UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Empirical Evaluation of Narrative Visualisation: A Dual Approach Using Heuristic and End-User Evaluation Methods

by Nina Errey

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Dr. Christy Jie Liang
and co-supervision of Dr. Tuck Wah Leong

University of Technology Sydney
Faculty of Faculty of Engineering and IT

February 2025

Certificate of Original Authorship

I, Nina Errey, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Computer Science, for the Faculty of Engineering and IT at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed prior to publication.

Nina Errey

February 2025

Acknowledgments

I would like to acknowledge the contribution of a number of people in relation to my thesis.

Firstly, I would like to thank to my supervisor, Dr. Christy Jie Liang, who has believed in my research and has provided me with on-going support even through the difficult times of COVID-19.

I would like to thank my co-supervisor Dr. Tuck Wah Leong, who with his magic ability has managed to lift my research to a higher level and has always provided support.

I would also like to thank my first supervisor, Dr. Didar Zowghi, who provided the foundational support to make this thesis a reality and believed in me even through trying circumstances.

I would like to thank the practitioners of narrative visualisation that have actively contributed to this thesis. Their interest and feedback have meant that this thesis is a product of their passion.

Finally, I would like to thank my family. This is for my dad Norman George Errey. RIP.

Statement of Format of Thesis

This thesis follows the format of a thesis by compilation and is structured as a composition of chapters and publishable works. It includes:

Chapter Two

N. Errey, J. Liang, T. W. Leong and D. Zowghi, "Evaluating narrative visualisation: a survey of practitioners,"

Published in International Journal of Data Science and Analytics 2023

https://doi.org/10.1007/s41060-023-00394-9

Chapter Three

N. Errey, D. Zowghi, T. W. Leong, Y. Wu, X. Yuan, W. Huang and J. Liang “Heuristics for evaluating narrative visualization: a validation study”

Published in the Journal of Visualisation 2025

https://doi.org/10.1007/s12650-025-01051-y

Chapter Four

N. Errey, J. Liang, T. W. Leong, Y. Chen, H. Vally, and C. M. Bennett, “Nudging with Narrative Visualisation: Communicating to a Young Adult Audience in the Pandemic”

Published in Proceedings of ACM Human-Computer Interaction CSCW, 2024

https://doi.org/10.1145/3686950

Chapter Five

N. Errey, Y. Chen, Y. Dong, Q. V. Nguyen, Y. Xiaoru, T. W. Leong and J. Liang "An Age-based Study into Interactive Narrative Visualization Engagement"

Published in IEEE Computer Graphics and Applications (pre-print) 2025

https://doi.ieeecomputersociety.org/10.1109/MCG.2025.3591817

Each publishable paper includes a section on the research approach and method specific to their work. The thesis format will follow the convention: Introduction, Chapter Two, Chapter Three, Chapter Four, Chapter Five, Conclusion and Implications.

All papers are primarily authored by the PhD candidate, Nina Errey, with support from supervisors Dr. Christy Jie Liang, Dr. Tuck Wah Leong, Dr. Didar Zowghi and other advisory academics according to topic.

Nina Errey

Table of Contents

List of Tables and Illustrations

List of Tables

List of Figures

Abbreviations:

SLR: Systematic Literature Review

HCI: Human-Computer Interaction

Thesis Abstract

Narrative visualisation is a form of visualisation that integrates storytelling with data visualisation. To be effective narrative visualisation must not miscommunicate information to its intended audience. Hence narrative visualisation should be systematically and rigorously evaluated to ensure that it is effective. An effective narrative visualisation is dually comprehensive and engaging for the reader. To ensure effectiveness, narrative visualisation evaluation methods must therefore look beyond conventional quantitative visualisation evaluation measures such as task completion time and error rate. This research is in two parts; inspection methods of evaluation and end-user testing methods of evaluation. These are two promising areas for narrative visualisation evaluation research.

The first part of this thesis investigates inspection methods of evaluation. Inspection methods of evaluation are a group of evaluation methods where a small number of experts inspect a user interface to discover gaps in its design. When evaluators employ a set of guidelines or heuristics as an aid, this is termed ‘heuristic evaluation.’ This research motivates the necessity for a set of heuristics to employ when inspecting narrative visualisation for evaluation. Then presents a heuristic framework for the evaluation of narrative visualisation. The second part of this thesis focuses on end-user testing in narrative visualisation. It comprises of two empirical experiments. The results of the end-user testing study contribute to a better grasp of how technologies such as narrative visualisations, using different strategies, can be better designed.

Keywords: narrative visualisation, data story-telling, data visualisation, evaluation, heuristic evaluation, empirical studies, end-user evaluation

Chapter One—Introduction

The narrative visualisation follows a long tradition of graphical representations that relay abstract information. The pioneer of graphical representation was William Playfair (1759 - 1823) who sought to replace conventional tables with the systematic visual representations with his “linear arithmetic.” He developed the first graphical representations of information such as scatterplots, time-series and multivariate displays [187]. Two centuries after Playfair came political scientist and statistician, Edward Tufte. He is described by the New York Times as the ‘Da Vinci of Data.’ He outlined the fundamental structures that information can be graphically represented. His seminal text ‘The Visual Display of Quantitative Information’[187]. is the starting point for an analysis of modern narrative visualisation. He rallied against unnecessary embellishments in graphics which he termed ‘chartjunk’. This resulted in leading media outlets using a level of restraint in their narrative visualisation designs [150].

In contrast to the minimalist graphics that were a product of Tufte’s anti-embellishment sentiment, graphic designers, most notably, Nigel Holmes designed feature-rich graphics. His designs appeared in magazines such as TIME, New Scientist and Fortune. He states that a graphic “must engage the reader’s interest [88].” He then demonstrated how embellished data through graphical imagery can engage a reader. Even though a general distrust of ‘beautiful’ graphics continued to be an issue for readers [150], their superiority in some circumstances is irrefutable. An example of this is a study by Bateman et al. where it was shown the addition of visual embellishment, specifically the work of Nigel Holmes compared to a plain chart allows users to recall data significantly better over time [25].

The interactive narrative visualisation is the subsequent step in the evolution of the narrative visualisation. Usually appearing online, they offer the user the opportunity to navigate a path through the dataset according to their interest and pace. “Therefore, each user acts like a director who composes his own story [116]” Their development has become richer with the introduction of enhanced client-side coding technologies such as JavaScript libraries or the vector animations of CSS3. These technologies have allowed for new opportunities for building interactive narrative visualisation. As practitioners embrace these technologies the propensity of interactive narrative visualisation will likely only increase.

One of the most pronounced hurdles in the development of interactive narrative visualisation is that they are laborious and time-consuming to build. An author of narrative visualisation described the development of interactive narrative visualisation as “melding the skills of computer science, statistics, artistic design and storytelling [169].” It is therefore that narrative visualisation is reserved for important information. Examples include explaining the extinction crisis to America’s debt ceiling [95, 100]. It is necessary that narrative visualisation is evaluated so that the critical information it contains is not misconstrued or deemed unengaging. Rigorous evaluation is key to ensuring effective narrative visualisation. The next section outlined further the necessity of rigorous narrative visualisation evaluation.

Motivation for the Research

In the public sphere combining data visualisation and storytelling is gaining momentum [169]. As narrative visualisation gains popularity it is crucial that it communicates effectively. To ensure effectiveness, the narrative visualisation development process necessitates a rigorous and systematic evaluation approach. Through evaluation the assurance of narrative visualisation effectiveness can be established [47].

Recently, there has been a growing emphasis on visualisation evaluation [65, 136]. When referring specifically to narrative visualisation, there is currently an ad-hoc approach to evaluation. For example, Nowak et al. proposed a qualitative, micro-phenomenological method when evaluating narrative visualisation, where, through a series of elicitation interviews the effectiveness of a narrative visualisation is established [137]. Conversely, Boy et al. approach narrative visualisation evaluation through quantitative data collection methods [33]. By relying on behavioural proxies, such as session times compared to reading times or click-through source, Boy et al evaluated the effectiveness of narrative visualisation quantitatively. Both the aforementioned methods are not suitable for narrative visualisation evaluation. While a micro-phenomenological evaluation method is detailed, it is not realistic to apply in an everyday professional setting [137]. Furthermore, although quantitative data collection methods can be considered valuable. The most appropriate approach for evaluation is advised as being best described as holistic [162].

A holistic approach to narrative visualisation evaluation is required since narrative visualisation is described as a complex and rich user experience and therefore cannot be adequately evaluated through web analytics alone. This does not discredit a quantitative approach, as it can outline initial indications and aid in the overall evaluation of narrative visualisation in a mass media context [33]. An effective narrative visualisation must be judged on both its ability to communicate to the individual but also as a communication device that is consumed by a mass audience. Therefore, it can be argued that web analytics metrics are needed for narrative visualisation evaluation, equally as much as qualitative evaluation methods. This research aims to develop a holistic evaluation approach for narrative visualisation. Currently, a comprehensive and coherent evaluation framework does not exist for narrative visualisation. The lack of such a foundational framework in this genre of visualisation has resulted in a lack of consensus in narrative visualisation development. Through creating an evaluation

framework that is designed specifically for narrative visualisation a robust and rigorous evaluation can be performed.

Objectives

This thesis is an effort to address the gap in literature surrounding narrative visualisation evaluation. The thesis is split into two parts. The two parts reflect each contribution. The overall aim is gain insight into how to holistically evaluate narrative visualisation and to better inform practitioners in their design practice.

The first aim of my research is to develop a framework for evaluating the effectiveness of narrative visualisation using inspection methods. This evaluation framework will be namely a heuristic framework. This will be validated where the objective is to refine and identify any weaknesses in the set of heuristics.

The second aim is end-user test narrative visualisation to examine effectiveness in different contexts. The second aim is a break from the first. Rather than focusing on the evaluation methods it focuses on narrative visualisation effectiveness in different contexts using different communication strategies.

The following table outlines the research objectives as they relate to research questions. Each research objective will be detailed further in this section.

No	**Research objectives**	**Research questions**
I	Explore the criteria that constitute an effective	RQ1: What are the criteria for an effective narrative visualisation?

	narrative visualisation	
II	Develop and refine systematic heuristic evaluation framework to evaluate narrative visualisation	RQ2: How can narrative visualisation be evaluated for its effectiveness using systematic heuristic evaluation framework?
III	Develop and refine end-user methods to evaluate narrative visualisation	RQ3: How can narrative visualisation be evaluated using end-users?

Table 1: Alignment of research objective and research questions

Research objective I: Explore the criteria that constitute an effective narrative visualisation

This study will investigate what is an effective narrative visualisation. This includes identifying and categorising the criteria that determine an effective narrative visualisation correlating it to its purpose. It is required that the criteria of effective narrative visualisation have been outlined before evaluation can be considered.

Research objective II: Develop and refine systematic heuristic evaluation framework to evaluate narrative visualisation

The development of a set of heuristics will be driven from the result of a literature review and practitioner feedback. The heuristics will focus on being universal and minimalistic.

It is then validated in two settings with experienced practitioners of narrative visualisation.

Research objective III: Develop and refine end-user methods to evaluate narrative visualisation

The investigation of end-user methods of evaluation is driven by empirical research. The primary aim of this objective is to better understand narrative visualisation audiences and explore various methods to evaluate their experience of narrative visualisation. Through rigorous end-user evaluation ultimately authors of narrative visualisation can be better informed when making design choices for their audience.

These three research objectives form the basis of the research I plan to perform. The significance of this work is that it provides foundational research in the area of narrative visualisation. As it is an emerging research area the opportunity exists to modulate the field and pioneer a fundamental theoretical framework that will encourage future research and development.

The following section will describe the research approach and proposed development of the evaluation framework.

Methods

This thesis has a mixed-method approach. The methods used for each study are discussed more deeply in each chapter in their own section to provide context around their use. In brief, Chapter Two's research design encompasses a systematic literature review and snowball sampling techniques to examine empirical experiments in narrative visualisation. This leads to the identification and inclusion of 17 relevant documents which are the basis of a set of elements to inform later research. Furthermore, Chapter Two is a lengthy and diverse chapter, where outside of a literature review it presents the results of a survey of practitioners and a set of interviews. The survey was informed by

the literature review and gave oversight of the current state of evaluation. The interviews provided in-depth understanding to the reasoning and cognitive processes that were apparent in the survey. The survey questions were analysed statically where appropriate. To analyse interviews we used a latent form of thematic analysis [35]. Here we analysed the data through underpinning concepts and assumptions which have been appropriated for narrative visualisation practitioners. For example, the theme of 'responsive' is a term that is understood by practitioners and means that a visualisation is developed to include unconventional browsing devices.

Chapter Three contains two separate experiments to validate the set of heuristics in different settings. The first experiment uses a survey instrument designed so that a practitioner can heuristically evaluate narrative visualisation. This instrument allows for externally hosted narrative visualisation to be pulled into the tool for evaluation. The data produced from this experiment was analysed for inter-rater reliability. This method was chosen as it illustrated the reliability of the heuristic evaluation framework to create repeatable evaluation metrics. The second experiment used an interview method. Differing from the first experiment the second experiment was in an informal group setting and allowed for qualitative feedback.

Chapter Four is a break from earlier chapters. This focuses on an empirical experiment to evaluate communication strategies in narrative visualisation from the user perspective. End-user testing requires stricter parameters to make concrete findings. We therefore used a model for measuring attitude to inform our research method. Namely the ABC model, this attitude measurement model is common in marketing analytics [59]. We quantitatively analysed end-user attitude before and after viewing a narrative visualisation stimulus. This method was appropriate for our aims and, to our knowledge the first time it was employed to evaluate narrative visualisation.

Chapter Five is similar to Chapter Four by focusing on end-user evaluation. In this chapter we are measuring engagement rather than attitude in end-users. To measure engagement we employ a visualisation engagement self-reporting questionnaire named VisEngage [92]. The aim of this experiment is to compare engagement across age cohorts. It is thus that VisEngage was appropriate for our aims as it customised to the visualisation domain. To analyse the data from this experiment we combined and compared engagement scores. The outcome of the experiment was determined by the difference, or lack of, difference between scores.

Key Findings

Chapter Two revealed multiple findings in regard to narrative visualisation evaluation practice. More detail and impact of findings are explained within the chapter. As an overview, we found basic statistics about evaluation, or there lack of within narrative visualisation practice. For example, we found that approximately half of practitioners include end-users in their evaluation. Another finding that motivated our later establishment of a heuristic framework was that 68% of practitioners did not use a set of guidelines or heuristics to evaluate narrative visualisation. We found some examples of novel evaluation methods that could warrant future research. One such example was the use of guerilla user testing. Once again, more detail can be found within the chapter.

Chapter Three presents findings on whether the heuristic framework established in Chapter Two is fit for purpose. The key finding is that our heuristic framework shows promise as a useful evaluation tool. It was found that practitioners found the set of heuristics insightful but would most likely be used in conjunction with other guidelines or heuristics. Two key themes emerged about where the heuristics would need to be complemented with other heuristic sets. These themes were, data analysis and usability. While the heuristic set is useful, it is too narrow in its focus on storytelling and therefore should be used in conjunction with other heuristic sets.

Chapter Four examined different communication strategies integrated into narrative visualisation and their effects on audience attitude. Multiple findings are reported in this chapter. Firstly, that the approach of using the ABC model to measure attitude change in narrative visualisation readers is a discovery that could potentially be used in other contexts. Secondly it was found that the interactively integrating the reader into a narrative visualisation narrative results in significant attitude change difference. It is widely regarded that those with strongly held beliefs cannot change their attitude by viewing narrative visualisation. To change the attitude of readers was not the aim the chapter. Instead we found that it was possible to remind the reader, or as we term it 'nudge' the reader. A 'nudge' is a term adopted from Thaler and Sunstein's seminal book, where it is described as 'a friendly push toward desired behaviour.' The possibility to 'nudge' an audience with narrative visualisation is a key finding of this thesis.

Chapter Five contains multiple findings that were a direct result of an empirical experiment investigating the effects of age on narrative visualisation engagement. Our primary finding from this experiment was that there is a significant difference between age groups and engagement. While arguably not particularly interesting as the result was expected, it is still important because it re-enforces the necessity for considering audience in narrative visualisation design.

Contributions of the Research

This research contributed to the field of visualisation in multiple instances. The contribution is not limited to the immediate results of experiments, but we hoped it would encourage greater emphasis on evaluation in narrative visualisation. Through increased dialogue between practitioners and researchers, we believe, rigorously evaluated and

Chapter Two represents two pivotal contributions. This chapter presents a heuristic framework for the evaluation of narrative visualisation. It is the first, to our knowledge,

that is available in extant literature. Our unique approach of forming the framework consisting of practice-led and thoroughly validated further highlights the importance of this contribution. While other guidelines for narrative visualisation do exist, none are informed by practitioners, who have valuable lived experience. Another important aspect of the heuristic framework is that it contains clear usage instructions that were informed by visualisation research and practice.

Many domain specific heuristic sets are not validated. By validating the heuristic framework it marks it as more established than other examples. This is because through validation, weaknesses and gaps can be found within the heuristics. The validation study in Chapter Three found the limitations of the heuristic set and simultaneously refined them. This is a key contribution as it solidifies the heuristic framework as a usable tool for evaluation.

In Chapter Four contributes a new way of regarding narrative visualisation. As a communication medium narrative visualisation is shown to be able to ‘nudge’ an audience. This is an important contribution as changing the attitude of an audience, albeit slightly, can be a serious mechanism in a democratic country. In this chapter we also establish a new approach to end-user testing narrative visualisation through using the ABC model. While this is a model that has been employed to evaluate other forms of visual media, this is the first time that it is used for the evaluation of narrative visualisation.

Finally, Chapter Five contributes by better informing practitioners of narrative visualisation. It is verifies the claim that audience age impacts the level of engagement in narrative visualisation. This final chapter also contributes a new method to evaluate narrative visualisation for end-users. Namely using the VisEngage model, where this

model has been used for interactive data visualisation, but not before in the context of narrative visualisation [92].

Structure of the Thesis

The structure of the thesis is as follows. Chapter Two serves as an introduction to the topic of narrative visualisation evaluation and investigated current evaluation practices. It then proposes a framework for heuristic evaluation stemming from a survey and a series of interviews with practitioners.

Chapter Three is dedicated to the validation of the heuristic framework proposed in Chapter Two. It does this by employing the proposed heuristic framework in two settings with professional practitioners of narrative visualisation. It is important to note that Chapter Three does not end the validation and establishment of the set of heuristics. Over time and usage they will grow and evolve. The work presented in Chapter Three is a step toward the establishment of a set of heuristics for narrative visualisation.

Chapter Four and Chapter Five are both concerned with end-user testing. These are separate from earlier chapters, however are still continue on the topic of narrative visualisation evaluation. The requirements of end-user testing dictate a different approach. Concrete findings need exacting parameters in the context of larger end-user testing experiments. We outline the limitations and the threats within each chapter. To conclude this thesis we summarise our findings, present future research avenues and once again outline the limitations present through-out the research.

Background

Definition of Narrative Visualisation

Narrative visualisation has two distinguishing features which differentiate this form of visualisation from other forms such as data visualisation or information visualisation.

Firstly, the purpose of the visualisation is a point of differentiation. Data visualisations are “are often conceived as tools that let people extract their own conclusions from the data [42].” An example of data visualisation is a Tableau dashboard. Information visualisation has a similar purpose and is described as an aid to make decisions, understand complex systems and find information that otherwise might remain hidden in data [123]. An example of a famous information visualisation is the London underground map. These forms of visualisation are tools that function primarily to visually represent datasets [131].

Narrative visualisation conveys a pre-meditated message to the reader and is therefore fundamentally different in purpose to forms of visualisation. narrative visualisation communicates through a process of curation by the author, who presents data with the purpose of transferring their intended message [91]. narrative visualisation implies a certain level of authorship, which is argued by Weber to represent the presence of a narrator [194].

Authorship is the second fundamental distinguishing feature of narrative visualisation. The role of a narrator does not need to be explicit and may be purely the fact that the data is embellished. In this context the term embellishment is any sort of decorative, illustrative addition to a representation which separates it from a computer generated, visualisation of pure data [31]. Studies suggest that embellished data improves reader comprehension and memorability [31]. It has however been long debated whether embellishment should be kept to a minimum. This concept was coined the ‘data/ink’ ratio by Tufte [187], who rallied against any extraneous embellishment that could distract from the data. The process of data embellishment is however vital to the development of narrative visualisation. It demonstrates that developing narrative visualisation is a series of aesthetic and rhetorical design decisions which influence reader interpretation and ultimately their comprehension.

Through analysis of narrative visualisation, it is possible to outline the components and functions defines this form of visualisation. The narrative components and structures that are characteristic to narrative visualisation have been described and categorised [169]. The sequence of story elements has been investigated, where effective narrative visualisation story sequences and scenarios are outlined [90]. One of the noted reader experience outcomes of narrative visualisation is that it can capture the attention and induce a state of flow in the reader. The factors that contribute to this state of flow have been defined [124]. Another function of narrative visualisation is the ability to frame concepts through a series of rhetorical choices that effect end-user interpretation. These rhetorical techniques have been investigated and effective communicative strategies for narrative visualisation development described [91]. Once the term narrative visualisation, its components and functions have been clearly defined it is possible to include subsets of visualisation terms inside this term. Narrative visualisation therefore encompasses other terms including; data story, data comic and information graphic.

Systematic Literature Review

I performed a systematic literature review with multiple aims. Firstly, to define what is an effective narrative visualisation. Secondly how is narrative visualisation evaluated and finally, what are known challenges in narrative visualisation evaluation. The results of this review informed later research, where, for example the criteria to evaluate narrative visualisation informed the survey questions for narrative visualisation practitioners.

A Systematic Literature Review (SLR) follows a rigorous procedure where a transparent, repeatable search strategy must be adhered to. It has three phases; planning, executing and reporting results.

In the planning phase, initially research questions are devised. From the research questions search keywords are identified and a search string is concatenated. The executing phase is where the search string is executed on selected databases. Once

results are retrieved a set of inclusion and exclusion protocol are applied. Once the protocol has been applied, the remaining studies are appraised for quality using a quality checklist. Once the primary search is complete the secondary search strategy is applied. The secondary search strategy includes the 'snowball effect' where the references of selected papers are analysed. Then a search of prolific authors and further a search in relevant publications results in a final set of selected studies. This is a rigorous approach, where as many relevant studies are found as possible.

The final phase, the reporting phase, is where selected studies are synthesized and analysed. In accordance with a data extraction form, each study was systematically reviewed and data collected according to the requirements of a data extraction form.

For a full list of selected studies see Appendix A. In this document, when references are made to the selected studies of SLR, I use letter S followed by the number allocated to the relevant study, (e.g. S1).

Definition and Measure of Effectiveness

In this review, I have collected definitions of effectiveness of narrative visualisation in different contexts. Ultimately, it depends on the definition of effectiveness as determined by the objective of the study. Effectiveness has been defined by the standard ISO-9241-11 in 1998, as the "accuracy and completeness with which users achieve specified goals [1]." To define effectiveness in the studies included in this review I have outlined the specified user-goals in the study. These goals were determined by my interpretation of each study. These goals are; preference, memorability, usability, engagement, first impression, empathy, recognition, cognitive load, risk perception and insight.

User goals can be classified as usability goals or user-experience goals [162]. The differentiation is that user-experience goals include an emotional response. These are argued by Saket et al.[162] to be crucial in the evaluation of effectiveness of a

visualisation. An effective narrative visualisation therefore corresponds directly to the user's ability to complete a goal which could vary from their ability to recall data in a memorability task, to their ability to create insights from the visualisation.

There is one measure of effectiveness that is pervasive is the accuracy of the representation of information as authored by the narrative visualisation practitioner, should correspond to the mental model simulated by the user. This means that successful explanatory visualisation must be reflected in the understanding that is gleaned by the user. This measure of effectiveness is described in Tversky's 'Cognitive Design Principles', who terms this the 'Principle of Congruence [16].'

Definition of an effective narrative visualisation: *the accuracy and completeness with which users achieve specified goals determined by the purpose of the visualisation. This must then be combined with the accuracy of the representation of information as authored by the narrative visualisation practitioner reflected in the mental model simulated by the end-user.*

This definition is central to my research and is fundamental to the proposed evaluation framework.

Criteria of Effective Narrative Visualisation

Once the definition of effectiveness had been established it was possible to devise a set of criteria that contribute to an effective narrative visualisation. These criteria were based

on the findings of the literature review and encompass only evidence-based outcomes of empirical experiments that were mentioned in the literature.

The criteria for an effective narrative visualisation are as follows:

1. Data Legitimacy through Identifying Real Sources

It is crucial to reveal the source of the information that is represented in narrative visualisation as it improves credibility and data legitimacy. This was found in S23 where information about the provenance of the source added credibility. Therefore, not merely adding the source, but further source analysis increases credibility. In S6 it was found that by revealing the source of the information presented in a visualisation, the perceived credibility is increased. Peck at al. (S6) found however, that less-educated participants did not consider the source of the visualisation as closely as those with more education. This suggests that narrative visualisation, regardless of source, should be built with accurate reflections of the data presented, as false data can lead to miscomprehension by the less visually literate.

2. Information density

Less literate audiences are overwhelmed by information dense visualisation (S6 and S19). Harrison et al. (S15) recommends that designers aim for a low to medium information complexity if they want to create an appealing narrative visualisation. Similar to the findings of S6, Harrison et al. (S15) observed that age significantly affected participants' preferred levels of visual complexity. In S19, the participant group was deliberately selected from an age range of 45 to 65 and the outcome of this study was that narrative visualisation can have a detrimental effect to communicating health risks. Damman et al. (S19) concluded that this "might suggest that effects were due to information overload."

3. Careful Use of Colour

The aesthetics of a visualisation, specifically its colourfulness requires careful authorship. It was shown in multiple studies (S7, S15, S27, S28 and S29) that colour was crucial in gaining the attention and interest of users. Borkin et al. (S28) found that colourful visualisations are more memorable. Harrison et al. (S15) observed that colourfulness of a narrative visualisation determines its appeal more than its visual complexity. It was observed, however that the use of colour in narrative visualisation is quite different to the use of colour in websites, where rather than a design feature it was a visual cue or an indicator. Harrison et al. (S15) writes that visualisations "rely on colour and colour composition as a design element much more than websites." Furthermore, Lonsdale et al. (S29) recommends "Colour should be used sparingly, with good contrast for text and images, and as an information tool (not as decoration)." Overly colourful visualisations are at risk of becoming childlike and the legitimacy of their data is compromised.

4. Chunking Principle

One of the key benefits of narrative visualisation is the ability to divide information into smaller portions that are easier to comprehend. This division of information into portions is called chunking and is referred to as the 'chunking principle.' Lonsdale et al. (S29) explains that cognitive load is reduced if information is presented in a limited amount of chunks. Working memory has a finite storage capacity so the optimum number is three to four chunks. The studies that mention chunking are S21, S25, S29 and S30. When comparing data comics to a visualisation in S21, Wang et al. concluded that data comics, because they divide the information into panels, which are therefore memorable chunks, are more memorable than other formats.

5. Left to Right Layout or 'Zigzag'

Sequential ordering of information in a conventional left-to-right, vertical layout was observed to be preferred by users. In an eye-tracking study (S30) where various layouts

were compared, the layouts with the least amount of 'workload' performed best. Workload in this context was measured through pupil dilation which was an indicator of cognitive load. Majooni et al. (S30) observed that a 'zigzag' form of layout enhance comprehension because it decreased workload. McKenna et al. (S32) investigated narrative flow in narrative visualisation, where scrolling down the page was the preferred mode of interaction. Similar to the findings of S30, users preferred the method that required the least amount of effort. It should be noted that this is for an audience that speaks modern European languages where left-to-right is the prevalent reading direction.

6. Relatable Content

Visualisations are more effective if the content they portray is relatable. Borkin et al. (S28) described on a fundamental level, the elements of effective narrative visualisation must have human recognisable imagery to be appealing. Human recognisable imagery was considered the use of either photographs or cartoons, which was referred to as a pictogram. This study had 261 participants viewing 2070 single panel graphics in a game-like scenario where they needed to recognise duplicate visualisations. Memorable visualisations were more likely to contain pictograms. "Since we are mostly attuned to natural scenes, it makes sense that some of the top memorable visualisations look closer to "nature" than the others. In these terms, we see that people may have more perceptual fluency with visualisations that at first glance appear to be more "natural" and that this fluency may be influencing memorability (S28)."

7. Personal Content 'Designing Conversations'

Relatable content can also be considered in a personal sense. Peck et al. (S6) interviewed rural community members about their comprehension of a series of visualisations that depicted drug use in the US. "The most recurring theme in our analysis were decisions framed or driven by personal experience (S6)." The personal framing of

data, which alters the attention toward or away from the specific data contained in a graphic, must be considered. Certain concrete elements of narrative visualisation, were suggested by Peck et al. (S6) to aid in the pursuit of personal goals such as adding a search bar or more realistic analogies of distance. Another recommendation in regard to creating relatable content is for the addition of a narrative element which would personally involve the user through a highly interactive ‘dialogue’ as described by Oh et al. (S18). A conversation that personally includes the user engages their imagination and thus retains their interest. As was observed by Bumester et al. (S8) interest is the primary driver when using a narrative visualisation. By creating content that personally interests the user or personally addresses the user through a dialogue results in narrative visualisation that are relatable and therefore more effective.

8. Cohesive/salient Design

Sudakov et al. (S11) write that a well-designed narrative visualisation “allows students to formulate understanding more rapidly.” The addition of images to text is thought to aid learning and is termed the ‘multimedia effect [122]’ Some studies contained in this review do not entirely support this theory, they do however, support a cohesive design where mixed media needs to be salient and relevant to the learning outcomes.

9. Representation Control

Moritz et al. (S5) found that users managed to solve tasks faster and more accurately when they are able to manipulate the representation of information. One of the key factors for the increase of task efficiency and accuracy was attributed to a reduction information density. When unnecessary information is hidden from view users can analyse information faster. Manipulation graphics allow the user to reorganise information which results in information coherence and spatial contiguity. These factors also contributed to increased task efficiency. Moritz et al. write “representation control

allows the viewers themselves to interactively adapt a given graphical representation to task requirements, thus increasing its ease of use (S5)."

10. Findability

Zwinger et al. (S9) found that the findability of narrative visualisation news organisations' website was the greatest hurdle preventing news readers accessing the visualisation.

11. Flow Factors

Effective narrative visualisation are able to hold the attention of readers through a series of 'flow factors.' These are studied in McKenna et al. (S32) and identified the factors that contribute to readers' engagement. The importance of 'flow factors' is described by McKenna et al. (S32) "each work functions as a "cognitive container" in which media add-ons work to hold reader attention rather than scatter it to external Web sources."

12. Careful authorship of Text or Addition of Text

The importance of well authored text that clearly conveys the health message is a vital component of an effective narrative visualisation in the health domain. Multiple studies found that slight variations in messaging can change the user experience and overall comprehension. In study S23 three iterations of a visualisation were presented to users with slight differentiations in messaging. The objective of the study was to find out if by communicating the effectiveness of a sugar tax, would increase support for the tax. By simplifying the visualisation and adding an introductory 'refutation', support for the tax increased among participants. The margin was from 45% to 49% which when extrapolated out to population level is substantive.

The assumption that a highly visual representation will be more accessible to a less literate audience has been questioned. Damman et al. (S19) found a detrimental effect from visualisation on user risk perception/comprehension. The challenges posed in the

communication of risk to lay audience is described by Spiegelhalter et al. [175] where he presents a series of guidelines to follow. When these guidelines are adhered to then a visualisation may be effective. The consequences of a visualisation, specifically in the health domain, that does not consider an audience with low visual literacy can have a detrimental, if not, damaging effect.

Existing Evaluation Frameworks

I have extracted five existing evaluation frameworks from the studies in the review. None of these frameworks are specifically devised for narrative visualisation and most are concerned with a particular aspect of narrative visualisation. They are therefore inadequate for the unique requirements of robust narrative visualisation evaluation.

Evaluation Frameworks	**Author**	**What for?**	**Ref.**
Data Ink Ratio	Tufte, ER	Printed diagrams	[187]
Visualizing Risk	Spiegelhalter, D., Pearson, M. and Short, I.	Risk visualisation	[175]
Cairo's Wheel	Cairo, A.	Aesthetic design	[42]
Evaluation Scenarios	Lam H, Bertini E, Isenberg P, et al.	Usability/Ease-of-Use	[109]

Elaboration Likelihood Model	Petty, R. E., & Cacioppo, J. T.	Attitude shifts/Persuasion	[146]

Table 2: Existing evaluation frameworks that have been employed to evaluate narrative visualisation

One of the most prominent forms of graphical evaluation is the data/ink ratio. This originates from Edward Tufte who, in 1983, proposed evaluating a graphic on its ability to present data using as little ink as possible [187]. The data-ink ratio was calculated by dividing the ink used for displaying the data by the total ink used in the graphic. The studies which cite the data-ink ratio are S2, S4, S13 and S28. In regard to visualisations the ability to convey information using minimal visual elements can be a key component of their effectiveness. The excess visual elements were termed by Tufte as 'chartjunk' which is referenced in the title of S13, 'Useful junk? The effects of visual embellishment on comprehension and memorability of charts' [25]. This well-cited study published in 2010 refutes the data-ink ratio which shows that visual embellishments are not detrimental to understanding. It is therefore that the data/ink ratio is inadequate as an evaluation framework for narrative visualisation.

The design of a visualisation is fundamental to its effectiveness to communicate, especially to communicate to the public. Spiegelhalter et al. is cited in S1, S19, S23 and S29 [175]. This paper is in regard to visualizing uncertainty. It outlines design guidelines of visualisation evaluation through a series of broad conclusions that can be applied to large audiences. What is advocated is an approach that is clear and accessible to all audiences. The broad conclusions devised by Speigelhalter et al. could be considered as design guidelines that can also be implemented as an evaluation framework. These guidelines are however in relation to the visualisation of risk. As narrative visualisation

is not necessarily a portrayal of risk, this evaluation framework is inadequate for a robust evaluation of narrative visualisation.

The design quality of a visualisation is also evaluated using Cairo's 'Visualisation Wheel Model [42].' One study (S4) assessed visualisations according to a series of 'quality dimensions'. These dimensions appear on the 'Visualisation Wheel'. Added quality dimensions of usability and ease-of-use are added to accommodate for interactivity. The outcome of the paper's findings is an exploratory model of the "overall quality" which was derived from an empirical experiment can be considered an evaluation framework. This model evaluates the design of the narrative visualisation but does not consider the content. This therefore means that this form of evaluation is inadequate for narrative visualisation.

Evaluating an interactive visualisations can deviate from design quality and focus on other dimensions such as ease-of-use or usability. Lam et al. derived seven evaluation scenarios based on a literature review of 800 visualisation publications [109]. This work is referenced in S26 which as described earlier in this paper investigates visualisation literacy. As is described by Lam et al. evaluation can range from visual data analysis and reasoning, to visualisation algorithm evaluation. narrative visualisation evaluation does not require all scenarios described by Lam et al. to be considered [109]. However, scenarios including user performance and user experience are pertinent. Therefore, due to the fact that these evaluation scenarios are not reflected in narrative visualisation, this form of evaluation is unsuitable for narrative visualisation.

The evaluation of visualisations can implement conventional visual communication theories such as Elaboration Likelihood Model (Petty and Cacioppo [146]). This model is cited in S3, S18, S22, S23 and S27. This model is pertinent to narrative visualisation as it describes attitude shifts. The evaluation of the message is based upon its ability to persuade the user. It suggests a dual-processing model where for example, visuals are

peripheral to text processing. In S27 titled ‘Putting Environmental Infographics Center Stage: The Role of Visuals at the Elaboration Likelihood Model’s Critical Point of Persuasion’ finds that the visual cues are found to be central elements and equal in importance to textual elements. This model is concerned with purely the ability for a visualisation to persuade, as narrative visualisation have more purposes that persuasion this evaluation framework is inadequate.

Finally, narrative visualisation are largely a web-based phenomenon and therefore website analytical evaluation frameworks have in the past been adopted accordingly. In S14, user-centred metrics were collected using the Gotz and Wen patterns of user-behaviour [76]. Four common categories of interactions were identified by Gotz and Wen; scan, flip, swap and drill-down. The objective of S14 was to evaluate the impact of initial narrative visualisation techniques on user-engagement with exploratory information visualisations. Therefore, to evaluate user-engagement a series of interactions were quantitatively analysed as they correlated with the interaction categories identified by Gotz and Wen. This evaluation framework does not include the emotional response of users as it is based only on their interactions rather than their inner-perceptions. As described earlier in this document, the rich and complex nature of narrative visualisation requires a holistic evaluation that includes the emotional response of users. It is therefore that this evaluation framework is inadequate for narrative visualisation evaluation.

