The Noisy Work of Uncertainty Visualisation Research: A Review

Harriet Mason

2024-07-19

Table of contents

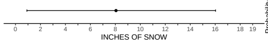
1 Background

From entertainment choices to news articles to insurance plans, the modern citizen is so over run with information in every aspect of their life it can be overwhelming. In this overflow of information, tools that can effectively summarize information down into simple and clear ideas become more valuable. Information visualisations remain one of the most powerful tools for fast and reliable science communication.

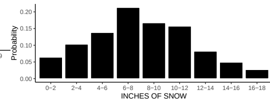
Visualization is an important step in exploratory data analysis and it is often utilised to **learn** what is important about a data set. Datasets such as Anscombe's quartet (Anscombe 1973) or the Datasaurus Dozen (Locke and D'Agostino McGowan 2018) highlight this power in visualisation. Additionally, visualisations allow for efficient and memorable communication. Even something as simple as sketching a distribution before recalling statistics or making predictions can greatly increase the accuracy of those measures (Hullman et al. 2018; **Goldstein2014?**).

Uncertainty visualisation is a relatively new field in research. Early papers that specifically reference "uncertainty visualisation" appear in the late 80s (Ibrekk and Morgan 1987), with geospatial information visualisation literature in the early 90s declaring this to be essential aspect of information display (MacEachren 1992; Carr, Olsen, and White 1992). These early experiments typically involved showing participants a distribution, such as those depicted in Figure **??**, and asking the viewers to extract a probability or average. Despite the new terminology visualisation of uncertainty has been present since the earliest times. For example, box plots or histograms can be considered to be displaying uncertainty in the sense of variability in observations sampled from a population distribution. Today, there is an abundance of publications on the topic which makes it is timely to construct a review of the field. In fact, there have already been several reviews published.

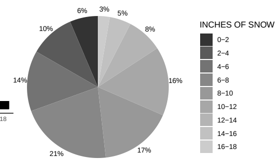
Reviews on uncertainty visualisation rarely offer tried and tested rules for effective uncertainty visualisation, but rather comment on the *difficulties* faced when trying to summarize



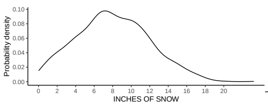
(a) Picture 1



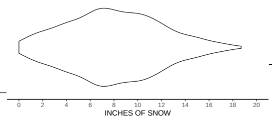
(b) Picture 2



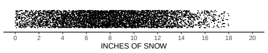
(c) Picture 3



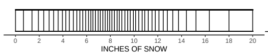
(d) Picture 4



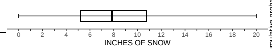
(e) Picture 5



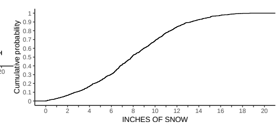
(f) Picture 6



(g) Picture 7



(h) Picture 8



(i) Picture 9

the field. Kinkeldey, MacEachren, and Schiewe (2014) found most experiments on the methods for uncertainty visualisation evaluation to be ad hoc, with no commonly agreed upon methodology or formalisation and no greater goal of describing general principals. Hullman (2016) commented on the difficulty in taking overarching themes from uncertainty visualisation, as several conflated issues make it unclear if subjects did poorly in an experiment because they misunderstood a visualisation, because the question was misinterpreted, or because they used a specific heuristic. Spiegelhalter (2017) commented that different plots are good for different things, and disagreed with the goal of identifying a universal "best" plot for all people and circumstances. Griethe and Schumann (2006) was unable to find common themes, but instead listed the findings and opinions of a collection of papers. L. Padilla, Kay, and Hullman (2022) summarized several cognitive effects that have repeatedly arised in uncertainty literature, however these effects were each discussed in isolation as a list of considerations an author might make. While these reviews are thorough in scope, none discuss how the existing literature contribute to the commonly state goal of uncertainty visualisation, scientific transparency. The problem faced by the literature is easily summarized with a famous quote by Henri Poincaré.

> "Science is built up of facts, as a house is built of

> stones; but an accumulation of facts is no more a science than a heap of stones is a house." - Henri Poincaré (1905)

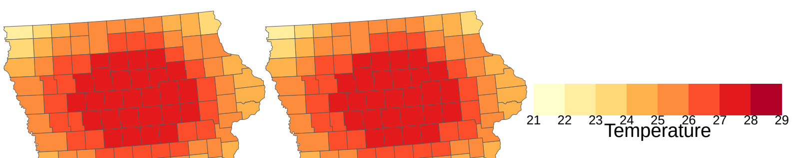
That is to say, despite the wealth of reviews, the field of uncertainty visualisation remains a heap of stones. This review attempts to address this issue by offering a novel perspective on the uncertainty visualisation problem, and hopefully laying the foundations on which we can build a house. This review is broken into several parts that each reflect a different approach to uncertainty visualisation. First we look at graphics that ignore uncertainty entirely and discuss why uncertainty should be included at all. Second, we look at methods that consider uncertainty to be just another variable and discuss the characteristics of uncertainty that make it a unique visualisation problem. Third, we look at methods that explicitly combine our estimate and its uncertainty and discuss if the visualisations created by these transformations are still "uncertainty visualisations". Fourth, we will discuss methods that implicitly include uncertainty by depicting a sample or original data in place of an estimate. Finally, we discuss how uncertainty visualisations can be effectively evaluated. When discussing each of these methods, we consider the *purpose* of uncertainty visualisation and comment on how effective each visualisation is at fulfilling that purpose.

Due to the fields origins and focus in geospatial information visualisation, there have been a large number of suggested variations on the choropleth map that allow authors to include uncertainty. We will use these maps to provide simple examples for each approach that can be easily compared to the "no uncertainty" choropleth map to better understand the costs and benefits of each approach. Despite the example of each method focusing on variations of the choropleth map, it is important to understand that the approaches we are discussing are universal and are not unique to maps.

2 Ignoring uncertainty

A good place to start might be at deceptively straight forward question, why should we include uncertainty at all?

Figure 2 depicts a choropleth map of the counties of Iowa. Each of these counties are colored according to an estimate of average daily temperature that was simulated so that the values followed a clear spatial trend (hot in the middle of the map, and cold on the outside). The variance of these estimates were simulated such that a hypothesis test would indicate the existence of a spatial trend in the low variance map, while the trend in the high variance map should be indistinguishable from noise. Is this distinction in validity of the spatial trend clear in in the map? Is the validity of the trend communicated through the visualisation?



(c) Choropleth Palette

(a) High Variance Data

(b) Low Variance Data

Figure 2: Two choropleth maps that depict the counties of Iowa where each country coloured acording to a simulated average temperature. Both maps depict a spatial trend, where counties closer to the center of the map are hotter than counties on the edge of the map. This trend is only technically there in the low variance condition, i.e. if we performed a hypothesis test we would conclude the trend exists. The high variance condition displays a spatial trend that could simply be the result of noise, which means the plot is displaying a false conclusion.

2.1 Signal-supression

The two choropleth maps appearing to be identical in Figure 2 highlights the need for uncertainty visualisations. Uncertainty visualisation is is required for transparency and this sentiment has been repeated. Some authors suggest uncertainty is important to include as it communicates the legitimacy (or illegitimacy) of the conclusion drawn from visual inference (Correll and Gleicher 2014; Kale et al. 2018; Griethe and Schumann 2006). Some authors have said that uncertainty should be included to degree of confidence or trust in the data (Boukhelifa et al. 2012; Zhao et al. 2023). Some authors directly connect uncertainty visualisation to hypothesis testing as it ensures the "validity" of a statement (Hullman 2020; Griethe and Schumann 2006), but allows for a proportional level of trust that is more detailed than the binary results of a hypothesis test (Correll and Gleicher 2014; Correll, Moritz, and Heer 2018). Some authors even go so far as to claim that failing to include uncertainty is akin to fraud or lying (Hullman 2020; Manski 2020).

This consensus leads us to understand that uncertainty visualisation is motivated by the need for a sort of "visual hypothesis test". A successful uncertainty visualisation would act as a "statistical hedge" for any inference me make using the graphic. This implies three key criteria for a good uncertainty visualisation:

1) Reinforce justified conclusions to encourage confidence in results
2) Prevent unjustified conclusions that are indistinguishable from noise
3) Perform tasks 1) and 2) in a way that is proportional to the level of confidence in those conclusions.

As Figure 2 showed, visualisations that are unconcerned with uncertainty have no issue showing justified signals, but struggle with the display of unjustified signals. Therefore, we call this approach "signal-supression" as it primarily differentiates itself from from a typical visualisation approach through criteria (2), that is, an uncertainty visualisation should prevent us from drawing unjustified conclusions. We will investigate several approaches for uncertainty visualisation in the following sections and in doing so, discuss how successful each approach is in fulfilling the signal-suppression criteria.

2.2 Alternative motivations

Of course, uncertainty visualisation is not only motivated by signal-supression.

It is not entirely true that uncertainty is only important for transparency, however it is not clear that the "uncertainty" provided in other cases is uncertainty at all. Some authors claim the purpose of uncertainty is not transparency, but rather important information for effective decision making (Ibrekk and Morgan 1987; L. Padilla, Kay, and Hullman 2022; Hullman 2016; Cheong et al. 2016; Boone, Gunalp, and Hegarty 2018; L. M. Padilla, Ruginski, and Creem-Regehr 2017). While this can be true, the uncertainty provided for this purpose is *not* the same as the uncertainty provided for transparency. It arguably may not be uncertainty at all. This is easiest to understand with an example.

Let's say my weather prediction app only provides me with the predicted daily rainfall when I am trying to decide if I should bring an umbrella. While my decision will be improved with the inclusion of uncertainty information, that is not because of a special quality of uncertainty, but rather because the uncertainty information can be used to calculate the true statistic of interest, the probability that it will rain at all. If I use uncertainty to improve decision making, it is not *true* uncertainty, rather it is supplementary material that is required to correct for an insufficient statistic. I was provided with was not the statistic required to make the decison.

3 Visualising Uncertainty as a Variable

Upon hearing that uncertainty needs to be included for transparency, the solutions may seem obvious. You may think "well, I will just add a dimension to my plot that includes uncertainty". This is not a silly choice, but lets look at what happens when we do this.

3.1 Bivariate Map

3.2 Uncertainty is not just another variable

Some authors do not describe uncertainty as a means of transparency, but rather as a variable of importance in of itself (Blenkinsop et al. 2000).

The true reason that uncertainty visualisation is differerent to normal information visualisation, is that normal information visualisation is designed to *find* signals in our data, while uncertainty visualisation should be designed to *supress* it.

