# CHS: Small: Developing a Probabilistic Grammar of Graphics for Flexible Uncertainty Visualization

Matthew Kay, University of Michigan

Journalists and scientists regularly communicate uncertainty information to the general public through visualization. When reporting on high-stakes and uncertain topics ranging from elections to natural disasters, journalists routinely employ visualizations incorporating uncertainty, allowing the public to experience uncertainty to improve understanding and decision-making: *Who will win the next presidential election? Should I evacuate in the face of potential flooding?* "Cones of uncertainty" are employed by the National Oceanic and Atmospheric Administration to communicate the uncertainty in hurricane path predictions [41], despite evidence suggesting the uncertainty in these predictions is misinterpreted by the public [48] and the existence of improved alternatives in the literature [44, 39]. Why does the adoption of more effective uncertainty visualization lag? One reason is that the construction of more sophisticated and effective uncertainty visualizations is more difficult, technically, than the construction of simple, less effective uncertainty visualizations like confidence intervals or prediction cones.

One domain that has seen a more creative approach to uncertainty visualization in journalistic circles is that of election forecasting. During the past several US elections, forecasters at *FiveThirtyEight.com* have directly displayed probability distributions representing their predictions for which party will control the House, Senate, or win the presidential election, often using elaborate interactive visualizations that allow readers to explore many facets of the predictions [12]. In the 2016 presidential election, the *New York Times* created a forecast "needle", which displayed the real-time predicted electoral college margin as election returns were being counted [23]. The New York Times election needle employed a sophis-

The New York Times election needle, as used in the 2016 election, **oscillated** in proportion to the uncertainty in the current predicted presidential vote margin as returns were counted in real time, an example of a **hypothetical outcome plot**.
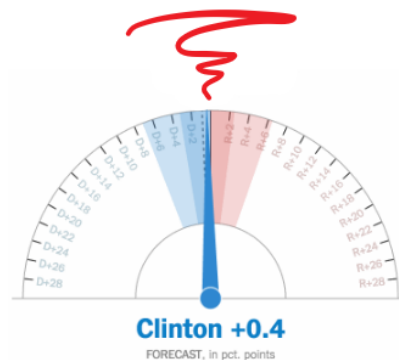

Hurricane error cones (left), employed by NOAA and journalists, are often misinterpreted by people (e.g., by thinking that they depict predicted size instead of uncertainty). Line ensembles (right), a **frequency framing** visualization, mitigate this misinterpretion.
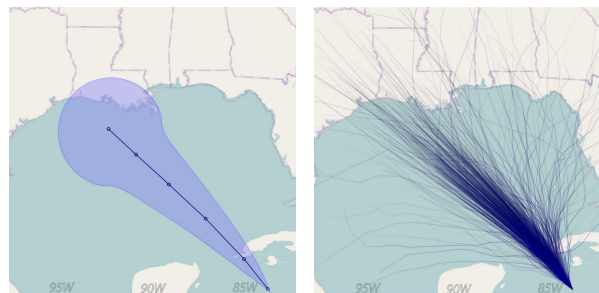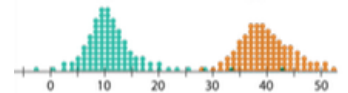

Compared to density plots, quantile dotplots (another **frequency framing** uncertainty visualization), are easier for people to estimate intervals from and yield higher-quality decisions in incentivized decision-making experiments.

Density plots
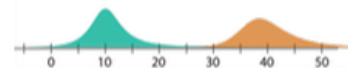
Quantile dotplots

Figure 1: Examples of several uncertainty visualizations one might like to support in a probabilistic grammer of graphics.

ticated animated uncertainty visualization—the needle jittered in proportion to the prediction's uncertainty—en example of a *hypothetical outcome plot* (HOPs) [26]. HOPs is an approach that

has been found to improve lay people's understanding of uncertainty by allowing them to experience uncertainty as natural frequencies, one animated outcome at a time [26, m5].

While more effective, uncertainty visualizations like hypothetical outcome plots are even harder to produce than traditional uncertainty visualizations: The New York Times election needle is a bespoke dynamic JavaScript application created by one of a small number of premier data journalism outfits; its implementation involves correctly handling probability distributions, sampling, and animation, and these aspects are intertwined in the implementation. Besides hypothetical outcome plots, effective—but difficult-to-implement—animated and static uncertainty visualizations exist across many domains, such as transit arrival prediction [m8, m1], hurricane path prediction [31, 44, 43, 20, 39], and medical risk communication [13]. This technical complexity is not only a barrier to less sophisticated designers (e.g., who do not have access to an in-house team dedicated to implementing visualizations), but is also a barrier to sophisticated designers who might wish to rapidly explore many possible visualization prototypes before settling on a single approach. Currently, therefore, effective uncertainty visualizations are **difficult to create** and the uncertainty visualization design space is **costly to explore**.

Various formalizations of information visualization construction into *algebras* [16, 37] or *grammars* [55] have sought to decrease the technical sophistication users need to create visualizations while simultaneously making it possible to explore a wide range of visualizations easily. The most successful of these have been based on the grammar of graphics [55], which underlies ggplot2 [51, 52] and Vega-Lite [49], and the similar VizQL algebra [38], which underlies Tableau. These algebras (and the toolkits based on them) are successful because they are not based on a set of predefined template visualizations (too restrictive), nor are they based on a low-level graphics API (unrestrictive—but requiring a large amount of code to create even simple visualizations). Instead, they operate at just the right level of abstraction, one which visualization designers also operate in: they formalize notions of *visual channels* (such as position, length, area, color), *geometries* (such as bars, lines, points, ribbons), and *statistical transformations* (such as means, histograms, kernel density estimators). These primitives can be recombined to create sophisticated visualizations easily, facilitating design exploration. However, none of these grammars formalize the notion of uncertainty or probability, requiring users to recombine statistical transformations and geometries (or worse, to transform input data outside of the visualization framework) in order to create effective uncertainty visualizations (both hypothetical outcome plots [26, m5] and quantile doplots [m8, m1] would require different preparations of the data in order to be visualized within these frameworks, making design exploration cumbersome).

