Value-Suppressing Uncertainty Palettes

Michael Correll
Tableau Research
mcorrell@tableau.com

Dominik Moritz
University of Washington
domoritz@cs.washington.edu

Jeffrey Heer
University of Washington
jheer@uw.edu

ABSTRACT

Understanding uncertainty is critical for many analytical tasks. One common approach is to encode data values and uncertainty values independently, using two visual variables. These resulting bivariate maps can be difficult to interpret, and interference between visual channels can reduce the discriminability of marks. To address this issue, we contribute Value-Suppressing Uncertainty Palettes (VSUPs). VSUPs allocate larger ranges of a visual channel to data when uncertainty is low, and smaller ranges when uncertainty is high. This non-uniform budgeting of the visual channels makes more economical use of the limited visual encoding space when uncertainty is low, and encourages more cautious decision-making when uncertainty is high. We demonstrate several examples of VSUPs, and present a crowdsourced evaluation showing that, compared to traditional bivariate maps, VSUPs encourage people to more heavily weight uncertainty information in decision-making tasks.

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, analyzing, and presenting 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
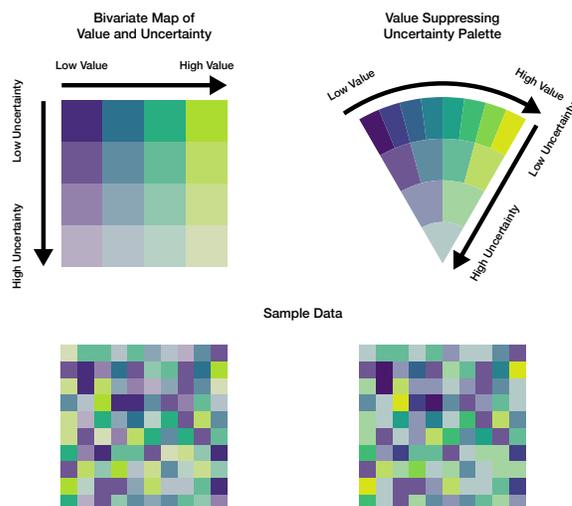
Figure 1: A standard bivariate map (left) and a VSUP (right), used to encode an identical 10x10 grid of random data. Both use the same visual channels to encode value (position along the Viridis [46] color map) and uncertainty (lightness and saturation). However, the VSUP uses a tree-like structure to allocate colors, defining more bins when uncertainty is low. This non-uniform budgeting affords better discrimination between values when uncertainty is low, even though the VSUP has fewer color bins (in this case, 15 to the bivariate map's 16). This tree-like structure also discourages analysis in regions where uncertainty may be unacceptably high.

mark properties. Due to the interference and interplay between different visual variables, bivariate maps may suffer from limited discriminability.

In this paper, we contribute **Value-Suppressing Uncertainty Palettes** (VSUPs) for integrating data and uncertainty information in visualizations. VSUPs intentionally alias together data values with high uncertainty, affording greater discriminability as uncertainty decreases. Traditional bivariate maps might be thought of as a 2D square, with differing outputs for each combination of value and uncertainty. In contrast, VSUPs can be conceptualized 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 data value. Figure 1 shows examples of both a traditional bivariate map and a VSUP.

We describe the motivations for VSUPs, provide examples of their utility for decision-making under uncertainty, and assess VSUPs in a crowdsourced experiment. Our results indicate

that VSUPs create close integration between uncertainty and data. This integration has a demonstrable effect on decision-making, leading viewers to make more cautious choices that avoid high uncertainty.

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. [42] and Buttenfield & 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.

Reviewing the state of the art in uncertainty visualization in the field, Greithe et al. [20] and Brodlie 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 [44]. 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 [24]), we limit the scope of our discussion to techniques which center on static charts.

Visual Variables for Uncertainty

One challenge of uncertainty visualization is that 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 increasingly difficult.

A further hurdle is that not all visual variables are well-suited for conveying uncertainty. MacEachren et al. [27] 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 not intuitively convey uncertainty.

Bivariate Maps

While uncertainty can be visualized entirely separately from the rest of the data (for instance, in a juxtaposed chart), this runs the risk that uncertainty information becomes *ignorable* [31]. It also complicates analysis, as interpreting the value and uncertainty of a given data point requires consulting two separate charts. To integrate uncertainty information into an analysis, we focus on the construction of *bivariate maps*, where value and uncertainty information is displayed simultaneously in a spatially co-located manner.

Bivariate maps can, in principle, be constructed from the combination of any two visual variables (such as shape and color, or size and texture). For example Ware's "textons" [49] overlay glyphs on top of regions to simultaneously encode two quantitative values. However, 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. Bivariate maps in these settings therefore typically rely on a visual channel related to color for encoding a secondary variable, such as opacity [39] or pixel noise [26].

Colors in univariate quantitative color maps should be sufficiently far apart as to be perceptually distinguishable [48], and vary in lightness as well as hue to afford an implicit ordering of value [2, 37]. Different choices of color maps can highlight different features of the data, and should be chosen with care [38]. These principles extend to bivariate color maps, with the additional consideration that our perception of color channels lacks *orthogonality*. That is, while we can perceive e.g., the height of a bar in a bar chart independently of its width, a particular property of a color (e.g., its redness, saturation, transparency, etc.) affect how its other properties are perceived [15, 50]. Many color channels such as hue and saturation are therefore perceptually *integral*, which complicates their use in bivariate maps.