It is rarely made clear exactly how uncertainty visualisation is different to normal information visualisation. Hullman (2016) commented that it is straightforward to show a value but it is much more complex to show uncertainty, but did not explain *why*. A visualisation is nothing more than computing a statistic, mapping that statistic to a visual feature and rendering that graphic. Nothing in this process suggests visualising an estimate of the variance should be any different to visualising an estimate of the mean, median, maximum, or of any other statistic, however this field *does* implicitly treat them different. There is no such thing as "maximum visualiation" or "median visualisation" so there *must* be something about uncertainty that causes authors to believe it is special. Here we seek to answer this question.

Uncertainty visualisation papers usually provide one of two primary justifications for their existence. The first justification is that uncertainty is fundamentally differently to other variables due to the psychological heuristics involved in interpreting uncertainty. Therefore uncertainty visualisation is different to normal visualisation as authors must consider the psychological effects of what they are visualising (Spiegelhalter 2017; Hullman et al. 2019). The second justification assets that uncertainty is an additional variable and it is of vital importance to interpreting an estimate. Therefore uncertainty visualisation is a high dimensional visualisation problem, as we need to figure out how to seamlessly add uncertainty into already existing graphics (**moritz2017trust?**; Griethe and Schumann 2006).

Why some authors pick one motivation over another is rarely explained and both leave the role of uncertainty in the visualisation unclear. The literature never explains if uncertainty should be treated as a variable, as metadata, or as something else entirely (Kinkeldey, MacEachren, and Schiewe 2014).

> Do we treat uncertainty as 'just another variable' to be visually represented or does it need to be treated differently? For example, when we picture a map showing air pressure distribution in combination with temperature, these two variables are certainly dependent on each other (in physical terms). The same is true with air pressure and its uncertainty, but we see a stronger dependency: Uncertainty can be seen as metadata of air- pressure which we argue makes a difference. Thus, we support Edwards and Nelson (2001, p. 35) who stated that '[p]erhaps data certainty information is unique and will require a new type of framework for designing symbolization'. Most studies that assess the usability of uncertainty visualisations do not contribute to this aspect since they test the retrieval of data and uncertainty separately, but from the perspective that there is nothing special about uncertainty. Traditionally, studies in this field focus on the ability to read both the map content and its uncertainty at the same time. This may be the mandatory criterion for a successful use of uncertainty, but the question that remains is whether this is sufficient to ensure that a user does not only have two separate values in mind but an integrated uncertain data value.

4 Explicitly Combining Uncertainty and Signal

4.1 Using Statistics

4.1.1 Exceedance Probability Map

4.1.2 What is an uncertainty visualisation?

4.2 Using Aesthetics

4.2.1 Value Supressing Uncertainty Palettes

4.2.2 There is no such thing as a test for everything

5 Implicitly Combining Uncertainty and Signal

5.1 Pixel maps

5.2 Just visualise your data

6 Evaluating uncertainty

7 Current methods

8 Testing signal supression

8.1 Current methods

There are many secondary benefits that come with this improved transparency, such as better decisions, more trust in the results and more confidence in the authors. These secondary benefits are, however, *not* the immediate goal of uncertainty. The following sections will discuss the issues and limitations in measuring uncertainty through these secondary metrics and provide suggestions as to how future studies should consider measuring uncertainty.

Trust is a by product of displaying uncertainty rather than the goal of it, and viewing the relationship in the converse direction can lead to misguided research. Considering trust, and not transparency, as the metric of importance in uncertainty communication can lead to a questionable subtext that argues against transparency, something that has been noticed by several other authors [Spiegelhalter (2017); ONeill2018]. Hullman (2020) found that author simultaneously argued that failing to visualise uncertainty was akin to fraud, but also many avoided uncertainty visualisation because they didn't want their work to come across as "untrustworthy". These authors are optimising *trust* rather than *transparency*, which means they opt to leave out uncertainty information when it does exactly what it is supposed to, decrease certainty in conclusions.

Science communication should be primarily concerned with accuracy. Setting trust and risk-aversion as the variables of interest implicitly encourages statisticians to set trust and risk-aversion as the primary goals of communication. The issue of trust being divorced from trustworthiness has been commented on by other authors (O'Neill 2018), however the issue still persists in the uncertainty visualisation literature. Zhao et al. (2023) displayed a several variations of a visualisation of a model prediction and its uncertainty and took participants using the model prediction as a sign of trust. They reported that visualising uncertainty information caused participants to trust the model in the low variance case, but the results in the high variance case were inconclusive. The discussion made it clear the authors thought the uncertainty information should make the visualisation more trustworthy, but conflating trust and the use of a prediction implied uncertainty information should somehow influence participants to use their own prediction, even though a prediction being uncertainty does not necessarily mean it is incorrect. Despite this, the authors seemed to assume that the uncertainty information *should* have an influence on that, showing they had not deeply considered *how* uncertainty information should influence the choices of the participants. (*Cite Gap 18: examples of studies where authors measure trust*)

A similar measure to trust is using "confidence" in an extracted value or a decision. Interestingly, "confidence" is also used to try and capture the clarity of a message in a normal visualisation. Confidence cannot simultaneously be a measure of clarity of visualisation *and* a way to capture the uncertainty expressed in a visualisation.

9 ———————- BELOW HERE IS UNSORTED —————————

10 Why is uncertainty special?

Interpretation and semantics experiments are seeking to identify a dimension (or visual task) that uncertainty naturally maps to. These experiments inherently view uncertainty as a variable that is separate to the variable on which we have mapped our signal. For example, lets say we have a map were maximum daily temperature is presented using points where the colour (red for hot and blue for cold) of the point is associated with the temperature, and the blurriness of that point is associated with the variance *of* that temperature. It is highly likely that our brains will not flatten that into a single variable depicting noise and signal, but rather *separately* extract the temperature (colour) and uncertainty (blur) information as two independent variables. If the variables are extracted separately, there is no guarantee that the uncertainty will act as an appropriate signal suppressor. This problem has been noticed by others in the field, that typically use this method (specifically in the spatial uncertainty context) and a desire for representations that integrate uncertainty and signal is one of the reasons for the invention of the value-suppressing uncertainty pallet (Correll, Moritz, and Heer 2018).

Value-suppressing uncertainty pallets (VSUP) were developed as a method that would allow the signal and the noise to be interpreted together such that insights gained by the viewers of a plot are appropriately *suppressed* by the uncertainty. Hence the name of the pallet. **?@fig-maps** depicted this map colouring approach and several other extensions on the typical choropleth map that differ in where in the visualisation process they combine the noise and signal information into a single concept of "valid signal". The first and most basic map is a simple choropleth map where each value (the colouring of each local government area) has no associated "uncertainty". The next map (which embodies the approach taken by the semantics experiments) is the bivariate map which maps the signal to colour value and the uncertainty to colour hue. If visualisation was performing signal suppression then that would mean the two dimensional space defined by colour value and colour hue can be mentally "flattened" into a single dimension of "valid value" that can captures the signal the our brain would need to be able to flatted this two dimensional space into a single space of "signal validity". The idea of two perceptual tasks flattening into one variable in the mind of the viewer may be wishful thinking, but it is not impossible given we are not certain on how the perceptual tasks are mapped within the human brain. Sterzik et al. (2023) found that when a value was mapped to the textures of stippling, hatching, and triangles, and found that the difference between two points on this one dimensional texture was actually a 2D space (likely "texture business" and "light/darkness"). That being said, if we look at the visual signals presented by the bivariate map, where the contrasting light and dark areas actually has no important meaning, it is unlikely this occurs for colour value and hue (or at least it doesn't occur in a way that is useful for uncertainty visualisation). Instead of hoping that uncertainty might collapse signal values into a single dimension, we can do some of that work ourselves, by using a VSUP which collapses the colour space such that high uncertainty values cannot be extracted. It is unclear how useful VSUPs are at actually combining signal and noise and therefore suppressing plot level insights, as they have only been tested on simple value extraction tasks that require evaluating a single point (Correll, Moritz, and Heer 2018; Ndlovu, Shrestha, and Harrison 2023) rather than looking for spatial relationships (which is arguably what maps are for). Following along with this trend, the next way we might consider visualising uncertainty is to combine uncertainty and signal at the earlier stage so the "suppressed signal" is represented by a single variable. This statistic can then be expressed in a one dimensional colour space, which is a method adopted by the Bayesian surprise metric map (Ndlovu, Shrestha, and Harrison 2023) and the exceedance probability map [Lucchesi2017].

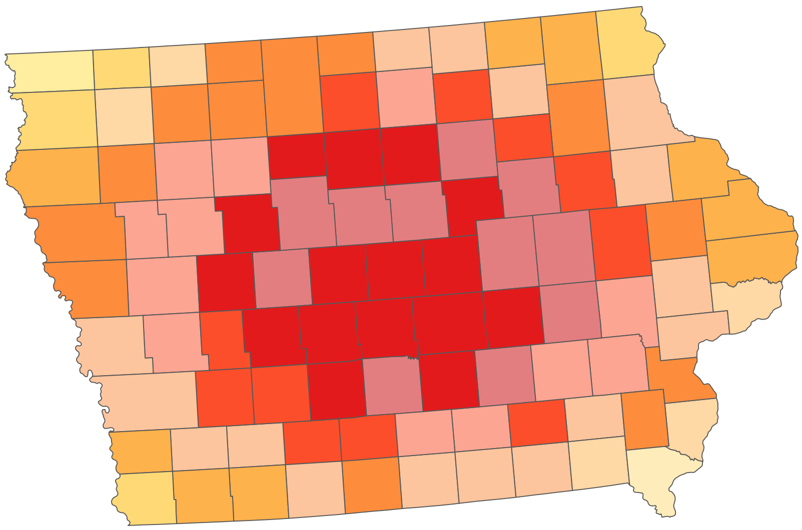


Figure 3: A visualisation of a typical choropleth map, as well as three other maps that are used to display uncertainty on the choropleth map. There is a high variance and low variance example for each type of map to show how well the technique supresses signal. Each map was created using the same data with the same base palette. At first glance, the high uncertainty bivariate and VSUP maps just look like maps with a low saturation colour palette, rather than map with high uncertainty. This visualisation makes it clear that supression methods that plot uncertainty to a second axis, such as hue, make uncertainty appear as a second variable, rather than a signal supression on our estimate.

