

Scoping the Terrain of Data Stories in the Social Media Wild

Ralph Vacca, Fordham University, ralph.vacca@fordham.edu
Camillia Matuk, New York University, cmatuk@nyu.edu
Peter Woods, University of Nottingham, peter.woods@nottingham.ac.uk
Kayla Desportes, New York University, kd90@nyu.edu
Anna Amato, New York University, ada437@nyu.edu
Megan Silander, Education Development Center, msilander@edc.org

Abstract: Drawing on previous research into the value of developing and sharing data stories on social media, we use this paper to examine how practitioners address a spectrum of interests and concerns in relation to their own data literacies within this media form. To do so, we analyzed 107 data story videos from TikTok and Instagram to explore what practices and communication techniques are apparent in social media data stories that exhibit features of data literacy. Through our analysis, we uncovered a series of digital storytelling techniques (e.g., speaking to the camera, using a green screen) that supported the creators' data science practices and communicative goals. This study contributes to the discourse on social media's role in data storytelling and literacy, providing guidance for future research and implications for the design of new data literacy learning experiences.

Introduction

Social media platforms have increasingly become spaces for the creation and dissemination of data stories - evidence-based arguments contextualized in human experience. These data stories represent a syncretic form of text, integrating data with personal experiences and various forms of social, cultural, and historical knowledge (Wilkerson et al., 2021). Central to data stories is an argumentative structure where informal reasoning is used to make claims, cite evidence, and explain reasoning (Kuhn, 1991; Means & Voss, 1996; Zohar & Nemet, 2002).

The construction of and engagement with data stories on social media platforms presents a unique learning opportunity to engage youth in critical data literacy - a set of skills that empower students to see themselves as active participants in data science, using data to understand and address societal issues, especially from marginalized perspectives (Philip et al., 2013). Data stories can enable students to exercise data literacy skills by connecting data with their lived experiences and cultural contexts, as well as fostering their participation in public discourse. For instance, Nguyen & Parameswaran (2023) observed that youth curate data for social media audiences by situating data in relatable contexts, remixing data types, and positioning their experiences as valid data points. Calabrese Barton et al. (2021) highlighted critical data practices among youth during the COVID-19 pandemic, including remixing (combining data sources for a comprehensive view), recontextualizing (critiquing data within personal and community contexts), and repositioning (viewing oneself as an agent rather than a mere statistic).

This study explores the multifaceted nature of data stories on social media, examining how they address a spectrum of interests and concerns, and the ways in which they are shaped by the goals and data literacy practices of their creators. Specifically, we ask: What practices and communication techniques are apparent in social media data stories that exhibit features of data literacy? Answering this question has implications for how we might design learning experiences to leverage real-world data storytelling practices.

Methods

Data collection

To explore what data literacy practices are apparent in social media data stories, we decided to focus on TikTok and Instagram, because both platforms have a user base that is largely comprised of youth and young adults. We first created dedicated Instagram and TikTok accounts, and then from those accounts we searched hashtags #dataviz; #datavisualization; #data; and identified additional, thematically related tags that frequently appeared with this tag, including #statistics; #datatok; #dataanalysis; #datascience. We also used the search terms: data story; data visualization; data argument; data narratives; data-driven storytelling. We then collected links to videos that appeared on the feed over a 21-day period in July 2023. One researcher scanned the 204 videos collected, and identified 107 that explicitly used data to support a specific communication goal (e.g., using maternal health data to vent about their own pregnancy experience). Of those videos 26% were from Instagram, and 74% from TikTok.

Data analysis

Our unit of analysis was a single video. We developed an initial set of codes to categorize the videos at a high level and then we analyzed the videos qualitatively to identify emergent sub-codes within each code (see Table 1). Our initial set of codes consisted of: (i) author posts; (ii) focus of content, (iii) embedded data type, (iv) techniques, (v) data practices, and (vi) goals. Two researchers engaged in all coding, and any discrepancies between coding was resolved through discussion. In coding videos for goals, we sought to describe the leading rhetorical objective of the data story video. For the data practices sub-codes, we adapted the critical data practices from Calabrese Barton et al. (2021) (e.g., remixing, recontextualizing, and repositioning) along with other work in analysis of data-narrative artifacts (AUTHORS, Year).

Table 1
Codebook for Analysis of Data Story Videos

Code	Description	Sub-Codes
Author posts	Frequency of data-focused posts in their feed, distinguishing between those with a significant presence of data-focused content vs. no other.	Data-dominant (i.e., 3+ data-focused posts); Not Data-dominant (< 3 data-focused posts).
Focus of content	Describes the main topical area of the data story video.	Economy; Gender/race; Health/wellness; Environment; Substance abuse; Generational; Violence; Children
Embedded data type	Describes the kind of data embedded into the data story video.	Qualitative; Quantitative; Mixed
Techniques	Describes the different media construction affordances used in the data story video	Speaking to camera; Green screen; Used data charts; Used animated data charts; Used clips and images to illustrate main points
Data Practices	Presence of (1) presenting/describing data, (2) adding a different context to the data or data claim, (3) mixing data across courses, (4) deconstructing a data claim, (5) connecting to a personal experience, or (6) making predictions.	(1) Present data; (2) Recontextualize data; (3) Remix data; (4) Deconstruct; (5) Personal experience; (6) Make predictions
Goals	Describes the leading (if more than one) rhetorical objective of the data story video.	Educate; Warn; Debate; Vent; Act

Results

Content and sources

Two-thirds (67%) of the videos we sampled originated from creators who had posted three or more data-centric videos; while one-third (33%) of our sample were by creators who had no similarly data-centric videos, based on a review of their video feed. Forty-six (46%) of the videos used exclusively quantitative data; 44% of the videos incorporated a mix of both quantitative and qualitative data; and 10%, used only qualitative data.

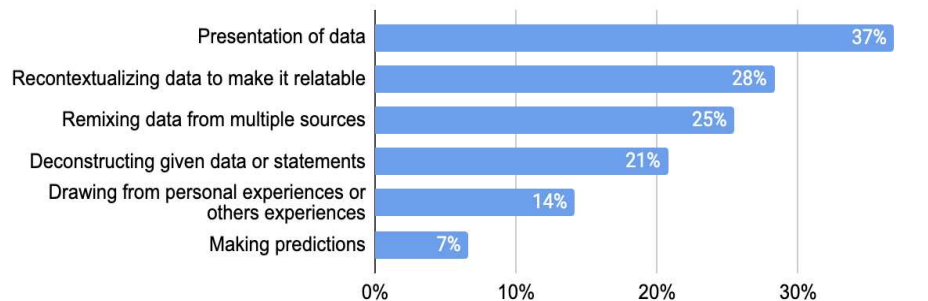
The thematic scope of the videos was notably diverse, with 26% addressing economic issues, including topics such as taxes, inflation, and the job market; 18% addressing issues of gender and race-based discrimination; 17% addressing topics in health and wellness; 9% addressing topics in environment and sustainability; and 8% addressing substance abuse, including smoking and drinking. Drawing comparisons across different generations was the main focus in 7% of the videos. Violence-related themes (e.g. school shootings) were the subject of 6%, and the remaining 4% focused on child education and wellbeing. This distribution underscores the wide-ranging interests and concerns addressed by social media users in their data-driven narratives.

Goals and relationship to practices

A majority of the videos, 51%, primarily aimed to disseminate information and educate the audience. Meanwhile, 17% of the videos had a clear intent to warn the audience, often about pressing issues or potential dangers. Another 17% of the videos were designed to engage the audience in debate. Notably, 12% of the videos served as outlets for venting or ranting, highlighting the role of these platforms in personal expression and the airing of grievances.

Finally, 3% of the videos encouraged viewers to take action on issues. This diversity of goals underscores the possibilities of social media as a tool for communication, education, debate, expression, and activism.

Figure 1
Data Practices Apparent in the Videos



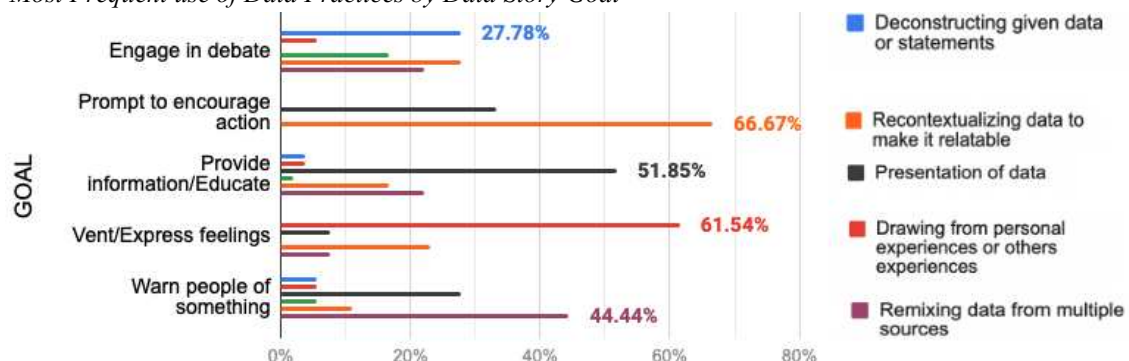
As evident in Figure 1, the videos demonstrated diverse data practices, 37% of the videos presented data; 28% recontextualized the data - for instance, in Figure 2a, the creator uses a fictional narrative to recontextualize data on how white households with a degree would be three times more wealthy than a black household with the same education. 25% remixed data from multiple sources. For example, in Figure 2b, the creator remixes data from four different data sources to discuss changes in student loan debt and impacts on students.

