

UNDERSTANDING AND SUPPORTING TRADE-OFFS IN THE
DESIGN OF VISUALIZATIONS FOR COMMUNICATION

by

Jessica Ruth Hullman

A dissertation proposal submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in The University of Michigan
2014

Doctoral Committee:

Assistant Professor Eytan Adar, Chair
Professor Thomas A. Finholt
Professor Paul J. Resnick
Professor Priti R. Shah

Acknowledgments

This chapter may be the hardest to write, as I suspect that anything I write will sound trite in comparison to the depth of gratitude and admiration I feel towards the mentors and others who have helped me achieve this dissertation. I also hesitate to add more front matter lest it should frame this work as something more monumental than it is: a few good ideas and promising lines of thought that could nonetheless stand more polishing, testing, and crystallization. But, as there is no adequate excuse for not writing this, here goes.

I am first of grateful to my advisor Eytan Adar, who has been an influential mentor and academic role model, and the best collaborator I could hope for. I appreciate the space Eytan has given me to develop my own ideas (provided I do so with rigor!) and the generosity he has shown with his own ideas, of which there have been many. Perhaps the most telling thing I could say is that I didn't really believe that computer scientists could be creative until I began working with Eytan. After a short time as his student I was convinced. It's been inspiring to watch him repeatedly produce brilliant ideas, practically on command, and to debate about and reason through the details of our various lines of work with him. I am and always will be proud to call myself his student.

I have had several other influential mentors during my time at SI. Priti Shah has been so generous with her deep intuitions around graphical reasoning. Tom Finholt has been a steady source of encouragement and strategist as I have worked my way through the program. I am grateful to have been one of the students under his wing. In my last year of the program, Paul Resnick has been a valuable collaborator with his clear vision. I am also indebted to Erin Krupka, whose dedication to excellence is a force to be reckoned

with. In a short time, I learned a great deal from her related to experimental design, statistical analysis, and writing. Finally, I owe thanks to Mick McQuaid, who encouraged me in my first year and helped me envision myself as a researcher back when I was a master's student.

Nick Diakopoulos has been a valuable research partner on multiple projects related to communicating with visualizations, including one presented here. Sean Munson has been a consistent source of academic advice and a close friend throughout my degree. I am grateful to Rayoung Yang, Tracy Liu, and Alissa Centivany for their support and needed camaraderie during my degree.

I owe huge thanks to my parents, from whom I get my work ethic. I could not have done this without their longtime support. Special thanks to my mom, Linda Hullman, for her ceaseless encouragement and for listening to me talk about research over so many dinners out.

Finally, thanks go to Cullen for injecting into my bleak doctoral student lifestyle large doses of nonsense, love, and plenty of walks in the woods, and to Ruthy, my fierce and loving wolfish sidekick.

TABLE OF CONTENTS

Acknowledgments	ii
List of Figures	vi
List of Tables	x
CHAPTER I Introduction	1
Understanding Visualization Design As Trade-offs	8
Methodology	15
Outline of the Dissertation and Findings	15
CHAPTER II Related Work	18
Guidelines for Visualization for Communication and Presentation	18
Visual Data Design and Interpretation By Broad Audiences	26
CHAPTER III Rhetorical Influences in Narrative Visualization	33
Bias and Rhetoric in Communication	36
Visualization Rhetoric Framework	41
Illustrating Visualization Rhetoric	62
Discussion	68
Conclusion	71
CHAPTER IV A Deeper Understanding of Sequence in Narrative Visualization	72
Related Work	79
Study: Patterns in Narrative Visualization Sequence	84
An Algorithmic Approach to Visualization Sequence Support	97

Validation Studies: User Perceptions of Sequences	106
Global Sequencing: Impacts of Parallel Presentation Structure	112
Integrating Sequence Support in Automated Presentation	117
Discussion	129
Conclusion	133
CHAPTER V Comparative Sample Plots: Visualizing Uncertainty for Complex	
Visualizations	134
Related Work	138
Uncertainty as Hypothetical Outcomes	143
Domain 1: Community Inference in Networks	148
Domain 2: Choropleth Maps of Voting Patterns	160
Discussion	166
Conclusion	174
CHAPTER VI Future Directions	176
CHAPTER VII Summary	179
References	181

LIST OF FIGURES

Figure I-I: An interactive visualization depicts geographical trends in government benefits to accompany a news article from the New York Times.....	2
Figure I-III: "512 Paths to the White House" presented on the New York Times website prior to the 2012 presidential election.	4
Figure I-IV: An interactive slideshow visualization created by a business consultancy is used to convey the results of a strategic analysis done for a client company.....	5
Figure I-V: A screenshot of an analyst using Tableau Software's automatic visualization suggestion features. Variables are dragged to the "shelves" above the graphical display and an optimized visualization appears.	6
Figure I-VI: A multiple view interactive visualization created by the New York Times graphics department summarizes how various demographic statistics changed during Mayor Koch's time in office.	10
Figure I-VII: Two visualizations from an interactive slideshow visualization created by the Guardian graphics department. The same color space is mapped differently depending on the data in each individual view.....	12
Figure II-I: "Afghanistan, Behind the Front Line" by the Financial Times consistently maps brightness to larger data values in different map views.	20
Figure II-II: Examples of visualizations that include extraneous (non-data representing) elements tested by Borkin et al. (2013).	21
Figure II-III: Snapshot of an animated transition from a scatterplot to a bar chart representation.	23

Figure II-IV: Views from a graphical history tool in Tableau, which enables analysts to save a representation of the actions taken in an interactive visualization analysis session.	25
Figure II-V: Two means visualized in a bar chart with error bars used to depict the range of possible values in which the mean is likely to fall in repeated trials.	31
Figure III-I: "How the Recession Changed Us" presented by the Atlantic Monthly.	47
Figure III-II: "Vehicle Sales" visualization by the Economist Graphic Detail.	48
Figure III-III: "House Democrats' Health Plan" presented on John Boehner's website (top), and "Organizational Chart of the Democrats' Health Plan" created by graphic designer Robert Palmer.	51
Figure III-IV: "A Peek into Netflix Queues" by the New York Times graphics department.	52
Figure III-V: "Speaker Pelosi's Energy Tax: A Bureaucratic Nightmare" presented by John Boehner's office website.	53
Figure III-VI: "Budget Forecasts, Compared to Reality" by the New York Times graphics department.	54
Figure III-VII: "The 2007 State of the Union Address" by the New York Times graphics department.	57
Figure III-VIII: "Mapping America: Every City, Every Block" by the New York Times graphics department.	63
Figure III-IX: Partial (left) and full view of "Poll Dancing: How accurate are poll predictions?" by David McCandless.	66
Figure IV-I: Two scatterplots apply the same colors to two different nominal variables.	77
Figure IV-II: "Show Me" dialogue provides example views for a user to choose from.	83
Figure IV-III: Reverse chronological (temporal) transition sequence in "A Historic Shift" by the NYT.	90
Figure IV-IV: Dimension walk (a comparison transition) in "Europe by the Numbers" by the Guardian.	93

Figure IV-V: Measure walk (a comparison transition) in "Europe by the Numbers" by the Guardian.....	94
Figure IV-VI: Diagram of NYT's "Copenhagen: Emissions, Treaties, and Impacts" showing parallelism.....	96
Figure IV-VII: Transition between depiction of latest recession (top) and economic boom on the 1990s (bottom). Note consistency in axes ranges.	99
Figure IV-VIII: Transition between depiction of latest recession (top) and Great Depression (bottom). Note change of axes ranges.	100
Figure IV-IX: Diagram of graph-based approach in which visualizations represent nodes. Edges (possible transitions) are labeled by type and weighted using a cost function and type weightings (denoted by * symbols) corresponding to user preferences	100
Figure IV-X: Experimental task presenting subjects with an initial visualization (left) and asking that they choose the better visualization to follow the first in a data presentation.	108
Figure IV-XI: Global sequences to support different hypothesized comparisons between consecutive visualizations (depicted with dotted lines).....	116
Figure IV-XII: Example of an input visualization list. Each row is assigned a unique <i>vid</i>	119
Figure IV-XIII: Diagram of graph-based approach. Edges represent transitions between visualization designs with unique vids and no data or mapping violations.	126
Figure V-I: Voting grid map displayed in the NYT shows voting patterns in the 2012 Presidential election by income level.	135
Figure V-II: Contour plots of resampled data convey the extent to which the contour predictions (of pH levels in rainfall) are affected by variation in the input data (Efron and Diaconis 1982).	142
Figure V-III: Overview of the process.	145

Figure V-IV: Two network diagrams, depicting the same node (4, lower left) being predicted members of two different communities (Communities C and E).....155

Figure V-V: Baseline network visualization, plus histograms of reliability estimates for network node membership questions. Incorrect responses to the first part of the question are omitted for clarity.159

Figure V-VI: A map grid distinguishes voting patterns in the U.S. by ethnicity (rows) and income levels (columns). The data represents predicted voting choices and turnout (Ghitza and Gelman 2013).162

Figure V-VII: Two boxplots showing different subsets of the 'mtcars' R data set, created with the default boxplot jitter function in R's ggplot2. Note the differing y-axis ranges..171

LIST OF TABLES

Table III-I: Rhetorical techniques by editorial layer, rhetorical form, and category.....	45
Table III-II: Top ten most frequently observed rhetorical design techniques in sample. ...	58
Table IV-I: Transition types and sample prevalence.	88
Table IV-II: Data representation, associated transition types, and relation to common visualization interactions.....	102
Table IV-III: Results of a multinomial logit regressing “chosen” transition on transition cost and an order indicator and three multinomial logits regressing "chosen" transition on transition types (spanning pairwise comparisons)	111
Table IV-IV: Expressiveness criteria adapted from Mackinlay (1986).....	121
Table V-I: Summary of error rates	166

CHAPTER I

Introduction

Visual depictions of abstract data (information visualizations) support the discovery and understanding of data insights, and increase the memorability and salience of information (e.g., Larkin and Simon 1987, Shah et al. 2005, Tufte 1983, Chua et al. 2006, Stone et al. 1997). As a communication medium, visualization enhances analysts, designers, and others' ability to convey messages based on data to large numbers of others across a variety of domains.

The practice of using visualizations to communicate insights in data to potentially large audiences is not novel. Newspapers and other print media publishers have a long history of using data graphics to emphasize particular points made in an article. With the interactive possibilities afforded by the web and a shift to digital sources of news and information, we are now witnessing widespread popular use of data graphics in online contexts where information is shared. As data representations, these visualizations naturally support data analysis by facilitating value comparisons and perceptually-based judgments of otherwise complex phenomena like trends or “typical” behavior. However, like the print news graphics that predate them, many of the visualizations that are created and presented on the web have clear communication purposes. These “communicative” visualizations are designed with the intention of conveying specific insights to audiences. For such graphics, the analytical operations that they support are more a means of achieving message conveyance than an end that a visualization user manipulates to achieve his or her own insight goals.

“Narrative” or storytelling visualizations bring elements of analytic, user-driven exploratory visualization into otherwise highly guided or “framed” information

presentations. This form of visualization practice has emerged as a new research area in the field of information visualization. The news media are an important source of examples used to motivate research. The New York Times graphics department designers (Figure I.I, Figure I.III) have used interactive visualizations to accompany news articles published on the site, and similar news themed visualizations regularly appear on the BBC website, the Guardian Datablog, and the Economist Graphic Detail blog. These companies employ graphics teams with designers trained in journalism, statistics, programming, and other relevant fields. Their work produces novel visualizations and even graphical formats that combine storytelling with data science to attract the attention of news readers.

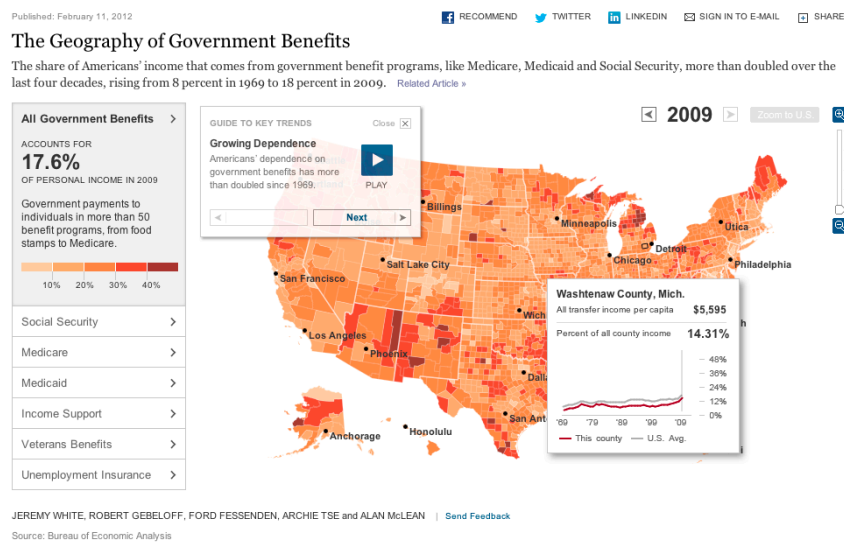


Figure I-I: An interactive visualization depicts geographical trends in government benefits to accompany a news article from the New York Times.

Also drawing attention among scholars and web users are data graphics presented on the websites of organizations that may have particular messages to convey to a group. These include political parties prior to an election, who often use data charts to communicate with voters or other groups of citizens (Figure I.II). The 2011 State of the Union address by Barack Obama was streamed online with a side panel that presented relevant data graphics. Non-profits (e.g. Kiva), national government or private organizations (e.g., Census Bureau) and global organizations (e.g., OECD) who store a

great deal of data are also increasingly presenting interactive data graphics on their websites. In addition their use as communication aids targeted at large public audiences, organizations have created visualizations under the guise of communicating with other organizations or groups. This subset of organizational use includes the graphics created to convey the Republican’s assessment of the Democratic Health Care plan in 2010 (Figure III-III), and the graphics used by the Tea Party during the election of 2008 (Figure I.II).



Figure I-II: Graphic presented by Sarah Palin’s campaign in the presidential election of 2008.

The societal significance of these practices lies in their potential to produce data presentations that act as “social artifacts” that symbolize complex issues and situations important to groups and individuals. The Health plan graphic achieved notoriety in a number of online forums, even provoking a counter-point graphic aimed at the Republican party (Figure III.III). “A Peek into Netflix Queues,” by the New York Times graphic department (Figure III.IV) quickly received over 100 comments discussing subtle assumptions held around the privacy of information

concerning entertainment choices, and the demographics that drive these choices. David McCandless’ “How I Stopped Worrying and Learnt to Love the Bomb (Kinda)” similarly provoked discussion around weapons of mass destruction, including of how to best convey communicate their impact to individual without direct experience.

These visualizations, and the discussions that occur around them in social media or social visualization sites, also shape beliefs and behavior around critical societal issues. For example, the New York Times’ “Paths to the White House,” launched shortly before

the election, depicted a decision tree that detailed 512 possible outcomes for the 2012 presidential election (Figure I.III). Individuals who studied the visualization as they watched the election could update their private forecasts directly and accurately using the external visual aid; some may have even changed their decision about whether or not to vote based on the outcome forecasted by the visualization. Government and other organizations and politicians have also turned to static and interactive graphics to inform and gain the public's support for their agendas, including the creation of jobs, saving of environmental resources, and state of public education.

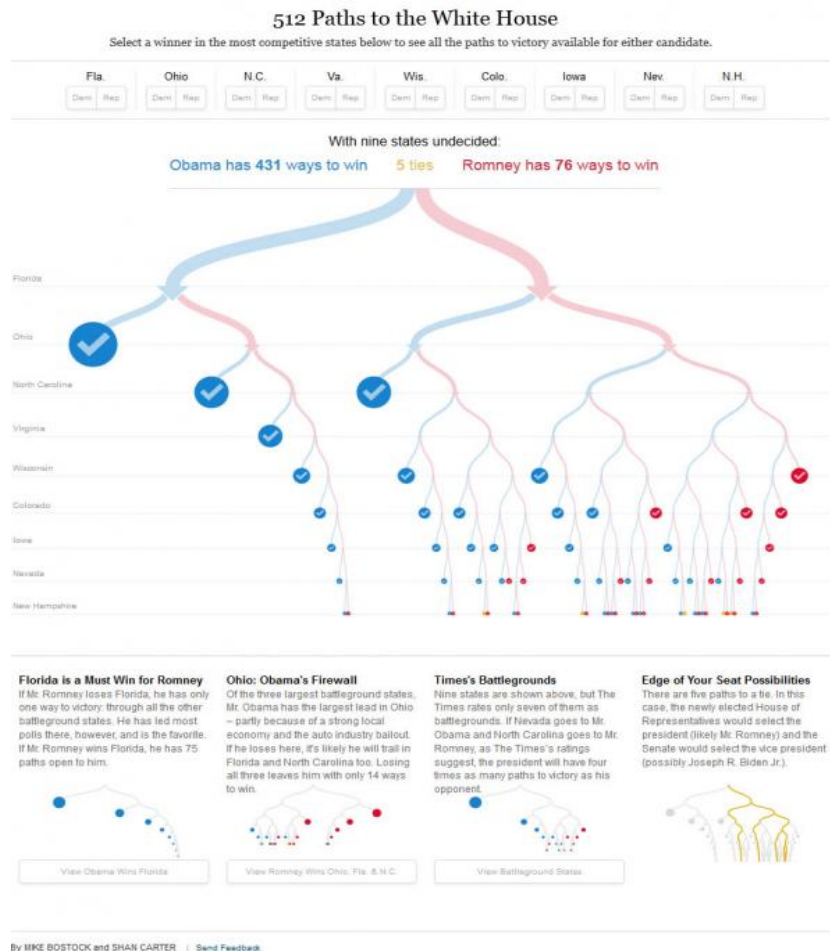


Figure I-III: "512 Paths to the White House" presented on the New York Times website prior to the 2012 presidential election.

Though less visible to the public eye, the use of data graphics use for communication is pervasive in business and other data driven industries, where the results of data analyses must be shared among analysts, clients, and other stakeholders (Elias et al. 2013). Here, the formats tend toward a few canonical types (such as slideshow-style presentations and narrative reports) yet like visualizations created by the media, these visualizations can be quite powerful in communicating insights from data analysis to larger numbers of business stakeholders. Figure I.IV depicts an interactive slideshow format for a business strategy report, created with Tableau Software. The salience of information presented in a well-designed visualization can convince others, such as business stakeholders, of the importance of particular decisions and strategies.

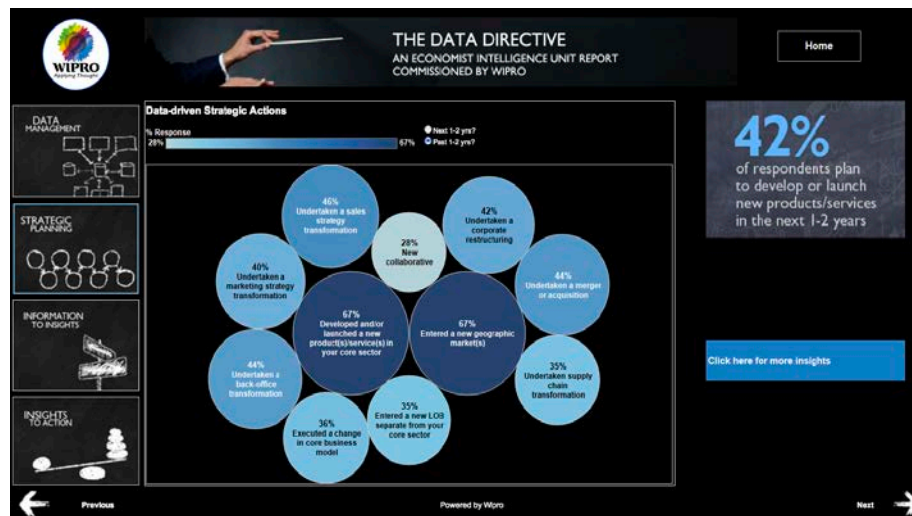


Figure I-IV: An interactive slideshow visualization created by a business consultancy is used to convey the results of a strategic analysis done for a client company.

The creation of visualizations for media-based, organizational, business, and even personal use is supported by a number of data visualization tools. These range from web and desktop software that non-expert designers can easily create visualizations (Many Eyes, Tableau, Microsoft Excel) to programming toolkits like D3, Protovis, Prefuse or Processing or visualization packages for statistical software like R's ggplot2, maps, etc. They do so by making it easier to map data to visual output, for example, with some even providing the visualization creator guidance for choosing an appropriate visual format

given a selected subset of data (Figure I.V). While the particular algorithms and the degree to which automated design features are supported may differ between systems, most share a common intention to ease the effort required to produce graphs by a wide range of visualization creators. Some of these tools, such as Microsoft Excel Charts and more recently, Tableau Software, as well as d3, are very likely supporting the creation of a sizeable percentage of the visualizations that are produced today.

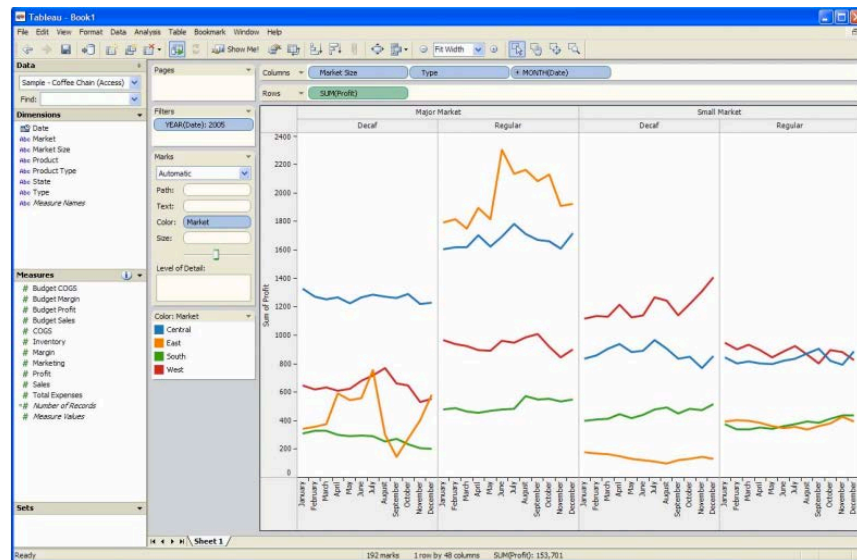


Figure I-V: A screenshot of an analyst using Tableau Software’s automatic visualization suggestion features. Variables are dragged to the “shelves” above the graphical display and an optimized visualization appears.

Despite the many environments in which visualizations for communication currently appear, there remains a need for frameworks and theories to shed light on how storytelling goals can be mapped to visualization features. The knowledge of the professional designers who create many of the most lauded visualizations, such as in popular news media, is tacit and thus unavailable to the many less experienced designers using existing software to create visualizations. Further, while visualization tools have evolved to offer various design optimizations to ease the act of creating a visualization given a data set (e.g., Mackinlay et al. 2007), many of these tools are framed from the perspective of analytical visualization guidelines rather than communication purposes.

For example, in considering particular visualizations that have been successful at engaging and communicating data to broad audiences, it becomes clear that some of their most notable design features remain unsupported in most current tools. Narrative visualizations, for example, are lauded for the way they guide a user through complex data visualizations using features like text annotation and sequential presentation (Segel and Heer 2010). Many of the most popular visualizations strike a balance between the complexity of a representation and its easy interpretability among end-users. These characteristics betray careful decisions made in the design, perhaps with awareness of how end-users would interpret the graph, yet it remains unclear what processes and information are being used to make these decisions. Consider, for example, if the “512 Paths to the White House” (Figure I.III) were shown in a treemap rather than a decision tree format. Both graphical formats are valid presentations for hierarchical data, yet little explicit knowledge is available in the visualization literature for guiding a designers’ decision between the two in this case. Would the graphic be as engaging as a treemap? The more condensed treemap format saves screen space, but is less likely to be so amenable to linear thinking processes that consider possible outcomes as trajectories of actions in multiple possible worlds. The clear hierarchy in the decision tree format makes it easy to hypothesize about different possible outcomes that are mapped to linear paths. The color choices further maintain established associations between Democrats and blue and Republicans and red, further supporting easy understanding. Other types of design decisions are also evident, such as those involved with identifying and representing the particular paths that are highlighted through small representations at the bottom of the decision tree. Finally, the text annotation at the top of the “513 Paths to the White House” graphic summarizes that Obama has 431 possible ways to win the election, to Romney’s 78. The large difference between these numbers of possible winning paths implies that Obama’s win is slightly more likely, yet a statistician might criticize the annotation as an over-simplification of the true probabilities of each candidate’s success, since these

numbers do not capture how probable any given single path is. Would further annotations that explicitly label each path with a predicted probability be a helpful addition to the visualization, or would this information overwhelm users? Such decisions about notions of probability and uncertainty related to data presentation are a common challenge in design.

This dissertation examines current visualization practice for evidence of where particular design needs are not yet supported, and presents empirically-derived knowledge and techniques to support these gaps. Specifically, the dissertation addresses a lack of design guidelines for using visualizations to convey particular messages, a lack of guidelines and support for designing sets of visualizations for presentation, and the need for more easily understandable uncertainty representations. The remainder of this introduction motivates the characterization of these challenges via the concept of design trade-offs, which is a device for framing the studies presented in this dissertation. This is followed by a summary of the dissertation's methodology and an outline of the chapters and key findings.

Understanding Visualization Design As Trade-offs

Consider an interactive visualization like that shown in Figure I.VI, which was presented in the New York Times as commentary on how mayor Edward Koch's tenure spanned important demographic changes to the NYC population. Subtle cues in the display suggest that the designers were faced with *design trade-offs*, decisions between multiple potentially conflicting design features. For example, the designers of this visualization undoubtedly faced decisions about what data from a larger data set to present, and what to omit. These decisions lead to support for some comparisons between data values, and consequently hypotheses, but at the expense of other comparisons and insights that might also be relevant. The designers reference several large data sets as sources for the visualization, including the 2010 Census Bureau data. The visualization

displays four time and or spatially-indexed measures from the Census data: the NYC wealth distribution for three time points spanning 20 years is shown on a map, and line and bar graphs display the white percentage of the population over 40 years, the murder rate over 48 years, and spending on housing over 20 years. The storytelling goal of the designers who created the visualization might be generally described as “New York’s population changed during Koch’s tenure,” with the title of the graphic promising to portray the differences. Yet if we consider the many demographic variables that changed during this time that were also undoubtedly available (such as unemployment rates, or immigration statistics, to name a few) a question arises of “Why were these variables selected?” The fact that only a limited set of variables were chosen from those available suggest an inevitable trade-off between the designers’ ability to fully realize their implied objective of displaying “The Difference Between Koch’s City and Today” and the need to create a display that would inform, but not overwhelm, the average user. The data that appears may have been chosen with a goal of including the measures that show the largest magnitude of change during Koch’s run, and/or other design goals such as maximizing the visual salience of the differences between values of a variable shown in the visualization, or maximizing the relevance of the selected data to current topics of interest in news about New York City.

The Difference Between Koch's City and Today

Related Article »

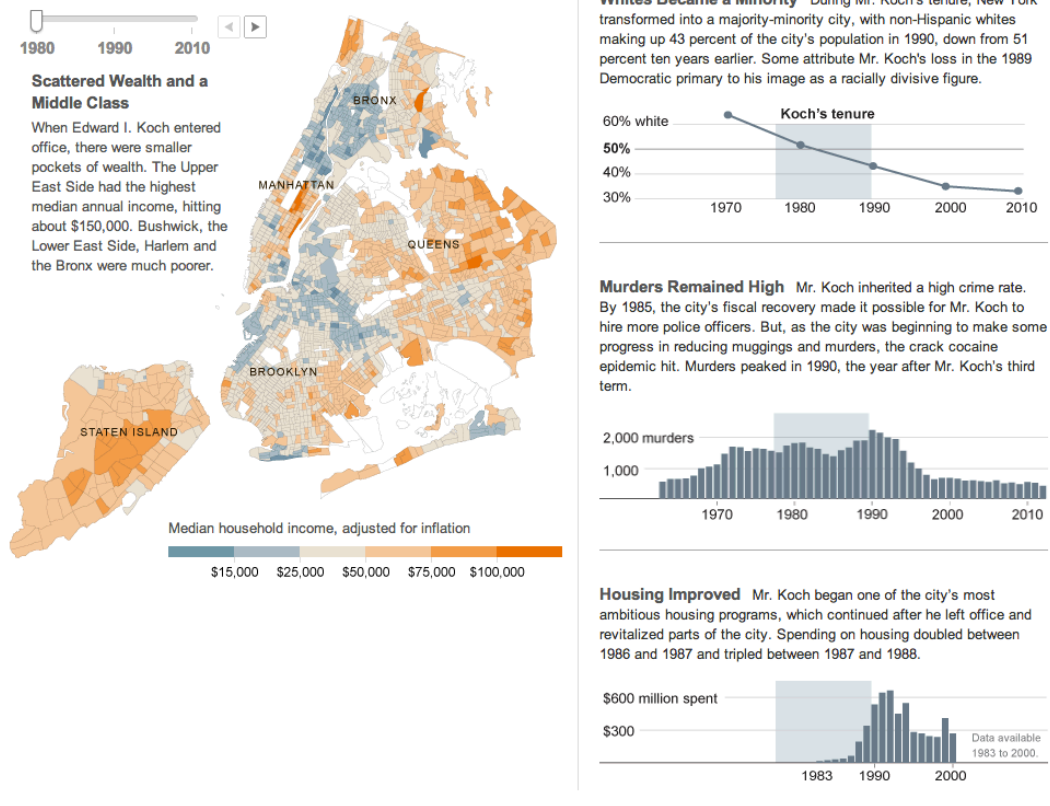


Figure I-VI: A multiple view interactive visualization created by the New York Times graphics department summarizes how various demographic statistics changed during Mayor Koch's time in office.

Interestingly, the necessity of omissions to limit the amount and complexity of data shown in such visualizations runs counter to other desirable features of news artifacts. For example, neglecting to present other relevant data counters the goals of a journalistic code of ethics that prioritizes transparency and objectivity in presenting data to the public (Kovach and Rosenstiel 2007). The goals of this code may have inspired other cues in the visualization that suggest a transparent presentation philosophy. For example, text annotations are used to describe an adjustment to the data to control for inflation (Figure I.VI lower left), and to cite the reason why data on housing is not shown for years before 1983 (Figure I.VI lower right).

The concept of a design trade-off does not only concern notions of transparent versus persuasive visualization design, or data omission versus inclusion. The design of the visualization in Figure I.VI is also likely to have required care on the designers' parts to balance "local" optimizations—those concerning the most effective data-to-visual mappings in the three distinct, *single* representations of NYC wealth data in 1980, 1990, and 2010 to the left of the screen—with "global" optimizations concerning the most effective data-to-visual mappings for the composite visualization comprised of these singular representations. Techniques for achieving a locally-optimal visualization typically automate design decisions in ways that maximize the accurate presentation of data in a single visualization that is being created, such as by mapping the data in a way that makes differences between values most visually distinct. These features are embedded in the visual mapping algorithms used in many visualization software tools, from Tableau Software to R's Ggplot2 package. In this example, if the designer had not taken care to design the mappings of the wealth data from 1980, 1990, and 2010 in a way that considered the properties of all three distributions, the maps are likely to have had qualitatively different color bins applied across the three views, preventing cross comparisons.

For an example of how this consequence might affect interpretation, consider the series of visualizations shown in Figure I.VII. These two visualization "slides" are displayed consecutively in a presentation on wealth and poverty in African from the Guardian. A close look at the use of the color scales in each visualization shows that the same colors are used for very different data values. For example, yellow indicates 65.0 – 82.1% proportions in the top visualization, and symbolizes much lower proportions in the bottom visualization, where yellow indicates 20 – 25%. While the difference between the highest and lowest percentage values in the second visualization are easier to see with this mapping, it is also possible that some end-users will not notice that a new mapping is applied since the same colors are re-used. Identifying the design that best balances these

competing considerations is likely to be difficult and time-intensive due to a lack of tools for modeling such trade-offs.

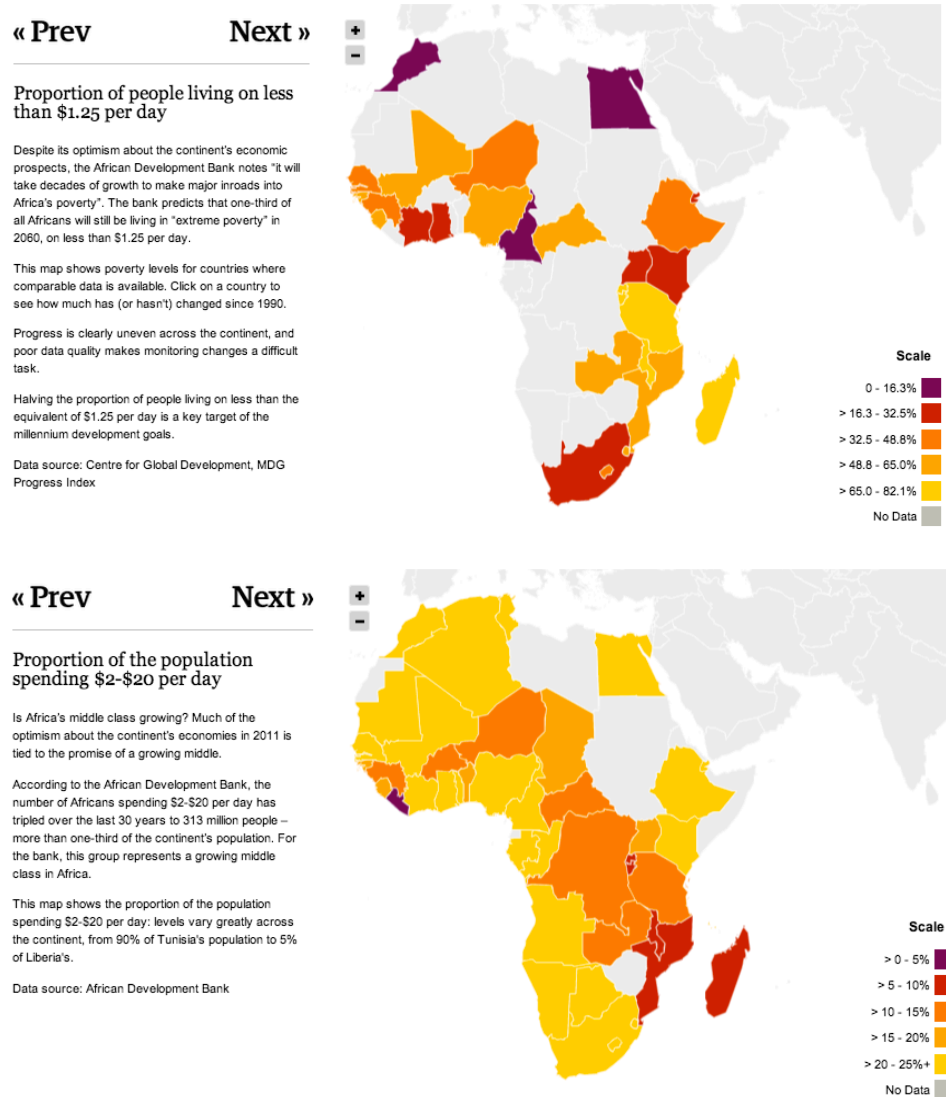


Figure I-VII: Two visualizations from an interactive slideshow visualization created by the Guardian graphics department. The same color space is mapped differently depending on the data in each individual view.

The slideshow format used in Figure I.VII introduces another form of trade-off: the question of the best order in which to present a series of separate yet related visualizations. The slideshow begins with a map visualization showing statistics on the current time period with (when the visualization was created; Percentage Growth in GDP

in 2010). After displaying related statistics describing the current conditions in a series of slides (including those shown in Figure I.VII), the slideshow concludes with a visualization of the Percentage Growth in GDP in 2012 by African country. This decision, to start in the current time and end with a future projection, is likely to impact the content and/or valence of the end-users' conclusions. Presenting the future projection first might even confuse users, preventing them from reaching an intended interpretation like "the economic situation in Africa is gradually improving."

Finally, the visualizations shown in both figures pose interesting questions about design decisions that involve communicating uncertainty to users. In designing visualizations for online audiences, designers must balance the accuracy benefits of faithfully presenting the data as an *approximation* of a real-world phenomena with the easier interpretability of a simpler approach to design. The question of how to display uncertainty remains challenging despite an established line of research in visualization uncertainty, as statistical concepts like uncertainty, variation, or reliability are challenging for most individuals to grasp (Tversky and Kahneman 1971). Nonetheless, uncertainty affects nearly all data sets as result of unavoidable error in collecting and modeling data. For example, if we again consider the data presented in the visualizations depicting wealth distributions for the three years in the left of Figure I.VI, uncertainty is likely to result from the way that data was collected. Income data like that available from the U.S. Census data is subject to the same sources of sampling error as other Census data, including underreporting of certain groups of citizens (such as those in nursing homes), and the possibility that the subset of households that were surveyed are not representative of the greater population (Census Bureau 2012). A researcher (John Logan of Brown University) is cited as the source of the income data that is shown, leaving it unclear if this data set is based on the Census data or a private survey. It is further unclear whether the data may have been transformed or modeled in other ways that introduce the potential for error. Canonical advice emphasizes the importance of providing full provenance information

with a visualization (Tufté 1983), yet the designer of Figure I.VI does not include modeling details nor visually depict information about potential uncertainty in the data. This may reflect an intentional decision based on the designer's awareness of the expectations or skills of many of the end-user. Would the average viewer of this visualization understand how to adjust her interpretation based on potential uncertainty? It is possible that the information was omitted in the interest of not confusing users with complex statistical descriptions or graphical displays of uncertainty.

The significance of this decision comes from the power visualizations have to shape our beliefs, and consequently, our behavior when it comes to phenomena like political elections, environmental issues, or debt. Consider, for example, the political choropleth maps that are often shown leading up to elections by news sources like the New York Times. Such maps display statistical predictions or raw data from citizen polls designed to forecast which candidate will win. When end-users take them seriously as indicative of the true political intentions of their fellow voters, they may modify their own behavior (for example, choosing not to vote if they see the winner to be a foregone conclusion). Hence, in many situations it is critical that properties like the "margin of error" are conveyed. The question arises of how this can be done when standard representations (like error bars for a bar chart) are likely to be misunderstood due to lower levels of statistical background (Belia et al. 2005) or are inapplicable to the chart format.

Trade-offs like those illustrated through Figure I.VI and Figure I.VII—between constrained and comprehensive presentation, transparent versus persuasive framing, singular versus set-wise design optimization, and uncertainty representation versus simpler summary estimates—motivate the work that comprises this dissertation. Using the trade-off as a focal point in the design process, the projects presented here seek to contribute to understanding around the design of visualizations for serving communication as well as analysis purposes, and to provide new knowledge artifacts and approaches for enhancing the current visualization design tool set.

Methodology

As this introduction has served to illustrate, critical decision points in the visualization design process can be thought of as trade-offs. The methods applied in this dissertation are motivated by the belief that to support the production of effective communicative visualizations, it is first necessary to examine the specific considerations that arise at these decision points in detail. Deeper understanding in turn enables modeling or operationalizing of the trade-off points and consequently, the development of techniques for helping designers negotiate the decisions. While various approaches could serve our initial goal of understanding design trade-offs, this dissertation will primarily apply methods that study the output of professional practice as a window into design decisions. Initial study of large samples of professionally-designed visualizations supports the process of inferring and then operationalizing design strategies for effectively negotiating trade-offs of interest. In the last study, an examination of expert practices for quantifying uncertainty serves as the formative study that motivates the development of a new visualization technique.

Throughout this dissertation, formative studies are complemented with experimental methods. Controlled experimentation is used to validate principles that are found and to evaluate the output of the proposed algorithm and techniques that make use of these principles. This “dual approach,” which combines study of design and reception, is warranted by the nature of design trade-offs as decisions between alternatives that are best solved using knowledge of how end-users will react to particular features. Our goal will be to learn something both about the form that these trade-offs take, and the effects that they have on user interpretations.

Outline of the Dissertation and Findings

This dissertation presents three projects to demonstrate how experts’ processes can be learned, formalized, and scaled to support the production of effective communicative

visualizations among broader audiences. This is accomplished by identifying and operationalizing professional's design strategies for use in the development of new design frameworks, techniques and models. Chapter II sets context for the themes explored throughout the work by summarizing forms of practice, challenges, and research in communicative visualization practice.

Chapter III focuses on communicative visualizations designed to persuade users to adopt specific interpretations. We present a study of professional "storytelling" visualization practice to provide insight into expert design strategies for guiding end-users' interpretations. The proposed rhetorical framework characterizes the persuasive dimension of visualization design by providing empirical evidence of several classes of rhetorical design strategies that trade-off comprehensive, impartial data presentation goals with intentions to persuade users toward intended interpretations. Classes of strategies are traced to editorial decisions in the design process, and to their expected effects on end-user interpretation. This chapter also contributes to understanding of how narrative design strategies interact with "extra-representational" factors like viewing codes. Visualization designers also benefit from these results.

