Authoring Data-Driven Videos with DataClips

Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Andres Monroy-Hernandez, and Pourang Irani

Fig. 1. Example of a data-driven video generated by a financial analysis using DataClips. Snapshot images illustrate the content of the video using 9 different clips: animated text, bar chart and line chart depicting the history of sales and salient events, pictograph and bar chart comparing production and sales, and unit pictograph, bar chart and donut chart illustrating sales during promotion.

Abstract—Data videos, or short data-driven motion graphics, are an increasingly popular medium for storytelling. However, creating data videos is difficult as it involves pulling together a unique combination of skills. We introduce DataClips, an authoring tool aimed at lowering the barriers to crafting data videos. DataClips allows non-experts to assemble data-driven "clips" together to form longer sequences. We constructed the library of data clips by analyzing the composition of over 70 data videos produced by reputable sources such as The New York Times and The Guardian. We demonstrate that DataClips can reproduce over 90% of our data videos corpus. We also report on a qualitative study comparing the authoring process and outcome achieved by (1) non-experts using DataClips, and (2) experts using Adobe Illustrator and After Effects to create data-driven clips. Results indicated that non-experts are able to learn and use DataClips with a short training period. In the span of one hour, they were able to produce more videos than experts using a professional editing tool, and their clips were rated similarly by an independent audience.

Index Terms—data video, narrative visualization, data storytelling, authoring tools, visualization systems

◆

1 INTRODUCTION

The information visualization community has recently focused attention on empowering data analysts and data enthusiasts in communicating insights through data-driven stories [27][36][44]. A wealth of data-driven stories can now be found online, as journalists, working for media outlets such as The New York Times [10] and The Guardian [7], as well as data enthusiasts [6], craft custom narrative visualizations for broad audiences [41]. Short data-driven motion graphics, also known as *data videos*, which combine both visual and auditory stimuli to convey a story, have garnered renewed attention [18]. Endowed with desirable properties, such as having a short duration and engaging visual effects and animations, data videos are a promising medium for conveying a data-driven narrative.

- *Fereshteh Amini and Pourang Irani are with University of Manitoba, Canada. Email: {amini, irani}@cs.umanitoba.ca.*
- *Nathalie Henry Riche, Bongshin Lee, and Andres Monroy-Hernandez are with Microsoft, E-mail: {nath, bongshin, amh}@microsoft.com.*

Yet, crafting data videos is not an easy task. It requires a significant amount of time and effort, a broad set of skills, dedicated software or programming capabilities, and often involves a plethora of tools. For example, the creation process for a data video [13] can span several days, involving people with different backgrounds (such as a data analyst generating the data and insights, a scripter crafting the narrative, along with designers and motion graphics professionals generating the video material), each of which may hinge on one or more specific software tools [28]. The goal of this work is to consolidate the creation of data videos through the major use of one tool, DataClips, and by lowering the skill level required to create data videos using common data visualizations and animations.

Amini et al. [18] recently explored the components and structure of a corpus of data videos, identifying different video sequences and how these sequences play in a narrative (i.e., establisher, initial, peak, and release). In this work, we take a step further and examine elemental video sequences of data videos composed using animated visualizations and infographics. We refer to these elemental units, or building blocks of a data video as *data-driven clips*. Our examination of over 70 professionally crafted data videos reveals the presence of seven major types of data clips across all videos. This led to the development of *DataClips*— a web-based tool allowing data enthusiasts to compose, edit, and assemble data clips to produce a data video without

possessing programming skills. We demonstrate that DataClips covers a wide range of data videos contained in our corpus, albeit limiting the level of customization of the visuals and animations. We also report on a qualitative user study with 12 participants comparing DataClips to Adobe Illustrator/After Effects software, commonly used to create data videos. Non-experts in motion graphics could create a larger number of data videos than those with expertise using the commercial tool, and with no loss in data video quality.

To summarize, our contributions are threefold: (1) the DataClips tool; (2) a library of data-driven clips that can be easily extended with new clips; and (3) a demonstration showing the ability to create data clips in an efficient manner.

2 RELATED WORK

In this section, we provide an overview of prior research around data storytelling and narrative visualization, as well as existing solutions for authoring narrative visualizations and data videos.

2.1 Data Storytelling & Narrative Visualization

The visualization literature is primarily focused on methods for visualizing data to facilitate in depth data analysis and exploration. Ideally, a data analyst can discover new insights from visualizations to communicate or convey a story about the data to its stakeholders. Recently, there has been an upsurge in work transforming data insights into visual stories [37], through various perspectives [27][36][38]. A number of visual analytics systems have also integrated storytelling features in their design (e.g., in-place annotations [25], exporting selected graphical history states [30]). However, supporting an easy creation of rich and diverse stories based on data insights is still an unsolved problem.

We are witnessing a growing research interest in studying storytelling techniques for creating more engaging and compelling data stories or *narrative visualizations*. In a design space analysis of 58 narrative visualizations, Segel and Heer [41] characterized data videos under different genres. Film and video-based data stories (i.e. data videos) are recognized among the seven genres of narrative visualization. Additional work has focused on specific elements of a narrative visualization. Bateman et al. [20] and Borkin et al. [21] specifically focused on understanding infographics and what makes them appealing or memorable to a large audience. By studying an online corpus of narrative visualizations from the area of journalism, Hullman et al. [32] identified categories of rhetorical techniques affecting reader interpretation. Researchers have also looked at methods of sequencing [34] and transitioning [19] elements in a narrative visualization.

Data videos have also been studied from the perspective of film narratives, a medium that bears significant similarity with data videos. Amini et al. [18] examined 50 data videos and teased apart the various dimensions of such a storytelling genre with respect to narratives in film or cinematography. Their results show that data videos exhibit clear narrative structures and use various presentation styles for animating and displaying data. While prior work has shed light on the possibilities of data videos, their structural constituents, and their use for mass appeal, there has been little research on enabling people to create such narrative visualizations.

2.2 Authoring Narrative Visualizations

The widespread adoption of infographics in fields such as data journalism has motivated researchers to investigate ways for making it easy to author narrative visualizations. One approach involves automatically generating explanatory visualizations from data [35][33][26]. This is possible by tailoring a specific data visualization type or dataset. In addition, the storytelling elements of the generated narrative visualizations are limited to annotations overlaid on data visualizations, not allowing for a rich data story.

