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

Category: Research

Paper Type: theory/model

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 can largely influence how audiences perceive the insights. Freytag's Pyramid is one of the most classic narrative structures that has been widely used in film and literature. While there are continuous recommendations and discussions on applying Freytag's Pyramid to data stories, there exists little systematic and practical guidance for data story creators on how to use Freytag's Pyramid for structured data story creation. To bridge this gap, we examined how existing practices apply Freytag's Pyramid through analyzing stories extracted from 103 data videos. We summarize our findings and propose 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 the story creation process. We evaluate the proposed design space through a workshop with 25 participants. The study result shows that our design space provides a clear framework for rapid storyboarding of data stories with the Freytag's Pyramid structure.

Index Terms—Freytag's Pyramid, Narrative Structure, Data-driven Storytelling, Design Space

1 Introduction

Data stories integrate various visual content (e.g., text, visualization, animation) to tell stories that convey data insights [46]. To create expressive data stories, designers carefully plan the narrative structure of the story. According to prior research, the narrative structure of a story can affect the audience's perception, understanding, and memory of it [28]. The narrative structure also decides the visual design considerations. For example, in visual storytelling, it is considered beneficial to use high-energy colors or increase the movement speed of the camera in the climax of a story to acquire a sense of tension [10].

There has been an increasing interest in studying the narrative structures of data stories [28, 29, 38, 46]. For example, Segel and Heer [46] identified three high-level structures (i.e., *drill-down story*, *martini glass structure*, and *interactive slideshow*) that provide readers with various levels of controls to manipulate 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, 21, 23, 43]. Freytag's Pyramid [25] is one of the most classic narrative structures that has been widely applied in film and literature [22, 25]. In this structure, a story starts with introducing the setting and then develops with a progressive increase of the tension that reaches the highest point at the climax and a following falling of the 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)*) Cohn et al. [20] redefined by analogy with traditional narrative structures, Animi et al. [2] coded the structures of 50 data videos and found a dominant sequence "E+I+PR+" that resembled the Freytag's Pyramid most. Moreover, Brent Dykes [23] proposed the *Data Storytelling Arc* model based on Freytag's Pyramid for telling data stories to drive business decisions. While these studies provide some summaries regarding the application of Freytag's Pyramid to data stories, there is little systematic work on how to create an expressive data story with Freytag's Pyramid structure.

To fill this gap, this work introduces a design space that concludes story patterns in key stages of Freytag's Pyramid: the *Setting*, *Rising-Climax*, and *Resolution* regarding three dimensions: (1) *narrative pattern* that provides narrative devices as initial ideas to assist the creative process; (2) *data flow* that provides strategies for the selection and organization of individual story pieces (i.e., facts backed up by data) to achieve a narrative pattern; and (3) *visual communication* that supports the selection of visual design strategies to enhance the presentation of a narrative pattern.

To derive the design space, we first collected a corpus of data stories with the Freytag's Pyramid structure from various online sources. Then, we narrowed down our analysis to the 103 data videos we collected based on that they have an author-driven narrative sequence that can better represent the intention of the creators, and the design of videos covers diverse visual forms that support us to extract abundant design strategies. To evaluate the effectiveness of our design space, we conducted a workshop with 25 participants. The participants were invited to create data stories with the Freytag's Pyramid structure. They were asked to create story outlines without our design space first, and then 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 materials. The method cards provided the definition, a case example, and a GIF demo for each category in our design space and can be found online: `https://vispyramid.github.io/method-cards/`. The results showed that our design space provided an efficient framework for participants to select, organize, and fill data facts into different stages of the Freytag's Pyramid while at the same time supporting designing visual contents throughout the structure.

2 Background and Related Work

In this section, we first introduce how Freytag's Pyramid is adapted to data stories as the background of our paper. Then we review relevant research works that guides creating narrative visualization.

2.1 Freytag's Pyramid for Data Stories

The narrative structure of a story determines the organization of story content [20]. Research shows that story authors consider the narrative structure essential to the expressiveness of a story and influence how audiences perceive the story [28]. Representative narrative structures have emerged across different domains such as drama, literature, and cinematography. The Greek philosopher Aristotle firstly examined the basic structure of drama and concluded that a story consists of a series of cause-and-effect events that go through stages of the beginning, middle, and end [53]. Based on Aristotle's model [23], Freytag [25] 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 raising and falling of tension, with the climax being the highest point of the tension, which forms a pyramid-shaped story structure. Freytag's Pyramid model was proved to be concise and powerful and has been used in a wide range of fields such as science communications, advertising, and games [22, 24, 44, 45].

Subsequently, the potential of applying Freytag's Pyramid to construct data stories has been found and actively discussed in the visu-

alization community [21, 23, 43, 50]. Cohn et al. [20] identified four major narrative categories (i.e., *Establisher (E)*, *Initial (I)*, *Peak (P)*, and *Release (R)*) refined from the stages of Freytag's Pyramid and other dramatic structures to analyze the structure of visual narratives. He further found the canonical "E-I-P-R" sequence resembles the Freytag's Pyramid most. Following Cohn's theory, Amini et al. [2] analyzed 50 professional data videos and coded the narrative categories of each video segment. They identified the most common structure "*E+I+PR+*" that consists of several units in the Establisher (E+) stage to ground the story, and several units in Initial (I+) stage to build up the tension and leads to a Peak (P), and finally wind up with multiple units in Release (R+) stage to convey the takeaway messages. Brent Dykes proposed a model called *Data Storytelling Arc* [23] based on Freytag's Pyramid for telling data stories to drive business decisions. Dykes's model includes four stages: *Setting*, *Rising insights*, *Aha Moment*, and *Solution and Next Steps* for describing the structure of a story. The *Setting* stage "provides the audience with the background information". The *Rising insights* stage is where "the subject of the analysis is explored at a deeper level", which serves as the build-up for the *Aha moment*. The *Aha moment* "shares the main finding or central insights". Finally, the story ends in the *Solution and Next Steps* stage that "guides the audience through the different options they have".

