

# Non-Uniform Bivariate Uncertainty Scales

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

## ABSTRACT

Uncertainty is a vital concern in many datasets. Visualizations often employ bivariate mappings as a way to present data and uncertainty values simultaneously. Yet, these bivariate maps can be difficult to interpret, and interference between visual channels can reduce the discriminability of marks. To address these issues, we contribute Value-Suppressing Uncertainty Maps (VSUMs). VSUMs allocate larger ranges of marks to data when uncertainty is low, and smaller ranges when uncertainty is high. This non-uniform budgeting of marks suppresses differences in uncertain data, but amplifies them for certain data, with the intended outcome of supporting more cautious patterns of decision-making. We demonstrate several examples of VSUMs, and present a crowdsourced evaluation showing that, compared to traditional bivariate maps, VSUMs encourage people to XYZ.

## ACM Classification Keywords

H.5.0. Information Interfaces and Presentation (e.g. HCI): General

## Author Keywords

Uncertainty Visualization; Color Perception; Thematic Maps; Semiotics.

## INTRODUCTION

Uncertainty is an inescapable component of collecting, presenting, and using data. A common goal in the communication of uncertainty is promoting *uncertainty-aware decisions*: the audience should be aware of the risks and rewards of certain decisions, modulate their confidence in their conclusions, and perhaps refrain from making a decision at all if there is too much uncertainty. A way that designers can contribute to this goal is by ensuring that uncertainty information is *well-integrated* with the rest of the data. That is, it should be difficult to discount or ignore the uncertainty in a dataset.

Simultaneous presentation of uncertainty and value necessitates the construction of a bivariate map — a relation, in terms of visual variables, between 2-tuples (value, uncertainty) and mark properties. The design of bivariate maps is difficult

and, due to the interference and interplay between different visual variables, often leads to maps with few discriminable categories.

In this paper, we contribute **Value-Suppressing Uncertainty Maps** (VSUMs) for integrating data and uncertainty information in visualizations. VSUMs intentionally alias together data values with high uncertainty, affording greater discriminability as uncertainty decreases. If traditional bivariate maps are a two-dimensional square, with differing outputs for each combination of value and uncertainty, VSUMs can be thought of as arcs: as uncertainty increases, values are mapped to smaller and smaller sets of outputs, culminating in a singularity where all inputs are mapped to an identical, highly uncertain mark regardless of their data value. Figure 1 shows examples of both a more traditional bivariate map and a VSUM.

We describe the motivations behind the use of VSUMs, examples of their utility for decision-making under uncertainty, and assess VSUMs in a crowdsourced experiment. Our experimental results indicate that VSUMs create close integration between uncertainty and data, promoting rapid but cautious decision-making.

## RELATED WORK

Despite the acknowledged importance of uncertainty in understanding data, explicit representations of uncertainty are often missing from visualizations [5]. This is partially due to the complexity of uncertainty as a concept. In typologies from Thomson et al. [41] and Battenfield & Beard [7], the authors note that many, occasionally contradictory, concepts can fall under the category of “uncertainty,” including data quality, sampling error, credibility, and provenance.

To our knowledge, there are no established standards for visualizing uncertainty in heatmaps. Reviewing the state of the art in uncertainty visualization in the field, Greithe et al. [18] and Brodlić et al. [6] present lists of potential techniques for conveying uncertainty in concert with data. Some of these options (interaction, animation, and sonification) are not applicable for static charts; even for dynamic charts, many users do not interact with charts in sufficient detail to recover uncertainty information [43]. While we acknowledge the potential utility of these techniques in uncertainty visualization (for instance, the animation used to convey sampling error in Hullman et al.’s Hypothetical Outcome Plots [21]), we limit the scope of our discussion to techniques which center on static charts.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

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

© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 123-4567-24-567/08/06...\$15.00

DOI: [http://dx.doi.org/10.475/123\\_4](http://dx.doi.org/10.475/123_4)

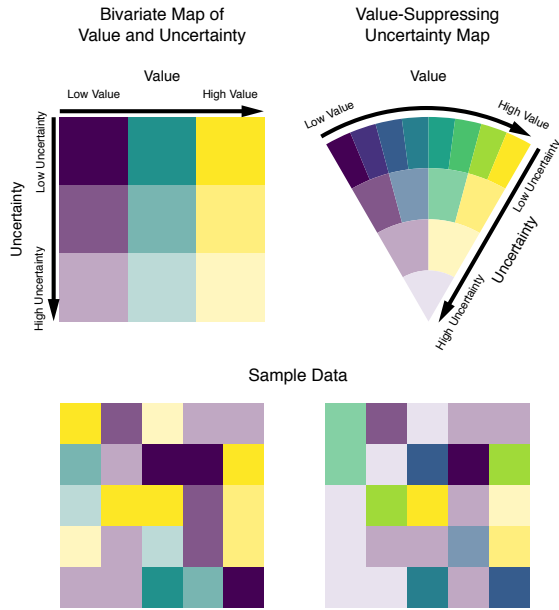


Figure 1: A standard bivariate map (left) and a VSUM (right), used to encode an identical 5x5 grid of random data. Both use the same visual channels to encode value (position along the Viridis [45] color map) and uncertainty (lightness and saturation), and have an equal standard of perceptual discriminability (at least 18 units of distance in CIELAB color space between colors). However, highly uncertain values result in colors that are very close together in the bivariate case, meaning the full bivariate map is only 9 bins under these constraints — a 4x4 map would result in colors that are perceptually too close together. By contrast, the VSUM intentionally reduces bins when uncertainty is high, eventually aliasing all highly uncertain values to the same color. This decision affords more distinct colors in other regions of the map, and so an increase in overall bins to 15. The resulting map suppresses the value of highly uncertain data, but increases the discriminability of data with low uncertainty.

### Visual Variables for Representing Uncertainty

The decision to explicitly encode uncertainty increases the dimensionality of the data, and so requires the use of (at least one) additional visual channel [6]. Uncertainty therefore inherently increases the visual complexity of a visualization. When the data are already complex to convey, and many of the more common or accurate visual variables are in use, allocating an additional dimension is non-trivial. As the number of dimensions increases, finding visual variables that are both perceptually accurate (in either estimation of quantity or discrimination of category) as well as perceptually separable from all the other encoding channels, becomes more and more difficult.

