A Design Space for Applying the Freytag's Pyramid Structure to Data Stories

Leni Yang, Xian Xu, XingYu Lan, Ziyan Liu, Shunan Guo, Yang Shi, Huamin Qu, and Nan Cao

Abstract— Data stories integrate compelling visual content to communicate data insights in the form of narratives. The narrative structure of a data story serves as the backbone that determines its expressiveness, and it can largely influence how audiences perceive the insights. Freytag's Pyramid is a classic narrative structure that has been widely used in film and literature. While there are continuous recommendations and discussions about applying Freytag's Pyramid to data stories, little systematic and practical guidance is available on how to use Freytag's Pyramid for creating structured data stories. To bridge this gap, we examined how existing practices apply Freytag's Pyramid by analyzing stories extracted from 103 data videos. Based on our findings, we proposed a design space of narrative patterns, data flows, and visual communications to provide practical guidance on achieving narrative intents, organizing data facts, and selecting visual design techniques through story creation. We evaluated the proposed design space through a workshop with 25 participants. Results show that our design space provides a clear framework for rapid storyboarding of data stories with Freytag's Pyramid.

Index Terms—Freytag's Pyramid, Narrative Structure, Narrative Visualization, Data Storytelling, Data Video

1 Introduction

Data stories integrate various visual content (e.g., text, visualization, and animation) to tell stories that convey data insights [56]. To create expressive data stories, designers carefully plan the narrative structures of their stories. According to prior research, the narrative structure of a story can affect the audience's perception, understanding, and memory of it [33]. The narrative structure also influences the considerations of visual design. For example, in visual storytelling, it is considered beneficial to use high-energy colors or to increase the movement speed of the camera in the climax of a story to elicit tension [11].

There has been an increasing interest in studying the narrative structures of data stories [33, 34, 46, 56]. For example, Segel and Heer [56] identified three high-level structures (i.e., *Martini Glass Structure*, *Interactive Slideshow*, and *Drill-Down Story*) that provide readers with various levels of control over the manipulation of the story presentation in interactive visualizations. In recent years, the visualization community has also been actively studying the potential of using Freytag's Pyramid and its variations in data storytelling [2, 24, 27, 50]. Freytag's Pyramid [29] is a classic narrative structure that has been widely applied in film and literature [25, 29]. Adhering to this structure, a story starts with introducing the setting, develops with a progressive increase of the tension that reaches the highest point at the climax, and is followed by falling tension until the end of the story. Similar structures have been found in data stories. Based on the four major narrative categories (i.e., *Establisher (E)*, *Initial (I)*, *Peak (P)*, and *Release (R)*) that Cohn et al. [23] redefined through an analogy with traditional narrative structures, Amini et al. [2] coded the structures of 50 data videos and found the dominant sequence "E+I+PR+" that closely resembled Freytag's Pyramid. Furthermore, Dykes [27] proposed the *Data Storytelling Arc* model for telling data stories to drive business decisions based on Freytag's Pyramid. Although these studies provide summaries regarding applying Freytag's Pyramid to data stories, little systematic work exists on how to create an expressive data story with Freytag's Pyramid.

To fill this gap, this work introduces a design space that concludes story patterns in key stages of Freytag's Pyramid—the *Setting*, the *Rising-Climax*, and the *Resolution*—regarding three dimensions: (1) *narrative pattern* that provides narrative devices as initial ideas to assist in the creative process, (2) *data flow* that provides strategies for selecting and organizing individual story pieces (i.e., facts backed up by data) to achieve a narrative pattern, and (3) *visual communication* that supports selecting visual design strategies to enhance the presentation of a narrative pattern. We systematically analyzed 103 high-quality data videos using the Freytag's Pyramid structure to derive the design space. While there are various genres of narrative visualization such as *Magazine Style*, *Slide Show*, and *Annotated Charts* [56], we narrowed down our analysis to data videos. Data videos, as one kind of author-driven data stories [56], have a linear path through the story and can represent the intention of the creators. Moreover, the design of videos covers diverse visual forms that assist us in extracting abundant design strategies which could be applied to other storytelling forms.

To evaluate the effectiveness of our design space, we conducted a workshop with 25 participants. The participants were invited to create data stories using Freytag's Pyramid. They were initially asked to create story outlines without our design space first, and they later refined their story outlines and accomplished their stories using our design space. We also provided a set of method cards derived from our design space as teaching material. The method cards provided the definition, a case example, and a GIF demo for each category in our design space, and they can be found on our website ([1]). The results of the workshop showed that our design space provided an efficient framework for participants to select, organize, and fill data facts into different stages of Freytag's Pyramid while at the same time supporting the design of visual content throughout the structure.

2 Background and Related Work

In this section, we introduce how Freytag's Pyramid is adapted to data stories and review relevant research on narrative visualization.

2.1 Freytag's Pyramid for Data Stories

Narrative structure is essential to the expressiveness and the audience's perception of a story [33]. Representative narrative structures have

[1] https://pyramidstory.idvxlab.com/

- *Leni Yang is with Intelligent Big Data Visualization Lab at Tongji University and Hong Kong University of Science and Technology. A part of this work was done when Leni Yang was a visiting student supervised by Nan Cao at Tongji University. E-mail: lyangbb@connect.ust.hk.*
- *Xian Xu, and Huamin Qu are with the Hong Kong University of Science and Technology. Email: xxubq@connect.ust.hk, huamin@cse.ust.hk.*
- *Xingyu Lan, Ziyan Liu, Yang Shi and Nan Cao are with are with Intelligent Big Data Visualization Lab at Tongji University. Nan Cao is the corresponding author. Email: {xingyulan, yangshi.idvx}@tongji.edu.cn, {ziyan.liu.design, nan.cao}@gmail.com.*
- *Shunan Guo is with Adobe Research. Email: sguo@adobe.com.*

Manuscript received 21 Mar. 2021; revised 13 June 2021; accepted 8 Aug. 2021. Date of publication 29 Sept. 2021; date of current version 22 Dec. 2021. Digital Object Identifier no. 10.1109/TVCG.2021.3114774

emerged across different domains such as drama, literature, and cinematography. Some structures conclude typical story plots in certain types of stories (e.g., [13, 20, 52]). For example, Campbell [20] proposed the Hero's journey structure, which defines 17 common stages in hero myth narratives such as "The Call to Adventure" and "Refusal of the Call". Other researchers have proposed more general narrative structures, such as the three-act structure [66], anachronous narrative structures [30], and parallel narratives [6]. The Greek philosopher Aristotle first examined the basic structure of drama and concluded that a story is a series of cause-and-effect events that progress through the beginning, middle, and end [66]. Based on Aristotle's model [27], Freytag [29] developed a "pyramid-based" dramatic structure with five stages: *Exposition*, *Rising action*, *Climax*, *Falling action*, and *Resolution*. This model extends the linear beginning-middle-end structure by including the rising and falling of tension via the climax, which forms a pyramid-shaped story structure. Freytag's Pyramid has proven to be powerful and been applied in many fields such as science communications, advertising, and games [25, 28, 53, 55].

Despite the variety of other narrative structures mentioned above and structures for other genres (e.g., the inverted pyramid for news [51] and essay structures [68]), we choose to focus our research on Freytag's Pyramid for data stories, given its popularity in the visualization community discourse [24, 27, 50, 61] and in data stories [2]. In particular, Cohn [23] identified four major narrative categories (i.e., *Establisher (E)*, *Initial (I)*, *Peak (P)*, and *Release (R)*) that were refined from the stages of Freytag's Pyramid and other dramatic structures to analyze the structure of visual narratives. He further found that the canonical "E-I-P-R" sequence closely resembles Freytag's Pyramid. Following Cohn's theory, Amini et al. [2] analyzed 50 professional data videos and found that "E-I-P-R" is the most common structure. Based on Freytag's Pyramid, Dykes [27] proposed a model called *Data Storytelling Arc* for telling data stories to drive business decisions. Dykes's model includes four stages: the *Setting* stage gives background information; the *Rising insights* stage goes into deeper analysis, which serves as the build-up for the *Aha moment*; the *Aha moment* reveals the main findings and central insights; and finally, the *Solution and Next Steps*.

These studies have shed light on applying Freytag's Pyramid in data storytelling. However, little systematic guidance is available for story creators on how to apply Freytag's Pyramid to data stories considering the creation process of a data story such as how to organize story pieces (i.e., facts backed up by data) and design visual content throughout the structure. Our work fills this gap by extracting story patterns from existing data stories with that structure. Notably, the four most observed narrative stages we identified from analyzing data stories align well with Cohn's "*E-I-P-R*" structure. To better capture the progression of story flows and simplify our design space, we further merged the *Rising action* and *Climax* stages as the *Rising-Climax* stage, considering that the *Climax* stage is not a standalone point in the story model but rather a culmination or the confluence of events in the *Rising action* [2]. For example, a story may present data facts in the *Climax* stage that have a substantial difference with that in the *Rising action* to create a twist, which usually leads to a climax. The relationship between data facts in these two stages is difficult to identify if they are not analyzed together. Specifically, our design space considers the following narrative stages:

- *Setting:* This stage provides contextual information of the data story and grabs the audience's attention.
- *Rising-Climax:* This stage builds the tension of the story and shows supporting facts that lead to the climax, which presents the central insights of the story.
- *Resolution:* This stage gives conclusions and take-away messages.

