Noisy Work: A Review of The Uncertainty Visualisation Literature

Harriet Mason

2/7/24

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.

I have not touched this section since the confirmation so it would be best to ignore it

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 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.

- industrial accidents have a "tollerable" 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 of "uncertainty" but most don't even seem to agree on what

2

I have thought more about the idea of defining uncertainty mathematically and I actually *DO* think it is possible. I want the record to show the reason I *did not* do this earlier is that *nobody* in the uncertainty literature has defined uncertainty mathematically, and I thought I was supposed to highlight problems in the literature review, not solve them. I have had an idea of how to formalise uncertainty after seeing Xiao-li Meng's work, but an uncertainty formalisation would solve *several* open questions he has postulated in previous papers, and would be so new it would need to be a paper in of itself, not as a sub-section in a literature review. A mathematical definition would be *completely* new and *very complicated*, and I think its difficulty is pretty clear by the fact that *nobody* and *I really mean nobody*, not spiegelhatter
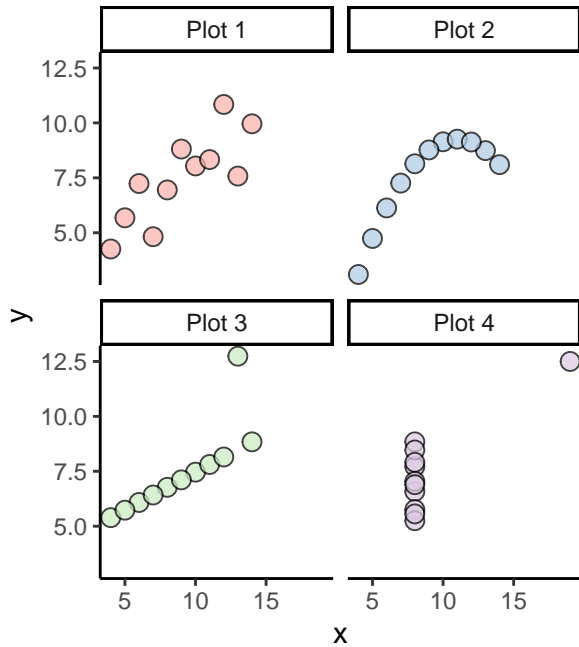
Figure 1: The four scatter plots that make up Anscombe's quartet. The four scatter plots are visually distinct but have the same mean, standard deviation, and correlation. The visualisation highlights the importance of plotting your data to identify interesting features that are hidden by other summary statistics.

uncertainty is. In this section we will establish exactly what we mean when we talk about uncertainty.

2.1 Coloquial 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 probabiltiy, some consider it to be an encompassing umbrella term of which probability is only an element, and some consider it to by 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 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.[1]

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

I am not sure how to collect these together and show that they are highly varied. This might be a situation for the tidytext clustering or something. There is an additional problem is that the definition for uncertainty is *completely* unrelated to the uncertainty that is quantified in the graphics. I think the tidytex can help with the definitions being all over the place, but I am not sure how to identify the lack of correlation between the colloquial definitions of uncertainty and the mathematical quantificaitons.

---

[1]Flesh out note - I think several of the sentences in this paragraph need examples (or quatified evidence) from the spreadsheet definition part

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.[2]

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), 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

I actually think I agree with the sentiment in the grammar of graphics that a taxonomy is more of an opinion and we can reclassify it and is therefore not thorough enough to remain unchanged and be build upon. That being said, if this is a literature review and all existing literature use written definitions or taxnomies, that is what I have to work with. The reason I selected the larger taxonomy is because the other taxonomies all *fit within it* and it can be used to define a space that shows the definitions are inconsistent.

---

[2]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.

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 untanlge the features of uncertainty:

- aleatory uncertainty (unavoidable randomness), epistemic uncertatiny (uncertainty about structure, still quantified), ontological uncertainty (uncertain about entire modelling process). Spiegelhalter (2017)
- measurement (uncertainty on a single varible), spatial (uncertainty in spatial data), temporal (uncertainty in a time data)
- aleatory (inherrent 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, intial state, model operation) and diagnostic (past/current) vs prognostic (forecasting)

- uncertainty is organised according to Satience (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)

Figure 2: Example of different uncertainty taxonomies
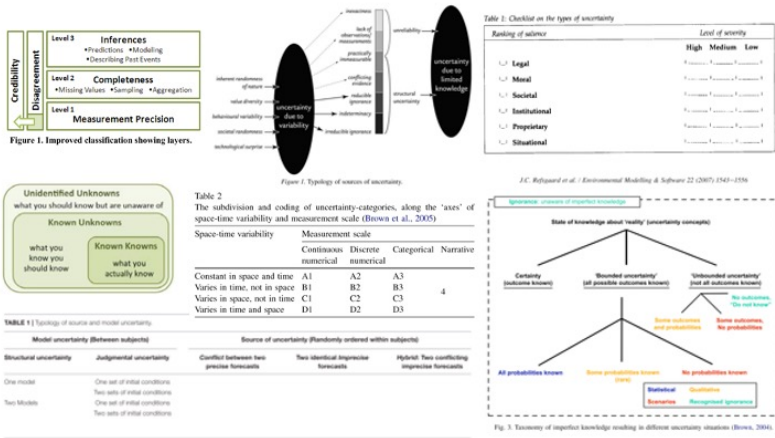
This taxonomy issue creates several problems in the literature. Ideally, if we wanted to combine the literature on uncertainty visualisation, each paper would state the same definition of uncertainty and explain which aspect of uncertainty they are visualising, so the experiments would shift around the space while the definition of uncertainty and what it contains would remains 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

This figure is not going to be in the final version of the paper. It is here as a placeholder. Ideally, I would replace with with a graphic that shows where all these taxonomies land in the location, level, nature taxonomy (since it is encompassing) and show they are very inconsistent and a new definition is pulled out for almost every paper (even in statistics papers)

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 commincate.

