Where's my data? Visualizing Missing Values in Time Series Data

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

Leave Authors Anonymous
for Submission
City, Country
e-mail address

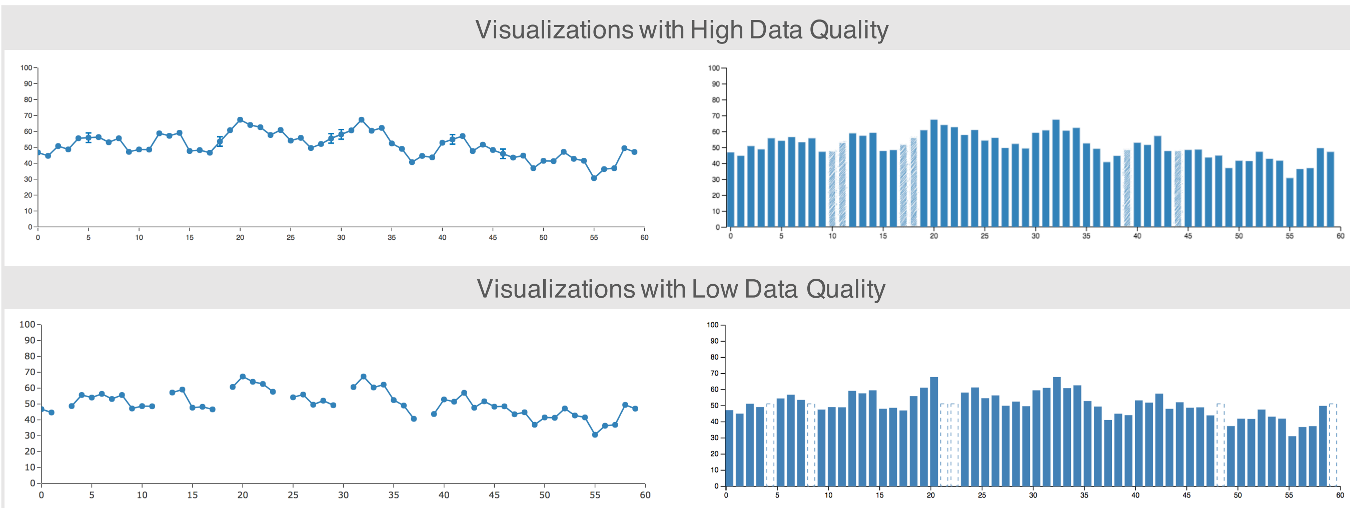


Figure 1. We measured factors influencing response accuracy, data quality, and confidence in interpretation for time series data with missing values. Our studies found that visualizations associated with uncertainty that preserve visual continuity lead to high perceived data quality, while representations that break continuity decrease these perceptions and can bias interpretation. These effects are mediated by the methods used to impute missing data.

ABSTRACT

Many real-world datasets are incomplete due to factors such as data collection failures or misalignments between fused datasets. Visualizations of incomplete datasets should allow analysts to draw conclusions from their data while effectively reasoning about the quality of the data and resulting conclusions. We examine how the methods used to impute and visualize missing data may influence analysts' perceptions of data and their confidence in their conclusions. Our experiments measure how different representations and imputation methods for visualizing incomplete datasets change analyst confidence and perceived data quality in time series visualizations. Our results provide preliminary guidance for visualization designers to consider when working with incomplete data in different domains and scenarios.

CHI'16, May 07–12, 2016, San Jose, CA, USA

ISBN 123-4567-24-567/08/06. . . $15.00

DOI: http://dx.doi.org/10.475/123_4

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; See http://acm.org/about/class/1998/ for the full list of ACM classifiers.

Author Keywords

Information Visualization; Graphical Perception; Time Series Data; Data Wrangling; Imputation

INTRODUCTION

Visualizations allow people to analyze and interpret data to understand current phenomena and help guide informed decisions. However, analysts often must make decisions using imperfect datasets. These datasets may be missing datapoints due to factors such as failures in the data collection pipeline, fusing data at different granularities, or censored data due to information privacy. As part of the data wrangling process, visualizations have several choices for dealing with missing data, including not encoding missing elements or *imputing* new data (calculating substitute values) based on existing data. Prior studies show that the ways we represent data influence how accurately people interpret and change their confidence in their data and results [16, 19, 33, 44]. We hypothesize that the ways we impute and visualize missing data may also bias analysts perceptions of that data.

We measure how imputation and visualization techniques influence perceived confidence and data quality for visualizing incomplete datasets. For most analysis tasks, analyst confidence in data and perceptions of data quality should decrease as the amount of missing data increases. We explore how three different categories of visual manipulations might manipulate perceived data quality: encodings indicating uncertainty, salient encodings that draw analyst attention, and encodings using visual metaphors of missing information (c.f., Fig. 2). We measure the effects of these visual attributes on perceived data quality, result confidence, and response accuracy in two common visualizations: line graphs and bar charts.

We also explore how methods of imputing missing values might additionally shift perceptions of data quality and bias responses. Systems use imputation to compute values that approximate missing datapoints to support analysis. As missing data is itself a type of data (it is information that no values are available for some unspecified reason), imputation allows systems to indicate where data is unexpectedly absent and provide principled approximations to avoid potential misinterpretation of absent data values [7]. Imputing values also allows systems to indicate potential threats to data quality by providing visual anchors the analysts can readily enumerate and contextualize quality errors [5, 45]. We focus on three common imputation methods encountered in current visualization systems: ad-hoc zero-filling, local interpolation, and marginal means (Fig. 3).

We compare imputation and visualization methods in two crowdsourced studies measuring the effects of these factors on analysts' accuracy, confidence in their conclusions, and perceived data quality. Our results indicate significant tradeoffs between different approaches for managing missing data. We found that uncertainty visualizations lead to higher perceived data quality and more accurate interpretation, but only when the visual continuity of the overall encoding is preserved. Using visual metaphors associated with missing value significantly degraded perceived quality and even led to misinterpretation. We discuss our findings in terms of how visualizations might leverage imputation and visualization to appropriately manipulate perceived data quality in different scenarios.

RELATED WORK

Missing data is typically a challenge associated with "dirty data"—datasets containing missing data, incorrect data, misalignments, and other such anomalies that may lead to erroneous conclusions [34]. Missing data can occur throughout the data lifecycle and has significant implications for analyst trust in data [42]. These implications can be especially problematic for data visualizations as little empirical understand exists to guide how visualizations can balance between indicating the presence of dirty data and not distracting from or biasing of the rest of the data [31]. Our research builds such knowledge by measuring the influence of various design factors for missing data using design choices that touch on various components of reasoning with dirty data: salience, uncertainty, and metaphor.

Methods for Analyzing Incomplete Data

Missing data can arise at all points in the data lifecycle, including during data capture, storage, updates, transmission, access, deletions, and purges [34]. A scraping process might fail due to an interrupted script, packet loss, or memory errors. Subsets of data may be withheld due to privacy considerations [20]. Part of the process of data wrangling [31, 23] is locating missing data and deciding how to manage it. In many cases, systems choose to *impute*—estimate a substitute value for—missing data to address potential anomalies affecting the quality of dataset coverage [38].

A broad variety of methods exist for data imputation (see Little and Rubin [36] and Lajeunesse [35] for surveys). For example, hot-deck imputation samples substitute values from the current signal while cold-deck imputation estimates values using other sources, such as related datasets [25] or domain heuristics [28]. Interpolation methods use weighted combinations of available data to statistically infer missing values using methods like linear interpolation, regression, and adaptive interpolation [24]. More complex imputation methods can integrate information about the processes used to generate the dataset [43] or use machine learning and related techniques to holistically estimate missing values [2].

