

V-FRAMER: Visualization Framework for Mitigating Reasoning Errors in Public Policy

Lily W. Ge
Northwestern University
Evanston, Illinois, USA
wanqian.ge@northwestern.edu

Matthew Easterday
Northwestern University
Evanston, Illinois, USA
easterday@northwestern.edu

Matthew Kay
Northwestern University
Evanston, Illinois, USA
mjskay@northwestern.edu

Evanthia Dimara
Utrecht University
Utrecht, Netherlands
evanthia.dimara@gmail.com

Peter Cheng
University of Sussex
Brighton, United Kingdom
p.c.h.cheng@sussex.ac.uk

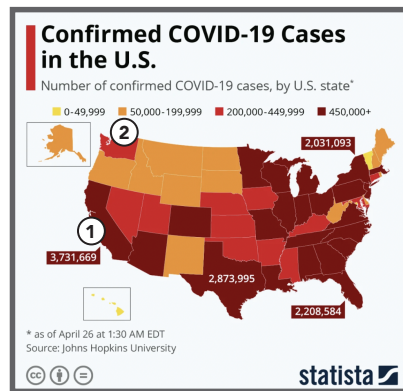
Steven L Franconeri
Northwestern University
Evanston, Illinois, USA
franconeri@northwestern.edu

Question: Did California (①) have a higher risk of COVID compared to Washington (②) around April, 2021?

Is the problem size different for subgroups?

Wrong answer: Yes

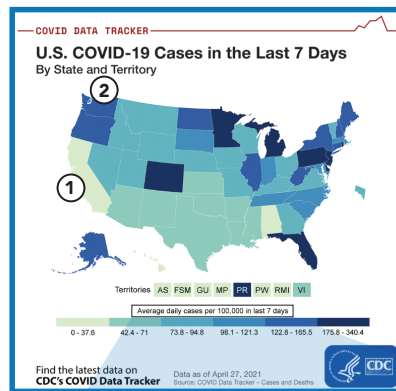
This map visualization is potentially misleading because it is **missing normalization** when making relative risk comparisons.



(A)

Correct answer: No

This map better represents relative risk by **normalizing it by population**.



(B)

Average daily cases per 100,000 in last 7 days

Figure 1: An example showing how a visualization could still lead to inaccurate conclusions about the data despite the visual components being grammatically correct (i.e., no visual distortions on scales, marks, or channels). To answer the policy-relevant question *Did California have a higher risk of COVID compared to Washington around April, 2021?*, one needs to consider the *Comparison Basis* and normalize COVID cases by population (real-world example map B [7, 68]). Not considering the comparison basis could lead to issues such as *Missing Normalization* (real-world example map A [41]) and result in completely different conclusions about the same data. This example focuses on one section of the V-FRAMER (one-pager in Figure 2).

ABSTRACT

Existing data visualization design guidelines focus primarily on constructing *grammatically-correct* visualizations that faithfully convey the values and relationships in the underlying data. However, a designer may create a grammatically-correct visualization that still leaves audiences susceptible to *reasoning misleaders*, e.g. by failing to normalize data or using unrepresentative samples. Reasoning misleaders are especially pernicious when presenting public policy data, where data-driven decisions can affect public health,

safety, and economic development. Through textual analysis, a formative evaluation, and iterative design with 19 policy communicators, we construct an actionable visualization design framework, V-FRAMER, that effectively synthesizes ways of mitigating reasoning misleaders. We discuss important design considerations for frameworks like V-FRAMER, including using concrete examples to help designers understand reasoning misleaders, and using a hierarchical structure to support example-based accessing. We further describe V-FRAMER's congruence with current practice and how practitioners might integrate the framework into their existing workflows. Related materials available at: <https://osf.io/q3uta/>.

CCS CONCEPTS

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

KEYWORDS

Visualization design guidelines, Reasoning, Framework, Public policy

ACM Reference Format:

Lily W. Ge, Matthew Easterday, Matthew Kay, Evanthia Dimara, Peter Cheng, and Steven L Franconeri. . *V-FRAMER: Visualization Framework for Mitigating Reasoning Errors in Public Policy*. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, May 11–16, 2024, Honolulu, HI, USA. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3613904.3642750>

1 INTRODUCTION

Imagine a policy maker tasked with providing recommendations on whether the public should wear masks in supermarkets. When inspecting the map shown in Figure 1.A, they see California and Texas in dark red, showing a high number of cases. This might tempt them to recommend that masking is especially important for those states. But because Figure 1.A shows *absolute* case counts, it is essentially a population map, and it is unsurprising that states with a larger population have a higher count. The map in Figure 1.B shows a far more useful view—normalized by population—which should better reflect the number of cases in an average supermarket, and would lead our policy maker to make more appropriate masking recommendations.

Visualizations can convey massive amounts of information to support data-based reasoning, but ineffective designs can lead their powers to backfire. As demonstrated in Figure 1, although map (A) can lead viewers to make poor decisions, it does follow typical visualization design guidelines intended to ensure that it is *grammatically correct*—it faithfully conveys the values and relationships in the underlying data. Grammatical violations of existing design guidelines tend to include visual distortions, such as inappropriate *y*-axis truncation of a bar chart (e.g., [18, 63]), which exaggerates differences. But the map in Figure 1.A does not contain grammatical violations. Its misleadingness stems from missing normalization, when normalizing by population is necessary to correctly answer the underlying domain-relevant question: *what is the COVID risk in each state?*

The general public is not typically trained to evaluate such relatively subtle differences when reasoning with data and is particularly susceptible to these issues. Designers, therefore, must construct visualizations that are not only grammatically correct, but also minimize potential reasoning errors to avoid misleading its viewers. In other words, it is essential to avoid a class of issues during visualization construction that could exist even in grammatically-correct visualizations, which we refer to as *reasoning misleaders*. Prior works have identified ways visualizations can mislead [19, 36, 43] but still lack an effective synthesis of design guidelines targeting these reasoning misleaders, making them harder to guard against in practice. Existing defenses against these issues mainly rely on expert knowledge scattered in the literature or left implicit in already-made, effective visualizations. External

tools assisting visualization construction (e.g., visualization linters [8, 21, 42]) typically target violations of guidelines on *grammatical* components (e.g., on scales, marks, or channels). Therefore, we need actionable guidelines that target the harder-to-discern reasoning misleaders.

We propose V-FRAMER, a Visualization Framework for Mitigating Reasoning Errors, situated in public policy and co-designed with and for policy communicators. We focus on the field of public policy to prioritize interventions where poor decisions have pernicious effects on high-stakes issues such as public health (e.g., mandatory masks), public safety (e.g., gun rights), and natural disaster prevention and response (e.g., hurricane forecasts). V-FRAMER effectively synthesizes ways of mitigating reasoning misleaders in an actionable, hierarchical structure, which was developed through a highly iterative process. We distill guidelines from both visualization and public policy literature to create a preliminary version of V-FRAMER (Section 4), which we refined iteratively with the expertise of 19 policy communicators during a formative evaluation (Section 5.1). Each interview session was composed of questions about their current practices before seeing V-FRAMER and their interactions with V-FRAMER after we showed it to them. Our before-and-after comparisons demonstrate that our final V-FRAMER covers the sets of considerations important in practice (Section 5.2). Additionally, we finalize important design objectives and describe how the final V-FRAMER meets those objectives. Based on V-FRAMER's congruence with practice, we discuss potential ways it can be integrated into existing workflows of policy communicators, such as through a checklist or an educational tool (Section 6). By offering a framework that spotlights issues that could exist even in grammatically-correct visualizations, we hope to strengthen support for better data-based reasoning with visualizations.

2 RELATED WORK

In this section, we discuss distinctions in how charts can misinform, review existing guidelines in visualizing data to point out how current focuses are insufficient, and discuss the integration of external representations such as visualizations in the policy analysis and communication process.

2.1 Misleading Visualizations

Prior work in visualization research investigated specifically how distortions of scales and visual encodings (e.g., *y*-axis truncation [11]) or improper mappings between grammatical components (e.g., mapping continuous data onto a perceptually discrete rainbow color scale [51]) can affect the perceived message of a visualization. Researchers have also compiled ways a chart can mislead, such as categorization of visualization mirages by McNutt et al. [43] and issues that can lead to misinformative visualizations by Lo et al. [36]. These categorizations offer valuable insights and lay the foundation for further investigations on visualization misinformation.

With the surge of visualization use during the global pandemic, more studies looked to real-world examples found in media and pinpointed especially problematic ways visualizations can mislead. Lee et al. [30] investigated *counter-visualizations*, which they defined to be “visualizations using orthodox methods to make unorthodox arguments”, and found that these seemingly well-formed

visualizations appear much more commonly in support of anti-mask arguments. Similarly, Lisnic et al. [33] analyzed misleading visualizations that appeared on Twitter during COVID-19 and showed that most misleading charts, in fact, do not violate design principles, but instead are misleading due to issues such as cherry picking or inappropriate causal inference. To test for people’s ability to identify visualization misinformation, Ge et al. [19] developed Critical Thinking Assessment for Literacy in Visualizations (CALVI) and discussed a *misleaders* set (i.e., decisions made in the construction of visualizations that can lead to conclusions not supported by the data). This set can be roughly separated into misleaders that can be more easily identified with adequate attention to the right part of the visualization (e.g., manipulation of scales) or misleaders that seemed harder to discern even when given attention (e.g., missing data) [19]. The misleaders that do not seem to rely as much on attention to identify were later incorporated into our framework as the majority set of *reasoning misleaders*, as described in Section 4.

Previous works on misleading visualizations suggest a key distinction in how charts can misinform: from grammatically-*incorrect* visualizations versus grammatically-*correct* visualizations. The former mislead by violations of basic design principles, while the latter can still mislead with no visual distortions or design violations (i.e., contains *reasoning misleaders*). Misleadingness from grammatically-correct visualizations share some commonalities with data-analysis issues in other fields such as statistics [16, 22, 49]. However, discussions of similar issues outside of the data visualization field either usually only cover a subset of the issues we are targeting or do not sufficiently examine the impacts on the resulting visual representation. Prior work within the visualization community targeting misleading visualizations were also more focused on the summarization or classification of related issues rather than providing guidelines that have a coverage of considerations important in practice. This necessitates further investigations on how to better support designers in navigating around potential reasoning misleaders in data visualizations.

