Character-Oriented Design for Visual Data Storytelling

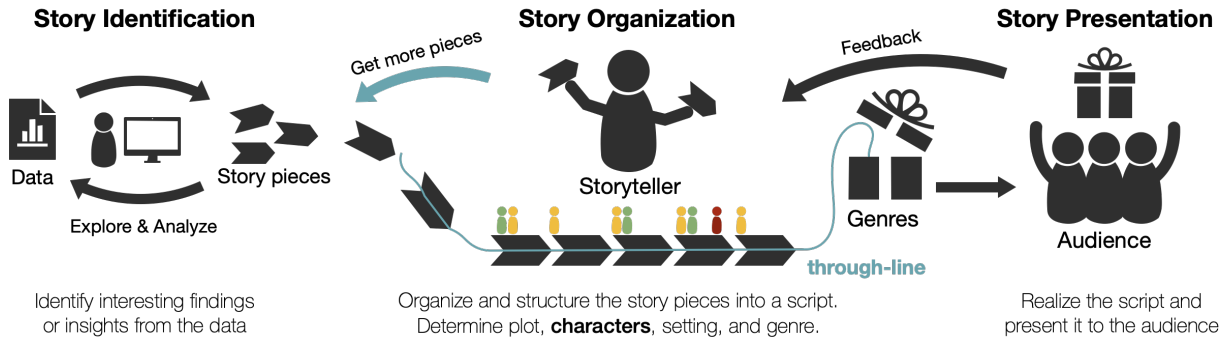Keshav Dasu, Yun-Hsin Kuo, and Kwan-Liu Ma

Fig. 1: Character-oriented visual storytelling consists of three stages: story identification, organization, and presentation. A key task in story organization is to determine what **main**, **supporting**, **antagonist** characters, etc. are and their relations to the plot.

Abstract— When telling a data story, an author has an intention they seek to convey to an audience. This intention can be of many forms such as to persuade, to educate, to inform, or even to entertain. In addition to expressing their intention, the story plot must balance being consumable and enjoyable while preserving scientific integrity. In data stories, numerous methods have been identified for constructing and presenting a plot. However, there is an opportunity to expand how we think and create the visual elements that present the story. Stories are brought to life by characters; often they are what make a story captivating, enjoyable, memorable, and facilitate following the plot until the end. Through the analysis of 160 existing data stories, we systematically investigate and identify distinguishable features of characters in data stories, and we illustrate how they feed into the broader concept of "character-oriented design". We identify the roles and visual representations data characters assume as well as the types of relationships these roles have with one another. We identify characteristics of antagonists as well as define conflict in data stories. We find the need for an identifiable central character that the audience latches on to in order to follow the narrative and identify their visual representations. We then illustrate "character-oriented design" by showing how to develop data characters with common data story plots. With this work, we present a framework for data characters derived from our analysis; we then offer our extension to the data storytelling process using character-oriented design. To access our supplemental materials please visit https://chaorientdesignds.github.io/.

Index Terms—Storytelling, Explanatory, Narrative visualization, Visual metaphor

◆

1 INTRODUCTION

Information, at times, can be abstract and intangible, which may lead to difficulties in communication. The beauty of visualization is captured in its ability to make the intangible tangible, the invisible visible, and the inaccessible accessible. Through visualization, we can utilize visual representations to embody complex and often large datasets, reveal hidden insights about both known and unknown phenomena, and afford a means to showcase findings as well as share insights with broader audiences. We, as data storytellers, are concerned with presenting these findings to large audiences. Stories and visual storytelling have been shared and consumed by our earliest ancestors. Some of the earliest forms of visual storytelling [29] played a role in communicating where rich sources of food can be located or where to avoid dangerous beasts. In visualization, we have utilized storytelling for a variety of communicative needs since it is effective for engagement [41], memorability [25, 56], and showing casualty [28].

As data storytellers, we play a role in capturing and sharing the wonder we see in data with others. In our stories, we are challenged to emphasize the scientific insights of our content and simultaneously engross [40] the audience with our narrative. The challenge of ensuring our content is both consumable and enjoyable while preserving scien-

tific integrity constrains our story design. These constraints can result in the audience having a difficult time understanding [7, 23] insights, topic relevancy, or where in the story to focus.

In data stories, numerous methods have been identified for constructing and presenting a plot. A story plot [28] is a narrative of events, with the emphasis falling on causality. The data storytelling process [44] can be viewed as three stages — identification, organization, and presentation. Typically, the first step resolves in the accumulation of a set of events (i.e., "story pieces"). These pieces are often the insights derived from either the collaborative efforts of data analysts and domain experts or the automation leveraged by statistics [62, 70]. The collection of events is guided by the shared intent of the author and analysts, which is the intention they seek to convey to the audience. This intention can take on many forms [50] (e.g., to inform, to educate, to entertain, or to explore) and centers the story. Next, in the organization stage, several narrative frameworks [44, 58, 63, 72] can assist us in sequencing these events into a cohesive story plot. During this stage, we need to ascertain several properties about these events, namely their relationship to one another and their ordering. We should end up having a structured outline of what we want to convey and the sequence in which to present them. Lastly, we have the presentation stage, where we give the look and feel to the story. There are many methodologies [34, 36, 38] at our disposal for tailoring our story for the target audience. However, it is within the presentation stage that there is an opportunity to expand how we view and design the visual elements that act out our story plots.

In our work, we are interested in data-driven, visual storytelling, particularly the characters that bring them to life. Data storytellers want to create rich experiences that evoke an emotional response, draw the audience in, and leave them with something to remember. Stories can

• Keshave Dasu, Yun-Hsin Kuo, and Kwan-Liu Ma are with the University of California, Davis. E-mail: {kdasu, yskuo, klma}@ucdavis.edu

be brought to life by characters; often they make a story captivating, enjoyable, memorable, and facilitate following the plot until the end. In other media, characters are often used as a bridge for the audience to cross into an unfamiliar and perhaps complex new worlds [14, 28, 67]. For example, in *Star Wars: Episode IV – A New Hope*, Luke Skywalker, the protagonist, is the bridge that leads the audience into the *Star Wars* world. Throughout the story, the audience learns more about the setting and the rules of this world (e.g., force) through his behaviors, rather than from a list of terminology. Through the lens of characters, the audience can gain an understanding of a world without prior knowledge. We are inspired to investigate the possibility of applying characters to convey scientific insights in data stories. A deeper understanding of data characters could address open data storytelling opportunities [40].

This work seeks to address the following — what a data character is, how we classify characters in data storytelling, and a space for how we can develop a data character and apply it within a story. Through the analysis of 160 existing data stories, we present a framework for data characters, where we identify three fundamental character roles: **main**, **supporting**, and **antagonist** characters. With this character-oriented design space, we further investigate how "conflicts" are contextualized in data storytelling, as they drive the narrative and can elevate a telling of a story [28, 67]. Notably, we find that designing an identifiable central character can support the author in aligning the story pieces with their intentions, arranging the sequence with a consistent message, and delivering this message to the target audience.

We consider our primary contributions are:

- a framework [1] for data characters that extend to the data storytelling process;
- a summary of storytelling terminology derived from a variety of storytelling and visualization literature as well as an assessment of data characters in the current literature; and
- case studies in various data story genres to demonstrate the applicability of our design space.

2 BACKGROUND

In this section, we walk through key storytelling terms and contextualize them for data storytelling. Storytelling structure and paradigms [14, 28, 67] have been ever present and are constantly evolving with new mediums and formats. However, we are interested in data stories and determining what applies to storytelling in general and how it can be translated and used for our purposes as *data* storytellers. Data stories can be presented in a variety of ways, such as infographics, comics, videos, virtual experiences, and so on.

2.1 What is a Story?

People have shared and told stories for ages [29] and consequently have postulated the rules and structures for effective storytelling. When considering where to focus on influences for data storytelling, we draw on elements from both written [28,67] and visual media [26]. In *Aspects of the Novel* by E. M. Forster [28], he analyzed the common aspects that all English-language novels share: *story*, *people*, *plot*, *fantasy*, *prophecy*, *pattern*, and *rhythm*. He describes a **story** as "a narrative of events arranged in their time sequence" and the **plot** as "also a narrative of events, with the emphasis falling on causality". Naturally, we look to see how these principles translate into data stories. With written and visual media, authors typically have more freedom and flexibility when creating their stories.

Data stories, however, tend to have less flexibility in that they often are constrained by "non-alterable non-fiction" [40]. Still, from existing literature [10, 40], we can surmise that common themes for both storytelling and data storytelling are: *characters*, *plot*, *theme*, *setting*, and *conflict*. The goal of most **data stories** is to reach a wide or targeted audience through the presentation of visualized findings or messages. The role of a storyteller [67] is to summarize all the narrative events such that the audience perceives this as a self-contained story.

_____

[1] https://chaorientdesignds.github.io/

2.2 Elements of a Story

The terminology for data storytelling as it relates to storytelling has some differences. We define a set of storytelling terms for consistency and clarity in Table 1. However, even within data stories, there are many varying explanations [10, 40, 44, 47, 61] of what *through-line, plot*, and *character* mean. Therefore, we provide a deeper explanation of those in this subsection.

