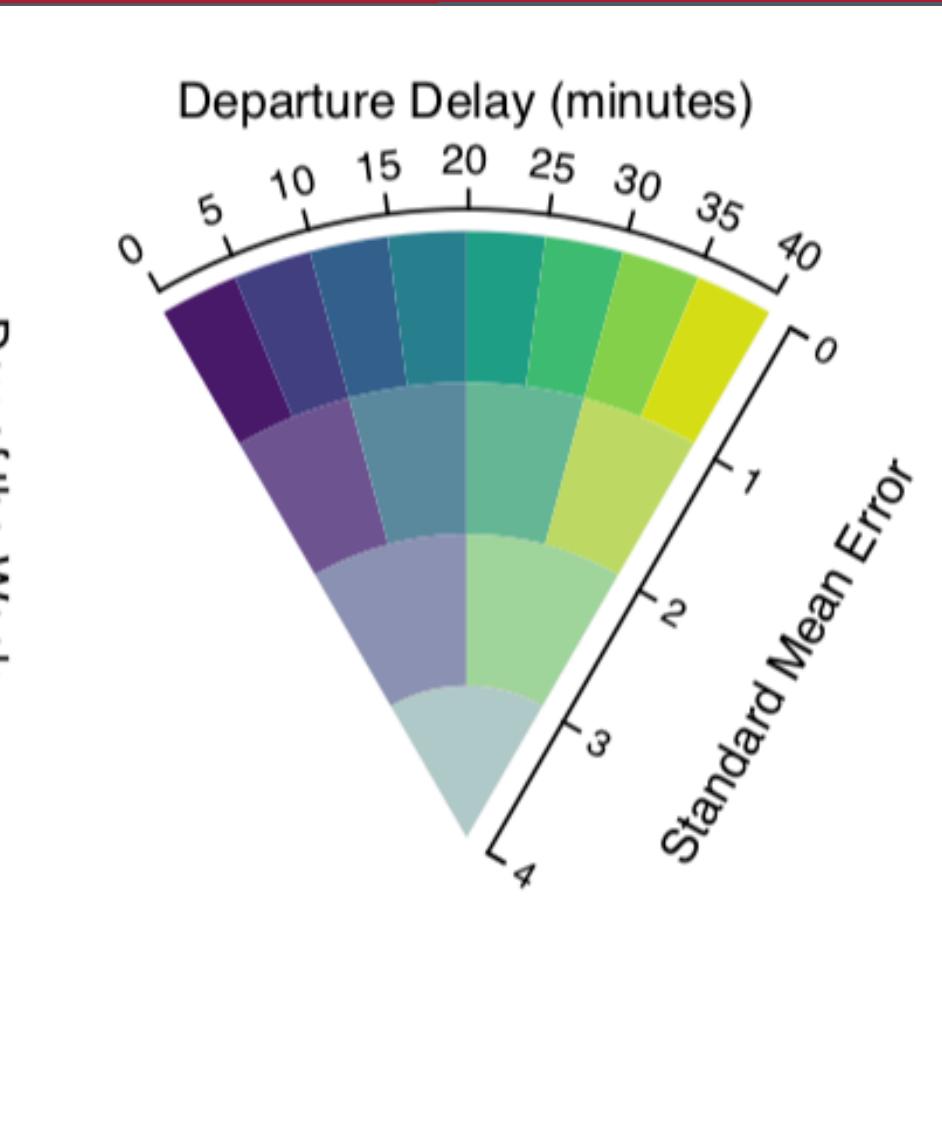
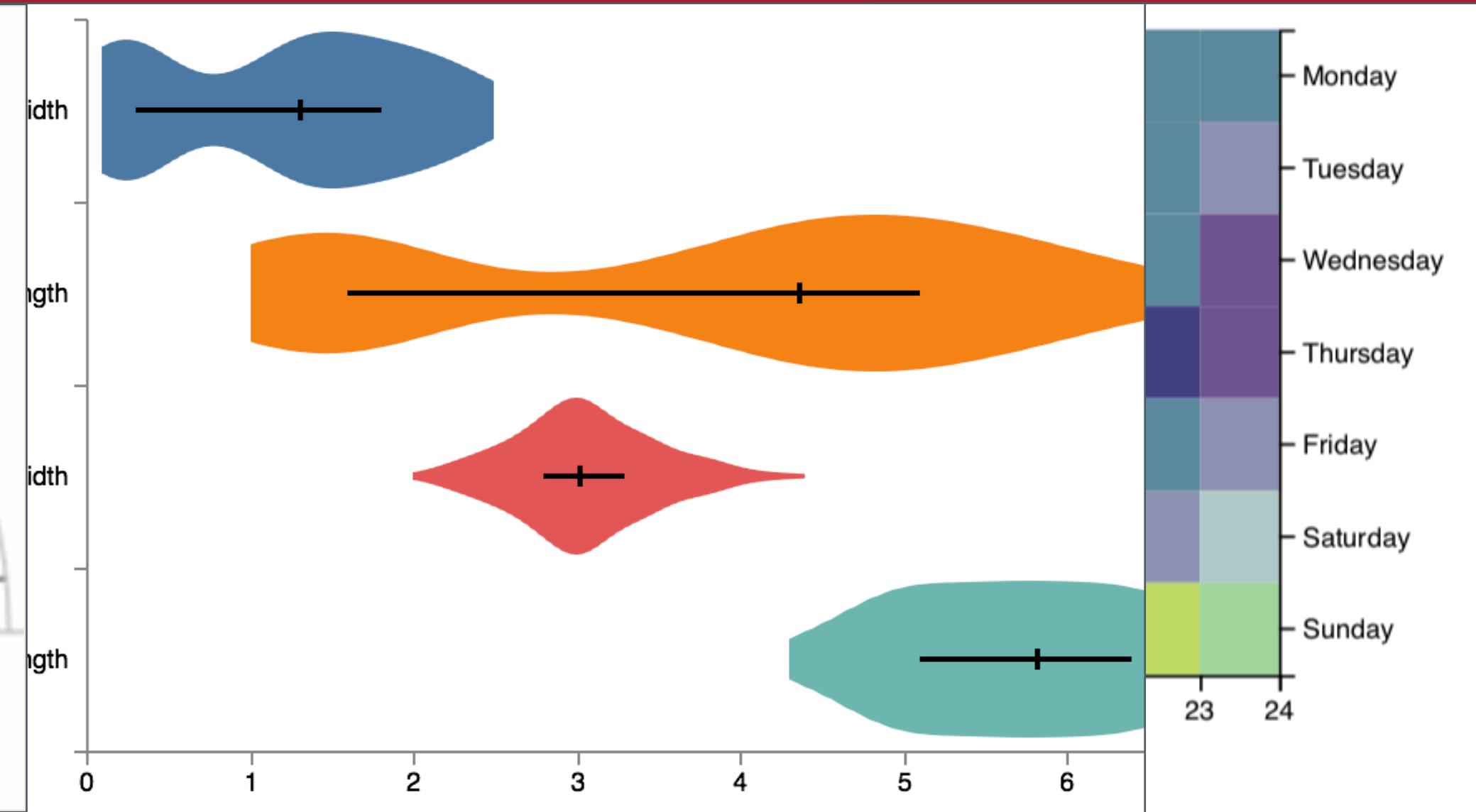
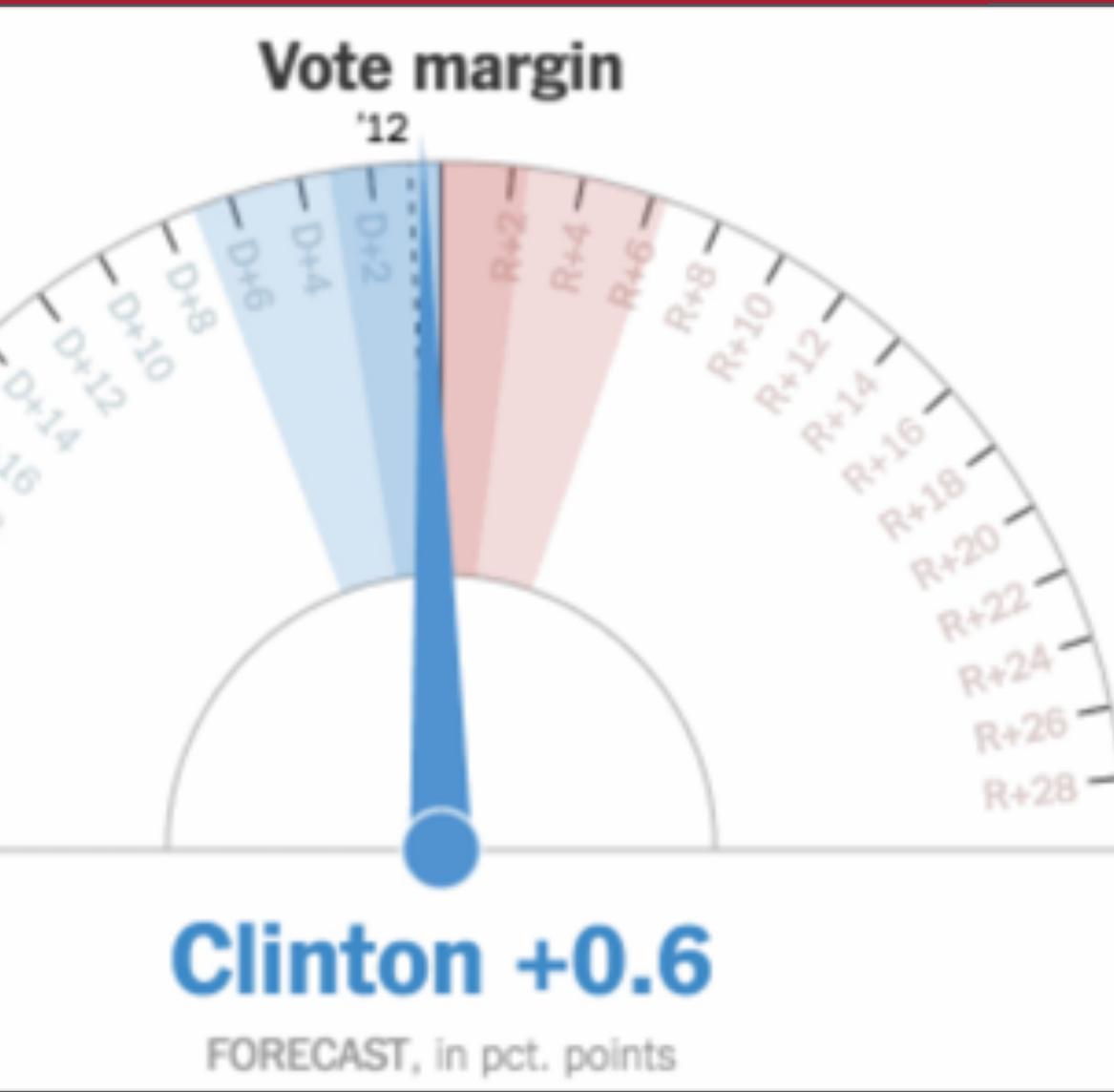


# 6.894: Interactive Data Visualization

# Visualizing Uncertainty

Arvind Satyanarayan



# The Lure of Incredible Certitude

*“Forthright characterization of scientific uncertainty is important in principle and in practice. Nevertheless, economists and other researchers commonly report findings with incredible certitude, reporting point predictions and estimates.”*

# The Lure of Incredible Certitude

*“The flashpoint of controversy has been the fact that research with weak assumptions typically yields bounds on quantities of interest rather than point inferences. Some scientists are comfortable reporting findings in the form of bounds and appreciate making explicit the tradeoff between strength of assumptions and strength of findings that the bounds make plain. However, many hold firm to the traditional practice of reporting point estimates and predictions, even though they may rest on fragile foundations or have obscure interpretations.”*

[Manski, 2019]

# The Lure of Incredible Certitude

*“All offices must try to impress the public with the quality of their work. Should too many doubts be raised, financial support from Congress or other sources may not be forthcoming. More than once has it happened that Congressional appropriations were endangered when it was suspected that government statistics might not be 100 percent accurate. It is natural, therefore, that various offices will defend the quality of their work even to an unreasonable degree.”*

[Morgenstern, 1963]

# The Psychological Necessity of Incredible Certitude?

*“You can't give the client a bound. The client needs a point.”*

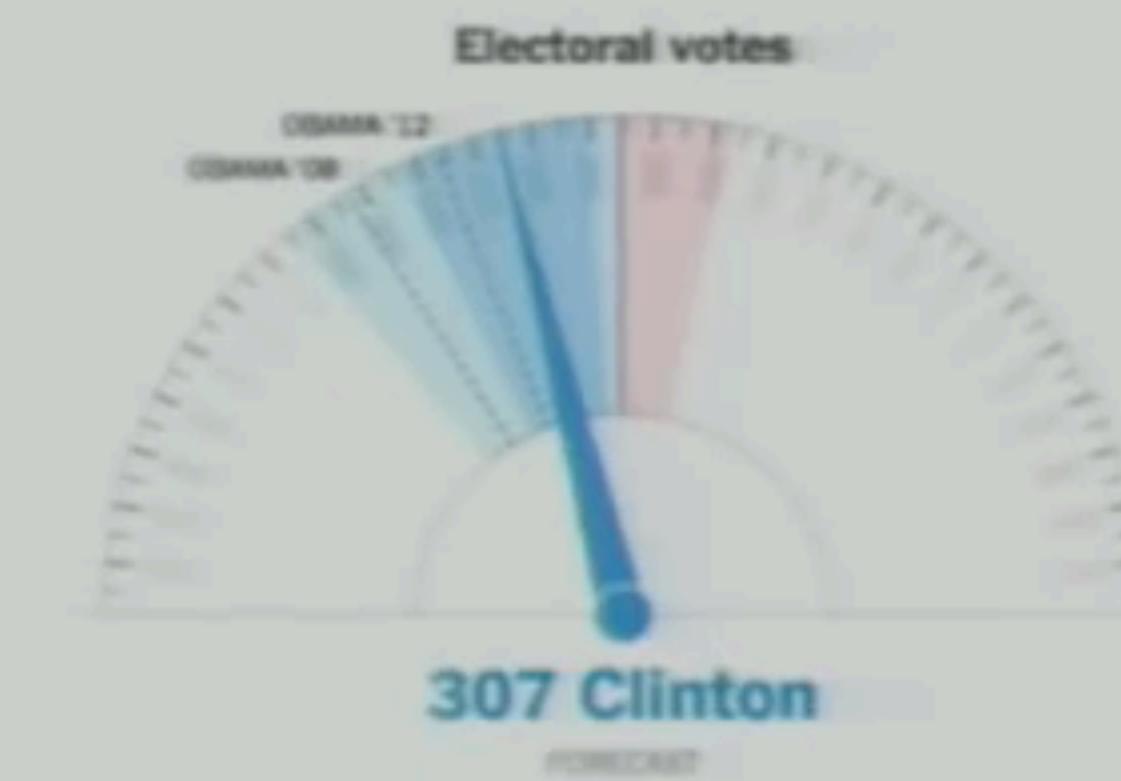
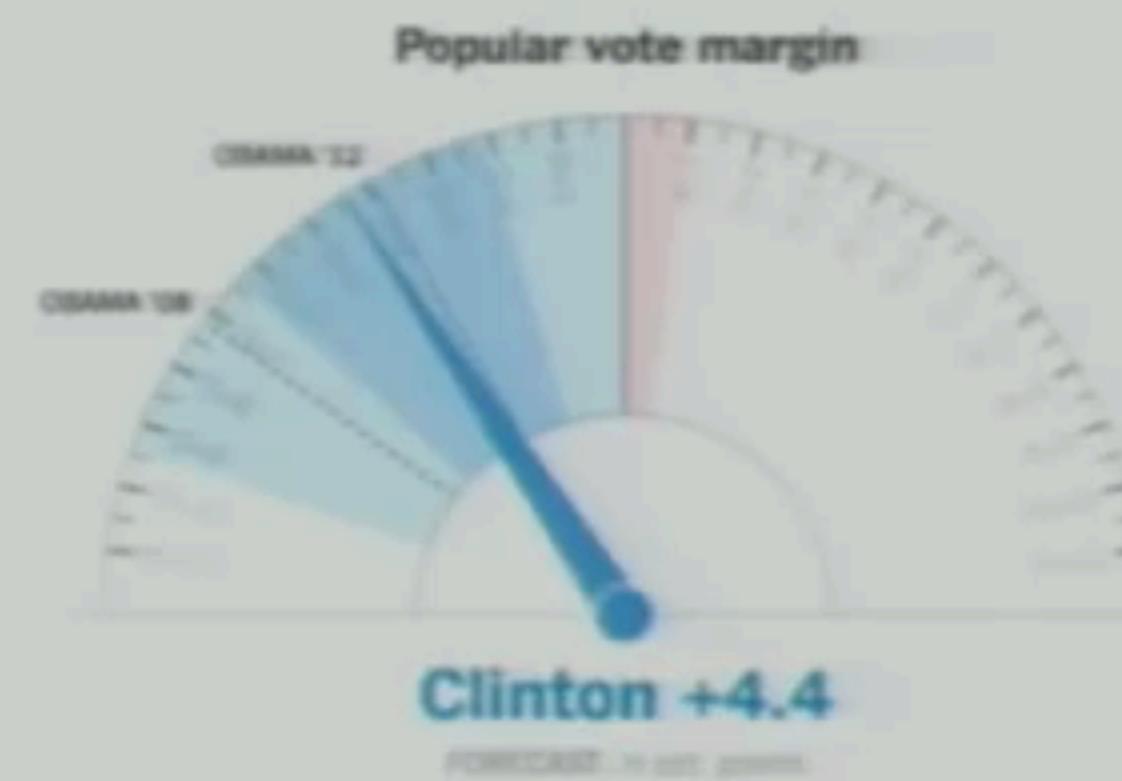
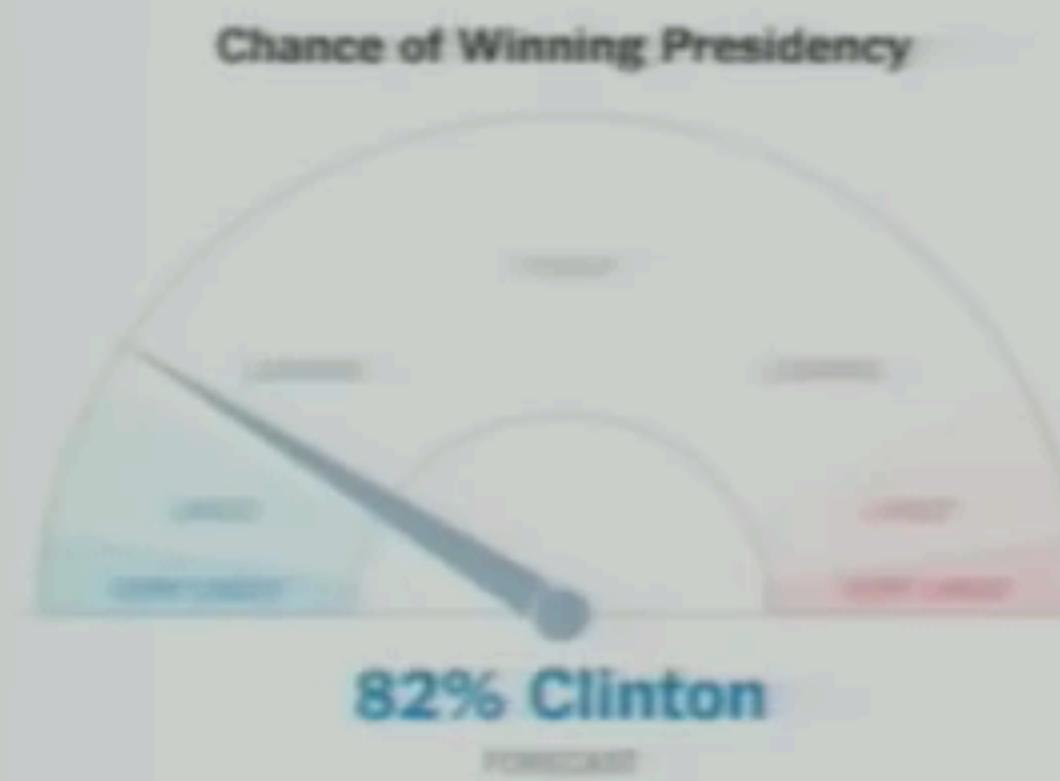
-Jerry Hausman, econometrician.

