Visualization Guardrails: Designing Interventions Against Cherry-Picking in Interactive Data Explorers

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

Zach Cutler
zcutler@sci.utah.edu
University of Utah
Salt Lake City, Utah, USA

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

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

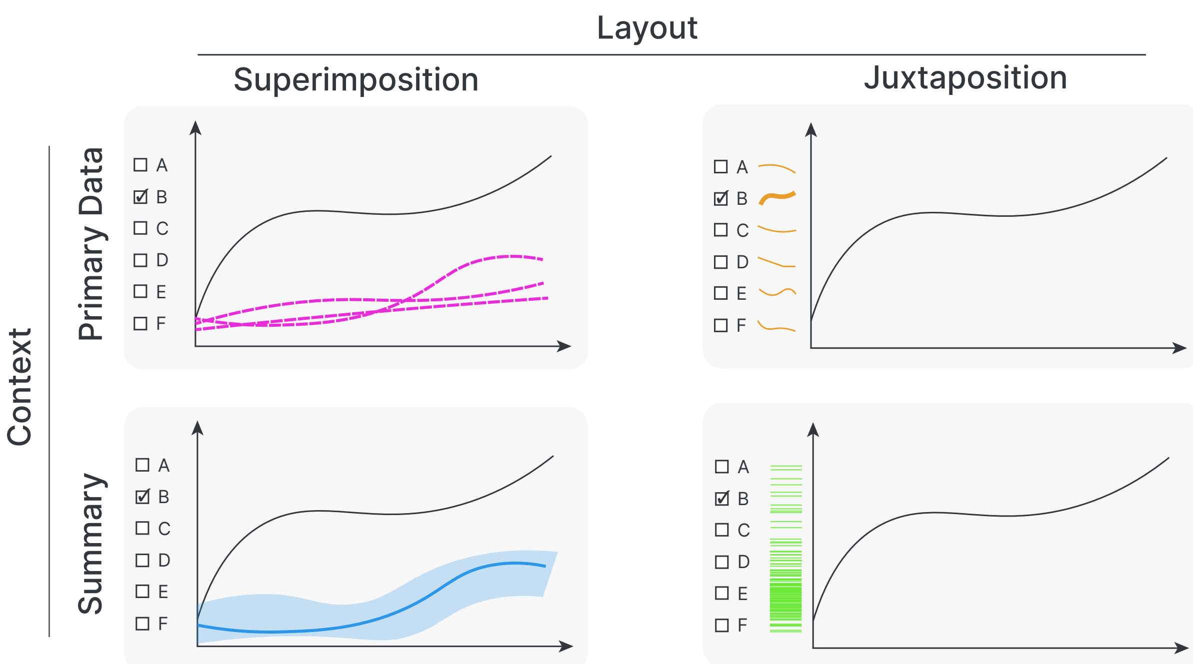


Figure 1: The design space of visualization guardrails against cherry-picking along two dimensions: what *context* is shown (primary data or a summary) and *layout*, or where it is shown (superimposed on or juxtaposed with the main chart). The figure shows the *main chart data* in black and the *context* in color. In all example charts, the user selects only one item (B) which, as shown by the guardrails, is an outlier in the data set.

Abstract

The growing popularity of interactive time series exploration platforms has made data visualization more accessible to the public. However, the ease of creating polished charts with preloaded data also enables selective information presentation, often resulting in biased or misleading visualizations. Research shows that these tools have been used to spread misinformation, particularly in areas such as public health and economic policies during the COVID-19 pandemic. Post hoc fact-checking may be ineffective because it typically addresses only a portion of misleading posts and comes too late to curb the spread. In this work, we explore using visualization design to counteract cherry-picking, a common tactic in deceptive visualizations. We propose a design space of *guardrails*—interventions to expose cherry-picking in time-series explorers. Through three crowd-sourced experiments, we demonstrate that guardrails, particularly those superimposing data, can encourage skepticism, though with some limitations. We provide recommendations for developing more effective visualization guardrails.

CHI '25, Yokohama, Japan

The definitive Version of Record was published in *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan*, https://doi.org/10.1145/3706598.3713385.

CCS Concepts

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

Keywords

Visualization, cherry-picking, general public visualizations, misinformation interventions.

ACM Reference Format:
Maxim Lisnic, Zach Cutler, Marina Kogan, and Alexander Lex. 2025. Visualization Guardrails: Designing Interventions Against Cherry-Picking in Interactive Data Explorers. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3706598.3713385

1 Introduction

Open data exploration platforms democratize access to data and visualizations of public importance. Examples include COVID-19 case dashboards on OurWorldInData [43], stock or cryptocurrency performance charts on Yahoo! Finance [74], or graphing platforms for various economic and policy indicators, such as Gapminder [58] or FRED [22]. Users not only create charts on these platforms for their own use, but also commonly share them on social media. While the popularity of data exploration platforms is a testament to their utility, the unconstrained and often unguided selection of data subsets and time periods that is commonly featured in such tools can lead to conclusions made based on *cherry-picked data*. Such data visualizations, when shared on social media, are often misrepresented in a way that supports wrong and, in the worst case, harmful arguments. Prior research has also shown that cherry-picking of items and time frames is an issue that contributed to misinformation arguments in over 40% of COVID-skeptic charts shared on Twitter, most of which were screenshots of data explorers [40].

It is important to acknowledge that the problem of cherry-picking in public-facing data explorers is a *wicked problem* [57]. Biased interpretations of data, and information in general, are entangled with the viewers' data and visualization literacy, data production quality, social and political movements, or the sense of belonging to a group with a strong ideology. This problem is hence wicked in the sense that, due to its complexity, it is resistant to any single solution. Rather, it requires continuous adaptation and innovation of approaches and interventions. In this work, we propose adding a novel type of intervention to the arsenal of tools against data-driven misinformation.

The most prevalent interventions for addressing misleading information focus on fact-checking, whether conducted by crowdsourcing or professional organizations [1]. However, these post hoc strategies come with significant limitations. Correcting every misleading post is nearly impossible, and unchecked misinformation can result in the implied truth effect, where uncorrected content is perceived as accurate [51]. Furthermore, because post hoc corrections and moderation inherently occur after the fact—sometimes delayed by several days [45]—the original misinformation often spreads widely before any correction can be made. Research also suggests that corrections are more effective when delivered by the same source that initially presented the misinformation [68] as opposed to other sources that may be dismissed as biased, indicating the value of enhancing the original content with contextual information. Prior research has found that data-driven misinformation arguments on social media, including cherry-picked data visualizations, often remain unaddressed by existing fact-checking interventions due to being rooted in factually accurate data and thus not being wrong, or viewed as personal opinions not worthy of debunking [39].

In this paper, we set out to explore the possibility of ante hoc interventions tackling cherry-picking in time series explorers from the perspective of visualization design, attempting to curb the issue before it has a chance to spread. We asked ourselves: **is there anything we can do when designing and implementing data exploration platforms to minimize misuse, misunderstandings, and misinterpretations?** To answer this question, we propose looking at the problem of misinterpretation and misuse of data explorers through the lens of threat modeling [59], and explore the possible interventions against cherry-picking in interactive visualizations. As one such intervention, we introduce the concept of visualization guardrails. Analogous to protective highway guardrails, which are useful when losing control of a vehicle but no hindrance in regular operation, **visualization guardrails automatically show the contextual data that expose cherry-picking, but do not interfere with interpretation if no cherry-picking is present**. We describe the ways contextual data could be visualized and integrated into existing charts by outlining the design space of guardrails.

We break down the issue of misinformative charts into two distinct yet equally important problems. Firstly, the *production problem*: data exploration interfaces make it very easy to create (perhaps, even nudge authors toward creating) cherry-picked views. The circumstances that may lead an individual to creating a cherry-picked chart can vary from a desire to maliciously misinform, to innocent ignorance, or even to a genuine reason to be interested in zooming in on a certain subset of data. In any case, however, it should be more difficult to end up with a potentially misleading view.

The second problem is the *reaction problem*: when the resulting view is then shared with others through social media, it may end up spreading the incorrect insight and convincing others. Moreover, charts created with data exploration platforms are typically adorned with a veneer of impartiality and reputability offered by the data exploration platform's logo and recognizable design, and hence may seem authoritative [39]. For an example, see the cherry-picked charts in the style of the Financial Times and Our World in Data in Figure 2, both reputable sources. Therefore, another major design goal for the guardrails is to introduce more nuance to views created with such reputable platforms.

To restate, our research questions are:

- What is the **design space** of guardrails against cherry-picking in data visualizations?
- Can we design guardrails that make cherry-picked charts **harder to produce**?

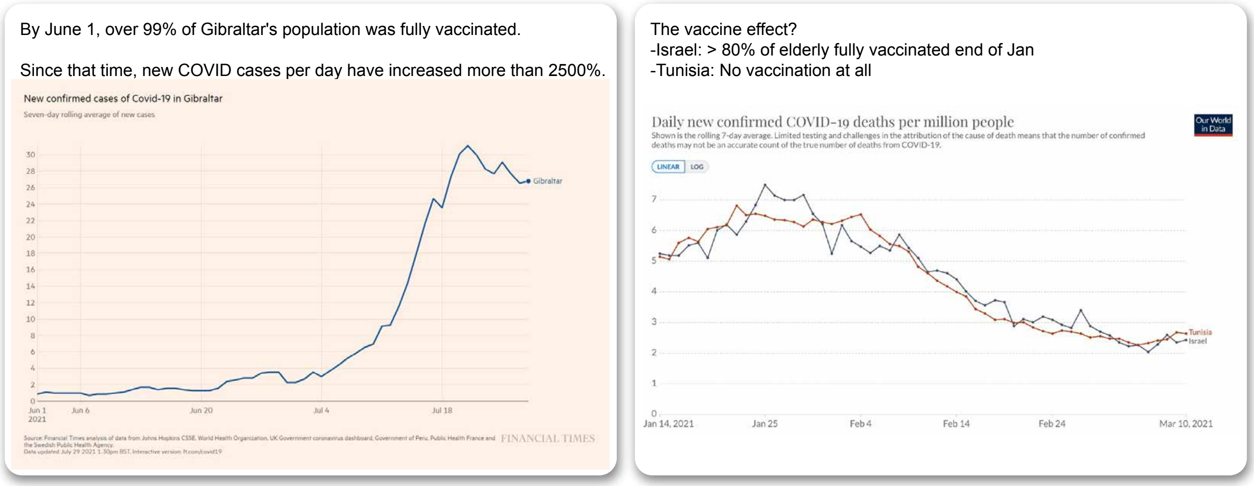


Figure 2: Examples of tweets that spread vaccine hesitancy using screenshots of time series explorers with cherry-picked data. The tweet on the left used the Financial Times explorer to show an increase in cases in a single highly-vaccinated country (Gibraltar), implying that vaccines are harmful. The tweet on the right used the OurWorldInData explorer [43] to show two countries of different vaccination levels with similar trends of deaths (Israel and Tunisia), implying that vaccines are ineffective.

- Can we design guardrails that make cherry-picked charts less convincing and lead to a **more skeptical reaction** from the audience?

In this work, we describe the design space of visualization guardrails and implement a prototype data explorer with four distinct guardrail instances. We limit our designs to line charts showing time-series data, which is a type of chart that is often used in debates on controversial topics (such as impacts of policy decisions on metrics of interest). Additionally, time-series line charts have been commonly used for cherry-picking in a social media context [40]. In order to investigate both the production and the reaction problems, we conducted three crowd-sourced experiments using our prototype of guardrails that target item cherry-picking, in which we attempt to mimic the real-world adversarial process of designing and sharing misinformative charts. In the first experiment, we challenged participants to create cherry-picked views using control and guardrail interfaces, thus red teaming our prototype. We then showed these charts to another set of participants in the second experiment and asked them to make a behavioral trust decision based on the chart, evaluating the resulting cherry-picking. We then conducted a third experiment to confirm our findings in a more controlled scenario in which we regulated the effect of cherry-picking egregiousness and misleading captions on guardrails. Our findings show that the **guardrails achieved the strongest desired effect of encouraging skepticism when guardrails are closest to the original visual language of the chart**. At the same time, guardrails are less effective when merely providing a clue about cherry-picking by showing a statistical summary or being plotted beside the main canvas. In such cases, a large part of the audience ignored the guardrails and instead focused on the main features of the visualization and the attached caption. Based on our results, we outline recommendations for designing effective guardrails against cherry-picking in data explorers.

