A SURVEY ON VISUAL DESIGN IN DATA-DRIVEN STORYTELLING

by

XINHUAN SHU

Department of Computer Science and Engineering

The Hong Kong University of Science and Technology

Supervised by Prof. Huamin Qu

September 2018, Hong Kong

TABLE OF CONTENTS

A SURVEY ON VISUAL DESIGN IN DATA-DRIVEN STORYTELLING

by

XINHUAN SHU

Department of Computer Science and Engineering

The Hong Kong University of Science and Technology

ABSTRACT

Communicating data insights effectively to a general audience has become a fundamental problem in the information explosion era. We have seen a growing interest in telling data-driven stories to alleviate this issue. Specifically, these data-driven stories are structured by findings from data and often incorporate visual representations into the process of narration. However, crafting such visual data stories is non-trivial and time-consuming. Particularly, effective visual design in a data story remains underexplored. Besides, narrative authoring tools are still in the infancy and do not keep pace with innovation in practice. These gaps catalyze the growth of research interest in the visualization field. Existing studies have shown typical visual design in data-driven storytelling, taking an initial step to ease the creation process. They indicate the potential to easily produce high-quality data stories for communication.

In this survey, we discuss the development of visual design in data-driven storytelling, as well as the authoring tools. The survey begins with an introduction to the motivation and challenges of data-driven storytelling. Then, based on the definition of the term *narrative*, we review visual design in three primary story components, namely, scenes, sequences, and transitions, and discuss how they compose a visual data story. After that, we investigate existing authoring tools and reflect on evaluation methodologies. The survey ends with current research gaps and future directions in data-driven storytelling.

CHAPTER 1

INTRODUCTION

In this chapter, we first briefly review the background of telling stories with data and visualization, and then discuss the motivations behind current visual narrative designs. After that, we summarize several typical challenges existing in the research field. Finally, we give an overview of this survey in the last section.

1.1 Background and Motivation

Although the term "data-driven storytelling" has become a buzzword and attracted great attention in recent years, telling stories with data has a long history and shows its strength in a variety of domains. For example, the first instance of data journalism was published in the Guardian in 1821, introducing schools in Manchester using the number of students and their costs [68]. In advertising, the Coca-Cola Company proposed a slogan "Six Million a Day" in 1925 to claim high sales and arouse interest among the public [8].

Nowadays, this trend is escalating as digital information is growing and becoming avaiable online at an exponential rate in the era of big data. Various tools have been developed to facilitate storytelling, such as data visualization. According to the analysis from Google News Lab [2], visualization in data journalism has helped reduce the world's complexity and assisted readers in making sense of data. Meanwhile, the proliferation of web-based visualization techniques (e.g., the rapid development of D3 [22]) has paved the way for the rise and easy sharing of visual data stories. Many different forms of storytelling emerge as shown in Figure 1.1, including animated infographics, videos, and interactive online visualizations. Data visualization as a powerful vector for communicating information is well recognized among practitioners.

The current research into data visualization has attached importance to exploratory data analysis. Many visual analytics systems aim to provide novel interfaces and interactions for data exploration, and some also integrate advanced machine learning and data mining algorithms to support powerful analysis [52]. Comparatively less attention has been paid to the communication function of visualization. However, with the increasing roles of visual data-driven stories and interactive visualization in data analysis, this line of research
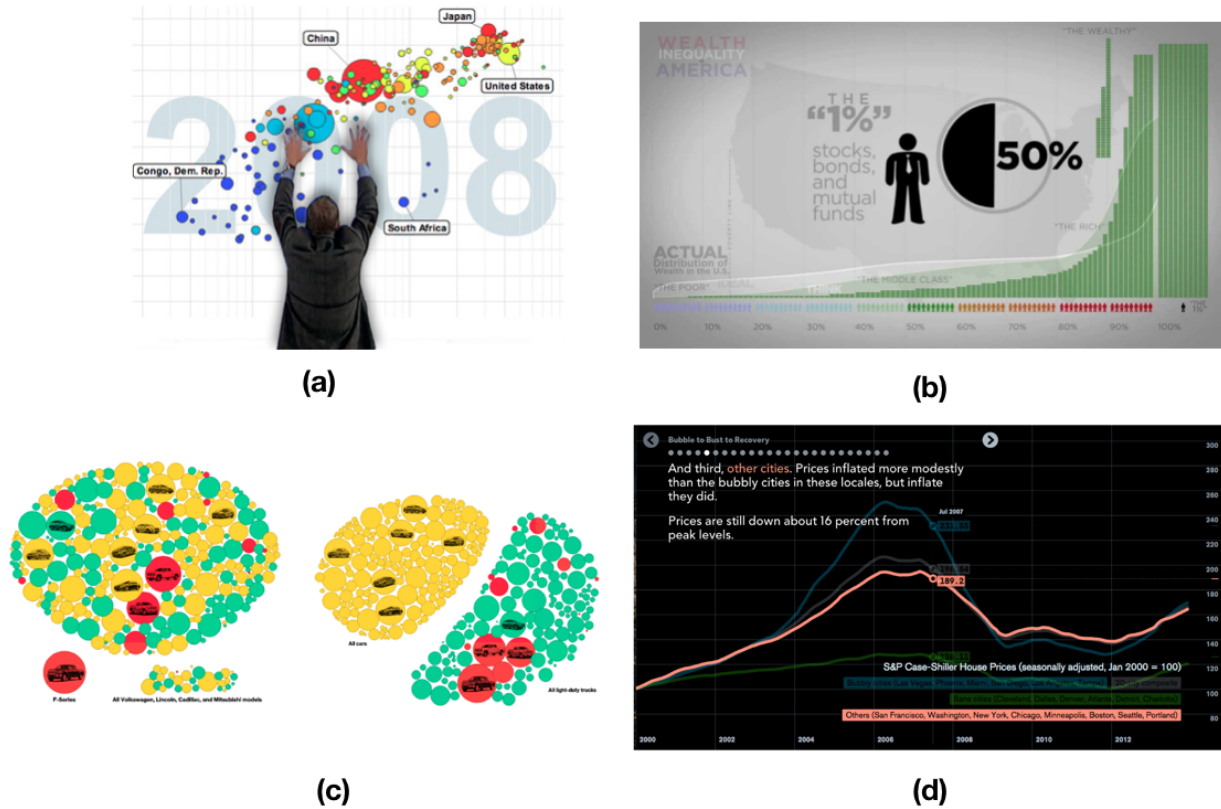
Figure 1.1. Different forms of storytelling. Rosling's talk (a) shows trends in human development [5]. A data video (b) reveals wealth inequality in America [16]. Online interactive visualizations from Bloomberg tell stories such as *Addicted to Trucks* (c) [13] and *Bubble to Bust to Recovery* (d) [9].

is attracting interest. The Dagstuhl seminar held in 2016 [26] invited both researchers and practitioners to dig into the topic of storytelling using visualization. Microsoft launched the "data-driven storytelling" project [1], which aims to understand the story design process and explore methods for easy creation. In spite of several breakthroughs in this field, there is still a plethora of research remaining to be done. For example, although some work [18, 25, 64] has identified a variety of techniques from existing stories, there is a lack of guidance on how to organize these techniques given a specific topic. In addition, how to evaluate the effectiveness of not only visual data stories but also visualization authoring tools is yet an open question.

1.2 Challenges

Storytelling is considered as the next step for visualization, specifically for it can present data in an interpretable way [47]. Thus, we have naturally come to issues of effective visual design in data-driven storytelling. In reference to the design implications informed by good graphics [66], visual design in data-driven narratives should better conform to these two

principles as well: one focuses on the content and form of the representation and the other pays attention to the comprehension and perception of the viewers. Therefore, we present three primary challenges resulting from the representation and evaluation of stories, and the limits of authoring tools.

The first challenge is how to tell stories clearly with engaging visual effects. On the one hand, the mission of a data story is to get a point across and to have the audience understand and trust the content [47]. Neither describing data and generating insights from an existing dataset nor arranging them in an interpretable way is an easy task. On the other hand, attractive visual effects contribute to a good narrative. A typical approach is to gain empirical knowledge from a curated collection of data stories. However, the derived design guidance is incomprehensive and constantly evolving. The way to inform an appropriate design space is still under explored.

The second challenge is how to evaluate the effectiveness of such visual data stories. While many interesting stories are widespread and well received in the online media such as Gapminder [5] and US wealth inequality [16], there is a lack of clearly defined metrics or evaluation methods. Typical measures used in visualization mostly consider time spent on given tasks and their response accuracy through a user study. Nevertheless, these are not pertinent to storytelling, which, instead, requires a crucial evaluation of comprehension, engagement, and memorability of viewers. It also gives rise to problems about comparing the effectiveness of different storytelling techniques.

Finally, the third challenge is how authoring tools could help support visual narrative designs and lower the barriers to craft data stories. Creating data stories is never a trivial task, relying on a broad set of tools and skills. Authors are expected to use dedicated software or even do the programming themselves when constructing a complete story. Existing tools are limited in many aspects. For example, professional softwares such as Adobe Illustrator or After Effects have steep learning curves and require users' capabilities to design and generate materials. Lightweight tools with a focus on single functions ask for extra tools to complete a story. It is challenging to strike a balance between design flexibility and efficiency of authoring tools.

1.3   Overview

There are many research issues for transforming data into visually shared stories. However, in this survey, we do not intend to use the term "storytelling" in a broad way that considers all data visualizations as stories. Specifically, this survey mainly focuses on a narrow

definition of a visual data story [50]. For example, we exclude exploratory data analysis visualizations which allow users to explore freely and interactively but provide little guidance. In addition, visualizations without enough annotations or written explanations can hardly be considered as well-told stories, since they require viewers to interpret the contents themselves. Therefore, our analysis is based on asynchronous stories with some form of author-specific narration. Our goal in thie paper is twofold: 1) to review what makes a compelling visual data story based on three primary story components; 2) to investigate ways to author narrative visualization. The rest of this survey is organized as follows.

In Chapter 2, we introduce some existing taxonomies for narrative visualization and visual storytelling. After that, we describe our taxonomy derived from the definition of the term "narrative".

In Chapter 3, we further illustrate how the visual design in data-driven narratives are developed based on three story components, namely, scenes, sequences, and connection, thus informing the design space. Various techniques are applied in different components to compose compelling and interpretable visual data stories. We then attempt to identify and categorize the design features with collected visualizations.

In Chapter 4, we investigate tools to help construct data-driven stories and discuss the evaluation methods used in these visualization authoring tools.

