

Noisy Work: A Review of The Uncertainty Visualisation Literature

Harriet Mason

2/27/24

In reviewing the literature of uncertainty visualisation

Table of contents

1	Introduction	1
2	Similar work	3
3	Defining Uncertainty	5
3.1	Colloquial Definitions	6
3.2	Taxonomy Definitions	7
3.2.1	Taxonomies of Uncertainty	7
3.2.2	Taxonomies of Visual Uncertainty	13
3.3	Contextual Information	16
4	General Attitude Towards Uncertainty (Uncertainty Communication Broadly)	20
5	Uncertainty Visualisation Literature Summary	23
5.1	Types of Uncertainty Visualisation Papers	23
5.2	Defining uncertainty with respect to inference	25
5.3	Uncertainty evaluation experiments repeat visualisation studies	29
5.4	Measuring uncertainty as noise, not signal	35
6	Great Examples	40
7	Future work	46
	Bibliography	48

1 Introduction

Edit note: this section is untouched since the confirmation. Might need a rewrite

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

2 Similar work

Edit note: I think with the addition of more summary papers I should just summarise common themes of issues the papers ran into instead of detailing the work paper by paper.

There are a wealth of uncertainty communication and visualisation literature reviews that already exist. Interestingly they all seem to have identified similar pitfalls or fallen into one themselves.

Kinkeldey, MacEachren, and Schiewe (2014) summarised the literature specifically for geospatial uncertainty visualisations and attempted to describe similarities in their characteristics. They found most experiments the methods for uncertainty visualisation evaluation to be adhoc, with no commonly agreed upon methodology or formalisation. They found that experiments took on an engineering approach, where plots were evaluated with the goal to improve a particular representation, rather than to investigate why particular representations do or do not work and create general principals. Authors would then ignore the engineering perspective they had taken on, and try to generalise beyond the specific constraints of their experiment. While they did discuss the results of some perceptual task experiments they found that even those results were somewhat conflicting and difficult to combine in any meaningful way. They

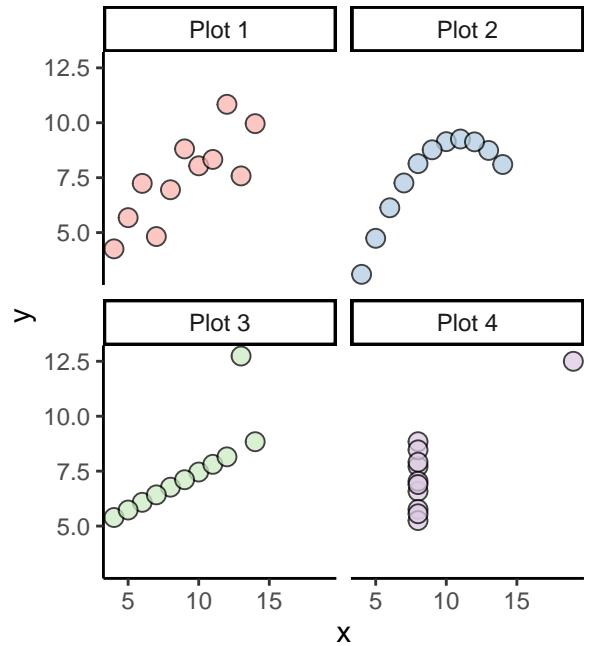


Figure 1: The four scatter plots that make up Anscombe’s quartet. The four scatter plots are visually distinct but have the same mean, standard deviation, and correlation. The visualisation highlights the importance of plotting your data to identify interesting features that are hidden by other summary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

also stress the importance of selecting the correct audience, as numerical and visual literacy can differ vastly between subpopulations so the results from one populations may not reflect the results from another. The paper contains an entire paragraph where experiments that compared box-plot representations are rattled off, each experiment finding results that conflicted with its predecessor. Ultimately they suggested the visualisation typologies move away from data types, uncertainty categories, and representation types and towards “task-centered typologies” to better enable generalisations of the findings.

Hullman (2016) reviewed the uncertainty visualisation literature more broadly (not specifically for geospatial uncertainty) and offered advice on methodology and analysis of results for other uncertainty visualisation experimenters. She found similar issues in the literature and also noticed the somewhat ad-hoc methodology though the lack of justification in the method for the particular questions participants were required to answer, and the lack of generalisability of comparing plots one by one to create a ranking. She also discussed the role of heuristics in the analysis of results and suggested authors look for signs of them in responses and ask subjects to describe their strategy in their response. What was particularly interesting was people’s tendency to ignore uncertainty in a plot all together, and use a pattern in the signal (such as distance between means) to answer questions that require uncertainty information. She also discussed the lack of clarity in exactly what “uncertainty” authors are trying to ask about. Whether the participants are supposed to answer according to some ground truth or their subjective probability is unclear, as is whether the *uncertainty* depicted represents the uncertainty of a sampling or population distribution. Ultimately all these issues become conflated and it becomes impossible to tell if subjects did poorly in an experiment because they misunderstood a visualisation, because the question was misinterpreted, or because they used a heuristic.

Spiegelhalter (2017) presents a literature review 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 be good at identifying trend, while another might be good for comparison. Therefore, if we change the goal of our visualisation, and therefore the metric we are trying to evaluate, we implicitly change the best visualisation as no visualisation will both cover and summarise enough information to be the best at answering every question. Finally, the visualisation advice he gives is more specifically about general visualisation practice rather than advice specifically for uncertainty visualisation.

Common sentiments such as avoiding chart junk, considering a table in place of a visualisation, use of narrative labels, managing emotion, less is more approach, etc, are the main concepts expressed. The only uncertainty specific advice given is that the graphic should allow “part-to-whole” comparison on an appropriate scale, which seems to be an extension of the framing concept discussed when speaking about uncertainty communication more broadly.

Griethe and Schumann (2006) provided a summary of the uncertainty visualisation literature and did not take any particular angle in the analysis. She makes the point that there is no universally accepted definition of uncertainty and no universally accepted way to display it. She defined uncertainty as “the degree to which the lack of knowledge about the amount of error is responsible for hesitancy in accepting results and observations without caution”. She then explains that there are two ways to visualise uncertainty, implicitly and explicitly. When a visualisation is implicit, the uncertainty is considered an uninteresting feature of the data, and instead of mapping some quantification of uncertainty to a visual feature, data is filtered out if it does not meet a specified “certainty” threshold. These statements clearly points towards a signal/noise dichotomy with normal visualisation and uncertainty visualisation. Additionally she notes that the motivation of the analysis is important, by stating “one has to answer the questions which kinds of uncertainty are relevant from the data’s perspective, which kinds are helpful from the perspective of the visualization steps and which kinds are practically extractable from the perspective of the available methods?”, and “The decision of which of these techniques to choose strongly depends on the intended goal” when commenting on which uncertainty visualisation techniques should be used. She clearly sees uncertainty as additional information or data, and therefore sees the uncertainty visualisation problem as a specific case of the high dimensional visualisation issues. She mentions that there are detailed classifications of uncertainty but the literature and subsequent techniques are not balanced, and instead concentrate on scalar values. In facing this lack of balance she creates a new classification system for uncertainty visualisation. She commented that uncertainty appears at every step in the analysis pipeline, and commented on some the mathematical definitions that seek to quantify uncertainty, but did not explain how these mathematical definitions can be combined across levels. Unfortunately this quantification is needed since she mentioned that in order to visualise uncertainty, it needs to be measurable.

These literature reviews highlight a common theme, the field of uncertainty visualisation is near impossible to navigate. There is so much noise in the experimental results that general patterns are difficult to identify and extract for a common purpose. This noise is attributed to several things, the results dependence on tasks, the difficulty in understanding the specifics of uncertainty and the conflation of signal and noise from many steps in the experimental design. Additionally these reviews often try to look at uncertainty visualisation from a specific perspective (e.g. geospatial or bayesian modelling) and find that even within these sub-fields the data is too noisy to come to any meaningful conclusions.

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, 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 to find some signal in the noise of the field. Ultimately, this paper should serve as a guide for future uncertainty visualisation authors to prevent their work from becoming noise in what is already an incredibly noisy field.

There are several frameworks that the uncertainty visualisation problem is approached from. One approaches it from the concept of risk, and tries to answer how to best express uncertainty in a visualisation such that the viewer can accurately extract and interpret the risks involved. Another approach is to consider uncertainty visualisation as a specific sub-field of high-dimensional data visualisation, and therefore uncertainty is just another variable that needs to be expressed.

3 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 find the best visualisation or expression of “uncertainty” but most don’t even seem to agree on what uncertainty is. In this section we will establish exactly what we mean when we talk about uncertainty.

3.1 Colloquial Definitions

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

will be accompanied by a highly specific quantification of uncertainty as a probability density distribution of a forecast when providing the visualisation. Ironically it seems that many of the papers that refer to themselves as “uncertainty” papers are just suffering from a form of uncertainty, a lack of precision.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 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.

3.2 Taxonomy Definitions

3.2.1 Taxonomies of Uncertainty

Edit Note: A lot of this is going to be moved to the uncertainty definition specific section of the vis lit review and the point here should just be “this is a mess” instead of trying to find any reason in it. My opinion and goal for this section has changed a lot since writing this (specifically deciding that the encapsulating taxonomy doesn't fix the problem and a maths definition is needed) so it is one of the sections that needs big rewrites.

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.

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

²I 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.

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

Examples of Taxonomies used to untangle the features of uncertainty:

- aleatory uncertainty (unavoidable randomness), epistemic uncertainty (uncertainty about structure, still quantified), ontological uncertainty (uncertain about entire modelling process). Spiegelhalter (2017)
- measurement (uncertainty on a single variable), spatial (uncertainty in spatial data), temporal (uncertainty in a time data)
- aleatory (inherent randomness in a forecast) vs epistemic (uncertainty in structure and parameters of statistical models) vs ontological (uncertainty about the entire modelling process as a description of reality)
- scientific methods (variance, distribution) vs human judgments (disagreement among parties)
- Random vs systematic uncertainty
- statistical (can provide a continuous pdf) vs bounded (know values will fall in a certain range) uncertainty
- accuracy vs precision
- Uncertainty from discovery, assessment and execution
- risk vs uncertainty
- “known knowns, known unknowns, and unknown unknowns”, (risk, uncertainty, ignorance)
- Source of uncertainty (model structure, parameters, forcing functions, initial state, model operation) and diagnostic (past/current) vs prognostic (forecasting)
- uncertainty is organised according to Sartre (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)

Edit note: I don't think I will keep this figure in the paper. The goal was to show they are inconsistent but I think a sentence or two in the text will achieve that.



Figure 1. Improved classification showing layers.

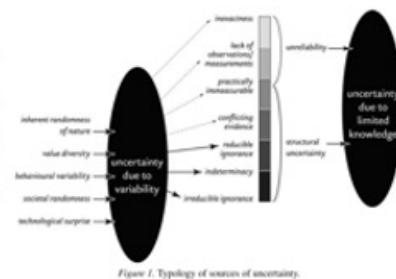


Table 1. Checklist on the types of uncertainty

Ranking of salience	Level of severity
1. Legal	High
2. Moral	Medium
3. Societal	Medium
4. Institutional	Low
5. Proprietary	Medium
6. Situational	Low

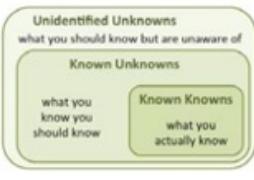


TABLE 1 | Typology of source and model uncertainty.

Model uncertainty (Between subjects)		Source of uncertainty (Randomly ordered within subjects)			
Structural uncertainty	Judgmental uncertainty	Conflict between two precise forecasts	Two identical imprecise forecasts	Hybrid: Two conflicting imprecise forecasts	
One model	One set of initial conditions Two sets of initial conditions	A1 B1 C1 D1	A2 B2 C2 D2	A3 B3 C3 D3	4
Two Models	One set of initial conditions Two sets of initial conditions				

Table 2
The subdivision and coding of uncertainty-categories, along the ‘axes’ of space-time variability and measurement scale (Brown et al., 2005)

Space-time variability	Measurement scale			
	Continuous numerical	Discrete numerical	Categorical	Narrative
Constant in space and time	A1	A2	A3	
Varies in time, not in space	B1	B2	B3	
Varies in space, not in time	C1	C2	C3	
Varies in time and space	D1	D2	D3	

J.C. Refsgaard et al. / Environmental Modelling & Software 22 (2007) 1543–1556

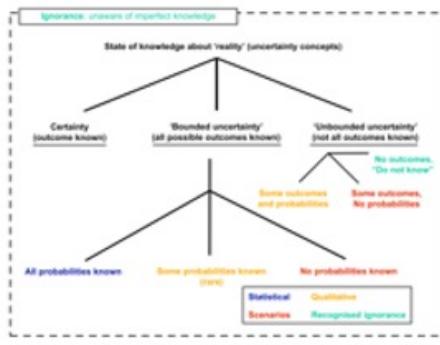


Fig. 3. Taxonomy of imperfect knowledge resulting in different uncertainty situations (Brown, 2004).

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

This does not mean there is nothing to be learned from these taxonomies, each in their own way show what aspect of uncertainty is important to certain cases. There are a handful of overlapping features that indicate which elements of uncertainty are of interest to statistics

and related applied fields. Walker et al. (2003) identified these similarities and used them to design a taxonomy that encapsulates a large proportion of the ad hoc definitions that currently overrun the literature, and is the definition of uncertainty that we will work with for the remainder of this paper.

Figure 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.

Edit note: Everything from this point down was used to try and provide some justification for why this dimensionality concept could work, but I have since changed my mind and am probably going to get rid of this section unless it fits somewhere else.

The *taxonomy* we will use in this paper may seem complicated and difficult to grasp at first glance, but it essentially boils down to three dimensions to consider. Uncertainty is a broad and complicated topic and while oversimplified taxonomies might help us hold the concept of uncertainty in our head, they do little to help with the practical question of how to quantify, visualise, or communicate uncertainty. Organising uncertainty according to its location, level and nature makes allows us to consider uncertainty in the 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 Schiwe 2014; Hullman 2016; Refsgaard et al. 2007). Figure 5 shows the connection between the typical

LOCATION

MODEL OUTCOME UNCERTAINTY

The accumulated uncertainty caused by all below locations

CONTEXT

Specifications of the problem & scope. This is where we identify of the boundary of the model

INPUTS

- Uncertainty about external driving forces, magnitude of their effects & system response
- Uncertainty of system data i.e. lacking complete information of all relevant variables

MODEL UNCERTAINTY

- Model structure uncertainty is about selected model being "correct"
- Model technical uncertainty is generated by software/hardware errors.

