The Noisy Work of Uncertainty Visualisation Research: A Review

Harriet Mason

2024-08-19

Table of contents

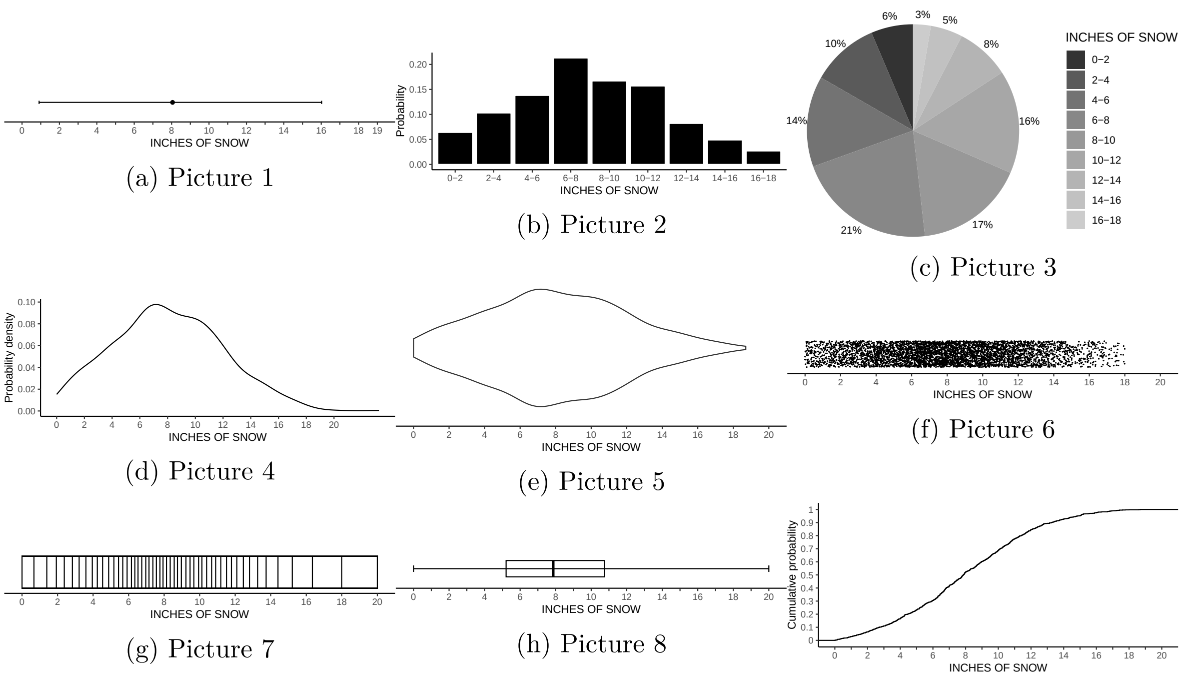
1 Background

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; Goldstein and Rothschild 2014).

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 1, 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 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 con-



(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

Figure 1: A replication of the the uncertainty visualisations shown by Ibrekk and Morgan (1987) in one of the earliest uncertainty visualisation experiments. This early experiment is a good example of many of the issues that are still common in uncertainty visualisation today. For example, the '95% confidence interval' is more accurately a '95% prediction interval'. Additionally graphics that depict different mathematical objects that are also different on their visual components are compared because of a percieved relation to uncertainty. So visualisations of the mean, PDF and CDF, are all discussed as though they all contain the relevant statistical information. The axis also have different scales, and visualisation methods that are now unpopular for displaying proportions, such as a pie chart, are used.

tribute 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.

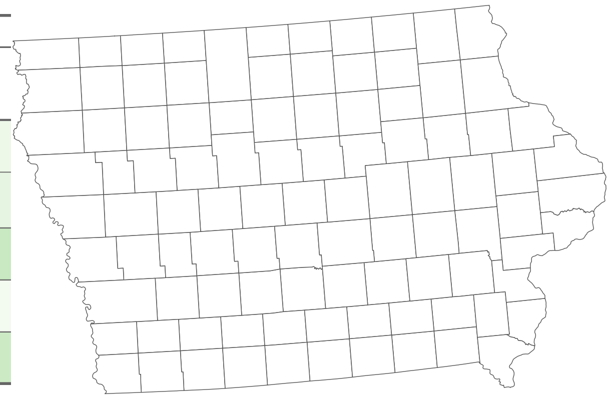
1.1 Spatial examples

There is an endless number of uncertainty visualisations and it would be impossible to show all of them in this review. While the theoretical approach we outline is useful to all uncertainty visualisations, regardless of their application, we feel that showing the variations on a single visualisation will help isolate the ideas we are trying to convey. 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 highlight the costs and benefits of each approach.

Figure 2 shows the first six rows of our example data, along with the geographical boundaries of our data set. The temperature variable was generated using a quadratic spatial trend, that is $Temperature = 28 - SLongitude^2 + SLatitude^2$ where the $SLongitide$ and $SLatitude$ are the scaled (to standard normal) longitude and latitude of the county's centroid. The high and low cases for the sampling variance are randomly sampled from a uniform distribution. The low standard errors are drawn from a $U_{[0,1]}$ distribution, while the high standard error cases are drawn from a a $U_{[1,2]}$ distribution. As we are dealing with an average, the sampling distribution would be normally distributed, so each county temperature estimate is assumed to come from a $N(Temp_i, SE_{case,i})$. This is the data we will be using in our spatial uncertainty examples for the rest of the paper.

			Standard Error	
ID	County	Average Temperature (°C)	High	Low
1	Hancock County	26.46	1.68	0.24
2	Wright County	27.31	1.61	0.40
3	Greene County	27.64	1.53	0.84
4	Cherokee County	25.54	1.22	0.11
5	Cedar County	25.72	1.14	0.78

(a) Data Table



(b) Map Boundaries

Figure 2: The the first 5 observations of the data used for the spatial uncertainty examples along with the boundaries of each county. The map boundaries are the Iowa county boundaries, however the 'temperature' data is not representative of the average temperature in Iowa. The temperature and standard error represent the average of the daily high temperature and the standard error of that average respectively.

2 Ignoring uncertainty

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

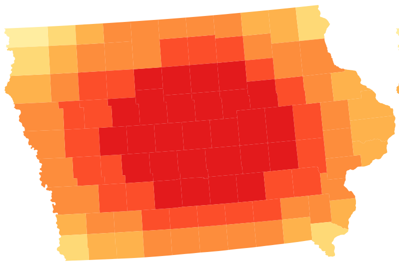
2.1 What is uncertainty

2.2 The choropleth map

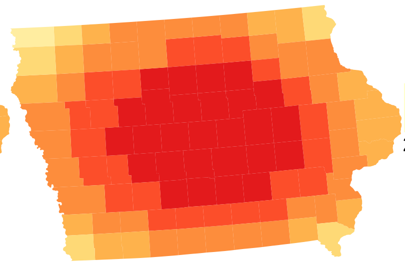
Figure 3 depicts a choropleth map of the counties of Iowa. Each of these counties are colored according to an estimate of average daily temperature that was simulated so that the values followed a clear spatial trend (hot in the middle of the map, and cold on the outside). The variance of these estimates 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?