As is described above each evaluation framework that has been implemented when evaluating narrative visualisation is inadequate. This does not mean they are inadequate in the context of the particular study, however, they are inadequate to provide a holistic and comprehensive evaluation approach to narrative visualisation.

Through adopting a broad approach, coupled with a rigorous search strategy, a complete sample should be ensured. Some papers may not be included due to their unavailability

in electronic resources. When composing the search string, based on the results of the pilot study, the terms ‘InfoVis’ and ‘Information visualisation’ were eliminated. These terms were eliminated as it defines a discipline that is differentiated from narrative visualisation. As described in section 1 there is a significant difference between information visualisation and narrative visualisation. According to Rendgen S., information visualisations are visual representations of pure data and are used in the fields of scientific research, digital libraries, data mining etc. [150]. In two studies the term ‘information visualisation’ was used in S14 where it was described as ‘narrative visualisation.’ Lyra et al. (S16) describes; “In this context, infographic is a more recent and popular type of information visualisation able to support learning.” This statement accommodated for the overlap of the term information visualisation and information graphic in deeming S16 a suitable study.

Data visualisation was not excluded from the search string as it was a widely used term that often overlaps with narrative visualisation. Data visualisation was used as a term that can be considered synonymous to narrative visualisation in 8 of the studies in this review; S2, S6, S7, S12, S17, S18, S26, S32.

Challenges in Evaluation

Many of the challenges met when evaluating narrative visualisation are common to all empirical research [47]. Some examples are picking the most suitable methodology or a sufficiently rigorous data collection process. Another challenge is the difficulty in obtaining an appropriate sample of participants, which adequately represent the user base. Also, there is the common challenge of data analysis where the results must be accurately synthesized and related to previous research and theories. Furthermore, the theories that the research is grounded upon may not yet be adapted to the emerging field of narrative visualisation.

More specifically the evaluation of narrative visualisation shares many challenges common to HCI empirical research. Carrol and Rosson wrote of a ‘claims analysis’ that listed the factors affecting the usability of interfaces, e.g. consistency in navigation, colour of rollovers etc.[48]. These ‘claims’ are often fundamental to the design of software or web development, however as described by Ellis et al. they are often difficult to articulate when producing ‘novel visualisation techniques [61].’ Carrol and Rosson’s claims analysis offers an ‘explanation’ of the scenario that the user is experiencing, but when that scenario is unique to the visualisation the explanation is unsubstantiated and difficult to repeat. The lack of a commonality of ‘claims’ results in the necessity of users to describe in their own words their experience.

Furthermore, narrative visualisation as a form of communication is implicitly novel and unique as the story or insight it portrays is particular to that visualisation. Therefore, it is challenging to rely on the generalized evaluation procedures that might not be appropriate to the individual visualisation. This limitation was noted in multiple studies where the universality of findings was questioned due to the localized materials [64, 124]. Nowak et al. describes how traditional evaluation methods and metrics are insufficient for capturing the “richness” of narrative visualisation [137]. Qualitative methods are recommended as they can evaluate the complex and holistic nature of experiences.

Qualitative methods are crucial when evaluating narrative visualisation as they are able to incorporate the personal perspective of users. The complex set of factors that intermix to inform attitudes and perceptions about visualisations can skew results. In a study by Peck et al.[143] it was found that personal connotations with the presented data may supersede many other dimensions of design. Personal connotations can also be the product of the practitioner [91]. Through framing and rhetorical techniques the information presented in a narrative visualisation can be distorted. This is however, arguably the purpose of narrative, which is as first described a tool to compel the user.

It is therefore not useful to measure the accuracy of data representation if that is not the objective of the visualisation and does not reflect its effectiveness.

When the effectiveness of a visualisation is defined by the mental simulation of the user, the challenge is how to measure it. North proposes measuring the insight gained by the user to evaluate visualisation [136]. He describes that “insight seems to capture the intuitive notion of visualisation’s purpose.” By investigating the user insight it is possible to gage what the visualisation has communicated in more exact terms rather than a Boolean response. The difficulties with this method is then to report the results in useful manner that can give a clear evaluation of the visualisation. North proposes counting the insights and coding them, which is time-consuming and requires more effort on behalf of the researchers.

As opposed to qualitative methods, controlled experiments are often preferred by researchers as they best enable rigorous measurements and can conclusively compare visualisations. The challenge is then to accurately portray the real-world circumstances that users interact with narrative visualisation. Carpendale writes that “no empirical method is perfect. That is, there is always a trade-off between generalizability, precision, and realism [47].” It is thus that the user testing of narrative visualisation should incorporate some ‘in the wild’ element where the usage is not purely based on the laboratory experiments and skew results.

The above mentioned challenges of narrative visualisation evaluation will each be considered in the final framework. There will be no all-encompassing solution however, the finalised framework will be able to make a series of recommendations and compromises which will consequently result in a more robust evaluation.

Heuristic Evaluation in Visualisation

Heuristics is at times considered to be interchangeable with design guidelines [182]. For the purposes of this work they are more specifically the method of expert review where the expert has a predetermined set of heuristics. This method of evaluation can be effective when a user-study is not affordable or practical. A heuristic evaluation is part of an iterative design process where it can be a relatively fast evaluation methodology. Another term for heuristic evaluation is ‘inspection methods’ which is differentiated from ‘testing methods’ which are user-studies.

Information visualisation researchers have proposed heuristics that are specific to a certain data domain e.g. ambient displays [132] or multiple view visualisations [21] or for a specific cognitive task [8]. The hurdle for information visualisation practitioners is to determine which set of heuristics is the most appropriate for the particular visualisation. In an exploratory study that investigated three different set of heuristics, it was found that the most minimal and therefore generalised set of heuristics was the most useful [26].

There has not been a set of heuristics that has been developed specifically for narrative visualisation evaluation. The value of a set of heuristics, regardless of the subject it is evaluating is indisputable. Santos et al writes “that heuristic evaluation is indeed suitable and produces useful results with a low investment even when performed by analysts not very experienced and hence it should be included in the practitioner’s evaluation toolkit” [26].

Zuk et al. write that heuristic sets for information visualisation can either take the form of a ‘tree organization’ where a ‘depth-first’ search approach is performed according to the requirements of the visualisation [203].

The finalised set of heuristics will be determined based on literature review and analysis real-world of narrative visualisation. Through studying a series of real-world narrative visualisation that have been deemed effective by the visualisation community, patterns will emerge which will contribute to the determining of the heuristics. Effective narrative

visualisation can be found through multiple sources including expert blogs, online forums and journal articles.

End-user Evaluation in Visualisation

User-studies are infamous for being costly and time-consuming [134]. As described by George-Palilonis et al. the development of narrative visualisation often does not lend itself to user-testing [74]. Some challenges given are the expected fast turnaround of projects, the limited number of people that are available to work on the project and finally the extra cost of paying for user-testing. This does not mean however that user-testing should not be part of the project lifecycle, instead, I propose that it must be systematic and time-efficient.

Some of the user-study methods that are used to evaluate narrative visualisation are quantitative and directly transported from website evaluation. These include session times and page views. As described by George-Palilonis et al. these metrics may not be useful when assessing the effectiveness of narrative visualisation [74]. It is posited that 'additional research and tools are needed'. It is not however suggested that these tools should *replace* the quantitative methods, rather they should be added to the toolbox of the evaluator. Instead, quantitative user-study data collection methods can be helpful in founding baseline assumptions which can be an indication of 'in the wild' usage. They can be easily compared with relative website evaluation benchmarks and often serve as initial evidence of the effectiveness of narrative visualisation. An example of this is the now infamous New York Times narrative visualisation 'Snowfall.' Which attracted 3.5 million page views and had a session time 10 times longer than the average [58].

User-study data collection methods that are qualitative such as questionnaires with likert-scale questions or open-ended questions are a popular form of data collection according to the literature review. These may be more appropriate instead of in-depth interviews which was the method of evaluation advocated by Nowak et al. [137]. These

methods may also be employed online without the user being present. It is also possible to quantitatively analyse the results of these methods, even open-ended questions can be codified. Subsequently these methods can be analysed and compared to benchmarks. The think aloud protocol or cognitive walk-throughs were also popular data collection methods, however require the presence of the user in a laboratory environment.

Narrative Visualisation Practice

There have been calls for further overlap and collaboration between major communities that work with data visualisation [141]. One community is the practitioners, who professionally author narrative visualisation. The other community is the academic community that research the role visualisation can play in storytelling with data. There is a widening gap between the two communities "as novel and innovative examples and genres of storytelling with data flourish in the media quite separately from the knowledge being built by the research community [133]."

The knowledge of data visualisation practitioners was recently probed [141]. It was found that tacit knowledge that is acquired through professional experience is often referred to and difficult to articulate. The practitioner, does not engage with a pre-defined process rather they have a conversation with the material or what is described as 'reflection-in-action [167]' Therefore an *in situ* decision making process means that a detailed task-based method for design and evaluation might not be appropriate for practitioners and we have devised our framework accordingly.

The first part of this thesis takes an 'in the wild' research method. This means carrying out research in an every-day and naturalistic environment [50]. This approach has certain disadvantages such as difficulty in reproducibility and easily translatable results [50]. We believe the advantages of this approach outweigh the disadvantages. The most

prominent advantage is the motivation for learning from evaluation in practice, appropriate in a practical setting as well as valuable to future academic research.

PART ONE

Inspection methods of Evaluation

Chapter Two—Evaluating Narrative Visualisation: A Survey of Practitioners

Synopsis

This chapter is foundational for my research on inspection methods of evaluation for narrative visualisation. Working together with collaborators I recognised an under-researched area where there is a lack of documented knowledge on the evaluation of narrative visualisation. When academic research was not prevalent, it was clear that practitioners were the knowledge experts. Their professional experience led to the primary contribution of this chapter which is a set of heuristics for narrative visualisation evaluation.

Published in International Journal of Data Science and Analytics 2023

https://doi.org/10.1007/s41060-023-00394-9

Abstract

Narrative visualisation is characterized by the integration of data visualisation and storytelling techniques. These characteristics provide challenges in its evaluation. Little is known about how these evaluation challenges are addressed by narrative visualisation practitioners. We surveyed experienced narrative visualisation 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 visualisation 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 visualisation systematically. Our practice-led heuristic framework couples first-hand practitioner experience with recent research literature. This work sheds light on how to address narrative visualisation evaluation to better inform both academic research and practice.

Introduction

Narrative visualisation is a form of visualisation that integrates storytelling with data visualisation. This storytelling aspect of narrative visualisation is a key differentiation from other forms of visualisation. To be effective the data-driven story conveyed by narrative visualisation, must be both engaging and compelling [108, 169]. When integrated seamlessly, narrative visualisation can be an effective communication device [75, 108]. It has been employed in a variety of contexts including recent important public health messaging, for example, explaining the benefits of social distancing in the transmission of COVID19 or the exponential growth in COVID19 outbreaks [130, 178].

However, the need for narrative visualisations to deliver engaging and compelling stories also results in challenges for its evaluation. Effective evaluation is key to ensuring that the complex information portrayed by the visualisation is not misconstrued or deemed unengaging. A lack of rigorous evaluation, particularly when the visualisation is used in critical situations, such as public health messaging, can have serious and detrimental consequences.

Conventional approaches to visualisation evaluation which measure task time and error rate are not sufficient in their evaluation. This is because they do not provide crucial insight into the user's comprehension or user experience [20, 137]. Qualitative methods, such as elicitation interviews and focus groups, have been used to gain deeper insights when evaluating narrative visualisations [67, 136]. However, qualitative approaches, such as the aforementioned examples, are too costly and laborious to use in a real-world setting [134, 137].

A possible solution (and compromise) is heuristic evaluation, which is described as being less costly, easy to implement and relatively fast compared to qualitative end-user study methods [134]. Introduced by Nielsen, heuristic evaluation can be used during all phases in the development of a product [134]. There have been several attempts at visualisation specific heuristics but none take into account the narrative characteristics particular to narrative visualisation [70, 181, 203]. In this paper, we propose a preliminary heuristic framework that can be used to specifically evaluate the narrative characteristics of this form of visualisation. Our set of heuristics aims to be general enough to be used for any narrative visualisation and is easy to understand and apply to promote wide use and reuse.

To develop our framework, we investigated the practices of narrative visualisation practitioners. The goal is to understand how these professionals conduct an evaluation *in situ* or 'in the wild'. The understanding gained can provide valuable insights into the

practical activities of developing and evaluating narrative visualisation in practice. To do this, we first conducted a survey with 63 narrative visualisation practitioners. The participants were recruited from the online forum named 'The Data Visualisation Society.' It is an active practitioner forum that includes visualisation practitioners from all genres of visualisation development, including narrative visualisation. Then, we conducted one-to-one interviews with a smaller group of 12 practitioners.

Our primary contribution is a heuristic evaluation framework that has detailed usage advice derived from our interviews, coupled with research literature. To conclude, we make a series of recommendations on how our heuristic framework can be practically implemented. This research bridges both the academic and practitioner communities, where we aim to better align narrative visualisation evaluation research and practice. To our knowledge, this is the first practice-led heuristic framework specific to the evaluation of narrative visualisation.

Related Work

In this section initially, we outline the measures of effective narrative visualisation. We then describe the methods with which to evaluate effective narrative visualisation. The last section motivates our approach for studying the practices narrative visualisation practitioners 'in the wild.'

Effective Narrative Visualisation

Gershon and Page were the first to note that storytelling could provide a valuable contribution to the area of visualisation [75]. Later, in 2010 Segel and Heer defined this contribution as a new genre of visualisation. They coined the term 'narrative visualisation' in their paper with the same title [169]. Fundamentally, narrative visualisation is a story "that primarily consists of visualisation steps, which can include text and images but essentially are based on data [107]." Other terms such as data-

driven storytelling and visual data stories are interchangeable, as they pertain to the goal of conveying a story based on data visualisation [9, 154].

Narrative visualisation has measures that influence effectiveness. The most prominent being that it must portray a comprehensible story. This measure of effectiveness was exemplified in one of two visualisation design principles and is termed the ‘Principle of Apprehension’, as opposed to the ‘Principle of Congruence’ [188]. The ‘Principle of Apprehension’ means that the representation of information as authored by the practitioner corresponds to the mental representation simulated by the reader. This is the primary measure of all forms of visualisation effectiveness, including narrative visualisation [45].

Another measure of effectiveness is user-experience. This measure refers to user-experience goals in visualisation evaluation outlined by Saket et al. including memorability, engagement, and enjoyment [20]. While memorability is straightforward in that it means the ability of the user to retrieve and recall information, engagement is a complex, multi-layered topic [92]. It can refer to session times and page views [33]. Conversely, user-engagement in the context of flow, can mean the resulting pleasurable emotions [180]. It has been shown that narrative visualisation can induce a state of flow when certain factors are present [124]. A flow state is where the end-user is absorbed by the narrative presented and therefore is deeply engaged. This deeper level of engagement is a measure of effective narrative visualisation.

Evoking an affective response, either pleasurable or otherwise, is crucial to narrative visualisation effectiveness [137, 169]. This response can aid problem solving and reduces boredom and frustration [148] . Design choices such as colour palette and rhetorical framing can affectively influence the reader [24, 91]. Finally, narrative visualisation effectiveness can be diminished when there is distrust from the reader due

to the poor data credibility, which can be partially solved by disclosing the data source and research methodology [143, 151].

The measures of effectiveness that define effective narrative visualisation provide the foundation for its evaluation. We now delve into the methods that seek to evaluate these measures of effectiveness.

Methods of Narrative Visualisation Evaluation

The importance of thorough and rigorous visualisation evaluation is undisputed by the academic visualisation community [136]. Conventionally, there are two methods to evaluate visualisations – inspection methods and end-user testing methods [201]. Inspection methods are examined as heuristic evaluation. End-user testing methods include representative end-users and are usually defined as either quantitative or qualitative [45].

End-user Testing Methods

Qualitative methods of end-user evaluation lend themselves to measures particular to effective narrative visualisation. This is because they provide a "richer understanding" of the comprehension and user-experience of the work [47]. Notable examples include the "walk-through" or the "think aloud" protocol and interviews [11, 137]. One study used focus groups to evaluate narrative visualisation, because they "enable us to obtain qualitative and affective information from participants easily [67]." 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 [47].

Quantitative user-data collection methods have been experimented with to evaluate storytelling in information visualisation [33]. These include session times and click-through source [33]. These user-centred metrics are directly transferred from web analytic frameworks and use low-level user-activity traces as signals which are

translated as user intentions [76]. They do not, however, provide crucial insight into user-engagement and experience. Studies examining end-user interaction with visualisation have also used eye-tracking [121]. This method of end-user evaluation is however notoriously cumbersome and alternatives solutions are advocated [101].

We have outlined here some end-user evaluation methods for visualisation and their associated challenges. The next section will move on to heuristic evaluation, which should, in theory, complement end-user testing [134].

Heuristic Evaluation

Heuristic evaluation is a common inspection methods of visualisation evaluation [11]It is described as a vital part of the visualisation practitioners' toolkit [26]. Tory and Möller in their summary of expert reviews recommend the use of heuristic evaluation for analyzing visualisation systems [65]. The recognition of the beneficial advantage of heuristic evaluation led several authors to propose sets of heuristics for visualisation. These are the 13 heuristics by Zuk et al. [203] and the 10 heuristics by Forsell and Johansson [70]. Specific to narrative visualisation, criteria for the evaluation of data-driven stories have been proposed [133]. These are a useful starting point, however, are quite high-level as they lack detail in their usage advice.

A framework-based approach to visualisation specific heuristics has been advocated [181]. This approach was compared to both a '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 [181]." Similarly, Zuk et al. suggested, "A *hierarchical* or *taxonomic* way of grouping may aid in selecting an appropriate set of heuristics [203]." A 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 [134]. Outside of the cost savings, other benefits include

gaining deeper insight than end-user studies and lowering the ‘intimidation barrier’ [65, 134].

Furthermore, heuristics serve as a systematic approach to group evaluation which has been shown to be more effective than unstructured group evaluation [87]. 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 [47]

Narrative Visualisation 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. VisCOMM [191] and VisInPractice [192]). These events are evidence that professional practice as an activity with its own methods and learning are recognized as valuable to the furthering of visualisation research. This is especially so for emerging forms of visualisation, such as narrative visualisation, where an established body of knowledge does not yet exist.

We investigate the evaluation practices of narrative visualisation practitioners, because their profession necessitates them to face the challenges of evaluation, where trade-offs between time/budget impact their decisions. We motivate our approach as we believe that there is value in studying their lived experience and tacit knowledge. Ultimately, this approach will help us better understand the purpose and role of narrative visualisation evaluation in a practical setting.

Research Method

Our study consisted of four steps. Firstly, we conducted a literature review to find out our survey items and interview discussion points. Secondly, we performed an online survey

of practitioners. Thirdly we conducted a series of semi-structured interviews. Our final step, described in our future work section, will validate our framework through a validation experiment. Supplemental material can be accessed here, this includes a series of interactive charts and raw data https://effectivenv.github.io/info/

Survey Design

We chose the research method of an online survey as it suited our aim of gaining insights from a substantial user-base of practicing narrative visualisation practitioners. 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 visualisation evaluated for?

RQ2 - How is narrative visualisation evaluated?

Regarding RQ1 we asked practitioners what they deemed to be the three most important elements of effective narrative visualisation. 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 that empirically evaluated narrative visualisation and various forms of visual or verbal storytelling. We performed a search of known databases including IEEExplore, ACM, and Google Scholar. Papers were also found using a snowballing method where we researched references and well-published authors in the visualisation discipline. We carefully selected high quality papers published in top venues, which also carry out empirical studies. The primary function of our list of elements was to serve initial survey items and discussion points in our interviews. (See Appendix B for Flow chart of survey questions)

The next section, corresponding to RQ2, asked practitioners about their evaluation practices. These questions were split into inspection methods and user-study 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 [103]. These authors have written a detailed series of papers on most aspects of survey design and development. Our survey was mobile phone accessible and included a standard consent form following research ethics guidelines provided by our organisation.

Survey Participants

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

Survey Data Analysis

The survey data collected was in most instances from multiple-choice questions, which were analysed quantitatively. We used thematic analysis to extract data from the open-ended feedback question and analyse responses. Thematic analysis was also applied to the questions where practitioners could input an 'other' option. The process of thematic analysis is latent as it analyses the data through underpinning concepts and assumptions which have been appropriated for narrative visualisation practitioners [35].

For example, ‘responsive’ is one theme that is taken from web development terminology referring to the ability for the visualisation 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.

Interview Design

The survey results provided insight into the evaluation practices of narrative visualisation practitioners. This is a useful foundation for understanding what evaluation methods are used, but it does not indicate how and why they are used. Our ultimate aim was to turn our list of ‘elements’ into a detailed heuristic evaluation framework that was coupled with usage advice.

Our interviews were semi-structured and mapped out questions to the research questions that guided our survey. We extracted the interviewee’s answers to our survey as a basis for our interview.

In the opening we asked regarding RQ1; a) What is an effective narrative visualisation? b) What is an ineffective narrative visualisation? 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 visualisation would change if the work was for a different audience? g) Do you think some characteristics would stay the same regardless of the audience? About RQ2 we asked; 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?

We found our interviewees through our survey, so therefore we could question the interviewee further on 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 – 5 years of experience, as well as having contributed to the development of at least 1 – 5 narrative visualisation.

Interview Participants

We also 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. 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 each participant was emailed their transcript post interview. See Table 1 for the self-reported characteristics of participants.

Interview Data Analysis

We pilot tested our interview and clarified some of the questions as a result of the pilot. The timing of each interview was between 40 minutes to 1 hour and was performed between June 2021 and January 2022. NVivo qualitative data analysis software was used to facilitate our coding and analysis. Three researchers independently coded one transcript and then compared codes. Two researchers were students and one was a professor.

The thematic analysis was a three stage process. Each interview transcript was coded incorporating both data-driven inductive coding and a top-down a priori approach. The

result was 29 nodes or codes supported by 10 sub-nodes. After initial coding, the author re-iterated all transcripts to identify and collate themes. In the final process the author exported NVivo codebooks and interview transcripts with NVivo's 'coding stripes' from selected interviews to discuss and clarify themes with the other researchers.

Findings

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

Background Questions

The first question we asked was how many years the practitioner had worked. Most practitioners have worked for more than 1 year as a narrative visualisation practitioner. With *n*= 4 or 6% indicating they worked less than 1 year. *n*= 30 practitioners or 48% indicated they worked from 1-5 years and *n*= 29 or 46% indicated they have been working for more than 5 years.

The second question we asked was how many narrative visualisation had the practitioners worked on. Most practitioners, *n*= 41 or 65% indicated they had developed more than 10 visualisations. *n*= 12 or 19% have developed 5 - 10 visualisations. *n*=10 or 16% have developed 1 - 5 visualisations. This shows that most practitioners are well experienced usually having multiple years of experience and developing more than 5 narrative visualisations (*n*=52 or 82%).

We asked which domain practitioners worked in. The three most dominant domains that became apparent from the 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 is the most common single domain for narrative visualisation practitioners to work in. This was not entirely unforeseen, as

many of the narrative visualisations that are publicly available are derived from news outlets. The second most common domain in which practitioners worked was education, with *n*=11 or 1% of practitioners. Finally, the health domain had *n*=3 or 5%.

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

PID	Gender	Domain	Exp. (years)	#Number of narrative visualisation developed	Country
P1	M	Other (Government)	> 5	> 10	Australia
P2	M	Other (Healthcare, Business, Education, Journalism, Entertainment, etc.)	> 5	> 10	US
P3	M	Journalism	1 - 5	5 - 10	Switzerland
P4	M	Journalism	1 - 5	5 -10	Finland

P5	M	Journalism	> 5	> 10	Australia
P6	M	Other (City Planning / Non Profit)	> 5	>10	US
P7	M	Journalism	1 - 5	> 10	UK
P8	M	Other (Energy and Climate Change)	1 - 5	5 - 10	Australia
P9	M	Journalism	1 - 5	5 - 10	Greece
P10	F	Journalism	1 - 5	> 10	Singapore
P11	M	Other (Management Consulting)	1- 5	5 - 10	Canada
P12	F	Other (Business)	1 - 5	> 10	US

Table 3: Demographic information about each study participant, labelled by participant ID (PID). All are self-reported characteristics except their country, which reflects their location at the time of interview

Inspection methods

Survey Results

Most 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 is used to inspect the visualisation, *n*=37 65%, answered 'no' and *n*=20 or 35% answered 'yes.' We asked practitioners regarding their informal inspection methods (*n*=37, *m*=9, *SD*=7). Group

discussion received significantly more responses (*n*=19 or 51%). This was followed by informal conversation (*n*=11 or 30%). Informal email received the least amount of responses (*n*=5 or 13%).

For those that identified as working in the journalism, domain is was significantly more that chose 'Group Discussion' as the preferred informal inspection method *n*=6 or 75% nominating this method (*n*= 8, *SD*=2.8)

The pre-defined criteria presented to practitioners were our list of elements (See Appendix A). From the 5 practitioners that chose 'Other', 4 responses could be thematically grouped under the term 'data accuracy'.

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 'hobby projects' (P4) and therefore published outside of a professional environment.

7 practitioners we interviewed chose 'group discussion' as a preferred informal

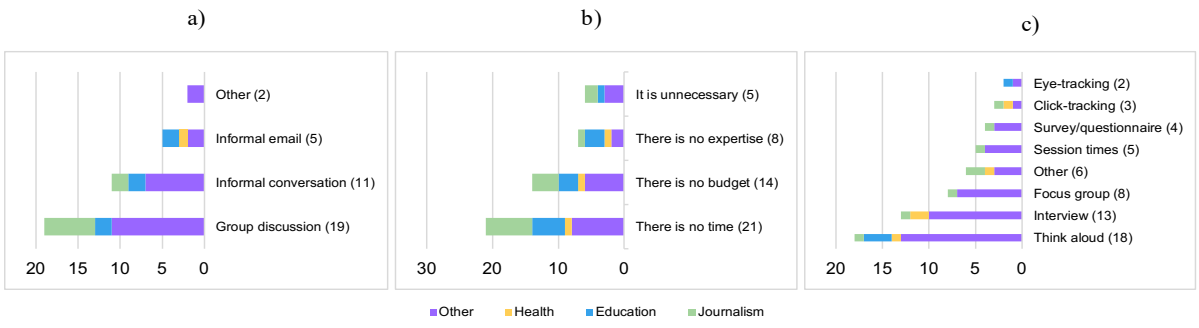


Figure 1: Survey results. a) Informal methods of evaluation using inspection methods b) Reasons for not employing end-user testing methods of evaluation c) End- user testing methods of evaluation

inspection evaluation method. We asked them why and found out that it was so that they

could access the skills and past knowledge of the team. “We have a team of people who do data visualisation daily for some years now and everyone has some experience of what works, what doesn't work” P3 or “You've got to rely on your past experience, through having expertise in the team and just general sense gathering, so 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 visualisation.

We asked practitioners why ‘data accuracy’ was added as a criterion for evaluating effective narrative visualisation and as illustrated by P12 it is an “ethical commitment,” where “integrity matters” P8. We have updated the heuristic framework to reflect this sentiment resulted in the addition of a ‘data accuracy and honesty’ heuristic in our evaluation framework.

User-study Methods

Survey Results

The responses to whether end-user testing was employed in narrative visualisation 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 test with end-users, they were then asked why there is no testing, they could select multiple reasons. Most practitioners (*n*= 21 or 41%) said there was ‘no time.’ The second most selected reason for no end-user testing was ‘no budget’ with *n*= 14 or 28%. The third most selected reason for not end-user testing was given as ‘it is unnecessary’ with *n*= 8 or 16%. *n*= 5 or 10% indicated that there was ‘no expertise.’

We asked what methods are employed when end-user testing narrative visualisation (*n*=50, *m*= 7, *SD*=6). The ‘think aloud protocol/walk-through’ method *n*= 18 or 30% was significantly higher than other methods. The second most popular method for end-user

testing with *n*= 13 or 22% are 'interviews.' Followed by 'focus groups' at *n*=8 or 13%. The only significantly low method for end-user testing was 'eye-tracking' which received *n*=2 or 4%. It is observed, 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.

Interview Analysis

We found that those working in journalism were much more likely not to employ end-user testing with 77% indicating they do not employ end-user testing. From our interviews with those working journalism, we found that for larger projects practitioners have employed end-user testing. 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 budget" P9.

We asked why the 'think/aloud' walk-through method of end-user testing was preferred. Practitioners, indeed desire to measure the emotional response of users, rather than relying on quantitative data collection methods. P1 explained that "it just captures the richness more than your digital collection method"

Social media was used as an 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. The dilemma lies in that the visualisation project has already been posted, and when asked further if they amend the project if it was criticized on social media "only if something is really wrong, that we will change it" P8.

Finally, 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 intensive user understanding” P10.

Heuristic Evaluation 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 experts to identify gaps in narrative visualisation 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 visualisation 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, user-experience, and credibility. ‘Composition’ was a suggested category name by multiple practitioners. This category encompasses the visual design aesthetic of narrative visualisation. It also includes information distribution and overall layout. ‘User-experience’ is a category that evolved from recent research in ‘user-experience focused’ evaluation [20]. Finally, ‘Credibility and trust’ is a separate category because of its unique and vital role in narrative visualisation effectiveness. There is, however, overlap between the categories and usage of each heuristic. The motivation for the delineation of each item will become clearer in the ‘Practical implementation’ section of 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 a

discussion point with practitioners and influenced our final heuristic framework. The responses for this question are thus described *n*=189, *m*=15, and *SD*=9.

Category	**Heuristic**
Composition the visual aesthetic, information distribution, and overall layout	Logical layout Information density Mindful use of colour Textual integration
User Experience how a user interacts with and experiences narrative visualisation	Cohesiveness Keep reader interest Interactivity Personally relatable content Easily recognisable content
Credibility and Trust data quality and ability for the narrative visualisation to be believed	Data source identified Data accuracy

Table 4: Proposed heuristic framework. Category with description and each Heuristics corresponding to that category

Composition

Logical layout *to not distract the reader*

In our survey *n*=33 or 17% chose this as an important heuristic for effective narrative visualisation. This is significantly higher than other heuristics. Of those that we interviewed 4 had nominated

'logical layout' as important for effective narrative visualisation. Initially, we intended this to refer to reducing the cognitive workload of the reader, which was an often mentioned key measure of effectiveness in visualisation research [37, 121, 124]. The original interpretation was summed up by P5 "logical layout, to be honest, that's just really about the good interface and visualisation design really."

In our interviews, the interpretation of logical layout usually received a broader scope beyond aesthetic layout. It was interpreted to mean creating intuitive designs that will not distract our readers out of the story. "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 and then it doesn't make any logical sense" P9.

A congruous aim to this heuristic was to minimize 'visually jarring' reader experiences. P10 asks while evaluating narrative visualisation "Is it visually jarring? Does it fit nicely? These are things that are very important to a reading experience." This finding was echoed in a paper by Brehmer et al. which studied timelines in visualisation [37]. Smooth timeline transitions that did not distract the reader by being 'visually jarring' were shown to increase effectiveness. "If you give the audience something that it's unexpected something that doesn't make sense for the human brain you will lose the attention" P9.

Information density *to guide the reader into the complexity*

It was pointed out, information-dense or complex data should not be avoided rather it must be introduced progressively. In narrative visualisation development practice, this means supplying an ‘entry-point’ to the data or another term mentioned, that has derived from user experience literature, “on-boarding the reader” P6. This process was described by P1 as “from very broad and slowly guiding the audience to more detailed information and drawdown.” In this particular case, the practitioner is echoing Shneiderman’s ‘Information seeking mantra.’ The mantra is, “overview first, zoom and filter, then details-on-demand [174].”

Through introducing the user, and guiding the user, complex information becomes accessible. P5 uses the example of the New York Times “I notice that the Times does it a lot where they will zoom in on a very little circle around a very specific part of it and explain what exactly is going on in that area.” P3 describes using a single axis as an entry-point to the data “So we can kind of enter into the more complex graphics later by starting with just one axis. And then they extend it and explain what the new axis actually means. So I'm really gently brought into the whole thing.” Another example of gently increasing the complexity of information is through textually explaining the context before introducing an information-dense visualisation. Reader reaction is described as “OK, I understand what's going on because I have the context” P7.

Mindful use of colour *regarding cultural and emotional connotations*

Colour 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 colours have within a broader societal context. “It’s very contextual and when I say contextual I don't mean in the in the sense like what kind of article you're writing but contextual in the sense of what culture you belong to” P4. “Colour is the most challenging thing in any visualisation, so you make a wrong. You choose the wrong colour and you 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' colour palette from a development tool named Flourish caused problems [69]. This was because the subject matter was taboo and the default colour palette could "re-enforce colour associations" P10. The consequence of not thinking about colour use could be regarded as disrespectful to certain ethnic groups.

When used deliberately colour association can be a powerful tool. It has been shown that colours have an affective response depending on the chosen palette [24]. This corresponds to psychology and colour theory which can we relate certain affective impressions to the properties colour palette.

Textual integration *for usability and intrigue*

Practitioners reflected on the necessity for text in narrative visualisation as a function of usability for graphics annotation without which the reader is "looking at geometric forms" P2. The textual content can provide quick access to data "to give them a brief overview of what they're seeing without delving into the real detail of it" P5.

Studies suggest that text can be as influential to reader outcomes as the corresponding visualisations [91, 105]. The title is described as being a vital component that should encapsulate the following narrative visualisation. "Whenever we design visualisation, we're trying to make the heading descriptive and conversational. We try to have it concise as opposed to having a title that's full of technical jargon. Then make sure people read the title and the subheading and can walk away knowing what the visualisation is about without even looking at the visualisation" P1.

In a study on narrative visualisation and rhetorical framing the authors asked if the practitioner intentionally incorporated "the power of rhetorical techniques" [91]. Our findings suggest, that rhetorical techniques are employed deliberately. P9 mentioned it should "allow the visualisation function as a written story." P5 explains "create a sense

of intrigue." Fundamentally, this highlights the role of narrative visualisation practitioner as a storyteller. As exemplified by P8, the text independently must tell a compelling story "like a movie script on TV or like you're watching a good TV series."

Reader-experience

Cohesiveness *because you can't visualise all the data*

A common thread in our interviews was that an ineffective narrative visualisation lacked focus. One example was when a practitioner-related a story of developing their first narrative visualisation. 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 visualisations to the readers without explaining why" P9.

The pitfalls of lack of focus are detrimental to the practitioner and the reader. The overall cohesiveness of narrative visualisation refers to the context from one data point to the next. The goal of maintaining consistency was a key principle used by practitioners. The structuring of the data into a coherent story sequence has been shown to improve the effectiveness of narrative visualisation [90]. Fundamentally, summed up by P8 "Every data point, every piece of information putting in there should be related to that issue or problem you're talking about and in the end if you're trying to draw a conclusion." Another practitioner exemplified the problem, "I find sometimes it can be hard to maintain a clear thread throughout the different sections of a longer visualisation" P6.

As was outlined by a study on the sequence in narrative visualisation, they often consist of a series of screens or episodes, rather than one stand-alone visualisation [90]. 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' [30]. 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 actually don't know what data chunking means” P11.

Keep readers interest *but first, gain their interest*

In our survey this heuristic gained significantly more responses with *n*=29 or 15%. Those that identified in the journalism domain selected this more than any other, which was expected when reader engagement is an important measure of effectiveness for this domain [74]. Furthermore, 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's you know, intriguing and makes people want to read” or as described by P11 the reader scans the article, and “That's where you could lose them, and that's where you want to catch them.”

Another practitioner mentioned how reading visualisation requires intellectual investment. “Especially when it's more complex and then for people to put in the effort they need to know that they will get something out of it” P3. The concept that the reader must ‘know they will get something out of it was outlined in an example. The practitioner explained that the reader was initially presented with an empty graph and the reader was left to think “I don't know why I should continue reading” P3. The practitioner then explained that, particularly from a “news perspective” it is important to remind readers why they should continue or they “can always jump away” P3.