Finally, we conclude some future research directions of visual design in data-driven narratives in Chapter 5.

CHAPTER 2

TAXONOMY

In this chapter, we first introduce two existing taxonomies given by Segel and Heer [64] and Tong et al. [65], respectively. The former is a highly cited paper and has contributed greatly to narrative visualization, while the latter is an up-to-date survey on storytelling and visualization adding a variety of new narrative methods and works. We aim to demonstrate their primary characteristics for an overview instead of explaining the technical details of each taxonomy. After that, we propose our taxonomy derived from the definition of the term "narrative".

2.1 Taxonomy by Segel and Heer

Segel and Heer [64] investigated 58 examples of narrative visualizations and formulated a design space containing three divisions of features, namely, genres, visual narrative tactics, and narrative structure tactics.
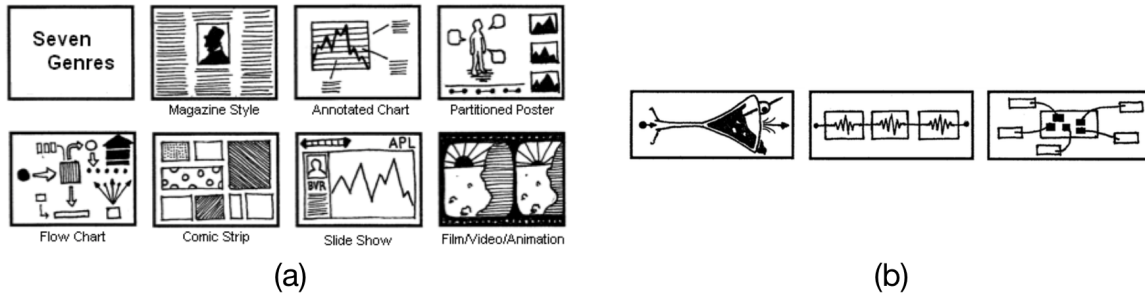

Figure 2.1. (a) Seven genres of narrative visualization. (b) Three schemes of structure. Left: martini glass style; middle: interactive slideshow; right: drill-down story [64].

- **Genres.** The first division identifies seven basic genres shown in Figure 2.1(a): magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show, and film/video/animation. These genres differ in the number of frames and the ordering of visual elements.
- **Visual Narrative Tactics.** The second division concerns visual mechanisms to facilitate the narrative. These methods are subdivided into three sections. *Visual structuring* describes the way to organize the story and points out the current position of the

5

viewer through visualization (e.g., progress bars). *Highlighting* directs attentions to the emphasized elements using color, motion, etc. *Transition guidance* aims to avoid disorienting the viewer when changing scenes.

- **Narrative Structure Tactics.** The third division considers both visual and non-visual techniques to assist the narrative. These ideas are captured in a three-way classification. *Ordering* corresponds to sequences viewers look at through narratives, including linear path, random access, and user-directed path. *Interactivity* allows viewers to manipulate the visualization, while *messaging* displays how visualization communicates insights with viewers combined visual cues and text descriptions (e.g., annotations).

This paper further characterize these design differences in term of the balance between author-driven and reader-driven approaches together with messaging and interactivity. Three common schemes can be classified as shown in Figure 2.1(b). *Martini glass structure* begins with an author-driven introduction to the narrative, followed by a reader-driven stage to allow interactive exploration. *Interactive slideshow* incorporates interactions within each slide, promoting a dialogue between these two approaches. *Drill-down story* allows readers to choose instances from a given theme, which prioritizes the reader-driven approach.

Though Segel and Heer offered a nice framework and first identified seven genres of narrative visualization, this work is more than 8 years old. The past few years have witnessed some newly emerging formats (e.g., ScrollyTelling [14] and SketchStory [49]) and a gradual shift of existing genres. These changes require a study on recent advancements in narrative visualization.

2.2 Taxonomy by Tong et al.

Tong et al. [65] conducted a survey targeting at storytelling literature in visualization, which is much more up to date and shows the evolving field. Besides, instead of a focus on journalism like Segel and Heer [64], the authors provided a comprehensive survey of storytelling literature with a coverage of scientific visualization, information visualization, and geo-spatial visualization.

In this survey, storytelling is considered both as an entity and a creative process. Thus, the taxonomy is first based on the logical notations of who, how, and why.

- **Who** plays the main role in storytelling for visualization? This survey identifies two dimensions, namely, *authoring tools* and *user engagement*. The former addresses the author who crafts the story, while the second is about the audience.

- **How** are stories told? Two categories are classified, i.e., *narrative* and *transition*. Transitions can be further subdivided into animated and static transitions.
- **Why** use storytelling for visualization? *Memorability* and data *interpretation* are important goals for storytelling. They address why authors present data in the form of stories.

The second dimension of this taxonomy refers to the sequence of events, which tracks the viewing path of readers through a narrative. Four categories are extracted. *Linear* indicates that viewers are supposed to follow the order prescribed by the author, while *user-directed path* allows readers to select or create the path. *Parallel* visualizes multiple paths at the same time, and *random access* (also known as *overview*) means no prescribed path.

The taxonomy in Figure 2.2 presents an overview of important elements in storytelling visualization. However, several dimensions still seem vague and intertwined. For example, as the paper stated itself, "narrative visuals contain the transitions between events", it is hard to distinguish *transition* from *narrative* when categorizing literacy in *how*. In addition, there is no work belonging to *interpretation* in this survey.

		Linear	User-Directed	Parallel	Overview
Who	Authoring Tools				
	User Engagement				
How	Narrative				
	Static Transitions				
	Animated Transitions				
Why	Memorability				
	Interpretation				

Figure 2.2. The classification of the storytelling literature proposed by Tong et al. [65]

2.3 Taxonomy for Visual Design in Data-driven Storytelling

Segel and Heer [64] took an empirical approach to investigate the design from an analysis of visualizations. Several works [20, 25, 53] have furthered this line of research and enriched the design space with emerging narrative techniques. However, the capability of visualization practitioners like data journalists to produce artifacts is limited to authoring tools, which did not cover cutting-edge techniques to make compelling stories. Tong et al. âĂŸs taxonomy did not specify concrete techniques to compose a story. Therefore, in this taxonomy, we consider the fundamental components of stories and explore designs with a look

into visualization practices and researches. Firstly, in view of the term "narrative", the definition in the Oxford English Dictionary [11] is "an account of a series of events, facts, etc., given in order and with the establishment of connections between them". Following this explanation, we propose our taxonomy for visual design in data-driven narratives, which consists of three components.

- **Scene**. The "scene" component concerns the explanation of data and the communication of messages. *Visualization types* work closely in conjunction with data. Despite diverse custom visualizations depending on different tasks, a few types appear frequently, e.g., maps, bar charts, and pictographs. *Textual annotation* includes descriptions beyond and on visualization, which assists both in narration and interpretation. *Highlighting* performed by changing graphical properties or leveraging close-up techniques draws readers' attention to the emphasized elements.

- **Sequence**. The second component reflects on the author-specific ordering in storytelling, which imposes a structure to a story and enhances the navigation for readers. *Linear* narratives are typical in data-driven storytelling. *Non-linear* narratives are well received in movies, which shed light on the design of visual data stories.

- **Transition**. The third component links separate story scenes with attractive visual effects. To convey a story smoothly, many techniques are explored such as camera motions, temporal and spatial controls, animated transtions, morphing, and metaphors.

Our taxonomy tries to better explore and summarize the content and representation of each story component. Thus, it could facilitate visualization selections. The organization of a story is somehow analogous with PowerPoint slides: each scene can be considered as a slide; the sequence indicates how authors organize multiple scenes or slides in a certain order; and the transitions are similar to visual effects when switching from one slide to another. Actually, some works, especially surveys, may address more than one theme. We have to recognize subjectivity when categorizing a diverse set of designs into a taxonomy and placing them by what we determine to be the main focus of the work. The detailed visual design techniques are described in Chapter 3.

CHAPTER 3

VISUAL NARRATIVE

Given the proliferation of data-driven storytelling in practice, it is of great importance to advance researches in visualization to identify factors that make visual data stories compelling and improve their performance in return. This chapter introduces existing works that focus on visual design in data-driven narratives. According to the role each work mainly serves in the storytelling process, we divide them into three categories, namely, scene, sequence, and transition. Considering a well-known principle in the design field, namely, "form follows function" [4], we then discuss main forms of visual narrative design and their corresponding functions in each category as shown in Table 3.1. The representative works are describe in detail with both benefits and drawbacks. Based on this taxonomy, we review current visualizaiton genres and reflect on how they compose a visual data story.

Categories		Form	Function
Story Component	Scene	Close-up, [34] Change graphical properties [69]	Highlight
		Text [41, 35, 60]	Annotation
		Pictogrphics [36, 72]	Visual
		Scale [59]	Consistency
	Sequence	Timeline [23]	Navigation
		Linear narrative, [40, 42, 46] Non-linear narrative [20, 44]	Structure
	Transition	Camera motions [67], Static transition [62], Animated transition [27, 29, 31, 37, 38, 56], Morphing [30], Metaphor [43, 51, 71],	Linking separated scenes

Table 3.1. Visual design in data-driven storytelling. In the *Form* column, we list forms used in the related papers. The *Function* column list the role the corresponding form serves.

3.1 Scene

One of the important components in data-driven storytelling is visual design in scenes, which aims to communicate events and facts in conjunction with data to a broad audience. Qualified visual representations are supposed to describe data and amplify cognition. With a curated collection of visualization artifacts online, some researchers conducted surveys to extract techniques and inform the design space.

Segel and Heer [64] took an initial step in classifying techniques and formulating the design from an analysis of 58 examples, as was mentioned in Section 2.2. This kind of work has shed light on further research in narrative visualization. Hullman and Diakopoulos [39] expanded their discussion including extra-representational influencers like contextual differences in interpretation. They distinguished four editorial layers for conveying meaning, which covered the whole creation process from the data, visual representation, textual annotation, to interactivity. Based on an observation of 51 professionally-produced narrative visualizations, five major forms of rhetorical techniques were identified in this genre of storytelling, namely, information access rhetoric, provenance rhetoric, mapping rhetoric, procedural rhetoric, and linguistic rhetoric. Specifically, these strategies of visualization rhetoric worked as an analytical framework to understand how visual techniques could prioritize particular interpretations. Since then, Stolper et al. [25] also adopted this empirical approach to enrich the range of techniques in constructing scenes, such as textual and audio narration, text annotation, labeling, and tooltips, and show their evolution over time. While many of these surveys investigate data stories but are agnostic to the data type, Brehmer et al. [23] provided a detailed analysis which is specific to timelines describing from three dimensions, namely, representation, scale, and layout. Thus, several fundamental elements embellishing scenes have been recognized through multiple iterations, such as annotations, representation, and highlights.