Another recurring design consideration when constructing color maps is whether to quantize data into a discrete set of colors, or to encode data using a continuous mapping. While continuous color maps afford greater fidelity in presenting values [32], non-linearity in human color perception introduces errors in extracting numeric values from continuous colors [2]. Quantizing a color map is therefore an exercise in balancing *perceptual* error and *quantization* error [13]. Discrete maps offer finer control over this balance, which can result in better performance in tasks involving heatmaps [34].

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: due to their additional complexity, Wainer & Francolini [47] reported high levels of error for participants even for "elementary" graph reading tasks, and additionally that bivariate legends were more difficult to memorize and internalize than their univariate equivalents. Therefore, bivariate maps in practice are often limited to a small set of output colors [36, 43] (say, a 4x4 matrix as in Figure 1).

Given that bivariate maps, for reasons of either practicality or perceptual fidelity, have only a limited budget of outputs, we adapt an insight from Dunn [14] that these limited cate-

gories ought to be assigned with regards to their importance to analysis tasks, rather than uniformly. Our Value-Suppressing Uncertainty Palettes apply this insight to data with uncertainty. Frequently, uncertain information ought to be weighted less heavily in a decision-making process, or have less prominence in an analysis, than information with high certainty.

Correll et al. [8, 9] use a precursor of VSUPs in their LayerCake genomics visualization tool, where marks representing 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 a "confidence fog" while highly certain values remain prominent. Other bivariate visualizations implicitly alias together uncertain values through perceptual integrality. 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, have similar 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 VSUPs makes the choice to alias values explicit to both the designer and the viewer, and results in a bivariate mapping with known perceptual properties.

VALUE-SUPPRESSING UNCERTAINTY PALETTES

Value-Suppressing Uncertainty Palettes (VSUPs) are a technique for creating bivariate maps of data *value* and *uncertainty*. VSUPs make two central assumptions about bivariate maps:

1. There is a limited *budget* for perceptual discriminability.

2. Differences among *certain* data are more germane than differences among *uncertain* data.

In some cases, these assumptions are 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 [41]. 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 variability of data values [35].

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. VSUPs codify this decision by *reducing the number of mark categories for representing value as uncertainty increases.* This means that data encoded using a VSUP will make visible only the largest of differences in uncertain data, but highlight comparatively small differences in value when uncertainty is low. VSUPs therefore act as both a filtering mechanism (in that values with too much uncertainty are 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 low quality or highly uncertain 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 [11].

VSUPs rely on an underlying **quantization tree** that governs how values are discretized. Values above a certain uncertainty threshold are mapped to a singular "root" node. As uncertainty decreases, the tree branches into leaves, which evenly divide the data domain. The data value of a parent is the midpoint of all of its children. These leaves can then branch again, up to a designer-specified stopping point. The layers quantize the uncertainty domain, and sibling nodes quantize the data domain. Figure 1 shows a VSUP with a tree with a branching factor of 2 and 4 layers, resulting in an output range with 15 distinct colors.

Since the visual channel used to encode uncertainty (e.g., saturation, lightness, transparency) often reduces the ability of people to distinguish colors, VSUPs have the added benefit of creating bivariate palettes that are perceptually easier to distinguish, by reducing color resolution in precisely the areas where color disambiguation would be most difficult anyway. For instance, the two closest colors in the VSUP in Figure 1 are 1.4 units farther apart in CIELAB color space than the two closest colors in the traditional bivariate map.

JavaScript code for generating VSUPs for use in D3 [3] charts is available at **https://github.com/uwdata/vsup**.

Design Considerations

There are multiple choices that designers must make before creating a VSUP. Specifically, they must choose which visual channels map to value (1), uncertainty (2), and the structure of the underlying VSUP tree (3).

When selecting a visual channel to represent uncertainty, we recommend channels with both a strong semiotic connection to uncertainty (such as those recommended by MacEachren et al. [27]) as well as a relatively large number of perceptually distinguishable levels (such as those evaluated by Boukhelifa et al. [4]). Here, we focus on a combination of increasing luminance and decreasing saturation, as in Hengl & Toomanian [23]. Correll & Gleicher [10] provide evidence that even audiences without statistical training can successfully interpret uncertainty information encoded in these channels.

When selecting a visual channel to represent value, the designer should consider the perceptual integrality of the value and uncertainty visual channels. For instance, we use the Viridis [46] color map frequently in this paper, as many standard color ramps (such as sequential ColorBrewer [22] ramps), following usual best practices, interpolate in both hue and luminance [48]. This interpolation in luminance interferes with our uncertainty encoding, introducing ambiguity. Ramps such as Viridis, which avoid very light and very dark colors, reduce this ambiguity.

When determining the quantization tree, the design should be careful to avoid having too many output colors. For reasons of memorability and complexity, Wainer & Francolini [47] suggest no more than 16 distinct outputs in bivariate maps. For a binary quantization tree, this results in a maximum tree depth of 4, as seen in Figure 1. In general, for tasks that require
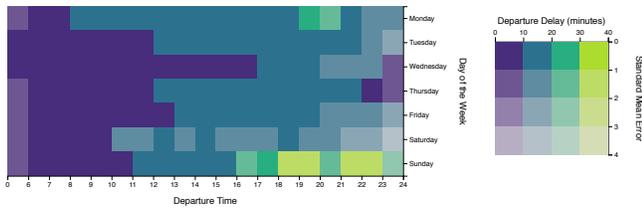
Figure 2: Average departure delay for different times of the day and days of the week visualized with a 2D uncertainty map. Horizontal position is the hour of scheduled departure, and vertical position is the day of the week.
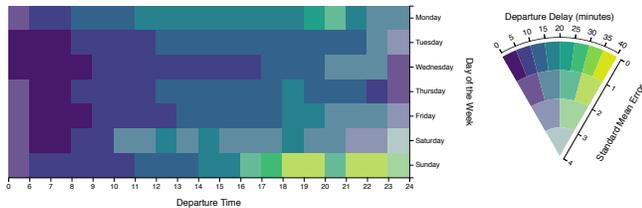