A further hurdle is that not all visual variables are well-suited for conveying uncertainty. MacEachren et al. [25] evaluate a number of visual variables with respect to their *semiotic fit* for representing uncertainty. They observe that certain visual variables such as blurriness and transparency seem to

have a more intuitive connection to uncertainty than other variables such as shape or hue. Unfortunately, Boukhelifa et al. [4] find that many visual variables habitually used for conveying uncertainty, such as blur and value, are also difficult to estimate. This results in a preference/performance gap where designers must choose between encoding uncertainty in a way that is intuitive but error prone, or use higher fidelity channels that may be more difficult to interpret.

### Bivariate Maps

#### Quantizing Color Scales

For spatial visualizations like choropleth maps, heatmaps, and treemaps, visual channels such as position and length are reserved for data variables other than value (such as geographic location or relative size). In these situations, data value is often encoded using color. Colors in such univariate quantitative color maps should be sufficiently far apart as to be perceptually distinguishable [47], and vary in lightness as well as hue to afford an implicit ordering of value [2, 35]. Different choices of color maps can highlight different features of the data, and should be chosen with care based on the data types and important features to be visualized [36].

Given these considerations, and the perceptual integrality of color channels [15, 48], the construction of *bivariate* color maps is difficult. Viewers need to be able to independently assess the variables of interest, but the resulting bivariate map should also be expressive and easy to interpret. Bivariate maps with these constraints are often limited to relatively few categories (say, a 3x3 matrix as in Figure 1) [34, 42]. In general, the quality of bivariate color maps is a multivariate measure involving consideration of not just the component color channels, but also the interpolation scheme and the color of the surround [1]. However, even simple bivariate maps can be difficult for a general audience to interpret [46].

#### VALUE-SUPPRESSING UNCERTAINTY MAPS

Value-Suppressing Uncertainty Maps (VSUMs) are a technique for creating bivariate maps of data *value* and *uncertainty*.

In some cases, this second assumption is violated. For instance, an analyst might be interested in “long tail risks” or other “black swan events,” where the impact of a value, no matter how uncertain, must be considered and planned for [40]. Other analysis tasks (such as filtering out outliers), require increased, rather than decreased, discriminability when uncertainty is high. However, for many information fusion tasks, the assumption is that uncertainty is related to data quality, or the uncertainty of data values [33]. As with null hypothesis significance testing, the aim would then be to avoid Type I errors, where patterns arising from statistical noise or measurement error are interpreted as significant signals.

If both of these assumptions hold, then it follows that the designer of a bivariate map should allocate more mark types to certain values, and fewer mark types to uncertain values. VSUMs codify this decision by *reducing the number of mark categories for representing value as uncertainty increases*. This means that data encoded using a VSUM will make visible

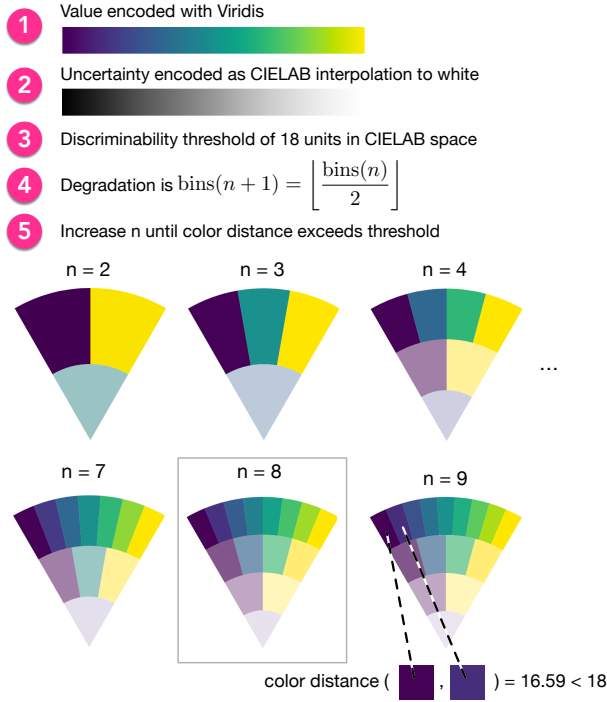


Figure 2: The process for creating a VSUM. VSUMs guarantee that no two marks in the resulting bivariate mapping are perceptually closer than some pre-specified threshold. This guarantee removes some of the issues of non-separability and ambiguity that are otherwise a concern when designing bivariate maps.

In this example two colors from the map for  $n = 9$  are too close in CIELAB space. Thus, we create a VSUM with  $n = 8$  color bins at the lowest level of uncertainty.

only the largest of differences in uncertain data, but highlight comparatively small differences in value when uncertainty is low. VSUMs therefore act as both a filtering mechanism (in that values with too much uncertainty are all mapped to the same glyph), as well as an implicit test of effect size (in that smaller and smaller changes in value are visible as uncertainty decreases). This strategy of dampening the low quality or high uncertainty values in maps in order to focus on more informative regions has measurable benefits, including the removal of statistically spurious visual patterns, and the highlighting of regions of interest [12].

## Method

The general VSUM construction algorithm is as follows:

1. **Select a visual variable to encode value.** In this paper, this variable is position along some established color scale.

Figure 2 shows this process in detail, and how it can be used to generate a bivariate map for presenting uncertainty. JavaScript code for generating VSUMs given arbitrary D3 [3] quantitative scale functions is available at [URL-REMOVED-FOR-REVIEW](#).

Correll et al. [9, 10] use a precursor of VSUMs in their LayerCake genomics visualization tool, where marks repre-

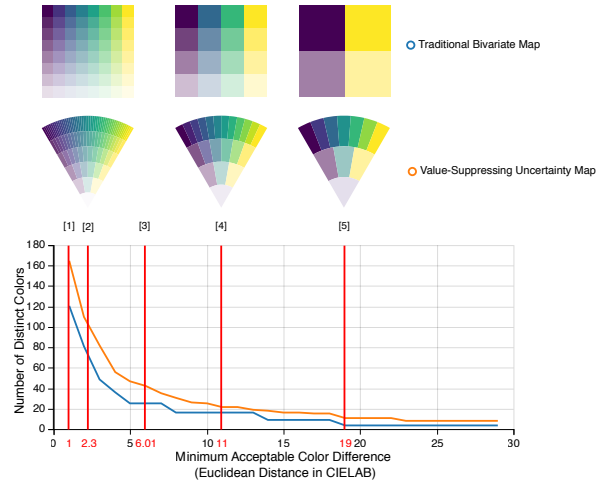


Figure 3: The number of discrete color categories in traditional bivariate color maps versus VSUMs, assuming value is encoded by the Viridis color map [45], and uncertainty is encoded by saturation/value (whiter values are more uncertain). Since these whiter uncertain colors are closer together in CIELAB space, and VSUMs intentionally require fewer color bins in those regions, VSUMs always have more color categories than the corresponding 2D bivariate map. The red lines denote: **1)** Theoretical 50% JND from CIELAB specification. **2)** 50% JND from controlled lab study of Mahy et al. [27]. **3)** 50% JND from Szafr et al.[38]. **4)** 50% JND for small visual targets from Stone et al. [37]. **5)** The threshold resulting in the simplest possible (2x2) bivariate map.