Through-Line. When we look at the storytelling process [20, 26, 67] and data storytelling process [44], both require ascertaining what the author wants to offer the audience. As authors of data stories, we want the audience to feel something [40] as they consume our story and leave with something [50] when they are finished. It is important for us to identify the core message that we want the audience to come away with. For example, a data story could seek to persuade the audience to take action against climate change by communicating the impact microplastics have on biodiversity. Alternatively, the story could intend to explain the flaws of a misconception, all types of plastic have the same effect on the environment. In both scenarios, the authors may communicate the life cycle of microplastics to the audience, while their intention influences the narrative. This message is often referred to as the theme of a story, or a recurring idea. A story can have many themes (e.g., biodiversity) but a through-line [26] is the *theme* (e.g., the impact of microplastics) that runs from the start to the end of a story. It is what helps keep the story on track and helps drive it from start to finish. A through-line interweaves the roles of the characters with the plot. It helps ensure consistency and continuity for the entire story.

Plot. A plot is a description of a set of events with a purpose [67]. Each event that transpires is causally connected and each event is essential to the overall story. As Poe [59] finds with his theory of "unity of effect", every element of a story should help create a single emotional impact. Namely, every element in our story must tie in with the through-line.

Character. A character is an entity that influences either itself or others. A story typically tracks what an entity wants [14], what the entity will do to get it, and what cost the entity will have to pay along the way. What pulls the character along is "desire". A character will take action in pursuit of a desire while learning new information about this desire. This new information causes a change in the course of actions and hence influences the story. A character in pursuit of a desire always encounters a struggle, which may cause a change in the character itself. The task of the storyteller is to present a change in a character or illustrate why that change did not occur. The task of this work is to show how this same process can be applied to data stories.

3 DATA STORYTELLING

Within visualization, there is a large body of work [44, 66] that utilizes and documents storytelling with data. Data storytelling and narrative visualization are two notable branches. Recently, the literature contained in these branches was further organized into three groups [66]. The works are sorted based on whether they (1) *address who are the main subjects involved* (authoring tools and audience), (2) *assess how stories are told* (narratives and transitions), or (3) *consider why storytelling is effective for visualization* (memorability and interpretation). This work seeks to contribute to *how stories are told* by systematically investigating and identifying distinguishable features of characters in data stories. The general storytelling process [26, 67] considers both the plot and the characters. We find little has been done to address the design of data characters, as there is a current focus on creating and conveying the data story plot. In this section, we review the notion of a data character and how it has been previously considered.

3.1 Data Storytelling Process

To better see the relationship between the storyteller and the data storyteller, we organize data storytelling literature into three stages, as shown in Figure 1, based on the process of the data storyteller [44] — identification, organization, and presentation of the story. At the beginning of developing a story, the data storyteller may not know what to discuss or share yet. A set of events (or story pieces [44]), such as

Storytelling Terminology	Description	Contextualized for Data Storytelling
Story or Narrative	Consists of many subsystems working together (e.g, characters, plot, and theme). It is an account of events arranged in their time sequence [28].	A combination of visualized findings or messages with connections such as temporal or causal relations [31, 40].
Theme	A recurring idea. A story can have many themes.	These appear as concepts paired with an intention. (i.e., to inform, persuade, entertain, comfort, explain, or terrorise [50]).
Through-Line	A single theme that runs from the start to the end of a story. It interweaves the roles of the characters with the plot.	One primary concept and intention that drives the story and motivates the characters' actions [50].
Plot	A combination of events and how those events are revealed. The arrangement (or the sequence) of events in the story.	A causal relationship [44] between a set of events, depicted by the behaviors and interactions [47] of data-driven visual elements.
Event or Story Piece	Atomic element of a story. It may focus on one character's status or behavior, or the relationships among multiple characters.	"Story pieces" [44] that are derived from data, provided by data analysts and domain experts, e.g., data facts and human insights.
Genre	Classification and organization of works into categories.	Magazine, annotated chart [11, 37, 52], partitioned poster, flow chart, comic strip [5], slide show [68], and film/video/animation [1, 9, 72]
Setting	A cluster of actual states of affairs or various events where the story takes place.	The devices and location where the story is presented [44] and the various scientific domains [47] where the insights were derived from.
Storytelling	To both share and provide an experience to an audience. To give the audience a form of knowledge [26, 67] that is both emotional and entertaining.	This is the same, with an additional condition that the story must be based on the data.
Audience	The targeted group of people who will receive the story	This is the same in data storytelling.

Table 1: Terminology with the definitions contextualized for data storytelling. These terms and their mappings were derived from a breadth of visual and written literature [26–28, 67], corroborated with the narrative and data storytelling literature [40, 44, 47, 50, 58, 61].

data facts or human insights, should be identified beforehand so that the data storyteller can figure out what to share.

Merely having a set of story pieces to report does not give a story. Through the relationships among story pieces (e.g., causality), the data storyteller can organize the story pieces into a story plot for presentation. However, if the storyteller is unclear on the relationships or why the subject matter is worth presenting, the audience may feel similarly when consuming the story. The organization stage is where data storytellers select story pieces, sequence them into a story plot, and leave with a structured outline for the story. Several frameworks offer systematic ways to sequence a story plot for various genres [5, 36, 50, 58, 72]. Here, we review the relevant literature on story plots from two aspects: the narrative structure and the communicative goal.

There is a popular set of three narrative structures [58] for data storytelling: the drill-down, the martini glass, and the interactive slideshow, which have been heavily discussed in this space. Recently, Yang et al. [72] provide guidelines for applying Freytag's Pyramid, a narrative structure that has been widely used in film and literature, to data stories. These structures are similar to what Truby describes as the *story movement* in the space of written storytelling [67]. Truby describes a set of narrative structure patterns that storytellers draw from nature to connect elements in a sequence (e.g., linear, meanderings, spiral, and branching). For data stories, the narrative structure addresses the story movement, yet the pace of this movement is often influenced by the audience's interactions [58]. For example, the drill-down structure can present multiple plot branches and even different stories together, thus enabling the audience to explore the story. Meanwhile, the alternatives often present a single linear narrative, which is easier for the author to maintain the plot consistency, and therefore has been the focus of automated reasoning [38, 42]. In terms of the communicative goal, Adegboyega and Heravi [50] identify seven types of story plots from data stories. Their work provides a great jumping-off point for story creation. In section 7, we expand on their work and describe generically how data characters may weave within these seven story plots.

It is important to note that the genre [58] and the setting [44] of a story should also be settled at this organization stage. Determining a genre (or genres) of the story can help data storytellers narrow down the choices of the ideal story plot and narrative structure, aligning with their intentions better. Segel and Heer [58] identify a set of genres for data stories: magazine, annotated chart [11, 37, 52], partitioned poster, flow chart, comic strip [5], slide show [68], and film/video/animation [1, 2, 9, 72]. Akin to other storytelling fields, these genres can be paired with each other. Many of these genres have been further developed, and methodologies and frameworks for using them have been offered. Refer to Tong et al. [66] for more details.

During the organization stage, the data storytellers may begin assess-

ing which visual elements will best achieve their communicative goals and support advancing the story. Understanding the relevancy of each visual element to the story theme often can help the data storytellers locate important story pieces [36]. Irrelevant story pieces in a story can become multiple story sub-plots. Without proper curation, the story likely encounters issues at the presentation stage, such as the audience struggling with comprehending the story and losing interest [7, 23]. This motivates our work to investigate how visual elements can be modeled as data characters. The characterization of story pieces with data characters may help reinforce the through-line and construct the story plot, supporting the data storytellers in delivering a clear message to the audience with their stories.

3.2 Characters in Data Stories

Prior research has addressed character-based [16, 17] or character-driven storytelling [12, 18]. Cai et al. [12] address the importance of character- and plot-driven storytelling and offer a hybrid system involving both approaches. Cavazza et al. [16–18] focus on character-driven stories and propose an engine that models character behaviors. Their work has a clear focus on automatic narrative generation, where characters in all of these works are presented as either human or virtual agents. These works focus on improving stories that either are rooted in traditional storytelling (i.e., not data-driven and allow for more creative freedoms of the author) or express the story content using virtual humanoid agents. However, data stories often are not expressed using human or virtual agents but rather through abstract representations. Data stories are also often rooted in non-alterable non-fiction, constraining the storytelling process. Our work contends that we should view and think of abstract visual representations as data characters, similar to how prior research applies techniques to virtual agents.

To provide recent examples of what we consider an effective data character, we look to the popular storyline visualization [6, 32, 64]. This visualization depicts an abstract line encoding that represents the progress of an entity during a temporal period. We can view this encoding as a character, drawing the attention of an audience. Furthermore, it can be used as a through-line and an entry point into understanding the change and growth of a single data-driven entity. Storyline visualizations contain multiple storylines, thus multiple characters, illustrating their relationships. Other examples include visualizations with properties that make them more identifiable or personable, such as anthropographics [8, 49]. These representations were shown to be effective in creating an emotional connection with underlying content [8, 43], while other evidence revealed their limitations of eliciting a specific emotion from an audience [46]. These works motivate a need for the data storytelling community to translate devices and structures from visual storytelling (i.e., film literature). We feel data characters may serve
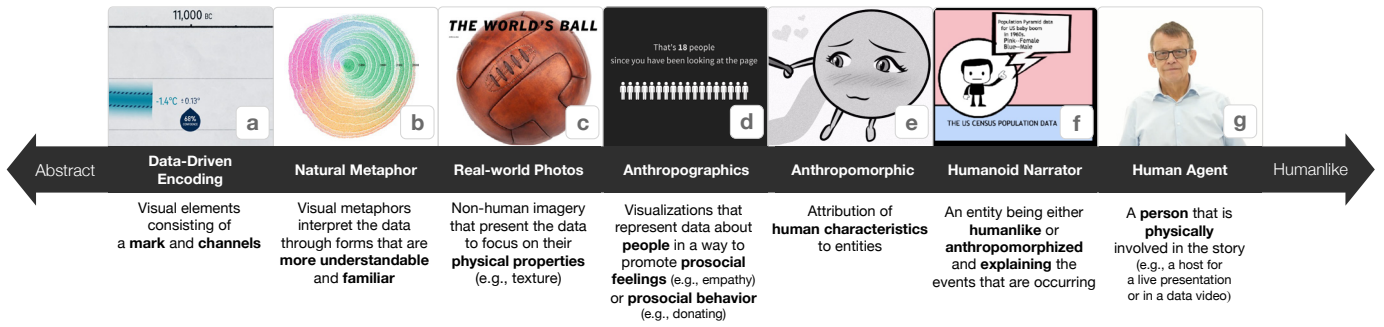
Fig. 2: Identifiable Central Character (ICC). In character-oriented visual data storytelling, we require a character that is central to the story and can be visually identified by the audience. From our analysis, we found this character's visual depiction can range from an abstract representation to a person. (a) a visual encoding such as a line to present the temperature [33] (b) Tree rings [21] showing the immigration (c) Soccer ball photos to show the texture progression in panels [65] (d) the anthropographic people [57] denote the loss to COVID-19 (e) a character that represents productivity from a data comic [71] (f) a humanlike character explaining the content [74] (g) Professor Hans Rosling [30] starring in a data video.

as a stepping stone to developing and eliciting an emotional connection with the audience, yet to be investigated in future work.

