# Interacting with Sensor Data to Understand Uncertainty in Air Quality Monitoring

Emily Pondaven[1], Zhiyu Liu[2], Eri Saikawa[3], Jason Jacques[1], and Kristin Williams[2,3]
ep247@st-andrews.ac.uk, sam.liu@emory.edu, eri.saikawa@emory.edu, jtj2@st-andrews.ac.uk, kristin.williams@emory.edu

[1]*School of Computer Science*
*University of St Andrews*
St Andrews, UK

[2]*Department of Computer Science*
*Emory University*
Atlanta, USA

[3]*Department of Environmental Sciences*
*Emory University*
Atlanta, USA

***Abstract*—Current air quality visualizations predominantly rely on map-based representations and Air Quality Index (AQI) color coding to show outdoor air quality data. As outdoor air quality concerns grow in importance—due to rising health risks—the need for greater transparency in scientific communication becomes increasingly critical. Showing uncertainty improves transparency, and allows for greater public understanding and informed decision-making. This paper compares sensor uncertainty visualizations along dimensions of context and interactivity. Map-based, time-based, and sensor-only are each presented in static, interactive, and animated formats. These include several novel visualizations such as a map-based hypothetical outcome contour plot. In a crowdsourced user study, we assess the clarity and usability of these nine visualizations, focusing on how well they communicate both the air quality data and associated uncertainties. We find that there are trade-offs for data interaction design: supporting accurate air quality judgments versus understanding data uncertainty. Participants were signficiantly more accurate AQI judgements when using static visualizations, but they had a significantly better understanding of uncertainty when using interactive visualizations. Our work informs visual interaction design to support interpretability, communicate uncertainty, and contextualize data to help with more confident judgments.**

***Index Terms*—Scientific communication, Interactive visualization, Sensor uncertainty, Air quality monitoring, Crowdsourcing.**

## I. Introduction

Environmental Scientists frequently use the Air Quality Index (AQI) in combination with map-based visualizations to communicate to the public the hazard levels posed by air pollutants [71]. Current visualization techniques follow national and international guidelines on what constitutes safe/acceptable levels of air pollutants. For example, the US Environmental Protection Agency (EPA) established an AQI that adopts a color-coded system to classify the hazard level posed by air pollutants such as particulate matter [71]. In addition to encoding hazard levels of air quality with color, local area conditions are visualized using national and global maps so that households can assess conditions in their local area to inform their daily activities like outdoor exercise.

Communicating uncertainty to the public is critical for establishing scientific transparency in visualized data and to ensure continued public trust in science [38]. We use the term *uncertainty* specifically to refer to the possibility that the data could be a range of values rather than a single value in line with prior work [31]. However, many common air quality visualizations do not encode or depict uncertainty information. One explanation may be that AQI visualization authors believe that 1) AQI visualizations should communicate a clear message drawn from an independent process that establishes expert confidence that the data supports the message, and 2) concerns that uncertainty information could obfuscate that message [31]. Researchers have previously identified sources of uncertainty with AQI data arising from sensor placement and the algorithms used to convert sensor readings to a spatial representation that can be rendered with map-based visualizations [58]. They, and others (e.g. [34], [36], [37], [58]), advocate for representing uncertainty information to better inform decision-making.

Air quality data has particular difficulty with uncertainty in the sensor reading itself. Over time, sensor readings can drift due to build up of contaminants, environmental changes, mishandling, etc. [4]. Previous research on geospatial visualization of air quality data advocates quantifying the uncertainty of sensor measurements and placement so that it can be communicated using visualization techniques [58].

Instead, we argue that interaction design can help with air quality sensing challenges by using similar methods for dealing with sensor uncertainty as part of user interaction. Like ambiguous sensing challenges for natural user interfaces for input and context-aware services [15], [43], air quality sensing is error-prone which can confuse decision-makers and cause performance problems if it is not adequately communicated. In addition to communicating uncertainty information through visualization design, we propose using interaction techniques to facilitate decision-makers' sense-making and reasoning under uncertainty. These techniques can support identifying and resolving ambiguity by involving the user or accumulating more information about sensed events [43]. Our approach goes beyond research on visualizing uncertainty in air quality data to engage with related work on interaction techniques for coping with uncertainty (see [47] for a discussion of the distinction between visualizations and interaction techniques).

This work was supported by the Halle Institute for Global Research at Emory University and the University of St Andrews

979-8-3315-8717-8/25/$31.00 ©2025 IEEE
DOI 10.1109/ICT4S68164.2025.00029

We use these strategies to supplement research on visualizing uncertainty to support communicating air quality data.

In this paper, we create a set of air quality visualizations and investigate their effectiveness for supporting 150 users with understanding data uncertainty in a crowdsourced study. We employ techniques from research on sensor uncertainty to vary our visualizations in terms of 3 levels of interactivity—static, animated, and interactive—and context—location, time, and sensor. In a crowdsourced study with 150 participants, we compare our visualizations in terms of supporting accurate interpretation of the AQI, comparison between two sensor streams with differing uncertainty, and confidence in air quality judgments. We find that there is a trade-off between supporting accurate AQI judgements and supporting understanding of data uncertainty. Our work contributes 1) a set of visualization techniques for communicating air quality using the US AQI, and 2) findings on the role of air quality visualization design in supporting accurate, informed, and trust-worthy scientific communication for better decision-making.

## II. Background

Below, we review related work on visualizing uncertainty in air quality data, and interaction techniques for coping with uncertainty in interaction design.

### A. Revealing Uncertainty

Uncertainty is present in most data [52]. It can enter scientific data at many processing stages, such as through measurement, analysis, and forecasting [52]. Data can also be fundamentally ambiguous when more context is needed to make sense of it [43]. One way to contextualize data is through visualization. Yet, data visualizations usually do not show uncertainty [31]. Drawing on [31], we use the following working definition of uncertainty: *the possibility that the observed value could take on a possible set of values*. Thus, visualizing uncertainty should convey the probability of a range of values instead of showing just one.

Air quality sensors themselves introduce uncertainty into data. So, we expand our working definition of uncertainty with [16]'s definitions of *sensor-wide uncertainty*. This is comprised of (1) instrument uncertainty due to the differences in monitor models and (2) sampling uncertainty due to the measurement methodologies [16]. These definitions align with the US EPA's Air Sensor Guidance [16]. In processing this data, its collection and cleaning impacts uncertainty: calibration, temporal averaging, and reference site correction are important to effectively reduce sensor-wide uncertainty [16]

Handling uncertainty in user interface design is challenging. In data visualization, authors frequently do not convey uncertainty because of 1) beliefs that it can be misleading or incomprehensible, 2) concern that non-experts cannot fully comprehend it, or 3) fear that it makes visualizations less credible [31]. Visualization practitioners describe multiple types of uncertainty associated with a single dataset, and these consists in different levels or layers of uncertainty [65]. When visualizing uncertainty, practitioners described using error bars, circles on a map, and colored branches for trees, but many expressed their frustration with finding helpful representations to communicate with [65]. Since uncertainty can occur at multiple levels of data and may even be inherently ambiguous, there is a need to cope with it in levels of system design [43]. Early work seeking to resolve uncertainty in sensor data streams found a role for interaction techniques in supporting appropriate resolution of data ambiguity [30], [43]. Later work supplemented interaction techniques with context, such as location and time [15]. More recent advances show how interaction techniques benefit from reasoning with and managing the uncertainty of multiple possible interaction states [5], [64], [72]. Here, we draw on these parallel bodies of work, in visualization and interaction techniques, to support interacting with sensor data to understand uncertainty in air quality monitoring.

Communicating uncertainty can improve user's confidence levels and decision-making. It can help with predicting true values, like bus arrival time [36]. In doing so, it helps with planning and making informed decisions, such as deciding when to walk to the bus station or estimating the risk of missing the next bus [36]. Interfaces that use fluctuating data, such as those that have a range of possible values, benefit from an "always-on" reading that encodes the uncertainty information in what is displayed [37]. Showing data fluctuate improves confidence in the measurement instrument such as showing body weight fluctuation trends on a bathroom scale [37]. Notably, users make more informed decisions with uncertainty information if they employ a decision aid [34]. Thus, being transparent about uncertainty information and helping with its interpretation is crucial to that data being integrated into day to day planning and activity.

