DOI: 10.1111/cgf.14815
Eurographics Conference on Visualization (EuroVis) 2023
D. Archambault, R. Bujack, and T. Schreck
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 42 (2023), Number 3

Data Stories of Water: Studying the Communicative Role of Data Visualizations within Long-form Journalism

M. Garreton[1], F. Morini[2,3], D. Paz Moyano[1], G.-C. Grün[2,4], D. Parra[1] and M. Dörk[2]

[1]Pontificia Universidad Católica de Chile , Santiago, Chile
[2]UCLAB, University of Applied Sciences Potsdam, Potsdam, Germany
[3]Södertörn University, Stockholm, Sweden
[4]TU Dortmund University, Dortmund, Germany

Abstract

We present a methodology for making sense of the communicative role of data visualizations in journalistic storytelling and share findings from surveying water-related data stories. Data stories are a genre of long-form journalism that integrate text, data visualization, and other visual expressions (e.g., photographs, illustrations, videos) for the purpose of data-driven storytelling. In the last decade, a considerable number of data stories about a wide range of topics have been published worldwide. Authors use a variety of techniques to make complex phenomena comprehensible and use visualizations as communicative devices that shape the understanding of a given topic. Despite the popularity of data stories, we, as scholars, still lack a methodological framework for assessing the communicative role of visualizations in data stories. To this extent, we draw from data journalism, visual culture, and multimodality studies to propose an interpretative framework in six stages. The process begins with the analysis of content blocks and framing elements and ends with the identification of dimensions, patterns, and relationships between textual and visual elements. The framework is put to the test by analyzing 17 data stories about water-related issues. Our observations from the survey illustrate how data visualizations can shape the framing of complex topics.

CCS Concepts

• ***Human-centered computing*** → *Visualization design and evaluation methods; Empirical studies in visualization;*

1. Introduction

Over the last decade, data journalism has gained popularity as a hybrid practice that combines different disciplines, including computer science, information design, and storytelling, to find stories in data and present them for public use [HL20]. This development is due to the rapid socio-technological transformation and maturation of the web, which enables journalists to explore innovative forms of news-making [HCHC20, KM13, RHDC18, SH10]. Among those, data stories have reached particular popularity. Data stories are a genre of long-form data journalism that combine text, infographics, statistical charts, and various other types of visual artifacts—such as photographs, videos, and illustrations—that ideally fit coherently into the main narrative [WEK18]. Data stories are characterized by a strong interdependence of textual and visual elements. Authors often resort to common patterns to blend every element seamlessly and provide rich context to readers [SH10]. Prior visualization research has paid close attention to this genre, either by investigating the rhetorical aspects of data visualization [HD11, Ali21, Hul19] or by focussing on the role of sequencing in establishing subgenres [SH10, HDR*13]. However, there has been little research on the interplay between data visualizations and the other elements present in a data story. In many cases, data visualizations appear to act as arguments promoting a particular understanding on a given issue by employing framing effects [HD11]. Here framing refers to the selective representation of a topic, which can shape audience perceptions or interpretations of the issue. Even less consideration has been given to methodological frameworks for the systematic assessment of interdependence among these elements. Text plays a fundamental role in providing context for visual elements and is equally important in shaping the reading experience [SH22]. However we know little about their respective roles and therefore we ask: *What is the communicative function of data visualizations in data stories and how do they relate to other story elements?*

This study focuses on data visualization and its interplay with other modes of representation such as text, photographs, videos, and illustrations. We adopted an interpretative approach and ran a content analysis on a selection of stories from different media outlets. Given the variety of examples, we devised a series of criteria for collecting a sample of data stories. First, to maintain a high quality standard, we limited our selection to award-winning data stories. Then, in order to ensure the comparability of their textual and visual representation, a topical focus was necessary for our selection. Data stories often provide context and can facilitate the understanding of demanding topics that can challenge our understanding, for

instance migration or climate change [Kni15]. Especially the latter has become a popular topic for data journalists all over the globe. *Water*—its scarcity, surplus, symbolic and practical value—is increasingly at the center of climate change coverage. Water can also be considered a hyperobject [Mor13] that stretches across spatial and temporal boundaries at a scale difficult to experience from a human perspective. Thus, reporting on water also means discussing several other connected problems such as the biodiversity crisis, health problems, and geopolitical tensions. Given these premises, we decided to use water as a topical focus for sampling data stories with which we finally obtained a corpus of 17 articles. We devised a framework for examining data stories and characterizing the communicative role of data visualizations. Our contribution is twofold: 1) a framework for the interpretative analysis of visualizations in data stories, and 2) observations from the close study of data visualizations in water-related data stories.

2. Background

This work relates to prior research on data visualization, human-computer interaction, journalism, media and communication.

2.1. Data stories as a multimodal artifact

Long-form digital journalism has gained momentum among academics and practitioners, captivating the audience by combining multimedia resources such as text, photographs, videos, and data visualizations like dynamic maps [Hii17]. By taking into account an expanded variety of long-form digital journalism, data stories are defined as "hybrid multimodal artifacts that weave numbers, words, images and design into a coherent whole" [WEK18]. This inherent hybridity of data stories with different semiotic modes interacting and cooperating makes these artifacts complex and interesting to analyze. These modes can be understood as a variety of expressive resources used to generate meaning that takes visual, verbal, pictorial, graphic, and other forms [Hii20, Bat17].

Research on multimodality has expanded beyond linguistics, now also concentrating on other forms of expression such as photography, diagrams, page layout, and typography [Hii20]. This development is essential to data visualization, as this broadening of the forms of expression also encompasses the data story and its components. As an attempt to extend the multimodal analysis to data visualization, a recent study focuses on the concept of canvas and how the properties of semiotic modes can appear in them [Hii20]. This type of analysis is of interest to us due to the decomposition of the data visualizations analyzed into different elements generated by the author. We will use this decomposition to analyze the components that make up data stories.

The fundamental component of journalistic stories is the narrative, which can be defined as a sequence of temporally structured events coherently related to each other with causal links [Tho93]. To generate these narratives, or stories, artifacts must have a coherent and cohesive relationship between the different modes, both textual and visual [WEK18]. The sequential structure of text and data visualization alternating with other visual elements is a common pattern in data stories.

2.2. The communicative effect of data stories

The relations between the different elements of a story shape the impression and understanding of the respective topic. Especially when applying data visualizations to represent social, cultural, or ecological issues, it is hard to maintain that data visualizations are unbiased, objective, or neutral [DFCC13]. On the contrary, they function as political artifacts that can challenge or naturalize particular worldviews [Nær20, DK20]. There is still little empirical work on this aspect; however, a forerunner in this type of research is the work by Hullman and Diakopoulos [HD11] providing a taxonomy of rhetorical uses of information visualizations.