Chapter IV takes a closer look at trade-offs that arise in designing communicative visualizations that rely on sequential presentation mechanisms, such as slideshow-style presentations or animations. Here, a creator must negotiate how to divide and present a set of data relationships across multiple data visualizations without losing the sense of coherence and completeness of a singular data graphic. A study of professional practice is combined with online experiments to identify, operationalize, and validate design principles for sequenced communicative visualizations. Studies indicate the need to ease the understandability of transitions in visualization presentations by minimizing the amount of conceptual change. Results also show how high-level structuring principles can benefit memory but may reduce possible comparisons. This knowledge forms the basis of a proposed graph-based algorithm for modeling sequence in the context of design support tools. This approach provides a novel approach to the problems that Chapter IV highlights

around needing to optimize for both local or “single visualization” design and for global “visualization sequence” or set design.

Chapter V demonstrates how statistical modeling practice can inform a new design technique that addresses the trade-off between faithfully presenting data as approximate or uncertain on the one hand, and creating a visualization that can be easily interpreted by users without statistical background on the other. Chapter V proposes and evaluates a new method for generating and visualizing hypothetical data samples in ways that support comparison between multiple sample plots. The proposed comparative sample plot approach conveys data uncertainty more directly than abstract yet conventional uncertainty annotations like error bars. Non-statistician end-users can produce more cautious and at times more accurate estimates of the reliability of data patterns through the use of a comparative sample plots method.

Chapter VI presents a vision for future work suggested by deeper study of design trade-offs in communicative visualization. Chapter VII summarizes the findings of this dissertation.

CHAPTER II

Related Work

This dissertation builds on prior work focused on using visualization to communicate, including designing visualizations for presentation to audiences who may lack advanced training in data analysis or statistics. Specifically, the presented studies address a gap in the literature related to explicit design guidance for narrative visualization creation, including automated support for sets of visualization for presentation, as well as a gap related to conveying uncertainty to non-expert audiences. The concept of design trade-offs that is used throughout this dissertation builds on prior conceptions of the visualization design process.

Guidelines for Visualization for Communication and Presentation

NARRATIVE VISUALIZATION

Early views of data graphics as communication aids describe these visualizations as highly constrained in the amount of data they show (such as only a few data points), as their usual purpose is to display summary statistics (Spence and Lewandowsky 1990). Others note that graphics made for presentation are best limited to very familiar graphical formats (ibid., Kosslyn 1985) as users are generally non-experts. The work of Tufte (1983) promotes examples of analytical and communicative graphics, yet focuses mostly on the value of design features that prioritize analysis (e.g., avoiding extraneous elements) in using visualizations to communicate messages about data.

More recently, Gershon and Page (2001) motivate communication and storytelling features in particular as an avenue for future research in visualization. The authors cite storytelling's potential to aid users in integrating presented information streams, and note the efficiency and intuitiveness of stories as a form of communication. Similar to the more recent work of Segel and Heer (2010), Gershon and Page draw analogies between film editing and visualization, though without explicitly translating film techniques to graphics visualization design. The authors also note features like highlighting and redundant messages across media as traits representative of storytelling graphics. However, their treatment is largely aimed at motivating research, rather than providing specific approaches to the design of storytelling visualizations.

Segel and Heer's (2010) characterization of the narrative visualization design space builds upon Gershon and Page's work, contributing a study of common design features and genres based on 58 examples from news media and organizations (e.g., Minnesota Employment Explorer, Gapminder). The authors used the sample to highlight commonly-used formats (genres) like interactive slideshows, drill down stories, and the martini glass format, in which a constrained author-driven presentation transitions to a more reader-driven exploratory phase later in the interaction session. Their work also illustrates features that tend to co-occur in narrative visualizations, distinguishing these graphics from more analytically motivated visualizations. These include annotation, visual highlighting, tacit tutorials, and progress bars, among other features that similarly serve to guide a user's interaction. The authors suggest the importance of semantic cues for guiding a user through a visualization, such as the usefulness of semantically consistent coloring (Figure II.I). The authors also note the relatively untapped potential for better understanding of how to present information in ways that leverage conventions of use, such as the left-to-right reading order of web pages.

The main contribution of Segel and Heer's work is to acknowledge the various design features and genres that distinguish narrative visualizations. The authors'

contribution stops short of providing specific guidelines for using the features they present, though two examples of professional design graphics are used to illustrate the features in situ. This leaves some area for speculation when it comes to the intended effects on interpretation of these features. There remains a need for theories that support tracing design decisions to their expected effects on the meaning made by a user of a narrative visualization. Chapter III's analysis of visualizations that result from rhetorical design strategies contributes to this gap by characterizing various design features' effects on interpretation via an analogy to "framing effects" in decision literature and related fields.

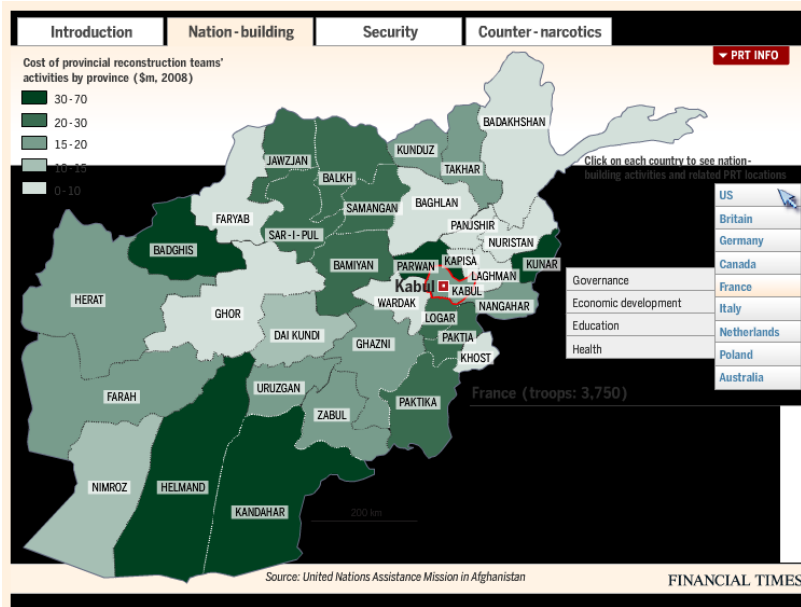


Figure II-I: "Afghanistan, Behind the Front Line" by the Financial Times consistently maps brightness to larger data values in different map views.

Controlled evaluations are one way to fortify the connection between communicative design strategies and visualization reception. For example, many of the features that Segel and Heer (2010) propose as semantically-intentioned design strategies, such as meaningful color mappings, visual highlighting, text messaging, or use of consistent visual layouts could be topics for controlled experimentation. The goal of such studies would be to provide more explicit design guidelines by providing predictive models for how particular features affect interpretations Yet while researchers generally

of understanding that structural properties for visualization presentations such as transitions between consecutive visualizations and high-level “global” presentation structures have. Chapter VI’s evaluation of a global transition structure also connects high-level structural storytelling properties to the user’s ability to make comparisons using a visualization. This contributes to existing work a new outcome measure that directly addresses how the comprehension process is affected by storytelling features, in contrast to techniques that focus only on low-level cognitive symptoms of such graphics (e.g., Borkin et al. 2013, Peck et al. 2013).

GUIDELINES FOR SETS OF VISUALIZATION FOR PRESENTATION

Chapter IV builds on several specific lines of visualization research in visualization transitions and graphical provenance by addressing open questions regarding how to use structure and transitions in communicating via sets of visualizations. Animated transitions are explored by Heer and Robertson (2007) as a method for supporting comprehension of complex transitions between two views in an interactive visualization (Figure II.III). By interpolating and animating intermediate steps between two visualizations, the technique makes it easier for a user to understand how the second view is derived from the first. The authors use two controlled experiments to assess the interpretive benefits of animated transitions, finding that users of animated transitions can better understand syntactic elements (tracing the properties of visual objects and their properties, such as color or size) and semantic elements (tracing the changes to the underlying data schema) of visualizations. These results are promising, but are limited to visualization presentations that are created with systems that support animating transitions, or by visualization designers who implement the transitions themselves. Many of the slideshow format graphics that are discussed in Segel and Heer’s (2010) work (e.g., Figure I.IV), do not use animated transitions. These graphics also present questions that extend beyond the scope of animated transitions, which are limited by the assumption

that a designer has already chosen some transition between views and simply wants to make the transition the least costly to users. How might a design choose between various possible transitions instead? Some researchers have noted that animations may also bring less positive impacts on interpretation, such as by resulting in more superficial understandings than static diagrams or text (Tversky et al. 2002). These limitations suggest exploration of additional approaches for easing the interpretability of complex transitions made in visualizations presentation.

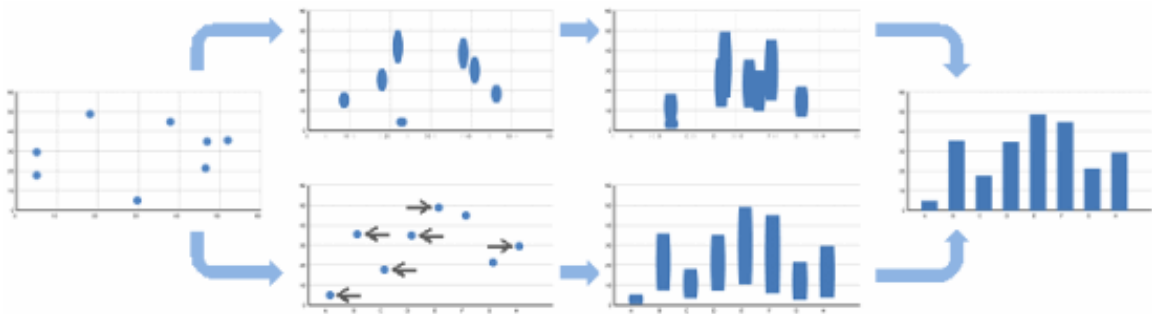


Figure II-III: Snapshot of an animated transition from a scatterplot to a bar chart representation.

Heer and Roberston introduce a taxonomy of transition types and posit several design guidelines for effective use of animated transitions, which may extend to use of transitions in visualization presentation more generally. However, the types they identify tend to be motivated by the analysis interactions that are possible with an exploratory visualization system, such as view transformations (panning or zooming), substrate transformations (axis rescaling), timesteps, and data schema changes. More recently, Heer and Shneiderman (2012) describe similar transition types in characterizing common “interactive dynamics” in visual analysis. These transitions too are motivated by interactive analysis processes which might not be directly relevant to other goals for visualization transitions, such as presenting a set of carefully selected visualizations together in a slideshow format.

Tools for visualization provenance are also intended to support communication and presentation. Graphical histories (Heer et al. 2008, Figure II.IV) is a feature created

for the Tableau visualization system. Graphical histories capture interactions with the tool for later presentation. Communication in this sense is constrained to the use of the “history” by a single analyst who wishes to recall how she used a complex analysis to make a discovery, or presenting a sequence of operations performed with a visualization tool to communicate insights to another analyst. Other provenance systems, such as VisTrails (Bavoli et al. 2005), similarly constrain communication goals to the case where an analyst wishes to leverage a past interaction history. The assumptions in such work, particularly when framed as a way for analysts to share their findings, raise the question of whether an insight is best conveyed by portraying the exact interaction trajectory that produced it. This question is important in considering communication of insights in data to stakeholders beyond other analysts. In cases like reports or presentation prepared by an analyst for a client or by a data journalist for an online audience, the knowledge of the audience regarding analysis methods may not match the analysts who created the presentation. It is possible that presenting a complex series of visualizations that played the role of intermediate steps in an analyst’s process of reaching an insight (such as graphs showing various transformations to a measure of interest, like log scaling or regression models) would confuse general audiences. It is also possible that some intermediate visualizations are not useful for helping an audience member to understand an insight, even if that audience member has sufficient training to understand a complex transformation. Chapter IV of this dissertation addresses these questions by motivating an understanding of transition and sequence as visualization properties with purely presentational goals, allowing for presentation orders that are distinct from the order of visualization viewing or creation during analysis.

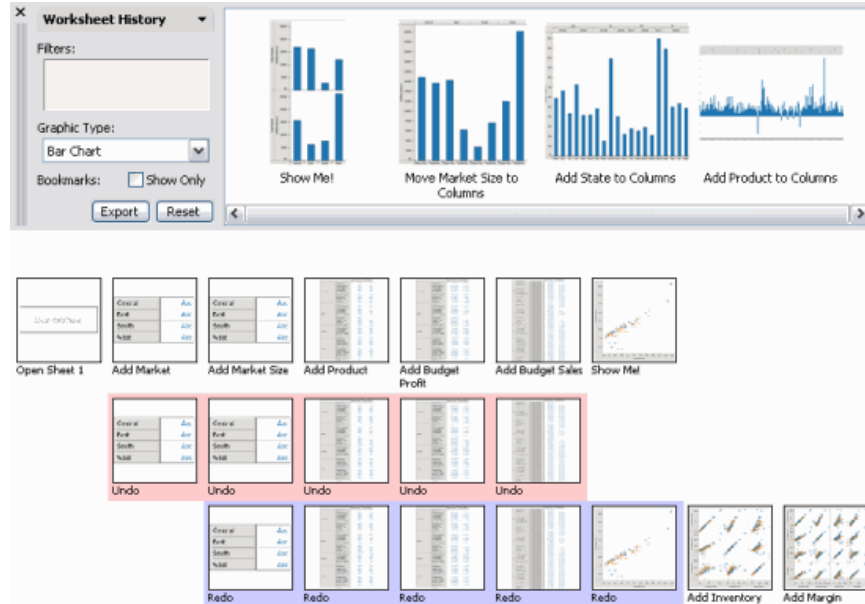


Figure II-IV: Views from a graphical history tool in Tableau, which enables analysts to save a representation of the actions taken in an interactive visualization analysis session.

Each of the studies presented in the chapters of this dissertation, and particular Chapter IV, touches on visualization presentations composed of multiple separate visualizations. This makes research from psychology and information visualization that notes the impact of other representations on graphical interpretation relevant. Multiple representations are frequently present in contexts like news media (e.g., text articles), and a combination of media characterizes many narrative visualizations themselves, like slideshows combining data graphics with text, images, and videos. Ainsworth (2006) presents a conceptual framework intended to provide understanding around how learners make use of multiple representations, including to complement or constrain interpretation and construct deeper understanding. Several of the guidelines presented as part of this framework are directly relevant to the goals of Chapters IV and V, which deal with specific questions around presenting multiple representations. Ainsworth notes that visualizations that appear in sets or with other media introduce new design considerations, including how much information is redundant across the representations, and in what order they are presented. Chapter IV's studies and proposed approach for

modelling sequence in visualization design specifically works to formalize design guidelines for such characteristics of visualization sets.

Visual Data Design and Interpretation By Broad Audiences

This dissertation follow prior work in emphasizes the notion of design trade-offs as a common type of decision encountered in designing visualizations to achieve communication goals. Several chapters of this dissertation address gaps in the existing visualization toolbox when it comes to supporting the design process of visualization creators that may include non-professional designers or data analysts. Additionally, visualizations that are created to communicate are often targeted at broader groups of end-users than were assumed in early visualization research. This dissertation presents a technique that extends prior work in designing uncertainty presentations appropriate for non-expert users.

SUPPORTING THE DESIGN PROCESS AMONG NON-PROFESSIONAL CREATORS

Existing characterizations of the design process for communicative visualizations include Kosara and Mackinlay (2013), Jain and Slaney (2011), and Cairo (2012), who draw analogies between communicative visualization design and the process that a journalistic uses to prepare a story. A communication goal is defined, and information is gathered that may support that goal. The designer then engages in a series of decisions that involve which subset of information to include, what format to present that information in, and how to frame, style or render the final result. The introduction of this dissertation presented examples of challenges that can affect these steps, such as achieving a balance between the amount and detail of presented information and the usability of the visualization. These decisions between potentially conflicting design alternatives, or trade-offs, frequently occur in designing visualizations for communication and online presentation. However, the algorithms that designers use to solve them are difficult to

document or even to apply in strict, rule-based ways. Agrawala et al. (2011) describe design principles for visual communication as “rules of thumb that might even oppose and contradict one another” (pg. 62). The knowledge that professional designers bring to bear in creating communicative visualizations may be utilized regularly without ever being clearly articulated (ibid., Vande Moere and Purchase 2011).

These characterizations largely align with more general theories and empirical findings from design research that construe design as reflective, exploratory, and creative (e.g., Schon 1995, Dorst and Cross 2001, Cross 2004). The goal of a design, or constraints that are applied, may itself shift as a result of a designer’s exploration of the design space (Dorst and Cross 2001, Schon 1995). Additionally, the space of possible designs tends to be large for many design problems, including visualization. Observations of expert practice suggest that designing a high quality communicative visualization is a highly time consuming process (Agrawala et al. 2011) that involves, for example, the generation of a large number of “practice” attempts. Munzner (2009) addresses the various threats to the validity of an information visualization that can result from inappropriate design decisions. Equally telling of the number of considerations that designers must be aware of in creating a visualization is the way that upstream errors can cascade to all downstream levels. For example, an inappropriate abstraction of the needs of web searchers (to see a display of the full connectivity graph) can motivate multiple clever solutions that are nonetheless confusing and more disorienting to web searchers than having no representation at all (Munzner 2009).

Overall, these portrayals of design suggest two avenues for research that aims to provide design support tools for designers who may not have extensive training or professional experience. First, the complexity of the design process as practiced by experts, and lack of clear articulations of their strategies suggests the need for work that provides explicit guidelines to help non-professionals create similarly effective visualizations. Vande Moere and Purchase (2011) argue this point by suggesting that

frequent articulation of design reasoning in presenting visualization research can help demystify what might otherwise be considered a “romantic” process. Byron and Wattenberg’s (2008) work on stream graphs, among others, are cited as evidence of how presenting detailed design reasoning around decisions made in designing a new visualization technique (including seemingly inconsequential decisions like color choice) can result in generalizable knowledge. Each of the studies presented in this dissertation includes some articulation of principles for effectively designing visualizations to achieve common goals. The goals for which this dissertation provides design guidelines include message conveyance, the coherence and understandability in a set of presented visualizations, and visualizing uncertainty to non-expert users.

Secondly, there is also room for more research into ways to embed design principles in visualization tools, as an alternative way to make them easily accessible to non-professional creators who may lack professionals’ experience. However, this integration task presents challenges even when design reasoning is clearly articulated. The notion of trade-offs between equally plausible designs as opposed to a single, “best” design (Agrawala et al. 2011, Vande Moere and Purchase 2011) and large space of possible designs suggests that principles for effective designs be implemented in tools in a way that supports consideration of multiple plausible designs. The notion of multiple plausible effective designs is supported by the proposed approach to modeling sets of visualizations for presentation in Chapter IV, which is designed to reduce the time and effort required for designers to explore the full space of design alternatives for single visualizations and the of possible linear paths for presenting a set of visualizations. Support for comparing possible designs to one another, in this case by reflecting on the probable framing of interpretive effects of different designs, is provided by the taxonomy of rhetorical design techniques presented in Chapter III.

CONVEYING COMPLEX DATA TO DIVERSE AUDIENCES

Early information visualization research (and various trajectories of more recent work) frames visualization as a tool for gaining insights into large scientific data sets among scientists or analysts who regularly engage in visual data analysis. Many of the examples of successful communicative visualizations discussed in recent research support interpretation among audiences that are larger and potentially more diverse than these expert populations. As noted above, unique to many communicative visualizations presented online today (such as those created by the New York Times graphics department) is their complexity, which goes beyond that which is prescribed for presentational data graphics in earlier work (Spence and Lewandowsky 1990). The shift in attitude among researchers about the capabilities of “non-experts” is described by the creators of IBM’s Many Eyes (Viegas et al. 2007) as a movement to “democratize” visualization by making the technology available to the broadest possible audience. This dissertation takes a similar viewpoint that non-expert analysts are capable of interpreting complex visualizations created by professional designers and of creating high quality communicative visualizations themselves given access to knowledge and tools.

Specifically, Chapter V of this dissertation addresses issues related to conveying approximations to broad audiences of end-users. While the definition of “effective” visualization for communication departs from notions of accurate interpretation alone, such as by also covering how well a visualization conveys an intended message, supporting accurate interpretation of the data values themselves remains critical as a design goal for communicative visualizations. The differences in interpretation that can result from a lack of data and graphical literacy on the part of visualizations users present challenges to visualization research. Statistical knowledge has long been assumed in visualization research, but the assumption that users can correctly interpret statistical summaries like confidence intervals or variance estimates may not hold for large groups of online visualizations users (Busse and Hong 2008, Viegas et al. 2008). A general tendency toward

insensitivity to sample size may be intentionally corrected by a scientist aware of the higher variation likely in a small sample, for example, but never enter the awareness of users who lack statistical training. Micallef et al. (2012) provide evidence that a finding that Euler diagrams positively impact the accuracy of Bayesian reasoning no longer holds when the user audience diversifies to include workers from Amazon's Mechanical Turk. These asymmetries in research results across different user populations motivate usable uncertainty visualizations as a key goal for visualization research (MacEachren 1995, Marx 2013).

The importance of uncertainty to visualization design is established in existing characterizations of the “visualization pipeline”¹. In information visualization research, this term is generally applied to the steps taken by a visualization designer (often a researcher), starting from an initial intent of capturing real-world phenomena and resulting in visualized data chosen based on its ability to depict this phenomena (first referenced by Card and Mackinlay 1997, further described in Munzer 2009, among others). The steps in this pipeline generalize to the same process that characterizes many inductive scientific practices that aim to generate knowledge about real-world phenomena. A population of interest is identified, with some target characteristic that is to be investigated in more detail through data analysis. This might take the form of a question or theory that a research poses about some facet of reality, such as theory about the impact of gender on math scores, or a query about how the average rainfall in some

¹ The term “visualization pipeline” has also been applied to the steps involved in rendering imagery produced by graphics software (e.g., algorithms for achieving realistic lighting on a simulated scene, Van Dem et al. 1994). However, the pipeline characterization referenced in information visualization generally describes a broader set of operations.

unknown location compares to another known location. Measures are chosen for sampling from the population in a way that captures this characteristic. A representation is chosen, and the data presented and, if necessary, described (or more generally, framed) such that the phenomena of interest is clearly perceivable. Descriptions of the pipeline frequently acknowledge the potential for bias or uncertainty to enter the creation process (e.g., Amar and Stasko 2004, Munzer 2009, Pang et al. 1996, MacEachren et al. 2005). For example, data capture is subject to statistical error, the noise or error that arises from uncontrolled sources of randomness that are generally unavoidable in data collection (Snedecor and Cochran 1989). Captured data might be cleaned, aggregated, or otherwise transformed so that it more clearly represents the target phenomena; however, this choice brings the potential for further bias if a selected method is not appropriate. Detailed information can be lost (such as when only averages of data variables over a population are displayed). The ubiquity of uncertainty in the pipeline makes the choice of representation for presenting information on uncertainty to users of a visualization a critical one.

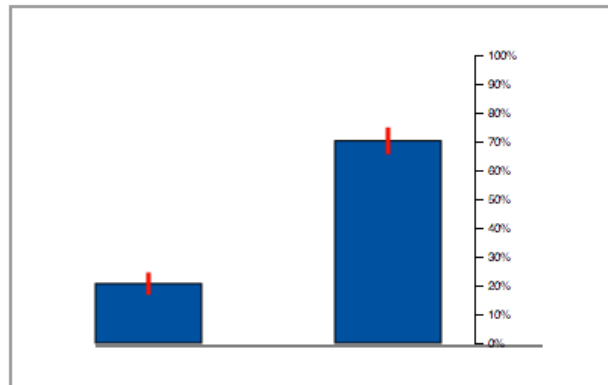


Figure II-V: Two means visualized in a bar chart with error bars used to depict the range of possible values in which the mean is likely to fall in repeated trials.

A common approach to communicating uncertainty is to convey the range of values that a statistic presented in a graph might plausibly take if the data collection and modeling process were to be repeated. Researchers have described how error bars, a relatively common uncertainty representation denoting confidence intervals, are

misinterpreted by a many visualization users, including experts (Belia et al. 2005, Marx 2013; Figure II.V). Other prior research has also articulated a risk that users of visualizations will discount uncertainty representations as peripheral (Buttenfield 1993). Our work on comparative sample plots addresses this gap by providing a generalizable technique for visualizing conveying uncertainty more directly as hypothetical samples. The comparative sample plots technique achieves the same goal as an error bar (depicting a confidence interval) but the more immediate presentation of the hypothetical samples as equally important visualizations. The directness of the technique makes it a more user-friendly method for populations that might lack statistical training in how to interpret the interval depicted by the error bar.

CHAPTER III

Rhetorical Influences in Narrative Visualization

As described in the preceding chapters, narrative visualizations introduce a unique set of design trade-offs, incorporating aspects of communicative and exploratory visualization into graphical displays that are often novel and complex. The most notable of these design tensions concerns what is described by Segel and Heer (2010) as a spectrum along which individual narrative visualizations can be located, where one end represents highly constrained, “author-driven” visualizations while the other represents more “reader-driven” visualizations, those that allow the user flexibility in the specific interaction trajectory and interpretation they experience.

Acknowledging that “author-driven” visualizations comprise a large number of information visualizations being presented in the media is an important step in the history of visualization research. The intentions of a creator have long taken a back seat to characterizations of visualizations that prioritize “letting the data speak” (Tufte 1983) by eliminating sources of bias or framing. However, existing characterizations of narrative visualization such as that contributed by Segel and Heer (2010) detail design features at a relatively high level (e.g., use of text for messaging). While such characterizations can be generalized to describe a large number of narrative visualizations, the persuasive effects achieved by many of the more “author-driven” of these visualizations remain unconnected to characterizations of design. Our understanding of how these visualizations are created to fulfill their presentational goals, and how narrative features are interpreted, remain limited.

In this chapter we examine the design and end-user interpretation of narrative visualizations in order to deepen understanding of how common design techniques represent rhetorical strategies that make certain interpretations more probable. We draw motivation from studies in semiotics, journalism, and critical theory that indicate particular rhetorical techniques used to communicate an intended message (Anderson 2000, Barthes 1978, de Souza 2005). Our work is further informed by evidence from decision theory, survey design, and political theory (Kahneman et al. 1982, Schwarz et al. 1985 and 1991, Saris and Sneiderman 2004) that suggests that subtle variations in a representation's rhetorical or persuasive techniques can generate large effects on users' interpretations of a message. Investigations related to InfoVis provide initial evidence that how data is framed or presented can significantly affect interpretation (Zacks and Tversky 1999).

We consider the design trade-offs that occur in the creation of a narrative visualization with reference to the various “rhetorical” or persuasive goals that these artifacts often evidence. This chapter contributes to InfoVis design and theory by providing insight into (1) the types and forms of use of particular rhetorical techniques in narrative visualizations, and (2) the interaction between those techniques and individual and community characteristics of end-users. The first contribution is a taxonomy of how particular design elements can be used strategically to directly or indirectly prioritize certain interpretations. This equips designers with a set of techniques for designing engaging narrative visualizations capable of communicating layered meanings. At the same time, the identification of classes of rhetorical techniques provides both designers and InfoVis researchers with a vocabulary for analyzing the underlying rhetorical functions of particular design strategies. These goals of visualizations remain under-discussed in many theoretical frameworks organized primarily around exploratory visualization.

Our second contribution is a set of concepts for understanding how these conventions interact with characteristics of the visualization interaction, end-user's knowledge, and the socio-cultural context. This stands to improve designers' awareness of how designs might be received differently by individual end-users and how they can cue shared cultural knowledge and associations. These "extra-representational" factors also tend to be neglected when designing or analyzing visualizations based on design principles such as those proposed by Tufte (2001, 2006). Researchers in InfoVis can benefit from a holistic understanding of visualization interpretation capable of providing insight into how particular interpretations arise as a result of interactions between a visualization, user mental models, and other external representations. This view is congruent with a distributed cognition model of InfoVis (Liu and Stasko 2010).

The remainder of this chapter is as follows: important terms related to rhetoric are defined and these concepts contextualized within InfoVis as well as semiotics, decision science, and political theory. We also describe our work in the context of research on narrative visualization. We begin our presentation of the Visualization Rhetoric framework by considering the process of narrative visualization creation, proposing that design decisions can be differentiated based on the components or "parts" of the visualization they pertain to (e.g., the data, the visual mapping, the textual framing). We use systematic study of a sample of narrative visualizations to trace the effects of specific editorial choices representing rhetorical strategies that give these visualizations their characteristic "framed" nature, in comparison to visualizations produced primarily for data depiction and analysis. Analytical devices for understanding the site of techniques and their interaction with end-user characteristics are also presented. The Illustrating Visualization Rhetoric section uses two case studies to demonstrate how an understanding of visualization rhetoric can provide insight for the analysis and design of narrative visualizations. The Discussion section reflects on themes that emerge from our analyses and highlights areas for future study.

Bias and Rhetoric in Communication

In this section we address the terminology used in the chapter and define visualization rhetoric. We then motivate the importance of our work and contextualize it with that of other relevant fields. This draws attention to the need for deeper understanding of visualization interpretation as it relates to rhetorical techniques and design.

A NOTE ON NOMENCLATURE

This chapter's focus on visualization rhetoric stands at the intersection of ideas of *bias* and user-designer relationships as understood in InfoVis, on the one hand, and theories of *rhetoric*, *framing* and author-reader interactions as elaborated in critical semiotic theories for literature, political rhetoric, and media artifacts on the other. *Bias*, *rhetoric*, *framing* (and the related literary term *perspective*) all describe how an interpretation arises from the interaction of representational, individual, and social forces. Differences can be traced mostly to superficial differences adhering in ordinary language. *Bias* is often defined in negatively connoted terms: "a systematic error introduced into sampling or testing by selecting or encouraging one outcome or answer over others" (Merriam-Webster). To *frame* an idea is typically more neutrally defined as to "form or articulate" (Oxford American) or "shape, construct" (Merriam-Webster). Similarly, the concept of perspective tends to be either neutrally or positively-connoted in literary and critical theory as a productive force in the telling of a story. The term *rhetoric* has a complex history, but has come to be associated with persuasion as a result of the implicit motivation of the speaker to gain other adherents to a preconceived view or conclusion (Bogost 2007).

We use the term rhetoric to refer to the set of processes by which intended meanings are represented in the visualization via a designer's choices and then shaped by individual end-user characteristics, contextual factors involving societal or cultural codes,

and the end-user's interaction. While this term may bring to mind negatively connoted notions of persuasion as bias common in some InfoVis literature, we seek to objectively describe the rhetorical nature of visualization design rather than to comment on the appropriateness of persuasion in visualization design.

INFORMATION VISUALIZATION

Despite its parallel meaning to terms like *rhetoric*, the pejorative term *bias* is more often found in InfoVis literature. Early theory emphasizes the analytic nature of graphical displays (e.g., Kosslyn 1989, Larkin and Simon 1987) as well as automated methods that optimize constraints imposed by human perceptual and cognitive abilities (e.g., Mackinlay 1986). Unequivocal designs are prioritized; “in the ideal case a chart or graph will be absolutely unambiguous, with its intended interpretation being transparent” (Cleveland 1994, pg. 192). Immediate clarity and minimal intervention on the part of the creator are emphasized (Tufte 2001). Where editorial choices must be made, designers are urged to provide detailed provenance information like the objective, time, and location of graph creation (Tufte 2006).

Some recent work in InfoVis has striven to overcome the narrow focus on optimizing visualization clarity and efficiency that dominated earlier work, acknowledging that interacting with a visualization involves thinking about and being influenced by factors beyond just the visual representation. Recent evaluation models (Munzer 2009) explicitly acknowledge that risks to validity can enter at levels beyond the visual encoding and interaction design, such as in characterizing the domain tasks and data. Additionally, several studies demonstrate that extra-representational preferences and conventions can influence interpretation, such as when the visual format cues interpretation frames (Bateman et al. 2010), or individual differences lead to differing visualization usage (Ziemkiewicz and Kosara 2009). As Norman (1999) describes, interpretations can be unpredictable when design elements may not immediately

communicate the designer's intended meaning as a result of influences on interpretation deriving from the end-user's context. Liu and Stasko (2010) frame the site of such differences via the mental model concept, arguing that the effects of such differences on interpretation have been underexplored in InfoVis. This supports a call for further consideration of visualization's role within webs of situated representations.

The visualization rhetoric model we propose is likewise motivated by an expanded view of visualization that takes into consideration under-acknowledged facets of design and interpretation. For instance, creating a visual representation necessitates simplification, as data is used to create an analytical abstraction that is transformed to a visual representation (Ziemkiewicz and Kosara 2009). Thus a rhetorical dimension is present in any design. Secondly, a designer's intentions may remain implicit and inarticulable by him or her, making it impossible to comply with the principle of providing full provenance. From the end-user's perspective, the pleasure of a concise, visual representation may be decreased if engaging with the visualization also requires sifting through explicit description of every design manipulation.

FRAMING IN DECISION AND OPINION FORMATION

Empirical studies in decision theory and political messaging provide evidence that even subtle changes in the rhetorical frame of an information presentation can significantly influence responses. In contrast to the rather dismissive viewpoint on intentional use of rhetorical devices in InfoVis literature, psychological, political and communication theorists have developed framing theory to investigate opinion formation processes in light of how people orient their thinking about an issue. Typically, these processes are viewed as responses to the use of particular communicative structures in messaging (e.g., Tversky and Kahneman 1981, Kahneman et al. 1982, Chong and Druckman 2007). Researchers seek to better understand "framing effects", situations where often small changes in the presentation of an issue or an event, such as slight

modifications of phrasing, produce measurable changes of opinion (Saris and Sneiderman 2004). Information representations can influence interpretation in diverse ways, such as by presenting a preliminary statistic before a decision (ibid.), or by manipulating the anchor points on a survey scale (Schwarz et al. 1985). Of particular relevance to InfoVis are findings that are explicitly visually based. For example, the amount of space provided between response choices in a scale can be interpreted as reflecting the underlying dimension and lead to different results when manipulated (Tourangeau et al. 2008). This literature further motivates a need to articulate and understand the implications of rhetorical strategies in visualization.

SEMIOTICS

Semiotics describes literary, visual, political, and other critical studies that examine how representations like texts, paintings, iconography, or media messaging can be decomposed into systems of signs. Signs—(defined as any material thing that stands for a non-present meaning, such as a word, color choice, or visual icon)—become meaningful through their interaction with other signs within a representation, as well as with signs that are culturally present (e.g., Barthes 1978). Semiotic theory has been introduced in HCI as an inspection method for interactive interfaces to help assess the designer-user meta-communication via the interactive artifact (de Souza 2005). First applied by Jacques Bertin (1984) as a tool for describing how information visualizations convey meaning, semiotic theories emphasize the communicative properties of visualizations alluded to in recent works (Viegas and Wattenberg 2006). This can serve designers seeking to better convey their intended messages (Anderson 2000) and increase their awareness of how design choices may affect interpretation. Semiotic theorists analyze the relationships between forms of media, their production, and the “modes of seeing” or interpretive conventions that they engender. The concept of viewing codes, including visual, textual, cultural, and perceptual (Chandler 2001), describes the implicit, often internalized

standards that support interpreting an artifact in a certain way. This motivates incorporating extra-representational factors like individual and group conventions into a visualization rhetoric framework.

NARRATIVE VISUALIZATION

In response to the growing number of online visualizations designed to convey a story, Segel and Heer's (2010) design space analysis presents three ways of distinguishing categories of narrative visualizations: (1) genres; (2) visual narrative tactics that direct attention, guide view transitions, and orient the user; and (3) narrative structure tactics such as ordering, interactivity, and messaging. Their contribution of abstract structures and genres provides a general framework that opens the discussion of narrative visualization to a wider range of examples. The framework also allows comparisons between visualizations based on how they structure users' interactions with data. We aim to expand the discussion of narrative visualizations to include the role of extra-representational influencers like individual, group, and contextual differences in interpretation. We outline additional visual and non-visual tactics used in narrative visualization, emphasizing how these represent omissions, additions, and implications.

Ziemkiewicz and Kosara (2009) contrast information visualization with visual representations. Narrative visualizations tend to be excluded from their model by criteria like non-trivial interactivity (allowing users to change the visual mapping parameters themselves) or non one-to-one mappings between the source domain and the visual output domain. In contrast, our work explores the dynamics of constrained interactivity and techniques like visual redundancy that are used to emphasize an intended meaning in narrative visualization. We also extend their discussion of information loss by considering the rhetorical effects of information omissions regardless of intention, based on our belief that the increased presence of such visualizations makes it important for InfoVis

researchers and practitioners to better understand how the editorial process of visualizing data necessarily constrains possible interpretations.

Visualization Rhetoric Framework

A primary contribution of this chapter is the development and demonstration of an analytical framework to guide discussion of the rhetorical aspects of InfoVis. In this section we present conceptual devices as well as the results of a large qualitative analysis used to identify specific rhetorical strategies used in InfoVis. We begin by describing the *editorial layers* of a visualization presentation where rhetorical choices are made, then describe the particular *visualization rhetoric techniques* identified in our analysis. A discussion of *viewing codes* follows, including aspects of *denotation* and *connotation*, which helps capture the role of end-users' implicit beliefs and knowledge in visualization interpretation.

EDITORIAL LAYERS

Editorial judgments, and thus rhetorical techniques, can enter into the construction of narrative visualizations from multiple paths. We distinguish between four **editorial layers** that can be used to convey meaning, including the data, visual representation, textual annotations, and interactivity. A given rhetorical technique might be applied to some layers more easily than others. Yet omissions, emphases, and ambiguity can be accomplished at each level. As the output of a designer's decision processes, a narrative visualization represents a sequence of choices to either add information (such as by adding suggestions of an intended message using textual annotations) or omit information (such as by omitting some variables or interactivity features). Distinguishing the possible sites of these choices paves the way for more recognition of their existence, and effects on end-user interpretations.

At the lowest level of the **data**, the creator of a visualization makes choices about the data source to represent, including what variables to include and which to leave out. Additional choices can further affect data, such as removing outliers, scaling, or aggregating values. Both of these particular data choices lead to loss of information in the final representation, yet are necessary choices in the act of visualization design (see Information Access, below). The **visual representation** layer carries traces of choices made about how the data will be mapped to the visual domain. Often, this mapping is lossy as a result of human visual perception abilities. For example, mapping a continuous variable to a gray scale leads to “lost” information due to human perception’s sensitivity and capability to distinguish different intensity levels (e.g., “just noticeable differences”). **Annotations** can be textual, graphical, or social, as in the inclusion of user comments in the overall presentation. Annotations have often been overlooked in InfoVis evaluation, yet serve an important role in many presentations that include visualization by focusing a user’s attention on specific areas in a graph. Finally, the **interactivity** of the visualization can be the site of choices that constrain a user’s interaction in ways that lead her to explore certain subsets of data. This can occur through navigation menus that limit the number of views of the data set that are possible, or linked search suggestions that likewise encourage the user to explore particular views over others. .

VISUALIZATION RHETORIC TECHNIQUES

We describe and present findings on the rhetorical strategies we observed in an extensive analysis of online narrative visualizations.

METHOD

We gathered a sample of fifty-one professionally-produced narrative visualizations, many from international news outlets like the New York Times (NYT) or BBC. In the interest of diversity we also included online visualizations from news magazines (e.g. The

Economist); local news providers (e.g. annarbor.com.); political outlets (e.g. Obama.org, website of the speaker of the house); and independent graphic designers known to publish their work in leading news outlets (e.g. David McCandless). Prior to coding, we familiarized ourselves with framing or bias techniques identified in semiotics (e.g., Barthes 1978, Bertin 1984, Chandler 2001, de Souza 2005), statistical presentation (e.g., Huff 1993, Tufte 2001), decision theory (e.g., Tversky and Kahneman 1981, Kahneman et al. 1982) and media and communication studies (e.g., Nelson and Oxley 1991). We iteratively coded particular techniques we observed referring to this set of theories as a guide, and relied on general knowledge of current events and how to interpret various graph formats as needed. We restricted our analysis to the details present in the visualization and their surrounding presentation. The saliency and primacy of the observed techniques were considered as the examples were coded. As coding progressed, we noted where techniques appeared to represent different implementations of the same basic function (e.g. thresholding data by removing values above or below predefined points). In such cases we labeled these “families” of similar techniques based on their simplest shared trait. The output of this analysis was a list of visualizations coded for each technique that appeared.

Affinity diagramming was then used to arrive at higher-level clusters of techniques. As in the case of creating families of low-level techniques, we decided against a formal, mutually-exclusive scheme in favor of groupings based on similarities in the underlying mechanism. This strategy was chosen primarily because it yielded four distinguishable categories that we felt best covered our critical observations: information access rhetoric functioning to limit the amount of information presented, provenance rhetoric functioning to provide background information, mapping rhetoric functioning to map elements of the visualization to non-explicit concepts, and procedural rhetoric functioning to constrain interaction over time. One remaining cluster of techniques was not clearly distinguishable based on a common mechanism, but was rather comprised of

methods that instead appeared to cluster based on an origin in linguistic rhetoric. We then tabulated patterns of frequency and co-occurrence of techniques in order to show the interrelatedness of the categories. Alternative schemes of rhetorical techniques may be possible for narrative visualizations. However, the representativeness of our sample leads us to believe that the categories below can serve as a guide for designers seeking to strengthen or subdue rhetorical effects. Table III.I presents the editorial layers, forms of rhetoric, categories of design techniques, and specific strategies surfaced by the study.

Table III-I: Rhetorical techniques by editorial layer, rhetorical form, and category.