2.2 Guidance on Narrative Visualizations

Researchers in the visualization community have attempted to provide theoretical bases and guidance for narrative visualizations. Segel and Heer [56] initiated the first step in constructing a design space of narrative visualization with three dimensions, namely, *genre*, *visual narrative*, and *visual structuring*. In addition, they proposed to structure data stories along a spectrum of author-driven and reader-driven approaches. They also discussed three structures—*Martini Glass Structure*, *Interactive Slideshow*, and *Drill-Down Story*—that use a mix of the author-driven and reader-driven approaches. Our work went into a deeper analysis of data videos, a typical form of author-driven data stories, that specifically use Freytag's Pyramid. We present more concrete design techniques tailored to stories with this structure. Bach et al. [8, 54] found 18 narrative patterns to provide guidance on how to achieve five general storytelling intents (i.e., argumentation, flow, framing, emotion, and engagement). Similarly, we considered the property of Freytag's Pyramid in which we adjusted and expanded Bach et al.'s narrative patterns. Specifically, we summarized narrative patterns that were tailored to achieve the storytelling intents in each stage of Freytag's Pyramid. Furthermore, we divide previous research in this area into two categories.

The first category of research focuses on the visual design strategies of narrative visualizations. Segel and Heer [56] first identified seven genres of narrative visualizations, including *Film/Video/Animation*, *Magazine Style*, *Slide Show*, *Flow Chart*, *Comic Strip*, *Annotated Chart*, and *Partitioned Poster*. Subsequently, design spaces are proposed regarding different genres (e.g., [9, 18, 39, 60, 64]). There are also studies on how to use visual embellishments to strengthen viewers' attention and memory (e.g., [10, 15, 16, 22, 32]). Studies on visual design strategies of data videos are most relevant to our work. Our analysis focused on data videos because data videos have unambiguous, author-driven story structures with narratives that eased the analysis process. Additionally, videos cover more diverse techniques of visual communication (e.g., static visual representations and animations). Amini et al. [4] proposed a taxonomy of eight animation types (e.g., creation and destruction) from over 70 videos, and they found that setup animation and pictographs in data videos can increase viewer engagement in the follow-up work [3]. Shi et al. [59] classified animation techniques in data videos on the basis of visual narrative strategies served by these animation techniques. Tang et al. [63] identified a taxonomy of animated transitions for smoothing narrative transitions in data videos. These works inspired some categories in our design space. Compared with these works, our work put more attention to story flows and the organization of data facts.

The second category of research studies how to structure narrative visualizations. Under this category, some studies surveyed structures used in data stories. Kosara [38] proposed an argument structure for data stories. Based on the analysis of time-orientated data stories, Lan et al. [40] concluded five narrative order patterns that are different from the linear order of story events. Other research under this category gave fine-grained guidance on how to improve the flexibility and smoothness of data story structures. McKenna et al. [46] identified and investigated seven flow-factors (e.g., navigation input, level of control, and navigation progress) that affect the reading experience on interactive data stories. For visualization stories in the form of a sequence of visualizations, Hullman et al. [33] summarized six transition types between adjacent visualizations. In a follow-up work, they found that people prefer hierarchical structures of visualization sequences that group visualizations with shared data properties, such as time period, measure, and spatial region [34]. To sum up, the structures used in data stories are diverse. Although these works contribute to increasing the expressiveness of narrative visualization structures, they do not fully explore the storytelling techniques for structures. We hope our work can complement the current research to increase the expressiveness of narrative visualization structures and inspire future works that investigate other structures.

3 Case Analysis

This section presents three cases from our corpus to give a preview of how the concepts in our design space (i.e., narrative stages, data flows, and visual communications) are presented in data stories, and how we deconstruct data stories in our analysis. The detailed methods of case collection and analysis are introduced in Section 4.1. The first two cases are videos and the third case is a web story, demonstrating how our

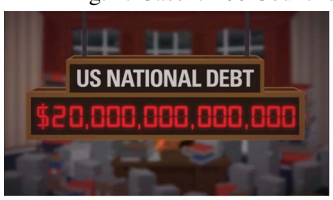
(a) **Setting:** Rosling introduces the visualization by explaining its components gradually.

(b) **Rising-Climax (i):** Rosling presents the changes of income and lifespan of 200 countries.

(c) **Rising-Climax (ii):** The climax reveals the remarkable progress that the world has made.

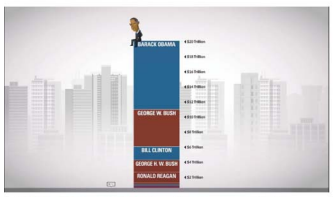
(d) **Resolution:** Rosling recaps the progress the countries achieved by replaying the animation.

Fig. 1: Case1: 200 Countries, 200 Years, 4 Minutes [49]



(a) **Setting:** The story first points out the issue of the U.S. national debt.

(b) **Rising-Climax (i):** The story shows how the debt is added up by each president's administration.



(c) **Rising-Climax (ii):** The climax presents the final amount of the debt.

(d) **Resolution:** The story predicts debt that would be brought by Trump administration.

Fig. 2: Case2: Donald Trump's $20 Trillion Problem [21]

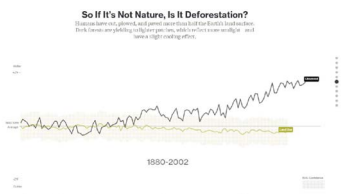
design space can be applied to a variety of data story forms. They are all from reputable sources and are popular among viewers. The three cases demonstrate the recurrent data flows and visual communications (marked in *italics*) in our design space.

3.1 Case 1: 200 Countries, 200 Years, 4 Minutes

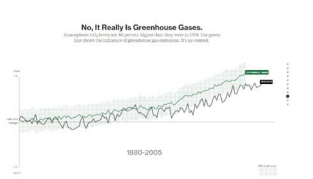
200 Countries, 200 Years, 4 Minutes was produced by BBC News in collaboration with Hans Rosling. This video tells the story of global development in 200 countries from 1810 to 2009 and has been viewed over nine million times.

Setting. Rosling commences by introducing a bubble chart that shows the life expectancy and individual personal incomes of 200 countries. As he explains the meaning of the y-axis (lifespan), the x-axis (income), and the bubbles (countries), these components gradually appear on the scene, as shown in Fig. 1. This introduction prepares the audience for understanding data insights conveyed with the visualization by *introducing the data attributes* encoded by the bubble chart. The animation of *building up the visualization* reduces the cognitive burden of the audience and helps them understand the visualization [3, 45]. **Rising-Climax.** After the visualization is set up, Rosling explores how the data *changes over time* from 1810 to 2009. While playing the animation, Rosling describes notable changes in the data and historical events that are related to the changes. When the story comes to 1948, the world was expected to witness a remarkable period wherein former colonies gained independence, got healthier, and were catching up with the development of western countries. The story then reaches its climax by presenting the fast progress of the world economy and health conditions. Rosling *speeds up* the animation to enhance the tension of the climax perceived by the viewers. **Resolution.** At the end of the story, Rosling recaps the *facts to be emphasized* in the story, that is, the marvelous progress that the countries achieved in 200 years. He does this by *replaying the visualization animation*.

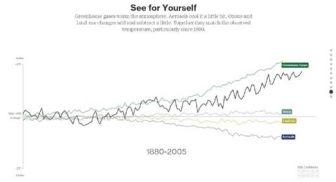
(a) **Setting:** The story introduces the global warming issue and the debate about the real cause of it.



(b) **Rising-Climax (i):** The story explores some factors, but none of them causes global warming.



(c) **Rising-Climax (ii):** The climax reveals that greenhouse is the only cause of global warming.



(d) **Resolution:** The story recaps the temperature changes brought by all factors.

Fig. 3: Case3: What's Really Warming the World? [12]

3.2 Case 2: Donald Trump's $20 Trillion Problem

Donald Trump's $20 Trillion Problem was produced by Visual Capitalist. This video reveals the significance of the United States national debt and has been viewed over one million times.

Setting. The story starts with a description of the obstacles facing the Trump administration. It then points out the most *significant data*—the colossal debt of 20 trillion dollars. An animation showing *counting numbers* that reach 20 trillion is presented to grab viewers' attention. **Rising-Climax.** The story then depicts the significance of the debt problem in detail. Specifically, the story uses blocks to represent the debt each president's administration acquired (Fig. 2). The blocks fall to the ground and stack up one by one as the story develops to show how the numbers have *added up* to the significant amount of debt. With the falling of each block, the animation becomes *increasingly intense* as each consecutive block throws up more dust from the ground and causes more intense shaking of the scene compared to the previous block. The video uses a close-up shot of each block, and the camera moves up to show new blocks appearing in the scene. Once blocks have all fallen, the camera zooms out from *the close-up shot to the overview* of all blocks to reveal the large amount of debt and bring a moment of awe. **Resolution.** Finally, the story displays how the economic plans proposed by the Trump administration would affect the U.S. debt. Based on the infrastructure investigation and tax cut plan, the *predicted debt data in the future* could increase to a number between 25.5 and 31.5 trillion dollars. A new block representing potential debt brought by Trump is *added to the existing pile of blocks* to create a visual predication of the future situation.

3.3 Case 3: What's Really Warming the World?

What's Really Warming the World from Bloomberg is a web story that explores what factors are causing global warming in the form of "scrollytelling" [57]. When it was published in 2015, it was viewed over two millions times [7]. We present this case to show the commonness of how Freytag's Pyramid is used in data videos and interactive data stories on web pages. Freytag's Pyramid underlies this data story, but it has fewer visual designs compared with data videos. To align with the diverse visual design methods presented in data videos, we also discuss how visual communication methods can enrich the original designs.

Setting. The story first provides a background paragraph about the global warming issue and the debate about what really causes it. This introduces the *context of data* which is presented in the story. To better evoke readers' concerns about the topic, the original story could include a documentary clip about natural disasters caused by global warming in the top of the webpage, which corresponds to the *related icons/cartoons/archival footage* strategy in our design space. **Rising-Climax.** The story then invites readers to scroll down to see how each factor influences climate change according to well-established climate models. The influence of each factor is presented by a line chart that compares the simulated trend of temperature changes brought by that factor and the actual temperature changes, as shown in Fig. 3.

When readers scroll down, factors such as solar and volcanoes appear one by one. However, in each chart, the simulated trend is far below the actual trend except for the last chart that shows the influence of greenhouse gases. Greenhouse gases turn out to be the only factor that has a simulated trend of temperature changes close to that of the true temperature changes. The story builds up its climax by *contrasting the influence of greenhouse gas and that of other factors*. The tension and surprise in the climax can be further enhanced by adding animation effects from our design space such as *shaking the scene*. **Resolution.** Finally, the story *re-emphasizes the fact* that greenhouse gases play a dominant role in global warming using *a summary chart* that shows the temperature changes brought by all factors.

4 Design Space

In this section, we present our study methodology of formalizing the design space and elaborate on the design space in detail.

4.1 Methodology

To understand how Freytag's Pyramid has been utilized in data stories, we surveyed and analyzed 103 data videos. Despite other forms of data storytelling, we decided to focus on data videos, given that videos can cover diverse techniques of visual communication (i.e., not only static visual representations but also animations and narratives). In addition, the results we derived from analyzing data videos can also be applied to other author-driven narrative visualizations such as slideshow [33, 56] and scrollytelling [57, 62]. We further discuss the generalizability of our study results in Section 6.2. We surveyed data videos from the lists of the previous studies [2, 56], reputable news agencies (e.g., Vox, The New York Times, and BBC News), and well-known video platforms (e.g., YouTube, Vimeo, Douyin, and Tencent Video). We used keywords such as "data-driven stories", "visual stories", and "graphical stories" to search for data videos in YouTube channels and Vimeo accounts maintained by the aforementioned news agencies. We also used keyword searching on video platforms such as YouTube, Vimeo, Douyin, and Tencent Video and subsequently processed videos with a high number of views. During this process, we ensured that the data videos we collected: (1) were data-driven, (2) contained at least one data visualization, and (3) used Freytag's Pyramid. We identified the narrative structure of each data story following the methodology of Amini et al.'s work [2]. Two authors segmented the data stories and coded the narrative stages independently. They cross-checked and refined the codes until they reached an agreement.

We found that *Setting*, *Rising*, *Climax*, and *Resolution* stages occurred in most of the data videos, which aligned with the findings in prior studies [2, 27]. We initiated the design space around on these four narrative stages. We further integrated the *Rising* and *Climax* stages as *Rising-Climax*, as stated in Section 2.1. We described the story patterns in each narrative stage from three dimensions: *narrative pattern*, *data flow*, and *visual communication*. Inspired by a prior study by Bach et al. [8], we built our design space around the *narrative pattern*, which directly implies the initial ideas of story creators and assists in the creative process. Furthermore, according to Lee and Riche [41], the creation of visual data stories involves selecting and organizing story pieces (i.e., facts backed up by data) and presenting them visually. With this in mind, we further investigated how story creators formalize their ideas into the selection and organization of story pieces (i.e., the *data flow* dimension) and the visual representations through which the narrative intent is communicated (i.e., the *visual communication* dimension).

Many of our codes were inspired by previous works. Specifically, when coding the *data flow* dimension, we referred to previous research on the organization of story pieces in sequencing visualization [33] and data comics [9]. For instance, we borrowed the transition type between consecutive visualizations, *Measure Walk*, from the study by Hullman et al. [33] when summarizing the data flow *contrasting in different measures*. In the *visual communication* dimension, we reviewed cinematography and digital storytelling techniques [5, 11, 14] to define the visual design strategies.

Note that none of the categories in each dimension was pre-defined. Rather, the categories were identified through an iterative process of mooting, merging, and refining, until exclusive narrative patterns with subordinated data flow and visual communication patterns emerged. Three authors with related backgrounds participated in the coding process. One has two years of experience in designing narrative visualizations, another has a screenwriting background and four years of experience in practice, and the remaining has two years of research experience in data visualization. They independently coded data videos using thematic analysis [17]. Codes for each narrative stage then went through at least three rounds of discussion to reach a consensus.

After we finished the initial version of the design space, we conducted several pilot studies to validate our design space with three domain experts (two graphic designers and one data journalist) to ensure that they could understand the items in our design space. After the pilot studies, we merged some related or redundant items. For example, the initial design space included *tilt*, *crane*, and *pan* to describe camera movements to shoot 3D visualizations. However, the designers suggested that these items are all low-level techniques for the same purpose of enhancing the audience's perceptions of the depth of the scene. Thus, we merged them into one called *showing depth with camera*. We also renamed some items to increase the clarity and conciseness of our design space with the help of experts. For example, the visual design strategy *Related icons/cartoons/archive footage* was changed from "Related icons/cartoons/historical images and video clips" because historical images and video clips are normally collectively called "archive footage" by designers. In the final version of our design space, each item is either frequently seen in the collected videos or is less frequent but considered as useful by experts and related literature. Related statistics can be found on our website.

4.2 Design Space Overview

This section gives the definition and functions of each dimension in our design space. Following this, the individual items in each narrative stage are discussed. Notably, the *data flow* and *visual communication* dimensions are centered around the *narrative pattern* dimension and accordingly provide strategies for each narrative pattern.

4.2.1 Dimension I: Narrative pattern

According to Bach et al. [8], a *narrative pattern* is "a low-level narrative device that serves a specific intent," which "can be used individually or in combination with others." *Narrative patterns* in data stories can further be interpreted as the counterparts to the classical plots of literary stories. For instance, the climax in a literary story often accompanies a plot of a fierce confrontation between the protagonist and the antagonist. Likewise, in data stories, a climax can be built by showing contrasting data facts (i.e., the narrative pattern *showing contrast*).

4.2.2 Dimension II: Data Flow

As stated in Section 4.1, we intend to describe how story authors select and organize story pieces (i.e., facts backed up by data) to convey their narrative intentions through the *data flow* dimension, which can be formally defined as *strategies for selecting and organizing data facts to apply narrative patterns*. Unless otherwise noted, the related low-level analysis tasks of a data fact can be any type (e.g., trend, value, and difference [67]), as we observed from our corpus. We identified two kinds of data flows from our analysis. One describes the characteristics or information presented in individual data facts. For example, the data flow *individuals behind data* for the *presenting concrete characters* narrative pattern in the *Setting* stage involves data facts that are related to persons. The other illustrates the connection between a group of data facts. For example, the data flow *contrasting in different granularity* for the narrative pattern *showing contrast* in the *Rising-Climax* stage indicates the organization of a set of comparative data facts in different granularity for contrast analysis. Although we intend to provide guidance on reflecting a narrative pattern through *data flow*, in practice story authors can either have a narrative pattern in mind and search for the corresponding data facts, or they could start from the characteristics of data facts they find valuable to see what narrative patterns can be incorporated into the story.

