strengths and weaknesses. An example of technique comparison is Li and North [14]. They compared dynamic query sliders to brushing histograms (implemented in otherwise identical systems) for interacting with geographic visualizations. In contrast, Kobsa [10] conducted a study comparing complete systems for tree visualization.

Evaluate a Visualization Technique, Tool, or Interaction Technique

In a single approach evaluation study, the researcher typically wishes to assess the strengths, weaknesses, and / or limitations of a particular visualization technique, tool, or interaction technique, without making direct comparison to a competing method. This type of study might be used when a designer has developed an approach for a specific problem and wants to identify ways to improve the design, or simply needs to validate that the approach does indeed support users to accomplish their tasks. It also might be used when the competing approaches are so primitive that comparison against them would be pointless. For example, in one study we wanted to evaluate a visualization-based photo browser designed for construction managers [27]. We chose to evaluate only the prototype interface rather than compare to a baseline system because the default interface used to accomplish the photo management tasks (a standard computer folder system) was clearly inadequate.

2.2 Specific Research Questions

Identifying the high-level study goal (section 2.1) is only a first step in user study design. The next step is to identify a more specific set of research questions or objectives. These research questions will similarly help in identifying research methods and metrics. Example research questions for each of the four major study types are given in Table 2.

A major challenge with defining research questions is that often the first iteration of a research question is much too vague, and therefore cannot help the researcher in narrowing down the methods [4]. For example, consider the research question, "Do multiple tag clouds improve information finding as compared to a single tag cloud?". This only makes sense if we already know answers to a bunch of additional context details, like who (?) will be using the tag clouds, for what task (?), in what domain (?), what exactly does information finding mean (?), and how will we define when the information finding has been successful (?). Precisely defining the research question is an important first step to understanding what techniques and metrics will be appropriate. Note that the example questions in Table 2 are part way there, but each one still needs to have some additional context defined.

Table 2: Example research question for each major study type.

Understand perceptual and cognitive principles	Which type of motion (blink, shake, or expand) attracts attention in one's peripheral vision most quickly, assuming equal motion frequencies?
Understand context	What data analysis challenges do teams of intelligence analysts encounter during in-person team meetings?
Compare visualization techniques, tools, or interaction techniques	Can users navigate horizontal tree layouts more quickly with a fisheye lens or an overview + detail display?
Evaluate one visualization technique, tool, or interaction technique	Do side-by-side treemaps enable people to answer tree comparison questions? What design improvements are necessary to make these comparison questions easier?

For studies with an understanding goal, research questions need to focus the study in order to make its scope manageable. For instance, in a perceptual study, the designer will typically choose a perceptual attribute (e.g. visual attention to motion in one's peripheral vision) and a metric (e.g. performance, specifically time to notice a motion event). The designer will then need to specify the conditions under which the phenomenon will be studied (e.g. young adults with normal vision, conducting a reading task under normal lighting conditions with no other distractions), which will help to identify how other factors will be controlled. For a context study, the designer would typically narrow down a focus area (e.g. understanding task requirements for a specific visualization problem, with a specific user population).

For studies with an evaluation goal, the research questions will similarly need to narrow down the context of the evaluation, including the precise users and tasks. Often one of the most difficult choices to make is what specific outcomes are most important to evaluate. Specific outcomes that might be considered are:

- Performance (efficiency, errors)
- Usability
- Learnability
- User experience / preference
- Utility of the feature set (e.g. which features are useful, not useful, missing, or require improvement)
- Support for insight generation
- Support for communication
- Support for learning
- Effect on the analytical reasoning process
- Effect on the collaboration process (including attributes such as efficiency / effectiveness / ease of group work, support information sharing and awareness, which in turn need to be precisely defined)

Typical studies focus on one or a small number of the above outcomes. Comparative evaluations will ask which technique / tool / interaction technique has the best result for the specific outcome and why. Single approach evaluations will typically ask whether the approach meets a threshold level of the outcome, or what the strengths and weaknesses of the approach are with respect to the outcome.

Early in a project, the research questions are often *exploratory* in nature. These types of questions help us to better understand the problem, and identify factors that are relevant. For example, exploratory questions might be worded as, “What factors are important for X?”, “What is X like?”, “How often does X happen?”, or “What are the conditions under which X is needed?”. *Understanding context* goals are nearly always exploratory in nature. At later stages of a project, we may ask *confirmatory* questions because we already have some specific hypotheses in mind. For example, we might ask, “Does Y cause X?” or “How much does X improve performance at task Y?” For a more detailed description of research question types and additional examples, see Easterbrook et al. [4].