These maps make the importance of combining uncertainty and signal in a single visual channel clear. A choropleth map will show signal that is not valid inference because of high uncertainty. At the other end of the spectrum, the bivariate map will show signal that is not always interesting because it forces us to interpret uncertainty and signal separately. As we move through these methods, it seems that the validity of any overarching insight becomes more visible at the cost of our ability

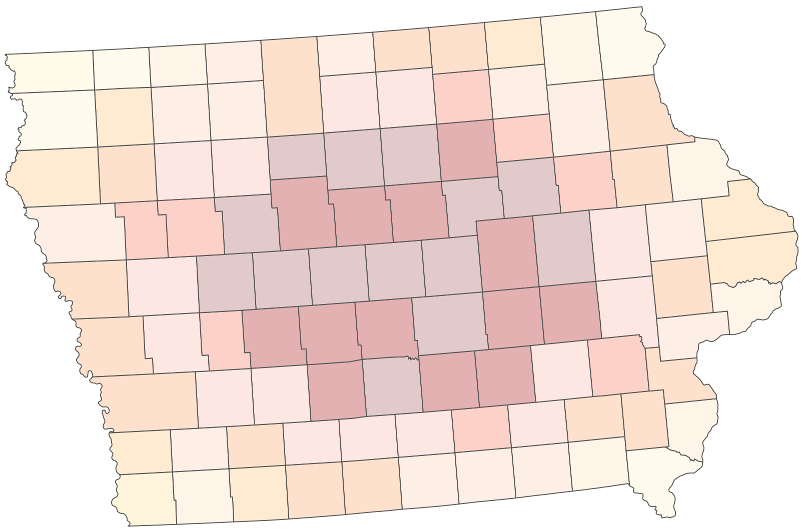


Figure 4: A visualisation of a typical choropleth map, as well as three other maps that are used to display uncertainty on the choropleth map. There is a high variance and low variance example for each type of map to show how well the technique supresses signal. Each map was created using the same data with the same base palette. At first glance, the high uncertainty bivariate and VSUP maps just look like maps with a low saturation colour palette, rather than map with high uncertainty. This visualisation makes it clear that supression methods that plot uncertainty to a second axis, such as hue, make uncertainty appear as a second variable, rather than a signal supression on our estimate.

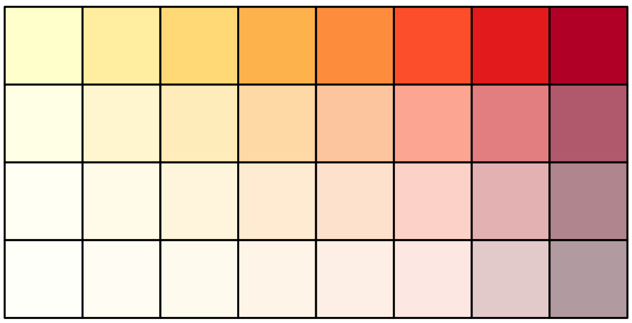


Figure 5: A visualisation of a typical choropleth map, as well as three other maps that are used to display uncertainty on the choropleth map. There is a high variance and low variance example for each type of map to show how well the technique supresses signal. Each map was created using the same data with the same base palette. At first glance, the high uncertainty bivariate and VSUP maps just look like maps with a low saturation colour palette, rather than map with high uncertainty. This visualisation makes it clear that supression methods that plot uncertainty to a second axis, such as hue, make uncertainty appear as a second variable, rather than a signal supression on our estimate.

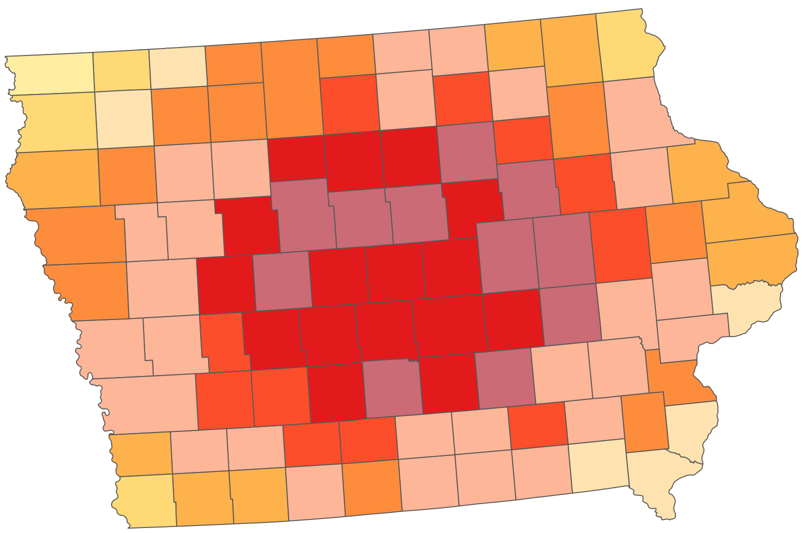


Figure 6: A visualisation of a typical choropleth map, as well as three other maps that are used to display uncertainty on the choropleth map. There is a high variance and low variance example for each type of map to show how well the technique supresses signal. Each map was created using the same data with the same base palette. At first glance, the high uncertainty bivariate and VSUP maps just look like maps with a low saturation colour palette, rather than map with high uncertainty. This visualisation makes it clear that supression methods that plot uncertainty to a second axis, such as hue, make uncertainty appear as a second variable, rather than a signal supression on our estimate.

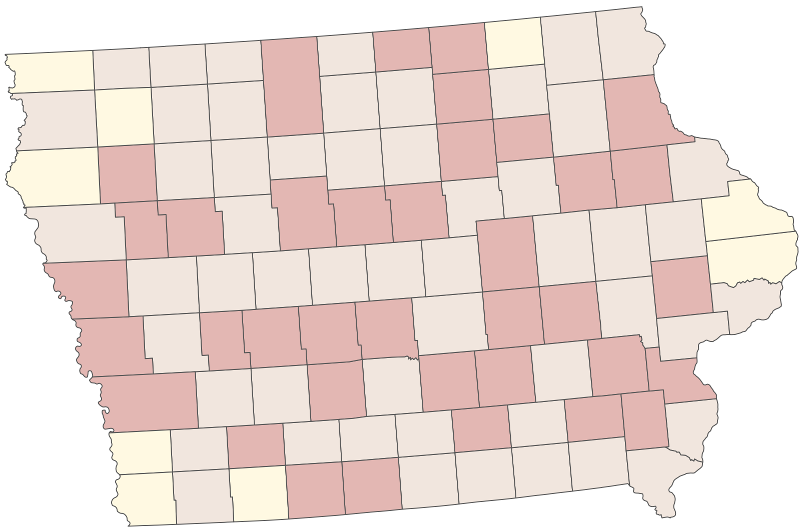


Figure 7: A visualisation of a typical choropleth map, as well as three other maps that are used to display uncertainty on the choropleth map. There is a high variance and low variance example for each type of map to show how well the technique supresses signal. Each map was created using the same data with the same base palette. At first glance, the high uncertainty bivariate and VSUP maps just look like maps with a low saturation colour palette, rather than map with high uncertainty. This visualisation makes it clear that supression methods that plot uncertainty to a second axis, such as hue, make uncertainty appear as a second variable, rather than a signal supression on our estimate.

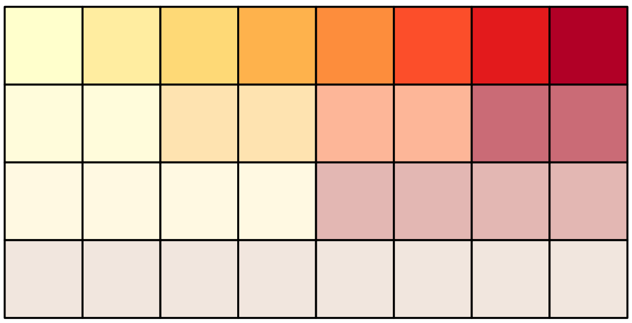


Figure 8: A visualisation of a typical choropleth map, as well as three other maps that are used to display uncertainty on the choropleth map. There is a high variance and low variance example for each type of map to show how well the technique supresses signal. Each map was created using the same data with the same base palette. At first glance, the high uncertainty bivariate and VSUP maps just look like maps with a low saturation colour palette, rather than map with high uncertainty. This visualisation makes it clear that supression methods that plot uncertainty to a second axis, such as hue, make uncertainty appear as a second variable, rather than a signal supression on our estimate.

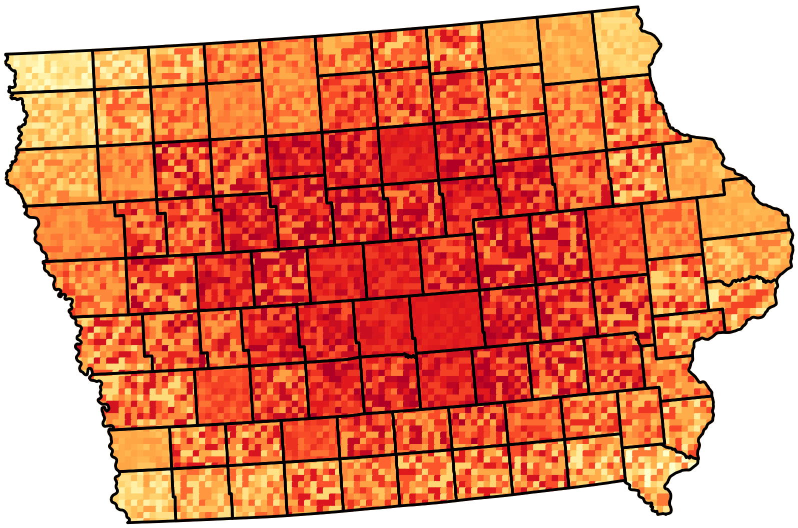


Figure 9: A visualisation of a typical choropleth map, as well as three other maps that are used to display uncertainty on the choropleth map. There is a high variance and low variance example for each type of map to show how well the technique supresses signal. Each map was created using the same data with the same base palette. At first glance, the high uncertainty bivariate and VSUP maps just look like maps with a low saturation colour palette, rather than map with high uncertainty. This visualisation makes it clear that supression methods that plot uncertainty to a second axis, such as hue, make uncertainty appear as a second variable, rather than a signal supression on our estimate.