This does not mean there is nothing to be learned from these taxonomies, each in their own way show what apsect 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 3 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 to 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.

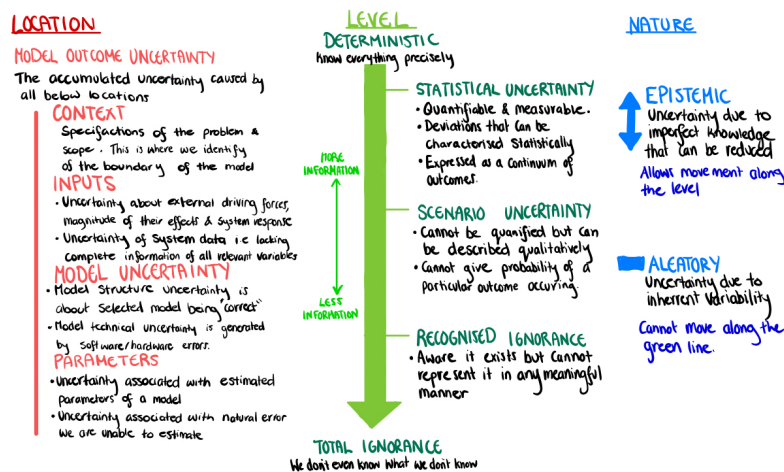*I need to fix the taxonomy illustration because inputs and model uncertainty need to be swapped*

Figure 3: 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,

I had this here to explain why these are the considerations people make in statistics and how they relate to elements of the taxonomy to try and provide some justification, but If I am just going to show the definitions are inconsitent, I might get rid of this section. I also largely included it so you guys can see what the taxonomies are considering

9

visualise, or communicate uncertainty. Organising uncertainty according to its location, level and nature makes allows us to consider uncertainty in the miriad 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 7 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 considred 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 visualsiation 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,
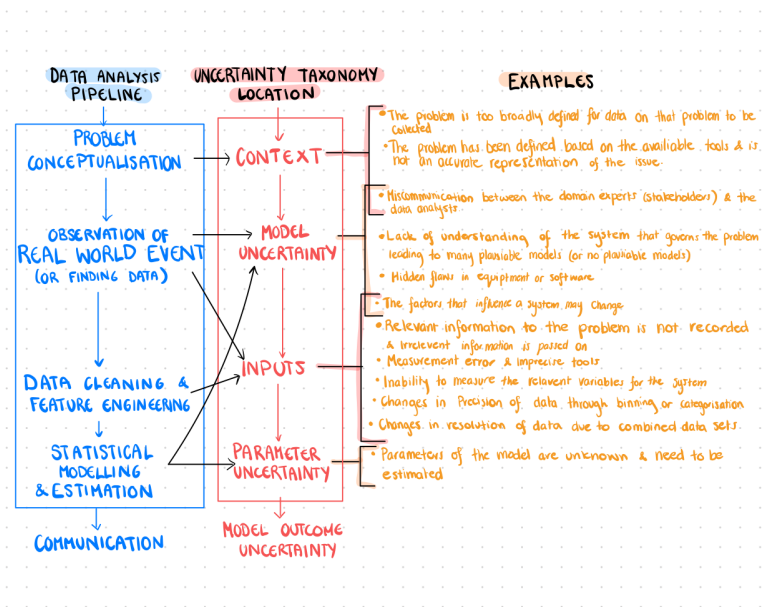
Figure 4: Illustration of the connection between the data generating process and the location of the uncertainty analysis. Several examples are collected from utypo and Munzner2009.

11

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 combined, 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 if 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 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

12

specific case and should be used going forward to define what aspect of the uncertainty space a contribution exists in.

