Evaluating Visualization techniques and tools: what are the main issues?

Beatriz Sousa Santos
DETI/IEETA - Universidade de Aveiro
3810-193, Aveiro, Portugal
bss@ua.pt

ABSTRACT

This position paper briefly addresses some relevant issues that should be considered whenever trying to evaluate any visualization tool or technique, which are in the author's opinion: motivation, evaluation methods, test data, collected data and data analysis.

Categories and Subject Descriptors

H.5.2 [User Interfaces]: *Evaluation/methodology*

Keywords

Information Visualization, evaluation, usability

1. INTRODUCTION

There are several definitions of Visualization that have been coined along the last twenty years, all of them including words such as insight, understanding and visual [1][2], which implies that we are dealing with a "human-in-the-loop problem" and thus complex by nature. This is perhaps the reason behind the fact that, although there are a lot of information visualization techniques and many systems that use them to visualize large amounts of information, there are comparatively few studies on the evaluation of those techniques and systems. However, now that the field of Information Visualization has matured and is applied to solve real problems, it is really important to know if these techniques and systems actually work. Furthermore, there is a need to know under what circumstances they should be used, how they compare and what tasks they best serve.

While there is not yet a body of knowledge on information visualization evaluation, we can find in literature works explicitly using evaluation in the process of designing a visualization system, evaluating specific systems and visualization techniques, as well as comparing alternative visualizations. Moreover, there are also authors making an effort to develop more systematic approaches to evaluation in information visualization and this complex problem has been the subject of journal special issues, panels and workshops in known conferences (e.g.[3]). Nevertheless, however interesting and relevant all these efforts may be, those who want to evaluate visualization techniques and systems, still struggle with a lack of specific techniques and methodologies to conduct the evaluation. Thus, currently, it seems that a reasonable approach could be to adapt methods from other disciplines as Human-Computer Interaction (HCI) and Social Sciences taking into account the specificities of the techniques and tools one is trying to evaluate.

In his reflection about measuring insight, North [4] writes that if the purpose of visualization is insight, the purpose of visualization evaluation is to determine to what degree visualizations achieve this purpose. If this claim is true, then evaluating visualizations should seek to determine how well visualizations generate insight. This is already a defying problem as North shows, yet visualizations should not only provide insight to their users, but also do it in reasonable time and with reasonable satisfaction. In fact, visualization evaluation is a multidimensional problem, involving at least three dimensions involved in usability evaluation (effectiveness, efficacy and satisfaction). Moreover, we could add other dimensions as accuracy, repeatability and robustness that ultimately influence the previous ones, but can be evaluated in a less user-centered approach.

In order to tackle such a complex problem we should look for help in several scientific areas, the more obvious being the ones already mentioned, Human-Computer Interaction and some Social Sciences as Ethnography. However, other disciplines can conceivably give a contribute, as Graphic Design. In fact, graphic designers are trained to convey information through the visual system and use a body of knowledge that the author believes might be very useful at least in evaluating the visual aspects of a visualization (that includes also interactive aspects).

Evaluating visualization techniques and tools confronts researchers with many questions; below, some comments are made to issues that must be precisely defined in order to have an evaluation methodology.

2. IMPORTANT ISSUES

There are some important issues that anyone facing the need to somehow evaluate a visualization technique or tool might have to consider. The first issue to think about is the motivation of the prospective evaluation (for instance, are we evaluating to inform the design/development process?). Then, it is necessary to choose the methods to use, and probably the wiser option is to use a combination of several methods in order to tap on their advantages and overcome their shortcomings. Most will imply the use of test data that can be synthetic or real. All evaluation methods involve collecting data (qualitative and/or quantitative) that have to be analyzed, possibly using statistical methods.

2.1 Motivation

This stage is crucial, it plays an important part in the choice of the methods to use, which will vary according to the phase in the development lifecycle, the available resources and other constraints.