sending genomic data with increasing uncertainty are mapped to a smaller and smaller set of increasingly grey colors, creating the effect of uncertain values retreating into “confidence fog” while highly certain values remain prominent. Other bivariate visualizations implicitly alias together uncertain values. For instance, if uncertainty is encoded by transparency, a maximally uncertain glyph may be entirely transparent, and so impossible to distinguish from any other maximally uncertain glyph. Other channels with a semiotic connection to uncertainty, such as saturation, value, blur, or size, also have deleterious effects on the disambiguation of colors and shapes. In both cases, the property of aliasing is *ad hoc*, and places no guarantees on the discriminability of colors. The binning and degradation approach of VSUMs makes the choice to alias values explicit to both the designer and the viewer, and results in a bivariate mapping with known perceptual properties.

## Design Considerations

There are multiple choices that designers must make before creating a VSUM. In particular, they must choose 1) which visual channels map to value and uncertainty, 2) a threshold of perceptual distinguishability (and, implicitly, a corresponding metric), and 3) a degradation function to determine how quickly uncertain values are aliased.

Often, the graphical conventions of a chart reserve variables such as position and area for other uses, and determine the visual channel to be used for value. In heatmaps, for instance,

value is regularly encoded using color, while vertical and horizontal position are reserved for location. In this paper, we primarily deal with the case where value is encoded using color. While there are some similar conventions for representing uncertainty, generally designers have more freedom in choosing a visual channel for uncertainty [24]. We recommend channels with both a strong semiotic connection to uncertainty as well as a relatively large number of perceptually distinguishable levels. Here, we focus on value/saturation (or, equivalently, transparency against a white background) and size. Correll & Gleicher [11] show that even audiences without statistical backgrounds can interpret uncertainty information encoded in these channels.

Another reason for limiting our discussion of bivariate mappings to color and size channels is that the perceptual distinguishability of these channels has been well-studied by prior work. CIELAB and other color spaces are designed to approximate perceptual uniformity—that is, distance in the color space should universally correspond with perceptual distance. Recent work [37, 38] has investigated the discriminability of colors in such spaces not only under laboratory conditions, but also among the disparate display conditions and color vision proficiency of crowdworkers. Similarly, formative work by Cleveland & McGill [8] assesses the discriminability of marks of different height, and this too has been extended both to a larger population of crowdworkers, and to the area channel specifically [20, 39].

We use values from these works as benchmarks for selecting thresholds of discriminability. The JNDs of these channels provides a useful minimum; in practice, we would wish to create maps with marks that are far more reliably distinguishable than the 50% or 75% discriminability rate of a JND. Figure 3 shows how these benchmarks can be used to select reasonable thresholds. Other factors, such as nameability and aesthetic preference, can be added in as additional weights impacting the desirability of certain color palettes, as in Gramazio et al. [16].

How quickly the number of bins decreases is another parameter of VSUMs. If this degradation is too fast, then we may only have a few discriminable levels of uncertainty. If this degradation is too slow, then we can be overwhelmed with colors. For instance, if we only decrement by one bin per level of uncertainty, there would be 1,485 different colors in a VSUM with a tolerance threshold of 1 unit in CIELAB space, far too much complexity for a discrete color mapping. In the examples in this paper, we halve the number of bins at each increasing level of uncertainty. When the number of initial bins at the highly certain region is a power of 2 (as in Fig. 1), this imparts a tree structure to the resulting VSUM. As uncertainty increases, bins in the previous level map to one and only one “parent” in the next uncertainty level, resulting in fewer crossings of color boundaries among rows.

## Examples

We present a set of examples showing how VSUMs can be applied to realworld datasets. The decision of VSUMs to allocate color bins asymmetrically amongst value and uncertainty regions affords greater discriminability in regions of the chart

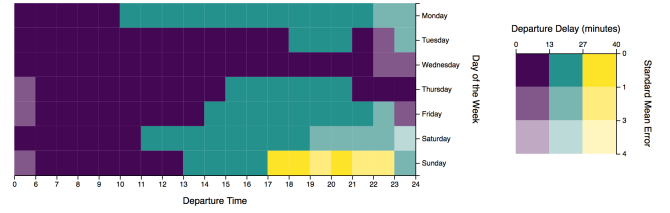


Figure 4: Average departure delay for different times of the day and days of the week visualized with a 2D uncertainty map. Horizontal position is the hour of scheduled departure, and vertical position is the day of the week.

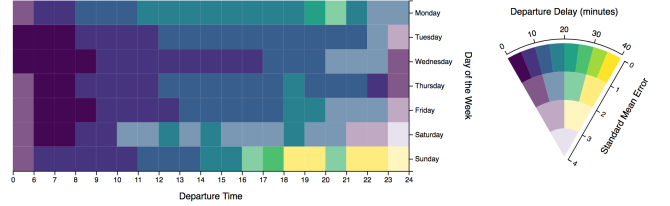


Figure 5: Flight delay data encoded with a VSUM.

that warrant closer scrutiny, while discouraging exploration of regions with noisy or unreliable signals.

## Air Travel Delay

Using a public data set from the U.S. Bureau of Transportation Statistics [31], we created a heatmap showing average flight delay for different times of the day and days of the week for U.S. carriers in January 2017. We use standard error ( $\sigma/\sqrt{n}$ ) as our measure of uncertainty. Starting from a minimum discriminability threshold of 18 units in CIELAB space, we created two alternate heatmaps of this delay information. Figure 4 shows a traditional bivariate map, whereas Figure 5 shows a VSUM generated under the same constraints.