2.3 Signal-supression

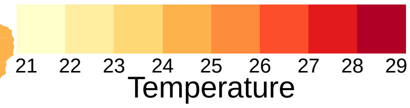
The two choropleth maps appearing to be identical in Figure 3 highlights the need for uncertainty visualisations. Uncertainty visualisation is is required for transparency and this sentiment has been repeated. Some authors suggest uncertainty is important to include as it



(a) Low Variance Data



(b) High Variance Data



(c) Choropleth Palette

Figure 3: 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.

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. Additionally, since the purpose of a visualisation is to give a quick "gist" of the information (Spiegelhalter 2017), this hedging needs to be communicated visually without any need for computation from the viewer. Additionally, Ndlovu, Shrestha, and Harrison (2023) found that participants applied the same methods they used for simple choropleth maps to complicated uncertainty maps even if that take away was invalid. Therefore, this hedging effect needs to be communicated simply through the visualisation. If we refer to the conclusion we draw from a graphic to be its "signal" and the variance that makes this signal harder to identify as the "noise", we can summaries this information into three key requirements. A good uncertainty visualisation needs to:

1) Reinforce justified signals to encourage confidence in results
2) Hide signals that are just noise to prevent unjustified conclusions

3) Perform tasks 1) and 2) in a way that is proportional to the level of confidence in those conclusions.

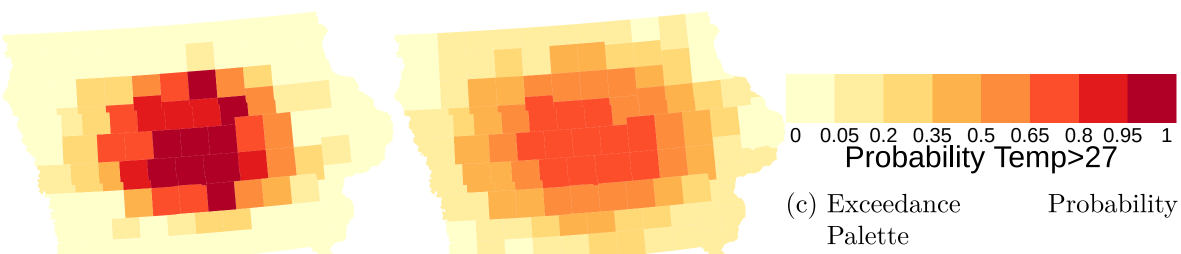
As Figure 3 showed, visualisations that are unconcerned with uncertainty have no issue showing justified signals, but struggle with the display of unjustified signals. Therefore, we coin this approach to uncertainty visualisation as "signal-suppression" since it primarily differentiates itself from from the "noiseless" visualisation approach through criteria (2). That is, the main difference between an uncertainty visualisation and a "normal" visualisation is that an uncertainty visualisation should prevent us from drawing unjustified conclusions.

2.4 Uncertainty as a signal

Uncertainty visualisation is not only motivated by signal-suppression, and we would be remiss to ignore these alternative motivations. Some authors claim the purpose of uncertainty is to improve 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). Other authors do not describe uncertainty as important for decision making, but rather explicitly state it as variable of importance in of itself (Blenkinsop et al. 2000). While uncertainty can provide useful information in decision making, it is important to recognize the "uncertainty" in these cases is not acting as "uncertainty" at all. It is acting as signal.

This is obvious for the cases where we are explicitly interested in the variance or error, as we are literally trying to draw conclusions about an "uncertainty" statistic. The same is true for "decision making" experiments, but it is less overt. This is easiest to understand with an example. Imagine you are trying to decide if you want to bring an umbrella with you to work. An umbrella is annoying to bring with you, so you only want to pack it if the chance of rain is greater than 10%. Unfortunately, your weather prediction app only provides you with the predicted daily rainfall. Therefore, your decision will be improved with the inclusion of uncertainty, *not* because uncertainty is important for your decision, but because it gives you the tools required to calculate the *actual* statistic you are basing your decision on. In this sense, uncertainty is no more "special" to decision making than weight is in a BMI calculation.

Uncertainty visualisation's made for these purposes should simply display the uncertainty statistic we are interested in, such as the variance, or probability of an event. This is precisely what we observe. Figure 4 depicts an exceedance probability map that has was designed as an alternative to the choropleth map to improve decision making under uncertainty (Kuhnert et al. 2018; Lucchesi, Kuhnert, and Wikle 2021). A keen viewer may notice that the "exceedance probability map" is actually just a choropleth map, only the statistic being displayed has changed. We do not believe this graphic be considered an uncertainty visualisation.



(a) Low Variance Data

(b) High Variance Data

(c) Exceedance Probability Palette

Figure 4: An exceedance probability map that depict the counties of Iowa where each country coloured acording to the probability that the average temperature exceeds 27. This map is a choropleth map where the variable of interest is a probability.

There seem to be two different definitions of "uncertainty visualisation" floating around in the literature. The first considers *any* visualisation of error, variance, or probability to be an uncertainty visualisation. The second believes an "uncertainty visualisation" is the output of a function that takes a normal visualisation as an input, and transforms it to include the uncertainty information. The former group believe the purpose of uncertainty visualisation to provide signal about a distribution, while the later believe it should act as noise to obfuscate a signal. The lack of explicit distinction between these two motivations leaves the literature muddled and reviewers struggle to understand if uncertainty should be treated as a variable, as metadata, or as something else entirely (Kinkeldey, MacEachren, and Schiewe 2014). This disagreement creates constant contradictions in what the literature considers to be an "uncertainty visualisation". For example Wilkinson (2005) mentions that popular graphics, such as pie charts and bar charts omit uncertainty, and Wickham and Hofmann (2011) suggests their product plot framework, which includes histograms and bar charts, should be extended to include 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). If you view uncertainty as a function applied to an existing graphic, then you would believe a pie chart and bar chart are not uncertainty visualisations, as they are yet to have the "uncertainty visualisation function" applied to them. If you view uncertainty as any graphic that depicts an "uncertainty statistic" then there are no limitations on which graphics can or cannot be uncertainty visualisations.