*“Ranges are for cattle. Give me a number.”*

-Lyndon B. Johnson, POTUS.

# Live Presidential Forecast

Updated 7:58:14 PM ET



The projections for each state are based on the votes reported so far and how those places have voted in previous elections.

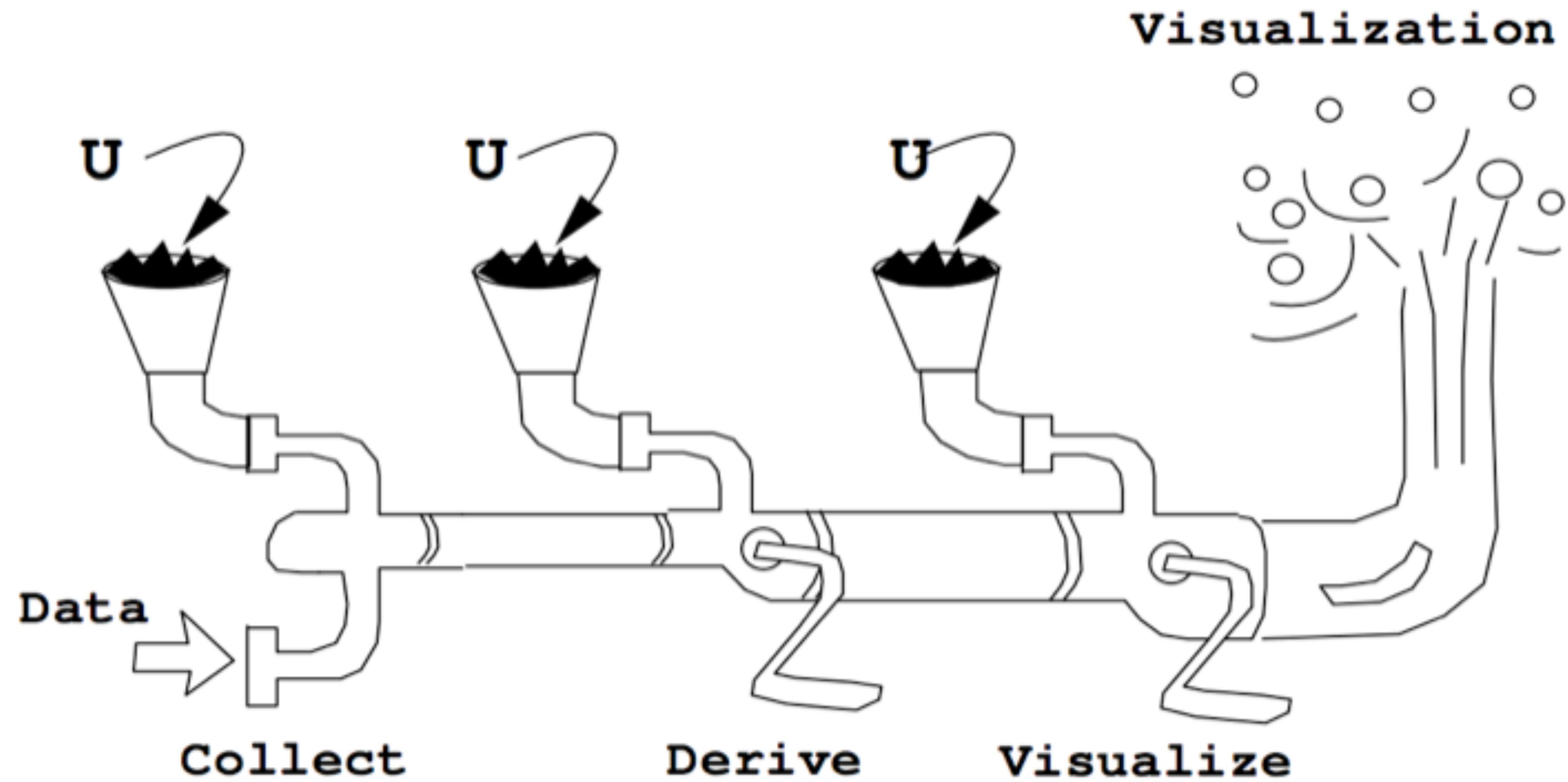
We're showing the closest states by default. [Show all states](#)

Name	Est. per cent of votes	Estimated margin
Michigan	0%	Clinton +6.5
New Mexico	0%	Clinton +6.4
Wisconsin	0%	Clinton +5.4



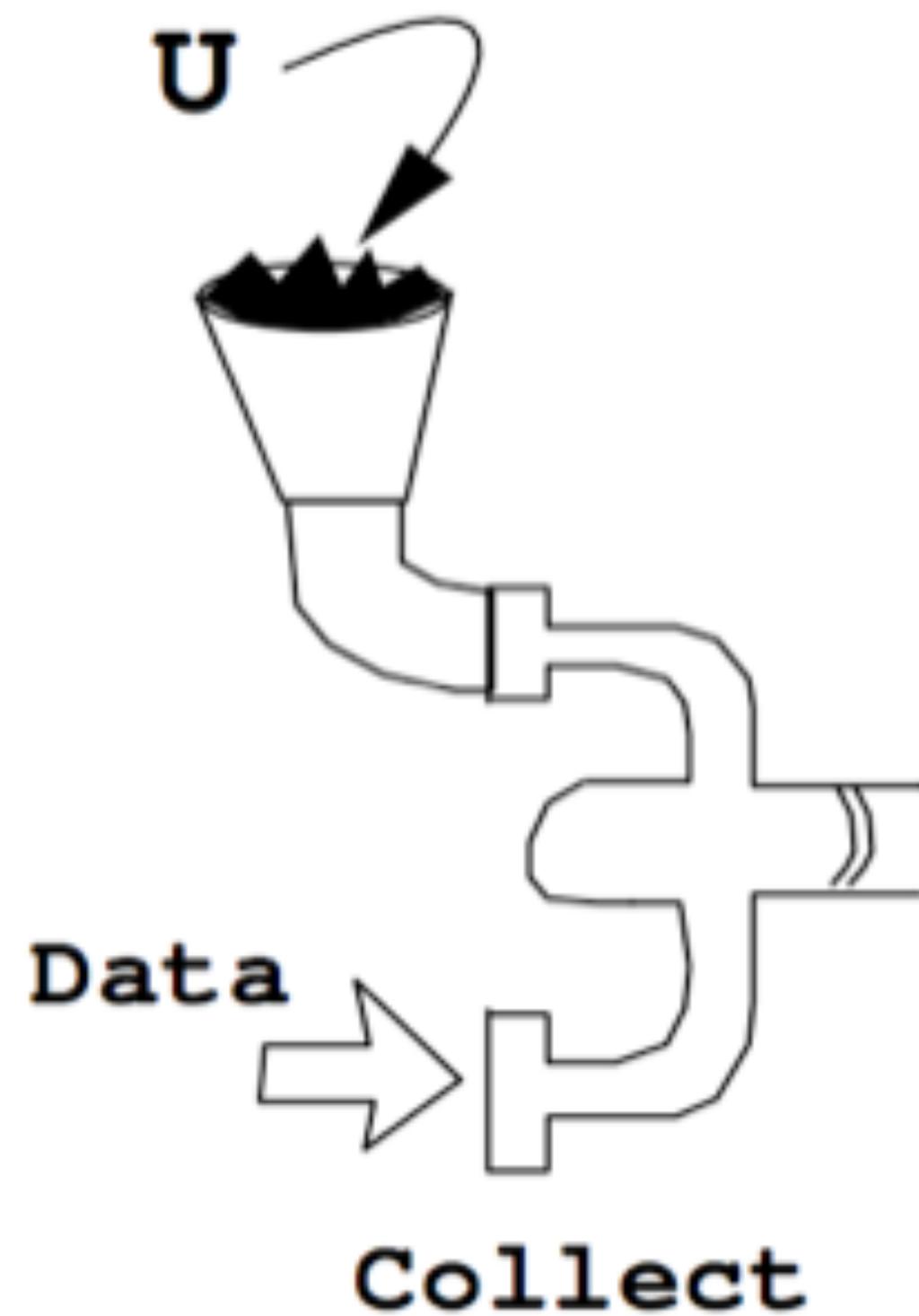
OPEN  
VIS 2017  
CONF

# Sources & Types of Uncertainty



# Sources & Types of *Uncertainty*

The possibility of many/other outcomes.

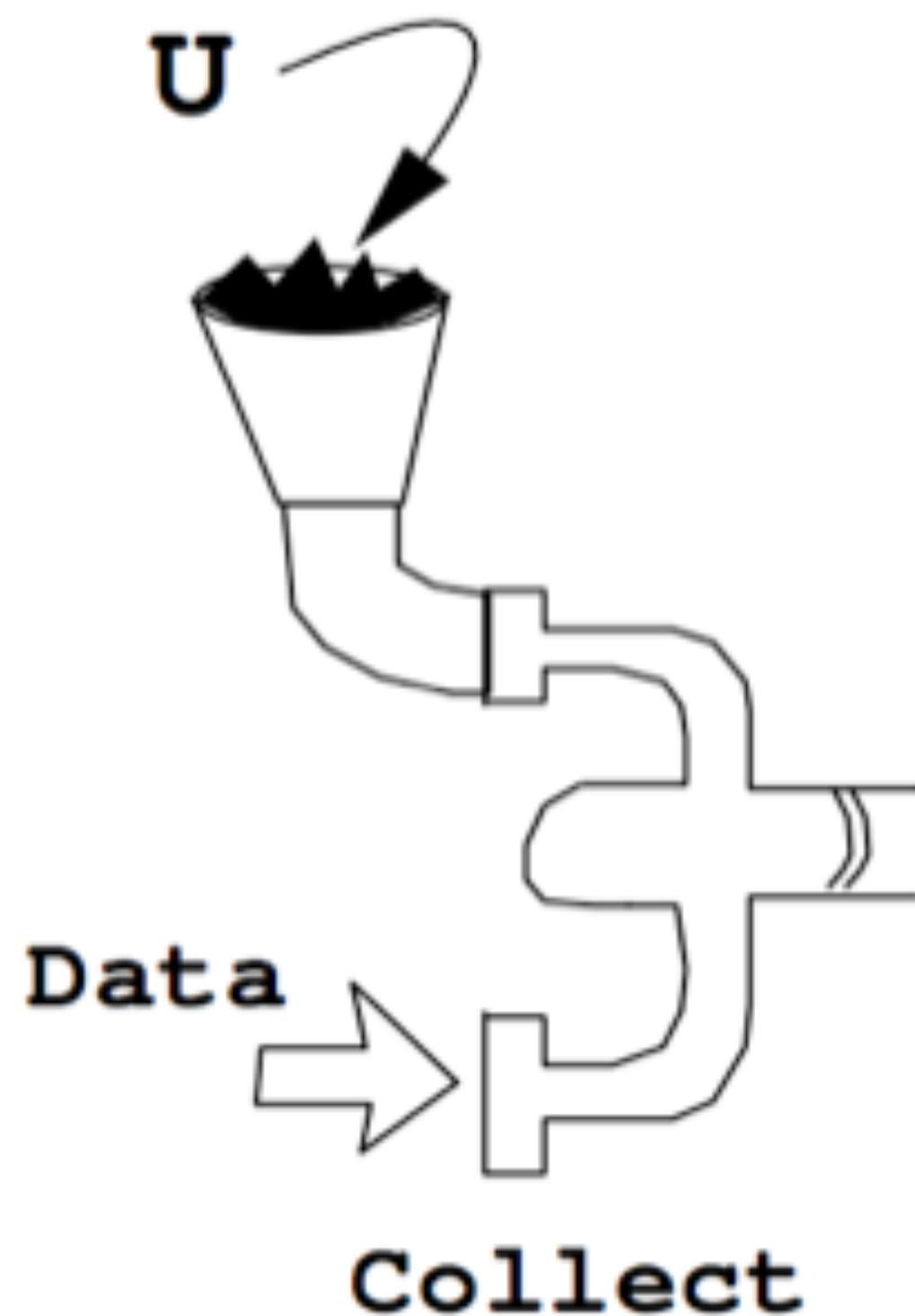


# Sources & Types of *Uncertainty*

*The possibility of many/other outcomes.*

## Measurement Uncertainty

How and how much should we sample the data?



Precision



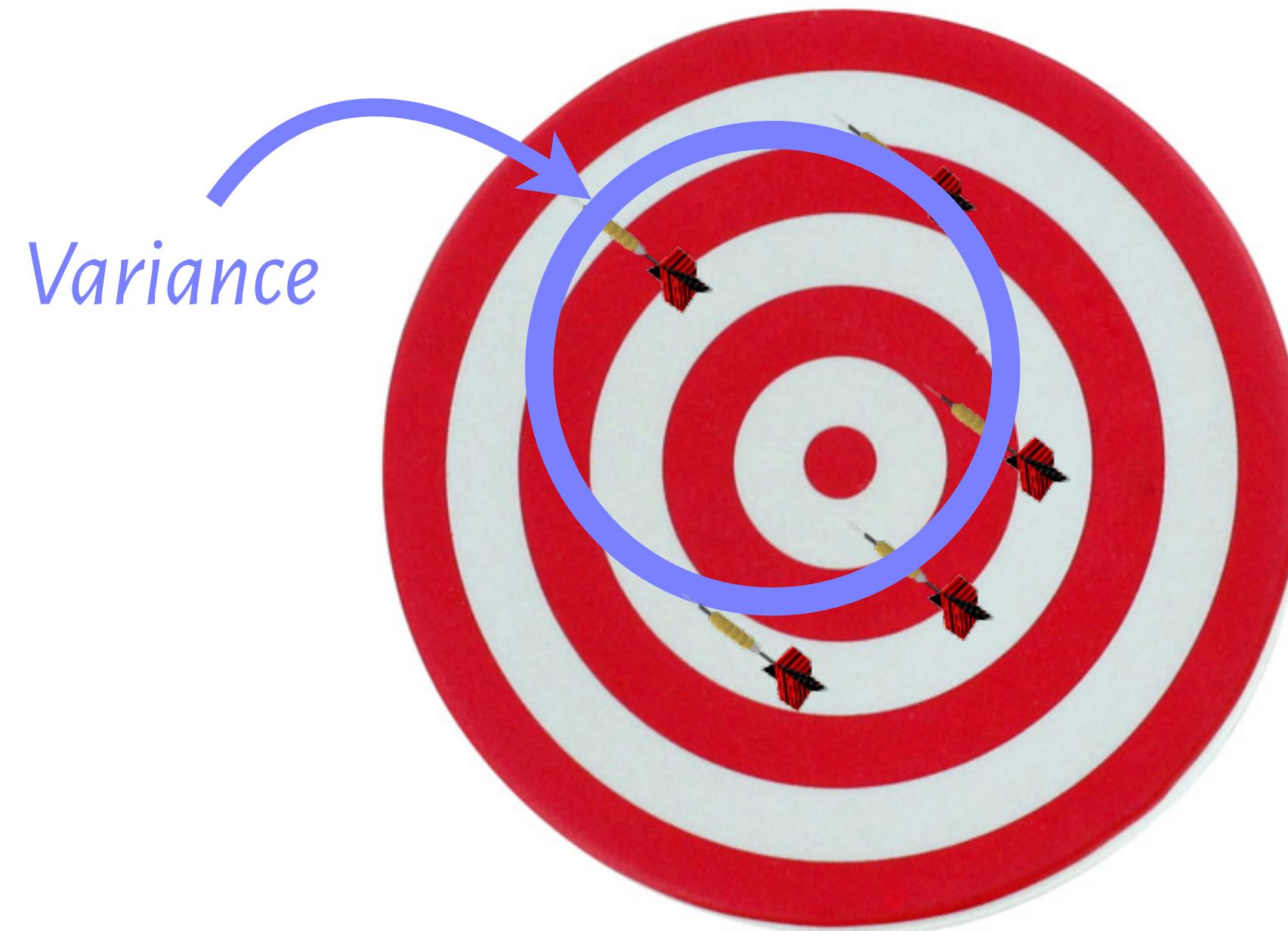
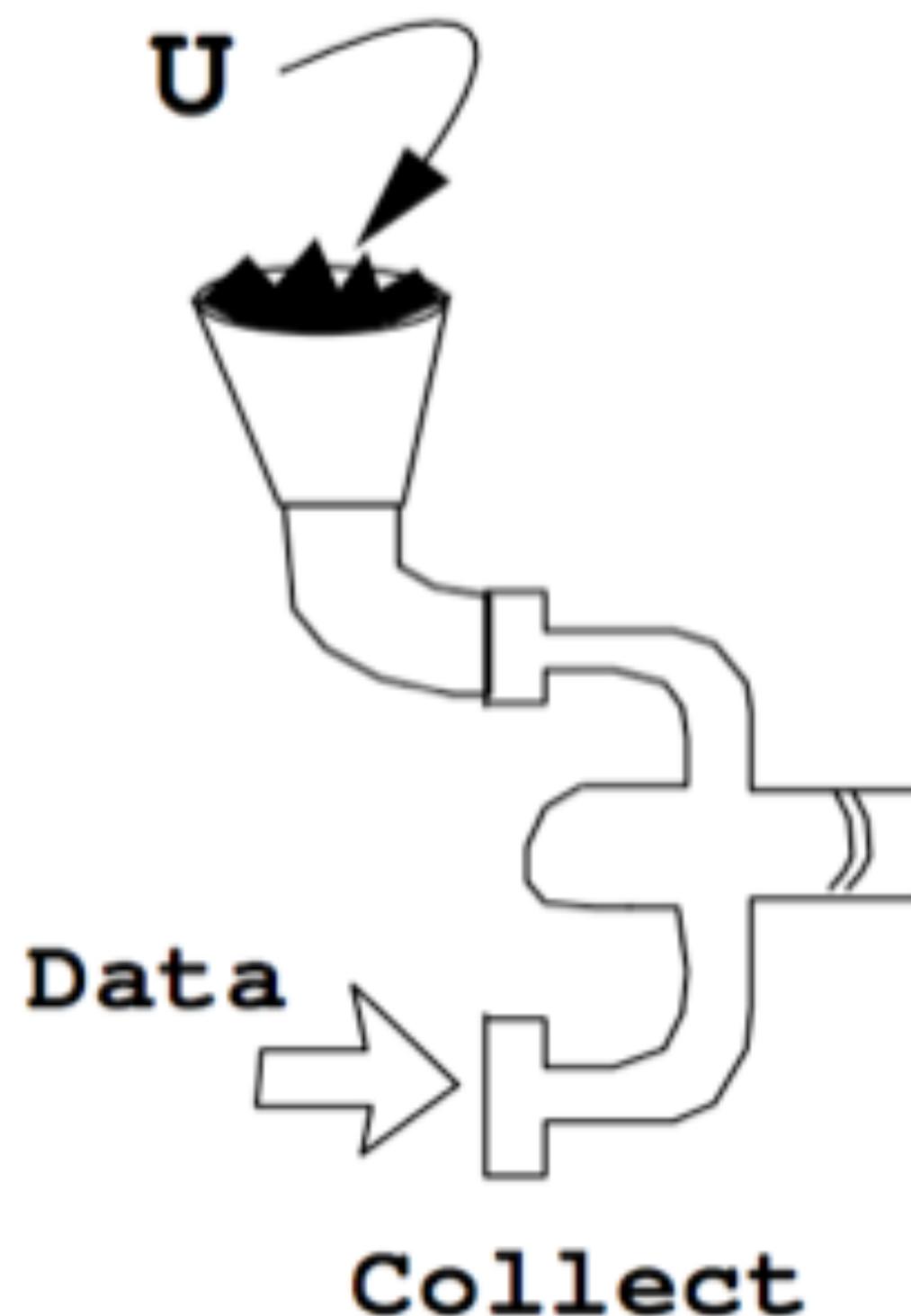
Accuracy

