



Evaluating narrative visualization: a survey of practitioners

Nina Errey¹ · Jie Liang¹ · Tuck Wah Leong¹ · Didar Zowghi¹

Received: 27 February 2023 / Accepted: 13 March 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract

Narrative visualization is characterized by the integration of data visualization and storytelling techniques. These characteristics provide challenges in its evaluation. Little is known about how these evaluation challenges are addressed by narrative visualization practitioners. We surveyed experienced narrative visualization practitioners to investigate their methods of evaluation. To gain deeper insight we conducted a series of semi-structured interviews with practitioners. We found that there is usually an informal approach to narrative visualization evaluation, where practitioners rely on prior experience and their peers for evaluation. Our study also revealed novel approaches to evaluation. We introduce a practice-led heuristic framework to aid practitioners to evaluate narrative visualization systematically. Our practice-led heuristic framework couples first-hand practitioner experience with recent research literature. This work sheds light on how to address narrative visualization evaluation to better inform both academic research and practice.

Keywords Narrative visualization · Visualization evaluation · Visualization design · Heuristic evaluation

1 Introduction

Narrative visualization is a subfield in visualization academic research and practice. Effective narrative visualization combines storytelling and data visualization to convey complex information in a comprehensible and compelling manner [1]. The proliferation of narrative visualization in news media attests to its increasing popularity. It has been employed in a number of contexts including explaining the effects of climate change or detailing the spread of COVID-19 [2, 3]. The important information contained in narrative visualization necessitates rigorous evaluation. Evaluation is key to ensuring that complex information portrayed in the narrative visualization is not misconstrued or deemed unengaging.

The evaluation of narrative visualization is fraught with challenges. Conventional approaches to visualization evaluation which measure task time and error rate are not sufficient in its evaluation. This is because they do not provide crucial insight into the user's comprehension or user experience [4, 5]. Qualitative methods, such as elicitation interviews and focus groups, have been used to gain deeper insights when evaluating narrative visualization [6, 7]. However, qualitative methods, such as the aforementioned examples, are potentially too costly and laborious in a practical setting. Furthermore, practice-led heuristics or evaluation guidelines, specific to narrative visualization, do not yet exist. Mystery surrounds the evaluation of narrative visualization in practice, where little is known about how these challenges are addressed outside of the laboratory.

In this work, our aim was dually to find out what current evaluation practices are, and how best can we support practitioners. To achieve our aim, we conducted a survey of 63 practitioners. These practitioners were recruited from an online forum named “The Data Visualization Society” [9]. It is an active practitioner forum that includes visualization practitioners from all genres of visualization development, including narrative visualization. Then, we conducted one-to-one interviews with a smaller group of 12 practitioners. The purpose of our interviews was to gain an understanding behind the reasoning and thought processes of practitioners

✉ Nina Errey
nina.z.errey@student.uts.edu.au

Jie Liang
jie.liang@uts.edu.au

Tuck Wah Leong
tuckwah.leong@uts.edu.au

Didar Zowghi
didar.zowghi@uts.edu.au

¹ School of Computer Science, University of Technology
Sydney, 15 Broadway, Ultimo, NSW 2007, Australia

when conducting evaluation. Our motivation for focusing on practitioners was that through their necessity and real-world experience, they have valuable insight into the practical evaluation of narrative visualization. This insight can be beneficial in solidifying foundational knowledge for both future narrative visualization practice and future academic research.

The results of our research indicated that there is often an ad hoc approach to narrative visualization evaluation. The lack of established guidelines or heuristics, mean that evaluation is rarely systematic. The most common method of evaluation is an informal group discussion. Practitioners relied on their colleagues and their past experience to inform their evaluation. We found that approximately half of practitioners employed end-users in their evaluation. Moreover, due to necessity, novel evaluation practices were employed such as guerrilla user-testing and social media. To aid practitioners in evaluation, we propose a preliminary set of heuristics specific to the evaluation of narrative visualization. These heuristics comprise of detailed usage advice derived from our interviews, coupled with research literature and survey results. To conclude, we make a series of recommendations about how our heuristics can be practically implemented, illustrated through the use of a case study. Our investigation findings and our preliminary set of heuristics are fundamental to establishing a community of practice for narrative visualization.

2 Related work

2.1 Narrative visualization evaluation

The importance of thorough and rigorous visualization evaluation is undisputed by the academic visualization community [5, 10, 11]. Conventionally, there are two methods to evaluate visualizations—inspection methods and end-user testing methods [11, 12]. Inspection methods are examined as heuristic evaluation. End-user testing methods include representative end-users and are usually defined as either quantitative or qualitative [13].

2.1.1 End-user testing methods

The field of human–computer interface (HCI) design retains a long-held certainty that end-users should play a fundamental role in the evaluation process [14]. Integrating end-users in both the formative stage and summative stage of a visualization development cycle has been shown to have beneficial outcomes [15]. Some general benefits include a greater understanding of user needs, and a broader perspective beyond what domain experts may offer [8]. Notwithstanding the recognized value in end-user testing, it does have weaknesses. Below we discuss the strengths and weaknesses

specific to qualitative and quantitative methods of end-user testing in relation to visualization research.

Qualitative methods of end-user evaluation lend themselves to measures particular to effective narrative visualization. This is because they provide a “richer understanding” of the comprehension and user experience of the work [11]. Notable examples include the “walk-through” or the “think aloud” protocol and interviews [4, 13]. One study used focus groups to evaluate narrative visualization because they “enable us to obtain qualitative and affective information from participants easily [7].” The drawbacks of qualitative end-user methods of evaluation are not just their costly, laborious nature; they are also difficult to replicate and difficult to quantifiably measure [11].

Quantitative user-data collection methods have been experimented with to evaluate storytelling in information visualization [16]. These include session times and click-through source [16]. These user-centered metrics are directly transferred from web analytic frameworks and use low-level user-activity traces as signals which are translated as user intentions [17]. They do not, however, provide crucial insight into user engagement and experience. Studies examining end-user interaction with visualization have employed eye-tracking [18]. This method of end-user evaluation is however notoriously cumbersome, and alternative solutions are advocated [19].

We have outlined here some end-user evaluation methods for visualization and their associated challenges. The next section will move on to heuristic evaluation, which should, in theory, complement end-user testing [8, 10].

2.1.2 Heuristic evaluation

Heuristic evaluation is a common inspection method of visualization evaluation [11, 13]. It is described as a vital part of the visualization practitioners’ toolkit [20]. Tory and Möller in their summary of expert reviews recommend the use of heuristic evaluation for analyzing visualization systems [10]. The recognition of the beneficial advantage of heuristic evaluation led several authors to propose sets of heuristics for visualization. These are the 13 heuristics by Zuk et al. [21] and the 10 heuristics by Forsell and Johansson [22]. The Forsell and Johansson et al. heuristic set is derived from a synthesis of existing heuristic sets, including the original usability heuristics described by Nielsen [8, 22]. Similarly, Zuk et al. presented a synthesis of pre-existing evaluation criteria, including the “information seeking” mantra first described by Shneiderman [21, 23]. While these heuristic studies are useful, they are related to system usability rather than the storytelling ability of the visualization.

Amini et al. outlined evaluation criteria for narrative visualization [24]. In their book chapter, each criterion for the evaluation of data-driven storytelling is described with ref-

erences to examples and academic research. The process for developing their evaluation criteria did not include direct consultation with practitioners. This differentiates our work, where practitioner feedback was fundamental to our study. Other examples of guidelines for data-driven storytelling have a focus on tools and code examples [25], or the overall design process [26].

A framework-based approach to visualization specific heuristics has been advocated [27]. This approach was compared to both “performance-based” and “process-based” approaches. These aforementioned approaches were not deemed suitable; instead, the framework approach was advocated as “the most generalizable and extensible of the three approaches” [27]. Similarly, Zuk et al. suggested “A hierarchical or taxonomic way of grouping may aid in selecting an appropriate set of heuristics [21].” Heuristic evaluation is part of an iterative design process where it can be a relatively fast evaluation methodology, and also has been shown to produce useful results when employed by non-experts [8]. Outside of the cost savings, other benefits include gaining deeper insight than end-user studies and lowering the “intimidation barrier” [8, 10]. Furthermore, heuristics serve as a structure to group evaluation which has been shown to be more effective than unstructured group evaluation [28]. Some disadvantages of heuristic evaluation have been listed as the lack of evaluation qualifications of the practitioner or difficulty to innovate while bounded by a set of rules [11].