When we use "uncertainty visualization" to refer to graphics that simply communicate a variance or probability, we are classifying visualisations by the data they display, not their visual features. Graphics, just like statistics, are not defined by their input data. A scatter plot that compares mean and a scatter plot that compares variances are both scatter plots. Given that there is no special class of visualisation for *other* statistics (such the median or maximum) there is no reason to assume visualisations that simply depict a variance, error, or probability to be special. Some authors implicitly suggest that that visualisations of variance or probability are differentiated due to the psychological heuristics involved in interpreting uncertainty (Hullman et al. 2019). While it is true that heuristics lead people to avoid uncertainty (Spiegelhalter 2017) there is no evidence that this psychological effect translates to issues with the visual representation of uncertainty. Again, given that we do not make these same visual considerations for other variables that elicit distaste or irrational behavior, there is no reason to assume this is what makes uncertainty visualisation so special.

This leads us to the conclusion that the visualisations made for the purpose of displaying information about uncertainty statistics are not uncertainty visualisations. These graphics are just normal information visualisations, and authors can follow existing principles of graphical design. We will focus on the perspective that uncertainty visualisation serves to obfuscate signal, and an uncertainty visualisation is a variation on an existing graphic that gives it the ability to suppress false signals.

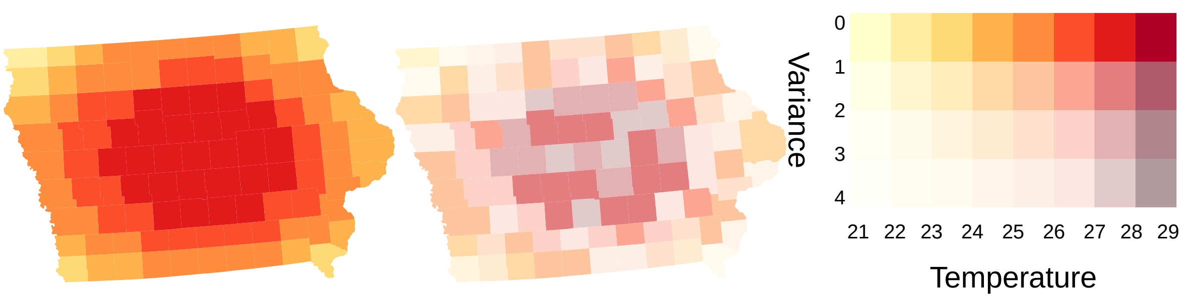
Of course, we do not believe there is anything wrong with explicitly visualizing variance, error, bias, or any other statistic used to depict uncertainty as a signal. Just like any other statistic, these metrics provide important and useful information for analysis and decisions. However, there is no interesting visualisation challenge associated with these graphics, and they do not require any special visualisation. The uncertainty in these graphics are acting as a signal variable, and they should be treated as such.

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 makes sense, as this is the simplest way to add uncertainty to an existing graphic is to simply map uncertainty to an unused visual channel.

3.1 The bivariate map

Figure 5 a variation of the choropleth map, where we have a two dimensional color palette. Not only is temperature mapped to hue, but variance is also mapped to saturation. While these two maps *do* look visually different (which was not the case in the choropleth map) the spatial trend is still clearly visible in both graphics. This means the uncertainty *is technically* being communicated, however the primary take away in the graphic is the spatial trend (that does not exist). The graphic did not hide the invalid signal, so it is not performing signal-suppression as we would like. At this point, it might be reasonable to ask, why? Why is including the uncertainty as a variable insufficient to achieve signal-suppression, and what changes should we make to ensure signal-suppression occurs?



(a) Low Variance Data (b) High Variance Data (c) Bivariate Palette

Figure 5: A bivariate map that depict the counties of Iowa where each county is coloured according to it’s average daily temperature and the variance in temperature. This map is a choropleth map with a two dimensional colour palette where temperature is represented by colour hue, and variance is represented by colour saturation. Even though uncertainty has been added to the graphic the spatial trend is still clearly visible in the case where the spatial trend could be attributed to noise.

3.2 Why this approach may (or may not) work

The difficulty in incorporating uncertainty into a visualisation is frequently mentioned but seldom explained. For example 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. Many authors seem to believe uncertainty visualisation is a simple high-dimensional visualisation problem, as the difficulty comes from working out how to add uncertainty into already existing graphics (Griethe and Schumann 2006). The problem with this approach to uncertainty visualisation is that it treats uncertainty the same as we would any other variable. However, Figure 5 makes it clear that simply including uncertainty as a variable is insufficient to perform signal-suppresion. If we cannot treat uncertainty as any other variable, what should we treat it as? We need to understand what uncertainty actually *is*, in order to understand how to integrate it into a visualisation.

3.2.1 It's a variable... it's a metadata... it's uncertainty?

Describing what uncertainty actually is, is surprisingly hard. Most authors simply avoid the problem and describe the characteristics of uncertainty, of which there are plenty. Often, 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. There are enough qualitative descriptors of uncertainty to fill a paper, but, none of this is particularly helpful in understanding how to integrate it into a visualisation.

Rather than trying to define uncertainty by what it *is* it may be easier to try and describe what uncertainty *is not*. Descriptive statistics 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. In 19th century England, *positivism* was the popular philosophical approach to science (positivists included famous statisticians such as Francis Galton and Karl Pearson). Practitioners of the approach believed statistics ended with descriptive statistics as science must be based on actual experience and observations (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. we need to assume that unobserved phenomena should be similar to observed phenomena (Otsuka 2023). Positivists abhore the assumption of the "uniformity of nature" as they believed referencing the unobservable is bad science. In other words, these scientists embraces descriptive statistics and shunned inferential statistics due to the inherent uncertainty that came with them. Uncertainty is a by-product of inference.

This history lesson illustrates what uncertainty actually is. At several stages in a statistical analysis, we will violate the "uniformity of nature" assumption. Each of these violations will impact the statistic we have calculated and push it further from the population parameter we wish to draw inference on. Uncertainty is the amalgamation of these impacts. If we do not violate the uniformity of nature assumption at any point in our analysis, we do not have any uncertainty.