# Sources & Types of *Uncertainty*

The possibility of many/other outcomes.

## Measurement Uncertainty

How and how much should we sample the data?



**Precision:** How much do the measurements agree with each other?



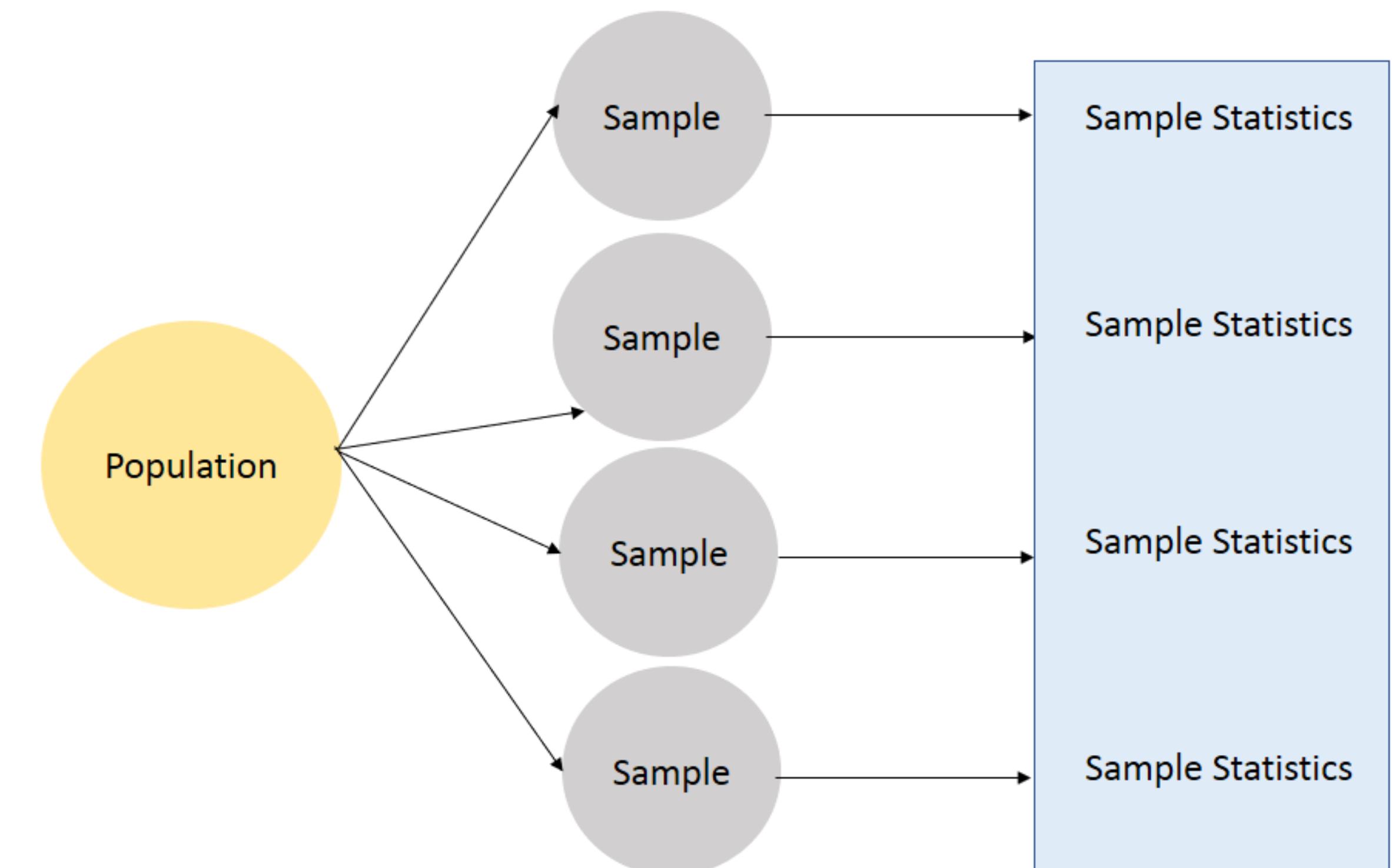
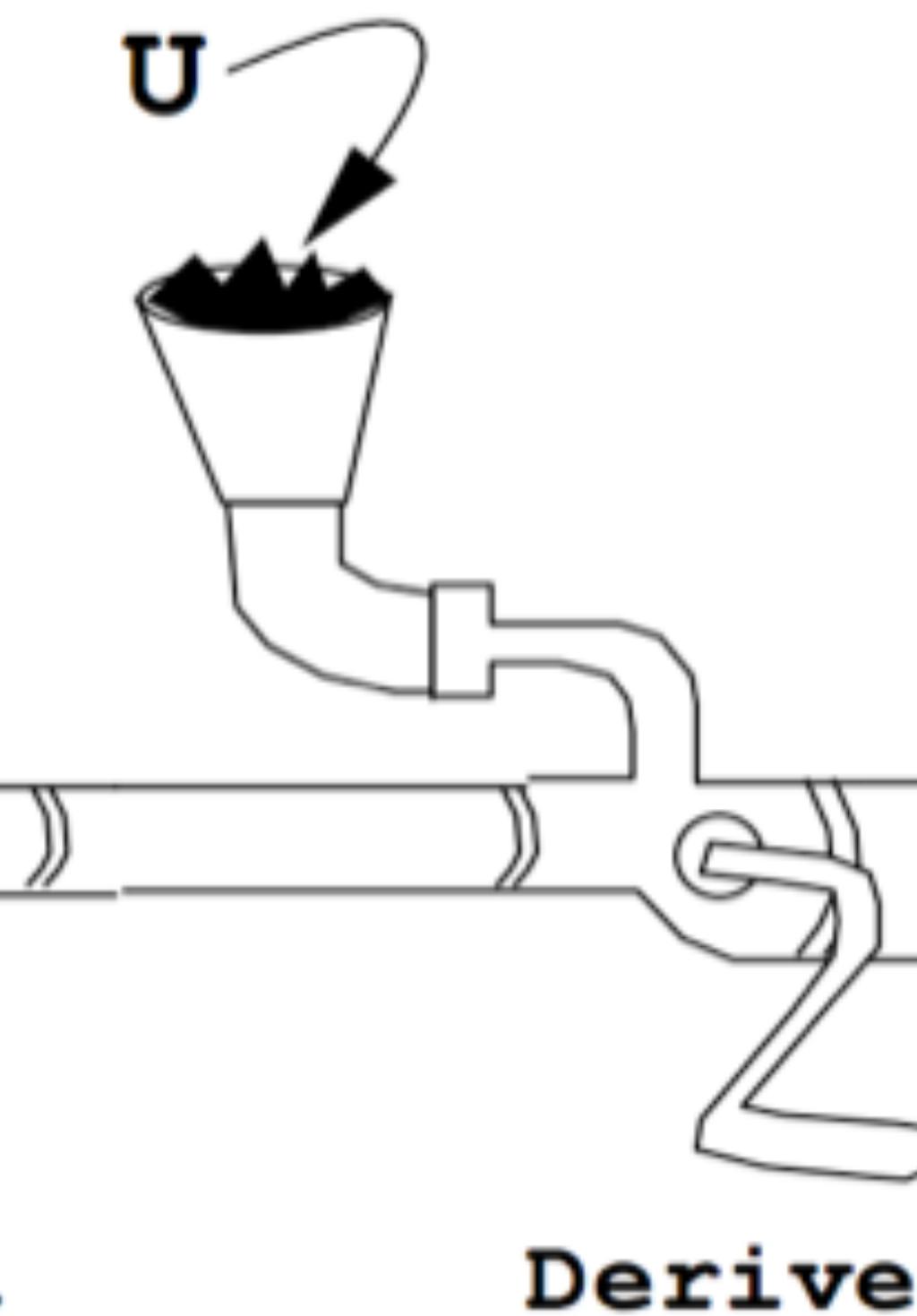
**Accuracy:** How much do the measurements reflect the true value?

# Sources & Types of *Uncertainty*

*The possibility of many/other outcomes.*

## Model Uncertainty

How does the data fit together?



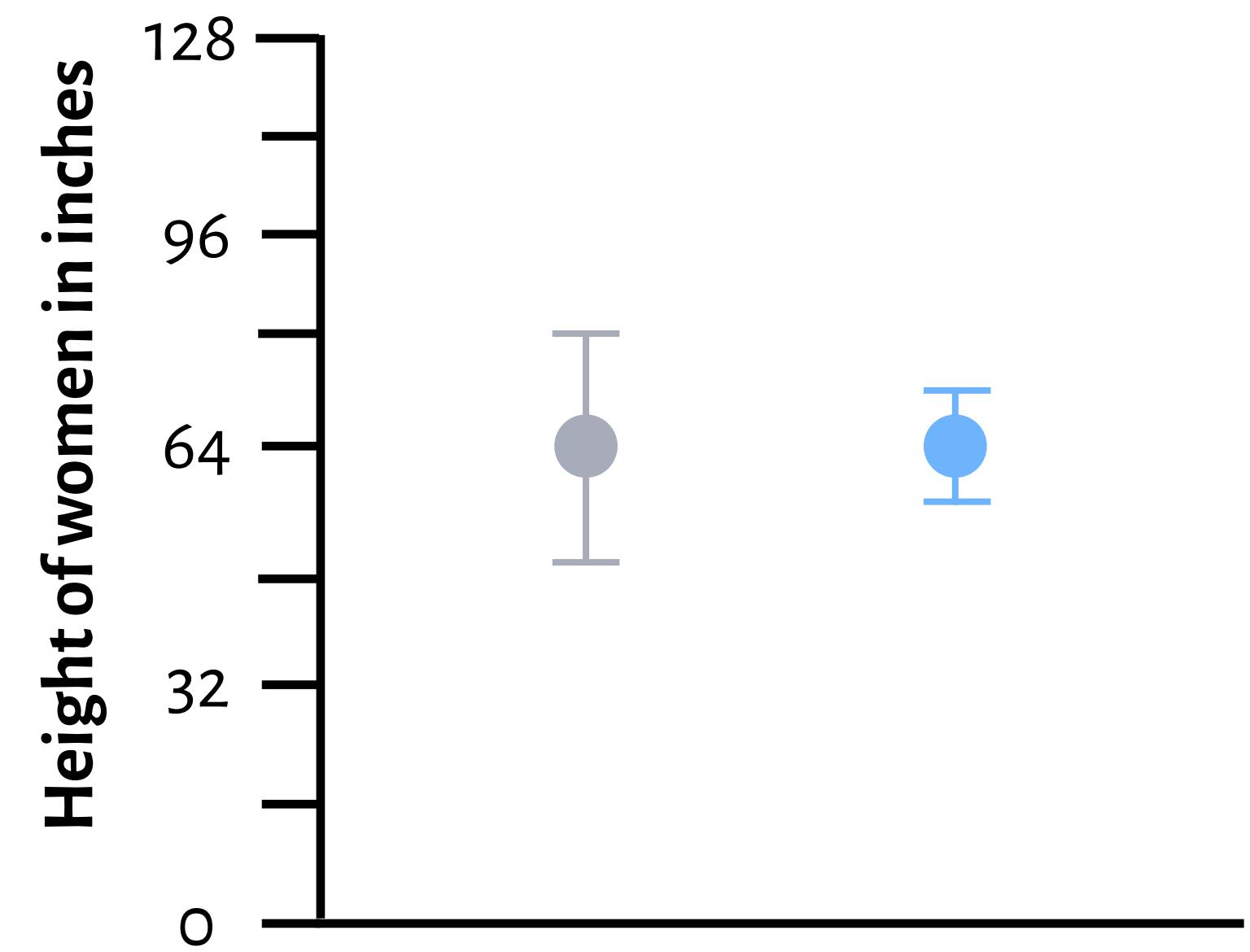
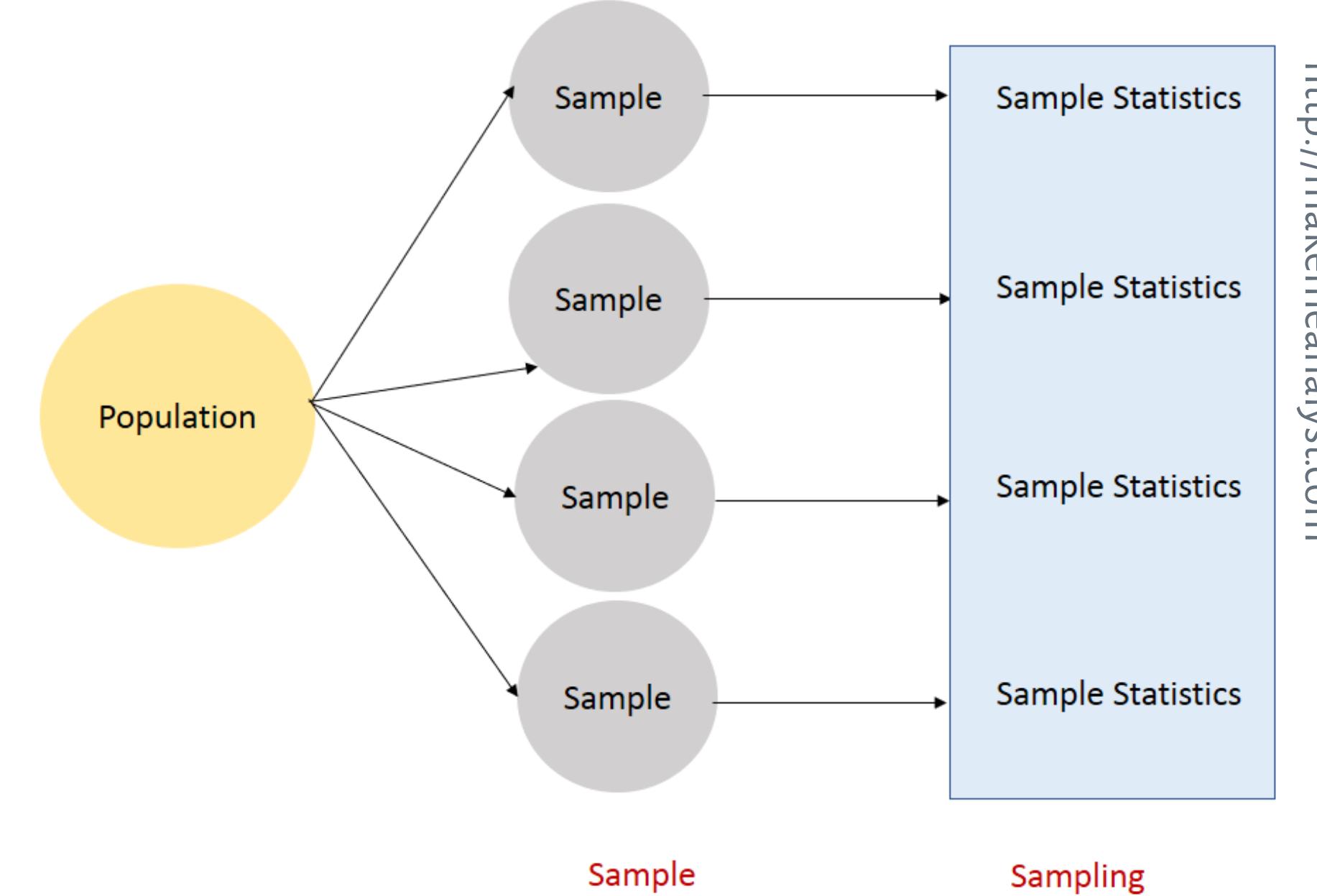
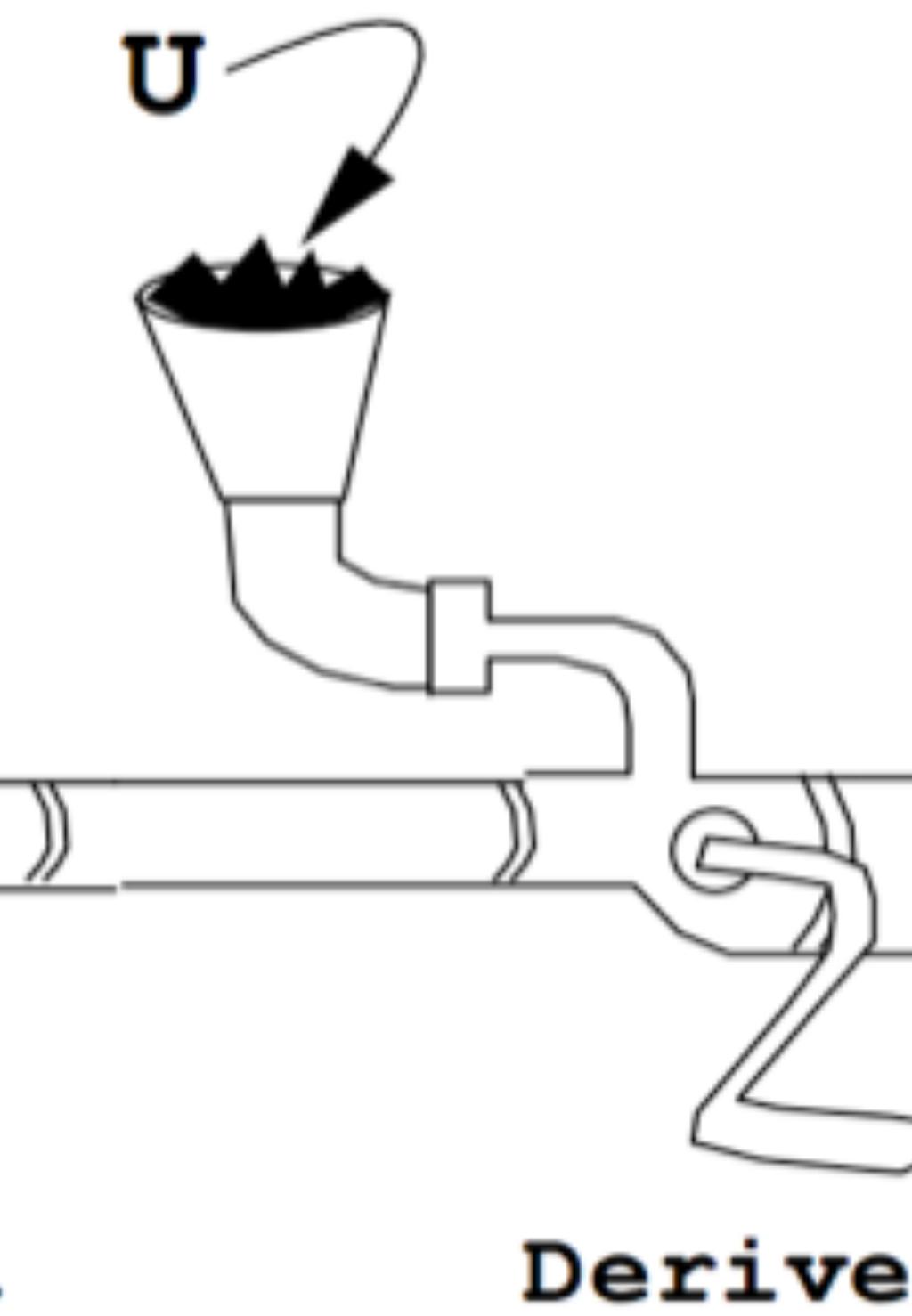
<http://makemeanalyst.com>