This work seeks to ask: *How would a grammar of graphics that incorporates uncertainty as a first class entity—a **probabilistic grammar of graphics**—change how we create uncertainty visualizations? Could a simple, consistent probabilistic grammar of graphics be constructed that easily supports a wide range of modern, effective uncertainty visualizations? Would a probabilistic grammar of graphics allow designers to more easily explore the space of uncertainty visualizations? Would a probabilistic grammar of graphics enable designers to create more correct uncertainty visualizations?* By implementing results of this proposal on top of ggplot2 and Vega-Lite, successful results would lead to improved support for uncertainty visualization in grammar of graphics-based visualization toolkits, making it easier for designers of all expertise levels to adopt effective uncertainty visualization techniques. This would provide substantial improvements to the way that uncertainty and risk are communicated to the public across a range of domains.

Intellectual Merit: This work addresses the following areas:

1. **Systematically cataloging existing uncertainty visualizations:** I will systematically col-

lect and analyze existing uncertainty visualizations from a variety of sources, including the academic literature and the web. For the academic literature, I will draw upon existing literature reviews of uncertainty visualization and risk visualization [35, 13, 21, 36], including my own work reviewing uncertainty visualization evaluation practices [m4]. However, many especially expressive examples of uncertainty visualization—particularly those employing sampling-based approaches similar to hypothetical outcome plots—have come from data journalists, as well. Therefore, I expand the scope of existing reviews to include "best of" lists of professional information visualizations (*e.g.* Kirk's monthly "best of" lists [29]). Finally, I will look at all visualizations produced in 2018 by a small number of well-known, high-quality data journalist teams, such as The New York Times' Upshot, FiveThirtyEight, and the Guardian. I will collect and classify these visualizations into a taxonomy of uncertainty visualizations.

2. **Development of a probabilistic grammar of graphics:** Given the visualizations collected in Aim 1, I will develop a specification grammar that encompasses as large a subset of those visualizations as possible while still being grounded in a grammar based on a combination of probability distributions with the grammar of graphics. As the primary goal of this grammar will be to facilitate rapid prototyping, the aim will be to find abstractions that generalize some of the specific instances of charts found while making it easy to quickly try out (or discover) new ways of presenting uncertainty. A secondary goal will be to facilitate the construction of more correct uncertainty visualizations; partly, this will be achieved through grounding the grammar in the specification of probability distributions as probabilistic graphical models, such that if an author wishes to (say) communicate a particular conditional probability, they would specify this in the grammar and the visualization system would determine how to do that. This will allow authors to easily swap out and explore different presentations while maintaining the correctness of the output visualizations.

3. **Understanding the expressiveness of the probabilistic grammar of graphics and the correctness of uncertainty visualization in the wild** I will re-specify professional, in-the-wild visualizations collected in Aim 1 using the grammar developed in Aim 2. This will enable two goals: first, to determine to the *expressiveness* of the probabilistic grammar of graphics; i.e., to what extent it is capable of generating a large variety of in-the-wild uncertainty visualizations. Second, to assess the correctness of in-the-wild visualizations. Because the grammar will be grounded in formal specification of probability distributions, it will possible to determine if a visualization actually communicates a particular uncertainty correctly. For example, one might ask: if a particular visualization is intended to depict a given conditional probability distribution (e.g., $P(A|B)$), does it? Reasoning about conditional probability can be difficult for people [22], and as described in the Background below, even experts can create visualizations that do not correctly convey the desired probabilities: depending on representation and specification, $P(A|B)$ might not be correctly depicted in a visualization.

4. **Understanding the usability of a probabilistic grammar of graphics** I will conduct a user study to assess whether or not visualization authors can rapidly and correctly create uncertainty visualizations using a probabilistic grammar of graphics. I will solicit participants who have been trained in a grammar-of-graphics-based toolkit, such as Vega-Lite [49]. I will have them use either that toolkit or a modified version of the same toolkit to create visualizations of particular distributions, and assess task difficulty and the correctness of created visualizations.

The proposed work will make the creation of sophisticated, easy-to-understand, correct un-

certainty visualizations routine for visualization designers, in the same way that the grammar of graphics already facilitates the rapid exploration of visualization design more generally. This work also raises the possibility for future automated visualization design systems to support the creation of verified-correct high-quality uncertainty visualization systems: automated visualization design systems based on grammar of graphics-like formalisms already exist [37, 40], but do not address uncertainty visualization directly. Provided with a similar formalism for uncertainty visualization, they might.

Broader Impacts of the Proposed Work: I will directly integrate the output of this research into teaching. I teach a 60-student Professional Master's class in Information Visualization. This class already extensively uses the grammar of graphics to teach students how to reason about and construct effective visualizations, and students are taught both the Vega-Lite [49] and Tableau [38] visualization toolkits. I also teach a module on uncertainty visualization in that class. I will integrate the probabilistic grammar of graphics into this curriculum, enhancing students' understanding of uncertainty visualization. The University of Michigan School of Information is also launching an online Master's degree as a series of Massic Open Online Courses (MOOCs), one of these will be a course on uncertainty visualization, which I will incorporate content from the results of this proposal into.

I will also build out probabilistic extensions to both Vega-Lite [49] and ggplot2 [52], popular grammar of graphics toolkits. I have a track record of creating and maintaining open source tools, including tidybayes [m6], an R package for visualizing Bayesian regression output in R; this package receives over 1000 downloads per month. With widely available methods for quickly building and prototyping uncertainty visualizations built on top of popular visualization toolkits, more effective uncertainty visualizations can proliferate. This will make the construction of correct, effective uncertainty visualization routine, which will in turn help the public make better decisions under uncertainty all the way from domains like transit decision-making [m8, m1] to medical decision-making [13] to decision-making during natural disasters [44, 20, 39] to decisions during elections [23].

1 Background and Related Work

To put the proposed research in context, here I provide a discussion of the most related work in uncertainty visualization and the construction of visualizations through the grammar of graphics.

Uncertainty Visualization: A starting point for distinguishing uncertainty visualization types is whether those visualizations use *intrinsic* or *extrinsic* representations of uncertainty [m8]. A common *extrinsic* representation is an annotation such as a confidence interval or confidence band. Such representations are notoriously difficult for people to correctly interpret [15], and often underperform other *intrinsic* representations, such as hypothetical outcome plots [26, m5], quantile dotplots [m8, m1], or line ensembles [44]. Intrinsic representations may vary from more abstract representations, such as densities [27], to *frequency framing* approaches, in which discrete outcomes are shown.

