

MIT Open Access Articles

Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Lee, Crystal et al. "Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online." Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, May 2021, Yokohama, Japan, Association for Computing Machinery, May 2021. © 2021 The Authors

As Published: <http://dx.doi.org/10.1145/3411764.3445211>

Publisher: Association for Computing Machinery (ACM)

Persistent URL: <https://hdl.handle.net/1721.1/131130>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution 4.0 International license



Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online

Crystal Lee
crystall@mit.edu
Massachusetts Institute of Technology
Cambridge, MA, USA

Tanya Yang
tanyang@mit.edu
Massachusetts Institute of Technology
Cambridge, MA, USA

Gabrielle Inchoco
ginchoco@wellesley.edu
Wellesley College
Wellesley, MA, USA

Graham M. Jones
gmj@mit.edu
Massachusetts Institute of Technology
Cambridge, MA, USA

Arvind Satyanarayan
arvindsatya@mit.edu
Massachusetts Institute of Technology
Cambridge, MA, USA

ABSTRACT

Controversial understandings of the coronavirus pandemic have turned data visualizations into a battleground. Defying public health officials, coronavirus skeptics on US social media spent much of 2020 creating data visualizations showing that the government’s pandemic response was excessive and that the crisis was over. This paper investigates how pandemic visualizations circulated on social media, and shows that people who mistrust the scientific establishment often deploy the same rhetorics of data-driven decision-making used by experts, but to advocate for radical policy changes. Using a quantitative analysis of how visualizations spread on Twitter and an ethnographic approach to analyzing conversations about COVID data on Facebook, we document an epistemological gap that leads pro- and anti-mask groups to draw drastically different inferences from similar data. Ultimately, we argue that the deployment of COVID data visualizations reflect a deeper sociopolitical rift regarding the place of science in public life.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in visualization; Visualization theory, concepts and paradigms; Social media; Ethnographic studies.**

KEYWORDS

digital ethnography, network analysis, Twitter, Facebook, data literacy, data visualization

ACM Reference Format:

Crystal Lee, Tanya Yang, Gabrielle Inchoco, Graham M. Jones, and Arvind Satyanarayan. 2021. Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online. In *CHI Conference on Human Factors in Computing Systems (CHI '21)*, May 8–13, 2021, Yokohama, Japan. ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3411764.3445211>



This work is licensed under a Creative Commons Attribution International 4.0 License.

CHI '21, May 8–13, 2021, Yokohama, Japan

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8096-6/21/05.

<https://doi.org/10.1145/3411764.3445211>

1 INTRODUCTION

Throughout the coronavirus pandemic, researchers have held up the crisis as a “breakthrough moment” for data visualization research [91]: John Burn-Murdoch’s line chart comparing infection rates across countries helped millions of people make sense of the pandemic’s scale in the United States [44], and even top Trump administration officials seemed to rely heavily on the Johns Hopkins University COVID data dashboard [70]. Almost every US state now hosts a data dashboard on their health department website to show how the pandemic is unfolding. However, despite a preponderance of evidence that masks are crucial to reducing viral transmission [25, 29, 105], protestors across the United States have argued for local governments to overturn their mask mandates and begin re-opening schools and businesses. A pandemic that affects a few, they reason, should not impinge on the liberties of a majority to go about life as usual. To support their arguments, these protestors and activists have created thousands of their own visualizations, often using the same datasets as health officials.

This paper investigates how these activist networks use rhetorics of scientific rigor to oppose these public health measures. Far from ignoring scientific evidence to argue for individual freedom, anti-maskers often engage deeply with public datasets and make what we call “**counter-visualizations**”—visualizations using orthodox methods to make unorthodox arguments—to challenge mainstream narratives that the pandemic is urgent and ongoing. By asking community members to “follow the data,” these groups mobilize data visualizations to support significant local changes.

We examine the circulation of COVID-related data visualizations through both quantitative and qualitative methods. First, we conduct a quantitative analysis of close to half a million tweets that use data visualizations to talk about the pandemic. We use network analysis to identify communities of users who retweet the same content or otherwise engage with one another (e.g., maskers and anti-maskers). We process over 41,000 images through a computer vision model trained by Poco and Heer [83], and extract feature embeddings to identify clusters and patterns in visualization designs. The academic visualization research community has traditionally focused on mitigating chartjunk and creating more intuitive visualization tools for use by non-experts; *better* visualizations, researchers argue, would aid public understanding of data-driven phenomena. However, we find that anti-mask groups on Twitter often create polished counter-visualizations that would

not be out of place in scientific papers, health department reports, and publications like the *Financial Times*.

Second, we supplement this quantitative work with a six month-long observational study of anti-mask groups on Facebook. The period of this study, March to September 2020, was critical as it spanned the formation and consolidation of these groups at the pandemic's start. Quantitative analysis gives us an overview of what online discourse about data and its visual representation looks like on Twitter both within and outside anti-mask communities. Qualitative analysis of anti-mask groups gives us an interactional view of how these groups leverage the language of scientific rigor—being critical about data sources, explicitly stating analytical limitations of specific models, and more—in order to support ending public health restrictions despite the consensus of the scientific establishment. Our data analysis evolved as these communities did, and our methods reflect how these users reacted in real time to the kaleidoscopic nature of pandemic life. As of this writing, Facebook has banned some of the groups we studied, who have since moved to more unregulated platforms (Parler and MeWe).

While previous literature in visualization and science communication has emphasized the need for data and media literacy as a way to combat misinformation [43, 47, 89], this study finds that anti-mask groups practice a form of data literacy in spades. Within this constituency, unorthodox viewpoints do not result from a deficiency of data literacy; sophisticated practices of data literacy are a means of consolidating and promulgating views that fly in the face of scientific orthodoxy. Not only are these groups prolific in their creation of counter-visualizations, but they leverage data and their visual representations to advocate for and enact policy changes on the city, county, and state levels.

As we shall see throughout this case study, anti-mask communities on social media have defined themselves *in opposition* to the discursive and interpretive norms of the mainstream public sphere (e.g., against the “lamestream media”). In media studies, the term “counterpublic” describes constituencies that organize themselves in opposition to mainstream civic discourse, often by agentively using communications media [37]. In approaching anti-maskers as a counterpublic (a group shaped by its hostile stance toward mainstream science), we focus particular attention on one form of agentive media production central to their movement: data visualization. We define this counterpublic's visualization practices as “counter-visualizations” that use orthodox scientific methods to make unorthodox arguments, beyond the pale of the scientific establishment. Data visualizations are not a neutral window onto an observer-independent reality; during a pandemic, they are an arena of political struggle.

Among other initiatives, these groups argue for open access to government data (claiming that CDC and local health departments are not releasing enough data for citizens to make informed decisions), and they use the language of data-driven decision-making to show that social distancing mandates are both ill-advised and unnecessary. In these discussions, we find that anti-maskers think carefully about the grammar of graphics by decomposing visualizations into layered components (e.g., raw data, statistical transformations, mappings, marks, colors). Additionally, they debate how each component changes the narrative that the visualization tells, and they brainstorm alternate visualizations that would better enhance

public understanding of the data. This paper empirically shows how COVID anti-mask groups use data visualizations to argue that the US government's response (broadly construed) to the pandemic is overblown, and that the crisis has long been over.

These findings suggest that the ability for the scientific community and public health departments to better convey the urgency of the US coronavirus pandemic may not be strengthened by introducing more downloadable datasets, by producing “better visualizations” (e.g., graphics that are more intuitive or efficient), or by educating people on how to better interpret them. This study shows that there is a fundamental epistemological conflict between maskers and anti-maskers, who use the same data but come to such different conclusions. As science and technology studies (STS) scholars have shown, data is not a neutral substrate that can be used for good or for ill [14, 46, 84]. Indeed, anti-maskers often reveal themselves to be more sophisticated in their understanding of how scientific knowledge is socially constructed than their ideological adversaries, who espouse naive realism about the “objective” truth of public health data. Quantitative data is culturally and historically situated; the manner in which it is collected, analyzed, and interpreted reflects a deeper narrative that is bolstered by the collective effervescence found within social media communities. Put differently, there is no such thing as dispassionate or objective data analysis. Instead, there are stories: stories shaped by cultural logics, animated by personal experience, and entrenched by collective action. This story is about how a public health crisis—refracted through seemingly objective numbers and data visualizations—is part of a broader battleground about scientific epistemology and democracy in modern American life.

2 RELATED WORK

2.1 Data and visualization literacies

There is a robust literature in computer science on data and visualization literacy, where the latter often refers to the ability of a person “to comprehend and interpret graphs” [65] as well as the ability to create visualizations from scratch [20]. Research in this area often includes devising methods to assess this form of literacy [4, 15, 66], to investigate how people (mis)understand visualizations [21, 22, 81], or to create systems that help a user improve their understanding of unfamiliar visualizations [1, 3, 20, 87, 95, 96]. Evan Peck et al. [82] have responded to this literature by showing how a “*complex tapestry of motivations, preferences, and beliefs* [impact] *the way that participants* [prioritize] *data visualizations*,” suggesting that researchers need to better understand users' social and political context in order to design visualizations that speak powerfully to their personal experience.

Linguistic anthropologists have shown that “literacy” is not just the ability to encode and decode written messages. The skills related to reading and writing are historically embedded, and take on different meanings in a given social context depending on who has access to them, and what people think they should and shouldn't be used for. As a consequence of such contingencies, these scholars view literacy as multiple rather than singular, attending to the impact of local circumstances on their way members of any given community practice literacy and construe its value [2]. Thus, media literacy, is not simply about understanding information, but about

being able to actively leverage it in locally relevant social interactions [51]. Building on these traditions, we do not normatively assess anti-maskers' visualization practices against a prescriptivist model of what literacy *should* be (according to, say, experts in human-computer interaction), but rather seek to describe what those practices actually look like in locally relevant contexts.

As David Buckingham [17] has noted, calls for increased literacy have often become a form of wrong-headed solutionism that posits education as the fix to all media-related problems. danah boyd [16] has documented, too, that calling for increased media literacy can often backfire: the instruction to “question more” can lead to a weaponization of critical thinking and increased distrust of media and government institutions. She argues that calls for media literacy can often frame problems like fake news as ones of personal responsibility rather than a crisis of collective action. Similarly, Francesca Tripodi [98] has shown how evangelical voters do not vote for Trump because they have been “fooled” by fake news, but because they privilege the personal study of primary sources and have found logical inconsistencies not in Trump’s words, but in mainstream media portrayals of the president. As such, Tripodi argues, media literacy is not a means of fighting “alternative facts.” Christopher Bail et al. [8] have also shown how being exposed to more opposing views can actually increase political polarization.

Finally, in his study of how climate skeptics interpret scientific studies, Frank Fischer [42] argues that increasing fact-checking or levels of scientific literacy is insufficient for fighting alternative facts. “While fact checking is a worthy activity,” he says, “we need to look deeper into this phenomenon to find out what it is about, what is behind it.” Further qualitative studies that investigate how ideas are culturally and historically situated, as the discussion around COVID datasets and visualizations are manifestations of deeper political questions about the role of science in public life.

2.2 Critical approaches to visualization

Historians, anthropologists, and geographers have long shown how visualizations—far from an objective representation of knowledge—are often, in fact, representations of power [50, 59, 76, 97]. To address this in practice, feminist cartographers have developed quantitative GIS methods to describe and analyze differences across race, gender, class, and space, and these insights are then used to inform policymaking and political advocacy [48, 62, 72]. Best practices in data visualization have often emphasized reflexivity as a way to counter the power dynamics and systemic inequalities that are often inscribed in data science and design [31, 32, 35]. Central to this practice is articulating what is excluded from the data [18, 77, 78], understanding how data reflect situated knowledges rather than objective truth [28, 49, 93], and creating alternative methods of analyzing and presenting data based on anti-oppressive practices [5, 34, 57]. Researchers have also shown how interpreting data visualizations is a fundamentally social, narrative-driven endeavor [38, 55, 82]. By focusing on a user’s contextual experience and the communicative dimensions of a visualization, computer scientists have destabilized a more traditional focus on improving the technical components of data visualization towards understanding how users interpret and use them [64, 68, 101].

Critical, reflexive studies of data visualization are undoubtedly crucial for dismantling computational systems that exacerbate existing social inequalities. Research on COVID visualizations is already underway: Emily Bowe et al. [13] have shown how visualizations reflect the unfolding of the pandemic at different scales; Alexander Campolo [23] has documented how the pandemic has produced new forms of visual knowledge, and Yixuan Zhang et al. [106] have mapped the landscape of COVID-related crisis visualizations. This paper builds on these approaches to investigate the epistemological crisis that leads some people to conclude that mask-wearing is a crucial public health measure and for others to reject it completely.