### B. Visualizing Uncertainty in Air Quality Data

Visualizations present air quality data by making its inherent complexity more accessible to the public and decision-makers. They effectively convey a large amount of information in a simpler form by leveraging visual perception to comprehend multi-dimensional information [57]. Potter et al. put forward a taxonomy that classifies uncertainty into multiple dimensions: 1D, 2D, 3D, and ND data visualization such as error bars, contour maps, and time-varying data, respectively [57]. Uncertainty visualization can have high dimensionality and be challenging to show with large dimensional data such as scientific visualization, geographical visualization, and geographic information science [65]. In geographical fields, data is often shown in 2D space when encoded on maps [57].

Maps are widely used to present air quality data. Platforms like AirNow[1] employ color-coded contours to represent estimated air quality levels. These geographic representations help with quickly identifying areas with better or worse air quality based on color. Widely used forms include heat maps, contour maps, and chloropeth maps [28]. However, alternative forms can be necessary depending on the context. Temporal air

[1] https://www.airnow.gov

quality data is often represented as a time series [29]. Radial charts have been introduced to show time using the plot's angle, although, uncertainty is rarely studied using these [29]. Radial plots are becoming more popular due to their ability to highlight cyclical trends like daily or seasonal patterns. Their circular layout can make periodic variations more intuitive and visually engaging, though they pose challenges when making precise comparisons or understanding their labeling [8]. Related, angular encodings techniques show data here and now—often termed *"nowcasting"w*ith a gauge-style widget showing highly location- and time-specific air quality information. This suggests that the design space of conveying context-specific air quality information to the public needs explored.

One source of uncertainty in geospatial visualization of air quality data is spatial sparsity. Data is collected by an air monitor network that, due to limited resources and geographical constraints, is impractical to install densely to create a large network. For example, California's network has slightly more than 250 stations [10]. While relatively dense, it doesn't cover the entire state [73]. San Jose, a city with $>1$ million residents, has only three monitors installed for an area of nearly 470 km$^2$ [10]. To cope with sparsity, data is estimated using spatial interpolation. Four common spatial interpolation methods—spatial averaging, nearest neighbor, inverse distance weighting, and Kriging—performed similarly across US states and only differed in California, because of the relatively high-density network of air monitors [73]. Using Kriging, researchers found that users demonstrated a "higher perception of risk" and had a tendency to make more informed decisions when uncertainty information was given [58].

### C. Encoding Uncertainty

Effectively communicating uncertainty has become more important as real-time data enable opportunities for real-time sensemaking and responsive coordination like that needed during the recent global pandemic (*c.f.*, [60]). While error bars are widely traditionally used in scientific visualizations, studies suggest that they can be misleading [12]. They are often misinterpreted as hard boundaries rather than probabilistic distributions [12]. Instead, gradient plots, violin plots, and box plots may be more effective [12]. Moreover, error bars are inappropriate for some visualizations and data. In these cases, encodings like color, texture, transparency, and glyphs offer alternatives. Colors and textures were found intuitive encodings [51]. Studies have also found that scaling the size of symbols at each data point is an acceptable approach for symbolic encodings [42], [62]. For spatial data, contour maps often use color-coded contours or thickness variations, while heatmaps can use whitespace or noise effects [6]. Animation techniques, mapped as speed or motion blurring, have also been used to represent uncertainty [53].

One interesting approach uses animation to encode uncertainty in hypothetical outcome plots (HOPs). Each animation frame represents a sample from a distribution estimated from the data [32]. This approach, displaying multiple possible realizations of data over time, forces viewers to recognize variability and uncertainty as part of their interpretation [32]. Instead of requiring users to exert cognitive effort to understand graphically encoded statistical concepts, users can directly extract frequency information with HOPs by leveraging their visual processing system [32]. In turn, this encourages probabilistic reasoning by prompting the user to consider the encoded variability of the presented data. HOPs have been found to outperform several static representations, including violin plots [32]. HOPs offer a promising avenue for visualization of uncertainty, but past literature has predominantly focused on simple line graphs.

## III. Communicating Air Quality

We created visualizations that depict air quality data uncertainty by 1) varying the level of interactivity support, and 2) varying the visual context used to show the data. Specifically, visualizations supported three levels of interactivity—static, animated, and interactive—and three visual contexts—spatial, temporal, and sensor-only—to yield nine distinct visualizations. This allowed us to explore how uncertainty in air quality data might be visually depicted using alternative spatio-temporal assumptions and revealed through user interaction. Table I summarizes how features of each visualization vary to show uncertainty for each interactivity level and context.

### A. Encoding Air Quality with Color Gradients

For each visualization, the main encoding of air quality is color. We convert $PM_{2.5}$ concentrations to the AQI and its associated color using the EPA's formula [71]. In all visualizations, this mapping is shown using a legend and a scale (see Figure 1). Depicting air quality data with color is more effective for clearly communicating the health risks posed by measured concentrations of pollutants instead of simply displaying their numerical values [2], [75]. While the US AQI employs a discrete color encoding scheme, we use a gradient of colors to show how a value may lie on a fuzzy boundary as part of our investigation into depicting uncertainty. Value-Suppressing Uncertainty Palettes—which intuitively encodes both the data and uncertainty within a single visual channel, i.e. color [13]—were considered but ultimately not used. This approach could obscure the air quality values due to its five-color range, making interpretation more difficult. Instead, uncertainty was represented using other encodings while color was reserved for air quality.

### B. Revealing Uncertainty with Interactivity

Visualizations varied in their support for user interaction by ranging from static, through animated, to fully interactive. We consider static visualizations as the most passive form as using them largely consists in perceiving their visual features. Next, we consider animated visualizations to offer an additional level of user engagement. They add an additional dimension of time to allow the user to passively explore the data. Finally, we created a set of user-controllable interactive visualizations that supported using input devices such as a mouse to control a slider, hover-based pop-ups, and menus. This gives the user agency over how the data is explored over time.

| Context | Interactivity Level: Static | Interactivity Level: Animated | Interactivity Level: Interactive |
|---|---|---|---|
| Spatial | Size and transparency of circle glyph | Animated HOP of map contours and color interpolation; Play and pause; Mouseover for interpolated value at that location | Slider changes displayed contour line and color interpolation; Mouseover for interpolated value at that location |
| Temporal | Polar bar height and threshold | Animated HOP of colors for seasonal bin; Play and pause; Mouseover for averaged value for the seasonal bin; Dropdown menu changes seasonal bins | Slider changes displayed color for seasonal bin; Mouseover for averaged value for the seasonal bin; Dropdown menu changes seasonal bins |
| Sensor-only | Width of needle | Animated HOP with two previous needle positions shown with reduced opacity; Play and pause | Slider changes the needle's angular position |

Uncertainty Representation

TABLE I

SUMMARY OF FEATURES USED TO ENCODE AND ENABLE EXPLORATION OF UNCERTAINTY (NOTE: COLOR IS BASED ON THE AQI COLOR ENCODING)

*1) Static Representations:* Static visualizations incorporate visual elements, such as color and glyphs, to show air quality uncertainty depending on the plot type, but they do not support user interaction with the visualization of air quality data.

*2) Animated Representations:* Animated visualizations automatically cycle through a range of AQI scenarios over time. In this approach, we follow [32] and generate visualization frames by repeatedly sampling from a normal distribution of AQI values. We randomly sample from a normal distribution to present the full range of possible AQI values given the measured value. This approach effectively creates the frames for an animated HOP. Each frame shows a possible AQI scenario drawn from the underlying distribution. We animate these frames to show the range of possible scenarios the empirical data could indicate to ensure users actively consider these (see [35] for example). While past literature focuses on simple line HOPs [32], [35], we develop novel visualizations for HOPs using contour lines and AQI color-changing HOPs to support >1 dimensional encoding. Integrating uncertainty directly into the plot in this way encourages users to actively consider it [35]. To guard against users mistakenly thinking the animation shows a time series (*c.f.*, [54]), the sensor-based visualization's past two frames remain in a lower opacity on the plot (see Figure 1C). Following [32], all animations have a frame speed of 400ms and present 100 different sampled frames [32]. Each frame in the animation is displayed independently, without smooth transitions between frames. This is commonly done in HOPs, as by [32], to emphasize the idea that every frame represents an independent possible AQI scenario (see Figures 1A and 1B).