Figure 2
Examples of Practices Recontextualizing (a) and Remixing (b).



Our analysis of the co-occurrence of goal and practice codes reveals how creators' goals were supported by distinct data practices (see Figure 3). For instance, in videos aimed at engaging in debate, 27.78% of creators employed the practice of deconstructing given data or statements, and an equal percentage (27.78%) used recontextualizing techniques. In turn, among videos with the goal of prompting action, a majority (66.67%) relied on recontextualizing data, emphasizing the importance of making data relatable to inspire viewer action.

Figure 3
Most Frequent use of Data Practices by Data Story Goal



Techniques and relationships to practices

The communication techniques used in the videos varied. In 31% of the videos, creators spoke directly to the camera, while in 26% of the videos creators used the green screening feature, which allows creators to lay a video

of themselves over pictures and charts to illustrate, or otherwise complement their narration. In 25% of the videos images of charts and data diagrams were either green screened or pasted on top of a video of a speaker. A further 20% of the videos employed animated data diagrams. Lastly, 14% of the videos used other embedded video clips and non-chart imagery to illustrate their main points and thoughts. These varied techniques underscore the creative potential of social media platforms, allowing users to choose from a range of techniques to effectively communicate their message and engage their audience. The analysis of co-occurrence patterns in the video content reveals that the practice of remixing data from multiple sources frequently coincided with the use of green screening, observed in 42.11% of the cases. In 25% of the videos, deconstructing data claims was paired with the technique of stitching one's video to an existing video. Drawing on personal experiences and speaking directly to the camera, occurring in 57.14% of the videos. Making predictions was another practice that commonly co-occurred with talking to the camera, seen in 60% of the cases. Lastly, the presentation of data was often accompanied by animations of visual data and data charts, as seen in 57.14% of the videos. These patterns indicate a strategic blending of techniques to enhance the effectiveness and appeal of data communication in social media videos.

Discussion

The data stories we analyzed were crafted with a variety of goals, ranging from engaging in debate and prompting activism to venting emotions and warning others. This diversity extends to the array of data literacy practices employed, which ranged from expository presentations to more complex and interpretative approaches, highlighting the rich potential of social media for diverse forms of data engagement. Notably, there was a connection between the specific goals of data stories and the distinct data practices utilized to support these goals. For example, in videos aimed at fostering debate, authors frequently engaged in the practice of deconstructing given data or statements. Such patterns demonstrate a strategic alignment of content goals with specific data practices, enabling creators to effectively communicate their intended message and engage their audience. Moreover, our results suggest that creators are adept at leveraging the creative potential of social media platforms, choosing from a range of techniques to effectively convey their message. This includes the use of green screening as opposed to speaking directly to the camera. Additionally, we found relationships between different techniques and data practices, indicating that creators were strategically employing techniques to enhance particular kinds of data practices like re-contextualizing data, remixing across data sources, or deconstructing claims. The connection between data practices, techniques, and rhetorical goals may be of particular interest in considering how we might extend the designs of existing data literacy learning tools to afford the creative and audience-centric approach found in social media platforms and experiences. Future research may delve deeper into how these relationships between goals, practices, and techniques can be harnessed to scaffold specific types of data literacy narrative projects in K-12 classrooms. For instance, advocacy projects may benefit from centering re-contextualization practices, while debates could focus on deconstructing data. This exploration could provide valuable insights into optimizing the use of social media as a tool for effective data storytelling and literacy.

This study is limited by the small sample of videos, which reduces the generalizability of the findings. While our approach to finding videos allowed us to identify a diverse range of content, a follow-up study might pursue a more focused investigation to allow a more nuanced analysis of data storytelling strategies for particular communication goals, within a specific content area, and using particular kinds of data.

References

- Calabrese Barton, A., Greenberg, D., Turner, C., Riter, D., Perez, M., Tasker, T., Jones, D., Herrenkohl, L. R., & Davis, E. A. (2021). Youth Critical Data Practices in the COVID-19 Multipandemic. *AERA Open*, 7, 23328584211041631.
- Kuhn, D. (1991). *The Skills of Argument*. Cambridge University Press.
- Means, M. L., & Voss, J. F. (1996). Who Reasons Well? Two Studies of Informal Reasoning Among Children of Different Grade, Ability, and Knowledge Levels. *Cognition and Instruction*, 14(2), 139–178.
- Nguyen, H., & Parameswaran, P. (2023). Meaning making and relatedness: exploring critical data literacies on social media. *Information and Learning Sciences*, 124(5/6), 149–167.
- Wilkerson, M., Finzer, W., Erickson, T., & Hernandez, D. (2021). Reflective Data Storytelling for Youth: The CODAP Story Builder. *Interaction Design and Children*, 503–507.
- Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, 39(1), 35–62.