Context comments

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 do with quantifying uncertainty, and the difficulties in communicating a complicated statistical model through a complicated map is already well discussed in the visualisation

I have this here to explain why context based taxonomies can be ignored when we are considering uncertainty. I did not go into too much detail because it would involve discussing visualisation which I am trying to avoid at this point. I actually think this quote from the grammar of graphics is exactly what I was trying to say. *"The system in this book is not a model for a GIS, however, because geography and statistics differ in a crucial respect. Geography is anchored in real space-time and statistics in abstract dimensions. This is a distinction along a continuum rather than a sharp break; after all, there is a whole field called spatial statistics (Cressie, 1991). But this difference in focus clearly means that a system optimized to handle geography will not be graceful when dealing with statistical graphics"*. Basically, my point is this section is still discussing the definitions of unertainty, we have not gotten to disussing visualisation so we are still in the abstract *statistics* domain. The only parts of spatial statistics that we care about in this domain are the statistical properties that come from spatial information, and *none* of those statistical properties generate uncertainty. This is not unique to the properties of spatial information, other statistics information that is important for modelling and assumptions, such as non-linearity, class levels, etc do not contribute to uncertainty. Operations on data, inherrent randomness, and missing information generate uncertainty, which is what we are talking about here. A confidence interval is a confidence interval whether it came from spatial data, temporal, or cross sectional data. *This does not mean I am saying spatial and temporal information is not valuable ever*, I am just saying it is not valuable *yet*. This information will be *highly relevant* when we discuss context and translating from the abstract domain back into real world for communication. Since we are still in

13

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 5. 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 adjustment in how we conceptualise and organise statistical methods would likely benefit other domains within the field.
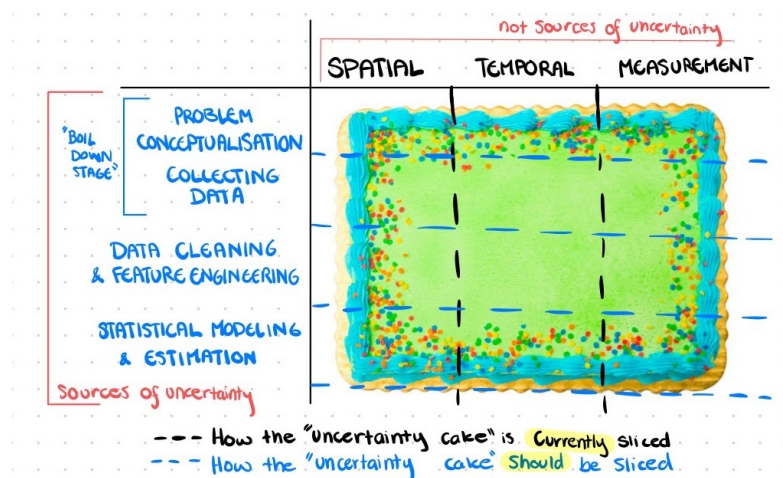

Figure 5: The metaphysical cake we are slicing differently

Maths comments I might relate back to an actual mathematical definition

It is important to note that this taxonomy is the closest we can get to mathematical definition of uncertainty. While individ-

In the "semantics" section of the book, Leeland Wilkinson provides a series of definitions of words related to uncertainty and mathematical definitions of examples of uncertainty, but those are not mathematical definitions of uncertainty. Every *mathematical* definition of uncertainty is a definition of *probability* OR a definition of an *example* of uncertainty, such as a confidence interval or PDF. I included this here, however, I actually now believe there

ual *expressions of uncertainty*, such as a confidence interval or a PDF are mathematically defined, the term *uncertainty* cannot be defined mathematically because it is a layperson word that has been adopted by statistics to describe any deviation from complete determinism, whether mathematically defined or not. While the broader field of uncertainty visualisation does need to be more closely inspected, specifically through the scope of mathematical definitions, this issue cannot be solved with a more specific definition of uncertainty through mathematics. Instead or establishing a new definition of uncertainty that is strict enough to prevent the current array of misunderstandings, we should focus on inspecting how the current literature *deviates* from established mathematical definitions, why these deviations occur, and understand how these deviations can lead to larger problems in uncertainty communication.

An explanation of the building blocks for my mathematical definition

The first problem in the conception and attempts to offer a definition of uncertainty, is that bias *and* variance are both technically uncertainty.

Basically the idea is that every analysis has some goal, and every choice we make that cuts out several alternatives is a "restriction" that trades the variance of multiple options into a potential bias. This occurs at *every stage* of the data analysis pipeline, from collecting data to communication. Bias comes from "incorrect information" being passed along the pipeline, while variance comes from "holding on to information". The randomness that is "irriduciable error" (if it exists or not is not relevant to our definition) cannot be differentiated from the bias and variance (I say both bias and variance because at the end of the pipeline it is impossible to know if error is being caused by bias or variance) unless we literally know the function that generated the data. We typically talk about bias and variance in terms of the model, but it exists at every stage of the pipeline. Since every step of the analysis pipeline has an almost infinite set of choices (there is an infinite set of things we can observe, an infinite number of choices that can be made in data cleaning, etc) each of these levels is functionally its *own*

I wrote this section between me deciding it would be possible to mathematically define uncertainty and me deciding it would be too large for this paper. I am aware that this is not a formal definition and I cannot say this is the unifying feature of uncertainty without some proof. But, this is something that *can be* defined in mathematical terms and *can* lead to an encompassing definition of uncertainty (i.e. it would be something that the groups of confidence intervals, error bars, and missing data using an information filtration) but that is beyond the scope of this paper. This way of thinking about uncertainty has resolved a lot of my issues with the field and a lot of other issues that have been raised by authors in their summary of the literature. This is very similar to Xiao-li Meng's work that asserts that bias and variance are actually the same thing, with a couple of changes.

random variable. When you make your choice you completely shrink the variance of that level to 0 (it basically goes from a distribution of choices to a single outcome of that distribution) and, the amount this *guess* is wrong by, is bias.