3 Empirical Approaches and Methods

User studies can be categorized according to the general methodological approach taken by the researcher, as well as by the specific data collection methods used.

3.1 High Level Approaches

I classify empirical user study approaches into four major categories, as shown in Table 3. Note that this set of approaches is not meant to be exhaustive (e.g., it is possible to have a qualitative experiment or a quantitative observational study), but instead is meant to capture the most common methods used currently in visualization. Numerous authors in other fields have described research methods involving human subjects in much greater depth. See McGrath [15], Lazar et al. [13], Creswell [3], or Easterbrook et al. [4] for a more comprehensive introduction and to gain an understanding of the philosophical differences among research traditions. For example, McGrath [15] provides a particularly useful taxonomy of general research approaches and the tradeoffs among them in terms of realism, generalizability, and precision. In contrast, my aim in this section is to focus specifically on visualization, and provide a small number of categories that describe the most common methods currently used in practice.

Table 3: Common empirical research approaches in visualization.

Quantitative Experiment	Makes a direct comparison between two or more controlled conditions and measures a quantitative difference between them. Results are analyzed using statistical hypothesis testing. Emphasis is on generic research results that can inform design of a whole class of visualization systems.
Qualitative Observational Study	Answers exploratory questions using mainly qualitative data gathered through techniques such as in-person or video observations, interviews, or journals. Groups may be compared but the partitioning into groups may be self-selected.
Inspection	A small number of experts inspect visualization tool(s), interface(s) or technique(s) using a pre-defined protocol and provide a report of their findings. Example inspection techniques include cognitive walkthroughs [17], heuristic evaluations [17], and abstract task evaluation [2].
Usability Study	Users complete tasks with a visualization tool, technique, or interaction method to assess whether it meets specified criteria. Normally used to assess whether a tool / technique / interaction method meets specific desired usability outcomes such as those in 2.2. Emphasis is on generating design improvements for a specific tool rather than generalizable design insights.

Combinations of the above approaches are also possible. For example, a quantitative experiment might also include the collection of qualitative data to help explain the results. As another example, in [22] we conducted an experiment with primarily qualitative observations, but also analyzed some quantitative differences between conditions.

3.2 Specific Data Collection Methods

Each of the study approaches will require data collection. A variety of data collection methods can be used, and these often are applicable across multiple approaches. Specific data collection methods used in visualization include:

- Direct observation
- Performance measurements
- Questionnaires
- Interviews
- Eyetracking
- Journaling
- Log analysis

For more detail on Eyetracking, see the Eyetracking chapter in this book. For other data collection methods, see Lazar et al. [13].

4. Relationship between Study Goals and Approaches

Table 4 illustrates the relationship between study goals and common research approaches, giving a short description and examples for each combination. The table reveals that all of the approaches can be used for multiple goals, and all of the goals can be achieved through more than one approach. However, not all combinations make sense. For example, Inspection is not possible unless there is a prototype system to evaluate, so it is not useful for the Understanding goals. Similarly, it would be difficult to understand the rich and complex nature of system use (Understanding Context) through a quantitative experiment since the approach is not exploratory. The text following the table describes advantages and challenges associated with each approach, with examples from my own experience.

Table 4 – Relationship between research approaches and study goals. Text in each cell describes a typical study of that type, with citations to examples. X's mark combinations that are unlikely to make sense.

	Understanding Goals		**Evaluation Goals**	
	Understand Perceptual & Cognitive Principles	**Understand Context**	**Compare Visualization Techniques, Tools, or Interaction Techniques**	**Evaluate a Visualization Technique, Tool, or Interaction Technique**
Quantitative Experiment	Typically a lab experiment with quantitative measures. Emphasis on generic insight into human attributes such as perception of different visual encodings. [6,24,25]	X	Lab or field experiment to compare two or more methods quantitatively (typically performance). Emphasis on results relevant to a specific domain or visualization type. [7, 10, 14]	X
Qualitative Observational Study	Lab or field observations to understand cognitive processes in data analysis. Emphasis on generic understanding rather than domain-specific design knowledge. [5]	Observations of naturalistic practices, often in field settings, and interviews, especially with domain experts. [9, 19, 26]	Compare methods in a lab study or field deployment, with a focus on qualitative outcomes. [22]	Observe use of a deployed system, ideally over a long time frame. [19, 20,21]
Inspection	X	X	Comparative evaluation by experts using a usability inspection approach. [23]	Single system evaluation by experts using a usability inspection approach. [28]
Usability Study	X	Evaluate usability of existing systems to establish requirements for a new system.	Usability study comparing multiple prototypes. Emphasis on identifying design improvements.	Prospective users conduct tasks with a prototype system. Emphasis on design improvements for that tool. [27]

