Uncertainty Visualization

Michael Correll  Tableau Research
Questions To Answer

What Does Uncertainty Mean?

How Should I Visualize It?

What Can Go Wrong?
WHAT DOES UNCERTAINTY MEAN, ANYWAY?
Things “Uncertainty” Can Mean

Doubt
Risk
Variability
Error
Lack of Knowledge
Hedging
...

...
Measurement Uncertainty

Sales of Widgets for Stores A and B

Widgets Sold

A

B

0 10 20 30 40 50 60 70 80

6
Forecast Uncertainty

Sales of Widgets for Quarters 1 and 2

Widgets Sold

<table>
<thead>
<tr>
<th>Quarter</th>
<th>Widgets Sold</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>70</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>
Model Uncertainty

Sales of Widgets for Quarters 1 and 2

Widgets Sold

Quarter

1
2
3

50
60
70
80
Decision Uncertainty

We Should Close Store A?

Widgets Sold

A

B
Uncertainty Vis Pipeline

Uncertainty Sources

**Measurement Uncertainty**: “We’re not sure what the data are”

**Forecast Uncertainty**: “We’re not sure what will happen to the data next”

**Model Uncertainty**: “We’re not sure how the data fit together”

**Decision Uncertainty**: “We’re not sure what to do with the data”
Measurement Uncertainty

[Graph showing data trends over time for different categories: Consumer, Corporate, and Home/Office. The x-axis represents order dates from November 2013 to November 2017, and the y-axis represents monetary values ranging from $0 to $60,000.]
Model Uncertainty

[Graph showing data for Consumer, Corporate, and Home Office over time]
Forecast Uncertainty
Uncertainty Visualization

There are different types and sources of uncertainty.

We can quantify or model our uncertainty.

The visual presentation of uncertainty can clash with cognitive and perceptual biases.
Terminology

Type I error
Type II error
Precision
Bias
Should I Bring an Umbrella?
Type I and II Errors

- Type I Error (false positive)
- Type II Error (false negative)

- True Positive
- True Negative
- False Positive
- False Negative
The Boy Who Cried Wolf
Did My Arrows Hit the Target?
Precision & Bias

Precision
Precision & Bias

Precision
Precision & Bias

Precision
Precision & Bias

Precision

Accuracy
Precision & Bias

**Precision**

**Accuracy**
Precision & Bias

Precision

Accuracy
Precision & Bias

Precision

Accuracy

Variability

Bias
What Does Uncertainty Mean?

Any one of a number of potentially interconnected quantitative, qualitative, or factors that affect the quality, reliability, or utility of your data or data-driven decisions. Anything that can cause you to be unsure about your data or how to use it.
What Does Uncertainty Mean?

Any one of a number of potentially interconnected quantitative, qualitative, or factors that affect the quality, reliability, or utility of your data or data-driven decisions. Anything that can cause you to be unsure about your data or how to use it.
Uncertainty Maps and Model Visualization

HOW SHOULD I VISUALIZE UNCERTAINTY?
Uncertainty Visualization Zoo

Intervals

### Intervals

<table>
<thead>
<tr>
<th></th>
<th>Bus Timeline</th>
<th>Route Timeline</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Density</strong></td>
<td><img src="image" alt="Density" /></td>
<td><img src="image" alt="Density" /></td>
</tr>
<tr>
<td><strong>Dotplot (20)</strong></td>
<td><img src="image" alt="Dotplot 20" /></td>
<td><img src="image" alt="Dotplot 20" /></td>
</tr>
<tr>
<td><strong>Dotplot (100)</strong></td>
<td><img src="image" alt="Dotplot 100" /></td>
<td><img src="image" alt="Dotplot 100" /></td>
</tr>
<tr>
<td><strong>Stripeplot (50)</strong></td>
<td><img src="image" alt="Stripeplot 50" /></td>
<td><img src="image" alt="Stripeplot 50" /></td>
</tr>
</tbody>
</table>

Hypothetical Outcome Plots
Missing Values

(a) Data Absent  (b) Color Points  (c) Color Points & Line Gradients  (d) Connected Error Bars  (e) Disconnected Error Bars  (f) Unfilled Points  (g) Unfilled Points & Line Gradients

