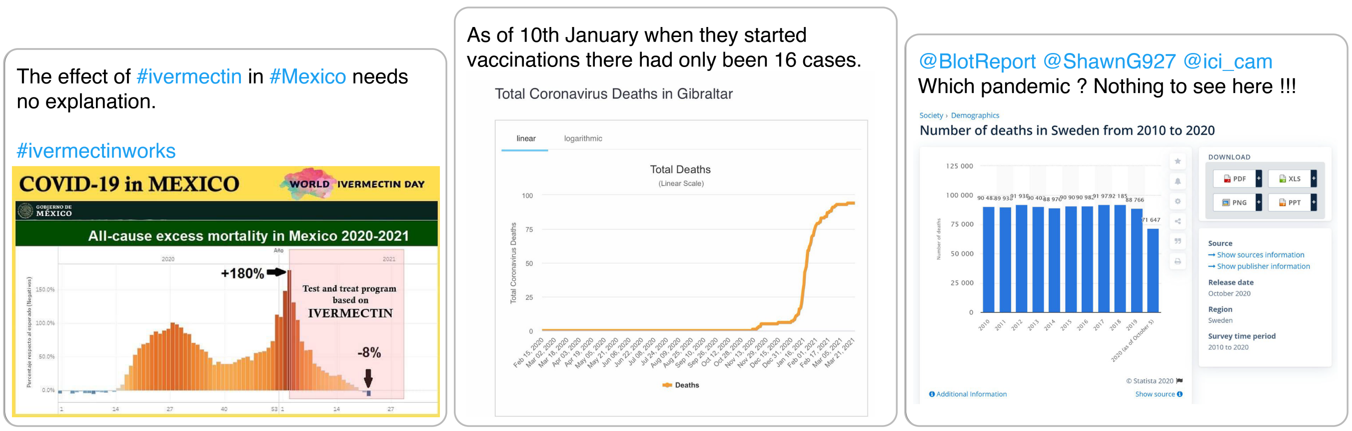
Misleading Beyond Visual Tricks: How People *Actually* Lie with Charts

Maxim Lisnic
maxim.lisnic@utah.edu
University of Utah
Salt Lake City, Utah, USA

Cole Polychronis
cole.polychronis@utah.edu
University of Utah
Salt Lake City, Utah, USA

Alexander Lex
alex@sci.utah.edu
University of Utah
Salt Lake City, Utah, USA

Marina Kogan
kogan@cs.utah.edu
University of Utah
Salt Lake City, Utah, USA



(a) Tweet 1419196852323422208 [removed] **(b) Tweet 1374425410834862084 [removed]** **(c) Tweet 1323425182044880896**

Figure 1: Examples of misleading visualizations on social media. (a) A screenshot of a COVID-19 case tracker chart originally posted by the government of Mexico (www.coronavirus.gob.mx). Additional annotations on the chart were added to attribute the fall in cases in January–March 2021 to the adoption of ivermectin, a medicine promoted as alternative treatment to COVID-19 but not shown to be effective in clinical trials [15, 57]. (b) A screenshot of a COVID-19 death chart from Worldometer, a data exploration site (www.worldometers.info/coronavirus). The poster suggests that the increase in cases is driven by vaccinations. (c) A screenshot of a chart showing the total number of deaths in Sweden between 2010 to 2020, reposted from Statista (www.statista.com). Although data from 2020 are incomplete, the poster states that the pandemic is exaggerated because the number of deaths is not sufficiently large.

ABSTRACT

Data visualizations can empower an audience to make informed decisions. At the same time, deceptive representations of data can lead to inaccurate interpretations while still providing an illusion of data-driven insights. Existing research on misleading visualizations primarily focuses on examples of charts and techniques previously reported to be deceptive. These approaches do not necessarily describe how charts mislead the general population in practice. We instead present an analysis of data visualizations found in a real-world discourse of a significant global event—Twitter posts with visualizations related to the COVID-19 pandemic. Our work shows that, contrary to conventional wisdom, violations of visualization design guidelines are not the dominant way people mislead with charts. Specifically, they do not disproportionately lead to reasoning errors in posters' arguments. Through a series of examples, we present common reasoning errors and discuss how even faithfully plotted data visualizations can be used to support misinformation.

CCS CONCEPTS

• **Human-centered computing** → **Visualization theory, concepts and paradigms**; **Empirical studies in collaborative and social computing**.

KEYWORDS

visualization, social media, misinformation, COVID-19

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM Reference Format:
Maxim Lisnic, Cole Polychronis, Alexander Lex, and Marina Kogan. 2023. Misleading Beyond Visual Tricks: How People *Actually* Lie with Charts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3544548.3580910

1 INTRODUCTION

Data visualizations created for the general public can help explain and summarize complex phenomena and otherwise incomprehensible amounts of data. On social media sites, users actively share and comment on data visualization posts as a form of collective sense-making. The insights gained from these visualizations, however, can mislead people by perpetuating misconceptions and misinformation. The visualization community has been primarily defining *misleading visualizations* as charts that interfere with the viewer's ability to accurately read off and compare values. The terms "deceptive", "misleading", "lying" are typically used to describe visualizations with visual tricks, such as truncated or inverted axes, or the violation of visualization guidelines and best practices, such as the use of unjustified 3D or problematic color maps [14, 49, 67]. This type of visual deception is rooted in the gap between the true value of data points used as input for the chart and the different values perceived by the viewer. This discrepancy has been called the "Lie Factor" [70].

The visualization community has studied these deceptive design techniques in detail [39], and "VisLies" (www.vislies.org), a long-running satellite event of the IEEE VIS, showcases some of the worst examples. Consequently, conventional wisdom may suggest that visualizations that are used to spread misinformation online would predominantly be ones that employ such visual tricks, and may imply that the solution may lie in promoting visual literacy, so that the general public can spot these deceptive techniques. More recent research, however, demonstrates that, in many cases, people who mistrust the scientific establishment propose data-driven arguments in support of their ideas and use what Lee et al. describe as *counter-visualizations*: visualizations using orthodox methods to make unorthodox arguments [34].

The question of what makes a visualization misleading or deceptive and how this can be fixed, however, remains. In an attempt to provide an answer, we collected, categorized, and analyzed 9,958 posts shared on Twitter that contain data visualizations related to the COVID-19 pandemic. Our work shows that common design violations in visualizations on social media are rare and occur at about the same rate in conspiracy- and misinformation-supporting posts as in neutral posts. We find that instead of using visualization design violations, actors who want to misinform draw attention to unexplained salient features of well-designed charts and assign meaning to them. We introduce the notion of *vulnerable visualizations*: visualizations created from accurate data with no clear intention to misinform but susceptible to supporting misinformation by not visualizing important context or not anticipating a biased reading. Through biased framing and annotations, vulnerable visualizations designed by authoritative sources can be misused to create a type of counter-visualizations described by Lee et al. [34].

Our paper makes several contributions. Firstly, we analyze the prevalence of visualization misinformation techniques in a large-scale real-world data set. To our knowledge, this is the first work to describe misleading visualizations among a sample of charts on a given topic that a general audience may see online, as opposed to a sample of examples previously identified as deceptive. Secondly, we introduce a typology of attributes of visual misinformation on social media. We illustrate the typology by presenting examples of specific instances of visualization-supported misinformation. Thirdly, we provide a theoretical framework for understanding how data visualizations can be used to reinforce misinformation arguments through the lens of inductive reasoning. Lastly, based on our findings, we propose design recommendations to safeguard charts and prevent their misinterpretation.

2 RELATED WORK

Our work is most closely related to research on online misinformation and data visualization. In this section, we highlight the importance of studying visualization as a potential vehicle of misinformation and the merits of using social media platforms as a setting for studying misleading visualizations.

2.1 Online Misinformation

Social media platforms allow people to quickly communicate to broad audiences, facilitating near-instant dissemination of important information. At the same time, rapid communication may offer opportunities to—intentionally or not—spread inaccurate, misleading, or even harmful information. Conventional wisdom suggests that, given enough users, a "wisdom of crowds" [65] would prevail and the spread of misinformation online would be stopped and corrected by the community. Research on online misinformation demonstrates that the crowds do detect and question false rumors [1, 43]. Recent work, however, has shown the limits of the online community to effectively self-correct, which has been successfully exploited by actors who intend to spread misinformation [61, 62].

Misinformation can spread broadly with the help of strategic information operations, but the initial narratives often arise naturally from the community as people try to make sense of an ongoing event [61]. The process of collective sense-making online is especially common during events that can be characterized by high uncertainty and limited accurate information, such as man-made [1, 20, 48] or natural disasters [63, 66, 71], or public health emergencies [16, 53].

In the context of public health-related crises, such as viral disease outbreaks, research confirms that people turn to social media platforms to discuss potential personal health risks with others [16]. This behavior has been especially prevalent during the COVID-19 pandemic: work by Pine et al. shows that people find it difficult to accurately assess risks independently based on existing data due to its inconsistencies, incompleteness, as well as scale [53]. These concerns encourage people to also turn to anecdotal evidence for risk assessment, which could be prone to personal biases. In our work, we examine people's attempts at reasoning with COVID-19 data in social media posts and highlight ways in which data inconsistencies and personal biases, among other reasons, can lead to perpetuation of misinformation.

During the COVID-19 crisis, both experts and novices share relevant data primarily in the form of data visualizations [77]. Interpretations of raw data and data visualizations are known to be dependent on context [28] and personal biases [51]. This fact underscores the importance of safeguarding against online audiences using data visualizations to make inaccurate conclusions, and recent research confirms that this is a pressing concern. Lee et al. [34] show that different users draw different conclusions from the same COVID-19 visualizations, and work by Zhang et al. [76] documents that many COVID-19 dashboard and visualization designers express concerns that their work could be misinterpreted or misused. A gap in knowledge, however, still exists about how how data visualizations on social media can contribute to the spread of misinformation. To address this gap, our work describes the attributes of data visualization posts that are deceptive and perpetuate misleading claims.

2.2 Misleading Visualizations

The first influential works that discuss deceptive charts in general context, such as Darrell Huff's 1954 book *How to Lie with Statistics* [22] and Edward Tufte's 1983 book *The Visual Display of Quantitative Information* [70], set the tone for the research and commentary on misleading visualizations for years to come [4, 9, 32, 33, 49, 67]. Tufte [70] introduces the notions of *graphical integrity* and *lie factor*, underscoring the importance of the size of the visual encoding matching the magnitude of the underlying value. Pandey et al. [49] confirm that common types of distortions that violate graphical integrity—truncated, inverted, or re-scaled axes—affect the viewers' perception of data, although more recent research shows that the magnitude of this effect is limited and context-dependent, especially among audiences with higher data literacy [12].

These types of visual tricks represent the primary way in which both the visualization research community and the general public have been thinking about misleading and deceptive charts. The "VisLies" (www.vislies.org) event has been an annual feature of the IEEE VIS conference since 2015 where researchers gather and discuss deceptive visualizations, primarily focusing on egregious data or plotting errors and distortion techniques. The general public also associates the term "misleading" with visual tricks and errors, as shown by recent work by Lo et al. [39] that examines 1,143 charts tagged as misleading, collected from search engines and social media sites. The results show that the vast majority of visualizations labeled as misleading exhibit errors in input data and choices of visual encoding, or employ distortion techniques. At the same time, only 7% of the charts collected by Lo et al. are "faithfully plotted" but have a misleading message, for example suggesting a correlation or omitting important contextual information.

This latter type of faithfully plotted visual misinformation was common in charts shared in online anti-mask groups at the onset of the COVID-19 pandemic [34]. Lee et al. [34] discuss that this finding indicates that visual misinformation is largely driven by an epistemological gap between communities, rather than a gap in visual literacy. Other research also adds support to the idea that interpretations of visualizations are largely context-dependent, whether the context is caused by personal beliefs and biases [51], introduced through biased framing and titles [28], or previous knowledge [74].

Yet, most existing research aimed at seeking solutions for misleading visualizations is still focused on identifying violations of design guidelines and blatant errors using automatic annotation and linting [14, 21, 41], although every work also acknowledges the importance of studying visualization errors that stem from biased reading. Several frameworks for thinking about data visualizations through the lens of cognitive biases have been proposed [5, 11]. However, a large-scale literature review by Dimara et al. [11] shows that, overall, research on cognitive biases in visualization has been limited, with existing work primarily focused on biases that affect the task of estimating values from a chart [11]. More recently, the visualization community has been advocating for thinking more broadly about goals of visualization as opposed to solely optimizing for perception precision [2, 8, 27]. Many potentially bias-prone visualization tasks identified in Dimara et al., however,—such as hypothesis assessment, causal attribution, and opinion reporting—have yet to be formally examined for biases [11].

An opportunity to reduce this research gap is presented by analyzing data from social media sites, where members of the general public commonly share their attempts at forming hypotheses and report their opinions. Research shows that social information has an impact on visualization perception [24], which underscores the importance of examining the role of visualizations in social media. To our knowledge, our work is the first large-scale analysis of visualization-supported reasoning on social media sites. Our research supports the finding by Lee et al. [34] that charts that are used to corroborate misinformation arguments largely conform to common visualization design guidelines, and goes further to present a typology of specific attributes that make such charts misleading and offer design recommendations.

3 METHODS

In order to explore the ways in which individuals might deceive others using visualizations, we collected and processed social media posts related to the COVID-19 pandemic from Twitter, as illustrated in Figure 2 and described in the sections below.

We consider this data set to be important and appropriate for our analysis for several reasons. Firstly, it is a large-scale data set of social media posts that is collected in an unbiased way with respect to the deceptiveness of visualizations and reflects the online visualization discourse as is. Secondly, the topic of the posts is an important prolonged crisis that has drawn visualizations from novices and experts alike: the COVID-19 pandemic.

3.1 Data Collection

We collected data generated during the pandemic on Twitter using the official Twitter COVID-19 streaming endpoint. This endpoint provides the full collection of tweets containing at least one of the 585 hashtags and keywords[1] that Twitter internally identified as being relevant to conversations on their platform related to the pandemic. We started streaming data from this endpoint on May 15, 2020 and stopped streaming on September 6, 2021. In that time, we collected 2.2 billion tweets related to the pandemic, totaling 22 terabytes of data (see Figure 2a). We then filtered this full stream

[1] Full list of filter terms available on the API documentation.

