

Data Guards: Challenges and Solutions for Fostering Trust in Data

Nicole Sultanum*
Tableau Research, Seattle, WA, USA

Dennis Bromley†
Tableau Research, Seattle, WA, USA

Michael Correll‡
Northeastern University, Portland, ME, USA

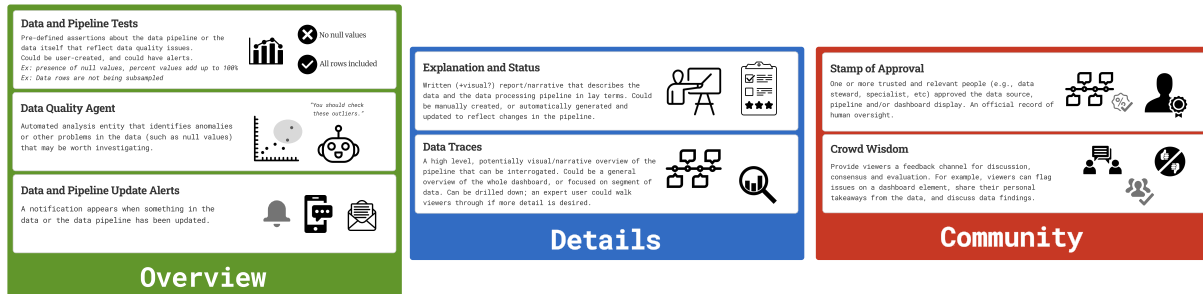


Figure 1: Proposed solutions for establishing trust in data, grouped into three categories: **Overview**, **Details**, and **Community**. Solutions were derived from a series of data-producer interviews based upon interviewee’s current data-trust obstacles and opportunities, and validated by data *consumers* in a card sorting exercise (Figure 2). Card design used in studies is illustrated above.

ABSTRACT

From dirty data to intentional deception, there are many threats to the validity of data-driven decisions. Making use of data, especially new or unfamiliar data, therefore requires a degree of trust or verification. How is this trust established? In this paper, we present the results of a series of interviews with both producers and consumers of *data artifacts* (outputs of data ecosystems like spreadsheets, charts, and dashboards) aimed at understanding strategies and obstacles to building trust in data. We find a recurring need, but lack of existing standards, for data validation and verification, especially among data consumers. We therefore propose a set of *data guards*: methods and tools for fostering trust in data artifacts.

1 INTRODUCTION

Errors in data-driven decision-making can have drastic consequences, from the copy and paste errors that some suggest led to over 100,000 excess deaths [15, 43] to the intentional deception that produced high-profile scandals in biology and psychology [23]. Despite the risks, people at all levels of data work must continue to make data-driven decisions in both personal and professional capacities. But people need to *trust* their data in order make decisions. How does one build trust, especially in new or unfamiliar data prepared or curated by others?

Prior work suggests that strategies for building trust in data are often informal (such as “eyeballing” of tabular data [2]), often undocumented [33], and that the resulting trustworthiness of data artifacts like visualizations can be difficult to measure [9]. Existing tooling for validation and verification are often targeted towards users who are creating or manipulating datasets [18], rather than the broader category of what prior work has called *data workers*: those who regularly work with data but do not identify as data scientists [25, 40]. These data workers may in turn be *consumers* of data artifacts: asked to trust in data they did not produce [40].

In this work, we contribute findings from a series of interviews to understand the struggles and strategies of data workers when vet-

ting or validating *data artifacts* (§3). Participants report frustration with current tooling to support these strategies, from the inability to verify high-level assumptions to the dependence on third parties to diagnose issues. Based on these findings and inspired by work in data transparency, we propose a set of consumer-centered protocols we term *data guards*: ways of fostering trust or transparency in data communication that are easy for *producers* (i.e., data artifact creators) to author, and for *consumers* to understand (§4). We end with a call to action for the visualization community to more tightly integrate data guards into the design, evaluation, and communication of data artifacts (§5).

2 RELATED WORK

Our work focuses on how *consumers* of data artifacts build trust, validate, or otherwise confirm insights they glean from data, and strategies that *producers* of these artifacts can employ to assist in or otherwise augment these trust-building processes. While there is extensive existing literature on data *cleaning* [7, 17, 26], data *profiling* [30], data *curation* [29], and otherwise avoiding *dirty* data [20] for those creating datasets, we focus on both different audiences (those who make use of these resulting “clean” datasets, who may or may not be the same populations as those who created them), and different tasks (confirming that a finding from, e.g., a table or dashboard, is sufficiently reliable).

