A Deeper Understanding of Sequence in Narrative Visualization

Jessica Hullman, Steven Drucker, Nathalie Henry Riche, Bongshin Lee, Danyel Fisher, and Eytan Adar

Abstract—Conveying a narrative with visualizations often requires choosing an order in which to present visualizations. While evidence exists that narrative sequencing in traditional stories can affect comprehension and memory, little is known about how sequencing choices affect narrative visualization. We consider the forms and reactions to sequencing in narrative visualization presentations to provide a deeper understanding with a focus on linear, "slideshow-style" presentations. We conduct a qualitative analysis of 42 professional narrative visualizations to gain empirical knowledge on the forms that structure and sequence take. Based on the results of this study we propose a graph-driven approach for automatically identifying effective sequences in a set of visualizations to be presented linearly. Our approach identifies possible transitions in a visualization set and prioritizes local (visualization-to-visualization) transitions based on an objective function that minimizes the cost of transitions from the audience perspective. We conduct two studies to validate this function. We also expand the 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 discuss how these insights can guide the design of narrative visualization and systems that support optimization of visualization sequence.

Index Terms—Data storytelling, narrative visualization, narrative structure.

1 Introduction

Storytelling is now a focus in visualization research and practice, as the study of narrative visualizations (e.g., [13][26]), development of automated data storytelling tools (e.g., [21]), and proliferation of narrative visualizations in news media attest. Supporting data story creation among those who may lack training in visualization design is particularly valuable, as these users may have domain expertise that allows them to produce useful insights into public data.

Story creation involves sequential processes of context definition, information selection, modality selection, and choosing an order to effectively convey the intended narrative. In using visualizations to tell a story, the events of interest are patterns in data sets represented in visualizations. A typical creation process involves using a tool like Tableau [31] or Microsoft Excel [19] to visually analyze data, and to generate visualizations via vector graphics or images for presentation. The story creator then must decide how to thread the representations into a compelling yet understandable sequence.

This structuring of evidence, combined with the choice of appropriate rhetorical strategies, is 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 [22], and the ability to recall the information later [32].

Research in narrative visualization points to visualization features that afford storytelling including guided emphasis (e.g., spatial ordering or partial animation [13][27] and structures for reader-driven storytelling (e.g., the Drill-down story [27]). 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 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? 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. We focus in particular on how 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 an outline of how automatic sequencing could be approached in designing systems to help non-designers navigate structuring decisions in creating narrative visualizations, such as by semi-automatically identifying and presenting more "effective" visualization sequences during a design session. 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 demonstrate insights for guiding the design of narrative visualizations and for informing systems to support visualization sequencing. These 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 conclude by discussing the implications of our findings

- *Jessica Hullman and Eytan Adar are with the University of Michigan. E-mail: {jhullman, eadar}@umich.edu.*
- *Steven Drucker, Nathalie Henry Riche, Bongshin Lee, and Danyel Fisher are with Microsoft Research. E-mail: {sdrucker, nathalie.henry, bongshin, danyelf}@microsoft.com.*

Manuscript received 31 March 2013; accepted 1 August 2013; posted online 13 October 2013; mailed on 4 October 2013.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.a

for the design of linear-style narrative visualization presentations and tools to support non-designers in creating narrative visualizations.

2 Related Work

2.1 Narrative Sequencing 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 [2]. 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., [32]). 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 [32]). Our global sequencing patterns study (Sect. 5.2) takes motivation from this approach. Pennington and Hastie [22] 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 [36]). 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.

2.2 Narrative Visualization

Existing research around narrative visualization creation includes systems for visualizing and sharing public and personally relevant data (e.g., [33]); supporting new interaction styles from rich media artifacts (e.g., [20]); and design space taxonomies to describe techniques used in exemplar professional artifacts [13][27]. The latter studies provide generalized advice for designing narrative visualizations. In addition to noting narrative formats that appear in interactive narrative visualizations such as the interactive slideshow [27], these studies describe how prioritization and sequencing of information can occur through spatial ordering, animation, and suggestive default views, among others [13] [27]. Yet, despite giving examples of successful structuring techniques, there is a lack of clearly outlined measures that creators can use to find the best sequence for visualizations among multiple possible sequences. We extend work in narrative visualization via an understanding of sequence informed by empirical analysis of professional visualizations as well as user validated measures and transition characterizations.