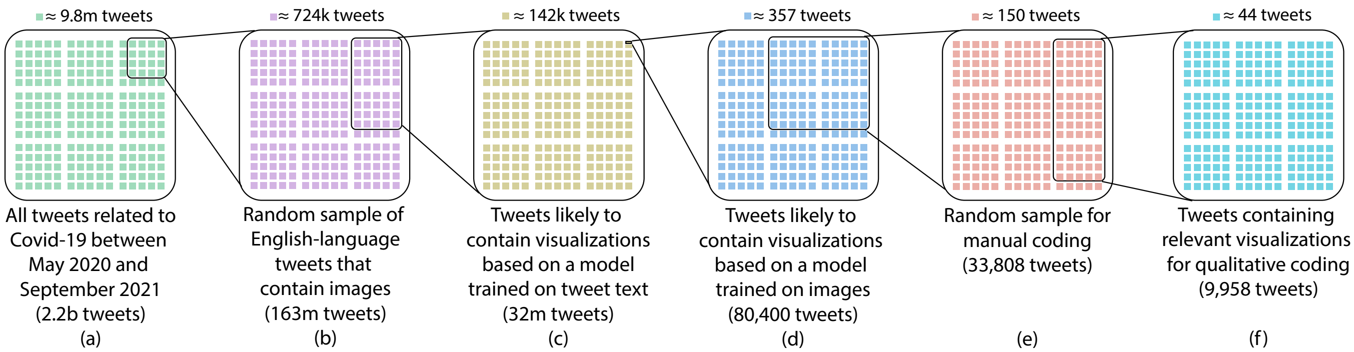


Figure 2: Illustration of the data collection and processing pipeline. (a) Initial set of all tweets related to COVID-19 collected between May 15, 2020 and September 6, 2021. (b) Random sample of original English language tweets that contained images. Original tweets indicate posts that are not retweets of other tweets of replies to existing tweets. (c) Tweets identified to contain visualizations based on analyzing tweet text using a Multinomial Naive Bayes classifier trained on tweet text. (d) Tweets further identified to contain visualizations based on a fine-tuned image recognition model. (e) Random sample taken from model outputs to reduce the size for manual qualitative coding. (f) Tweets identified to contain relevant data visualizations after removing images misclassified as visualizations and visualizations not pertaining to the COVID-19 pandemic through manual inspection. This set is used for qualitative coding and analysis.

down to only English language tweets that contained images (although not explicitly visualizations) and removed any retweets or replies. From this set of tweets that potentially contain visualizations and meet the rest of our collection and scoping criteria, we extracted two data sets. First, we extracted the full first week of data May 15–22, 2020 (19,214 tweets), which we call our train-test data set. This data set was used to train machine learning models to be used for data processing (described in the following subsection). The first two authors manually labeled whether or not tweets in the train-test data set contained visualizations, resulting in a data set that had 8,170 positive examples (tweets that did share visualizations) and 11,044 negative examples (tweets that did not share visualizations). Second, we randomly sampled from the remaining data to create a data set of 162.8 million tweets (Figure 2b). It is this data set that we applied our data processing models to in order to arrive at our final data set, used to answer our research questions.

3.2 Data Processing

Retrieving the full set of over 162 million images from Twitter's photo storage system Blobstore [13] would take weeks of clock time. In order to scalably detect visualizations from our random sample of the data (Figure 2b), we applied two machine learning models, described in greater detail in sections 3.2.1 and 3.2.2. First, we predicted if a tweet is likely to contain a visualization based on the content of its message. If this model predicted that the tweet has likely shared a visualization, we then retrieved that image from the Blobstore and applied a visualization detection model.

3.2.1 Predicting the Presence of a Visualization. In order to predict whether a tweet is likely to share a visualization, we trained a Multinomial Naive-Bayes classifier (scikit-learn v0.24.1) on the train-test data set. Specifically, we embedded the text of tweets using a bag-of-words approach and supplemented it with metadata features: the number of hashtags, emojis, user mentions, and capitalized words. This model had an accuracy of 0.75 on the train-test data (precision = 0.70, recall = 0.71). To better understand this accuracy, we also labeled all of the visualizations in our train-test data set to explore whether this model was more likely to miss certain kinds of visualizations. We found that common visualization types (point and bar charts, tables, maps, etc.) are missed about 4.29% of the time. Rarer visualization types, such as isotype charts or network diagrams, are missed about 14.30% of the time. This suggests that our model performs better at detecting simple and more common charts, but this bias is negligible given our large sample size.

Applying this model as a filtering step reduced the number of tweets for which we had to retrieve images by about 80%, producing the data set shown in Figure 2c. We retrieved the image files between September 6, 2021 and September 8, 2021. Consequently, although some tweets may have since been removed or their authors may have been suspended, all posts in our analysis were public as of September 2021 and spent sufficient time on the platform to potentially be widely circulated.

3.2.2 Classifying Images as Visualizations. In order to determine whether or not images were visualizations, we used the Inception V3 Model, developed at Google [68] and pre-trained on ImageNet [10]. We used a transfer learning approach and fine-tuned this model with our train-test data set, resulting in an accuracy of 0.73 (precision = 0.71, recall = 0.65). Applying this model to the data set shown in Figure 2c and filtering any tweets that were not identified as sharing a visualization left us with a data set of 80,400 tweets that shared some kind of visualization (Figure 2d). We then qualitatively coded a subset of these as described in the section below.

3.3 Qualitative Coding

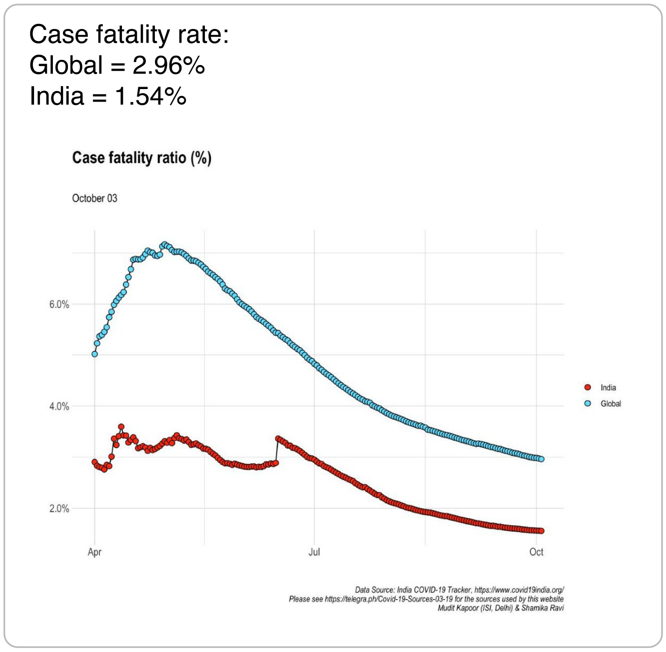
To arrive at our final corpus of tweets with visualizations, we first took a random sample (33,808 tweets) to reduce the amount of data to be manually coded, and then conducted multiple stages of qualitative coding. In the first stage, we manually examined the outputs of automated data processing to eliminate any irrelevant posts

Category	Code	Count	Percent
(a) Source of visualization	Unknown or created by author	4,453	44.7%
	Screenshot of a static chart from reputable source	2,646	26.6%
	Screenshot of an interactive dashboard from reputable source	2,860	28.7%
(b) Text polarity	Tweet text: neutral	8,419	84.5%
	Tweet text: support measures	517	5.2%
	Tweet text: oppose measures	954	9.6%
	Tweet text: promote alternative measures	68	0.7%
	Annotation native to chart: neutral	266	2.7%
	Annotation native to chart: support measures	45	0.5%
	Annotation native to chart: oppose measures	58	0.6%
	Annotation native to chart: promote alternative measures	17	0.2%
	Annotation added to chart: neutral	164	1.6%
	Annotation added to chart: support measures	44	0.4%
	Annotation added to chart: oppose measures	121	1.2%
	Annotation added to chart: promote alternative measures	16	0.2%
	Title on chart: neutral	8,816	88.5%
	Title on chart: support measures	120	1.2%
	Title on chart: oppose measures	29	0.3%
	Title on chart: promote alternative measures	4	0.0%
(c) Visualization design violation	Truncated axis	116	1.2%
	Dual axis	541	5.4%
	Value as area/volume	494	5.0%
	Inverted axis	57	0.6%
	Uneven binning	10	0.1%
	Unclear encoding	40	0.4%
	Inappropriate encoding	18	0.2%
(d) Reasoning errors	Cherry-picking data	514	5.2%
	Cherry-picking timeframe	69	0.7%
	Setting an arbitrary threshold	453	4.5%
	Causal inference	691	6.9%
	Suggesting data validity issues	80	0.8%
	Failure to account for data validity issues	65	0.7%
	Failure to account for statistical nuance	105	1.1%
	Misrepresentation of scientific studies	26	0.3%
	Incorrect reading of chart	10	0.1%
Total		**9,958**	**100.0%**

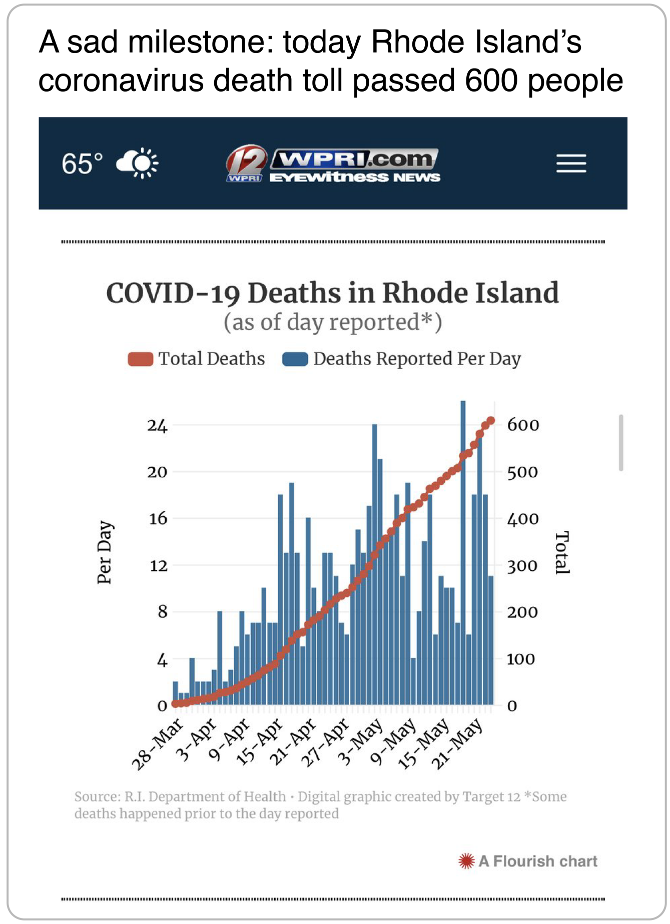
Figure 3: The codebook used to manually annotate the 9,958 relevant visualization tweets and the frequencies of codes. The codebook includes information about the (a) source of the visualization, (b) presence and polarity of textual components of the post and the chart, (c) presence of any visualization design violations, as well as (d) commonly occurring reasoning errors. Each post is described by one code from (a), and any number of codes from (b), (c), and (d).

for future annotation. Two of the authors, with the help of a (paid) undergraduate student, reviewed the tweets to remove (1) posts that were incorrectly labeled as visualizations, which primarily consisted of images of text and diagrams, and (2) posts that contained visualizations that do not pertain to the COVID-19 pandemic, for instance charts showing stock performance without mentioning the pandemic. We held multiple discussions during this process to decide how to deal with any edge cases. This process yielded 9,958 posts that contained relevant visualizations for the next stage.

In the second stage, we developed a codebook and used it to annotate the 9,958 posts, as well as created a closely related typology derived from the codebook. First, the first author coded a sample of 400 visualization tweets through an open-coding process [64], and developed the initial set of codes. The codebook includes the following categories of codes: (a) source information, or whether the visualization was a screenshot of a chart from the government or a media outlet or an interactive dashboard, or was created by an unknown author, (b) the presence and opinion polarity of the tweet text, the chart title, native annotations (annotations added to chart by the visualization creator), and added annotations (annotations added to a screenshot of a chart by a third party), (c) any existing violations of common visualization guidelines, as well as whether the choice of visual encoding is appropriate for a given type of data, and whether it is generally possible to understand the chart given the available information, and (d) commonly occurring types of reasoning errors. In order to refine the codes, particularly those pertaining to the reasoning error category, two authors applied the codebook to a sample of 200 posts independently and held iterative discussions. At the end of this process, we achieved thematic saturation and no additional issues were identified. The finalized codebook, shown in Figure 3, was used by the first author to annotate the entire sample of visualizations. Then, we grouped related codes based on the common features to create a typology of visualization post attributes. We include the results of the annotation stage in the supplemental



(a) Tweet 1313069181873528834



(b) Tweet 1264592009219715072

Figure 4: Post with visualizations containing violations of common visualization design guidelines that are not identified as misleading. Importantly, we cannot state that they are unambiguously *not* misleading, as previous research has shown the deceptive potential of design violations [49]. (a) A chart with a truncated axis that a viewer could misinterpret to show an almost 0% case fatality rate in India but accurately described by the author. (b) A chart with a dual axis that a viewer could misinterpret to show that daily deaths are on the order of hundreds but accurately interpreted by the author.

materials. In accordance with Twitter API Terms of Service[2], we provide tweet IDs that can be rehydrated to retrieve the contents of the post, as long as the post is still publicly accessible on Twitter.

3.4 Identifying Misleading Visualizations

In an attempt to present an unbiased analysis rather than one that reflects the authors' personal opinions about the deceptiveness of specific charts, we used the tweet text to identify deceptive charts by applying the following definition of misleading visualizations: visualizations that the public uses as the basis of opinionated conclusions that are entirely incorrect or contain significant reasoning errors. There are several reasons to this definition. Firstly, operationalizing deceptiveness in this way allows to provide actionable steps to combat deceptiveness by identifying specific reasoning errors. Secondly, we found that the decision of whether a certain visualization is misleading or not is *highly personal*, as it represents a judgement of whether a chart could theoretically deceive someone.