This idea is very similar to the bias-variance trade off, but it is slightly different because we are assuming there is not a trade-off perse, but that bias is just "cashed in" variance. If selections at each stage are made using feedback, we can reduce the bias without increasing the variance, but it "costs" some data. Assumptions let you reduce the variance with a low data "cost" but there is a higher chance of being incorrect (since there is less validation) and introducing bias. Flexible models like neural networks can reduce the variance but it comes at a high data "cost". If you have a *lot* of data it is better to avoid as many assumptions as possible because increased certainty that comes from more data can only effect error that is still in the form of variance (i.e. we are considering alternatives). This is well established in modelling, it is why big tech companies try to collect as much data as they can and then put it in a neural network and hope they have enough data and enough feedback to avoid the bias at *all* stages of the analysis. Recording *everything* and putting it in the most flexible modle possible means we don't even have to put any poles down, even at the data collection stage

Validation methods (such as cross validation, assumption checks, etc) help to guide the choices we make in the data analysis pipeline, but they can never offer any kind of certainty or truth considering validation methods themselves are also only estimates. Once a variance has been cashed in for bias, its error cannot be reduced by more data.

When we consider the "uncertainty" of a model we usually only consider the variance of the final stage, the prediction error for *a specific model and a specific set of features*, and ignore the variance that has been converted into bias by our selections that was introduced at earlier stages of the pipeline. Since uncertainty is basically "how true is this estimate or statement", if we had not cashed in our variance for a single outcome of bias, it would appear as variance. The climate scenario uncertainty plots are an example of bias that was converted back to variance

for the sake of the visualisation (time series forecasts typically assume away the bias by stating "we are assuming previous trends will continue into the future"). Prediction error is a useful metric because it allows us to see if the bias introduced in earlier stages expressed as variance through an error, however it still cannot account for bias introduced at the sampling stage as that bias also exists in the test set.

Prediction error is not always feasible or useful in a statistical analysis (such as when we produce a visualisation) so it is unclear exactly *how* we should reintroduce that bias as variance to communicate the pair as uncertainty. It is not uncommon for uncertainty to only be considered at the final stage of an analysis as a confidence interval or estimate range, but this is typically misleading as earlier bias has not been converted back to variance. Factors that typically decrease the variance, such as sample size, can only have an effect on something that is still in a state of variance. This means that an established model that is build on more and more data will have confidence intervals that are incredibly small but if it is build on some bias (such as a biased sample) the confidence intervals will become misleading and will be too small to *actually* predict future data. It is not uncommon for true results or models produced by a different teams to sit entirely outside of the confidence band of another. By hiding the selections we made at earlier stages, when we traded that variance in for bias, we functionally made *point predictions* (they were based on educated guess but point predictions none the less) but *these* point predictions are often invisible to us.

A Forest analogy

Let us consider the data analysis pipeline a dense forest that you need to cross to get to your house. To prevent yourself from getting *too* lost you decide to bring some poles and rope so you can mark where you are going. You know your rope is long enough to get you to your house with enough wiggle room that you can reach the 5 houses either side of your house. This problem is drawn in Figure 6a. Figure 6b depicts the typical data analysis pipeline and resulting uncertainty using this forest analogy. In this method, we navigate the forest by

putting one of our poles down every time we notice the trees have changed (i.e. we have move to a new section of the forest). As we move down the line and we put our poles down, we cannot be certain we are always walking in the direction of our house, but we can do small checks or validation to make sure we are vaguely heading in the right direction. When we get to the final stage, we have enough rope remaining to get to the three houses directly in front of us, and we state that band as our "uncertainty". Figure 6c depicts the bias variance trade off in the case where we make no decisions, record as much information as possible, and check every single model with every single variable and keep every alternative as a possible option. This method will result in no bias (although it is often impossible to check the entirety of these spaces so it cannot) but a huge amount of variance. In an effort to avoid putting down any poles, we have managed to navigate the forest, but have essentially no idea which of the 10 houses are ours. Of course, with sufficient feedback we can whittle down this vast array of choices (which is what a neural network can do) but feedback from an increased sample size can only influence variance, it cannot influence bias that was previously variance. Figure 6d shows an alternative way to perform the data analysis process to consider variance at stages beyond model uncertainty (such as the scenario uncertainty that comes with different human choices in climate change models). In this scenario you navigate the forest to a certain point, and then clone yourself three times and each put down a new pole at different distances to each other. All four versions of yourself will then walk in an identical fashion from the pole to the end of the forest, and each have your own path with your own confidence interval. This is how the climate scenario uncertainty is constructed.