While an exhaustive survey of imputation methods is beyond the scope of this paper, understanding the relationship between different imputation choices and perceived data quality is critical for visualizing missing data. As Babad and Hoffer note, even if data values can be inferred with reasonable accuracy, it is important for analysts to understand when and where missing data occurs [7]. Missing data can have a significant impact on inference and decision making and can lend context to analyses. Most significantly for our work, missing data is a key component of *data quality*, a measure of the trust and suitability of data for addressing a given problem [41].

Time series data has specific considerations for data quality (see Gschwandtner et al. for a survey [27]). For example, non-uniform sampling may force interpolations. Joining data across two temporal sources with different granularities can lead to misalignment [6]. Measures taken at the same time may conflict. Since data is typically continuous, violations to trends may be especially salient. We use time series data as it is commonly used in both real-world analysis and empirical studies for visualization and these factors make it an important special case for understanding the implications of missing data for temporal analysis.

Visualizing Incomplete Data

Wong and Varga refer to missing data as *black holes* in a visualization–"a dark area of the cognitive workspace that by the absence of data indicates that one should take care [50]." They argue that it is unclear when and how visualizations should replace missing data to support sensemaking, yet it is clear that people should be able to detect and reason about missing data. Many visualization systems support data quality analysis, including quality change over time [11], data preprocessing [8], and highlighting missing, incorrect, or imputed values [9, 10, 21]. For example, Visplause [5] supports data quality inspection for time series data in order to assist analysts in inferring potential causes of missing data. Wrangler [32] uses statistical methods to help analysts impute missing values. xGobi [45], MANET [49], and VIM [46] offer visualization suites that allow analysts to understand the amount of missing data and compare different imputation methods.

Many visualization systems oriented towards specific domains or datatypes automatically process missing data. Some visualizations provide little to no visual indication of imputed data. For example, Turkay et al. [47] substitute missing values with the feature mean. Systems in meteorology [18] and psychology [28] interpolate missing data based on domain heuristics. Other systems leverage visual saliency to manipulate whether analyst attention is drawn to imputed values. For example, TimeSearcher uses brightly colored marks to indicate locations of missing values [14]. Restorer uses grayscale to reduce the salience of missing spatial data, and luminance to interpolate imputed values [48]. However, the influence of imputation and the corresponding visualization methods used in these systems is not well understood. We ground our exploration of imputation in the practices currently used in missing data visualization.

Graphical Perception

Prior studies in graphical perception show how the methods used to visualize data change our interpretation of that data. For example, studies show that visualization design changes our abilities to estimate and compare statistical values [3, 15, 22, 29] and shift our confidence in those estimations [1]. As imputed values represent uncertain information, we draw on prior findings in uncertainty visualization to inform our study. Specific visual attributes, such as luminance, blurriness, and sketchiness, can indicate uncertainty in data and shift people's confidence in their conclusions [13, 16, 33, 37]. Presenting data as "sketchy" additionally increases engagement with and willingness to critique data [51], which may have interesting ramifications for perceived data quality. Individual values can shift statistical perceptions of data [17], indicating imputed points introducing variation may potentially bias analyses. As many imputation methods provide no quantifiable measure of uncertainty, we evaluate encodings that both present either the level or the existence of uncertain information.

A handful of prior studies have explicitly evaluated the influence of visualization methods on perceptions of data quality. Xie et al. [52] measure how to communicate data quality in high dimensional data using size, brightness, and hue, and found hue and size to be strong channels for encoding quality information. Eaton et al. [20] compared how different methods for visualizing missing data in line graphs influenced accuracy and confidence in response for point-comparison and trend estimation recall tasks. They substituted missing values with zero, rendered no marks for missing data (*data absent*), and used gapped circles to indicate missing data. They found that people interpreted confidently even when critical data was missing, but found no significant differences between methods. Participants expressed an overall preference for visualizations that explicitly indicated missing data. Andreasson and Riveiro [4] conducted a similar study comparing the effects of absent data, fuzziness, and annotated absent data on analyst confidence in a decision making task. Their results showed that people had a strong preference for conditions with annotated absent data and a strong dislike for fuzziness.

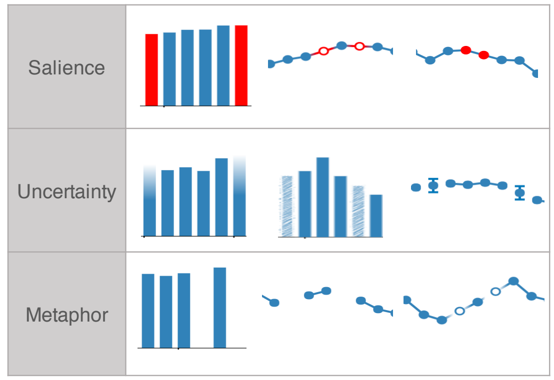


Figure 2. Our studies looked at three distinct categories of visualizations that could encode imputed values: salient encodings that draw attention to imputed values, uncertainty encodings, and encodings that use visual metaphors of absence.

Our work looks to significantly extend these findings by separating effects of imputation methods such as the zero-filling in [20] from visualization methods, considering variable numbers of missing values, and leveraging a wider variety of visualization methods. We also evaluate bar charts in addition to line graphs, where not drawing missing data is indistinguishable from zero values. Prior studies indicate that the kinds of information people synthesize across bars and lines can vary [53], and these differences may significantly impact perceptions of missing data.

CONDITIONS & HYPOTHESES

Data quality concerns how suitable a given dataset is to solve a problem or make a decision. Dimensions of data quality include several factors related to a data source (e.g., accessibility, volume, and relevance) and others that relate to perceptions of the dataset (e.g., completeness, credibility, and reliability) [41]. While analysts must consider factors of a data source when choosing a dataset, the visualizations used to analyze data directly influence perceptions of the data. In this study, we measure how imputation and visualization choices impact response bias and perceptions when data is incomplete. We focus on completeness, credibility, reliability, and confidence in analysis as key indicators of data quality.

We tested three categories of visualization type for communicating missing data that we encountered in the systems discussed in §Related Work. The first category focused on making missing data **salient** by leveraging bright colors to attract participant attention to missing data points (e.g., [10, 14]). The second category emphasized that missing data was **uncertain** by using encodings correlated with uncertainty perception (e.g., [4, 9]). The third category used encodings corresponding to **metaphors** for missing data as some portion of their visual representation is missing (e.g., [4, 20]). As these semantically related to incompleteness, we anticipate that these encodings will also degrade data quality perceptions. Some tested manipulations were hybrids of these categories that examine dependencies across conditions. To mirror prior studies, we included a condition where missing data was entirely absent as a baseline.

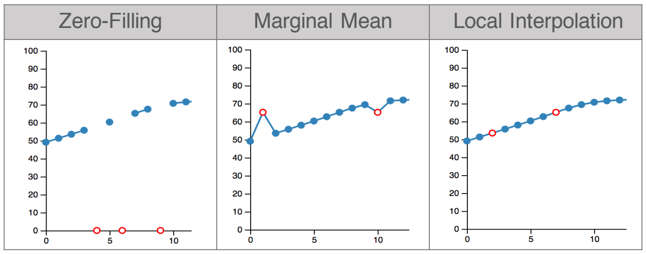


Figure 3. We measured effects of three different imputation methods on data interpretation: zero-filling (substituting missing values with zeros), marginal means (substituting with the mean of the available data), and local interpolation (linear interpolating between adjacent available datapoints).