To test this hypothesis empirically, four annotators—two authors and two students affiliated with our academic institution—independently coded a sample of 400 posts, labeling them on a 5-point Likert scale [37] ranging from "not at all deceptive" to "extremely deceptive". To evaluate the inter-rater reliability, we calculated Krippendorff's α score [30] which had a value of 0.243. Converting our 5-point Likert scale results to a binary value of "not at all deceptive" and "deceptive" yielded an α of 0.351. Krippendorff recommends discarding results with $\alpha < 0.667$ [30], which indicates that our results had a very low degree of agreement. We held follow-up discussions and reviewed examples of disagreement and found that different annotators focus on different features of the

[2]https://developer.twitter.com/en/developer-terms/more-on-restricted-use-cases

chart, are biased by personal beliefs surrounding the pandemic, and have disagreements about the significance of common visualization design guidelines. The results of our experiment show that it is difficult to predict whether a visualization can mislead its audience. Instead, we used social media traces—such as tweet text—as evidence that a chart has been misinterpreted by the viewer.

Specifically, by applying the *text polarity* codes (shown in Figure 3), we can distinguish between COVID-19 data visualization posts on Twitter that are neutral—such as status updates and factual observations—and posts in which the author uses the visualization as the primary basis of support for a certain argument. For the purposes of this work, we refer to the latter group as *opinion tweets*. Opinion tweets often offer biased and incomplete interpretations of attached data visualizations. Therefore, we consider opinion tweets that contain reasoning issues in the way the author interprets the visualization (identified using the *reasoning error* codes shown in Figure 3) to be *misleading tweets*.

Although this approach allows us to identify visualizations that either mislead the tweet author or were used by the author in an attempt to mislead others, we cannot state that rest of the visualizations in our data set are unambiguously and universally *not* misleading. In particular, many visualizations that violate design guidelines still have the potential to deceive their viewers, as shown by prior research [49], even if we do not observe tweet text that misinterprets it. Figure 4 illustrates several examples of tweets that share *potentially misleading visualizations*.

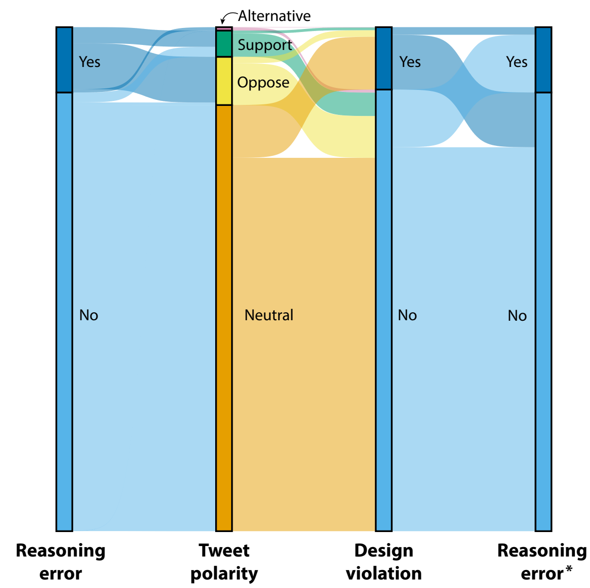


Figure 5: A flow diagram showing the relationship between tweet text polarity, presence of visualization design violations, and presence of reasoning errors. We observe similar proportions of tweets with and without design violations across text polarities. We also observe similar marginal frequencies of design violations and reasoning errors. The reasoning error column is shown twice, indicated by the asterisk.

3.5 Ethical considerations in social media research

It is important to acknowledge the ethical implications of collecting, analyzing, and sharing social media posts in our work. Tweet authors—even those with public accounts—may not be aware of their data being collected and highlighted for research purposes. At the same time, we want to offer the reader transparency into our research methods by showing compelling examples of types of visualization posts and by providing the opportunity to further explore the discussion around them. As a solution, we present anonymized examples of posts appearing throughout the paper but also offer the ability to follow the link and explore the tweet as it appears on the platform in its original context.

We consider it valuable also to include tweets that have since been deleted in our analysis. Many such tweets had been circulating on the platform for months before being removed, highlighting the importance of studying potential platform interventions and visualization design considerations to prevent the spread of misinformation. In an attempt to preserve the authors' right to remove their content, however, we only offer tweet IDs in our supplemental materials instead of the complete tweet data. Therefore, tweets deleted by the time of publication or removed later will not be available for future data collection.

4 FINDINGS

Our annotated data enables us to offer descriptive statistics of the data set as well as a typology of attributes of tweets with misleading visualizations. In this section, we present our findings that are derived from analyzing the statistics, patterns, and specific examples of tweets after applying the codebook shown in Figure 3. An interactive browser of our data set and descriptive statistics is also available at https://hashtag-misleading.netlify.app/.

4.1 Quantitative Overview

In this section we present a quantitative overview of the data set in general and misleading tweets in particular, based on our qualitative coding. Figure 3 presents a breakdown of all codes in the data set.

4.1.1 Source of Visualization. Data visualizations in tweets come from a variety of sources, with the majority (55.3%) being screenshots of existing charts from media outlets, government websites and presentations, and data exploration websites. Data exploration websites—such as OurWorldInData [58] and Worldometer—and interactive dashboards that accompany COVID-related search results on Google offer interactive visualization dashboards that allow users to selectively plot variables and data points. The remaining 45% of tweets include data visualizations for which we were unable to unambiguously identify sources.

4.1.2 Text Polarity. In terms of the polarity of observations, 62% of all opinion tweets oppose conventional measures such as masks, lockdowns, and vaccination, or deny the existence or severity of the crisis. 34% of opinion tweets support measures and government intervention. The remaining 4% may not actively oppose measures

Attribute	Opinion tweets						Neutral tweets	
	Support measures		Oppose measures		Alternative measures			
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Reasoning errors:								
Cherry-picking	144	27.9%	380	39.8%	51	75.0%	*Not applicable*	
Setting an arbitrary threshold	8	1.5%	445	46.6%	0	0.0%		
Causal inference	218	42.2%	415	43.5%	58	85.3%		
Issues with data validity	5	1.0%	136	14.3%	4	5.9%		
Failure to account for statistical nuance	76	14.7%	28	2.9%	1	1.5%		
Misrepresentation of scientific studies	6	1.2%	16	1.7%	8	11.8%		
Incorrect reading of chart	8	1.5%	2	0.2%	0	0.0%		
Any reasoning error	321	62.1%	904	94.8%	66	97.1%		
Construction attributes:								
Use of annotations on chart	87	16.8%	193	20.2%	34	50.0%	413	4.9%
Reframing screenshots of existing charts:								
Screenshot of static chart	192	37.1%	125	13.1%	9	13.2%	2,320	27.6%
Screenshot of interactive dashboard	141	27.3%	399	41.8%	20	29.4%	2,300	27.3%
Any screenshot	333	64.4%	523	54.8%	29	42.6%	4,620	54.9%
Violations of visualization design guidelines:								
Truncated axis	18	3.5%	19	2.0%	1	1.5%	78	0.9%
Dual axis	19	3.7%	49	5.1%	3	4.4%	470	5.6%
Value as area/volume	13	2.5%	33	3.5%	0	0.0%	448	5.3%
Inverted axis	2	0.4%	12	1.3%	2	2.9%	41	0.5%
Uneven binning	5	1.0%	2	0.2%	0	0.0%	3	0.0%
Unclear encoding	8	1.5%	14	1.5%	1	1.5%	17	0.2%
Inappropriate encoding	1	0.2%	11	1.2%	0	0.0%	6	0.1%
Any design violation	54	10.4%	129	13.5%	7	10.3%	1,043	12.4%
Total	**517**	**100.0%**	**954**	**100.0%**	**68**	**100.0%**	**8,419**	**100.0%**

Figure 6: The results of qualitative coding by polarity of the tweets. Neutral tweets are posts that do not offer explicit or implicit interpretations of data, and therefore reasoning error codes do not apply. Percent columns do not sum up to 100 as a tweet can have none, one, or multiple attributes. The prevalence of design violations is consistent across opinion tweets and neutral tweets. Moreover, potentially misleading design violations in opinion tweets are much less common than reasoning errors.

but instead support alternative measures and medications, primarily hydroxychloroquine and ivermectin.

4.1.3 Visualization Design Violations. In our annotated data, we find that only 12% of all posts contain data visualizations that violate common visualization design guidelines. The prevalence of such posts across tweets that contain reasoning errors and those that do not is similar with 11% and 13%, respectively. Overall, the most common design guidelines violations among tweets with reasoning errors include the use of dual axes (5.4% of all posts), encoding of quantities as area or 3D shapes (5.0%), truncated axes (1.2%), inverted axes (0.6%), with other miscellaneous violations covering another 0.5% of posts.

These results can be contrasted with Lo et al.'s survey of charts tagged as "misleading", found through search engines and social media platforms [39]. Based on the authors' supplemental materials, 57% of charts have "visualisation design" or "perception" issues that roughly translate to our definition of visualization design violations. Among those, 18% of charts have a truncated axis or otherwise inappropriate axis range, 12% use area or 3D encoding, 7% have dual axes, and 2% have inverted axes. Not surprisingly, we see, overall, proportionally fewer visualization design violations, since we sampled visualizations from an online discourse not limited to misleading charts. As seen in Figure 5, however, the prevalence of visualization design violations does not vary based on whether the post is neutral or opinionated, whether there are any reasoning errors in the interpretation, and what side of the argument the post supports. Using a Pearson's chi-squared test, we also find no statistically significant relationship between tweet polarity and presence of design violations ($\chi^2(3) = 3.2046$, $p = 0.3611$), or between presence of reasoning errors and presence of design violations ($\chi^2(1) = 1.5722$, $p = 0.4494$). Moreover, using a McNemar's chi-squared test we find no statistically significant difference in the marginal frequencies of design violations and reasoning errors across all tweets ($\chi^2(1) = 1.514$, $p = 0.2185$). This finding suggests that—relative to the prevalence of their use in support of biased and misinformation conclusions—the issue of design violations is overrepresented in research and in discussions of deceptive visualizations.

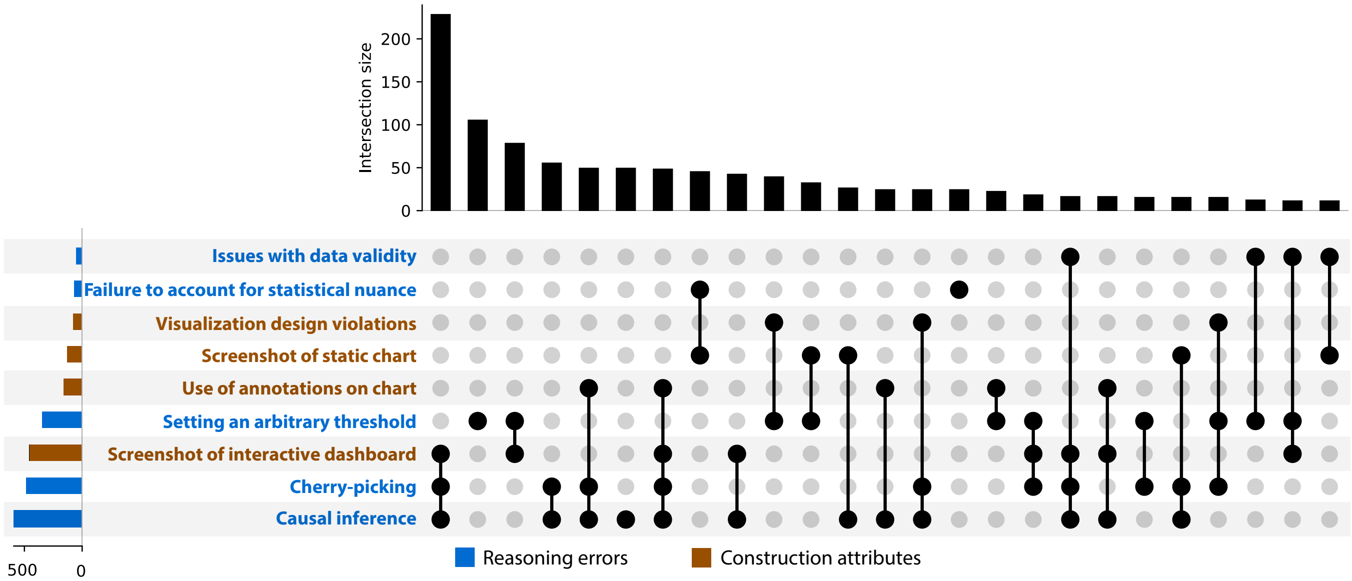


Figure 7: The prevalence and overlap of visualization tweet attributes among tweets with reasoning errors shown as an UpSet plot [36]. The horizontal bars show the total prevalence of a given attribute. The vertical bars reflect the number of tweets that have the exact combination of attributes highlighted by black dots directly below. We show the 25 most common overlapping attribute sets that describe 84% of the data. The remaining 16% are described by combinations that are more rare. For instance, reasoning errors such as Misrepresentation of Scientific Studies and Incorrect Reading of Chart are not part of the 25 most common attribute sets and are omitted.

4.2 Typology of Misleading Visualization Attributes

In this section, we describe the tweets we can identify as misleading by offering a typology of visualization post features. The typology is derived by grouping related codes. We described each post in terms of two types of attributes: attributes of its argument's reasoning—such as cherry-picking or improper causal inferences—and attributes of the post's construction—like the use of annotations on the chart. The attributes are not mutually exclusive and each post can contain none, one, or many attributes. In fact, as discussed in more detail below, certain combinations of attributes can potentially interact with one another. Figure 6 provides a summary of the prevalence of individual attributes by polarity, and Figure 7 illustrates frequencies and common combinations of attributes in an UpSet plot [36]. As we introduce the various attributes of the typology throughout this section, we return to the figures to discuss relevant observations.