2.2 Visualization Design Guidelines

Visualization research increasingly prioritizes the study of intuitive designs that should be accessible to broad audiences (for a review, see [18]). Moreover, existing work guides the choice of which graph type to choose to maximize perceptual precision when reading values [5, 9] or judging correlations [20], maximizing the discriminability of color palettes [59], or creating effective designs for prescribed lower-level perceptual tasks [48, 55]. Much of this advice has also been formalized within rule-based recommender systems which provide more guidance, including APT [38], SAGE [53], Show Me within Tableau [39], Voyager [71], and Draco [45].

In an effort to correct visualization designs that go astray, Hopkins et al. [21] developed VisualLint, which identifies erroneous elements in a visualization and annotates its components. Chen et al. [8] designed a linter and fixer framework, VizLinter, that detects issues that deviate from well-recognized visualization design principles and fixes the visualization accordingly. Visualization linters work well in identifying violations of existing design guidelines precisely because the well-known principles refer to relatively generic, *grammatical* components of visualizations, such as scales, marks,

or channels. Besides linters, Kristiansen et al. [29] have developed recommendation systems mainly resolving issues on grammatical components. Others have encouraged visualization skepticism, or re-examination, during design [14, 37, 43]. But re-examinations ultimately rely on the examiner being able to identify reasoning misleaders that often appear in practice but which no existing linters, recommenders, or frameworks have complete coverage of. Advice or tools that primarily help with the construction of grammatically-correct visualizations cannot adequately guard against misleading, but still grammatically-correct, visualizations.

2.3 Visualizations in Public Policy

Policy problems are often referred to as “wicked problems” [2, 46, 50, 60] because they are complex, high-stakes, ill-defined, and do not have single correct answers. Despite this, policy analysts have defined many of the core phases involved in policy problem-solving to assist in the analysis process [4, 25]. For example, an eight-step process [4] used for policy analysis includes: defining the problem, assembling evidence, constructing the alternatives, selecting the criteria, projecting outcomes, considering trade-offs, narrowing and deciding, and finally clearly conveying a prescription. At its core, policy analysis is essentially complex problem-solving, which can also be assisted with techniques that aid problem-solving and reasoning in general. Such techniques include using external representations (including visualizations) [1, 13, 28, 73], which can help with considering and learning about complex ideas. In support of using visualizations to assist policy decision-making, Ruppert et al. [54] argued that visualizations should be incorporated in policy analysis stages to facilitate communication between different stakeholders including policy analysts, domain experts, and public stakeholders (e.g., general public). Yet, there is inadequate guidance on how to effectively use visualizations to support sound policy reasoning in the general public.

In practice, expert practitioners in the field—such as Hans Rosling [52] and journalists or news outlets [17, 47, 65] who specialize in explaining complex or data-heavy topics—leverage visualizations to communicate policy-relevant data. Some of this policy communication expertise has been formalized as “chart choosers” for effective visualizations to highlight a given type of data pattern. For example, the Financial Times [64] introduced a breakdown of chart types by the underlying data relationships, which they note was inspired by a similar project, the Graphic Continuum [58], that also seeks to guide graphic choices. Others focus on the diversity, equity, and inclusion aspects in data visualizations, such as the Do No Harm Guide from the Urban Institute [57].

Still, these visualization design guidelines primarily focus on the construction of grammatically-correct visualizations. We address this lack of clear guidance by developing an actionable framework that (1) situates in public policy and explicitly aims to cover data visualization considerations important for avoiding reasoning misleaders in practice, and (2) is co-created with its user base (i.e., policy communicators).

3 V-FRAMER DESCRIPTION AND WALK-THROUGH

Before describing V-FRAMER’s development process (Section 4), we first give an overview of the final framework. In this section, we walk through the hierarchical structure as well as the categories within the final V-FRAMER (one-pager¹ shown in Figure 2).

Data Considerations. The top level (A in Figure 2) of the hierarchy is composed of *Data Consideration* categories: *considerations that are most relevant when working with data to avoid reasoning misleaders in resulting visualizations*. The three data consideration categories are *Data Representativeness* (Section 3.1), *Comparison Basis* (Section 3.2), and *Distributions* (Section 3.3).

Example Policy-making Stages. The second level (B in Figure 2) is composed of example *Policy-making Stages: policy questions that data could answer*. Each example policy-making stage follows a data consideration category. For example, *What variable(s) show(s) the problem size?*, *Is the problem size different for subgroups?*, and *Is the problem size worsening and at what speed?*

Reasoning Misleaders. The bottom level (C in Figure 2) is composed of *Reasoning Misleaders: issues in grammatically-correct visualizations that can lead to conclusions not supported by the data*. The reasoning misleaders in the set are *Missing Data*, *Cherry Picking*, *Missing Normalization*, *Inappropriately Aggregating*, *Concealing Uncertainty*, and *Inadequately Representing Uncertainty*. Each reasoning misleader directly follows an example policy-making stage, which represents relevant questions for consideration. V-FRAMER illustrates the impact of each reasoning misleader through a table of visualization examples, created from combinations of reasoning misleaders and common chart types. A reasoning misleader in a visualization cannot be identified solely by examining the visualization or the data it contains. One must also consider the underlying analytic question, evaluating whether the visualization design could lead to an inaccurate perceived message for that specific question. Moreover, it is important to note that the table of visualization examples is intended as visual demonstrations, not as an exhaustive list (refer to Section 5.2 for details on the design objectives). Here, we walk through the final V-FRAMER by describing representative examples from each combination, which contains illustrations of potentially *Misleading* and *Better* examples (both indicated with their respective colors in the following descriptions).

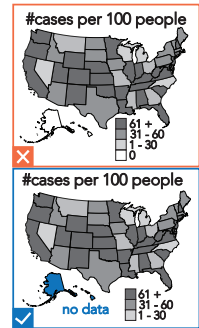
3.1 Data Representativeness

This data consideration requires one to reason about whether the data sample and variable of interest are representative of the population and the problem, respectively. It is important to indicate potential biases of the data-generating process and ensure the presented data is representative of the population. The variable of interest should also provide an adequate measure of the problem. The policy-making stage example is *What variable(s) show(s) the problem size?* Specific reasoning misleaders that pertain to *Data Representativeness* are *Missing Data* and *Cherry Picking*; both include potentially-biased samples or variables and hinder the accuracy of the conclusions drawn from the presented data.

What variable(s) show(s) the problem size? The communication goal in this example stage is to convey the state of variable(s) of interest, which would be indicative of the problem size. For example, temperature over time can convey climate change, and questions like “what time frame would provide adequate context?” can be relevant to ensure the presentation is representative of the problem. Thus, it is under the *Data Representativeness* consideration category. We detail its relevant reasoning misleaders below.

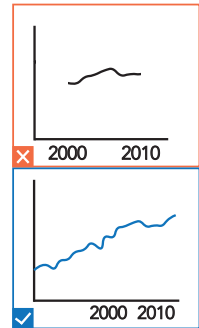
Missing Data *Not indicating missing data*

can lead the viewer to inaccurate impressions of the data. For instance, missing values may be defaulted to zeros either by choice or through the visualization authoring tool. However, if a choropleth map that shows COVID-19 infection rates has a region that was coded as zero due to missing data, it can mislead viewers into thinking there are in fact no cases in that particular region (combination of *Missing Data* and *Map* in Figure 2). Thus, in the case of missing data, it is better to *use salient visual features to indicate incomplete data*, such as adding direct annotations [61] or indicating uncertainty [56].



Cherry Picking *Only presenting a subset of data*

can be potentially misleading. In the case of climate change, there may be certain periods of time where the change is relatively small but the overall trend is still increasing. If only the time frame with a relatively stable trend is shown, then that can lead to misinterpretations of the data, such as the top example shown on the right (combination of *Cherry Picking* and *Line* in Figure 2). For a more complete understanding, *it is essential to include all important context in data*, such as the entire trend instead of a biased subset [36].



3.2 Comparison Basis

This data consideration is relevant when making comparisons between different groups. It is important to compare groups under a fair comparison basis and ensure the scale is informative. The policy-making stage example is *Is the problem size different for subgroups?* The specific reasoning misleaders for *Comparison Basis* are *Missing Normalization* and *Inappropriately Aggregating*; both disregard subgroup differences and could lead to inaccurate conclusions.

Is the problem size different for subgroups? The goal in this stage is to communicate comparison between subgroups. For instance, it is common to compare regional subgroups in COVID-19 data: “how do the risks of COVID-19 in different states compare to each other” or “which states are more impacted by COVID-19 and would need more strict mask mandates” can all be relevant questions that require relative comparisons between subgroups.

¹This one-pager is resized from supplemental materials to fit vertically in the paper.