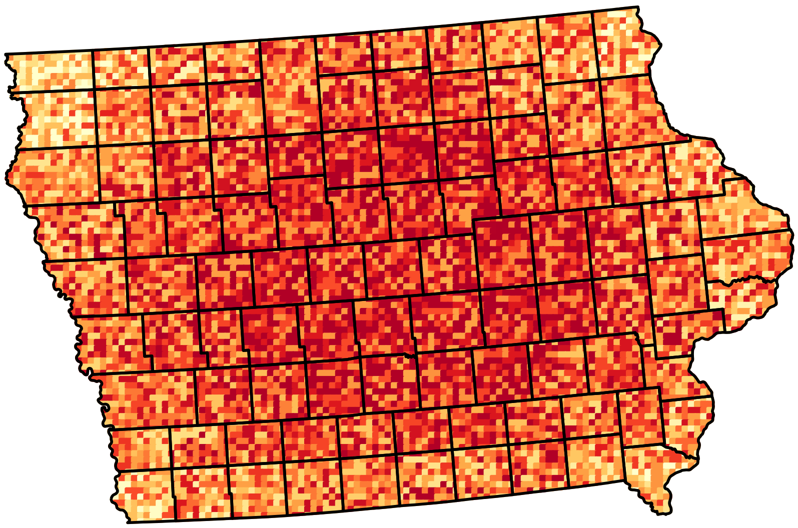


Figure 10: A visualisation of a typical choropleth map, as well as three other maps that are used to display uncertainty on the choropleth map. There is a high variance and low variance example for each type of map to show how well the technique supresses signal. Each map was created using the same data with the same base palette. At first glance, the high uncertainty bivariate and VSUP maps just look like maps with a low saturation colour palette, rather than map with high uncertainty. This visualisation makes it clear that supression methods that plot uncertainty to a second axis, such as hue, make uncertainty appear as a second variable, rather than a signal supression on our estimate.

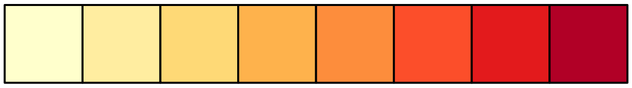


Figure 11: A visualisation of a typical choropleth map, as well as three other maps that are used to display uncertainty on the choropleth map. There is a high variance and low variance example for each type of map to show how well the technique supresses signal. Each map was created using the same data with the same base palette. At first glance, the high uncertainty bivariate and VSUP maps just look like maps with a low saturation colour palette, rather than map with high uncertainty. This visualisation makes it clear that supression methods that plot uncertainty to a second axis, such as hue, make uncertainty appear as a second variable, rather than a signal supression on our estimate.

to extract particular values of signal or noise. Therefore, given that the primary goal of visualisation *is* insights (North 2006), visualisation authors should err on the side of representing suppressed signal as a single variable, rather than visualising uncertainty separately using two different channels.

10.1 Uncertainty visualisations vs information visualisation

If uncertainty is inherently different, we would expect evaluation studies of uncertainty visualisations to produce different results to similar evaluation studies on "normal" variables. This does not seem to be the case. Uncertainty is noise, but most evaluation experiments measure it as signal. The previous section highlights a unique and fascinating problem faced by uncertainty visualisation.

Even though uncertainty does not exist without a signal, that does not mean every visualisation that includes a concept related to uncertainty (such as a visualisation or error or variance) must be designed for signal supression. However, if a visualisation is not designed for signal supression, it should be treated the same as any other variable. These visualisation are not "special" in the sense that signal supression visualisations are.

Hullman et al. (2019) found that the majority of papers evaluate visualisations on performance (approximately 65% of the papers they surveyed) or interpretation and semantics (approximately 17%) and both of these evaluation goals will run into problems because of their conceptualisation of uncertainty.

By far the most common metric used is accuracy, which is used by approximately 36% of evaluation studies (Hullman et al. 2019).

Visualisation authors are almost unanimous in commenting that the "information" in two plots must be the same in order for the visual techniques to be compared (Cleveland and McGill 1984; Kinkeldey, MacEachren, and Schiewe 2014). Kinkeldey, MacEachren, and Schiewe (2014) adopts an existing definition also that suggests two graphics are informationally equivalent if all the information in one plot is inferable from the other and vice-versa, but adds that two plots are computationally equivalent if that information can be extracted from both plots with similar easy and speed. Many visualisation authors ignore this concept and simply compare two visualisations because they both belong to the class of "uncertainty visualisations". Ibrekk and Morgan (1987) compared a 6 visualisations of a PDF, a box plot, a CDF and a mean with a confidence interval on this basis. They found that people are better at extracting the mean from a plot when they are shown a plot that contains a mean with a confidence interval than when they are shown a box plot, or other visualisations that did not allow for the mean to just be read off the plot. Hullman, Resnick, and Adar (2015) compared the static error bars and violin plots of the marginal distributions of two variables (A and B) to an animated plot that depicted outcomes of the joint distribution of A and B in each frame and found that the visualisation of the joint distibution was better at answering questions about the joint distribution than the visualisations of the marginal distribution. Hofman, Goldstein, and Hullman (2020) commented that "theoretically" the sampling distribution of the mean and the prediction interval of a new observation are equal "so long as one knows the sample size", but does not seem to provide participants with that sample size, or recognise the assumptions and background knowledge that would be required to compare the two. Hofman, Goldstein, and Hullman (2020) and Zhang et al. (2022) compared prediction and sampling distributions because they are both "uncertainty" that is typically depicted around the mean. They found that people are better at answering questions about a prediction interval when shown a prediction interval instead of a sampling distribution.

This collection of examples starts to paint a pretty clear picture. Visualisations with rather shocking information asymmetry are regularly compared because they are both "uncertainty visualisations".

If two graphics are visually different but identical in the information they contain, they must differ in how that information is depicted. Once we have a mathematical expression of uncertainty, the visualisation of that uncertainty is theoretically identical to the visualisation process of any other variable. For simple tasks such as value extraction, there is a hierarchy to perceptual tasks where extracting visual information in some forms is easier than others. The hierarchy was originally established 40 years ago by Cleveland and McGill (1984), below is an updated version summarised by Vanderplas, Cook, and Hofmann (2020):

1) Position along a common scale.
2) Position along a non-aligned scale.
3) Length, direction, angle, slope
4) Area
5) Volume, density, curvature
6) Shading, colour saturation, colour hue
7) Discriminable shape
8) Indiscriminable shape

This hierarchy is a good general rule, however it can change from person to person (Davis et al. 2022) Additionally, there are other graphical rules to consider such as gestalt principles, broader methods of extraction, and attention principles (Vanderplas, Cook, and Hofmann 2020). These established visualisation concepts allow us to anticipate the ease with which certain pieces of information will be extracted from a plot. We can use these concepts to understand the computational complexity of a graphic. A bizarre feature of the uncertainty visualisation literature is that it does not work to build upon these existing principles or identify the ways in which uncertainty visualisations may diverge from these rules. These building block concepts of visualisation are seldom mentioned.

It is difficult to find examples of uncertainty visualisation experiments where the plots do contain the same information, however when they do, the results align with existing information visualisation research. Technically, a PDF and a mean with confidence intervals both have enough information to extract the mean of the distribution, however they both have a very different computational cost. To extract the mean using a PDF, a participant would need to identify the point along the x axis that splits the area under the curve in half. If a participant is provided with a mean with a confidence interval, extracting the average is a simple task of reading the position on an aligned scale. Ibrekk and Morgan (1987) found that when asking participants for the "best estimate" (which they thought should be interpreted as the mean of the distribution) of a skewed distribution, participants provided the mean when given a mean with confidence intervals and the mode when given a PDF (Ibrekk and Morgan 1987). Similar results to this occur over and over again in the uncertainty visualisation literature. Gschwandtnei et al. (2016) found that visualisations where the start time of an interval could literally be read off the plot (error bars, centred error bars, and ambiguation) performed better than the plots (accumulated probability, gradient, and violin) where the start time involved some guesswork because the drop off was gradual. Cheong et al. (2016) found that participants were better at answering questions when they were explicitly given the relevant probability in text rather than when they needed to read it off a map. (*Cite Gap 15: Examples of replicated perceptual task experiments*).

These results show that uncertainty is not technically different to any other variable. When trying to anticipate the results of these studies, we can use the same principles of information equivalence and difficulty of relevant mental tasks to understand which plots will outperform others. This does not mean that visualising uncertainty as a signal is incorrect or bad, it just means that uncertainty as a signal is no different from any other information visualisation.

- decision making experiments are usually displaying the wrong signals

The final method used by authors is to just explicitly ask about uncertainty and signal information separately. Sanyal et al. (2009) mapped uncertainty to dots and signal to a 3D surface and asked participants to identify areas of high and low signal and high and low uncertainty. Participants were not asked to combine that information in any way, and the signal and the noise were treated as separate variables. Correll and Gleicher (2014) asked participants to separately extract the mean and variance from four uncertainty visualisations. These methods explicitly view the uncertainty and signal as two separate variables that should be extracted from a plot, and not two variables that should be interpreted together. Even viewing these questions as a routine check to make sure the signal information isn't impacted by the uncertainty is counter intuitive, because the whole point *of* the uncertainty is to impact the signal information.

11 What even is an uncertainty visualisation?

11.1 The literatures definition of uncertainty visualisations

What exactly is an uncertainty visulisation is never explained and certainly does not have a consistent definition. For example Wilkinson (2005) mentions that popular graphics, such as pie charts and bar charts omit uncertainty, however at least one or both of these charts are used in a significant number of uncertainty visualisation experiments (Ibrekk and Morgan 1987; Olston and Mackinlay 2002; Zhao et al. 2023; Hofmann et al. 2012). Wickham and Hofmann (2011) suggests their product plot framework, which includes histograms, should have a way to measure uncertainty, but does not consider that a histogram is *already* a depiction of PDF, something that is often considered an uncertainty visualisation by other authors.

11.1.1 What is an uncertainty visualisation?

Kinkeldey, MacEachren, and Schiewe (2014) categorised uncertainty according to five criteria depicted in Figure 12 which considers if a visualisation is implicit or explicit, intrinsic or extrinsic, visually integrable or seperable, coincident or adjacent, and static or dynamic. A similar version of this taxonomy was presented by L. Padilla, Kay, and Hullman (2022) who commented that visualisation can be organised into two categories, "graphical annotations of distributional properties" and "visual encodings of uncertainty" which seems to functionally align with the intrinsic/extrinsic distinction by Kinkeldey, MacEachren, and Schiewe (2014). Griethe and Schumann (2006) organised uncertainty visualisations into two cases (1) a hypothesis test was preformed to confirm the validity of the visualisation and (2) the visualisation has uncertainty depicted. Kristin Potter, Rosen, and Johnson (2012) organised several existing uncertainty visualisations into groups based on the dimensionality of the data (1D, 2D, 3D, and No Dimension) and the dimensionality of the uncertainty (Scalar, Vector, Tensor). However, because the term "PDF", function that is used to describe a random variable, is used to describe both the data and the uncertainty for all dimensions. Grewal, Goodwin, and Dwyer (2021) created a taxonomy that mapped uncertainty visualisations to some point in a 2D space defined by the "domain expertise" and "continuum of discreteness" (that scaled from "point estimate" to "continuous distribution").