Although extracting design guidance by empirical studies is deemed as a feasible approach, it faces several problems. The design knowledge is derived from a small sample of visualization examples and is also continually evolving, thus leading to the incompleteness [54]. And there exists inevitable subjectivity when grouping and summarizing techniques. Apart from conducting surveys to inform the design space, the visualization research community has proposed new methods to optimize the scene representations recently.

Nearly all the above surveys argued that text **annotations** play an important role in supplementing visualization and directing readers' attention. Many efforts have been made to create annotated information visualizations for online news [41, 35]. Hullman et al. [41]

proposed the Contexifier system targeting at providing context for business news as shown in Figure 3.1. It first automatically creates a timeline graph of one companyâĂŹs stock. Other articles with relevant contents about this company feed information to the additive customized annotations for the line graph visualization. The selection and layout of those annotations in the graph depended on an underlying algorithm in view of three factors, namely, linguistic relevancy of the topic, visual saliency based on the data series, and analysis of article volume about historical events.
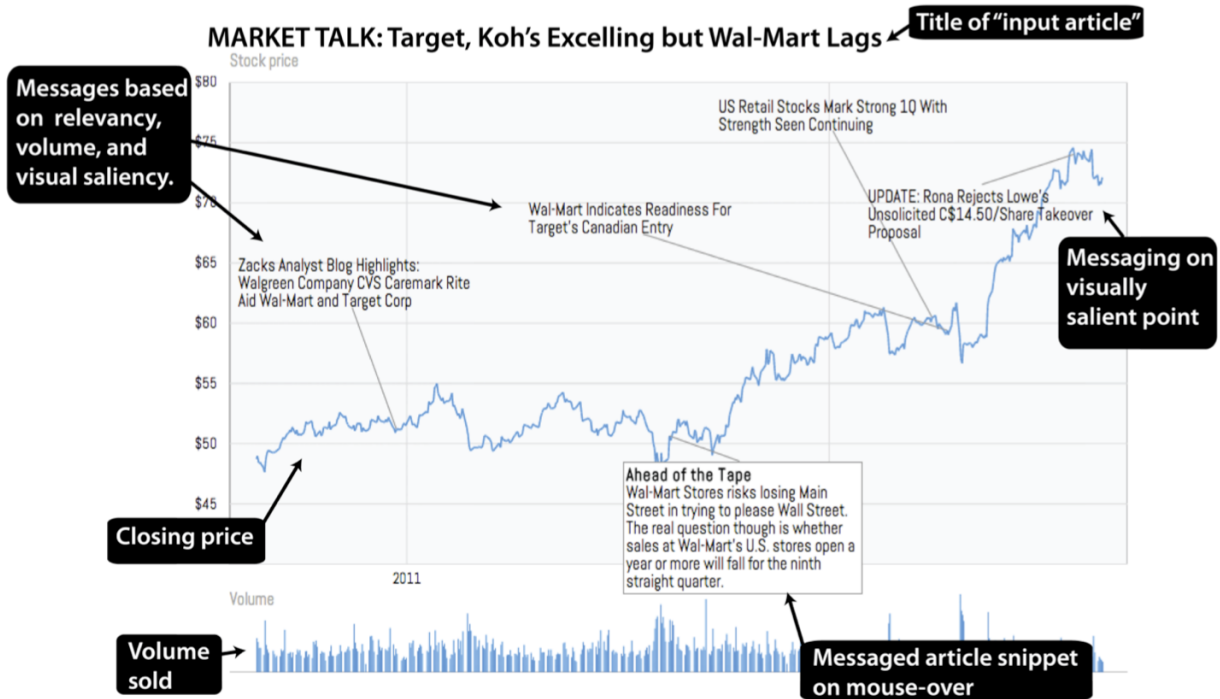

Figure 3.1. An annotated visualization produced by Contextifier [41].

Since Contextifier takes a step towards annotated news visualization but is limited to the stock time series using line charts, Gao et al. [35] further contributed a news generation pipeline to this area. And the NewsView system implemented this pipeline to automatically generate interactive and annotated maps, which was the most prevalent form of these visualization.

In addition to automated generation, Ren et al. [60] developed ChartAccent, an authoring tool that allows users to augment charts with annotations easily. The current system provided manual and data-driven annotations based on the design space informed by 106 annotated charts. However, the problem in the reproduction study is that default data-driven annotations were not ideal most times. It reflects on the challenge that the initial design space might be incomprehensive, leading to the ineffectiveness of the tool.

Among a variety of narrative visualizations, most rely on a few basic visualization forms, such as the above line graphs, bar charts, and maps, as well as **infographics** (also known

as pictographs) [18]. Some researches [19] also attest to the prominence of pictographic representation by triggering viewers' engagement compared with standard charts through a crowd-sourced online study. Therefore, it attracted research interests in easing the creation of data-driven infographics [45, 72] and embedding them to provide compelling scenes [17, 36].

Figure 3.2. A use case to change the original column chart to an infographic style version with customized marks with InfoNice [72].

Another frequent technique adopted by authors when creating scenes is to perform element **highlighting**. There are many methods found in visualization practices, such as wrapping highlighted items with a shape and fading out unessential parts to emphasize important ones. Waldner et al. [69] enhanced the focus to pop out from its context in a large dynamic scene using a strong visual attractor, flicker. Some movie narrative techniques could be used for reference in data-driven storytelling. For example, fisheye views [34] as a close-up highlight element details to the audience.

Despite these promising design guidelines, an often-overlooked aspect is the wide existence of multiple views. Effective single views may constitute inconsistent multiple scenes, leading to slow and error-prone interpretation. Qu and Hullman presented a qualitative study to investigate specific **consistency** constraints and exceptions [59, 58]. As shown in Figures 3.3, there were more validations than exceptions considering constrains specialized to scales (e.g., color scale, x or y scale). The exceptions indicated conditions where specific design rationales should defend inconsistencies and override constraints. These findings provided a foundation for automated inconsistency detection mechanisms. Nevertheless, this work mainly started from the perspective of authors to create multiple scenes and did not study the impacts inconsistency leave to viewers.

In summary, many surveys have identified basic elements and techniques from analysis of visualization artifacts. Several works develop tools to either ease the creation or augment
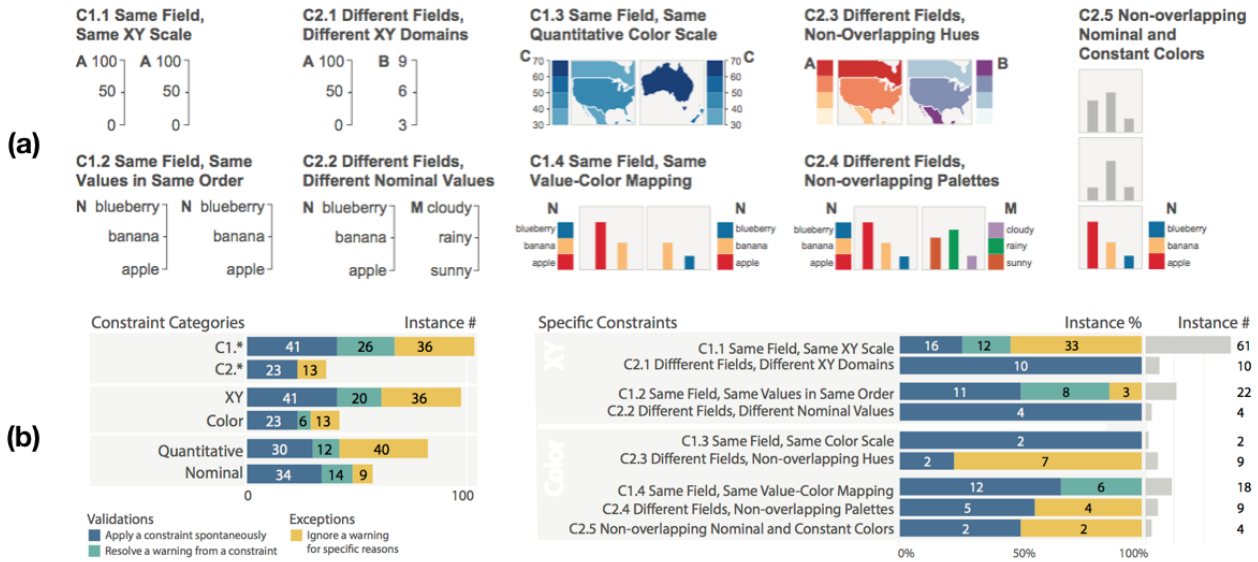
Figure 3.3. Qu and Hullman conducted quantative studies for XY and color scales to investigate consistency principle (a) and counted validation and exception (b) [59].

representations. And the research community also begins to explore high-level design principles such as consistency to further embellish storytelling. However, the expressiveness and effectiveness are evaluated based on a small sample of both users and visualizations in most cases. And there remain research efforts to explore more design considerations. For example, audio narration like background music is indispensable in some narrative forms, but few works have advanced in this direction.

3.2 Sequence

Conveying a narrative with data visualization requires story creators to thread representations into a compelling yet understandable sequence. This form of author-specified ordering differentiates itself from open-ended explorations by providing navigation aids frequently with explicit structures. Several surveys [25, 64] initialize and further enlarge a variety of techniques used in visualization artifacts to communicate structures and navigation to viewers. These techniques include next/previous buttons, scrolling, breadcrumbs, section header buttons, menu selection, timelines, and geographic maps. Besides, some works focus on the narrative structures. Based on Cohn's theory of visual narratives [28], Amini et al. [18] identified the structures from a qualitative analysis of 50 data videos, which consisted of four major narrative categories, i.e., Establisher (E), Initial (I), Peak (P), Release (R). An example with detailed explanation for each category is shown in Figure 3.4. According to the findings, they noted that most data videos followed this method, and the Initial category played the most prominent role in data videos.
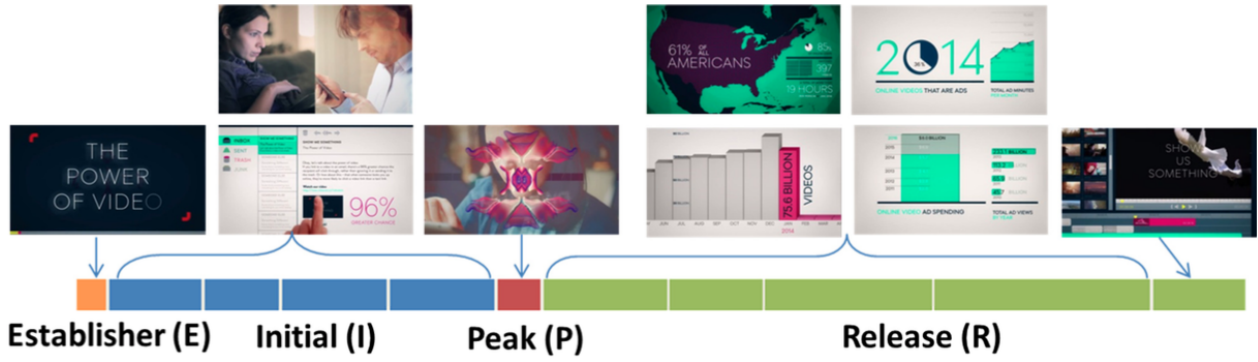
Figure 3.4. Visual narrative structures: *Establisher* provides referential information; *Initial* "sets the action or event in motion"; *Peak* reflects on the most important events; and *Release* shows "the aftermath of the Peak" [18].