We focus on three areas of prior work: analyses of work practices and sensemaking among data workers—particularly data consumers; analyses of errors, omissions, and verification strategies in data-driven decision-making; and tooling and strategies for verification and validation in analytics.

2.1 Data Workers and Trust-Building

Prior works on trust and verification, such as Kandel et al. [17] and surveys of data quality management such as Liu et al. [26], often focus on data scientists charged with the *creation* of high-quality datasets. We instead shift our attention to a particular subset of “data workers” [2, 8, 40] that we call *consumers*: that is, data workers who make use of data artifacts created by others.

In a study on the related process of data profiling by analysts, Ruddle et al. [33] notably claim that “most analysts perform profiling in an ad-hoc manner, following an undocumented process that makes data profiling more an art than a science.” This finding aligns with Bartram et al.’s [2] survey of data munging and federation

*e-mail: nsultanum@tableau.com

†e-mail: dbromley@tableau.com

‡e-mail: m.correll@northeastern.edu

practices of data workers, which found that qualitative processes like “eyeballing” were prominent in building trust in data, often in conjunction with extensive use of spreadsheet tools as ways to provide a sense of direct control or ownership over their data. The participants in the more general analysis of analytic processes in Kandogan et al. [19] similarly report that “that 80 percent of any effort is getting the data together and understanding data quality”, but that challenges of provenance and communication can result in data quality concerns that require “checks and balances” to detect and correct. Of note is that data quality issues do not only arise either in the creation of datasets or in the interpretation of final visualizations, but rather that, as per Sacha et al. [34], uncertainties occur at all stages of the analytics process, with impacts that flow between stages. Chu et al. [7] further emphasize that data cleaning and data verification is complex and generally qualitative.

Beyond the steps involved in verification or trust-building, the very notion of what is meant by data quality and trustworthiness is multifaceted. For instance, in a survey of a similar participant pool as our own (viz., “those who use data”), Wang and Strong [42] identify 20 dimensions under 4 categories that their participants associated with data quality, several which touch on aspects of trust: accuracy of data (e.g., believability, accuracy, reputation), relevancy of data (e.g. timeliness, value-add), representation of data (e.g., interpretability, ease of understanding), and accessibility of data.

2.2 Smells and Mirages: Threats to Data Reliability

While we acknowledge lack of trust arising from intentional deception [24], unintentional trust issues can arise through an unreliable finding or a misinterpretation of a visualization. McNutt et al. [27] introduce the term “visualization mirage” to refer to erroneous conclusions emerging from cursory readings of a visualization without additional scrutiny. In line with our notion of “data guards”, McNutt et al. propose testing regimes in order to surface potential data quality or uncertainty issues in visualizations. Drawing from the software engineering metaphor of “code smells” (patterns in code that often indicate potential issues), Shome et al. [36] and Foidl et al. [12] propose categorizations for “data smells”, that can point to potential issues with data quality and interpretation. Kim et al. [20] also propose a “taxonomy of dirty data” to illustrate ways that data (especially manually entered or curated data) can be invalid.

Many existing visual analytics systems lack tools or workflows for detecting potentially unreliable or erroneous conclusions. This gap can result in high error rates when assessing the accuracy of insights from visual analytics [45], rates that are only partially improved by including uncertainty information [35]. In keeping with the highly contextual nature of assessing data quality, we note work by Song et al. [39] suggesting the ways in which data quality information is visualized (e.g., whether missing values are elided or indicated with imputed values) can influence the perceived reliability of data sets. The inclusion of explicit visualizations of missing data can even impact the strategies that analysts use when making sense of data [38]. The heterogeneous nature of analytical strategies can, however, exacerbate issues caused by a lack of verification in analytics. For instance, Yanai and Lercher [44] find that participants who approach a dataset with a specific hypothesis in mind were less likely to notice dramatic data quality issues.