The authors of these taxonomies did not intend for them to be complete descriptions of uncertainty visualisations, however they do paint a picture of the what authors consider to be important. There are common themes in the taxonomies that ensure we differentiate between a sample and a mass, consider if uncertainty should be depicted with the estimate or as a second variable, identify the parallel of hypothesis testing, and consider the dimensionality, sample size and precision depicted by the visualisation. These taxonomies may highlight perceived

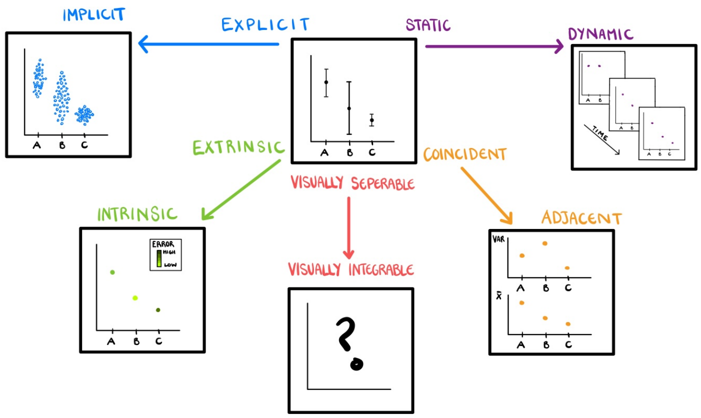


Figure 12: Visualisation of Kinkeldey, MacEachren, and Schiewe (2014) uncertainty visualisation taxonomy. It can be seen that most of the categories can be considered a change to the typical error bar chart. The taxonomy categorises the visualisations based on five criteria, implicit (a sample) or explicit (a depiction of mass); Intrinsic (alter existing symbols symbols to represent uncertainty) or extrinsic (add new objects to represent uncertainty); visually integral or separable (the uncertainty can be separated from the data and read independently) ; coincidence (uncertainty and data are represented in the same plot) or adjacent; static or dynamic (animated, interactive etc). Many of the distinctions conflate changes of the input data and statistics calculated and with visual changes, highlighting how defining uncertainty and visualisation together conflates several elements of a graphic.

important differences, but they do not help us understand what *is* or *is not* an uncertainty visualisation. For that we need to understand *why* uncertainty visualisation exists as a sub-field at all.

11.1.2 What is uncertainty?

Uncertainty is a concept that is famously hard to pin down, with constantly shifting definitions that mean it is quantified differently from author to author.

The ramifications of this is seen throughout the literature, as a survey of visualisation authors cited "not knowing how to calculate uncertainty" as one of the primary reasons they did not include it in visualisations (Hullman 2020). Some works (Hullman et al. 2018; Maceachren et al. 2012; Thomson et al. 2005) get around this definition issue by focusing narrowly on specific terms defined mathematically, such as probability, variance, error, or precision. Others (Griethe and Schumann 2006; Pang, Wittenbrink, and Lodha 1997; Pham, Streit, and Brown 2009; Boukhelifa et al. 2017) include broader loosely related elements, such as missing values, but ignore these concepts when translating that uncertainty to a visualisation. Existing mathematical definition of uncertainty (or related concepts) have been unable to unify this range of diverse concepts. For example, Thomson et al. (2005) suggests a mathematical formula for *examples* of uncertainty, and information theory tries to quantify uncertainty using the idea of entropy, but the disconnect between the broad concept of uncertainty and what we can reliably quantify remains a problem.

Some authors prefer to define uncertainty in qualitative terms, often through the use of taxonomies. In these cases, uncertainty is split using an endless stream of ever changing boundaries, such as whether the uncertainty is due to true randomness or a lack of knowledge (Spiegelhalter 2017; Hullman 2016; Walker et al. 2003), if the uncertainty is in the attribute, spatial elements, or temporal element of the data (Kinkeldey, MacEachren, and Schiewe 2014), whether the uncertainty is scientific (e.g. error) or human (e.g. disagreement among parties) (Benjamin and Budescu 2018), if the uncertainty is random or systematic

(Sanyal et al. 2009), statistical or bounded (Gschwandtnei et al. 2016; Olston and Mackinlay 2002), recorded as accuracy or precision (Griethe and Schumann 2006; Benjamin and Budescu 2018), which stage of the data analysis pipeline the uncertainty comes from (Walker et al. 2003), how quantifiable the uncertainty is (Spiegelhalter 2017; Walker et al. 2003), etc. In their own way, each of these approaches show an aspect of uncertainty that an author felt was important to differentiate.

What is interesting about all these views of uncertainty, is that they all imply that it is a latent feature of our data. That is, it is something that can be revealed through exploration, and not something that needs to be explicitly calculated. The constant stream of new definitions that are unable to capture the true essence of uncertainty might be a result of this false assumption that underpins ever definition. Uncertainty cannot be defined without a signal because uncertainty does not exist without a signal, is a by-product of inference.

11.1.2.1 Uncertainty as a by-product of inference

It is easy to see uncertainties relationship to inference when we consider what might *not* be considered uncertain rather than just trying to think about what *is*.

Descriptive statistics simply describe our sample as it is and summarizes large data down into an easy to swallow format. Descriptive statistics are not seen as the primary goal of modern statistics, however, this was not always the case. Around the 19th century in England, *positivism* was the popular philosophical approach to science (positivists included famous statisticians such as Francis Galton and Karl Pearson) and practitioners of the approach believed statistics ended with descriptive statistics as science must be based on actual experience and observations, therefore anything that refers to the unobservable (such as new observations or population statistics) is not true science (Otsuka 2023). In order to make statements about population statistics, future values, or new observations we need to perform inference, which requires the assumption of the "uniformity of nature" (i.e. that unobserved phenomena should be similar to observed phenomena) (Otsuka 2023). This subtle shift, from descriptive statistics to inferential statistics was shunned during the positivism era *due to the fact it introduced the unknowable*, or in other words, uncertainty.

This approach to uncertainty is embedded in the idea that descriptive statistics do not have uncertainty, which some readers may disagree with. Specifically, because it means uncertainty is *not* a latent attribute of data, but rather an attribute of a specific hypothesis or estimate. Deniers of this fact follow a consistent logical path and it is easy to identify the common mistake. We know that variance and probability are typically considered types of "uncertainty" *and* descriptive statistics can have variance and probabilities *therefore* descriptive statistics must have uncertainty. The flaw in this logic comes from the first step, assuming that the tools with which we measure uncertainty *are* uncertainty in of themselves. This confusion is common and there are many papers that spend a great deal of time clarifying the difference. Begg, Welsh, and Bratvold (2014) highlight that uncertainty is related to not knowing a specific value, while variability refers to the range of values a quantity can take at different locations, times or instances. Spiegelhalter (2017) made sure to comment on the difference between precise random events (such as the probability associated with a coin flip), and uncertainty (such as the estimated probability associated with a coin that might be biased). The variance of a sample variance can be calculated and know, therefore it is not uncertain but rather it a precise description of dispersion. If we were to discuss drawing a new observation, or estimating the true mean of a population *then* the variance would become relevant in our discussions of uncertainty.

Several authors have connected the issues around defining uncertainty to inference, however it is often discussed as a *task* or *goal* dependence. Multiple authors have commented on the need to consider quantifying and expressing uncertainty at every stage of a project as the "goal" shapes every step of the analysis (Kinkeldey, MacEachren, and Schiewe 2014; Hullman 2016; Refsgaard et al. 2007). Otsuka (2023) suggested that the process of observing data to perform statistics is largely dependent on our goals, because the process of boiling real world entities down into probabilistic objects (or "probabilistic kind" as he puts it) depends on the relationship we seek to identify with our data. Meng (2014) commented what is kept as data and what is tossed away is determined by the motivation of an analysis and what was previously noise can be shown to become signal depending on the the question we seek to answer. Kale, Kay, and Hullman (2019) discussed how the choices we make in our analysis impact our outcomes and introduce uncertainty. Carlin and Moreno-Betancur (2023) mentions that each research question can be can be categorised as descriptive, predictive, or causal, each of which has its own appropriate statistical methods and motivation agnostic model selection leads to statistical analysis that is devoid of meaning. Wallsten et al. (1997) argue that the best method for evaluating or combining subjective probabilities depends on the uncertainty the decision maker wants to represent and why it matters. Fischhoff and Davis (2014) looks at uncertainty visualisation for decision making decides that we should have different ways of communicating uncertainty based off what the user is supposed to do with it. The importance of inference when discussing uncertainty is never directly acknowledge, but it is always present. It is clear that the relationship between uncertainty and inference is noticed at every stage in an analysis, however combining these uncertainties into a single "uncertainty" value is near impossible (Spiegelhalter 2017).

With this understanding it becomes clear to see why uncertainty is tied to an endless string of examples in the data analysis pipeline. Uncertainty examples include imputed data, model selection, inherent randomness, biased sampling, etc, not because these things *are* uncertainty, but because they *create* uncertainty when we perform inference. Whether or not these elements are relevant is highly dependent on what statistic you are trying to draw inference on, and by extension, the purpose of your visualisation.

11.2 What is an uncertainty visualisation?

The concept of uncertainty as a by-product of inference, not a latent feature of data, extends naturally to visualisation. There are two primary reasons for making a visualisation, to perform exploratory data analysis (EDA) or communication. Communication involves identifying a signal we want to communicate and designing the visualisation that best conveys that, while EDA involves creating a versatile visualisation without an explicit purpose and using it to extract several signals.