4.1 Quantitative Experiments

I have conducted quantitative experiments for both understanding perceptual principles (e.g. to understand perceptual tasks with scatterplot and landscape visualizations of multidimensional data [24,25]) and for evaluating visualization techniques (e.g. new approaches to schedule visualization [7]). An advantage of quantitative studies is that they can provide convincing statistical evidence that one technique outperforms another. The downside is that they provide only a pinhole view since the study conditions must be strictly controlled. This typically means that the results are only reliable for the specific task and user population tested. As a result, a large collection of studies on the same topic is necessary in order to gain any holistic understanding. For example, our studies of landscape and scatterplot spatializations of multidimensional data sets revealed that scatterplots often outperformed landscape displays. However, we must note that our understanding of this topic is limited to two specific abstract tasks (memory of specific views, and finding regions within a specified value range) with an undergraduate student population. While we anticipate that these results are likely to be more generalizable than this, we cannot be certain that scatterplot views will outperform landscapes in all cases.

This example also brings up another important challenge: choosing an appropriate benchmark task. Because of the number of repeat trials required for statistical analysis, it is often impossible to test a large number of tasks within one study, so the set of tasks must be selected carefully. It can also be a challenge to choose tasks that are both ecologically valid (i.e. match what people do in the real world) but also easily measurable (i.e. they have a correct answer that can be validated, and ideally the measurements have some intermediate levels of granularity). Time is often a better primary metric than accuracy since it is measured on a continuous scale rather than a binary (correct / incorrect) one. However, even when time is the main metric, a clearly defined task is still necessary so that it is clear when a trial should end and the timer should stop. Abstraction versus domain specificity is another important factor to consider. In our landscape / scatterplot studies we chose tasks that were abstracted away from any particular domain, in order to increase generalizability of the results. This is both a benefit and a drawback: it is more *likely* that the results apply widely, but not *certain* that they apply in particular domain such as document analysis for intelligence analytics. For further reading, Lazar et al. [13] provide a good introduction to designing and analyzing a quantitative experiment.

4.2 Qualitative Observational Studies

Qualitative observational studies can be used for any of the goal types, as illustrated in Table 4. For example, we conducted a qualitative field study to understand the current work practices of interdisciplinary building design teams [26] (understanding context). In the laboratory, we used qualitative analysis to understand the process of visualization construction by novices [5] (understanding cognitive principles). In these projects, the research questions were exploratory (i.e. we did not know in advance which factors might be important for design). Qualitative methods allowed us to gain a holistic understanding of current practices. Qualitative analysis can also be useful in evaluation, to learn the complexities of when and why a visualization technique works / does not work.

Qualitative methods are very well established in the social sciences. The most common methods in Visualization are Grounded Theory (a general approach that involves building a theory to explain behaviour by observing undisturbed real world activities) and Content Analysis (a more specific method for analyzing the content of interviews, questionnaire responses, video observations, blog comments, etc.). These methods typically involve iteratively *coding* the material (i.e. assigning named categories to statements or observed events) until a cohesive description of the results emerges. Codes may be either based on existing knowledge and theory (fixed coding), or may emerge during the analysis (open coding). See Lazar et al. [13] or Creswell [3] for a more detailed introduction to qualitative methods and details on the coding process. In addition, Isenberg et al. [8] provide a visualization-specific perspective and Shneiderman and Plaisant [21] describe a specific approach for longitudinal evaluation of deployed visualization tools.

Despite their benefits, qualitative approaches do not come without costs. Because they are more susceptible to researcher bias than quantitative experiments, it is important to follow a structured and established method. Collecting the data can be time consuming, especially in a field study, because you cannot predict when interesting things will happen. Therefore you may have to observe many hours of repeated or irrelevant activities in order to capture the interesting and relevant ones. Analyzing the data is even more time consuming and it is not obvious at the outset what findings will be interesting and novel. For example it took me several months of analysis before the data from the building design field study began to make any sense, and many of the early "findings" turned out to have been already reported by other researchers in different domains. Therefore, it took several iterations to identify findings that were novel and interesting to the research community, and to structure those findings in a cohesive way. New researchers often feel lost in the analysis process; there is typically a stage where it feels as though you are drowning in data with no possible way of making sense of it all. If you find yourself in this state, do not despair! Persistence pays off.