A data character can be the visual elements that are "performing" the data story [47]. There are some direct references [47, 61, 74] on the role a data character assumes; however, a story can not exist without characters. When designing data characters, we are looking for visual entities that relate to the theme of the story and advance the story plot. A data character could naturally take on rules of visual encoding, the well-established combination of marks and channels with their respective mappings to data. A data character should consider properties that delve deeper into the communicative effort and that help the audience bridge the gap between science and the story. Thus, a foundation for the properties of a data character can be derived from works that address visualization design [8, 49, 61, 63, 74] and visual metaphors [13, 21, 22, 45, 51, 54, 55, 69]. There is limited research that addresses the role of a data character in data storytelling. In the following section, we provide a framework to classify data characters and their behaviors.

4 Data Character Framework

In this section, we describe the development of our framework for data characters that was derived through discussions with storytelling experts and from the analysis of existing data stories. In section 5, we present our framework for how we classify data characters and the behaviors we identified. We then describe a space for how we can develop a data character and apply it within a story. We refer to this space as *character-oriented design*, to be elaborated in section 6 and section 7.

4.1 Framework Derivation

To better understand the nature of data characters, the forms that they assume, and their behavior in data storytelling, we created a corpus of 160 data stories and analyzed these stories using a codebook we developed. The corpus was made by merging several other corpora [35, 43, 48, 73] and updated to include newer stories. The merged corpus size was reduced based on repeats and any stories that were no longer accessible (i.e., required Adobe Flash Player or went offline). We further filtered down the set to primarily visual stories. In these stories the content and core messages are communicated via data-driven visualizations, animations, or videos. We excluded those heavily dependent on multiple paragraphs of text, where the visualizations served as annotations or figures rather than the driving narrative force. Newer stories were sourced from either accredited data storytelling sites (e.g., Bloomberg) or the Information is Beautiful awards [24]. We also added data stories that addressed minority domains in the merged corpus.

To develop the codebook, which we used to analyze our data story corpus, we consulted with experts in literary media and data storytelling, including published authors. Through multiple open-ended consultations and discussions, we gained an understanding of their process of character creation and character development. These insights laid the foundation for our codebook, including identifying a fundamental set of character roles and their effects on the narrative. We then shared the codebook with two separate groups of data visualization and data storytelling experts, whose feedback helped further translate and contextualize these insights for data storytelling. In each session, we exchanged our findings and translations from the literary consultations. Each session took approximately 2–3 hours. This process broadened the dimensions of the codebook. For example, we include a new type, real-world photos, for identifiable central characters, to be illustrated in subsection 5.1. We also expand the types of **antagonist**, to be elaborated in subsection 5.3. The visualization experts and data story practitioners validated the finalized codebook. Using the codebook, two researchers independently coded the corpus. They met for three sessions to compare and discuss any mismatches until reaching a consensus. The finalized codebook and corpus are on our website [2].

From our analysis, we derive that the role that visual elements and visualizations serve in data storytelling falls into two states: (1) *given an existing visualization, how can we adapt it to emphasize and explain a finding* [62] and (2) *given a finding, how can we visualize it* [60]. Both states share the communication goal of data storytelling; however, they present two different starting points for what we view as characters. The former has an existing character that will be developed and altered to depict the plot, whereas the latter starts from scratch. This work delves deeper into the former. We investigate how the 160 data stories would be framed with data characters. We find that the constant involvement of a data character in all the story pieces can support the audience in seeing and tracking continuity in a story. Future work may investigate concrete strategies for maintaining and presenting the connections between story pieces (i.e., through-line) in a story, including how to elicit an emotional connection to the stories.

5 Character Roles & Behaviors

In written storytelling [14, 67], there are many character roles in addition to the main character, such as a deuteragonist or a love interest. In the simplest form of a story, there would be a protagonist, a character that drives the plot forward, and an antagonist, a character that stands in direct opposition to the protagonist [14, 67]. For simplicity, in this work, we only focus on just three essential roles: **main**, **supporting**, and **antagonist** characters. It is documented that a main character need not be the protagonist, as often there are other characters that can advance the story, even the antagonist is included [14]. In a data story, the protagonist could be the audience, as they interact with data stories and can be the ones that drive the story plot forward, whereas the main character remains a visual element in the story. Future work may investigate the nuances and interplay of characters and potentially data character roles.

---

[2]https://chaorientdesignds.github.io/

5.1 Main Character

The core of our data characters is a concept. While concepts are often intangible ideas, let alone visual, this cannot be the case for the **main character (MC)**. After all, data storytelling is a visual medium. The **MC** is the device that the audience can rely on to make sense of and contextualize what is transpiring; therefore, the **MC** should be visually present in the story. More precisely, we need a visual element that can be identified by the audience as the center of the story, to which we refer as an **identifiable central character (ICC)**. The ICC becomes the vehicle that visually navigates the audience through the story plot to the conclusion. After reviewing our corpus of data stories, we find the visual representation of the **MC** (i.e., potential ICCs) ranges from an abstract representation to a real person. As shown in Figure 2, we present the common forms we identified from our analysis.

The ICC can be any visual encoding, visualization, or set of visual elements, so long as it is central to communicating the core message. It is up to the author on what visual representation will suit their narrative needs the best. For example, in the data story about the U.S. immigration from Cruz et al. [21], the ICC is a natural metaphor, as shown in Figure 2b. The story begins with the authors revealing this data-driven natural metaphor without any context; the audience thus may not be able to understand the inherent meaning.

When the story begins, our **MC** starts in the "ordinary world" or an expected state [14]. In this initial state, the audience should be primed with the relevant context and understanding to decode what is visually presented. The introduction of the character must get across the relevant context, and background information as well as decoding information to relate the visual element to what they mean. Furthermore, it sets up the motivation and purpose of this character. It should also introduce doubts, the uncertainty, and begin to challenge the character in its purpose. This context drives the data character's actions and sets up expectations for the audience. For example, the U.S. immigration story opens up by introducing a large **tree ring** with a description of how it represents U.S. immigration. However, the explanation for what the colors mean is initially left unanswered, prompting the audience to scroll onward to learn more about what they are seeing.

The behaviors of a character are driven by its desire to reach a personal goal [14]. The desire is determined by the author by factoring in their intent, the through-line, and the story pieces. The lens of the audience into a story is through the **MC**. The **MC** has a **desire** that it seeks throughout the course of the story. For simplicity, the other two characters either aid the **MC** in attaining its desire or are in opposition. Those in opposition will bring the **MC** to change and evolve throughout the story, and the audience will learn what those changes mean. In the case of Figure 2b, the **MC** may desire to *inform* the audience about population growth and demographic changes over the years in the United States. As the audience scroll onwards, the authors introduce more characters in the story, illustrated in the following subsection.

Although it is ideal for a **MC** to also be the sole identifiable central character, ICC, there are exceptions. With other forms of storytelling, technically, there can be identifiable characters that are central to the story (e.g., *The Lord of the Rings*). It is important to be cautious when having multiple ICCs in a story, as they will compete for the attention of the audience. This may result in issues with the delivery of the data storyteller's core message. Introducing more characters will increase the complexity of the story, but depending on the communicative goal, it can be effective in conveying the intended message.

5.2 Supporting Characters

A **supporting character (SC)** brings out dimensions of the **MC** and helps push it towards its desire. **SCs** must have some relationships with either the **MC** or the **antagonist**, but do not require a relationship among themselves, as shown in Figure 6. For simplicity and the sake of clear story illustration, we recommend not to have a **SC** that supports both the **MC** and the **antagonist**; however, it is possible (e.g., the betrayer in storytelling). The **SC** needs not actively work towards supporting the **MC** to pursue its desire; rather, it should never intentionally impede the **MC**. We found **SCs** in data stories are often
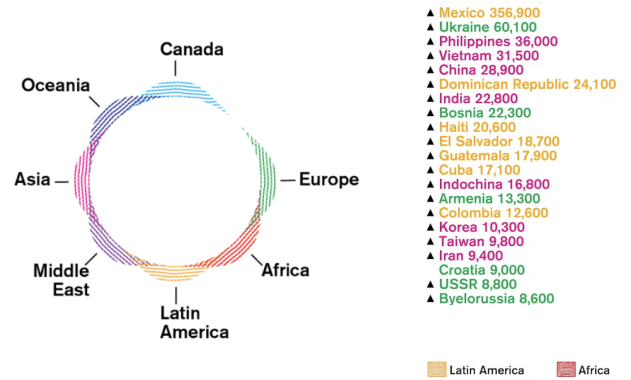

Fig. 3: An example of two **SCs**. The **MC** is a natural metaphor illustrating the U.S. immigration [21]. The **SCs** of this story include a legend and additional details of where immigrants are emigrating from. These **SCs** provide extra information for the audience to better understand the **MC**.

tasked with providing missing context, extra information, and even alternative representations of the data.