Since EDA is the visual parallel to descriptive statistics, it is performed without an explicit hypothesis which means there *is* no uncertainty in the visualisation. Similarly to our issues with descritive statistics, this is not well understood by the uncertainty visualisation community. Some authors simply confuse varianc for uncertainty. For example K. Potter et al. (2010) aimed to create a summary plot that "concisely presented data with uncertainty information" to create an exploratory visualisation tool that visualised uncertainty. Other authors regonise inference will occur (in some shape or form) and believe uncertainty *should* be visualised but do not recognise *how* uncertainty would be visualised. Hullman and Gelman (2021) argued that there is no such thing as a "model-free" visualisation, therefore visualisation require robust visualisations of uncertainty as we are always performing inference. Griethe and Schumann (2006) commented that "if visualization is used as a means to explore a data volume or to communicate its contents the uncertainty has to be included". While we agree people cannot prevent themselves from performing inference, that does not mean it is possible for uncertainty to be included in an EDA visualisation. A versatile visualisation such as a scatter plot allows for a viewer to consider several hypothesis at once, each of which will be a different inferential statistic with a different distribution that depicts its uncertainty. It is likely impossible to suppress all possible signals at once.

12 How do we evaluate uncertainty visualisations?

If asking direct questions about uncertainty causes us to treat it as a signal, how do we evaluate uncertainty as *noise*? When we ask the viewer of a plot to look at data and extract a value, we are asking them to perform inference on that value. There will be noise associated with that answer and that is uncertainty. If we ask direct question about some uncertainty metric, we have turned the uncertainty into signal because that is what the participants are drawing inference on.

12.1 Current attempts to measure uncertainty

There is also a swath of studies that are aware a question that boils down to a value extraction experiment, or a question that should be answered by literally ignoring uncertainty information is not what we want when we consider uncertainty visualisations. These papers often try to ask a question that should utilise both the uncertainty and signal in the response, however this is rarely what actually occurs. This method typically results in in cryptic or confusing questions that create a large amount of noise on the interpretation side of the analysis (Hullman 2016).

Some authors opt for asking slightly vague questions that imply a use of uncertainty, but compare it to a ground truth that is very specific. Ibrekk and Morgan (1987) asked participants for the "best estimate" which was evaluated in accuracy by comparing it to the mean, however the "best estimate" depends on the loss function we are using, and a loss function of minimised error was not implied by the question. Hofmann et al. (2012) showed two distributions in 20 different visualisations (a line-up protocol) using a jittered sample, a density plot, a histogram, and a box plot and asked participants. Participants were asked to report in which of the plots was "the blue group furthest to the right" The experiment set up is shown in Figure 13. The participants answers were then compared to a ground truth where the correct plot had a blue distribution with a right shifted mean. By comparing the results to a ground truth statistic and marking participants as "wrong" or "right", the error from the participants that had an alternative interpretation to the concept of "furthest right" was conflated with the error from a the visualisation choice. These papers make it unclear if the participants got the answers wrong because they misunderstood the question or because of something related to the plot. Therefore, this method leads to inconclusive results about the plot design, and is not advised.

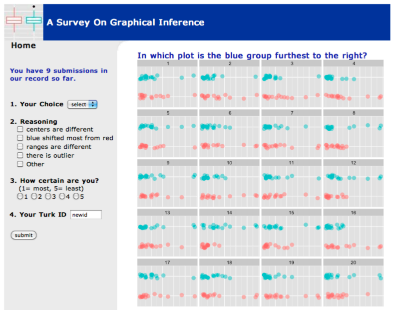


Figure 13: This shows the user interface for the experiment performed by Hofmann et al. (2012). The question of "furthest to the right" is open to interpretation.

Another method used by authors is to ask a deterministic question about a random event. L. M. Padilla, Ruginski, and Creem-Regehr (2017) provided participants with a visualisation of the cone of uncertainty and asked then to "decide which oil rig will receive more damage based on the depicted forecast of the hurricane path". The cone of uncertainty provides a 60% confidence interval for the location of the eye of a hurricane, which allows us to know the area where the eye of the storm will go, it does not given any information about the intensity of a storm, the size of a storm, or even if a location will be hit. This inclusion of determinism seems to cause the authors to stumble themselves, as they are not consistent with their assumptions. In their first experiment L. M. Padilla, Ruginski, and Creem-Regehr (2017) indicated the correct answer was to assume that the storm was equally intense no matter how far from the centre of the distribution an oil rig was, however answering their third experiment correctly hinged on assuming the intensity of the storm at a particular point (which in this experiment they phrased as damage) *does* change in intensity as you move away from the centre of the distribution. Given these conflicting assumptions, it is unclear how the participants were supposed to adjust the probabilistic path information to answer a deterministic question about which oil rig would receive the most damage. Other authors have commented on the complexity of communicating hurricane risk because the path, storm surge and wind speed are all important and cannot be ignored (Spiegelhalter 2017). The flip side of this is asking participants for a deterministic answer to a probabilistic question. Correll and Gleicher (2014) asked participants "how likely is candidate B to win the election?" when the two distributions indicated voter preference. Participants were not able to answer the question about likelihood in term of probability, but were instead given seven options from 1=Outcome will be most in favour of A to 7=Outcome will be most in favour of B. The ground truth statistic for this question was a scalar multiple of Cohen's d, indicating participants were supposed to incorporate uncertainty information using a very specific formula that was likely unknown to them but assumed to be used implicitly.

These examples are a bit complete mishmash of methods, however they point to a larger issue that goes beyond decision making, trust and confidence experiments. Authors *have no idea* how to evaluate the effects of uncertainty in an uncertainty visualisation.

12.2 How to test uncertainty as noise

So, the current methods of measuring or understanding the role of uncertainty in a visualisation is questionable at best, however this is not because visualisation authors are missing the mark, but rather uncertainty is *particularly* difficult to express in a visualisation. In simple estimates or verbal communication, the signal is often easy to identify because it is what we are explicitly saying. Unlike statistical models, visualisations are used in both data exploration and communication. This means what exactly is a *signal* in any particular visualisation is hard to identify, since we often let the visualisation *tell us* what the signal is. Additionally, you cannot add noise to *every single possible* signal one might take from a visualisation. Two people looking at the same visualisation might, just by chance, develop two entirely different insights and draw inference on two completely different statistics. These unique and fascinating challenges that are faced by uncertainty visualisation have been completely untouched by the literature. This section will cover some interesting research in uncertainty visualisation and suggestions for better ways to measure uncertainty.

12.2.1 Qualitative Studies

Alternatively visualisation research could shift away from the accuracy concept all together ask questions that allow for open ended responses. This method can enlighten authors as to *how* the uncertainty information was used by the participants. Hofmann et al. (2012) tried to capture this by asking participants why they considered a particular plot to be more “right shifted”, however this qualitative assessment does not seem to have made it into the final paper. Daradkeh (2015) presented participants with ten investment alternatives and asked participants “from among available alternatives, which alternative do you prefer the most”, and were asked to think aloud and consider the uncertainty in their decision making. The experimenters goal was to observe and organise the methods people use when making decisions in the face of uncertainty. This study was an excellent example in a useful experimental design. They highlighted the specific aspects of uncertainty that participants typically considered, such as the range of outcomes that are above/below a certain threshold, minimum and maximum values, the risk of a loss, etc, and mapped where in the decision making process participants made these considerations. Data visualisation is commonly utilised as a tool in data exploration, so it is not uncommon for a data analyst to make a plot with only a vague goal and pull out a large number of adjacent observations. This experimental framework could replicate this process.

12.2.2 Just noticeable signal

It could be argued that a well done uncertainty visualisations should have an imperceptible signal unless the signal would be identified with a hypothesis test, almost like a reverse line up protocol, but this idea also has some issues that should be considered. The reject or do not reject concepts in hypothesis testing do not offer a complete image of uncertainty, and exploration of uncertainty visualisation largely stems from a desire to move away from this binary framework.

Additionally, setting up an uncertainty visualisation where the participants are expected to notice the signal once the data

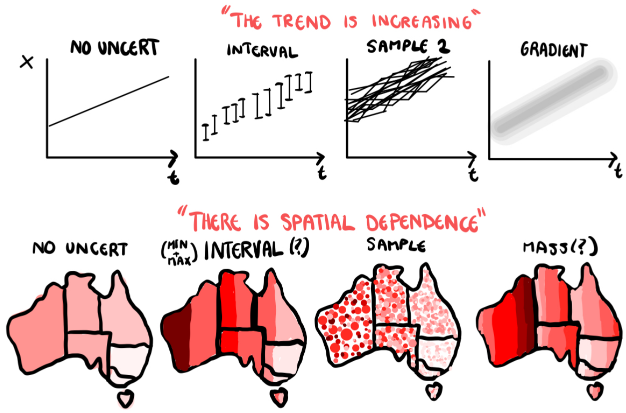


Figure 14: The visualisation showing the kind of signal that we would expect an uncertainty plot to supress

behind the visualisation passes a hypothesis test implies the signal *is* noticeable to a human at that level. (**Patrick2023?**) compared people ability to recognise patterns in a residual plot to typical statistical tests and found human viewers looking at a plot were less sensitive than the typical residual tests. These experiments utilised the line-up protocol which has been suggested as a method to check if perceived patterns are real or merely the result of chance (Buja et al. 2009; Wickham et al. 2010; Chowdhury et al., n.d.). This concept bears similarity to the goal of uncertainty visualisation, but it is not quite the same. Figure 15 shows the conceptual difference between the line-up protocol and uncertainty visualisation.

12.3 Other elements of uncertainty visualisations

- Things that are true for both normal visualisation and uncertainty, but of particular interest for uncertainty

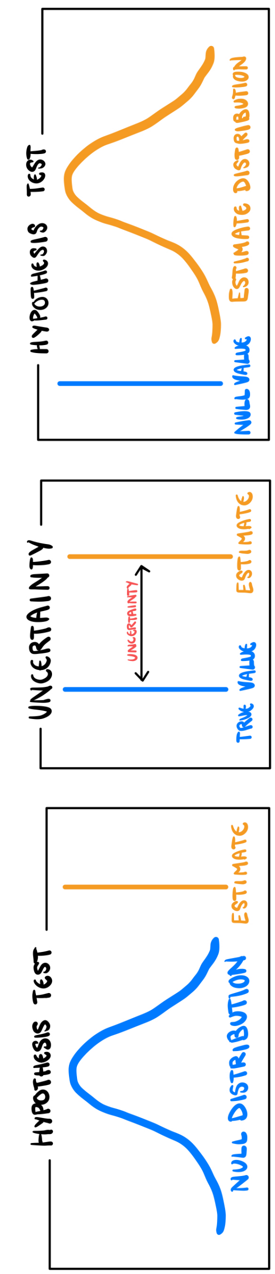


Figure 15: A line-up protocol displays the uncertainty about the null and identifies if the true data plot is identifiable (and therefore significantly difference). This can be considered the graphical equivalent of a standard hypothesis test. Uncertainty is frequently used to descibe the area around the estimate that we think might contain the true value, and we assesses if the null of “no signal” is within this plot (e.g. if the error bars overlap with zero). Uncertainty is technically the distance between the true value and

12.3.1 Heuristics

Additionally, these heuristics and biases can change depending on the larger scope of the graphic and the population we are communicating with (Spiegelhalter 2017; Kinkeldey, MacEachren, and Schiewe 2014).