Regarding the role of authors and readers, data stories can be considered a particular combination of explorative and communicative visualizations [SH10]. To understand the complexity of how people interpret narrative stories, "viewing codes" [Cha07] define them as the cultural, perceptual, cognitive, and psychological lenses that shape the conventions of what users expect to be communicated by specific visualizations.

Research on journalism has extensively studied how textual devices can generate specific orientations towards a topic [AJ21, Ali21]. This is known as framing and is an expanding research area in data visualization [vBMK*20]. Framing can be defined as the process through which people develop a particular conceptualization of a topic or reorient their thinking about it [CD07]. A narrower definition of framing in communication is that it seeks to select some aspects of reality and make them more prominent in communication, which determines how people understand, remember, and evaluate how to act on a problem [Ent93]. The elements comprising framing in communication are phrases, images, and presentation styles, which reveal what the communicator considers relevant to an issue by emphasizing certain aspects [Dru01].

We seek to extend the concept of framing to visualization in data stories and understand framing as the process in which particular aspects of a phenomenon are—deliberately or not—presented in more relevant or salient ways and thus promote particular evaluations, interpretations, and decisions [BLW11, vBMK*20]. Framing has been studied more extensively in images and illustrations [PBDSdV15, PHFH19, KL05, Ken01], which can serve as a starting point for framing analysis of data visualization. Some authors claim images are more powerful framing tools than text [KL05, PHFH19] since people tend to accept the visual frame with little hesitation. This could be explained by the stronger emotional and immediate cues of visual forms [RD11, BMG*10].

3. Interpretative framework for visualizations in data stories

Our framework rests on two different, yet complementary approaches. First, we use the perspective proposed by Entman [Ent93] who identified main *elements of framing* that a text can contain:

- Problem definition: The issue that is at stake.
- Causal interpretation: The forces that create the problem.
- Moral evaluation: Judgments about motivation and fault.
- Treatment recommendation: Suggesting and justifying solutions.

A data story may perform all four or a subset of these framing elements and it could be that particular story components address

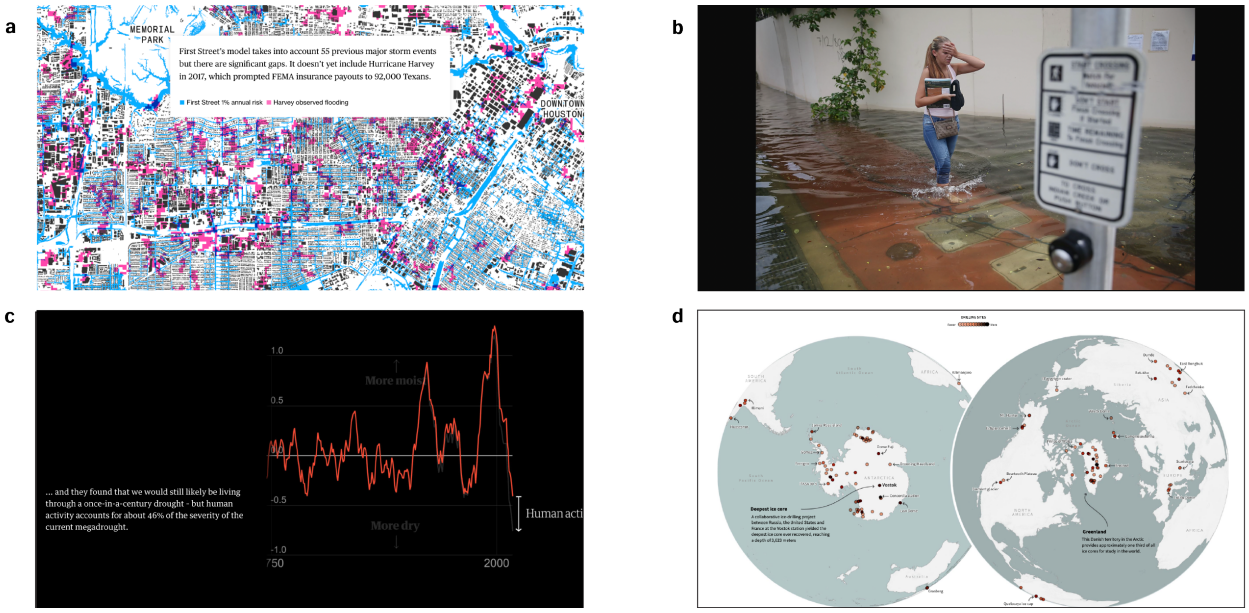

Figure 1: *A selection of data stories of water-related issues, including first row* [Blo20, Blo21] *and in second row* [The21a, Reu21b]

particular framing elements. Second, the *levels of visual framing* by Rodriguez and Dimitrova [RD11] suggest a four-tiered model for identifying and analyzing visuals as: 1) denotative systems, 2) stylistic-semiotic systems, 3) connotative systems, and 4) ideological representations. The levels of visual framing have been proposed for photographs, however, they can also be transferred to data visualization [vBMK*20], since framing can be understood as a window of reality and it can directly impact how environmental discourse is understood. The levels of visual framing offer clues for understanding how meaning is constructed in a visualization by examining the different visual characteristics and how they are organized and connected. For our research, we follow a two-fold methodology that integrates elements of framing [Ent93] and levels of visual framing [RD11] as particularly appropriate for our research on visualizations in the context of data stories. The elements of framing can help understand the communicative functions of visualizations with regard to a given issue, e.g., if they address the problem definition or causal interpretation. Whereas the levels of visual framing turn our attention to the particular characteristics of the visual form. This combination of both approaches lets us appraise both form and function of visualizations in data stories.

We surveyed 17 data stories focusing on water-related topics, with the intention of observing the communicative function of data visualizations, and their interdependence with other modes of representation. For this purpose, we drew on archives of two major data journalism competitions: Sigma Awards and SND (Society for News Design) Awards. First, we searched for all the articles awarded in 2021 and 2022 to select stories with relevant keywords such as *water*, *flood*, *drought*, *sea*, *ocean*, *river*, *precipitation*, *rain*, *lake*, and *glacier*. From this initial selection we obtained a sample of 122 stories. We reviewed each of the items to ensure that they correspond to three key requirements: The article had to adhere to the definition of a data story and exhibit its fundamental characteristics (i.e., integration of text, data visualization, and other visual elements such as photographs, illustrations, videos); the topical focus had to be on water-related issues; finally, the article had to be written in English. Based on these criteria, we obtained a corpus of 17 stories published by news outlets worldwide (4), in USA (7), Hong Kong (2), Singapore (2), Australia (1), and UK (1). This limited sample allowed us to perform qualitative in-depth analysis of each story, including the careful consideration of each visual and textual element as unique units. Moreover, we considered a total of 117 data visualizations across the 17 stories.