Continuing with the story of the U.S. immigration, we left off with the introduction of the **MC**, a data-driven natural metaphor of a **tree ring**. As the story advances, we encounter other characters that offer *extra information* and *alternate representations* to help the audience gain a deeper understanding of the **tree ring**. As shown in Figure 3, we see a **legend** that decodes the visual encodings, allowing the audience to understand what the color means and infer more insights about the **tree ring**. The **tree ring** is also accompanied by a **list of locations**, providing extra information on the geographic origins of individuals immigrating to the U.S. These devices help the audience understand and interpret our **MC** as the story progresses.

5.3 Antagonistic Character / Force

The difference between an **antagonistic force (AF)** and an **antagonist character (AC)** is that an **AF** is an ethereal presence that exists but is not seen directly, whereas an **AC** is visually present in the story. We found that when the **antagonist** appears as a visual character in data stories, it is often the **MC** driving the story forward. As shown in Figure 4-(d), there is a depiction of an **AC** in the data story "Out of Sight" about the U.S. drone strikes in Pakistan. The **AC** in this story are the grey lines that symbolize individual U.S. drone strikes.

If there is no **antagonist**, the **MC** will achieve its desire unimpeded. In the context of visual data storytelling, the **antagonist** needs to be present to stop the **MC** from pursuing its desire. *How can a visualization have a "villain" or "antagonistic force", and what does that imply?* In our framework, the **AF** can represent *misunderstandings* and *misconceptions*. A misconception is a mistaken belief or having the wrong idea. For example, it is a misconception to hold the belief that the earth is flat. A misunderstanding is having different interpretations of a meaning. As an example, thinking that a rainbow color map implies weather data would be a misunderstanding. Throughout the story, these **AFs** or the **AC** prevent the **MC** from achieving its resolution. In the context of the U.S. immigration data story, the antagonist is not visually present and therefore is an **antagonistic force**. The force manifests as misconceptions being hurled at the **MC** about a lack of context and understanding of where people immigrated from.

From our analysis, we find the **antagonist** in data stories often is a force. We further identify three forms that these **AFs** assume to represent misconceptions and misunderstandings — *lack of value*, *lack of context*, and *lack of trust*.

Lack of value. The **AF** creates external conflicts with the characters in the form of questions, such as "why is the design meaningful?" or "why do microtubules matter?", driving the story plot to address these questions. In a data story about colorized math equations [3], as shown in Figure 4-(a), the **AF** is constantly questioning the value of the design and its usability. As a result, the story attempts to motivate and demonstrate the design is effective.
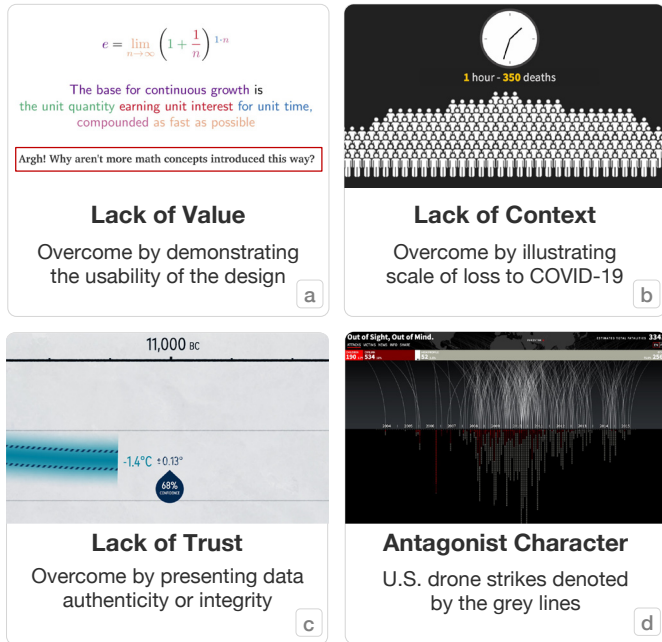
Fig. 4: Examples of **Antagonist** Forces (a–c) and Character (d) in DS. (a) The story about colorized math equations [3] argues the design **value**. (b) The story "Pace of Death" [57] provides the **context** for how many have died from COVID-19. (c) The video "Degrees of Uncertainty" [33] attempts to demonstrate the **integrity** of the data analysis for climate change. (d) The story "Out of Sight, Out of Mind" [39] depicts the antagonist as drone strikes.

Lack of context. The characters face external conflicts due to missing information, leading to a misunderstanding or a complete misconception. Often this kind of conflict is resolved by showing the "scale" of a phenomenon (e.g.,, the scale of loss, the scale of gun violence, or the scale of climate change). For instance, as shown in Figure 4-(b), the data story "Pace of Death" addresses the **AF** by depicting the number of people who passed away, in a given time interval, due to COVID-19.
Lack of trust. The **AF** introduces internal conflicts where the characters must prove or refute claims on their integrity (e.g., data authenticity, uncertainty, or credibility). In our analysis, we find this can be seen in the form of uncertainty visualizations or visualizations that delve into how the "black box" of a machine learning model behaves.

In all these instances, the **antagonist** is constantly attempting to challenge the **main character** and introducing conflicts to prevent the **MC** from reaching its desire.

5.4 Conflict & Tension

The rationales behind developing and understanding data characters include (1) structuring story content via characters helps filter out irrelevant information and leaves the data storytellers only with the content serving their communicative goal and (2) it unlocks the device of *conflicts*, akin to the debate in the context of science-based works. The difference between an explanation and a story is a conflict. Here, we illustrate more on internal and external conflicts.

Why do we want conflicts in a science-based story? Wouldn't that obfuscate, if not detract from, the messages we want to get across? The conflict is a device that helps the audience understand the motivation of the story and can lead to an appreciation for the endeavor or relevancy of what is being communicated. In other forms of storytelling, the conflict is used as a means to bring the audience into the story and to form an emotional connection to the characters, such that the audience cares or want the character to succeed in fulfilling its desire. From our analysis, we find the conflict assumes two forms, *internal* and *external* conflicts. Internal conflict exists when a character struggles with their own opposing desires or beliefs. It happens within the character and drives its development. External conflict sets a character against something (e.g., the **antagonist**) or someone (e.g., the reader)
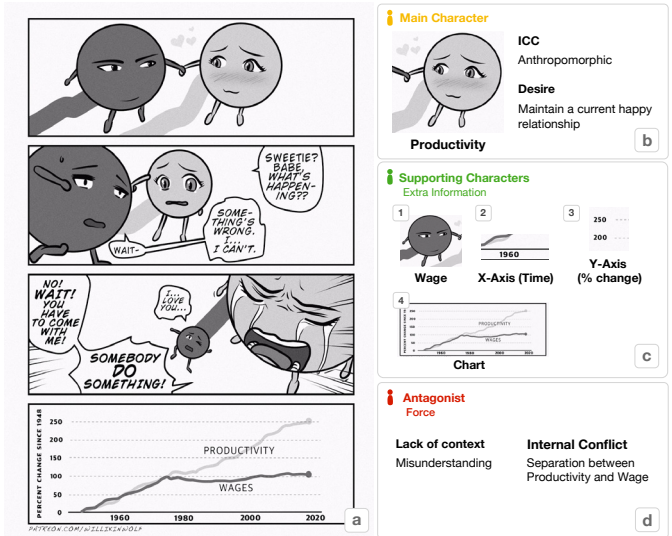

Fig. 5: The data comic "Something's wrong", created by Willikin Woolf [71] and included in the data comic gallery [4]. (a) Original data comic (b) The **MC** – **productivity** (c) The **SCs** – (c1) **wage**, (c2) **x-axis** with the year, (c3) **y-axis** with percentage changes, and (c4) the line **chart** with the same colors as **productivity** and **wage**. (d) The antagonist is an unaddressed force (**AF**), i.e., the reason behind the separation.

beyond its control. External forces stand in the way of a character's motivations and create tension as it tries to reach its goal. A majority of the data stories that we review tend to contain external conflicts where the story characters are fighting external forces to either refute claims or provide the missing context to remove a misconception or misunderstanding. The story resolves once the conflicts are addressed.

5.5 Data Comic – "Something's wrong"

To better understand data characters, their roles, and the conflict, we will go through a data comic and address the questions as follows.
- Who is the **MC**?
- Does the story have ICC? If so, what kind of ICC it is?
- Who are the **SCs**?
- What is the **antagonist**? How does it introduce the conflict?
- What is the relationship between the plot and the characters?
- What is the main theme, i.e., the through-line?

Data comic [5], inspired by the visual language of comics, is a rising and popular genre for presenting information effectively. We will analyze a data comic *Something's wrong* by Willikin Woolf [71], featured on the data comic gallery curated by Bach et al. [4]. This story depicts the relationship between two identifiable comic characters. The **MC** is likely the "productivity" character, as shown in Figure 5-(b). This is because from Figure 5-(a) panel 3, we see the panel focuses more on the despondent **productivity**, rather than the "wage" character. We may infer that the desire of **productivity** is to maintain the current happy relationship with "wage", as seen in Figure 5-(a) panel 1. The visual representation of the **MC** (i.e., ICC) is anthropomorphism.