Like data feminists, anti-mask groups similarly identify problems of political power within datasets that are released (or otherwise withheld) by the US government. Indeed, they contend that the way COVID data is currently being collected is non-neutral, and they seek liberation from what they see as an increasingly authoritarian state that weaponizes science to exacerbate persistent and asymmetric power relations. This paper shows that more critical approaches to visualization are necessary, and that the frameworks used by these researchers (e.g., critical race theory, gender analysis, and social studies of science) are crucial to disentangling how anti-mask groups mobilize visualizations politically to achieve powerful and often horrifying ends.

3 METHODS

This paper pairs a quantitative approach to analyzing Twitter data (computer vision and network analysis) with a qualitative approach to examining Facebook Groups (digital ethnography). Drawing from social media scholarship that uses mixed-methods approaches to examine how users interact with one another [6, 10, 19, 58, 75], this paper engages with work critical of quantitative social media methods [9, 103] by demonstrating how interpretive analyses of social media discussions and computational techniques can be mutually re-enforcing. In particular, we leverage quantitative studies of social media that use network analysis to understand political polarization [6], qualitative analysis of comments to identify changes in online dialogue over time [104], and visualization research that reverse-engineers and classifies chart images [88, 104].

3.1 Twitter data and quantitative analysis

3.1.1 Dataset. This analysis is conducted using a dataset of tweet IDs that Emily Chen et al. [27] assembled by monitoring the Twitter streaming API for certain keywords associated with COVID-19 (e.g., “coronavirus”, “pandemic”, “lockdown”, etc.) as well as by following the accounts of public health institutions (e.g., @CDCGov, @WHO, etc). We used version 2.7 of this dataset, which included over 390M tweets spanning January 21, 2020–July 31, 2020. This dataset consists of tweet IDs which we “hydrated” using twarc [36] into full tweets with associated metadata (e.g., hashtags, mentions, replies, favorites, retweets, etc.).

To identify tweets that primarily discuss data visualizations about the pandemic, we initially adopted a strategy that filtered for tweets that contained at least one image and keyword associated with data analysis (e.g., “bar,” “line,” but also “trend,” “data”). Unfortunately, this strategy yielded more noise than signal as most images in the resultant dataset were memes and photographs. We

therefore adopted a more conservative approach, filtering for tweets that explicitly mentioned chart-related keywords (i.e., “chart(s)”, “plot(s)”, “map(s)”, “dashboard(s)”, “vis”, “viz”, or “visualization(s)”). This process yielded a dataset of almost 500,000 tweets that incorporated over 41,000 images. We loaded the tweets and their associated metadata into a SQLite database, and the images were downloaded and stored on the file system.

3.1.2 Image classification. To analyze the types of visualization found in the dataset, we began by classifying every image in our corpus using the mark classification computer vision model trained by Poco and Heer [83]. Unfortunately, this model was only able to classify 30% of the images. As a result, we extracted a 4096-dimensional feature embedding for every image, and ran k-means clustering on a 100-dimensionally reduced space of these embeddings, for steps of k from 5–40. Two authors manually inspected the outputs of these runs, independently identified the most salient clusters, and then cross validated their analysis to assemble a final list of 8 relevant clusters: line charts, area charts, bar charts, pie charts, tables, maps, dashboards, and images. For dimensionality reduction and visualization, we used the UMAP algorithm [71] and iteratively arrived at the following parameter settings: 20 neighbors, a minimum distance of 0.01, and using the cosine distance metric. To account for UMAP’s stochasticity, we executed 10 runs and qualitatively examined the output to ensure our analyses were not based on any resultant artifacts.

3.1.3 Network analysis. Finally, to analyze the users participating in these discussions, we constructed a network graph: nodes were users who appeared in our dataset, and edges were drawn between pairs of nodes if a user mentioned, replied to, retweeted, or quote-tweeted another user. The resultant graph consisted of almost 400,000 nodes and over 583,000 edges, with an average degree of 2.9. To produce a denser network structure, we calculated a histogram of node degrees and identified that two-thirds of the nodes were degree 1. We then computed subgraphs, filtering for nodes with a minimum degree of 2 through 10 and found that degree 5 offered us a good balance between focusing on the most influential actors in the network, without losing smaller yet salient communities. This step yielded a subgraph of over 28,000 nodes and 104,000 edges, with an average degree of 7.3. We detected communities on this network using the Louvain method [12]. Of the 2,573 different communities detected by this algorithm, we primarily focus on the top 10 largest communities which account for 72% of nodes, 80% of edges, and 30% of visualizations.

3.2 Facebook data and qualitative analysis

3.2.1 Digital ethnography. While qualitative research can involve clinical protocols like interviews or surveys, Clifford Geertz [45] argues that the most substantial ethnographic insights into the cultural life of a community come from “**deep hanging out**,” i.e., long-term, participant observation alongside its members. Using “lurking,” a mode of participating by observing specific to digital platforms, we propose “**deep lurking**” as a way of systematically documenting the cultural practices of online communities. Our methods here rely on robust methodological literature in digital ethnography [30, 69], and we employ a case study approach [92]

to analyze these Facebook groups. To that end, we followed five Facebook groups (each with a wide range of followers, 10K–300K) over the first six months of the coronavirus pandemic, and we collected posts throughout the platform that included terms for “coronavirus” and “visualization” with Facebook’s CrowdTangle tool [33]. In our deep lurking, we archived web pages and took field notes on the following: posts (regardless of whether or not they included “coronavirus” and “data”), subsequent comments, Facebook Live streams, and photos of in-person events. We collected and analyzed posts from these groups from their earliest date to September 2020.

Taking a case study approach to the interactional Facebook data yields an analysis that ultimately complements the quantitative analysis. While the objective with analyzing Twitter data is statistical representativeness—we investigate which visualizations are the most popular, and in which communities—the objective of analyzing granular Facebook data is to accurately understand social dynamics within a singular community [92]. As such, the Twitter and Facebook analyses are foils of one another: we have the ability to quantitatively analyze large-scale interactions on Twitter, whereas we analyze the Facebook data by close reading and attending to specific context. Twitter communities are loosely formed by users retweeting, liking, or mentioning one another; Facebook groups create clearly bounded relationships between specific communities. By matching the affordances of each data source with the most ecologically appropriate method (network analysis and digital ethnography), this paper meaningfully combines qualitative and quantitative methods to understand data visualizations about the pandemic on a deeply contextual level and at scale.

3.2.2 Data collection & analysis. Concretely, we printed out posts as PDFs, tagged them with qualitative analysis software, and synthesized themes across these comments using grounded theory [61]. Grounded theory is an inductive method where researchers collect data and tag it by identifying analytically pertinent themes. Researchers then group these codes into higher-level concepts. As Kathy Charmaz [26] writes: these “methods consist of systematic, yet flexible guidelines for collecting and analyzing qualitative data to construct theories ‘grounded’ in the data themselves,” and these methods have since been adapted for social media analysis [85]. While this flexibility allows this method to respond dynamically to changing empirical phenomena, it can also lead to ambiguity about how new data fit with previously identified patterns. Digital ethnography also requires a longer time horizon than quantitative work in order to generate meaningful insights and, on its own, does not lead immediately to quantifiable results. These limitations are a major reason to use both qualitative and quantitative approaches. Following Emerson et al. [40], we employ an integrative strategy that weaves together “exemplars” from qualitative data alongside our interpretations. We have redacted the names of individual users and the Facebook groups we have studied, but we have preserved the dates and other metadata of each post within the article where possible.

3.2.3 Note on terminology. Throughout this study, we use the term “anti-mask” as a synecdoche for a broad spectrum of beliefs: that the pandemic is exaggerated, schools should be reopening, etc. While groups who hold these beliefs are certainly heterogeneous, the

mask is a common flashpoint throughout the ethnographic data, and they use the term “maskers” to describe people who are driven by fear. They are “anti-mask” by juxtaposition. This study therefore takes an emic (i.e. “insider”) approach to analyzing how members of these groups think, talk, and interact with one another, which starts by using terms that these community members would use to describe themselves. There is a temptation in studies of this nature to describe these groups as “anti-science,” but this would make it completely impossible for us to meaningfully investigate this article’s central question: understanding what these groups mean when they say “science.”

4 CASE STUDY

In the Twitter analysis, we quantitatively examine a corpus of tweets that use data visualizations to discuss the pandemic, and we create a UMAP visualization (figure 1) that identifies the types of visualizations that proliferate on Twitter. Then, we create a network graph (figure 2) of the users who share and interact with these data visualizations; the edges that link users in a network together are retweets, likes, mentions. We discover that the fourth largest network in our data consists of users promulgating heterodox scientific positions about the pandemic (i.e., anti-maskers). By comparing the visualizations shared within anti-mask and mainstream networks, we discover that there is no significant difference in the kinds of visualizations that the communities on Twitter are using to make drastically different arguments about coronavirus (figure 3). Anti-maskers (the community with the highest percentage of verified users) also share the second-highest number of charts across the top six communities (table 1), are the most prolific producers of area/line charts, and share the fewest number of photos (memes and images of politicians; see figure 3). Anti-maskers are also the most likely to amplify messages from their own community. We then examine the kinds of visualizations that anti-maskers discuss (figure 4).

This leads us to an interpretive question that animates the Facebook analysis: how can opposing groups of people use similar methods of visualization and reach such different interpretations of the data? We approach this problem by ethnographically studying interactions within a community of anti-maskers on Facebook to better understand their practices of knowledge-making and data analysis, and we show how these discussions exemplify a fundamental epistemological rift about how knowledge about the coronavirus pandemic should be made, interpreted, and shared.

4.1 Visualization design and network analysis

4.1.1 Visualization types. What kinds of visualizations are Twitter users sharing about the pandemic? Figure 1 visualizes the feature embeddings of images in our corpus, with color encoding clusters revealed and manually curated through k-means. Each circle is sized by the engagement the associated tweet received calculated as the sum of the number of favorites, replies, retweets, and quote tweets. Our analysis revealed eight major clusters: line charts (8908 visualizations, 21% of the corpus), area charts (2212, 5%), bar charts (3939, 9%), pie charts (1120, 3%), tables (4496, 11%), maps (5182, 13%), dashboards (2472, 6%), and images (7,128, 17%). The remaining 6,248 media (15% of the corpus) did not cluster in thematically coherent

ways. Here, we characterize salient elements and trends in these clusters.

Line charts represent the largest cluster of visualizations in our corpus. There are three major substructures: the first comprises line charts depicting the exponential growth of cases in the early stages of the pandemic, and predominantly use log-scales rather than linear scales. Charts from John Burn-Murdoch at the *Financial Times* and charts from the nonprofit Our World in Data are particularly prominent here. A second substructure consists of line charts comparing cases in the United States and the European Union when the US was experiencing its second wave of cases, and the third consists of line charts that visualize economic information. This substructure includes line charts of housing prices, jobs and unemployment, and stock prices (the latter appear to be taken from financial applications and terminals, and often feature additional candlestick marks). Across this cluster, these charts typically depict national or supranational data, include multiple series, and very rarely feature legends or textual annotations (other than labels for each line). Where they do occur, it is to label every point along the lines. Features of the graph are visually highlighted by giving some lines a heavier weight or graying other ones out.

Maps are the second largest cluster of visualizations in our corpus. The overwhelming majority of charts here are choropleths (shaded maps where a geographic region with high COVID rates might be darker, while low-rate regions are lighter). Other visualizations in this cluster include cartograms (the size of a geographic region is proportional its number of COVID infections as a method of comparison) and symbol maps (the size of a circle placed on a geographic region is proportional to COVID infections). The data for these charts span several geographic scales—global trends, country-level data (the US, China, and the UK being particularly salient), and municipal data (states and counties). These maps generally feature heavy annotation including direct labeling of geographic regions with the name and associated data value; arrows and callout boxes also better contextualize the data. For instance, in a widely shared map of the United Kingdom from the *Financial Times*, annotations described how “[t]hree Welsh areas had outbreaks in meatpacking plants in June” and that “Leicester, which is currently in an enforced local lockdown, has the second-highest rate...” These maps depict a wide range of data values including numbers of cases/deaths, metrics normalized per capita, rate of change for cases and/or deaths, mask adherence rates, and the effect of the pandemic on greenhouse gas emissions. Interestingly, choropleth maps of the United States electoral college at both the state- and district-level also appear in the corpus, with the associated tweets comparing the winner of particular regions with the type of pandemic response.

