

Misinformation by Visualization: What Do We Learn From Misinformative Visualizations?

Leo Yu-Ho Lo¹, Ayush Gupta¹, Kento Shigyo¹, Aoyu Wu¹, Enrico Bertini², and Huamin Qu¹

¹Hong Kong University of Science and Technology
²Northeastern University

Abstract

Data visualization is powerful in persuading an audience. However, when it is done poorly or maliciously, a visualization may become misleading or even deceiving. Visualizations give further strength to the dissemination of misinformation on the Internet. The visualization research community has long been aware of visualizations that misinform the audience, mostly associated with the terms “lie” and “deceptive.” Still, these discussions have focused only on a handful of cases. To better understand the landscape of misleading visualizations, we open-coded over one thousand real-world visualizations that have been reported as misleading. From these examples, we discovered 74 types of issues and formed a taxonomy of misleading elements in visualizations. We found four directions that the research community can follow to widen the discussion on misleading visualizations: (1) informal fallacies in visualizations, (2) exploiting conventions and data literacy, (3) deceptive tricks in uncommon charts, and (4) understanding the designers’ dilemma. This work lays the groundwork for these research directions, especially in understanding, detecting, and preventing them.

CCS Concepts

• **Human-centered computing** → **Information visualization**;

1 Introduction

Data visualization has become a part of our daily information consumption. The general public consumes information from visualizations appearing in newspapers, on television, and on the Internet. Some of these visualizations have the purpose of swaying the audience toward a particular agenda. For example, they are used in political promotions, business advertisements, donation programs, and even on electricity bills to persuade us to reduce our energy consumption [TS08]. The study by Pandey *et al.* showed that data visualization has a significant effect on fortifying a message or prompting a change [PMN*14]. When visualizing the supporting data faithfully, these usages are good demonstrations of the power of data visualization. However, when the data does not support the intended claims, it is tempting to distort the visualizations to make them look supportive, which is misleading.

The existence of misleading visualization cases has long been documented and discussed. Back in the 1950s, before personal computers, the book *How to Lie with Statistics* by Huff [Huf54] collected examples of misleading charts from the newspapers. Nowadays, we have a much larger data source on the Internet. Still, the discussions among the research community are still focused on a limited set of cases – most commonly, truncated axis, area encoding, and 3D charts [PRS*15, Cai15, CH17, MK18, Sza18, MWN*18, ZS19, HCS20, LO20]. We imagine that taking a broad look at the

problem will help further the discussion, and perhaps we may find commonalities in other cases that may lead to breakthroughs on these long-discussed cases.

To widen our horizon on misleading visualization, we have collected 129,125 visualization images from the Internet through keyword searches from three major search engines and four social media platforms. We open-coded the collected images until we reached saturation, *i.e.*, after a period of no discovery of any new type of issues. At the end of the coding process, we had examined 6,500 images (~5% of the collected images), and among these images, 1,979 images (~30% of examined images) were candidate images—visualizations that are the target of this study—and 1,143 (~58% of candidate images) were tagged with issues. We discovered 74 different types of issues and organized them into a taxonomy of 12 categories. These 12 categories further fit into a five-stage visualization process. Figure 1 shows the 74 types of issues in a five-stage visualization process.

From the taxonomy and the notes taken during the coding process, we identified four directions that would enrich the discussion on misleading visualizations: (1) informal fallacies in visualizations, (2) exploiting conventions and data literacy, (3) deceptive tricks in uncommon charts, and (4) providing design guidelines by understanding the designers’ dilemma.

Informal fallacy sneaks in at the input and interpretation stages

of the visualization process. For example, cherry-picking or double counting the data and false analogies or causal linkages in the message interpreted from the visualization. We suspect that most of the informal fallacies in argumentation also apply to visualizations. Secondly, miscommunication happens when the convention is broken implicitly, and it happens with visualizations, too. There are many implicit expectations when viewing a chart, for example, expecting a linear scale but the data is shown in log scale, and expecting consistent intervals between tick marks but the scale is changing along the axis. Charts become misleading when the implicit assumptions are not met. Thirdly, a word cloud is a common visualization form, but it is rarely reported as misleading. Are word clouds immune to being misleading? Or, are misleading word clouds stealthy enough to go unspotted? Are there any other visualization forms that share this commonality? These are the questions we are eager to answer, and they may lead us in a new research direction. Lastly, not all misleading visualizations intend to mislead. We have seen examples where the visualization authors have made their best design effort to avoid misleading the audience, given the limitation of the underlying data and presentation form. For example, the age groups in the dataset are unevenly split, so avoiding an inconsistent binning size is challenging, or a projection of the spherical Earth onto a 2D world map is unavoidably distorted. Understanding these dilemmas in misleading visualizations is an important step toward providing practical guidelines for visualization practitioners.

2 Related Work

2.1 Misinformation

The Internet is filled with false information in the post-truth era [Key04]. The current study of misinformation has been focused on the textual content and propagation network, but the equally important visual content is being neglected [WD17]. The false information that has been circulated online during the pandemic is an unfortunate demonstration of the key role played by visual content. “Counter-visualizations,” a term created by Lee *et al.* [LY1*21] to describe the visualizations used in unorthodox ways, are widely circulated on social media platforms to challenge public health measures. The COVID-19-related counter-visualizations and their associated discussions on social media platforms show that the use of a powerful persuasive device—data visualization—can work against scientific reasoning. We echo the authors’ call that it is a matter of urgency for the research community to pay more attention to the visualizations created by non-experts and those circulated online.

2.2 Lie with Visualizations

Studies on misleading visualizations began a long time ago, before the Internet. Back in the 1950s, Huff’s classic statistics book, *How to Lie with Statistics*, included misleading visualizations collected from newspapers, illustrating how journalists applied tricks to distort the actual meaning of the underlying data [Huf54]. He showed that the problem of misleading visualizations is highly related to the nature of statistical communication. This was at a time when personal computers had not yet been invented.

In the age of computer-drawn visualizations, the discussion continued with wider coverage by Tufte in the 1980s, among others. In his book *The Visual Display of Quantitative Information*, Tufte

introduced the terms “graphical integrity” and “lie factor,” stating that visual encoding should proportionally represent the numerically measured values in a visualization [TGM83]. Despite the rule’s simplicity, it captures many basic cases of misleading visualizations—for example, truncated axis, area encoding, and 3D pie charts. However, the rule’s controversy is related to its oversimplicity on the cases not covered, like an inverted axis and cases where violating the rule to present the data visually is unavoidable, such as the valid use of log scale and 2D map projection.

In a recent book by Cairo entitled *How Charts Lie*, the discussion is more focused on spotting the dubious data underlying the visualizations and the intended message of the visualizations’ authors [Cai19]. Using the examples collected from current affairs and the Internet, the author revealed how visualization creators could hide the underlying dubious data and present their intended message to the audience. Instead of misleading the audience by manipulating the visual elements that can be spotted from the visualizations themselves, like truncated axis, or 3D charts, manipulating the data is much harder to spot. As the author quoted Ronald Coase, “If you torture data long enough, it’ll always confess to anything.” These examples also reflect the reality of the digital world we live in. The abundance of content keeps us busy, so we have no time to verify it. This type of misleading content is impossible to identify from the visualization image alone without understanding the context.

Monmonier’s *How to Lie with Maps* [Mon91] follows the spirit of the 1950s classic and focuses on how cartographic visualizations may provide misleading information. Maps are one of the most common visualization types, and we are familiar with the distorted two-dimensional projection of the three-dimensional spherical Earth. The commonly used Mercator projection has long resulted in a misunderstanding of Africa’s size, which is largely understated on the projected 2D world map. This case reminds us that the creators are not necessarily ill-intended when making visualizations that might eventually mislead people. Instead, it is a design trade-off in a dilemma situation. Technological advancements may provide solutions to these dilemmas. For example, Google Maps has been designed to show the actual spherical size on a 2D screen. We encountered more examples of these design dilemmas, and they are not limited to maps. They are in various visual forms and have different rationales inherited from the underlying data that cause these design dilemmas.

The above books are good collections of misleading visualization examples. This work aims to complement and expand their work to provide broader coverage of misleading visualizations on the Internet. From our collected visualization examples, we found that some misleading visualizations indeed try to add visual cues to avoid misleading the audience. Despite the best efforts of the designers, it is not enough to prevent the audience from getting the wrong message. Understanding these dilemmas will be a step forward for the community in developing practical guidelines for designers to make better design decisions.