4.2.1 ***Reasoning Errors.*** Reasoning errors (RE) contain commonly occurring logical attributes that can form the basis of a misleading argument.

RE 1: Cherry-picking. Visualization posts are characterized by cherry-picking when the main conclusion is consistent with the incomplete evidence presented but likely would not be generalizable with more representative evidence.

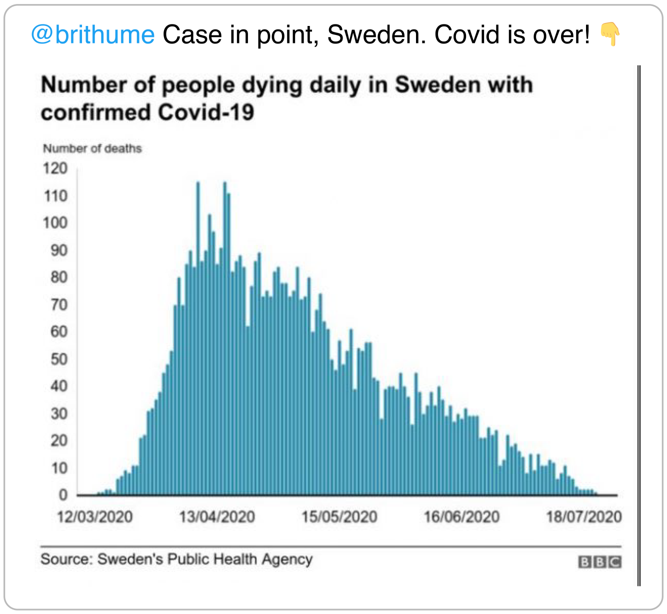
We can distinguish between two types of cherry-picking: cherry-picking of the data points and of the time frame. Figure 8a shows an example of data point cherry-picking, in which the author argues against the implementation of COVID-19 measures by providing a single data point as evidence. Sweden—a country that did not enforce strict lockdowns and reported fewer deaths and cases than some countries that did—is very commonly used as cherry-picked evidence in COVID-skeptic posts.

The post in Figure 8b is an example of time frame cherry picking. The post compares the case curves in the US in the Summers of 2020 and 2021, before and after vaccine availability. The limited time frame allows the author to omit the sharp fall in cases during the initial vaccine distribution in the first half of 2021.

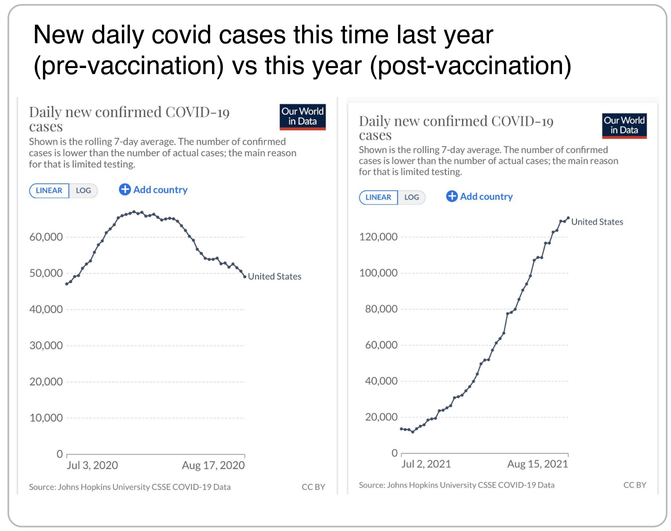
Presenting one data point in support of an argument is not universally misleading. There is a balance between cherry-picking evidence for an argument and providing an illustrative example that condenses large-scale data into a zoomed-in and easy-to-understand chart. However, although an illustrative example is useful to *explain* a phenomenon, the post should offer a way to get additional evidence to prove its *validity*.

RE 2: Setting an Arbitrary Threshold. A common attribute of misleading tweets in our data set is the author setting an arbitrary threshold against which a certain phenomenon is judged. The threshold can be stated explicitly as a number, or be visual as an annotation in a chart.

In the context of public health crises, such as disease outbreaks like the COVID-19 pandemic, the lack of an official threshold provides opportunities for people to define their own. According to *A*



(a) Tweet 1289587370082234370 [removed]



(b) Tweet 1427307226264489991

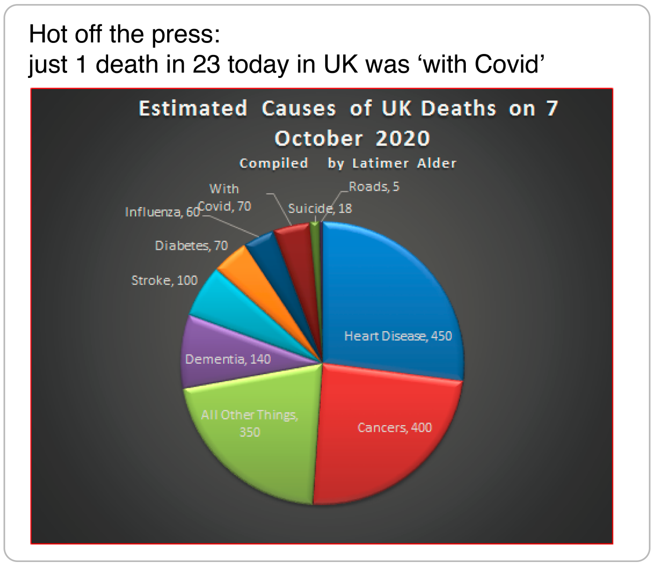
Figure 8: Examples of posts that employ *cherry-picking*. (a) An author shares a screenshot from the BBC showing a COVID-19 death curve from one country that had fewer government-mandated restrictions relative to their neighbors (Sweden) approaching zero and arguing that COVID-19 restrictions are unnecessary altogether. In this post, the argument is more effective because of the omission of comparable neighbor countries that experienced fewer COVID-19 deaths, or other countries with limited restrictions that experienced more deaths. (b) A user shares COVID case curves for the United States for the periods of early July through mid August for the years 2020 (before the vaccination campaign) and 2021 (during the vaccination campaign). Because the number of cases in August 2021 is higher than in August 2020, the user suggests that the vaccination campaign failed. This example carefully selects the time frame that most effectively supports the argument, omitting a large drop in cases in Spring 2021.

Dictionary of Epidemiology [31], a pandemic is "an epidemic occurring worldwide, or over a very wide area, crossing international boundaries and usually affecting a large number of people", whereas an epidemic is "an illness... ...clearly in excess of normal expectancy." The terms "large number" and "excess of normal" in these definitions imply that the level of seriousness of a disease outbreak is to be determined holistically and leaves room for disagreement.

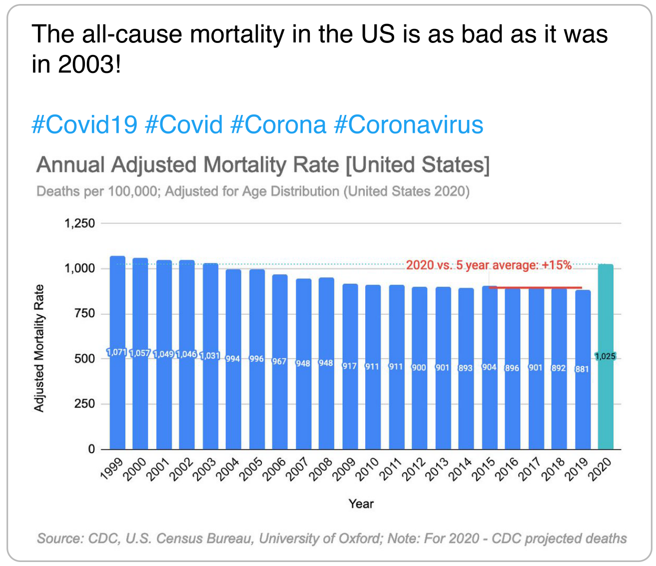
As seen in Figure 6, this type of reasoning error is especially common in COVID-skeptic posts, as they attempt to redefine the threshold of seriousness of the pandemic. For instance, the post in Figure 9a implies that COVID-19 is not a serious concern because "only" 1 in 23 deaths were caused by it. This post also shows data for a single day, in an example of time frame cherry-picking. The tweet from Figure 9b is part of a conversation discussing that the ongoing pandemic does not warrant measures because the mortality rate in 2020 is "only" 15% higher than the previous five years and comparable to 2003. In both examples, the author takes advantage of the fact that in the context of personal health, people's level of risk aversion and cost-benefit calculations are highly personal. Additionally, the attached data visualizations are used to visually exaggerate the effect. One of the pie chart's largest sections is "All other causes", which makes the COVID-19 deaths appear relatively small. The bar chart includes data as far back as 1999, when the all-cause mortality was higher than in 2020.

RE 3: Causal Inference. Tweet authors often assign cause-and-effect relationships in an attempt to explain certain salient features of a chart and evaluate them. Causal relationships are typically evaluated either by themselves against an author-defined satisfactory threshold as discussed previously, or against another inferred causal relationship. Causality inferred from a visualization can be especially misleading in cases when the data are cherry-picked. This approach is used frequently: as seen from Figure 7, causal inference and cherry-picking are among the most commonly co-occurring attributes of a post.

The author of Figure 10a evaluates the effectiveness of vaccines by highlighting that the vaccination start date in Uruguay preceded a large spike in cases, implying a causal relationship. Although causal inference is common in posts with all types of arguments, Figure 6 shows that this reasoning error is especially common in tweets that promote alternative measures and attempt to prove their effectiveness. For instance, the author of Figure 10b provides two examples of cause-and-effect relationships: one of vaccines in Israel and one of ivermectin in Zimbabwe. Since the Zimbabwe case curve is lower, the author argues that ivermectin is more effective.

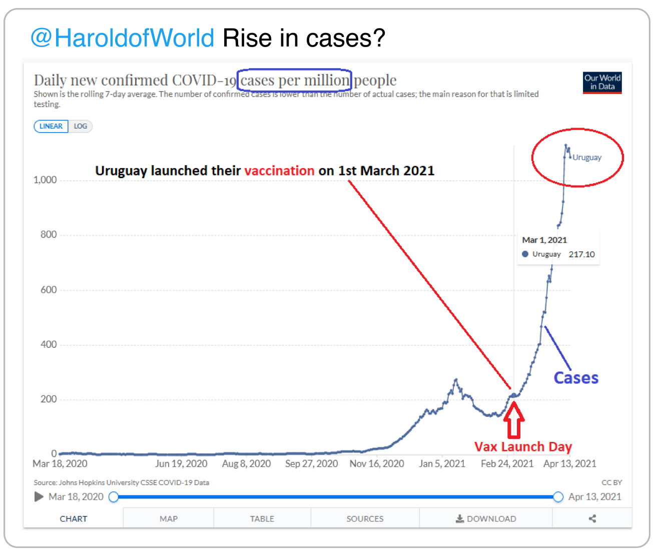


(a) Tweet 1313881187191054338

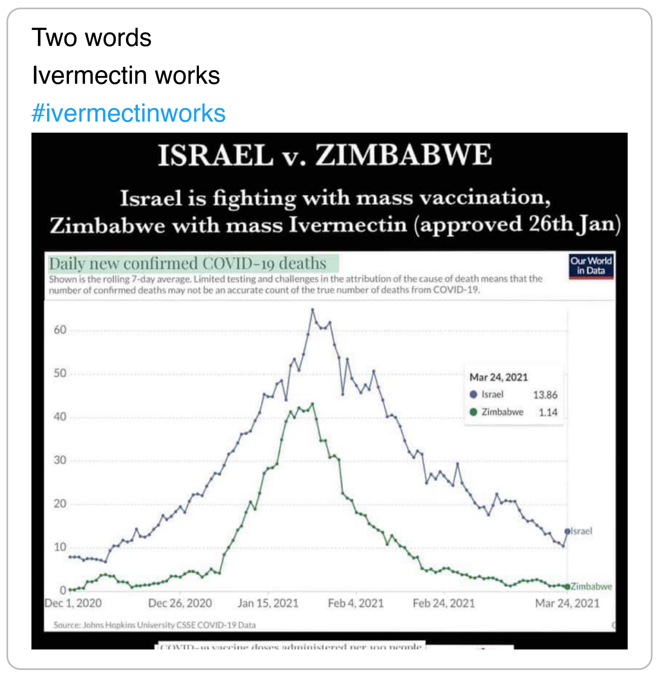


(b) Tweet 1375113309922291712

Figure 9: Examples of posts in which the user's argument hedges on the introduction of an *arbitrary threshold*. (a) A user shares data on causes of UK deaths and argues that since "only" 1 in 23 deaths was caused by COVID-19, it is not a significant problem. (b) An author shares a chart of annual mortality rate in the US and argues that an increase in deaths of "only" 15% is not significant enough.

