

# Noisy Work: A Review of The Uncertainty Visualisation Literature

Harriet Mason

1/30/24

## 1. Introduction

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. So overwhelming in fact, that it is not uncommon for people to become so overwhelmed by the information they refuse to engage with it at all. In this overflow of information, tools that can effectively summarise information down into simple and clear ideas become more valuable. While websites that summarise insurance plans or hotel options down into a clear table make a living off it, few tools are as powerful or effective as a visualisation. Information visualisations remain one of the most powerful tools for fast and reliable science communication. Effectively leveraging visualisations to make them more effective is therefore key in facilitating simple and clear communication of science and other ideas.

Think back to the last time you made some sort of data visualisation. What was the purpose of that visualisation? Was it to better understand your data? Was it to help you make a decision? Was it to communicate that decision to someone else? Now think about the last time you expressed some form of uncertainty. Was it a set of numerical confidence intervals? Maybe they were expressed as a set of values in a table. Did you consider visualising your uncertainty instead? There are many stages in our analysis that benefit from the power of data visualisation, however this does not mean it is always done with success. Visualization is an important step in exploratory data analysis and it is often utilised to **learn** what is important about a data set. The importance of data driven discovery is highlighted by data sets such as Anscombe's quartet (Anscombe 1973) or the Datasaurus Dozen (Locke and D'Agostino McGowan 2018). Each of the pairwise plots in these data sets have the same summary statistics but strikingly different information when visualised. Anscombe quartet is shown in Figure 1, because describing the data is never the same as seeing it. Instead of having to repeatedly check endless hypothesis to find interesting numerical features, visualisations **tell** us what is important about our data. This powerful aspect of data visualisation is poorly or seldom

used in later stages when we are communicating our findings, specifically with respect to uncertainty.

- Suh et al. (2023) found that the main issue with deployment of AI was communication between data science and subject matter experts, since subject matter experts didn't want technical presentations that didn't explain exactly when and how a model would fail.

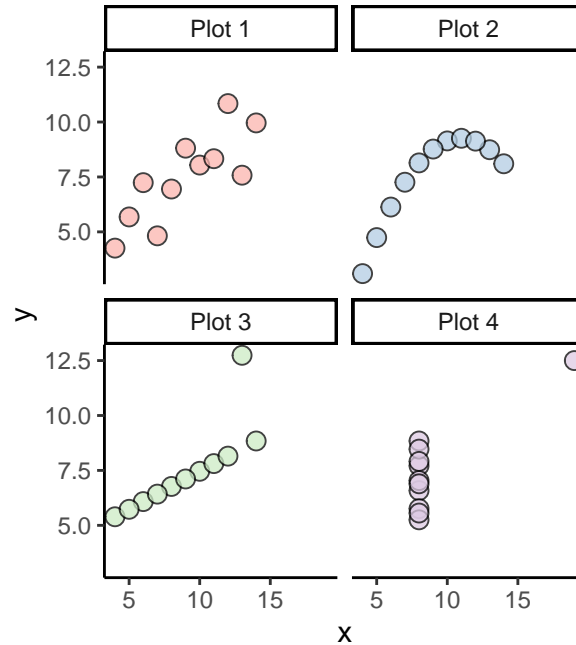


Figure 1: The four scatter plots that make up Anscombe's quartet. The four scatter plots are visually distinct but have the same mean, standard deviation, and correlation. The visualisation highlights the importance of plotting your data to identify interesting features that are hidden by other summary statistics.

- industrial accidents have a “tolerable” risk chart based on risk and number of fatalities Spiegelhalter (2017)

## 2. Defining Uncertainty

Uncertainty visualisation is not made any easier by the fact that the term “uncertainty” lacks a commonly accepted definition in the literature. Spiegelhalter (2017) even commented that the definition of uncertainty essentially “depends on who you ask”. The disarray of organisation and definitions within the uncertainty literature is a common sentiment expressed by reviewers (Spiegelhalter 2017; Kinkeldey, MacEachren, and Schiewe 2014). This mishmash of

terminology leads to a large body of work, all claiming to finding the best visualisation or expression of “uncertainty” but most don’t even seem to agree on what uncertainty is. In this section we will establish exactly what we mean when we talk about uncertainty.

## 2.1 Colloquial Definitions

The most common definition of uncertainty used in papers that discuss visualising it are colloquial. Most definitions make some nod towards variance, precision, trust, error, missing values or another related concepts. Since *uncertainty*, unlike *probability*, does not have a strict definition that is established in mathematics, there is large variety in its usage. Some consider it to be synonymous with probability, some consider it to be an encompassing umbrella term of which probability is only an element, and some consider it to be something else entirely. This lack of a definition leads to uncertainty being swapped out freely with other vague words such as ambiguity and risk. Definitions presented in papers that discuss uncertainty visualisation want it to be encompassing and include everything a layperson might consider when trying to consider “uncertainty”. The issue with these broad definitions is that in order to visualize something, it needs to be quantified, and in order to quantify something, we need to know exactly what it is we are counting. This discrepancy appears regularly in the uncertainty visualisation literature. Often a vague definition such as “uncertainty is anything that isn’t deterministic”, or even a refusal to define uncertainty at all in an experiments introduction will be accompanied by a highly specific quantification of uncertainty as a probability density distribution of a forecast when providing the visualisation. Ironically it seems that many of the papers that refer to themselves as “uncertainty” papers are just suffering from a form of uncertainty, a lack of precision.<sup>1</sup>

This lack of precision has caused several pernicious issues in the literature, including different representation of uncertainty being compared despite not being realistic substitutes (with no acknowledgement of this from the authors); a lack of understanding as to *why* uncertainty should be included in a visualisation; a large swath of literature focused on a minute aspect of the entire uncertainty space with little attention to creative solutions to real world visualisation problems; and a swath of newly designed visualisations that are disconnected from the problems that they hope to solve. These problems means the literature either needs to transition away from the broad umbrella term “uncertainty”, or consider the full range of sources of uncertainty and how they can be communicated when designing an experiment.<sup>2</sup>

The most encompassing definition of uncertainty I have seen comes from Walker et al. (2003) who define uncertainty as **“any deviation from the unachievable ideal of completely deterministic knowledge of the relevant system”**. This definition encapsulates many concepts adjacent to randomness such as predictions, probability mass distributions (PMF),

---

<sup>1</sup>Flesh out note - I think several of the sentences in this paragraph need examples (or quantified evidence) from the spreadsheet definition part

<sup>2</sup>I am not sure if I should mention this here because I go into depth in the literature review and these comments might come across as unjustified if they are left in this section. Maybe this is more of a conclusion paragraph.

estimate error, and any data that is not a set of deterministic outcomes. This is the definition of uncertainty I will use for the rest of this report.

## 2.2 Taxonomy Definitions

### 2.2.1 Taxonomies of Uncertainty

While colloquial definitions of uncertainty are useful when trying to communicate with laypeople who don't understand strict statistical definitions, these definitions are not helpful when we are trying to work out exactly what should be shown on a plot. When authors try to make a considerable effort to define what we mean when we talk about uncertainty, they often express it in the form of a taxonomy.

When reading any literature about uncertainty or an adjacent topic you will be overrun with more ways to organise uncertainty than you will know what to do with. The problem is so pervasive that it is almost impossible for two randomly selected papers that both claim to be about uncertainty will use the same definition. What is worse is that the reason for using one taxonomy over another, or why a new taxonomy was established in the first place is rarely discussed. This leaves new authors in the field with a wide array of choices and no information that would allow them to determine which is most effective for them. The complete mess of this literature makes it more enticing for authors to just define their own definition of uncertainty, which only adds to this problem. This is not done in ignorance of the issue either, many papers that comment on the unending list of uncertainty definitions will still provide a new taxonomy of their own design. I would almost find this situation comical if I were not one of the people forced to sort through this mess.