PARAMETERS

- Uncertainty associated with estimated parameters of a model
- Uncertainty associated with natural error we are unable to estimate

LEVEL

DETERMINISTIC
know everything precisely

MORE INFORMATION
↑
↓
LESS INFORMATION

TOTAL IGNORANCE

We don't even know what we don't know

STATISTICAL UNCERTAINTY

- Quantifiable & measurable.
- Deviations that can be characterised statistically
- Expressed as a continuum of outcomes.

SCENARIO UNCERTAINTY

- Cannot be quantified but can be described qualitatively
- Cannot give probability of a particular outcome occurring.

RECOGNISED IGNORANCE

- Aware it exists but cannot represent it in any meaningful manner

NATURE

EPISTEMIC

Uncertainty due to imperfect knowledge that can be reduced
Allows movement along the level

ALEATORY

Uncertainty due to inherent variability
Cannot move along the green line.

Figure 3: I need to fix the taxonomy illustration because inputs and model uncertainty need to be swapped. 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.

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

The level axis forces us to identify how quantifiable the uncertainty is. This element is important because establishing how quantifiable uncertainty is informs us how it can be communicated, something that Spiegelhalter (2017) also identified in the form of “precision”. Statistical uncertainty can be quantified, so it can be expressed using a PDF, a variance, an error estimate, or something similar while outcome uncertainty cant be quantified but each scenario may be simulated. Acknowledged ignorance can only be stated in the form of an assumption. Visualisations of climate change scenario uncertainty typically combine the outcome uncertainty of carbon emissions with the 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. (*find the citation for this*) 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) (*I have other papers*) considered the method by which uncertainty was quantified (e.g. by a single value, a function, simulated

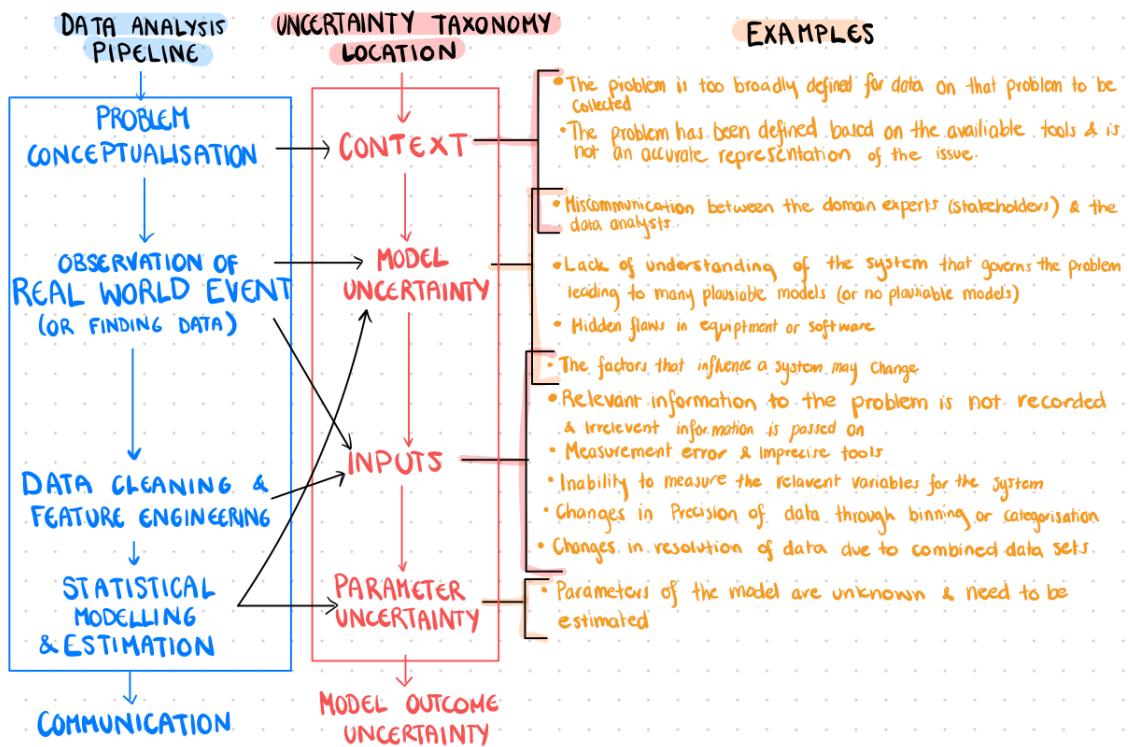


Figure 4: Illustration of the connection between the data generating process and the location of the uncertainty analysis. Several examples are collected from utypo and Münzner2009.

outcomes, etc), which can be thought of as a more precise categorisation of the level axis. This taxonomy also ignores contextual information about the data, such as spatial or temporal considerations. This flexibility means that the use of this framework is not limited by a specific case and should be used going forward to define what aspect of the uncertainty space a contribution exists in.

It is important to note that this taxonomy is thorough but it is not a mathematical definition of uncertainty. While individual *expressions of uncertainty*, such as a confidence interval or a PDF are mathematically defined, the term *uncertainty* has not been defined mathematically. This is largely because it is a layperson word that has been adopted by statistics to describe any deviation from complete determinism, however this does not mean that a mathematical definition that encompasses missing values, alternative models, and prediction intervals *cannot* be found. However, that task is far beyond the scope of this work. In this work, we will investigate the array of misunderstandings and focus on inspecting how the current literature *deviates* from already established mathematical definitions, why these deviations occur, and understand how these deviations can lead to larger problems in uncertainty communication.

3.2.2 Taxonomies of Visual Uncertainty

Edit Notes: Similar to the “similar works” section, I think this section would be better off with a short sentence about each paper and comments on their similarities because a paragraph for each paper is too much

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

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

Kristin Potter, Rosen, and Johnson (2012) organised several existing uncertainty visualisations into groups based on the dimensionality of the data (1D, 2D, 3D, and No Dimension) and the

dimensionality of the uncertainty (Scalar, Vector, Tensor). However, because the term “PDF”, a statistical object that describes a random variable that is typically a one dimensional function, is used to describe both the data and the uncertainty for all dimensions, it is hard to understand how this dimensionality should work.

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

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

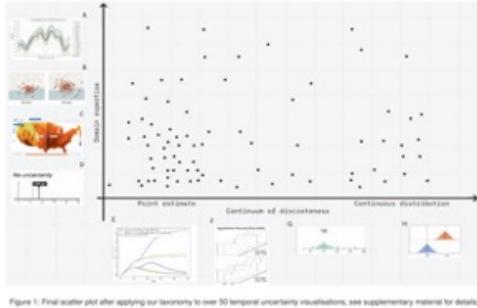


Figure 1: Final scatter plot after applying our taxonomy to over 50 temporal uncertainty visualisations, see supplementary material for details.

Data Dim.	Uncertainty Dimensionality								
	Scalar			Vector			Tensor		
1D	[62]	[77]	[85]	[82]					
2D	[7]	[13]	[14]	[22]	[27]	[8]	[9]	[33]	[53]
	[30]	[31]	[33]	[33]	[45]	[67]	[66]	[92]	[97]
	[43]	[49]	[53]	[51]	[56]				
	[60]	[64]	[69]	[72]	[78]				
	[79]	[83]	[83]	[91]					
	[95]	[10]	[17]	[28]	[82]				
3D	[12]	[20]	[19]	[18]	[12]	[5]	[50]	[53]	[52]
	[46]	[47]	[50]	[54]	[55]	[92]			
	[50]	[58]	[71]	[72]	[73]				
	[75]	[80]	[81]	[86]	[87]				
	[93]	[96]	[10]	[82]	[61]				
ND	[2]	[23]	[26]	[32]	[90]				

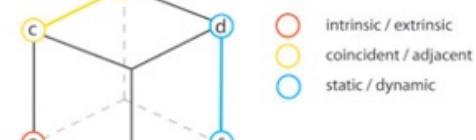


Figure 1. UVIS³ ('Uncertainty Visualisation cube') for categorisation of uncertainty significance in visualisations

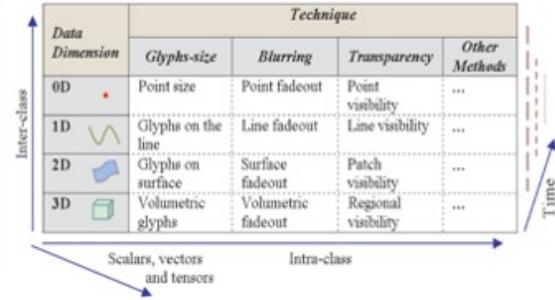


Figure 5: Example of different uncertainty visualisation taxonomies

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 “uncer-

tainty 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 visualisation taxonomy when they claim “future typologies should take different categories of tasks into account (1) communication tasks (2) analytical tasks (3) exploratory task”, a common typology for information visualisation in general. The process of understanding and estimating uncertainty requires knowledge of the data, the statistical methods used to make an estimate, and the assumptions of a model. Visualising the statistics that represent uncertainty should be no different than depicting the statistics that represent any other element of a graphic, and therefore there seems to be no reason to have a separate taxonomy for uncertainty visualisation.

3.3 Contextual Information

Edit Note: This section should probably be moved to the visualisation specific section, although that depends on the tone that I end up with after the rewrites of the earlier sections.

Some uncertainty work, specifically uncertainty visualisation, focus on the context surrounding the uncertainty. These are things like spatial data, temporal data, category data, etc. While other discussions of uncertainty may consider this distinction more relevant to their work, here we will consider this contextual information. Contextual information is important for interpretation and understanding of graphics, but it does not in of itself 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, inherent randomness, and assumptions that create bias 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.

Papers that stress the importance of spatial or temporal contexts may be conflating contextual information with uncertainty. Figure 7 depicts the role of contextual information in graphics, as something that helps us understand the information in the statistical graphic. 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. Kinkeldey, MacEachren, and Schiwe (2014) discussed how uncertainty can be represented by three components attribute (what) position (where) and temporal (when) and that studied typically deal with uncertainty around attribute but rarely position and time, however it is never specified what considerations should cause attribute uncertainty to be different to position or temporal uncertainty. Kay et al. (2016) did an experiment that showed uncertainty around bus arrival times, however the visualisation used in the experiment, shown in Figure 6, is indistinguishable from most work that would be considered “attribute” 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 literature, and in later sections we will discuss how this concept is related to uncertainty visualisation.

This does not mean we are saying spatial and temporal information is never of importance. Spatial information is *highly relevant* when we translate from the abstract domain back into real world for communication. This position is best explained by Wilkinson (2005) in *The Grammar of Graphics*:

“Geography is anchored in real space-time and statistics in abstract dimensions.
This is a distinction along a continuum rather than a sharp break... but this difference in focus clearly means that a system optimized to handle geography will not be graceful when dealing with statistical graphics.”

This paper focuses on uncertainty, specifically uncertainty that is still in the the abstract *statistics* domain. Since we are still in the abstract domain, things that are anchored in real space-time create confusion that is evident in the literature. Spatial and temporal uncertainty papers limit how they communicate their uncertainty because they are considering what can



Figure 6: The uncertainty visualisation used by Kay et al. (2016) to show the uncertainty around a bus arrival time prediction.

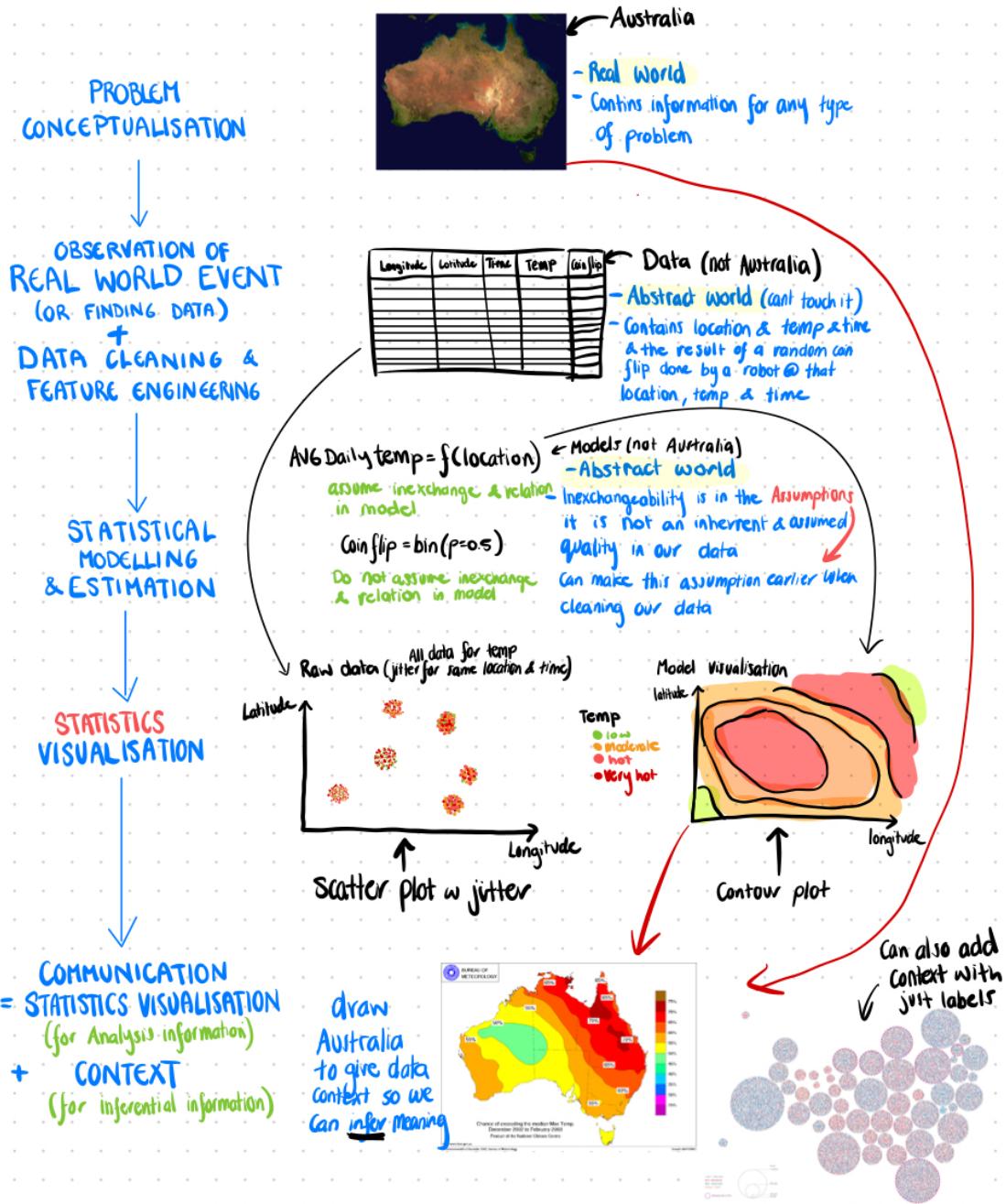
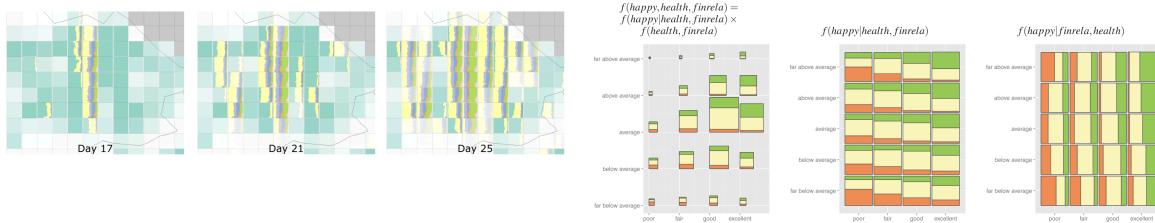


Figure 7: Data pipeline with visualisations to show why we are ignoring contextual information for the time being. The visualisations that depict the data are similar to typical statistical graphics. The purpose of the spatial information is to provide context to the information we extract from the statistical aspect of the graphic.

be visualised inside the context provided by maps and, not what they should *try* to visualise based on the uncertainty abstraction. Additionally, since uncertainty is defined in the abstract domain, of the issues faced by spatial uncertainty visualisations are likely shared by other graphics that only differ in contextual information.