We draw our tested imputation methods from three methods we observed in existing visualization systems. **Zero-filling** substituted a single value (0) for all missing data points, as in many commercial systems. **Local interpolation** linearly interpolated between adjacent available items, as in [28, 48]. **Marginal means** replaced each missing data value with the mean of all available signals, as in [21, 47]. For our data, zero-filling introduced the highest deviation from the original dataset, marginal means the second, and local interpolation the lowest. While we experimented with more complex interpolation methods, we found no significant differences in our results between those methods and the three selected. Figure 3 provides examples of the tested imputation and visualization categories.

Based on these conditions, we hypothesized that:

H1–Perceived data quality and response accuracy will both degrade as the amount of missing data increased.

H2–Salient visualization methods will generate higher perceived data quality and lower response accuracy than uncertainty or metaphor methods.

H3–Local interpolation will lead to higher perceived confidence, quality and response accuracy than marginal means or zero-filling as takes into account current trends in dataset.

H4–Imputed values will lead to higher perceived data quality than absent values.

H1 stems from the idea that completeness is a key indicator of data quality and provides a quality check for our experiment. **H2** arises from certainty and completeness as aspects of data quality. As salient visualizations provide no visual indications associated with either completeness (as in the metaphor condition) or uncertainty (as in the uncertain condition), we anticipate it will lead to higher perceived quality. This corresponds with observations from Andreassen and Riveiro [4] who found evidence that "fuzzy" visualizations, correlated with uncertainty [13], were not well-liked for decision making with missing data. We predict **H3** on the basis of potential biases introduced by zero-filled and mean values and that local interpolation will create plausible variation in imputed values. This aligns with Correll and Heer's findings that values outside of a distribution can bias statistical perceptions in data [17]. **H4** stems directly from Eaton et al. [20], who showed a preference for visualizations using explicit visual indications of missing data compared to absent values.

METHODS

We ran two 7 (visualization type) $\times$ 3 (imputation method) $\times$ 4 (percentage of missing data) full factorial within-participants studies to measure how visualization and imputation influence time series analysis, focusing on two conventional visualizations: line and bar graphs. Each study followed the same general procedure. Specific differences between the two studies are discussed in their respective sections. For each study, we had three independent variables—visualization type, imputation method, and percentage of missing data—and five dependent variables—accuracy, confidence in response, data credibility, data reliability, and data completeness.

Stimuli & Tasks

We generated each graph as a 1000 $\times$ 400 pixel graph using D3 [12] and Plot.ly [30]. Figure 1 shows examples of these graphs. Each graph visualizes 60 values representing the frequency of Tweets collected per minute over an hour. This context provided a concrete problem scenario where we often find missing data in the real-world due to failures in data collection. We simulated missing data completely at random (MCAR) by randomly removing a subset of values in each graph ($0\%, 10\%, 20\%$, or 30%). We replaced these values with an imputed value computed using one of the three imputation methods described in Conditions & Hypotheses (zero-filling, local interpolation, or marginal mean). The 0% removed condition provided a baseline for measuring changes to our dependent variables due to data removal. The imputed values were then rendered using one of the seven candidate visualization methods per graph type (c.f., Figures 4 and 6).

Above each graph, we provided a brief sentence contextualizing the data, a statement encouraging participants to complete the questions as quickly and accurately as possible, and a counter indicating current progress through the study. Below each graph, we enumerated five questions, answered using radio buttons:

1. Were there more Tweets on average in the first or second half-hour?
2. How confident are you in your response?
   1–Extremely Unconfident, 5–Extremely Confident
3. How credible is this data?
   1–Extremely Uncredible, 5–Extremely Credible
4. How complete is this data?
   1–Extremely Incomplete, 5–Extremely Complete
5. How reliable is this data?
   1–Extremely Unreliable, 5–Extremely Reliable

We chose to use averages as our evaluative task as it forced participants to consider information from all points in the dataset and mitigated changes to the correct response and task difficulty introduced by randomly removing values. Questions 2 through 5 provide different perspectives on the perceived quality of the data and conclusions.

Tested Data

Both noise and task difficulty may influence perceptions in our study: noisier signals may change the effects of different imputation methods and confidence in a conclusion may correlate with difficulty. To control for these concerns, we used synthetic datasets to provide control over noise and difficulty. Each graph contained 60 y-values ranging from $y = 0$ to $y = 100$. X-values corresponded to each value's order within the dataset (i.e., $x = 0$ to $x = 59$). We computed the y-values by first generating a signal from structured random noise [40]. We generated signals using five different noise levels and considered noise as a random effect in our analyses. We then used a constraint-based optimization to adjust the mean difference between the first and last thirty points while minimizing deviation from the original random signal to control difficulty. We separated the means of the first and last half hour by 6.0, randomly selecting which half hour was highest. We used this difference threshold as it achieved desirable response accuracy in prior studies [3].

Each graph visualized a randomly selected dataset from 93 total datasets generated using this method. To assist with reproducing our results, the datasets and experimental infrastructures are available online at `removed.for.review`.

Procedure

Our study consisted of five phases: (1) consent, (2) screening, (3) instructional tutorial, (4) formal study, and (5) demographic questionnaire. Each participant first provided informed consent to participate in the study in accordance with our IRB protocol. We then screened participants for color vision deficiencies using a set of four Ishihara plates. Participants then received instructions about the study and were serially shown examples of each of the seven visualization conditions with one missing value. Each stimuli in the tutorial explained that some data was missing and that we had "guessed" at the values and described how we visualized "guessed" values. They had to correctly identify the half-hour with the highest average value for each condition before beginning the formal study.

The formal study consisted of 87 trials presented serially (84 from our factorial design and 3 engagement checks). To mitigate effects from changing the visualization paradigm, we blocked stimuli by visualization method and randomized the order of blocks. Within each block, participants saw all twelve combinations of missing data (0%, 10%, 20%, and 30%) and imputation method (zero-filling, local interpolation, and marginal mean) presented in random order. Each stimuli visualized a random dataset, with each dataset occurring at most once per participant. We constructed three engagement checks with 0% missing data where the average between halves of the dataset differed by 20. These engagement checks were added to blocks 2, 4, and 6.

After completing the formal study, participants completed a demographic questionnaire, which included an opportunity for open-ended comments, and were compensated for their participation.

Measures & Analysis

We used three primary measures to analyze participant responses: accuracy (Question 1), perceived confidence in their answer (Question 2) and perceived quality (Questions 3-5). Unless otherwise specified, our main analysis used a repeated measures analysis of covariance (ANCOVA) to test for main and interaction effects with question order and noise treated as random effects and the actual and imputed difference between means as covariates. In both experiments, our response data was normally distributed. To control for Type I errors in planned comparisons across conditions, we used Tukey's Honest Significant Difference test with $\alpha = .05$ for post-hoc analyses, and Dunnett's test with $\alpha = .05$ to test against baseline conditions. We elected not to use response time as a measure for this study. While understanding the effects of missing data on analysis speed is an interesting question, the inclusion of our subjective measures and use of crowdsourcing make it less reliable given our experimental apparatus.

Accuracy can be measured as identifying the mean either with or without the imputed values. For our data, these two means differed in only 4.6% of trials. Because imputed values are not actual data and the tutorial explicitly indicated these values are "guesses," we measured accuracy based on the mean of available non-missing datapoints. We did not instruct participants to ignore or consider this data in their judgments in order to ascertain natural intuitions about working with missing data. To capture potential bias due to imputed values and subsampled means, we include both the difference between mean of the available data and differences in means inclusive of the imputed values as covariates in our analyses.