# Sources & Types of *Uncertainty*

The possibility of many/other outcomes.

## Model Uncertainty

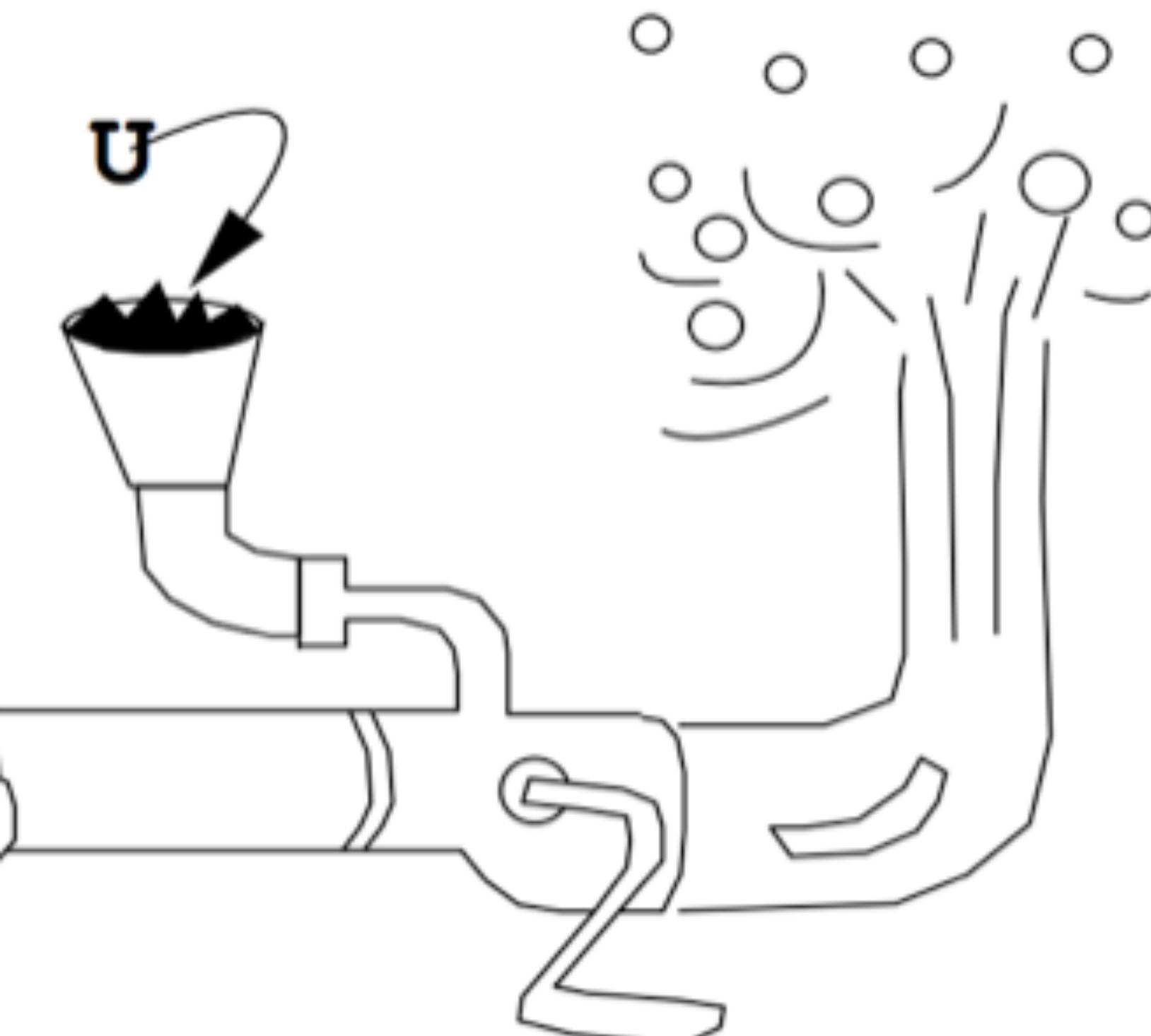
How does the data fit together?



# Sources & Types of *Uncertainty*

The possibility of many/other outcomes.

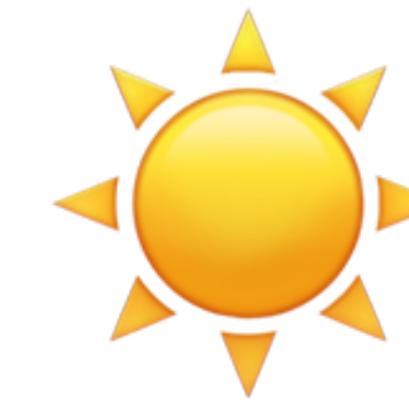
## Visualization



**Visualize**

## Decision/Forecast Uncertainty

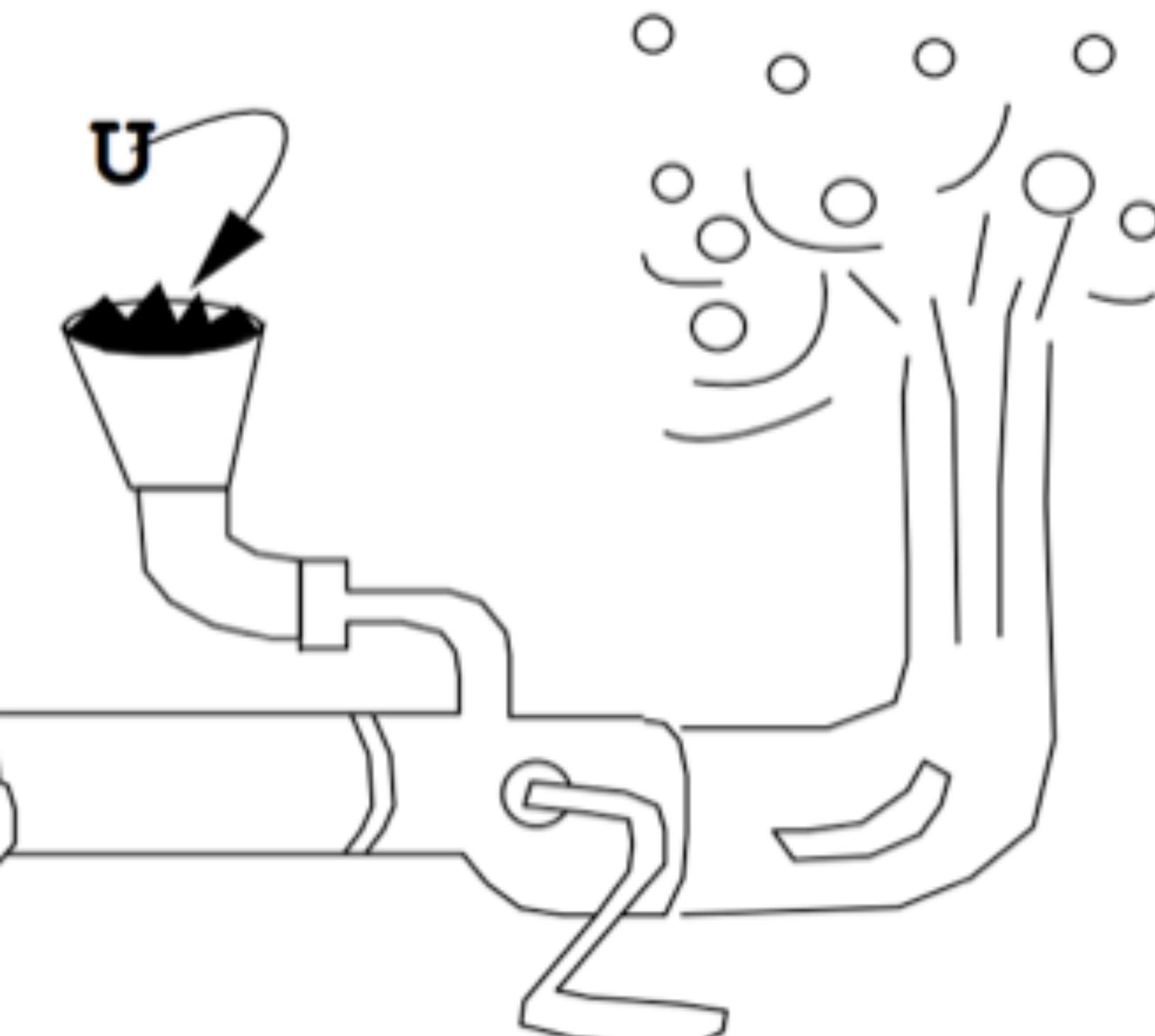
How do I assess the risk or error?




# Sources & Types of *Uncertainty*

The possibility of many/other outcomes.

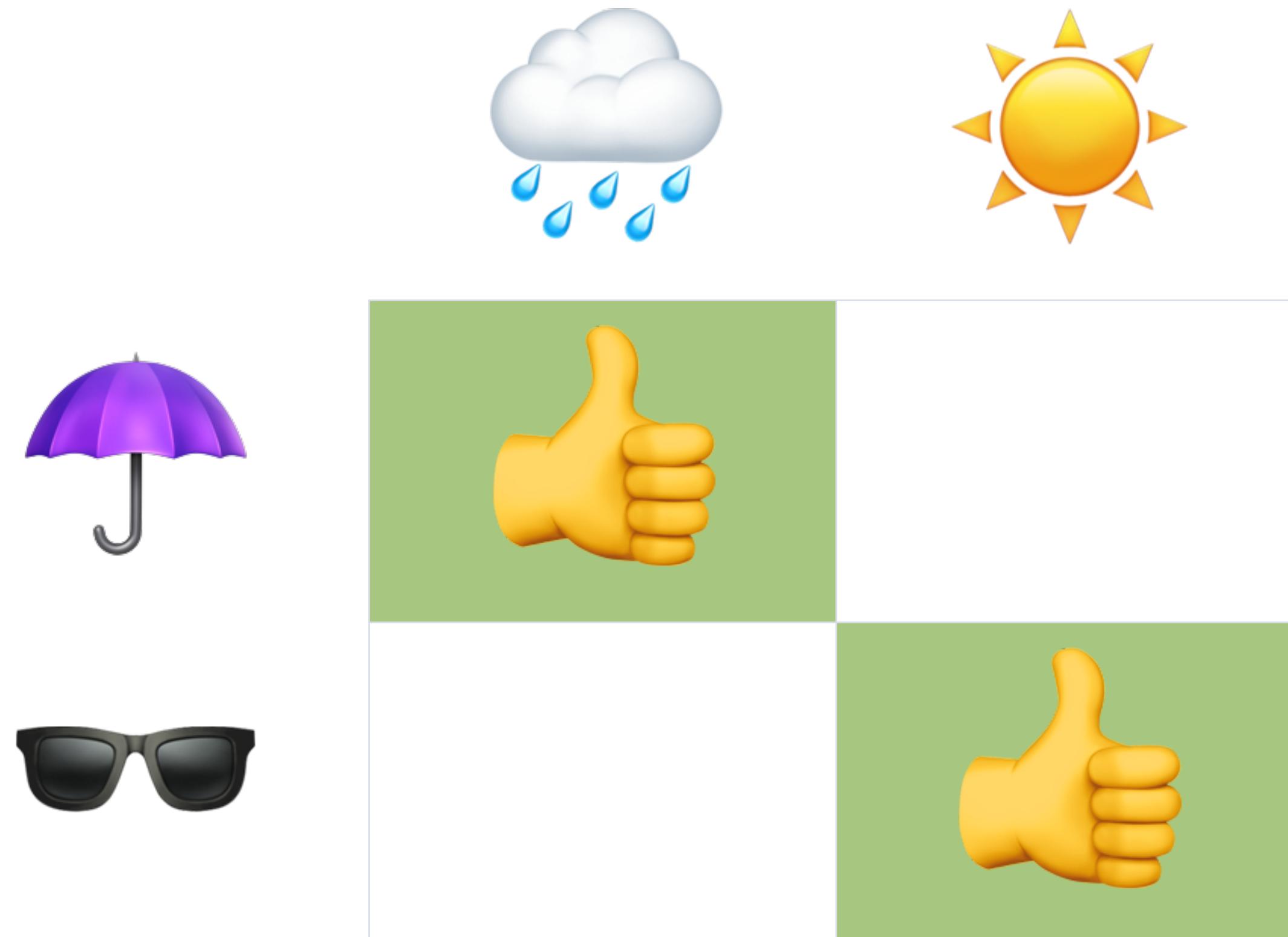
## Visualization



**Visualize**

## Decision/Forecast Uncertainty

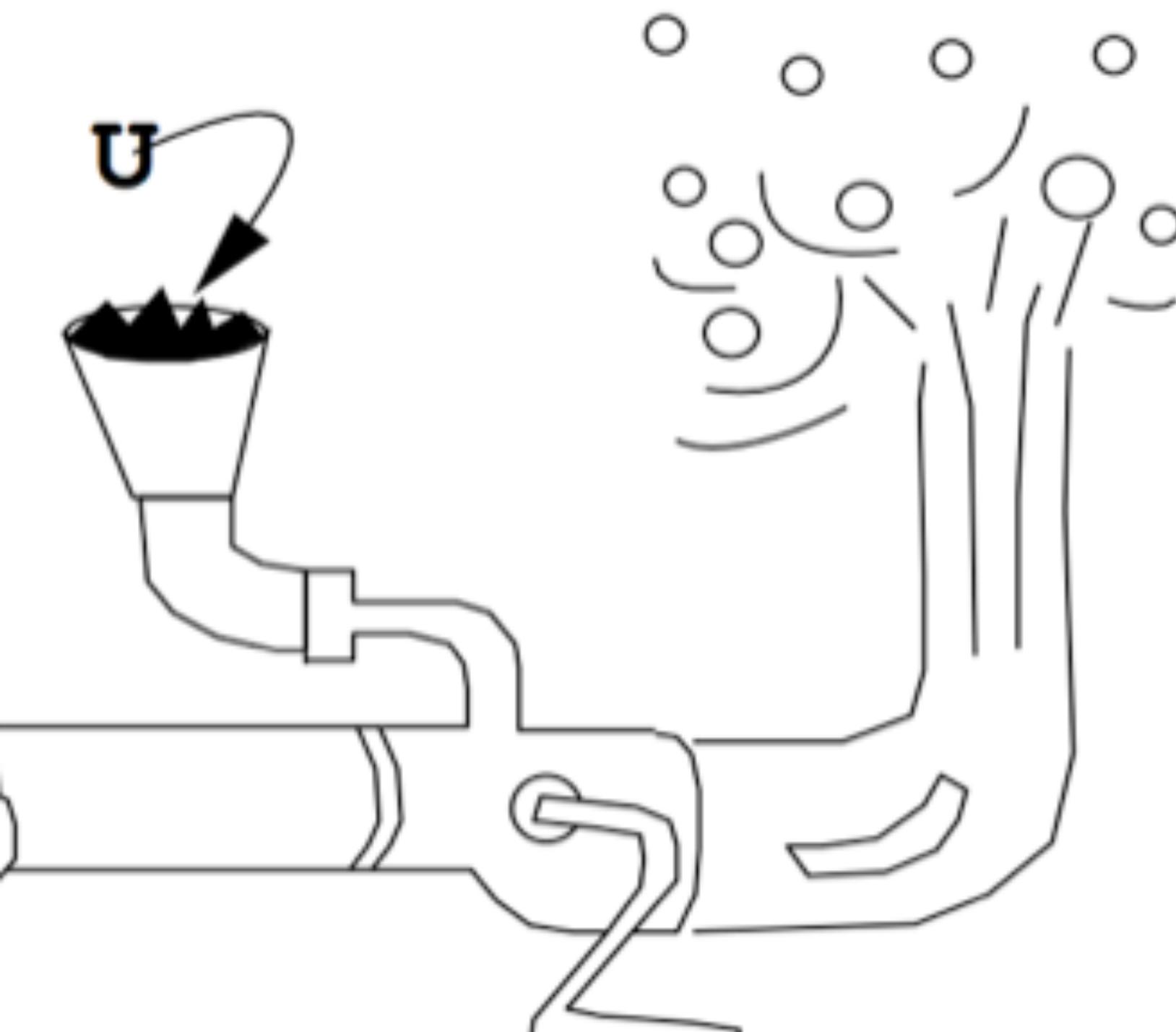
How do I assess the risk or error?