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 conclusions that are not necessarily specific to spatial or temporal data, and it may lead us to miss relevant information that is outside our domain. This is not only the case for uncertainty data. Slingsby, Reeve, and Harris (2023) discussed using a gridded glyphmap for spatial modelling, however the statistical elements are identical to a fluxuation diagram depicting $f(\text{infection}|\text{time}, \text{age}, \text{longitude}, \text{latitude})$ presented in Wickham and Hofmann (2011). The plot by Slingsby, Reeve, and Harris (2023) is simply a special case of the graphic by Wickham and Hofmann (2011), where a separate visualisation is made for each time point, the continuous function $\text{infection}|\text{age}$ has replaced the categorical variable *happy*, a fade is added for the population size, and a map is shown in the background. Slingsby, Reeve, and Harris (2023) presents other maps in the paper, however they are also special cases of the product plots framework. This is not to say this extension is uninteresting, on the contrary we believe identifying cases where particular graphics are useful is a worthwhile endeavour, but these plots seem to be have been developed unaware of the existing statistical framework. When we separate the field of visualisation along lines that may not be of particular relevance, we increase the likelihood of “reinventing the wheel” in each sub-dicipline.



4 General Attitude Towards Uncertainty (Uncertainty Communication Broadly)

Edit note: this section needs a huge rewrite. The idea of this section blends way too much with the measurement discussions in the “noise and signal” section. I have already taken paragraphs from here and just pasted it in the noise section so it might not be coherent. Additionally this section goes on SEVERAL long rambles that realistically should have been a sentence or not included at all.

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 of 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 ambiguity) 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 detailed 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 and 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.

Edit note: Currently these notes are just from spiegelhater because this is the only paper I have moved notes over from. Other sources will be used here lol. Working out what exactly they are communicating is a bit of a pain though.

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 information 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 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 is 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 uncertainty 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 ambiguity-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 literature, 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 bizarre 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*. There have been other discussions on the appropriateness of discussing risk-aversion as a bias (Vranas 2000), but few have made the point that if uncertainty holds some *value* there is not technically a “right” decision at all. 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.

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 visualisation, specifically with respect to the “signal suppression” 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 fascinating challenges that are faced by viewing uncertainty visualisation through the lens of “signal suppression” 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.

5 Uncertainty Visualisation Literature Summary

The previous sections of this paper established several issues in uncertainty communication at large. In this section, we will discuss the current literature on uncertainty visualisation, how the uncertainty communication issues more broadly manifest in the specific case of visualisation, and what general lessons we can take from the existing work.

5.1 Types of Uncertainty Visualisation Papers

Edit notes: Originally I was planning to have the discussion following this section relate to these “types” of papers, but that didn’t end up happening. Depending on what I do with later sections, this might need a rewrite or to just be removed, since I’m not sure how relevant this is to the main ideas of the paper.

Uncertainty visualisation papers can be organised according to the *goal* of the experiment. Evaluation experiments are the standard rule for visualisations because the human brain is not as reliable as mathematical calculation. Therefore, user studies often aim to assess the limitations, biases, and heuristics of our mental calculator so that we can better understand the problems we may encounter when we plot our data. This is not to say any paper that suggests a visualisation without an evaluation experiment is completely lacking in justification, and there are many papers that suggest a novel visualisation without an evaluation study. Sometimes these papers are a preliminary step in finding a solution for common problems and intend to evaluate the visualisation in later work. Until that later work is done, it is often difficult to accept one particular representation. While justification could come in the form of discussing established concepts in visualisation, such as the hierarchy of perceptual tasks, even these may need an evaluation study to back up their claims.

The reasoning for this is obvious. There are a large number of heuristics and biases that are not obvious to us when designing visualisation. Additionally, these heuristics and biases can change depending on the larger scope of the graphic and the population we are communicating with. Additionally, since there is often a myriad of ways to visualise any particular mathematical object, to adopt specific visualisation that is not already common practice needs reasonable justification, lest we run into a range of unforeseen heuristic pitfalls.

The remainder of section 4 will focus on the results and methods of uncertainty visualisation evaluation experiments. There are six main types of evaluation experiments in the uncertainty visualisation literature, they are:

- Association perceptual tasks: These experiments identify perceptual tasks that are/could be

associated with uncertainty, and test its association with uncertainty, check it's number of distinguishable levels, and/or check how quickly/accurately people can draw the uncertainty information from the graphic.

- Value perceptual tasks: Participants are shown some depiction of uncertainty and are asked to extract a value, make a comparison, or do something simple to show that the graphic is readable.
- Heuristic Checks: These experiments are checking if a perceptual bias that has been identified through public use of a particular plot, other evaluation experiments, or gut feeling identified through public use of a plot or in a different.
- Plot Comparisons: These papers describe an uncertainty visualisation evaluation experiment, where they compare two or more uncertainty visualisations on a specific set of questions. These questions can be about specific statistics (e.g. What is the probability that A>B?) or about a secondary goal such as trust or decision making (e.g. How much do you trust this estimate?).
- Qualitative analysis: These experiments show participants an uncertainty visualisation for an open ended or specific task, and asks them to describe how the visualisation was used to come to a conclusion, or they are asked to describe what they see in the visualisation.
- Method papers: These papers discuss issues with the evaluation methods of uncertainty visualisation and suggest changes or improvements, typically with results indicating the benefit of the change.

These experiments are not mutually exclusive within a paper, and a single study will often contain several experiments, usually from several of these categories. These distinctions are important because each have their own issues in how they relate to uncertainty more broadly. Below we will go through the five key issues in uncertainty visualisation, describe how they manifest in the uncertainty visualisation literature, and offer some tentative solutions to these problems.

We will also comment on some of the uncertainty visualisation papers that suggest a visualisation without an evaluation experiment. Some of these plots may offer a solution to the unique problems we mention in this paper and these visualisations will be discussed in “Part 6: Great Examples”. We will also consider these papers to illustrate the issues that visualisations are aware of and working to fix, and discuss whether or not these directions of research are likely to bear fruit, and why.

5.2 Defining uncertainty with respect to inference

When we consider mapping uncertainty, a reasonable question that should come to mind is “uncertainty about what?”. Every single thing concept or activity in statistics does not have uncertainty. For example, descriptive statistics cannot have uncertainty as they are just describing statistics as they are, the sample mean is the sample mean, we are not uncertain about its value. A sample variance describes dispersion, which is also how we describe uncertainty, but the sample variance in of itself does not have any uncertainty. Most of the mathematical uncertainty concepts have this idea embedded in them. For example, a confidence interval

is an interval drawn around the sample mean that will theoretically contain the *population mean* 95% of the time. This issue has been noticed several times, however it is often brought up as a *task* dependence rather than an understanding that uncertainty is inherently linked to inference on a particular statistic. Wallsten et al. (1997) argue that the best method for evaluating or combining subjective probabilities depends on the uncertainty the decision maker wants to represent and why it matters. Munzner (2009) created a nested model for visualisation that highlighted how the first mistake that can be made in a visualisation is in the problem characterisation, and failing to do it well can cause downstream effects and damage the effectiveness of a visualisation. Fischhoff and Davis (2014) looks at uncertainty visualisation for decision making decides that we should have different ways of communicating uncertainty based off what the user is supposed to do with it. Health have a wide range of metrics used to communicate different risks (@ Spiegelhalter 2017) indicating they utilise new statistics based off the specific aspect of risk that is relevant to the issue. This should make it clear that uncertainty only exists when we are performing inference. Therefore it only makes sense to describe uncertainty in terms of something to be uncertain *about*, it is not some a latent feature of statistics and data.

The idea that uncertainty can only be defined in the presence of a motivating question is well grounded in most areas of statistics. The entire process of data analysis, from deciding what should be observed as data through to communicating that data in a plot is governed by human decision and the goal of an analysis. At the philosophical level, applied statistics is simply taking real world entities and boiling them down into probabilistic objects, an ontological process that is largely dependent on our goals (Otsuka 2023). When we move onto data provenance the issue persists, as what is kept as data and what is tossed away is determined by the motivation of an analysis and what was previously noise can be shown to become signal depending on the resolution of the question (Meng 2014). After moving onto modelling this issue continues as each research question can be categorised as descriptive, predictive, or causal, each of which has its own appropriate statistical methods and motivation agnostic model selection leads to signal devoid of meaning (Carlin and Moreno-Betancur 2023). Even at the final stages of visualisation a lack of understanding of the motivating question make it difficult to untangle what is signal and what is noise, leaving many uncertainty visualisation studies with conflicting results (Kinkeldey, MacEachren, and Schiewe 2014). These cases highlight that uncertainty is defined at *every* stage in relation to our motivating question, from data collection to visualisation. Discussions of uncertainty cannot be had if we are not clear *what* we are uncertain about. Once it is established what we are uncertainty about, we can consider the other elements of uncertainty that need to be defined.

We feel the need to make this point clear because a *large* amount of uncertainty research seems to be confused by the purpose of uncertainty visualisation, what it is related to, and why we visualise it. Often, the authors are even more confused about what they are doing than the participants. There are enough papers that fail to understand these basic principals of uncertainty that this entire paper could simply be a list with explanations. Instead, we have a handful of highlights presented below.

- Hofman, Goldstein, and Hullman (2020) asked questions about discusses the difference in

showing inferential or outcome (prediction) intervals and then found that participants shown the inferential distribution performed poorly when asked questions about predictions. They found that showing the relevant distribution to the question reduced the error. This result should be astoundingly obvious if it was well understood that uncertainty is not some latent variable and is only relevant for specific ground truth statistics. This paper's *perceived* importants highlights an underlying belief that uncertainty is linked to a specific variable or data set, not the variable we are drawing inference on.

- L. M. K. Padilla et al. (2021) found that high uncertainty in the model estimates (calculated uncertainty) and low forecaster confidence (which is typically an expression of suspected bias in the model) both caused participants to have decreased confidence in their results and suggested modelers express both if they are relevant. Kale, Kay, and Hullman (2019) discussed the importance of communicating decisions made in the data analysis pipeline and being aware of the alternatives. While there is nothing wrong with the core ideas of these papers, the fact that they were published indicates that the general population was not already aware that choices introduced early in the data analysis pipeline create bias and therefore uncertainty in our final values. This indicates a general failure to see bias and variance both as sources of uncertainty.
- Wickham and Hofmann (2011) suggests their product plot framework, which includes histograms, should have a way to measure uncertainty, but does not consider that a histogram is *already* a depiction of mass and would already be considered an uncertainty visualisation were our statistic of interest the population mean. To suggest that a histogram needs uncertainty bounds is similar to suggesting a density or box plot need uncertainty bounds.
- Griethe and Schumann (2006) commented that "if visualization is used as a means to explore a data volume or to communicate its contents the uncertainty has to be included". K. Potter et al. (2010) aimed to create a summary plot that "concisely presented data with uncertainty information". These papers fail to realise the exploration step of an analysis, which includes descriptive statistics, exploratory data visualisation, and unsupervised machine learning techniques, is performed without a prior hypothesis. Therefore, it is performed to *suggest* a direction for future inference work or communication, which *do* contain uncertainty and does not contain uncertainty within itself. This confusion would not occur if it was understood that uncertainty does not exist without inference.
- Ibrekk and Morgan (1987) asked participants for the "best" estimate while displaying a population density function. The participants provided the mode instead of the mean, although there was no indication the mean was what was being asked about.
- Boukhelifa et al. (2012) tried to quantify the strength of the intuitive connection between "sketchiness" and uncertainty. Since they had seen literature that explained uncertainty was task specific, they tested the connection for multiple "tasks", what they failed to understand, is that the "task" dependency which is likely just a proxy of the evaluation experiment problem discussed in the previous section. This misunderstanding led to "sketchiness" being used to depict uncertainty in the categories of bar charts, in graphic networks, and in train lines, with little consideration or explanation as to *how* these things would be uncertainty. I imagine it is quite hard to be "uncertain" about the existence of a train line. Participants rightfully assumed the sketchiness therefore represented something else, such alternative options or simply

ignored it.

The reality is that a lot of uncertainty visualisation authors do not seem to have an intuitive understanding of uncertainties connection to inference, when inference is being performed, and how to design experiments that capture this relationship.