Area charts feature much heavier annotation than line charts (though fewer than maps). Peaks, troughs, and key events (e.g., when lockdowns occurred or when states reopened) are often shaded or labeled with arrows, and line marks are layered to highlight the overall trend or depict the rolling average. When these charts reflect data with a geographic correspondence, this data is often at a more local scale; line charts typically depict national or supranational data, and area charts more often visualized data at the state or county level. Notable subclusters in this group include the viral “Flatten the Curve” graphic, stacked area/streamgraphs, and “skinny bar” charts (charts of temporal data that closely resemble

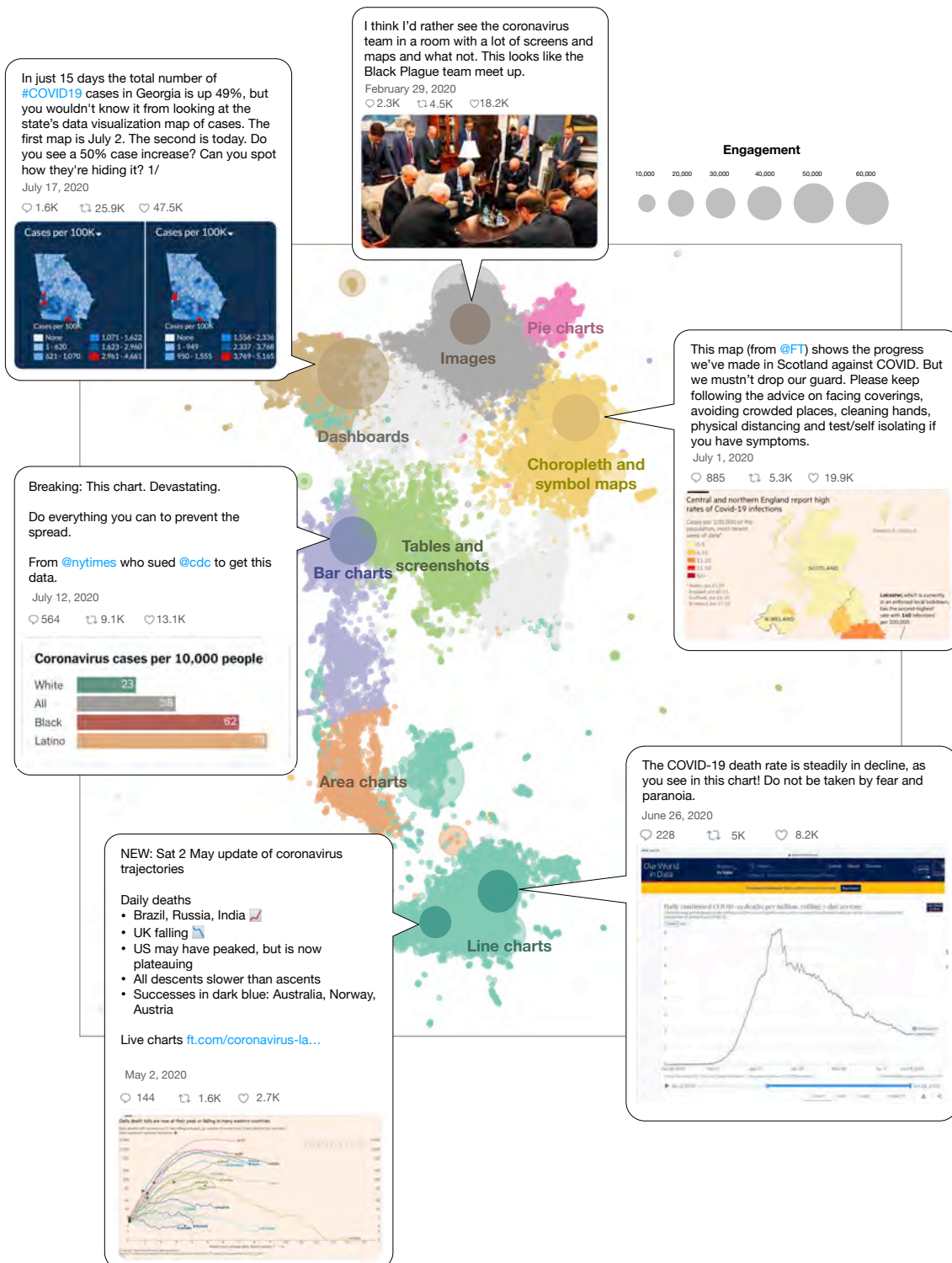


Figure 1: A UMAP visualization of feature embeddings of media found in our Twitter corpus. Color encodes labeled clusters, and size encodes the amount of engagement the media received (i.e., the sum of replies, favorites, retweets, and quote tweets).

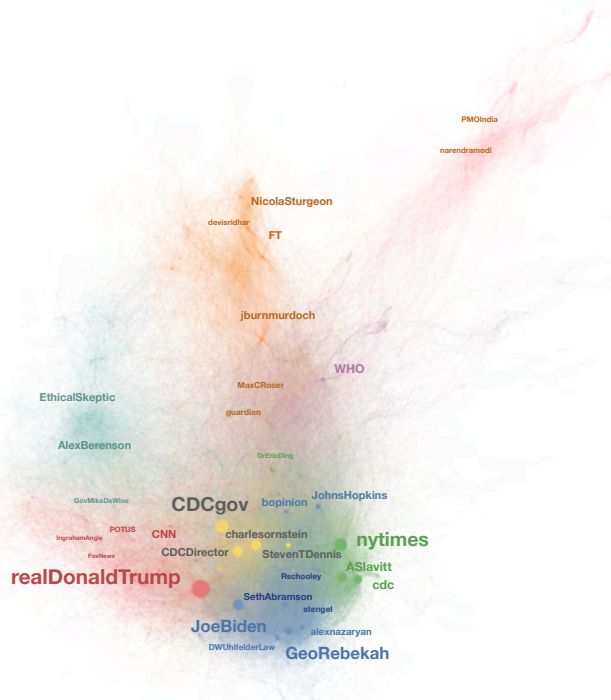


Figure 2: A network visualization of Twitter users appearing in our corpus. Color encodes community as detected by the Louvain method [12], and nodes are sized by their degree of connectedness (i.e., the number of other users they are connected to).

area charts, but use bar marks with narrow widths. Charts from the *New York Times* are especially prominent examples of the latter category—particularly screenshots of a red chart that was featured on the mobile front page.

Bar charts are predominantly encode categorical data and are more consistently and more heavily annotated than area charts. In addition to the annotations described for area charts (direct labeling of the tops of bars, labeled lines and arrows), charts in this cluster often include concise explainer texts.

These texts include some form of extended subtitles, more descriptive axis tick labels, or short passages before or after the bar chart that contextualize the data. Visually, the cluster is equally split between horizontal and vertical charts, and both styles feature a mix of layered, grouped, and stacked bars. Bar chart “races” (e.g., those developed with the Flourish visualization package) are one of the more frequently recurring idioms in this cluster. These are horizontal bar charts depicting the total number of cases per country, and animated over time.

Dashboards and images. While the remaining clusters are thematically coherent, we did not observe as rich a substructure within them. The dashboard cluster is overwhelmingly dominated by screenshots of the Johns Hopkins dashboard, and the image cluster is primarily comprised of reaction memes featuring the photos or caricatures of heads of state.

4.1.2 User networks. What are the different networks of Twitter users who share COVID-related data visualizations, and how do they interact with one another? Figure 2 depicts a network graph of Twitter users who discuss (or are discussed) in conversation with the visualizations in Figure 1. This network graph only shows users who are connected to at least five other users (i.e., by replying to them, mentioning them in a tweet, or re-tweeting or quote-tweeting them). The color of each network encodes a specific community as detected by the Louvain method [12], and the graph accounts for the top 10 communities (20,500 users or 72% of the overall graph). We describe the top six networks below listed in order of size (i.e., number of users within each network). While we have designated many of these communities with political orientation (e.g., left- or right-wing), these are only approximations; we recognize that these terms are fundamentally fluid and use them primarily as shorthand to make cross-network comparisons (e.g., mainstream political/media organizations vs. anti-mask protestors).

1. American politics and media (blue). This community features the American center-left, left, mainstream media, and popular or high profile figures (inside and outside of the scientific community). Accounts include politicians (@JoeBiden, @SenWarren), reporters (@joshtpm, @stengel), and public figures like Floridian data scientist @GeoRebekah and actor @GeorgeTakei. The user with the most followers in this community is @neiltyson.

2. American politics and right-wing media (red). This community includes members of the Trump administration, Congress, and right-wing personalities (e.g., @TuckerCarlson). Several accounts of mainstream media organizations also lie within this community (@CNN, @NBCNews), which reflects how often they mention the President (and other government accounts) in their coverage. Notably, these are official organizational accounts rather than those of individual reporters (which mostly show up in the previous group). Several mainstream media organizations are placed equally between these two clusters (@NPR, @washingtonpost). The user with the most followers in this community is @BarackObama.

3. British news media (orange). The largest non-American network roughly corresponds to news media in the UK, with a significant proportion of engagement targeted at the *Financial Times*’ successful visualizations by reporter John Burn-Murdoch, as well as coverage of politician Nicola Sturgeon’s coronavirus policies. The user with the most followers in this community is @TheEconomist.

4. Anti-mask network (teal). The anti-mask network comprises over 2,500 users (9% of our network graph) and is anchored by former *New York Times* reporter @AlexBerenson, blogger @EthicalSkeptic, and @justin_hart. *The Atlantic*’s @Covid19Tracking project (which collates COVID-19 testing rates and patient outcomes across the United States) and @GovMikeDeWine are also classified as part of this community. Governor DeWine of Ohio is not an anti-masker, but is often the target of anti-mask protest given his public health policies. Anti-mask users also lampoon *The Atlantic*’s project as another example of mainstream misinformation. These dynamics of intertextuality and citation within these networks are especially important here, as anti-mask groups often post screenshots of graphs from “lamestream media” organizations (e.g., *New York Times*) for the purpose of critique and analysis. The user with the most followers in this community is COVID skeptic and billionaire @elonmusk.

Table 1: Descriptive Statistics of Communities

Community #	Verified Users as % of Total Users	In-Network Retweets as % of Total Retweets	Original Tweets as % of Total Tweets
1	8.39	73.30	22.12
2	14.36	75.45	44.75
3	22.92	89.32	34.00
4	10.56	82.17	37.12
5	12.33	58.29	21.57
6	8.94	70.97	37.46

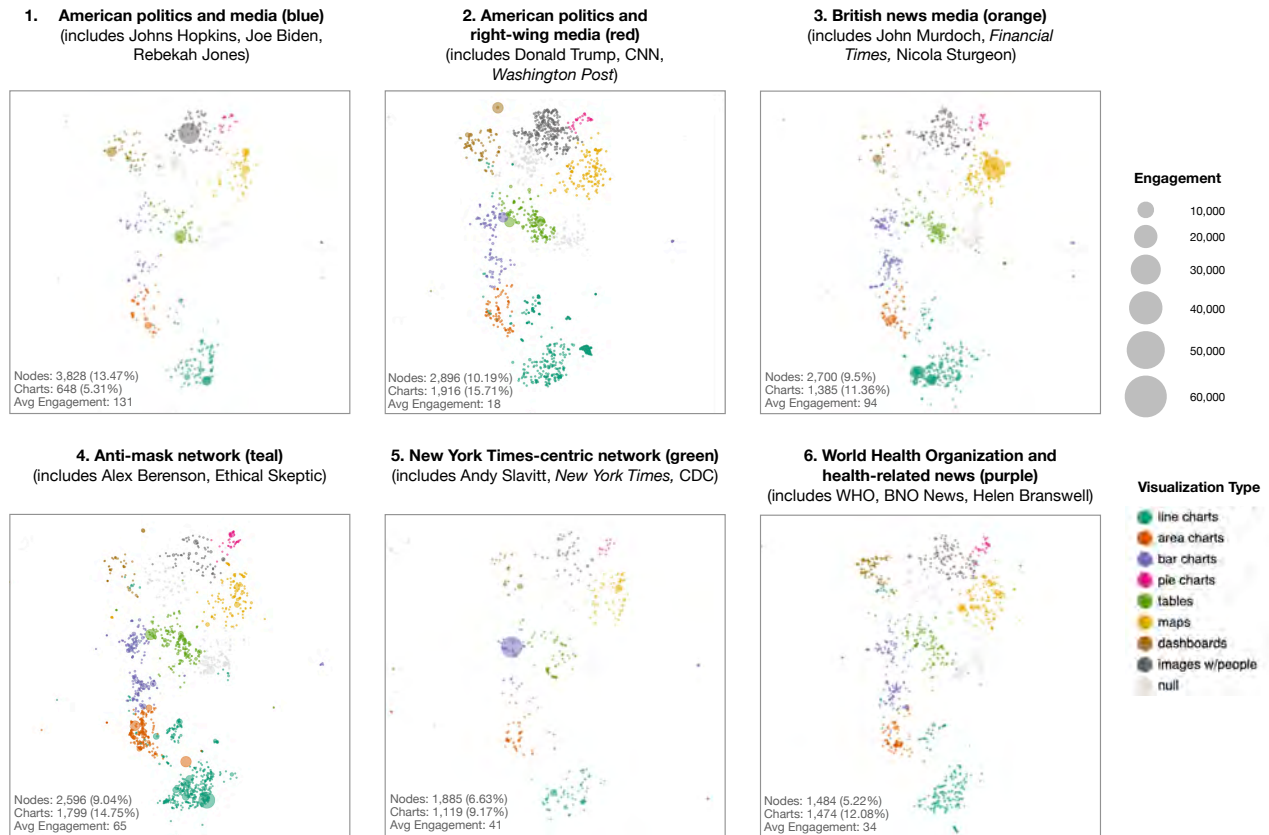


Figure 3: Visualizing the distribution of chart types by network community (with top accounts listed). While every community has produced at least one viral tweet, anti-mask users (group 6) receive higher engagement on average.