Figure 3: Flight delay data encoded with a VSUP.

finer-grade distinctions between uncertainty, the depth of the tree could be increased. For tasks involving finer-grade distinctions between value, the branching factor of the tree could be increased. However, these factors increase the number of colors, and so designers should moderate between legend complexity and data fidelity. Beyond these two parameters, non-uniform binning would allow the designer to target particular distributions or important subregions of the data that would otherwise be difficult to visualize in a limited palette, as in Dunn's [14] dynamic bivariate maps. As an additional form of control, prior tools using VSUP-like techniques, such as LayerCake [8], allow the analyst to control the bins in the tree through interaction with the color legend.

EXAMPLES

We present a set of examples showing how VSUPs can be applied to real-world datasets. VSUPs allocate color bins asymmetrically amongst value and uncertainty regions, affording greater discriminability in regions of the chart 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 [33], 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. We created two alternate heatmaps of this delay information. Figure 2 shows a traditional bivariate map, whereas Figure 3 shows a VSUP generated under the same constraints.

Both maps illustrate a similar 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 VSUP, in contrast,
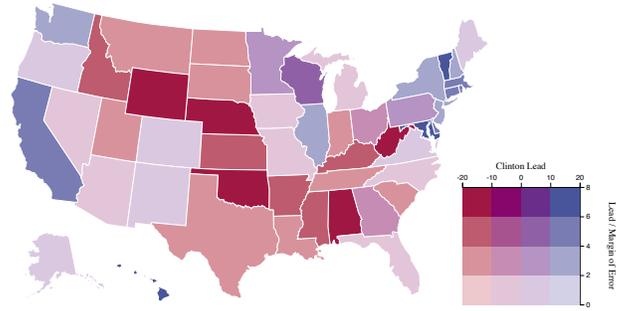

Figure 4: Polling data prior to the 2016 U.S. Presidential Election, encoded in a traditional 2D bivariate map. The redness and blueness show the polling lead for Trump and Clinton, respectively. But many polls had high margins of error, creating uncertainty about the election results.
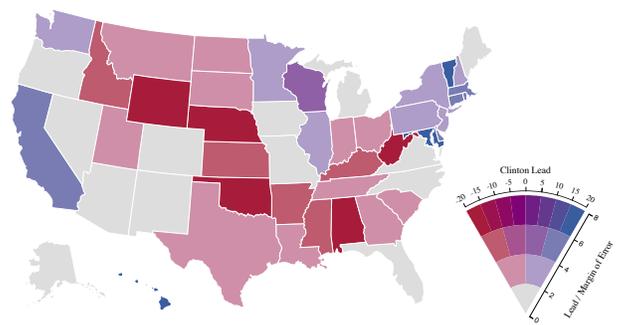

Figure 5: Polling data prior to the 2016 U.S. Presidential Election, encoded with a VSUP.

has sufficient color resolution to show a quasi-linear trend. Uncertainty is high only for a few bins for which there are small number of flights, either on the weekends or late at night. The VSUP does not obscure any significant trends in the high uncertainty data, despite having fewer color categories overall.

2016 Pre-Election Polling

Using a public data set from Wikipedia [51] of the statewide polls that were conducted closest to the 2016 U.S. presidential election, we created a choropleth map showing which candidate was leading in the polls. We calculated uncertainty as the size of the lead in terms of margin of error (lead/margin). The AP style guide for reporters suggests that a candidate leading by more than two margins of error is said to be "significantly ahead" in the polls [17]. As the lead becomes closer to the margin of error, there is more doubt about the electoral outcome, and smaller leads are less definitive.

By providing more color categories when the margin of error was relatively small, the VSUP map shows that, in many of the states leaning towards Hillary Clinton, her lead was quite small. The VSUP map, by limiting inferences about which candidate was leading when the uncertainty is high, also prevents the viewer from making strong claims about the outcome of the election in highly volatile states, even if a state is leaning one way or the other. Given the relative volatility in polling data, this caution is warranted.

EVALUATION

We performed two crowdsourced experiments on Amazon's Mechanical Turk to evaluate the effectiveness of VSUPs 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. Each experiment focused on one of two separate 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 a value of 0.1 and an uncertainty of 0.2."

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 minimize danger. E.g., "place your 5 ships on the safest locations on the board."

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 [21] as a pre-test, 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 for participation, for a target rate of $8/hour. After completion of the main tasks, we solicited demographic information, including a risk aversion assay from Mandrik & Bao [29]. Participants from the first experiment were excluded from participating in the second.

Experimental materials, including data tables and stimuli generation code, are available at https://github.com/uwdata/vsup-paper.

Identification Experiment

We wished to determine what effects, if any, our various design considerations had on performance at a basic analysis task: using a bivariate legend to interpret a heatmap. There are a number of design decisions to be made even after one has determined which two visual channels will be used to encode value and uncertainty. In particular:

1. Whether or not to *juxtapose* or *superimpose* the data and uncertainty maps.

2. Whether or not to *discretize* each visual channel, or encode values *continuously*.

3. Whether or not to use a *square* legend, which gives equal area to each bin, or a *wedge* legend, which reduces in size as uncertainty increases.

