The Influence of Data Storytelling on the Ability to Recall Information

Dominyk Zdanovic
Science, Policy and Information Studies
Department of Communication & Psychology
Aalborg University Copenhagen
Copenhagen, Denmark

Tanja Julie Lembcke
Science, Policy and Information Studies
Department of Communication & Psychology
Aalborg University Copenhagen
Copenhagen, Denmark

Toine Bogers
Science, Policy and Information Studies
Department of Communication & Psychology
Aalborg University Copenhagen
Copenhagen, Denmark
toine@hum.aau.dk

ABSTRACT

With ever-increasing amounts of complex data, we need compelling ways to distill this information into meaningful, memorable and engaging insights. Data storytelling is an emerging visualization paradigm that aims to "tell a story" with data in order to elicit deeper reflections in an effective manner. However, the effects of adding a narrative to a visualization on the memorability of the information remain speculative. Based on a review of related work, we synthesize a framework of data storytelling principles with concrete actions for every principle. We use this framework to design an online, controlled experiment to test compare traditional data visualizations with data storytelling visualizations in terms of their effects on short-term and long-term recall of information displayed in the visualizations. In general, despite long-held assumptions in the visualization community, we find no significant differences in recall between traditional visualizations and data storytelling visualization. However, we find indications that the cognitive load induced by different chart types and self-assessed prior knowledge on the chart topics could possibly have a moderating effect on information recall.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in visualization**; Visualization theory, concepts and paradigms; Visualization design and evaluation methods.

KEYWORDS

Data storytelling, data visualization, recall, memory, information interaction, learning effect

ACM Reference Format:
Dominyk Zdanovic, Tanja Julie Lembcke, and Toine Bogers. 2022. The Influence of Data Storytelling on the Ability to Recall Information. In *Proceedings of the 2022 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '22), March 14–18, 2022, Regensburg, Germany.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3498366.3505755

CHIIR '22, March 14–18, 2022, Regensburg, Germany

ACM ISBN 978-1-4503-9186-3/22/03…$15.00
https://doi.org/10.1145/3498366.3505755

1 INTRODUCTION

Traditional data visualization techniques for displaying raw data in a concise, graphical, and understandable way to deduce patterns, trends and general information have been around for decades. However, with ever-increasing amounts of complex data at our disposal, we need new and effective means of keeping up with this data deluge and distilling it into meaningful, memorable and engaging insights. *Data storytelling* (DS) is one such approach that allows us to compile data by telling a 'story' with numbers. The underlying assumption is that by conveying "data not just in numbers or charts, but as a narrative that humans can comprehend" [18], it is easier to tap into the emotional and decision-making areas of the brain, rather than just presenting facts and numbers like traditional data visualization techniques do [11].

However, researchers have only started to examine the effects and perceived benefits of data storytelling over traditional data visualization. While some work has shown indications of a positive effect of data storytelling visualizations on engagement [40], other work has shown no meaningful differences in terms of engagement [4, 5] or empathy [21, 25]. The picture on the influence of data storytelling principles on memorability of the information presented in the visualization is similarly muddled. One study found no difference in short-term recall between data storytelling and traditional visualizations, but data storytelling visualizations with relevant visual embellishments did seem to fair better in terms of long-term recall [2]. Despite the lack of empirical evidence, it is often assumed that applying data storytelling principles increases memorability and engagement [21]. If we increasingly use data storytelling visualizations in academic, business, and political contexts to communicate important information and provide an impact on the reader, one important criteria is that readers are also able to recall the key information in those visualizations. It is therefore important to more thoroughly investigate the effects that data storytelling visualizations can have on readers' recall of important information.

In this paper, we take another step towards determining the influence of data storytelling principles on the memorability (or recall) of information presented in visualizations. Based on a review of related work, we synthesize a framework of data storytelling principles with concrete actions for every principle. We then use this framework to design and conduct an online, controlled experiment to test compare traditional data visualizations with data storytelling visualizations in terms of their effects on short-term and long-term recall of information displayed in the visualizations. In our experiment, we focus on one of the most popular types: data visualizations annotated with narrative elements. Our setup enables us to answer the following research questions:

RQ1 How do data storytelling visualizations relate to traditional data visualizations and which principles characterize both techniques?
RQ2 How does the application of data storytelling principles to data visualization influence short-term recall of information?
RQ3 How does the application of data storytelling principles to data visualization influence long-term recall of information?

In general, despite long-held assumptions in the visualization community, we find no significant differences in short- or long-term recall between traditional visualizations and data storytelling visualization. However, we find indications that the cognitive load induced by different chart types and self-assessed prior knowledge on the chart topics could possibly have a moderating effect on information recall.

The rest of this paper is organized as follows. We review the related work on data visualization, data storytelling and their perceived effects on the next section. In Section 3, we present a framework of principles for producing data storytelling visualizations along with concrete actions, and contrast it with traditional data visualization principles. Sections 4 and 5 describe respectively the methodology and results of our online, controlled experiment. We discuss our findings and conclude in Section 6.

2 BACKGROUND

2.1 Data visualization

According to Tufte [36], *visualization* is a means to visually present measured quantities with specific visual elements e.g. lines, shapes, colors, words, symbols and similar. Traditional visualizations include (but are not limited to) line charts, pie charts, bar charts, area charts, bubble charts, scatter plots, tree maps, heat maps. Tufte introduced the term *graphical excellence* when talking about graphical displays to communicate complex ideas or data in a precise, clear and efficient manner. Later, Tufte [37] also coined the term 'data-ink ratio', which is the proportion of ink that is used to present actual data compared to the total amount of ink (or pixels) used in the entire display. He argued that the data-ink ratio should be as high as possible to facilitate better understanding of the visualization. While Tufte's approach of creating visualizations is considered "purist, traditional" and is still used to this day [30], there is also growing body of work that argues against his minimalist principles [2, 16, 30].

Data visualization is the research field that is focused on visually representing various types of data, and is strongly related to *information visualization*. It extends the traditional approach to visualizing data so it is more approachable and supports communication to broader audiences [13, 29]. There is no exact consensus on what constitutes data and information visualization [1]. Even though the term was originally coined by Robertson et al. [29], some researchers even consider Tufte as one one of its disciples.

Other seminal work in the field of (data) visualization includes the work by Cleveland and McGill [9] on how easily we can visually decode information from different visualization types. They compared different visualizations on their potential to effectively communicate data to readers. Among other things, they found that it is easier for people to decode data visualizations based on position along common scales than by using shading or color saturation. Later, this work was replicated and confirmed by Heer and Bostock [15] on a larger scale using crowdsourcing.