A study on verbal storytelling found that when listeners listened to story it could induce a state of ‘enchantment’ which increased story memorability [179]. This is similar to narrative visualisation where readers can be induced into a state of ‘flow’ [124]. In the aforementioned study, this was achieved through page scrolling parallel to text appearance on a browser. In practice this form of narrative visualisation is termed ‘scrolly-telling [127].’‘ Scrolly-telling’ is a combination of the words scrolling and telling and appears normally on news outlet websites as long-form articles with mixed media associated with it [127].

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. If the reader has “to scroll many times to read the story” or “you don't want them to stop halfway through visualisation, and it's like, I 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. People have finite attention spans. I think that there's going to be a lot of fall-off” P11.

Interactivity *only when the reader desires to drill down*

The consensus of interactivity usage by narrative visualisation practitioners was that it is not necessary for the effectiveness of narrative visualisation. Multiple practitioners described interactivity as ‘bells and whistles’ (P5 and P12). For example “I would say only if it serves the purpose of telling the story. 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 piece has to be ‘skimmable’ P3. This means that the reader can scan quickly over the content or ‘skim’ over the content without feeling they have missed information. “Because you can really simply have any static news article without any interactivity, but still be very informative and interesting” P4. 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.”

For an expert audience interactivity was advocated. “You add more interactivity for people who are experts to drill down to understand how it's working” P4. It has been shown that interactive visualisations can facilitate data retention in readers [129].

Therefore, if a reader desires to learn more, the option should be available, however without detracting from the remainder of the narrative visualisation.

A further benefit of interactivity is that it can promote learning through fostering a sense of enjoyment and curiosity. “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.

Personally, relatable content *to reach the reader*

Evidence suggests that personally, relatable content is the primary driver for gaining and retaining reader attention [67, 143]. “If the designer couldn’t make it relatable, the piece doesn't have a chance at all to reach the intended audience” P1 or as exemplified 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 out.” The dilemma for the narrative visualisation practitioner is to engage the reader by appealing to their frames of understanding and reference.

An example put forward by one practitioner was when they needed to show air quality around the world, which is usually associated with parts-per-million (PPM). This is a data type not easily understood outside of the scientific community. “Then I found this research that said if you have that amount of PPM, that means you are smoking this many cigarettes per day. Like if I tell you that if your air quality is this, it means that you are smoking 10 cigarettes per day for one year” P4.

The thought process of taking abstract data and applying it to the relatable frame of reference is a fundamental part of narrative visualisation development and seems to be largely intuitively incorporated. Illustrated by P5 “Usually it's fairly self-evident if something is working or not.”

Easily recognisable content *for universality*

The use of imagery that is easy to recognize because it relates to everyday objects has been shown to increase visualisation effectiveness [32]. By presenting the information in visual metaphors the reader can mentally map concepts and analogies [190]. These can bridge cultural divides and can aid in universal comprehension. Easily recognisable 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. “It's very important to use real-world conventions. I think, especially with new kinds of charts and unfamiliar things, it can take people a while to understand what's going on. So to be able to route it in something that they already know, like a clock, which everyone is very familiar with, it helps them just digest it faster” P12.

The digital clock-face exemplifies easily recognisable content as it is a common visual metaphor. Multiple practitioners nominated a narrative visualisation that uses data from the ‘American Time Use Survey’, as particularly effective [197]. It simulates a working day for men and women. Each dot represents a person where they move according to their location at that time of the day. The clock-face clicks over automatically and the user can choose to pause or increase the pace or slow down the movement of dots. “Everyone including children know how to read the time. We have a mental connection from this 12-hour time format. There's a mental model already in our heads” P1. Using the widely recognized clock-face motif, complex data on American time use was easily comprehended. See figure 3 for a screenshot.

Credibility and Trust

Data source identified *it’s non-negotiable*

“It seems sensible that the ethics maintained in journalism should be upheld in visual data storytelling as well [112].” “I wouldn't even necessarily say that that's good narrative visualisation, that's just good journalism” P5. “It's non-negotiable” P2. This was the

consensus of practitioners who agreed that every single piece of work required the original data source. When possible the methodology of how the data was manipulated and presented, should be included. Practitioners described that simply linking to the data is not adequate when developing narrative visualisation, because the story is product of processed data and its' representation may not be inherently comprehendible from the source. The process is summed up with "You know, if you don't believe me here is the data, you go analyse it" P11. This sentiment is echoed in academic research wherein a study by Peck et al. it was found identifying the data source of a visualisation increased credibility with some audiences [143].

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 visualisation may look precise but isn't necessarily accurate [41]. Furthermore, due to their inaccurate data, which has been emphasized, it is particularly compelling. The challenge is then for the practitioner to represent the data without manipulating to the extent it is no longer accurate. As explained by P12 "try to show what the data 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 [105]. 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 a 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 day. There's no way you could scale that proportion. So they use the log to sort of like, you know, tie everything together. Its necessary distortion, but it's also a distortion and some people looking at that, feel that it's misleading" P10.

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 visualisation on the internet.' For the rationale behind removing 'data chunking' see previous section under 'cohesiveness.'

'Findability of the visualisation on the internet' was the least popular heuristic in our survey of practitioners n=4 or 2%. 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 visualisation does not necessarily have to have a large audience which is required to be broadcast on the internet. For example "if you're talking about a visualisation that's being developed specifically for some really small niche group of researchers or something like that, then it doesn't matter" P5. Another reason was that search engine optimization is considered not part of the development process for narrative visualisation. As suggested by P9 "this is more a marketing part so maybe I would take that out."

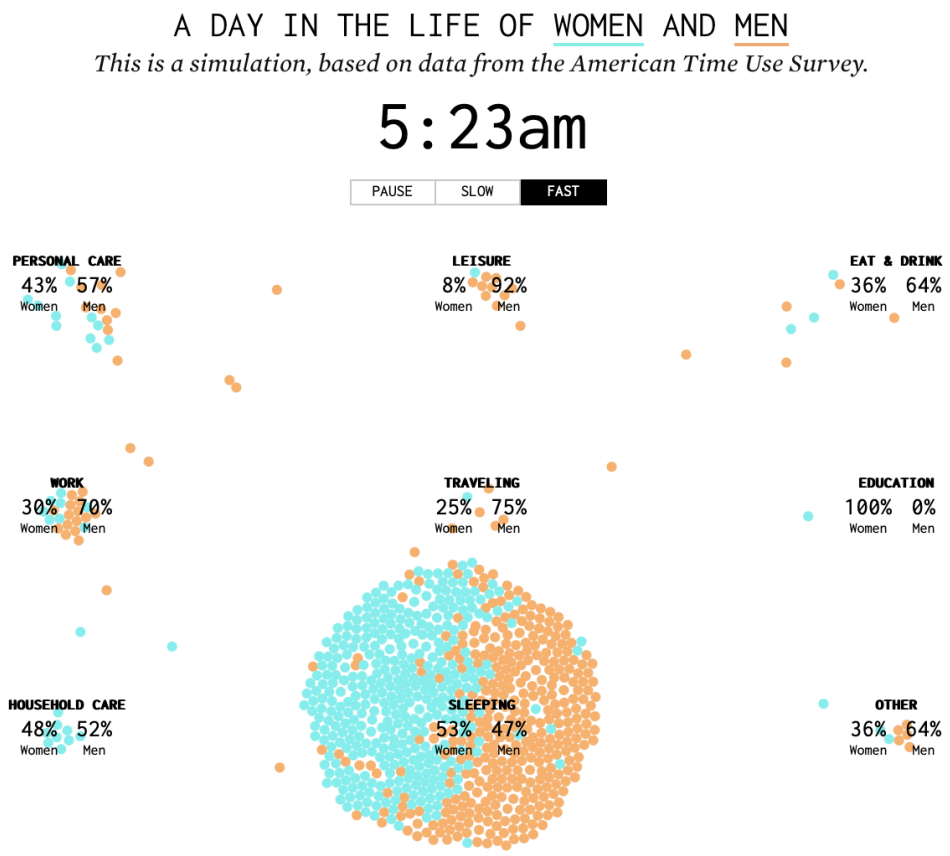


Figure 2: Screenshot from the Day in the Life of Women and Men, used with permission from author Nathan Yau [197].

Case Study: A Day in the Life of Women and Men

'A Day in the Life of Women and Men' authored by Nathan Yau, is based on data from the 'American Time Use Survey [197].' 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 visualisation was mentioned in two interviews as effective narrative visualisation. As described by P1, the "choice of using movement to represent real world activities is perfect." Considering the exemplary nature of this

narrative visualisation we deem it suitable as a case study. Categorically, we relate each heuristic to the narrative visualisation to demonstrate how our heuristic framework could be applied. See Figure 3 for a screenshot of the narrative visualisation.

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 colour'* is shown through the fact that the practitioner has used neutral colours for each gender. Stereotypical colours 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 visualisation. 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 visualisation 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 visualisation. *'Interactivity'* is used as suggested by our interviewees. Instead of being necessary for the comprehension of the narrative visualisation, 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 recognisable 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 has not only given the source of the data and the method to process it, but also tutorials and tools used.

Implementation of Framework

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 [38]. 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 visualisation recognised the value in knowledge sharing and listed some of the benefits. For example, it could “accelerate a lot of decisions” P10 or they could “make the whole process repeatable, reproducible” P1. Santos et al. extends this point, where heuristics can facilitate the replication of skills and a “common ground for the comparison among works [26].”

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 phased evaluation method as described by Nielsen [134]. Rather, our heuristic

framework should be implemented as described by Carpendale, as foundational checklist. The practitioner can keep the heuristics in mind during the development process at various phases of the project [47].

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.

Limitations and Future Work

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 visualisation. 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 selected primarily from one online community forum. The community forum ‘The Data visualisation Society’ does not include the entire narrative visualisation practitioner community and therefore may skew the results. ‘The Data Visualisation Society’ does however have a substantial user base and is as close as possible to being representative of the practitioner community.

A substantial portion (58%) 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 visualisation practitioners are rarely familiar with the term heuristics [141]. 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 [134]

Our future research will validate our heuristic framework through a validation experiment. The primary aim will be to find out whether our framework is useful and usable to practitioners of narrative visualisation. We acknowledge that many heuristic sets are not validated and it is an important step in heuristic establishment [85]. We plan to couple our validation experiment with content analysis of narrative visualisation. This will outline if the heuristics we mentioned are apparent in effective narrative visualisation.

Our research has discovered new research opportunities in the area of narrative visualisation evaluation. For example, the novel evaluation method of ‘guerilla testing’ could be an interesting avenue of research. Another interesting area could be the use of social media feedback as an evaluation tool, specifically Twitter. A revealing cross-disciplinary study could investigate the process of data abstraction into relatable frames of reference. For example, one interviewee described how abstract air pollution data was equated into cigarettes smoked per day. To make complex data accessible and compelling is no easy feat. It is clear, further studies are needed to examine this cognitive process and others shown by experienced practitioners of narrative visualisation.

Conclusion

Our study of narrative visualisation practitioner evaluation practices ‘in the wild’ found usually ad hoc evaluation approaches are employed. We propose a preliminary heuristic framework so that the evaluation process can be streamlined. Through coupling real-world practices with academic research, we introduce the foundation for further research into the best practices and heuristic evaluation of narrative visualisation evaluation.

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 Visualisation Society’ for allowing us to post our survey to their forum.

Chapter Three—Heuristics for Evaluating Narrative Visualisation: A Validation Study

Synopsis

The contents of chapter three are closely aligned with the previous chapter. It is an obvious next step to validate if the set of heuristics that were presented in chapter two. The decision to make this a separate research work was the depth of the heuristics set validation. Working with collaborators it was determined that the set of heuristics were required to be implemented in different contexts and using different scenarios. It is thus that the validation was split into two separate studies, with multiple experiments in each study. All the participants that took part in validation were experienced practitioners of narrative visualisation.

Published in the Journal of Visualisation 2025

https://doi.org/10.1007/s12650-025-01051-y

Abstract

Narrative visualisation integrates data visualisation and narrative techniques to convey a compelling story. Narrative visualisation is notoriously difficult to evaluate. One solution is heuristic evaluation, using a domain-specific set of heuristics. This paper validates a set of heuristics proposed specifically for evaluating narrative visualisation. We conducted studies with experienced narrative visualisation practitioners in both summative and formative settings. We found that the set of heuristics showed promise in a summative setting, where similar responses evidenced that the set of heuristics could provide reliable evaluation metrics. Furthermore, in a formative setting, implementing the set of heuristics was reported to be useful in the design process; however, due to their limited focus, we recommend that it be implemented in conjunction with other evaluation guidelines.

Introduction

The capacity for narrative visualisation to convey complex, important information compellingly has resulted in it being an increasingly popular communication medium. It has been employed in various contexts, ranging from explaining the extinction crisis to America's debt ceiling [7, 96]. However, the means by which practitioners can conduct rigorous evaluations of their narrative visualisation remain problematic.

Conventional empirical evaluation methods and metrics are not sufficient in evaluating the experiential nature of narrative visualisation. Narrative visualisation practitioners use narrative techniques, appropriated from classical narrative disciplines such as literature, comics, and film [10, 17, 91]. Such techniques can create rich, affective reader experiences that help create a story in the reader's mind [159]. Empirical studies are

needed in narrative visualisation evaluation as it is a distinctive form of visualisation that bridges the divide between traditional storytelling and the disparate technicalities of data visualisation.

One possible method to evaluate narrative visualisation is heuristic evaluation. First described by Nielsen in 1994, it involves usability experts who review a system and judge how well it meets the objectives of a list of predefined guidelines – heuristics [134]. The visualisation community has recognized that heuristic evaluation could be a valuable tool for evaluation [117, 61]. Indeed, various sets of heuristics have been proposed for information visualisation [118, 119, 91]. Recently, a set of heuristics specifically for evaluating narrative visualisation was proposed [63]. Unlike other visualisation heuristic sets, this set incorporates the storytelling aspects of narrative visualisation. It was created from a survey of practitioners, who were integral in their establishment. This heuristic set has yet to be assessed as advantageous to practitioners when evaluating narrative visualisation.

In this paper, we describe how we validated our domain-specific set of heuristics and outline the results of this validation. We conducted studies in both summative and formative settings to ascertain whether the heuristics could be useful in different contexts and at different stages in the development cycle. Summative evaluation is described as a ‘seal of approval’ when evaluating a visualisation [61]. In our summative study, we tasked multiple evaluators to individually evaluate a published, or a working prototype of a narrative visualisation [134] The formative study differs in the development phase. Occurring much earlier in the project cycle, the purpose of our formative study was to investigate whether the set of heuristics can assist with the early design of the narrative visualisation. In this study, we tasked practitioners to create a wireframe of a narrative visualisation. Wireframe creation is a common step in the formative stages of data visualisation development [141].

We found that when implemented in a summative setting, the set of heuristics resulted in a good level of inter-rater reliability in evaluator responses. Consistent evaluation responses meant that this evaluation method could potentially standardise narrative visualisation evaluation as it evidences a reliable evaluation method.

In a formative setting, practitioners reported the heuristic set useful. This was because it provided structure to their design process. In both studies, we found that practitioners mentioned that they would integrate other guidelines or heuristics to complement the heuristic set. They reasoned that the domain-specific set of heuristics was too narrowly focused on the story-telling aspects of narrative visualisation. With this research, we contribute an important step toward empirically validating a set of heuristics for narrative visualisation evaluation. To conclude this paper, we identify opportunities for further refinement of the heuristics and future research related to narrative visualisation evaluation.

Related Work

In this section, we define narrative visualisation, summarize recent research, and explore the challenges in its evaluation. We then describe domain-specific heuristic evaluation and its relationship with visualisation. Finally, we motivate our work by describing the importance of validating domain-specific heuristics.

Narrative Visualisation

Narrative visualisation is defined as a story consisting primarily of visualisation steps, which can include text and images but essentially is based on data [107]. Segel and Heer were the first to coin the term and map the narrative visualisation design space [169] They categorized narrative visualisation into overarching genres, such as comic strips or slide shows. Narrative visualisation research has further detailed aspects of the

design space, including story sequences, rhetorical strategies and factors that encourage reader engagement [90, 91, 124]. While delineating the unique and complex characteristics of narrative visualisation, more understanding is required in to how it can be practically evaluated.

In an attempt to capture the richness and complexity of narrative visualisation, novel evaluation methods have been investigated. Some examples of novel evaluation methods studied include elicitation interviews, focus groups, and reaction cards [67, 125, 137]. The aforementioned studies all incorporate end-users. Currently, many narrative visualisation practitioners do not employ end-user testing in their development process [63]. This is largely due to drawbacks in end-user testing such as it being expensive, laborious, and time-consuming. One evaluation method that does not include end-users is heuristic evaluation. This benefit, among others, gives us reason to believe that heuristic evaluation could be a promising solution to the challenge of practically evaluating narrative visualisation.

Heuristic Evaluation for Visualisation

Initially introduced by Nielsen, heuristic evaluation is an inspection method of evaluation [134]. Inspection methods of evaluation are a group of evaluation methods where a small number of experts inspect a user interface to discover gaps in its design. Heuristic evaluation differs from other inspection methods because it incorporates a list of guidelines or heuristics in the evaluation procedure. Nielsen described heuristic evaluation as a ‘discount’ evaluation method as it is quick to perform and does not require end-users [134]. This method is said to be particularly helpful early in the design process, however, it can be used at all stages of development. Heuristic evaluation is most effective when the heuristics are domain-specific. Similar to other domains, domain-specific heuristics are recommended for visualisation evaluation [181, 203].

Tory and Möller were the first to empirically establish that expert reviews, using a set of heuristics or guidelines, are a valuable tool in the domain of visualisation [186]. This is because heuristic evaluation can bridge the gap between visualisation design and technical implementation [128, 129]. Another benefit is that heuristic evaluation can provide structure to the visualisation evaluation process [186]. Furthermore, methodologies that adopt a heuristic approach are evidenced to be useful for visualisation evaluation. For example, the colloquially named 'ICE-T' methodology was found to be promising when valuing a visualisation [193]. Alternatively, combining heuristics with other methods such as question-based scoring, has been found to be a robust visualisation evaluation approach [84]. Another usage for heuristics is as an effective aid to educate students about the concepts and foundations of visualisation [163].

Finding an appropriate set of heuristics is one of the main challenges for visualisation evaluation. Multiple sets of heuristics exist for visualisation evaluation, each with a different intent and focus. For example, Amar and Stasko identify heuristics designed to cover known gaps in visual analytics processes [8]. To assess a visualisation's usability, Forsell and Johansson compiled a list of published heuristics and tested them on a collection of usability problems to identify 10 heuristics for information visualisation usability evaluation [70]. All aforementioned heuristics are not specific to narrative visualisation evaluation. While evaluation criteria and design guidelines have been proposed for narrative visualisation they are not designed primarily for heuristic evaluation [104, 133]. It is therefore we focus on our set of heuristics because they are specifically proposed for narrative visualisation heuristic evaluation [63].

Carpendale described multiple disadvantages of heuristic evaluation for visualisation. For example, the heuristics may be neither easily understandable nor straightforwardly applicable by practitioners who are not experienced evaluators [47]. Moreover, being

bounded by a set of rules could result in heuristics stifling the creativity of the designer [47]. Another notable issue is that domain-specific heuristics are narrowly focused meaning usability problems may be overlooked [47]. Such limitations in heuristic evaluation led Nielsen to recommend that heuristic evaluation should, optimally, complement end-user testing [134].

One barrier to the adoption of heuristics is translating the heuristic evaluation method into a realistic operational process [98]. Practical difficulties arise, such as time constraints [98]. In the early phases of development, clear objectives and variables might not yet be well defined. Effective heuristics function as a way of focusing attention on important aspects of the visualisation and saving time by adding structure to evaluation [186]. In this work, we adopt Scriven's definition of formative evaluation [168]. Fundamentally, formative evaluation means detecting deficiencies in an early to intermediate version of an application to improve its design. Carpendale described how heuristics can act as a checklist to keep in mind during the formative stages of visualisation development [47].

The goal of summative evaluation is to assess or compare a post-deployment or a near-deployment prototype [168]. Summative evaluation is necessitated in contexts such as industry awards or student marking. The results of the summative evaluation must be reliable so that they can be objective and repeatable.

Validating Domain-specific Heuristics

Validating domain-specific heuristics is crucial because, without empirical validation, weaknesses in the heuristics might not be identified. Weak heuristics can result in improper or poorly evaluated applications. Hermati et. al systematically reviewed domain-specific heuristics [85]. They found a documented lack of empirical validation in establishing a set of domain-specific heuristics.

Formalised methods for the validation of heuristics are not yet widely implemented [85]. Hermati et. al found that the most common form of validation consisted of the application of the heuristics by experts, and analysing their outcomes. Other validation methods included comparing the outcomes of heuristic evaluation to end-user evaluation outcomes. In this study, we have adopted the former method of validation. A similar method to validating heuristics was used in multiple visualisation evaluation studies [84, 193].

Therefore, our validation approach consisted of applying the heuristics by experts and analysing their outcomes. We focus on the usefulness and practicality of the heuristic set. Using realworld examples, settings and practitioners we are able to bridge the divide between academia and visualisation practice

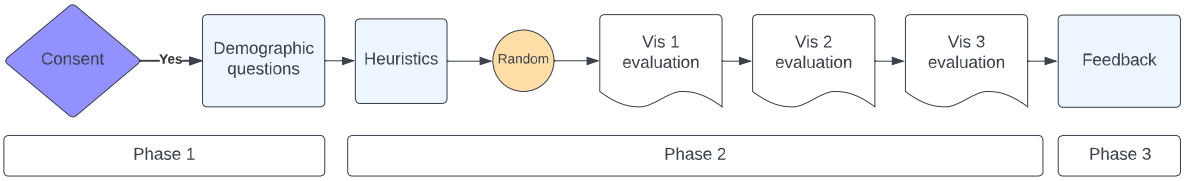


Figure 3: Procedure for summative study experiment. Phase 1 consists of consent and demographic questions. Phase 2 consists of heuristic introduction, random order of three visualisations. Phase 3 consists of qualitative feedback questions.

Research Method

This work aimed to validate a set of heuristics for the evaluation of narrative visualisation. We performed two studies in two varied evaluation settings to achieve this aim. The two evaluation settings were; summative and formative. To see raw data, please visit the supplementary material:

https://osf.io/6aytu/?view_only=33469c21cfdd40b68b14dd910c7af6a0

The Set of Heuristics

The set of heuristics that was the focus of this study was developed in three phases [63]The first phase focused on finding empirical research in narrative visualisation and storytelling. The concrete findings from the reviewed research were then synthesized into a set of foundational topics featured in a survey of practitioners. The results of the survey then influenced a series of interviews where the set of heuristics were discussed. This iterative process of developing a set of heuristics is as described in literature on visualisation evaluation [47].

The heuristic set was minimal in its design and similar to the original usability heuristic set described by Nielsen [134]. A challenge in establishing a set of heuristics is that it must be specific in its instruction yet simultaneously general so that it can be applied universally. The set is split into three over-arching categories; ‘composition’, ‘reader experience’ and ‘credibility and trust.’ The ‘composition’ category relates to the visual aesthetic, information distribution, and overall layout. The ‘reader experience’ category relates to how a reader interacts with and experiences narrative visualisation. Finally, ‘credibility and trust’ relates to data quality and ability for the narrative visualisation to be believed. In order to perform a validation on the set of heuristics we adapted it into a series of questions. Similar adaptations were successful in evaluating programming languages [160]. Please see Table 5 for the set of heuristics adapted into questions.

Composition	Does the narrative visualisation flow logically?
	Does the narrative visualisation slowly guide the reader into complexity?
	Do the colours of the narrative visualisation have any unwanted cultural or emotional connotations?
	Does the text, specifically the title, aid the reader in understanding the narrative visualisation and intrigue the reader to read further?
Reader Experience	Does all data visualized relate to the focus of the narrative visualisation?
	Does the narrative visualisation retain interest without being too long?
	Does the narrative visualisation function without the need for interactivity?
	Does the narrative visualisation personally relate to the reader's frame of reference?
	Does the narrative visualisation use recognisable universal metaphors?
Credibility and Trust	Does the narrative visualisation identify a data source?
	Does the narrative visualisation present the data in a confusing manner that could lead to misinterpretation from the reader?

Table 5: The set of heuristics for evaluating narrative visualisation adapted into questions. The first column contains the question ID, the second column contains the upper-level category, the third column contains the questions

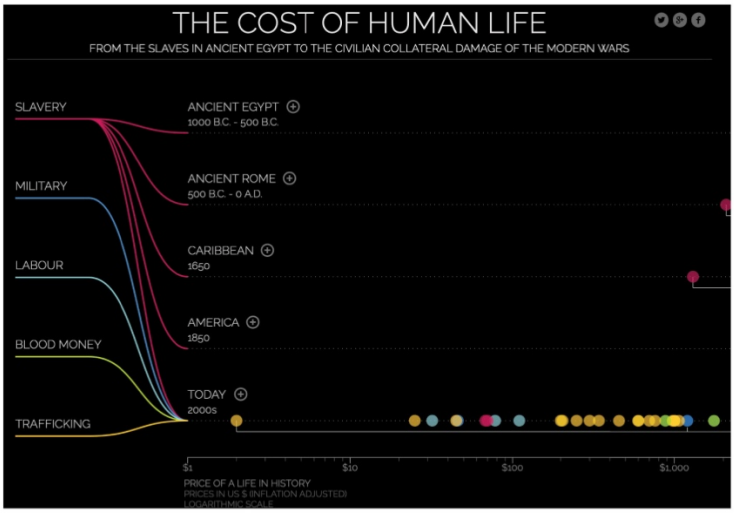
Study One - Summative Evaluation

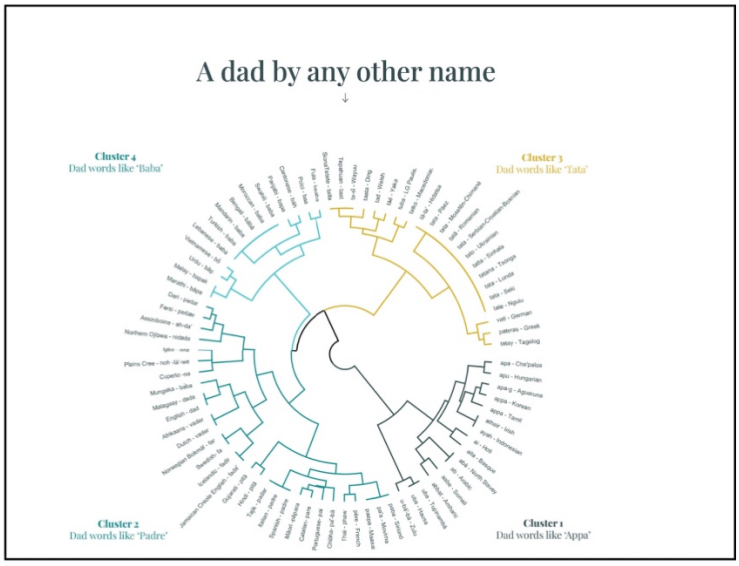
The aim of this first study was to ascertain if a set of heuristics could be used to evaluate narrative visualisation with reliable outcomes when evaluating a post-deployment or a near deployment prototype. We also asked practitioners about the usefulness and feedback on the set of heuristics.

Procedure

This experiment consisted of three phases. The introduction phase, the evaluation phase, and lastly the qualitative feedback phase. See Figure 1 for a diagram of the experiment procedure. Our survey instrument was hosted on the Qualtrics survey platform. During the introduction phase, the participant first was presented with a consent form, where ethics information was provided. If the participant consented they would then be asked demographic questions including, gender, years worked, and designation. After completing the demographic questions, the participant was presented with a set of heuristics adapted into questions. Please see Table 1 for the questions that were presented in the introductory phase, and also used during the evaluation phase of the experiment.

The second phase, the evaluation phase, consisted of presenting the participants with three separate narrative visualisation examples. We used a pseudorandom order of visualisations to minimize potential ordering effects. There were feedback questions were not mandatory to answer.


THE COST OF HUMAN LIFE
FROM THE SLAVES IN ANCIENT EGYPT TO THE CIVILIAN COLLATERAL DAMAGE OF THE MODERN WARS
SLAVERY
MILITARY
LABOUR
BLOOD MONEY
TRAFFICKING
ANCIENT EGYPT
1000 B.C. - 500 B.C.
ANCIENT ROME
500 B.C. - 0 A.D.
CARIBBEAN
1650
AMERICA
1850
TODAY
2000s
$1
$10
$100
$1,000
PRICE OF A LIFE IN HISTORY
PRICES IN US $ (INFLATION ADJUSTED)
LOGARITHMIC SCALE


A dad by any other name
Cluster 4
Dad words like 'Baba'
Cluster 3
Dad words like 'Tata'
Cluster 2
Dad words like 'Padre'
Cluster 1
Dad words like 'Appa'

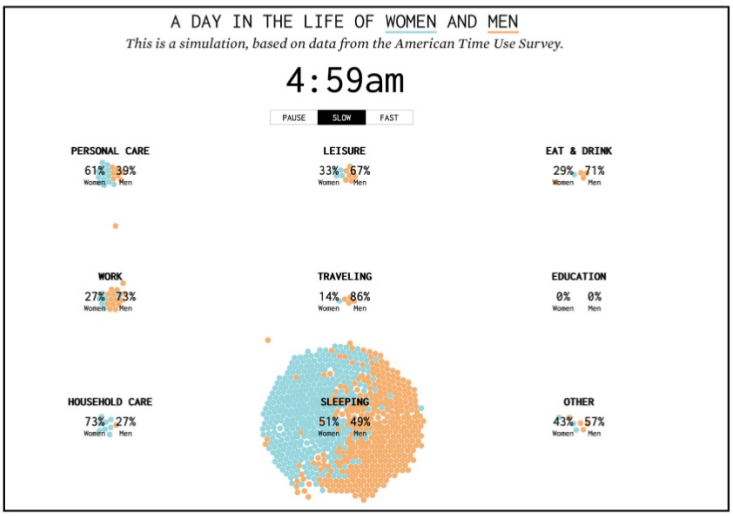

A DAY IN THE LIFE OF WOMEN AND MEN
This is a simulation, based on data from the American Time Use Survey.
4:59am
PAUSE
SLOW
FAST
PERSONAL CARE
61% 39%
LEISURE
33% 67%
EAT & DRINK
29% 71%
WORK
27% 73%
TRAVELING
14% 86%
EDUCATION
0% 0%
HOUSEHOLD CARE
73% 27%
SLEEPING
51% 49%
OTHER
43% 57%
Women
Men

Figure 4: Three narrative visualisation examples used in the evaluation stage of the summative study. From the top; Visualisation A by Carlo Zapponi “Has the cost of a human life changed?” [199], Visualisation B by Nathan Yau “A Day in the life of Men and Women” [197], Visualisation C is an example of ineffective visualisation devised by the authors of this paper.

Materials

Three narrative visualisations were selected as examples in our study. We name each; visualisation A, visualisation B, and visualisation C. We will explain our reasoning for selecting each visualisation below. All visualisations were used with permission from their authors.

visualisation A won the Information is Beautiful Challenge award for best interactive visualisation. Titled “Has the cost of a human life changed?” by Carlo Zapponi, it compared the cost of human life over the centuries [199]. Visualisation A could be described as a ‘martini glass’ structured narrative visualisation [17]. This is where the narrative visualisation is structured linearly and then concludes with an explorative, interactive visualisation on the last screen. This structure is quite common for narrative visualisation [169]. Due to the award-winning status of this narrative visualisation, it served as an exemplary example for this study.

Visualisation B differs from visualisation A in structure and topic. It was selected because it was deemed effective by narrative visualisation practitioners [63]. Based on the American Time Use survey and titled “A day in the life of women and men” by Nathan Yau [197]. It does not require scrolling and is animation based. It was deemed particularly effective by practitioners due to multiple factors, including the lack of scrolling, the unique animation and the interactive element, which gave the control over the animation to the reader. As observed by Segel and Heer, most interactive narrative

visualisation is a hybrid of both read-driven and author-driven content, which is illustrated succinctly in Visualisation B [169].

Visualisation C was developed by the authors of this paper as it is not ethical to use a published example as an ineffective example. Published narrative visualisation is usually attributed to a particular author where evaluating it as ineffective would be to the detriment to the author. Therefore to evaluate an ineffective example, it was necessary to develop an ineffective example. The ineffective example was based on characteristics of ineffective narrative visualisation described by practitioners [63]. Some of these characteristics are included in Visualisation C. Characteristics of ineffective narrative visualisation contained in Visualisation C included a particularly complex chart, namely a circular dendrogram, at the opening of the narrative visualisation, an inappropriate title and a lack of cited data source. Visualisation C is an example of a 'scrollytelling' narrative visualisation [170]. Scollytelling means that through the reader scrolling the story is incrementally revealed in a series of screens. Due to the structured, linear reader experience, it is an author-driven narrative visualisation. See Figure 4 for selected screenshots of each visualisation.

Participants

We emailed 48 practitioners of narrative visualisation. These practitioners were recruited from an active practitioner forum called the Data Visualisation Society. Each practitioner had indicated that they are a contributor to narrative visualisation in their personal introduction to the forum. We received 28 responses. From the 28 that started the survey a further 11 did not complete the experiment. The final total equalled 17 participants, who had completed the entire experiment. Further details on the challenges of recruiting and requesting time from professionals in narrative visualisation can be found in the 'Limitations' section of this chapter.

Of those who completed the experiment, 4 were female and 13 were male. No participants indicated they were binary/third gender or preferred not to say. We asked participants how many years they had worked with narrative visualisation. The largest group was 8 participants, who indicated they had worked with narrative visualisation for 10 years or more. We asked participants what their primary designation was. 5 participants indicated they worked primarily as a practitioner, 2 as designers, 5 editors and 5 as 'other.' Some other designations included analyst, consultant and design director. For the remainder of this work, we refer to the 17 participants in the summative study with a P1- ID number.

Study Two – Formative Evaluation

Heuristic evaluation is reported to assist in all phases of development. The aim of our second study was to find out if the set of heuristics could aid in early development of narrative visualisation. Normally in early phases of narrative visualisation development low-fidelity wireframes are produced, these serve to support later development and their audience is usually restricted to the development team. In this study, we investigate if implementing the set of heuristics could improve design. At this early stage, improved design is difficult to quantify so we relied on feedback of practitioner participants.

Procedure

We ran our two separate experiments with similar procedures. The first experiment was conducted with six practitioners, who were split into two groups. Each group contained three practitioners. We approximated that the skill sets of group members were equally distributed. Therefore each group contained a designer, editor, and data specialist. It has been observed that a group discussion is most effective when group participants have clearly allocated skills to contribute to the discussion [38].

One group was deemed the control group (condition 1). The control group did not use a set of heuristics to evaluate their prescribed narrative visualisation task. This group was given 10 minutes, where the scenario was explained, their expected outcome described and any questions availed. Then they were allotted 30 minutes to complete the task described in the task section of this document. The second group followed a similar procedure, with the inclusion of an intervention (condition 2). They were presented with the set of heuristics which was described in detail for 5 minutes before their allotted 30 minutes of time began. Once the allotted time was complete we questioned the practitioners on their experiences. Finally, we followed-up shortly after the experiment asking the practitioner to rate their 'satisfaction with their design.'

The second experiment followed a similar procedure to the first. It differed in having only two highly-experienced participants and a different task. Similar to experiment 1, experiment 2 gave one participant the set of heuristics (condition 2) and one participant was deemed the control (condition 1).

Intervention

To implement the set of heuristics as an intervention in our experiment conditions, we adapted the heuristics into a series of questions. Their instruction was to implement them as a checklist. See Table 5 for the heuristic questions adapted for the experiment.

Experiment One

Task

We developed a hypothetical narrative visualisation development scenario. The scenario was aimed to resemble a task the practitioners faced in their day-to-day work. We determined an appropriate task from an examination of their previously published visualisation work. The data that the task was based upon is freely available. If the

practitioners desired to use imagery they were encouraged to source the images from iStock photo so that we could purchase their copyright.

The task for experiment 1 (E1) was as follows. A not-for-profit organization named ‘Dementia Care Singapore’ has requested a narrative visualisation that informs Singaporean middle aged audiences about dementia. Dementia is on the rise in Singapore, due to its aging population. Middle aged Singaporeans should act imminently to reduce the risk of being diagnosed with dementia. ‘Dementia Care Singapore’ would like to bring attention to the 5 health considerations to reduce the risk of dementia.