2.3 Instrumenting Trust: Profiling and Verification

Given the prominence of data quality concerns in analytics but the lack of standard repeatable and quantitative methods for verifying data quality, there have been a variety of proposed design interventions to assist users in either profiling, wrangling, or verifying data. Kandel et al.’s [18] *Profiler* visualizes potential anomalies in data that may indicate underlying data quality issues or errors in reshaping; Jannah’s [16] *MetaReader* surfaces errors and warnings such as potential mismatches between data columns and inferred seman-

tic roles (e.g., a primary key); Bors et al.’s [3] *MetricDoc* surfaces data quality metrics at both global and individual value levels (e.g., a “plausibility” value flagging potential status as an outlier or data quality concern); and the *Ferret* system by Lange et al. [23] surfaces features that might indicate manipulation in tabular data. However, these systems are mostly designed for those *creating* data artifacts. Relatively fewer works focus on consumers of data artifacts, who have domain expertise but not necessarily data processing expertise or data curation access [40]. Notable exceptions include the metric-based “risk gauges” [46] or quality scores [3], the “metamorphic testing” approach in McNutt et al. [27] to highlight the robustness of visual patterns in charts, and as in Fan et al. [10] to detect potentially deceptive visualization designs. We draw inspiration from these similarly motivated, but conceptually distinct, design interventions when proposing our own data guards.

3 CHARACTERIZING DATA TRUST NEEDS

We conducted two rounds of interviews. In the first round, we interviewed eight expert dashboard *producers* (P01-P08), i.e., professionals who create dashboards for corporate and general public consumption, to understand how they address trust in their dashboard designs and what challenges they (and their users) face in the process (§3.1 and §3.2). Using thematic analysis [5], we organized this feedback into themes that we use to propose seven classes of design solutions, which we collectively call *data guards* (§4.1). We followed up with a second round of interviews with ten professional data *consumers* (C01-C10) to solicit their trust-related experiences when perusing data artifacts for decision making (e.g., dashboards and other forms of dynamic data reports), which helped substantiate challenges (§3.1) and desiderata (§3.2) identified in the first round. We also performed a card-sorting exercise with these participants to encourage ideation and prioritization around our seven proposed data guards (§4.2). Study instruments can be found in Supplemental Materials.

3.1 Barriers for Data Trust

We first present six categories of *barriers* to trust in data (B1-B6) that emerged from our interviews. That is, challenges or contextual factors that complicate the process of building trust.

(B1) Data is heavily context-dependent. A challenge reported by producers working as external consultants — but also a natural consequence of the division of labor between producers and consumers — is a lack of shared domain knowledge [29] and common ground [31]. Lack of familiarity with domain idiosyncrasies such as how to calculate bond rate denominators (e.g., *360 days? Exact days in year, including leap days? Days excluding weekends?* (P03)) can result in errors that would not be caught by generalizable data quality metrics and is important for validation. Yet, domain information is difficult to transfer: “*Sometimes the people who are building these dashboards or building these data sources don’t really know the business as intimately. There’s definitely a disconnect and there’s just a bit of education that needs to happen.*” (C05).

(B2) Detecting issues requires a discerning eye. The opacity of data processing pipelines in data artifacts like dashboards means that consumers might only notice issues when they see something that defies their expectations (e.g., outliers). This was often reported as an intuition based on experience rather than a systematic approach, akin to “data smells” [36, 12]: “*this doesn’t smell right*” (C08). Novice consumers are therefore less likely to detect issues when they arise. If discrepancies are not visible or data not inspectable, trust can be a (perhaps unjustified) default: “*I do trust the data because I don’t have a choice. is that fair? (..) Maybe just because there’s no other way to get that [data], I can’t know the answer, I can’t verify it.*” (C02).

(B3) Data trust builds on human relationships. Data work is collaborative. Most producers (5/8) stated working closely with

consumers (and proxies) to not only bridge the domain knowledge gap, but also build interpersonal trust; e.g., “*They trust me because it’s me. And that’s the professionalism you build up (...). I’ve spent the time to reconcile, the data, I’ve thought about this and the 10 different ways that you’re going to use it.*” (P07). Consumers also state relying on interpersonal trust as part of their assessment of data trust [40], e.g., “*If I was looking at a [data source], and they said [anonymous] certified this I’d be cool. I trust it. Because he knows what he’s talking about.*” (C02); or a community assessment, e.g., “*What gives me confidence is the lack of user feedback that something is wrong.*” (C10). However, when no such relationship exists, trust may be negatively affected, e.g., “*Who’s presenting it? Has this person credibly done this before?*” (C09).

(B4) Trust is hard to build, easy to lose. Beyond human relationships, the ongoing experience consumers have with their data artifacts also influences their perception of trust: “*Some of it comes down to day-to-day working in these tools (...), and feeling like it’s always a consistent experience to use it and that I see the same things each time.*” (C08). Conversely, a lack of experience may induce mistrust: e.g., “*I have definitely doubted newer dashboards that have come out*” (C07); as well as those facing constant issues with particular sources, and having mistrust as a starting point: “*I wish that my mental model was trust but verify. It is not. It is distrust until I can prove it.*” (C02).