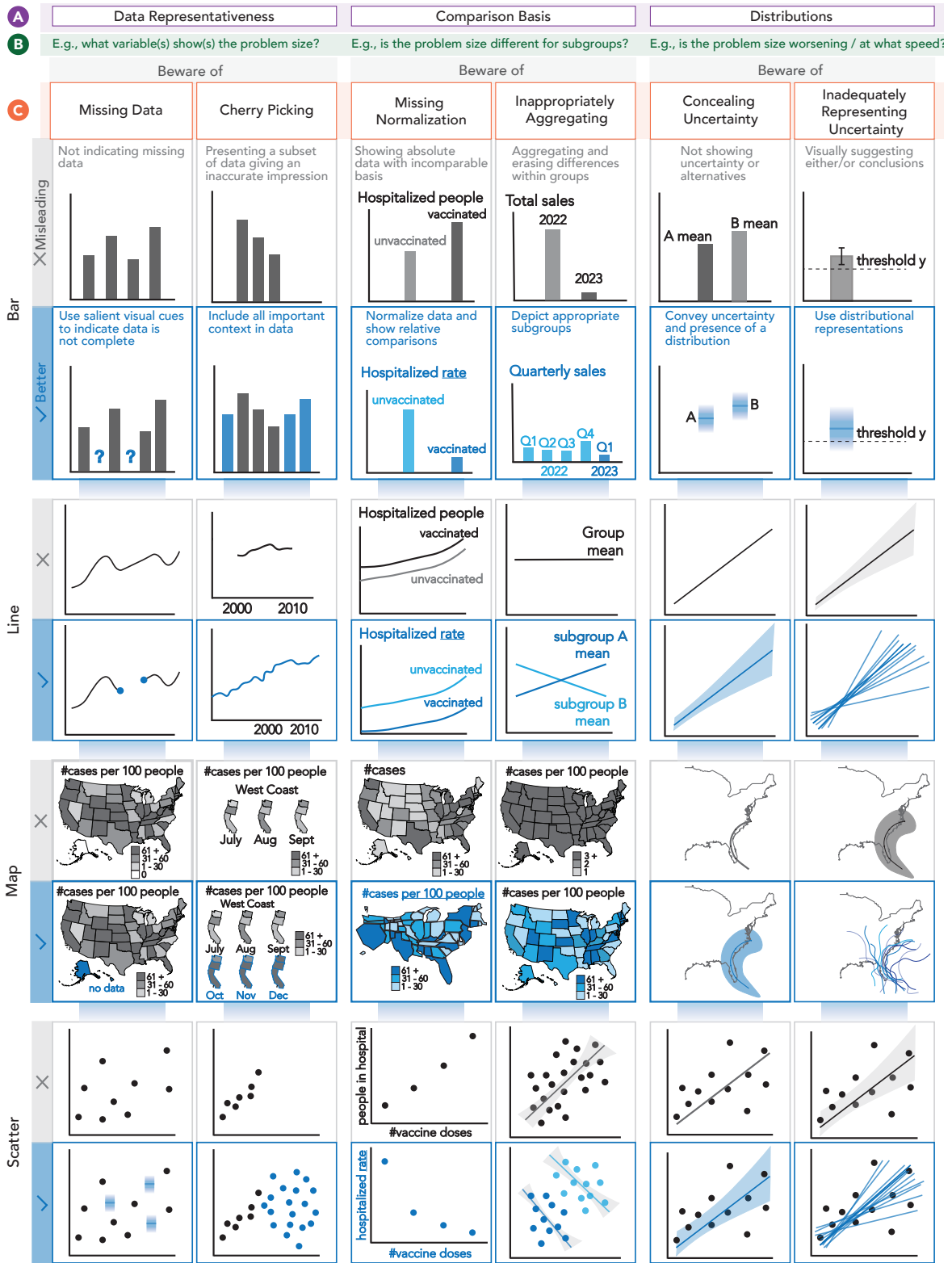
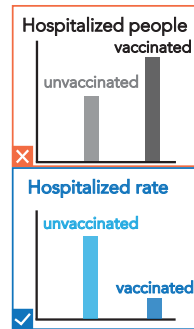


Figure 2: V-FRAMER and its hierarchical components: **A** Data considerations, **B** Example policy-making stages, and **C** Reasoning misleaders with the visualization examples table.

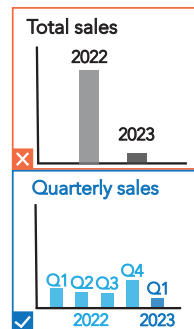
Thus, it is under the *Comparison Basis* consideration category. We detail its relevant reasoning misleaders below.

Missing Normalization When subgroups are under comparison, **absolute value comparisons with an incomparable basis** are often uninformative and may lead to incorrect conclusions. For instance, if one were to show absolute numbers of people in the hospital who are vaccinated versus unvaccinated, more *people* in the hospital would be vaccinated, such as the top example shown on the right (combination of *Missing Normalization* and *Bar* in Figure 2). This is because the majority of the population is vaccinated, so



the absolute counts of hospitalized vaccinated people would naturally outweigh unvaccinated people (analogous to Figure 1.A). However, the conclusion is flipped if we instead consider hospitalized *rates* rather than absolute counts. Out of the people who are vaccinated, fewer are in the hospital compared to people who are unvaccinated. Thus, in this case, it is better to **normalize data when making relative comparisons** (also recall Figure 1.B).

Inappropriately Aggregating When working with data that contains subgroups, whether or not to aggregate or how to aggregate is another important consideration. **Aggregation that erases differences within groups** can lead to vastly different impressions of data. For instance, if it were currently the end of March in 2023, it would not be meaningful to aggregate and compare total sales between 2022 and 2023, because 2023 is not over yet. The level of aggregation (annual) could lead to inaccurate impressions, as shown on the right (combination of *Inappropriately Aggregating* and *Bar* in Figure 2). In this case, it is better to **depict subgroup differences with appropriate granularity** (e.g., by using quarterly sales instead).



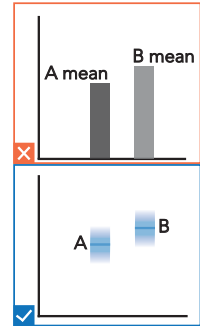
3.3 Distributions

This data consideration requires one to think about distributions rather than merely point estimates. It is often more informative to convey the values of each variable associated with different outcomes and their chances of occurrence. The policy-making stage example is *Is the problem size worsening and at what speed?* Reasoning misleaders that are relevant to *Distributions* are *Concealing Uncertainty* and *Inadequately Representing Uncertainty*; misleadingness could come from either not showing uncertainty or showing a representation that can still lead to falsely certain conclusions.

Is the problem size worsening and at what speed? This example stage can be applicable when making projections to predict the future state of a problem, which is often uncertain. For instance, during the peak of COVID-19, one may need to predict the trend of cases in preparation for informed decisions amid the rapidly

changing circumstances. Thus, this stage is under the *Distributions* category, as one would need to consider alternative projections or a distribution of possibilities. We detail its relevant reasoning misleaders below.

Concealing Uncertainty **Completely disregarding uncertainty** such as not showing distributions can lead to falsely certain conclusions. This could be making predictions based on a regression line that do not have uncertainty bands. A viewer unaware of uncertainty and the underlying distributions would be more tempted to make falsely certain predictions. Or in a comparison of two groups, point estimates of the means might show a gap between them [70], but this gap may become less meaningful when distributions are shown instead. The examples shown here demonstrate this (combination of *Concealing Uncertainty* and *Bar* in Figure 2)—**showing the presence of a distribution can support more accurate conclusions**, such as using gradients, which is a way to show uncertainty that could be more generalizable to a variety of chart types [24] and also discourages binary interpretations [12].



Inadequately Representing Uncertainty Not all uncertainty representations lead to desirable results. Some suggest extremely dichotomous conclusions (i.e., **visually suggesting either-or conclusions**), such as the top example with the hurricane forecast cone shown on the right (combination of *Inadequately Representing Uncertainty* and *Map* in Figure 2). Although there is a visual presentation of a distribution, the clear cut off with the cone can lead viewers to more easily conclude that if they are not within the cone, then they are safe from the impact of the hurricane [6, 34, 66], which can lead to fatal consequences. **More distributional representations**, instead, can mitigate dichotomous ways of thinking [12, 35].



4 PRELIMINARY FRAMEWORK CONSTRUCTION

In the development of V-FRAMER, we first constructed a **preliminary version** (see supplemental materials) with preliminary design objectives derived from related work. This was an attempt to avoid the possible scenario of important categories not coming up when participants discussed examples based on recent memory during the formative evaluation. Additionally, we used the preliminary version of V-FRAMER in the interviews to elicit feedback for iterative refinements (Section 5). Based on the results of the formative evaluation, we constructed the **final version** of V-FRAMER, as described in Section 3 and in Figure 2.

Here, we describe the construction process (Figure 3) of the preliminary V-FRAMER. We chose a one-page format for the framework, because we aimed to present the guidelines in a centralized place for ease of access and transfer.

4.1 Preliminary Design Objectives

The three preliminary design objectives (pDO) derived from a review of related work served as a basis for the development of the preliminary version of V-FRAMER. Along with the framework content, these three pDO were also candidates for refinement during the formative evaluation.

pDO.1 Explicitly integrate data visualization and public policy. As detailed in Section 2.2 and Section 2.3, most visualization design guidelines are either not grounded in public policy or primarily focus on the creation of merely grammatically-correct visualizations (e.g., [43, 58, 59, 64]). Thus, we focus on issues that can still exist in grammatically-correct visualizations, situate V-FRAMER in public policy, and explicitly connect relevant visualization considerations with policy considerations.

pDO.2 Provide a highly directed process for guided usage. As described in Section 2.3, expert policy analysts have defined step-by-step processes (e.g., [4]) that are commonly-used guides for policy analysis. This approach seem to provide more structure for the often complex and ill-defined policy problems. Thus, we base the structure of preliminary V-FRAMER around a directed process, similar to how policy analysts are trained.

pDO.3 Demonstrate examples with concrete illustrations. As mentioned in Section 2.3, external representations can help with the understanding of complex ideas (e.g., [1, 13, 28, 73]), which is also applicable to ill-defined, and complex, policy problems. Additionally, the reasoning misleaders that we are targeting seem to already be harder to identify [19]. Thus, we show concrete visual illustrations on V-FRAMER to help explain the presented concepts.

With these preliminary design objectives in mind, we detail the construction process for preliminary V-FRAMER in Section 4.2, where we note satisfaction of corresponding preliminary design objective(s) in parentheses when relevant.

4.2 Coding and Preliminary Construction

To distill preliminary sets of categorizations spanning both data visualization and public policy (pDO.1), we drew upon several sources: *Factfulness* [52], *A Practical Guide for Policy Analysis* [4], a synthesis of visualization *misleaders* from prior work [19], and real-world examples of misleading visualizations from VisLies meetups held in conjunction with the IEEE VIS conference [67]. We aimed to extract three sets of categories in total: one set from data visualization, one set from public policy, and one set to explicitly integrate the two (pDO.1). We adopted a team-based coding approach [40] with a total of six coders including a lead coder. The iterative process involved regular meetings [40] where the coding team reviewed and refined the codes (i.e., categories) and definitions as appropriate.² The category-specific inclusion and exclusion criteria and mapping processes are outlined here. The final definitions are detailed in Section 3.

²See supplemental materials for documentation on the earlier phases of the coding process.