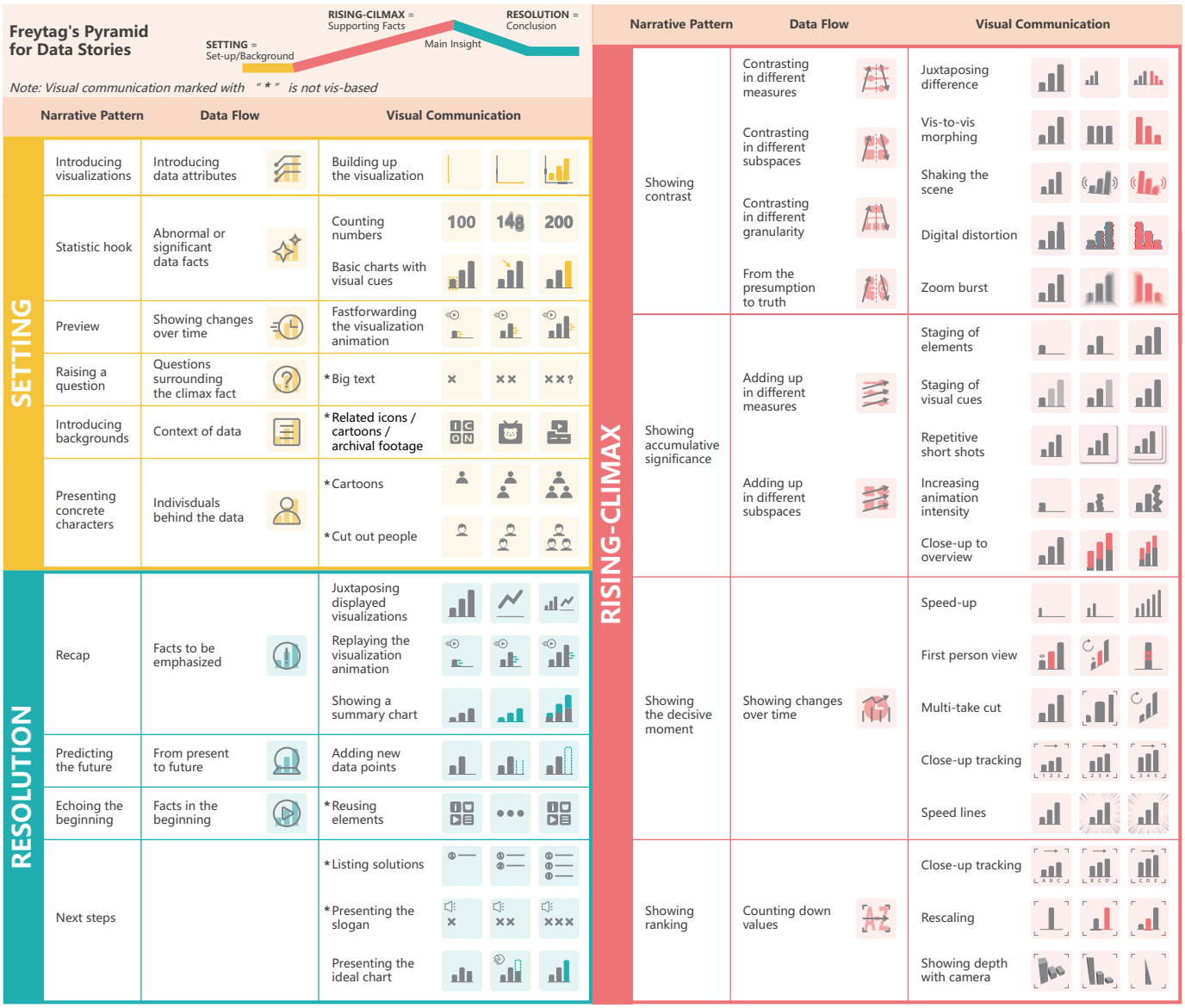


Fig. 4: Design space for creating data stories with Freytag's Pyramid. The illustration of Freytag's Pyramid refers to the idea [26].

4.2.3 Dimension III: Visual Communication

The *visual communication* dimension describes how *narrative patterns* are displayed to the audience, which is formally defined as *visual design techniques to enhance the presentation of narrative patterns*. As our data stories are collected by analyzing data videos, we can cover a wide range of visual design techniques such as the animation effects, camera movements, and editing techniques used in the videos. Specifically, our *visual communication* dimension mainly covers the selection of multimedia content such as presenting *cut-out people* for the narrative pattern *presenting concrete character* and animation designs such as *shaking the scene* for the narrative pattern *showing contrast*. Although visualizations are one of the main visual content types in data stories, we do not propose guidelines for selecting visualizations based on the types of data used in the story, which has already been extensively investigated by the existing literature [48].

4.3 Setting

We identified six narrative patterns to be applied in the *Setting* stage.

Introducing visualizations. This pattern explains the visual encodings of the visualizations to guide the audience through how they can read the charts in a story. When the story contains uncommon or new visualization designs, providing such an introduction is important. Otherwise, the audience may flounder and give up listening [42]. Data flow. The corresponding data flow of this narrative pattern is *introducing data attributes*. For example, in the first case (Fig. 1), the narrator, Rosling, introduces the two attributes, life expectancy and income per person, which are represented in the x-axis and y-axis of the chart, respectively. Visual communication. The visual communication technique *building up the visualization* presents each part of the visualization progressively when the corresponding design is explained. This technique can reduce the cognitive burden of the audience [45], facilitate their comprehension of the visualization, and increase their engagement [3].

Statistic hook. Statistic hook is a storytelling technique in which interesting numbers or visualizations are presented at the beginning of the story to grab attention. Data flow. The corresponding data flow is to select *abnormal or significant data facts* that reflect the importance of the topic of the story. Generally, statistic hooks are unexpected findings such as a sudden spike or dip in time-series data [27] or summary statistics (e.g., sum and maximum values) Visual communication. When presenting statistic hooks, the designer may use *counting number*, an animation showing a number counting up instead of presenting only the final value to arouse the audience's interest in the result. For a statistic hook with relatively more complex findings such as a noticeable trend found in temporal data, designers can use *basic charts with visual cues* that the audience typically finds familiar and highlight the findings with visual cues [37].

Preview. Preview is a foreshadowing storytelling technique that gives a hint in advance of what is to come later in the story [69]. This narrative pattern is often used by movie trailers for establishing an expectation within the audience on how the story will develop [69]. Data flow. Preview is commonly observed in stories that show changes in time-series data. These stories usually start by describing the general change of data from the starting point to the end. They then go back to the starting point and describe the details chronologically. Accordingly, we name the data flow as *showing changes over time*. Visual communication. The visual communication used is *fastforwarding the visualization animation*, which goes through visualizations from the beginning of time to the end quickly with animation.

Raising a question. This narrative pattern poses a question directly to the audience to evoke their curiosity and induces suspense for the subsequent exploration of the answer [8]. Data flow. The corresponding data flow is *questions surrounding the climax fact*, with the answer to the question being revealed at the *Climax* stage of the story. By an analogy with literary stories, the question can be compared to the main conflict of a story, and the climax should resolve the conflict [44]. Visual communication. Story authors can use *big text* on a full screen to emphasize the question. The occurrence and disappearance of the text also serves as a transition signal of the story progressing to the next stage (i.e., from *Setting* to *Rising-Climax*).

Introducing backgrounds. This narrative pattern provides the contextual information of the story, such as its motivation or its general topic. Data flow. In data stories, the contextual information often relates to how the data is collected, who are the data subjects, and what tasks the data analysis is for. We refer to these types of story pieces as the data flow *context of data*. Visual communication. The commonly identified visual communication approach to convey the abstract information is to use *related icons/cartoons/archive footage*. Moreover, when the data is related to some historical or trending events, the cartoons and archive footage (e.g., news clips) are often used for event recap.

Presenting concrete characters. Presenting concrete characters allows the story to begin with a personal view, and this can include telling an anecdote. The characters can also be imaginary figures. Such a pattern can enhance the engagement of the audience as they can relate better to people than to abstract data [9, 27]. Data flow. The corresponding data flow is the *individuals behind the data*—facts related to data about people. For instance, the video *All the Medalists: Men's 100-Meter Freestyle* starts with the description that Nathan Adrian won gold in the 100 meter freestyle race in the 2012 Olympics. Visual communication. The visual communication techniques include *cut-out people* and *cartoons*, both of which provide the audience with concrete images of the characters.

4.4 Rising-Climax

We identified four narrative patterns in the *Rising-Climax* stage.