Examples of Taxonomies used to untangle the features of uncertainty:

- aleatory uncertainty (unavoidable randomness), epistemic uncertainty (uncertainty about structure, still quantified), ontological uncertainty (uncertain about entire modelling process). Spiegelhalter (2017)
- measurement (uncertainty on a single variable), spatial (uncertainty in spatial data), temporal (uncertainty in a time data)
- aleatory (inherent randomness in a forecast) vs epistemic (uncertainty in structure and parameters of statistical models) vs ontological (uncertainty about the entire modelling process as a description of reality)
- scientific methods (variance, distribution) vs human judgments (disagreement among parties)
- Random vs systematic uncertainty
- statistical (can provide a continuous pdf) vs bounded (know values will fall in a certain range) uncertainty
- accuracy vs precision
- Uncertainty from discovery, assessment and execution

- risk vs uncertainty
- “known knowns, known unknowns, and unknown unknowns”, (risk, uncertainty, ignorance)
- Source of uncertainty (model structure, parameters, forcing functions, initial state, model operation) and diagnostic (past/current) vs prognostic (forecasting)
- uncertainty is organised according to Satiencia (legal moral, societal, institutional, proprietary, situational) and level of severity (high, medium, low)
- Level of precision in epistemic uncertainty: 1) numbers given to appropriate levels of precision 2) a distribution range 3) a measure of statistical significance 4) verbal quantifiers to numbers 5) refusal to give a number unless the evidence is good enough Spiegelhalter (2017)

This taxonomy issue creates several problems in the literature. Ideally, if we wanted to combine the literature on uncertainty visualisation, each paper would state the same definition of uncertainty and explain which aspect of uncertainty they are visualising, so the experiments would shift around the space while the definition of uncertainty and what it contains would remain stationary. This would make it easy for the typical reader to use the literature to find an appropriate visualisation for the particular category or case of uncertainty. Instead, one has to comb through the method of a particular design to understand if the uncertainty you are working with is amenable to this method. Another issue this creates is that the “space” of uncertainty visualisation is unclear. This means that several areas of uncertainty visualisation that have a real need for creative visualisation tools are left untouched, while the visualisations that are considered “uncertainty visualisations” simply because they are an expression of a probability mass function dominate the literature. Finally the taxonomies do little to help data analysts or statisticians untangle, estimate and understand the uncertainty that may be in their own projects. The endless set of constantly changing definitions implies their *is no* set meaning to the term “uncertainty” and something that cannot be defined certainly cannot be quantified for communication or estimation. This leads people to be less likely to visualise uncertainty due to its imprecision and difficulty to communicate.

This does not mean there is nothing to be learned from these taxonomies, each in their own way show what aspect of uncertainty is important to certain cases. There are a handful of overlapping features that indicate which elements of uncertainty are of interest to statistics and related applied fields. Walker et al. (2003) identified these similarities and used them to design a taxonomy that encapsulates a large proportion of the ad hoc definitions that currently overrun the literature, and is the definition of uncertainty that we will work with for the remainder of this paper.

Figure 2 is an illustration of the taxonomy presented by Walker et al. (2003). In this taxonomy, there are three things we need to consider for each “uncertainty” we encounter through the modelling process. First, we need to consider the source of the uncertainty. Is this uncertainty coming from the data (for example, from an inaccurate measurement), from the array of model choices we have, or from the assumptions on a parameter? This is the *location* of the uncertainty. Second, consider how well you can quantify this uncertainty. Do you know exactly

how much measurement error there is in each observation or are you not even aware if there is a measurement error? This is the *level* of your uncertainty, and it ranges from discrete to total ignorance. Finally, consider how this uncertainty came into existence. Is it a result of a naturally random process (aleatory) or is it due to imperfect information and could be improved (epistemic). This is the *nature* of your uncertainty. Walker et al. (2003) then goes on to describe mapping our uncertainty in a 3D space that is defined by its location, level, and nature, but I think the taxonomy is more easily understood as a series of questions we should consider when we are trying to quantify uncertainty. Location asks “where and at what stage in the modelling process is this uncertainty coming from”, level asks “how well can we quantify this uncertainty”, and nature asks “can we reduce this uncertainty with improves knowledge?”. While this taxonomy specifically applies to the uncertainty in the model, it is easy to fit most of the uncertainty we would consider quantifying into this format.

*I need to fix the taxonomy illustration because inputs and model uncertainty need to be swapped*

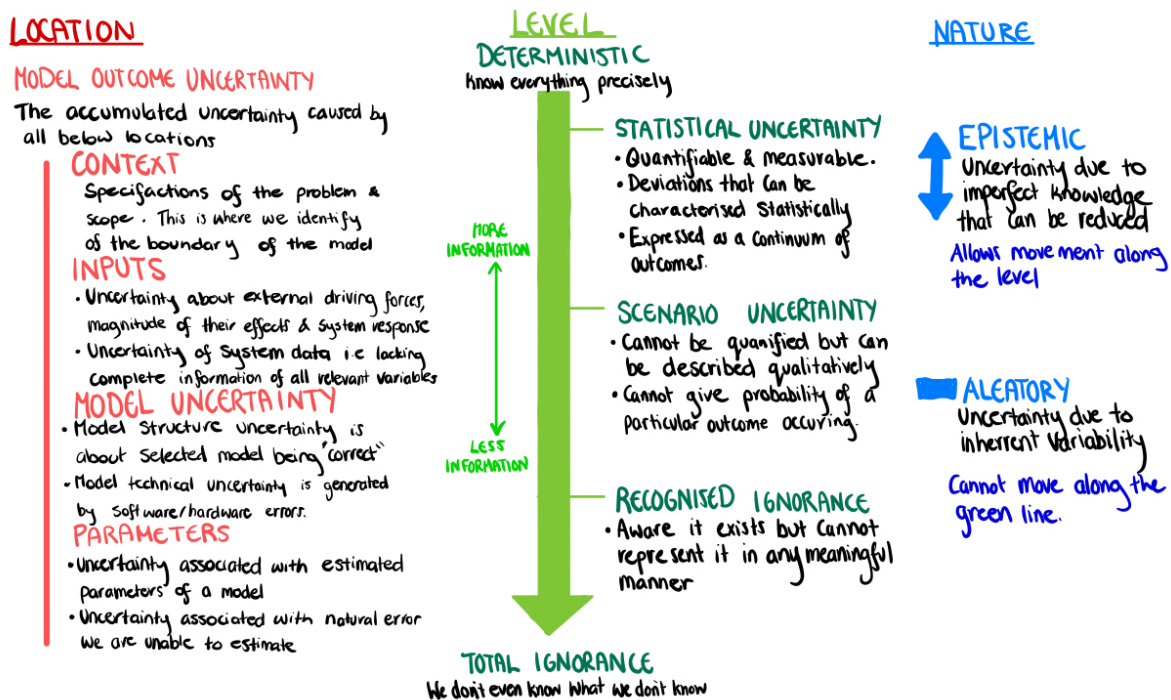


Figure 2: Depicts an illustration of the taxonomy described in Walker et al. (2003). From right to left the drawing shows the location, level and nature of uncertainty with examples of that category underneath. A specific source of uncertainty from the location can be mapped to a level of ignorance that can increase or decrease (i.e. moving up or down the green line) depending on the nature of the uncertainty. Identifying the location, level and nature of your uncertainty allows you to better understand it.