The influence of the "goal" of an analysis

Prediction error is also useful because the *goal* of our analysis is clearly defined in a metric. This is not always the case. 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
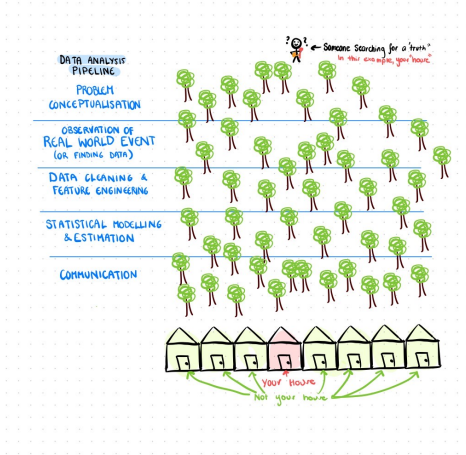
(a) Data analysis Goal
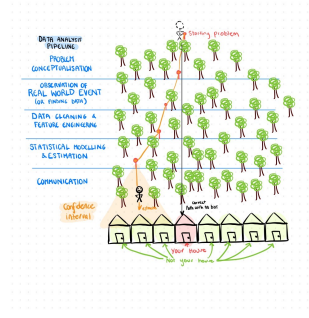
(b) The typical modelling process
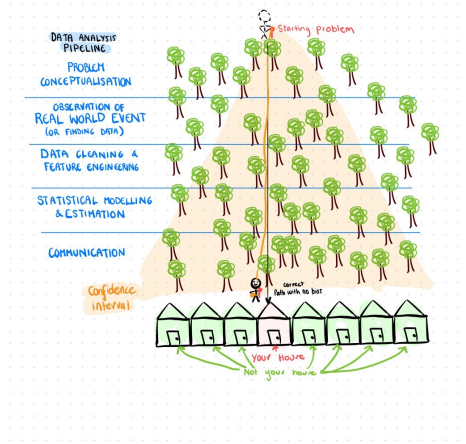
(c) Making zero assumption at every stage

(d) An example of "scenario uncertainty" with a confidence interval

Figure 6: An analogy of walking through the forest to help understand how our conceptualization of uncertainty and bias can influence what we consider to be uncertainty and what we consider to be bias.

19

possibility of measurement error or biased inputs? These questions start to highlight another invisible problem in uncertainty definitions, signal and noise can only be untangled in the presence of a motivating question.
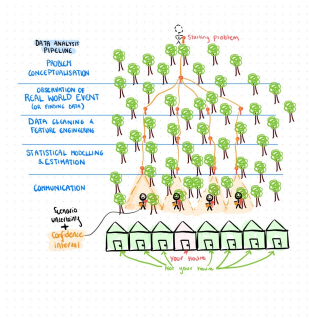
The idea that uncertainty can only be defined in the presence of a motivating question is well grounded in most areas of statistics. The entire process of data analysis, from deciding what should be observed as data through to communicating that data in a plot is governed by human decision and the goal 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 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. 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.

This concept is a little harder to define mathematically, but we *can* consider that every analysis has some metric that is the target "goal" of the analysis. Refusing to define this target goal from the begining does not mean you avoid this problem, it just increases the bias and variance issues described earlier, because there is no way to check assumptions or correct the variance with data. This is actually why I was thinking about the concept of *suffient visual statistics*. If a graphic has a piece

of information mapped to it (which we should know because we made the graphic ourselves) then that information can be extracted with a percecptual task and the graph has a *sufficient visual statistic*. I also think the idea of sufficient statistics is useful because it establishes that we can only know if we have extracted all the relevant information from our *if a metric we are collecting data for has been defined*. If a piece of information cannot be extracted with a perceptual task, it can only be found with assumptions and we introduce a bias at the communication stage. If the information does exist but it reuires a heuristic or some collection of perceptual tasks that might be harder for some people than others, we have maintained variance at the communication stage. A complicated visualisation with a *huge* amount of information can technically be used to answer a lot of questions, but the mental overload would create variance. This is why it is important to have your *goal* in mind when designing a visualisation and not assume it can or should be able to do everything. I suspect we can work out retroactively which visualisation contain which visual statistics, by systematically working our way through the grammar of graphics and then information that is contained in simple perceptual tasks.

Even if this cannot be used to give a mathematical definition of uncertainty, it is still soemthing that *should* be considered. This might be obvious when it is read, but the conversation about whether or not we should consider what was previously variance to be variance when we communicate our analysis, or if avoiding defining a *goal* metric can have detrimental effects on our communication are things that are rarely discussed although it is important.

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

Di said this has already been shown and I believe it, but then I don't understand why the uncertainty visualisation literature seems to ignore it. This is what I was trying to show with the plot and question section of the excel spreadsheet. I was trying to highlight that when you ask participants to answer a question, to mark them as correct or incorrect you need to define a metric, and if the plot was created using a pipeline that *ignores* that metric, it will be passing on incorrect information and lead to mistakes when it is being read. Expecting any plot to be able to answer any question that has the basis of some random metric simply because the metric is related to uncertainty and so is the plot seems ridiculous to me, but its harder to find a paper that *doesn't* do this than a paper that *does*.

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 give 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 "continum of discreteness" that scaled from "point estimate" to "continuous distribtion".

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.

Figure 7: Example of different uncertainty visualisation taxonomies

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 visualisaiton 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 visualsiation 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 (Uncertainty Communication Boradly)

It would be impossible to discuss uncertainty visualisation without mentioning the broader conversation around uncertainty communication. The conversation around uncertainty communication is often centered around the psychology is communicating risks.