After the selection, we analyzed the stories in our sample following six steps inspired by previous research and based on the elements and levels of (visual) framing. Our analysis has been exploratory applying a qualitative research design approach in an iterative fashion. Six coders worked in tandems over the course of six months. They started by reading and coding the stories individually and met regularly to discuss their progress and coding effort in order to reach consensus about coding results. In the first months, this process of individual and then group iteration allowed us to build a qualitative framework. Over the last two months, we tested it with the stories reviewed in this study. During this time, coders regularly met to coordinate and reflect on the methodological process and refine the respective steps. It is during these negotiations when the interpretative methodology emerged, while examining data stories and exchanging impressions and insights. The six steps of the analytical process are illustrated in Figure 2. The process starts with coders identifying and naming individual content blocks and ends with arranging all observations in a matrix, across 19 fundamental dimensions of visualizations in data stories.

Step 1—Identify content: In the first step, we assigned different media types to all content blocks. The process consisted of observ-

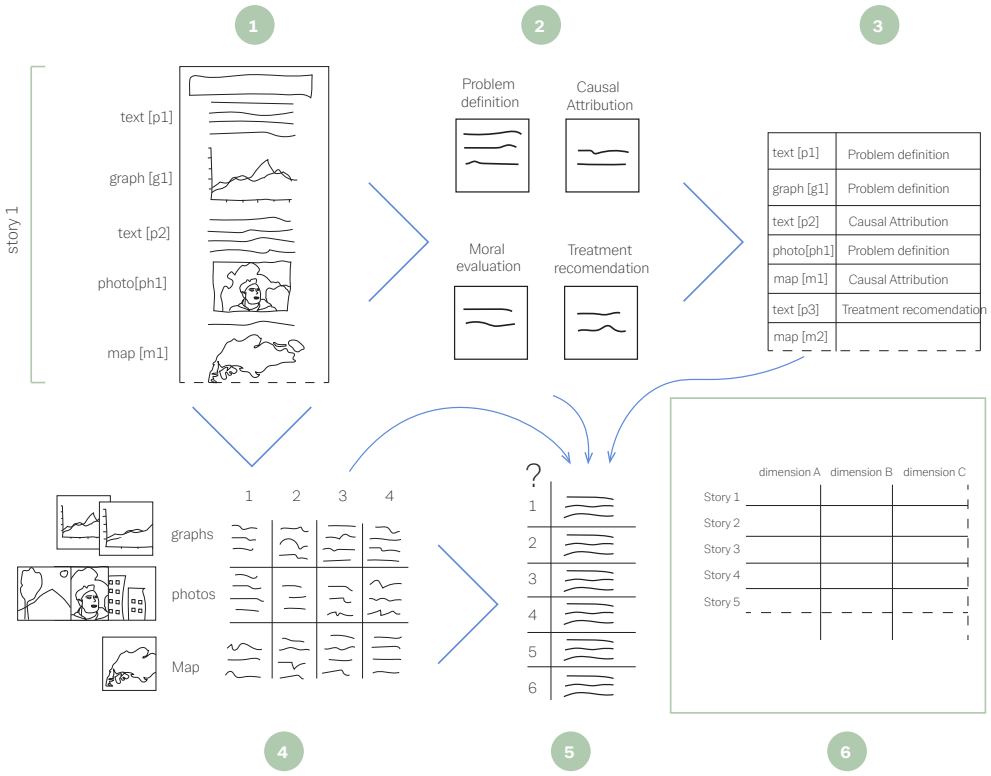

Figure 2: *The analysis of this study included six steps. The first five were carried out for each of the data stories separately, and in a final step we integrated all the observations into a matrix. In step (1) we visually identified the content blocks. In (2) we describe the framing elements and then (3) assign a framing element to each content block. In step (4) we grouped the visual representations by type (photographs, maps, graphs, etc.) and analyze the levels of visual framing for each group [RD11]. In step (5), we recognize patterns and relationships among the stories and formulate overarching questions. To answer them we used the observations from all the previous stages. Finally in step (6), we collected the findings of the stories we analyzed in a matrix with 19 dimensions that emerged from the previous steps.*

ing each element within the story flow and label it according to its form. We identified several types of content blocks: paragraphs of text, illustrations, photographs, videos, charts, diagram, and maps. The goal of this step was to understand how each story is visually structured and how content blocks relate to each other.

Step 2—Recognize framing elements: In step two, we relied on Entman's concept of framing [Ent93] to recognize what elements of framing were at play: Problem definition, causal attribution, moral evaluation, and treatment recommendation. For each of the stories within the corpus, coders briefly described in one paragraph if and how each of the elements was present in the respective story. This allowed coders to reach consensus and prepare for the next step.

Step 3—Classify content blocks: Whereas the classic definition of framing typically focuses on text, we distinguished between three components according to the definition of data stories: Text, visuals, and data visualization. Building on the description of framing elements from Step 2, we evaluated each content block—text, visuals, and data visualization—to decide if one of the framing elements applied and classified them accordingly. This allowed us to identify the most recurrent framing elements, both for the overall corpus as well as by mode of representation.

Step 4—Assess visual framing levels: After analyzing which framing elements were included in the narration of data stories, we then moved on to also investigate the visual framing deployed. To do this, we took apart the linear story flow and grouped the visual components together by type—photographs, maps, graphs, illustration. This allowed us to approach each group of visual components holistically to examine the extent of visual framing within each type across multiple components. To examine the visual components, we relied on the four-level framework for the analysis of visual framing [RD11] that investigates visuals as denotative (level 1), stylistic-semiotic (level 2), and connotative systems (level 3) as well as ideological representations (level 4). This allowed us to identify the tone of data visualizations and other visuals.

Step 5—Reach consensus: In this phase, coders worked initially in tandem to consolidate their independent findings on their sample subset. Then, all coders jointly formulated and refined the analysis questions to derive observations and comparisons across the entire

sample, in addition to the story-by-story analysis done in previous steps. From this discussion six questions emerged:

1. How do the visual content blocks compare/relate to each other?
2. How do they help frame the story?
3. What is the relationship between visual and textual elements?
4. How is the larger concept of water represented?
5. What is the status of people in the story?
6. Which emotions are conveyed by the story?