The *taxonomy* we will use in this paper may seem complicated and difficult to grasp at first glance, but it essentially boils down to three dimensions to consider. Uncertainty is a broad and complicated topic and while oversimplified taxonomies might help us hold the concept of uncertainty in our head, they do little to help with the practical question of how to quantify, visualise, or communicate uncertainty. Organising uncertainty according to its location, level and nature makes allows us to consider uncertainty in the myriad of ways it can arise in a data analysis project and better understand *how* this uncertainty should be communicated. With a detailed understanding of the sources of uncertainty and the ways it can be quantified, we can better understand the gaps in the current set of visualisation tools and assess if these gaps need to be filled.

The *location* axis of the taxonomy lines up neatly with the typical data analysis pipeline. Multiple authors in the uncertainty literature have commented on the need to consider quantifying and expressing uncertainty at every stage of a project (Kinkeldey, MacEachren, and Schiewe 2014; Refsgaard et al. 2007). Figure 3 shows the connection between the typical data analysis pipeline and the location element of the uncertainty taxonomy, along with examples of how uncertainty could arise from each of these steps. The relevance of every stage of the uncertainty taxonomy is backed up by both theoretical and practical work in statistics and adjacent fields. Munzner (2009), in her nested model for visualisation, comments on how even something as low level as poor problem abstraction can cause down stream effects and become impossible to ignore. Otsuka (2023) highlights that the choices made in the ontological process of boiling real world events down into statistical objects can influence later stages of analysis due to what was considered signal and what was considered noise. Meng (2021) stresses the importance of treating data as an input from another process with noise and variance, rather than as a neutral input. High level stages such as model selection and estimation have obvious introductions of uncertainty through the choice of models and nature of estimation [‘I’m not sure if I need a citation here because this feels obvious’]. Location is also acknowledged as an important element of consideration in the taxonomies of several visualisation papers [‘I will probably need a citation for which ones’].

The *level* axis forces us to identify how quantifiable the uncertainty is. This element is important because establishing how quantifiable uncertainty is informs us how it can be communicated, something that Spiegelhalter (2017) also identified in the form of “precision”. Statistical uncertainty can be quantified, so it can be expressed using a PDF, a variance, an error estimate, or something similar while outcome uncertainty cant be quantified but each scenario may be simulated. Acknowledged ignorance can only be stated in the form of an assumption. Visualisations of climate change scenario uncertainty typically combine the outcome uncertainty of carbon emissions with the and statistical uncertainty of the noisy measurement and modelling system to create a cohesive illustration of the uncertainty surrounding climate change. It is commonly noted that each uncertainty must be discussed in isolation, but combining the uncertainty from every stage is near impossible (Spiegelhalter 2017). The level axis and climate change visualisations provide tools for how to consider combining these uncertainties. Uncertainties that are the same level but from multiple locations can likely be combined to establish an overall model uncertainty. While uncertainties at different levels cannot be com-

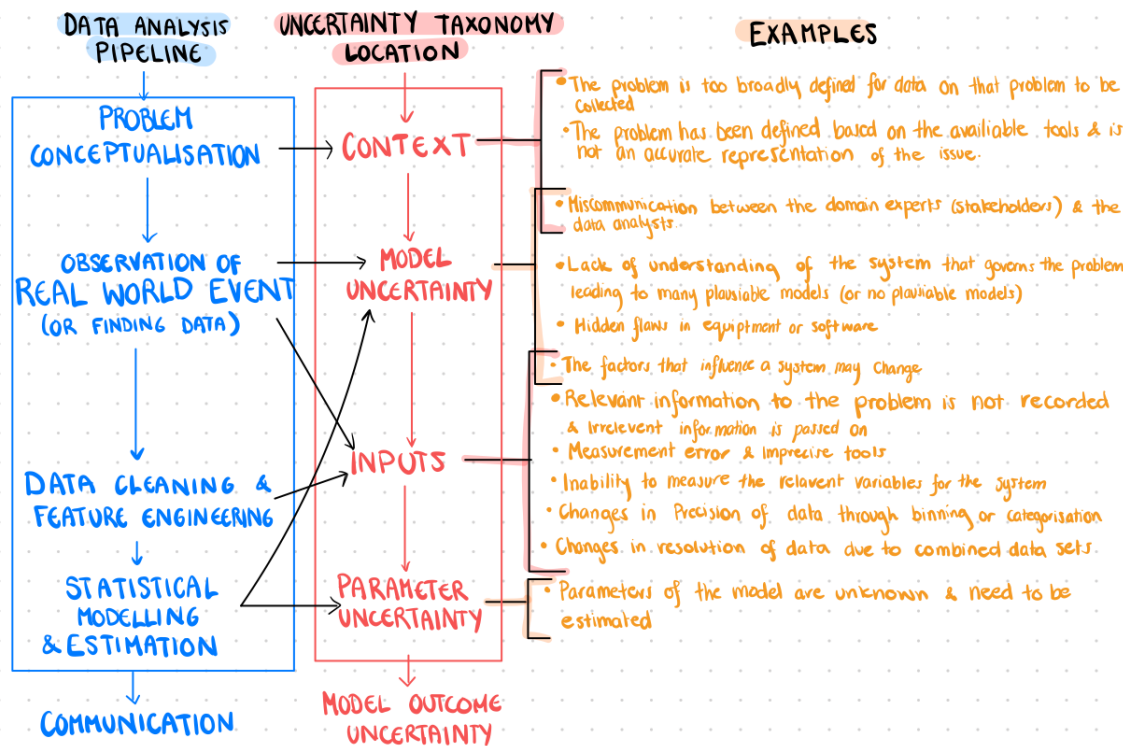


Figure 3: Illustration of the connection between the data generating process and the location of the uncertainty analysis. Several examples are collected from utypo and Munzner2009.



bined, they can still be visualised simultaneously as shown in the climate scenario uncertainty visualisations.

*The nature axis* allows us to easily communicate if the uncertainty can be reduced or if it needs to be accepted. While this level is not of particular importance to the visualisation of uncertainty, that is not the only purpose of the taxonomy. The nature of the uncertainty is an important element in expressing uncertainty, especially to stakeholders, and connecting the nature to the location makes it easier to understand *how* uncertainty can be reduced.

Mapping uncertainty into the 3D space defined by location, level, and nature can expand beyond the narrow case of statistical modelling it was defined for. (\_\_\_\_?) included “consensus” as a taxonomy element, however it was not included in the final definition, but it can easily be added as a location that would align with the communication stage of the data pipeline. Several authors [(+\_?)*Spiegelhalter (2017)*] considered the method by which uncertainty was quantified (e.g. by a single value, a function, simulated outcomes, etc), which can be thought of as a more precise categorisation of the level axis. This taxonomy also ignores contextual information about the data, such as spatial or temporal considerations. This flexibility means that the use of this framework is not limited by a specific case and should be used going forward to define what aspect of the uncertainty space a contribution exists in.