2.2 Data storytelling

Data storytelling is a more recent visualization paradigm that aims to combine a *narrative* with data in order to "tell a story" and elicit deeper reflections on the data the visualization represents [8, 19]. As argued by Chatman [7] and Zhang [38], narratives consist of two parts: (1) the story, which is shown in the narrative; and (2) the discourse, or how a narrative is shown. The story is further subdivided into two topics: events (actions, happenings) and existents. In simpler terms, the story is the *what* in a narrative that is depicted, the discourse is *how*.

Linking more 'traditional' data visualization approaches and storytelling was an idea coined by Knaflic [19]. He argued that, while data itself can be hard to understand, within it there is a story that one can bring up to life in order to communicate the core conclusions represented by a visualization more efficiently. In turn, this could support more efficient decision-making. Ryan [32], another proponent of data storytelling, described to this approach as *information compression*: smashing complicated information into manageable pieces by focusing on what is most important and then pretending it is whole and bound entirely within the visualization(s) used to illustrate the message.

According to a study by Nowak et al. [26], using narrative techniques to tell the data story also appears to pay off. In their study, they asked participants to describe both visualizations with data storytelling elements and more traditional visualizations. The descriptions of the data storytelling visualizations tended to resemble the intended narrative, while the descriptions of traditional visualization focused more on decoding and usability flaws in the visualizations, suggesting that narrative techniques work and are helpful in getting the message across.

Segel and Heer [33] argued that most visualization tools aim to produce visualizations for exploration and analysis only, thereby leaving it to the reader to find the story behind the graphic. They analyzed a large number of visualizations in the context of online journalism and defined three distinct features of this visual design space features: genre, visual narrative tactics and narrative structure tactics. In addition, based on their analysis they also proposed three common types of storytelling approaches: the martini glass structure, the interactive slideshow and the drill-down story. Another systematic analysis was performed by Hullman and Diakopoulos [17], who introduced their visualization rhetoric framework. They

[1]We will treat data visualization and information visualization as identical concepts in the rest of this paper.

analyzed which visual design tactics can result in the addition or remove of information, along with how these design choices changes the rhetoric of a visualization.

Data storytelling principles have been applied successfully in different domains. For instance, Thöny et al. [35] describe how the application of data storytelling principles to geographic visualization systems can help to better guide the audience and expand the accessibility of the information in these systems to a more general audience [22, 39]. In the field of journalism, data storytelling principles have also been applied successfully [27] as well as analyzed systemically in order to determine what makes an award-winning data story [24].

2.3 Effects of data storytelling

Only in recent years have researchers started to examine the effects of data storytelling over traditional data visualization on different constructs, such as engagement, empathy, and recall.

Engagement. Boy et al. [5] evaluated the impact of using narrative visualization techniques and storytelling on *engagement* with exploratory information visualizations. They conducted three web-based field experiments on a popular news outlet and on a popular visualization gallery website. Contrary to their expectations, narrative visualization techniques were not found to increase user engagement with exploratory visualizations [4, 5]. Zhao et al. [40] investigated the effectiveness of the comic strip narrative style for data-driven storytelling. They focused in particular on the partitioning (i.e., chunking complex visuals into manageable pieces) and the sequence of visualizations (i.e., organizing them into a meaningful order). Through a qualitative study, they found that the data comic format did increase engagement with and perceived usability of the visualizations.

Empathy. Liem et al. [21] examined the influence of data storytelling visualizations on another construct: *empathy* (or prosociality) in an immigration context. They compared three different visualizations: (1) personal visual narratives designed to evoke empathy; (2) structured visual narratives about groups of people; and (3) a traditional data visualization without narrative elements. They compared these three conditions in two crowdsourcing studies on the participants' attitudes towards immigration, but found no evidence that the two data storytelling conditions elicited more positive attitudes towards immigration than the traditional data visualization. Their work confirmed the findings Boy et al. [6], who compared anthropomorphized embellishments to standard embellishment for narrative visualizations and found a similar lack of influence on empathy. These findings were confirmed by Morais et al. [25], who tested an information-rich anthropomorphized data storytelling visualization against a simple bar chart in a crowdsourcing study and asked participants to donate hypothetical money. They found a similarly small effect on empathy.

Recall. In the visualization community, data storytelling is often assumed to increase memorability and, as a result, *recall* of information [21]. Storytelling in general has been shown to have an impact on information recall. Bower and Clark [3] tested the recall of words embedded in stories versus recall from a random list of words and found that participants who constructed stories remembered six to seven times more words than participants in the list condition. In a related experiment, Heath and Heath [14, pp. 42-44] asked students to give short presentations on crime statistics and asked them afterwards to write down everything they could remember from the speeches. They found that while 63% of students remembered the stories associated with statistics, only 5% could remember an individual statistic.

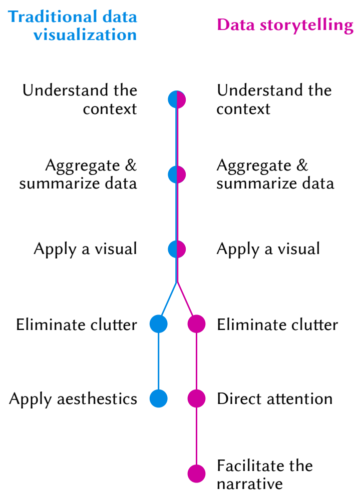
Bateman et al. [2] investigated the influence of visual embellishments, such as color enhancements, textual pointers and explanations, on both short and long-term recall of information, and compared this to the 'minimal' visualization style advocated by Tufte [36]. Bateman et al. [2] found that there were no differences between the two conditions in terms of short-term recall—immediately after examining the visualizations. However, after two to three weeks, participants recalled the charts with visual embellishments significantly better. Messages beyond the written text were significantly more often received in the condition with visual embellishments, and participants also found such charts more attractive, enjoyable, and easier to read and remember [2]. Li and Moacdieh [20] extended the work by Bateman et al. [2] by applying time limits to examining the visualizations. They found that both time limits and chart type significantly affected short-term recall performance. Chart type was also found to affect the time participants needed to review them as charts with visual embellishments required less time to review them [20]. This was later confirmed by Ragan et al. [28], who also discovered that it was the visual embellishments that were directly related to the data which improved recall, whereas unrelated visuals had the opposite effect.

Visual embellishments can contribute to weaving a narrative into a visualization, but this is not required. In contrast to the work by Bateman et al. [2]—which did not focus on data storytelling—we specifically focus on data storytelling visualizations and compare them to traditional data visualizations in order to determine the effect of the presence of a narrative on recall.

3 A FRAMEWORK OF DATA STORYTELLING PRINCIPLES

In order to reliably examine the effects of data storytelling visualizations on recall as compared to traditional data visualizations, we need a reproducible approach for generating data storytelling visualizations. A handful of researchers have come up with different guides—suited to different domains and usage contexts—on how to generate high-quality data visualizations, and how to effectively add a narrative to these visualizations. Our starting point in producing a framework for producing data storytelling visualizations was the work by Knaflic [19], Ryan [31], Segel and Heer [33], Hullman and Diakopoulos [17], Bach et al. [1], and Echeverria et al. [12], from which we collected a variety of of visualization and storytelling principles.