Both maps illustrate a similar temporal trend: flight delays are shorter towards the beginning of the day, increasing on average over time. The traditional bivariate map affords only a coarse examination of this trend: two visible “blocks” of color, early in the day, and later in the afternoon. The VSUM, in contrast, has sufficient color resolution to show a quasi-linear trend. The additional color resolution for uncertainty information in VSUMs also promotes caution when considering outliers: the small number of flights on Saturdays, or departing late at night, make standard error for these delays relatively high, and the connected values more suspect than the rest of the chart.

Using a dataset from a group of virologists interested in the population dynamics of viral mutation [30], we visualized the variability in viral genomes across 25 populations of the simian immunodeficiency virus (SIV). Next generation sequencing (NGS) allows biologists to collect and analyze large amounts of genomics data. However, these techniques often create data quality issues, as they must align large numbers of potentially ambiguous “reads” of relatively small numbers of base pairs. In the dataset here, it is important not only to identify hotspots in the viral genome (locations with high rates of mutation or variability), but to not be distracted by false positives. The traditional bivariate map (Figure 6) affords the analysis of



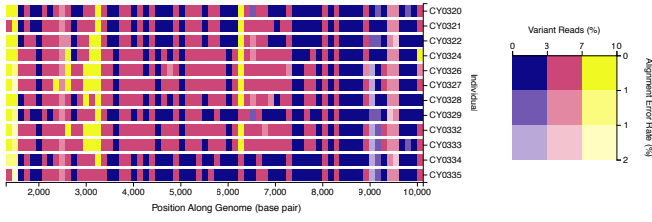


Figure 6: Variability for 12 different viral populations of the SIV virus, encoded with a traditional 2D bivariate map. Horizontal position denotes location along the SIV genome. Each row is the viral population of a different infected animal. Error in aligning short reads can lead to poor quality data, and so the amount of these errors is encoded as the uncertainty.

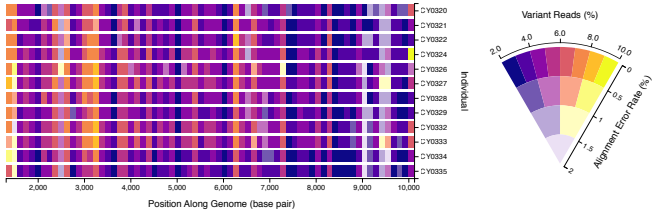


Figure 7: Variability for 25 different viral populations of SIV, encoded with a VSUM. The magnitude of mutation rate, as well as the generally poorer quality of data towards the end of the sequence, is more readily visible than with the traditional bivariate color map.

only a few, large hot spots. Extending the “confidence fog” metaphor used in viral genome browsers like LayerCake [9], VSUMs (Figure 7) enable biologists to explore interesting patterns while not being distracted by noise. Smaller scale “hot spots” that are mapped to the same magenta color in the traditional map are visible as bright orange lines. Pink hot spots towards the end of the genome (around reference nucleotide 9,500) that could be plausibly considered potential regions of interest in the 2D map, are encoded as highly uncertain and of ambiguous value in the VSUMs. In fact, these regions have systematically low coverage; the lack of reliable data for these regions should discourage analysts from giving their apparent pattern equal consideration from the stronger signals earlier in the genome.

## EVALUATION

We performed a crowdsourced experiment on Amazon’s Mechanical Turk to evaluate the effectiveness of VSUMs for integrating uncertainty and value information in visualizations. This focus on integration meant that we limited our experimental tasks to scenarios where the participants needed to consider both value and uncertainty before making a decision. We gave participants two main tasks:

1. An **identification** task, where we gave participants charts with value and uncertainty information, and asked them to locate specific regions. E.g., “click on the region of the chart with highest uncertainty *and* lowest value.”
2. A **prediction** task, inspired by the game Battleship, where we gave participants charts with both *forecast* and *forecast*

*uncertainty* information, and asked the participants to place tokens on the board in order to optimize expected value. E.g., “place your 4 ships on safe locations on the board.”

We discuss the two experimental tasks in more detail in the following sections. Experimental materials, including data tables, are available at [URL-REMOVED-FOR-REVIEW](#).

For each task, participants saw one of three types of heatmaps. Either two **juxtaposed** maps side-by-side, with separate encodings for value and uncertainty, a traditional **2D bivariate** map, where value and uncertainty are orthogonally encoded, or a **VSUM**. Figure 8 shows examples of these factor levels. Other factors included the size of the map (either a 4x4 or 8x8 grid), and (for the prediction task), the question framing (attack or defense). Secondary factors were evenly blocked across stimuli, resulting in a 3x2 within subjects design for the identification task, and a 3x2x2 within subjects design for the prediction task. Participants reported fatigue effects in piloting, and so we limited the number of replicates in this study: participants saw 18 stimuli (6 for each map type) for the identification task, and 12 (4 for each map type) for the prediction task.

In all cases, we generated color maps based on the requirement that no two colors in the resulting color scheme were closer than 18 units of Euclidean distance in CIELAB space. This threshold results in a 9-color bivariate map and 15-color VSUM for all color scales used in this study. Given the size of the marks used in our study, this threshold distance is sufficient for a predicted discriminability rate of over 95% [37]. Other choices of threshold, and differing univariate scales, would result in different maps (see Fig. 3).

## Identification Task

For the identification task, we gave participants bivariate representations of uncertainty, and asked them to identify a location meeting a given criteria, e.g., “Click on the location with the greatest uncertainty.” There were three potential criteria, mapping to the three vertices of a VSUM: 1) high value, low uncertainty, 2) low value, low uncertainty, and 3) high uncertainty. Each map contained four locations, one of each combination of maximum and minimum value and uncertainty. The correct answer was always one of these locations. Other locations in the grid were created pseudo-randomly, with less extreme ranges of uncertainty and probability. We measured performance both in terms of accuracy (did the participant select the correct point) as well as response time.