The taxonomies that cannot be contained within the Walker et al. (2003) taxonomy are not a new uncertainty dimensions, but rather they are conflating adjacent information with uncertainty. A good example of this is the distinction between spatial, temporal, and estimated uncertainty. When we create statistical models and quantify uncertainty, the spatial or temporal aspects are often boiled away (Otsuka 2023), and the only element of it that remains is the inexchangeability of our data. For this reason, there realistically is no such thing as “spatial” uncertainty in the same sense that there is no such thing as “photograph” uncertainty or “coin toss” uncertainty or “baby birth” uncertainty. Maps and photographs boil down to the same statistical objects in the same way that coin tosses and baby births boil down to the same statistical objects. Once we have removed the physical nature these objects are near identical when considered statistically. Of course the nature of them can change based on what we choose to observe (i.e. if we consider a coin landing on its side to be an outcome or if we consider intersex babies as an outcome in births) but this means the contextual information of our data, e.g. whether or not it is spatial or temporal, is only of importance at two stages of the data analysis pipeline. It is first important when we *boil it down* to our data through *observation* and again when we *map our statistical information back to its real world information* to *communicate* it. The only way the context uniquely contributes to uncertainty is in the “boil down” stage, however spatial or temporal uncertainty never refers to that specific cause of uncertainty. It is typically in reference to measurement error, data aggregation, statistical modelling, or other steps that occur *after* the boil down stage, and are therefore not unique to spatial or temporal data. While the spatial or temporal features *may* become relevant at the final communication stage, however this has little to do with quantifying uncertainty, and the difficulties in communicating a complicated statistical model through a complicated map

is already well discussed in the visualisation literature, and in later sections we will discuss how this concept is related to uncertainty visualisation.

In this sense we are not proposing that the current conceptualisation of data according to its physical properties is necessarily wrong, or that it leads to incorrect conclusions, we are merely suggesting for a different way of slicing the metaphysical cake of uncertainty that better aligns with what is *actually happening* when we perform a statistical analysis. This metaphysical cake is illustrated below as Figure 4. In changing what the actual boundaries are when we work with these problems, we improve our understanding of what we are actually doing when we conceptualise uncertainty, and in the process allow for more creative solutions creativity (e.g. by borrowing concepts from spatial visualisation to visualise uncertainty in photographs generated). However, since uncertainty is almost synonymous with the concept of “statistics”, this need for an adjustment in philosophy is likely needed in several domains within the field.

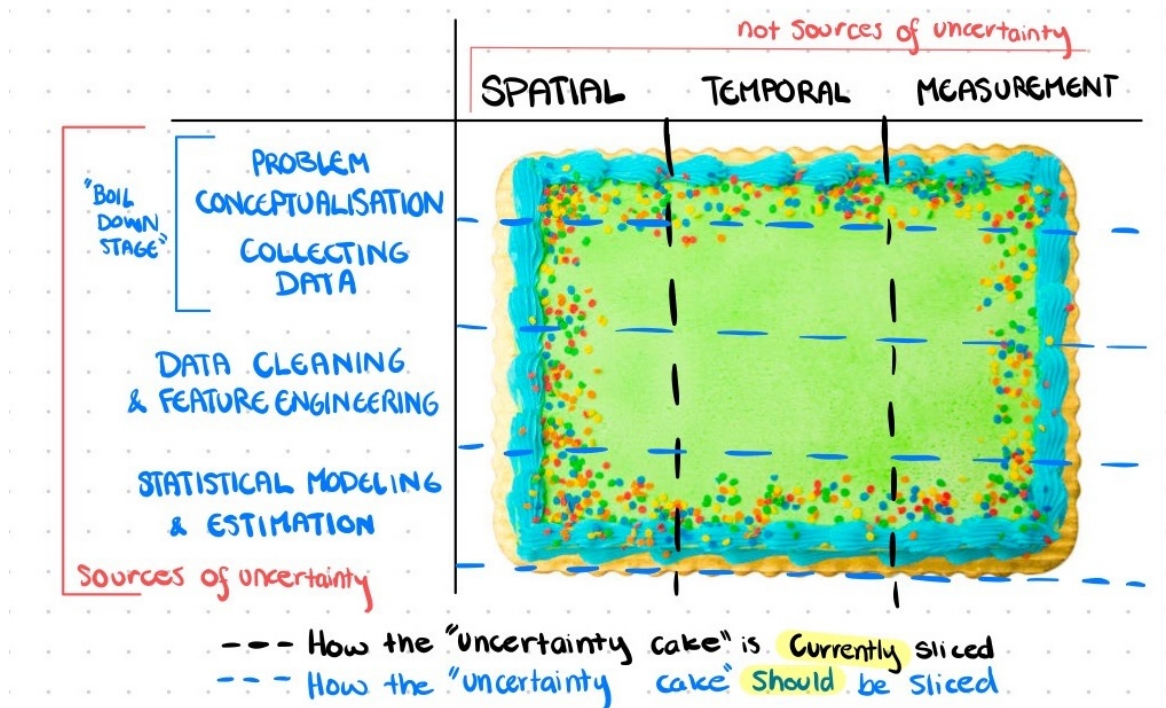


Figure 4: The metaphysical cake we are slicing differently

*Should I add in a bit where uncertainty papers are organised according to which uncertainties they visualise or should I leave that until the lit review section?*

### 2.2.2 Taxonomies of Visual Uncertainty

There are some taxonomies of visual uncertainty. These taxonomies seem to seek to define “uncertainty visualisation” and a combined term rather than as two separate words. We largely disagree with this conceptualization of uncertainty visualisations as it blurs the line between the mathematical estimations and assumptions (the uncertainty part) and the visual depiction of those mathematical objects (the visualisation part). While we will not focus on these taxonomies, it is still important to mention them and highlight the reasons they are not good practice.

There are a small handful of taxonomies (or typologies) that are specifically for uncertainty visualisations. Kinkeldey, MacEachren, and Schiewe (2014) identified five categories of uncertainty based on the work of other literature reviews of uncertainty visualisation. The five categories are: (1) explicit/implicit (directly mapping or showing multiple outcomes) (2) intrinsic/extrinsic (using existing symbols e.g. colour value, or new objects e.g. grids) (3) visually integral/separable (can or cannot be separated from the data) (4) coincidence/adjacent (if data and uncertainty are in integrated or separate views) (5) static/dynamic (animation/interaction). Of the groups they identified, they actually only used (4), (2), and (5), left out (1) because most visualisations are explicit, and (3) corresponds to (1) in most cases (Kinkeldey, MacEachren, and Schiewe 2014).

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”, a statistical object that describes a random variable that is typically a one dimensional function, is used to describe both the data and the uncertainty for all dimensions, it is hard to understand how this dimensionality should work.

Grewal, Goodwin, and Dwyer (2021) created a taxonomy that mapped uncertainty visualisations to some point in a 2D space that was defined by the axis “domain expertise” and “continuum of discreteness” that scaled from “point estimate” to “continuous distribution”.

Thomson et al. (2005) titled their paper “A typology for visualizing uncertainty” however it is a taxonomy for the specific types of uncertainty that we may want to visualise and a way to calculate that uncertainty. While this paper is more about mathematics than visualisation, it is clearly written as a precursor to visualisation.

These taxonomies are useful in the sense that they often identify areas where there is an over or under supply of uncertainty visualisations. Additionally taxonomies do highlight a concept of uncertainty dimensionality that was not directly covered in the uncertainty taxonomy, however the distinction between uncertainty as a single value versus as a PDF could be considered a sub-consideration of uncertainty that’s level is statistical so it can be quantified. While uncertainty visualisation taxonomies are not the standard practice (generating a taxonomy of uncertainty and then developing methods to depict that uncertainty is) (Kinkeldey,

MacEachren, and Schiewe 2014), it is still important to acknowledge the existence of “uncertainty visualisation typologies” and why they may miss the mark in what is needed from the uncertainty visualisation literature.

