Research statement

1 APPROACH

I work in the fields of human-computer interaction and information visualization, and focus on problems related to **communicating complex data and concepts to end-users,** particularly **communicating uncertainty.** I conduct this work across many application domains, from personal health informatics (including sleep sensing and weight management) to everyday sensing and prediction contexts (like weather forecasting and transit arrival time prediction). I am driven by a thirst for problem areas where existing solutions baffle users: every new confusion surrounding how a system *is* carries fresh insight into how it *should be*.

My approach to research is multidisciplinary and informed by my background not only in computer science, but also visual arts and information visualization. I combine a strong aesthetic design sense with an engineering and computer science background, building systems that are well-engineered, useful, and informed by research, but also clear and aesthetically pleasing. I draw upon methods in **human–computer interaction, visual design, information visualization, perception, and statistics**. I build and deploy novel user-centered computing systems to answer my specific research questions. When working across domains, I often collaborate with experts, as in my work on understanding sleep quality [J06,C04,C05,A05] and weight change [C06]. I am keenly interested in work that has impact on real users, motivating my primary research program in uncertainty visualization for lay people [J04,J07,C10,C15], but also side projects, such as my work on gender representation in image search [C08].

I believe impact beyond publications is important. I have released research software such as PVT-Touch [A05,J06,C05], the ARTool R package [R01], and the tidybayes R package [R02] that have been downloaded by thousands of other researchers. This experience has also lead me to **value code as an avenue for research impact**: my tidybayes package, which is designed for visualizing results from Bayesian analyses in R, has allowed me to disseminate my work on uncertainty visualization to practicing scientists working in diverse fields (including psychology, ecology, and finance). I have given invited talks on uncertainty visualization to practitioner-oriented audiences, like Tapestry (http://youtu.be/E1kSnWvqCwo) and OpenVisConf (http://youtu.be/vqzO-9LS0G4). Where permitted by Human Subjects oversight, I have released data and analyses from studies as GitHub repositories with R code (archived on long-term repositories, like zenodo.org), aiding reproducibility and meta-analysis. Altogether, these practices broaden the impact of my research.

Much of my work falls into two broad categories: building and evaluating better techniques for **communicating uncertainty** and developing tools and methods for **usable statistics**.

2 COMMUNICATING UNCERTAINTY

People are exposed to sensing and prediction on a daily basis (*How many steps did I take today? How long until my bus shows up? Will it rain today?*). Uncertainty is inherent to these systems and usually poorly communicated—but it need not be. I employ methods from human–computer interaction to elicit users' needs related to uncertainty, then design and evaluate novel uncertainty visualizations to support those needs.

2.1 Uncertainty visualization for non-experts

A core part of my work in uncertainty visualization is in **developing uncertainty displays for non-experts which are grounded in literature on human perception and statistical reasoning**. I also ground my work in real-world decision contexts, as the ultimate goal of uncertainty communication is to help people make better decisions in the real world.

A project that exemplifies my approach is a series of multi-methods studies I conducted on uncertainty visualization for transit arrival prediction [C10,C15]. This included in-person interviews of real users at bus stops, paper prototyping, online surveys, and an incentivized decision-making experiment. During this work I developed a novel uncertainty visualization type, the **quantile dotplot** (Fig. 1). Quantile dotplots are a discrete outcome uncertainty visualization—that is, they depict probabilities through discrete outcomes rather than continuous probabilities—an approach that is inspired by work in medical risk communication suggesting that people reason better about probability when it is depicted as discrete possible outcomes (Ancker *et al.*, 2006). This project included two online experiments to assess the effectiveness of quantile dotplots alongside other common uncertainty visualizations, such as intervals, densities, cumulative distribution functions (CDFs), and displays without uncertainty at all. The first study, published at CHI 2016 [C10], found that quantile dotplots can make people's estimates of probabilities more precise, perhaps through the use of subitizing (Choo & Franconeri, 2014), people's ability to quickly and accurately count small numbers of dots.