Frequency framing approaches are motivated by literature suggesting that people better reason about Bayesian probability when presented with probabilities as natural frequencies instead of abstract probabilities [22] (*e.g. 7 out of 10* instead of *70%*). The classic example of a frequency framing approach is an *icon array* [13], which displays uncertainty in a categorical variable using a matrix of possible outcomes (such as dots or icons of individuals) and which is commonly used in medical risk communication. I extended the icon array approach to apply to continuous random variables by introducing *quantile dotplots*, which improve probability estimation and decision-

making over intervals or densities [m8, m5]. Hullman *et al.* introduced an animated frequency framing uncertainty visualization, hypothetical outcome plots (HOPs), which also improve performance over common static and non-frequency framing approaches [26, m5]. Building on these recent advances in effective, *intrinsic* uncertainty visualizations, an important goal of the proposed work is to support intrinsic, frequency-framing uncertainty visualization approaches in particular, as evidence increasingly suggest these approaches better align with how people reason about uncertainty. At the same time, correctly constructing frequency framing uncertainty visualizations is often technically more difficult than simple extrinsic displays, like confidence intervals.

The Grammar of Graphics: The Grammar of Graphics is a systematic way of describing the elements in a visualization, such as data, geometries, visual encodings,[1] statistical transformations, and scales [55, 51, 52]. As an example, a simple scatterplot of variable A against variable B might be described by a specification like this:

$$\text{encodings}(x \ position \leftarrow A, y \ position \leftarrow B) \ +$$
$$\text{geometry}(point)$$

Similarly, a bar chart of the same data might be:

$$\text{encodings}(x \ position \leftarrow A, height \leftarrow B) \ +$$
$$\text{geometry}(bar)$$

Importantly, encoding additional variables into other visual channels typically does not require substantial changes to a design. For example, a scatterplot showing two different groups based on a variable $C \in \{c_1, c_2\}$ might simply map the variable C onto the *color* channel, generating a scatterplot where points are colored different depending on which group they below to (c_1 or c_2):

$$\text{encodings}(x \ position \leftarrow A, y \ position \leftarrow B, color \leftarrow C) \ +$$
$$\text{geometry}(point)$$

While this works well for straightforward visualizations of data points, visualizing uncertainty in toolkits like ggplot2 requires the use of a whole variety of different geometries and statistics, such as densities, violin plots, histograms, dotplots, intervals, and ribbons [52][2]. These different plot types, if viewed within a probabilistic grammar of graphics, should not be so different from each other, but rather could be viewed as different variants of a much smaller set of geometries, with probability distributions assigned to different visual channels.

Frequency and proportion visualization grammars: Other visualization formalisms have been proposed for the creation of visualizations of *frequencies* or *proportions*; i.e. visualizations that depict counts or proportions within a dataset. *Product plots* [53] generalize a variety of plot types, including mosiac plots, treemaps, and stacked bar charts within a framework that specifies charts

---

[1]I use *visual encodings* in place of *aesthetics*—as they are called in ggplot2—to be more consistent with terminology elsewhere in the visualization literature.

[2]but see, in particular, the ggplot2 function reference: https://ggplot2.tidyverse.org/reference/ and the proliferation of alternative geometries for uncertainty and probability visualization in the ggplot2 ecosystem, such as the ggridges package: https://cran.r-project.org/web/packages/ggridges/vignettes/introduction.html

in terms of conditional probabilities and a small number of types of *layouts*. While flexible, this framework requires variables either to be discrete or to have been discretized; thus, continuous random variables are not fully supported (visualizations like probability densities, for example, are not specifiable as product plots). Frequency-framming visualizations of uncertainty—which are known to be particularly effective [m1, m8, 13, 26, m5, 44, 43]—are also not supported. Further, this framework is not expressed in terms of geometries and visual channels, making it difficult to integrate into the grammar of graphics. However, the fact that a wide variety of plot types can be expressed in terms of conditional probability within this framework does demonstrate the potential for directly integrating probability notation into plot specification grammars.

The *Atom unit visualization* grammar [45] is another visualization grammar for the specification of frequency data; it is focused on the construction of visualizations in which each data point is displayed on-screen as a dot (hence "unit"), which encompasses a range of frequency-framing uncertainty visualizations. Atom also generalizes a wide variety of visualization types, such as treemaps [28] and dotplots [54], and does so primarily by incorporating sophisticated types of *layouts* into the specification. Like product plots, this approach precludes common uncertainty visualizations like densities, and does not directly act within a grammar of graphics like approach (e.g., allowing different geometries to be combined into a single display). Unlike product plots, it does not incorporate probability distribution specification.

Like *product plots*, the goal of our grammar is to build upon a formal specification of probability. Close relatives to this approach outside of visualization can be found in Bayesian probabilistic programming languages, such as BUGS [33, 34], JAGS [46], and Stan [17]. Such languages allow users to directly specify Bayesian graphical models in terms of probability distributions[3]. Incorporating similar probabilistic programming capabilities into a visualization framework would allow the specification of graphics output in terms of probability distributions.

2 Preliminary Work

I have studied the properties of frequency framing uncertainty visualizations [m5, m3, m4, m8, m1], and developed *quantile dotplots* [m8, m1], a novel static frequency-framing alternative to density plots and cumulative distribution functions. In transit decision-making, quantile dotplots improve probability estimation and decision quality over other displays [m8, m1]. In addition to developing and testing novel uncertainty visualizations, I have developed and maintained tidybayes [m6], an R package designed to aid in the visualization of results from Bayesian models fit using Markov Chain Monte Carlo (MCMC) methods in probabilistic programming languages like BUGS [33, 34], JAGS [46], and Stan [17]. This package acts as glue between the output of those modelling languages and ggplot2 [52]. It will be part of the infrastructure used to implement a complete probabilistic grammar of graphics: Currently, tidybayes assists in steps for preparing the output of MCMC with ggplot2. However, different uncertainty visualizations still require different data preparation within tidybayes, and it is still not possible to specify probability distributions to visualize directly within ggplot2. My work in Aim 2 will close this gap.

3 Aim 1: Catalog existing uncertainty visualizations

In order to develop a comprehensive probabilistic grammar of graphics, it is necessary to scope out a design space of effective uncertainty visualizations. The goal of Aim 1 is to determine the bounds of that design space:

---

[3]Strictly speaking, Stan allows users to specify a log probability function, but in many ways has a similar user interface to those of the more declarative BUGS and JAGS, which explicitly involve the specification of graphical models

Research Questions: What kinds of uncertainty visualizations exist in the literature that should be supported by a probabilistic grammar of graphics? What kinds of uncertainty visualizations exist in the wild—particularly in use by professional data journalists—that should be supported by a probabilistic grammar of graphics? What properties do these visualizations have that better enable lay understanding of probability, and which are therefore important to support?

