Visualizing Uncertainty

Alex Kale
Visualizing uncertainty

"Uncertainty must be displayed if it is to be reasoned with and incorporated into the visual analytics process. In existing visualizations, much of the information is displayed as if it were ‘true.’”
(Thomas and Cook 2005)

Why should we show uncertainty?

• Guard against overconfidence (i.e., ignoring uncertainty)
  • Insensitivity to sample size
  • Not acknowledging assumptions

• Visualizations often used for inferences & decisions
  • People may infer different amounts of uncertainty when omitted
Uncertainty: “the possibility of multiple outcomes”
Terminology

Precision
Precision

Accuracy
Precision

Terminology

Accuracy

Variance

Bias
Pang, Wittenbrink, Lodha. Approaches to Uncertainty Visualization.

Figure 1: This visualization pipeline shows the introduction of uncertainty (U) from models and measurements, uncertainty from transformation processes, and uncertainty from the visualization process itself.
Example: Quantifying Uncertainty

Imagine that we took many measurements of the concentration of a solute in seawater.

Data (in *parts per billion*): [132.24, 131.72, 129.93, 128.60, 129.16, 127.76, 134.45, 127.03, 133.20, 133.64, 129.34, 131.96, 131.55, 131.76, 120.39, 127.49, ... ]

Q: How much do these readings vary?
Q: How much do these readings vary?
Example: Quantifying Uncertainty

Imagine that we took many measurements of the concentration of a solute in seawater.

Data (in parts per billion): [132.24, 131.72, 129.93, 128.60, 129.16, 127.76, 134.45, 127.03, 133.20, 133.64, 129.34, 131.96, 131.55, 131.76, 120.39, 127.49, ... ]

Q: How do our predictions of the level in the entire ocean vary?
Q: How do our predictions of the level in the entire ocean vary?
Visualization Approaches

Glyphs and summary plots
- Error bars
- Boxplots

Visual variables
- Lightness/value, area
- Blur/focus, “sketchiness”

Set of discrete outcomes
- Beeswarm plot, dotplot
- Hypothetical outcome plots
Visualization Approaches

What is “sketchiness”? 

Boukhelifa et al. (2012), Wood, Jo et al. (2012)
Visualization Approaches

Summaries vs discrete outcomes

Cox et al. (2013)
Visualization Approaches

Hypothetical Outcome Plots

Present uncertainty as samples over time, where each sample is a new frame in an animated visualization

Hullman et al. (2015)
Visualization Approaches

Hypothetical Outcome Plots

If those numbers are correct, sampling error could produce this debate instead...

...and leave these candidates out.

Where Boys Outperform Girls in Math: Rich, White and Suburban Districts

Applicable to arbitrary encodings
Animating Uncertainty is Controversial

Bad?

Better?
Framing Probability as Frequency

<table>
<thead>
<tr>
<th>M</th>
<th>Tu</th>
<th>W</th>
<th>Th</th>
<th>F</th>
</tr>
</thead>
</table>

Frequency framing improves reasoning
4 out of 5 times, vs 80%
(Gigerenzer & Hoffrage 1995)
Visualization Approaches

Quantile Dotplots

Kay et al. (2016)
Bus riders made more accurate probability estimates with 20 dot QDPs than with 100 dot QDPs or densities. Better quality decisions with 50 and 20 dot QDPs vs densities, intervals, or text.

Kay et al. (2016), Fernandes et al. (2019)
Evaluating Uncertainty Visualizations
“The advantage for uncertainty estimates depends critically on how they are expressed. It is crucial that the expression is compatible with both the decision task and cognitive processes of the user.”

Joslyn & LeCleric (2013)
Does the vis support the user’s task?

Is the model of uncertainty appropriate to answer the user’s question?

What is the level in the entire ocean?

vs

How much do the readings vary?
Does the vis support the user’s task?

Can the user extract the needed information from the visualization?

How confident are you that $\mu_B > \mu_A$?

What is probability that $B > A$?

Are $A$ and $B$ correlated?
Is the vis expressive & can users infer meaning?