After merging similar principles, we devised a combined model of the principles behind both traditional data visualizations and data storytelling visualizations, which is shown in Figure 1a. It shows that while the initial stages of the two approaches are similar, they start diverging at the principle *Eliminate the clutter.* At this step, eliminating clutter for data storytelling visualizations explicitly means removing all unnecessary, non-data related information



(a) Principles of DV and DS

Phases	Principles	Concrete actions
Explore the data	**Understand the context** Define the audience and the purpose of the visualisation	– Brainstorm & research – Find data
	Aggregate & summarize data Define the main data points to deliver the purpose	– Clean data – Manipulate data if applicable
Craft the visualization	**Apply a visual** Translate data into a visualisation	– Select an appropriate graph type
	Eliminate clutter Reduce the cognitive load	– Apply Gestalt principles – Remove data labels, data markers, grid, legend, tick marks, axis label
	Direct attention Ensure visual guidance for the user	– Remove unnecessary data and push necessary, but non-focal data to the background – Choose preattentive attributes for focal data points e.g. orientation, shape, size, line length, hue, intensity, curvature
Tell the story	**Facilitate the narrative** Deliver the message	– Apply narrative text labels to focal data points – Ordering & Interactivity – Clear title delivering the main insight or outcome

(b) Data storytelling framework

Figure 1: Figure 1a (left) shows the where the principles behind creating traditional data visualizations vs. data storytelling visualizations are the same and where they diverge. Figure 1b (right) shows our framework for creating data storytelling visualizations as synthesized from the literature.

completely—tick marks, labels, grids and other to minimize the cognitive load to the minimum. However, for traditional visualizations, such 'drastic' measures need not always be applied. If there is no message or outcome explicitly stated, traditional visualizations still require 'background' information such as axis labels, the legend, and a grid with corresponding values for the user to be able to explore each data value. Nevertheless, for traditional data visualizations these also tend to be edited to minimize the cognitive load, such as by choosing a simple background grid, or by adjusting the breaks for numeric values on the axes.

After their divergence, the next stage in creating a data storytelling visualization is to *direct attention* to specific data points or trends rather than explicitly displaying the whole dataset. In contrast, the last stage of traditional data visualization is to *apply aesthetics*, which includes choosing appropriate colors, readable fonts, or general stylistic themes. This stylization is also a part of data storytelling visualization, but more for the purpose of drawing attention to certain data points by, for instance, intensifying the color of important points while attenuating non-focal (yet necessary) data points.

Finally, in data storytelling visualizations we need to *facilitate the narrative* by adding narrative elements. This is a core principle of data storytelling: creating a story by explaining the trends with insights or by highlighting clear outcomes in titles or other annotations. This guided explanation of the data is in contrast to traditional visualizations, which requires readers to explore the visualization on their own and draw their own conclusions.

To facilitate the reproducible generation of data storytelling visualizations, we expanded on each principle in Figure 1a by listing concrete actions to be taken in that stage. The complete framework for generating data storytelling visualizations is shown in Figure 1b. We furthermore grouped the principles together into three main phases, resulting in a framework with three levels:

(1) **Phases** reflect the thematic focus of the principles they comprise and are three-fold: *Explore the data*, *Craft the visualization* and *Tell the story*.
(2) **Principles** represent the individual steps needed to reach the final visualization, which include (1) *Understand the context*, (2) *Aggregate & summarize the data*, (3) *Apply a visual*, (4) *Eliminate clutter*, (5) *Direct the attention*, and (6) *Facilitate the narrative*.
(3) **Concrete actions** further elaborate on the principles. These actions should be understood as measures to take only if applicable. Whether an action is required, depends on the data, domain, context, and the desired outcome.

We applied our own framework to produce the visualizations used in online, controlled experiment, which will be described in more detail in the next section.

4 METHODOLOGY

4.1 Design

In order to test the influence of data storytelling visualization vs. traditional data visualization on recall, we designed a controlled experiment where we manipulated a single independent variable: *visualization type* with traditional data visualization (DV, condition 1) and data storytelling visualization (DS, condition 2) as its two conditions. Our dependent variables were the short-term and

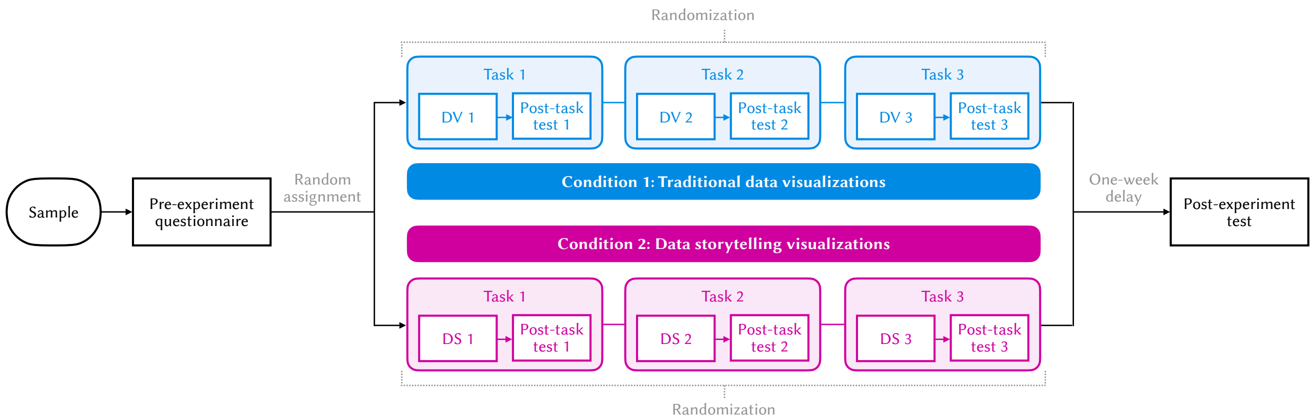


Figure 2: Illustration of our between-group design with two conditions (traditional data visualization vs. data storytelling visualizations). After a pre-experiment questionnaire, participants are randomly assigned to one of these two conditions. All participants are presented with three different visualizations in randomized order and tested on their short-term recall of the information in these visualizations. After a one-week delay, participants are tested on their (long-term) recall once more.

long-term recall of the information contained in these visualizations We chose a between-group design, because we expected a strong learning effect if participants were to be shown both a DS and DV visualization on the same topic. It is very likely that information learned from the first visualization is simply reinforced when viewing the second visualization, making it impossible to accurately measure information recall after the second visualization.