The **SCs** are wage, the x- and y-axes, and the chart, indicated in Figure 5-(c1) to (c4). As shown in Figure 5-(a) panels 1-3, **wage** seems eager to stay with **productivity**. Meanwhile, the **axes** and the **chart** are only introduced in the last panel. The **axes** help us understand more about **wage** and **productivity** for providing the additional context of their percentage change over time, from similar trends to drastic separation. This is because the colors of both characters correspond to the line colors in the **chart**.

The **antagonist**, in this story, would be the reason behind this separation. Visually absent in the story, this **AF** remains unaddressed or unexplained either, at least on this page of the comic. The **AF** introduces the conflict to **productivity** in the form of a separation from **wage**. The through-line between all these events is to explain the relationship between **wage** and **productivity**. A possible theme could be
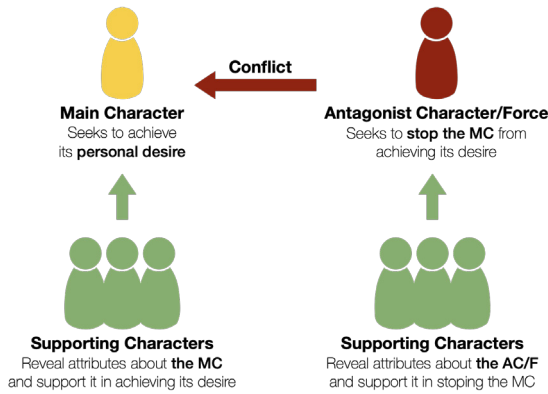
Fig. 6: Character Web. The relationships between main, supporting, and antagonist characters. The main character (MC) has a desire and is the focus of the story. The antagonist force (AF) tries to prevent the MC from achieving its desire. Supporting characters (SC) reveal dimensions about either the MC or AC/F.

to persuade the audience into thinking that these two characters should not drift apart and have a linear relationship.

6 CHARACTER ORIENTED DESIGN SPACE

From our framework and analysis of data stories, we have unpacked a data character, identified common representations of data characters, defined properties of basic data character roles, and contextualized conflict in data storytelling. In this section, we describe how data characters, as defined in this work, can be developed as well as how they could fit into the broader data storytelling space. When we refer to this space as *character-oriented design*, we intend for it to be used as a guide for helping data storytellers in developing or identifying characters in their stories.

6.1 Character Motivation

To begin, we first outline our process for character creation, which was also informed by the aforementioned discussions with experts. We consider a **character** as the lens for our audience to view the abstract, which often is a complex concept, we intend to convey. How a character is defined determines both the presentation of the data and the story focus. A data character is a derivation of a character, thus it must adhere to some core storytelling principles: (1) a data character must serve a narrative goal and (2) a data character must have a **desire**. We define **desire** as the rationale that motivates a character pursuing a goal. By assigning a desire to a data character, we should ensure the behaviors of the character are associated and align with this base desire, throughout the data story. That is, the character is consistent in its behaviors as it pertains to achieving this desire. Contextualized for data storytelling, this desire is the communication of some scientific concept, linked to an existing dataset with varying intentions. We offer the following steps as a suggestion for developing a data character and their **desire**.

1. Identify the meta-concept (through-line).
2. Distill the concept into smaller more distinguishable aspects.
3. Relate these aspects to the data (i.e., via examples or non-examples).
4. Rank which aspects best illustrate the concept.
5. Select the visual representation.

To not lose sight of the original intention of the story, and to track whether our characters are in line with their desires, it is good to make use of a through-line (Table 1). As discussed in subsection 3.1, data stories can be composed of many plots and subplots, steering the original narrative in many directions. A through-line in data storytelling would be the core concept that ties these subplots together, providing guidance and relating back to the original narrative. Identifying a through-line requires the assessment of the story pieces for what common theme or over-encompassing plot could link these pieces all together. If we have five disjoint story pieces, then the through line has to be the commonality among the pieces. Alternatively, if we cannot find a common link

between these pieces, we should then consider re-evaluating what we seek to convey. In our case, for data storytelling, this commonality is a scientific concept that each individual piece relates to. We can utilize the seven story plot types [50] along with the seven genres [58] to identify a central theme, through-line, for our story. As for the situations where certain story pieces address disparate concepts, one of them must be prioritized. The remaining groups should either be omitted [36] or be part of another story.

Premise & Through-Line: When organizing a data story, the first step is to identify what is the through-line between our story pieces. Namely, *what is the meta-concept that ties the messages we want to convey?* If we know our through-line, we may understand how our characters relate to the plot. Then, we can create a premise, or a small road map, of the entire story. A premise is the simplest expression of the story being told and presents a sense of the MC as well as the outcome of the story.

Possibilities: Once we have an understanding of the premise and through-line, we should identify what is possible in this premise. What types of genres do we have access to, what structures can we make use of, and what setting will it take place in?

Challenges + Problems: Identifying the unique challenges and problems ties to conveying the content of our data story. These problems could be the audience lacking familiarity with what is being expressed, the subject matter is too abstract, there are difficulties simplifying the science, and so on. We need to note what the main misunderstanding or misconceptions about the story content are.

After answering these points, we have developed a foundation for our main character and supporting characters. The next challenge is determining what should be our main character and how to develop this from our through-line and antagonist.

6.2 Character Creation

Naively, a data character inherits the properties of a data-driven visual element It would likely contain a set of attributes and behaviors that relate to explaining the concept. A character can be a lens into a meta-concept or scientific domain. Initially, a character represents an idea in a world full of misconceptions, misunderstandings, and uncertainty. The role of a character in the story is often to explore its relationship to this perceived world and attempt to overcome misconceptions, asserting its place. Contextualized for communicating scientific information, a data character would be born from data or namely analytics. As discussed in subsection 5.1, the motivation of a character is its desire , which would be the core message to get across. Thus, it is important to understand what concepts represent a data character. Depending on the complexity of the core concept, we often may rely on multiple characters to best convey it. This is due to concepts having many dimensions that should be viewed and expressed with varying lenses. The concept we prioritize and give the focal lens would be the basis of a main character. Collections of smaller aspects that exemplify and illustrate concepts as a collection would result in supporting characters. Aspects from the data thatmay challenge the concept (i.e., contradictions) can lead to potential antagonists.

We suggest the best starting point is the meta concept that contains the smaller sub-concepts, i.e., the through-line. Once we identify the meta concept, we may create sub-groups of the findings to express more distinguishable aspects of the meta concept. We then can investigate how each sub-group illustrates the aspects of the meta concept. Precisely, we inspect whether each sub-group serves a direct example, or perhaps a non-example (i.e., contradiction), of the meta concept. These relations allow us to rank each sub-group based on their capability of demonstrating these aspects. Some sub-groups are stand-alone, some aspects may show details of phenomena in others, while some may introduce doubt or uncertainty. This gives us a basis for the main, supporting, and antagonistic characters. This can also be the point where data storytellers begin to design the visual elements that will portray these characters.
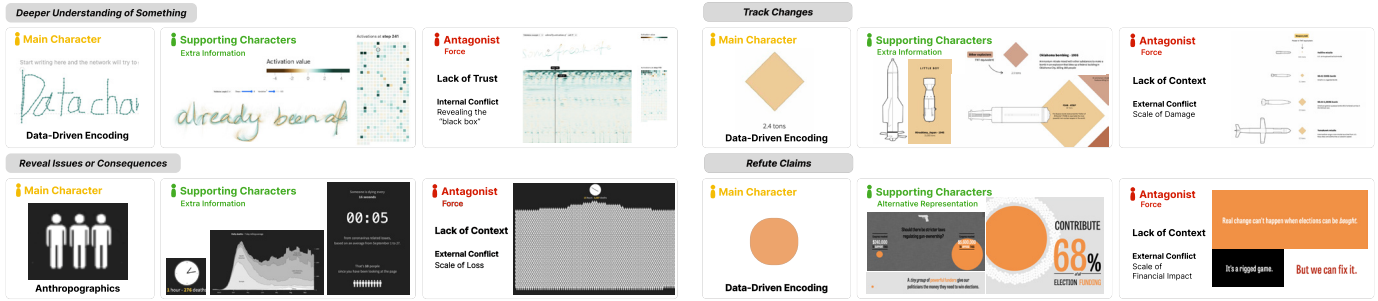
7 CHARACTERS IN STORY PLOTS

Fig. 7: Plot Types and Data Characters. For four of the five plot types, we illustrate through existing data stories how data characters appear and behave. The first column represents the **main character** followed by **supporting** and **antagonistic forces**. Image sources: [15, 19, 53, 57].

From our analysis of the corpus, we identify several patterns in how characters often behave in the stories. In this section, we present four concrete story examples to describe how the relationships of the character roles, as shown in Figure 6, could be contextualized in the story plots. Here, we first introduce the initial seven story plots for data stories, identified by Ojo and Heravi [50]:

1. Refute claims.
2. Reveal unintended consequences.
3. Reveal information of personal interest.
4. Enable deeper understanding of a phenomenon.
5. Reveal anomalies and deficiencies in systems.
6. Reveal information about an entity in increasing levels of detail.
7. Track changes in systems.