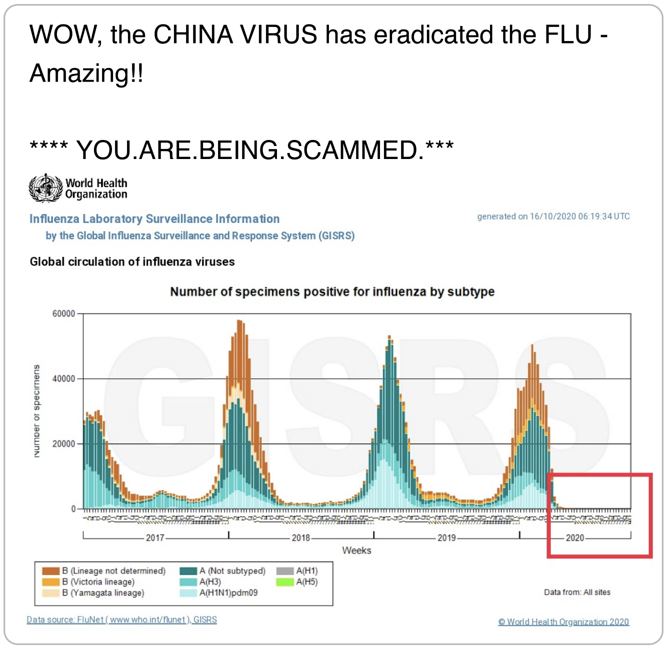


(a) Tweet 1382439566058065927 [removed]

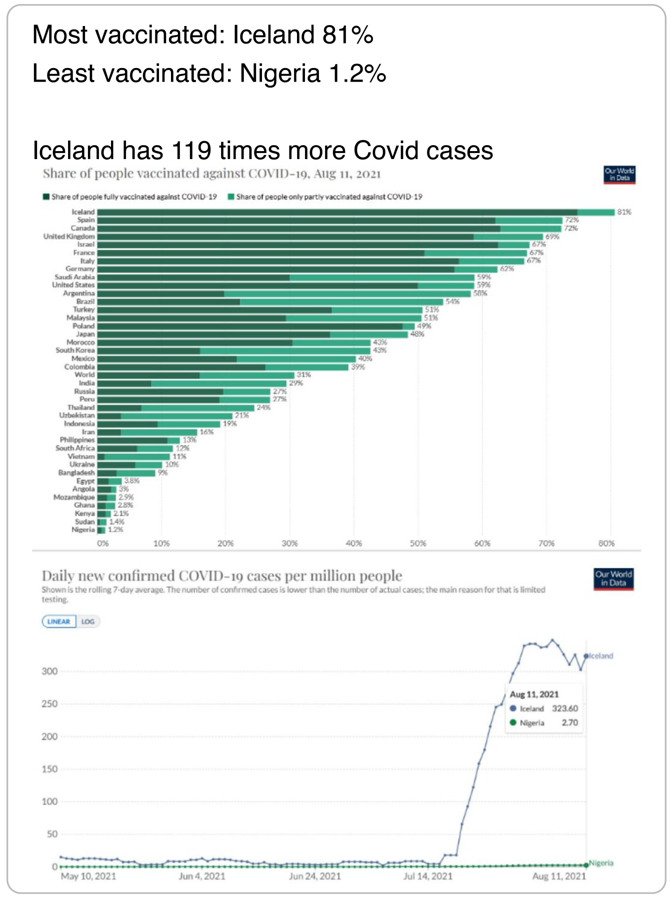


(b) Tweet 1375342512576077825 [removed]

Figure 10: Examples of posts implying *causal relationships* from limited and anecdotal data. (a) A tweet author suggests that the increase in COVID-19 cases in Uruguay—a prominent feature of the chart—was caused by the vaccination campaign. The user adds annotations to a COVID-19 dashboard screenshot to explain their reasoning. (b) An author shares a screenshot of a COVID-19 dashboard that shows that Israel is experiencing more COVID-19 deaths than Zimbabwe. The author states that the discrepancy is due to ivermectin being more effective than vaccinations.



(a) Tweet 1317061948228460546



(b) Tweet 1425925153322635276

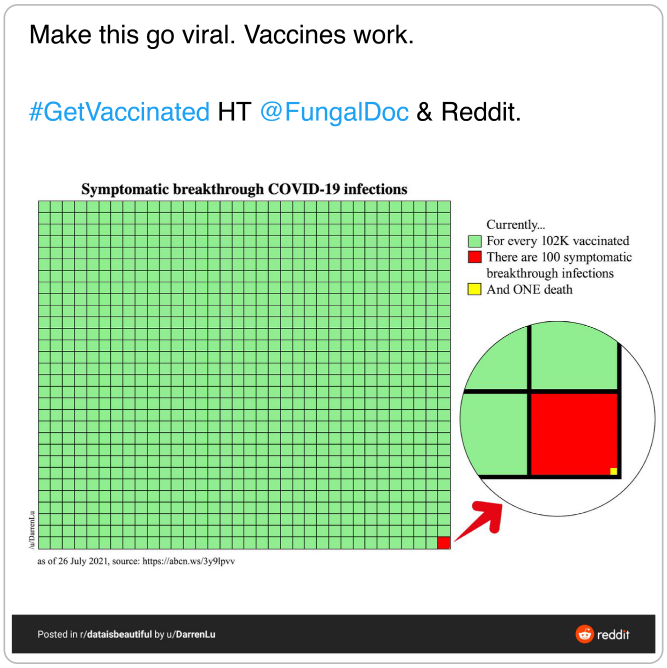
Figure 11: Examples of misleading arguments that suggest or do not account for *data validity*. (a) An author shares a WHO chart of flu cases and highlights a period of no new cases of flu. The author suggests that the COVID-19 pandemic is not real and the data are manipulated by miscounting flu cases as COVID-19. (b) A user shows two charts to highlight the fact that although Iceland is more vaccinated than Nigeria, it is experiencing more COVID-19 cases, implying that vaccines are not effective. The user fails to account for the fact that Iceland has a much higher testing rate, making it unreasonable to compare the two countries.

RE 4: Issues with Data Validity. During a fast-developing and novel crisis, issues may arise with the quality and consistency of data. In a developing situation, the lack of uncertainty communication may result in widespread confusion about what data can be used for inductive reasoning and how. If important caveats about data accuracy and data interpretation are not explicitly provided on charts [38], the viewer is left to trust the data to the level that supports their prior beliefs. The omission of such caveats results in two opposing strategies of using visualizations and data validity issues as basis of one's argument: appealing to data issues when they are not present, and ignoring them when they are.

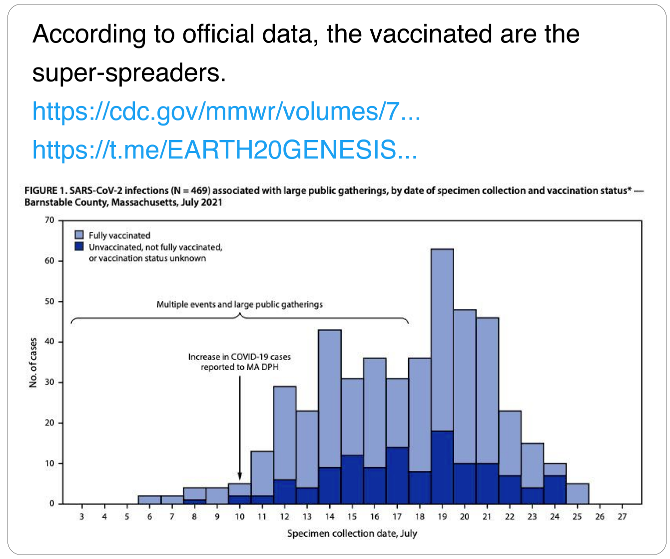
The first strategy, illustrated in Figure 11a, consists of the poster pointing out a salient feature of a chart, suggesting that it is caused by a data validity issue, and providing an explanation that supports a conspiracy. The author suggests that influenza being mistaken for COVID-19 is the primary reason for the small number of flu cases in 2020, the implication being that the pandemic is exaggerated.

On the other hand, one may reject verifiable data concerns if they provide support for their argument. In Figure 11b, the note on the chart explains that "[...] limited testing [...] means that the number of confirmed deaths may not be an accurate count." Nonetheless, the author points out that highly vaccinated Iceland has 119 times more cases than a low-vaccinated Nigeria and makes the argument that vaccination is not useful or potentially harmful. In this case, the data caveat likely has a large impact on the case chart. According to OurWorldInData [58], the source of the case chart, on the date highlighted by the author (August 11, 2021) Iceland performed over 200 times as many tests per capita as Nigeria did (8.93 versus 0.04 per 1,000 people).

As a result, data collection anomalies can be misused as evidence in support of misinformation, whereas explainable phenomena may be abused to sow potentially unwarranted distrust in data quality and reject an argument.



(a) Tweet 1420023807318843395



(b) Tweet 1422695341686677504

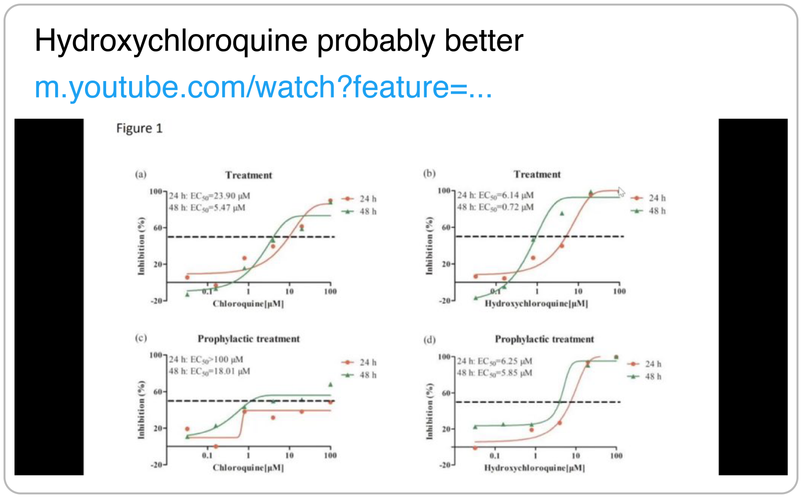
Figure 12: Examples of posts in which users *fail to account for statistical nuance* in their visualizations. (a) A user posts a chart showing the frequency of COVID-19 infections and deaths in the vaccinated population. The user concludes that vaccines are effective based on this chart alone, which is not possible without comparing the data to those from a control group. (b) An author reposts a chart from a CDC report [3] showing COVID-19 cases in Barnstable County, Mass. broken down between vaccinated and nonvaccinated. The author suggests that since there are more vaccinated cases, the vaccinated are "super-spreaders." The author fails to account for the high proportion of vaccinated in the general population. This caveat is highlighted in the text of the CDC report [3]: however, it is not communicated in the shared figure.

RE 5: Failure to Account for Statistical Nuance. Besides ignoring data issues, users often do not account for important statistical nuance in visualizations if doing so helps support their argument.

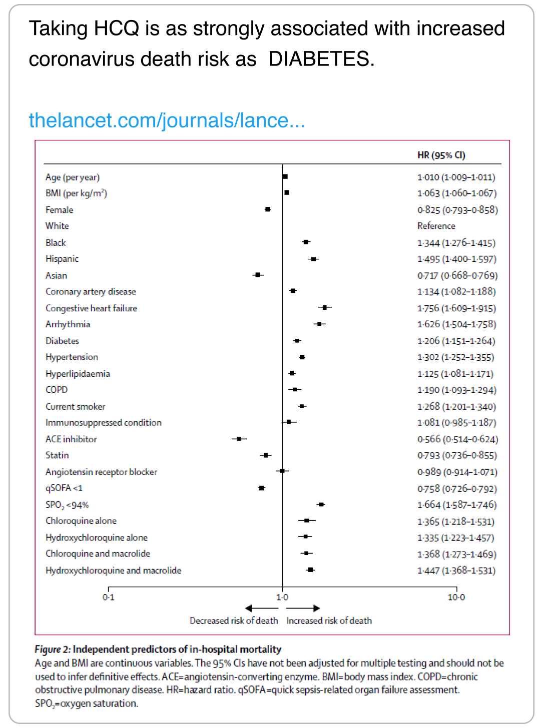
For instance, consider an experiment in which a given effect is evaluated by measuring outcomes for treatment and control groups. One common issue observed in our data set is users providing data showing the outcome of a single group in an experiment and judging the efficacy of the treatment against an arbitrary level of "goodness" rather than against the other group. In Figure 12a, the author argues that vaccines work by showing that among a population of 102,000 vaccinated, only one death has been reported. However, without knowing the death rate of the general population or of the nonvaccinated population, it is not possible to make conclusions about the efficacy of the treatment.

Another related issue is the *base rate fallacy*: charts providing outcome counts for both treatment and control groups without information about their relative sizes or about the general population. For instance, the attachment from Figure 12b is a chart that often has accompanied COVID-skeptic arguments. The chart comes from a report published by the CDC [3]. The chart shows that the majority of COVID-19 cases in the surveyed population were among previously vaccinated people. The report generally argues for increased COVID prevention measures and highlights the idea that the then newly emerged Delta variant is highly transmissible. When taken out of context, however, this chart can be used to support the idea that people are more likely to get and spread the disease if vaccinated. The report lists multiple limitations to the study in the discussion section, including a note that "..., data from this report are insufficient to draw conclusions about the effectiveness of COVID-19 vaccines against SARS-CoV-2, including the Delta variant, during this outbreak. As population-level vaccination coverage increases, vaccinated persons are likely to represent a larger proportion of COVID-19 cases." This caveat is not communicated on the chart itself, which quickly became a shareable artifact on social media and was interpreted outside of this context.

RE 6: Misrepresentation of Scientific Studies. Promoting scientific literacy has long been discussed as a way to inoculate the general public against misinformation [60]. In an increasingly complex world, it is not feasible for everyone to develop scientific expertise in relevant domains, and therefore one of the main goals of increasing scientific literacy is to encourage the population to cautiously trust science. Science educators describe an ideal of scientific literacy among lay audiences as the right balance between the extremes of believing any form of scientific authority and believing nothing at all [19, 45].



(a) Tweet 1239776019856461824



(b) Tweet 1264224112311844864

Figure 13: Examples of posts in which users *misrepresent scientific studies* through visualizations. (a) An author argues for the use of hydroxychloroquine in treatment of COVID-19. The shared figure is from an early 2020 in vitro study [75]. Later work by Lee et al. [35] notes that that although the evidence of the efficacy of hydroxychloroquine in vitro was promising, large-scale randomized clinical trials conducted afterwards have demonstrated low efficacy. The authors also discuss that the large number of hydroxychloroquine studies has been influenced largely by political pressure [35]. (b) A user argues against the use of hydroxychloroquine noting that it leads to an increased risk of mortality. The figure comes from a study that has since been retracted [42] due to concerns about the veracity of the data.

Many misleading tweets in our data set lie close to one of the two extremes. Most of the reasoning attributes described above are characteristic of the users's tendency to believe nothing but their personal experiences and observations, but we also identify a type of reasoning in which users accept any scientific findings that align with their prior beliefs at face value and exaggerate their interpretation. For instance, users share figures from studies on the efficacy of certain types of medication that have not yet been peer-reviewed, reproduced, or otherwise scrutinized, e.g., by approval for use in most countries. Figure 13 illustrates examples of such tweets.

In this form of reasoning, instead of rejecting scientific authority in favor of pseudo-science, users selectively exaggerate the importance of singular scientific results that confirm their beliefs.