Key Challenges: The diversity of uncertainty visualization types may make categorization difficult. Finding the right balance of generalization will be key: are violin plots [18] different from density plots, or just different parameterizations of a similar plot? What about icon arrays and quantile dotplots? Categorizing animated approaches may also present a challenge: are there common strategies used when depicting uncertainty using animation (e.g., as a metaphor for drawing samples from a distribution)? Or do these approaches defy some higher-level categorization and formalization?

This aim will proceed by collecting and systematically categorizing uncertainty visualizations from a variety of sources. Two main sources of visualizations will be used:

1. **Uncertainty visualizations in the academic literature:** I will start with review articles on uncertainty visualization, such as a review I conducted recently [m4], other reviews in the visualization literature [35, 36] and in the risk communication literature [13, 21]. I will also collect uncertainty visualization articles by searching for "uncertainty visualization", "risk visualization", "probability visualization", and related terms on scholarly search engines and corpora, such as Google Scholar, the ACM Digital Library, and IEEE Xplore.

2. **Professional visualizations on the web:** I will draw upon visualization "best of" lists to find exemplar professional visualizations on the web. For example, Andy Kirk, a professional visualization designer, regularly publishes visualization "best of" lists on his blog, *Visualising Data*: http://www.visualisingdata.com/blog/. David McCandless, a visualization designer, produces the yearly *Information is Beautiful* awards, a curated list of the best visualizations of the year across eight categories (https://www.informationisbeautifulawards.com/about). Maarten Lambrechts, a data journalist, maintains a catalog of esoteric visualizations he calls *xenographics* (https://xeno.graphics/), which includes a category for uncertainty. In addition to combing these archives, I will look through archives of visualizations from premier data journalism outfits, such as The New York Times' Upshot, FiveThirtyEight, Bloomberg, and the Guardian.

Uncertainty visualizations obtained from this systematic review will be systematically categorized across several dimensions using affinity diagramming techniques. This will proceed in a combination of a top-down and bottom-up fashion: it will begin with some categories drawn from the literature, but also allow categorizations and dimensions to emerge from the data. Some such categories might include:

- Intrinsic or extrinsic representation used (*e.g.* intervals versus densities)
- Visual encodings used: units (*e.g.* icon arrays [13], unit visualizations [45], quantile dotplots [m8, m1]), length (histograms), area (mosaics, densities, product plots [53], pie charts), color [32, 19]
- Use of frequency framing
- Denominators in frequency framing visualizations [13]
- Epistemic or aleatory uncertainty depicted [42]
- Use of animation (*e.g.* HOPs-like approaches [26, m5]) or sampling

- Use of metaphor (*e.g.* risk communication theatres [47]), New York Times Election Needle [23])
- Use of visual depictions that suggest randomness: *e.g.* random order in icon arrays [13], dithering in areal uncertainty on maps [32].

Visualizations will be systematically categorized within the dimensions above and others which emerge from the review. These dimensions will define a design space of uncertainty visualizations that I might wish to support in Aim 2. Design dimensions (and particular values within each dimension) will be prioritized, in order to identify the aspects of the uncertainty visualization design space that are most crucial to cover. In particular, I will focus on uncertainty visualization types that are both known to be effective in the literature and which are not well-supported by existing grammar of graphics approaches: this means, for example, that if necessary, some common but less effective visualization types (like intervals) may be left aside if necessary. While not a given, this narrowing process may be necessary to ensure a final grammar that is consistent and expressive without being overly complex.

4 Aim 2: Develop and implement a probabilistic grammar of graphics

Given the visualizations collected in Aim 1 and their systematic categorization, I will develop a specification grammar that encompasses as large a subset of those visualizations as possible while still being grounded in a grammar based on probability distributions.


Research Questions: How would a probabilistic grammar of graphics that incorporates uncertainty as a first class change how we create uncertainty visualizations? How would it need to be constructed to support a variety of effective, modern uncertainty visualization techniques, including frequency framing and animation?

Key Challenges: Supporting a diversity of uncertainty visualization types within a single consistent grammar may be difficult. As in aim 1, finding the right balance of generalization will be key: are violin plots [18] different from density plots, or just different parameterizations of a similar plot? What about icon arrays and quantile dotplots? Animation also presents a challenge: what aspects of animation are necessary to be encoded in such a grammar, if any?


As the primary goal of this grammar will be to facilitate rapid prototyping, the aim will be to find abstractions that generalize some of the specific instances of charts found while making it easy to quickly try out (or discover) new ways of presenting uncertainty. As an example of what this could look like, consider the following three possible visualization specifications, each of which includes a conditional probability distribution:
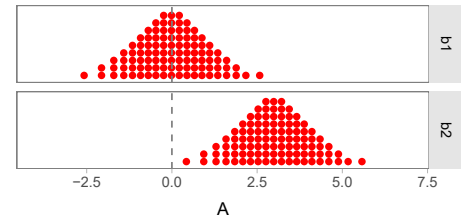
(1) encodings($x\ position \leftarrow A, y\ position \leftarrow B, height \leftarrow P(A|B)$) +
     geometry($block$)
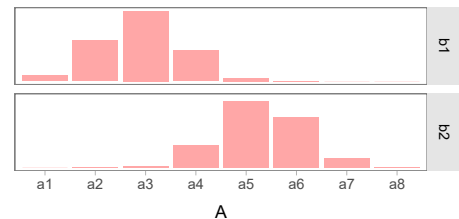
(2) encodings(x position $\leftarrow A$, y position $\leftarrow B$, height $\leftarrow P(A|B)$) +
     geometry($block$)
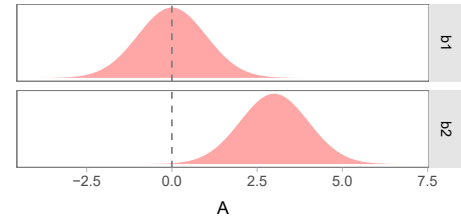
(3) encodings(x position $\leftarrow A$, y position $\leftarrow B$, height $\leftarrow P(A|B)$) +
     geometry($unit$)

These example charts, created in ggplot2, involve much more complex specifications than the hypothetical probabilistic grammar of graphics specifications shown above (each uses a different geometry, required different data preparation, etc). Yet, in a probabilistic grammar, these charts are quite similar. (1) and (2) have the same specification, but differ only because A is a discrete variable in (1) and a continuous variable in (2). Yet, the basic encoding is the same: probability is mapped onto *height*, and a *block* geometry (which could result in bars, densities, violin plots, and perhaps even other chart types) is used. The only difference with (3) is that a *unit* geometry is used: probability is still mapped to height, but an appropriate data representation of the random variable A is prepared in order to function with a *unit* geometry.