The aforementioned studies have shed lights on the importance of the application of Freytag's Pyramid in data storytelling. However, there exists little systematic guidance for story creators on how to apply the 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 such a gap by extracting story patterns from existing data stories with the structure. Notably, the four most commonly observed narrative stages we identified from analyzing data stories align well with the "*E+I+PR+*" structure in Cohn's theory. To better capture the progression of the story flow and simplify our design space, we further merge the *Rising actions* and the *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 actions*. 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 gives supporting facts leading to the climax that shares the main finding or the central insight of the story.
- *Resolution:* This stage gives conclusions and take-away messages.

2.2 Guidance on Narrative Visualizations

Researchers in the visualization community have been attempting to provide theoretical basis and guidance for narrative visualizations. Research in this area can be broadly divided into two categories. The first category of research focuses on the visual design strategies of narrative visualizations. Design spaces are proposed regarding different genres of narrative visualizations, such as data-GIFs [49], data comics [8], and timeline [16]. There are also studies on how to use visual embellishments to strengthen viewers' attention and memory (e.g., [9, 13, 14, 19, 27]). Most relevant to our work are those that study the visual design strategies of data videos. Amini et al. [4] concluded a taxonomy of 8 animation types (e.g., creation, destruction, etc.) and their implementations in 8 basic visualizations (e.g., line/area chart, bar chart, pictograph, etc.) from over 70 videos. In follow-up work, Animi et al. [3] found the usage of setup animation and the pictographs in data videos can increase viewer engagement. Tang et al. [51] classified the data-driven animated transitions (e.g., staged animation) and non-data-driven ones (e.g., fade-in/off) for smoothing the narrative transition in data videos. These works are inspirational to identify the dimensions of our design space. The data stories we used in this paper are also extracted from analyzing data videos. However, rather than focusing on the creation process of data videos, our work put more attention to the story flow and the organization of data facts.

Fig. 1: Case1: 200 Countries, 200 Years, 4 Minutes [41]

The second category of the research attempts to provide guidance on how to structure narrative visualizations. Segel and Heer proposed three structures, *Martini Glass Structure*, *Interactive Slideshow*, and *Drill-Down Story* that give the user varying degrees of control over interactive visualizations to manipulate the content and order of the story presentation [46]. McKenna et al. [38] identified and investigated seven flow-factors (e.g., navigation input, level of control, navigation progress, etc.) that affect the reading experience on interactive data stories. For visualization stories in the form of a sequence of visualizations, Hullman et al. [28] summarized transition types (i.e., dialogue, temporal, causal, granularity, comparison, and spatial) between adjacent visualizations. In the follow-up work, they found a hierarchical structure of visualization sequence was preferred when people were asking to sequence a set of visualizations. Specifically, people prefer to group visualizations with shared data properties such as time period, measure, and spatial region, and then order the views within these groups [29]. Their findings help create smooth story flows by reducing the transition costs between the visualization. Although the aforementioned work provides guidance on improving the smoothness and flexibility of the structures of data stories, the storytelling techniques for structures are not fully explored. We hope that our study on how to apply the time-tested dramatic structure of Freytag's Pyramid to the data video stories can complement the current research for increasing the expressiveness of narrative visualization structures.

3 Case Analysis

In this section, we present selected cases from our corpus to give a preview of how the concepts in our designs space (i.e., narrative stages, types of data flows and visual communications) are presented in the data stories, and how we decompose data stories in our analysis. The detailed methods of case collection and analysis are introduced in subsection 4.1. In particular, The first two cases are videos. The third case is from web to demonstrate how our design space can be applied to other data stories forms in addition to videos. 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*) that use the narrative patterns concluded in our design space.

3.1 Case1: 200 Countries, 200 Years, 4 Minutes

200 Countries, 200 Year, 4 Minutes was produced by BBC News collaborating with Hans Rosling. In this video, Hans Rosling tells a story of global development by presenting the changes of two representative indicators, that is, the life expectancy and income per person, of 200 countries from 1810 to 2009.

Setting. In the beginning, Rosling introduces a bubble chart that shows the life expectancy and income per person 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

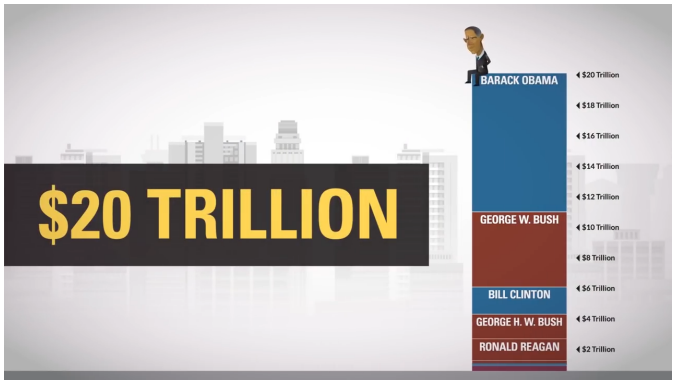


Fig. 2: Case2: Donald Trump's $20 Trillion Problem [18]

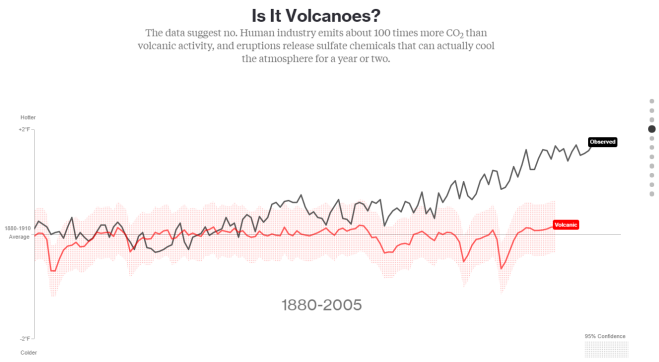


Fig. 3: Case3: What's Really Warming the World? [11]

for understanding data insights conveyed by the visualization through *introducing the data attributes* encoded by the scatter plot. In the introduction, the animation of *building up charts components gradually* reduces the cognitive burden of the audiences and help them understand the visualization [3, 37]. **Rising-Climax.** After the visualization is setup, Rosling dives deeper into the part of how the data *changes over time* from 1810 to 2009. While playing the animation, Rosling describes notable changes in the data and the historical events that are related to the changes. When the story comes to 1948, the world was going to witness a remarkable period when the former colonies gained independence, got healthier, and were catching up with the development of western countries. The story then reaches the climax when the fast progress of the world economy and health conditions is presented, meanwhile, Rosling *speeds up* the animation to enhance the tension of the climax perceived by the viewers. **Resolution.** At the end of the story, Rosling *replays the visualization animation* to recap the *facts to be emphasized* in the story, that is, the marvelous progress of the countries achieved in 200 years.