2.3 Academic Research on Misleading Visualizations

The motivation of this work is to open up more research directions for the study of misleading visualizations and look into the

latest developments in misleading tricks being used in visualizations. Several previous works have summarized from the literature the potential pitfalls when creating visualizations [BE15, MKC20], optical illusions [BP06], and cognitive biases [DFP*18]. Each of these works includes a rich set of potential causes of misleading visualizations. Still, the study of misleading visualizations has focused on a few tricks inspired by real-world examples [CH17].

Among many misleading tricks, truncated axis, inverted axis, area encoding, dual-axis, and rainbow colors are the most commonly picked for the studies, usually by heuristics or from the authors' experiences [PRS*15, Cai15, CH17, MK18, Sza18, MWN*18, ZS19, HCS20, LO20]. The notorious inverted axis visualization titled "Gun deaths in Florida" has been frequently discussed. In contrast, in our labeled set of real-world examples, inverted axis is a comparatively rare occurrence. Why is that the case? On the other hand, are there tricks that we have overlooked but widely used? This work aims to better understand the domain of misleading visualizations from real-world examples and guide our future research on developing detection and prevention measures [HCS20].

3 Method

To build up an understanding of the domain of misleading visualizations, we followed the grounded theory method (GTM) for human-computer interaction studies described by Muller [Mul14]. GTM is a widely used method in HCI studies for understanding uncharted domains. It is implemented through continuous data collection, coding the data, and revising the developing theory until theoretical saturation is reached, also known as "no further surprises."

3.1 Data Collection

GTM starts with data collection, but collecting a large set of real-world examples of misleading visualizations was challenging. Fortunate or not, compared to the sea of faithful visualizations, the existence of misleading visualizations is relatively sparse. For the aforementioned book authors, collection relies on spotting misleading visualizations encountered in daily life, on the Internet, or in the news media. Such a collection takes a long time and substantial devoted effort, and in the end, the collected dataset may not have representative coverage because of personal bias in the collection process.

We turned our attention to the reported cases on the Internet through web search engines and social media platforms. There are three major search engines, (1) Google, (2) Bing, and (3) Duck-DuckGo, and four major social media platforms, (1) Facebook, (2) Twitter, (3) Pinterest, and (4) Reddit. For the keyword terms, we combined the ten adjectives or verbs derived from the synonyms of "misleading" and five nouns derived from the synonyms of "visualization" to form 50 keyword terms. We applied these keyword searches on search engines and social media platforms. The exception to this was Reddit, which has a subforum named *r/dataisugly* dedicated to discussing misleading or poorly made visualizations. We found that the discussion posts on this subforum were more relevant than the results of a keyword search. Therefore, instead of using a keyword search on Reddit, we crawled all the discussion threads on *r/dataisugly*. By crawling the keyword search results from three search engines, three social media platforms, and the

r/dataisugly on 24th February 2021, a total number of 129,125 images, including irrelevant images, were collected, and the composition is shown in Table 1. While this method provides representative coverage, it is limited to the reported cases, and the unspotted cases are uncovered. We suspect that there are many more stealthy cases not being spotted. It remains an important question to explore in future work: what are the characteristics of stealthy misleading visualizations? We have included further discussions of this question, summarized from our findings, in section 6.

3.2 Open-Coding

Since the collected images are mixed with irrelevant images—images that are not visualizations—we filtered the collection first before open-coding. To clarify the terminology for describing the coding process, we use "image" to refer to image files and "visualization" to refer to the charts in the image. Besides relevancy, we further narrowed down the focus of the study to single-viewed visualizations and included only the common chart types, *i.e.*, bar, line, circle, area, map, and point, by following the chart type taxonomy proposed by Borkin *et al.* [BVB*13]. Out of curiosity, the previous study by Borkin stated a counter-intuitive fact that text visualizations are rarely seen in their collected dataset. We wanted to check if this was also the case for our collected images and therefore also included text visualization as our study targets.

To ensure the ordering of the collected images would not affect the coding process, we first performed a random shuffle on the order of all the collected images and then fetched images in batches for filtering and coding. The open coding process is:

1. Fetch a batch of 500 images from the randomly shuffled images.
2. Filter out irrelevant images.
3. Filter out visualizations with multiple views.
4. Filter out visualizations that are not one of the following types: (1) bar, (2) line, (3) circle, (4) area, (5) map, (6) point, or (7) text.
5. Examine each of the images and the associated social media post and web page to figure out the issues in the visualizations and add one or multiple tags to it.

During the coding process, we found that, although we started with the intention of studying misleading visualizations, the collected visualizations included a lot of uninformative visualizations. This type of visualization does not mislead, but the reader cannot get any information from it either. When we read a visualization, we expect to get information from it, but uninformative visualizations simply leave readers confused. For example, a chart with no axis titles nor units is not comprehensible and provides no useful information regardless of whether it is drawn accurately or not (Figure 3e). We found that the study of misleading visualizations is inseparable from these uninformative visualizations. They are both failed cases of data visualization. To avoid assuming the intention of the visualization authors, we use the collective term "issues" to refer to both (1) *distortion*, where the reader gets a different message than the information in the data, and (2) *confusion*, where the reader cannot get information from the visualizations at all. These two types of issues, respectively, result in **misinformative** and **uninformative** visualizations.

Two of the authors open-coded the first two batches, a total of

Table 1: Numbers of collected images by data source and keywords. Relevancy—the number of candidate images in 100 return results—reflects the relevant images after removing the “junk” returned by data sources. Accuracy—the number of tagged images in 100 return results—reflects the misinformative or uninformative visualizations returned by data sources. *r/dataisugly* is a subforum of Reddit dedicated to the discussion of misleading or poorly made visualizations. The highest values are highlighted in bold.

Platforms	# of Images (%)	# of Examined (%)	# of Candidates (Relevancy)	# of Tagged (Accuracy)	Adjectives / Verbs	# of Images (%)	# of Examined (%)	# of Candidates (Relevancy)	# of Tagged (Accuracy)	Nouns	# of Images (%)	# of Examined (%)	# of Candidates (Relevancy)	# of Tagged (Accuracy)
<i>r/dataisugly</i> *	6862 (5.31%)	330 (5.08%)	227 (68.79%)	201 (60.91%)	Lie	21805 (16.89%)	1137 (17.49%)	239 (21.02%)	96 (8.44%)	Chart	34070 (26.39%)	1752 (26.95%)	640 (36.53%)	320 (18.26%)
Google	35488 (27.48%)	1738 (26.74%)	672 (38.67%)	376 (21.63%)	Fraud	17085 (13.23%)	887 (13.65%)	171 (19.28%)	61 (6.88%)	Graph	27433 (21.25%)	1399 (21.52%)	711 (50.82%)	381 (27.23%)
DuckDuckGo	28938 (22.41%)	1445 (22.23%)	498 (34.46%)	240 (16.61%)	Misleading	16290 (12.62%)	793 (12.20%)	389 (49.05%)	292 (36.82%)	Visualization	23031 (17.84%)	1144 (17.60%)	260 (22.73%)	155 (13.55%)
Bing	17224 (13.34%)	861 (13.25%)	159 (18.47%)	90 (10.45%)	Deceive	12410 (9.61%)	605 (9.31%)	109 (18.02%)	55 (9.09%)	Vis	21027 (16.28%)	986 (15.17%)	51 (5.17%)	32 (3.25%)
Pinterest	15435 (11.95%)	824 (12.68%)	59 (7.16%)	30 (3.64%)	Deceptive	11927 (9.24%)	593 (9.12%)	203 (34.23%)	127 (21.42%)	Viz	16702 (12.93%)	889 (13.68%)	90 (10.12%)	54 (6.07%)
Twitter	13846 (10.72%)	714 (10.98%)	272 (38.10%)	170 (23.81%)	Dubious	11127 (8.62%)	581 (8.94%)	133 (22.89%)	54 (9.29%)					
Facebook	11332 (8.78%)	588 (9.05%)	92 (15.65%)	36 (6.12%)	Dishonest	8908 (6.90%)	440 (6.77%)	134 (30.45%)	74 (16.82%)					
					Misrepresent	8092 (6.27%)	395 (6.08%)	155 (39.24%)	100 (25.32%)					
					Misguide	7751 (6.00%)	385 (5.92%)	110 (28.57%)	45 (11.69%)					
					Misinform	6868 (5.32%)	354 (5.45%)	109 (30.79%)	38 (10.73%)					
Total	129125	6500	1979	1143										

