

# Emerging and Recurring Data-Driven Storytelling Techniques: Analysis of a Curated Collection of Recent Stories

Charles D. Stolper, Bongshin Lee, Nathalie Henry Riche, and John Stasko

**Abstract**—Storytelling with data is becoming an important component of many fields such as graphic design, the advocacy of causes, and journalism. New techniques for integrating data visualization into narrative stories have now become commonplace. Authors are enabling new reader experiences, such as linking textual narrative and data visualizations through dynamic queries embedded in the text. Novel means of communicating position and navigating within the narrative also have emerged, such as utilizing scrolling to advance narration and initiate animations. We advance the study of narrative visualization through an analysis of a curated collection of recent data-driven stories shared on the web. Drawing from the results of this analysis, we present a set of techniques being employed in these examples, organized under four high-level categories that help authors to tell stories in creative ways: communicating narrative and explaining data, linking separated story elements, enhancing structure and navigation, and providing controlled exploration. We describe the benefits of each storytelling technique along with a number of example applications of the ideas through recent data-driven stories. Additionally, we discuss the trends we observed as well as how the field has evolved and grown. Finally, we conclude with a discussion of areas for future research.

**Index Terms**—Storytelling, narrative visualization, information visualization.



## 1 INTRODUCTION

THE art of storytelling with visualization has rapidly evolved over the past few years. Advancements in web-based visualization technology (e.g., the rapid adoption of D3 [7]) fostered novel dynamic data-driven storytelling on the web. Authors have introduced a variety of novel narrative visualization techniques and refined existing ones while narrative visualizations have continued to grow in popularity amongst both authors and readers. Some of these stories even include innovative techniques that are new to the information visualization research community. For example, *Visualizing MBTA Data* [9] tells the story of Boston’s public transit using a range of interesting storytelling and visualization techniques including clever linkages between text and visualization. We speculate that the developers of exploratory, analytic visualizations could benefit by learning about and adopting ideas from many of these stories.

In this work, we set out to investigate the current state of data-driven storytelling. Our goal was to identify and describe the spectrum of techniques that today’s data-storytellers are using, and to inform the design of future data-storytelling tools. In doing so, we provide a curated collection of interesting examples of storytelling with data visualization and the techniques each uses. We reflect on how authors currently tell stories to identify the aspects that have been under-explored by the research community and that have great potential.

We report on a qualitative analysis of 45 stories, including news organizations’ online presences (e.g., The New York

Times, The Economist, FiveThirtyEight), popular visualization blogs (e.g., Visual.ly, FlowingData, Eagereyes), and storytelling tools’ websites (e.g., Tableau Public, Infogr.am). To have a more focused and detailed discussion, we reviewed storytelling through Lee et. al’s lens [23], excluding those extreme “reader-driven” storytelling examples that lack an author-defined “plot.” Our set of stories ranges from more conventional narrative visualizations, such as long articles including visualizations or data-driven videos, to novel or experimental examples of data-driven storytelling. The results of our analysis are 20 visual, data-driven storytelling techniques falling under four broad categories of aids that authors use to help tell their stories. In describing each of the categories, we provide example stories from the dataset that exemplify the category and the associated techniques used to achieve it. We also discuss the trends we observed across these different techniques and how the field has evolved since its beginnings. We conclude with a discussion of research opportunities especially for the design of authoring tools for data-driven storytelling.

The contributions of this article are fourfold: 1) Descriptions and examples of 20 data-driven storytelling techniques grouped under four broad categories common to this area, 2) A curated collection of 45 author-driven narrative visualization stories coded with the techniques and categories, 3) An analysis of how the area has evolved in recent years and the techniques that appear to be growing in popularity, and 4) Research opportunities specifically for the design of authoring tools for data-driven storytelling derived from our analysis.

## 2 RELATED WORK

### 2.1 Understanding Narrative Visualization

Recognizing the increasing need and use of visualization in telling stories with data, the visualization research community

- Charles D. Stolper and John Stasko are with the School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, 30308. E-mail: {charles.stolper,stasko}@cc.gatech.edu.
- Bongshin Lee and Nathalie Henry Riche are with Microsoft Research, One Microsoft Way, Redmond, WA 98052. E-mail: {bongshin,nath}@microsoft.com.

has made several efforts to better understand and characterize this topic. In 2010, Segel and Heer [28] introduced the term *narrative visualization* to describe asynchronous storytelling using data-driven graphics. Through an analysis of 58 examples of narrative visualization, they identified seven *genres* of narrative visualization: magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show, and film/video/animation, and presented three design patterns for interactive stories: the martini glass structure, interactive slideshows, and drill-down stories. Our analysis differs from theirs in that we focus on more detailed storytelling techniques (particular mechanisms or operations) rather than the broader genres and design patterns they examined. For example, we found stories where multiple different techniques were used in a particular genre or pattern from their analysis. Hullman and Diakopoulos [14] expanded the discussion of narrative visualization, studying the rhetorical devices used in narrative visualization through the analysis of 51 example narrative visualizations. We aim to further advance this discussion to incorporate the novel data-driven storytelling techniques that have emerged since these two articles were published.

More recently, other researchers have conducted analyses to investigate ways to help people create data-driven stories more easily and effectively. In an effort to understand the effect of sequences in narrative visualization to inform the design of an automatic sequencing algorithm, Hullman et al. [16] provide a focused analysis of transitions between scenes in 42 “explicitly-ordered” stories. These stories included slideshows (interactive, recorded live and archived online, or merely archived online), animated data videos, and interactive timelines. Inspired by the growing number of data videos, Amini et al. [4] analyzed 50 professionally designed data videos, one of the seven genres in [28]. They extracted the most salient constituents from these videos as a first step toward developing a novel tool that can enable people to generate compelling data stories using video. Choe et al. [9] analyzed 30 recorded presentation videos to specifically examine visualization types and annotations for Quantified-Self presenters used to convey their insight during their presentation. As discussed in the future opportunities section later, our analysis further informs the design and development of a storytelling system that help people create compelling data-driven stories.

Several researchers have discussed broader research opportunities and challenges in using visualization as a presentation and communication medium. Kosara and Mackinlay [21] provide a high-level overview of storytelling research, calling for more attention to storytelling with data. They describe different storytelling scenarios, list noteworthy examples, and discuss many future research opportunities in this area. Ma et al. [24] discuss the challenges and possibilities of telling stories using scientific visualization. Lee et al. [23] present a visual data storytelling process (or VDSP) involving roles such as data analysts, scripters, editors, presenters, and audiences. They also take a narrower view of what constitutes a visual data story than Segel and Heer in order to foster a more focused discussion. (In selecting the stories for our corpus, we used Lee et al.’s narrower scope of visual data stories [23], thereby excluding the more extreme cases of “reader-driven” storytelling that [28] discusses.)

## 2.2 Systems to Create Narrative Visualization

Just as a better understanding of narrative visualization techniques can help inform the design of visualization storytelling tools, available tools can help define and popularize techniques. While a

number of systems for more-easily telling data-driven stories exist, little research has been conducted in this area. Here we briefly summarize example research and commercial systems that influenced or might incorporate various data-storytelling techniques.

Slide-based presentation systems such as Microsoft PowerPoint [26] and Apple Keynote [5] have been used for visualization-based storytelling (mainly in a synchronous setting). Visualizations are usually embedded as static figures because of their rudimentary support for visualization. These systems have strong support for textual narrative and annotations, but tend to have little support for interaction beyond navigating between slides.

In response to this limited interaction with the data and chart components, the Tableau system has recently incorporated storytelling aspects into its software. The feature, which is called Story Points [11] (Figure 4), enables an author to create scene-based stories each consisting of visualizations or dashboards. Story Points allows the author to customize the interaction on a given dashboard, enabling stories with author-specified interaction. In Story Points, each scene is given a title-caption, which is displayed on the breadcrumbs at the top of the display along with navigation buttons. If the scene allows interaction by the reader, a reset button is provided to return the scene to the author’s initially-provided display. The New York Times recently described the internal tool (Mr. Chartmaker) they use to create and embed charts in their stories [3] including support for annotations, highlighting, adapting charts for different screen sizes, and providing authors with the ability to include interactive elements.

Additional research efforts have enabled people with little or no programming skills to create narrative visualization, supporting interactions and animations. Timeline JS [19] allows an author to create interactive timelines from various data inputs. The tool generates slides consisting of text, images, videos, and maps along with an interactive timeline that can be used as navigation between the slides. TimeLineCurator [12] is a web-based tool for creating Timeline JS projects. It helps people easily generate a visual timeline by automatically extracting temporal events from freeform text and enabling them to interactively curate the events. The VisualEyes tool [10] provides an interface for creating data-driven stories. The interface allows the author to add visualizations, images, text, and maps, all coordinated through the use of interactive widgets. While the interface does not require programming, many of the customization settings are lower-level and benefit from significant experience with the system. Ellipsis [27] was designed to allow the author to incorporate storytelling techniques (such as data-driven annotation, triggered animation, and incremental data loading) into their stories with reduced coding.

Due to the popularity of infographics, the last decade has seen an emergence of tools supporting their creation. These tools range from those supporting the creation of a simple form containing a single chart, to more sophisticated ones including multiple visuals, text, and limited interactions. For example, Infogr.am [2] and Infoactive [1] are two websites for creating interactive infographics. Each allows an author to upload data in order to generate various chart types, including controlled interaction such as filtering using legends. The author can add section headers, narrative text, and images in order to convey the story. Each also supports a variety of themes that modify the color scheme, text font, and header styles. We aim to expand the understanding of narrative visualization categories (and associated visual storytelling techniques), informing the design of these and future storytelling systems.

SketchStory [22] is a rare system designed explicitly for synchronous data storytelling. It allows an author to create a story in advance and then use simple gestures to present the story to an audience. SketchStory includes support for interaction such as brushing and linking between visualizations and handwritten annotations. The gestures, however, make the system impractical for asynchronous storytelling. While we do not analyze any synchronous stories, the techniques and categories we identify from the asynchronous narrative visualization would be still useful for the design of synchronous storytelling systems.