Researches in the visualization community have studied this category in depth. Hullman et al. [40] provided a deeper understanding of sequence in the form of linear, "slideshow-style" presentations. Starting from a qualitative analysis of 42 narrative visualizations, they gained insights on how to structure visualization storytelling. The first finding pointed out six transition types distinguished by changes to data dimensions and used to drive transformations between visualizations. Another finding describes higher-level strategies that local transitions (visualization-to-visualization) respond to a small number of changes to data dimensions, while global transitions (involving multiple local transitions) are required to maintain consistency in the form of parallelism. These results helped them build a semi-automatic graph-driven approach for visualization sequence supports. An objective function is proposed with an aim to minimize the transition costs considering audience's perceptions. After that, Hullman et al. [42] further studied impacts that transitions bring to longer visualization sequences. Their results indicated a hierarchical structure preferred by humans beginning with grouping subsets of visualizations with shared data properties.

Based on these works, Kim et al. [46] presented GraphScape, a directed graph model to extend automated reasoning about visualization similarity and sequencing. Instead of a focus on conceptual transition types as Hullman et al. did, GraphScape leverages the Vega-Lite language and works on a more sophisticated cost function balancing local and global transitions. Thus, this model works both as a generative tool (e.g., given a started visualization, traverse the graph and provide a sequence for this series of visualizations) and an evaluation tool (e.g., given a sequence of visualizations, measure their costs). However, this work currently implements at a level of visualization specification and does not go deep into the data level.

However, above works all focus on linear sequences. In practice, narrative visualizations do not necessarily conform to the linear narration. For example, a well-received data
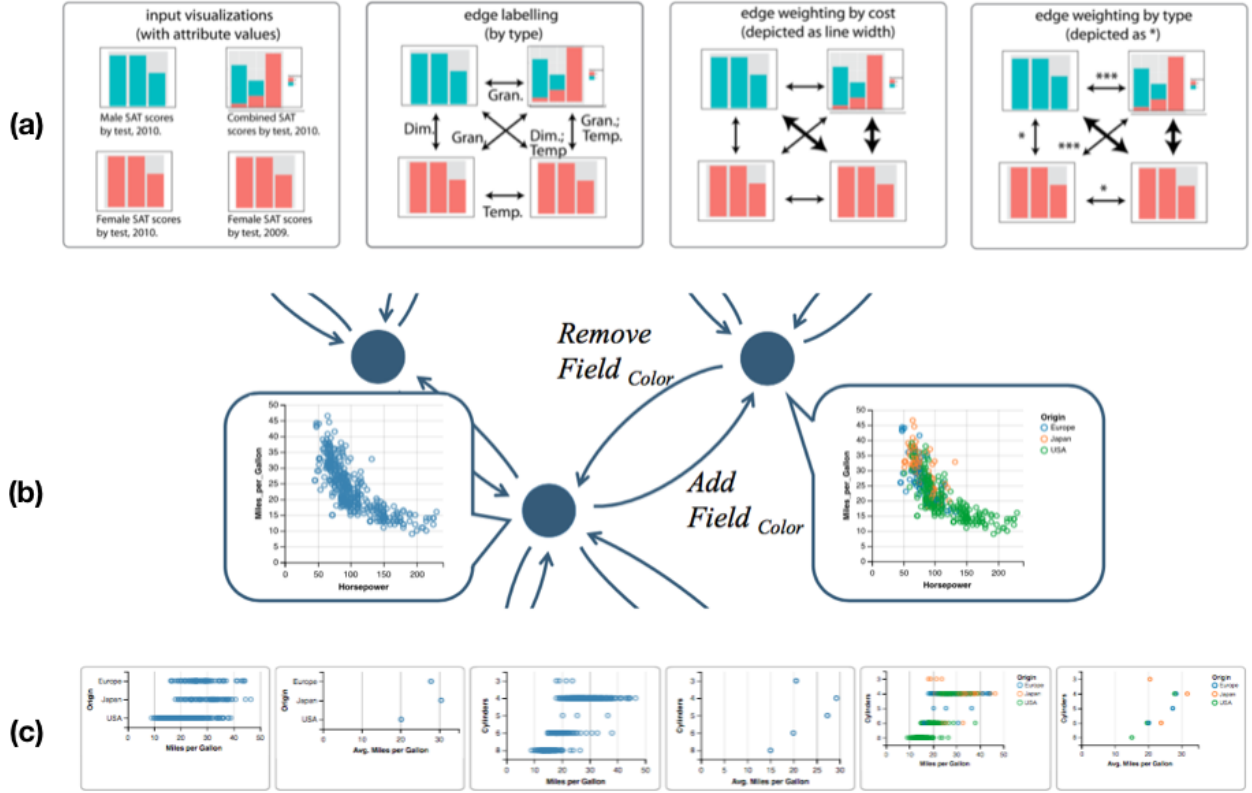
Figure 3.5. (a) Diagram of the graph-based approach where each visualization represents a nodes and edges stand for possible transitions whose weights are calculated by the cost function [40]. (b) The GraphScape model extended from (a) [46]. (c) An example of a viusalizaion sequence.

video named "Wealth Inequality in America" [16] does not adopt a linear narrative as well. It is based on comparisons among three conditions, i.e., the ideal condition, what American think, and the reality. Recently, Kim et al. [44] presented Story Explorer to visualize non-linear narratives in movies. It provided a curated collection of non-linear narrative techniques, including flashbacks, flashforwards, bidirectional flashes, zigzags, follow-the-hero, and so on. Bach et al. [20] identified flashback as a critical narrative in data comics. Wang et al. [71] applied foreshadowing to present video clickstream data. However, most of them are not studied in the field of data-driven storytelling and few tools support to embed these techniques for visualization sequencing.

3.3 Transition

A narrative typically contains an account of multiple events or facts described in a series of ordered scenes. Linking these separated story scenes is crucial, especially providing transitions with attractive and engaging visual effects. Segel and Heer [64] pointed out *transition guidance* as an important dimension in visual narrative tactics. They found the common

techniques in narrative visualization borrowed from films, e.g., camera motions, animated transitions. As such, much practice and research have studied this category in depth.

Camera motions remain an ever-present mechanism in many visual data stories. The main advantage of imagining a virtual camera in data-driven narratives is its innate nature, as it conforms to the way how humans perceive the real world. Thus, it can easily elicit viewers' engagement by triggering their emotions. For example, in the "Global Wealth Inequality" video [6], when describing the richest 1% have accumulated 43% of our world's wealth, the authors moved the camera quickly along the bar but still took some time to reach the top. It easily provoked emotions and emphasized the wealth gap between the rich and the poor to the audience. In addition to camera viewing, other techniques such as zooming and panning are used frequently as well. Research in this area concentrates more on optimizing the smooth effects and is maturing rapidly. For instance, Wijk and Nuij [67] presented a model to handle smooth viewing with a focus on simultaneous zooming and panning. However, the evaluation of the effect was based on the perception and did not consider the cognitive aspects which also played an important role in this kind of animation.
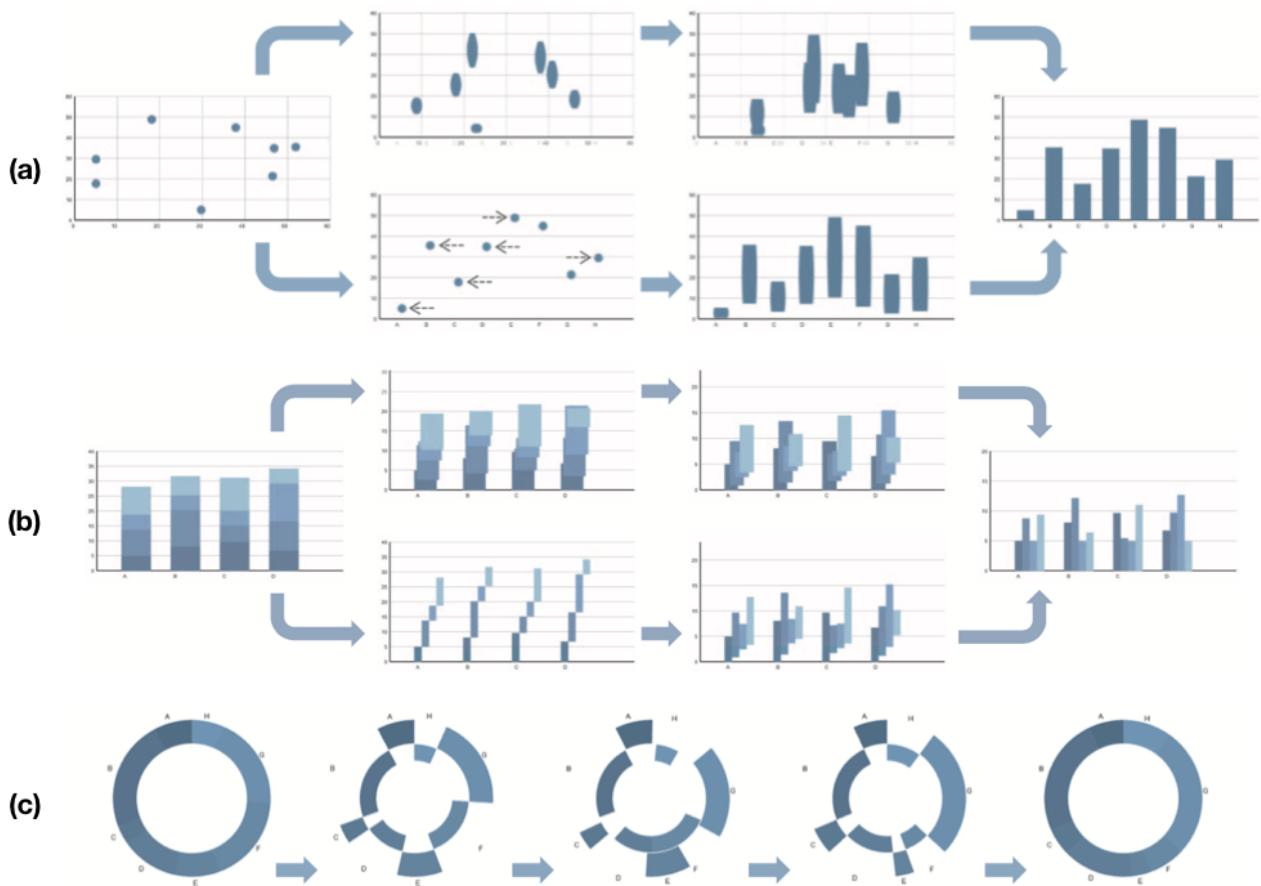

Figure 3.6.   Animated transitions between statistical data graphics. (a) Animating from a scatter plot to a bar chart. (b) Animating from stacked bars to grouped bars. (c) A multi-stage animation of changing values in a donut chart [38].