**r/dataisugly* is not crawled with keywords

1,000 images, independently by following the procedure mentioned above and then discussing the definition of the tags. The undecidable cases were put into a discussion with a third author. Examples of these undecidable cases are whether it is valid to truncate the y-axis of a line chart or visualization that require extra context to be understood that is not included in the associated social media post or web page. The authors discussed and looked for more context on the Internet to resolve these cases. In this way, the two coders iteratively discussed and refined the definition of each tag until they reached Cohen’s $\kappa > 0.7$. Among the first two batches of 1,000 images, 303 were candidate images, 174 were tagged, and 52 tags were discovered. Following the codebook, one of the coders continued to code images batch by batch. Upon discovering new tags, the two coders discussed their definitions and checked if they overlapped with any previously discovered tags to ensure the tags were mutually exclusive. The coding process continued until the saturation criterion was reached, *i.e.*, no new tag was found in the last 100 tagged images. In total, 6,500 images were examined, 1,979 of them were candidate images, 1,143 were tagged with at least one issue, and 74 tags were discovered. To ensure correctness, the coders then performed a within-tag consistency check. The tagged images are publicly accessible on <https://leoyuholo.github.io/bad-vis-browser>, and the shuffled list of 129,125 images and the open-coding results can be found on OSF: <https://osf.io/wghxd>.

4 Data Sources, Keywords and Limitations

Table 1 shows the number of results returned from different sources. Since search engines return results from the same pool of content and users on social media platforms tend to cross-post the same content onto different platforms, these data sources are not mutually exclusive. One of these cases is the notorious inverted axis visualization “Gun deaths in Florida.” From the labeled 19 examples of inverted axis, the “Gun deaths in Florida” visualization was repeated 11 times. Such a dramatic case with the reversal effect of an inverted axis has been widely cross-posted across the Internet.

The Reddit subforum *r/dataisugly* has both high relevancy and accuracy, reflecting it as a highly concentrated discussion forum on misinformative and uninformative visualizations. Since all these return results are reported cases associated with the keywords or the discussion topics, the unreported cases on the low relevancy and accuracy sources may be prevalent but not spotted or discussed. The study by Lee *et al.* [LYI*21] on both Twitter and Facebook is a reminder of such prevalence that the collected dataset cannot cap-

ture. The keywords “misleading,” “misrepresent,” and “deceptive” have the highest relevancy and accuracy. Surprisingly, the keyword “lie” has low accuracy despite most related books being titled with it. Many of the examined images are related to financial data and mention “Charts Don’t Lie” in the associated social media posts, which may refer to the serial financial advice book authored by Ryan [Rya14]. For the nouns, visualizations are most commonly referred to as “chart” and “graph,” while the short forms “vis” and “viz” are not. Studying the visualizations circulating on the Internet is an important yet neglected aspect of understanding information disorder [WD17]. The reported statistics may inform future studies on choosing keywords and selecting data sources.

Despite our best efforts to gather images from the widest available sources, our coverage was still limited by the following factors:

1. Although there are visualizations in languages other than English, the majority of the visualizations are in English.
2. The visualizations are static images. Videos and animated visualizations are not covered.
3. The sampled population from different data sources has overlaps. Same visualization or its variants—with annotations, corrections, or framing—may be posted more than once. For example, we have seen the notorious “Gun deaths in Florida” and its variants 11 times during the coding process.
4. The population reflects only the reported cases. Issues that are less discussed or less aware are underrepresented because the coding process relies on explicit reporting and the coders’ awareness to spot the issues in the visualizations. For example, color blind unfriendly is more ubiquitous than the only reported case in the tagged images.

In the coding process, we discovered 74 issues in visualizations. Examples of the 15 most common issues can be found in the Appendix. Among these common issues, we can locate the frequently discussed issues of *Truncated Axis* (1st), *3D* (2nd), *Dual-axis* (4th), and *Area Encoding* (12th). If *Area Encoding* is combined with *Pictorial Area Encoding*, it ranks at 8th. Despite that, there are more common issues missing from the discussions. The following issues can be misleading: *Misrepresentation* (5th), *Inconsistent Tick Interval* (7th), *Not Data* (9th), *Selective Data* (10th), and *Dubious Data* (11th). These issues can also be confusing in that the readers cannot get any information from the visualizations: *Missing Title* (3rd), *Missing Axis Title* (6th), *Missing Legend* (8th), *Missing Value La-*

Stages	Categories	Tags	Bar(574)	Line(416)	Circle(193)	Area(163)	Map(121)	Point(53)	Text(2)
Input (262)	Garbage-in (152)	Not Data (50)	1	26					
		Selective Data (40)	6	22	1	10			1
		Dubious Data (37)	14	16	4	2	1		
		Non Sequitur (24)	9	3	7	3			1
		Too Few Data Points (1)		1					
	Lie with Statistics (110)	Discretized Continuous Variable (28)	3				24	1	
		Missing Normalization (26)	10	9			7		
		Inappropriate Item Order (23)	20		3				
		Inappropriate Metric (9)	2	3		1	1	2	
		Questionable Prediction (9)	1	5		2		1	
Trend Line on Random Data (7)			1					6	
Visualization Design (459)	Choice of Axis (324)	Inappropriate Use of Accumulation (5)	2	2		1		1	
		Inappropriate Aggregation Granularity (2)					1	1	
		Two-Way Normalization (1)	1						
		Truncated Axis (171)	155	7		9			
		Dual Axis (78)	20	58					
		Inappropriate Axis Range (32)	5	25		2			
		Inverted Axis (19)	2	3		12	1	1	
	Choice of Chart (71)	Log Scale (9)	2	7					
		Extended Axis (7)	4			3			
		Data of Different Magnitudes (6)	3	3					
Color Mess (64)	Linear Scale on Exponential Data (2)		1		1				
	Overusing Colors (32)	6	14	9	1	1	1		
	Indistinguishable Colors (31)	5	8	9	5	3	1		
	Color Blind Unfriendly (1)			1					
	Plotting (525)	Incomplete Chart (286)	Inappropriate Use of Line Chart (23)	3	20				
			Inappropriate Use of Pie Chart (17)			17			
			Confusing Chart Type (16)	5	4	1	1	4	1
			Misusing Circular Layout (7)	7					
			Inappropriate Use of Stacked (5)	4			1		
			Inappropriate Use of Bar Chart (2)	2					
Inappropriate Use of Scatterplot (1)								1	
Inconsistency (159)		Overusing Colors (32)	6	14	9	1	1	1	
		Indistinguishable Colors (31)	5	8	9	5	3	1	
		Color Blind Unfriendly (1)			1				
Chaotic Canvas (80)	Missing Title (94)	36	22	14	6	12	4		
	Missing Axis Title (62)	33	23		4				
	Missing Legend (52)	7	7	6	6	15	10		
	Missing Value Labels (35)	2		29	3	1			
	Missing Axis (25)	14	9		2				
	Missing Axis Ticks (17)	6	6		4		1		
	Missing Units (1)	1							
	Misrepresentation (65)	Missing Units (1)	1						
		Misrepresentation (65)	38	2	13	11	1		
		Inconsistent Tick Intervals (52)	13	31		6		2	
Inconsistent Binning Size (25)		12	3	3		6	1		
Changing Scale (10)		4	5		1				
Violating Color Convention (4)		2	1	1					
Inconsistent Grouping (1)				1					
Perception (198)	Inconsistent Tick Labels (1)		1						
	Inconsistent Value Labels (1)				1				
	Cluttering (19)		17	1			1		
	Confusing Legend (16)	6	2	1			7		
	Plotting Error (14)	7	2	3	2				
	Missing Abbreviation (10)	8	2						
	Plotting Out of Chart (8)	4	3		1				
Visual Illusion (198)	Misalignment (7)	7							
	Illegible Text (6)	1	2	3					
	3D (105)	42	5	50	4	3	1		
	Area Encoding (35)	2	1	5	17	3	7		
	Ineffective Color Scheme (28)	5		2		20	1		
	Pictorial Area Encoding (19)				19				
	Inappropriate Use of Smoothing (4)		4						
	Distractive Value Labels (3)	3							
	Map Projection Distortion (2)					2			
	Inappropriate Aspect Ratio (1)		1						
False Linkage (28)	Sine Illusion (1)		1						
	Invalid Comparison (16)	8	8						
	Correlation Not Causation (7)		6				1		
	Pattern Seeking (5)		5						
	Power of Words (28)	Misleading Claim (17)	7	5		1	2	2	
		Misleading Annotation (7)	5	1		1			
		Misleading Title (3)			1	1	1		
		Misleading Value Labels (1)			1				
	Hidden Information (22)	Hidden Distribution (9)	4					5	
		Overplotting (8)	1			7			
Hidden Uncertainty (4)			3	1					
Hidden Population Size (1)		1							

Figure 1: Taxonomy of the 74 issues discovered in the 12 categories, which are further grouped into the five analytics stages. The tag descriptions can be found in the Appendix.

blems (13th), *Inappropriate Axis Range* (14th), and *Overusing Colors* (15th). These issues are not only related to the perceptual understanding of a chart, like *3D* or *Area Encoding*, but also involve a wider spectrum of causes, like input data, missing chart elements, axis choice, and scalability. We have discovered 74 issues, but they are not exhaustive to the issue space. Putting the issues into a taxonomy will help us learn more about the issue space and understand what causes visualizations to be misinformative or uninformative.