After being surveyed on their pre-existing knowledge about the three visualization topics, each participant was randomly assigned to either the DS or DV condition and shown three different visualizations from that condition in random order, as shown in Figure 2. Section 4.2.1 describes which topics we selected and how we created two different types of visualizations for each topic.

After each of these three visualization tasks, participants were presented with a post-task questionnaire containing five questions designed to test their immediate, *short-term* recall of the information included in the preceding visualization. To test the participants' *long-term* recall, participants were sent a follow-up message one week after completing the main experiment inviting them to complete the post-experiment test of their knowledge of the three visualization topics, This one-week interval between the two tests was chosen based on related work [2, 20, 28].

We believe this study design allows us to compare DS and DV visualizations based on the recall of the information therein contained, both immediately after and after a delay period. While the authors' home university only has an Institutional Review Board available for medical studies, we had a senior faculty member review the study design and propose changes where necessary.

4.2 Materials and Equipment

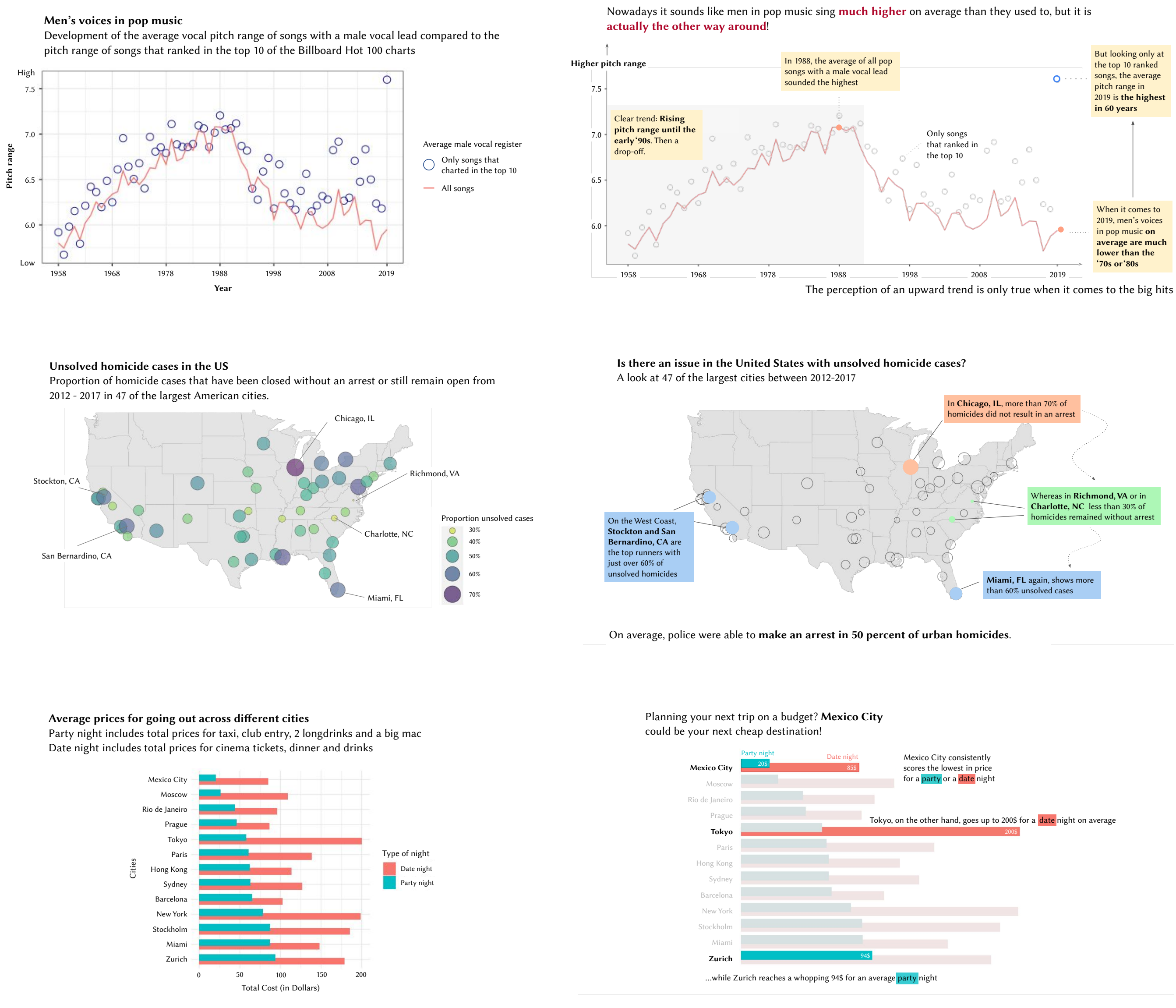
Due to Covid-19 restrictions, we conducted our experiment online. We used Gorilla[2], an online platform for designing and conducting experiments. Participant recruitment took place using Prolific[3], an online participant recruitment platform. Both services can be connected so that participants recruited on Prolific are automatically sent to the linked online experiment on Gorilla. A reminder was automatically sent from Gorilla through Profilic to participants one week after they completed the initial experiment.

4.2.1 Visualization development. We used different tools for developing the three DS and DV visualizations in each condition. R was used to develop the baseline DV visualizations for three different topics, after which Sketch[4], a vector graphics editor, was used to edit the visualizations and add the narrative elements necessary to transform them into DS visualizations.

We selected three different datasets to create our three visualizations and created two variants of each visualization, one for each condition. Both the DV and DS visualizations are about the same topic and provide the same information, with the DS visualization being a narrative modification of the baseline DV visualization. We purposely selected topics that the general public can be expected not to have any does prior knowledge of so as to avoid introducing bias to our results. If many participants already know about topic X, then it becomes impossible to separate prior knowledge from knowledge acquired by studying the visualization.

Men's voices in pop music. The dataset for the first visualizations originated on a journalism site called *The Pudding*. The original dataset[5] contains data about the changes in the pitch of male singing voices over time. We used this dataset to show how the pitch range of male singers has changed over the past six decades and to disprove the the claim that it seems that nowadays male singers are singing "higher than ever before" [34]. Figure 3 shows the baseline DV visualization (left) created based on this dataset by following the principles outlined in Figure 1 as well as the DS visualization (right) created based on this baseline visualization. We created a line chart to show the change in pitch over time and combined it with a scatterplot for the top-10 hits.

[2] http://gorilla.sc/
[3] https://www.prolific.co/
[4] http://www.sketch.com/
[5] https://github.com/the-pudding/data/tree/master/names-in-songs



(a) Traditional data visualization

(b) Data storytelling visualization

Figure 3: Overview of the three visualization sets used in our experiment. Traditional data visualization (left) and data storytelling visualization (right) are shown for *Men's voices in pop music* (top), *Unsolved US homicides* (middle) and *Costs of a night out* (bottom).