Animated transitions are another promising technique to connect story scenes and facilitate perception of changes between related data visualizations. As a typical example like *Green Honey* [7] shows, when viewers scroll down, the transitions will be triggered by small points' flying to their new position according to different colors. Heer and Robertson [38] studied the design and effectiveness of animated transitions between statistical data visualizations such as bar charts, pie charts, and scatter plots as shown in Figure 3.6. They crafted a taxonomy of seven transition types and proposed a visualization system, DynaVis for assessing the efficacy. The finding showed that animated transitions can improve graphical perception. However, not all animated transition scenarios are significantly different. For example, staged transitions only gain modest performance compared with linear transitions. Also, in this work, they just identified the design space among three common statistical data graphics and did not cover a wide range of transition scenarios. Some researches expand it to tree structures. Plaisant et al. [56] proposed SpaceTree breaking the transition into 3 stages. Guilmaine et al. [37] further compared four animated transitions in a radial tree visualization, i.e., linear, staged, hierarchical, and a hybrid animation mixed with the staged and hierarchical approaches. The results indicate the hierarchical transitions in tree visualizations outperform other techniques significantly. However, the evaluation was just based on a target-tracking task and required to be further validated. Another research focus on animated transitions is to evaluate the effectiveness of such techniques like temporal distortion [29] and staggering effects [27] in visual tracking. The findings showed the benefits using these techniques, but the extent depended on the conditions and transition types.

Besides, animated transitions in **3D space** is much eye-catching. Some software like PowerPoint and Keynote has already embedded such transitions such as cube and box. Elmqvis et al. [31] integrated animated rotations in 3D space, somewhat akin to rolling the dice, in scatterplot matrix navigation.

Morphing is a special effect in animations which "changes one shape into another through a seamless transition" [10]. This technique has been applied to several online visual data stories, such as EPFL Data Monolith [3] and The Rhythm of Food [12]. They typically use points as basic elements. The previous data graphic is separated into multiple points, then these points animate to their new positions and form a new data visualization. Similarly, Drucker and Fernandez [30, 55] presented SandDance to explore the effectiveness of animating between unit representations as shown in Figure 3.7.
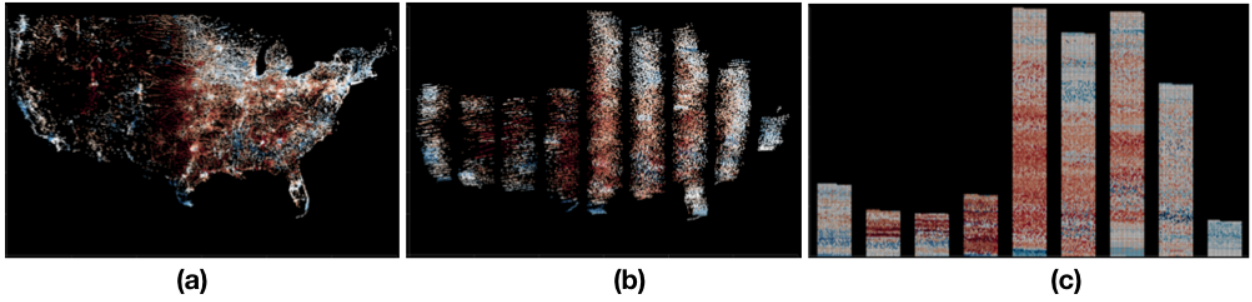
Figure 3.7. A sequence explores the 2010 American election [30]. (a) Scatter plots are colored by voting percents on the map. (b) The path interpolates between the starting and ending states. (c) Bar charts are binned by logitude.

Visual metaphors show wonderful performance in data-driven storytelling and presentation but require much creativity. Visual Sedimentation [43] presents an attractive design metaphor for visualizing data streams in Figure 3.8. It simulates the real-world sedimentation processes where objects fall because of gravity and then aggregate into strata over time. As for data streams which have incoming data continually, sedimentation fits a lot since data is encoded as objects and a force model controls the falling rate. Wang et al. [71] also used this metaphor to visualize video clickstream data. Lin and Vuillemot [51] explored spirograph designs for ambient display of Tweets. However, one problem for these visual metaphors is that they increase pressure for viewers to accurately extract the data. It might be not appropriate for strict data exploration.

Figure 3.8. The Visual Sedimentation metaphor is applied to a bar chart (left), a pie chart (center), and a bubble chart (right) [43].

Despite multiple animated transition techniques have been created and implemented in the visualization research, Amini et al. [17] found such transitions are uncommon in data videos today. They suspected the complexity of realizing them with some programming skills and the lack of support in existing video editing tools contribute to the deficiency. Thus, they opted to support such kind of transitions as shown in Figuire 3.9. However, they

only extended chart transitions [38] to and from pictographs. Their finding also indicated a gap that data visualization tools are not keeping up with the pace of innovations, as Google News Lab pointed out [2].

Clip Type → Vis Type ↓	Creation	Cycling	Accumulation	Embellished	Annotation	Transition	Drill & Roll	Destruction
Line/Area Chart	Grow axis & first line	Morph line paths	Grow each line	Add icon to line grow	Display reference line	Transform line chart into pictograph	Transform data point into filled icon	Remove lines & axis
Bar Chart	Grow axis & first bar or set of bars	Adjust bar lengths	Grow each stack of bars	Add icon to bar grow	Highlight bars	Transform bar chart into pictograph	Transform bar into filled icon	Remove bars & axis
Pie/Donut Chart	Draw donut layout & grow 1st slice	Adjust slice level	Grow slices	Add icon to donut chart	Display information label	Transform bar chart into donut chart	Transform slice into filled icon	Shrink slices & Remove donut layout
Map	Draw map borders	Change colored regions	Color each set of regions	Add icon to map	Point to map regions	Transform map into pictograph	Transform region into filled icon	Wipe map borders
Pictograph	Grow axis & draw 1st set of icons	Adjust the number of icons	Draw each set of icons	Add icon to pictograph	Display information label	Transform pictograph into other vis	Transform icons into other vis	Destroy icons
Colored Pictograph	Color 1st set of icons	Change color	Color each set of icons	Add icon to colored pictograph	Point to icons	Transform icons into other vis	Transform icons into other vis	Decolor icons
Scaled Icon	Draw icon layout	Adjust icon size	Draw and size each icon	Add icon to scaled icon	Flash/move icon	Transform icons into other vis	Transform icons into other vis	Shrink icons until they disappear
Filled Icon	Draw icon layout	Adjust icon fill level	Draw and fill each icon	Add icon to filled icon	Display information label	Transform icons into other vis	Transform icons into other vis	Empty icons

Figure 3.9.   Clip types as a function of visualization types (rows) and animation types (columns) supported by DataClips [17].

However, the use of animation always arouses a heated discussion [33]. Robertson et al. [62] compared two alternative trend visualizations based on static representations in Figure 3.10 with the animated Gapminder Trendalyzer. The finding showed that trend animation can be challenging both in analysis and presentations. Although animations made participants enjoyable and exciting, it resulted in many errors. However, some works [19, 38] argued that animation can boost understandability of data insights and increase focused attention. Thus, without a series of unified evaluation criteria, it is hard to assess the effectiveness of visual data stories.
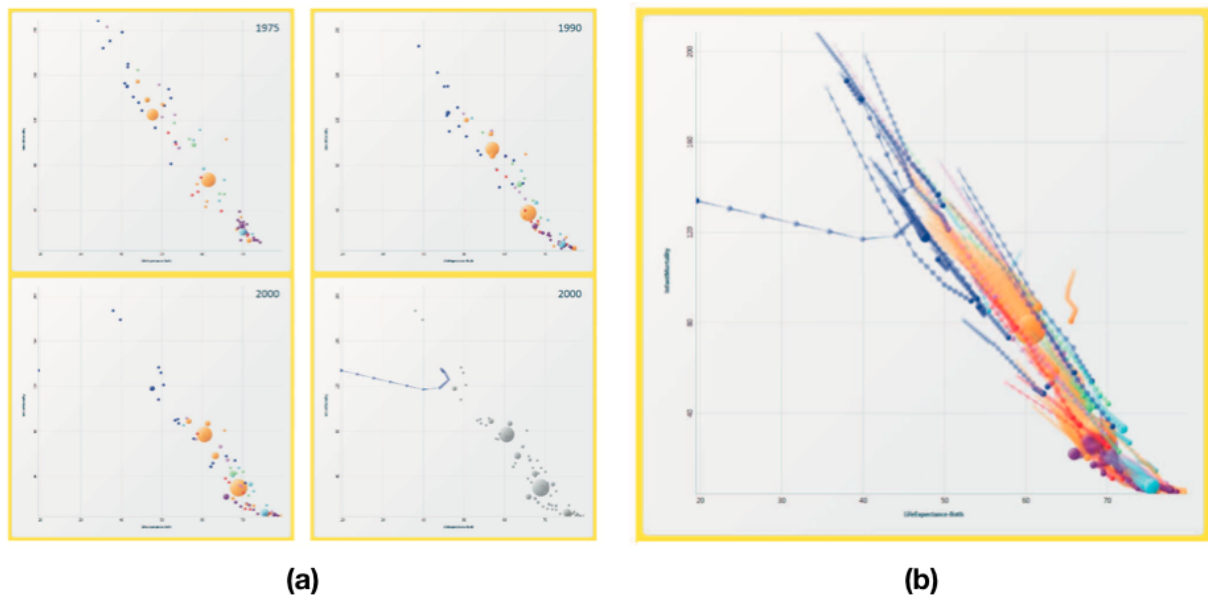
Figure 3.10. Two alternative static trend visulizations [62]. (a) A small multiples display the trend traces side by side. (b) One display shows traces of all trends overlaid simultaneously.