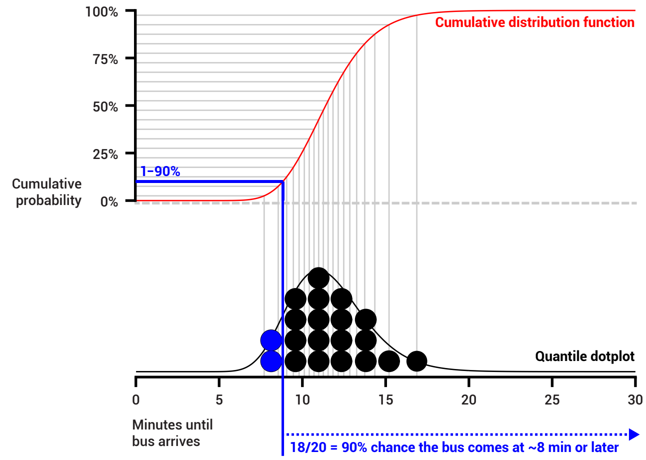


Fig. 1. Construction of a *quantile dotplot*, an uncertainty visualization technique I developed as an alternative to probability density plots and cumulative distribution functions. This technique allows fast and accurate estimation of intervals through counting [C10] and yielded higher decision quality compared to other approaches in an incentivized decision-making experiment [C15]. This example depicts a probabilistic prediction of bus arrival time and shows how one can count hypothetical busses to determine the chance that the bus arrives after any point on the timeline: for example, if one arrived at the bus stop 8 minutes from now, one would miss 2/20 hypothetical buses, leaving an 18/20 = 90% chance of making the bus.

However, an important problem in uncertainty visualization is that perceptual effectiveness rarely tells the whole story: **just because people can extract the probability value they need to make the best decision does not mean they know how to use that probability to make the best decision**. Therefore, I followed up my initial study with an incentivized experiment in which I measured decision quality: in simulated bus-catching scenarios, participants were given uncertainty displays of predicted arrival times and an incentive structure (e.g., a few cents per minute for staying out home out of the rain and a bonus for reaching their destination on time), then asked when they would go to the bus stop. We compared people's performance (the expected payoff given their choices) to a baseline (the expected payoff given an optimal strategy). An important aspect of this study that I was instrumental in developing was the use of a measure of decision quality to assess people's performance on different visualization types; I developed the decision quality metric we used and analyzed the final results of the study. Quantile dotplots yielded the best quality decisions, and people were also able to improve over the course of many trials. Importantly, variance in performance was also lower with quantile dotplots: even the worst performers did well, a crucial quality for uncertainty visualizations for lay audiences. In comparison, typical uncertainty visualizations (like densities and intervals) had little improvement. This work was published at CHI 2018 [C15] and received a Best Paper Honorable Mention.

In other work, I have investigated systems for tracking and communicating sleep quality (Best Paper Award at UbiComp 2012 [C04]), people's perceptions of error in weight scales (Best Paper Award at UbiComp 2013 [C06]), and patient-centered ways of communicating results (and their uncertainty) in self-experimentation systems for health [J05]. I also collaborate regularly with Jessica Hullman on uncertainty visualization work under the banner of our co-directed cross-institutional lab, the Midwest Uncertainty Collective (http://mucollective.co). In addition to papers listed above, I have collaborated on several papers with Jessica and her students, on which I have been involved in helping form research questions, shape study designs, and advise on statistical analyses [J04,J07,C17].

3 USABLE STATISTICS

My ongoing work in communicating uncertainty has strongly affected how I have come to communicate results of my own work—using Bayesian estimation with a focus on understanding effect sizes, trying more faithfully to reflect the uncertainty of the scientific process. In addition to adopting these approaches, I have advocated for them through a paper at InfoVis 2015 (Best Paper Honorable Mention [J02]) and at CHI 2016 (Best Paper [C11]), and by co-organizing several workshops and special interest groups on transparent statistical communication in human–computer interaction [W03,W04,W06]. I strongly believe that **improved methods and practices in statistics can be achieved by treating statistical practice as a user-centered design problem**.