**Expressiveness**: How well does the visualization present all the facts (and only the facts)?
Mackinlay (1986)
What is the information format?

Does the visualization present data in a way that aligns with human reasoning abilities?
Can users decode the data from the chart?

**Effectiveness**: How easily can a human estimate the values from the visualization?
Mackinlay (1986)

<table>
<thead>
<tr>
<th>Quantitative</th>
<th>Ordinal</th>
<th>Nominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Position</td>
<td>Position</td>
</tr>
<tr>
<td>Length</td>
<td>Density</td>
<td>Hue</td>
</tr>
<tr>
<td>Angle</td>
<td>Saturation</td>
<td>Texture</td>
</tr>
<tr>
<td>Slope</td>
<td>Hue</td>
<td>Connection</td>
</tr>
<tr>
<td>Area</td>
<td>Textures</td>
<td>Containment</td>
</tr>
<tr>
<td>Volume</td>
<td>Connection</td>
<td>Density</td>
</tr>
<tr>
<td>Density</td>
<td>Containment</td>
<td>Saturation</td>
</tr>
<tr>
<td>Saturation</td>
<td>Length</td>
<td>Length</td>
</tr>
<tr>
<td>Hue</td>
<td>Angle</td>
<td>Angle</td>
</tr>
<tr>
<td>Texture</td>
<td>Slope</td>
<td>Slope</td>
</tr>
<tr>
<td>Connection</td>
<td>Area</td>
<td>Area</td>
</tr>
<tr>
<td>Containment</td>
<td>Volume</td>
<td>Volume</td>
</tr>
<tr>
<td>Shape</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Presenting Uncertainty When Position Encoding is Occupied
Juxtaposition

Data Map

Uncertainty Map
Superposition

Bivariate Map
Value-Suppressing Uncertainty Palettes

Correl et al. (2018)
Value-Suppressing Uncertainty Palettes

Kay (in press)
Presenting Uncertainty for Multivariate Data
Multivariate Data

**Example:** MR spectroscopy

Measure the resonance frequencies of molecules in the brain.

**Voxels:** spatial resolution of about a cubic centimeter

Model concentrations of different molecules.

Data format: One row per voxel; One column per molecule.
Parallel Coordinates

One line per voxel (spatial unit)
Parallel Coordinates for Distributions

Show multivariate distribution for each voxel
Step 1: Interpolation

**Input:** Two distributions

\[ x_1 = \text{N}(\mu_1, \sigma_1^2) \]
\[ x_2 = \text{N}(\mu_2, \sigma_2^2) \]

**Output:** \( \mu \) and \( \sigma^2 \) across horizontal axis \( a \in [0, 1] \)

\[ \mu_a = (1 - a)\mu_1 + a\mu_2 \]
\[ \sigma_a^2 = (1-a)^2\sigma_1^2 + a^2\sigma_2^2 \]
Step 2: Calculate Probability Density

**Input**: $\mu_a$ and $\sigma_a^2$

**Output**: Probability density over horizontal span $a$ and vertical span $b$

$$PC(a, b) = \frac{1}{\sqrt{2\pi} \sigma_a^2} \exp\left[-\frac{(b-\mu_a)^2}{2\sigma_a^2}\right]$$
Emphasizing Means

Fig. 6

Parallel coordinates plots of four MR spectroscopy metabolites. From left to right, means are increasingly emphasized. Left: a sigmoidal PC plot of the same data shown in Figure 4 with two additional variables (glutamate and n-acetylaspartate). Center: the estimated PDF mapped into PC space, with means emphasized according to their uncertainty. Right: direct visualization of the PDF.
Implications for Brush Selection

- Compute integral within range of selection
- Define threshold (e.g., 95%) to determine what lines are selected
Limited for Correlated Variables

\[ \rho = .9 \quad \rho = 0 \quad \rho = -.9 \]
Thanks!

Alex Kale

http://students.washington.edu/kalea

@AlexKale17

kalealex

Slides adapted from Jessica Hullman & Michael Correl