

Visualization Literacy in the Age of Big Data: Vital Skills for Modern Media Consumption

Mara Solen



Fig. 1: A frequency word cloud made from various definitions of the term “visualization literacy”.

Abstract—Access to information is increasing rapidly, and data visualizations are becoming increasingly common as a way to communicate summaries of large datasets, especially in media designed for the general public. Accompanying this rise in visualization is a need for visualization literacy. In this work we survey the current body of research on visualization literacy to find current research themes and guide future work. We found three themes of visualization literacy: fundamental, believability, and engagement. Believability and engagement are not traditionally discussed in visualization literacy research. We analyzed participant pools for visualization literacy user studies and found them to be lacking in diversity in gender and sex, age, education level, and recruitment country.

Index Terms—Information visualization, literacy, education, survey.

1 INTRODUCTION

Written language has been in use for thousands of years. For the majority of its existence, only a small set of people were able to understand and construct text. These people were likely wealthy and had access to resources and education that allowed them to learn the skills of reading and writing, otherwise known as literacy. Literacy among the general population has been increasing significantly over the past century due to the increasing accessibility of education. However, this change has been less noticeable in families with lower income [25].

Information, including but not limited to written language, has seen an increase in collection and accessibility in recent history. In order to allow people to understand information, particularly that which is domain-specific, not human-readable, or of large scale, it is increasingly common to visualize it. Additionally, information visualizations are

becoming more commonly used by organizations with large audiences like news outlets and social media pages which is making them more accessible to the general public.

With this method of communicating information comes a new form of literacy: visualization literacy. Analogous to written literacy, early on, visualizations were only comprehensible to highly educated individuals, and now with their rising popularity the general population is gaining more exposure. However, visualization literacy in the general public is low [5] despite its prevalence.

Visualization education is primarily accessible to privileged individuals. While some researchers have worked to address this gap in visualization literacy, they often create online courses taught in English [23] that are inaccessible to those without a strong, consistent internet connection and fluency in English. Further, three quarters of the work into visualization education for specific groups focuses on teaching young children despite adults likely struggling as well. For example older adults who have difficulty comprehending novel chart types and prefer simple ones, even if considered less effective by the visualization community [14].

The contributions of this paper are as follows: (1) a survey of existing visualization literacy literature, (2) a set of themes found in the

surveyed literature, (3) a new definition for visualization literacy that encompasses all the themes that were found, and (4) gaps in participants of visualization literacy user studies.

2 RELATED WORK

Despite my best efforts, we were unable to find any existing survey papers on visualization literacy. However, we will discuss some papers that are not primarily survey papers but that included significant background sections.

Boy et al. [4] constructed a definition for visualization literacy by starting with the definition of the general term literacy, defined in the Oxford dictionary as “the ability to read and write”, and modifying the definition to be specific to the topic of visualization. They discuss earlier work in the field, such as that in the cognitive processes involved in reading graphs and how it is an iterative process rather than a straight-forward serial process. A finding to come out of this past work is that a person’s understanding of a visualization depends on their expertise in the topic being visualized. However, this finding is less prevalent in more visually literate people. While this work is well cited, it focuses on a narrow, low-level view of visualization literacy and does not discuss higher level but related skills such as critically evaluating visualizations.

Börner et al. [3] completed a literature review of visualization literacy as part of their work to develop a framework to guide further research. They described much of the previous work and noted that none of it explicitly discussed the construction of visualizations. Similar to Boy et al. [4], the focus of this work is on low-level tasks and categorizations.

Peck et al. [29] took a different approach to visualization literacy. They discuss what makes people living in rural Pennsylvania interested in certain charts and not others and how attention can be improved without employing biased or ineffective visual encodings. These biases may be different or more pronounced in populations that are less literate in visualization than in environments such as universities or large urban centers where design studies are commonly done, even for tools meant for novice users like LineUp [13].

3 PROCESS

There were three stages in our process. The first stage was to collect literature based on seed papers and inclusion criteria. The second stage was to analyze the collected literature for themes using thematic analysis. The third stage was to conduct an analysis of the participant pools used in user studies in the collected literature.

3.1 Paper Collection

For this work, we found relevant literature from different research venues using keyword search and backward and forward chaining from seed papers. We then surveyed and analyzed the papers. The seed papers were *Data is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania* [29] and *A principled way of assessing visualization literacy* [4]. A total of 34 papers that discussed visualization literacy were included. Inclusion criteria was that the paper had some mention of visualization literacy, including using different terms such as “understanding” [34] and “perceptions of” [29] visualization.

After reading the papers selected by the first phase of paper collection, we sought out additional papers on topics that are not typically considered visualization literacy but that have interesting overlap: believability of visualizations and engagement in visualizations. The seed papers for these are *Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online* [22], *Data is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania* [29].

Data is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania [29] was a seed paper in the first stage for finding papers that explicitly mention visualization literacy. In the second stage, it was a seed paper for literature that did not explicitly mention visualization literacy but did discuss the believability of visualizations or engagement in visualizations.

Papers that did not discuss visualization literacy, believability of visualizations, or engagement in visualizations were excluded. *A Comparative Evaluation on Online Learning Approaches using Parallel Coordinate Visualization* [20] discusses visualization literacy but only in the introduction as a segue to visualization education and it did not discuss the believability of visualizations or engagement in visualizations. *Effect of Adaptive Guidance and Visualization Literacy on Gaze Attentive Behaviors and Sequential Patterns on Magazine-Style Narrative Visualizations* [2] used visualization literacy as an independent variable to find ways to support those with low visualization literacy skills, but did not develop the research of visualization literacy itself. *The Correlation between Users’ Cognitive Characteristics and Visualization Literacy* [24] discussed visualization literacy at an individual level using cognitive characteristics which was not the focus of this study.

3.2 Thematic Analysis

We used thematic analysis on the collected papers to find themes. We first used thematic analysis on the definitions of visualization literacy that were used in the surveyed papers, the results of which are in Table 2. We then used thematic analysis on the main body of each paper, the results of which are shown in Table 4 in the appendix.

Our process for thematic analysis started with reading the collected papers and using informal open coding to develop an initial set of themes. We then iterated over the set of themes and grouped them into major and minor themes. The codes used and the final major themes are presented in Figure 2.

3.3 Audience Analysis

Additionally, we analyzed the papers for diversity of participant pools. In the initial reading of the surveyed papers, we noticed potential gaps in the participant pools of user studies. For each of the papers with user studies we collected the provided demographics of the participant pools. We enumerated the groups that were studied and then reported on informal observations made about their gender and sex, age, education level, and recruitment country, providing context for future work.

4 THEMES

We established three themes from all papers surveyed: fundamentals, believability, and engagement. Fundamentals describes how a person interprets and constructs the basics of visualizations. For example, fundamentals includes extracting a piece of data from a visualization. Fundamentals was the most commonly mentioned theme.