(B5) Data definitions are often unclear or ambiguous. Consumers commented that understanding calculations and metrics at a deeper level was a constant challenge and cited various sources of ambiguity, including duplicate or scattered data sources, e.g., “*So I look at reports in [data source 1] and [data source 2]. I do my forecasting on a sheet. And then I also look at the [data source 3]. There is not one single source of truth or one number that I can depend on*” (C05); but also conceptual alignment issues, with different stakeholders interpreting the same data differently: e.g., “*Even within departments, we’re talking different languages. ‘Attrition’ to one person may mean something different; to me it means numbers, while for Human Resources it is attrition of people*” (C05). Consumers often cited a lack of transparency on relevant metadata and documentation to help clarify these ambiguities: “*I could know what made the number if somebody wrote a definition down*” (C02).

(B6) Environments change and processes break. Changes in the data and pipeline introduce risk and can have a significant impact on trust: “*If something made it into production that shouldn’t, it was done when the data updated*” (P06). They explain: “*There have been cases where there were assumptions made in a workbook that might have no longer been true or we didn’t take into account things that came in later on down the line.*” (P06). From a consumer’s perspective, there are also contextual changes: “*Where something changes in an upstream system (...), like business logic. It’s very hard, the business users don’t necessarily tell the people in charge of the data pipes that this changed happened, or they don’t realize that this change happened and therefore suddenly the values are weird*” (C10). As in **(B5)**, a lack of documentation is a primary concern, but also a lack of change tracking or alerts.

3.2 Mitigation Strategies and Desiderata

We collated strategies that our participants followed to establish or repair trust in data artifacts, plus ideas for improvement. We structure these themes as design goals **(G1-G5)**, each associated with one or more corresponding barriers they are meant to overcome.

(G1) Understand domain knowledge (B1, B2, B5). Producers work closely with consumers when building data artifacts. Given the demands on domain knowledge, some also argue it is a “[consumer] job as the one who owns the data” (P05) to validate data artifacts, and that consumers should be given better tools to assess issues e.g., “*put the stop sign that says ‘you got to look at this’.* We can show you where the numbers come from, but we can’t look at

it for you.” (P01). Conversely, consumers also desired tools to help them navigate contextual ambiguity of data artifacts: “*It would be lovely to be able to say ‘I see 20 different calculations called engagement. These are all the different ways that engagement is being calculated’.*” (C06). All these point to a need for domain-aware tools that can better support collaboration and error diagnosis.

(G2) Provide provenance and metadata (B2, B4, B5). Consumers commented wanting more metadata: e.g., “*I want a data definition. Where did this come from? Did it come from [my expected source]? Did it come from somewhere else?*” (C02). This is sometimes addressed in dashboards as “*information icons (...)* with a tool tip on to explain some of the metrics and (...) say, ‘*Hey, if you’re struggling with these metrics, then here’s what they do, and how they work*’ ” (P04), but it is not always present. Alternatively, some consumers also reported going back to the source data to diagnose issues: “*I go in the system and I look at the raw data. I need to understand if there’s a definition problem, or a logic problem. (...)*” (C02). This suggests consumers may benefit from some form of data source and pipeline description in a suitable format and appropriate level of detail (e.g., in situ “*metadata blocks*” [1] or even peripheral “*datasheets*” [13]).

(G3) Communicate high-level trust signals (B1, B2, B3). While comprehensive details are useful for building initial trust and “*chasing smells*”, they are not always needed to maintain existing trust. For example, with data quality controls and trusted oversight in place, having a “*certified*” stamp is helpful: “*When there is investment in a data analyst core and data governance, that’s been vetted and approved everywhere it needs to be approved, then I want to trust it*” (C07). Other visual signals like breadcrumbs for filtering operations may also support data awareness that helps build trust, so “*you’re seeing what actually that view contains*” (P07); or warnings that something is “*in a draft state*” (P04).

