

Proposed Working Memory Measures for Evaluating Information Visualization Tools

Laura Matzen, Laura McNamara, Kerstan Cole, Alisa Bandlow, Courtney Dornburg, & Travis Bauer
Sandia National Laboratories
PO Box 5800
Albuquerque, NM 87185
505-844-1505

lematze@sandia.gov, lamcnam@sandia.gov, kscole@sandia.gov, abandlo@sandia.gov,
ccdornb@sandia.gov, tlbauer@sandia.gov

ABSTRACT

The current information visualization literature highlights design and evaluation processes that are highly variable and situation dependent, which raises at least two broad challenges. First, lack of a standardized evaluation criterion leads to costly re-designs for each task and specific user community. Second, this inadequacy in criterion validation raises significant uncertainty regarding visualization outputs and their related decisions, which may be especially troubling in high consequence environments like those of intelligence analysts. We seek ways to standardize the "apples and oranges" of the extant situation through tools based upon general principles of human cognition. Theoretically, information visualization tools enable the user to see information in a way that should attenuate the user's memory load and increase the user's task-available cognitive resources. By using general cognitive abilities, like available working memory resources, as a dependent measure, we propose standardized evaluative capabilities can be generalized across contexts, tasks, and user communities.

Keywords

Information visualization, Usability evaluation, Cognitive load, Working memory.

1. INTRODUCTION

The current information visualization literature highlights design and evaluation processes that are highly variable and situation dependent [1, 5, 6]. One important reason for this is that the field of information visualization is not focused on a particular domain or type of data. Instead, researchers have identified a range of knowledge tasks that lend themselves to visualization technology, including exposing uncertainty, concretizing relationships, formulating cause and effect, determining domain parameters, multivariate explanation, testing hypotheses, answering previously unforeseen questions, looking at data from different perspectives, and discovering patterns [4].

A wide variety of user communities can benefit from visualization technology that supports these knowledge tasks. In particular, visual analytics tools stand to benefit intelligence analysts. The amount of information that is theoretically available to intelligence analysts is a double-edge sword: although analysts are

privity to extensive, often proprietary datasets, it can be difficult to retrieve, assess, aggregate, and interpret the massive amounts of data that such databases contain, particularly considering the tight timelines that analysts frequently face. Ideally, visual analytics tools should make more information available, more easily and rapidly, than the current suite of search, notation, and storage tools that analysts typically have on their desktops. Perhaps most importantly, these tools must minimize cognitive load, to ensure that analysts' resources are fully available for making sense of complex information sources.

Many visual analytics tools have been aimed at the Intelligence Community (IC). However, formal validation of these tools is rare for a range of reasons. Ideally, validation would involve a controlled experiment to assess the impact of information visualization tools on intelligence products used for decisions in key real-world events, including the completeness, reliability, and accuracy of the assessments derived with and without visual analytics software. Such formal studies are difficult for a variety of reasons. For one thing, intelligence analysis is often task and analyst-specific, leading to difficulty in generalizing findings. Secondly, ground truth can be difficult to define. Collecting historical user effectiveness data requires accessing and understanding complex data sets that are often sensitive, proprietary, even classified. Additionally, some data sets lack ground truth by their very nature. For example, terrorism threat analysis accuracy may be indeterminate if the threat is never realized.

This is not to say that information visualization researchers have not tried to develop evaluative techniques to assess usability and usefulness. However, such *in-situ* usability studies and controlled experiments are "helpful but take significant time and resources," [6] and moreover, do not generalize across conditions and contexts, leading to costly re-designs for each project and specific user community.

A key reason for the cost of well-controlled usability studies is the lack of standardized metrics to evaluate different information visualization tools across users and domains. Given the demands intelligence analysis and other complex analytic tasks place on cognitive resources, we propose using cognitive workload measures as a standardized evaluation metric for information visualization tools. Measures of cognitive workload have been well-characterized by psychology and human factors research. We suggest that evaluation techniques focused on cognitive processing measures could provide metrics applicable across tools, tasks, and datasets. Evaluative principles that rely on cognitive processing, and not on domain-specific analysis

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

BELIV'10, April 10–11, 2010, Atlanta, GA, USA.

Copyright 2010 ACM 978-1-4503-0007-0...\$5.00.

outcomes, can lead to more cost-effective design principles for information visualization software in the IC and in other domains.

2. COGNITIVE LOAD EVALUATION

Cognitive load refers to the level of cognitive processing resources necessary to complete a task [2, 3, 4]. Cognitive load measures have been used to test performance in a variety of domains [3]. One method for measuring cognitive load is to assess *working memory*, or the brain's ability to acquire and maintain information for short time periods. Working memory is the "theoretical construct that has come to be used in cognitive psychology to refer to the system or mechanism underlying the maintenance of task-relevant information during the performance of a cognitive task" [7].