5. New York Times-centric network (green). This community is largely an artifact of a single visualization: Andy Slavitt (@ASlavitt), the former acting Administrator of the Centers for Medicare and Medicaid Services under the Obama administration, posted a viral tweet announcing the *New York Times* had sued the CDC (tagged with the incorrect handle @cdc instead of @CDCGov). The attached bar chart showing the racial disparity in COVID cases was shared widely with commentary directly annotated onto the

graph itself, or users analyzed the graph through quote-tweets and comments. The user with the most followers in this community is @NYTimes.

6. World Health Organization and health-related news organizations (purple). This community consists of global health organizations, particularly the @WHO and its subsidiary accounts (e.g., @WHOSEARO for Southeast Asia). The user with the most followers in this community is @YouTube.

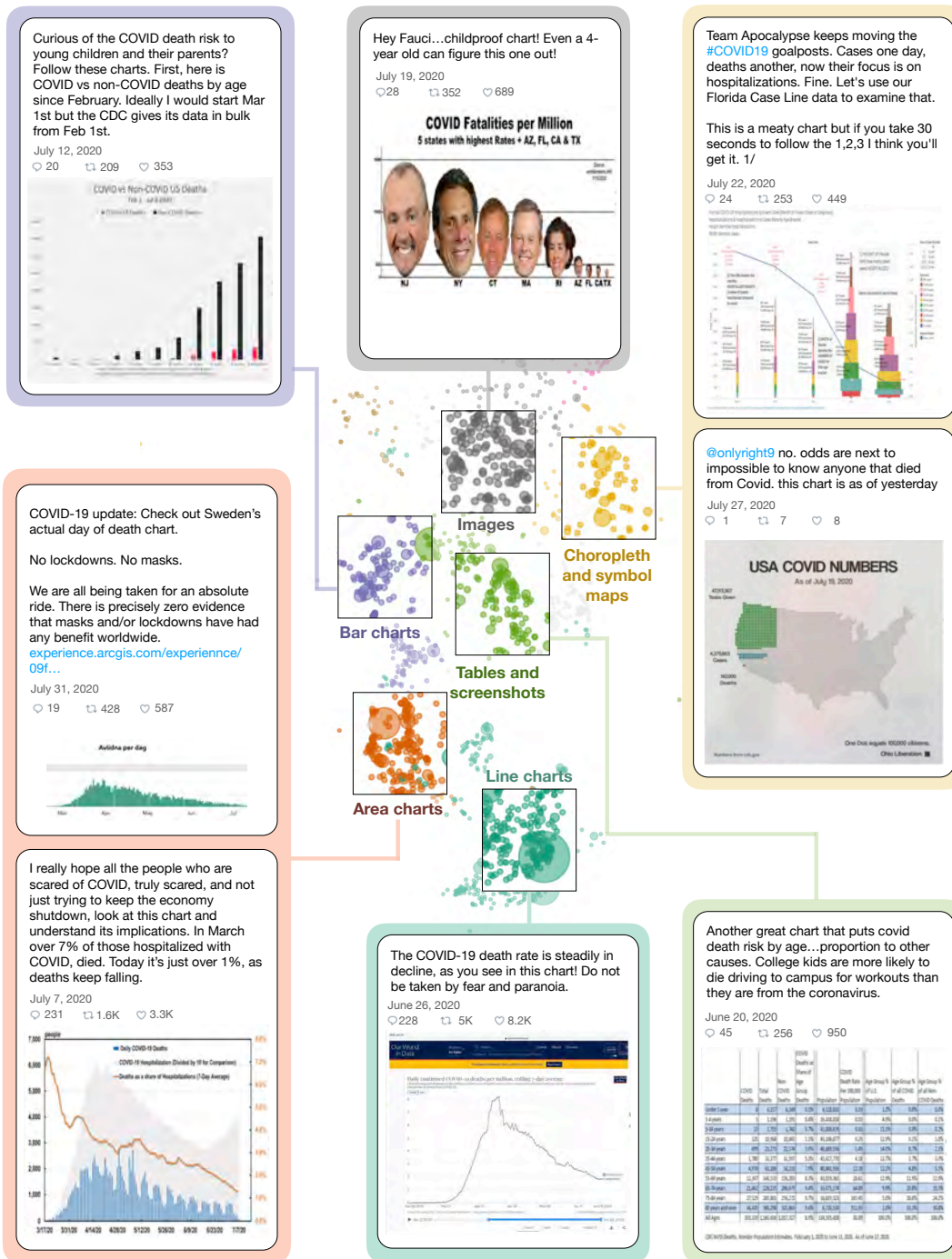


Figure 4: Sample counter-visualizations from the anti-mask user network. While there are meme-based visualizations, anti-maskers on Twitter adopt the same visual vocabulary as visualization experts and the mainstream media.

4.1.3 Descriptive statistics of communities. Table 1 lists summary statistics for the six largest communities in our dataset. There are three statistics of interest: the percentage of verified users (based on the total number of users within a community), the percentage of in-network retweets (based on a community’s total number of retweets), and the percentage of original tweets (based on a community’s total number of tweets). Twitter verification can often indicate that the account of public interest is authentic (subject to Twitter’s black-boxed evaluation standards); it can be a reasonable indication that the account is not a bot. Secondly, a high percentage of in-network retweets can be an indicator of how insular a particular network can be, as it shows how often a user amplifies messages from other in-network members. Finally, the percentage of original tweets shows how much of the content in that particular community is organic (i.e., they write their own content rather than simply amplifying existing work). Communities that have users who use the platform more passively (i.e., they prefer to lurk rather than comment) will have fewer original tweets; communities that have higher levels of active participation will have a higher number of original tweets as a percentage of total tweets.

The networks with the highest number of in-network retweets (which can be one proxy for insularity) are the British media (89.32%) and the anti-mask networks (82.17%), and the network with the highest percentage of original tweets is the American politics and right-wing media network (44.75%). Notably, the British news media network has both the *highest* percentage of verified users (22.92%), the *highest* percentage of in-network retweets (89.32%), and the fourth-highest percentage of original tweets (34.00%). As the third largest community in our dataset, we attribute this largely to the popularity of the graphs from the *Financial Times* from a few sources and the constellation of accounts that discussed those visualizations. While other communities (anti-mask, American politics/right-wing media, and WHO/health-related news) shared *more* visualizations, this network shared fewer graphs (1,385) that showed the *second-highest* level of engagement across the six communities (averaging 94 likes, retweets, or mentions per visualization). The network whose visualizations garner the highest level of engagement is the American politics and media network (131 likes, retweets, or mentions per visualization), but they only shared about half of the visualizations (648) compared to their British counterparts.

Through the descriptive statistics, we find that the anti-mask community exhibits very similar patterns to the rest of the networks in our dataset (it has about the same number of users with the same proportion of verified accounts). However, this community has the second highest percentage of in-network retweets (82.17% of all retweets) across the communities, and has the third-highest percentage of original tweets (37.12%, only trailing the World Health Organization network, at 37.46%).

4.1.4 Cross-network comparison of visualization types. Figure 3 depicts the distribution of visualization types by each community, along with descriptive statistics on the numbers of users, charts, and average engagement per tweet. These scatterplots show that there is little variance between the types of visualizations that users in each network share: almost all groups equally use maps or line, area, and bar charts. However, each group usually has one viral visualization—in group 3 (British media), the large yellow circle

represents a map from the *Financial Times* describing COVID-19 infections in Scotland; in group 5 (*New York Times*), the large purple circle in the center of the chart represents the viral bar chart from Andy Slavitt describing the racial disparities in COVID cases. The visualizations with the highest number of engagements in each of the six communities is depicted in figure 1.

Overall, we see that each group usually has one viral hit, but that the anti-mask users (group 4) tend to share a wide range of visualizations that garner medium levels of engagement (they have the third-highest number of average engagements in the six communities; an average of 65 likes, shares, and retweets). As a percentage of total tweets, anti-maskers have shared the second-highest number of charts across the top six communities (1,799 charts or 14.75%). They also use the *most* area/line charts and the *least* images across the six communities (images in this dataset usually include memes or photos of politicians). These statistics suggest that anti-maskers tend to be among the most prolific sharers of data visualizations on Twitter, and that they overwhelmingly amplify these visualizations to other users within their network (88.97% of all retweets are in-network).

4.1.5 Anti-mask visualizations. Figure 4 depicts the data visualizations that are shared by members of the anti-mask network accompanied by a select tweets from each category. While there are certainly visualizations that tend to use a meme-based approach to make their point (e.g., “Hey Fauci...childproof chart!” with the heads of governors used to show the rate of COVID fatalities), many of the visualizations shared by anti-mask Twitter users employ visual forms that are relatively similar to charts that one might encounter at a scientific conference. Many of these tweets use area and line charts to show the discrepancy between the number of projected deaths in previous epidemiological and the numbers of actual fatalities. Others use unit visualizations, tables, and bar charts to compare the severity of coronavirus to the flu. In total, this figure shows the breadth of visualization types that anti-mask users employ to illustrate that the pandemic is exaggerated.

4.2 Anti-mask discourse analysis

The Twitter analysis establishes that anti-maskers are prolific producers and consumers of data visualizations, and that the graphs that they employ are similar to those found in orthodox narratives about the pandemic. Put differently, anti-maskers use “data-driven” narratives to justify their heterodox beliefs. However, a quantitative overview of the visualizations they share and amplify does not in itself help us understand *how* anti-maskers invoke data and scientific reasoning to support policies like re-opening schools and businesses. Anti-maskers are acutely aware that mainstream narratives use data to underscore the pandemic’s urgency; they believe that these data sources and visualizations are fundamentally flawed and seek to counteract these biases. This section showcases the different ways that anti-mask groups talk about COVID-related data in discussion forums. What kinds of concerns do they have about the data used to formulate public policies? How do they talk about the limitations of data or create visualizations to convince other members in their physical communities that the pandemic is a hoax?

4.2.1 Emphasis on original content. Many anti-mask users express mistrust for academic and journalistic accounts of the pandemic, proposing to rectify alleged bias by “following the data” and creating their own data visualizations. Indeed, one Facebook group within this study has very strict moderation guidelines that prohibit the sharing of non-original content so that discussions can be “guided solely by the data.” Some group administrators even impose news consumption bans on themselves so that “mainstream” models do not “cloud their analysis.” In other words, **anti-maskers value unmediated access to information and privilege personal research and direct reading over “expert” interpretations.** While outside content is generally prohibited, Facebook group moderators encourage followers to make their own graphs, which are often shared by prominent members of the group to larger audiences (e.g., on their personal timelines or on other public facing Pages). Particularly in cases where a group or page is led by a few prominent users, follower-generated graphs tend to be highly popular because they often encourage other followers to begin their own data analysis projects, and comments on these posts often deal directly with how to reverse-engineer (or otherwise adjust) the visualization for another locality.

Since some of these groups are focused on a single state (e.g. “Re-open Nevada”), they can fill an information gap: not every county or locality is represented on data dashboards made by local newspapers, health departments, or city governments—if these government entities have dashboards or open data portals at all. In such cases, the emphasis on original content primarily reflects a grassroots effort to ensure access to pandemic-related data where there are no alternatives, and only secondarily serves to constitute an alternative to ideologically charged mainstream narratives. In the rare instances where mainstream visualizations are shared in such a group, it is usually to highlight the ways that mainstream analysis finally matches anti-mask projections, or to show how a journalist, government official, or academic can manipulate the same data source to purposefully mislead readers.

In order to create these original visualizations, users provide numerous tutorials on how to access government health data. These tutorials come either as written posts or as live screencasts, where a user (often a group administrator or moderator) demonstrates the process of downloading information from an open data portal. During the livestream, they editorialize to show which data are the most useful (e.g., “*the data you can download [from the Georgia Health Department website] is completely worthless, but the dashboard—which has data that everyday citizens cannot access—actually shows that there are no deaths whatsoever,*” July 13, 2020). Since many local health departments do not have the resources to stand up a new data system specifically for COVID-19, some redirect constituents to the state health department, which may not have granular data for a specific township available for public use. In the absence of data-sharing between states and local governments, users often take it upon themselves to share data with one another (e.g., “[redacted] *brings us this set of data from Minnesota. [...] Here it is in raw form, just superimposed on the model,*” May 17, 2020) and they work together to troubleshoot problems with each dataset (e.g., “*thanks. plugging [sic] in new .csv file to death dates is frustrating but worth it,*” May 2, 2020).