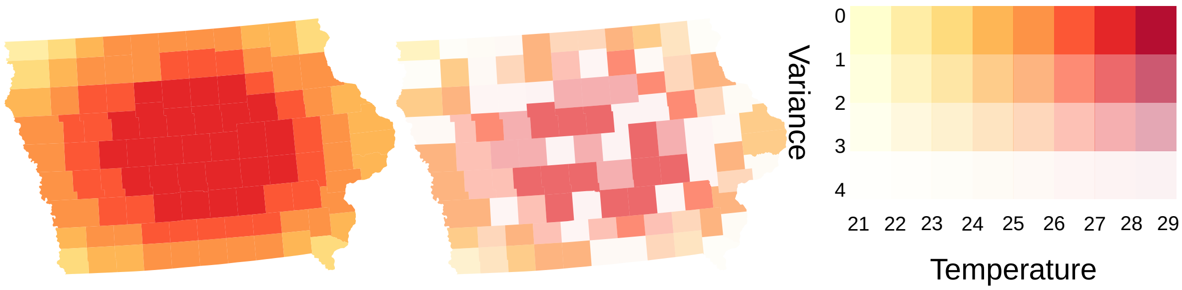
This interpretation of uncertainty indicates that the uncertainty on a statistic is not of value in of itself. Uncertainty is metadata about our statistic that is required for valid inference. This means uncertainty should not be visualised by itself and we should seek to display signal and uncertainty together as a "single integrated uncertain value" (Kinkeldey, MacEachren, and Schiewe 2014). This aspect of uncertainty visualization makes it a uniquely difficult problem. This is something frequently mentioned

3.2.2 Visualising the "single integrated uncertain value"

Typically, when making visualisations, we want the visual channels to be separable, that is, we don't want the data represented through one visual channel to interfere with the others (Smart and Szafir 2019). Mapping uncertainty and signal to separable channels allows them to be read separately, which does not align with the goal of communicating them as a "single integrated channel". Additionally, visualizing uncertainty and signal separately allows the uncertainty information to simply be ignored, which is a pervasive issue in current uncertainty visualisation methods (L. Padilla, Kay, and Hullman 2022). We can see this problem in Figure 5, as it sends the message "this data has a spatial trend and the estimates have a large variance" as we read the signal and the uncertainty separately.

This means effective uncertainty visualisation should be leveraging integrability. That is, the visual channels of the uncertainty and the signal would need to be separately manipulable, but read as a single channel by the human brain. While most visual aesthetics *are* separable, there are some variables that have been shown to be integrable, such as color hue and brightness (Vanderplas, Cook, and Hofmann 2020). When visualizing uncertainty using its own visual channel, we can also consider visual semiotics and make sure to map uncertainty to intuitive visual channels, such as mapping more uncertain values to lighter colors (Maceachren et al. 2012).

Figure 6 is an example of a variations of Figure 5 where uncertainty is mapped to transparency, and temperature is mapped to color hue to leverage these visualisation concepts. This method achieves signal suppression quite well. The spatial trend is clearly visible in the low variance case and that trend it becomes much harder to identify in the high variance case. While this was an effective approach for this graphic, relying on integrability may not give us the amount of control we want over our signal-supresison. Without a strong understanding of how these visual channels collapse down into a single channel, relying on integrability could create unintended consequences such as displaying phantom signals or hiding justified signals.



(a) Low Variance Data (b) High Variance Data (c) Transparency Palette

Figure 6: A bivariate map that depict the counties of Iowa where each county is coloured acording to it's average daily temperature and variance. This map is a variation on the previous bivariate map where instead of variance being mapped to colour saturation, it is mapped to transparency.

Additionally, multi-dimensional colour palettes can make the graphics harder to read and hurt the accessibility of the plots (Vanderplas and Hofmann 2015).

There is another reason Figure 6 is better at signal-suppression than Figure 5, and it may not be due to integrability. Colour value has a second desirable quality for signal-supression, which is that the colours become harder to distinguish as the value decreases. This means high uncertainty values are harder to differentiate than low uncertainty values. This implicit feature of colour value can generalised to other aesthetics by transforming the visual feature space ourselves.

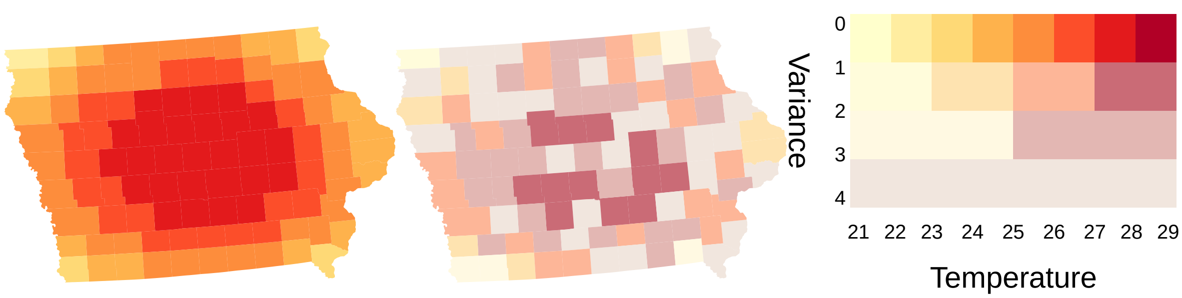
4 Visualising uncertainty and signal in a new space

Instead of hoping that uncertainty might collapse signal values into a single dimension, we can do some of that work ourselves, and uncertainty visualisation authors already have.

4.1 Value Supressing Uncertainty Palettes

The Value Suppressing Uncertainty Palette (VSUP) (Correll, Moritz, and Heer 2018), was designed with the intention of preventing high uncertainty values from being extracted from a map. Since the palette was designed with the extraction of individual values in mind and it has only been tested on simple value extraction tasks (Correll, Moritz, and Heer 2018) or search tasks (Ndlovu, Shrestha, and Harrison 2023), it is unclear how effective the method is at suppressing broader insights such as spatial trends.

Figure 7 is a visualisation of the Iowa temperature data using a VSUP to color the counties. The low uncertainty case still has a visible spatial trend, while the spatial trend in the high uncertainty map has functionally disappeared. This means the VSUP has successfully suppressed the spatial trend in the data. However the spatial trend may not be the only signal of concern in our graphic. Now we must return to the original signal-suppression criteria and ask ourselves if they have all been met. Are all the justified signals reinforced, while all the unjustified signals are suppressed? Is a graphic that performs perfect signal-suppression even possible?



(a) Low Variance Data (b) High Variance Data (c) VSUP Palette

Figure 7: A map made with a VSUP. The counties of Iowa are coloured acording to it's average daily temperature and the variance in temperature, although the space they have been mapped to is slightly warped. Similar to the bivariate map, temperature is mapped to hue while variance is mapped to saturation. This map successfully reduces the visibility of the spatial trend in the high uncertainty case while maintaining the visibility of the spatial trend in the low uncertainty case.

4.2 What can and cannot be suppressed?

The methods used by the VSUP bring to light a slight problem with uncertainty visualisation. Specifically that uncertainty and the purpose of visualisation are somewhat at odds with one another. There are two primary motivations behind visualisation, communication and exploratory data analysis (EDA). Communication involves identifying a signal we want to communicate and designing a visualisation that best conveys that, while EDA involves creating a versatile visualisation using it to extract several signals. If we are designing an uncertainty visualisation for communication then we can just suppress the specific signal we are seeking to communicate. In the map example, we would consider Figure 7 to be a success as the only signal we are concerned with is the spatial trend. However, it is not uncommon for authors to express a desire for uncertainty visualisations that perform signal-suppression in visualisations made for EDA (Sarma et al. 2024; Griethe and Schumann 2006). For uncertainty visualisation for EDA to work, we would need to assume that suppressing individual estimates using their variance should naturally extend to broader suppression of plot level insights. Unfortunately it is not clear whether or not uncertainty visualisations for EDA are possible.

