

# Shaping Data: Scaling Data Visualizations

Oliver Bieh-Zimmert

**Abstract**—Generating the maximum number of visual patterns by uncovering the entire space of possible visual designs remains a challenge within the construction process of information visualization. Users interact with different mindsets consisting of design, data analysis, application development, and hardware resource usage. Therefore, they desire a flexible and productive interface that keeps them clued into the design process without requiring knowledge of the underlying technical system. A general model was applied to a prototype to demonstrate the benefits and restrictions of this construction process and to contribute toward bringing different mindsets together.

**Type of presentation:** Research contribution – Visual analytics

**Key words:** visual analytics, visual scalability, modeling, big data, data flow

## INTRODUCTION

Data scientists, decision makers, information designers, and data-driven journalists face the same challenge of seeking meaningful patterns in large datasets [17][26][10]. Supporting this seeking process through an interactive visual analytical system can help these professionals of different mindsets uncover insights, especially from unknown datasets [15]. However, the number of visual encodings—and thus, the space of possible visualization designs—is extremely large [14]. The aforementioned users desire a simple and intuitive visual system that delivers meaningful representations quickly [14]. Currently, this requires a different skill set involving design, data analysis, application development, and hardware resource allocation. Specifically, technical and programming skills are necessary to exploit the full potential of visual mapping within the design process of data visualization. However, this can lead to a time-consuming process involving the development of custom visualizations, which is a deviation from the actual design process. In the worst case, only a subset or very highly aggregated datasets are used to shorten the creation process [2]. This harms the exploration process, meaning that “people spend substantially more time working with small datasets than massive ones [5].”

Work in this direction is rooted in the “visual programming” paradigm, which embodies access to the creation of unlimited visualization designs. A number of software systems, development frameworks, or libraries with a notion of the data-flow process have been implemented [1][4][7][13], but their simplification is still an open issue. The

goal of this paper tries to describe a blueprint that shifts from a programming-oriented approach to one that is more modeling-oriented in order to minimize the cognitive distance between creation and the resulting data visualization.

A closer look at the process of creating data visualizations and their influencing factors is necessary to answer the above research questions. The contribution puts up a “no developing, more modeling” position for discussion that redirects users of visual systems to focus on the design process through a generic interface without requiring to have programming knowledge.

## 1 CROSS-DOMAIN UNDERSTANDING

The design of visual representations stays within the spectrum of fully capable visual programming environments and pre-defined model sets of visualizations accounting for basic user needs and task requirements [22].

Cross-domain advances in Astronomy [18], Biology [23], Business [20], Epidemiology [3], Healthcare [6], Manufacturing [12], Security [21], and Software development [24] show the increasing individuality in understanding data and the use of information visualization. However, this individuality comes at the cost of productivity, resulting from a time-consuming process of programming. In general, the more flexible a visual system, the more time is required by a user to be productive with it. In contrast, the easier the system design, the less likely it is that a user will have the necessary control over system customization [11]. Constructing a visual structure within the range between flexibility and productivity require a formal

description encompassing the possible number of visual designs in preparation of a semantic, syntactic, and lexical model specification. This description follows the knowledge creation process (Fig. 1) to construct and apply a model for pattern discovery [27]. While flexibility is described as generating the maximum number of visual designs is productivity defined as the quick and understandable construction of data visualizations with minimal knowledge efforts [11]. For this purpose, a prototype will be developed and analyzed under the practical implications to fulfill the flexibility and productivity needs.

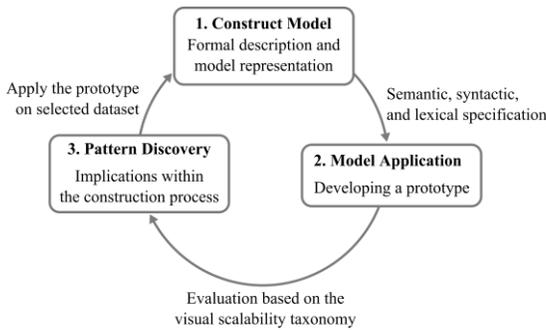


Fig. 1. Knowledge creation process

## 2 VISUAL ENCODING

Visual analytical systems aim to provide the capability of generating the maximum number of possible representations of an underlying dataset. If the total number of visual perspectives is limited (i.e., set of only predefined charts), the resulting number of patterns is restricted as well. This can affect the discovery of insights, decisions making, telling of data-driven stories, or finding responses to hypothetical questions.