Unsolved US homicides. In an investigative story published by the Washington Post, Mellnik et al. [23] provided insights into the status quo on unsolved homicides in the US. The authors used data from nearly 55,000 homicides between 2012 and 2017. The data included the location of the murder, whether an arrest was made and basic demographic information about each victim. We decided to summarize the dataset[6] by creating a map-based visualization which—for the 50 largest American cities—displays the relative frequency of solved homicide cases using different-sized circles.

Costs of a night out. The last dataset used in our controlled experiment is about the costs of a night out in different cities across the world. The dataset[7] was used to display a grouped horizontal bar chart of a list of cities and the average prices of going out on both a party night and a date night.

4.3 Participants

We used purposive sampling to recruit the participants for our study using Prolific. Prolific allowed us to establish certain pre-screening

[6]https://github.com/washingtonpost/data-homicides/

[7]https://data.world/makeovermonday/2018w48

criteria which emerged out of necessity due to the nature of the three visualization pairs we developed. We recruited participants who met the following requirements:

- **Education level.** Participants had to have a bachelor degree or higher.
- **Nationality.** US citizens were excluded because they were more likely to have prior knowledge about the US homicide visualizations.
- **Vision.** The visualizations we developed were not entirely suitable for people suffering from color blindness, so these were removed from consideration.
- **Language fluency.** In order to understand our instructions, question and labels in the visualizations, participants were required to be fluent in English.

We are aware that especially the first two requirements could hurt the generalizability of our findings. Apart from participants having to be 18 or older to be able to participate through Prolific, we did not adopt any age-related recruitment criteria, because we are not aware of any work that has shown a meaningful influence of age of how people interpret data visualizations.

We recruited 130 participants in total. Based on pre-screening using our pre-experiment questionnaire (described in more detail in Section 4.4.1), 24 participants were excluded, because they were too knowledgable about at least one of the three topics. Another 12 participants were dropped from the dataset, because they did not complete the post-experiment questionnaire to complete the entire experiment. Of the remaining 94 participants, 44 were randomly assigned to the DV condition and 50 to the DS condition. All participants were financially compensated for their time according to the guidelines provided by Prolific. Our final sample contained participants from various nationalities, which Portugal ($n = 26$), Poland ($n = 13$), and Italy ($n = 11$) making up the biggest groups. Participants's gender was distributed fairly equally with 44 female and 50 male participants. Our participants' ages ranged from 19 to 56 years old with an average age of 29. We found no significant differences in recall performance based on these demographic characteristics.

4.4 Procedure

4.4.1 Pre-experiment questionnaire. After agreeing to participate in our experiment on Prolific, participants were informed about the purpose, structure, and duration of the study. They were then directed to the online experiment on Gorilla and asked to sign a consent form. All participants were automatically anonymized by Prolific and only entered into Gorilla with a user ID. In the first step, the pre-test questionnaire, participants were asked to answer three Likert-scale questions about their prior knowledge about the three visualization topics.

4.4.2 Post-task tests. As shown in Figure 2, participants were randomly assigned to one of the two conditions after completing the pre-test questionnaire. Each condition consisted of three tasks, corresponding to the three visualizations described in Section 4.2.1, with the tasks assigned in random order. After each of these three visualization tasks, participants were presented with a post-task test of their knowledge of the visualization. For each of the three topics, the same post-task test questionnaires were used, regardless of the condition.

The post-task questionnaires containing five questions designed to test their short-term recall of the information included in the preceding visualization, four of which were forced-choice questions about important elements of the visualization [10]. An example question for the bottom visualizations in Figure 3 is "*Is the following statement true or false: In general, across all cities, it is more expensive to go on a date night than to go to a party?*"[8]. When formulating these questions, we made sure that none of them could only be answered using the information from the narrative elements in the DS visualizations. Each post-task test concluded with an open question on each topic asking the participant to write down any additional pieces of information they could remember about the preceding visualization. These open questions were intended in order to examine any additional information that participant could recall, but was not tested in the forced-choice questions. Answers to the open questions were analyzed thematically.

After completing the three tasks, participants were directed back to Prolific. Participants were then told that the second part of the experiment would take place in seven days. To ensure that the they would come back to complete the second part of the experiment, participants were only paid after completing both parts of the experiment.

4.4.3 Post-experiment test. After one week, all participants received a notification through Prolific asking them to head back to Gorilla to complete the post-experiment test questionnaire. This questionnaire consisted of four forced-choice questions targeting the same visualization facts as the post-task questionnaires in the main experiment. However, to avoid easy recognition of their previous answers, we paraphrased each question and their answer options. We also concluded the post-experiment test questionnaire with an open question. Finally, participants were redirected to Prolific for the final time and reimbursed for their participation.

5 RESULTS & ANALYSIS

5.1 Short-term recall

We first compare our two conditions—traditional data visualization (DV) and data storytelling (DS)—on their effects on immediate, short-term recall. Each correctly answered forced-choice question was awarded with one point. With three tasks and four questions per task, this meant that participants could score between 0 and 12 points on their short-term recall test.

Recall performance of our participants was relatively high for both conditions. DS visualization resulted in the highest scores ($M = 9.74$, $SD = 1.85$), but the traditional DV condition was close in performance ($M = 9.48$, $SD = 1.71$). well as for the DS condition, as also shown in Figure 4a. In order to test for a significant difference between the mean scores, we first checked both score distributions for normality. Both Q-Q plots and the Shapiro-Wilk test for normality indicated that the scores in both the DV condition ($W(44) = .89$, $p < .001$) and the DS condition were not normally

[8] A complete list of both the post-task and post-experiment questions along with other relevant data can be found at https://doi.org/10.5281/zenodo.5821922.

