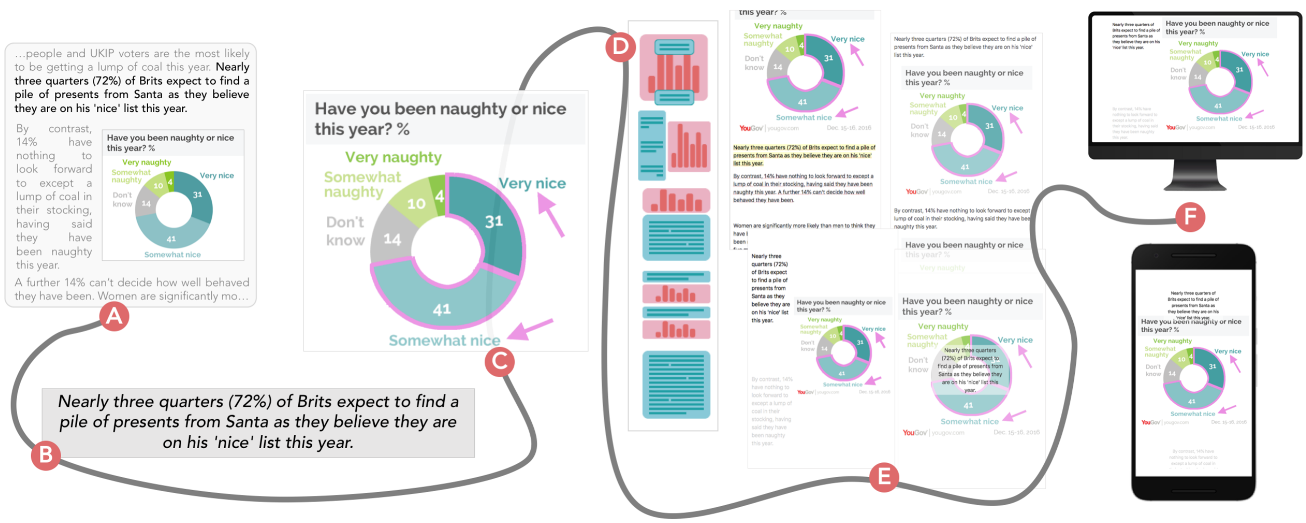
Leveraging Text-Chart Links to Support Authoring of Data-Driven Articles with *VizFlow*

Nicole Sultanum
Fanny Chevalier
University of Toronto
{nicolebs,fanny}@cs.toronto.edu

Zoya Bylinskii
Adobe Research
Cambridge, MA, USA
bylinski@adobe.com

Zhicheng Liu
University of Maryland
College Park, MD, USA
leozcliu@umd.edu



Source article: https://yougov.co.uk/topics/politics/articles-reports/2016/12/21/14-brits-think-they-are-santas-naughty-list

Figure 1: Transforming (A) a static data-driven article (with text and supporting charts) into a dynamic article by leveraging text-chart links. Given (B) a text segment, authors can link it to (C) a related chart, and add related chart highlights (in pink); these links are then used by various (D) storytelling layouts that authors may use to generate (E) dynamic renditions of the text-chart content to be (F) displayed on various platforms.

ABSTRACT

Data-driven articles—i.e., articles featuring text and supporting charts—play a key role in communicating information to the public. New storytelling formats like scrollytelling apply compelling dynamics to these articles to help walk readers through complex insights, but are challenging to craft. In this work, we investigate ways to support authors of data-driven articles using such storytelling forms via a *text-chart linking* strategy. From formative interviews with 6 authors and an assessment of 43 scrollytelling stories, we built *VizFlow*, a prototype system that uses text-chart links to support a range of dynamic layouts. We validate our text-chart linking approach via an authoring study with 12 participants using *VizFlow*, and a reading study with 24 participants comparing versions of the same article with different *VizFlow* intervention levels. Assessments showed our approach enabled a rapid and expressive authoring experience, and informed key design recommendations for future efforts in the space.

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445354

CCS CONCEPTS

• **Human-centered computing** → **Information visualization**; **Visualization systems and tools**; **Empirical studies in visualization**; • **Applied computing** → Publishing.

KEYWORDS

Scrollytelling, Data Stories, Data-Driven Articles, Text-Chart Links

ACM Reference Format:
Nicole Sultanum, Fanny Chevalier, Zoya Bylinskii, and Zhicheng Liu. 2021. Leveraging Text-Chart Links to Support Authoring of Data-Driven Articles with *VizFlow*. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3411764.3445354

1 INTRODUCTION

We live in an era of information overload and fast pace of information exchange based on data. Forming an opinion on complex topics such as climate change, gun control, or the state of a pandemic, often requires consuming a range of data-driven articles from verified news sources to personal blogs. Also known as "textual narratives" [50] and "magazine style narrative visualizations" [41], we define *data-driven articles* as visual data stories [31] that feature primarily textual narratives, containing claims and insights backed by data and illustrated with data visualizations. While traditionally static, these articles have become increasingly interactive over the years as communicators experimented with a variety of storytelling genres and narrative patterns such as infographics [46], interactive slide shows [41], and data videos [2] to help readers understand data [3, 33, 41, 50].

Scrollytelling [42, 43] is one such emerging approach where dynamic updates to the article content are triggered by scrolling; notable examples include the New York Times article called *"Snow Fall: The Avalanche at Tunnel Creek"*[1] (which is attributed as a pioneer of the genre [42]) and the popular educational article *"A visual introduction to machine learning"*[2]. A key design idea in these dynamic data-driven articles is connecting portions of the text body to corroborating chart components using visual cues, such as highlighting and animations. Through these carefully crafted connections, authors create a compelling and enjoyable experience for readers while also guiding them through complex insights and supporting data understanding.

Curating such dynamic experiences requires significant programming skills with current solutions. This demands an interdisciplinary team with technical, design, and communication backgrounds working on a single piece for a considerable period of time, making the process overly complex and time consuming [11]. While prior research has focused on the authoring of visualizations [38, 54], fewer works have looked at data-driven articles as a target authoring medium, and most works examined them from a reader's perspective, e.g., assessing reader engagement across different content presentation styles [33, 57]. To our knowledge, no works have looked at holistically supporting *authors* to create dynamic data-driven articles more easily.

In this work, we seek to lower the barriers to create dynamic data-driven articles without programming expertise. We build on the key notion of text and chart connections, or *text-chart links*, to separate content from presentation (i.e., chart interventions and storytelling layouts) and enable a more modular and extensible approach to authoring of data-driven storytelling. The separation between semantics and presentation has been proposed in a visualization authoring context [37] and the notion of text-chart links has been applied to data-driven articles to some extent [14, 35, 57], but they have yet to be explored in the context of authoring and controlling presentation aspects of data-driven articles.

To help inform this authoring experience, we conducted interviews with 6 authors from major media outlets (§3.1). Their feedback reflected an increasing readership on mobile platforms and a desire for less exploratory and more guided approaches, leading us to shift focus to scrollytelling stories. We followed with a survey of scrollytelling articles, which informed content presentation requirements and common layouts (§3.2). To consolidate insights gathered from these investigations on supporting the authoring of dynamic data-driven articles, we designed a prototype system called *VizFlow* (§4). *VizFlow* makes use of author-defined text-chart links to render content according to a range of layouts (Fig. 1).

To evaluate *VizFlow*, we recruited 12 professional communicators and researchers in relevant fields to create dynamic versions of existing static data-driven articles (§5.1). Authors crafted dynamic articles within a short amount of time, appreciated the ease-of-use of *VizFlow* and the balance of speed and expressiveness enabled by the tool, and commented on the tool's potential for supporting communicators. We also evaluated *VizFlow* from a reader's perspective, inviting 24 participants from the general population to comment on 3 versions of the same article, including 2 *VizFlow* versions with different levels of dynamic behaviour and a static baseline (§5.2). Most participants (75%) preferred the *VizFlow* versions and found them more helpful and appealing.They also provided context on what dynamic features supported their preferences and why. Combined results of these studies helped validate our modular approach to authoring. Beyond serving as a test bed for our text-chart linking approach, evaluations of *VizFlow* also surfaced design lessons and relevant avenues of future work (§6), including spaces for automation, opportunities to support readers, and author guidance towards more balanced storytelling choices.

Our contributions include: (1) design directions for the authoring of dynamic data-driven articles, based on an assessment of scrollytelling articles and author feedback; (2) *VizFlow*, a prototype system and test bed for the use of text-chart links to guide different storytelling layouts; (3) evaluations of *VizFlow* with authors and readers, demonstrating the utility, expressiveness, and effectiveness of our approach, which helped inform (4) a set of lessons learned as additional design recommendations for future authoring systems.

2 RELATED WORK

Terms like "data stories" and "data-driven stories" have been broadly used to refer to journalistic pieces based on data analysis [48], which may or may not include graphical components. In visualization research, Lee et al. [31] define "visual data stories" as including a set of story pieces "backed up by data" where pieces are "visualized to support an intended message" and "presented in a meaningful order". Within that, we define *data-driven articles* as *visual data stories that are primarily guided by text and supported by visualizations*. Terms like "magazine style" [41], and "textual narratives" [50] have also been used for stories featuring "visualizations embedded in narrative text" [53]; in contrast, our scope also entails more analytical data-rich documents [4] (e.g., data analysis reports) that might not have an inherent narrative structure but are still data-driven.

Following, we review past work on authoring and reading of data-driven articles. In particular, we focus on dynamic presentation strategies, and how links between text and graphics have been leveraged in this medium.

[1] https://www.nytimes.com/projects/2012/snow-fall/index.html
[2] http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

2.1 Authoring data-driven articles

Past works [31, 37] outlined a three-stage process for visual storytelling authoring: (1) *exploration* of data sets to inform potential stories; (2) *drafting*, to outline a plot and assess possible ways of communicating the story; and (3) *presentation*, to flesh out content and visual identity. Our work focuses on the two latter components, drafting and presentation, with an emphasis on stories with a substantial presence of text. Related research focusing on these same components includes tools to create visualizations [32, 34, 38] and to author specific storytelling genres such as data sketches [56], data comics [23], and data-driven videos [2], but offer little means to add text for context. Ellipsis [37] and Chen et al.'s story synthesis framework [10] feature text annotations as an integral part of their design, but are tailored towards narrative visualizations and not data-driven articles. Finally, authoring tools like Tableau Stories [51] and GeoTime [14] do support the creation of data-driven articles, but offer little freedom in the presentation aspects.

Most tools to customize presentation of data-driven articles require some scripting knowledge. Idyll [12] is a markup language for web-based data stories, including interactive widgets and a number of interactive scrollytelling layouts for text and visuals. Latif et al. [29] also proposed a markup-based framework to connect text and visualizations, which was applied to graph data and node-link based visualizations [30]. No-programming tools like Adobe Spark [1] and others [15, 45, 47] support content creation with interactive storytelling for the web, but are mostly limited to images and videos and are not designed with data visualization in mind. To our knowledge, our work is the first to offer a no-programming solution to the authoring of dynamic data-driven articles.

2.2 From visual storytelling to scrollytelling

Efforts in visual storytelling have often turned to online articles to help inform the design space of visual data stories [33, 41, 50]. Segel and Heer [41] reviewed 58 articles and proposed 7 genres of narrative visualizations, including magazine-style narrative visualizations, but also a variety of reader-driven genres that allow exploratory freedom. Online articles have since gravitated towards more guided and author-driven storytelling forms. Tse [55] reports how this shift took place in the New York Times as they found the majority of readers did not engage with interactive segments and "just want to scroll", thus leading to recommendations to trigger dynamic changes (e.g., animations and transitions) based on scroll.

This scroll-based storytelling format, aptly called *scrollytelling*, has seen increased interest and uptake in recent years [42, 43]. The format has also been widely applied to data-driven articles, as attested by later assessments of online articles in visual storytelling [33, 42, 50]. An advantage of this medium is its linear nature [16], making it a good fit to narrative [7]. In addition, scrolling is found to align well with reader expectations and how "we interact with computers" [43]. Past assessments on reader perceptions also suggest that scrollytelling is enjoyable and preferable over static presentations [33]. On the other hand, scrollytelling experiences are complex and costly to produce [42], and can be easily misused [26] (e.g., "scrolljacking"). There is an opportunity to bridge the complexity gap while preserving good scrollytelling practices.