To summarize, our paper makes several contributions:

- A novel conceptual framework for tackling issues of misuse and misinterpretations of data visualizations through the lens of threat modeling.
- A definition of the design space of visualization guardrails that protect against cherry-picking.
- A set of crowd-sourced user studies exploring the effects of guardrails on the production of and reaction to visualizations of cherry-picked data, and the resulting recommendations for implementing visualization guardrails in public-facing data explorers.

2 A Threat Modeling Framework for Visualizations

In this section, we propose adopting the concept of *threat modeling* from computer security as a way to approach the problem of data visualizations misused in support of misinformation. By applying the threat modeling framework to cherry-picking, we demonstrate that threat modeling is useful both as a mental model for surfacing vulnerabilities in data interfaces and as a practical guide for identifying appropriate interventions. We believe that this model is applicable to a variety of other problems as well, such as misinterpretations of statistics or incorrect causal inferences, and we urge researchers to explore such applications in future work.

2.1 What is Threat Modeling?

The Threat Modeling Manifesto, put forward by a group of security researchers aiming to promote security and privacy during software development, defines threat modeling as "analyzing representations of a system to highlight concerns about security and privacy characteristics" [9]. Threat modeling allows the researcher to identify things that can go wrong, pinpoint design issues, and inform mitigation measures.

In his book *Threat Modeling: Designing for Security* [59], Adam Shostack proposes a four-step framework for threat modeling, which involves answering four questions: (1) What are you building?, (2) What can go wrong?, (3) What should you do about it?, and (4) Did you do a decent job?

Example outcomes of threat modeling in the security context could be identifying opportunities for hackers to access confidential information in a data base, or designing interventions against denial-of-service attacks. However, although a number of specific computer security-related checklists or domain-specific heuristics threaten modeling itself, it is a value- and principle-driven approach. Therefore, it is highly adaptable to a wide variety of security and privacy issues, as well as (as we will argue below) issues of data and visualization misuse.

2.2 Applying Threat Modeling to Visualization Threats

Next, we go over the four steps of the framework with the goal of outlining the issue of misuse of interactive visualizations used by the general public to support misinformation, as well as motivating the application of threat modeling to this issue.

2.2.1 What Are You Building? For the purposes of this work, we focus on general public-facing interactive time series exploration portals created by local government agencies such as state and county epidemiology tracking dashboards, news organizations such as Yahoo! Finance [74], or specialized data exploration platforms such as OurWorldInData [43]. In this step, the data explorer platform governance should outline the core functionality and values of the platform that should be protected and not compromised during the threat modeling exercise. We assume that examples of such core functionality could include instant access to data for everyone, freedom of exploration without major restrictions, ability to export and share views and data, among others.

2.2.2 What Can Go Wrong? To answer this question, we can turn to previous work that outlined reasoning errors in social media users' interpretations of data visualizations [40]: cherry-picking favorable subsets of data, assigning causality to salient features of charts, or not accounting for common statistical fallacies. In this work, we focus on *cherry-picking*—one of the most often used tactics—as an illustrative example. Some of the core functionality in data explorers described in the previous section, such as access to data and freedom of exploration, result in the danger of making selective choices and emphasizing those results, pointing to a possibility of cherry-picking. Figure 2 shows two instances of tweets using cherry-picked data explorer charts to spread misinformation and to promote vaccine hesitancy.

2.2.3 What Should You Do About It? Shostack outlines four possible paths of action: (1) *accept* that there is an issue and do nothing, (2) *eliminate* the feature causing an issue, (3) *transfer* the responsibility to the user, or (4) *mitigate* the issue [59]. All four strategies are feasible for data explorer platforms. An example of acceptance is simple: one could do nothing. Eliminating a feature could take the form of restricting problematic interactions by, for instance, not allowing plotting any two time series on the same chart. Transferring responsibility to the user may involve requiring data or visualization literacy evaluations, or a checklist that assists a user in evaluating whether their selection is misleading, but may be unrealistic to realize in platforms accessible to the general public.

Although we urge platform designers to consider all of the above strategies when performing threat modeling of their tools, in this paper we will pursue the goal of *mitigation* of the cherry-picking threat and design *guardrails* against it. We argue that of the four possible actions, mitigation is the ultimate goal of misinformation interventions. Firstly, acceptance, or inaction against misinformation is not productive. Secondly, we believe that the benefits of the "problematic" features (i.e., unrestricted freedom of exploration and ability to take screenshots and share) are high, and therefore we opt to explicitly maintain such features and not pursue elimination. And lastly, in the adversarial context of misinformation, we posit that any transfer of responsibility strategy would be futile due to the high possibility of bad-faith actors purposefully seeking out cherry-picked views.

2.2.4 Did You Do a Decent Job? The goal of this last step of threat modeling is to evaluate the success of the chosen intervention strategy. In order to answer this question, we conducted three rounds of evaluation studies, described in Sections 6, 7, and 8. Firstly, we conducted a study challenging participants to produce cherry-picked views with and without our guardrails in an approach similar to using a Red Team in cybersecurity. This enabled us to conduct a second experiment in which a different set of participants were asked to review the charts produced in the first study. Lastly, we conducted a third, more controlled experiment, modulating the severity of cherry-picking in the chart.

3 Related Work

In this section, we first discuss the role of cherry-picking of information for the purposes of spreading misinformation. Then, we present an overview of existing work in data visualization on designing interventions against fallacies and cognitive biases.

3.1 Cherry-Picking and Questionable Research Practices

In his testimony before the US House of Representatives, climate scientist Richard Somerville described cherry-picking as "[making] selective choices among competing evidence, so as to emphasize those results that support a given position, while ignoring or dismissing any findings that do not support it" [60]. By analogy, in the context of visualization we define cherry-picking as **plotting data that support a given position, while not plotting data that do not.**

Cherry-picking favorable data or results has long been acknowledged as a questionable research practice across the scientific community, alongside HARKing (hypothesizing after the results are known) and p-hacking [2, 10]. Cherry-picking, or selectively showing only information that supports a given argument, is an outcome that could be unintentional and point to ignorant or inattentive practices, or be intentional and reflect malicious intent to misinform.

Unintentional cherry-picking is a common behavior in judgement under uncertainty and could stem from a variety of cognitive biases, such as availability bias (focusing only on information that is readily available) [64], confirmation bias (focusing on information that supports prior beliefs) [52], and anchoring bias (focusing on information presented first) [65]. **Intentional cherry-picking** is a strategy shown to support misinformation arguments about climate change denial [12], vaccine hesitancy [19], and is, more broadly, one of the hallmark tactics of denialism [31].

Cherry-picking is a practice that is not limited to data and visualizations. Quote-mining refers to cherry-picking a quote or a statement out of its original context [31]. Quoting others out of context has been practiced for a variety of purposes over the course of history, ranging from political propaganda [44] to misusing critics' quotes in misleading advertisements [56]. Both cherry-picking data and quote-mining statements can be described as sharing half-truths, meaning these tactics are difficult to debunk since they are based in truth [4]. It then follows that in order to combat the issue of focusing on a subset of information, an intervention must either raise awareness about or explicitly show or summarize the missing context.

3.2 Interventions Against Fallacies in Data Visualizations

Previous work on interventions against biases in visualization largely focuses on professional and scientific visual analytics, and specifically the forking paths problem [54], the multiple comparisons problem [76], inaccurate model specifications [34] , and imbalance across variables viewed [66]. Another potential source of bias in visual analytics arises when analysts fail to fully account for the uncertainty in their estimates. Even though uncertainty visualization and our proposed guardrails tackle different problems, both approaches share the goal of visually adding context to a chart. We further explore the connections between these approaches in Section 4.2.4, after introducing our design space.

Technical interventions that have been proposed to tackle the exploratory problems primarily depend on the analyst's good faith, including strategies that automatically score their biasedness [66] similar to algorithmic approaches to detect cherry-picking in big data from the database literature [3, 37, 38], or including visualizations of the analyst's process [67]. In the context of the spread of online misinformation, we cannot typically rely on users to track their own biasedness and reliability—approaches that would fall under the strategy of transferring the responsibility described in Section 2.2.3.

Interventions that target biases and fallacies in narrative visualizations or at the audience level include using textual warnings against assuming that correlation equals causation [36], attaching multiple views to combat visualization mirages [78], adding interactive linking between text and data [77], as well as design alternatives for highlighting the truncation of the vertical axis [14]. Although the visualization community has raised concerns about the role of cherry-picked charts in the spread of misinformation across numerous studies [24, 25, 40, 41], to the best of our knowledge, this is the first work specifically attempting to design interventions against cherry-picking.

4 Designing Guardrails

In this section we describe our approach to characterizing the design space of visualization guardrails and outline the specifics and variations of the resulting design space.

4.1 Design Process

To develop a broad set of ideas, we engaged in a parallel prototyping process, where each of the authors independently developed multiple designs [18]. Before the start of the design process, the authors put together the design brief. We agreed that the main goal of the designs would be to provide missing context and promote skepticism in the viewer [36]. Additionally, the designs should be:

- **Nonobtrusive**: there should be no restrictions on exploration and selections, and every commonly allowed selection must be as visually salient with a guardrail as without. This goal stems from the fact that we target mitigation, rather than elimination of a feature (Step 3 of threat modeling).
- **Undemanding**: the guardrail should not directly slow down the user with pop-ups, questionnaires, or assessments. Making an author complete evaluations would more closely resemble transfer of responsibility (Step 3 of threat modeling).
- **Tamper-evident**: it should be difficult to get rid of the guardrail, for instance by cropping a screenshot. This goal is motivated by previous work [40] that showed that most of misleading charts shared on Twitter were screenshots using various levels of cropping.

In order to come up with initial guardrail designs, each of the authors independently created sketches based on the requirements. The sketches were either free-form or on top of examples of cherry-picked views, as identified by previous work on misleading visualizations [40]. We provide all of our sketches in supplemental materials. The first author then reviewed and organized the sketches into common themes and ideas, and all authors discussed the results and used them to describe the possible design space. Figure 1 presents an overview of the resulting design space along two dimensions: **context**, or *what* is shown, and **layout**, *where* it is placed.

4.2 Design Space

Given our design requirements and the problem at hand, we identified that the task of designing a guardrail against cherry-picking in a data exploration platform is related to presenting a helpful *visual comparison.* Gleicher et al. described three ways objects could be visually compared: by superimposition, by juxtaposition, or by explicit encoding of differences [27, 28]. Explicitly encoding a difference would, however, involve using a different representation of the selected items. For example, when comparing COVID cases in two countries, explicitly encoding differences could entail creating a derived dimension that subtracts the cases of the countries, and visualizing this derived dimension instead of the original data. As a result, this approach limits the saliency of selections (the original data is no longer shown) and violates our goal of nonobtrusiveness. However, the other two visual comparison strategies—*superimposition* and *juxtaposition*—fit our design requirements and describe where the guardrail could be placed.

The visual comparison strategies describe the **layout** of the guardrails, or where the guardrail is shown. The other dimension of the resulting design space is **context**, or what is being shown. We categorize our design sketches into two types of context: the guardrail can either show *primary data* in the same units, level of aggregation, and visual language as the main data, but potentially sampled to a small set of items or *visual summaries*—transformations and aggregations of the data, or additional data that provides a summary context (e.g., a market index for stock data).

4.2.1 Layout. We describe two main types of guardrail placement: superimposition and juxtaposition, as illustrated in Figure 1. **Superimposed** guardrails exist on the same canvas and scales as the main chart data. A designer of a Superimposed guardrail has two main decisions to make. The first one is defining what contextual data should be shown as the guardrail for the given scenario and domain. For instance, to tackle an instance of item cherry-picking, a COVID-19 data explorer showing Sweden's cases would probably plot other Nordic countries, such as Denmark, Norway, and Finland as well. The second design decision to consider is the treatment of the vertical scale: specifically, whether the axis should be adjusted or not in cases when the contextual data would go above or below the original frame. Examples of this design variation can be seen in Figure 3. The axis could be zoomed out to include all contextual detail, thereby sacrificing saliency or detail of the main selection. Alternatively, the scale could be kept as is, and instead would include a visual indication that there is out-of-frame context.

One of the main advantages of the Superimposed guardrails is that, by virtue of being plotted together with the main data, they are not just tamper-evident but virtually tamper-proof: it would be very difficult to remove the guardrails from the view. Additionally, Superimposed guardrails exist on the same scale and units as the main data and offer an opportunity to directly compare the values of the main chart data and the contextual data.

