

1 **The Role of Data Journalists as Educators**

2  
3  
4 MANDI CAI, Northwestern University, United States

5  
6 MATTHEW KAY, Northwestern University, United States

7  
8  
9 In 2020, data journalists reported on complex, dynamic issues like COVID-19 and the U.S. presidential election, publishing articles that  
10 used data to visualize a narrative and articles that explicitly instructed audiences on how to interpret data and data visualizations.  
11 Data journalism may very well be instrumental in developing the public’s familiarity with and literacy in data and data visualizations,  
12 but it remains unclear how to define or measure literacy cultivated by data and visual journalism. It also remains unclear how best to  
13 support data journalists in teaching their readers. Moreover, data journalists frequently report on polarizing topics involving data that  
14 is subject to widespread disagreement. This raises the question of how to teach data and data visualization literacy while considering  
15 opposing values and ideologies among audiences. I outline existing definitions of data and data visualization literacy, and propose  
16 some future directions for how data journalists could support this literacy among their readers and partner with educators.  
17  
18

19 Additional Key Words and Phrases: data literacy, visualization literacy, data journalism

20  
21 **ACM Reference Format:**

22  
23 Mandi Cai and Matthew Kay. 2024. The Role of Data Journalists as Educators. In *CHI ’24 Workshop on Toward a More Comprehensive*  
24 *Understanding of Visualization Literacy, May 11–16, 2024, Honolulu, Hawaii*. ACM, New York, NY, USA, 5 pages. <https://doi.org/XXXXXXX.XXXXXXX>  
25  
26

27  
28 **1 INTRODUCTION**

29  
30 The year 2020 was a monumental year for data journalists. As the COVID-19 virus swept the world, data journalists,  
31 alongside researchers and public health officials, immediately spun up dashboards showing COVID-19 data and used this  
32 data to publish critical stories about the pandemic’s impact [1, 26, 27]. These dashboards were accompanied by many  
33 explanatory articles on how to interpret COVID-19 data and predictions. For example, The Star Tribune published an  
34 article explaining why models of COVID-19’s impact on Minnesota varied drastically in the number of projected deaths  
35 [8]. The New York Times published articles that explained how to interpret a logarithmic line chart and what it meant  
36 to “flatten the curve” [7, 25]. In addition to a pandemic that forced the entire world to understand COVID-19 models  
37 and data visualizations, a U.S. presidential election took place in 2020. Noteworthy examples of how data journalists  
38 attempted to cultivate data and visualization literacy in the elections domain included articles explaining how political  
39 campaigns use data on voters to inform their campaign tactics [11] and how to read election maps [21, 22]. Throughout  
40 2020, data journalists and the broader news media were repeatedly thrust into the role of helping diverse audiences  
41 make sense of data and data visualizations amid mixed messaging and misinterpretations. However, in a fast-paced  
42  
43  
44

---

45 Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not  
46 made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components  
47 of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to  
48 redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

49 © 2024 Association for Computing Machinery.  
50 Manuscript submitted to ACM

newsroom, it is not always clear what the educational value of data and visual journalism is or how to measure it. In a study by Fink and Anderson [10], the data journalists interviewed said they “knew little about how their audiences interacted with their stories.”

I propose exploring **the vital role that data journalists play as educators of data and data visualization literacy, and how to teach and measure this literacy in a polarized media landscape when the data itself is controversial.** Content encountered in the news media, whether it be through online news websites, broadcast and cable television, or social media platforms, serve as opportunities for audiences to engage and reason with data and visualizations on topics that impact their daily life and community. However, this relevance to people’s lived experiences also means that the meaning of the data may be subject to disagreement and even politicization, like numbers of COVID-19 cases or vote tallies during elections. Indeed, several studies [15, 18, 23] have called for visualization researchers to factor in the values that individuals hold and the contexts individuals are embedded in when they engage with data and data visualizations. Making matters more complicated, data and data visualizations in the news are often evolving stories that change as more information is known.

## 2 EXISTING FRAMEWORKS FOR DATA AND VISUALIZATION LITERACY

To define *data literacy*, Calzada Prado and Marzal [6] propose a framework for librarians that segments data literacy into understanding data, finding and obtaining data, reading, interpreting and evaluating data, managing data, and using data. Across data literacy frameworks, there are **explicit mentions of data literacy as the ability to use data in the process of evidence-based thinking to solve real-world problems, as well as an emphasis on the importance of the data’s context** [28, 29]. For example, Wolff et al. [29] envision data literacy as a pool of skills and knowledge that includes specialized skills unique to the specific context that the individual is using the data in. Gummer and Mandinach [14] describe a conceptual framework of what data literacy looks like for teachers who can use data to inform their teaching practices and track the progress of students. Given the mass amounts of recorded data across a variety of contexts, Bhargava [3] argues for conceptualizing data literacy as “literacy in the age of data”, or “the desire and ability to constructively engage in society through or about data” [3].