The above examples show the potential power of a probabilistic grammar of graphics, even in creating simple charts. Much of the work in this aim will lie in carefully, incrementally extending a grammar like the one above to cover a suitable range of chart types identified in Aim 1.
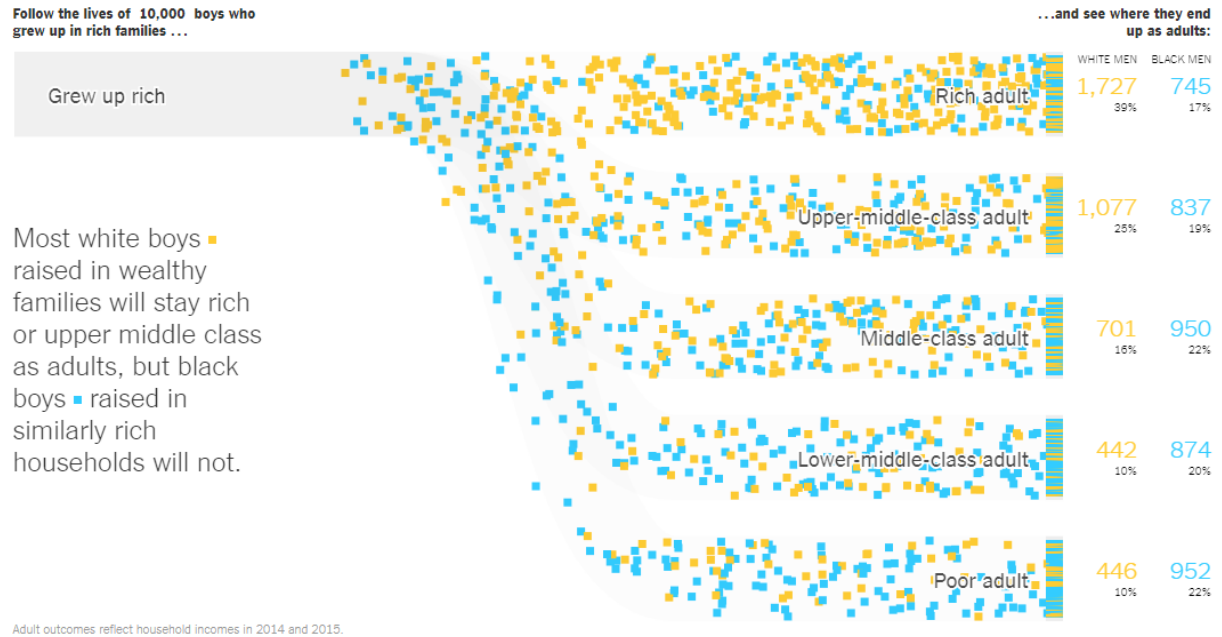
A key challenge lies in specifying animation within this framework. Consider 2, which uses animation to display realizations from one distribution "flowing" into another. This particular "flowing" visual metaphor is a common metaphor for sampling from a probability distribution. Given a graphical model, this maps easily onto a partial specification; something like: $P(race|grew\ up\ rich)$ "flows into" $P(race|income\ class \land grew\ up\ rich)$. Since I will gather a large variety of visualizations, including animated visualizations, it should be possible to ensure that simple constructs like this generalize across this set, or develop more appropriate and general constructs as necessary when they do not.

5   Aim 3: Re-specify existing visualizations in the uncertainty visualization grammar and assess the correctness of those visualizations

One goal of developing this grammar will be to facilitate the construction of more correct uncertainty visualizations. Given the database of visualizations from Aim 1 and the uncertainty visualization grammar developed in Aim 2, I will re-specify as many of the existing visualizations as possible using that grammar. This will aid in answering the following research questions:
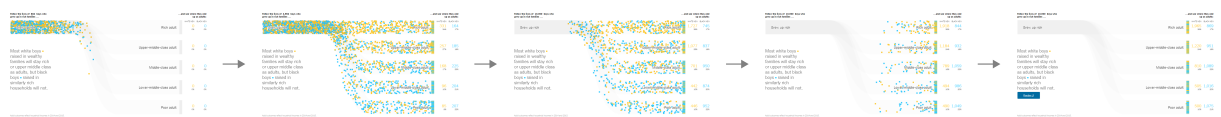
Research Questions: Could a simple, consistent probabilistic grammar of graphics be constructed that easily supports a wide range of modern, effective uncertainty visualizations? What

Figure 2: An example of a professionally-produced probabilistic visualization [14] (https://nyti.ms/2GGpFZw) I would both like to be able to support **and** improve. This visualization falls victim to the *fallacy of the transposed conditional*: by normalizing bar heights on the right-hand side, it depicts $P(race|income\ class \wedge grew\ up\ rich)$ only, so one cannot see $P(income\ class|race \wedge grew\ up\ rich)$: imagine if very few boys of either race who grew up rich ended up in a lower class, then $P(income\ class \neq rich|race = black \wedge grew\ up\ rich)$ would be low even if $P(race = black|income\ class \neq rich \wedge grew\ up\ rich)$ is high. It is worth noting that by reading the numbers in the table to the right it is possible to verify that this is not the case for this particular visualization, but this is difficult to verify from the visualization alone. A probabilistic grammar of graphics would have users explicitly specify the desired distributions, avoiding this error, and would create a similar animation sequence automatically, where this example was custom built in JavaScript.

is the state of existing uncertainty visualizations in the wild: are they correctly communicating desired uncertainties?

Key Challenges: A key aspect of this aim lies in re-expressing existing visualizations in the newly-developed probabilistic grammar of graphics. This may require deconstructing or obtaining datasets from existing visualizations, or simulating datasets in order to recreate similar visualizations, potentially a large undertaking.

Because the grammar is grounded in formal specification of probability distributions, it will be able to be used to determine if a visualization actually communicates a particular uncertainty correctly. While re-specifying each visualization, we will also write down which conditional probabilities the designer appeared to intend to communicate, and then assess whether or not those distributions are correctly communicated in the visualization itself. For example, one might ask: does a visualization based on the joint probability distribution $P(A,B)$ communicate $P(A|B)$? Depending on representation and specification, $P(A|B)$ might not be visible in the visualization even if it is intended to be from the communication goals of the visualization. For example, in Figure 2, the distribution of adult income given race is not directly depicted, despite this being central to the story of the original article the visualization is from [14]. I will re-express as many visualizations as possible within the dataset from Aim 1, and then assess the extent to which errors appear to be made in published, professional visualizations of probability and uncertainty.