The disadvantages of Superimposed guardrails include the danger of overplotting—a lot of contextual information in the main frame of the visualization may not scale with many items selected. This problem could be alleviated by dynamically adjusting the size of the comparison set as an author chooses more items. Another problem is that it might not always be obvious which data items or time frames are useful "important context." A system could leverage metadata (such as regions of the world for country data, or sectors for financial data) to make such a determination. A generic

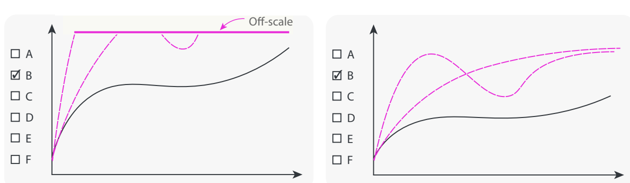


Figure 3: Two design variations of vertical scale treatment in Superimposed guardrails. The example on the left preserves the scale of the main chart data, emphasizing the original selection but truncating the off-scale context. The design on the right adjusts the scale to fit the guardrails.

implementation that does not require additional data could leverage statistical information, i.e., by including a set of representative examples of the data set.

The idea of the **Juxtaposed** layout of guardrails is to leave the main canvas of the visualization unchanged. Instead, we provide contextual information in a separate, juxtaposed view to the side, above, or below the chart. When designing Juxtaposed guardrails, the decision of their placement depends on the underlying data and target issue. For instance, if the goal of the guardrail is to give an indication that a climate change-denying author chose a small fraction of the time frame of ocean temperature data, the guardrail could run along the "problematic" dimension—in this case, the horizontal axis right below the chart (as seen in the example on the left in Figure 4).

In Juxtaposed guardrails, the decision about what constitutes important context data is less central than in Superimposed guardrails. Specifically, juxtaposition allows the designer to show *all* of the data points that would fit into the frame. Similarly, a Juxtaposed guardrail is easy to combine with other guardrails, since, unlike with Superimposed guardrails, there is no issue of overplotting the main chart.

However, Juxtaposed approaches are croppable and thus not tamper-evident (a design goal). Implementations could use strategies to reduce the croppability, for example, by embedding a Summary directly into the axis, such that cropping out the guardrail would also involve cropping out part of the axis. Alternatively, the designer could wrap the chart and the guardrails with a frame that could be indicative of parts left out. In general, however, even croppable designs could provide a degree of protection, as it would be much easier for an online audience to highlight cherry-picking if they could point to the full chart in replies or community notes.

4.2.2 Context. We distinguish between two types of guardrails' contextual data: they could show Primary Data of the same type as the main chart data, or they could use aggregated or transformed data in the form of a Summary.

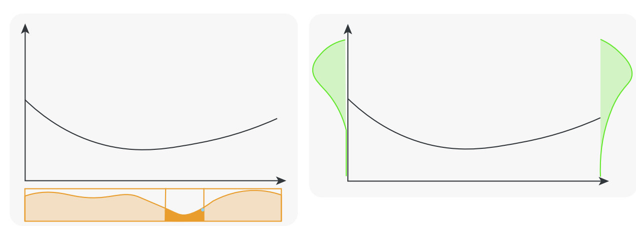


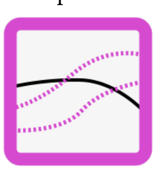
Figure 4: Examples of Juxtaposed guardrail variations for time frame cherry-picking. On the left is an example of Juxtaposed Primary Data : the miniature view below shows the entire time frame and highlights the cherry-picked period of a dip in value. On the right is a periphery plot [47]—an instance of Juxtaposed Summary . The peripheries show that the value of interest is much higher in the periods before and after the selection.

Primary Data guardrails provide contextual data by directly plotting the primary data—meaning data at the same level of granularity and of the same units as the main data in the chart. Primary Data guardrails are shown in the top row of Figure 1. The main advantage of Primary Data guardrails is their simplicity (both to implement and to understand): while they should be visually distinct from the main data, they still use the units as the chosen items and need virtually no explanation to be understood. This advantage is especially strong in a Superimposed Primary Data guardrail: it utilizes the same visual encoding and exists on the same axes as the main data, and thus its meaning is self-explanatory.

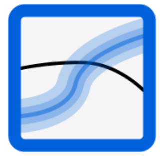
Summary guardrails, on the other hand, condense the primary data into a summary form for the purposes of providing context. They are shown in the bottom row in Figure 1. There are many possibilities for the exact implementation of Summary guardrails, and, consequently, the designer should deliberate over what type of summary is the most meaningful, given data and domain. For instance, a simpler Summary could provide the average, interquartile range, or extrema of the data set. A more complicated Summary could show information about the variance or the shape of the distribution of the data.

Compared to Primary Data guardrails, the Summary guardrails are more compact and help mitigate the problem of overplotting by consolidating all of the contextual data into a single visual representation. On the downside, however, Summary guardrails may be more complex to understand as they represent a departure from the visual encoding of the main chart and may involve an uncommon visualization type. Therefore, it would be advisable to train the viewer to read the chart using annotations or training modules.

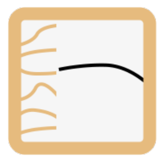
4.2.3 Visual Techniques and Implementations. As alluded to in the previous sections, there are a number of design alternatives and decisions that one should consider when designing what the guardrails look like and what data they are composed of. Specifically, while our design space describes the general affordances of different guardrail types, each type of guardrail type could be realized using different visual encodings. In this section, we present an overview of specific implementation variations and related visualization techniques from prior work.



The **Superimposed Primary Data** guardrail is arguably the simplest: as it involves automatically plotting a subset of contextual data, there are few design choices to be made. The key decision in this guardrail design involves defining a contextual data set that is effective yet not overly large, akin to determining the optimal set of forecasts to display in a Multiple Forecast Visualization. [49]. We observe two simple implementations of this technique in data explorers online. First is the scatter plot on Gapminder [58] that automatically includes translucent data points for the entire available data set; however, the opacity of the context is adjustable and could be completely removed. The second example is Google Search: when looking up a macroeconomic or demographic metric for a given country, the resulting chart typically includes two other regions for reference. For instance, at the time of writing this paper, looking up "population of the US" returns a chart with the populations of the US over the past century, as well as those of Russia and Mexico in fainter lines.



Designers of **Superimposed Summary** guardrails have more liberty in implementation, as there are many types of visual summaries to choose from. When tackling time frame cherry-picking, the Superimposed Summary could take form of lagged variance information. Similar designs have long been popular in financial trading with the purpose of encoding past volatility of a financial instrument alongside its value, and include Bollinger Bands [8] and Standard Deviation Channels, or Envelope Channels. In the context of item cherry-picking, as discussed previously, Superimposed Summaries could take the form of a simple average, interquartile range, or extrema information of the contextual data using a representation similar to that of contour boxplots proposed by Whitaker et al. [71]. Aside from statistical summaries, a Superimposed Summary could aggregate the contextual data into a background heatmap using techniques such as DenseLines [46]. Superimposed Summaries could also simply draw on common domain-specific benchmarks, such as stock market indices in the financial sector.



The main goal of **Juxtaposed Primary Data** guardrails is to show omitted items or time periods of the main data next to the main chart. An example of previously proposed visualization technique that could serve as a guardrail for time cherry-picking is stack zooming [32]. In the context of data explorers, Juxtaposed Primary Data could benefit from being miniature as to not be too large to gist or end up distracting from the main view. To achieve this, the specific implementations could be chosen from the set of charts designed to take up little space, such as sparklines [63] or horizon charts [30] in instances where the sign of the data provides meaningful context. Additionally, although this type of guardrail does not require a precise definition of "context data set" as Superimposed Primary Data, space constraints may necessitate decisions about what data is seen immediately beside the visualization. Juxtaposed Primary Data guardrails often can play a dual role and take the form of scented widgets [72]. For instance, a sparkline can serve as the preview of an item's data next to each item's selection checkbox—a technique commonly used in financial data explorers. In guardrails against time cherry-picking, the horizontal axis selection slider could instead be a miniature chart of the entire period which the user could brush (as seen in the example on the left in Figure 4). An example of an existing implementation of this is the exchange rate explorer on CoinMarketCap.com [7]: a user could zoom in onto a short period of, for instance, Bitcoin price chart by brushing over a miniature view below the main chart.

The guardrail type with the largest set of implementation alternatives is **Juxtaposed Summary**, as there are virtually no restrictions on what could be shown, as long as it provides useful context. One subset of design alternatives consists of variations of focus-and-context visualizations, or designs that allow the user to see both detail and overview simultaneously [11]. This could include distorted views, such as Multistream from Cuenca et al. [15]. Morrow et al.'s Periphery Plots [47] provide examples of nondistorted variations of this guardrail that could be helpful in the context of time cherry-picking (seen in the example on the right of Figure 4). Designs similar to periphery plots could similarly be used to tackle item cherry-picking—in this case, however, the Summary in the periphery would be calculated based on the omitted items rather than omitted time frame. The space of ways to show the distribution of contextual data is large and the choice of a specific visual encoding should depend on the features of the underlying distribution, as well as the expected audience's visual literacy, as these types of charts are typically less common in visualizations for general public. Correll provides a helpful analysis of advantages and disadvantages of distribution visualizations as well as their combinations as raincloud plots [13]. The example in Figure 1 as well as our prototype employ a vertical strip plot, but other designs we have considered include a box plot, a violin chart, or an inverted histogram (shown in the example on right in Figure 4).

4.2.4 Relatedness to Uncertainty Design Spaces. Many of the visual techniques that can be used as guardrails are also commonly used to visualize uncertainty, albeit with different goals and interpretations. The task of exposing cherry-picking is related to visualizing uncertainty: both intend to give a clue about, or cast doubt in, the completeness of the data displayed [26]. However, although uncertainty designs typically address questions like "How *reliable* is this estimate?" or "What is the spread of possible outcomes around *this point*?", guardrails shift the focus to "How *representative* is this value?" and "How does this data point compare to the *rest of the available data*?"

As a result of this difference in goals, there is only a partial overlap in visual techniques for guardrails and for uncertainty visualizations. The shared visual techniques consist of *distributional visualizations* [50] techniques: a box plot could be used to either visualize the uncertainty of an estimate or summarize the omitted data as a guardrail. Similarly, an ensemble visualization [69] of uncertainty could communicate the various forecasts of snowfall for the season, whereas a visually similar Superimposed Primary Data guardrail could show the historical snowfall data for previous seasons as context. However, many guardrail techniques do not have equivalent uncertainty visualizations, such as the Juxtaposed Primary Data visualization of the selected time period in Figure 4 that simply plots the entire available time series alongside the selection, rather than samples from a distribution. The reverse is also true: direct encodings of uncertainty in the visual channel are not applicable to the guardrail design space—such as modifying the mark of interest by blurring it proportionally to the uncertainty [42]—since there exists no quantifiable dimension of uncertainty. As discussed previously, guardrails do not communicate the uncertainty properties of the value of interest itself and have the goal of not modifying the existing mark.

Building on this distinction, our design space uniquely focuses on describing the opportunities for **augmenting existing visualizations**—a goal not addressed by previous related design spaces. Prior typologies of uncertainty visualization [50, 61, 62] primarily offer a categorization of the source of uncertainty (measurement error vs. credibility) or specific visual properties of uncertainty visualization techniques (whether they are animated, 1- or 2-dimensional, have dichotomous boundaries or are "fuzzy"). In contrast, our proposed design space categorizes visualization techniques in a way that describes opportunities and trade-offs of incorporating them into an existing canvas in a nonobtrusive way.

5 Prototype Implementation

With the goal of evaluating the overall merits of guardrails, we implemented a prototype data explorer and therefore had to make decisions about what it would look like and what data it would use. As a starting point, we chose one simple visual technique as the implementation for each major guardrail type. In the designing of the prototype, we prioritized *ecological validity*: therefore, we used real-world data, and aimed to mimic the functionality and look of existing data explorer platforms, such as OurWorldInData's COVID-19 explorer [43]. Screenshots of our prototype implementation can be seen in Figures 5, 10, 12, as well as in supplemental materials. A sandbox version of our prototype is available online.

We chose to design for (and evaluate with) two different data sets that are commonly used in public-facing data exploration tools: COVID-19 death count data by country retrieved from OurWorldInData [43] and individual stock performance data retrieved from Yahoo! Finance [74]. We anonymized the country and stock names to limit the influence of preconceived ideas about COVID-19 or particular stocks. We limited the number of items available to select to 15 or less to simplify the tasks in evaluation. We also chose to limit our guardrails prototypes to item-cherry-picking, excluding time-cherry-picking from our design and our study.