To compile an initial set of issues relevant to *Reasoning Misleaders*, we reviewed prior literature [19] for a synthesis of related issues as well as real-world examples of misleading visualizations [67] (A in Figure 3). From the list of 11 misleaders categorized by Ge et al., we first retained categorizations that were **not grammatical violations** (grammatical violations result in visual distortions on or improper mappings between scales, marks, or channels). For instance, the *Manipulation of Scales* categories were excluded because they were visual distortions or manipulations. In contrast, *Missing Normalization* was retained because it involves no visual distortion but can still lead to inaccurate conclusions. Following this inclusion criterion, we retained 5 out of the 11 categories from prior work [19] in our reasoning misleaders set (i.e., *Missing Data*, *Cherry Picking*, *Missing Normalization*, *Inappropriate Aggregation*, and *Concealed Uncertainty*). To evaluate this set against other sources, we examined real-world examples of misleading visualizations from VisLies meetups (2015 - 2021) [67] held in conjunction with the IEEE VIS conference and found that most of the relevant issues from those real-world examples fit into this set. We only came across

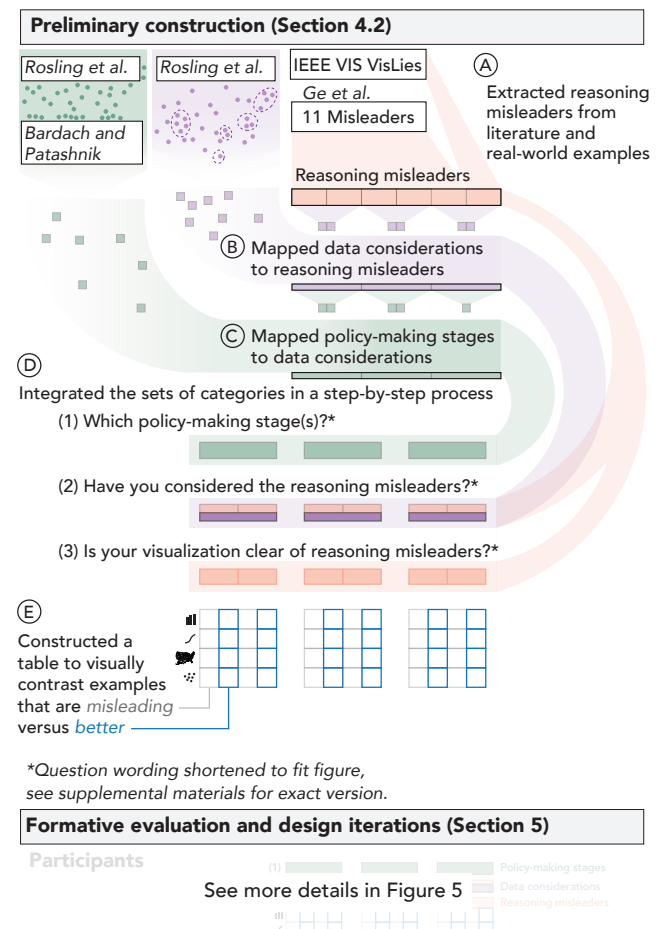


Figure 3: The construction process for the preliminary version of V-FRAMER. The preliminary V-FRAMER was used as the first version in the formative evaluation (Section 5).

one example that did not fit well, which was a hurricane forecast cone [66] and categorized it as *Inadequately Representing Uncertainty*. This additional reasoning misleader category was to account for potentially-misleading uncertainty representations rather than merely no uncertainty at all (as implied by *Concealed Uncertainty*). The result was a set of 6 reasoning misleaders.

To extract an initial set of *Data Consideration* categories, we conducted document analysis using *Factfulness* [52]. It contains practical knowledge in communicating policy problems with visualizations that goes beyond typical guidelines about grammatical visual components. With the first-author as the lead coder, we iteratively performed open-coding [15] and categorized key points from *Factfulness* into groups considering their relevant data characteristics. We especially focused on *data* characteristics rather than visual elements because the aim of the framework targets issues that can happen even in grammatically-correct visualizations (i.e., assuming the visual elements are well-designed). Data characteristics we looked for include data types (e.g., time series), data transformations (e.g., rate), data biases (e.g., cherry picking), or extrapolation (e.g., predictions under uncertainty). In order to extract the categories of *Data Considerations* most relevant to avoiding reasoning misleaders, we considered the categorizations in the context of the set of reasoning misleaders (Ⓔ in Figure 3) and retained the ones that can be **directly mapped to reasoning misleaders**. For the categories that did not map well, we further refined them. For example, *Uncertainty* was merged into *Distributions* because both mapped to the same two reasoning misleaders: *Concealed Uncertainty* and *Inadequately Representing Uncertainty*. The coding process was highly iterative with regular meetings to refine the codebook [40], resulting in a set of 3 data considerations.

For the distillation of *Policy-making Stages*, we started with a commonly-used guide, *A Practical Guide for Policy Analysis* [4]. This guide was used as the initial codebook [40] to iteratively code key points from *Factfulness* [52]. In order to ensure the policy-making stages are questions that data visualizations could help answer, we iteratively mapped the policy-making stages to the data consideration categories (Ⓒ in Figure 3) and retained the ones that can be **directly mapped to data considerations**.³ The result was a set of 3 policy-making stages. The categories extracted from these initial sources were only considered preliminary and were candidates for revision during the formative evaluation (Section 5).

To provide more structure and better integrate the three sets of categories we extracted, we organized them in a step-by-step process in the framework: (1) *What policy-making stages are you conveying with data?* (featuring policy-making stages), (2) *Have you considered the reasoning misleaders corresponding to your policy-making stage above?* (featuring data considerations connected with their corresponding reasoning misleaders), (3) *Is your visualization clear of potential reasoning misleaders?* (featuring reasoning misleaders) (Ⓓ in Figure 3). This step-by-step process was inspired

³For instance, “consider the causes of the problem” was excluded. A potential reasoning misleader, “inferring unsupported causations”, we initially added only as a potential error in the causation stage was kept in during the iterations merely for the purpose of brainstorming techniques that may be useful for mitigation. But later we further confirmed that it is more about the lack of knowledge of the causal structure in the domain than the property of the visualization itself [19].

by the process seen in the practical guide [4] and our attempt to situate the framework within public policy (pDO.1 and pDO.2).

We then constructed a visualization examples table to demonstrate potential visualization techniques that can help mitigate the effects of the reasoning misleaders (pDO.3). We aimed for the examples to be easy to understand and able to fit in our one-page format. Thus, the table contained representative example demonstrations and was not meant to be exhaustive. We looked to prior work [31] for a set of common chart types to support example construction. The top four data visualization types in news outlets ranked by Lee et al. [31] are choropleth map, bar chart, line chart, and bubble chart. Instead of bubble chart, we included scatterplot, since it is essentially a base version of a bubble chart. As a result, the examples table is made up of combinations of the 6 reasoning misleaders and 4 common chart types (Ⓔ in Figure 3 and more details in Figure 2).

This preliminary V-FRAMER⁴ with the step-by-step process was the first version used in the formative evaluation.

5 FORMATIVE EVALUATION AND DESIGN ITERATIONS

We conducted a formative evaluation to (1) analyze the congruence of our preliminary framework with practitioners’ knowledge, (2) incorporate feedback from policy communicators to iteratively refine the preliminary V-FRAMER, and (3) finalize important design objectives for V-FRAMER. Similar to the iterative co-design process employed by prior work [69], this was not meant to be a controlled comparative study. The aim of the iterative process was to arrive at a framework that is not only grounded in literature, but also congruent with practitioners’ knowledge through co-design.

Participants We started recruitment by contacting professional policy communicators known to the authors. From there, we recruited by snowball sampling, encouraging participants to forward the recruitment material to their professional circles. At the same time, we publicly posted our recruitment material through online platforms such as organizational Slack channels and mass emailing systems. The recruitment material invited anyone who works in public policy and communicates data to schedule an interview via an online scheduler. The scheduler contained screening questions to ask potential participants to briefly describe their professional role and whether they are based in the U.S. before they can confirm their appointment time.

We had 19 participants respond to our call, and all participants were based in the United States. The participants either study or work professionally⁵ in public policy and communicate policy-relevant data (Figure 4). Participants worked in both private and public sectors, with roles including: data associate, data analyst, research associate. Policy problems our participants work on include: housing data analysis and policy, tax policy, public health, transportation, and water equity. Upon successful completion of the interview, participants were offered 30 USD as compensation for their time.

⁴Although the policy stage related to causes of the problem was deemed out of scope (discussed in footnote³), we included it in the framework version used in the interviews (explicitly indicated as out of scope) to hear any thoughts our participants may have on causation within public policy. For related discussion, see Section 7.1.

⁵Including self-reported part time work ($n=1$).

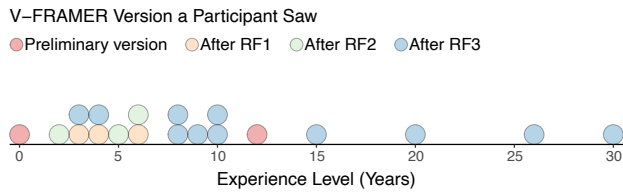


Figure 4: Years of experience working in public policy reported by participants and colored by the version of V-FRAMER they interacted with during their interview. Two participants self-identified as students at the time of the interview, one of whom did not provide years of experience (coded as 0 years).