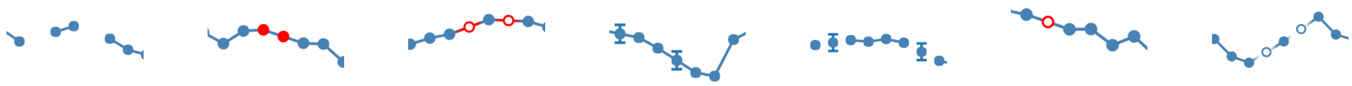
Our four questions relating to perceived data and analysis quality were framed as 5-item Likert scales. We constructed a three-item scale from responses about perceived data reliability, credibility, and completeness to investigate perceptions of data quality. We used this new scale as a dependent measure in-place of the component questions.

Participants

We collected data from 115 U.S. participants on Amazon's Mechanical Turk ($\mu_{age} = 33.9, \sigma_{age} = 10.7$, 54 female, 58 male, 3 DNR). All participants reported normal or corrected to normal vision. To ensure honest participation and task understanding, we excluded any participants who answered more than two of the three engagement check stimuli incorrectly. This resulted in two exclusions (4%) for Experiment One and none in Experiment Two. In each experiment, participants had two hours to complete the study and took an average of 12 minutes.

EXPERIMENT ONE: LINE GRAPHS

Line graphs are among the most common methods for visualizing data. We tested three factors we hypothesized may effect missing data interpretation in line graphs: visualization type, percentage of missing data, and imputation method. Our visualization types consisted of seven different graphs types that manipulated some combination of the point marks and lines themselves. We anticipated that since the connection between values in a line chart is salient, manipulating the lines may have different effects from manipulating the points alone.



(a) Data Absent (b) Color Points (c) Color Points & Line Gradients (d) Connected Error Bars (e) Disconnected Error Bars (f) Unfilled Points (g) Unfilled Points & Line Gradients

Figure 4. We tested seven different visualization methods that manipulated point and line encoding for missing data. These stimuli show examples of all seven conditions at our lowest noise level.

Figure 4 shows the seven tested visualization designs: (a) Data Absent (baseline), (b) Color Points, (c) Color Points with Line Gradients (where the imputed value and its connections are both colored red), (d) Connected Error Bars, (e) Disconnected Error Bars (where the line does not pass through imputed points), (f) Unfilled Points, and (g) Unfilled Points with Line Gradients (where the line passing through imputed data is alpha-blended). Drawing on our three target visualization categories, the color conditions and unfilled points manipulated salience, error bar conditions showed uncertainty, and the unfilled point conditions and color with gradient used metaphor. Two of these designs (Fig. 4c and f) represent hybrids between salience and metaphor encodings. Adding information about variance or range to error bars could add additional relevant statistical information to supplement participants' responses beyond that available in the other methods. To avoid potential confounding effects from this information, error bars simulated a fixed margin of error equal to the difference between means in the original data (6 units).

Line Graph Results

We collected data from 63 U.S. participants on Amazon's Mechanical Turk ($\mu_{age} = 33.9, \sigma_{age} = 10.9$, 27 female, 34 male, 2 DNR). We excluded two participants for missing two or more of the engagement check trials, resulting in 5,124 trials. We looked at the effects of visualization type, imputation method, and percentage of missing data on each dependent measure using a full factorial three-factor (visualization, imputation, missing amount) rmANCOVA, with question order and noise level treated as random factors and differences in present data means and means inclusive of imputed values as covariates. Figure 5 summarizes our results for accuracy and perceived data quality. Complete data is available on the project page (`removed.for.review`).

Overall, participants correctly identified the half-hour with the higher mean in 84.6% of trials. We found a significant main effect of percentage of missing data on accuracy ($F(3,5038) = 5.54, p < .001$). As predicted, accuracy degraded as the amount of missing data increased.

We found a strong correlation between perceived data credibility, completeness, and reliability (Cronbach's $\alpha = 0.90$). We constructed a three-item scale using these factors to describe data quality. We found significant main effects on data quality due to percentage of missing data ($F(3,5038) = 420.98, p < .0001$), visualization method ($F(6,5038) = 22.83, p < .0001$), and imputation method ($F(3,5038) = 80.39, p < .0001$). As predicted in **H1**, as the amount of missing data increased, perceived confidence decreased. We found a marginal effect of imputed means as our covariate ($F(1,5038) = 3.36, p < .07$), which suggests that imputed values may have biased perceived data quality. Connected error bars led to higher perceived data quality than all other visualization types, whereas the data absent condition used in prior studies led to lower perceived quality than all other types ($\mu_{error} = 3.63 \pm 0.08, 95\% CI$ vs. $\mu_{absent} = 2.99 \pm 0.07$). Local interpolation had higher perceived quality than marginal means ($\mu_{local} 3.43 \pm 0.05$ to $\mu_{means} = 3.32 \pm .05$), and both methods had higher perceived quality than zero-filling ($\mu_{zero} = 2.96 \pm .05$).

We also found three significant interaction effects on data quality: visualization type and missing amount ($F(18,5038) = 2.23, p < .003$), visualization type and imputation method ($F(12,5038) = 2.36, p < .006$), and imputation method and missing amount ($F(12,5038) = 10.69, p < .0001$). As the amount of missing data increased, connected error bars preserved higher perceived quality than all other conditions. Connected error bars with local interpolation had the highest perceived quality overall ($\mu_{error,local} = 3.87 \pm .12$), while unfilled points with zero-filling had the lowest ($\mu_{unfill,zero} = 2.78 \pm .13$). Overall, local interpolation led to the highest perceived quality; however, color points with line gradients had significantly lower overall perceived quality for local interpolation than all other visualization methods ($\mu_{color,local} = 3.31 \pm .13$). Connected error bars and zero-filling also had significantly higher perceived quality than all other zero-filling conditions ($\mu_{error,zero} = 3.25 \pm .14$). As expected, the 0% missing data baseline had significantly higher overall perceived quality than any other conditions. Connected error bars with any amount of missing data had higher perceived confidence than all data absent conditions as well as 20% or 30% missing data for all remaining visualization types, including disconnected error bars. Both 20% and 30% zero-filling had significantly lower perceived quality than all combinations of missing amount and imputation method.

We found significant main effects on perceived confidence in response due to percentage of missing data ($F(3,5038) = 71.18, p < .0001$) and imputation method ($F(2,5038) = 18.28, p < .0001$), and a marginal effect of visualization type ($F(6,5038) = 1.88, p < .08$). As expected, confidence decreased as more data was removed. Participants reported significantly higher confidence when using local interpolation than marginal means ($\mu_{local} = 3.60 \pm .05$ to $\mu_{mean} = 3.47 \pm .05$), and lowest confidence in zero-filling ($\mu_{zero} = 3.30 \pm .06$). We also found a significant interaction effect of percentage missing and imputation ($F(6,5038) = 4.55, p < .0001$) and a marginal interaction of visualization and imputation ($F(12,5038) = 1.65, p < .08$), with zero-filled color