Heuristic checks are useful because they look at unknown pitfalls that might exist in interpretation of current plots. Since the hypothesis for these experiments are usually quite specific, e.g. "do people perceive a an outcome that is within the bar as more likely than one outside it, even if both outcomes are the same distance from the mean?". This means they are less likely to fall into the trap of trying to answer questions that are *far* too broad to be answered with a single experiment (e.g. "is a scatter plot better at showing uncertainty than a box plot"). This work also provides useful insights for experiments by highlights pitfalls participants might fall into when they review the results of evaluation experiments (Hullman 2016). Newman and Scholl (2012) found that participants were more likely to view points within the bar as more likely than points outside of the bar in bar charts with error bars. Similar effects have been identified in other types of uncertainty displayed. L. M. Padilla, Ruginski, and Creem-Regehr (2017) found that points that were on an outcome of an ensemble display were perceived as more likely than points not on an outcome, even when the point that was not on a specific outcome of the ensemble was closer to the mean of the uncertainty distribution. The sine illusion can cause the confidence interval of a smoothed sine curve to seem wider at the peaks than the troughs, causing us to underestimate uncertainty associated with changing values (Vanderplas and Hofmann 2015).

12.4

In a similar vein, experiments that verify smaller aspects of plot design might be more useful to the field in the long run because it helps contribute to a larger working theory of "how do we see visualisations". Many visualisation experiments try to compare two plots with several differences, but do not seem to be interested in the mechanisms by which we extract information from visualisations. Small perceptual tasks that seek to answer small but highly relevant questions (for example, if colour hue and colour value can be perceived as a single signal suppressed variable) would be useful to the field.

13 Future work

This paper has identified gaps in the uncertainty visualisation literature that could be filled to progress the field.

Each new development should be accompanied by a mathematical definition of the uncertainty being addressed. Ideally, a mathematically definition of uncertainty that allows us to combine these components would be developed, but in the absence of that, authors should be more specific about what aspect of "uncertainty" they are covering with their visualisation.

The concept of uncertainty should be formalised within the grammar of graphics. This formalisation would allow uncertainty visualisation authors to have a clear understanding of what is or is not an uncertainty visualisation. Additionally placing uncertainty visualisation in the framework that is used to understand existing information visualisation research would help authors understand when existing methods can be used to explain their results. incorporating uncertainty into the grammar of graphics will also give a more precise concept of the information contained within a plot. Other fields of science employ marginal changes when designing experiments to ensure it is well understood *what* aspect of their experiment is contributing to their results, and a better sense of what "marginal" is in the case of uncertainty visualisation would greatly help the field. (*XXX Is data pipeline connected with the grammar of graphics? Should this be a recommendation?*)

Experimental practices on uncertainty visualisation need to be standardised. If we are going to consider uncertainty as noise, not signal, there needs to be a way to identify this signal suppression in an experimental design. As the literature currently exists, there is no way to combine papers to get a meaningful sense of how uncertainty information is understood by a viewer.

There is also the possibility that uncertainty visualisation evaluations will need to swap to a qualitative methodology where participants are allowed to freely comment on what they notice in graphics until we establish how the existence of noise can be observed.

If an uncertainty visualisation researcher would prefer to perform experiments rather than formalise methods, there are options there too. It would be interesting to know if any perceptual tasks that can be mapped to two different visual tasks condense into a single dimension when looking for overarching signal in a plot. Alternatively, the task dependency many authors in uncertainty visualisation mention would be a useful direction to consider. It is clear that the the number of potential tasks that can be performed on a visualisation increases with with the number of observations. A single observation is limited to value extraction, two observations can be compared, multiple observations allow for shapes or global statistics to be extracted. The interaction between sample size and task is of particular interest to the uncertainty visualisation community, as uncertainty can be expressed through multiple observations using a sample, or through a single value using an error. Of course, this is limited by the fact that there also isn't a definition for what is a "task" and given the mess created by the lack of formalisation in uncertainty visualisation, it may be wise to formalise that concept before performing these experiments. Amar, Eagan, and Stasko (2005) suggested a taxonomy for information visualisation based on the types of tasks we use visualisations for and suggest 10 "analytical primitives" that we can then map to visualisations, which could be a good starting point. Regardless, these are directions of research would be fruitful to the uncertainty visualisation community even if it appears on the surface to be research that is only beneficial to the "normal" visualisation community. (*XXX Not sure what this paragraph is recommending?*)

Bibliography

Amar, Robert, James Eagan, and John Stasko. 2005. "Low-level components of analytic activity in information visualization." *Proceedings - IEEE Symposium on Information Visualization, INFO VIS*, 111–17. https://doi.org/10.1109/INFVIS.2005.1532136.

Anscombe, F. J. 1973. "Graphs in Statistical Analysis." *The American Statistician* 27 (1): 17–21. https://www.tandfonline.com/doi/abs/10.1080/00031305.1973.10478966.

Begg, Steve H., Matthew B. Welsh, and Reldar B. Bratvold. 2014. "Uncertainty vs. variability: What's the difference and why is it important?" *SPE Hydrocarbon Economics and Evaluation Symposium*, no. May: 273–93. https://doi.org/10.2118/169850-ms.

Benjamin, Daniel M., and David V. Budescu. 2018. "The role of type and source of uncertainty on the processing of climate models projections." *Frontiers in Psychology* 9 (MAR): 1–17. https://doi.org/10.3389/fpsyg.2018.00403.

Blenkinsop, Steve, Pete Fisher, Lucy Bastin, and Jo Wood. 2000. "Evaluating the perception of uncertainty in alternative visualization strategies." *Cartographica* 37 (1): 1–13. https://doi.org/10.3138/3645-4v22-0m23-3t52.

Boone, Alexander P., Peri Gunalp, and Mary Hegarty. 2018. "Explicit versus actionable knowledge: The influence of explaining graphical conventions on interpretation of hurricane forecast visualizations." *Journal of Experimental Psychology: Applied* 24 (3): 275–95. https://doi.org/10.1037/xap0000166.

Boukhelifa, Nadia, 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): 2769–78. https://doi.org/10.1109/TVCG.2012.220.

Boukhelifa, Nadia, Marc Emmanuel Perrin, Samuel Huron, and James Eagan. 2017. "How data workers cope with uncertainty: A task characterisation study." *Conference on Human Factors in Computing Systems - Proceedings* 2017-May (May): 3645–56. https://doi.org/10.1145/3025453.3025738.

Buja, Andreas, Dianne Cook, Heike Hofmann, Michael Lawrence, Eun Kyung Lee, Deborah F. Swayne, and Hadley Wickham. 2009. "Statistical inference for exploratory data analysis and model diagnostics." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367 (1906): 4361–83. https://doi.org/10.1098/rsta.2009.0120.

Carlin, John B., and Margarita Moreno-Betancur. 2023. "On the uses and abuses of regression models: a call for reform of statistical practice and teaching." http://arxiv.org/abs/2309.06668.

Carr, Daniel B., Anthony R. Olsen, and Denis White. 1992. "Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data." *Cartography and Geographic Information Systems* 19 (4): 228–36. https://doi.org/10.1559/152304092783721231.

Cheong, Lisa, Susanne Bleisch, Allison Kealy, Kevin Tolhurst, Tom Wilkening, and Matt Duckham. 2016. "Evaluating the impact of visualization of wildfire hazard upon decision-making under uncertainty." *International Journal of Geographical Information Science* 30 (7): 1377–1404. https://doi.org/10.1080/13658816.2015.1131829.

Chowdhury, Niladri Roy, Dianne Cook, Heike Hofmann, Mahbubul Majumder, Eun-Kyung Lee, and Amy L Toth. n.d. "Using Visual Statistical Inference to Better Understand Random Class Separations in High Dimension, Low Sample Size Data."

Cleveland, William S., and Robert McGill. 1984. "Graphical perception: Theory, experimentation, and application to the development of graphical methods." *Journal of the American Statistical Association* 79 (387): 531–54. https://doi.org/10.1080/01621459.1984.10478080.

Correll, Michael, and Michael Gleicher. 2014. "Error bars considered harmful: Exploring alternate encodings for mean and error." *IEEE Transactions on Visualization and Computer Graphics* 20 (12): 2142–51. https://doi.org/10.1109/TVCG.2014.2346298.

Correll, Michael, Dominik Moritz, and Jeffrey Heer. 2018. "Value-suppressing uncertainty Palettes." *Conference on Human Factors in Computing Systems - Proceedings* 2018-April: 1–11. https://doi.org/10.1145/3173574.3174216.

Daradkeh, Mohammad. 2015. "Exploring the use of an information visualization tool for decision support under uncertainty and risk." *ACM Inter-*

national Conference Proceeding Series 24-26-Sept. https://doi.org/10.1145/2832987.2833050.

Davis, Russell, Xiaoying Pu, Yiren Ding, Brian D. Hall, Karen Bonilla, Mi Feng, Matthew Kay, and Lane Harrison. 2022. "The Risks of Ranking: Revisiting Graphical Perception to Model Individual Differences in Visualization Performance." *IEEE Transactions on Visualization and Computer Graphics* PP: 1–16. https://doi.org/10.1109/TVCG.2022.3226463.

Fischhoff, Baruch, and Alex L. Davis. 2014. "Communicating scientific uncertainty." *Proceedings of the National Academy of Sciences of the United States of America* 111: 13664–71. https://doi.org/10.1073/pnas.1317504111.

Grewal, Yashvir, Sarah Goodwin, and Tim Dwyer. 2021. "Visualising Temporal Uncertainty: A Taxonomy and Call for Systematic Evaluation." *IEEE Pacific Visualization Symposium* 2021-April (April): 41–45. https://doi.org/10.1109/PACIFICVIS52677.2021.00013.

Griethe, Henning, and Heidrun Schumann. 2006. "The Visualization of Uncertain Data: Methods and Problems." *Proceedings of SimVis '06* vi (August): 143–56. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Visualization+of+Uncertain+Data:+Methods+and+Problems#0.