5 Taxonomy

Our first intuitive attempt at axial coding was to group issues by their concerned chart elements. For example, *Dual Axis* concerns the axis, hence it is grouped with other issues related to the axis, such as *Truncated Axis*, *Missing Axis*, and *Inconsistent Tick Intervals*. Following the same approach, the issues related to data are grouped together. For example, the *Data of Different Magnitudes* issue is grouped with *Selective Data*. These groups are related to (1) chart elements: axis, color, legend, and annotation, (2) grammar of graphics [Wic10]: data, statistical transformation, and aesthetic mapping, and (3) other: plotting and message. Although the elements of these groups share similarities, they do not help to clarify why a visualization misleads or when the misleading element was introduced.

Our second attempt was guided by the notes taken during the coding process when examining the visualizations and the associated social media posts or web pages. One of the visualizations examined is associated with a blog post explaining the dual axis that states “Dual axes time series plots may be ok sometimes after all” [Eli16]. The *Dual Axis* issue is highly related to the issue *Data of Different Magnitudes*, where a *Dual Axis* is trying to solve the dilemma of plotting two data series of different magnitudes. Take Figure 3g as an example. When forcing the two series to be plotted on the same y-axis, the smaller magnitude series gets dominated by the larger one, and the reader can barely see it exists. Instead of grouping *Data of Different Magnitudes* with other data-related issues, it is associated with other issues concerning the **Choice of Axes**. This grouping gives us a clearer picture of why a visualization becomes misleading. Other decisions on the **Choice of Axes** that resulted in misleading visualizations include *Inappropriate Axis Range*, *Log Scale*, *Linear Scale on Exponential Data*, and the trio *Truncated/Extended/Inverted Axis*. Figure 3 shows the example visualizations for this category, and a description of these tags can be found in the Appendix.

Missing Axis, *Missing Axis Title*, and *Missing Axis Ticks* are related to the axis, but they are not design decisions. Putting them alongside *Truncated Axis* or *Log Scale* would blur the picture rather than clarify it. Instead, we put all issues related to missing parts of the chart into the category **Incomplete Chart**. This grouping provides a better foundation to develop theories upon. We can further fit these categories into a five-stage visual analytics process modified from McNutt *et al.* [MKC20]. These stages are (1) input stage, (2) visualization design stage, (3) plotting stage, (4) perception stage, and (5) interpretation stage.

5.1 Input Stage: Data Curation and Wrangling

The **Garbage-in** category borrows the idiom “garbage in, garbage out” to describe the issues related to inputting faulty data into the

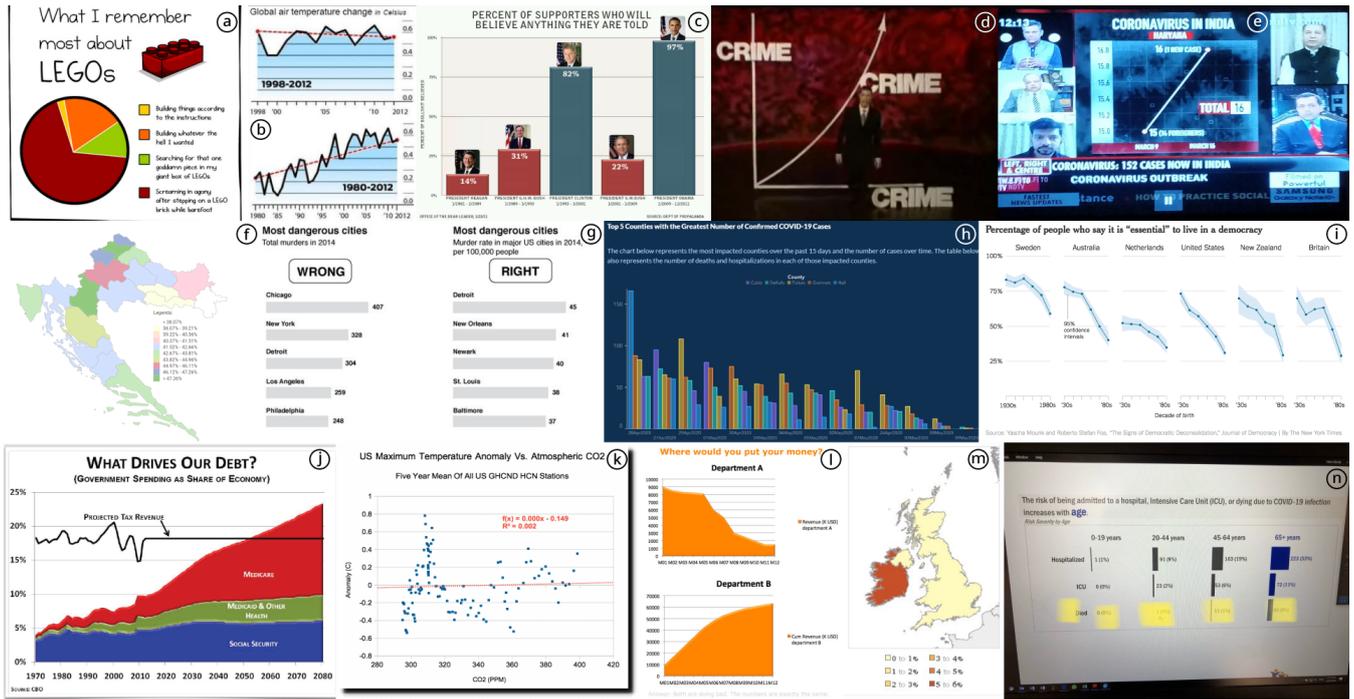


Figure 2: Issues in the Input Stage. These issues are (a) Not Data, (b) Selective Data, (c) Dubious Data, (d) Non Sequitur, (e) Too Few Data Points, (f) Discretized Continuous Variable, (g) Missing Normalization, (h) Inappropriate Item Order, (i) Inappropriate Metric, (j) Questionable Prediction, (k) Trend Line on Random Data, (l) Inappropriate Use of Accumulation, (m) Inappropriate Aggregation Granularity, and (n) Two-way Normalization. See the Appendix for an explanation of each visualization, accompanied by a high-resolution version of the image.

process, which means outputting a faulty visualization is unavoidable. *Not Data* are visualizations that are drawn to illustrate concepts rather than visualizing data. Readers may mistake these as being drawn according to actual data, but this is not the intention. Similarly, readers expect *Non Sequitur* visualizations to reveal some information from the data, but they do not. They are mostly created for decorative purposes. On the other hand, even when the visualization is visualizing data, it can be based on *Selective Data*, also known as cherry-picking [Kla08]. The visualization selects the partial truth to tell and hides any inconvenient elements. These pickings can be picking a specific period, specific comparison opponents, or selecting a particular set of data points. *Dubious Data* are reported cases that cast doubt on the underlying data or data source. *Too Few Data Points* is an issue that arises when connecting two dots with a straight line to form a line chart.