Step 6—Consolidate: The goal of this final step is to consolidate previous observations into one single matrix. To achieve this coders worked together to address the questions raised in the previous step. Discussing the questions led to a system of dimensions suitable to characterize the communicative function of visualizations in the data stories of the sample. All questions from Step 5 were translated into individual dimensions and different possible values for each dimension. For example, the question "How is the larger concept of water represented" was translated into the dimensions "presence of water" and "role of water". The dimension "presence of water" could have the values "minimal" or "prominent"; the dimension "role of water" could have the values "threat", "resource", "witness of climate change", or "devastating power". One or more dimensions for each question have been integrated into a visual matrix comprising 19 dimensions (see Figure 3). The dimensions serve different scopes within our analysis and range from defining the role of people and water within the story, to judging the visual coherence among visual blocks. Single dimensions are described in detail in the following section. Compiling the matrix helped us in consolidating our analysis across the stories' sample and ultimately discussing emerging patterns concerning the role of data visualizations in relation to other content blocks.

4. Dimensions

To characterize the communicative function of data visualization and how it relates to other story elements, we developed a classification system comprising several dimensions. This has been distilled following our bottom-up categorization process and, when possible, we mobilized existing taxonomies from related domains and adapted them to our research. To this extent, part of the dimensions are concerned with observing framing effects and agenda setting in the analyzed stories [Wea07, Ent93]. Additional dimensions consider the role of human and non-human actors in the narrative, the coherence of visual elements, and geographic scale of the story setting. Below we discuss each group of dimensions, their respective categories and frequencies. We use colors to differentiate them. The number in parentheses used after naming a dimension refers to the total number of stories in which it was identified. In Figure 3, we list all the dimensions identified in each of the stories.

4.1. Presence and content of framing elements

The first set of dimensions encode the position of framing elements within the story. We created a dimension for each of the four framing elements and determined which components of the article are responsible for it. These dimensions are critical to detect the recurrence of framing patterns across our sample of stories. After defining which story components are responsible for framing, we also identified their main topic.

Presence of framing elements				
	Problem definition	text	17	
		visuals	13	
		datavis	17	
		not included	0	
	Causal attribution	text	15	
		visuals	5	
		datavis	8	
		not included	1	
	Moral evaluation	text	11	
		visuals	1	
		datavis	2	
		not included	6	
	Treatment recomm.	text	10	
		visuals	1	
		datavis	1	
		not included	7	

Table 1: *Framing elements are present across all content blocks. However, problem definition and causal attribution are prominent, especially for visuals and data visualizations.*

Problem definition refers to the issue at hand and its costs and benefits for other actors involved in the story. Across our sample, *problem definition* concerns a variety of topics: Food security, flooding, drought, heavy rainfall, melting glaciers, sea-level rise, weather extremes, wildfires, and biodiversity crisis. In the majority of cases, the problem is defined by the excess or scarcity of water. For food security and biodiversity crisis the focus is on the long term effect of water scarcity or excess on animals and land, whereas for the remaining ones the focus is on extreme events and their social, economic, and environmental impact.

Causal attribution defines which forces are at the root of the problem and is used to attribute responsibility to one or more agents. In our sample, causal attribution is attributed variably to humans and nature. We identified seven main responsible agents: Hydro-power, climate change / global warming, unsustainable agriculture, high dependency, heavy rainfall, lack of investment, urbanization. Especially for unsustainable agriculture, high dependency, lack of investment, and urbanization the responsibility is clearly on human agents and social conflicts.

Moral evaluation evaluates causal agents and their effects. As the name suggests, the reader is likely to form a moral judgment and becomes able to trace causality between story components. In our stories, moral evaluation is attributed to six different agents: Politics, inefficient infrastructures, climate change or increased global emissions, human influence or activity, and social inequality. All these agents are connected variably to human activity across different domains: Society, politics, and the economy.

Treatment recommendation is used to share solutions or suggest strategies in order to address the problem or its cause. We found five different types of treatment recommendations: Limit global warming or reduce carbon emissions, prepare for change, support ecosystems, build or improve infrastructure, develop policies or political solutions. Given the clear focus on climate change, all these strategies are concerned with either protection or adap-

Content of framing elements	Problem def.	food security	2	
		floodings	7	
		drought	4	
		heavy rainfall	1	
		melting glaciers	3	
		sea level rise	1	
		weather extremes	1	
		wildfires	2	
		not given	0	
	Causal attribution	hydropower	2	
		climate change	11	
		agriculture	2	
		high dependency	3	
		rainfall	1	
		no investments	1	
		urbanisation	1	
		not given	1	
	Moral evaluation	political body	2	
		infastructure	1	
		climate change	4	
		human influence	6	
		social inequality	3	
		not given	4	
	Treatment recomm.	limit warming	2	
		change	4	
		protection	3	
		infrastructure	5	
		develop policies	3	
		not given	7	

Table 2: *Floodings, droughts, and climate change are considered key problems across most of the stories in our sample. Additionally, humans are considered responsible for these extreme events.*

tation. Only two types of solutions are focused more broadly on system change: Limit global warming or reduce carbon emissions, and develop policies or political solutions.

The framing elements are presented in the data stories both in the text as well as in the visual content and data visualizations (see Table 1). We identified that all the articles contain a *problem definition*. Of the total of 17 stories analyzed, all used text and data visualization to outline the problem. At the same time the vast majority also did so through the other visual blocks (13). When we look at the *causal attribution*, only one of the articles does not mention them; we see that many use text (15) and almost half do so through data visualization (8). On the other hand, the framing elements corresponding to *moral evaluation* and *treatment recommendation* are present in fewer stories, but when they appear, it is through text (moral evaluation: 11 and treatment recommendation: 10). Only one story uses visual content and data visualization to communicate moral evaluation and recommend treatments. When observing the topics addressed by the stories in our corpus, we notice the prevalence of certain issues (see Table 2). The most frequent problems are flooding (7) and drought (4), followed by glacier melting (3), food security (2) and wildfires (2). As for the causes, these are also varied but with a clear majority attributed to climate change

Scale of storyline	Local	13	
	Regional	10	
	National	9	
	International	3	
	Planetary	3	

Table 3: *Local and regional stories are the most common across our sample. In some cases, authors combine more than one scale within the same story.*

or global warming (11). In relation to moral evaluation, the stories most frequently relate it to human influence (6) and to climate change and the increase of carbon emissions (4). That is to say that these stories highlight the responsibility of human influence and activity for the cause of the problem and/or the problem itself. Finally, the stories that present treatment recommendations do so by proposing to improve infrastructure (5) and prepare for change (4), while only two stories recommend limiting global warming.