3.4 Combining Multiple Components

In the above sections, we identify three major components to compose a data-driven narrative, namely, scenes, sequences, and transitions. We learn from both research and practice to inform the design space of each component, respectively. Actually, a visual data story is not required to explicitly comprise all three parts. Take seven genres of narrative visualization (Figure 2.1) for example. A magazine style which embeds visualization in a page of text or an annotated chart usually has a single frame. Thus, they may clearly address the scene design. The sequence and transition deal more with external factors such as text narration. Authors may consider the narrative sequence for the layout of partitioned posters ("multi-view visualizations"), flow charts, or comic strips, but are likely to employ a simple static transition between multiple frames. When authoring slide shows or film, these three components all play important roles. The scene design gives a basic representation to communicate data insights. The narrative sequence works for the goal of storytelling such as comparisons and trends, thus amplifying cognition. Advanced animated transition effects engage a broad range of viewers.

We also examine the story components with new emerging genres like ScrollyTelling [14] and SketchStory [49]. First, ScrollyTelling asks viewers to scroll down to explore visual stories. It is similar to slide shows but provide more transition techniques such as scrolling interactions and animation effects. SketchStory spans annotated charts, partitioned posters, and videos. It allows authors to sketch along the real-time presentation with pen and touch

20

interactions as shown in Figure 3.11. It enriches the storytelling with audio narration and real-time feedback.

Figure 3.11. Telling a real-time story using SketchStory [49].

However, we have seen that the proliferation of different narrative visualization genres and richness of techniques they employ in practice are largely limited to the authoring tools. Many research works have developed novel design and interactions. Most of them are not embedded into existing tools and require extra programming capabilities which practitioners tend not to equip with.

CHAPTER 4

NARRATIVE AUTHORING TOOLS

As summarized at the end of Chapter 3, by learning from strategies used in practice and reflecting on the problems when crafting stories, the research community has devoted great efforts to building tools that help users to easily visualize data and communicate narratives. However, with the proliferation of narrative authoring tools, it arouses another challenge to evaluate and compare the strengths and weakness among these tools. A feasible approach is to adjust the different evaluation standards from the perspective of the research contribution. Therefore, in this chapter, we review the related narrative authoring tools with regards to their contributions, evaluation criteria, and evaluation methods in detail, and discuss the research opportunities concerning different parts.

4.1 Approaches

Different approaches feature prominently in tools intended to craft data-driven narratives. The approach selection is closely related to the contribution of the tool. In the following, we give an overview of existing tools that facilitate storytelling.

4.1.1 Automatic Generation

Automatic tools can be found in the applications of annotations and presentation orders. Contextifier [41] and NewsViews [35] are both used to construct annotated visualization in news to contextualize text articles. Some works [40, 46] leverage the directed graph model to automatically reason visualization sequencing in a linear order. However, no existing tool could craft a rich data story automatically. Most of them augment the storytelling in a specific element to some extent. This form provides hints for tool builders to determine which element in data-driven narratives could be improved by automatic generation and which asks for human-in-the-loop tools.

4.1.2 Authoring tools

There already exists a multitude of general-purpose tools to create data graphics or videos, such as Adobe Creative Suite, Microsoft Excel and Power BI. However, most of them have

a steep learning curve and are difficult for users to manipulate data directly. Thus, crafting data-driven narratives is much time-consuming with these tools.

Recent narrative authoring tools [17, 63] are more expressive and efficient with a curated collection of visual design choices. For example, Satyanarayan and Heer [63] instantiated a storytelling model in Ellipsis, which provides an interface for users to add elements to augment data storytelling in visualization. Amini et al. [17] implemented DataClips, an authoring tool to craft a complete data video with a library of data-driven clips. These clips contain popular visualization types with corresponding animation types and can be easily extended to new clips. Some dedicated tools are built for specific storytelling elements. For instance, ChartAccent [60] aims to support data-driven annotation. InfoNice [72] enables easy creation of data-driven infographics. A list of narrative authoring tools with their major functions is summarized in Table 4.1. However, the proliferation of narrative authoring tools brings about difficulties and challenges for users to select the appropriate tool and evaluate its effectiveness.

4.1.3 Programming

It is certainly possible to craft narrative visualization by programming. Actually, many computer science researchers work on this way and also provide a large palette of open-source libraries and packages, such as D3.js, processing, and Vega. Notably, there even exist libraries specific to some function, e.g., swoopyDrag.js or labella.js for data-driven annotation and labelling, and anime.js for animation. Iterative design between creating visualization with D3 and editing it with Illustrator is doable as well with the Hanpuku bridge tool [21]. However, although these programming languages lend much freedom to customize data-driven narratives precisely, many visualization practitioners like data journalists are not equipped with capable programming skills. They require an interactive interface to author the story directly and intuitively. Therefore, many research efforts concentrate on developing better tools to ease the creation.

4.2 Evaluation Criteria

Actually, information visualization evaluation always presents difficulties and has been discussed frequently among the research community. Many works [57, 48] contribute to establishing a series of evaluation criteria and methods. The BELIV workshop series focuses on the challenges of evaluation in visualization. However, the evaluation targeting at narrative

authoring tools faces another obstacle. For example, typical evaluation is based on controlled experiments to compare users' performance such as completion time and error rates. But it is challenging to assess these criteria in view of an authoring system which supports users' creative work. It is hard to calculate time and errors for a visual story creation. What's more, the primary problem is the lack of an appropriate comparative system. Most of novel authoring systems are designed and developed due to the insufficient functions of existing systems. It is reasonable that the newly-developed tool should outperform the existing one. However, users must have the diverse knowledge of tools. Like, a professional Adobe Illustrator designer gets accustomed to using this software but may not have enough time to learn the full capacities of the new authoring system, where the tutorial is mostly limited to a short study session. It brings difficulties to control essential factors and set tasks during the evaluation experiment. Therefore, a series of evaluation criteria is required to guide the approach.

Amini et al. [32] recently proposed a variety of criteria to consider when evaluating a narrative authoring tool. It is natural that the list does not cover an exhaustive set of criteria. It still sheds light on our thinking. We propose a tailored list of criteria combined with their work and evaluation standards derived from the previous survey of relevant tools. And with an increasing trend of narrative authoring tools in the future, the study of criteria for evaluating data-driven storytelling tools will grow at length. Currently, they can be grouped to eight aspects as follows.

- **Expressiveness**. The tool provides with a number of design choices, including various visualization types, annotation, and interactions. Users are allowed to create a range of possible visual data stories with the system.
- **Flexibility**. It relates to the extent to which an author can extend the given library of design choices and make custom designs. This criterion demands a balance between creative work and ease of creation.
- **Guidance**. Some tools target at novice users who are likely to have problems when creating stories. It would be of great help to provide design support such as making layout suggestions during the process.
- **Efficiency**. It concerns the speed to produce a data-driven story using the tool, i.e., users can create how many stories within the limited time. This criterion is of great importance to some kinds of users such as journalists who need to produce stories frequently and quickly.
- **Usability**. Given an expected data-driven story, the tool should provide a series of functions for users to craft it.

- **Learnability**. The learning curve of the storytelling tool are closely related to the number of functions and intuitive interactions the tool provides. Qualified documents and abundant tutorial materials can also help ease the learning.

- **Integration**. An authoring tool for narrative visualization is supposed to be integrated into current workflow and assist in authors' creative work. This criterion asks tool builders to understand the methods authors are using to create data-driven stories.

- **Iteration**. The process to craft a data-driven story inevitably involves several iterations. It assesses the extent to which the tool supports iterative controls.

Actually, a qualified narrative authoring tool is not necessarily required to satisfy all of these criteria. Good tool designers should strike a balance and make trade-offs among these criteria when they build such tools. Suitable evaluation standards should depend upon the research contribution that the system intends to achieve. For example, if the contribution of a tool is to help automatically construct annotated visualizations to contextualize data-rich texts (e.g., [41]), flexibility may be not adopted for the evaluation. But when the system aims to allow users to realize a broad range of designs (e.g., [72]), the evaluation should attach great importance to the expressiveness criterion.

4.3   Evaluation Methods

Ren et al. [61] discussed the evaluation methods used in many visualization authoring systems whose main research contribution lies on âĂIJexpressiveâĂİ designs. They summarized six approaches to conduct evaluation in these works. We, here, mainly follow their methodology but just focus on storytelling tools as discussed in Section 4.1. Since general visualization authoring systems do not necessarily equip with the capability of storytelling, they mostly provide visualization for data exploration and analysis like Microsoft Excel. Authors need to edit the generated visualization with other tools to support narratives. Although some tools like DataInk [73] focus on binding data with Information graphics, their main contribution lies on creative design not data-driven storytelling. Notably, we include InfoNice [72], because it is incorporated into an analytics tool, bridging the gap between data exploration and presentation. Another exception is SketchStory [49], which allows presenters to freeform sketch on a digital whiteboard during the process. Therefore, we summarize the evaluation methods of narrative authoring tools in Table 4.1, which is different from the one of Ren et al. in some aspects.

	Approaches	Empirical Study	Reproduct-ion Study	Free-Form Study	Comparative Study	Gallery
Contextifier[41]	automated	survey			y	
NewsViews[35]	automated	survey				
Sequence[40]	automated	survey				
GraphScape[46]	automated					
DataClips[17]	authoring	survey		y	video design	
ChartAccent[60]	authoring	survey	y			
Timeline Storyteller[15]	authoring	survey				
Ellipsis[63]	authoring	interviews		y		y
InfoNice[72]	authoring				y	y
SketchStory[49]	authoring	wizard-of-oz	y	y	y	
Hanpuku[21]	program					3 cases

Table 4.1. Evaluation methods used for each of the narrative authoring tools. "y" means the evaluation method was conducted for this tool.