To create these prototypes, we needed to select a specific design instance of each guardrail. As discussed above in Section 4.2.3, the exact visual language and criteria for selecting contextual data are decisions the designer has to make depending on the domain and problem. In our case, in the **Superimposed Primary Data** condition we defined contextual items to be countries from the same region in the Viral scenario; and stocks from the same industry in the Stocks scenario. For **Superimposed Summary**, we showed the average of all items as a line and the interquartile range as a shaded area (Figure 10). For a more realistic presentation, we labeled the average line as "Market Index" in the Stocks scenario.

In the **Juxtaposed Primary Data** condition, we chose to provide a sparkline of each item next to its label (Figure 5). We filled in the area under the line chart in light color, which made it easier to compare magnitudes among the small multiples that are stacked vertically relative to just lines. Lastly, for the **Juxtaposed Summary** guardrail we implemented a stripplot showing data of all items across all time points, with each tick representing the value of one item on a given day. The ticks are shown with slight transparency to account for over-plotting, and we use color-coding to distinguish positive from negative values—a distinction relevant for the Stocks scenarios. A shaded funnel denotes which part of the global scale is currently shown on the main canvas, and also makes it more difficult to crop the plot without leaving evidence of tampering. For both of the Summary guardrails, we added a sentence explaining what the shaded area or the stripplot denote.

6 Study 1: Production

In order to evaluate whether the guardrails make it *more difficult to cherry-pick data*, we conducted a crowd-sourced experiment

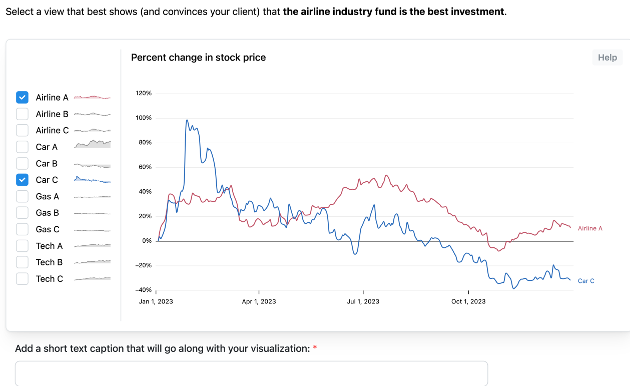


Figure 5: Screenshot of our prototype implementation of data explorer in the Study 1 experimental setup. Shown is the Stocks B scenario with a Juxtaposed Primary Data guardrail.

challenging our participants to use our prototype data exploration platform with and without guardrails in place. One goal of this study was to evaluate whether participants find it more difficult to create cherry-picked visualizations with guardrails. The other goal of this study was to produce a data set of cherry-picked visualizations to be used in a subsequent reaction study (see Section 7). This section describes our methodology and results. Figure 6 presents a visual overview of the study procedure in the form of a data comic [70].

6.1 Methods

The experiment investigated how participants used our data explorer to create cherry-picked visualizations, both with and without guardrails.

During the study, participants were tasked with using our data explorer to create a visualization supporting a given proposition by selecting data in a line chart and writing a caption, mimicking the act of sharing a cherry-picked data explorer-based visualization on social media. Each participant was randomly assigned one guardrail condition, and they would perform half of their tasks with that guardrail and half without any guardrail, as controls. To prioritize ecological validity, the study employed anonymized real-world data and we realistic scenarios of malicious cherry-picking observed in recent events: such as downplaying the seriousness of a deadly viral disease or exaggerating the profitability of a specific security. Therefore, we constructed four data scenarios presented to the participants: two based on anonymized COVID-19 fatalities, reframed as a fictional viral disease (Viral A & B), and two using anonymized individual stock performance data (Stock A & B). In the Viral scenarios, participants acted as public health officials tasked with promoting not-so-effective policies. In the Stock scenarios, participants imagined themselves as financial advisors promoting underperforming funds at the direction of a supervisor. The prompt and user interface is shown in Figure 5.

The study was reviewed by our institution's IRB and deemed exempt from full board review. Before the main study, we conducted two pilots. After a first in-person pilot with two students, we made adjustments to clarify scenario descriptions, add a help button, and fix a visual bug. A second pilot with 10 Prolific participants confirmed the tasks were understandable and feasible for remote users, requiring no further changes. We recruited 130 participants from Prolific for the main study. The survey was conducted using the reVISit framework [17]. We logged the interaction data using the Trrack library [16] and used the NASA TLX [29] questionnaire to assess participants' subjective workload after each task. At the experiment's conclusion, we debriefed the participants about the guardrails and asked whether they noticed them and felt they influenced the tasks. The median completion time was 15.5 minutes, and participants were compensated $5 (average hourly rate of $19). The study stimulus is available online.

The study produced 520 visualizations and captions. After the experiment, two authors independently reviewed the submitted visualizations and captions with the goal of removing unsuitable submissions for the follow-up study. They agreed on inclusion and exclusion for 395 evaluations, discussed and resolved disagreements on 125, and ultimately excluded 18% of submissions. Exclusions were due to irrelevant selections, unintelligible captions, or overly generic responses (e.g., "Chart"). The final data set consisted of 427 submissions.

Quantitative analysis involved paired t-tests to compare the number of clicks and NASA TLX scores between tasks with and without guardrails. Qualitative analysis examined captions and poststudy feedback to identify recurring themes and insights.

6.2 Findings

As a result of Study 1, our participants generated 520 cherry-picked visualizations and captions, of which 427 passed our quality review and were used in further analysis. The submissions spanned two scenarios and four guardrails, as well as the control condition. All resulting submissions—visualizations and captions—are available for review online.

6.2.1 Quantitative Results. Figures 7 and 8 summarize the within-subject differences in the number of clicks per task and the NASA TLX survey results and the subjective influence of guardrails on task difficulty. **We find that Superimposed conditions made cherry-picking more effortful, but Juxtaposed guardrails made navigating the data easier.**

Participants in the Juxtaposed Primary Data condition required significantly fewer clicks (21 versus 32 in Control; $T(38) = -4.00, p < 0.0001$) and less time to explore the data (106 seconds versus 143 seconds in Control; $T(38) = -2.12, p = 0.041$). They also reported better subjective performance (28.9 versus 39.7; $T(38) = -2.45, p = 0.019$) and lower mental demand (48.8 versus 55.8; $T(38) = -1.84, p = 0.075$). Similar "benefits" were observed in the Juxtaposed Summary condition for performance (26.5 versus 31.7; $T(52) = -1.95, p = 0.055$). Additionally, Figure 8 shows that fewer than 5% of participants in the Juxtaposed Primary Data and Juxtaposed Summary conditions reported that the guardrails made their task "much more difficult." **This indicates that our juxtaposed guardrails actually make cherry-picking easier to achieve,** which is the opposite of our goal.

Those in the Superimposed Summary condition, however, reported higher mental demand (51.3 versus 43.5; T(47)=2.10, p=0.041).

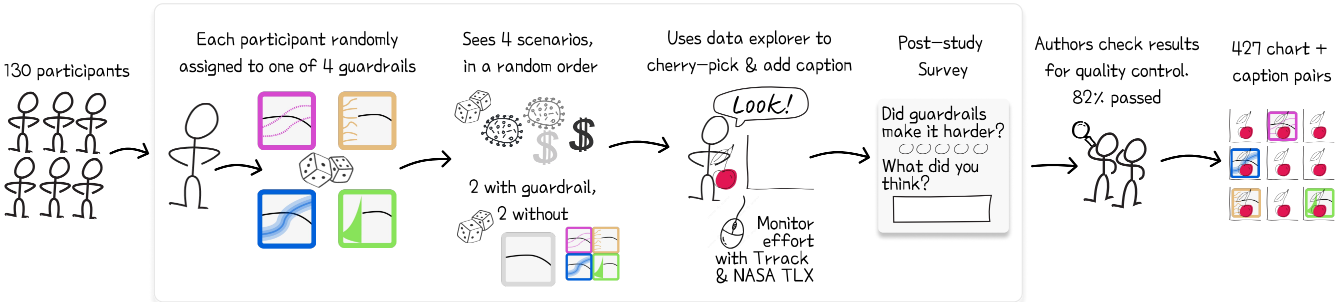


Figure 6: Data comic showing the study design in Study 1: Production. The study design employed a between-subjects approach for the guardrail condition and within-subjects approach for the task scenarios.

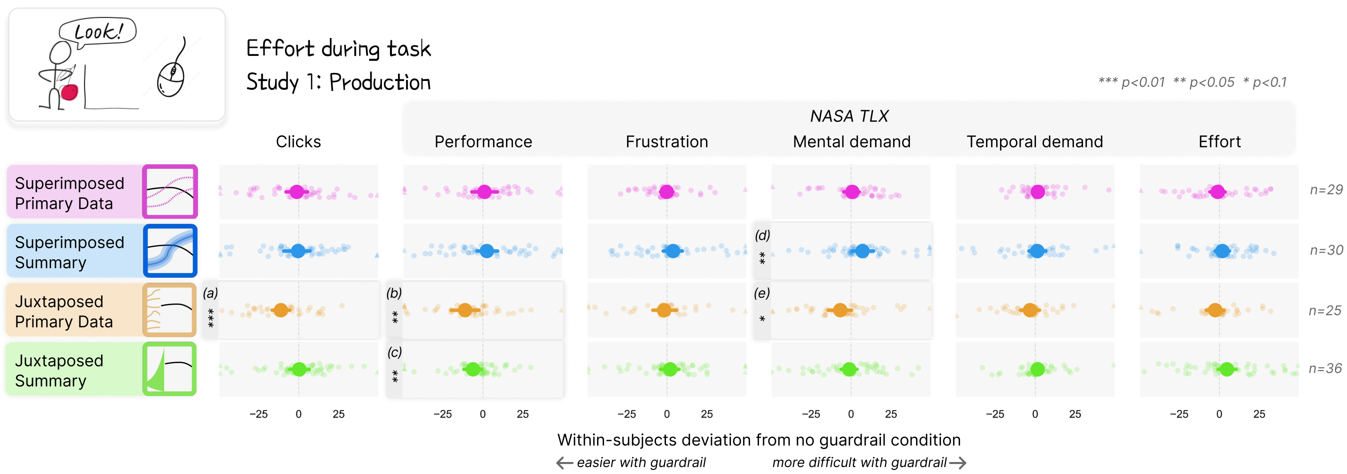


Figure 7: Juxtaposed guardrails make it easier to complete the cherry-picking task: participants require fewer clicks to complete the task (a), report better performance (b, c), and slightly lower mental demand (e). Superimposed Summary leads to a higher reported mental demand (d). Shown are the mean difference between the metrics for guardrail and control tasks for each participant, as well as 95% bootstrapped confidence intervals. For clarity, large outliers (<5% of data) are shown as triangle markers on the edges. For the NASA TLX Performance metric, lower means better [29].

Figure 8 further illustrates that participants in Superimposed conditions more frequently rated the guardrails as making tasks "more difficult" or "much more difficult", indicating that **superimposed conditions support our goal of making cherry-picking more difficult**.

6.2.2 Qualitative Results. Consistent with our quantitative results, participants noted that the **Juxtaposed guardrails made cherry-picking easier**: *"It made it easier to visualize without constantly turning off and on each data set to conceptualize which ones would work best."* This aligns with the design of Juxtaposed Primary Data, which provides a clear overview beneficial for both authors and audiences of cherry-picked visualizations.

In contrast, **participants found Superimposed guardrails to be more obstructive**. Those in the Superimposed Summary condition expressed frustration in text responses and one participant remarked that the guardrails *"interfered with my ability to cherry-pick the data I needed to."* Another participant noted, *"I couldn't easily show that one investment was better than others without hiding some truth."* Two participants even refused to perform the task, with one caption simply stating: *"I can't. I'd be a liar. [Option A] is terrible."*

Despite these challenges, qualitative analysis of captions revealed participants' adaptability. When guardrails exposed context that debunked cherry-picking, **participants often shifted focus**. For instance, one caption downplayed the magnitude of infections under Policy A and emphasized trends: *"Policy A's peak comes fast but comes back down just as fast" (link)*. Similarly, when promoting Airline stocks, a participant reframed average returns as a sign of stability: *"Over time, growth in stock prices in the airline industry has either been consistent with or outperformed the market average, thus being the most predictable" (link)*. Others incorporated world knowledge to explain away an underperforming stock: *"Covid kind of ruined airlines but now that it's getting less and less prevalent we're going to see a resurgence" (link)*.