### 3 ANALYSIS METHOD AND PROCESS

Our goals in conducting this analysis were to better understand the state-of-the-art data-driven storytelling techniques that authors are using in everyday practice, to identify new developments, and to reflect how the field has evolved recently. Our first step was to identify a corpus of data stories on which to conduct our analysis. To generate this corpus, we followed the methodology used by Segel and Heer [28], Hullman and Diakopoulos [14], and Hullman et al. [16]: We began by surveying a number of popular blogs’ “best visualization of 2014” lists [17], [18], [20], [30], [31] looking for those that incorporate stories. Additionally, we included a selection of popular or staff recommended stories from the Visual.ly community. To identify those techniques that existing tools support, we included highlighted examples from a number of tools’ webpages, including Tableau Story Points, Timeline JS, and Infogr.am. Finally, we supplemented this set with stories from frequent visual data journalism sources such as The New York Times, The Washington Post, The Guardian, Bloomberg, FiveThirtyEight, and The Economist in order to counterbalance standard practice against the novelty effect that might be present in the “best of” and “recommended” lists.

Our corpus is admittedly not comprehensive or representative of the field as a whole, since no collection short of all examples could be. We simply sought to examine a broad suite of stories and include particular examples that would be relatively representative of the current state of the field and that, in many cases, the community has found especially interesting. Throughout the article we will refrain from making specific quantitative ratios or judgments about a technique’s prevalence. Instead, we make more general estimates on the commonality and frequency of different techniques based on the larger set of examples we considered when constructing our more focused collection.

Unlike the broader scopes of Segel and Heer and Hullman et al., we were specifically looking for asynchronous, author-driven, data stories. Due to the difficulty in defining the boundary between author-driven and reader-driven stories, we used a proxy condition: the stories must have some form of author-specified ordering and narrative. The ordering condition excluded the “drill-down” stories described by Segel and Heer that give the reader a complete freedom to choose the order that they experience the story, but it included the flow-chart genre they described that implies an author-chosen ordering even if the reader ignores it. The narrative condition excluded the majority of dashboards on Tableau Public for example. We did not limit stories merely to slide-shows and videos, as Hullman et al. do, as this would exclude many of the scrolling-based stories that dominate the current landscape. Ultimately, our stories align to the category “Self-running presentations for a large audience” defined by Kosara and Mackinlay [21].

To ensure that the pieces were data-driven, we mandated that stories include a form of glyph-based data visualization. This had the effect of excluding much of the output of FiveThirtyEight, which has a penchant for including data tables rather than visualizations in much of their work. Finally, we had many discussions amongst the authors regarding “list of fact” infographics (for example, *Girl Power*<sup>1</sup> and *The Science of Happiness*<sup>2</sup>). This genre is very popular online, often consisting a tall, vertical canvas on which a designer places a series of textual blurbs and small visuals around a theme. We differentiated these “list of fact” infographics from the infographics medium as a whole because this specific sub-medium lacks authorial narrative. We chose to exclude them for this reason. In the end, the resulting dataset we document here consists of 45 stories, which were published after Segel and Heer’s analysis including 1 from 2011, 3 from 2012, 8 from 2013, 18 from 2014, and 15 from the first three months of 2015.

We began by analyzing each story, open tagging the narrative visualization features that we found in each. Once all of the stories were analyzed, we merged equivalent tags as well as noticed that some tags were a higher level of abstraction than other tags. We grouped the lower-level tags under the higher level tags, generating our set of data storytelling *categories* and *techniques*. We then recoded the entire set of stories based upon these final tags. The results of this analysis can be found in Table 1 and we use the following section to describe the categories and techniques in greater detail. An accompanying website<sup>3</sup> provides an interactive version of the table and includes direct links to each story.

Some of the techniques we identified are innovative and new (e.g., linking between the text and chart, using menus as breadcrumbs, or creatively constraining the reader through scrolling), while others are well known (e.g., tooltips or brushing and linking between visualizations). Our goal was to document all the techniques authors used, whether new or old, such that our dataset and analysis can act as a snapshot of the state of asynchronous, author-driven, visualization-driven storytelling. Later in this article we compare the set of stories we selected to those in Segel and Heer [28] to help understand how the field has changed and grown since their article was published. As time passes, undoubtedly new techniques will emerge and become more common. Further analysis following the same described method will generate equivalent snapshots in the future. Such a series of analyses enables direct comparison between different time points, allowing for identifying novel techniques and categories within the space of author-driven, visualization-driven storytelling.

### 4 VISUALIZATION-DRIVEN STORYTELLING TECHNIQUES

In this section, we enumerate the different techniques we identified under four broad categories:

- Communicating Narrative and Explaining Data
- Linking Separated Story Elements
- Enhancing Structure and Navigation
- Providing Controlled Exploration

Each category is a key component of data-driven storytelling and helps differentiate these stories from more general exploratory

1. <http://www.nssfblog.com/infographic-girl-power/>  
 2. <http://www.webpagefx.com/blog/general/the-science-of-happiness/>  
 3. <http://www.cc.gatech.edu/gvu/ii/ds/charts/corpus.html>

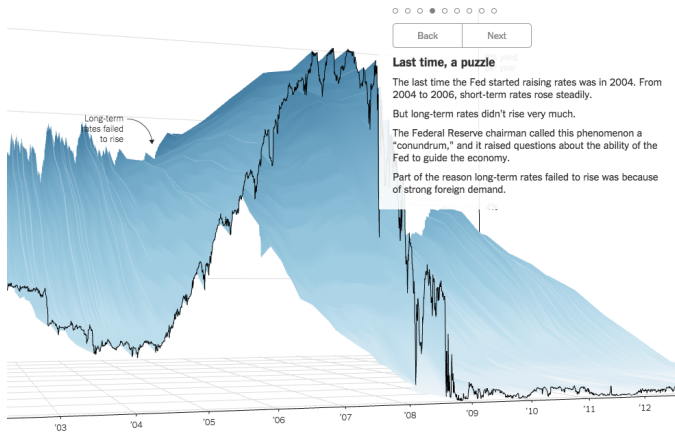


Fig. 1: A scene from *The Yield Curve*<sup>[1]</sup>, a story from The New York Times. This scene includes a slide title, textual narrative, and dot-style breadcrumbs. In addition, the annotation and arrow on the chart stay with the data point even as the reader freely rotates the 3D model.

data visualization tools. While most stories in our dataset employ techniques from multiple of these categories, the specific techniques themselves vary strongly in frequency within their category and across the collection. In the subsections that follow, we describe each of the categories and its techniques in detail, providing specific examples from the story collection that exemplify the technique. This section serves as more of a documentation and explanation of the categories and techniques. The following section then reflects on the technique’s prevalence, growth, and evolution over the last few years since the initial articles about data-driven storytelling [14], [28] appeared.

In the descriptions below, we use the term “visualization” to refer to the interactive pictorial representations of data that occur in stories. We employ the term “chart” when the pictorial representation is static. We also use the general term story “element” as a more general notion including both those pictorial representations as well as text, headers, annotations, and so on.

#### 4.1 Communicating Narrative and Explaining Data

As we focused our analysis on asynchronous stories without accompanying human presentation, nearly every story in our dataset has some form of author narration. These techniques enable authors to use the storytelling methods associated with other mediums, such as writing, in conjunction with data to help explain and communicate their messages. We found a variety of techniques used to convey their stories.

The simplest technique was simply using long **textual narrative** to convey key points, interspersing the visualizations throughout. A large number of the stories we observed utilized this technique. Note that long-form textual narration did not always constitute paragraphs in a magazine-style article. Many slide shows include longer narrative components, including *What the Jobs Report Really Means*<sup>[24]</sup> and *How Sunspots Control Global Weather*<sup>[17]</sup>, where each slide contains long narrative text to accompany the visualization. (Throughout this article we refer to the example narrative websites via a superscripted, bracketed number to an item in Appendix A. Conventional bracketed numbers still refer to cited items in the References.)

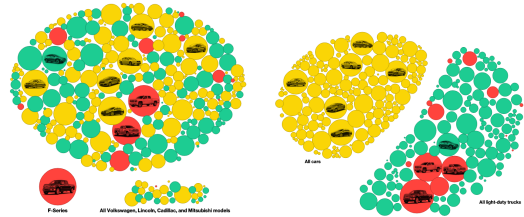


Fig. 2: Two scenes from *Addicted to Trucks*<sup>[34]</sup>, a story from Bloomberg. These consist of the single set of glyphs repositioned into different groups when the user scrolls to a certain position. The authors have also added text annotations directly on the charts.

The communication of narrative and explanation of data need not be confined to text—multiple forms of media can be employed. Data videos and recorded demonstrations such as *Wealth Inequality in America*<sup>[35]</sup> and *Will Saving Poor Children Lead to Overpopulation?*<sup>[20]</sup> use **audio narration** rather than written narration. This enables the author to more closely tie together the elements of the story (as the narrative is temporally linked to the visual elements) but in exchange for making it more difficult for the reader to experience the story at his or her own pace.

Moving past narrative text and audio narration, many stories include textual annotations to help communicate narrative and present data. In some cases, these annotations are **labeling** above or below a visualization or titling a slide or section. *Zero Hour Contracts in Four Charts*<sup>[38]</sup>, *The Evolution of Israeli Politics*<sup>[19]</sup>, and *The Yield Curve*<sup>[1]</sup> (Figure 1) use some variant of this technique. These titles are often differentiated from the rest of the text using a different typeface or font-style.

In some cases, the annotations help guide the reader through the author’s intended narrative structure. **Flowchart arrows** connect components of the story when the author’s intended ordering may be unclear. For example, *Pulp Fiction in Chronological Order*<sup>[40]</sup> and *Comprehensive History of Philosophy*<sup>[41]</sup> both use flowchart arrows to help guide the reader between the chronological events.