2.3 Text-chart links for data-driven articles

While works on authoring of data-driven articles are relatively scarce, more attention has been dedicated to assess the value of explicit connections between text and visuals for reader understanding. A study on automatic text-table linking found that readers perusing documents with explicit cell highlighting significantly outperformed those without highlighting, even when automated highlighting was only 50% accurate [22]. Other studies [27, 53] found that readers with low visual abilities benefited more significantly from adaptive chart highlights in data-driven articles. A study that compared reader experiences with static and dynamic versions of data-driven articles (namely, "stepper" and "scroller" forms) [33] found the dynamic versions were preferred by readers. A later study that compared two content layouts (slideshow and vertical) with and without interactive linking [57] found that linking had a significant positive effect on user engagement. All these works underscore the value of dynamic storytelling forms and visual highlighting of data-driven articles to support reading.

Following, several works leveraged text-chart connections to support chart highlighting in data-driven documents. Many have proposed strategies to extract links between text segments and charts, including a crowdsourcing method to consolidate text-mark references from multiple annotators [25] and automatic techniques based on regular expressions [35] and rule-based matching [22]. Past research also showcased the utility of text-chart links via practical use cases, including facilitated reading of dense data reports with tables [4, 22], an interactive sports story with coupled text and interactive charts side by side [35], and an interactive integration of statistical multiverse analysis within academic publications [13]. This diversity showcases the large application space that can benefit from our proposed approach.

3 AUTHORING PRACTICES

To inform the design of our authoring framework, we conducted interviews with six authors (A1–A6) of data-driven articles and learned about their workflows and preferences. Following their feedback, we collected and analyzed 43 data-driven articles from various online sources, to discover popular presentation formats.

3.1 Formative interviews with authors

To understand current workflows, tools and best practices for scrollytelling, we conducted formative interviews with six authors: three professional communicators from two major news outlets (A1–A3), two research communicators from a non-profit organization (A4-A5), and one independent blogger (A6). We present relevant findings alongside respective takeaways (**T1**-**T6**).

3.1.1 Procedure and Tasks. After a search for published data-driven articles (both static and dynamic), we sent targeted emails to authors to invite them to a 30-60 minute semi-structured interview. Authors were asked to describe their workflow and design decisions when authoring one of their past articles, as well as to comment on general design guidelines and tools for authoring such articles. Participants were compensated with a $20 gift card. Interviews were audio recorded, transcribed, and organized via a thematic analysis. Emerging themes informed the takeaways described below.

3.1.2 Findings on authoring processes. Participants outlined similar authoring workflows to those mapped by past works on visual storytelling processes [11, 31, 37]. They reported starting from an interesting dataset to explore, or a particular question to answer with data. Key story points emerge as they unravel interesting findings, which they then seek to organize in a linear fashion. Authors reported typically beginning a draft allocating "a chart per point" alongside a set of initial charts, and then flesh out the text around those points and visuals: *"I would use the charts to structure the main points I want to make. And then the writing is a way to move between them and contextualize them."* (A1). This calls for the **support of a content skeleton organized around sections and charts (T1)**.

Participants reported their "fleshing out" process often entailed revisiting the data, e.g., to deepen analysis, fill in story gaps, break down a complex insight, and improve presentation, leading to new and refined text, charts and structure. The authoring process may thus benefit from **support of iterative refinement of content (T2)**, allowing for sections to be reorganized and charts updated.

Participants mentioned regularly using various visualization tools, including Tableau (A6), Excel (A5, A6), R (A4), D3 (A1, A2), and Illustrator (A1–A3), showcasing a diversity of practices. Most of them acknowledged having enough freedom to choose what tools to use in their practice, and while they felt comfortable and sufficiently empowered by them, some also claimed reluctance to try and learn different ones. This raises potential challenges due to lack of common standards, prompting us to strive for **wider compatibility within a diverse visualization ecosystem (T3)**.

3.1.3 Findings on storytelling choices. All six participants reported having experience authoring traditional (i.e., static) data-driven articles, while five had also authored dynamic data-driven articles (including interactive visualizations and scrollytelling articles). Despite the appeal of the latter, participants pointed to several challenges for dynamic stories, aptly summarized as: *"The majority of the work is static [because] it's easier to get a static chart to work, it's easier to read, it's easier to get it to load on every single device"* (A3).

The first challenge is technical complexity, including a need for programming knowledge, significant time and personnel resources (as more complex projects tend to be a team effort [11]), and lack of integration with publishing platforms. This entails a need to **lower the technical barriers to integrate dynamic behaviour (T4)**.

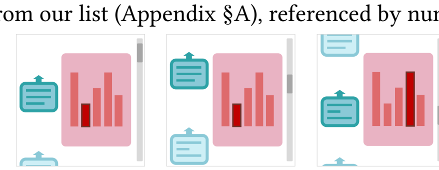
The second challenge is designing for consumption. Authors argued that while interactivity and dynamic behaviour are important to engage readers, it should be *"clear and quick"* (A4) and provided in an *"extremely guided way"* (A1): *"for the most part, people are just scrolling, you know? it's very rare for people to pause and actually click on something, pause and actually hover over something, pause and do anything at all."* (A1). This suggests that dynamic behaviour should **emphasize passive user actions (T5)**, making scrollytelling interventions particularly appealing.

The third challenge is mobile support, an emerging concern in visual storytelling research [16, 19]. Authors stated 40% (A5) to 80% (A2) of their readers consume content from mobile devices, requiring considerations for how their visualizations would look on handheld devices. Content display and dynamic behaviour should therefore be considered with **mobile friendliness (T6)** in mind, e.g., accounting for the possibility of limited screen real estate.

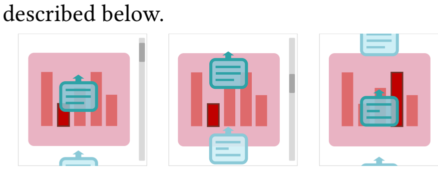
3.2 Assessment of data-driven scrollytelling articles

Following the interviews, we chose to narrow our focus to scrollytelling formats, given their relevance to current authoring practices **(T5)**. Of over 140 articles collected from past visual storytelling assessments [19, 33, 50] and from popular online sources (e.g., The New York Times and The Pudding), we analysed 43 data-driven scrollytelling articles that satisfied our inclusion criteria. This includes 26 articles from past assessments [19, 33, 50] and 17 new additions. Inclusion criteria were: (a) being data-driven (i.e., containing data visualization graphics) and (b) containing at least one scrollytelling segment (i.e., dynamic behaviour mapped to scroll interaction). We catalogued common presentation features to help inform reasonable defaults for dynamic behaviour (§A and Fig. 7 in the Appendix). For a more targeted assessment of highlighting techniques, we chose to focus on classic charts only and excluded articles featuring primarily spatial visualizations (e.g., maps and 3D renderings). We report key scrollytelling features that most impacted our system design — content layouts, chart types, and scroll-based effects (i.e. dynamic highlights).

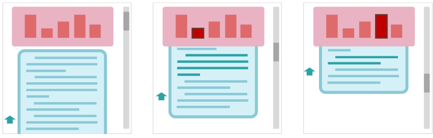
3.2.1 Scrollytelling layouts. This feature refers to the way charts and text are arranged relative to each other and how they leverage scroll dynamics. While largely consistent within an article, these layouts were often alternated with non-dynamic segments (e.g., static paragraphs or blocks of text), or slightly modified to indicate section breaks (e.g., alternating text location from left to right). Mobile support was prevalent, with most articles featuring adapted layouts for mobile aspect ratios. Below, we outline the most common scrollytelling layouts and their adaptations for mobile **(T6)**. Layout prevalence (when applicable) is reported as **D** for desktop screen ratio and **M** for mobile screen ratio. We also include example articles from our list (Appendix §A), referenced by number *(Ex: #)*.



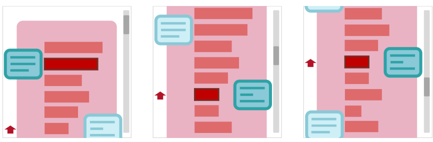
Side by side [**D**: 19 (*44%*) | **M**: 1 (*2%*)]: Chart and texts are displayed side by side on the screen, with text usually on the left. Text is split into smaller segments (typically ranging from a sentence to a paragraph) that slide up and down as the reader scrolls. Charts update to match the current visible text segment. This layout is not well suited to mobile displays and most such articles (16/19) featured an alternate mobile layout: the two most common were *snippets over chart* (6/16) *(Ex: 26, 35)* and *snapshots* (4/16) *(Ex: 23, 24)*, both described below.



Snippets over chart [**D**: 16 (*37%*) | **M**: 16 (*37%*)]: The chart is displayed on the background, with smaller text segments shown as snippet blocks sliding over the chart as the page scrolls. This layout is well adapted to mobile screens and most articles (10/16) retained the same layout between the two screen ratios *(Ex: 13, 37)*.

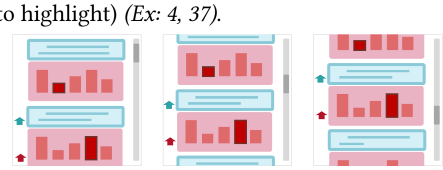


Pinned chart [**D**: 4 (*9%*) | **M**: 2 (*4%*)]: The chart is in a fixed position on the screen as an overlay (often on top), while text scrolls behind it. This layout has the look and feel of a traditional article while maintaining relevant visuals in sight *(Ex: 33, 40)*.



Long chart [**D**: 6 (*14%*) | **M**: 0 (*0%*)]: A long vertical chart that takes several scrolls to navigate, and reveals locally-relevant text content upon scroll. In our collection they were used for interactive lists and timelines while interweaving scrolling navigation with storytelling; they also featured deeply customized charts, all of which required non-standard adaptations to mobile *(Ex: 31, 41)*.

Long form [**D**: 13 (*30%*) | **M**: 15 (*34%*)]: Continuous segments of text that are not scrolling-responsive and resemble traditional articles. In our collection, they were often present after dynamic sections (e.g., *side by side* and *snippets over chart*) and contained the bulk of the article *(Ex: 15, 20)*. Charts in long form segments were usually static, although some featured interactive elements (e.g., click/tap or hover to highlight) *(Ex: 4, 37)*.



Snapshots [**D**: - (-*%*) | **M**: 9 (*21%*)]: A common mobile adaptation layout where each of the multiple text-related highlights for a single chart are displayed as static charts accompanied by its respective text and arranged sequentially. Equivalent desktop screen ratios included *side by side* (4/9) *(Ex: 6,10)*, *snippets over chart* (2/9) *(Ex: 12,14)*, *pinned chart* (2/9) *(Ex: 17, 40)* and *long chart* (1/9) *(Ex: 8)*.

3.2.2 Chart types and scroll effects. We catalogued chart types present in scrollytelling articles and found a similar distribution as reported in past visualization web scraping efforts [6, 36, 40]: *line charts* (14/43, or 32% of articles), *bar charts* (13, 30%), *force-based layouts* (10, 21%), *timelines* (6, 14%), *scatterplots* (5, 12%), and *area charts* (5, 12%). We note, however, the marked presence of force-based layouts which are often paired with dynamic changes and entail technical complexity *(Ex: 1, 6, 10, 26)*.

Finally, we catalogued scroll effects, i.e., interventions on chart and text triggered by scroll transitions. Considerably less scroll effects were used on text, and a majority of articles (23, 53%) did not feature any. We thus focus on chart scroll effects, divided into two groups: *highlighting* effects, which encompass additive and in-place style changes such as revealing (29, 67%) and hiding a chart (13, 30%), emphasising (16, 37%) and de-emphasising (11, 25%) chart elements; and *transformation* effects, which modify chart element shape and data mapping, such as repositioning elements (23, 53%), morphing elements, (e.g., from a circle to a bar) (13, 30%) and rescaling axes (7, 16%). While transformation effects require direct control of chart elements, highlighting effects can be applied as overlays and enable wider compatibility **(T3)**.