To support diverse data stories and to lower the barriers for creating narrative visualizations, Satyanarayan and Heer [40] introduced Ellipsis based on a set of abstractions for storytelling with data visualizations. The graphical user interface of Ellipsis allows people to import data visualizations and add storytelling elements to create multiple scenes. However, such a tool does not consolidate the various features necessary to craft a complete data video, based on the concept of a data clips—a ubiquitous element of major data videos.

2.3 Video-based Storytelling

Video-based storytelling is an active research topic and such media are also referred to as "annotated videos" or "multimedia presentations" [39]. Authoring a video-based story involves developing a narrative using a collection of media assets and added annotations. Bulterman and Hardman [24] identified the key authoring problems to address when designing an authoring environment for video-based stories (e.g., the ease of creating unique videos). We argue that similar paradigm can be extended and considered for authoring data videos.

Shen et al. [42] have developed a video-based authoring system that suggests candidates for the "next scene" based on semantic relationships between scenes. Similarly, video story creation tools for non-experts, such as iMovie Trailers [8] and Animoto [3], provide templates that help novices follow a fixed narrative structure and arrange captured content. However, these tools rely on people to decide on their own the appropriate types of elements and how to include them in their stories. To eliminate this burden from authors (especially since we are targeting non-expert authors), we built DataClips based on the concept of predefined story abstractions in our library of data clips, allowing non-experts to rapidly generate a variety of data stories.

3 DATACLIPS

3.1 Motivations

The building blocks of data videos are individual data-driven video sequences, or data clips, each targeting a specific insight of the story conveyed by an animated visualization. Many data videos found online are produced by a dedicated department, e.g., The Guardian visuals [7], or crafted by an independent company [13]. Through an informal interview [28] with the directors of the company who has created several data videos [13] as well as data journalists at the recent Dagstuhl seminar on data-driven storytelling [16], we learned that one minute of data video, excluding data analysis and insights extraction, takes about a week's worth of work from a scripter and an experienced motion graphic designer. Iterating over the video material is costly as each sequence involves several hours of work using Adobe Illustrator [1] for the visual designs and After Effects [2] for the animations. Thus, a significant time is spent upfront on scripting and storyboarding; but iteration is often unavoidable as clients have trouble envisioning the final product without experiencing earlier versions. As design and animations are customized, updating a video with new data also requires a significant amount of time. While these comments may not be representative of the creation of all existing data videos, they give an idea of the overhead and skills required for their creation. We aim at lowering the barriers to authoring data videos, to help a wider audience to use this storytelling medium.

3.2 Usage Scenarios and Target Audience

We closely engaged with two professionals who communicate stories supported by data on a regular basis: Kate, an investigative journalist, and Matt, a finance manager. Both are experts in data analysis but have no expertise with programming or video editing. We met three times with them, gathering their usage scenarios, relevant data, and creating data videos included on our companion website [17].

Rapid video prototyping tool. Kate works for a national news outlet. Her role is to find data and facts, and analyze them to craft news stories on a variety of topics. Kate finds data videos and animated visualizations effective for telling a data story to her TV Channel audience on its website or on social media. She rarely creates them herself, however, because they require a substantial amount of time and resources from a dedicated department in her company. She saw the greatest opportunities for an authoring tool to support her (1) to quickly craft

Fig. 2. Annotated screenshot of DataClips tool interface: a) saved clip sequences, b) clip preview and sequencing panel, c) the clip library panel, d) clip configuration panel, e) import new data, f) clear all clips in preview/sequencing panel, g) category of clips for filling pictographs, h) data configuration options and corresponding input boxes, and i) helper images including numbered items corresponding to the input boxes, j) visual and animation configuration options and corresponding input fields.

short data videos for informal breaking news to be shared on social media sites, and (2) as a prototyping tool to experiment with different narratives and ease communication with her graphics department when producing a high-end data video.

Data video clips authoring tool. Matt's role is to report on financial results and opportunities for a series of products. Matt spends about a fourth of his time compiling presentations to report to executives in the company. Matt found short animated visualizations (illustrating a single insight) the most compelling to "*bring dry and static charts to life*" in presentations and reports. He mentioned that dues to time constraints, he does not create such animations in Microsoft PowerPoint or other tools, especially as they are tedious to update (for each quarter and each product). Figure 1 shows clips created with our tool based on Matt's data and insights to support a story of sales and the evolution of promotional events affecting the sales.

3.3 Design Considerations

Considering motivations and scenarios, we settled on four design considerations (DCs) for an authoring tool for non-programmers and non-video producers. Our premise is that the author has already collected and analyzed data to extract a set of insights for the video.

(DC1) Lower the barrier for authoring data videos. We strive to strike a balance between predefined templates and customizable data-driven videos. Our target audience includes those who have collected and explored their data but are unlikely to have the skills or time to master visual design skills or video editing software. While video templates (e.g., iMovie Trailers) are easiest to create, they are unlikely to cover the wide range of stories people can tell with their data [18]. We propose to rely on a set of templates for short video clips that authors can populate with their data and sequence together.

(DC2) Emphasize pictographs. The use of pictographs or isotypes is heavily present in data videos [8]. Such icon-based data visualizations, reinforces data semantics, may require less interpretation time and increase story retention [20][29]. However, most editing tools [1] today only support manual graphical creation, leading to inaccurate visual encodings. Our goal is to support the creation of accurate animated pictographs by generating them from data.

(DC3) Support data-driven attention cues. Attention cues and strategies are extensively used in data videos to engage viewers and guide their attention during the delivery of a story [18]. For example, it is common to progressively disclose annotations while highlighting related elements within a data visualization. We aim at supporting the creation of data-driven attention cues, enabling authors to import them along with the data, rather than adding them manually on a case-by-case basis. We also propose to include animated transitions between visualizations [31]. Such animated transitions are uncommon in data videos today as they are complicated to craft.

(DC4) One-of-a-kind data video. Our first goal is to provide an authoring tool for novices with a reasonable level of customization: to easily create videos with a different look and feel. For example, rather than enabling users to select the animation timing and behavior of each individual element of a visualization (as PowerPoint does), we chose to enable users to have controlled timing for sets of elements (e.g. axes and bars in a bar chart). Our architecture is modular: it allows advanced users, able to produce code, to easily extend the capabilities of the tool by adding clips.