*3) Interactive Representations:* Interactive visualizations support mouse based interaction to change the displayed features. This was primarily accomplished using mouse hover and a slider. When the mouse was hovered over a colored area of the visualization, a tool-tip showed the numerical value $PM_{2.5}$ concentration that was used to determine the location's AQI value, and so, its color encoding. Figure 1B shows an example of this mouse hover interaction. A slider offers the user additional agency over the displayed information and supports interacting with the range of hypothetical plots possible for the measured data value. The slider's control of scenarios—ranging from best to worst case—supports interacting with the range of possible AQI values underpinning the single air quality value typically presented. Our work differs from approaches using side-by-side optimistic and pessimistic forecasts [63], by using a slider to show every $5^{th}$ percentile (from the $5^{th}$ to $95^{th}$ percentiles) of a normal distribution centered around the value used in the static plots. Our assumption of a normal distribution was important from an interaction design perspective, as the jump between extreme changes—in contour lines on the spatial plot, in color on the temporal plots, or needle movement in the sensor-only plot—proved confusing. We assume averages come from a normal distribution since evaluating the most effective distribution for representing the empirical data would take us far from our research questions. Specifically, transitioning

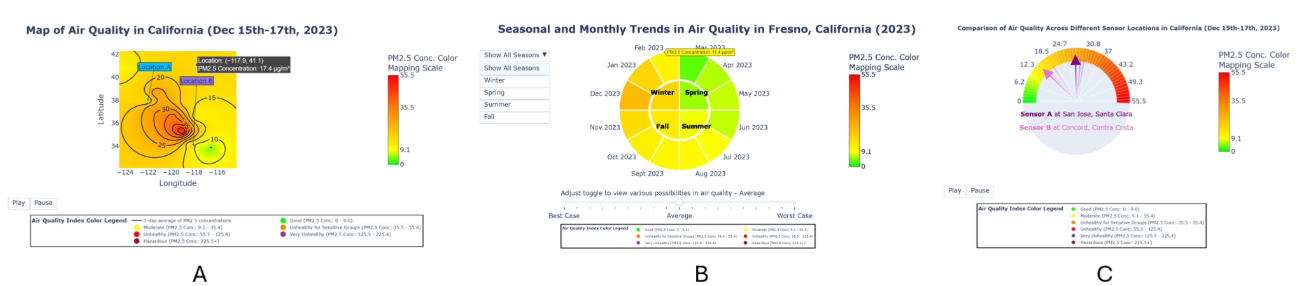


Fig. 1. Examples of non-static uncertainty visualization interfaces created for user study, including the animated map-based (A), interactive time-based (B), and animated sensor-only (C) representations. The text over screenshots A and B show the hover functionality. All visualizations show the PM2.5 concentration color mapping scale and the air quality color legend

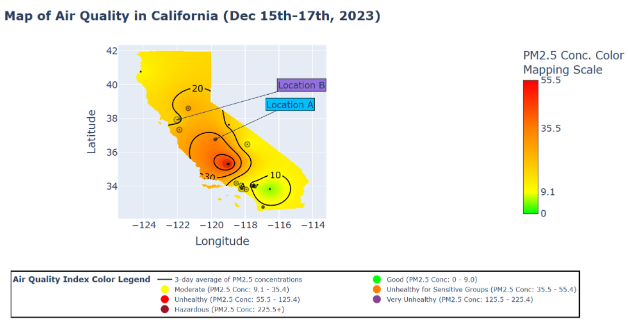


Fig. 2. Map-based static uncertainty visualization interface containing the PM2.5 concentration color mapping scale and the air quality color legend. Note that text has been enlarged for improved clarity.

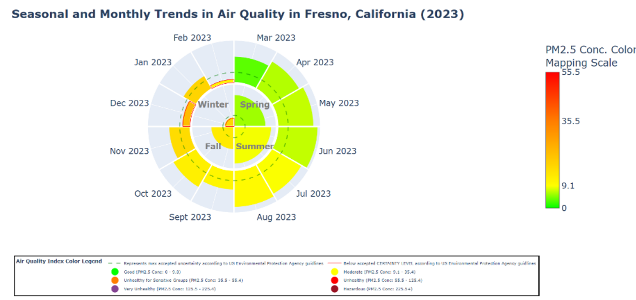


Fig. 3. Time-based static uncertainty visualization interface containing the PM2.5 concentration color mapping scale and the air quality color legend. Note that text has been enlarged for improved clarity.

between larger steps in the uncertainty distribution would likely require more time to make sense of and would merit its own study. Instead, showing every $5^{th}$ percentile created a smoother transition between frames.

### C. Showing Context to Inform Interpretation of Uncertainty

We use visualization to show data in context to support making sense of sensor-data uncertainty. We vary context according to spatial, temporal, and sensor assumptions. To make these assumptions apparent, we use geo-spatial maps, radial plots, and gauge-style visualizations to show this context.

*1) Map-based Visualizations:* We use geospatial maps to depict the spatial assumptions of AQI values to contextualize the sensor data. Currently, the EPA's air quality map, such as those provided by AirNow[2], uses colored contours to represent estimated air quality levels (see [21] for research examples). We build on this common technique by incorporating labeled contour lines to show the $PM_{2.5}$ spatial average for an area (see Figure 2). For the static visualization, uncertainty is shown using circle glyphs at the location of the sensor. We encoded uncertainty values using the boldness of the glyph's outline and the circle's radius. We further multiply radii using a scaling factor to ensure their clear visibility. Specifically, the larger and more transparent the circle glyph, the more uncertain the air quality data at that sensor location. This is in line with prior work comparing symbolic uncertainty encodings: the larger the glyph size, the higher the perceived uncertainty [42], [62]. Since this relation would need to be understood by any user using the visualization, a secondary encoding is used to avoid misinterpretation from solely relying on size. We follow [42] by using fuzziness as a secondary encoding of uncertainty. A darker circumference shows more certainty, consistent with [42]. We combine glyph fuzziness and circle size to encode uncertainty while enabling access to the visualization's color encoding: each circle's interior could be transparent enough to view the AQI-based color coding in a sensor's location.

*2) Time-based Visualizations:* We use radial plots to depict AQI values over time to contextualize the sensor data at a particular location. Following [29], the plot's angle represents time. It is structured into two concentric sections: the central section segmented into seasons and the outer into months following the Gregorian calendar (see Figure 1B). This hierarchy provides both a seasonal overview and a detailed monthly breakdown to improve readability and support comparisons of seasonal variations in $PM_{2.5}$ concentrations and sensor uncertainty (*c.f.*, [67], [77]).

The static representation (see Figure 3) uses a polar bar chart to maintain consistency with the other gauge-like visualization while adding a visual channel to represent uncertainty: bar height. Here, inverted height mapping is implemented to represent uncertainty in each section of the radial plot, i.e. taller bars indicate greater certainty in the air quality data. This design choice aligns with research suggesting that larger visual features attract attention and reduce cognitive load [41], [50], [59], [78]. Since it is important for users to quickly identify months and seasons with more reliable air quality data, emphasizing certainty through increased bar height makes the data more salient. Unlike the static, map-based visualization—where uncertainty encoding does not affect the visibility of the AQI color to represent air quality—the polar bar chart attracts visual attention to larger features [59] to direct users' attention to more certain data encoded with higher bars. Thus, this mapping was deliberately chosen despite contradicting our motivations for the map-based visualizations uncertainty encoding [42], [62]. However, as the uncertainty heights may be misinterpreted due to this difference, further features were added to improve clarity. Firstly, a red dotted line is added to represent the maximum acceptable uncertainty of $PM_{2.5}$ concentration as given by the EPA guidelines [48]. In addition, a red border is added to the bars below this line to imply that they have an uncertainty higher than that of the acceptable level. Since red is often perceived as a warning color [14], we use this color to reinforce the message that this data is not certain enough.