4.2.2 Critically assessing data sources. Even as these users learn from each other to collect more data, they remain critical about the circumstances under which the data are collected and distributed. Many of the users believe that the most important metrics are missing from government-released data. They express their concerns in four major ways. First, there is an ongoing animated debate within these groups about which metrics matter. Some users contend that deaths, not cases, should be the ultimate arbiter in policy decisions, since case rates are easily “manipulated” (e.g., with increased testing) and do not necessarily signal severe health problems (people can be asymptomatic). The shift in focus is important, as these groups believe that the emphasis on cases and testing often means that rates of COVID deaths by county or township are not reported to the same extent or seriously used for policy making. As one user noted, “*The Alabama public health department doesn’t provide deaths per day data (that I can tell—you can get it elsewhere). I sent a message asking about that. Crickets so far,*” (July 13, 2020).

Second, users also believe that state and local governments are deliberately withholding data so that they can unilaterally make decisions about whether or not lockdowns are effective. During a Facebook livestream with a Congressional candidate who wanted to “use data for reopening,” for example, both the candidate and an anti-mask group administrator discussed the extent to which state executives were willing to obscure the underlying data that were used to justify lockdown procedures (August 30, 2020). To illustrate this, the candidate emphasized a press conference in which journalists asked the state executive whether they would consider making the entire contact tracing process public, which would include releasing the name of the bar where the outbreak started. In response, the governor argued that while transparency about the numbers were important, the state would not release the name of the bar, citing the possibility of stigmatization and an erosion of privacy. This soundbite—“*we have the data, but we won’t give it to you*”—later became a rallying cry for anti-mask groups in this state following this livestream. “*I hate that they’re not being transparent in their numbers and information they’re giving out,*” another user wrote. “*They need to be honest and admit they messed up if it isn’t as bad as they’re making it out to be. [...] We need honesty and transparency.*”

This plays into a third problem that users identify with the existing data: that datasets are constructed in fundamentally subjective ways. They are coded, cleaned, and aggregated either by government data analysts with nefarious intentions or by organizations who may not have the resources to provide extensive documentation. “*Researchers can define their data set anyhow [sic] they like in absence of generally accepted (preferably specified) definitions,*” one user wrote on June 23, 2020. “*Coding data is a big deal—and those definitions should be offered transparently by every state. Without a national guideline—we are left with this mess.*” The lack of transparency within these data collection systems—which many of these users infer as a lack of honesty—erodes these users’ trust within both government institutions and the datasets they release.

Even when local governments do provide data, however, these users also contend that the data requires context in order for any interpretation to be meaningful. For example, metrics like hospitalization and infection rates are still “*vulnerable to all sorts of issues that make [these] data less reliable than deaths data*” (June 23, 2020),

and require additional inquiry before the user considers making a visualization. In fact, there are multiple threads every week where users debate how representative the data are of the population given the increased rate of testing across many states. For some users, random sampling is the only way to really know the true infection rate, as (1) testing only those who show symptoms gives us an artificially high infection rate, and (2) testing asymptomatic people tells us what we already know—that the virus is not a threat. These groups argue that the conflation of asymptomatic and symptomatic cases therefore makes it difficult for anyone to actually determine the severity of the pandemic. “We are counting ‘cases’ in ways we never did for any other virus,” a user writes, “and we changed how we counted in the middle of the game. It’s classic garbage in, garbage out at this point. If it could be clawed back to ONLY symptomatic and/or contacts, it could be a useful guide [for comparison], but I don’t see that happening” (August 1, 2020).

Similarly, these groups often question the context behind measures like “excess deaths.” While the CDC has provided visualizations that estimate the number excess deaths by week [25], users take screenshots of the websites and debate whether or not they can be attributed to the coronavirus. “You can’t simply subtract the current death tally from the typical value for this time of year and attribute the difference to Covid,” a user wrote. “Because of the actions of our governments, we are actually causing excess deaths. Want to kill an old person quickly? Take away their human interaction and contact. Or force them into a rest home with other infected people. Want people to die from preventable diseases? Scare them away from the hospitals, and encourage them to postpone their medical screenings, checkups, and treatments [...] The numbers are clear. By trying to mitigate one problem, we are creating too many others, at too high a price” (September 5, 2020).

4.2.3 Critically assessing data representations. Even beyond downloading datasets from local health departments, users in these groups are especially attuned to the ways that specific types of visualizations can obscure or highlight information. In response to a visualization where the original poster (OP) created a bar chart of death counts by county, a user commented: “the way data is presented can also show bias. For example in the state charts, counties with hugely different populations can be next to each other. The smaller counties are always going to look calm even if per capita they are doing the same or worse. Perhaps you could do a version of the charts where the hardest hit county is normalized per capita to 1 and compare counties that way,” to which the OP responded, “it is never biased to show data in its entirety, full scale” (August 14, 2020).

An ongoing topic of discussion is whether to visualize absolute death counts as opposed to deaths per capita, and it is illustrative of a broader mistrust of mediation. For some, “raw data” (e.g., counts) provides more accurate information than any data transformation (e.g., death rate per capita, or even visualizations themselves). For others, screenshots of tables are the most faithful way to represent the data, so that people can see and interpret it for themselves. “No official graphs,” said one user. “Raw data only. Why give them an opportunity to spin?” (June 14, 2020). These users want to understand and analyze the information for themselves, free from biased, external intervention.

4.2.4 Identifying bias and politics in data. While users contend that their data visualizations objectively illustrate how the pandemic is no worse than the flu, they are similarly mindful to note that these analyses only represent partial perspectives that are subject to individual context and interpretation. “I’ve never claimed to have no bias. Of course I am biased, I’m human,” says one prolific producer of anti-mask data visualizations. “That’s why scientists use controls... to protect ourselves from our own biases. And this is one of the reasons why I disclose my biases to you. That way you can evaluate my conclusions in context. Hopefully, by staying close to the data, we keep the effect of bias to a minimum” (August 14, 2020). They are ultimately mindful of the subjectivity of human interpretation, which leads them to analyzing the data for themselves.

More tangibly, however, these groups seek to identify bias by being critical about specific profit motives that come from releasing (or suppressing) specific kinds of information. Many of the users within these groups are skeptical about the potential benefits of a coronavirus vaccine, and as a point of comparison, they often reference how the tobacco industry has historically manipulated science to mislead consumers. These groups believe that pharmaceutical companies have similarly villainous profit motives, which leads the industry to inflate data about the pandemic in order to stoke demand for a vaccine. As one user lamented, “I wish more of the public would do some research into them and see how much of a risk they are but sadly most wont [sic]—because once you do and you see the truth on them, you get labeled as an ‘antivaxxer’ which equates to fool. In the next few years, the vaccine industry is set to be a nearly 105 billion dollar industry. People should really consider who profits off of our ignorance” (August 24, 2020).

4.2.5 Appeals to scientific authority. Paradoxically, these groups also seek ways to validate their findings through the scientific establishment. Many users prominently display their scientific credentials (e.g., referring to their doctoral degrees or prominent publications in venues like Nature) which uniquely qualify them as insiders who are most well-equipped to criticize the scientific community. Members who perform this kind of expertise often point to 2013 Nobel Laureate Michael Levitt’s assertion that lockdowns do nothing to save lives [67] as another indicator of scientific legitimacy. Both Levitt and these anti-mask groups identify the dangerous convergence of science and politics as one of the main barriers to a more reasonable and successful pandemic response, and they construct their own data visualizations as a way to combat what they see as health misinformation. “To be clear. I am not downplaying the COVID epidemic,” said one user. “I have never denied it was real. Instead, I’ve been modeling it since it began in Wuhan, then in Europe, etc. [...] What I have done is follow the data. I’ve learned that governments, that work for us, are too often deliberately less than transparent when it comes to reporting about the epidemic” (July 17, 2020). For these anti-mask users, their approach to the pandemic is grounded in a more scientific rigor, not less.

4.2.6 Developing expertise and processes of critical engagement. The goal of many of these groups is ultimately to develop a network of well-informed citizens engaged in analyzing data in order to make measured decisions during a global pandemic. “The other side says that they use evidence-based medicine to make decisions,” one user wrote, “but the data and the science do not support current

actions” (August 30, 2020). The discussion-based nature of these Facebook groups also give these followers a space to learn and adapt from others, and to develop processes of critical engagement. Long-time followers of the group often give small tutorials to new users on how to read and interpret specific visualizations, and users often give each other constructive feedback on how to adjust their graphic to make it more legible or intuitive. Some questions and comments would not be out of place at all at a visualization research poster session: “*This doesn’t make sense. What do the colors mean? How does this demonstrate any useful information?*” (July 21, 2020) These communities use data analysis as a way to socialize and enculturate their users; they promulgate data literacy practices as a way of inculcating heterodox ideology. The transmission of data literacy, then, becomes a method of political radicalization.

These individuals as a whole are extremely willing to help others who have trouble interpreting graphs with multiple forms of clarification: by helping people find the original sources so that they can replicate the analysis themselves, by referencing other reputable studies that come to the same conclusions, by reminding others to remain vigilant about the limitations of the data, and by answering questions about the implications of a specific graph. The last point is especially salient, as it surfaces both what these groups see as a reliable measure of how the pandemic is unfolding and what they believe they should do with the data. These online communities therefore act as a sounding board for thinking about how best to effectively mobilize the data towards more measured policies like slowly reopening schools. “*You can tell which places are actually having flare-ups and which ones aren’t,*” one user writes. “*Data makes us calm.*” (July 21, 2020)

Additionally, **followers in these groups also use data analysis as a way of bolstering social unity and creating a community of practice.** While these groups highly value scientific expertise, they also see collective analysis of data as a way to bring communities together within a time of crisis, and being able to transparently and dispassionately analyze the data is crucial for democratic governance. In fact, the explicit motivation for many of these followers is to find information so that they can make the best decisions for their families—and by extension, for the communities around them. “*Regardless of your political party, it is incumbent on all of us to ask our elected officials for the data they use to make decisions,*” one user said during a live streamed discussion. “*I’m speaking to you as a neighbor: request the data. [...] As a Mama Bear, I don’t care if Trump says that it’s okay, I want to make a decision that protects my kids the most. This data is especially important for the moms and dads who are concerned about their babies*” (August 30, 2020). As Kate Starbird et al. have demonstrated, strategic information operations require the participation of online communities to consolidate and amplify these messages: these messages become powerful when emergent, organic crowds (rather than hired trolls and bots) iteratively contribute to a larger community with shared values and epistemologies [94].

Group members repost these analyses onto their personal timelines to start conversations with friends and family in hopes that they might be able to congregate in person. However, many of these conversations result in frustration. “*I posted virus data from the CDC, got into discussion with people and in the end several straight out voiced they had no interest in the data,*” one user sighed. “*My post*

said ‘Just the facts.’ [screenshot from the CDC] People are emotionally invested in their beliefs and won’t be swayed by data. It’s disturbing” (August 14, 2020). Especially when these conversations go poorly, followers solicit advice from each other about how to move forward when their children’s schools close or when family members do not “follow the data.” One group even organized an unmasked get-together at a local restaurant where they passed out t-shirts promoting their Facebook group, took selfies, and discussed a lawsuit that sought to remove their state’s emergency health order (September 12, 2020). The lunch was organized such that the members who wanted to first attend a Trump rolling rally could do so and “*drop in afterward for some yummy food and fellowship*” (September 8, 2020).

4.2.7 Applying data to real-world situations. Ultimately, anti-mask users emphasize that they need to apply this data to real-world situations. The same group that organized the get-together also regularly hosts live-streams with guest speakers like local politicians, congressional candidates, and community organizers, all of whom instruct users on how to best agitate for change armed with the data visualizations shared in the group. “*You’re a mom up the street, but you’re not powerless,*” emphasized one of the guest speakers. “*Numbers matter! What is just and what is true matters. [...] Go up and down the ladder—start real local. Start with the lesser magistrates, who are more accessible, easier to reach, who will make time for you.*” (July 23, 2020)

These groups have been incredibly effective at galvanizing a network of engaged citizens towards concrete political action. Local officials have relied on data narratives generated in these groups to call for a lawsuit against the Ohio Department of Health (July 20, 2020). In Texas, a coalition of mayors, school board members, and city council people investigated the state’s COVID-19 statistics and discovered that a backlog of unaudited tests was distorting the data, prompting Texas officials to employ a forensic data team to investigate the surge in positive test rates [24]. “*There were over a million pending assignments [that were distorting the state’s infection rate],*” the city councilperson said to the group’s 40,000+ followers. “*We just want to make sure that the information that is getting out there is giving us the full picture.*” (August 17, 2020) Another Facebook group solicited suggestions from its followers on how to support other political groups who need data to support lawsuits against governors and state health departments. “*If you were suddenly given access to all the government records and could interrogate any official,*” a group administrator asked, “*what piece of data or documentation would you like to inspect?*” (September 11, 2020) The message that runs through these threads is unequivocal: that data is the only way to set fear-bound politicians straight, and using better data is a surefire way towards creating a safer community.