3.2 Case2: Donald Trump's $20 Trillion Problem

Donald Trump's $20 Trillion Problem was produced by Visual Capitalist, an online publisher that specializes in data-driven visual content on global trends, investing, technology, and economy. This video reveals the significance of the United States national debt and has been viewed over 1 million times.

Setting. The story starts with the description of obstacles that the Trump administration would have to tackle and then points out the most looming issue - the colossal 20 trillion dollars debts. A animated counter showing *counting numbers* until reaching 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 administration brought as shown in Fig. 2. The blocks fall on the ground and stack together one by one as the story develops. With the falling of each block, the animation becomes *increasingly intensive* as the consecutive block raising more dust on the ground and causing more intensive shaking of the scene compared to the previous block. Meanwhile, the video uses a close-up shot of each block and the camera moves up to show new blocks appearing in the scene. When blocks have all fallen, the camera zooms out from *the close-up shot to the overview* of all blocks. Finally, the overview shows how the numbers are *added up* to the significant amount of debt. **Resolution.** In the end, the story discusses how the economic plans proposed by the Trump administration would affect the U.S. debt. Based on the infrastructure investigating 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 visually present the prediction of the future situation.

3.3 Case3: What's Really Warming the World?

What's Really Warming the World from Bloomberg is a story that explores what factors are causing global warming in the form of "scrollytelling" [47]. We present this case to show the commonness of how Freytag's Pyramid structure is presented in data videos and interactive data stories on web pages. This data story underlied the Freytag's Pyramid structure but has less visual designs compared to what we found in data videos. To align with the diverse visual design methods presented in data videos, in the following, we also discuss how using certain visual communication methods can enrich the original designs.

Setting. The story first introduces the backgrounds through a text paragraph about the global warming issue and the debate about what really causes the global warming. To better evoke the readers' concerns on the topic, the original story can include a 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 the 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 orbital changes, solar and volcanoes appear one by one. However, in each presented chart, the simulated trend is far below the actual trend, except for the last chart representing the influence of the greenhouse gases. It turns out the greenhouse gases are the only factor that the simulated trend of temperature changes it brings is close to the true temperature changes. The story builds up its climax by *contrasting the influence of the greenhouse gas and that of other factors*. The tension and surprise in the climax can further be enhanced by adding animation effects such as *shaking the scene* from our design space. **Resolution.** Finally, the story gives *a summary chart* that shows the temporal changes brought by all factors to *re-emphasize the fact* that the greenhouse gases play a dominant role in global warming.

4 Design Space

In this section, we first present our study methodology for formalizing the design space of data stories that follow Freytag's Pyramid model. Then, we present an overview of the three dimensions in our design space, including *Narrative Pattern*, *Data Flow*, and *Visual Communication*. Next, we elaborate on the design space in detail according to the order of the three narrative stages, including *Setting*, *Rising-Climax*, and *Resolution*.

4.1 Methodology

To understand how the Freytag's Pyramid structure has been utilized in data stories, we collected a corpus of data stories from the lists of previous studies [2, 46], reputable news outlets (e.g., Vox, The New York Times, BBC News), and well-known video platforms (e.g., YouTube and Vimeo). During this process, we ensured that the data stories we collected are data-driven and contain at least one data visualization. Also, the stories should be author-driven [46] and have a non-ambiguous narrative structure. Then, we analyzed the narrative structures of the data stories following the methodology proposed by Amini et al. [2] and identified those data stories that use the Freytag's Pyramid structure.