3.1 Tools to support reliable data exploration and statistical reporting

In collaboration with researchers at Inria, I have been exploring interactive techniques that build on the idea of multiverse analysis. To address bias and overfitting introduced into an analysis when researchers conduct data exploration but report only a single model, *multiverse analysis* (Steegen *et al.*, 2016) suggests that scientists should report a set of reasonable analyses rather than a single analysis. However, these sets can be large (100s or 1000s of different combinations of data transformation and modeling decisions), so such a report is difficult to convey within the confines of a traditional research paper. Building on the notion of explorable explanations (Victor 2011), we developed *explorable multiverse analysis reports*, interactive papers for reporting multiverse analyses (CHI 2019 Best Paper [C19]). Apart from helping develop the general concept of EMARs, one of my primary contributions to this work was a technique to allow readers of a paper to interactively specify their own priors on a Bayesian analysis. The naïve approach—refitting the model with a new prior as the user interact with the paper contents—is a time-consuming procedure that would not be possible in an interactive setting for models fit using Markov chain Monte Carlo, as many Bayesian models are. Instead, I developed an approach in which a small number of similar models (typically ~8–12) can be pre-fit and then used to drive interactive prior exploration by using a mixture of model posteriors weighted according to their marginal likelihoods. This allows readers to tweak the prior for a focal parameter in a model (e.g., the mean difference between two treatments) in order to see how sensitive the paper's conclusions are to the prior (Fig. 2), which helps address an important limitation of traditional Bayesian analyses: the author's prior may not be the reader's prior. Currently, I am working with a former student (Abhraneel Sarma, now a PhD student at Northwestern) in developing an R package to make it easy for other researchers to build EMARs.

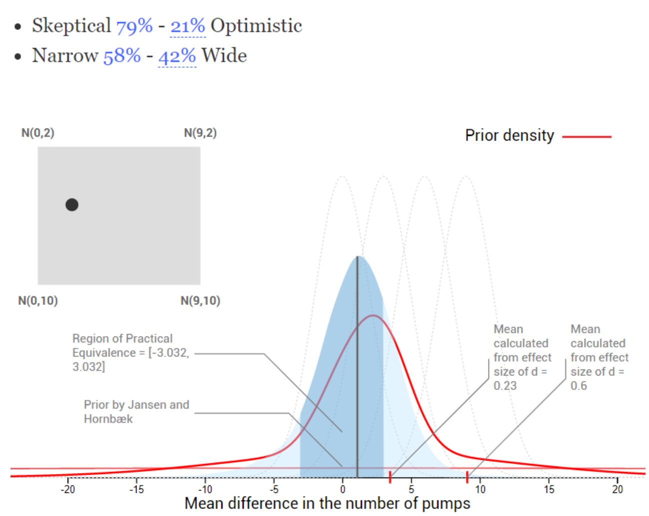


Fig. 2. Excerpt from an interactive *explorable multiverse analysis report* (EMAR), an interactive academic paper in which the reader can adjust aspects of the analysis. Here, the reader can change the prior on a Bayesian analysis and see the effects of this change in real time (the gray box in the upper left and the text at the top can be directly manipulated to change the prior) [C19]. An interactive example of such a report is available here: https://explorablemultiverse.github.io/examples/prior/. Examples of other types of EMARs are available here: https://explorablemultiverse.github.io/