The **Lie with Statistics** category is statistical transformations that manipulate the data into shapes that convey the intended messages. Without the purpose of grouping data points, a *Discretized Continuous Variable* unnecessarily becomes a categorical variable, and the values near the boundaries of the splits become prominently different. A typical example is applying categorical rainbow colors to continuous variables when sequential colors can be used instead. This exaggerates the boundary cases. Take the color scheme in Figure 2f as an example, when 46.11% is red and 46.12% is blue, that 0.01% difference becomes very significant, which is misleading. The common issue of population maps is an example of *Missing Normalization*. It shows meaningless comparisons when the numbers are in absolute values instead of per capita. *Two-way*

Normalization does not miss out on normalization, but arbitrarily interprets a two-way table by dividing the numbers by their column or row total [Cai19]. *Inappropriate Item Order* can hide patterns or show misleading trends by putting the items in an arbitrary order, like Figure 2h. *Inappropriate Metric* is a carefully picked metric, threshold, or definition of grouping to support unconvincing claims. *Trend Line on Random Data* is a regression line on weakly correlated data that misleads people into thinking there is a trend. *Questionable Prediction* extrapolates the current trend beyond the present time without stating how the projection is made. *Inappropriate Use of Accumulation* can hide a recent declining trend by showing the accumulated historical values. *Inappropriate Aggregation Granularity* is similar to applying an inappropriate magnification level on a microscope, both of which lead to the wrong conclusion or false findings.

5.2 Visualization Design Stage: Choices in Visual Encoding

Choice of Axis is the most common visualization design decision. While visualizing two series with different magnitudes in one chart, a decision is made on the choice of the axis. *Dual Axis* (Figure 3b) and *Data of Different Magnitudes* (Figure 3g) have the same underlying problem: two series with different magnitudes. The series with a larger magnitude will dominate the one with a smaller magnitude, and the reader can hardly see its changes. While plotting the two series with two different axes, the chart becomes misleading if the reader is unaware. Even when we fix the problem to only one axis, choosing an appropriate axis range is not easy [Wit19, CBF20]. Choosing an *Inappropriate Axis Range* (Figure 3c) has the same over-exaggerated or understated result as *Trun-*

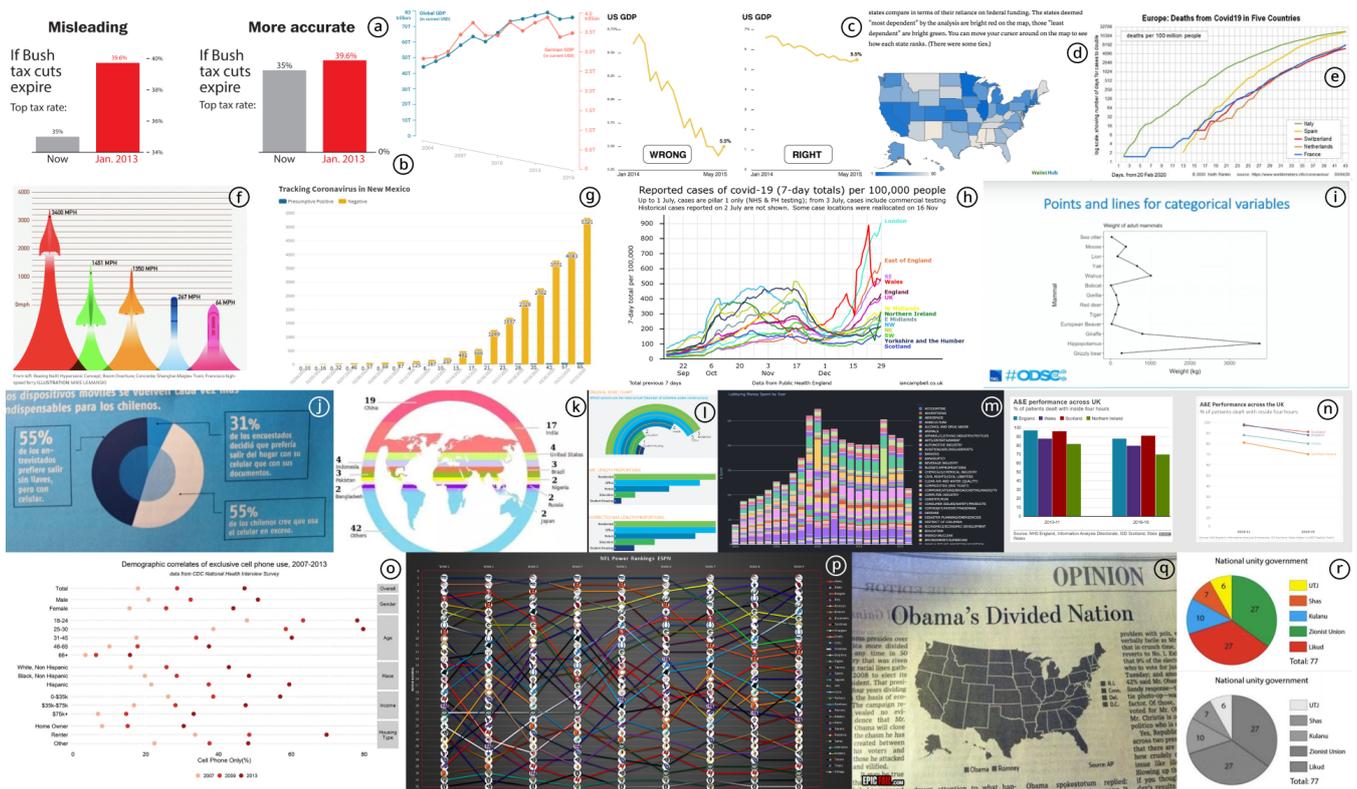


Figure 3: Issues in the Visualization Design Stage. These issues are (a) Truncated Axis, (b) Dual Axis, (c) Inappropriate Axis Range, (d) Inverted Axis, (e) Log Scale, (f) Extended Axis, (g) Data of Different Magnitudes, (h) Linear Scale on Exponential Data, (i) Inappropriate Use of Line Chart, (j) Inappropriate Use of Pie Chart, (k) Confusing Chart Type, (l) Misusing Circular Layout, (m) Inappropriate Use of Stacked, (n) Inappropriate Use of Bar Chart, (o) Inappropriate Use of Scatterplot, (p) Overusing Colors, (q) Indistinguishable Colors, and (r) Color Blind Unfriendly. See the Appendix for an explanation of each visualization, accompanied by a high-resolution version of the image.

cated/Extended Axis (Figure 3a and Figure 3f). An *Inverted Axis* is a decision on the axis direction. Conventionally, the axis develops from smaller numbers, moving upward to larger numbers, but this is not the case when the number represents a ranking. Readers expect smaller numbers at the top instead of at the bottom. The same principle applies to color saturation and brightness. Figure 3d is a choropleth map with an inverted color scheme on ranking data. The solid blue represents the least dependent states, and grey—a less saturated color—represents the most dependent states while the readers expect the opposite. *Log Scale* and *Linear Scale on Exponential Data* are two sides of the same coin, so when should you apply log scale and when should you not? Again, this relates to the convention, visualization literacy, and the readers' expectation. Plotting exponential growth data on a linear scale shows fewer meaningful patterns, but a log scale can be misleading if the readers focus only on the curve.

The common cause of inappropriate **Choice of Chart** is applying the wrong chart type to unsuitable data. Putting categorical data on axes or swapping the x and y-axis are *Inappropriate Use of Line Chart* (Figure 3i). Plotting a pie chart on data that has no part-to-whole relationship is the most common cause of *Inappropriate Use of Pie Chart*. *Inappropriate Use of Stacked* is when too many layers are stacked, making it incomprehensible for the reader. *Misusing Circular Layout* attempts to use curved bars in a comparison

or render length-encoded lines in curves, resulting in the readers getting an inaccurate comparison. A chart has a *Confusing Chart Type* because it partly belongs to a common chart but also deviates from it. These charts are hard to decode or understand and leave readers in confusion. Similarly, readers cannot comprehend the *Inappropriate Use of Bar Chart and Scatterplots*.

When colors in the visualization are indistinguishable, it is a **Color Mess**, and the reader cannot get information from the visualizations. *Overusing Colors* in a visualization can only overwhelm the reader, and the visualization becomes hideous to look at. *Indistinguishable Colors*, on the other hand, when the same or very similar colors are used for different categories, causing the readers to mistake one category for another. For color blind people, a *Color Blind Unfriendly* visualization is similar to indistinguishable colors.