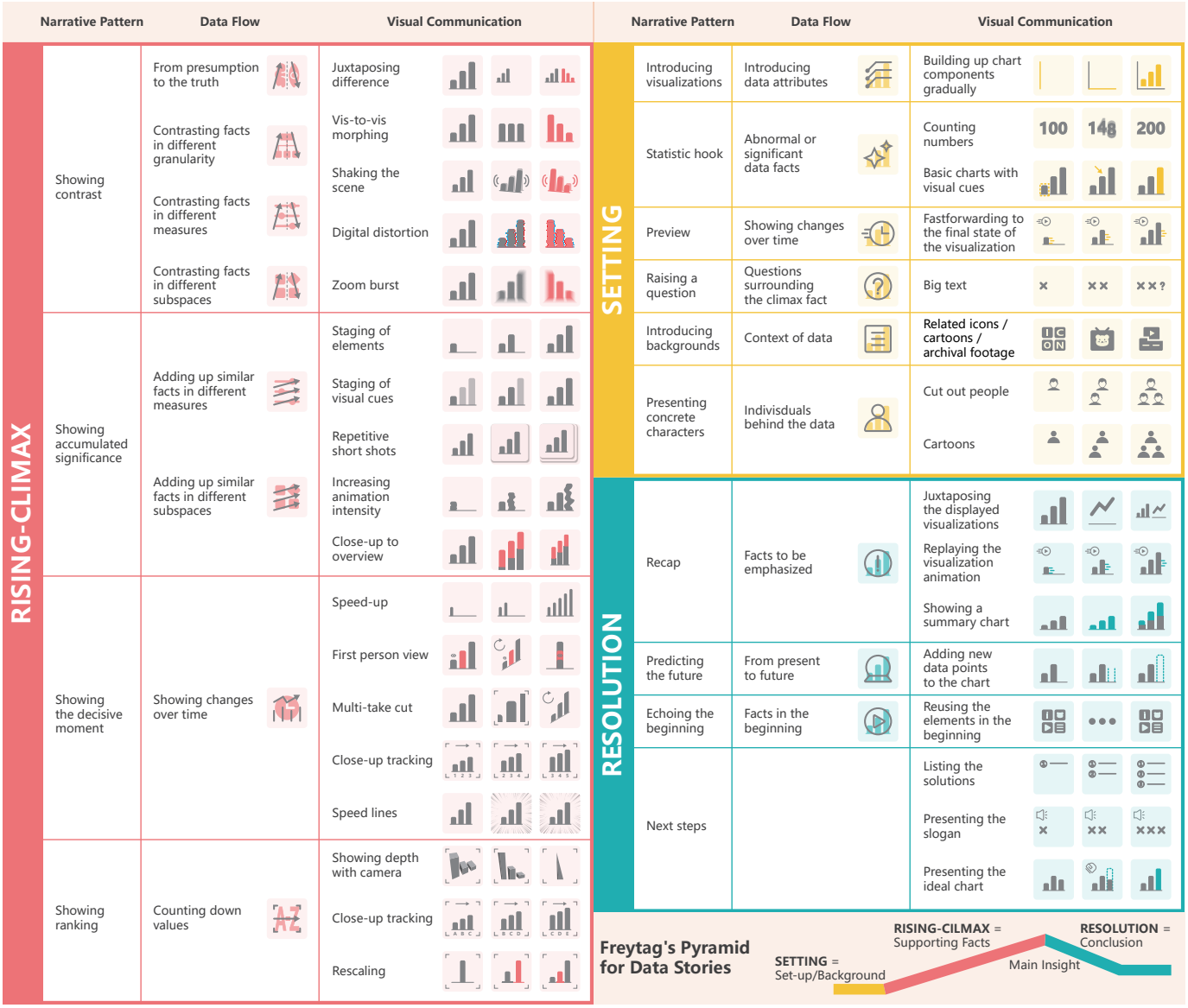


Fig. 4: Design space for creating data stories having the Freytag's Pyramid.

Specifically, three authors first segmented the data stories and coded the narrative stages independently. Then, they cross-checked the codes and refined the codes until reaching an agreement. We found *Setting*, *Rising*, *Climax*, and *Resolution* stages occurred in the majority of the data stories, which aligns with the findings in prior studies [2, 23]. We further narrowed our analysis to the 103 data videos we collected as videos cover more diverse techniques of visual communication (e.g., static visual representations, animations, narratives). Thus, our study results derived from analyzing data videos could also be applied to other author-driven narrative visualizations such as slideshow [28] and scrollytelling [47]. We initiated the design space around on these four narrative stages. We further integrate *Rising* and *Climax* stages as *Rising-Climax* as stated in Section 2.1 for simplicity.

We describe the story patterns in each narrative stage from three dimensions: the *narrative pattern*, *data flow*, and *visual communication*. Inspired by a prior study by Bach et al. [7], we build our design space around the *narrative pattern* that directly implies the initial ideas of story creators and assists the creative process. Furthermore, according to Lee and Riche [33], the creation process of visual data stories involves selecting and organizing story pieces (i.e., facts backed up by data) and presenting them through visual representations. With this in mind, we further investigate how story creators formalize their ideas into the selection and organization of story pieces (i.e., *data flow* dimension) and the visual representations through which the narrative intent is communicated (i.e., *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 [28] and data comics [8]. For instance, we borrowed the transition type between consecutive visualizations, *Measure Walk*, from the study by Hullman et al. [28] when summarizing the data flow *contrasting findings in different measures*. In the visual communication dimension, we reviewed cinematography and digital storytelling techniques [5, 10, 12] to define the visual design strategies.

Note that none of the categories in each dimension were pre-defined, but were identified through an iterative process of mooting, merging, and refining until exclusive narrative patterns with subordinated data flow and visual communication patterns emerged. Specifically, three authors with related backgrounds participated in the coding process. In particular, one has over 2 years of experience in designing narrative visualizations, one has the screenwriting background and over 4 years of experience in practice, and the other has over 2 years of research

experience in data visualization. They independently coded the data videos using the thematic analysis [15]. Then, codes for each narrative stage went through at least 3 rounds of discussions to reach the consensus.

4.2 Design Space Overview

In this subsection, we give the definition and functions of each dimension in our design space before diving into the individual items in each narrative stage. Notably, the data flow and visual communication dimension center around the narrative pattern dimension and provide strategies for each narrative pattern accordingly.

4.2.1 Dimension I: Narrative pattern

According to Bach et al. [7], the definition of *narrative pattern* is "a low-level narrative device that serves a specific intent", which "can be used individually or in combination with others". The *narrative patterns* in data stories can further be interpreted as the counterparts of classical plots of literary stories. For instance, the *Climax* in a literary story often comes with a plot of a fierce confrontation between the protagonist and the antagonist. Likewise, in data stories, a climax could 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 could be any type (e.g., trend, value, difference, etc.) [54] as what 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 a data fact that is related to a person. The other kind of data flow illustrates the connection between a group of data facts. For example, the data flow *contrasting findings 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 intended to provide guidance on reflecting a narrative pattern through *data flow*, in practice, story authors could either have a narrative pattern in mind and search for the corresponding data facts, or starting from the characteristics of data facts they found valuable to see what narrative patterns can be incorporated in the story.

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 from analyzing data videos, we were able to 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 *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 in data stories, we do not propose guidelines on selecting visualizations based on the types of data used in the story, which has already been extensively investigated by the existing literature [40].

4.3 Setting

We identified six narrative patterns that can be applied in the *Setting* stage as shown in the top right yellow section in Fig. 4.

Introducing visualizations Introducing visualizations explains the visual encodings of the visualizations to guide the audience on how to 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 [34]. 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, before going to the details of the story. Visual communication. The visual communication technique *building up chart components gradually* presents each part of the visualization progressively in the scene when the corresponding design is explained. This technique can reduce the cognitive burden of the audience [37], facilitates their comprehension of the visualization, and increases their engagement [3].