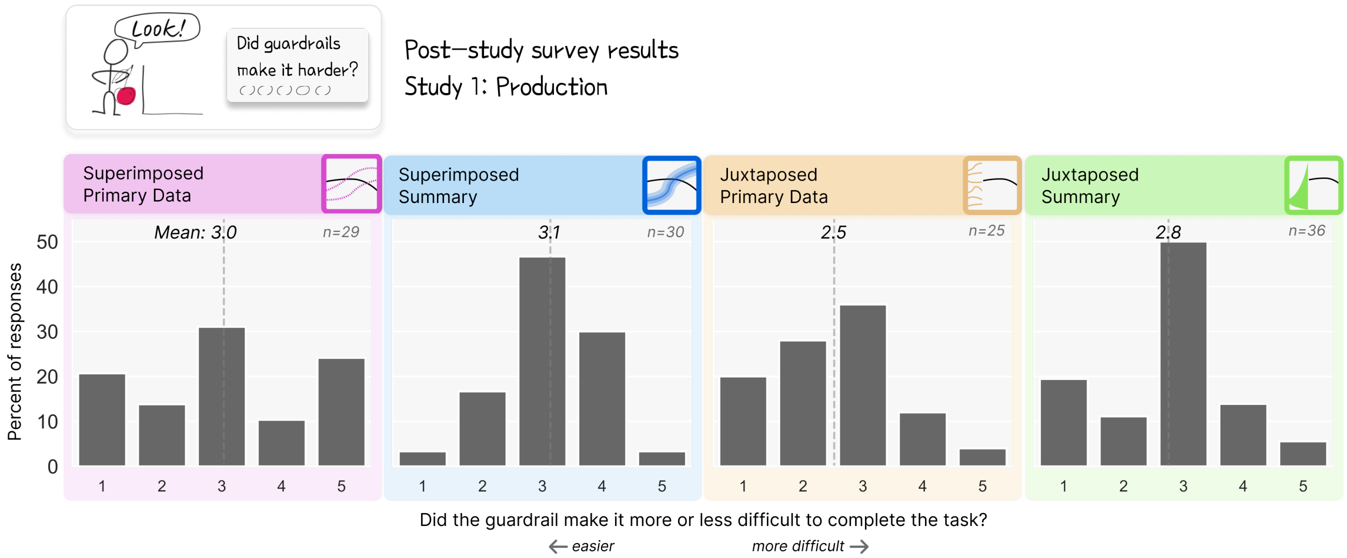


Figure 8: Participants in the Superimposed Primary Data condition described the guardrail to make cherry-picking "much more difficult" more often than those in any other condition; however, the responses are very split. On average, participants reported Juxtaposed guardrails to make it slightly easier to complete the task.

7 Study 2: Reaction

In order to evaluate whether the guardrails make *cherry-picked data less convincing*, we conducted a second crowd-sourced experiment asking another set of participants to review charts created by the participants of Study 1. This section describes our methodology and results. Figure 9 presents a visual overview of the study procedure in the form of a data comic [70].

7.1 Methods

The study was reviewed by our institution's IRB and deemed exempt from full board review. Similar to Study 1, this study was implemented using the reVISit framework [17], and full instructions are available in the supplemental materials. Before the main study, we conducted three pilot studies on Prolific with five participants each to test the clarity of the scenario and questionnaire. Based on their feedback, we made minor adjustments to the UI and task language.

For the main study, we recruited 160 English-speaking participants from Prolific. Participants were randomly assigned to one of four scenarios from Study 1 and shown five visualizations in random order—one from each condition (4 guardrails and 1 control), selected from 427 charts created in Study 1. The interface in Study 2 was the same as Study 1, with controls disabled and the addition of captions. Figure 10 shows the experimental setup.

To measure guardrails' effect on how convincing the cherry-picked claims are, we presented participants with a hypothetical decision-making scenario inspired by a trust game [6], an approach from behavioral economics. Trust game-based hypothetical decision-making and investment scenarios have been proposed to measure trust in visualizations [20], and more recently have been adapted in visualization work [21, 23, 53, 73, 75]. Participants in the Stock scenario were asked to imagine they were selecting an investment portfolio in a way that maximizes their profits. They were shown a visualization from Study 1 that promoted a certain investment, and were asked to decide how much they would invest ($0 to $100) in the recommended industry based on the available information. In the Viral scenario, participants were asked to imagine they were traveling to a virus-affected area and had to buy insurance to minimize their financial risk. They were shown a visualization from Study 1 that downplayed the risk of virus and were asked to choose how much health insurance they would buy (also $0 to $100) based on the available information.

Participants were also asked to provided a brief rationale for their choices and completed Likert-scale questions assessing trustworthiness, persuasiveness, clarity, and likelihood of sharing the visualization. In the debrief, they were also asked whether they noticed the guardrails, and whether they understood their meaning. The median completion time was 10 minutes, and participants were compensated $2.50 (median hourly rate of $15/hr). The study stimulus is available online.

We analyzed the data quantitatively using repeated measures ANOVA and post hoc t-tests with the Benjamini-Hochberg [5] procedure for multiple comparisons. The scripts and the results of all statistical tests are available in the supplemental materials. We also qualitatively reviewed the justification and feedback text to identify key themes.

7.2 Findings

Figure 11 presents the monetary action results by guardrail and scenario. Guardrails had observable effects on skepticism in three of the four conditions. **Primary Data guardrails performed slightly**

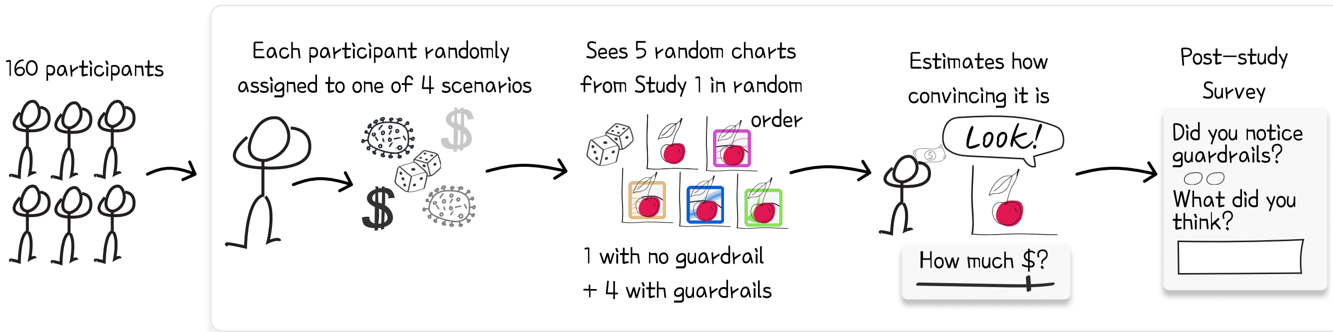


Figure 9: Data comic showing the study design in Study 2: Reaction. The study design employed a between-subjects approach for the task scenarios and within-subjects approach for the guardrail condition.

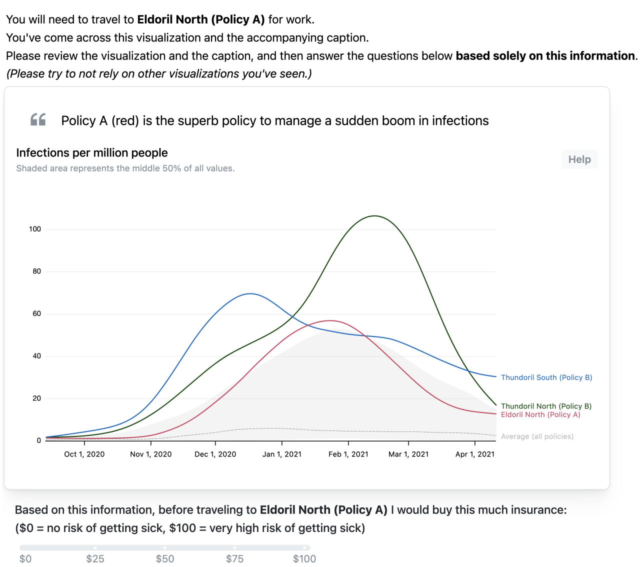


Figure 10: Screenshot of our prototype implementation of data explorer in the Study 2 experimental setup. Shown is the Viral A scenario with Superimposed Summary .

more effectively than Summary guardrails, with some participants finding the visual Summaries more difficult to interpret, leading them to ignore these guardrails. The Primary Data guardrails appeared to be more intuitive, as their visual encoding resembled the main data selections, and few participants expressed confusion.

7.2.1 Quantitative Results. Repeated measures ANOVA tests showed **significant differences in monetary decision between guardrails and control conditions in the Viral A scenario** ($F(4, 152) = 7.79, p < 0.001$). post hoc t-tests revealed that participants in the Superimposed Primary Data condition spent, on average, \$63 on insurance, compared to \$45 in the Control condition ($T(38) = 4.28, p = 0.001$). Superimposed Summary (mean \$55, $T(38) = 1.99, p = 0.076$) and Juxtaposed Primary Data (mean \$56, $T(38) = 2.72, p = 0.024$) also led to higher insurance purchases compared to Control. Despite this, the poststudy survey revealed that **most participants did not fully consider the guardrails**: 38% of participants did not notice them, and 23% noticed them but did not understand their meaning. As a result, we do not observe statistically significant results in the other scenarios.

7.2.2 Qualitative Results. **Participants who did notice the guardrails directly referenced the guardrails in their rationales**. For example, a participant in the Superimposed Primary Data condition stated: *"[The chart] shows [option A] to be the worst one out of the lines shown."* Another noted in the Superimposed Summary condition: *"[the infection rate] is still considerably higher than the average..."* These findings suggest that the guardrails may blend into the visualization, depend on data literacy, or be overlooked due to added visual complexity. Even among those who understood the guardrails, some chose to focus on the main data instead, with one participant stating: *"I just followed the figures and how they were either rising or falling"*.

8 Study 3: Controlled Reaction

We conducted a third crowd-sourced experiment to further explore how guardrail effectiveness varies with the severity of cherry-picking, while also accounting for the influence of redeeming factors. In contrast to Study 2, in this experiment we prioritize internal validity by removing captions and controlling the data displayed in each scenario, which allowed us to isolate the impact of guardrails from other factors. Below, we describe our methodology and findings. Figure 12 presents a visual overview of the study procedure and screenshots of the conditions in the form of a data comic [70].

8.1 Methods

For Study 3, we used a simplified set of five stocks from the same Yahoo! Finance [74] data set as in Study 2, selecting stocks with nonoverlapping time series to highlight the highest return in any data subset. Unlike in previous studies, we removed the caption to isolate the effect of the guardrails. To measure guardrail effects, we asked a 5-point Likert question, "The visualization supports the idea that stock X yielded the highest returns in 2023", followed by an open-text rationale and an attention check.