2.1.3 Narrative visualization practice

It has been recognized that there are barriers to knowledge production and use between the academic and practitioner communities. In the VIS community, this recognition has led to events that cross this divide (e.g., VisAp [29] and Vis-InPractice [30]). These events are evidence that professional practice as an activity with its own methods and learning, is recognized as valuable to the furthering of visualization research. This is especially so for newer forms of visualization, such as narrative visualization, where an established community of practice does not yet exist.

This study aims to investigate whether a community of practice is appropriate for narrative visualization. This aim is founded on the notion that a community promotes discussion and encourages a willingness to share ideas [31]. The concept of a community of practice was first introduced by Lave and Wenger [32] and further expanded by Wenger in his book “Communities of Practice [31].” It is through the process of sharing information and experiences with the group, that members learn from each other, and have an opportunity to develop personally and professionally. In areas such as business and health care, communities of practice are integral in fostering and maintaining domain knowledge [33].

In this study, we motivate our approach as we believe that there is value in studying practitioners’ lived experience and tacit knowledge. In this context, tacit knowledge is a combination of intellectual knowledge, cognitive skill and manual skill. Tacit knowledge is described “as the ability to make decisions in the absence of written rules [34].” This form of knowledge is characterized by the implicit decisions made in the development process, which are, by definition, difficult to articulate.

Data visualization practice is multi-disciplinary, where an individual or a team are required to have skills in computer science, data analytics, design, project management, etc. It has been suggested that data visualization practice requires at least “eight hats” or rather eight different areas of knowledge [35]. Moreover, narrative visualization practitioners must be adept storytellers [36], where narrative structures and reader experiences must be considered along with functional usability considerations associated with data visualization. In this work, we attempt to garner the substantial accumulated knowledge of narrative visualization practitioners so we can better inform both practitioners and researchers.

3 Research methodology

To achieve our aim of investigating how narrative visualization is evaluated in practice, we performed an online survey of practitioners. To gain deeper insight we conducted a series of semi-structured interviews relying on our survey findings to prompt and aid practitioners to articulate their knowledge. Supplemental material can be accessed here; this includes a series of interactive charts and raw data <https://effectivenv.github.io/info/>.

3.1 Survey design

The survey consisted of both multiple-choice questions and open-ended questions. Our survey was designed to offer the option for the practitioner to add feedback to complement multiple-choice questions. The following two research questions guided our survey:

RQ1 - What is narrative visualization evaluated for?

RQ2 - How is narrative visualization evaluated?

Regarding RQ1 we asked practitioners what they deemed to be the three most important elements of effective narrative visualization. We offered the practitioner a list of elements to choose from and the option to add their own. This list of elements was derived from an analysis of research literature which empirically evaluated narrative visualization and various forms of visual or verbal storytelling. We performed a search of known databases including IEEEExplore, ACM, and Google Scholar. We carefully selected peer-reviewed papers

published in reputable venues. The inclusion criteria were as follows; (1) that the paper included an empirical experiment on either visualization, visual storytelling or verbal storytelling, (2) was either a journal paper or a conference paper, (3) was published before 2009. We removed all irrelevant papers and duplicates based on an analysis of the abstract and keywords. Our finalized list of elements referenced 17 papers. The list is not exhaustive, where the primary function of our list of elements was to serve as initial survey items and discussion points in our interviews. Selected papers are referenced in Table 3.

The next section, corresponding to RQ2, asked practitioners about their evaluation practices. These questions were split into inspection methods and end-user testing methods. The evaluation methods for both categories were derived from the research literature. At the end of the survey, practitioners could opt to take part in a future interview for further research. The survey was hosted on the Qualtrics platform. The format followed the survey design checklist outlined by Kitchenham and Pfleger, is widely accepted as establishing best practice for survey design [37].

3.2 Survey participants

We recruited practitioners from an online forum named “The Data Visualization Society [9].” At the time of writing this forum had approximately 13,000 members. We searched the “Introductions” channel to find practitioners of narrative visualization. We approached practitioners that had proven work experience in contributing to the development of narrative visualization. As a secondary process for vetting participants, we directly messaged each potential respondent via the forum, and if they did not believe they had contributed to the development of narrative visualization, then they either answered in the negative or did not respond to our inquiry.

3.3 Survey data analysis

The survey data collected were in most instances from multiple-choice questions, which could be quantitatively analyzed. We used thematic analysis to extract data from the open-ended feedback question. Thematic analysis was also applied to the questions where practitioners could input an “other” option. The process of thematic analysis is latent as it analyzes the data through underpinning concepts and assumptions which have been appropriated for narrative visualization practitioners [38]. For example, “responsive” is one theme that is taken from web development terminology referring to the ability for the visualization to render correctly on an unconventional browser. Two coders coded one open-ended feedback question independently and compared codes for inconsistencies. Once the codes were agreed upon, the author coded the other open-ended feedback questions.

3.4 Interview design

The survey results provided insight into the evaluation practices of narrative visualization practitioners. This is a useful foundation for understanding what evaluation methods are used, but it does not indicate how and why they are used. We found our interviewees through our survey, so therefore we could question the interviewee further on the reasoning behind their survey responses. We had 35 respondents that were willing to take part in an interview. We screened them to find those with a minimum of 1 year of experience, as well as having contributed to the development of at least 5 narrative visualizations.

Our interviews were semi-structured and had three distinct phases. See Table 1 for a description of interview questions and motivations.

3.5 Interview participants

We aimed to have an international participation base; however, most participants were from primarily English-speaking countries. Unintentionally half of the participants were from the domain of journalism. Out of those that were suitable and did not decline our invitation we recruited 12 interviewees.

Each participant agreed to the consent form provided, and their interview was recorded online using the Zoom platform. Interview transcripts were automatically generated using Microsoft Stream and then closely read to make sure errors, or misquotations were eliminated. Ethics approval was provided by our organization. See Table 2 for the self-reported characteristics of participants.

3.6 Interview data analysis

We pilot tested our interview and slightly re-worded multiple questions due to the feedback from our pilot interview. The timing of each interview was between 40 min to 1 h and was performed between June 2021 and January 2022. NVivo qualitative data analysis software was used to facilitate our coding and analysis.

The thematic analysis was a three-stage process. Three researchers independently coded one transcript and then compared codes. Once finalized, each subsequent interview transcript was similarly coded incorporating both data-driven inductive coding and a top-down priori approach. The result was 29 upper-level codes supported by 10 lower-level codes. After initial coding, the author reiterated over all transcripts to identify and collate themes. The final stage consisted of the author exporting NVivo codebooks and interview transcripts with NVivo’s “coding stripes” from selected interviews where themes were discussed and clarified with fellow researchers.

Table 1 Semistructured interview question logic informed by participant's survey responses

Phase	RQ	Aim	Questions
1	RQ1	To find out what comprises effective and ineffective narrative visualization	a) What is an effective narrative visualization? b) What is an ineffective narrative visualization?
2	RQ1	To illuminate and detail our heuristic framework with the consideration of audience for latter questions (f, g)	c) These are the characteristics you considered most important in the survey Why did you choose them? d) Do you think any characteristics are missing? e) Do you think any characteristics do not belong? f) Do you think the characteristics of effective narrative visualization would change if the work was for a different audience? g) Do you think some characteristics would stay the same regardless of the audience?
3	RQ2	To find out reasoning behind chosen evaluation method	h) This is the method(s) you said you use to evaluate your work with the expert review Why do you evaluate using that method? i) This is the method(s) you said you use to evaluate your work when end-user testing Why do you evaluate using that method? j) Hypothetically would you use a defined set of guidelines for design and evaluation?