### 2.1 Shaping Data Representations

In shaping, the creation of data representations is conducted by shaping data attributes to visual properties. Shaping is defined as the process of combining attributes (A), operations (O), and visual properties (V) based on a composite model representation for uncovering meaningful patterns from the underlying dataset.

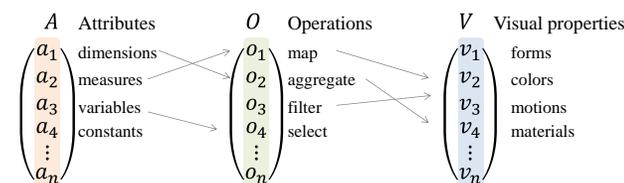


Fig. 1. Composition of attributes, operations, and visual properties

Attribute mapping, including their applied operations on visual properties, can be described through the following function:

$$\begin{aligned}
 d(x) \dots \text{data function} & & s(x) \dots \text{shape function} \\
 d: A \rightarrow O & & s: O \rightarrow V \\
 s \circ d: A \rightarrow V, x \mapsto (s \circ d)(x) := s(d(x)) & & (1)
 \end{aligned}$$

The function encompasses the available compositions of attributes, operations and visual properties to provide the entire set of possible visual encodings. This allows users to choose from these quantified set identifying patterns in the creation process via the interplay of human cognition and information visualization [19].

## 3 MODELING VISUAL REPRESENTATIONS

A model representation built based on an interactive user interface must be suited to a generalized system for creating data visualizations, as well as support intuitive usage with minimal knowledge. This demands a syntactic and semantic model that represents model elements and allows for their direct manipulation through rapid, incremental, and reversible operations so that users can learn by a layered or spiral approach [25]. This allows users to learn a set of model elements, which they can practice on and expand their knowledge based on direct feedback. "What the user needs to pay attention to is the components or behavior of the components that the user can see in an actual application environment, not the syntax of low-level computer-oriented language." [16]

The following model abstraction in Fig. 1 covers the shaping process of visual properties, pre-attentive attributes, and patterns. The model arranges selected visual properties within a group. In this way, a pre-attentive attribute includes a chain of visual properties that can be summarized in an overall pattern. This model can be applied to selected datasets and used operations.

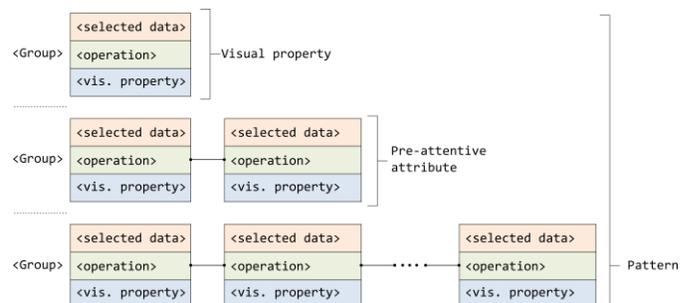


Fig. 1: Shaping process in design stage

The goal of this approach is to divide the construction process into atomic model elements. These model elements can be used flexibly to build individual visual representations. Furthermore, the shaped elements can connect abstract data with visual characteristics [9].

These benefits help bring together the different mindsets of visual design, data analysis, application development, and hardware resource use, and share their knowledge of constructing meaningful insights.