3.4 User Interface

With the considerations above in mind, we implemented DataClips (http://hci.cs.umanitoba.ca/projects-and-research/details/dataclips). Its interface (Figure 2) is composed of three panels:

1. Clip Library, populated with a set of data-driven clips we describe in detail in section 4;

2. My Clips, a workspace panel where clips are previewed and sequenced to form a longer video, and;

3. Clip Configuration panel, where users can assign data to each individual clip and customize its visual appearance.

We illustrate the main components and features of DataClips through the creation of a short video. Let us imagine Emma, InfoVis paper co-chair this year, who would like to create a short data-driven video to illustrate statistics on the conference attendance and its evolution over the past five years. Emma has gathered the data into a spreadsheet and collected a number of insights to communicate the evolution of the number and gender of authors over the years.

3.4.1　Familiarization and Data Import

Emma connects to DataClips on the web and takes a first look at an example story including five clips already loaded into the tool (Figure 2a). She goes through the story sequence in the workspace (Figure 2b) and selects each clip in turn, which triggers the selected clip's configuration options to appear below the workspace in the configuration panel (Figure 2d). Then, she explores the Clip library panel by placing her mouse pointer on the images in Figure 2c to play the animations.

Emma decides to import her data (Figure 2e). DataClips currently supports a specific data format and does not handle data manipulation within the tool. Thus, she copies and pastes her data table from a spreadsheet software (Table 1) into the data import form in DataClip. The example sequence in her workspace is populated with the new data and she can now review each clip loaded with her data values. Clips requiring specific data columns such as ones figuring maps appear unavailable as her table does not contain geographical data. Emma clears her workspace (Figure 2f) and starts from scratch.

Year	Number of Authors	Number of Males	Number of Females	Notes on Trends
2012	471	306	165	30% Increase in Authorship
2013	303	158	145	US Recession
2014	510	301	209	InfoVis in Paris

Table 1. Fictitious sample dataset for InfoVis attendees.

3.4.2　Data Configuration

Emma's attention moves to the filled stick figure among the pictograph-based clips category (Figure 2g). She wants to see what it would look like with the percentage of female attendees. She drags the clip to the workspace, and upon dropping it into the My Clip panel, the animation plays with an automatic assignment of her data columns. However, the DataClips default assignment does not show the number of females, but the number of males instead. She selects the clip, which brings up its configuration options in the panel below.

Emma notices that the data binding (Figure 2h) is composed of two columns: (1) a column with all her data attribute names separated based on post-processing of the data types (i.e., dimensions or categories vs. measures or values) and (2) a column containing three input boxes which are populated by the system based on the type of imported data columns. To ease the data configuration, a helper image of the visualization conveys the binding between data column and visual encoding (Figure 2i). Emma replaces the column "Number of Males" by the column "Number of Females" in the first box to populate the clip with her target data column and continues by adding an additional clip: a line chart with arrow annotations. Each clip has a different set of input boxes depending on its data requirement. Overall, there are six types of input boxes to bind to data, the ones noted with a * are required for each clip, others are optional depending on the clip:

1. Categories*: column names of categorical data attributes such as date-time or geolocation. Emma places her column "year" in this box for creating her line chart.

2. Selected Categories: a specific value of the attribute selected in above (for filtering purposes). For example, dragging and dropping a subset of years allows her to use only this subset for her line chart instead of all available ones.

3. Values*: column names of numerical data attributes. For example, Emma selects "Number of Authors" for populating the y-axis of the line chart clip she has added to her clip sequence panel.

4. Base Values: are used by clips depicting ratio of values out of a total (e.g., percentage). Emma populated the filled icon with "Number of Females" as values and "Number of Authors" as base values.

5. Drill-down/Roll-up Values: animations involving drill-down and roll-up operations require an additional data attribute to be specified.

For example, after creating her line chart, Emma could select and add a "line to pictograph" clip that will drill-down into a specific year and show the percentage of females. She would then drag the year value into this column.

6. Annotations: column name containing textual annotations associated with specific values. For the line chart Emma is creating, she places the column name "Notes on Trends" in this box. As the line is drawn, the animation pauses and displays the annotation if present.

3.4.3　Visual Configuration

Each clip also has a set of options for the configuration of its visuals (Figure 2j). The visual configuration is composed of two columns: a column with options to select from, and a column with fields and widgets to adjust parameters. Overall, there are five options depending on the clip type:

1. Color and background: allows selection of color palettes and background image for each clip. The selection populates the next clips dropped in the workspace.

2. Icon: allows selection of different icons for pictographs.

3. Axes and orientation: allows displaying axes (present or not) and changing orientation (vertical or horizontal)

4. Title and legend: modifies the visibility, content, and positioning of title and legends displaying attributes' values.

5. Animation style and timing: style refers to the ordering and staging of the animations for the visual elements, which vary for each clip. For example, bars in a bar chart can appear "staggered" or "together." The user also sets the timing of the entire clip.

3.4.4　Clip Sequencing and Export

Emma has now created two clips. However, she would like to first play the line chart showing the evolution in number of authors, and then display the number of females as a filled pictograph for the current year. To rearrange the order of the clips, Emma simply drags the line chart into the first position. Emma then saves her sequence (stored locally in her browser for later edits) and exports the current version as a video file that she saves to her disk (Figure 2e).

3.5　Iterative Design

To investigate the usability of the interface, we performed an hour-long usability session with five users from diverse backgrounds: two storytelling experts with no expertise in video editing and programming, one graphic designer, and two motion graphics editing experts. We asked our participants to reproduce an existing video [14] from a printed storyboard, and observed usability issues as they executed the task with DataClips. We iterated over the interface design as follows:

Interface layout and icons. We rearranged the position and visibility of the three major panels and their content to better match the observed authoring workflow: **selection → sequencing → configuration.** We initially used static icons representing each clip within the library panel and organized them by type of clip as described in section 4.2. They were organized by role in the narrative rather than by visualization types, and used static icons to convey the fact that they apply to any data. However, participants spent a long time finding the clip they had in mind, and icons failed to depict the actual animations. Thus, we designed animated icons that would show the clip format as the user hovers the cursor over the icon.

Assignment of data attributes. Perhaps the most salient change in the interface was the data configuration. Participants were unable to understand the terminology and assign data attributes to input boxes configuring the clip. We iterated over several terms, but none was satisfactory. To solve this, we created helper images including numbered items corresponding to the input boxes (Figure 2i). We also pre-populated each box based on the column types.

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

Fig. 3. Taxonomy of clip types as a function of visualization types (rows) and animation types (columns). Icons and descriptions inside cells show one example implementation for each type.

4 A LIBRARY OF DATA-DRIVEN CLIPS

In this section, we describe our methodology for selecting the visualizations and clips to create DataClips' library.