4.3 Inspection

In HCI, usability inspection is generally thought of as a "discount" method of usability evaluation since asking a small number of experts (typically around five) to review a prototype typically takes less time than running a study with end users. Inspection techniques have been used successfully in visualization and are typically able to identify some strengths, problems, and limitations for a specific design. However, the quality of the results depends on several factors. Zuk et al. [28] reported that the findings of heuristic evaluation (one type of usability inspection, where the interface is evaluated with respect to several pre-defined heuristics) depended highly on the heuristics used and the types of evaluators that were chosen. In particular, they suggest using visualization-specific heuristics rather than just usability heuristics, and including experts from usability, visualization, and the application domain. We came to similar conclusions in our own previous research [23].

In general I have found that inspection techniques provide a quick way to identify likely problems with a visualization design. Conducting an inspection typically identifies many issues that were already known (e.g. missing functionality that was left out of a research prototype in the interest of time), but also nearly always reveals some problems that were not anticipated. However, the problems that are identified tend to be at a somewhat superficial and generic level. Unless some domain experts are included, and given realistic tasks for their domain, it is very hard to predict how well a system will perform in real use. Therefore this approach seems most useful for formative (early) evaluation of prototypes, where the emphasis is on finding problems and improving the design, rather than summative (late) evaluation, where the emphasis is on validation.

4.4 Usability Studies

A usability study is a study with end users that has a primary emphasis on identifying problems to directly improve the design of a next version, or to validate that users can accomplish certain tasks with a design. A usability study may use many of the same data collection approaches as quantitative experiments or qualitative studies, but is typically less formal and rigorous. For example, we took a usability study approach to evaluate the construction photo browser [27] described earlier because our primary objectives were to identify design problems and verify whether users could accomplish designated tasks with the interface.

Usability studies are generally quicker to set up and analyze as compared to more rigorous studies, yet provide better evidence of what will happen in real world use as compared to inspection. In visualization research, it is very difficult for results of a usability study to be a research contribution on their own due to the lack of rigor. For example, it may be hard to provide definitive evidence of which features are essential for each task. However, usability studies can provide insight to improve design, and evidence to validate design ideas. In non-research settings, iterative prototyping and evaluation via usability testing may be the most effective method to ensure a successful final design.

5 Lessons Learned

Over the years, I have learned numerous lessons about conducting user studies:

- **Know your goal, one goal per study**. Too many studies start with a very vague goal (e.g., "I need to evaluate my technique") or too many goals. Clearly defining the goal (and research question) at the start of user study planning helps to narrow down the research methods and metrics that will be useful. Similarly, trying to answer too many questions at once can make a study unwieldy. Often it is worth considering running a set of related studies, each one answering just one or two specific questions.
- **Choose tasks carefully**. One of the biggest challenges with study design in visualization is the choice of task. Most importantly, the task needs to be ecologically valid (i.e. relevant to real world use of the visualization). Particularly in controlled studies, it can be easy to get caught up in finding a task that is easily measurable, losing track of whether it represents a realistic scenario of use.
- **Pilot the instructions**. Particularly for open-ended tasks, the wording of the instructions can have a very strong impact on participant behaviour. For example, in a tabletop collaboration study [22], we observed that small wording changes or ordering of instructions impacted how much responsibility each participant took for their part of the work, how they prioritized their ultimate objectives, and how they chose to work together ver-

sus separately. Some bias of people's work practice is inevitable, but testing the instructions through pilot studies can help to minimize this effect.