Believability describes how a person can be critical or trusting of visualizations. Huynh et al.’s [17] definition of visualization literacy used the word “critical”, and then discussed the ability of their participants to critically think about visualizations. Peck et al. [29] looked at this topic as part of a larger investigation on the reliability of visualizations. Lee et al. [22] also discussed the topic of trust in visualizations.

Engagement describes whether a person is engaged and what makes a person more or less engaged in a visualization. Peck et al. [29] examined engagement in visualization and found results suggesting that personal interest in a chart makes a person more engaged and more likely to rate the usefulness of the chart highly.

4.1 Fundamentals

Fundamentals are likely common due to the connection to the traditional meaning of literacy: the ability to interpret and construct written language. As it has been the most common theme in the existing literature, it has been the most heavily researched and developed.

The term *graph literacy* was used instead of visualization literacy during some early work in the field, such as in “Graph Literacy: A Cross-Cultural Comparison” [12]. Galesic et al. developed a graph literacy scale and used it to compare “the ability to understand graphically presented information” [12] between German and American participants. They found that both countries performed similarly and that, in each country, one third of the population had low graph literacy ability.

In 2014, Boy et al. [4] noted that it was challenging to quantify people’s visualization-reading abilities. They also noted that the idea

Table 1: CD = Completion Date. HR = Hours Required. The milestones of the project, shown as the expected and actual numbers.

Milestone	Expected CD	Expected HR	Actual CD	Actual HR	Notes
Select around 25 papers	Oct. 30th	15 hours	Oct. 30th	10 hours	
Read papers	Nov. 10th	30 hours	Nov. 8th	20 hours	Reading was made more efficient by scanning the papers for relevant sections
Generate paper framework	Nov. 16th	10 hours	Nov. 16th	20 hours	Generating the framework, findings, and contributions of the paper took more time than expected
Complete draft	Dec. 5th	15 hours	Dec. 12th	40 hours	
Create presentation	Dec. 12th	5 hours	Dec. 13th	5 hours	
Finish the paper	Dec. 17th	5 hours	Dec. 17th	10 hours	

Table 2: The definitions used for visualization literacy in the surveyed papers. 11 papers did not include definitions for visualization literacy at all, 12 papers had one definition, and 4 papers had multiple definitions. All unique definitions were extracted and analysed to determine whether or not they mentioned each of the themes. The definitions are first sorted by the number of papers that use them and then sorted alphabetically.

Definition: the ability...	Papers	Fundamentals	Believability	Engagement
and skill to read and interpret visually represented data in and to extract information from data visualizations	[3, 7, 23, 24]	read, interpret, extract		
to make meaning from and interpret patterns, trends, and correlations in visual representations of data	[3, 5, 10, 31]	make meaning, interpret		
to read, comprehend, and interpret graphs	[29, 31]	read, comprehend, interpret		
to use well-established data visualizations (e. g., line graphs) to handle information in an effective, efficient, and confident manner	[4, 11]	handle		
to confidently create and interpret visual representations of data	[1]	interpret, create		
to confidently use a given data visualization to translate questions specified in the data domain into visual queries in the visual domain, as well as interpreting visual patterns in the visual domain as properties in the data domain	[3, 4]	translate, interpreting		
to critically read and construct data visualizations	[17]	read, construct	critically	
to read and construct visual representations to make meaning of data and to support the understanding of datasets through data visualization types (e.g. scatter graph, geo map), data variables (i.e. qualitative, quantitative) and graphic variable types (e.g. shape, size, color)	[30]	read, construct		
[read] charts and graphs	[27]	read		
to understand and appropriately handle data visualizations	[17]	understand, handle		
to use common data visualizations in an efficient and confident manner	[2]	use		

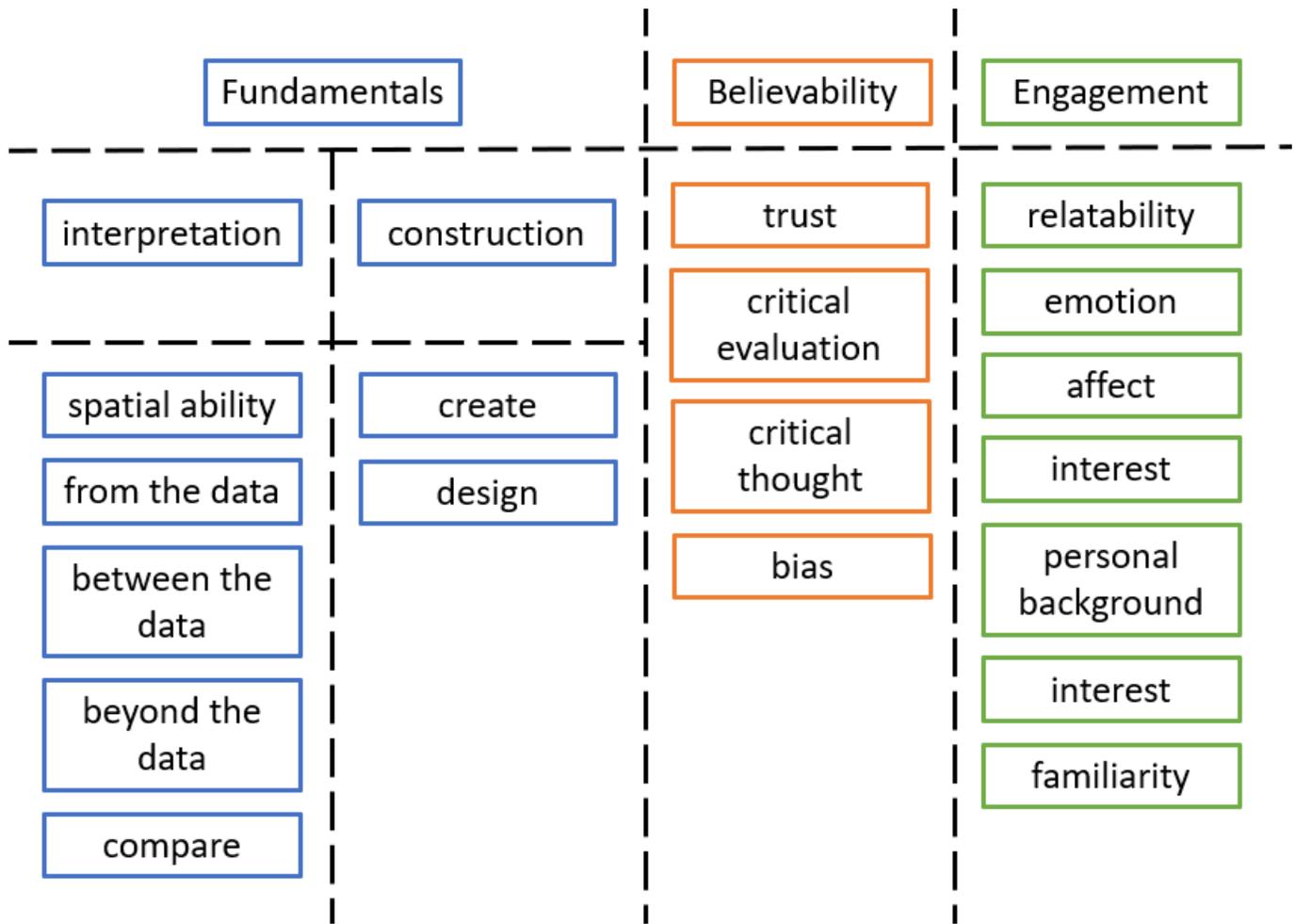


Fig. 2: An overview of the coding process, showing the initial codes (second and third rows) and the themes that eventually came from them (first row).

of interpreting visualizations can be broken into different levels as. For example, having the ability to extract basic information from a chart does not guarantee that someone can identify trends and relationships. In response, they developed a useful method for assessing visualization literacy that visualization designers can use to support the creation of future visualization literacy assessment tools.