RE 7: Incorrect Reading of Chart. In rare cases, visual distortions on the chart directly cause the viewers to arrive at inconsistent conclusions. Figure 14a illustrates an example that seems to show many fewer cases in Canada compared to the US, whereas the visual differences in the map are mostly caused by the inconsistent granularity of data between countries (province level for Canada and county level for the US). The author interprets the differences to be caused by COVID-19 restrictions.

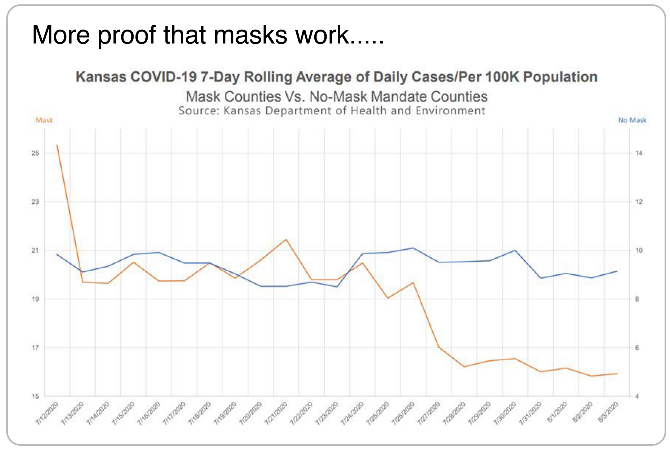
In another example in Figure 14b, the author attaches a dual axis line chart of COVID-19 cases in counties with and without mask mandates. Whereas the case numbers are actually higher in mask mandate counties than in counties without a mandate, the unequal relative scales of the axes make the line appear lower, which was misinterpreted by the user.

Previous research on misleading visualizations has been primarily focused on this type of error [49]. However, we find incorrect reading of charts to be the least common reasoning error in our analysis, as seen in Figure 6.

4.2.2 ***Construction Attributes.*** Construction attributes (CA) are graphical and textual methods that communicate or emphasize the message of the post. In misleading visualization tweets, construction attributes describe channels used to introduce or exaggerate reasoning errors described previously.

CA 1: Use of Post Text. Prior research by Kong et al. shows that the framing of a visualization greatly influences the viewers' interpretations [28]. In another study, the authors find that the

(a) Tweet 1322031826744020992



(b) Tweet 1291564622152183809

Figure 14: Examples of posts in which users *incorrectly read the charts* that violate common visualization design guidelines. (a) A user posts a map of COVID-19 cases in North America. The data in the map are not adjusted by population and is presented at different granularities: by county in the US and by province in Canada. As a result, the map looks much sparser in Canada than in the US. It is not possible to make precise readings from the map due to distortions, but the author nonetheless concludes that stricter COVID-19 measures in Canada are the cause of the visual discrepancy. (b) An author posts a dual axis chart of COVID-19 cases in Kansas broken down by counties with and without mask mandates. Both axes are truncated in such a way relative to each other (15–25 cases per 100k for counties with mask mandates, shown in orange, and 4–14 cases per 100k for counties without mask mandates), that even though counties with mask mandates have a higher number of cases, the line associated with them appears lower. The poster likely did not notice the vastly different scales and argues for the use of masks using this chart.

information from the title is more readily recalled by the viewers than the information from the chart itself [29]. This phenomenon provides an opportunity to deceive and mislead users by providing biased interpretations attached to the chart.

In the context of social media as a platform for sharing visualizations, including an interpretation via the post text is an intrinsic feature. In every tweet, authors add text to either provide their own interpretations of the chart or additional context that is important for their argument. In many examples, tweet text is used to frame a visualization in a way that makes it deceptive.

CA 2: Use of Annotations on Chart. Another channel for providing interpretations and context for a visualization is the use of annotations directly on the chart. Among opinion tweets, 21% of visualizations include textual or graphical annotations. Lin et al. [38] introduce the concept of *data hunches* and describe a design space for adding important context about the data representativeness to the data visualization itself. For instance, annotations added to Figure 10a provide additional information about the start of the vaccination campaign, whereas annotations on Figure 11a highlight the lack of flu cases—an important salient feature of the chart.

Although annotations on a chart can be essential to its understanding and stem from expert knowledge, they can also be deceptive and suggest relationships and caveats that do not exist. In their paper, Lin et al. warn about the potential for harm and argue for the use of annotations only within "groups of experts that are supported by networks of trust" [38]. In the context of charts shared for general audiences online, our research shows that annotations are likely to be misused and mislead the audience.

CA 3: Reframing Screenshots of Existing Charts. We find that the majority of visualizations in opinion tweets (58%) are screenshots of existing charts from reputable sources. One the one hand, this points to the ubiquity and ease of availability of COVID-19 visualizations and data. As an attribute of a misinformation tweet, however, reframing a reputable chart could provide an illusion of impartiality and trust on the part of the author of the tweet, as well as plausible deniability in case their argument is proven false. Aside from static charts from reputable sources, users often repost charts from interactive dashboards. Such dashboards let the user individually select regions to be plotted on the same chart and compared against each other. Although this approach allows one to freely explore the data, such tools could encourage cherry-picking of data points and faulty comparisons between them. From Figure 7, we can see that the combination of the use of interactive dashboards, cherry-picking, and causal inference is the most common set of attributes among posts with any reasoning errors.

Researchers have expressed concern that with the increasing prevalence and complexity of interactive visualizations, nonexpert viewers are more likely to misinterpret visualizations created by experts and identify spurious correlations [46]. Recent work also confirms that public health visualization dashboard designers should

(a) Tweet 1373985666225213440 [removed]

(b) Tweet 1388578805636157440

Figure 15: Examples of posts in which violations of common visualization design guidelines potentially exaggerate the argument. (a) A user highlights the fall in COVID-19 hospitalizations in the UK and attributes it to vaccination. The perceived scale of the decrease is likely exaggerated by the truncated axis and by the vertically elongated scale of the chart. (b) An author shares a dual axis chart of vaccinations and deaths in India and suggests a causal relationship between the two, which is likely exaggerated by the use and scale of the dual axis.

consider the possibility that their work can be used to mislead people [76]. The authors underscore the importance of expert knowledge in correct understanding of pandemic visualizations and seek ways to communicate important context.

CA 4: Violations of Common Visualization Design Guidelines. As seen from Figure 6, charts that violate common visualization design guidelines—for instance, those that use truncated axes or 3D figures—are not disproportionately used to support opinionated arguments. Moreover, results in Figure 6 show that explicit incorrect readings of charts caused by design violations are rare compared to other errors in visualization interpretation. Nonetheless, research shows that these techniques have the potential to affect the accuracy of viewers' perception of the chart [49].

Our findings suggest that even though design violations are not central to visualization-supported misinformation, they may help exaggerate the intended message. For instance, Figure 15a shows a post that attributes the drop in COVID-19 cases in the UK to the vaccination campaign. We would expect the perceived effect to be potentially stronger than the actual effect because the figure has a truncated axis [49]: upon first glance, a viewer might incorrectly infer that cases in the UK have almost completely disappeared. In another example, the post in Figure 15b assigns a cause-and-effect relationship to the rise of COVID-19 cases in India and increasing vaccinations. A dual axis chart is a common way to highlight a spurious correlation [72] that can make the association appear stronger and may even suggest a one-to-one relationship between cases and vaccines administered.

5 DISCUSSION

In this section, we consider the implications of the results of our analysis on the study of deceptive visualizations. We provide recommendations for general-audience visualization design as well as for the direction of future research.

5.1 Visual Misinformation Beyond Design Violations

We were surprised to discover that widely studied common visual deception techniques are not the main driver of visualization-supported misinformation online. The vast majority of both all

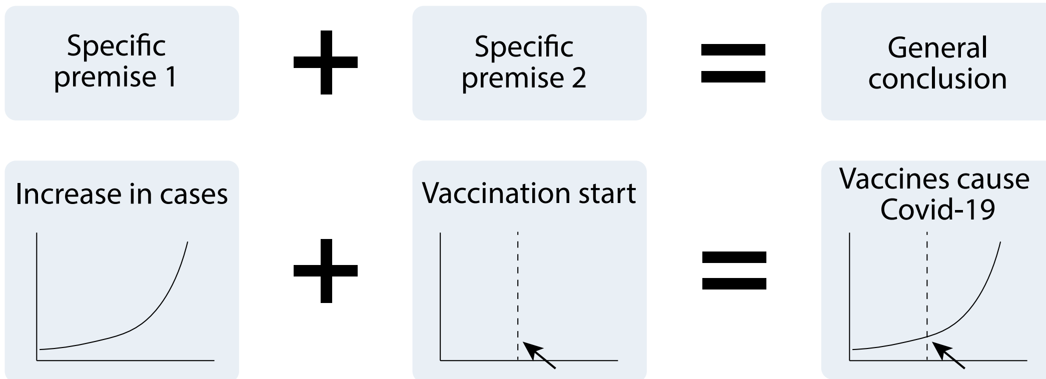


Figure 16: An example of inductive reasoning in a visualization tweet, similar to the tweet in Figure 10a. A user starts with a first premise: an existing chart showing an increase in COVID-19 cases. Through tweet text and chart annotations, the user provides a second premise: a vaccination campaign started around the same time the increase started. The user then suggests the conclusion: the vaccination start was the cause of the case increase. This attempt at creating a general conclusion from specific premises is logically consistent, but the conclusion is weakly supported by the limited premises.

charts in our data set in general (88%) and all charts with COVID-skeptic interpretations in particular (87%) do not have any features that violate common visualization design guidelines. The remaining 12% of visualizations violate visualization design guidelines by, for example, using truncated or inverted axes. However, these features are not typically used to support misinformation arguments and seem to occur at similar rates in opinion and neutral posts.

In posts that comment on the severity of the ongoing crisis, visualization design violations can be helpful to exaggerate the effect and help argue that, for instance, the impacts of the crisis are not sufficiently severe. However, the general argument of the post would still hold even if such techniques were not present, like in examples in Figure 15. Previous research suggests that design violations may exaggerate or diminish a message to a certain extent [49], but our analysis shows that they do not form the basis of visual misinformation and they are not disproportionately misinterpreted by tweet authors. In contrast, many data visualizations that fully conform to design guidelines can be used to support effective misinformation arguments.

Design violations are not more prevalent in online misinformation for several reasons. Firstly, whereas people may misinterpret the chart after a quick glance, engaging with it for a longer period of time required to attach it to a post and write text to go with it allows them to read the chart more carefully. Secondly, except for outright fabricated data, the information required to interpret the chart correctly is still present even in badly designed charts. We speculate that the possibility to read values accurately—albeit with difficulty—makes it easier for the audience to point out the mistake and to debunk the chart by leveraging their "collective intelligence" [48]. Consequently, it might be more difficult to make a deceptive argument with a mis-designed chart than a well-designed chart that contains more subtle reasoning flaws.

Most visualizations attached to our tweets of interest do not just conform to design guidelines but are also screenshots of charts from reputable sources, such as government and media outlets. Such visualizations are not intentionally created to be deceptive but rather are presented in a way that supports a common misinformation argument or conspiracy surrounding the crisis. This fact suggests that, depending on framing, conversation context, and anticipated audience, even some faithfully plotted visualizations are *vulnerable* to misinterpretation.

5.2 Visual Misinformation as Weak Inductive Reasoning

One helpful way to think about how framing can make a visualization deceptive is to consider how most, if not all, arguments with reasoning errors are constructed. Typically, users posit an idea supported by a chart and context. The salient features of the visualization and users's context form premises that weakly support an *inductive argument*—in other words, factual observations are used to derive a more general principle. Figure 16 illustrates the structure of an inductive argument schematically.

Misinformation arguments of this form generally do not contain formal logical fallacies, as the conclusion always logically follows from the presented premises. Deductive arguments can be either true or not true, whereas inductive arguments are defined to either be "cogent" or "not cogent"—in other words, plausible and not plausible [26]. The strength of an inductive argument, or the measure of how plausible the conclusion is, depends on the completeness and strength of the premises. Most empirical knowledge is also derived through inductive reasoning and is almost never definitive, which philosopher David Hume identified as a problem many centuries ago [25]. As a solution, in order to distinguish between scientific and pseudo-scientific theories, Karl Popper introduced the standard of *falsifiability*: a determination of whether a theory can be logically contradicted [56]. Popper describes pseudo-science as a "confirmation bias machine" that provides theories that are very good at offering explanations for all phenomena but do not present conditions under which the theory could be refuted [55].

Modern misinformation largely resembles pseudo-scientific theories of the past and is not always amenable to fact-checking because the premises are typically based on true data. Cook et al. [7]

present an analysis of climate change misinformation arguments through the lens of logic and reasoning and discuss that many climate denialist claims are plausible inductive arguments that are presented definitively. The authors discuss potential ways to invalidate such arguments through falsifiability, such as identifying hidden premises or implicit conditions necessary for the conclusion to hold. If shown to be not true or not plausible, the conclusion becomes not consistent or less plausible.

5.3 Hidden Premises Make Visualizations Vulnerable

In the context of visualization, we can consider potential hidden premises in posts in our analysis. As we have seen in examples throughout the paper, a typical visualization-based weak inductive argument often hinges on various implicit assumptions: posts omit a larger sample size in favor of cherry-picked data points, omit notes about impactful events that are expected to have an effect on the data, and also omit important caveats about data quality and uncertainty. Existing charts often do not communicate any of such data uncertainties as salient features and present data with an illusion of precision or certainty [23]. This false sense of accuracy and certainty likely empowers users to make definitive deceptive claims using visualizations as "scientific evidence."

In many cases, support for a misinformation conclusion may be minimized by explicitly visualizing the conditions or caveats the author wrongly assumes to corroborate their reasoning. For instance, in the case of the chart from the CDC report in Figure 12b, showing the population-level vaccination coverage would indicate to the audience that infection is not more prevalent in the vaccinated population; in the case of the chart that compares Iceland and Nigeria in Figure 11b, introducing the uncertainty of death counts due to limited testing as a graphical property of the chart may warn users against making comparisons. Therefore, although most visualizations in our analysis are not themselves deceptive, they are *vulnerable to deception*: they do not visualize important context and do not anticipate a biased reading and thus have no defenses against misinterpretations and supporting existing common misconceptions.