6 Aim 4: Evaluate the usability of a probabilistic grammar of graphics

Without a formal and reliable way of reasoning about uncertain data and their visual representation simultaneously, I posit that more errors in uncertainty representation construction will happen. A better formalism should allow the construction of a better tool, which should result in better visualizations. This leads to the following research questions:

Research Questions: Would a probabilistic grammar of graphics allow designers to more easily explore the space of uncertainty visualizations? Would a probabilistic grammar of graphics enable designers to create more correct uncertainty visualizations?

Key Challenges: Evaluating the value of novel, complex toolkits is always a challenge: one cannot recruit users who are experts in a new toolkit, so learning effects are difficult to tease apart from differences in expressiveness or usability when comparing a new tool to an old one. That said, this is a problem largely in a conservative direction: that is, at worst, it should cause us to underestimate the improvement incurred by the new toolkit. In addition, some comparative measures, like differences in the correctness of visualizations produced, are less likely to exhibit this problem.

This aim will consist of two experiments designed to assess the extent to which the developed probabilistic grammar of graphics facilitates the efficient construction of effective uncertainty visualizations. Participants for these experiments will be recruited from current and former Master's students who have taken my Master's-level information visualization class. These students are by-and-large professional Master's students, many of whom have some industry experience in data analysis and/or data visualization roles, making them reasonable proxies for understanding professional users. These participants will also have been trained in the use of the grammar of graphics as that material is taught in the Information Visualization class.

The **first experiment** will focus on the task of recreating existing visualizations. From the database of uncertainty visualizations constructed in Aim 1 and recreated in the probabilistic grammar of graphics in Aim 3, I will select a small number (2–8) of visualizations. Participants

will be shown those visualizations and then asked to recreate those visualizations using either a vanilla grammar of graphics toolkit (ggplot2 or Vega-Lite) or an implementation of the probabilistic grammar of graphics built on top of one of those toolkits (see the Timeline, below).

This experiment will focus on understanding difficulty of use of the grammar. As such, the outcomes measures will be:

- Time on task
- Completion rate (e.g. number of completed visualizations)
- Task load, measured using the NASA TLX (Task Load Index) [24]
- Qualitative assessment of understanding using a think-aloud protocol

Participants will be randomly assigned to one of two conditions: probabilistic or non-probabilistic grammar of graphics (i.e. it will be a between-subjetcs experiment). A pilot study (see Timeline) will be conducted, followed by a Bayesian power analysis [30] to determine the number of subjects for the full study, and the study plan and analysis code will be pre-registered to prevent the inadvertent introduction of researcher degrees of freedom into the final analysis.

The **second experiment** will focus on correctness. Similar to the first experiment, participants will be given uncertainty visualizations to recreate taken from the catalog in Aim 1. However, participants will not be shown the visualizations themselves; rather, the dataset and communication goal will be described, and participants will be asked to recreate it. The recreated visualizations will then be scored on correctness by two researchers. Researchers will be blind to which condition each participant was assigned to.

This latter experiment will help us assess whether or not errors in communicating uncertainty can be mitigated by using a probabilistic grammar of graphics.

7 Educational and Research Integration Plan

I will use the developed uncertainty visualization grammar in my teaching. I regularly teach a 60-student Master's-level class in information visualization; in this course, I already introduce students to the grammar of graphics (through Vega-Lite [49]) and to principles of uncertainty visualization in a separate module. However, these modules are presently somewhat disconnected, and effective strategies for integrating uncertainty visualization into what students have learned about the grammar of graphics are scarce. Developing a probabilistic grammar of graphics is an opportunity to better integrate those aspects of the curriculum.

In addition, I have taught a Doctoral seminar on Bayesian regression and communicating uncertainty from Bayesian models. Woven throughout this course is content on the effective visual communication of uncertainty from statistical models. Currently, this content employs heavy use of tidybayes [m6] and the grammar of graphics, as instantiated within ggplot2 [52]. A probabilistic grammar of graphics implemented on top of ggplot2 would facilitate the pervasive uncertainty visualization content in this course, and help me better train upcoming scientists in the effective communication of the results of their work.

From both of these classes, I will report on the effectiveness of employing a probabilistic grammar of graphics in teaching uncertainty visualization to a venue such as the Pedagogy of Data Visualization Workshop [11].

In addition to the above activities, I will mentor PhD students and Master's students involved in developing and carrying out the technical research work in this proposal. I will also continue providing opportunities for undergraduates and underrepresented groups to participate in research:

I currently am mentoring an undergraduate through the University of Michigan's Undergraduate Research Opportunity Program (UROP), and in the past I have mentored a Computing Research Association Distributed Research Experiences for Undergraduates (DREU) student from an underrepresented group; the latter student contributed substantially to some of my early work on uncertainty visualization [m8]. I regularly mentor professional Master's students, and currently have three Master's students working on projects related to uncertainty visualization.

8 Evaluation and Milestones

Research Evaluation: The proposed research includes quantitative and qualitative evalutions at various stages. We will consider the development of a probabilitistic grammar of graphics successful if we can demonstrate its expressiveness: that is, if we can re-express many complex uncertainty visualizations (including animated uncertainty visualizations) within a single, coherent, and compact notation. This is not only a qualitative assessment (the succinctness and clarity of notation), but also a quantitative one: we will assess what proportion of uncertainty visualizations we have collected are expressable within the framework we develop.

In addition to assessment of expressiveness, quantitative usability assessments will be used to evaluate whether or not the developed notation—as instantiated in an existing toolkit like ggplot2 [52] or Vega-Lite [49]—allow experienced visualization designers to more quickly and *correctly* produce complex uncertainty visualizations. This will be considered a success if we learn that (1) participants are capable of creating uncertainty visualizations more easily, with no loss in quality or (2) if participants are capable of creating complex uncertainty visualizations that are more correct (not falling victim, for example, to the fallacy of the transposed conditional).

Educational Integration Evaluation: To assess in-class content, I will use standard anonymous course evaluations administered at the University of Michigan, which allow for the inclusion of course-specific questions (allowing me to specifically ask about content related to uncertainty visualization). I will compare course evaluations to previous iterations of each class.