After seeing that visualization literacy was being treated as two rigid categories: novice and expert, Maltese et al. [26] investigated how much visualization literacy differs among their population which consisted mostly of university students and faculty. Across their participants, there was not a large difference in visualization literacy. They found that even those with a background in science, technology, engineering, and math (STEM) struggled with basic visualization interpretation.

Lee et al. [23] took the visualization literacy assessment method guidelines proposed by Boy et al. [4] and implemented them into a concrete assessment consisting of a set of visualizations and multiple-choice questions called the Visualization Literacy Assessment Test (VLAT). The test, which follows the assessment method in focusing on interpreting visualizations, was validated by five visualization experts and then tested on just under two hundred participants. The authors found that the tool was reliable and propose that it can be used in future work for quantifying visualization literacy ability and for improving visualization education. Similar assessments have been developed as well, such as the treemap literacy test created by Firat et al. [10]. Interestingly, Firat includes construction in their definition of treemap literacy while Lee does not include it in their definition of visualization literacy.

Children have been a focus of many visualization literacy papers due to the desire to teach visualization skills while people are young. Alper et al. [1] collected visualization methods used in elementary school textbooks to see how children are currently learning about visualization. They found that textbooks commonly used types of visualization that were understudied in the visualization community. They developed a tool to help with the teaching and learning of visualization interpretation with emphasis on teaching the children to understand different levels of abstraction by gradually moving from concrete ideas to abstract ones. The tool was successful, leaving teachers surprised by its effectiveness and contemplating what may have made it as effective as it was. Chevalier et al. [7] also examined visualization literacy in children and make many statements about what is expected, believed to be, or present versus what is lacking in elementary school education. For example, they found that visualizations are used extensively in elementary school classrooms to communicate information or teach other subjects but that there is not an equivalent focus on how to understand them. They also propose that the standard definition of visualization literacy be expanded to include the construction of visualizations, as they report that up until their paper most work in visualization literacy was on the interpretation of visualizations.

Gäbler et al. [11] continued with the work to support children in learning to interpret visualizations by creating an educational game. The game was found to interest the participants, but the authors found that success in the game did not necessarily transfer to real-life settings. Huynh et al. [17] also made an educational game to teach visualization literacy in the classroom. They found similar results to Gäbler et al. [11]

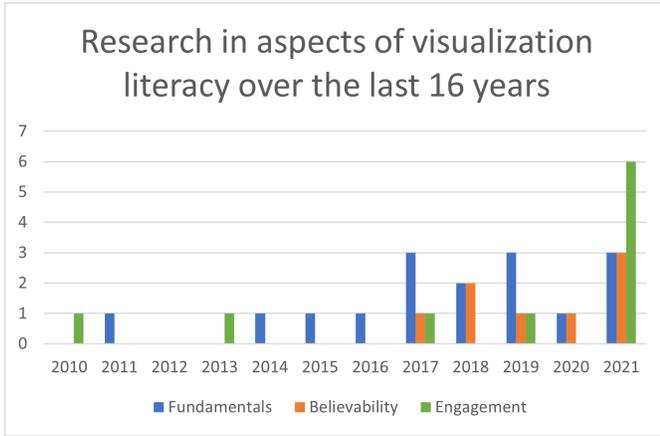


Fig. 3: A grouped bar chart showing the number of papers included in this survey that discussed each of the themes for each year since 2010. Fundamentals papers were found from every except 2010, 2012, and 2013 with an increase beginning in 2017. Believability papers were first found from 2017 with an increase in 2021. Engagement papers started in 2010 and were found in just under half of the years with a significant increase in 2021.

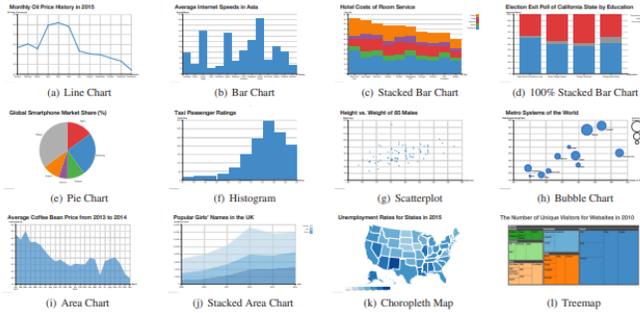


Fig. 4: The charts used in VLAT to assess visualization literacy [23].

in that the results about the children learning visualization skills were inconclusive.

Patients with prostate cancer, typically older men, are another group that has been studied for their understanding of visualizations. Hakone et al. [14] developed a tool, called PROACT, for communicating health risks to these patients. They found that using visualization to convey information was effective in general, especially for those with low numeracy and visual literacy, and additionally that it was more effective to use simple, static visualizations rather than complex, interactive ones.

Sultana et al. [31] focused their efforts on a different group of people: those living in rural Bangladesh. These researchers were particularly interested in how traditional methods for communicating data visually differ from prominent Western methods. They found that the participants commonly used concrete units in their visualizations and that it was common to combine art and data. They suggest that, rather than always focusing on standardizing visualization methods, instead it is more effective to situate them: to develop visualizations with the target audience's background in mind.

Börner et al. [5] evaluated the visualization literacy of the general public. They conducted an investigation on museum visitors who were shown various visualizations and then asked a set of questions. One of those questions was about how to read the visualization, relating to interpretation, and another was about what kind of dataset would warrant that type of visualization, relating to construction. They found that much of the population is unable to interpret the visualizations they showed and that most people could not think of datasets for which the

Exploring your information further

These interactive graphs give you a different way of seeing and thinking about the benefits of different treatments. Feel free to explore these graphs using your mouse.