4. If discretizing, whether to employ a *VSUP* or a *standard* quantization scheme.

For the identification task, we gave participants a 5x5 heatmap, and asked them to click on a region of the heatmap that encodes a particular (value, uncertainty) pair. To support comparability across conditions, we chose a set of target pairs such that each target mapped to a unique color across all encoding types. This resulted in $(4+4+2+1)=11$ valid targets. For each trial, we placed one example of each of these 11 targets randomly in a heatmap; the remaining 13 cells of the heatmap were randomly sampled from this target list. Thus there could be multiple correct locations in each trial, deterring a strategy of looking for "oddball" colors.

We selected this task both as a way of assessing the ability of people to successfully encode and decode values using a bivariate map, and to provide training for the prediction task (since we reused the bivariate maps in that task). We measured performance both in terms of accuracy (did the participant select the correct point) as well as response time.

We recruited 24 participants for this task: 12 female, 12 male, ($M_{age}= 31$, $SD_{age} = 7.8$). Each of the participants saw the data encoded as each of 8 types of bivariate encoding and legend, with 8 replications, for a total of 64 stimuli. Figure 6 shows these factor levels.

Hypotheses

We had two initial hypotheses, informed by prior work:

1. Juxtaposed maps, by introducing a second search task to the identification task (searching for the proper values in two, rather than one map), would have poorer performance than other conditions.

2. Continuous maps require more specific and accurate color matching between value and color, and so would have poorer performance than discrete conditions.

We had no strong hypotheses about the shape of the legend (square vs. wedge), but, as the wedge shape makes more uncertain bins in the legend smaller (and thus potentially harder to disambiguate), we included it as a check against VSUPs which explicitly alias uncertain values. For this identification task, which did not involve reasoning with uncertainty, but simply matching colors to a color legend, we also did not have strong hypotheses about performance differences between VSUPs and traditional discrete bivariate maps.

Results

We performed a repeated measures ANOVA on our results to measure the effect of juxtaposed versus superimposed maps, and continuous versus discrete legends, on accuracy. Participant ID was included as a random effect. Figure 7 shows the performance of each of our 8 conditions.

The results support our first hypothesis. Juxtaposition was a significant effect ($F(1, 166) = 5.5, p = 0.02$). A post-hoc t-test confirmed ($t = 2.5$, $p = 0.01$) that superimposed charts ($M = 0.58$, $SD = 0.50$) performed significantly better than juxtaposed ($M = 0.51$, $SD = 0.49$) ones. For our identification task, juxtaposition requires at least two search tasks (one to locate squares with the correct data value, another to locate ones with the correct uncertainty value), which introduces additional error compared to conditions where value and uncertainty are displayed simultaneously.

The results also support our second hypothesis. Discretization was a significant effect ($F(1, 166) = 30, p < 0.01$). A post-hoc t-test confirmed ($t = 6.1$, $p < 0.01$) that charts with discrete
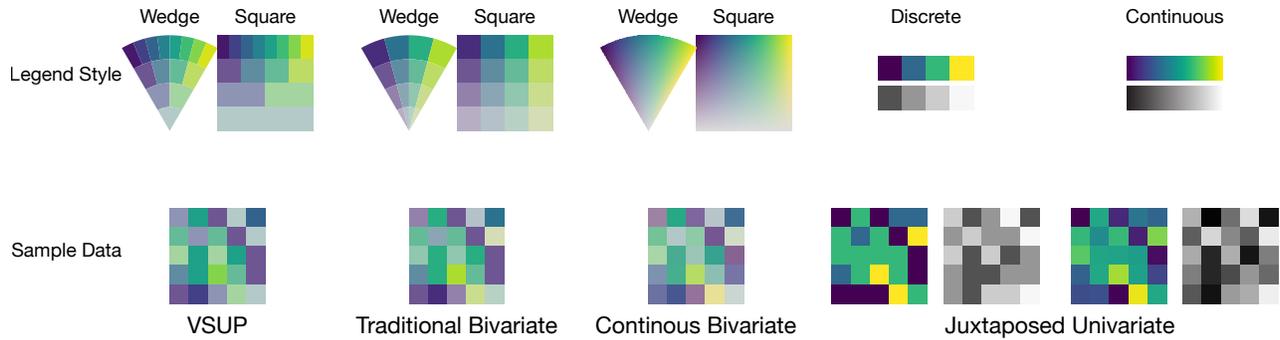
Figure 6: The 8 conditions from the identification experiment. Juxtaposed maps require participants to make an error-prone connection between areas in two separate maps in order to make a decision that integrates value and uncertainty. Traditional bivariate maps integrate both value and uncertainty. VSUPs attempt to improve on traditional bivariate maps by reducing color resolution as uncertainty increases, discouraging conclusions based on noisy or imprecise data.
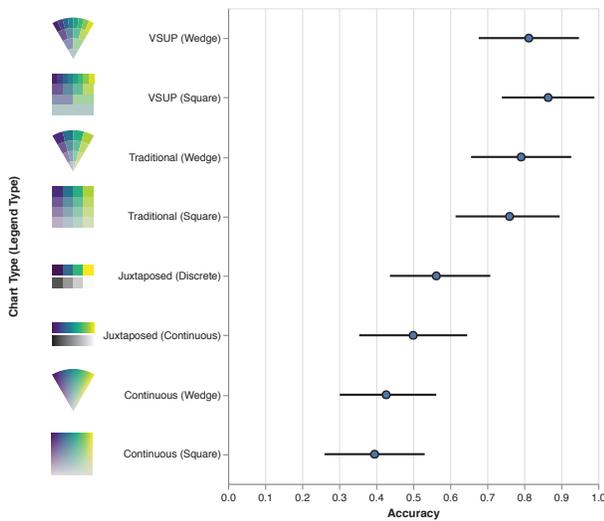

Figure 7: Accuracy results for the identification experiment. For examples of each condition, see Figure 6. Juxtaposing two univariate maps for both value and uncertainty requires an error-prone search task for identification tasks. Continuous rather than discrete bivariate maps requires an error-prone color encoding and estimation task. Discrete bivariate maps, both VSUPs and otherwise, avoid these issues. The confidence intervals are bootstrapped 95% CIs of trimmed means.

bins ($M = 0.63$, $SD = 0.48$) performed significantly better than charts with continuous color maps ($M = 0.47$, $SD = 0.5$). The lack of quantization bias in continuous maps is countered by the perceptual error in precisely estimating value from color. Relying on a discrete set of output colors simplifies this task.