*3) Sensor-only Visualizations:* Inspired by the EPA's AirNow.gov gauge-style AQI widget[3], the sensor-only visualizations use a similar gauge design to compare air quality

[2] https://www.airnow.gov

[3] https://www.airnow.gov/aqi-widgets/

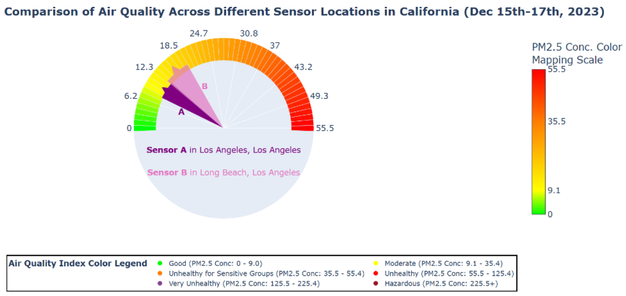


Fig. 4. Sensor-only static uncertainty visualization interface containing the PM2.5 concentration color mapping scale and the air quality color legend. Note that text has been enlarged for improved clarity.

from two sensor locations. The plot type was chosen for its simplicity in visualizing a single data stream with a needle and its common use in eco-feedback systems for goal-setting and comparisons [40], [46], [56]. Gauge charts are widely used in contexts like energy consumption and are easy to understand, even for non-technical users [7], [70]. In all three gauge plots, the outer semi-ring shows all $PM_{2.5}$ concentrations with the AQI color scale, with the needle's angle indicating the air quality level (see Figure 1C). Prior studies have used gauge charts for normative comparisons [40]. Rather than comparing sensor data to an average, we directly compare two sensor locations to allow users to determine which has better air quality. For the static visualization, the uncertainty is represented by the needle width, with greater uncertainty represented by a thicker needle (see Figure 4). This follows research on encoding uncertainty via size as in the static map-based visualization. The needle's width spans the range of possible values, where the width is double the uncertainty measures values calculated to represent error bars (see next Section for more detail).

## IV. Data Processing and Characterizing Uncertainty

Below we outline our processing steps for generating air quality data, which include choosing the data source, selecting an interpolation method, and calculating the related uncertainties. These steps are crucial for creating the visualizations.

### A. Air Quality Data

$PM_{2.5}$ concentration can pose significant health consequences and is one of a set of common air pollutants that inform air quality assessments [4]. We used data published in the official daily summary sheets of $PM_{2.5}$ concentrations collected by the EPA [3]. We focused on California's 2023 dataset because it provided the most recent and comprehensive year-long coverage. Additionally, the state's frequent summer wildfires would likely produce air quality visualizations that varied more than other states, as $PM_{2.5}$ concentrations fluctuate seasonally. Within the California dataset, we chose data from air monitors around Fresno because we could use the same data to show year-long trends over time, fluctuations in its spatial dispersion, and oscillations between air quality categories even in the same day.

### B. Temporal Resolution

The EPA dataset provides data with two levels of temporal resolution depending on the method used to interpret raw samples: FRM and FEM. FRM monitors are developed to a clearly defined standard for particulate matter concentration (either $PM_{2.5}$ or $PM_{10}$) set by the federal government and known for its high reliability, but their filter-based measurement method results in low temporal resolution as 24 hours are needed to collect the sample [1], [22], [74]. In comparison, FEM monitors are developed using new technologies that can support higher temporal resolution through continuous sampling such as an hourly temporal resolution [1], [22], [45]. Often FRM and FEM are used interchangeably since both can be used to monitor compliance with National Ambient Air Quality Standards [22]. In this study, we use the 24-hour average of $PM_{2.5}$ concentrations from continuous monitors collected using FEM to create the temporal resolutions needed for our study: either averaged over 3 days or 3 months. The online data source provided FRM monitor data every few days, typically every 3 days, so it was necessary to average the continuous monitor data to synchronize the dates. For the time-based visualization, a 3-month average was used to highlight long-term quality trends for decision-making. We further processed this data to calculate a measure of uncertainty by comparing these values with the FRM monitors at the same location and on the same day. If no FRM value was available at a given location, the corresponding FEM data point was disregarded, as its uncertainty could not be determined. Conversely, if only an FRM measurement was available, its uncertainty was assumed to be zero, as FRM is considered the reference standard. This decision means that uncertainty is only calculated when comparing FEM data to the more reliable FRM data, ensuring that uncertainty values are only assigned when there's an accurate reference point for comparison.

### C. Spatial Resolution

Collected data is limited by the spatial resolution of air quality monitors. This results in a sparse dataset. Yet, higher spatial resolution is needed to leverage this data for local decisions. To meet this need, spatial interpolation imputes synthetic values to depict air quality across the entire space including areas lacking sensor coverage. There are four main techniques for spatial interpolation, but none of these have been found clearly better [19]. In this study, we selected Kriging due to its greater flexibility in parameter tuning. For example, we could specify the shape of the semi-variogram, which models how spatial correlation changes with distance. The semi-variogram shape influences how values are interpolated, with different shapes (e.g., spherical, exponential, Gaussian) impacting the smoothness and accuracy of estimated values. Choosing an appropriate shape can lead to more precise estimates at unsampled locations by better capturing the spatial structure of air quality variations. To choose the shape of the semi-variogram, we applied Leave-One-Out Cross-Validation due to the limited sensor data which assesses different shapes by systematically omitting sensor readings and predicting their values to compare accuracy. A limitation of Kriging is that in areas where sensor data varies significantly between neighboring locations, interpolation can smooth out extreme values since it prioritizes spatial consistency. This sometimes hides the actual recorded concentrations. As a result, the true $PM_{2.5}$ concentration at a sampled location may differ from what is displayed on the interpolated map, potentially affecting how users perceive air quality at certain locations.

### D. *Quantifying sensor uncertainty*

To quantify sensor uncertainty, we use root mean square error (RMSE), represented as the following equation [27]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{1}$$

where:

- $n$ is the number of air quality data observations,
- $y_i$ is the observed value (i.e. the continuous sensor air quality data value) for the $i$-th observation,
- $\hat{y}_i$ is the actual value (i.e. the gold standard FRM monitor data) for the $i$-th observation.

The method calculates a measure of the difference between the continuous sensor air quality data and the gold standard FRM monitor data at the same site. As FRM monitor data is often not collected daily while continuous sensors are 24-hour averages, it is necessary to synchronize the dates. Based on the difference in days between FRM monitor data collections - usually 3 days - this is achieved by calculating the average of the continuous monitor data spanning between the preceding and subsequent FRM monitor data collection dates.

## V. Crowdsourced Evaluation

We evaluated the effectiveness of our visualizations by crowdsourcing a controlled comparison. We varied uncertainty data context and interactivity. For context, we used 3 conditions: map-based, time-based, and sensor-based. For levels of interactivity, we used 3 levels: static, animated, and interactive. In total, we designed 9 visualizations. To assess these, we compared them in terms of intelligibility of the AQI, interpretability of data uncertainty, and overall respondent confidence in their judgment. We detail our study more below.

### A. *Crowdsourcing Visualization Evaluations*

Crowdsourcing can offer rapid access to large numbers of diverse participants at relatively low cost [26]. Despite lacking many traditional experimental controls, crowdsourcing has seen increasing use as a tool for research. Successful applications have been demonstrated in areas such as social science [55], education [18], natural language processing [68], image labeling [61], [69], and more [33], [39]. Notably, work has shown that crowd-based studies can be equivalent to lab studies especially in the area of visualization and graphical perception [11], [17], [18], [24], [25], [66]. Crowdsourced studies can reproduce the results seen in lab-based equivalents of visualization studies, efficiently and often at lower cost.

