Towards a Foundation for Information Visualization Engineering

Lars Grammel[1] and Margaret-Anne Storey[2]
University of Victoria

ABSTRACT

Despite much progress, it remains challenging to develop new visualization systems, predict their qualities, and understand design trade-offs. We propose an empirical framework for Information Visualization (InfoVis) theories consisting of a context space, a visualization space, visualization metrics, and visualization principles. Using this framework, we identify 5 possible steps to advance InfoVis theory. While the underlying ideas we discuss will be familiar to readers, we hope that expressing them in a systematic framework will contribute to the discussion of InfoVis theory, identify challenges that still need to be addressed, and shed light on how to integrate different theories in InfoVis.

KEYWORDS: Information Visualization, InfoVis Engineering, InfoVis Science, Theory, Framework

1 INTRODUCTION

Despite progress in Information Visualization (InfoVis), it remains challenging to develop new visualization systems, predict their qualities, and understand the trade-offs between designs. Researchers have recognized the need for InfoVis theory to address these challenges [8, 9, 11, 12, 15]. Thomas and Cook, in particular, expressed the need to move from craftsmanship to engineering, and to develop a theory for visual representations and interaction techniques [12]. We believe that practitioners must be able to apply visualization principles grounded in systematic empirical work. To this end, we propose that InfoVis science and InfoVis engineering be considered distinct areas of InfoVis:

InfoVis Science *is the systematic gathering of knowledge about InfoVis and the organization of that knowledge into testable visualization principles.*

InfoVis Engineering *is the creative application of scientific visualization principles to design or develop InfoVis systems*.

While both InfoVis science and engineering require us to build visualizations and work with visualization principles, they differ in their goals. InfoVis science is concerned with creating, understanding and refining visualization principles. In InfoVis Science, specific visualizations are created to gather empirical data in user studies. In contrast, InfoVis engineering is concerned with creating visualizations that are used in real-world settings and address practical problems. It employs visualization principles to build such visualizations for a specified context. This distinction emphasizes why InfoVis theories are crucial: they are the extracted essence of InfoVis science knowledge. We believe that InfoVis engineering could be substantially improved if theories were more carefully articulated and validated.

[1] email: lars.grammel@gmail.com
[2] email: mstorey@uvic.ca

2 A FRAMEWORK FOR EMPIRICAL INFOVIS THEORY

While many InfoVis theories and taxonomies already exist (e.g. [1, 3, 4, 6, 9, 14]), each theory focuses only on certain aspects of visualization design and it is challenging to apply them in combination to address practical design decisions. To help understand how different InfoVis theories are connected and which additional theories are required, we outline a framework into which InfoVis theories can be placed. The framework consists of four elements: context space, visualization space, visualization metrics, and visualization principles. While these four elements have been addressed individually in the literature, we believe that it is important to explicitly articulate them and understand how they are connected.

The **visualization space** consists of all visualizations that can possibly be constructed. The visualization space contains aspects such as graphical design and interaction. It addresses aspects at various levels of abstraction ranging from elemental perceptual properties through complex visualization forms and structures.

The **context space** is the set of all environmental factors outside the visualization itself, but which influence the outcomes of visualization metrics. It includes, for example, the setting in which a visualization takes place, values of data, output devices, task or presentation goals, and the user's background, knowledge and capabilities. For example, the setting could be casual, such as a visual arts exhibition, or formal, such as a work environment. Similarly, assumptions can often be made about the background of the users, e.g. that they are interested in visual arts. When evaluating visualizations, researchers may be able to explore and control the effects of context factors. However, in InfoVis engineering, visualizations will be selected or designed for a specific context.

A **visualization metric** is a measure of some property of a given visualization in a specific context. One example of a visualization metric is the measure of how aesthetically pleasing a Streamgraph of movie revenues is for visitors to a website [2]. Here, the visualization is the Streamgraph, the context contains the box office revenues of movies (*data*) and the potential visitors of the website (*users*), and the metric is the perceived aesthetic pleasure. Visualization metrics will often need to take both the context and visualization into account, because they are based on emergent properties of the human-computer system such as cognition [9].

Visualization principles predict and explain how context factors and visualization attributes affect metrics. They also capture the trade-offs that arise when optimizing visualization metrics. For example, a possible principle could encapsulate the fact that bar charts are more effective for part-to-whole analysis than pie charts [5]. Visualization principles should be testable so that we can increase our confidence in them by conducting empirical studies.

All four elements of the framework include concepts at different levels of abstraction. For example, in the context space a task can be modeled at task, sub-task, action and event level [6]. Similarly, metrics exist at a perceptual level, e.g. understanding how precisely color can be perceived [14], or at an application level, such as counting the number of insights gained using a concrete visualization [10]. By connecting the different levels of abstraction in the four framework elements, we might be able to integrate more detailed concepts and principles with more abstract ones. For example, Gotz and Zhou proposed a taxonomy that links the different levels of abstraction in tasks [6].

3 InfoVis Science

The framework outlined above provides us with a template in which existing InfoVis theories and research results can be placed to show what has already been covered, what needs to be addressed, and how the different pieces fit together. We now propose some steps to using this framework to further InfoVis research.