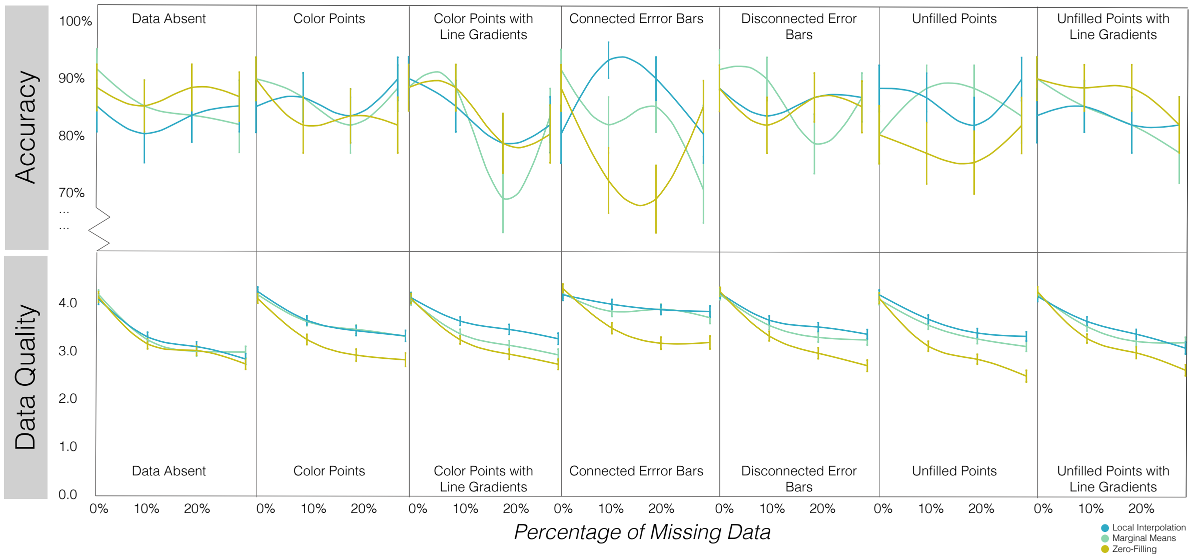


Figure 5. We compared accuracy and data quality across visualization, percentage of missing data, and imputation type for line graphs (blue = local interpolation, green = marginal means, gold = zero-filling). Perceived data quality was highest for local interpolation and connected error bars.

points ($\mu_{pointcolor,zero} = 3.09 \pm .17$) leading to significantly lower confidence overall than all local interpolation conditions and than marginal means with color and line gradients ($\mu_{color,means} = 3.55 \pm .14$).

Line Graphs–Synthesis of Results

Our results support **H1** on all dependent measures–as the percentage of missing data increased, accuracy, perceived quality, and confidence in analysis decreased. We also found support for **H4** as the data absent condition led to significantly lower perceived data quality and confidence than all other visualization conditions. We found partial support for **H3** as local interpolation led to significantly higher confidence and perceived data quality. We found no evidence of accuracy bias from imputation methods, consistent with Eaton et al. [20].

However, our results showed an effect counter to that predicted in **H2**. We anticipated that using uncertain encodings would lead to lower perceived quality while salient imputed values would lead to high perceived quality. However, our results showed the opposite effect: participants regarded color with line gradients as of lower quality, while connected error bars robustly supported the highest overall perceived data quality and confidence even with significant amounts of missing data (Fig. 5). This conclusion ran contrary to Andreasson and Riveiro [4] who found that fuzziness, another correlate for uncertainty, led to the lowest overall preference for decision making tasks. The robustness demonstrated by connected error bars as the amount of missing data increased suggests that connected error bars may preserve perceived quality even as actual quality decreases, which could bias decision making. As this assumption is grounded in descriptive statistics and a lack of an effect, further testing is needed to determine the validity of this observation.

One potential explanation for the high perceived quality with connected error bars is that participants see error bars as providing additional explanatory information to support their analysis even though the error bars encoded no usable statistical information. However, the effect only held for connected error bars, where the line through the points was preserved. To better understand the influence of error bars, we explored a wider range of uncertainty encodings using a different visual channel in Experiment Two.

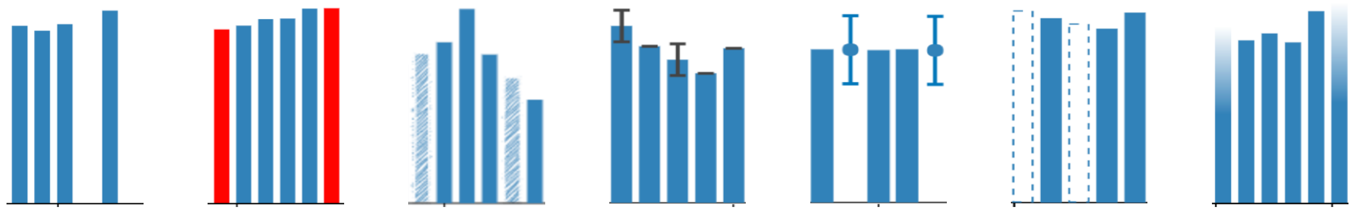
EXPERIMENT TWO: BAR CHARTS

We extended our line graph study to more deeply explore the influence of missing data imputation and visualization. Bar charts provide an interesting case for visualizing missing data as they use bar height to encode data rather than position and connection and shift the data patterns people observe [53]. Many techniques for bar charts also visualize absent data, zero-filling, and $y = 0$ the same way, which may change biases and quality perceptions compared to line graphs.

As in the lines study, we evaluated seven visualization designs. We prioritized encodings correlated with uncertainty to better understand our Experiment One results, including some methods associated with uncertainty that do not explicitly communicate the absolute amount of uncertainty (e.g., sketching and dashes). Figure 6 seven shows the seven tested visualization designs: (a) Data Absent, (b) Color Bars, (c) Sketched Bars, (d) Bars with Error Bars, (e) Points with Error Bars, (f) Unfilled Bars with Dashed Outlines, and (g) Alpha-blended Gradient Bars. Error bars again approximated a 6% margin of error, and gradients used this amount to define the blend radius. Color bars manipulated salience; error bar conditions, sketchiness, gradients, and dashed outlines correlated with uncertainty; and dashed outlines, points with error bars, and gradient bars used metaphor. Three conditions (points with error bars, dashed outline bars, and gradient bars) are hybrids of metaphor and uncertainty.

Bar Graph Results

We collected data from 52 U.S. participants on Mechanical Turk ($\mu_{age} = 33.8$, $\sigma_{age} = 10.5$, 28 female, 24 male). All par-



(a) Data Absent (b) Color Bar (c) Sketched Bar (d) Bars with Error Bars (e) Points with Error Bars (f) Bars with Dashed Outlines (g) Alpha Gradient Bars

Figure 6. We tested seven different visualization methods that manipulated bar encoding for missing data. These stimuli show examples of all seven conditions at our lowest noise level.

ticipants answers at least two of the engagement check stimuli correctly, resulting in 4,368 trials. We looked at the effects of visualization type, imputation method, and amount of missing data on each dependent measure using a full-factorial three-factor (visualization, imputation, and percentage missing) rmANCOVA, with question order and noise level treated as random factors and differences in present data means and means inclusive of imputed data as covariates. Figure 7 summarizes our results.

People correctly identified the higher average half-hour in 88.5% of trials. We found significant main effects on accuracy due to percentage of missing data ($F(3,4282) = 18.76, p < .0001$), visualization type ($F(6,4282) = 3.46, p < .003$), and imputation method ($F(2,4282) = 6.99, p < .001$). As predicted by **H1**, performance decreased as the amount of missing data increased. Sketched bars significantly outperformed data absent ($\mu_{sketch} = 86.5\% \pm 3.1\%$), and both sketched and color bars significantly outperformed points with error bars ($\mu_{color} = 88.6\% \pm 2.9\%$ vs. $\mu_{errorpoints} = 82.5\% \pm 3.4\%$). Local interpolation significantly outperformed zero-filling ($\mu_{local} = 89.7\% \pm 1.8\%$ vs. $\mu_{zero} = 83.8\% \pm 2.1\%$), but we found no significant performance difference between either method and marginal means ($\mu_{means} = 85.7\% \pm 2.1\%$). We found a significant interaction effect of the percentage of missing data and imputation ($F(6,4282) = 2.28, p < .04$). We again found that no missing data had significantly higher accuracy across all imputation methods. Local interpolation provided robust accurate analyses: local interpolation with 30% missing values had significantly higher performance than 10% missing zero-filling or 30% missing marginal means ($\mu_{local,30\%} = 90.9\% \pm 2.9\%$ vs. $\mu_{zero,10\%} = 83.0\% \pm 3.8\%$ and $\mu_{means,30\%} = 82.7\% \pm 3.8\%$).