# Sources & Types of *Uncertainty*

The possibility of many/other outcomes.

## Visualization



**Visualize**

## Decision/Forecast Uncertainty

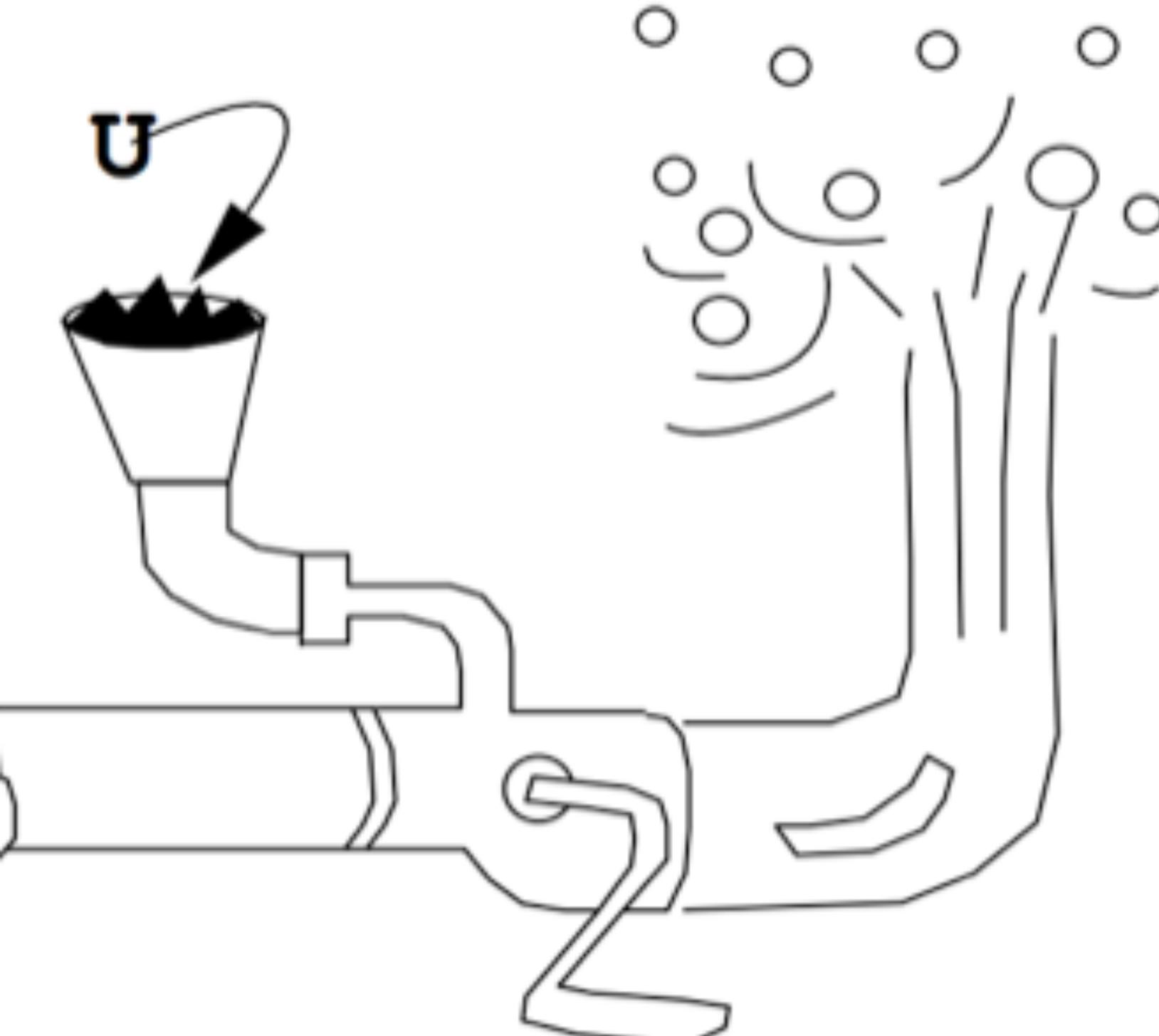
How do I assess the risk or error?

<b>Type II False Negative</b>	<b>Type I False Positive</b>

# Sources & Types of *Uncertainty*

The possibility of many/other outcomes.

**Visualization**



**Visualize**

**Never confuse Type I and II errors again:**

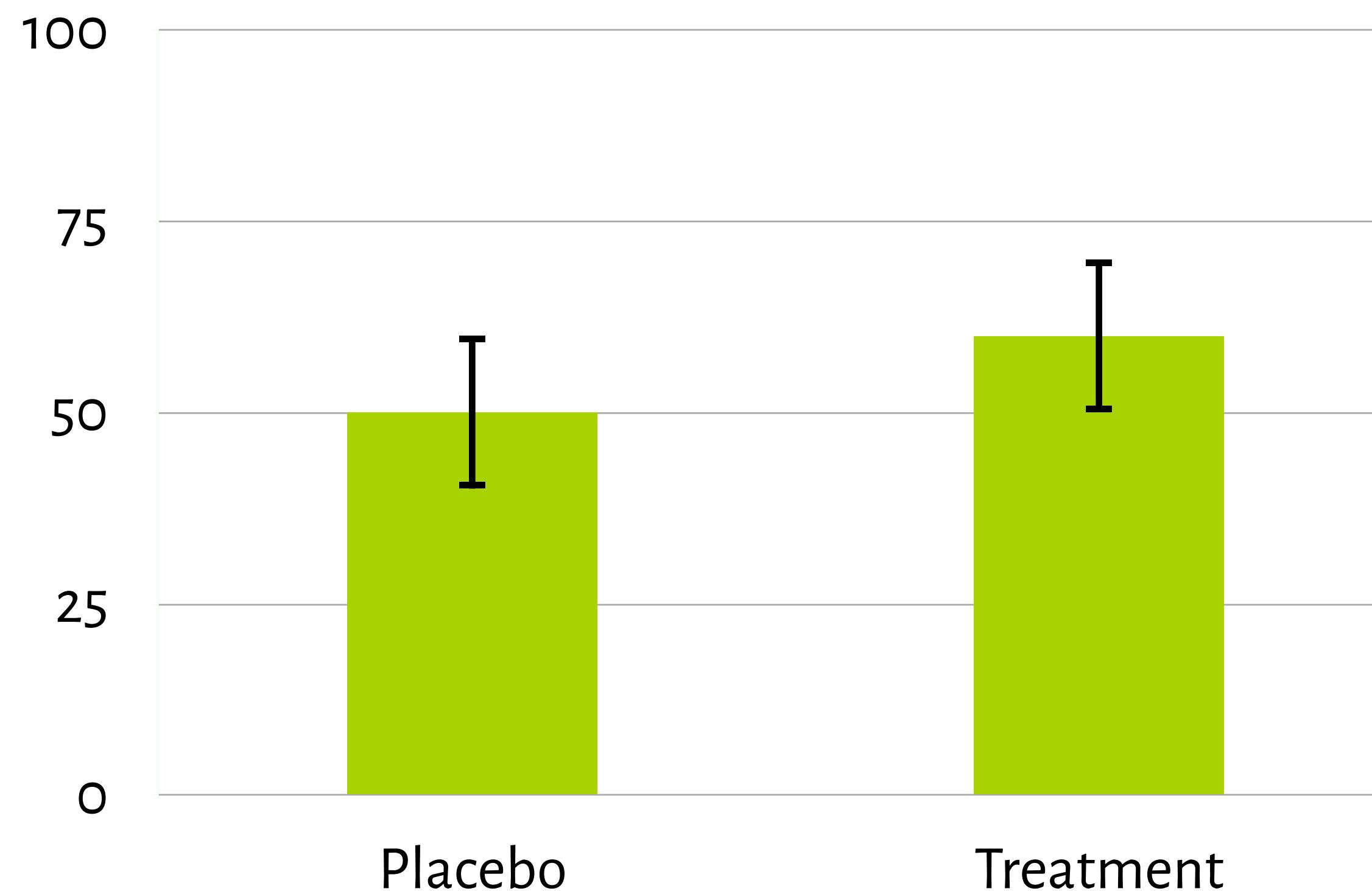
**Just remember that the Boy Who Cried Wolf caused both Type I & II errors, in that order.**

**First everyone believed there was a wolf, when there wasn't. Next they believed there was no wolf, when there was.**

**Substitute "effect" for "wolf" and you're done.**

Kudos to @danolner for the thought. Illustration by Francis Barlow  
"De pastoris pueri et agricolis" (1687). Public Domain. Via [wikimedia.org](https://commons.wikimedia.org)

# Visualizing Uncertainty: Glyphs

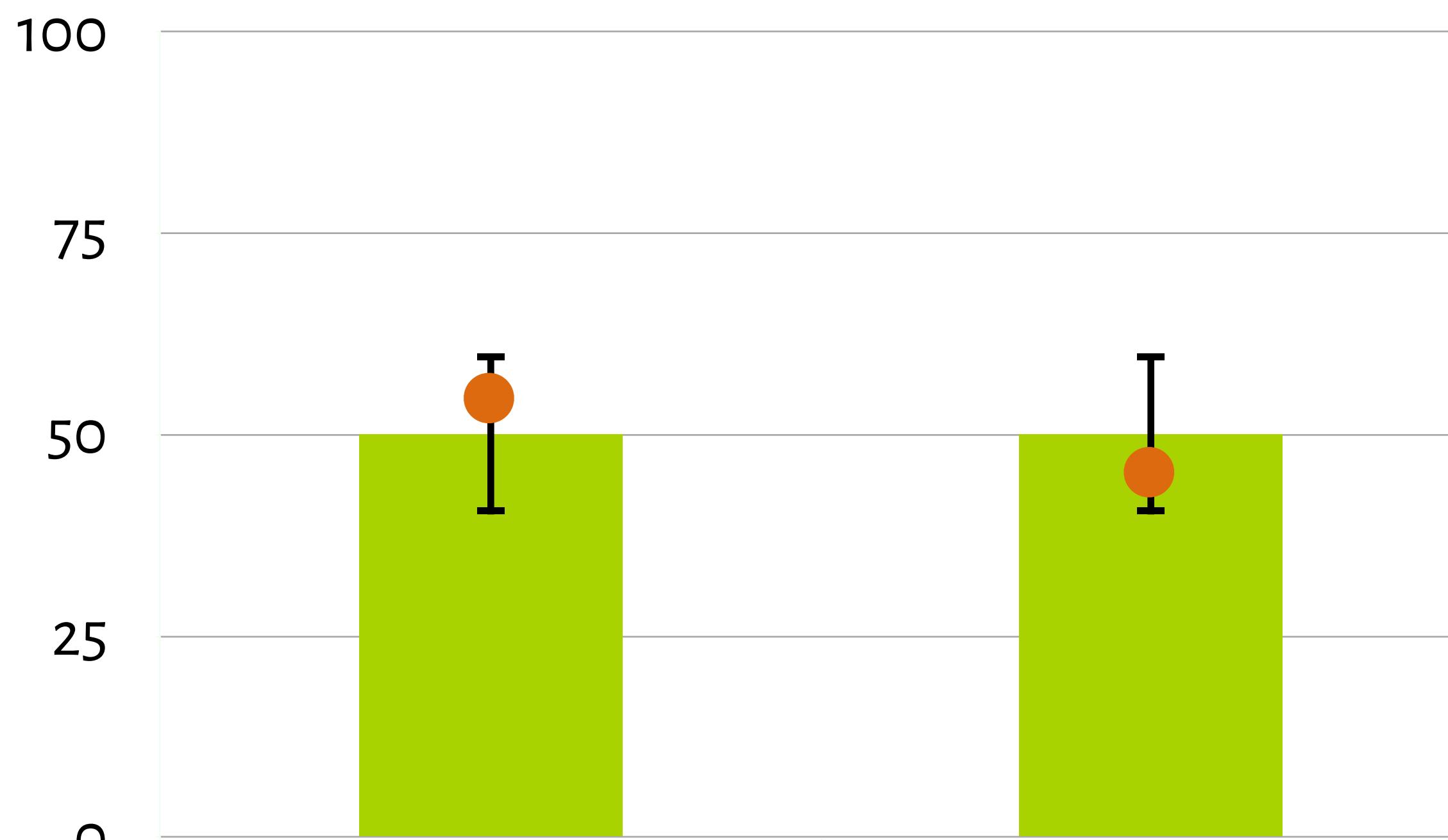


## Expressiveness?

Does it express *all* the facts and *only* the facts?

- ✗ Error bars aren't consistently used to visualize the same measure (standard deviation, standard error, IQR, 95% CI).

# Visualizing Uncertainty: Glyphs



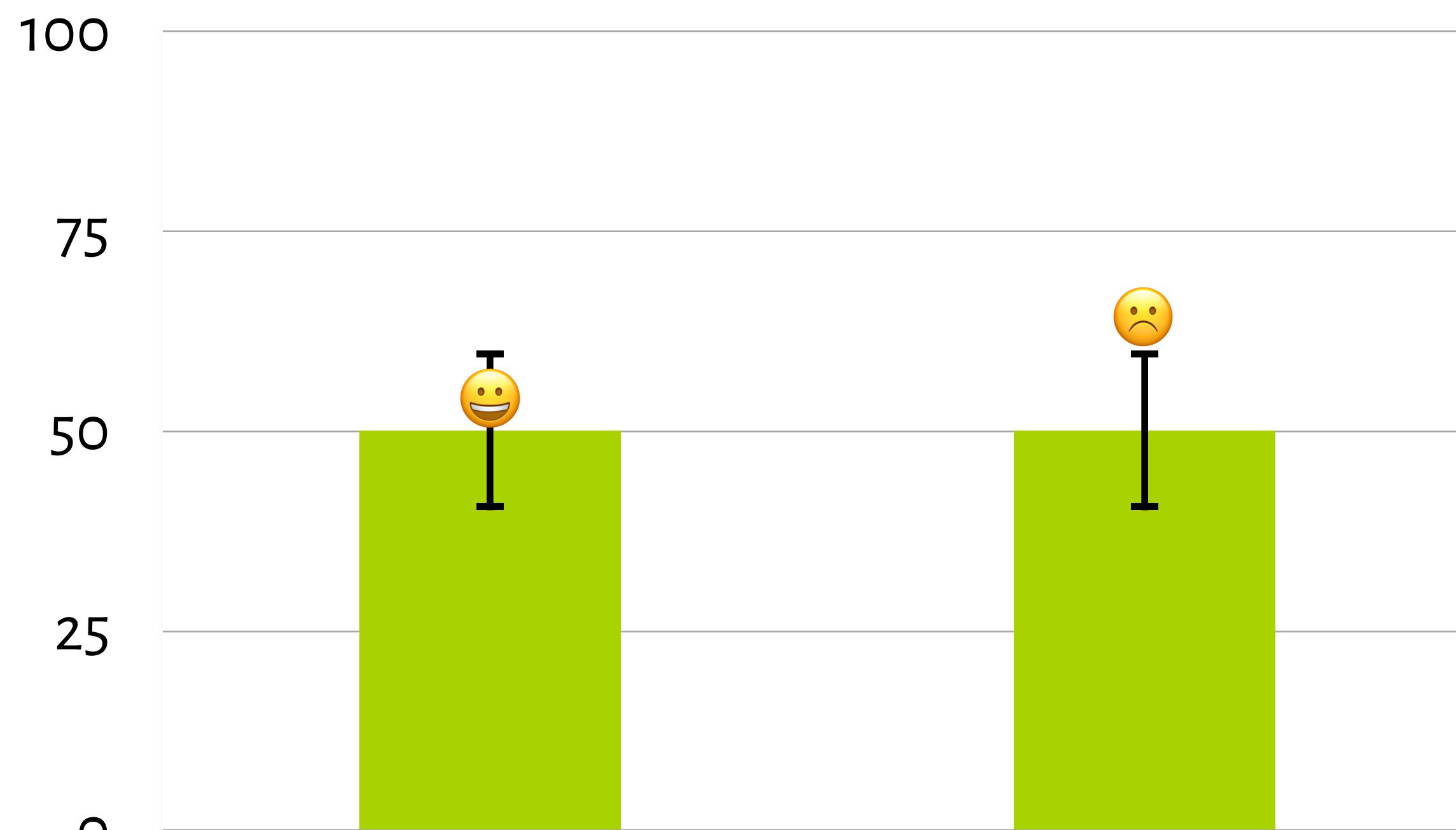
## Expressiveness?

Does it express *all* the facts and *only* the facts?

- ✗ Error bars aren't consistently used to visualize the same measure (standard deviation, standard error, IQR, 95% CI).
- ✗ Within-the-bar bias: people perceive points falling within the bar as more likely than those that lie outside.