4 *VIZFLOW*

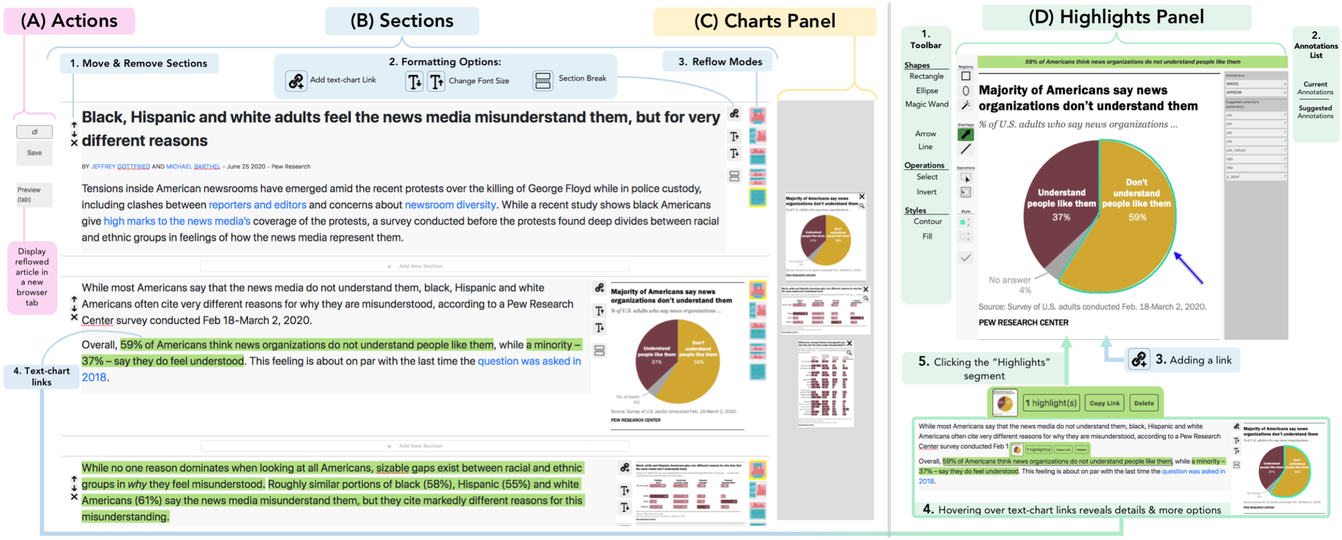
Following our formative assessments, we created *VizFlow*, a proof of concept tool to help authors create dynamic data-driven articles without coding. It was designed to resemble a regular text editor with added storytelling features, instrumented by the notions of *reflow modes*, *sections*, and *text-chart links*. Our assessment of scrollytelling stories informed 6 storytelling layouts (§3.2.1), 5 of which are implemented by *VizFlow* : *snippets over chart*, *side by side*, *pinned chart*, *long form* and *snapshots* (Fig. 4). In *VizFlow* we call them *reflow modes*, and they represent the different layouts authors can apply to *sections*. These, in turn, correspond to different partitions of the story and define a scope for unique reflow modes to be applied (Fig. 4). Reflow modes can be changed with a single click, allowing authors to freely experiment with different storytelling formats. Such seamless layout changes required a semantic representation to separate the underlying message to be conveyed from its presentation, which we achieved via author-defined *text-chart links*: they encode meaningful semantic connections between text segments and individual charts, and each reflow mode makes use of them in a different way.

VizFlow was also intended as a test bed to evaluate the text-chart linking approach via interactive user tasks. To this end, we implemented a minimal set of features to support end-to-end authoring with text-chart links, striving to provide enough functionality to reduce the burden of peripheral tasks such as chart highlighting.

In the following segments, we present the features and design rationale of *VizFlow* and how it may be used within an authoring workflow. We refer to the takeaways **(T1-T6)** presented earlier (§3.1) to further justify design decisions. We also briefly discuss implementation details.

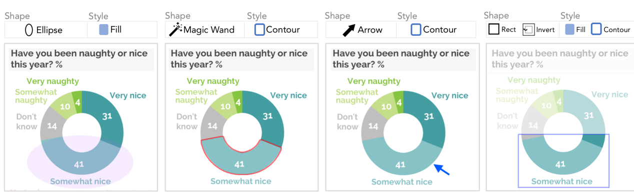
4.1 Adding content and structure

After opening *VizFlow* (Fig. 2), authors are greeted with a ready-to-edit blank section, where they can start recording key insights and corresponding charts generated as part of their data analysis. Authors may start typing an outline of points to develop later, and possibly assign each point to a new section. New sections can be added via the `Add new section` buttons shown in-between sections, or by splitting an existing section in two using the `Section Break` option (Fig. 2(B)) or its respective keyboard shortcut. In this way, *VizFlow* **supports creating content skeletons organized around sections and charts (T1)**. Standard text editing options — bold, italics, and font size — are available via keyboard shortcuts and follow author experiences with common text editors.



Source article: https://www.pewresearch.org/fact-tank/2020/06/25/black-hispanic-and-white-adults-feel-the-news-media-misunderstand-them-but-for-very-different-reasons/

Figure 2: Authoring view of *VizFlow* features a (B) section-based text editor with (B.1) options to rearrange sections, (B.2) format content and (B.3) choose reflow modes. Text-chart links are highlighted green (B.4), with link details visible on hover (D.4). The (D) highlights panel features (D.1) tools to create and style standard annotations. Created annotations and suggestions are listed on the right (D.2).



Source: https://yougov.co.uk/topics/politics/articles-reports/2016/12/21/14-brits-think-they-are-santas-naughty-list

Figure 3: Examples of combinations of *VizFlow* annotation shapes and styles applied to the same pie chart segment.

As writing unfolds, authors can **iteratively refine content (T2)** by freely rearranging ↑↓ and removing ✕ sections (Fig. 2(B.1)).

Accompanying charts can be added by directly copy-pasting images or their respective URLs into the editor, which are then displayed on the *charts panel* to the right (Fig. 2(C)). We designed *VizFlow* to accept charts as raster images, a compatibility compromise to **support wider compatibility within a diverse visualization ecosystem (T3).** This facilitates the addition of any chart, regardless of the visualization toolset used to create it, which in turn allows users to create dynamic versions of most pre-existing static articles available online.

4.2 Linking text and charts

After a few paragraphs of text have been fleshed out and charts have been uploaded into the editor, authors can link text and charts via the `Add link` option on a selected text snippet (or via its associated keyboard shortcut). After selecting the corresponding chart from a gallery list, the *highlights panel* (Fig. 2(D)) appears, displaying the selected chart (and the corresponding text snippet on top) along with options to annotate the chart. Annotation options were guided by our assessment of chart types (§3.2.2) and suitability to raster images (i.e., overlays). They include a small but expressive set of annotations that can be customized in shape (`rectangle`, `ellipse`, `arrow`, `line` and `magic wand`, i.e., colour-based region selection) and style (`fill` and `contours`) (Fig. 3). Annotation regions also allow for `inverse selections` which can be used for mask-like effects. A few possible compositions are illustrated in Fig. 3. To further streamline user annotations, *VizFlow* offers automated annotation suggestions of common chart elements (Fig. 2(D.2)), including data marks such as bars and pie segments, and chart labels such as title, legend and axis labels. These suggestions are offered as a separate annotations list under active annotations, and can be easily previewed via hover, added to the list of active annotations with a click, or ignored when not appropriate.

After adding their desired annotations, authors can confirm their selections ✓ and are returned to the authoring view (Fig. 2(B)). The linked chart now appears to the right of the section and the linked text segment is highlighted in green (e.g., bottom section in Fig. 2(B)). Hovering over linked segments allows authors to get both a preview of the annotations (shown directly over the section chart) as well as a context menu with editing options for the text-chart link (Fig. 2(D.4)). From there, authors can edit link details, e.g., re-access the highlights panel to edit its corresponding annotations (Fig. 2(D.5)) or re-link the segment to a different chart. They can also `Copy Link` a segment, replicating text-link properties (chart and annotations) to another segment of text. With certain reflow

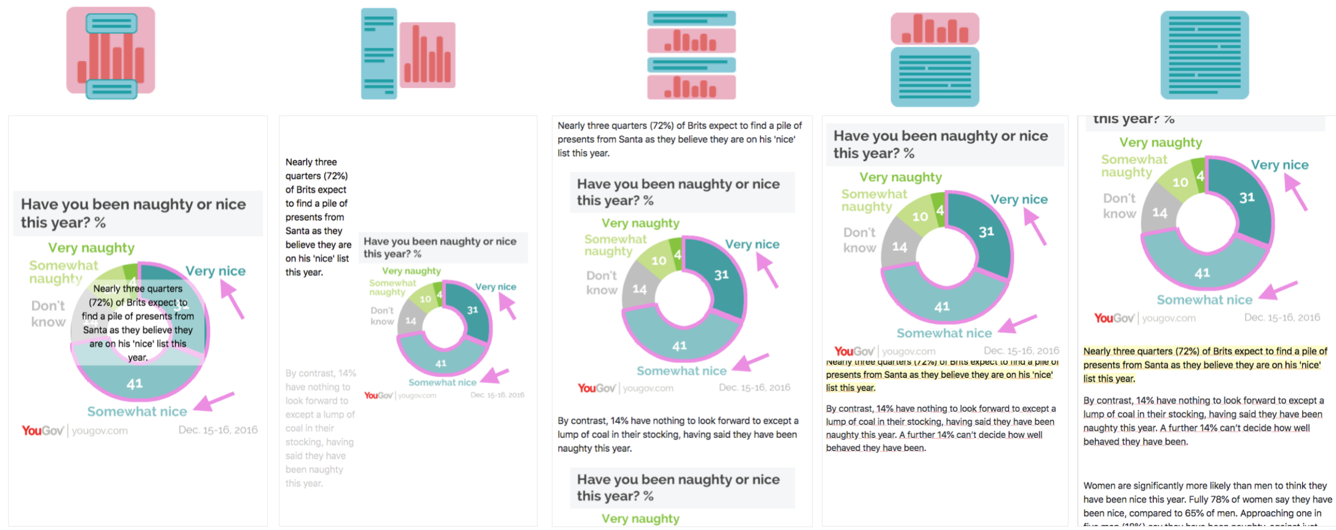


Figure 4: Reflow modes implemented by *VizFlow*, based on scrollytelling layouts. From left to right: *snippets over chart*, *side by side*, *snapshots*, *pinned chart* and *long form*. For *pinned chart* and *long form* modes, chart highlights are activated on text hover (yellow text highlights).

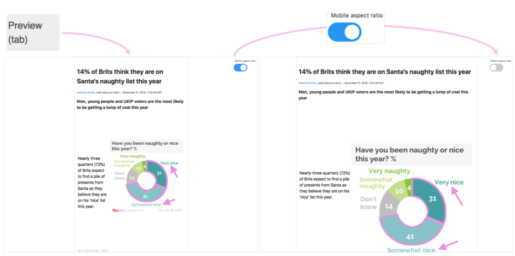


Figure 5: Preview tab of reflowed article, accessible from the "Preview (tab)" button in the authoring view (Fig. 2(A)). It features an aspect ratio switch to emulate how content may be viewed on desktop or mobile.

modes, this provides the ability to create highlighting sequences akin to animations when text segments are linked back-to-back.