In all of the above techniques, the annotations are distinct from the charts. Conversely, many authors include **text annotation on visualizations** themselves. These annotations can direct the reader’s attention to those aspects that the author deems crucial to the story. These on-chart annotations can still come in a variety of forms, however. In *342,000 Swings Later*<sup>[13]</sup>, the authors mark various groups of elements in the visualization such as swings per at-bat, swings per hit, and swings per home run on the visualization rather than in the textual narrative. In *The History and Future of Everything—Time*<sup>[27]</sup>, the authors mark interesting events along the animated timeline. In some cases these events are duplicated in the audio narration, but not always. Approximately half of the stories we analyzed included some form of textual annotation directly on the chart, so it is quite a common technique.

Clearly, annotation is an important aspect of storytelling with data. Hullman et al. [15] study annotation on charts carefully, and determine different types and purposes of annotations. They also created a system that can automatically produce annotated line graphs of a stock’s behavior.

Many of the stories we analyzed provided interactive explanation in the form of **tooltips**: additional details provided to the reader when he or she mouses over an element of the visualization.



These tooltips vary in level of detail though. In some stories, such as *In Gaza, a Pattern of Conflict*<sup>[8]</sup> or *Bubble to Bust to Recovery*<sup>[25]</sup> (Figure 3), the tooltips are simply the values at a given point in the visualization. Others, such as *The World Above Us*<sup>[45]</sup>, have tooltips containing other information, in this case details about each specific satellite. One of the views in *Selvecity*<sup>[30]</sup> has tooltips that are simply larger thumbnails of the tiny thumbnails that make up the visualization. Both Tableau and Infogr.am support tooltips, and the stories in our set made with those tools, such as *The Case for a New Type of Minimum Wage*<sup>[21]</sup> (Figure 4) and *Beating Stanford in Fundraising*<sup>[4]</sup>, respectively, demonstrate that.

Other times, rather than using text to draw the reader’s attention, authors use graphical properties to perform **element highlighting**. This can be done by wrapping the highlighted elements with a shape, as the author of *Money*<sup>[33]</sup> does. Other times an author might emphasize important features while de-emphasizing the rest of a visualization, as the author of *Bubble to Bust to Recovery*<sup>[25]</sup> does in Figure 3 to highlight the “other cities” line or the author of *Winning Over Virginia*<sup>[32]</sup> does to highlight Northern Virginia on the map or the associated bar chart (Figure 5). Sometimes, rather than just fading out the color as both of those authors did, the author may choose to only color those elements that are being highlighted and maybe even add some text annotations and arrows like the authors of *The Case for a New Type of Minimum Wage*<sup>[21]</sup> (Figure 4) and *The Yield Curve*<sup>[1]</sup> (Figure 1) chose to do.

## 4.2 Linking Separated Story Elements

Data-driven stories typically contain multiple story elements including text in various forms and one or more visualizations or charts. Making connections between elements is crucial, particularly when data represented in visualizations is used to explain aspects of the story. We found three fundamental types of techniques that authors use to link different story elements such as text and visualization.

The notion of “brushing and linking” between visualizations, where clicking on an item or items in one view highlights those same items in other views, is very common in open-ended visualization systems that help people explore data. It is therefore noteworthy that we found only a few instances of stories utilizing

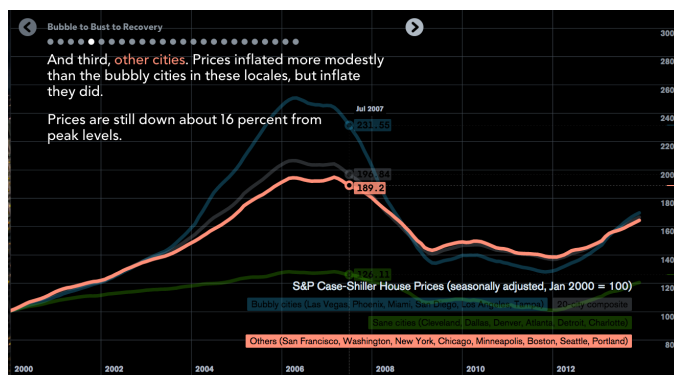


Figure 3: *Bubble to Bust to Recovery*<sup>[25]</sup>, a story from Bloomberg. This demonstrates a slide show structure using dots as breadcrumbs, highlighting elements, tooltips, and a consistent color scheme between the textual narrative and the visualization. The legends acts also as a dynamic query widget for filtering the charts.

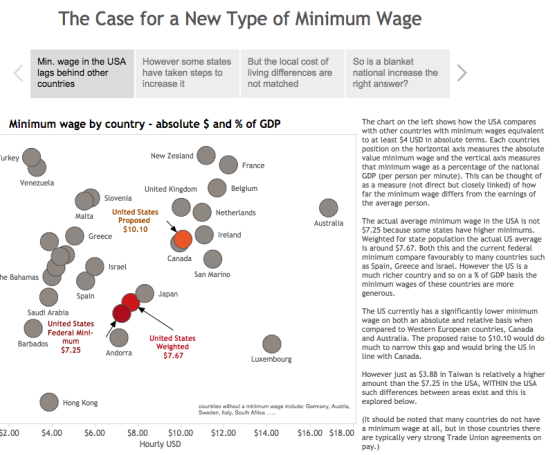


Figure 4: *The Case for a New Type of Minimum Wage*<sup>[21]</sup>, a story created using Tableau Story Points. This scene provides examples of textual narrative, slides with titles as breadcrumbs, element highlighting, and textual annotation on the chart.

the technique of **linking elements through interaction**. The only stories in our collection where interaction on one visualization is mapped to changes in another were *4th Down: When to Go for It and Why*<sup>[12]</sup>, *The Nuclear Age*<sup>[43]</sup>, *The Race Gap in America’s Police Departments*<sup>[7]</sup>, and *Visualizing MBTA Data*<sup>[9]</sup>.

In the first example, two different plots over the same data space (yards-for-a-first-down and position-on-the-field) are linked such that mousing over a point in one highlights the corresponding point in the other. *The Nuclear Age*<sup>[43]</sup> (Figure 6) consists of two visualizations: a stacked bar chart of nuclear warheads by year and country and a Gantt chart of years as a nuclear power by country. Brushing over the stacked bar chart displays the precise values for each of the country segments of the given year as labels on the Gantt chart. Clicking on a country in the Gantt chart filters the bar chart to just that country. The last story mentioned above includes many instances of paired visualizations (usually with one being an embellished form of Boston’s MBTA subway map).

Connections through interaction can extend to other types of relationships as well. Authors have started incorporating brush-

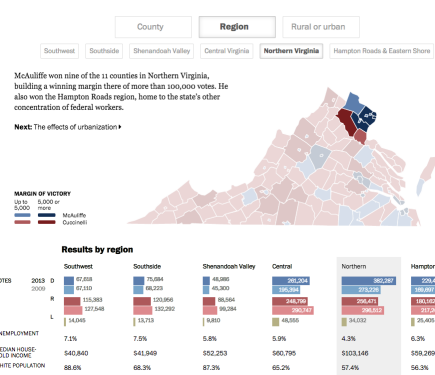


Figure 5: A scene from *Winning Over Virginia*<sup>[32]</sup>, a story from The Washington Post. This scene demonstrates breadcrumbs with button-looking scene titles, a nested linear structure, a consistent color mapping between all of the visualization components, and author highlighting of elements in the visualizations.

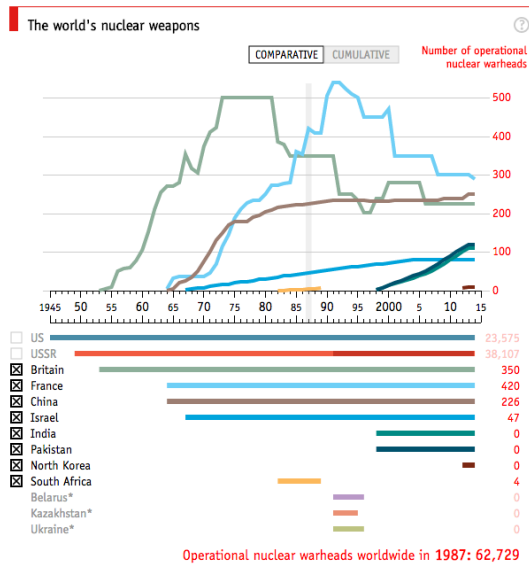


Fig. 6: *The Nuclear Age* [43], a story from The Economist. This scene includes linked visualizations between the line chart and Gantt chart, consistent colors between charts, tooltips (the figures for grey-highlighted portion of the line chart are displayed by the bar chart), and dynamic query widget (whether to display the top visualization as a comparative line chart or a cumulative stacked-bar chart).

ing and linking not just between visualizations, but between the narrative text and visualizations as well. *Visualizing MBTA Data* [9] incorporates this technique in a number of ways. Hovering over certain marked phrases in the text can trigger highlighting, annotations, filtering, or opening a details panel within the visualization. The author also uses this same approach to explain how to use some of the interactive visualizations by highlighting components when the reader mouses over them in the instructions. *Road Map* [3] (Figure 7) uses this technique to filter the custom-built word clouds based on attributes described in the text. The “numbered streets” filter is applied after the reader has clicked on the label in the narrative.

Interaction is just one way to link separated story elements. We also found multiple examples of **linking elements through color**. Authors usually accomplish this through a consistent color mapping between attributes that appear in multiple visualizations. In *In Gaza, a Pattern of Conflict* [8], for example, all of the charts of total rocket launches are in grey, while all of the charts of deaths are in brown. In *Where We Came From and Where We Went* [2] (Figure 8), the authors assign each of the regions of the United States (i.e., West, South, Midwest, and Northeast) a color (yellow, blue, green, purple) and keep that color scheme consistent throughout the piece in all 50 of the charts as well as its breadcrumb-map (discussed in Section 4.3).