4.2.1 There is no uncertainty in EDA

Earlier we established that uncertainty is a by-product of inference, which means without inference, there is no uncertainty. Often EDA is used to give us an understanding of our data and identify which signals are worth pursuing. In this sense, EDA is the visual parallel to descriptive statistics, as it is performed without an explicit hypothesis which means there is no inference, and by extension, there is no uncertainty.

Some authors recognize inference will occur (in some shape or form) and believe uncertainty *should* be visualised but do not recognize *how* uncertainty would be visualised. Hullman and Gelman (2021) argued that there is no such thing as a "model-free" visualisation, therefore all visualisations require uncertainty as we are always performing inference. However, even something as simple as calculating the uncertainty that is used to suppress our visualisation is not model free, as we need to identify if the sampling variance or the sample variance is more appropriate (Hofman, Goldstein, and Hullman 2020). While we agree that people cannot prevent themselves from performing inference, this does not mean it is possible to include uncertainty in a visualisation designed for EDA. However, this does mean we should endeavor for a versatile uncertainty visualisation method that is able to perform signal-suppression on all the signals displayed in the visualisation.

Uncertainty visualisation for EDA would be possible if we designed a plot in such a way that suppressing individual estimates using their variance would naturally extend to broader suppression of plot level insights. This assumption is commonly made by visualisation researchers in normal visualisation experiments (North 2006), however achieving it would largely depend on the methods use to perform signal suppression. By manually combining the values in

Figure 7, we violated this requirement, so the VSUP is not versatile enough to act as an uncertainty visualisation for EDA.

4.2.2 The limitations of explicitly visualising uncertainty and signal

The lack of versatility of the VSUP is easy to see with a simple example. Let's say we have a graphic that depicts a set of coefficients from a linear regression and the value of the coefficient is shown using a single color. We want to know "Which of these coefficients are different from 0?" as well as "Which of these coefficients are different from each other?". To answer this question we do a series of t-tests on these estimates.

All of the individual t-tests of fail to reject the null hypothesis that the coefficients are different from 0. We then make a visualisation that suppresses this signal and ensures that all of the estimates are visually indistinguishable from 0. We then do a comparison of two means t-test and find that several of the values need to be visually distinguishable from each other. The VSUP method must pick a single color for each estimate, and these colors must be *either* visually distinguishable or indistinguishable from each other. We cannot perform signal-suppression on both these signals simultaneously.

This example highlights a fundamental problem with the VSUP that extends to the bivariate map as well. When we blend these colors, we need to decide at what level of *uncertainty* to blend these colors together. Even though the bivariate map does not explicitly combine color values at certain variance levels, the mapping of variance to color saturation does this implicitly. That is, at certain saturation values the colors in a bivariate map are imperceptibly different to the human brain and appear as though they are mapped to the same value. At this point, it is irrelevant whether or not the colours are technically different, they are the same color in the human brain. Which hypothesis are suppressed and which are not largely depends on the method used to combining colors in the palette (Kay 2019). The VSUP here used a tree based method as that is what was used by Correll, Moritz, and Heer (2018), but there are alternatives that are more appropriate for different hypothesis.

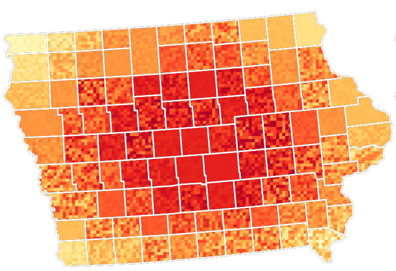
If we only use a single value to express each signal-suppressed statistic, we will always need to decide which signals we suppress and which we do not. However, If we could express the statistic of a cell using multiple colors, this limitation may disappear entirely.

5 Implicitly Combining Uncertainty and Signal

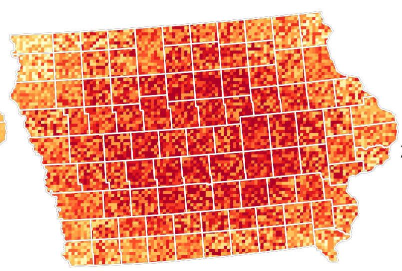
There is technically a stage of our analysis where the estimate and variance are not separate, when we only have a sample. Rather than trying to figure out how to combine signal and uncertainty into a single color, we can just display a sample instead and allow the viewer to extract *both* the estimate and the variance.

5.1 Pixel map

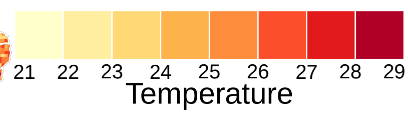
Figure 8 displays a pixel map (Lucchesi, Kuhnert, and Wikle 2021), which is a variation of the choropleth map where each area is divided up into several smaller areas, each colored using outcomes from the larger area's temperature sampling distribution. The spatial trend is clearly visible in the low variance case, but functionally disappears in the low variance case. While the spatial trend is just barely visible in the high uncertainty case, it is much harder to see. This means the graphic also achieves the third criteria for signal-suppression, i.e. our difficulty in seeing the distribution is proportional to the level of uncertainty in the graphic.



(a) Low Variance Data



(b) High Variance Data



(c) Pixel Map Palette

Figure 8: A pixel map of the counties of Iowa. In this map, each county is broken up into several smaller regions and coloured according to a potential daily temperature, given the its average daily temperature and its sampling distribution. This results in each county being represented by a sample rather than a single value. In this graphic, we can clearly see the spatial trend in the low variance case, while the spatial trend is much harder to identify in the high variance case.

It is clear that the pixel-map is not only "suppressing" the false information, but it is doing so by simulating *more* information. The efficacy of this method means that visualisations of simulated samples pop up repeatedly in the literature, with examples including samples that are animated over time (Hullman, Resnick, and Adar 2015; Blenkinsop et al. 2000), pixel-maps, and spaghetti time series plots. Not only does this method help readers understand the plot level "gist", it is also unlikely to damage the viewers ability to extract individual estimates. Extracting global statistics, such as the mean or variance, from a sample can be done with relative each, especially when those values are mapped to color (Franconeri 2021). Therefore, the pixel map performs signal suppression, without sacrificing the viewers ability to extract general statistics, unless those statistics *should* be harder to extract due to the uncertainty in the value. So is this the best uncertainty visualisation? And if so, why?