4.3 Reflowing data-driven articles

Inspired by our formative interviews with authors (§3.1), we designed *VizFlow* to **emphasize passive user actions (T5)** by invoking scrollytelling interventions. Based on our corresponding analysis of scrollytelling layouts (§3.2.1), we implemented 5 of the 6 most common layouts discovered (with the exception of *long chart* due to its custom nature) including dynamic layouts and layouts that support **mobile-friendliness (T6)**. In *VizFlow*, they are called "reflow" modes due to the way they leverage text-chart links to redefine content flow. For snippets over chart and side by side, text content is split into sentence-level snippets positioned at the center or on the left; text is continuously scrolled, featuring one snippet at a time, and activates associated annotations on the chart when a sentence encompasses a text-chart link. For pinned chart and long form, text is rendered as paragraph blocks, with linked text shown as underlined segments that reveal chart annotations on hover. For snapshots, text paragraphs are split based on the span of text-charts links, with static annotated versions of the linked chart displayed immediately below each.

For each section of an article, authors can easily assign one of the 5 reflow modes and freely **iterate between these options (T2)**. *VizFlow*'s **section-based approach (T1)** allows distinct reflow modes to be associated with different sections of the article and increases flexibility of dynamic behaviors. Authors can see complete reflowed articles via a separate `Preview` window, accessible from the authoring view (Fig. 2(A)) as a new browser tab that automatically synchronizes with content changes (Fig. 5). This allows authors to easily **iterate over design alternatives (T2)** by experimenting with dynamic reflow configurations, thus **lowering the technical barriers to integrating dynamic behaviour (T4)**. Having writing, editing, and design (*WED*) tightly interwoven is noted to make for good journalism [16]. The preview window further supports **mobile friendliness (T6)** by letting authors preview content on a mobile or desktop aspect ratio via the `mobile` switch (Fig. 5). A preview of a complete reflowed article is shown in Fig. 8 (Appendix).

4.4 Implementation details

VizFlow is implemented as a JavaScript React and SlateJS application on a Python Flask server. Synchronization between the authoring view and preview tab is supported by the Broadcast Channel API, allowing different browser sessions to share data; a data share broadcast is triggered after every update in the authoring view, resulting in seamless content updates on the preview window. Authored *VizFlow* articles can also be accessed directly via a standalone preview window, serving as a general reading platform.

The suggestion module leverages LEAF-Net's chart parsing module [9], a Mask-RCNN model [18] trained to extract bounding boxes for data and labels marks from chart images. This model was trained on and is able to recognize marks from: pie charts, bar charts, boxplots, line graphs and scatter plots [9]. To reduce clutter and ease navigation, small hard-to-select marks were filtered out (e.g., tick labels) and remaining marks were sorted by position (left to right, top to bottom). Given the non-interactive processing times (60-90 seconds), charts were pre-processed when possible and their extractions stored for later reuse.

5 EVALUATION

To assess our approach to leverage text-chart links for creating reflowed articles, we conducted two validation studies. The first study looked at author impressions on our concept and consisted of a task-based interactive session to reflow a traditional data-driven article using *VizFlow*. The second study assessed reader perceptions on different versions of reflowed articles via an online questionnaire. We detail these two efforts in the following subsections.

5.1 Author perspectives on *VizFlow*

We conducted a qualitative assessment of *VizFlow* through an interactive task. We recruited 12 individuals with diverse backgrounds to take part in the study. Participants included: 5 communicators (C1–C5), i.e., professionals with newsroom experience, 3 of whom took part in our formative interview study (A1–A3); and 7 researchers (S1–S7), i.e., experts in related fields such as as data visualization, storytelling, graphics design experts and data science. Each received a $30 gift card in their local currency for participation.

5.1.1 Procedure and task. Sessions encompassed (a) a walkthrough of *VizFlow*, (b) an interactive reflow task, where participants used *VizFlow* to create a reflowed version of a traditional (i.e., static) data-driven article, followed by (c) a System Usability Scale (SUS) questionnaire [8], and (d) a semi-structured interview to reflect on their experiences with the tool. For the reflow task, each participant was assigned 1 of 4 publicly available static data-driven articles (Fig. 6). Articles were pre-loaded into *VizFlow*, with text content provided as a single section and supporting charts in the charts panel (Fig. 2(C)), but no text-chart links. Participants were asked to link to all the charts in the article, and to use reflow modes however they found appropriate. Interview questions covered impressions of the tool and the concept, their design decisions when creating the dynamic article, their opinions on reflow modes, and thoughts about the future of this concept as informed by their practices.

Due to time constraints with their professional engagements, sessions with the communicator group (C1-C5) were synchronous and tailored for a 1-hour duration. Participants in the researcher group (S1-S7), on the other hand, were given the opportunity to complete the exercise on their own, after watching a 7min video walkthrough of *VizFlow* and completing the reflow task at their own pace. In the end, both groups completed the same task, and interviews covered the same topics.

The articles used in the study were selected for content diversity, featuring a variety of writing styles and chart types (e.g., bar, pie, line, bubble and interval) (Fig. 6). To narrow our search space and expedite sessions for our professional participants, we first looked for published articles they authored that fulfilled the above criteria. We ended up with 3 of our 4 selected articles being authored by, and assigned to C1-C3. All other participants were unfamiliar with their assigned articles.

Interviews were recorded, transcribed, and organized via thematic analysis. We also analyzed and compared reflowed articles in terms of the authors' structural and creative choices, and computed overall usability assessment from the SUS questionnaire. In the following segments, we report findings from our analysis, contextualized by frequency of mentions of qualitative feedback in a (#/12) format. Frequency counts include only explicit mentions and should thus be seen as a lower bound. Links to author-reflowed articles are provided in the Supplemental materials.

5.1.2 Limitations. We considered alternative study setups and their pros and cons. We opted against a comparative study given the lack of a fair baseline: direct manipulation tools like Adobe Spark [1] rely on templates and do not offer the same level of control of our text-chart linking approach, while programming-based approaches like Idyll [12] would entail a steep learning curve for some users. We also provided pre-drafted articles instead of asking participants to write articles from scratch. While the latter would be more ecologically valid, we would have extra confounds in comparing author preferences and workflows, while adding significant complexity to the task that would be peripheral to text-chart linking.

5.1.3 Overall impressions. **Opinions on *VizFlow* were generally positive**, corroborated by a SUS score of 77 (i.e., comfortably above the average SUS score, 68). Participants stated they found it easy, intuitive, and straightforward to learn and use (7/12). The notion of content sectioning was familiar to many, akin to content management systems (CMS) used in newsrooms and blog platforms (C3) and Jupyter notebooks (S3, S4).

Participants found *VizFlow* useful and timely (5/12). Communicators acknowledged its value for **lowering barriers to create "scrollers"**, which are *"annoying to do manually"* (C1), that it can be **particularly useful to novices** (C3, C5) such as *"editorial reporters [because] you don't need a developer"* (C3) as well as a **quick prototyping tool for data stories**: *"there's always a question of what would this look like as a scrollyteller (...) And this lets you answer that question very quickly"* (C1). A researcher in visual storytelling liked that *VizFlow* was *"focused on really storytelling"* (S7) instead of just visualization authoring, and that it was *"very needed"* (S7).

As for the limitations, some participants **missed more fine grained control over the presentation**, e.g., layout customizations (S4) and better control of pacing with spacing between sections (S2), to a lack of more exploratory modes of interaction, e.g., embedded annotations and callouts when hovering over chart elements (S3). More importantly, a few raised **limitations about the linking approach itself**, on *"being forced into thinking in terms of each sentence needs to be its own annotation, or each annotation needs to be a complete sentence"* and that it would be important to *"think about ways to break out of that model"* (C1). Also, while the linking and reflow approach may confer simplicity to the process, it **may constrain creativity**: *"the paradigm is still, I have some text, I have some charts and I'm going to try to link parts of the chart to parts of the text (...) rather than "let's start with a graphic and figure out how we want to ride around it", if there's like a more nonlinear*

ID	Title	Source	# Words	# Charts
Art-A	What Happened to America's Political Center of Gravity?	NYT [link]	731	5
Art-B	U.S. Lags in Coronavirus Testing After Slow Response to Outbreak	NYT [link]	505	3
Art-C	Election 2019: Britain's most diverse Parliament	BBC [link]	463	6
Art-D	Black, Hispanic and white adults feel the news media misunderstand them, but for very different reasons	Pew [link]	738	3

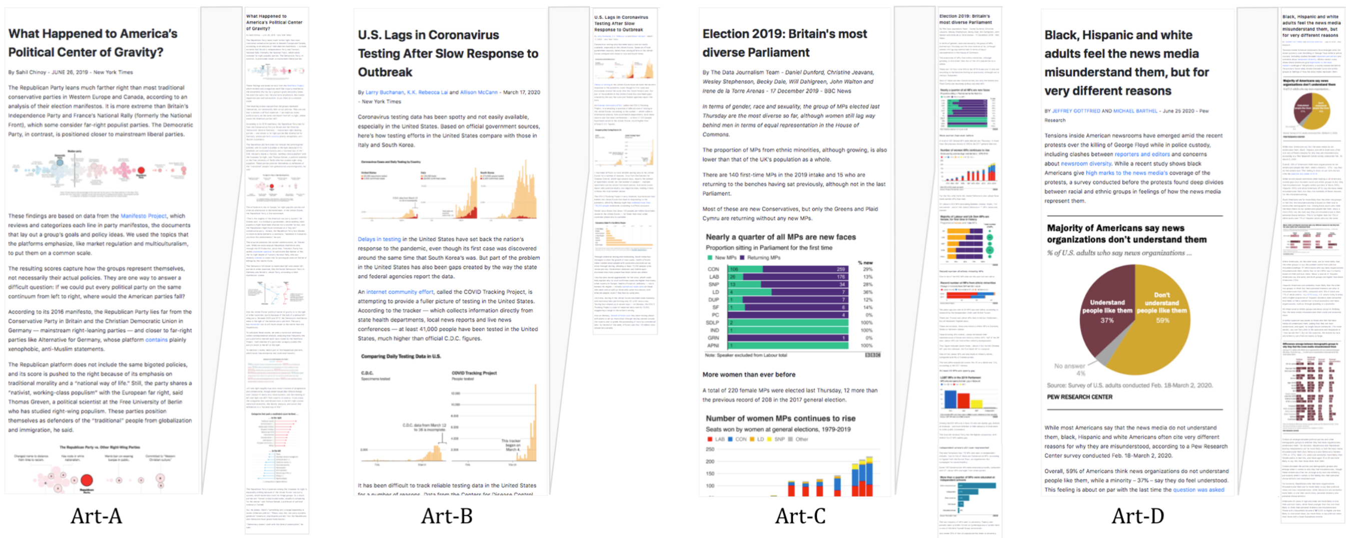


Figure 6: Overview of articles used in the author and reader studies, plus features and links to the original source article. Each article illustration features a close-up (left) and small-scale full overview (right).

way of telling the story" (C1). Another related downside is the potential for overuse, as *"the default in the tool is to get you to make a section that has a highlight in it"*, but *"a lot of the charts are fairly straightforward"* and *"don't always need highlights"* (C2).

5.1.4 Authoring workflows. The researchers who created reflowed versions of articles on their own (S1–S7) reported spending between 30 minutes and 2 hours on the task, including time to experiment and learn the tool. With in-session support (i.e., usability guidance and support on error recovery), communicators (C1-C5) were given up to 25 minutes to work on the articles, and all of them were able to complete the task within the allotted time. This suggests **articles can be reflowed in a fairly efficient manner** using *VizFlow*.

Reported (researcher) and observed (communicator) **workflows were generally similar**, including for those working with their own articles. After reading the article, some participants began splitting content into sections while others alternated between splitting sections, looking for text segments to link, and choosing reflow modes as they went. On a first pass, participants added text-chart links as connections between charts and text were found, e.g., *"Whenever I see numbers, I will look for the same numbers on the plot, and if I had the opportunity to highlight them I would highlight them"* (S4), but several tweaks would be made to improve flow and pacing. We can expect this workflow to be **more iterative if participants use *VizFlow* to author from scratch**: *"we are writing the piece and making the charts all in tandem (..) you're constantly tweaking the chart, you're exporting it out, you are making a JPEG or you are running a script, and then you are re-importing it into the tool, and then you're seeing how it looks in the piece"* (C2).