Missing Values

Visualizations with High Data Quality

Visualizations with Low Data Quality

Special Case: Implicit Uncertainty
Special Case: Implicit Uncertainty
Special Case: Implicit Uncertainty
Uncertainty Vis Pipeline

1) Quantify Uncertainty
2) Choose a free visual variable
3) Encode uncertainty with the variable
SNAP

Data Map

Uncertainty Map
Uncertainty Vis Pipeline

1) Quantify Uncertainty
2) Choose a free visual variable
3) Encode uncertainty with the variable
Uncertainty Vis Pipeline

1) Quantify Uncertainty
2) Choose a free visual variable
3) Encode uncertainty with the variable
4) Unify the Data Map and Uncertainty Map
How to Unify?

Data Map

Uncertainty Map
Juxtaposition

Data Map

Uncertainty Map
Superposition
Superposition

Bivariate Map
Superposition

Uncertainty Vis Pipeline

1) Quantify Uncertainty
2) Choose a free visual variable
3) Encode uncertainty with the variable
4) Unify the Data Map and Uncertainty Map
Uncertainty Vis Pipeline

1) Quantify Uncertainty
2) Choose a free **visual variable**
3) Encode uncertainty with the variable
4) Unify the Data Map and Uncertainty Map
Semiotics of Uncertainty

Ceci n’est pas une pipe.
The Variable Matters!

![Chart showing 2016 Senate Seats with a significant majority in blue for Republicans (R), fewer in red for Democrats (D), and very few in gray for Independents (I).]
The Variable Matters!

2016 Senate Seats

R  D  I
Semiotics of Uncertainty
Semiotics of Uncertainty

Series #1: General Uncertainty by Visual Variable
Fuzziness Juxtaposition
Fuzziness Superposition
Size Juxtaposition
Size Superposition
“Sketchiness”


“Sketchiness”


Encoding Uncertainty

Some visual variables (like fuzziness and value) have a **semiotic connection** to uncertainty.

However, intuitive variables may not always be accurately interpreted!
Model Visualization
Polling Data

I am sorry that we didn't poll all 63 million Trump voters SUSAN

SUSAN @Sue4the5
Repeating to @Amy_Siskind @pppolls
"survey of 572 registered voters" This is a sample of 63 million voters who support Trump? What a crock of shit.

8:06 AM - 1 Nov 2017

1,373 Retweets 6,231 Likes
The NYT Needle
News Will Be Flashed from the Tower of The Times Building on Tuesday Night.

The results of the election next Tuesday night will be flashed by electric light from the tower of the Times Building, so that the mob around people will be able to tell which of the candidates they favor.

To provide the aid which the voters can use them easily and comfortably, a telegraph machine will be installed in the basement of the Times Building and the bulletin displayed on signals stretched from the north side of the building. There will be a similar service at the Market office of The Times, 135 West 12th Street.

The electric signals from the tower of the Times Building will be flashed from a point 300 feet above the street level. A light will be used to indicate the vote of each candidate and the letters will be displayed in white light. The letters M, W, B, and V will indicate where McCullin has been elected, a steady light in the east will indicate Jerome's election, and a steady light in the north will indicate Hearst's election.

Jerome's election will be indicated by a steady light in the west. A light in the north will indicate Osborn's election. A light in the south, moving from east to west, will indicate Sturtevant's election.
Election Bulletins
BY BOMBS.
TUESDAY NIGHT
THE TRIBUNE
will send up from the roof of the
GREAT NORTHERN HOTEL
hourly, shells containing blue and red stars—
effectively on the hour—at 7, 8, 9, 10, 11 p.m.
12 midnight, 1 and 2 a.m. Wednesday morning,
unless election is decided earlier, in which
cause twelve bombs will be sent up in rapid suc-
cession. Blue to indicate McKinley’s election.
Red to indicate Bryan’s election.
SIX BOMBS EVERY HOUR.
The first bomb sent up, if blue, indicates the
returns in COOK COUNTY as that hour are
favorable to McKinley; if red, unfavorable to Bryan.
After sixty seconds two bombs will be sent up
in rapid succession, and will indicate, if blue,
that returns from ILLINOIS favor McKinley;
if red, Bryan.
After sixty seconds more three bombs will be
sent up in rapid succession, and if blue will in-
dicate that at that hour returns from the entire
country favor McKinley; if red, Bryan. Each
bomb bursts high in the air, scattering a shower
of stars.
Polling Data