5 DISCUSSION

Anti-maskers have deftly used social media to constitute a cultural and discursive arena devoted to addressing the pandemic and its fallout through practices of data literacy. Data literacy is a quintessential criterion for membership within the community they have created. The prestige of both individual anti-maskers and the larger Facebook groups to which they belong is tied to displays of skill in accessing, interpreting, critiquing, and visualizing data, as

well as the pro-social willingness to share those skills with other interested parties. This is a community of practice [63, 102] focused on acquiring and transmitting expertise, and on translating that expertise into concrete political action. Moreover, this is a subculture shaped by mistrust of established authorities and orthodox scientific viewpoints. Its members value individual initiative and ingenuity, trusting scientific analysis only insofar as they can replicate it themselves by accessing and manipulating the data firsthand. They are highly reflexive about the inherently biased nature of any analysis, and resent what they view as the arrogant self-righteousness of scientific elites.

As a subculture, anti-masking amplifies anti-establishment currents pervasive in U.S. political culture. Data literacy, for anti-maskers, exemplifies distinctly American ideals of intellectual self-reliance, which historically takes the form of rejecting experts and other elites [53]. The counter-visualizations that they produce and circulate not only challenge scientific consensus, but they also assert the value of independence in a society that they believe promotes an overall de-skilling and dumbing-down of the population for the sake of more effective social control [39, 52, 98]. As they see it, to counter-visualize is to engage in an act of resistance against the stifling influence of central government, big business, and liberal academia. Moreover, their simultaneous appropriation of scientific rhetoric and rejection of scientific authority also reflects longstanding strategies of Christian fundamentalists seeking to challenge the secularist threat of evolutionary biology [11].

So how do these groups diverge from scientific orthodoxy if they are using the same data? We have identified a few sleights of hand that contribute to the broader epistemological crisis we identify between these groups and the majority of scientific researchers. For instance, they argue that there is an outsized emphasis on *deaths* versus cases: if the current datasets are fundamentally subjective and prone to manipulation (e.g., increased levels of faulty testing, asymptomatic vs. symptomatic cases), then deaths are the only reliable markers of the pandemic’s severity. Even then, these groups believe that deaths are an additionally problematic category because doctors are using a COVID diagnosis as the main cause of death (i.e., people who die because of COVID) when in reality there are other factors at play (i.e., dying with but not because of COVID). Since these categories are fundamentally subject to human interpretation, especially by those who have a vested interest in reporting as many COVID deaths as possible, these numbers are vastly over-reported, unreliable, and no more significant than the flu.

Another point of contention is that of lived experience: in many of these cases, users do not themselves know a person who has experienced COVID, and the statistics they see on the news show the severity of the pandemic in vastly different parts of the country. Since they do not see their experience reflected in the narratives they consume, they look for hyperlocal data to help guide their decision-making. But since many of these datasets do not always exist on such a granular level, this information gap feeds into a larger social narrative about the government’s suppression of critical data and the media’s unwillingness to substantively engage with the subjectivity of coronavirus data reporting.

Most fundamentally, the groups we studied believe that **science is a process, and not an institution**. As we have outlined in the

case study, these groups mistrust the scientific establishment (“Science”) because they believe that the institution has been corrupted by profit motives and politics. The knowledge that the CDC and academics have created cannot be trusted because they need to be subject to increased doubt, and not accepted as consensus. In the same way that climate change skeptics have appealed to Karl Popper’s theory of falsification to show why climate science needs to be subjected to continuous scrutiny in order to be valid [42], we have found that anti-mask groups point to Thomas Kuhn’s *The Structure of Scientific Revolutions* to show how their anomalous evidence—once dismissed by the scientific establishment—will pave the way to a new paradigm (“As I’ve recently described, I’m no stranger to presenting data that are inconsistent with the narrative. It can get ugly. People do not give up their paradigms easily. [...] Thomas Kuhn wrote about this phenomenon, which occurs repeatedly throughout history. Now is the time to hunker down. Stand with the data,” August 5, 2020). For anti-maskers, valid science must be a process they can critically engage for themselves in an unmediated way. *Increased* doubt, not consensus, is the marker of scientific certitude.

Arguing that anti-maskers simply need more scientific literacy is to characterize their approach as uninformed and inexplicably extreme. This study shows the opposite: users in these communities are deeply invested in forms of critique and knowledge production that they recognize as markers of scientific expertise. If anything, anti-mask science has extended the traditional tools of data analysis by taking up the theoretical mantle of recent critical studies of visualization [31, 35]. Anti-mask approaches acknowledge the subjectivity of how datasets are constructed, attempt to reconcile the data with lived experience, and these groups seek to make the process of understanding data as transparent as possible in order to challenge the powers that be. For example, one of the most popular visualizations within the Facebook groups we studied were unit visualizations, which are popular among anti-maskers and computer scientists for the same reasons: they provide more information, better match a reader’s mental model, and they allow users to interact with them in new and more interesting ways [80]. Barring tables, they are the most unmediated way to interact with data: one dot represents one person.

Similarly, these groups’ impulse to mitigate bias and increase transparency (often by dropping the use of data they see as “biased”) echoes the organizing ethos of computer science research that seeks to develop “technological solutions regarding potential bias” or “ground research on fairness, accountability, and transparency” [7]. In other words, these groups see themselves as engaging deeply within multiple aspects of the scientific process—interrogating the datasets, analysis, and conclusions—and still university researchers might dismiss them in leading journals as “scientifically illiterate” [74]. In an interview with the Department of Health and Human Services podcast, even Anthony Fauci (Chief Medical Advisor to the US President) noted: “one of the problems we face in the United States is that unfortunately, there is a combination of an anti-science bias [...] people are, for reasons that sometimes are, you know, inconceivable and not understandable, they just don’t believe science” [41].

We use Dr. Fauci’s provocation to illustrate how understanding the way that anti-mask groups think about science is crucial to grappling with the contested state of expertise in American democracy. In a study of Tea Party supporters in Louisiana, Arlie Russell

Hochschild [52] explains the intractable partisan rift in American politics by emphasizing the importance of a “deep story”: a subjective prism that people use in order to make sense of the world and guide the way they vote. For Tea Party activists, this deep story revolved around anger towards a federal system ruled by liberal elites who pander to the interests of ethnic and religious minorities, while curtailing the advantages that White, Christian traditionalists view as their American birthright. We argue that the anti-maskers’ deep story draws from similar wells of resentment, but adds a particular emphasis on the usurpation of scientific knowledge by a paternalistic, condescending elite that expects intellectual subservience rather than critical thinking from the lay public.

To be clear, we are not promoting these views. Instead, we seek to better understand how data literacy, as a both a set of skills and a moral virtue championed within academic computer science, can take on distinct valences in different cultural contexts. A more nuanced view of data literacy, one that recognizes multiplicity rather than uniformity, offers a more robust account of how data visualization circulates in the world. This culturally and socially situated analysis demonstrates why increasing access to raw data or improving the informational quality of data visualizations is not sufficient to bolster public consensus about scientific findings. Projects that examine the cognitive basis of visualization or seek to make “better” or “more intuitive” visualizations [60] will not meaningfully change this phenomenon: anti-mask protestors already use visualizations, and do so extremely effectively. Moreover, in emphasizing the politicization of pandemic data, our account helps to explain the striking correlation between practices of counter-visualization and the politics of anti-masking. For members of this social movement, counter-visualization and anti-masking are complementary aspects of resisting the tyranny of institutions that threaten to usurp individual liberties to think freely and act accordingly.

6 IMPLICATIONS AND CONCLUSION

This paper has investigated anti-mask counter-visualizations on social media in two ways: quantitatively, we identify the main types of visualizations that are present within different networks (e.g., pro and anti-mask users), and we show that anti-mask users are prolific and skilled purveyors of data visualizations. These visualizations are popular, use orthodox visualization methods, and are promulgated as a way to convince others that public health measures are unnecessary. In our qualitative analysis, we use an ethnographic approach to illustrate how COVID counter-visualizations actually reflect a deeper epistemological rift about the role of data in public life, and that the practice of making counter-visualizations reflects a participatory, heterodox approach to information sharing. Convincing anti-maskers to support public health measures in the age of COVID-19 will require more than “better” visualizations, data literacy campaigns, or increased public access to data. Rather, it requires a sustained engagement with the social world of visualizations and the people who make or interpret them.

While academic science is traditionally a system for producing knowledge within a laboratory, validating it through peer review, and sharing results within subsidiary communities, anti-maskers reject this hierarchical social model. They espouse a vision of science that is radically egalitarian and individualist. This study forces

us to see that coronavirus skeptics champion science as a personal practice that prizes rationality and autonomy; for them, it is *not* a body of knowledge certified by an institution of experts. Calls for data or scientific literacy therefore risk recapitulating narratives that anti-mask views are the product of individual ignorance rather than coordinated information campaigns that rely heavily on networked participation. Recognizing the *systemic* dynamics that contribute to this epistemological rift is the first step towards grappling with this phenomenon, and the findings presented in this paper corroborate similar studies about the impact of fake news on American evangelical voters [98] and about the limitations of fact-checking climate change denialism [42].

Calls for media literacy—especially as an ethics smokescreen to avoid talking about larger structural problems like white supremacy—are problematic when these approaches are deficit-focused and trained primarily on individual responsibility. Powerful research and media organizations paid for by the tobacco or fossil fuel industries [79, 86] have historically capitalized on the skeptical impulse that the “science simply isn’t settled,” prompting people to simply “think for themselves” to horrifying ends. The attempted coup on January 6, 2021 has similarly illustrated that well-calibrated, well-funded systems of coordinated disinformation can be particularly dangerous when they are *designed* to appeal to skeptical people. While individual insurrectionists are no doubt to blame for their own acts of violence, the coup relied on a collective effort fanned by people questioning, interacting, and sharing these ideas with other people. These skeptical narratives are powerful because they resonate with these people’s lived experience and—crucially—because they are posted by influential accounts across influential platforms.

Broadly, the findings presented in this paper also challenge conventional assumptions in human-computer interaction research about who imagined users might be: visualization experts traditionally design systems for scientists, business analysts, or journalists. Researchers create systems intended to democratize processes of data analysis and inform a broader public about how to use data, often in the clean, sand-boxed environment of an academic lab. However, this literature often focuses narrowly on promoting expressivity (either of current or new visualization techniques), assuming that improving visualization tools will lead to improving public understanding of data. This paper presents a community of users that researchers might not consider in the systems building process (i.e., supposedly “data illiterate” anti-maskers), and we show how the binary opposition of literacy/illiteracy is insufficient for describing how orthodox visualizations can be used to promote unorthodox science. Understanding how these groups skillfully manipulate data to undermine mainstream science requires us to adjust the theoretical assumptions in HCI research about how data can be leveraged in public discourse.

What, then, are visualization researchers and social scientists to do? One step might be to grapple with the social and political dimensions of visualizations at the *beginning*, rather than the end, of projects [31]. This involves in part a shift from positivist to interpretivist frameworks in visualization research, where we recognize that knowledge we produce in visualization systems is fundamentally “multiple, subjective, and socially constructed” [73]. A secondary issue is one of uncertainty: Jessica Hullman and Zeynep Tufekci

(among others) have both showed how *not* communicating the uncertainty inherent in scientific writing has contributed to the erosion of public trust in science [56, 100]. As Tufekci demonstrates (and our data corroborates), the CDC’s initial public messaging that masks were ineffective—followed by a quick public reversal—seriously hindered the organization’s ability to effectively communicate as the pandemic progressed. As we have seen, people are not simply passive consumers of media: anti-mask users *in particular* were predisposed to digging through the scientific literature and highlighting the uncertainty in academic publications that media organizations elide. When these uncertainties did not surface within public-facing versions of these studies, people began to assume that there was a broader cover-up [99].

But as Hullman shows, there are at least two major reasons why uncertainty *hasn’t* traditionally been communicated to the public [54]. Researchers often do not believe that people will understand and be able to interpret results that communicate uncertainty (which, as we have shown, is a problematic assumption at best). However, visualization researchers also do not have a robust body of understanding about *how*, and when, to communicate uncertainty (let alone how to do so effectively). There are exciting threads of visualization research that investigate how users’ interpretive frameworks can change the overarching narratives they glean from the data [55, 82, 90]. Instead of championing absolute certitude or objectivity, this research pushes us to ask how scientists and visualization researchers alike might express uncertainty in the data so as to recognize its socially and historically situated nature.

In other words, our paper introduces new ways of thinking about “democratizing” data analysis and visualization. Instead of treating increased adoption of data-driven storytelling as an unqualified good, we show that data visualizations are not simply tools that people use to understand the epidemiological events around them. They are a battleground that highlight the contested role of expertise in modern American life.