## Prediction Task

For the prediction task, we gave participants the rules of a game like Battleship. Greis et al. [17] employ these game-like experimental tasks to assess how different visual designs communicate uncertainty information, which can be abstract or complex, to the general audience. In our task, the participant and a (fictional) adversary have to place tokens representing ships on a spatial grid. Without direct access to the location of the opponent’s ships, the player then has to place tokens representing missile strikes on their opponent’s grid. Players have to place all their tokens before continuing. If a player fires a missile at a square containing a ship, then that ship

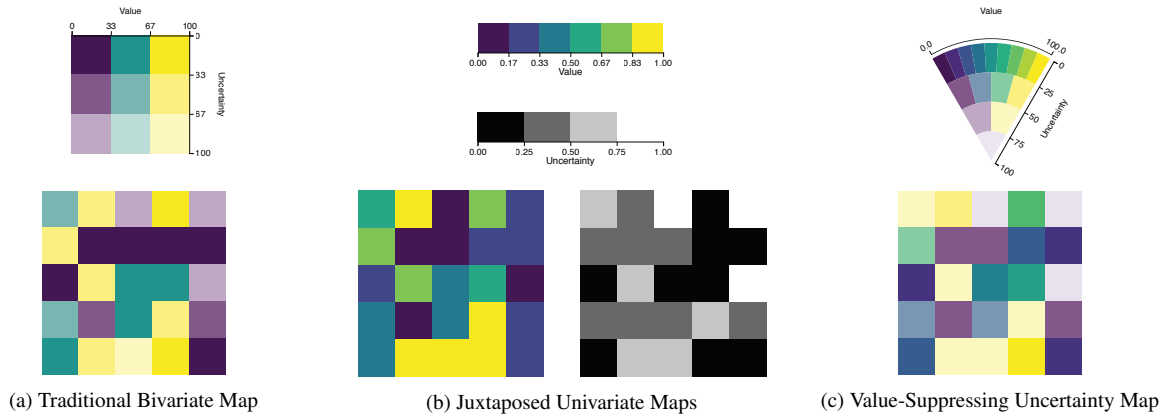


Figure 8: The three graph types in our evaluation. Bivariate color ramps ensure that there are orthogonal mappings for each combination of value and uncertainty, but, since color channels are not separable, can afford only a few discrete colors before color categories become unacceptably close, perceptually. Juxtaposed maps, by keeping these channels separate, afford a larger range of colors, but require a visual search task to recover both value and uncertainty in a region. VSUMs have many of the advantages of bivariate maps, but assign more color categories to certain data, at the expense of ambiguity when uncertainty is high.

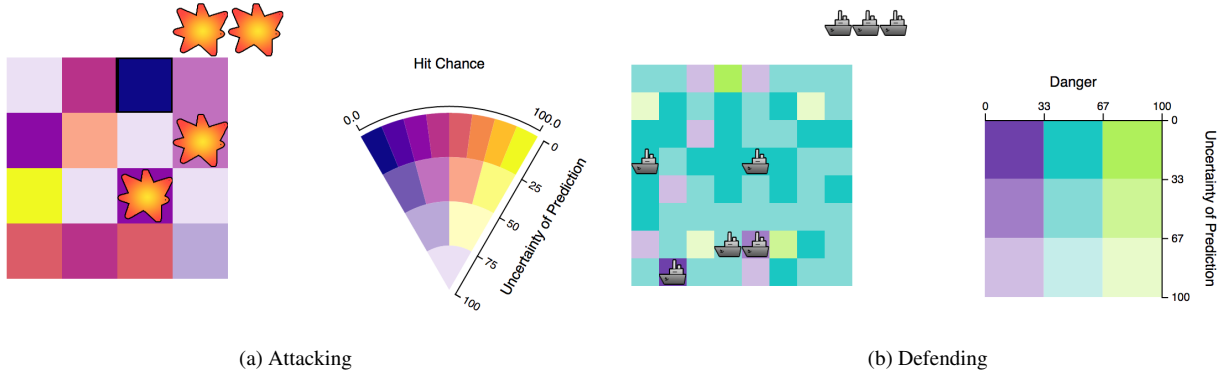


Figure 9: The two framings of the Prediction task. In the “attack” framing, the participant is given a map of predictions of the location of enemy ships, along with the uncertainty in those predictions. The participant should place their missile strikes on locations with a high probability of containing a ship, and with high certainty in this probability. The “defend” framing is the opposite task: the participant has a list of likely missile locations, and ought to place their ships on locations with low probability of attack, and high certainty in this probability.

is sunk. The objective is to maximize the number of enemy ships hit, while minimizing the number of your own ships that are hit. In our task, participants were given a chart encoding predictions of the location of enemy ships (in the attacking case) or potential missile strikes (in the defending case). The *value* component of the prediction was the likelihood of an item being present in a particular region. The *uncertainty* component was the confidence in this prediction.

We included both a **gain** framing (the attacker’s task: given predictions of where your opponent’s ships are, where should you launch your missiles?), and a **loss** framing (the defender’s task: given predictions of where your opponent’s missiles will fall, where should you place your ships?) to account for systematic differences in human judgments under uncertainty. Tversky & Kahneman [44] illustrate that framings in terms of gains or losses produce reliably different outcomes. In general, the function of the value of gains is concave, but is

convex for losses [22]. That is, avoiding a large loss is often more valuable than striving for a similarly large gain. To avoid the potential for (incorrect) transfer learning, the attacking and defending tasks used different value color maps (Plasma and Cool) than the identification task, and from the opposite framing in this task. Figure 9 shows these differing color maps and framings.

Each map contained a minimum number of extreme locations with each combination of very high and very low value and uncertainty. High expected value, high uncertainty locations were “risky.” As a balance to the risky locations, there were also an equal number of “safer” locations, with slightly less expected value, but also less uncertainty. The remaining locations of the map varied pseudo-randomly in the middle ranges.

The ideal strategy from a value-maximizing standpoint would be to place tokens on areas with the highest expected value, ig-

noring the uncertainty information. However, as with roulette and other similar games of chance, the variability in expected value is relevant when considering where to place bets [28]. Over the short term, a player may seek out high expected value but risky locations, or, if risk averse, accept lower expected value but lower variability locations. We therefore measured the distribution of both *value* and *uncertainty* of the tokens placed by the participants.

### Hypotheses

Across both the identification and prediction tasks, we had several hypotheses, stemming from our belief that VSUMs promote better *integration* between uncertainty and value information, and also encourage *caution* by highlighting the ambiguity or untrustworthiness introduced by uncertain data. In particular:

1. Participants would be **faster and more accurate** when completing the identification task using a VSUM as compared to the other map types. Juxtaposed maps, by adding an additional search task to the lookup task, would have especially poor performance compared to VSUMs.
2. Participants would choose targets with **less uncertainty** in the prediction task when using a VSUM compared to the other maps types. This would result in a tradeoff where they would also choose targets with **less expected value** than the other map types. Again, juxtaposed maps, by increasing the difficulty of retrieving both variables simultaneously, would highlight this difference.