5.3 Plotting Stage: Drawing on the Canvas

A chart may still be readable with minor chart elements missing, but an **Incomplete Chart** is incomprehensible when missing one or more elements that are crucial to the understanding of the meaning of the chart. A visualization with a *Missing Title* keeps the reader wondering what the visualization is about and what message it is intended to convey. *Missing Axis Title* makes the meaning of the axis ambiguous. *Missing Axis* and *Missing Legend* can be compensated for with labels on the visual marks. But, when all of these

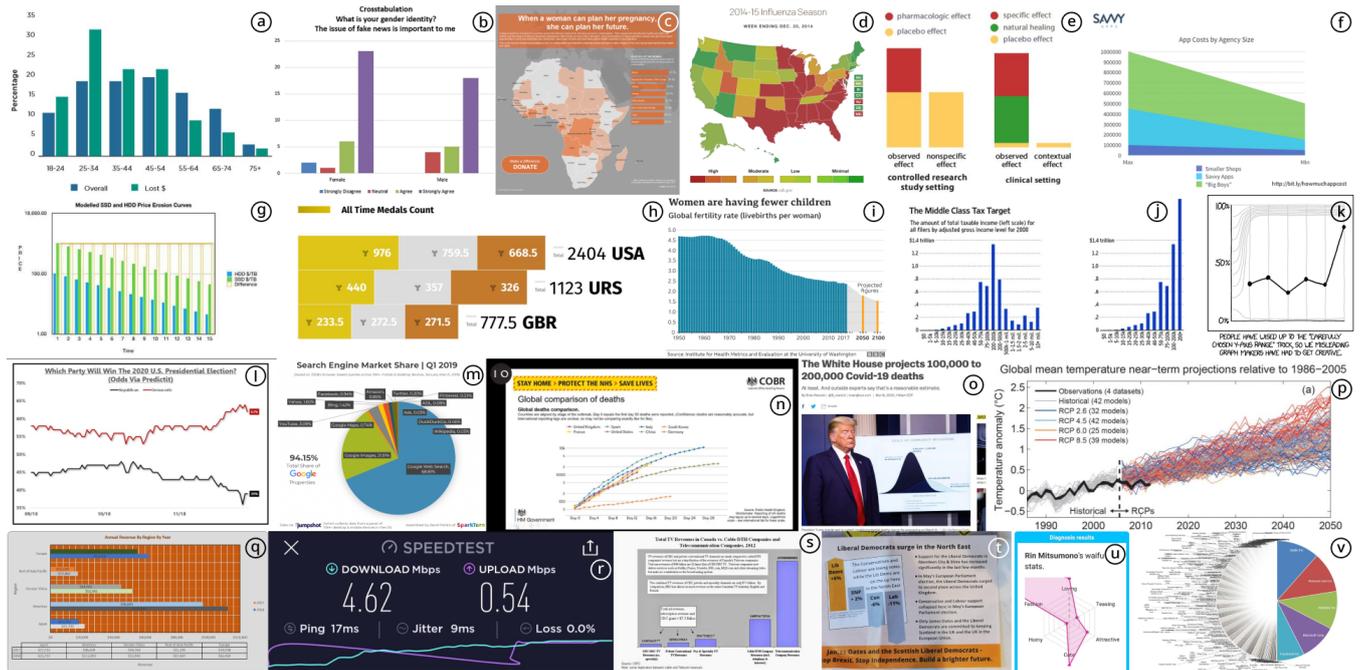


Figure 4: Issues in the Plotting Stage. These issues are (a) Missing Title, (b) Missing Axis Title, (c) Missing Legend, (d) Missing Value Labels, (e) Missing Axis, (f) Missing Axis Ticks, (g) Missing Units, (h) Misrepresentation, (i) Inconsistent Tick Intervals, (j) Inconsistent Binning Size, (k) Changing Scale, (l) Violating Color Convention, (m) Inconsistent Grouping, (n) Inconsistent Tick Labels, (o) Inconsistent Value Labels, (p) Cluttering, (q) Confusing Legend, (r) Plotting Error, (s) Missing Abbreviation, (t) Misalignment, (u) Plotting Out of Chart, (v) Illegible Text. See the Appendix for an explanation of each visualization, accompanied by a high-resolution version of the image.

elements are missing, readers cannot get any information from the visualization (Figure 4d).

Inconsistency occurs in arbitrarily drawn charts that break our conventional understanding of the concepts. Readers expect the charts to be drawn according to our shared understanding of the world. For example, drawing to scale, following convention, and using the same units across different values. *Misrepresentation* is when data values are drawn disproportionately or not to scale. It is commonly spotted as value labels that fail to match the visually encoded geometric objects. We expect the tick intervals in a chart to be consistent, but charts with *Inconsistent Tick Intervals* violate this expectation and show false patterns. Malicious visualization authors can use this trick to create a distorted visual perception. *Inconsistent Binning Size* means varying the boundaries of the binning groups to put more or fewer items into specific groups, resulting in a manipulated statistics calculation. *Changing Scale* means the scale changes midway. It can be treated as a special case of inconsistent tick intervals. The tick intervals change in the later part of the axis with or without any visual cues. *Violating Color Convention* is a color mismatch between colors and the category they represent. *Inconsistent Tick Labels* are tick labels that are not in the same format. *Inconsistent Value Labels* are the labels that are inconsistently annotated. *Inconsistent Grouping* is when some entities are grouped while others are not. It makes the dominating entity look less dominant by splitting it into smaller members while the other groups remain grouped.

Visualizations with a **Chaotic Canvas** are unreadable. The rendered visualization is glitchy or messed up. *Cluttering* is a scalability

issue related to too many data points or series clumping together. A chart with a *Confusing Legend* is difficult to comprehend its encoding or incomprehensible at all. A *Plotting Error* is a glitch in a chart as a result of software bugs or other reasons. *Plotting Out Of Chart* means plotting a data value that goes beyond the axis range, therefore, its geometric mark is out of the chart area. *Misalignment* of the items or labels results in poor readability, making it hard to perform a comparison. *Missing Abbreviation* makes the text unnecessarily long, and *Illegible Text* is the text that overlaps itself or other text and becomes unreadable.

5.4 Perception Stage: Visually Perceiving the Visualization

The only category in this stage is **Visual Illusion**. The visualization is drawn to scale, but we perceive it differently. *Unjustified 3D* is a perspective distortion technique. The closer it is, the larger it looks, despite being the same size in the 3D perspective. The closer objects are perceived as more prominent, although they may be the same or smaller. It allows more ink for smaller objects in 3D bar charts. *Area Encoding and Pictorial Area Encoding* are not linear encodings. According to Stevens' power law, the exponent for area is ~0.7, which means linearly encoding data values as areas leads the readers to consistently underestimate the values [Ste75]. *Ineffective Color Scheme* are rainbow colors, categorical colors on sequential data, and sequential colors on categorical data. Most are related to colors used on a choropleth. Dichotomy colors can hide the underlying distribution (Figure 5q). A light red (50.1%) can conceal all the blue (49.9%), which is most commonly seen in choropleth maps concerning political voting. The commonly used Mercator projection suffers from *Map Projection Distortion*, and com-