Statistic hook Statistic hook is a storytelling technique that presents some interesting numbers or visualizations 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-oriented data [23] or summary statistics (e.g., sum, maximum values, etc.) For example, in the third case (Section 3.2), the story starts by showing how large the United States national debt is to attract the audience' attention to the severity of the U.S. debt crisis. Visual communication. When presenting the statistic hooks, the designer may use *counting number*, an animation showing a number counting up instead of presenting the final value to arouse the audience' interests in the result. For 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 (e.g., glow, arrow, etc.) [32].

Preview Preview is a foreshadowing storytelling technique that gives an advance hint of what is to come later in the story [55]. This narrative pattern is often used by movie trailers for establishing the expectation of the audience on how the story will develop [55]. 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. Then, they 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 used visual communication is *fastforwarding to the final state of the visualization* which goes through visualizations from the beginning of time to the end quickly with animation.

Raising a question This narrative pattern, as suggested by its name, poses a question directly to the audience to evoke their curiosity and sets suspense for the subsequent exploration of the answer [7]. 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 analogy with literary stories, the question can be compared to the main conflict of a story, and the climax should resolve the conflict [36]. Visual communication. Story authors can use *big text* to emphasize the question on full screen. The occurrence and disappearance of the text also serve as a transition signal of the story progressing to the next stage (i.e., from *Setting* to *Rising-Climax*).

Introducing backgrounds This narrative pattern focuses on providing contextual information of the story, such as the motivation or the general topic of the story. Data flow. In literary stories, contextual information often refers to the time, place, or surroundings of the events. In data stories, however, the contextual information often relates to how the data was collected, the data subject, and what tasks the data analysis is for. We refer this type of story pieces as the data flow *context of data*. Visual communication. The commonly identified visual communication approach to convey the abstract information is by using *related icons, cartoons, and 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 begins the story with a personal view, such as telling an anecdote of someone. 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 [23]. Data flow. The corresponding data flow is the *individuals behind the data* - data facts related to data about some persons. 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 in the 2012 Olympics, and then tells the data of the Olympic records in men's 100-meter freestyle. 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 that is illustrated in the left red section of Fig. 4.

Showing contrast Showing contrast presents data facts that have a substantial difference, which creates a plot turn that usually leads to the climax. Data flow. The common data flows of *showing contrast* are *contrasting facts in different measures* which shows the contrast of different measured values for the same categories in the data and *contrasting facts in different subspaces* which shows the contrast of the measured values across different categories. A measure is a numerical data column on which certain aggregations (e.g., SUM, AVG) can be performed [28]. A subspace is a subset of the data by using filters for any data dimensions that contain qualitative, categorical information [54]. For example, a story may compare the increasing of average income and the decreasing of happiness index of a country, with the average income and happiness index being two measures of the dataset with opposite trends; or compare the average income across different countries, where the data of each country is a subspace of the dataset. Another data flow of showing contrast is *contrasting facts in different granularity* in a hierarchical data set, such as comparing the average income across countries in the world and cities of a specific country. Besides contrasting data facts, storytellers also present data that contrary to common beliefs to form a sense of contrast, which we refer to as *from presumption to the truth*. For example, the data video *Wealth Inequality in America* firstly presents the assumed wealth distribution among Americans from a survey before showing the true wealth distribution that is more skewed than what people assumed. Visual communication. We observed five visual communication techniques 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 to another. *Shaking the scene* adds vibrations to the view in climax as an analogy to the feeling of shock. *Digital distortion* is from the glitch art [39] that adds digital noises to the scene to mimic a crash of the filming devices that 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. This narrative pattern is often demonstrated through *adding up similar facts in different measures* or *adding up similar facts 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. Specifically, *staging of elements* lets the visual component (i.e., part of, or the whole of a visualization) representing each data fact progressively appear in the scene with the development of the story. *Staging of visual cues* employs visual cues (e.g., glow, arrow, etc.) 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]. *Close-up to overview* consists of a series of close-up views at 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. *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 intensive rhythm of the story [10]. *Increasing animation intensity* augments the exaggeration of the animation as the story progresses. For example, in the case2 in Fig. 2, when the blocks representing data facts fall down one by one, the consecutive block raises more dust and causes a more intensive shaking of the scene compared to the previous one.

Showing the decisive moment Showing the decisive moment of a story 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. We refer to the corresponding data flow as *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, which acts as a signal of the upcoming changes by increasing the pace of the story and raising the expectation of the audience. *First person view* puts the visualization in a three-dimensional scene and makes the data trend as a roller coaster track in which the reader then rides from one end to the other. When audiences' views move up and down following the trend, they have an immersive experience on the fall and rise of data. *Multi-take cut* demonstrates the short period of time right before the climax from various camera angles around the chart, which adds a sense of movement or frenetic excitement [52]. *Close-up tracking* horizontally moves the camera 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 is originally used in Manga art to enhance the intensity of an action moment [26] .

Showing ranking Showing ranking is a narrative approach of sorting things according to certain criterion and reveals them one by one to establish a sense of suspense. Data flow. The corresponding data flow *counting down values* typically includes data fact(s) of story subjects from those ranked low to ranked high. Visual communication. We identified three visual communication techniques. *Close-up tracking* horizontally moves the camera to show items lined up in a rank chart from the one ranking the lowest to the one ranking the highest. *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 three-dimensional. It applies camera movement techniques such as crane, tilt, and pan to shot each item, thus enhances the audience's perceptions on the size of each item to evoke a sense of amazement [10].

4.5 Resolution

In the *Resolution* stage, we identified four narrative patterns as presented in the left blue section of Fig. 4.

Recap Recap helps renew 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 new chart to summarize the findings scattered in the story. For instance, the video *7 Billion: How Did We Get So Big So Fast?* [42] uses an animated bar chart to show how the global population of different continents

changes over time in the *Rising-Climax* stage and uses an area chart to display the overall trend of changes in the *Resolution* stage.