Interview Procedure The first author (i.e., interviewer) conducted the interview sessions over Zoom, which ranged from approximately 30 minutes to 79 minutes, averaging about 45 minutes per participant. The interviewer first asked participants to read the consent form, and then answered any questions they may have. After gaining consent from the participants, the interviewer presented a slide deck for them to follow along as the interviewer proceeded with the semi-structured questions. All of the interview sessions contained two distinct sections (i.e., *before* and *after* the introduction of our framework) involving a total of three primary tasks: (1) before introducing our framework, we aimed to understand their current practices and challenges in policy communication, (2) after introducing the most up-to-date V-FRAMER, we asked them to apply their example(s) to our proposed framework (i.e., verbally walking through a step-by-step process that starts with identify the relevant stage among the *policy-making stages*, then consider the potential reasoning misleaders in the context of each *data consideration*, and finally examining the visualization examples in the relevant combinations of *reasoning misleaders* and chart types), and (3) reflecting on their experiences using our framework. The last 2 minutes of the interview session were dedicated to a demographic survey. The interview protocol can be found in supplemental materials.⁶

Methods Throughout the interview process, we considered and incorporated feedback from participants to iteratively refine our preliminary V-FRAMER. We also conducted thematic analysis using both inductive and deductive approaches to investigate patterns in the data [3]. The first author anonymized and split the participants' transcribed responses based on their answers relevant to the interviewer's questions. Afterwards, the first author (i.e., main coder) and another author began discussing excerpts and derived initial codes together. Because each interview session had two distinct sections, each transcription was then split into two (*before* and *after* the introduction of our framework) for analysis. The main coder used the preliminary categories distilled in Section 4 as the codebook [40] to analyze participants' responses to questions *before* any discussion of V-FRAMER. This was to identify current considerations and challenges in policy communication, without the direct influence of any framework. We used them as one proxy for evaluating congruence between practitioners' knowledge and our

framework. Similarly, the main coder then analyzed participants' responses to questions *after* seeing V-FRAMER. This was to serve as another proxy for investigating (1) how congruent the preliminary V-FRAMER is to practitioners' knowledge, and (2) whether or how the framework could be integrated into existing workflows. Note that not all participants consented to the inclusion of their selected transcriptions, in which case we only include paraphrased or aggregated insights. To further protect the anonymity of our participants, we have used *[brackets]* with more generic terms to abstract the details from the participants. The generic terms we use as replacements still retain the excerpt's necessary meaning.

5.1 Iterative Refinements of V-FRAMER

Before discussing the congruence of V-FRAMER with practitioners' knowledge, we first describe the refinement process and how V-FRAMER changed in response to participants' feedback in between interviews (Figure 5⁷).

Refinement 1 (RF1): Addition of higher-level categories. The first two participants who used preliminary V-FRAMER to walk through their example both expressed hesitancy on how to apply their example to a particular policy-making stage. For instance, P1 mentioned that *“a lot of these stages can sort of go together.”* This hesitancy suggested that the users may need more direction in choosing their most relevant policy-making stage, and prompted us to add a higher-level categorization for the policy stages, grouping them into *past and current*, or *future state of the problem*.

Refinement 2 (RF2): Addition of scaffolding and interactivity. After interviewing three more participants, we accumulated more evidence that participants were not following the expected “vertical reading order” in the step-by-step process (starting with the most relevant policy-making stage then drilling down to its associated reasoning misleaders and examples). The higher-level categories discussed in RF1 were added to particularly assist in providing more direction, but they did not offer sufficient guidance either. Instead, it seemed to add a layer of restraint for some participants. For instance, P5 brought up that, regarding the *future state of the problem* category that mainly covered the *Distributions* data consideration, *“sometimes you are evaluating the current or past state of something based on a sample... so it's not just about future”*. Thus, we removed the higher-level categories from RF1. As another attempt to ensure V-FRAMER satisfies pDO.2, we included an interactive version of the framework to more directly lead participants through the intended step-by-step process. We still presented the one-page V-FRAMER after the interactive version to elicit any additional feedback.

Refinement 3 (RF3): Addition of guiding arrows on one-pager. We noticed that participants generally found the one-page framework to be easy to understand and were not as confused on the reading order after first going through the interactive version. This observation indicates that pDO.2 was sufficiently satisfied. This also suggests that using more scaffolded methods, such as interactivity, when first introducing the framework to people has value of its own. Participants who saw both the interactive and the one-page version also found the one-pager to be helpful. P7, for example, mentioned that *“it's nice to have a one-pager to pass onto people.”*

⁶<https://osf.io/q3uta/>

⁷Wording changes were considered minor and not detailed. See supplemental materials for the exact versions.

Another participant expressed that it is helpful to have everything in one place to assist with data quality checks. Thus, although interactivity appeared to assist in the understanding of V-FRAMER, the one-pager should also stand alone. As our attempt to make the one-pager stand alone and more explicitly indicate the intended step-by-step process, we added guiding arrows (see Figure 5).

Throughout the rest of the interviews, we kept the same format: the interactive version preceding the one-pager with its guiding arrows. This is to further evaluate the framework's preliminary design objectives and its congruence with practitioners' knowledge.

5.2 Congruence of V-FRAMER with Practice and Final Design Objectives (DO)

As detailed in the interview procedure and method of analysis, each interview session was separated into *before* introducing the framework (i.e., discussion on their current practices and challenges) and *after* introducing the framework (i.e., discussion focusing on the content of the framework). We describe how well V-FRAMER captures practitioners' knowledge from these two aspects and how the formative evaluation informed the final set of design objectives.

5.2.1 Congruence of V-FRAMER with practitioners' considerations of data and potential reasoning misleaders.

Before introducing V-FRAMER. Participants' responses to questions before seeing the framework can serve as additional data to evaluate our preliminary framework, since they responded only based on their prior knowledge. After conducting the interviews, we mapped participants' responses to the categories on V-FRAMER and found that each example from participants' current considerations of data and challenges fit into at least one category from our set of *Data Considerations* or *Reasoning Misleaders*. Among the sets of categories, more participants discussed considerations related to comparison basis and its associated reasoning misleaders. Collectively, discussions before introducing our framework during the interviews covered all of the *Data Consideration* and *Reasoning Misleader* categories but one: *Inappropriately Representing Uncertainty*. There was generally a match between the data consideration and its associated reasoning misleaders. However, considerations related to distributions and its associated reasoning misleaders were not discussed as much as others, but we later found these to be equally important to keep in the framework (see Section 7.3).

After introducing V-FRAMER. There were notably more discussions around all *Data Consideration* and *Reasoning Misleader* categories, and each category contained relevant examples from participants. Many participants began thinking of more examples from their own experience that resonated with the content in V-FRAMER when they reached the visualization examples table. One participant especially acknowledged that the visual illustrations could help make the concept of reasoning misleaders more concrete. The increased engagement and examples from participants suggested that the visual demonstrations of the effects of the reasoning misleaders seemed to assist in understanding. Thus, **we retain pDO.3 that V-FRAMER satisfied as DO.3: demonstrate visual examples to illustrate otherwise abstract concepts.**

The categories of data considerations generally mapped to its associated reasoning misleaders across participants' examples. Although all of the data considerations and reasoning misleaders

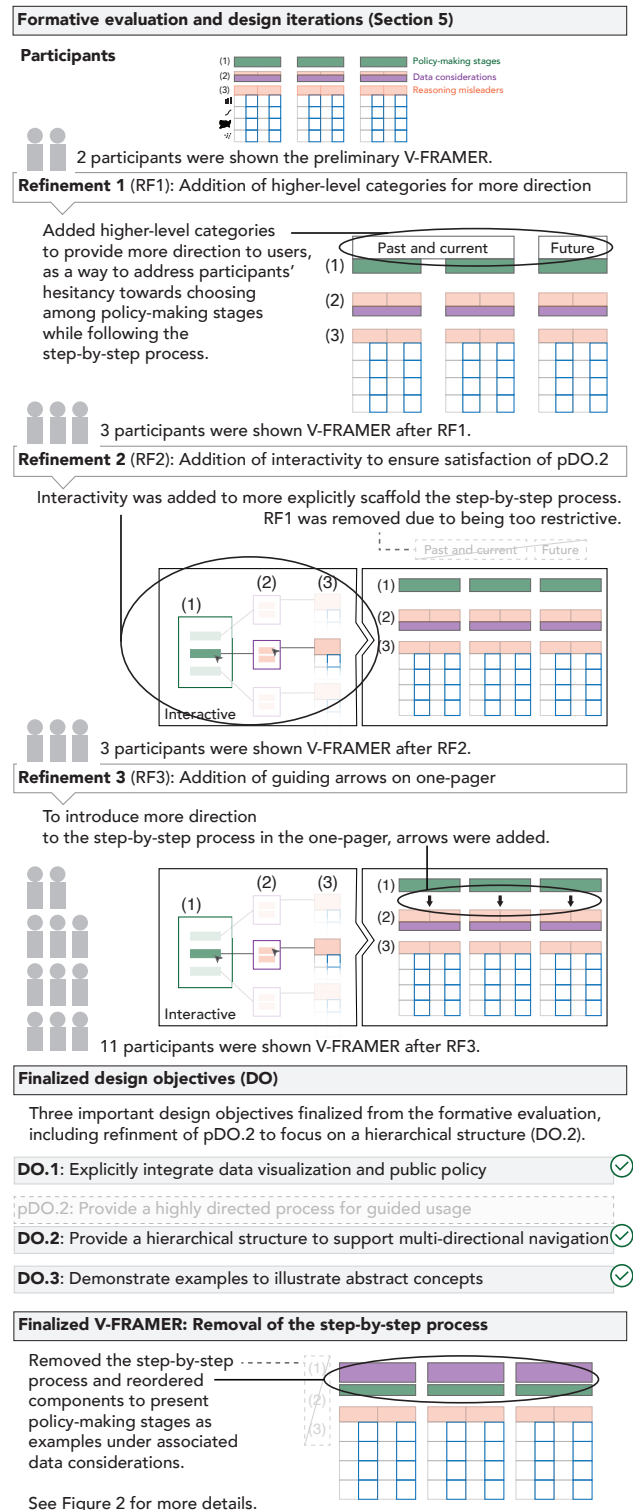


Figure 5: Iterative framework refinements informed by the formative evaluation with policy communicators. We finalized a set of design objectives, which informed the construction of the final V-FRAMER.

demonstrated relevance to our participants' examples, *Comparison Basis* and *Data Representativeness* seemed to be especially applicable. Consequently, many participants brought up examples that considered the reasoning misleaders *Missing Normalization* and *Inappropriately Aggregating*, which makes sense, since many worked with geographical data that involved different subgroups that required making relative comparisons under a comparable basis. In relation to *Data Representativeness*, some participants mentioned some data collection challenges. Specifically, P4 mentioned that it “oftentimes [is] harder to get [smaller jurisdictions] to respond to our survey, because they have much smaller teams... so that missing data, especially this concept of representativeness is a real challenge.” The reasoning misleader *Cherry Picking* has also been frequently considered as a potential challenge. For instance, P1, who worked in public health, mentioned that their graphs showing COVID-19 trends usually “begin in March, 2020, so you can see the actual whole entire trend, instead of breaking down like this week... then this week, that doesn't really show you a good comparison.”