We performed a second ANOVA among the superimposed discrete charts to determine the effect of legend shape (wedge or square) and quantization scheme (VSUP or standard) on performance, with participant ID as a random factor. We did not find a significant effect for either the legend shape ($F(1,70) = 0.04$, $p = 0.84$) or the quantization scheme ($F(1,70) = 1.4$, $p = 0.24$).
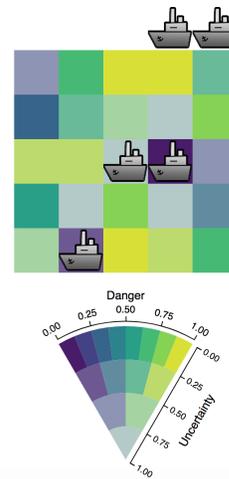

Figure 8: The prediction task. The participant has a list of locations, and ought to place their ships on locations with low probability of attack, and high certainty in this probability. Ships above the heatmap have yet to be placed.

Prediction Experiment

For the prediction task, we gave participants the rules of a game similar to Battleship. Greis et al. [18] 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 5x5 spatial grid, with the expectation that certain squares will be hit by missiles. Players have to place all their tokens before continuing. The objective is to minimize the number of your own ships that are hit. In our task, participants were given a map representing the predictions of missile strikes in each location on the grid. The *value* component was the ship's *danger* if placed on the square. The *uncertainty* component was the confidence in this prediction. Other studies of uncertainty representation, such as in Cox et al. [12], have used "prediction + prediction

uncertainty" stimuli to elicit differences in decision-making between visualizations of uncertainty.

Our stimuli were created by randomly sampling from values that fell within each of the 16 bins of the 4x4 2D bivariate color map. This resulted in 16 samples. The remaining 9 samples were "bad" values, with low safety and high certainty. This created 4 quartile categories of uncertainty, with a skew towards the highest quartile. This stimuli design meant that, while participants had at least one "safe" square (low danger with high certainty), they were forced to make at least some guesses in other quartiles.

We selected this task in order to promote risk-averse behavior. Tversky & Kahneman [45] illustrate that framings in terms of gains or losses produce reliably different outcomes. In particular, there is greater perceived value in avoiding large losses as opposed to striving for a large gain [25]. Our results from the prior study indicate that discrete, non-juxtaposed maps outperformed the other bivariate maps we selected, so we limited our study to only 4 types: square and wedge bivariate maps. While other map types might produce different patterns of decision-making (for instance, the juxtaposed map might encourage participants to ignore uncertainty information altogether), their low accuracy for our previous information fusion tasks led us to discard them, as it would be difficult to distinguish between different patterns of predictions caused by the design, and different patterns caused by simply misreading the heatmap. Having fewer conditions also afforded a within-subjects design that controlled for the variation in interpersonal differences in strategies and risk-aversion, while limiting the potential effects of learning and fatigue from large numbers of stimuli.

The ideal strategy from a value-maximizing standpoint would be to place tokens on areas with the lowest predicted danger (highest expected value), ignoring 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 [30]. A risky player would choose guesses with high expected 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. We therefore measured the distribution of both *value* and *uncertainty* of the tokens placed by the participants.

We recruited 24 participants for this task: 12 female, 12 male, (M_{age}= 37, SD_{age} = 9.8). Our selected square and wedge maps were either VSUPs or traditional bivariate maps, for a 2 (square or wedge legend) x 2 (VSUP or standard) factorial design, with 6 replications, for a total of 24 stimuli. Prior to the main task, we included a short replication of the identification task from the prior experiment (with 12 stimuli) for training and exclusion purposes. 3 people with unacceptably low accuracy on the training task (6%, 6%, and 25% accuracy compared to a mean of 70%) were excluded from analysis.

Hypotheses
We had two hypotheses, stemming from our belief that VSUPs promote better *integration* between uncertainty and value infor-
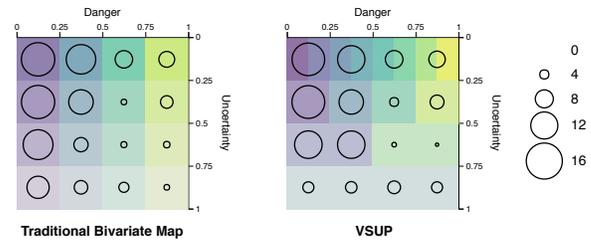

Figure 9: 2D histogram of placements from the prediction experiment. For each trial, participants were asked to place 5 tokens on heatmaps encoding both predicted safety, and uncertainty in these predictions. There were intentionally more tokens to place than there were "safe and certain" locations (top left corner). When looking at traditional 2D maps (left), participants favored safe but uncertain locations (bottom left corner). When looking at VSUPs (right), participants favored locations that were less safe, but more certain.)

mation, and encourage *caution* by highlighting the ambiguity introduced by uncertain data. In particular:

1. Participants would avoid targets with **high uncertainty** when using a VSUP.

2. This would result in a tradeoff where they would also choose targets with **higher danger** when using a VSUP.

As with the prior experiment, we had no strong hypotheses for square vs. wedge-shaped legends, but included both as a check against the potential implicit VSUP-like properties of wedge-shaped legends.