9 Timeline of Proposed Work

	Year 1				Year 2				Year 3			
	Fa	Wi	Sp	Su	Fa	Wi	Sp	Su	Fa	Wi	Sp	Su
Aim 1: Catalog visualizations	▓	▓	▓		▓							
Aim 2: Develop grammar		▓	▓	▓	▓	▓	▓					
Aim 3: Re-express vises in grammar			▓	▓	▓	▓	▓	▓	▓			
Aim 4: Evaluate with users						*pilot*			▓	▓	▓	▓
Implementation — *Vega-Lite*			▓		*first version complete*							
						refinement						
ggplot2					▓	▓	▓		▓			
Teaching — *Master's*						▓				▓		
MOOC									▓			
Phd										▓		

The first implementation of the grammar will be created on top of Vega-Lite, and timed to have a first version completed just before I teach the Master's-level visualization class in the beginning of Winter of the second year of the project. This will allow me to teach the grammar to students

(who will already be learning Vega-Lite anyway), and continue refining both the grammar and the implementation. This first teaching run-through will also be an opportunity to pilot studies for Aim 4, giving advance testing and time to work out bugs before testing the system in a controlled study for Aim 4 in the third year of the project. For participants in the evaluation studies, I will recruit students from the Master's course I teach in Year 3, in which students will be taught the grammar of graphics.

10 Broader Impacts of the Proposed Work

This work has several broader impacts:

Societal Outcomes, Toolkits: This work has the potential to make the production of effective uncertainty visualizations easier and more widespread, improving the communication of uncertainty and risk across a range of domains. If toolkits developed in this work are adopted by data journalists, improving their uncertainty communication, the ability of audiences to understand and act upon uncertainty in news reports will be improved. This has wide-reaching implications for data journalism that impacts everyday decision-making, from election outcomes to forecasts of extreme weather events, such as floods or hurricanes. Better uncertainty visualization toolkits integrated into statistical modeling workflows (such as R, ggplot2 [52], and tidybayes [m6]) would raise standards of scientific communication, ultimately improving the way that scientists and laypeople interpret and act upon the results of research across a range of domains.

Publications: Besides instantiating the results of this work in freely-available, open source toolkits, I will publish results in scholarly conferences and journals, such as the InfoVis track of the IEEE VIS Conference, where accepted papers appear in the IEEE journal Transactions in Visualization and Computer Graphics (TVCG). I also publish in the ACM SIGCHI Conference on Human Factors in Computing systems, the premier venue in Human-Computer Interaction. I have a track record of publishing in both venues.

11 Investigator Qualifications

I have substantial experience in the development and controlled evaluation of uncertainty visualizations (*e.g.* [m1, m8, m5, m4, m3, m2]). In addition, I have published two open source R packages [m10, m6] on the Comprehensive R Archive Network (CRAN) [25], which entails following strict package development standards [25, 50], gaining the approval of archive maintainers, and being responsive to package maintenance issues in order to maintain status in the archive. The tidybayes [m6] package, in particular, represents my initial work at integrating Bayesian statistical analysis into a grammar of graphics workflow, and currently garners approximately 1000 downloads per month. My background positions me well both to create a formal description of uncertainty visualization within the grammar of graphics, and to build and deploy toolkits implementing this vision that have the potential to be adopted more widely.

12 Results of Prior NSF Support

Dr. Kay is a PI on NSF grant IIS 1815790: CHS: Small: Collaborative Research: Validating and Communicating Model-based Approaches for Data Visualization Ability Assessment ($238,848; 09/01/2018–08/31/2021). This project uses studies of visualization effectiveness to inform our understanding of the abilities and biases of viewers, both individually and collectively. **Intellectual Merit**: Using a combination of experiments, statistical modeling, and interview studies, this work challenges long-standing assumptions about visualization effectiveness, and will lay a foundation for future experiments that account for differences in visualization reading ability. **Broader Impact**:

This work will support a broader educational goal of using robust statistical modeling techniques in experimentation, through course modules that can be integrated into existing data visualization courses, and through outreach activities that allow individuals to see how well they perform visualization tasks compared to others who have taken the experiments. As this project was recently funded, no publications have yet resulted from it.

References

References to the work of the PI are indicated with an *m*.

References to our work

[m1] M. Fernandes, L. Walls, S. Munson, J. Hullman, and M. Kay. Uncertainty Displays Using Quantile Dotplots or CDFs Improve Transit Decision-Making. In *Conference on Human Factors in Computing Systems - CHI '18*, 2018.

[m2] M. Greis, J. Hullman, M. Correll, M. Kay, and O. Shaer. Designing for uncertainty in hci: When does uncertainty help? In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, pages 593–600, New York, NY, USA, 2017. ACM.

[m3] J. Hullman, M. Kay, Y.-S. Kim, and S. Shrestha. Imagining replications: Graphical prediction & discrete visualizations improve recall & estimation of effect uncertainty. *IEEE transactions on visualization and computer graphics*, 24(1):446–456, 2018.

[m4] J. Hullman, X. Qiao, M. Correll, A. Kale, and M. Kay. In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE transactions on visualization and computer graphics*, 2018.

[m5] A. Kale, F. Nguyen, M. Kay, and J. Hullman. Hypothetical outcome plots help untrained observers judge trends in ambiguous data. *IEEE transactions on visualization and computer graphics*, 2018.

[m6] M. Kay. tidybayes: Tidy data and geoms for Bayesian models. *R package version 1.0.3*, 2018.

[m7] M. Kay and J. Heer. Beyond weber's law: A second look at ranking visualizations of correlation. *IEEE transactions on visualization and computer graphics*, 22(1):469–478, 2016.

[m8] M. Kay, T. Kola, J. R. Hullman, and S. A. Munson. When (ish) is My Bus? User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 5092–5103, 2016.

[m9] M. Kay, G. L. Nelson, and E. B. Hekler. Researcher-Centered Design of Statistics: Why Bayesian Statistics Better Fit the Culture and Incentives of HCI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, pages 4521–4532, New York, New York, USA, 2016. ACM Press.

[m10] M. Kay and J. O. Wobbrock. ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs. *R package version 0.10.4*, 2016.

References

[11] Pedagogy of data visualization. Workshop at IEEE VIS 2016. http://vgl.cs.usfca.edu/pdvw/2016/abstracts.html.

[12] Forecasting the race for the Senate. FiveThirtyEight.com. https://projects.fivethirtyeight.com/2018-midterm-election-forecast/senate, 2018.

[13] J. S. Ancker, Y. Senathrajah, R. Kukafka, and J. B. Starren. Design Features of Graphs in Health Risk Communication : A Systematic Review. *Journal of the American Medical Informatics Association*, 13(6):608–619, 2006.