Study 3 included three conditions: *Correct* (control), where all data was shown and participants were prompted about the actual top-performing stock; *Incorrect*, where the best-performing stock

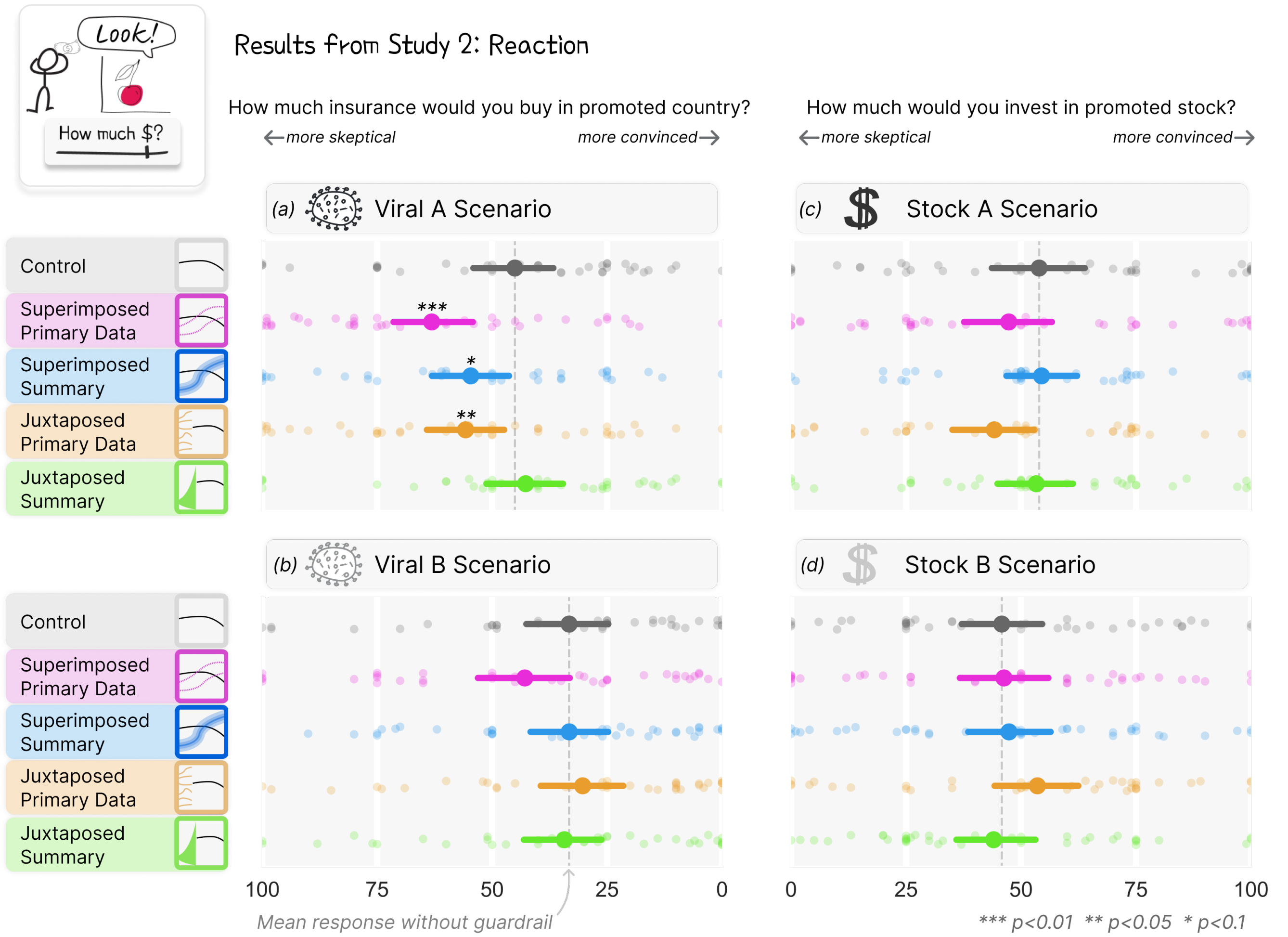


Figure 11: Most guardrails had a significant effect on making the viewers skeptical of the cherry-picked charts in the Viral A scenario (a). However, the effect was smaller in other scenarios (b–d). The effects vary highly by guardrail and scenario. Shown are average values and bootstrapped 95% CI (n = 1000) and individual data points in the background. Note the inverse scale between the Viral and Stock scenarios: the monetary action in the Viral scenarios involves making a decision about insurance purchase (less insurance = trusting cherry-picking), whereas in the Stock scenarios it involves making an investment (more investment = trusting cherry-picking).

was hidden and participants were prompted about the second-best stock; and *More Incorrect*, where the top two stocks were hidden and participants were prompted about the the third-best stock, which performed below the average of all stocks. In all cases, the prompted stock was the highest performing "salient" stock, but the guardrails could be used to discover that the chart was cherry-picked. We conducted a first pilot with 30 participants, after which we rephrased the survey questions for clarity. We then conducted a second pilot with 150 participants and performed a power analysis to calculate the required sample size for detecting meaningful differences across conditions with a target power of 0.8 and an alpha level of 0.05.

For the main study, we recruited 675 Prolific participants, all fluent in English, who were randomly assigned to one of the three correctness conditions and one of five guardrail conditions, i.e., each participant saw only a single chart. The median completion time was 3 minutes, and participants were compensated $0.75 (average hourly rate of $15). The study stimulus is available online. We again analyzed Likert responses with ANOVA and post hoc t-tests, applying the Benjamini-Hochberg procedure [5] to adjust for multiple comparisons, and reviewed participants' text responses.

8.2 Findings

8.2.1 Quantitative Results. Figure 13 summarizes Study 3 results for participants' agreement with the statement that the cherry-picked item is the highest overall. Our ANOVA results showed statistically significant differences due to both cherry-picking severity ($F(2, 672) = 31.39, p < 0.0001$) and guardrail condition ($F(4, 670) = 43.18, p < 0.0001$), confirming that **guardrails influence skepticism toward cherry-picking**, with more severe cherry-picking yielding stronger effects. As expected, we observe

Figure 12: Data Comic and conditions in Study 3: Controlled Reaction. Participants were asked to rate the following prompt on a 5-point Likert scale ('Strongly disagree' to 'Strongly agree') "The visualization supports the idea that stock C yielded the highest returns in 2023." Screenshots show the visualization seen by participants in the "More Incorrect" scenario, in which the top two true best stocks are hidden. The Incorrect scenario only hides the top stock, whereas the Correct scenario shows all stocks.

no effects in the *Correct* condition, indicating that guardrails do not have an effect on correct interpretations.

Post hoc t-tests with Benjamini-Hochberg FDR correction revealed that, compared to Control, the Superimposed Primary Data guardrail had a large, significant effect in both *Incorrect* and *More Incorrect* conditions ($T(88) = 10.88, p < 0.0001$ and $T(77) = 12.68, p < 0.0001$). The Superimposed Summary and Juxtaposed Primary Data guardrails also had significant effects but only in the *More Incorrect* scenario ($T(70) = 3.61, p = 0.002$ and $T(73) = 2.39, p = 0.058$, respectively). Similar to Study 2, the **Superimposed Primary Data guardrail was the most effective against cherry-picking, while Juxtaposed and Summary guardrails had smaller effects**. Juxtaposed Summary, the

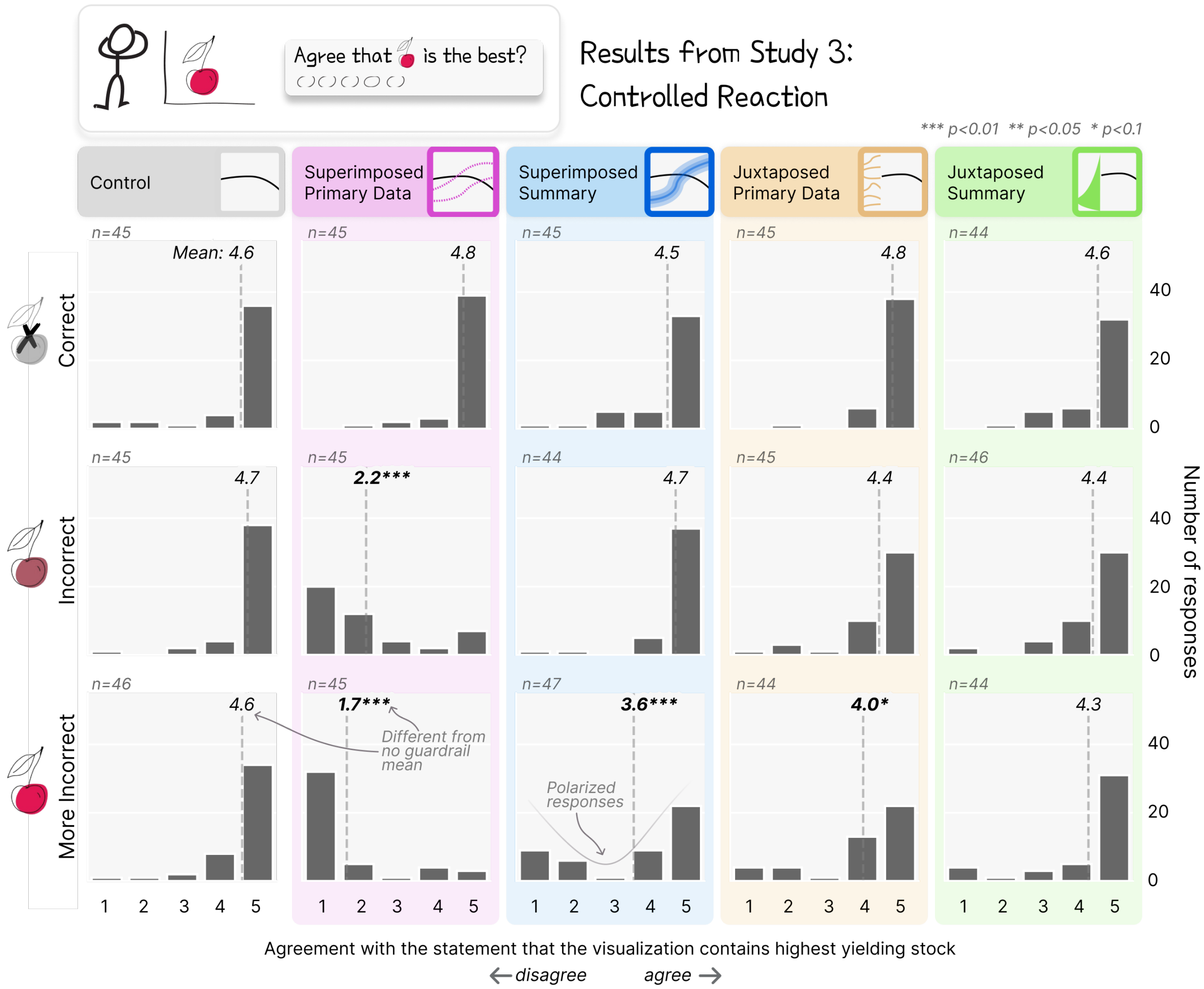


Figure 13: In Study 3, Superimposed Primary Data has a very strong effect of encouraging skepticism in both cherry-picked conditions compared to no guardrail. Superimposed Summary and Juxtaposed Primary Data have a significant effect in the More Incorrect condition. As expected, guardrails have no effect in the Correct, not cherry-picked condition. Note the highly polarized distributions of responses: participants either strongly agree or disagree.

most visually distinct guardrail, produced results closest to control. Notably, **the distribution of responses in all conditions is highly polarized**: almost all responses either strongly agreed or strongly disagreed with the statement. Several conditions exhibit a bimodal response distribution, suggesting that guardrails are likely effective when noticed and understood, but the challenge lies in capturing the viewer's attention and understanding.

8.2.2 Qualitative Results. We reviewed the open-text responses in the Superimposed Summary of *More Incorrect* condition, which was highly bimodal and allows us to better understand the source of an audience's disagreement. Participants' explanations reinforce the notion that guardrails are effective when they successfully capture both the viewer's attention and their understanding. We find that 13 of 15 responses that correctly disagreed (1 or 2 on the Likert scale) explicitly referenced the guardrail showing that the cherry-picked stock was below the industry average. One participant stated, *"Stock C was below the industry average at the end of the year, meaning it couldn't have possibly had the highest return in the industry."*

In contrast, only 9 of 31 responses that incorrectly agreed (4 or 5 on the Likert scale) referenced the average line, but none were able to integrate it into their reasoning. These participants merely acknowledged its presence and primarily focused on comparing the cherry-picked stocks, as illustrated by one explanation: *"I looked at the industry average. And then I also looked at all the other stocks, C looks like it's been doing better than any other one."* These results point to the fact that, in order to be effective, more complex guardrails should be both highly visually salient and directly cue the user about how to integrate them into higher level chart comprehension [55].

9 Discussion and Design Recommendations

Our findings indicate that guardrail designs have the potential to mitigate cherry-picking in data explorers, though they come with limitations and important considerations. In this section, we discuss these findings and offer design recommendations for effective guardrail implementation.

Recommendation 1: Prioritize simpler guardrails that maintain the original visual language. Our design procedure and crowdsourced studies uncover a tension between designing a guardrail that effectively summarizes contextual information and maintaining its alignment with the original chart's visual language, as summarized in Figure 14. The closer the guardrail is to the original chart visually, the easier it is for the audience to notice and understand it. At the same time, guardrails that use the same visual encoding for individual items can lead to overplotting. Our studies show that *Primary Data* guardrails are the most effective, primarily due to the fact that they are easier for audiences to notice and understand in both real-world and controlled scenarios. Additionally, they do not require viewers to independently extract high-level patterns, such as integrating the distributional statistics into their interpretation, a process highly affected by individual differences [55]. Therefore, we recommend prioritizing guardrails that closely match the visual encoding of the main data and using *Primary Data* guardrails as the default. However, in some cases the domain in question may require larger amounts of information to be used in the guardrail or could benefit from specific statistical summaries. When using a *Primary Data* guardrail is not an option, we suggest that *Summary* guardrails should always be paired with tutorials or detailed annotations to help users interpret and apply them correctly.