Interview phase, research question (RQ), corresponding aim and corresponding questions

Table 2 Demographic information about each study participant, labelled by participant ID (PID)

PID	Domain	Exp	Num	Location
P1	Government	> 5	> 10	Australia
P2	Healthcare	> 5	> 10	USA
P3	Journalism	1–5	5–10	Switzerland
P4	Journalism	1–5	5–10	Finland
P5	Journalism	> 5	> 10	Australia
P6	City Planning	> 5	> 10	USA
P7	Journalism	1–5	> 10	UK
P8	Energy/Climate Change	1–5	5–10	Australia
P9	Journalism	1–5	5–10	Greece
P10	Journalism	1–5	> 10	Singapore
P11	Management Consulting	1–5	5–10	Canada
P12	Business	1–5	> 10	USA

Self-reported characteristics: domain, years of experience, number of narrative visualization developed and location

4 Results and findings

When reporting survey results n represents total number, m represents mean and SD represents standard deviation. We applied statistical analysis where appropriate.

4.1 Background questions

The first question we asked was how many years the practitioner had contributed to the development of narrative visualization. Most practitioners had worked for more than one year. With $n=4$ or 6% indicating they worked less than one year. $n=30$ practitioners or 48% indicated they worked from 1 to 5 years and $n=29$ or 46% indicated they had worked for more than five years.

The second question we asked was how many narrative visualizations had the practitioners contributed to the development of. Most practitioners, $n=41$ or 65% indicated they had developed more than 10 visualizations. $n=12$ or 19% had developed 5–10 visualizations. $n=10$ or 16% had developed 1–5 visualizations. These data suggest that most practitioners are quite experienced usually having multiple years of experience and developing more than five narrative visualizations ($n=52$ or 82%).

We asked which domain practitioners worked in. The three most dominant domains that became apparent from current research literature were: journalism, health, and education. We, therefore, gave four options to practitioners: journalism, health, education, and other. $n=13$ or 21% indicated they work in journalism. This was the most common single domain for narrative visualization practitioners to work in.

This was not entirely unforeseen, as many of the narrative visualizations that are publicly available are derived from news outlets. The second most common domain in which practitioners worked was education, with $n=11$ or 17% of practitioners. Finally, the health domain had $n=3$ or 5%.

Most practitioners identified as working in the other domain, or multiple domains $n=36$ or 57%. Some examples of the other domains include climate change, business research, retail, and paleontology. We observed that often the boundaries of a domain are not clearly delineated, where practitioners mentioned they work in overlapping domains, such as both health and education.

4.2 Inspection methods

4.2.1 Survey results

The majority of practitioners had a colleague or an external expert inspect their work before they released it, $n=57$ or 90%. When asked if a set of pre-defined criteria was used to inspect the visualization, $n=37$, answered “no” and $n=20$ answered “yes.” We asked practitioners regarding their informal inspection methods ($n=37$, $m=9$, $SD=7$). Group discussion received significantly more responses ($n=19$) compared to the other methods of informal evaluation. Group discussion accounts for 51% of the 37 practitioners that indicated they did not use pre-defined criteria to evaluate their work. This was followed by informal conversation ($n=11$). Informal email received the least amount of responses ($n=5$).

The two practitioners that indicated that they used other informal inspection methods, one mentioned “informal conversation,” while another mentioned the use of “Google Documents.” We asked practitioners what pre-defined criteria they used to evaluate their work. From the five practitioners that chose other as a criterion, 4 responses could be thematically grouped under the term “data accuracy.”

4.2.2 Interview analysis

Of the practitioners, we interviewed all except one had a colleague or external expert inspect their work. The one practitioner who did not have a colleague or external expert review their work explained that their work was often best described as a “hobby project” and therefore published outside of a professional environment.

Seven practitioners we interviewed had chosen “group discussion” as their informal inspection evaluation method. We asked them why and discovered that it was so that they could access the skills and past knowledge of their team. “We have a team of people who do data visualization daily for some years now and everyone has some experience of what works and what doesn’t work” P3. “You’ve got to rely on your past experience, through having expertise in the team and just

general sense gathering, hence the discussions” P5. From our survey and confirmed by our interviews we can conclude that usually practitioners use an informal group discussion when inspecting narrative visualization.

We asked practitioners why “data accuracy” was considered a criterion for evaluating effective narrative visualization. It was explained by P12 that it is an “ethical commitment,” where integrity is a key priority to practitioners. We have updated our heuristic framework to reflect this sentiment. The heuristic that corresponds to this feedback is named “data accuracy and honesty.”

4.3 End-user testing methods

4.3.1 Survey results

The responses to whether end-user testing was employed in narrative visualization evaluation, were almost equally split. $n=31$ or 49% responding with “yes” and $n=32$ or 51% responding with “no.” When practitioners indicated they did not evaluate with end-users, they were then asked why. Most practitioners, ($n=21$) said there was “no time.” The second most selected reason for no end-user testing was “no budget” ($n=14$). The third most selected reason for not end-user testing was given as “there is no expertise” ($n=8$). Finally, “it is unnecessary” was the least selected reason provided ($n=5$).

From the 31 practitioners that indicated they did employ end-user testing, we asked what methods did they use ($n=50$, $m=7$, $SD=6$). The “think aloud protocol/walk-through” method ($n=18$) was significantly higher than other methods. The second most popular method for end-user testing was “interviews” ($n=13$). Followed by “focus groups” ($n=8$). The only method for end-user testing that was significantly unlikely to be employed was “eye-tracking” ($n=2$). We found, however, all forms of quantitative end-user testing including; “session times,” “survey/questionnaire,” “click-tracking,” and “eye-tracking,” received a lower than average amount of responses.

4.3.2 Interview analysis

We found that practitioners that worked in journalism did not often employ end-user testing. From our interviews, however, we observed a trend where despite answering in the negative in our survey, actually end-user testing had been employed. This trend was noted by 4 interviewees P3, P5, P9, and P10. The difference is that they would end-user test “for a large project that was not time sensitive, so not in the news cycle” P5 or as another practitioner explained “not for a story as we need to have the budget” P9. We asked why the “think/aloud” walk-through method of end-user testing was preferred. Practitioners desire to measure the emotional response of users, rather than relying on quantitative data collection methods.

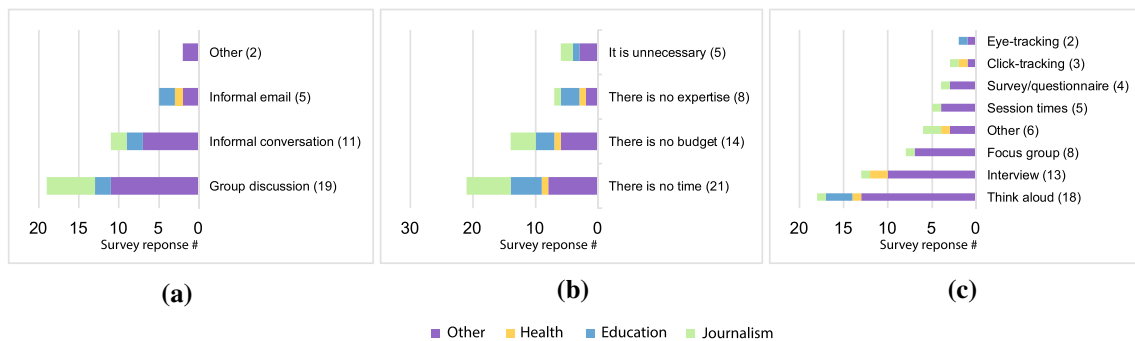


Fig. 1 Survey results. **a** Informal methods of evaluation using inspection methods. **b** Reasons for not employing end-user testing methods of evaluation. **c** End-user testing methods of evaluation

P1 explained that “it just captures the richness more than your digital collection method.”

Missing from our analysis of research literature yet popular with practitioners was the use of social media as an end-user evaluation tool. Particularly Twitter, where practitioners believed any issues would be picked up by followers and reported back. “Social media is a good feedback loop in terms of telling you what is really wrong” P8. One dilemma we observed was that the visualization project must be publicly available and therefore officially published to be accessible by social media. When asked further if they amend the project if it was criticized on social media “only if something is really wrong, then we will change it” P8.

A novel end-user evaluation method that was mentioned was guerilla testing. This form of end-user testing means randomly approaching members of the public that have no context to the project. Here the practitioner described watching closely, while users interacted with the project and asking them “why did you stop there?” or “why did you click there?” The reason for choosing this method of evaluation was to gain a deeper perspective which was described as providing a “really intense understanding of the user” P10.