(G4) Highlight data and pipeline changes (B4, B5, B6). Change awareness is key. A standard practice is to communicate when data was last updated, e.g., “*Everything is updated at a different cadence, and so in the dashboard, we literally have a footer under every KPI ‘last update by’*” (C10). This informs that there was a change but not what has changed, leading consumers to seek assistance: “*Sometimes I’ll have that conversation with whoever made the updates to know exactly what they updated*” (P06). Consumers might also keep references to past data and pipeline versions (akin to version control): “*I will always take snapshots of the report that I culled my numbers from. So I can link it back to, ‘well on this day and time this is what the report said’*” (C05). Arguably, automation could support or supplant some of these manual strategies: “*it would be cool if [my dashboard] could recognize that something funky happened between yesterday and today and flag that*” (C01)

(G5) Leverage and mediate social trust (B1, B3). Peer relationships play an important role in the data ecosystem as a primary fallback when trust collapses: “*If I don’t [have data governance], I’m gonna ask people I trust*” (C02). Individuals in trusted roles are also sought after to troubleshoot issues, e.g., “*I actually looked for the person who created the dashboard in order to get me that information*” (C03). Peer activity also provides trust signals to consumers: “*you’ve got this dashboard that has two views in the past six months [that should not be trusted]; and there’s this one dashboard that is being only used once a year, but it’s by [the CEO]”* (C03). Ultimately, data intelligence and automation are unlikely to replace trusted human relationships. However, there is little technology support to mediate these relationships currently, and the design space of social and collaborative aspects of data trust remains largely unexplored; e.g., “*I don’t want to have to find the person who can answer the ‘why’. I want this ‘why’ to be automatically redirected to the person who could answer me*” (C03).

	C01	C02	C03	C04	C05	C06	C07	C08	C09	C10
Stamp of Approval (Community)		#4	#1	#2	#2	#1	#4	#1	#1	#3
Explanation and Status (Detail)	#1	#2	#3	#1	#2		#1	#2	#4	#5
Data Traces (Detail)		#1		#1			#2	#4	#6	#4
Data Quality Agent (Overview)	#3	#3			#1	#4	#5		#3	#1
Data and Pipeline Tests (Overview)	#4	#7				#3	#6	#1	#2	#2
Crowd Wisdom (Community)	#2	#5	#2	#2	#3	#2	#3	#3	#5	#7
Data and Pipeline Update Alerts (Overview)	#5	#6					#7	#2	#7	#6

Figure 2: Card sorting results for different data-trust solutions broken out by participant and solution and colored by rank (from #1 to #7). Individual boxes indicate which rank was selected by whom for each solution. Solutions are sorted vertically by recursive ranking: highest number of rank #1's first, followed by rank #2's, and so on.

4 DATA GUARDS

We use the strategies we observed and the gaps in existing processes to propose seven classes of design interventions that we collectively call *data guards* (Fig. 1). These solution ideas were designed as potential trust-enabling additions to data artifacts, with some inspired by current mitigation strategies and some proposing new features. We explain our data guards alongside relevant design goals (§4.1) and share consumer feedback on these ideas (§4.2).

4.1 Data Guards Ideation

Our seven strategies fall into three general clusters (Figure 1). **Overview** strategies provide high-level trust signals for monitoring (1-3). **Details** strategies allow for deep data dives (4-5). And, lastly, **Community** strategies leverage trusted peer relationships (6-7):

1. **Data and Pipeline Tests (G1, G3, G4)**. Inspired by the notion of software assertions and data profiling checklists [33], these strategies outline predefined tests over the data, the data pipeline, and the charts to surface data quality and presentation issues, such as null values, percentages not amounting to 100%, truncated y-axis, or values outside domain thresholds. Rules could potentially be crafted by consumers and reused across data artifacts.

2. **Data Quality Agent (G1, G3)**. Building off of existing data quality [3] or insight reliability [46] metrics, these strategies entail automated analyses to detect “data smells”, thus pointing to potential anomalies or other problems in the data (such as outliers or redundant variables) that consumers may find worth investigating.

3. **Data and Pipeline Change Alerts (G3, G5)**. These strategies employ notification to communicate and summarize changes to the data or the data pipeline since the last time the user interacted with the data artifact. These summaries could augment existing metrics or notifications of data “freshness” [4].

4. **Explanation and Status (G1, G2, G4)**. Drawing from existing efforts to better describe and contextualize data in machine learning, such as Model Cards [28], Data Cards [32] and Data Sheets [13], these strategies employ written, potentially visual, reports describing underlying data and calculations in lay terms. These explanations could be manually created by producers, or automatically generated and updated to reflect changes in the pipeline.

5. **Data Traces (G1, G2)**. These strategies provide a drill-down view of the data pipeline (potentially in narrative form) focusing on a visible data slice, e.g., from selected points in a chart, or from a data filter. Visualizations of data flows and provenance in tools like VisTrails [6] have shown promise for teaching and communication of unfamiliar visualizations [37]. These interventions could also be smaller scale, such as communicating the specific form of aggregation used to generate a mark in a visualization [21].