A recent and useful technique is to link not just visualizations through color, but to link the narrative text and visualization through color as well. At the simplest level, this consists of sharing color-mappings between text and visualization. This often takes the form of the name of a category in the text being colored the same as the category’s glyph in a visualization. *Bubble to Bust to Recovery* [25] (Figure 3) demonstrates this technique in action. When discussing the “other cities” category, the author has matched that label with the color of the line, more closely tying

Across all groups, @numbered streets are among the most common. The share of streets with #presidents names is highest among the bottom 40 percent and the suffixes street and avenue account for about three-quarters of all homes — more than twice as many as in any other group.



Fig. 7: *Road Map* [3], a story from The New York Times. Here the reader has clicked on the dynamic query widget embedded in the textual narrative to filter the word cloud visualization to just numbered streets.

the narrative to the visualization. The reader is immediately able to discern what item in the visualization the author is referencing in the text.

Three recent stories in our set chose not to link multiple charts through brushing, opting instead to have a single set of glyphs that animate into the different charts that the authors use to tell the story. This **linking elements through animation** has a stronger effect on retaining the reader’s context, as studied by Heer and Robertson [13]. In these examples, the transitions between the charts (or glyph arrangements) are triggered by reaching certain scrolling positions. When the reader scrolls to a certain point, the glyphs animate to their new positions. In *Green Honey* [28], the glyphs are small circles of different colors that are arranged into various color spaces and then grouped by terms used in the names. *Addicted to Trucks* [34] (Figure 2) follows a very similar pattern, where each glyph represents a model of automobile and the glyphs are rearranged in a variety of ways. *342,000 Swings Later, Jeter Calls it a Career* [13] uses a slight variant on this technique. The story contains a single set of glyphs, but the reader begins zoomed into only one of the glyphs. As the reader scrolls, the story semantically zooms out and more and more of the glyphs are revealed.

### 4.3 Enhancing Structure and Navigation

Our penultimate category, enhancing structure and navigation, is closely tied to one of our inclusion criteria—requiring the stories to contain some form of author-specified ordering. A key difference of storytelling from open-ended exploration is that authors impose a structure to a story and frequently provide navigation aids. Even within this notion, however, the variety of techniques authors employ demonstrates the creativity and versatility in how they encourage that structure.

A simple technique, often aligned with a slide show style presentation, is the use of **next/previous buttons**. Here, the buttons and ordering of the slides imposes a strong linearity. For example, *The Case for a New Type of Minimum Wage* [21], *Bubble to Bust to Recovery* [25], *The Yield Curve* [1], and *Egypt in Turmoil* [5] (Figures 4,3,1, and 9) are all instances of slide-shows that use previous/next buttons.

An inherent linear ordering also occurs in vertically tall stories that require multiple pages. The reader experiences the story by scrolling down the page. Visualizations are incorporated into a textual narrative and appear at the appropriate time (e.g., *Obama’s*

*Health Law*<sup>[23]</sup>, *In Gaza, a Pattern of Conflict*<sup>[8]</sup>, or *Beating Stanford*<sup>[4]</sup>). This **scrolling** technique was very common in our set and has inspired its own new genre. The technique, sometimes more explicitly called “scrollytelling” (or “scrolljacking”) has become very popular over the last few years [6]. As the name implies, scrollytelling stories unfold as the reader scrolls; scrolling triggers changes to the visualization itself rather than simply moving the elements up or down on the screen. For example, in *342,000 Swings*<sup>[13]</sup>, while the text scrolls vertically as expected, the visualization continually zooms out as the reader scrolls down.

*Visualizing MBTA Data*<sup>[9]</sup> is an interesting twist on scrollers. Rather than having scrolling trigger any changes like scrolljacking, a Boston subway map visualization moves down to stay on screen as the reader scrolls through a timetable visualization of each train’s movement throughout a day. Because this visualization is linked with the timetable visualization (as we discussed in Section 4.2), the reader may experience a simulated scrolljacking experience: the subway map visualization appears to be changing as the reader is scrolling when the mouse passes over the timetable visualization.

One technique common amongst vertically tall stories that use a single set of glyphs (e.g., *Green Honey*<sup>[28]</sup>, *Addicted to Trucks*<sup>[34]</sup>, and *342,000 Swings Later, Derek Jeter Calls It a Career*<sup>[13]</sup>) are short pieces of text at various points, somewhat equivalent to the text of a slide. As the reader scrolls, these annotations scroll in the standard way, even while the visualizations only change at certain trigger locations. This technique provides the reader with the appropriate feedback that the scrolling is working as intended while allowing the author to tell the story.

Another technique to enhance navigation that we observed across the collection is the use of **breadcrumbs** to communicate where in the story the reader is. In most cases these breadcrumbs also provide the reader to direct access to corresponding locations in the story. Authors used a variety of techniques for including breadcrumbing from the trivial to the complex. A common breadcrumb style uses dots to represent each slide or scene. The current slide is indicated by some form of highlighting, and each dot acts as a shortcut, taking the reader to the associated slide. *Bubble to Bust to Recovery* (Figure 3), *Green Honey*<sup>[28]</sup>, *The Yield Curve*<sup>[1]</sup> (Figure 1) all use dots to communicate the readers’ position.

A number of stories use a series of **section header buttons** to convey structure and assist navigation through stories. Tableau’s Story Points [11] has adopted this technique for the stories that it produces. The author titles each of the scenes, and these titles are placed at the top of the story to help control movement through the narrative. *How Sunspots Control Global Weather*<sup>[17]</sup> and *The Case for a New Type of Minimum Wage*<sup>[21]</sup>, two Story Points examples, exhibit this technique (see Figure 4). Other stories such as *Winning Over Virginia* (Figure 5) use more explicit section buttons to control navigation .

*Where We Came From and Where We Went*<sup>[2]</sup> (Figure 8) faced a unique navigation challenge, namely, having too many sections. To handle all 50 states as separate sections and provide direct navigation, the authors chose to use **menu selection**. While it does not have dedicated navigation buttons, only one section can be selected and each section can be directly accessed. As the reader scrolls through the story, the menu also automatically updates based on the current section (U.S. state). Note that this menu-for-navigation technique provides flexibility for the authors to better tell their story while retaining context for the reader, because the menu itself and story can have different orderings of the story

topics. The menu lists each of the states alphabetically to make it easier for the reader to find a state, yet the authors highlighted three states (California, Nevada, and Florida) by placing them at the top of the story and the remaining states’ sections in alphabetical order after the three.

We also observed other non-linear navigation techniques in our dataset. In *Nelson Mandela’s Extraordinary Life*<sup>[44]</sup> and *Egypt in Turmoil*<sup>[5]</sup> (Figure 9), both made using Timeline JS, the only visualization in the story is the **timeline** at the bottom of the display that acts as the navigational aid for the rest of the story. Each slide is represented in the timeline by either a flag (for points) or a bar (for durations) and as the reader moves through the story the respective glyph is highlighted on the timeline. As with other techniques, clicking the glyph in the timeline jumps the reader to that scene in the story.

The authors of *Where We Came From*<sup>[2]</sup> (Figure 8), discussed just above, did not solely employ a menu selection for choosing states. Under each section heading, the authors also include a **geographic map** with the current state highlighted. This map is interactive, allowing the reader to jump to any of the states using it. In this manner, the map acts as a secondary navigational aid for the story.

Structure and navigation are clearly two essential elements of data-driven storytelling. As such, other researchers have studied this category in depth. For example, from a qualitative analysis of 42 professional, slideshow-style narrative visualizations, Hullman et al. proposed a graph-driven approach that finds “effective” visualization sequences based on an objective function that minimizes the transition costs from the audience perspective [16].

#### 4.4 Providing Controlled Exploration

Unlike pure exploratory visualization, narrative visualization has been characterized by limited or constrained reader interaction with visualizations. While studies have found reasons to doubt the traditional assumption that exploration inherently increases reader engagement [8], the appeal of adding more flexible exploratory

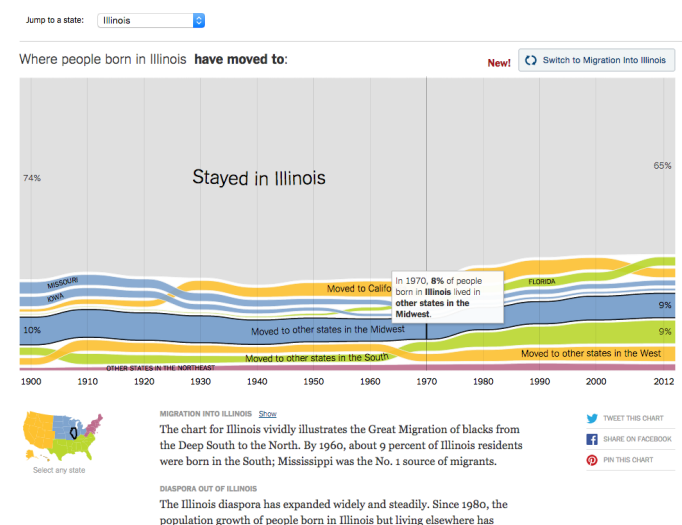


Fig. 8: *Where We Came From and Where We Went*<sup>[2]</sup>, a story from The New York Times. This shows examples of using a combobox and map as breadcrumbs, tooltips and dynamic query widgets for controlled interaction, textual narrative, and linked elements through a consistent color mapping of geographic regions.



## Timeline: Egypt in Turmoil

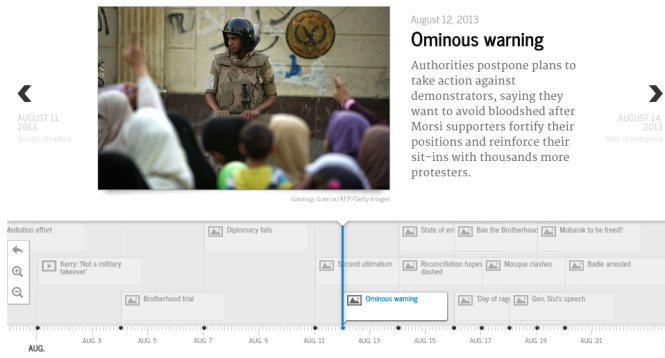


Fig. 9: *Egypt in Turmoil*<sup>[5]</sup>, a story from Aljazeera America using Timeline JS. This utilizes the slide show structure, with slide titles, textual narrative, and non-linear breadcrumbing using a timeline.

capabilities to stories remains. However, enabling full reader exploration in a story has risks. If the reader is able to change a visualization significantly, the data displayed may no longer be consistent with the narrative surrounding it. Furthermore, the reader may find themselves lost in the data, either from data-overload or from having over-filtered. In light of this, data-driven storytelling often relies on providing exploratory elements that are controlled or constrained.

