

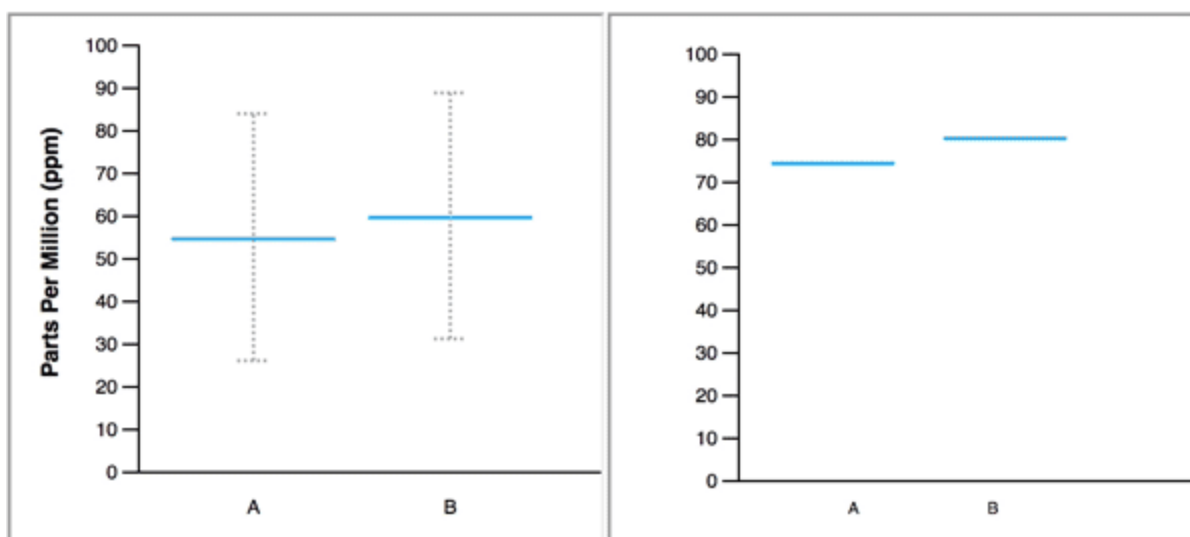
Hypothetical Outcome Plots: Experiencing the Uncertain

 medium.com/hci-design-at-uw/hypothetical-outcomes-plots-experiencing-the-uncertain-b9ea60d7c740

January 26, 2016

Top highlight

If you are like most people, including many data analysts, interpreting visualizations of uncertainty feels hard and abstract. This article describes Hypothetical Outcome Plots (HOPs), a promising approach to visualizing uncertain data for general audiences and analysts alike. Rather than showing a continuous probability distribution, HOPs visualize a set of draws from a distribution, where each draw is shown as a new plot in either a small multiples or animated form. HOPs enable a user to experience uncertainty in terms of countable events, just like we experience probability in our day to day lives.



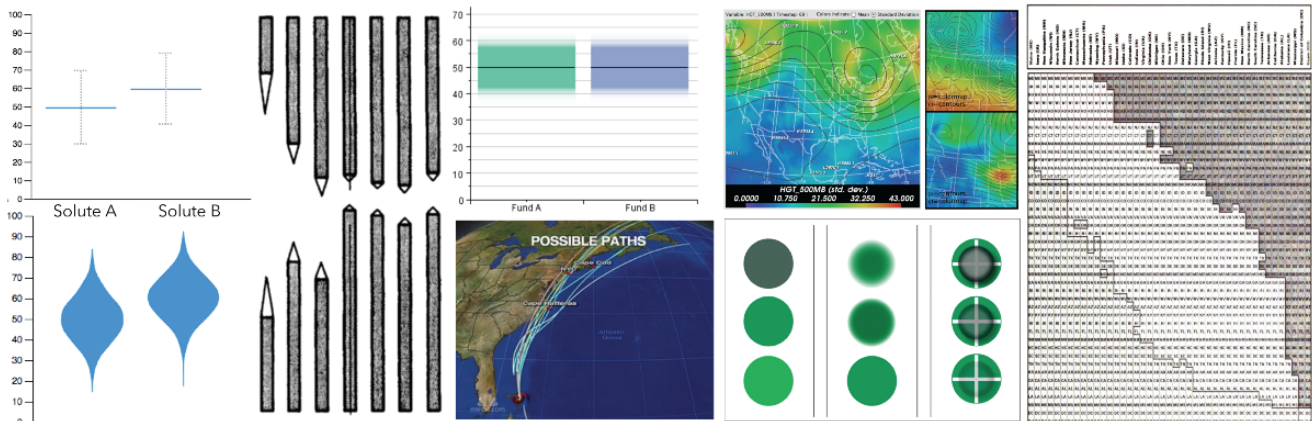
How likely is it that B will be greater than A if more draws are taken? Error bars vs. HOPs