Participants

We recruited 6 practitioners from a professional data-storytelling studio. The practitioners retained varied levels of experience and expertise. All participants identified as female. For the remainder of this work, we refer to all participants in the formative study with a P2- ID number. Three participants were in the control group that were not given the set of heuristics and three participants were in the group that were given the set of heuristics.

Experiment Two

Task

Similar to the first experiment, we devised a hypothetical scenario which resembled a task the practitioners might face in their day-to-day work. The scenario was as follows.

An Australian government agency named Cyber.gov.au has requested a narrative visualisation that informs an older audience about online scams. Scammers are using more sophisticated techniques than ever before. Cyber.gov.au desires that older Australians comprehend the magnitude of the threat and also learn the ‘impersonation

red flags.' These are a set of guidelines that can help older people determine the legitimacy of an email, text message or phone call.

Participants

We recruited two highly experienced data visualisation experts. Each had 11 years of professional experience as data visualisation specialists in a range of contexts, including government, finance and health domains. One identified as female and the other identified as male. The male was given the set of heuristics before and the female practitioner was deemed the control, and not given the set of heuristics.

Results

Study One – Summative Evaluation

Participant responses

We compiled the responses for each heuristic categorised by visualisation. We aggregated the responses of participants to determine group average and if there were any outliers. Participant responses were generally consistent with the group average. After performing the inter-quartile range rule, one extreme outlier was determined. P1-4, who submitted the lowest aggregated responses at 83. We did not remove the extreme outlier from the dataset as after analysis the participant gave valid responses in-line with other participants, albeit somewhat lower.

The evaluation phase of the survey consisted of 11 heuristics questions per visualisation. Therefore there were 33 evaluation questions in total. The possible aggregated range of responses is 132, where if a participant gave a 1 to each question, the lowest possible score is 33, or if they gave a 5 for each question the highest possible score is 165. Our

reported range of aggregated responses equalled 49, the lowest score was 83 and the highest was 132.

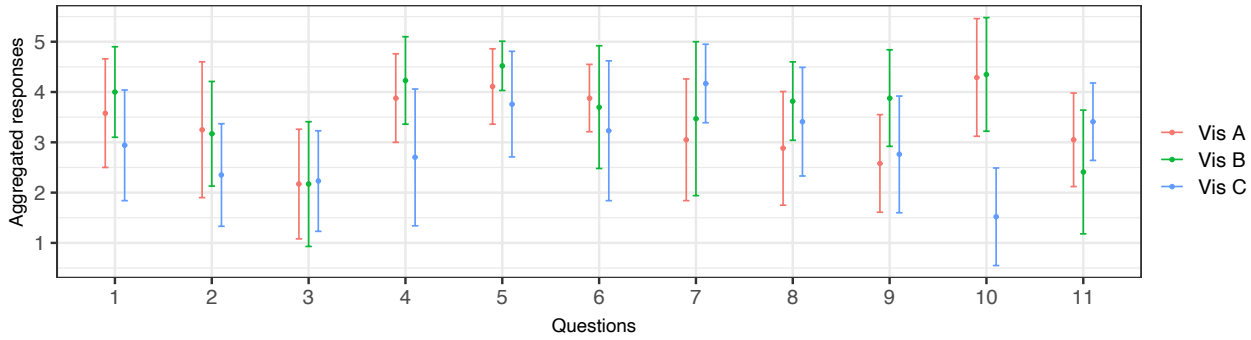


Figure 5: Means comparison chart for each heuristic question, grouped by visualisation example, with aggregated responses on y axis, bars representing standard deviation per heuristic question

Inter-rater reliability

We tested inter-rater reliability using the intra-class correlation coefficient (ICC). A high degree of reliability was found between the participant responses. The average measure ICC was .88, t(32)= 8.70, p<.001. With a confidence interval of 95%=.82-93. The ICC .88 indicates a 'good reliability' as the ICC value is between 0.75 and 0.9 [106]. The 'model' used was the 'Two-Way Random-Effects Model,' to generalize results. This is providing participants shared similar characteristics, such as professional expertise in contributing to the development of narrative visualisation. The 'type' we selected was the 'mean of k raters' due to the multiple participants that took part in the survey. Finally we chose the 'definition' of 'absolute agreement' as that suited our aim finding reliable responses to the heuristic questions posed.

We also calculated inter-rater reliability at the individual heuristic level to assess whether the participants' responses were more reliable for some heuristics than for others. The analysis revealed that a good level of reliability was found for each heuristic outside of

question 7. Question 7 was ‘Does the narrative visualisation function without the need for interactivity?’ and it received a ‘moderate reliability’ score[106]. It’s average measure of ICC was .67, $t(32)= 3.16$, $p<.056$. See Figure 5 for the standard deviation for each question. This chart illustrates that question 7 had the highest level of deviation and therefore displays the lowest level of inter-rater reliability.

Qualitative Feedback

The last phase of the survey asked feedback questions. 13 of 17 (76%) participants indicated they ‘somewhat agreed’ that the set of heuristics were useful for evaluation. A further two indicated ‘they neither agreed nor disagreed’ that the set of heuristics was useful, while two ‘somewhat disagreed’ that the set of heuristics was useful.

We asked if the participants would likely use the set of heuristics with complementary guidelines. 12 of 17 (71%) indicated they ‘somewhat agreed’ a further four, ‘neither agreed nor disagreed’ and one participant ‘somewhat disagreed.’ We asked an open text question about what other guidelines participants would use in conjunction with the set of heuristics. This question received responses that could be categorized into two major themes. Firstly responses could be thematically grouped under ‘data’ as illustrated by P1-5 who suggested “Data focus.” Further to the point that data accuracy was not addressed in the heuristic set P1-2 wrote, “Accuracy, emphasis, and message.” Another participant mentioned two specific sets that they would use, P1-11 “I might use ICE-T or Perceived Visual Informativeness.” A second theme that was mentioned by participants was ‘usability’ as described by P1-17 “A checklist that focused on visualisation best-practice, accessibility, and UX.”

We asked if there was any heuristic that the participant would change. P1-8 wrote, “I feel like the question on interactivity could be more nuanced because some of the visuals aren't really interactive.” This is in-line with the finding that question 7 received the lowest

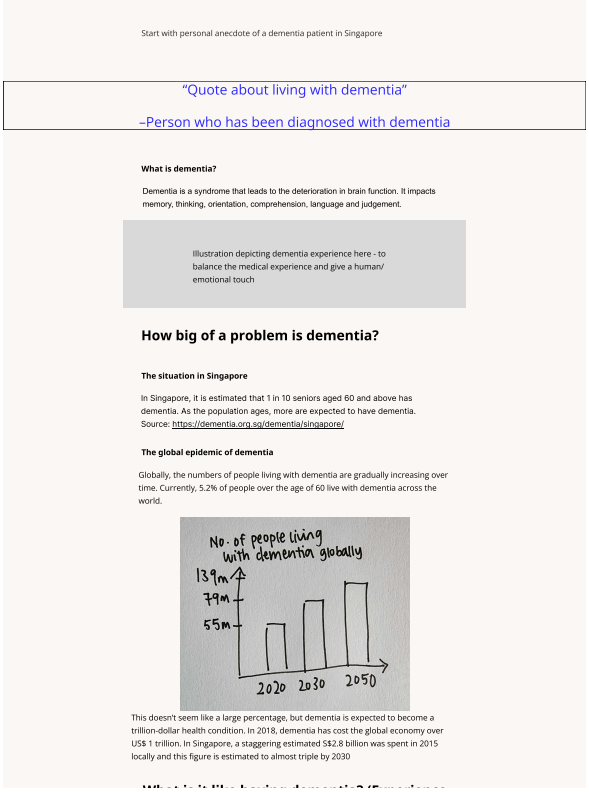
degree of reliability. This question asked ‘Does the narrative visualisation function without the need for interactivity?’ Another comment of interest was from P1-9 who wrote “The question after data and data source seems somewhat not related to the other questions. It's more of a technical/transparency thing. Of course, it is important, but maybe it should be part of a different/technical checklist.” P1-11 wrote, “It doesn't really address persuasiveness.”

The last question on the survey asked participants if there were any final thoughts on the set of heuristics. P1-17 answered “I love the idea of a checklist to evaluate the efficacy of a narrative visualisation but I feel like this list is too basic and misses some important areas of focus.” Conversely, P1-15 responded with “Conceptually, a very good approach homing in on narrative data visualisation specifically. Smart list, useful for practitioners who are either actively developing a narrative visualisation or critiquing a finished one.”

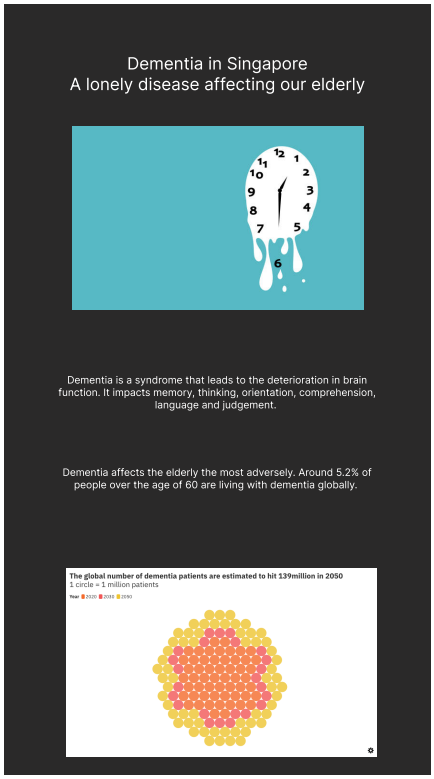
In the experiment we asked participants if they wanted to comment after each heuristic question. It was not mandatory to leave a comment and therefore often they did not. We analysed each comment that were left to gain greater insight into the cognitive processes of the evaluators. We examined question 7 as it received the lowest level of reliability. We found that there may have been some confusion in regard to question 7 which asked ‘Does the narrative visualisation function without the need for interactivity?’ For example P1-15 wrote in response to visualisation A “I assume you're referring to the chart page at the end. Interactivity was indeed necessary to advance past the sentence slides at the beginning.”

The deliberately ineffective characteristics of narrative visualisation contained in visualisation C were reflected in the comments left by participants. Question 2 asked ‘Does the narrative visualisation slowly guide the reader into complexity?’ We

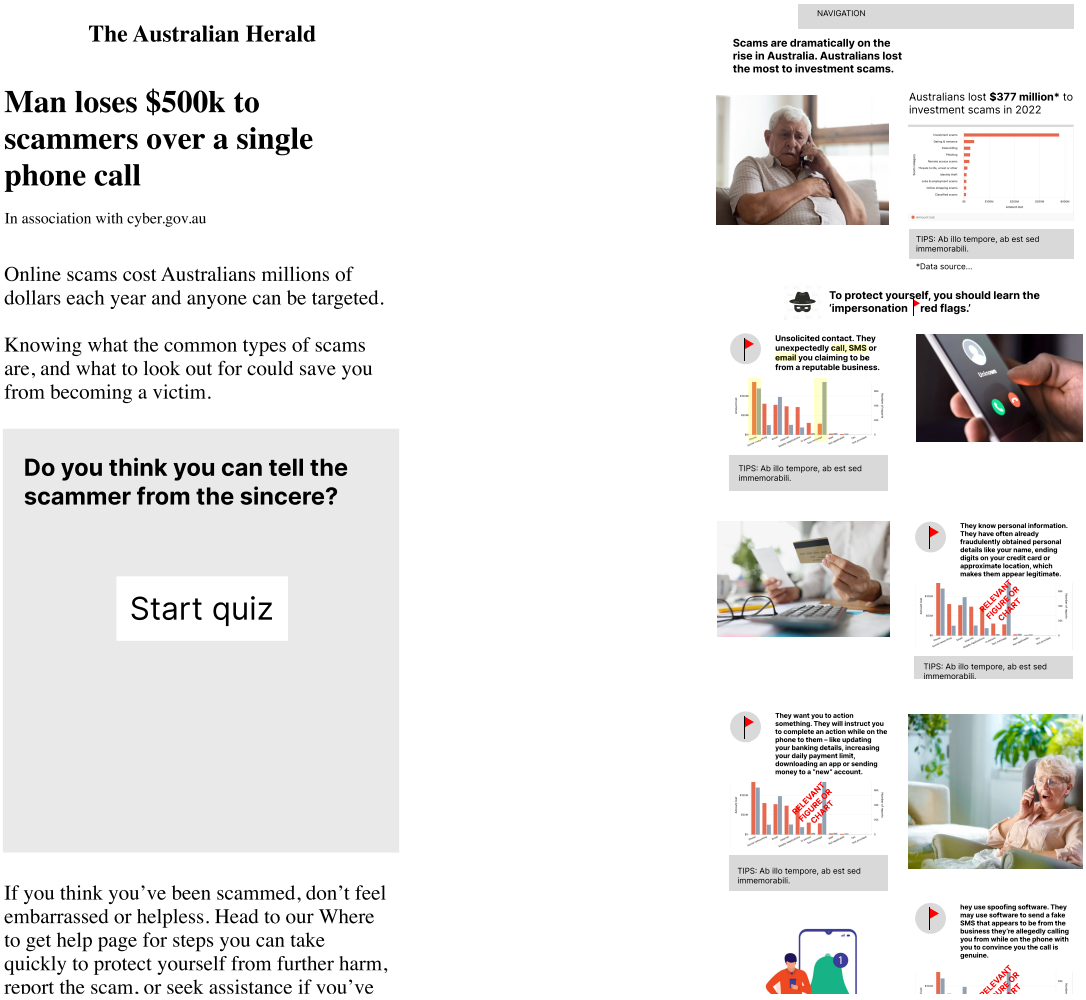
deliberately placed an overly complex chart at the beginning of visualisation C. It seems that the overly complex chart was noticed and assessed as such by participants. P1-11 commented that visualisation C “Pretty much just jumps right into the complexity and then the narrative is after.” P1-7 commented “You are given the diagram without much introduction and it is confusing.”



E1-1 (Control)



E1-2 (Heuristics)



E2-1 (Control) E2-2 (Heuristics)

Figure 6: Wireframes developed by practitioners. E1-1) Experiment 1 – control, E1-2) Experiment 1 – with heuristics, E2-1) Experiment 2 – control and E2-2) Experiment 2 – with heuristics

Study Two – Formative Evaluation

Observations

Figure 6 presents the four wireframes that the practitioners produced.

We asked participants if they found the heuristic set useful for formative evaluation. P2-7 reported "it's nice to have something formal rather than just working on intuition." P2-5 explained that "it's a very useful list that we would use. But we would add on other things that might be more applicable." We asked the participants that used the

heuristics, about their usage process. P2-7 who was part of the second experiment (E2) described that the heuristics functioned as a reference “whenever I make a design decision I can refer back to it and, ask which box does it tick?” Those from the first experiment (E1) reported to “check through the list to make sure that we had all the things were present” (P2-6). The E1 control group, reported repeatedly discussing the data and re-reading the supplied task. This indicated a less structured approach taken by the control group. Our data suggests that in both experiments, the group that used the heuristic set had a more structured approach to design and evaluation of their wireframe.

While, the extensive expertise of all participants should be noted, there is the possibility that some heuristics were missed in the design phase. We asked participants if they would have forgotten any heuristic had it not been on the list. Remembering to identify the data source was mentioned in both experiments. (P2-6) who used the heuristic set explained how they “frantically searched for the data sources, making sure we had them somewhere.” It can be observed that in the wireframes designed without the heuristic set, the data source was omitted on their respective wireframes. All participants agreed that identifying a data source is required for a published narrative visualisation.

The E1 control group’s wireframe concluded with a map where readers could search for help. In the E2 control wireframe, readers were asked to take part in a quiz. These are both examples of creative ways to interactively engage the reader. We asked why these elements were included in their wireframes. P2-8 explained why she included a quiz, “older people are the audience and everyone loves a quiz, especially them.” We asked the group from E1 why they included a map, “we wanted to say that, you can get yourself checked and there are resources to help you” (P2-2). While the wireframe that was heuristically evaluated in E1 did include a quiz-like section, audience consideration was not specifically mentioned by the group that used the set of heuristics.

Satisfaction with design

Post experiment we asked practitioners, via email, to rate their satisfaction with the design with a score of 10. Where 1 was the worst and 10 was the best. The purpose of this approach was to give participants a broad scope of responses, beyond a five-point Likert scale. We found that those who did use the heuristic set were more satisfied with their design (N=4, M=7, SD=0) compared to those who did not (N=4, M=5.25 SD=0.5).

The qualitative data supported our quantitative result, where those who used the set of heuristics were generally more positive in their responses. P2-6, who did use the set of heuristics said, “I'm quite happy with what we produced in half an hour.” The restricted time limit was mentioned by all participants when asked about their reasoning for their given rating. P2-8, who did not use the set of heuristics, reported “I needed more time.” This was echoed by P2- 3, who described their wireframe as “incomplete.”

Discussion and Future Work

Reflections on the Assessment.

Our analysis of the results suggested that the set of heuristics showed promise for the summative evaluation of narrative visualisation. In the summative study we found that evaluators provided reliable results when evaluating narrative visualisation examples. We found that there was a good level inter-rater reliability when we compared the evaluation responses from evaluators. The intra-class correlation coefficient was .88, where a value between 0.75 and 0.9 is deemed ‘good reliability’ [106]. While, there were some discrepancies on an individual heuristic level, the data suggests generally consistent results. When considering that summative evaluation is used in contexts such as industry awards, or student marking, consistent and reliable evaluation metrics mean

the heuristic set might prove a useful tool for narrative visualisation summative evaluation.

In our formative study we found that the practitioners did find the heuristic set useful. We received a generally positive response to the integration of the heuristics in the design process. We found that the heuristic set provided structure to the design process. The heuristics functioned as a reference where crucial aspects, such as integrating a data source, would be otherwise forgotten if the heuristics were not referred to. Furthermore, those that used the heuristic set in their wireframe design, reported greater satisfaction with their outcome. This suggests that the heuristic set improved design, from the perspective of the practitioner. One observed drawback of integrating heuristics in the formative stage of development, was that creative audience engagement did not seem to be equally considered by the group that used the heuristic set compared to the control group. More research is required to examine if creative audience engagement is less considered when integrating heuristics into the narrative visualisation design process.

Complementary Guidelines.

Practitioners indicated that they would likely use a complementary set of guidelines or heuristics in conjunction with the set studied here. It was clear that the practitioners desired a more ‘applicable’ set of guidelines or heuristics, which focused on aspects such as usability and data visualisation. It is not feasible that one set of heuristics covers all possible aspects of narrative visualisation evaluation, therefore we have recommended heuristics that might work well in conjunction with the set we have studied.

Usability is widely researched in relation to information visualisation. One possible complementary set of usability heuristics, tailored to information visualisation, which could potentially translate to narrative visualisation, was proposed by Forsell and Johansson [70]. Forsell and Johannson’s set of heuristics are a synthesis of multiple

heuristics that include, among others, the 10 usability heuristics proposed by Nielsen[134]. Fittingly, Forsell and Johannson's set of heuristics were analysed and recommended to not be used exclusively, as they were lacking in aspects such as aesthetics [189].

Another aspect that was not covered in the studied heuristic set, but suggested by practitioners was data visualisation. Data visualisation best practices and guidelines are many and varied. P1-11 mentioned two possible options; "ICE-T or Perceived Visual Informativeness." Firstly, Perceived Visual Informativeness, is an evaluation measure that focuses on the quality of visual evidence. It requires further research to its suitability as complementary to the heuristic set we proposed. At this point in time, it has been studied primarily in the health domain [102]. Secondly, ICE-T is a methodology, which is based on a heuristic set adapted into a series of questions similar to the approach we have taken in this work. The data visualisation emphasis of the ICE-T methodology means that it could potentially integrate quite favorably with the set of heuristics we have focused on in this work. The primary drawback is that there could potentially be some redundancies, specifically in the 'Credibility and Trust' upper-level category. This category is comparatively data-focused and could be compared to the 'Confidence' category in the ICE-T methodology list of questions.

Refinement of the Heuristic Set.

We found that there was confusion and misunderstanding concerning question 7 'Does the narrative visualisation function without the need for interactivity?' This question received the lowest level of inter-rater reliability. When we investigated the qualitative data, we found further evidence that this question caused confusion. Interaction is a notoriously ambiguous term in visualisation research. Retrospectively, we should have clearly adopted an established definition of interaction for visualisation, "where interaction is the interplay between a person and a data interface involving a data-related

intent [145].” We believe that citing this definition would have largely cleared the confusion that this question generated. For example, practitioners questioned whether simply scrolling is classed as an interaction, and, as the definition suggests, it is indeed an interaction.

Questions 10 and 11 which come under the upper-level category ‘Credibility and Trust’ garnered multiple comments from practitioners. Their appropriateness was questioned, as they are more data-focused rather than storytelling-focused. Future iterations of the narrative visualisation-specific set of heuristics could relate them to a complementary data-focused set of heuristics.

Future Research Opportunities.

A fruitful avenue for future research is further investigation into different complementary sets of heuristics for the evaluation of narrative visualisation. The ultimate goal is to determine a suite of heuristics that encompass multiple aspects of narrative visualisation, including data visualisation, and usability. A meta-analysis would verify the most suitable usage for each stage of evaluation from formative to summative stages. We predict, that such a study, would greatly benefit the design and evaluation of future narrative visualisation.

All participants who took part in this work had professionally contributed to the development of narrative visualisation for multiple years. Therefore, the results of this work are indicative of expert advice. It stands to reason that the set of heuristics would influence students of narrative visualisation differently. Rather than focusing the attention of an experienced practitioner on the important aspects of the visualisation’s design, the set of heuristics would act as a means to educate students. A similar usage has been advocated by Santos et. al [163]. Possible future work could examine if the set of

heuristics can foster design skills and provide a useful learning experience for students of narrative visualisation.

Exploratory Evaluation

The purpose of summative evaluation has been questioned, as it is produced at too late a stage to influence the design of the visualisation. It is alleged that not much is gained from simply stating one visualisation is better than another, rather an exploratory approach is recommended [61]. Exploratory evaluation asks ‘What have we learned?’ Answering such a question should not be limited to a stage in development. The participatory process to the development of our set of heuristics means, that with each summative heuristic evaluation, the practitioner acting as evaluator, could potentially reflect as to how it could influence their work. We encourage practitioners to share and evaluate the work of their peers. Through an open and supportive dialogue narrative visualisation design and evaluation will be better informed.

In this work, we contribute to the narrative visualisation research area by performing a robust and rigorous validation of domain-specific heuristics. Establishing heuristics for specific domains should not stop once the heuristics are proposed. This is an important step toward establishing a set of heuristics for the evaluation of narrative visualisation, it is, however, not the final step. Through an iterative approach, the set of heuristics will modulate over time to suit this ever-evolving research area.

Limitations

A common issue with visualisation evaluation studies is the dichotomy of localisation and generalisation. While it is not feasible to study all possible scenarios, we have attempted to use varied scenarios in different stages of development, with topically different materials. For example we chose both formative and summative settings, with different

materials in each. The threat then is that the findings in this work are from too varied scenarios to draw concrete conclusions. This is a well-known challenge in evaluation research. We have recognized the fact, and compensated where possible.

The number of participants and the variance in gender are further limitations to this study. The low participant numbers reflect the challenge of finding available professional practitioners of narrative visualisation. It should be noted other work in the visualisation research area have similar numbers of participants. Where Wall et al. recruited 15 participants and Väätäjä et al. recruited five [189, 193]. Our inclusion criteria were quite narrow, where we required a proven record in contributing to the development of narrative visualisation. This meant recruiting suitable participants was particularly challenging. Moreover, the gender disparity in both studies were not ideal. Our first study was primarily male participants, where the second study consisted primarily of females. Future work will recruit more varied participants and in greater numbers.

Allocating, even small blocks of time, from busy professional practitioners is problematic. According to feedback from practitioners, a lack of time was the primary reason so few completed the survey. We have at all points in this research, considered the time pressures felt by practitioners. We extend our gratitude to the practitioners that have given their time and valuable expertise to this research.

Conclusion

This work presents a step toward empirically validating a domain-specific set of heuristics for the evaluation of narrative visualisation. We asked practitioners to heuristically evaluate narrative visualisations, and they produced reliable evaluation metrics. Furthermore, the heuristics were reported to assist in the formative stages of narrative visualisation development. We found that the set of heuristics most likely would

need to be used in conjunction with data visualisation or usability-focused heuristics or guidelines, to encompass all aspects of narrative visualisation.

Acknowledgements

We thank all practitioners that participated in our study. We thank Nathan Yau and Carlo Zapponi for allowing us to use their work.

.

PART TWO

End-user methods of Evaluation

Chapter Four—Nudging with Narrative Visualisation

Synopsis

The fourth chapter of my thesis is a research work that was a collaboration with epidemiologists. A break from earlier chapters the second part of my thesis explores the end-user perspective on narrative visualisation evaluation. Working with epidemiologists resulted in a context for an experiment that was focused on mask wearing during the COVID-19 pandemic. Their input verified the accuracy of the data presented and the motivation for the work. The limitation was however that it was too narrowly focused on the context of the pandemic and therefore criticised for not reporting universal findings.

Published in Proc. ACM Hum.-Comput. Interact. CSCW, 2024

https://doi.org/10.1145/3686950

Abstract

The compelling aspect of effective narrative visualisation consequentially results in the potential to shift the attitude of an audience. However, there is much to understand about how narrative visualisation can be best designed and utilized to influence audiences. This paper focuses on an empirical experiment where we examined the effects of two communication strategies - anthropomorphism and personal identification - on a young adult audience. In particular, we wanted to understand which strategy, when integrated into narrative visualisations can nudge young adults' attitude towards greater consideration in the context of the COVID-19 pandemic. The results of this study contribute to better grasp of how technologies such as narrative visualisations, using different strategies, can be better designed to deliver more targeted messaging.

Introduction

Narrative visualisation is an increasingly popular communication medium. It combines data visualisation and storytelling techniques to engage readers with complex issues. For example, it has been used to explain the effects of climate change or inflation [7, 99]. Despite the widespread adoption of narrative visualisation to communicate critical information, there is limited empirical evidence about how its design can impact audience attitude. Empirical evidence suggests that narrative visualisation can persuade an audience and solidify their prior beliefs [10, 86, 114, 161]. What is lacking is empirical research into targeted communication strategies that can be integrated into the design of narrative visualisation. Through investigating communication strategies we can determine how best we can target the design of narrative visualisation to optimally impact the attitude of a specific audience.

Viewing visualisation is evidenced to have a small effect on audience attitude, which theoretically could have tangible results if extrapolated to population level [151]. In this

work we refer to a deliberate shift in audience attitude as a ‘nudge.’ Stemming from behavioural psychology, nudging is a friendly push to encourage desired behavior[183]. To our knowledge, nudging is yet to be explored in association with narrative visualisation. In this study, we examine the effects of two different communication strategies integrated into narrative visualisation, to nudge the attitude of a specific demographic.

We selected young adults as our target demographic because they are identified as a group with a propensity for low compliance with public health measures aimed at curbing the spread of COVID-19 [23, 51, 135, 158]. Several studies have demonstrated that young adults’ adherence to virus mitigation measures may be more motivated by community rather than personal preservation [39, 173]. By appealing to a young adult’s altruistic intentions, we examine the effects of targeted narrative visualisation on altering their attitude.

The first of the two communication strategies that we investigated was anthropomorphism. This strategy replaces visualized data with images of people and leverages the audience’s motivation to support their social circle and wider community. Differentiating our study from previous research that explored this topic, we have incorporated detailed imagery which reflect the target demographic and therefore appeals to their frame of reference [34]. The second communication strategy we studied is derived from game theory, specifically role enactment [144]. In this strategy we posit that interactive role enactment, will increase empathy in the target audience through a communication strategy we named personal identification. Essentially this means a viewer imagines themselves as a character in the narrative presented, where, using interactivity, they are personally involved with the narrative. Previous research has shown that role enactment does increase willingness to help in users, however this has yet to be evidenced in the context of narrative visualisation [144, 176].

In this work, we present an empirical experiment with 1084 young adult participants where we investigated nudging with narrative visualisation. Our results suggest that the personal identification communication strategy can influence the attitude of a young adult audience significantly. This is the first visualisation study, to our knowledge, that has empirically investigated how targeted narrative visualisation can be utilized as a tool for nudging. Our primary contribution is a deeper understanding of how communication strategies integrated into narrative visualisation can be used to influence attitude. Additionally, this study sheds light on communicating to a young adult demographic, and how designers of narrative visualisation can benefit from a tailored messaging approach.

Related work

We focus on related work in two areas (1) communication strategies in narrative visualisation and (2) nudging and health communication.

Communication Strategies in Narrative Visualisation

Research into narrative visualisation has grappled with comprehending and distinguishing this form of visualisation. Previous studies have mapped the design space [37, 55, 91, 169], described the design process [107, 117, 164] and examined implementations [64, 124]. While largely focusing on authoring, relatively few studies investigate the audiences' perspective. Some notable exceptions include empirical experiments on role of prior knowledge [86] and the effect of narrative visualisation structures [114]. Our work aims to highlight the important consideration of the audience. More specifically, how narrative visualisation designers might design more effectively by speaking to a particular audience.

To achieve our aim, we integrated two communication strategies into narrative visualisation. Interchangeable with the term 'communication strategy' is the term 'narrative pattern.' A narrative pattern is described as a "low-level narrative device that

serves a specific intent"[18]. Some examples of alternative communication strategies include asking the audience to 'make a guess' about visualized data or open data exploration, where the audience can explore a complex interface [18, 169]. While both the aforementioned strategies can engage the reader, they do not emotionally connect with the reader. In this study we focus on communication strategies which are theorized to induce an empathic reader response [18].

It has been suggested that data visualisation designers anchor their graphics in empathy by creating a visual connection between abstract data and living human beings[83]. One empirical study that used icons replacing abstract data was not successful at increasing a significant empathetic response [34]. Earlier studies, however, found, that compared to a chart, showing pictures of orphan children to potential charity-donors had a positive effect on donation amounts [73]. Visualisation using icons or pictographs has shown to encourage engagement, while not specifically empathy [177]. Furthermore, it has been evidenced in other areas of research such as robotics, lifelike expressive characters are more effective than icons in encouraging empathy [155]. At its most basic level, empathy means eliciting an affective response to someone else's experience [34].

We differentiate our study from previous work on anthropomorphism as we include photographic imagery that reflects the intended audience. One study which focused on anthropomorphism and graphics, termed 'anthropographics,' highlighted the lack of empirical evidence including end-users [128]. As first suggested by Morais et al. we hypothesize that photographs of *relatable* people, can elicit an empathetic response greater than that of abstract data [128]. Indeed anthropographics have been studied in relation to COVID-19 visualisation, where a lack of empirical research is emphasized [118]. This study is an effort to address the lack of empirical evidence in a newly established research area consisting of anthropomorphism and graphics. Throughout the remainder of this study, we call this communication strategy 'anthropomorphism.'

Narratives involve sharing the perspective between the reader and the featured characters portrayed in the narrative. Role-enactment occurs when the reader "imagines themselves as the characters and temporarily replaces their real-life identities and roles with the characters [144]." This cognitive process may enhance the persuasiveness of a narrative because individuals are likely to feel a sense of responsibility for the character. In the context of serious games, it has been evidenced that role enactment increases the gamer's willingness to help[144]. In the aforementioned study a game was devised concerning a humanitarian scenario, which measured willingness to help by amount which would be hypothetically donated. A further study, using the same scenario found that adding interactivity significantly increased the amount that would be donated [176].

Interactivity facilitates a reciprocal communication pathway with the narrative, where the reader is required to actively engage and identify with characters. Fundamental to the elaboration likelihood model, it is theorised that interactivity increases the persuasiveness of the message [146]. This study investigates the effects of adding interactivity to narrative visualisation, however interactivity is not the focus of our study, rather a tool to integrate the 'personal identification' communication strategy.

Nudging and Health Communication

Thaler and Sunstein introduced the concept of nudging in their book 'Nudge: improving decisions about health, wealth and happiness' in 2008 [183]. It is a concept that is based upon adaptive design of a decision environment, also known as 'choice architecture.' A nudge can consist of three heuristics that can be applied: anchoring, representativeness and availability. We will focus on the *availability* heuristic. We present an example of risk so that the audience is reminded of the risk [183]. Previous studies on the effects of nudges with a risk aversion approach in the health domain, have had positive results [46, 156].

In the context of COVID-19, the UK government actioned a specialized group of behavioural psychologists named the 'Nudge Unit [78].' The Nudge Unit, during the first outbreak of COVID-19 introduced concepts such as singing 'Happy Birthday' when washing your hands. The ethics about pushing members of the public to make choices desired by the government has been the subject of debate, where the differentiation between an individual's autonomy and 'oughtonomy' are blurred [53, 54]. Nudging is observed to be implemented for profitable gains by private companies, subverting the aim of aiding the public make healthy decisions [202]. Furthermore, if the messaging seems to impinge too much on the freedoms of the reader, it may trigger a defensive response. Known as psychological reactance, this occurs when individuals perceive that their freedom is threatened, and try to re-establish their freedom [36]. Health messages specifically can have an unintended consequence of psychological reactance, where for example, smokers counter-argued anti-smoking messages and reinforced their habit [62]. In this study we do not aim to completely change audience attitude, indeed such an aim might result in the opposite effect. Instead, our goal is finding evidence of a small to moderate attitude shift that indicates a friendly push which is representative of a nudge.

Context, Audience and Intervention

To achieve our goal of investigating the design of narrative visualisation to nudge an audience, we designed four narrative visualisation conditions on the topic of the spread of COVID-19 and the importance of wearing a mask.

Context: mask wearing to curb COVID-19 transmission

In this study we present the risk of spreading COVID-19, when not wearing a mask. Mask wearing could be considered a provocative topic. In the United States, for example, mask mandates have become a symbol for authoritarianism [13]. It is however, a convenient topic as relative to other virus intervention measures, mask wearing is

unobtrusive to an individual's daily life [68]. They are credited with globally preventing infections and ultimately, the deaths of multitudes of people [68, 89, 115]. Wearing a mask is evidenced to be beneficial in fighting respiratory infections, such as COVID-19, that are transmitted through droplets and aerosols [6, 140]. This study is focused on the United Kingdom and therefore reflects how masks are broached only in this country. It stands to reason mask wearing is considered differently in different countries.

Audience: Young Adults

Our target audience were chosen for their unique characteristics, which were highlighted during the COVID-19 pandemic. Studies have shown that outside of some anti-social individuals, young adults generally want to protect the older members of their community and the perception that young adults were deliberately violating public health measures were inaccurate [39, 135, 173]. Particularly, in the context of epidemics, young adults are shown to feel a sense of social responsibility to the vulnerable members in their community [39, 173].