We investigate trade-offs and added benefits of exposing viewers to the full possible distribution of a data's value by varying exposure through levels of interactivity, and by varying context, through visualization forms. To do so, we developed exploratory hypotheses,

- **H1** Less animation will support more accurate interpretation of the AQI. Specifically, we expect static visualization to do the best, followed by interactive, and animated.
- **H2** Greater interactivity will support better understanding of uncertainty. Specifically, we expect interactive visualizations to do the best, followed by animated, and finally static.
- **H3** Greater context will support more accurate and confident decision-making. Specifically, we expect spatial visualizations to do the best, followed by temporal, and finally, sensor-only.

We used 3 survey questions to measure evidence for each of our hypotheses. They asked for 1) the AQI interpretation of a specific value (*i.e.*, *good, moderate, etc.*), 2) to compare data from two sensors and choose which was more reliable, and 3) to rate confidence in the answers to 1 and 2 on a 7 point Likert scale.

### B. *Procedure*

Based on the 9 visualizations, we designed a set of basic demographic and background questions, trial questions, and post-trial questions. For the basic demographic and background questions, we adopted the questions from the US Census and the National Assessment of Educational Progress [9], [49]. These background questions were designed to elicit self-reported mathematical proficiency, helping us to understand the relationship between the interpretation of the visualization and numerical literacy.

To introduce the trials, we designed one training task based on the US EPA's AQI color scheme. We applied the specified translation to the scale to align the color scheme with the $PM_{2.5}$ concentration level [71]. We asked each respondent to identify a labeled portion of the scale. This provided both a baseline for comparison and an objective filter for whether participants could complete the study. Participants were then asked to answer three questions about each of the visualizations, alongside each visualization interface. In total, we designed 27 trial questions for the respondents. Due to the large number of potential comparisons, we chose to fully randomized the study trials. Finally, participants were asked to identify their preferred visualization and give optional comments on the visualization or study.

To minimize the impact of differing backgrounds and abilities between participants unduly influencing relative results we elected to use a within-subjects design. Equally, to avoid ordering effects, the order of the nine visualizations along with their corresponding trial questions was shuffled using the Fisher-Yates algorithm.

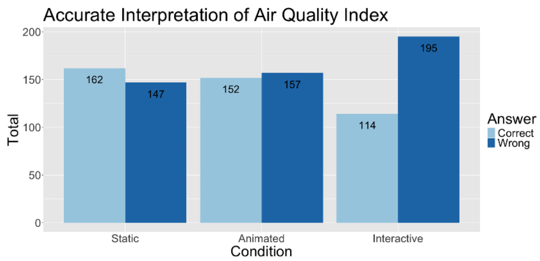


Fig. 5. Participants more accurately interpreted the AQI using static visualizations, but were substantially more inaccurate when using interactive uncertainty visualizations.

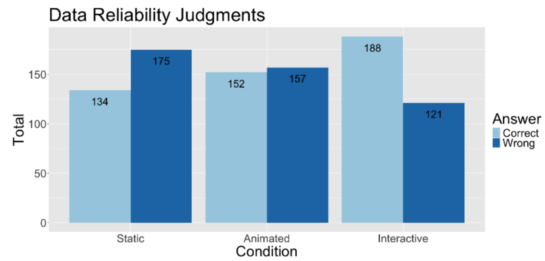


Fig. 6. Participants were better able to compare the reliability data reported by two sensors when using interactive visualizations than when using static or animated visualizations.

### *C. Participants*

We recruited 150 crowdworkers using Amazon Mechanical Turk. Due to the nature of the study, focused on US air quality metrics, we recruited only US-based crowdworkers. Based on testing by the research team, the estimated task completion time was approximately 20 minutes. Average earnings on Amazon Mechanical Turk is estimated to be $6.19/hr [23]. Following recommended practice, we targeted our payments above this level at U.S. minimum wage [76] and listed the task for $2.50.

Subsequent to data collection, we filtered responses on the basis of a) correct completion of the training task and b) completion time in greater than 7 minutes, a little over 1/3rd of our estimated completion time. Of the 150 responses, 103 participants correctly completed the training task, the fastest of which completed the task in 8 minutes 16 seconds. Regardless of their performance in the training task, or their completion time, all respondents were paid.

Of the 103 participants that successfully completed the AQI training task, 70 were male and 30 female; no respondents chose a non-binary option or declined to answer. All but one of these respondents indicated education to bachelors or higher, with a single participant indicating that their highest level of education was at the high school level.

## VI. Findings

### *A. Air Quality Index Interpretability*

As we expected, the static visualizations supported more accurate interpretation of the air quality index (AQI; 162 correct answers across the three layouts versus 147 incorrect) than for the other two visualization conditions. Notably, and to our surprise, participants were more inaccurate than accurate at interpreting the AQI using both the animated (152 correct verses 157 incorrect) and the interactive visualizations (114 correct verses 195 incorrect). While we expected interactive visualizations to outperform animated visualizations, this was not the case. A Binomial Regression using the Logit link function and a mixed effects model revealed the visualization condition had a significant impact on the correct interpretation of the AQI: $\chi^2(2, N = 103, p{<}0.001)$. Compared to static visualizations, we found that participants were only 0.3 times as likely (95% confidence interval [0.20,0.46]) to correctly interpret the AQI using interactive visualizations than static and 0.88 as likely (95% confidence interval [0.60, 1.28]) using animated visualization. That is, participants were 3 times more likely to make accurate AQI interpretations using static visualizations than interactive (95% confidence interval [2.16, 4.97]) and 1.14 more accurate than when using animated (95% confidence interval [0.78, 1.66]).

### *B. Air Quality Data Uncertainty*

In line with our expectations, the interactive visualizations supported more accurate comparisons of data reliability between 2 sensors (188 correct verse 121 incorrect). In line with our hypotheses, this was followed by animated visualizations (152 correct) and then, static visualizations (134 correct). Surprisingly, participants were more likely to make inaccurate comparisons of data reliability between 2 sensors than they were to make accurate comparisons when using animated or static visualizations. A Binomial regression using the Logit link function and a mixed effects model revealed the visualization condition had a significant impact on comparing data reliability between two sensors: $\chi^2(2, N = 103, p{<}0.001)$. We found that participants were 1.7 times as likely to correctly compare the reliability of data reported by two sensors when using interactive visualizations than when using animated (95% confidence interval [1.12, 2.46]), and 2.1 times more likely when using interactive visualizations than when using static visualizations (95% confidence interval [1.43, 3.16]).

### *C. Confidence in Air Quality Data Judgments*

We expected both the greater context of the location and season in the AQI visualizations to impact participants' confidence in their AQI interpretations and reliability judgments. While participants were less likely to correctly interpret the map condition their mean reported confidence was slightly higher than for other visualizations (Map: *mean*=5.8, *sd*=0.88, *range*=3-7; Time: *mean*=5.7, *sd*=0.94, *range*=2-7; Sensor: *mean*=5.7, *sd*=0.97, *range*=2-7). Further, the combined standard deviation over the three variants lower (0.86 map $<$0.94 time $<$0.98 sensor). This may be due to increased familiarity with this representation. While we did not observe much difference in participants' perceived confidence (see Figure 7). An Aligned Rank Transform followed by an RMANOVA found a significant main effect of visualization type on confidence:

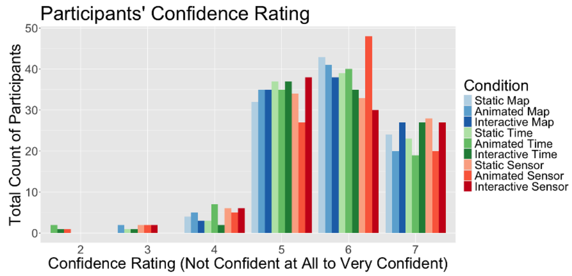


Fig. 7. Despite high numbers of inaccurate AQI interpretations and data reliability comparisons, study participants were confident in their judgments overall.