[14] E. Badger, C. C. Miller, A. Pearce, and K. Quealy. Extensive data shows punishing reach of racism for black boys. *New York Times*, 19, 2018.

[15] S. Belia, F. Fidler, J. Williams, and G. Cumming. Researchers misunderstand confidence intervals and standard error bars. *Psychological methods*, 10(4):389–96, dec 2005.

[16] J. Bertin, W. J. Berg, and H. Wainer. *Semiology of graphics: diagrams, networks, maps*, volume 1. University of Wisconsin press Madison, 1983.

[17] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76(1), 2017.

[18] M. Correll and M. Gleicher. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE transactions on visualization and computer graphics*, 20(12):2142–2151, 2014.

[19] M. Correll, D. Moritz, and J. Heer. Value-Suppressing Uncertainty Palettes. *Conference on Human Factors in Computing Systems - CHI '18*, 2018.

[20] J. Cox, D. House, and M. Lindell. Visualizing Uncertainty in Predicted Hurricane Tracks. *International Journal for Uncertainty Quantification*, 3(2):143–156, 2013.

[21] R. Garcia-Retamero and E. T. Cokely. Communicating Health Risks With Visual Aids. *Current Directions in Psychological Science*, 22(5):392–399, 2013.

[22] G. Gigerenzer and U. Hoffrage. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102(4):684–704, 1995.

[23] Gregor Aisch, Nate Cohn, Amanda Cox, Josh Katz, and Adam Pearce. Live Presidential Forecast. *New York Times*, Nov. 2016.

[24] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[25] K. Hornik. The comprehensive R archive network. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(4):394–398, 2012.

[26] J. Hullman, P. Resnick, and E. Adar. Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences about Reliability of Variable Ordering. *PLOS ONE*, 10(11):e0142444, Nov. 2015.

[27] H. Ibrekk and M. Morgan. Graphical Communication of Uncertain Quantities to Nontechnical People. *Risk Analysis*, 7(4):519–529, 1987.

[28] B. Johnson and B. Shneiderman. Tree-maps: A space-filling approach to the visualization of hierarchical information structures. In *Visualization, 1991. Visualization'91, Proceedings., IEEE Conference on*, pages 284–291. IEEE, 1991.

[29] A. Kirk. Best of the visualization web, March 2018. http://www.visualisingdata.com/2018/06/best-visualisation-web-march-2018/.

[30] J. K. Kruschke and T. M. Liddell. The Bayesian New Statistics: Hypothesis testing , estimation, meta-analysis, and planning from a Bayesian perspective. *Psychonomic bulletin & review*, pages 1–53, 2016.

[31] L. Liu, A. Boone, I. Ruginski, L. Padilla, M. Hegarty, S. Creem-Regehr, W. Thompson, C. Yuksel, and D. House. Uncertainty Visualization by Representative Sampling from Prediction Ensembles. *IEEE Transactions on Visualization and Computer Graphics*, (99):1, 2016.

[32] L. R. Lucchesi and C. K. Wikle. Visualizing uncertainty in areal data with bivariate choropleth maps, map pixelation and glyph rotation. *Stat*, 2017.

[33] D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The bugs project: Evolution, critique and future directions. *Statistics in medicine*, 28(25):3049–3067, 2009.

[34] D. J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter. Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, 10(4):325–337, 2000.

[35] A. M. MacEachren. Visualizing uncertain information. *Cartographic Perspective*, 13(13):10–19, 1992.

[36] A. M. Maceachren, A. Robinson, and S. Hopper. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, 32(3):139–160, 2005.

[37] J. Mackinlay. Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)*, 5(2):110–141, 1986.

[38] J. Mackinlay, P. Hanrahan, and C. Stolte. Show me: Automatic presentation for visual analysis. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1137–1144, 2007.

[39] M. Mirzargar, R. T. Whitaker, and R. M. Kirby. Curve boxplot: Generalization of boxplot for ensembles of curves. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2654–2663, 2014.

[40] D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe, and J. Heer. Formalizing visualization design knowledge as constraints: Actionable and extensible models in draco. *IEEE transactions on visualization and computer graphics*, 2018.

[41] N. National Hurricane Center. Definition of the nhc track forecast cone, nov. 2018. http://www.nhc.noaa.gov/aboutcone.shtml.

[42] A. O'Hagan, C. . E. Buck, A. . Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow. *Uncertain Judgements: Eliciting Experts' Probabilities*. John Wiley & Sons, Ltd., 2006.

[43] L. M. Padilla, G. Hansen, I. T. Ruginski, H. S. Kramer, W. B. Thompson, and S. H. Creem-Regehr. The influence of different graphical displays on nonexpert decision making under uncertainty. *Journal of Experimental Psychology: Applied*, 21(1):37–46, 2015.

[44] L. M. Padilla, I. T. Ruginski, and S. H. Creem-Regehr. Effects of ensemble and summary displays on interpretations of geospatial uncertainty data. *Cognitive Research: Principles and Implications*, 2(1):40, 2017.

[45] D. Park, S. M. Drucker, R. Fernandez, and N. Elmqvist. ATOM: A Grammar for Unit Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 2626(c), 2017.

[46] M. Plummer et al. Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124. Vienna, Austria, 2003.

[47] E. Rifkin and E. Bouwer. *The illusion of certainty: Health benefits and risks*. Springer, 2007.

[48] I. T. Ruginski, A. P. Boone, L. M. Padilla, L. Liu, N. Heydari, H. S. Kramer, M. Hegarty, W. B. Thompson, D. H. House, and S. H. Creem-Regehr. Non-expert interpretations of hurricane forecast uncertainty visualizations. *Spatial Cognition & Computation*, 16(2):154–172, 2016.

[49] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer. Vega-Lite: A Grammar of Interactive Graphics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):341–350, 2017.

[50] R. C. Team. Writing r extensions. *R Foundation for Statistical Computing*, 1999.

[51] H. Wickham. A layered grammar of graphics. *Journal of Computational and Graphical Statistics*, 19(1):3–28, 2010.

[52] H. Wickham. *ggplot2: elegant graphics for data analysis*. Springer, 2016.

[53] H. Wickham and H. Hofmann. Product Plots. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2223–2230, Dec. 2011.

[54] L. Wilkinson. Dot Plots. *The American Statistician*, feb 1999.

[55] L. Wilkinson. *The Grammar of Graphics*. Statistics and Computing. Springer-Verlag, New York, 2005.