One specific consideration is that the health behaviors of young adults are deeply influenced by their peers [79]. Where their perceived social norms of others are important antecedents of health-related behaviours [28]. In a study on the effects of anti-drug messaging, it was

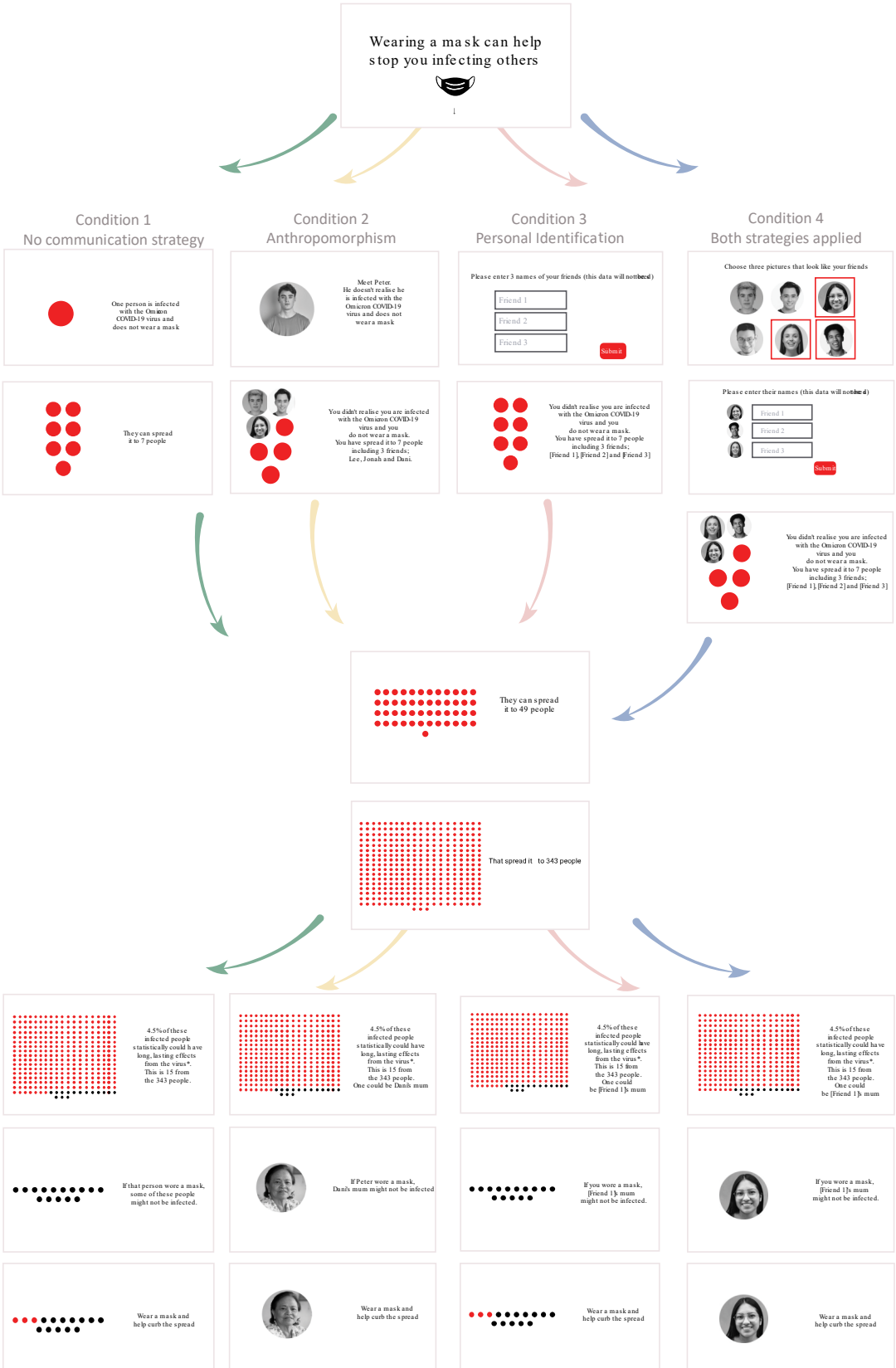


Figure 7: Flow diagram containing narrative visualisation designs of the four experiment conditions.

found adolescents were more persuaded by messaging that emphasized the social effects, than physical effects of drug use [166].

In a later study it was found adolescents were persuaded by antismoking advertising emphasizing social disapproval rather than health risks [142]. We therefore frame our message through highlighting social relationships, which are so important to a young

adult audience, with the consequences of infecting an elderly person. We theorize these aspects will speak to this demographic.

Intervention: Narrative Visualisation

This section describes the narrative visualisation variations we designed and subsequently used in our experiment. The dataset the narrative visualisations were based upon, consisted of recent research into the effects of the Omicron strain of COVID-19 and the likelihood of contracting long COVID [171, 196]. The accuracy of the visualized data was epidemiologically verified by members of the research team who are qualified in the area. It should be noted that this study was undertaken in September 2022, where the dataset presented will likely be outdated, as is the ever-changing context of COVID-19. The narrative visualisation design is an interactive article which is based primarily on scrolling, and has been observed to be popular form of COVID-19 visual communication [200].

From the four conditions, the first was a control condition, which did not contain a targeted communication strategy. This narrative visualisation opened with a title screen and an illustrated face mask. The title was adjusted for maximum impact, in-line with research that suggests that the title of a visualisation deeply influences reader recall [105]. The title spoke directly and imperatively to the audience, where it stated “wearing a mask can help stop you infecting others.” Furthermore, the colour palette selected was deliberately chosen for its specific affective response. Studies suggest that red and black

insight a ‘disturbing’ response from the reader [24]. The user was then required to scroll down and the visualized data were sequentially revealed. The user was presented with a hypothetical scenario about an individual that was infectious with the Omicron variant of COVID-19. This particular variant, at the time of writing, had an estimated basic reproduction number of between, six to eight [171]. Where a basic reproduction number denotes the expected number of cases directly generated from a single case. Accordingly, we chose a basic reproduction number of seven in the midpoint of this range to use in our communications. To express the uncertainty of the situation, we repeatedly highlighted the hypothetical nature of the narrative. The core message was that if an individual does not wear a mask, they potentially could spread the virus to a large number of people. This was illustrated in three steps, where the number of infections increased exponentially, from seven, to 49 to 343. The final screen then showed that 4.5% of those infected could have “statistically have long, lasting effects from the virus.” This is 15 people of the 343 overall amount of infected people. To add credibility and trust to the visualisation, we referenced the source of the 4.5% statistic in the narrative visualisation [196]. The 4.5% long COVID statistic is a direct finding of a large UK based study. It is, however, representative of one study only and statistics regarding long COVID may vary. To maintain consistency between conditions, the narrative visualisations that were presented in conditions two, three and four had the same title and structure as the first condition.

The second condition contained the anthropomorphism communication strategy. To apply this strategy, images of young adults replaced dots and were given generic culturally appropriate names e.g. Peter. The images featured were bought from iStock photo and copyright provided [147]. To diminish the affective response from certain colours that were not in our chosen colour palette, all images were converted to greyscale. Considerations were taken to present images of young adults that were familiar to the target demographic. In the final screen of this condition, the dots were

replaced by a photograph of a woman who was described as the mother of one of the portrayed characters.

The third condition contained the personal identification communication strategy. After the initial title screen, the user was asked to input three names of their friends. We clearly stated that the names of their friends would not be stored, and privacy concerns availed. The inputted names of their three friends would appear throughout the text of the narrative including the final screen. In the final screen the text integrated the name of one of their friends' names stating it could be this friend's mother who may be at risk of long COVID.

The fourth condition contained both anthropomorphism and identification strategies. Similar to condition three, the user was requested to input three friends' names. In addition, they needed to select images of their friends to correspond with their names. This final screen used one of the selected friend images and stated that it was their mother who was infected with long COVID. See Figure 7 for a diagram of narrative visualisation conditions.

Research Method

To test which communication strategy was most effective at nudging an audience we conducted a study using a crowd-sourcing platform. Please refer to the supplemental material [1]to see raw data and narrative visualisation source code.

Experiment Design

To measure the effectiveness of a nudge in our study we employed direct attitude measurement. While other forms of measuring affect on audiences, such as

[1] https://osf.io/fykwv/?view_only=465f6221b5c04fbdb964b3425b04098d

persuasiveness, could be used, we selected direct attitude measurement because it has a singular pursuit of measuring audience attitude [126]. In this paper we define attitude as “a psychological tendency that is expressed by evaluating a particular entity with some degree of favour or disfavor” [59]. Direct attitude measurement asks participants explicitly about their attitude toward a topic and has been widely used in psychology and the behavioural sciences [120]. This approach does, however, have a number of limitations. For example, participants might not be aware of their attitude toward a topic [77]. Another concern is the possibility of impression management, where a participant misrepresents their attitude to appear favorable to the researcher [120]. To address the former limitation we have used a universally known topic. To address the latter we have implemented an anonymous survey.

Multiple models of attitude have been proposed, some examples include; the Vector Model, the Cognitive-Affective-Conative Model and the ABC model [44, 60, 165]. Generally, all these models have similar components. The most influential of these models is the ABC model[60]. This model has three components; the affective component; the behavioural component and cognitive component. The affective component is the emotional response of the participant. The behavioral component consists of actions or observable responses in regard to a topic. Finally, the cognitive component refers to thoughts and beliefs a participant has about a topic.

In our survey we ask Likert scale questions that correspond with each ABC component, to calculate an overall attitude score. We score responses from strongly agree (+3) to strongly disagree (-3). To measure the effect of the narrative visualisation, we compare the participant’s attitude score pre-stimulus and post-

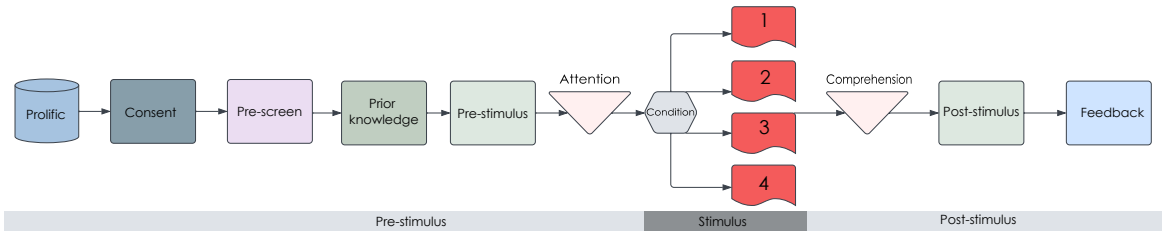


Figure 8: Flow chart of experiment procedure

stimulus. The success of the nudge is determined by the difference between the pre-stimulus attitude score and post- stimulus attitude score. If it is statistically significant compared to the other experiment conditions, it is deemed effective.

Survey Design

Our survey instrument was iteratively developed. The interdisciplinary research team had multiple discussions which resulted in three major iterations. Our primary goal was to develop a comprehensive survey instrument that would accurately evaluate audience prior knowledge and attitude. An example of where a question changed due to our research team discussions, was the affective component question which originally mentioned ‘mask mandates.’ Despite the necessity to ask an emotionally loaded, affective question, the specific term ‘mask mandate’ was deemed to have too many political connotations. Instead, the updated question referred to ‘civic duty’ which is a less politicized term.

We further refined our survey by seeking feedback from external sources. At conception stage our study was presented at an academic symposium where we asked students in the target demographic for feedback on the survey questions, experiment conditions and research methodology. In accordance with their feedback we enhanced the experiment conditions by identifying data source and re-wording questions. The next step was a small series of user-tests following the think aloud protocol, where users spoke aloud their experience while using the survey instrument. We ran our pilot on the crowd-sourcing platform Prolific, a United Kingdom (UK) based crowdsourcing platform

focusing on academic studies [149]. We allocated 10 participants per condition and determined that the survey instrument was technologically sound. Ethics approval was received from our organization.

Experiment Procedure

Our experiment procedure consisted of three phases (see Figure 2): pre-stimulus, stimulus, and post-stimulus.

Pre-stimulus

The participants were directed from the crowd-sourcing platform (Prolific), to the survey platform[149]. They were then asked to consent to the experiment, and were validated using a pre-screener demographic question. This question verified that the age of the participant was between 18-25 years. Then they were presented with three prior knowledge Likert scale questions. These questions ascertained the prior knowledge and beliefs of the participant. Through adding these Likert scale questions together, we could calculate an overall prior knowledge score. This prior knowledge score is a co-variate required for our statistical analysis. We then asked three pre-stimulus attitude questions. These questions correspond to the three components of attitude; affect, behavior and cognition (ABC). These questions were all Likert scale varying from strongly agree to strongly disagree. We calculate their pre-stimulus attitude score by adding affective question + behavioral question + cognitive question. Our ABC questions were as follows; Affective - *Do you believe it is your civic duty to wear a mask?* Cognitive - *Do you believe that long COVID is a serious risk to people in your community?* Behavioral - *How likely are you to wear a mask in a public place such as a shopping center?* Finally, before viewing the stimulus participants were asked an attention check question;

Stimulus

Participants were randomly assigned one of four narrative visualisation conditions. The narrative visualisation was embedded into Qualtrics survey platform using an iframe. The narrative visualisations were hosted on an external server and used jQuery JavaScript framework to achieve the communication strategy functionality [61].

Post-stimulus

To verify if the participant had comprehended the narrative visualisation, we asked a comprehension check question directly after the stimulus which was- *From the above data story, what is the percentage of infected people that will have long, lasting effects from the virus?* The participant was then asked the same attitude questions as pre-stimulus. The sum of these Likert scale questions determined our post-stimulus attitude score. The final question was a qualitative feedback question.

Participants

For each condition, we recruited 271 participants through a reputable UK based crowd-sourcing platform named Prolific[149]. Using Prolific's pre-screening tool, we constrained the participation to young adults between the ages of 18 – 25 years old, who were located in the UK. Using Prolific's study distribution tool we randomly allocated a balanced gender sample to each condition. This meant that the platform allocated an equal number of male and females to each condition.

Our participant sample size was a result of a power calculation with a goal of a 90% confidence level and a 5% margin of error. The population was calculated based on the young adult population size in the UK in 2022 (4.15 million [62]). Our ideal sample size was calculated at approximately 271 participants per condition. With four conditions, our total ideal sample size was calculated at 1084 participants.

Quantitative Analysis

We first analyse attitude change differentiation to determine communication strategy effect per condition. Then we analyse overall attitude shift.

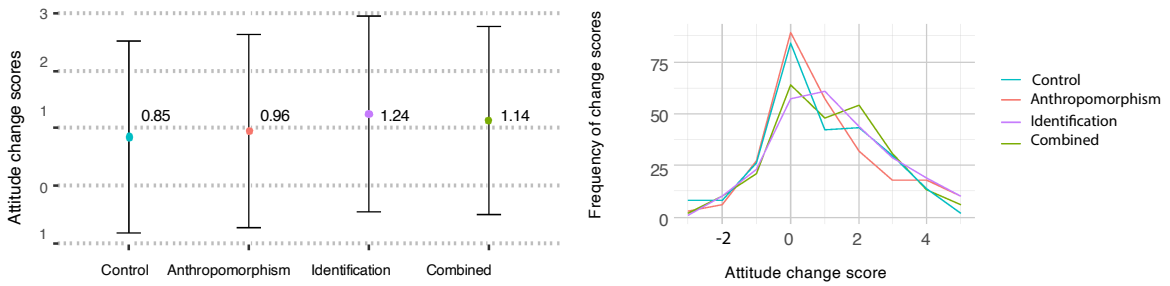
Hypotheses

We expected to observe differences between the four conditions, with respect to attitudes towards mask wearing. Specifically, our hypotheses were as follows:

H0 There is no association between narrative visualisation that integrates a communication strategy and attitude change.

H1 Narrative visualisation with anthropomorphised data, will have a higher positive attitude change in participants toward mask wearing in young adults compared to a narrative visualisation without anthropomorphised data.

H2 Narrative visualisation that integrates personal identification, will have a higher positive attitude change in participants toward mask



A)

B)

Figure 9: A) Means comparison of attitude change difference per condition showing standard error bars B) Frequency of attitude change difference per condition

wearing in young adults compared to a narrative visualisation without personal identification.

H3 Narrative visualisation that integrates personal identification and anthropomorphised will have a higher positive attitude change in participants toward mask wearing in young adults compared to a

narrative visualisation without combined communication strategies applied.

Analysis of Communication Strategies

Out of the 1084 participants (271 per condition), 18 failed the comprehension check. Our total number of valid participants was 1066. To investigate which experiment condition had the greatest effect, our dependent variable needed to be the attitude difference score. This was calculated by subtracting the post-stimulus attitude score from the pre-stimulus attitude score. Using the inter-quartile range rule, we identified outliers in the attitude difference score data. This procedure normalized the data so that the attitude score scale was between minimum -3 to maximum 5. After this procedure a Gaussian-like distribution was achieved in our attitude score data, which was assessed for near homoscedasticity. The number of participants after removing outliers was 1021.

To investigate if there was a statistical significant difference between groups we performed a one-way between groups analysis of

variance (ANOVA). The result of which showed a statistical significance with a confidence level of 95%, $F(3,1017) = 2.91$, $p=0.03$. Post hoc analysis using Tukey's adjusted post hoc test for multiple comparisons found that the mean value of attitude change score was significantly higher for *identification (condition 3)* compared to *control (condition 1)* ($p = 0.03$, 95% C.I. = [-0.77, -0.02]). Other comparisons did not show statistically significant differences in mean attitude change scores.

There was also no statistically significant difference in mean attitude change scores *control* condition and *anthropomorphism (condition 2)* condition (p=0.87) and *control* condition and *combined (condition 4)* condition (p=0.19). There was also no statistically significant difference in mean attitude change scores between *anthropomorphism* condition and *identification* condition (p=0.20) or between *identification* condition and *combined* condition (p=0.90). This result suggests that personal identification alone is the most effective at changing attitude of young adults toward mask wearing.

To factor prior-knowledge as a co-variate, we ran one-way ANCOVA (analysis of covariance). The result suggests that there is a similar effect even when controlling for prior-knowledge, $F(3,1016) = 3.03$, $p < 0.02$. The effect of the prior-knowledge on results is using Cohen's d effect size guidelines is small $\eta 2 = .00$. We can therefore conclude that prior-knowledge and beliefs do not significantly influence attitude change.

Figure 3:A shows the mean of attitude change per condition. The mean of *anthropomorphism* condition (m=.96) is only slightly higher than the mean of *control* condition (m=.85) and the difference is not statistically significant. This suggests that narrative visualisation with anthropomorphism applied is not observably more effective in changing attitude compared with a narrative visualisation that has no communication strategy applied. Therefore we cannot support **H1** according to our experiment results.

We can observe in Figure 3:A that mean difference is higher than all other conditions for *identification* condition (m=1.24). This data suggests that the personal identification communication strategy is most effective at positively shifting the attitude of participants. **H2,** is therefore affirmed due to the significantly higher positive attitude change score compared to the control condition.

The mean of *combined* condition (m=1.14) is lower than *identification* condition (m=1.24). The attitude change score is not significant between *control* condition and *combined* condition (p=0.19). We cannot support **H3** where the combined

communication strategies of anthropomorphism and personal identification do not significantly increase positive attitude scores relative to individually applied strategies. What the data suggest from this result is that combining communication strategies might slightly hinder attitude change for some participants, rather than aiding it.

Analysis of Attitude Shift

To examine if there is a significant change in attitude between pre-stimulus and post-stimulus, we performed a Paired Samples t-test on all conditions combined. We calculated pre-stimulus/post-stimulus amount by adding the ABC response questions together where each question can receive the lowest score of 6 (strongly disagree) and highest 12 (strongly agree). With a 95% confidence level, the results suggested a significant difference in the scores for pre-stimulus (M=28.84, SD=4.18) and post-stimulus (M=29.89, SD=4.34) conditions; ($p = 0.001$). The effect size is moderate, according to Cohen's d guidelines ($\eta 2 = 0.62$) and suggests that narrative visualisation does have a positive effect on attitude. This result is in-line with earlier research which showed narrative visualisation is moderately persuasive [3].

Figure 10 illustrates the frequency at which different attitude change scores appeared per condition. We can observe that most participants did not alter their attitude and therefore resulted in 0 for their attitude change score. *Identification* condition altered 61 participants' attitude by +1. This is more than any other condition. The second highest change of attitude score was *combined* condition where 54 participants altered their

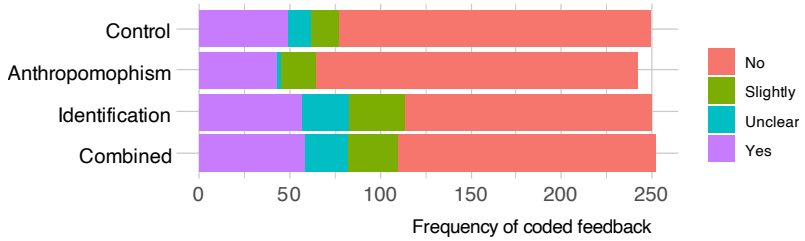


Figure 10: Stacked bar chart showing frequency of each feedback sentiment coded as yes, no, slightly or unclear, per condition

attitude by +2. This data suggests that *identification* moderately altered the attitude of many participants (61 participants +1 attitude change score) and therefore was the most effective communication strategy for nudging an audience. Meanwhile, *combined* strategies altered the attitude of the audience to a greater extent however, for a smaller number of participants (54 participants +2 attitude change score).

Qualitative Analysis of Audience Feedback

We asked participants “has your opinion on mask wearing changed after viewing the data story? Why or why not?” It was not mandatory to respond to the question and from the 1066 participants that answered the survey without failing the comprehension check, 993 wrote a feedback response. Our goal was to shed light on the cognitive or affective processes that participants drew upon when interacting with each narrative visualisation condition. To qualitatively analyse our data we adopted both a top down priori approach and an inductive theming approach [35]. We determined upper-level over-arching categories for each response, then inductively coded lower-level themes. Two coders independently coded responses and any inconsistencies were discussed. Firstly we coded each response either; yes, no, slightly or unrelated. From the 993 responses 630 said no their opinion had not changed. 207 responses said their opinion had changed. 91 reported their opinion had changed slightly and a further 65 gave feedback that did not clearly answer the question.

Positive Opinion Change

From the 207 participants that indicated their opinion had positively changed, most were from the *identification* and *combined* conditions. This is in-line with the attitude score analysis in section 5. See figure 4 for a chart illustrating each feedback response sentiment per condition. We investigated the reasoning for their positive opinion change. Through iterative analysis two primary reasons emerged; persuaded by evidence and

an empathetic response. These reasons were not always clearly delineated as often the evidence inspired an empathetic response. We therefore, grouped feedback responses on whether they largely mentioned the data or focused their own personal response to the data.

Participants highlighted how the data visualized in the narrative visualisation persuaded their opinion. Whether abstract data is presented as a dot or an image of a person, did not seem to influence the participants greatly. Some feedback responses did however highlight how data visualisation can change attitude through visually presenting data. “The information provided is much more powerful as it is presented visually with dots as a data story, rather than just paragraphs of information.” Furthermore, another participant wrote “I knew it was important to wear a mask, seeing the statistics in a digestible format made me understand the implications more.”

91 of the 207 participants directly indicated their opinion had changed due to an increase in empathy (44%). This was encouraging as increased empathy was ultimately the principal aim of the narrative visualisation. The sentiment was exemplified in this response, “it has changed because I don't want to infect anyone's parents so I will keep others in mind and wear masks more often indoors more after reading this.”

An empathetic response due to the *identification* communication strategy was frequently referred to. Participants described how they established an emotional connection with the narrative using names of friends that were close to them. For example; “I felt that using my friends names worked better, especially as the "mum" was actually a parent who has just been diagnosed with a health issue.” Another example; “The data was revealing and the connection to my friend’s life made it relatable.”

Negative or Negligible Opinion Change

We investigated the reasons why participants indicated they did not change their opinion. The most frequently mentioned reason for lack of opinion change was that the participant was already aware of the data presented. This reason accounted for 288 of 630 (44%) of participants that indicated they did not change their opinion. For example, “No, I was already taking my precautions and the story validated my decision to wear masks in public.” Thaler and Sunstein described how a nudge based on the *availability bias* operates through reminding people of true probabilities [7]. Here people assess the likelihood of risk by asking how readily examples come to mind. By presenting an example of COVID-19 spread, even if the participant already is aware of the risk, an opportunity to remind them, reinforces their belief and improves their accurate probability judgement. Exemplified in this comment “… it does serve as a gentle reminder that when in very crowded areas, especially amongst the elderly or vulnerable, that a mask may be beneficial to the wider community.”

The second most frequently mentioned reason for lack of opinion change was that the data presented in the narrative visualisation was not compelling. This accounted for 148 of the 630 (23%) of participants that indicated they did not change their attitude. For example, “Not really, the story (excluding the statistics) seems generic and made up.” Other reasons for a lack of attitude change were the prevalence of vaccinations and disbelief in the severity of COVID-19 or effectiveness of masks. One participant explained their disbelief of the severity of COVID-19, “I believe that we have to continue our lives as normal and compare COVID to a "bad" case of the flu.” These instances evidence that even a slight opinion change, for some participants, is unlikely.

In the analysis of feedback responses we found instances of reactance, for example; “I don't like the guilt tripping, I feel like I would be less likely to wear a mask.” We expected to observe this phenomena, and evidence from our data suggests it is rare [36]. We found 11 instances that suggested reactance from 993 responses (1%).

Discussion

In this section, we provide recommendations on communication strategy integration based on our study. We then move to discussing the ethical ramifications of nudging with narrative visualisation and the phenomena of reactance.

Recommendations for Communication Strategy Integration

Personal identification is the key to reaching a young adult audience. In the personal identification communication strategy we asked young adults to identify data points using names of their friends. The names of their friends then appeared through-out the narrative, placing the participant within the narrative. Our hypothesis was that young adults would be influenced by their altruistic motivations to protect their community, where through a deliberate process of role-enactment an affective, empathetic response would be evoked. The qualitative data suggested that in many instances, this was the case. For example, “I felt that using my friends names worked better, especially as the ‘mum’ was actually a parent who has just been diagnosed with a health issue.”

Other less contentious topics would result in different attitudinal change outcomes. Nevertheless, based on our results, our study joins the growing set of studies that found role-enactment results in an affective, albeit small audience response [144, 176]. However, the size of differences should be put in their context. Attitudes towards COVID-19 are often politicized and are at times, strongly held. Consequently, it is unlikely to observe large effects on these strongly held attitudes and small but notable effects, in this context, might be more important than they look. If an author decides to use personal identification integrated into narrative visualisation as a manipulative tool, it necessitates increased responsibility. Authors must be cognizant of the effects their design choices may have. Our results highlight how narrative visualisation authors could use the

personal identification communication strategy to encourage prosocial decision-making, specifically in a young adult audience.

Switching abstract data with relatable imagery is not as effective for nudging. Young adults did not change their attitude significantly when we applied the anthropomorphism communication strategy relative to other conditions. We hypothesized that relatable imagery would result in a different outcome compared to an earlier study that showed anthropomorphism to have little effect.. The qualitative data analysis indicated that anthropomorphic graphics were rarely mentioned. This suggests that switching abstract data with imagery, even when the imagery is of reflective of the audience, does not seem to greatly influence the persuasiveness of a narrative visualisation. We suspect more complex communication strategies are better suited to a young adult audience. Anthropomorphism is an example of a communication strategy that is not adequately responsive for a young adult audience.

Combining communication strategies leads to complexity and less effective messaging: One potential theory why the combined condition was less successful at nudging an audience, could be that too many communication strategies result in an opaque message. This is in-line with previous research where the cost of increased interactivity distracts the user from message comprehension[110]. When delivering a nudge, particularly when it is concerned with provocative topics such as mask wearing, we recommend an approach that is explicit in its intentions. If intent is opaque, a narrative visualisation is received with greater scepticism. This was exemplified by one participant's feedback who received the combined condition; "this just felt like more propaganda."

Reactance and Ethics of Nudging

Health communication can result in two different attitudinal outcomes: (a) agreement with the message's recommendations evidenced by a positive response, or (b) rejection of the message demonstrated by defensive reactance and denial. In our study we found evidence of reactance where participants felt that they had been manipulated by the narrative visualisation. This is not the desired response from the narrative visualisation, especially when participants stated they were less likely to wear a mask after viewing the narrative visualisation. When delivering a nudge, particularly when it is concerned with provocative topics such as mask wearing, we recommend an approach that is explicit in its intentions to avoid a defensive response. If intent is opaque, a narrative visualisation is received with greater scepticism. One potential theory why the *combined* condition was less successful at changing the attitude of the audience, could be that too many communication strategies result in an opaque message. Here, the fundamental message is obscured due to an overly complex communication approach which results in a less convincing narrative visualisation.

For 99% of those that participated in our study, no reactance was reported. Rather, participants that indicated their opinion had changed, mentioned that the evidence swayed their opinion and increased their empathetic response. Despite these seemingly desirable outcomes, it does have adverse implications. Theoretically, falsified data might have a similar result if communicated using similar communication strategies. In a study on anti-science and anti-mask COVID-19 visualisations, it was observed that often conventional data visualisation techniques were implemented [113]. These so-called 'counter-visualisations' used data visualisation to manipulate their audience to believe a message that was antagonistic to government health orders. 'Counter-visualisations' were, in some cases, quite successful at gaining traction with audiences [113]. The motivation of a nudge should be based on empirical research and expert opinion. Otherwise, outcomes may not be effective, or even may be harmful, as they are not built on a firm, scientific evidence base. Nudging an audiences' attitude could potentially be

a powerful mechanism, especially in a democratic society and we hope that our findings are implemented ethically in the future.

Limitations

Our study was limited to four experiment conditions. Each condition could be deemed not entirely orthogonal with the other conditions. This is a notorious challenge for research in visualisation evaluation, where modality and messaging must vary between conditions, however only to the extent that can be quantifiably measured. Our aim in the experiment conditions was to make each condition with as limited variation as possible. To diminish differences between experiment conditions, we deliberately did not use colour photographs in the anthropomorphic conditions. The reasoning for this was to reduce affective colour responses. Similarly, all textual integration was as closely replicated as possible between conditions, to reduce the influence of rhetorical framing [91]. Finally, the crowd-sourcing platform assigns participants randomly to each condition. For this reason some conditions were assigned participants that generally were more positive toward wearing masks before viewing the stimulus. This may have impacted our results. It should be noted, however, no one condition had a significantly higher positive pre-stimulus perception of mask wearing compared to the others.

A common problem for evaluation studies is the dichotomy of generalisability and localisation. Our chosen topic is taboo, however this is similar to other studies on narrative visualisation that have employed controversial topic such as immigration or obesity [114, 139]. It is challenging to find a topic that qualifies investigation that does not have contextual pre-conceived notions. At the time of the experiment in September 2022, the topic of mask wearing and COVID-19 was one that resonated with those on the research team, particularly those who are epidemiologically qualified. While topics are many, this particular one was where we had the expertise and motivation to focus on as our selected case study.

Optimally, we would have created many more experiment conditions where a multitude of variations could be experimented. Communication strategies are many, where we selected these two, as they were deemed most promising to achieve our study aims. It stands to reason that other communication strategies are equally as promising research avenues but were overlooked in our literature review. Furthermore, other demographics, outside of our target of young adults may benefit from the communication strategies we presented. The necessity to limit the scope of the study dictated that we focus on one demographic. Similarly, we limited our study to participants from the UK. In the early stages of this study we investigated other nationalities, however it was beyond the scope of this study to equally investigate multiple nationalities. To encourage future research, we have freely shared the raw code from our experiment conditions so that it can be adapted for other audiences.

Conclusion and Future Work

Integrating personal identification in narrative visualisation induces an affective response in a young adult audience. This communication strategy can shift the attitude of a young adult audience significantly more than if a targeted communication strategy was not applied. When we combined communication strategies it resulted in a shift in attitude, albeit with moderate success. In comparison, the identification communication strategy resulted in more widespread, attitude shift that was less dramatic. A slight attitude shift could be otherwise described as a nudge, where in this study, we gently reminded participants of the importance of wearing masks. These findings could be implemented with other messaging and have the potential to be quite a powerful mechanism. Even a small attitude shift could potentially translate to meaningful outcomes if extrapolated to a population level. Future research can expand on our results to examine alternative implementations, audiences and the ramifications of nudging with narrative visualisation.

Chapter Five— An Age-Based Study into Narrative Visualisation Engagement

Synopsis

Following the previous chapter that explored end-user evaluation of narrative visualisation, is the fifth chapter that further investigates end-user evaluation. The previous chapter targeted narrative visualisation to a single demographic, however this chapter asks if the assumptions contained in the earlier chapter are correct. The aspect of audience preference and behaviours is, an under-explored area in narrative visualisation research. Categorising an audience by age is a fundamental means of differentiating audience types. The motivation for this study was to shed light on the assumptions that are held by narrative visualisation practitioners and researchers, including myself.

Published in IEEE Computer Graphics and Applications (pre-print) 2025

https://doi.ieeecomputersociety.org/10.1109/MCG.2025.3591817

Abstract

Research has shown that an audiences' age impacts their engagement in digital media. Interactive narrative visualization is an increasingly popular form of digital media that combines data visualization and storytelling to convey important information. However, audience age is often overlooked by interactive narrative visualization authors. Using an established visualization engagement questionnaire, we ran an empirical experiment where we compared end-user engagement to audience age. We found a small difference in engagement scores where older age cohorts were less engaged than the youngest age cohort. Our qualitative analysis revealed that the terminology and overall understanding of interactive narrative patterns integrated into narrative visualization was more apparent in the feedback from younger age cohorts relative to the older age cohorts. We conclude this paper with a series of recommendations for authors of interactive narrative visualization on how to design inclusively for audiences according to their age

Introduction

Demographic characteristics such as an audiences' age can impact how they interact and engage with digital media [97, 198]. Interactive narrative visualisation, or data storytelling is a form of digital media that has been shown to compel and explain information through an engaging end-user experience [137, 157, 172]. It has thus been used to communicate critical topics such as the effects of climate change or election outcomes [66, 184]. Despite conveying important information, interactive narrative visualisation is often designed without audience age in mind. One potential reason is that creating interactive narrative visualisation is a labour-intensive process that requires

expertise and creativity [169]. It is, therefore, researchers have concentrated on aiding authors to alleviate the challenges involved with their creation. Consequently, narrative visualisation researchers have yet to fully explore the requirements of audiences of interactive narrative visualisation. The lack of research into narrative visualisation audiences has led authors to make assumptions about audience behaviours and preferences.

In this study we attempt to address a fundamental knowledge gap on the question on how audience age impacts engagement in interactive narrative visualisation. We examine age-related issues that impact engagement in interactive narrative visualisation. Moreover we qualitatively analyse why significant differences exist. We aim to give a concrete, evidence-based answer to whether authors of narrative visualisation should prioritize audience age when designing interactive narrative visualisation and how this should be done. To achieve our aim, we performed an empirical experiment. We developed three narrative visualisation examples that employed different interactive narrative patterns tailored specifically for engagement [27]. We randomly assigned 2400 participants to one of the three narrative visualisation examples and measured their engagement using VisEngage [92]. VisEngage is a self-reporting questionnaire specifically developed to measure engagement in visualisation. The age groups were split into generations. A younger audience with an age of 18-27, a middle-younger audience between the ages of 28-44, an older audience of 44-60, and finally, the oldest age cohort consisting of over 60. These age groups were determined by generation boundaries [57].

The results of our study revealed that there is a significant difference between older audiences and younger audiences' engagement in narrative visualisation. The greatest difference was between the two age extremes, where the 18-27 age cohort showed increased engagement compared to the 60+ age cohort. From our qualitative analysis,

it was found that younger audiences were more observant of the interactive techniques employed that encouraged engagement. The terminology used by younger audiences was distinctly different from their older counterparts, where they described the cognitive processes involved in their interactive engagement. Older audiences were not so discerning and reported their negative engagement was due to the interactive device causing confusion or distraction. To our knowledge, this is the first study that investigates age groups and interactive narrative visualisation engagement. We conclude the work by presenting a series of recommendations for designing interactive narrative visualisation inclusively based on our findings.

Related Work

Narrative Visualisation

Segel and Heer, who coined the term ‘narrative visualisation [169].’ They proposed a design space outlining genres and structures [169]. Further foundational research on narrative visualisation established an analytic framework by defining rhetorical techniques and possible transitions for story-sequencing [90, 91]. Recent advances in narrative visualisation authoring processes include an evaluation framework, machine-guided workflows, and generative AI co-creation systems [12, 63, 72].

While there is much research on the authoring process of narrative visualisation, its impact on audiences is relatively under-explored. When viewing visualisation, it has been found that an audience’s personal beliefs impact their viewing experience [143]. Furthermore, the story structure can influence end-user engagement in a narrative visualisation [114]. More recently, a series of ‘narrative patterns’ were observed in narrative visualisation, where the author's intent was correlated with a narrative device integrated into a data-driven story [18]. Each one has been studied and shown, with varying degrees of success, to encourage audience engagement. For example, a study

into the effects of elicitation techniques found that asking the audience to ‘Make a Guess’ about presented data in a narrative visualisation can produce feelings of ‘surprise and interest [157].’ These emotions are fundamental to end-user engagement. Another study found multiple benefits of employing a narrative pattern that speaks directly to the audience [172]. Named ‘Breaking the Fourth Wall,’ interacting with this narrative pattern increased audience self-story connection, engagement, and information recall by including the reader inside the presented narrative [172]. Lastly, data exploration was described as a narrative pattern to encourage end-user engagement; however, it has had inconclusive results. The addition of a storytelling element to an interactive exploratory data visualisation did not increase the time spent viewing the page [33]. Conversely, exploring data in a gamified context, did increase end-user engagement [55].