Showing contrast. Showing contrast presents data facts that are substantially different, which creates a plot turn that leads to the climax. Data flow. The common data flows for *showing contrast* are *contrasting in different measures* and *contrasting in different subspaces*. The former presents the contrast of different measured values for the same categories in the data, whereas the latter depicts the contrast of the measured values across different categories. A measure is a numerical data column on which certain aggregations (e.g., SUM and AVG) can be performed [33]. A subspace is a subset of the data created by using filters for any data dimensions that contain qualitative, categorical information [67]. For example, a story may compare the increase of average income and the decrease of happiness index for a certain country, with the average income and happiness index being two measures of the dataset with opposite trends. Additionally, it can compare the average income across different countries, where the data of each country is a subspace of the dataset. Another data flow for showing contrast is *contrasting in different granularity* in a hierarchical data set, such as comparing the average income worldwide and the average income of a specific country. Besides contrasting data facts, storytellers can present data that is contrary to common beliefs to form a sense of contrast, which we refer to as *from the presumption to truth*. Visual communication. We observed five visual designs that enhance the audience's feeling of surprise when presenting the contrast. *Juxtaposing difference* presents visualizations of data facts side by side while *vis-to-vis morphing* demonstrates the transformation from the visualization of one data fact into another. *Shaking the scene* adds vibrations to the view during the climax as an analogy to the feeling of shock. *Digital distortion* is from the glitch art [47] that adds digital noises to the scene to mimic a crash of the filming devices which overturns viewers' expectations. *Zoom burst* makes the surrounding areas of the climax scene blurred while keeping the center sharp as the scene bursts toward the audience.

Showing accumulative significance. Showing accumulative significance increases the intensity of a story by showing similar data facts repetitively. Data flow. It is often demonstrated through *adding up in different measures* or *adding up in different subspaces* to present similar data facts from different perspectives. Visual communication. We found five visual communication techniques that help increase the tension of the story. *Staging of elements* lets the visual component (i.e., the visualization in part or whole) representing each data fact appear progressively in the scene as the story develops. *Staging of visual cues* employs visual cues (e.g., glow and arrow) to highlight each visual component as the story progresses. These two techniques increase the audience's excitement progressively by showing cumulative elements in a static shot [5]. *Repetitive short shots* swiftly switches between visualizations with the same visual encoding to present a series of similar facts in different facets of the dataset to create an intense story rhythm [11]. *Increasing animation intensity* augments the exaggeration of the animation as the story progresses. For example, in the second case (Fig. 2), when the blocks representing data facts fall one by one, the consecutive block raises more dust and causes a more intensive shaking of the scene compared to the previous one. *Close-up to overview* consists of a series of close-up views on each visual component and an overview of all components displayed on the screen. The audience's anticipation increases as each close-up view shows up and the final overview brings a moment of awe.

Showing the decisive moment. This narrative pattern triggers the climax by bringing the audience to an important and special moment. Such a moment can either be a semantically meaningful time (e.g., the 2008 financial crisis) or a point when a qualitative change occurs (e.g., when China became the top export country in 2009). Data flow. The corresponding data flow is *showing changes over time*. For example, in the first case (Fig. 1), Rosling introduces the marvelous progress of the countries achieved in 200 years by showing how their life expectancy and income per person increased over time. Visual communication. We found five visual communication techniques that enhance the feeling of excitement and intensity at the decisive moment. *Speed up* increases the animation speed in a short period of time right before the climax. This acts as a signal of the upcoming changes by increasing the pace of the story and raising an expectation within the audience. *First person view* puts the visualization in a 3D scene and makes the data trend like a roller coaster track that the reader then rides from one end to the other. When the audience's view moves up and down following the trend, they have an immersive experience with the fall and rise of data. *Multi-take cut* demonstrates the short period right before the climax when various camera angles around the chart add to a sense of movement or frenetic excitement [65]. *Close-up tracking* moves the camera horizontally to gradually reveal the data at the next time point, which increases the audience's anticipation of the upcoming event and the feeling of excitement when the climax is revealed. *Speed lines* adds radial, twinkling lines when the decisive moment is coming, which was originally used in manga art to enhance the intensity of an action moment [31].

Showing ranking. Showing ranking is a narrative approach of sorting things according to some criteria and revealing them one by one to establish a sense of suspense. Data flow. The corresponding data flow *counting down values* typically includes data facts of story subjects from low to high ranking. Visual communication. We identified three visual communication techniques. *Close-up tracking* moves the camera horizontally to show items lined up in a rank chart from the lowest to

highest rank. *Rescaling* gradually brings the items to the audience's sight and changes the scale of the canvas to fit the size of the new coming item. *Showing depth with camera* is used when the rank chart is 3D. It applies camera movement techniques such as crane, tilt, and pan to shoot each item. Thus, it enhances the audience's perceptions of the size of each item to evoke a sense of amazement [11].

4.5 Resolution

In the *Resolution* stage, we identified four narrative patterns.

Recap. Recap helps reinforce the audience's memory of the main messages communicated in the story. Data flow. Its corresponding data flow is *facts to be emphasized*, which recalls the data facts that are indispensable to the conclusion of the story. Visual communication. We identified three visual communication techniques that aim to help the audience quickly review the story. *Juxtaposing displayed visualizations* places key visualizations side by side on the screen again. *Replaying the visualization animation* provides a fast review of major changes of data over time in the story. *Showing a summary chart* introduces a chart to summarize the findings scattered in the story.

Predicting the future. Storytellers apply this narrative approach to provide a glimpse of future data trends. Data flow. The data flow *from present to future* includes data facts of the predicted data. Visual communication. Data stories often present the predicted data by *adding new data points* to the visualization, which allows the audience to investigate the future data in the context of historical data.

Echoing the beginning. Echoing the beginning revisits story content presented at the beginning, which is a common literary narrative technique [1]. Data flow. The data flow employed here is *facts in the beginning*, with the data facts at the beginning of the story reappearing at the end. Visual communication. The visual design is usually kept similar to what is displayed at the beginning, which we refer to as *reusing elements*; thus, the audience can easily make connections between the beginning and the end.

Next steps. This narrative pattern provides solutions to the discussed problem and encourages the audience to take action. Data flow. It requires the storytellers to use the external domain knowledge beyond data to provide recommendations or opinions. The data flow can vary from case to case; therefore, we left the data flow of this narrative pattern blank in our design space. Visual communication. The visual communication techniques include *listing the solutions* and *presenting the slogan*, which demonstrate the solutions clearly and encourage the audience to take action. Another technique is *presenting the ideal chart*, which demonstrates the target data, such as a quarterly sales target, to the audience for stating the goal.

5 Workshop

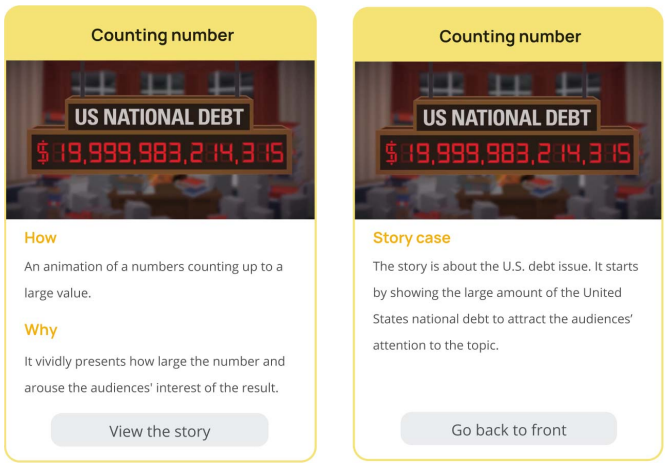
We ran a workshop investigate: G1) the usefulness of the design space; G2) how the design patterns in our design space are used by participants; and G3) whether the design space is easy to use.

5.1 Participants

We recruited 25 participants (13 females) aged between 19 and 30 by advertising on online social media platforms. The participants included college students, researchers, and professionals with backgrounds including design, art, journalism, computer science, bioengineering, and urban planning. Their proficiency in creating data stories also differed greatly, ranging from no experience (56%), up to 1 year (28%), 1-2 years (8%), to 2-5 years (8%). Before the workshop, 24% of the participants had never heard of Freytag's Pyramid, 60% knew similar concepts, 8% heard of this structure, 4% were familiar with this structure, and 4% often used it.

5.2 Teaching Materials and Data

During the workshop, we provided a set of method cards as the teaching material of our design space. All method cards are available on our website. The design of the method cards was inspired by the Napa Cards [43] and IDEO Method Cards [35]. The intention was to facilitate participants' understanding of each category of our design space with examples. The method cards are organized by narrative stages and narrative patterns. Each method card illustrates a type of data flow or visual communication from three perspectives (as shown in Fig. 5): how to use it, why it is useful, and examples of usage. The datasets we used in the workshop for participants to create data stories related to two topics of general interest: *top 10 killer diseases in the world* and *the obesity problem in the world*. Given the diversity of data facts we observed in the two datasets, we hypothesized that they could lead to various stories. Following the suggestions from previous studies [2, 9, 59], instead of giving participants only the raw data, we also provided various categories of pre-extracted data facts (e.g., outlier, rank, and distribution [67]) and their corresponding visualizations to focus participants' attention on planning narrative strategies.