## REFERENCES

- [1] G. Abram, L. Treinish. An extended data-flow architecture for data analysis and visualization. *ACM SIGGRAPH Computer Graphics*, Vol. 28, Issue 2, pp. 17-21, 1995. DOI=<http://dx.doi.org/10.1145/204362.204366>.
- [2] O. Bieh-Zimmert, C. Felden. Shaping unlimited patterns: a vision of state-of-the-art visual scalability. *Proceedings of MEDES '14*, pp. 72-77, 2014. DOI=<http://dx.doi.org/10.1145/2668260.2668279>.
- [3] O. Bieh-Zimmert, C. Koschtial, C. Felden. Representing multidimensional cancer registry data. *i-Know '13, Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies*, pp. 35-38, 2013. DOI=<http://dx.doi.org/10.1145/2494188.2494222>.
- [4] M. Bostock, V. Ogievetsky, J. Heer. D<sup>3</sup> Data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 17, Issue 12, 2011. DOI=<http://dx.doi.org/10.1109/TVCG.2011.185>.
- [5] C. Chabot. Demystifying visual analytics. *IEEE Journal of Computer Graphics and Applications*, Vol. 29, Issue 2, pp. 84-87, 2009. DOI=<http://dx.doi.org/10.1109/MCG.2009.23>.
- [6] L. Chittaro, C. Combi, G. Trapasso. Data mining on temporal data: a visual approach and its clinical application to hemodialysis. *Journal of Visual Languages and Computing*, Vol. 14, Issue 6, pp. 591-620, 2003. DOI=<http://dx.doi.org/10.1016/j.jvlc.2003.06.003>.
- [7] D. S. Dyer. A Dataflow Toolkit for Visualization, *IEEE Computer Graphics and Applications*, Vol. 10, Issue 4, pp. 60-69, 1990.
- [8] 2002. DOI=<http://dx.doi.org/10.1109/INFVIS.2002.1173156>.
- [9] S. Few. *Now you see it: simple visualization techniques for quantitative analysis*. Analytics Press, Oakland CA, 2009.
- [10] J. Gray, L. Bounegru, L. Chambers. *The Data Journalism Handbook: How Journalists Can Use Data to Improve the News*. O'Reilly Media, Inc., Sebastopol, CA, USA, 2012.
- [11] A. G. Forbes, T. Höllerer, G. Legrady. Behaviorism: a framework for dynamic data visualization. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 16, Issue 6, pp. 1164-1171, 2010. DOI=<http://dx.doi.org/10.1109/TVCG.2010.126>.
- [12] M. Haage, K. Nilsson, On the scalability of visualization in manufacturing. In *Proceedings of ETFA '99, 7th IEEE International Conference on Emerging Technologies and Factory Automation*, Vol. 1, pp. 43-51, 1999. DOI=<http://dx.doi.org/10.1109/ETFA.1999.815337>.
- [13] P. Haeberli. ConMan: a visual programming language for interactive graphics. *Computer Graphics*, Vol. 22, Issue 4, pp. 103-111, 1988. DOI=<http://dx.doi.org/10.1145/378456.378494>.
- [14] J. Heer, M. Bostock, V. Ogievetsky. A tour through the visualization zoo. *Communications of the ACM*, Vol. 53, Issue 6, pp. 59-67, 2010. DOI=<http://dx.doi.org/10.1145/1743546.1743567>.
- [15] J. Heer, B. Shneiderman. Interactive dynamics for visual analysis. *Communications of the ACM*, Vol. 55, No. 4, pp. 45-54, 2012. DOI=<http://dx.doi.org/10.1145/2133806.2133821>.
- [16] M. Hirakawa, T. Ichikawa. Advances in visual programming. *Proceedings of the Second International Conference on Systems Integration*, pp. 538-543, 1992. DOI=<http://dx.doi.org/10.1109/ICSI.1992.217309>.
- [17] D. A. Keim, J. Schneidewind, H. Ziegler. Challenges in visual data analysis. *Proceedings of Information Visualization (IV '06)*, pp. 9-16, 2006. DOI=<http://dx.doi.org/10.1109/IV.2006.31>.
- [18] T. Luciani, R. Hachley, D. Q. Oliphant, B. A. Cherinka, G. E. Marai. Pixel-based overlays for navigating a galaxy of observations. *IEEE Symposium on Large Data Analysis and Visualization*, pp. 137-138, 2007. DOI=<http://dx.doi.org/10.1109/LDAV.2011.6092337>.
- [19] R. E. Patterson, L. M. Blaha, G. Grinstein, et. al. A human cognition framework for information visualization. *Computer and Graphics*, Vol. 42, pp. 42-58, 2014. DOI=<http://dx.doi.org/10.1016/j.cag.2014.03.002>.
- [20] K. B. Pratt, G. Tschapek. Visualizing concept drift. *KDD '03, Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 735-740, 2003. DOI=<http://dx.doi.org/10.1145/956750.956849>.
- [21] P. Ren, Y. Gao, Z. Li, Y. Chen, B. Watson. IDGraphs: intrusion detection and analysis using stream compositing, *IEEE Computer Graphics and Applications*, Vol. 26, Issue 2, pp. 28-39, 2006. DOI=<http://dx.doi.org/10.1109/MCG.2006.36>.
- [22] L. Ren, F. Tian, X. Zhang, L. Zhang. DaisyViz: A model-based user interface toolkit for interactive information visualization systems. *Journal of Visual Languages and Computing*, Vol. 21, Issue 4, pp. 209-229, 2010. DOI=<http://dx.doi.org/10.1016/j.jvlc.2010.05.003>.
- [23] H.-J. Schulz, A. M. Uhrmacher, H. Schuhmann. Visual analytics for stochastic simulation in cell biology, *i-KNOW '11 Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, pp. 1-8, 2011. DOI=<http://dx.doi.org/10.1145/2024288.2024345>.
- [24] M. Sensalire, P. Ogao, A. Telea. Classifying desirable features of software visualization tools for corrective maintenance. *SoftVis '08, Proceedings of the 4th ACM symposium on Software visualization*, pp. 87-90, 2008. DOI=<http://dx.doi.org/10.1145/1409720.1409734>.
- [25] B. Shneiderman. Direct manipulation: a step beyond programming languages. *Computer*, Vol. 16, Issue 8, pp. 57-69, 1983, DOI=<http://dx.doi.org/10.1109/MC.1983.1654471>.
- [26] R. Spence. *Information Visualization – Design for Interaction*. 2<sup>nd</sup> edition, Pearson Education Limited, Harlow, Essex, England, 2007.
- [27] L. Xiao, J. Gerth, P. Hanrahan. Enhancing visual analysis of network traffic using a knowledge representation. *IEEE Symposium on Visual Analytics Science and Technology*, pp. 107-114, 2006. DOI=<http://dx.doi.org/10.1109/VAST.2006.261436>.