In the realm of *visualization literacy*, Boy et al. [4], Börner et al. [5], and Lee et al. [19] have provided definitions of visualization literacy that all specify one’s ability to read, understand, and make meaning from the visual representations of data, with Ge et al. [12] extending past definitions to include a critical thinking component – “the ability to read, interpret, and reason about erroneous or potentially misleading visualizations.” Börner et al. [5] note a gap in best practices for teaching visualization design and present a process model for data visualization construction with the goal of creating a visualization for some purpose. Visualization and data literacy inevitably overlap as it is necessary to understand the data being visualized.

Several of the aforementioned frameworks mention the importance of considering socio-cultural contexts. Both Wolff et al. [29] and Gummer and Mandinach [14] echo this sentiment, stating that teaching data should be “responsive to cultural differences that might affect individual learner’s view of the world” [29] and “expand beyond a cognitive focus on knowledge and skills to include beliefs/values, identity, and epistemic elements” [14]. Indeed, one’s identity can impact how they make sense of data and data visualizations. Ericson et al. showed how political partisanship moderates participants’ subjective interpretations of COVID-19 data and data visualizations [9]. Another study by Kahan et al. [16] showed that participants with higher numeracy – one’s ability to interpret and use numbers, often probabilities,

105 to draw valid inferences [20, 24] – exhibited more polarized interpretations of data on a ban on carrying concealed  
106 weapons in public and its relationship to crime numbers, compared with data on skin rashes and its relationship to skin  
107 cream use. Individuals with higher numeracy used their increased cognitive ability to bend their interpretation of the  
108 concealed carry data to conform with their own political leanings.  
109

### 110 3 PATHS FORWARD

111 **How do these frameworks translate to the learning experiences of viewing data and data visualizations**  
112 **in the news? How might we adapt them for data journalists who are interested in fostering data and data**  
113 **visualization literacy? To support data journalists, it is crucial to define what successful learning looks like when**  
114 **a reader engages with their data and visual journalism. This metric will likely depend on what purpose specific**  
115 **audiences have in mind when reading these articles, and data journalists' perceptions of their own roles as**  
116 **educators and how they service the public.**  
117  
118  
119

120 For example, if a reader comes to a news article with forecast visualizations of COVID-19 case trends for the next two  
121 weeks, what skills would this reader need to have in order to make an appropriate decision from these visualizations  
122 that aligns with that reader's risk tolerance? Some of the required skills and knowledge would translate across the  
123 reader's experiences with similar types of visualizations, but some are also specific to that specific topic, decision, and  
124 the reader's current priorities. Future research can intentionally dissect understanding and usage of data and data  
125 visualizations within contentious topics in news media across participants of varying values, ideology, and political  
126 affiliations, taking note of the external evidence that individuals bring in to support their reasoning. The public does not  
127 interpret data and data visualizations in a vacuum. They may engage with it in a news article, draw some inferences,  
128 and continue evolving their understanding of the data visualization as it relates to some larger narrative. It is thus worth  
129 asking, in the context of our media landscape, whether teaching data and data visualization literacy should include  
130 **fostering one's ability to discuss a visualization and acknowledge the uncertainty inherent in any quantified**  
131 **representation of dynamic, complex issues.** It is not a cleanly scoped, one-to-one interaction with a visualization  
132 that is objective and final – the visualization is part of a larger existing dialogue.  
133  
134  
135  
136  
137

138 Data journalists' recognition of their roles as data and visualization educators also needs to be investigated. Should  
139 they choose to take on an educator mindset, it remains an open question of how to train data journalists with this  
140 mindset and equip them with teaching objectives when they are juggling other roles and responsibilities in a fast-paced  
141 newsroom. Researchers, educators, and journalists could partner to explore formats for articles on topics of interest to  
142 readers that explicitly target the development of certain data and visualization literacy skills. It would be worthwhile to  
143 work with some readers directly to figure out what they want to learn and how they successfully learn, perhaps in a  
144 series of workshops that could also function as trust building opportunities between journalists and readers.<sup>1</sup>  
145  
146

147 Another path forward is one in which educators adapt data and visual journalism as materials in their classroom  
148 curricula. Challenges in teaching data visualization, described by Bach et al. [2], include “communicating the need  
149 for visualization education” and “retaining motivation during learning visualization”. Perhaps focusing visualization  
150 curriculum around current events in the news could communicate the need and provide motivation to students who  
151 may be impacted by these issues and seek to dissect them further through visualization. Educators can tie in widely  
152  
153

154 <sup>1</sup>There has been a documented decrease in public trust in the news in the United States since the 1950's, which could impact impressions of data and data  
155 visualizations communicated by news outlets [13, 17].  
156

discussed data sets in the news into their lesson plans after teaching students to work with smaller, simpler data sets, even explicitly acknowledging and explaining current points of confusion in these data sets among the broader public that data journalists detect. Data journalists could monitor public interest and responses to their coverage and forward observations to educators to build classroom lesson plans around.