Prior work on visualization transitions includes Heer and Robertson's [10] 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.

2.3 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 [8] 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 [16]. 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 [7]. 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.

3 Patterns in Narrative Visualization Sequence

3.1 Motivating Scenario

Many narrative visualizations that researchers point to 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., [13][27]). 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 paper, 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 [16]. We next 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 in Sect. 4.

3.2 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)?

3.2.1 Study Design

42 narrative visualizations created between 2006 and 2012 were compiled (full list in supplementary file). We seeded the set with visualizations in an independently-curated sample of New York Times (NYT) and Guardian interactives [23]. Additional examples came from visualization blogs and repositories (e.g., visualizing.org) and well-known news sources (e.g., BBC). We included only visualizations with non-ambiguous sequencing cues like numbered slides or steps across linked views, a "Next," "→," or "Continue" button, or a "Play" button for a self-running video or slideshow. These features had to occur without additional navigational choices. Interactive slideshows formed the largest format in our sample (23/42), with other presentations including animated data videos (7/42) and 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 are fairly unambiguous in a slideshow-style presentation, the constituent states of smooth animated narrative visualizations are more difficult to identify. A *visualization state* has been defined as a set of parameters applied to data [14], or the settings of interface widgets in a visualization environment along with the application content [11]. We define a *narrative visualization state* as an informationally-distinct visual representation and transitions as state changes after [10]. 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 often divides into one state per slide, a single slide can represent multiple states if it contains animation within single numbered slides. Rather than counting the states in smooth animations, we focus on 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.

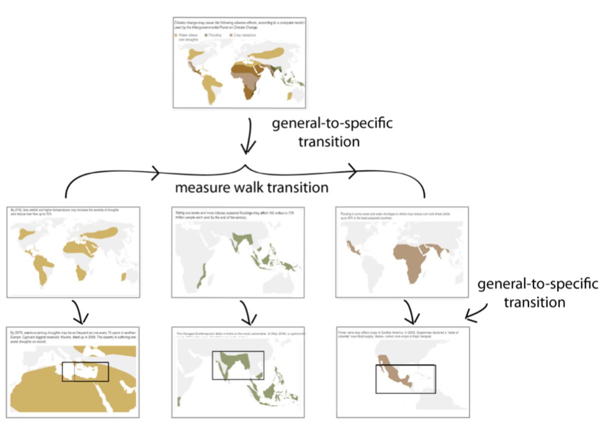

Fig. 1. Parallelism in sequencing in the NYT's "Copenhagen: Emissions, Treaties, and Impacts: Possible Impacts" interactive [3]. Three general-to-specific transitions detail three possible climate outcomes (drought, flooding, crop shortage), which at a higher level comprise a measure walk sequence.

Table 1. Transition types with 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	(29/42)	88.1%
	Reverse chronological	(11/42)	
	Future chronological	(12/42)	
Causal	Explicit Cause	(7/42)	23.8%
	Alternative Reality	(3/42)	
Granularity	General to Specific	(28/42)	71.4%
	Specific to General	(16/42)	
Comparison	Dimension Walk	(20/42)	64.3%
	Measure Walk	(19/42)	
Spatial	Spatial Proximity	(10/42)	23.8%

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.

3.2.2 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 (see Table 1). 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 asked in one state is followed by a visualization that answers that question. *Temporal* transitions involve orderings of visualization states based on a time variable associated with the data in each (see Fig. 2). These include standard *chronology* 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*). 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 explicit mention that income influences voting. *Granularity* transitions order visualization states based on the level of detail or degree of filtering of data they involve, such as from an overview plot of industry stock performances to a plot focused on stocks in a single industry (see also Fig.

1, 2). Finally, in *Comparison* transitions, either the independent variable (i.e., dimension) or the dependent variable (i.e., measure) is held constant while the other is changed. This can show how populations differ for a given outcome, or provide multiple perspectives on a single population or dimension, respectively (see Fig. 2). *Spatial* transitions are a subset of comparison transitions where the same dependent variable is shown for different spatial areas in sequence. Table 1 lists transition types and the sample frequency.