### Results

Rather than standard means, Cleveland & McGill [8] use 25% trimmed means to report effect sizes in graphical perception tasks. Trimmed means remove the first and last quartiles and then compute the mean from the remaining points. This filtering is important for graphical perception tasks (especially crowdsourced tasks) where error distributions may be long-tailed [20]. Trimmed means violate the sampling assumptions required for many null-hypothesis significance tests. Thus, while we report effect sizes in terms of trimmed means, we perform inferential statistical tests on the standard means. Asterisks in our charts denote significant differences, as the reported by a post-hoc Tukey’s test of Honest Significant Difference (HSD).

#### Participants

We limited our population to Turkers from the United States, with a prior task approval rate of at least 90%. As the experimental tasks required multi-hue color perception, we presented participants with a set of Ishihara plates [19] as a pretest, and excluded participants who either misidentified the values in the plates, or who self-reported as having a color vision deficiency (CVD) in the post-test. Based on piloting, we paid participants \$2 dollars for participation, for a target rate of \$8/hour. After completion of the main tasks, we solicited demographic information. We recruited a total of 24 participants for this experiment: 4 female, 19 male, and 1 who declined to state ( $M_{\text{age}} = 32$ ,  $SD_{\text{age}} = 7.5$ ). All reported having either college degrees, or some college experience. Average self-reported experience with interpreting graphs and charts was low (1.9

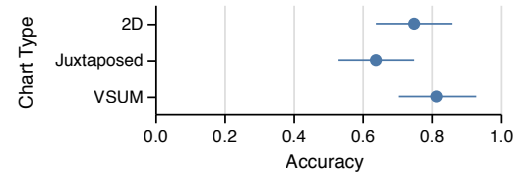


Figure 10: The effect of chart type on accuracy for the identification task. VSUMs are similarly accurate for simple search tasks as more traditional bivariate visualizations, despite containing additional colors and categories, and having a more complex relationship between uncertainty, value, and color than traditional charts. The confidence intervals are bootstrapped 95% CIs of trimmed means.

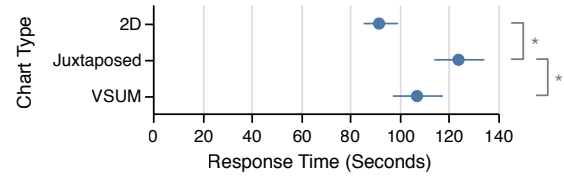


Figure 11: The effect of chart type on response time for the identification task. Juxtaposed maps of value and uncertainty introduce a visual search task, increasing the time to complete even simple tasks involving both value and uncertainty information. The confidence intervals are bootstrapped 95% CIs of trimmed means.

on a 5 point Likert scale, with 1 indicating “No experience”), and no participant reported experience greater than 3 out of 5.

#### Identification Task

For the identification task, our primary measures of performance were accuracy and response time. Our hypothesis was that VSUMs would exhibit better performance than the other charts.

We performed a one-way ANOVA investigating the effect of chart type (juxtaposed, full bivariate, or VSUM) on accuracy. We did not find a significant effect ( $F(2, 859) = 2.4$ ,  $p = 0.09$ ), although VSUMs had the highest accuracy. Figure 10 illustrates this result.

We also performed a one-way ANOVA on the effect of chart type on response time. In this case, we found a significant main effect ( $(F, 2, 859) = 5.0$ ,  $p < 0.001$ ). A Tukey’s HSD found that participants were significantly slower at performing the task on juxtaposed maps, as opposed to the other map types. Figure 11 illustrates this result.

These results only partially support our hypothesis. There was no significant difference in terms of accuracy between different maps, partially due to the simplicity of the identification task: participants only had to find a single target in a heatmap, with extreme values in both data and uncertainty axes. However, our results do provide evidence that juxtaposed maps, by introducing an additional search task to the problem of assessing the value and uncertainty of a location, require significantly more time to interpret.

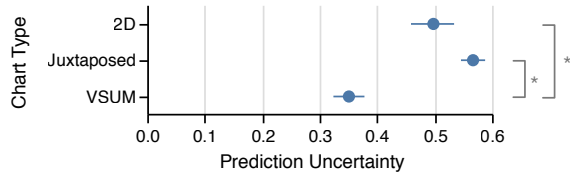


Figure 12: The effect of chart type on the average uncertainty in predictions for the prediction task. By reducing the number of color categories as uncertainty increases, VSUMs encourage more caution in predictions, making people less likely to consider data with strong, but spurious patterns. The confidence intervals are bootstrapped 95% CIs of trimmed means.

### Prediction Task

For the prediction task, where there is no dominating strategy, our primary measure was the distribution of guesses, in terms of both their average value, and their average uncertainty. A risky player would choose guesses with high value, regardless of the uncertainty of those points. A more conservative guesser might eschew high-risk, high-reward locations, resulting in a lower average value of guesses, but also lower uncertainty. Our hypothesis is that VSUMs, by communicating uncertainty information at the expense of value information, would promote more conservative guesses than the other chart types.

We performed a one-way ANOVA of chart type on average value of guesses. We found no significant effect ( $F(2, 861) = 0.27, p = 0.8$ ), indicating that participants were generating similar expected values from their guesses across conditions.

However, an ANOVA of chart type on average *uncertainty* of guesses did find a significant main effect ( $F(2, 861) = 61, p < 0.001$ ). A post-hoc Tukey’s HSD confirmed that participants had guesses with significantly less uncertainty with VSUMs than with both other chart types. Additionally, participants’ guesses on juxtaposed charts were significantly riskier than with the other chart types. Figure 12 illustrates this result.

As with the identification task, a one-way ANOVA also indicated that chart type had a significant effect on response time ( $F(2, 861) = 13, p < 0.001$ ). A Tukey’s HSD confirmed that, as with the previous task, response time was significantly longer with juxtaposed maps (trimmed mean response time of 23 seconds) compared to the other chart types (both with trimmed mean response time of 18 seconds).

Despite the documented effect of framing in prior work, a one-way ANOVA of task frame on expected value found no significant main effect ( $F(1, 862) = 0.01, p = 0.9$ ), and the trimmed means were quite similar (expected value of 0.76 per guess for attackers, versus 0.75 per guess for defenders).