3.2 Tools to help novices build Bayesian models

I believe Bayesian analysis allows researchers to shift to a more nuanced view of practical effect sizes and an accumulation of knowledge. However, it remains locked in something of a specialist mode, as Bayesian models are typically specified using probabilistic programming languages like JAGS (http://mcmc-jags.sourceforge.net/) and Stan (http://mc-stan.org/). A better understanding of researchers' needs *and* the underlying analyses is required for more user-friendly tools to emerge. To help address this problem, I helped develop analysis templates for researchers familiar with common statistical analyses (e.g., ANOVAs) to translate their analyses into Bayesian models [C18]. Through a qualitative study in which users translated previous analyses into Bayesian analyses, we found that the templates helped researchers translate their analyses into a Bayesian approach, but that challenges remain: in particular, users unfamiliar with Bayesian statistic found it difficult to set priors for their models, an area I am tackling in future work (below).

4 FUTURE WORK

The core of my work lies in **building visualizations (and particularly uncertainty visualizations) that lay people can actually understand**. This thrust involves not only studying the principles and properties of effective uncertainty visualizations, but also building tools to help visualization designers adopt more effective uncertainty visualization techniques. In addition, to correctly assess whether these visualizations are usable by wide audiences, I am **developing methods for assessing visualization literacy that are grounded in human perception**. Finally, I am **developing tools for usable, reliable statistical analysis that researchers can actually understand.**

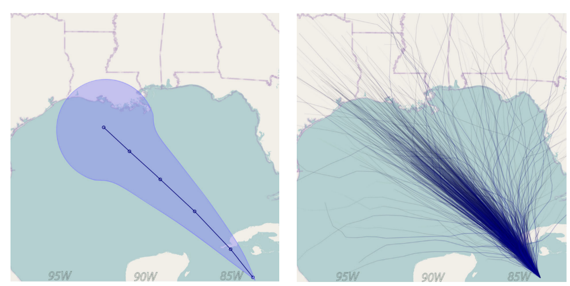
4.1 A probabilistic grammar of graphics

When reporting on high-stakes and uncertain topics ranging from elections to natural disasters, journalists routinely employ uncertainty visualizations in an attempt to help the public answer important questions: *Who will win the next presidential election? Should I evacuate in the face of potential flooding?* Apart from some specialized data journalism outfits—like FiveThirtyEight and the New York Times' Upshot—most journalists adopt conventional but problematic uncertainty displays. For example, journalists use *cones of uncertainty* to illustrate hurricane path predictions, despite evidence that uncertainty cone visualizations are misinterpreted by the public, and the fact that known-better alternative uncertainty visualizations exist (Padilla *et al.*, 2017). This is not surprising: constructing sophisticated and effective uncertainty visualizations—such as quantile dotplots [C10,C15] or animated hypothetical outcome plots (Hullman *et al.*, 2015; [J07])—is more technically difficult than constructing common but less effective uncertainty visualizations, like intervals.

Inspired by the success of probabilistic programming languages, I am building a **probabilistic grammar of graphics** that integrates uncertainty into visualization specification as a first-class object. By formalizing the specification of uncertainty visualizations, it will make it easier to explore the design space of uncertainty visualizations and to conduct systematic research into the effectiveness of different uncertainty visualizations in different contexts, thereby giving a

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

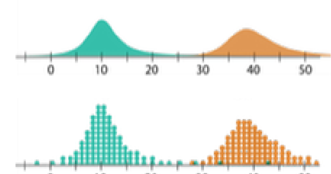


Fig. 3. Examples of several uncertainty visualizations one might like to support in a probabilistic grammar of graphics.

formal grounding to my other work on uncertainty visualization. I received an NSF Small to pursue this work this year, and one of my PhD students (Xiaoying Pu) has already developed a draft grammar integrating conditional probability specifications into the grammar of graphics. As the project evolves, we will evaluate the potential of the grammar to make it easier for visualization designers to quickly construct uncertainty visualizations. We will also use the grammar to assess uncertainty visualization *correctness*, as one issue with uncertainty visualization is that it is often easy for designers to construct subtly incorrect uncertainty visualizations. Ultimately, I hope to automate or semi-automate the construction of effective uncertainty visualizations given a dataset, task, and context.