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 labelled automatically given simple matching of data variables against common temporal formats and sorting. Similarly, visualizations of data with hierarchical variables or spatial coordinates could be labelled automatically for Granularity and Spatial transitions, respectively. Comparison transitions can be inferred either by relying on conventions in existing systems for distinguishing dimensions from measures (e.g., [30]) or by using conventions of the graphical format to infer which variable is the independent dimension and which is the dependent measures, such as by looking at the axes mappings in a scatterplot, where the *x-axis* is typically reserved for a independent measure. We focus on implicit types in the sequencing approach that we outline as these types can be inferred more easily.

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 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 [5]. An example 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 1). We explore the impact of parallelism on user ratings and comprehension of visualization narratives in a study presented in Sect. 5.2.

The insights that 1) local transitions are frequently based on a small number of changes to data dimensions and 2) parallelism of sequence patterns can be observed at the global level leads to a general observation that *maintaining consistency across transitions* appears to be an important principle in structuring visualization storytelling. 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 partial animation, a technique for easing the comprehensibility of transitions [10]. Maintaining consistency through gradual changes between consecutive visualizations in narrative presentations enables comparisons between slides, helping to balance the necessary juxtapositions that must occur in order for the story to proceed not unlike animating a transition can support understanding (e.g., [10]). 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.

4 An Algorithmic Approach to Visualization Sequence Support

We propose a graph-driven approach to finding effective sequences for narrative visualizations informed by our analysis. The 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 in Sect. 3. Inputs and stages are shown in Fig. 2. 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 incorporating further prioritization of some sequence types over others, then validate the approach using user evaluation and explore additional optimizations in Sect. 5.

4.1 Defining Data Attributes for Transition Labelling

We observed that many explicit transition types surfaced in our study were based on *single changes* to one of the data attributes used to generate a visualization. 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. This aspect of our approach resembles models for visual exploration that describe transformations that occur in pipelines (functions used in visualization reaction) [14]

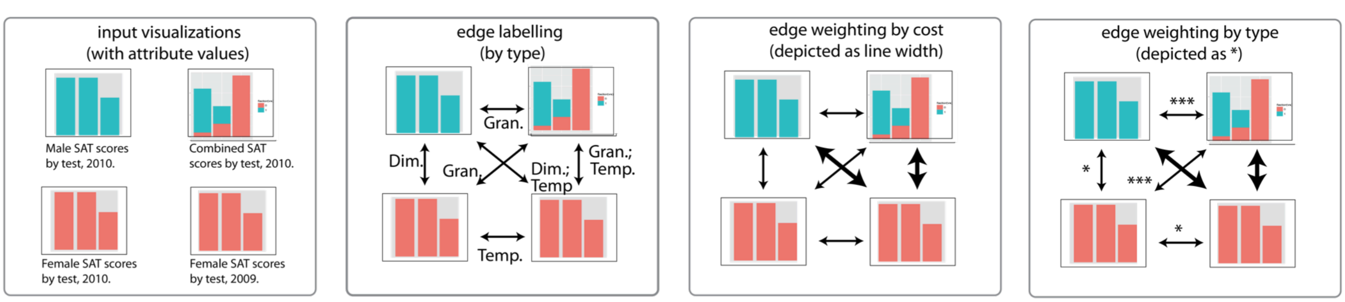

Fig. 2. Diagram of graph-based approach in which visualizations represent nodes. Edges (possible transitions) are labeled by type and weighted using a cost function and type weightings (denoted by * symbols) corresponding to user preferences.

including as directed graphs that can be compared to semi-automatically create new visualizations [26]. 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 modelled the conceptual flow of data between pipeline actions from a system perspective, our interest is primarily in user reactions to conceptual change over transitions.

We begin by considering each visualization state as a node in the graph that is represented using four attributes. These include a dependent (or outcome) variable, an independent variable, a time variable, and a set of hierarchical relations. 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 Granularity 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 granularity changes involve steps between different levels in a data-defined hierarchy, or can be achieved by filtering.