5.2.2 Congruence of V-FRAMER with practitioners' considerations in policy-making stages.

We were able to map participants' examples from both **before** and **after** the introduction of V-FRAMER to the *Policy-making Stages*, and all of the stages contained at least one participant example. However, compared to the congruence described in Section 5.2.1, there were noticeably more instances where participants' examples fit into a policy-making stage but did not fit well with its associated data consideration and reasoning misleaders, or vice versa. Particularly, for the policy-making stage *Is the problem size worsening and at what speed?* and its associated data considerations and reasoning misleaders, there were examples of projections that mapped well. However, there were also instances where participants were more interested in a past trend for the problem and did not look at distributions, or evaluating a current state of the problem based on a sample of the population and did not necessarily need to evaluate how the problem changed. This was also apparent from participants' remarks during the interviews. Comments from the rest of the participants (after RF3) aligned with the earlier comments regarding the policy-making stages—although most participants were able to choose the stage(s) most relevant to their own examples, some still had difficulties immediately making a clear connection. For instance, P12 expressed that “we do all of these [policy-making stages] sort of at different time points.” This suggests that the step-by-step process originally introduced to satisfy pDO.2, rather than providing more guidance, may be too restrictive to users.

Even though some participants did not immediately fit their example into one of the stages, comments on the general relevance suggest that the stages were still consistent with their considerations. In particular, P11 stated that “we are asking these questions similarly, but in a way that is like a little bit more specific to... the context in which we are looking at.” This general consistency suggests that the policy-making stages should not simply be removed from the framework. However, the non-perfect correspondence of the stages for some of our participants does suggest that we should make the stages less prescriptive. The less prescriptive stages should also not unnecessarily break designers' workflows or their own conceptualizations of a particular policy problem.

Thus, also considering the value of the visual examples described in DO.3 above, **we refined pDO.2 (provide a highly directed process for guided usage) with a new DO.2: provide a hierarchical structure to support multi-directional navigation.** The hierarchical structure includes better support for starting with the examples table to gain a better understanding (“example-based accessing”). To satisfy the refined DO.2, we first merged the original steps 2 and 3 from the preliminary V-FRAMER to reduce redundancy. Then, we swapped the policy-making stages with data considerations to offer the stages as examples (Figure 5 *Finalized V-FRAMER*). This was to more clearly indicate that, although relevant, the stages should only be considered as examples and may not perfectly correspond to specific conceptualizations of policy problems. With these refinements, V-FRAMER satisfied DO.2.

Overall, our integration of visualization-related categories and policy-related categories have facilitated meaningful connections between the two fields, as seen through the interviews. Several participants also pointed out additional challenges in policy communication, such as the lack of standards in industry. This further necessitates a standardized, actionable synthesis of guidelines that is also easily accessible to policy communicators. Thus, **we retain pDO.1 that V-FRAMER satisfied as DO.1: explicitly integrate data visualization and public policy.** This hierarchical structure and the final V-FRAMER were described in Section 3.

6 HOW MIGHT V-FRAMER BE INTEGRATED INTO EXISTING WORKFLOWS?

We identified three salient potential integrations of V-FRAMER: (1) as a checklist, (2) brainstorming tool, and (3) educational tool. We also describe an example demonstration of use for each, which was inspired by the ways in which our participants interacted with V-FRAMER during the formative evaluation.

6.1 As a Checklist

Over half of the participants commented on the potential of using V-FRAMER to assist in data quality checks. For instance, P1 noted that “[the framework] has a lot of the key things that we need to take a look at before anything goes out.” P4 remarked “I think once I read it, a lot of things clicked in my mind of challenges we address. I don't think I oftentimes think about all the challenges at once.” Many participants expressed strong needs in going through a more systematic check before releasing information to the public. Our final DO.2, which focuses on hierarchical structuring, also supports V-FRAMER's utility as a checklist. Users can freely access components in the hierarchy as they perform quality checks, such as starting with the data considerations (top-down) or the examples table (bottom-up, example-based accessing).

Demonstration of Use Imagine designer-*X* who is examining the health of the economy and has already created a map visualization that shows the absolute number of people unemployed in each state. Before publishing the map, designer-*X* does a quality check of the created visualization. Scanning through the visual examples in the table on V-FRAMER, designer-*X*'s attention is caught by the contrast between the completely opposite impressions given by the bar charts under *Missing Normalization*. Upon further examination, designer-*X* realizes that one chart is showing the number

of people in the hospital that are vaccinated, while the *Better* version shows the rate. Designer-X then reasons through why this is the case: it makes sense that there are more people in the hospital that are vaccinated, because the majority of the population is already vaccinated. Looking to make comparisons between states with different working populations, designer-X draws connections to the example policy-making stage *Is the problem size different for subgroups?* and the data consideration *Comparison Basis*. Understanding the issue, designer-X corrects the map visualization by showing unemployment rates instead.

6.2 As a Brainstorming Tool

The second-most frequently mentioned potential integration is using V-FRAMER in the brainstorming process, before finalizing a design for a visualization. For example, P5 said the framework “*would be useful in [the] first iteration of making a visual*”. P12 remarked that the framework could help “think through other ways to visualize” when communicating data and mentioned that their team would have brainstorming sessions particularly about how to show uncertainty. Our final DO.3 especially supports this observation of potential V-FRAMER integration in the brainstorming process. The visualization examples table could create points of discussion as they actively think about how to best avoid the reasoning misleaders, which can further inspire alternative examples or ways of mitigation.

Demonstration of Use Imagine designer-Y who is looking to show the change in exam scores for a local school but has not created a visualization yet. After first thinking about showing the change in mean exam scores, designer-Y discusses with the team about whether or how to show that there is a distribution of exam scores rather than just a mean value. Someone suggested using error bars. Using V-FRAMER in their data meeting, the team looks through the examples for *Concealing Uncertainty* and *Inadequately Representing Uncertainty* reasoning misleaders under *Distributions*. The visualization examples prompt them to rule out only showing mean values and brainstorm ways of showing uncertainty other than using error bars. They first try to use gradients like the examples shown under *Distributions*, and then discuss other visualization types for distributional representations such as violin plots, swarm plots, or ridgeline plots. Ultimately, the team decides to use swarm plots as they show all of the underlying data points.

6.3 As an Educational Tool

Several participants also commented on the potential for V-FRAMER to assist in training more junior analysts. For instance, P14 pointed out that “*having this information at hand is really helpful, especially for younger analysts who are joining the team and might be taking over work... it’s just like a reminder for best practice*.” Besides training others, it could also be applicable in self-learning contexts to strengthen skill sets. Namely, P5 pointed out that “*I think a lot of the utility is just in consciously having to articulate things that I kind of assume that I’m doing and thinking*.” Thinking about using it as a way to practice the related concepts, they expressed that “*it’s useful to me to be trained in this... Here’s a set of questions that are really important to ask yourself. Go through and practice it*.” Our final DO.1 and DO.3 support this potential integration. The inclusion of

policy-making stages could help designers connect visualization techniques to the context they are working in, and the visualization examples table is helpful in explaining abstract concepts like the reasoning misleaders, which could assist in understanding.

Demonstration of Use Imagine designer-Z, a junior analyst studying public policy, is taking a class on policy communication. The instructor gives students a lab assignment in which each student gets a different policy question. Using V-FRAMER, the students must use the reasoning misleaders to construct a misleading visualization that purports to answer their policy question. Students then pair up and exchange their misleading examples. Within pairs, students must identify the reasoning misleader in their partner’s example and propose fixes for it using V-FRAMER. Designer-Z is assigned a question about how global temperature changes over time. They scan through the reasoning misleader categories, spotting two examples of line charts listed under *Cherry Picking*. The correct example uses a time range that conveys enough context to show an increasing overall trend. To make a misleading chart for their data, designer-Z does the opposite, visualizing a short time frame in which the temperature stays generally constant. Designer-Z then exchanges their example with designer-V. They fix each other’s misleading examples by first using V-FRAMER to narrow down the relevant reasoning misleader for the associated policy question, then using the examples to come up with solutions. The instructor provides feedback on the correctness and quality of the fixes. This engaged process (i.e., active learning [44]) of actively reflecting, identifying related reasoning misleaders, fixing the issues, then receiving feedback helps students grasp related concepts and assess their own understanding.

7 DISCUSSION

7.1 In Pursuit of Causality

Recall that even though deemed out of scope, “inferring unsupported causation” was included as a potential discussion point in the interviews in case participants had thoughts regarding causality (see footnote⁴). Since that was not the focus of the interviews, most participants did not comment on the issue of inferring unsupported causation. For those that did, some participants’ comments suggested that they are typically not the ones trying to communicate what drove a policy solution but focused on communicating the resulting policy solution instead. Others commented that some of the tools they created or the data they presented were intended to help people drive their own policy decision, so they do not try to communicate a particular cause of the problem. From the limited information we observed during the interviews, it seemed that considerations regarding causality may be more relevant during internal communication (e.g., to determine what factors caused the problem). Although prior work in the visualization community have studied causal support and how visual displays may influence viewers’ causal conclusions (e.g., [26, 27, 72]), there were not enough evidence to conclude that communication of causality is a primary goal when the policy communicator’s audience is the general public. However, we acknowledge the importance of causal inference in making policy prescriptions, and future studies that focus more on this aspect of policy communication could expand the scope of such communication frameworks.

7.2 (Dis)aggregation and Data Privacy

Many of our participants mentioned that they work with census data, which can raise data privacy concerns. For example, disaggregating too much may put certain groups at risk of privacy issues. Yet, by recommending against *Inappropriately Aggregating*, V-FRAMER may exacerbate such issues—in fact, inappropriately *disaggregating* is a concern when taking privacy into account. This highlights the complex nature of some reasoning misleaders—other context-dependent considerations may interact in complex ways with concerns about misleadingness. There is no simple fix: e.g., aggregating more may lead to privacy-preserving displays that are potentially misleading; aggregating less may lead to privacy-violating displays that may be more accurate. Text complementing visualizations, as a medium for providing more context (e.g., annotations), has been previously studied and found to add value in interpretation [62], and we also expect explanatory text to assist in a viewer’s reasoning. Specifically, the addition of context through explanatory text could potentially help alleviate some of the concerns like data privacy. Future work can start with our data consideration and reasoning misleader sets to identify the more context-dependent ones to draw out the interplay between context and visualization guidelines, which can help with the potential extensions of V-FRAMER or the development of new guidelines.