Predicting the future Storytellers apply this narrative approach to provide a glimpse of the future trends of the data. 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 chart*, which allows the audience to investigate the future data in the context of historical data.

Echoing the beginning *Echoing the beginning* revisits story contents presented at the beginning, which is a common narrative technique used in literary stories [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 the elements in the beginning*, 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 an 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 demonstrates the solutions clearly and encourages the audience to take actions. Meanwhile, *presenting the ideal chart* demonstrates the target data such as a quarterly sales target to the audience for stating the goal.

5 Workshop

To evaluate our design space, we ran a workshop with 25 participants to investigate: G1) the perceived usefulness of the design space; G2) how design patterns in our design space were used in creating data stories; G3) whether the design space is easy to use.

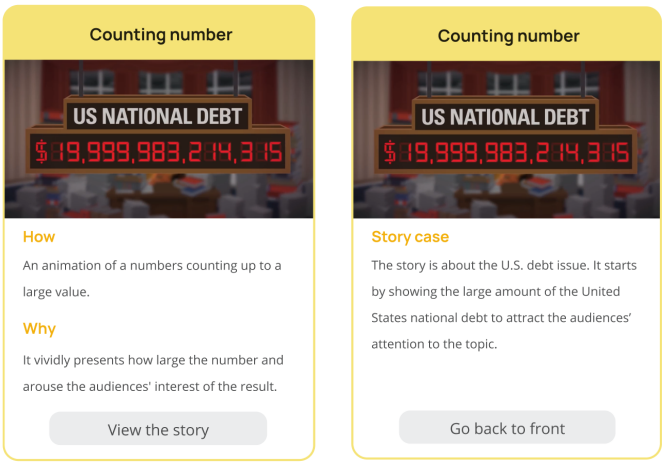
5.1 Participants

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

5.2 Teaching Materials and Data

During the workshop, we provided a set of method cards as the teaching materials of our design space. The design of the method cards was inspired by the Napa Cards [35] and IDEO Method Cards [30]. The intention was to facilitate participants' understanding of each category of our design space with examples, and boost inspiration on the use of narrative patterns, data flows, and visual communication strategies in data story creation. The method cards are organized by the 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. In particular, for each category in the data flow dimension, the method cards provide the definition and an example of how a story could organize data facts into the corresponding data flow. For the visual communication dimension, the method cards introduce the definition, give an example, and also a GIF demo to show how the narrative pattern can be displayed to the audience. All method cards are available at our website [1]. The datasets we used in the workshop for participants to create data stories related to two topics of general interest: *top*

[1] https://vispyramid.github.io/method-cards/



(a) (b)

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

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. Instead of giving participants only the raw data, we also provided various categories of pre-extracted data facts (e.g., outlier, ranks, distribution, etc. [54]) and their corresponding visualizations to get participants' attention more focused on planning narrative strategies following suggestions of previous studies [2, 8].

5.3 Procedure

The workshop lasted about 4 hours. We began with a 30-min introduction to the narrative visualization, the Freytag's Pyramid, and the concept of storyboard for planning the story. To let the audience understand how this structure could be applied to data stories, we showed two example data stories that were structured based on Freytags' Pyramid. We decomposed the stories into different narrative stages and gave a detailed illustration of how each stage was implied in the narratives. After the introduction, all participants were divided into groups of 2-3 people on a voluntary basis, which ended up in 12 groups of participants. Each group was asked to create stories with the Freytag's Pyramid structure based on either the given datasets or any online datasets on their own initiative.

To have a thorough investigation on how our design space could help story creators construct data stories from story planning to the final accomplishment, we gave each group of participants 40 minutes to explore the dataset and create a story outline with key points in each stage before diving into the design space. After that, we gave instructions on how to use our design space and the method cards on our website. Participants were then asked to refine their story outlines using our design space freely. Once a group fixed their story outline, we instructed them to sketch a storyboard to present the visualizations of their stories (as shown in Fig. 6). Participants were also allowed to write down their visual design ideas that were not able to be demonstrated through sketching, such as transitions between scenes and animation effects. The entire sketching process lasted around 2.5 hours. When finished, 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-interviews with the participants once a group finished drawing, and some interviews were conducted after the workshop. We asked participants for their feedback on using Freytag's Pyramid for data stories, in particular, the difficulties they encountered when drafting the story outlines without the design space and whether our design space helped improve their stories.

5.4 Analysis and Results

This section discusses the findings and feedback from the workshop and semi-interviews regarding our three study goals (G1–G3).

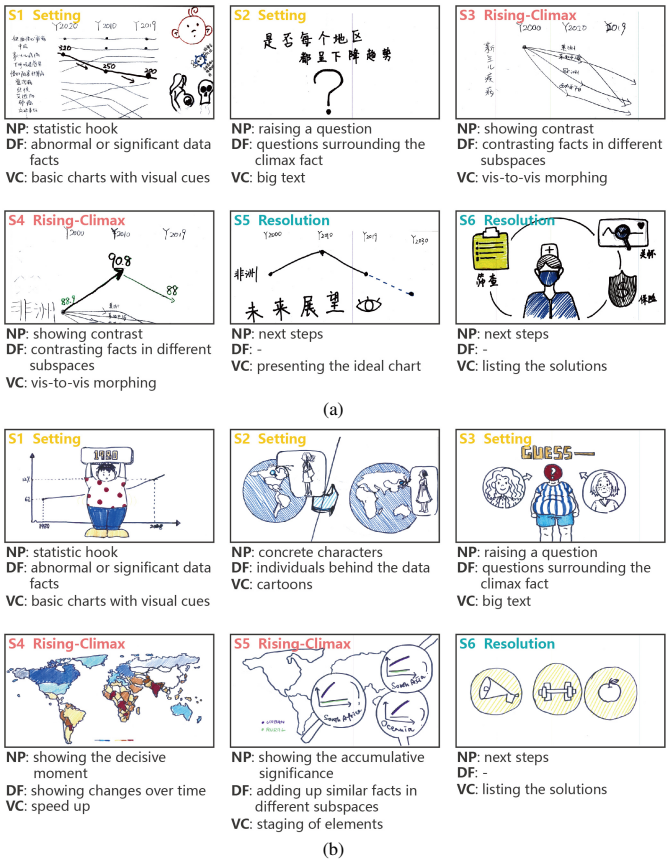


Fig. 6: Storyboards from the workshop: (a) the neonatal conditions over the world (©Xinyang Shan, Danchen He, Yisu Wang), and (b) the trend of global obesity (©Jiaohan Wu, Xiaozhou Ye). The used patterns are noted under each scene including the narrative pattern (*NP*), data flow (*DF*), and visual communication (*VC*).