5 Heuristic framework

From our survey we determined that inspection methods of evaluation are often employed; however, usually they are employed informally, without adhering to guidelines, or criteria. Our heuristic framework is aimed at enabling practitioners to identify gaps in narrative visualization development in a structured manner. Rather than presenting a minimal set of heuristics, we delve deeper into the actual usage and best practice of each heuristic. We corroborate or contradict research literature on narrative visualization evaluation where appropriate. As a response to the feedback from practitioners, we have categorized each heuristic into three high-level categories. These categories are: composition, reader experience, and credibility and trust. See Table 3

for an outline our heuristic framework. In this table each heuristic is allocated an upper-level category, title, description, referenced source material, and the amount of survey responses it received.

“Composition” was a suggested category name by multiple practitioners. This category encompasses the visual design aesthetic of narrative visualization. It also includes information distribution and overall layout. “Reader experience” is a category that stems from recent research in “user experience focused” evaluation [50]. Rather than user experience, when developing narrative visualization, considerations must be observed that apply specifically to the narrative aspects of narrative visualization. “Credibility and trust” is a separate category because it has a unique and vital role in narrative visualization effectiveness. There is overlap between our categories and usage of each heuristic. The motivation for the delineation of each item will become clearer in the “Implementation of framework” section this paper.

In our survey, we asked respondents to nominate their 3 most important elements with 12 options including one “other” option. The results of this question served as discussion points in our interviews with practitioners and influenced our final heuristic framework. Our chosen terminology is primarily for the benefit of practitioners. We found academic terms such as “heuristics” are not widely understood by practitioners, and therefore, we were required to modulate our vocabulary to suit practitioners. When reporting survey results n represents total number, m represents mean, and SD represents standard deviation. The responses for this question are thus described $n=189$, $m=15$, and $SD=9$.

5.1 Composition

Logical layout to not distract the reader

In our survey $n=33$ chose this as an important heuristic for effective narrative visualization. This is significantly higher than other heuristics. This heuristic refers to the arrangement and organization of graphical and textual elements in a design. An oft-mentioned key measure of effectiveness

Table 3 The proposed practice-led, heuristic framework for narrative visualization evaluation

Category	Heuristic and description	Source literature	Survey #
<i>Composition</i> The visual aesthetic, information distribution, and overall layout	<i>Logical layout</i> A layout refers to the arrangement and organization of graphical and textual elements in a design. A logical layout is one that reduces the reader's cognitive workload	McKenna et al. [39] Brehmer et al. [40]	33
	<i>Information density</i> When too much information is presented it is termed "information dense." Overly information dense narrative visualization is difficult to comprehend, particularly for less literate audiences	Majooni et al. [18] Damman et al. [41]	11
	<i>Mindful use of color</i> Mindful use of color refers to the affective response of color palettes	Bartram et al. [42]	12
	<i>Textual integration</i> Refers to the important influence text has on the reader interpretation, specifically titles	Kong et al. [43]	8
	<i>Cohesiveness</i> Cohesiveness refers to the overall coherency of the visualization by maintaining context from one data point to the next. This has been shown to facilitate understanding	Hullman et al. [44] Brehmer et al. [40]	18
	<i>Retain interest</i> Originally from a study in the news industry that placed importance on the session times of readers. Alternatively, it refers to a deeper level of interest or "enchantment" specifically in regard to storytelling	George-Palinos et al. [45] Sturm [46]	29
<i>Reader Experience</i> How a reader interacts with and experiences narrative visualization	<i>Interactivity</i> Interactive visualization can facilitate data retention with users	McKenna et al. [39] Moritz et al. [47]	8
	<i>Personally relatable content</i> When the reader can personally relate to the content provided it has been shown to be more memorable and impactful	Peck et al. [48] Figueiras [7]	13
	<i>Easily recognizable content</i> When imagery is recognizable, it has been shown that memorability is improved	Borkin et al. [49]	18
	<i>Data source identified</i> When a data source is identified, it increases credibility and trust	Peck et al. [48]	18
<i>Credibility and Trust</i> Data quality and ability for the narrative visualization to be believed	<i>Data accuracy and honesty</i> Refers to the ethical representation of data	Not originally derived from literature	-

Upper-level categories, heuristic and description, source literature and number of survey responses

in visualization research, a logical layout can reduce the cognitive workload of the user [18, 39, 40]. The primary aim when regarding this heuristic was to minimize “visually jarring” user experiences. For example, P10 asks while evaluating narrative visualization “is it visually jarring? Does it fit nicely?” This finding was echoed in a paper by Brehmer et al. which studied timelines in visualization [40]. Smooth timeline transitions that did not distract the user by being “visually jarring” were shown to increase effectiveness.

Furthermore, narrative visualization can induce a state of flow in the reader [39]. The layout of a narrative visualization is enhanced through implementing a series of “flow factors [39].” Some of these “flow factors” were echoed in our interviews, where practitioners aimed to arrange elements so that reader attention was not lost due to functionality. Fundamentally, “if the reader has to think about anything else other than the story, you will lose him. If he has to think, why should I press this button or not? Should I scroll or not? Should I scroll horizontally or vertically? Then we are losing him” P9.

Information density to guide the reader into the complexity

This heuristic was derived from empirical evidence which indicated overly information dense visualization was difficult to comprehend, particularly for less literate audiences [41]. To overcome this challenge, practitioners explain that information-dense or complex data should not be avoided; rather, it must be introduced progressively. Practically, this means supplying an “entry-point” to the data or as described by P6 “on-boarding the reader.” This process was described by P1 as “slowly guiding the audience to more detailed information.” P7 explains “we do the complex chart at the end of the article because we want the reader to explore it and to spend time on it.”

Practitioners explained through introducing and guiding the user, complex information becomes accessible. P3 reasoned that by using a single axis as an entry-point to the data the user is “really gently brought into the whole thing.” Another example of gradually increasing the complexity of information is by textually explaining the context before introducing an information-dense visualization.

Mindful use of color regarding cultural and emotional connotations

Mindful use of color refers not only aesthetically to visualization, but also color indicators to aid user comprehension. When used deliberately color association can be a powerful tool. It has been shown that colors have an affective response depending on the chosen palette [42]. This corresponds to psychology and color theory which can we relate certain affective impressions from a color palette.

Color is described by practitioners as being the “most powerful design tool we have. It’s also probably the hardest to do well” P2. The reasoning behind this is the cultural and emotional connotations that colors have within a broader societal context. “It’s very contextual and when I say contextual I

don’t mean in the sense of what kind of article you’re writing but in the sense of what culture you belong to” P4. “Color is the most challenging thing in any visualization. If you choose the wrong color, you can skew the story or the meaning of the story, or even worse, you manipulate the readers’ emotion” P9. For example, one practitioner mentioned that the default color palette from a development tool named Flourish [51], caused problems. This was because the subject matter was taboo and the default color palette could “re-enforce color associations” P10.

Textual integration for usability and intrigue

Evidence suggests that text can be equally as influential to reader outcomes as graphical representations [43]. A descriptive title is a vital component to an effective narrative visualization [43]. P1 described the thought process when devising a suitable title, “whenever we design a visualization, we’re trying to make the heading descriptive and conversational. We try to make it concise as opposed to having a title that’s full of technical jargon. Then we make sure people can read the title and can walk away knowing what the visualization is about without even looking at the visualization.” The title functions as a quick access to the data visualized, where it can “give them a brief overview of what they’re seeing without delving into the real detail of it” P5.

Beyond the title, the text integrated into a narrative visualization, provides a rhetorical frame of reference for the reader. In a study on narrative visualization and rhetorical framing, the authors asked if the practitioner intentionally incorporated “the power of rhetorical techniques” [52]. Our interviews suggest that rhetorical techniques are employed deliberately. P9 mentioned they should “allow the visualization function as a written story” or as explained by P5, the text must “create a sense of intrigue.” These comments highlight the role of the narrative visualization practitioner as a storyteller.