Most research into general uncertainty communication seems to focus specifically on how we communicate risks. Risk and uncertainty are not perfect synonyms of one another. Spiegelhalter (2017) makes the distinction clear by specifying *risk* as *a precise random event* (such as a coin flip) while *uncertainty* (also sometimes referred to as abiguity) is imprecision (such as the estimated probability associated with a coin that might be biased). For this reason, risk is often seen as a specific case of uncertainty and that well defined sub area has a lot of

This section is supposed to be an explanation of the general literature, but the definition problem makes it hard to sort the general literature. While there is that overview from spiegelhatter, it is mostly a discussion on risk communication, communicating ambiguous information is not discussed much. The focus on risk (uncertainty as signal) and not uncertainty (uncertainty as noise) is an issue, but the bigger problem is caused by the vagueness of uncertainty issue. Since a paper will be titled "uncertainty communication" it could be any number of things mathematically, but 90% of the time is it just communicating risks or uncertainty as signal (and very rarely uncertainty as noise). I can only tell which one a paper is *once I am in the method section* and even then it is sometimes not explained. Also, since I haven't been using this metric since the beginning I need to go back and check this for all the papers. I also think this section would benefit from the table, I can quantify how many papers communicate uncertaitny as signal

detailled research. Some markers of probability that have common uses, probabilities such as 0 (impossible), 0.5 (a fair coin flip), 1 (certain) are easy for people to have an internal sense of *hullman paper*. It is incredibly hard to communicate small risks, because people cannot differentiate between 1/100 an 1/1000, and communicating low probability but high impact risks are particularly challenging (Spiegelhalter 2017). This issue can be somewhat alleviated by communicating the likelihood of an event in relation to another event on a similar scale (such as being struck by lightning) (Spiegelhalter 2017). The framing of how we communicate risks matters due to the affect heuristic, so there will be a difference in peoples behaviours if, for example, you tell people they have a 90% chance of survival vs a 10% chance of death, and when expressing a frequency, a larger numerator communicates a larger risk (Spiegelhalter 2017). The reference class of a particular risk should be explicitly stated to avoid confusion, for example if we are communicating the chance of rain, the time period over which we are expecting this chance of rain (i.e. 50% chance today or this week) needs to be specified (Spiegelhalter 2017). The problem with a risk focus is that it is only a small piece of the uncertainty pie, and very few of these findings can be reasonably assumed to go beyond risk communication. Even less can be assumed to translate to uncertainty visualisation.

this is all currently from the spieglehatter paper because that is the only paper I have moved notes over from mendeley before I started doing the summary write up. this will not all be from one paper, dont worry.

Risk and uncertainty are not only different in how they are communicated, but also why they are communicated. Risk communication focuses on communicating probability around events. These events are typically a binary event that is known to be random, so a version of communication that does not mention risks (e.g. telling someone they will or will not get cancer instead of communicating their risks about it) inherently carries with it a lack of transparency. Additionally risk communication is often communicated to avoid or informa about an unfavorable outcome, so framing can be important if we want to influence decisions in a certainty way. These motivations do not extend to the communication of uncertainty information

This section still needs some citations from papers that communicate uncertainty, but the are *hard* to organise because of the ambiguous uncertainty definition problem. So they will all be communicating something different but you cant find out what it is until you are knee deep in the method section.

more broadly. While uncertainty information *is* often used to improve decisions, it seems the main motivation in its communication is to offer transparency to the viewer about a statistical analysis. That is, it acts as a form of statistical "hedging" for signals found in an analysis. Interviews with experts in statistic back up this primary motivation, as ignoring uncertainty information of often expressed as being similar to fraud or lying and the goal of "improving decisions" is only seen as a secondary outcome of an appropriately hedged signal (Hullman 2020; Manski 2020). This concept of uncertainty means that the work on communicating risk does not cover the field of uncertainty communication as a whole. It covers the specific case where the signal itself *is* uncertain, but this work does not naturally extend to the goal of hedging a signal (i.e. when we want to communicate uncertainty as noise). Unfortunately understanding the relevant noise seems to be its own problem, as statisticians and visualisation authors seem to be unaware or confused in how to calculate the uncertainty related to an analysis (Hullman 2020). Additionally, communicating uncertainty such that it conveys the appropriate singal supression involves combining uncertainty from multiple different sources, something that is considered to be difficult by even experts in uncertainty communication (Spiegelhalter 2017).

There are some elements of risk communication that seem to translate to uncertaitny communication more broadly. Using words to communicate risk is discouraged because people can misinterpret the actual risk involved, so when possible numerical estimates of risk should be communicated, however this can make information harder to understand for those with poor numeracy skills (Spiegelhalter 2017). People performing better and preferring with numerical estimates of risk translates to uncertainty communication more broadly *inc (citations for this)*. Risk communication also needs to have clear objectives, use plain language, limit information to only what is necessary, and segment the audience to allow for differences in interest and knowledge (Spiegelhalter 2017). These general concepts of communication also extend to communication of uncertainty. On the other hand, it is not sensible to assume that probability specific concepts such as framing, intuitiveness of probabilities, or reference classes are still relevant when we consider uncertainty

communication as signal supression, not as a synonym of risk or probability communication. In this same vein, most uncertainty communication papers very often discuss risk-aversion as something that needs to be considered despite it being unclear exactly how risk-aversion is related to uncertainty communication *(citations for the papers that do this)*.

