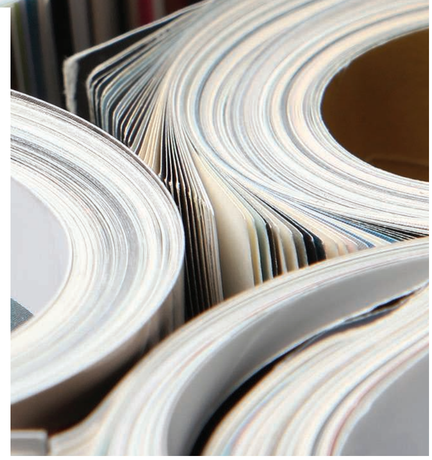
Advances in Visualization Recommender Systems



Klaus Mueller, Stony Brook University

This installment of Computer's series highlighting the work published in IEEE Computer Society journals comes from IEEE Transactions on Visualization and Computer Graphics.

As we take part in the revolution of big data, visualization has become an important mechanism for enabling human-in-the-loop data analytics and data-driven decision making. Since visual representations of data and information greatly appeal to the fast cognitive facilities of the human brain, they form a powerful interface to connect the human to the data and any advanced machine-learning algorithms that could be operating on them. This human–machine discourse is commonly referred to as *visual analytics*.

Data visualization begins with the data that are transformed into a graphical representation by a given visualization method. At this point, the rendered visualization is just an image, which the human observer's perceptual and cognitive facilities subsequently convert into insight. But not all visualizations succeed in this crucial latter task, and many do not even entice a human to engage, that is, look and spend the effort to study it. While the latter is not yet fully understood, it is related to aesthetics and graphic appeal.

Selecting the best visualization for a given data configuration and analytical goal is difficult. This might be surprising to some and is one of the reasons that work in this research area is so important. The visualization design begins with deciding what basic chart is best suited for the type (that is, categorical, numerical, temporal, string) and size (that is, number of columns and rows) of the data at hand. The design then continues with choosing the right set of colors, the appearance of marks, graphical detail, aspect ratio, chart orientation, and so on.

Digital Object Identifier 10.1109/MC.2019.2918513
Date of publication: 30 July 2019

EDITOR **RON VETTER**
University of North Carolina Wilmington; vetterr@uncw.edu

There are two principal factors that govern the graphical design of a data visualization and promote a human's understanding of the data and the information they represent, namely, *expressiveness* and *effectiveness*. A visualization is expressive when it is able to communicate all information in the data; it is effective when it does so better than an alternative visualization.

Since a mainstream user cannot be expected to possess the kind of expertise needed to design the most expressive and effective chart for his or her data, automating this design task has been an active area of research. In fact, several research systems on this topic had already emerged in the late 1980s. These were essentially rule-based systems informed by contemporary fundamental research on visual information encoding. While effective, rule-based systems have the inherent limitations that the coding of rules is a tedious process and executing a rule-based system can be prone to combinatorial explosion.

The featured spotlight article is "Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco" (*IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 438–448, 2019) by Dominik Moritz, Chenglong Wang, Greg L. Nelson, Halden Lin, Adam M. Smith, Bill Howe, and Jeffrey Heer. It addresses this problem in a unique and forward-looking manner by taking advantage of advanced machine learning and a set of empirical studies on visualization designs that recently became available. These studies produced many user-rated visualization designs and represent a considerable amount of visualization design knowledge and best practices that the authors use effectively. Another enabling technology is Vega-Lite, a high-level grammar-for-visualization design that was developed in the authors' research laboratory and described in a different set of articles. Vega-Lite maps data to properties of graphical marks, that is, points or bars, and an associated compiler then automatically produces visualization components including axes, legends, and scales.

The ensuing democratization of insight has the propensity to propel both science and society forward.

The Draco system, which is the subject of this article, uses answer-set programming in conjunction with a weight-learning method to represent and apply the visualization design knowledge gained through the aforementioned empirical studies. The design knowledge is encoded as a set of constraints used to design future visual representations. Draco uses the answer-set solver to evaluate the set of constraints when generating or suggesting a visualization. Some constraints can be soft, which enables tradeoffs for using a particular visual design instead of another. In such cases, these soft constraints are weighted, allowing the system to compose a list of the best candidates. The weights can be set manually or learned with a RankSVM classifier. Users can then input their data, and Draco recommends a list of visualizations, ranked by their model-based preference scores.

The system is also notable in that researchers can use Draco to encode findings they generate from their own (and other) empirical studies. These findings can then be captured as constraints to construct, augment, or update a new or existing visualization recommender and validator system.

These are undoubtedly exciting times for data visualization use and research. The emerging web-scale community databases of visualization designs are great enablers in the quest to expressively and effectively design visualization of data and bring it to the mainstream arena. The development of Draco, and other systems that may emerge, empowers everyone to generate compelling, yet truthful, visualizations of data. They can then share this information with others, enabling them to form their own opinion on the substance of the data and the information conveyed. The ensuing democratization of insight has the propensity to propel both science and society forward and lead to a better informed world as a whole.

KLAUS MUELLER is a professor with the Computer Science Department, Stony Brook University; a senior scientist with the Computational Science Initiative, Brookhaven National Laboratory; and the editor-in-chief of *IEEE Transactions on Visualization and Computer Graphics*. Contact him at mueller@cs.stonybrook.edu.