Table 2 relates this schema to common interactive dynamics in visual analytics as defined by [12]. 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 create small multiples displaying related dependent variables for a data group). Table 2 also describes the schema using a standardized data representation - that of the R package ggplot2 [35], which is based on Wilkinson's Grammar of Graphics system for visualization characterization [37].

4.2 Objective Function: Maintaining Consistency

Taking a graph-based approach in which links (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.

4.2.1 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 link (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 vsiualization node into the first visualization node. The more transformations it takes to convert a first visualization to a second, the harder we expect it to be for users to infer a connection between them. This could make comparing the visualizations in a meaningful way more difficult, consistent with research in preserving mental models across transitions [10]. We examine this assumption about transformation cost through user studies in Sect. 5.1.

Formally, transformation cost is the total number of changes to the independent variable, dependent variable, time, and level of granularity 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 2, 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 results 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

Table 2: Data representation, transition types, and relation to common visualization interactions as described in [12] and realization in ggplot2 [35]. The way in which a given transition is realized in these frameworks can vary depending on the properties of the input data set.

Representation	Relevant 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 measures)
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	Granularity	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)

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.

4.2.2 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 as equally different to two visualizations representing a temporal transition? The visual information analysis mantra [29] "overview, zoom & filter, detail on demand" suggests that general-to-specific transitions are preferable, but this has not been empirically evaluated, and other questions remain. How do both of these types compare to a granularity-based transition such as a general-to-specific transition, or a change in the dimension being shown? Systematic preferences for some transitions over others could be incorporated into the above approach using type weightings. We examine perceptions of local transitions types in Sect. 5.1.

4.3 Automatic Global Sequencing

A final question is how particularly effective global sequences can be inferred. 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? A user study in Sect. 5.2 addresses this remaining question.

5 Evaluating User Perceptions of Sequences

How do end-users of linear narrative visualization presentations perceive the types and "costs" of transitions? We examine user perceptions of local transitions types, then consider global strategies.

5.1 Local Transitions: Transformation Cost and Transition Type Weighting

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 granularity 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 granularity transitions over measure transitions, based on the greater conceptual distance between visualizations showing two different dependent variables.

5.1.1 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 [9]. 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

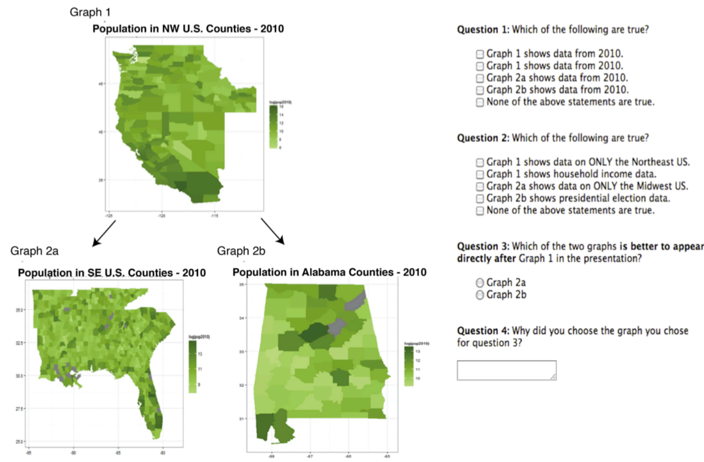

Fig. 3. Experimental task presenting participants with an initial visualiza-

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 (labelled Graph 1) and asks that they choose between the other two visualizations (labelled 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?" 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).

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 granularity.

Cost Constant trials: In 17 Cost Constant HITs, we tested four transition types: temporal (chronological, reverse chronological), comparative dimension walk, comparative measure walk, and granularity 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 granularity 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 granularity 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 [6].

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

Table 3 (left): Two multinomial logits regressing "chosen" transition on transition costs (1, 2, 3) and order indicator, allowing comparison of all three costs. Table 4 (right): Results of three logits regressing "chosen" transition on transition types (spanning all possible comparisons).

Multinomial Logits on "Cost Varying" Trials(N=357)

IV | Baseline	Cost=1	Cost=2
Order Indicator	0.348 (0.249)	0.348 (0.249)
Cost=1	Baseline	2.85 (0.262)***
Cost=2	-2.85 (0.262)***	Baseline
Cost=3	-3.56 (0.725)***	-0.711 (0.766)
L.R. Test	(X^2=15.4)***	(X^2=15.4)***
McFadden R^2	0.10	0.10

Notes: Logit: standard errors are in parentheses. Significant at: * 10 %, ** 5 %, *** 1 % level.