6. **Stamp of Approval (G3, G5)**. These strategies rely on an explicit record of human oversight. For instance, an official record that relevant experts approved the data source, pipeline and views in this data artifact and a means to reach out to these experts.

7. **Crowd wisdom (G5)**. While the stamp of approval strategy above uses the metaphor of vetting, crowd wisdom makes a less

prescriptive use of a shared community shared for discussion and consensus. Consumers could flag issues, share their personal take-aways from the data, and discuss data findings (as with collaborative visualization environments like ManyEyes [41]). Other forms of leverage of social data could be the use of records of interaction [11] to guide users to important areas of a visualization, or externalizing other people’s expectations of relationships in data [22].

4.2 Consumer Feedback on Data Guards

Consumer rankings for the data guards are shown in Fig. 2. Due to time constraints, some participants chose not to rank certain items, hence the white spaces; some also chose to place two items in the same rank, hence the repeated rankings. In this context, the item most frequently ranked first was **Stamp of Approval** (4 times). Consumers who already have similar experience working with “certified” data sources appreciate the simplicity of having a human-vetted high-level trust signal that did not require additional investigation or action when “*at that uber level where I really don’t touch data*” (C09). Similarly, “*Can I find the data? Can I find the right dashboard? Can I understand the dashboard? And if not, who do I contact? Realistically, given the amount of time, that would be the extent of what I could do.*” (C03).

The next most commonly top-ranked items were **Explanation and Status** (3 times) and **Data Traces** (2 times). They provide more in-depth trust-related information at the expense of more time and effort, aligning with past findings on data worker practices [2, 40]. Consumers argued these strategies support both building and up-holding trust: “*For Explanation and Status, if you don’t use a dashboard for a while, if you’re new to a dashboard, if you’re just finding out about it if you’re nearly hired, the first question you’re going to ask is ‘how do we get to these numbers?’*” (C07); and “*when I find something to be untrustworthy, and I then need to either convince myself it should be trusted or prove to somebody that it is not accurate, the Data Trace tells me how to do that.*” (C02).

While these strategies in the **Details** cluster feature closer to the top, findings show that all proposed data guards were deemed valuable to at least some consumers. All strategies were placed on Top two at least once, and all three clusters have items ranked first. This suggests that consumer trust needs are diverse and that a multi-pronged approach to fostering trust may be most helpful.

5 DISCUSSION AND CLOSING THOUGHTS

Trust in data artifacts is a multifaceted and nebulous concept. Beyond the issues of dirty data or intentional deception, trust in data can be lost as data circulates to new audiences, new artifacts are built on top of old data, or institutional needs or processes change. In general, we see a need for ways to provide context, awareness, and guidance over data artifacts, especially for *consumers* of data artifacts who may have limited access to underlying data collection and curation processes that generated the artifact, who may be encountering the artifact for the first time, or, even, who may have additional domain knowledge and context that suggest threats to the validity or applicability of the data they are given.

Data guards are broad categories of designs to address these unmet trustbuilding needs. While aspects of these strategies have been proposed in prior work, our work reframes and prioritizes them with consumers in mind. Future research is required to further explore how these strategies can be instantiated into analytics tools, and in particular the promises or limitations of automation [14] to supplement the current highly manual and qualitative processes in trust.

These strategies can come at a cost. Data guards are themselves *data artifacts*, which require the same sort of trustbuilding and validation to show their validity. Too much complexity could overwhelm users, and spurious or invalid alerts could erode trust, just as spurious or invalid data undermined our participants’ trust in the artifacts they encountered.