## 4 CV OLIVER BIEH-ZIMMERT

Oliver Bieh-Zimmert is breaking the existing limitations of visual scalability on the horizon of seeking unlimited patterns in massive data sets. He consults enterprises and organizations in Data Visualization and Big Data Analytics with a cross-disciplinary background in information systems, data modeling, and interactive data representation. He is publishing articles and talks about the latest discoveries in shaping large data sets as doctoral student at the Freiberg University of Mining and Technology and was winner of the Data Science Summit 2013 in Redwood, California.

**E-Mail:** [oliver.bieh-zimmert@visual-telling.com](mailto:oliver.bieh-zimmert@visual-telling.com)

### 4.1 Awards and Talks

- First prize at the [Data Science Summit Redwood](#) visualization challenge in Redwood, California, 2013
- Blog about data visualization: [Visual-Telling](#)

### 4.2 Senior Data Science and Visualization Expert

- Since July 2011: Senior Data Scientist at the T-Systems Multimedia Solutions GmbH
- Since Oct. 2010: Ph. D. student at the University of Resources Freiberg with specialization on high scalable data visualizations
- Oct. 2009 – Jun. 2011: Research associate at the Center of Applied Research and Technology, Dresden
- Mai 2007 – May 2009: Research associate at the University of Applied Sciences Dresden.
- Sept. 2008 – Sept. 2009: Scholar of Entrepreneurship SmartSoft at the University of Technology Dresden

### 4.3 Relevant Publications

- [Shaping Unlimited Patterns: A Vision for State-of-the-Art Visual Scalability](#), MEDES '14 6th International Conference on Management of Emergent Digital EcoSystems, Buraidah, Saudi Arabia, 2014
- [Representing Multidimensional Cancer Registry Data](#), Proceedings of i-KNOW, 2013
- [Job of a Data Scientist](#), "Es fasziniert mich, welchen Einfluss Daten auf unseren Alltag haben.", Frankfurter Allgemeine Zeitung, 6th of March 2014, No. 55/10 D 3