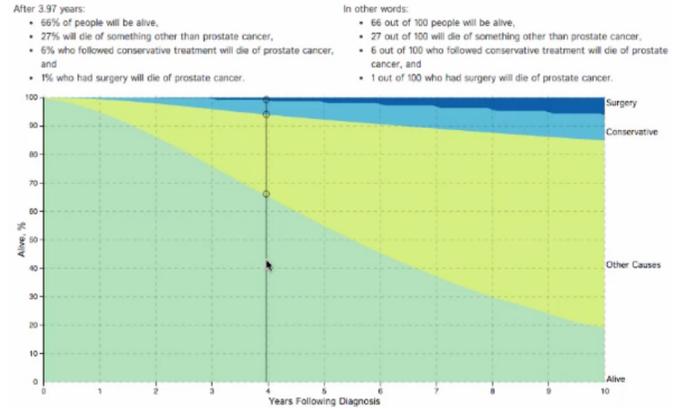
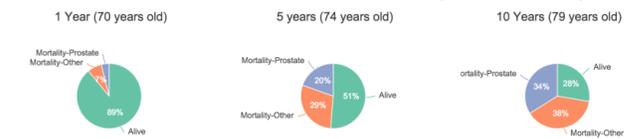


Fig. 5: Part of an early version of PROACT, when the authors were using more complex chart types [14].

What do I do next?

What threats am I facing?

- Your risk of dying from prostate cancer:
- Your risk of dying from other causes:



What are the possible treatments for me?

- Surgery
- Conservative management
- Radiation
- Other treatments

0% risk of dying from prostate cancer in the next 10 years.
34% risk of dying from prostate cancer in the next 10 years.
PROACT does not include this information. (Why?)
PROACT does not include this information. (Why?)

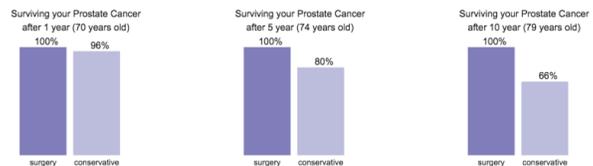


Fig. 6: Part of the final version of PROACT, when the authors were using multiple simpler chart types [14].

less common visualization types could be used.

Börner et al. [3] wrote a visualization literacy framework paper that included a set of definitions, conceptual frameworks, exercises, and assessments to describe visualization literacy. They note that none of the visualization literacy assessments that existed at the time discussed the construction of visualizations, so they thought it important to include discussion of visualization construction in their work. They propose the data visualization literacy framework (DVL-FW) to support further research and assessment design.

This survey of research into the fundamentals of visualization literacy shows that the interpretation of visualization is at the forefront of researcher's minds. Current literature covers the teaching and assessment of visualization interpretation skills and understanding where the differences in these skills come from, however the teaching and assessment of visualization construction is much less common. The majority of the surveyed papers do not consider this aspect or discuss why they did not consider it. Both Börner et al.'s framework paper [3] and Chevalier et al.'s observations and reflections [7] have noted the lack of research on constructing visualizations. Börner et al.'s framework paper addressed this by including it in their framework.

Chevalier et al. [7] state that visualization literacy needs to focus

more on the construction of visualizations. They make an analogy to reading literacy which is taught via both reading *and* writing, and propose that future educational efforts in visualization should include construction. Finally, they suggest that the definition of visualization literacy should be expanded to include this aspect.

4.2 Believability

With increased access to information, it is important to find accurate information and to know what information to trust. Chen et al. have found those with low health literacy are less likely to trust health information from specialists and that the general public may be receiving health information from unqualified sources such as friends or blogs [6]. As a form of information communication, data visualizations are contributing to the trust and distrust of the general public. Visualizations have been found to decrease misconceptions gained from reading texts [28], but can they also cause their own misconceptions? Visualizations can be misleading both intentionally and unintentionally [8].

There has been significant work in believability in the visualization community in recent years. Peck et al. [29] looked into how rural Pennsylvanians, a group who often struggle with economic and infrastructure challenges, perceive visualizations. Participants were shown ten charts and asked to rank them by how personally useful they found them. After the first ranking, the sources of the articles were shown to the participants who were then asked to rank them again. Interestingly, this experiment found that less educated participants were less likely to change their rankings after seeing the sources.

Heyer et al. [16] investigated how eliciting prior knowledge, particularly knowledge around provocative topics, could influence misconceptions in visualization. They conducted an experiment that found that visualization was more persuasive than just text, but that eliciting prior knowledge did not affect persuasiveness. Participants also partook in reflections on the activities in the experiment. They consistently noted that they had a different mental model after the experiment. However, most participants did not doubt the data or the visualizations and those who did doubt the data or the visualizations seemed to have had firsthand experience or prejudice that caused the doubt. This shows that visualizations that are effective at informing opinions ineffective at changing opinions. The authors suggest that the tasks of informing and opinion changing should be considered separately from each other.

To study the difference between how articles with visualizations and articles without visualizations clarify misperceptions, Mena [28] conducted an online experiment. They looked into both the effect on misconceptions of including visualizations and of prior knowledge on the topic. Mena found that including visualizations reduced misperceptions for those with below-average prior knowledge. However, the visualizations in the articles were not made to be misleading and studies on showing manipulated visualizations to those with below-average prior knowledge may be useful future research.

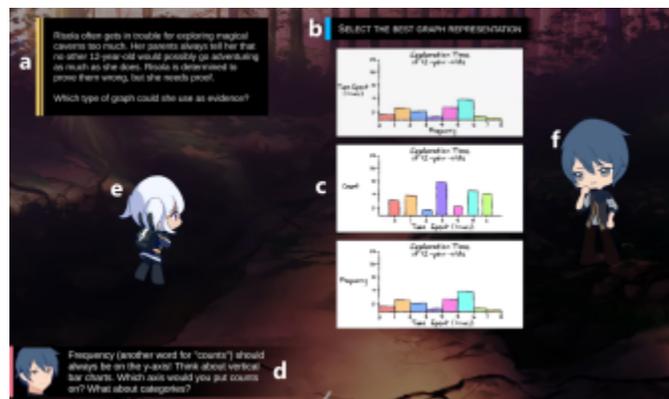


Fig. 7: A screenshot from the narrative educational game designed by Huynh et al. showing dialog and charts [17].

Huynh et al. [17] developed an educational visualization-based game

for children, similar to Gäbler et al. [11], however they also discussed the ability to critically examine and design visualizations. They discuss thinking critically, that is, considering why each visual encoding choice was made, as an important part of both construction and interpretation of visualizations.

One of the ways Jena et al. [18] claims that emergent, or inexperienced, users can be supported in their learning of visualization is by expanding visualization education outside of niche academic communities. Doing this would make knowledge on visualization more accessible, however making visualization courses must be done carefully. Jena et al. mention that being able to understand visualizations is not enough, but that people also need to be able to critically evaluate them.

Mansoor et al. [27] argue that the studies of visualization literacy and bias in visualization are threads of research that should be merged. They show how these threads could be merged in two ways. First, visualization literacy research could be integrated into the study of various types of bias in visualization. Second, including questions about bias in visualization literacy assessments could help determine whether participants are novices or experts. Chevalier et al. [7] make a similar claim. They propose that the definition of visualization literacy should include the ability to critique the representation of the data in a given visualization and understand if it is misleading.