REFERENCES

- [1] B. Bach, E. Freeman, A. Abdul-Rahman, C. Turkay, S. Khan, Y. Fan, and M. Chen. Dashboard design patterns. *IEEE TVCG*, 29(1):342–352, 2022. 3
- [2] L. Bartram, M. Correll, and M. Tory. Untidy data: The unreasonable effectiveness of tables. *IEEE TVCG*, 28(1):686–696, 2021. 1, 4
- [3] C. Bors, T. Gschwandtner, S. Kriglstein, S. Miksch, and M. Pohl. Visual interactive creation, customization, and analysis of data quality metrics. *Journal of Data and Information Quality (JDIQ)*, 10(1):1–26, 2018. 2, 4
- [4] M. Bouzeghoub. A framework for analysis of data freshness. In *Proceedings of the 2004 international workshop on Information quality in information systems*, pp. 59–67, 2004. 4
- [5] V. Braun and V. Clarke. *Thematic analysis.*, chap. 4, pp. 57–71. American Psychological Association, 2012. 2
- [6] S. P. Callahan, J. Freire, E. Santos, C. E. Scheidegger, C. T. Silva, and H. T. Vo. Vistrails: visualization meets data management. In *Proceedings of ACM SIGMOD 2006*, pp. 745–747, 2006. 4
- [7] X. Chu, I. F. Ilyas, S. Krishnan, and J. Wang. Data cleaning: Overview and emerging challenges. In *Proceedings of ACM SIGMOD 2016*, pp. 2201–2206, 2016. 1, 2
- [8] A. Crisan, B. Fiore-Gartland, and M. Tory. Passing the data baton: A retrospective analysis on data science work and workers. *IEEE TVCG*, 27(2):1860–1870, 2020. 1
- [9] H. Elhamedadi, A. Gaba, Y.-S. Kim, and C. Xiong. How do we measure trust in visual data communication? In *2022 IEEE Evaluation and Beyond-Methodological Approaches for Visualization (BELIV)*, pp. 85–92. IEEE, 2022. 1
- [10] A. Fan, Y. Ma, M. Mancenido, and R. Maciejewski. Annotating line charts for addressing deception. In *Proceedings of ACM SIGCHI 2022*, pp. 1–12, 2022. 2
- [11] M. Feng, C. Deng, E. M. Peck, and L. Harrison. Hindsight: Encouraging exploration through direct encoding of personal interaction history. *IEEE TVCG*, 23(1):351–360, 2016. 4
- [12] H. Foidl, M. Felderer, and R. Ramler. Data smells: Categories, causes and consequences, and detection of suspicious data in ai-based systems. In *Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI*, pp. 229–239, 2022. 2
- [13] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. D. Iii, and K. Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021. 3, 4
- [14] J. Heer. Agency plus automation: Designing artificial intelligence into interactive systems. *Proceedings of the National Academy of Sciences*, 116(6):1844–1850, 2019. 4
- [15] T. Herndon, M. Ash, and R. Pollin. Does high public debt consistently stifle economic growth? a critique of reinhart and rogooff. *Cambridge journal of economics*, 38(2):257–279, 2014. 1
- [16] H. M. Jannah. Metareader: A dataset meta-exploration and documentation tool, 2014. 2
- [17] S. Kandel, A. Paepcke, J. M. Hellerstein, and J. Heer. Enterprise data analysis and visualization: An interview study. *IEEE TVCG*, 18(12):2917–2926, 2012. 1
- [18] S. Kandel, R. Parikh, A. Paepcke, J. M. Hellerstein, and J. Heer. Profiler: Integrated statistical analysis and visualization for data quality assessment. In *Proceedings of ACM AVI 2012*, pp. 547–554, 2012. 1, 2
- [19] E. Kandogan, A. Balakrishnan, E. M. Haber, and J. S. Pierce. From data to insight: work practices of analysts in the enterprise. *IEEE computer graphics and applications*, 34(5):42–50, 2014. 2
- [20] W. Kim, B.-J. Choi, E.-K. Hong, S.-K. Kim, and D. Lee. A taxonomy of dirty data. *Data mining and knowledge discovery*, 7:81–99, 2003. 1, 2
- [21] Y. Kim, M. Correll, and J. Heer. Designing animated transitions to convey aggregate operations. *Computer Graphics Forum*, 38(3):541–551, 2019. 4
- [22] Y.-S. Kim, K. Reinecke, and J. Hullman. Data through others’ eyes: The impact of visualizing others’ expectations on visualization interpretation. *IEEE TVCG*, 24(1):760–769, 2017. 4
- [23] D. Lange, S. Sahai, J. M. Phillips, and A. Lex. Ferret: Reviewing Tabular Datasets for Manipulation. *Computer Graphics Forum*, 42(3):187–198, June 2023. 1, 2