This misunderstanding is not due to laziness or the fault of the authors, but rather is likely caused by the absence of a strict definition of uncertainty. The imprecise and confusing set of existing definitions are creating a field in which the authors themselves do not know what they are testing. Uncertainty examples include imputed data, model selection, inherent randomness, biased sampling, etc, not because these things *are* uncertainty, but because they *create* uncertainty *when we perform inference*. These things are not “uncertainty” in of themselves, but rather contribute to the distance between the final estimate and the statistic we want to perform inference on. The field of uncertainty communication is in desperate need of a unifying mathematical definition that firmly identifies exactly how each of these things contribute to uncertainty. This may be a considerable task, but as of right now, the list of things that are “uncertainty” continues to grow, and the best methods of understanding and quantifying them are ad-hoc at best. Thomson et al. (2005) suggests a mathematical formula for *examples* of uncertainty, Meng (2014) mathematically defined the variance introduced to a model by the array of model choices, information theory tries to quantify uncertainty using the idea of entropy. None of these existing methods are thorough enough for an analyst to understand what causes uncertainty, and quantify it for communication. Realistically, this problem is well known, and put simply by Freeman Dyson in his famous Birds and Frogs speech:

Rigorous theorems are the best way to give a subject intellectual depth and precision. Until you can prove rigorous theorems, you do not fully understand the meaning of your concepts. - Freeman Dyson (Birds and Frogs speech)

This limitation becomes apparent when you look at the limited scope of what the uncertainty visualisation literature has *actually* focused on visualising. Almost every uncertainty visualisation paper contains a broad and encapsulating definition of uncertainty in the introduction, followed by a visualisation of a normally distributed probability density function.

Edit notes: Need some way to show uncertainty visualisation papers provide sweeping definitions and then visualise a simple PDF everytime.

In order to visualise something, it needs to be quantifiable. In order to quantify something, it needs to be mathematically defined. The broad classes of what is considered “uncertainty” is not currently quantifiable in any way that is not ad-hoc. The only aspects of uncertainty that are currently quantifiable are confidence intervals, prediction intervals, and related terms. Even then, these quantifications often only capture the uncertainty in our model or its assumptions, it ignores bias often introduced in earlier stages of the analysis. Broader and more complicated concepts such as the effects of assumptions, imputed missing variables, and model choices remain difficult to meaningfully quantify beyond ad-hoc methods. This means that the current literature of uncertainty visualisation papers are doing one of two things:

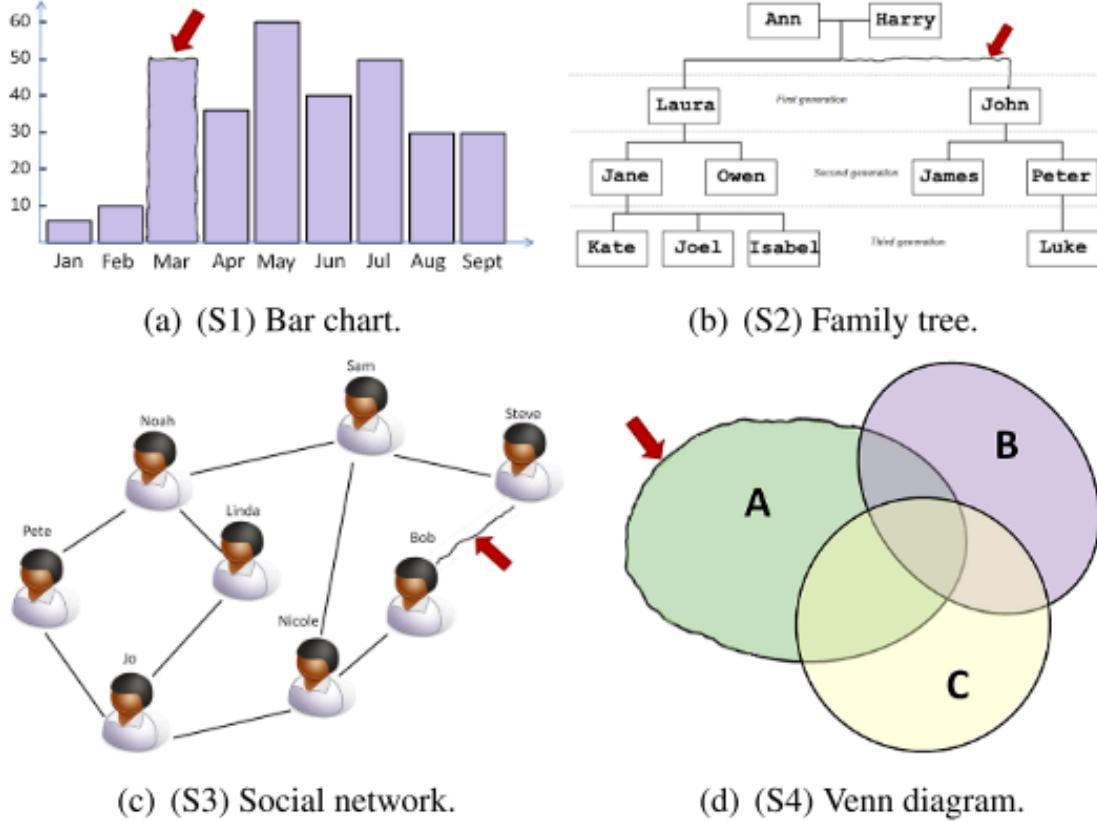


Fig. 6. The four abstract scenarios used in the study.

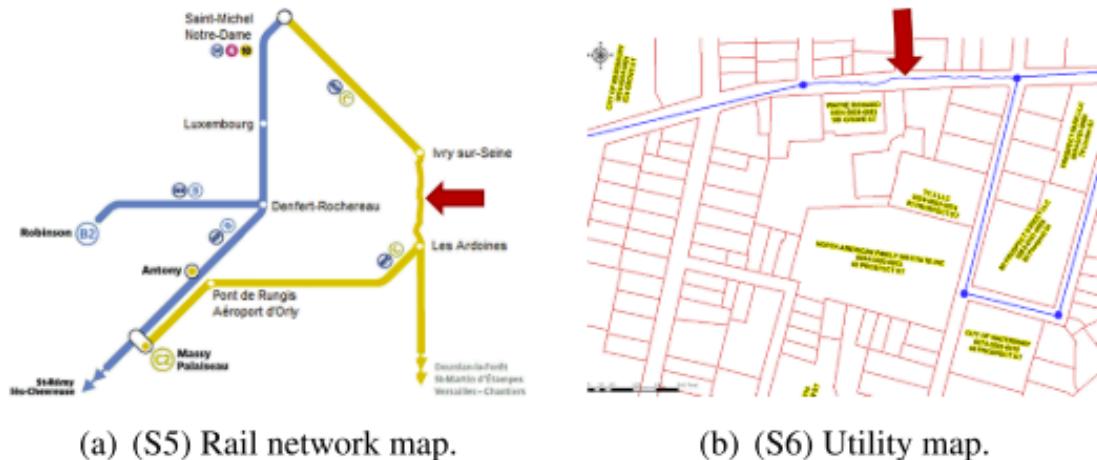


Fig. 7. The two non-abstract scenarios used in the study.

Figure 8: The graphics displayed by Boukhelifa et al. (2012) to identify if there is an intuitive connection between sketchiness and uncertainty.

- 1) Evaluating multiple visual representations of a single mathematical expression of uncertainty (e.g. a PDF, confidence interval or error). These papers are purposefully vague, and would be clearer in their intention if the title was aligned with the actual mathematical object they are visualising. For example, a paper that is titled “evaluation of uncertainty visualisations” that only considers PDF visualisation should realistically be titled “evaluation of PDF visualisations”.
- 2) Evaluating multiple visual representations of *multiple* different mathematical expression of uncertainty. Since these experiments visualise different mathematical objects, some of which are closer to the ground truth statistic, the results of these papers can be easily predicted with an understanding of how to get the ground truth statistic from a plot and the perceptual tasks required to do so.

It is clear that the vague definition is causing across the board misunderstandings of uncertainty, confusing experiment methodologies and motivations, and a literature that is difficult to navigate. This is also causing a large body of experiments that have easily predictable results that can be easily explained by the existing visualisation. This problem will be the focus of the following section.

5.3 Uncertainty evaluation experiments repeat visualisation studies

The first issue in uncertainty visualisation is there does not seem to be a clear rule for what makes a particular visualisation a depiction of uncertainty. This problem seems to start at the lack of the mathematical definition of uncertainty which then bleeds into a lack of a definition for an “uncertainty visualisation”. For example Wilkinson (2005) mentions that popular graphics, such as pie charts and bar charts omit uncertainty, however at least one or both of these charts are used in a large number of uncertainty visualisation experiments. How can a chart omit uncertainty according to the grammar of graphics, but be used to describe it in several evaluation studies? The answer to this question is related to the definition problem discussed earlier, but to understand how, we will need to go through the grammar of graphics, and its absence in the uncertainty visualisation literature.

Edit notes: The “in order to visualise uncertainty you need to quantify it and the quantification does not align with the definition” point could also work here since this section is now about value extraction.

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 generates an estimated density function, which is then depicted on a plot using a line, which is then converted into information in your head using perceptual tasks and

heuristics. 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 is. Computer science considered two pieces of information to be equivalent if they come from the same data. This definition of information is much looser and is contained within the mathematical definition of information. Thankfully, we can turn to the grammar of graphics to try and organise and untangle the information in our graphics.

When creating a graphic there are several tasks that must be completed in a specific order, regardless of whether or not we are working with a uncertainty visualisation or not. Wilkinson (2005) depicts these steps as a pipeline, shown in Figure 9. This pipeline shows every step from data observation to rendering of a graphic, but we are going to focus on the tail end of the pipeline specifically on statistics, geometry, coordinates, and aesthetics.

Edit notes: Grammar of graphics is just a placeholder, Di suggested to replace this with my own pipeline visualisation based on some data wrangling papers.

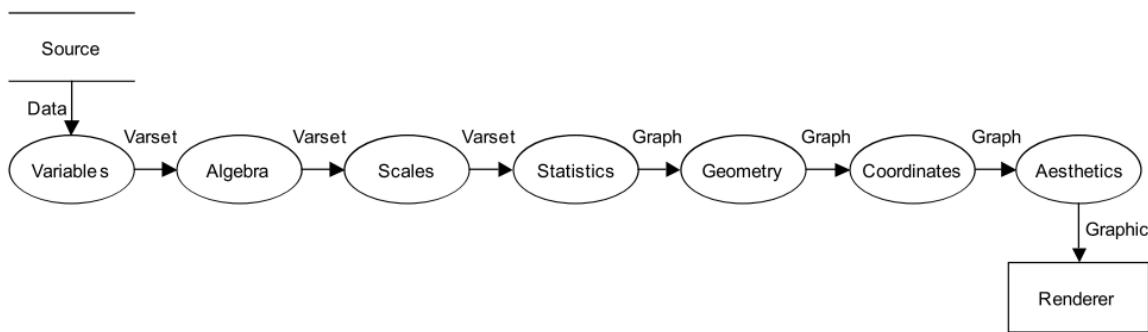


Figure 2.2 From data to graphic

Figure 9: The grammar of graphics data analysis pipeline

In order to visually communicate uncertainty we must:

- 1) Mathematically quantify uncertainty in some way.
- 2) Visually express this quantification using some geometry, axis, aesthetic, and rendering.
- 3) Interpret the visual expression to extract the information using perceptual tasks.

Each of these steps have a large amount of research independent of the “uncertainty visualisation” subfield. There are numerous ways to depict uncertainty mathematically. Examples of

mathematically defined objects that are used to express uncertainty include: confidence intervals, sampling distributions, bias, variance, prediction error, value ranges, quantiles etc. These definitions all exist in mathematics independent of the field of visualisation. Once we have some quantification of uncertainty, we need to express it visually. There are numerous options in combinations of geometry, axis, and aesthetics and modern graphical software will allow most of these combinations. There is also a large amount of research in information visualisation and psychology into how the choice of visual expression impacts a person's ability or ease with which they extract relevant information. In this sense, once we have a mathematical expression of uncertainty, the visualisation of that uncertainty is identical to the visualisation process of any other variable, or at least this is how it is treated in the uncertainty visualisation experiments.

This distinction is important because lower levels of the pipeline influence higher levels. Visualisation authors are almost unanimous in commenting that the information in two plots must be the same in order for the visual techniques to be compared (Cleveland and McGill 1984; Kinkeldey, MacEachren, and Schiewe 2014) *this citation has way more, check confirmation.* Kinkeldey, MacEachren, and Schiewe (2014) adopts a definition from the 80s that suggests two graphics are informationally equivalent if all the information in one plot is inferable from the other and vice versa, and computationally equivalent if that information can be extracted from both plots with similar ease and speed. This definition illustrates a similar idea to the grammar of graphics pipeline, shown in Figure 9. First the mathematical information in a plot is established, and then that mathematical information is visually represented. Two plots that depict mathematical information that cannot be inferred from each other, cannot be compared on their visual elements. To prevent this requirement from being *too* limiting, when using a ground truth statistic we can consider two plots to be informationally equivalent if both visualisations depict a sufficient statistic. For example, a depiction of a sample and its average are equivalent if we are trying to infer the population mean. This does not mean they are computationally equivalent. Additionally, two plots cannot be compared on their visual elements if they contain different information as well. The reason for this is obvious. If you compare two graphics and one depicts the mean and the other depicts a median and you ask viewers to extract the median value, the graphic that *actually* depicts that statistic will perform better. People cannot perceive information that is not depicted.

Computational equivalency in visualisation is also something that has been well researched. For simple tasks such as value extraction, there is a hierarchy to perceptual tasks where extracting visual information in some forms is easier than others. The hierarchy established 40 years ago by Cleveland and McGill (1984) is:

- 1) position along a common scale.
- 2) position along a non-aligned scale.
- 3) length.
- 4) angle/slope.
- 5) area.
- 6) volume.

7) colour

Edit notes: need to add the notes from the susan lit review that has the more recent perceptual task list

This hierarchy has had additions and adjustments over the years. For example, this hierarchy is a general rule, but the hierarchy can change from individual to individual. Additionally, there are other graphical rules to consider such as gestalt heuristics, broader methods of extraction, and attention principles. These established visualisation concepts allow us to anticipate the ease with which certain pieces of information will be extracted from a plot.