Editorial layer	Rhetorical form	Category	Techniques
Data	Information Access	Omission	<ul style="list-style-type: none"> • Not citing sources • Ambiguous definitions • Value, axis thresholding • Omitting outliers
Data		Metonymy	<ul style="list-style-type: none"> • Variable selection • Aggregation and categorization
Annotation	Provenance	Data provenance	<ul style="list-style-type: none"> • Citing/linking sources • Additional facts/references • Methodology citation • Exception annotation
Annotation, visual representation		Representing uncertainty	<ul style="list-style-type: none"> • Error bars • Describing inferential limits • Forecast annotation • Expressions of doubt
Annotation		Creator identification	<ul style="list-style-type: none"> • Author bio • Personal anecdote
Visual representation	Mapping	Obscuring	<ul style="list-style-type: none"> • Gratuitous 3rd dimension • Violating discriminability • False cause-and-effect • Double axes
Visual representation		Visual metaphor and metonymy	<ul style="list-style-type: none"> • Suggestive spatial mapping • Typographic, color mapping • Visual noise
Visual representation		Contrast	<ul style="list-style-type: none"> • Visual contrast • Variable splice
Annotation, visual representation		Classification	<ul style="list-style-type: none"> • Grouping by size, color, etc. • Consistent typographic mapping • Significance equations
Annotation, visual representation		Redundancy	<ul style="list-style-type: none"> • Disaggregation of homogenous values/marks
Annotation	Linguistic	Typographic emphases	<ul style="list-style-type: none"> • Italics and bolding
		Irony	<ul style="list-style-type: none"> • Rhetorical questions • Deliberate understatement • Quotation marks
		Similarity	<ul style="list-style-type: none"> • Analogy • Metaphoric statements • Parallelism • Simile • Double entendre
		Individualization	<ul style="list-style-type: none"> • Apostrophe

Interactivity

Procedural

Anchoring

- Default views, comparisons
- Spatial ordering
- Partial animation

INFORMATION ACCESS RHETORIC

The first decisions made by a visualization designer often concern what data to represent. To simplify complex ideas in a visual representation it is often helpful to keep distracting or irrelevant information to a minimum (e.g., Mayer et al. 2005). **Omission** techniques are the least likely to be explicitly indicated by a visualization, yet can be inferred from data that are available given ample contextual information. Assuming that most professional producers of online visualizations are aware of the importance of data provenance, *neglecting to cite data sources* or other important provenance information or *defining variables ambiguously* can be considered omissions. These may be motivated by *knowledge assumptions* of the end-user, such as when a complex statement is made without explicit reference to intermediate clauses. In *The Atlantic's* 'How the Recession Changed Us' (Figure III.I), the overall message about negative effects of the recession assumes that end-users intuit several non-explicit propositions in decoding the iconography and statistics. The number of times that the word 'uncertainty' appeared in the New York Times, for example, only makes sense in the graphic if one assumes that mentions of uncertainty in articles equates to economic-related risks and recession. Omissions may also result from a desire to simplify complex phenomena by excluding complicating information from the visual representation, as in the case of *thresholding values* or *omitting exceptional cases*. A visual representation occurs in *axis thresholding*, in which the values most important to communicate a pattern through comparison are used to set the range of the axis, so that higher or lower values that may be relevant but complicate the message are not shown.

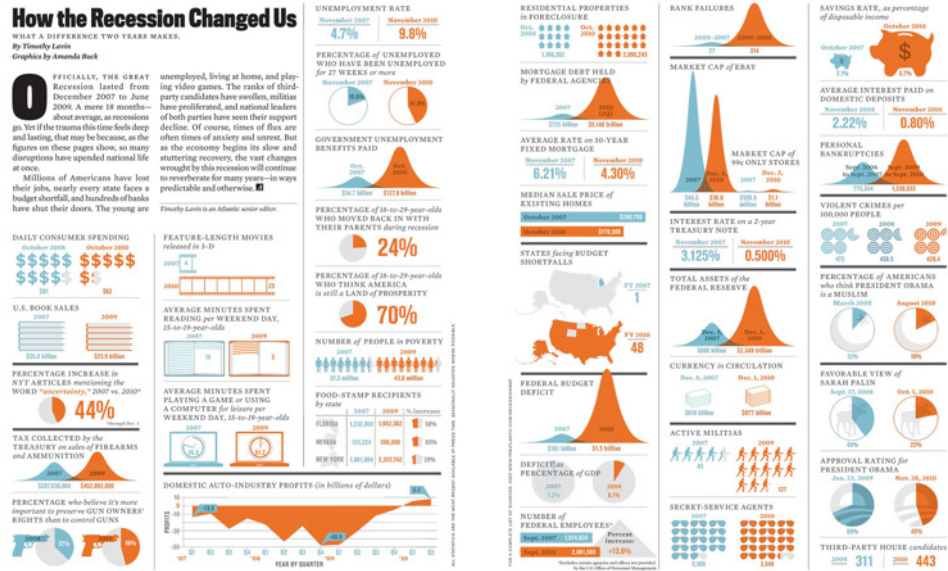


Figure III-I: “How the Recession Changed Us” presented by the Atlantic Monthly.

Omission or information loss choices can also be transferred to the end-user via *filtering* capabilities like search bars that allow a user to select a subset of data. Intentional information loss has been discussed on the part of the designer (Huff 1993, Tufte 2001, Ziemkiewicz and Kosara 2009), but has been underexplored from the perspective of user-driven filtering. The increasing prevalence of narrative visualization suggests that user-driven information loss or avoidance may be a fruitful area for research.

Metonymy techniques that manipulate part-whole relationships serve simplification as well. At the basest level, the *selection of variables* to visualize involves creating a subset of a larger data set to present a simplified visual representation of chosen features. *Averaging* techniques like mean, median, and clustering similarly substitute simpler representations for a wider range of values, as do *textual and visual summaries*. *Categorizing, binning, or aggregating* values can be used to make an intended effect more apparent. An Economist graph on car sales (Figure III.II) depicts only ‘light vehicles’ for some countries’ data, yet all sales for other countries.

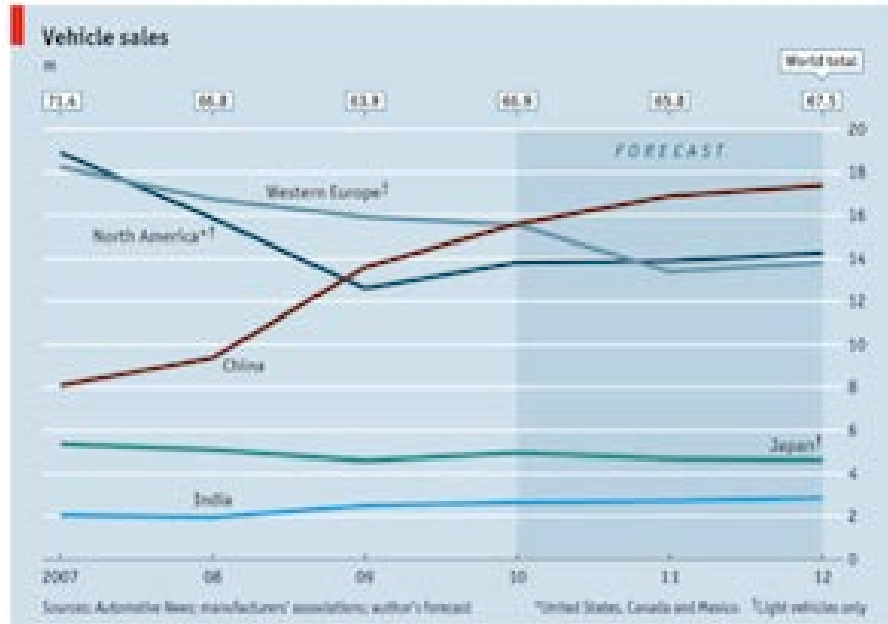


Figure III-II: "Vehicle Sales" visualization by the Economist Graphic Detail.

PROVENANCE RHETORIC

Similar to objectivity values in InfoVis, journalistic codes of ethics emphasize the journalist's duty to remain impartial and present information as clearly as possible (Kovach and Rosenstiel 2007). A number of visualization rhetoric techniques observed in our sample work to signal the transparency and trustworthiness of the presentation source to end-users. Doing so conveys a respect for the audience and reaffirms a journalist's public interest motive, strengthening the journalist's credibility (ibid.). **Data provenance** strategies include *citing and/or linking data sources, additional references, methodological choices, and relevant facts*, as well as *annotating exceptions and corrections*, thus achieving goals proposed by Tufte for graph provenance (Tufte 2006). Several of these methods are depicted at the bottom of Figure III.II.

Representing uncertainty can be accomplished through visual representations like *error bars*, yet appeared more often in our sample via textual means. These included *descriptions of inferential limits* (i.e. confidence intervals), *"leap-of-faith" or forecast annotations* explicitly labelling the point in a graph where data are extrapolated, or

expressions of doubt regarding potential conclusions (see Figure III.II, tag line below title). The dominance of textual uncertainty representations suggests an intriguing comparison between these visualizations and the visually-based ways of denoting uncertainty that have been developed in InfoVis and statistical graphics, such as error bars or confidence envelopes (e.g., Wainer 2009). The reliance on textual means may indicate a lack of adequate methods or commonly understood codes for visually representing uncertainty to non-experts (Skeel et al. 2009).

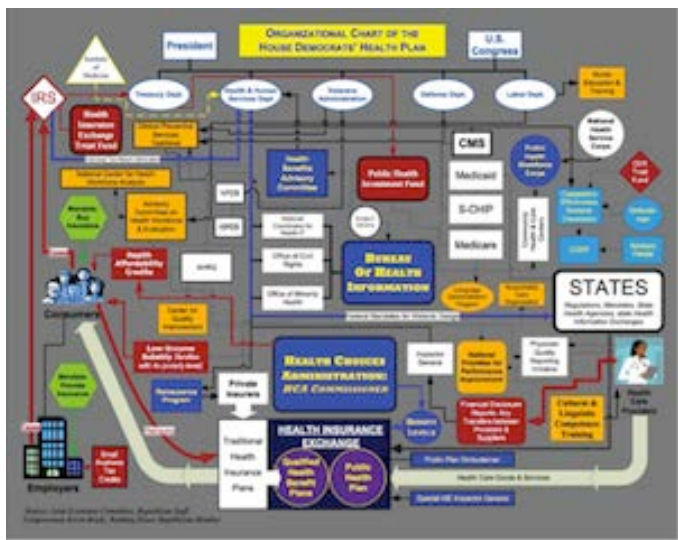
Finally, in some cases explicit steps are taken to signal the **identification** of a visualization's designer. While author-designers are usually credited for their work, in some cases additional information is provided, through *author bios* or *personal anecdotes*.

MAPPING RHETORIC

Mapping rhetoric refers to manipulating the information presentation via the data-to-visual transfer function, the constraints that determine how a piece of information will be translated to a visual feature. **Obscuring** can result from introducing “noise” into a representation, often on a perceptual level, such as in the case of adding a *gratuitous third dimension*. Other means of obscuring are applications of non-essential *sizing transformations* that violate *discriminability limits*. This may mean making some elements too small for judgment, *oversizing* to the point of overwhelming the presentation, or obscuring a value's true position on an axis. More subtly, non-intentional obscuring occurs when a designer *neglects to map information to the most salient visual judgment types* as suggested by work like (Cleveland and McGill 1984). Noise can be introduced on a semantic level, by implying *false cause-and-effect* relationships or by using complex design tactics like the *double-axis*, which experts have noted are difficult to decode even when properly used (Wainer 2009; see Figure III.II and Figure III.IX).

Visual metaphor and metonymy maps visual signs to non-present or implicit meanings. Some of these are interpreted automatically due to congruence with embodied

experience, such as *suggestive spatial mappings* like “left = past, right = future” or “up = more or better, down = less or bad” (Lakoff 1990). *Typographic mappings* and *color mappings* pair visualized patterns to categories via visualization components, such as by applying red and blue font colors representing political parties to statistics in an election-themed visualization. *Visual noise* is a visual metaphor technique that can also serve to obscure. It has become popular in recent years through visualizations like the visually confusing graphics by political party representatives of political parties to represent the “confused” policies of the opposing group (Figure III.III, top). Visual noise can be used more subtly as well, as in David McCandless’ ‘Poll Dancing’ visualization (Figure III.IX) or more obviously as in the ‘Democrats’ Health Plan’ graphic (Figure III.III, top), which prompted the response graph that appeared to be motivated in part by the goal of creating a distinctly non-noisy graph (Figure III.III, bottom).



Organizational Chart of the House Democrats' Health Plan

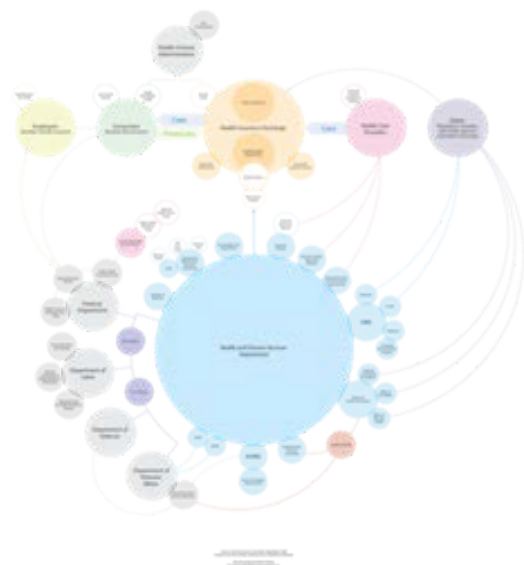


Figure III-III: "House Democrats' Health Plan" presented on John Boehner's website (top), and "Organizational Chart of the Democrats' Health Plan" created by graphic designer Robert Palmer.

Contrast techniques can serve ambiguity, as in the juxtaposition of oppositional pieces of information that occur in *visual contrasts* or *variable splices*. In these cases, information that is not obviously associated with target variables is included, adding an additional layer of perspective on an issue. An example can be found in the NYT interactive visualization entitled 'A Peek Into Netflix Cues' (Figure III.IV). The title and two variables of rental lists and movie rank variables are mapped to the important visual dimensions of spatial position and color. These mappings imply an overall message

organized around geographic patterns in top rentals. However, a choice was made to include the less obviously relevant critic meta-scores for each movie, along with a sample NYT review of each, to the left of the map frame. The result is an implication that this information may generate further insight through comparisons with the geographic patterns. Scanning comments attached to the visualization validates that such comparisons did occur among users.



Figure III-IV: "A Peek into Netflix Queues" by the New York Times graphics department.

Classification can be accomplished through *grouping by size, position, or color* (see Figure III.III, bottom). *Consistent typographic manipulations* of font sizes and styles and *equations of significance* presented in a legend-like format to highlight certain values can also classify information within a visualization. Such classifications can show clusters of priority or importance.

Redundancy techniques emphasize by *disaggregating homogenous values or visual marks*. The repetition of identical labels, or the disaggregation of values with little variance or similar functions or relationships between them, can be used both to emphasize as well as to create *visual noise*. In a second politically-themed graph from John Boehner's office on a new energy tax, a label of 'Higher prices' is used repeatedly in labels placed closed to

one another, presumably to emphasize the economic ramifications of the plan on taxpayers over combining the labels into one (Figure III.V). We note that the bijective or one-to-one mapping from the data to the target (visual) domain required in Ziemkiewicz and Kosara's (2009) taxonomy for information visualization is violated in nearly all occurrences of redundancy.



Figure III-V: "Speaker Pelosi's Energy Tax: A Bureaucratic Nightmare" presented by John Boehner's office website.

LINGUISTIC-BASED RHETORIC

Multiple techniques closely resembled rhetorical devices that derive from conventions of language usage. These techniques tended to be (but were not exclusively)

implemented at the textual layer, albeit with several exceptions. **Typographic emphases** like font *bolding* or *italicizing* derives meaning from conventions long associated with typography.

Irony is a basic literary and artistic strategy that sets up a discordance between the literal meanings of a statement and an alternative implied meaning. Visualizations in our sample often used *rhetorical questions* with irony, which has an effect of engaging the user's attention by directly addressing her, while at the same time using the question in order to imply its inverse. These tend to be used in titles to sarcastically set the stage for a user to arrive at an obvious interpretation. This is the case in 'Budget Forecasts, Compared With Reality' where a prominent textual annotation above the visualization poses the question "How accurate have past White House budget forecasts been?" despite numerous other annotations explicitly describing inaccuracies in forecasts (Figure III.VI). *Quotation marks* and *deliberate understatement* accomplish similar objectives.

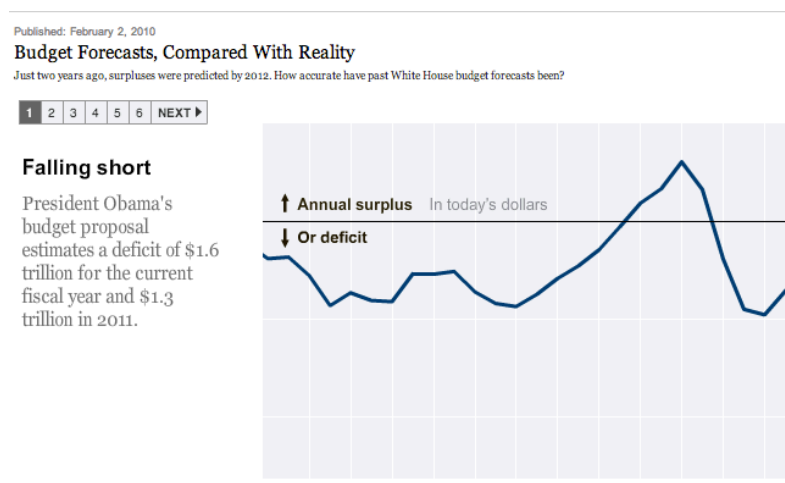


Figure III-VI: "Budget Forecasts, Compared to Reality" by the New York Times graphics department.

Similarity techniques resemble *contrast* techniques except that the comparison between two entities is motivated by assumed similarities between them. One method is *analogy*, in which a comparison is made in order to provide insight into the lesser known of two entities. *Metaphoric statements* equate two ideas or values by labelling or directly asserting that one *is* the other, as in the visualization titled 'Speaker Pelosi's National

Energy Tax: A Bureaucratic Nightmare' (Figure III.V). *Parallelism* involves expressing two linguistic statements or visual features to show that they are equal in importance. An example occurs in 'How the Recession Changed Us' (Figure III.I), through the juxtaposition of infographics of roughly the same size representing different data yet each framed around negative implications of the recession. *Simile* resembles analogy and parallelism but the goal tends to be for effect and emphasis of a similarity relationship. *Double entendre* hinges on a linguistic or visual similarity alone that is used to unite two ideas or entities. David McCandless' 'Poll Dancing' visualization (Figure III.IX) uses both, in the title and vertical visual format.

Finally, **individualization** techniques represent ways to directly address or appeal to the user as an individual. These techniques are similar to directly addressing a person using a second-person tense in language. This can increase interest and ease processing on the part of the user. *Apostrophe* is the direct address of the end-user in the title and annotations attached to a visualization, including rhetorical questions and suggested goals as mentioned above. More subtle means of individualization observed in our sample include providing alternative exploratory functions like *sorting and filtering methods* (Fig. 3.6) and *phrasing or imagery framed from an individual-citizen level view*, such as using people icons and phrasing like 'Buy Insurance' that is framed from the ordinary citizen view in the 'Organizational Chart of the House Democrats' Health Plan' (Figure III.III, top), in which labels like 'Higher Prices' that feature prominently across the top of the graph are framed sympathetic to the citizen tax-payers' perspective. Such techniques suggest that the user adopt a "Cartesian" cultural viewing code that privileges the individual (see Viewing Codes, below for further reflection on this).

PROCEDURAL RHETORIC

"Procedural rhetoric" is based in an artifact's procedural mode of representation, in other words, the expression of meanings through rule-based representations and

interactive functions (Bogost 2007). For instance, Diakopoulos et al. (2011) use procedural rhetoric in the form of game mechanics to drive attention in an interactive information graphic. The techniques we present here are similar to Segel and Heer's (2010) suggestions of interactivity features for storytelling in visualizations, yet are framed from the perspective of the editorial emphases and omissions they represent. This perspective opens them up for critical analyses of their rhetorical functions.

Anchoring techniques primarily direct a user's attention in a way that subsequently helps convey a message. *Default views* provide an initial point of interpretation anchored to the default visual configuration (e.g., Figure III.VI). *Fixed comparisons* present some information by default so that users can contrast this information with other values in the visualization. This can increase engagement via individualization when values suggested for comparisons are more likely to be salient to a user, such as in Figure III.VII, which presents several terms related to contested political issues that appeared in the speech transcript. Yet this technique also encourages a user to look for trends related to a particular data value over other potential comparisons in the larger data set. The fact that widely known methods for judging the 'visual significance' of a trend (as one might judge statistical significance) are lacking among most users becomes a particular risk. *Spatial ordering* leverages reading and scanning conventions to prioritize some information (Segel and Heer 2010). *Animations* leverage time to suggest a story, and *partial animation* that pauses or ends on particular views prioritizes through a "climactic" effect. More subtle means of anchoring include *search suggestions* or direct or implied *goal suggestions*, prompting the user to examine particular parts of the data rather than explore freely.

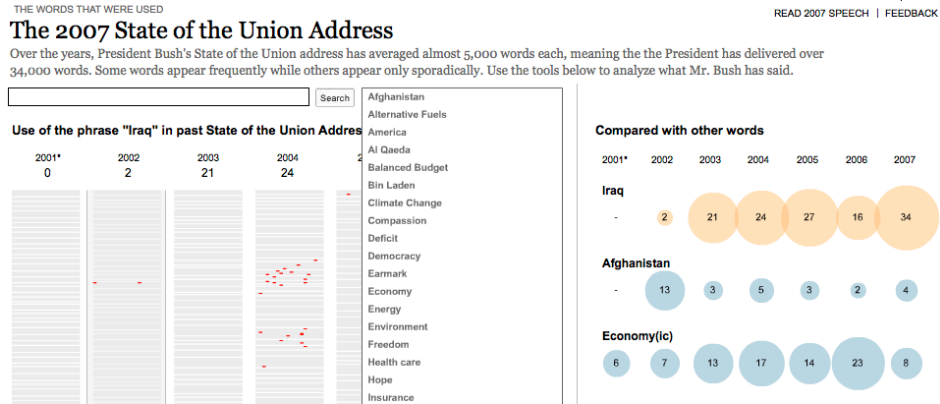


Figure III-VII: "The 2007 State of the Union Address" by the New York Times graphics department.

More explicitly interactive techniques include **filtering**, through search bars or menuing that constrain the data depiction based on a user's preferences for certain information (this also appears in individualization techniques). Search bars are likely to be effective in engaging a user to explore data based on how the personalization of information increases the salience of the message being presented (e.g., Skinner et al. 1994). Menu choices that appear by default can also help users find the most interesting comparisons or views in a visualization using the information gained by designers who have already thoroughly explored the data in the design process.

PATTERNS OF OCCURRENCE

While the output of our coding is indicative of the distribution of techniques found within our particular sample of narrative visualizations (i.e. many drawn from journalism outlets), a sample from other genres of visualization would likely produce a different distribution. Still, our results allowed comparisons of differences in the frequency of specific techniques, as well as co-occurrence trends. The top ten most prevalent techniques (ranked by frequency) were grouping by color, aggregating values, suggestive spatial mappings, goal suggestions, bolded fonts, data source citations, metaphoric statements, color mappings, apostrophe, and variable splices (Table III.II).

Table III-II: Top ten most frequently observed rhetorical design techniques in sample.

Rank	Technique
1	Color grouping
2	Value aggregation
3	Spatial mappings
4	Goal suggestions
5	Bold fonts
6	Data source citations
7	Metaphoric language
8	Color mappings
9	Apostrophe
10	Variable splice

A conclusion to be drawn from this ranking concerns the way that many of these techniques represent common strategies in a wide variety of data visualizations, based on their perceptual salience (e.g., spatial mappings, grouping by color) or their common use in other facets of communication (e.g., metaphoric statements). The fact that standard communication strategies can pave the way for potentially significant rhetorical effects may partially result from our observation that they often appeared in combination. A designer might opt to use many less obvious framing strategies to convey a visualization story, so as to reduce the appearance of bias that can result from extreme usage of a single strategy.

This ranking excludes several techniques that affected nearly all visualization, albeit to different degrees. These are variable selection, default views, knowledge assumptions, and visual contrasts. These naturally occur very frequently (e.g., an infinite number of variables cannot be visualized; a starting view for the visualization must be chosen; some knowledge must be assumed to communicate at all, such as a rudimentary ability to read charts; the goal of visualization is to compare data using vision). An insight to be gleaned from even these, however, arises when one considers that possible alternatives do exist, but appear to be unconventional. Choosing a default view, for example, may be unavoidable, but the choice of a single default view for all users is not a

given. Designers might dynamically choose default views in cases where the goal of the visualization is less specifically focused on a single intended interpretation. This particular implementation was not observed however.

Some techniques appeared together quite frequently. Data source citations tended to appear with other provenance techniques (i.e., methodology citations) more often than they appeared alone. While knowledge assumptions are on some level unavoidable, analogy, parallelism or other linguistic-based similarity techniques nearly always occurred with more extreme assumptions. An example is the title ‘The Arab Powder Keg’, which assumes that the user is familiar with the powder keg reference. Again, however, we note that this trend is not inevitable. A designer wishing to create a chart likely to be understood by the largest number of users could annotate the presentations with definitions in smaller type so as to include users without the requisite prior knowledge. Another notable pattern was the tendency for rhetorical questions to be used with implicit goal suggestions. In these cases, a question was posed that was most easily interpreted as ironic or pedantic in light of other annotations that directly instructed users to look for particular patterns.

A pronounced pattern throughout our analysis was the observation that the effectiveness of individual strategies depends on references to other layers of the presentation. This occurs despite the way that some categories are more closely associated with certain editorial layers (i.e., linguistic rhetoric mapping to annotations). A clear example is described below for the ‘Poll Dancing’ visualization (Figure III.IX), where a double-entendre in the title depends on several visual metaphors in the graph. This highlights the nature of narrative visualizations as multimedia artifacts that can’t easily be reduced to visualization alone.

VIEWING CODES

The concept of **viewing codes** is an adaption of theories presented in semiotics (e.g., Barthes 1978) that capture how attributes of the receiver of an artifact influence interpretation. Viewing codes are the cultural, perceptual, cognitive, and psychological lenses that guide how an end-user (or community) interprets a representation. This concept sheds light on the constraints imposed on end-user interpretations by habits and beliefs that are not explicitly contained in the visualization but rather implied by visualization elements. Below, we discuss how a distinction between **denotation** and **connotation** becomes important with regard to discussions of viewing codes.

In semiotic studies, codes are thought of as systems of related conventions, accumulated over time, that correlate signifiers, or symbols or representations, with signifieds, or meanings (Chandler 2001). In InfoVis, for example, the conventions that dictate what end-users expect to be communicated by given visualization formats are codes. Bar graphs, for example, are conventionally associated with discrete trends, while line graphs are associated with temporal trends. Prior experience with these graph types informs expectations when faced with a new graph. When non-temporal data are graphed in a line graph, users tend to frame their interpretations of the data using language associated with trends, such as “as a person gets taller they become more male” (Zacks and Tversky 1999).

Cultural codes describe the social norms and wider beliefs of a culture that a designer can target to suggest a particular interpretation. **Individual-level codes** can be higher-cognitive constraints (e.g., abilities) or more emotionally-based patterns of reaction. Empirical literature demonstrates how individual differences deriving from spatial intelligence (e.g., Carroll 1993) as well as prior knowledge can affect visualization interpretation (Conati and Maclaren 2008, Ziemkiewicz and Kosara 2009) and even bias perception (Henderson and Ferreira 2004). For example, individuals differ in their interests and prior knowledge regarding various types of news. Consequently, these

differences lead to differences in how users interpret the implications of the story in a narrative visualization.

Perceptual codes constrain what is salient to the user given human visual perception tendencies, such as gestalt principles of continuation, common fate, and closure (Wertheimer 1939). Perceptual tendencies can combine with internalized knowledge to form additional types of codes such as **textual codes**, the conventions associated with the presentation and interpretation of text. With regard to online information visualizations, these include the common positioning of the title either in the top center or top left of the presentation, the inclusion of source and designer credits toward the lower right or left hand corners of the layout, as well as the assumed left-to-right reading style in many Western cultures noted by Segel and Heer (2010). Similarly, **aesthetic codes** combine perceptual as well as shared yet subjective preferences for a particular style of presentation. In the tradition of visualization design that prioritizes high data-ink ratios, minimalist techniques such as colorless backgrounds and an avoidance of non-necessary ornamentation create a particular aesthetic code that can affect a user's judgment of the quality of a visualization.

A given element of a visualization-based presentation (whether textual, visual, or a combination) can activate individual or cultural viewing codes in several ways. *Denotation* refers to descriptive elements, including either textual or visual statements (such as iconography) that directly attribute features to objects. In the above example of users' differing expectations of bar versus line graphs, the height of the bars directly conveys the value for each bar's group for the y -axis variable (e.g., cost, score, or another quantity of interest). Likewise, the location of the points comprising the line directly conveys the value of the y -axis in the line graph. Users familiar with how to read a bar and line graph use this straightforward mapping to interpret the data. *Connotation*, however, refers to cases where a secondary symbol cues, but does not directly associate, a meaning. This

form of communication better describes why users of a bar graph are more likely to interpret the data as discrete rather than a temporal trend, while line graphs tend to evoke temporal interpretations regardless of the data (Zacks and Tversky 1999). Users have come to associate each graph type with particular data types (discrete categories and temporal trends), and the format itself activates the code of this expectation despite the lack of explicit reference.

Illustrating Visualization Rhetoric

Two case studies are used to demonstrate the kinds of insights that the visualization rhetoric framework provides into the interaction of specific design strategies, their communicative functions, and the extra-representational factors that constrain them. The first example, 'Mapping America: Every City, Every Block' highlights how the editorial layers described above can be used to convey meaning, and how specific techniques employed at these levels represent omissions and emphases of some data over others. The second example, 'Poll Dancing', demonstrates how viewing codes can be cued through design elements in practice, either through direct communication (denotation) or implicit suggestion (connotation).

'MAPPING AMERICA' VISUALIZATION

The New York Times

Mapping America: Every City, Every Block

Find something interesting? Share this view

Browse local data from the Census Bureau's American Community Survey, based on samples from 2005 to 2009. Because these figures are based on samples, they are subject to a margin of error, particularly in places with a low population, and are best regarded as estimates.

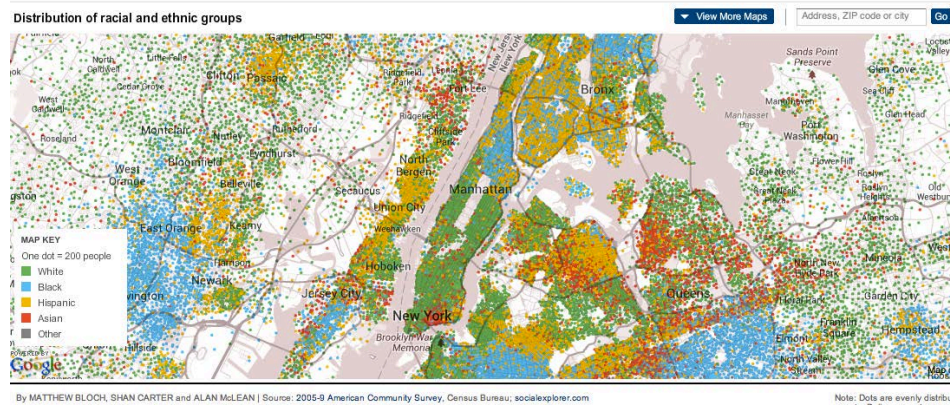


Figure III-VIII: "Mapping America: Every City, Every Block" by the New York Times graphics department.

The United States Census represents a nation-wide attempt to provide an objective view of the demographic distribution of the country. The New York Times Graphic Department's 'Mapping America: Every City, Every Block' interactive visualization depicts 2010 U.S. Census results. Rhetorical techniques are employed at the four different editorial layers of the visualization described above to convey the comprehensiveness of the data collection. At the level of the data, the choice to use actual census results rather than third-party summaries of the data conveys the truthfulness of the visualization as a non-biased depiction. The annotation layer communicates this choice. In this example, social annotations are provided in the form of comments in the right side bar that draw attention to important features and suggest conclusions based on the data. The annotation layer is also leveraged in this example for **data provenance** purposes, through a *methodology citation* behind the depiction as well as specific *data source citations*. The latter citations may betray *knowledge assumptions* on the part of the designer who wishes to appeal to a user's prior knowledge of the scope of the census data collection. In the context of visual journalism, such techniques shape users' interactions and interpretations by *signalling transparency* such that various beliefs associated with objective information

visualizations as a journalistic standard (Kovach and Rosenstiel 2007) are cued. Another annotation works as an **uncertainty representation** that conveys impartiality by referencing *the inferential limits* imposed by a margin of error. The *redundancy* in the title annotation phrasing, “Every City, Every Block” emphasizes the comprehensiveness to the portrayal. Similarly, techniques using the interactivity layer include a *default zoomed-out view* of all of New York City (the largest US city and presumed home of the default New York Times user) and additional zooming features for gaining an even more holistic view of the country. A *search bar* allows users to explore data for any US region using addresses, zip codes, or city names of personal significance to them. Together, these choices convey a sense that the visualization provides a relatively unobstructed presentation of all information necessary to decode the patterns inhering in the data. The depicted story of the spatial distribution of ethnic groups is further supported by consistent mappings, such as of groups to colors that are applied identically to data points in the multiple views.

Yet like any visualization, less impartial choices are evident as well. The choice to represent the families part of the ‘Housing and Families’ category with a single variable on ‘Same-Sex Couples’ represents an example of information access rhetoric through **metonymy**, as it omits other families like two parent or single person households. If additional data was available from the source but the designers excluded it, this choice can be read as an implicit suggestion to end-users that they are expected to find this information more interesting than other family-based variables. The visual representation carries further emphases on particular views of data. The choice of which variables are *mapped to salient pre-attentive channels* (Ware 2008) leads those variables to be more salient in the end-user’s interpretation. Here, the use of color leverages the pre-attentive qualities of this visual encoding channel to represent racial and ethnic groups, subtly privileging this information.

As described above, interactivity can be used to promote exploration of specific subsets of the wider range of available information, subtly privileging some information over other information. For example, an emphasis is put on the race and ethnicity information by a default view that **anchors** users' interpretations so that they are most likely to be formed based on this dimension of the data. By clicking on a 'View More Maps' button in the example, users are taken to a *menu* of additional choices, which enforce the priority of the Race and Ethnicity view by listing this first, making it more likely that users will interact with these views as a result of common navigational conventions. Exploring these additional variables reveals some *ambiguity in variable definitions*; the requirements for membership in the Race and Ethnicity categories of 'Foreign-born population' and 'Asian population' are not explained, leaving uncertainty as to what extent these groups overlap. While ambiguity techniques can function oppositely to omission techniques by providing a user with the possibility of several differing interpretations, they also omit more specific information such that a user is prevented from knowing with certainty whether her interpretation is supported. Faced with ambiguity, a user is able to choose for herself which definition or reading of a visualization element to assume. She may default to the definition that better supports an interpretation cued by her individual viewing codes, or unique knowledge and beliefs. This can work in favor of an intended interpretation on the part of the designer, such as in cases where providing the full unambiguous information might eliminate the plausibility of a highly engaging yet flawed interpretation.

'POLL DANCING' VISUALIZATION

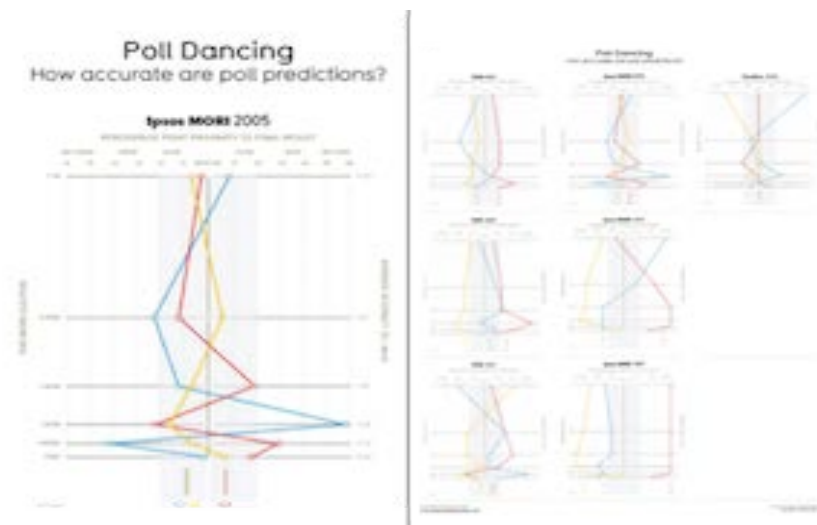


Figure III-IX: Partial (left) and full view of "Poll Dancing: How accurate are poll predictions?" by David McCandless.

A second example shows more clearly how extra-representational constraints can also significantly influence an end-user's interpretation. David McCandless' 'Poll Dancing: How accurate are poll predictions?' (Figure I.I) visualization summarizes the accuracy of political poll predictions from several years and polling agencies in a small multiples presentation of vertical line graphs. In each individual graph of one agency's predictions over a year, colored bars representing the political parties are drawn to connect data points positioned on the y-axis according to the amount of time prior to the election and on the x-axis according to whether the predictions fell over (to the right) or under (to the left) of a centered vertical line representing complete accuracy (or error of zero). Despite the apparent straightforwardness of the representation, analysis from a rhetorical standpoint provides insight into several layers of meaning implied as a result of design choices. Which of these alternate levels of meaning an individual user prioritizes depends on the **viewing codes** that constrain the interpretation, representing a second important insight that can be gained from rhetorical analysis. In the 'Poll Dancing' visualization, the framing of the poll predictions as 'dancing' in the title annotation lines brings to mind cultural associations with dancing as well as potential associations that stem from a user's

unique beliefs and knowledge about dancing. On a more basic level, the word ‘dancing’ combines with the juxtaposition of the visually-jagged line graphs in a *visual-linguistic metaphor*. Another type of **visual metaphor** is evident in that the variation, or directionality and distance to the center ‘accuracy’ line of the colored lines in the individual graphs, results in a *visual noise* effect. This effect is connected to the dancing association cued by the title based on a similarity between the *parallelism* inherent in the perceptual approximation of movement achieved by the jagged lines and the movement in dancing. In this case, the brightly-colored lines also naturally pop out against the muted grey and white background as a result of a perceptual codes. An *aesthetic code* that equates minimalism with representational impartiality may have motivated the colorless background and low contrast annotations.

Returning to the central metaphor, based on her prior experience and associations with political poll predictions, a user might interpret the association drawn between political poll predictions and the act of dancing as a light-hearted presentational technique that does not necessarily comment on the value of political poll predictions. On the other hand, a user with a more skeptical prior orientation to poll predictions might interpret the dancing connection as implying a frivolous or amusing aspect that suggests the results should not be taken seriously. Hence, differences internalized in *individual codes* can significantly alter the message an end-user interprets.

Another possible level of meaning can also be inferred given the specific design elements and consideration of additional associations that might be created by the title and visual representation. The title ‘Poll Dancing’ implicitly *connotes* the identically-pronounced term ‘pole dancing’, referring to a form of entertainment and exercise that traditionally takes place in strip clubs. As such, a second form of metaphorical substitution, *double-entendre*, is used to cue a double-meaning to any users who are aware of the existence and term for ‘pole dancing’ in English. This meaning may gain further

support through another **visual metaphor** cued by the choice to orient the line graphs vertically and to center the colored lines around the straight vertical line representing zero error. Users familiar with pole dancing may associate this vertical line with the pole that a pole dancer orients her movement around. This *connotation*, if cued in an end-user with a negative association with ‘pole dancing’ deriving from cultural stereotypes associated with the activity, might lead to an interpretation of the visualization’s message as an even stronger value judgment on the worth of political poll prediction. This results from the way these negative associations with pole dancing are metaphorically transferred to political poll predictions.

Interestingly, *connotation* as that described above depends on *denotational* communication of meaning, as the denoted signs are used in connotation to imply a non-present meaning (Chandler 2001). In the above example, the implication of pole dancing achieved by the vertical representation of the central “pole” relies on the same element that plays a directly descriptive role by representing the zero point (or accurate prediction).

Discussion

The study of narrative visualizations offers an opportunity for increasing understanding of the complementary relationship between explorative and communicative dimensions in InfoVis. We suggest several important considerations for this space highlighted by our analysis, and note areas that may be fruitful for future exploration.

The effects of subtle rhetorical manipulation of information has generated sometimes surprising results in decision theory and political and communication studies. Applying a similar experimental approach to narrative visualizations is a natural parallel. Our work sets the stage for such studies by providing a taxonomy of specific information presentation manipulations used in narrative visualizations. Formal models that have

been developed to capture the formation of user opinions as dependent on personal attitudes (Saris and Sneiderman 2004) similarly motivate future modelling of combined effects of rhetorical techniques and personal and cultural viewing codes on a user's interpretation in narrative visualization.

