Understanding Data-Driven Visual Encodings through Deconstruction

John Thompson* John Stasko†

Georgia Institute of Technology

Abstract

Our ongoing work aims to better understand the relationship between visual marks and data encodings that comprise innovative data visualizations. By innovative, we refer to data visualizations that in part extend or deviate from conventional visualization techniques. Authors of these types of visualizations engage and inform their audience by incorporating novel data-bound compositions. In our work, we deconstruct a set of these visualizations into their elemental data-driven marks by employing a semi-structured coding process. This process serves as the initial step for defining a visual grammar that can act as the building blocks for creative data visualizations. The primary purpose of this model is to define the set of graphics, data-bindings, and visual styles that researchers may then incorporate into a data visualization authoring tool. Here, we present our analysis approach to deconstruct the set of visualizations and how we cataloged the coded data for dissemination among the visualization community. We then highlight interesting trends observed within the collection of visualizations analyzed thus far.

1 Introduction

In order to engage their audience, authors of creative or artistic data visualizations frequently go beyond what one finds in typical visualization charts and graphs. By presenting the observer with unique visuals, authors can intrigue and captivate their audience with interesting data. However, in order to create such unique visuals, an author must manually compose a scene graph using vector graphics tools or custom visualization code that provides the desired level of expressivity. Thus, the author's resulting design can operate outside of the visual grammars supported by currently available visualization authoring tools.

In order for a visual grammar to support the authoring of such creative data visualizations, it should at least support the diversity of data-bindings that are currently being employed by practitioners. The best-case scenario, if it exists, is for a visual grammar to abstractly account for yet to be conceived data-to-visual encodings.

We deconstructed a set of creative visualizations into the data elements and data attributes conveyed in the visualization. Furthermore, we categorized each visual mark and attribute driven by data; based on the transformation and encoding employed by the designer. By surveying and analyzing innovative data visualizations, our findings will drive a visual grammar that encompasses the currently conceived data-bindings from creative practitioners. Moreover, if patterns and abstractions exist within this initial grammar, we plan to synthesize our findings into a model that can flexibly account for future creative data mappings.

2 Related Work

Early work on visual grammars from Wilkinson [7] and Wickham [6] uses algebraic transforms to describe the functions that bind data and graphics. Wickham's grammar was motivated by the creation of an authoring tool, much as ours has been. Both these visual grammars follow a paradigm of a data-to-visual pipeline that results in template-style charts. We hope to achieve results where our model is grounded in real data, however it can account for unconceived data-bindings.

Recent work has sought to inspect and survey the area of creative or artistic data visualizations in order to validate this mode of visual communication. Kosara uses a sublimity scale to describe visualizations, with pragmatic and artistic visualizations book-ending that spectrum [2]. Lau and Vande Moere describe how the interdisciplinary field of information aesthetics helps to accentuate the underlying meanings in the datasets, therefore conveying more significance to the observer [3]. These two works are mainly taxonomies of the creative data visualization discipline. Viegas and Wattenberg also provide a survey of the area, but they also examine the techniques artists use to convey meaning in visual analysis [5]. A more recent study by Byrne et al. furthered this research pursuit to understand the acquired codes of meaning conveyed in data visualizations and infographics [1]. This approach methodically analyzes a set of visualizations to draw a synthesized understanding of the techniques employed by practitioners. Our approach is similar, yet our goal differs by seeking to identify, document, and synthesize the methods that designers use to creatively bind data-to-visual elements in the scene graph.

3 Collection of Visualizations

Our analysis covers 68 examples that showcase a wide variety of data visualization for public or casual use. These visualizations exemplify designs that are innovative, appealing, beautiful and informative. By focusing on this genre of visualizations, our analysis aims to expose a visual grammar that either conforms, evolves, or deviates from the information visualization community's canon of visual grammars. The visualizations are taken from three main sources: the Kantar Information Is Beautiful Awards of 2014[1], published data graphics from the New York Times & Guardian during 2014[2], and the Visual.ly Staff Picks of 2014[3].

We chose these specific visualization galleries in order to compile a set of data compositions that will reveal a grammar of innovative visual data-encodings. Each of the three galleries chosen are from reputed sources. Given that each source reviews the visualizations within their gallery, we felt that the aesthetics, creativity, clarity, and data would be of a higher caliber than if we personally sampled visualizations from the wild.

In order to build a visual grammar based on the visual encodings encountered in the three collections, we deconstructed each visualization to its primitive components.