5.2 Reader experience

Cohesiveness to maintain context and focus

Cohesiveness refers to the overall coherency of the narrative visualization by maintaining context from one data point to the next. The structuring of the data into a coherent story sequence has been shown to improve the effectiveness [44]. The goal of maintaining coherency was affirmed by practitioners as one of great importance. P8 exemplified this goal, explaining “every data point, every piece of information put in there should be related to the issue or problem you’re talking about.” Another practitioner explained the challenge of maintaining cohesiveness, “I find sometimes it can be hard to maintain a clear thread throughout the different sections of a long visualization” P6.

A common topic when discussing cohesiveness, was that an ineffective narrative visualization lacked focus. One

example was when a practitioner recounted their first narrative visualization development experience. The scope of the piece of work continued to grow and therefore became unmanageable. “I wanted to visualize all the data so I just dropped 1000 visualizations to the readers without explaining why” P9.

As was outlined by a study on the sequence in narrative visualization, they often consist of a series of screens or episodes, rather than one stand-alone visualization [44]. The practitioner has the task of splitting the information into conceptually separate episodes ordered in a sequence. This process is academically referred to as “chunking” [53]. Initially “data chunking” was a separate heuristic; however, due to practitioner feedback, it was deemed unnecessary. It seems “chunking” is not widely understood outside of academia. For example, “I don’t know what data chunking means” P11.

Retain interest *but first, gain their interest*

In our survey this heuristic gained a significant amount of responses ($n=29$). When we asked practitioners about this heuristic, they highlighted that it was important to first gain the attention of the reader before retaining it. As illustrated by P5 “you need to open with something that is intriguing and that makes people want to read” or as described by P11 the reader scans the article, and that’s where you want to catch them.”

Another practitioner mentioned how reading a narrative visualization requires intellectual investment. P3 explained “when people put in the effort they need to know they will get something out of it.” The concept that the reader must “know they will get something out of it” was outlined in an example. P3 showed a narrative visualization that opened with an empty graph. P3 then explained that, particularly from a “news perspective” it is important to clearly indicate to readers why they should continue or they “can always jump away.” Presenting readers with an empty graph is therefore not recommended.

We found that a substantial amount of practitioners that work in journalism selected this heuristic in the survey ($n=9$). The reasoning for this is that in news media, often the value of a narrative visualization is judged on the session times or the “page views” that the visualization attracts [45].

On a more profound level, however, retain reader interest could be compared to a study on verbal storytelling. This study found listening to a story can induce a state of “enchantment” which consequently, increased story memorability [46]. This is similar to narrative visualization where readers can be induced into a state of “flow” [39]. In the aforementioned study, this was achieved through content progressively revealing itself as the reader scrolls vertically. In practice, this form of narrative visualization is termed “scrolly telling.” A combination of the words scrolling and telling, “scrolly telling” is commonly featured on news media websites [54].

The use of “scrolly telling” is a double-edged sword. While it can induce undivided attention in the reader, it can be overused. Practitioners pointed out that the reader lose interest when having to scroll too much. “You don’t want them to stop halfway through visualization, because they don’t care anymore” P18. “Scrolly telling” was described as an “overly structured experience” where it “constrains the user experience and forces them through a long sort of linear process” P11.

Interactivity *only when the reader desires to drill down*

It has been shown that interactive visualizations can facilitate data retention in readers [47]. Therefore, if a reader desires to learn more, the option should be available, however without detracting from the remainder of the narrative visualization. For an expert audience, interactivity was advocated. “You add more interactivity for people who are experts to drill down so they can understand how it works” P4.

A further benefit of interactivity is that it can promote learning by fostering a sense of enjoyment and curiosity [47]. “Some people want to click around and the people who want to click around aren’t going to perceive that as work” P2. “Interactivity keeps the excitement or the curiosity going and can promote it longer” P1.

Conversely, interactivity was described by practitioners as not necessary for the effectiveness of narrative visualization. Multiple practitioners described interactivity as “bells and whistles” (P5 and P12). For example “I think a lot of times people throw on all the bells and whistles and making things move is entertaining. But, if it doesn’t tell the story and get the point across, it can be distracting” P5. This was especially so for those working in journalism that note the visualization has to be “skimmable.” This means that the reader can scan quickly over the content without feeling they have missed information. Because you can simply have any static news article without any interactivity, and it still can be very informative and interesting” P4.

Personally relatable content *to reach the reader*

Empirical studies suggest that personally, relatable content is the primary driver for gaining and retaining reader attention [7, 48]. As explained by P1 “if the designer couldn’t make it relatable, the piece doesn’t have a chance to reach the intended audience.” Further illustrated by P12 “because if you’re not living in those spreadsheets and understanding what all that data means, it takes a long time to figure it out.” The dilemma for the narrative visualization practitioner is to engage the reader by appealing to their frames of understanding and reference.

An example given by one practitioner was when they needed to compare international air quality data. Air quality data are usually related in parts-per-million (PPM). This metric is used to describe particle saturation and is not easily understood outside of the scientific community. The practitioner explained their ideation and design process, “I found

this research that said if you are exposed to a certain amount PPM, that means you are smoking this many cigarettes per day. I can therefore show you that your air quality is equal to smoking about 10 cigarettes per day for one year” P4.

Easily recognizable content *for universality*

Visualization effectiveness is increased when it features imagery that relates to every-day objects [49]. The reader not only prefers content that they can recognize; they also remember it more easily [49]. Some examples of recognizable content are photographs and cartoons. “It’s very important to use real world conventions” P12. By presenting the information in visual metaphors the reader can mentally map concepts and analogies [55]. These can bridge cultural divides and can aid in universal comprehension. Easily recognizable content “means that people can recognize the symbols easily so it will be a universal story, no matter if the reader is in Greece or in Australia” P9.

5.3 Credibility and trust

Data source identified *it’s non-negotiable*

It has been shown that identifying the source of visualized data leads to greater trust in a narrative visualization [56]. Practitioners generally agreed that every single piece of work required a reference to its original data source. “It’s non-negotiable” P2. Furthermore, “I wouldn’t even necessarily say that that’s good narrative visualization, that’s just good journalism” P5.

In addition to identifying data source, the methodology of how the data was manipulated and visualized should be included. The reasoning for including the methodology was that it may not be inherently clear how the narrative was constructed by simply linking to the data source. As summed up by one practitioner “if you don’t believe me here is the data, you go analyze it” P11. Peck et al conducted an empirical study where part of the study included revealing the data source of a visualization. They found identifying the data source of a visualization increased credibility with some audiences [48].

Data accuracy and honesty *to avoid misinformation*

Due to their reoccurrence as a theme in our survey and interviews, data accuracy and honesty became necessary to include as a heuristic. Data accuracy was considered a best practice by practitioners. “People need to trust what they’re looking at” P8. Cairo pointed out often compelling narrative visualization may look precise but is not necessarily accurate [57]. The challenge is then for the practitioner to represent the data without manipulating to the extent it appears no longer accurate. As explained by P12 “try to show what the data is showing.”

Conversely, data honesty means representing the data so that it can be interpreted correctly, therefore data accuracy is not the aim, and indeed it can detract from effectiveness

[43]. Acting ethically means not only acting honestly and virtuously but also considering and minimizing the potential errors of interpretation. One apt example was given by P10 who described an often misunderstood Covid-19 case chart. “It’s actually quite misleading because it was a log scale. It had all of the countries in a single screen when actually between countries the disparity in the numbers is tremendous. You’re in USA is talking about hundreds of thousands every day and then Singapore you have two or three a day. There’s no way you could scale that proportion. So they use a log scale to visualize everything together. It is a necessary distortion, but still, a distortion and some people looking at that, feel that it is misleading” P10.

5.4 Superfluous heuristic elements

Some elements that were on our original list do not appear in our heuristic framework. These elements are; “data chunking” and “findability of the visualization on the internet.” The rationale behind removing “data chunking” is explained in Sect. 5.2. “Findability of the visualization on the internet” was the least popular heuristic in our survey of practitioners and received a significantly small amount of responses ($n=4$). This heuristic element originated from an empirical study that found that narrative visualizations were often overlooked due to poor internet findability [58].

We asked practitioners why they thought that this heuristic was not popular. P4, P8, P5, P11, and P2 all indicated that the cause was due to audience considerations. An effective narrative visualization does not necessarily have a large audience. For example “if you’re talking about a visualization that’s being developed specifically for some really small niche group, then it doesn’t matter” P5. Another reason was that search engine optimization is considered not part of the development process for narrative visualization. As suggested by P9 “this is more a marketing part so maybe I would take that out.”