While our study did not assess how authoring practices would change if working on an article from scratch, we asked participants how they envisioned their workflows may be affected. Most participants in the researcher group (6/12) believed their writing style may change, to be more connected to the chart (S2, S5, S6), more chart-centered (S1, S4), more analytical (S3).

5.1.5 Reflow mode and layout choices. **Dynamic reflow modes were prevalent among participants.** Out of all 103 sections created, 28 (27%) were mapped to snippets over chart and another 28 to side by side. All participants used at least one of these dynamic modes. Most participants prioritized one mode over another: 4 participants used only snippets over chart, 4 used only side by side, and 4 used both. Choices tended to be based on personal preference, but many still acknowledged that each mode had its pros and cons, *"all made degree of sense"* (C4), and the choice should depend on the situation, e.g., the best fit for a chart's aspect ratio or mobile display. Some argued that dynamic modes could be less suited than the static ones for more straightforward stories (C1), and care must be taken not to overuse them (C2).

Pinned chart and *snapshots*, i.e., **the static reflow modes, were used less often** and were only mapped to a total of 4 and 8 sections, respectively. This may be attributed to an initial desire to experiment with the more dynamic *VizFlow* modes, but as participants were fairly positive towards these layouts we could expect higher uptake in the long term. They were particularly appreciated as a less distracting option that still offered reader support (8/12), albeit with room for improvement, e.g., make pinned chart interaction more mobile friendly, by making it sensitive to scroll instead of click/tap (3/12). Finally, the long form reflow mode was used in 35 sections (33%), mostly for plain text sections and only occasionally featuring static charts (3/35). Beyond serving as a reasonable default, the **long form mode was used as a buffer between other reflowed sections** to confer some *"breathing room"* (S2) and help set a more balanced *"rhythm"* (S5).

These findings illustrate the **diversity of preferences and design considerations** and underscores the **value for an expanded and improved list of reflow modes**. Suggestions for additional modes included a dynamic mode where snippets appear as scroll-triggered annotations over the chart (C3), a more space-efficient version of snapshots (C1), and a static layout where the colour of linked text matches its corresponding chart entities (S3). Other ideas for improvement entailed more fine grained control over layouts, such as alternating location of text on side by side (S4, S6), adjusting white space between sections (S2), and adjusting size and position of charts in both static and scrollytelling modes (C5).

5.1.6 Links and highlights. To check whether linking behaviours are consistent across participants, we conducted an analysis of link overlap between participant-reflowed versions of the same article. Each article was assigned to three participants; for every segment of text containing a link, we counted how many of the three participants tagged that same segment. If a link A tagged by one participant spanned two or more links B and C from another participant, then B and C each counted towards a separate individual linking segment, and A was counted as a match for both segments. Of the 81 total linked text segments across our four articles: 30 (37%) were tagged by all three participants, 24 (29%) by two participants, and 27 (33%) by one; that is, about 66% of segments were tagged by at least two participants. Articles with **higher agreement**, *Art-C* and *Art-D*, **follow a more data-centric, report-like writing style** with clearer mentions of chart elements, whereas *Art-A* and *Art-B* are presented in a more narrative style with more peripheral connections to the charts. Even in the cases where participants tagged the same text segment, they did not always match it to the same chart — about 30% of text segments in *Art-B* and 33% in *Art-A* were linked to different charts. While an author starting from scratch will often have a clear interpretation of how text and charts go together, this assessment shows that there is still **room for subjectivity and creativity in choosing text segments to link**, enough to make full automation of text-chart linking challenging [22, 25, 35].

We also assessed choices of highlighting styles. Participants produced 263 highlights across 145 links (about 1.8 highlights per link). Out of those, 57% used `contour` styles, 35% used masks (i.e., `inverse selection` plus `fill`), 20% used `fill`, and only 13% used `arrow` and `lines`. Similarly to reflow modes, choices were mostly **guided by personal preference**, but were largely **consistent within an article**. While this can be partly attributed to *VizFlow* caching and applying the most recent style to created shapes, several participants stated deliberately seeking consistency in their highlights (5/12), an aspect identified in past research [24].

As for our range of supported highlights, a few said the *"arrows and boxes"* (C2) were expressive enough and they did not feel *"short handed"* (S2), but others wanted **highlights that felt more *"native to the graphic"***, e.g., magnify (S6, S3, C4, C5) and animating chart elements (C4, C5). A communicator also emphasized the importance of **fine grained control over presentation aspects**: *"a lot of the work that we did relied on things looking quite polished (..) from an aesthetic graphics perspective. We'd spend an inordinate amount of time thinking about colours and line weights and how much to curve this arrow and you know whether it should be a straight line or how thick the arrowhead should be"* (C1).

5.2 Reader perspectives on reflowed articles

After learning the benefits and shortcomings of our approach from authors, we wanted to assess whether readers found value in the types of articles created with *VizFlow*. While past studies found that dynamic formats [33] and linking [57] have a positive effect on readers, they did not assess why these elements are considered appealing. Understanding reader preferences can inform best practices for authoring and the design of authoring tools. To do so, we conducted a survey study with 24 reader participants who were asked to compare and provide qualitative feedback on 3 versions of the same article with different degrees of reflow intervention:

(1) *Dynamic*: Mostly scrollytelling-based reflow layouts, including snippets over chart and side by side.
(2) *Light*: Mostly static-based reflow layouts, including pinned chart and snapshots.
(3) *Baseline*: Mimics the layout of the original article without any text-chart links.

We used the same 4 articles from the authoring evaluation. For the dynamic versions we repurposed author-created reflowed articles from the prior study with minor tweaks for improved uniformity and flow. From these, we created the light versions by replacing most of the scrollytelling segments with static reflow modes (pinned chart and snapshots), while retaining the same text-chart links. The baseline versions were created to mimic the layout of the original static article without links. All three versions of each article were provided exclusively through *VizFlow*'s preview mode. Links to all 12 reader interventions (4 articles $\times$ 3 versions) are provided in the Supplemental Materials.

Each participant was presented with three versions of the same article (dynamic, light, and baseline) in a counterbalanced order to avoid ordering effects. From the six possible order permutations of 3 versions, we derived 24 independent trials (4 articles $\times$ 6 presentation orders). Participants were asked to (a) list pros and cons of each version based on their "open-ended impressions from a reader perspective", and then to (b) pick a favourite version and explain why. Responses were analyzed via open and axial coding. Participants received a $5 gift card in their local currency for participation.

5.2.1 Participants. The 24 readers (R1-R24) were recruited from the general population via mailing lists and social media posting. They included 14 women, 8 men, and 2 who did not specify gender. Ages varied from 18-25 (9), 26-35 (12) and 36-45 (3). Occupations included students (16), post doctoral fellows (3), professionals in a variety of fields (4) and a family manager (1). Self reported familiarity with data visualization ranged from unfamiliar (2) and neutral (10) to familiar (6) and very familiar (6).

5.2.2 Findings. Participants submitted an average of 300 words (2-3 paragraphs) of feedback. Most participants preferred reflowed

Version	Baseline	Light	Dynamic	# Words	# Charts
Art-A	0	3	3	731	5
Art-B	3	1	2	505	3
Art-C	0	4	2	463	6
Art-D	2	2	2	738	3
Total	**5**	**10**	**9**		

Table 1: Reader-preferred versions for each article.

articles: 10 chose the light version, 9 chose the dynamic version, while only 5 preferred the baseline version. Reasons to favour the light and dynamic versions most often cited the highlights and how they helped guide the reading (7/18) and added more information (4/18), as well as improved story flow (3/18). Those who preferred the baseline version appreciated the clean (2/5) and concise (2/5) presentation, and that it encouraged one's own analysis (1/5) and was easier to navigate and skim (2/5). The light version was said to strike a balance between the dynamic and baseline versions (4/10) featuring highlights in static form or on demand, within a familiar and navigable layout. The above suggests that reflow strategies can add value to a diverse set of readers, and there is room to investigate **personalized reflowed versions to suit reader preferences**.

Beyond personal preferences, feedback also suggests that readers would choose different reflow versions depending on the task, level of analysis, and the article content. A breakdown of version preferences for each article (Table 1) shows that participants who read chart-heavy articles (*Art-A* and *Art-C*) tended to favour reflowed versions, whereas the baseline version was preferred by most readers of the simpler (*Art-B*) and more text-oriented (article *Art-D*) articles. This indicates that **different types of articles may benefit from different levels of reflow**.

Further analysis of pros and cons also revealed that despite personal preferences, **participants consistently noted the merits of the three different versions**. Readers cited occasions in which they would choose alternative reflow modes, (e.g., getting a quick overview versus close reading) which we summarize below.

The baseline version was found to be familiar (7/24), more concise (5) and easier to skim and scroll (4), but also requiring more time (5) and effort (6) since highlights are not easily provided (10). It was found effective to obtain a ***"global picture of the article"*** (R24), easier *"to go back and forth"* (R14), and to let readers *"analyse the figures entirely and see if reached the same conclusions"* (R2).

The dynamic version favoured chart analysis (13/24) through a guided (12), interactive (5), easy (3) and engaging (3) format; though potentially confusing (5), distracting (4), hard to skim due to scrollytelling interventions (6), misguiding when overemphasising visuals over text (4), and entailing a learning curve for unfamiliar readers (5). It was found best suited to ***"short attention spans"*** (R6) and for *"a step by step explanation of what is in a graph"* (R20).

Finally, the light version supported chart analysis (19/24) while providing a clear (7), familiar (3) and well paced (5) layout more amenable to skimming (3). It was also found potentially long (4), repetitive (10), and misguiding (3). It reportedly enabled ***"no distraction in terms of complex scrolling"*** (R16) but *"still had a lot of graphics that were broken down in a more understandable way"* (R3).

6 DISCUSSION AND FUTURE WORK

Our evaluations suggest that an authoring framework based on text-chart links has a lot of potential. While the concept of separating semantics and presentation is not new [37], applying these concepts to data-driven articles called for a deeper understanding of how this particular medium is structured and created. Even though our proof of concept, *VizFlow*, was heavily informed by scrollytelling media as a first iteration, our overall strategy is fundamentally modular and can accommodate a range of storytelling formats that leverage text with supporting visuals. Also, while our main focus was lowering technological barriers, the tool was found relevant even for more technically experienced authors as a quick means to prototype story flows. Our reader assessment also revealed that there is a wide and diverse range of layout preferences and that the readership could benefit from the ability to seamlessly transition across different presentation formats on demand. Our assessment also highlighted weaknesses of our approach, namely, a potential to stifle creativity by conditioning storytelling to the creation of text-chart links. This reflects the ever-existing tension between supporting simplicity and providing flexibility within creativity support tools (i.e., low threshold and high ceiling [44]) and deserves further investigation.

Below is a summary of lessons learned from our studies, in the form of design recommendations (**D1–D8**) to inform related efforts on text-chart linking for authoring dynamic data-driven articles. Several lessons align with authoring recommendations for interactive articles [20]. We also discuss remaining questions within each, which we hope will encourage further research.

Support early drafting and ideation stages (D1). While our evaluation with authors suggests our proposed concept would extend well in the wild, we did not directly assess the impact on authoring workflows when working on articles from scratch. From formative findings, we learned that authors spend significant time iterating on their stories; in the evaluation, they envisioned writing styles could change to make full use of the added dynamic features, and some were interested in using *VizFlow* as a tool for prototyping dynamic behaviour. We argue there is value in supporting these interventions earlier in authoring workflows, and that it may better guide storytelling design. Some automated support, e.g., in reducing the burden of manually linking text and charts, is an important related concern, which we further detail under **D4**.

Aesthetics matter (D2). Both communicators and researchers in our evaluation study underscored the importance of stories that "look nice" in increasing reader engagement. These concerns ranged from the visual design of the charts and their transition effects to the general pacing of the article and spacing between sections; they also align well with recent research on visual aesthetics in digital stories and their effect on engagement and learning [17]. Future efforts can focus on providing more fine-grained control of presentation aspects, including options to further customize reflow modes (e.g., choosing whether to place text on the left or right on a side by side layout), to more polished and powerful chart authoring capabilities, which we discuss under **D3**.