Acknowledging the distinction between denotation and connotation contributes to InfoVis design and theory by highlighting an epistemological tension that invades many narrative visualizations. This trade-off is between techniques of "objective" charts informed by transparency ideals on the one hand, and the layers of connoted interpretation that can seep into or co-opt the basis of objectivity via rhetorical strategies on the other. The 'Poll Dancing' example leverages the visual representation to precisely depict trends in forecasting. At the same time, connoted meanings imply that poll predictions may be best characterized as "entertaining" rather than rigorous or scientific. The fact that both modes are possible within the same space may explain why such visualizations are engaging in ways that is difficult for numeric representations alone to achieve. The intriguing tension or interplay that results from combining seemingly oppositional techniques may help explain how rhetoric can exert a positive influence in visualizations. Future work includes devising means of assessing narrative visualizations such that these positive influences are recognized, while still acknowledging the potential for rhetorical decisions to negatively affect a user's accurate interpretation of data.

A frequent example of such a productive tension in our sample is that some narrative visualizations appear to be concerned with presenting their work as credible even in cases where the journalist may have taken some liberties in preparing the graphic. This is likely the influence of journalistic notions of transparency, where creators are expected to be upfront about their knowledge as well as what they don't know (Kovach and Rosenstiel 2007). In many examples, the journalist's presence is explicitly stated, such as through notes about how a visualization contains 'predictions' or 'forecasts' at the bottom of the graph (see Figure III.II). These acknowledgements may play a double role in

the sense that they strengthen the sense of the journalist's or designer's integrity despite explicitly pointing to a lack thereof. This observation dovetails with the observation that codes or conventions appear to operate in narrative visualizations. Not only do transparency clues suggest that an end-user should believe the specific interpretation being emphasized in the visualization, they also implicitly suggest to users a preferred way of making similar decisions when viewing other visualizations. Insight from critical media and semiotic studies suggests that such codes are dynamic systems that change over time (Chandler 2001). Many professionally produced narrative visualizations form part of a larger system of meaning and rhetoric, knowledge of which guides an informed user on how to interpret the particular example. By giving more attention to the development, maintenance, and propagation of such conventions in information visualization, researchers and designers alike stand to gain control over dimensions of interpretation that have remained mostly unaccounted for or underexplored.

A related discussion prompted by this chapter concerns the degree of intentionality that can be assumed behind the rhetorical effects achieved in narrative visualization. In analysis we noted all possible, although not necessarily intended, framing effects of design choices. We have demonstrated how particular trade-offs, such as between transparent presentation and persuasive framing of a message, appear to be evidenced by visualizations in our sample. However, future studies would do well to connect insights arrived at through inference from design outcomes (also including Segel and Heer's work) with more direct observation of designers' processes. Interview methods or participant observation can provide insight into the extent to which visualization creators are cognizant of rhetoric and related trade-offs. In any case, the power of rhetorical techniques to manipulate user interpretations calls for greater consideration of the responsibility that designers have to consider the possibly unintended effects their choices. Pre-existing design methodologies, such as scenario creation, could be adapted to support considerations of how elements of a viewing code might shape interpretations in a

particular direction. Similarly, greater reflection on ambiguous design elements could provide insight into how these aspects contribute to multiple plausible interpretations.

Finally, our analysis concentrated on professionally designed visualizations, yet a larger number of visualizations are created by non-professionals. A specific aim for future work concerns the possibility for integrating rhetorical and communicative features into existing visualization tools that are accessible to the public online, including collaborative visualization systems (Chinchor and Pike 2009, Thomas and Cook 2005). Introducing knowledge around rhetorical features and trade-offs could allow analysts and other data workers who lack professional design training to better achieve their communication goals.

Conclusion

The rhetorically-based framework presented here summarizes many ways in which narrative visualizations can be used to prioritize certain interpretations of data through design. Within the framework we have presented, it becomes clearer how trade-offs related to the persuasive dimension of design are likely to impact end-user interpretations. The concept of viewing codes, comprised of the many attributes of the visualization user and his or her context of use on various levels, is proposed as an additional set of constraints mediating interpretation. These themes demonstrate the necessity for models of interpretation in the development of design support tools, a theme that will be explored further in the subsequent chapter.

CHAPTER IV

A Deeper Understanding of Sequence in Narrative Visualization

In using visualizations to tell a story, the events of interest are patterns in data sets represented in visualizations, and the “presentation style” for these visualized patterns is the graphical format and the retinal mappings (such as the assignment of data values to locations, colors, or other visual variables). Chapter III characterized the creation of narrative visualizations as a series of editorial choices. Design decisions are made sequentially, often between alternative possible alternatives, in order to define the context for the visualization, select the most appropriate information and modality for a storytelling goal, and to select a sequence with which to narrate or render information.

Chapter III’s emphasizes the effects that even “minor” design choices in the editorial process, like posing a rhetorical question, can have on users’ interpretations of data. This inspires a deeper investigation of one particular set of decisions that are likely to constrain narrative visualization interpretation: those choices about how to use presentation order effectively to convey a message about data.

In what is likely to be a ‘typical’ creation process, an individual uses a tool like Tableau or Microsoft Excel to visually analyze data, and to generate visualizations via vector graphics or images for presentation. The individual must decide how to thread the representations into a compelling yet understandable sequence. Chapter III comments on this final stage of data story creation by introducing “procedural rhetoric,” referring to design decisions that concern information rendering by dictating the range of interactivity that a visualization provides end-users. Techniques observed in Chapter III’s study

included default views and partial animation. Elsewhere, Segel and Heer (2010) note structuring techniques like the Drill-down story, which similarly concerns the presentation structure and sequencing of the visualized content.

Comparable textual techniques for structuring evidence, combined with the choice of appropriate rhetorical strategies, are referred to as “the art of storytelling” among literary scholars. Evidence from cognitive psychology suggests that structural aspects, including the sequence in which information is delivered, play an important role in effective storytelling. Whether trial evidence or fictional narratives, the sequencing and forms of grouping used in a narrative affect the meaning that is constructed, the judgments that are consequently made by the audience (Pennington and Hastie 1992) and the ability to recall the information later (Thorndyke 1977). Yet much is still to be learned about the principles that govern effective structuring of transitions between consecutive visualizations in narrative presentations, and how different tactics for sequencing visualizations are combined into global strategies in common formats like slideshow presentations.

A gap also exists in current understanding around how end-users’ perceptions are affected by sequencing choices in narrative visualization. What characteristics make a sequence of visualizations successful in the eyes of users, as well as the designer? Other challenges stem from the potential conflict between narrative or sequence (“global”) design considerations and design optimizations framed around singular visualizations (“local”) that are included in many current systems. The Tableau visualization system, for example, relies on canonical automated visualization presentation techniques which model the effectiveness of different data-to-visual mappings for graphical perception as well as the faithfulness with which they express desired relationships (Mackinlay et al. 2007, Mackinlay 1986). The risk occurs when the individual visualizations in a set, which might each be best represented with a distinct graphical format and retinal variable mappings, introduce complexities to interpretation as a set. The task to the end-user who

must interpret the set is likely to be complicated by applications of the same retinal value to different data values (see Figure I-VII, reproduced again below), for example. Equally difficult to interpret are visualizations that map the same data variable in two different but overlapping ways (e.g., overlapping qualitative color range).

In using visualizations to tell a story, the events of interest are patterns in data sets represented in visualizations, and the “presentation style” for these visualized patterns is the graphical format and the retinal mappings (such as the assignment of data values to locations, colors, or other visual variables). Chapter III characterized the creation of narrative visualizations as a series of editorial choices. Design decisions are made sequentially, often between alternative possible alternatives, in order to define the context for the visualization, select the most appropriate information and modality for a storytelling goal, and to select a sequence with which to narrate or render information.

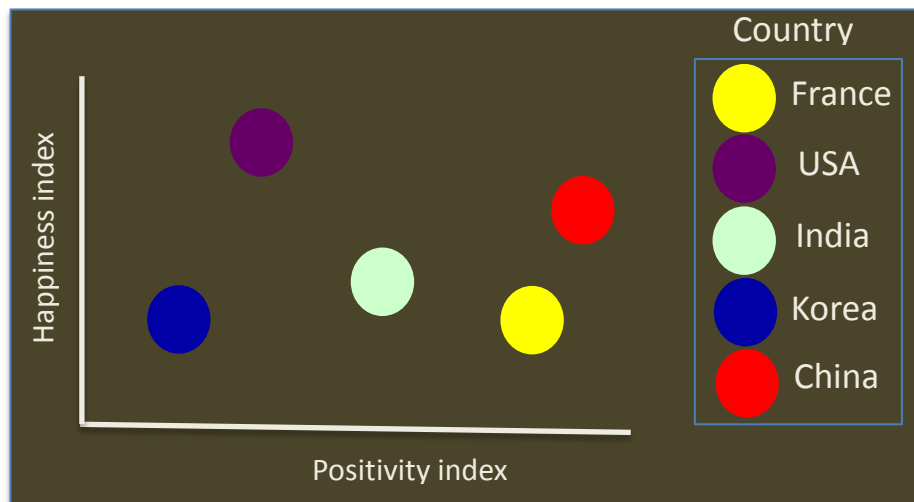
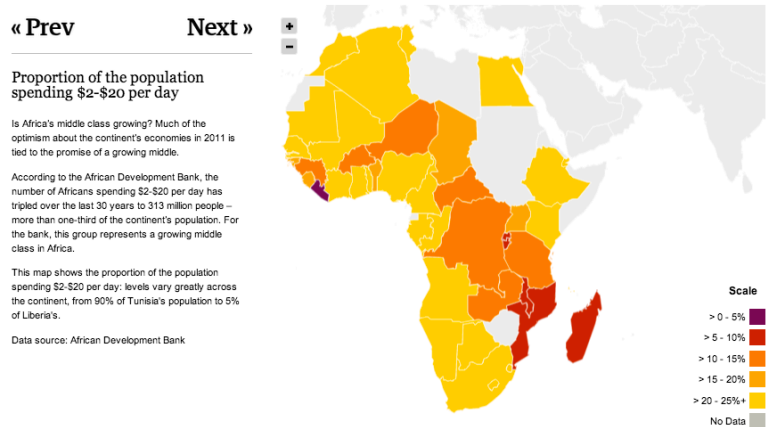
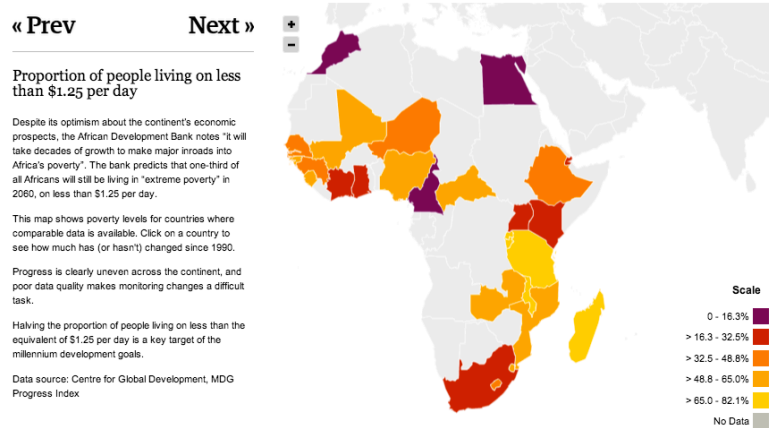
Chapter III’s emphasizes the effects that even “minor” design choices in the editorial process, like posing a rhetorical question, can have on users’ interpretations of data. This inspires a deeper investigation of one particular set of decisions that are likely to constrain narrative visualization interpretation: those choices about how to use presentation order effectively to convey a message about data.

In what is likely to be a ‘typical’ creation process, an individual uses a tool like Tableau or Microsoft Excel to visually analyze data, and to generate visualizations via vector graphics or images for presentation. The individual must decide how to thread the representations into a compelling yet understandable sequence. Chapter III comments on this final stage of data story creation by introducing “procedural rhetoric,” referring to design decisions that concern information rendering by dictating the range of interactivity that a visualization provides end-users. Techniques observed in Chapter III’s study included default views and partial animation. Elsewhere, Segel and Heer (2010) note structuring techniques like the Drill-down story, which similarly concerns the presentation structure and sequencing of the visualized content.

Comparable textual techniques for structuring evidence, combined with the choice of appropriate rhetorical strategies, are referred to as “the art of storytelling” among literary scholars. Evidence from cognitive psychology suggests that structural aspects, including the sequence in which information is delivered, play an important role in effective storytelling. Whether trial evidence or fictional narratives, the sequencing and forms of grouping used in a narrative affect the meaning that is constructed, the judgments that are consequently made by the audience (Pennington and Hastie 1992) and the ability to recall the information later (Thorndyke 1977). Yet much is still to be learned about the principles that govern effective structuring of transitions between consecutive visualizations in narrative presentations, and how different tactics for sequencing visualizations are combined into global strategies in common formats like slideshow presentations.

A gap also exists in current understanding around how end-users’ perceptions are affected by sequencing choices in narrative visualization. What characteristics make a sequence of visualizations successful in the eyes of users, as well as the designer? Other challenges stem from the potential conflict between narrative or sequence (“global”) design considerations and design optimizations framed around singular visualizations (“local”) that are included in many current systems. The Tableau visualization system, for example, relies on canonical automated visualization presentation techniques which model the effectiveness of different data-to-visual mappings for graphical perception as well as the faithfulness with which they express desired relationships (Mackinlay et al. 2007, Mackinlay 1986). The risk occurs when the individual visualizations in a set, which might each be best represented with a distinct graphical format and retinal variable mappings, introduce complexities to interpretation as a set. The task to the end-user who must interpret the set is likely to be complicated by applications of the same retinal value to different data values (see Figure I.VII, reproduced again below), for example. Equally

difficult to interpret are visualizations that map the same data variable in two different but overlapping ways (e.g., overlapping qualitative color range).



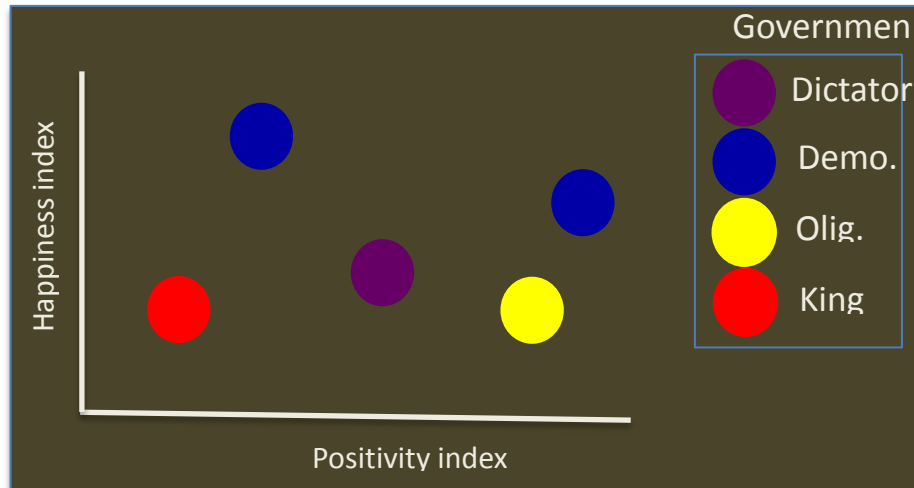


Figure IV-I: Two scatterplots apply the same colors to two different nominal variables.

With the popularity of narrative visualization among individuals who may lack design or statistical expertise yet have important domain knowledge to contribute, a deeper understanding of sequence could pave the way for tools and systems that support more effective story structuring. Creators might be business analysts, marketing professionals, small business owners, or any number of other occupations in which data plays an important role. Or, they may be web users who are passionate about data and particular causes (e.g., politics, environmentalism, etc.). These users are creating data graphics using the template-based GUI's included in reporting tools like spreadsheet applications (e.g., Microsoft Excel), more advanced domain-general visualization tools (e.g., Tableau) or even presentation software (e.g., Microsoft PowerPoint). These visualizations reach even greater numbers of individuals who might be other stakeholders in the business or industry, or online audiences who encounter the data graphics in blog posts or social media.

We focus in particular on how a particular subset of popular data visualization use—linear, slideshow-style presentations—can benefit from knowledge on the effects of sequencing styles on user perceptions and message communication. These may include slideshows based on series of data representations for live presentation as well as

interactive visualization slideshows presented online. A central contribution of our work is to outline an approach for automatic sequencing support that could help non-designers make structuring decisions in creating narrative visualizations. This includes a novel algorithm for semi-automatically identifying and presenting more “effective” visualization sequences during a design session. In outlining a sequence-based algorithm, we motivate the need for specific measures for negotiating conflicts between global and local automated presentation techniques.

First, to gain empirical knowledge on the forms that structure and sequence take in narrative visualization, we conducted a qualitative analysis of 42 professional narrative visualizations. Our results inform a graph-driven approach that identifies possible transitions in a visualization set (represented as nodes in a graph) and prioritizes visualization-to-visualization transitions (represented as weighted links) based on an objective function that minimizes the cost of transitions from the audience perspective. We conducted two large studies to validate this function as well as to expand our approach with additional knowledge of user preferences for different types of local transitions and the effects of global sequencing strategies on memory, preference, and comprehension. Our results include a relative ranking of types of visualization transitions by the audience perspective and support for memory and subjective rating benefits of visualization sequences that use parallelism as a structural device. We use additional design examples to motivate the need for specific means of addressing conflicts with local visualization optimization algorithms (e.g., Mackinlay 1986). We conclude by discussing the implications of our findings for the design of linear-style narrative visualization presentations and tools to support non-designers in creating narrative visualizations.

Related Work

NARRATIVE SEQUENCE AND STYLING

Our work is motivated by the systematic analysis of narrative in cognitive psychology. Researchers have empirically demonstrated that stories are perceived as being made of conceptually-separable episodes or sub-goals in a chain of actions that form the story's plot (Black and Bower 1979). Stories are thought to contain microstructure via the particular details of an event and macro-structure via the relationship of those events to one another in the plot (e.g., Thorndyke 1977). We make an analogy between story episodes and visualization states in narrative visualizations, which must also be sequenced to form a larger presentation.

Many psychological theories of narrative are grounded in experiments showing the importance of structure and sequence to story reception. Studies have shown that subjects are sensitive to suprasentential, or between-sentence, structure in a narrative, and use it to guide comprehension and recall. Such experiments typically test subjects understanding and recall for “scrambled” or randomly sequenced stories in comparison to those presented in “normal” order (e.g., temporal sequencing or groupings by causal implications (Thorndyke 1977).

Our global sequencing patterns study takes motivation from this approach. Pennington and Hastie (1992) show that grouping court evidence by sub-stories leads to more confident and unanimous decisions among jurors over evidence that is presented haphazardly (e.g., with grouping based on motives rather than temporal proximity). These results may be due to story understanding being a constructive process in which audience members summon up explanations so as to choose between decision alternatives (see also Wilensky 1980). While these authors assume a “correct” story, our approach takes a more conservative stance by assuming that more than one compelling sequence may be effective to narrate a set of visualizations. Yet just as jurors in a trial must learn and choose among

decision alternatives in order to generate the most likely story, creators of narrative visualizations must infer viable transitions between visualizations and make judgments about which are most persuasive to use in a story. By inquiring into transition principles and how end-users react to them, we intend to support this aspect of the story creation.

NARRATIVE VISUALIZATION RESEARCH

Existing research toward widely supporting creation of narrative visualizations includes systems for visualizing and sharing public and personally relevant data (e.g., Viegas et al. 2007) supporting new interaction styles from rich media artifacts (e.g., Rich Interactive Narratives) and design space taxonomies that describe techniques used in exemplar professional artifacts (e.g., Segel and Heer 2010, Chapter III). The latter studies provide generalized advice for designing narrative visualizations gained from professional examples. In addition to noting narrative formats that appear in interactive narrative visualizations such as the interactive slideshow (Segel and Heer 2010), these studies describe how prioritization and sequencing of information can occur through spatial ordering, animation, and suggestive default views, among others (Segel and Heer 2010, Chapter III). In earlier work, Gershon and Page (2001) describe how storytelling could benefit information visualization. They describe how animation could be used to present time-based information in sequence. They also propose more generally how a series of dynamic visualization views can be arranged as sequential frames, where each view is designed to display small enough amounts of information so as not to cognitively overwhelm a user. Yet, despite giving examples of structuring techniques that professional designers have refined, there is a lack of clearly outlined metrics that creators can use to find the best sequence for visualizations among multiple possible sequences. We expand prior work in narrative visualization via an understanding of sequence informed by empirical analysis of professional visualizations and user validated metrics and transition characterizations.

Prior work on visualization transitions includes Heer and Robertson's (2007) study of animated transitions in statistical graphics. Though they focused primarily on the effect of animation and staging of transitions taken as given, we note parallels between our principle of maintaining consistency and the guidelines they propose. The taxonomy of transition types we identify in professional narrative visualizations offer an end-user perspective of conceptually-based transitions (i.e., changes to the data being shown), providing a counterpoint to the types that Heer and Robertson define from a system representation of schematic and syntactic operations applied to data. We expand on their observations of transitions based in timesteps, filtering, and data schema changes, elaborating how users perceive these and other conceptual changes that occur in transitions.

THE ROLE OF ALTERNATIVES IN DESIGN

Our intention to inform the design of tools for supporting narrative visualization creation is motivated by design research demonstrating the importance of exploration of alternative designs among creators. Researchers like Duncker (1945) have shown that individuals often fixate on a single or narrow range of potential solutions early in a design process. Studies of successful design processes, however, indicate that generating and considering alternatives supports better understanding of the design specification: constraints and guidelines that are not in the initial specification but which help dictate what makes for a desirable design (Kolodner and Wills 1993). These insights have been applied most recently in ad design studies that find that parallel prototyping techniques that involve early generation of diverse examples produce better quality designs than techniques based in iteration and refining of a single design (Dow et al. 2011). We note that the time constraints operating on creators of narrative visualization presentations like data slideshows make it unlikely that all possible sequencings for telling a given story from a visualization set will be explored. The risk is that the creator uses a less compelling

sequence than they might. Having a better understanding what drives sequencing choices in narrative visualization, and a user-validated approach for algorithmically identify and prioritizing possible sequences is one way to work towards supporting exploration in the narrative visualization design process.

AUTOMATED VISUALIZATION PRESENTATION

The canonical approach to providing automatic support for generating effective visualizations can be found in systems like Tableau (tableausoftware.com), nee Polaris (Stolte et al. 2002). Here, a “Show Me” technique (Mackinlay et al. 2007, Figure IV.II) combines a table algebra with embedded models based on best practices to automatically identify and suggest to a user the best visualization format for representing a selected set of relational data. As proposed by Mackinlay (1986), the best graphical design is that which “expresses a set of relations and their structural properties effectively.” The approach, which can be framed as a constraint satisfaction problem, relies on “expressiveness criteria” constrain the search space of possible graphical designs by identify the set of graphical languages that can express desired information without contributing confounds (e.g., avoid encoding nominal variables with saturation as users will falsely interpret the variable levels as ranked). These are combined with “effectiveness criteria” that use information on the accuracy of perception given different mark properties to determine whether a graphical language exploits capabilities of the output medium and human perception. These criteria rely on prior results on the accuracy of interpreting numerical information depending on what retinal variable is used to encode the data (such as position, size, or angle) (Cleveland and McGill 1984). While many visualization systems offer some ability to reuse commands on new visualizations, which can help maintain consistency in encodings across related visualizations, more focused support for creating an effective set of visualizations for presentation is lacking. Later in this chapter, we expand the canonical approach to automated presentation to integrate

sequence considerations by extending the expressiveness and effectiveness criteria to encompass visualization sets.



Figure IV-II: "Show Me" dialogue provides example views for a user to choose from.

A more recent addition to the literature on automatically creating and evaluating the effectiveness of a visualization is a “visual embedding” model (Demiralp et al. 2014). These are functions from data points to a space of visual primitives that measurably preserve data structure in the perceptual mapping. Structure is preserved when pairwise distances between data points are preserved in the visual mapping. In either case, the end result is a visualization designed to support accurate communication of data and comparisons between data values within the visualization. In proposing a sequence support algorithm that can be integrated with the canonical APT approach, we rely on pairwise comparisons between data values within a single graphic similarly to the visual embeddings approach as a way to model the local consequences of globally (or sequence) motivated changes to the design of a data graphic.

Our work is also relevant to graph-based approaches that make it possible to transfer prior actions used in creating visualizations to automatically generate new visualizations during the same session (e.g., Scheidegger et al. 2007, Bavoli et al. 2005, Jankun-Kelly et al. 2007). These approaches focus on comparing representations of different graphics that results from iterative processes (or pipeline). By modeling pipelines as directed graphs where nodes represent particular visualization states, it is possible to compute the difference between two analogous visualizations (such as those that share many features but a few differences), and then apply the difference operation to a third visualization with similar properties to one of the original visualizations (Scheidegger et al. 2007). Our approach similarly relies on difference operations applied to pairs of visualization specifications that represent unique visualization states created during a use session. However, we are interested not in learning how a difference operation can be applied to a new visualization to save time, but instead how more effective narrative visualization presentations can be created given support for modeling effectiveness at the level of a set of visualizations.

Study: Patterns in Narrative Visualization Sequence

MOTIVATING SCENARIO

Many of the notable narrative visualizations pointed to by researchers are created by professional designers who draw on advanced training in journalism, graphic design, statistics, and other relevant fields to create compelling presentations (e.g., Hullman and Diakopoulos 2011, Segel and Heer 2010). Yet in numerous scenarios, non-designers create presentations from visualized data for the purpose of communicating a narrative of interest to a stakeholder or group. A marketing analyst or other data consultant may present clients with data presentations that describe the state of the market for a product, or the results of a change made to the client business strategy, product, or website. In many such cases, these individuals must first make sense of data themselves to distil

important points for a presentation, capture these points in data representations like visualizations, and then sequence these representations in a linear presentation. In this chapter, we consider the latter stage in this process, namely the act of sequencing selected visualizations. When the creator lacks design training, this can be a time-consuming trial-and-error process.

We argue that analysts using narrated data presentations could be helped by tools for identifying effective sequences for visualizations. Considering alternative paths through a set of visualizations is likely to enable a more compelling final artifact based on the importance of design alternatives in creation (Kolodner and Wills 1993, Dow et al. 2011). In the following section we describe an analysis of professional narrative visualizations that we used in order to identify what makes a good sequence. Our observations inform an algorithmic approach to identifying sequences introduced later in this chapter.

QUALITATIVE ANALYSIS

To inform the design of a tool that suggests good story structures with insights on the strategies of professional designers, we conducted a qualitative analysis of the structural aspects of 42 examples of explicitly-guided (i.e., unambiguously linearly ordered) professional narrative visualizations. The study poses several questions about sequencing in professional narrative visualization presentations:

- What types of changes (transition types) drive between-visualization transitions in linear narrative visualizations?
- Are there general characteristics that are shared among the common types of transitions?
- How do strategies for *local* (visualization-to-visualization) transitions compare to *global* transitions (patterns involving multiple local transitions)?

STUDY DESIGN

Forty-two visualization-based stories were compiled, starting from visualizations in an independently-curated sample of New York Times and Guardian interactives (Rooze 2011). These were supplemented with examples from visualization blogs and repositories (e.g., visualizing.org) and other well-known news sources (e.g., BBC) that are looked to as sources for high quality visualization presentations. We included only visualizations with non-ambiguous sequencing cues like numbered slides or steps, a “Next,” “→,” or “Continue” button, or a “Play” button for a self-running video or slideshow. In other words, we looked for visualizations where these features occurred the presence of additional navigational choices. While interactive slideshows intended for online presentation formed the largest format in our sample (23/42), other presentation formats included animated data videos (7/42), animated interactive timelines (6/42), live narrated visualization presentations (1/42), and static slideshows archived online but originally intended for live presentation (5/42).

While the individual “states” that comprise a visualization sequence is relatively unambiguous in a slideshow-style presentation, smooth animated narrative visualizations are more difficult to break down into their constituent visualization states. A visualization state has previously been defined as a set of parameters applied to data (Jankun-Kelly et al. 2007) or the settings of interface widgets in a visualization environment along with the application content (Heer et al. 2008). We draw from these definitions to define a “narrative visualization state” as an informationally-distinct visual representation in a presentation. Our definition of a state does not consider different portions of a single static visualization to be unique states. Though static visualizations are likely to be processed sequentially (such as if labels suggest that users examine data in a particular order), coding these would require more arbitrary judgments on how to divide static graphs. While a slideshow composed of unique static slides typically divides into one story unit per slide, a single slide could represent multiple units if it contains animation within

single slides. Rather than counting the number of states in a smooth animation, our interest is in noting changes from one transition form to another. For instance, we are interested in when a series of chronological transitions showing population estimates for different time slices (possibly spanning many states) changes to another transition form. The time-based transition sequence might give way to a transition where the measure or measure changes to GDP per capita while time stays constant.

Coding proceeded as follows: two coders first informally analyzed visualizations in the set with a focus on those aspects of the presentations that suggested how consecutive states in a data story are prioritized or ordered. Over several iterations, various categories of state-to-state order emerged. A coding protocol that captured these aspects was created and discussed by both coders. Visual interaction strategies that appeared relevant to sequencing, such as animated transitions between states, were also noted. Ten visualizations were randomly drawn from the set and coded independently by both coders, and the protocol updated upon reconciliation of disagreements. The remaining visualizations were then coded independently.

Additionally, we analyzed “global” structuring tactics spanning longer sequences of visualizations in a presentation. Coding first at the local level of visualization-to-visualization transitions allowed us to work up to observations at a global presentational level in a final collaborative coding. This entailed reviewing the combinations of transitions that occurred in each presentation to note patterns indicating global sequencing strategies.

Table IV-I: Transition types and sample prevalence.

Category	Transition Types	Sample Frequency	Total
Dialogue	Question & Answer	(4/42)	16.7%
	Who, What, When, Where, Why, How	(3/42)	
Temporal	Simple chronological	(28/42)	88.1%
	Reverse chronological	(11/42)	
	Future chronological	(12/42)	
Causal	Explicit Cause	(7/42)	23.8%
	Alternative Reality	(3/42)	
Spatial	Spatial Proximity	(10/42)	23.8%
Hierarchical	General to Specific	(28/42)	71.4%
	Specific to General	(16/42)	
Comparison	Dimension Walk	(20/42)	64.3%
	MeasureWalk	(19/42)	

DESIGN IMPLICATIONS

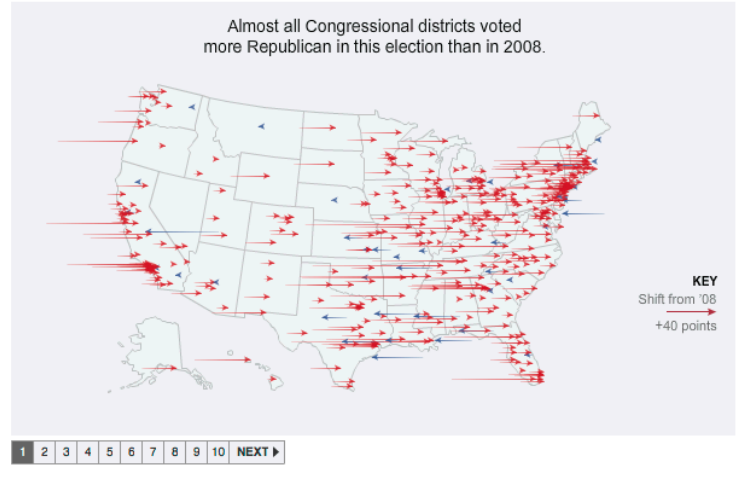
Several insights that emerged from our analysis inform the design of an algorithmic approach that we describe below for identifying sequencing possibilities in narrative visualization. The first implication consists of a set of *transition types* characterizing the difference between the data shown in one visualization and another that directly follows it (Table IV.I). A key aspect of the types we observed is that each represents a single change in one dimension of a data representation from one slide (visualization) to the next. As such, the types imply a data-dependent intention behind sequencing choices. Five primary categories of transition types that share this characteristic emerged from coding.

In *Dialogue* transitions, a question is explicitly posed in one state (such as a slide in an interactive slideshow or a frame in an animation), and is immediately followed by a visualization that answers that question. *Temporal* transitions are orderings of visualization states based on a time variable associated with the data in each (Figure IV.III). These include standard *chronology* (forward progression through time) as well as moving from back in time from one visualization to the next (*reverse chronological*) or forward in time to a visualization that shows a future projection (e.g., *future chronological*).

Published: November 3, 2010

A Historic Shift

Almost all Congressional districts voted more Republican in this election than in 2008.



Published: November 3, 2010

A Historic Shift

The shift reversed the Obama-driven wave of 2008 ...



A Historic Shift

... and reversed the movement toward Democrats in 2006 when the Iraq war was a top issue.

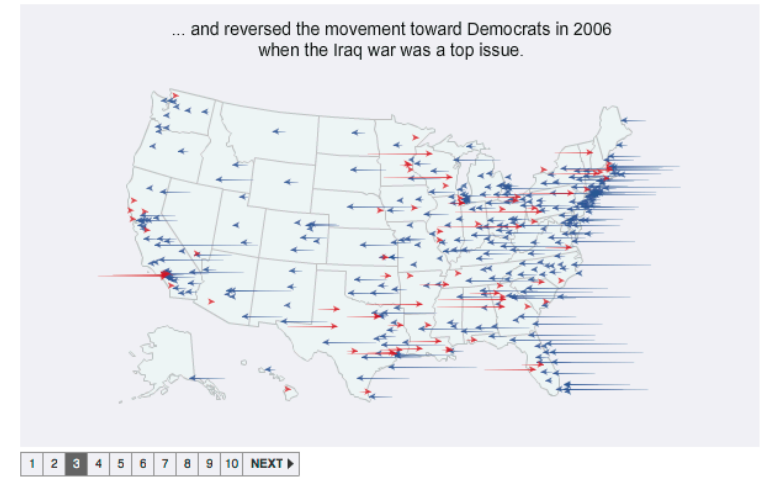


Figure IV-III: Reverse chronological (temporal) transition sequence in "A Historic Shift" by the NYT.

In *Causal* transitions, one visualization state follows another to explicitly hypothesize a causal relationship. For example, a bar chart of voting likelihood by region could be followed by a bar chart of voting likelihood by income along with an explicit statement like “Income drives voting patterns.” *Hierarchical* transitions order visualization states based on the level of detail or degree of filtering of data they involve. Our definition of this transition type encompasses existing definitions of more specific forms of transitions that involve hierarchy that are noted in research on analytical visualization operations. We code visualization transitions in our sample as hierarchical transitions that resemble drill-down transitions as described by Segel and Heer (2010): presentations of a general theme followed by the viewer’s choice of specific instances of this theme. However, the explicitly ordered examples we observed in our sample did not provide such as choice to the user. An example is a map of the world followed by a view of a map of a specific location (Figure IV.VI). Our definition of hierarchical transitions also includes view transformations (defined as movement of a camera through a virtual space by Heer and Roberston (2007)) like zooming, and filter transformations in which data elements are added or removed from the display without changes to the underlying data schema. For example, we label as hierarchical a sequence of increasingly detailed visualizations (e.g., a choropleth map of the world with the mean GDP of continents indicated followed by the same map frame with the GDP of countries indicated). Our definition spans these multiple prior transitions based on the observational perspective of our qualitative study, which considers the output visualization rather than the output visualization and the series of steps that created the visualization.

In *Comparison* transitions, either the independent variable (i.e., dimension per database terminology, e.g., Agrawal et al. 1997) or the dependent variable (i.e., measure per database terminology, *ibid.*) is held constant while the other is changed. Again, our definition is applied somewhat broadly as our coding is based on existing examples,

requiring inference of how designers' conceive of independent and dependent variables in their data in some cases. A dimension walk can show how populations or another form of independent variable differs for a given outcome, as in Figure IV.IV in which a transition is made between a view of life expectancy at birth for males in a set of European countries to a view of life expectancy at birth for females in the same countries. A measure walk provides multiple perspectives on a single population or dimension by cycling through different outcome or dependent variables (Figure IV.V). *Spatial* transitions are a subset of comparison transitions where the same dependent variable is shown for different spatial areas in sequence. Table IV.I lists transition types and the sample frequency.

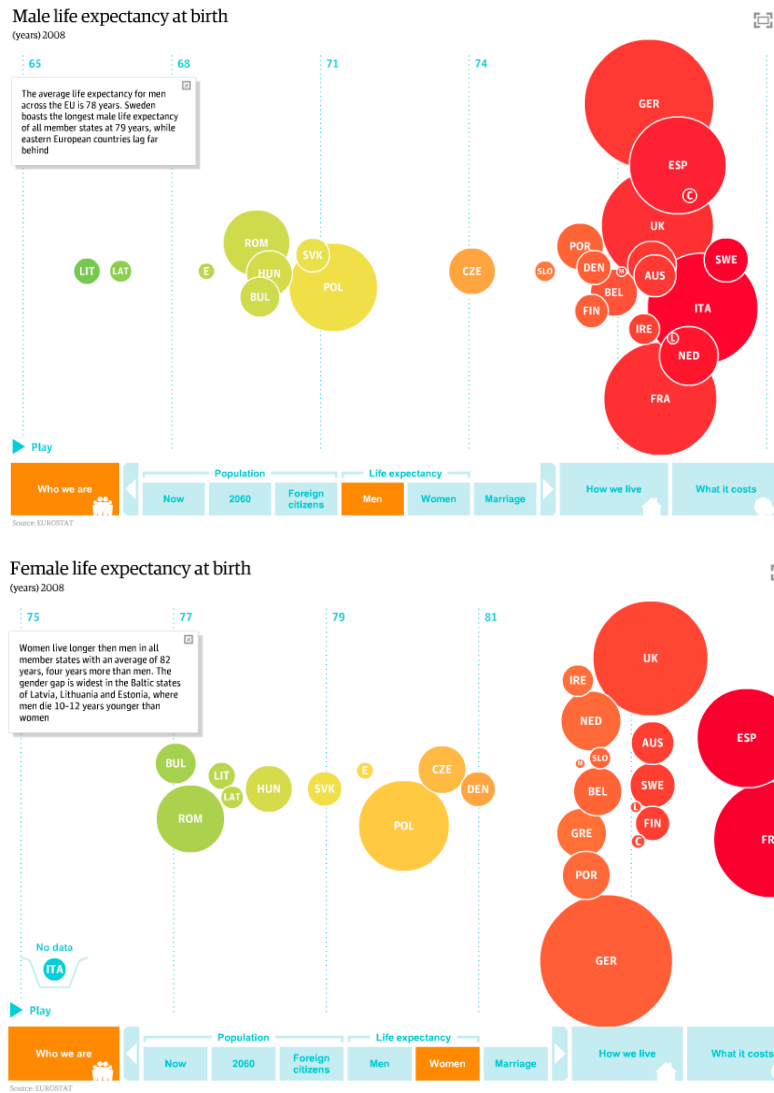


Figure IV-IV: Dimension walk (a comparison transition) in "Europe by the Numbers" by the Guardian.

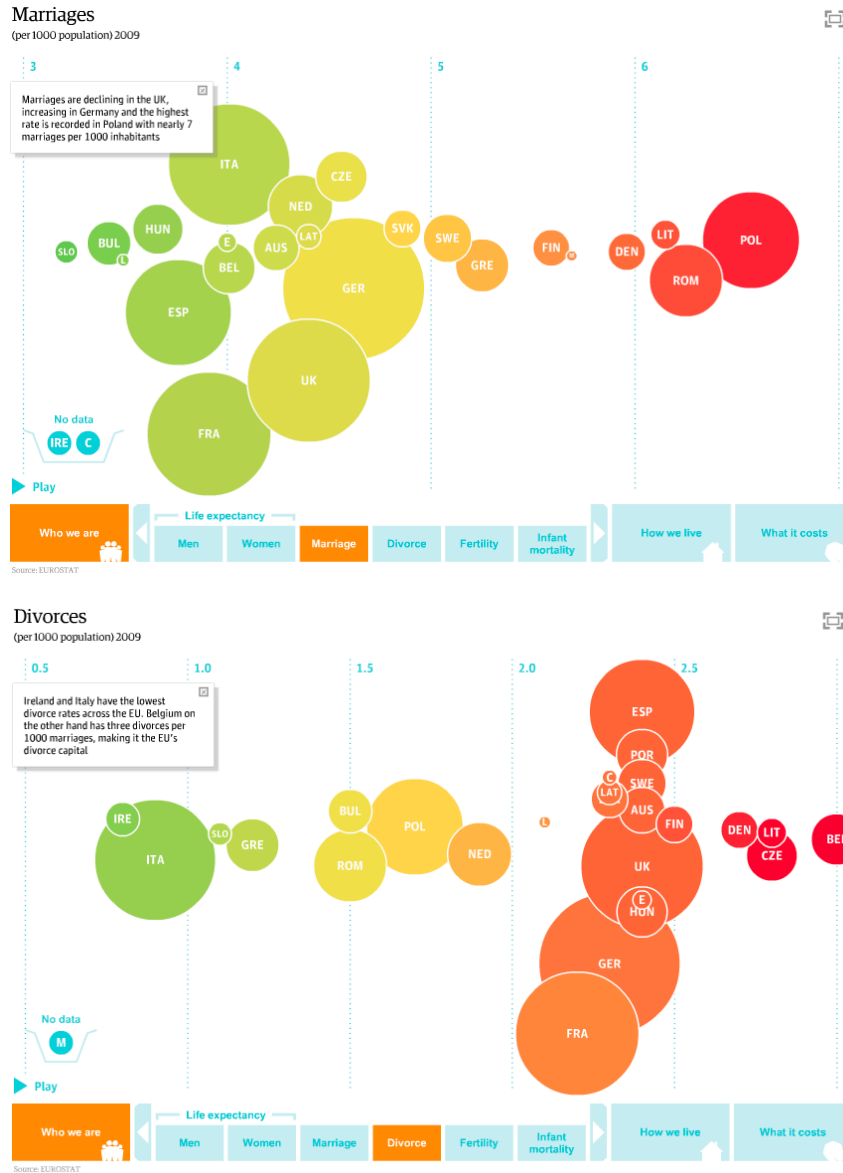


Figure IV-V: Measure walk (a comparison transition) in "Europe by the Numbers" by the Guardian.