6 Case study: a day in the life of women and men

In this section we present a case study that illustrates how our heuristic framework is reflected in an effective narrative visualization. We focus on an example that was deemed effective by two interviewees, P1 and P11. “A Day in the Life of Women and Men” authored by Nathan Yau, is based on data from the “American Time Use Survey” [59]. It simulates a working day for men and women. Each dot represents a person where cyan represents women and orange represents men. They move according to their location at various times of the day. The clock-face ticks over automatically and the reader can choose to pause, increase, or slow down the

movement of dots. This particular narrative visualization was mentioned in two interviews as effective narrative visualization. As described by P1, the “choice of using movement to represent real world activities is perfect.” Considering the exemplary nature of this narrative visualization we deem it suitable as a case study. Categorically, we relate each heuristic to the narrative visualization to demonstrate how our heuristic framework could be applied. See Figure 3 for a screenshot.

Firstly, when considering, “Logical layout” we can observe that the information is presented without any “visually jarring” elements. In regard to “Information density,” we can see that the author has not presented all the information immediately, rather through the passing of time, specifically minutes, the information is categorically revealed. Interestingly, “Mindful use of color” is shown through the fact that the author has used neutral colors for each gender. Stereotypical colors such as pink for women and blue for men are avoided. “Textual content” is apparent in the title, which is both clear and impactful. Further evidence of careful textual content integration, is the short conversational paragraph that follows the visualization. Here the author outlines what they found surprising from the data. This paragraph serves as a conversation with the reader, asking them if they found this surprising too.

“Cohesiveness” is illustrated by the fact that if we go to the original data source there are more than 100 different activities listed in the “American Time Use Survey.” The author has simplified this list to 9 over-arching activities. This decision is an example of how the author has maintained a focused, cohesive message. The “Retain interest” heuristic can be initially observed in the attention grabbing title. Then as the visualization unfolds, attention is retained through the use of movement illustrating the difference between genders and their correlating time-use. We must also note that scrolling is kept to a minimum and not required to read the narrative visualization. “Interactivity” is used as suggested by our interviewees. Instead of being necessary for the comprehension of the narrative visualization, it provides a way for the reader to explore the data. By pausing the data the reader can stop and examine it, at their own pace. By making the dots move faster or slower, it gives the reader the opportunity to optimize their experience. “Personally relatable content” is observed by the fact that everyone has a frame of reference when it comes to how they spend their day. The reader asks “how do I spend my time?” We all must sleep, we all must eat and drink etc. “Easily recognizable content” is observable, specifically in the correlation between the clock-face and the movement of dots. As described by P1, a clock-face is a universally understood motif, “everyone including children know how to read time.” Finally, both heuristics in the category of “Credibility and trust” are addressed. The author

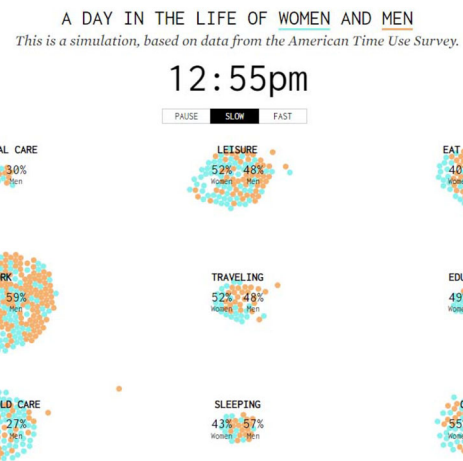


Fig. 2 Screenshot of a narrative visualization titled “A Day in the Life of Women and Men” by Nathan Yau [59]

has not only given the source of the data and the method to process it, but also tutorials and tools used.

As was

7 Implementation of framework

To achieve our goal of aiding practitioners to evaluate their narrative visualization, we outline how it can be implemented in a professional setting. In our survey of practitioners, we found that 90% of practitioners had a colleague or external expert review their work. Informal group discussion was significantly the most employed inspection method of evaluation. Group discussion and the resulting group decisions are usually superior to individual decisions. This is due to the assembly effect, where the decision is qualitatively and quantitatively superior to individual judgement [60]. The primary pitfall of group discussion is that it draws solely on the experiences of the particular development team. When we proposed a hypothetical set of guidelines, or heuristics to practitioners they were universally accepted. Practitioners of narrative visualization recognized the value in knowledge sharing and listed some of the benefits of a community of practice. For example, it could “accelerate a lot of decisions” P10 or they could “make the whole process repeatable, reproducible” P1. Santos et al. extend this point, where heuristics can facilitate the replication of skills and a “common ground for the comparison among works [20].”

The question is then how would this framework be systematically implemented when an informal approach was preferred. We propose that our heuristics are not implemented as a phased evaluation method as described by Nielsen [8]. Rather, our heuristic framework should be implemented as described by Carpendale, as a foundational checklist. The

practitioner can keep the heuristics in mind during the development process at various phases of the project [11].

Our upper-level categories aim to make the heuristic framework easier to keep in mind. We found that most practitioners were pressed for time and upper-level categories act as over-arching themes that require consideration in project development. With our framework in mind, it can be applied, for example, in the initial conception phase. Our framework can guide practitioners as an internal checklist. Indeed, practitioners mentioned they used their own internal checklist. The process of which is outlined as, “I go through to make sure that I address different aspects” P3. Furthermore, rather than a free-flowing group discussion, a structured discussion that iterates through each heuristic could streamline the process.

8 Discussion and future work

Our investigation of the evaluation methods of narrative visualization practitioners found that rarely systematic methods are used in evaluation. The reason for a lack of a systematic process was that practitioners preferred to rely on their team and their past experience to inform their evaluation. We believe this finding suggests practitioners have an understanding of evaluation that is particular to their circumstances. Where rather than summative or formative, evaluation is explorative [61]. Explorative evaluation helps the evaluator grasp new concepts and ideas, which are useful. This is different from formative evaluation which focuses on improving the design of a particular visualization, or summative evaluation, which seeks a “seal of approval” once completed [61]. Our study evidences that practitioners desire to learn and share their knowledge where the establishment of a community of practice would be an appropriate mechanism in this context. We suggest that future academic research in the area of narrative visualization evaluation leverages this finding. Rather than prescribing processes, or even tools, future academic research should include practitioners in their research methodologies. This approach would have the additional benefit of strengthening ties between both communities.

In this work we have presented a practice-led heuristic framework, informed by practitioners and principally for the use of practitioners. This is different from similar evaluation criteria, where practitioners were not the primary influence in their formation [24]. A key benefit of our approach is that it theoretically should be apposite to the purpose of informing practitioners, as it directly reflects the advice of practitioners themselves. The heuristic framework was influenced and amended according to the input of practitioners through an

iterative development process. As evidenced by the fact that two original heuristic elements proved to be superfluous according to our survey and interviews. These elements, therefore, did not appear in our final heuristic framework. Similarly, we added a heuristic because it was a common theme that appeared in both our survey and interviews. Our finalized heuristic framework is therefore directly representative of practitioner feedback.

In our study we found approximately half of practitioners employ end-users in their evaluation. This is at odds with the recommendation that end-users should be included in the visualization development cycle [15]. The reason that twenty-one practitioners gave for not employing end-users was that there was no time. We believe a streamlined, structured approach might make end-user testing less time-consuming. One noteworthy observation, however, was a move toward greater end-user evaluation. Multiple practitioners had employed end-user testing, albeit in larger, more complex projects. These practitioners had initially indicated they did not employ end-user testing in our survey, however, had since changed their methods when asked in an interview. The trend toward greater employment of end-user testing is encouraging. Our ambition is that through this research, future evaluation methods, such as end-user evaluation, can benefit from a better informed approach.

A practice-led approach generates new perspectives and can lead to interesting avenues of research. Based on this work, some examples of interesting avenues of research are as follows. Firstly, the novel evaluation method of guerilla testing could be an interesting avenue of research. Traditional end-user testing is at odds with this direct and spontaneous approach; therefore, it might prove to be a genuine solution to the lack of realism often observed in laboratory end-user testing. Another interesting area could be the use of social media feedback as an evaluation tool, specifically Twitter. Understanding the role of social media and similar approaches can help us better understand the purpose and role of evaluation in a practical setting. Ultimately this will lead to the development of methods that are more applicable in these settings.