An interaction technique that has become common to constrain a readers’ ability to adjust visualization components is the use of **dynamic queries** [29]. We found dynamic queries in a variety of forms in the dataset. In one section of *Selfiecity*<sup>[30]</sup>, the reader can change which of the four city’s selfies are displayed and how those photos are displayed (“normal,” “cropped,” or “cropped and rotated”). At end of the story, the authors of *Addicted to Trucks*<sup>[34]</sup> (Figure 2) allow the reader to change the color scheme and grouping attribute of one visualization independently, but constrain the choices to only four options: “car/truck,” “origin,” “major brand,” and “gainers/losers.”

The dynamic query portion of *Addicted to Trucks*<sup>[34]</sup> (Figure 2) is also an example of the **embedded exploratory visualizations** technique. This technique is used when the the authors have included a more exploratory visualization, often defined by a large amount of user interaction, within the story. *Gay Rights in the US, State by State*<sup>[22]</sup> also utilizes the technique, including a large exploratory visualization at the top of the story and then discussing the details of the story as it progresses.

A final technique in the controlled exploration category is a **separate exploratory visualization** with which the reader can interact. We debated whether this inclusion of this technique violates our scope of not allowing totally open, reader-driven exploration. This technique is slightly different, however, because the story includes a strong narrative and the authors have chosen to link to an open-ended exploratory visualization. For example, in *Selfiecity*<sup>[30]</sup>, *Oscars 2015: Does winning best director kill your career?*<sup>[18]</sup>, *A Better Way to Find the Best Flights and Avoid the Worst Airports*<sup>[39]</sup>, and *Stop and Frisk All but Gone from New York*<sup>[11]</sup>, the authors provide a link to a separate exploratory visualization to distinctly separate the open exploration from the story itself. This technique can be seen as an evolution of the “martini glass” structure identified by Segel and Heer. Though

some authors do include these more exploratory views at the end of the stories (e.g. *Addicted to Trucks*<sup>[34]</sup> (Figure 2)), several are choosing to exclude the exploratory view from the narrative and treat it as a separate piece entirely rather than as a component of the narrative.

## 5 REFLECTIONS ON THE GROWTH OF THE FIELD

In this section we reflect on the categories and techniques identified above. Listing each technique as we have done may provide an impression that they align one-to-one with story examples. This is far from the truth, however. Below, we first discuss the observation that many stories employ multiple techniques in parallel to achieve their narrative goals. Next, we reflect on how the different categories and techniques we identified relate to the genres, design patterns, and narrative techniques presented in prior research in this area. Finally, we examine the initial set of stories listed in Segel and Heer’s pioneering article [28] and identify the techniques used in each. This analysis provides a view into how the field has evolved since that work was published.

### 5.1 Composing Multiple Techniques

Data-Driven storytelling techniques do not exist in isolation. As one would expect, most stories in our set use a technique from all or many of the categories. For instance, *Gaza, a Pattern of Conflict*<sup>[8]</sup> uses scrolling to enhance structure and navigation, textual narrative to communicate the narrative, and uses color to link its visualizations. Additionally, it was somewhat common in our dataset that stories would use multiple techniques within a category. For example, *Strong Hiring Still Isn’t Bringing Pay Raises*<sup>[14]</sup> includes both traditional long-form textual narrative and brief audio narration. Stories such as *What the Jobs Report Really Means*<sup>[24]</sup> and *Wealth Inequality in America*<sup>[35]</sup> use both color and animation to link components together.

In addition to straight linear orderings, authors have begun to use nesting techniques. For example, *Winning Over Virginia*<sup>[32]</sup> (Figure 5) demonstrates that one slide can have sub-slides such as the counties or regions in that story. In some cases these subplots might consist of a sequence of distinct visualizations, such as the map slide shows embedded within *A Nation Divided*<sup>[10]</sup>. In other cases, the author might reference a series of insights taken from the same visualization, as was the case with *The Yield Curve*<sup>[1]</sup> (Figure 1). In still other cases, the author might add more and more detail to a given visualization to produce narrative. In *Bubble to Bust to Recovery*<sup>[25]</sup>, for example, the author uses a sequence of slides to add one data series at a time, the third of which is depicted in Figure 3.

Authors taking advantage of scrolling navigation have used a variety of means of integrating the navigation technique with other techniques. For example, in *A Nation Divided*<sup>[10]</sup> the authors have embedded slideshows of maps of Germany (colored by various features) within the vertical narrative. Rather than the reader advancing the slideshows using next and previous buttons, the reader advances the slides using the same scrolling navigation as the surrounding narrative text. The reader advances through these map-slides *horizontally* by scrolling *vertically* along the page. This use of “scrolljacking” instead of next and previous buttons strengthens the linearity of the authors’ story. However, there can be some confusion for the reader about the proper way to advance the story through the slideshows. *Selfiecity*<sup>[30]</sup>, *Where We Came From*<sup>[2]</sup>, and *Green Honey*<sup>[28]</sup> each integrate their scrolling navigation with their navigation widgets. As the

Segel & Heer Codes	Our Techniques
Genre	
Magazine Style	(Majority of stories)
Annotated Graph/Map	Text annotation on visualization
Partitioned Poster	
Flow Chart	Flowchart arrows
Comic Strip	
Slide Show	Next/previous buttons, some scrolling
Film/Video/Animation	Narration
Visual Narrative	
Visual Structuring	
Establishing Shot/ Splash Screen	
Consistent Visual Platform	
Progress Bar/Timebar	
Checklist Progress Tracker	Breadcrumbs, Section header, buttons, Menu selection, Timeline, Geographic map
Highlighting	Highlighting elements
Close-Ups	
Feature Distinction	
Character Direction	
Motion	
Audio	
Zooming	
Transition Guidance	Linking through color, Linking through animation
Familiar Objects (with Cuts)	
Viewing Angle	
Viewer (Camera) Motion	
Continuity Editing	
Object Continuity	
Animated Transitions	
Narrative Structure	
Ordering	(Focus on linear stories)
Random Access	
User-Directed Path	
Linear	
Interactivity	
Hover Highlighting / Details	Tooltips
Filtering / Selection / Search	
Navigation Buttons	Next/previous buttons
Very Limited Interactivity	Dynamic query, Embedded exploratory visualization
Explicit Instruction	
Tacit Tutorial	
Stimulating Default Views	
Messaging	
Captions / Headlines	Labeling
Annotations	Text annotation on visualization
Accompanying Article	Separate exploratory visualization
Multi-Messaging	
Comment Repetition	
Introductory Text	Textual narrative
Summary / Synthesis	Textual narrative

TABLE 2: The different types of classifications in Segel and Heer [28] and connections to techniques from our analyses.

user scrolls through the stories, the story’s breadcrumbs, section headers, or menu are updated to be at the reader’s current position. In the case of *Selfiecity* [30], the smooth transitions provided by scrolling are reinforced by the page smoothly scrolling to a section when its section header is clicked.

Almost all of the examples in our collection contain rich combinations of techniques, as is evident in Table 1. This growing complexity of data-driven stories illustrates how the field is maturing and authors are utilizing the different techniques available to them to build even more sophisticated narrative presentations.

## 5.2 Alignment with Segel and Heer’s Genres and Codes

Segel and Heer [28] examined their examples and seeded their codes (classifications) using McCloud’s taxonomy of Visual elements [25] and further analysis they performed. We did not seed our initial code set and our codes were the result only of our analysis. As a result of these differences, some of our codes align quite closely with one of Segel and Heer’s codes, while in other cases the codes do not align all that closely. Table 2 shows the connections between our techniques and their categories.

At the highest level of Segel and Heer’s codes were the Genre, Visual Narrative, and Narrative Structure categories. Most of the alignment between our code sets was among the Visual Narrative and Narrative Structure categories. The codes in these two categories were often examples of what we have been calling techniques. In contrast, the genre codes are not specific techniques in the same way, instead they are often defined by the existence of a set of techniques. For example, a Slide Show is often evidenced

by Next and Previous buttons and a flowchart by Flowchart Arrows. Applying their genre codes to our dataset, we found that many of our stories were Magazine Style stories, with plenty of Narrative Text surrounding the visualization elements. We also had a few Flow Charts, Slide Shows, and Videos.

We would argue, however, that an additional genre has developed since Segel and Heer’s work: the scroller. Leveraging the reader’s scrolling has become a common practice for visualization stories, and a distinct addition to authors’ interaction toolbox. In one sense, scrollers such as *Green Honey* [28], *342,000 Swings Later* [13], and *Addicted to Trucks* [34] (Figure 2) are the hybrid of Segel and Heer’s slide-show and magazine style genres: the reader transitions between scenes through scrolling rather than through navigational buttons. This changes the interaction from discrete next and previous buttons or a collection of dot breadcrumbs to a continuous scrolling interaction. One reason for the adoption may be to make the stories more accessible for phones and tablets, which have seen massive adoption over recent years [32]. Dots are too small to select with touch, removing navigational UI elements such as buttons frees up valuable screen real-estate, and scrolling naturally corresponds to the commonly-used flick gestures on touch-enabled devices. Finally, scrollers more strongly enforce the linearity of their stories, discouraging (or even disabling) the reader to skip through the story and providing greater control to the author on how the story is consumed.