4.3.1 Empirical Study

Brehmer et al. [24] introduced empirical study in a broad way with a multitude of methods, including observational field studies, interviews, and so on. They encouraged pre-design empirical studies for information visualization. One obvious benefit was to characterize work practices and associated problems, thus informing the design space. In the data-driven storytelling field, many works present exemplary instances through a pre-design empirical study. A favorite way is to analyze a curated collection of existing stories. For example, Brehmer et al. first surveyed 263 timelines and identified 14 design choices with three dimensions [23] to build Timeline Storyteller [15]. Amini et al. [18] attained empirical knowledge with two exploratory studies. The first was to identify high-level structures with key components based on a qualitative examination of 50 online data videos through the cinematography lens. Then, they observed the creation process of 13 experienced storytellers. The derived design space worked as the cornerstone for DataClips [17] they developed for authoring data videos. Another way to understand empirical knowledge is to involve end users. For example, Satyanarayan and Heer [63] conducted formative interviews with experienced story makers to learn the general authoring process and limitations of current tools.

SketchStory [49] conducted a Wizard of Oz study and usage tests with participants [70] to iterate the design.

In summary, when designing authoring tools, empirical studies play a critical role both in informing design spaces and justifying trade-offs. They provide an initial understanding about the tool framework.

4.3.2   Reproduction Study

In a reproduction study, participants are asked to reproduce a copy of the given visual data stories. This study is often served as a tutorial to familiarize subjects with the authoring tool, which first examines its learnability. ChartAccent [60] and SketchStory [49] both featured a reproduction study. It should be noted that this study can encounter the usability issues. For example, participants complained that ChartAccent [60] didnâĂŹt support z-order manipulation. SketchStory [49] iterated the system design after the reproduction study and free-from study.

However, the main weakness is that a reproduction study does not conform to the reality. Such tools are built to support creative work and author novel data-driven narratives. The finding from a reproduction study can not totally demonstrate the usability.

4.3.3   Free-Form Study

A free-form study requires participants to author their own visual data stories with the tool. DataClips [17], Ellipsis [63], and SketchStory [49] all utilized this evaluation. This study simulates the real-world usage scenario and can be used to assess multiple aspects of the authoring tool, such as expressiveness, learnability, and usability. In this experiment, it should be noted that there is likely to be a phase before creation for users to get familiar with data and facts [17], as well as generate and sketch ideas [49].

However, a free-form study tends to last limited time duration. That means the participants may not have enough time to master all functions the tool provides, as well as create a variety of visualization artifacts. Therefore, the expressiveness canâĂŹt be demonstrated fully in this study.

4.3.4   Comparative Study

Tool builders are likely to conduct a controlled experiment to compare their tools with existing commercial software tools. It can be used to highlight the usability and efficiency.

For example, Wang et al. compared InfoNice [72] with Data-Driven Guides [45] and measured time to create an infographic-style visualization with the same data using these two tools. Lee et al. [49] employed SketchStory to do real-time presentations as compared to the traditional way using PowerPoint. They measured the engagement for both the audience and presenter and found sketching facilitate storytelling in that scenario. Amini et al. [17] asked participants to create data video using DataClips against the combination of Adobe Illustrator and Adobe Effects, because there was no such similar tool to craft data videos directly. Thus, it brings a problem that the experiment is not under strict controls. It is hard to find an appropriate evaluation matrix to compare tools accurately. We have remarked that research prototypes contribute to novel features and may not incorporate some engineering functions such as âĂIJundo/redoâĂĪ. However, in practice, these functions do work. Thus, it may decrease the impression among participants and also increase the creation time in the comparative study.

4.3.5 Gallery

As we have mentioned in Section 4.3.3, subjects can only master limited capability to show expressiveness of the tool in limited time. Researchers recently have thought to provide a gallery themselves to demonstrate expressiveness. Ellipsis [63], InfoNice [72], and Hanpuku [21] all offer galleries for review. The benefits are obvious in expressiveness and efficiency. However, a major concern might be the target users canâĂŹt produce such visualization artifacts as the authors do. Detailed tutorials and instruments should be complemented. Since the gallery is provided by the tool builder, it can not show the usability and learnability of tool. It is supposed to be combined with other evaluation methods.

CHAPTER 5

CONCLUSION AND FUTURE WORK

The emergence and fast development of web-based visualization technologies paves the way for a proliferation of online visual data-driven stories. In this survey, we review visual design in data-driven storytelling and also pay attention to tools that facilitate authoring visual data stories. After introducing two taxonomies respectively based on narrative genres and storytelling roles in visualization, we organize our taxonomy based on three primary components to guide visual design in composing data-driven narratives. Then we investigate narrative authoring tools to assist in crafting high quality visualization artifacts and reflect on methodologies adopted to evaluate these tools.

The research in data-driven storytelling is a growing discipline. There is a large palette of visual data storytelling techniques emerging and recurring in a multitude of forms ranging from animated infographics and videos to interactive online visualizations. However, visualization tools are not keeping up with the pace of innovation. In addition, research on understanding factors to compose high quality data stories should be extended to encompass more advanced techniques, e.g., non-linear narratives and animation. Nowadays, there are efforts being made to understand visual data-driven stories and build narrative authoring tools to facilitate storytelling. We specifically summarize a few challenges and potential research directions which are worthwhile studying further as follows.

Understanding of stories. Some research has demonstrated essential elements of data-driven storytelling. However, most of them analyze explicit techniques, such as text annotations to augment visualization, and timelines or progress bars to provide navigation aids. With visual data stories becoming more abundant and creative, there is certainly room for improvement and new findings. For example, little attempt has been made to understand stories from the perspective of cognitive psychology and perception. How does a compelling story engage a broad audience? Does it ignite emotions among the viewers? If yes, is there any common technique such as visualizing personal data or providing situated design? In addition, few researches have worked on interpreting non-linear narratives and adding audio narration in data-driven storytelling, whereas many acclaimed data videos or comics [16, 20] employ these patterns and convey good stories. How to better incorporate these implicit techniques into the visual design of data-driven stories can be a potential direction for future exploration.

Support from authoring tools. As mentioned in Section 4, there is an increasing trend of authoring visualization tools, whereas these tools are not keeping pace with innovation. For example, current authoring tools only provide limited animation, which makes it uncommon to see such animated transitions in practice. Besides, they fail to support guidance or warnings during the authoring process. This is mainly due to the incomplete understanding of design choices, like consistency constraints. Another important factor is the end-users who actually craft visual stories for data insight communication. They can be experts or novice users. Thus, different design considerations should be recognized in view of their expertise. For experts, how to integrate these visualization tools into their regular workflows to improve efficiency might be a major concern. While for non-experts, tool builders should provide sufficient tutorials for the use of the tool. Guidance or creation suggestions during the process are welcome as well.

BIBLIOGRAPHY

[1] Data-driven storytelling. [Online]. Available: https://www.microsoft.com/en-us/research/project/data-driven-storytelling/

[2] Data journalism in 2017. [Online]. Available: https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf?_ga=2.84903257.176494904.1535965689-1811352806.1535012924

[3] Epfl data monolith. [Online]. Available: https://vimeo.com/194404753

[4] Form follows function. [Online]. Available: https://en.wikipedia.org/wiki/Form_follows_function

[5] Gapminder. [Online]. Available: https://www.gapminder.org/

[6] Global wealth inequality. [Online]. Available: https://www.youtube.com/watch?v=uWSxzjyMNpU

[7] Green honey. [Online]. Available: http://muyueh.com/greenhoney/

[8] A history of coca-cola advertising slogans. [Online]. Available: https://www.coca-colacompany.com/stories/coke-lore-slogans

[9] How americans die. [Online]. Available: https://www.bloomberg.com/graphics/dataview/how-americans-die/

[10] Morphing. [Online]. Available: https://en.wikipedia.org/wiki/Morphing

[11] Oxford english dictionary. [Online]. Available: http://www.oed.com/

[12] The rhythm of food. [Online]. Available: http://rhythm-of-food.net/

[13] Scientific proof that americans are completely addicted to trucks. [Online]. Available: https://www.bloomberg.com/graphics/2015-auto-sales/

[14] Scrollytelling. [Online]. Available: http://vallandingham.me/scroll_talk/examples/

[15] Timeline storyteller. [Online]. Available: https://timelinestoryteller.com/

[16] Wealth inequality in america. [Online]. Available: https://www.youtube.com/watch?v=QPKKQnijnsM

[17] F. Amini, N. H. Riche, B. Lee, A. Monroy-Hernandez, and P. Irani, "Authoring data-driven videos with dataclips," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 501–510, Jan. 2017.

[18] F. Amini, N. Henry Riche, B. Lee, C. Hurter, and P. Irani, "Understanding data videos: Looking at narrative visualization through the cinematography lens," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15.   New York, NY, USA: ACM, 2015, pp. 1459–1468. [Online]. Available: http://doi.acm.org/10.1145/2702123.2702431

[19] F. Amini, N. H. Riche, B. Lee, J. Leboe-McGowan, and P. Irani, "Hooked on data videos: Assessing the effect of animation and pictographs on viewer engagement," in *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, ser. AVI '18.   New York, NY, USA: ACM, 2018, pp. 21:1–21:9. [Online]. Available: http://doi.acm.org/10.1145/3206505.3206552

[20] B. Bach, Z. Wang, M. Farinella, D. Murray-Rust, and N. Henry Riche, "Design patterns for data comics," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18.   New York, NY, USA: ACM, 2018, pp. 38:1–38:12. [Online]. Available: http://doi.acm.org/10.1145/3173574.3173612

[21] A. Bigelow, S. Drucker, D. Fisher, and M. Meyer, "Iterating between tools to create and edit visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 481–490, Jan. 2017.

[22] M. Bostock, V. Ogievetsky, and J. Heer, "D3: Data-driven documents," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2301–2309, Dec. 2011.

[23] M. Brehmer, B. Lee, B. Bach, N. H. Riche, and T. Munzner, "Timelines revisited: A design space and considerations for expressive storytelling," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 9, pp. 2151–2164, Sep. 2017.

[24] M. Brehmer, S. Carpendale, B. Lee, and M. Tory, "Pre-design empiricism for information visualization: Scenarios, methods, and challenges," in *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, ser. BELIV '14.   New York, NY, USA: ACM, 2014, pp. 147–151. [Online]. Available: http://doi.acm.org/10.1145/2669557.2669564

[25] N. H. R. J. S. C. D. Stolper, B. Lee, "Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories," Tech. Rep.,

2016. [Online]. Available: https://www.microsoft.com/en-us/research/publication/emerging-and-recurring-data-driven-storytelling-techniques-analysis-of-a-curated-collection-of-

[26] S. Carpendale, N. Diakopoulos, N. H. Riche, and C. Hurter, "Data-driven storytelling (dagstuhl seminar 16061)," *Dagstuhl Reports*, vol. 6, no. 2, pp. 1–27, 2016.