Candidate A is ahead of Candidate B in the polls, with 55% of the likely voters*
Polling Data

Candidate A is ahead of Candidate B in the polls, with 55% of the likely voters*

*poll of 100 people, margin of error +/-5
Monte Carlo Approach

Candidate A is ahead of Candidate B in the polls, with 55% of the likely voters*

*poll of 100 people, margin of error +/-5
A Likely Voter
Poll
Actual Election?
Actual Election?
Actual Election?
Actual Election?
Pangloss Plot

Candidate A is ahead of Candidate B in the polls, with 55% of the likely voters*

*poll of 100 people, margin of error +/-5
Biden is *favored* to win the election

We simulate the election 40,000 times to see who wins most often. The sample of 100 outcomes below gives you a good idea of the range of scenarios our model thinks is possible.

- **Trump wins**: 10 in 100
- **Biden wins**: 89 in 100
- **TIE**: 1

**Note**: Don’t count the underdog out! Upset wins are surprising but not impossible.
Biden is *favored* to win the election

We simulate the election 40,000 times to see who wins most often. The sample of 100 outcomes below gives you a good idea of the range of scenarios our model thinks is possible.

- **Trump wins**
  - 10 in 100

- **Biden wins**
  - 89 in 100

*Don’t count the underdogs out! Upset wins are surprising but not impossible.*
Model Visualization

Building models is necessary to quantify uncertainty

It is important to communicate the variability in model outcomes

Dynamic or ensemble displays can help communicate complex models
How Should I Visualize Uncertainty?

Choose an appropriate visual variable based on the domain, literacy, and expertise of your audience. Be mindful that any display of uncertainty inherently increases the complexity of your visualization, and that there is a preference/performance gap.
How Should I Visualize Uncertainty?

Choose an appropriate visual variable based on the domain, literacy, and expertise of your audience. Be mindful that any display of uncertainty inherently increases the complexity of your visualization, and that there is a preference/performance gap.

IT DEPENDS
WHAT CAN GO WRONG WHEN VISUALIZING UNCERTAINTY?

Cognitive and Perceptual Biases and Disfluencies
### Forecast for Seattle, WA

<table>
<thead>
<tr>
<th>Fri Nov 30</th>
<th>Sat Dec 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Daytime High</strong></td>
<td><strong>Daytime High</strong></td>
</tr>
<tr>
<td>44°F</td>
<td>44°F</td>
</tr>
<tr>
<td>38°F</td>
<td>39°F</td>
</tr>
<tr>
<td>36°F</td>
<td>34°F</td>
</tr>
<tr>
<td>30°F</td>
<td>33°F</td>
</tr>
<tr>
<td><strong>Nighttime Low</strong></td>
<td><strong>Nighttime Low</strong></td>
</tr>
<tr>
<td>41°F</td>
<td>39°F</td>
</tr>
<tr>
<td>33°F</td>
<td>36°F</td>
</tr>
</tbody>
</table>
“The high tomorrow will be 44, and the low will be 38”
Deterministic Construal Error

Probabilistic data is misinterpreted as being deterministic.
Cone of Doom

Cone of Doom

Spaghetti/Ensemble Plots
Spaghetti/Ensemble Plots

Things That Can Wrong

People Confuse Uncertainty with Certainty
Which Stock To Buy?

Company A

Company B
Neither!
What Swag Should We Send?

Fake Insights

Wu Wei
Pareidolia
Have People Made Up Their Mind About Obama?
Lineups Protocol

Buja et al. Statistical inference for exploratory data analysis and model diagnostics.
Royal Society, 2009.

Not everyone here is innocent.
Lineups Protocol

Lineups Protocol!