The first issue with these taxonomies is that most of these taxonomies are created by observing existing uncertainty visualisations. This means under researched gaps in the taxonomy could be due to a lack of need just as well as a lack of attention, and the benefit of trying to create visualisations that land in uncharted areas of the taxonomy are unclear.

The second issue is that there does not seem to be anything these taxonomies offer that would not be better established by separate taxonomies for uncertainty and visualisation. These taxonomies seem to depict uncertainty visualisation as a field within itself, rather than a specific case of information visualisation that requires more attention to mathematics than usual. The difference between “uncertainty visualisations” and “data visualisations” is not technically in the visual element, it is mathematical. Kinkeldey, MacEachren, and Schiewe (2014) almost acknowledges this in their own paper that discusses an uncertainty visualisation taxonomy when they claim “future typologies should take different categories of tasks into account (1) communication tasks (2) analytical tasks (3) exploratory task”, a common typology for information visualisation in general. The process of understanding and estimating uncertainty requires knowledge of the data, the statistical methods used to make an estimate, and the assumptions of a model. Visualising the statistics that represent uncertainty should be no different than depicting the statistics that represent any other element of a graphic, and therefore there seems to be no reason to have a separate taxonomy for uncertainty visualisation.

### **3.General Attitude Towards Uncertainty + Its communication**

It is difficult to read a paper that discusses uncertainty that does not make a nod towards uncertainty bias. In the case of uncertainty visualisation, these nods are not always relevant. Similarly to how contextual information,

I don't think this is worth a section but it is mentioned a lot in . - the role of uncertainty visualisation is to communicate accurate information - problem of communicating to a group thing (avoid communicating inaccurate information) - how people feel about uncertainty is irrelevant - natural disasters make people uncomfortable but to include it in a conversation about whether or not we should express it is ridiculous. - Numerical risks: Use absolute risks (but relative risk when catastrophic event), for single unique event use percent chance, express change as a proportion, frequency, or percentage (and be clear about reference class), avoid framing bias by showing risk with and without the outcome, keep the denominator fixed when making comparisons, explicit about time interval, For more knowledgeable audiences consider providing epistemic uncertainty and confidence in analysis Spiegelhalter (2017)

- General issues when communicating risks: clear objectives, segment audience into target groups, use plain language and limit info to only what is necessary, allow for different

levels of interest knowledge and numeracy (a gist level, numerical level, then evidence and uncertainty) Spiegelhalter (2017)

- compares a bunch of stuff to the chance of getting struck by lightning since they are on a common scale (which is kind of funny) Spiegelhalter (2017)
- hard to communicate small risks because people struggle to discriminate between 1 in 10 000 and 1 in 100 000. Common solution is to use a logarithmic scale (if drawn vertically these are known as risk ladders). design of scale important because risk is associated with position rather than magnitude Spiegelhalter (2017)
- heath had a wide range of different metrics they adopted to communicate chronic risk of adverse events in the future Spiegelhalter (2017)
- using words to communicate numerical risk is generally discouraged and using words without numerical interpretation can lead to lower accuracy and higher risk aversion Spiegelhalter (2017)
- the Affect heuristic comes to play in how people understand risk Spiegelhalter (2017)
- avoids the topic of risk at times of acute crisis because numerical risk would be inappropriate Spiegelhalter (2017)
- positive vs negative framing matters, can try to avoid it Spiegelhalter (2017)
- no preference between percent or frequency although (small interaction with numeracy skills though) Spiegelhalter (2017)
- Need careful specification of the reference class (e.g. % chance of rain should be fractions of situations like this in which it rains - over both time and location) Spiegelhalter (2017)
- Need to make time interval clear (50% each year) Spiegelhalter (2017)
- understanding better in a table than text Spiegelhalter (2017)
- if expressing as a frequency, note that larger numerators suggest a larger risk, extreme version of this is ignoring the denominator all together Spiegelhalter (2017)
- low-probability high-impact events are particularly difficult to communicate risks about Spiegelhalter (2017) # 4 LITERATURE SUMMARY
- methods used in uncertainty visualisation evaluation remain ad hoc. studies approach issue from a usability perspective instead of asking WHY representations DO or DO NOT work. Studies do not follow any methodology commonly agreed upon. Kinkeldey, MacEachren, and Schiewe (2014)
- Lack of formalisation and rigour in empirical methods is an issue that is much broader than uncertainty visualisation (extends to info-vis and related domains). Therefore this paper is a starting point, not a definitive summary Kinkeldey, MacEachren, and Schiewe (2014)

- The question in these papers is often “does method A work better than method B” which is an engineering approach, where the goal is to improve a particular product rather than create general principles Kinkeldey, MacEachren, and Schiewe (2014)
- Authors then try to generalise beyond the specific constraints of the test even though the framing does not allow for it Kinkeldey, MacEachren, and Schiewe (2014)
- Alternatively, we should ask how/why does method A work better than method B, which grounds work in perceptual and cognitive theory and provides a framework Kinkeldey, MacEachren, and Schiewe (2014) # 4.?? Similar work (other literature reviews)
- If you just keep using rankings, you need to compare everything one by one forever to get an idea of the space Hullman (2016)
- people sometimes ignore uncertainty and instead look at the pattern (e.g. distance between means) and use a heuristic to answer the question. be proactive about possible heuristics, look for signs of heuristics in responses, ask subjects to describe their strategy, consider including degree of heuristic as a dependent variable Hullman (2016)
- People believe in the law of small numbers but don't believe in the law of large numbers Hullman (2016)
- experiencing an event makes your perceived probability higher than if you see a description of it Hullman (2016)
- Blur and fog are suggested because they are metaphors for uncertainty, however the usefulness of these metaphors is rarely investigated Kinkeldey, MacEachren, and Schiewe (2014)
- Frequency thing means that more effective communication to participants in a study means more effective science and we have a roll on effect Hullman (2016)
- Probability judgements are less relative than other psychophysical judgements, 0%, 50% and 100% all have well understood meanings. Subjective probability distributions are spontaneous, and subjects are constructing a distribution upon being asked rather than articulating a fully formed mental representation of their beliefs. Therefore it is subject to heuristics. consider anchoring a scale with familiar probabilities . not sure if being asked for the entire distribution helps, e.g. the graphical one reduces noise but asking people for an interval estimation leads to overprecision and overconfidence Hullman (2016)
- If the goal is to compare visualisation, ground truth is easy, if the goal is to evaluate how accurate the subjective probability distributions are, the ground truth becomes complex. ground truth issue e.g. if you provide a sample from a distribution, is the true value the mean of that sample or the population mean? Hullman (2016)
- believed that offering enough decisions with varying rewards you can observe the true subjective probability, however whether or not probability is the right representation of subjective uncertainty is a topic of debate Hullman (2016)