[27] F. Chevalier, P. Dragicevic, and S. Franconeri, "The not-so-staggering effect of staggered animated transitions on visual tracking," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2241–2250, Dec. 2014.

[28] N. Cohn, "Visual narrative structure," *Cognitive Science*, vol. 37, no. 3, pp. 413–452, 2013.

[29] P. Dragicevic, A. Bezerianos, W. Javed, N. Elmqvist, and J.-D. Fekete, "Temporal distortion for animated transitions," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 2009–2018. [Online]. Available: http://doi.acm.org/10.1145/1978942.1979233

[30] S. Drucker and R. Fernandez, "A unifying framework for animated and interactive unit visualizations," Microsoft Research, Tech. Rep., 2015. [Online]. Available: https://www.microsoft.com/en-us/research/publication/a-unifying-framework-for-animated-and-interactive-unit-visualizations/

[31] N. Elmqvist, P. Dragicevic, and J. Fekete, "Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1539–1148, Nov. 2008.

[32] G. B. C. E. F. Amini, M. Brehmer and B. Wiederkehr., *Evaluating data-driven stories & storytelling tools.* A K Peters/CRC Press, 2018.

[33] D. Fisher, *Animation for Visualization: Opportunities and Drawbacks.* Eds. Sebastopol: OâĂŹReilly, 2010, ch. 19, pp. 329–352.

[34] G. W. Furnas, "Generalized fisheye views," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '86. New York, NY, USA: ACM, 1986, pp. 16–23. [Online]. Available: http://doi.acm.org/10.1145/22627.22342

[35] T. Gao, J. R. Hullman, E. Adar, B. Hecht, and N. Diakopoulos, "Newsviews: An automated pipeline for creating custom geovisualizations for news," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, pp. 3005–3014. [Online]. Available: http://doi.acm.org/10.1145/2556288.2557228

[36] P. Goffin, J. Boy, W. Willett, and P. Isenberg, "An exploratory study of word-scale graphics in data-rich text documents," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 10, pp. 2275–2287, Oct. 2017.

[37] D. Guilmaine, C. Viau, and M. J. McGuffin, "Hierarchically animated transitions in visualizations of tree structures," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, ser. AVI '12. New York, NY, USA: ACM, 2012, pp. 514–521. [Online]. Available: http://doi.acm.org/10.1145/2254556.2254653

[38] J. Heer and G. Robertson, "Animated transitions in statistical data graphics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1240–1247, Nov. 2007.

[39] J. Hullman and N. Diakopoulos, "Visualization rhetoric: Framing effects in narrative visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2231–2240, Dec. 2011.

[40] J. Hullman, S. Drucker, N. H. Riche, B. Lee, D. Fisher, and E. Adar, "A deeper understanding of sequence in narrative visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2406–2415, Dec. 2013.

[41] J. Hullman, N. Diakopoulos, and E. Adar, "Contextifier: Automatic generation of annotated stock visualizations," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13. New York, NY, USA: ACM, 2013, pp. 2707–2716. [Online]. Available: http://doi.acm.org/10.1145/2470654.2481374

[42] J. Hullman, R. Kosara, and H. Lam, "Finding a clear path: Structuring strategies for visualization sequences," *Comput. Graph. Forum*, vol. 36, no. 3, pp. 365–375, 2017.

[43] S. Huron, R. Vuillemot, and J. Fekete, "Visual sedimentation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2446–2455, Dec. 2013.

[44] N. W. Kim, B. Bach, H. Im, S. Schriber, M. Gross, and H. Pfister, "Visualizing nonlinear narratives with story curves," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 595–604, Jan. 2018.

[45] N. W. Kim, E. Schweickart, Z. Liu, M. Dontcheva, W. Li, J. Popovic, and H. Pfister, "Data-driven guides: Supporting expressive design for information graphics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 491–500, Jan. 2017.

[46] Y. Kim, K. Wongsuphasawat, J. Hullman, and J. Heer, "Graphscape: A model for automated reasoning about visualization similarity and sequencing," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. New York, NY, USA: ACM, 2017, pp. 2628–2638. [Online]. Available: http://doi.acm.org/10.1145/3025453.3025866

[47] R. Kosara and J. Mackinlay, "Storytelling: The next step for visualization," *Computer*, vol. 46, no. 5, pp. 44–50, May 2013.

[48] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale, "Empirical studies in information visualization: Seven scenarios," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 9, pp. 1520–1536, Sep. 2012.

[49] B. Lee, R. H. Kazi, and G. Smith, "Sketchstory: Telling more engaging stories with data through freeform sketching," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2416–2425, Dec. 2013.

[50] B. Lee, N. H. Riche, P. Isenberg, and S. Carpendale, "More than telling a story: Transforming data into visually shared stories," *IEEE Computer Graphics and Applications*, vol. 35, no. 5, pp. 84–90, Sep. 2015.

[51] Y. Lin and R. Vuillemot, "Spirograph designs for ambient display of tweets," *IEEE VIS*, 2013.

[52] R. May, P. Hanrahan, D. A. Keim, B. Shneiderman, and S. Card, "The state of visual analytics: Views on what visual analytics is and where it is going," in *Proc. IEEE Symp. Visual Analytics Science and Technology*, Oct. 2010, pp. 257–259.

[53] S. McKenna, N. H. Riche, B. Lee, J. Boy, and M. Meyer, "Visual narrative flow: Exploring factors shaping data visualization story reading experiences," *Computer Graphics Forum (EuroVis '17)*, vol. 36, no. 3, pp. 377–387, 2017.

[54] D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe, and J. Heer, "Formalizing visualization design knowledge as constraints: Actionable and extensible models in draco," *IEEE Transactions on Visualization and Computer Graphics*, p. 1, 2018.

[55] D. Park, S. M. Drucker, R. Fernandez, and N. Elmqvist, "Atom: A grammar for unit visualizations," *IEEE Transactions on Visualization and Computer Graphics*, p. 1, 2018.

[56] C. Plaisant, J. Grosjean, and B. B. Bederson, "Spacetree: supporting exploration in large node link tree, design evolution and empirical evaluation," in *Proc. INFOVIS 2002. IEEE Symp. Information Visualization*, Oct. 2002, pp. 57–64.

[57] C. Plaisant, "The challenge of information visualization evaluation," in *Proceedings of the Working Conference on Advanced Visual Interfaces*, ser. AVI '04. New York, NY, USA: ACM, 2004, pp. 109–116. [Online]. Available: http://doi.acm.org/10.1145/989863.989880

[58] Z. Qu and J. Hullman, "Evaluating visualization sets: Trade-offs between local effectiveness and global consistency," in *BELIV*, 2016, pp. 44–52.

[59] Z. Qu and J. Hullman, "Keeping multiple views consistent: Constraints, validations, and exceptions in visualization authoring," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 468–477, Jan. 2018.

[60] D. Ren, M. Brehmer, B. Lee, T. Hollerer, and E. K. Choe, "Chartaccent: Annotation for data-driven storytelling," in *Proc. IEEE Pacific Visualization Symp. (PacificVis)*, Apr. 2017, pp. 230–239.

[61] D. Ren, B. Lee, M. Brehmer, and N. Henry Riche, "Reflecting on the evaluation of visualization authoring systems." IEEE, October 2018. [Online]. Available: https://www.microsoft.com/en-us/research/publication/reflecting-on-the-evaluation-of-visualization-authoring-systems/

[62] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko, "Effectiveness of animation in trend visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1325–1332, Nov. 2008.

[63] A. Satyanarayan and J. Heer, "Authoring narrative visualizations with ellipsis," *Comput. Graph. Forum*, vol. 33, no. 3, pp. 361–370, 2014.

[64] E. Segel and J. Heer, "Narrative visualization: Telling stories with data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1139–1148, Nov. 2010.

[65] C. Tong, R. C. Roberts, R. Borgo, S. Walton, R. S. Laramee, K. Wegba, A. Lu, Y. Wang, H. Qu, Q. Luo, and X. Ma, "Storytelling and visualization: An extended survey," *Information*, vol. 9, no. 3, p. 65, 2018.

[66] B. Tversky, J. B. Morrison, and M. Bétrancourt, "Animation: can it facilitate?" *Int. J. Hum. Comput. Stud.*, vol. 57, no. 4, pp. 247–262, 2002.

[67] J. J. van Wijk and W. A. A. Nuij, "A model for smooth viewing and navigation of large 2D information spaces," *IEEE Transactions on Visualization and Computer Graphics*, vol. 10, no. 4, pp. 447–458, Jul. 2004.

[68] B. C. Veglis, A., "Towards a taxonomy of data journalism," *Journal of Media Critiques*, vol. 3, no. 11, pp. 109–118, 2017.

[69] M. Waldner, M. L. Muzic, M. Bernhard, W. Purgathofer, and I. Viola, "Attractive flicker – guiding attention in dynamic narrative visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2456–2465, Dec. 2014.

[70] J. Walny, B. Lee, P. Johns, N. H. Riche, and S. Carpendale, "Understanding pen and touch interaction for data exploration on interactive whiteboards," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2779–2788, Dec. 2012.

[71] Y. Wang, Z. Chen, Q. Li, X. Ma, Q. Luo, and H. Qu, "Animated narrative visualization for video clickstream data," in *SIGGRAPH ASIA 2016 Symposium on Visualization*, ser. SA '16. New York, NY, USA: ACM, 2016, pp. 11:1–11:8. [Online]. Available: http://doi.acm.org/10.1145/3002151.3002155

[72] Y. Wang, H. Zhang, H. Huang, X. Chen, Q. Yin, Z. Hou, D. Zhang, Q. Luo, and H. Qu, "Infonice: Easy creation of information graphics," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: ACM, 2018, pp. 335:1–335:12. [Online]. Available: http://doi.acm.org/10.1145/3173574.3173909

[73] H. Xia, N. Henry Riche, F. Chevalier, B. De Araujo, and D. Wigdor, "Dataink: Direct and creative data-oriented drawing," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: ACM, 2018, pp. 223:1–223:13. [Online]. Available: http://doi.acm.org/10.1145/3173574.3173797