#### 4 CONCLUSION

Data and visual journalism is an opportunity for the public to understand the relevance of data and data visualizations, as well as sharpen their data and visualization literacy skills. The inherently polarizing nature of many topics covered by data journalism underscores the importance of considering diverse perspectives and ideologies when teaching these skills. Data journalists could collaborate with researchers and educators in data and visualization literacy to understand their audiences, create pieces that intentionally foster certain competencies, and provide materials and audience observations to educators for classroom use. As Bach et al. [2] point out in their call to action, there is a need for more interdisciplinary communities to discuss approaches to teaching visualization literacy – this is a community that data journalists can be a part of.

#### REFERENCES

- [1] Stephanie Adeline, Connie Hanzhang Jin, Alyson Hurt, Thomas Wilburn, Daniel Wood, and Ruth Talbot. 2020. Tracking the coronavirus around the U.S.: See how your state is doing. Retrieved February 28, 2024 from <https://www.npr.org/sections/health-shots/2020/09/01/816707182/map-tracking-the-spread-of-the-coronavirus-in-the-u-s>
- [2] Benjamin Bach, Mandy Keck, Fateme Rajabiyazdi, Tatiana Losev, Isabel Meirelles, Jason Dykes, Robert S. Laramée, Mashael AlKadi, Christina Stoiber, Samuel Huron, Charles Perin, Luiz Morais, Wolfgang Aigner, Doris Kosminsky, Magdalena Boucher, Søren 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. <http://arxiv.org/abs/2308.07703> arXiv:2308.07703 [cs].
- [3] Rahul Bhargava. 2015. *Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data*. Technical Report. <https://datapopalliance.org/wp-content/uploads/2015/11/Beyond-Data-Literacy-2015.pdf>
- [4] Jeremy Boy, Ronald A. Rensink, Enrico Bertini, and Jean-Daniel Fekete. 2014. A Principled Way of Assessing Visualization Literacy. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (Dec. 2014), 1963–1972. <https://doi.org/10.1109/TVCG.2014.2346984>
- [5] Katy Börner, Andreas Bueckle, and Michael Ginda. 2019. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences* 116, 6 (Feb. 2019), 1857–1864. <https://doi.org/10.1073/pnas.1807180116>
- [6] Javier Calzada Prado and Miguel Ángel Marzal. 2013. Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. *Libri* 63, 2 (Jan. 2013), 123–134. <https://doi.org/10.1515/libri-2013-0010>
- [7] Kenneth Chang. 2020. A Different Way to Chart the Spread of Coronavirus. Retrieved February 28, 2024 from <https://www.nytimes.com/2020/03/20/health/coronavirus-data-logarithm-chart.html>
- [8] Michael Corey and C. J. Sinner. 2020. 360 deaths or 22,000? Why Minnesota’s COVID-19 models are so different. Retrieved February 28, 2024 from <https://www.startribune.com/360-deaths-or-22-000-why-minnesota-s-covid-19-models-are-so-different/568966211/>
- [9] Jonathan D Ericson, William S Albert, and Ja-Nae Duane. 2022. Political affiliation moderates subjective interpretations of COVID-19 graphs. *Big Data & Society* 9, 1 (Jan. 2022), 205395172210806. <https://doi.org/10.1177/20539517221080678>
- [10] Katherine Fink and C. W. Anderson. 2015. Data Journalism in the United States. *Journalism Studies* 16, 4 (July 2015), 467–481. <https://doi.org/10.1080/1461670X.2014.939852> Publisher: Routledge \_eprint: <https://doi.org/10.1080/1461670X.2014.939852>.
- [11] Minami Funakoshi, Elizabeth Culliford, and Wen Foo. 2020. How political campaigns use your data. Retrieved March 7, 2024 from <https://www.reuters.com/graphics/USA-ELECTION/DATA-VISUAL/yxmvjgojvr/>
- [12] Lily W. Ge, Yuan Cui, and Matthew Kay. 2023. CALVI: Critical Thinking Assessment for Literacy in Visualizations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–18. <https://doi.org/10.1145/3544548.3581406>
- [13] Paul Gronke and Timothy E. Cook. 2007. Disdaining the Media: The American Public’s Changing Attitudes Toward the News. *Political Communication* 24, 3 (Aug. 2007), 259–281. <https://doi.org/10.1080/10584600701471591>
- [14] Edith S. Gummer and Ellen B. Mandinach. 2015. Building a Conceptual Framework for Data Literacy. *Teachers College Record* 117, 040305 (April 2015), 1–22.
- [15] Helen Ai He, Jagoda Walny, Sonja Thoma, Sheelagh Carpendale, and Wesley Willett. 2023. Enthusiastic and Grounded, Avoidant and Cautious: Understanding Public Receptivity to Data and Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2023), 1435–1445. <https://doi.org/10.1109/TVCG.2023.3326917>