ACKNOWLEDGMENTS

The authors thank Stephan Risi, Maeva Fincker, and Mateo Monterde for their assistance with quantitative methods and supplemental material. We also thank the members of the Visualization Group (especially Jonathan Zong, Alan Lundgard, Harini Suresh, EJ Sefah, and the Fall 2020 UROP cohort: Anna Arpacı-Dusseau, Anna Meurer, Ethan Nevidomsky, Kat Huang, and Soomin Chun). This manuscript benefited from the insights of Rodrigo Ochigame, Hannah LeBlanc, Meghan Kelly, Will Deringer, Blakeley P. Hoffman, Mariel García-Montes, and the comments of anonymous reviewers. This project was supported by NSF Award 1900991, NSF Dissertation Improvement Grant 1941577, an SSRC Social Data Dissertation Fellowship, and the MIT Programs for Digital Humanities.

REFERENCES

- [1] Eytan Adar and Elsie Lee. 2020. Communicative Visualizations as a Learning Problem. *IEEE Transactions on Visualization and Computer Graphics* (2020), 1–11. <https://doi.org/10.1109/TVCG.2020.3030375>
- [2] Laura M. Ahearn. 2004. Literacy, Power, and Agency: Love Letters and Development in Nepal. *Language and Education* 18, 4 (Sept. 2004), 305–316. <https://doi.org/10.1080/09500780408666883>
- [3] Alexandra P. Alberda. 2020. COVID-19 Data Literacy is for Everyone. <https://medium.com/nightingale/covid-19-data-literacy-is-for-everyone-46120b58cec9>
- [4] Basak Alper, Nathalie Henry Riche, Fanny Chevalier, Jeremy Boy, and Metin Sezgin. 2017. Visualization Literacy at Elementary School. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 5485–5497. <https://doi.org/10.1145/3025453.3025877>
- [5] Anti-Eviction Mapping Project. 2019. *(Dis)location Black Exodus*. <http://archive.org/details/dislocationblackexodus>
- [6] Ahmer Arif, Leo Graiden Stewart, and Kate Starbird. 2018. Acting the Part: Examining Information Operations Within #BlackLivesMatter Discourse. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 1–27. <https://doi.org/10.1145/3274289>
- [7] Association for Computing Machinery. 2020. ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT). <https://facctconference.org/>
- [8] Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Mehrout, and Alexander Volfovsky. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115, 37 (Sept. 2018), 9216–9221. <https://doi.org/10.1073/pnas.1804840115> Publisher: National Academy of Sciences Section: Social Sciences.
- [9] Nancy K. Baym. 2013. Data not seen: The uses and shortcomings of social media metrics. *First Monday* 18, 10 (Oct. 2013). <https://doi.org/10.5210/fm.v18i10.4873>
- [10] Manon Berriche and Sacha Altay. 2020. Internet users engage more with phatic posts than with health misinformation on Facebook. *Palgrave Communications* 6, 1 (April 2020), 1–9. <https://doi.org/10.1057/s41599-020-0452-1> Number: 1 Publisher: Palgrave.
- [11] James S. Bielo. 2019. “Particles-to-People...Molecules-to-Man”: Creationist Poetics in Public Debates. *Journal of Linguistic Anthropology* 29, 1 (2019), 4–26. <https://doi.org/10.1111/jola.12205> eprint: <https://anthrosource.onlinelibrary.wiley.com/doi/pdf/10.1111/jola.12205>
- [12] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, 10 (Oct. 2008), P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008> Publisher: IOP Publishing.
- [13] Emily Bowe, Erin Simmons, and Shannon Mattern. 2020. Learning from lines: Critical COVID data visualizations and the quarantine quotidian. *Big Data & Society* (July 2020). <https://doi.org/10.1177/2053951720939236> Publisher: SAGE PublicationsSage UK: London, England.
- [14] Geoffrey C Bowker and Susan Leigh Star. 2000. *Sorting Things Out: Classification and Its Consequences*. MIT Press, Cambridge, MA.
- [15] 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>
- [16] danah boyd. 2018. You Think You Want Media Literacy... Do You? <https://points.datasociety.net/you-think-you-want-media-literacy-do-you-7cad6af18ec2>
- [17] David Buckingham. 2017. Fake news: is media literacy the answer? <https://davidbuckingham.net/2017/01/12/fake-news-is-media-literacy-the-answer/>
- [18] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research* 81 (2018), 1–15.
- [19] Jean Burgess and Ariadna Matamoros-Fernández. 2016. Mapping sociocultural controversies across digital media platforms: one week of #gamergate on Twitter, YouTube, and Tumblr. *Communication Research and Practice* 2, 1 (Jan. 2016), 79–96. <https://doi.org/10.1080/22041451.2016.1155338>
- [20] 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> Publisher: National Academy of Sciences Section: Colloquium Paper.
- [21] Katy Börner, Adam Maltese, Russell Nelson Balliet, and Joe Heimlich. 2016. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization* 15, 3 (July 2016), 198–213. <https://doi.org/10.1177/1473871615594652> Publisher: SAGE Publications.
- [22] André Calero Valdez, Martina Ziefle, and Michael Sedlmair. 2018. Studying Biases in Visualization Research: Framework and Methods. In *Cognitive Biases in Visualizations*, Geoffrey Ellis (Ed.). Springer International Publishing, Cham, 13–27. https://doi.org/10.1007/978-3-319-95831-6_2
- [23] Alexander Campolo. 2020. Flattening the Curve; Visualization and Pandemic Knowledge. <https://sifk.uchicago.edu/news/flattening-the-curve-visualization-and-pandemic-knowledge/>
- [24] Joe Carroll. 2020. Texas Governor Says Positive-Rate Surge Under Investigation. *Bloomberg* (Aug. 2020). <https://www.bloomberg.com/news/articles/2020-08-13/texas-governor-says-positivity-rate-surge-under-investigation>
- [25] CDC. 2020. Coronavirus Disease 2019 (COVID-19). <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/cloth-face-cover-guidance.html>
- [26] Kathy Charmaz. 2006. *Constructing grounded theory*. Sage Publications, London ; Thousand Oaks, Calif.

- [27] Tinggui Chen, Qianqian Li, Peihua Fu, Jianjun Yang, Chonghuan Xu, Guodong Cong, and Gongfa Li. 2020. Public Opinion Polarization by Individual Revenue from the Social Preference Theory. *International Journal of Environmental Research and Public Health* 17, 3 (Feb. 2020). <https://doi.org/10.3390/ijerph17030946>
- [28] Angèle Christin. 2016. From daguerreotypes to algorithms: machines, expertise, and three forms of objectivity. *ACM SIGCAS Computers and Society* 46, 1 (March 2016), 27–32. <https://doi.org/10.1145/2908216.2908220>
- [29] Derek K. Chu, Elie A. Akl, Stephanie Duda, Karla Solo, Sally Yaacoub, Holger J. Schünemann, Derek K. Chu, Elie A. Akl, Amena El-harakeh, Antonio Bognanni, Tamara Lotfi, Mark Loeb, Anisa Hajizadeh, Anna Bak, Ariel Izcovich, Carlos A. Cuello-Garcia, Chen Chen, David J. Harris, Ewa Borowiack, Fatimah Chamseddine, Finn Schünemann, Gian Paolo Morgano, Giovanna E. U. Muti Schünemann, Guang Chen, Hong Zhao, Ignacio Neumann, Jeffrey Chan, Joanne Khabsa, Loyal Hneiny, Leila Harrison, Maureen Smith, Nesrine Rizk, Paolo Giorgi Rossi, Pierre AbiHanna, Rayane El-khoury, Rosa Stalteri, Tejan Baldeh, Thomas Piggott, Yuan Zhang, Zahra Saad, Assem Khamis, Marge Reinap, Stephanie Duda, Karla Solo, Sally Yaacoub, and Holger J. Schünemann. 2020. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet* 395, 10242 (June 2020), 1973–1987. [https://doi.org/10.1016/S0140-6736\(20\)31142-9](https://doi.org/10.1016/S0140-6736(20)31142-9) Publisher: Elsevier.
- [30] E. Gabriella Coleman. 2014. *Hacker, Hoaxer, Whistleblower, Spy: The Many Faces of Anonymous*. Verso, London; New York.
- [31] Michael Correll. 2019. Ethical Dimensions of Visualization Research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, Glasgow, Scotland Uk, 1–13. <https://doi.org/10.1145/3290605.3300418>
- [32] Sasha Costanza-Chock. 2020. *Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice*. MIT Press, Cambridge, MA. <https://papers.ssrn.com/abstract=3189696>
- [33] CrowdTangle Team. 2020. CrowdTangle. <https://apps.crowdtangle.com/crystalcovidvis/lists/pages>
- [34] Data for Black Lives COVID-19 Movement Pulse Check April 2020. Technical Report.
- [35] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data Feminism*. The MIT Press, Cambridge, Massachusetts.
- [36] Documenting the Now. 2016. Twarc. <https://github.com/DocNow/twarc>
- [37] John Downey and Natalie Fenton. 2003. New Media, Counter Publicity and the Public Sphere. *New Media & Society* 5, 2 (June 2003), 185–202. <https://doi.org/10.1177/1461444803005002003> Publisher: SAGE Publications.
- [38] Joseph Dumit and Regula Valérie Burri. 2008. Social studies of scientific imaging and visualization. In *The Handbook of Science and Technology Studies*, Edward J. Hackett, Olga Amsterdamska, Judy Wajcman, and Michael Lynch (Eds.). MIT Press, Cambridge, MA, 297–317.
- [39] Omri Elisha. 2011. *Moral Ambition: Mobilization and Social Outreach in Evangelical Megachurches*. Number 12 in Anthropology of Christianity. University of California Press, Berkeley, CA. OCLC: ocn695683676.
- [40] Robert M. Emerson, Rachel I. Fretz, and Linda L. Shaw. 2011. *Writing ethnographic fieldnotes* (2nd ed ed.). The University of Chicago Press, Chicago.
- [41] Anthony Fauci. 2020. Dr. Anthony Fauci: "Science is Truth". <https://www.hhs.gov/podcasts/learning-curve/learning-curve-05-dr-anthony-fauci-science-is-truth.html> Last Modified: 2020-06-17T12:32:04:00.
- [42] Frank Fischer. 2019. Knowledge politics and post-truth in climate denial: on the social construction of alternative facts. *Critical Policy Studies* 13, 2 (April 2019), 133–152. <https://doi.org/10.1080/19460171.2019.1602067>
- [43] Kirstin Fontichiaro and Jo Angela Oehrli. 2016. Why Data Literacy Matters. *Knowledge Quest* 44, 5 (2016), 21–27. <https://eric.ed.gov/?id=EJ1099487> Publisher: American Association of School Librarians.
- [44] Jason Forrest. 2020. How John Burn-Murdoch's Influential Dataviz Helped the World Understand Coronavirus. <https://medium.com/nightingale/how-john-burn-murdochs-influential-dataviz-helped-the-world-understand-coronavirus-6cb4a09795ae>
- [45] Clifford Geertz. 1998. Deep Hanging Out. *New York Review of Books* (Oct. 1998). <https://www.nybooks.com/articles/1998/10/22/deep-hanging-out/> Section: Ideas.
- [46] Lisa Gitelman (Ed.). 2013. *"Raw Data" Is an Oxymoron*. MIT Press, Cambridge, MA.
- [47] Andrew M. Guess, Michael Lerner, Benjamin Lyons, Jacob M. Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan Sircar. 2020. A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences* 117, 27 (July 2020), 15536–15545. <https://doi.org/10.1073/pnas.1920498117> ISBN: 9781920498115 Publisher: National Academy of Sciences Section: Social Sciences.
- [48] Susan Hanson. 1992. Geography and Feminism: Worlds in Collision? *Annals of the Association of American Geographers* 82, 4 (1992), 569–586. <https://doi.org/10.1111/j.1467-8306.1992.tb01718.x> _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8306.1992.tb01718.x>.
- [49] Donna Haraway. 1988. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14, 3 (1988), 575. <https://doi.org/10.2307/3178066>
- [50] J B Harley. 1989. Deconstructing the map. *Cartographica: The International Journal for Geographic Information and Geovisualization* 26, 2 (June 1989), 1–20. <https://doi.org/10.3138/E635-7827-1757-9T53> Publisher: University of Toronto Press.
- [51] Renee Hobbs (Ed.). 2016. *Exploring the Roots of Digital and Media Literacy through Personal Narrative*. Temple University Press. <https://doi.org/10.2307/j.ctvrf898z>
- [52] Arlie Russell Hochschild. 2016. *Strangers in Their Own Land: Anger and Mourning on the American Right*. New Press, New York.
- [53] Richard Hofstadter. 1966. *Anti-Intellectualism in American Life* (1st edition ed.). Vintage, New York.
- [54] Jessica Hullman. 2020. Why Authors Don't Visualize Uncertainty. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (Jan. 2020), 130–139. <https://doi.org/10.1109/TVCG.2019.2934287>
- [55] Jessica Hullman and Nicholas Diakopoulos. 2011. Visualization Rhetoric: Framing Effects in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (Dec. 2011), 2231–2240. <https://doi.org/10.1109/TVCG.2011.255>
- [56] Jessica Hullman, Xiaoli Qiao, Michael Correll, Alex Kale, and Matthew Kay. 2019. In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (Jan. 2019), 903–913. <https://doi.org/10.1109/TVCG.2018.2864889>
- [57] Meghan Kelly. 2019. Mapping Syrian Refugee Border Crossings: A Feminist Approach. *Cartographic Perspectives* 93 (Nov. 2019), 34–64–34–64. <https://doi.org/10.14714/CP93.1406> Number: 93.
- [58] Scott F. Kiesling, Umashanthi Pavalanathan, Jim Fitzpatrick, Xiaochuang Han, and Jacob Eisenstein. 2018. Interactional Stancetaking in Online Forums. *Computational Linguistics* 44, 4 (Dec. 2018), 683–718. https://doi.org/10.1162/coli_a_00334
- [59] Rob Kitchen, Martin Dodge, and Chris Perkins. 2011. Introductory Essay: Power and Politics of Mapping. In *The Map Reader*. John Wiley & Sons, Ltd, 387–394. <https://doi.org/10.1002/9780470979587.ch50> Section: 5.1 _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470979587.ch50>.
- [60] Robert Kosara. 2016. An Empire Built On Sand: Reexamining What We Think We Know About Visualization. In *Proceedings of the Beyond Time and Errors on Novel Evaluation Methods for Visualization - BELIV '16*. ACM Press, Baltimore, MD, USA, 162–168. <https://doi.org/10.1145/2993901.2993909>
- [61] Robert Kozinets. 2019. Integrating: Combining Telescopic and Microscopic Understandings. In *Netnography: The Essential Guide to Qualitative Social Media Research* (3 ed.). SAGE Publications, London.
- [62] Mei-Po Kwan. 2002. Feminist Visualization: Re-envisioning GIS as a Method in Feminist Geographic Research. *Annals of the Association of American Geographers* 92, 4 (Dec. 2002), 645–661. <https://doi.org/10.1111/1467-8306.00309>
- [63] Jean Lave and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press, Cambridge, UK. Google-Books-ID: CAVIOrW3vYAC.
- [64] Sukwon Lee, Sung-Hee Kim, Ya-Hsin Hung, Heidi Lam, Youn-Ah Kang, and Ji Soo Yi. 2016. How do People Make Sense of Unfamiliar Visualizations?: A Grounded Model of Novice's Information Visualization Sensemaking. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan. 2016), 499–508. <https://doi.org/10.1109/TVCG.2015.2467195> Conference Name: IEEE Transactions on Visualization and Computer Graphics.
- [65] Sukwon Lee, Sung-Hee Kim, and Bum Chul Kwon. 2017. VLAT: Development of a Visualization Literacy Assessment Test. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (Jan. 2017), 551–560. <https://doi.org/10.1109/TVCG.2016.2598920>
- [66] R. Lengler. 2006. Identifying the Competencies of 'Visual Literacy' - a Prerequisite for Knowledge Visualization. In *Tenth International Conference on Information Visualisation (IV'06)*. Tenth International Conference on Information Visualisation (IV'06), 232–236. <https://doi.org/10.1109/IV.2006.60> ISSN: 2375-0138.
- [67] Zenobia Lloyd. 2020. Q&A: Michael Levitt on why there shouldn't be a lockdown, how he's been tracking coronavirus. *The Stanford Daily* (Aug. 2020). <https://www.stanforddaily.com/2020/08/02/qa-michael-levitt-on-why-there-shouldnt-be-a-lockdown-how-hes-been-tracking-coronavirus/>
- [68] Adam Maltese, Dubravka Svetina, and Joseph Harsh. 2015. Research and Teaching: Data Visualization Literacy: Investigating Data Interpretation Along the Novice-Expert Continuum. *Journal of College Science Teaching* 45, 01 (2015). https://doi.org/10.2505/4/jcst15_045_01_84
- [69] Annette N. Markham. 2013. Fieldwork in Social Media. *Qualitative Communication Research* 2, 4 (Dec. 2013), 434–446. <https://doi.org/10.1525/qcr.2013.2.4.434>
- [70] Ed Mazza. 2020. Top Trump DHS Official Asks Twitter To Help Him Find Coronavirus Outbreak Map. *Huffington Post* (Feb. 2020). https://www.huffpost.com/entry/ken-cuccinelli-coronavirus-map_n_5e54dd40c5b66729c618d35 Section: Politics.