These transition types can be distinguished based on whether they require an explicit interpretation of the data applied by the creator (which we refer to as *explicit transition types*), or are inferable from the data attributes themselves using conventions based on data types or graphical formats (which we refer to as *implicit transition types*). For example, Question & Answer transitions require that a creator has a priori classified visualization states by what question(s) each answers, and Causal transitions similarly require creator input on what variables or patterns are causal within and across

visualizations in the set. Chronological transitions, on the other hand, could be labeled automatically given simple matching of data variables against common temporal formats and sorting. Similarly, visualizations of data with explicitly linked hierarchies between data in two views (e.g., filter operations) or with spatial coordinates could be labeled automatically for Hierarchical and Spatial transitions, respectively. We propose that comparison transitions can be inferred either by relying on conventions in existing systems for distinguishing dimensions from measures based on data type (e.g., Stolte et al. 2002), or by using conventions of the graphical format to infer which variable is the independent dimension and which is the dependent measure (e.g., the *x-axis* of a scatterplot is typically reserved for an independent variable (dimension); the color or graduated symbol size of a choropleth map is typically reserved for the dependent measure). We focus on implicit types in the sequencing approach that we outline as these types can be inferred more easily to enable an automated approach.

Another finding describes higher-level or global strategies for sequencing visualizations. We noted that designers occasionally repeated a pattern comprised of two or more transition types as defined in Table IV.I, as if to lend consistency to the presentation's structure as well as to equate different parts of a presentation. We refer to this occurrence as *transition parallelism* based on its resemblance to linguistic parallelism, in which a syntactic structure is repeated in a text, often to equate the importance of two concepts or statements (Corbett and Connors 1998). This characteristic might be alternatively referred to as "sequence compressibility," to capture the "chunking" effect it achieves through encoding several transitions as singular units in a higher-level structure. An example of transition parallelism occurs in the NYT interactive "Copenhagen: Emissions, Treaties, and Impacts," in which three possible climate futures of water stress, flooding, and crop reduction are each investigated. The three possible effects are combined via a *measure walk*. At a local level, the slides for each climate effect include a

general-to-specific transition from a global color-coded map to a specific affected region, followed by a reverse-chronological transition to an image that represents a past symptom of the region's vulnerability along with a comment on likely future effects for the region (Figure IV.VI). We explore the impact of parallelism on user ratings and comprehension of visualization narratives in a final study presented below.

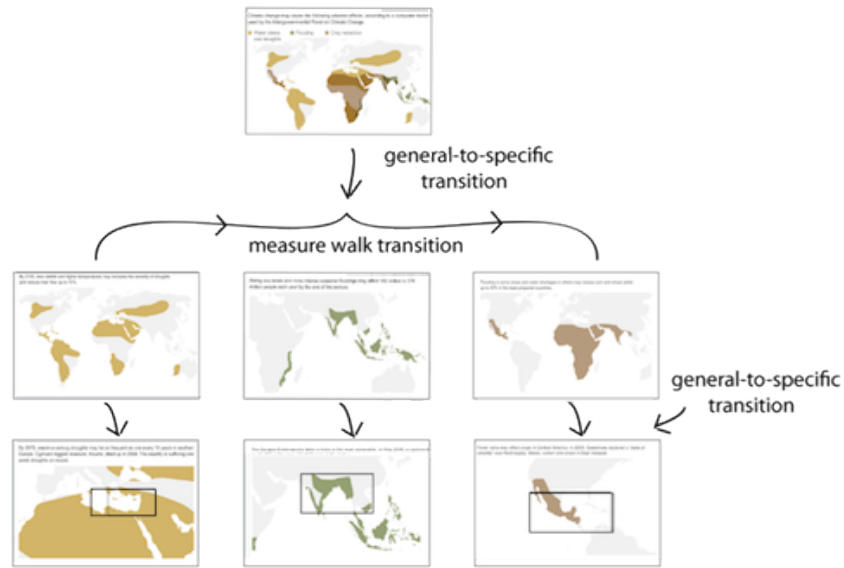


Figure IV-VI: Diagram of NYT's "Copenhagen: Emissions, Treaties, and Impacts" showing parallelism.

The insights that 1) local transitions are frequently based on a small number of changes to data dimensions and 2) parallelism of sequence patterns is used at a global (multiple sequence) level leads to a general observation that **maintaining consistency across transitions** is an important principle in structuring visualization storytelling. We define maintaining consistency as a goal to minimize changes to the data schema in transitioning from one visualization to the next. In many of the transitions we observed, multiple dimensions of a visualization (including both data dimensions like independent or dependent variables, as well as chart format) were held constant across two or more multiple states, such that a limited amount of information changed at a time in transition from one visualization to the next. For example, rather than transitioning to a bubble chart of the GDP of North African countries in 2000 to a bubble chart of the GDP per

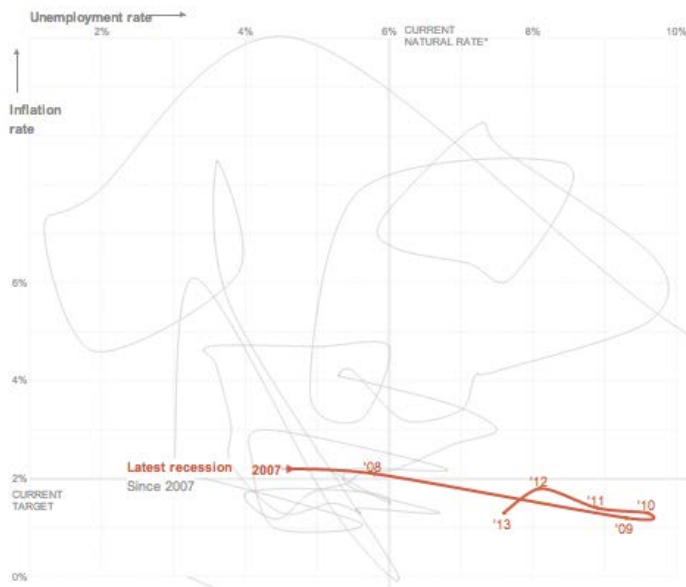
capita of the same countries in 2010, designers tended to choose one dimension (such as time) and maintain the others (independent variable, dependent variable, etc.). When multiple aspects of a representation did occur between consecutive states, slide shows that included animation often used animated transitions, a technique for easing the comprehensibility of transitions (Heer and Robertson 2007). We hypothesize that maintaining consistency through gradual changes between consecutive visualizations in narrative presentations enables comparisons between slides, helping to balance the necessary “jumps” or juxtapositions that must occur in order for the story to proceed. A series of nearly identical visualizations may be perceived as boring, but the introduction of new unknowns must proceed slowly enough that the user can comprehend the sequence and does not become cognitively overloaded. Considering psychological theories of narrative understanding, maintaining a certain amount of consistency between states is likely to make it easier for users to generate the explanations that tie the patterns represented by visualizations into a coherent story.

An Algorithmic Approach to Visualization Sequence Support

Drawing on the insights from our above analysis, we propose a graph-driven approach to finding effective sequences for narrative visualizations. The goal of this approach is to help designers find effective presentation orders for visualizations, such as those that reduce the amount of conceptual change between one visualization and the next (for example, the transition in Figure IV.VII, which depicts two economic periods that are closer in time and which are depicted with common axes ranges, to the transition in Figure IV.VIII, in which the time difference between the economic periods is larger and the axes ranges are different).

We begin by assuming a simple narrative visualization creation model, in which the visualization creator generates the desired single visualizations with an existing system

(e.g., Microsoft Excel, IBM’s Many Eyes, etc.) and the sequence support tool is used as an extension to one of these tools to aid in the design of the presentation sequence. Later in this chapter, we discuss additional considerations required given a more complex narrative visualization design process using a tool that support automated suggestion of the best locally optimized graphic (e.g., Tableau). The described below approach specifies a format for representing different visualization states as nodes in a graph so as to allow an algorithm to compare nodes and label potential transitions using the types outlined above. Inputs and stages are shown in Figure IV.IX. An objective function based on the principle of maintaining consistency is then used to apply weights to edges (transitions) in the graph to allow assessment of the quality of transitions at the local level. We consider the potential for further prioritizing of some sequence types over others, and for supporting user input in the form of desired data comparisons that are used to help prioritize some sequences over others.



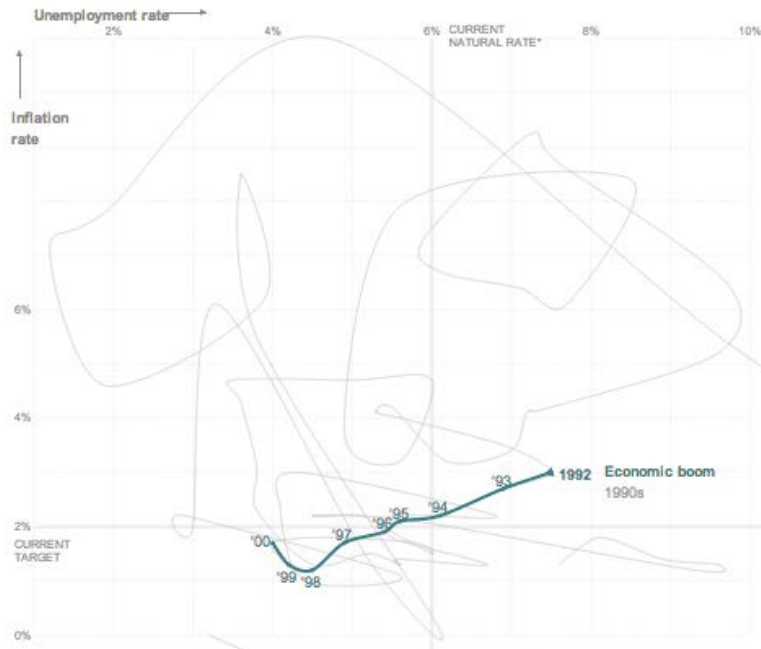
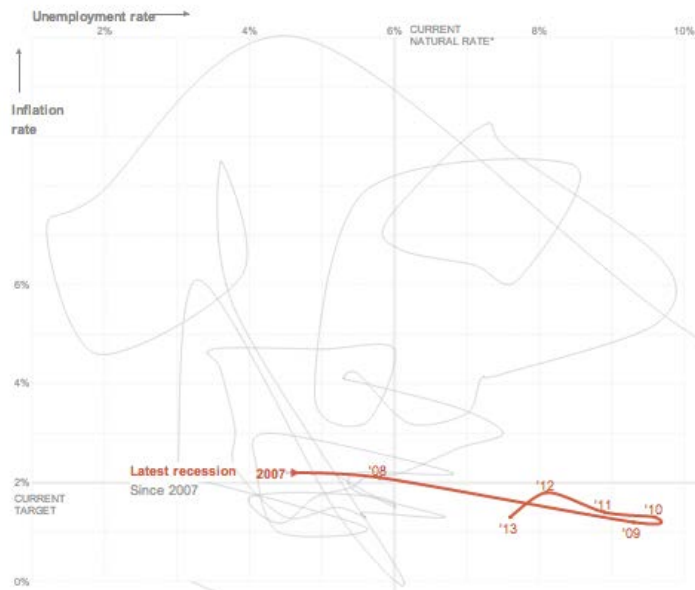


Figure IV-VII: Transition between depiction of latest recession (top) and economic boom on the 1990s (bottom). Note consistency in axes ranges.



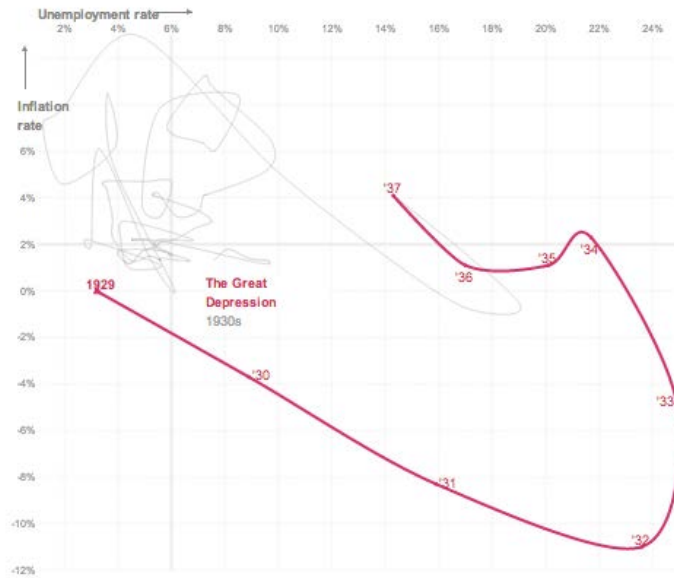


Figure IV-VIII: Transition between depiction of latest recession (top) and Great Depression (bottom). Note change of axes ranges.

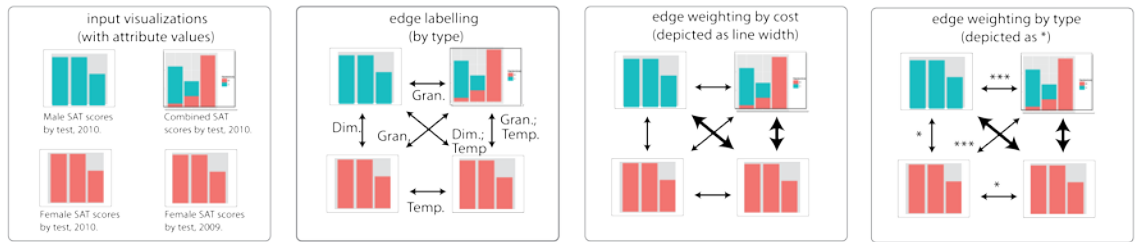


Figure IV-IX: Diagram of graph-based approach in which visualizations represent nodes. Edges (possible transitions) are labeled by type and weighted using a cost function and type weightings (denoted by * symbols) corresponding to user preferences

DEFINING DATA ATTRIBUTES FOR TRANSITION LABELING

We observed a “single change” basis to the explicit transition types in our study. This led us to believe that if we were to identify the set of important data-based attributes along which change tends to occur in visualization-to-visualization transitions, we could infer transitions by comparing pairs of visualizations based on how their attribute values differ. For example the two left-most visualizations in the diagram in Figure IV.IX show SAT scores received in 2010 by males and females, respectively. In this case, the visualizations might be created by applying a filter to the underlying variables of SAT

scores based on a second gender variable. A system could be designed to recognize and label this type of relationship as a dimension walk, as the two views show data for the same measure but for mutually-exclusive populations.

This aspect of our approach resembles models that describe transformations that occur in pipelines (functions used in visualization creation) (Bavoli et al. 2005, Jankun-Kelly et al. 2007) including as directed graphs that can be compared to semi-automatically create new visualizations (Scheidegger et al. 2007). Yet our focus on narrative visualizations differs from a focus on visualizations generated through user-controlled transforms in an analysis setting. While prior work has modeled the conceptual flow of data between pipeline actions from a system perspective, our interest is primarily in user reactions to conceptual change over transitions.

These include a dependent (or outcome) variable, an independent variable, a time variable, and a set of hierarchical relations within or between data variables. Attribute values are defined using data characteristics such as variable types or system-defined labels and information on the data-to-visualization mapping. Hierarchical relations can be encoded through common hierarchies implied in a data type, such as the Roman calendar system; in hierarchical dependencies between several nominal variables such as a variable for car maker (e.g., Ford) and a related variable for car model (e.g., Focus); or by applying filters applied to a given variable to create subsets. Filtering can also occur by applying operations to the visual view only (e.g., zooming) so that only a subset of data is visible. Time variables are often recognizable independent of the representation, such as through date-time formatting applied to given variables in a data set. Additionally, for some plots, dependent and independent variable attributes can be inferred through their mappings to particular positional and retinal visual variables in a given visualization type. In common 2D visualizations like bar charts and scatterplots, the vertical positioning of a data point often corresponds to the dependent variable and the horizontal to the independent variable.

By characterizing each graph node (visualization state) using the four attributes (independent variable, dependent variable, time, and hierarchical level), it becomes possible for a graph-based algorithm to label potential edges (transitions) between nodes as Temporal, Comparative, or Hierarchical transitions (as well as subsets of these types) by looking for simple relationships between pairs of states. The specific comparison transitions of a measure walk and a dimension walk represent changes in a dependent (or outcome) variable and an independent variable, respectively. Temporal transitions involve changes in a time dimension of data, while hierarchical changes involve steps between different levels in a data-defined hierarchy, or can be achieved by filtering.

Table IV-II: Data representation, associated transition types, and relation to common visualization interactions as described in Heer and Shneiderman (2012) and realized in ggplot2.

Representation	Transition Types	Relevant Interactions	ggplot2 Realization
Dependent variable	Comparative – Measure walk	Sort, Derive, Navigate (Distortion), Coordinate (small multiples)	Data variable, Stat (e.g., logarithm), Facet (e.g., small multiples showing related metrics)
Independent variable	Comparative – Dimension walk	Filter (independent variable, such as with query widget), Navigate (scroll, pan), Coordinate (small multiples)	Data variable, Data filter (e.g., one group at a time), Facet (e.g., small multiples by group variable)
Time	Temporal	Filter (direct selection, slider), Coordinate (small multiples)	Data variable, Data filter (e.g., filter data frame by subset of year variable), Facet (e.g., small multiples by year)
Hierarchical relation	Hierarchical	Filter (direct selection, query widget, slider), Navigate (overview & detail, zoom, semantic zoom), Derive (aggregate)	Data variable, Data filter (e.g., show aggregate then filter to one group), Stat (e.g., expand width of histogram bins), Scale (e.g., show smaller scale)

Table IV.II relates this schema to common interactive dynamics in visual analytics as defined by Heer and Shneiderman (2012). For example, a measure walk could be realized in two states where the second represents a sorting or derivation of the first, such

as going from a standard birth rate to a normalized rate, or the second is achieved through a distortion navigation or view coordination (faceting to creating small multiples displaying related dependent variables for a data group). Table 5.2 also describes the schema using a standardized data representation - that of the R package ggplot2 (Wickham and Wickham 2007) which is based on Wilkinson et al.'s (1999) Grammar of Graphics system for visualization characterization. Below, we discuss several specific schemes for integrating such as an algorithm into visualization software with local optimization support.

OBJECTIVE FUNCTION: MAINTAINING CONSISTENCY

Taking a graph-based approach in which edges (transitions) between visualization states (nodes) are inferred by comparing relevant data attributes between the nodes makes it possible to identify possible local (visualization-to-visualization) sequences in a set of visualization states. Yet, without a means of prioritizing transitions, the approach is likely to identify a very large number of transitions even for a relatively small set of visualizations. For example, labelling possible transitions in a set of just 10 visualization states with up the 4 data inferable transition types results in up to 360 labels for 90 transitions. We thus sought a means of filtering the set of possible transitions between visualization sets by relying on edge weighting via an objective cost function.

MAINTAINING CONSISTENCY

Based on our observation of maintaining consistency as an apparent principle used by professional designs, we define an objective function of **transformation cost** that assigns a cost to each possible edge (transition) between two nodes (visualizations states) in the graph. The cost function captures the amount of difference between the attribute values of each visualization node, where difference is measured by the number of changes required to transform the second visualization node into the first visualization node. The more transformations it takes to convert a first visualization to a second, the harder we

expect that it will be for users to infer a connection between the two. This could make comparing the visualizations in a meaningful way more difficult, consistent with research in preserving mental models across transitions (Heer et al. 2007). We examine this assumption about transformation cost in the first user study below.

As a general formulation, transformation cost is the total number of changes to the independent variable, dependent variable, time, and level of hierarchy required to transform a first visualization to a second visualization in a state-to-state transition irrespective of the type of transition. For example, if we consider two bar charts shown in Figure IV.IX, one depicting male SAT scores by test in 2010 and one showing female SAT scores by test in 2009, we assign a transformation cost of 2 representing a transformation of the male independent variable to the female and a transformation of the temporal variable from 2010 to 2009 (a reverse chronological transition). If the female bar chart instead showed TOEFL scores, a cost of 3 would result based on the additional measure transition. To standardize the *unit* of change that equates to a transformation cost of “1” along any single dimension, we suggest that transformation cost should be calculated relative to the full set of parameters describing each visualization rather than in absolutes. For example, the time stamps associated with data for some visualizations might differ in 10 year increments. If the earliest time point is 30 years before the latest time point, but other data sets are only 10 years apart in time, then one might map a transformation cost of “1” to a 10 year difference in time, and higher cost to a 30 year difference. We control for such *within-dimension* differences in cost unit in our studies below, and discuss possible elaborations in the Discussion section.

Assigning a cost function after labelling all possible state-to-state transitions enables filtering to a smaller set of potentially simpler transitions. This filtered set might be presented to a user in an interface for supporting end-user sequencing of narrative visualizations.

PRIORITIZING TRANSITION TYPES

In identifying possible transitions, the transformation cost function treats transition types as equally effective. But do audiences of narrative visualizations regard two visualization states representing a measure walk transition equally to two visualizations representing a temporal transition? The visual information analysis mantra (Shneiderman 1996) suggests that general-to-specific transitions are preferable, but this has not been empirically evaluated, and other questions remain. Systematic preferences for some transitions over others could be incorporated into the above approach using type weightings. We examine user perceptions of local transitions types in the first user study below.

USER INPUT

One of the characteristic features of narrative visualizations is that design decisions tend to be driven by storytelling goals, rather than the more traditional design goal of “letting the data speak” (Tufte 1983). Based on decision science literature that frames story interpretation as a constructive process in which aspects of the information sequence directly affect the meaning constructed by a receiver (Pennington and Hastie 1992, Wilensky 1980), we propose an optional user input model based on desired comparisons. According to this model, visualization creators with clear a priori storytelling goals specify these goals in the form of a set of ranked comparisons between data variables that help convey their intended messages. For example, a visualization creator may wish to highlight a temporal increase in a particular measure, like unemployment, in telling a story about the consequences of a change to leadership in a certain location. The graph-based algorithm we propose would apply these constraints by checking that either single visualizations support the desired comparisons, and if not, that visualizations containing the variables specified as a desired comparison be consecutively sequenced to support that comparison (such as a visualization depicting ‘Unemployment

by county in 1989' appearing directly before a visualization depicting 'Unemployment by county in 1990').

AUTOMATIC GLOBAL SEQUENCING

A final remaining question is how an approach described above can infer global sequences in a set of visualizations that are likely to result in effective linear narrative visualization presentations. For example, how might a tool identify sequences that make use of parallelism, and what information should be used to determine whether a particular form of parallelism is appropriate? We address this remaining question through a second user study.

Validation Studies: User Perceptions of Sequences

We use a large two-part study on Amazon's Mechanical Turk (MTurk) to ask two questions about local transitions:

1. How do users react to the level of consistency between two consecutive visualizations in a presentation?
2. Do users show systematic preferences for temporal, comparative and hierarchy transitions when multiple possible transitions are possible from the same initial visualization?

With regard to 1, we specifically examine how users respond to the transition cost of a visualization transition independent of its type. We vary transformation cost between two candidate transitions to examine how users' choices are affected by cost (referred to below as *Cost Varying* trials). To answer question 2, we control cost in the second half of our study, and examine how choices are affected by type (referred to as *Cost Constant* trials).

Our hypotheses are as follows:

H1: Users will consistently prefer lower cost transitions to higher cost transitions, regardless of transition type.

H2: Users will consistently prefer dimension, temporal, and hierarchical transitions over measure transitions, based on the greater conceptual distance between visualizations showing two different dependent variables.

DATA AND STIMULI

A data set describing characteristics of 3109 U.S. counties across 48 contiguous states was obtained by combining 2010 Census Bureau data with 2012 presidential election data made available by the Guardian Data Blog. This set was supplemented by historical census data dating back to 1790, election-themed data from polls conducted earlier in 2012, and election results from 2008. A set of 74 visualizations was created using the R `ggplot2` package, across common chart types like bar charts, line charts, density histograms, country (U.S.) and state maps, scatterplots, and bubble charts.

Our goal was to create sets of three visualization stimuli of the same type (e.g., map), where two visualizations represent two possible transitions relative to an initial visualization. We use these stimuli in a Mechanical Turk human intelligence task (HIT) that presents users with the initial visualization (labeled Graph 1) and asks that they choose between the other two visualizations (labeled Graph 2a and Graph 2b) as possible following states in a data presentation: “Which of the two graphs is better to appear directly after Graph 1 in the presentation?” (Figure IV-X). The two visualizations to be chosen within each set of three included either 1) alternatives of two different costs when considered with respect to the first visualization (“Cost Varying” HITs), or 2) alternatives of two different types but with cost held constant (Cost Constant HITs).

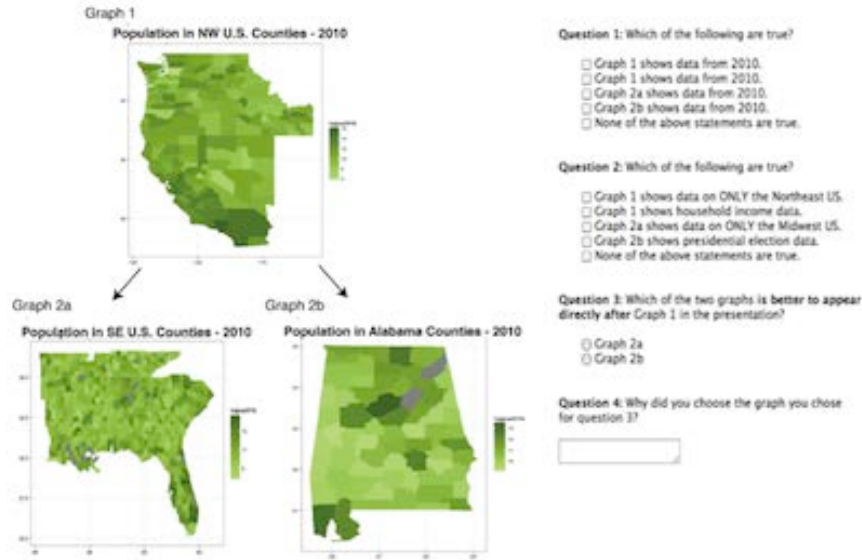


Figure IV-X: Experimental task presenting subjects with an initial visualization (left) and asking that they choose the better visualization to follow the first in a data presentation.

Cost Varying trials: The Cost Varying HITs varied the cost of the two visualizations presented as options to follow Graph 1. Fifteen of the 18 Cost Varying HITs included one visualization with a transition of cost “1” (for example, a change in the region shown only) and the other visualization with a cost of “2” relative to the first visualization (for example, a change in the region and the measure shown). Three HITs included a visualization of cost “1” and a visualization of cost “3” relative to the first visualization (for example, a change in the region, the measure, and the time period). We included these higher cost alternatives to include cases where one visualization was markedly different from the first and might represent a surprising transition. All alternatives were balanced over the 4 transition types of temporal, dimension walk, measure walk, and hierarchical.

Cost Constant trials: In 17 Cost Constant HITs, we tested four transition types: temporal (chronological, reverse chronological), comparative dimension walk, comparative measure walk, and hierarchical transitions (general-to-specific or specific-to-general). These transitions have a transition cost of 1 for the single dimension along which the change occurs. We chose these four types because they are implicitly conveyed by data

characteristics, rather than requiring creator input. To reduce the number of factors in this initial study, we do not distinguish subtypes of temporal and hierarchical transitions (e.g., reverse chronology), nor are Spatial transitions distinguished as a subset of Dimension transitions. However, we maintained separate variables for the comparative types of dimension and measure walks. Both of these types compare one view of data to another that is equal in the time period and the level of the hierarchy or resolution (e.g., country-level data), but may display a large conceptual difference based on the strong human tendency to distinguish between causal and outcome components of phenomena (Diaconis 2006).

In both Cost Varying and Cost Constant HITs we used the same syntax and chart format with a set of visualizations of a given type (e.g., same color and shape) unless changes were necessitated by the chart format (e.g., shape changes for different countries in a map).

EXPERIMENTAL PROCEDURE

The “Cost Varying” and “Cost Constant” HITs were launched as a combined series of 35 HITs with a \$0.10 reward. Each HIT began with an intro page describing that the worker would be presented with a data visualization and asked to decide which of two additional visualizations should follow the first in a data presentation (slideshow). It was stressed in the initial description and on the later “choice” page that the subject should not consider the quality of the individual visualizations in her choice. Additionally, it was explained that the subject would start the task with an additional bonus reward of \$0.15. If the subject’s choice of visualization matched the visualization chosen by the majority of other workers who saw the same stimuli set, the subject would retain the full \$0.15; otherwise, they would lose the \$0.15 bonus. This “punishment agreement” incentivization technique has been shown to produce significantly higher quality responses on MTurk (Shaw et al. 2011).

Upon consenting to participate, a subject was required to correctly answer a question about the goal of the task. She was then presented with the 3 graphs labeled “1”, “2a”, and “2b” (see Figure IV.X). After answering two “information extraction” questions which we required to verify that the subject paid attention, she was asked a multiple choice question, “Which of the two graphs is better to appear directly after Graph 1 in the presentation?” with “Graph 2a” and “Graph 2b” as the only choices.

RESULTS

143 total workers completed the 875 HITs (trials) in the study, taking an average of 118 seconds per trial. We omitted 179 (20.4%) of the 875 trials where subjects answered at least one of the information extraction questions incorrectly, leaving 696 observations. We insured 1) that randomization of HIT order in the sequence and presentation order of the 2a and 2b visualizations in any single HIT was successful; and 2) that there were no significant differences in the time taken by subjects to complete the task based on whether transformation cost varied or not (M : 114.6 s vs. 121.3, $t=-1.56$, $p=0.12$).

Effects of transition consistency (transformation cost): We first examined question 1, whether a lower transformation cost between the two visualizations in a sequence resulted in a preference for that sequence over higher cost alternatives. Table IV.III, left, displays the results of a multinomial logit models run with the R package *mlogit*, which enabled us to compare the costs to one another while accounting for the fact that a participant could complete multiple trials. “Transition choice” (a binary variable indicating whether a visualization transition represented by Graph 2a or 2b was chosen) is regressed on transformation cost of “1,” “2,” and “3” to distinguish whether effects differ by cost levels. Omitted from the results is a dummy variable called “present” included to account for the constrained set of cost alternatives available in a trial. The reported models in Table IV.III differ only in which cost is set to the baseline category. Results indicate that while participants are much less likely to choose a higher cost transition relative to a transition

with a cost of “1,” there is no observable difference in a participant’s likelihood to prefer a transition with a cost of “2” to one with a cost of “3.” The order in which the visualization appeared in the choice (#1 or #2) is included as a predictor.

Table IV-III: Results of a multinomial logit regressing “chosen” transition on transition cost and an order indicator (left); three multinomial logits regressing "chosen" transition on transition types (spanning pairwise comparisons)

GLMM <u>Probit</u> on “Cost Varying” Trials		GLMM <u>Probits</u> on “Cost Constant” Trials			
(Intercept)	1.173***	IV Baseline	Metric	Segment	Temporal
Cost=2	-3.24 (0.166)***	Order Indicator	0.358 (0.104)***	0.358 (0.104)***	0.358 (0.104)***
Cost=3	-3.58 (0.334)***	Metric	Omitted	0.547 (0.264)**	-0.047 (0.321)
Order Indicator	0.316 (0.161)***	Segment	-0.547 (0.264)**	Omitted	-0.593 (0.315)*
Controls	Yes	Temporal	0.047 (0.321)	0.593 (0.315)*	Omitted
N (Observations)	714	Granularity	-1.02 (0.203)***	-0.471 (0.244)*	-1.065 (0.315)***
AIC	305.7	Controls	Yes	Yes	Yes
Notes: <u>Probit</u> : standard errors are in parentheses. Significant at: * 10 %, ** 5 %, *** 1 % level.		N	678	678	678
		AIC	857.4	857.4	857.4

Effects of transition types: We next considered whether participants displayed equivalent levels of preference for temporal, comparative, or hierarchical-based transitions when cost was held constant. Table IV.III reports the results of three multinomial logit models run on Cost-Constant trials. These models were run identically to the Cost-Varying models, except that the covariate of interest was transition type rather than cost, and again only the baseline category to be compared against differs across the three models.

Our interest is in whether preferences for one type over another can be observed, as this would be useful in a sequence support tool for suggesting transitions. Interpreting the results for each type with reference to the baseline transition comparison allows us to assess relative preferences for transition types. We find that a temporal transition is preferred over hierarchical, dimension, or measure transitions (all $p < 0.01$). Both dimension and measure transitions are preferred over hierarchical transitions as well (both $p < 0.01$). No preferences exist between a dimension and measure transition. Results

can be summarized as follows (“>” indicates that the type to the left was preferred over the type to the right, and “|” represents no preference):

Temporal > (Dimension | Measure) > Hierarchical

We also observed an order effect based on whether a visualization was in the first or second position from left in the layout. Hence, contextual factors (such as bias toward the last visualizations seen) may influence interactions with narrative visualizations.

Global Sequencing: Impacts of Parallel Presentation Structure

Our qualitative study suggested the global strategy of parallelism, or repetition of certain local level transition sequences within a visualization presentation. Here, we use a between-subjects study to ask: Does using parallelism in a global sequence benefit presentation audience members, in the types of patterns that are understood and/or ability to remember a visualization story? This provides information with which we can evaluate whether global strategy effectiveness can be modelled simply by summing local transition costs, or whether additional objective functions for global sequencing are required.

DATA AND STIMULI

The primary difference between the prior study and this one is that participants in this experiment are shown an entire presentation, rather than only one transition (e.g., two visualizations) at a time. We begin with a set of visualizations that displays the following characteristics, which we expect to be common in many presentations: the set includes data on two (or more) high level concepts or “groupings,” with each grouping being associated with multiple visualizations in the set, and each visualization in one grouping having a counterpart visualization in the other grouping which differs only based on the grouping dimension. In our study the grouping dimensions is time period (1900 and 2010), but other examples might be presidential candidates (e.g., Obama

election results by region versus Romney election results by region), or even two levels within a hierarchal dataset (e.g., various labor statistics by continent and by city). We kept format the same across all visualizations (using bubble charts) to allow us to examine sequence effects in a controlled setting. The visualizations we use are all bubble chart visualizations that display fertilizer usage by state for three spatial regions: the full U.S., the Eastern U.S., and the Western U.S. time periods. The visualizations are alike except that the 1900 charts display 1900 population data from our Census data set (reabeled as Fertilizer Usage to prevent strong effects of prior knowledge in the task) using blue circles and the 2010 charts display 2010 population data using green circles. In each chart, the size of the bubble and the position along the y -axis (the only labeled axis) are both set to a scaled version of the population statistic for that state in either 1900 or 2010.

We examine two main forms of parallelism described in the Design Implications section above, and depicted through examples in Figure IV.IV and Figure IV.V: a dimension walk and a measure walk strategy, plus several variants derived from these which deviate from the *perfect* repetition of local transition patterns of the first two. The measure walk strategy, which w

e refer to as a *between-group* sequence, interleaves visualizations from the two groups such that a measure for one group always appears directly before the same measure for the other group. A dimension walk strategy, which we refer to as a *within-group* sequence, keeps the visualizations corresponding to each high-level group in consecutive sequence (e.g., three 1900 visualizations followed by three 2010 visualizations). Our expectation is that the between-group sequence will support comparisons between the two groups for each measure. On the other hand, the within-group visualizations will support comparisons between measures within each higher level group. Noting that these sequence types both include one or more transitions with costs greater than one, we also include several variants of the between- and within-group

strategies, but where the sequences were revised to potentially enable additional comparisons and reduce the overall costs associated with the sequence. However, this requires breaking the “perfect” parallelism of the first two sequences (Figure IV.XI).

Our hypotheses are as follows:

H1: Non-reverse treatments (between and within-group sequences) will be rated as more understandable and less difficult to explain than reverse treatments.

H2: Performance on between- and within-group comparison questions will differ by treatment.

H2a: Subjects who see between-group sequences will perform better on average on between-group questions.

H2b: Subjects who see within-group sequences will perform better on average on within-group questions.

H3: No differences for treatment will be found for accuracy on the null comparison questions.

H4: Memory will be better for non-reversed sequence treatments.

We note that confirming H1 and H4 would suggest that computing global cost by summing local transition costs is not optimal. This is because the within-reverse and between-reverse treatments have *lower* costs than the non-reverse treatments when global cost is computed as the sum of local costs. Instead, another objective function(s) to capture global sequence preferences may be needed.

EXPERIMENTAL PROCEDURE

82 Master’s students from a large university were recruited and given an \$8 Amazon gift card for participating. An initial screen described that participants would view a presentation of data visualizations that was designed to communicate a story about the data, and would be asked several questions about the content. After answering a multiple-choice question that ensured understanding of the task goal, the participant

viewed a self-advancing presentation of the six visualizations corresponding to one of the four treatments. Each visualization was shown for 8 seconds before the page advanced. Hints that remained visible during the presentation explained the presentation format and prompted participants to pay attention to how the data in each visualization changed from state to state.

After viewing the presentation, the participant answered a question to verify he or she paid attention to graph labels, and provided a free-text explanation of why he or she thought the visualizations appeared in the order they did. The participant provided 7 point Likert ratings in response two questions: “How easy was it to come up with a reason for why the visualizations were put in the order they appeared in?” and “Assuming the presentation is designed to communicate a story about the data, how easy is it to understand the presentation?” The participant was then given a second, unannounced opportunity to watch the timed presentation, followed by a page that presented eight True/False questions. Each question asked about a trend that was apparent only in comparing two of the 6 visualizations to one another, which may or may not have appeared consecutively in the sequence. While 15 total visualization-to-visualization comparisons were possible within the group of six visualizations, we focused on a set of eight comparisons that included three within-group comparisons (e.g., Eastern U.S. vs. Western U.S. in 1900), three between-group comparisons (e.g., Eastern U.S. 1900 vs. Western U.S. in 2010), and two “null” comparisons, which asked about a trend between two visualizations that did not appear in consecutive order in any of the treatment sequences.

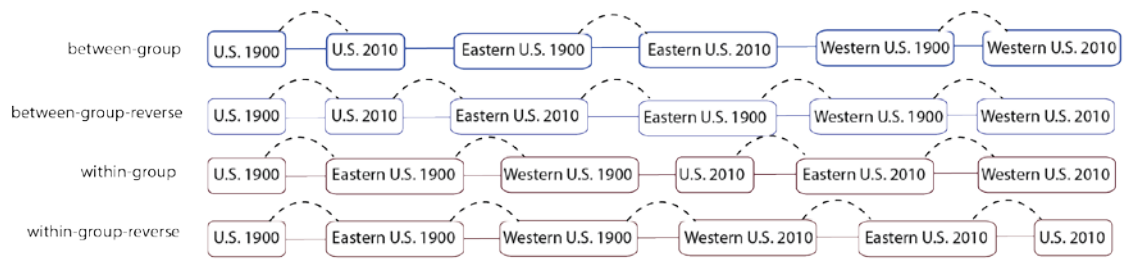


Figure IV-XI: Global sequences to support different hypothesized comparisons between consecutive visualizations (depicted with dotted lines).

RESULTS

82 subjects completed the task, averaging 711 seconds. Removing data from subjects who incorrectly answered the verification question left data from 73 subjects for analysis. We first checked whether ratings on the difficulty in explaining a visualization and how understandable the presentation was differed based on whether the sequence exhibited “perfect” parallelism (e.g., was *not* a reverse sequence treatment). Regarding H1, ratings for the difficulty of explaining the presentation higher for reverse treatments (M : reverse=4.79, non-reverse=4.03), yet this difference was not significant ($t=-1.85$, $p=0.06$). Ratings for the understandability of the presentation did not significantly differ (M : reverse=4.12, non-reverse=4.56; $df=67$, $t=-1.25$, $p=0.21$).

We next examined whether accuracy on the between, within, and null comparison questions differed based on sequence type. While accuracy on the between-group questions was better among subjects who saw a between-group sequence (including reversed) (M : 0.92 vs. 0.86) and accuracy on within-group questions was higher for subjects who saw a within group sequence (including reverse) (M : 0.87 vs. 0.84), t-tests for between and within question accuracy indicated no significant differences by treatment ($df=69$, $t=1.58$, $p=0.12$ for accuracy on between group questions, $df=70$, $t=-0.57$, $p=0.57$ for accuracy on within group questions). As H3 expects, no treatment-based differences existed for accuracy on null comparisons (comparing pooled between-group treatments

with pooled within-group treatments; $df=70$, $t=-0.26$, $p=0.80$ comparing pooled between and pooled within treatments).