A bizarre feature of uncertainty visualisation literature is that it does not work to build upon these existing principle or identify the ways in which uncertainty visualisations may diverge from these rules, but rather it seems to exist with complete indifference to the research in information visualisation at large. The most common uncertainty visualisation evaluation study asks participants to perform some value abstraction either explicitly by asking participants to report values or vaguely by asking participants to find regions of least certainty, rank plots, report their confidence, and/or make decisions according to uncertainty information. Experimenters then compare these results based on the participants accuracy, preference, confidence, and/or speed Hullman (2016). Most commonly, these value abstraction papers compare two different visualisations of uncertainty on a set of these questions, however, since there is no rules for what is or is not an uncertainty visualisation the graphics that are compared often differ in both computation *and* information. Hullman et al. (2018) asks participants questions that would require a prediction interval but provided a sampling distribution and argued that the two could be considered equivalent so long as participants understood the difference between a confidence interval and a prediction interval, knew the sample size, and knew the formula to convert the confidence interval to a prediction interval. The authors themselves note that this is ridiculous, which begs the question. Uncertainty visualisaiton papers only seem to require that the plots depict “uncertainty” which has already been established to be poorly defined. The complete lack of formality is easy to see if we actually plot the visualisations compared across evaluation experiments according to the grammar of graphics *in the figure below:*

Edit note: need to include a graphic that illustrates the wide spread of mathematical information and visual techniques that are visualised in these experiments.

Reading the uncertainty visualisation literature will make the importance of information and computation incredibly salient. If plots that have different information are compared, whichever directly depicts the ground truth statistic (or has information that can easily be converted to that ground truth statistic) will perform the best *cite*. If plots that have the same information are compared, the plot that requires a perceptual task that is higher on the hierarchy will perform the best *cite*. Therefore, when it comes to any task that involves value abstraction, many uncertainty visualisation evaluation studies are simple repeating already established results in visualisation.

Several uncertainty visualisation literature reviews have commented on the immense difficulty of combining results across experiments, but none have been able to pinpoint the reason. Kinkeldey, MacEachren, and Schiewe (2014) suggested a task based taxonomy that organised the plots by the question or purpose of the visualisation, and even suggested that communication tasks (such as value retrieval of uncertainty values) should just follow traditional cartography rules, Spiegelhalter (2017) simply ignored the existence of uncertainty visualisation as a subfield and just suggested we follow established rules within uncertainty visualisation, Hullman (2016) said that commented on the large amount of noise generated from uncertainty visualisation evaluation methods that led to a difficulty in combining results, Griethe and Schumann (2006) did not meaningfully combine the work and instead listed findings in the field, but also did not comment on the lack of generalisable results. While we largely agree with the difficulties of these authors, we do not agree that the solution is a task based taxonomy or better questions. While we agree the lack of attention to tasks and appropriate questions has generated problems in the field, there is a much more pressing issue, which is that uncertainty visualisation authors don't seem to understand *why* uncertainty visualisation is different to typical visualisations.

The class of “uncertainty visualisation” does not seem to have any rules for what belongs in the group and what makes it different to a “typical” visualisation is rarely discussed. Instead one of two general motivations for the field are provided in the subtext.

1) Uncertainty is fundamentally differently to other variables due to psychological heuristics involved in the interpretation of uncertainty (e.g. risk aversion). 2) Uncertainty is not of interest in of itself and is typically layered on at the end of a the visualisation pipeline, so uncertainty visualisation is the field of adding error information into already established graphics.

Neither of these motivations are remotely relevant when we are evaluating the simple task of value extraction. The psychological effects of uncertainty are only relevant for decision making and have nothing to do with value abstraction, and the information overload issue is not specific to uncertainty and is simply an example of a larger dimensionality issue. Therefore, these explanations for the difference fall short of explaining *why* uncertainty should be treated any differently than any other variable.

There is also the common view that uncertainty as a dimensionality problem. This perspective is not quite right and the literature seems to be about 50/50 on whether uncertainty is a psychological or dimensionality problem. Kinkeldey, MacEachren, and Schiewe (2014) also identified this issue when they highlighted that the literature makes it unclear if uncertainty is a variable in of itself, something that should be interpreted with the main variable, or if it is metadata. It is important to keep in mind that just because information is in a graphic, that does not mean it will be “seen”. The phenomena of inattentional blindness shows that there is no perception without attention and it is powerful enough that participants can fail to be aware of random objects appearing on a screen or a gorilla walking through a basketball game (Simons and Chabris 1999; Mack 2003). Therefore if information is mapped to graphical elements that are so low on hierarchy they can be ignored, they might as well not be there at all. The people who don't plot uncertainty because they think its unimportant and the people who relegate uncertainty to the lowest ranking aesthetic on the list are both saying

the same thing, “I think the uncertainty is unimportant”. Including uncertainty is worth very little if no attention is left to see it. This does not mean we cannot put a *large* amount of information in a graphic. Glyph maps can be used to depict the multi-dimensional information in spatial-temporal data by mapping line plots to map locations, but we still need to decide what information is important. The map can either present trends in the global or local variance depending on whether or not the line plot is scaled globally or locally (Wickham et al. 2012), however smaller details are almost impossible to convey. No matter how much information we try and put in a plot, there will always be only a handful of key takeaways.

This issue has come up in many different forms in the uncertainty visualisation literature. Typically as a “task dependence” on the performance of uncertainty visualisations. It is repeatedly identified as a problem in previous reviews of the uncertainty visualisation literature (Kinkeldey, MacEachren, and Schiwe 2014; Hullman 2016) as well as across many sub domains and applications (Wallsten et al. 1997; 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) takes this approach by considering 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. Kinkeldey, MacEachren, and Schiwe (2014) suggested that the nature of the task plays an important role and believed the task dependence for uncertainty visualisation could explain some of the inconsistency in outcomes.

Despite these results, it is clear that authors intuitively *feel* that uncertainty is different even though this difference has not been apparent in existing uncertainty visualisation experiments. Kinkeldey, MacEachren, and Schiwe (2014) also noticed that it is not clear whether or not uncertainty is just another variable, as many value extraction papers treats it as such in their experimental design and uncertainty may need to be in a variable class of its own. Hullman (2016) commented that it is straightforward to show a value but it is much more complex to show uncertainty, but did not explain *why* uncertainty information faces more difficulty. So, if we cannot see any difference to uncertainty information when it comes to value abstraction, *when* is uncertainty information different?

5.4 Measuring uncertainty as noise, not signal

Uncertainty is noise. This may be an obvious statement, but what is not obvious is that the existing literature sees noise as a signal. When we directly ask “What is the variance of this variable” the variance *becomes* the signal. Noise and signal are not an inherent property of variance, or meta-data or the aspects of an analysis that are usually secondary considerations, it is a property of the data that *promotes* a message we are trying to infer and the data that *suppresses* it. The replication problem in uncertainty visualisation, where experiments are simply repeating established visualisation work, comes from the misunderstandings established

in previous sections. Uncertainty is poorly defined, therefore authors do not understand it is not an inherent property of the data and is actually related to a particular inference question, this incorrect view that uncertainty is always variance or probabilities leads to the class of “uncertainty visualisations” that depict different mathematical objects or visual features with little in the sense of consistent rules, this complicated class has caused visualisation authors to miss that they are simply repeating the established evaluation results from normal information visualisations. This situation is a bit of a mess. It also highlights a unique and fascinating problem faced by uncertainty visualisation. If asking direct questions about uncertainty causes us to treat it as a signal, how do we evaluate uncertainty as *noise*?

Uncertainty evaluation studies are not all value extraction. There are cases where authors *know* the uncertainty information is relevant, but are not quite sure how to capture it. These studies most commonly evaluate a visualisation on its ability to convey information for a decision or how well it induces trust. The motivation behind these experiments are obvious, if we cannot measure uncertainty *directly* maybe we can measure its secondary affects. The issue is that in order to say if some plots are *better* or *worse* than others, is to ascribe value to uncertainty information.

Decision making questions that typically involve uncertainty discuss the concept of “risk-aversion” as though it is a mistake. The irony of an uncertainty visualisation pper expressing the importance of uncertainty visualisation, but setting up a ground truth or optimum strategy in which participants *should* ignore the uncertainty information is not lose on us. This perspective does not make sense because it implies participants should make decisions entirely based on the average. A possible soltion for this is to provide a utility framework for a particular experiment (Hullman 2016), however it is unclear how easily participants could adopt a utility framework that is different to their own. Cheong et al. (2016) introduced a very simple optimum strategy into their experiment on bushfire uncertainty visualisations, where participants should choose to evacuate a house as soon as its likelihood of catching fire is 50% or higher. Interestingly, despite this very obvious and simple tactic to maximize payout from the experiment, it seems like many participants did not adopt it, instead acting much more warily as though they were considering whether or not they would *actually* evacuate in the event of a fire. These papers also very rarely consider what the optimum decision would be in a given scenario, and a difficulty with anticipating the best strategy is conflated with the failings of the uncertainty visualisation. Ultimately, we agree with Hullman (2016) who suggested that what determines an approriate ground truth is largely a philosophical exercise.

Authors also occasionally use trust as a proxy for uncertainty. Similar to measuring uncertainty through decision quality, this method also has its own series of issues. 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 transparency (*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 by 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. Considering trust, and not transparency, as the metric of importance in uncertainty communication leads to a questionable subtext that argues against transparency. Hullman (2020) found that author simultaneously argued that failing to visualise uncertainty was akin to fraud, but also many avoided uncertainty visualisation because they didn't want their work to come across as "untrustworthy". 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 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, some studies use a ground truth in these trust experiments, which makes for a very confusing read on exactly how participants are supposed to use the information. - Zhao et al. (2023) displayed a model prediction with uncertainty and took participants using the model prediction as a sign of trust. They reported that visualising uncertainty information caused participants to trust the model in the low variance case, but the results in the high variance case were inconclusive. The discussion made it clear the authors thought the uncertainty information should make the visualisation more trustworthy, but conflating trust and the use of a prediction implied uncertainty information should somehow influence participants to use their own prediction. If the Bureau of Meteorology was uncertain about their rain prediction, that would not be sufficient conditions for me to decide to make my own rain predictions. Despite this, the authors seemed to assume that the uncertainty information *should* have an influence on that, showing they had not deeply considered how uncertainty information should influence the choices of the participants.

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

There are other studies that try to include uncertainty information without directly asking for it but do not go down the route of asking for a decision or levels of trust. These papers often try to ask a question that should utilise both the uncertainty and signal in the response. This is an interesting approach, but these papers are often still looking for a "signal" which is specified in the form of a "ground truth" that is used to evaluate the plot. Unfortunately this

usually comes in the form of slightly cryptic questions that create a large amount of noise on the interpretation side (Hullman 2016).

- Hofmann et al. (2012) showed two distributions in 20 different visualisations (a lineup protocol) using a jittered sample, a density plot, a histogram, and a box plot. Participants were asked to report in which of the plots was “the blue group furthest to the right” and to provide reasoning from a multiple choice list and grade their certainty. The experiment set up is shown in Figure 10. The participants answers were then compared to a ground truth where the correct plot had a right shifted mean. By comparing the results to a ground truth statistic and marking participants as “wrong” or “right”, the error from the participants that had an alternative interpretation to the concept of “furthest right” was conflated with the error from a the visualisation choice.
- L. M. Padilla, Ruginski, and Creem-Regehr (2017) made several confusing assumption when determining the ground truth of their experiments that were assessing storm visualisations. Experiment 2 assumed that the participants should provide responses that indicated they perceived the intensity of the storm to be independent of the uncertainty distribution of its location. This assumption seems counterintuitive as best, however experiment 3 asked participants to “decide which oil rig will receive more damage based on the depicted forecast of the hurricane path” for which they *were* supposed to incorporate uncertainty information and not assume they were independent. The authors seemed to be unaware of these confounding mistakes in their assumptions of how the participants were supposed to utilise the uncertainty information.
- Sanyal et al. (2009) mapped uncertainty to dots and signal to a 3D surface and asked participants to identify areas of high and low signal and high and low uncertainty. Participants were not asked to combine that information in any way, and the signal and the noise were treated as separate variables.
- Correll and Gleicher (2014) asked participants to extract the mean and variance from four uncertainty visualisations, bar charts with error bars, box plots, gradient plots, and violin plots (that were adjusted to make sure the mathematical information they were showing was the same) from two side by side distributions. Participants were also asked to answer a question that was supposed to incorporate both the signal and the noise, such as “How likely is candidate B to win the election?” when the two distributions indicated voter preference. Participants were not able to answer the question about likelihood in term of probability, but were instead given seven options from 1=Outcome will be most in favor of A to 7=Outcome will be most in favor of B. The ground truth statistic for this question was a scalar multiple of Cohen’s d, indicating participants were supposed to incorporate uncertainty information using a very specific formula that was likely unknown to them but assumed to be used implicitly. - Cheong et al. (2016) tested multiple different visual representations of uncertainty for representing the likelihood of a house being burned down based on its location.