We may consider "phenomenon", "entity", and "system" as exchangeable main characters, depending on the communicative goal. Consequently, we collapse the second and the fifth plot types into *reveal anomalies, issues, or unintended consequences*. We merge the fourth and the sixth plot types into *deeper understanding of something*, as these two story plots explore a character (e.g. phenomenon or entity) in depth. By factoring in the character roles, we integrate these seven story plots into the following five types:

1. Refute claims.
2. Reveal anomalies, issues, or unintended consequences.
3. Reveal information of personal interest.
4. Deeper understanding of something.
5. Track changes.

In the remaining section, we describe how data characters may be woven into each of these story plots, including a generic through-line, the desire of **MC**, **AF**, and a possible premise of the story plot. We focus on **AF** as we find it to be the majority of **AF/C** in our corpus, whereas **AC** is rather intuitive due to its visibility. While these examples do not represent the only way to construct such a story plot, they illustrate how we may pair characters and story plots together to organize the story.

Refute claims. The through-line can be to persuade the audience into believing the inverse of the claim. The **MC** desires only to prove the inverse of the claim. The **AF** may take the form of misconceptions or misunderstandings that support the claim. Consequently, the conflict, when resolved by the **MC**, helps refute the claim. *Premise:* the **MC** heads towards the state, where the world is opposite of the claim. The **AF** causes conflicts to prevent the **MC** from reaching its desire. By overcoming the conflict, the **MC** fulfills its desire and refutes the claim.

For example, we examine the story, *Money Wins Elections*, as shown in Figure 7. The through-line is a claim that "money wins elections", with other concepts addressing corruption in the U.S. government. The **MC** is a **point mark** that represents a vote, where its size encodes financial investments. We may infer the claim the **MC** seeks to refute is that "all votes are equal" or that "nothing can decide election outcomes", as the **MC** desires to persuade the audience that money can buy election results. The **AF** is the misconception that causes external conflicts. The **MC** resolve the conflict by demonstrating that the scale of financial investment can influence the election outcome.

Reveal anomalies, issues, and unintended consequences. The through-line is theme-dependent and affected by the disposition of the author towards the consequences or unexpected events. The desire of the **MC** is to maintain the world state as expected. The **AF** introduces an action or event that creates conflict with the world **MC** expected. While the **MC** addresses this conflict, the result causes the **MC** to deviate from the direction it expected.

Referring to Figure 7, we see an instance of this type of story, *The Pace of Death*. The through-line is COVID-19, with a sub-concept of mortality rate. The **MC** is a person depicted by an anthropographic **icon**. There are several **SCs** providing extra information about the person, such as a **clock** illustrating how many people pass away after a period of time. The **MC** is initially under the impression that the mortality rate is not high. As the story advances, the **AF** reveals what the true mortality rate looks like. One of the **SCs** is a **timer** that reveals the scale of loss that has been incurred while the audience reads this story; thus, we may infer the primary intent of the author is to terrorize the audience to call attention to the severity of COVID-19.

Reveal information of personal interest. The through-line is often to inform the audience about some information. The desire of **MC** is to showcase the personal interest, which is the author's in this case [50], whereas the **AF** is to dismiss the value. In this story plot, the **MC** tries to bring attention to a topic and motivate its value. The **AF** introduces conflicts by casting doubt on the topic's value. By overcoming the conflict, the **MC** persists in showing why the topic is interesting.

Deeper understanding of something. The through-line is to explain something (e.g., a phenomenon) to the audience. The desire of **MC** is to explain the mechanisms behind the topic as simply as possible, whereas the **AF** are misunderstandings about this topic. The **MC** attempts to explain to the audience, however, the **AF** introduces hurdles based on misunderstandings. To overcome the conflict, the **MC** must unpack the content further, until the author feels the story goal is achieved.

In the story, *Four Experiments in Handwriting with a Neural Network*, the **MC** is the **handwritten phrase** of the user, as shown in Figure 7. The **MC** seeks to better understand how a neural network learns to the handwriting style. The through-line is about neural networks with sub-concepts in handwriting and generative models. The **AF** comes in the disguise of an internal conflict, the black box nature of neural networks. To overcome this conflict, the **MC** unpacks the black box, the layers of the neural network.

Track changes. The through-line is often to inform the audience of changes in a character (e.g., an entity or a system). The is the same for the desire of **MC**. However, **AF** presents obstacles to MC in the attempt of causing changes. In this story plot, the **MC** is trying to stay "unchanged" but is presented with obstacles, resulting in conflicts. To overcome these conflicts, the MC must change.

This story plot can be seen in the data story, *How powerful was the Beirut blast?*, as shown in Figure 7. The through-line is the devastation that an explosion causes. The **MC** is a data-driven **diamond mark** that represents an explosion measured in TNT equivalent (e.g., it starts at 0.01 tons). As the story progresses, the **mark** grows larger and larger, and the narrative changes to convey how different blasts compare. The **AF** for this story appears as a lack of context, a misconception of the magnitude of how devastating various bombs or explosive accidents are. The **MC** overcomes this conflict by comparing these incidents.

8 DISCUSSION

The goal of our work is to illustrate how characters can be utilized to frame abstractions and communicate insights or findings through the story plot. Through our framework, we have identified specific features, relationships, and roles of data characters. From these findings, we describe a space for developing data characters and applying them in data stories, which we refer to as a character-oriented design space. Within this space, we hope to motivate data storytellers to view visual elements as characters with narrative goals, rather than data-bound abstractions to be explained. By treating visual elements as characters, we organize visual element(s) into those that are driving the story (MC), supporting the story (SC), and those contribute via contradiction (AC/F). We further investigate how data characters weave into representative story plots for data stories. The scope of this work is to offer a framework for understanding data characters and a design space to serve as the foundation for developing data characters. Our work introduces new considerations for the data storyteller: (1) they must develop the characters to best tell their story, (2) the roles (i.e., MC, SC, AC/F) for their characters, and (3) the number of characters needed to tell their story. In this section, we discuss the audience as data characters and character roles that are specific to data stories.

Audience as a character. There are many character roles outside the ones discussed in this work, and stories can utilize these roles in a myriad of ways that we did not address. When data stories are interactive, it becomes more "reader-driven" as the audience is now a part of the story, often the protagonist [58]. A protagonist is defined as the character that moves the story plot forward. In this role, the audience is not always visually represented as a character, they may control the **main character** and even at times could be an **antagonistic force**, but may not be an ICC themselves. For example, the audience may play the role of the "devil's advocate" towards either the **main character** or the **antagonist**. In some instances, the audience can be visible and represented as an avatar or other ICC depiction, as shown in Figure 2. The audience can also serve as secondary key characters, known as a deuteragonist [14, 67], they can play a role akin to either the **main character**, **antagonist**, or even a neutral agent.

Data story specific roles. A potential direction for future work would be identifying character roles unique to data storytelling. A method that could be applied to identify such roles is character archetypes, as shown in Table 2, which are templates for generic characters based on certain types of behaviors and patterns. This device can give the storyteller guidance on the specific nature and behaviors that the MC, SC, or AC/F will take on. Archetypes differ in other storytelling media as they focus on types of people and the human condition. In novels and screenplays [14, 26, 67], some examples of archetypes include: father, queen, mentor, warrior, and lover. Often these archetypes come with strengths, inherent weaknesses, and understandable relationships to help build a story. However, often it is not a focus of data stories to discuss or explore the human condition.

From our analysis, we suggest a starting place for character archetypes would be systems, anomalies, entities, and phenomena. The properties of the data can indicate some behaviors. For example, hierarchy suggests depth, temporality may imply change, and spatiality could imply closeness or bonds. We may extend this thinking by taking data types into account, such as nominal, categorical, numerical, and their pairings. Multiple numerical datasets could contrast with one another. We discuss three potential data character archetypes; *overview*, *parental*, and *cluster*.

The archetype of the *overview* would be a character that knows the broad picture of the relationships between other characters, but need not give an opinion, similar to an observer. An example can be a visual analytic system for managing production lines. With stories centered around prediction, we can pull out a *parental* archetype, where the parent has multiple attributes presented as a formula to describe their child. The story could be interested in the parent-child relationship, what is the relationship's strength, how the child affects the parent, how external negative factors that the child faces also impact the parent, etc. This sort of archetype could be useful for explainable AI. Another potential character archetype would be *clusters*. A pairing of numerical

Character Terminology	Description
Character	A visual entity that influences itself and others and serves a narrative purpose.
Desire	The rationale that motivates a character's actions.
Conflict	Arises when a character while pursuing their desire faces an obstacle. The pressure that is applied to the MC forces change.
Archetypes	Patterns within an entity; the behaviors they exhibit that are essential in how they interact with others.
Main Character (MC)	MC has the central desire and serves to contextualize or introduce the story/domain.
Supporting Character (SC)	This character role complements either the MC or AC. They provide a means to see more depth about the MC or AC. The desire of SC can align with either the AC or MC, but will not go against it.
Antagonist Character or Force (AC/F)	AC seeks to prevent the MC from reaching their desire. It causes conflict with MC. The story plays out when the conflict is resolved.

Table 2: Character-specific storytelling terminology. These terms and their mappings were derived from a breadth of visual and written storytelling literature [14, 26–28, 67].

and categorical can result in clusters. They are commonly used in visualization to denote relationships and close associations. The properties of a cluster can reveal commonalities, uniqueness, and how they are affected by the change (e.g., does the group stay or split).

While there are many character roles identified in other forms of storytelling, it is unclear how relevant these are to data stories. This work lays the foundation by providing a fundamental set of data character roles (i.e., MC, SC, AC/F), investigates how they weave into data stories, and discusses how the space of data characters could be expanded. We hope this work presents an opportunity within the space of data characters to extend past the roles identified in this work.