A brief demonstration

Let's start with an example. The two figures to the left show the same data set, a set of measured concentrations of two chemicals in many water samples. The figure to the left shows the mean concentration of each chemical with an interval expected to contain 95% of future samples of A and B that are collected. The figure to the right is an animated hypothetical outcomes presentation where each frame is a draw from the distribution of A and B. Try estimating how likely it is that B will have a greater concentration than A in future water samples, first with the error bars visualization, and then with HOPs (Answer given below). With HOPs, you can estimate the *reliability* of $B > A$ by counting how many frames show $B > A$, but this information is not accessible in the error bars. Notice also how the values of A and B move together: the chemicals are correlated. From HOPs, we can infer variable dependencies as well as probabilities.

Background

Visualizing uncertainty in data — in the form of variance, precision, accuracy, reliability or related concepts — has a relatively long history. Francis Galton visualized a hypothetical distribution of heights back in 1869; visualizations were used in China in the earlier 1800's to show predictions related to children's health. Researchers and practitioners have since proposed many techniques.



Techniques for visualizing uncertainty (clockwise from top left): error bars, , , , a , , possible hurricane paths on a news weather forecast, violin plots.

However, many visualizations that we encounter don't show uncertainty at all. According to a future agenda for research and development of visualization tools: "There is no accepted methodology to represent potentially erroneous information ... There is no agreement on factors regarding the nature of uncertainty, quality of source, and relevance to a particular decision or assessment". How can visualizing uncertainty still be a largely unsolved topic?

Description versus experience

Interpreting visualizations of uncertainty may feel difficult and abstract. Why is this? For starters, most uncertainty presentations provide a static description of a probabilistic process. For example, error bars might convey a 95% confidence interval, standard error, or the standard deviation of a random variable. But contrast this for a moment with how you face uncertainty — the potential for multiple possible outcomes — in your day-to-day life. You want to eat at a restaurant but expect there might be a wait on certain nights. You know that the time it takes you to commute home can vary. You perceive sequences of events, and you build a sense of expectancy — or non-expectancy — specific to different types of events. Often, you don't have to think much to do this. Intuition lets us take in a complex uncertain scenario and output a decision quickly and often with confidence.

Intuition is not, however, magical. It arises from a stream of past experiences, with similarities to new situations in which we need to act. Properties of the new situation cue accumulated wisdom of how different actions have played out in the past, and suddenly we know what to do.

Learning from repeated experience is at the heart of statistics, especially frequentist statistics where models with known probabilities are used to make statements that directly quantify the uncertainty of events. As the mathematician Laplace said, “the theory of probabilities is basically just common sense reduced to calculus; it makes one appreciate with exactness that which accurate minds feel with a sort of instinct, often without being able to account for it.”

A 95% confidence interval depicts a range of values such that if repeated samples were taken and the confidence interval computed for each sample, 95% of these ranges would contain the true value, or population mean. A variable might be created to capture the probability that an observed datum is accurate. Both cases refer to a hypothetical experience in which some data or a summary of data is repeatedly generated and recorded. But the end-user is not involved in this experience.