We found a strong correlation between perceived data credibility, completeness, and reliability ($\alpha = .95$). We constructed a three-factor scale describing data quality using these measures. We found significant main effects on data quality due to the percentage missing ($F(3,4282) = 715.84, p < .0001$), visualization method ($F(6,4282) = 38.00, p < .0001$), and imputation ($F(2,4282) = 50.38, p < .0001$). Perceived quality decreased with the amount of missing data. We found that bars with error bars had significantly higher perceived quality than all other conditions ($\mu_{error} = 3.50 \pm .09$), while gradient, color, and sketched bars had significantly higher perceived quality than the three metaphor conditions: outlined bars, points with error bars, and data absent ($\mu_{gradient} = 3.28 \pm .09$, $\mu_{color} = 3.25 \pm .10$ and $\mu_{sketched} = 3.12 \pm .09\%$ vs. $\mu_{outline} = 2.84 \pm .09$, $\mu_{errorpoints} = 2.68 \pm .09$, and $\mu_{dataabsent} = 2.83 \pm .09$, Fig. 7). A Dunnett's test found that sketched bars, gradient bars, color bars, and bars with error bars all led to greater perceived data quality the baseline data absent condition. Local interpolation and marginal means had significantly higher perceived quality than zero-filling ($\mu_{local} = 3.22 \pm .05$ and $\mu_{means} = 3.18 \pm .05$ vs. $\mu_{zero} = 2.82 \pm .05$).

We also found interaction effects between visualization type and percentage of missing data ($F(18,4282) = 5.00, p < .0001$), visualization type and imputation method ($F(12,4282) = 9.15, p < .0001$), and imputation method and percentage missing ($F(6,4282) = 7.91, p < .0001$). For all visualization types, the baseline condition (no data missing) had significantly higher perceived quality than all other conditions. Overall, bars with error bars had the highest perceived quality at all levels, while points with error bars had the lowest. We also found that perceived quality was significantly higher for bars with error bars at 30% missing data than points with error bars at 10% missing ($\mu_{error,30\%} = 3.23 \pm .18$ vs. $\mu_{errorpoints} = 3.09 \pm .13$). Bars with error bars using interpolation or marginal means had significantly higher perceived quality overall than points with error bars with any interpolation methods.

Response confidence significantly differed for visualization type ($F(6,4282) = 11.35, p < .0001$), percentage missing ($F(3,4282) = 174.5, p < .0001$), and imputation method ($F(2,4282) = 15.97, p < .0001$). A Dunnett's test found people were more confident analyzing sketched bars, gradient bars, color bars, and bars with error bars than the baseline data absent condition. A Tukey's HSD revealed that sketched bars and bars with error bars led to significantly more confidence responses than all three metaphor encodings ($\mu_{sketch} = 3.39 \pm .11$ and $\mu_{error} = 3.37 \pm .11$ vs. $\mu_{outline} = 3.06 \pm .11$, $\mu_{errorpoints} = 2.89 \pm .11$, $\mu_{dataabsent} = 3.01 \pm .11$). Gradient and color bars led to more confident judgments than points with error bars ($\mu_{gradient} = 3.24 \pm .11$ and $\mu_{color} = 3.29 \pm .11$). Participants were more confident with local interpolations than with marginal means ($\mu_{local} = 3.34 \pm .07$ vs. $\mu_{means} = 3.21 \pm .07$) and zero-filling ($\mu_{zero} = 2.99 \pm .07$).

We found significant interaction effects of visualization type and imputation ($F(12,4282) = 4.15, p < .0001$) and percentage missing and imputation ($F(6,4282) = 6.39, p < .0001$), and a marginal effect of visualization type and amount of

Figure 7. Our results for accuracy and data quality in bar charts across visualization type, percentage of missing data, and imputation type (blue corresponds to local interpolation, green to marginal means, and gold to zero-filling). Overall, bars with error bars and sketched bars led to higher performance than metaphor and data absent conditions, even though the uncertain information provided did not provide additional relevant data. We found a similar pattern for perceived data quality, where metaphor encodings performed poorly regardless of imputation.

missing data ($F(18,4282) = 1.47, p < .09$). We found two clusters of perceived confidence for local interpolation: bars with error bars, sketched bars, gradient bars, and colored bars significantly outperformed outlined bars and points with error bars. People were significantly more confident with sketched bars and bars with error bars using either local interpolation or marginal means than they were with interpolated points with error bars ($\mu_{errorpoints,local} = 2.90 \pm .19$).

Bar Graphs–Synthesis of Results

Our results support **H1**: as the amount of missing data increased, accuracy, confidence in result, and perceived data quality decreased. We also found partial found support for **H3**—local interpolation outperformed zero-filling and led to higher quality perceptions and confidence—and for **H4**—data absent was always in the lowest performing encoding groups. We also found that imputing data using marginal means or zero-filling may bias participants' responses when larger amounts of data are missing.

Unlike line charts, we found evidence of interpretation bias for response accuracy across different conditions. People were more likely to estimate values correctly when using sketching and color than with point error bars or absent data. Absent data in bar charts is perceptually indistinguishable from zero-valued data in all but the error bar conditions, which may have partially biased estimates. This hypothesis is further supported as local interpolation significantly outperformed zero-filling even with three times the amount of missing data. However, zero-filled point error bars are visually distinguishable from absent data. This suggests a second hypothesis about the source of this bias: attentional selection. The process used to visually average information may be impacted by the point condition's use of height and position encodings. Based on known mechanisms of visual attention (see Gleicher et al. [26] for a discussion), we assume that people first visually select for the non-missing data and then average over this selected data. The shift for points suggests that inconsistent encoding interferes with people's abilities to do this. Future testing is needed to verify this hypothesis.

We again found evidence inverting **H3**—uncertainty encodings provided the highest perceived confidence and data quality scores. Bars with error bars, despite not supporting the most accurate estimations and providing no useful supplemental statistics, supported the highest perceived quality and confidence. Sketching, another uncertainty encoding, also led to high perceived confidence, despite its similarity to fuzziness [4]. Outlined bars, however, led to lower perceptions of quality, despite semantic correlations with uncertainty [13]. This evidence combined with Experiment One suggests that the blending uncertainty and preserved visual continuity between imputed and available marks lead to higher perceptions of data quality and mitigates potential bias from imputed data. That is, when imputed data is visualized in ways that create a visual discontinuity between sequential marks, perceived quality decreases. This hypothesis is also consistent with **H4**—data absent bars consistently led to low perceived quality.

DISCUSSION

We measured the effect of missing data on interpretation accuracy, response confidence, and data quality in time series data across 14 visualization methods and three imputation types. Our results show:

- Perceived data quality and analysis confidence generally degrades as the amount of missing data increases.
- Visualizations associated with uncertainty tend to support higher perceptions of data quality than salience or metaphor-based encodings.

- Metaphor visualizations can significantly degrade perceptions of data quality, confidence, and even lead to incorrect responses if they break the visual continuity of a visualization.
- Local interpolations reduce bias and lead to higher perceptions of quality and confidence in the analysis.
- Electing to not impute missing values can bias interpretations and leads to low perceptions of quality and confidence.

