

**Introducing additional domain-specific measures in evaluating visual analytic tools**  
Position Paper for the IEEE VAST “Metrics for the Evaluation of Visual Analytics”  
workshop

Mark A. Whiting, Carrie Varley, Jerome Haack  
Pacific Northwest National Laboratory  
9-12-2007

The new field of visual analytics needs to evolve measures for evaluation that enable technology and tool developers to better understand and accommodate needs of visual analytic tool users—information analysts. To decide which measures to develop and use, the visual analytic community must blend evaluation components from both the technology and the domain-specific perspectives.

From *Illuminating the Path, the R&D Agenda for Visual Analytics* (Thomas and Cook 2006) we see the following focus areas for the field:

- Analytical reasoning techniques that enable users to obtain deep insights that directly support assessment, planning and decision making
- Visual representations and interaction techniques that take advantage of the human eye’s broad bandwidth pathway into the mind to allow users to see, explore, and understand large amounts of information at once
- Data representations and transformation that convert all types of conflicting and dynamic data in ways that support visualization and analysis
- Techniques to support production, presentation, and dissemination of the results of an analysis to communicate information in the appropriate context to a variety of audiences.

With our community’s experience in visualization, data representation and management, and to a lesser extent, production and presentation, we have developed some evaluative measures for software featuring various characteristics of these specialties (Mackinlay 1986; Chen and Yu 2000; Plaisant 2004). Our current challenge, then, is to blend current and evolving technology evaluation measures with evaluative measures drawn directly from domains of interest. Here, we will present some of those measures drawn from intelligence analysis. We will discuss possible application of these to the IEEE Visual Analytics Science and Technology symposium’s contest, as the contest is explicitly modeled after (but abstracted from) intelligence analysis problems.

Valuable information to apply to our process may be drawn from the discussion of analytical lessons learned in (Cooper 2005). He draws from a list of questions developed by veteran analyst Charles Allen (Allen 2001), presented and discussed below. It is important to remember that in the VAST contest, entries are provided by tool builders, and in most cases, not professional analysts. The reasons motivating Allen’s questions may not be intuitive to a development team not working closely with analysts. However, an introduction to such issues that are critical to successful intelligence analysis is important so that tool builders are best prepared to address them.

**“What set of hypotheses was being considered? Was the set comprehensive, or was there bias in the selection of hypotheses? What *a priori* probability was attached to each hypothesis? Again, was there bias?”**

One anomaly that can be introduced by complex visual analytic tools is the bias introduced by the characteristics of the tool itself. For example, if a tool analyzes text documents, clusters them, and produces images of cluster maps, users might be biased by the predominant clusters being shown, attaching significance to these clusters by virtue of their coincidence, rather than assigning importance through a more analytic process. Similarly, the kind of data that a tool processes well may bias the user to assign importance to information revealed only in consideration of that data, as opposed to all of the data that must be considered.

Possible VAST contest measures:

*Did the tool facilitate exploration of a comprehensive set of hypotheses?*

*Did the tool use itself introduce a bias in the selection of hypotheses?*

*Did the tool help overcome user bias in the selection of hypotheses?*

**“Was there a good understanding about the observables that were expected to differentiate between the hypotheses? Was intelligence collection requested on the basis of these differentially diagnostic observables?”**

Differentiators in hypotheses assessment would seem to be key features for visual analytic tools to highlight and help provide user focus.

Possible VAST contest measures:

*When multiple hypotheses were considered, did the tool facilitate identification of differentiators? Did the analyst have the ability to track and manage differentiators and intelligence collection needs in the tool?*

**“Were all of the available data considered? How were the data weighted? What degree of credibility was accorded the sources?”**

Again, tools that cannot process heterogeneous data may bias users toward the particular pieces of information revealed in limited data sources. Data weighting may erroneously occur as a result of visual prominence in a display.

Possible VAST contest measures:

*What kinds and amounts of data can be considered by the tool? Does the tool create a bias for certain kinds of data?*

*Is there data credibility assessment or tracking in the tool?*

**“Was the possibility of deception considered and accounted for?”**

Deception can occur in several fashions (Bell and Whaley 1991). Identification of deceptive patterns or features should be incorporated into visual analytic tools, although approaches for this are a matter of current research.

Possible VAST contest measures:

*How are deceptive data identified, assessed, and managed in the tool?*

**“Was the analytic process logically correct? Was the confidence in rendered judgments correctly estimated? If so, and if the confidence was low, was additional collection requested?”**

Analysis has an intrinsic basis in evidence. The presentation of evidence should strongly support judgments, and be available for people reviewing the product to follow the flow of logic and understand suppositions and conclusions. Ideally, a tool should identify gaps in knowledge and gaps where there is little information, both of which result in collection requirements.

Possible VAST contest measures:

*Are formal (or non-formal, specific) analytic processes supported in the tool?*

*Is confidence assessed and managed in the tool?*

*Are information gaps identified by the tool?*

*Are collection requests supported and managed in the tool?*

**“Were the judgments presented in a timely and adequate manner?”**

This question needs to be answered in the context of the task situation. In the VAST contest, hard deadlines are established, and the tasks have been structured so that time to complete is not a critical concern for users (other than just getting the results submitted on time). Future contests may have different timing situations. This would then involve an assessment of how well the tool supported analyses under various time constraints.

Possible VAST contest measures:

*Does the tool perform well in short-duration assignments? Long-duration? What features are well suited for each situation?*

Even though practically all of the measures above call for qualitative assessment, they take another step in assembling a set of evaluative measures of keen interest to the user community. The next challenge is determining an appropriate balance between the technology-based perspective and the domain-specific perspective for evaluating visual analytic tools.

## Bibliography

Allen, C. (2001, September 12, 2007). "Intelligence: Cult, Craft, or Business?" Program on Information Resources Policy, from [http://pirp.harvard.edu/pubs\\_pdf/allen/allen-i01-1.pdf](http://pirp.harvard.edu/pubs_pdf/allen/allen-i01-1.pdf).

Bell, J. and B. Whaley (1991). Cheating and Deception. New York, St. Martin's Press.

Chen, C. and E. Yu (2000). "Empirical Studies of Information Visualization: A Meta-Analysis." International Journal of Human-Computer Studies **53**(5): 851-866.

Cooper, J. (2005). Curing Analytic Pathologies: Pathways to Improved Intelligence Analysis. Washington, DC, Center for the Study of Intelligence.

Mackinlay, J. (1986). "Automating the Design of Graphical Presentations of Relational Information." ACM Transactions on Graphics **5**(2): 110-141.

Plaisant, C. (2004). The Challenge of Information Visualization Evaluation. Proceedings of the Working Conference on Advanced Interfaces, New York, ACM Press.

Thomas, J. J. and K. Cook, Eds. (2006). Illuminating the Path: a Visual Analytics Agenda. IEEE Computer Graphics and Applications.