With regard to Segel and Heer’s Visual Narrative category, there was overlap between the two coding schemes, but the two analyses focused on different aspects in more depth. For example, since our set revolved primarily around linear stories, our coding scheme goes into more depth on progress tracking within a linear story. The breadcrumb, section headers, timeline, and geographic map positioning techniques are all grouped together under Segel and Heer’s Checklist Progress Tracker code. On the other hand, since their dataset included a wider selection of genres, Segel and Heer had seven codes for highlighting elements (Close Ups, Feature Distinction, Character Direction, Motion, Audio, and Zooming) whereas we have only the one. Finally, the linking techniques we identified are related to Segel and Heer’s Transition Guidance category. Linking through color is related to Segel and Heer’s Familiar Objects and Object Continuity code and Linking through Animation is related to Segel and Heer’s Camera Motion and Animated Transitions codes.

The codes in Segel and Heer’s Narrative Structure category most closely align with the techniques that we identified. Since our stories were explicitly linear, we did not share Segel and Heer’s Ordering codes. However, we shared a variety of codes in the Interactivity and Messaging categories. Segel and Heer’s Hover Highlighting / Details code aligns with our Tooltip technique. Both of our coding schemes have a Navigation Buttons code. We went into more detail with Controlled Exploration, making it a category rather than a technique. By focusing on fewer open-ended stories, we did not have techniques for Explicit Instruction, Tacit Tutorials, or Stimulating Default Views. Segel and Heer’s Messaging category aligns closely with our Communicating Narrative category. Our Labeling, Annotation, and Textual Narrative techniques map to Segel and Heer’s Captions/Headlines, Annotations, and Intro/Summary Text codes. The only major difference was that Segel and Heer had a code for an Accompanying Article, whereas we found the opposite: data-driven stories would occasionally have a link to a full exploratory version of a visualization, an evolution of Segel and Heer’s martini glass structure.

### 5.3 Evolution of the Field

In order to identify trends in the use of data-driven storytelling techniques, we have analyzed Segel and Heer’s collection of stories [28] using our set of techniques. We were unable to analyze all of the stories in their set. 12 of the stories are no longer available at the provided URLs, and 6 of the stories are behind the Financial Times’s paywall. We were therefore able to analyze 40 of the 58 stories. Of those 40, we would argue that 15 are not stories based on our criteria from Section 3. Our discussion therefore addresses the remaining 25 stories in Segel and Heer’s corpus. Table 1 shows these 25 stories broken down via the techniques we identified earlier in Section 4. An accompanying website<sup>4</sup> provides an interactive version of the table.

The major theme in the the development over the past five years has been the integration of more and more components into single stories. As is telling from the different codes developed, Segel and Heer found that many of their stories linked to a related article. In contrast, our set of stories almost always include visualization elements embedded within textual narratives. There appears to be less of a need to separate these two elements. The addition of multiple elements (whether these are narrative or additional visualizations) tends to require that these be linked in order to retain context. Therefore, we have also seen a similar rise in the use of techniques from the Linking Elements category. It is unclear whether the rise of scrolling as a navigation mechanism is the cause or the effect of the rise in co-located storytelling elements. However, as discussed above, the Scroller genre has certainly developed in tandem with the rise. Similarly, there seems to be a lowered use of Navigation Buttons in favor of Scrolling, again signaling the pattern of keeping elements on the same screen. Finally, it is worth mentioning that many of the techniques in the Communicating Narrative (Labeling, Annotation, Tooltips, and Highlighting Elements) category have all remained popular between both corpora.

### 5.4 Recent Notable Stories

Storytelling with data continues to evolve and develop, much of it outside the visualization research community. Since our initial gathering of the collection of stories, a number of new examples have appeared that make interesting use of the techniques identified earlier. We review a number of stories below that we believe provide creative designs and approaches for communicating narrative stories using data.

Clever uses of scrolling continue to appear, often combined with next/previous controls to assist different styles of navigation through a story. *What’s Really Warming the World*<sup>[3]</sup> (Figure 10) provides a line graph that is persistent through the story, and as the viewer moves through the story, the lines for potential other causes of global warming appear. Each different cause is assigned a unique color as well. Other notable uses of scrolling can be found in *California’s Getting Fracked*<sup>[1]</sup> and *How China’s Economic Slowdown could Weigh on the Rest of the World*<sup>[5]</sup>. (The superscripted, bracketed numbers in this section refer to stories in Appendix B.)

We found multiple new applications of animation in stories as well. The stories *What’s Really Warming the World*<sup>[3]</sup>, *California’s Getting Fracked*<sup>[1]</sup>, and *The Fallen of WW II*<sup>[2]</sup> all employ animation to link elements and enliven a story. *How China’s Economic Slowdown could Weigh on the Rest of the World*

#### Is it All Three of These Things Combined?

If it were, then the response to natural factors should match the observed temperature. Adding the natural factors together just doesn’t add up.

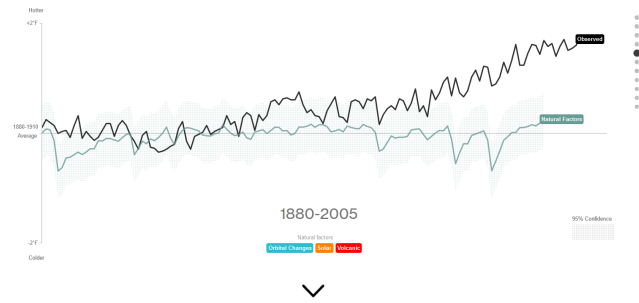


Fig. 10: *What’s Really Warming the World*<sup>[3]</sup>, from Bloomberg. This employs scrolling, animation, linking through color, and controlled exploration through dynamic queries and embedded visualizations.

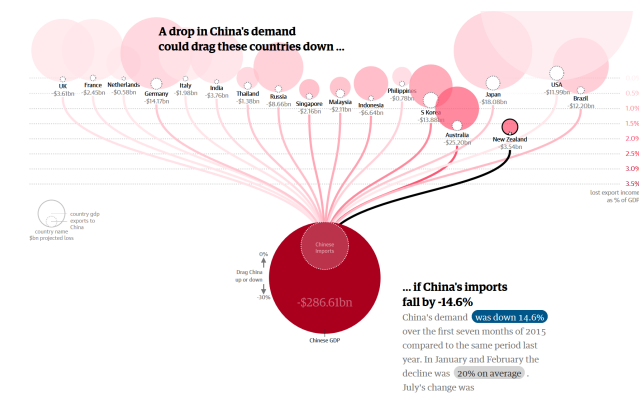


Fig. 11: *How China’s Economic Slowdown could Weigh on the Rest of the World*<sup>[5]</sup>, a story from the Guardian. The story employs multiple techniques including tooltips, labeling, and linking through the color, but especially noteworthy is a clever linking of text items and visualization through animation to represent countries’ falling economies.

*World*<sup>[5]</sup> (Figure 11) cleverly employs a dropping animation to communicate how particular country’s GDP may fall as a reaction to economic problems in China.

The techniques of labeling, tooltips, and element highlighting continue to be prevalent among this new collection of stories we identified, with the majority of stories applying all of those techniques.

Controlled exploration also continues to be a recurring theme in many examples. Dynamic query interactions are provided by *Out of Sight, Out of Mind*<sup>[4]</sup> and *What’s Really Warming the World*<sup>[3]</sup> and embedded exploratory visualizations are present in *California’s Getting Fracked*<sup>[1]</sup> and *The Fallen of WW II*<sup>[2]</sup>.

Two recent example stories that have received considerable attention in social media and in the visualization community uniquely apply one of our identified techniques. *The Fallen of WW II*<sup>[2]</sup> is primarily a video story and uses narration throughout the presentation. The story *Out of Sight, Out of Mind*<sup>[4]</sup> uses multiple techniques, but the story centers around a timeline of notable drone strike incidents in Pakistan.

4. <http://www.cc.gatech.edu/gvu/ii/dds/charts/segel.html>

## 6 OPPORTUNITIES FOR AUTHORING TOOLS

While strong advances have occurred within data-driven storytelling over the past few years, more research needs to be conducted on how to effectively leverage the advantages of traditional narrative-based storytelling and those of interactive visualization together. Kosara and Mackinlay [21] as well as Lee et al. [23] provide a set of broad research opportunities, calling for more research on data-driven storytelling. In this section, we specifically reflect on opportunities to support a wide audience to create data-driven stories.

As we described in the Related Work section, a number of authoring systems have appeared over the past few years such as Infogr.am or Timeline.js. However, many of the newest techniques we have identified in this paper are not supported by existing systems. We present a number of opportunities specifically for designing data-driven storytelling authoring tools.

### 6.1 Controlled Reader Interaction and Experience

While we agree with Lee et al.'s [23] view that the most extreme cases of “reader-driven storytelling” should not be considered storytelling and there is no guarantee that interaction inherently increases reader engagement [8], authors can utilize well-designed interaction to make their readers’ experiences more effective and, simply, more fun. The challenge faced by data-storytelling researchers, however, is how to support authors to find the right balance between full exploration systems and static stories. One possible solution is to make it easier for authors to configure the degree of possible interaction with underlying data. Some of this will come from furthering the research on supporting event-driven interaction (e.g., Ellipsis [27]) and some of this will come from simplifying and incorporating tools for mapping dynamic query widgets to elements of the story.

The story-consuming experience itself can also be reader-specific. For example, some readers might prefer a slide show to a scroller or vice versa; some readers might prefer static graphics with select annotations while others might be more partial to an interactive chart with tooltips. Animating charts can be fun the first time a reader experiences a story, but readers may find them tedious on subsequent readings. In other words, readers may want to choose alternate techniques within a single category than what the author initially chose. How to effectively leverage the reader’s preferences while still remaining true to the author’s story could prove to be an interesting and valuable research direction.