4.2. Scale of storyline

This dimension maps the scale of the narrative, intended as the political, cultural, geographical, and social extent at which the story is told. We identified five levels of scales, as visible in Table 3. Data stories can be local [Reu21b]) , regional, national, international, or planetary (e.g., Figure 1.d, [Reu21b]). The first three types identify stories concerned with small or medium scale problems, often recounted at the level of a community (local), or country-wide (national). The last two types include examples from more than one country up to a planetary scale. It is not uncommon for stories to have more than one scale, if the author decides to transition between different magnitudes of the same problem. When observing how these scales are represented in our stories, we identified that most of them are represented locally (13) and/or regionally (10), as shown in Table 3. Both international and planetary scales are less common, with three stories each.

4.3. Presence and roles of actors

This dimension evaluates the presence and roles of actors in the story. It considers human actors as well as non-human entities and critically assesses how they are described. For this study, we mainly considered two sub-categories: *water* and *people*. Table 4 summarizes both groups with their dimensions.

Water is portrayed in various ways: As a threat (7), resource (7), witness of climate change (7), or devastating power (5). In the first case, water is presented as a future threat to humans, cities, and the environment. When water is presented as a resource, the focus is instead on its importance for human life. The last two types denote water as directly connected to climate change. When water is witness of climate change, authors frame it as a clear proxy for global warming. Instead, when presented as a devastating power, it is often implied how water is capable of destroying infrastructures and the natural environment.

People are consistently included in almost all stories. When we recognize them as affected (11), people are present in the story to

Actors	Water	Role	Threat	7	
			Resource	7	
			Witness	7	
			Devastating	5	
			None	0	
		Presence	Minimal	2	
			Prominent	14	
			None	1	
	People	Role	Affected	11	
			Experts	11	
			Politics	3	
			None	2	
		Presence	Minimal	10	
			Prominent	5	
			None	2	

Table 4: *Water and people are treated very differently in our sample. Water has a prominent role and is often depicted as a threat. People are less central and represented as victims or science experts. Political actors are very seldomly mentioned.*

Tone	Story	Neutral	8	
		Negative	9	
		Positive	0	
	Visuals	Neutral	7	
		Negative	9	
		Positive	1	
	Dataviz	Neutral	9	
		Negative	7	
		Positive	1	

Table 5: *Story, visuals, and data visualizations share negative and neutral tones across our sample. Only few stories have positive or hopeful content.*

Relation between	Text, visuals, datavis	well-connected	15	
		not connected	2	
		overlaps flow	9	
		alternating flow	0	
	Visual blocks	consistent	15	
		inconsistent	2	

Table 6: *Overall data stories are well designed. Content is well-connected and visually consistent.*

illustrate the impacts and effects water has on them (e.g., Figure 1.b). In some cases people report from their own experience and how water-related problems impact their life. Scientists—or scientific experts (11)—usually take an advisory role and add a scientific perspective to the topic. People are recognized as political actors (2) when their role in the story explicitly concerns policy-making or activism; they are described as active subjects capable of making and/or influencing political decisions. People are not always present in the story, when this is the case, their role is encoded as "none" (2). When we compare the presence of water and people in the stories, we recognize that the former, as expected, is much more prominent (14) than the latter (5).

4.3.1. Tone of narrative

The tone of the narrative is concerned with its emotional affordances. This dimension considers text, visual, and data visualization alike, as visible in Table 5. Ultimately, we were able to distill three different types for this dimension that apply invariably to textual and visual elements: negative tone, neutral tone, and positive tone. When the tone is negative, text and visual elements are used to convey an alarming or depressing picture (e.g., Figure 1.c). For visuals, this usually suggests a certain use of color or perspective that induce negative feelings. When the tone is neutral, components are used to simply describe an event or portray its context. Finally, a positive tone is observed when elements in the stories are used to convey hopeful or constructive content.

After examining our corpus of data stories, we can conclude that the tone used in the stories is either neutral or negative and this is consistent with the visual and data visualization blocks. The visuals are slightly more negative than neutral while with the data visualization it is the other way around; neutral and negative. The story that has a positive tone is presented through data visualizations and visual blocks, not through the story.

4.4. Relation between elements

The last two dimensions are concerned with the interdependence of the story's components (see Table 6). The first dimension evaluates *the relation between text and visuals*, i.e., how text and other elements interact within the story context. Story elements can be *well-connected* or *not connected*. In the first case, the visual and text blocks are well intertwined, for example, with the text referring to the charts or the photos showing people mentioned in the text. In the second case, the visual blocks and text blocks are not well connected to one another, i.e., the text does not refer to the charts or the addition of visual elements does not seem to follow a clear narrative purpose. The second dimension highlights how different blocks are laid out on the article page. Story blocks can *overlap* or *alternate*. When elements overlap, this means that the story is composed in whole or in parts by visual, textual or data visualization elements that flow over others (e.g., Figure 1.a). An example is a map overlaid with text boxes and annotations allowing the story to be followed. Conversely, when the elements alternate, it means that the content blocks are arranged one after the other without overlaps. Visual elements do not exclusively refer to text, they also refer to each other. Therefore, the third dimension determines how visual components relate to each other. In this case, we decided to discuss consistency versus inconsistency. When visual components are consistent the design of visual blocks is consistent across the data story or the visual blocks are graphically layered on top of each other. For example, photographs or charts overlapping maps. If visual components are inconsistent the design of visual blocks varies across the story leading to confusion, using different types of encoding, color choices or contrasting visual elements.

When looking at the frequency of relationships between elements, we see that for the most part texts are well connected to visuals. Likewise, when examining visual and data visualization blocks,

they are also mostly consistent with each other. With respect to the layout we find both overlapping and alternating arrangements.

5. Patterns

For all 17 stories surveyed, we were able to apply our dimensions and recognize repeating patterns. Categorizing all items across our sample allowed us to shed light on the communicative role of data visualization as well as other key elements of water-related data stories. The distribution matrix in Figure 3 is key in reading these patterns and understanding their relevance within the genre. In the following, we outline the main recurring observations.

Structure of data stories. Water related data stories are all characterized by a variety of media integrated in the story. This integration is not superficial and limited to single pairs of artifacts and text, but several types of multimedia elements are woven in together and constitute the final outcome. In most of our stories visuals, text, and data visualization are well connected. Interactivity reinforces the interweaving between different elements. In several occasions annotations or text appear above or intertwined with visualizations, creating a stronger connection between the two content blocks. They are not independent blocks, but they reference each other tightly as a coherent whole. The strategy of overlapping and mixing different content blocks is used often across our sample and constitutes a distinguished layout. In some cases, for instance, boxes of texts are used as overlays for maps (e.g., Figure 1.a) as the user scrolls or clicks through the story. In this case, visualization is used to establish a visual base upon which the story unfolds **[Kon20, ABC20]**. Whenever stories were composed of different visual elements, we found them to be mostly visually consistent. Color encoding, typography, and layout are all important in maintaining consistency. Creators of data stories seem to pay particular attention to ensuring a coherent visual structure Data visualizations are very detailed and specifically designed for data stories, although interactivity is quite limited, resorting to click or scroll. The visual components are often quite elaborate and well produced.