Weak inductive reasoning is not unique to misinformation arguments, or in our case specifically COVID-skeptic arguments. 25% of tweets with reasoning errors in our data consist of posts that support commonly accepted methods of COVID-19 prevention and treatment and follow similar reasoning, as for instance the post in Figure 12a that attempts to prove the efficacy of vaccines. In this example, however, the conclusion is corroborated by multiple scientific studies about vaccine efficacy [18, 47, 54, 69]. The omission of the control group outcomes on the chart in this case is not central to the argument. In fact, this omission has the opposite effect and makes the argument less convincing, which implies that understanding the problem of visual misinformation not only helps prevent misinformation but also improves the effectiveness of official crisis communication. By visualizing a more complete set of premises for the anticipated conclusion, a chart can provide stronger support to the take-away and leave fewer avenues for vaccine-skeptic attacks.

The author or the audience of a visualization post may not consider these important caveats to their interpretations that are not explicitly shown in a vulnerable visualization for several reasons. One reason is confirmation bias, or the tendency to interpret information in a way that reinforces prior beliefs [44]. If the conclusion matches existing beliefs, there is little to no incentive for the viewer to challenge their conclusion. Another reason is the process of social influence, or conformity to the demands of one's social environment [6]. This process is especially relevant in the context of our analysis—social media—where due to filter bubbles and personalized suggestions, users often end up seeing posts from, and interacting with, only one side of the discussion [50]. In such a setting, challenging assumptions of a post's reasoning can be perceived as antagonistic or may fail to get a stranger's attention [17].

5.4 The Role of Data Exploration Websites

Screenshots of charts from data exploration tools—such as the OurWorldInData COVID explorer [58] and Worldometer—are common in our data set and in the examples throughout this work. On the one hand, their design provides easy access to COVID-19 data and allows even nonexpert users to freely explore and share their findings. At the same time, this freedom of exploration can lead to many of the reasoning errors we have discussed, such as cherry-picking, causal inference, and failure to account for data validity issues.

With their design centered on comparing COVID-19 data from different regions, data exploration tools encourage many types of inferences that are prone to misinterpretation. This type of visualization is useful to contextualize data and answer personally meaningful questions: for instance, it helps infer health risks in a travel destination by comparing it to a familiar local baseline. However, plotting any subset of data on one chart also encourages potentially inaccurate inferences related to evaluating the effects of restrictions and interventions. Such inferences can be both unintentional, as an attempt to make sense of salient differences in the chart, or an intentional way of spreading disinformation by cherry-picking data.

Our results in Figure 6 show that more than 40% of all COVID-skeptic charts in the COVID-19 discourse on Twitter are screenshots of data exploration dashboards. This high prevalence may be explained by their popularity and ease of use. In the absence of explicit data caveats, the space of interactions offered in the dashboard is likely interpreted as the space of data representations that are valid for inference. As a result, users choose any subset of data that supports their existing beliefs or reinforces their (maybe not sincerely) held position.

In the case of OurWorldInData [58], data exploration tools are a relatively recent development: usually, the website provided in-depth data stories supported by interactive graphics, such as on the environmental impacts of food[3] or an analysis of the impact of vaccinations on COVID death rates[4]. We suspect that information presented in this way—as a guided narrative—is less susceptible to accidental misinterpretation. We highlight, however, the tension that platforms that support citizen science face. On the one hand, they inspire creativity, freedom of exploration, and democratize data. At the same time, such freedom can lead to a proliferation of conclusions based on weak premises and misinformation.

[3]https://ourworldindata.org/environmental-impacts-of-food
[4]https://ourworldindata.org/covid-deaths-by-vaccination

5.5 Design Recommendations

In this section, we summarize ways in which visualization designers can safeguard their charts or interactive dashboards and prevent them from being vulnerable to misinterpretations. Our design recommendations are practical methods for introducing hidden premises of potential interpretations of a visualization.

As we have briefly discussed before, in many cases it would be possible to **add a very salient feature to the chart that would render the misinformation argument not logical**. For example, the addition of population-level vaccination information to the CDC chart in Figure 12b would make it visually apparent that the proportion of COVID-19 patients who were vaccinated does not drastically exceed the proportion of the general population that is vaccinated. Even though the authors did consider the potential misinterpretation of their report and added a text note in an attempt to prevent it, including it directly in the chart as a visual feature would help make the important context more portable and resilient to viral spread. Charts should **make information about data caveats and accuracy more salient and dynamic**. For instance, many charts from data exploration websites in our analysis include a note in small font explaining that actual case numbers are likely much higher than reported, due to limited testing in some regions. However, this note is not a prominent feature of the chart and is not reflected in the visual design of the case curve. Additionally, the note is typically a generic statement that appears on all charts in the same form. The note could be more effective if the information about the testing limitations in a given region was chart-dependent and changed dynamically as viewers switch between regions.

Similarly, charts could **communicate the uncertainty of data estimates** to prevent viewers from assigning misleading meaning to data anomalies. Work by Hullman [23] discusses why visualization designers typically do not include uncertainty in their work products, including concerns that uncertainty can confuse the viewers, obfuscate the message of the plot, or undermine the designer's credibility. As our findings show, however, it is potentially beneficial to spread awareness about data imprecision to limit its support of pseudo-scientific arguments. Lin et al. [38] argue that "data hunches" should be explicitly yet distinctly communicated as part of the chart and should be considered in decision-making. Data exploration tools should prevent visualization of multiple items at the same time in cases when it is known that the items have a vastly different representativeness of the underlying phenomenon, as is, for example, the case when countries vary significantly in their testing strategies or data collection methodologies (Figures 10b and 11b).

In summary, we urge visualization designers to take into account potential misinterpretations of their charts and address them directly in the visual array. Our work shows that if important notes about data collection and use are not communicated as salient visual features in the chart, they are likely to be ignored by the general audience. Designers should consider what information is not shown in the chart and could be filled in by the viewers's beliefs and biases. Misinformation typically converges from many individual ideas to fewer, more common narratives [40, 73]. Therefore, visualization creators should be able to review common existing misconceptions on the topic and consider whether the visualization could be used to support them.

6 LIMITATIONS AND FUTURE WORK

Our research is subject to several important limitations. Firstly, our work is based only on data pertaining to the COVID-19 pandemic discourse on Twitter in English. However, we expect our findings to be generalizable to crisis and noncrisis situations where human behavior and data-driven policy choices have a significant effect on the economic and public health outcomes. Secondly, the specific affordances of a given social media platform may have an effect on the type of content that is shared and widely spread on it. In the case of Twitter, the strict character limit in tweet text may encourage users to post data visualizations, particularly ones with many annotations. The absence of replies in the feed view makes potentially important context added in the discussion not immediately available to the viewer. The fact that tweets are broadcasted in a general feed rather than common information spaces [59], make Twitter optimal for sharing information to reach a more diverse audience compared to, for instance, Facebook [52]. At the same time, lack of such common information spaces can also strip visualizations of their context, making them potentially more misleading. Thirdly, we are able to identify only explicit instances of chart misinterpretation as evidenced by social media trace data. Although the author of the tweet may not have explicitly misrepresent the chart, some of their audience might make inaccurate conclusions that were not shared, which could be valuable to explore in future research. Nonetheless, we believe studying misleading tweets that state their ideas explicitly is important, as this contributes to perpetuation of misinformation arguments.

We hope our contributions help direct future research in the spaces of deceptive visualization and online misinformation by offering a novel way of thinking about the notion of "lying with charts." Future work should formally analyze the replies to misleading social media posts, as well as explore the best ways to represent hidden premises into charts and evaluate the effectiveness of adding safeguards to vulnerable visualizations.

7 CONCLUSION

We collected, categorized, and organized thousands of data visualization posts from Twitter related to the COVID-19 pandemic to describe how people mislead with visualizations in practice. The results of our work show that visualization manipulation and the lie factor are not the main drivers of visual misinformation online. About 12% of charts we analyzed violate common visualization design guidelines, however, they are not typically used to support misinformation. Most COVID-skeptic data visualization posts on Twitter use faithfully plotted charts, accurate data, and make logically consistent arguments. Instead, tweet authors use salient features of visualizations as the premise for plausible inductive arguments that promote misinformation. In order to prevent a visualization from being vulnerable to such attacks, visualization designers should include safeguards in the form of important contextual information and uncertainty. Aside from optimizing the visualization design for its intended purpose, the designer should also anticipate how a biased viewer may use and misuse the chart.

ACKNOWLEDGMENTS

We wish to thank Sunny Siu, Jack Wilburn, Vineet Pandey for their contributions. This work is supported by the National Science Foundation (IIS 2041136).

REFERENCES

[1] Ahmer Arif, John J. Robinson, Stephanie A. Stanek, Elodie S. Fichet, Paul Townsend, Zena Worku, and Kate Starbird. 2017. A Closer Look at the Self-Correcting Crowd: Examining Corrections in Online Rumors. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW)*. ACM, Portland, OR, USA, 155–168. https://doi.org/10.1145/2998181.2998294

[2] Enrico Bertini, Michael Correll, and Steven Franconeri. 2020. Why Shouldn't All Charts Be Scatter Plots? Beyond Precision-Driven Visualizations. In *2020 IEEE Visualization Conference (VIS)*. IEEE, Salt Lake City, UT, USA, 206–210. https://doi.org/10.1109/VIS47514.2020.00048

[3] Catherine M. Brown, Johanna Vostok, and Hillary Johnson. 2021. Outbreak of SARS-CoV-2 Infections, Including COVID-19 Vaccine Breakthrough Infections, Associated with Large Public Gatherings - Barnstable County, Massachusetts, July 2021. *MMWR. Morbidity and Mortality Weekly Report* 70 (2021), 1059–1062. https://doi.org/10.15585/mmwr.mm7031e2

[4] Alberto Cairo. 2015. Graphics Lies, Misleading Visuals. In *New Challenges for Data Design*, David Bihanic (Ed.). Springer, London, UK, 103–116. https://doi.org/10.1007/978-1-4471-6596-5_5

[5] 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, Switzerland, 13–27. https://doi.org/10.1007/978-3-319-95831-6_2

[6] Robert B. Cialdini and Noah J. Goldstein. 2004. Social Influence: Compliance and Conformity. *Annual Review of Psychology* 55 (2004), 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015

[7] John Cook, Peter Ellerton, and David Kinkead. 2018. Deconstructing Climate Misinformation to Identify Reasoning Errors. *Environmental Research Letters* 13, 2 (2018), 024018. https://doi.org/10.1088/1748-9326/aaa49f

[8] Michael Correll. 2022. Are We Making Progress In Visualization Research?. In *2022 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)*. IEEE, Oklahoma City, OK, USA, 1–10. https://doi.org/10.1109/BELIV57783.2022.00005

[9] Michael Correll and Jeffrey Heer. 2017. Black Hat Visualization. In *Workshop on Dealing with Cognitive Biases in Visualisations (DECISIVe)*. IEEE, Phoenix, AZ, USA, 1–4.

[10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Miami, FL, USA, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[11] Evanthia Dimara, Steven Franconeri, Catherine Plaisant, Anastasia Bezerianos, and Pierre Dragicevic. 2020. A Task-Based Taxonomy of Cognitive Biases for Information Visualization. *IEEE Transactions on Visualization and Computer Graphics* 26, 2 (2020), 1413–1432. https://doi.org/10.1109/TVCG.2018.2872577

[12] Jannetje E. P. Driessen, Daniël A. C. Vos, Ionica Smeets, and Casper J. Albers. 2022. Misleading Graphs in Context: Less Misleading than Expected. *PLOS ONE* 17, 6 (2022), 1–9. https://doi.org/10.1371/journal.pone.0265823

[13] Twitter Engineering. 2012. Blobstore: Twitter's in-House Photo Storage System.

[14] Arlen Fan, Yuxin Ma, Michelle Mancenido, and Ross Maciejewski. 2022. Annotating Line Charts for Addressing Deception. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, New Orleans, LA, USA, 1–12. https://doi.org/10.1145/3491102.3502138

[15] Luis Ignacio Garegnani, Eva Madrid, and Nicolás Meza. 2022. Misleading Clinical Evidence and Systematic Reviews on Ivermectin for COVID-19. *BMJ Evidence-Based Medicine* 27, 3 (2022), 156–158. https://doi.org/10.1136/bmjebm-2021-111678

[16] Xinning Gui, Yubo Kou, Kathleen H. Pine, and Yunan Chen. 2017. Managing Uncertainty: Using Social Media for Risk Assessment during a Public Health Crisis. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, Denver, CO, USA, 4520–4533. https://doi.org/10.1145/3025453.3025891

[17] Aniko Hannak, Drew Margolin, Brian Keegan, and Ingmar Weber. 2014. Get Back! You Don't Know Me Like That: The Social Mediation of Fact Checking Interventions in Twitter Conversations. *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)* 8, 1 (2014), 187–196.

[18] Anne M. Hause. 2021. Safety Monitoring of an Additional Dose of COVID-19 Vaccine — United States, August 12–September 19, 2021. *MMWR. Morbidity and Mortality Weekly Report* 70 (2021), 1379–1384. https://doi.org/10.15585/mmwr.mm7039e4

[19] Friederike Hendriks, Dorothe Kienhues, and Rainer Bromme. 2016. Trust in Science and the Science of Trust. In *Trust and Communication in a Digitized World: Models and Concepts of Trust Research*, Bernd Blöbaum (Ed.). Springer International Publishing, Cham, Switzerland, 143–159. https://doi.org/10.1007/978-3-319-28059-2_8

[20] Thomas Heverin and Lisl Zach. 2012. Use of Microblogging for Collective Sense-Making during Violent Crises: A Study of Three Campus Shootings. *Journal of the American Society for Information Science and Technology* 63, 1 (2012), 34–47. https://doi.org/10.1002/asi.21685

[21] Aspen K. Hopkins, Michael Correll, and Arvind Satyanarayan. 2020. VisuaLint: Sketchy In Situ Annotations of Chart Construction Errors. *Computer Graphics Forum* 39, 3 (2020), 219–228. https://doi.org/10.1111/cgf.13975

[22] Darrell Huff. 1993. *How to Lie with Statistics* (reissue ed.). W. W. Norton & Company, New York, NY, USA.