We propose using working memory metrics to evaluate information visualization tools by assessing what proportion of a user's working memory resources are consumed by using the tool as opposed to making sense of the data that the tool is designed to illuminate. We suggest that *effective information visualization tools should minimize the cognitive demands stemming from finding and manipulating raw data*. Effective tools should free the analyst's working memory resources for making sense of the information. In making this suggestion, we agree with Huang et al. (2008), that typical visualization performance measures compare performance differences in response time and accuracy, but fail to capture the requisite mental effort [4]. Thus, while performance measures might be equivocal across two visualizations, users may have to expend greater mental effort to compensate for a bad visualization. If an analyst devotes more cognitive resources to navigating a difficult interface or interpreting a confusing visualization, s/he will have fewer resources available for analyzing the data. This could decrease the likelihood of gaining insight and increase the chance of error.

3. PROPOSED METHODOLOGY FOR WORKING MEMORY MEASURES

To assess the feasibility of working memory metrics for evaluating cognitive load, we are conducting a series of dual-task experiments. Dual task paradigms, common in psychology and human factors research, require participants to complete two simultaneous tasks. They are either directed to focus on both tasks equally, or to devote effort to one primary task at the expense of the other secondary task. Participants can perform well on the secondary task only when they have excess cognitive resources not consumed by the primary task. Secondary task measures have been shown to be more sensitive measures of workload than primary tasks alone [8, 9]. In our evaluation studies, professional analysts will complete a series of simple information retrieval and summarization problems, similar to the kinds of tasks that intelligence analysts conduct in their daily work. Some of the analysts will use traditional read-and-search methods with raw text data, while others will solve the problems using a prototype visual analytics tool built on Sandia's Titan visualization framework (see www.vtk.org). The analysts' secondary task will be a well-validated working memory task called the Sternberg task [10]. The Sternberg task requires participants to remember target items (such as three random letters) and detect the targets in a string of distractor items. Accuracy and reaction times will be compared across conditions to assess the relative working memory burden imposed by using

the visualization tool. If the visualization tool is more effective than the traditional read-and-search approach, it should reduce the primary cognitive burden, and support better Sternberg task performance. Our primary hypothesis is that good performance on the secondary working memory task indicates the software does not impose a high cognitive burden. Thus, in a real-world analysis task, the analyst would have more resources available for high-level engagement with the data set. In other words, secondary working memory task performance should be indicative visualization tool effectiveness.

4. CONCLUSION

We believe evaluation approaches incorporating working memory measures will help researchers assess, at the very least, if different tool designs place unwanted burden on users' cognitive processing. Such evaluation approaches should also have reliability and validity across temporal, contextual, and user-related variances. We also believe such measures can be used to evaluate many types of visual analysis tools without re-developing application-specific metrics. An effective interface should minimize demand on cognitive resources, enabling users to focus on high-level analysis and sense-making. In future studies, we hope to demonstrate that this approach may allow: 1) effective comparisons across different users, data sets, and analysis tasks and 2) informative evaluations of visualization tools without the need to re-develop tool- and application-specific evaluation metrics and research designs.

5. REFERENCES

- [1]. Amar, R and Stasko, J., "Knowledge Precepts for Design and Evaluation of Information Visualizations," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 11, No. 4, July/August, pp. 432-442, 2005.
- [2]. Brunken, R., Plass, J.L., and Leutner, D. Assessment of Cognitive Load in Multimedia Learning with Dual-Task Methodology: Auditory Load and Modality Effects, *Instructional Science*, vol. 32, pp. 115-132, 2004.
- [3]. Brunken, R., Steinbacher, S., Plass, J.L., and Leutner, D. Assessment of Cognitive Load in Multimedial Learning Using Dual-Task Methodology, *Experimental Psychology*, vol. 49, pp. 109-119, 2002.
- [4]. Huang, W., Eades, P. and Hong, S. Beyond time and error: A cognitive approach to the evaluation of graph drawings, *Proc. 2008 Conference on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization*, 2008.
- [5]. Munzer, T. A. *Nested Model for Visualization Design and Evaluation*, *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, pp. 921-928, 2009.
- [6]. Plaisant, C., Fekete, J., and Grinstein, G. Promoting insight-based evaluation of visualizations: From contest to benchmark repository. *IEEE Transactions on Visualization and Computer Graphics* 14 (1), pp. 120-134, 2008.
- [7]. Shah, P. and Miyake, A. Models of working memory: An introduction. In A. Miyake and P. Shah (eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, Cambridge University Press, Cambridge, UK, pp 1-27, 1999.
- [8]. Gawron, V. (2008). Human performance, workload, and situational awareness handbook. Boca Raton, FL: Taylor & Francis.
- [9]. Meshkati, N., Hancock, P.A., & Rahimi, M. (1989). Techniques of mental workload assessment. In: J. Wilson (Ed.). *Evaluation of Human Work: Practical Ergonomics Methodology*. London: Taylor and Francis.
- [10]. Wierwille, W.W., and Eggemeier, F.T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors*, 35, 263-282.