5.2 Show me the data

The pixel map is not the best uncertainty visualisation, but it is trying to *imitate* the best uncertainty visualisation. The best uncertainty visualisation is the visualisation that best captures the limitations of of our raw data. As we discussed in previous sections, we can consider uncertainty to be "the amalgamations of the impacts of violations to the assumption of the uniformity of nature". It's not a definition that rolls off the tongue, but we can work with it.

Thankfully, this definition of uncertainty aligns nicely with all the concepts that are included in the uncertainty umbrella. Some works (Hullman et al. 2018; Maceachren et al. 2012; Thomson et al. 2005) focus narrowly on specific terms with mathematical definitions, such as probability, confidence intervals, variance, error, or precision. These works are only concerned with quantifying the final impact of uncertainty on our statistics. That is, how large should the bound around our statistic be, such that our "true" statistic can be inferred. Others (Griethe and Schumann 2006; Wilkinson 2005; Pang, Wittenbrink, and Lodha 1997; Pham, Streit, and Brown 2009; Boukhelifa et al. 2017) include broader loosely related elements, such as missing values, reliability, model validity, or source integrity. These broader and harder to quantify concepts are concerned about potential sources of uncertainty, that is, they describe violations to the assumption of the uniformity of nature.

What this means is that we have two types of uncertainty, but one is more "processed" than the other. Quantifiable uncertainties are just assumption violations expressed as an effect on our final statistic. The disconnect between these two expressions of uncertainty creates a huge problem for authors trying to visualise it. 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).

There are two reasons we might leave our uncertainty as an assumption violation rather quantifying the effect. The first reason is that we may be unable to translate the assumption violation to a quantifiable uncertainty. There is no blanket rule that allows us to reliably quantify all uncertainty for every statistic, although some researchers have tangled with the idea. 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 they ignore the disconnect between the broad concept of uncertainty and what we can reliably quantify. Some authors don't believe that it is even possible to quantify all the assumption violations. Spiegelhalter (2017) mentioned that combining the uncertainties that appear at each stage of an analysis into a single "uncertainty" value is near impossible.

The second reason to leave uncertainty as a potential violation of our assumptions, is that we might not know the final statistic we are seeking to calculate. This is the case for visualisations made for EDA, and a large number of developments in EDA visualisation have been in displaying these difficult to quantify violations. For example, Tierney and Cook (2023) builds upon the tidy data principles to allow users to handle missing values. This includes data plots with a missing value "shadow" that allows visualisation authors to identify if the variables used in a plot have any structure in their missing values, which would contribute to uncertainty.

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.

This relationship between uncertainty and the "purpose" of our analysis is littered throughout the literature. 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. 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.

This makes it very difficult to move quantified uncertainty through the the layers of our analysis, especially when designing a visualisation for EDA. If we don't know what the final statistic is, we cannot quantify the effects of our assumptions. Therefore, often the best uncertainty visualisation is not an "uncertainty visualisation" at all, but simply the most accurate depiction of our raw data as it gives us a good idea of it's limitations.

This does not mean that visualizing raw data instead of implementing sampling techniques will always prevent insignificant signal from getting through. Buja et al. (2009) illustrated how groups that appear linearly separable in a linear discriminant analysis (LDA) visualisation of the data can actually be the result of a LDA performed on too many variables, something that was not clear from the visualisation until the line-up protocol was implemented. However just showing the data it is simple but effective option for uncertainty visualisation that that seems to be largely overlooked. While it is not always possible, it should always be considered as an effective uncertainty visualisation when the raw data is available.

6 Evaluating uncertainty visualisations

If we want to make conclusions about how effective any uncertainty visualisation method is we need to look at the results of evaluation experiments. Unfortunately the illustrative methods we have used thus far, i.e. showing a graphic and saying “wow look at this”, are lacking if we want any generalizable results. However, despite the abundance of uncertainty visualisation evaluation experiments, existing literature reviews have struggled to synthesise them into any common rules (Kinkeldey, MacEachren, and Schiewe 2014; Hullman 2016).

Here we discuss common evaluation methods, why these methods might struggle to create a cohesive set of recommendations for uncertainty visualisations, and consier how we to best evaluate visualisations on their ability to perform signal-suppression.

6.1 Current methods

Including uncertainty in a visualisation comes with many secondary benefits. Examples of these benefits include better decisions, more trust in the results and the ability to extract additional statistics, such as the variance. Ultimately, these secondary benefits are not the primary goal of uncertainty visualisation, and evaluating uncertainty visualisations on these criteria often has unintended consequences.

6.1.1 Value extraction of uncertainty statistics

Uncertainty visualisations are often evaluated based on how accurately (Hullman et al. 2019) viewers can separately extract the estimate and the variance (Kinkeldey, MacEachren, and Schiewe 2014). This means a significant chunk of evaluation studies boil down to showing a participant a visualisation and asking questions such as “what is the variance of X?”, or “what is the mean of X?”. This seems like a relatively straight forward approach, and it is similar to how non-uncertainty visualisations are evaluated, but is this appropriate for uncertainty visualisations? The role of uncertainty is rarely evaluated in these studies as the graphics are often compared on the basis of being “uncertainty visualisations” (Ibrekk and Morgan 1987; Hullman, Resnick, and Adar 2015; Hofman, Goldstein, and Hullman 2020), a class that has no established definition. By shifting the focus of our inference from $\bar{X}$ to $Var(X)$ or $P(X)$ we end up evaluating visualisations on their ability to convey uncertainty statistics, rather than on uncertainties ability to suppress statistics. This leads to a series of experiments where the uncertainty is evaluated as a signal even if that was not the goal of the experiment.

The problem with evaluating uncertainty as a signal are identical to the problems associated with displaying uncertainty as a signal. There is no reason to assume uncertainty would behave any differently to any other variable when we evaluate them in this way. For example, Ibrekk and Morgan (1987) found that participants were more accurate at extracting a statistic when it could be directly read off the graphic, than when it required an area estimate (which is the case if using the PDF), or when there was no visual indicator for the statistic at all (which is the case when using the CDF of an asymmetric function). Hullman, Resnick, and Adar (2015) found that a visualisation that allows viewers to count outcomes to estimate a probability outperformed one that required a complicated area calculation. Hofman, Goldstein, and Hullman (2020) and Zhang et al. (2022) found that participants were better at answering questions about a prediction intervals when shown a prediction interval instead of a sampling distribution. Gschwandtnei et al. (2016) found that graphics where the required statistic could be directly read off the plot outperformed those that involved guesswork due to a gradually decreasing line. Cheong et al. (2016) found that participants made better decisions when they were explicitly given the relevant probability in text rather than when they needed to read it off a map. It is well established that extracting information from a graphic using a perceptual task will always be less accurate than explicitly reading the value provided in text form (Cleveland and McGill 1984).