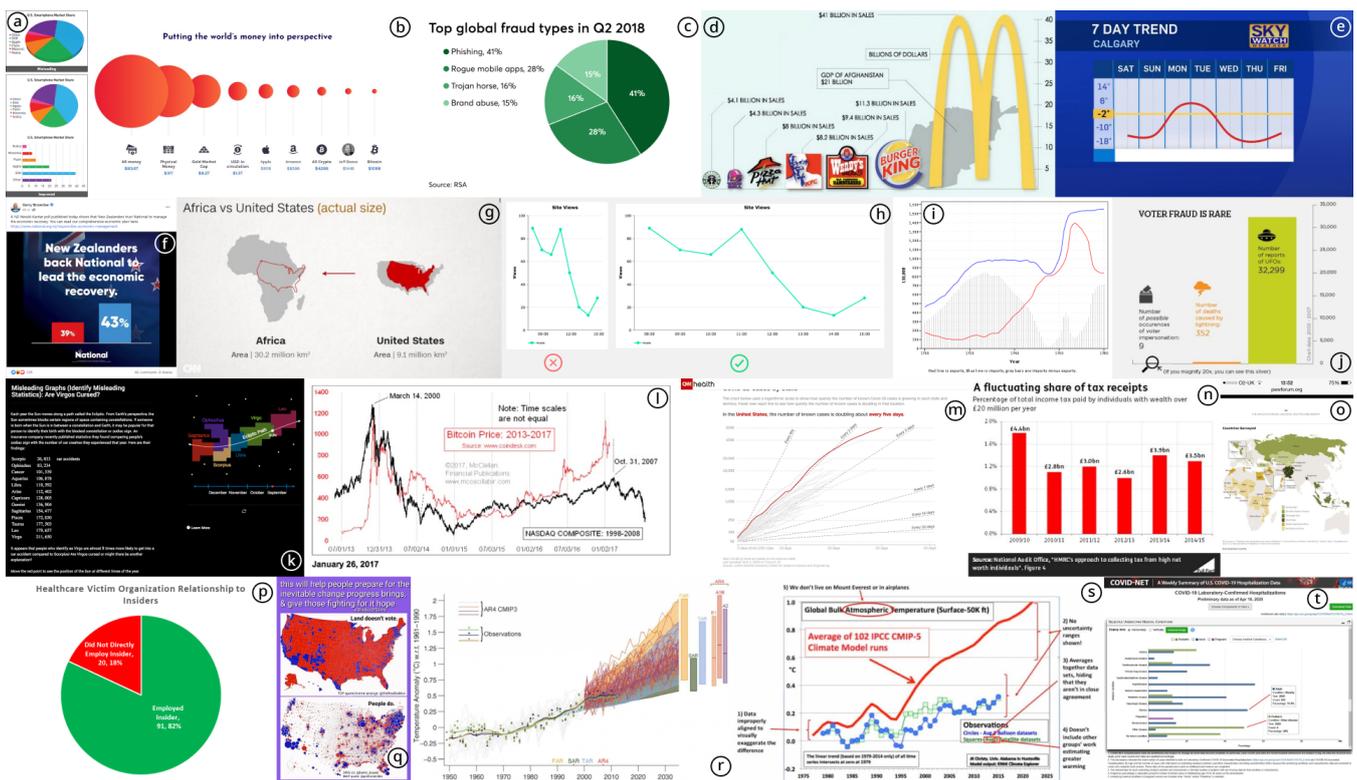


Figure 5: Issues in the Perception and Interpretation Stage. These issues are (a) 3D, (b) Area Encoding, (c) Ineffective Color Scheme, (d) Pictorial Area Encoding, (e) Inappropriate Use of Smoothing, (f) Distractive Value Labels, (g) Map Projection Distortion, (h) Inappropriate Aspect Ratio, (i) Sine Illusion, (j) Invalid Comparison, (k) Correlation Not Causation, (l) Pattern Seeking, (m) Misleading Claim, (n) Misleading Annotation, (o) Misleading Title, (p) Misleading Value Labels, (q) Hidden Distribution, (r) Overplotting, (s) Hidden Uncertainty, (t) Hidden Population Size. See the Appendix for an explanation of each visualization, accompanied by a high-resolution version of the image.

paring the area is inaccurate on these 2D maps. However, we did not find any deliberately deceptive cases. *Distractive Value Labels* are the value labels that affect perceived data size. The size of value labels provides visual cues that even dominate the data-encoded geometric objects. *Inappropriate Use of Smoothing* is the interpolation between data points. It creates non-existing data points. *Sine Illusion* is one of the 143 optical illusions [BP06]. There may be a lot of these going unspotted because most people are not aware of them. They are hard to detect by the human eyes but are perhaps easier for machines to spot. We came across a blog post in the coding process that discusses why the middle series of stacked area charts are consistently underestimated [Zei10]. Both line chart and stacked area chart are suspected of this illusion. An *Inappropriate Aspect Ratio* is commonly seen in line charts, but it can also be applied to other chart types [CMXF20].

5.5 Interpretation Stage: Comprehending the Message

When visualization is being used for communication, we need to remain skeptical about its message. It can put two or more things together and suggest a **False Linkage**. It can also use the **Power of Words** to mislead the interpretation of the visualization. Or, you may interpret something differently if it keeps any crucial **Hidden Information** from you.

False Linkage is putting unrelated or incomparable items

together. *Invalid Comparison* is like comparing apples to oranges by linking unrelated items by comparison or incomparable units/normalization. It is used as a red herring, like comparing the chances of voter fraud (Figure 5j). *Correlation Not Causation* is trying to lead the reader to draw a causality conclusion by falsely linking events to explain the data and give a cause to the observed data. *Pattern Seeking* is to seek a correlation from historical data to try to predict what happened in the past will happen in the future just because the pattern matches.

Power of Words is when the text description on the chart does not match the message conveyed by the chart. The chart may be faithfully plotted, but the text is misleading. *Misleading Claim* is the use of a chart to make a misleading claim by their interpretation through text description. *Misleading Annotation* is when the annotation in the chart is misleading. *Misleading Title* is when the title does not match the message in the chart. Figure 5p has *Misleading Value Labels* that the interpretation at first glance is a pie chart with sectors 91.82% and 20.18%.

Hidden Information is when some critical information that is not being shown on the chart leads to the drawing of incorrect conclusions or being used to support a misleading claim. *Hidden Distribution* is an underlying distribution that is not shown in the visualization, causing a misconception of the actual picture. Charts

with *Overplotting* issues have visual marks in the same position. The marks that are being overlapped are hidden from the readers, resulting in understated areas (Figure 5r). *Hidden Uncertainty* is the case that uncertainty is not being visually represented when it matters. *Hidden Population Size* is when the population size matters to the statistical summary but is not shown.

6 What Have We Learned From Misinformative Visualizations?

Throughout the process of coding and forming the taxonomy, we found four promising future research directions to help better understand misleading visualizations.

Understanding the designers' dilemma. In some misleading cases, we found that the visualization authors have already tried their best not to mislead the reader by providing visual cues to mitigate any misleading effect. Yet, the resulting visualization is still misleading. Given a data series mixed with dense historical data and sparse projected data, despite the best efforts of the author to avoid misleading the readers by providing visual cues, the chart with inconsistent tick intervals in Figure 4i still misleads readers. A similar difficulty arises when the binning or intervals, like age or income groups, are unevenly set during the data collection process (Figure 4j). The case of dual axis has the root cause of visualizing two data series of different magnitudes in one chart. The smaller magnitude series is dominated when both series are put under a unified axis scale (Figure 3g). The discussions of these issues are valuable guidance for practitioners [Eli16, Mut18]. Besides the problems inherited from the data, the limitations may come from the presentation medium. Projecting the spherical Earth onto a 2D world map means distortion is unavoidable (Figure 5g). Finding and understanding these dilemmas in visualization designs is an important step toward developing practical design guidelines. The great design of railway maps is possible only when the designers realize the constraints they are facing [Ven12].

The role of conventions and data literacy education. Breaking convention is a major source of misperception [CH17]. The visualizations we encountered during the coding process reminded us of the convention changes from culture to culture and time to time. A good example of such a case is the use of log scales in charts concerning COVID-19 cases. At the beginning of the outbreak, people found the log scale misleading because it flattened the curve. However, to people who had learned to read log scale charts, not using log scale is also misleading due to its exponential nature (Figure 3h). The discussion of breaking conventions is always fuzzy because we do not have an agreed set of conventions [CE20]. It becomes even trickier when considering its shifting nature across cultures and time. Data literacy education plays a vital role in shaping our expectations on how to read visualizations. Like the log scale example, people who learned to read log scale charts expect exponential data to be plotted in a log scale, while others find log scale misleading rather than informative. Clarifying the existing conventions and their variations across cultures is essential to progressing our discussion on broken conventions. The same importance lies in shaping the conventions through data literacy education.

The unexplored space of informal fallacies in visualization. Data visualizations are seldom standalone or self-explanatory. They

are often used in the context of arguments, and informal fallacies in arguments also apply to visualizations. It happens in the input and interpretation stages. Invalid input is fated to result in a false visualization. Selectively choosing the most favorable data to visualize falls into the cherry-picking fallacy (Figure 2b). Even with valid data, Figure 2i shows how to create a straw man by carefully crafting a metric to make a visualization supporting a misleading claim. When interpreting a voting result visualization, misinterpreting the territorial area as population creates an illusion of a landslide victory (Figure 5q), which falls into red herring fallacy. Other fallacies are the false analogy fallacy (as *Invalid Comparison* in Figure 5j), post-hoc fallacy (as *Correlation Not Causation* in Figure 5k), and slippery slope fallacy (as *Hidden Uncertainty* in Figure 5s). Informal fallacies in visualizations have very different representations and are hard to spot, so we suspect more—if not all—of the informal fallacies are applicable to visualizations.