- she also discusses graphical interface (users have to draw the probability distribution) and framing probabilities as natural frequencies are better choices because they reduce noise and stuff. @Hullman2016
- Consider incorporating utility functions. Include “probability-coherence” checks Hullman (2016)
- evidence that people do not follow the axioms of probability when conveying subjective probabilities. How do you deal with this as an analyst (e.g. if the probabilities do not sum to 1 Hullman (2016)
- If someone answers a question correctly it is hard to tell if it was because of the information provided or if the person was using a heuristic Hullman (2016)
- You cant tell if subjects in an experiment do poorly because the visualisation was bad or because the question was misinterpreted Hullman (2016)
- Visualisations risk: consider a good summary table as a visualisation, use multiple formats because no single representation suits all members of an audience, illuminate graphics with words and numbers, design graphics to allow a part-to-whole comparison on an appropriate scale, narrative labels are important (show magnitude through tick marks), use narrative, images and metaphors to gain/retain attention but dont arouse undue emotion, assume low numeracy of a general public audience and adopt a less-is-more approach by reducing the need for inference and making clear and explicit comparison, be cautious about interactivity and animations (may introduce unnecessary complexity), avoid chart junk (like 3d bar charts), assess the needs of the audience, and iterate towards a final design. Spiegelhalter (2017)
- Graphical features that improve accuracy, features that facilitate behavioural change, and features that viewers like may be different (are not necessarily the same). Spiegelhalter (2017)
- also should consider graphical literacy Spiegelhalter (2017)
- graphs are good for gists and different graphics are good at different things (i.e. trends vs comparison etc) but hard to avoid framing (show information as part of a whole) Spiegelhalter (2017)
- scattering in pictograph increases the impression of randomness but makes it difficult to count Spiegelhalter (2017)
- people with high numeracy count pictographs but low numeracy assess area Spiegelhalter (2017) # 4.1 Plot Comparison Papers # 4.1.1 The problem of incomparable plots Visual inference can be seen as a process that combines multiple fields, from mathematics to psychology, to convey meaning. Unfortunately which tasks belong to mathematics, computer science, or psychology is poorly defined. The uncertainty visualisation taxonomies make the blur between these fields. A visualisation is, in a lot of ways, a statistic, or at the very least statistics are calculated prior to the visualisation being calculated, however

this may be invisible to the user. A density plot does not just “appear” from the data, usually a smoothing function (that should be calibrated according to the sample size however how this should be done is a different matter of dispute) generates an estimated density function (statistics) which is then depicted on a plot using a line (design or computer science) which is then converted into information in your head using perceptual tasks and heuristics (psychology).

These overlapping but distinct fields create a high burden of spanning expertise to understand if two visualisation are equivalent in each field. This is largely caused by the concept of “information” differing dramatically between fields. According to mathematics, a sufficient statistic for a parameter contains the same amount of information as an entire sample. Therefore, if you are trying to estimate a population mean, the sample mean and the entire data set are both equivalent, however a similar statistic, such as the median, is not and will become a worse and worse approximation for the mean the more skewed your underlying distribution in. Computer science considered two pieces of information to be equivalent if they come from the same data [This is the justification that I have come across in several computer science field papers, I am not sure if this is true for the whole field so this is definitely a question for Sarah]. This definition of information is much looser and is contained within the mathematical definition of information.

Most uncertainty visualisation papers show that the definition of information *should* be the one defined by mathematics, not the one defined by computer science. A close look at the visual inference literature shows this to be true, but a large amount of noise is added to the research through visual heuristics. This is simply because, if one visualisation depicts a “significant statistic” and another visualisation shows a similar statistic, then the graphic that shows the significant statistic holds more relevant information to the question and will lead to more accurate interpretations. If the information that is depicted on a plot cannot be used to *mathematically* generate a best estimate for a specific statistic, I do wonder how the visual system would be expected to cover that gap if *not* for heuristics.

While mathematical information *should* be considered the starting point in considering if two visualisations are equivalent, it is not the only consideration. A large amount of work in perceptual tasks, attention, and psychology surrounding charts shows that all the information depicted is not paid attention to in equal measure. Elements such as colour can have sizable impacts on the way a graphic is perceived. Mathematics identifies the information that is *shown* while psychology identifies the difference in information *received*. A failure to understand this distinction leads to graphics that are different not only in the way information is depicted but in the information itself. This problem is pervasive in uncertainty visualisation literature.

- The big data paradox shrinks confidence intervals but magnifies bias (Bradley et al. 2021) because the confidence intervals shrink due to the sample size but the imprecise target group creates bias.



## 4.?? Noise vs Signal (question +)

The first issue that arises with uncertainty research is that uncertainty is frequently treated as a signal, not noise. This discrepancy is also noted by Kinkeldey, MacEachren, and Schiewe (2014) who, in their own literature review on uncertainty methods, comment that the literature seems unsure if uncertainty is simply another variable or if it should be treated differently. This is a fair question to ask. If uncertainty is simply a signal we are trying to convey, why should it be treated any different to any other signal in a visualisation?

- if the experiment DOES use noise as a signal, it is often very unclear how the participants should include that in their response and what would be a valid response from participants (e.g. the likelihood to how much question and the cone of uncertainty questions). separate to the data, uncertainty has its own goal. Is it to impact confidence in the estimate (like a bayesian thing)? to suppress signal (like a barely noticeable difference line up plot thing)? to facilitate more efficient decisions (so accounting for risk thing)?
- will it impact their decision at all? there are a couple of papers where you can answer all the questions by just ignoring the uncertainty information because of the way they have set it up. e.g. difference in means when the variance is the same, experiments that check if someone can still get the signal and make sure the uncertainty information isn't in the way.
- The literature also does not discuss if uncertainty is another variable, or if it needs to be treated differently (i.e. is it metadata of another variable). Most papers studies on uncertainty don't consider this issue, and the question remains, should someone have two separate values, or an integrated view uncertain data view Kinkeldey, MacEachren, and Schiewe (2014)

This is best seen with a simple look at some papers (Make into table) Questions where uncertainty is a signal

Questions where uncertainty is a noise

## 4.?? Greater Infoviz Issues

- Since there are so many papers that treat uncertainty as a signal, not a noise, issues in the uncertainty visualisation literature also help to understand some difficulties in the information visualisation literature as large.
- Suggests that it is straightforward to show a value but much more complex to show uncertainty Hullman (2016) (I think this is interesting in relation to how people SEE uncertainty visualisation. Why is uncertainty hard to visualise?)
- Authors provide little justification for their chosen response models (e.g. absolute accuracy vs relative measures) Hullman (2016)

- When asking questions, researchers should focus on events that can be repeated (because of a misunderstanding of confidence intervals) Hullman (2016)
- Uncertainty defined as a by product of the process of making a visualisation, from data collection to communicating Hullman (2016)
- Some questions (e.g. what are the chances that the no.6 bus will arrive first) you can elicit the viewers subjective probability distribution Hullman (2016)
- communication tasks: This category comprises map reading tasks involving data and uncertainty value retrieval (which location is most uncertain?). For tasks from this category, visualisation techniques can be chosen following the traditional rules from cartography. Kinkeldey, MacEachren, and Schiewe (2014) (basically saying why dont you just use what is already here if you are treating uncertainty as a signal, which I agree with)
- papers usually have same variance if there is a signal question, and use the mean as a proxy to answer the question. Really should ask a changing signal question where the signal IS impacted by the uncertainty information.
- Weirdly When setting their ground truth it seems like a lot of papers EXPECT participants to completely ignore uncertainty information as noise. Like the participants are actually considered to be incorrect if they dampen their signal understanding because of noise considerations which I find FUNNY.
- I think the questions where uncertainty is a signal, not a noise, allow the authors of visualisation experiments to ignore the context and motivation that usually governs a visualisation. # 4.?? Tasks and Motivating Questions (uncertainty about.. .what?)
- “cannot assess the quality of risk communication unless the objectives are clear” circles back to motivation. He assumes we are fulfilling a duty to inform. Spiegelhalter (2017) ## 4.?? What are you uncertain about? A mathematical considerations
- Should add plot comparison mathematical framework example here
- take a plot and work through the entire example as a flow chart thing ## 4.1?? In other statistics What is meant by “uncertainty” may seem obvious to some, but when you attempt to quantify or visualise it you will quickly find yourself asking, “uncertainty about... what?”. Do you mean uncertainty on an estimate? On a forecast? How many steps ahead is this forecast? Are we only considering the uncertainty in the estimate or in the parameters or are we considering the possibility of measurement error or biased inputs? Signal and noise can only be untangled in the presence of a motivating question.