F(2,816)=7.65, p<0.001. Subsequent pairwise comparisons using Tukey's adjustment method found significant differences between the map and sensor conditions (*t(816)*=-2.75, $p$<0.05) and the sensor and time conditions (*t(816)*=3.78, $p$<0.001). Visual inspection of our results suggests that the animated sensor was largely contributing to differences across the visual context types (see Figure 7).

We expected the greater context to improve participant's interpretation of the AQI. Contrary to expectations, the spatial condition supported less accurate answers (207 incorrect) than the temporal condition (165 incorrect) and sensor only condition (163 incorrect). A Binomial Regression using the Logit link function and a mixed effects model revealed the contextual condition had a significant impact on the correct interpretation of the AQI: $\chi^2(2, N = 103, p{<}0.001)$. Compared to temporal and sensor contexts, we found participants were half as likely to correctly interpret the AQI using spatial context (0.55 as likely than sensor only context, with 95% confidence interval of [0.37, 0.81]; and 0.56 as likely than temporal context, with 95% confidence interval of [0.38, 0.83]).

### *D. Discriminating Uncertainty*

Interestingly, for the 47 respondents who provided incorrect responses to the training question, just 4% for the static conditions, 5% for the interactive conditions, and 4% for the animated conditions indicated that they were unable to interpret the visualization across all trials. This is likewise reflected in the reported confidence score, with only 4% of trials (all variants) indicating a lack of confidence in the response. Further, of those trials reported with low confidence, participants were actually correct in 34.5% of cases. Due to the small numbers of these cases, we do not claim any particular interpretation on this aspect of our results.

### *E. Self-reported Mathematical Ability*

Our survey included three questions to indicate self-reported mathematical ability. We asked respondents to indicate whether they would be able to calculate an average (mean), percentage (restaurant tip), and describe properties of a triangle (geometry). Of our 103 respondents who passed our training task, 85 reported being able to complete at least one of these tasks. 60 participants indicated that they would be able to do at least two of these tasks. Overall, respondents indicated a mean ability of 1.7 (*SD* = 1.1). When also considering the indicated confidence of their report, responses were approximately normal.

To identify the effect of self-reported mathematical ability on our participants' interpretations of our visualizations, we conducted Pearson correlations. Self-reported mathematical ability had a small positive correlation (0.12) with the participants' confidence in interpreting the visualizations. Surprisingly, self-reported mathematical ability had a moderate *negative* correlation ($-0.30$) with the number of correct responses. This negative correlation was present for both the ability to correctly interpret an absolute (weak, $-0.13$) and relative (nominal, $-0.03$) reading.

### *F. The Potential of Gradients for the AQI*

The official color scheme for the AQI uses discrete color bands. Our use of a more fine-grained color banding potentially offers improved access to the data, through an additional encoding, but at the cost of increased complexity in interpreting that data. For our sensor-based visualization—which closely replicates the clean air scale visualization—we gave respondents the option to indicate that they were unable to interpret the visualization. This enabled us to collect data on whether participants faced challenges with interpreting the enhanced gradients. We use these responses to isolate the effect of more finely graduated color scales on the participant's ability to correctly discern the underlying data. None of our 103 respondents indicated an inability to interpret the static EPA-style sensor visualization, though 2% and 3% of trials indicated some challenges when interpreting the interactive and animated visualizations, respectively. Further, only 2% (static), 2% (interactive), and 3% (animated), of trials indicated a lack of confidence in the associated answers. Only in the interactive trials were these answers actually incorrect. Thus, respondents are capable of interpreting data represented using more granular color scales than used by the US EPA.

### *G. User Preferences*

Our respondents were given the opportunity to identify their most and least preferred visualization representation. For brevity, we presented only the type of visualization, not the level of interactivity. Nearly 40% of respondents indicated that the map-based visualizations were the least preferred. As the map-based visualizations elicited the most incorrect responses, the trouble experienced by participants could be expected. However, while the sensor visualization elicited the most frequently correct responses, respondents narrowly indicated that they most preferred the time-based visualizations (35%). Overall, there was no clear leader with map-based visualizations coming a close second (34%) and time-based closely following (31%).

## VII. Discussion

### *A. Balancing AQI Judgments with Understanding Uncertainty*

We found that participants more accurately interpreted the AQI when using static visualizations, but they were better at comparing the uncertainty of data from two sensors when using interactive visualizations. Notably, participants were 33% more inaccurate at interpreting the AQI when using the interactive visualizations compared to static. This suggests there may be a trade-off between interpreting the AQI and understanding the uncertainty associated with the underlying data. We speculate that this effect may be due to participants overly focusing on outliers—or interesting and specific aspects—in the data, rather than the overall probability distribution, and so, may impact the overall assessment. Alternatively, respondents may have overly focused on aspects that support their preexisting biases; particularly in contentious subject areas such as air quality and public health data.

Participants were more confident using the map-based visualization than the other types of visual contexts. Participants are likely familiar with map-based data representations as they regularly encounter similar visualizations through weather reports which increasingly include air-quality updates, such as pollen counts [20]. For our time-based visualizations, however, the additional context of seasonal fluctuations in $PM_{2.5}$ in the visualizations resulted in the least confident answers. The radial plot was most likely novel for the participants. In particular, our use of longer bars for lower uncertainty may have been unexpected, and may have lowered participants' reported confidence rating. At the same time, the animated sensor visualization had qualitatively distinct confidence judgements, but these were averaged out in our analysis. These varying uncertainty encodings may be interacting with visual context and support for interactivity to reveal a much more complex impact on user confidence. Future work should investigate interaction effects of levels of interactivity and visual context and their impact on confidence in one's own judgments.

Surprisingly, we found that self-reported mathematical ability negatively correlated with correct responses. This raises questions about the use of novelty in scientific communications. While our study had a clear air quality theme, unexpected presentation types may have negatively influenced perception. This may suggest that more basic and familiar presentation types offer a more effective communication medium for reporting absolute values to visualization users. However, our results also highlight that animation offers an important design dimension for facilitating interpretation of uncertainty. Notably, this contrasts with the interactive plots where designers may surrender some control of this dimension to the user. The ability to direct the viewer's sequential and cumulative attention may be important for understanding uncertainty, but interactive control may risk fixation on specific data points rather than aggregated probabilities.

### B. Interpreting AQI Colors

Unexpectedly, participants were significantly less accurate when interpreting AQI values with interactive visualizations compared to the other two interaction levels. When looking more closely at the results for map-based, time-based, and sensor-only interactive plots, map-based interactive visualizations had the greatest gap between incorrect and correct answers for interpreting AQI values. One possible explanation for this is related to the limitation of Kriging. After using the spatial interpolation technique, air quality data at a specific location is sometimes hidden if neighboring data is very different. Consequently, the AQI interpretation for a given location might not accurately reflect the actual $PM_{2.5}$ concentration visible on the map-based visualization. In our study, we asked about the AQI color at a sensor location that could have appeared green, but instead, appeared yellow. We used FRM monitor data for AQI coloring, which resulted in a limited number of data points. This spatial sparsity may have resulted in the specific sensor location's $PM_{2.5}$ data being concealed after applying Kriging. Improving the granularity of this technique or exploring alternative spatial interpolation methods could potentially address this issue. Further investigation into these factors would be valuable to better understand the high number of incorrect responses associated with this visualization type.

### C. Participant Demographics

While our anticipated effect size was small, we chose to use a similar size participant pool as with other crowdsourced visualization studies (e.g. [25]). Despite this constraint, we were still able to detect relatively small differences in our data. However, the size of our participant pool and Amazon Mechanical Turk pose limitations for our study demographics. Notably, our participant pool was mostly college-educated. For the purpose of studying scientific communication to the public, future work should investigate whether our findings hold for people with other education levels and gender identity. To give context for our sample, only about $1/3$ of the US population has at least a college degree [44]. Also, our participant pool skewed male. Further, it did not include anyone who identified as non-binary. Future work should examine whether our findings hold for a more diverse population.