4.1 Methodology

To compose a library of clips that would enable the creation of a broad range of data videos, we examined a corpus of over 70 data videos available on news media, government and research center websites, visualization blogs, and online video portals such as YouTube.com and Vimeo.com. We used several search keywords such as "data video," "animated infographic," "infographic video," "motion infographic," etc. and processed the top returned results. We kept videos that (1) presented arguments supported by data, and (2) included at least one data visualization.

We proceeded to segment each video from this corpus into clips. We grouped the clips into different categories based on the clip's role in the narrative (e.g., introducing by setting up the scene, explaining a fact in data using annotation, etc.). We then counted the occurrences of each type of clip, excluding the same animated visualization of the same data. For example, the data video in [12] includes an animated bar chart with growing bars three different times throughout the video, but it reuses the same dataset; hence, we counted it a single time.
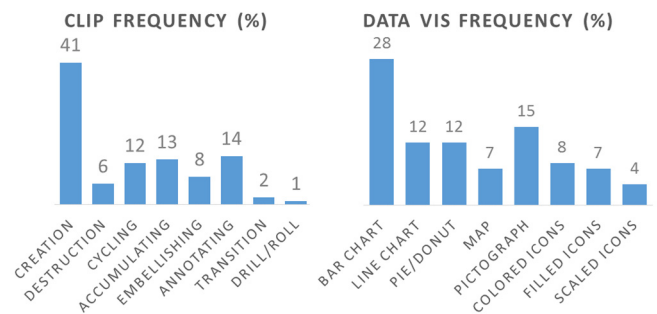
Fig. 4. Frequency of types of clips and data visualizations.

Our analysis led to seven types of clips that we describe below, each applied to eight types of visualizations most commonly found in the videos. Figure 3 gives an overview of the visualization type by type of clip. Note that each clip × visualization combination has different variations. For example, a line chart can be created by drawing both axes and lines together, or one after the other. Figure 4 shows the frequency of clip types and visualization types in our corpus.
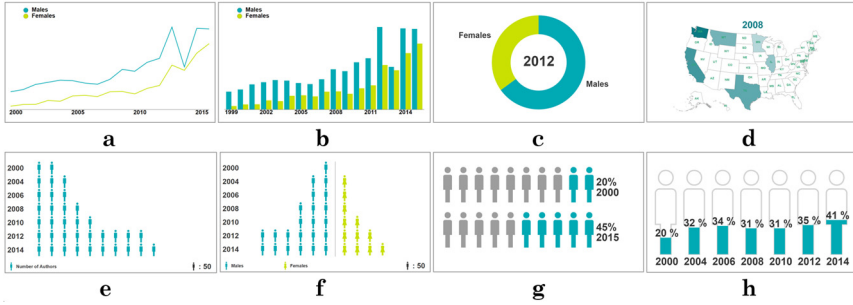
Fig. 5. Screenshot of the eight visualization type supported by DataClips. On the top row are standard charts: (a) Line chart, (b) Bar, (c) Donut, and (d) Map; and on the bottom row are pictograph-based representations: (e) Tally Pictograph, (f) Tally Pictograph-Comparison, (g) Colored Pictographs, and (h) Filled Pictographs.
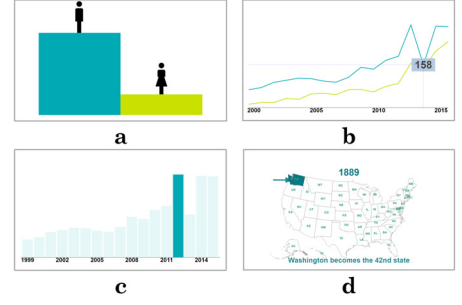
Fig. 6. Annotation clips: (a) male and female icon embellishments above bars, (b) line chart with annotated value and reference lines, (c) bar chart with highlight, and (d) US map with arrow annotation for a given state.

4.2   Visualization Types

DataClips supports eight types of visualization: four standard charts, and four pictographs representations (DC2), most commonly found in data videos.

Standard charts (maps, bar, line, and donut charts). Figure 5a–d shows standard charts in DataClips and the first column in Figure 5 shows the relative percentages found in our corpus. Note that, in line with [8], more than half (59%) of the total data visualizations we observed are standard charts.

Pictograph representations. Figure 5e–h present pictographs or iso-types (International System of TYpographic Picture Education), encoding data using pictorial representations (icons). In the simplest form, a pictograph or pictorial unit bar graph [23] divides the value to encode into equal portions, each represented by one icon (Figure 5e). Different categories can be distinguished by changing the icon shape or color (Figure 5f). A variation of pictographs uses colored icons (Figure 5g) to encode ratios and percentages. The proportion is represented by coloring n icons out of a total of m representing the total value. Note that these representations may result in approximations.

Other iconic representations used to compare numerical values over time or for different attributes are icons scaled based on value or partially filled to encode percentages (Figure 5h). These representations are engaging, as their animation mimics physical objects growing or filling out. However, the effectiveness of these encodings is questionable as our perception is not accurate for estimating areas [43]. We opted to implement the filled icon, most commonly seen in data videos. The icon area filled with color encodes a value n out of a total value m. Increasing (or decreasing) n causes the icon to appear to be filling up (or empting out).

4.3   Clip Types

We briefly describe seven types of clip that we gathered empirically from existing data videos, covering a wide range of animations and attention cues (DC3). Note that a subset of these clip types correspond to animations identified in the taxonomy presented in [31].

1. Creation and destruction: these clips provide animated sequences to create a visualization (e.g., staggered appearance of bars) or to destroy it (e.g., staggered disappearance of bars).

2. Cycling: these clips cycle through years to convey the evolution of values as dynamic changes of the visualization (e.g., iterating over percentages of females as filled icon year after year).

3. Accumulating: these clips gradually add data attributes to the visualization. Most common ones are bar charts starting off with one series (e.g. # of males) and then adding a second one, (e.g. # of females).

4. Transitions: these clips are rarely data-driven in existing videos. Most transitions in these videos are a combination of destroy/create

clips rather than staged transitions, attempting to match elements of both visualizations as described in [31]. We suspect that they are rarely done due to the complexity of realizing them accurately in existing video editing tools. We opted to support these clips in DataClips and extended existing chart transitions [31] to and from pictographs.

5. Drill-down and roll-up: these clips support the transition to a subset of the data visualized in a previous clip. For example, drill-down is used when transitioning from a point in a line chart (for a specific year) to an icon-filled representation of the percentage of female attendees for this specific year.