It is unclear how participants are supposed to incorporate uncertainty information into their responses when uncertainty visualisation authors themselves are unsure. Other authors interpret this difficulty as a combination of the uncertainty and task complexity (Kinkeldey, MacEachren, and Schiewe 2014), and not as a general confusion. All these examples sever

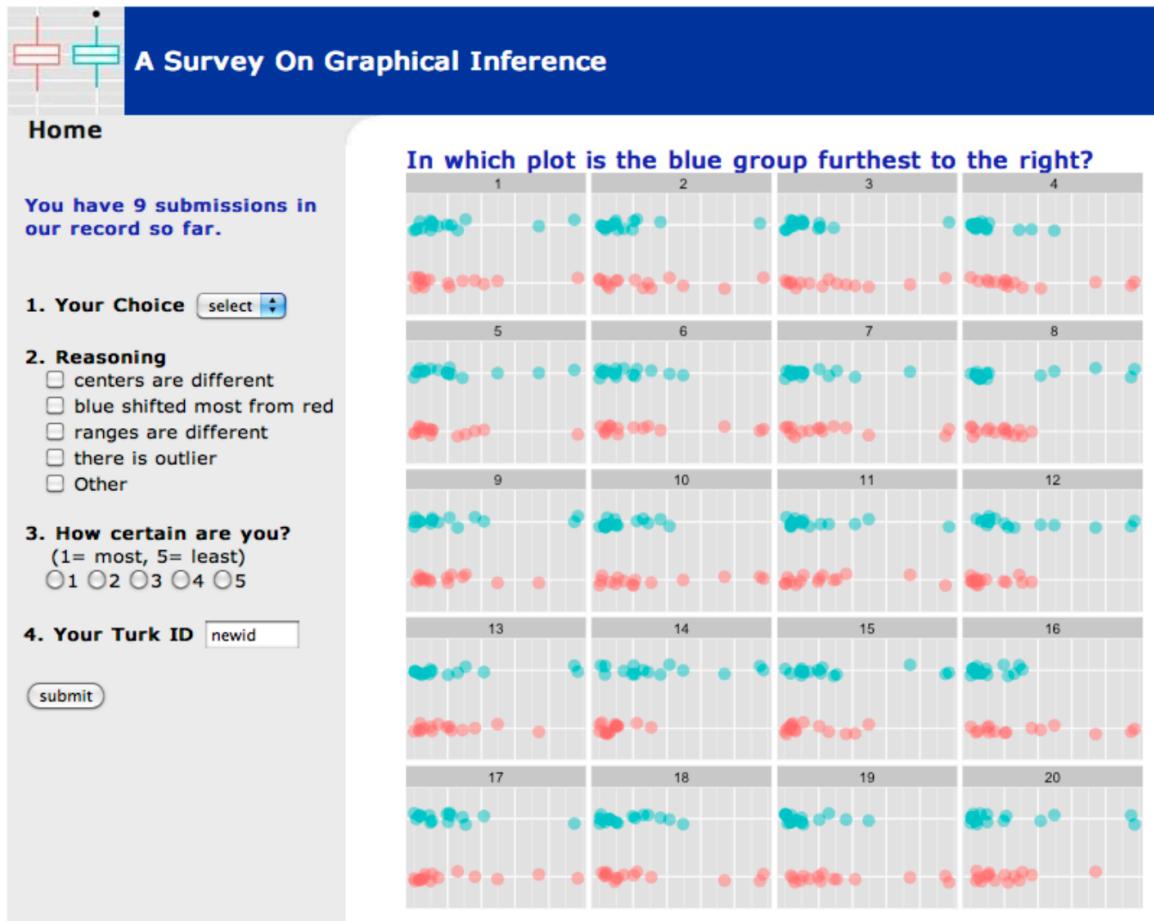


Figure 10: This shows the user interface for the experiment performed by Hofmann et al. (2012). The question of “furthest to the right” is open to interpretation.

to illustrate that *we dont know how to measure uncertainty* and if we don't know how to measure it, *we cannot evaluate the quality of a visualisation*. It could be argued that a well done uncertainty visualisations should have an imperceptible signal unless the signal would be identified with a hypothesis, but this idea also has a miriad of issues that come with it. There is already a series of issues with the reject/do not reject concepts in hypothesis testing, so it is already questionable to set the results of a hypothesis test as the ground truth in an uncertainty visualisation experiment, however these may be additional problems that go beyond the binary nature of the issue. (**Patrick2023?**) compared people ability to recognise patterns in a residual plot to typical statistical tests and found human viewers looking at a plot were less sensitive than the typical residual tests. These experiments utilised the lineup protocol which has been suggested as a method to check if perceived patterns are real or merely the result of chance (Buja et al. 2009; Wickham et al. 2010; Chowdhury et al., n.d.). This concept bears similarity to the goal of uncertainty visualisation, except uncertainty visualisation wants to achieve this within a single plot.

The problem of trying to assess uncertainty without knowing *why* we are depiting uncertainty has been noticed by other authors. Spiegelhalter (2017) noted we “cannot assess the quality of risk communication unless the objectives are clear”.

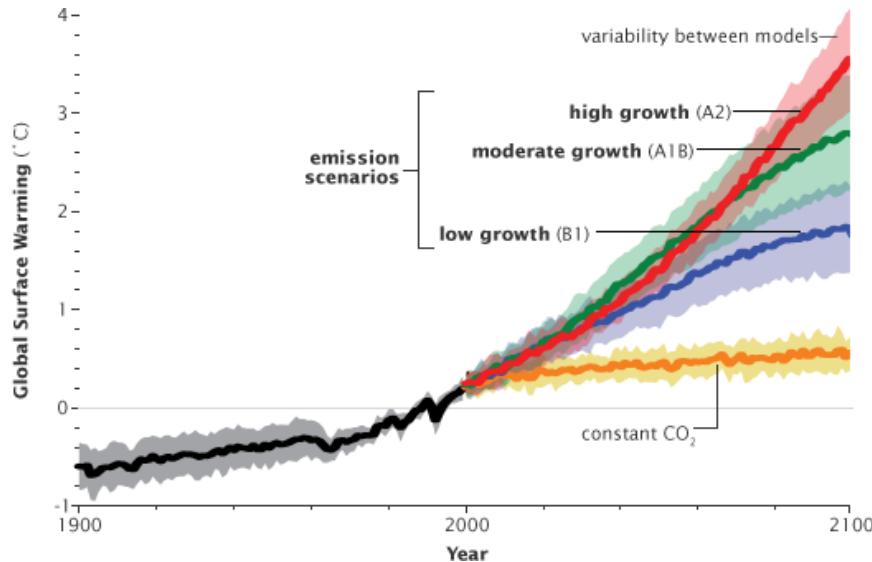
6 Great Examples

Edit notes: this section seems to go off a bit on dimensionality of visual tasks and stuff that might be better put in the noise and signal section about measuring uncertainty.

Of course we do not believe every uncertainty visualisation is unusable and every experiment is build upon misguided assumptions. Despite the common problems detialled in the previous sections, there is some interesting work in the uncertainty visualisation space. This can come in the form of an uncertainty visualisation that identifies or fixes pitfalls in the previous sections (without identifying them) or does an experiment that identifies useful information that avoids these pitfalls.

Uncertainty that is created by steps that are further down the pipeline are often ignored by the uncertainty visualisation literature, however there are visualisations that make these considerations. If uncertainty had a more encapsulating mathematical definition, the uncertainty created at earlier stages might be able to be visualised with the model uncertainty that is typically expressed through confidence intervals. That being said, even with a definition that allows us to combine these uncertainties into a single value, it still might be desirable to see which stage of the analysis is creating uncertainty in the visualisaiton. One example of a visualisation that includes that information is visualisations of climate scenario uncertainty, as shown in **?@fig-climatescenario**. Uncertainty is introduced into this model by the choices of individuals which is difficult to model or predict and is also unlikely to be constant over time.

Therefore specific scenarios are modelled and shown, each with their own model uncertainty. This highlights a method that can be used to visualise uncertainty from earlier in the pipeline that is often forgotten or ignored via assumptions.



{fig-

climatescenario}

A lot of work that identifies heuristics or biases in plots do not hinge on the previous assumptions. This work also provides useful insights for experiments by highlights pitfalls participants might fall into when they review the results of evaluation experiments (Hullman 2016). Newman and Scholl (2012) found that participants were more likely to view points within the bar as more likely than points outside of the bar in bar charts with error bars. Similar effects have been identified in other types of uncertainty displayed. L. M. Padilla, Ruginski, and Creem-Regehr (2017) found that points that were on an outcome of an ensemble display were perceived as more likely than points not on an outcome, even when the point that was not on a specific outcome of the ensemble was closer to the mean of the uncertainty distribution. The sine illusion can cause the confidence interval of a smoothed sine curve to seem wider at the peaks than the troughs, causing us to underestimate uncertainty associated with changing values (Vanderplas and Hofmann 2015)

There are a number of visualisation methods that do take the approach of uncertainty as signal suppression. These visualisation examples try to establish uncertainty as something that should be interpreted *with* the signal and take several approaches to do so. Figure 11 depicts the array of approaches that have been used to visualise uncertainty in a map.

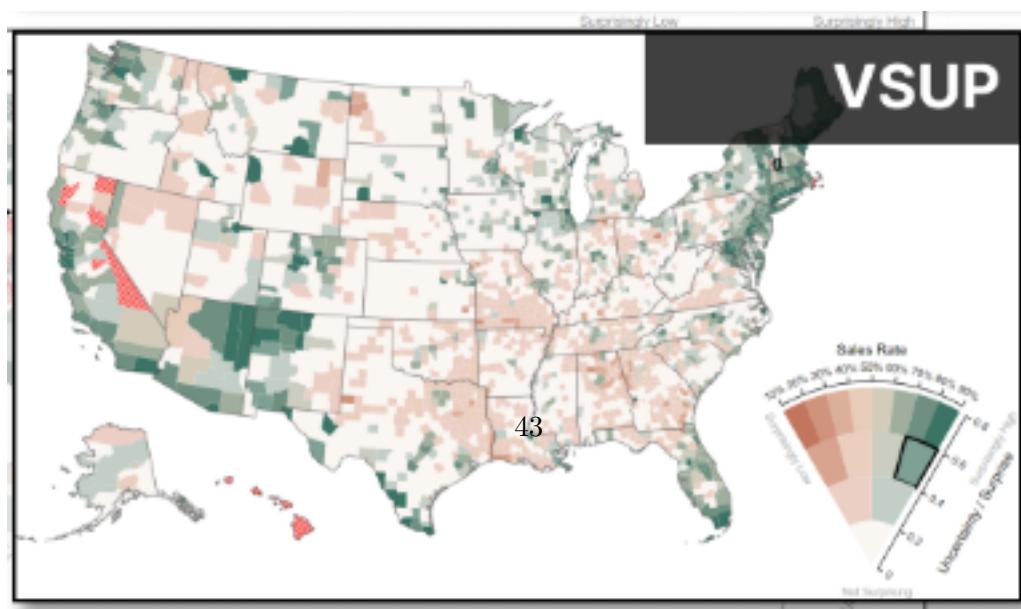
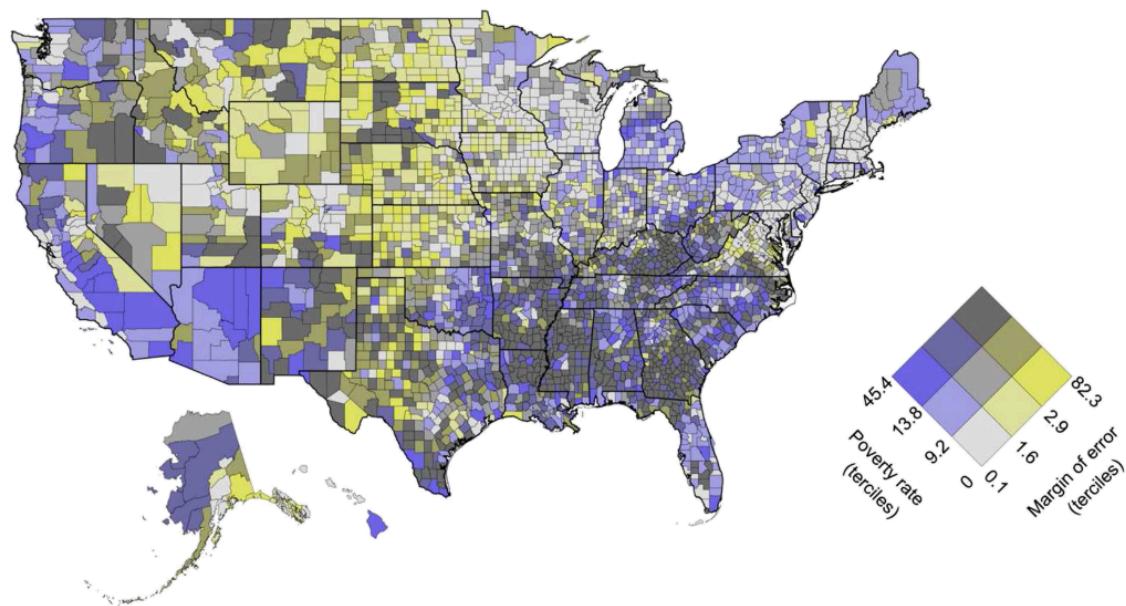
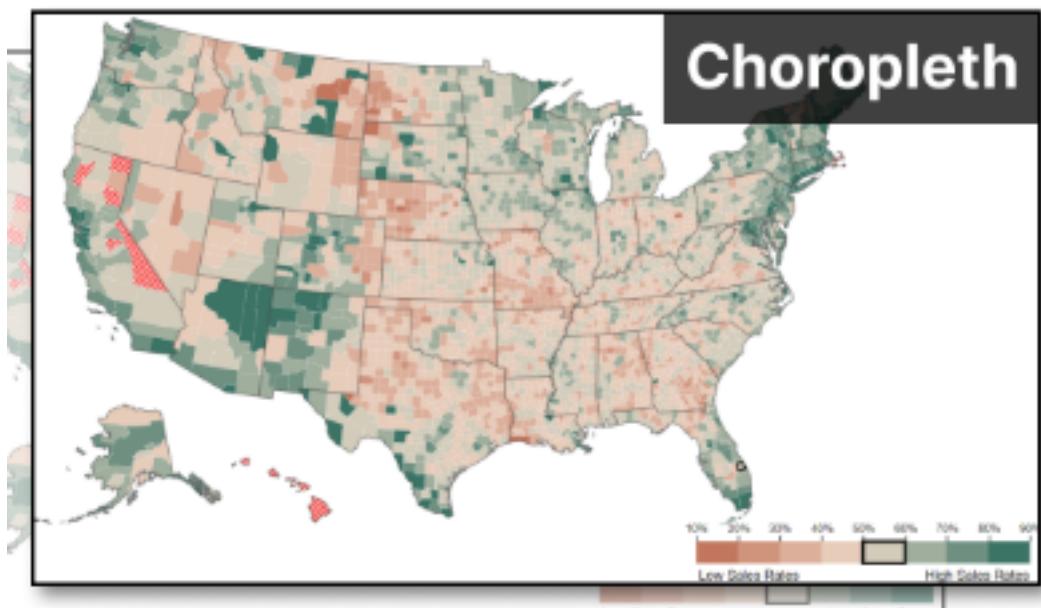
The first approach to embedding noise in the signal is to adjust the signal visualisation so that it includes the noise. This method has been particularly popular with map visualisations. To visualise uncertainty such that it does its job of signal suppression the uncertainty and the signal need to be interpreted in conjunction with each other. There are several suggested ways to go about this, most of which assume the “no uncertainty” version of the visualisation is a

simple choropleth map. There are two methods to include uncertainty in a map. The first is to include uncertainty as error and map our signal and noise to a two dimensional *colour* space, which is the method adopted by the bivariate map and the value suppressing uncertainty pallet (Correll, Moritz, and Heer 2018). These methods rely on the assumption that colour hue and value exist mentally as a two dimensional colour space and that distance in this colour space is linear (or at least appears linear using our scale). However, for uncertainty to be interpreted in the visualisation *with* the signal, our brain would need to be able to flatten this two dimensional space into a single space of “signal validity”, and there is simply no evidence that we can do this. This visualisation method brings up an interesting concept though. If there are two aesthetics we can use that are flattened into a single channel in our mind, then plotting measurement and error to each of those aesthetics is a viable method to visualise uncertainty. The idea of two perceptual tasks flattening into one variable in the mind of the viewer may be wishful thinking, but it is not impossible given we are not certain on how the perceptual tasks are mapped within the human brain. Sterzik et al. (2023) found that when a value was mapped to the textures of stippling, hatching, and triangles, and found that the difference between two points on this one dimensional texture was actually a 2D space (likely “business” and light/darkness). This idea of two aesthetics flattening into one cohesive variable in the perceptual space is particularly interesting for uncertainty visualisation.