Finally, we calculated total error for the memory task by summing the number of visualizations (out of six) that were incorrectly sequenced in the memory task. H4 predicts that memory for the original presentation sequence will be better if the sequence uses “perfect” (non-reverse) parallelism. Results confirmed the difference. An ANOVA indicated significant differences between individual treatments ($F(3,69)=5.59$, $p=0.002$). TukeyHSD tests comparing the four individual treatments identified significantly better memory for the original sequence in the between-group treatment compared to either the between-group reverse or within-group reverse treatments (adjusted $p=0.04$ and $p=0.007$, respectively), as well as significantly better memory for the original sequence in the within-group treatment compared to the within-group reverse treatment (adjusted $p=0.02$) and marginally better memory for within-group compared to the between-group reverse treatment (adjusted $p=0.09$).

Integrating Sequence Support in Automated Presentation

As discussed above, our proposed graph algorithm assumes a relative simple design scenario for narrative visualization presentations, in which visualizations are created a priori and the algorithm is used to suggest effective sequences, with optional user input capturing storytelling goals. We now discuss modifications for integrating our algorithm into existing visualization systems that provide semi-automated local (singular) visualization design suggestions. This context calls for several new objective functions to negotiate potential conflicts between localized and global visualization design optimization.

In this section, we first outline an APT-style generation algorithm that can be used to populate the space of possible visualization designs for each desired visualization. Pattern expressiveness is introduced as a measure for capturing the extent to which a

singular visualization design expresses possible pairwise comparisons between data points. A pattern effectiveness score is calculated by applying a perceptual model to this measure. We expand upon the graph algorithm described above by outlining a specific process for determining preferable sequences given a large space of possible singular designs.

AUTOMATED PRESENTATION DESIGN SCENARIO

We assume a visualization creator has obtained a multivariate dataset, which might include any mix of Boolean, nominal, ordinal, geographical, and quantitative variables. The creator also specifies an “input visualization list” of the data to be used to create visualizations in the presentation. Each item in this list (referred to as the visualization id, or *vid*) is itself a list of data variables that the creator wishes to combine in each single visualization, similar to the ranked lists of variable names used as input to Tableau’s Show Me feature (Mackinlay et al. 2007). An example of a visualization input list is shown in Figure IV.XII. While our prototype implementation supports data variable labels, the input visualization list could alternatively include lists of MySQL-style statements to support more sophisticated data subsetting.



country lat, country long, GDP 2010
country lat, country long, GDP 2012
Poverty 2010, %Unemp. 2010, GDP 2010
Government Type, %Unemp. 2010, %Health Ins. 2010
Government Type, GDP 2010, GDP 2012
country lat, country long, GDP Growth 2012
country lat, country long, Birth Rate 2010
country lat, county long, Death Rate 2010
Government Type, GDP Per Capita 2010, Country
Country, Poverty 2010, %Subsidized Housing

...

Figure IV-XII: Example of an input visualization list. Each row is assigned a unique *vid*.

A visualization generator uses this list to populate the space of possible visualization designs given each ranked list in the input list, relying on existing expressiveness and effectiveness rules (Mackinlay 1986). Each visualization design becomes a node in an undirected graph. Edges are created according to transition expressiveness constraints, and weighted using transition effectiveness criteria. Paths are then identified using a greedy approach that also takes into account pattern effectiveness as a function to be maximized.

We describe the components of the approach in more detail below, as they are used in a prototype implementation of the technique.

SPECIFIC APPROACH

VISUALIZATION DESIGN GENERATION

For our prototype, we assume that each *vid* in the visualization input list contains three variables from the data set that the creator wishes to use to create a single visualization (e.g., latitude, longitude, GDP Per Capita 2010; GDP Per Capita 2010,

Unemployment Rate 2010, Percent Health Care Coverage 2010). The visualization design generator proceeds through this list, applying APT-style expressiveness and effectiveness criteria to assign the best 2D visualization format (mark type) to the first two variables in the list. Our current implementation supports 2D mark types of choropleth and graduated symbol maps, bar graphs, and scatterplots. Additional expressiveness criteria are applied to identify the best retinal variable for visualizing the third variable specified for that visualization. Our implementation supports qualitative color mappings, sequential color mappings, divergent color mappings, qualitative shape, and circular area. For example, if the specified data list for the visualization are the three quantitative variables of GDP Per Capita 2010, Unemployment Rate 2010, and Percent Health Care Coverage 2010, the visualization design generator assigns a scatterplot with mappings to position along x - and y -axes as the best 2D mark type. According to graphical perception research, circular area is more accurate for conveying numerical data than is color. The visualization generator assigns circular area as the retinal variable to which the Percent Health Care Coverage 2010 data is mapped. Alternative retinal variable mappings (such as sequential color mapping for the example) are also retained. Table IV.IV shows the mark type and retinal variable mappings supported by our expressiveness criteria.

Table IV-IV: Expressiveness criteria adapted from Mackinlay (1986).

2D Var. Type	2D Mark Type	Var. 3 (Retinal) Type	Retinal Var. Mapping	Binning	Levels
Quant., Quant.	Scatterplot	Nominal, Boolean	Qualitative color, symbol shape	Given	Given
		Quant. (!all x > 0)	Diverging color	Quantile, Equal Interval	4-8
		Quant. (all x > 0)	Sequential color, circular area	Quantile, Equal Interval	4-8
Nominal, Quant.	Bar	Nominal, Boolean			
		Quant. (!all x > 0)	Diverging color	Quantile, Equal Interval	4-8
		Quant. (all x > 0)	Sequential color	Quantile, Equal Interval	4-8
Latitude, longitude	Map	Nominal, Boolean			
		Quant. (!all x > 0)	Diverging color	Quantile, Equal Interval	4-8
		Quant. (all x > 0)	Sequential color, circular area (graduated symbols)	Quantile, Equal Interval	4-8

The visualization generator then produces a set of design alternatives with the selected mark type and each possible retinal variable mapping. This set can range in size, potentially including a large number of designs (e.g., 100+) depending on the number of possible realizations for the retinal variable mapping. Qualitative mappings (those generated for nominal variables) typically result in the smallest number of design alternatives based on the determinate nature of the number of levels for the nominal variable. For example, the nominal variable “Government Type,” referring to the type of

rule in a country, can take 11 possible values. This limits the possible design alternatives to the set of qualitative color mappings that contain 11 values (e.g., two ColorBrewer schemes). If the mark type selected given the other two visualizations in the set is a scatterplot, any qualitative shape schemes that contain at least 11 distinct shapes are also used to create design alternatives.

In the case of quantitative data-to-retinal variable mappings, the data values must first be binned so that they can be mapped to a finite number of retinal variable realizations (e.g., distinct color shades in a sequential or diverging color scheme, or distinct circular area sizes). For sequential and divergent color mappings, we create designs by systematically varying the number of pre-specified levels (i.e., bins) applied to the data values from 4 to 8 based on the typical number of items that can be stored in working memory (Miller 1956). We also systematically vary the type of binning algorithm applied for color, using either Jenks natural breaks, quantile binning, or equal interval binning. Jenks natural breaks is an iterative data classification method for determining the best binning given the observed distances between sets of values (Jenks et al. 1971). Quantile binning computes the cumulative distribution function (CDF) for the set of values, assigning equal proportions of the values to each of four quantiles. Equal interval binning divides the range of possible values into equal size bins. For size mappings, we find bins using both quantile binning and equal interval binning.

Each design is captured as a design specification, a json object with fields containing the design index (a combination of the *vid* or index of the item in the input visualization list and a unique number distinguishing that design from other possibilities for the input item, e.g., “0_1”); a 2D mark type; position frame information for the *x*- and *y*-axes (specifying the range of data values to be mapped to either the width or height of the visualization frame); the frame information for the retinal variable applied to the third

variable, including the set of bin levels describing the range of values in each bin, and the realization of that bin as a color, circular area, or shape.

Finally, the design generator applies a retinal variable effectiveness criteria by evaluating the extent to which the visualization design supports pairwise comparisons between data values. We define a *pattern expressiveness* score, which is the percentage of all pairwise comparisons between data values for which the two values in the pair are mapped to distinguishable visual elements. Because we use ColorBrewer for all of our color mapping schemes, for designs that use color mappings for the third variable we simply calculate the percentage of data value pairs that are mapped to two different bins (i.e., the color applied to convey the data in each bin is guaranteed to be distinguishable by non-color blind users from all other bin colors in the scheme). For nominal variables, we count two distinct data values that take the same level as distinguishable (e.g., if two different countries take the value ‘Democratic’ for Government Type, we consider those two values comparable given the determinate nature of the data categorization). For quantitative variables, two distinct values that are assigned to the same bin (and therefore, color or size realization) are not counted as a distinguishable pair. For circular area mappings, we use an interpolation function to map bin levels (represented as the mean of the range of values assigned to that bin) to distinguishable circular areas. Our function applies the Stevens’ exponent identified in graphical perception research for area judgments (0.7; Stevens 1969). Stevens’ power law describes the increase in stimulus magnitude required to result in an equal increase in perceived magnitude (Stevens 1957).

The pattern expressiveness score for each design is captured in the design specification.

GRAPH REPRESENTATION

Each visualization design becomes a node in an undirected graph (Figure IV.XIII:). Two indexes are assigned as node attributes: the *vid* is used to keep track of the

item from the input visualization list that the design was created to depict and the *design index* is recorded for the specific design (e.g., '0-0' represents the first possible design created for the first item in the input visualization list, '14-120' represents the 120th design created for the 14th visualization in the input visualization list, etc.). Other node attributes include the pattern effectiveness score, the 2D mapping type, the positional frames of the first two variables in the list item, the retinal mapping type applied to the third variable in the list item, and the retinal frame applied to the third variable.

Edges are created between each two nodes in the graph (i.e., visualization designs) so long as the pair does not violate any of the following three *sequence expressiveness criteria*:

- *Unique input visualization violation*: The *vid* for which node *a* was designed is the same as the input visualization list item for which node *b* was designed.
- *Mapping violation*: Two different data variables (e.g., GDP in 2010 and GDP per capita in 2010) are mapped to retinal variables in identical ways. Two retinal variable mappings are considered identical if each unique *visual-data realization* (e.g., color, symbol size, shape where color, size, or shape are used to convey data value, shape) that appears in one design is also used in the other design. For example, consider a first visualization design that is a scatterplot showing teachers' salaries (position along *y*-axis) by their days on the job (position along *x*-axis), with the shape of the points mapped to circles, squares, triangles, and bars to convey the teachers' political leanings (e.g., Democratic, Republican, Libertarian, Green party). A second visualization design displays a scatterplot of teachers' bonuses (*y*-axis) by their salaries (position along *x*-axis), with the same shapes mapped to their university affiliations (e.g., Univ. of Michigan,

Michigan State, Ohio State, MIT). By applying the same retinal variable mappings (shapes to points in a scatterplot), some end-users may mistakenly assume that the shapes in the second visualization are also depicting political parties. Those who do not make this assumption face the need to relearn that the same shapes stand for different categories.

- *Data violation:* The same data variable (e.g., GDP in 2010) is mapped in two distinct ways. A retinal variable mapping is considered distinct if it contains no overlapping visual-data realizations (i.e., none of the same colors, shapes, or sizes where color, shape, or size is used to depict data values). As an example, consider a first visualization design which is a map depicting U.S. states colored using a three-level sequential blue mapping (from royal blue to white) to show unemployment rates. A second design is a bar graph that depicts U.S. states (mapped to position along x -axis) and their GDPs (mapped to position along y -axis), with bars colored by their unemployment rate using a five level sequential blue mapping (from royal blue to white). Despite using the same colors as the end points for a sequential color mapping of the unemployment rates, these two designs are different in the way they map the same data for some data points to shades of blue. This type of violation may confuse users if they intuitively try to apply the mapping they learned from the first visualization to interpreting the second visualization.

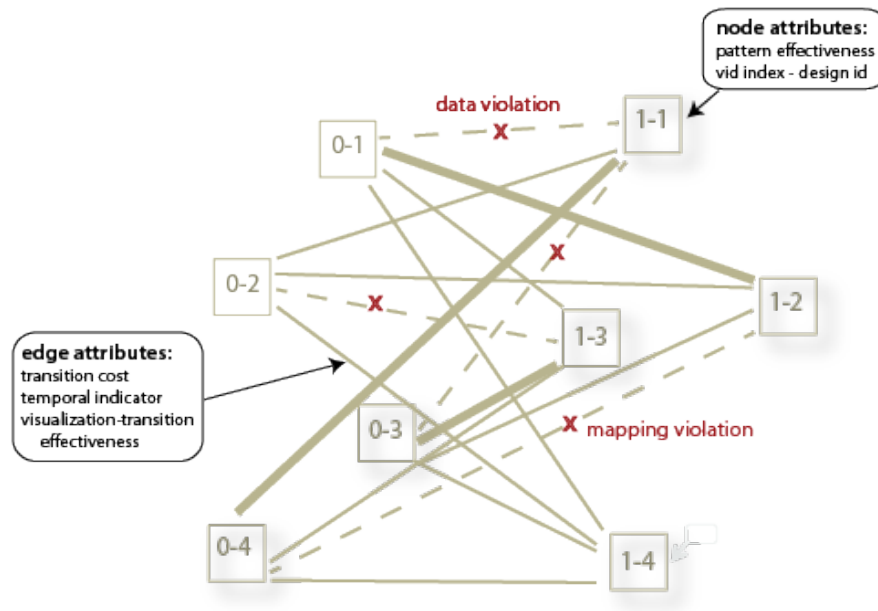


Figure IV-XIII: Diagram of graph-based approach. Edges represent transitions between visualization designs with unique vids and no data or mapping violations.

All edges that are created are indexed with a unique edge index, and assigned a “*vid_pair*” edge attribute that indicates the two vids for each node in the pair (e.g., “0_2” indicates that one node is created to depict the first list of three ranked variables in the input visualization list, and the second node depicts the third list of ranked variables in the input visualization list). The *vid_pair* enables filtering to only the edges for a transition between two specific intended visualizations.

We assign several more edge attributes for later use in filtering the possible paths representing visualization sequences. A *transition cost* captures the relative number of variables (out of the total number in each ranked list in the input visualization list, which is 3 in our demo prototype) that appear in both the visualization designs for that edge. Specifically, we calculate transition cost as the number of changing variables over the total number of variables (e.g., an edge for two visualizations that share only one variable results in $2/3 = 0.67$ transition cost).

Figure IV.XIII provides an example diagram of the approach containing designs for two *vids* from the input list. Widths of the edges in Figure IV.XIII represent transition costs.

PATH IDENTIFICATION

Overview. We propose a greedy approach to searching the graph for paths where each visualization design and transition is “maximally effective” according to measures of visualization and transition effectiveness. We capture the effectiveness of each single visualization design associated with an edge with the pattern effectiveness measure described above. Transition effectiveness is operationalized as transition cost.

Detailed Summary. We begin by identifying the maximally effective visualization transitions for each pairing of *vids*. We first combine the pattern effectiveness scores of each of the two visualization designs in each pair (i.e., edge) with the transition cost of that edge. We compute the *visualization-transition effectiveness* score as the sum of the pattern effectiveness scores divided by the cost of that transition. This enables us to know the score of the maximally effective edges in the graph for each pairing of items in the input visualization list.

We then apply a greedy search to the space of possible permutations (i.e., orderings) of the *vids*. We first generate all possible orderings. That is, if the input visualization list contained 11 items (i.e., lists of three variables indicating the data to be shown in a particular visualization), then we are generating all possible permutations of the set $(0, 1, \dots, 11)$. For each possible ordering of *vids* (e.g., 2, 1, 7, 3, 11, 10, 4, 6, 9, 8, 5), we identify a path for each unique first edge in the sequence (e.g., 2, 1) that achieves the maximum score for that pairing. For example, for the first transition (2, 1) specified in the example ordering we identify all edges between a visualization design with a *vid* of 2 and a visualization design with a *vid* of 1. Each of these edges initializes a path. We continue constructing each of these paths by checking whether any instantiations of the second *vid*

in the path (in the example, designs corresponding to *vid 1*) are associated with maximum scoring paths between that item and the next in the sequence (e.g., designs with *vid 7*). If not, we select the edge with the maximum score, and move to the next index in the list (e.g., 3), etc. Once the entire path has been constructed, we evaluate the path by finding the mean visualization-transition effectiveness score for that path.

Adjustment for creator specified comparisons. If the visualization creator specifies a list of desired data variable or subset comparisons as described above, than any desired comparisons that are not supported by any single visualization can become consecutive order constraints. More specifically, the vids for the two data variable lists that contain the variables that are to be compared (e.g., GDP Per Capita 2010 from list “2” and GDP Per Capita 2012 from list “1”) can be specified as a required consecutive transition (e.g., “2_1”) in any generated paths. This serves to filter the set of possible orderings early in the graph search phase of the algorithm.

Additional optimization. The approach can be further refined by applying additional constraints to evaluation of the possible paths, such as by identifying paths that use more preferred transitions or global structuring techniques. For example, temporal transitions (or other encoded types) can be preferred as an additional constraint, as suggested by the preference for these transition identified in our sequence type validation study above. For instance, our prototype implementation supports temporal transition identification using a simple *temporal transition boolean*. This identifies edges where the three variables in each visualization design are identical with the exception of the year associated with one more of the variables. For example, we note a temporal transition for an edge between a design that depicts country centroid latitude and longitude with GDP Per Capita 2010 and a design that depicts country centroid latitude and longitude with GDP Per Capita 2013. Temporal transition identification is supported by a variable naming scheme that we assume for our prototype implementation, in which all data

variables in the input data set include the date for when the data refers to at the end of the variable name (e.g., GDP_Per_Capita_2010).

Paths with parallel structure (“sequence compressibility”) can be identified given a more complete transition type labeling. Though not supported in our prototype implementation, this could be achieved by having the visualization creator specify a label from the set (dimension, measure) for each variable or MySQL-style statement in the input visualization list. In addition to temporal transition inference as described above, added hierarchical inference features (such as a means of identifying where two MySQL-style statements in the input visualization list refer to subsets of the same data), could be used to label hierarchical transitions. It then becomes possible to label each path through the visualization design space as a sequence of types (e.g., H-HT-T-D-M-T-D-M-HT-DT) and apply pattern mining techniques to these strings to capture those with repeated type sequences.

We have also experimented with capturing the similarity between variables that change for low cost transitions using term overlap calculations. We calculate the number of terms that are shared between the two variables that differ between two designs, and divide this number of shared terms by the total number of terms in the longer variable name, after removing temporal indicators. This information is used to calculate a term overlap score that can be used to prefer transitions for which the two designs display distinct yet related variables (e.g., GDP and GDP Per Capita).

Discussion

We summarize the sequence approach above, addressing how our studies’ insights can be integrated into existing literature, and key implications of our work.

ALGORITHMICALLY IDENTIFYING EFFECTIVE SEQUENCES

The graph-driven approach we propose includes an objective function for minimizing local (visualization-to-visualization) costs of transitions. Each visualization

state becomes a node represented by several attributes (independent and dependent variables, time, and level of hierarchy), and a graph including possible type-labeled edges (types of local transitions) is constructed by comparing the attribute values for each pair of nodes. Graph edges are weighted with the transformation cost calculated for those two nodes, and an additional weighting based on type applied to choose between sequences of the same cost. Our first study's finding of a strong preference for lower cost transitions at a local level supports the importance of first weighting by cost, such as to filter a large set of possible transitions in a sequence support system. The additional systematic differences in preferences based on type that were uncovered supports also weighting edges by type to identify sequences.

The results of our global sequencing study suggest a need for more sophisticated global constraints than simply summing local transition costs to determine the best path through a graph of weighted visualization transitions. While our results regarding how comparisons are affected by sequence were inconclusive, if further study confirms a link between consecutive sequence and comparison, then a sequence support system could take the comparisons that the visualization designer wants to make as input, and use these as constraints in identifying the best sequence. Finally, the improved sequence memorability for sequences with "perfect" parallelism, rather than those that reverse local transition patterns, suggests benefits to also automatically identifying and prioritizing sequences that use parallelism. In the context of approaching automatic sequencing as a graph search, a promising approach would be to infer graph motifs (patterns in local transition type) (e.g., Wernicke 2006) from string representations of paths through the graph, and then identify paths that contain the motifs two or more times.

LIMITATIONS

We evaluated temporal and hierarchical transitions as singular types without distinguishing subtypes like chronological and reverse chronological transitions. Yet

differences in perceptions and preferences may exist between subtypes (e.g., a preference for going forward in time rather than backward). We also did not distinguish spatial transitions from other independent variable changes but it is possible that participants' reactions to the spatial subtypes are somewhat distinct from other forms of independent variable transitions.

Future studies should determine the extent to which explicit guidance about the reasoning behind a transition can overcome sequence effects. For example, can annotations added to visualizations in an interactive slideshow, or a presenter's statements in a live presentation, overcome the effects on the audience of a complex transition?

As noted in the description of the approach above, there may be ambiguity in the particular decision rules used in transition labelling under a given grammar. Factorial crowdsourced user studies in which transition labels are removed is one avenue for distinguishing the conceptual differences between visualizations to resolve discrepancies in rankings transitions in implementing automatic sequence support for narrative visualization.

IMPLICATIONS AND FUTURE WORK

A primary intention of our work has been to demonstrate how sequencing can be systematically approached in narrative visualization, such as in system design. Future work should further evaluate how to best combine information on local transition costs, type weightings, and global constraints like parallelism.

A related question is whether animating a transition can balance the potential negative effects of a costly transition. Suppose that several visual representations that a creator wishes to include in a presentation require costly transitions. Can animating these transitions reduce the negative effects for presentation users?

Additionally, by relating our approach to the grammar of graphics (Wilkinson et al. 1999) and standard visualization interactions (Heer and Shneiderman 2012) we

demonstrate how decision rules for labelling transitions might be defined. At the same time, the results of our qualitative approach on observed transitions can be compared to the types of interactions that we did not observe, such as transitions achieved simply by sorting. Doing so could enable deeper understanding of the differences between communicative and exploratory visualization, as well as potentially suggest forms of transitions that could be used in guided interactive narrative visualizations that are designed to suggest a given conclusion by walking a user through analysis step by step.

An important avenue for future work is to evaluate algorithms for reconciling potential conflicts between sequence optimization and automated optimizations that suggest the most effective single visualization (e.g., Mackinlay 1986, Mackinlay et al. 2007). We have presented the details of an implementation of such an algorithm which extends the canonical expressiveness and effectiveness criteria to also include global (sequence) considerations. The pattern effectiveness measure that we design provides a means of capturing how well a single visualization supports pattern finding by an end-user. At the same time, global considerations like transition costs and sequence compressibility (parallel structure) are identified. Combining these two types of measures is required to negotiate local versus global design trade-offs.

Finally, our work has implications for designers of narrative visualizations. Our global sequencing results provide some suggestion that sequential order supports comparisons between presented visualizations. The common “interactive slideshow” format for narrative visualizations could be adapted to allow more comparisons in cases where they are relevant. Navigational choices beyond “Previous” and “Next” buttons (such as an “Up” and “Down” option), could enable comparisons with visualizations that do not appear directly before or after the visualization of focus. Doing so may increase how much is learned from visualized data without resulting in an entirely “reader-driven” presentation where the user is presented with an overwhelming number of navigation options.

Conclusion

This chapter has made several contributions to understanding of narrative visualization by presenting theories around the role of sequential order and the construction of sets of visualizations for presentation. Study of professional narrative visualization presentations was used to identify a set of key transition types popular in these artifacts, as well as tendencies toward minimizing the number of changes to key data attributes (including independent, dependent, hierarchical, and temporal variables). A graph-based model was outlined with the understanding that modelling sets of visualizations as nodes in a graph and transitions as edges could enable automatic identification of “good” sequences using objective functions like transformation cost minimization and a preference for parallel structure. This model was validated through studies of user perceptions of sequence effects. We contextualize the graph-based approach against prior visualization research with a specific focus on integrating global design considerations for narrative visualization with local design optimization. The results contribute new metrics and patterns that can aid researchers and practitioners in evaluating narrative visualization designs, and motivated a vision of how systems might use the model to provide semi-automated, guided support for visualization creation.

CHAPTER V²

Comparative Sample Plots: Visualizing Uncertainty for Complex Visualizations

A theme that emerges in the prior chapters is that many critical design decisions concern comparisons that one wishes to support with a visualization. The importance of setting up the right comparisons aligns with references to the advantages of visualizations as external visual memory aids that enable offloading of otherwise effortful cognitive operations to the perceptual system (e.g., Larkin and Simon 1987). Supporting comparisons occurred in the form of selecting and juxtaposing subsets of data in suggestive ways in Chapter III's investigation of rhetorical visualization, and through sequential ordering and design consistencies between visualizations in a set in Chapter IV. In this chapter, we consider a new challenge related to this theme: how a designer can design visualizations to support intended comparisons to convey to users that data is *uncertain* without generating confusion. Like other decisions explored in this dissertation, this consideration too often involves trade-offs. Users who are made aware that data are estimates based on sampling are less likely to form unsupported opinions based on data that they encounter. Yet supporting awareness of concepts like variation or margin of error is difficult given most individuals' difficulties with statistical concepts (e.g., Tversky and Kahneman 1971), and a lack of support for visualizing uncertainty among many visualization formats.

Consider Figure V.I, which depicts an excerpt from a post that appeared in a New York Times political blog around the time of the 2012 Presidential election (Gelman and

² This chapter is an expanded version of: Hullman, J., Adar, E., Resnick, P., Ghitza, Y., and Gelman, A. "Comparative Sample Plots: Visualizing Uncertainty for Complex Visualizations" (in preparation).

Feller 2012). The post describes how in addition to a rich versus poor divide that has driven Republican versus Democratic voting among many Americans in recent elections, the 2012 election exit polls also provide evidence of a party split among richer voters. A grid of four visualizations displays the percentage of Republican vote in each state for different income brackets. The distribution of colors by state in each individual map view supports the hypothesis that higher income groups tend to vary more in their voting behavior. By comparing the percentages for the same state across different income levels, it is also possible to infer which states show a party divide based on income, versus which states tend to favor the same party regardless of income. Similar map grids are presented to display voting differences by age group, gender, and race, supporting further inferences that some demographic groups tend to vary more in their voting preferences than others.

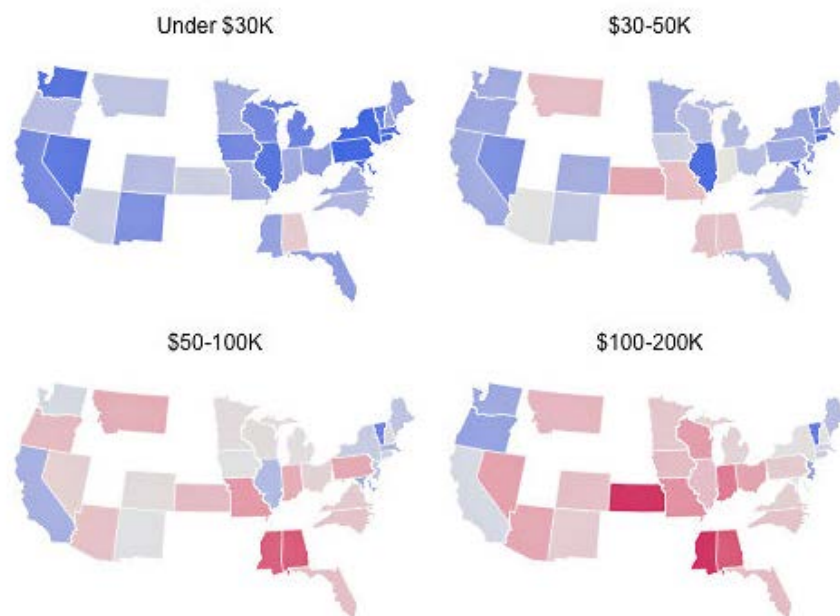


Figure V-I: Voting grid map displayed in the NYT shows voting patterns in the 2012 Presidential election by income level.

A reader of the New York Times post might form generalized opinions about voting patterns and demographics based on the presented evidence of group divisions. The statistical and political modeling experience of the posts' authors and the critical

editorial process that the post is likely to have gone through to appear in the Times reduces the likelihood that the hypotheses it presents will support inaccurate conclusions. However, readers may be at risk of taking too seriously some of the differences shown in the grids for certain demographic groups. For example, an apparently stronger Republican favoring among high income groups in Kansas versus Montana may disappear in future elections based on the expectedly high levels of variation for both these states, which have small populations. A reader with an interest in one or both of these locations might take this visual difference as proof of a difference between the two populations. How might the designer of these visualizations have conveyed these types of subtle yet important differences in estimate reliability using the map visualization? The study presented in Chapter III surfaced evidence of designers' use of uncertainty annotations to note where visualized data is subject to a margin of error, but it is unclear whether just mentioning the possibility of error would be enough.

In this chapter, we address these questions and others related to visualization trade-offs that involve communicating the approximate nature of data. Our intention is to provide insight into alternative methods for representing uncertainty that are appropriate for use in online visualizations presented to end-users who are not necessarily experts. We begin by acknowledging that the interpretations of visualized data that any end-user draws can only be as accurate as the data itself. The term *uncertainty* refers to cases where an interpretation based on data is *not* entirely reliable, as a result of ambiguity or imprecision in judged probabilities, due to a lack of evidence, conflicting evidence, or unreliable evidence (Curley and Yates 1989). It can stem from random variation (or statistical error) that occurs naturally in taking multiple measurements, such as when multiple rainfall readings are averaged to predict the monthly rainfall rate for a location. Or, error may be systematic, like when an instrument takes overly high readings, or a model (e.g., spatial interpolation) has a particularly high error rate for certain types of input data (e.g., outliers). The ubiquity of error and variation in data has made uncertainty

visualization a critical issue in visualization and geovisualization (MacEachren 1992, Marx 2013, etc.)

Confidence intervals, which quantify statistical error and convey precision associated with a calculation (Cumming and Finch 2005), are the basis of many visual uncertainty representations. Error bar representations are commonly added to bar or line charts, where they depict a distribution of possible values around a central, most likely estimate. However, graphical uncertainty annotations like error bars, confidence envelopes, and other summary plots have been criticized for requiring levels of statistical background that novice and even expert users may lack (Belia et al. 2005, Marx 2013). To understand how the error associated with one estimate shown in a bar chart compares to another requires understanding how to interpret the interval created by the error bars, which can itself be ambiguous (e.g., standard error or a 95% confidence interval). The interpretation can also be unintuitive, leading to common misunderstandings. For example, a 95% confidence interval is described as such because if one were to repeat the experiment an infinite number of times and construct a 95% confidence interval around the parameter to be estimated each time, then 95% of these intervals would contain the true value. However, the intervals are more commonly interpreted as the interval will contain the true value 95% of the time. Uncertainty glyphs are also difficult to apply to more complex formats. Alternative uncertainty representations like color, saturation, or blur are difficult for users to perceive and quantify (Kosara 2011, MacEachren et al. 1998). Specific representations of uncertainty have been created for some complex visualization formats (Holzhüter et al. 2010, Talbot et al. 2009) but rarely generalize.

This chapter proposes and studies a methodology for producing, visualizing, and presenting plots of hypothetical data samples to convey uncertainty. We contribute detailed demonstrations of how a comparative sample plot technique can be applied in two common data modeling scenarios: dealing with uncertainty from missing data in social network diagrams, and dealing with uncertainty in model predictions for voting

behavior in a presidential election. This requires development of techniques both for generating hypothetical samples and for making the sample plots visually comparable. For each technique, exploratory user studies confirm the potential usefulness of comparative sample plots for conveying uncertainty to users who are not necessarily expert data analysts.

Related Work

UNCERTAINTY VISUALIZATION

Understanding and integrating uncertainty information is a critical yet challenging part of visual analysis (Marx 2013). Taxonomies frame uncertainty as symptomatic of processes used to create a visualization, from data collection to dissemination (MacEachren et al. 2005, MacEachren 1992, Pang et al. 1996, Thomson et al. 2005, Skeels et al. 2008). We are primarily interested in measurement errors, which affect the outcome of models applied to data yet can be difficult to convey.

As described above, graphical formats for confidence intervals, like error bars of

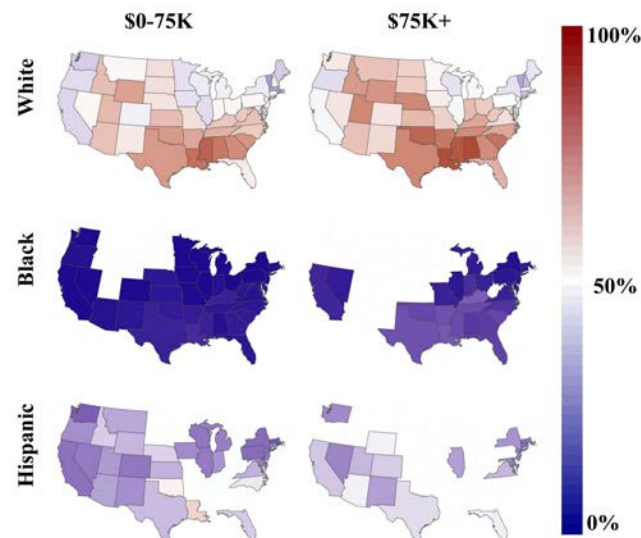


Figure 6.5: A map grid distinguishes voting patterns in the U.S. by ethnicity (rows) and income levels (columns). The data represents predicted voting choices and turnout (Ghitza and Gelman 2013).

envelopes, are perhaps the most common visual representation of statistical error but can be misinterpreted, even by experts (e.g., Belia et al. 2005, Marx 2013). Retinal variables like saturation, blur, and transparency have been explored in geographic visualization (e.g., MacEachren 1992), though results suggest that perceivers have difficulty quantifying these visual effects (Kosara 2011). Methods are needed that can depict uncertainty simultaneously with data (MacEachren 1992) to avoid users' tendencies to dismiss uncertainty information as peripheral (Buttenfield 1993).

We explore presentations comprised of multiple hypothetical samples as an alternative format that avoids separating uncertainty information from the data. This approach is inspired by research on dynamic uncertainty representation in geospatial applications (Aerts et al. 2003, Bastin et al. 2002, Elschlaeger et al. 1997, Goodchild et al. 1994). Elschlaeger et al. (1997) demonstrated how the quality of Digital Elevation Model (DEM) data affects predicted costs for new highways. They stochastically generated multiple visualizations representing possible outcomes and sequenced these via animation. They contribute a method for logically ordering realizations and interpolating intermediate scenes, but did not evaluate their method or discuss generalization to other problems. Evans (1997) conducted a comparison between the use of color saturation to display value certainty levels on land cover maps with maps that displayed only highly certain data and a "flickering" map that alternated between showing all data and only highly certain data. The flickering maps were found to be helpful overall, though annoying to some users.

Bastin et al. (2002) proposed generating and animating realizations via a fuzzy membership function to convey membership likelihoods for map and image categories. The authors note, however, that introducing a fuzzy membership function greatly increases the dimensionality of the visualization space, as fuzzy classification yields a set of continuous membership maps in contrast to a single map produced by hard classification. Additionally, while the information in the produced membership maps can

be combined in a single visualization, such visualizations tend to be more difficult to interpret, as there are no established representations of fuzzy set memberships. The visual mappings that are available to show the memberships, such as color, may conflict with the preferred visual variable for displaying the predicted community memberships. Bastin et al. (2002) focus on linkage between potentially complex visualizations of continuous membership maps in order to provide the analyst with tools to make sense of the potentially unfamiliar methods and representations. Alternatively, we focus in our first demonstration below on the simpler scenario in which uncertainty is depicted for hard classifications using the comparative sample plots technique. While our technique also contributes additional complexity in the form of presenting more plots showing hypothetical samples, our technique does not require adding additional visual mappings to display fuzzy memberships. This is likely to simplify the visual judgment of a user for any single visualization. In addition to this distinction, we generalize the approach of displaying possible outcomes beyond geospatial data modeling scenarios.

GRAPHICAL STATISTICAL INFERENCE

Visualization and the comparison of multiple views of data are important in exploratory data analysis (EDA) (Tukey 1988, Wang Baldanado et al. 2000), as this supports identifying subtle relationships in data and developing inferential models. Our work is motivated by statistical approaches to using visual comparisons in model evaluation. Gelman (2004) proposes an inferential framework for using graphical displays to assess a model that has been applied to data, such as a statistical distribution. “Model-congruent” data is simulated and compared to the observed data to judge the model’s appropriateness or fit. Permutation bootstrapping has been proposed as a basis for a visual hypothesis test method, in which visualization users judge a series of plots with regard to a particular hypothesis (Buja et al. 2009, Majumder et al. 2013, Wickham et al. 2010, Hoffman et al. 2012). If an analyst can pick the observed data from out of a line-up

including 19 “null plots” representing the null hypothesis that no pattern exists, a standard significance test has been approximated.

While that approach is analogous to statistical hypothesis testing, our approach is more analogous to statistical confidence intervals. Classical confidence intervals are generated by making some assumptions about the process by which the observed data sample was generated, which allows inferences about the sampling distribution of some property of the sample (such as the mean) that would result if the same sampling process were run many times. Rather than analytically derive a sampling distribution for a particular property of the sample, we empirically generate alternative hypothetical samples. While comparisons between the observed and simulated data are of interest in the hypothesis testing approaches, visually comparing the resampled plots to one another is the important operation in our approach. In this way, our work is closer to Diaconis’ and Efron’s (1982; Figure V.II) use of contour plot visualizations for exploring effects of variation in input data. In Diaconis and Efron’s work, an original sample of 2,000 measurements of the *pH* value (representing the acidity) of every rainfall recorded at nine weather stations over a two year period were bootstrapped to produce additional samples of 2,000 observations. Each sample was fed as input to a spatial interpolation method (kriging), and the results shown in the four maps reproduced in Figure V.II. As the authors describe, the maps (and underlying samples) allow the analyst to estimate the variability that the contours inferred by kriging would show if many additional sets of 2,000 observations could be collected and compared. Large regions of low acidity on the original map may become much smaller “islands” on the additional maps, or vice versa.

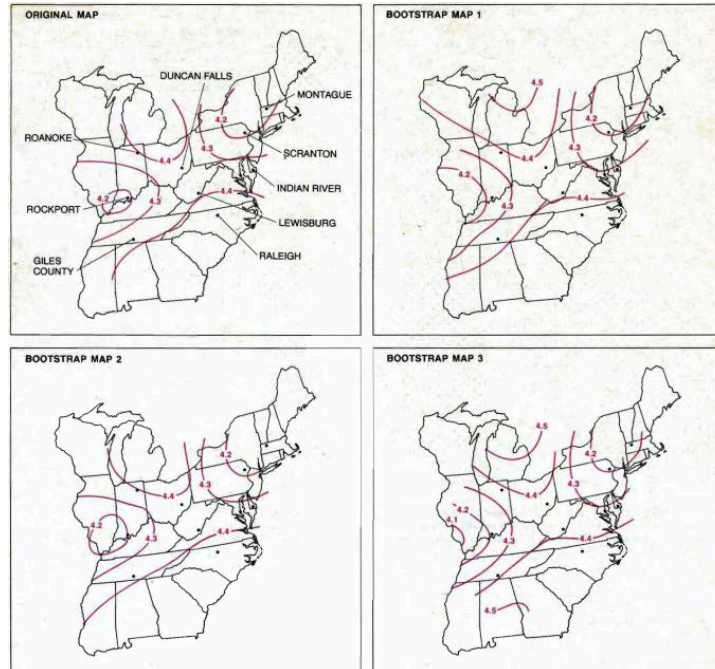


Figure V-II: Contour plots of resampled data convey the extent to which the contour predictions (of pH levels in rainfall) are affected by variation in the input data (Efron and Diaconis 1982).

Recent work on resampling-based graphical inference for InfoVis (Wickham et al. 2010, Majumder et al. 2013) has neglected to address design considerations for achieving the required visual consistency across plots. We provide several specific methods for maintaining visual stability for clustered network diagrams and choropleth maps, and discuss the generalization of the visual stability requirement. These methods can be mapped back onto graphical inference techniques for hypothesis testing. We evaluate the potential for displaying larger numbers of visualizations (e.g., ~ 100) than have typically been used in statistical visualization. Finally, while prior work in graphical inference has been criticized for not addressing perceptual models that may influence the visual test's power, we explicitly examine the effects of different levels of perceptibility on CSP's usefulness.

NETWORK UNCERTAINTY AND DYNAMIC VISUALIZATION

The task of conveying uncertainty for community-clustered graphs has been addressed using various imputation and fuzzy membership techniques. For the reasons

described in the above section, we focus on hard classification, but adapt an imputation technique (Adamic and Adar 2001) to draw samples from the space of probable networks. These samples represent the same group of nodes, but have varying edge properties, and thus community predictions differ. This approach allows for inference of the variability of community predictions, which can result from inaccuracy in the input graph (e.g., missing edges). It also supports many common applications in which “hard” community algorithms are used (e.g., nodes are classified into one cluster), in contrast to “fuzzy” classification methods for networks (e.g., Vehlow et al. 2013).

Our presentation of a technique for supporting visual comparisons across clustered network diagrams resembles dynamic graph visualization approaches for maintaining object similarity across multiple temporal states of a network (Yee and Fisher 2001, Purchase et al. 2006). However, the goal of depicting temporal evolution of non-clustered graphs that motivates these approaches differs from our objective of supporting visual comparisons despite changes in community membership across non-temporal graph instantiations. To our knowledge, this problem has not been addressed in prior work.