6. Annotations: animations cover a variety of techniques that guide the viewer's attention to selected portions of the visualizations and reveal specifics annotations. Commonly found examples include highlighting or filtering elements of visualizations, adding a graphical or textual annotation on a specific part of a visualization, or integrating reference lines and numerical values to charts. Data videos also include annotation clips in which icons are overlaid on top of standard data visualizations. Icon annotations are used to distinguish between different types of attributes (Figure 6a) or simply as an embellishment to make the clip more engaging by personifying abstract data visualizations.

7. Multiple views: clips consisting of multiple views appeared less frequently in the data video corpus we examined; these feature several visualizations at once either side-by-side or as an overview+detail setup. Due to their large screen real-estate and the small portions of videos containing them, we decided not to include them.

4.4   Library Coverage

To demonstrate that the principles behind DataClips can lead to a wide range of videos, we reproduced clips from a subset of the corpus of over 70 data videos we analyzed. We demonstrate that Dataclips can recreate about 87% of our corpus, albeit small differences in visual design (e.g., icons placed *beside* growing bars instead of *atop*). We describe our three levels analysis below:

Full coverage (31% of our corpus) refers to the ability to recreate every data-driven clip included in a video with the current library of implemented clips in DataClips.

Minor changes required (56%) refers to reproducibility with minor changes to the implemented clips, or, the same animation is achievable by replacing the original clip with another similar one selected from the current library of implemented clips in DataClips.

Major changes required (13%) refers to the implementation of new clips or new visualizations. An example of such a video is [5], including custom data visualizations.

Figure 7 shows snapshots from two videos we recreated with DataClips. Our companion website demonstrates how DataClips supports the authoring of these videos.
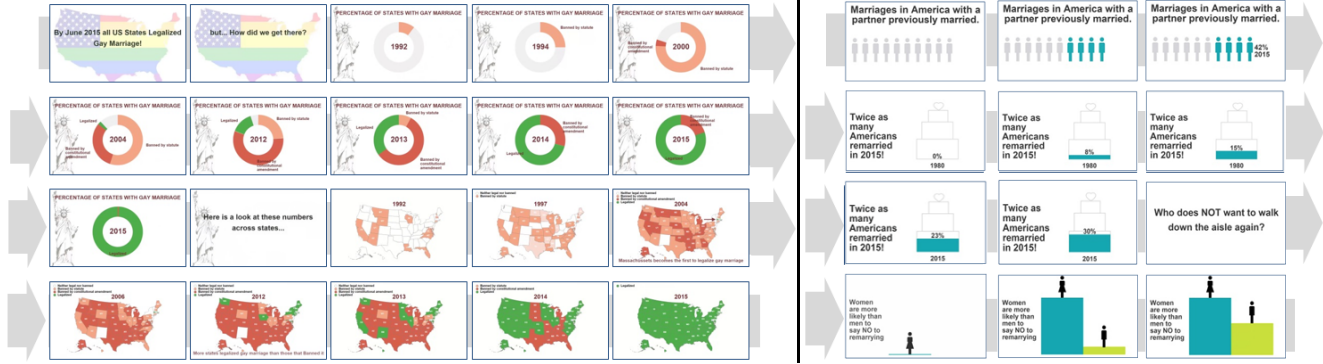
Fig. 7. Videos recreated about same sex marriage by The New York Times [15] (left) and statistics on remarriage by The Guardian [14] (right).

4.5 Implementation

DataClips is a web application using a traditional client-server architecture with HTML, JavaScript, and CSS, as well as d3.js [22] for animated visualizations. The client side follows a model–view–presenter paradigm, using Backbone.js [4]. Each clip implements a model and a view. This structure makes it easy to add a new data clip to the application (DC4) (see website for code examples).

5 EVALUATION

We conducted a user study to evaluate if non-experts could create data videos using DataClips, and gain insights on how the authoring experience and output would compare to videos created with professional tools. Our study was a between-subjects design with two sample groups: one group of participants used DataClips; the other group used Adobe Illustrator and After Effects, commonly used to create data videos. We asked participants to generate data-driven clips based on a list of insights and accompanying dataset we provided. We report our qualitative observations during this process and provide insights on the quality of the videos generated from both groups by asking 40 different volunteers to rate them.

5.1 Participants

We recruited 12 participants (4 males, 8 females; aged 18–35) through advertisements on university bulletin boards and email announcements. The six participants who used DataClips had over 3 years of experience in creating charts using Excel, but had no experience in creating videos. The six participants in the Adobe Illustrator and After Effects group had over two years of experience with this software, and had created videos before. We rewarded participants with $50 at the beginning of the session, independent of their performance.

5.2 Data and Experimental Material

We extracted data and insights from a data video on drug usage published by The Guardian [11]. We selected this video because output and data were publicly available, and it contained 20 facts of general interest that did not require a specific sequence. The dataset contains a set of statistics based on surveying 15,500 drug users via a global drug survey. Instead of asking participants to create an entire video following a narrative supported by a sequence of insights, we asked them instead to create individual clips for a subset of facts they selected, which could potentially be assembled later. We felt this enabled to keep the study to a reasonable duration and avoided the skills of participants in crafting a compelling narrative to interfere too much with the quality of the final outcome. To ensure variety, we selected the subset of ten insights most different from each other, out of 20 presented in the original video. Participants used the material in a spreadsheet document, with each insight and corresponding data in separate sheets. Table 2 shows an example insight and corresponding dataset sample provided to the participants in the study.

Insight: In both US and UK, there are more Cannabis users than Tobacco or Energy Drinks:

Dataset:

Country	% Cannabis	% Tobacco	% Energy Drinks
UK	91	85	79
US	89	76	78

Table 2: Example of one insight and its corresponding data sample

The study was run in the lab, using a computer with a 1600x1900 screen resolution. In the Adobe group, participants were given the option to use their own laptop and other equipment like trackpad, stylus, etc. All used Adobe Illustrator and After Effects CC 2015. Two of them used their own laptop. In both groups, participants had access to the internet in case they required to download images or icons. We also provided them with sketching materials, such as blank sheets of paper, color markers, and pens.

5.3 Procedure

We ran 2-hour long sessions. The experimenter asked participants to think aloud and was present in the room to observe. The experimenter started the session by showing four data videos with a high number of views and including diverse types of data visualizations and animations. We divided the study in two main phases: (1) *idea generation and sketching,* and *(*2) *authoring*.