In the same sense that risk and uncertainty are different, risk-aversion and amiguity-aversion (which can be considered the uncertainty version of risk aversion) are also not the same. Risk-aversion is the tendency of people to prefer outcomes of low uncertainty to those outcomes with high uncertainty despite the outcomes with high uncertainty having a higher or equal expected outcome. In the case of communicating risk, the only reason this heuristic would be relevant was if risk-aversion was a mistake that needed to somehow be corrected for, rather than a heuristic that *reveals the utility of decreased uncertainty*. Risk-aversion is often treated as a mistake or something that should be avoided by the uncertainty visualisation literature even though that is not necessarily true, depending on what is causing the risk-aversion (e.g. is it due to poorer estimates or increased awareness of negative outcomes). If a particular communication method leads people to place a greater value on certainty or make worse estimates, that is the aspect of importance, not risk-aversion which conflates these factors. The work that advocates for transparency in risk-communication gives you whiplash when it's subtext argues that transparency leads to "incorrect" conclusions. Risk-aversion has a large number of confounding factors that make it difficult to understand if it is something to be considered in risk communication, ambiguity-aversion has a similar problem. Ambiguity-aversion is the tendency of people to prefer to take on certain risks rather than unknown risks. It is essentially risk-aversion applied to risk. For this reason discussions of ambiguity-aversion have the same issues as the conversations about risk-aversion, where they do not consider ambiguity-aversion to be an estimate of the utility of decreased uncertainty, it contains a subtext that argues against transparency, and it is considered a valuable insight in of itself despite conflating many factors.

The idea that risk and ambiguity aversion reveal the utility of decreased variance seems to seldom be considered in the lit-

erature, therefore it should be mentioned here, otherwise this issue might continue. However, considering risk aversion is at its core, a trade off between bias and variance, it is bizzare to assume it to be a mistake when statisticians perform these trade offs all the time. If risk-aversion is truly "irrational" behaviour, every statistical textbook that discusses the use of a biased but consistent estimator in a finite sample case (a common example would be specific cases of the maximum likelihood estimator) should be tossed out. Discussing risk and ambiguity aversion as though it is some diversion from a perfectly rational, mathematical choice, implies there is *no value to a decrease in variance which is not even true within mathematics*. The concept of risk aversion treats risk as "noise", as though it is something that only exists to be ignored, but if the common place use of biased MLE's is an indication of anything, it is that *ignoring uncertainty* is the irrational choice.

Trust is another concept that is related to uncertainty communication, but fact that it is often front and center in the discussion has a somewhat ominous subtext. If the purpose of displaying uncertainty information is to appropriately hedge a signal with noise, then it should be assumed that trust is only related to uncertainty communication through increased transparency and honesty. It does seem to be the case that uncertainty communication increases trust because it is a proxy for *(add citations)*. Unfortunately a large proportion of the literature discusses trust as something that is directly related to uncertainty visualisation rather than as an observable product. In viewing uncertainty communication as directly related to trust, not related through the proxy of transparency, several unobserved variables are conflated. The amount of uncertainty that is present in the information, whether or not the information is trustworthy or makes sense, prior beliefs of the participants, and the trustworthiness of the source are a few examples of variables that are conflated when authors directly consider trust to be of direct interest. Similarly to the issues in risk-aversion directly considering trust, and not transparency, as the metric of importance in uncertainty communication leads to a questionable subtext that argues against transparency. Being concerned about audiences perception of trust, without first establishing if what we are communicating is *trustworthy*, leads

to the implication that details that result in information not being considered trustworthy should be avoided.

Authors that imply that a decrease in trust or an increase in ambiguity or risk aversion are metrics of importance in of themselves do not understand the purpose of uncertainty communication as methods to increase transparency. Science communication should be primarily concerned with accuracy, setting trust and risk-aversion as the variables of interest implicitly encourages statisticians to set trust and risk-aversion as the primary goals of communication. The issue of trust being divorced from trustworthiness has been commented on by other authors (O'Neill 2018), however the issue still persists in the uncertainty visualisation literature (Zhao et al. 2023). Additionally, it seems that risk-aversion being divorced from the value of risk has gone unnoticed.

While this is a focus on uncertainty visualisation, these issues in uncertainty communication need to be acknowledged because they bleed into the uncertainty visualisation literature. Uncertainty visualisation are compared on their ability to communicate explicit risk, elicit trust, and prevent risk or ambiguity avoidance, continuing these issues in uncertainty communication at large. While uncertainty visualisation can make it easier to communicate the complicated details of uncertainty information, it carries some challenges that are unique to the visualisaition, specifically with respect to the "signal supression" concept. In simple estimates or verbal communication, the signal is often easy to identify because it is what we are explicitly saying. Visualisations are used in both data exploration and communication. This means what exactly is a *signal* in any particular visualisation is hard to identify, since we often let the visualisation *tell us* what the signal is. Additionally, you cannot add noise to *every single possible* signal one might take from a visualisation. Two people looking at the same visualisation might, just by chance, develop two entirely different insights. These unique and facinating challenges that are faced by viewing uncertainty visualisaiton through the lense of "signal supression" have been almost completely untouched by the literature.