Our results again partially support our hypotheses. We did not observe statistically different patterns of *expected value*, but did find that VSUMs produce different patterns of *acceptable uncertainty*. That is, VSUMs promote avoidance of uncertain values, without notably impacting the expected value. Juxtaposed maps, by contrast, make the retrieval of uncertainty information significantly more difficult; it is possible that participants ignore or downweight it when making predictions,

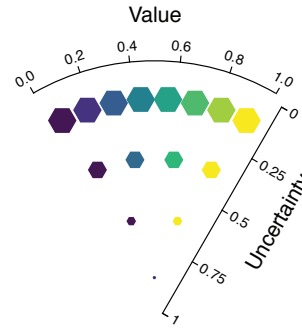


Figure 13: A VSUM where value is mapped to color, and uncertainty is mapped to glyph area. As glyphs become smaller, it becomes more difficult to distinguish their colors [37]. VSUMs, by reducing the number of color categories as uncertainty increases, account for this ambiguity. On the right, the VSUM has been used to encode a 5x5 grid of random data.

resulting in riskier predictions overall. By making the consideration of uncertainty values integral to the design of the chart, VSUMs appear to promote more cautious decisions.

### DISCUSSION

The results from our initial evaluation of VSUMs are promising, showing that even a general audience can make use of uncertainty information in a reasonable way. The way that uncertainty information is presented can have a measurable impact on decision-making. The VSUM strategy of only “showing” differences when uncertainty is sufficiently low is somewhat analogous to the inferential statistics such as effects tests. VSUMs might therefore be used to promote more caution in judgments, and lead analysts away from spurious signals in data.

Bivariate maps have an inherent limitation in that visual channels are often difficult to attend to separately or orthogonally. Bivariate maps are also complex to interpret. Systems which simultaneously display large amounts of value and uncertainty data to general audiences (such as Pangloss [29]) avoid them for precisely these reasons, relying instead on juxtaposed maps. Our results indicate that juxtaposition produces additional problems, forcing users that wish to integrate uncertainty and data to locate an item of interest, and then re-locate that same item in a visually distant map, which may have few visual landmarks in common. This additional step before information fusion is reflected in longer response times, and patterns of decision-making that are less mindful of risk. Given the deleterious effects of latency on data analysis [23], designers ought to carefully consider the tradeoffs between expressiveness of visual encoding, and usability.

### Limitations and Future Work

In general, VSUMs excel in scenarios where we wish to make firm guarantees about the discriminability of different marks in a chart, but must also represent, in as much detail as possible, two, potentially orthogonal variables of interest. VSUMs have improved data resolution over naïve bivariate maps by



assuming that uncertain values ought to have less visual impact than highly certain values. VSUMs therefore act both as a kind of filtering device as well as a bivariate representation. This assumption holds in scenarios like the virology dataset (Figure 7), where the expectation is that highly uncertain data is unreliable or should otherwise be downweighted in analysis. If the analyst has a different interpretation of uncertainty, or wishes to quickly and orthogonally analyze the distributions of uncertainty and value, other strategies, such as juxtaposed maps, may be more appropriate. VSUMs are designed for the *integration* of value and uncertainty. Designers should take care in considering when and how this integration is desirable.

Our experiments dealt with cases where both uncertainty and value were represented by color. The perceptual non-separability of color channels is well-known [15, 48], and so the concept of a limited “budget” of distinguishable marks easier to quantify and illustrate. In principle, a VSUM can be created for any combination of visual variables. All that is required is a perceptual model of the interaction between these two variables. Where these models exist, as with the interaction between size and color [37], the creation of VSUMs is straightforward. Where these models do not exist, or where the perceptual interaction is too complex to efficiently model, experimental work remains to be done before VSUMs can be considered a feasible design strategy. We are currently experimenting with the creation of VSUMs using these other channels, as in Figure 13.

## Conclusion

Often, uncertainty, data quality, or confidence are considered separate information from the data itself, relegated to tooltips or visually distant supplemental charts. We believe, in contrast, that uncertainty information ought to be directly integrated with charts of data itself. This integration introduces additional complexity in the design and presentation of data. Value-Suppressing Uncertainty Maps represent one strategy for dealing with this complexity, by assigning marks properties in a way that supports the disambiguation of values in data where uncertainty is low, but suppresses these judgments when uncertainty is high. This decision of how to allocate visual variables promotes patterns of decision-making that make responsible use of uncertainty information, discouraging comparison of values in unreliable regions of the data, and promoting comparison in regions of high certainty.

## REFERENCES

1. Jürgen Bernard, Martin Steiger, Sebastian Mittelstädt, Simon Thum, Daniel Keim, and Jörn Kohlhammer. 2015. A survey and task-based quality assessment of static 2D colormaps. In *SPIE/IS&T Electronic Imaging*. International Society for Optics and Photonics, 93970M–93970M.
2. David Borland and Russell M Taylor. 2007. Rainbow color map (still) considered harmful. *IEEE computer graphics and applications* 27, 2 (2007).
3. Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D3: Data-Driven Documents. *IEEE transactions on visualization and computer graphics* 17, 2 (2011), 2301–2309.
4. 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.
5. Nadia Boukhelifa and David John Duke. 2009. Uncertainty visualization: why might it fail?. In *CHI’09 Extended Abstracts on Human Factors in Computing Systems*. ACM, 4051–4056.
6. Ken Brodrie, Rodolfo Allendes Osorio, and Adriano Lopes. 2012. A review of uncertainty in data visualization. In *Expanding the frontiers of visual analytics and visualization*. Springer, 81–109.
7. Barbara Buttenfield and M Kate Beard. 1994. Graphical and geographical components of data quality. *Visualization in geographic information systems* (1994), 150–157.
8. William S Cleveland and Robert McGill. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association* 79, 387 (1984), 531–554.
9. Michael Correll, Adam L Bailey, Alper Sarikaya, David H O’Connor, and Michael Gleicher. 2015. LayerCake: a tool for the visual comparison of viral deep sequencing data. *Bioinformatics* (2015), btv407.
10. Michael Correll, Subhadip Ghosh, David O’Connor, and Michael Gleicher. 2011. Visualizing virus population variability from next generation sequencing data. In *Biological Data Visualization (BioVis), 2011 IEEE Symposium on*. IEEE, 135–142.
11. M Correll and Michael Gleicher. 2013. Error bars considered harmful. *IEEE Visualization Poster Proceedings*. IEEE (2013).
12. Michael Correll and Jeffrey Heer. 2017. Surprise! Bayesian Weighting for De-Biasing Thematic Maps. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 651–660.
13. Çağatay Demiralp, Michael S Bernstein, and Jeffrey Heer. 2014. Learning perceptual kernels for visualization design. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 1933–1942.
14. Richard Dunn. 1989. A dynamic approach to two-variable color mapping. *The American Statistician* 43, 4 (1989), 245–252.
15. W. R. Garner and G. L. Felfoldy. 1970. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology* 1, 3 (1970), 225–241.
16. Connor C Gramazio, David H Laidlaw, and Karen B Schloss. 2017. Colorgical: Creating discriminable and preferable color palettes for information visualization. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 521–530.