In the idea generation phase, the experimenter first introduced datasets and insights using a written sheet. Then participants were asked to review each insight and rapidly sketch ideas for an animated visual representation. The experimenter asked participants to generate storyboard-like sketches focusing on: (i) convey the insight to the general public; (ii) select a visual representation that fits the data and insight best; and (iii) think about possible types of animations, and additional text or images required. Participants were encouraged to ask questions if they needed clarification about the data or insights. At the end, the experimenter asked them to describe the storyboards they had created.

In the authoring phase, the experimenter instructed participants to select and implement as many of their sketched ideas as possible in one hour. To motivate them to create high quality videos, we notified them that all of the videos they produced would enter a contest and the author with the highest ratings would win a prize. In the DataClips group, the experimenter first demonstrated the capabilities of the tool using an automated step-by-step tutorial included with the tool, using Intro.js [9]. Participants could also ask questions about the system and the instructor provided additional explanations. Note that this training lasted from 15 to 20 minutes in addition to the one hour authoring phase. The experimenter concluded the session with a semi-structured interview asking participants about their overall authoring experience and the issues they encountered, if any.
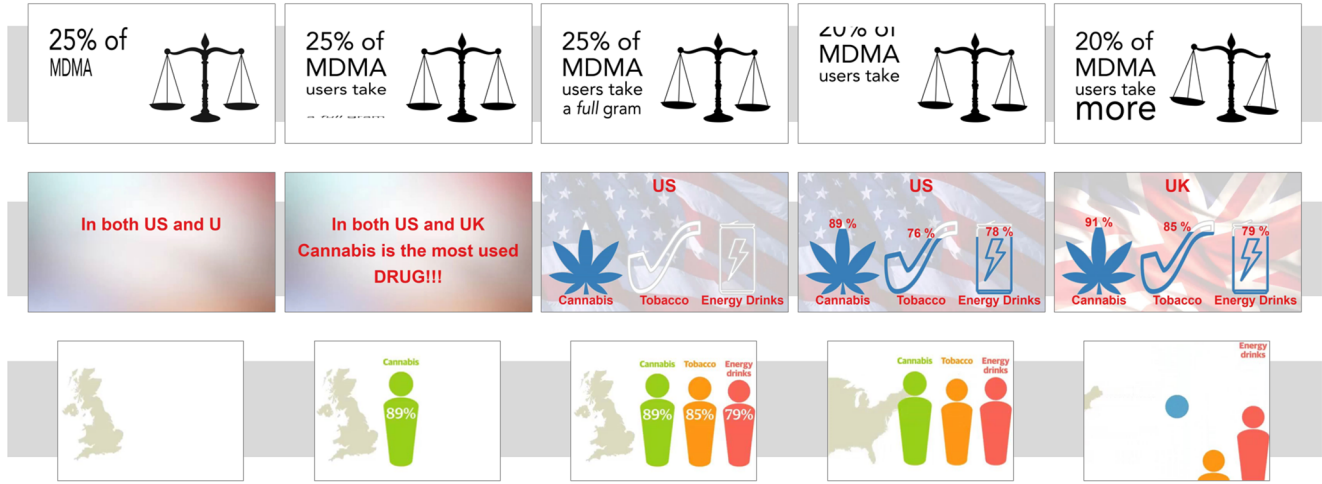
Fig. 8. Examples of clips created by our participants in the DataClips group (top), Adobe software group (middle), and The Guardian example we found online (bottom) based on the same dataset [11].

5.4 Results

We recorded video and audio of sessions (which participants consented to) and took notes during sessions. We analyzed these to extract: (1) selected insights and ideas with rationale, (2) pros and cons vocalized by participants, and (3) the duration required for creating each video clip. We also analyzed the artifacts produced by the participants (storyboards and videos) and asked a group of 40 volunteers to rate their quality. Study material and artifacts are available on our companion website.

5.4.1 Generated Data Videos

We collected a total of 31 videos (each composed of one or more clips, see Figure 8). Table 3 presents a set of interesting differences in the quantity and nature of videos created in both groups.

To gather an independent opinion on their quality, we asked a separate group of 40 volunteers to view each video clip and rate them from 1 (very poor) to 10 (excellent). We randomly presented each of the videos generated, including clips extracted from the original video from The Guardian (but without voice narration). We asked each of the 40 volunteers to rate each video on a printed questionnaire. We asked volunteers to simply provide their "overall impression" of each video. These rankings would inform us of the following: (1) whether videos generated by non-experts using DataClips were of sufficient presentation quality to equal that of videos created by experts using the Adobe suite; and (2) whether there were any trade-offs in using DataClips, e.g., would participants create more videos with our tool, but of lesser perceived quality. We summarize the outcome of these rankings below, instead of presenting them in a separate discussion, to provide a holistic view of the rankings in the context of the results from authoring the clips.

More clips with more variations with DataClips. Participants using DataClips were non-experts in video editing but still generated more clips than did experts using Adobe software. As we expected, the average time to create the first clip was considerably shorter with DataClips. However, we were surprised to observe that clips created with Adobe did not employ custom data visualizations (beyond standard charts) given the freedom offered to the users. Overall, we counted seven types of visualization covered in DataClips whereas only three different ones were used in videos created with the Adobe software.

Data-driven clips. Perhaps one objective measure of the quality of a data-driven video is its accuracy regarding the data it conveys. As data binding comes for free with DataClips, all visualizations occurring in the videos were accurate without additional effort from the user. However, with Adobe software, participants did not always create accurate visual encodings. When asked, participant 2 (PA2) replied by saying

that "*it does not matter if the bar height is not exact.*" Pointing to the resulting video, PA2 also commented that "*this is not supposed to be read by machines but by humans who don't care about the exact numbers.*" While this issue may not prove crucial for professional designers or data analysts with a strong background in visual perception, it is certainly concerning for non-experts.

Similar rating between sources. We find that volunteers' overall impression of videos created with DataClips was equivalent to that of videos generated using the Adobe tools, and even more interestingly, equivalent to video clips extracted from the initial video produced by The Guardian (note however that we removed the voice narration to compare to our other two conditions). Figure 9 illustrates the ratio of average rankings between the three conditions. Through calculating the weighted rank average, we find 36% of the rankings were in favor of DataClips, 35% on average in favor of the Adobe tools, and 29% for the data videos from The Guardian. This outcome indicates that the perceived quality of videos with DataClips matched that of the other two conditions, despite the former being created by non-experts.