Results
Consistent with our first experiment, we found no significant effect of legend shape on either uncertainty ($F(1,61) = 0.01$, $p = 0.92$) or value ($F(1,61) = 3.1$, $p = 0.08$) of guesses. This result suggests that wedge- and square-shaped legends promote similar patterns of decision-making. Overall, we prefer to employ the wedge-shaped legend for VSUPs, and the square-shaped legend for traditional maps, as it makes the conceptual differences between the two more apparent.

The results partially support our first hypothesis. We performed a repeated measures ANOVA on our results to measure the effect of VSUP versus standard quantizations, and square-versus wedge-shaped legends, on average uncertainty in bets. We found no significant effect of quantization scheme on average guess uncertainty ($F(1,61) = 0.05$, $p = 0.83$). This result indicates that there does not appear to be a uniform pattern of risk aversion between the two scale types. However, given our quartile-based stimuli design, the central tendency of uncertainty would not necessarily capture caution in guesses. To capture differences in this non-normal distribution, and to test our belief that VSUP users would avoid highly uncertain regions, we performed a one-sided two-sample, Kolmogorov-Smirnov test on the distributions of uncertainty in guesses in the VSUP and traditional bivariate conditions. We found a significant difference ($D = 0.5$, $p = 0.03$), and an inspection of the distribution shows that participants using VSUPs were less

likely to make guesses with very high uncertainty. Figure 9 shows this result.

The results support our second hypothesis. A similar ANOVA for average danger value of guesses found a significant effect of quantization scheme ($F(1, 59) = 17$, $p < 0.01$). A post-hoc t-test confirmed ($t = 2.3$, $p = 0.02$) that guesses made with VSUPs had significantly higher average danger ($M = 0.32$, $SD = 0.17$) (and so lower expected value) than traditional maps ($M = 0.29$, $SD = 0.18$).

These two results, taken together, suggest that VSUPs led participants to moderate their guesses: avoiding highly uncertain regions, but willing to "gamble" more on regions of middling certainty (as in Figure 9). In our design, where these gambles were usually poor (due to the predominance of "bad" cells), this moderation resulted in guesses with higher average danger. However, when expected value and uncertainty are orthogonal, this moderation can be beneficial, or at least not harmful. Indeed, for guesses with middling and low uncertainty (quartiles 1-3), where this orthogonality was present, there was no significant difference ($t = 1.56$, $p = 0.11$) in average danger between VSUP ($M = 0.30$, $SD = 0.24$) and traditional ($M = 0.29$, $SD = 0.24$) conditions.

DISCUSSION

Our experimental results have a number of implications for designers wishing to employ bivariate maps for encoding uncertainty. Both of our experiments show that the way that uncertainty information is presented can have a measurable impact on decision-making. VSUPs make a number of design choices in presenting uncertainty. Firstly, they present uncertainty information simultaneously (superimposed) with value, rather than in juxtaposed charts. Secondly, they encode value and uncertainty via discrete, quantized bins, rather than continuously. Lastly, they are designed to alias and suppress uncertain values, rather than certain values. Our experiments provide evidence about the potential impact of each of these choices on performance for information fusion tasks. We will discuss the impact of our experimental findings on the consideration of each of these design decisions in turn.

Juxtaposition vs. Superposition

Gleicher et al. [16] propose three major categories for visualizations that support comparison: juxtaposition (where two variables are visualized in adjacent charts), superposition (where the two variables are encoded simultaneously in a single chart) and explicit encoding (where the two variables are reduced to a single calculated variable, which is directly visualized). Juxtaposition has a number of advantages for bivariate data of this sort: firstly, there are no issues of the integrality of different components of color perception, as there are separate maps for both uncertainty and value. Secondly, juxtaposed displays allow viewers to separately consider patterns in value and uncertainty. The literature also suggests that superposition, which requires a bivariate scale, results in poor accuracy even for basic identification tasks [47].

However, our results indicate that juxtaposed visualizations have a substantial drawback for information fusion tasks. Integrating information from juxtaposed heatmaps is error-prone, as it requires searching for connections in two separate charts. In our experiments, the error introduced by juxtaposition outweighed the benefits of having simpler univariate scales rather than bivariate ones (accuracy of 58% vs. 51% for superimposed and juxtaposed charts, respectively). In our experimental data, where value and uncertainty were not well-correlated, there were no reliable visual landmarks in common amongst the value and uncertainty charts, which may have made it difficult for participants to perform quick correspondence between value in one chart, and uncertainty in the other. Given this error-prone information fusion, we recommend either the use of superposition or explicit encoding for information fusion tasks involving uncertainty. At the very least, designers should use interaction (such as highlighting across charts) to make comparison in dense juxtaposed charts easier.

Discrete vs. Continuous Encodings

When using color to encode data, designers must make the decision whether or not to separate the data into discrete classes of color outputs, or to employ a continuous and class-less scale. This, in turn, introduces a trade-off between *quantization error* (the sampling error introduced by binning the data) and *perceptual error* (the error introduced by the viewer as they decode the color). There is also a cost in terms of expressivity and consistency: a designer might wish for quantitative, continuous information to be represented continuously [28].