- 209 [16] Dan M. Kahan, Ellen Peters, Erica Cantrell Dawson, and Paul Slovic. 2017. Motivated numeracy and enlightened self-government. *Behavioural*  
210 *Public Policy* 1, 1 (May 2017), 54–86. <https://doi.org/10.1017/bpp.2016.2>
- 211 [17] Jonathan McDonald Ladd. 2010. The Role of Media Distrust in Partisan Voting. *Political Behavior* 32, 4 (Dec. 2010), 567–585. <https://doi.org/10.1007/s11109-010-9123-z>
- 212 [18] Crystal Lee, Tanya Yang, Gabrielle D Inchoco, Graham M. Jones, and Arvind Satyanarayan. 2021. Viral Visualizations: How Coronavirus Skeptics  
213 Use Orthodox Data Practices to Promote Unorthodox Science Online. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing*  
214 *Systems*. ACM, Yokohama Japan, 1–18. <https://doi.org/10.1145/3411764.3445211>
- 215 [19] Sukwon Lee, Sung-Hee Kim, and Bum Chul Kwon. 2017. VLAT: Development of a Visualization Literacy Assessment Test. *IEEE Transactions on*  
216 *Visualization and Computer Graphics* 23, 1 (Jan. 2017), 551–560. <https://doi.org/10.1109/TVCG.2016.2598920>
- 217 [20] Jordana M. Liberali, Valerie F. Reyna, Sarah Furlan, Lilian M. Stein, and Seth T. Pardo. 2012. Individual Differences in Numeracy and Cognitive  
218 Reflection, with Implications for Biases and Fallacies in Probability Judgment. *Journal of Behavioral Decision Making* 25, 4 (Oct. 2012), 361–381.  
219 <https://doi.org/10.1002/bdm.752>
- 220 [21] Betsy Mason. 2020. Opinion | This Election Map Will Be Everywhere. Don't Let It Fool You. Retrieved March 7, 2024 from <https://www.nytimes.com/interactive/2020/10/30/opinion/election-results-maps.html>
- 221 [22] The Learning Network. 2020. What's Going On in This Graph? | 2020 Presidential Election Maps. Retrieved March 7, 2024 from <https://www.nytimes.com/2020/11/19/learning/whats-going-on-in-this-graph-2020-presidential-election-maps.html>
- 222 [23] Evan M. Peck, Sofia E. Ayuso, and Omar El-Etr. 2019. Data is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania. In  
223 *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–12. <https://doi.org/10.1145/3290605.3300474>
- 224 [24] Ellen Peters, Daniel Västfjäll, Paul Slovic, C K Mertz, Ketti Mazzocco, and Stephan Dickert. 2006. Numeracy and Decision Making. *Psychological*  
225 *Science* 17, 5 (2006), 407–413.
- 226 [25] Siobhan Roberts. 2020. Flattening the Coronavirus Curve. Retrieved February 28, 2024 from [https://www.nytimes.com/article/flatten-curve-](https://www.nytimes.com/article/flatten-curve-coronavirus.html)  
227 [coronavirus.html](https://www.nytimes.com/article/flatten-curve-coronavirus.html)
- 228 [26] The New York Times. 2020. Coronavirus in the U.S.: Latest Map and Case Count. Retrieved February 28, 2024 from [https://www.nytimes.com/](https://www.nytimes.com/interactive/2023/us/covid-cases.html)  
229 [interactive/2023/us/covid-cases.html](https://www.nytimes.com/interactive/2023/us/covid-cases.html)
- 230 [27] The Texas Tribune. 2020. How coronavirus impacted Texas: Hospitalizations, vaccinations, cases and deaths. Retrieved February 28, 2024 from  
231 <https://apps.texastribune.org/features/2020/texas-coronavirus-cases-map/>
- 232 [28] Phil Vahey, Louise Yarnall, Charles Patton, Daniel Zalles, and Karen Swan. 2006. Mathematizing Middle School: Results From a Cross-Disciplinary  
233 Study of Data Literacy. In *American Educational Research Association Annual Conference*. 1–15.
- 234 [29] Annika Wolff, Daniel Gooch, Jose J. Cavaro Montaner, Umar Rashid, and Gerd Kortuem. 2016. Creating an understanding of data literacy for a  
235 data-driven society. *The Journal of Community Informatics* 12, 3 (2016), 9–26.  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260