17. Miriam Greis, Passant El Agroudy, Hendrik Schuff, Tonja Machulla, and Albrecht Schmidt. 2016. Decision-Making under Uncertainty: How the Amount of Presented Uncertainty Influences User Behavior. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. ACM, 52.
18. Henning Griethe, Heidrun Schumann, and others. 2006. The visualization of uncertain data: Methods and problems.. In *SimVis*. 143–156.
19. LeGrand H Hardy, Gertrude Rand, and M Catherine Rittler. 1945. Tests for the detection and analysis of color-blindness. I. The Ishihara test: an evaluation. *JOSA* 35, 4 (1945), 268–275.
20. Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 203–212.
21. Jessica Hullman, Paul Resnick, and Eytan Adar. 2015. Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PloS one* 10, 11 (2015), e0142444.
22. Daniel Kahneman and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society* (1979), 263–291.
23. Zhicheng Liu and Jeffrey Heer. 2014. The effects of interactive latency on exploratory visual analysis. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 2122–2131.
24. Alan M MacEachren. 1992. Visualizing uncertain information. *Cartographic Perspectives* 13 (1992), 10–19.
25. 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.
26. Jock Mackinlay. 1986. Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)* 5, 2 (1986), 110–141.
27. Marc Mahy, L Eycken, and André Oosterlinck. 1994. Evaluation of uniform color spaces developed after the adoption of CIELAB and CIELUV. *Color Research & Application* 19, 2 (1994), 105–121.
28. Leonard Mlodinow. 2009. *The drunkard's walk: How randomness rules our lives*. Vintage. 81–104 pages.
29. Dominik Moritz, Danyel Fisher, Bolin Ding, and Chi Wang. 2017. Trust, but Verify: Optimistic Visualizations of Approximate Queries for Exploring Big Data. In *ACM Human Factors in Computing Systems (CHI)*. <http://idl.cs.washington.edu/papers/trust-but-verify>
30. Shelby L O'Connor, Ericka A Becker, Jason T Weinfurter, Emily N Chin, Melisa L Budde, Emma Gostick, Michael Correll, Michael Gleicher, Austin L Hughes, David A Price, and others. 2012. Conditional CD8+ T cell escape during acute simian immunodeficiency virus infection. *Journal of virology* 86, 1 (2012), 605–609.
31. Bureau of Transportation Statistics. 2016. Airline Delays, Cancellations and Tarmac Times. (2016). [http://www.rita.dot.gov/bts/data\\_and\\_statistics/by\\_mode/airline\\_and\\_airports/airline\\_delay.html](http://www.rita.dot.gov/bts/data_and_statistics/by_mode/airline_and_airports/airline_delay.html)
32. Lace Padilla, P Samuel Quinan, Miriah Meyer, and Sarah H Creem-Regehr. 2017. Evaluating the Impact of Binning 2D Scalar Fields. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 431–440.
33. Maria Riveiro. 2007. Evaluation of uncertainty visualization techniques for information fusion. In *Information Fusion, 2007 10th International Conference on*. IEEE, 1–8.
34. Philip K Robertson and John F O'Callaghan. 1986. The generation of color sequences for univariate and bivariate mapping. *IEEE Computer Graphics and Applications* 6, 2 (1986), 24–32.
35. Bernice E Rogowitz and Alan D Kalvin. 2001. The "Which Blair Project": a quick visual method for evaluating perceptual color maps. In *Visualization, 2001. VIS'01. Proceedings*. IEEE, 183–556.
36. Bernice E Rogowitz, Lloyd A Treinish, Steve Bryson, and others. 1996. How not to lie with visualization. *Computers in Physics* 10, 3 (1996), 268–273.
37. Maureen Stone, Danielle Albers Szafr, and Vidya Setlur. 2014. An engineering model for color difference as a function of size. In *Color and Imaging Conference*, Vol. 2014. Society for Imaging Science and Technology, 253–258.
38. Danielle Albers Szafr, Maureen Stone, and Michael Gleicher. 2014. Adapting color difference for design. In *Color and Imaging Conference*, Vol. 2014. Society for Imaging Science and Technology, 228–233.
39. Justin Talbot, Vidya Setlur, and Anushka Anand. 2014. Four experiments on the perception of bar charts. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 2152–2160.
40. Nassim Nicholas Taleb and Mark Blyth. 2011. The black swan of Cairo: How suppressing volatility makes the world less predictable and more dangerous. *Foreign Affairs* (2011), 33–39.
41. Judi Thomson, Elizabeth Hetzler, Alan MacEachren, Mark Gahegan, and Misha Pavel. 2005. A typology for visualizing uncertainty. In *Electronic Imaging 2005*. International Society for Optics and Photonics, 146–157.
42. Bruce E Trumbo. 1981. A theory for coloring bivariate statistical maps. *The American Statistician* 35, 4 (1981), 220–226.

43. Archie Tse. 2016. Why We Are Doing Fewer Interactives. (2016). <https://github.com/archietse/malofiej-2016/blob/master/tse-malofiej-2016-slides.pdf> Malofiej.
44. Amos Tversky and Daniel Kahneman. 1985. The framing of decisions and the psychology of choice. In *Environmental Impact Assessment, Technology Assessment, and Risk Analysis*. Springer, 107–129.
45. StÃfan van der Walt and Nathaniel Smith. 2015. mpl colormaps. <https://bids.github.io/colormap/>, (2015).
46. Howard Wainer and Carl M Francolini. 1980. An Empirical Inquiry Concerning Human Understanding of Two-Variable Color Maps. *American Statistician* (1980), 81–93.
47. Colin Ware. 1988. Color sequences for univariate maps: Theory, experiments and principles. *IEEE Computer Graphics and Applications* 8, 5 (1988), 41–49.
48. Colin Ware. 2012. *Information visualization: perception for design*. Elsevier.