Recommendation 2: Carefully identify potential targets and contextual information. Across all types, constructing a guardrail demands a clear and careful definition of "contextual information" that would debunk cherry-picking, whether it is demographically comparable subsets, or countries with similar climates. Defining this context correctly may require a careful examination of the domain and consulting with a domain expert. Additionally, an evaluation of the domain and existing misinformation can help determine which types of cherry-picking should be addressed in the first place. For example, climate change misinformation often involves cherry-picking specific time frames [12], while COVID-19 conspiracy theories typically cherry-pick data from items (e.g., countries) [40]. In this process, we recommend following the threat modeling steps outlined in Section 2: reflect on the system goals, identify vulnerabilities, propose and evaluate interventions.

Recommendation 3: Implement guardrails to deter cherry-picking and enhance platform usability. Our results from Study 1 show that *Superimposed* and *Juxtaposed* guardrails have different impacts on authors' experience using the data explorer. Similar to previous work evaluating visualization composition, we found that participants found it easier to integrate superimposed information [33, 48, 55]. *Superimposed* guardrails thus made it harder to create cherry-picked views and led to less convincing presentations. This could push misinformation actors toward less reputable platforms, while helping unintentional cherry-pickers recognize their biases. In contrast, participants cited that *Juxtaposed* guardrails made it easier to explore the data. Although this may aid careful cherry-picking, we found no evidence of exploration ease leading to more misleading visualizations. Since *Superimposed* and *Juxtaposed* guardrails can be effectively combined, we recommend using both to prevent cherry-picking while improving data exploration.

Recommendation 4: Continually adapt to ever-evolving misleading strategies. Importantly, we found that cherry-picking is an *adversarial process*. Creators adapt to effective guardrails and find new ways to cherry-pick. Previous research shows that people often crop visualization screenshots, add misleading text and annotations [40]. Our experiments confirm that authors shift focus to specific salient data features to influence readers' interpretations. As a result, it is essential to continually adapt to ever-evolving misinformation arguments and to iteratively update the designs to ensure their clarity. Analyzing usage logs from the data explorer could offer insights into the data subsets creators focus on, guiding future guardrail improvements.

10 Conclusion and Future Work

In this paper, we describe an approach to designing technical interventions against the misuse of data visualizations in support of misinformation. We examine cherry-picking in visualization through the lens of threat modeling and describe the design space of *guardrails*: interventions incorporating contextual data that would expose cherry-picking if it is there, and not interfere with interpretation if not. Our experiments find that guardrails make it more difficult to create cherry-picked charts and encourage viewer skepticism; however, the difficulty in implementing successful guardrails lies in drawing the audiences attention to them.

We are hopeful that data exploration platforms adopt similar interventions in their designs. A review of visualization flaws from designers' perspective by Lan and Liu [35] highlights multiple stages in the design workflow where such flaws can emerge, uncovering opportunities for more targeted interventions. Implementing guardrails and other visualization threat mitigation strategies in practice would allow future work to examine the role that they play in complex real-world contexts. As discussed, misinformation using data visualizations is adversarial and a wicked problem. Because of this, studying real-world adoption of guardrails would be especially important: while the results of our experiments show moderate effects of guardrails encouraging skepticism, it is challenging to predict the exact effects of guardrails on online data discourse.

We hypothesize that beyond influencing an individual in isolation, guardrails could have indirect effects in a world where the general public is familiar with their use. For instance, guardrails could provide evidence that triggers a fact-checking discussion on social media, while guardrails cropped out of a screenshot could alert the audience of tampering attempts. Guardrails surfaced in a chart, even if missed by the original poster, can be referenced by community notes and replies to fact-check the post without any external information. The use of guardrails against misinformation could also eventually be associated with reputable sources and serve as a trustworthiness indicator in and of itself [39]. Aside from adapting and evolving their tactics, it is also likely that malicious actors would migrate to other, nonguardrailed platforms, and create a demand for "alternative" data exploration sites.

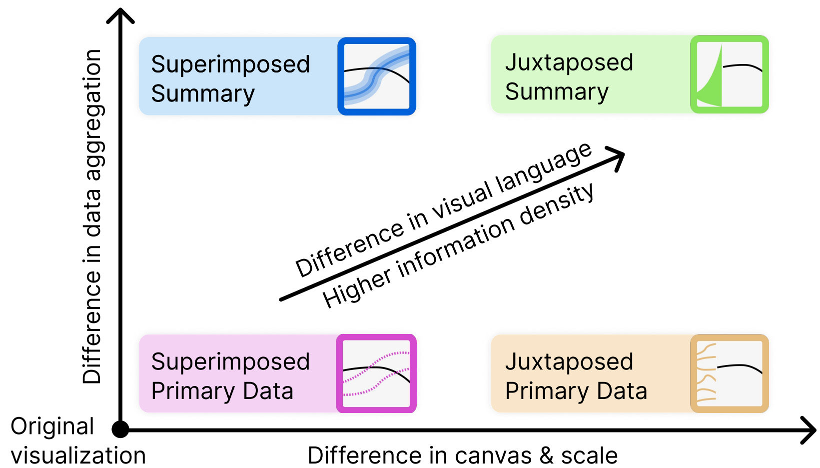


Figure 14: This overview highlights the design trade-offs in the guardrails space, as identified through our design process and evaluation studies. By moving beyond the data aggregation level and chart canvas of the original visualization, designers can incorporate more of the potentially omitted context and information, allowing for a richer guardrail representation. However, this shift also moves away from the visual language of the original chart, which may make it harder for the audience to accurately interpret the guardrail. Balancing this tension is crucial when designing guardrails in data explorers.

In addition to studying the effects of guardrail adoption, future work should also examine strategies to make guardrails—in particular Juxtaposed designs—less amenable to cropping, as well as the effects of combining multiple forms of guardrails. Future research should also investigate automated methods of determining appropriate domain data or aggregation type to be used as context.

Acknowledgments

We thank the reviewers for their constructive feedback throughout the revision process, which helped substantially strengthen the manuscript. This work is supported by the National Science Foundation (IIS 2041136, IIS 1751238, and CNS 2213756).

References

[1] Zhila Aghajari, Eric P. S. Baumer, and Dominic DiFranzo. 2023. Reviewing Interventions to Address Misinformation: The Need to Expand Our Vision Beyond an Individualistic Focus. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–34. doi:10.1145/3579520

[2] Chittaranjan Andrade. 2021. HARKing, Cherry-Picking, P-Hacking, Fishing Expeditions, and Data Dredging and Mining as Questionable Research Practices. *The Journal of Clinical Psychiatry* 82, 1 (Feb. 2021), 25941. doi:10.4088/JCP.20f13804

[3] Abolfazl Asudeh, H. V. Jagadish, You (Will) Wu, and Cong Yu. 2020. On Detecting Cherry-Picked Trendlines. *Proceedings of the VLDB Endowment* 13, 6 (Feb. 2020), 939–952. doi:10.14778/3380750.3380762

[4] Alberto Barchetti, Emma Neybert, Susan Powell Mantel, and Frank R. Kardes. 2022. The Half-Truth Effect and Its Implications for Sustainability. *Sustainability* 14, 11 (June 2022), 6943. doi:10.3390/su14116943

[5] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57, 1 (1995), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

[6] Joyce Berg, John Dickhaut, and Kevin McCabe. 1995. Trust, Reciprocity, and Social History. *Games and Economic Behavior* 10, 1 (July 1995), 122–142. doi:10.1006/game.1995.1027

[7] Binance. 2013. CoinMarketCap. Binance.

[8] John Bollinger. 2002. *Bollinger on Bollinger Bands.* McGraw-Hill, New York.

[9] Zoe Braiterman, Adam Shostack, Jonathan Marcil, Stephen de Vries, Irene Michlin, Kiim Wuyts, Robert Hurlbut, Brooke S.E. Schoenfeld, Fraser Scott, Matthew Coles, Chris Romeo, Alyssa Miller, Izar Tarandach, Avi Douglen, and Marc French. 2020. Threat Modeling Manifesto. https://www.threatmodelingmanifesto.org/.

[10] Fionn Büttner, Elaine Toomey, Shane McClean, Mark Roe, and Eamonn Delahunt. 2020. Are Questionable Research Practices Facilitating New Discoveries in Sport and Exercise Medicine? The Proportion of Supported Hypotheses Is Implausibly High. *British Journal of Sports Medicine* 54, 22 (Nov. 2020), 1365–1371. doi:10.1136/bjsports-2019-101863

[11] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman. 1999. *Readings in Information Visualization: Using Vision to Think.* Morgan Kaufmann Publishers Inc., Burlington, MA.

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

[13] Michael Correll. 2023. Teru Teru Bōzu: Defensive Raincloud Plots. *Computer Graphics Forum* 42, 3 (2023), 235–246. doi:10.1111/cgf.14826

[14] 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.* ACM, Honolulu HI USA, 1–12. doi:10.1145/3313831.3376222

[15] Erick Cuenca, Arnaud Sallaberry, Florence Y. Wang, and Pascal Poncelet. 2018. MultiStream: A Multiresolution Streamgraph Approach to Explore Hierarchical Time Series. *IEEE Transactions on Visualization and Computer Graphics* 24, 12 (Dec. 2018), 3160–3173. doi:10.1109/TVCG.2018.2796591

[16] Zachary T Cutler, Kiran Gadhave, and Alexander Lex. 2020. Trrack: A Library for Provenance Tracking in Web-Based Visualizations. In *IEEE Visualization Conference (VIS).* IEEE, Salt Lake City, UT, 116–120. doi:10.1109/VIS47514.2020.00030

[17] Yiren Ding, Jack Wilburn, Hilson Shrestha, Akim Ndlovu, Kiran Gadhave, Carolina Nobre, Alexander Lex, and Lane Harrison. 2023. reVISit: Supporting Scalable Evaluation of Interactive Visualizations. In *2023 IEEE Visualization and Visual Analytics (VIS).* IEEE, Melbourne, Australia, 31–35. doi:10.1109/VIS54172.2023.00015

[18] Steven P. Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L. Schwartz, and Scott R. Klemmer. 2011. Parallel Prototyping Leads to Better Design Results, More Divergence, and Increased Self-Efficacy. *ACM Transactions on Computer-Human Interaction* 17, 4 (Dec. 2011), 18:1–18:24. doi:10.1145/1879831.1879836

[19] Ève Dubé, Jeremy K. Ward, Pierre Verger, and Noni E. MacDonald. 2021. Vaccine Hesitancy, Acceptance, and Anti-Vaccination: Trends and Future Prospects for Public Health. *Annual Review of Public Health* 42, 1 (April 2021), 175–191. doi:10.1146/annurev-publhealth-090419-102240

[20] Hamza Elhamdadi, Aimen Gaba, Yea-Seul Kim, and Cindy Xiong. 2022. How Do We Measure Trust in Visual Data Communication?. In *2022 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV).* IEEE, Oklahoma City, OK, USA, 85–92. doi:10.1109/BELIV57783.2022.00014

[21] Hamza Elhamdadi, Lace Padilla, and Cindy Xiong. 2022. Using Processing Fluency as a Metric of Trust in Scatterplot Visualizations. *arXiv* 1 (2022), 8 pages. doi:10.48550/ARXIV.2209.14340

[22] Federal Reserve Bank of St. Louis. 1997. FRED, Federal Reserve Economic Data. Federal Reserve Bank Of St. Louis.