No singular definition of interactive narrative visualisation has been established. In this study, we use a broad, inclusive definition for interaction in visualisation, as described by Dimara and Perin [56]. Their definition is thus, “Interaction for visualisation is the interplay between a person and a data interface involving a data-related intent [56].” To avoid being too vague, we refer to interactive narrative visualisation, where an interactive modality such as scrolling, clicking, or inputting end-user-generated data is a key element of the narrative visualisation. This differs interactive narrative visualisation from narrative visualisation, such as data videos or data comics, where interactive modality is not required [9, 18]. Examples of interactive narrative visualisation are primarily published by media outlets on the internet, where technologies such as JavaScript are fundamental to their delivery. In the next section, we investigate engagement and how it is measured for visualisation.

Engagement

The human-computer interface (HCI) community has long considered engagement a fundamental concept in user-centered design [14]. Visualisation research, in comparison, has relatively recently begun to seriously regard engagement [92, 119, 162]. The definition of engagement is often ambiguous and dependent on discipline [14]. We adopt an HCI definition of engagement, which centers on the quality of the user experience and on the positive aspects of the interaction, particularly the phenomena associated with being captivated by technology [14]. O'Brien et al. listed dimensions of engagement as including aesthetic appeal, novelty, perceived challenge, feedback, motivation, and affect [138]. Alternatively, engagement can be viewed as a continuum from low to high [119]. Furthermore, engagement in narrative visualisation has been viewed in the context of flow and fluid interaction [124].

Numerous methods have been proposed to measure engagement in the field of visualisation. McKenna et al. adopted a self-reporting questionnaire for measuring engagement for a visual story experiment [124]. Boy et al. evaluated engagement by analyzing time spent on interaction and user input [33]. Time spent assessing visualisation using reaction cards has also been examined to measure engagement in visualisation [125]. Nowak et al. used elicitation interviews to examine factors including emotional affect and engagement in narrative visualisation [137]. A purpose-built method for measuring engagement in visualisation was proposed by Hung and Parsons named VisEngage. VisEngage is a self-reporting questionnaire based on the user-engagement scale adapted for visualisation [92]. Similar adaptations were successful in other domains, such as social networking applications and games [22, 195]. A questionnaire comprising 22 questions, VisEngage addresses 11 engagement characteristics, where each characteristic corresponds to two questions. VisEngage is a relatively robust method to measure end-user engagement in visualisation [92].

Age-based Research, Myths and Assumptions

Previous work on different age cohorts in visualisation usually focused on either the very old or the very young. For example, visualisation research with children has investigated pedagogical approaches for visual literacy and visualisation design [29, 93]. Research into elderly audiences has examined aspects of accessibility and comprehension [3, 43]. A lack of understanding of different age groups needs is a known and oft-cited dilemma in visualisation research [198]. Where it has been noted that in visualisation empirical research, often the age of study participants is not reported, which could skew results [40]. Relatively little research within the visualisation domain is cross-generational, where age groups are compared without focusing on either age extreme.

In the HCI field, age-based research has found significant differences between age groups. For example, the time taken to perform input modalities of users compared to the age group found older adults to be significantly different to their younger counterparts [49]. The older group was comparatively less competent when compared to the younger age group in terms of digital gameplay competency [153]. Furthermore, significant differences were found in avatar preferences compared to age group [152]. While these are evidence-based findings that differentiate age groups and their digital activities, many myths perpetuated in grey literature are not founded on scientific evidence.

In a report by the Interactive Advertising Bureau in the United Kingdom (UK) it was found that younger age cohorts find interactive advertisements more appealing than their older counterparts [94]. This claim was further documented in grey literature, which for example, reported that interactive content is “key to engaging young audiences [111].” However, in a report by the NN Group, unnecessary interactivity and flashy graphics are found to be ‘annoying’ by young adults [81]. The NN Group report explained that young adults are digital natives, who are “people raised in a digital, media-saturated world” and distinctly different to their older counterparts. Young audiences deemed ‘digital natives’

were more confident and less patient with user interfaces, according to the NN Group Report [81].

Older adults are described as wary of technology. For example, older adults are supposedly less likely to prefer gamified user experiences and prefer text-based content [5, 185]. Moreover, older adults are reported to be less confident with user interfaces and are "hesitant to explore [111]." Research into usability for older audiences is still lacking, where their preferences and behaviors are not adequately considered [111, 198]. While et al. introduced the term GerontoVis, which encapsulates data visualisation design that primarily focuses on older adults, which they describe as a largely overlooked area of visualisation research [198]. We motivate this work by assessing the aforementioned assumptions about behaviors and preferences relating to engagement and audience age, contextualized for interactive narrative visualisation.

Research Method

We conducted a crowd-sourced study using three narrative visualisation examples as a stimulus to achieve our research aim of investigating whether age impacts end-user engagement in interactive narrative visualisation. To see the interactive narrative visualisation example code and raw data, please see the supplementary material[2]

Experiment Design

One of the primary challenges when designing an evaluation experiment is the dichotomy of localization and globalization. It is important to have results that can be globalized and, therefore, universally applicable. Conversely, it is necessary to have strict experiment parameters to report concrete results. To address this challenge, we

[2] https://osf.io/6a5rp/?view_only=fc3f33f0407f4e19a14e8be6ca2e1526

developed three different narrative visualisation examples that were similar in length but varied in topic. Each example is different in its data, messaging, and interactive narrative pattern. The intent of the integrated interactive narrative patterns was, however, similar – engagement [27].

The research team iteratively developed three narrative visualisation examples. Each example was inspired by publicly accessible interactive narrative visualisations from reputable publishers. Publishers that influenced our designs include The New York Times, ABC Australia, and The Pudding [80, 82, 184]. By no means the only publishers of interactive narrative visualisation, each of the aforementioned publishers is commended in online journalism awards for their interactive narrative visualisation [15].

Design One: 'Make a Guess'

The first narrative visualisation example, Design One, used a 'Make a Guess' interactive narrative pattern [27]. This pattern encourages engagement by stimulating the curiosity of an audience [157]. The audience is asked to guess an answer to a question, and the answer to the question is then revealed, affirming or disaffirming the accuracy of their answer. The objective of the 'Make a Guess' narrative pattern is that the audience questions their perception of reality by revealing a mismatch between perception and the actual data. An example of the 'Make a Guess' narrative pattern is a New York Times story on education titled 'You Draw It: How Family Income Predicts Children's College Chances [80].'

In our study, the narrative visualisation design example, which we refer to as Design One, 'Make a Guess', was based upon a dataset from the WWF's Living Planet Report 2022 [71]. It opened by asking the participant if they could answer this question; 'What do you think is the percentage of decline of wildlife populations since 1970?' Underneath the question was a sliding bar set by default to 20% and a button stating, 'find out.' Once

the participant clicked on the button, if the sliding bar was set to any number under 69%, the participant would receive the same message - wildlife decline was more than their estimate. If they estimated above 69%, they were answered with a 'you are close.' The default sliding bar amount, set at 20%, was a deliberate ploy for the user to estimate a lower value. See Figure 11:A for a diagram of Design One.

Design Two: 'Breaking the Fourth Wall'

The second narrative visualisation example, Design Two, used the 'Breaking the Fourth Wall' interactive narrative pattern. 'Breaking the Fourth Wall' is a term often cited in cinema and literature disciplines. In interactive narrative visualisation, a direct question is asked of the audience, normally to input personal data. This creates a 'self-story connection,' which has been found to encourage engagement, as it includes the user within the story [19]. The narrative visualisation design example that inspired this study was a finalist in the online journalism awards. Published by the ABC Australia Story Lab, the narrative visualisation is titled 'See how global warming has changed the world since your childhood [184].' In our study, Design Two broke the fourth wall by asking the user to input their name. Specifically, the user was asked to 'please enter your first name (this data will not be stored)' so that privacy concerns were availed with the assurance that data relating to the user's name would not be stored. Design Two opened by stating, 'How likely are you to be murdered by a serial killer?' The user was asked to enter their first name click submit. Once submitted, a screen appeared with '[name] have you ever watched a horror movie and found yourself scared? Should you be scared of serial killers?' This is followed by a basic bar chart explaining '1% of homicides in the UK are about 5.' The narrative sequentially revealed itself as the user

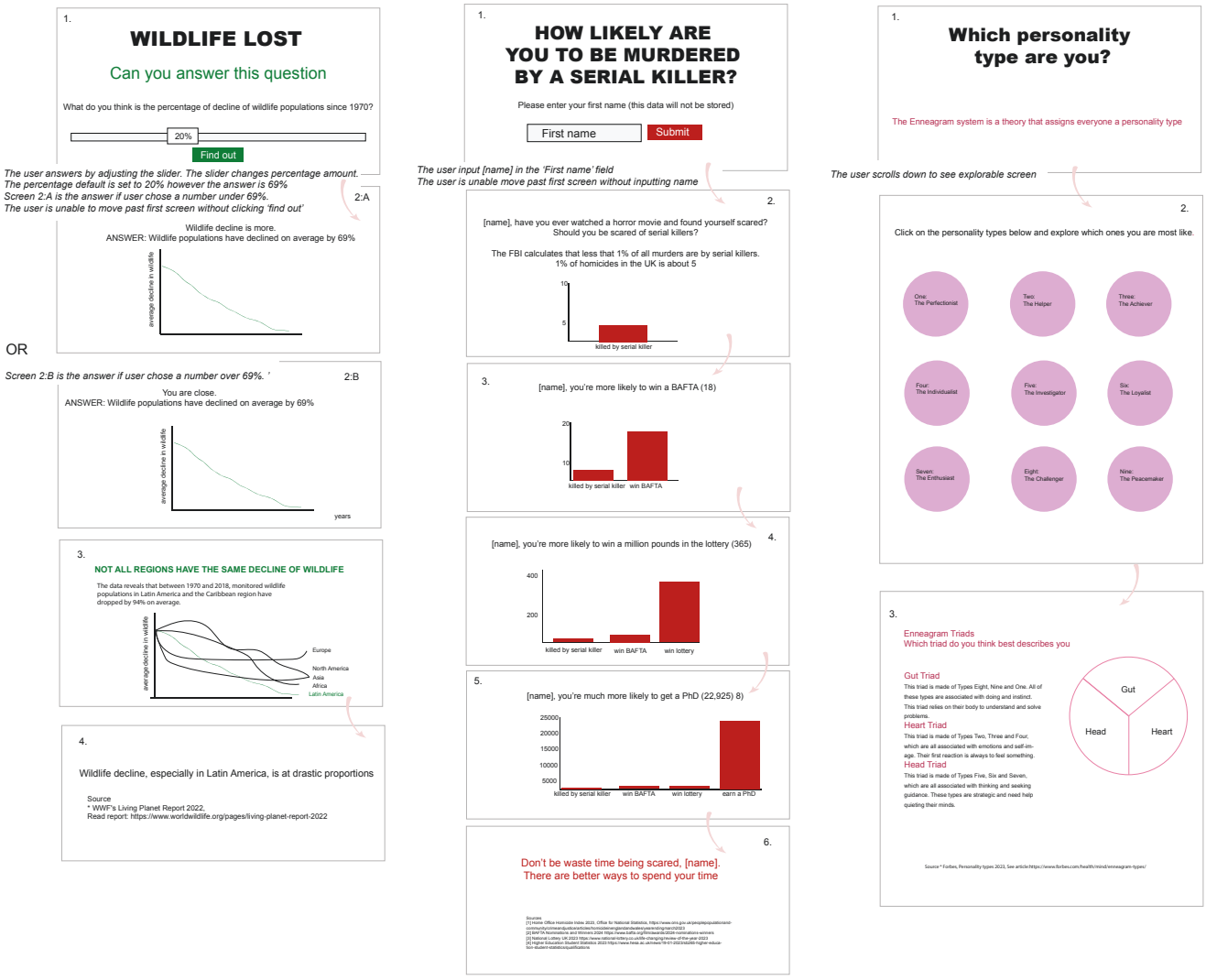


Figure 11: A diagram of interactive narrative visualisation design examples A) Design One: 'Make a Guess' B) Design Two: 'Breaking the Fourth Wall' and C) Design Three: 'Exploration'

scrolled down the screen.' All data sources are referenced at the end of the narrative visualisation. The participants that received Design Two were asked a comprehension check question: 'how many people earned a PhD?' See Figure 11:B for a diagram of Design Two.

Design Three: 'Exploration'

We refer to the third interactive narrative visualisation design example as 'Design Three.' Differentiating from the previous two examples, Design Three integrated a narrative pattern that encouraged data exploration. The audience is asked to freely explore data so that they can create their narrative. Such an experience is described as a 'reader-driven' narrative visualisation. Design Three was inspired by a narrative visualisation that appeared in a digital publication called 'The Pudding.' The specific interactive narrative visualisation was titled 'A Visual Guide to the Aztec Pantheon', which explained Aztec iconography [82]. Similar in interface design to the Pudding example, Design Three encouraged users to click on the interface to explore information. The example asked users, 'Which personality type are you? The Enneagram system is a theory that assigns everyone a personality type.' The user was then asked to click on the personality types to find out which one they most like. See Figure 11:C for a diagram of Design Three.

Survey Design

The survey instrument was adapted from the VisEngage engagement questionnaire [92]. The survey instrument contained 22 questions, where the 11 engagement characteristics were allocated two questions each. For clarity, the wording of each question mentions a 'data story' rather than a narrative visualisation. The participant could answer on a 7-point Likert scale from strongly disagree to agree strongly. Examples of the questions are as follows; 'While using this data story, I found its look and feel to be pleasing' or 'The content or message of this data story was interesting to me.' As described by Hung and Parsons, an overall engagement score can be achieved by adding together the results of each question [92]. Strongly disagree is allocated a one and strongly agree is allocated a 7. Therefore, the maximum engagement score is 154, corresponding to high engagement, and the minimum is 22, corresponding to low engagement.

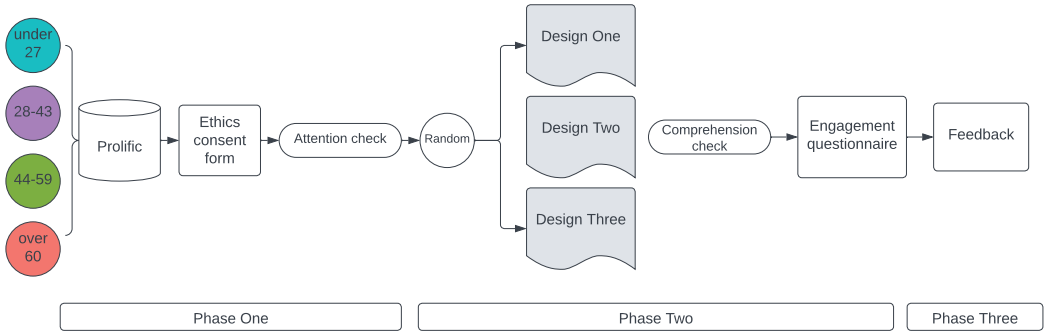


Figure 12: A flow diagram of experiment procedure. Experiment procedure: Prolific platform; ethics consent form; attention check, interactive narrative visualisation designs; comprehension check; engagement questionnaire and feedback question

Experiment Procedure

We conducted the experiment on the Prolific crowdsourcing platform. The experiment was in three phases. The first phase was where the participant exited Prolific and moved to the Qualtrics survey platform. They were asked to read and consent to the consent form, where ethics details were attached. In the next step, participants were asked for their Prolific ID, which was automatically inserted, and an attention check question. If the participant failed to consent, add their Prolific ID, or failed the attention check, their token was revoked, and they were returned to Prolific.

The second phase was where the participant was asked to 'please interact with the data story and then answer the questions below.' One of the three randomly allocated interactive narrative visualisation designs was presented using an iFrame, where the interactive narrative visualisation was hosted on an external server. After the iFrame, we posed a comprehension question to ensure participants had interacted with the narrative visualisation. After the comprehension check question, the participant answered the 22 engagement questions. All questions were on a Likert scale, and all were mandatory.

The final phase of the experiment was a qualitative feedback question that asked, ‘Did you feel you were engaged in the data story? Why or why not?’ This question was not mandatory. The participant could then either submit a response or move to the next step, which returned them to the Prolific platform.

Participants

We split each age cohort according to what are often described as ‘generations [4, 57].’ Generational research is a foundational topic in social sciences; however, we would highlight that the names used to describe generations can be loaded with stereotypical connotations. The objective of this research is not to perpetuate stereotypes associated with generational labels, and therefore we are not using the commonly used labels. We will refer to each age cohort by their age and refrain from using labels to diminish stereotypical connotations.

The youngest cohort consisted of ages ranging between 18-27. This age bracket saw participants born on or before 1997. Due to limitations with the crowd-sourcing platform, the youngest participant allowable age was 18. The second cohort had ages ranging between 28-43, where their birth year was on or between 1981 – 1996. This age cohort came to adulthood during the first years of the new millennium. Born between 1965-1980, this age cohort is between 44-59 years old. Finally, the oldest cohort consisted of participants with ages 60+ with a birth year of or above 1965.

Our participant sample size was a result of a power calculation with a goal of a 95% confidence level and a 4% margin of error. The population was calculated based on the adult population size in the UK in 2022 [2]. Our ideal sample size was calculated at approximately 601 participants per age cohort, therefore, with three design examples with equally distributed participants with 200 in each group, our total ideal sample size was approximately 2400 participants.

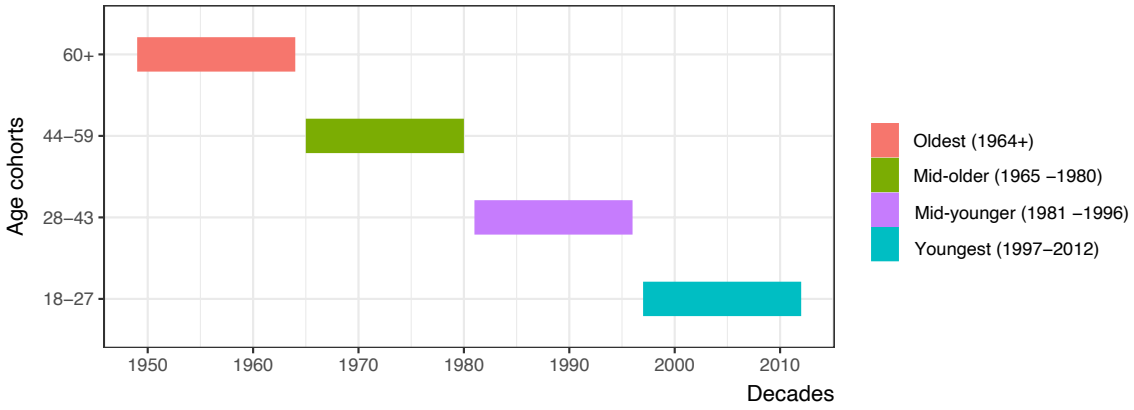


Figure 13: Depiction of age cohorts and the years of birth [57]. Note: chart is current in the year 2024.

Results

Hypotheses

We expected to observe differences between the four age cohorts while factoring in the effect of the narrative visualisation design examples. We firstly affirmed if a significant difference exists, specifically, our alternate hypothesis was as follows:

H1: There is a significant difference in engagement score and age cohort

Quantitative Analysis

Out of the 2400 participants, 77 failed the comprehension check. To investigate which age cohort had the greatest engagement, our dependent variable needed to be the overall engagement score. This was calculated by adding together all question responses in the VisEngage questionnaire as recommended by Hung and Parsons [92]. We normalized the engagement score data removing extreme outliers. Outliers were

identified with engagement scores below 50 and above 155, where we judged that their extreme response patterns indicated likely response bias. The number of outliers amounted to 43 participants, less than 2% of participants. A Shapiro-Wilk test was performed, and the result was not significant. Approximately equal variances were tested using the engagement score as the dependent variable in Levene's test, the result of which was not significant. The final number of participants totalled 2280 participants.

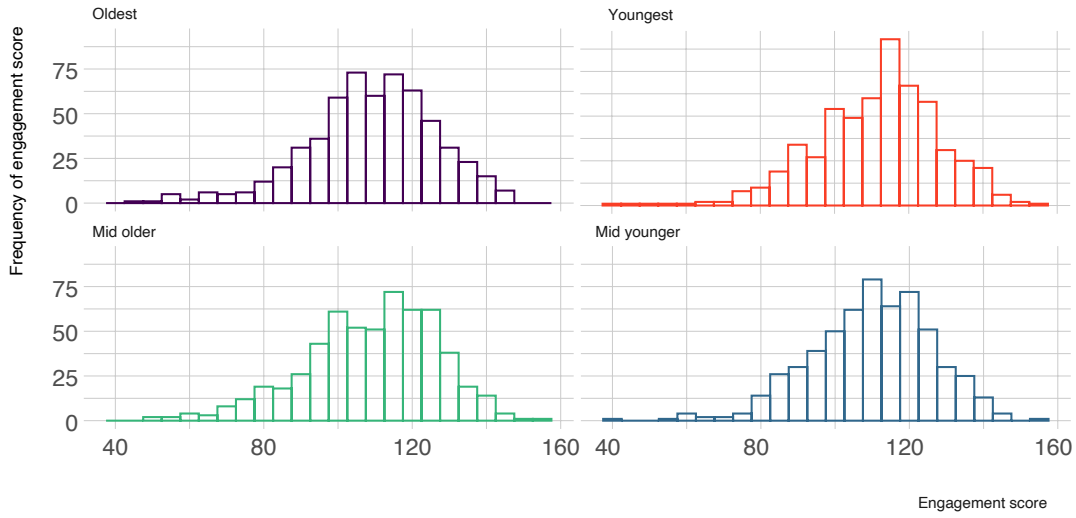


Figure 14: Small multiple of histograms displaying frequency of engagement score by age cohort

We ran a one-way ANOVA to compare the effect of age and engagement score. The one-way ANOVA revealed that there was a statistically significant difference in engagement between at least two groups (F(3, 2426) = [2.81], p = [0.03]). Tukey's HSD test for multiple comparisons found that the mean value of the engagement score was significantly different between the 60+ age cohort and 18-24 age cohort with a confidence interval of 95% (p = 0.05, 95% C.I. = [-5.1, 0.02]). All other age cohorts between groups means comparisons showed no significant differences. Figure 14 shows the frequency engagement scores per age cohort for all combined design examples.

Displaying that the frequency of higher engagement scores are in the youngest age cohorts.

According to our test results we could accept our research hypothesis (H1). We then measured effect size of age on engagement score. The effect size of age cohort on engagement score, as measured by Cohen's f, was 0.06, indicating a small effect size (95% C.I. = [0.01, 0]) [52]. We analysed the interaction effect of age cohort and narrative visualisation design example on participant engagement scores. The two-way ANOVA revealed that there was not a statistically significant interaction between the effects of age cohort and narrative visualisation design example (F(6, 2355) = [1.52], p = [0.17). Therefore, when factoring in the narrative visualisation design examples, they did not interact with participant engagement score.

Finally, overall engagement in the interactive narrative visualisation was positive for all age cohorts. As stated in section 'Survey Design,' the minimum possible engagement score was 22. The mean engagement score was 110 for all age cohorts combined. Furthermore, the median engagement score was 112. This data reveals that the majority of participants were positively engaged in interactive narrative visualisation.

Qualitative Analysis

We investigated the thought processes of participants by analyzing their responses to a qualitative feedback question. The survey instrument asked, 'Did you feel you were engaged in the data story? Why or why not?' The aim was shed light on the cognitive reasoning and reflections of participant on their engagement with the interactive narrative visualisation. This question was not mandatory, and we received 2278 responses. To qualitatively analyse the data, we adopted an inductive theming approach using latent theming [35]. We inductively coded lower-level themes determined by upper-level themes. Initially we determined upper-level themes by word frequency matched to

potential engagement related issues. Examples such as ‘attention,’ ‘color’ and ‘interactive’ were deemed as upper-level themes according to their relative high frequency in the qualitative data. Two coders independently coded responses, and any inconsistencies were discussed.

Interactivity

‘Interactive’ was mentioned seven times by the 60+ cohort. This was in contrast to the 18-27 age cohort, who mentioned ‘interactive’ or ‘interactivity’ 49 times. We analysed the exact phrases that participants used in the 18-27 cohort relating to interactivity. Our analysis revealed that the younger audience attributed their engagement to interactivity, for example, “yes because it was interactive” or I was engaged as it was an interactive task” (both comments from 18-27 age cohort, Design One).

The interactive device in Design Two, aimed to include the audience in the story and thus encourage engagement. 18-24 age cohort recognized that this was the aim of the interactive device, for example, “Yes as it was interactive and by using my name felt personal” (18-27 age cohort, Design Two). We found further evidence of perceptiveness in Design Three from the 18-27 age cohort “I did feel I was engaged in the data story as I had to click to find the information as well as scroll for more information” (18-27 age cohort, Design Three). The goal of the interactive device in Design One:, ‘Make a Guess’ was to illustrate a mismatch between audience expectation and reality. For example, ‘The decline is higher’ which made me feel engaged” (18-27 age cohort, Design One). While we observed that 18-24 age cohort was relatively more aware of the interactive devices this does not mean other age groups were oblivious, only their perceptive feedback was less frequent. For example, “I felt engaged as there was an interactive question where I could enter what I thought to be the answer. This made the impact of learning the true answer heavier as I was engaging with the story” (44-59 age cohort, Design One) or “I found the personality types interesting flip over to read and to associate

the descriptions with the images you gave for the personality" (60+ age cohort, Design Three).

When recognized by the 60+ age cohort, there were instances where the interactive device had the opposite effect of encouraging engagement and reported as a distraction or confusion, "My name personalized the story initially, but the repetition of it became quite annoying and distracting" (60+ age cohort, Design Two) or "I was distracted by, initially, not realizing I had to scroll down the box to gain more information" (60+ age cohort, Design Two) or "I was engaged, mainly because at first I was not sure what was going on. I was mainly confused" (60+ age cohort, Design Two). In the younger age cohorts, there were no observable reports of the interactive devices causing distraction or confusion. The younger age cohorts did report that they preferred an easy-to-use interface, where the interactive device did not detract from the user experience. As described here, "it would have been better if you did not have to scroll down" (28-43 age cohort, Design One), or it "felt like a bit of a gimmick, why not just have the information under the pictures instead of needing to click nine times" (18-27, Design Three). The data suggests that regardless of the audiences' age, usability influences engagement. For the older age cohorts, however, the effects of poor usability cause greater effect than a minor aggravation, but feelings of confusion and distraction.

Cognition

In older audiences, it has been shown that complex visualisations can be cognitively demanding, requiring users to remember and interpret multiple information pieces simultaneously [97]. We found evidence that older audiences expressed a preference for less complex visualisation, for example, "I don't find it particularly easy to interpret graphs or charts and I loathe Venn diagrams, so, I possibly had to concentrate more than other participants in order to ensure that I was interpreting the information correctly"

(60+ age cohort, Design Two) or “I found the charts a little confusing to begin with - possibly my age!” (44-59 age cohort, Design One).

Furthermore, older audiences mentioned that they were required to revisit the narrative visualisation to fully comprehend it; “My biggest problem with it was the need to scroll down. At first I didn't realize there was more data below” (60+ age cohort, Design One). The youngest age cohort did not report missing key information due to issues related to visual complexity. This reveals older audiences might miss crucial information if the representation is too complex or the interactive device is not clearly marked.

Older audiences mentioned legibility issues, particularly in regard to font sizes. One suggestion was larger, clearer text “I felt engaged because of the topic. I didn't see the answer at first so I would make this bigger and, in a pop-up,” (60+ age cohort, Design One) or “The graphics were too big or font size was too small” (44-59 age cohort, Design Two). Font size resulting in poor legibility is a documented issue for older audiences in interface design, but often overlooked in the context of visualisation design [198].

Aesthetic Appeal

We examined the responses of the participants who reported they were not engaged. The primary reason, reported by the 18-27 age cohort, was criticism of the aesthetic appeal of the narrative visualisations. For example, Design One, had a black background, which was described as ‘dated’, where for example it was stated, “It felt quite outdated especially with the colors” (18-27, Design One). Design Three, was described as ‘cluttered’, for example, “No as the text for the personality types 1-9 was cluttered” (18-27, Design Three). These comments reveal the importance of sound aesthetic design, particularly for an interactive narrative visualisation aimed at a younger audience.

One difference observed between the older and younger age cohorts was their preference toward text integrated into the narrative visualisation. Younger audiences preferred less text that was divided into smaller sections, for example “I felt that the gradual reveal of information meant that it was easier to compartmentalize statistics and different pieces of information rather than looking at a solid block of text, it felt more intuitive” (18-27, Design One). Older audiences preferred more text that provided context to the interactive narrative visualisation. For example, “It was thought provoking and I felt it needed more pages to explain what has been lost and why” (60+ age cohort, Design One).

Finally, we coded 1437 strongly positive responses. This is in line with our quantitative data analysis, where overall positive engagement was reported for all age cohorts, however slightly less for the older age cohorts. The 60+ age cohort reported 326 instances of positive engagement, 338 by the 44-59 age cohort, 382 by the 28-43 age cohort, and 391 times by the 18-27 age cohort. This reflected a generally positive opinion of engagement in the narrative visualisation examples across all age cohorts.

Discussion and Future Work

Designing Narrative Visualisation for Older Audiences

Older audiences are more attuned to usability difficulties.

Older age cohorts reported feeling distracted and confused by the interactive narrative patterns integrated into the narrative visualisation. These negative reactions could explain the lower engagement scores of the older age cohort compared to the youngest age cohort. It is important to note that all age cohorts desired ease of use. The primary difference between age cohorts was the extent of the negative reaction to usability-related concerns. For example, older age cohorts reported not recognizing functionality

in the interactive narrative visualisation, such as scrolling. Younger age cohorts recognized that they were required to scroll but preferred that it was not required. The findings of this study highlight the importance for narrative visualisation authors to prioritize usability for all audiences, however, especially if the visualisation is aimed at an older aged audience. Moreover, for older age cohorts, important information contained in an interactive narrative visualisation might be missed if the interactive device or data representation is too complex.

Older audiences prefer more text and larger font

In the study we found that older audiences asked for more text for context to the narrative presented. We suggest that authors of narrative visualisation provide the option for the end-user to access text about the presented topic. While it might not be necessary for primary messaging contained within the narrative visualisation, more text provides context to older audiences, where assumed knowledge might not be present.

Qualitative data analysis revealed that older audiences requested larger text for legibility. Legibility and more generally accessibility is a known issue for older audiences[97]. Specifically in interactive narrative visualisation, where interactive devices are integrated with the intent to engage an audience, fundamental accessibility concerns might be overlooked.

Designing Narrative Visualisation for Younger Audiences

Younger audiences understand interactive narrative patterns.

Qualitative data analysis indicated a stark difference in the terminology used by the youngest age cohort relative to the older age cohorts. Interactivity was mentioned much more frequently with the youngest age cohort, and usually in a positive light. Furthermore, the youngest age cohort seemed to have a perceptive understanding of

how the interactive narrative pattern achieved its intent of encouraging engagement. For example, the younger age cohort's responses to Design Three, 'Exploration' specifically outlined how exploring the data through clicking was more engaging than simply reading it. Our study sheds light on the depth of understanding of interactive devices held by the youngest age cohort, illustrated by their descriptive feedback.

The results of this study do not disprove the NN Group report, which stated it was a myth that young adults "crave multimedia and innovative design [81]." Rather, the results of this study give credence to the fact that young adults are accustomed and therefore more understanding of interactive devices. Increased engagement is, therefore largely due to the young adults' ability to perceive the intent of the author. Whereby recognizing that as an audience, younger age cohorts are expected to be engaged therefore they are engaged. For future authors of interactive narrative visualisation targeted towards younger age cohorts, it is recommended to use interactive narrative patterns with explicit intent. Duplicitous or superfluous use of interactive narrative patterns would likely be recognized and thus could result in lowered engagement.

Younger audiences appreciate aesthetics and less text

Aesthetics could potentially not be prioritized in interactive narrative visualisation. Specifically, when machine-driven workflows or AI co-creation tools are used for visualisation generation, aesthetic appeal might be considered of lesser importance to functionality and accuracy. Our study highlighted that if the intended audience is a younger adult audience, aesthetic appeal directly impacts their engagement. It is recommended that authors of narrative visualisation segment their information thoughtfully. We observed that younger audiences preferred an interface that was less cluttered and thus easy to digest. Furthermore, while it is helpful to use automated tools

for narrative visualisation generation, the role of the author is still vital to aesthetically evaluate the overall design and flow of the narrative visualisation.

Interactive Narrative Visualisation has a Broad Appeal.

One positive outcome of our study is the apparent appeal of interactive narrative visualisations. The mean engagement score for all age groups was 110, where the lowest possible score was 22, and the highest was 152. The positive mean average indicates overall positive engagement in interactive narrative visualisation. In addition, the qualitative analysis evidenced a largely positive reaction, where 63% of feedback responses reported a strongly positive engagement. This finding evidences that as a communication medium, interactive narrative visualisation can engage a broad audience. Authors of interactive narrative visualisation should not shy away from designing narrative visualisation for older audiences. The study presented here shows that, when designed inclusively, interactive narrative visualisation is an engaging medium for all age demographics.

Future Work

Fruitful future avenues of research could consider individual engagement characteristics. We have simply added the engagement 22 questions, where each of the 11 engagement characteristics received two corresponding questions, in the VisEngage questionnaire to achieve a final engagement score. We added them together as it was recommended by Hung and Parsons [92]. However, it could be interesting to investigate if individual characteristics appeal to age groups differently.

As the study of narrative visualisation audiences is still an emerging area of research, we suggest other demographics that could be worthy as a focus of investigation. Other demographics divided by gender, education, and location could prove to be interesting

avenues of investigation. The empirical evidence reported in this study will result in more effective interactive narrative visualisation as it can better inform future interactive narrative visualisation design and research. Empirical visualisation research can overlook the age of their participant base, this study shows that age can impact experiment results and should be reported [40]. We hope the findings of this study encourages authors and researchers to seriously consider audience age when designing or researching interactive narrative visualisation in the future.

Limitations

There are several limitations to this study. The foremost limitation is that there are but three narrative visualisation design examples. Optimally, we would have used a multitude of examples. However, the scope of this study dictated a limit of three. It should be noted, however, that each design example uses one of the three interactive narrative patterns that are described for engagement [27]. The objective of this study was a broad approach, where we used differing topics, narrative patterns, and designs. It is unfeasible to study all possible combinations of topics, narrative patterns, and designs. We believe that the three examples we developed were adequate to achieve our study's aims.

Another notable limitation is that this study is only representative of an audience based in the UK. The availability of the oldest age cohort from countries outside the UK and the US was specifically challenging and disappointing to the international research team. The uneven distribution of older participant country locations resulted in a decision to focus the experiment on participants from the UK. Rather than a skewed result, we prefer our results to concretely representing the behaviors of peoples from one locale. Furthermore, the premise of our study is to question whether different demographics engage differently with narrative visualisation; therefore, it stands to reason that the

locale of participants might impact study results. Comparing audience engagement across multiple countries is outside the scope of this work. For future researchers, we have provided our designs and code from the interactive narrative visualisation examples in the supplementary material and we encourage researchers to replicate this study in other languages or locales.

Conclusion

To communicate effectively, content authors are required to recognize the needs, preferences, and behaviours of their intended audience. The outcomes of this study suggest that audience age significantly impacts their engagement in interactive narrative visualisation. Older audiences that are in the 60+ age cohort find that interactive narrative patterns integrated into narrative visualisation cause usability difficulties. Younger age cohorts do not experience the same response when presented with interactive narrative patterns. Younger age cohorts recognize and appreciate interactive narrative patterns and are thus more engaged than their older counterparts. Our results lead to valuable implications for designing future interactive narrative visualisation.

Chapter Six—Conclusion and Implications

Conclusion

This thesis contributes by providing a comprehensive investigation into evaluating effective narrative visualisation. The first part of the thesis examined inspection methods of evaluation, namely heuristic evaluation. The perspective that we encompassed at all stages within the first part of this thesis was that of the practitioner. Through adopting a practice-led approach we have developed a heuristic framework that has real-world implications. Via a process of first literature review, practitioner survey and then a series of semi-structured interviews, it was possible to develop a heuristic framework that reflected the experience and objectives of practitioners. The heuristic framework was then validated in two separate settings using different scenarios, The breadth of our approach resulted in a rigorously developed set of heuristics. To our knowledge these are the first set of practice-led heuristics for the evaluation of narrative visualisation.