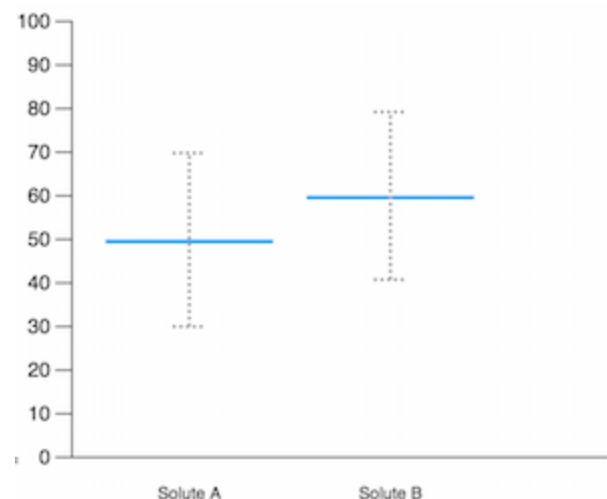
Being the subject of the “experiencing” can make all the difference. Having direct experience of an uncertain process can lead to different decisions compared to getting only a description of that process.

Probability problems, and the power of frequency

We want to give users a way to experience probability, but there is a fundamental disconnect between the abstract, numerical probability calculated by statistics and the way we experience uncertainty. This disconnect has even led some statisticians to conclude that “probability does not exist.” So how do we expect people to reason about probability in a useful way?

“When called upon to judge probability, people actually judge something else and believe they have judged probability” — Daniel Kahneman, 2011

When asked directly, most people *can* answer questions that ask for probabilities. But they often do so using *heuristics*, a form of intuition that provides a mental shortcut for hard decisions. Heuristics work by substituting a complex decision involving multiple parameters and uncertainties with an easier decision. Imagine viewing a chart (like the one to the left) displaying the amount of two chemicals in water samples with error bars depicting 95% predictive intervals (intervals expected to contain 95% of the values collected in many future samples). Can you answer the question *How likely is it that B will be greater than A if many more samples are gathered?*



How likely is it that B will be greater than A if many more draws are taken?

We can avoid the complicated task of incorporating the variance in our estimate by using a simpler cue — say, the difference between the means — to guess how reliable the $B > A$ pattern is. Big differences are reliable, small differences are unreliable. This heuristic works when the data resembles the prior information on which it is based, but results in many more errors in cases that are different, such as small but reliable differences.

Flawed reasoning has been reproduced across many decision-making experiments demonstrating heuristics in human reasoning. However, these problems can be greatly reduced by framing the data differently. In particular, researchers have found that people of all backgrounds make better decisions when probabilities are presented as : counts of successes (or failures) given some total number of events. For instance, one might present the above comparison using a frequency statement that describes the probability ($Pr(B>A)=0.75$) in terms of counts (“In 3 out of 4 cases, $B>A$ ”).

But can an uncertainty visualization really make it possible for users to experience the uncertainty through concrete, countable outcomes? Yes!

Hypothetical Outcome Plots

Hypothetical outcome plots consist of multiple individual plots (frames), each of which depicts one draw from a distribution. In its simplest form, we simply: (1) Draw a sample of hypothetical outcomes (draws) from a distribution; (2) for each, make a plot that becomes one frame in an animated or small multiples presentation.

For example, the visualization to the left uses HOPs to depict the same data shown in the previous error bars visualization. Now try estimating how likely it is that B will be greater than A in new samples. The visualization below depicts the number of sunny days in two different cities using an abstract representation of the probability density functions (PDFs) next to a HOPs visualization. Imagine that you are considering a February vacation, and need to decide whether Los Angeles or Orlando is a better bet for warm, sunny days. What’s the likelihood that Los Angeles will have more sunny days in a week? *(See below for answers to both questions!)*