[Newman & Scholl, 2012]  
[Correll & Gleicher, 2014]

# Visualizing Uncertainty: Glyphs



## Expressiveness?

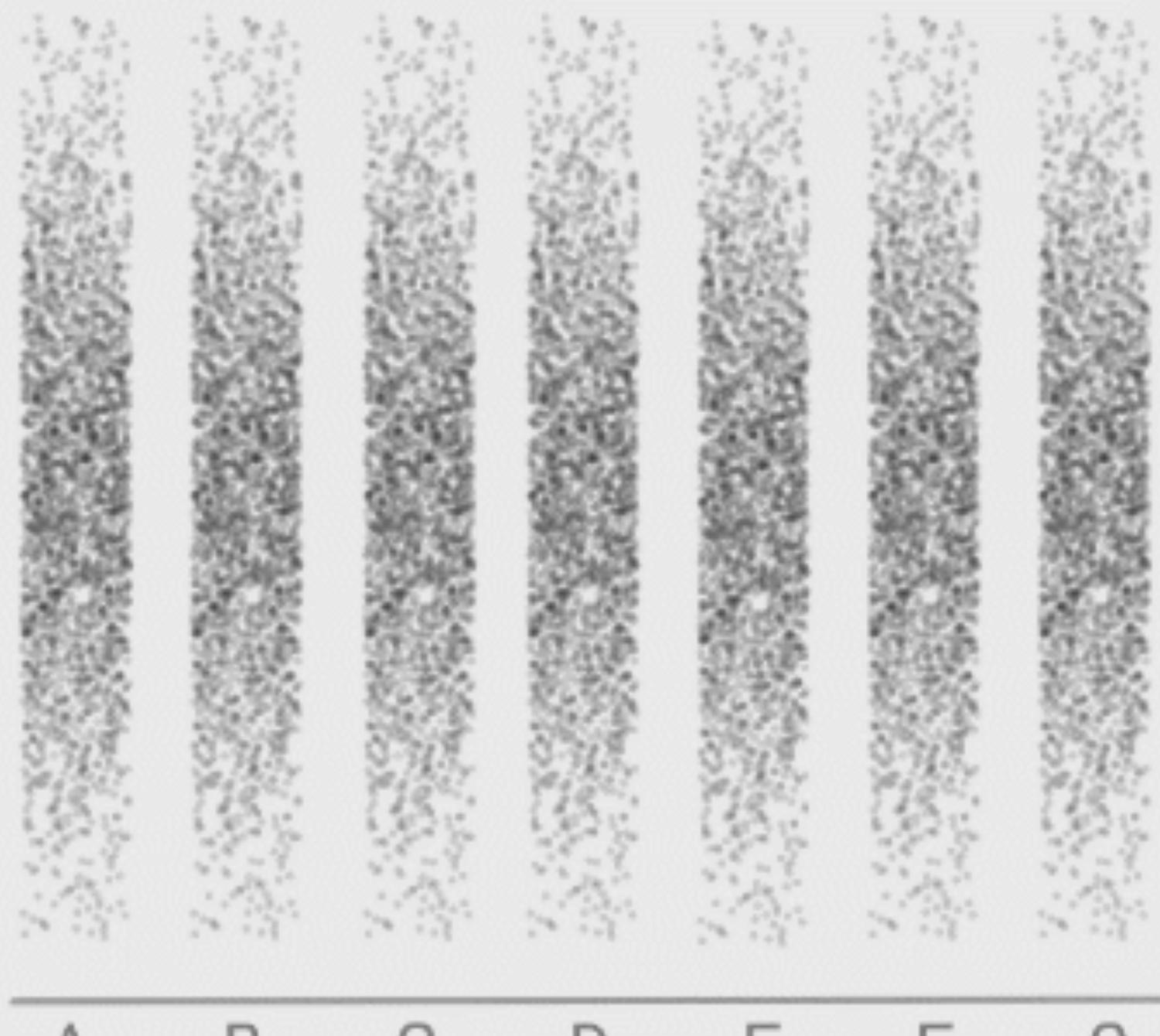
Does it express *all* the facts and *only* the facts?

- ✗ Error bars aren't consistently used to visualize the same measure (standard deviation, standard error, IQR, 95% CI).
- ✗ Within-the-bar bias: people perceive points falling within the bar as more likely than those that lie outside.
- ✗ Binary bias: people perceive values to either be in or out of the margins of error.

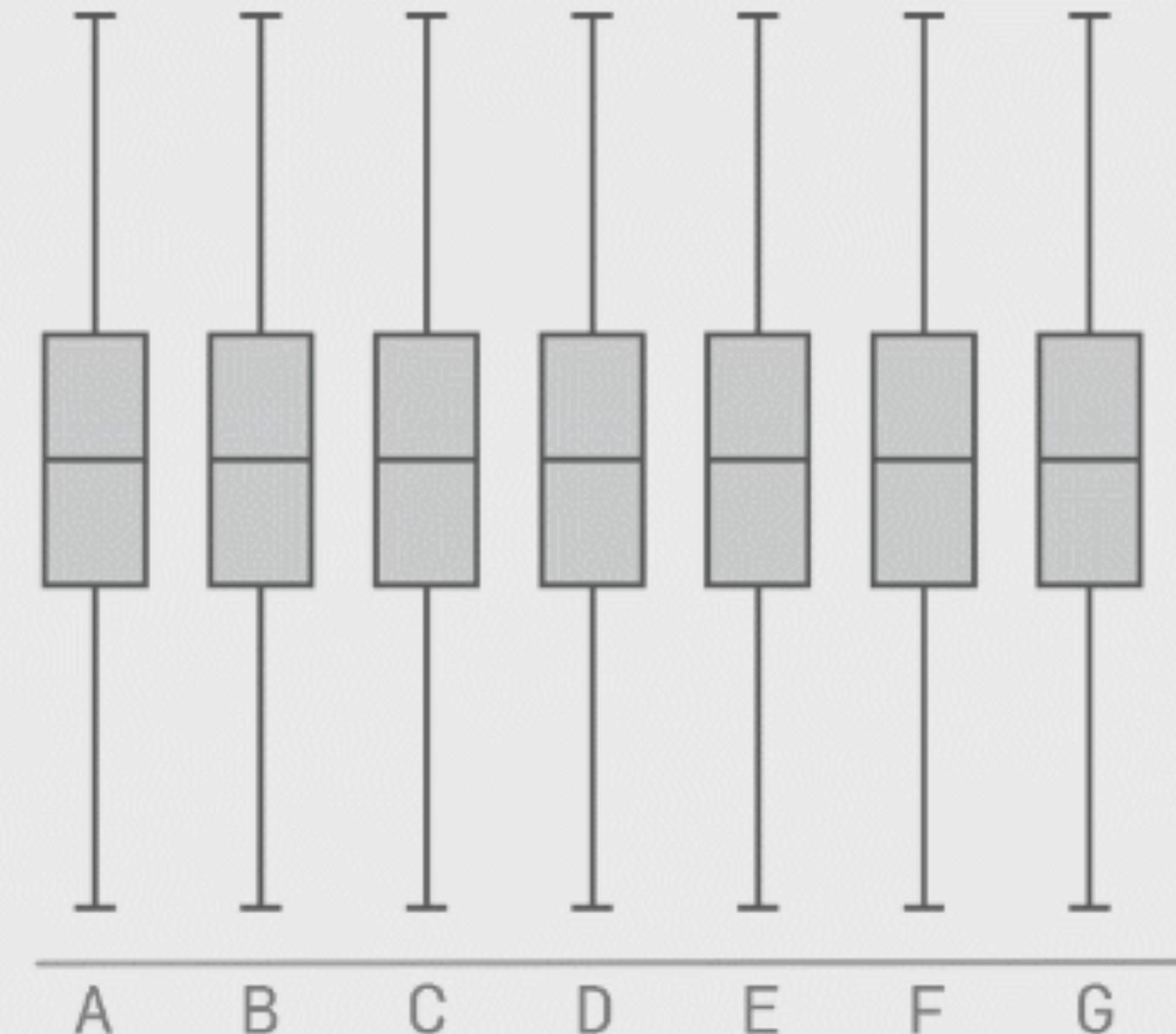
[Newman & Scholl, 2012]  
[Correll & Gleicher, 2014]

# Visualizing Uncertainty: Visual Variables

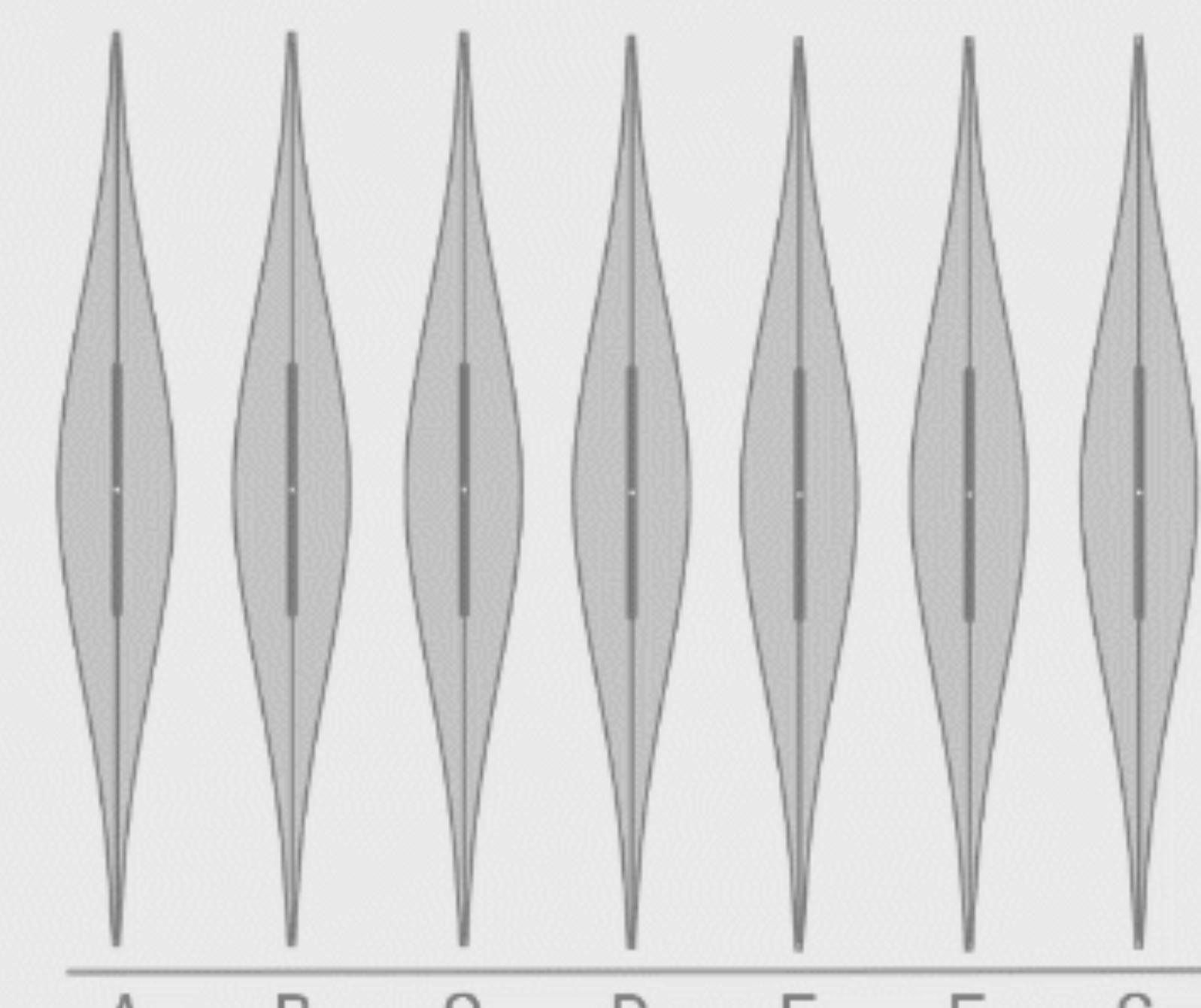
**Raw Data**



**Box-plot of the Data**



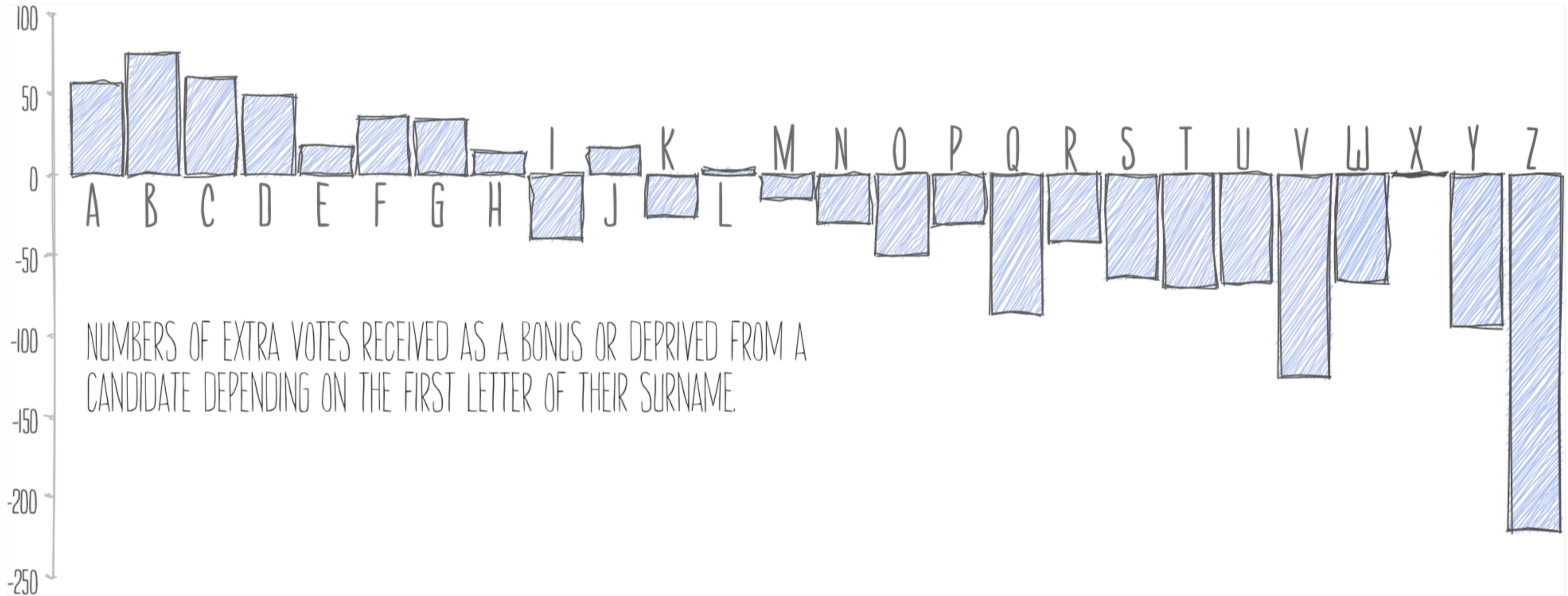
**Violin-plot of the Data**



[Matejka & Fitzmaurice, 2017]

[Correll & Gleicher, 2014]

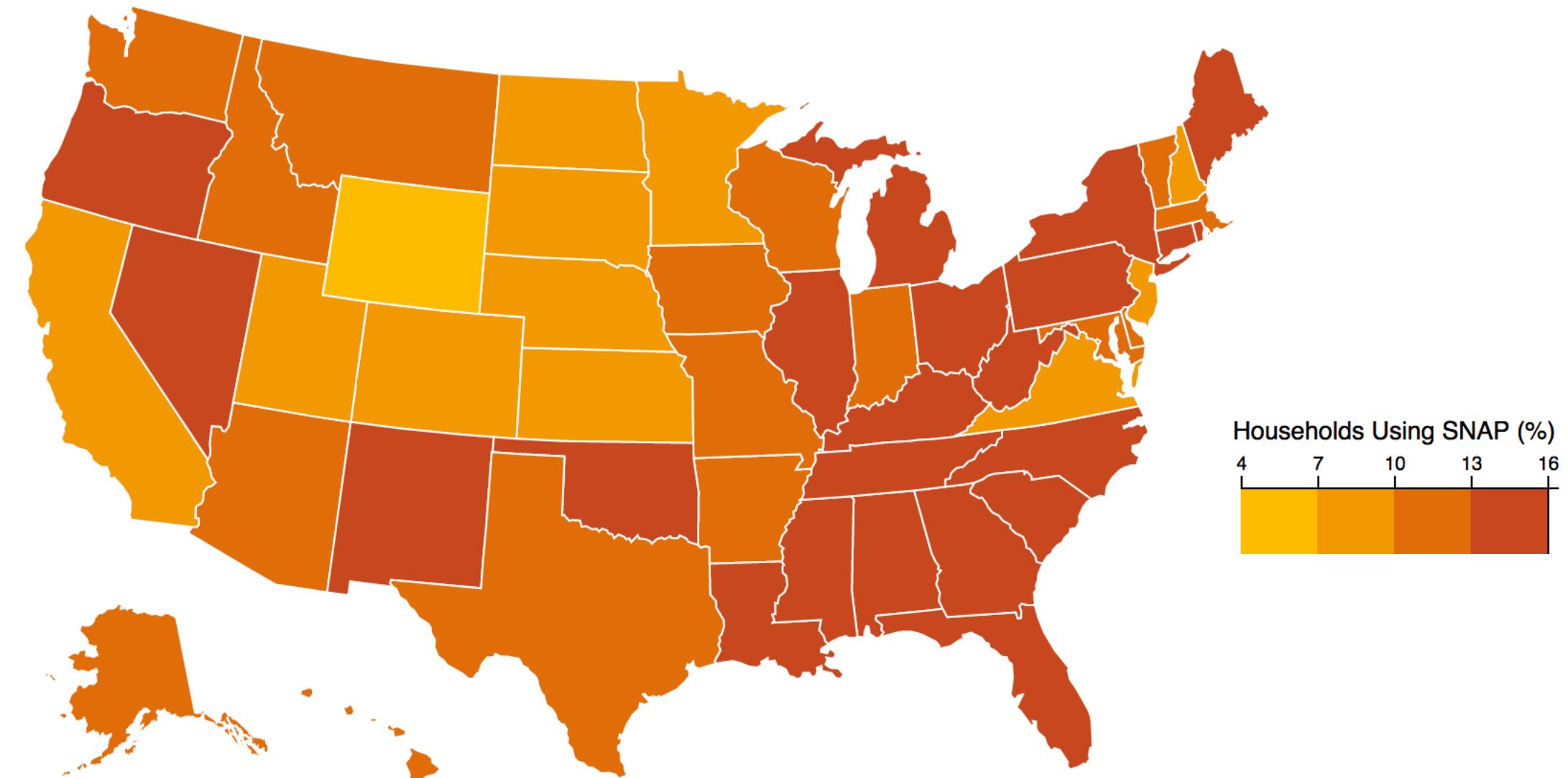
# Visualizing Uncertainty: Visual Variables