Multinomial Logits on "Cost Constant" Trials (N=338)

IV | Baseline	Measure	Dimension	Granularity
Order Indicator	0.327 (0.116)***	0.327 (0.116)***	0.327 (0.116)***
Measure	Baseline	0.006 (0.177)	0.599 (0.195)***
Dimension	-0.006 (0.177)	Baseline	0.593 (0.206)***
Temporal	0.550 (0.220)***	0.556 (0.200)***	1.15 (0.232)***
Granularity	-0.599 (0.195)***	-0.593 (0.206)***	Baseline
L.R. Test	(X^2=483.0)***	(X^2=483.0)***	(X^2=483.0)***
McFadden R^2	0.53	0.53	0.53

(e.g., same color and shape) unless changes were necessitated by the chart format (e.g., shape changes for different countries in a map).

5.1.2 Experimental Procedure

The Cost Varying and Cost Constant HITs were launched as a combined series of 35 HITs with a \$0.10 reward. Each 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 participant should not consider the quality of the individual visualizations in her choice. Additionally, it was explained that she would start the task with an additional bonus reward of \$0.15. If the participant's choice of visualization matched the visualization chosen by the majority of other workers who saw the same stimuli set, she would retain the full \$0.15; otherwise, they would lose the \$0.15 bonus. This "punishment agreement" incentivization technique has been shown to produce higher quality responses on MTurk [28].

After consenting, a participant had to correctly answer a question about the task goal. She was then presented with three graphs labelled "1," "2a," and "2b" (see Fig. 3). After answering two "information extraction" questions that verified that the participant paid attention, she answered a multiple choice question, "Which of the two graphs is better to appear directly after Graph 1 in the presentation?" where "Graph 2a" and "Graph 2b" were the only choices.

5.1.3 Results

143 total participants 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 particpants 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 to complete the task based on whether transformation cost varied or not (M: 114.6s vs. 121.3s, t=-1.56, p=0.12).

Effects of transition consistency (transformation cost): We first examined whether a lower transformation cost between the two visualizations in a sequence resulted in a preference for that sequence over higher cost alternatives. Table 3 displays the results of two 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 3 differ only in which cost is set to the baseline category. Results indicate that while particpants 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.

Effects of transition types: We next considered whether participants displayed equivalent levels of preference for temporal, comparative, or granularity-based transitions when cost was held constant. Table 4 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 granularity, dimension, or measure transitions (all p<0.01). Both dimension and measure transitions are preferred over granularity 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) > Granularity

Additionally, we see a significant order effect of whether a visualization was in the first or second position from the left in the layout (the order indicator). Hence, other contextual factor (such as bias toward more recently seen visualizations) may influence interactions with narrative visualizations.

5.2 Sequencing: Impacts of Parallelism

In our qualitative study, we observe 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.

5.2.1 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 (relabelled 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 labelled 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 Sect. 3.2 and depicted through examples in Fig. 3: a measure walk and a dimension 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 we 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 both of these sequence types 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 (Fig. 3).

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**: Participants who see between-group sequences will perform better on average on between-group questions.
- **H2b**: Participants 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 functions to capture global sequence preferences may be needed.

5.2.2 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.

Lastly, the participant saw the set of original visualization laid out in random order. She was asked to input the original order of visualizations in the presentation in numbered boxes, in order to test for memory differences based on sequence type.