However, for bivariate color maps, where often the color channels for each variable are not perceptually separable, it can be difficult to decode the color into its component parts. In our identification experiment this perceptual decoding error was high. Continuous maps performed the worst of all of our other bivariate map types (47% accuracy vs. 63% accuracy for discrete maps). Echoing the results of Padilla et al. [34], we hold that the error introduced by quantization, which is known and bounded, has performance benefits even for continuous data, if the perceptual error is sufficiently high. Bivariate maps, where the integrality of channels raise this perceptual error, are a prime example of such a scenario. We advise designers to avoid continuous bivariate color scales when possible, and instead employ discrete scales with a relatively small number of categories.

Risk Averse vs. Risk Seeking Decision-Making

It is often difficult for the general audience to successfully make use of uncertainty information. In order to minimize the complexity of uncertainty, uncertain information is often visualized in a binary way, where values either meet some explicit uncertainty standard, or do not. These binary encodings for uncertain information, such as error bars for estimates, or "cones of uncertainty" for hurricane paths, result in binary patterns of decision-making that do not accurately reflect the underlying uncertainty in the data [12, 10]. On the other hand, presenting more detailed uncertainty information may overwhelm or confuse the viewer [19].

VSUPs are in a middle ground between sparse but simple, and dense but complex presentations of uncertainty. At high levels of uncertainty, VSUPs have only one output color, and so explicitly prevent any decoding of value. This narrowing of

options is reflected not only in the quantization scheme, but also in the shape of the legend, narrowing as the uncertainty increases. We selected this design to promote *risk-averse* decision-making. That is, people should be hesitant to make strong predictions or decisions about values when uncertainty is high. Our prediction experiment, drawing from Kahneman & Tversky's prospect theory [25], was further designed to elicit avoidance of uncertain data, by framing the predictions in terms of minimizing losses, rather than maximizing gains. This framing was reflected in participants mostly making safe and certain guesses (low danger, high certainty). Where the differences in decision-making occurred were where the participants had exhausted the set of safe and certain locations, and were forced to value either certainty or expected value. A risk seeker would choose uncertain locations if the reward were high enough, whereas a risk-avoider would select locations with less uncertainty, even with a lower payoff. Participants using VSUPs were likelier to make these risk-avoiding decisions than when they were using traditional bivariate maps. Beyond VSUPs, this result indicates that the designers of visualizations can have fine-grained control over the types of decision-making their visualizations promote. Visualization designers should consider not just fidelity to the underlying dataset, but also the decision strategies their designs support.

Limitations and Future Work

VSUPs assume that uncertain values ought to have less visual impact than highly certain values. VSUPs therefore act both as a kind of filtering device as well as a bivariate representation. This assumption holds in scenarios like the polling dataset (Figure 5), 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. VSUPs 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, 50], and so the concept of a limited "budget" of distinguishable marks easier to quantify and illustrate. In principle, a VSUP 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 [40], the creation of VSUPs 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 VSUPs can be considered a feasible design strategy. We are currently experimenting with the creation of VSUPs using these other channels, such as color and size.

VSUPs, like other mappings involving quantization, are also highly sensitive to the binning scheme used. Uncertain data with values that fall very close to bin boundaries may end up having large color differences, even though they may be close in value. The comparison to effects tests is therefore imperfect,

and a particular item's encoding in a VSUP cannot be used as a proxy for a test of statistical significance. As with other quantized forms of data analysis, including histograms and geospatial binning, analysts and designers should consider the distribution of values before quantizing.

Conclusion

Uncertainty, data quality, or confidence values are often considered separately from the data itself, relegated to tooltips or visually distant supplemental charts. We contend, in contrast, that uncertainty information ought to be directly integrated within a shared chart. This integration introduces additional complexity in the design and presentation of data. Value-Suppressing Uncertainty Palettes represent one strategy for dealing with this complexity, by assigning mark 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 II. 2007. Rainbow color map (still) considered harmful. *IEEE Computer Graphics and Applications* 27, 2 (2007). DOI: http://dx.doi.org/10.1109/MCG.2007.323435

3. Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2301–2309. DOI: http://dx.doi.org/10.1109/TVCG.2011.185

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. DOI: http://dx.doi.org/10.1109/TVCG.2012.220

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. DOI: http://dx.doi.org/10.1145/1520340.1520616

6. Ken Brodlie, 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. DOI: `http://dx.doi.org/10.1007/978-1-4471-2804-5_6`

7. Barbara Buttenfield and M Kate Beard. 1994. Graphical and geographical components of data quality. *Visualization in geographic information systems* (1994), 150–157.

8. 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. DOI: `http://dx.doi.org/10.1093/bioinformatics/btv407`

9. Michael Correll, Subhadip Ghosh, David O'Connor, and Michael Gleicher. 2011. Visualizing virus population variability from next generation sequencing data. In *2011 IEEE Symposium on Biological Data Visualization (BioVis)*. IEEE, 135–142. DOI: `http://dx.doi.org/10.1109/BioVis.2011.6094058`

10. Michael Correll and Michael Gleicher. 2014. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2142–2151. DOI: `http://dx.doi.org/10.1109/TVCG.2014.2346298`

11. 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. DOI: `http://dx.doi.org/10.1109/TVCG.2016.2598618`

12. Jonathan Cox, Donald House, and Michael Lindell. 2013. Visualizing uncertainty in predicted hurricane tracks. *International Journal for Uncertainty Quantification* 3, 2 (2013). DOI:`http://dx.doi.org/10.1615/Int.J.UncertaintyQuantification.2012003966`

13. Michael W Dobson. 1973. Choropleth maps without class intervals?: a comment. *Geographical Analysis* 5, 4 (1973), 358–360. DOI: `http://dx.doi.org/10.1111/j.1538-4632.1973.tb00498.x`

14. Richard Dunn. 1989. A dynamic approach to two-variable color mapping. *The American Statistician* 43, 4 (1989), 245–252. DOI:`http://dx.doi.org/10.2307/2685372`