4 Analysis

Our analysis method is a semi-structured coding of the visualizations. Due to the sheer size and variety of the visual encodings present in the set, we required a framework and seed categories to retain structure in our analysis. Munzner's recently published model of information visualization seemed appropriate [4]. The

*e-mail: jrthompson@gatech.edu

†e-mail: stasko@cc.gatech.edu

[1]http://www.informationisbeautifulawards.com

[2]http://collection.marijerooze.nl

[3]http://visual.ly

Component	Attribute	Description	Values
Data Object	**semantic meaning**	the meaning of the data, what it represents	N/A (varies based on data)
	type of data	defined by relationship to other data	Elements, Attributes, Links, Positions, Grids
	category	defined by possible algorithmic transformations	None, Quantitative, Categorical, Ordinal, GeoPoint, GeoShape
Transform	**derivation**	manipulation of a data object into another set	Count, Bins, Average, Nest, Sum, Similarity, Comparison
Visual Encoding	**visual mark**	elemental graphics that compose a scene graph	Line, Area, Point, Figure, Shape, Text, Bar, Control
	data-driven encoding	stylistic or geometric properties of the mark based on data	Length Scale, Position [X and Y], Area Scale, Color Hue Color Saturation, Color Luminance, Tilt/Angle, Movement, Figure Shape, Fill, Repeated Mark, Stroke Width
	purpose	objective of visual encoding	Explanatory, Compositional

Table 1: The taxonomy used in semi-structured coding.

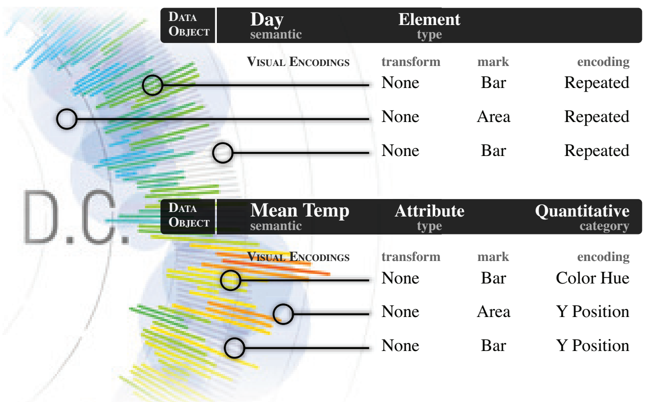

Figure 1: Example of coding for *Weather Radials* visualization.

model's use of data, transforms, and encodings as functional objects in the visualization pipeline provided us with a framework from which to build.

In order to help accurately analyze each visual component that is bound to data, we developed an analysis tool. The tool supports the capture of snippets (as seen in Figure 1) for each visual encoding and then allows for input of the datum attributes, semantics, and data transforms. The current taxonomy used in our analysis is shown in Table 1. We plan for this taxonomy to expand or re-fold through further analysis. Each visualization encountered so far can be broken down into the data objects, transforms, and visual encodings specified in this table. Figure 1 demonstrates part of the analysis for *Weather Radials*[4], a visualization of a year's worth of daily temperature and precipitation records in a collection of cities.

So far we have analyzed 37 of these visualizations. The results of the analysis on that subset is a total of 761 visual encodings and a total of 271 data objects.

The product of our analysis, the meta-data from these annotations, exposes detailed information on each encoding. This meta-data will help in our synthesis of a visual grammar for novel data visualization techniques. Finally, the findings can be shared with the community as a catalog of data-to-visual encodings. Visualization researchers and practitioners can search through different types and styles of data-to-visual encodings that apply to their own projects.

[4] http://weather-radials.com/

5 Discussion

We hope that this work will help to inform the specification and design of future data visualization authoring tools. Through our initial findings we have already discovered a few interesting results. First, text marks are the most prevalent: 285 (37.5%) of the 761 encodings. This leads us to believe that future data visualization research needs to involve the harmonious placement of text within a visualization.

The most common visual encoding was the repeated mark, commonly used for enumerating marks based on a data element. Position data-bindings were the next most common. However, the majority of these visual encodings are compositional, meaning the author used the position of a mark to clarify its relation to another element. This leads us to believe that many position-wise encodings are in fact based on the placement of related visual marks instead of being directly driven by data. We believe that insights similar to these along with the synthesis of our findings into a visual grammar will inform the specification of an expressive visualization authoring tool. Furthermore, as we learn from the creative techniques employed by practitioners, we see this work as a form of communication between practitioners and the research community.

Acknowledgements

This work supported in part by the National Science Foundation via award IIS-1320537.

References

[1] L. Byrne, D. Angus, and J. Wiles. Acquired Codes of Meaning in Data Visualization and Infographics: Beyond Perceptual Primitives. *IEEE Trans. on Visualization and Computer Graphics*, 22(1):509–518, 2016.

[2] R. Kosara. Visualization criticism - The missing link between information visualization and art. *Proceedings of the International Conference on Information Visualisation*, pages 631–636, 2007.

[3] A. Lau and A. V. Moere. Towards a model of information aesthetics in information visualization. In *Proceedings of the International Conference on Information Visualisation*, pages 87–92, 2007.

[4] T. Munzner. *Visualization Analysis and Design*. CRC Press, 2014.

[5] F. B. Viégas and M. Wattenberg. Artistic Data Visualization: Beyond Visual Analytics. *Online Communities and Social Computing*, 4564(HCII 2007, LNCS 4564):182–191, 2007.

[6] H. Wickham. A Layered Grammar of Graphics. *Journal of Computational and Graphical Statistics*, 19(1):3–28, 2010.

[7] L. Wilkinson. *The Grammar of Graphics*. Springer, 1999.