The biggest failing of this evaluation method is not the predictable outcomes, but that it encourages us to see successful examples of signal-suppression as a failing. Blenkinsop et al. (2000) commented that visually integrable depictions of uncertainty should be avoided, as they decrease the viewers confidence in their extracted data values. This conclusion is antithetical to the goals of signal-suppression and occurs because these methods evaluate uncertainty as a signal, not as noise.

6.1.2 Trust, confidence, and risk aversion

Trust is a by-product of displaying uncertainty and it commonly measured in uncertainty evaluation studies (Hullman et al. 2019). 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]. Science communication should be primarily concerned with accuracy.

Setting trust as the variable of interest implicitly encourages statisticians to set trust and as the primary goal of communication. Evaluating visualisations on trust conflates trust and transparency and ultimately discourages signal-suppression. We can see this effect pop up in the results of evaluation studies. For example, Zhao et al. (2023) found that participants were more trusting of model estimates with low uncertainty, but this effect did not carry over to estimates with high uncertainty. Despite decreased trust being a desired outcome of signal suppression, the authors discussion implied this result was not desired. This perspective ends up extending to visualisation authors as well. 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". This is a classic example of the negative impacts of placing direct importance on *trust* rather than *transparency*. In cases of high uncertainty, authors will opt to leave out uncertainty information because it decreases confidence in the authors conclusions. This is the end result of designing uncertainty visualisations for increased trust, rather than signal-suppression.

Another metric that is similar to trust is the participants confidence in their decision or extracted value. Confidence has many of the same issues as trust, but it has an additional confounding factor. In non-uncertainty visualisation evaluation experiments, "confidence" is used as a proxy for the clarity of the visualisation. Confidence cannot simultaneously be a measure of clarity of visualisation *and* a way to capture the uncertainty expressed in a visualisation. Uncertainty visualisations conflate these two measures when they ask about confidence.

Risk-aversion is another secondary effect of uncertainty visualization that is used to evaluate uncertainty visualisations (Hullman et al. 2019). Risk aversion is an economics term used to describe an agent who would chose a random variable with a lower expected payout because it also has a lower variance. Risk aversion is considered irrational behaviour because it is a deviation from the behavior of a rational agent. Comparing participant responses to that of a ration agent has even been suggested as a benchmark for uncertainty visualization experiments (Wu et al. 2023). However, just like trust, evaluating graphics on risk-aversion discourages signal-suppression. Rational agents *by definition* should *ignore* uncertainty information. Risk aversion is considered to be irrational because it means the economic agent *is* considering the uncertainty information. The only cases where a rational agent should not ignore the uncertainty information is when the uncertainty is signal, not noise. Designing graphics that encourage choices that align with that of a rational agent, is to encourage graphics that do not include uncertainty at all.

For these reasons we do not believe trust, confidence, or risk-aversion are useful measures to evaluate uncertainty visualisations. While they are designed to capture the secondary effects of uncertainty, using them as primary measures of visualisation performance is in direct conflict with designing visualisations for signal-suppression.

6.1.3 Questions that attempts to capture signal-suppression

There is a collection of studies that seem to be aware of the issues behind using trust or value extraction to evaluate uncertainty visualisations. These studies try to measure the effect of some kind of "single integrated value" but the methodology is often ad-hoc with varying levels of success.

The first method is what we call the "vague question" approach. The authors of these studies will ask the participants a question that implies that they should use a use of uncertainty, but compare the readers answer to a very specific ground truth. This means the participants are being *evaluated* as though they are performing a value extraction task, but they are not being asked the *questions* that are asked in the value extraction task. Ultimately this approach results in strangely cryptic questions that create a large amount of noise due to the varied interpretation of the questions (Hullman 2016). For example Hofmann et al. (2012) showed study participants 20 plots, where each plot displayed two distributiosn (as a pair of jittered samples, density plots, histograms, or box plots) and asked them to identify the plot where "the blue group furthest to the right". The mean of the two groups was used to decide which group was actually “furthest to the right” and the participants were evaluated against that ground truth. In another example Ibrekk and Morgan (1987) asked participants for the “best estimate”, but the participant’s responses were evaluated against the mean of the distribution. Ultimately, the term “best” is up to the users interpretation, and the estimate that minimized the sum of squared errors was not implied by the question. This vague question approach leads to inconclusive results, as we are left unclear if it was the phrasing of the question or the plot design that caused the participants to answer incorrectly.

A variation of the “vague question” problem, is a series of studies that ask questions that are impossible to answer using the information given to the participants. These studies frequently expect participants to give deterministic answers for probabilistic questions. For example Correll and Gleicher (2014) showed participants the distribution of voter preferences for two candidates in an election, and aksed them “how likely is candidate B to win the election?”. 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. In another example, 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. 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). Interestingly it is unclear how the participants were supposed to answer this question, as L. M. Padilla, Ruginski, and Creem-Regehr (2017) did three experiments and the first and third experiment had conflicting assumptions in their ground truth. In the first study, participants were required assume to the storm was equally intense reguardless of the probability of the probability of the oil rig being hit (i.e. more likely to be hit *does not* mean more damage) but in the second study required participants to assume the opposite (i.e. more likely to be hit *does* mean more damage). Similar to the vague question problem, these studies seem to be aware that we should be evaluating an uncertainty visualization based on a deterministic observation (such as our example of “does this map have a spatial trend”) but are unsure how to incorporate or evaluate it.

While these approaches are certainly a step in the right direction, the experiments end up having far too much noise in their results. Additionally, by having a ground truth, they end up implicitly asking users to either treat uncertainty as a signal, or to ignore it entirely, neither of which is advisable for evaluating graphics on signal-suppression.

6.2 Testing signal supression

If we cannot ask direct questions about uncertainty, we cant measure the secondary effects of uncertainty, and we can't ask indirect questions about the uncertainty, how are we supposed to evaluate uncertainty visualisations? How do you measure something that disappears the second you look directly at it? To evaluate uncertainty visualisations, we need an experimental design that evaluates uncertainty as noise, not as signal. This means we need to measure *uncertainty's impact on the signal*, not the uncertainty itself.

6.2.1 Comparing to hypothesis tests

The most obvious way to evaluate uncertainty visualizations is to compare the visualisations to statistical tests. If a graphic was performing signal-suppression we would expect the signal to be harder to see at higher levels of uncertainty. The ideal outcome is a an uncertainty visualisation where the signal is only perceivable when it would be identified by a hypothesis test. Evaluating visualisations as though they are akin to hypothesis tests is a well established concept in visualisation. The lineup protocol is a good example of this approach, which is a confirmatory visualisation tool that can be used to check if perceived patterns are real or merely the result of chance (Buja et al. 2009; Wickham et al. 2010). The motivation behind the lineup protocol is similar to the motivations behind signal-suppression, although the lineup protocol is more explicitly tied to a specific hypothesis. Weihao Li and VanderPlas (2024) compared standard statistical tests to the lineup-protocol, and evaluated the visualisations using the power curves that are typical for hypothesis testing. In a similar vein, Kim et al. (2019) investigated how different uncertainty visualisation methods influenced user's prior beliefs, and evaluated the graphics by comparing their results to those from Bayesian inference. This approach is similar to the lineup protocol as it also evaluates a visualisation by comparing it to an analogous statistical calculation, however it does so using a different statistical philosophy.