### 6.2 Smart, Dynamic, Data-Driven Annotations

As is clear from the overwhelming number of stories that make use of the technique in both Segel and Heer’s and our corpora, annotation plays a crucial role in asynchronous data storytelling as the surrogate for the storyteller. As pervasive as it is, however, annotation is far from a “solved problem.” There are three properties of any annotation: the content (e.g., text or value), the location, and the time (i.e., when the annotation should appear and disappear). Authors need to be able to control each of these aspects for each of their annotations. As authors provide more and more freedom to their readers in the course of their pieces, there is also the risk that the annotations become inconsistent, irrelevant, or confusing. Should the reader zoom or filter a view such that the data an annotation refers to is no longer displayed, then the annotation may no longer be consistent with the visualization. The converse is also true, authors may want to provide annotations when certain conditions are met, such as zooming deep into a view

or when certain data is selected. The author may want the text of an annotation to change based on reader interaction.

While we chose not to include the story *What Are The Odds of Going Pro*<sup>5</sup> in our dataset due to its lack of an author-specified ordering, it provides a good example of annotations that adapt as the user interacts with the visualization: each of the four annotations in the story’s display adapts as the reader selects different gender-sport pairings. The annotation in Figure 1, from *The Yield Curve*<sup>[1]</sup>, stays in place above the data point it references even as the reader rotates the 3D model. Satyanarayan and Heer [27] provide some initial work into this task, yet how to effectively enable this level of authorial control over annotations will continue to be a promising research challenge going forward.

### 6.3 Navigation through Scrolling

As discussed in the previous section, scrolling has become a pervasive and powerful technique used in data-driven storytelling. Many of the most innovative stories in our collection employ scrolling in creative ways. However, only casual discussions of the merits of using scrolling as a navigational structure have been written [6]. We need to better understand the strengths and weaknesses of using scrolling as an interaction mechanism, especially with visualization. Furthermore, we need to incorporate scrolling into storytelling tools: either through providing authors variety in their linear structure techniques or through tools specifically designed to harness this technique.

### 6.4 Effectiveness-informed Authoring

In our analysis, we witnessed the effect that tools have on the techniques that are used by authors: stories created using Tableau Story Points all had slideshows with section title breadcrumbs; Timeline JS stories were all slideshows using timeline breadcrumbing; the visualizations within stories created with Infogr.am all had tooltips. Storytelling tools could make it easy for people to create compelling data-driven stories as well as less-effective ones. For example, excepting *Winning Over Virginia*<sup>[32]</sup>, the stories that use nested navigational structures did not indicate this explicitly in their breadcrumbs. Instead, these authors opted to treat each timestep as a distinct scene. It would be worthwhile to explore which of these alternatives is more effective, taking into account that there is a spectrum between explicitly presenting the nested structure and hiding it. Furthermore, a wide variety of means can express this structure, some of which authors may not even have implemented yet. Adding animation to these structures, as authors do, further complicates matters. While animations usually help readers retain context [13], specifying animation often increases the complexity of both storytelling and the tools for doing so. To design a storytelling tool that encourages authors towards the most effective techniques, the research community needs to better understand their strengths, weaknesses, and ideal applications, and compare the effectiveness of the techniques within a category in a more controlled setting.

## 7 CONCLUSION

Data-driven storytelling continues to evolve and authors are developing new ways to support narratives with visualization. In this article we investigated the state-of-the-art in data-driven storytelling, analyzing a large collection of data stories to identify the techniques that today’s data-storytellers use. In contrast to

5. <http://www.tableau.com/public/gallery/odds-going-pro>

Segel and Heer’s original analysis [28] that centered on the design space and genres within narrative visualization as a whole, our analysis focused on the more specific techniques used in author-constructed, data-driven storytelling. We identified 20 visual storytelling techniques organized into four broad categories: communicating narrative and explaining data, linking separated story elements, enhancing structure and navigation, and providing controlled exploration. We then reflected on these findings, discussed the evaluation of and trends within the community, and enumerated a number of possibilities for future research.

This work advances the discussion of narrative visualization by describing the state-of-the-art of the medium. Through identifying specific techniques to be incorporated into stories, this work informs the design of future tools that enable storytellers to effectively create data-driven pieces. We as a research community should be making a greater effort to fully leverage visualization as a storytelling medium. Alongside improving our understanding of the techniques and categories that authors have been using, we need to provide storytellers with better tools and support for creating compelling data-driven stories. Finally, we must inspire and encourage authors to continue exploring ever more creative means of telling visual data stories.

## ACKNOWLEDGMENTS

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

## 8 REFERENCES

- [1] Infoactive. <https://infoactive.co>.
- [2] Infogr.am. <https://infogr.am>.
- [3] G. Aisch. Seven Features You’ll Want In Your Next Charting Tool. <http://vis4.net/blog/posts/seven-features-youll-wantin-your-next-charting-tool>.
- [4] 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*, CHI ’15, pages 1459–1468, 2015.
- [5] Apple Corp. Keynote for Mac. <https://www.apple.com/mac/keynote>.
- [6] M. Bostock. How To Scroll. <http://bost.ocks.org/mike/scroll/>, Nov. 2014.
- [7] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data-Driven Documents. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2301–2309, Dec. 2011.
- [8] J. Boy, F. Detienne, and J.-D. Fekete. Storytelling in information visualizations: Does it engage users to explore data? In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI ’15, pages 1449–1458, 2015.
- [9] E. K. Choe, B. Lee, and m. c. schraefel. Characterizing Visualization Insights from Quantified-Selfers Personal Data Presentations. *IEEE Computer Graphics and Applications*, 35(4):27–38, 2015.
- [10] B. Ferster. VisualEyes. <http://www.viseyes.org/>.
- [11] E. Fields. 8.2 Preview: Story Points. <http://www.tableau.com/about/blog/2014/5/82-preview-tell-story-your-data-story-points-30761>.
- [12] J. Fulda, M. Brehmer, and T. Munzner. Timelinecurator: Interactive authoring of visual timelines from unstructured text. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):300–309, Jan 2016.
- [13] J. Heer and G. Robertson. Animated Transitions in Statistical Data Graphics. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1240–1247, Nov. 2007.
- [14] J. Hullman and N. Diakopoulos. Visualization Rhetoric: Framing Effects in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2231–2240, Dec. 2011.
- [15] J. Hullman, N. Diakopoulos, and E. Adar. Contextifier: automatic generation of annotated stock visualizations. In *2013 ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI ’13, pages 2707–2716, April 2013.
- [16] J. Hullman, S. Drucker, N. Riche, B. Lee, D. Fisher, and E. Adar. A Deeper Understanding of Sequence in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2406–2415, Dec. 2013.
- [17] A. Kirk. 10 Significant Visualisation Developments: January to June 2014. <http://www.visualisingdata.com/index.php/2014/08/10-significant-visualisation-developments-january-to-june-2014>.
- [18] A. Kirk. 10 Significant Visualisation Developments: July to December 2014. <http://www.visualisingdata.com/index.php/2014/12/10-significant-visualisation-developments-july-december-2014>.
- [19] Knight Foundation. Timeline JS. <http://timeline.knightlab.com>.
- [20] R. Kosara. The State of Information Visualization, 2015. <https://eagereyes.org/blog/2015/the-state-of-information-visualization-2015>.
- [21] R. Kosara and J. Mackinlay. Storytelling: The Next Step for Visualization. *Computer*, 46(5):44–50, May 2013.
- [22] B. Lee, R. Kazi, and G. Smith. SketchStory: Telling More Engaging Stories with Data through Freeform Sketching. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2416–2425, Dec. 2013.
- [23] B. Lee, N. H. Riche, P. Isenberg, and S. Carpendale. More Than Telling a Story: a Closer Look at the Process of Transforming Data into Visually Shared Stories (In Press). *Computer Graphics and Applications, IEEE*, 2015.
- [24] K.-L. Ma, I. Liao, J. Frazier, H. Hauser, and H.-N. Kostis. Scientific Storytelling Using Visualization. *IEEE Computer Graphics and Applications*, 32(1):12–19, Jan. 2012.
- [25] S. McCloud. *Making comics: storytelling secrets of comics, manga and graphic novels*. Harper, New York, 1st ed edition, 2006.
- [26] Microsoft Corp. Microsoft powerpoint. <http://products.office.com/en-us/powerpoint>.
- [27] A. Satyanarayan and J. Heer. Authoring narrative visualizations with ellipsis. In *Computer Graphics Forum*, volume 33, pages 361–370. Wiley Online Library, 2014.
- [28] E. Segel and J. Heer. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, Dec. 2010.
- [29] B. Shneiderman. Dynamic queries for visual information seeking. *IEEE Software*, 11(6):70–77, Nov. 1994.
- [30] D. Skau. 14 Of 2014’s Best Information Designers and Animators. <http://blog.visual.ly/14-2014s-best-information-designers-animators>.
- [31] N. Yau. The Best Data Visualization Projects of 2014. <http://flowingdata.com/2014/12/19/the-best-data-visualization-projects-of-2014-2>.
- [32] K. Zickuhr and L. Rainie. Tablet and E-reader Ownership. <http://www.pewinternet.org/2014/01/16/tablet-and-e-reader-ownership/>.