Overall, these findings are encouraging, as they indicate that participants who were non-expert in video editing generated more videos with DataClips than experienced participants did with professional Adobe software, without any apparent differences in viewers' rating.

Authoring Tool(s)	Clip Sequences	Average # of Clip Sequences	Average # of Clips	Distinct Insights (out of 10)	Distinct Vis Types	Average Time (first video)	Average Time (all videos)
DataClips	23	3.83	2.5	8	7	9.4	15.8
Adobe Software	8	1.33	1.2	5	3	45.1	37.5

Table. 3. Numbers for distinct items used as well as the average time (in minutes) it took participants to create each video.
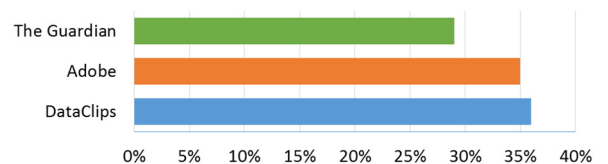

Fig. 9. Ratio of the average viewer rankings for videos produced with DataClips, Adobe Illustrator/After Effects, and clips from the initial video from The Guardian.

5.4.2 Authoring Experience

Our analysis of recordings led to three major insights regarding the use of DataClips to create data-driven videos.

Learnability. We confirmed that all six non-expert participants using DataClips were able to learn and use the main features of DataClips with a short training period. The data configuration was perhaps the most difficult for them to understand. However, we observed that participants could correct their mistakes without intervention from the experimenter and find configuration options for each clip without much difficulty. We also hypothesize that a part of the difficulties participants encountered was due to the lack of familiarity with the dataset, as we observed a lot of back and forth between spreadsheet and list of column names in the DataClips panel.

Rapid ideation. We observed that 11 clips (out of 23) created by participants with DataClips had some changes compared to their initial storyboard. While four of these were relatively minor changes, due to lack of control provided by DataClips regarding the layout or type of animation effects, other changes were more substantial. We observed at least three particular instances in which participants changed their design upon investigation of the clip library. Their comments indicated that they saw a more visually interesting clip or a clip better suited for their data. For example, regarding the use of a standard pie chart, PD2 stated that *"I guess I can show the percentage better using this [the filled icon clip]."* We believe this ability for non-experts to ideate on how to convey data insights is important.

Rapid prototyping. The average number of videos generated by the group with Adobe software was less than 2. All of the videos also only featured standard data visualizations and animations. All participants reported on the lack of time for them to achieve what they wanted. PA1, PA4, and PA5 commented that *"if I had more time, I could have […]"* PA1 remarked that *"if I had more time, I could have made the video more sophisticated."* From their comments, we also noted that most of these participants avoided picking certain insights (e.g., trend data) due to the amount of work it would require them to produce the video. We did not observe any of these issues with the participants using DataClips. These results suggest that even a more experienced video editing audience could benefit from using DataClips as a rapid prototyping tool before polishing videos with professional software suites like Adobe Illustrator/After Effects.

6 DISCUSSION AND LIMITATIONS

Results of our study show that our six non-expert video editing participants could generate more videos with DataClips than expert video editing participants with professional software. In addition, ratings tend to indicate that the perceived quality of videos created with both tools is equivalent. Additional insights on the authoring process appear to confirm that DataClips can lower the barrier to creating data-driven videos and possibly fulfill both of our intended usage scenarios.

However, as with all qualitative studies with *small sample size*, these results should be treated with caution and DataClips warrants further evaluation to confirm if our initial insights apply more generally. In particular, we did not compare DataClips to other software designed for enabling non-experts to create data videos. Note that during a pilot, we had initially included a group in which non-expert participants used Microsoft PowerPoint to create data videos, reasoning that this software might be used by non-experts to create animated data visualizations. However, we decided to discard the group, as our two pilot participants struggled to create even a single data-driven clip.

Furthermore, we opted to use Adobe software due to the *lack of an existing tool* supporting the same capabilities as in DataClips. The purpose of Adobe software is much broader than authoring data videos and proper utilization of all its features requires a larger time window.

Another limitation of our study is the *evaluation of the quality of the videos* generated. Assessing data-driven storytelling media (and data-driven videos) remains an open research question in our community. Our study provides a first step into assessing data video quality but does not delve into all relevant metrics (e.g., engagement, memorability). Thus, we cannot make any assertions regarding the effectiveness of communication and evaluating the quality of data videos remains an open question.

Our observations also pointed to the need for *several iterations of the tool*. For example, participants asked for more control over the layout, which could easily be enabled via direct manipulation. Similarly, it is also possible to add other features like "the ability to include a voice narrations" to add richness to the videos produced. A less straightforward iteration relates to control over the animation effects and the addition of variations of clips. There is a tradeoff between adding several variations of a clip to the library versus providing a generic clip with more configuration settings. In the first case, searching through a large library might become cumbersome and overwhelm first-time users. In the second case, the task of configuring a high number of parameters may also become cumbersome, and labeling them meaningfully for use by non-experts is not trivial. We aim at iterating over the design to strike the right balance.

7 CONCLUSION AND FUTURE WORK

As interest in presenting data moves beyond the confines of data analysts, more general-purpose tools are needed to allow the easy creation of data-driven stories. In this paper, we introduce DataClips, an authoring tool for creating data videos, aimed at non-experts. We developed DataClips based on our close examination of 70 data videos available in the mainstream media and developed by reputed data journalism organizations. From this exploration, we identified the major components necessary for creating compelling data videos. This includes a significant library of data clips, which allow presenting data-driven insights using different visualization styles (including pictographs) as well as different methods for engaging viewers through the use of motion graphics. We report on our design rationale for DataClips, its implementation details, and a qualitative evaluation. From the latter, we find that non-experts can create data videos having the same visual caliber as those created using commercial animation tools (not necessarily for data videos) but also can create more videos in a limited time than tools currently used for creating such videos.

In future work, we envision two main research directions. First, we aim at enabling non-experts to craft compelling narratives [19] for data-driven videos by offering a set of templates geared toward different styles of insights based on the intended message. Second, we will deepen our understanding of what makes compelling data videos. We aim at exploring different evaluation methods and metrics to attempt to better capture the characteristics that make good data videos.

8 REFERENCES

[1] Adobe illustrator cc. http://www.adobe.com/ca/products/illustrator.html/, AdobeIllustrator. [Online; accessed 31-March-2016].