4.2 Measuring visualization literacy using psychophysical models

As much of my research is concerned with effective communication to lay audiences, I am particularly concerned with the measurement of visualization literacy as a way of understanding (and explaining) variance in visualization interpretation ability. In collaboration with Lane Harrison (at Worcester Polytechnic Institute), I was awarded an NSF Small to develop measures of visualization literacy based on individuals' performance in visualization perception tasks. As part of this grant, I we have begun constructing hierarchical Bayesian models of individuals' performance on an extensive set of visualization perception tasks, and plan to use these models to develop a smaller, more manageable set of tasks that are highly predictive of visualization literacy as measured using the full set of tasks. This will yield a validated, perceptually-grounded method for quickly measuring visualization literacy.

4.3 Statistically reliable exploratory visual analytics tools

A related issue to multiverse analysis reporting is that of bias introduced to analyses caused by the use of exploratory visual analytics tools, which may allow an analyst to wander a *garden of forking paths* of possible analyses and inadvertently take exploratory findings as if they are confirmatory (introducing bias to their results). My PhD student Xiaoying Pu and I published a proposal for a design space for the creation of more reliable visual analytics tools [C16]. The key insight of this work is that the problems caused by exploratory visual analytics are analogous to overfitting in statistical analysis and machine learning: analysts, in exploring a dataset, are implicitly developing models in their minds of the underlying data, and may overfit if they slice the data too finely. Given that, techniques used to address overfitting in statistics and machine learning (e.g., regularization) could be introduced into exploratory visual analytics tools to help address overfitting there. Imagine, for example, exploratory tools in which increasing regularization is applied to the data the more the analyst explores, gradually smoothing out any signal the more they slice the dataset. I am excited to explore questions about how this regularization should be done, how it should be surfaced to the user, how it can be designed to integrate into analysts' workflows, and whether it can help analysts make more reliable inferences during exploratory data analysis.

4.4 User-centered tools for Bayesian modeling

Following my work on Bayesian analysis templates and interactive prior exploration, I plan to build user-centered statistical tools to help researchers conduct Bayesian analysis, with a particular focus on prior-setting, sensitivity analysis, and model checking. This will build upon my own expertise in information visualization and the R statistical programming language (e.g., authoring the tidybayes R package: [R02]). This work will instantiate known best practices from the statistical literature, such as the principled Bayesian workflow advocated by Gabry *et al.* (2017). I will build accessible, interactive R-based suite of tools for Bayesian modeling, including a strong visualization and educational component to help researchers understand the modeling choices they make, set reasonable priors for models, and interpret the results of their analyses. While building upon my work in *communicating* uncertainty, this work will also advance the state of the art in interfaces for *inputting* uncertainty, for example, through developing interactive visualizations for *prior elicitation* and *model checking*, building on the techniques I developed for *explorable multiverse analysis reports*.

5 REFERENCES

Alphanumeric references (e.g. [C09], [J04]) refer to my work and can be found on my CV. Other references are below.

Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: a systematic review. *J Am Med Inform Assoc*, 13(6), 608–618.

Choo, H., & Franconeri, S. L. (2014). Enumeration of small collections violates Weber's law. Psychonomic Bulletin & Review, 21(1), 93–99.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2017). Visualization in Bayesian workflow. arXiv preprint. https://arxiv.org/abs/1709.01449v5

Hullman, J., Resnick, P., & Adar, E. (2015). Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences about Reliability of Variable Ordering. PloS One, 10(11).

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

Silver, N. (2012). The Weatherman Is Not a Moron. The New York Times. Retrieved from http://www.nytimes.com/2012/09/09/magazine/the-weatherman-is-not-a-moron.html

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. Perspectives on Psychological Science, 11(5), 702–712.

Victor, B. (2011). Explorable Explanations. http://worrydream.com/ExplorableExplanations/