[23] Aimen Gaba, Zhanna Kaufman, Jason Cheung, Marie Shvakel, Kyle Wm. Hall, Yuriy Brun, and Cindy Xiong Bearfield. 2024. My Model Is Unfair, Do People Even Care? Visual Design Affects Trust and Perceived Bias in Machine Learning. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (Jan. 2024), 327–337. doi:10.1109/TVCG.2023.3327192

[24] 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*. ACM, Hamburg Germany, 1–18. doi:10.1145/3544548.3581406

[25] Lily W Ge, Matthew Easterday, Matthew Kay, Evanthia Dimara, Peter Cheng, and Steven L Franconeri. 2024. V-FRAMER: Visualization Framework for Mitigating Reasoning Errors in Public Policy. *Proceedings of CHI 2024, ACM Conference on Human Factors in Computing Systems* 1, 1 (2024), 15 pages. doi:10.1145/3613904.3642750

[26] N. Gershon. July-Aug./1998. Visualization of an Imperfect World. *IEEE Computer Graphics and Applications* 18, 4 (July-Aug./1998), 43–45. doi:10.1109/38.689662

[27] M. Gleicher. 2018. Considerations for Visualizing Comparison. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 413–423. doi:10.1109/TVCG.2017.2744199

[28] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D. Hansen, and Jonathan C. Roberts. 2011. Visual Comparison for Information Visualization. *Information Visualization* 10, 4 (2011), 289–309. doi:10.1177/1473871611416549

[29] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, Peter A. Hancock and Najmedin Meshkati (Eds.). Human Mental Workload, Vol. 52. North-Holland, Amsterdam, Netherlands, 139–183. doi:10.1016/S0166-4115(08)62386-9

[30] Jeffrey Heer, Nicholas Kong, and Maneesh Agrawala. 2009. Sizing the Horizon: The Effects of Chart Size and Layering on the Graphical Perception of Time Series Visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, Boston, MA, 1303–1312. doi:10.1145/1518701.1518897

[31] Mark Hoofnagle and Chris Jay Hoofnagle. 2007. What Is Denialism? *SSRN Electronic Journal* 1, 1 (2007), 14 pages. doi:10.2139/ssrn.4002823

[32] Waqas Javed and Niklas Elmqvist. 2010. Stack Zooming for Multi-Focus Interaction in Time-Series Data Visualization. In *Proceedings of the IEEE Symposium on Pacific Visualization (PacificVis '10)*. IEEE, Taipei, Taiwan, 33–40. doi:10.1109/PACIFICVIS.2010.5429613

[33] Waqas Javed, Bryan McDonnel, and Niklas Elmqvist. 2010. Graphical Perception of Multiple Time Series. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '10)* 16, 6 (2010), 927–934.

[34] Alex Kale, Ziyang Guo, Xiao Li Qiao, Jeffrey Heer, and Jessica Hullman. 2023. EVM: Incorporating Model Checking into Exploratory Visual Analysis. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2023), 1–11. doi:10.1109/TVCG.2023.3326516

[35] Xingyu Lan and Yu Liu. 2025. "I Came Across a Junk": Understanding Design Flaws of Data Visualization from the Public's Perspective. *IEEE Transactions on Visualization and Computer Graphics* 31, 1 (Jan. 2025), 393–403. doi:10.1109/TVCG.2024.3456341

[36] Po-Ming Law, Leo Yu-Ho Lo, Alex Endert, John Stasko, and Huamin Qu. 2021. Causal Perception in Question-Answering Systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–15. doi:10.1145/3411764.3445444

[37] Yin Lin, Brit Youngmann, Yuval Moskovitch, H. V. Jagadish, and Tova Milo. 2021. On Detecting Cherry-Picked Generalizations. *Proceedings of the VLDB Endowment* 15, 1 (Sept. 2021), 59–71. doi:10.14778/3485450.3485457

[38] Yin Lin, Brit Youngmann, Yuval Moskovitch, H. V. Jagadish, and Tova Milo. 2022. OREO: Detection of Cherry-Picked Generalizations. *Proceedings of the VLDB Endowment* 15, 12 (Aug. 2022), 3570–3573. doi:10.14778/3554821.3554846

[39] Maxim Lisnic, Alexander Lex, and Marina Kogan. 2024. 'Yeah, This Graph Doesn't Show That': Analysis of Online Engagement with Misleading Data Visualizations. In *SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM, Honolulu, HI USA, 1–14. doi:10.1145/3613904.3642448

[40] Maxim Lisnic, Cole Polychronis, Alexander Lex, and Marina Kogan. 2023. Misleading Beyond Visual Tricks: How People Actually Lie with Charts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–21. doi:10.1145/3544548.3580910

[41] 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. doi:10.1111/cgf.14559

[42] A. M. MacEachren, R. E. Roth, J. O'Brien, B. Li, D. Swingley, and M. Gahegan. 2012. Visual Semiotics & Uncertainty Visualization: An Empirical Study. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (Dec. 2012), 2496–2505. doi:10.1109/TVCG.2012.279

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

[44] Milton Mayer and Richard J. Evans. 2017. *They Thought They Were Free: The Germans, 1933-45*. The University of Chicago Press, Chicago.

[45] Nicholas Micallef, Vivienne Armacost, Nasir Memon, and Sameer Patil. 2022. True or False: Studying the Work Practices of Professional Fact-Checkers. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (March 2022), 1–44. doi:10.1145/3512974

[46] Dominik Moritz and Danyel Fisher. 2018. Visualizing a Million Time Series with the Density Line Chart. *arXiv:1808.06019 [cs]* 1, 1 (Aug. 2018), 9 pages. arXiv:1808.06019 [cs]

[47] Bryce Morrow, Trevor Manz, Arlene E. Chung, Nils Gehlenborg, and David Gotz. 2019. Periphery Plots for Contextualizing Heterogeneous Time-Based Charts. In *2019 IEEE Visualization Conference (VIS)*. IEEE, Vancouver, BC, Canada, 1–5. doi:10.1109/VISUAL.2019.8933582

[48] Brian Ondov, Nicole Jardine, Niklas Elmqvist, and Steven Franconeri. 2019. Face to Face: Evaluating Visual Comparison. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 861–871. doi:10.1109/TVCG.2018.2864884

[49] Lace Padilla, Racquel Fygenson, Spencer C. Castro, and Enrico Bertini. 2022. Multiple Forecast Visualizations (MFVs): Trade-offs in Trust and Performance in Multiple COVID-19 Forecast Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 1–11. doi:10.1109/TVCG.2022.3209457

[50] Lace Padilla, Matthew Kay, and Jessica Hullman. 2022. Uncertainty Visualization. In *Computational Statistics in Data Science*, Hao Helen Zhang, Richard Levine, Thomas Lee, and Walter Piegorsch (Eds.). Vol. 21. Wiley, Hoboken, NJ.

[51] Gordon Pennycook, Adam Bear, Evan T. Collins, and David G. Rand. 2020. The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. *Management Science* 66, 11 (2020), 4944–4957. doi:10.1287/mnsc.2019.3478

[52] Scott Plous. 2007. *The Psychology of Judgment and Decision Making*. Mcgraw-Hill Higher Education, New York.

[53] Timothy J. Prestby. 2023. Trust in Maps: What We Know and What We Need to Know. *Cartography and Geographic Information Science* 52, 1 (Nov. 2023), 1–18. doi:10.1080/15230406.2023.2281306

[54] Xiaoying Pu and Matthew Kay. 2018. The Garden of Forking Paths in Visualization: A Design Space for Reliable Exploratory Visual Analytics : Position Paper. In *2018 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)*. IEEE, Berlin, Germany, 37–45. doi:10.1109/BELIV.2018.8634103

[55] Ghulam Jilani Quadri, Arran Zeyu Wang, Zhehao Wang, Jennifer Adorno, Paul Rosen, and Danielle Albers Szafir. 2024. Do You See What I See? A Qualitative Study Eliciting High-Level Visualization Comprehension. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–26. doi:10.1145/3613904.3642813

[56] Laura Reina. 1996. Why Movie Blurbs Avoid Newspapers.. https://www.thefreelibrary.com/Why+movie+blurbs+avoid+newspapers-a018662749.

[57] Horst W. J. Rittel and Melvin M. Webber. 1973. Dilemmas in a General Theory of Planning. *Policy Sciences* 4, 2 (June 1973), 155–169. doi:10.1007/BF01405730

[58] Hans Rosling and Zhongxing Zhang. 2011. Health Advocacy with Gapminder Animated Statistics. *Journal of Epidemiology and Global Health* 1, 1 (2011), 11. doi:10.1016/j.jegh.2011.07.001

[59] Adam Shostack. 2014. *Threat Modeling: Designing for Security*. Wiley, Indianapolis, IN.

[60] Richard Somerville. 2011. Climate Science and EPA's Greenhouse Gas Regulations. Testimony before the US House of Representatives Committee on Energy and Commerce Subcommittee on Energy and Power. https://www.govinfo.gov/content/pkg/CHRG-112hhrg66704/html/CHRG-112hhrg66704.htm.

[61] Chase Stokes, Chelsea Sanker, Bridget Cogley, and Vidya Setlur. 2024. From Delays to Densities: Exploring Data Uncertainty through Speech, Text, and Visualization. *Computer Graphics Forum* 43, 3 (June 2024), e15100. doi:10.1111/cgf.15100

[62] Judi Thomson, Elizabeth Hetzler, Alan MacEachren, Mark Gahegan, and Misha Pavel. 2005. A Typology for Visualizing Uncertainty. In *Visualization and Data Analysis 2005*, Vol. 5669. SPIE, Bellingham, WA, 146–157. doi:10.1117/12.587254

[63] Edward Tufte. 2001. *The Visual Display of Quantitative Information* (2nd ed.). Graphics Press, Cheshire, CT, USA.

[64] Amos Tversky and Daniel Kahneman. 1973. Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology* 5, 2 (Sept. 1973), 207–232. doi:10.1016/0010-0285(73)90033-9

[65] Amos Tversky and Daniel Kahneman. 1974. *Judgment under Uncertainty: Heuristics and Biases*. Vol. 185. Cambridge University Press, Cambridge, UK.

[66] Emily Wall, Leslie M. Blaha, Lyndsey Franklin, and Alex Endert. 2017. Warning, Bias May Occur: A Proposed Approach to Detecting Cognitive Bias in Interactive Visual Analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, Phoenix, AZ, 104–115. doi:10.1109/VAST.2017.8585669

[67] Emily Wall, Arpit Narechania, Adam Coscia, Jamal Paden, and Alex Endert. 2022. Left, Right, and Gender: Exploring Interaction Traces to Mitigate Human Biases. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (Jan. 2022), 966–975. doi:10.1109/TVCG.2021.3114862

[68] Nathan Walter and Riva Tukachinsky. 2020. A Meta-Analytic Examination of the Continued Influence of Misinformation in the Face of Correction: How Powerful Is It, Why Does It Happen, and How to Stop It? *Communication Research* 47, 2 (March 2020), 155–177. doi:10.1177/0093650219854600

[69] Junpeng Wang, Subhashis Hazarika, Cheng Li, and Han-Wei Shen. 2019. Visualization and Visual Analysis of Ensemble Data: A Survey. *IEEE Transactions on Visualization and Computer Graphics* 25, 9 (Sept. 2019), 2853–2872. doi:10.1109/TVCG.2018.2853721

[70] Zezhong Wang, Jacob Ritchie, Jingtao Zhou, Fanny Chevalier, and Benjamin Bach. 2021. Data Comics for Reporting Controlled User Studies in Human-Computer Interaction. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (Feb. 2021), 967–977. doi:10.1109/TVCG.2020.3030433

[71] R. T. Whitaker, M. Mirzargar, and R. M. Kirby. 2013. Contour Boxplots: A Method for Characterizing Uncertainty in Feature Sets from Simulation Ensembles. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (Dec. 2013), 2713–2722. doi:10.1109/TVCG.2013.143

[72] Wesley Willett, Jeffrey Heer, and Maneesh Agrawala. 2007. Scented Widgets: Improving Navigation Cues with Embedded Visualizations. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '07)* 13, 6 (2007), 1129–1136. doi:10.1109/TVCG.2007.70589

[73] Cindy Xiong, Lace M. K. Padilla, Kent Grayson, and Steven Franconeri. 2019. Examining the Components of Trust in Map-Based Visualizations. *EuroVis Workshop on Trustworthy Visualization (TrustVis)* 1, 1 (2019), 5 pages. doi:10.2312/TRVIS.20191186

[74] Yahoo! 1997. Yahoo! Finance. Yahoo! Inc.

[75] Fumeng Yang, Chloe Rose Mortenson, Erik Nisbet, Nicholas Diakopoulos, and Matthew Kay. 2024. In Dice We Trust: Uncertainty Displays for Maintaining Trust in Election Forecasts Over Time. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–24. doi:10.1145/3613904.3642371

[76] Emanuel Zgraggen, Zheguang Zhao, Robert Zeleznik, and Tim Kraska. 2018. Investigating the Effect of the Multiple Comparisons Problem in Visual Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, Montreal, QC, Canada, 479. doi:10.1145/3173574.3174053

[77] Chengbo Zheng and Xiaojuan Ma. 2022. Evaluating the Effect of Enhanced Text-Visualization Integration on Combating Misinformation in Data Story. In *2022 IEEE 15th Pacific Visualization Symposium (PacificVis)*. IEEE, Tsukuba, Japan, 141–150. doi:10.1109/PacificVis53943.2022.00023

[78] Qian Zhu, Leo Yu-Ho Lo, Meng Xia, Zixin Chen, and Xiaojuan Ma. 2022. Bias-Aware Design for Informed Decisions: Raising Awareness of Self-Selection Bias in User Ratings and Reviews. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–31. doi:10.1145/3555597