9 CONCLUSION & FUTURE WORK

Our goal with storytelling is to engage the audience while preserving the scientific integrity of the content. A data story should provide an entry point for the audience, a cohesive plot, and a cast of characters to lead the audience along to where the storyteller intends. As storytellers, we want the audience to create an emotional connection with the story and leave with the intended message. By exploring the role data characters play, we believe characters present the pathway to this goal.

We review 160 data stories and identify features of data characters, the roles they assume, types of antagonists in data stories, types of conflicts for data stories, and the relationships data characters have among one another, and offer a framework for data characters and design space for developing characters and applying them in a data story. From the perspective of a data storyteller, we show the role of characters and their importance in data storytelling with consideration of where character design should occur. We introduce the idea of an identifiable central character (ICC) as a device that data storytellers can use to select their MC, illustrated through our case studies, the relationships between the MC, SC, and AC/F in the context of data stories, and provide an outline for how characters can be woven with five types of data story plots.

For future work, we suggest exploring general patterns in common data stories in terms of broadening character roles. We believe, by providing a discussion on the relationships between characters and data-driven visual entities, we can achieve a more consistent language among designers. Stories require a through-line to connect all the story pieces together cohesively. However, these connections should remain clear to the audience. This would require some identifiable central character or characters for the audience to contextualize the presented information and latch on to as the story unfolds. The relationship between the plot and the story is not complete without the characters. The effective weaving of the two gives us the story. Our desire with this manuscript is for the readers to view and think in terms of characters when creating either a data story, narrative, or explanatory visualizations.

FIGURE CREDITS

Figure 1 is a partial recreation of Fig. 2 from [44], which is in the public domain.

Figure 2 (a) Image credit: Neil Halloran, 2021, How Sure Are Climate Scientists, Really? As seen in https://www.youtube.com/watch?v=R7FAAfK78_M. (b) as seen in Fig. 9 from [21], creative commons license, an updated graphic can be seen in https://web.northeastern.edu/naturalizing-immigration-dataviz/.
(c) worldcupballs.info and Adidas, 2014, The World's Ball, as seen in https://www.nytimes.com/interactive/2014/06/13/sports/worldcup/world-cup-balls.html. (d) Manas Sharma, Simon Scarr, and Gurman Bhatia, 2020, The pace of death, as seen in https://www.reuters.com/graphics/HEALTH-CORONAVIRUS/DEATHS/xlbpgobgapq/index.html. (e) Willikin Wolf, 2019, Something's Wrong, as seen in https://twitter.com/WillikinWolf/status/1176006515968241665. (f) as seen in Fig. 10 from [74]. (g) Gapminder.org, creative commons license, as seen in https://www.gapminder.org/answers/will-saving-poor-children-lead-to-overpopulation/.

Figure 3 Image credit: Pedro Cruz, John Wihbey, Avni Ghael, and Felipe Shibuya, as seen in https://web.northeastern.edu/naturalizing-immigration-dataviz/.

Figure 4 (a) Kalid Azad, 2019, Colorized math equations, as seen in https://betterexplained.com/articles/colorized-math-equations/.
(b) Manas Sharma, Simon Scarr, and Gurman Bhatia, 2020, The pace of death, as seen in https://www.reuters.com/graphics/HEALTH-CORONAVIRUS/DEATHS/xlbpgobgapq/index.html. (c) Neil Halloran, 2021, How Sure Are Climate Scientists, Really? As seen in https://www.youtube.com/watch?v=R7FAAfK78_M. (d) Pitch Interactive, 2004, Out of sight, out of mind, as seen in https://drones.pitchinteractive.com/.

Figure 5 Image credit: Willikin Wolf, 2019, Something's Wrong, as seen in https://twitter.com/WillikinWolf/status/1176006515968241665.

Figure 7 (a) Shan Carter, David Ha, Ian Johnson, Chris Olah, 2016, as seen in https://distill.pub/2016/handwriting/. (b) Marco Hernandez, Simon Scarr, 2020, How powerful was the Beirut blast?, as seen in https://graphics.reuters.com/LEBANON-SECURITY/BLAST/yzdpxnmqbpx/index.html. (c) Manas Sharma, Simon Scarr, and Gurman Bhatia, 2020, The pace of death, as seen in https://www.reuters.com/graphics/HEALTH-CORONAVIRUS/DEATHS/xlbpgobgapq/index.html. (d) Tony Chu, 2014, Money Wins Elections, as seen in http://letsfreecongress.org/.

ACKNOWLEDGMENTS

REFERENCES

[1] 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 ACM Conference on Human Factors in Computing Systems*, pp. 1459–1468. ACM, New York, NY, 2015. doi: 10.1145/2702123.2702431 3

[2] 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. doi: 10.1109/TVCG.2016.2598647 3

[3] K. Azad. Colorized math equations. https://betterexplained.com/articles/colorized-math-equations/, 2019. 5, 6

[4] B. Bach. https://datacomics.github.io/. Accessed: 2022-02-22. 6

[5] B. Bach, Z. Wang, M. Farinella, D. Murray-Rust, and N. Henry Riche. Design patterns for data comics. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–12. IEEE, New York, NY, 2018. doi: 10.1145/3173574.3173612 3, 6

[6] T. Baumgartl, M. Petzold, M. Wunderlich, M. Hohn, D. Archambault, M. Lieser, A. Dalpke, S. Scheithauer, M. Marschollek, V. M. Eichel, N. T. Mutters, H. Consortium, and T. V. Landesberger. In search of patient zero: Visual analytics of pathogen transmission pathways in hospitals. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):711–721, 2021. doi: 10.1109/TVCG.2020.3030437 3

[7] J. Boy, F. Detienne, and J.-D. Fekete. Storytelling in information visualizations: Does it engage users to explore data? In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 1449–1458. ACM, New York, NY, 2015. doi: 10.1145/2702123 1, 3

[8] J. Boy, A. V. Pandey, J. Emerson, M. Satterthwaite, O. Nov, and E. Bertini. Showing people behind data: Does anthropomorphizing visualizations elicit more empathy for human rights data? In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 5462–5474. ACM, New York, NY, 2017. doi: 10.1145/3025453.3025512 3, 4

[9] J. D. Bradbury and R. E. Guadagno. Documentary narrative visualization: Features and modes of documentary film in narrative visualization. *Information Visualization*, 19(4):339–352, 2020. doi: 10.1177/1473871620925071 3

[10] R. Bran. Message in a bottle telling stories in a digital world. *Procedia-Social and Behavioral Sciences*, 2(2):1790–1793, 2010. doi: 10.1016/j.sbspro.2010.03.986 2

[11] C. Bryan, K.-L. Ma, and J. Woodring. Temporal summary images: An approach to narrative visualization via interactive annotation generation and placement. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):511–520, 2016. doi: 10.1109/TVCG.2016.2598876 3

[12] Y. Cai, C. Miao, A.-H. Tan, and Z. Shen. A hybrid of plot-based and character-based interactive storytelling. In *Technologies for E-Learning and Digital Entertainment*, pp. 260–273. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. 3

[13] Z. Cai, Y.-N. Li, X. S. Zheng, and K. Zhang. Applying feature integration theory to glyph-based information visualization. In *IEEE Pacific Visualization Symposium*, pp. 99–103. IEEE, New York, NY, 2015. doi: 10.1109/PACIFICVIS.2015.7156363 4

[14] J. Campbell. *The hero with a thousand faces*, vol. 17. New World Library, 2008. 2, 4, 5, 9

[15] S. Carter, D. Ha, I. Johnson, and C. Olah. Experiments in handwriting with a neural network. *Distill*, 2016. doi: 10.23915/distill.00004 8

[16] M. Cavazza, F. Charles, and S. J. Mead. Characters in search of an author: Ai-based virtual storytelling. In *International Conference on Virtual Storytelling*, pp. 145–154. Springer, 2001. 3

[17] M. Cavazza, F. Charles, and S. J. Mead. narrative representations and causality in character-based interactive storytelling. *Proceedings of CAST 2001*, pp. 139–142, 2001. 3

[18] F. Charles, S. J. Mead, and M. Cavazza. Character-driven story generation in interactive storytelling. In *Proceedings Seventh International Conference on Virtual Systems and Multimedia*, pp. 609–615. IEEE, 2001. 3

[19] T. Chu. Money wins elections. http://letsfreecongress.org/, 2014. 8

[20] E. Coats. Pixar's 22 rules of storytelling. https://www.aerogrammestudio.com/2013/03/07/pixars-22-rules-of-storytelling/. Accessed: 2022-02-22. 2

[21] P. Cruz, J. Wihbey, A. Ghael, F. Shibuya, and S. Costa. Dendrochronology of u.s. immigration. *Information Design Journal*, 25(1):6–20, 2019. doi: 10.1075/idj.25.1.01cru 4, 5, 10

[22] K. Dasu, T. Fujiwara, and K.-L. Ma. An organic visual metaphor for public understanding of conditional co-occurrences. In *2018 IEEE Scientific Visualization Conference (SciVis)*, pp. 1–5. IEEE, 2018. 4

[23] K. Dasu, K.-L. Ma, J. Ma, and J. Frazier. Sea of genes: A reflection on visualising metagenomic data for museums. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):935–945, 2020. doi: 10.1109/TVCG.2020.3030412 1, 3

[24] Data Visualization Society. Information Is Beautiful Awards. https://www.informationisbeautifulawards.com/. Accessed: 2023-06-29. 4

[25] N. M. Dudukovic, E. J. Marsh, and B. Tversky. Telling a story or telling it straight: The effects of entertaining versus accurate retellings on memory. *Applied Cognitive Psychology*, 18(2):125–143, 2004. doi: 10.1002/acp.953 1