Uncertainty as Hypothetical Outcomes

We describe how comparative sample plots can be generated and used to quantify and visualize uncertainty for observed or modeled data. We then discuss why we expect the approach to have advantages over showing a single visualization.

GENERATING HYPOTHETICAL SAMPLES

The introduction of computation and resampling methods like bootstrapping provided statisticians with ways to overcome several constraints on calculating estimates and their reliability. The use of computers for resampling provided an alternative to manual calculations of the uncertainty associated with a parameter (Davison and Hinkley 1997). As a general approach, it frees researchers from the assumptions that 1) data

conform to a Gaussian distribution, and 2) the only available statistical measures are those whose theoretical properties can be analyzed mathematically (Diaconis and Efron 1983).

There are two ways to generate samples:

1. *Bootstrapping (non-parametric) approach*: When the appropriate model (i.e., statistical distribution) is not known, such as when the observed dataset is small or skewed, then non-parametric resampling with replacement (bootstrapping) can be applied. The input data set is resampled with replacement many times (100 or more), producing hypothetical datasets with the same number of observations as the input dataset. The technique can be used directly for samples assumed to come from independent, identically distributed observation. Established adjustment techniques enable applying a bootstrap in complex scenarios where data display known correlations, such as repeated measures from a single subject.
2. *Model-based (parametric) approach*: When a model for input data is known, it can be used to generate new, hypothetical samples. Parameters of the model can be estimated from the available dataset. The model can then be “run” on hypothetical inputs to produce new samples. The resulting samples can communicate the range of values that a given parameter is predicted to take, even if the original dataset did not span the entire range. A simple example is generating samples using the mean and standard deviation of a dataset modeled as coming from a Gaussian distribution. We demonstrate how model-based sample generation can be used for two more complex models: a social network clustered into communities, and a model that predicts voter choice and turnout for the 2008 presidential election.

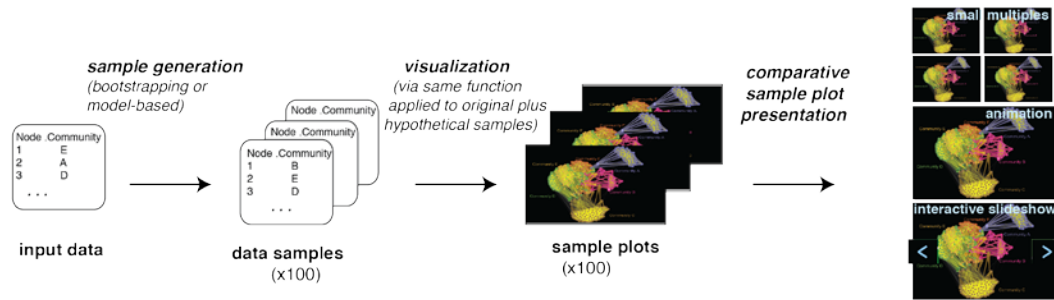


Figure V-III: Overview of the process.

COMPARATIVE PRESENTATION OF GENERATED SAMPLES

Generated samples can be visualized using the same visual mapping function that would otherwise be applied to data. We refer to the visualization of each generated sample as a sample plot. A method for maintaining *visual stability* is necessary, so that plots of different generated samples can be compared without complex visual decoding by the user. For example, a visualization function may be locally optimized to find the best visual mapping given a single input data set. This can result in the same data elements being assigned different locations or retinal values (such as different colors) across sample plots unless the technique is adapted to make the plots more visually comparable. In the Discussion section below we provide a general formulation of the visual stability requirement.

The cognitive limits of a human observer pose questions regarding *how* sample plots should be presented, and *how many*. We propose and evaluate the effectiveness of three presentation mechanisms that vary in 1) how many plots are presented, 2) at what rate, and 3) with what level of interactive control (Figure V.III). A *small multiples* presentation presents a small set of randomly drawn sample plots (e.g., four), affording comparisons between the outcomes represented in each. An *animation* presents a randomly ordered sequence of 100 sample plots at a frame rate of 20 FPS. Animation allows presenting more possible realizations of data than could static small multiples, and facilitates perceptions of changes between states provided there is consistent use of retinal

variables (Robertson et al. 2008). Finally, prior work predicts benefits for understanding uncertainty in data if interactive capabilities allow a user to guide his or her own understanding (Rheingans 1992). An *interactive slideshow visualization* enables browsing of a large set (e.g., 100) of sample plots in the same random order as the animation, but through the use of forward (\rightarrow) and back (\leftarrow) control buttons.

HYPOTHESES

We compare the relative effectiveness of these presentation formats as applied to network diagrams and choropleth map using between-subjects user studies, which we present below. These studies focus on several key measures of interpretation accuracy. Firstly, we calculate the rates of perceptual errors by treatment type by comparing observed responses to a ground truth perception of the data. The ground truth is the answer that a “perfect perceiver and integrator” of the information is expected to produce. In the case of the baseline, this is simply the correct value (whether on a binary or continuous scale) for the target attribute of a data point. For the comparative sample plots presentations, we assume that the perfect perceiver correctly perceives the value for the target attribute in each individual plot, and then integrates those values using unweighted averaging. Our expectation for perceptual accuracy ($H_{\text{Perception}}$) is that perceptual errors will be more frequent as the number of sample plots and rate of presentation increases; hence:

$$\textit{Baseline} < \textit{Small Multiples} < \textit{Interactive Slideshow} < \textit{Animation}$$

We are interested in how well individuals can recognize the true level of reliability for a given data pattern using the different presentation formats. We test reliability estimation accuracy by determining a ground truth reliability level for each possible outcome (out of two total) that a user is asked to compare. Our study protocol first poses a question of which of two possible outcomes are more likely. The participants are then asked to predict the number of times the outcome they report as more likely in this first

question would persist if 100 additional data samples collected in the exact same way were available. Relative and absolute frequency elicitation frames such as this (e.g., “In 100 cases, how many times would X occur?”) have been shown to reduce random noise affecting judgments of uncertainty (e.g., “What is the probability that X will occur?”) (Price 1998) and to increase Bayesian reasoning (Gigerenzer & Hoffrage ’95) over probability formats.

We determine the true level of reliability using a set of 100 generated hypothetical samples. Our expectation regarding reliability estimation ($H_{\text{Reliability}}$) is that individuals will be more accurately infer the true reliability level as the number of plots and interactivity of the presentation format increases.

We provide the baseline users with the option of refraining from making the reliability judgment with no penalty. This option is provided as we expect users of the singular baseline visualization to recognize the difficulty of estimating the reliability of data patterns given a single data sample (see descriptions of study visualizations below for more information on indirect information that might inform some users’ reliability estimates). We expect the following pattern of accuracy:

Baseline < Small Multiples < Animation < Interactive Slideshow

More specifically, we suspect that the better accuracy that results from viewing more samples in the animation and interactive slideshow presentations will result in part from the fact that the small multiples presentation can suggest a true reliability only in increments of 25%. For example, if two plots show the same outcome and the two other plots show the opposite outcome, then a rational user of the presentation will assume the true reliability to be 50%; if three plots show the first outcome the estimate of the true reliability will be 75%, etc. We test these expectations in analyzing our study results below.

Domain 1: Community Inference in Networks

Network (i.e., node-link) diagrams depict relational datasets, such as social relationships (e.g., Heer and boyd 2005, Perer and Shneiderman 2008). Network diagrams support analyzing how connections cluster nodes into communities, which are determined by a community finding algorithm (e.g., the “Louvain” method (Blondel et al. 2008)). Communities are often visualized with convex hulls encircling member nodes (e.g., (Heer and boyd 2005, Perer and Shneiderman 2008)). However, noise in data acquisition can lead to missing edges, like when some individuals do not consistently update connections or join a network (e.g., a friend does not use Facebook). Misclassification into communities can result from such noise in input data. The binary nature of community membership in most visualization schemes, in which communities are grouped in hulls and/or by color, can prevent users from understanding how sensitive outcomes are to variation in the input data.

We describe how hypothetical network samples can be generated by adapting existing imputation techniques for networks, resulting in samples that are slight variations of the observed network and share all its key structural properties. Next, we describe how plots can be generated for these hypothetical samples in a way that permits visual comparisons between them. Finally, we present study of the impact of the comparative plots on users’ ability to answer questions and assess the reliability of their answers.

DATA AND POSSIBLE OUTCOME GENERATION

The starting sample dataset was the real egonet of an adult Facebook user. . We isolated the graph’s largest connected component for use in the experiment (350 nodes, 3492 edges, with an avg. degree of 19.95, a density of .0517, and modularity (Newman 2006) of .483 when applying the Louvain community detection algorithm (Blondel et al. 2005)).

We generated additional hypothetical samples that each included some additional edges, using a model-based approach. We estimated a model from the observed 350 node graph, a model with a similarity parameter for each pair of nodes that are not linked in the graph. We applied a normalized version of the similarity metric described by Adamic and Adar (2001) (specifically, using a form that only takes into account structural features:

$$similarity(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log |\Gamma(z)|}$$

where x, y and z are nodes and $\Gamma(k)$ indicates the neighbors of k . Intuitively, the similarity increases with the number of shared connections, with extra weight for those shared neighbors that have few other neighbors. We normalized the similarity into a pseudo-probability in the range of 0 to 1 by dividing the largest value the similarity may take in the network and then by the largest observed similarity. A new hypothetical graph was generated by including all of the original edges and including each additional edge with probability equal to its normalized similarity score. (An alternative approach would omit observed edges from the original graph with some probability but we did not do that.) This procedure was run 100 times. Resultant graphs had edge counts from 4903 to 5057 (M: 4979), density: .08 to .0827 (M: .082), avg. degree: 28.02 to 28.89 (M: 28.46), and a modularity ranging from .385 to .41 (M: .397).

To support easy comparison of data elements across multiple resampled visualization requires maintaining retinal variables and other visual encodings across representations. We accomplished that by maintaining (1) a single layout, with each node having a fixed position across all plots of hypothetical samples, and (2) by maintaining a large degree of color consistency, so that observed color differences between plots always conveyed semantic meaning of a node being in different communities in the two plots.

To do this, we used the information about which nodes appear together across the set of sample graphs to create a “co-community graph,” which contained an edge for every pair of nodes that appeared together in the same community partition at least once. Edges

were weighted by the number of sample graphs where the pair of nodes were predicted to be in the same community (e.g., if nodes A and B were in the same community 50 times, the edge would have a weight of 50). Edges between node pairs that appeared together in less than half of the samples were eliminated, and the resulting network was drawn in GUESS (Adar 2006) using a standard force-directed layout. Because nodes that often were in the same community formed cliques, the force-directed layout pulled nodes within the same community together but further apart from other communities. Nodes that belonged to multiple communities tended to be placed between the communities to which they might belong, preserving the semantics of the network layout. This layout defined fixed coordinates for all nodes across all plots.

Next, we assigned colors to the weakly connected components in that co-community graph. Colors were assigned using ColorBrewer (Harrower and Brewer 2003) to achieve maximal orthogonality, to make it easier for users to visually distinguish them. Each node's "preferred" color was recorded based on the color assigned to its component in the co-community graph.

To ensure that communities were assigned consistent coloring we determined a stability measure for each node in the following way. An increasing threshold was applied to edges in the co-community graph to eliminate edges that did not appear together at least k times. Each time we incremented k we noted when a node became disconnected and assigned it a stability score of k . Intuitively, those communities that were consistent across all hypothetical samples never became disconnected as they were consistent across all samples, and thus their nodes received the maximum possible stability score.

In each sample plot, a community was assigned the preferred color of the most stable node in that community (ties were broken by the arbitrary numerical id of the node). Intuitively, this had the effect of keeping nodes that were stably together in the same communities, across many samples, the same color in all the plots. Those nodes that

moved between communities took on the “stable” color of the communities to which they were assigned in that particular sample.

For each of the 100 plots a modified metaball was drawn underneath it to take the place of a convex hull (both due to aesthetic concerns but more practically because the hull misrepresents cluster size when an outlier node “pulls” the hull boundary to a far position). The 100 images were used to create a small multiples presentation containing four randomly drawn network diagrams, an animation, and an interactive slideshow visualization as described above. Examples of the diagrams can be seen in Figure V.III and Figure V.IV.

STUDY QUESTIONS AND OBJECTIVES

Viewing nodes relative to social communities is a common task in using network diagrams (e.g., Heer and boyd 2005). Here, the community membership predicted for a given node might be of interest (i.e., with what community is an individual most associated?) Another relevant comparison concerns the size of communities relative to one another. We selected four nodes, each of which might be associated with one of two communities, and four communities to use in node membership questions and community size comparisons.

Each user answered all eight questions: four node membership questions and four community size comparison questions. For each, the user was also asked to assess the reliability of their answer. Samples of each type of question are show below.

- **Node membership.** “Which community is Person 4 more likely to be part of, community C or community E?”
- **Reliability of node membership.** “Imagine that you have access to 100 samples of data (100 network diagrams). Estimate the number of times out of 100 that Person 4 will be part of the community that you selected above?”

- **Community size comparison.** “Which community is bigger (has more nodes), A or C?”
- **Reliability of community size comparison.** “Imagine that you have access to 100 samples of data (100 network diagrams). Estimate the number of times out of 100 that the answer you chose above will be larger.”

As described above, we framed the reliability question using a relative frequency based format rather than a probability-based format (e.g., “What is the probability that the community you chose will also be larger if more data is collected?”), based on experimental evidence that frequency formats reduce random error among estimates (Price 1988, Gigerenzer and Hoffrage 1995).

The node membership question asks for a binary judgment. For baseline users presented with a single plot, answering should amount to decoding the color of the node and matching that color to the community colors to identify the correct label. For users of comparative sample plots, this same judgment calls for a more complex process, in which the user notes the node’s predicted community (color) in each sample plot, then integrates these responses across all plots.

In choosing questions we aimed to include a range of different ‘true reliability’ levels, calculated as described above. High reliability associated with one of two possible outcomes might be one outcome occurring in 85-90% of a set of new samples; low reliability is when each of two outcomes are roughly equally likely (50%). We selected nodes and communities to use in four node membership and four community size comparison questions.

EXPERIMENTAL PROCEDURE

Tasks were run as a between-subjects experiment on Amazon’s Mechanical Turk (AMT) with a base reward of \$1.00. Study participants were told that the network diagrams depict predicted groupings with a social network. Participants who saw more

than one diagram were informed that the multiple diagrams represent different hypothetical samples of the same network for which communities have been predicted. Participants were instructed on how to use the network layout and required to identify the mean node count in communities in a sample network. A practice question familiarized participants with the two-part question format. Participants in the baseline treatment were told they could answer “NA” to the second part of a question (the reliability question) without being penalized in their reward or bonus if they did not feel equipped to estimate the reliability of a pattern from the single visualization they were shown. This option was included so as not to force inaccurate reliability estimates among baseline participants.

The eight two-part questions were divided across two screens. Participants earned a \$0.15 bonus if one randomly selected part of their answer for each question matched the majority response (if part one) or came within 15 of the majority response (if part two) for the same question/visualization. Baseline users who responded “NA” for the second part were told that the first part of their answer would be scored.

RESULTS

199 participants completed the task (mean time: 644.9s). 25 participants with answers that conveyed a misunderstanding of the question (an answer less than 50 for the second question part) were removed.

Table 1 summarizes the results. Overall, comparing community sizes was considerably more difficult than classification of nodes' community membership (perceptual error rates 0.41 vs. 0.17 ; $t(339)=-11.2, p<.001$). The plots did not directly convey the number of nodes in a community and the node count per community was also large ($\mu: 45$), so some of the community size comparisons may have been difficult to judge without manually counting the nodes.

Table V-I: Summary of error rates for network task. Bolded cells indicate a significantly lower error rate in a pairwise comparison with the base, single plot condition ($\alpha = 0.05$, TukeyHSD correction).

Question type	Presentation type	Perceptual error (frequency of wrong community selected)	Reliability error (mean of absolute difference from true reliability)	Signed reliability error (mean of difference from true reliability)
Node membership	Base	0.25	26.3	0.1
	sm mult	0.12	20.8	-6.7
	Anim	0.11	21.5	-6.1
	Slideshow	0.19	24.9	-1.7
Community size	Base	0.40	41.2	16.3
	sm mult	0.41	44.0	14.2
	Anim	0.39	44.1	12.1
	slideshow	0.42	43.6	15.7

For community size comparisons, there was no significant difference in perceptual errors among the conditions ($F(3,170)=.15$, $p>.10$). For node membership questions, users of the baseline visualization made twice as many errors as users of the animation and small multiples ($F(3,170)=4.6$, $p<.01$). (No differences existed between the interactive slideshow error rates and other presentation types.)

For reliability estimates on community size comparisons, there was no significant difference between conditions ($F(3,170)=.94$, $p>.10$). In all conditions, users tended to overestimate reliability, and by similar amounts.

For reliability estimates of individual nodes' community membership, not all conditions were the same ($F(3,170)=5.8$, $p<.001$). The estimates from users in the small multiples condition were significantly closer to the true reliability scores than users in the base condition. In contrast to reliability estimates for community size comparisons, there was no clear pattern in the direction of estimation errors; if anything, users of the animation and small multiples tended to underestimate reliability. The small multiples also performed better than the interactive slideshow, in contrast to our expectation.

Interestingly, only 4 users out of 52 responded “NA” for a node membership or community size question. This is somewhat surprising considering that the baseline visualization provided no direct signal of reliability.

DETAILED RESULTS AND INTERPRETATION

We see no clear evidence of comparative sample plots increasing perceptual error as predicted. Indeed, the animation and small multiples versions led to fewer errors in assessing node membership. This led us to question if there might also be some perceptual benefit to seeing slight variations of a visualization that balances the added complexity. One possible explanation is that comparative sample plots may support easier correction of initial misjudgments of visualized data. When there is a just barely perceptible feature in a single plot, that feature may be more noticeable in some of the plots of hypothetical samples.

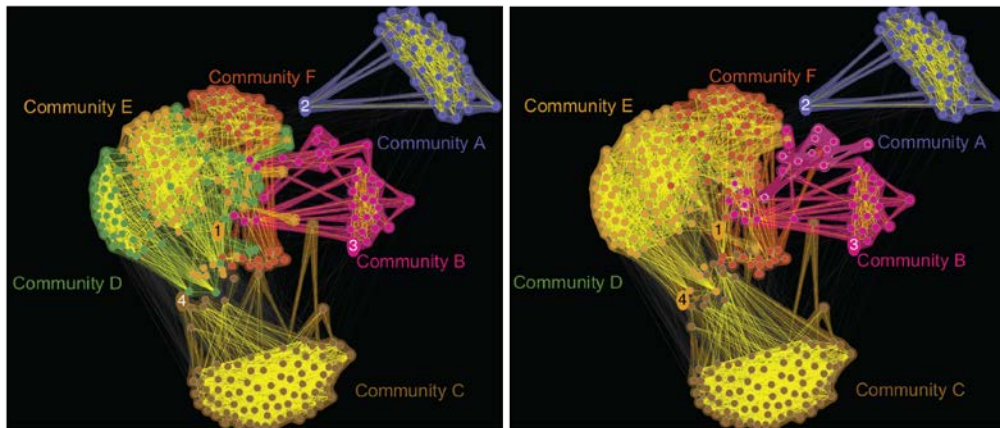


Figure V-IV: Two network diagrams, depicting the same node (4, lower left) being predicted members of two different communities (Communities C and E).

For some of the community membership questions that were posed to users, the communities to choose between were very similar in color (e.g., light orange E vs. brown C) while in others they were more distinct (e.g., bright pink B vs. brown C). There were also differences in how geographically close the two communities were in the network

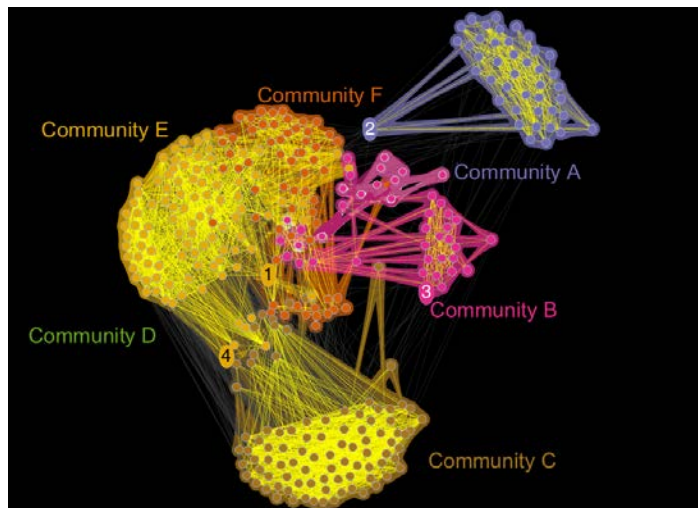
layout (see Figure V.IV for an example). It is more difficult to distinguish the node's color (and consequently the community) when the colors are less distinct and the communities farther apart. We found that the primary difference in between error rates by presentation format occurred for a comparison involving very similar colored, distant communities C (brown) and E (light orange). In this case, baseline users make more than three times the errors that animation users did ($\mu_{\text{anim}}=.09, \mu_{\text{cont}}=.32$). For the other three questions, where the communities were always adjacent and colors more distinct, there were no differences in error rates ($F(3,170)=1.9, p>.10$).

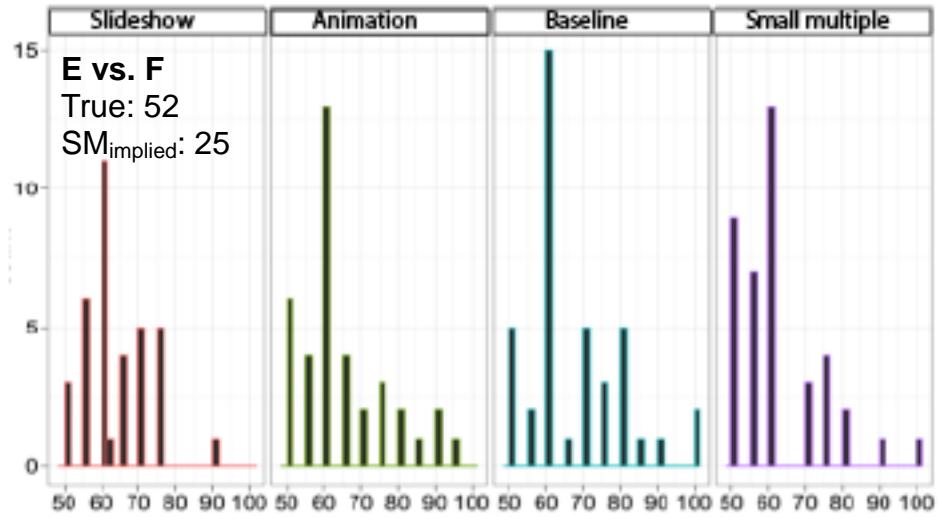
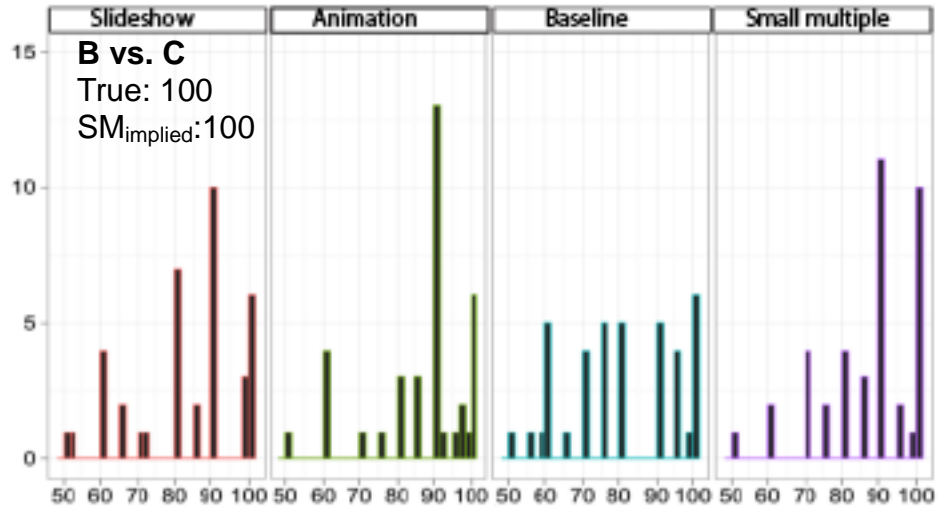
While our results support some advantages over the baseline of the animation and small multiples, the interactive slideshow tended to perform on par with the baseline. Logged data on how many samples were looked at by users of the interactive slideshows showed considerable variation (μ : 184.6; *median*: 14, *min*: 1, *max*: 1235). Slideshow users who viewed more plots had considerably lower perceptual and reliability estimation errors than those who viewed only a single plot ($t=5.44, df=28, p<.001$ and $t=4.55, df=34, p<.001$, respectively).

Our expectation is that the users of the small multiples presentation based their reliability estimates on the percentage of new samples suggested by the patterns across the four randomly chosen plots they were shown. To investigate whether our results support this expectation, we examined the pattern of reliability estimates by question among users of each presentation format. Figure V-V displays the single visualization shown to baseline users above four faceted histograms, each of which shows the number of users (y -axis) who submitted a reliability estimate (x -axis). (Users who did not choose the correct community label for the first part of that node membership question are omitted for clarity). With each set of histograms we display the two signals that we expect to influence the slideshow and animation users and small multiples comparative sample plot

users, respectively: the true reliability as defined using 100 sample plots (True), and the reliability level suggested by the four plots included in the small multiples ($SM_{implied}$).

The results shown in Figure V-V suggest that in general, our expectations for the small multiples reliability estimates hold true. By comparing the $SM_{implied}$ to the histograms, it is apparent that the small multiples users were considerably more likely to estimate in accordance with this signal, with the exception of the comparison between E or F. In this case, the $SM_{implied}$ value for the true answer was only 25, which was not a valid response given our question framing, which instructed users that their reliability estimates should be between 50 and 100. For this question (2nd histogram set from top) many small multiples users submitted of or near 50. The small multiples users' estimates for the comparisons for which the $SM_{implied}$ was 100 may provide evidence that some users were uncomfortable with this level of reliability. For these comparisons (first and fourth histogram set from top) a response of 90 is more common than 100.





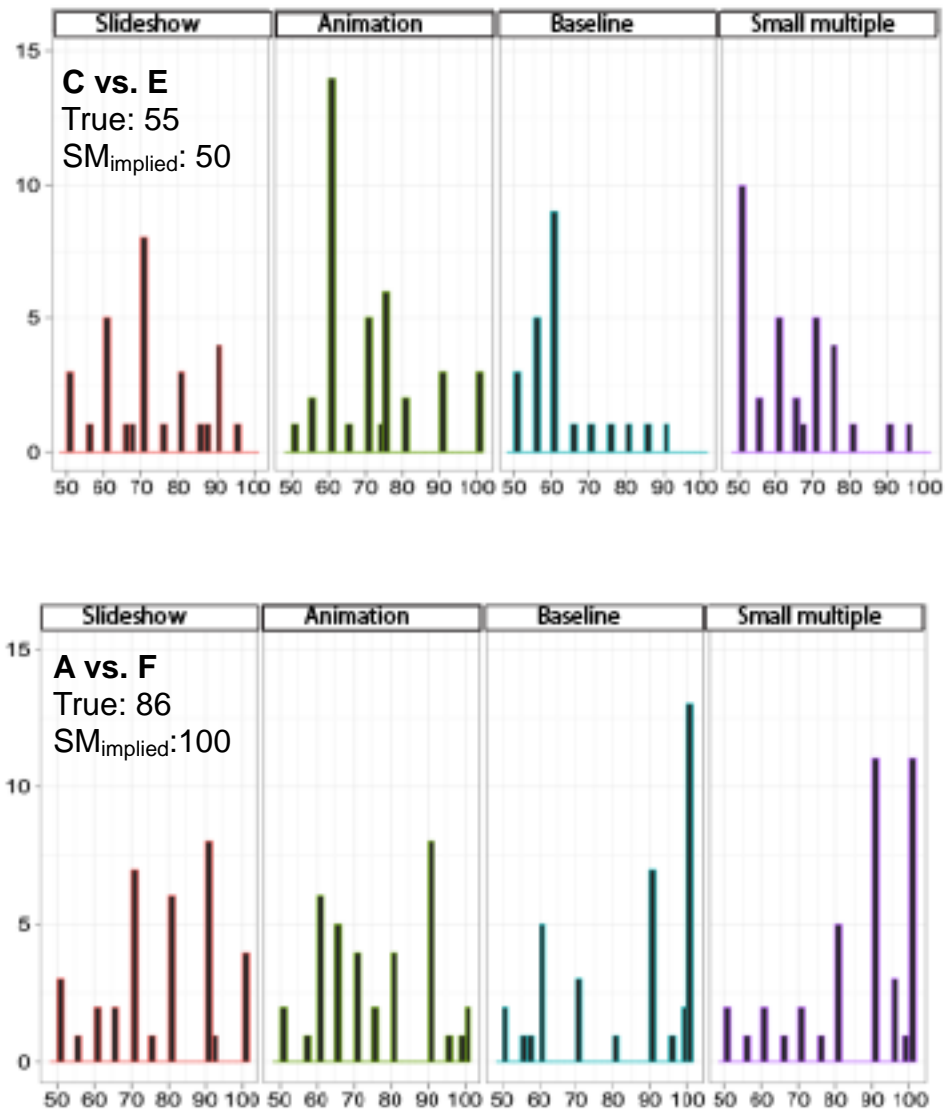


Figure V-V: Baseline network visualization, plus histograms of reliability estimates for network node membership questions. Incorrect responses to the first part of the question are omitted for clarity.

Figure V-V’s histograms also provide some evidence of a hesitancy among users of the animation and slideshow formats to provide “extreme” reliability estimates, which we define as estimates of either 50 or 100. These values marked the ends of the reliability input scale. The medians of the animation and slideshow users never span these ends of the scale, though in the median responses of baseline and small multiples users do include estimates of 50 and 100. It is possible that users perceive the difficulty of the perceptual

task (judging 100 plots) and are less likely to report extreme responses; however, this hypothesis requires more formal testing to be evaluated.

In the following section, we use a second domain demonstration (choropleth maps of voting patterns) to further examine the impact of perceptual features on perceptual error and reliability estimate accuracy.

Domain 2: Choropleth Maps of Voting Patterns

American voting behavior is a frequent topic in political science research. Predictive techniques have been applied to outcomes like public opinion formation and vote choice and voter turnout (Ghitza and Gelman 2013). Model predictions are of great interest to the public, especially during elections when choropleth maps displaying predicted voting patterns often appear in large news publications and popular political blogs (e.g., nytimes.com, fivethirtyeight.com, Figure V.I).

In analyzing predictions, large differences between groups are frequently of interest, such as in voting choices between voters with different income levels or among ethnic groups in a given state (e.g., the different preference for Democratic vs. Republican candidates among rich White voters and low-income White, Black, and Hispanic voters, respectively, in the 2012 election). It may be unwise, however, to make much of a large predicted difference for all rich vs. all poor voters in Illinois if there is little correlation between income and voter choices across states. For example, if richer voters in some states show an overall Democratic majority, while in other states this demographic shows an overall Republican majority, then a difference in Illinois may be an artifact of the high variance in votes from richer voters in general. In the case of voting patterns, higher variance for some groups may stem from more indecision among these voters, from the sampling process used to capture their votes, or from other systematic differences. Small differences can also be of great interest to the public, such as when a small advantage of a candidate among one group may be the key to a candidate winning an election, provided

that the small difference is a reliable one that holds consistently in all states (e.g., Obama's advantage among female Hispanic voters in the 2008 election). Hence, the reliability of a prediction is not determined solely by the size of the difference that is predicted, but is also driven by its consistency. We investigate how using comparative sample plots may aid users' understandings of uncertainty in choropleth maps of predicted voting choice and turnout for the 2008 presidential election (Ghitza and Gelman 2013).

DATA AND POSSIBLE OUTCOME GENERATION

We fit the same statistical models detailed in (Ghitza and Gelman 2013), modeling 2008 turnout (whether somebody voted or not) and vote choice (whether they voted for the Democratic candidate Barack Obama or the Republican candidate John McCain) as a function on a set of geographic and demographic covariates. Instead of using marginal maximum likelihood estimation, as detailed in the original paper, we fit the models using Stan (Stan Development Team, 2013), which implements the No U-Turn sampler (Hoffman and Gelman, in press), an extension to Hamiltonian Monte Carlo sampling (Duane et al. 1987). Under this alternative modeling framework, it is trivial to pull samples of our quantities of interest from the posterior distribution, which is the key step in visualizing uncertainty within the currently discussed resampling context. The models were each fit with 6 chains, run for 1000 iterations. We saved the final 500 iterations of each chain and randomly sampled 100 of those saved iterations for inclusion in the experiment. Each saved iteration serves as a hypothetical sample for our comparative sample plots. Each of these iterations is a model that can be used to predict turnout and choice in different states for different demographic groups. Our estimates account for the uncertainty of the statistical model, which in turn takes into account the uncertainty of our estimates due to sample size, along with the survey weights included in each of the surveys we use.

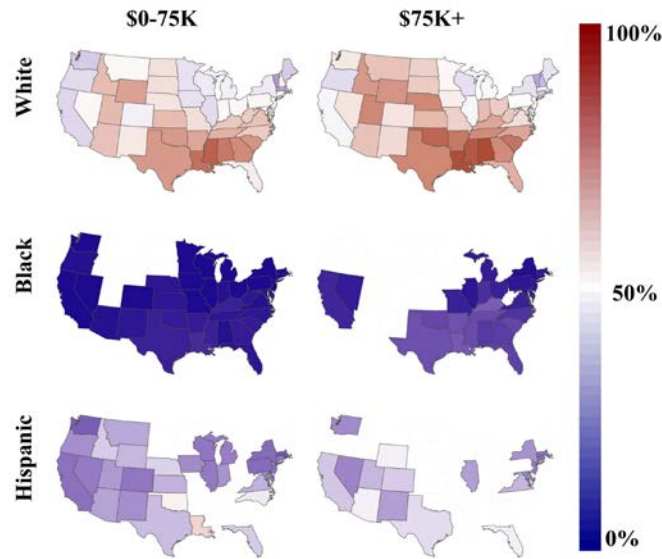


Figure V-VI: A map grid distinguishes voting patterns in the U.S. by ethnicity (rows) and income levels (columns). The data represents predicted voting choices and turnout (Ghitza and Gelman 2013).

VISUALIZATION

We display vote choice for six ethnicity/income subgroups in each of the lower 48 states, among people who voted in the election (Figure V.V). Here, visual consistency in the location of a given group is maintained across plots by fixing the groups' locations within the grid of maps (e.g., the row and column index), and the location of each state within each map. Only colors vary between plots, and the semantics of colors also remain fixed between plots. Our color scheme smoothly transitions from dark blue to white to dark red. This matches the standard color scheme used in most visualizations of voting in the popular press, where blue and red reflect Democratic and Republican vote choice, respectively. The region in which a given majority party vote (Democratic or Republican) is most fragile (the region closest to 50% party vote) is mapped to white, which we found best rendered visible subtle differences between Democratic and Republican majorities that are often of interest in common usage of voting maps. We originally experimented with a blue-purple-red color scheme, but found that the current scheme more clearly shows which candidate has majority support in each group—instead of having to

distinguish between similar shades of purple, the transition at 50% support goes from light blue to light pink.

We note, however, that as a result of this choice, other small differences occurring in other regions of the color spectrum (e.g., two shades of deep blue representing Democratic majorities) might vary in difficulty compared to judging the same difference in the region close to 50%. Using a gray-scale color scheme is an alternative scheme that we expect would produce more uniform perceptual error rates regardless of the region of the scale. However, that mapping also loses the party association with red and blue, potentially complicating the visual decoding task.

The baseline, single visualization displayed the predicted means for each group using the map grid. We did not create a small multiples display as each plot already included a grid of six U.S. maps and thus small multiples would have been too visually complex. We did implement animation and interactive slideshow visualizations.

EXPERIMENTAL PROCEDURE

We again focus our study on how comparative sample plots support typical comparisons using these maps: comparing voting patterns (percentage vote for John McCain) for the same state but different income levels and/or demographic groups. Each group for comparison is therefore a combination of a state, ethnicity, and income level. Questions again took a two-part format, such as:

- **Voting percentage:** “Which group has a higher voting percentage for McCain: white people in Nebraska who earn up to \$75K, or hispanic people in Nebraska who earn up to \$75K?”
- **Reliability of voting percentage.** “Imagine that you have access to 100 samples of data (100 map grids). Estimate the number of times out of 100 that the answer you chose above will have a higher value?”

We again chose comparisons for questions with a goal of including a range in the true reliability. We also ensured a range in the level of true difference in the voting percent-ages across groups, so as to ask about both large and small differences, including several for groups that were predicted to split their votes nearly evenly between the two candidates.

The experimental procedure followed that of the network study with one addition. Users first completed a perceptual calibration task of five trials in which they compared the values (represented by colors) of different states within a single example map. These enabled the user to practice doing the color comparisons and locating states. It also allowed us to gather data on what color differences were noticeable. This information provides a proxy for perceptual difficulty, allowing us to examine whether the challenges associated with a given visual judgment affect the performance of the different presentation types using an empirically-based threshold for a just noticeable difference (JND). The perceptual task included trials where users compared groups with very small actual voting percentage differences. Specifically, we tested a true difference of 0, 0.02, 0.03, 0.04, and 0.10, based on our own visual assessments that the perception threshold would fall somewhere below 0.04.

RESULTS

120 participants completed the task (mean time: 1056.2s). Seven participants with answers below 50 for a reliability question were removed. Again we find that many interactive slideshow visualization users interacted with only one visualization (32%).

We next examined whether presentation type appeared to affect perceptual error and reliability error overall. Perceptual error was slightly higher with the baseline, followed by the slideshow and small multiples, but not significantly so ($F(2,110)=2.1$, $p>0.10$). These results and those mentioned later in this section are shown in Table V.II.

Reliability error showed no significant differences from presentation type either ($F(2,110)=1.0, p>0.10$). The fact that mean signed reliability was close to mean absolute reliability means that people almost always overestimated reliability. The signed reliability error was significantly lower in the animation condition ($F(2,110)=3.2, p<0.05$), meaning that animation users overestimated less, and sometimes underestimated reliability. Again, most baseline users opted to submit reliability estimates (33 out of 35 total).

Given the results of our perceptual calibration task, the true differences between the voting percentages of some groups that we asked about were unlikely to be perceptible to users of the baseline visualization. We noted an increase in errors from an effect size of 0.03 to 0.02 (14% and 63% respectively) that suggests that the effect size that defines a just noticeable difference is located within this interval. As noted above, this interval is approximate, as our color scale ranged from red to blue, and different regions of this scale may vary slightly in the effect size that is detectable. However, we use the lower end (0.02) as the threshold beyond which most users are likely to not perceive a given effect size. We examined perceptual error and reliability estimates for these comparisons between groups where the depiction in the baseline was below the JND threshold. We note that users of the comparative sample plots are more likely to see some plot that shows a larger difference. For the five questions in which the true difference was less than 0.02, the animation resulted in approximately 25% fewer perceptual errors than the baseline ($F(2,110)=4.7, p<0.05, p_{adj}<0.01$). No differences were found for levels of absolute reliability error ($F(2,110)=0.6, p>0.10$). For signed reliability error, we saw marginally lower levels among users of the animation than the baseline by approximately 33% ($F(2,110)=3.3, p<0.05, p_{adj}=0.07$). Turning to consider the seven questions with more perceivable differences (>0.02), we saw no differences in the levels of error based on presentation type ($F(2,110)=0.9, p>0.10$). No differences were found for reliability estimation error, either for absolute error.

Table V-II: Summary of error rates. Bolded cells indicate a significantly lower error rate in a pairwise comparison with the base, single plot condition ($\alpha = 0.05$, TukeyHSD correction).

	Presentation type	Perceptual error (frequency of wrong group selected)	Reliability error (mean of absolute difference from true reliability)	Signed reliability error (mean of difference from true reliability)
Overall	Base	46.6	24.7	16.7
	Anim	39.6	22.7	10.5
	Slideshow	41.5	24.3	15.3
Questions with < JND in baseline	Base	0.61	31.1	24.0
	Anim	0.45	29.3	15.7
	Slideshow	0.55	31.7	23.9
Questions with > JND in baseline	Base	0.36	20.1	6.2
	Anim	0.36	17.9	4.8
	Slideshow	0.32	18.9	5.5

Discussion

Our work proposes that generating and visualizing many hypothetical outcomes can help people make reliability assessments that are more humble and more aligned with statistical models of the reliability of those effects. Our studies confirmed that, in some cases, and with some presentation formats, comparative sample plots had that effect. However, that was not true for all comparisons that users were asked to complete. We briefly discuss limitations of our results and summarize a sensitivity analysis. We then discuss potential reasons for the differential advantages we observed for comparative sample plots.

LIMITATIONS

While this chapter provides important support for apparent improvements to individuals' abilities to recognize reliability levels using comparative sample plots under some conditions, we do not elicit our participants' interpretations of the levels of reliability they support. This leaves ambiguity as to whether most individuals could translate the

reliability information into a more complex assessment of the significance of the data pattern as influenced by data uncertainty. Studying the downstream effects of using comparative sample plots versus other forms of uncertainty representation (or lack thereof) is a critical next step for future work.