Lee et al. [22] discuss the critical evaluation of visualizations in their analysis of twitter data on anti-mask discourse. They enumerate many ways in which the twitter users are critical of visualizations, from the method of data collection to the visual encodings used. The authors describe how the users have a strong ability to critique visualizations and have high visualization and data literacy. They discuss how being critical of information is a key part of being literate in visualization, but that researchers must be careful who they consider to be illiterate in this aspect as users may be skilled in critical thinking and have different reasons to interpret visualization differently from others. For example, the twitter users in their studies had very different ideas of what made sources reliable, which is outside the scope of visualization literacy.

Considering the critical evaluation of visualizations as part of visualization literacy is supported by the literature. The increase in information available to the public and concern in organizations and individuals manipulating this information has made people from a variety of backgrounds interested in knowing how to detect biases in what they read. This skill of critical evaluation of visualizations has become an important aspect to visualization literacy, and should be considered part of the standard definition in future research.

Further, being able to reproduce visualizations is one way that a person can show a visualization is misleading and it is one of the ways anti-mask twitter users are able to inform others. The addition of critical evaluation to the definition of visualization literacy further supports the addition of construction.

4.3 Engagement

Why do people choose to engage with visualizations? Some people engage because they are asked to and they may be compensated for complying. However, people also engage with visualizations due to a variety of types of personal interest. Some visualizations are nice to look or cover fun topics which draw people in, while others contain crucial data and trends that people feel they need to know. People may also feel pressured to engage in visualizations and information in general, such as a person who is uninterested in watching sports shows but engages in visualizations of post-game summaries to avoid becoming an outcast of their social group. Does engaging with visualization for these different reasons change how a person interprets them?

First, what could cause people to not engage with visualization? One reason someone might not engage is because they are informationally poor. Yu [35] analysed the information practices of this population. They found that the participants failed to engage with information for many reasons. The participants were sometimes uninterested, but they were also blocked from information due to poor financial status, general information literacy, and access to points where the information could be retrieved. Expanding on the lack of skills, participants had

limited knowledge of and experience in how to work with data which led to them being unable to use data to support their decisions. We suspect that the participants would have similar issues engaging with visualizations, as data access is fundamental to visualization.

A significant part of Peck et al.'s [29] work in rural Pennsylvania was investigating which visualizations the participants found most useful to them. The main outcome of the work was referenced in the title: people find visualizations that relate to them personally to be more engaging. This result supports that of Sultana et al. [31] in that visualization research should focus on designing situated visualizations rather than standardized ones in order for users to find the visualizations engaging. Lee et al. [22] saw that anti-mask twitter users were very engaged with data and visualization around mask-related and other COVID mandates. Strong opinions on political topics, particularly those that directly effect the user, seem to cause very strong engagement.



Fig. 8: The station used for constructing and interacting with visualizations based on previously entered data, designed by Pepler et al. [30].

Pepler et al. [30] implemented a data visualization literacy station within a museum and studied how they could get visitors to engage with it. They created an interactive station that allowed users to enter their own personal data and then choose how to visualize it. Users could also see the data and visualizations of other recent users and could compare themselves to others. They also attempted to make the data entry process inclusive of those from different countries and support both individual and group data entry so that people would not go unrepresented and therefore be stopped from engaging. They confirmed the results from Peck et al. that people were more engaged with personalized data and further suggest that viewing visualizations in social settings can increase engagement.

Emotion is another factor in engagement. Harrison et al. [15] conducted an experiment to understand how emotional priming would affect performance with visualization. They ran a study with 963 participants from Amazon's Mechanical Turk (MTurk) where each participant was emotionally primed and then shown a set of simple charts and asked to quickly make judgements. The results showed that emotional priming did have an effect on a participant's judgement and that positive priming increased performance, suggesting that having different reasons for engaging in visualizations may cause different outcomes.

Lan et al. [21] looked at the inverse of Harrison et al. : can visualizations (infographics in particular) be designed to cause specific emotions in readers? The authors first studied the types of emotions that could be produced with infographics and which aspects of infographics caused these emotions. They then created a set of design guidelines specifically for drawing out different emotions from users. We suspect that these guidelines help both infographic designers and infographic readers. They help infographic designers to convey the right feelings and they

may allow informed readers to see how the designer may be attempting to influence them.

Pre-existing emotions and intentionally-emotional visualizations can influence how people interpret charts, but what emotions arise when viewing a typical chart? Van Koningsbruggen et al. [33] found that participants would describe a traditional visualization as "just a graph", indicating that they had not felt emotions while analyzing it. The authors found that participants actively tried to separate their emotions from the visualization, due to it seeming improper and not permitted to involve emotions in data analysis tasks. These participants were asked to view the charts in study conditions and we suspect that there may have been more emotion involved had the participants come across the visualizations naturally and were not being observed.

Some visualizations naturally cause strong emotions, such as those used for communicating health information. Hakone et al. [14] observed that these visualizations, such as PROACT, need to have a carefully designed narrative flow to allow users to make objective and informed health decisions. They found that ensuring a patient is calm before showing information is important.

Huynh et al.'s [17] educational game for learning visualization investigated how to make children more engaged in visualization. While they did not improve the real-world visualization skills of the children, they were able to find that gamification increased engagement. Further, they found that the version of the game with a narrative was more engaging than the one without. We wonder if other age groups would engage more with visualization systems if they had narratives built in.

Engaging with visualization may not be considered a skill in the surveyed literature, but a person's engagement can have an effect on the ways they perceive a visualization. Future research, particularly that which uses visualization literacy assessments, should take into account how engaged in the visualizations the participants are either by developing personalized visualizations or by measuring engagement. Additionally, visualization design studies for novice or general users should put renewed effort into designing situated and engaging visualizations to ensure that visualization literacy research is done independently of engagement.

4.4 A New Definition

The ability to interpret was included in every definition of visualization literacy and makes up the majority of the literature in the field. However, researchers have shown that there are more aspects to visualization literacy than interpretation. Supported by the body of research described above, we propose a new definition of visualization literacy that includes interpretation, construction, believability, and engagement with visualization: the ability to critically interpret and construct engaging visualizations.

It is unlikely that, going forward, a single definition will be unanimously agreed upon and used in research. However, we hope that these themes are at the very least discussed in future work, as they are all vital to understanding how people interact with visualization.

5 AUDIENCE ANALYSIS

16 of the surveyed papers had user studies. The demographics of the participant pools of each individual user study is available in Table 3. The papers that directly contribute to visualization literacy (direct contributors) are analysed first and the others (indirect contributors) second. Findings from the two groups are compared.

5.1 Direct Contributors

Two of the studies that makes direct contributions to visualization literacy recruited participants from Amazon's Mechanical Turk (MTurk). The assessment by Boy et al. [4] does not provide demographics of the participants. A survey of participants in MTurk found that the majority were younger and more well educated than participants from other services [19], implying that MTurk may not recruit a diverse set of participants.

The remaining seven user studies in this category provide more demographics on their participants and will be the focus of this analysis. Five of these seven studies included the gender or sex of the participants

Table 3: Descriptions of participant pools using provided demographics. Papers that explicitly contribute to visualization literacy research are above the double line.