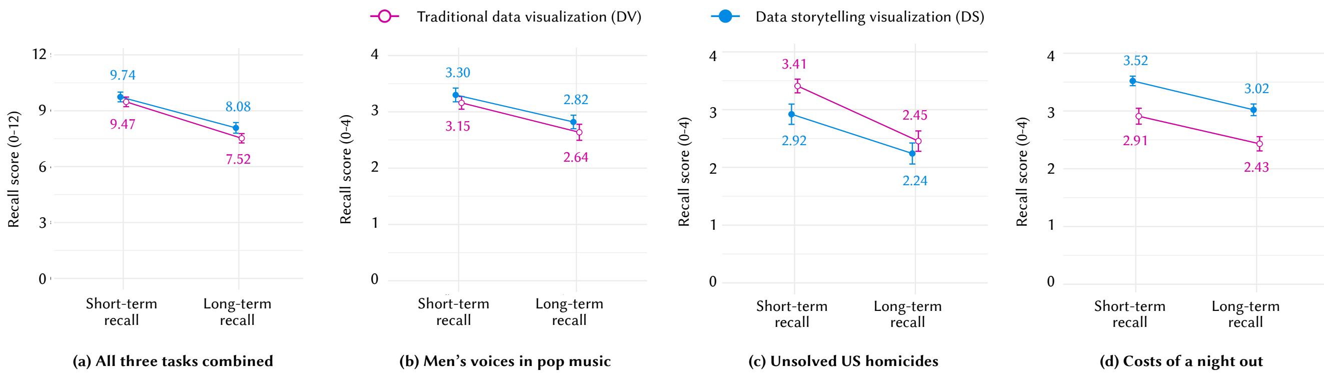


Figure 4: Short- and long-term recall performance in our two conditions for (a) all three tasks combined, (b) men's voices in pop music, (c) unsolved US homicides, and (d) costs of a night out.

distributed ($W(50) = .89$, $p < .001$). As a result, we used the Mann-Whitney U test to test for significant differences between the two conditions. Short-term recall scores summed over all three tasks combined were higher in the DS condition (*mean rank* = 49.97) than in the DV condition (*mean rank* = 44.69). However, a Mann-Whitney U test revealed that this difference was not statistically significant ($U(N_{DV} = 44, N_{DS} = 50) = 1223.5$, $z = 0.96$, $p = .34$).

Next, we examined the short-term recall score differences for each task separately, which are shown in Figure 4b-d. For *Men's voices in pop music*, participants in the DS condition scored better on recall, but this difference was not significant ($U(N_{DV} = 44, N_{DS} = 50) = 1248$, $z = 1.21$, $p = 0.23$). Perhaps surprisingly, the traditional data visualization for *Unsolved US homicides* outperformed the DS visualization, but this difference was also not significant ($U(N_{DV} = 44, N_{DS} = 50) = 875$, $z = 1.84$, $p = 0.07$). However, for the *Costs of a night out* task, we do find that short-term recall performance in the DS condition (*mean rank* = 55.90) was significantly higher than in the DV condition (*mean rank* = 37.95) according to a Mann-Whitney U test ($U(N_{DV} = 44, N_{DS} = 50) = 1520.0$, $z = 3.46$, $p < .001$).

5.2 Long-term recall

Calculating long-term recall performance on the post-experiment questionnaire was done in the same way as for short-term recall with a score range of 0 to 12 points. In general, overall recall performance dropped for both conditions after a week, with the difference between the two conditions becoming bigger. However, the same patterns present for short-term recall were also present for long-term recall. The DS visualizations ($M = 8.08$, $SD = 2.01$) again resulted in the higher scores than the traditional DV condition ($M = 7.52$, $SD = 1.66$). well as for the DS condition, as also shown in Figure 4a.

Long-term recall scores were also not normally distributed, so we used a Mann-Whitney U test again to test for significant differences between the conditions. Long-term recall scores aggregated over all three tasks were higher in the DS condition (*mean rank* = 51.03) than in the DV condition (*mean rank* = 43.49). However, this difference was again not statistically significant ($U(N_{DV} = 44, N_{DS} = 50) = 1276.5$, $z = 1.35$, $p = .17$).

Examining the long-term recall score differences for each individual visualization task (Figure 4b-d) showed the same patterns as for short-term recall. For *Men's voices in pop music*, participants in the DS condition scored better on recall, but this difference was not significant ($U(N_{DV} = 44, N_{DS} = 50) = 1211$, $z = 0.89$, $p = 0.38$). Again, the traditional data visualization for *Unsolved US homicides* outperformed the DS visualization, but this was also not significant ($U(N_{DV} = 44, N_{DS} = 50) = 985$, $z = 0.89$, $p = 0.37$). For the *Costs of a night out* task, long-term recall performance was significantly higher in the DS condition (*mean rank* = 55.92) than in the DV condition (*mean rank* = 37.93) according to a Mann-Whitney U test ($U(N_{DV} = 44, N_{DS} = 50) = 1521.0$, $z = 3.45$, $p < .001$).

5.3 Delay effects

Unsurprisingly, recall performance worsens after a week across conditions and tasks, as visible in all subcharts in Figure 4. We performed Wilcoxon signed-rank tests to determine whether this drop-off was significant, separated by condition. For the DV condition, the drop-off in recall performance after a week was indeed statistically significant according to a Wilcoxon signed-rank test ($N = 44$, $z = 5.18$, $p < 0.001$). For the DS condition, this difference was also significant ($N = 50$, $z = 5.13$, $p < 0.001$). The drop-off in recall performance is also significant for each of the three individual tasks, as shown in Figure 4. This suggests that our choice to separate short-term and long-term recall by a week was long enough to observe a significant drop in recall performance.

Another relevant question is whether the gap between the two conditions widens after a week or whether it stays the same. The difference between short-term and long-term recall is indeed smaller for the DS condition ($M_{DS} = 1.66$, $SD = 1.66$) than for the DV condition ($M_{DV} = 1.95$, $SD = 1.74$), which suggests that data story telling visualizations could mitigate the effects of memory attenuation better than traditional visualizations. However, this difference was not statistically significant according to a Mann-Whitney U test $U(N_{DV} = 44, N_{DS} = 50) = 1012.5$, $z = 0.67$, p = 0.50).

5.4 Influence of prior knowledge

Even though we filtered out participants that were too knowledgable about at least one of the three topics according to their self-assessment in the pre-experiment questionnaire, prior knowledge may still influence recall performance. Prior knowledge about a topic, even a little, could be reinforced upon seeing a visualization, which could make it easier to recall information from the visualization. To examine the possible influence of prior knowledge, we correlated their self-assessed knowledge about a task topic and the short-term recall scores they achieved. We further split our analysis by experimental condition and used Spearman's ρ as our correlation coefficient. For *Men's voices in pop music* and *Unsolved US homicides*, there was no significant correlation between prior knowledge and short-term recall score. On the *Costs of a night out* task, there was a significant, moderately negative correlation between self-assessed prior knowledge and short-term recall after being exposed to the traditional data visualization ($\rho(44) = -0.34$, $p < .05$). This suggests that people who thought they knew more about the costs of a night in different cities around the world performed worse on the recall test. However, in the data storytelling condition, there was no significant correlation between prior knowledge and short-term recall performance ($\rho(50) = -0.13$, $p = .34$). The discrepancy between prior knowledge and short-term recall scores may be due to incorrect preconceptions about costs being harder to dislodge, perhaps because people had different cities in mind when assessing their prior knowledge. Perhaps the presence of narrative elements in the DS condition is enough to do away with these misconceptions, where the baseline DV visualization is not able to succeed in this.