(a) (b)

Fig. 5: Example of a method card for explaining a category in the design space: (a) front and (b) back.

5.3 Procedure and Analysis Approach

The workshop was conducted offline and lasted about four hours. We began with a 30 minute introduction to narrative visualization, Freytag's Pyramid, and the storyboard for planning a story. After the introduction, participants were divided into 12 groups of two to three people on a voluntary basis. Each group was asked to create stories with Freytag's Pyramid based on either the given datasets or any online dataset. To thoroughly investigate how our design space could help story creators construct data stories, we gave each group of participants 40 minutes to explore the dataset and create a story outline with key points in each stage of Freytag's Pyramid before using the design space. Each group of participants wrote down their story outlines. Afterward, we gave a 15 minute demonstration on how to use our design space and method cards. Participants were then asked to refine their story outlines using our design space freely. The refinement stage took about 20 minutes. Once a group finalized their story outline, we instructed them to sketch a storyboard to present the visualizations of their stories (all storyboards can be found on our website). Participants could write down their visual design ideas that could not be demonstrated through sketching, such as transitions between scenes and animation effects. The sketching process lasted around 2.5 hours. The workshop was video-recorded. Finally, we asked each group to introduce their story content and design ideas.

After the workshop, participants filled a questionnaire regarding the usefulness and usability of our design space using a 7-point Likert scale. We also conducted semi-structured interviews with the participants for their feedback on using Freytag's Pyramid and the design space for data stories once a group finished drawing. Some interviews were conducted after the workshop through online meetings. Interview questions were centered around: (1) their opinions on using Freytag's Pyramid for data stories; (2) the difficulties they encountered when drafting the story outlines without the design space; (3) whether and, if applicable, how our design space helped improve their stories; and (4) their comments on the usefulness and clarity of the design space

and our teaching materials. Each interview took approximately 10-15 minutes, and audio was recorded for subsequent analysis. Two authors independently coded the story outlines and storyboards collected from participants to investigate what items from our design space were used by each group. Specifically, we classified frames of each storyboard into different stages of Freytag's Pyramid and coded narrative patterns, data flows, and visual communications that were used in each stage. Based on the coding results, we counted the frequency of each category in our design space (the number of groups who used it), which can be found on our website. We also compared the story outlines participants created with and without using the design space.

5.4 Results

This section discusses the workshop and interview results.

G1: Design space usefulness

In general, participants found that following certain structures made the design process smooth, and they acknowledged that Freytag's Pyramid provided them with a compelling way to communicate their insights compared with reporting data in a flat style. Some participants also mentioned that stories with such a structure were more acceptable and memorable for the viewers and had potential for creating viral videos. Participants also agreed that our design space was useful 1 7 ($M = 6.08, SD = 1.82$) for creating stories with Freytags' Pyramid. Specifically, the usefulness of the *narrative pattern* dimension was rated 1 7 ($M = 6, SD = 1.81$), the *data flow* dimension was rated 1 7 ($M = 5.72, SD = 1.67$) and the *visual communication* dimension was rated 1 7 ($M = 5.88, SD = 1.71$). Participants expressed their appreciation on the structure of our design space, which they believed provided systematic guidance in the story creation process (G8P1: *"It provides a clear plan for me to construct my data stories."*) and was easy for beginners to use (G4P1: *"For beginners like me, it serves as a set of story templates..."*).

We examined the usefulness of our design space by comparing the story outlines that participants made before and after using the design space. Generally, 11 out of the 12 groups of participants changed their story outlines. Among the 11 groups, one group failed to give their story outline before using the design space. The changes participants made can be classified into the following aspects:

Enriching the story plots with new inspirations. We found that six groups added narrative patterns including *preview* (1 out of 6 groups), *presenting concrete characters* (1), *statistic hook* (2), and *raising a question* (3) in the *Setting* stage of their stories to make the opening of their stories more engaging. For example, group 5's story was about the severity of neonatal conditions in Africa. They originally planned to start with a direct introduction to the topic. Inspired by the *statistic hook* in our design space, they added a multiple line chart showing trends of the number of deaths caused by different diseases. The chart presented a notable decreasing trend in the number of deaths caused by neonatal conditions, which could lead the audience's attention to this particular disease. They then asked a question "Is the number of deaths caused by neonatal conditions decreasing in every region around the world?" to evoke the curiosity of the audience. We also found that participants added narrative patterns in the *Resolution* stage of their stories to emphasize take-away messages. Specifically, we found that one group added the narrative pattern *recap* to re-emphasize the issues they discussed in their story, and seven groups added the narrative pattern *next steps* at the end of their stories. For example, group 6 told the story about the severity of obesity around the world, and at the end, they encouraged the audience to choose a healthy lifestyle.

Modifying the story from one with a flat style to one that builds up an outstanding climax. We found that six groups changed their *Rising-Climax* stage to use narrative patterns in our design space, while stories of other groups were already aligned with our design space. Specifically, three groups used the narrative pattern *showing contrast*, two used *showing accumulative significance*, and one used *showing ranking*. Among these groups, four groups added more data facts to build up a stronger climax. For example, group 4's story was about the rise of obesity in China. Before using our design space, they only discussed the increasing trend of Chinese adults with obesity. After seeing the narrative pattern *showing contrast* and the data flow *contracting in different subspaces* in our design space, they added the trend of obesity in other countries before showing that in China, which presented a significant contrast that China had a much steeper upward trend than other countries and resulted in a stronger climax. Furthermore, we found that two groups removed some data facts and kept related ones that formed narrative patterns in our design space to make the key messages in the climax more outstanding (G3P2: *"We tried to put all findings into the story, but to clearly convey the climax, we removed irrelevant ones."*).

Adding new visual design ideas. Some groups also thought of visual communication strategies to be used when deciding their story outlines. We found that four groups explicitly wrote down visual communication strategies they never conceived before learning the design space. This indicated that our design space provided new visual design ideas for participants. Other groups did not write down their original plans of visual designs when preparing their story outlines but told us that the design space did provide them with new design ideas, and they used them when sketching their storyboards (G11P1: *"I never thought of gradually increasing the animation intensity to build up the tension of the story."*; G12P1: *"Most visual communication techniques are new to me."*).

G2: Design space usage

The storyboards from participants were diverse in their content. For example, from the seven storyboards that tell stories about the dataset *top 10 killer diseases in the world*, we saw a wide range of topics, from a comparison of the worldwide number of death caused by different diseases to a focus on a particular disease such as neonatal conditions and Alzheimer's disease. We found that every narrative pattern was used. As most narrative patterns have only one corresponding data flow, the data flows were also mostly used, with only *contrasting in different granularity* and *from the presumption to truth* for the narrative pattern *showing contrast* being exceptions. Participants who used this narrative pattern suggested that they selected other data flows because they found those brought story topics that were more of their interests or brought a stronger climax. In addition, we found that most visual communications (24/34) were used. When we asked why participants did not choose alternative visual communication strategies, they suggested that they took the aesthetic aspects of their stories and the difficulties to achieve a strategy into consideration. Interestingly, we found that some items from our design space were used more often together than others. For example, five groups used both *statistic hook* and *raising a question* in the *Setting* stage. Two groups used both *staging of elements* and *staging of visual cues* for the narrative pattern *showing accumulative significance*. To know whether these findings identify general strategies or whether there are other general strategies, further studies can conduct workshops with more participants and collect more data. Unexpectedly, two groups creatively combined different narrative patterns in the *Rising-Climax* stage and formed stories with two climaxes: a minor warm-up one followed by a core climax. It showed that our design space could enhance the expressiveness of story structures.

G3: Ease of use

The usability of our design space was positively rated 1 7 ($M = 6.04, SD = 1.79$). Participants praised the clarification of visual communication approaches by using a sequence of key frames at the right side of each category, as shown in Fig. 4. They also appreciated the examples illustrated on the method cards on the website (G7P1: *"Looking at the examples on the website, I can better imagine what my design would look like."*). Some participants also mentioned how our design space facilitated the team communication in the creative process (G6P1: *When I want to tell the design I want to achieve, I just need to point at the corresponding item in the design space.*). For improvements of our method cards, some participants desired a mobile version of the website, and others wanted to see more examples.

6 Discussion

In this section, we discuss the implications for applying narrative structures to data stories as well as our limitations and future work.