2.2 Evaluation methods

Several alternatives to evaluate visualizations exist, each having its applicability, advantages, and disadvantages. Some methods widely used in Human-Computer Interaction seem readily adaptable and have been used to evaluate information visualization tools and techniques (e.g. observation); others, can be more difficult to use, as heuristic evaluation, which involves a list of heuristics fine-tuned to the situation. The author has been using evaluation methods imported from Human-Computer Interaction such as observation techniques, controlled experiments, and questionnaires. However, these methods constrain user thinking (as asserted in [4]) and performance, which is specially undesirable if the evaluation is performed in order to inform the project/development of a tool or technique. In this case, more “open-ended” strategies should be used as a first approach to the evaluation. It may be useful to organize discussion sessions where the tool or technique is presented and participants are asked to freely use it, find problems, criticize any aspect, and give suggestions as how to improve it. The author has organized sessions with two types of participants, people having a profile similar to the target users and graphic designers, with interesting results, in elicitation of new ideas with the former type of participants and in criticizing visual aspects with the latter.

2.3 Test data

Basically, two fundamental choices can be made: synthetic or real data, nevertheless, several intermediate types of data can be used. Ultimately, the evaluation of a visualization technique should be performed with real data, but it is reasonable to begin by using fully specified and systematically controlled data structures embedded in synthetic data. The use of computer generated data is flexible and allows the detection of errors and inaccuracies of the visualization technique to be evaluated in a way much easier than using real data. In some applications it may require a lot of modeling and may be only approximated, however it is the only method that allows a complete knowledge of the “ground truth”. No matter what the used test data, these should be clearly defined in order to comply with the needs of reproducibility and verifiability inherent to any scientific endeavor. Moreover, there should be common test data sets widely available to the visualization community, as to allow the comparison among visualization techniques.

2.4 Collected data

All evaluation methods produce some type of output data. For instance in controlled experiments, they reflect the dependent variables.

It is important to consider the nature of the collected data and the level of representation (measuring scale) used. As to nature, data can be discrete or continuous. According to level of representation, data can be qualitative (categorical) or quantitative (numerical). Categorical data can be measured solely through nominal or ordinal scales, whereas quantitative data can also be measured using interval or ratio scales. For instance, questionnaires often produce categorical data (sometimes coded as numbers, e.g. from 1 to 7, that should not be treated as quantitative).

The type of data has a direct influence on the statistical methods to be used. Unfortunately, techniques that can only be used with the highest level of representation (e.g. average and standard deviation) are often misapplied to data that were actually collected using a measuring scale corresponding to a lower representation level (e.g. ordinal data).

2.5 Data analysis

The choice of analysis techniques has a great impact on the credibility of the obtained results. It must be adequate to the type of data (sample size, distribution, nature, measuring scale, etc). Often, the widely used and well known parametric statistical methods are not applicable.

The author considers Exploratory Data Analysis [5] an interesting first approach to data analysis, since it provides general information on the structural relations, showing the amplitudes, asymmetries, localizations, outliers, etc.; it also usually provides some clues to further analysis, namely on the statistical methods to be used to test the original hypothesis, or ideas on other hypothesis.

3. CONCLUSIONS

The problem of evaluating visualization tools and techniques has long intrigued the author, which strongly believes it is absolutely fundamental to develop effective and systematic methods to tackle this defying problem in order to allow Information Visualization to fulfill its promise and serve a wide community of users in diverse areas. Moreover, this problem urgently needs attention by the visualization community, and all the efforts toward tackling it are valid, even if they are at first too naïve and unsophisticated.

4. REFERENCES

[1] Card, S., Mackinlay, J., Shneiderman, B. (eds). 1999. Readings in Information Visualization: Using Vision to Think. Morgan Kaufman.

[2] Brodlie, K., Carpenter, L., Earnshaw, R., Gallop, J. Hubbold, R., Mumford, A., Osland, C. Quarendon, P. 1992. Scientific Visualization, Techniques and Applications. Springer Verlag.

[3] Chen, C., and Cherwinski M. Ed. 2000. Introduction to the Special Issue on Empirical evaluation of information visualizations. Int. J. Human–Computer Studies, vol. 53, no. 5, pp. 631-635.

[4] North, C. 2005. Toward Measuring Visualization Insight. IEEE Computer Graphics and Applications, vol. 11, no. 4, pp. 443-456.

[5] Hoaglin, D., Mosteller, F., and Tukey, J. 1983. Understanding Robust and Exploratory Data Analysis. John Wiley & Sons.