Gschwandtnei, Theresia, Markus Bögl, Paolo Federico, and Silvia Miksch. 2016. "Visual Encodings of Temporal Uncertainty: A Comparative User Study." *IEEE Transactions on Visualization and Computer Graphics* 22 (1): 539–48. https://doi.org/10.1109/TVCG.2015.2467752.

Hofman, Jake M., Daniel G. Goldstein, and Jessica Hullman. 2020. "How Visualizing Inferential Uncertainty Can Mislead Readers about Treatment Effects in Scientific Results." *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3313831.3376454.

Hofmann, Heike, Lendie Follett, Mahbubul Majumder, and Dianne Cook. 2012. "Graphical Tests for Power Comparison of Competing Designs." http://www.public.iastate.edu/.

Hullman, Jessica. 2016. "Why evaluating uncertainty visualization is error prone." *ACM International Conference Proceeding Series* 24-October: 143–51. https://doi.org/10.1145/2993901.2993919.

———. 2020. "Why Authors Don't Visualize Uncertainty." *IEEE Transactions on Visualization and Computer Graphics* 26 (1): 130–39. https://doi.org/10.1109/TVCG.2019.2934287.

Hullman, Jessica, and Andrew Gelman. 2021. "Designing for Interactive Exploratory Data Analysis Requires Theories of Graphical Inference." *Harvard Data Science Review*, 1–70. https://doi.org/10.1162/99608f92.3ab8a587.

Hullman, Jessica, Matthew Kay, Yea Seul Kim, and Samana Shrestha. 2018. "Imagining Replications: Graphical Prediction Discrete Visualizations Improve Recall Estimation of Effect Uncertainty." *IEEE Transactions on Visualization and Computer Graphics* 24 (1): 446–56. https://doi.org/10.1109/TVCG.2017.2743898.

Hullman, Jessica, Xiaoli Qiao, Michael Correll, Alex Kale, and Matthew Kay. 2019. "In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation." *IEEE Transactions on Visualization and Computer Graphics* 25 (1): 903–13. https://doi.org/10.1109/TVCG.2018.2864889.

Hullman, Jessica, 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). https://doi.org/10.1371/journal.pone.0142444.

Ibrekk, Harald, and M. Granger Morgan. 1987. "Graphical Communication of Uncertain Quantities to Nontechnical People." *Risk Analysis* 7 (4): 519–29. https://doi.org/10.1111/j.1539-6924.1987.tb00488.x.

Kale, Alex, Matthew Kay, and Jessica Hullman. 2019. "Decision-making under uncertainty in research synthesis: Designing for the garden of forking paths." *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290605.3300432.

Kale, Alex, Francis Nguyen, Matthew Kay, and Jessica Hullman. 2018. "Hypothetical Outcome Plots Help Untrained Observers Judge Trends in Ambiguous Data." *IEEE Transactions on Visualization and Computer Graphics* 25 (1): 892–902.

Kinkeldey, Christoph, Alan M. MacEachren, and Jochen Schiewe. 2014. "How to assess visual communication of uncertainty? a systematic review of geospatial uncertainty visualisation user studies." *Cartographic Journal* 51 (4): 372–86. https://doi.org/10.1179/1743277414Y.0000000099.

Locke, Steph, and Lucy D'Agostino McGowan. 2018. *datasauRus: Datasets from the Datasaurus Dozen.* https://CRAN.R-project.org/package=datasauRus.

MacEachren, Alan M. 1992. "cartographic perspectives Visualizing Uncertain Information." *Cartographic Perspectives*, no. 13: 10–19.

Maceachren, Alan M., 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): 2496–2505. https://doi.org/10.1109/TVCG.2012.279.

Manski, Charles F. 2020. "The lure of incredible certitude." *Economics and Philosophy* 36 (2): 216–45. https://doi.org/10.1017/S0266267119000105.

Meng, Xiao Li. 2014. "A trio of inference problems that could win you a nobel prize in statistics (if you help fund it)." *Past, Present, and Future of Statistical Science*, 537–62. https://doi.org/10.1201/b16720-52.

Ndlovu, Akim, Hilson Shrestha, and Lane T Harrison. 2023. "Taken by Surprise? Evaluating How Bayesian Surprise & Suppression Influences Peoples' Takeaways in Map Visualizations." In *2023 IEEE Visualization and Visual Analytics (VIS)*, 136–40. IEEE.

Newman, George E., and Brian J. Scholl. 2012. "Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias." *Psychonomic Bulletin and Review* 19 (4): 601–7. https://doi.org/10.3758/s13423-012-0247-5.

North, Chris. 2006. "Toward measuring visualization insight." *IEEE Computer Graphics and Applications* 26 (3): 6–9. https://doi.org/10.1109/MCG.2006.70.

O'Neill, Onora. 2018. "Linking Trust to Trustworthiness." *International Journal of Philosophical Studies* 26 (2): 293–300. https://doi.org/10.1080/09672559.2018.1454637.

Olston, C., and J. D. Mackinlay. 2002. "Visualizing data with bounded uncertainty." *Proceedings - IEEE Symposium on Information Visualization, INFO VIS* 2002-Janua: 37–40. https://doi.org/10.1109/INFVIS.2002.1173145.

Otsuka, Jun. 2023. *Thinking About Statistics: The Philosoph-*

ical Foundations. 1st ed. New York: Routledge. https://doi.org/10.4324/9781003319061.

Padilla, Lace M., Ian T. Ruginski, and Sarah H. Creem-Regehr. 2017. "Effects of ensemble and summary displays on interpretations of geospatial uncertainty data." *Cognitive Research: Principles and Implications* 2 (1). https://doi.org/10.1186/s41235-017-0076-1.

Padilla, Lace, Matthew Kay, and Jessica Hullman. 2022. "Computational Statistics in Data Science." In, 405–26. John Wiley & Sons.

Pang, Alex T., Craig M. Wittenbrink, and Suresh K. Lodha. 1997. "Approaches to uncertainty visualization." *Visual Computer* 13 (8): 370–90. https://doi.org/10.1007/s003710050111.

Pham, Binh, Alex Streit, and Ross Brown. 2009. "Visualization of information uncertainty: Progress and challenges." In *Advanced Information and Knowledge Processing*, 36:19–48. Springer-Verlag London Ltd. https://doi.org/10.1007/978-1-84800-269-2_2.

Potter, K., J. Kniss, R. Riesenfeld, and C. R. Johnson. 2010. "Visualizing summary statistics and uncertainty." *Computer Graphics Forum* 29 (3): 823–32. https://doi.org/10.1111/j.1467-8659.2009.01677.x.

Potter, Kristin, Paul Rosen, and Chris R. Johnson. 2012. "From quantification to visualization: A taxonomy of uncertainty visualization approaches." *IFIP Advances in Information and Communication Technology* 377 AICT: 226–47. https://doi.org/10.1007/978-3-642-32677-6_15.

Refsgaard, Jens Christian, Jeroen P. van der Sluijs, Anker Lajer Højberg, and Peter A. Vanrolleghem. 2007. "Uncertainty in the environmental modelling process - A framework and guidance." *Environmental Modelling and Software* 22 (11): 1543–56. https://doi.org/10.1016/j.envsoft.2007.02.004.

Sanyal, Jibonananda, Song Zhang, Gargi Bhattacharya, Phil Amburn, and Robert J. Moorhead. 2009. "A user study to compare four uncertainty visualization methods for 1D and 2D datasets." *IEEE Transactions on Visualization and Computer Graphics* 15 (6): 1209–18. https://doi.org/10.1109/TVCG.2009.114.

Spiegelhalter, David. 2017. "Risk and uncertainty communication." *Annual Review of Statistics and Its Application* 4:

31–60. https://doi.org/10.1146/annurev-statistics-010814-020148.

Sterzik, Anna, Monique Meuschke, Douglas W. Cunningham, and Kai Lawonn. 2023. "Perceptually Uniform Construction of Illustrative Textures." http://arxiv.org/abs/2308.03644.

Thomson, Judi, Elizabeth Hetzler, Alan MacEachren, Mark Gahegan, and Misha Pavel. 2005. "A typology for visualizing uncertainty." *Visualization and Data Analysis 2005* 5669 (March 2005): 146. https://doi.org/10.1117/12.587254.

Vanderplas, Susan, Dianne Cook, and Heike Hofmann. 2020. "Annual Review of Statistics and Its Application Testing Statistical Charts: What Makes a Good Graph?" https://doi.org/10.1146/annurev-statistics-031219-041252.

Vanderplas, Susan, and Heike Hofmann. 2015. "Signs of the Sine Illusion — Why We Need to Care Signs of the Sine Illusion — Why We Need to Care" 8600. https://doi.org/10.1080/10618600.2014.951547.

Walker, W. E., P. Harremoes, J Rotmans, J. P. Van Der Sluijs, M. B. A. Van Asselt, P Janssen, and M. P. Krayer Von Krauss. 2003. "Defining Uncertainty." *Integrated Assessment* 4 (1): 5–17. https://www.narcis.nl/publication/RecordID/oai:tudelft.nl:uuid:fdc0105c-e601-402a-8f16-ca97e9963592.

Wallsten, Thomas S., David V. Budescu, Ido Erev, and Adele Diederich. 1997. "Evaluating and combining subjective probability estimates." *Journal of Behavioral Decision Making* 10 (3): 243–68. https://doi.org/10.1002/(sici)1099-0771(199709)10:3%3C243::aid-bdm268%3E3.0.co;2-m.

Wickham, Hadley, Dianne Cook, Heike Hofmann, and Andreas Buja. 2010. "Graphical inference for infovis." *IEEE Transactions on Visualization and Computer Graphics* 16: 973–79. https://doi.org/10.1109/TVCG.2010.161.

Wickham, Hadley, and Heike Hofmann. 2011. "Product plots." *IEEE Transactions on Visualization and Computer Graphics* 17 (12): 2223–30. https://doi.org/10.1109/TVCG.2011.227.

Wilkinson, Leland. 2005. *The Grammar of Graphics (Statistics and Computing)*. Berlin, Heidelberg: Springer-Verlag.

Zhang, Sam, Patrick Ryan Heck, Michelle Meyer, Christopher

F Chabris, Daniel G Goldstein, and Jake M Hofman. 2022. "An Illusion of Predictability in Scientific Results."

Zhao, Jieqiong, Yixuan Wang, Michelle V. Mancenido, Erin K. Chiou, and Ross Maciejewski. 2023. "Evaluating the Impact of Uncertainty Visualization on Model Reliance." *IEEE Transactions on Visualization and Computer Graphics* PP (X): 1–15. https://doi.org/10.1109/TVCG.2023.3251950.