5.2.3 Results

82 participants completed the task in an average of 711s. Removing those who incorrectly answered the verification question left 73 participants. We first checked whether ratings on the difficulty in ex-

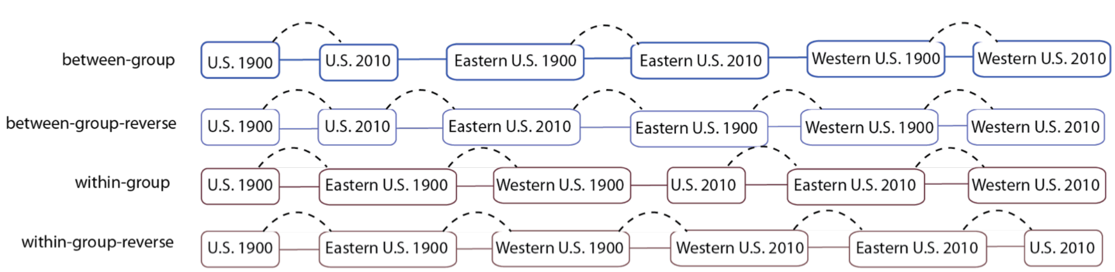

Fig. 4. Global sequences to support different hypothesized comparisons between consecutive visualizations (depicted with dotted lines).

plaining 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 only marginally 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 participants who saw a between-group sequence (including reversed) (M: 0.92 vs. 0.86) and accuracy on within-group questions was higher for participants 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).

6 DISCUSSION

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

6.1.1 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-labelled 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., [34]) and then search the space of global sequences for those that repeat particular motifs.

6.1.2 Limitations

We evaluated temporal and granularity 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 Sect. 4, 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.

6.1.3 Implications and Future Work

Our work describes in depth how supporting narrative sequencing can be systematically approached in visualization systems. Future work should further evaluate how to best combine local transition costs, type weightings, and global constraints like parallelism.

A related question is whether animating a transition (e.g., [10]) can balance the potential negative effects on user perceptions of a costly transition.

By relating our approach to the grammar of graphics [37] and standard visualization interactions [12], 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 explore how the sequence optimization described here can interface with optimizations designed to suggest the most effective individual visualization (e.g., [17][18]), including how potential conflicts between single visualization vs. sequence models can be resolved. We see this is a crucial open question in providing semi-automatic support for visualization creation as narrative visualization becomes more prominent.

Our work also 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 enable additional comparisons where relevant. Navigational choices beyond "Previous" and "Next" buttons (such as an "Up" and "Down" option) could support comparisons with visualizations that do not appear directly before or after the visualization of focus, potentially increasing the amount that is learned from visualized data while still guiding interaction.

References

[1] L. Bavoli, S. Callahan, P. Crossno, J. Friere, C. Scheidegger, C. Silva, and H. Vo, "VisTrails: Enabling Interactive Multiple-View Visualizations," *IEEE Vis '05,* pp. 135–142, 2005.

[2] J.B. Black and G.H. Bower. "Episodes as chunks in narrative memory." *J. of Verbal Learning and Verbal Behavior*, vol. 18, pp. 309-318, 1979.

[3] J. Bronzan, A. Cox, X.G.V., and K. Quealy. "Cophenhagan: Emissions, Treaties, and Impacts (Possible Impacts)". http://www.nytimes.com/interactive/2009/12/05/world/climate-graphic-background.html?_r=0#tab=2.

[4] B. Buxton, *Sketching User Experiences: Getting the Design Right and the Right Design*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007.

[5] E.P.J. Corbett and R.J. Connors, *Classical Rhetoric for the Modern Student*, 4th ed. Oxford University Press, USA, 1998.

[6] P. Diaconis, "Theories of Data Analysis: From Magical Thinking Through Classical Statistics," in *Exploring Data Tables, Trends, and Shapes*, D.C. Hoaglin, F. Mosteller, and J.W. Tukey, Eds. John Wiley & Sons, Inc., pp. 1–36, 2006.

[7] S.P. Dow, A. Glassco, J. Kass, M. Schwarz, D.L. Schwartz, and S.R. Klemmer, "Parallel prototyping leads to better design results, more divergence, and increased self-efficacy," *ACM TOCHI*, vol. 17, no. 4, pp. 18:1–18:24, 2010.

[8] K. Duncker. "On problem-solving." *Psychological Monographs*, vol. 58, no. 5, 1945.

[9] Guardian, Datablog. http://www.guardian.co.uk/news/datablog.

[10] J. Heer and G. Robertson, "Animated Transitions in Statistical Data Graphics," *IEEE TVCG (Proc. InfoVis)*, vol. 13, no. 6, pp. 1240–1247, 2007.