6 DISCUSSION & CONCLUSIONS

In this paper, we presented the results of an online, controlled experiment designed to compare traditional data visualizations with data storytelling visualizations in terms of their effects on short-term and long-term recall of information displayed in the visualizations.

We did not find any significant differences in either short-term or long-term recall between our two experimental conditions, even though in general average recall scores were higher after being exposed to data storytelling visualizations, both immediately after completing the tasks and after a delay period of one week. This means we cannot confirm assumptions held by parts of the data visualization community that data storytelling has a positive effect on recall of information [21].

When drilling down into the three different visualization tasks, we do find indications that chart type has a moderating influence on the effects of including narrative elements on top of a baseline visualization. For our *Costs of a night out* task, where the baseline visualization was a bar chart, recall performance was significantly higher in the data storytelling condition. In the case of *Men's voices in pop music*, where the baseline visualization was combination of a line chart and a scatterplot, the DS condition also saw higher recall scores, but not significantly so. However, for the *Unsolved US homicides*—a map visualization combined with proportional circular areas—adding narrative elements appeared to decrease recall performance, albeit not significantly so. Earlier work by Heer and Bostock [15] and Cleveland and McGill [9] has shown that chart types differ in the cognitive load they invoke when reading and decoding them, and we believe this is a plausible explanation for our results. According Heer and Bostock [15], bar charts are the easiest charts to decode, followed by line charts, with circular area comparisons being significantly more difficult to decode. Adding narrative elements to a chart is likely to increase the cognitive effort required to process the visualization. If too much effort has to go into decoding the baseline visualization, then adding an extra narrative layer on top could simply overload the reader and hurt recall performance. While the influence of cognitive effort on recall remains conjectural at this point, the work by Heer and Bostock [15] and Cleveland and McGill [9] makes it a plausible explanation that merits further study.

Another possible influence on recall performance could be the personal relevance of the visualization's topic for the participant—it is possible that reading about the *Costs of a night out* in different cities was simply more interesting to participants than the other topics, resulting in higher scores. Finally, one could also argue that by adding a narrative to the visualization, cognitive effort is reduced, because the reader does not have to extract the key findings from the visualizations on their own anymore. At any rate, more work is needed to better tease apart the influence of chart type and narrative elements on recall.

When considering the difference between short-term and long-term recall, we found that for all tasks and conditions there was a significant drop-off in recall performance. Without reinforcement of the information presented in the visualizations, this is to be expected. The drop-off in performance was steeper for the traditional data visualizations than for the data storytelling condition, which could suggest that narrative embellishments could help mitigate memory attenuation better than only the baseline visualizations. However, this difference was not significant; perhaps a longer delay period would have shown an even stronger effect. In their work, Bateman et al. [2] used a delay period of 2-3 weeks and they did find a significant difference in long-term recall (but not in short-term recall). However, they focused only on visual embellishments and had a different experimental setup, so their results are not directly comparable.

Finally, while in most cases prior knowledge of a topic did not seem to have an influence on recall performance, the picture is not completely clear. For visualizations that are less cognitively demanding, there is a chance that narrative elements could be more successful in correcting incorrect prior held beliefs. However, more focused work is needed to be able to draw any strong conclusions about this potential relationship.

6.1 Future work

There are many fruitful avenues of future work that follow from our experimental findings, foremost being the possible relationship between chart type complexity and data storytelling. We found sufficient indications that chart types that required more cognitive effort to decode did not benefit from adding narrative elements to them. However, we only compared three different chart types, so our results are far from conclusive. Future work should look into using the work by Heer and Bostock [15] to select more and more different chart categories by cognitive load and then examine the benefits of data storytelling for each of them. It would also be interesting to examine whether a chart's cognitive load has an impact on engagement, empathy or other attributes. Explicitly measuring cognitive load and modeling it as a dependent variable would also be a useful next step.

Future studies should also consider the influence of motivation for participating in studies like ours, as also argued by Knaflic [19]. Our participants received a monetary reimbursement, but placing the study in a different context could lead to more valid results. One such example could be situating it in a learning environment, where the motivation is to achieve certain learning goals. In general, prior interest in a topic—as opposed to prior knowledge—could also influence people's performance and should be measured in future experiments. Including the participants' cultural background could also be prudent, as previous research has indicated that the cultural context can affect visualization perception [17].

When producing our data visualizations, we were guided by our framework of data storytelling principles synthesized from related work. However, we limited our data storytelling style to the so-called 'Martini Glass' structure [33]. Other narrative storytelling styles could potentially lead to different results.

6.2 Conclusions

While more work needs to be done, through our framework of data storytelling principles and the findings from our online, controlled experiments, we have taken a solid next step towards examining the influence of data storytelling visualizations on the recall of information highlighted in those visualizations.

REFERENCES

[1] Benjamin Bach, Moritz Stefaner, Jeremy Boy, Steven Drucker, Lyn Bartram, Jo Wood, Paolo Ciuccarelli, Yuri Engelhardt, Ulrike Koeppen, and Barbara Tversky. 2018. Narrative design patterns for data-driven storytelling. In *Data-driven storytelling*. AK Peters/CRC Press, 107–133.

[2] Scott Bateman, Regan L. Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher Brooks. 2010. Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts. In *CHI '10: Proceedings of the 28th International Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 2573–2582. https://doi.org/10.1145/1753326.1753716

[3] Gordon H Bower and Michal C Clark. 1969. Narrative stories as mediators for serial learning. *Psychonomic science* 14, 4 (1969), 181–182.

[4] Jeremy Boy. 2015. *Engaging the People to Look Beyond the Surface of Online Information Visualizations of Open Data*. Ph.D. Dissertation. Telecom ParisTech.

[5] Jeremy Boy, Francoise Detienne, and Jean-Daniel Fekete. 2015. Storytelling in Information Visualizations: Does It Engage Users to Explore Data?. In *CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1449–1458. https://doi.org/10.1145/2702123.2702452

[6] Jeremy Boy, Anshul Vikram Pandey, John Emerson, Margaret Satterthwaite, Oded Nov, and Enrico Bertini. 2017. Showing People Behind Data: Does Anthropomorphizing Visualizations Elicit More Empathy for Human Rights Data?. In *CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 5462–5474. https://doi.org/10.1145/3025453.3025512

[7] Seymour Chatman. 1978. *Story and discourse : narrative structure in fiction and film*. Cornell University Press Ithaca, N.Y. 277 p. pages.

[8] Siming Chen, Jie Li, Gennady Andrienko, Natalia Andrienko, Yun Wang, Phong H. Nguyen Cagatay, and Turkay. 2020. Supporting Story Synthesis: Bridging the Gap between Visual Analytics and Storytelling. *IEEE Transactions on Visualization and Computer Graphics* 26, 7 (2020), 2499–2516. https://doi.org/10.1109/tvcg.2018.2889054

[9] William S. Cleveland and Robert McGill. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. Amer. Statist. Assoc.* 79, 387 (1984), 531–554. https://doi.org/10.1080/01621459.1984.10478080

[10] Douglas W. Cunningham and Christian Wallraven. 2012. *The task*. CRC Press, 39–50.