The case study we presented highlighted how our heuristic framework is apparent in effective narrative visualization. More research is required into if our heuristic framework, implemented according to our recommendations, does in fact result in effective narrative visualization. Our future research will validate our heuristic framework through a validation experiment. Our aim will be to find out whether our framework is useful and usable to practitioners of narrative visualization. We acknowledge that many heuristics are not validated and it is an important step in heuristic establishment [62].

9 Limitations

As with most surveys, to control for sampling bias can be challenging. We targeted practitioners who personally introduced themselves as being experienced in developing narrative visualization. As was the case, their introductions were brief and can lead to misinterpretation. When directly contacting a practitioner, if they believe that they were not suitable, then they would either not reply or reply in the negative. This process sifted out most unsuitable practitioners.

The practitioners were primarily from one online community forum. The community forum “The Data Visualization Society” [9] does not include the entire narrative visualization practitioner community. “The Data Visualization Society” [9] does however have substantial user base and is as close as possible to being representative of the practitioner community. It has a large portion of academic or students of visualization. We appealed to participants that did not have an obvious academic background. There is however overlap between the two communities.

A substantial portion of practitioners we interviewed, identified as working in the journalism domain. This has skewed our results to favor the perspective of those working in that particular domain. Further investigation is needed into whether a different practitioner domain ratio would impact the final heuristic framework.

It has been recognized that visualization practitioners are rarely familiar with the term heuristics [63]. Our survey and interview used the term “pre-defined criteria” or “guidelines” as these terms are a more appropriate terminology for our study participants. We aim to introduce the term heuristic to the vernacular of practitioners as its precise meaning of a cognitive short-cut best describes our evaluation framework [64].

10 Conclusion

Our study of narrative visualization evaluation practice found usually practitioners employ an ad hoc evaluation approach. Their preferred method of evaluation was an informal group discussion. Practitioners relied on the accumulative knowledge and experience of their team to evaluate narrative visualization. At times, novel approaches to evaluation were adopted such as the use of social media or guerilla user-testing. To aid practitioners to streamline their evaluation, and encourage a “common ground for the comparison among works [20],” we propose a preliminary, practice-led heuristic framework so that inspection evaluation methods can be streamlined. Through coupling real-world practices with academic research, we introduce the foundation for a community of practice for narrative visualization.

Acknowledgements We thank all practitioners that participated in our survey and interviews for their time, participation, and feedback. We also would like to thank the administrator of “The Data Visualization Society” for allowing us to post our survey to the forum. We thank Nathan Yau for allowing us to feature his work.

Author Contributions All authors contributed to the study conception and design. The first draft of the manuscript was written by Nina Errey, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding No funding was received for conducting this study.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

1. Page, N., Gershon, W.: What storytelling can do for information visualization. *Commun. ACM* **44**(8), 31–37 (2001). <https://doi.org/10.1145/381641.381653>
2. NASA, A Degree of Concern. <https://climate.nasa.gov/news/2878/a-degree-of-concern-why-global-temperatures-matter/> (2019). Accessed 27 February 2023
3. Stevens, H.: Why outbreaks like coronavirus spread exponentially, and how to “flatten the curve”, *Washington Post*, <https://www.washingtonpost.com/graphics/2020/world/corona-simulator/> (2020). Accessed 27 February 2023
4. Saket, B., Endert, A., Stasko, J.: Beyond usability and performance: a review of user experience-focused evaluations in Visualization, in *BELIV*. Baltimore (2016). <https://doi.org/10.1145/2993901.2993903>
5. North, C.: Toward measuring visualization insight. *IEEE Comp. Graph. Appl.* **26**(3), 6–9 (2006). <https://doi.org/10.1109/MCG.2006.70>
6. Nowak, S., Bartram, L., Schiphorst, T.: A micro-phenomenological lens for evaluating narrative visualization, in *BELIV*. Berlin (2018). <https://doi.org/10.1109/BELIV.2018.8634072>
7. Figueiras, A.: How to tell stories using visualization, in *International conference on information visualisation*, Lisborn, (2014). <https://doi.org/10.1109/IV.2014.78>
8. Nielsen, J.: Heuristic Evaluation. In: Nielsen, J., Mack, R. (eds.) *Usability inspection methods*, pp. 25–63. Wiley, New York (1994)
9. Data Visualization Society: <https://www.datavisualizationsociety.org> (2023). Accessed 27 February 2023
10. Tory, M., Moller, T.: Evaluating visualizations: do expert reviews work? *IEEE Comp. Graph. Appl.* **25**(5), 8–11 (2005). <https://doi.org/10.1109/MCG.2005.102>
11. Carpendale, S.: Evaluating information visualizations, in *Information visualization lecture notes in computer science*, vol. Lecture, pp. 19–45. Notes in computer science, Springer, Berlin (2008)
12. Zhu, Y.: Measuring effective data visualization, in *Advances in visual computing*, Lake Tahoe (2007)
13. Andrews, K.: Evaluating information visualisations, in *BELIV*, Venice (2006). <https://dl.acm.org/doi/10.1145/1168149.1168151>
14. Norman, D., Draper, S.: *User-centered system design: new perspectives on human-computer interaction*. L. Erlbaum Associates Inc., Hillsdale (1986)
15. Hix, D., Swan, J., Gabbard, J., McGee, M., Durbin, J., King, T.: *User-centered design and evaluation of a real-time battlefield visu-*