Title	First Author, Year, Citation	Description of Participant Pool
A Principled Way of Assessing Visualization Literacy	Boy, 2014, [4]	40 participants recruited from MTurk for each of multiple experiments.
Data Visualization Literacy: Investigating Data Interpretation Along the Novice-Expert Continuum	Maltese, 2015, [26]	202 participants (54% female, mean age = 25), predominantly college and university undergraduates (68%) and graduate students (9%).
Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors	Börner, 2016, [5]	273 participants, all science museum visitors visiting museums in the United States.
Visualization Literacy at Elementary School	Alper, 2017, [1]	21 participants, 6 in kindergarten and 15 in grade 2 at a French immersion school in an upper-class district in Seattle.
VLAT: Development of a Visualization Literacy Assessment Test	Lee, 2017, [23]	191 participants recruited from MTurk, 105 females and 86 males, ages ranging from 19 to 72 with a mean of 37, all participants had completed a high school education, 42% had a bachelor's degree, 15% had a master's or doctoral degree.
Diagram Safari: A Visualization Literacy Game for Young Children	Gäbler, 2019, [11]	23 participants, all in grade 4, 13 female and 10 male.
Designing Narrative-Focused Role-Playing Games for Visualization Literacy in Young Children	Huynh, 2021, [17]	33 participants, all ages 11 to 13 with a mean of 12.09, 16 female and 17 male.
Cultivating data visualization literacy in museums	Peppler, 2021, [30]	195 participants in 74 groups with an average of 3 people per group. 8% were Latines, 74% were white, 11% reported multiple racial identities, 10% were black, 3% were Asian, 1% selected "other" for racial background, and 1% chose not to provide race information. 53% were female, 44% were male, 3% were non-binary in gender, and 1% chose not to provide gender information. 52% were age 17+, 21% were age 7-10, 11% were age 14-16, 10% were below age 7, 6% were age 11-13, and 1% did not provide age information.
How poor informationally are the information poor?	Yu, 2010, [35]	73 participants, from a variety of locations on the rural-urban scale in north China, 20 of which have specific demographics included in the paper. 17 were male, 3 were female. Ages ranged from 24 to 74 with a mean of 42. Most participants had not completed a senior secondary program and one participant had a college education. Participants had occupations such as farmer (4), taxi driver (1), timber trader (5), rural family inn owner (6), village leader (1), retired headmaster (1), bus driver (1), university technician (1), or were laid off (2).
Influencing Visual Judgment through Affective Priming	Harrison, 2013, [15]	963 participants recruited from MTurk.
PROACT: Iterative Design of a Patient-Centered Visualization for Effective Prostate Cancer Health Risk Communication	Hakone, 2017, [14]	8 participants, 6 patients and 2 doctors. All patients were male and age 65+.
Data is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania	Peck, 2019, [29]	42 participants, all rural Pennsylvanians recruited from either a local university, a local construction site, or a local farmers market. Most participants identified as liberal on the liberal-conservative scale. Most participants had some personal experience with drug abuse. The participants on average had a high school education and some college credit and varied consistently in age from 18-64 with some participants being age 65+.
Treemap Literacy: A Classroom-Based Investigation	Firat, 2020, [10]	25 participants, 2 female and 23 male. All were university students, 14 undergraduate students, 4 master's students, and 7 PhD students. Participant ages ranged from 18 to 38.
Pushing the (Visual) Narrative: the Effects of Prior Knowledge Elicitation in Provocative Topics	Heyer, 2020, [16]	160 participants, 103 male, 56 female, and 1 other recruited from MTurk.
"It's Just a Graph" – The Effect of Post-Hoc Rationalisation on InfoVis Evaluation	van Koningsbruggen, 2021, [33]	24 participants, 10 female and 14 male with an average age of 35.21. All participants were European with 21 of them being Dutch.
Smile or Scowl? Looking at Infographic Design Through the Affective Lens	Lan, 2021, [21]	245 participants, recruited from MTurk. Ages range from 18 to 79 with an average age of 37.39.

and they were consistently able to find participant pools with equal numbers of females and males. Six of these seven studies included the age of the participants, but they were heavily skewed towards the younger side. This is partially because three of the six studies were specifically done on children. However, even the studies that were not focused on children had low average ages.

Out of the four of these seven studies that were not focused on children, two included the education details of the participants. All participants had completed a high school education and the majority had some level of university education. One of these studies included race and the majority of participants were white.

All of these studies were conducted in the United States. Most of these studies show indicators of wealth among the participants: the elementary school study by Alper et al. [1] states that it was conducted in a school located in an upper-class district, those with university degrees are typically wealthier than those without [32], and those with the time and money to visit museums may be wealthier than average as well.

This informal analysis shows that visualization literacy user studies have been limited to those who are WEIRD (Western, educated, industrialized, rich, and democratic), which is not representative of the world population. Because of this, results found from the studies should not be considered conclusive of the global population.

5.2 Indirect Contributors

There are eight papers in the indirect contributors category. Three of the studies in these papers recruited participants from MTurk. The assessment by Harrison et al. [15] does not provide demographics of the participants, while the other two MTurk studies provide demographics.

The seven user studies with demographics will be the focus of this analysis. Four of these seven studies included the gender or sex of the participants and were not limited based on gender or sex. These four studies were strongly skewed towards male participants. Six of these seven studies included the age of the participants and were not limited based on age. These six studies were typically evenly distributed with limited numbers of participants above age 65.

Three of these seven studies included the education details of the participants. The participants from Firat et al.'s [10] study all had some university education. Many but not all of the participants from Peck et al.'s [29] study had completed high school and at least some college credit. The majority of participants in Yu's [35] study did not complete a high school education. None of these studies included race information.

Six out of eight of these studies were conducted in the United States with the remaining two being in China and Europe. This informal analysis shows that user studies that indirectly contribute to visualization literacy research tend to be limited to those who are WEIRD, but are not always. Again, results found from the studies should not be considered conclusive of the global population.

5.3 Comparison

Direct contributors had more diverse participants in terms of gender and sex. Indirect contributors to visualization literacy were skewed towards male participants, likely because most of the groups the researchers recruited from were male-dominated. Due to this imbalance, results found in these studies may not hold true for women and gender minorities.

Indirect contributors had wider distributions of age and were less skewed towards younger participants than the direct contributors. Indirect contributors also had more diverse participants in terms of education level and countries recruited from, although only one of the studies recruited participants from a non-Western country.

Future visualization literacy research should focus on recruiting more diverse participants. They should focus on recruiting older, less educated, and non-Western participants in order to make more global claims about visualization literacy which is important to ensure all people can benefit from new research.

6 DISCUSSION AND FUTURE WORK

Visualization literacy is a challenging topic to scope, and it would be easy to let it encompass all of visualization. In this paper, we focused on work in understanding the state of, measuring, and teaching visualization literacy in the global population.