[2] Adobeaftereffects cc. http://www.adobe.com/ca/products/aftereffects.html/, Adobe AfterEffects. [Online; accessed 31-March-2016].

[3] Animoto. http://animoto.com/.

[4] Backbone.js. http://backbonejs.org/, Backbonejs. [Online; accessed 31-March-2016].

[5] The fallen of world war ii. http://www.fallen.io/ww2/, Fallen. [Online; accessed 31-March-2016].

[6] FLOWINGDATA. http://flowingdata.com/.

[7] The Guardian Datablog. http://www.theguardian.com/data/.

[8] iMovie. https://www.apple.com/mac/imovie/.

[9] Intro.js. http://introjs.com/, Introjs. [Online; accessed 31-March-2016].

[10] The New York Times. http://www.nytimes.com/.

[11] Drug use: 20 things you might not know. http://www.theguardian.com/-society/datablog/video/2012/mar/15/drugs-use-animation-20-facts, The Guardian, 2012. [Online; accessed 01-July-2015].

[12] Will taxing the rich fix the deficit? https://youtu.be/FC5Gkox-1QY/, Learn Liberty, 2012. [Online; accessed 01-July-2015].

[13] Children of recession. http://rightcolours.com/portfolio–unicef-inno-centi-report.html, RightColors, 2014. [Online; accessed 26-June-2015].

[14] Four-in-ten couples are saying "I Do," again. https://youtu.be/-EKYOWHtaPQE, Pew Research Center, 2014. [Online; accessed 01-July-2015].

[15] Gay marriage state by state: From a few states to the whole nation. http://www.nytimes.com/interactive/2015/03/04/us/gay-marriage-state-by-state.html?_r=0, The New York Times, 2015. [Online; accessed 26-June-2015].

[16] Data-driven storytelling. https://www.dagstuhl.de/en/program/calendar/-semhp/?semnr=16061/, Dagstuhl Seminar, 2016. [Online; accessed 31-March-2016].

[17] Dataclips project. http://hci.cs.umanitoba.ca/projects-and-research/de-tails/dataclipsj, DataClips, 2016. [Online; accessed 31-March-2016].

[18] F. Amini, N. H. Riche, B. Lee, C. Hurter, and P. Irani. Understanding data videos: Looking at narrative visualization through the cinematography lens. In *Proc. CHI*, pages 1459–1468. ACM Press, 2015.

[19] D. Badawood and J. Wood. A visual language to characterise transitions in narrative visualization. In *Posters Compendium of InfoVis*, 2013.

[20] S. Bateman, R. L. Mandryk, C. Gutwin, A. Genest, D. McDine, and C. Brooks. Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In *Proc. CHI*, pages 2573–2582. ACM Press, 2010.

[21] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister. What makes a visualization memorable? *IEEE Trans. Visualization and Computer Graphics (InfoVis '13)*, 19(12):2306–2315, 2013.

[22] M. Bostock, V. Ogievetsky, and J. Heer. D³: Data-driven documents. *IEEE Trans. Visualization and Computer Graphics*, 17(12):2301–2309, 2011.

[23] W. C. Brinton. *Graphic presentation*. Ripol Classic Publishing House, 1939.

[24] D. C. Bulterman and L. Hardman. Structured multimedia authoring. *ACM Ttran. Multimedia Computing, Communications, and Applications*, 1(1):89–109, 2005.

[25] R. Eccles, T. Kapler, R. Harper, and W. Wright. Stories in geotime. *Information Visualization*, 7(1):3–17, 2008.

[26] T. Gao, J. Hullman, E. Adar, B. Hecht, and N. Diakopoulos. Newsviews: An automated pipeline for creating custom geovisualizations for news. In *Proc. CHI*, volume 23. ACM Press, 2014.

[27] N. Gershon and W. Page. What storytelling can do for information visualization. *Communications of the ACM*, 44(8):31–37, 2001.

[28] K. Gluic. Skype interview. http://rightcolours.com/, 2015, September 22.

[29] S. Haroz, R. Kosara, and S. L. Franconeri. Isotype visualization–working memory, performance, and engagement with pictographs. In *Proc. CHI*, pages 1191–1200. ACM Press, 2015.

[30] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE Trans. Visualization and Computer Graphics (InfoVis '08)*, 14(6):1189–1196, 2008.

[31] J. Heer and G. G. Robertson. Animated transitions in statistical data graphics. *IEEE Trans. Visualization and Computer Graphics*, 13(6):1240–1247, 2007.

[32] J. Hullman and N. Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *IEEE Trans. Visualization and Computer Graphics*, 17(12):2231–2240, 2011.

[33] J. Hullman, N. Diakopoulos, and E. Adar. Contextifier: automatic generation of annotated stock visualizations. In *Proc. CHI*, pages 2707–2716. ACM Press, 2013.

[34] J. Hullman, S. Drucker, N. H. Riche, B. Lee, D. Fisher, and E. Adar. A deeper understanding of sequence in narrative visualization. *IEEE Trans. Visualization and Computer Graphics (InfoVis '13)*, 19(12):2406–2415, 2013.

[35] E. Kandogan. Just-in-time annotation of clusters, outliers, and trends in point-based data visualizations. In *Symp. on Visual Analytics Science and Technology*, pages 73–82. IEEE, 2012.

[36] R. Kosara and J. Mackinlay. Storytelling: The next step for visualization. *IEEE Computer*, 46(5):44–50, 2013.

[37] 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 *IEEE Computer Graphics and Applications, in press*, 2015.

[38] 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, 2012.

[39] B. Meixner, K. Matusik, C. Grill, and H. Kosch. Towards an easy to use authoring tool for interactive non-linear video. *Multimedia Tools and Applications*, 70(2):1251–1276, 2014.

[40] A. Satyanarayan and J. Heer. Authoring narrative visualizations with ellipsis. *Computer Graphics Forum (EuroVis '14')*, 33(3):361–370, 2014.

[41] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Trans. Visualization and Computer Graphics (InfoVis '10)*, 16(6):1139–1148, 2010.

[42] E. Y.-T. Shen, H. Lieberman, and G. Davenport. What's next?: emergent storytelling from video collection. In *Proc. SIGCHI*, pages 809–818. ACM, 2009.

[43] E. R. Tufte and P. Graves-Morris. *The visual display of quantitative information*, volume 2. Graphics press Cheshire, CT, 1983.

[44] W. Wojtkowski and W. G. Wojtkowski. Storytelling: its role in information visualization. In *European Systems Science Congress*, 2002.