Support integration with visualization authoring tools (D3). As we focused on authoring and on the relevance of text-chart links, we purposefully kept our highlighting capabilities simple. But our assessment of scrollytelling articles plus participant feedback show that there is much to gain by integrating visualization authoring tools into the workflow. It would enable effects more native to the chart (e.g., repositioning, morphing) as well as more complex chart layouts (e.g., force layouts). In particular, it would also better support the notion of using consecutive text-chart links as keyframes of a visual sequence, in something akin to slideshow transitions or video animations. We should also acknowledge that this would add complexity to authoring efforts, and striking a balance between simplicity and flexibility on this front may prove to be a significant undertaking. First steps could allow *VizFlow* to link to visualizations created with more accessible tools like Tableau [52], Vega-Lite [39] and Observable notebooks [21], providing vector-based representations that are more amenable to native effects, transitions, and interaction. That said, in view of the diverse charting practices reported by authors in our studies, we argue that vector chart authoring options should be offered as an addition — and not a replacement — to our more chart-agnostic approach, i.e., still allowing authors work with their current external tools even if only leveraging limited highlighting capabilities.

Provide automation support while maintaining authoring freedom (D4). There is potential to automate many tasks within our workflow that would be especially useful in early authoring stages. Real-time automatic detection of text-chart links could provide link suggestions while authors are drafting content, and also help re-map or maintain linkage consistency as content evolves. Another opportunity is detecting what chart elements should be highlighted within a text-chart link, which builds on existing efforts to enable automatic chart element extraction [9, 36, 40]. In the interest of visual consistency [24], a third opportunity for automation is enabling automatic style consistency across all charts when a chart is modified. *VizFlow* could also be used as a tool for annotating articles in support of training the kinds of automated models proposed above. However, we underscore that authoring is a fundamentally creative effort and there is significant room for subjectivity even within simpler tasks, as evidenced by our assessment of link overlap (§ 5.1.6). As such, these interventions should not be implemented as fully autonomous actions, but rather as supporting features offered to authors to help expedite their workflows.

More (reflow modes) is better (D5). While our reflow options were limited to popular layouts based on our scrollytelling assessment, our author and reader evaluations hint at a wide range of preferences and reading tasks. Author evaluation also found that quick experimentation with storytelling layouts was one of the highlights of their experience, suggesting that having more options readily available would be valuable to authoring, and is something our extensible framework easily enables. For instance, additional storytelling formats could include horizontal scrollers *(Ex: 40)*, dynamic direct linkage between text and charts [28, 49], and scroll-based versions of traditional steppers [33, 51, 57] *(Ex: 42)*. Reader assessments suggest that static text-chart layouts are also worth considering, e.g., a small multiples version of the snapshots mode *(Ex: 37)*.

Empower readers with choice and tailored support (D6). Our reader study showed that preferences are diverse, depending on reader goals at a given time (e.g., whether skimming or delving into details). Allowing readers to reflow an article to a format that matches their reading style or task is likely to be beneficial, and something our proposed approach could support. Similar to the idea of adapting visualizations to different devices [19], we propose that adapting storytelling formats to different readers and reading contexts has potential to significantly transform the reading experience. It opens up space for future explorations not only in the adaptive visualizations space [5, 27, 53] but also in defining what a relevant spectrum of storytelling modes for data-driven articles may look like given preferences and reading tasks of a wider audience.

Support authors in making balanced authoring choices (D7). While readers appreciated the easy experimentation enabled by our approach, authors warned for a potential to overuse reflow modes that would lead to distracting or frustrating reader experiences. An important next step after obtaining a better understanding of reader preferences and tasks listed above (**D6**) is providing actionable and integrated guidance to authors to avoid such authoring pitfalls and make more informed storytelling choices.

Support collaboration (D8). A core tenet of our no-coding approach was to reduce dependence on programming expertise and forced team effort. However, many stories still demand synergistic creative efforts between various experts such as data scientists, journalists, and graphic designers [11]. Supporting collaborative authoring is especially key in professional settings like newsrooms, to ensure that time using the tool is spent on authoring and creation and not on figuring out how to effectively collaborate around it.

7 CONCLUSION

In this work, we explored an approach to storytelling authoring of data-driven articles that separates semantics and presentation in the form of text-chart links and storytelling layouts. Our approach was grounded on existing authoring practices, which informed the design and development of *VizFlow*, a proof of concept authoring tool to create dynamic data-driven articles. Our evaluation findings with authors and readers underscore the value of this approach towards more efficient and effective addition of dynamic behaviour to data-driven articles, and help make a wider variety of consumption formats more readily available for the medium. Our work also contributed to a deeper understanding of scrollytelling approaches, which to our knowledge, had not yet been systematically assessed.

It has never been more critical to deliver fact-based and data-driven information to the general public. Ultimately, we hope our work and proposed directions for future investigation can contribute to empower both authors and readers towards a more data-aware and well informed world.

ACKNOWLEDGMENTS

We thank the anonymous reviewers, Naeemul Hassan, Ben Shneiderman and Vidya Setlur for their feedback to improve this manuscript, and Sumit Shekhar for his support with LEAF-Net. We also thank the exceptionally talented and insightful authors who took part in our studies.

REFERENCES

[1] Adobe. 2020. Adobe Spark. https://spark.adobe.com/
[2] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Andres Monroy-Hernandez, and Pourang Irani. 2016. Authoring data-driven videos with dataclips. *IEEE transactions on visualization and computer graphics* 23, 1 (2016), 501–510.

[3] Benjamin Bach, D Stefaner, Jeremy Boy, Steven Drucker, Lyn Bartram, Jo Wood, Paolo Ciuccarelli, Yuri Engelhardt, Ulrike Koeppen, and Barbara Tversky. 2018. Narrative design patterns for data-driven storytelling. In *Data-Driven Storytelling*. CRC Press, Boca Raton, FL, USA, 107–133.

[4] Sriram Karthik Badam, Zhicheng Liu, and Niklas Elmqvist. 2018. Elastic documents: Coupling text and tables through contextual visualizations for enhanced document reading. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 661–671.

[5] Oswald Barral, Sébastien Lallé, Grigorii Guz, Alireza Iranpour, and Cristina Conati. 2020. Eye-Tracking to Predict User Cognitive Abilities and Performance for User-Adaptive Narrative Visualizations. In *Proceedings of the 2020 International Conference on Multimodal Interaction*. ACM, Utrecht, Netherlands, 163–173.

[6] Leilani Battle, Peitong Duan, Zachery Miranda, Dana Mukusheva, Remco Chang, and Michael Stonebraker. 2018. Beagle: Automated extraction and interpretation of visualizations from the web. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal, Canada, 1–8.

[7] Mike Bostock. 2014. How to Scroll. https://bost.ocks.org/mike/scroll/

[8] John Brooke. 1996. . CRC press, Boca Raton, FL, USA, Chapter SUS: a "quick and dirty" usability scale, 189.

[9] Ritwick Chaudhry, Sumit Shekhar, Utkarsh Gupta, Pranav Maneriker, Prann Bansal, and Ajay Joshi. 2020. LEAF-QA: Locate, encode & attend for figure question answering. In *The IEEE Winter Conference on Applications of Computer Vision*. IEEE, Aspen, CO, USA, 3512–3521.

[10] Siming Chen, Jie Li, Gennady Andrienko, Natalia Andrienko, Yun Wang, Phong H Nguyen, and Cagatay Turkay. 2018. Supporting story synthesis: Bridging the gap between visual analytics and storytelling. *IEEE transactions on visualization and computer graphics* 26, 7 (2018), 2499–2516.

[11] Fanny Chevalier, Melanie Tory, Bongshin Lee, Jarke van Wijk, Giuseppe Santucci, Marian Dörk, and Jessica Hullman. 2018. From Analysis to Communication: Supporting the Lifecycle of a Story. In *Data-Driven Storytelling*. AK Peters/CRC Press, Boca Raton, FL, USA, 169–202.

[12] Matthew Conlen and Jeffrey Heer. 2018. Idyll: A markup language for authoring and publishing interactive articles on the web. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. ACM, Berlin, Germany, 977–989.

[13] Pierre Dragicevic, Yvonne Jansen, Abhraneel Sarma, Matthew Kay, and Fanny Chevalier. 2019. Increasing the transparency of research papers with explorable multiverse analyses. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow, Scotland, UK, 1–15.

[14] Ryan Eccles, Thomas Kapler, Robert Harper, and William Wright. 2008. Stories in geotime. *Information Visualization* 7, 1 (2008), 3–17.

[15] Honkytonk Films. 2020. Klynt. http://www.klynt.net/

[16] Mario Garcia. 2019. *The Story*. Thane & Prose, New York, NY, USA.

[17] Esther Greussing and Hajo G Boomgaarden. 2019. Simply bells and whistles? Cognitive effects of visual aesthetics in digital longforms. *Digital Journalism* 7, 2 (2019), 273–293.

[18] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*. IEEE, New York, NY, USA, 2961–2969.

[19] Jane Hoffswell, Wilmot Li, and Zhicheng Liu. 2020. Techniques for Flexible Responsive Visualization Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13.

[20] Fred Hohman, Matthew Conlen, Jeffrey Heer, and Duen Horng Polo Chau. 2020. Communicating with Interactive Articles. *Distill* 5, 9 (2020), e28. https://distill.pub/2020/communicating-with-interactive-articles/

[21] Observable HQ. 2020. Observable. https://observablehq.com/

[22] Dae Hyun Kim, Enamul Hoque, Juho Kim, and Maneesh Agrawala. 2018. Facilitating document reading by linking text and tables. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. ACM, Berlin, Germany, 423–434.

[23] Nam Wook Kim, Nathalie Henry Riche, Benjamin Bach, Guanpeng Xu, Matthew Brehmer, Ken Hinckley, Michel Pahud, Haijun Xia, Michael J McGuffin, and Hanspeter Pfister. 2019. Datatoon: Drawing dynamic network comics with pen+touch interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow, Scotland, UK, 1–12.

[24] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2017. Internal and external visual cue preferences for visualizations in presentations. *Computer Graphics Forum* 36, 3 (2017), 515–525.

[25] Nicholas Kong, Marti A Hearst, and Maneesh Agrawala. 2014. Extracting references between text and charts via crowdsourcing. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, Toronto, Canada, 31–40.

[26] Robert Kosara. 2016. The scrollytelling scourge. https://eagereyes.org/blog/2016/the-scrollytelling-scourge

[27] Sébastien Lallé, Dereck Toker, and Cristina Conati. 2019. Gaze-Driven Adaptive Interventions for Magazine-Style Narrative Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 1, 1 (2019), 1–10.

[28] Sébastien Lallé, Tiffany Wu, and Cristina Conati. 2020. Gaze-Driven Links for Magazine Style Narrative Visualizations. In *2020 IEEE Visualization Conference (VIS)*. IEEE, Salt Lake City, UT, USA, 1–4.

[29] Shahid Latif, Diao Liu, and Fabian Beck. 2018. Exploring Interactive Linking Between Text and Visualization. In *EuroVis (Short Papers)*. IEEE, Brno, Czech Republic, 91–94.

[30] Shahid Latif, Kaidie Su, and Fabian Beck. 2019. Authoring Combined Textual and Visual Descriptions of Graph Data. In *Proc. Eurographics Conf. Visualization*. IEEE, Porto, Portugal, 1–5.

[31] Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. 2015. More than telling a story: Transforming data into visually shared stories. *IEEE computer graphics and applications* 35, 5 (2015), 84–90.

[32] Zhicheng Liu, John Thompson, Alan Wilson, Mira Dontcheva, James Delorey, Sam Grigg, Bernard Kerr, and John Stasko. 2018. Data Illustrator: Augmenting vector design tools with lazy data binding for expressive visualization authoring. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal, Canada, 1–13.

[33] Sean McKenna, N Henry Riche, Bongshin Lee, Jeremy Boy, and Miriah Meyer. 2017. Visual narrative flow: Exploring factors shaping data visualization story reading experiences. *Computer Graphics Forum* 36, 3 (2017), 377–387.