6.1 Design Space for Freytag's Pyramid

Our work applied Freytag's Pyramid from the field of narratology to narrative visualizations. We analyzed 103 data videos with this structure and proposed a design space that characterizes the narrative patterns, data flows, and visual communication techniques that help construct data stories. When formalizing the design space, we found that many techniques in data stories can be mapped to traditional storytelling forms. For example, in film and literature, the climax is often built through a fierce confrontation, while in data stories, the climax is usually built by showing contrasting data facts. However, we noticed that creating data stories faces unique challenges compared with traditional stories. For example, data as the main "character" of data stories is more abstract than traditional characters, which makes crafting a meaningful and understandable narrative structure difficult. In our workshop, some participants found selecting data facts from the dataset and arranging them into an engaging plot difficult before they were given the design space. Our design space alleviates this problem by guiding the selection and organization of data facts.

6.2 The Generalizability of the Design Space

This section discusses the generalizability of our design space. We start by discussing how our design space can be applied to other genres of narrative visualization [56] in addition to data videos. First, our resultst can be easily generalized to Magazine Style and Slide Show because these two forms use a linear storyline where Freytag's Pyramid can be applied. The visual design strategies in our design space can be smoothly applied in these two genres with visual elements such as text, images, and animation that are covered by data videos. Similarly, the design space can be applied to scrollytelling [57, 62], an emerging web-based author-driven storytelling form. Second, Comic Strip and Flow Chart, which also follow a linear storyline, can use our design space. However, they present their storylines using carefully designed layouts. Therefore, future work is needed to expand our design space, taking layout into consideration. On the other hand, applying our results to Annotated Chart and Partitioned Poster is challenging because they do not have a clear sequence of frames in their representations.

Next, although our study results come from data videos with Freytag's Pyramid, our design space can be applied in combination with less rigid storytelling structures, such as Martini Glass Structure and Drill-Down Story [56] that allow users to choose among multiple storylines, or communicate author-intended messages prior to prompting the user to freely explore the visualization. Storytellers can apply our design space to design a specific storyline or the author-intended messages in these structures. Additionally, our design space can be used together with other rhetorical techniques for story structures, such as the manipulation of the temporal progress. For example, Freytag's Pyramid can be applied to time-orientated data stories by giving a preview of the latest data fact in the *Setting* stage and then going into the *Rising-Climax* stage by introducing facts in temporal order, with the climax being the most important moment. This structure aligns with the "trace-back" pattern of the time-oriented data story summarized in Lan et al.'s study [40], wherein five narrative orders that are different from the linear order of the story events are derived. Future work may consider conducting a deeper analysis of how the temporal order of data facts can be used together with Freytag's Pyramid by collecting data stories with the two structures.

6.3 How Designers Structure Data Stories

When selecting data facts for the narrative structure, designers consider the emotional value of the data apart from their logic connections, as participant G3P2 noted, "*We tried to find which data facts are most surprising to the audience first.*" Similarly, some participants mentioned that they arranged data facts in a way that built anticipation or created suspense. We also found that the participants generally put much effort into attracting audiences at the beginning of the story. For example, five groups used both *statistic hook* and *raising a question* in the *Setting* stage. According to the participants' feedback, they felt that reading data could sometimes be less interesting to the audience. Therefore, raising the audience's interest was their primary concern. Moreover, we found that some participants extended the basic pyramid structure and created nested story structures in the workshop. Specifically, two groups designed two climaxes in their stories by embedding a small pyramid structure into a larger one that formed the overall story.

6.4 Implications for Automatic Data Story Generation

Researchers have developed tools to generate data stories automatically [36, 58, 70]. Existing tools can generate data stories with logical connections between data facts [58], but how to generate a data story that has an expressive structure remains unclear. Although the design space may provide some directions on the automatic generation of a structured data story, technical challenges still exist according to our observations. First, extracting a core message and its supporting data facts from a dataset to create a meaningful storyline is challenging. It requires the machine to understand the semantics information in the data. Second, selecting the most appropriate visual communication techniques is non-trivial. Participants in our workshop considered several subjective factors when determining the design. For example, they decided which design method was most "suitable" or "beautiful" for the scene. They also considered the coherence of different visualizations and scenes. Understanding aesthetics is challenging for computers. Third, evaluating the quality of the generated data stories is not easy because the effectiveness of data stories is mostly subjective and is difficult to quantify. Given these challenges, we propose that future automatic tools should take data interpretation, computational aesthetics (e.g, [71]), and computational semantics into the design considerations.

6.5 Limitations and Future Work

Our design space is derived from a finite corpus of 103 data videos. It can be expanded by collecting more cases and including more design factors. For instance, our design space can be further expanded by mapping narrative patterns of stories onto visualization tasks. We observed some connections between narrative patterns and visualization tasks, including *showing the decisive moment* and visualization task *find extreme* or *find anomalies* [19] in time-series data. Additional works are required to collect more cases to identify more possible relationships. We believe the strategies concluded in the design space can be applied to other author-driven linear narrative visualizations, but there may be potential challenges specific to other scenarios that require further exploration. In addition, although participants in the workshop were satisfied with using Freytag's Pyramid for data stories, this structure is only one of all possible data story structures. Future work can assess and compare the effectiveness of different data story structures. Moreover, we expect that our study and prior studies for data story authoring tools (e.g., the tool for scrollytelling [62]) are complementary. For instance, by making the story structure explicit and adding design recommendations for each stage of the structure, data story authoring tools could streamline the creation process. However, the actual effectiveness and the challenges of realizing such a mechanism require future work.

7 Conclusion

We propose a design space for applying Freytag's Pyramid to data stories through an analysis of 103 data videos with this structure. The design space concludes narrative patterns that serve the narrative intents of each stage in Freytag's Pyramid. For each narrative pattern, the design space further summarizes the strategies for selecting and organizing data facts to reflect the narrative patterns, and the visual design techniques to support the presentation of the narrative patterns. The workshop results indicate that our design space provides guidance with a clear framework to facilitate the creation of data stories with Freytag's Pyramid. We hope our design space can help designers, data analysts, journalists, and data enthusiasts to convey their data insights with engaging stories as well as give implications for the development of data story authoring tools.

Acknowledgments

Nan Cao is the corresponding author. This work was supported in part by the National Natural Science Foundation of China (62072338) and NSF Shanghai 20ZR1461500.

REFERENCES

[1] G. Adamo. Beginnings and endings in novels. *New Readings*, 1:83–104, 1995.

[2] F. Amini, N. Henry Riche, B. Lee, C. Hurter, and P. Irani. Understanding data videos: Looking at narrative visualization through the cinematography lens. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 1459–1468. Association for Computing Machinery, New York, NY, USA, 2015.

[3] F. Amini, N. H. Riche, B. Lee, J. Leboe-McGowan, and P. Irani. Hooked on data videos: assessing the effect of animation and pictographs on viewer engagement. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pp. 1–9, 2018.

[4] F. Amini, N. H. Riche, B. Lee, A. Monroy-Hernandez, and P. Irani. Authoring data-driven videos with dataclips. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):501–510, 2016.

[5] D. Arijon. *Grammar of the film language*. Silman-James Press, 1991.

[6] L. Aronson. *The 21st century screenplay: a comprehensive guide to writing tomorrow's films*, vol. 2011. Silman-James Press, 2010.

[7] S. Award. Shorty awards:what's really warming the world? finalist in data visualization. https://shortyawards.com/8th/whats-really-warming-the-world, 2010. Accessed: 2021-02-01.

[8] B. Bach, D. Stefaner, J. Boy, S. Drucker, L. Bartram, J. Wood, P. Ciuccarelli, Y. Engelhardt, U. Koeppen, and B. Tversky. Narrative design patterns for data-driven storytelling. In *Data-Driven Storytelling*, pp. 107–133. CRC Press, 2018.

[9] B. Bach, Z. Wang, M. Farinella, D. Murray-Rust, and N. Henry Riche. Design patterns for data comics. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 1–12, 2018.

[10] S. Bateman, R. L. Mandryk, C. Gutwin, A. Genest, D. McDine, and C. Brooks. Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 2573–2582, 2010.

[11] B. Block. *The visual story: Creating the visual structure of film, TV and digital media*. CRC Press, 2013.

[12] Bloomberg. What's really warming the world? https://www.bloomberg.com/graphics/2015-whats-warming-the-world/, 2015. Accessed: 2021-02-01.

[13] C. Booker. *The seven basic plots: Why we tell stories*. A&C Black, 2004.

[14] D. Bordwell, K. Thompson, and J. Smith. *Film art: An introduction*. McGraw-hill Education New York, 2017.

[15] R. Borgo, A. Abdul-Rahman, F. Mohamed, P. W. Grant, I. Reppa, L. Floridi, and M. Chen. An empirical study on using visual embellishments in visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2759–2768, 2012.

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

[17] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.

[18] M. Brehmer, B. Lee, B. Bach, N. H. Riche, and T. Munzner. Timelines revisited: A design space and considerations for expressive storytelling. *IEEE Transactions on Visualization and Computer Graphics*, 23(9):2151–2164, 2016.