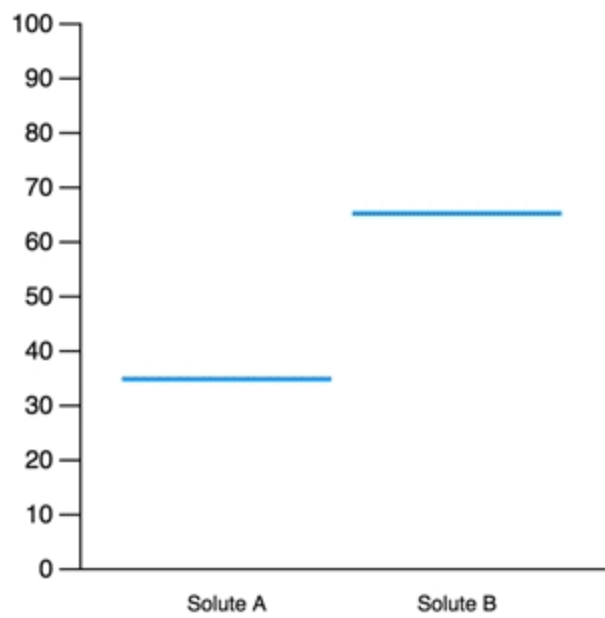
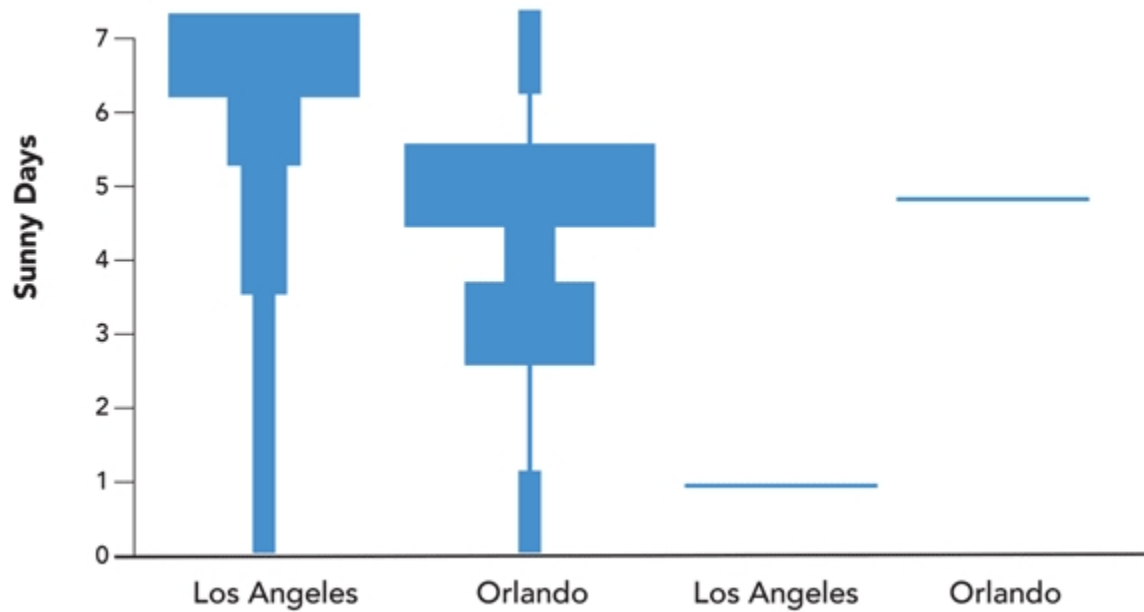


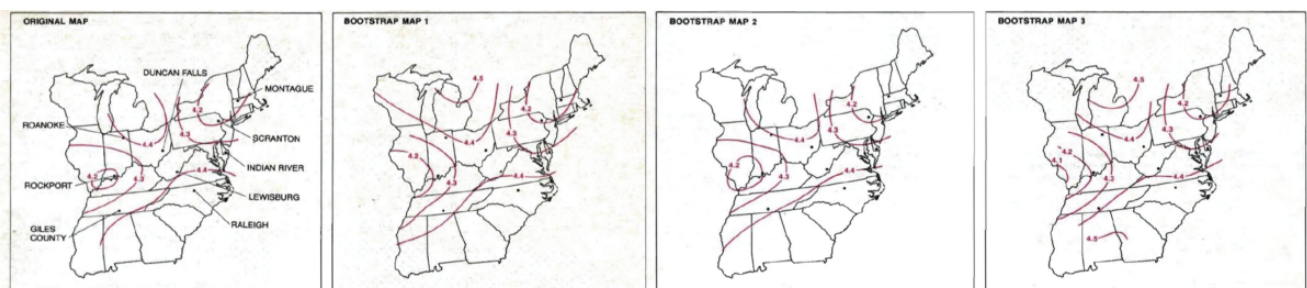
Image for post

How likely is it that B will be greater than A if many more draws are taken?