[34] Honghui Mei, Yuxin Ma, Yating Wei, and Wei Chen. 2018. The design space of construction tools for information visualization: A survey. *Journal of Visual Languages & Computing* 44 (2018), 120–132.

[35] Ronald Metoyer, Qiyu Zhi, Bart Janczuk, and Walter Scheirer. 2018. Coupling story to visualization: Using textual analysis as a bridge between data and interpretation. In *23rd International Conference on Intelligent User Interfaces*. ACM, Tokyo, Japan, 503–507.

[36] Jorge Poco and Jeffrey Heer. 2017. Reverse-engineering visualizations: Recovering visual encodings from chart images. *Computer Graphics Forum* 36, 3 (2017), 353–363.

[37] Arvind Satyanarayan and Jeffrey Heer. 2014. Authoring narrative visualizations with ellipsis. *Computer Graphics Forum* 33, 3 (2014), 361–370.

[38] Arvind Satyanarayan, Bongshin Lee, Donghao Ren, Jeffrey Heer, John Stasko, John Thompson, Matthew Brehmer, and Zhicheng Liu. 2019. Critical reflections on visualization authoring systems. *IEEE transactions on visualization and computer graphics* 26, 1 (2019), 461–471.

[39] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. 2017. Vega-Lite: A Grammar of Interactive Graphics. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)* 23, 1 (2017), 341–350. http://idl.cs.washington.edu/papers/vega-lite

[40] Manolis Savva, Nicholas Kong, Arti Chhajta, Li Fei-Fei, Maneesh Agrawala, and Jeffrey Heer. 2011. Revision: Automated classification, analysis and redesign of chart images. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, Santa Barbara, CA, USA, 393–402.

[41] Edward Segel and Jeffrey Heer. 2010. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics* 16, 6 (2010), 1139–1148.

[42] Doris Seyser and Michael Zeiller. 2018. Scrollytelling–an analysis of visual storytelling in online journalism. In *2018 22nd International Conference Information Visualisation (IV)*. IEEE, Fisciano, Italy, 401–406.

[43] Bill Shander. 2020. The Past, Present, and Future of Scrollytelling. https://medium.com/nightingale/the-past-present-and-future-of-scrollytelling-10dd37dc1003

[44] Ben Shneiderman. 2007. Creativity support tools: Accelerating discovery and innovation. *Commun. ACM* 50, 12 (2007), 20–32.

[45] Shorthand. 2020. Shorthand, a digital storytelling platform. https://shorthand.com/

[46] Waralak V Siricharoen. 2013. Infographics: the new communication tools in digital age. In *The international conference on e-technologies and business on the web (ebw2013)*. IEEE, Bangkok, Thailand, 169–174.

[47] Codevise Solutions. 2020. Pageflow: Interactive multimedia storytelling with ease. https://www.pageflow.io/

[48] Florian Stalph. 2018. Classifying Data Journalism: A content analysis of daily data-driven stories. *Journalism Practice* 12, 10 (2018), 1332–1350.

[49] Markus Steinberger, Manuela Waldner, Marc Streit, Alexander Lex, and Dieter Schmalstieg. 2011. Context-preserving visual links. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2249–2258.

[50] Charles D. Stolper, Bongshin Lee, N. Henry Riche, and John Stasko. 2016. *Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories*. Technical Report. Microsoft Research, Washington, USA.

[51] Tableau. 2020. Stories. https://help.tableau.com/current/pro/desktop/en-us/stories.htm

[52] Tableau. 2020. Tableau. https://www.tableau.com/

[53] Dereck Toker, Cristina Conati, and Giuseppe Carenini. 2018. User-adaptive support for processing magazine style narrative visualizations: Identifying user characteristics that matter. In *23rd International Conference on Intelligent User Interfaces*. ACM, Tokyo, Japan, 199–204.

[54] Chao Tong, Richard Roberts, Rita Borgo, Sean Walton, Robert S Laramee, Kodzo Wegba, Aidong Lu, Yun Wang, Huamin Qu, Qiong Luo, et al. 2018. Storytelling and visualization: An extended survey. *Information* 9, 3 (2018), 65.

[55] Archie Tse. 2016. Why we are doing fewer interactives. https://github.com/archietse/malofiej-2016/blob/master/tse-malofiej-2016-slides.pdf
[56] Haijun Xia, Nathalie Henry Riche, Fanny Chevalier, Bruno De Araujo, and Daniel Wigdor. 2018. DataInk: Direct and creative data-oriented drawing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal, Canada, 1–13.
[57] Qiyu Zhi, Alvitta Ottley, and Ronald Metoyer. 2019. Linking and Layout: Exploring the Integration of Text and Visualization in Storytelling. *Computer Graphics Forum* 38, 3 (2019), 675–685.

A CATALOG OF SCROLLYTELLING ARTICLES

(1) M. Lee. **Green Honey**. 2013. http://muyueh.com/greenhoney/
(2) T. Chu. **Let's Free Congress / Money wins Elections**. 2013. http://letsfreecongress.org/
(3) A. Dant, H. Fairfield. **Fewer Helmets, More Deaths**. 2014. http://www.nytimes.com/interactive/2014/03/31/science/motorcycle-helmet-laws.html
(4) G. Murphy. **A Visual Analysis of Battle at the Berrics**. 2014. http://www.georgelmurphy.com/berrics/
(5) M. Barry, B. Card. **Visualizing MBTA Data: An Interactive Exploration of Boston's Subway System**. 2014. http://mbtaviz.github.io/
(6) G. Aisch. **The Clubs that Connect The World Cup**. 2014. http://www.nytimes.com/interactive/2014/06/20/sports/worldcup/how-world-cup-players-are-connected.html
(7) . **How the Recession Reshaped the Economy, in 255 Charts**. 2014. https://www.nytimes.com/interactive/2014/06/05/upshot/how-the-recession-reshaped-the-economy-in-255-charts.html
(8) J. Treat, A. Scalamogna. **We'll Have What They're Having**. 2014. https://www.nationalgeographic.com/foodfeatures/diet-similarity/
(9) S. Yee, T. Chu. **A Visual Introduction to Machine Learning**. 2015. http://www.r2d3.us/visual-intro-to-machine-learning-part-1/
(10) A. Pearce et al.. **Scientific Proof that Americans are Completely Addicted to Trucks**. 2015. http://www.bloomberg.com/graphics/2015-auto-sales/
(11) J. Ashkenas et al.. **How the U.S. and OPEC Drive Oil Prices**. 2015. http://www.nytimes.com/interactive/2015/09/30/business/how-the-us-and-opec-drive-oil-prices.html
(12) E. Roston et al.. **Bloomberg Carbon Clock**. 2015. http://www.bloomberg.com/graphics/carbon-clock/
(13) K. Collins. **Why Infectious Bacteria Are Winning**. 2015. http://qz.com/576057/why-infectious-bacteria-are-winning/
(14) E. Roston, B. Migliozzi. **What's Really Warming the World?**. 2015. http://www.bloomberg.com/graphics/2015-whats-warming-the-world/
(15) R. Spencer. **Fleeing Syria for Europe: Safaa's fatal journey**. 2015. http://s.telegraph.co.uk/graphics/projects/safaas-journey/index.html
(16) P. Minczeski et al.. **How Fed Rates Move Markets**. 2015. http://graphics.wsj.com/reacting-to-fed-rates/
(17) S. Yaccino et al.. **This is Who Republican Presidential Contenders Follow on Twitter**. 2015. https://www.bloomberg.com/politics/graphics/2015-who-republican-candidates-follow/
(18) I. Boudway, C. Lindblad. **The Year Ahead 2016: 50 Companies to Watch**. 2016. http://www.bloomberg.com/graphics/year-ahead-2016/
(19) H. Anderson, M. Daniels. **Film Dialogue from 2,000 Screenplays, Broken Down by Gender and Age**. 2016. http://polygraph.cool/films/
(20) B. Gardiner et al.. **The Dark Side of Guardian Comments**. 2016. https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments
(21) R. Goldenberg. **Started From The Bottom**. 2016. http://russellgoldenberg.github.io/started-from-the-bottom/
(22) H. Park, T. Griggs. **Could Trump Really Deport Millions of Unauthorized Immigrants?**. 2016. https://www.nytimes.com/interactive/2016/11/29/us/trump-unauthorized-immigrants.html
(23) L. Gamio et al.. **1,000 Times Gold**. 2016. https://www.washingtonpost.com/graphics/sports/olympics/the-1000-medals-of-the-united-states/
(24) J. Abelson et al.. **Clash in the Name of Care**. 2016. https://apps.bostonglobe.com/spotlight/clash-in-the-name-of-care/story/
(25) M. Daniels, M. Beuoy. **Most Unlikely Comebacks: Using Historical Data To Rank Statistically Improbable Wins**. 2017. http://polygraph.cool/nba/
(26) R. Goldenberg, D. Kopf. **Making it Big**. 2017. https://pudding.cool/2017/01/making-it-big/
(27) A. Thomas. **Free Willy and Flipper by the Numbers**. 2017. https://pudding.cool/2017/07/cetaceans/
(28) Outside in America team. **Bussed out: How America moves its homeless**. 2017. https://www.theguardian.com/us-news/ng-interactive/2017/dec/20/bussed-out-america-moves-homeless-people-country-study
(29) J. Diehm, A. Thomas. **Women's pockets are inferior**. 2018. https://pudding.cool/2018/08/pockets/
(30) S. Vickars. **What makes a titletown?**. 2018. https://pudding.cool/2018/11/titletowns/
(31) R. Goldenberg. **Life after Death on Wikipedia**. 2018. https://pudding.cool/2018/08/wiki-death/
(32) V. Shenoy. **Kohli's path to 10,000 ODI runs, charted**. 2018. https://shapeofthegame.com/kohli-races-to-10000/
(33) W. Cai. **War of Worlds**. 2018. http://fingfx.thomsonreuters.com/gfx/rngs/NORTHKOREA-USA-KIM-TRUMP/010070JM16P/index.html
(34) R. Goldenberg. **The World through the eyes of the US**. 2018. https://pudding.cool/2018/12/countries/
(35) J. Holder, A. Hern. **Bezos's empire: how Amazon became the world's most valuable retailer**. 2018. https://www.theguardian.com/technology/ng-interactive/2018/apr/24/bezoss-empire-how-amazon-became-the-worlds-biggest-retailer
(36) J.P. Corbeil et al.. **Comment decortiquer la victoire de la CAQ?**. 2019. https://www.ledevoir.com/documents/special/18-10_resultats_election_qc_2018/index.html
(37) C. Guiborg et al.. **Falling through the gap: How the UK's gender pay gap sizes up**. 2019. https://www.bbc.co.uk/news/resources/idt-a524dd3a-c09c-4f09-bc03-c5006d75ef96
(38) R. Goldenberg, A. Thomas. **How many high school stars make it in the NBA?**. 2019. https://pudding.cool/2019/03/hype/
(39) C Aguilar Garcia, P. Whiteside. **Why 7,000 people die needlessly every day**. 2019. https://news.sky.com/story/why-7-000-people-die-needlessly-every-day-11770982
(40) S. Chinoy, J. Ma. **How Every Member Got to Congress**. 2019. https://www.nytimes.com/interactive/2019/01/26/opinion/sunday/paths-to-congress.html
(41) A. Barret. **The Evolution of the American Census**. 2020. https://pudding.cool/2020/03/census-history/
(42) S.M. Bell et al.. **Genomic analysis of COVID-19 spread. Situation report 2020-03-27**. 2020. https://nextstrain.org/narratives/ncov/sit-rep/2020-03-27
(43) China Power. **How does water security affect China's development?**. 2020. https://chinapower.csis.org/china-water-security/