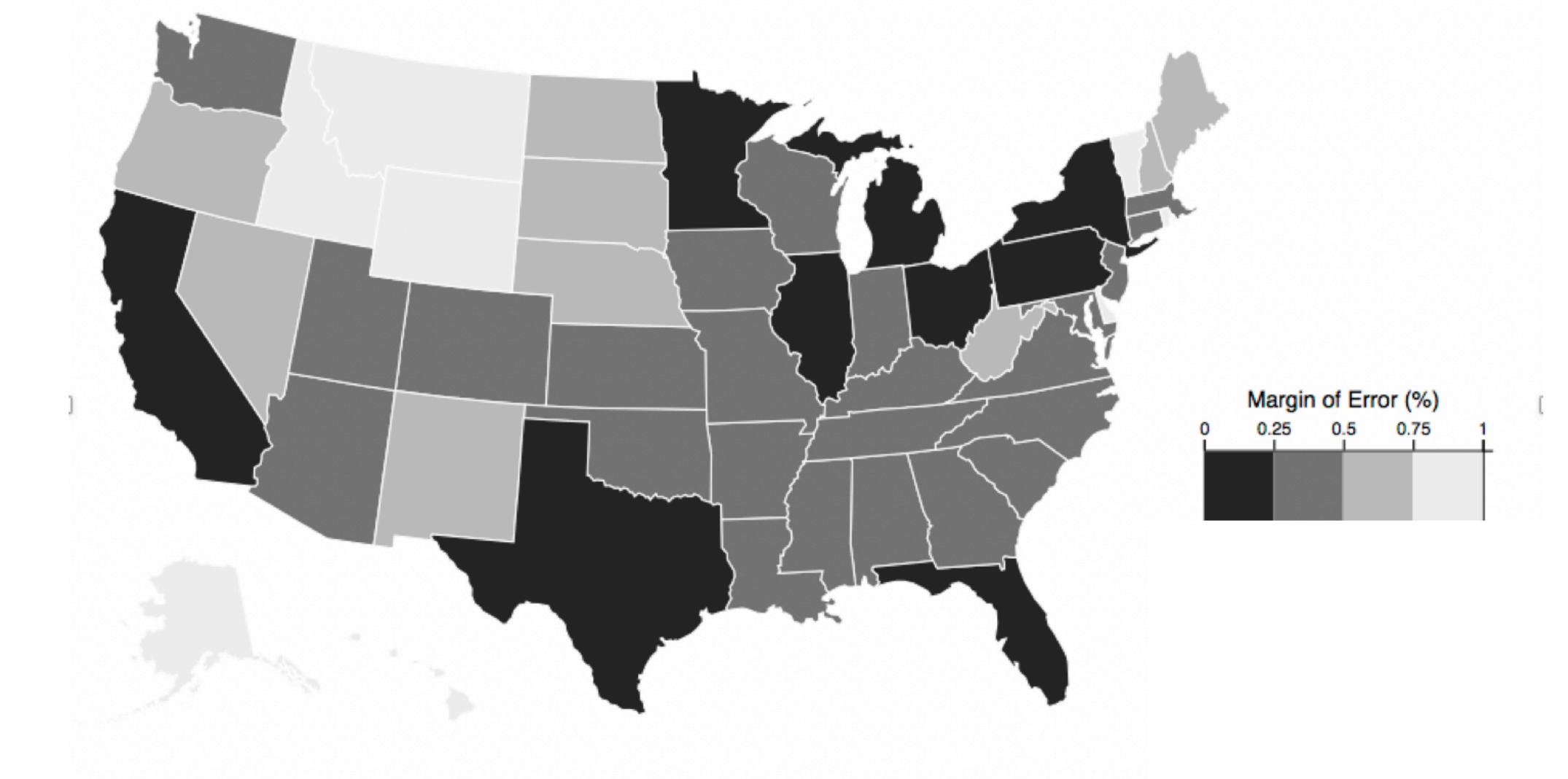
[Wood et al., 2012]

[Boukhelifa et al., 2012]

# Visualizing Uncertainty: Visual Variables



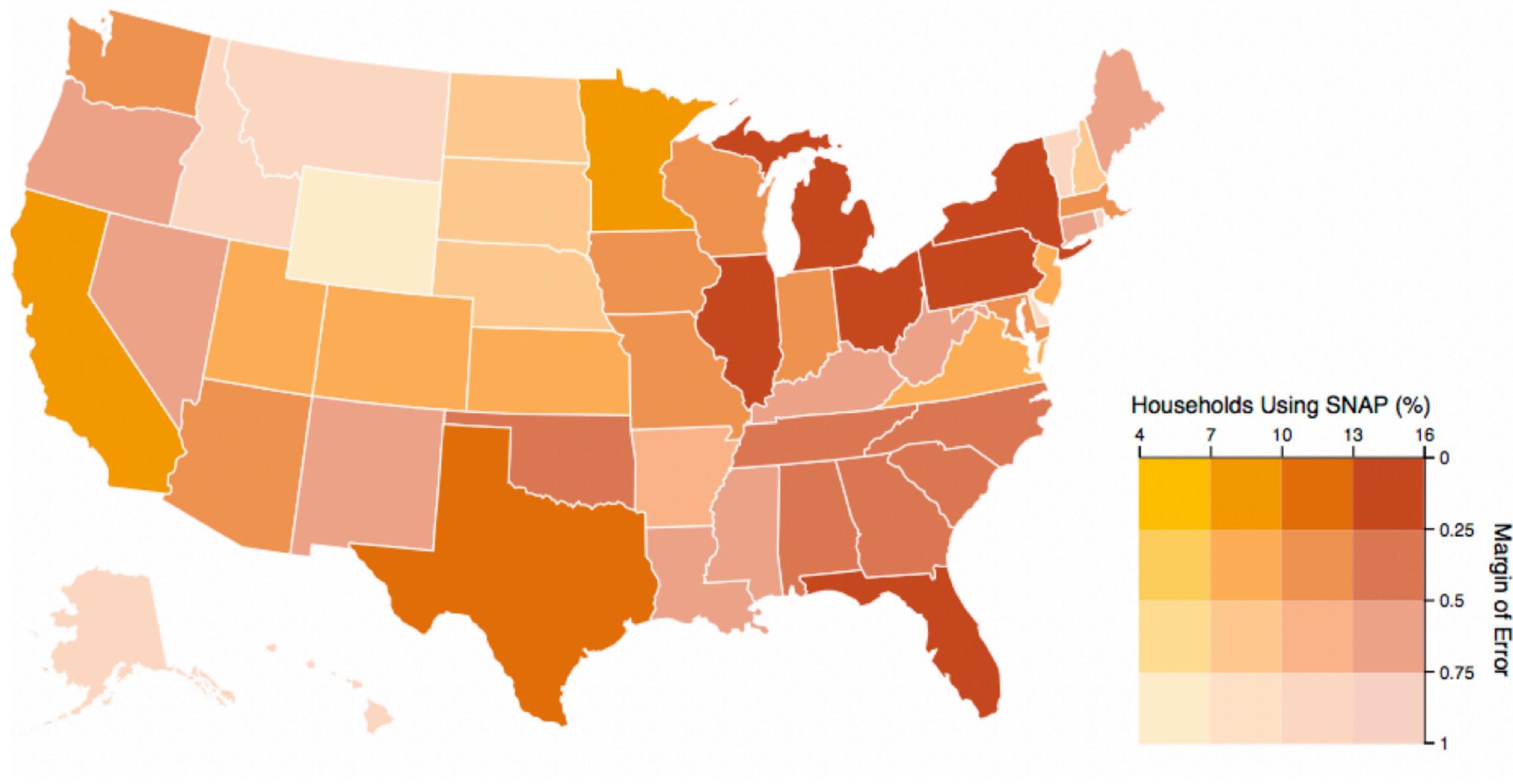
Data Map



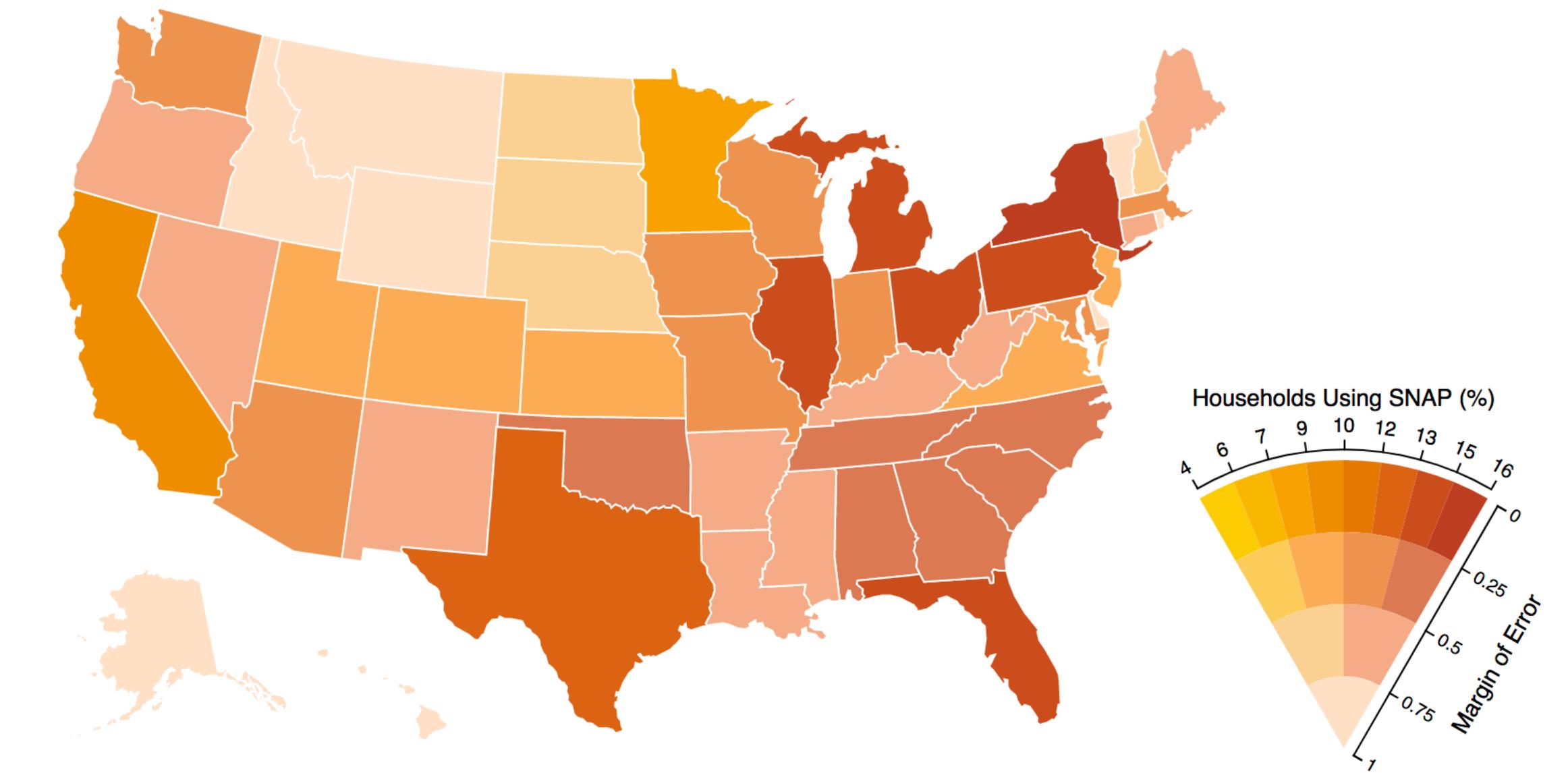
Uncertainty Map

[Correll, Moritz, & Heer, 2018]

# Visualizing Uncertainty: Visual Variables



Bivariate Map (Data + Uncertainty)



Value-Suppressing Uncertainty Map

[Correll, Moritz, & Heer, 2018]

# Visualizing Uncertainty: Set of “Draws”

The New York Times

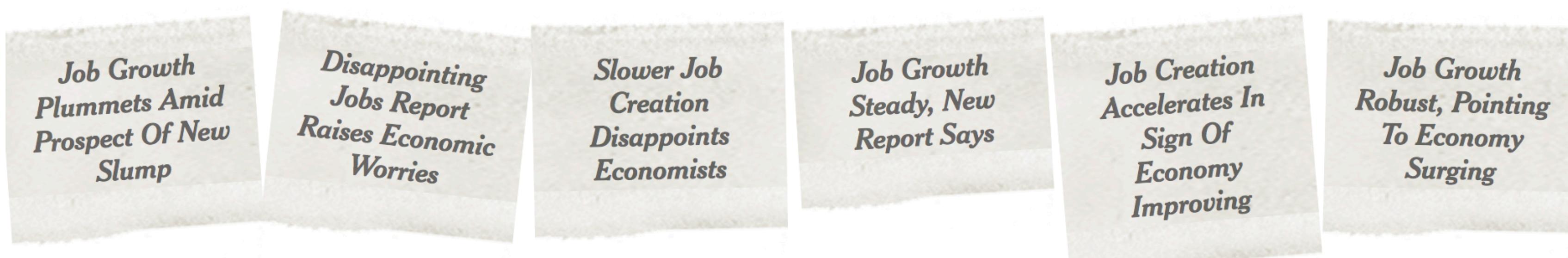
: The Upshot

STATISTICAL NOISE

## How Not to Be Misled by the Jobs Report

If the economy actually added 150,000 jobs last month, it would be possible to see any of these headlines:

The jobs number is just an estimate, and it comes with uncertainty.



Under 55,000 jobs  
4% chance

55,000 to 110,000  
19% chance

110,000 to 140,000  
19% chance

160,000 to 190,000  
19% chance

190,000 to 245,000  
19% chance

245,000+  
4% chance

# Visualizing Uncertainty: Set of “Draws”

The New York Times

TheUpshot

STATISTICAL NOISE

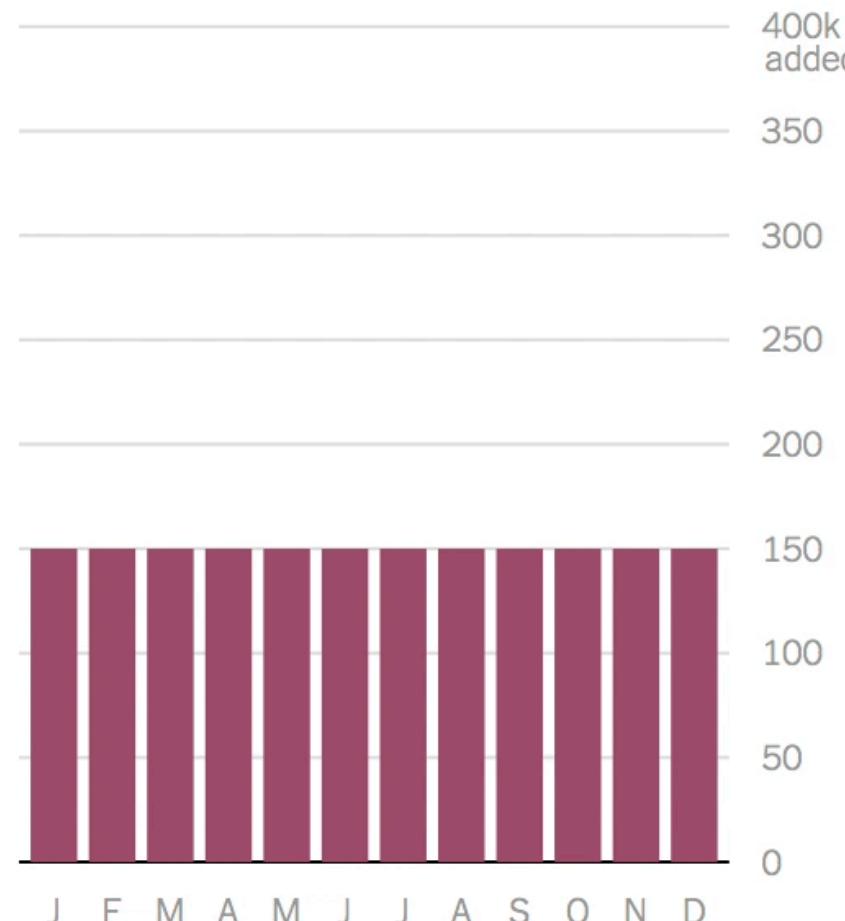
## How Not to Be Misled by the Jobs Report

If the economy actually added 150,000 jobs last month, it would be possible to see any of these headlines:

The jobs number is just an estimate, and it comes with uncertainty.

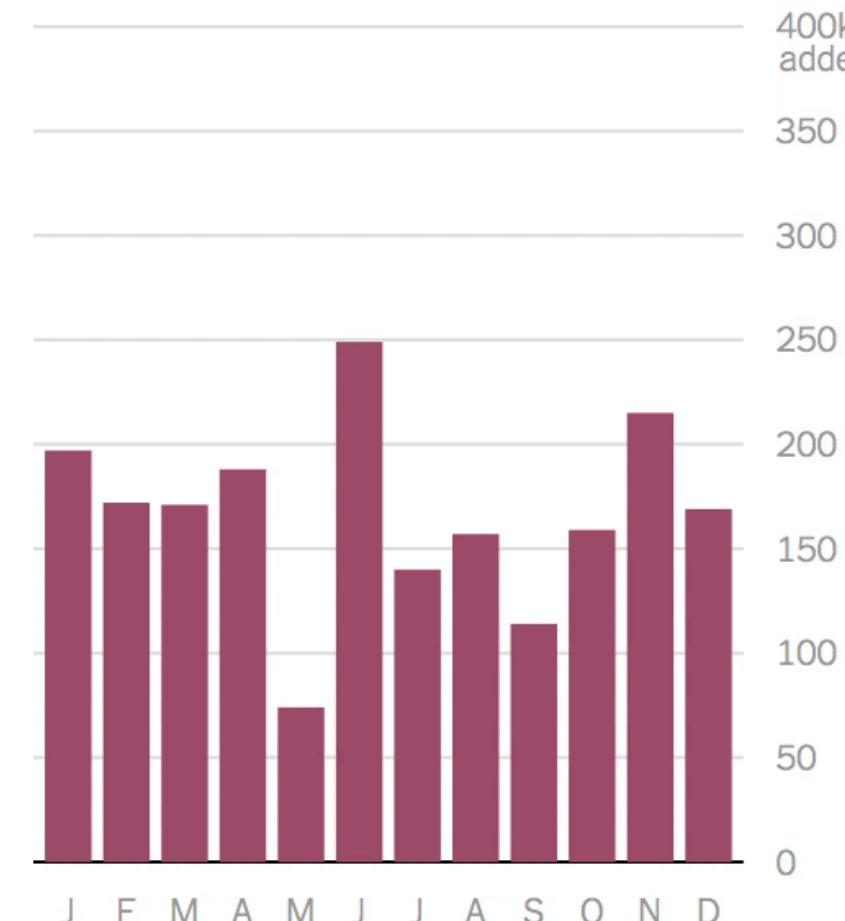


If job growth **were actually steady**  
over the last 12 months...