[11] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala, "Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation," *IEEE TVCG (Proc. InfoVis)*, vol. 14, no. 6, pp. 1189 –1196, 2008.

[12] J. Heer and B. Shneiderman. "Interactive Dynamics for Visual Analytics," *CACM,* vol. 55, no. 4, pp. 45-54, 2012.

[13] J. Hullman and N. Diakopoulos, "Visualization Rhetoric: Framing Effects in Narrative Visualization," *IEEE TVCG (Proc. InfoVis)*, vol. 17, no. 12, pp. 2231–2240, 2011.

[14] T.J. Jankun-Kelly, K.-L. Ma, and M. Gertz, "A Model and Framework for Visualization Exploration," *IEEE TVCG*, vol. 13, no. 2, pp. 357–369, 2007.

[15] W. Kintsch, T.S. Mandel, and E. Kozminsky, "Summarizing scrambled stories," *Memory & Cognition*, vol. 5, no. 5, pp. 547–552, 1977.

[16] J. Kolodner and L.M. Wills, "Case-Based Creative Design," *AAAI Spring Symposium on AI and*, pp. 50–57, 1993.

[17] J. Mackinlay. Automating the Design of Graphical Presentations of Relational Information. *IEEE TVCG*, vol. 5, no. 2, pp. 110–141, 1986.

[18] J. Mackinlay, P. Hanrahan, and C. Stolte. Show Me: Automatic Presentation for Visual Analysis. *IEEE TVCG*, vol. 13, no. 6, pp. 1137–1144, 2007.

[19] Microsoft. (2003). Microsoft Excel. Redmond, WA: Microsoft.

[20] Microsoft. (2012). Rich Interactive Narratives. Bangalore, India: Microsoft.

[21] NarrativeScience. http://www.narrativescience.com.

[22] N. Pennington and R. Hastie, "Explaining the evidence: Tests of the Story Model for juror decision making," *Journal of Personality and Social Psychology*, vol. 62, no. 2, pp. 189–206, 1992.

[23] M. Rooze. http://marijerooze.nl/thesis/graphics.

[24] M. Sadoski, A. Paivio, N. J. U. (eds, T. Models, and P. Of, "Reference: A Dual Coding Theoretical Model of Reading."

[25] R.C. Schank and R.P. Abelson, Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures. Psychology Press, 1977.

[26] C. Scheidegger, H. Vo, D. Koop, J. Freire, and C. Silva, "Querying and Creating Visualizations by Analogy," *IEEE TVCG,* vol. 31, no. 6, 2007.

[27] E. Segel and J. Heer, "Narrative Visualization: Telling Stories with Data," *IEEE TVCG (Proc. InfoVis)*, vol. 16, no. 6, pp. 1139–1148, 2010.

[28] A. Shaw, J. Horton, and D. Chen, "Designing Incentives for Inexpert Human Raters," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 1871330, Jun. 2011.

[29] B. Shneiderman, "The eyes have it: a task by data type taxonomy for information visualizations." *Proc. VL*, pp. 336-343, 1996.

[30] C. Stolte, D. Tang, and P. Hanrahan. Polaris: A system for query, analysis, and visualization of multidimensional relational databases. *IEEE TVCG,* vol. 8, no. 1, pp. 52-65, 2002.

[31] Tableau Software. (2012). Tableau. Seattle, WA: Tableau Software.

[32] P.W. Thorndyke, "Cognitive structures in comprehension and memory of narrative discourse," *Cognitive Psychology*, vol. 9, no. 1, pp. 77–110, 1977.

[33] F.B. Viegas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon. "ManyEyes: a Site for Visualization at Internet Scale." *IEEE TVCG (Proc. InfoVis),* vol. 13, no. 6, pp. 1121–1128, 2007.

[34] S. Wernicke, "Efficient Detection of Network Motifs," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 3, no. 4, pp. 347–359, 2006.

[35] H. Wickham, "A Layered Grammar of Graphics," *Journal of Computational and Graphical Statistics*, vol. 19, no. 1, pp. 3–28, 2010.

[36] R. Wilensky, *Understanding goal-based stories*. Garland Pub., 1980.

[37] L. Wilkinson, The Grammar of Graphics (2^{nd} ed.). Springer, 2005.