- **Abstract away from the tool in quantitative experiments**. Quantitative experiments can compare either visualization or interface techniques, or complete visualization tools. Generally, those that compare complete tools have fewer generalizable outcomes, since it is impossible to isolate which features or design attributes made one tool better. It is often better to implement one system and then add and remove features of interest to understand their effects in isolation. Note, however, that *qualitative* evaluations of complete tools can provide some of this understanding.
- **User studies are not only for evaluation**. As illustrated by the examples in this chapter, user studies are not just for the purpose of evaluating visualizations. Understanding characteristics of users and the context surrounding the use of visualizations is just as important to achieving a successful design.
- **Use a variety of methods**. Qualitative and quantitative methods each have their place. For example, field studies are very realistic, but lack in precision, whereas lab experiments are very precise (in terms of measurements), but limited to specific (and often simplistic) tasks in a controlled environment and thus not realistic [15]. In my early work, I focused strictly on quantitative approaches, putting great value on the results being scientifically conclusive. However, I soon realized that my understanding was incomplete, because each study revealed only a small part of the picture. Qualitative methods can help to fill in the gaps and answer research questions with a more exploratory nature. Therefore, a good solution is to use either mixed-method designs or conduct multiple studies using complementary approaches.
- **Visualization is not that different**. Research methods for interacting with people have been around for a long time. Fields such as HCI, Psychology, and Social Sciences have documented these methods well, and many of the methods are very similar across disciplines. Rather than reinventing the wheel, we should aim to understand existing research approaches, adopt them for our purposes, and then tweak them where necessary. Visualization does have some specific challenges, especially the need to access expert users and examine how tools support very complex analysis tasks. However, these are not totally unique to visualization, and can be addressed by established methods such as longitudinal field studies.

6 Where to Go from Here?

Over the last decade, user studies have become the de facto standard for understanding users and evaluating the effectiveness of visualization tools and techniques. As a community, we have gradually adopted a greater variety of user study methods from HCI and other fields. By and large, these methods have been very effective and have not required substantial modifications to work for visualization. But the variety of adopted methods is not very large; for example, the vast majority of visualization user studies are still quantitative experiments. Use of qualitative and other approaches has only just begun. As a community, we should continue to explore a variety of research methods from other disciplines and bring them into our repertoire. One example of a method that might provide deeper understanding is ethnography. Among the few qualitative studies done in visualization, few could be considered a true ethnography, where the researcher is not only an observer of human activities but also an active participant who is immersed in the culture. Closest is probably Sedlmair et al.'s [19] research with automotive engineers, where the researchers worked closely with the target users over a three-year period. More work of this nature should be encouraged. More generally, the visualization community will benefit from a greater number of pre-design studies and long-term deployment studies that can provide a rich and detailed understanding of how visualizations can fit into work practices. In addition, online studies (e.g. through Mechanical Turk [1]) offer the potential to collect certain types of data faster, less expensively, and with more participants than laboratory studies. These should be considered as an alternative or complementary approach to laboratory studies with straightforward, controlled tasks.

References

1. Amazon Mechanical Turk [online]. Available: https://www.mturk.com/mturk/welcome [Accessed: June 15, 2012].
2. C. Ardito, P. Buono, M.F. Costabile, and R. Lanzilotti. Systematic Inspection of Information Visualization Systems. In *Proceedings of the 2006 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization*, pp. 1-4, 2006.
3. J.W. Creswell. *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson/Merrill Prentice Hall, 2007.
4. S. Easterbrook, J. Singer, M.A. Storey, and D. Damian. Selecting Empirical Methods for Software Engineering Research. In Shull et al., eds., *Guide to Advanced Empirical Software Engineering*, Springer-Verlag London Limited, 2008, Section III, pp. 285-311..
5. L. Grammel, M. Tory, and M.A. Storey. How information visualization novices construct visualizations. In *IEEE Transactions on Visualization and Computer Graphics*, volume 16, no. 6, pp. 943-952, Nov./Dec. 2010.
6. C.G. Healey and J.T. Enns. Large Datasets at a Glance: Combining Textures and Colors in Scientific Visualization. In *IEEE Transactions on Visualization and Computer Graphics*, volume 5, no. 2, pp. 145-167, April-June 1999.
7. D. Huang, M. Tory, S. Staub-French, and R. Pottinger. Visualization Techniques for Schedule Comparison. In *Computer Graphics Forum*, volume 28, no. 3, pp. 951-958, June 2009.
8. P. Isenberg, T. Zuk, C. Collins, and S. Carpendale. Grounded Evaluation of Information Visualizations. In *BELIV 2008: BEyond time and errors: novel evaLuation methods for Information Visualization*, 2008.
9. Y. Kang and J. Stasko. Characterizing the Intelligence Analysis Process: Informing Visual Analytics Design through a Longitudinal Field Study. In *Proceedings of IEEE Visual Analytics Science and Technology*, pp. 21-30, October 2011.
10. A. Kobsa. User Experiments with Tree Visualization Systems. In *IEEE Symposium on Information Visualization*, pp. 9-16, 2004.
11. A. Komlodi, A. Sears, and E. Stanziola. Information Visualization Evaluation Review, *ISRC Tech. Report, Dept. of Information Systems, UMBC-ISRC-2004-1*, 2004. Reported in [Plaisant].
12. H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical Studies in Information Visualization: Seven Scenarios. In *IEEE Transactions on Visualization and Computer Graphics*, 30 Nov. 2011, IEEE Computer Society Digital Library, <http://doi.ieeecomputersociety.org/10.1109/TVCG.2011.279>