[19] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, 2013.

[20] J. Campbell. *The hero with a thousand faces*, vol. 17. New World Library, 2008.

[21] V. Capitalist. Donald trump's $20 trillion problem. https://www.youtube.com/watch?v=tgpfswDHA_Y, 2017. Accessed: 2021-02-01.

[22] N. Cawthon and A. V. Moere. The effect of aesthetic on the usability of data visualization. In *Proceedings of the International Conference Information Visualization*, pp. 637–648. IEEE, 2007.

[23] N. Cohn. Visual narrative structure. *Cognitive science*, 37(3):413–452, 2013.

[24] K. Cole, Nussbaumer. The structure(s) of story. http://www.storytellingwithdata.com/blog/2020/5/21/the-structures-of-story, 2020. Accessed: 2021-03-08.

[25] J. E. Cutting. Narrative theory and the dynamics of popular movies. *Psychonomic Bulletin & Review*, 23(6):1713–1743, 2016.

[26] B. Dykes. Data storytelling: Separating fiction from facts. https://www.forbes.com/sites/brentdykes/2016/07/13/data-storytelling-separating-fiction-from-facts-2/?sh=786bcab73a7c, 2016. Accessed: 2021-02-01.

[27] B. Dykes. *Effective Data Storytelling: How to Drive Change with Data, Narrative and Visuals*. John Wiley & Sons, 2019.

[28] S. J. ElShafie. Making science meaningful for broad audiences through stories. *Integrative and comparative biology*, 58(6):1213–1223, 2018.

[29] G. Freytag. *Technique of the drama: An exposition of dramatic composition and art*. S. Griggs, 1895.

[30] G. Genette. *Narrative discourse: An essay in method*, vol. 3. Cornell University Press, 1983.

[31] H. Hayashi, G. Matsuda, Y. Tamamiya, and K. Hiraki. Visual effect of" speed lines" in manga: An experimental study on spatial attention. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, 20(1), 2013.

[32] J. Hullman, E. Adar, and P. Shah. Benefitting infovis with visual difficulties. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2213–2222, 2011.

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

[34] J. Hullman, R. Kosara, and H. Lam. Finding a clear path: Structuring strategies for visualization sequences. *Computer Graphics Forum*, 36(3):365–375, 2017.

[35] IDEO. Method cards. https://www.ideo.com/post/method-cards, 2003. Accessed: 2021-02-01.

[36] Y. Kim, K. Wongsuphasawat, J. Hullman, and J. Heer. Graphscape: A model for automated reasoning about visualization similarity and sequencing. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 2628–2638, 2017.

[37] H.-K. Kong, W. Zhu, Z. Liu, and K. Karahalios. Understanding visual cues in visualizations accompanied by audio narrations. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 1–13, 2019.

[38] R. Kosara. An argument structure for data stories. In *Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Short Papers*, pp. 31–35, 2017.

[39] X. Lan, Y. Shi, Y. Zhang, and N. Cao. Smile or scowl? looking at infographic design through the affective lens. *IEEE Transactions on Visualization and Computer Graphics*, 27(6):2796–2807, 2021.

[40] X. Lan, X. Xu, and N. Cao. Understanding narrative linearity for telling expressive time-oriented stories. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 1–13, 2021.

[41] B. Lee, N. H. Riche, P. Isenberg, and S. Carpendale. More than telling a story: Transforming data into visually shared stories. *IEEE Computer Graphics and Applications*, 35(5):84–90, 2015.

[42] S. Lee, S.-H. Kim, Y.-H. Hung, H. Lam, Y.-a. Kang, and J. S. Yi. How do people make sense of unfamiliar visualizations?: A grounded model of novice's information visualization sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):499–508, 2015.

[43] B. Lyn, B. Jeremy, C. Paolo, D. Steven, E. Yuri, K. Ulrike, S. Moritz, T. Barbara, and W. Jo. Napa cards. http://napa-cards.net/#info, 2016. Accessed: 2021-02-01.

[44] MasterClass. Mastering story arc: How to structure a climax. https://www.masterclass.com/articles/how-to-structure-a-climax, 2020. Accessed: 2021-03-18.

[45] P. D. Mautone and R. E. Mayer. Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology*, 93(2):377, 2001.

[46] S. McKenna, N. Henry Riche, B. Lee, J. Boy, and M. Meyer. Visual narrative flow: Exploring factors shaping data visualization story reading experiences. *Computer Graphics Forum*, 36(3):377–387, 2017.

[47] R. Menkman. *The glitch moment (um)*. Institute of Network Cultures Amsterdam, 2011.

[48] T. Munzner. *Visualization analysis and design*. CRC press, 2014.

[49] B. News. Hans rosling's 200 countries, 200 years, 4 minutes - the joy of stats - bbc four. https://www.youtube.com/watch?v=jbkSRLYSojo, 2010. Accessed: 2021-02-01.

[50] Policyviz. Policyviz: Story structure. https://policyviz.com/2017/03/21/story-structure/, 2019. Accessed: 2021-03-08.

[51] H. Pöttker. News and its communicative quality: The inverted pyramid—when and why did it appear? *Journalism Studies*, 4(4):501–511, 2003.

[52] V. Propp. *Morphology of the Folktale*, vol. 9. University of Texas Press,

2010.
[53] K. A. Quesenberry and M. K. Coolsen. What makes a super bowl ad super? five-act dramatic form affects consumer super bowl advertising ratings. *Journal of Marketing Theory and Practice*, 22(4):437–454, 2014.
[54] N. H. Riche, C. Hurter, N. Diakopoulos, and S. Carpendale. *Data-driven storytelling*. CRC Press, 2018.
[55] B. Rolfe, C. M. Jones, and H. Wallace. Designing dramatic play: Story and game structure. In *Proceedings of Conference on Human Computer Interaction*, pp. 448–452, 2010.
[56] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010.
[57] D. Seyser and M. Zeiller. Scrollytelling–an analysis of visual storytelling in online journalism. In *Proceedings of International Conference Information Visualisation*, pp. 401–406. IEEE, 2018.
[58] D. Shi, X. Xu, F. Sun, Y. Shi, and N. Cao. Calliope: Automatic visual data story generation from a spreadsheet. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):453–463, 2020.
[59] Y. Shi, X. Lan, J. Li, Z. Li, and N. Cao. Communicating with motion: A design space for animated visual narratives in data videos. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 1–13, 2021.
[60] X. Shu, A. Wu, J. Tang, B. Bach, Y. Wu, and H. Qu. What makes a data-gif understandable? *IEEE Transactions on Visualization and Computer Graphics*, 27(02):1492–1502, 2021.
[61] J. Smith. Is your data story actually a story? `https://medium.com/nightingale/is-your-data-story-actually-a-story-3d1fa52394d9`, 2019. Accessed: 2021-03-08.
[62] N. Sultanum, F. Chevalier, Z. Bylinskii, and Z. Liu. Leveraging text-chart links to support authoring of data-driven articles with vizflow. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 1–17, 2021.
[63] J. Tang, L. Yu, T. Tang, X. Shu, L. Ying, Y. Zhou, P. Ren, and Y. Wu. Narrative transitions in data videos. In *Proceedings of the IEEE Visulization Conference Short Papers*, 2020.
[64] T. Tang, J. Tang, J. Hong, L. Yu, P. Ren, and Y. Wu. Design guidelines for augmenting short-form videos using animated data visualizations. *Journal of Visualization*, 23(4):707–720, 2020.
[65] TV-Tropes. Multi-take cut. `https://tvtropes.org/pmwiki/pmwiki.php/Main/MultiTakeCut`, 2020. Accessed: 2021-03-13.
[66] J. Upton. *Critical observations on Shakespeare*. G. Hawkins, 1748.
[67] Y. Wang, Z. Sun, H. Zhang, W. Cui, K. Xu, X. Ma, and D. Zhang. Datashot: Automatic generation of fact sheets from tabular data. *IEEE Transactions on Visualization and Computer graphics*, 26(1):895–905, 2019.
[68] N. Warburton. *The basics of essay writing*. Routledge, 2020.
[69] L. Wenchao, W. Yun, Z. Haidong, and Q. Huamin. Improving engagement of animated visualization with visual foreshadowing. In *Proceedings of the IEEE Visulization Conference Short Papers*, 2020.
[70] A. Wu, Y. Wang, X. Shu, D. Moritz, W. Cui, H. Zhang, D. Zhang, and H. Qu. Ai4vis: Survey on artificial intelligence approaches for data visualization. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2021. doi: 10.1109/TVCG.2021.3099002
[71] A. Wu, L. Xie, B. Lee, Y. Wang, W. Cui, and H. Qu. Learning to automate chart layout configurations using crowdsourced paired comparison. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 1–13, 2021.