The idea that uncertainty can only be defined in the presence of a motivating question is well grounded in most areas of statistics. The entire process of data analysis, from deciding what should be observed as data through to communicating that data in a plot is governed by human decision and the goal of an analysis. At the philosophical level, applied statistics is simply

taking real world entities and boiling them down into probabilistic objects, an ontological process that is largely dependent on our goals (Otsuka 2023). When we move onto data provenance the issue persists, as 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 resolution of the question (Meng 2014). After moving onto modelling this issue continues as each research question can be categorised as descriptive, predictive, or causal, each of which has its own appropriate statistical methods and motivation agnostic model selection leads to signal devoid of meaning (Carlin and Moreno-Betancur 2023). Even at the final stages of visualisation a lack of understanding of the motivating question make it difficult to untangle what is signal and what is noise, leaving many uncertainty visualisation studies with conflicting results (Kinkeldey, MacEachren, and Schiewe 2014). These cases highlight that uncertainty is defined at *every* stage in relation to our motivating question, from data collection to visualisation. Discussions of uncertainty cannot be had if we are not clear *what* we are uncertain about. Once it is established what we are uncertainty about, we can consider the other elements of uncertainty that need to be defined.

#### 4.1?? General task notes

- Mentioned aleatory vs epistemic uncertainty, in the visualisation literature aleatory uncertainty is more commonly the focus. also mentions dependence and joint pdfs as a consideration Hullman (2016)
- Most studies involve intrinsic approaches such as colour, transparency, and those that used extrinsic techniques typically used glyphs, error bars, grid based techniques, or contouring Kinkeldey, MacEachren, and Schiewe (2014)
- most studies used coincident approaches, static visualisations, are applied to a specific domain and dont have generalised results Kinkeldey, MacEachren, and Schiewe (2014)
- The tasks for participants were most commonly value retrieval (some where uncertainty and data value are retrieved separately, others where they are some separately). also also be asked to extract a relative value Kinkeldey, MacEachren, and Schiewe (2014)
- How does the complexity of uncertainty relate to the categories of user and task Kinkeldey, MacEachren, and Schiewe (2014)
- Several studies provide evidence that the usability of uncertainty representations can be highly user and task dependent Kinkeldey, MacEachren, and Schiewe (2014)
- The nature of the task plays an important role for the usability of uncertainty techniques, this may explain many of the inconsistent outcomes from the studies under review since two studies assessing different techniques and user groups are not comparable Kinkeldey, MacEachren, and Schiewe (2014)

- Doing a follow up publication to discuss issues with reasoning and decision making based on uncertainty visualisations which is not considered or discussed in this paper Kinkeldey, MacEachren, and Schiewe (2014)
- Main take away is that “we need to systematize future empirical studies on uncertainty visualisation to better enable comparison and generalisation of the findings”
- One way to advance this goal is a taxonomy, however existing taxonomies are focused on data types, uncertainty categories, and representation types Kinkeldey, MacEachren, and Schiewe (2014)
- he lists a bunch of tasks that were tested with uncertainty visualisations at the end Kinkeldey, MacEachren, and Schiewe (2014)
- risk is uncertainty as a signal, and when uncertainty communication is spoken about this is typically what is discussed, uncertainty as signal suppression is what is usually meant when we talk about visualisations but that area is functionally ignored. This idea, that uncertainty can only be defined in the face of a motivating question, is well grounded in the literature but seldom explicitly mentioned

In section 2.1 it was mentioned that uncertainty must be defined within a specific motivating question, otherwise it inherently does not make sense. A large difficulty with the uncertainty visualisation evaluation studies is that this rule is not followed. There are a shockingly large number of evaluation studies that seem to pay no attention to the information that is relevant to the question they are asking, and show participants a selection of seemingly random visualisations from a statistical point of view.

- inferential uncertainty and outcome uncertainty ARE NOT THE SAME THING??? they visualise DIFFERENT DISTRIBUTIONS

We define two primary motivations for uncertainty visualisation. 1) To prevent deterministic conclusions from a random signal (uncertainty as noise) 2) To convey information about a variance, probability, or other random event (uncertainty as signal)

Uncertainty as signal papers have incredibly predictable results Uncertainty as noise papers should follow a similar protocol to the line up papers

Notes from previous sections that were moved here - The concept of uncertainty being task dependent is *particularly* salient for uncertainty visualisation, and it is repeatedly identified as a problem in previous reviews of the uncertainty visualisation literature (Kinkeldey, MacEachren, and Schiewe 2014; Hullman 2016) as well as across many sub domains and applications (Wallsten et al. 1997; Munzner 2009; Fischhoff and Davis 2014; Meng 2021; Amar, Eagan, and Stasko 2005). The fact that this conclusion is repeatedly reached shows both the importance and the lack of acknowledgement this concept receives. - Fischhoff and Davis (2014) discusses how decisions with different levels of flexibility should have different uncertainty visualisations. If the user has no choice they should simply be made aware of the risk, if they have several

options, the user needs to see the uncertainty in each, if they need to come up with the options themselves, they need to see how things work and how the system fails. - 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. - Statistics is, at its core, the study of uncertainty. Therefore discussing uncertainty visualisation as a separate sub domain to “normal” data visualisation is inherently confusing. What is typically meant by “uncertainty” visualisation is “noise”, that is, we want to present some signal cushioned by its natural variance. Unfortunately, this distinction between “signal” and “noise” is entirely goal dependent. - Bella et al. (2005) found that most participants were ignorant to the fact that error bars are used for both confidence intervals and standard error bars, two wildly different indicators of precision. (**also the paper sherry sent me**)

- Need to compare to an appropriate ground truth, but this is a philosophical exercise Hullman (2016) ## 4.?? Incomparable plots
- But when different techniques are to be compared, it is important that the scenarios and datasets are informationally equivalent, i.e. according to Larkin and Simon (1987, p. 67) ‘[t]wo representations are informationally equivalent if all of the information in the one is also inferable from the other, and vice versa’. A goal in testing, then, is often to determine whether they are also computationally equivalent, or whether one depiction has an advantage over another, i.e. ‘[t]wo representations are computationally equivalent if they are informationally equivalent and, in addition, any inference that can be drawn easily and quickly from the information given explicitly in the one can also be drawn easily and quickly from the information given explicitly in the other, and vice versa’ (Larkin and Simon 1987, p. 67). Kinkeldey, MacEachren, and Schiewe (2014)
- The way we ask questions and the types of questions we ask are selected with little justification. This paper makes suggestions to reduce the noise in the data from these papers. Hullman (2016)
- Most commonly check absolute measures of accuracy, where accuracy is abs(subjective error - actual probability). Some papers also check relative measures of accuracy and ask subjects to find regions of least certainty or to rank targets by uncertainty. studies also consider response time and confidence rating. some studies also ask users for their own expression of uncertainty (e.g. by asking when sketches are uncertainty and ranking various depictions of uncertainty). Some studies ask subjects to make some decision using the data. sometimes things like complexity of task, degree of visual overload, ease of use, visual appeal, and preferences Hullman (2016) # 4.?? Source of Uncertainty
- rarely used extension is aggregation of uncertainty over an area (retrieve overall estimation from a spatial distribution of uncertainty). Other tasks include comparisons, rankings. also have a task called “search” where participants have to identify entities that fulfill certainty characteristics (high or low values) Kinkeldey, MacEachren, and Schiewe (2014)