Which location is likely to have more sunny days in a week?

The idea of presenting multiple possible outcomes to convey uncertainty is not entirely new. In early work on bootstrapping (a computational method for estimating parameters from data), Efron and Diaconis included maps showing predicted rainfall levels for multiple resampled data inputs. Within statistics education, simulation has been used to convey concepts like sampling distributions and confidence intervals. The “dance of the p-values” is a graphical simulation of the limits of significance testing.

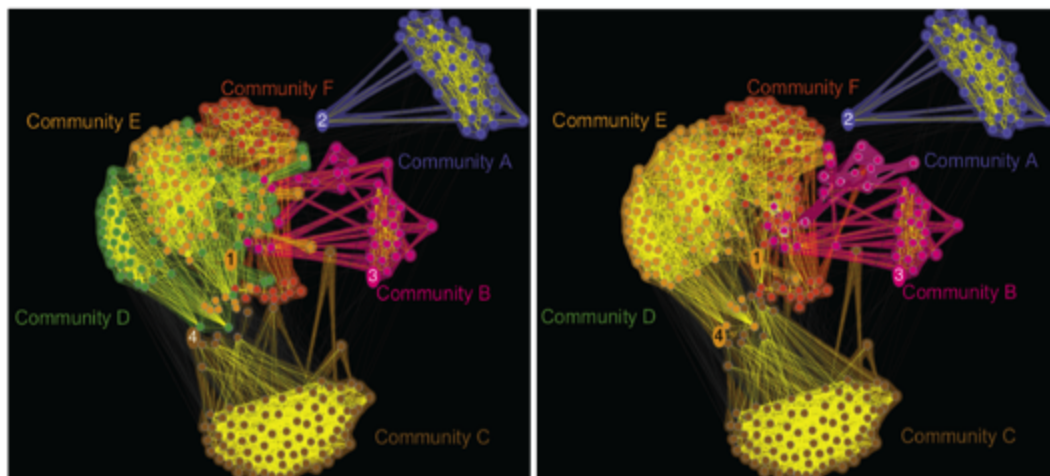


Simulating outcomes gives the user a more concrete way to think about probability distributions or statistical constructs. Instead of using simulated outcomes as a stepping stone to help people understand static visualizations, we think that in many cases HOPs are a good *substitute* for these visualizations, outside of just educational or complex modeling settings.

Making HOPs work

Simulating draws: There are two steps to creating effective HOPs. We start with a set of observed data. More specifically, we need some notion of the distribution that produced our observed data so that we can make draws. How we arrive at this distribution is flexible. We can use non-parametric bootstrapping, i.e., resampling with replacement from our original

data set to create a large number of replicates that we assume to be sampled from the same underlying distribution as the original data. Or, we can use our observed data to infer a model, such as a Gaussian with a certain mean and standard deviation, then take draws from the model. The bootstrapping literature is full of techniques, both frequentist and Bayesian, for a number of data types. Regardless of what we choose, our goal is to generate *representative* hypothetical outcomes, those that really could have resulted from the same process that produced our observed data.



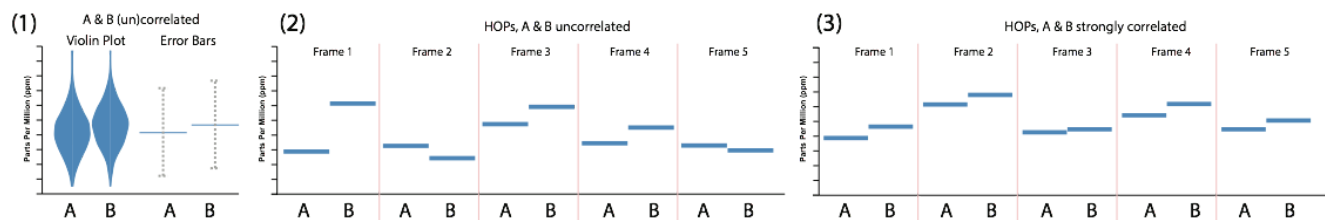
Two frames from an animated HOP of a clustered social network. Note how node 4 is in Community C on the left, and Community E on the right.