[23] Jessica Hullman. 2020. Why Authors Don't Visualize Uncertainty. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2020), 130–139. https://doi.org/10.1109/TVCG.2019.2934287

[24] Jessica Hullman, Eytan Adar, and Priti Shah. 2011. The Impact of Social Information on Visual Judgments. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, Vancouver, BC, Canada, 1461–1470. https://doi.org/10.1145/1978942.1979157

[25] David Hume. 1793. *An Inquiry Concerning Human Understanding. A Dissertation on the Passions. An Inquiry Concerning the Principles of Morals. The Natural History of Religion.* T. Cadell, London, UK.

[26] Patrick J. Hurley. 2014. *A Concise Introduction to Logic.* Cengage Learning, Boston, MA, USA.

[27] Lauren Klein. 2022. What Data Visualization Reveals: Elizabeth Palmer Peabody and the Work of Knowledge Production. *Harvard Data Science Review* 4, 2 (2022), 1–34. https://doi.org/10.1162/99608f92.5dec149c

[28] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2018. Frames and Slants in Titles of Visualizations on Controversial Topics. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, Montreal, QC, Canada, 1–12. https://doi.org/10.1145/3173574.3174012

[29] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2019. Trust and Recall of Information across Varying Degrees of Title-Visualization Misalignment. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, Glasgow, UK, 1–13. https://doi.org/10.1145/3290605.3300576

[30] Klaus Krippendorff. 2004. *Content Analysis: An Introduction to Its Methodology.* SAGE, Thousand Oaks, CA, US.

[31] John M. Last. 2000. *A Dictionary of Epidemiology.* Oxford University Press, USA, New York, NY, USA.

[32] Claire Lauer and Shaun O'Brien. 2020. The Deceptive Potential of Common Design Tactics Used in Data Visualizations. In *Proceedings of the ACM International Conference on Design of Communication (SIGDOC)*. ACM, Virtual, 1–9. https://doi.org/10.1145/3380851.3416762

[33] Claire Lauer and Shaun O'Brien. 2020. How People Are Influenced by Deceptive Tactics in Everyday Charts and Graphs. *IEEE Transactions on Professional Communication* 63, 4 (2020), 327–340. https://doi.org/10.1109/TPC.2020.3032053

[34] 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 *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, Yokohama, Japan, 1–18. https://doi.org/10.1145/3411764.3445211 arXiv:2101.07993 [cs]

[35] Zelyn Lee, Craig R. Rayner, Jamie I. Forrest, Jean B. Nachega, Esha Senchaudhuri, and Edward J. Mills. 2021. The Rise and Fall of Hydroxychloroquine for the Treatment and Prevention of COVID-19. *The American Journal of Tropical Medicine and Hygiene* 104, 1 (2021), 35–38. https://doi.org/10.4269/ajtmh.20-1320

[36] Alexander Lex, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister. 2014. UpSet: Visualization of Intersecting Sets. *IEEE Transactions on Visualization and Computer Graphics (InfoVis)* 20, 12 (2014), 1983–1992. https://doi.org/10.1109/TVCG.2014.2346248

[37] Rensis Likert. 1932. A Technique for the Measurement of Attitudes. *Archives of Psychology* 22 140 (1932), 55–55.

[38] Haihan Lin, Derya Akbaba, Miriah Meyer, and Alexander Lex. 2022. Data Hunches: Incorporating Personal Knowledge into Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 504–514. https://doi.org/10.1109/TVCG.2022.3209451

[39] Leo Yu-Ho Lo, Ayush Gupta, Kento Shigyo, Aoyu Wu, Enrico Bertini, and Huamin Qu. 2022. Misinformed by Visualization: What Do We Learn From Misinformative Visualizations? *Computer Graphics Forum* 41, 3 (2022), 515–525. https://doi.org/10.1111/cgf.14559

[40] Alice E Marwick and William Clyde Partin. 2022. Constructing Alternative Facts: Populist Expertise and the QAnon Conspiracy. *New Media & Society* 00, 0 (2022), 1–21. https://doi.org/10.1177/14614448221090201

[41] Andrew McNutt, Gordon Kindlmann, and Michael Correll. 2020. Surfacing Visualization Mirages. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, Honolulu, HI, USA, 1–16. https://doi.org/10.1145/3313831.3376420

[42] Mandeep R. Mehra, Frank Ruschitzka, and Amit N. Patel. 2020. Retraction—Hydroxychloroquine or Chloroquine with or without a Macrolide for Treatment of COVID-19: A Multinational Registry Analysis. *The Lancet* 395, 10240

(2020), 1820. https://doi.org/10.1016/S0140-6736(20)31324-6

[43] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. 2010. Twitter under Crisis: Can We Trust What We RT?. In *Proceedings of Workshop on Social Media Analytics (SOMA)*. ACM, Washington, DC, USA, 71–79. https://doi.org/10.1145/1964858.1964869

[44] Raymond S. Nickerson. 1998. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology* 2, 2 (1998), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

[45] Stephen P. Norris. 1997. Intellectual Independence for Nonscientists and Other Content-Transcendent Goals of Science Education. *Science Education* 81, 2 (1997), 239–258. https://doi.org/10.1002/(SICI)1098-237X(199704)81:2<239::AID-SCE7>3.0.CO;2-G

[46] Mehrnoosh Oghbaie, Michael J. Pennock, and William B. Rouse. 2016. Understanding the Efficacy of Interactive Visualization for Decision Making for Complex Systems. In *IEEE Systems Conference (SysCon)*. IEEE, Orlando, FL, USA, 1–6. https://doi.org/10.1109/SYSCON.2016.7490526

[47] Samantha M. Olson. 2021. Effectiveness of Pfizer-BioNTech mRNA Vaccination Against COVID-19 Hospitalization Among Persons Aged 12–18 Years — United States, June–September 2021. *MMWR. Morbidity and Mortality Weekly Report* 70 (2021), 1483–1488. https://doi.org/10.15585/mmwr.mm7042e1

[48] Leysia Palen, Sarah Vieweg, Sophia B. Liu, and Amanda Lee Hughes. 2009. Crisis in a Networked World: Features of Computer-Mediated Communication in the April 16, 2007, Virginia Tech Event. *Social Science Computer Review* 27, 4 (2009), 467–480. https://doi.org/10.1177/0894439309332302

[49] Anshul Vikram Pandey, Katharina Rall, Margaret L. Satterthwaite, Oded Nov, and Enrico Bertini. 2015. How Deceptive Are Deceptive Visualizations? An Empirical Analysis of Common Distortion Techniques. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, Seoul, South Korea, 1469–1478. https://doi.org/10.1145/2702123.2702608

[50] Eli Pariser. 2011. *The Filter Bubble: What The Internet Is Hiding From You*. Penguin Books Limited, London, UK.

[51] 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 CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, Glasgow, UK, 1–12. https://doi.org/10.1145/3290605.3300474

[52] Shruti Phadke and Tanushree Mitra. 2020. Many Faced Hate: A Cross Platform Study of Content Framing and Information Sharing by Online Hate Groups. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, Honolulu, HI, USA, 1–13. https://doi.org/10.1145/3313831.3376456

[53] Kathleen H Pine, Myeong Lee, Samantha A. Whitman, Yunan Chen, and Kathryn Henne. 2021. Making Sense of Risk Information amidst Uncertainty: Individuals' Perceived Risks Associated with the COVID-19 Pandemic. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, Yokohama, Japan, 1–15. https://doi.org/10.1145/3411764.3445051

[54] Ian D. Plumb. 2022. Effectiveness of COVID-19 mRNA Vaccination in Preventing COVID-19–Associated Hospitalization Among Adults with Previous SARS-CoV-2 Infection — United States, June 2021–February 2022. *MMWR. Morbidity and Mortality Weekly Report* 71 (2022), 549–555. https://doi.org/10.15585/mmwr.mm7115e2

[55] Karl Raimund Popper. 2002. *Conjectures and Refutations: The Growth of Scientific Knowledge*. Psychology Press, East Sussex, UK.

[56] Karl Raimund Popper. 2002. *The Logic of Scientific Discovery*. Psychology Press, East Sussex, UK.

[57] Gilmar Reis, Eduardo A.S.M. Silva, Daniela C.M. Silva, Lehana Thabane, Aline C. Milagres, Thiago S. Ferreira, Castilho V.Q. dos Santos, Vitoria H.S. Campos, Ana M.R. Nogueira, Ana P.F.G. de Almeida, Eduardo D. Callegari, Adhemar D.F. Neto, Leonardo C.M. Savassi, Maria I.C. Simplicio, Luciene B. Ribeiro, Rosemary Oliveira, Ofir Harari, Jamie I. Forrest, Hinda Ruton, Sheila Sprague, Paula McKay, Christina M. Guo, Karen Rowland-Yeo, Gordon H. Guyatt, David R. Boulware, Craig R. Rayner, and Edward J. Mills. 2022. Effect of Early Treatment with Ivermectin among Patients with Covid-19. *New England Journal of Medicine* 386, 18 (2022), 1721–1731. https://doi.org/10.1056/NEJMoa2115869

[58] Hannah Ritchie, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, Diana Beltekian, and Max Roser. 2020. Coronavirus Pandemic (COVID-19). https://ourworldindata.org/coronavirus.

[59] Kjeld Schmidt and Liam Bannon. 1992. Taking CSCW Seriously: Supporting Articulation Work. *Computer Supported Cooperative Work* 1 (1992), 7–40.

[60] Aviv J. Sharon and Ayelet Baram-Tsabari. 2020. Can Science Literacy Help Individuals Identify Misinformation in Everyday Life? *Science Education* 104, 5 (2020), 873–894. https://doi.org/10.1002/sce.21581

[61] Kate Starbird, Ahmer Arif, and Tom Wilson. 2019. Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW)*. ACM, Austin, TX, USA, 127:1–127:26. https://doi.org/10.1145/3359229

[62] Kate Starbird, Jim Maddock, Mania Orand, Peg Achterman, and Robert Mason. 2014. Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter After the 2013 Boston Marathon Bombing. *IConference* 00, 0 (2014), 1–9. https://doi.org/10.9776/14308

[63] Kate Starbird, Leysia Palen, Amanda L. Hughes, and Sarah Vieweg. 2010. Chatter on the Red: What Hazards Threat Reveals about the Social Life of Microblogged Information. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW)*. ACM, Savannah, GA, USA, 241–250. https://doi.org/10.1145/1718918.1718965

[64] Anselm Strauss and Juliet M. Corbin. 1990. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Sage Publications, Inc, Thousand Oaks, CA, USA.

[65] James Surowiecki. 2005. *The Wisdom of Crowds*. Knopf Doubleday Publishing Group, New York, NY, USA.

[66] Jeannette Sutton, Leysia Palen, and Irina Shklovski. 2008. Backchannels on the Front Lines: Emergent Uses of Social Media in the 2007 Southern California Wildfires. *Proceedings of the 5th International ISCRAM Conference* 00, 0 (2008), 1–9.

[67] Danielle Albers Szafir. 2018. The Good, the Bad, and the Biased: Five Ways Visualizations Can Mislead (and How to Fix Them). *Interactions* 25, 4 (2018), 26–33. https://doi.org/10.1145/3231772

[68] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. 2016. Rethinking the Inception Architecture for Computer Vision. In *IEEE International Conference on System, Computation, Automation and Networking (ICSCA)*. IEEE, Las Vegas, NV, USA, 2818–2826. https://doi.org/10.1109/CVPR.2016.308

[69] Mark W. Tenforde. 2022. Effectiveness of mRNA Vaccination in Preventing COVID-19–Associated Invasive Mechanical Ventilation and Death — United States, March 2021–January 2022. *MMWR. Morbidity and Mortality Weekly Report* 71 (2022), 459–465. https://doi.org/10.15585/mmwr.mm7112e1

[70] Edward R. Tufte. 1983. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, USA.

[71] Sarah Vieweg, Amanda L. Hughes, Kate Starbird, and Leysia Palen. 2010. Microblogging during Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, Atlanta, GA, USA, 1079–1088. https://doi.org/10.1145/1753326.1753486

[72] Tyler Vigen. 2017. *Spurious Correlations*. Hachette Books, New York, NY, USA.

[73] Liang Wu, Fred Morstatter, Xia Hu, and Huan Liu. 2016. Mining Misinformation in Social Media. In *Big Data in Complex and Social Networks*. Chapman and Hall/CRC, London, UK.

[74] Cindy Xiong, Lisanne Van Weelden, and Steven Franconeri. 2020. The Curse of Knowledge in Visual Data Communication. *IEEE Transactions on Visualization and Computer Graphics* 26, 10 (2020), 3051–3062. https://doi.org/10.1109/TVCG.2019.2917689

[75] Xueting Yao, Fei Ye, Miao Zhang, Cheng Cui, Baoying Huang, Peihua Niu, Xu Liu, Li Zhao, Erdan Dong, Chunli Song, Siyan Zhan, Roujian Lu, Haiyan Li, Wenjie Tan, and Dongyang Liu. 2020. In Vitro Antiviral Activity and Projection of Optimized Dosing Design of Hydroxychloroquine for the Treatment of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *Clinical Infectious Diseases* 71, 15 (2020), 732–739. https://doi.org/10.1093/cid/ciaa237

[76] Yixuan Zhang, Yifan Sun, Joseph D. Gaggiano, Neha Kumar, Clio Andris, and Andrea G. Parker. 2023. Visualization Design Practices in a Crisis: Behind the Scenes with COVID-19 Dashboard Creators. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 1037–1047. https://doi.org/10.1109/TVCG.2022.3209493

[77] Yixuan Zhang, Yifan Sun, Lace Padilla, Sumit Barua, Enrico Bertini, and Andrea G Parker. 2021. Mapping the Landscape of COVID-19 Crisis Visualizations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, Yokohama, Japan, 1–23. https://doi.org/10.1145/3411764.3445381