[26] S. Field. *Screenplay: The foundations of screenwriting*. Delta, 2005. 2, 3, 9

[27] E. J. Fink. *Dramatic story structure: A primer for screenwriters*. Routledge, 2014. 3, 9

[28] E. M. Forster. *Aspects of the Novel*. RosettaBooks, 2010. 1, 2, 3, 9

[29] C. Fritz and G. Tosello. The hidden meaning of forms: methods of recording paleolithic parietal art. *Journal of archaeological method and theory*, 14(1):48–80, 2007. doi: 10.1007/s10816-007-9027-3 1, 2

[30] GAPMINDER.ORG. https://www.gapminder.org/answers/will-saving-poor-children-lead-to-overpopulation/. Accessed: 2022-02-22. 4

[31] N. Gershon and W. Page. What storytelling can do for information visualization. *Communications of the ACM*, 44(8):31–37, 2001. doi: 10.1145/381641.381653 3

[32] R. Gove. Automatic narrative summarization for visualizing cyber security logs and incident reports. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1182–1190, 2021. doi: 10.1109/TVCG.2021.3114843 3

[33] N. Halloran. How sure are climate scientists, really? https://www.youtube.com/watch?v=R7FAAfK78_M, 2021. 4, 6

[34] J. Heer and G. Robertson. Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1240–1247, 2007. doi: 10.1109/TVCG.2007.70539 1

[35] F. Hohman, M. Conlen, J. Heer, and D. H. P. Chau. Communicating with interactive articles. *Distill*, 5(9):e28, 2020. 4

[36] J. Hullman and N. Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2231–2240, 2011. doi: 10.1109/TVCG.2011.255 1, 3, 7

[37] J. Hullman, N. Diakopoulos, and E. Adar. Contextifier: automatic generation of annotated stock visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2707–2716. ACM, New York, NY, 2013. doi: 10.1145/2470654.2481374 3

[38] 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. doi: 10.1109/TVCG.2013.119 1, 3

[39] P. Interactive. Out of sight, out of mind. https://drones.pitchinteractive.com/. 6

[40] P. Isenberg, B. Lee, H. Qu, and M. Cordeil. Immersive visual data stories. In *Immersive Analytics*, pp. 165–184. Springer, New York, NY, 2018. 1, 2, 3

[41] J.-A. Kang, S. Hong, and G. T. Hubbard. The role of storytelling in advertising: Consumer emotion, narrative engagement level, and word-of-mouth intention. *Journal of Consumer Behaviour*, 19(1):47–56, 2020. doi: 10.1002/cb.1793 1

[42] Y. Kim, K. Wongsuphasawat, J. Hullman, and J. Heer. Graphscape: A model for automated reasoning about visualization similarity and sequencing. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 2628–2638. ACM, New York, NY, 2017. doi: 10.1145/3025453.3025866 3

[43] X. Lan, Y. Wu, Y. Shi, Q. Chen, and N. Cao. Negative emotions, positive outcomes? exploring the communication of negativity in serious data stories. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3517530 3, 4

[44] 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. doi: 10.1109/MCG.2015.99 1, 2, 3, 10

[45] Y.-N. Li, D.-J. Li, and K. Zhang. Metaphoric transfer effect in information visualization using glyphs. In *Proceedings of the International Symposium on Visual Information Communication and Interaction*, pp. 121–130. ACM, New York, NY, 2015. doi: 10.1145/2801040.2801062 4

[46] J. Liem, C. Perin, and J. Wood. Structure and empathy in visual data storytelling: Evaluating their influence on attitude. In *Computer Graphics Forum*, pp. 277–289. Wiley Online Library, 2020. 3

[47] K.-L. Ma, I. Liao, J. Frazier, H. Hauser, and H.-N. Kostis. Scientific storytelling using visualization. *IEEE Computer Graphics and Applications*, 32(1):12–19, 2012. doi: 10.1109/MCG.2012.24 2, 3, 4

[48] 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. doi: 10.1111/cgf.13195 4

[49] L. Morais, Y. Jansen, N. Andrade, and P. Dragicevic. Showing data about people: A design space of anthropographics. *IEEE Transactions on Visualization and Computer Graphics*, 28(3):1661–1679, 2022. doi: 10.1109/TVCG.2020.3023013 3, 4

[50] A. Ojo and B. Heravi. Patterns in award winning data storytelling: Story types, enabling tools and competences. *Digital journalism*, 6(6):693–718, 2018. doi: 10.1080/21670811.2017.1403291 1, 2, 3, 7, 8

[51] S. Petridis and L. B. Chilton. Human errors in interpreting visual metaphor. In *Proceedings of on Creativity and Cognition*, p. 187–197. ACM, New York, NY, USA, 2019. doi: 10.1145/3325480.3325503 4

[52] D. Ren, M. Brehmer, B. Lee, T. Höllerer, and E. K. Choe. Chartaccent: Annotation for data-driven storytelling. In *IEEE Pacific Visualization Symposium*, pp. 230–239. IEEE, New York, NY, 2017. doi: 10.1109/PACIFICVIS.2017.8031599 3

[53] Reuters. How powerful was the beirut blast? https://graphics.reuters.com/LEBANON-SECURITY/BLAST/yzdpxnmqbpx/index.html, 2020. 8

[54] J. S. Risch. On the role of metaphor in information visualization. *arXiv preprint arXiv:0809.0884*, 2008. 4

[55] A. Sallaberry, Y.-c. Fu, H.-C. Ho, and K.-L. Ma. Contact trees: Network visualization beyond nodes and edges. *PLOS ONE*, 11(1):1–23, 01 2016. doi: 10.1371/journal.pone.0146368 4

[56] H. Ç. Sarıca and Y. K. Usluel. The effect of digital storytelling on visual memory and writing skills. *Computers & Education*, 94:298–309, 2016. doi: 10.1016/j.compedu.2015.11.016 1

[57] S. Scarr, M. Sharma, and G. Bhatia. Covid-19: The pace of death. https://www.reuters.com/graphics/HEALTH-CORONAVIRUS/DEATHS/xlbpgobgapq/index.html, Sep 2020. 4, 6, 8

[58] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010. doi: 10.1109/TVCG.2010.179 1, 3, 7, 9

[59] D. Shen. Edgar Allan Poe's Aesthetic Theory, the Insanity Debate, and the Ethically Oriented Dynamics of ""The Tell-Tale Heart"". *Nineteenth-Century Literature*, 63(3):321–345, 12 2008. doi: 10.1525/ncl.2008.63.3.321 2

[60] 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. doi: 10.1109/TVCG.2020.3030403 4

[61] Y. Shi, C. Bryan, S. Bhamidipati, Y. Zhao, Y. Zhang, and K.-L. Ma. Meetingvis: Visual narratives to assist in recalling meeting context and content. *IEEE Transactions on Visualization and Computer Graphics*, 24(6):1918–1929, 2018. doi: 10.1109/TVCG.2018.2816203 2, 3, 4

[62] A. Srinivasan, S. M. Drucker, A. Endert, and J. Stasko. Augmenting visualizations with interactive data facts to facilitate interpretation and communication. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):672–681, 2018. doi: 10.1109/TVCG.2018.2865145 1, 4

[63] C. D. Stolper, B. Lee, N. H. Riche, and J. Stasko. Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories. Technical report, Microsoft Research, Washington, USA, 2016. 1, 4

[64] Y. Tanahashi and K.-L. Ma. Design considerations for optimizing storyline visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2679–2688, 2012. doi: 10.1109/TVCG.2012.212 3

[65] T. N. Y. Times. The world's ball. https://www.nytimes.com/interactive/2014/06/13/sports/worldcup/world-cup-balls.html, 2014. 4

[66] C. Tong, R. Roberts, R. Borgo, S. Walton, R. S. Laramee, K. Wegba, A. Lu, Y. Wang, H. Qu, Q. Luo, and X. Ma. Storytelling and visualization: An extended survey. *Information*, 9(3), 2018. doi: 10.3390/info9030065 2, 3

[67] J. Truby. *The anatomy of story: 22 steps to becoming a master storyteller*. Farrar, Straus and Giroux, 2008. 2, 3, 4, 9

[68] Q. Wang, Z. Li, S. Fu, W. Cui, and H. Qu. Narvis: Authoring narrative slideshows for introducing data visualization designs. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):779–788, 2019. doi: 10.1109/TVCG.2018.2865232 3

[69] S. Wang, Y. Tanahashi, N. Leaf, and K.-L. Ma. Design and effects of personal visualizations. *IEEE Computer Graphics and Applications*, 35(4):82–93, 2015. doi: 10.1109/MCG.2015.74 4

[70] 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, 2020. doi: 10.1109/TVCG.2019.2934398 1

[71] W. Woolf. https://twitter.com/WillikinWolf/status/1176006515968241665. Accessed: 2022-02-22. 4, 6

[72] L. Yang, X. Xu, X. Lan, Z. Liu, S. Guo, Y. Shi, H. Qu, and N. Cao. A design space for applying the freytag's pyramid structure to data stories. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):922–932, 2022. doi: 10.1109/TVCG.2021.3114774 1, 3

[73] L. Yang, X. Xu, X. Lan, Z. Liu, S. Guo, Y. Shi, H. Qu, and N. Cao. A design space for applying the freytag's pyramid structure to data stories. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):922–932, 2022. doi: 10.1109/TVCG.2021.3114774 4

[74] Z. Zhao, R. Marr, and N. Elmqvist. Data comics: Sequential art for data-driven storytelling. Technical report, Univ. of Maryland, 2015. 4, 10