As described above, we chose to use a relative frequency format for our reliability elicitation questions based on evidence of this format's advantages over probability formats. However, we note that by framing reliability as the number of times out of 100 that the observed pattern held, it is possible that animation and interactive slider users were given an advantage, as they had access to the same number of hypothetical samples. To test the robustness of our results against a more general format, we ran a sensitivity analysis with the choropleth maps study materials. We substituted the relative frequency format elicitation prompt with a probability-based likert question: "What are the chances that the group your chose above would continue to have a higher percent if more data was gathered?: Random, Slightly better than random, Better than random, Much better than random, Completely certain." We see the same patterns of results for reliability estimation (lower signed reliability estimate error and overall reliability estimate error), albeit by lower margins (10%, 20% advantage of animation).

We did not include any surveys to capture how familiar or engaged users were with the content shown in the visualizations. Especially in the case of the choropleth voting maps, however, it is possible that users with strong political interests or prior knowledge would be better able to use the visualizations based on experience considering specific group comparisons of voting outcomes. Particularly when an effect is subtle, it is reasonable to think that more politically-savvy users would find comparative sample voting maps more useful in identifying small but potentially important differences between groups.

CONSIDERATIONS IN USING COMPARATIVE SAMPLE PLOTS

PRESENTATION FORMAT

The advantages of comparative sample plots depended in part upon the presentation format. In contrast to predictions of some prior work (Rheingans 1992), giving the user interactive control of a slideshow resulted in worse performance than presenting sample plots as small multiples or rapidly in an animation. The range in how many visualizations users of the interactive slideshow viewed suggests that engagement may be required in order for interactivity to benefit users of resample visualizations. Users were more accurate when they viewed more plots, but outside of research labs and professional settings, this might be hard to prompt.

A somewhat surprising finding was that animation and small multiples sometimes led to more accurate perception of presented information compared to the baseline. Plots of hypothetical samples appear to prompt correction of what might otherwise be misperceived patterns or relationships. The greater range in possible outcomes depicted in a set of hypothetical samples compared to a single sample explains why this corrective function is possible. The advantage was primarily visible in cases where the visual difference between the two values was subtle in the plot of the original sample. A thorough perceptual model is required to predict cases where an effect size is unlikely to be perceptible. Users of comparative plots also tended to be more humble in their reliability estimates, giving lower estimates than users of single plots. Occasionally their estimates were too low. Given the overall predisposition of people to be overconfident in the reliability of findings from statistical models and visualizations (Soll and Klayman 2004), a visualization technique that compels people make lower estimates of reliability may be valuable, even if it makes them overly cautious on occasion.

VISUAL STABILITY REQUIREMENT

As described above, a key requirement in using comparative sample plots is to maintain visual stability between hypothetical sample plots. We presented two domain-specific demonstrations that called for different approaches and varying amounts of design manipulation to achieve. In the case of the choropleth maps, the manipulation required to support comparisons across sample plots was minimal. The primary visual attributes that identified data across hypothetical samples were the spatial positions of the predictions for each state-income-ethnicity group. State positions were fixed within each U.S. map, and the positions of the U.S. maps representing different income-ethnicity were fixed in the grid format. This is a default property of the faceting command in R's `ggplot2` package as long as the same attributes are used to create the facets in each hypothetical sample plot. While color was a primary visual variable for displaying the predicting voting percentages, a decision was made prior to plotting to display the full range of values that the percentage could take (0 to 100%). A percentage scale has a standard range and so it is not unusual to present the full 0 to 100% range even if no single sample plot in the set included a prediction that achieved the maximum or minimum value on this scale (e.g., 100%). Thus, we simply supplied the default domain 0 to 100% as in input to the coloring function which we applied to the specific percentage values observed in each hypothetical sample.

In other cases, however, the value of a target attribute may vary between sample plots (e.g., pH levels identified in rainfall as in Diaconis and Efron's (1983) example discussed above, which may not have a standardized scale range). Many default visualization functions that are applied in graphics systems (such as in R and similar mathematical and statistical modeling packages, Microsoft Office Products like Excel, Tableau Software, IBM's Many Eyes, etc.) apply *dynamic scale definition* techniques which results in different realizations of slightly different sets of values, even when the underlying type of measure (e.g., pH level) stays the same. We refer to Wilkinson's (1999) definition of a *scale* as a function that measures the contents of a frame for a set of values

that are to be plotted. The *frame* is the set of tuples that ranges over all possible values that the set of data observations take given some variable of interest. Frames play a critical role in mapping data values to the visual domain by serving as reference structures for how aesthetics are applied to those values (e.g., determining the number and properties of color values that will be used to display the levels of a categorical variable). When the scales used in a visualization are defined using the frame at the level of the single visualization (i.e., the set of values to be shown in a single plot), the resulting properties of that scale will vary with the frame's content (i.e., with those particular data values). Hence, when two different sets of values for the same underlying measure (e.g., *miles per gallon*) are plotted in two separate visualizations, then the properties of the scale computed for each of the two visualizations may differ (e.g., different minimum and maximum on a y -axis, as in Figure V.VI). Had the end-points for the voting percentage scale used in the choropleth map sample plots not been set prior to mapping the data, then maintaining visual stability would entail setting each individual plot's scale properties for the voting percentage to the properties defined by calculating a scale on the union of all the frames across the set of plots. For example, if a value of 99% for the McCain voting percentage was identified as the maximum McCain voting percentage for the entire set of values across all plots, and a minimum value of 2% was observed across the same data set, then each plot should use a color mapping function with a fixed domain that ranges from 2% to 99%.

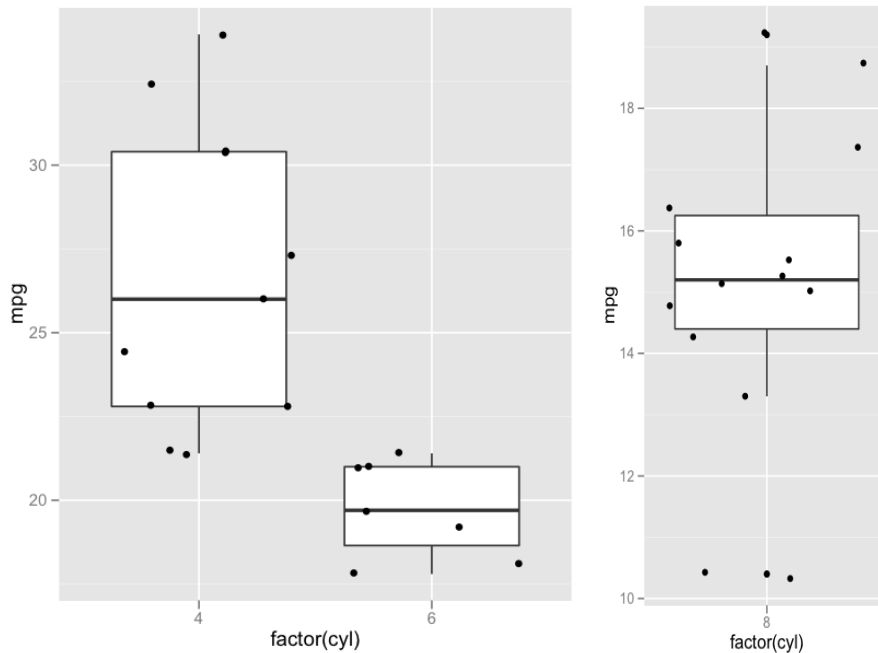


Figure V-VII: Two boxplots showing different subsets of the 'mtcars' R data set, created with the default boxplot jitter function in R's ggplot2. Note the differing y-axis ranges.

The visual design manipulation required to maintain visual stability across the set of hypothetical network diagrams did require overcoming dynamic scale definition, as well as another form of “local” or “single visualization based” optimization. To support comparisons between community memberships across hypothetical diagrams required stabilizing the colors applied to communities across the set of diagrams. It was necessary to fix the number of possible color values to be shown in each diagram using the maximum number of communities identified in any single hypothetical network, so that even if no nodes fell in that community, the colored label for that community could still be shown identically across all plots (overcoming dynamic scale definition as defined above). However, further adjustments were also required as a result of the locally based definitions of the communities themselves (which were defined by default entirely by the nodes that were classified as being classified in the same group for that particular network). We use the term *dynamic attribute definition* to refer to this and other cases in which the presence and properties of an attribute that a user will judge in the plots is

determined by default at the level of a single sample (i.e., the data to be shown in a single hypothetical network diagram determines how many communities are present, and the “definition” of each community which in this case might be thought of as the number of nodes and the properties of nodes that are classified as members of that community). The difference between dynamic attribute definition and dynamic scale definition is that dynamic scale definition results in different properties being used to realize a given measure, but does not change the meaning of the measure itself (e.g., miles per gallon means the same thing in both plots in Figure V.VI). Under default conditions that give rise to dynamic attribute definition, the meaning of an attribute that the user will judge can change (e.g., a community that is created by a group of nodes in one network diagram may not exist in any recognizable form in another network diagram).

Our technique for fixing stabilizing community colors for the network diagram comparative sample plots thus required a method for overcoming the dynamic attribute definition of the communities. In this case, we “forced” a stable attribute definition for each of a finite set of communities by creating a co-community graph from all of the co-classifications of pairs of nodes across all hypothetical networks. The co-community graph represents a “union” across all hypothetical networks, similar to the union operation used to find the range of a scale as described for choropleth maps above. By capturing information about all hypothetical networks to be shown as separate diagrams (100 in this case), the co-community graph provided the information needed to fix the definitions of a finite set of communities which would be consistently colored across network diagrams. Specifically, this was achieved by identifying the graph’s weakly connected components, and then identifying the most stable node in each weakly connected component, and assigning a unique color to each of these nodes. The mapping function used to assign colors to the remaining nodes in each hypothetical network then consisted of simply noting which of the “stable nodes” a new node was classified with, and assigning the stable node’s color to the new node.

We generalize the above visual stability considerations as the requirement that scale definitions be fixed under conditions of dynamic scale definition, and that attribute definitions be fixed under conditions of dynamic attribute definition. The specific methods that are used to overcome either form of dynamic optimization will depend on the functions that are used to create the outputs that are to be visualized within a given modeling and mapping process. However, in general the definitions of both forms of dynamic optimization require first identifying which target attributes will be judged by the user, and then defining the union operation and the functions that will operate on the union to achieve fixed visual scales and definitions for these attributes.

Based on this definition, we note that visual formats which contain more rigid definitions for how a given set of values will be displayed (e.g., how states in a map of the U.S. will be shown relative to one another, how to map paired sets of observations in two dimensions using a scatterplot) are simpler cases for applying comparative sample plots. However, as we have shown through the network diagrams, even in cases where the default modeling and visualization technique lacks rigid attribute definitions and visual realizations of the attributes, it is possible to attain visual stability.

JUDGMENT HEURISTICS

The willingness of single plot users to report reliability estimates even without any direct indicators of uncertainty in the plots supports prior findings that concepts like reliability, uncertainty, and probability are often misunderstood by non-statisticians (e.g., Tversky and Kahneman 1971). We hypothesize that users of single plots followed a heuristic that the reliability of an observed difference was proportional to the size of the observed difference. Sometimes this heuristic works. But if, for example, a mean difference between two datasets is small but the variance within each is very small, the result may still be reliable. When the mean difference is large but the within dataset variability is very large it is also not a reliable difference. Examination of users' reliability

reports for particular questions seem to be broadly consistent with that hypothesis, with more benefits from comparative sample plots when there was a mismatch between effect sizes and variability of those effects. Further experimentation designed to explicitly test this hypothesis would be necessary to assess it rigorously, and is explored in the next chapter.

Conclusion

We have presented comparative sample plots, a technique for presenting uncertainty in visual plots by presenting multiple plots of alternative, hypothetical samples. The technique helps people identify some features that are hard to notice in a single plot. More importantly, it gives them a way to assess the reliability of a feature, whether it would be stably present in slight variations of the original dataset.

Comparative sample plots can be adapted to many data modeling and visualization scenarios where uncertainty visualization is currently challenging. Applying the approach to a new type of plot involves two challenges. The first is generating hypothetical samples. This can be done through bootstrapping (resampling from the observed sample) or by estimating a statistical model's parameters from the observed sample and the using the model to generate additional samples.

The second challenge is to figure out what to make visually stable across plots and what will be allowed to vary. In our comparative network diagrams, we fixed the graph layout and devised a way to make the colors stable for the more stable communities, while allowing the colors to vary between plots for nodes that moved between communities. In our choropleth maps, we fixed the shape and layout of the regions to be colored and fixed the mapping of colors to quantities, with the changing quantities between samples causing the actual colors to vary. We provide a generalization of this requirement by defining dynamic scale definition and dynamic attribute definition.

We have also explored applying comparative sample plots to treemaps of normally distributed data. Streamgraphs, voronoi diagrams, and various other complex visualization formats are also used to compare data values that are subject to uncertainty, making them good candidates for future applications and study. Simpler visualization formats like scatterplots are also amenable to a comparative sample plot approach. It will be important for future work to assess whether animated or interactive presentations of hypothetical samples can outperform existing uncertainty representations like error bars.

CHAPTER VI

Future Directions

Future research in supporting communicative visualization practice among broad audiences will provide further knowledge and tools for data storytelling using narrative formats. Additionally, future work will further develop and validate techniques for scaffolding data reasoning among end-users that are not expert analysts.

ENHANCING NARRATIVE DATA COMMUNICATION

The chapters on visualization rhetoric and sequence in narrative visualization presented in this dissertation demonstrate approaches for operationalizing narrative design techniques. Systematic study is occurring to better understanding the interaction between visualization and text, and to model this interaction in automated tools. Systems have been presented for automating annotation of visualizations for communicative and analysis purposes (Hullman et al. 2013, Kandogan 2012). A task for future work is to conduct user studies that complement these design foci via a model for predicting how text references will impact interpretations of a visualization. For example, controlled studies could be used to assess the extent to which text framing that suggests an interpretation of data that is *not* supported by the visualization can still shape interpretations. Prior study has concluded that graphical plus text descriptions of social information presented with a visualization can shape new users perceptions of the visualized data, even when the information is biased (Hullman et al. 2011). The broader class of framing techniques suggested by the visualization rhetoric framework pose similar

questions around the impacts on interpretation of more traditional forms of textual rhetorical framing, like rhetorical questions and suggestive metaphors.

Another promising area for future work relates to tools for supporting data sharing, such as the sharing of visualizations in social media like personal blogs (Danis et al. 2008). A key challenge in this space is to preserve provenance information to support accurate interpretations of the data as it moves between online contexts.

The chapter on sequence modeling highlights the importance of presentation order for message conveyance and in particular for supporting associations between distinct data representations. This chapter motivates future research in context-adaptive animation of transitions in communicative visualization presentations for narrative formats like interactive slideshows. Current formats provide a simple linear sequence for visualizations, and do not typically tailor the possible next steps from a given visualization slide based on the individual's interaction trajectory. Future work will explore features that enable more possible paths through an interactive slideshow, such as through up and down buttons. These suggestions could be customized based on the prior interaction sequence of an end-user. For example, a user who has viewed visualizations that tend to focus on economic statistics could be presented with the choice of viewing subsequent visualizations that provide more granular results on these statistics. This form of customized sequencing would reduce the typical transition cost between visualizations.

SCAFFOLDING DATA REASONING

Our work on comparative sample plots motivates further exploration of more direct, interactive methods for visualizing uncertainty, as an alternative to abstract graphical annotations like error bars. Future research will explore how enabling interaction with uncertainty representations, such as hypothetical sample plots, supports better understanding. Additionally, there is a need for the development of perceptual models for describing how visual factors impact an end-users ability to acknowledge

uncertainty. For example, how do different presentations of the same data (scatterplot versus boxplot, for example) impact interpretations of the statistical properties of the data?

Our work on comparative sample plots indicates that heuristics may define how judgments of uncertainty-related concepts like pattern reliability are formed. Future work will explore what other decision strategies influence common visual judgment tasks, such as predictions about the typical behavior of a variable shown in a set of hypothetical samples (e.g., the most likely value for next year's GDP for the U.S. given hypothetical samples based on this year's data).

Other forms of scaffolding for semantic cognition, or the process by which visualization users infer meaning, are also needed to better ensure effective use of visualizations in online environments. One example that has been recently proposed is the use of concrete scale representations to help users understanding scales that include very large or small values (Chevalier et al. 2013). There is an opportunity for creating automated tools that can produce this sort of semantic scaffolding for reasoning about data scale. Another area for future work concerns how individuals arrive at causal understandings of visualized data. Causal hypotheses were often suggested by features of the narrative visualizations examined in creating the visualization rhetoric framework. Future research might explore ways to debias causal interpretations of data, as well as the ways in which graphical formats and other narrative techniques subtly prompt causal hypotheses.

CHAPTER VII

Summary

The goal of this dissertation was to contribute to communicative visualization practice by providing tools and knowledge to help visualization creators negotiate design trade-offs. In studying narrative visualizations, I identified two primary types of design trade-offs: those related to which information to omit versus which information to present in order to achieve storytelling goals, and the trade-off between presenting information simultaneously to support the maximum number of data comparisons versus presenting the information sequentially to gradually convey a message about data. The visualization rhetoric framework presented to address the former trade-off operationalizes many visualization design strategies that I observed to be common in visualization interpretation, but that remain absent from existing design taxonomies. In studying persuasive examples of narrative visualization, I also found that many strategies appear to appeal to contextualized knowledge, such as aesthetic norms or cultural expectations. By acknowledging the importance of these “extra-representational” factors to how a communicative visualization is interpreted, the visualization rhetoric work extends prior research by providing theory for describing the expectations associated with graphical formats, for example (Zacks and Tversky 1999, Best et al. 2001, Ziemkiewicz and Kosara 2009) as well as less explored factors like cultural knowledge.

In studying the impact of presentation sequence in narrative visualization, I found that online audiences strongly preferred visualization transitions that minimized the amount of conceptual change between two visualizations. I also observed systematic preferences for certain types of transitions. Audiences found temporal transitions, in

which one data visualization is followed by a second that is identical except for the time period of the data, to be the easiest to understand. I present the concept of parallel transition structure (sequence compressibility), referring to the repetition of a transition structure. For example, a hierarchical transition (from a high level view to a more specific subset, such as a choropleth map of the North American continent followed by a map of the United States) might be repeated several times in a row, with each new use presenting a new subset of the data (e.g., the prior hierarchical transition is followed by a parallel transition between a map of Europe followed by a map of Germany). I find that this form of repetition in sequenced visualization presentations such as slideshow formats positively impacts end-users' abilities to explain and remember the presentation sequence. The proposed graph approach for automating presentation sequence suggestions is a novel contribution to visualization literature, which has largely overlooked the impact of presentation order on interpretation. While still to be rigorously evaluated with users, my work on extending automatic presentation techniques to consider global (sequence) as well as local (singular visualization) considerations in particular contributes a new perspective on accepted design support approaches that are used in visualization systems like Tableau (Mackinlay et al. 2007, Mackinlay 1986).

The comparative sample plots technique is a promising step for the area of uncertainty visualization, as our study results suggest that the technique can support better recognition of pattern reliability (such as when a pattern is or is not reliable) among broad audiences of visualization users. The results we observe indicate that our technique leads users to be more cautious, and to make fewer errors in perceiving information. In some instances, end-users more accurately judge the reliability of data patterns with comparative sample plots. While the effect sizes of some results are small, the technique is an important contribution to research in uncertainty visualization given the many challenges in getting most individuals to recognize uncertainty effects (e.g., Tversky and Kahneman 1971).

References

- Adamic, L. and Adar, E. Friends and Neighbors on the Web. *Social Networks* 25, (2001), 211–230.
- Aerts, J., Clarke, K., and Keuper, A. Testing Popular Visualization Techniques for Representing Model Uncertainty. *Cartography and Geographic Information Science*, (2003), 249–261.
- Agrawala, M., Li, W., and Berthouzoz, F. Design Principles for Visual Communication. *Communications of the ACM* 54, 4 (2011), 60-69.
- Amar, R., and Stasko, J. Knowledge precepts for design and evaluation of information visualizations. *Visualization and Computer Graphics, IEEE Transactions on*, 11, 4 (2004), 432-442.
- Andersen, P.B. What semiotics can and cannot do for HCI. *Knowledge Based Systems* 14, (2000), 419–424.
- Bal, M. *Narratology: Introduction to the theory of narrative*. University of Toronto Press (1997).
- Barthes, R. *Image-Music-Text*. Hill and Wang, 1978.
- Bateman, S., Mandryk, R.L., Gutwin, C., Genest, A., McDine, D., and Brooks, C. Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. *Proceedings of the 28th international conference on Human factors in computing systems*, (2010), 2573–2582.
- Bavoil, L., Callahan, S., Crossno, P., Freire, J., Scheidegger, C., Silva, C., & Vo, H. Vistrails: Enabling interactive multiple-view visualizations. In *Visualization, 2005. VIS 05. IEEE* (2005), 135-142).
- Bederson, B.B., Shneiderman, B., and Wattenberg, M. Ordered and quantum treemaps: Making effective use of 2D space to display hierarchies. *ACM Trans. Graph.* 21, 4 (2002), 833–854.

- Belia, S., Fidler, F., Williams, J., and Cumming, G. Researchers misunderstand confidence intervals and standard error bars. *Psychological methods* 10, 4 (2005), 389–396.
- Bertin, J. *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press, 1984.
- Best, L, Smith, L, & Stubbs, D. Graph use in psychology and other sciences. *Behavioural processes*, 54, 1 (2001), 155-165.
- Borgo, R., Abdul-Rahman, A., Mohamed, F., Grant, P. W., Reppa, I., Floridi, L., & Chen, M. An empirical study on using visual embellishments in visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 18, 12 (2012), 2759-2768.
- Borkin, M. Vo, A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. What makes a visualization memorable?. *Visualization and Computer Graphics, IEEE Transactions on*, 19, 12 (2013), 2306-2315.
- Bransford, J, & Schwartz, D. Rethinking transfer: A simple proposal with multiple implications. *Review of research in education*, 24, (1999), 61-100.
- Bruls, M., Huizing, K., and Wijk, J. van. Squarified Treemaps. *In Proceedings of the Joint Eurographics and IEEE TCVG Symposium on Visualization*, Press (1999), 33–42.
- Buja, A., Cook, D., Hofmann, H., et al. Statistical inference for exploratory data analysis and model diagnostics. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 367, 1906 (2009), 4361–4383.
- Buttenfield, B.P. Representing Data Quality. *Cartographica: The International Journal for Geographic Information and Geovisualization* 30, 2 (1993), 1–7.
- Buxton, B. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2007.
- Card, S.K., Mackinlay, J., and Shneiderman, B. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, 1999.
- Census Bureau. Source and Accuracy of Estimates for Income, Poverty, and Health Insurance Coverage in the United States: 2012. Available at http://www.census.gov/hhes/www/p60_245sa.pdf. Retrieved Nov. 22, 2013.
- Chandler, D. *Semiotics: The Basics*. Routledge, 2001.

- Chen, Y., Yang, J., and Ribarsky, W. Toward effective insight management in visual analytics systems. *Visualization Symposium, IEEE Pacific*, IEEE Computer Society (2009), 49–56.
- Chong, D. and Druckman, J.N. Framing Theory. *Annual Review of Political Science* 10, 1 (2007), 103–126.
- Chua, H.F., Yates, J., and Shah, P. Risk avoidance: Graphs versus numbers. *Memory & cognition* 34, 2 (2006), 399.
- Chuah, M.C. and Roth, S.F. Visualizing common ground. In *Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on* (2003), 365-372.
- Cialdini, R. *Influence: Science and Practice*. (5th edition). Boston: Pearson. (2009).
- Clark, H., and Brennan S. Grounding in communication. In Resnick LB, Levine RM and Teasley SD (Eds). *Perspectives on Socially Shared Cognition*. APA Books: Washington, (1991), 127–149.
- Cleveland, W.S., Diaconis, P., and McGill, R. Variables on scatterplots look more highly correlated when the scales are increased. *Science* 216, 4550 (1982), 1138–1141.
- Cleveland, W., and McGill, R. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79, 387 (1984), 531-554
- Cleveland, W.S. *The Elements of Graphing Data*. Hobart Press, 1994.
- Conati, C. and Maclaren, H. Exploring the Role of Individual Differences in Information Visualaization.
- Corbett, E.P.J. and Connors, R.J. *Classical Rhetoric for the Modern Student*. Oxford University Press, USA, 1998.
- Cross, N. Expertise in design: An overview. *Design Studies* 25, 2 (2004), 427-441.
- Demiralp, C., Scheiddegger, C., Kindlmann, G., Laidlaw, D., and Heer, J. *IEEE Computer Graphics and Applications* (2014).
- Diaconis, P. Theories of Data Analysis: From Magical Thinking Through Classical Statistics. In D.C. Hoaglin, F. Mosteller and J.W. Tukey, eds., *Exploring Data Tables, Trends, and Shapes*. John Wiley & Sons, Inc., 2006, 1–36.

- Diakopoulos, N., Kivran-Swaine, F., and Naaman, M. Playable Data: Characterizing the Design Space of Game-y Infographics. (2011).
- DiMicco, J., Karhalios, K., and McKeon, M. Telling Stories with Data. Workshop at *IEEE InfoVis 2010* (2010).
- DiMicco, J., Diakopoulos, N., Hullman, J., Karhalios, K., and Perer, A. Telling Stories with Data II: The Next Chapter. Workshop at *IEEE InfoVis 2011* (2011).
- Dorst, K., and Cross, N. Creativity in the design process: co-evolution of problem-solution. *Design studies*, 22, 5 (2001), 425-437.
- Dow, S.P., Glassco, A., Kass, J., Schwarz, M., Schwartz, D.L., and Klemmer, S.R. Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Trans. Comput.-Hum. Interact.* 17, 4 (2010), 18:1-18:24.
- Ehlschlaeger, C.R., Shortridge, A.M., and Goodchild, M.F. Visualizing spatial data uncertainty using animation. *Computers & Geosciences* 23, 4 (1997), 387-395.
- Evans, J. In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences* 7, 10 (2003), 454-459.
- Fagerlin, A., Wang, C., and Ubel, P.A. Reducing the influence of anecdotal reasoning on people's health care decisions: is a picture worth a thousand statistics? *Medical decision making: an international journal of the Society for Medical Decision Making* 25, 4 (2005), 398-405.
- Gershon, N., & Page, W. What storytelling can do for information visualization. *Communications of the ACM*, 44, 8 (2001), 31-37.
- Harrower, M., and Brewer, C. ColorBrewer. org: an online tool for selecting colour schemes for maps. *Cartographic Journal, The*, 40, 1 (2003), 27-37.
- Heer, J. and Agrawala, M. Design considerations for collaborative visual analytics. *Information Visualization* 7, 1 (2008), 49-62.
- Heer, J. and Bostock, M. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. *Proceedings of the 28th international conference on Human factors in computing systems*, (2010), 203-212.

- Heer, J. and Boyd, D. Vizster: Visualizing Online Social Networks. *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*, IEEE Computer Society (2005), 5–.
- Heer, J., Mackinlay, J., Stolte, C., and Agrawala, M. Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1189 –1196.
- Heer, J. and Robertson, G. Animated Transitions in Statistical Data Graphics. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1240–1247.
- Heer, J., Viégas, F.B., and Wattenberg, M. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Communications of the ACM* 52, 1 (2009), 87–97.
- Hegarty, M., and Waller, D. Individual Differences in Spatial Abilities. In P. Shah and A. Miyake (Eds), *The Cambridge handbook of visuospatial thinking*. Cambridge University Press, 2005.
- Hegarty, M., Kriz, S., & Cate, C. The roles of mental animations and external animations in understanding mechanical systems. *Cognition & Instruction*, 21, 2003.
- Huff, D. *How to Lie with Statistics*. W. W. Norton & Company, 1993.
- Hullman, J., Adar, E., and Shah, P. The impact of social information on visual judgments. ACM Press (2011), 1461.
- Hullman, J., Adar, E., and Shah, P. Benefitting InfoVis with Visual Difficulties. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011).
- Hullman, J. and Diakopoulos, N. Visualization Rhetoric: Framing Effects in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2231–2240.
- Hullman, J., Diakopoulos, N., and Adar, E. Contextifier: Automatic Generation of Annotated Stock Visualizations. ACM Press (2013).
- Hutchins, E. *Cognition in the Wild*. MIT Press, Cambridge, MA, 1994.
- Inbar, O., Tractinsky, N., and Meyer, J. Minimalism in information visualization: attitudes towards maximizing the data-ink ratio. *Proceedings of the 14th European conference on Cognitive ergonomics: invent! explore!*, ACM (2007), 185–188.

- Jain, R., & Slaney, I. M. Micro Stories. (2011). Available at <http://dlib2.computer.org/portal/web/computingnow/content?g=53319&type=article&urlTitle=micro-stories-and-mega-stories>.
- Jankun-Kelly, T.J., Ma, K.-L., and Gertz, M. A Model and Framework for Visualization Exploration. *IEEE Transactions on Visualization and Computer Graphics* 13, 2 (2007), 357–369.
- Jenks, G. and Caspall, F.C. Error on Choroplethic Maps: Definition, Measurement, Reduction. *Annals of the Assoc. of American Geographers* 61, 2 (1971), 217–244.
- Jones, J., Chang, R., Butkiewicz, T., and Ribarsky, W. Visualizing uncertainty for geographical information in the global terrorism database. (2008), 69830E–69830E.
- Kahneman, D., and Frederick, S. Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin and D. Kahneman (Eds), *Heuristics of Intuitive Judgment: Extensions and Applications*. Cambridge University Press, 2002.
- Kahneman, D., Slovic, P., and Tversky, A. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press, 1982.
- Kahneman, D., and Tversky, A. Subjective probability: A judgment of representativeness. *Cognitive psychology* 3, 3 (1972), 430-454.
- Kahneman, D. and Tversky, A. On the Psychology of Prediction. *Psychological Review* 80, 4 (1973), 237-251.
- Kandogan, E. Just-in-time annotation of clusters, outliers, and trends in point-based data visualizations. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on* (2012), 73-82.
- Kintsch, W., Mandel, T.S., and Kozminsky, E. Summarizing scrambled stories. *Memory & Cognition* 5, 5 (1977), 547–552.
- Kolodner, J. and Wills, L.M. Case-Based Creative Design. In *AAAI Spring Symposium on AI and*, (1993), 50–57.
- Kosara, R. Visualization Criticism - The Missing Link Between Information Visualization and Art. *Information Visualisation, International Conference on*, IEEE Computer Society (2007), 631–636.

- Kosara, R., & Mackinlay, J. Storytelling: The Next Step for Visualization. *Computer (Special Issue on Cutting-Edge Research in Visualization)* 46, 5, (2013), 44–50.
- Kosara, R., & Ziemkiewicz, C. Do Mechanical Turks dream of square pie charts?. In *Proceedings of the 3rd BELIV'10 Workshop: BEyond time and errors: novel evaluation methods for Information Visualization* (2010).
- Kosslyn, S.M. Understanding charts and graphs. *Applied Cognitive Psychology* 3, 3 (1989), 185–225.
- Kovach, B. and Rosenstiel, T. *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect, Completely Updated and Revised*. Three Rivers Press, 2007.
- Lakoff, G. *Women, Fire, and Dangerous Things*. University Of Chicago Press, 1990.
- Larkin, J.H. and Simon, H.A. Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science* 11, 1 (1987), 65–100.
- Latour, B. Visualization and Cognition: Drawing Things Together. In H Kucklick (ed.) *Knowledge and Society Studies in the Sociology of Culture Past and Present*, Jai Press vol. 6, 1-40.
- Leitner, M. and Buttenfield, B.P. Guidelines for the Display of Attribute Certainty. *Cartography and Geographic Information Science* 27, 1 (2000), 3–14.
- Liu, Z. and Stasko, J.T. Mental models, visual reasoning and interaction in information visualization: a top-down perspective. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 999–1008.
- MacEachren, A.M., Robinson, A., Gardner, S., Murray, R., Gahegan, M., and Hetzler, E. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic. INFORMATION SCIENCE* 32, (2005), 160.
- MacEachren, A.M. Visualizing uncertain information. *Cartographic Perspectives* 13, 13 (1992), 10–19.
- Mackinlay, J. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)* 5, (1986), 110–141.
- Manovich, L. Cultural analytics: Visualizing cultural patterns in the era of more media. *Domus* (923 (2009)).

- Mayer, R. Multimedia learning: Are we asking the right questions? *Educational Psychologist* 32 (1997), 1-19.
- McCloud, S. *Understanding Comics*. HarperCollins (1994).
- Micallef, L., Dragicevic, P., & Fekete, J. Assessing the Effect of Visualizations on Bayesian Reasoning Through Crowdsourcing. *Visualization and Computer Graphics, IEEE Transactions on*, 18, 12 (2012), 2536-2545.
- Miller, G. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63, 2 (1956), 81.
- Miller, N. and Campbell, D.T. Recency and primacy in persuasion as a function of the timing of speeches and measurements. *The Journal of Abnormal and Social Psychology* 59, 1 (1959), 1-9.
- Munzner, T. A nested model for visualization design and validation. *Visualization and Computer Graphics, IEEE Transactions on*, 15, 6 (2009), 921-928.
- Nelson, T.E. and Oxley, Z.M. Issue Framing Effects on Belief Importance and Opinion. *The Journal of Politics* 61, 4 (1999), 1040-1067.
- Norman, D.A. *The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex, and Information Appliances Are the Solution*. The MIT Press, 1999.
- North, C. Toward measuring visualization insight. *Computer Graphics and Applications, IEEE*, 26, 3 (2006), 6-9.
- North, C., Chang, R., Endert, A., Dou, W., May, R., Pike, B., & Fink, G. Analytic provenance: process+ interaction+ insight. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems* (2011), 33-36.
- Olston, C. and Mackinlay, J.D. Visualizing Data with Bounded Uncertainty. *University of California, Santa Cruz*, (2002), 37-40.
- Pang, A., Wittenbrink, C., & Lodha, S. Approaches to uncertainty visualization. *The Visual Computer*, 13, 8 (1997), 370-390.
- Pennington, N. and Hastie, R. Explaining the evidence: Tests of the Story Model for juror decision making. *Journal of Personality and Social Psychology* 62, 2 (1992), 189-206.

- Perer, A. and Shneiderman, B. *Integrating Statistics and Visualization: Case Studies of Gaining Clarity during Exploratory Data Analysis*. 2008.
- Pousman, Z., Stasko, J. T., & Mateas, M.. Casual information visualization: Depictions of data in everyday life. *Visualization and Computer Graphics, IEEE Transactions on*, 13, 6 (2007), 1145-1152.
- Purchase, H., Hoggan, E., and Gorg, C. How important is the Mental map? an empirical investigation of a dynamic graph layout algorithm. In *Graph Drawing*, 2006.
- Robertson, G., Card, S., and Mackinlay, J. Information visualization using 3D interactive animation, *Comm. of the ACM*, vol. 36, 1993.
- Robertson, G., Fernandez, R., Fisher, D., Lee, G., and Stasko, J. Effectiveness of Animation in Trend Visualization, *IEEE TVCG*, vol. 14, Nov. 2008.
- Sack, W. Aesthetics of information visualization. *Context Providers: Conditions of Meaning in Media Arts* (2011).
- Sadoski, M., Paivio, A., (eds, N.J.U., Models, T., and Of, P. Reference: A Dual Coding Theoretical Model of Reading.
- Saris, W.E. and Sniderman, P.M. *Studies in public opinion: attitudes, nonattitudes, measurement error, and change*. Princeton University Press, 2004.
- Schank, R.C. and Abelson, R.P. *Scripts, Plans, Goals, and Understanding: An Inquiry Into Human Knowledge Structures*. Psychology Press, 1977.
- Schön, D. Designing as reflective conversation with the materials of a design situation. *Knowledge-Based Systems*, 5, 1 (1992), 3-14.
- Schnotz, W. Commentary: Towards an integrated view of learning from text and visual displays. *Educational psychology review*, 14, 1 (2002), 101-120.
- Schwarz, N., Hippler, H.-J., Deutsch, B., and Strack, F. Response Scales: Effects of Category Range on Reported Behavior and Comparative Judgments. *Public Opinion Quarterly* 49, 3 (1985), 388 –395.
- Segel, E. and Heer, J. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1139–1148.
- Seife, C. *Proofiness: The Dark Arts of Mathematical Deception*. Viking Adult, 2010.

- Shaw, A., Horton, J., and Chen, D. *Designing Incentives for Inexpert Human Raters*. Social Science Research Network, Rochester, NY, 2011.
- Scheidegger, C., Vo, H., Koop, D., Freire, J., & Silva, C. Querying and creating visualizations by analogy. *Visualization and Computer Graphics, IEEE Transactions on*, 13, 6, (2007), 1560-1567.
- Shneiderman, B. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Trans. Graph.* 11, 1 (1992), 92–99.
- Skeels, M., Lee, B., Smith, G., and Robertson, G. Revealing uncertainty for information visualization. *Proceedings of the working conference on Advanced visual interfaces*, ACM (2008), 376–379.
- Snedecor, G.W. and Cochran, W.G. *Statistical Methods*. Iowa State University Press, 1989.
- Souza, C.S. de. *The Semiotic Engineering of Human-Computer Interaction*. The MIT Press, 2005.
- Spence, I. Visual Psychophysics of Simple Graphical Elements. *Journal of Experimental Psychology: Human Perception and Performance* 16, 4 (1990), 683–692.
- Spence, I. & Lewandowsky, S. Graphical perception. *Modern methods of data analysis* (1990): 13-57.
- Spence, I., & Lewandowsky, S. Displaying proportions and percentages. *Applied Cognitive Psychology*, 5, 1 (1991), 61-77.
- Sprague, D., and Tory, M. Motion in casual InfoVis and the interrelation between personality, performance, and preference (2009).
- Stevens, S. On the psychophysical law. *Psychological Review* 64, 3 (1957), 153–181.
- Stevens, S. On predicting exponents for cross-modality matches. *Perception & Psychophysics*, 6, 4 (1969), 251-256.
- Thomas, J.J. and Cook, K.A., eds. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr, 2005.
- Thorndyke, P.W. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology* 9, 1 (1977), 77–110.
- Tufte, E.R. *The Visual Display of Quantitative Information*. Graphics Press, 1983.

- Tufte, E.R. *Beautiful Evidence*. Graphics Press, 2006.
- Tversky, A. and Kahneman, D. Belief in the Law of Small Numbers. *Psychological Bulletin* 76, 2 (1971), 105-110.
- Tversky, A. and Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cognitive psychology* 5, 1 (1973), 207-232.
- Tversky, A. and Kahneman, D. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review* 90, 4 (1983), 293-315.
- Van Dam, A., Feiner, S. K., Hughes, J. F., & Phillips, R. L. *Introduction to computer graphics* 55. Reading: Addison-Wesley (1994).
- Vande Moere, A., and Purchase, H. On the role of design in information visualization. *Information Visualization*, 10, 4 (2011), 356-371.
- Viegas, F.B., and Wattenberg, M. Communication-Minded Visualization: A Call to Action. IBM Research (Technical Report), 2001.
- Viegas, F.B., Wattenberg, M., Van Ham, F., Kriss, J., and McKeon, M. Manyeyes: a site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics*, (2007), 1121-1128.
- Viegas, F.B., Wattenberg, M., McKeon, M., Van Ham, F., and Kriss, J. Harry potter and the meat-filled freezer: A case study of spontaneous usage of visualization tools. *Proc. HICSS*, (2008).
- Wainer, H. *Picturing the Uncertain World: How to Understand, Communicate, and Control Uncertainty through Graphical Display*. Princeton University Press, 2009.
- Wang Baldonado, M.Q., Woodruff, A., and Kuchinsky, A. Guidelines for using multiple views in information visualization. *Proceedings of the working conference on Advanced visual interfaces*, ACM (2000), 110-119.
- Ware, C. *Visual Thinking: for Design*. Morgan Kaufmann, 2008.
- Wattenberg, M. and Kriss, J. Designing for social data analysis. *IEEE Transactions on Visualization and Computer Graphics*, (2006), 549-557.
- Wattenberg, M. Visualizing the stock market. *CHI '99 Extended Abstracts on Human Factors in Computing Systems*, ACM (1999), 188-189.

- Wickham, H., Cook, D., Hofmann, H., and Buja, A. Graphical inference for infovis. *Visualization and Computer Graphics, IEEE Transactions on* 16, 6 (2010), 973–979.
- Wickham, H. A Layered Grammar of Graphics. *Journal Of Computational And Graphical Statistics* 19, 1 (2010), 3–28.
- Wilensky, R. *Understanding goal-based stories*. Garland Pub., 1980.
- Yee, K., and Fisher, D. Animated Exploration of Dynamic Graphs with Radial Layout, *INFOVIS*, 2001.
- Zacks, J. and Tversky, B. Bars and lines: a study of graphic communication. *Memory & Cognition* 27, 6 (1999), 1073–1079.
- Ziemkiewicz, C. and Kosara, R. Preconceptions and individual differences in understanding visual metaphors. *Computer Graphics Forum (Proc. EuroVis)*, 2009.
- Ziemkiewicz, C., Crouser, R. J., Yauilla, A. R., Su, S. L., Ribarsky, W., & Chang, R. How locus of control influences compatibility with visualization style. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on* (2011), 81-90.
- Ziemkiewicz, C., & Kosara, R. Embedding information visualization within visual representation. In *Advances in Information and Intelligent Systems* (2009), 307-326.