				Past surveys			Text-Chart Layout (Desktop)					Text-Chart Layout (Mobile)							Scroll Effects (Chart)						
	Title	Author	Year	ID (McKenna et al.)	ID (Stolper et al.)	ID (Hoffswell et al.)	Side by side	Snippets over chart	Pinned chart	Long chart	Long form	Side by side	Snippets over chart	Pinned chart	Long chart	Long form	Snapshots	No adaptation	Reveal	Hide	Emphasize	De-emphasize	Reposition	Morph	Rescale Axes
1	Green Honey	M. Lee	2013	56	28		-	x	-	-	-	-	-	-	-	-	-	x	x	x	x	x	x	-	-
2	Let's Free Congress / Money wins Elections	T. Chu	2013	59			x	-	-	-	-	-	-	-	-	x	-	-	x	-	-	-	-	-	-
3	Fewer Helmets, More Deaths	A. Dant, H. Fairfield	2014	3			x	-	-	-	-	-	-	-	-	-	-	x	x	x	-	-	-	x	x
4	A Visual Analysis of Battle at the Berrics	G. Murphy	2014	5			-	-	-	-	x	-	-	-	-	-	-	-	x	x	-	-	-	x	-
5	Visualizing MBTA Data: An Interactive Exploration of Boston's Subway System	M. Barry, B. Card	2014	10	9		x	-	-	x	x	-	-	-	-	x	-	-	-	-	-	-	-	-	-
6	The Clubs that Connect The World Cup	G. Aisch	2014	57		1	x	-	-	-	-	-	-	-	-	x	x	-	x	x	-	-	x	-	-
7	How the Recession Reshaped the Economy, in 255 Charts		2014				-	x	-	-	-	-	x	-	-	-	-	-	x	-	x	x	x	x	-
8	We'll Have What They're Having	J. Treat, A. Scalamogna	2014			47	-	-	-	x	-	-	-	-	-	x	x	-	-	-	-	-	-	-	-
9	A Visual Introduction to Machine Learning	S. Yee, T. Chu	2015	1			x	-	-	-	-	x	-	-	-	-	-	-	x	-	-	x	x	x	-
10	Scientific Proof that Americans are Completely Addicted to Trucks	A. Pearce et al.	2015	2	34		x	-	-	-	-	-	-	-	-	x	x	-	-	-	-	-	x	-	-
11	How the U.S. and OPEC Drive Oil Prices	J. Ashkenas et al.	2015	12			-	x	-	-	-	-	x	-	-	-	-	-	x	-	x	x	-	x	-
12	Bloomberg Carbon Clock	E. Roston et al.	2015	32			-	x	-	-	-	-	-	-	-	x	x	-	x	x	-	-	-	-	x
13	Why Infectious Bacteria Are Winning	K. Collins	2015	43			-	x	-	-	-	-	x	-	-	-	-	-	x	x	-	-	x	x	-
14	What's Really Warming the World?	E. Roston, B. Migliozzi	2015	47	48		-	x	-	-	-	-	-	-	-	x	x	-	x	x	-	-	-	-	-
15	Fleeing Syria for Europe: Safaa's fatal journey	R. Spencer	2015	64			x	-	-	-	x	-	-	-	-	x	-	-	-	-	-	-	-	-	-
16	How Fed Rates Move Markets	P. Minczeski et al.	2015	75			-	x	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-
17	This is Who Republican Presidential Contenders Follow on Twitter	S. Yaccino et al.	2015			53	-	-	x	-	-	-	-	-	-	x	x	-	-	-	x	x	-	-	-
18	The Year Ahead 2016: 50 Companies to Watch	I. Boudway, C. Lindblad	2016	30			x	-	-	-	-	-	-	-	-	x	-	-	-	-	x	x	-	-	-
19	Film Dialogue from 2,000 Screenplays, Broken Down by Gender and Age	H. Anderson, M. Daniels	2016	46			-	-	x	-	x	-	-	-	-	-	-	x	x	-	-	-	x	x	-
20	The Dark Side of Guardian Comments	B. Gardiner et al.	2016	51			-	x	-	-	x	-	-	-	-	-	-	-	x	-	-	x	-	x	x
21	Started From The Bottom	R. Goldenberg	2016	62			x	-	-	-	-	-	-	x	-	-	-	-	x	x	-	-	x	x	-
22	Could Trump Really Deport Millions of Unauthorized Immigrants?	H. Park, T. Griggs	2016				-	x	-	-	x	-	x	-	-	-	-	-	-	x	x	x	x	-	-
23	1,000 Times Gold	L. Gamio et al.	2016				x	-	-	-	-	-	-	-	-	x	x	-	-	-	x	-	x	-	-
24	Clash in the Name of Care	J. Abelson et al.	2016				x	-	-	-	-	-	-	-	-	x	x	-	x	-	-	-	x	-	x
25	Most Unlikely Comebacks: Using Historical Data To Rank Statistically Improbable Wins	M. Daniels, M. Beuoy	2017	61			x	-	-	-	-	-	-	-	-	-	-	x	x	-	-	-	-	-	-
26	Making it Big	R. Goldenberg, D. Kopf	2017				x	-	-	-	x	-	x	-	-	-	-	-	-	-	x	-	x	-	-
27	Free Willy and Flipper by the Numbers	A. Thomas	2017				-	x	-	-	-	-	x	-	-	-	-	-	x	-	-	-	-	-	-
28	Bussed out: How America moves its homeless	Outside in America team	2017				-	x	-	-	x	-	-	-	-	-	-	x	x	-	-	-	x	-	x
29	Women's pockets are inferior	J. Diehm, A. Thomas	2018				-	x	-	-	-	-	x	-	-	-	-	-	x	-	x	-	x	-	-
30	What makes a titletown?	S. Vickars	2018				x	-	-	-	-	-	x	-	-	-	-	-	x	x	x	-	x	x	x
31	Life after Death on Wikipedia	R. Goldenberg	2018				-	x	-	x	x	-	x	-	-	-	-	-	x	x	x	-	-	-	-
32	Kohli's path to 10,000 ODI runs, charted	V. Shenoy	2018				x	-	-	-	-	-	x	-	-	-	-	-	x	-	-	-	x	x	x
33	War of Worlds	W. Cai	2018			10	-	-	x	-	-	-	-	x	-	-	-	-	x	-	-	x	-	-	-
34	The World through the eyes of the US	R. Goldenberg	2018			20	-	-	-	x	-	-	-	x	-	-	-	-	-	-	x	x	-	-	-
35	Bezos's empire: how Amazon became the world's most valuable retailer	J. Holder, A. Hern	2018			27	x	-	-	x	-	-	x	-	-	-	-	-	-	-	-	-	x	-	-
36	Comment decortiquer la victoire de la CAQ?	J.P. Corbeil et al.	2019				x	-	-	-	-	-	x	-	-	-	-	-	x	x	-	-	x	x	-
37	Falling through the gap: How the UK's gender pay gap sizes up	C. Guiborg et al.	2019				-	x	-	-	x	-	-	-	-	x	-	-	x	-	x	-	x	-	-
38	How many high school stars make it in the NBA?	R. Goldenberg, A. Thomas	2019			21	-	-	-	-	x	-	-	-	-	x	-	-	-	-	-	-	x	-	-
39	Why 7,000 people die needlessly every day	C Aguilar Garcia, P. Whiteside	2019				-	x	-	-	x	-	x	-	-	-	-	-	x	-	x	-	x	x	-
40	How Every Member Got to Congress	S. Chinoy, J. Ma	2019			14	-	-	x	-	-	-	-	-	-	x	x	-	-	-	-	-	-	-	-
41	The Evolution of the American Census	A. Barret	2020				x	-	-	x	-	-	x	-	-	-	-	-	x	-	-	-	-	-	-
42	Genomic analysis of COVID-19 spread. Situation report 2020-03-27	S.M. Bell et al.	2020				x	-	-	-	-	-	-	-	-	-	-	-	x	x	x	x	x	-	-
43	How does water security affect China's development?	China Power	2020				-	x	-	-	x	-	x	-	-	-	-	-	x	-	x	-	x	-	-

Figure 7: Scrollytelling articles and subset of catalogued features.

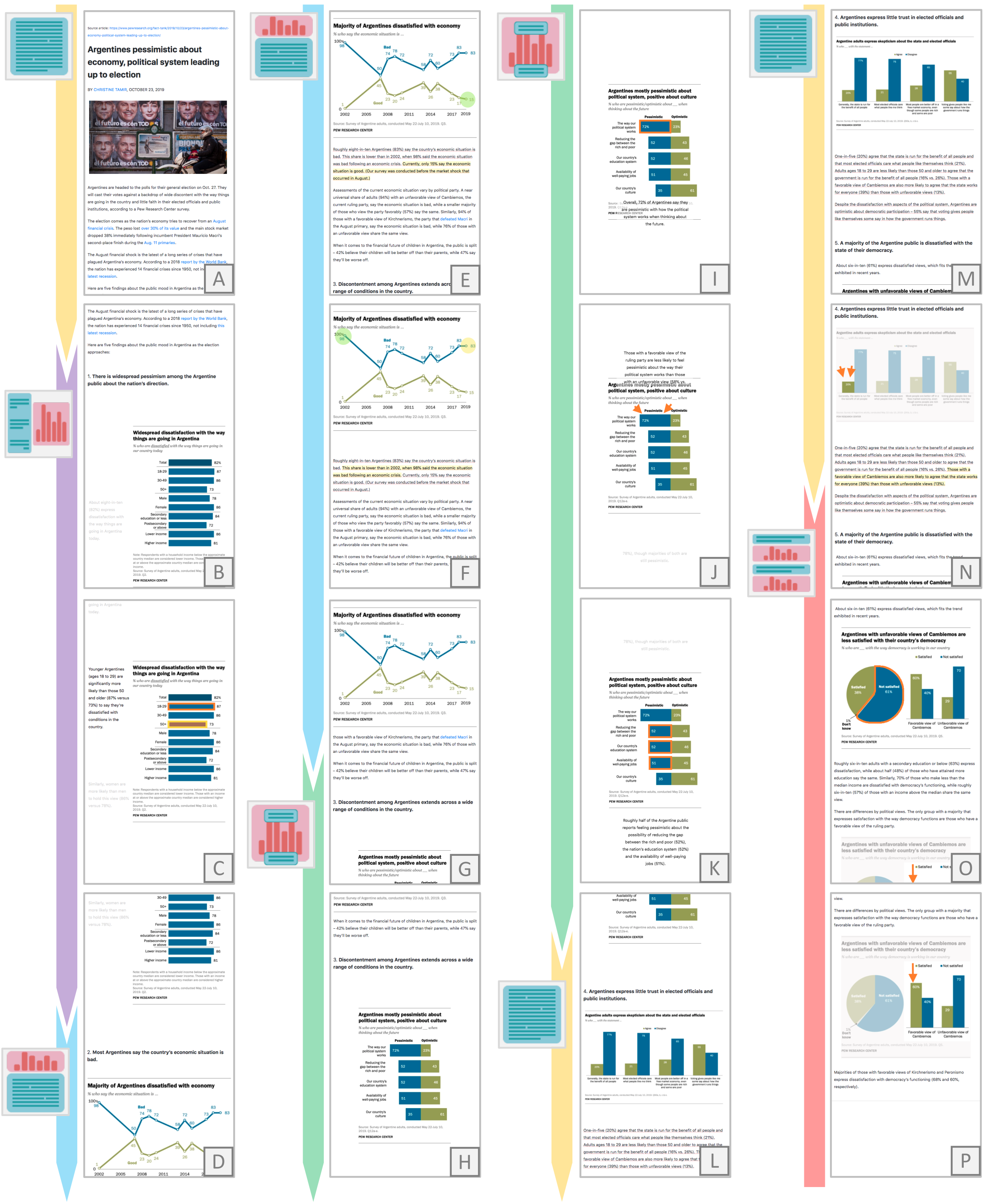


Figure 8: Preview of a complete reflowed article, featuring all reflow modes and transitions between sections. Live *VizFlow* previews can be accessed here: [http://vizflow.cs.toronto.edu/reflow?story=chi2021-demo1] [http://vizflow.cs.toronto.edu/reflow?story=chi2021-demo2]