Supporting comparison through visual stability: For HOPs to communicate uncertainty, the user needs to easily compare outcomes across frames. As designers, we need to ensure *visual stability*: we should construct the individual frames so that the visual encodings (e.g., color mappings, axes range, layouts) stay consistent. For example, in animated HOPs like those shown as alternatives to error bars, any given *y*-axis position must stand for the same value across all frames. This is trivial with many common statistical plots, but can require clever solutions when generating HOPs using optimization-based visualization approaches. For example, in a dynamic network visualization, the placement of nodes can vary over slightly different input networks because the layout is driven by the clustering of nodes, which changes as the link structure changes. To fix node position across frames in a HOPs presentation of the clustered graph we use information about the entire set of hypothetical clustered graph outcomes to find the best global layout.

Why HOPs work

Of course, there are a few drawbacks to HOPs. Dynamically presenting draws introduces *sampling error*: the user views a finite number of draws, sacrificing precision compared to viewing a model that summarizes a very large or even infinite number of draws. The user also has to combine information across multiple frames. This can feel difficult, and without summary marks added to show properties like the mean, users can have a hard time inferring such information when variance is high.

Despite the drawbacks, the potential advantages of HOPs are exciting and far ranging. Unlike most uncertainty visualizations, HOPs do not require the addition of new visual variables (such as blur) or graphical annotations. This is a powerful property that makes HOPs generalizable to different data inputs and visual encodings. In our work, we've applied HOPs to network diagrams showing community structure, maps visualizing rainfall or household radon levels, treemaps of hierarchical data, and choropleth maps depicting election outcomes. Further, in contrast to error bars, violin plots, and most static visualizations of uncertainty, HOPs naturally convey properties of the *joint distributions* of multiple variables, as shown in the first demonstration of HOPs above. Note that for both independent and correlated variables, the error bars and violin plots would be identical.



HOPs look different for independent versus correlated variables, but violin plots and error bars can't express joint probabilities.

HOPs support mechanistic processes like counting, while still allowing for less deliberative “gisting” in which the user simply watches the animated outcomes, relying on their visual system to draw out patterns. When you make judgments using HOPs like the animated examples in this post, do find yourself counting events of interest, or watching and intuiting? Our observations of how people use HOPs make us think that people may be combining counting and perceptual approximation. For example, a viewer might count the number of frames in which B is higher than A for a short time, then estimate the total frames they viewed to guess the probability. Is one strategy superior, and can we design HOPs to prompt certain strategies?

Our work so far leads us to believe that HOPs are especially useful for helping non-expert users reason about uncertainty. In experiments we've run, users recruited through Amazon's Mechanical Turk can use HOPs without training to make most common judgments of single variable distributions as well as they can with error bars or violin plots. When it comes to multivariate judgments like how reliable a perceived difference between two random variables is, untrained HOPs users are far, far more accurate than those using the standard representations. Perhaps because of these properties, the New York Times and other designers have recently used dynamic HOPs-like presentations to illustrate uncertainty in employment projections and political elections. Our current research aims to provide design principles for optimizing the parameters of HOPs for these and other applications. For example, what is the best frame rate for presenting draws, and how does it interact with properties of the underlying distribution (such as the “rareness” of the pattern of interest)? Research in how well a person's visual system can do at extracting information from moving targets outside of focused attention (ensemble processing) informs our work. By adding

summary marks and interactive features to HOPs we can create dynamic visualizations that show the same data properties as static visualizations while expressing other properties that otherwise go unnoticed.

Want to learn more about HOPs? Check out [our research on HOPs](#) or our upcoming talk at the [2016 OpenVis conference](#).

This post was authored by Jessica Hullman. Her collaborators on HOPs include Paul Resnick and Eytan Adar.

Answers

A brief demonstration: $P(B>A) = 0.95$ (95%)

$P(B>A) = 0.75$ (75%)

$P(\text{more sunny days in Los Angeles than Orlando}) = 0.55$ (55%)