The second part of this thesis has potential to have far reaching implications. The contribution of a new way of employing narrative visualisation means that it could be potentially implemented for the specific purpose of ‘nudging’ an audience. Slightly changing the attitude of an audience may not initially seem as particularly impactful, however if extrapolated to a population level then it can potentially be quite a powerful mechanism. It is therefore necessary that nudges are ethically deployed, where it is based solidly on scientific evidence. In our study we used the context of wearing mask to hinder the spread of COVID-19. The evidence that mask wearing saved many lives is substantial and continues to be an important defence in the fight against COVID-19. The

second experiment in the latter part of this thesis addressed myths and assumptions perpetuated about the preferences and behaviours of age cohorts. The result of this experiment was significant. We found that older age cohorts engage less with interactive narrative visualisation. We believe that the findings of this study can better inform designers of narrative visualisation and therefore a valuable contribution to narrative visualisation research area.

Limitations and Future Research

There were some marked limitations that were pervasive through the research. The most prominent is the dichotomy of localisation and generalisation. A common issue with visualisation evaluation studies. This threat is particularly problematic for studies in narrative visualisation, where each, individual narrative visualisation is separate in its aims and presentation. Therefore, it could be argued that the result from the experiment are only related to the narrative visualisation that was used within that experiment. This in turn diminishes the universality of findings. We have recognised that this is a challenge and where possible have attempted to diversify the materials in our research. For example, the summative validation study used three different examples to evaluate, where each example was varied in structure and aim. It should be noted that it is impossible to study all scenarios and while it is accurate that the universality of findings could be questioned, we have attempted to overcome this limitation.

Another limitation that applied mostly to the opening chapters of this thesis, was in regard to control for sampling bias. We targeted practitioners who personally introduced themselves as being experienced in contributing to the development of narrative visualisation. 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 second part of the thesis did not have the threat of control sampling bias as participants were randomly selected via a crowd-sourcing platform. The crowd-sourcing platform did however prove to have some substantial limitations. This was apparent in both user-studies, where we were forced to focus on participants located in the UK. This was due to the fact that the crowd-sourcing platform is located in the UK and therefore the largest pool of participants were located in that locale. While it could be argued that the impact of having only participants from the UK is not greatly detrimental as culturally and linguistically the UK is in some ways similar to Australia, we believe that this is actually not the case. We have therefore made all code and other research materials freely available, where perhaps in the future the user-studies could be carried out in Australia, or other countries.

Finally it should be noted that a substantial portion (58%) of practitioners we interviewed, identified as working in the journalism domain. This has skewed our results to favour the perspective of those working in that particular domain. Further investigation is needed into whether a different practitioners domain ratio would impact the final heuristic framework.

Topics for Future Research

Heuristic Framework with Non-Experts

A promising avenue for further research involves exploring the possibility of implementing the heuristic framework in a classroom setting. This has been employed with information visualisation heuristics and shown to have beneficial results [163]. The heuristics operated as a teaching tool that could foster design skills in students. At this point in time all practitioners that we have investigated related to the heuristic set were experienced professionals. It stands to reason that non-experts could have a different experience

when implementing the heuristics. Rather than an informal checklist to remind the user of best practice, the heuristics could function as a guide and teaching aid.

Heuristic Framework Tailored to Narrative Visualisation Subsets

As new forms of narrative visualisation emerge so should the heuristic framework evolve to accommodate them. Similar to the usability heuristics proposed by Nielsen, the heuristics proposed here are minimalist [134]. They can be adapted and expanded to be tailored to subsets of narrative visualisation. Some potential subsets such as interactive narrative visualisation and scrollytelling have requirements that are distinct. These would require an adapted heuristic framework. Future work could even predict

Communication Strategies

Otherwise referred to as ‘narrative patterns [133]’ communication strategies are devices integrated into narrative visualisation with specific intent. In the third chapter of this thesis two communication strategies that encourage empathy were selected. These were namely, ‘personal identification’ and ‘anthropomorphism’, There are however 13 other communication strategies described in the seminal book ‘Data-Driven Storytelling [133].’ Some examples of potentially exciting communication strategies to study include, studying the effects of ‘flow management’ where for example data is revealed deliberately at a progressively slower or faster pace to encourage an affective response [133]. Another example is data exploration as communication strategy deliberately intent on encouraging engagement in the audience. Recently other communication strategies mentioned have been investigated with varying success, namely the ‘make a guess’ narrative pattern and the ‘breaking the fourth wall’ narrative pattern [157, 172].

Addressing Other Myths and Assumptions

Another important area for future investigation is to address other myths and assumptions about narrative visualisation audiences. In this thesis I have investigated

age based myths and assumptions as they relate to engagement in narrative visualisation. In grey literature surrounding narrative visualisation, such as blogs or YouTube videos, other concepts that could be classed as myths are perpetuated. Some audience preference and behaviours myths include designing for specific gender or designing for an education level or culture. Without scientific investigation, it is not possible to accurately determine the needs of the audience. When narrative visualisation is conveying vital information, authors that are not properly informed could design less effective narrative visualisation. Therefore evidence-based advice and recommendations is required so that authors of narrative visualisation can design for their audience.

Investing Other 'Nudge' Types and Ethical Guidelines

In this thesis I looked at one type of 'nudge' namely the 'availability heuristic.' The 'Availability' heuristic relates to how common or visible or familiar something is perceived to be. The greater the commonness, then the greater the perceived frequency or incidence, and also the greater sense of trust in the communication. Other heuristics for nudges include 'Anchoring and Adjustment' defined as using a known/comparable fact and adjusting it to estimate or decide about something which is unknown [183]. It would be potentially interesting to see if the 'Anchoring and Adjustment' nudge type is effective as it is popular among authors. Finally, another 'nudge' type is 'Representativeness' which is described as how similar something is thought to be in relation to a perceived stereotype or assumption [183]. This concept could be employed in various implementations. Further research would be needed to find out the exact contextual benefits 'Representativeness' could have.

Shifting the attitude of an audience is a powerful mechanism and could be valuable in contexts outside of purporting health advice. Contexts where nudging with narrative visualisation could be seen as valuable could include product marketing and political

lobbying. These are contexts where shifting an audience attitude may not be done with their best interests at heart. It is therefore imperative that greater thought and guidelines for ethically integrating nudges in narrative visualisation are created. This is perhaps the most impactful of the avenues of future work that have been presented in this section of my thesis. I believe that clear ethical guidelines for nudging with narrative visualisation would be an impactful and valuable future work project.

Investigating Novel Evaluation Methods

In the first part of this thesis practitioners uncovered some novel approaches to evaluation. These could be exciting avenues of research. The most fascinating novel evaluation approach was described as ‘guerilla user-testing.’ In this instance no laboratory is required, instead the evaluator approaches members of the public in a public space and asks them if they would consent to taking part in a test. This form of evaluation has the benefit of spontaneity and fluidness. This evaluation method has yet to be investigated in relation to visualisation generally, and narrative visualisation more specifically. Finally, another novel form of evaluation was the use of social media to evaluate ‘on-the-fly’ where the author would update the narrative visualisation according to feedback in a series of short feedback loops.

Summary

This thesis makes several noteworthy contributions to the existing body of research in the field of narrative visualisation. These contributions encompass a comprehensive exploration of various methods of evaluation.

First, this study extends the scholarship on heuristic evaluation of narrative visualisation, focusing on the practitioners that author narrative visualisation. By integrating their feedback and expertise, the proposed heuristic framework for evaluation is a practical tool that can be used in a real-life setting. Unlike many domain-specific heuristic sets this

set has been validated. Important to identify weaknesses and gaps in the heuristics, validation is crucial step to establishing heuristics.

Second, the thesis introduces a new means to employ narrative visualisation, namely to ‘nudge’ an audience. To the authors knowledge, nudging is yet to used in relation to narrative visualisation. It has the potential to be a powerful mechanism, where slightly changing the attitude of an audience could have real-world consequences. This study was able evaluate narrative visualisation effectiveness in a tailored audience setting resulting in evidence-based recommendations for narrative visualisation authors. Furthermore this thesis contributed greater understanding of communication strategies integrated into narrative visualisation and their effectiveness. This is still an emerging area of literature and this study provides the foundation for later work in the area.

Third, the thesis addresses myths and assumptions about narrative visualisation audiences. While it is important to tailor narrative visualisation to an audience, it is still important to fully understand the audience. By investigating one of the most fundamental aspects of audience categorisation, namely age, this work will lead to better informed designers of narrative visualisation.

In summary, this research contributes to a holistic approach to narrative visualisation evaluation. It enriches our knowledge and appreciation of narrative visualisation. Through encouraging a conversation about evaluation narrative visualisation will be more effective. The integral role that practitioners played in this thesis means that it has both a practical and theoretical contributions. As an important medium of communication, which is increasingly growing in popularity, it is vital that narrative visualisation is both comprehensible and engaging.

Bibliography

[1] International Standards Organization (ISO). 2018. Iso 9241-11:2018(En): Ergonomics of Human-System Interaction — Part 11: Usability: Definitions and Concepts. Retrieved from https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en.

[2] Office for National Statistics (ONS). 2022. Population Estimates for the Uk, England, Wales, Scotland and Northern Ireland: Mid-2021. Retrieved April, 2024 from https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2021.

[3] Zahra Shakeri Hossein Abad, Guenther Ruhe and Mohammad Noaeen. 2016. Requirements Engineering Visualization: A Systematic Literature Review. *Proceedings - 2016 IEEE 24th International Requirements Engineering Conference, RE 2016*, 6-15. http://dx.doi.org/10.1109/RE.2016.61

[4] Tracie O. Afifi, Samantha Salmon, Tamara Taillieu, Ashley Stewart-Tufescu, Janique Fortier and S. Michelle Driedger. 2021. Older Adolescents and Young Adults Willingness to Receive the Covid-19 Vaccine: Implications for Informing Public Health Strategies. *Vaccine* 39, 3473-3479. http://dx.doi.org/10.1016/j.vaccine.2021.05.026

[5] Abmatic AI. 2023. Personalizing the Website for Different Age Groups. Retrieved April, 2024 from https://abmatic.ai/blog/personalizing-website-for-different-age-groups.

[6] Julia Aledort, Nicole Lurie, Jeffery Wasserman and Samual Bozzette. 2007. Non-Pharmaceutical Public Health Interventions for Pandemic Influenza: An Evaluation of the Evidence Base. *BMC Public Health* 7, 208-208. http://dx.doi.org/0.1186/1471-2458-7-208

[7] Ben Hubbard Alissa J. Rubin, Josh Holder, Noah Throop, Emily Rhyne, Jeremy White and James Glanz. Produced by Josh Williams, Sarah Almukhtar, Rumsey Taylor, Josh Keller, Falih Hassan, Ehab Al-Rikabi. 2022. Extreme Heat Will Change Us. Retrieved 17/07, 2023 from https://www.nytimes.com/interactive/2022/11/18/world/middleeast/extreme-heat.html.

[8] Robert Amar and John Stasko. 2004. A Knowledge Task-Based Framework for Design and Evaluation of Information Visualizations. *Proceedings - IEEE Symposium*

on Information Visualization, INFO VIS, 143-149. http://dx.doi.org/10.1109/INFVIS.2004.10

[9] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Christophe Hurter and Pourang Irani. 2015. Understanding Data Videos: Looking at Narrative Visualization through the Cinematography Lens. In *Proceedings of Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Seoul, Republic of Korea, 1459–1468. http://dx.doi.org/10.1145/2702123.2702431

[10] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Jason Leboe-McGowan and Pourang Irani. Year. Hooked on Data Videos: Assessing the Effect of Animation and Pictographs on Viewer Engagement. In *Proceedings of International Conference on Advanced Visual*. 1-9. http://dx.doi.org/10.1145/3206505.3206552

[11] Keith Andrews. Year. Evaluating Information Visualisations. In *Proceedings of BELIV*.

[12] Victor Nikhil Antony and Chien-Ming Huang. 2024. Id.8: Co-Creating Visual Stories with Generative Ai. *ACM Trans. Interact. Intell. Syst.* 14, 3, Article 20. http://dx.doi.org/10.1145/3672277

[13] Lauren Aratani. 2020. How Did Face Masks Become a Political Issue in America? Retrieved September, 2022 from https://www.theguardian.com/world/2020/jun/29/face-masks-us-politics-coronavirus.

[14] Simon Attfield, Gabriella Kazai, Mounia Lalmas and Benjamin Piwowarski. 2011. Towards a Science of User Engagement (Position Paper).

[15] Online Journalism Awards. 2020-2023. Retrieved 11 September, 2024 from https://awards.journalists.org.

[16] Tversky B. 2006. *Cognitive Design Principles for Automated Generation of Visualizations*. Lawrence Erlbaum Associates,

[17] Benjamin Bach, Mandy Keck, Fateme Rajabiyazdi, Tatiana Losev, Isabel Meirelles, Jason Dykes, Robert S. Laramee, Mashael AlKadi, Christina Stoiber, Samuel Huron, Charles Perin, Luiz Morais, Wolfgang Aigner, Doris Kosminsky, Magdalena Boucher, Soren Knudsen, Areti Manataki, Jan Aerts, Uta Hinrichs, Jonathan C. Roberts and Sheelagh Carpendale. 2023. Challenges and Opportunities in Data Visualization Education: A Call to Action. *IEEE Transactions on Visualization and Computer Graphics* 30, 1, 649-660. http://dx.doi.org/10.1109/TVCG.2023.3327378

[18] Benjamin Bach, Zezhong Wang, Matteo Farinella, Dave Murray-Rust and Nathalie Henry Riche. 2018. Design Patterns for Data Comics. In *Proceedings of 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing

Machinery, Montreal QC, Canada, Paper 38. http://dx.doi.org/10.1145/3173574.3173612

[19] S. Sandra Bae, Rishi Vanukuru, Ruhan Yang, Peter Gyory, Ran Zhou, Ellen Yi Luen Do and Danielle Albers Szafir. 2023. Cultivating Visualization Literacy for Children through Curiosity and Play. *IEEE Transactions on Visualization and Computer Graphics* 29, 1, 257-267. http://dx.doi.org/10.1109/TVCG.2022.3209442

[20] S Bahador, A Endert and J Stasko. Year. Beyond Usability and Performance: A Review of User Experience-Focused Evaluations in Visualization. In *Proceedings of BELIV*.

[21] Michelle Q.Wang Baldonado, Allison Woodruff and Allan Kuchinsky. 2000. Guidelines for Using Multiple Views in Information Visualization. *Proceedings of the Workshop on Advanced Visual Interfaces*, 110-119. http://dx.doi.org/10.1145/345513.345271

[22] Firdaus Banhawi and Nazlena Mohamad Ali. 2011. *Measuring User Engagement Attributes in Social Networking Application*.

[23] Soubhik Barari, Stefano Caria, Antonio Davola, Paolo Falco, Thiemo Fetzer and Stefano Fiorin. 2020. Evaluating Covid-19 Public Health Messaging in Italy: Self-Reported Compliance and Growing Mental Health Concerns. *Mental Health Weekly Digest*, 369. http://dx.doi.org/10.1101/2020.03.27.20042820

[24] Lyn Bartram, Abhisekh Patra and Maureen Stone. 2017. Affective Color in Visualization. In *Proceedings of Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Denver, Colorado, USA, 1364–1374. http://dx.doi.org/10.1145/3025453.3026041

[25] Scott Bateman, Regan L. Mandryk, Carl Gutwin, Aaron Genest, David McDine and Christopher Brooks. Year. Useful Junk?: The Effects of Visual Embellishment on Comprehension and Memorability of Charts. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*. ACM. http://dx.doi.org/https://dl.acm.org/doi/10.1145/1753326.1753716

[26] Beatriz Quintino Ferreira Beatriz Sousa Santos, and Paulo Dias. 2015. Heuristic Evaluation in Information Visualization Using Three Sets of Heuristics: An Exploratory Study Beatriz. *Human-Computer Interaction* 67, 259-270. http://dx.doi.org/10.1007/978-3-319-20901-2

[27] Moritz Stefaner Benjamin Bach, Jeremy Boy, Steven Drucker, Lyn Bartram, Jo Wood, Paolo Ciuccarelli, Yuri Engelhardt, Ulrike Köppen, Barbara Tversky. 2018. Narrative Design Patterns for Data-Driven Storytelling. In *Data-Driven Storytelling*, Christophe Hurter Nathalie Henry Riche, Nicholas Diakopoulos, Sheelagh Carpendale Ed. CBC Press, 107-133.

[28] Alan Berkowitz. 2004. The Social Norms Approach: Theory, Research and Annotated Bibliography. 2004.

[29] Fearn Bishop, Johannes Zagermann, Ulrike Pfeil, Gemma Sanderson, Harald Reiterer and Uta Hinrichs. 2020. Construct-a-Vis: Exploring the Free-Form Visualization Processes of Children. *IEEE Transactions on Visualization & Computer Graphics* 26, 01, 451-460. http://dx.doi.org/10.1109/tvcg.2019.2934804

[30] J Black and G Bower. 1979. Episodes as Chunks in Narrative Memory. *Journal of Verbal Learning and Verbal Behavior* 18, 3, 309 - 318.

[31] R. Borgo, A. Abdul-Rahman, F. Mohamed, P. W. Grant, I. Reppa, L. Floridi and Chen Min. 2012. An Empirical Study on Using Visual Embellishments in Visualization. *IEEE Transactions on Visualization and Computer Graphics* 18, 12, 2759-2768. http://dx.doi.org/https://doi.org/10.1109/TVCG.2012.197

[32] Michelle Borkin, Azalea Vo, Zoya Bylinskii, Phillip Isola, Student Member, Shashank Sunkavalli, Aude Oliva, Hanspeter Pfister and Senior Member. 2013. What Makes a Visualization Memorable? *IEEE Transactions on Visualization and Computer Graphics* 19, 12, 2306 – 2315. http://dx.doi.org/https://doi.org/10.1109/TVCG.2013.234

[33] J Boy, F Detieene and JD Fekete. Year. Storytelling in Information Visualizations: Does It Engage Users to Explore Data? In *Proceedings of Conference on Human Factors in Computing Systems*. http://dx.doi.org/https://doi.org/10.1145/2702123.2702452

[34] Jeremy Boy, Anshul Vikram Pandey, John Emerson, Maraget Satterthwaite, Oded Nov and Enrico Bertini. Year. Showing People Behind Data: Does Anthropomorphizing Visualizations Elicit More Empathy for Human Rights Data? In *Proceedings of 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. http://dx.doi.org/10.1145/3025453.3025512

[35] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2, 77 - 101.

[36] J Brehm and S Brehm. 1981. *Psychological Reactance: A Theory of Freedom and Control*. Academic Press, San Diego.

[37] M Brehmer, B Lee, B Bach, N Riche and T Munzner. 2017. Timelines Revisited: A Design Space and Considerations for Expressive Storytelling. *IEEE Transactions on Visualization and Computer Graphics* 23, 9, 2151 - 2164. http://dx.doi.org/https://doi.org/10.1109/TVCG.2016.2614803

[38] J Brilhart and G Galances. 1998. Decision Making in Small Groups, Nineth Edition. In *Effective Group Discussion*, McGraw Hill, USA, 237-258.

[39] Samantha Brooks, Rebecca Webster, Louise Smith, Lisa Woodland, Simon Wessely and Niel Greenberg. 2020. The Psychological Impact of Quarantine and How to Reduce It: Rapid Review of the Evidence. *Lancet* 395, 912–920. http://dx.doi.org/10.1016/S0140-6736(20)30460-8

[40] Alyxander Burns, Christiana Lee, Ria Chawla, Evan Peck and Narges Mahyar. 2023. Who Do We Mean When We Talk About Visualization Novices? *Conference on Human Factors in Computing Systems - Proceedings*. http://dx.doi.org/10.1145/3544548.3581524

[41] A Cairo. 2013. Emotional Data Visualization: Periscopic's "U.S. Gun Deaths" and the Challenge of Uncertainty. Retrieved September, 2023 from https://www.peachpit.com/articles/article.aspx?p=2036558.

[42] A Cairo. 2016. *What We Talk About When We Talk About Visualization*. New Riders,

[43] Gabriela Cajamarca, Valeria Herskovic, Andrés Lucero and Angeles Aldunate. 2022. A Co-Design Approach to Explore Health Data Representation for Older Adults in Chile and Ecuador. In *Proceedings of 2022 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, 1802–1817. http://dx.doi.org/10.1145/3532106.3533558

[44] Bobby Calder and Richard Lutz. Year. An Investigation of Some Alternatives to the Linear Attitude Model. In *Proceedings of Proceedings of the Third Annual Conference of the Association for Consumer Research*.

[45] S Card, J Mackinlay and B Shneiderman. 1999. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufman, Burlington MA.

[46] M Caris, H Labuschagne, M Dekker, M Kramer, M Agtmael and C Vandenbroucke-grauls. 2018. Nudging to Improve Hand Hygiene. *Journal of Hospital Infection* 98, 4, 352-358. http://dx.doi.org/10.1016/j.jhin.2017.09.023

[47] Sheelagh Carpendale. 2008. Evaluating Information Visualizations. In *Information Visualization*, Andreas Kerren et al. Eds. Springer Berlin Heidelberg, Berlin, Heidelberg, 19-45.

[48] John M. Carroll and Mary Beth Rosson. 1992. Getting around the Task-Artifact Cycle: How to Make Claims and Design by Scenario. *ACM Transactions on Information Systems* 10, 2, 181-212. http://dx.doi.org/https://doi/10.1145/146802.146834

[49] Diana Carvalho, Maximino Bessa, Luís Magalhães and Eurico Carrapatoso. 2016. Age Group Differences in Performance Using Diverse Input Modalities: Insertion Task Evaluation. In *Proceedings of XVII International Conference on Human Computer Interaction*. Association for Computing Machinery, Salamanca, Spain, Article 12. http://dx.doi.org/10.1145/2998626.2998664

[50] Alan Chamberlain, Andy Crabtree, Tom Rodden, Matt Jones and Yvonne Rogers. 2012. Research in the Wild: Understanding 'in the Wild' Approaches to Design and Development. In *Proceedings of Designing Interactive Systems Conference*. Association for Computing Machinery, Newcastle Upon Tyne, United Kingdom, 795–796. http://dx.doi.org/10.1145/2317956.2318078

[51] Alison Cohen, Lindsay Hoyt and Brandon Dull. 2020. A Descriptive Study of Covid-19–Related Experiences and Perspectives of a National Sample of College Students in Spring 2020. *Journal of Adolescent Health* 67, 3, 369-375. http://dx.doi.org/10.1016/j.jadohealth.2020.06.009

[52] J Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences.* NY: Routledge Academic, New York.

[53] Shlomo Cohen. 2013. Nudging and Informed Consent. *Am J Bioeth* 13, 6, 3-11. http://dx.doi.org/10.1080/15265161.2013.781704

[54] Ignaas Devisch. 2011. Progress in Medicine: Autonomy, Oughtonomy and Nudging. *Journal of Evaluation in Clinical Practice* 17, 857-861. http://dx.doi.org/0.1111/j.1365-2753.2011.01716.x

[55] Nicholas Diakopoulos, Funda Kivran-Swaine and Mor Naaman. Year. Playable Data: Characterizing the Design Space of Game-Y Infographics. In *Proceedings of Conference on Human Factors in Computing Systems*. 1717-1726. http://dx.doi.org/10.1145/1978942.1979193

[56] Evanthia Dimara and Charles Perin. 2020. What Is Interaction for Data Visualization? *IEEE Transactions on Visualization and Computer Graphics* 26, 1, 119-129. http://dx.doi.org/10.1109/TVCG.2019.2934283

[57] Michael Dimock. 2019. Defining Generations: Where Millennials End and Generation Z Begins. Retrieved April, 2024 from https://www.pewresearch.org/short-reads/2019/01/17/where-millennials-end-and-generation-z-begins/.

[58] David Dowling and Travis Vogan. 2015. Can We “Snowfall” This? *Digital Journalism* 3, 2, 209-224. http://dx.doi.org/10.1080/21670811.2014.930250

[59] Alice Eagly and Shelly Chaiken. 1993. *The Psychology of Attitudes*. Harcourt, Brace and Jovanovich, Fort Worth.

[60] Alice H. Eagly and Shelly Chaiken. 2007. The Advantages of an Inclusive Definition of Attitude. *Social Cognition* 25, 5, 582--602. http://dx.doi.org/10.1521/soco.2007.25.5.582

[61] Geoffrey Ellis and Alan Dix. Year. An Explorative Analysis of User Evaluation Studies in Information Visualisation. In *Proceedings of 2006 AVI workshop on BEyond time*

and errors: novel evaluation methods for information visualization. ACM, 1-7. http://dx.doi.org/https://doi.org/10.1145/1168149.1168152

[62] David Erceg-Hurn and Lyndall Steed. 2011. Does Exposure to Cigarette Health Warnings Elicit Psychological Reactance in Smokers? *Journal of Applied Social Psychology* 41, 219-237. http://dx.doi.org/10.1111/j.1559-1816.2010.00710.x

[63] Nina Errey, Jie Liang, Tuck Wah Leong and Didar Zowghi. 2023. Evaluating Narrative Visualization: A Survey of Practitioners. *International Journal of Data Science and Analytics*. http://dx.doi.org/10.1007/s41060-023-00394-9

[64] Matteo Farinella, Dave Murray-rust Nathalie, Henry Riche, Benjamin Bach, Zezhong Wang and Shunming Wang. Year. Comparing Effectiveness and Engagement of Data Comics and Infographics. In *Proceedings of CHI*. 1-12. http://dx.doi.org/10.1145/3290605.3300483

[65] Nathalie Henry Riche Bongshin Lee Christophe Hurter Pourang Irani Fereshteh Amini. 2005. Understanding Data Videos: Looking at Narrative Visualization through the Cinematography Lens. *Hydroinformatics*, 265-280. http://dx.doi.org/10.1201/9781420038002.ch15

[66] Tom Fevrier and Marie Patino. 2024. French Left Alliance Wins Most Seats, Far Right Fails to Carry First-Round Victory. Retrieved 11 September, 2024 from https://www.bloomberg.com/graphics/2024-french-election-results/.

[67] Ana Figueiras. Year. How to Tell Stories Using Visualization. In *Proceedings of International Conference on Information Visualisation*. 18-26.

[68] Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Thomas A Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah and Jeffrey W. Eaton. 2020. Estimating the Effects of Non-Pharmaceutical Interventions on Covid-19 in Europe. *Nature* 584, 7820, 257-261. http://dx.doi.org/10.1038/s41586-020-2405-7

[69] Flourish. 2023. Retrieved September, 2023 from https://flourish.studio.

[70] Camilla Forsell and Jimmy Johansson. 2010. An Heuristic Set for Evaluation in Information Visualization. In *Proceedings of International Conference on Advanced Visual Interfaces*. Association for Computing Machinery, Roma, Italy, 199–206. http://dx.doi.org/10.1145/1842993.1843029

[71] World Wildlife Fund. 2022. Living Planet Report 2022. 2022.

[72] S. Guo G. Wu, J. Hoffswell, G. Y. -Y. Chan, R. A. Rossi and E. Koh. 2024. Socrates: Data Story Generation Via Adaptive Machine-Guided Elicitation of User Feedback. *IEEE Transactions on Visualization and Computer Graphics* 30, 1, 131-141. http://dx.doi.org/10.1109/TVCG.2023.3327363

[73] Alexander Genevsky, Daniel Västfjäll, Paul Slovic and Brian Knutson. 2013. Neural Underpinnings of the Identifiable Victim Effect: Affect Shifts Preferences for Giving. *The Journal of Neuroscience* 33, 43, 17188–17196. http://dx.doi.org/10.1523/JNEUROSCI.2348-13.2013

[74] Jennifer George-Palilonis and Mary Spillman. 2013. Storytelling with Interactive Graphics: An Analysis of Editors' Attitudes and Practices. *Visual Communication Quarterly* 20, 1, 20-27. http://dx.doi.org/10.1080/15551393.2013.765777

[75] Nahum Gershon and Ward Page. 2001. What Storytelling Can Do for Information Visualization. *Communications of the ACM* 44, 8, 31-37 %U https://dl.acm.org/doi/10.1145/381641.381653.

[76] D Gotz and Z Wen. Year. Behavior-Driven Visualization Recommendation. In *Proceedings of International Conference on Intelligent User Interfaces*. 315 - 324.

[77] Walter Gove and Michael Geerken. 1977. Response Bias in Surveys of Mental Health: An Empirical Investigation. *American Journal Of Sociology* 82, 6, 1289-1317.

[78] Institute for Government. 2022. Nudge Unit. Retrieved December, 2022 from https://www.instituteforgovernment.org.uk/explainers/nudge-unit.

[79] Scott Graupensperger, Christine Lee and Mary Larimer. 2021. Young Adults Underestimate How Well Peers Adhere to Covid-19 Preventive Behavioral Guidelines. *Journal of Primary Prevention* 42, 3, 309-318. http://dx.doi.org/10.1007/s10935-021-00633-4

[80] Amanda Cox Gregor Aisch, Kevin Quealy 2015. You Draw It: How Family Income Predicts Children's College Chances. Retrieved 11 Septembet, 2024 from https://www.nytimes.com/interactive/2015/05/28/upshot/you-draw-it-how-family-income-affects-childrens-college-chances.html.

[81] Nielsen Norman Group. 2017. Designing for Young Adults (Ages 18–25). 2017.

[82] Luc Guillemot Gwendal Uguen. 2022. A Visual Guide to the Aztec Pantheon. Retrieved 11 September, 2024 from https://pudding.cool/2022/06/aztec-gods/.

[83] Marrisa Harrison and A E Hall. 2010. Anthropomorphism, Empathy, and Perceived Communicative Ability Vary with Phylogenetic Relatedness to Humans. *Social, Evolutionary, and Cultural Psychology* 4, 1.

[84] Marti A. Hearst, Paul Laskowski and Luis Silva. 2016. Evaluating Information Visualization Via the Interplay of Heuristic Evaluation and Question-Based Scoring. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5028-5033. http://dx.doi.org/https://doi.org/10.1145/2858036.2858280

[85] Setia Hermawati and Glyn Lawson. 2016. Establishing Usability Heuristics for Heuristics Evaluation in a Specific Domain: Is There a Consensus? *Applied Ergonomics* 56, 34-51. http://dx.doi.org/https://doi.org/10.1016/j.apergo.2015.11.016

[86] Jeremy Heyer, Nirmal Kumar Raveendranath and Khairi Reda. Year. Pushing the (Visual) Narrative: The Effects of Prior Knowledge Elicitation in Provocative Topics. In *Proceedings of 2020 CHI Conference on Human Factors in Computing Systems (CHI '20).* 1-14. http://dx.doi.org/10.1145/3313831.3376887

[87] R Hirokawa. 2009. Consensus Group Decision-Making, Quality of Decision, and Group Satisfaction: An Attempt to Sort “Fact” from “Fiction”. *Communication Studies,* 33, 2, 407-415.

[88] Nigel Holmes. 1991. *Designer's Guide to Creating Charts and Diagrams*. Watson-Guptill Publications,

[89] Solomon Hsiang, Daniel Allen, Sébastien Annan-Phan, Kendon Bell, Ian Bolliger, Trinetta Chong, Hannah Druckenmiller, Luna Yue Huang, Andrew Hultgren, Emma Krasovich, Peiley Lau, Jaecheol Lee, Esther Rolf, Jeanette Tseng and Tiffany Wu. 2020. The Effect of Large-Scale Anti-Contagion Policies on the Covid-19 Pandemic. *Nature* 584, 7820, 262-267. http://dx.doi.org/10.1038/s41586-020-2404-8

[90] Jessica Hullman, Steven Drucker, Nathalie Henry Riche, Bongshin Lee, Danyel Fisher and Eytan Adar. 2013. A Deeper Understanding of Sequence in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics* 19, 12, 2406 - 2415. http://dx.doi.org/10.1109/TVCG.2013

[91] Jessica Hullman and Nick Diakopoulos. 2011. Visualization Rhetoric: Framing Effects in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics* 17, 12, 2231-2240. http://dx.doi.org/10.1109/TVCG.2011.255

[92] Ya-Hsin Hung and Paul Parsons. Year. Assessing User Engagement in Information Visualization. In *Proceedings of Conference on Human Factors in Computing Systems*. http://dx.doi.org/10.1145/3027063.3053113

[93] Elaine Huynh;, Angela Nyhout;, Patricia Ganea; and Fanny Chevalier. 2021. Designing Narrative-Focused Role-Playing Games for Visualization Literacy in Young Children. *IEEE Transactions on Visualization and Computer Graphics* 27, 2, 924-934. http://dx.doi.org/10.1109/TVCG.2020.3030464

[94] IAB.UK. 2024. Interactive Ads More Appealing to Teens, Study Shows. Retrieved 11 September, 2024 from https://www.iabuk.com/news-article/interactive-ads-more-appealing-teens-study-shows.

[95] K. Daigle J. Janicki, S. Kiyada and A. Rao. 2022. On the Brink. Retrieved from https://www.reuters.com/graphics/GLOBAL-ENVIRONMENT/EXTINCT/lbvgggdgevq/.

[96] J. Janicki, Daigle, K., Kiyada, S., and Rao, A. 2022. On the Brink. Retrieved October, 2024 from https://www.reuters.com/graphics/GLOBAL-ENVIRONMENT/EXTINCT/lbvgggdgevq/.

[97] Kate Finn Jeff Johnson. 2017. *Designing User Interfaces for an Aging Population : Towards Universal Design*. Morgan Kaufmann,

[98] Robin Jeffries, James R. Miller, Cathleen Wharton and Kathy M. Uyeda. 1991. User Interface Evaluation in the Real World: A Comparison of Four Techniques. *Conference on Human Factors in Computing Systems - Proceedings* 91, c, 119-124. http://dx.doi.org/10.1145/108844.108862

[99] Ben Spraggon Julian Fell, Matt Liddy, Tim Leslie. 2022. What's Your Personal Rate of Inflation? Retrieved 17/07, 2023 from https://www.abc.net.au/news/2022-11-16/cost-of-living-personal-rate-of-inflation/101326768.

[100] D. Wong and M. Duhalde K. Lee. 2023. America's Debt: What Us$31 Trillion Looks Like. Retrieved 2024 from

[101] Nam Wook Kim, Zoya Bylinskii, Michelle A. Borkin, Aude Oliva, Krzysztof Z. Gajos and Hanspeter Pfister. 2015. A Crowdsourced Alternative to Eye-Tracking for Visualization Understanding. In *Proceedings of 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, Seoul, Republic of Korea, 1349–1354. http://dx.doi.org/10.1145/2702613.2732934

[102] Andy J. King, Jakob D. Jensen, LaShara A. Davis and Nick Carcioppolo. 2014. Perceived Visual Informativeness (Pvi): Construct and Scale Development to Assess Visual Information in Printed Materials. *Journal of Health Communication* 19, 10, 1099-1115 %U http://www.tandfonline.com/doi/abs/1010.1080/10810730.10812013.10878004.

[103] B Kitchenham and S Pfleeger. 2002. Principles of Survey Research Part 3: Constructing a Survey Instrument. *ACM Special Interest Group on Software Engineering* 27, 2, 20 - 24.

[104] Cole Nussbaumer Knaflic. 2015. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley %* http://doi.wiley.com/10.1002/tdm_license_1.1 %@ 978-1-119-00225-3 978-1-119-05525-9 %U https://onlinelibrary.wiley.com/doi/book/10.1002/9781119055259,

[105] H Kong, Z Liu and K Karahalios. Year. Trust and Recall of Information across Varying Degrees of Title-Visualization Misalignment. In *Proceedings of Conference on Human Factors in Computing Systems*.

[106] Terry K. Koo and Mae Y. Li. 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine* 15, 2, 155-163.

[107] Robert Kosara and Jock Mackinlay. 2013. Storytelling: The Next Step for Visualization. *Computer* 46, 5, 44 -50. http://dx.doi.org/10.1109/MC.2013.36

[108] Robert Kosara and Jock D. Mackinlay. 2013. Storytelling: The Next Step for Visualization. *Computer* 46, 44-50.

[109] H. Lam, E. Bertini, P. Isenberg, C. Plaisant and S. Carpendale. 2012. Empirical Studies in Information Visualization: Seven Scenarios. *IEEE Transactions on Visualization and Computer Graphics* 18, 9, 1520-1536.

[110] Heidi Lam. 2008. A Framework of Interaction Costs in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics* 14, 6, 1149–1156.

[111] Grace Lau. 2024. Why Interactive Content Is the Key to Engaging Young Audiences. Retrieved 11 September, 2024 from https://blog.scoop.it/2024/01/25/why-interactive-content-is-the-key-to-engaging-young-audiences/.