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 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 Uncertainty Visualisation Literature Summary

4.1 Similar work

There are a wealth of uncertainty communication and visualisation literature reviews that already exist.

Kinkeldey, MacEachren, and Schiewe (2014) summarised the literature specifically for geopsatial uncertainty visualisations and attempted to describe similarities in their characteristics. They found large difficulties in the ability

- 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 eigour in empirical methdos 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 toal is to improve a particular product rather than create general principals 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 cognative theory and provides a framwork Kinkeldey, MacEachren, and Schiewe (2014)

Again, this is why I had that table, because these papers are impossible to organise without some metrics. Things that are highly important for uncertainty as I understand it, are not mentioned at all or only mentioned knee deep in the method, so reading and organising these papers such that the information can be combined is a bit of a nightmare. If I *dont* find this information uncertainty, I will end up with the same problem as *all* the other uncertainty visualisation literature reviews, which is that the information is *all conflicting and it is impossible to identify any results.* The other literature reviews *clearly state this.*

31

- 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)

- 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 signes 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 dont believe in the lar of large numbers Hullman (2016)

- experiencing an event makes your percieved probability higher than if you see a description of it Hullman (2016)

- 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 farmiliar 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 incorperating 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)

Spiegelhalter (2017) presents an literature reivew of uncertainty communication as a whole and considers uncertainty visualisation as a subfield of communication as a whole. In this review he has several key findings that are of particular interest. First, the best visualisation is highly dependent on your audience, and depending on how broad your audience is, there may not be a best visualisation. No single representation will be accessible to all members of an audience, and even if most audience members can read a plot, they may not read it in the same way. For example, people with high numeracy skills count pictographs, but people with low numeracy assess area, and seemingly unrelated elements may not equally influence all these methods of reading, for example scattering in a pictograph makes counting particularly more difficult. For these reasons, it is better to

understand your audience and iterate towards a final design, rather than only consider one option. Second, the best visualisation for one task is unlikely to be the best visualisation for another task. Graphics are good for "gists" and different graphics are good for different "gists", e.g. one plot may we good at identifying trend, while another might be good for comparison Spiegelhalter (2017). Therefore, if we change the goal of our visualisation, and therefore the metric,

Visualisations that improve accuracy, visualisations that improve behavioural change, and visualisation that viewers like may not necessarily all be the same visualisation Spiegelhalter (2017).

He highlights a theme that is a common theme in the uncertainty visualisation literature, which is no visualisation will cover enough information, graphics are good for "gists" and different graphics are good at different things (e.g. identifying trends vs comparison). we consider a summary table as a visualisation

illuminate graphics with words and numbers, design graphics to allow a part-to-whole comparison on an appropriate scale, include appropriate narrative labels (such as magnitudes through tick marks), use 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),

These literature reviews highlight a common theme, In contrast to these literature reviews, we do not have a specific domain or tool analysis we intend to perform. Rather, the point of this literature review is to understand the *purpose* of visualising uncertainty (i.e. signal supression), and understand if the current visualisation tools are sufficient to achieve that purpose and if not, where there is need for improvement. Additionally, rather than discussing "uncertainty visualisation" as its own field, we view it as a subfield of both uncertainty and visualisation, and understand if this change in conceptualisation can alleviate or

help us understand some of the issues faced by previous authors attempting to summarise the uncertainty visualisation literature.

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 percieved. Mathematics identifies the information that is *shown* while psychology identifies the difference in information *recieved*. 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 respose 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 supress signal (like a barely noticiable 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 isnt 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 dont 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 commucating 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?)

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
- tasks e.g. "What is the average of X" (how well lined up you are moving down the data pipeline) and resolution (uncertainty is very often moving back a resolution level in statisitcs - also can be an assumption - bootstrap is a replacement for an assumption, not a resolution change, but bootstrapping is centered on the data) e.g. "What is the noise around the estimate of the value of X"
- "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)

4.1?? General task notes

- heath had a wide range of different metrics they adopted to communicate chronic risk of adverse events in the future Spiegelhalter (2017) (use this for motivation based

statistics since they define new statistics based off what they need)

- Mentioned alleatory vs epistemic uncertainty, in the visualisation literature alleatory 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 retrival (some where uncertainty and data value are retrived 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 cateogies of user and task Kinkeldey, MacEachren, and Schiewe (2014)

- Several studies probide 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 insonsitent outcomes from the studies under review since two studies assessing sifferent 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 visualisationsat 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 supression 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 evalusation 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 a 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 approripate ground truth, but this is a philisophical exercse Hullman (2016) ## 4.?? Incomparable plots

- But when different techniques are to be compared, it is important that the scenarios and datasets are information-

ally 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 abolute measures of accuracy, where accuracy is abs(subjective error - actual probabiltiy). 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. ometimes 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)

43

- 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
  - – This has some good maps and some bonkers ones

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 dont 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.

———. 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.

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.

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.

———. 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.

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

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.

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. 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.

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.