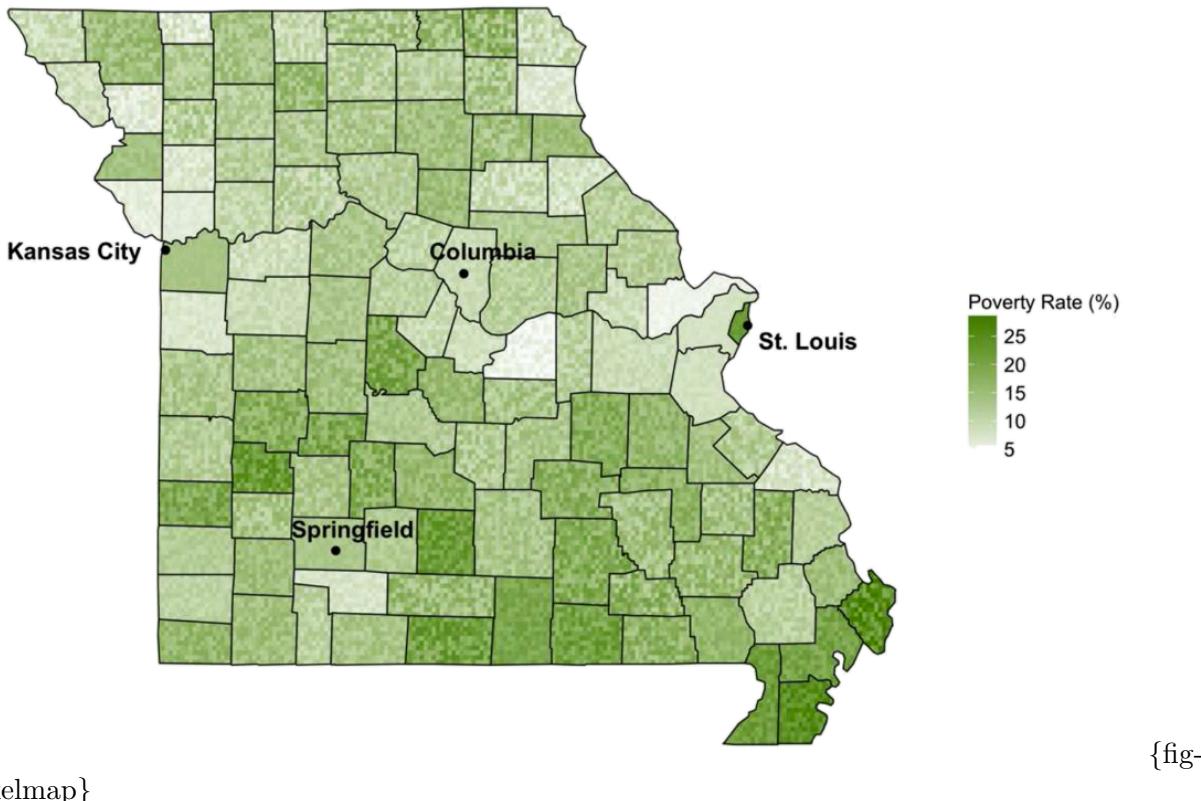
An alternative method to visualising uncertainty in a map is to combine uncertainty and signal at the earlier stage so the “suppressed signal” is represented by a single variable. This statistic can then be expressed in a one dimensional colour space, which is a method adopted by the Bayesian surprise metric map (Ndlovu, Shrestha, and Harrison 2023) and the exceedance probability map [Lucchesi2017]. These methods do not require us to compress colour hue and value into a single variable to interpret signal and noise as a single variable, it is done explicitly for us before the visualisation. These methods require more specificity in *how* the uncertainty contains important information.

These maps make the importance of combining uncertainty and signal in a single visual channel clear. A chloropleth map will show signal that is not valid inference because of high uncertainty. At the other end of the spectrum, the bivariate map will show signal that is not always interesting because it forces us to interpret uncertainty and signal differently. The value suppressing uncertainty pallet was clearly designed with hypothesis testing in mind, because it combines highly uncertain values. The main difference between the VSUP and the bivariate colour map is actually the subtle combination of signal and noise through discernability. In this sense, the uncertainty visualisation maps that combine uncertainty and signal before it is conveyed visually simplify the visualisation and make it easier to interpret value suppressed signal. This is what is done by the exceedence probability map and the Bayesian surprise metric map. They are more explicit in how the uncertainty should suppress the signal since they make the specific hypothesis explicit.

Another method that has been used is to visualise samples from the distribution we are trying to draw inference on instead of estimates along with a variance or error. This has been used in maps with the pixel map [Lucchesi2017], but is more commonly used in animation with the



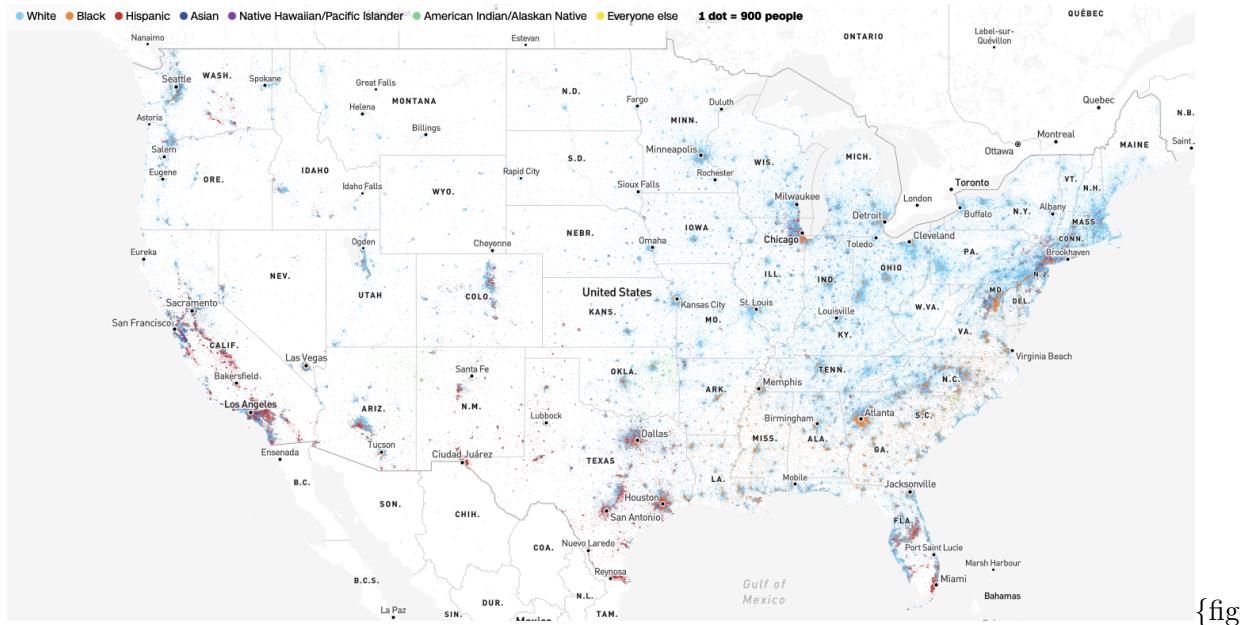
HOPs plots (Hullman, Resnick, and Adar 2015) or similar concepts (Blenkinsop et al. 2000). The difficulty with resampling methods is that they are typically conflated with animation even though this does not necessarily need to be true, as can be seen in the `?@fig-pixelmap`. Even though Kale et al. (2018) found that users were more sensitive to the underlying trend when shown the HOPs plot over the static outcomes plot, this effect went away when the speed of the animation increased, indicating that performance difference might be due to overplotting in the static case and a smaller sample size in the static condition would produce the same result. Therefore, so long as the sample size of the number of outcomes is kept to a level where individual samples are still visible, there is no evidence that an outcome plot *needs* to be animated. Visualisations that opt to express a signal as a sample rather than an estimate have the potential to suppress signal since it is not explicitly visualised, however this has yet to be shown in evaluation experiments. This is not to say that visualisations of mass would not be able to perform signal suppression, but a sample can easily be expressed using aesthetics such as colour on a map and mass visualisation often struggling with issues such as over or under smoothing.



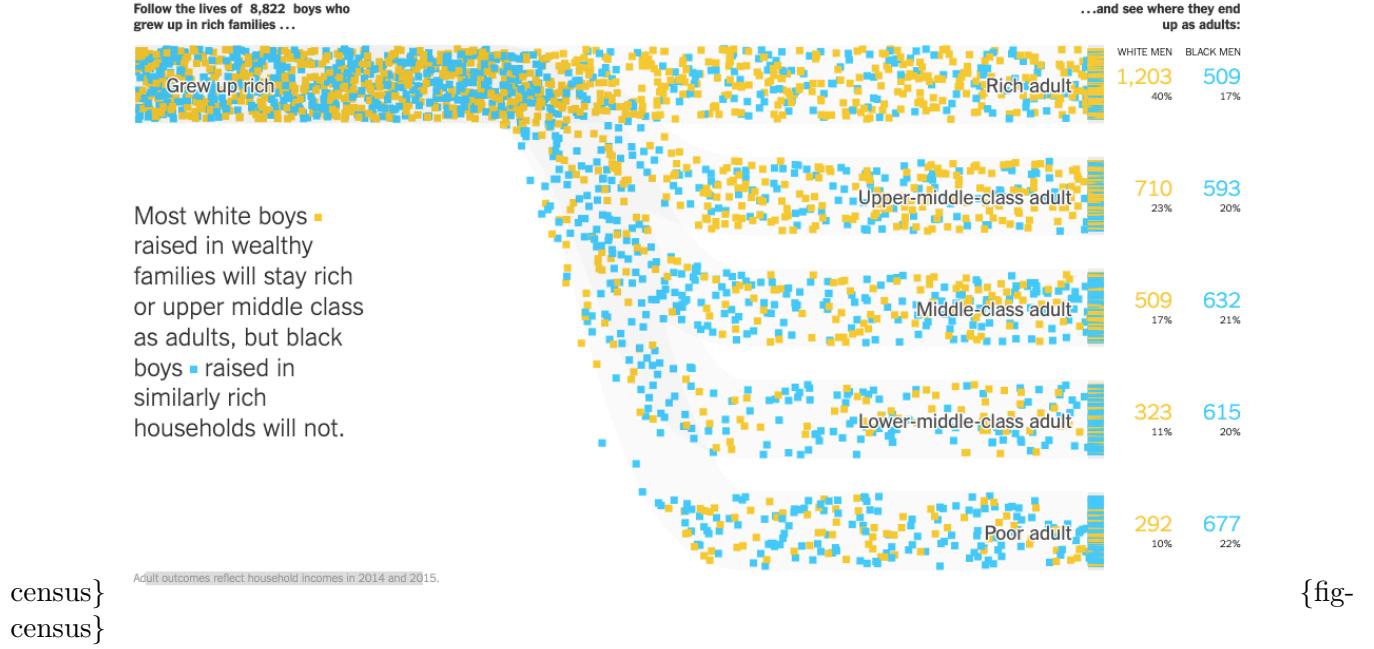
`pixelmap}`

A final method to perform signal suppression is simply to visualise the data if it is available. An example of this is shown in `?@fig-census` for racial distributions spatially in america, an interactive version of the plot can be found [here](#). This map shows the typical causes of uncertainty in a spatial model, i.e. regions where data is sparse, ethnically diverse areas, uneven

distribution of points within boundaries, etc, but it avoids the need to create a visualisation with a specific signal in mind. This is the technique typically employed by exploratory data analysis, which means it's lack of a specified signal means there is both *no* uncertainty (since we are technically not performing inference) but in the event we *do* implicitly perform inference, there is some hedging. In this sense, the best uncertainty visualisation you can get without specifying a signal you want to convey is visualising the data itself. The census dot map's addition of interactivity also allows users to zoom in and see the details that caused "uncertainty" in the form of inconsistent colours at lower resolutions when they were zoomed out. The raw data can also be used in a similar method to the HOPs plot, where a statistic of importance is implied by the visualisation, but the signal is suppressed using the data itself instead of resampling methods. This method is used by the New York Times with their class mobility plot shown in fig, an animated version of which can be found [here](#). This does not mean that visualising raw data instead of implementing sampling techniques is always a valid uncertainty visualisation that will prevent insignificant signal from getting through. Buja et al. (2009) illustrated how groups that appear linearly separable in a linear discriminant analysis (LDA) visualisation of the data can actually be the result of a LDA performed on too many variables, something that was not clear from the visualisation until the lineup protocol was implemented.



{fig-



These great examples show there are visualisations that are being made that have a good conceptualisation of uncertainty and express it as signal suppression, however this work is not being done formally within visualisation. These industry visualisation are likely made with a large amount of back and forth between different levels of an organisation, something that is important for good visual communication (Spiegelhalter 2017).

7 Future work

This paper has identified several issues in the uncertainty visualisation literature and with those issues, we have several suggestions for future work that would likely be fruitful for the field.

Our first suggestion for future work is a mathematical definition of uncertainty. Specifically, a definition that has a place for all the concepts that currently sit under the “uncertainty” umbrella, explains how these concepts relate to uncertainty, and also allows us to quantify them in a way where they can be combined or visualised. This work would not only benefit uncertainty visualisation, but also statistics more broadly. There is currently a lot of work that quantifies the uncertainty through bias and variance at various stages of the data analysis pipeline separately. Elements such as imputed data, assumptions, sampling methods, and analyst decisions all have their own independent work, quantification and discussions, but methods to bridge between these concepts are few and far between. A conceptual framework that allows us to combine these results would be incredibly beneficial to the field.