First, we need to understand and organize and agree upon relevant factors in the context space. For some factors (e.g., the presentation goal or task) initial taxonomies have been developed (e.g. [6]), while for other factors (e.g., user's visualization knowledge and capabilities) more work is required. Similarly, the visualization space needs to be categorized so it can be systematically explored. While numerous models have been proposed for this purpose (e.g. [1, 3, 4]), extending and integrating these models to ensure they are complete and appropriate requires further study. Working taxonomies of context factors and the visualization space are important because they enable us to use a shared and well-defined vocabulary in different studies and measurements.

Second, we need to find and standardize useful visualization metrics (e.g. measuring insight [10]) and develop appropriate scales of measurement for them. In addition, we need to refine our measurement techniques so that we ensure that we record both the context and visualization space, and that results can be reliably reproduced. Documenting context, visualization and measurement results using well-defined taxonomies facilitates making connections between different studies and enables sharing anonymized data online. This could in turn help with meta-analyses of multiple studies. We believe that such meta-analyses would allow us to make conclusions about broader ranges of context factor values, and would increase the confidence in the results gathered in the different studies.

Third, we need to conduct empirical studies to provide the evidence to develop, refine and refute visualization principles. While established methodologies such laboratory experiments and field studies have a well earned place in evaluating visualization systems, we believe that remote studies conducted over the internet [7] are also very promising and important. Such studies allow us to increase the number of participants by several orders of magnitude.

Fourth, based on results from the empirical studies, we need to find more abstract principles of information visualization that allow us to understand and predict the outcomes of visualization metrics. For elementary perceptual aspects, such principles have already been developed and tested [14]. However, we need principles that explain the effects of visual forms and structures [15]. This is a major challenge because predictive principles are hard to formulate and validate given the amount of potentially relevant context factors. A good starting point could be placing existing guidelines (e.g. [1, 5, 13]) into the framework as visualization principles, and testing them in empirical studies.

Finally, a major challenge is systematically integrating visualization principles across different levels of abstraction. For example, perceptual principles that explain how we perceive the length of lines are obviously related to principles explaining how well a bar chart works for part-to-whole analysis. However, the latter is affected by other perceptual and cognitive principles, and it is not clear how these principles interact. It is also possible that additional effects come into play on visual structures and forms [15] and as emergent properties of the human-computer system [9]. It is important to understand how several principles can be used together to predict trade-offs at the level of visualization, e.g. how a cumulative line in a sorted bar chart affects the task performance for part-to-whole analysis [5].

4 Conclusion

We outlined a framework that can be used to help us understand how different InfoVis theories are connected and to evaluate the need for additional theories. It consists of four components: context space, visualization space, visualization metrics and visualization principles. We described 5 potential steps that use this framework to further InfoViz research and should provide a better foundation for InfoVis engineering. We hope that this position statement contributes to the discussion of what InfoVis theory is, what questions still need to be addressed, and how to integrate different theories in InfoVis.

Acknowledgements

We would like to thank the members of the CHISEL and VisID groups, in particular Melanie Tory, Chris Bennett, Bradley Blashko and Patrick Gorman, for their valuable feedback and editing suggestions. This work was funded by an IBM CAS PhD Fellowship.

References

[1] Jacques Bertin. *Semiology of Graphics*. Univ. of Wisc. Press, 1983.

[2] L. Byron and M. Wattenberg. Stacked graphs–geometry & aesthetics. *IEEE Trans. on Visualization and Comuter. Graphics*, pp. 1245–1252, 2008.

[3] S.K. Card and J. Mackinlay. The structure of the information visualization design space. In *InfoVis' 97*, 1997.

[4] Ed H. Chi. A Taxonomy of Visualization Techniques using the Data State Reference Model. In *InfoVis '00*, 2000.

[5] Stephen Few. *Now you see it: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press, 2009.

[6] David Gotz and Michelle X. Zhou. Characterizing users' visual analytic activity for insight provenance. *Information Visualization*, 8(1):42–55, 2009.

[7] Jeffrey Heer and Michael Bostock. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *CHI '10*, pp. 203–212, New York, NY, USA, 2010. ACM.

[8] Chris Johnson et al. NIH-NSF Visualization Research Challenges Report, 2006.

[9] Z. Liu, N. Nersessian, and J. Stasko. Distributed cognition as a theoretical framework for information visualization. *IEEE Trans. on Visualization and Computer Graphics*, pp. 1173–1180, 2008.

[10] Chris North. Toward measuring visualization insight. *IEEE Comput. Graph. Appl.*, 26(3):6–9, 2006.

[11] Helen C. Purchase, Natalia Andrienko, T. J. Jankun-Kelly, and Matthew Ward. Theoretical foundations of information visualization, pp. 46–64, 2008.

[12] J.J. Thomas and K.A. Cook. Illuminating the path: The research and development agenda for visual analytics. *IEEE Computer Society*, 2005.

[13] Edward R. Tufte. *Envisioning Information*. Graphics Press, Cheshire, CT, 1990.

[14] Colin Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

[15] Caroline Ziemkiewicz and Robert Kosara. Beyond Bertin: Seeing the forest despite the trees. *Computer Graphics and Applications (CG&A), Visualization Viewpoints*, 30(5):6–10, 2010.