G1: Design space usefulness

In the interview, we asked participants' opinions on using the Freytag's Pyramid structure to construct data stories. In general, participants found it was nature to design their stories following certain structures and acknowledged that Freytag's Pyramid was concise and friendly for beginners in data storytelling (G5P1: *"Everyone can use it!"*, G6P1: *"The structure is clear, I know what to talk about in each stage of my story."*). Second, participants found the narrative structure provided them with a compelling way to communicate their insights compared to reporting data in a flat style (G3P2: *"We tried to put all findings into the story, but to clearly convey the climax in the structure, we removed irrelevant ones."*. Some participants also mentioned stories with underlying narrative structure were more acceptable and easier to remember for general public and could be potentially used in viral videos, which further confirmed the benefit of incorporating Freytag's Pyramid into data stories.

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 ($M = 5.88, SD = 1.71$), respectively. Participants expressed their appreciation on the overall 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."*, G6P2: *"I had lots of thoughts on the story plots and designs, this framework helped me clear my mind."*) and was easy for beginners to use (G4P1: *"For beginners like me, it serves as a set of story templates providing different choices for me to try and combine."*). We further looked into how our design space could help participants structure the story through by comparing the story outlines that participants made before and after using the design space. 11 out of 12 groups of participants changed their story outlines by adding (mostly), deleting or relocating data insights in their outline.

From the interview feedback, we found our design space is helpful from the following perspectives: (1) data facts selection and organization (G10P1: *"It helps us allocate our insights into proper positions in the story flow"*, G5P1: *"At first, we had no idea where to start our analysis. The data flows provide starting points"*); (2) enriching the story plots with new inspirations (G10P2: *"In the beginning, we didn't have an proper entry point...We found the 'raising a question' in the design space was quite suitable."*, G9P1: *"The 'showing contrast' reminded us of our previous findings about...We looked it again and found a good story."*). (3) introducing new visual design approaches (G11P1: *"I never thought of the idea of gradually increasing the animation intensity to build up the tension of the story. It is really eye-catching."*, G12P1: *"Most visual communication techniques are new to me."*). One designer G11P2 noted that *"The design space is comprehensive. Though all visual effects I can eventually think out by myself, it provides quick reference."*; (4) sparking new design ideas through the combination of different visual communications. For example, one group used both *shaking the scene* and *juxtaposing the difference* to support the *showing contrast* narrative pattern, and one group used both *close-up tracking* and *speed-up* to augment the *showing the decisive moment* narrative pattern. Some participants concerned that some visual effects could be hard to implement in practices. Some participants also worried about their lack of knowledge on data analysis for extracting important findings that could fit into the data flow in the design space.

G2: Design space usage

We coded all the storyboards to investigate how the patterns in our design space were used. We found every narrative pattern was used. The dominant patterns and their frequency in each stage were as follows: *statistic hook* (8) and *raising a question* (6) in the setting stage, *showing contrast* (8), *showing accumulated significance* (8) and *showing the decisive moment* (4) in the rising-climax stage, and *next steps* (12) in the resolution stage. In addition, we found most visual communications (24/34) were used. The dominant visual communications and their frequency for each narrative pattern were: *listing the solutions* (5) and *presenting the ideal chart* (3) for *next steps*, *juxtaposing difference* (5) for *showing contrast*, *speed-up* (3) for *showing the decisive moment*, *staging of elements* (3) and *staging of visual cues* (3) for *showing accumulated significance*, and *using charts with visual cues* (6) for *statistic hook*. Most data flows (14/17) were also used. The dominant data flows and their frequency for each narrative pattern were: *presenting contrasting facts in different subspaces* for *showing contrast* and *adding up similar facts in different measures* (8) for *showing accumulated significance*. We also identified an animation effect that was not covered by our design space as a few participants added a magnifier to emphasize related regions in the visualizations. Surprisingly, we found 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 how the expressiveness of story structures could be enhanced by applying our design space.

G3: Ease of use

The usability of our design space was positively rated 1▁▁▁▂▃▅▆7 ($M = 6.04, SD = 1.79$). Participants praised on clarifying the meaning of each visual communication approach through a sequence of key frames at the right side of each category as shown in Fig. 4. Meanwhile, they also appreciated the examples illustrated in 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 the potential improvements of our method cards, some participants desired an mobile version of the website to investigate the card in cell phone, and some participants liked 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 Structure

In this work, we introduced the structure-Freytag's Pyramid from the field of narratology to narrative visualizations. We analyzed 103 data videos with the Freytag's Pyramid structure and proposed a design space that characterizes the narrative patterns, data flows, and visual communication techniques that help construct data stories with the structure. When formalizing the design space, we found that many techniques in data stories can be mapped to those in traditional storytelling forms, such as fictions and films. For example, in fictions and films, the climax is often built through a fierce confrontation, while in data stories climax is usually built by showing contrasting data facts. We also observed many storytelling methods, such as foreshadowing and repetition, in both data stories and traditional stories. However, we also noticed that creating data stories faces unique challenges compared to traditional stories. For example, data as the main "character" of the data stories can be far more abstract than real characters in traditional stories, which brings more difficulties to crafting a narrative structure that is meaningful and understandable to the audiences. In our workshop, some participants stated that they found it difficult to select data facts from the dataset and arrange them into an engaging plot before given the design space. Our design space successfully alleviates this problem with the summary of data flow patterns observed in existing practices guiding the selection and organization of data facts.