The concept of uncertainty should also be formalised within the grammar of graphics. Wilkinson (2005) referred to uncertainty as “semantics”, however the depth of integration required to understand problems in uncertainty visualisation suggest they *need* to be formally placed within the framework. Displaying a distribution is not as simple as displaying the data. For example, the distinction between a sample and a smoothed density function is not cut and dry. One may view them both as the same data but one is represented as a point, the other is binned according to some function and then represented as a line. However, this distinction blurs as n gets larger and larger and, depending on the collision modifier used, depictions of the sample will appear identical to depictions of the mass. This is obviously true due to large sample theory, the backbone of statistics. The issue of blurred elements of the grammar of graphics in the case of distribution visualisations does not end here. Visualisations that depict mass such as a confidence interval, a boxplot, a PDF, a letterbox plot, a violin plot, a histogram, a quantile dot plot, etc, all show some version of the mass of a variable at different resolutions, but whether these variables differ in something as low down in the grammar pipeline as the variable stage, or later at the geometry stage is unclear. The lack of formalisation in this way allows authors to compare graphics that *seem* to be similar mathematically but are actually conveying very different information. Not only would a formalisation of uncertainty visualisation within the grammar of graphics allow us to iron out some of these confusions, but they would make it easier to understand when existing visualisation methods apply and can be used to explain our results.

If we are going to consider uncertainty as noise, not signal, there needs to be a way to identify this signal suppression in an experimental design. It is clear that value retrieval and questions with a specific ground truth are not appropriate, as uncertainty visualisation will simply follow the existing visualisation rules. Testing a secondary feature of a plot, that is, how strong a signal is becomes a little more complicated. Line up protocols have been used in adjacent work that look at the strength of the signal in a plot, and this idea of identifying if plots have some “barely noticeable differences” could be utilised. There is also the possibility that uncertainty visualisation evaluations will need to swap to a qualitative methodology where participants are allowed to freely comment on what they notice in graphics until we establish how the existence of noise can be observed.

Additionally, information visualisation needs a more precise concept of the difference between two plots, and that framework should be utilised in experiments. Other fields of science employ marginal changes when designing experiments to ensure it is well understood *what* aspect of their experiment is contributing to their results. Visualisation has the grammar of graphics, however we have already discussed how this can break down in the case of distribution visualisation.

If a visualisation researcher would prefer to perform experiments rather than formalise methods, the task dependency many authors in uncertainty visualisation mention would be a useful direction for research. Unfortunately the current visualisation evaluation literature is *far* too noisy to allow us to identify what might be the cause of this task dependence. It is unclear if it is a conflation between task and the ground truth/displayed statistics concepts discussed earlier,

or an actual difference in the way we perceive graphics. A good starting place for this idea would be to identify if there is task dependency in perceptual tasks, since all perceptual task research to date has been done on value retrieval. That is, an experiment that compares things like area, position, colour and other perceptual tasks on value extraction, comparisons, and other simple tasks. It is also clear that the number of potential tasks that can be performed on a visualisation increases with the number of observations. A single observation is limited to value extraction, two observations can be compared, multiple observations allow for shapes or global statistics to be extracted. The interaction between sample size and task is of particular interest to the uncertainty visualisation community, as uncertainty can be expressed through multiple observations using a sample, or through a single value using an error. Of course, this is limited by the fact that there also isn't a definition for what is a "task" and given the mess created by the lack of formalisation in uncertainty visualisation, it may be wise to formalise that concept before performing these experiments. Amar, Eagan, and Stasko (2005) suggested a taxonomy for information visualisation based on the types of tasks we use visualisations for and suggest 10 "analytical primitives" that we can then map to visualisations, which could be a good starting point. Regardless, these are directions of research that would be fruitful to the uncertainty visualisation community even if it appears on the surface to be research that is only beneficial to the "normal" visualisation community.

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>.
- Blenkinsop, Steve, Pete Fisher, Lucy Bastin, and Jo Wood. 2000. "Evaluating the perception of uncertainty in alternative visualization strategies." *Cartographica* 37 (1): 1–13. <https://doi.org/10.3138/3645-4v22-0m23-3t52>.
- Boukhelifa, Nadia, Anastasia Bezerianos, Tobias Isenberg, and Jean Daniel Fekete. 2012. "Evaluating sketchiness as a visual variable for the depiction of qualitative uncertainty." *IEEE Transactions on Visualization and Computer Graphics* 18 (12): 2769–78. <https://doi.org/10.1109/TVCG.2012.220>.
- Buja, Andreas, Dianne Cook, Heike Hofmann, Michael Lawrence, Eun Kyung Lee, Deborah F. Swayne, and Hadley Wickham. 2009. "Statistical inference for exploratory data analysis and model diagnostics." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367 (1906): 4361–83. <https://doi.org/10.1098/rsta.2009.0120>.
- Carlin, John B., and Margarita Moreno-Betancur. 2023. "On the uses and abuses of regression models: a call for reform of statistical practice and teaching." <http://arxiv.org/abs/2309.06668>.
- Cheong, Lisa, Susanne Bleisch, Allison Kealy, Kevin Tolhurst, Tom Wilkening, and Matt

- Duckham. 2016. “Evaluating the impact of visualization of wildfire hazard upon decision-making under uncertainty.” *International Journal of Geographical Information Science* 30 (7): 1377–1404. <https://doi.org/10.1080/13658816.2015.1131829>.
- Chowdhury, Niladri Roy, Dianne Cook, Heike Hofmann, Mahbubul Majumder, Eun-Kyung Lee, and Amy L Toth. n.d. “Using Visual Statistical Inference to Better Understand Random Class Separations in High Dimension, Low Sample Size Data.”
- Cleveland, William S., and Robert McGill. 1984. “Graphical perception: Theory, experimentation, and application to the development of graphical methods.” *Journal of the American Statistical Association* 79 (387): 531–54. <https://doi.org/10.1080/01621459.1984.10478080>.
- Correll, Michael, and Michael Gleicher. 2014. “Error bars considered harmful: Exploring alternate encodings for mean and error.” *IEEE Transactions on Visualization and Computer Graphics* 20 (12): 2142–51. <https://doi.org/10.1109/TVCG.2014.2346298>.
- Correll, Michael, Dominik Moritz, and Jeffrey Heer. 2018. “Value-suppressing uncertainty Palettes.” *Conference on Human Factors in Computing Systems - Proceedings* 2018-April: 1–11. <https://doi.org/10.1145/3173574.3174216>.
- Fischhoff, Baruch, and Alex L. Davis. 2014. “Communicating scientific uncertainty.” *Proceedings of the National Academy of Sciences of the United States of America* 111: 13664–71. <https://doi.org/10.1073/pnas.1317504111>.
- Grewal, Yashvir, Sarah Goodwin, and Tim Dwyer. 2021. “Visualising Temporal Uncertainty: A Taxonomy and Call for Systematic Evaluation.” *IEEE Pacific Visualization Symposium* 2021-April (April): 41–45. <https://doi.org/10.1109/PACIFICVIS52677.2021.00013>.
- Griethe, Henning, and Heidrun Schumann. 2006. “The Visualization of Uncertain Data: Methods and Problems.” *Proceedings of SimVis '06* vi (August): 143–56. <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Visualization+of+Uncertain+Data:+Methods+and+Problems#0>.
- Hofman, Jake M., Daniel G. Goldstein, and Jessica Hullman. 2020. “How Visualizing Inferential Uncertainty Can Mislead Readers about Treatment Effects in Scientific Results.” *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3313831.3376454>.
- Hofmann, Heike, Lendie Follett, Mahbubul Majumder, and Dianne Cook. 2012. “Graphical Tests for Power Comparison of Competing Designs.” <http://www.public.iastate.edu/>.
- Hullman, Jessica. 2016. “Why evaluating uncertainty visualization is error prone.” *ACM International Conference Proceeding Series* 24-Octobre: 143–51. <https://doi.org/10.1145/2993901.2993919>.
- . 2020. “Why Authors Don’t Visualize Uncertainty.” *IEEE Transactions on Visualization and Computer Graphics* 26 (1): 130–39. <https://doi.org/10.1109/TVCG.2019.2934287>.
- Hullman, Jessica, Matthew Kay, Yea Seul Kim, and Samana Shrestha. 2018. “Imagining Replications: Graphical Prediction Discrete Visualizations Improve Recall Estimation of Effect Uncertainty.” *IEEE Transactions on Visualization and Computer Graphics* 24 (1): 446–56. <https://doi.org/10.1109/TVCG.2017.2743898>.
- Hullman, Jessica, Paul Resnick, and Eytan Adar. 2015. “Hypothetical outcome plots outper-

- form error bars and violin plots for inferences about reliability of variable ordering.” *PLoS ONE* 10 (11). <https://doi.org/10.1371/journal.pone.0142444>.
- Ibrekk, Harald, and M. Granger Morgan. 1987. “Graphical Communication of Uncertain Quantities to Nontechnical People.” *Risk Analysis* 7 (4): 519–29. <https://doi.org/10.1111/j.1539-6924.1987.tb00488.x>.
- Kale, Alex, Matthew Kay, and Jessica Hullman. 2019. “Decision-making under uncertainty in research synthesis: Designing for the garden of forking paths.” *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3290605.3300432>.
- Kale, Alex, Francis Nguyen, Matthew Kay, and Jessica Hullman. 2018. “Hypothetical Outcome Plots Help Untrained Observers Judge Trends in Ambiguous Data.” *IEEE Transactions on Visualization and Computer Graphics* 25 (1): 892–902.
- Kay, Matthew, Tara Kola, Jessica R. Hullman, and Sean A. Munson. 2016. “When (ish) is my bus? User-centered visualizations of uncertainty in everyday, mobile predictive systems.” *Conference on Human Factors in Computing Systems - Proceedings*, 5092–5103. <https://doi.org/10.1145/2858036.2858558>.
- 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>.
- Mack, Arien. 2003. “Inattentional Blindness: Looking Without Seeing.” *Current Directions in Psychological Science* 12 (5): 180–84.
- 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/rssc.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>.
- Ndlovu, Akim, Hilson Shrestha, and Lane T Harrison. 2023. “Taken by Surprise? Evaluating How Bayesian Surprise & Suppression Influences Peoples’ Takeaways in Map Visualizations.” In *2023 IEEE Visualization and Visual Analytics (VIS)*, 136–40. IEEE.
- Newman, George E., and Brian J. Scholl. 2012. “Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias.” *Psychonomic Bulletin and Review* 19 (4): 601–7. <https://doi.org/10.3758/s13423-012-0247-5>.
- 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>.

- Padilla, Lace M. K., Maia Powell, Matthew Kay, and Jessica Hullman. 2021. “Uncertain About Uncertainty: How Qualitative Expressions of Forecaster Confidence Impact Decision-Making With Uncertainty Visualizations.” *Frontiers in Psychology* 11 (January). <https://doi.org/10.3389/fpsyg.2020.579267>.
- Padilla, Lace M., Ian T. Ruginski, and Sarah H. Creem-Regehr. 2017. “Effects of ensemble and summary displays on interpretations of geospatial uncertainty data.” *Cognitive Research: Principles and Implications* 2 (1). <https://doi.org/10.1186/s41235-017-0076-1>.
- Potter, K., J. Kniss, R. Riesenfeld, and C. R. Johnson. 2010. “Visualizing summary statistics and uncertainty.” *Computer Graphics Forum* 29 (3): 823–32. <https://doi.org/10.1111/j.1467-8659.2009.01677.x>.
- Potter, Kristin, Paul Rosen, and Chris R. Johnson. 2012. “From quantification to visualization: A taxonomy of uncertainty visualization approaches.” *IFIP Advances in Information and Communication Technology* 377 AICT: 226–47. https://doi.org/10.1007/978-3-642-32677-6_15.
- Refsgaard, Jens Christian, Jeroen P. van der Sluijs, Anker Lajer Højberg, and Peter A. Vanrolleghem. 2007. “Uncertainty in the environmental modelling process - A framework and guidance.” *Environmental Modelling and Software* 22 (11): 1543–56. <https://doi.org/10.1016/j.envsoft.2007.02.004>.
- Sanyal, Jibonananda, Song Zhang, Gargi Bhattacharya, Phil Amburn, and Robert J. Moorhead. 2009. “A user study to compare four uncertainty visualization methods for 1D and 2D datasets.” *IEEE Transactions on Visualization and Computer Graphics* 15 (6): 1209–18. <https://doi.org/10.1109/TVCG.2009.114>.
- Simons, Daniel J., and Christopher F Chabris. 1999. “Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events.” *Perception* 28 (9): 1059–74.
- Slingsby, Aidan, Richard Reeve, and Claire Harris. 2023. “Gridded Glyphmaps for Supporting Spatial COVID-19 Modelling.” <https://www.staff>.
- Spiegelhalter, David. 2017. “Risk and uncertainty communication.” *Annual Review of Statistics and Its Application* 4: 31–60. <https://doi.org/10.1146/annurev-statistics-010814-020148>.
- Sterzik, Anna, Monique Meuschke, Douglas W. Cunningham, and Kai Lawonn. 2023. “Perceptually Uniform Construction of Illustrative Textures.” <http://arxiv.org/abs/2308.03644>.
- Thomson, Judi, Elizabeth Hetzler, Alan MacEachren, Mark Gahegan, and Misha Pavel. 2005. “A typology for visualizing uncertainty.” *Visualization and Data Analysis* 2005 5669 (March 2005): 146. <https://doi.org/10.1117/12.587254>.
- Vanderplas, Susan, and Heike Hofmann. 2015. “Signs of the Sine Illusion — Why We Need to Care Signs of the Sine Illusion — Why We Need to Care” 8600. <https://doi.org/10.1080/10618600.2014.951547>.
- Vranas, Peter B. M. 2000. “Gigerenzer’s normative critique of Kahneman and Tversky.” *Cognition* 76 (3): 179–93. [https://doi.org/10.1016/S0010-0277\(99\)00084-0](https://doi.org/10.1016/S0010-0277(99)00084-0).
- 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](https://doi.org/10.1002/(sici)1099-0771(199709)10:3%3C243::aid-bdm268%3E3.0.co;2-m).
- Wickham, Hadley, Dianne Cook, Heike Hofmann, and Andreas Buja. 2010. “Graphical inference for infovis.” *IEEE Transactions on Visualization and Computer Graphics* 16: 973–79. <https://doi.org/10.1109/TVCG.2010.161>.
- Wickham, Hadley, and Heike Hofmann. 2011. “Product plots.” *IEEE Transactions on Visualization and Computer Graphics* 17 (12): 2223–30. <https://doi.org/10.1109/TVCG.2011.227>.
- Wickham, Hadley, Heike Hofmann, Charlotte Wickham, and Dianne Cook. 2012. “Glyph-maps for visually exploring temporal patterns in climate data and models.” *Environmetrics* 23 (5): 382–93. <https://doi.org/10.1002/env.2152>.
- Wilkinson, Leland. 2005. *The Grammar of Graphics (Statistics and Computing)*. Berlin, Heidelberg: Springer-Verlag.
- 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>.