Does there exist a type of chart that is immune to misleading the audience? There are chart types that do not apply to certain types of issues. For example, axis-related issues do not apply to pie charts because there are no axes in pie charts. We wonder if a chart type is immune to misleading tricks. Conversely, can we always make a chart misleading? From our collected samples, we have only found two cases of uninformative word clouds (*Non Sequitur* and *Missing Legend*). Still, we did not find any reported instances of misleading word clouds. Word cloud seems to be a candidate that is immune to misleading the audiences. Unfortunately, a previous synthetic study has shown that word clouds can be injected with manipulated data without it being spotted [WCHB10]. We are afraid that no chart is immune to misleading the audience. Instead, some are stealthier and are rarely spotted, which is even more dangerous. This brought our attention to similar cases of sine illusion and distorted map projections. Both cases are more prevalent than expected as they were not pointed out and discussed as frequently as others. Finally, instead of designing a type of chart that is immune to misleading the audience, we must rely on the authoring process to better prevent misleading elements from sneaking into the visualizations. That is, fortifying each stage of the visualization pipeline, validating and verifying the produced visualizations. The preliminary work on visualization linter [MK18, ZS19, MKC20, HCS20, CSX*21] is a promising direction, and this work completes the missing link between the linters and real-world examples [HCS20].

7 Conclusion

Visualization has persuasive power, and this power can be used to guide people in the wrong direction. This work aims to enrich the discussion on misleading visualizations. From the collected dataset and the taxonomy, we found a large issue space that leads to the resulting uninformative or misleading visualization. The ever-increasing use of visualization in public discourse as part of arguments also falls into informal fallacies. On one side, educating the general public on better data and visualization literacy is an urgent matter that is also welcomed by students [BW17]. On the other side, expanding the discussion scope will help us learn more about the nature of misleading visualizations. We will continue our research in this area, opening up the aforementioned research directions and developing preventative measures for countering misleading visualizations.

References

- [BE15] BRESCIANI S., EPPLER M. J.: The pitfalls of visual representations: A review and classification of common errors made while designing and interpreting visualizations. *Sage Open* 5, 4 (2015), 2158244015611451. 3
- [BP06] BACH M., POLOSCHEK C. M.: Optical illusions. *Adv Clin Neurosci Rehabil* 6, 2 (2006), 20–21. 3, 9
- [BVB*13] BORKIN M. A., VO A. A., BYLINSKII Z., ISOLA P., SUNKAVALLI S., OLIVA A., PFISTER H.: What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2306–2315. 3
- [BW17] BERGSTROM C. T., WEST J.: Calling bullshit in the age of big data. *Calling Bullshits. Data Reasoning in a Digital World. Available online at: <http://callingbullshit.org>* (2017). 10
- [Cai15] CAIRO A.: Graphics lies, misleading visuals. In *New challenges for data design*. Springer, 2015, pp. 103–116. 1, 3
- [Cai19] CAIRO A.: *How Charts Lie: Getting Smarter about Visual Information*. WW Norton & Company, 2019. 2, 6
- [CBF20] CORRELL M., BERTINI E., FRANCONERI S.: Truncating the y-axis: Threat or menace? In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–12. 6
- [CE20] CHEN M., EDWARDS D. J.: “isms” in visualization. In *Foundations of Data Visualization*. Springer, 2020, pp. 225–241. 10
- [CH17] CORRELL M., HEER J.: Black hat visualization. In *Workshop on Dealing with Cognitive Biases in Visualisations (DECISiVe)*, *IEEE VIS* (2017). 1, 3, 10
- [CMXF20] CEJA C. R., MCCOLEMAN C. M., XIONG C., FRANCONERI S. L.: Truth or square: Aspect ratio biases recall of position encodings. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 1054–1062. 9
- [CSX*21] CHEN Q., SUN F., XU X., CHEN Z., WANG J., CAO N.: Vizlinter: A linter and fixer framework for data visualization. *IEEE transactions on visualization and computer graphics* 28, 1 (2021), 206–216. 10
- [DFP*18] DIMARA E., FRANCONERI S., PLAISANT C., BEZERIANOS A., DRAGICEVIC P.: A task-based taxonomy of cognitive biases for information visualization. *IEEE Transactions on Visualization and Computer Graphics* 26, 2 (2018), 1413–1432. 3
- [Eli16] ELLIS P.: Dual axes time series plots may be ok sometimes after all, Aug 2016. URL: <http://freerangestats.info/blog/2016/08/18/dualaxes>. 5, 10
- [HCS20] HOPKINS A. K., CORRELL M., SATYANARAYAN A.: Visualint: Sketchy in situ annotations of chart construction errors. *Computer Graphics Forum* 39, 3 (2020), 219–228. 1, 3, 10
- [Huf54] HUFF D.: *How to lie with statistics*. WW Norton & Company, 1954. 1, 2
- [Key04] KEYES R.: *The post-truth era: Dishonesty and deception in contemporary life*. Macmillan, 2004. 2
- [Kla08] KLASS G.: Just plain data analysis: Common statistical fallacies in analyses of social indicator data. *Statlit.org* (2008). 6
- [LO20] LAUER C., O'BRIEN S.: The deceptive potential of common design tactics used in data visualizations. In *Proceedings of the 38th ACM International Conference on Design of Communication* (2020), pp. 1–9. 1, 3
- [LYI*21] LEE C., YANG T., INCHOCO G. D., JONES G. M., SATYANARAYAN A.: Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), pp. 1–18. 2, 4
- [MK18] MCNUTT A., KINDLMANN G.: Linting for visualization: Towards a practical automated visualization guidance system. In *Vis-Guides: 2nd Workshop on the Creation, Curation, Critique and Conditioning of Principles and Guidelines in Visualization* (2018). 1, 3, 10
- [MKC20] MCNUTT A., KINDLMANN G., CORRELL M.: Surfacing visualization mirages. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–16. 3, 5, 10
- [Mon91] MONMONIER M.: *How to lie with maps*. University of Chicago Press, 1991. 2
- [Mul14] MULLER M.: Curiosity, creativity, and surprise as analytic tools: Grounded theory method. In *Ways of Knowing in HCI*. Springer, 2014, pp. 25–48. 3
- [Mut18] MUTH L. C.: Why not to use two axes, and what to use instead, May 2018. URL: <https://blog.datawrapper.de/dualaxis/>. 10
- [MWN*18] MORITZ D., WANG C., NELSON G. L., LIN H., SMITH A. M., HOWE B., HEER J.: Formalizing visualization design knowledge as constraints: Actionable and extensible models in draco. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 438–448. 1, 3
- [PMN*14] PANDEY A. V., MANIVANNAN A., NOV O., SATTERTHWAITHE M., BERTINI E.: The persuasive power of data visualization. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 2211–2220. 1
- [PRS*15] PANDEY A. V., RALL K., SATTERTHWAITHE M. L., NOV O., BERTINI E.: How deceptive are deceptive visualizations? an empirical analysis of common distortion techniques. In *Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems* (2015), pp. 1469–1478. 1, 3
- [Rya14] RYAN S.: *Charts Don't Lie: 4 Untold Trading Indicators and How to Make Money with Them*. CreateSpace Independent Publishing Platform, 2014. 4
- [Ste75] STEVENS S.: Psychophysics: Introduction to its perceptual, neural, and social prospects. 8
- [Sza18] SZAFIR D. A.: The good, the bad, and the biased: Five ways visualizations can mislead (and how to fix them). *Interactions* 25, 4 (2018), 26–33. 1, 3
- [TGM83] TUFTE E. R., GRAVES-MORRIS P. R.: *The visual display of quantitative information*. Graphics press Cheshire, CT, 1983. 2
- [TS08] THALER R. H., SUNSTEIN C. R.: Nudge: improving decisions about health. *Wealth, and Happiness* 6 (2008), 14–38. 1
- [Ven12] VENETIKIDIS A.: Making sense of maps, Sep 2012. URL: https://www.ted.com/talks/aris_venetikidis_making_sense_of_maps. 10
- [WCHB10] WICKHAM H., COOK D., HOFMANN H., BUJA A.: Graphical inference for infovis. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 973–979. 10
- [WD17] WARDLE C., DERAKHSHAN H.: Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe* 27 (2017). 2, 4
- [Wic10] WICKHAM H.: A layered grammar of graphics. *Journal of Computational and Graphical Statistics* 19, 1 (2010), 3–28. 5
- [Wit19] WITT J. K.: Graph construction: an empirical investigation on setting the range of the y-axis. *Meta-psychology* (2019). 6
- [Zei10] ZEIGEN: Lying with charts 102: Deceptions of stack, Aug 2010. URL: <https://web.archive.org/web/20130322115008/http://www.zeigen.com/blog/2010/08/lying-charts-102/>. 9
- [ZS19] ZHENG W., SHERIF Y.: Visualization linter. *Course Project* (2019). 1, 3, 10