Although our design space is derived from data videos, we believe that it can also be easily applied to other author-driven narrative visualizations. Because data videos are a representative genre of author-driven data stories [46] and cover diverse visual representations (i.e., both static and animated), the design patterns we observe can cover a wide range of storytelling forms from slideshow [28] to scrollytelling [47]). However, since data videos mostly convey stories in an author-driven way, they involve few interactions between data stories and the audience, which is also a key design factor in the creation of interactive data stories. In the future, we plan to continue examining the role of interactions and other design factors (e.g., colors, chart layout, etc.) in the pyramid structured storytelling. We also plan to study the use of more types of classic narrative structure in data stories, such as Hero's Journey [17] and parallel narratives [6].

6.2 How Designers Structure Data Stories

When selecting data facts for the narrative structure, designers would consider the emotional value of the data apart from their logic connections. As participant G3P2 noted: *"When exploring the dataset, 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 (G5P1: *"We present the data about the top death causes of the world and many development countries first to let the audiences assume...then we show the data of Africa which tells a different story"*). Second, we found that the designers generally put much effort into attracting audiences at the beginning of the story. For example, three groups combined multiple narrative patterns (e.g., using both statistic hook and raising a question) from the *Setting* stage to get the attention of audiences. According to the participants' feedback, they felt reading data could sometimes be less interesting to the audiences, therefore bringing the audience's interest was their primary concern. Third, we found some participants extended the basic pyramid structure and created nested story structures in the workshop. For instance, one group of participants designed two climaxes in their stories by embedding a delicate pyramid structure into a larger one that formed the overall story.

6.3 Implications for Automatic Data Story Creation

In recent years, many researchers in the visualization community attempt to develop tools to generate data stories automatically [31, 48]. While existing tools are able to identify logical structures behind the data facts [48], it is still unclear how to generate an expressive data story. Through constructing a design space for data stories with the pyramid structure, our work provides a certain theoretical foundation for generating expressive data stories automatically. For example, after detecting the logic between data facts, the system can organize the data facts into data flows of the pyramid structure and choose corresponding visual templates to form a structured data story. While the design space may provide some directions on 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 not easy. It requires the machine to understand the semantics and contextual information in the data. In our workshop, many participants completed this task based on their own domain knowledge and personal judgment on what data facts are valuable. Second, selecting the most appropriate visual communication techniques from the design space is non-trivial. According to our observation, human designers usually consider a bunch of subjective factors when determining the optimal design solutions. For example, they would decide which design method is most "suitable" or "beautiful" for the scene. They would also consider the coherence of different visualization or scenes. Understanding such knowledge around aesthetics is challenging for computer systems. Third, evaluating the quality of the generated data stories with the pyramid structure is not easy, since the narrative stages are mostly associated with people's subjective feelings and can be difficult to quantify. Given these challenges, we propose that future tools aiming to generate data stories automatically should incorporate more knowledge about data interpretation, computational aesthetics, and computational semantics into the design considerations.

6.4 Limitations and Future Work

As our design space is derived from a corpus of 103 data videos, it may not capture all techniques of creating data stories with the Freytags' Pyramid structure. The design space could be refined and expanded by collecting more cases and including more design factors. For instance, it can be further refined to structure stories of certain data types or cover other designs strategies such as using the change of colors to indicate the plot development, which is common in films [10] but under-explored in data stories. Second, though we believe the strategies concluded in the design space could be applied to other author-driven linear narrative visualizations such as slide-show style presentation or scrollytelling, there may be potential challenges that are specific to other scenarios that require further exploration. Third, while our workshop indicates that our design space can facilitate the creation of data stories with Freytag's Pyramid, more rigorous assessment needs to be done to quantify the actual effectiveness of the patterns and compare which patterns are more preferable. Our future work includes conducting a systematic evaluation on the patterns in our design space and exploring the potential of applying other narrative structures to data storytelling.

7 Conclusion

We propose a design space for applying the narrative structure Freytag's Pyramid to data stories through the analysis of 103 data videos with this structure. The 103 stories are decomposed into key stages in the structure, and the design space concluded narrative patterns that serve the narrative intents of each stage. For each narrative pattern, the design space further summarized the strategies for selecting and organizing data facts to reflect the narrative patterns in the data flow dimension, and the visual design techniques to support presentation of the narrative patterns in the visual communication dimension. The workshop results indicate that our design space provides guidance with a clear framework to facilitate the creation process of data stories with Freytag's Pyramid. Participants in the workshop also appreciated the idea of infuse the data stories with the power of narrative structures to engage the audiences and increase the communication effectiveness. We hope our design space can help designers, data analyst, journalists, or any data enthusiast to convey their data insights through engaging stories, as well as give implications on the development of data story authoring tools.

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] 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.

[8] 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.

[9] 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.

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

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

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

[13] 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.

[14] 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.

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

[16] 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.

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

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

[19] 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.

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

[21] 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.

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

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

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

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

[26] 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.

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

[28] 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.

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

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

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

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

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

[38] 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.

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

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

[41] 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.

[42] NPR. 7 billion: How did we get so big so fast? — skunk bear. `https://www.youtube.com/watch?v=VcSX4ytEfcE`, 2011. Accessed: 2021-02-01.

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

[44] 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.

[45] 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.

[46] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] J. Tang, L. Yu, T. Tang, X. Shu, L. Ying, Y. Zhou, P. Ren, and Y. Wu. Nar-

rative transitions in data videos. In *Proceedings of the IEEE Visulization Conference Short Papers*, 2020.

[52] TV-Tropes. Multi-take cut. `https://tvtropes.org/pmwiki/pmwiki.php/Main/MultiTakeCut`, 2020. Accessed: 2021-03-13.

[53] J. Upton. *Critical observations on Shakespeare*. G. Hawkins, 1748.

[54] 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.

[55] 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.