While avoiding bias is a critical element of effective visualization, we find that the ways systems impute and visualize missing data can manipulate perceived data quality and confidence in results. Whether ideal perceptions of quality are high or low is likely dependent on parameters of the data, problem, and domain (see §Application of Results).

We found preliminary evidence that high confidence and perceived quality depends on mulitple factors. Our studies show that visualizing imputed datapoints using uncertainty encodings that preserve the overall visual paradigm of available data lead to the highest perceived data quality and confidence in result. We see this in connected error bars in line graphs and bars with error bars and sketching in bar graphs. This conclusions runs somewhat contrary to prior work on preference in decision support [4]; however, it is inline with research on uncertainty and trust (see Sacha et al [44] for a discussion). Specifically, visual explanations of errors, such as those indicated by error bars, may increase trust in a system [19]. This trust may manifest in increased confidence and perceived data quality. Examining further design components of uncertainty visualization could offer further insight into this phenomena (e.g., alternative designs [33] and cognitive biases [16]).

Linear interpolation consistently produced the highest confidence and perceived data quality, while zero-filling led to the lowest. These results indicate that sophisticated imputation methods may be worth added complexity to avoid response bias in scenarios requiring high analyst confidence and perceived quality. Default fixed-value filling can bias perceptions of data and lead to incorrect conclusions. We offer first steps towards a systematic understanding of the role of imputation in visualization; however, our studies focused on a simple domain, single statistical task, and relatively smooth signals. Further evaluation is needed to more deeply understand the effects of imputation in visualization.

Limitations & Future Work

Our study significantly extends prior work on visualizing missing data, exploring new visualization methods, effects of imputation, and using different chart types. However, we made several simplifying assumptions in our experiment. For example, our data generation tended to favor datasets that seldom dipped near zero in order to minimize deviation from the original structured noise. This may have helped prime people to discern between zero-filled values as missing data rather than $y = 0$ values. Additionally, the smooth characteristics of our signals meant there were few outliers in the datasets. Our narrative scenario used a familiar but simple and low-risk task. While these choices allowed us strong control over our tested conditions to encourage general understanding, future testing should extend our work to real-world datasets and scenarios to better understand the impact of these choices.

Further, we tested a small set of possible imputation and visualization methods, drawing inspiration from visualization tools that actively manage missing data. However, we found few tools explicitly discuss missing data management. Future work should extend to a broader set of visualization and imputation methods, such as multiple imputation and machine learning-based approaches (e.g., expectation maximization) to understand their broader utility for data in different domains.

Application of Results

We anticipate the desired level of perceived quality will vary based on data and domain contexts. While our results enumerate how design choices for visualizing incomplete datasets might modulate perceived quality, a number of factors may inform the desirable level of perceived quality, including:

Decision Risk: Accuracy is often paramount in high risk situations. Missing values may jeopardize that accuracy. Visualization systems can encourage caution in interpreting flawed datasets by using representations that avoid bias and appropriately decrease perceived data quality and decision confidence.

Data Fusion: Combining data from multiple sources may mean that the overall anticipated data quality varies between those sources. In these scenarios, systems can leverage domain knowledge to guide analysts to visualizations that modulate confidence and apparent quality appropriately for each source.

Confidence in Imputation: Individual imputation methods may differ in how faithfully they represent the original data. In scenarios like cold-deck imputation, analysts may know how well the imputed data mirrors the original data. Visualizations can leverage this knowledge to choose methods that adapt perceived quality proportionally to the imputation quality.

Intent to Persuade: Persuasive visualizations typically require the audience to trust in the argument illustrated by the data. Such visualizations should optimize perceived data quality and interpretation confidence to achieve this goal.

CONCLUSION

Data is often incomplete due to factors like deletions, misalignments, or collection failures. Missing data elements can complicate data interpretation and comparisons. We used time series data with missing values to measure how missing data influences data interpretation, their confidence in that interpretation, and the perceived quality of the data. We found that the design choices and interpolation methods used to represent data significantly influence analysts' perceptions of data. Our results enumerate design trade-offs for designers to consider when crafting behaviors for handling missing data in visualization tools. We offer systematic insight that will enable designers to make informed decisions about how to represent and interpolate the missing data points tailored to the demands and domains of their stakeholders.

ACKNOWLEDGMENTS

Omitted for blind review.

REFERENCES

1. Muhammad Adnan, Mike Just, and Lynne Baillie. 2016. Investigating time series visualisations to improve the user experience. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5444–5455.
2. Shilpi Ahuja, Mary Roth, Rashmi Gangadharaiah, Peter Schwarz, and Rafael Bastidas. 2016. Using Machine Learning to Accelerate Data Wrangling. In *Data Mining Workshops (ICDMW), 2016 IEEE 16th International Conference on*. IEEE, 343–349.
3. Danielle Albers, Michael Correll, and Michael Gleicher. 2014. Task-driven evaluation of aggregation in time series visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 551–560.
4. Rebecca Andreasson and Maria Riveiro. 2014. Effects of visualizing missing data: an empirical evaluation. In *Information Visualisation (IV), 2014 18th International Conference on*. IEEE, 132–138.
5. Clemens Arbesser, Florian Spechtenhauser, Thomas Mühlbacher, and Harald Piringer. 2017. Visplause: Visual Data Quality Assessment of Many Time Series Using Plausibility Checks. *IEEE transactions on visualization and computer graphics* 23, 1 (2017), 641–650.
6. Aleks Aris, Ben Shneiderman, Catherine Plaisant, Galit Shmueli, and Wolfgang Jank. 2005. Representing unevenly-spaced time series data for visualization and interactive exploration. *Lecture notes in computer science* 3585 (2005), 835.
7. Yair M Babad and Jeffrey A Hoffer. 1984. Even no data has a value. *Commun. ACM* 27, 8 (1984), 748–756.
8. Jürgen Bernard, Tobias Ruppert, Oliver Goroll, Thorsten May, and Jörn Kohlhammer. 2012. Visual-interactive preprocessing of time series data. In *Proceedings of SIGRAD 2012; Interactive Visual Analysis of Data; November 29-30; 2012; Växjö; Sweden*. Linköping University Electronic Press, 39–48.
9. Markus Bögl, Peter Filzmoser, Theresia Gschwandtner, Silvia Miksch, Wolfgang Aigner, Alexander Rind, and Tim Lammarsch. 2015. Visually and statistically guided imputation of missing values in univariate seasonal time series. In *Visual Analytics Science and Technology (VAST), 2015 IEEE Conference on*. IEEE, 189–190.
10. Christian Bors, Markus Bögl, Theresia Gschwandtner, and Silvia Miksch. 2017. Visual support for rastering of unequally spaced time series. In *Proceedings of the 10th International Symposium on Visual Information Communication and Interaction*. ACM, 53–57.
11. C Bors, T Gschwandtner, and S Miksch. 2014. QualityFlow: Provenance Generation from Data Quality. In *Poster Proceedings of the EuroGraphics Conference on Visualization*.
12. Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D3: Data-Driven Documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)* (2011). `http://vis.stanford.edu/papers/d3`
13. Nadia Boukhelifa, Anastasia Bezerianos, Tobias Isenberg, and Jean-Daniel Fekete. 2012. Evaluating sketchiness as a visual variable for the depiction of qualitative uncertainty. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2769–2778.
14. Paolo Buono, Aleks Aris, Catherine Plaisant, Amir Khella, Ben Shneiderman, H Hochheiser, and B Schneiderman. 2005. Interactive pattern search in time series. In *Proc. SPIE*, Vol. 5669. 175–186.
15. Michael Correll, Danielle Albers, Steven Franconeri, and Michael Gleicher. 2012. Comparing averages in time series data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1095–1104.
16. Michael Correll and Michael Gleicher. 2014. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 2142–2151.
17. Michael Correll and Jeffrey Heer. 2017. Regression by Eye: Estimating Trends in Bivariate Visualizations. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 1387–1396.
18. Suzana Djurcilov and Alex Pang. 1999. Visualizing gridded datasets with large number of missing values (case study). In *Proceedings of the conference on Visualization'99: celebrating ten years*. IEEE Computer Society Press, 405–408.
19. Mary T Dzindolet, Scott A Peterson, Regina A Pomranky, Linda G Pierce, and Hall P Beck. 2003. The role of trust in automation reliance. *International Journal of Human-Computer Studies* 58, 6 (2003), 697–718.
20. Cyntrica Eaton, Catherine Plaisant, and Terence Drizd. 2005. Visualizing missing data: Graph interpretation user study. *Human-Computer Interaction-INTERACT 2005* (2005), 861–872.
21. Sara Johansson Fernstad and Robert C Glen. 2014. Visual analysis of missing dataâĂŤTo see what isn't there. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on*. IEEE, 249–250.
22. Johannes Fuchs, Fabian Fischer, Florian Mansmann, Enrico Bertini, and Petra Isenberg. 2013. Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3237–3246.
23. Tim Furche, Georg Gottlob, Leonid Libkin, Giorgio Orsi, and Norman W Paton. 2016. Data Wrangling for Big Data: Challenges and Opportunities.. In *EDBT*. 473–478.
24. Jun Gao. 2006. Adaptive Interpolation Algorithms for Temporal-Oriented Datasets. In *Temporal Representation and Reasoning, 2006. TIME 2006. Thirteenth International Symposium on*. IEEE, 145–151.