- [71] Leland McInnes, John Healy, and James Melville. 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (Dec. 2018). <http://arxiv.org/abs/1802.03426>
- [72] Sara L. McLafferty. 1995. Counting for Women. *The Professional Geographer* 47, 4 (Nov. 1995), 436–442. https://doi.org/10.1111/j.0033-0124.1995.436_j.x
- [73] Miriah Meyer and Jason Dykes. 2019. Criteria for Rigor in Visualization Design Study. *IEEE Transactions on Visualization and Computer Graphics* (2019), 1–1. <https://doi.org/10.1109/TVCG.2019.2934539>
- [74] Bruce L. Miller. 2020. Science Denial and COVID Conspiracy Theories: Potential Neurological Mechanisms and Possible Responses. *Journal of the American Medical Association* 324, 22 (Dec. 2020), 2255. <https://doi.org/10.1001/jama.2020.21332>
- [75] David Moats and Erik Borra. 2018. Quali-quantitative methods beyond networks: Studying information diffusion on Twitter with the Modulation Sequencer. *Big Data & Society* 5, 1 (June 2018), 205395171877213. <https://doi.org/10.1177/2053951718772137>
- [76] Erik Mueggler. 2011. *The Paper Road: Archive and Experience in the Botanical Exploration of West China and Tibet*. University of California Press, Berkeley, CA.
- [77] Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism* (1 edition ed.). NYU Press, New York.
- [78] Mimi Onuoha. 2016. The Library of Missing Datasets. <http://mimionuoha.com/the-library-of-missing-datasets>
- [79] Naomi Oreskes and Erik M. Conway. 2010. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. Bloomsbury Press, New York, NY. OCLC: 461631066.
- [80] Deokgun Park, Steven M. Drucker, Roland Fernandez, and Niklas Elmqvist. 2018. Atom: A Grammar for Unit Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 24, 12 (Dec. 2018), 3032–3043. <https://doi.org/10.1109/TVCG.2017.2785807>
- [81] Paul Parsons. 2018. Promoting Representational Fluency for Cognitive Bias Mitigation in Information Visualization. In *Cognitive Biases in Visualizations*, Geoffrey Ellis (Ed.). Springer International Publishing, Cham, 137–147. https://doi.org/10.1007/978-3-319-95831-6_10
- [82] Evan M. Peck, Sofia E. Ayuso, and Omar El-Etr. 2019. Data is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300474>
- [83] Jorge Poco and Jeffrey Heer. 2017. Reverse-Engineering Visualizations: Recovering Visual Encodings from Chart Images. *Computer Graphics Forum* 36, 3 (June 2017), 353–363. <https://doi.org/10.1111/cgf.13193>
- [84] Theodore M. Porter. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press, Princeton, NJ.
- [85] John Postill and Sarah Pink. 2012. Social Media Ethnography: The Digital Researcher in a Messy Web. *Media International Australia* 145, 1 (Nov. 2012), 123–134. <https://doi.org/10.1177/1329878X1214500114> Publisher: SAGE Publications Ltd.
- [86] Robert N Proctor. 2011. *Golden Holocaust: Origins of the Cigarette Catastrophe and the Case for Abolition*. University of California Press, Berkeley, CA.
- [87] Puripant Ruchikachorn and Klaus Mueller. 2015. Learning Visualizations by Analogy: Promoting Visual Literacy through Visualization Morphing. *IEEE Transactions on Visualization and Computer Graphics* 21, 9 (Sept. 2015), 1028–1044. <https://doi.org/10.1109/TVCG.2015.2413786> Conference Name: IEEE Transactions on Visualization and Computer Graphics.
- [88] Manolis Savva, Nicholas Kong, Arti Chhajta, Li Fei-Fei, Maneesh Agrawala, and Jeffrey Heer. 2011. ReVision: automated classification, analysis and redesign of chart images. In *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*. ACM Press, Santa Barbara, California, USA, 393. <https://doi.org/10.1145/2047196.2047247>
- [89] Dietram A. Scheufele and Nicole M. Krause. 2019. Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences* 116, 16 (April 2019), 7662–7669. <https://doi.org/10.1073/pnas.1805871115> Publisher: National Academy of Sciences Section: Colloquium Paper.
- [90] E Segel and J Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (Nov. 2010), 1139–1148. <https://doi.org/10.1109/TVCG.2010.179>
- [91] Ben Shneiderman. 2020. Data Visualization's Breakthrough Moment in the COVID-19 Crisis. <https://medium.com/nightingale/data-visualizations-breakthrough-moment-in-the-covid-19-crisis-ce46627c7db5>
- [92] Mario Luis Small. 2009. 'How many cases do I need?': On science and the logic of case selection in field-based research. *Ethnography* 10, 1 (March 2009), 5–38. <https://doi.org/10.1177/1466138108099586> Publisher: SAGE Publications.
- [93] Wonyoung So and Fábio Duarte. 2020. Cartographers of North Korea: Who are they and what are the technical, political, and social issues involved in mapping North Korea. *Geoforum* 110 (March 2020), 147–156. <https://doi.org/10.1016/j.geoforum.2020.02.008>
- [94] Kate Starbird, Ahmer Arif, and Tom Wilson. 2019. Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 127:1–127:26. <https://doi.org/10.1145/3359229>
- [95] Christina Stoiber, Florian Grassinger, Margit Pohl, Holger Stitz, Marc Streit, and Wolfgang Aigner. 2019. *Visualization Onboarding: Learning How to Read and Use Visualizations*. preprint. Open Science Framework. <https://doi.org/10.31219/osf.io/c38ab>
- [96] Yuzuru Tanahashi, Nick Leaf, and Kwan-Liu Ma. 2016. A Study On Designing Effective Introductory Materials for Information Visualization. *Computer Graphics Forum* 35, 7 (2016), 117–126. <https://doi.org/10.1111/cgf.13009> <https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13009>
- [97] Emma Teng. 2004. *Taiwan's Imagined Geography: Chinese Colonial Travel Writing and Pictures, 1683-1895*. Harvard Univ Asia Center. Google-Books-ID: ufetNF9d7Y4C.
- [98] Francesca Tripodi. [n.d.]. *Searching for Alternative Facts*. Technical Report. Data & Society Research Institute. 64 pages.
- [99] Zeynep Tufekci. 2020. Opinion | On Masks and Clinical Trials, Rand Paul's Tweeting Is Just Plain Wrong. *The New York Times* (Nov. 2020). <https://www.nytimes.com/2020/11/06/opinion/sunday/coronavirus-masks.html>
- [100] Zeynep Tufekci. 2020. Opinion | Why Telling People They Don't Need Masks Backfired. *The New York Times* (March 2020). <https://www.nytimes.com/2020/03/17/opinion/coronavirus-face-masks.html>
- [101] Fernanda B. Viegas and Martin Wattenberg. 2006. Communication-minded visualization: A call to action [Technical forum]. *IBM Systems Journal* 45, 4 (2006), 801–812. <https://doi.org/10.1147/sj.454.0801> Conference Name: IBM Systems Journal.
- [102] Etienne Wenger. 1998. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press, Cambridge, UK; New York. OCLC: 318240189.
- [103] Angela Xiao Wu and Harsh Taneja. 2020. Platform enclosure of human behavior and its measurement: Using behavioral trace data against platform episteme. *New Media & Society* (July 2020), 146144482093354. <https://doi.org/10.1177/1461444820933547>
- [104] Sarita Yardi and danah boyd. 2010. Dynamic Debates: An Analysis of Group Polarization Over Time on Twitter. *Bulletin of Science, Technology & Society* 30, 5 (Oct. 2010), 316–327. <https://doi.org/10.1177/0270467610380011> Publisher: SAGE Publications Inc.
- [105] Vincent Yi-Fong Su, Yung-Feng Yen, Kuang-Yao Yang, Wei-Juin Su, Kun-Ta Chou, Yuh-Min Chen, and Diahn-Warng Perng. 2020. Masks and medical care: Two keys to Taiwan's success in preventing COVID-19 spread. *Travel Medicine and Infectious Disease* (June 2020), 101780. <https://doi.org/10.1016/j.tmaid.2020.101780>
- [106] Yixuan Zhang, Yifan Sun, Sumit Barua, Enrico Bertini, and Andrea G. Parker. 2020. *Mapping the Landscape of COVID-19 Crisis Visualizations*. Preprint. Open Science Framework. <https://doi.org/10.31219/osf.io/kd3y9>