[11] Brent Dykes. 2015. Data storytelling: What it is and how it can be used to effectively communicate analysis results. https://hstalks.com/article/619/data-storytelling-what-it-is-and-how-it-can-be-use/ Accessed on 11/30/2020.

[12] Vanessa Echeverria, Roberto Martinez-Maldonado, and Simon Buckingham Shum. 2017. Towards Data Storytelling to Support Teaching and Learning. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction* (Brisbane, Queensland, Australia) *(OZCHI '17)*. Association for Computing Machinery, New York, NY, USA, 347–351. https://doi.org/10.1145/3152771.3156134

[13] Nahum Gershon and Ward Page. 2001. What Storytelling Can Do for Information Visualization. *Communications of the ACM* 44, 8 (August 2001), 31–37. https://doi.org/10.1145/381641.381653

[14] Chip Heath and Dan Heath. 2007. *Made to Stick: Why Some Ideas Survive and Others Die*. Random House, New York, NY.

[15] Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 203–212. https://doi.org/10.1145/1753326.1753357

[16] Steven Heller. 2006. *Nigel Holmes: On Information Design*. Jorge Pinto Books Inc.

[17] Jessica Hullman and Nicholas Diakopoulos. 2011. Visualization Rhetoric: Framing Effects in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2231–2240. https://doi.org/10.1109/tvcg.2011.255

[18] Miro Kazakoff. 2020. The Next Chapter in Analytics: Data Storytelling. https://mitsloan.mit.edu/ideas-made-to-matter/next-chapter-analytics-data-storytelling Accessed on October 5, 2021.

[19] C.N. Knaflic. 2015. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley.

[20] Huiyang Li and Nadine Moacdieh. 2014. Is "chart junk" useful? An extended examination of visual embellishment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 58, 1 (2014), 1516–1520. https://doi.org/10.1177/1541931214581316

[21] J. Liem, C. Perin, and J. Wood. 2020. Structure and Empathy in Visual Data Storytelling: Evaluating their Influence on Attitude. *Computer Graphics Forum* 39, 3 (2020), 277–289. https://doi.org/10.1111/cgf.13980

[22] Artur Lugmayr, Erkki Sutinen, Jarkko Suhonen, Carolina Islas Sedano, Helmut Hlavacs, and Calkin Suero Montero. 2016. Serious storytelling – a first definition and review. *Multimedia Tools and Applications* 76, 14 (2016), 15707–15733. https://doi.org/10.1007/s11042-016-3865-5

[23] Ted Mellnik, Steven Rich, Kimbriell Kelly, and Wesley Lowery. 2018. Homicide database: Mapping Unsolved Murders in Major U.S. Cities. *The Washington Post* (June 2018). https://www.washingtonpost.com/graphics/2018/investigations/unsolved-homicide-database/

[24] Ronald Metoyer, Qiyu Zhi, Bart Janczuk, and Walter Schreier. 2018. Coupling Story to Visualization. *23rd International Conference on Intelligent User Interfaces* (2018). https://doi.org/10.1145/3172944.3173007

[25] Luiz Morais, Yvonne Jansen, Nazareno Andrade, and Pierre Dragicevic. 2021. Can Anthropographics Promote Prosociality?A Review and Large-Sample Study. In *CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, Article 611. https://doi.org/10.1145/3411764.3445637

[26] Stanislaw Nowak, Lyn Bartram, and Thecla Schiphorst. 2018. A Micro-Phenomenological Lens for Evaluating Narrative Visualization. *2018 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)* (2018). https://doi.org/10.1109/beliv.2018.8634072

[27] Adegboyega Ojo and Bahareh Heravi. 2017. Patterns in Award Winning Data Storytelling. *Digital Journalism* 6, 6 (2017), 693–718. https://doi.org/10.1080/21670811.2017.1403291

[28] Eric Ragan, Alyssa Peña, and Lane Harrison. 2020. Memorability of Enhanced Informational Graphics: The effects of design relevance and chart type on recall. *Interdisciplinary Journal of Signage and Wayfinding* 4 (2020). https://doi.org/10.15763/issn.2470-9670.2020.v4.i1.a54

[29] George G. Robertson, Stuart K. Card, and Jack D. Mackinlay. 1993. Information Visualization Using 3D Interactive Animation. *Commun. ACM* 36, 4 (April 1993), 57–71. https://doi.org/10.1145/255950.153577

[30] Lindy Ryan. 2016. *Visual Communication and Literacy*. Elsevier Science & Technology, San Francisco, 109–129.

[31] Lindy Ryan. 2016. *The Visual Imperative: Creating a Visual Culture of Data Discovery*. Elsevier Science & Technology.

[32] Lindy Ryan. 2016. *Visual Storytelling with Data*. Elsevier Science & Technology, San Francisco, 131–151.

[33] Edward Segel and Jeffrey Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1139–1148. https://doi.org/10.1109/tvcg.2010.179

[34] The Pudding. 2019. Are Men Singing Higher in Pop Music? https://pudding.cool/2019/08/register/ Accessed on October 7, 2021.

[35] Matthias Thöny, Raimund Schnürer, René Sieber, Lorenz Hurni, and Renato Pajarola. 2018. Storytelling in Interactive 3D Geographic Visualization Systems. *ISPRS International Journal of Geo-Information* 7, 3 (2018), 123. https://doi.org/10.3390/ijgi7030123

[36] Edward R. Tufte. 1983. *The Visual Display of Quantitative Information*. 197 pages.

[37] E. R. Tufte. 1990. Data-Ink Maximization and Graphical Design. *Oikos* 58, 2 (1990), 130. https://doi.org/10.2307/3545420
[38] Yangjinbo Zhang. 2018. Converging Data Storytelling and Visualisation. In *Entertainment Computing – ICEC 2018*, Esteban Clua, Licinio Roque, Artur Lugmayr, and Pauliina Tuomi (Eds.). Springer International Publishing, Cham, 310–316.
[39] Yangjinbo Zhang and Artur Lugmayr. 2019. Designing a User-Centered Interactive Data-Storytelling Framework. *Proceedings of the 31st Australian Conference on Human-Computer-Interaction* (2019). https://doi.org/10.1145/3369457.3369507
[40] Zhenpeng Zhao, Rachael Marr, Jason Shaffer, and Niklas Elmqvist. 2019. Understanding Partitioning and Sequence in Data-Driven Storytelling. *Information in Contemporary Society Lecture Notes in Computer Science* (2019), 327–338. https://doi.org/10.1007/978-3-030-15742-5_32