- [24] M. Lisnic, C. Polychronis, A. Lex, and M. Kogan. Misleading beyond visual tricks: How people actually lie with charts. In *Proceedings of ACM SIGCHI 2023*, pp. 1–21, 2023. 2
- [25] J. Liu, N. Boukhelifa, and J. R. Eagan. Understanding the role of alternatives in data analysis practices. *IEEE TVCG*, 26(1):66–76, 2019. 1
- [26] S. Liu, G. Andrienko, Y. Wu, N. Cao, L. Jiang, C. Shi, Y.-S. Wang, and S. Hong. Steering data quality with visual analytics: The complexity challenge. *Visual Informatics*, 2(4):191–197, 2018. 1
- [27] A. McNutt, G. Kindlmann, and M. Correll. Surfacing visualization mirages. In *Proceedings of ACM SIGCHI 2020*, pp. 1–16, 2020. 2
- [28] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, and T. Gebru. Model cards for model reporting. In *Proceedings of ACM FAccT 2019*, pp. 220–229, 2019. 4
- [29] M. Muller, I. Lange, D. Wang, D. Piorkowski, J. Tsay, Q. V. Liao, C. Dugan, and T. Erickson. How data science workers work with data: Discovery, capture, curation, design, creation. In *Proceedings of ACM SIGCHI 2019*, pp. 1–15, 2019. 1, 2
- [30] F. Naumann. Data profiling revisited. *ACM SIGMOD Record*, 42(4):40–49, 2014. 1
- [31] G. M. Olson and J. S. Olson. Distance matters. *Human-computer interaction*, 15(2-3):139–178, 2000. 2
- [32] M. Pushkarna, A. Zaldivar, and O. Kjartansson. Data cards: Purposeful and transparent dataset documentation for responsible ai. In *Proceedings of ACM FAccT 2022*, pp. 1776–1826, 2022. 4
- [33] R. A. Ruddle, J. Cheshire, and S. J. Fernstad. Tasks and Visualizations Used for Data Profiling: A Survey and Interview Study. *IEEE TVCG*, pp. 1–12, 2024. 1, 4
- [34] D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis, and D. A. Keim. The Role of Uncertainty, Awareness, and Trust in Visual Analytics. *IEEE TVCG*, 22(1):240–249, 2016. 2
- [35] A. Sarma, X. Pu, Y. Cui, E. T. Brown, M. Correll, and M. Kay. Odds and Insights: Decision Quality in Exploratory Data Analysis Under Uncertainty. preprint, Open Science Framework, 2024. 2
- [36] A. Shome, L. Cruz, and A. Van Deursen. Data smells in public datasets. In *Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI*, pp. 205–216, 2022. 2
- [37] C. T. Silva, E. Anderson, E. Santos, and J. Freire. Using vistrails and provenance for teaching scientific visualization. *Computer Graphics Forum*, 30(1):75–84, 2011. 4
- [38] H. Song, Y. Fu, B. Saket, and J. Stasko. Understanding the effects of visualizing missing values on visual data exploration. In *2021 IEEE Visualization Conference (VIS)*, pp. 161–165. IEEE, 2021. 2
- [39] H. Song and D. A. Szafr. Where’s my data? evaluating visualizations with missing data. *IEEE TVCG*, 25(1):914–924, 2018. 2
- [40] M. Tory, L. Bartram, B. Fiore-Gartland, and A. Crisan. Finding their data voice: Practices and challenges of dashboard users. *IEEE Computer Graphics and Applications*, 43(1):22–36, 2021. 1, 2, 3, 4
- [41] F. B. Viegas, M. Wattenberg, F. Van Ham, J. Kriss, and M. McKeeon. Manyeyes: a site for visualization at internet scale. *IEEE TVCG*, 13(6):1121–1128, 2007. 4
- [42] R. Y. Wang and D. M. Strong. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4):5–33, 1996. 2
- [43] J. Watkins, W. Wulaningsih, C. Da Zhou, D. C. Marshall, G. D. Sylianteng, P. G. D. Rosa, V. A. Miguel, R. Raine, L. P. King, and M. Maruthappu. Effects of health and social care spending constraints on mortality in england: a time trend analysis. *BMJ open*, 7(11):e017722, 2017. 1
- [44] I. Yanai and M. Lercher. A hypothesis is a liability. *Genome Biology*, 21(1):231, s13059–020–02133–w, Dec. 2020. 2
- [45] E. Zraggen, Z. Zhao, R. Zeleznik, and T. Kraska. Investigating the Effect of the Multiple Comparisons Problem in Visual Analysis. In *Proceedings of ACM SIGCHI 2018*, pp. 1–12. ACM, 2018. 2
- [46] Z. Zhao, E. Zraggen, L. De Stefani, C. Binnig, E. Upfal, and T. Kraska. Safe visual data exploration. In *Proceedings of ACM SIGMOD 2017*, pp. 1671–1674, 2017. 2, 4