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. DOI:`http://dx.doi.org/10.1016/0010-0285(70)90016-2`

16. Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D Hansen, and Jonathan C Roberts. 2011. Visual comparison for information visualization. *Information Visualization* 10, 4 (2011), 289–309. DOI: `http://dx.doi.org/10.1177/1473871611416549`

17. N Goldstein. 1994. *The Associated Press Stylebook and Libel Manual. Fully Revised and Updated*. ERIC.

18. 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. DOI: `http://dx.doi.org/10.1145/2971485.2971535`

19. Miriam Greis, Jessica Hullman, Michael Correll, Matthew Kay, and Orit Shaer. 2017. Designing for Uncertainty in HCI: When Does Uncertainty Help?. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 593–600. DOI: `http://dx.doi.org/10.1145/3027063.3027091`

20. Henning Griethe, Heidrun Schumann, and others. 2006. The visualization of uncertain data: Methods and problems. In *SimVis*. 143–156.

21. 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. DOI: `http://dx.doi.org/10.1364/JOSA.35.000268`

22. Mark Harrower and Cynthia A Brewer. 2003. ColorBrewer. org: an online tool for selecting colour schemes for maps. *The Cartographic Journal* 40, 1 (2003), 27–37.

23. Tomislav Hengl and Norair Toomanian. 2006. Maps are not what they seem: representing uncertainty in soil-property maps. In *Proc. Accuracy*. 805–813.

24. 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. DOI: `http://dx.doi.org/10.1371/journal.pone.0142444`

25. Daniel Kahneman and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society* (1979), 263–291. DOI: `http://dx.doi.org/10.2307/1914185`

26. Lydia R Lucchesi and Christopher K Wikle. 2017. Visualizing uncertainty in areal data with bivariate choropleth maps, map pixelation and glyph rotation. *Stat* (2017). DOI:`http://dx.doi.org/10.1002/sta4.150`

27. 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. DOI: `http://dx.doi.org/10.1109/TVCG.2012.279`

28. Jock Mackinlay. 1986. Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)* 5, 2 (1986), 110–141.

29. Carter A Mandrik and Yeqing Bao. 2005. Exploring the concept and measurement of general risk aversion. *NA-Advances in Consumer Research Volume 32* (2005).

30. Leonard Mlodinow. 2009. *The drunkard's walk: How randomness rules our lives*. Vintage. 81–104 pages.

31. Dominik Moritz, Danyel Fisher, Bolin Ding, and Chi Wang. 2017. Trust, but Verify: Optimistic Visualizations of Approximate Queries for Exploring Big Data. In *Proceedings of the 33rd annual ACM conference on Human factors in computing systems*. ACM. DOI: `http://dx.doi.org/10.1145/3025453.3025456`

32. Jean-Claude Muller. 1979. Perception of Continuously Shaded Maps. *Annals of the Association of American Geographers* 69, 2 (1979), 240–249. DOI: `http://dx.doi.org/10.1111/j.1467-8306.1979.tb01254.x`

33. 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`

34. 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. DOI: `http://dx.doi.org/10.1109/TVCG.2016.2599106`

35. Maria Riveiro. 2007. Evaluation of uncertainty visualization techniques for information fusion. In *10th International Conference on Information Fusion*. IEEE, 1–8. DOI:`http://dx.doi.org/10.1109/ICIF.2007.4408049`

36. 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. DOI: `http://dx.doi.org/10.1109/MCG.1986.276688`

37. 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. DOI: `http://dx.doi.org/10.1109/VISUAL.2001.964510`

38. 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. DOI: `http://dx.doi.org/10.1063/1.4822401`

39. Robert E Roth, Andrew W Woodruff, and Zachary F Johnson. 2010. Value-by-alpha maps: An alternative technique to the cartogram. *The Cartographic Journal* 47, 2 (2010), 130–140. DOI: `http://dx.doi.org/10.1179/000870409X12488753453372`

40. Maureen Stone, Danielle Albers Szafir, 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.

41. 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.

42. 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.

43. Bruce E Trumbo. 1981. A theory for coloring bivariate statistical maps. *The American Statistician* 35, 4 (1981), 220–226. DOI:`http://dx.doi.org/10.2307/2683294`

44. Archie Tse. 2016. Why We Are Doing Fewer Interactives. (2016). `https://github.com/archietse/malofiej-2016/blob/master/tse-malofiej-2016-slides.pdf` Malofiej.

45. 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. DOI: `http://dx.doi.org/10.1007/978-3-642-70634-9_6`

46. Stéfan van der Walt and Nathaniel Smith. 2015. mpl colormaps. `https://bids.github.io/colormap/`, (2015).

47. Howard Wainer and Carl M Francolini. 1980. An Empirical Inquiry Concerning Human Understanding of Two-Variable Color Maps. *American Statistician* (1980), 81–93. DOI:`http://dx.doi.org/10.2307/2684111`

48. Colin Ware. 1988. Color sequences for univariate maps: Theory, experiments and principles. *IEEE Computer Graphics and Applications* 8, 5 (1988), 41–49. DOI: `http://dx.doi.org/10.1109/38.7760`

49. Colin Ware. 2009. Quantitative texton sequences for legible bivariate maps. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (2009). DOI: `http://dx.doi.org/10.1109/TVCG.2009.175`

50. Colin Ware. 2012. *Information visualization: perception for design*. Elsevier.

51. Wikipedia. 2017. Statewide opinion polling for the United States presidential election, 2016. (2017). `https://en.wikipedia.org/wiki/Statewide_opinion_polling_for_the_United_States_presidential_election,_2016`