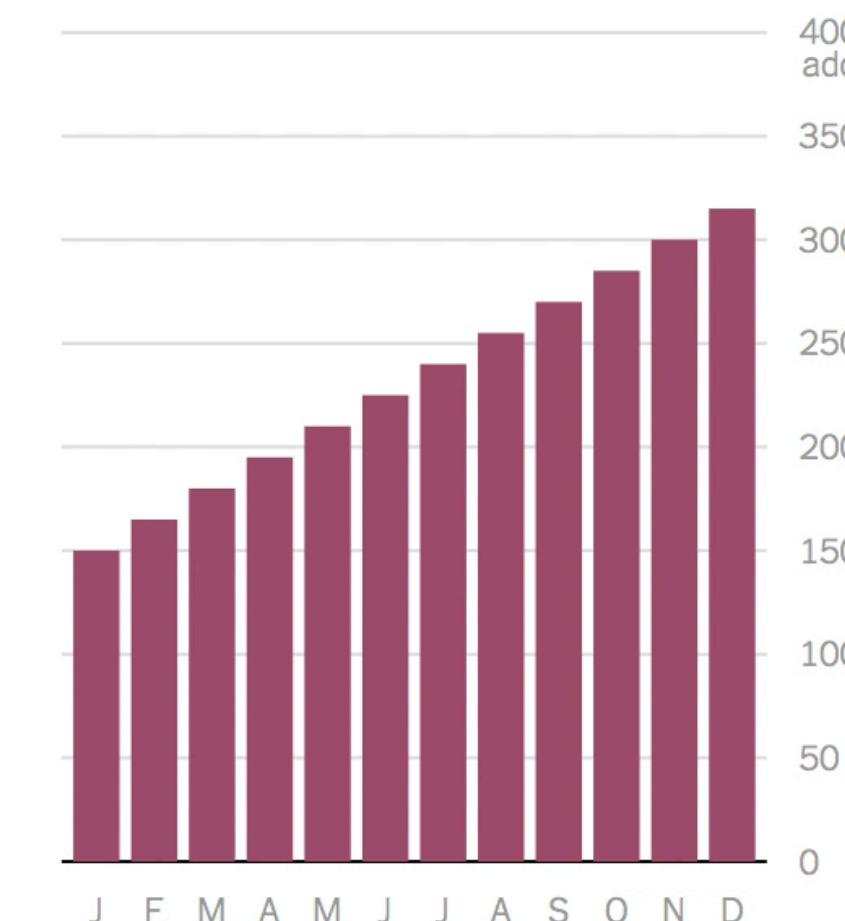


...the jobs report  
could look like this:

Play

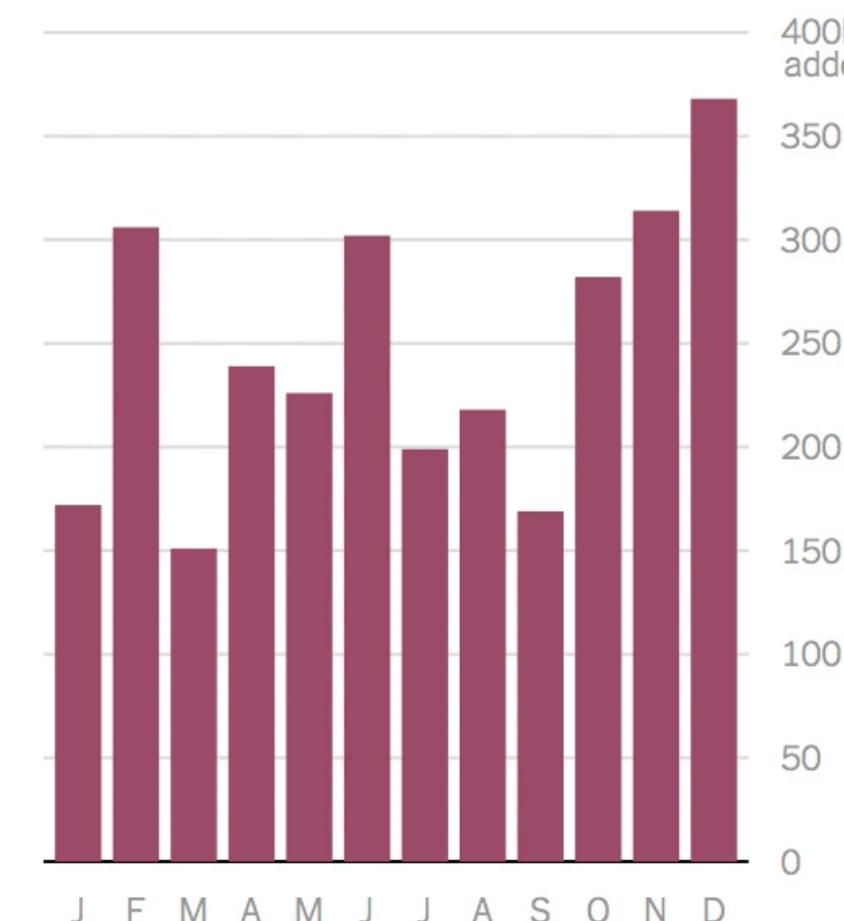


If job growth **had been accelerating...**

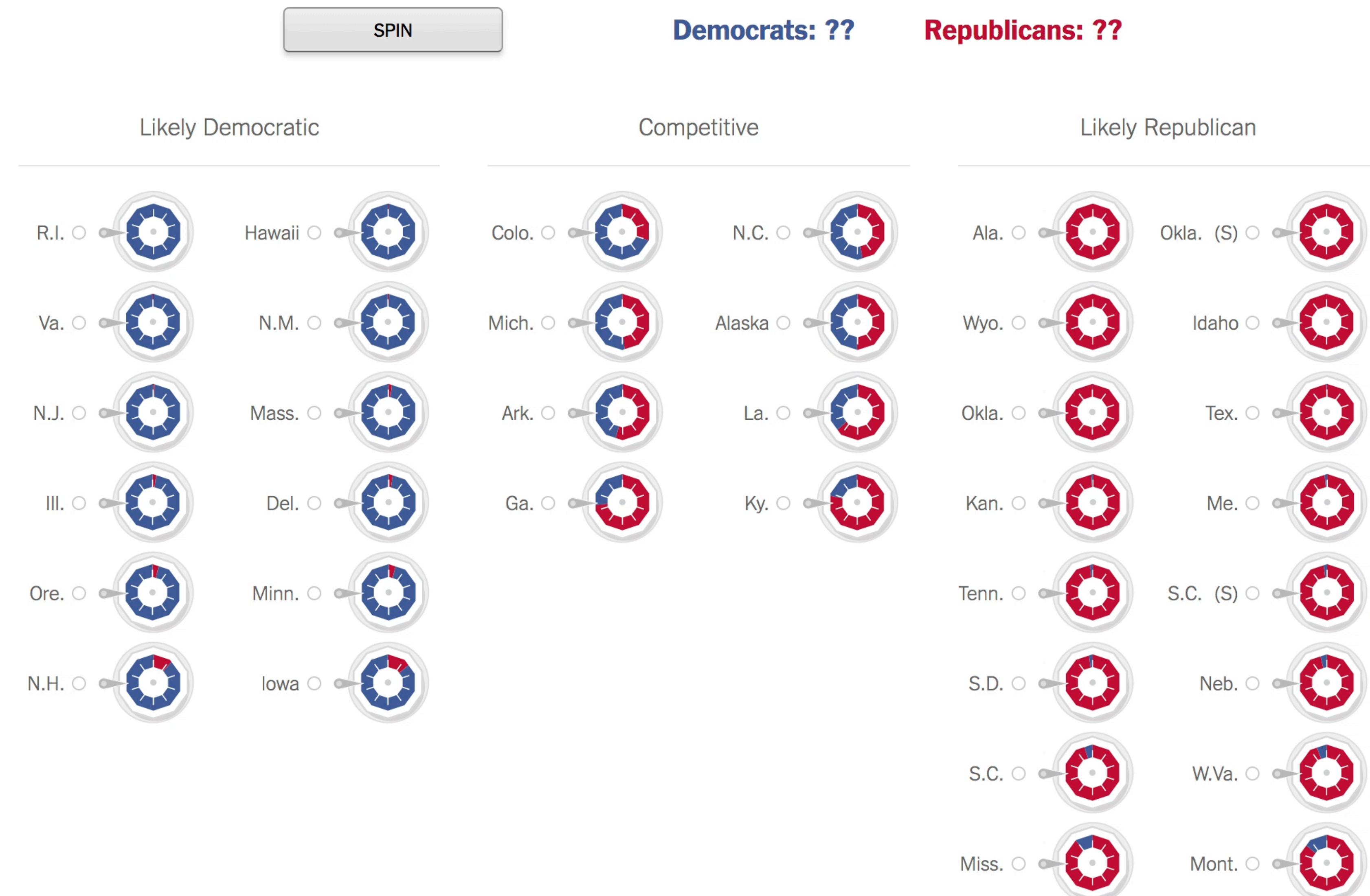


...the jobs report  
could look like this:

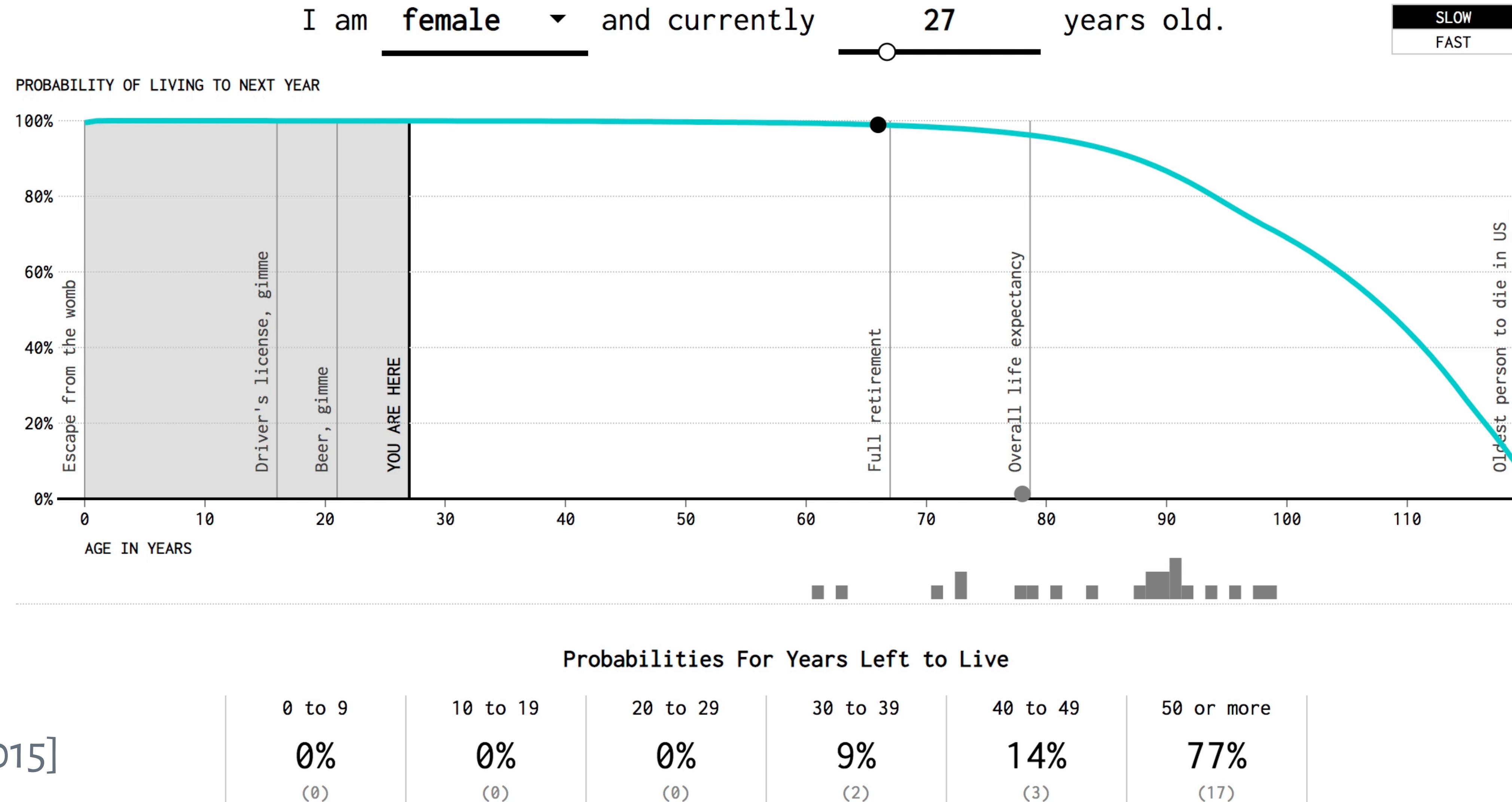
Play



# Visualizing Uncertainty: Set of “Draws”



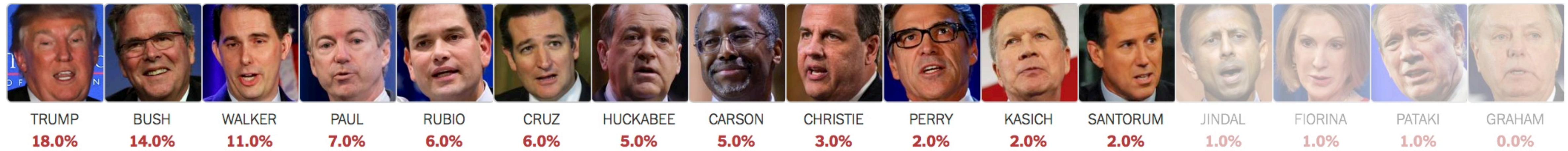
# Visualizing Uncertainty: Set of “Draws”



# Visualizing Uncertainty: Set of “Draws”

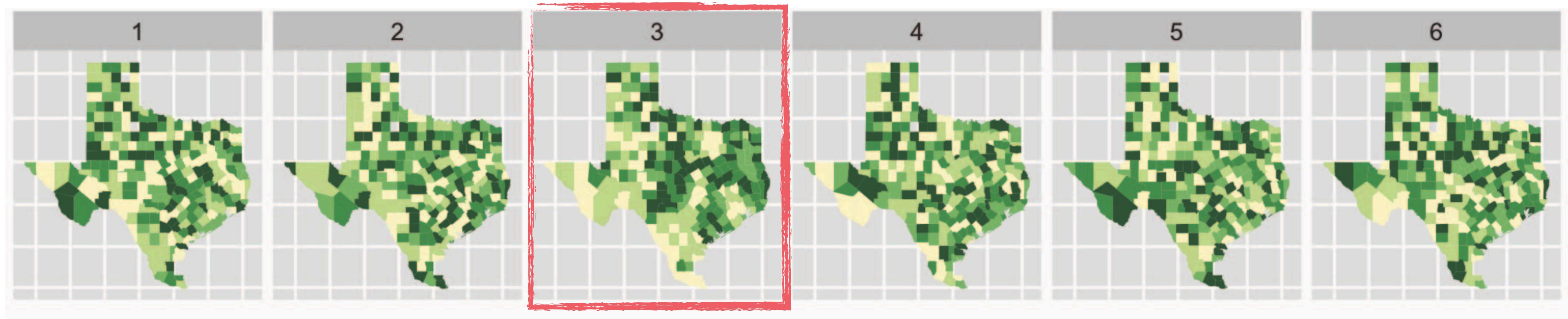
Here's a simulation of who could be in and who could be out if the candidates' averages were rounded to the nearest whole number.

**If the averages are correct, but rounding is to the nearest whole number:**



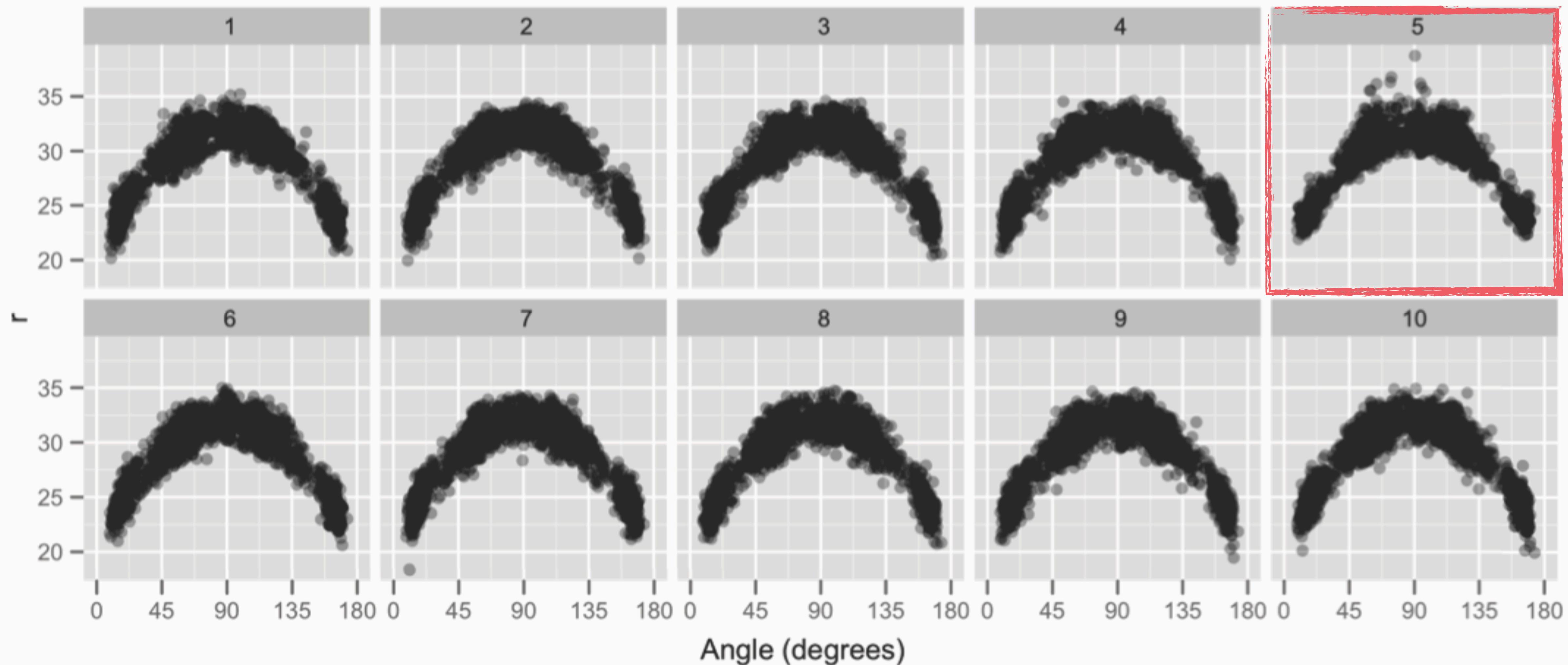
Rounding to fewer decimal places could be welcome news for candidates on the cusp like Mr. Santorum (who has already called the debate rules “a miscarriage”), Mr. Kasich or Mr. Jindal.

# Rorschach and Visual Lineups



[Wickham et al., 2010]

# Rorschach and Visual Lineups



[Wickham et al., 2010]

Error and uncertainty can be difficult to understand, and can require a background in statistics or high data literacy.

People have cognitive and perceptual biases that can yield poor or error-prone decision-making when working with uncertain data.