## 9 APPENDIX A: STORIES ANALYZED

- [1] G. Aisch and A. Cox. A 3-D View of a Chart That Predicts The Economic Future: The Yield Curve. <http://www.nytimes.com/interactive/2015/03/19/upshot/3d-yield-curve-economic-growth.html>, 2015.
- [2] G. Aisch, R. Gebeloff, and K. Quealy. Where We Came From and Where We Went, State by State. <http://www.nytimes.com/interactive/2014/08/13/upshot/where-people-in-each-state-were-born.html>, 2014.
- [3] G. Aisch and B. Marsh. Road Map. [http://www.nytimes.com/interactive/2015/01/29/sunday-review/road-map-home-values-street-names.html?\\_r=0](http://www.nytimes.com/interactive/2015/01/29/sunday-review/road-map-home-values-street-names.html?_r=0), 2015.
- [4] A. Ambrose. Beating Stanford in Fundraising. [https://infogr.am/beyond\\_crowdfunding](https://infogr.am/beyond_crowdfunding), 2013.
- [5] A. America. Egypt in Turmoil. <http://america.aljazeera.com/articles/timeline-egypt-inturmoil0.html>, 2013.
- [6] G. Arnett. The Numbers Behind the Worldwide Trade in Drones. <http://www.theguardian.com/news/datablog/2015/mar/16/numbers-behind-worldwide-trade-in-drones-uk-israel>, 2015.
- [7] J. Ashkenas and H. Park. The Race Gap in America’s Police Departments. <http://www.nytimes.com/interactive/2014/09/03/us/the-race-gap-in-americas-police-departments.html>, 2014.
- [8] J. Ashkenas, A. Tse, and K. Yourish. In Gaza, a Pattern of Conflict. <http://www.nytimes.com/interactive/2014/07/31/world/middleeast/in-gaza-a-pattern-of-conflict.html>, 2014.
- [9] M. Barry and B. Card. Visualizing MBTA Data. <http://mbtaviz.github.io/>, 2014.
- [10] L. Borgenheimer, P. Blickle, J. Stahnke, S. Venohr, C. Bangel, M. Young, F. Mohr, M. Biedowicz, and A. M. . A Nation Divided. <http://zeit.de/feature/german-unification-a-nation-divided>, 2014.
- [11] M. Bostock and F. Fessenden. Stop and Frisk All but Gone from New York. <http://www.nytimes.com/interactive/2014/09/19/nyregion/stop-and-frisk-is-all-but-gone-from-new-york.html>, 2014.
- [12] B. Burke, S. Carter, J. Daniel, T. Giratikanon, and K. Quealy. 4th Down: When to Go for It and Why. <http://www.nytimes.com/2014/09/05/upshot/4th-down-when-to-go-for-it-and-why.html>, 2014.
- [13] S. Carter, J. Ward, and D. Waldstein. 342,000 Swings Later, Derek Jeter Calls It a Career. <http://www.nytimes.com/interactive/2014/09/14/sports/baseball/jeter-swings.html>, 2014.

- [14] B. Casselman. Strong Hiring Still Isn't Bringing Pay Raises. <http://fivethirtyeight.com/datalab/strong-hiring-still-isnt-bringing-pay-raises/>, 2015.
- [15] B. Casselman. What Wal-Mart's Pay Increase Means for the Economy. <http://fivethirtyeight.com/datalab/what-wal-marts-pay-increase-means-for-the-economy/>, 2015.
- [16] H. Enten. Boston Breaks Snowfall Record with Amazing Finish. <http://fivethirtyeight.com/datalab/boston-breaks-snowfall-record-with-an-amazing-finish/>, 2015.
- [17] M. Francis. How Sunspots Control Global Weather. <http://public.tableausoftware.com/views/SunSpotsStory/SunspotsTheWeather>, 2014.
- [18] W. Franklin, T. Griggs, and G. Arnett. Oscars 2015: Does Winning Best Director Kill Your Career? <http://www.theguardian.com/film/ng-interactive/2015/feb/20/what-it-really-means-to-win-the-oscars-best-director>, 2015.
- [19] A. L. G. The Evolution of Israeli Politics. <http://www.economist.com/blogs/economist-explains/2015/03/economist-explains-11>, 2015.
- [20] Gapminder Foundation. Will Saving Poor Children Lead to Overpopulation? <http://www.gapminder.org/videos/will-saving-poor-children-lead-to-overpopulation/>, 2014.
- [21] P. Gilks. The Case for a New Type of Minimum Wage. [http://public.tableausoftware.com/views/MinimumWage\\_3/MinimumWage-StoryPointsEdition](http://public.tableausoftware.com/views/MinimumWage_3/MinimumWage-StoryPointsEdition), 2014.
- [22] Guardian US Interactive Team. Gay Rights in the US, State by State. <http://www.theguardian.com/world/interactive/2012/may/08/gay-rights-united-states>, 2012.
- [23] E. Klein, M. Yglesias, L. McGann, E. Barkhorn, A. Rockey, L. Williams, K. Keller, J. Posner, M. Bell, Y. Victor, R. Mark, A. Palanzi, T. Whiting, J. S. Maria, J. Fong, T. Whiting, A. Katakam, M. Thielking, A. Espuelas, E. Caswell, D. Stanfield, J. Posner, and J. Posner. Obama: The Vox Conversation. <http://www.vox.com/a/barack-obama-interview-vox-conversation/obama-domestic-policy-transcript>, 2014.
- [24] M. C. Klein. What the Jobs Report Really Means. <http://www.bloomberg.com/dataview/2013-11-08/>, 2013.
- [25] M. C. Klein. Bubble to Bust to Recovery. <http://www.bloomberg.com/dataview/2014-02-25/bubble-to-bust-to-recovery.html>, 2014.
- [26] M. C. Klein. How Americans Die. <http://www.bloomberg.com/dataview/2014-04-17/how-americans-die.html>, 2014.
- [27] Kurzgesagt. The History and Future of Everything – Time. <https://www.youtube.com/watch?v=2XkV6IpV2Y0>, 2013.
- [28] M. Lee. Green Honey. <http://muyueh.com/greenhoney/>, 2013.
- [29] Lemonly. What is Excessive Alcoholism Costing Our Economy? <http://americanaddictioncenters.org/hidden-costs-of-alcoholism/>, 2015.
- [30] L. Manovich, M. Stephaner, M. Yazdani, D. Baur, D. Goddemeyer, A. Tifentale, N. Hochman, and J. Chow. Selfiecity. <http://selfiecity.net/>, 2014.
- [31] D. McCandless, D. Swain, M. McLean, M. Quick, and C. Miles. Byte-Sized. <http://www.bbc.com/future/story/20130621-byte-sized-guide-to-data-storage>, 2013.
- [32] T. Mellnik, K. Park, and R. Johnson. Winning Over Virginia. <http://www.washingtonpost.com/wp-srv/special/local/2013-elections/demographics/>, 2013.
- [33] R. Monroe. Money. <http://xkcd.com/980/>, 2011.
- [34] A. Pearce, B. Migliozzi, and D. Ingold. Scientific Proof that Americans are Completely Addicted to Trucks. <http://www.bloomberg.com/graphics/2015-auto-sales/>, 2015.
- [35] Politzane. Wealth Inequality in America. <https://www.youtube.com/watch?v=QPKKQnijnsM>, 2012.
- [36] K. Quealy and M. Sanger-Katz. Obamacare: Who was Helped Most. <http://www.nytimes.com/interactive/2014/10/29/upshot/obamacare-who-was-helped-most.html>, 2014.
- [37] G. S., P. K., A. C. M., and L. P. Battle Scars. <http://www.economist.com/blogs/graphicdetail/2014/07/daily-chart-21>, 2014.
- [38] A. Sedghi. Zero Hour Contracts in Four Charts. <http://www.theguardian.com/news/datablog/2015/feb/25/zero-hour-contracts-in-four-charts>, 2015.
- [39] N. Silver. A Better Way to Find the Best Flights and Avoid the Worst Airports. <http://fivethirtyeight.com/features/fastest-airlines-fastest-airports/>, 2015.
- [40] N. Smith. Pulp Fiction in Chronological Order. <http://visual.ly/pulp-fiction-chronological-order>, 2012.
- [41] SuperScholar.org. Comprehensive History of Philosophy. <http://superscholar.org/comp-history-philosophy/>, 2015.
- [42] The Data Team. The Agony of Greece. <http://www.economist.com/blogs/graphicdetail/2015/03/daily-chart-0>, 2015.
- [43] The Data Team. The Nuclear Age. <http://www.economist.com/blogs/graphicdetail/2015/03/interactive-daily-chart>, 2015.
- [44] TIME Staff. Nelson Mandela's Extraordinary Life. <http://world.time.com/2013/12/05/nelson-mandelas-extraordinary-life-an-interactive-timeline/>, 2013.
- [45] D. Yanofsky and T. Fernholtz. The World Above Us. <http://qz.com/296941/interactive-graphic-every-active-satellite-orbiting-earth/>, 2014.

## 10 APPENDIX B: RECENT ADDITIONAL STORIES

- [1] A. Flagg, S. Craig, and A. Bruno. California's Getting Fracked. <http://www.facesoffracking.org/data-visualization/>, 2015.
- [2] N. Halloran. The Fallen of WWII. <http://www.fallen.io/ww2/>, 2015.
- [3] E. Rolston and B. Migliozzi. What's Really Warming the World? <http://www.bloomberg.com/graphics/2015-whats-warming-the-world/>, 2015.
- [4] P. I. team. Out of Sight, Out of Mind.
- [5] C. Zapponi, S. Clarke, H. Bengtsson, T. Griggs, and P. Inman. How China's Economic Slowdown Could Weigh on the Rest of the World. <http://www.theguardian.com/world/ng-interactive/2015/aug/26/china-economic-slowdown-world-imports>, 2015.



**Charles D. Stolper** is a Computer Science PhD Candidate in the School of Interactive Computing at the Georgia Institute of Technology. His research interests are in Information Visualization, Graph Visualization, Data Science, and Human-Computer Interaction. His work focuses on identifying aspects of and implementing visualization toolkits.



**Bongshin Lee** is a Senior Researcher at Microsoft Research. She received her PhD in Computer Science from the University of Maryland at College Park in 2006. Her research interests include Human-Computer Interaction, Information Visualization, and User Interfaces and Interaction Techniques. She is currently focusing on developing innovative ways for people to create visualizations, interact with their data, and share data-driven stories visually.



**Nathalie Henry Riche** is a researcher at Microsoft Research since December 2008. Nathalie holds a Ph.D. in Computer Science from the University of Paris Sud & Inria, France and the University of Sydney, Australia. Her research interests span the entire information visualization spectrum: from designing interfaces for scientists to visually explore their data, to designing interfaces for the general public to share data-driven stories.



**John Stasko** is a Professor in the School of Interactive Computing at the Georgia Institute of Technology. He is a widely published author in the areas of information visualization, visual analytics, and human-computer interaction. Stasko is an IEEE Fellow, an ACM Distinguished Scientist, and a member of the ACM CHI Academy. In 2013, he was General Chair of the IEEE VIS Conference.