There has been significant work in related fields, such as individual cognitive differences in visualization. The goal of this paper was to investigate our understanding of visualization literacy, and not the question of why certain individuals or groups may lack visualization literacy. Early work in visualization literacy was done using different terms, such as "visualization comprehension". In 2005, Velez et al. [34] investigated how cognitive characteristics, in particular spatial ability, affects understanding of visualizations as well as the differences in understanding they found between men and women. Similarly, Lee et al. [24] examined the correlations of visualization literacy with the cognitive ability of numeracy, the cognitive motivation of need for cognition, and two cognitive styles which are described as "a person's characteristic mode of perceiving, thinking, remembering, and problem solving". These papers focused on finding reasons why people may have low visualization literacy which is important work for visualization in general but was not the focus of this survey.

Given how young visualization literacy research is, all of the themes we found could be investigated much more. Fundamentals research should be expanded to include the construction of visualizations and how that skill interacts with the ability to interpret visualizations. How strong is the general public's ability to construct visualization? Does learning to interpret visualizations include skills in constructing visualizations?

Believability research could investigate how to change opinions on believability, create trust with users, and which groups are considered to be illiterate in visualization versus literate but with different views. Which visualizations are effective at changing people's opinions? How can we determine whether people are unable to see that certain sources are unreliable or if they have different standards of reliability?

Engagement research should be done to investigate how to build narratives into visualization systems, how to situate visualizations even when they cover a global topic, and the differences in engagement between viewing visualizations in social settings or alone. How can narratives be built into visualizations that are not naturally narrative? What visualization techniques show users both a global picture and a local version that emphasizes personal impact? Even if researchers do not consider believability and engagement to be vital parts of visualization literacy, they should at least discuss believability and engagement and why they are not relevant to their research.

Future work could also focus on the interplay between the themes identified in this paper: fundamentals, believability, and engagement. For example, how strongly does being skilled in each of these themes correlate to being skilled in the others? Is believability taught separately from the fundamentals or is it an inherent aspect of fundamentals?

Future user studies need to do a better at studying true novices and underrepresented groups, particularly in gender and sex, age, education level, and recruitment country. Further, future studies should be more consistent with reporting demographics of participants. Researcher's must consider and discuss whether their visualization research is really helpful for everyone, whether they are really designing for novice users, and when should they be calling for better visualization literacy.

We were unable to complete all of the readings that may have been relevant to this paper due to time constraints. In the future, I'd like to read and include information from books such as *Data Feminism* [9].

7 CONCLUSION

We surveyed twenty-four papers and then identified and discussed three themes: fundamentals, believability, and engagement. We found that two themes, believability and engagement, were not typically examined in work on visualization literacy. We propose that they be considered in standard visualization literacy research and provide a new definition for visualization literacy that encompasses all of these themes. Participant pools in visualization literacy user studies were analysed to find gaps in the literature. Gaps were identified in gender, age, education level,

and recruitment country diversity. We suggest that future visualization literacy user studies include more demographics about their participants so gaps can be more easily identified and remedied.

REFERENCES

- [1] B. Alper, N. H. Riche, F. Chevalier, J. Boy, and M. Sezgin. Visualization literacy at elementary school. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 5485–5497. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3025877
- [2] O. Barral, S. LallÉ, A. Iranpour, and C. Conati. Effect of adaptive guidance and visualization literacy on gaze attentive behaviors and sequential patterns on magazine-style narrative visualizations. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 11(3-4):1–46, 2021.
- [3] K. Börner, A. Bueckle, and M. Ginda. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences*, 116(6):1857–1864, 2019.
- [4] J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete. A principled way of assessing visualization literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1963–1972, 2014. doi: 10.1109/TVCG.2014.2346984
- [5] K. Börner, A. Maltese, R. N. Balliet, and J. Heimlich. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization*, 15(3):198–213, 2016. doi: 10.1177/1473871615594652
- [6] X. Chen, J. L. Hay, E. A. Waters, M. T. Kiviniemi, C. Biddle, E. Schofield, Y. Li, K. Kaphingst, and H. Orom. Health literacy and use and trust in health information. *Journal of Health Communication*, 23(8):724–734, 2018. PMID: 30160641. doi: 10.1080/10810730.2018.1511658
- [7] F. Chevalier, N. H. Riche, B. Alper, C. Plaisant, J. Boy, and N. Elmqvist. Observations and reflections on visualization literacy in elementary school. *IEEE computer graphics and applications*, 38(3):21–29, 2018.
- [8] M. Correll and J. Heer. Black hat visualization. In *Workshop on Dealing with Cognitive Biases in Visualisations (DECISiVe)*, IEEE VIS, 2017.
- [9] C. D’ignazio and L. F. Klein. *Data feminism*. MIT press, 2020.
- [10] E. E. Firat, A. Denisova, and R. S. Laramée. Treemap Literacy: A Classroom-Based Investigation. In M. Romero and B. Sousa Santos, eds., *Eurographics 2020 - Education Papers*. The Eurographics Association, 2020. doi: 10.2312/eged.20201032
- [11] J. Gäbler, C. Winkler, N. Lengyel, W. Aigner, C. Stoiber, G. Wallner, and S. Kriglstein. Diagram safari: A visualization literacy game for young children. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, CHI PLAY '19 Extended Abstracts, p. 389–396. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3341215.3356283
- [12] M. Galesic and R. Garcia-Retamero. Graph literacy: a cross-cultural comparison. *Medical Decision Making*, 31(3):444–457, 2011.
- [13] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. Lineup: Visual analysis of multi-attribute rankings. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2277–2286, 2013. doi: 10.1109/TVCG.2013.173
- [14] A. Hakone, L. Harrison, A. Ottley, N. Winters, C. Gutheil, P. K. J. Han, and R. Chang. Proact: Iterative design of a patient-centered visualization for effective prostate cancer health risk communication. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):601–610, 2017. doi: 10.1109/TVCG.2016.2598588
- [15] L. Harrison, D. Skau, S. Franconeri, A. Lu, and R. Chang. Influencing visual judgment through affective priming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, p. 2949–2958. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2470654.2481410
- [16] J. Heyer, N. K. Raveendranath, and K. Reda. Pushing the (visual) narrative: the effects of prior knowledge elicitation in provocative topics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2020.
- [17] E. Huynh, A. Nyhout, P. Ganea, and F. Chevalier. Designing narrative-focused role-playing games for visualization literacy in young children. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):924–934, 2021. doi: 10.1109/TVCG.2020.3030464
- [18] A. Jena, M. Butler, T. Dwyer, K. Ellis, U. Engelke, R. Kirkham, K. Marriott, C. Paris, and V. Rajamanickam. The next billion users of visualization. *IEEE Computer Graphics and Applications*, 41(2):8–16, 2021.
- [19] S. H. Kimball. Survey data collection; online panel efficacy. a comparative study of amazon mturk and research now ssi/survey monkey/opinion access. *Journal of Business Diversity*, 19(2), 2019.
- [20] B. C. Kwon and B. Lee. A comparative evaluation on online learning approaches using parallel coordinate visualization. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, p. 993–997. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2858036.2858101
- [21] X. Lan, Y. Shi, Y. Zhang, and N. Cao. Smile or scowl? looking at infographic design through the affective lens. *IEEE Transactions on Visualization and Computer Graphics*, 27(6):2796–2807, 2021.
- [22] C. Lee, T. Yang, G. D. Inchoco, G. M. Jones, and A. Satyanarayan. Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–18, 2021.
- [23] S. Lee, S.-H. Kim, and B. C. Kwon. Vlat: Development of a visualization literacy assessment test. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):551–560, 2017. doi: 10.1109/TVCG.2016.2598920
- [24] S. Lee, B. C. Kwon, J. Yang, B. C. Lee, and S.-H. Kim. The correlation between users’ cognitive characteristics and visualization literacy. *Applied Sciences*, 9(3):488, 2019.
- [25] L. Li, X. Chen, D. Wu, and H. Li. Effects of attending preschool on adolescents’ reading literacy: Evidence from the ethnic minority children in china. *Children and Youth Services Review*, 116:105211, 2020. doi: 10.1016/j.childyouth.2020.105211
- [26] A. V. Maltese, J. A. Harsh, and D. Svetina. Data visualization literacy: Investigating data interpretation along the novice—expert continuum. *Journal of College Science Teaching*, 45(1):84–90, 2015.
- [27] H. Mansoor and L. Harrison. Data visualization literacy and visualization biases: Cases for merging parallel threads. In *Cognitive Biases in Visualizations*, pp. 87–96. Springer, 2018.
- [28] P. Mena. Reducing misperceptions through news stories with data visualization: The role of readers’ prior knowledge and prior beliefs. *Journalism*, 0(0):14648849211028762, 2021. doi: 10.1177/14648849211028762
- [29] E. M. Peck, S. E. Ayuso, and O. El-Etr. *Data is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania*, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2019.
- [30] K. Peppler, A. Keune, and A. Han. Cultivating data visualization literacy in museums. *Information and Learning Sciences*, 2021.
- [31] S. Sultana, S. I. Ahmed, and J. M. Rzeszutowski. Seeing in context: Traditional visual communication practices in rural bangladesh. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3):1–31, 2021.
- [32] S. Tavernise. Education gap grows between rich and poor, studies say. *The New York Times*, 9, 2012.
- [33] R. van Koningsbruggen and E. Hornecker. “it’s just a graph”—the effect of post-hoc rationalisation on infovis evaluation. In *Creativity and Cognition*, pp. 1–10, 2021.
- [34] M. Velez, D. Silver, and M. Tremaine. Understanding visualization through spatial ability differences. In *VIS 05. IEEE Visualization, 2005.*, pp. 511–518, 2005. doi: 10.1109/VISUAL.2005.1532836
- [35] L. Yu. How poor informationally are the information poor? evidence from an empirical study of daily and regular information practices of individuals. *Journal of Documentation*, 2010.