7.3 An Anti-uncertainty Feedback Loop

During the formative evaluation, uncertainty is generally perceived as important, but not typically conveyed in public policy. This is for several reasons seen during our interviews and in prior work [23], including: (1) communicators seemed to believe that uncertainty would be harder for the audience to interpret; for instance, P5 said that “*I operate off the assumption that people aren’t going to take the time to, or don’t want to, or are not going to look at what we are showing, and then want to consider margins of error*”, and (2) limited skills on the team to convey uncertainty; for instance, P4 remarked that it “*is hard for our team... to try to figure out how to appropriately map uncertainty, especially when we are primarily communicating to non data experts.*” Although the reasoning misleaders related to uncertainty were not discussed as much as the others, our participants largely agreed that uncertainty is important to consider—it just may be harder to interpret or convey. Thus, we still think *Distributions* and its associated reasoning misleaders are valuable parts of V-FRAMER and should not simply be removed. However, help is limited if the team lacks the desire or necessary skills to follow what is outlined in V-FRAMER.

We point out a negative feedback loop that hinders uncertainty consideration. Several participants mentioned that uncertainty is often not shown to their audience under the assumption that it would be hard to understand. This assumption would lead to less practice in conveying uncertainty, and the lack of practice can ultimately lead to not having the desired skill sets to convey uncertainty. The lack of skill sets then leads to not being able to adequately convey uncertainty to their audience. This feedback loop also unconsciously trains the audience to not expect uncertainty, subsequently leading people to be unfamiliar with uncertainty depictions.

Breaking this feedback loop can greatly advance efforts in conveying uncertainty to the general public. One effort in achieving

this goal may be improving the general public’s *uncertainty literacy* in visualizations. Another route is to tackle it from the designers’ side. Only a framework presenting the considerations and techniques that go into avoiding such reasoning misleaders may not be enough. Further investigations should facilitate uncertainty communication by supporting teams that might realize the importance but do not have access to the necessary skill sets to do so.

7.4 Feasibility of Perceived “Neutrality”

During data analysis, we also looked at what underlying policy communication goals were relevant and considered important to our participants. One goal that emerged was “not providing recommendations to policy makers”, but instead aiming to help policy makers make informed decisions. A driving motivation for this particular goal seemed to be the need to remain neutral, not privileging one particular policy option over another. This emphasis on “informing not recommending” seem to largely support the use of tools such as dashboards that enable users to interactively explore and filter data to assist in understanding. However, this notion of “neutrality” glaringly contradicts prior conversations in the visualization community on whether data or visualizations can be neutral. Such discussions have repeatedly pointed out that the data generating process is necessarily biased, as data itself is not a naturally occurring phenomenon [10]. Additionally, by the nature of the visualization construction process, the designer has to make choices about data representations, which can affect viewers’ interpretations [32]. However, some of our participants’ responses seem to suggest that there is still a perceived “neutrality” that may be impossible to achieve, in which case it is crucial to raise the awareness of the inevitable non-neutrality of data visualizations. This may require future explorations to expand the reasoning misleaders set to account for steps that even precede the data considerations to include *data generation* misleaders, which might help surface these tensions around the feasibility of “neutrality” in visualizations.

7.5 Limitations and Future Work

Using *Factfulness* [52] as one of the initial data sources for the preliminary framework construction has certain limitations. Although *Factfulness* offers practical insights that go beyond the typical visualization guidelines on grammatical visual components, it does rely on one expert’s experiences. However, our before-and-after comparisons from the formative evaluation suggest that V-FRAMER does cover important considerations in practice. Specifically, each data consideration generally mapped to its associated reasoning misleaders, and none of the participants mentioned an obviously missing category before or after seeing the framework. There was also a general consistency between the set of example policy-making stages on V-FRAMER and the set of policy-relevant problems across the participants. Future research could investigate alternative starting points, such as a systematic review of wider collections of already-made, effective visualizations in the wild, which may lead to organizations of data considerations that differ in granularity.

Grounding the framework in public policy with a participant pool based in one country has inevitable limitations on the framework’s applicability to a wider community and to domains with similar data considerations. However, this focus allowed us to deepen the

discussion around policy communicators. While we recruited in one country, we did put effort into curating diverse perspectives, which can come from the policy problems participants work on rather than the participants' geographical location. Expanding the framework to a more general audience would broaden the impact but does not necessarily affect the aim of the current V-FRAMER, which is specifically developed for and with a specialized audience. Future empirical studies can investigate ways of maximizing the framework's utility, starting with the potential integrations of V-FRAMER (Section 6). It may also be interesting to deploy V-FRAMER to document its actual usage over time, which could offer alternative insights for improving and expanding the framework.

8 CONCLUSION

We contribute V-FRAMER, a framework iteratively co-designed through a formative evaluation with 19 policy communicators. Situated in public policy, V-FRAMER explicitly aims to cover data visualization considerations important in avoiding reasoning misleaders in practice—a class of issues that can misinform viewers even in grammatically-correct visualizations. V-FRAMER's hierarchical components include a set of data considerations, each accompanied by an example policy-making stage to provide additional context. V-FRAMER also includes a table of visualization examples comprising reasoning misleaders and different chart types. Our findings indicate that these visualization examples are useful for making the abstract concept of reasoning misleaders more tangible. This further informs our recommendations for potential framework integrations to support different use cases and workflows, such as a data quality checklist or an educational tool for junior analysts. We hope our framework begins to lay the foundation for improved data-based reasoning with visualizations, going beyond the need for mere grammatical correctness in visualization design.

ACKNOWLEDGMENTS

Along with the Northwestern MU Collective, the Design Cluster program at the Center for HCI and Design, and the Visual Thinking Lab, we extend our gratitude to Amritha Anupindi, Elizabeth Burslem, Mandi Cai, Yuan Cui, Elizabeth Durango-Cohen, Maryam Hedayati, Hyeok Kim, Taewook Kim, Sheng Long, Abhraneel Sarma, Jon Schwabish, Elizabeth Tipton, Fumeng Yang, and Eric Zaslow for their feedback on this work. We thank the participants for their time and the anonymous reviewers for their helpful comments. This work was supported in part by a grant from the National Science Foundation (IIS-1901485). The following statements are included by Ge, in accordance with the NSF Graduate Research Fellowship Program Administrative Guide (NSF 23-075): This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2234667. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- [1] Shaaron Ainsworth. 2006. DeFT: A conceptual framework for considering learning with multiple representations. *Learning and instruction* 16, 3 (2006), 183–198.
- [2] John Alford and Brian W Head. 2017. Wicked and less wicked problems: a typology and a contingency framework. *Policy and Society* 36, 3 (2017), 397–413.

- [3] Mohammed Ibrahim Alhojailan and Mohammed Ibrahim. 2012. Thematic analysis: A critical review of its process and evaluation. *West east journal of social sciences* 1, 1 (2012), 39–47.
- [4] Eugene Bardach and Eric M Patashnik. 2019. *A practical guide for policy analysis: The eightfold path to more effective problem solving*. CQ press, California, USA.
- [5] Jacques Bertin. 1983. *Semiology of Graphics*. University of Wisconsin Press, Wisconsin, USA.
- [6] Kenneth Broad, Anthony Leiserowitz, Jessica Weinkle, and Marissa Steketee. 2007. Misinterpretations of the “Cone of Uncertainty” in Florida during the 2004 Hurricane Season. *Bulletin of the American Meteorological Society* 88, 5 (2007), 651 – 668. <https://doi.org/10.1175/BAMS-88-5-651>
- [7] CDC. 2021. U.S. COVID-19 Cases in the Last 7 Days by State and Territory. <https://twitter.com/CDCgov/status/1387509465013006336>.
- [8] Qing Chen, Fuling Sun, Xinyue Xu, Zui Chen, Jiazhe Wang, and Nan Cao. 2022. VizLinter: A Linter and Fixer Framework for Data Visualization. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2022), 206–216. <https://doi.org/10.1109/TVCG.2021.3114804>
- [9] William S Cleveland and Robert McGill. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association* 79, 387 (1984), 531–554.
- [10] Michael Correll. 2019. Ethical Dimensions of Visualization Research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300418>
- [11] Michael Correll, Enrico Bertini, and Steven Franconeri. 2020. Truncating the Y-Axis: Threat or Menace?. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3313831.3376222>
- [12] Michael Correll and Michael Gleicher. 2014. Error Bars Considered Harmful: Exploring Alternate Encodings for Mean and Error. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2142–2151. <https://doi.org/10.1109/TVCG.2014.2346298>
- [13] Richard Cox. 1996. *Analytical reasoning with multiple external representations*. Ph.D. Dissertation. University of Edinburgh UK.
- [14] Marian Dörk, Patrick Feng, Christopher Collins, and Sheelagh Carpendale. 2013. Critical InfoVis: Exploring the Politics of Visualization. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) (CHI EA '13). Association for Computing Machinery, New York, NY, USA, 2189–2198. <https://doi.org/10.1145/2468356.2468739>
- [15] Satu Elo and Helvi Kyngäs. 2008. The qualitative content analysis process. *Journal of Advanced Nursing* 62, 1 (2008), 107–115. <https://doi.org/10.1111/j.1365-2648.2007.04569.x>
- [16] Stephen Few. 2019. *The Data Loom: Weaving Understanding by Thinking Critically and Scientifically with Data*. Analytics Press, California, USA.
- [17] FiveThirtyEight. 2022. Data Visualization – FiveThirtyEight. <https://fivethirtyeight.com/tag/data-visualization/>.
- [18] Steven L. Franconeri, Lace M. Padilla, Priti Shah, Jeffrey M. Zacks, and Jessica Hullman. 2021. The Science of Visual Data Communication: What Works. *Psychological Science in the Public Interest* 22, 3 (2021), 110–161. <https://doi.org/10.1177/15291006211051956> PMID: 34907835.
- [19] Lily W. Ge, Yuan Cui, and Matthew Kay. 2023. CALVI: Critical Thinking Assessment for Literacy in Visualizations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, 1–18. <https://doi.org/10.1145/3544548.3581406>
- [20] Lane Harrison, Fumeng Yang, Steven Franconeri, and Remco Chang. 2014. Ranking visualizations of correlation using weber's law. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 1943–1952.
- [21] Aspen K. Hopkins, Michael Correll, and Arvind Satyanarayan. 2020. VisualLint: 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. 1954. *How to Lie with Statistics*. 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] Christopher H Jackson. 2008. Displaying Uncertainty With Shading. *The American Statistician* 62, 4 (2008), 340–347. <https://doi.org/10.1198/000313008X370843>
- [25] Bruce S Jansson. 2017. *Empowerment series: Becoming an effective policy advocate*. Cengage Learning, Boston, MA, USA.
- [26] Alex Kale, Yifan Wu, and Jessica Hullman. 2021. Causal support: Modeling causal inferences with visualizations. *IEEE transactions on visualization and computer graphics* 28, 1 (2021), 1150–1160.
- [27] Smiti Kaul, David Borland, Nan Cao, and David Gotz. 2021. Improving visualization interpretation using counterfactuals. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 998–1008.