## VIII. Conclusion

We presented a set of air quality data representations that vary in their support for interactivity and context to make sense of uncertainty using visualization techniques. We used these to explore how well they supported accurate air quality judgments, comparing certainty across two sensor data streams, and facilitating confidence in one's own judgment. We found by supporting interactivity with sensor data uncertainty, participants were more accurate when comparing the certainty of two sensor data streams than when using static or animated visualizations, but they were also the less accurate at interpreting the AQI when using the interactive versions. While static visualization supported more accurate AQI interpretations, we found that animated versions did not differ markedly from static in supporting AQI interpretations. Our work shows that when communicating uncertainty to the public, care must be taken to balance support for understanding uncertainty with helping users make decisions in the context of that uncertainty.

ACKNOWLEDGMENT

Collaborative Research Grant funding was provided by the Halle Institute for Global Research at Emory University and the University of St Andrews. Support was also provided by an Independent Research Grant through the Office of Undergraduate Research Programs through Emory University's College of Arts and Sciences.

REFERENCES

[1] Code of federal regulations, title 40, parts 50-59, 2025.

[2] AAMER, H., MUMTAZ, R., ANWAR, H., AND POSLAD, S. A very low cost, open, wireless, internet of things (iot) air quality monitoring platform. In *2018 15th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT (HONET-ICT)* (2018), pp. 102–106.

[3] Air data: Air quality data collected at outdoor monitors across the us. https://www.epa.gov/outdoor-air-quality-data.

[4] Air sensor guidebook. https://cfpub.epa.gov/si/si_public_record_report .cfm?Lab=NERL&dirEntryId=277996.

[5] APPERT, C., AND BEAUDOUIN-LAFON, M. Swingstates: adding state machines to the swing toolkit. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2006), UIST '06, Association for Computing Machinery, p. 319–322.

[6] BRODLIE, K., ALLENDES OSORIO, R., AND LOPES, A. *A Review of Uncertainty in Data Visualization*. Springer London, London, 2012, pp. 81–109.

[7] BROWN, C. J., AND MARKUSSON, N. The responses of older adults to smart energy monitors. *Energy policy 130* (2019), 218–226.

[8] BURCH, M., AND WEISKOPF, D. *On the Benefits and Drawbacks of Radial Diagrams*. Springer New York, New York, NY, 2014, pp. 429–451.

[9] BUREAU, U. C. Educational Attainment, Undergraduate Field of Degree. https://www.census.gov/programs-surveys/acs/.

[10] Quality assurance air monitoring site information. https://ww2.arb.ca.gov/applications/quality-assurance-air-monitoring-site-information.

[11] CASLER, K., BICKEL, L., AND HACKETT, E. Separate but equal? a comparison of participants and data gathered via amazon's mturk, social media, and face-to-face behavioral testing. *Computers in Human Behavior 29*, 6 (2013), 2156–2160.

[12] CORRELL, M., AND GLEICHER, M. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics 20* (2014), 2142–2151.

[13] CORRELL, M., MORITZ, D., AND HEER, J. Value-suppressing uncertainty palettes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), CHI '18, Association for Computing Machinery, p. 1–11.

[14] DANGER, E. The colour handbook: How to use colour in commerce and industry, 1987.

[15] DEY, A. K., AND MANKOFF, J. Designing mediation for context-aware applications. *ACM Trans. Comput.-Hum. Interact. 12*, 1 (mar 2005), 53–80.

[16] EILENBERG, S. R., SUBRAMANIAN, R., MALINGS, C., HAURYLIUK, A., PRESTO, A. A., AND ROBINSON, A. L. Using a network of lower-cost monitors to identify the influence of modifiable factors driving spatial patterns in fine particulate matter concentrations in an urban environment. *journal of exposure science & environmental epidemiology* (2020).

[17] FLEISCHER, A., MEAD, A. D., AND HUANG, J. Inattentive responding in mturk and other online samples. *Industrial and Organizational Psychology 8*, 2 (2015), 196–202.

[18] FOLLMER, D. J., SPERLING, R. A., AND SUEN, H. K. The role of mturk in education research: Advantages, issues, and future directions. *Educational Researcher 46*, 6 (2017), 329–334.

[19] FONTES, T., AND BARROS, N. Interpolation of air quality monitoring data in an urban sensitive area: The oporto/asprela case. *Revista da Faculdade de Ciência e Tecnologia da Universidade Fernando Pessoa* (01 2010).

[20] GIBERSON, J. E., MURRAY, L. W., AND FRENZ, D. A. Pollen counts in the television media. In *Allergy and Asthma Proceedings* (1997), vol. 18, OceanSide Publications, p. 19.

[21] HABEEB, R., AND JAVAID, S. Using gis interpolation technique to assess urban air quality and evolve remediation strategies, 11 2019.

[22] HALL, E. S., KAUSHIK, S. M., VANDERPOOL, R. W., DUVALL, R. M., BEAVER, M. R., LONG, R. W., AND SOLOMON, P. A. Integrating sensor monitoring technology into the current air pollution regulatory support paradigm: Practical considerations. *Am. J. Environ. Eng 4*, 6 (2014), 147–154.

[23] HARA, K., ADAMS, A., MILLAND, K., SAVAGE, S., CALLISON-BURCH, C., AND BIGHAM, J. P. A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, Apr. 2018), CHI '18, Association for Computing Machinery, pp. 1–14.

[24] HAUSER, D. J., AND SCHWARZ, N. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods 48*, 1 (Mar. 2016), 400–407.

[25] HEER, J., AND BOSTOCK, M. Crowdsourcing graphical perception: using mechanical turk to assess visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2010).

[26] HIRTH, M., JACQUES, J., RODGERS, P., SCEKIC, O., AND WYBROW, M. Crowdsourcing technology to support academic research. In *Evaluation in the Crowd. Crowdsourcing and Human-Centered Experiments* (2017).

[27] HODSON, T. O. Root mean square error (rmse) or mean absolute error (mae): When to use them or not. *Geoscientific Model Development Discussions 2022* (2022), 1–10.

[28] HOGRÄFER, M., HEITZLER, M., AND SCHULZ, H.-J. The state of the art in map-like visualization. *Computer Graphics Forum 39*, 3 (2020), 647–674.

[29] HUBER, C., NAGEL, T., AND STUCKENSCHMIDT, H. Experiencing data on location: A case study of visualizing air quality for citizens. *KN-Journal of Cartography and Geographic Information 73*, 2 (2023), 97–108.

[30] HUDSON, S. E., AND NEWELL, G. L. Probabilistic state machines: Dialog management for inputs with uncertainty. In *Proceedings of the 5th annual ACM symposium on User interface software and technology* (1992), pp. 199–208.

[31] HULLMAN, J. Why authors don't visualize uncertainty. *IEEE transactions on visualization and computer graphics 26*, 1 (2019), 130–139.

[32] HULLMAN, J., RESNICK, P., AND ADAR, E. Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PLoS One 10*, 11 (Nov. 2015), e0142444.

[33] JACQUES, J. T., AND KRISTENSSON, P. O. Design strategies for efficient access to mobile device users via amazon mechanical turk. In *Proceedings of the First ACM Workshop on Mobile Crowdsensing Systems and Applications* (New York, NY, USA, 2017), CrowdSenSys '17, Association for Computing Machinery, p. 25–30.

[34] JOSLYN, S. L., AND LECLERC, J. E. Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *Journal of Experimental Psychology: Applied 18*, 1 (2012), 126 – 140.

[35] KALE, A., NGUYEN, F., KAY, M., AND HULLMAN, J. Hypothetical outcome plots help untrained observers judge trends in ambiguous data. *IEEE transactions on visualization and computer graphics 25*, 1 (2018), 892–902.

[36] KAY, M., KOLA, T., HULLMAN, J. R., AND MUNSON, S. A. 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* (2016), pp. 5092–5103.

[37] KAY, M., MORRIS, D., SCHRAEFEL, M., AND KIENTZ, J. A. There's no such thing as gaining a pound: Reconsidering the bathroom scale user interface. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (2013).

[38] KERR, J. R., SCHNEIDER, C. R., FREEMAN, A. L. J., MARTEAU, T., AND VAN DER LINDEN, S. Transparent communication of evidence does not undermine public trust in evidence. *PNAS Nexus 1*, 5 (12 2022), pgac280.