25. William Githungo, Silvery Otengi, Jacob Wakhungu, and Edward Masibayi. 2016. Infilling Monthly Rain Gauge Data Gaps with Satellite Estimates for ASAL of Kenya. *Hydrology* 3, 4 (2016), 40.

26. Michael Gleicher, Michael Correll, Christine Nothelfer, and Steven Franconeri. 2013. Perception of average value in multiclass scatterplots. *IEEE transactions on visualization and computer graphics* 19, 12 (2013), 2316–2325.

27. Theresia Gschwandtner, Johannes Gärtner, Wolfgang Aigner, and Silvia Miksch. 2012. A taxonomy of dirty time-oriented data. *Multidisciplinary Research and Practice for Information Systems* (2012), 58–72.

28. Kerem Gülensoy, Caterina Gawrilow, and Tatiana von Landesberger. 2014. Visual exploration of dirty activity sensor and emotional state data from psychological experiments. In *Proceedings of the 14th International Conference on Knowledge Technologies and Data-driven Business*. ACM, 19.

29. 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*. ACM, 1303–1312.

30. Plotly Technologies Inc. 2015. Collaborative data science. (2015). `https://plot.ly`

31. Sean Kandel, Jeffrey Heer, Catherine Plaisant, Jessie Kennedy, Frank van Ham, Nathalie Henry Riche, Chris Weaver, Bongshin Lee, Dominique Brodbeck, and Paolo Buono. 2011a. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization* 10, 4 (2011), 271–288.

32. Sean Kandel, Andreas Paepcke, Joseph Hellerstein, and Jeffrey Heer. 2011b. Wrangler: Interactive visual specification of data transformation scripts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3363–3372.

33. Matthew Kay, Tara Kola, Jessica R Hullman, and Sean A Munson. 2016. When (ish) is my bus?: User-centered visualizations of uncertainty in everyday, mobile predictive systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5092–5103.

34. Won Kim, Byoung-Ju Choi, Eui-Kyeong Hong, Soo-Kyung Kim, and Doheon Lee. 2003. A taxonomy of dirty data. *Data mining and knowledge discovery* 7, 1 (2003), 81–99.

35. Marc J Lajeunesse, J Koricheva, J Gurevitch, and K Mengersen. 2013. Recovering missing or partial data from studies: a survey of conversions and imputations for meta-analysis. *Handbook of Meta-analysis in Ecology and Evolution* (2013), 195–206.

36. Roderick JA Little and Donald B Rubin. 2014. *Statistical analysis with missing data*. John Wiley & Sons.

37. Alan M MacEachren, Robert E Roth, James O'Brien, Bonan Li, Derek Swingley, and Mark Gahegan. 2012. Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2496–2505.

38. Heiko Müller and Johann-Christph Freytag. 2005. *Problems, methods, and challenges in comprehensive data cleansing*. Professoren des Inst. Für Informatik.

39. Michael J Paul and Mark Dredze. 2011. You are what you Tweet: Analyzing Twitter for public health. *Icwsm* 20 (2011), 265–272.

40. Ken Perlin. 2002. Improving noise. In *ACM Transactions on Graphics (TOG)*, Vol. 21. ACM, 681–682.

41. Leo L Pipino, Yang W Lee, and Richard Y Wang. 2002. Data quality assessment. *Commun. ACM* 45, 4 (2002), 211–218.

42. Line Pouchard. 2016. Revisiting the data lifecycle with big data curation. *International Journal of Digital Curation* 10, 2 (2016), 176–192.

43. Susanne Rässler. 2016. Data fusion: identification problems, validity, and multiple imputation. *Austrian Journal of Statistics* 33, 1&2 (2016), 153–171.

44. Dominik Sacha, Hansi Senaratne, Bum Chul Kwon, Geoffrey Ellis, and Daniel A Keim. 2016. The role of uncertainty, awareness, and trust in visual analytics. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 240–249.

45. Deborah F Swayne and Andreas Buja. 1998. Missing data in interactive high-dimensional data visualization. *Computational Statistics* 13, 1 (1998), 15–26.

46. Matthias Templ, Andreas Alfons, and Peter Filzmoser. 2012. Exploring incomplete data using visualization techniques. *Advances in Data Analysis and Classification* 6, 1 (2012), 29–47.

47. Cagatay Turkay, Arvid Lundervold, Astri Johansen Lundervold, and Helwig Hauser. 2012. Representative factor generation for the interactive visual analysis of high-dimensional data. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2621–2630.

48. Ray Twiddy, John Cavallo, and Shahram M Shiri. 1994. Restorer: A visualization technique for handling missing data. In *Proceedings of the conference on Visualization'94*. IEEE Computer Society Press, 212–216.

49. Antony Unwin, George Hawkins, Heike Hofmann, and Bernd Siegl. 1996. Interactive graphics for data sets with missing valuesâĂŤMANET. *Journal of Computational and Graphical Statistics* 5, 2 (1996), 113–122.

50. BL William Wong and Margaret Varga. 2012. Black holes, keyholes and brown worms: Challenges in sense making. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 56. SAGE Publications Sage CA: Los Angeles, CA, 287–291.
51. Jo Wood, Petra Isenberg, Tobias Isenberg, Jason Dykes, Nadia Boukhelifa, and Aidan Slingsby. 2012. Sketchy rendering for information visualization. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2749–2758.
52. Zaixian Xie, Shiping Huang, Matthew O Ward, and Elke A Rundensteiner. 2006. Exploratory visualization of multivariate data with variable quality. In *Visual Analytics Science And Technology, 2006 IEEE Symposium On*. IEEE, 183–190.
53. Jeff Zacks and Barbara Tversky. 1999. Bars and lines: A study of graphic communication. *Memory and Cognition* 27 (1999), 1073–1079.