- [28] David Kirsh. 2010. Thinking with external representations. *AI & society* 25, 4 (2010), 441–454.
- [29] Yngve S. Kristiansen, Laura Garrison, and Stefan Bruckner. 2021. Semantic Snapping for Guided Multi-View Visualization Design. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 43–53. <https://doi.org/10.1109/TVCG.2021.3114860>
- [30] Crystal Lee, Tanya Yang, Gabrielle D 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 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 607, 18 pages. <https://doi.org/10.1145/3411764.3445211>
- [31] Sukwon Lee, Sung-Hee Kim, and Bum Chul Kwon. 2017. VLAT: Development of a Visualization Literacy Assessment Test. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 551–560. <https://doi.org/10.1109/TVCG.2016.2598920>
- [32] Elsie Lee-Robbins and Eytan Adar. 2023. Affective Learning Objectives for Communicative Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 1–11. <https://doi.org/10.1109/TVCG.2022.3209500>
- [33] Maxim Ljusic, 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* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, 21 pages. <https://doi.org/10.1145/3544548.3580910>
- [34] Le Liu, Alexander P. Boone, Ian T. Ruginski, Lace Padilla, Mary Hegarty, Sarah H. Creem-Regehr, William B. Thompson, Cem Yuksel, and Donald H. House. 2017. Uncertainty Visualization by Representative Sampling from Prediction Ensembles. *IEEE Transactions on Visualization and Computer Graphics* 23, 9 (2017), 2165–2178. <https://doi.org/10.1109/TVCG.2016.2607204>
- [35] Le Liu, Lace Padilla, Sarah H. Creem-Regehr, and Donald H. House. 2019. Visualizing Uncertain Tropical Cyclone Predictions using Representative Samples from Ensembles of Forecast Tracks. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 882–891. <https://doi.org/10.1109/TVCG.2018.2865193>
- [36] 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>
- [37] Giorgia Lupi. 2017. Data humanism: the revolutionary future of data visualization. <https://www.printmag.com/article/data-humanism-future-of-data-visualization/>.
- [38] Jock Mackinlay. 1986. Automating the Design of Graphical Presentations of Relational Information. *ACM Trans. Graph.* 5, 2 (apr 1986), 110–141. <https://doi.org/10.1145/22949.22950>
- [39] Jock Mackinlay, Pat Hanrahan, and Chris Stolte. 2007. Show Me: Automatic Presentation for Visual Analysis. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1137–1144. <https://doi.org/10.1109/TVCG.2007.70594>
- [40] Kathleen M. MacQueen, Eleanor McLellan, Kelly Kay, and Bobby Milstein. 1998. Codebook Development for Team-Based Qualitative Analysis. *CAM Journal* 10, 2 (1998), 31–36. <https://doi.org/10.1177/1525822X980100020301>
- [41] Niall McCarthy. 2021. Confirmed COVID-19 Cases in the U.S. <https://www.statista.com/chart/20978/coronavirus-cases-us-map/>.
- [42] Andrew McNutt and Gordon Kindlmann. 2018. Linting for visualization: Towards a practical automated visualization guidance system. https://c4pgv.dbvis.de/McNutt_Kindlmann_2018.pdf. VisGuides: 2nd Workshop on the Creation, Curation, Critique and Conditioning of Principles and Guidelines in Visualization.
- [43] Andrew McNutt, Gordon Kindlmann, and Michael Correll. 2020. Surfacing Visualization Mirages. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3313831.3376420>
- [44] Joel Michael. 2006. Where's the evidence that active learning works? *Advances in Physiology Education* 30, 4 (2006), 159–167. <https://doi.org/10.1152/advan.00053.2006> PMID: 17108243.
- [45] Dominik Moritz, Chenglong Wang, Greg L. Nelson, Halden Lin, Adam M. Smith, Bill Howe, and Jeffrey Heer. 2019. Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 438–448. <https://doi.org/10.1109/TVCG.2018.2865240>
- [46] B Guy Peters. 2017. What is so wicked about wicked problems? A conceptual analysis and a research program. *Policy and Society* 36, 3 (2017), 385–396.
- [47] Washington Post. 2022. Washington Post | FlowingData. <https://flowingdata.com/tag/washington-post/>.
- [48] Ghulam Jilani Quadri and Paul Rosen. 2022. A Survey of Perception-Based Visualization Studies by Task. *IEEE Transactions on Visualization and Computer Graphics* 28, 12 (2022), 5026–5048. <https://doi.org/10.1109/TVCG.2021.3098240>
- [49] Alex Reinhart. 2012. Statistics Done Wrong.
- [50] Horst WJ Rittel and Melvin M Webber. 1973. Dilemmas in a general theory of planning. *Policy sciences* 4, 2 (1973), 155–169.
- [51] Bernice E Rogowitz and Lloyd A Treinish. 1996. Why should engineers and scientists be worried about color. *IBM Thomas J. Watson Research Center, Yorktown Heights, NY* (1996), 13 pages.
- [52] Hans Rosling, Ola Rosling, and Anna Rosling Rönnlund. 2018. *Factfulness: Ten Reasons We're Wrong About the World - and Why Things Are Better Than You Think*. Flatiron Books, New York, NY, USA.
- [53] Steven F. Roth, John Kolojechick, Joe Mattis, and Jade Goldstein. 1994. Interactive Graphic Design Using Automatic Presentation Knowledge. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA) (CHI '94). Association for Computing Machinery, New York, NY, USA, 112–117. <https://doi.org/10.1145/191666.191719>
- [54] Tobias Ruppert, Jens Dambusch, Michel Krämer, Tina Balke, Marco Gavaneli, Stefano Bragaglia, Federico Chesani, Michela Milano, and Jörn Kohlhammer. 2015. *Visual Decision Support for Policy Making: Advancing Policy Analysis with Visualization*. Springer International Publishing, Cham, 321–353. https://doi.org/10.1007/978-3-319-12784-2_15
- [55] Bahador Saket, Alex Endert, and Çağatay Demiralp. 2018. Task-based effectiveness of basic visualizations. *IEEE transactions on visualization and computer graphics* 25, 7 (2018), 2505–2512.
- [56] Abhraneel Sarma, Shunan Guo, Jane Hoffswell, Ryan Rossi, Fan Du, Eunyee Koh, and Matthew Kay. 2023. Evaluating the Use of Uncertainty Visualisations for Imputations of Data Missing At Random in Scatterplots. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 602–612. <https://doi.org/10.1109/TVCG.2022.3209348>
- [57] Jonathan Schwabish and Alice Feng. 2021. Do no harm guide: Applying equity awareness in data visualization. <https://www.urban.org/research/publication/do-no-harm-guide-applying-equity-awareness-data-visualization>.
- [58] Jon Schwabish and Severino Ribeca. 2022. The Graphic Continuum. <https://www.severinoribeca.one/portfolio-item/the-graphic-continuum/>.
- [59] Samuel Silva, Beatriz Sousa Santos, and Joaquim Madeira. 2011. Using color in visualization: A survey. *Computers & Graphics* 35, 2 (2011), 320–333.
- [60] Herbert A. Simon. 1996. *Social Planning: Designing the Evolving Artifact*. MIT Press, Cambridge, MA, 139–167.
- [61] Hayeong Song and Danielle Albers Szafir. 2019. Where's My Data? Evaluating Visualizations with Missing Data. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 914–924. <https://doi.org/10.1109/TVCG.2018.2864914>
- [62] Chase Stokes, Vidya Setlur, Bridget Cogley, Arvind Satyanarayan, and Marti A. Hearst. 2023. Striking a Balance: Reader Takeaways and Preferences when Integrating Text and Charts. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2023), 1233–1243. <https://doi.org/10.1109/TVCG.2022.3209383>
- [63] 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.
- [64] Financial Times. 2022. Visual Vocabulary. <https://ft-interactive.github.io/visual-vocabulary/>.
- [65] The New York Times. 2022. Graphics - The New York Times. <https://www.nytimes.com/spotlight/graphics>.
- [66] VisLies. 2017. Vis Lies 2017 Gallery. <https://www.vislies.org/2017/gallery/>.
- [67] VisLies. 2021. IEEE VIS VisLies Meetups. <https://www.vislies.org/2021/>.
- [68] VisLies. 2021. VisLies 2021 Gallery. <https://www.vislies.org/2021/gallery/>.
- [69] Zezhong Wang, Lovisa Sundin, Dave Murray-Rust, and Benjamin Bach. 2020. Cheat Sheets for Data Visualization Techniques. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376271>
- [70] Jeremy Bennet Wilmer and Sarah Horan Kerns. 2022. What's really wrong with bar graphs of mean values: variable and inaccurate communication of evidence on three key dimensions. <https://doi.org/10.31219/osf.io/av5ey>.
- [71] Kanit Wongsuphasawat, Dominik Moritz, Anushka Anand, Jock Mackinlay, Bill Howe, and Jeffrey Heer. 2016. Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 649–658. <https://doi.org/10.1109/TVCG.2015.2467191>
- [72] Cindy Xiong, Joel Shapiro, Jessica Hullman, and Steven Franconeri. 2020. Illusion of Causality in Visualized Data. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2020), 853–862. <https://doi.org/10.1109/TVCG.2019.2934399>
- [73] Jiajie Zhang. 1997. The nature of external representations in problem solving. *Cognitive science* 21, 2 (1997), 179–217.