Distance vs. angle for 3 point shots by the LA Lakers. One plot is the real data. The others are generated according to a null hypothesis of quadratic relationship.

Distance vs. angle for 3 point shots by the LA Lakers. One plot is the real data. The others are generated according to a null hypothesis of quadratic relationship.

Negative Results

People tend to analyze patterns and make decisions, even if there is “nothing to see.”

Negative or null results can correspond to weak and non-robust visual patterns across a model space.
Things That Can Go Wrong

People Confuse Uncertainty with Certainty

People Confuse Signal with Noise
Base Rate Fallacy

1% of the villagers are werewolves

80% of werewolves are allergic to silver.

10% of innocent villagers are allergic to silver.

If a villager is allergic to silver, what’s the probability they are a werewolf?
Bayes’ Law

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]
Bayes’ Law

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

\[ P(\text{🐺} | +\text{Test}) = \frac{P(+\text{Test}|\text{🐺})P(\text{🐺})}{P(+\text{Test})} \]
Bayes’ Law

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

\[ P(\text{Wolf} | +\text{Test}) = \frac{P(+\text{Test}|\text{Wolf})P(\text{Wolf})}{P(+\text{Test})} \]

\[ P(+) = P(+ \wedge \text{Wolf})P(\text{Wolf}) + P(+ \wedge \neg \text{Wolf})P(\neg \text{Wolf}) \]
Bayes’ Law

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

\[ P(\text{🐺} | +\text{Test}) = \frac{P(+\text{Test}|\text{🐺})P(\text{🐺})}{P(+\text{Test})} \]

\[ P(+) = P(+ \wedge \text{🐺})P(\text{🐺}) + P(+ \wedge \sim \text{🐺})P(\sim \text{🐺}) \]

\[ P(+) = 0.01 \times 0.8 + 0.99 \times 0.1 \]

\[ P(+) = 0.107 \]

\[ P(\text{🐺} | +) = \frac{0.8 \times 0.01}{0.107} \approx \boxed{0.075} \]
Problems

People are bad at this.

People who should be good at this are bad at it.

How you present the problem affects how bad people are at it.
How To Present Probabilities

<table>
<thead>
<tr>
<th>Less Intuitive</th>
<th>Probability</th>
<th>P(A) = 0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage</td>
<td>60% chance of A</td>
<td></td>
</tr>
<tr>
<td>Natural Frequency</td>
<td>3 out of 5 times, A happens.</td>
<td></td>
</tr>
</tbody>
</table>

Quantile Dot Plots

Base Rate Fallacy

Pangloss Dot Plot?

52% of a poll of 50 likely voters support Candidate A. Margin of error +/- 5%.

This chart shows 50 possible elections, given this poll result.
Things That Can Wrong

People Confuse Uncertainty with Certainty

People Confuse Signal with Noise

People Confuse Probabilities with ???
What Can Go Wrong?

Uncertainty can be difficult to understand and require a statistical background and high numeracy. Additionally, cognitive and perceptual biases can result in people making poor or error-prone decisions from uncertain data.
What Can Go Wrong?

Uncertainty can be difficult to understand and require a statistical background and high numeracy. Additionally, cognitive and perceptual biases can result in people making poor or error-prone decisions from uncertain data.
Questions To Answer

What Does Uncertainty Mean?

How Should I Visualize It?

What Can Go Wrong?
Questions To Answer

What Does Uncertainty Mean?
LOTS OF THINGS

How Should I Visualize It?
IT DEPENDS

What Can Go Wrong?
A LOT
Wrap Up

Uncertainty can happen at all stages of the analysis process, from data collection to final decision-making.
Wrap Up

Variables like blur and transparency can be intuitive for showing uncertainty, but hard to decode.
Wrap Up

Consider using discrete samples to show variation and uncertainty in a model
Wrap Up

Consider when uncertainty is high enough that doing *nothing* is the right thing to do.
Topics I Didn’t Cover

Uncertainty Quantification

Uncertainty Visualization Evaluation

Visualization Verification

… lots more
Questions?

Michael Correll  Tableau Research