[39] KITTUR, A., CHI, E. H., AND SUH, B. Crowdsourcing user studies with mechanical turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2008), CHI '08, Association for Computing Machinery, p. 453–456.

[40] Kohlbrecher, J., Hakobyan, S., Pickert, J., and Grossmann, U. Visualizing energy information on mobile devices. In *Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems* (2011), vol. 2, IEEE, pp. 817–822.

[41] Luna, K., Nogueira, M., and Albuquerque, P. B. Words in larger font are perceived as more important: explaining the belief that font size affects memory. *Memory 27*, 4 (2019), 555–560.

[42] MacEachren, A. M., Roth, R. E., O'Brien, J., Li, B., Swingley, D., and Gahegan, M. Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2496–2505.

[43] Mankoff, J., Hudson, S. E., and Abowd, G. D. Providing integrated toolkit-level support for ambiguity in recognition-based interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2000), CHI '00, Association for Computing Machinery, p. 368–375.

[44] McElrath, K., and Martin, M. Bachelor's Degree Attainment in the United States: 2005 to 2019. Tech. Rep. ACSBR-009, U.S Census Bureau, 2021.

[45] Mou, C.-Y., Hsu, C.-Y., Chen, M.-J., Chen, Y.-C., et al. Evaluation of variability in the ambient pm2. 5 concentrations from fem and frm-like measurements for exposure estimates. *Aerosol and Air Quality Research 21*, 4 (2021), 200217.

[46] Murugesan, L., Hoda, R., Salcic, Z., and Verma, P. Policy recommendations to induce behavioural changes through interactive energy visualisation. In *2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)* (2018), pp. 1091–1096.

[47] Myers, B. A. *Pick, Click, Flick! The Story of Interaction Techniques*. ACM, 2024.

[48] Nadzir, Utbah Rabuan, m. S. M., Sham, S. Z. A., Bahri, S. B. I. W. S., Borah, J., Majumdar, S., Lei, T. M. T., Ali, S. H. M., Wahab, M. I. A., and Yunus, N. H. M. Evaluations of low-cost air quality sensors for particulate matter (PM2.5) under indoor and outdoor conditions. *Sens. Mater. 35*, 8 (Aug. 2023), 2881.

[49] "national assessment of educational progress". https://nces.ed.gov/nationsreportcard/.

[50] Orquin, J. L., Lahm, E. S., and Stojić, H. The visual environment and attention in decision making. *Psychological Bulletin 147*, 6 (2021), 597–617.

[51] Padilla, L., Kay, M., and Hullman, J. Uncertainty visualization, Apr 2020.

[52] Padilla, L., Kay, M., and Hullman, J. *Uncertainty Visualization*. Wiley, 2022, ch. 22.

[53] Pang, A., Wittenbrink, C., and Lodha, S. Approaches to uncertainty visualization. *The Visual Computer 13* (10 1996).

[54] Pang, A. T., Wittenbrink, C. M., and Lodha, S. K. Approaches to uncertainty visualization. *Vis. Comput. 13*, 8 (Nov. 1997), 370–390.

[55] Paolacci, G., and Chandler, J. Inside the turk: Understanding mechanical turk as a participant pool. *Current Directions in Psychological Science 23*, 3 (2014), 184–188.

[56] Pereira, L., and Nunes, N. Understanding the practical issues of deploying energy monitoring and eco-feedback technology in the wild: Lesson learned from three long-term deployments. *Energy Reports 6* (2020), 94–106.

[57] Potter, K., Rosen, P., and R. Johnson, C. From quantification to visualization: A taxonomy of uncertainty visualization approaches. In *Proceedings of the 2012 IFIP International Federation for Information* (2012).

[58] Preston, A., and Ma, K.-L. Communicating uncertainty and risk in air quality maps. *IEEE Transactions on Visualization and Computer Graphics 29*, 9 (2022), 3746–3757.

[59] Proulx, M. J. Size matters: large objects capture attention in visual search. *PLoS One 5*, 12 (Dec 2010), e15293.

[60] Reinhart, A., Brooks, L., Jahja, M., Rumack, A., Tang, J., Agrawal, S., Al Saeed, W., Arnold, T., Basu, A., Bien, J., et al. An open repository of real-time covid-19 indicators. *Proceedings of the National Academy of Sciences 118*, 51 (2021), e2111452118.

[61] Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., and Tomlinson, B. Who are the crowdworkers? shifting demographics in mechanical turk. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2010), CHI EA '10, Association for Computing Machinery, p. 2863–2872.

[62] Sanyal, J., Zhang, S., Bhattacharya, G., Amburn, P., and Moorhead, R. J. A user study to compare four uncertainty visualization methods for 1D and 2D datasets. *IEEE Trans. Vis. Comput. Graph. 15*, 6 (Nov. 2009), 1209–1218.

[63] Schneider, M., McDowell, M., Guttorp, P., Steel, E. A., and Fleischhut, N. Effective uncertainty visualization for aftershock forecast maps. *Nat. Hazards Earth Syst. Sci. 22*, 4 (Apr. 2022), 1499–1518.

[64] Schwarz, J., Mankoff, J., and Hudson, S. Monte carlo methods for managing interactive state, action and feedback under uncertainty. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (2011), pp. 235–244.

[65] Skeels, M., Lee, B., Smith, G., and Robertson, G. Revealing uncertainty for information visualization. *Sage Journals Information Visualization* (2010).

[66] Smith, S. M., Roster, C. A., Golden, L. L., and Albaum, G. S. A multi-group analysis of online survey respondent data quality: Comparing a regular usa consumer panel to mturk samples. *Journal of Business Research 69*, 8 (2016), 3139–3148.

[67] Smith, T., and Bookhagen, B. Assessing uncertainty and sensor biases in passive microwave data across high mountain asia. *Remote Sensing of Environment 181* (2016), 174–185.

[68] Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. Cheap and fast – but is it good? evaluating non-expert annotations for natural language tasks. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (Honolulu, Hawaii, Oct. 2008), M. Lapata and H. T. Ng, Eds., Association for Computational Linguistics, pp. 254–263.

[69] Sorokin, A., and Forsyth, D. Utility data annotation with amazon mechanical turk. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (2008), pp. 1–8.

[70] Tarabieh, K. A., Elnabarawy, I. O., Mashaly, I. A., and Rashed, Y. M. The power of data visualization: A prototype energy performance map for a university campus. In *Sustainable Human–Building Ecosystems*. 2015, pp. 194–203.

[71] U.S. Environmental Protection Agency. Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI). Tech. Rep. EPA-454/B-24-002.

[72] Williams, K., Hammer, J., and Hudson, S. The iot codex: A book of programmable stickers for authoring and composing embedded computing applications. In *2023 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (2023), IEEE, pp. 1–11.

[73] Wong, D. W., Yuan, L., and Perlin, S. A. Comparison of spatial interpolation methods for the estimation of air quality data. *Journal of Exposure Science & Environmental Epidemiology* (2004).

[74] Yanosky, J. D., and Maclntosh, D. L. A comparison of four gravimetric fine particle sampling methods. *Journal of the Air & Waste Management Association 51*, 6 (2001), 878–884.

[75] Yau, N. *Visualize this: the FlowingData guide to design, visualization, and statistics*. John Wiley & Sons, 2021.

[76] Young, J., and Young, K. Don't Get Lost in the Crowd: Best Practices for Using Amazon's Mechanical Turk in Behavioral Research. *Journal of the Midwest Association for Information Systems (JMWAIS) 2019*, 2 (Oct. 2019).

[77] Zhao, N., Liu, Y., Vanos, J. K., and Cao, G. Day-of-week and seasonal patterns of pm2.5 concentrations over the united states: Time-series analyses using the prophet procedure. *Atmospheric Environment 192* (2018), 116–127.

[78] Zhu, T., and Yang, Y. Research on mobile learning platform interface design based on college students' visual attention characteristics. *PLoS One 18*, 7 (Jul 2023), e0283778.