13. J. Lazar, J.H. Feng, and H. Hochheiser. *Research Methods in Human-Computer Interaction*. United Kingdom: John Wiley & Sons Ltd., 2010. ISBN-10: 0470723378, ISBN-13: 978-0470723371.
14. Q. Li and C. North. Empirical Comparison of Dynamic Query Sliders and Brushing Histograms. In *IEEE Symposium on Information Visualization*, pp. 147-153, 2003.
15. J.E. McGrath. Methodology Matters: Doing Research in the behavioral and social sciences. in *Readings in Human-Computer Interaction: Toward the Year 2000*, R.M. Baecker, J. Grudin, and W.A.S. Buxton, eds. pp. 152-169, 1995.
16. T. Munzner. A Nested Model for Visualization Design and Validation. In *IEEE Transactions on Visualization and Computer Graphics*, volume 15, no. 6, pp. 921-928, November 2009.
17. J. Nielsen and R.L. Mack, eds. *Usability Inspection Methods*. New York: John Wiley & Sons, 1994, ISBN 0-471-01877-5.
18. C. Plaisant. The Challenge of Information Visualization Evaluation. In *Proceedings of the working conference on Advanced Visual* I*nterfaces*, pp. 109 – 116, 2004.
19. M. Sedlmair, P. Isenberg, D. Baur, M. Mauerer, C. Pigorsch, and A. Butz. Cardiogram: Visual Analytics for Automotive Engineers. In ACM Conference on Human Factors in Computing Systems, pp. 1727-1736, May 2011.
20. J. Seo and B. Shneiderman. Knowledge discovery in high dimensional data: Case studies and a user survey for the rank-by-feature framework. In *IEEE Transactions on Visualization and Computer Graphics*, volume 12, no. 3, pp. 311-322, May/June 2006.
21. B. Shneiderman and C. Plaisant. Strategies for Evaluating Information Visualization Tools: Multi-dimensional In-depth Long-term Case Studies. In *BELIV 2006: BEyond time and errors: novel evaLuation methods for Information Visualization*, 2006.
22. A. Tang, M. Tory, B. Po, P. Neumann, and S. Carpendale. Collaborative Coupling over Tabletop Displays. *In ACM CHI 2006*, pp. 1181-1190, Apr. 2006.
23. M. Tory and T. Möller. Evaluating Visualizations: Do Expert Reviews Work. In *IEEE Computer Graphics and Applications*, volume 25, no. 5, pp. 8-11, Sept./Oct. 2005.
24. M. Tory, D.W. Sprague, F. Wu, W.Y. So, and T. Munzner. Spatialization Design: Comparing Points and Landscapes. In *IEEE Transactions on Visualization and Computer Graphics*, volume 13, no. 6, pp. 1262-1269, Nov./Dec. 2007.
25. M. Tory, C. Swindells, and R. Dreezer. Comparing Dot and Landscape Spatializations for Visual Memory Differences. In *IEEE Transactions on Visualization and Computer Graphics*, volume 16, no. 6, pp. 1033-1040, Nov. / Dec. 2009.
26. M. Tory, S. Staub-French, B. Po, and F. Wu. Physical and digital artifact-mediated coordination in building design. In *Journal of Computer Supported Cooperative Work*, volume 17, no. 4, pp. 311-351, Aug. 2008.
27. F. Wu and M. Tory, "PhotoScope: Visualizing Spatiotemporal Coverage of Photos for Construction Management," ACM Conference on Human Factors in Computing Systems, pp. 1103-1112, 2009.
28. T. Zuk, L. Schlesier, P. Neumann, M.S. Hancock, and S. Carpendale. Heuristics for Information Visualization Evaluation. In *BELIV 2006: BEyond time and errors: novel evaLuation methods for Information Visualization*, 2006.