- Whether to use classed or unclassed uncertainty is rarely discussed Kinkeldey, MacEachren, and Schiewe (2014). e.g. land map with remotely sensed images has ambiguity in the boundaries, which can result from multiple sources (land cover class, measurement error, images from different dates). If you combine them all it might not make sense for someone who only needs one.
- I think the dimension of the uncertainty should be considered every time there is aggregation from multiple sources of uncertainty and you want to maintain those sources. You can project the uncertainty down into one dimension if the source information is in of itself unimportant. # 4.?? Importance (and unimportance) of Secondary and Meta information
- Specifies uncertainty can be represented by three components attribute (what) position (where) and temporal (when) Kinkeldey, MacEachren, and Schiewe (2014)
- Studies dealing with uncertainty information typically deal with attribute but rarely position and time Kinkeldey, MacEachren, and Schiewe (2014)
- 10/44 studies involve multiple types of uncertainty Kinkeldey, MacEachren, and Schiewe (2014)
- When we boil down data into a statistical object, we boil away many aspects on context. Birth and coin flips become an object that is one in the same, and often the visualisation
- Why are spatial and temporal data considered special? What is the statistical element that remains after they have been away? Inexchangeability. Temporal data is one dimensional inexchangeability, spatial data has two dimensional inexchangeability.
  - There are other things that have inexchangeability, and not every question related to spatial or temporal data requires the inexchangeability.
  -
- Yes the data type is relevant but its not the most important thing
  - one with temporal, one with that temporal paper but you didn't know it was temporal, one with the bus time cdf
  - screenshot from these three papers, the connecting factor is the question, not the data type
  - The data type: spatial, temporal, etc, usually is mentioned as the
  - Not a feature of the data, but it is a feature of the question “how do I show uncertainty on a map?”. This is a fine question to ask, but the lack of specificity in what is being done, and the zoning in on functionally irrelevant elements leads the literature as a huge mess.
- Hurricane risk is not only the path, but also the storm surge and wind speed. Spiegelhalter (2017)

- coloured maps for earthquake risk are easily interpretable Spiegelhalter (2017) # 5 Great Examples
- best things you have seen and why they are the best
- can't have a best practices because you need experimental results and my point is that the experiments aren't testing the right things
- pixel map
  - Similar to what could be good but its computation isnt quite what I would like
  - heaps of sample depictions are animated and that is not always possible, but this is a clear indicator of how you can do a sample visualisaiton
  - hasnt been tested so it is unclear if this actually works at signal supression
- VSUP
  - The purpose of uncertainty is clear in the design and it is signal supression
  - Supression is not the only uncertainty feature we should be interested in (shape is also important which is why I have beef with over smoothing and something visualisation is very good to express) but by being clear in its goal it knows what it is
  - It addresses several key issues with the bivariate map that arrise in general and especially when it is used for visualising uncertainty
- the census dot map (has been taken down but CNN replicated it [here](#))
  - I love a good depiction of a sample
  - You can see the sparse regions and the pop
  - it is similar to the pixel map (could also be singal supression but it is not tested)
  - the sample depiction works well with the interactivity and you can see the signal that is appropriate for a particular level of zoomed in. If you zoom out the points turn into solid colours and show clear signal.
- The new york times class mobility and race animated plot [here](#)
- the animated surface classifications from that satelite images experimental paper
- the climate change scenario uncertainty. There are quite a few of these:
  - [e.g.1](#)
  - [e.g.2](#)
  - there are a few examples on the [climate change scenario wiki](#)
  - Good example of uncertainty that has uncertainty from two different sources (human decision in the form of inputs and statistical uncertainty), one of which cant be quantified.
  - a good example in how you can combine two different types of uncertainty

Notes - Since a lot of the good visualisations are made by random people creating a new visualisation (likely through back and forth communication which is HOW you get good visualisations) - A lot of these show the data instead of an inference which is probably better for this scenario because sample size is important for uncertainty. It is not always appropriate and sometimes dots are not exactly equal to one draw. - comments about smoothing in density plots and bin width in histograms. I do think the sample size should be communicated to people. While it is not the only thing that matters it seems kind of pointless to go to such lengths to hide it. - Sometimes points are used for things that are NOT samples which can get confusing and honestly I don't like it.

## 5. Future work/Conclusion

### Appendix (or information I am holding)

#### Experiment paper checklist

General - Specific Field of Research - Question asked - Risk or Uncertainty: - Noise vs signal (risk vs uncertainty) - Major task goal - e.g. identify a bias or misunderstanding, compare two visualisation methods - Minor task goal - Communication(value retrieval) - Analytical(complex consideration) - Exploration(free for all)

Uncertainty Considerations - Source - Level (statistical-ignorance) - Nature (Epistemic/Aleatory)

Visual Uncertainty taxonomy - Explicit/implicit - (directly mapping or showing multiple outcomes) - Intrinsic/Extrinsic - (using existing symbols e.g. colour value, or new objects e.g. grids) - Visually integral/separable - (can or cannot be separated from the data) - Coincidence/adjacent - (if data and uncertainty are in integrated or separate views - Static/dynamic - (animation/interaction)

Other plot considerations - Dimension of data - Dimension of uncertainty - Feature mappings

Extra (not necessarily recorded) - Metrics used (and recorded) - Possible Heuristics - What is the ground truth? - Participant literacy

### 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>.



- Bella, Sarah, Fiona Fidler, Jennifer Williams, and Geoff Cumming. 2005. “Researchers misunderstand confidence intervals and standard error bars.” *Psychological Methods* 10 (4): 389–96. <https://doi.org/10.1037/1082-989X.10.4.389>.
- Bradley, Valerie C., Shiro Kuriwaki, Michael Isakov, Dino Sejdinovic, Xiao Li Meng, and Seth Flaxman. 2021. “Unrepresentative big surveys significantly overestimated US vaccine uptake.” *Nature* 600 (7890): 695–700. <https://doi.org/10.1038/s41586-021-04198-4>.
- 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>.
- 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>.
- 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>.
- 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>.
- 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>.
- . 2021. “Enhancing (publications on) data quality: Deeper data minding and fuller data confession” 184 (4): 1161–75. <https://doi.org/10.1111/rssa.12762>.
- Munzner, Tamara. 2009. “A nested model for visualization design and validation.” *IEEE Transactions on Visualization and Computer Graphics* 15 (6): 921–28. <https://doi.org/10.1109/TVCG.2009.111>.
- Otsuka, Jun. 2023. *Thinking About Statistics: The Philosophical Foundations*. 1st ed. New York: Routledge. <https://doi.org/10.4324/9781003319061>.
- 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](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>.
- 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>.
- Suh, Ashley, Gabriel Appleby, Erik W. Anderson, Luca Finelli, Remco Chang, and Dylan Cashman. 2023. “Are Metrics Enough? Guidelines for Communicating and Visualizing Predictive Models to Subject Matter Experts.” *IEEE Transactions on Visualization and Computer Graphics* 14 (8). <https://doi.org/10.1109/TVCG.2023.3259341>.
- 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>.
- Walker, W. E., P. Harremoes, J. Rotmans, J. P. Van Der Sluijs, M. B. A. Van Asselt, P. Janssen, and M. P. Kraye 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](https://doi.org/10.1002/(sici)1099-0771(199709)10:3%3C243::aid-bdm268%3E3.0.co;2-m).