Role of data visualization. Data visualization is mostly used to frame the problem and its causal agents. However, it is rarely used to communicate treatment recommendations. In our corpus of stories only one of them **[The21b]** uses data visualization to suggest a solution. Data visualizations tend to function as a way to provide evidence of a problem and its causes, supporting with data what is described textually. To this extent, data visualization is more neutrally connoted. Whereas other visual artifacts and text often convey a negative tone (e.g., Figure 1.f), data visualization is slightly more often used as a neutral element within stories (e.g., Figure 1.d). Designers systematically employ bland color schemes, typography, and visual elements, making the data visualization grammar decisively more neutral, especially when compared with other visual artifacts. This finding suggests that charts are used as a contextualizing element, rather than an explicitly rhetorical one. Moreover, data visualization represents problems and causes (as framing elements) and across our corpus these are essentially negative and alarming issues (flooding, drought, climate change, unsustainable agriculture, etc). Therefore we observe an inconsistency or discrepancy between the neutral tone conveyed by the visualization and the negative content it actually represents. Maps are central to water-related data stories and it is possible to distinguish some main use cases among the items in our sample. In some cases, maps are used to georeference a place such as a lake, river, or mountain (e.g., Figure 1.g). This allows users to understand the geographical context of the events. Sometimes maps do not contain data at all and are not meant to provide additional information, but they serve as guidance through the narrative. For instance, in **[Al 20]** and **[Reu21a]** the visualizations depict rivers and readers use vertical scroll to follow the downstream flow encountering pieces of the narrative.

Interdependence between frames and tone. *Problem definitions* span all types of content blocks, while text and data visualization are also used to determine *causal attribution*. However, when looking at frames in relation to the tone of content blocks a discrepancy is noticeable: Text and visuals are more often considered to have a negative tone, while data visualization is predominantly connoted as neutral. This pattern highlights that both text and visuals are used explicitly to mobilize emotions, for example, through photographs representing desperate people in the midst of devastating floods or fields of crops burned by the sun **[Reu20, Reu21a, Blo20]**. Data visualization is presented as neutral, even when it is used to frame negative content. This difference in tone possibly signals the complementary roles of different content blocks: Authors tend to provide more layers of information for the benefit of the reader.

Representation of water. The main problems discussed in water-related data stories are floods and droughts. These issues are complex and have widespread causes and effects. Authors often resort to scaling down the problem at a local level to make it more understandable and relatable. However, in some cases authors also choose to switch between local and global perspectives, going from regional stories to planetary overviews (e.g., Figure 1.i). This strategy allows them to show that problems associated with water, although they have affect locals' lives, have also a crucial planetary component. The stories in our sample mostly use maps to represent water. These make it easier to situate problems and their effects in a specific territory, contributing to bind them to human experience and make them easier to grasp for readers. Therefore, water is always visually present in stories, specifically as a threat or necessary, but scarce resource for human subsistence.

Presence of people. Unlike water, people have a less prominent presence in our corpus. They are represented mainly through photographs, which tend to portray people as victims of floods, droughts, or other water-related threats (e.g., Figure 1.c). Even when scientific experts are present in stories, they usually only support the *problem definition*. Political actors are largely absent. This tendency suggests that data stories related to water are focused on reporting incidents, problems, trends, or extreme events as consequences of climate change that potentially threaten our society. Authors offer little room for solutions or in discussing the role of policy-making or activism in regards to climate change issues. By portraying people as victims, authors often dismiss their role as causal agents of climate change. There is no emphasis on making people responsible for the problems, and certainly less on making them participate in the solutions.

Figure 3: *Distribution of the dimensions for each of the 17 stories we analyzed. From left to right:* **[Kon20, Sou21, Reu20, Sou20, Reu21a, ABC20, The21c, Blo20, Blo21, Reu21b, New21, The21a, NPR21, Al 20, The20, The21b, USA21]**

6. Conclusion

In this paper, we have introduced an interpretative framework to understand the communicative role of data visualization within data stories. This methodology sets out to systematically interrogate how charts and maps are used in relation to other narrative components (e.g., text, photographs, videos) and which communicative function and framing role they fulfill. While our framework can be applied to data stories about any topic, for the purpose of this study to test it, we selected and analyzed a sample of 17 data stories focusing on water-related topics. The results of this initial analysis, already shed light on recurring patterns. The visual structure of data stories is similar across our sample, with visual and textual blocks closely tied to each other, for example, with the text referencing a chart. Data visualization is often used to frame problems, rather than to discuss solutions. While often used to represent critical issues, the tone of data visualization is predominantly neutral, rather than positive or negative. People are often represented through photographs, rather than through visualizations. The role of people in the stories is likely to be passive, for instance as victims of floods or droughts. Water, instead, occupies a prominent role in the analyzed data stories and is often portrayed as a threat. Maps are particularly important to depict at which scale water is causing problems. For this purpose, local scales are favored over international or planetary overviews, likely reflecting the authors' interest to make a story relevant to their respective readerships in communities and regions. These findings establish a ground for further examination and critical discussion of data stories on pressing issues such as water issues, biodiversity loss, or the climate crisis.

More consistency and constructiveness. All the stories across our sample focus on framing water-related issues where data visualization plays a key role: Maps and charts are designed to provide the societal or geographical context as well as the extent of the events that are taking place. We argue that data stories, rather than being limited to a neutral tone, would benefit from visualizations designed more consistently (both in terms of visuals and content) and in tune with the overall tone of the narrative. Data visualization is rarely used to represent potential solutions. Whereas water-related problems are often critical, we recommend that authors invest more care in discussing possible solutions. In the design of data stories *treatment recommendation* should be given the same space and attention of *problem framing* and *moral evaluation*. This could support a shift from an informative function of data stories to providing a more constructive approach with regards to the issues at hand. There is evidence that solutions journalism makes people feel more in control of the situation and hence more likely to take action [TL22]. Data visualization has the—currently underutilized—potential to support problem solving and support readers in developing a constructive perspective on difficult problems.

Two types of visual structure. Another point for discussion concerns the visual structure of data stories. Across our sample, we identified two main layout types: Overlap and alternate flow. The former is designed to convey a holistic image of several textual and visual blocks. The latter is effective in delivering content orderly to the reader. Whereas effective, both structures revolve around text as the core element for the story, taking little advantage of visualizations' flexibility and potential for interactivity. We suggest to better interweave visual artifacts and use data visualizations as a foundational block when planning data stories. In some cases, making data visualization or other visual artifacts more central could be a good way to foster readers' engagement with the content.