[112] Bongshin Lee, Nathalie Henry Riche, Petra Isenberg and Sheelagh Carpendale. 2015. More Than Telling a Story: Transforming Data into Visually Shared Stories. *IEEE Computer Graphics and Applications* 35, 5, 84-90. http://dx.doi.org/https://doi.org/10.1109/MCG.2015.99

[113] Crystal Lee, Tanya Yang, Gabrielle Inchoco, Graham M Jones and Arvind Satyanarayan. Year. Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online. In *Proceedings of CHI*. http://dx.doi.org/10.1145/3411764.3445211

[114] Johannes Liem, Charles Perm and Jo Wood. 2020. Structure and Empathy in Visual Data Storytelling: Evaluating Their Influence on Attitude. *Computer Graphics Forum* 39, 3, 277-289. http://dx.doi.org/10.1111/cgf.13980

[115] W Lyu and GL Wehby. 2020. Community Use of Face Masks and Covid-19: Evidence from a Natural Experiment. *Health Affairs,* 39, 8, 1419-1425.

[116] M Mast M Burmester, R Tille, W Weber, . Year. How Users Perceive and Use Interactive Information Graphics: An Exploratory Study. In *Proceedings of Proceedings of the International Conference on Information Visualisation*.

[117] Kwan Liu Ma, Isaac Liao, Jennifer Frazier, Helwig Hauser and Helen Nicole Kostis. 2012. Scientific Storytelling Using Visualization. *IEEE Computer Graphics and Applications* 32, a, 12-19. http://dx.doi.org/10.1109/MCG.2012.24

[118] Sorapure Madeleine. 2022. Anthropographics in Covid-19 Simulations. *Big Data & Society* 9, 1, 1-9.

[119] Narges Mahyar, Sung-Hee Kim and Bum Chul Kwon. Year. Towards a Taxonomy for Evaluating User Engagement in Information Visualization. In *Proceedings of IEEE VIS 2015*. Workshop.

[120] Gregory Maio and Geoffrey Haddock. 2010. *The Psychology of Attitudes and Attitude Change*. SAGE Publications Ltd, Los Angeles.

[121] Azam Majooni, Mona Masood and Amir Akhavan. 2018. An Eye-Tracking Study on the Effect of Infographic Structures on Viewer's Comprehension and Cognitive Load. *Information Visualization* 17, 3, 257-266. http://dx.doi.org/10.1177/1473871617701971

[122] Richard E. Mayer. 2005. Cognitive Theory of Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning*, Richard Mayer Ed. Cambridge University Press, 31-48.

[123] Riccardo Mazza. 2009. Introduction to Visual Representations. In *Introduction to Information Visualization*, Springer London, London, 1-15.

[124] S. McKenna, N. Henry Riche, B. Lee, J. Boy and M. Meyer. 2017. Visual Narrative Flow: Exploring Factors Shaping Data Visualization Story Reading Experiences. *Computer Graphics Forum* 36, 3, 377-387. http://dx.doi.org/10.1111/cgf.13195

[125] Tanja Mercun. 2014. Evaluation of Information Visualization Techniques: Analysing User Experience with Reaction Cards. *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*. http://dx.doi.org/10.1145/2669557.2669565

[126] A Meschtscherjakov, M Gärtner, A Mirnig, C Rödel and M Tscheligi. Year. The Persuasive Potential Questionnaire (Ppq): Challenges, Drawbacks, and Lessons Learned. In *Proceedings of Persuasive Technology: 11th International Conference*.

[127] Sarah Mittenentzwei, Laura A. Garrison, Eric Mörth, Kai Lawonn, Stefan Bruckner, Bernhard Preim and Monique Meuschke. 2023. Investigating User Behavior in Slideshows and Scrollytelling as Narrative Genres in Medical Visualization. *Computers & Graphics*. http://dx.doi.org/10.1016/j.cag.2023.06.011

[128] L Morais, Y Jansen, N Andrade and P Dragicevic. 2020. Showing Data About People: A Design Space of Anthropographics. *IEEE Transactions on Visualization and Computer Graphics* 28, 3, 1661-1679.

[129] Julia Moritz, Hauke S. Meyerhoff, Claudia Meyer-Dernbecher and Stephan Schwan. 2018. Representation Control Increases Task Efficiency in Complex Graphical Representations. *PLOS ONE* 13, 4, e0196420. http://dx.doi.org/10.1371/journal.pone.0196420

[130] Goodreau; Pollock; Birnbaum; Hamilton; Morris. 2020. Can't I Please Just Visit One Friend? Retrieved September, 2023 from http://statnet.org/COVID-JustOneFriend/.

[131] Tamara Munzner. 2014. *Visualization Analysis and Design*. A K Peters/CRC Press,

[132] Elizabeth L. Murnane, Xin Jiang, Anna Kong, Michelle Park, Weili Shi, Connor Soohoo, Luke Vink, Iris Xia, Xin Yu, John Yang-Sammataro, Grace Young, Jenny Zhi, Paula Moya and James A. Landay. 2020. Designing Ambient Narrative-Based Interfaces to Reflect and Motivate Physical Activity. *Conference on Human Factors in Computing Systems - Proceedings*, 1-14. http://dx.doi.org/10.1145/3313831.3376478

[133] Christophe Hurter Nathalie Henry Riche, Nicholas Diakopoulos, Sheelagh Carpendale. 2018. *Data-Driven Storytelling*. CBC Press,

[134] Jakob Nielsen. 1994. *Usability Inspection Methods*. John Wiley and Sons, New York.

[135] Amy Nivette, Denis Ribeaud, Aja Murray, Aannekatrin Steinhoff, Laura Bechtiger, Urs Hepp, Lilly Shanahan and Manuel Eisner. 2021. Non-Compliance with Covid-19-Related Public Health Measures among Young Adults in Switzerland: Insights from a Longitudinal Cohort Study. *Social Science & Medicine* 268, 1 -9. http://dx.doi.org/10.1016/j.socscimed.2020.113370

[136] C. North. 2006. Toward Measuring Visualization Insight. *IEEE Computer Graphics and Applications* 26, 3, 6-9. http://dx.doi.org/10.1109/MCG.2006.70

[137] 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)*, 11-18. http://dx.doi.org/10.1109/BELIV.2018.8634072

[138] Heather L. O'Brien, Elaine G. Toms, E. Kevin Kelloway and Elizabeth Kelley. 2009. Developing and Evaluating a Reliable Measure of User Engagement. *Proceedings of the American Society for Information Science and Technology* 45, 1, 1-10. http://dx.doi.org/10.1002/meet.2008.1450450258

[139] Jeeyun Oh, Hayoung Sally Lim, Jacob Copple and Emily Chadraba. 2018. Harnessing the Persuasive Potential of Data: The Combinatory Effects of Data Visualization and Interactive Narratives on Obesity Perceptions and Policy Attitudes. *Telematics and Informatics* 35, 6, 1755-1769. http://dx.doi.org/10.1016/j.tele.2018.05.004

[140] World Health Organisation. 2019. Non-Pharmaceutical Public Health Measures for Mitigating the Risk and Impact of Epidemic and Pandemic Influenza. 2019.

[141] P. Parsons. 2022. Understanding Data Visualization Design Practice. *IEEE Transactions on Visualization & Computer Graphics* 28, 01, 665-675. http://dx.doi.org/10.1109/TVCG.2021.3114959

[142] Cornelia Pechmann, Guangzhi Zhao, Marvin Goldberg and Ellen Reibling. 2003. What to Convey in Antismoking Advertisements for Adolescents: The Use of Protection Motivation Theory to Identify Effective Message Themes. *Journal of Marketing* 67, 2, 1–18. http://dx.doi.org/10.1509/jmkg.67.2.1.18607

[143] Evan Peck, Sofia Ayuso and Omar El-Etr. Year. Data Is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania. In *Proceedings of Conference on Human Factors in Computing Systems*. http://dx.doi.org/doi.org/10.1145/3290605.3300474

[144] Wei Peng, Mira Lee and Carrie Heeter. 2010. The Effects of a Serious Game on Role-Taking and Willingness to Help. *Journal of Communication* 60, 4, 723-742. http://dx.doi.org/10.1111/j.1460-2466.2010.01511.x

[145] Evanthia Dimara; Charles Perin. 2020. What Is Interaction for Data Visualization? *IEEE Transactions on Visualization and Computer Graphics* 26, 1, 119-129. http://dx.doi.org/10.1109/TVCG.2019.2934283

[146] Richard Petty and John Cacioppo. 1986. The Elaboration Likelihood Model of Persuasion. *Advances in experimental social psychology* 19, 123-204. http://dx.doi.org/10.1016/S0065-2601(08)60214-2

[147] iStock Photo. Retrieved 15 May 2023, from https://www.istockphoto.com/.

[148] R W Picard. 1997. *Affective Computing*. MIT Press, Cambridge.

[149] Prolific. Retrieved 15 May, 2023 from https://prolific.co/.

[150] S. Rendgen. 2012. *Information Graphics*. Taschen,

[151] J Reynolds, M Pilling and T Marteau. 2018. Communicating Quantitative Evidence of Policy Effectiveness and Support for the Policy : Three Experimental Studies. *Social Science & Medicine*, 218, 1-12. http://dx.doi.org/10.1016/j.socscimed.2018.09.037

[152] Mark Rice, Ranieri Koh, Quintessence Liu, Qixiang He, Marcus Wan, Vanessa Yeo, Jamie Ng and Wah Pheow Tan. 2013. Comparing Avatar Game Representation Preferences across Three Age Groups. In *Proceedings of CHI '13 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, Paris, France, 1161–1166. http://dx.doi.org/10.1145/2468356.2468564

[153] Mark Rice, Lih Jie Yau, Jeremy Ong, Marcus Wan and Jamie Ng. 2012. Intergenerational Gameplay: Evaluating Social Interaction between Younger and Older Players. In *Proceedings of CHI '12 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, Austin, Texas, USA, 2333–2338. http://dx.doi.org/10.1145/2212776.2223798

[154] Nathalie Henry Riche, Christophe Hurter, Nicholas Diakopoulos and Sheelagh Carpendale. 2018. *Data-Driven Storytelling*. A K Peters CRC Press, Boca Raton, Florida.

[155] Lauel Riek, Tal-chen Rabinowitch, Bhismadev Chakrabarti and Peter Robinson. Year. How Anthropomorphism Affects Empathy toward Robots. In *Proceedings of HRI*. http://dx.doi.org/10.1145/1514095.1514158

[156] Nicole Robitaille, Nina Mazar, Claire Tsai, Aver Haviv and Elisabeth Hardy. 2021. Increasing Organ Donor Registrations with Behavioral Interventions: A Field Experiment. *Journal of Marketing* 85, 3, 168-183. http://dx.doi.org/10.1177/0022242921990070

[157] Milad Rogha, Subham Sah, Alireza Karduni, Douglas Markant and Wenwen Dou. 2024. The Impact of Elicitation and Contrasting Narratives on Engagement, Recall and Attitude Change with News Articles Containing Data Visualization. *IEEE Transactions on Visualization and Computer Graphics* PP, 1-14. http://dx.doi.org/10.1109/TVCG.2024.3355884

[158]

[159] M Ryan. 2004. Will New Media Produce New Narratives? In *Narrative across Media: The Languages of Storytelling*, University of Nebraska Press, 337-358.

[160] Caitlin Sadowski and Sri Kurniawan. 2011. Heuristic Evaluation of Programming Language Features: Two Parallel Programming Case Studies. *Proceedings of the 3rd ACM SIGPLAN workshop on Evaluation and usability of programming languages and tools*, 9-14. http://dx.doi.org/https://dl.acm.org/doi/10.1145/2089155.2089160

[161] Yumiko Sakamoto, Samar Sallam, Aaron Salo, Jason Leboe-Mcgowan and Pourang Irani. Year. Persuasive Data Storytelling with a Data Video during Covid-19 Infodemic: Affective Pathway to Influence the Users' Perception About Contact Tracing Apps in Less Than 6 Minutes. In *Proceedings of IEEE Pacific Visualization Symposium*.

[162] Bahador Saket, Alex Endert and John Stasko. 2016. Beyond Usability and Performance: A Review of User Experience-Focused Evaluations in Visualization. *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, 133-142. http://dx.doi.org/https://dl.acm.org/doi/10.1145/2993901.2993903

[163] Beatriz Sousa Santos, Beatriz Quintino Ferreira and Paulo Dias. 2016. Using Heuristic Evaluation to Foster Visualization Analysis and Design Skills. *IEEE Computer Graphics and Applications* 36, 1, 86-90.

[164] Arvind Satyanarayan and Jeffrey Heer. 2014. Authoring Narrative Visualizations with Ellipsis 33, 3, 361-370. http://dx.doi.org/10.1111/cgf.12392

[165] Leon Schiffman and Leslie Kanuk. 2004. *Consumer Behaviour 8th Ed*. Pearson Education, New York.

[166] Denise Schoenbachler and Tommy Whittler. 1996. Adolescent Processing of Social and Physical Threat Communications. *Journal of Advertising* 25, 4, 37–54.

[167] D.A. Schön. 1992. *The Reflective Practitioner: How Professionals Think in Action (1st Ed.)*. Routledge,

[168] M Scriven. 1967. The Methodology of Evaluation. In *Perspectives of Curriculum Evaluation*, Rand McNally, 39 - 83.

[169] Edward Segel and Jeffery Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics* 16, 6, 1139-1148. http://dx.doi.org/10.1109/TVCG.2010.179

[170] Doris Seyser and Michael Zeiller. 2018. Scrollytelling – an Analysis of Visual Storytelling in Online Journalism. *2018 22nd International Conference Information Visualisation (IV)*, 401-406. http://dx.doi.org/10.1109/iV.2018.00075

[171] F Shearer, C Bennett, H Vally, J McCaw and N Golding. 2022. New Covid Variants May Be More Transmissible but That Doesn't Mean the R0 – or Basic Reproduction Number – Has Increased. Retrieved September, 2022 from https://theconversation.com/new-covid-variants-may-be-more-transmissible-but-that-doesnt-mean-the-r0-or-basic-reproduction-number-has-increased-186826.

[172] Yang Shi, Tian Gao, Xiaohan Jiao and Nan Cao. 2023. Breaking the Fourth Wall of Data Stories through Interaction. *IEEE Transactions on Visualization and Computer Graphics* 29, 1, 972-982. http://dx.doi.org/10.1109/TVCG.2022.3209409

[173] Eunha Shim, Gretchen Chapman, Jeffery Townsend and Alison Galvani. 2012. The Influence of Altruism on Influenza Vaccination Decisions. *J R Soc Interface* 9, 74, 2234-2243. http://dx.doi.org/10.1098/rsif.2012.0115

[174] B. Shneiderman. Year. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of Proceedings 1996 IEEE Symposium on Visual Languages*. 336-343. http://dx.doi.org/10.1109/VL.1996.545307

[175] David Spiegelhalter, Mike Pearson and Ian Short. 2011. Visualizing Uncertainty About the Future. *Science* 333, 6048, 1393-1400 %U https://www.science.org/doi/1310.1126/science.1191181.

[176] Sharon Steinemann, Elisa Mekler and Klau Opwis. Year. Increasing Donating Behavior through a Game for Change: The Role of Interactivity and Appreciation. In

Proceedings of 2015 CHI conference on Computer-Human Interaction in Play (CHI PLAY '15). 319-329. http://dx.doi.org/10.1145/2793107.2793125

[177] Robert Kosara Steve Haroz, Steven L. Franconeri. Year. Isotype Visualization – Working Memory, Performance, and Engagement with Pictographs. In *Proceedings of CHI '15*.

[178] Harry Stevens. 2020. Why Outbreaks Like Coronavirus Spread Exponentially, and How to "Flatten the Curve". Retrieved September, 2023 from https://www.washingtonpost.com/graphics/2020/world/corona-simulator/.

[179] Brian W. Sturm. 1999. The Enchanted Imagination: Storytelling's Power to Entrance Listeners. *American Library Association.org* 2.

[180] A Sutcliffe. 2009. Designing for User Engagement: Aesthetic and Attractive User Interfaces. *Synthesis Lectures on Human-Centered Informatics 2,* 1, 1, 1–55.

[181] Alvin Tarrell, Ann Fruhling, Rita Borgo, Camilla Forsell, Georges Grinstein and Jean Scholtz. 2014. Toward Visualization-Specific Heuristic Evaluation. In *Proceedings of Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*. Association for Computing Machinery, Paris, France, 110–117. http://dx.doi.org/10.1145/2669557.2669580

[182] Alvin Tarrell, Ann Fruhling, Rita Borgo, Camilla Forsell, Georges Grinstein and Jean Scholtz. 2014. Toward Visualization-Specific Heuristic Evaluation. *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, 110-117. http://dx.doi.org/https://dl.acm.org/doi/10.1145/2669557.2669580

[183] Richard Thaler and Cass R Sunstein. 2008. *Nudge : Improving Decisions About Health, Wealth, and Happiness*. Yale University Press, New Haven.

[184] Joshua Byrd Tim Leslie, Nathan Hoad. 2019. See How Global Warming Has Changed the World since Your Childhood. Retrieved 11 September, 2024 from https://www.abc.net.au/news/2019-12-06/how-climate-change-has-impacted-your-life/11766018.

[185] TopTal. 2019. Age before Beauty – a Guide to Interface Design for Older Adults. Retrieved April, 2024 from https://www.toptal.com/designers/ui/ui-design-for-older-adults.

[186] M Tory and T Moller. 2005. Evaluating Visualizations : Do Expert Reviews Work ? *IEEE Computer Graphics and Applications* 25, 5, 8 - 11.

[187] E.R. Tufte. 2001. *The Visual Display of Quantitative Information Aesthetics and Technique in Data Graphical Design.* Graphic Press,

[188] B Tversky, J Morrison and M Betrancourt. 2002. Animation : Can It Facilitate ? *Human-Computer Studies* 57, 247–262.

[189] Heli Väätäjä, Jari Varsaluoma, Tomi Heimonen, Katariina Tiitinen, Jaakko Hakulinen, Markku Turunen, Harri Nieminen and Petri Ihantola. 2016. Information Visualization Heuristics in Practical Expert Evaluation. *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, 36-43. http://dx.doi.org/https://doi.org/10.1145/2993901.2993918

[190] I; Chen Viola, M; Isenberg, T. 2020. Visual Abstraction. In *Foundations of Data Visualization*, Springerlink, 15-37.

[191] Viscomm. 2023. Retrieved September, 2023 from https://www.viscomm.io/.

[192] VisInPractice. 2021. Retrieved September, 2023 from http://ieeevis.org/year/2021/info/visinpractice.

[193] Emily Wall, Meeshu Agnihotri, Laura Matzen, Kristin Divis, Michael Haass, Alex Endert and John Stasko. 2019. A Heuristic Approach to Value-Driven Evaluation of Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 25, 1, 491-500.

[194] Wibke Weber. 2020. *Exploring Narrativity in Data Visualization in Journalism*. Amsterdam University Press,

[195] Eric N. Wiebe, Allison Lamb, Megan Hardy and David Sharek. 2014. Measuring Engagement in Video Game-Based Environments: Investigation of the User Engagement Scale. *Computers in Human Behavior* 32, 123-132. http://dx.doi.org/https://doi.org/10.1016/j.chb.2013.12.001

[196] Jacqui Wise. 2022. Covid-19: Two Million People in the Uk Are Estimated to Be Experiencing Long Covid,. *The BMJ* 377, 1391. http://dx.doi.org/10.1136/bmj.o1391

[197] Nathan Yau. 2019. A Day in the Life of Women and Men. Retrieved September, 2022 from https://flowingdata.com/2019/03/06/women-men-timeuse/.

[198] R. Jordan Crouser and Ali Sarvghad Zack While. 2024. Gerontovis: Data Visualization at the Confluence of Aging. *Computer Graphics Forum* 43, 3. http://dx.doi.org/https://doi.org/10.1111/cgf.15101

[199] Carlo Zapponi. 2014. Has the Cost of Human Life Changed? Retrieved September, 2024 from https://humancost.carlo.im.

[200] Yixuan Zhang, Yifan Sun, Lace Padilla, Sumit Barua, Enrico Bertini and Andrea G Parker. Year. Mapping the Landscape of Covid-19 Crisis Visualizations. In *Proceedings of CHI '21*. 1-23. http://dx.doi.org/10.1145/3411764.3445381

[201] Ying Zhu. Year. Measuring Effective Data Visualization. In *Proceedings of Advances in Visual Computing*. 652-661.

[202] Shoshana Zuboff. 2019. The Reality Business. In *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.*, Public Affairs, New York, 199-232.

[203] Torre Zuk, Lothar Schlesier, Petra Neumann, Mark S. Hancock and Sheelagh Carpendale. 2006. Heuristics for Information Visualization Evaluation. In *Proceedings of 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*. Association for Computing Machinery, Venice, Italy, 1–6. http://dx.doi.org/10.1145/1168149.1168162

Appendices

Appendix A - Systematic Literature Review

Collected Literature

S1	Buljan, I., Malički, M., Wager, E., Puljak, L., Hren, D., Kellie, F., Marušić, A. (2018). “No difference in knowledge obtained from infographic or plain language summary of a Cochrane systematic review: three randomized controlled trials.” Journal of Clinical Epidemiology, 97, 86–94. https://doi.org/10.1016/j.jclinepi.2017.12.003
S2	Borgo, R., Abdul-Rahman, A., Mohamed, F., Grant, P. W., Reppa, I., Floridi, L., & Chen, M. (2012). “An empirical study on using visual embellishments in visualisation.” IEEE Transactions on visualisation and Computer Graphics, 18(12), 2759–2768. https://doi.org/10.1109/TVCG.2012.197
S3	Ancker, J. S., Chan, C., & Kukafka, R. (2009). “Interactive graphics for expressing health risks: Development and qualitative evaluation.” Journal of Health Communication, 14(5), 461–475. https://doi.org/10.1080/10810730903032960
S4	Locoro, A., Cabitza, F., Actis-Grosso, R., & Batini, C. (2017). “Static and interactive infographics in daily tasks: A value-in-

	use and quality of interaction user study." Computers in Human Behavior, 71, 240–257. https://doi.org/10.1016/j.chb.2017.01.032
S5	Moritz, J., Meyerhoff, H. S., Meyer-Dernbecher, C., & Schwan, S. (2018). "Representation control increases task efficiency in complex graphical representations." PLoS ONE, 13(4), 1–15. https://doi.org/10.1371/journal.pone.0196420
S6	Peck, E. M., Ayuso, S. E., & El-Etr, O. (2019). Data is personal: Attitudes and perceptions of data visualisation in rural Pennsylvania. Conference on Human Factors in Computing Systems - Proceedings, 1–12. https://doi.org/10.1145/3290605.3300474
S7	Newell, R., Dale, A., & Winters, C. (2016). "A picture is worth a thousand data points: Exploring visualisations as tools for connecting the public to climate change research." Cogent Social Sciences, 2(1), 1–23. https://doi.org/10.1080/23311886.2016.1201885
S8	Burmester, M., Mast, M., Tille, R., & Weber, W. (2010). "How users perceive and use interactive narrative visualisation: An exploratory study." Proceedings of the International Conference on Information visualisation, 361–368. https://doi.org/10.1109/IV.2010.57

S9	Zwinger, S., Langer, J., & Zeiller, M. (2017). “Acceptance and usability of interactive infographics in online newspapers.” Proceedings - 2017 21st International Conference Information visualisation, IV 2017, 176–181. https://doi.org/10.1109/iV.2017.65
S10	Rasch, T., & Schnotz, W. (2009). “Interactive and non-interactive pictures in multimedia learning environments: Effects on learning outcomes and learning efficiency.” Learning and Instruction, 19(5), 411–422. https://doi.org/10.1016/j.learninstruc.2009.02.008
S11	Sudakov, I., Bellsky, T., Usenyuk, S., & Polyakova, V. V. (2016). “Infographics and Mathematics: A Mechanism for Effective Learning in the Classroom.”s Primus, 26(2), 158–167. https://doi.org/10.1080/10511970.2015.1072607
S12	de Haan, Y., Kruikemeier, S., Lecheler, S., Smit, G., & van der Nat, R. (2017). “When Does an Infographic Say More Than a Thousand Words?” Journalism Studies, 19(9), 1293–1312. https://doi.org/10.1080/1461670x.2016.1267592
S13	Bateman, S., Mandryk, R. L., Gutwin, C., Genest, A., McDine, D., & Brooks, C. (2010). “Useful junk? The effects of visual embellishment on comprehension and memorability of charts.” Conference on Human Factors in Computing

	Systems - Proceedings, 4, 2573–2582. https://doi.org/10.1145/1753326.1753716
S14	Boy, J., Detienne, F., & Fekete, J.-D. (2015). Storytelling in Information visualisations. 1449–1458. https://doi.org/10.1145/2702123.2702452
S15	Harrison, L., Reinecke, K., & Chang, R. (2015). Infographic Aesthetics. 1187–1190. https://doi.org/10.1145/2702123.2702545
S16	Lyra, K. T., Isotani, S., Reis, R. C. D., Marques, L. B., Pedro, L. Z., Jaques, P. A., & Bitencourt, I. I. (2016). Infographics or Graphics+Text: Which material is best for robust learning? Proceedings - IEEE 16th International Conference on Advanced Learning Technologies, ICALT 2016, 366–370. https://doi.org/10.1109/ICALT.2016.83
S17	Boy, J., Pandey, A. V., Emerson, J., Satterthwaite, M., Nov, O., & Bertini, E. (2017). Showing People Behind Data. 5462–5474. https://doi.org/10.1145/3025453.3025512
S18	Oh, J., Lim, H. S., Copple, J. G., & Chadraba, E. K. (2018). Harnessing the persuasive potential of data: The combinatory effects of data visualisation and interactive narratives on obesity perceptions and policy attitudes.

	Telematics and Informatics, 35(6), 1755–1769. https://doi.org/10.1016/j.tele.2018.05.004
S19	Damman, O. C., Vonk, S. I., Haak, M. J. Van Den, Hooijdonk, C. M. J. Van, & Timmermans, D. R. M. (2018). Patient Education and Counseling The effects of infographics and several quantitative versus qualitative formats for cardiovascular disease risk , including heart age , on people ' s risk understanding. Patient Education and Counseling, 101(8), 1410–1418. https://doi.org/10.1016/j.pec.2018.03.015
S20	Crick, K., & Hartling, L. (2015). Preferences of Knowledge Users for Two Formats of Summarizing Results from Systematic Reviews : Infographics and Critical Appraisals. 47, 1–9. https://doi.org/10.1371/journal.pone.0140029
S21	Farinella, M., Nathalie, D. M., Riche, H., Bach, B., Wang, Z., & Wang, S. (2019). Comparing Effectiveness and Engagement of Data Comics and Infographics. 1–12.
S22	Lee, E., & Kim, Y. W. (2016). Effects of infographics on news elaboration , acquisition , and evaluation : Prior knowledge and issue involvement as moderators. https://doi.org/10.1177/1461444814567982
S23	Reynolds, J. P., Pilling, M., & Marteau, T. M. (2018). Social Science & Medicine Communicating quantitative evidence of

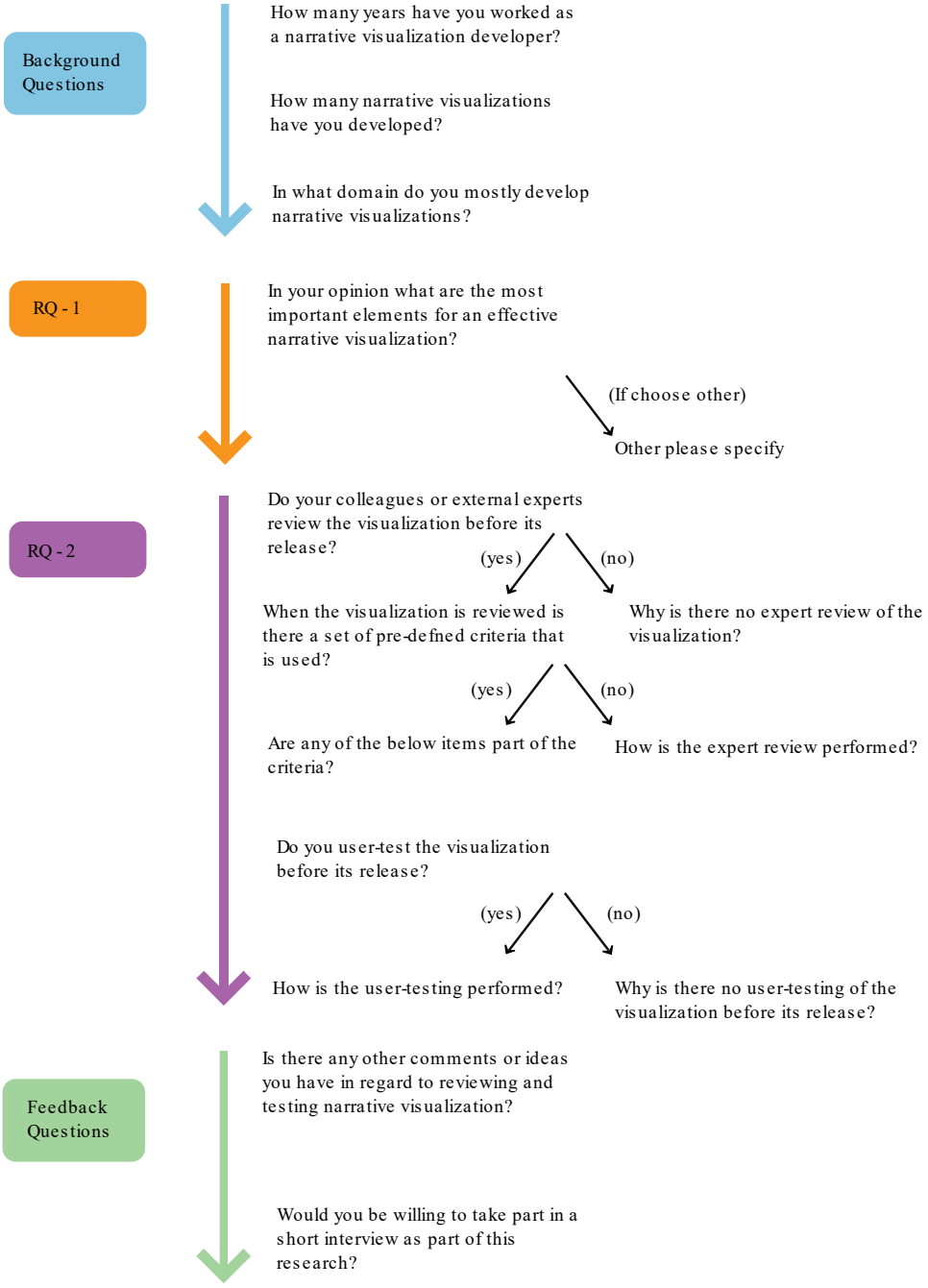
	policy effectiveness and support for the policy : Three experimental studies. Social Science & Medicine, 218(April), 1–12. https://doi.org/10.1016/j.socscimed.2018.09.037
S24	Occa, A., & Suggs, L. S. (2016). Communicating Breast Cancer Screening With Young Women : An Experimental Test of Didactic and Narrative Messages Using Video and Infographics Communicating Breast Cancer Screening With Young Women : An Experimental Test of Didactic and Narrative Messages. UHCM, 21(1), 1–11. https://doi.org/10.1080/10810730.2015.1018611
S25	Gallagher, S. E., Dulain, M. O., Mahony, N. O., Kehoe, C., Mccarthy, F., Morgan, G., … Morgan, G. (2017). Instructor-provided summary infographics to support online learning online learning. Educational Media International, 3987, 1–19. https://doi.org/10.1080/09523987.2017.1362795
S26	Börner, K., Maltese, A., & Balliet, R. N. (2016). Investigating aspects of data visualisation literacy using 20 information visualisations and 273 science museum visitors. 15(3), 198–213. https://doi.org/10.1177/1473871615594652
S27	Lazard, A., & Atkinson, L. (2015). Infographics Center Stage : The Role of Visuals at the Elaboration Likelihood Model ' s Critical Point of Persuasion. https://doi.org/10.1177/1075547014555997

S28	Borkin, M. A., Member, S., Vo, A. A., Bylinskii, Z., Isola, P., Member, S., … Member, S. (2013). What Makes a visualisation Memorable ? IEEE Transactions on visualisation and Computer Graphics, 19(12), 2306–2315. https://doi.org/10.1109/TVCG.2013.234
S29	Lonsdale, S., & Graham, R. (n.d.). Visualizing the The impact terror threat : of communicating security information to the general public using infographics and motion graphics . 37–72.
S30	Majooni, A., Masood, M., & Akhavan, A. (2018). An eye-tracking study on the effect of infographic structures on viewer's comprehension and cognitive load. Information visualisation, 17(3), 257–266. https://doi.org/10.1177/1473871617701971
S31	Diakopoulos, N., Kivran-Swaine, F., & Naaman, M. (2011). Playable data: Characterizing the design space of game-y infographics. Conference on Human Factors in Computing Systems - Proceedings, 1717–1726. https://doi.org/10.1145/1978942.1979193
S32	McKenna, S., Henry Riche, N., Lee, B., Boy, J., & Meyer, M. (2017). Visual Narrative Flow: Exploring Factors Shaping Data visualisation Story Reading Experiences. Computer

	Graphics Forum, 36(3), 377–387. https://doi.org/10.1111/cgf.13195

Appendix B - Survey Study

Figure 15: Flow chart of survey questions



Appendix C - Nudging Study

SPSS outputs

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for ... Lower Bound
1		257	.85	1.657	.103	.64
2		260	.96	1.673	.104	.75
3		255	1.24	1.696	.106	1.03
4		249	1.14	1.626	.103	.94
Total		1021	1.05	1.668	.052	.94
Model	Fixed Effects			1.664	.052	.94
	Random Effects				.089	.76

		95% Confidence Interval for Mean Upper Bound	Minimum	Maximum	Between-Component Variance
1		1.05	-3	5	
2		1.16	-3	5	
3		1.45	-3	5	
4		1.34	-3	5	
Total		1.15	-3	5	
Model	Fixed Effects	1.15			
	Random Effects	1.33			.021

Levene Statistic			df1	df2	Sig.
Difference	Based on Mean	.281	3	1017	.839
	Based on Median	.139	3	1017	.937
	Based on Median and with adjusted df	.139	3	1008.125	.937
	Based on trimmed mean	.244	3	1017	.866

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	24.214	3	8.071	2.916	.033
Within Groups	2814.622	1017	2.768		
Total	2838.836	1020			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Difference
Tukey HSD

(I) Condition	(J) Condition	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-.109	.146	.878	-.49	.27
	3	-.395*	.147	.037	-.77	-.02
	4	-.292	.148	.198	-.67	.09
2	1	.109	.146	.878	-.27	.49
	3	-.285	.147	.209	-.66	.09
	4	-.183	.148	.602	-.56	.20
3	1	.395*	.147	.037	.02	.77
	2	.285	.147	.209	-.09	.66
	4	.103	.148	.900	-.28	.48
4	1	.292	.148	.198	-.09	.67
	2	.183	.148	.602	-.20	.56
	3	-.103	.148	.900	-.48	.28

*. The mean difference is significant at the 0.05 level.

Condition	N	Subset for alpha = 0.05	
		1	2
1	257	.85	
2	260	.96	.96
4	249	1.14	1.14
3	255		1.24
Sig.		.194	.213

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 255.186.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Appendix D – Act Your Age Study

Statistical test results in R One-way ANOVA

```
                Df Sum Sq Mean Sq F value Pr(>F)
Age              3   2426   808.7   2.812  0.038 *
Residuals     2319 666940   287.6
```

Post-Hoc Tests

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Total ~ Age * Condition, data = g_data2)

$Age
             diff       lwr        upr     p adj
MY-Y   -1.3438596 -3.791238 1.10351861 0.4920813
MO-Y   -2.4011007 -4.849554 0.04735259 0.0569581
O-Y    -2.4210526 -4.868431 0.02632562 0.0537764
MO-MY  -1.0572411 -3.505694 1.39121224 0.6833222
O-MY   -1.0771930 -3.524571 1.37018527 0.6700082
O-MO   -0.0199519 -2.468405 2.42850142 0.9999967
```

Levene's Test

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value  Pr(>F)
group   3  2.4589 0.06109 .
      2275
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
        Shapiro-Wilk normality test

data:  aov_residuals
W = 0.99233, p-value = 1.366e-09
```