- alization virtual environment. *Proceed. IEEE Virtual Real.* (1999). <https://doi.org/10.1109/VR.1999.756939>
16. Boy, J., Detieene, F., Fekete, J.: Storytelling in information visualizations: does it engage users to explore data?, in *Conference on human factors in computing systems*, Seoul (2015). <https://doi.org/10.1145/2702123.2702452>
 17. Gotz, D., Wen, Z.: Behavior-driven visualization recommendation, in *International conference on intelligent user interfaces*, Sanibel Island (2009). <https://doi.org/10.1145/1502650.1502695>
 18. Majooni, A., Masood, M., Akhavan, A.: An eye-tracking study on the effect of infographic structures on viewer's comprehension and cognitive load. *Inf. Visual.* **17**(3), 257–266 (2018). <https://doi.org/10.1177/1473871617701971>
 19. Kim, N., Bylinskii, Z., Borkin, M., Oliva, A., Gajos, K., Pfister, H.: A crowdsourced alternative to eye-tracking for visualization understanding, in *Conference on human factors in computing systems*, Seoul (2015). <https://doi.org/10.1145/2702613.2732934>
 20. Santos, B., Ferreira, B., Dias, P.: Heuristic evaluation in information visualization using three sets of heuristics: An exploratory study. *Human-Comp. Inter.* **67**(6), 259–270 (2001)
 21. Zuk, T., Schlesier, L., Neumann, P., Hancock, M., Carpendale, S.: Heuristics for information visualization evaluation, in *BELIV*. New York (2006). <https://doi.org/10.1145/1168149.1168162>
 22. Forsell, C., Johansson, J.: An heuristic set for evaluation in information visualization, in *The International conference on advanced visual interface*, Rome (2010). <https://doi.org/10.1145/1842993.1843029>
 23. Shneiderman, B.: Eyes have it: a task by data type taxonomy for information visualizations, in *IEEE Symposium on visual languages*, Boulder. Colorado (1996). <https://doi.org/10.1109/VL.1996.545307>
 24. Amini, F., Brehmer, M., Boulduan, G., Elmer, C., Widerkehr, B.: Evaluating data-driven stories and storytelling tools. In: Henry Riche, N., Hurter, C., Diakopoulos, N., Carpendale, S. (eds.) *Data-driven storytelling*, pp. 249–286. Taylor & Francis, Boca Raton (2018)
 25. Yau, N.: *Visualize this: the FlowingData guide to design, visualization, and statistics*. Wiley, Hoboken (2015)
 26. Nussbaumer Knaflic, C.: *Storytelling with data: a data visualization guide for business professionals*. Wiley, Hoboken (2015)
 27. Tarrell, A., Forsell, C., Fruhling, A., Grinstein, G., Borgo, R., Scholtz, J.: Toward visualization-specific heuristic evaluation, in *BELIV*. Paris (2014). <https://doi.org/10.1145/2669557.2669580>
 28. Hirokawa, R.: Consensus group decision-making, quality of decision, and group satisfaction: an attempt to sort “fact” from “fiction”. *Commun. Stud.* **33**(2), 407–415 (2009). <https://doi.org/10.1080/10510978209388445>
 29. VisAp. <https://visap.net/2022> (2022). Accessed 27 February 2023
 30. VisInPractice. <https://ieevis.org/year/2022/info/visinpractice> (2022). Accessed 27 February 2023
 31. Wenger, E.: *Communities of practice: learning, meaning, and identity*. Cambridge University Press, Cambridge (1999)
 32. Lave, J., Wenger, E.: *Situated learning: legitimate peripheral participation*. Cambridge University Press (1991). <https://doi.org/10.1017/CBO9780511815355>
 33. Li, L., Grimshaw, J., Nielsen, C., Judd, M., Coyte, P., Graham, I.: Use of communities of practice in business and health care sectors: a systematic review. *Implement. Sci.* (2009). <https://doi.org/10.1186/1748-5908-4-27>
 34. Magnus, J., Morgan, M.: Lessons from the tacit knowledge experiment, in *Methodology and tacit knowledge*, pp. 375–381. Wiley, Chichester (1999)
 35. Kirk, A.: *Data visualization: a successful design process*. Packt Publishing Limited, Olton (2012)
 36. Segel, E., Heer, J.: Narrative visualization: telling stories with data. *IEEE Trans. Visualiz. Comp. Graph.* **16**(6), 1139–1148 (2010). <https://doi.org/10.1109/TVCG.2010.179>
 37. Kitchenham, B., Pfleeger, S.: Principles of survey research in Part 3: constructing a survey instrument. *ACM Special Inter. Group Softw. Eng.* **27**(2), 20–24 (2002). <https://doi.org/10.1145/511152.511155>
 38. Braun, V., Clarke, V.: *Qualitat. Res. Psychol. Using thematic analysis in psychology* **3**(2), 77–101 (2006). <https://doi.org/10.1191/1478088706qp063oa>
 39. McKenna, S., Henry Riche, N., Lee, B., Boy, J., Meyer, M.: Visual narrative flow: Exploring factors shaping data visualization story reading experiences, in *Eurographics conference on visualization*, Barcelona (2017). <https://doi.org/10.1111/cgf.13195>
 40. Brehmer, M., Lee, B., Bach, B., Riche, N., Munzner, T.: Timelines revisited: a design space and considerations for expressive storytelling. *IEEE Trans. Visualiz. Comp. Graph.* **23**(9), 2151–2164 (2017). <https://doi.org/10.1109/TVCG.2016.2614803>
 41. Damman, O., Vonk, S., Haak, M., Hooijdonk, C., Timmermans, D.: The effects of infographics and several quantitative versus qualitative formats for cardiovascular disease risk, including heart age, on people's risk understanding. *Patient Educat. Counsel.* **101**(8), 1410–1418 (2018). <https://doi.org/10.1016/j.pec.2018.03.015>
 42. Bartram, L., Patra, A., Stone, M.: Affective color in visualization, in *Conference on human factors in computing systems*, Denver (2017). <https://doi.org/10.1145/3025453.3026041>
 43. Kong, H., Liu, Z., Karahalios, K.: Trust and recall of information across varying degrees of title-visualization misalignment, in *Conference on human factors in computing systems*, Glasgow (2019). <https://doi.org/10.1145/3290605.3300576>
 44. Hullman, J., Drucker, S., Henry Riche, N., Lee, B., Fisher, D., Adar, E.: A deeper understanding of sequence in narrative visualization. *IEEE Trans. Visualiz. Comp. Graph.* **19**(12), 2406–2415 (2013). <https://doi.org/10.1109/TVCG.2013.119>
 45. George-Palilonis, J., Spillman, M.: Storytelling with interactive graphics: an analysis. *Visual Commun. Quart.* **20**(1), 20–27 (2013). <https://doi.org/10.1080/15551393.2013.765777>
 46. Sturm, B.: The enchanted imagination: storytelling's power to entrance listeners. *School Library Media Annual* **2**(July), 1523–4320 (1999)
 47. Moritz, J., Meyerhoff, H., Meyer-Dernbecher, C., Schwan, S.: Representation control increases task efficiency in complex graphical representations. *PLoS ONE* **13**(4), 1–15 (2018). <https://doi.org/10.1371/journal.pone.0196420>
 48. Peck, E., Ayuso, S., El-Etr, O.: Data is personal: Attitudes and perceptions of data visualization in rural Pennsylvania. In: *Conference on human factors in computing systems*, Glasgow (2019). <https://doi.org/10.1145/3290605.3300474>
 49. Borkin, M., Vo, A., Bylinskii, B., Isola, P.: What makes a visualization memorable? *IEEE Trans. Visualiz. Comp. Graph.* **19**(12), 2306–2315 (2013). <https://doi.org/10.1109/TVCG.2013.234>
 50. Saket, B., Endert, A., Tasko, J.: Beyond usability and performance: a review of user experience-focused evaluations in visualization, in *BELIV*. Baltimore (2016). <https://doi.org/10.1145/2993901.2993903>
 51. Flourish, Flourish data studio <https://flourish.studio/> (2022) Accessed 27 February 2023
 52. Hullman, J., Diakopoulos, N.: Visualization rhetoric: framing effects in narrative visualization. *IEEE Trans. Visualiz. Comp. Graph.* **17**(12), 2231–2240 (2011). <https://doi.org/10.1109/TVCG.2011.255>
 53. Black, J., Bower, G.: Episodes as chunks in narrative memory. *J. Verb. Learn. Verbal Behav.* **18**(3), 309–318 (1979). [https://doi.org/10.1016/S0022-5371\(79\)90173-7](https://doi.org/10.1016/S0022-5371(79)90173-7)

54. Seyser, D., Zeiller, M.: Scrollytelling - An analysis of visual storytelling in online journalism, in International conference on information visualisation, Fisciano (2018). <https://doi.org/10.1109/iV.2018.00075>
55. Viola, I., Chen, M., Isenberg, T.: Visual abstraction, in Foundations of data visualization, pp. 15–37. Springer, New York City (2020)
56. Lee, B., Henry Riche, N., Isenberg, P., Carpendale, S.: More than telling a story: transforming data into visually shared stories. *IEEE Comp. Graph. Appl.* **35**(5), 84–90 (2015). <https://doi.org/10.1109/MCG.2015.99>
57. Cairo, A.: Emotional data visualization: Periscopic’s “U.S. Gun Deaths” and the challenge of uncertainty, <https://www.peachpit.com/articles/article.aspx?p=2036558> (2013) Accessed 27 February 2023
58. Zwinger, S., Langer, J., Zeiller, M.: Acceptance and usability of interactive infographics in online newspapers, in International conference on information visualization, London (2017). <https://doi.org/10.1109/iV.2017.65>
59. Yau, N.: A day in the life of women and men, flowing data. <https://flowingdata.com/2019/03/06/women-men-timeuse/> (2019) Accessed 27 February 2023
60. Brillhart, J., Galances, G.: Decision making in small groups, Ninth Edition, in Effective Group Discussion, pp. 237–258. McGraw Hill (1998)
61. Ellis, G., Dix, A.: An explorative analysis of user evaluation studies in information visualisation. *BELIV* (2006). <https://doi.org/10.1145/1168149.1168152>
62. Hermawati, S., Lawson, G.: Establishing usability heuristics for heuristics evaluation in a specific domain: is there a consensus? *Appl. Ergon.* **56**, 34–51 (2016). <https://doi.org/10.1016/j.apergo.2015.11.016>
63. Parsons, P.: Understanding data visualization design practice. *IEEE Trans. Visualiz. Comp. Graph.* **28**(1), 665–675 (2022). <https://doi.org/10.1109/TVCG.2021.3114959>
64. APA, Heuristics, American psychological association. <https://www.apa.org/pubs/highlights/peeps/issue-105> (2017) Accessed 27 February 2023

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.