Limitations. The strength of the presented interpretative framework is the abundance of dimensions which allowed us to study data stories in depth and evaluate specific cases reliably. One obstacle we encountered during our work concerns the number of dimensions and the possibility to expand them indefinitely. By iterating over our system of dimensions we managed to set out its key elements. However, further scrutiny should be paid to the choice of dimensions and their importance in defining data stories. Similarly, since we analyzed aspects such as tone of story components, we had to deal with variable degrees of subjectivity in interpreting some of the elements. Although trying to mitigate subjective interpretation by iterating over the coding multiple times and with different coders, it is impossible and perhaps undesirable to entirely remove personal perspectives. When it comes to understanding the tone or message behind a visual or textual artifact, interpretation plays an important role. For this reason our interpretative framework is explicitly designed to accommodate subjectivity, while fostering consolidation of coders' interpretation. It is important to acknowledge that our findings are based on a relatively small set of data stories on a specific topic. To further validate the viability and study the transferability of the framework, more such studies on a wider range of topics need to be carried out.

Future research. Aside from its limitations, this systematic approach to data stories and the communicative function of data visualization therein poses a great opportunity for future work in visualization research and design. In this study, we report on the rationale of water-related data stories and on the communicative function of data visualization. We do so by rolling out an interpretative framework that comprises several key dimensions of data stories and allowing for comparison across news outlets and genres. As data stories develop further, both in terms of form and content, more thorough categorization systems are needed. There is the potential for this framework to be expanded and include several other aspects such as the role of interactivity, relevance of storytelling in maps [Rot21], the type and consistency of visual grammar across different types of artifacts, and the recurrence of visual models. Additionally, more content specific dimensions could be added to include other topics and types of data stories. To this extent, we hope this framework will be applied to other topics beyond water-related issues as important steps forward to its consolidation.

Acknowledgements

We thank our colleagues Linda Freyberg and Ilias Kyriazis for their feedback on a draft version of this paper. We are also very grateful for the thoughtful suggestions provided by the reviewers on the submitted manuscript. The research was in part supported by ANID - Doctorate Grant, Millennium Science Initiative Program Code ICN17 002 (IMFD) and the project *VIDAN: Visual and Dynamic Arrangements of Narratives* (13FH126PX6) funded by the German Federal Ministry of Education and Research (BMBF).

References

[ABC20] ABC NEWS (AUSTRALIA): Devoured, 2020. https://www.abc.net.au/news/2020-02-19/australia-bushfires-how-heat-and-drought-created-a-tinderbox/11976134,Last accessed on 2022-11-28. 8, 9

[AJ21] APPELGREN E., JÖNSSON A. M.: Engaging citizens for climate change—challenges for journalism. *Digital Journalism 9*, 6 (2021), 755–772. 2

[Al 20] AL JAZEERA: Saving the nile, 2020. https://interactive.aljazeera.com/aje/2020/saving-the-nile/index.html,Last accessed on 2022-11-28. 8, 9

[Ali21] ALIEVA I.: How american media framed 2016 presidential election using data visualization: The case study of the new york times and the washington post. *Journalism Practice* (2021), 1–27. 1, 2

[Bat17] BATEMAN J. A.: Issues for information design. *Information design: research and practice* (2017). 2

[Blo20] BLOOMBERG: Mapping america's underwater real estate, 2020. https://www.bloomberg.com/graphics/2020-flood-risk-zone-us-map/,Last accessed on 2022-11-28. 3, 8, 9

[Blo21] BLOOMBERG NEWS: Redlined, now flooding, 2021. https://www.bloomberg.com/graphics/2021-flood-risk-redlining/,Last accessed on 2022-11-28. 3, 9

[BLW11] BRANTNER C., LOBINGER K., WETZSTEIN I.: Effects of visual framing on emotional responses and evaluations of news stories about the gaza conflict 2009. *Journalism & Mass Communication Quarterly 88*, 3 (2011), 523–540. 2

[BMG*10] BATEMAN S., MANDRYK R. L., GUTWIN C., GENEST A., MCDINE D., BROOKS C.: Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2010), pp. 2573–2582. 2

[CD07] CHONG D., DRUCKMAN J. N.: Framing theory. *Annual review of political science 10*, 1 (2007), 103–126. 2

[Cha07] CHANDLER D.: *Semiotics: the basics*. Routledge, 2007. 2

[DFCC13] DÖRK M., FENG P., COLLINS C., CARPENDALE S.: Critical infovis: Exploring the politics of visualization. In *alt.chi 2013: Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems* (2013), ACM, pp. 2189–2198. 2

[DK20] D'IGNAZIO C., KLEIN L. F.: *Data feminism*. MIT press, 2020. 2

[Dru01] DRUCKMAN J. N.: The implications of framing effects for citizen competence. *Political behavior 23*, 3 (2001), 225–256. 2

[Ent93] ENTMAN R. M.: Framing: Towards clarification of a fractured paradigm. *McQuail's reader in mass communication theory 390* (1993), 397. 2, 3, 4, 5

[HCHC20] HOHMAN F., CONLEN M., HEER J., CHAU D. H. P.: Communicating with interactive articles. *Distill* (2020). 1

[HD11] HULLMAN J., DIAKOPOULOS N.: Visualization rhetoric: Framing effects in narrative visualization. *IEEE transactions on visualization and computer graphics 17*, 12 (2011), 2231–2240. 1, 2

[HDR*13] HULLMAN J., DRUCKER S., RICHE N. H., LEE B., FISHER D., ADAR E.: A deeper understanding of sequence in narrative visualization. *IEEE Transactions on visualization and computer graphics 19*, 12 (2013), 2406–2415. 1

[Hii17] HIIPPALA T.: The multimodality of digital longform journalism. *Digital journalism 5*, 4 (2017), 420–442. 2

[Hii20] HIIPPALA T.: A multimodal perspective on data visualization. *Data visualization in society* (2020). 2

[HL20] HERAVI B. R., LORENZ M.: Data journalism practices globally: Skills, education, opportunities, and values. *Journalism and Media 1*, 1 (2020), 26–40. 1

[Hul19] HULLMAN J.: Why authors don't visualize uncertainty. *IEEE transactions on visualization and computer graphics 26*, 1 (2019), 130–139. 1

[Ken01] KENSICKI L. J.: Deaf president now! positive media framing of a social movement within a hegemonic political environment. *Journal of Communication Inquiry 25*, 2 (2001), 147–166. 2