We can use a similar evaluation method to evaluate uncertainty visualisations, however, special care would need to be taken for their versatility requirement. While lineup protocols often have a specific hypothesis that is used to create the null distribution, uncertainty visualisations are intended to be more flexible and can be created without visualizing a specific null distribution. Human viewers using the lineup protocol are less sensitive to deviations from the null hypothesis than the typical statistical tests (Weihao Li and VanderPlas 2024). This means swapping to a visual format and giving viewers the ability to check several hypothesis at once can reduce the effectiveness of the hypothesis test. Therefore it is highly likely that uncertainty visualisations would not produce identical results to standard hypothesis tests or even the lineup-protocol. Additionally, uncertainty visualisations would need to be simultaneously checked against several hypothesis that come from a variety of null distributions.

6.2.2 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", even though this qualitative analysis did not make 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. 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 identified where in the decision making process participants made these considerations.

6.2.3 Heuristics

While experiments that explicitly identify heuristics in current methods are not technically measuring signal-suppression, they are still a useful consideration to keep in mind when designing experiments for signal suppression. Heuristic checks are look at unknown pitfalls that might exist in interpretation of current plots (Hullman 2016) and can change depending on the larger scope of the graphic and the population we are communicating with (Spiegelhalter 2017; Kinkeldey, MacEachren, and Schiewe 2014).

Several heuristics are of particular importance for uncertainty visualisations, and are likely to impact how well different methods performing signal suppression. 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). 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 an outcome was closer to the center of the distribution (and therefore more likely) (L. M. Padilla, Ruginski, and Creem-Regehr 2017). This can be considered an extension of the within bar bias, where participants looking at a bar chat with error bars view outcomes within the bar as more likely than those outside it (Newman and Scholl 2012). Several studies have found that viewers use a heuristic where they compare the distance between two estimates to estimate if they are different, and in doing so, ignore the uncertainty information (L. Padilla, Kay, and Hullman 2022).

These heuristics have the potential to create noise in signal suppression evaluations, and should be kept in mind when designing an evaluation experiment.

7 Future work

This paper has identified gaps in the uncertainty visualisation literature that could be filled to progress the field. We formalized the uncertainty visualisation problem, and in doing so highlighted an untouched area of uncertainty visualisation research.

Experimental practices on uncertainty visualisation need to be standardised. Uncertainty visualisations for decision making treat uncertainty as signal, while visualisations for signal-suppression treat uncertainty as noise. 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. Researchers need to ensure that when they identify the motivation behind their visualisation technique, and that their evaluation methods match the motivations of the paper. Additionally *evaluation methods that evaluate uncertainty as noise need to be developed.*

Experimenters should consider evaluating visual aesthetics on how well they suppress information in a graphic. Research into separability and integrability is of particular interest to uncertainty visualisation, as it allows interference from the uncertainty variable. When designing experiments, authors often choose aesthetics that are visually distinguishable, uncertainty visualisation authors should consider doing the opposite.

Software that allows users to easily perform signal suppression should be developed. Existing uncertainty visualisation methods view a distribution as its own object and there are no software options for the "an uncertainty visualisation is a function of a visualisation" philosophy.

Bibliography

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.

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.

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.

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

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.

Franconeri, Steven L. 2021. "Three Perceptual Tools for Seeing and Understanding Visualized Data." *Current Directions in Psychological Science* 30 (5): 367–75. https://doi.org/10.1177/09637214211009512.

Goldstein, Daniel G., and David Rothschild. 2014. "Lay understanding of probability distributions." *Judgment and Decision Making* 9 (1): 1–14.

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, 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.

Kay, Matthew. 2019. "How Much Value Should an Uncertainty Palette Suppress if an Uncertainty Palette Should Suppress Value? Statistical and Perceptual Perspectives," October. https://doi.org/10.31219/osf.io/6xcnw.

Kim, Yea Seul, Logan A. Walls, Peter Krafft, and Jessica Hullman. 2019. "A Bayesian cognition approach to improve data visualization." *Conference on Human Factors in Computing Systems - Proceedings*, 1–14. https://doi.org/10.1145/3290605.3300912.

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.

Kuhnert, P. M., D. E. Pagendam, R. Bartley, D. W. Gladish, S. E. Lewis, and Z. T. Bainbridge. 2018. "Making management decisions in the face of uncertainty: A case study using the Burdekin catchment in the Great Barrier Reef." *Marine and Freshwater Research* 69 (8): 1187–1200. https://doi.org/10.1071/MF17237.

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

Lucchesi, Lydia, Petra Kuhnert, and Christopher Wikle. 2021. "Vizumap: an R package for visualising uncertainty in spatial data." *Journal of Open Source Software* 6 (59): 2409. https://doi.org/10.21105/joss.02409.

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.

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 Philosophical 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.

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.

Sarma, Abhraneel, Xiaoying Pu, Yuan Cui, Michael Correll, Eli T Brown, and Matthew Kay. 2024. "Odds and Insights: Decision Quality in Exploratory Data Analysis Under Uncertainty." In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. CHI '24. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3613904.3641995.

Smart, Stephen, and Danielle Albers Szafir. 2019. "Measuring the separability of shape, size, and color in scatterplots." *Conference on Human Factors in Computing Systems - Proceedings*, 1–14. https://doi.org/10.1145/3290605.3300899.

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.

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.

Tierney, Nicholas, and Dianne Cook. 2023. "Expanding Tidy Data Principles to Facilitate Missing Data Exploration, Visualization and Assessment of Imputations." *Journal of Statistical Software* 105 (7): 1–31. https://doi.org/10.18637/jss.v105.i07.

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.

Weihao Li, Emi Tanaka, Dianne Cook, and Susan VanderPlas. 2024. "A Plot Is Worth a Thousand Tests: Assessing Residual Diagnostics with the Lineup Protocol." *Journal of Computational and Graphical Statistics* 0 (0): 1–19. https://doi.org/10.1080/10618600.2024.2344612.

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.

Wu, Yifan, Ziyang Guo, Michails Mamakos, Jason Hartline, and Jessica Hullman. 2023. "The Rational Agent Benchmark for Data Visualization." https://arxiv.org/abs/2304.03432.

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.