Appendix

Table 4: The 27 papers considered for this survey with whether they were included (“Used?”), contained a definition for visualization literacy (“Definition?”, if yes and the cell is red then the definition of visualization literacy does not match what is discussed in the paper) discussed fundamentals, discussed believability, and discussed engagement. For the theme columns, a white cell means the theme was not discussed in the associated paper, a green cell means the theme was only discussed in the content of the associated paper and was not included in the provided definition of visualization literacy, a blue cell means the theme was discussed in both the content of the associated paper and the provided definition of visualization literacy, and a purple cell means the theme was discussed in the provided definition of visualization literacy but not in the content of the associated paper. 14 papers discuss fundamentals, 7 papers discuss believability, and 8 papers discuss engagement. The papers are first sorted numerically by year then sorted alphabetically by author. It is noteworthy that papers on fundamentals were more common during earlier years while papers discussing believability or engagement were more common during later years, suggesting a shift in interests.

Title	First Author, Year, Citation	Used?	Definition?	Fundamentals	Believability	Engagement
How poor informationally are the information poor? Evidence from an empirical study of daily and regular information practices of individuals	Yu, 2010, [35]	yes	no			
Graph Literacy: A Cross-Cultural Comparison.	Galesic, 2011, [12]	yes	no			
Influencing visual judgment through affective priming	Harrison, 2013, [15]	yes	no			
A Principled Way of Assessing Visualization Literacy	Boy, 2014, [4]	yes	yes			
Data Visualization Literacy: Investigating Data Interpretation Along the Novice-Expert Continuum	Maltese, 2015, [26]	yes	no			
Investigating Aspects of Data Visualization Literacy Using 20 Information Visualizations and 273 Science Museum Visitors	Börner, 2016, [5]	yes	yes			
A Comparative Evaluation On Online Learning Approaches Using Parallel Coordinate Visualization	Kwon, 2016, [20]	no				
Visualization Literacy at Elementary School	Alper, 2017, [1]	yes	yes			
Black hat visualization	Correll, 2017, [8]	yes	no			
PROACT: Iterative Design of a Patient-Centred Visualization for Effective Prostate Cancer Health Risk Communication	Hakone, 2017, [14]	yes	no			
Development of a visualization literacy assessment test	Lee, 2017, [23]	yes	yes			
Observations and Reflections on Visualization Literacy in Elementary School	Chevalier, 2018, [7]	yes	yes			
Data Visualization Literacy and Visualization Biases: Cases for Merging Parallel Threads	Mansoor, 2018, [27]	yes	yes			
Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments	Börner, 2019, [3]	yes	yes			
Diagram safari: A visualization literacy game for young children	Gäbler, 2019, [11]	yes	yes			
The Correlation between Users’ Cognitive Characteristics And Visualization Literacy	Lee, 2019, [24]	no				

continued on next page

Title	First Author, Year, Citation	Used?	Definition?	Fundamentals	Believability	Engagement
Data is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania	Peck, 2019, [29]	yes	yes			
Treemap Literacy: A classroom-Based Investigation	Firat, 2020, [10]	yes	yes			
Pushing the (Visual) Narrative: The Effects of Prior Knowledge Elicitation in Provocative Topics	Heyer, 2020, [16]	yes	no			
Effect of Adaptive Guidance and Visualization Literacy on Gaze Attentive Behaviors and Sequential Patterns on Magazine-Style Narrative Visualizations	Barral, 2021, [2]	no				
Designing Narrative-Focused Role-Playing Games for Visualization Literacy in Young Children	Huynh, 2021, [17]	yes	yes			
Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online	Lee, 2021, [22]	yes	no			
The Next Billion Users of Visualization	Jena, 2021, [18]	yes	no			
“It’s Just a Graph” – The Effect of Post-Hoc Rationalisation on InfoVis Evaluation	van Koningsbruggen, 2021, [33]	yes	no			
Smile or Scowl? Looking at Infographic Design Through the Affective Lens	Lan, 2021, [21]	yes	no			
Cultivating data visualization literacy in museums	Peppler, 2021, [30]	yes	yes			
Seeing in Context: Traditional Visual Communication Practices in Rural Bangladesh	Sultana, 2021, [31]	yes	yes			