[KL05] KING C., LESTER P. M.: Photographic coverage during the persian gulf and iraqi wars in three us newspapers. *Journalism & Mass Communication Quarterly 82*, 3 (2005), 623–637. 2

[KM13] KOSARA R., MACKINLAY J.: Storytelling: The next step for visualization. *Computer 46*, 5 (2013), 44–50. 1

[Kni15] KNIGHT M.: Data journalism in the uk: A preliminary analysis of form and content. *Journal of Media practice 16*, 1 (2015), 55–72. 2

[Kon20] KONTINENTALIST: A river drained: Fish, rice, and food security in the mekong, 2020. https://cdn-images.kontinentalist.com/static-html/food-security-mekong-river-hydropower-dam-climate-change/index.html,Last accessed on 2022-11-28. 8, 9

[Mor13] MORTON T.: *Hyperobjects: Philosophy and Ecology after the End of the World*. U of Minnesota Press, 2013. 2

[Nær20] NÆRLAND T. U.: 4. the political significance of data visualization: Four key perspectives. *Data visualization in society* (2020), 63. 2

[New21] NEW YORK TIMES: This glacier in alaska is moving 100 times faster than normal, 2021. https://www.nytimes.com/interactive/2021/04/13/climate/muldrow-glacier-alaska-mount-denali.html, Last accessed on 2022-11-28. 9

[NPR21] NPR: Who will pay to protect tech giants from rising seas?, 2021. https://apps.npr.org/sea-level-rise-silicon-valley/,Last accessed on 2022-11-28. 9

[PBDSdV15] POWELL T. E., BOOMGAARDEN H. G., DE SWERT K., DE VREESE C. H.: A clearer picture: The contribution of visuals and text to framing effects. *Journal of communication 65*, 6 (2015), 997–1017. 2

[PHFH19] PARROTT S., HOEWE J., FAN M., HUFFMAN K.: Portrayals of immigrants and refugees in us news media: Visual framing and its effect on emotions and attitudes. *Journal of broadcasting & electronic media 63*, 4 (2019), 677–697. 2

[RD11] RODRIGUEZ L., DIMITROVA D. V.: The levels of visual framing. *Journal of Visual Literacy 30*, 1 (2011), 48–65. URL: https://doi.org/10.1080/23796529.2011.11674684, arXiv:https://doi.org/10.1080/23796529.2011.11674684, doi:10.1080/23796529.2011.11674684. 2, 3, 4

[Reu20] REUTERS: China's glaciers in retreat, 2020. https://graphics.reuters.com/CLIMATE-CHANGE/CHINA-GLACIER/rlgvdamqnpo/,Last accessed on 2022-11-28. 8, 9

[Reu21a] REUTERS: Devoured, 2021. https://graphics.reuters.com/GLOBAL-ENVIRONMENT/SAND-POYANG/qzjpqxxabvx/,Last accessed on 2022-11-28. 8, 9

[Reu21b] REUTERS: Scientists scramble to harvest ice cores as glaciers melt, 2021. https://graphics.reuters.com/CLIMATE-CHANGE/ICE-CORES/zjvqkjkjlvx/,Last accessed on 2022-11-28. 3, 6, 9

[RHDC18] RICHE N. H., HURTER C., DIAKOPOULOS N., CARPENDALE S.: *Data-driven storytelling*. CRC Press, 2018. 1

[Rot21] ROTH R. E.: Cartographic design as visual storytelling: synthesis and review of map-based narratives, genres, and tropes. *The Cartographic Journal 58*, 1 (2021), 83–114. 10

[SH10] SEGEL E., HEER J.: Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics 16*, 6 (2010), 1139–1148. 1, 2

[SH22] STOKES C., HEARST M.: Why more text is (often) better: Themes from reader preferences for integration of charts and text, 2022. URL: https://arxiv.org/abs/2209.10789, doi:10.48550/ARXIV.2209.10789. 1

[Sou20] SOUTH CHINA MORNING POST: China's worst floods in decades, 2020. https://multimedia.scmp.com/infographics/news/china/article/3094790/china-floods/index.html,Last accessed on 2022-11-28. 9

[Sou21] SOUTH CHINA MORNING POST: China floods: how henan province was inundated by historic rainfall, 2021. https://multimedia.scmp.com/infographics/news/china/article/3142519/index.html,Last accessed on 2022-11-28. 9

[The20] THE PUDDING: How is flooding affecting your community?, 2020. https://pudding.cool/projects/flooding/visuals, Last accessed on 2022-11-28. 9

[The21a] THE GUARDIAN: What tree rings reveal about america's megadrought, 2021. https://www.theguardian.com/us-news/ng-interactive/2021/jun/17/tree-rings-america-megadrought-visual,Last accessed on 2022-11-28. 3, 9

[The21b] THE STRAITS TIMES: Saving singapore's shores, 2021. https://www.straitstimes.com/multimedia/graphics/2022/01/singapore-protect-sea-levels-rise/index.html, Last accessed on 2022-11-28. 8, 9

[The21c] THE WASHINGTON POST: Klamath river, 2021. https://www.washingtonpost.com/nation/interactive/2021/klamath-river-basin-drought/,Last accessed on 2022-11-28. 9

[Tho93] THOMAS S.: Allan bell, the language of news media. oxford: Blackwell, 1991. pp. xv 277. 52.95(*pap*.18.95). - roger fowler, language in the news: Discourse and ideology in the press. london: Routledge, 1991. pp. xi 254. 59.95(*pap*.16.95). *Language in Society 22*, 1 (1993), 115–121. doi:10.1017/S004740450001695X. 2

[TL22] THIER K., LIN T.: How solutions journalism shapes support for collective climate change adaptation. *Environmental Communication 0*, 0 (2022), 1–19. URL: https://doi.org/10.1080/17524032.2022.2143842, doi:10.1080/17524032.2022.2143842. 10

[USA21] USA TODAY: Downpour, 2021. https://www.usatoday.com/in-depth/news/investigations/2021/11/30/climate-change-extreme-rainfall/8550366002/, Last accessed on 2022-11-28. 9

[vBMK*20] VAN BEEK L., METZE T., KUNSELER E., HUITZING H., DE BLOIS F., WARDEKKER A.: Environmental visualizations: Framing and reframing between science, policy and society. *Environmental Science & Policy 114* (2020), 497–505. 2, 3

[Wea07] WEAVER D. H.: Thoughts on agenda setting, framing, and priming. *Journal of communication 57*, 1 (2007), 142–147. 5

[WEK18] WEBER W., ENGEBRETSEN M., KENNEDY H.: Data stories: Rethinking journalistic storytelling in the context of data journalism. *Studies in communication sciences 2018*, 1 (2018), 191–206. 1, 2