When more COVID-19 data doesn't equal more understanding

March 4 2021, by Daniel Ackerman

MIT researchers found that Covid-19 skeptics on Twitter and Facebook — far from being "data illiterate" — often use sophisticated data visualization techniques to argue against public health precautions like mask mandates. Credit: Jose-Luis Olivares, MIT

Since the start of the COVID-19 pandemic, charts and graphs have

helped communicate information about infection rates, deaths, and vaccinations. In some cases, such visualizations can encourage behaviors that reduce virus transmission, like wearing a mask. Indeed, the pandemic has been hailed as the breakthrough moment for data visualization.

But new findings suggest a more complex picture. A study from MIT shows how coronavirus skeptics have marshalled data visualizations online to argue against public health orthodoxy about the benefits of mask mandates. Such "counter-visualizations" are often quite sophisticated, using datasets from official sources and state-of-the-art visualization methods.

The researchers combed through hundreds of thousands of social media posts and found that coronavirus skeptics often deploy counter-visualizations alongside the same "follow-the-data" rhetoric as public health experts, yet the skeptics argue for radically different policies. The researchers conclude that data visualizations aren't sufficient to convey the urgency of the COVID-19 pandemic, because even the clearest graphs can be interpreted through a variety of belief systems.

"A lot of people think of metrics like infection rates as objective," says Crystal Lee. "But they're clearly not, based on how much debate there is on how to think about the pandemic. That's why we say data visualizations have become a battleground."

The research will be presented at the ACM Conference on Human Factors in Computing Systems in May. Lee is the study's lead author and a Ph.D. student in MIT's History, Anthropology, Science, Technology, and Society (HASTS) program and MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL), as well as a fellow at Harvard University's Berkman Klein Center for Internet and Society. Co-authors include Graham Jones, a Margaret MacVicar Faculty Fellow in

Anthropology; Arvind Satyanarayan, the NBX Career Development Assistant Professor in the Department of Electrical Engineering and Computer Science and CSAIL; Tanya Yang, an MIT undergraduate; and Gabrielle Inchoco, a Wellesley College undergraduate.

As data visualizations rose to prominence early in the pandemic, Lee and her colleagues set out to understand how they were being deployed throughout the social media universe. "An initial hypothesis was that if we had more data visualizations, from data collected in a systematic way, then people would be better informed," says Lee. To test that hypothesis, her team blended computational techniques with innovative ethnographic methods.

They used their computational approach on Twitter, scraping nearly half a million tweets that referred to both "COVID-19" and "data." With those tweets, the researchers generated a network graph to find out "who's retweeting whom and who likes whom," says Lee. "We basically created a network of communities who are interacting with each other." Clusters included groups like the "American media community" or "antimaskers." The researchers found that antimask groups were creating and sharing data visualizations as much as, if not more than, other groups.

And those visualizations weren't sloppy. "They are virtually indistinguishable from those shared by mainstream sources," says Satyanarayan. "They are often just as polished as graphs you would expect to encounter in data journalism or public health dashboards."

"It's a very striking finding," says Lee. "It shows that characterizing antimask groups as data-illiterate or not engaging with the data, is empirically false."

Lee says this computational approach gave them a broad view of

COVID-19 data visualizations. "What is really exciting about this quantitative work is that we're doing this analysis at a huge scale. There's no way I could have read half a million tweets."

But the Twitter analysis had a shortcoming. "I think it misses a lot of the granularity of the conversations that people are having," says Lee. "You can't necessarily follow a single thread of conversation as it unfolds." For that, the researchers turned to a more traditional anthropology research method—with an internet-age twist.

Lee's team followed and analyzed conversations about data visualizations in antimask Facebook groups—a practice they dubbed "deep lurking," an online version of the ethnographic technique called "deep hanging out." Lee says "understanding a culture requires you to observe the day-to-day informal goings-on—not just the big formal events. Deep lurking is a way to transpose these traditional ethnography approaches to digital age."

The qualitative findings from deep lurking appeared consistent with the quantitative Twitter findings. Antimaskers on Facebook weren't eschewing data. Rather, they discussed how different kinds of data were collected and why. "Their arguments are really quite nuanced," says Lee. "It's often a question of metrics." For example, antimask groups might argue that visualizations of infection numbers could be misleading, in part because of the wide range of uncertainty in infection rates, compared to measurements like the number of deaths. In response, members of the group would often create their own counter-visualizations, even instructing each other in data visualization techniques.

"I've been to livestreams where people screen share and look at the data portal from the state of Georgia," says Lee. "Then they'll talk about how to download the data and import it into Excel."

Jones says the antimask groups' "idea of science is not listening passively as experts at a place like MIT tell everyone else what to believe." He adds that this kind of behavior marks a new turn for an old cultural current. "Antimaskers' use of data literacy reflects deep-seated American values of self-reliance and anti-expertise that date back to the founding of the country, but their online activities push those values into new arenas of public life."

He adds that "making sense of these complex dynamics would have been impossible" without Lee's "visionary leadership in masterminding an interdisciplinary collaboration that spanned SHASS and CSAIL."

The mixed methods research "advances our understanding of data visualizations in shaping public perception of science and politics," says Jevin West, a data scientist at the University of Washington, who was not involved with the research. Data visualizations "carry a veneer of objectivity and scientific precision. But as this paper shows, data visualizations can be used effectively on opposite sides of an issue," he says. "It underscores the complexity of the problem—that it is not enough to 'just teach media literacy.'" It requires a more nuanced sociopolitical understanding of those creating and interpreting data graphics."

Combining computational and anthropological insights led the researchers to a more nuanced understanding of data literacy. Lee says their study reveals that, compared to public health orthodoxy, "antimaskers see the pandemic differently, using data that is quite similar. I still think data analysis is important. But it's certainly not the salve that I thought it was in terms of convincing people who believe that the scientific establishment is not trustworthy." Lee says their findings point to "a larger rift in how we think about science and expertise in the U.S." That same rift runs through issues like climate change and vaccination, where similar dynamics often play out in social media

discussions.

To make these results accessible to the public, Lee and her collaborator, CSAIL Ph.D. student Jonathan Zong, led a team of seven MIT undergraduate researchers to develop an interactive narrative where readers can explore the visualizations and conversations for themselves.

Lee describes the team's research as a first step in making sense of the role of data and visualizations in these broader debates. "Data visualization is not objective. It's not absolute. It is in fact an incredibly social and political endeavor. We have to be attentive to how people interpret them outside of the scientific establishment."

 More information: Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online. doi.org/10.1145/3411764.3445211

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology