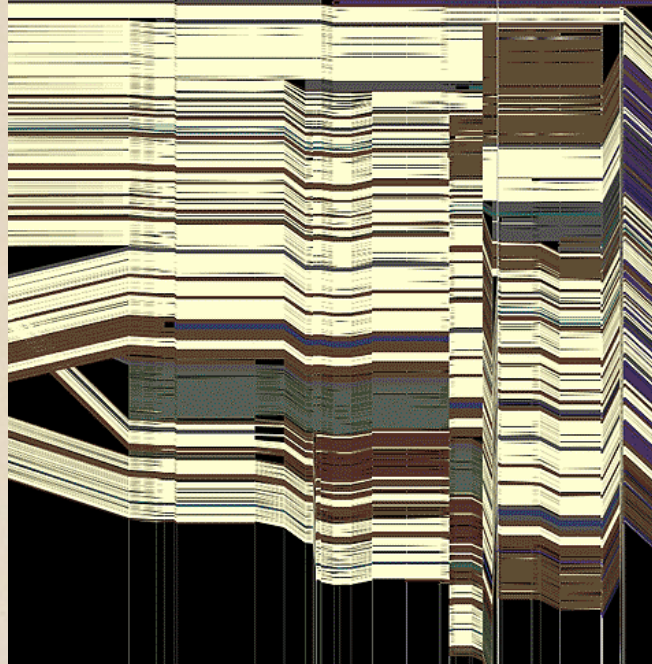
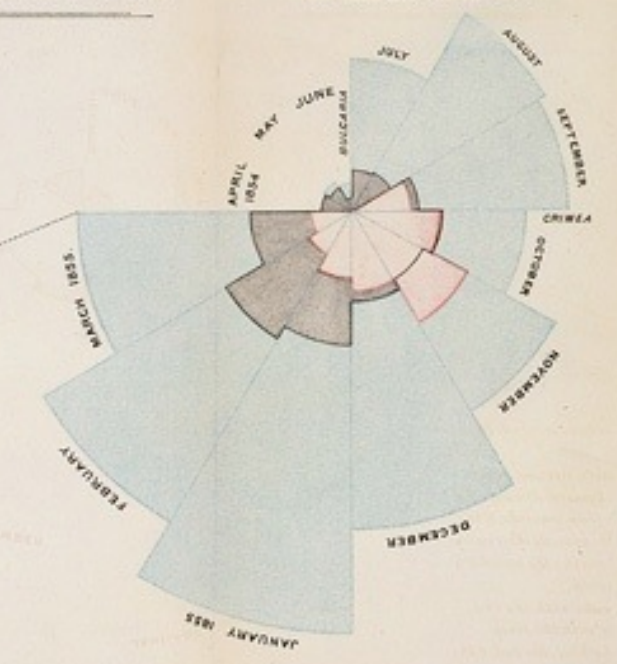


CSE 412 - Intro to Data Visualization

# Uncertainty Visualization



Jane Hoffswell University of Washington

(with significant material from Michael Correll)

# Guest Lecture: Scalable Visualization

Fri Mar. 5 - Guest: Jeffrey Heer (UW)

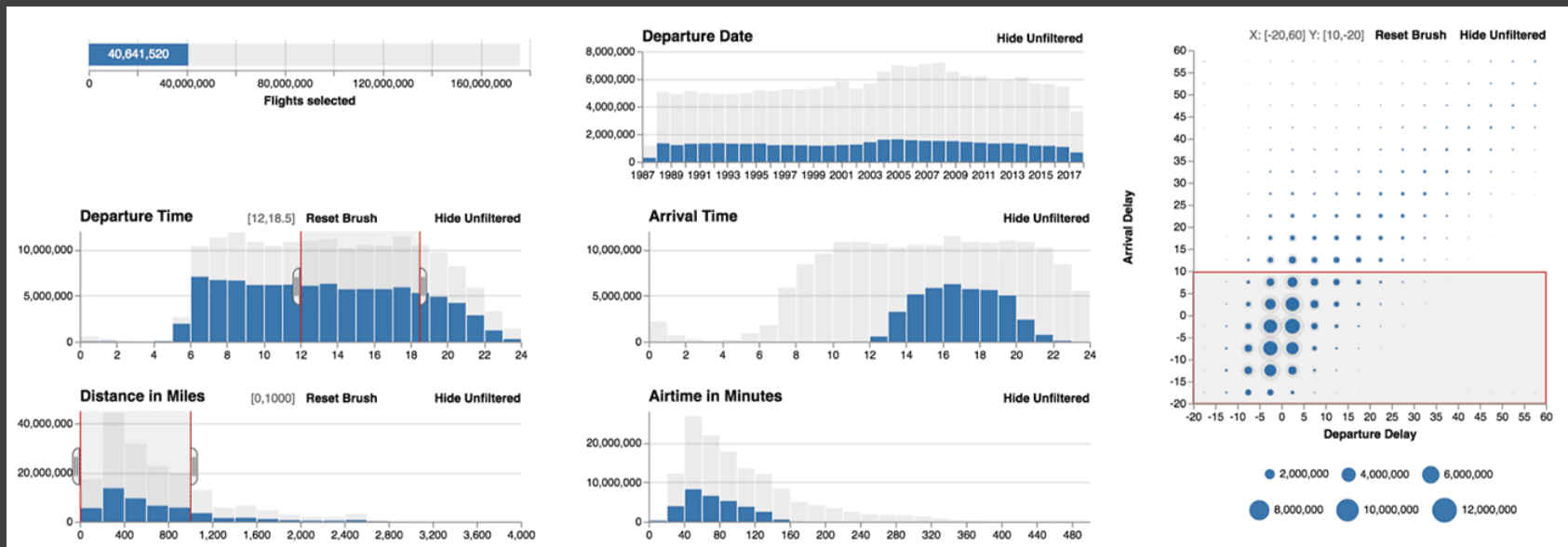


Image: "Falcon: Balancing Interactive Latency and Resolution Sensitivity for Scalable Linked Visualizations." CHI 2019.

# Questions to Answer

What does uncertainty mean?

How should I visualize it?

What can go wrong?

**What does uncertainty mean?**

# What does uncertainty mean?

Doubt

Risk

Variability

Error

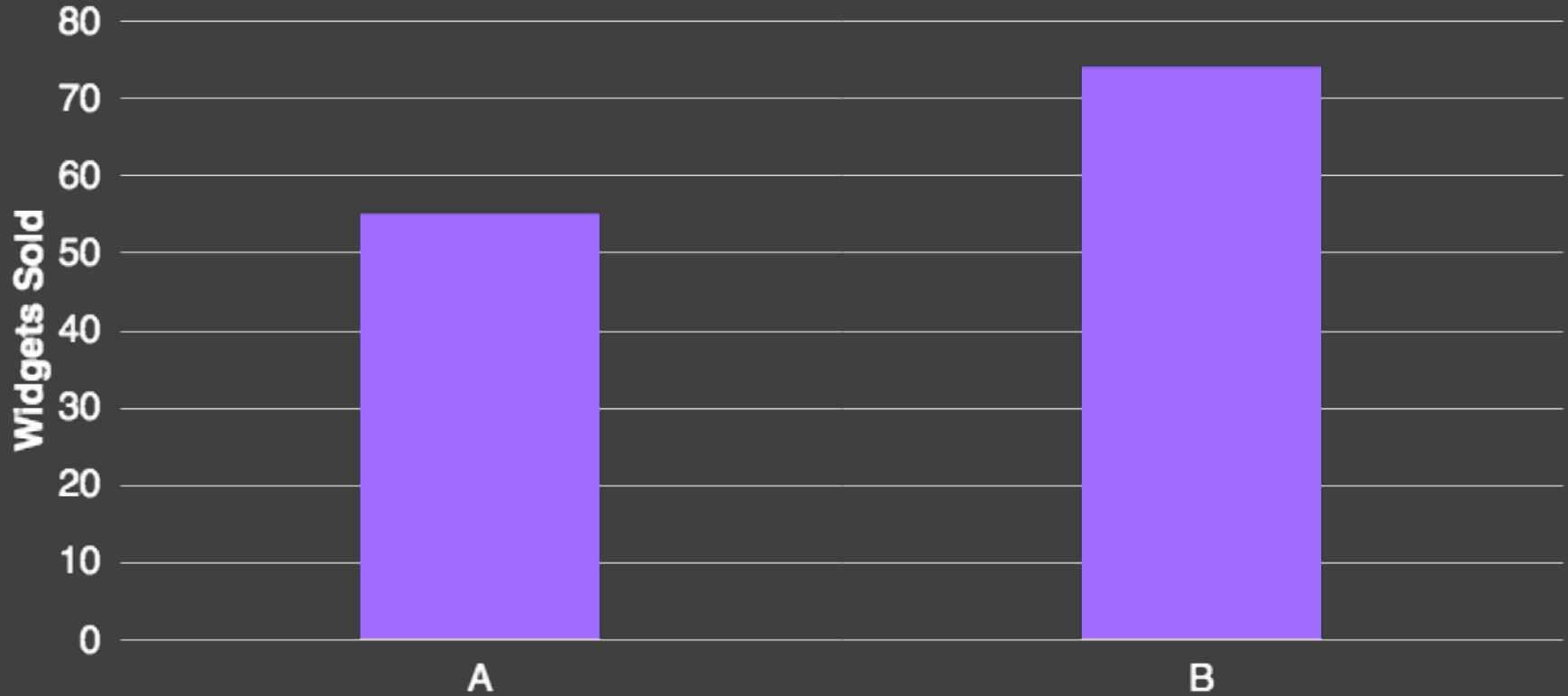
Lack of Knowledge

Hedging

...

# What does uncertainty mean?

Sales of Widgets for Stores A and B



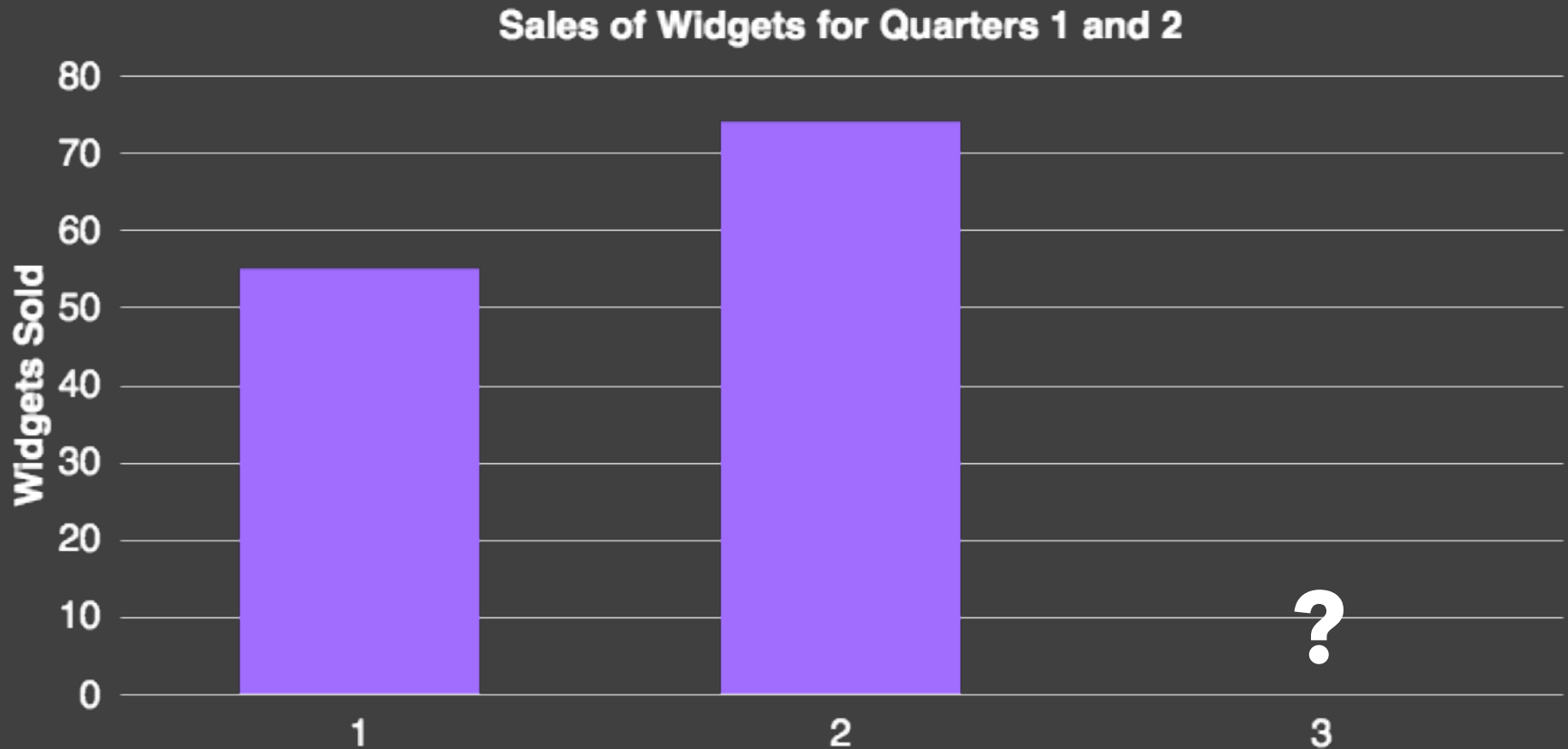
# What does uncertainty mean?



## Measurement Uncertainty:

"We're not exactly sure what the values in the data are."

# What does uncertainty mean?

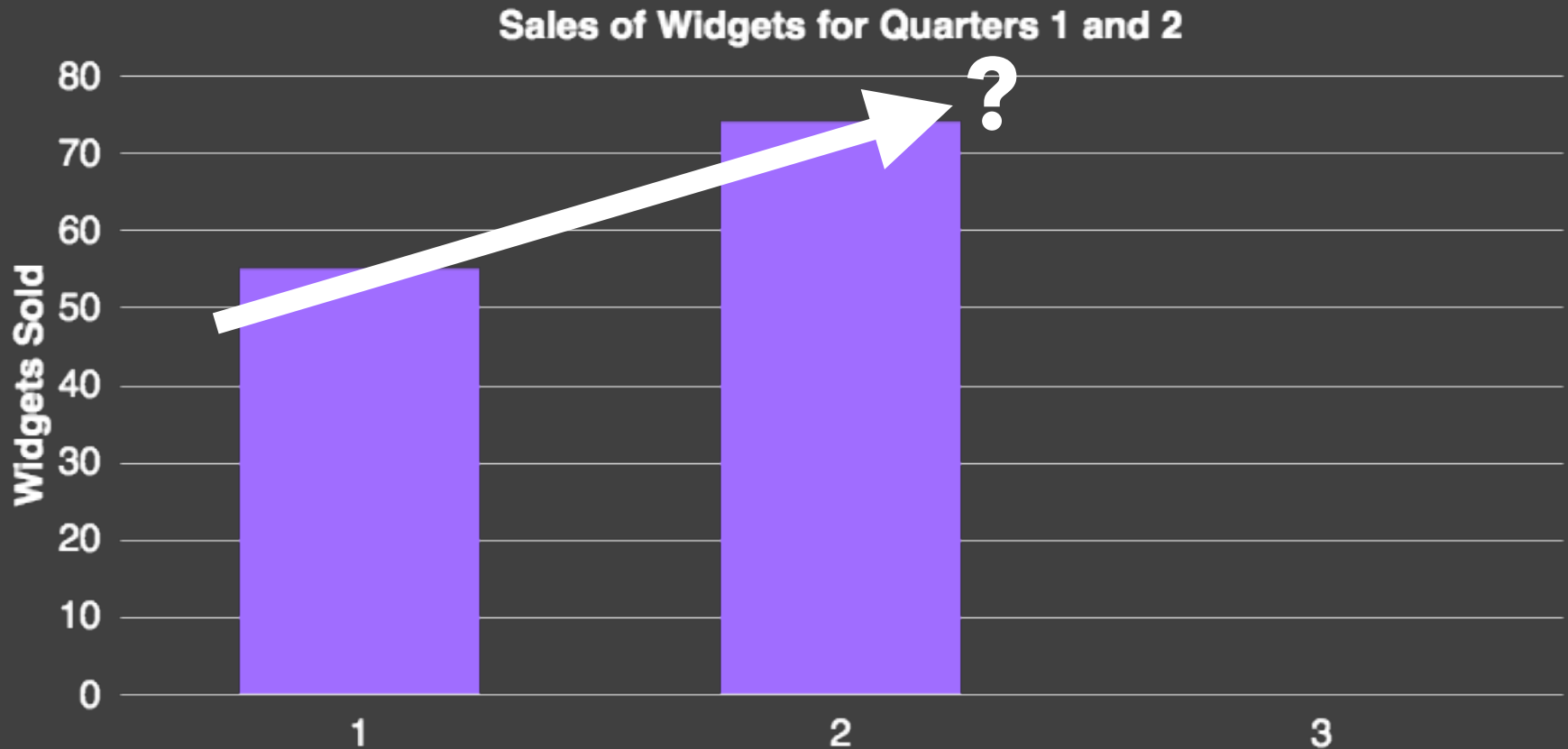


## Forecast Uncertainty:

"We're not exactly sure what will happen to the data next."



# What does uncertainty mean?

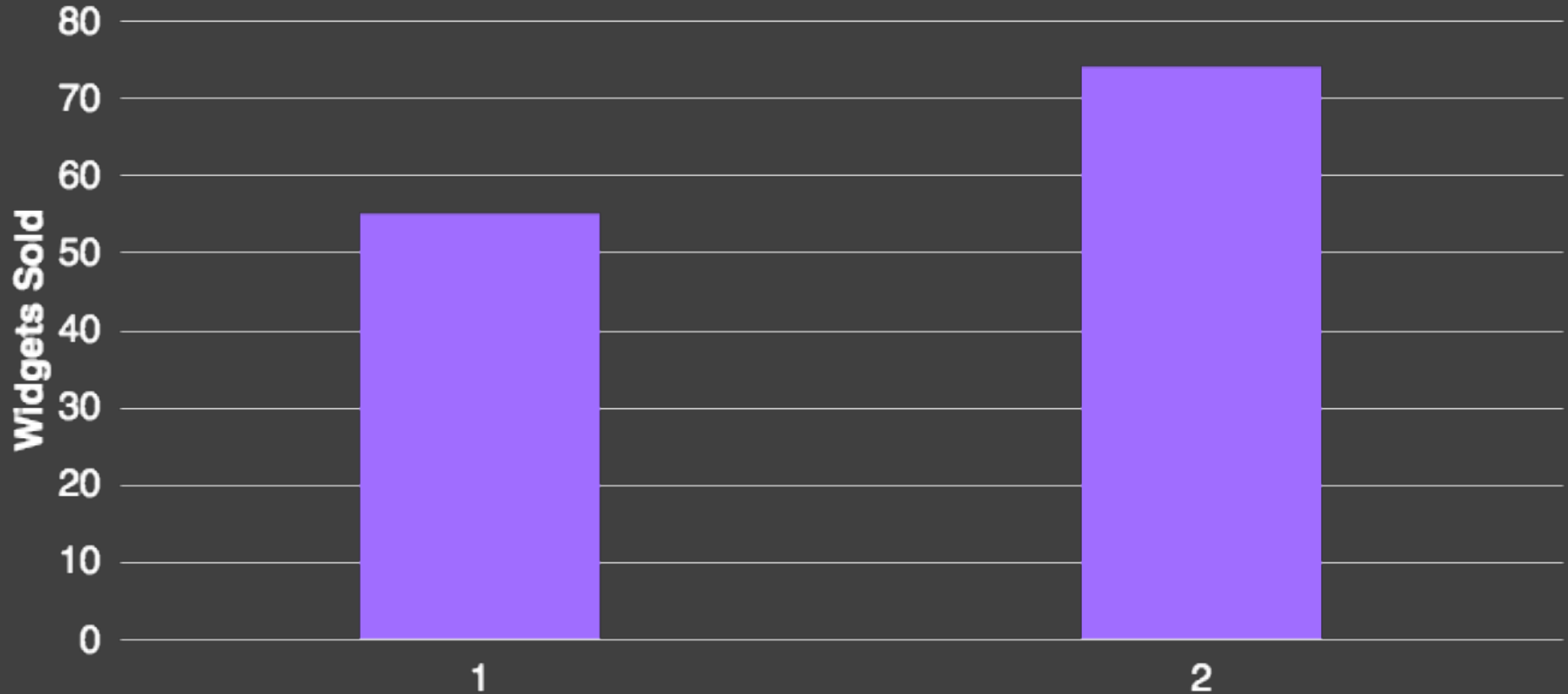


## Model Uncertainty:

"We're not exactly sure how the data fits together."

# What does uncertainty mean?

We should close store A ?



## Decision Uncertainty:

"We're not exactly sure what to do with the data."

# What does uncertainty mean?

## **Measurement Uncertainty:**

"We're not exactly sure what the values in the data are."

## **Forecast Uncertainty:**

"We're not exactly sure what will happen to the data next."

## **Model Uncertainty:**

"We're not exactly sure how the data fits together."

## **Decision Uncertainty:**

"We're not exactly sure what to do with the data."

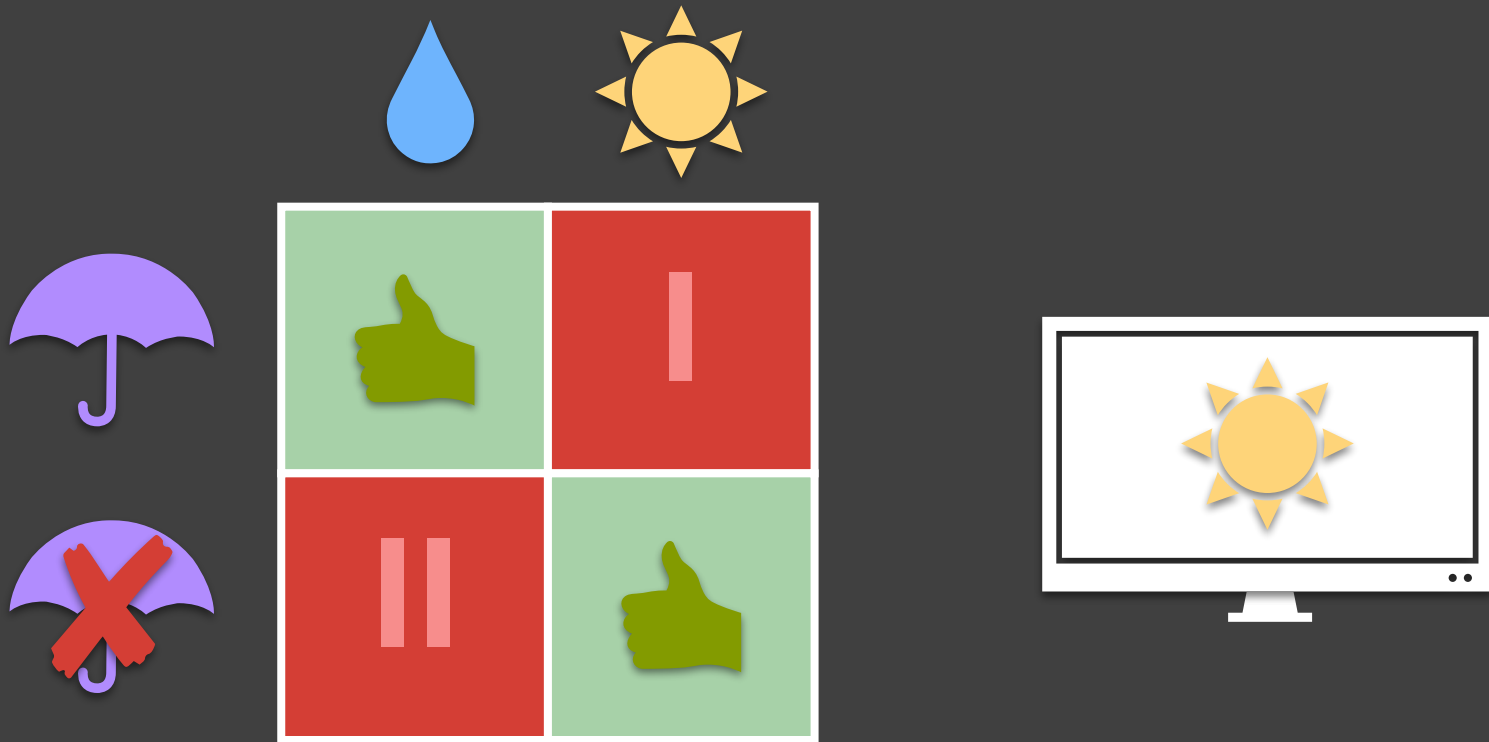
# Type I and II Errors

Should I bring an umbrella today?



# Type I and II Errors

Should I bring an umbrella today?



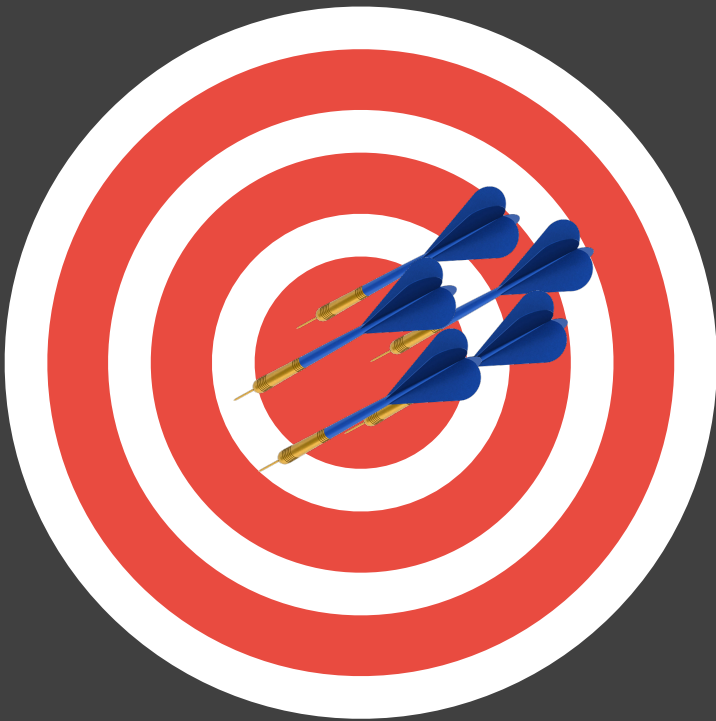
# Precision and Bias

**Precision**



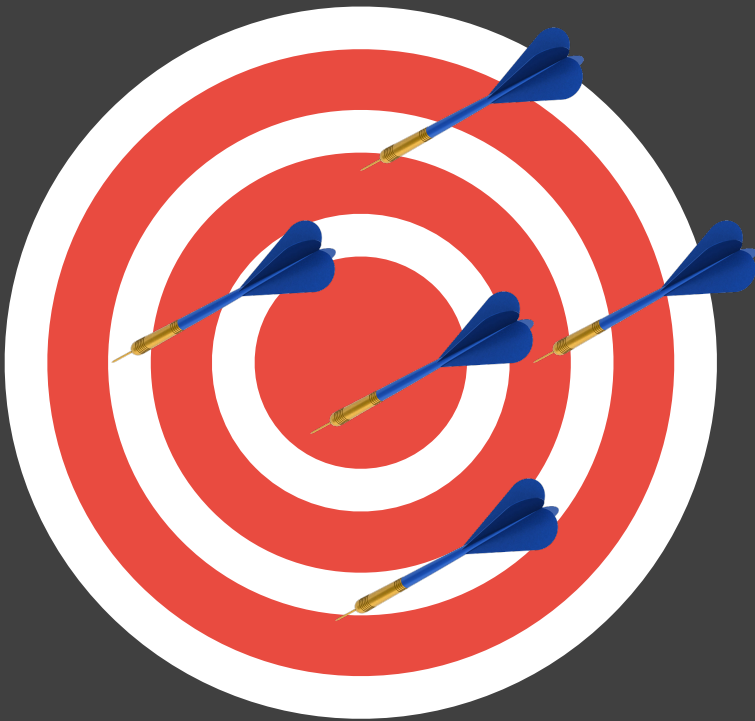
# Precision and Bias

**Precision**



# Precision and Bias

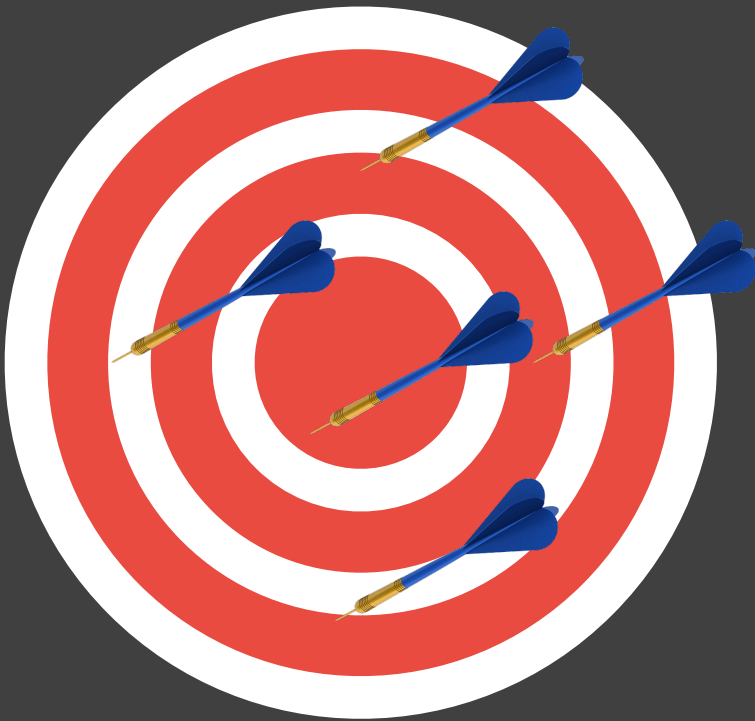
**Precision**





# Precision and Bias

**Precision**

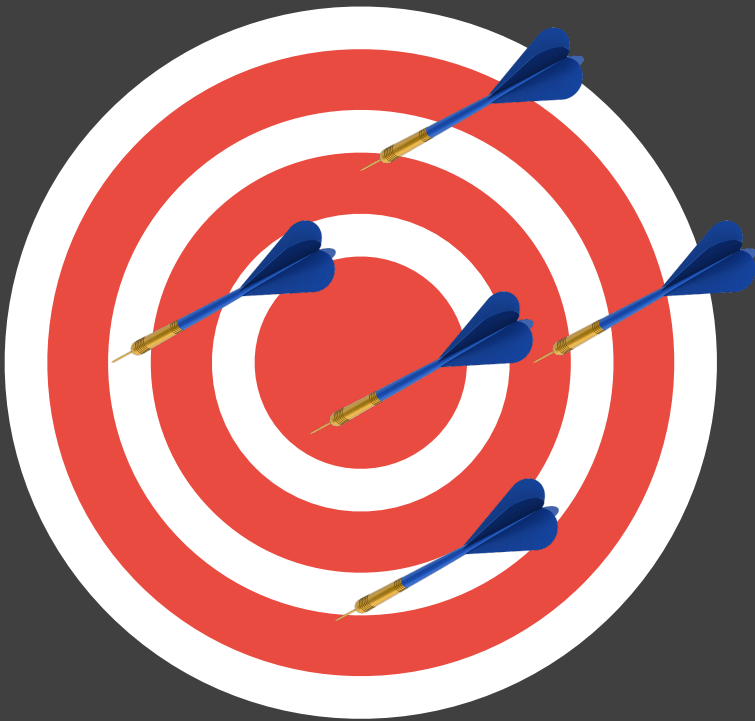


**Accuracy**



# Precision and Bias

**Precision**

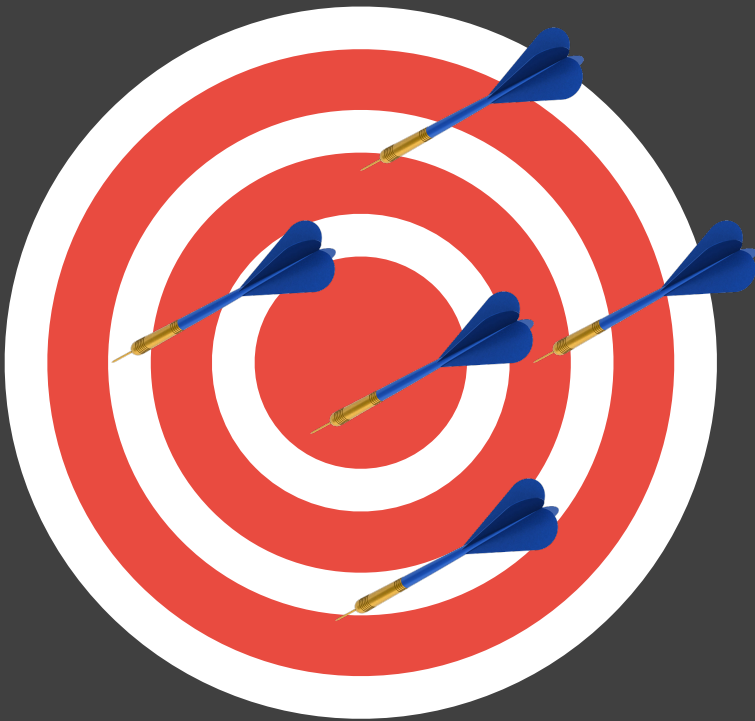


**Accuracy**

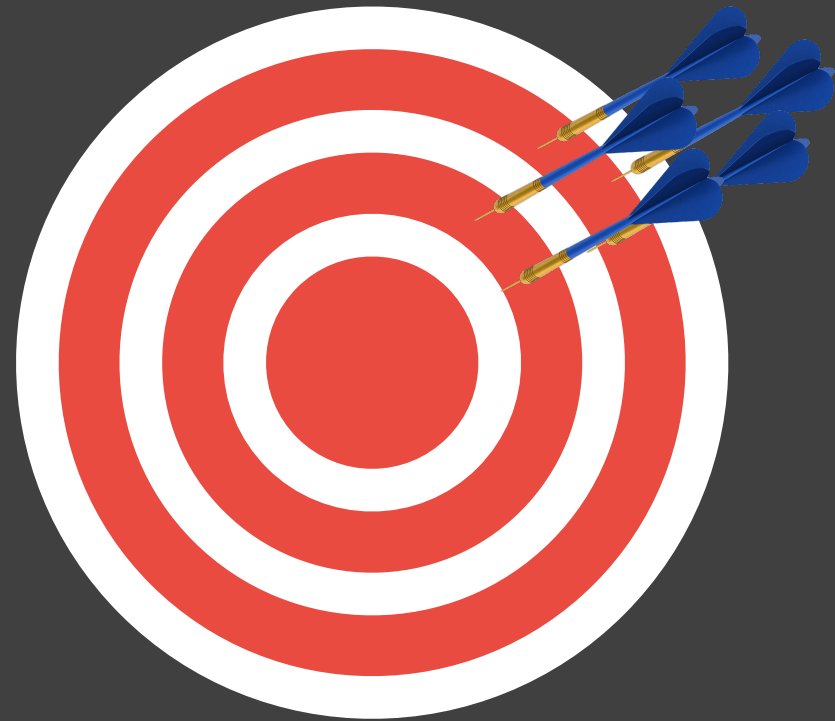


# Precision and Bias

**Precision**

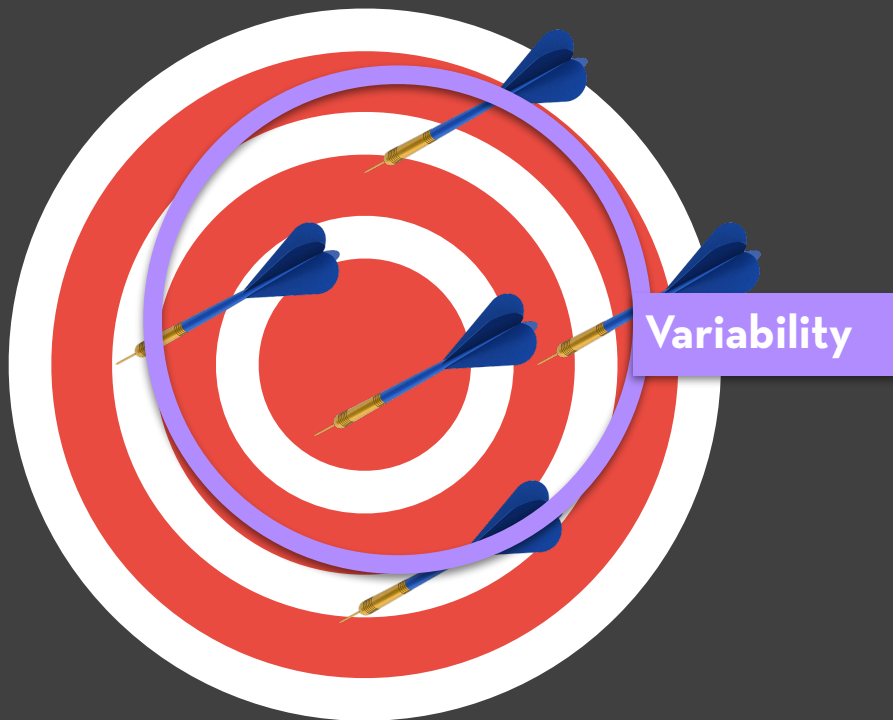


**Accuracy**

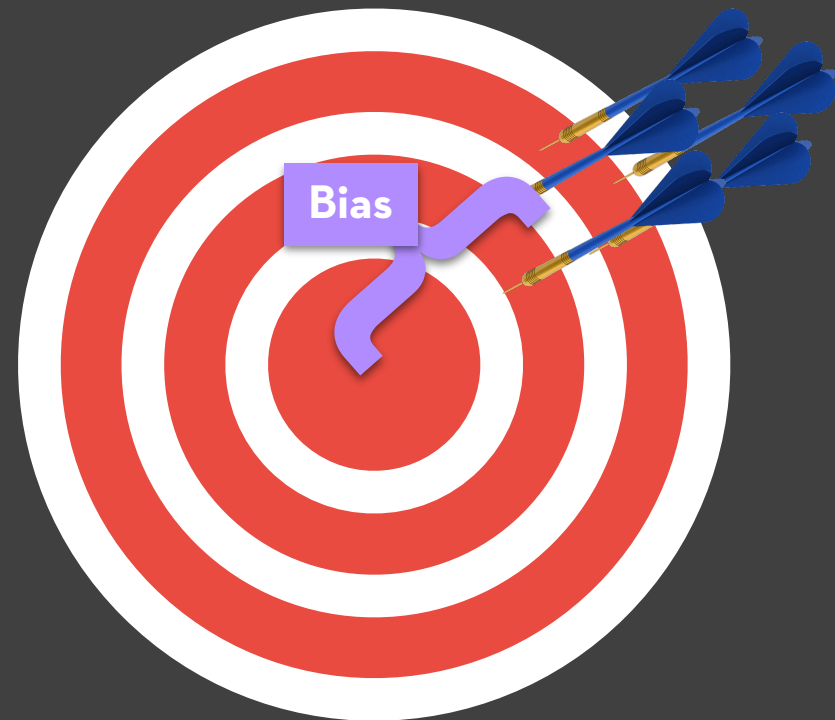


# Precision and Bias

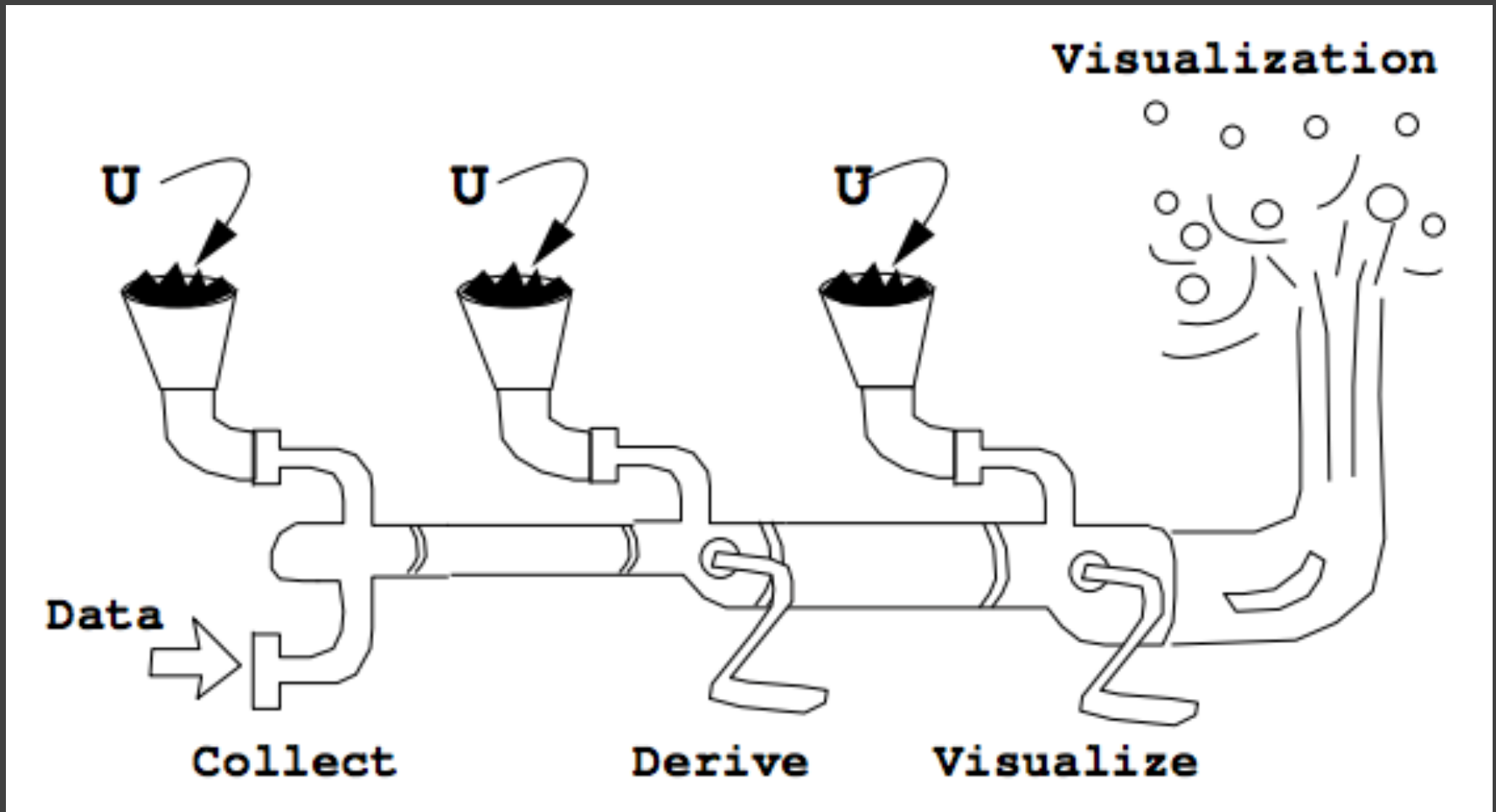
**Precision**



**Accuracy**



# Uncertainty Visualization Pipeline



"Approaches to Uncertainty Visualization." Pang et al. The Visual Computer, 1997.

# 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.

**LOTS OF THINGS**

**How should I visualize uncertainty?**



# Uncertainty Visualization Zoo

## The UncertaintyVis Browser

An Interactive Survey of Uncertainty Visualisation Research Publications.

Styling adapted from Tominski & Aigner's [The TimeViz Browser](#).

# of Research Papers: 286

Search: Title, Abstract, Authors, & Year.

How to use filters:

- Show: Interested!
- Neutral: I don't care.
- Hide: I'm not interested!

Publication Details

Publication Type

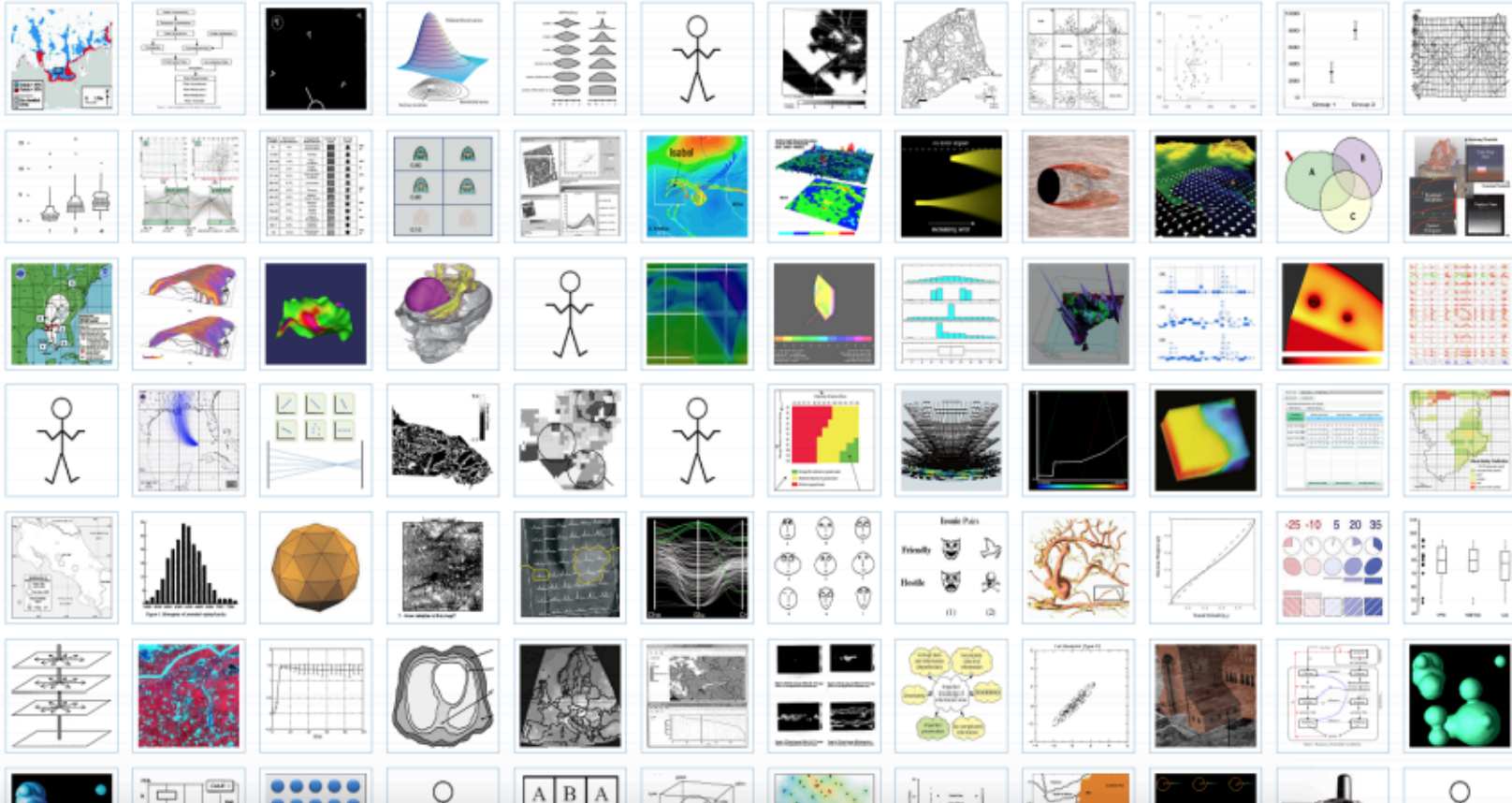
- Journal
- Conference
- Poster
- Thesis
- Tech Report
- Book

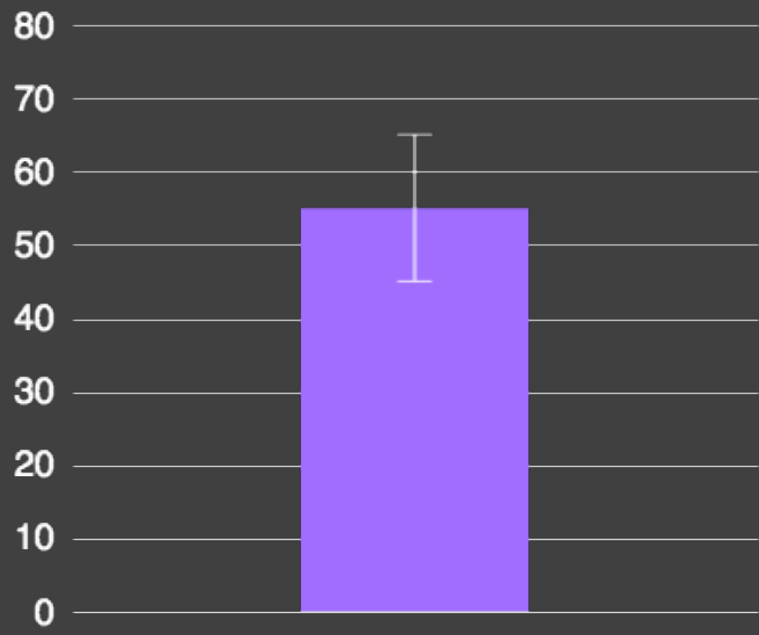
Publication Venue

- Cartography
- Psychology
- Health
- Human Factors
- Decision Scie.
- Info Vis
- Sci Vis
- Cyber Security
- Ubi Comp
- Statistics
- Data Science

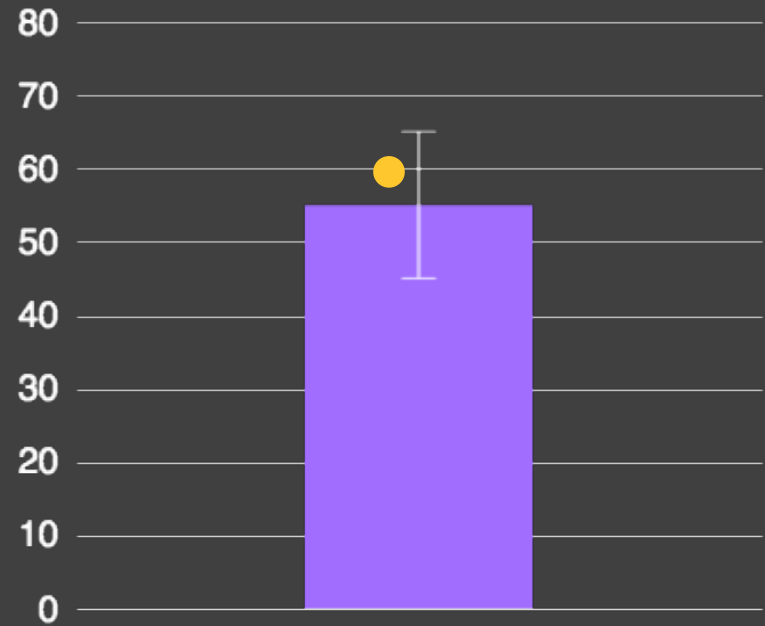
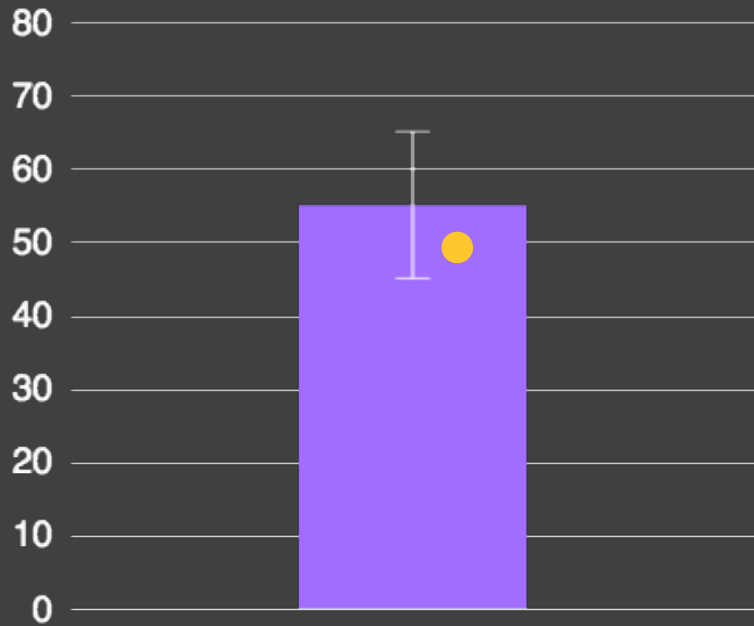
Application Domain

- Aviation
- GIS
- Medical Scie.



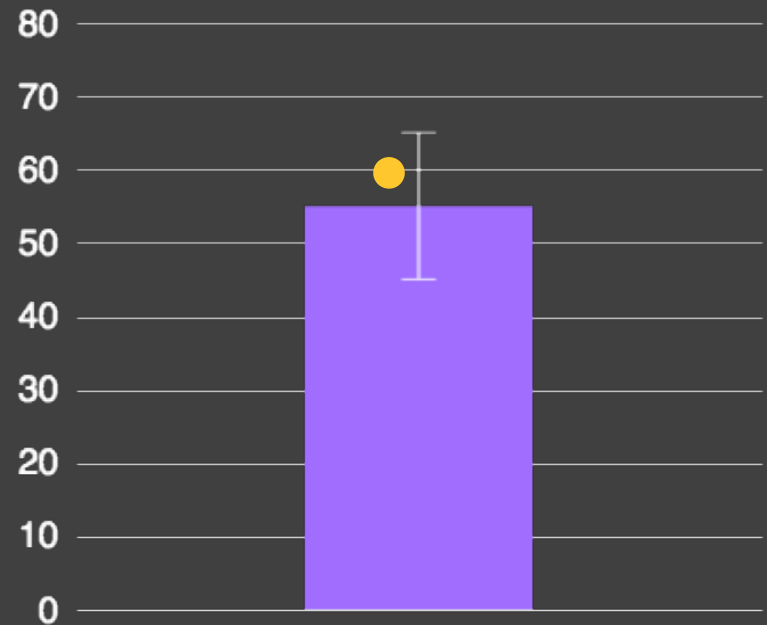
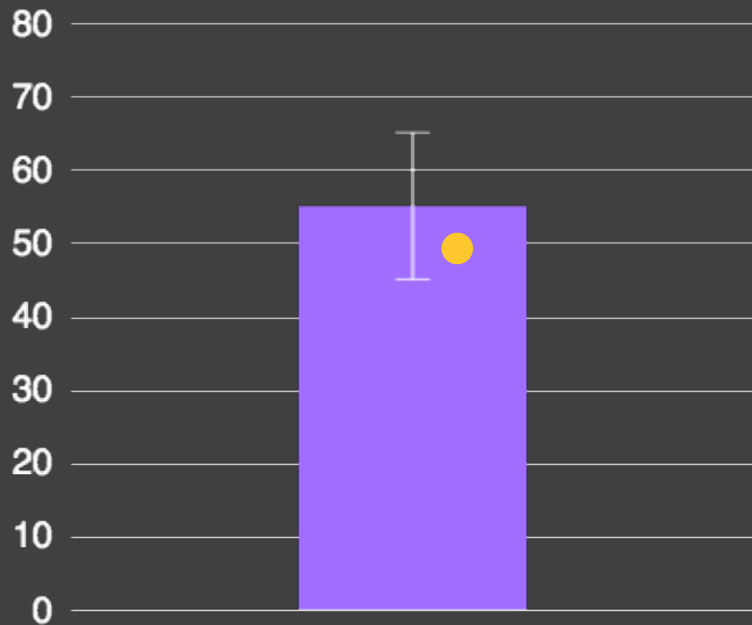


# Which point is more likely?



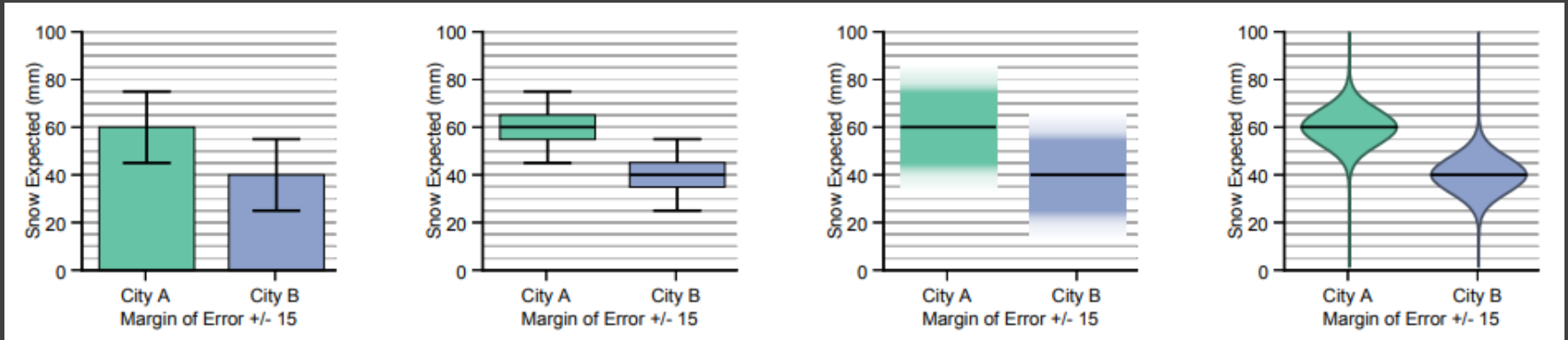
# Within-the-bar Bias

Values "within" the bar are considered more likely









# Intervals

[Correll et al. '14]



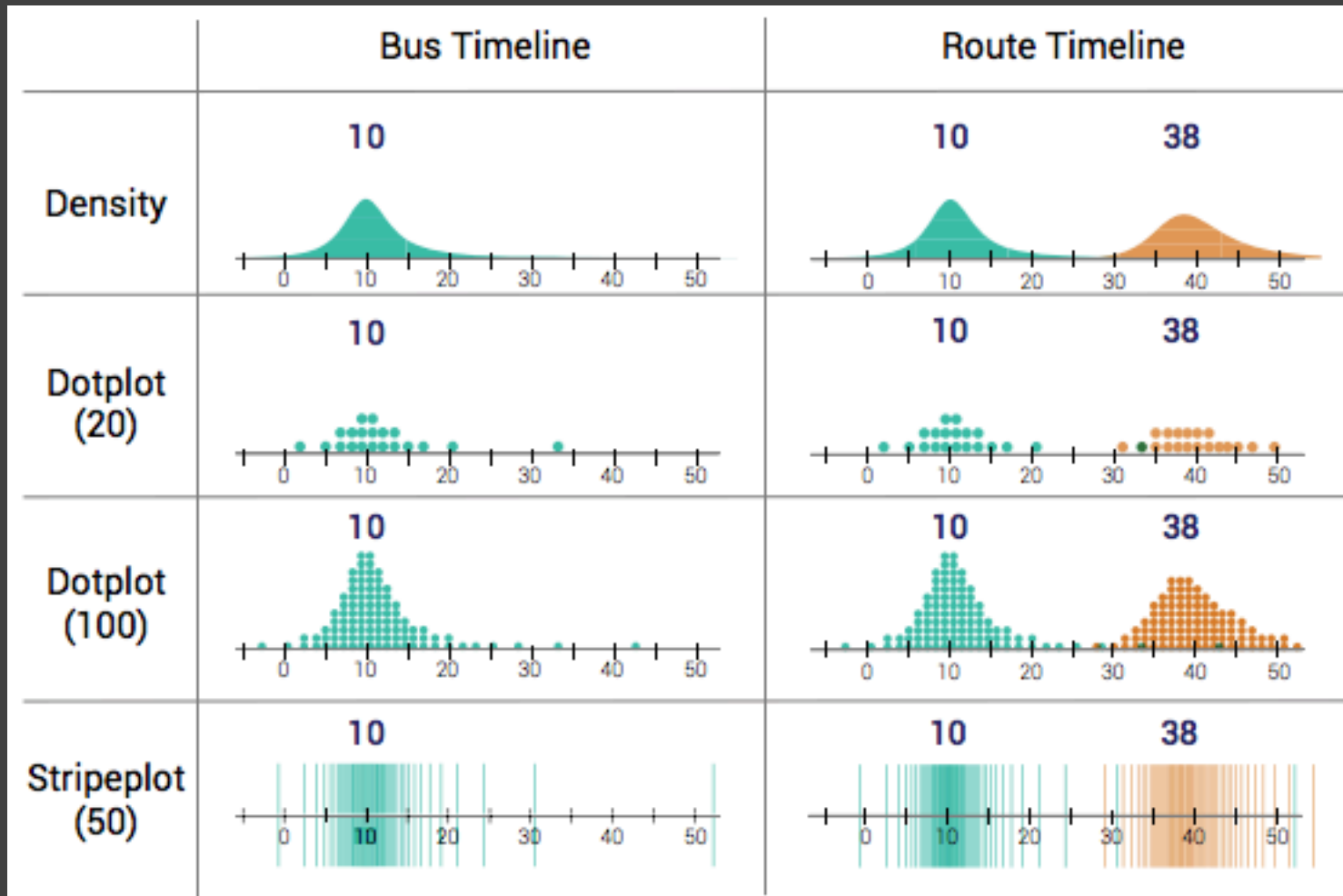
# Intervals

[Kay et al. '16]

	Density	Stripeplot	Density+Stripeplot	Dotplot(20)	Dotplot(50)	Dotplot(100)
shows discrete, countable events						
fast counting in tails		●		●	●	●
fast counting in body				●		
directly estimate density	●	●	●		●	●
directly estimate quantiles		●	●	●	●	●
tight densities drawn consistently		●		●		
project to axis		●				
easily assess range (min/max)	●	●			●	●
easily assess mode	●		●	●	●	●

# Intervals

[Kay et al. '16]



# Hypothetical Outcome Plots

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



...the jobs report **could look like this:**





# Missing Data

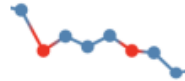
[Song et al. '18]



(a) Data Absent



(b) Color Points



(c) Color Points & Line Gradients



(d) Connected Error Bars



(e) Disconnected Error Bars



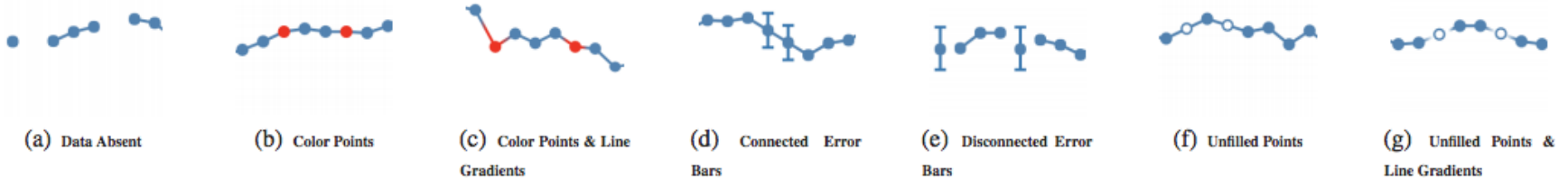
(f) Unfilled Points



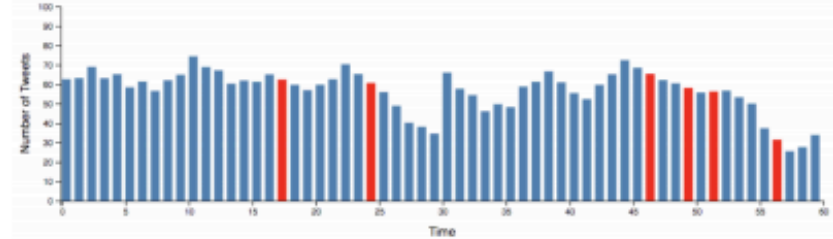
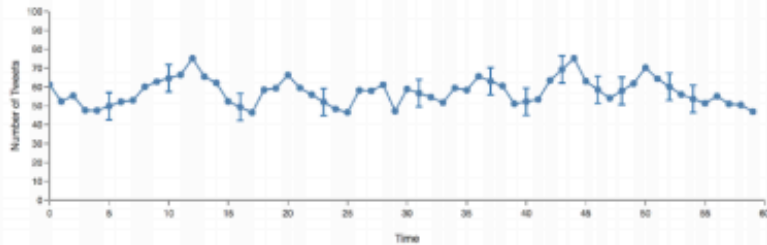
(g) Unfilled Points & Line Gradients

# Missing Data

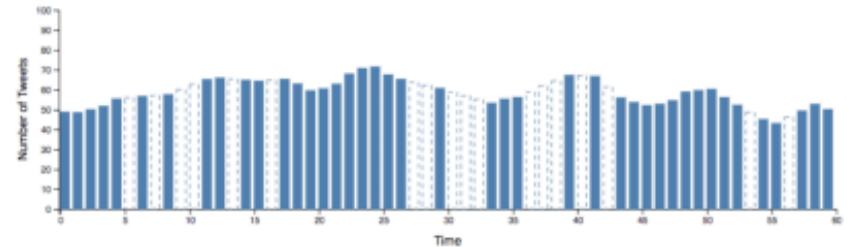
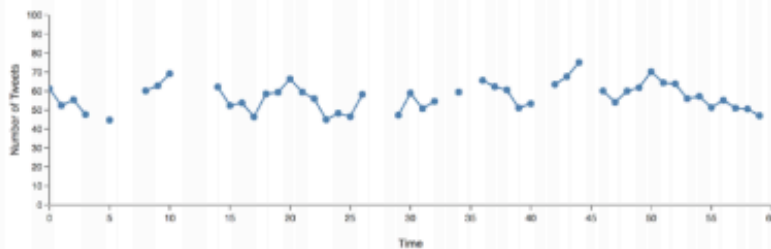
[Song et al. '18]



## Visualizations with High Data Quality



## Visualizations with Low Data Quality

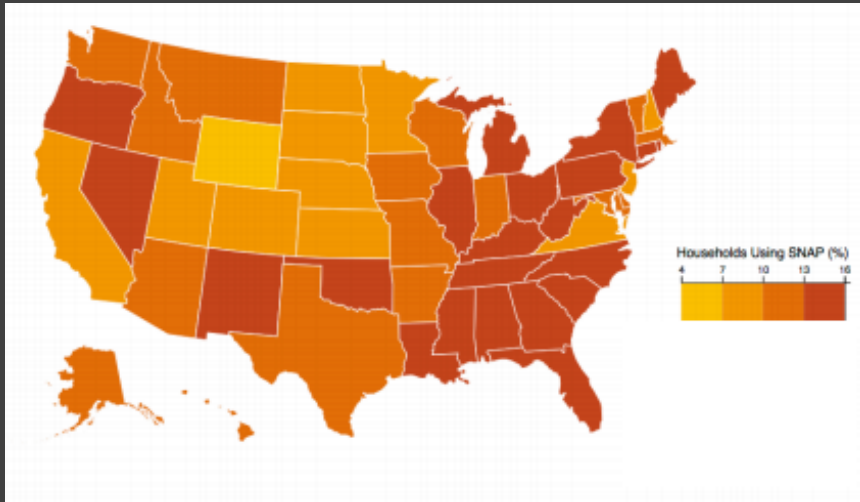


# Encoding Uncertainty

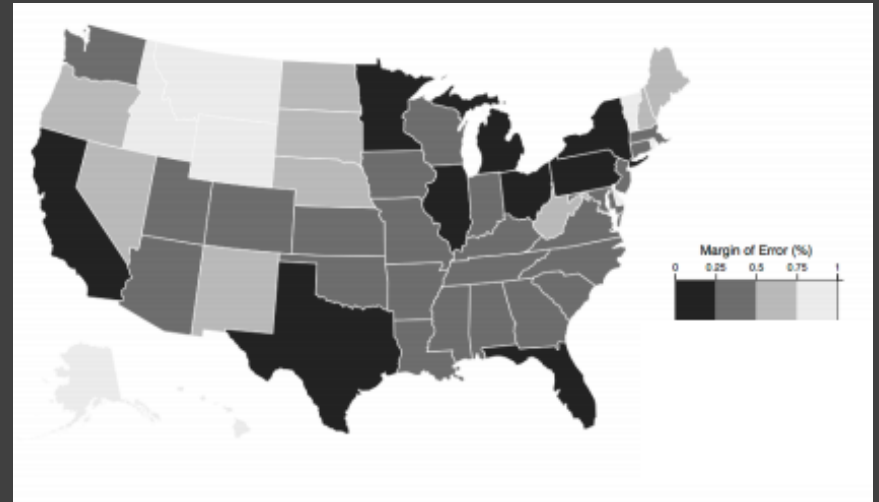
Given some geographic data with some level of uncertainty, how should we visualize it?

# Encoding Uncertainty: Juxtaposition

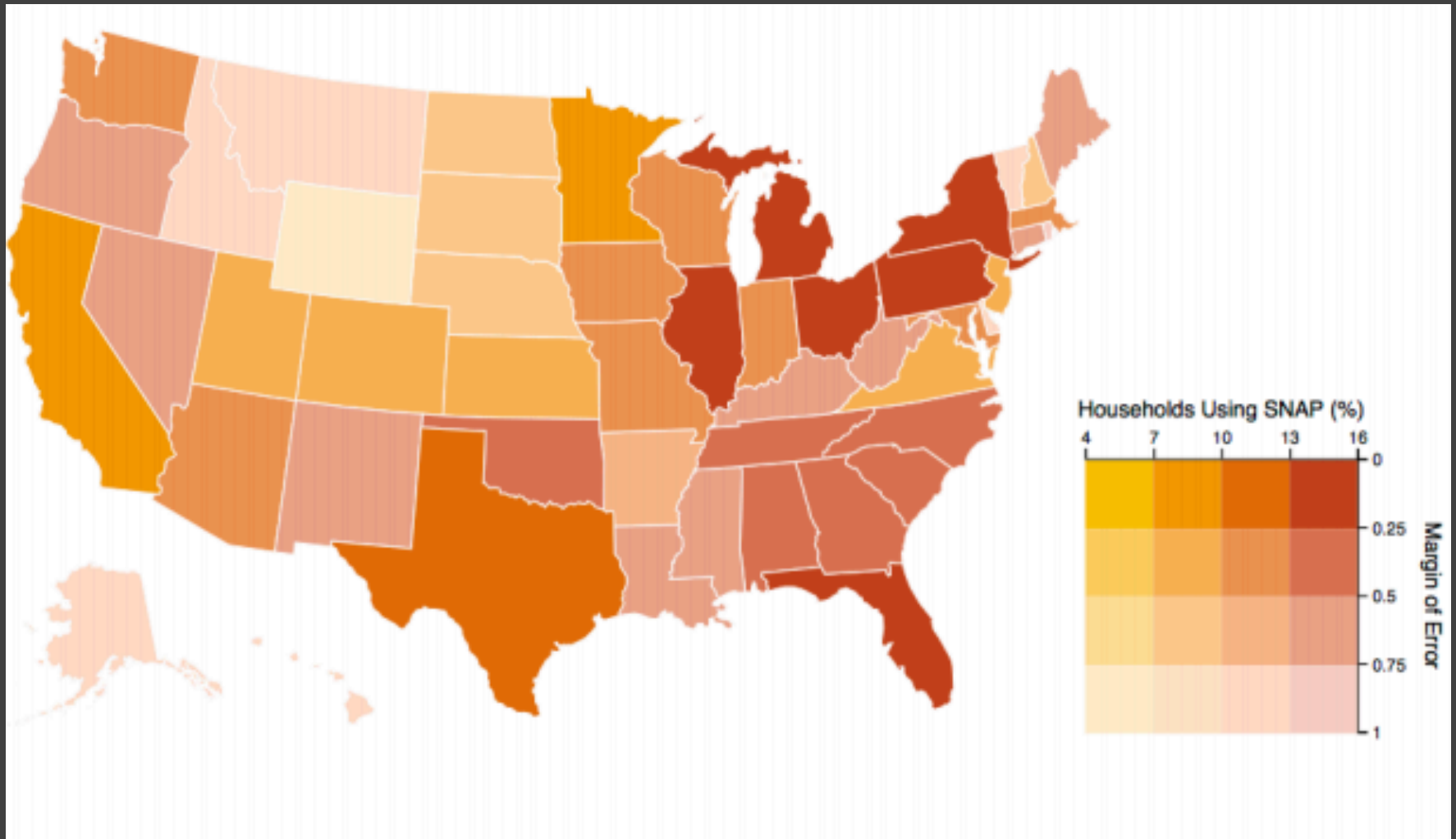
Data Map



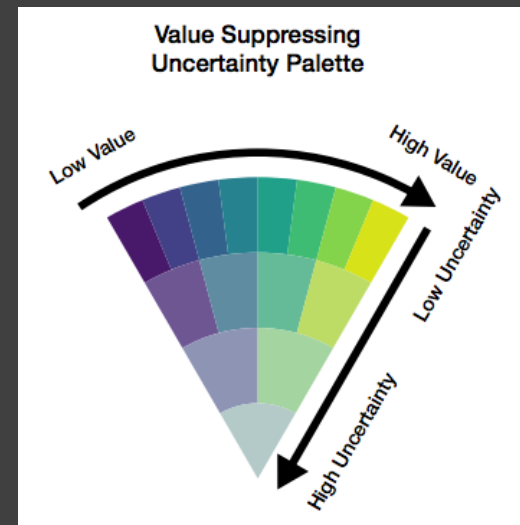
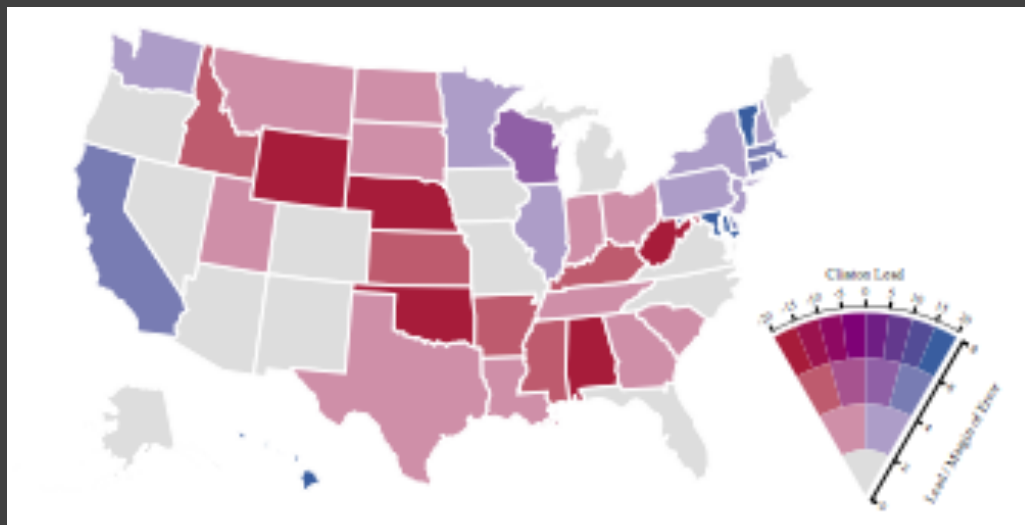
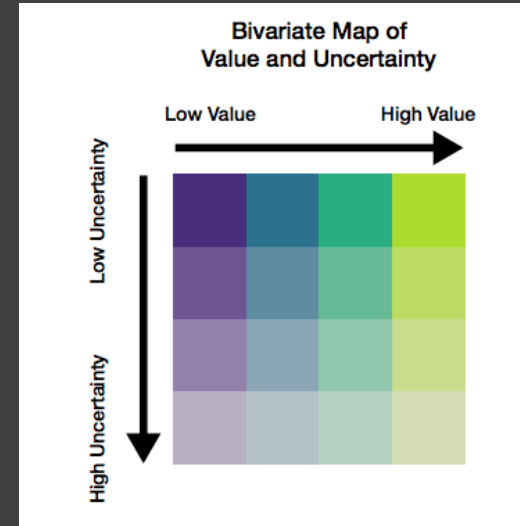
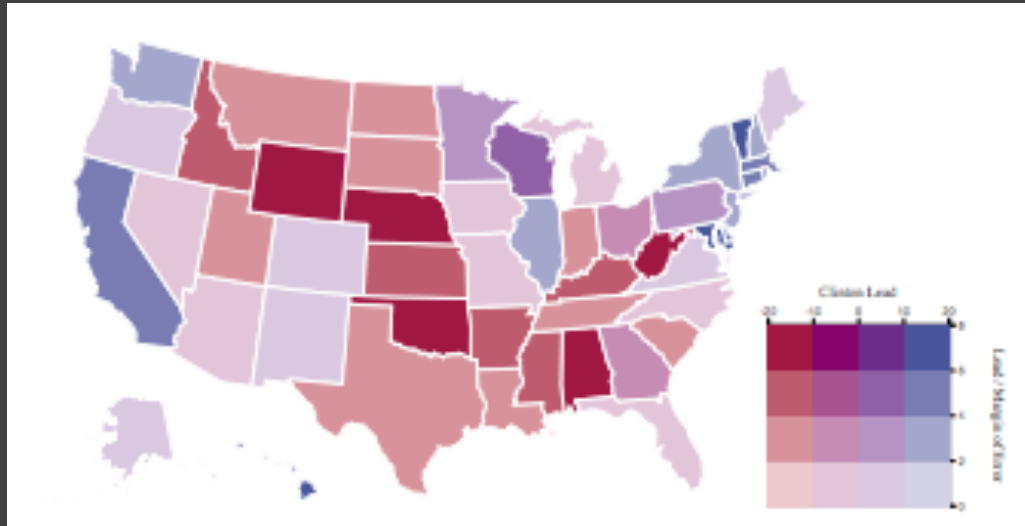
Uncertainty Map



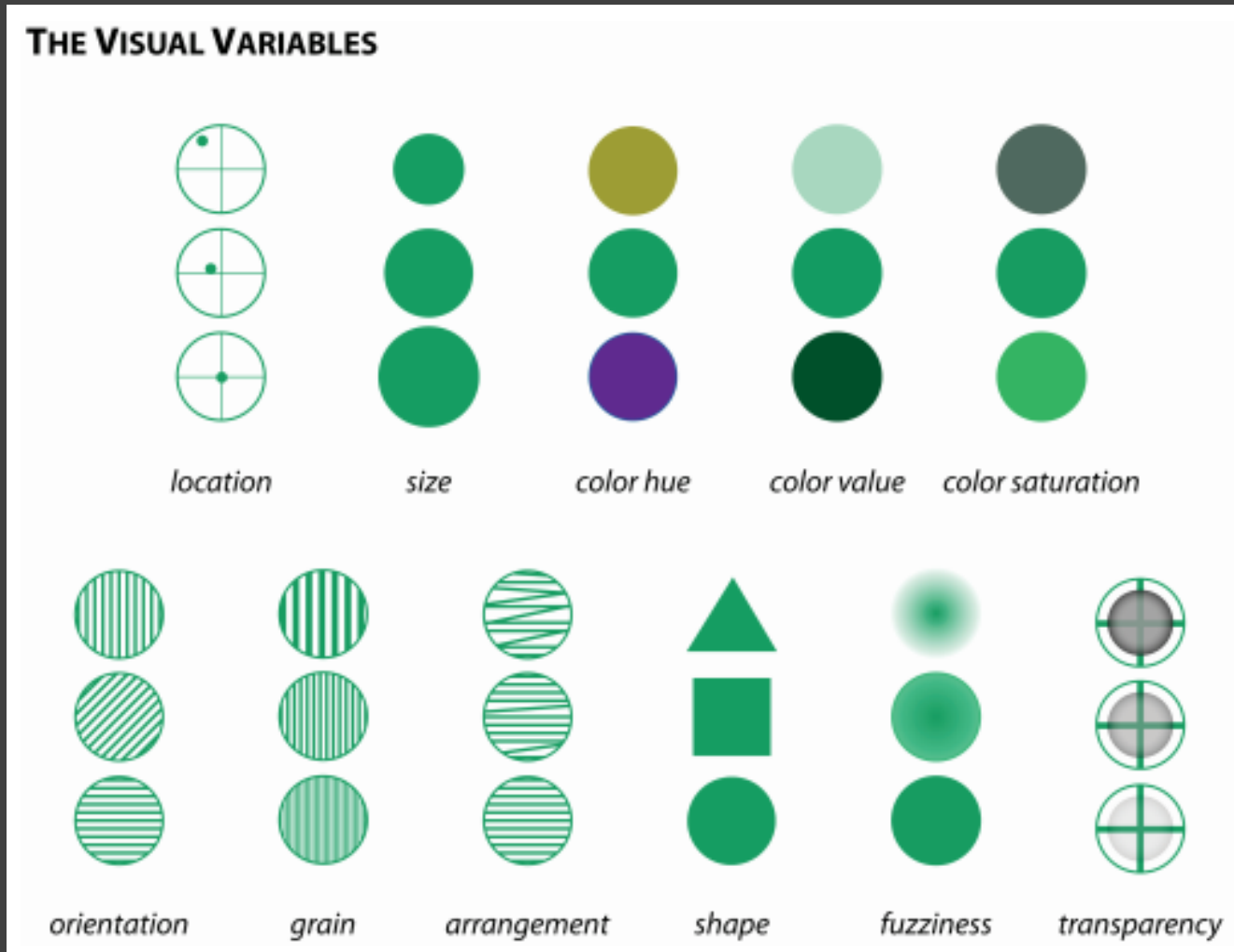
# Encoding Uncertainty: Superposition



# Value-Suppressing Uncertainty Palettes

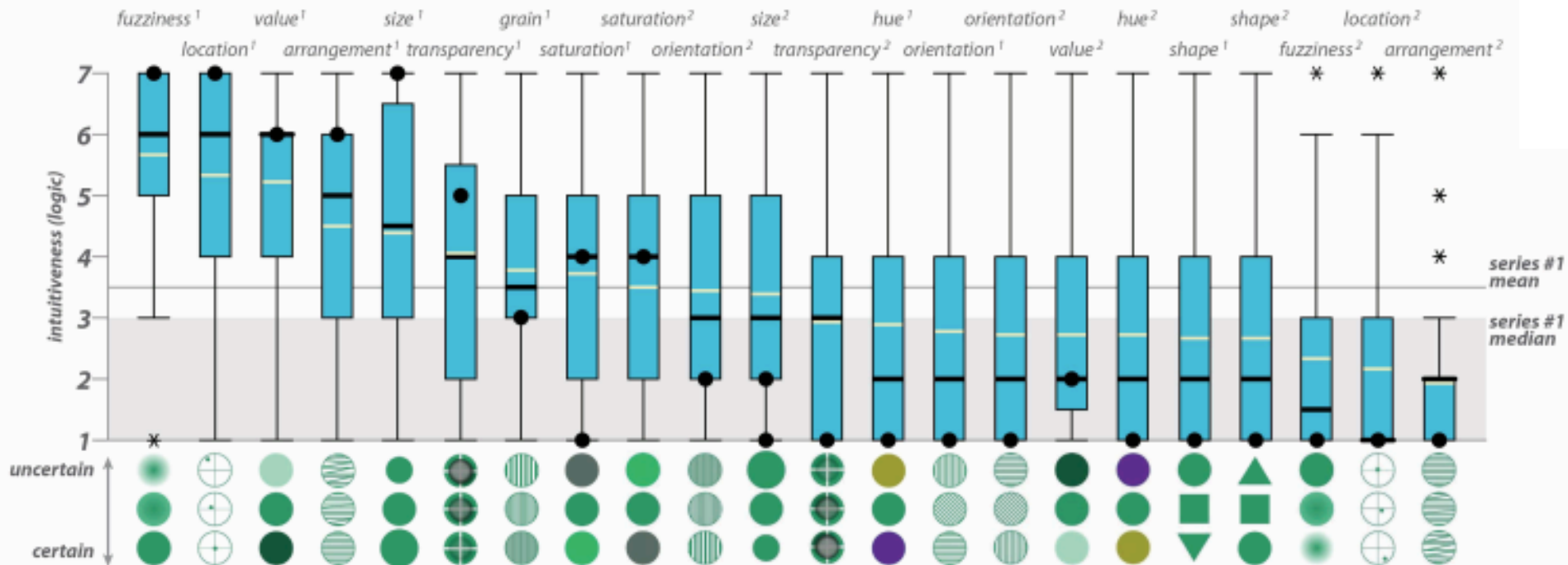


# Semiotic of Uncertainty



# Semiotic of Uncertainty

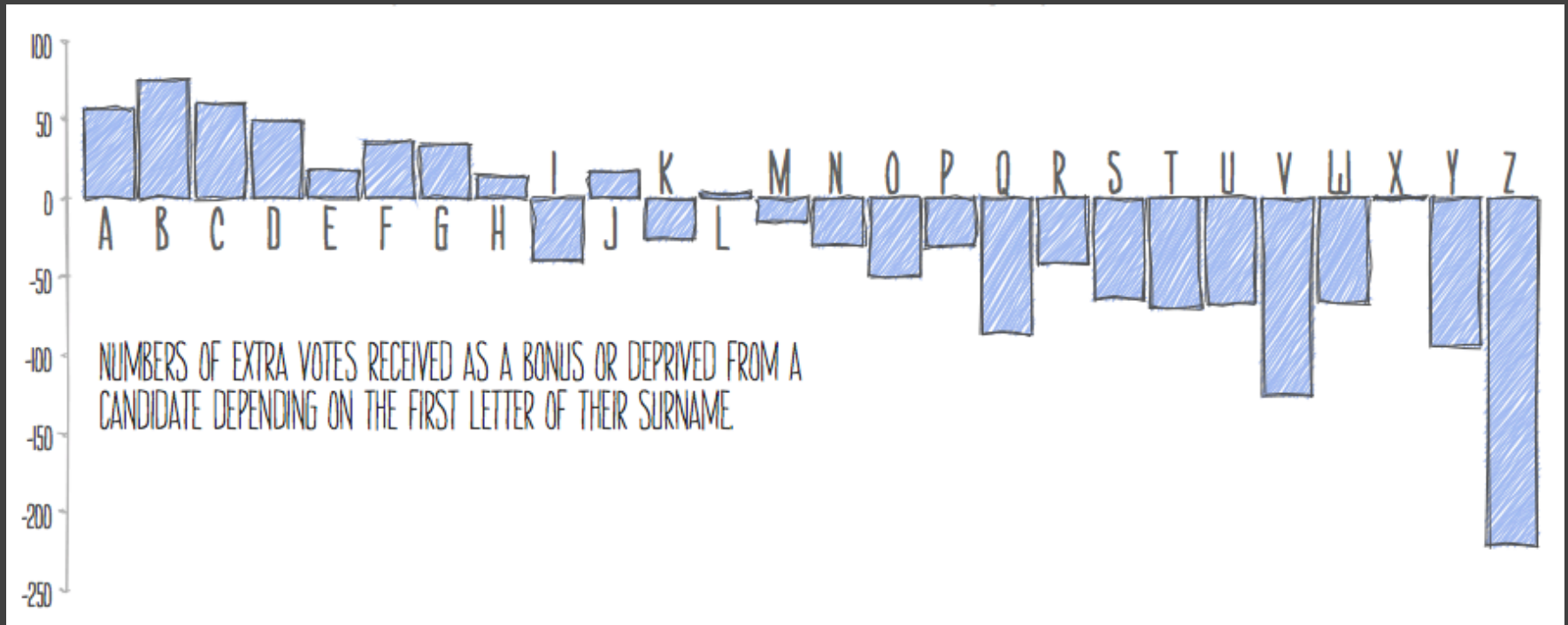
## SERIES #1: GENERAL UNCERTAINTY BY VISUAL VARIABLE





# "Sketchiness"

[Wood et al. '12]  
[Boukhelifa et al. '12]



# 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*

# Polling Data

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

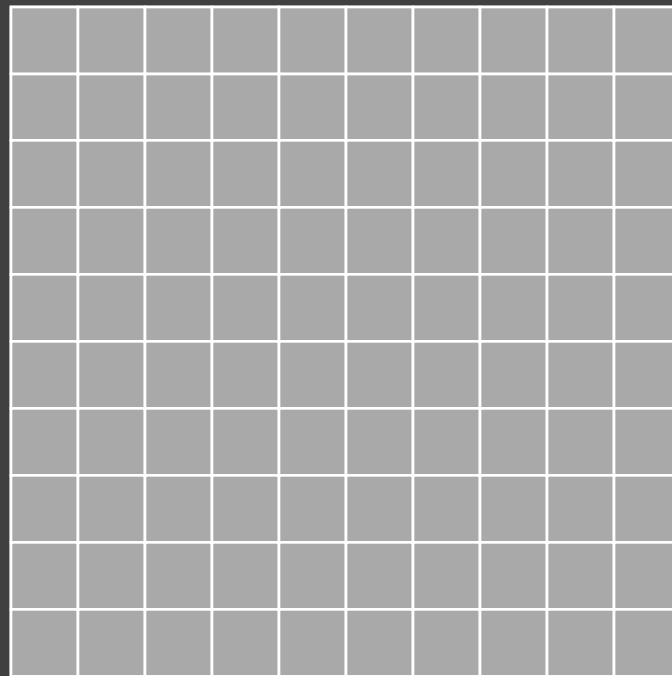


*\*poll of 100 people, margin of error +/-5*

# Polling Data

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

Poll results

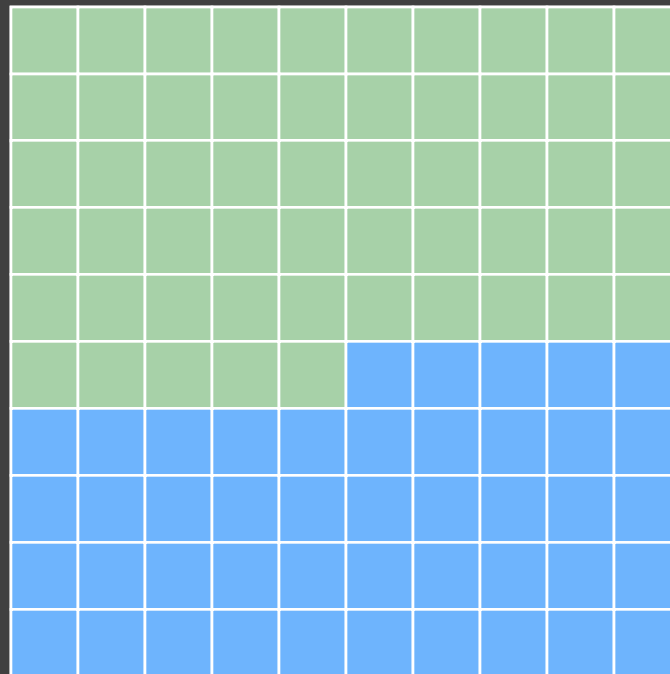


*\*poll of 100 people, margin of error +/-5*

# Polling Data

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

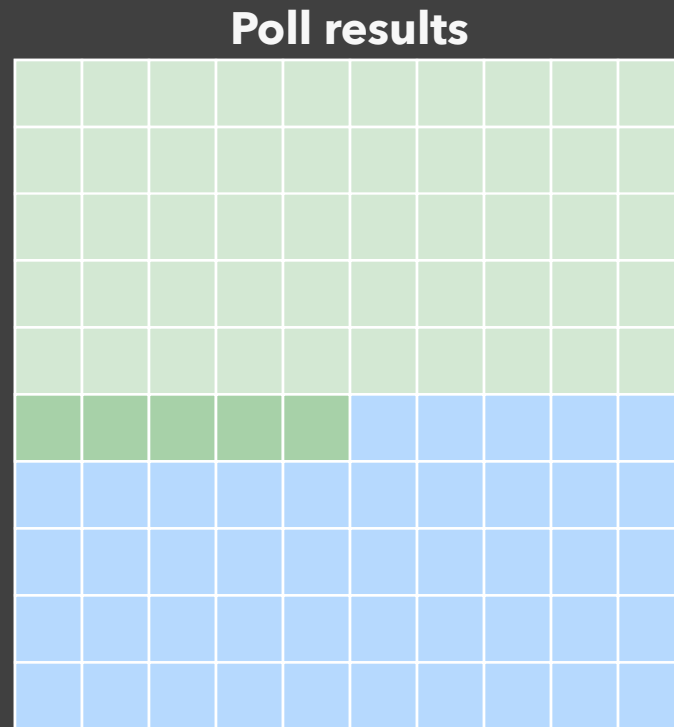
Poll results



\*poll of 100 people, margin of error +/-5

# Polling Data

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

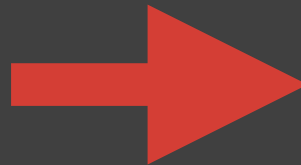
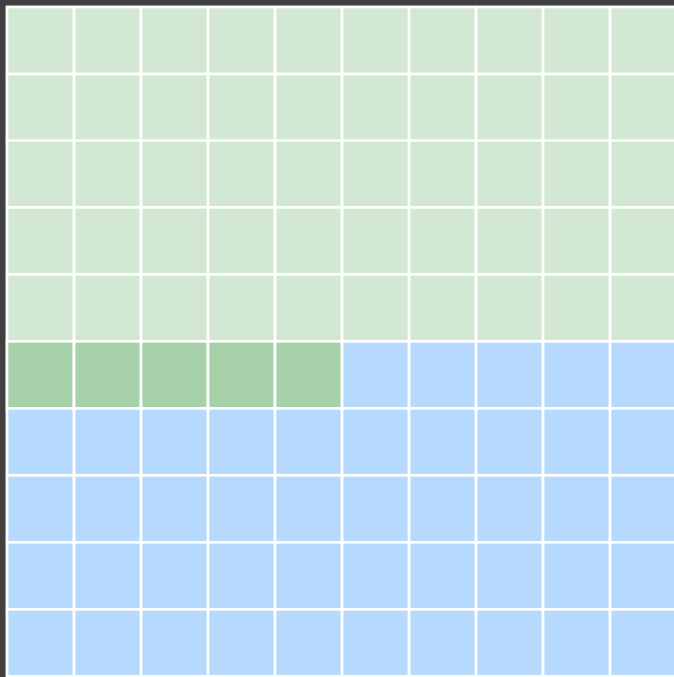


\*poll of 100 people, margin of error +/-5

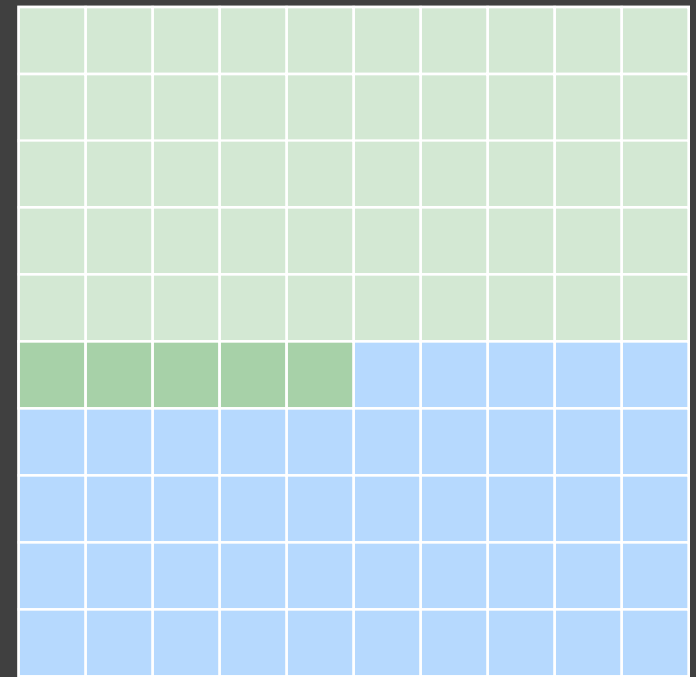
# Polling Data

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

Poll results



Election results



*\*poll of 100 people, margin of error +/-5*

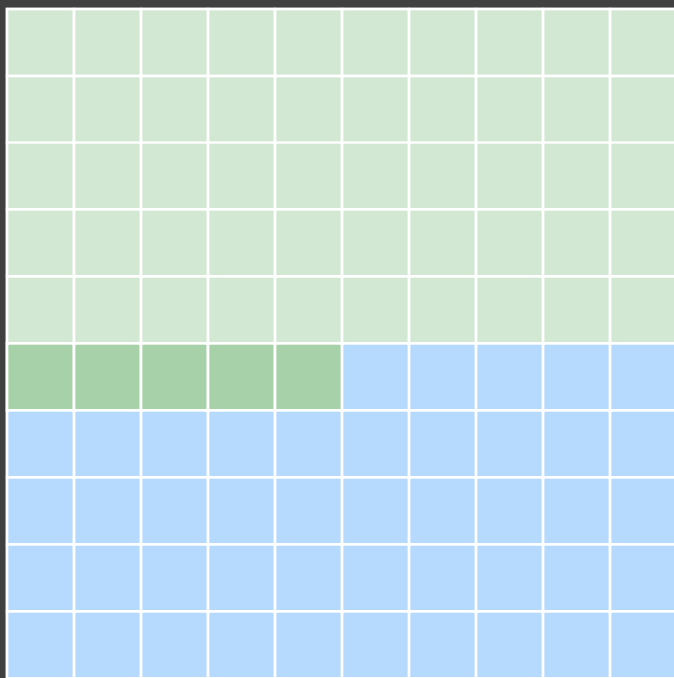




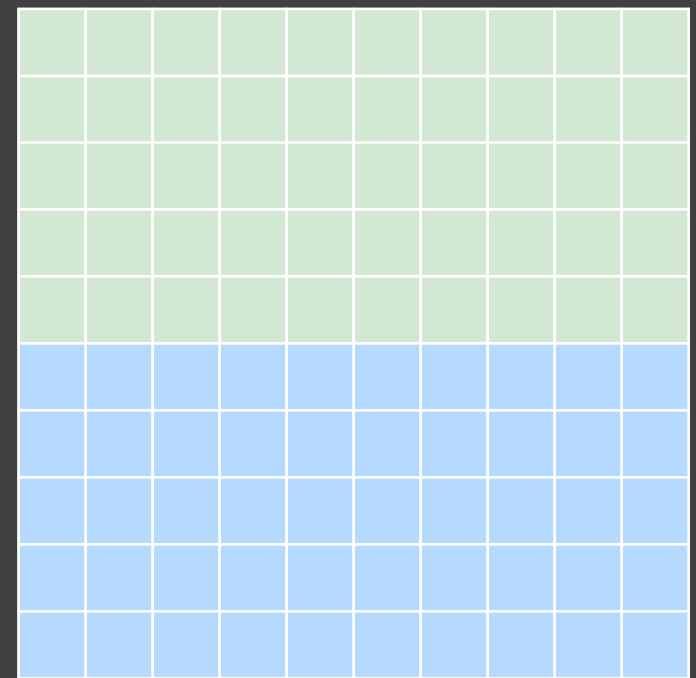
# Polling Data

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

Poll results



Election results

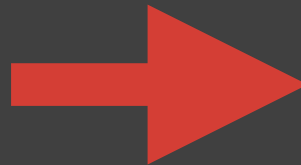
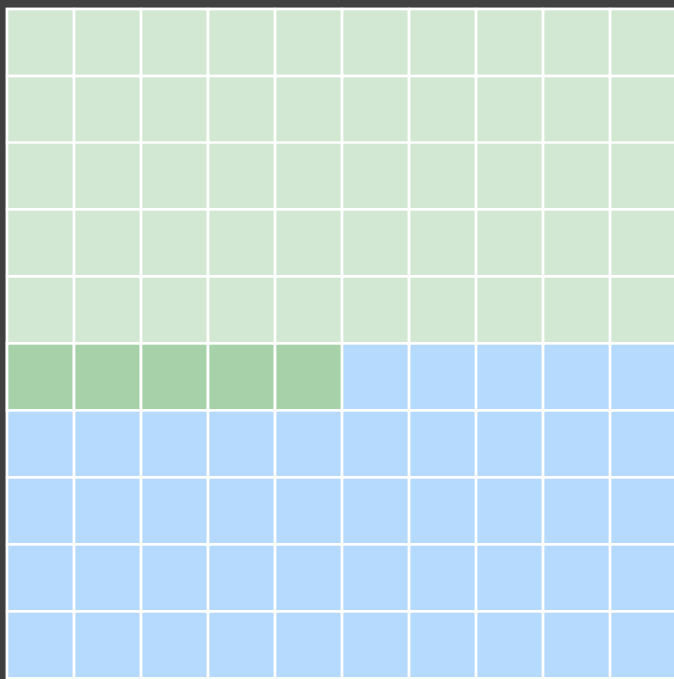


\*poll of 100 people, margin of error +/-5

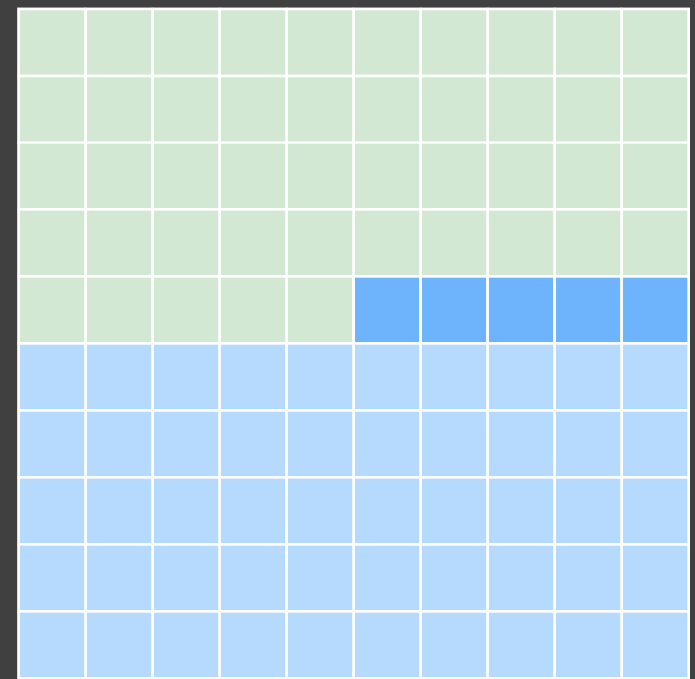
# Polling Data

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

Poll results



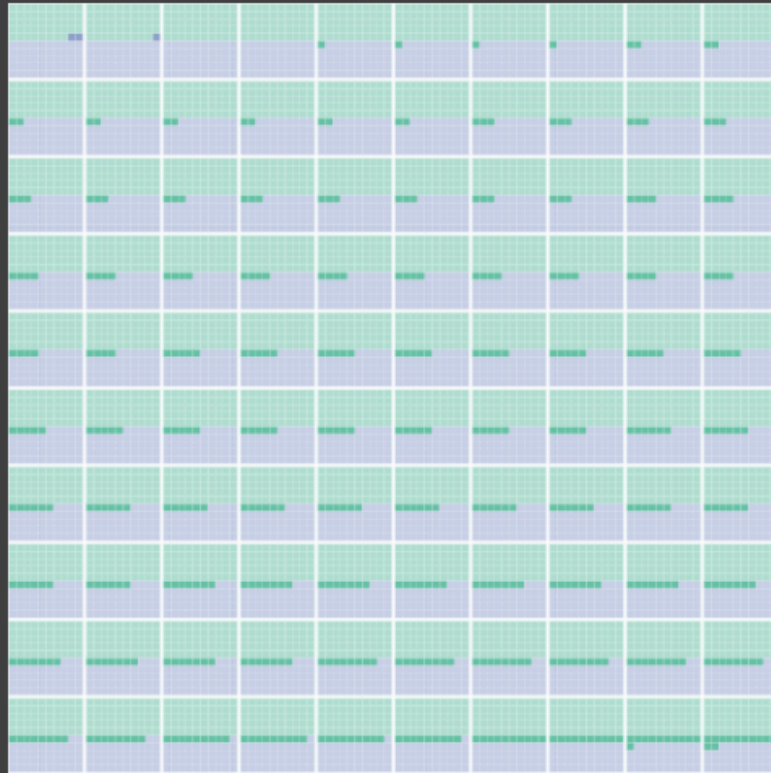
Election results



\*poll of 100 people, margin of error +/-5

# Pangloss Plot

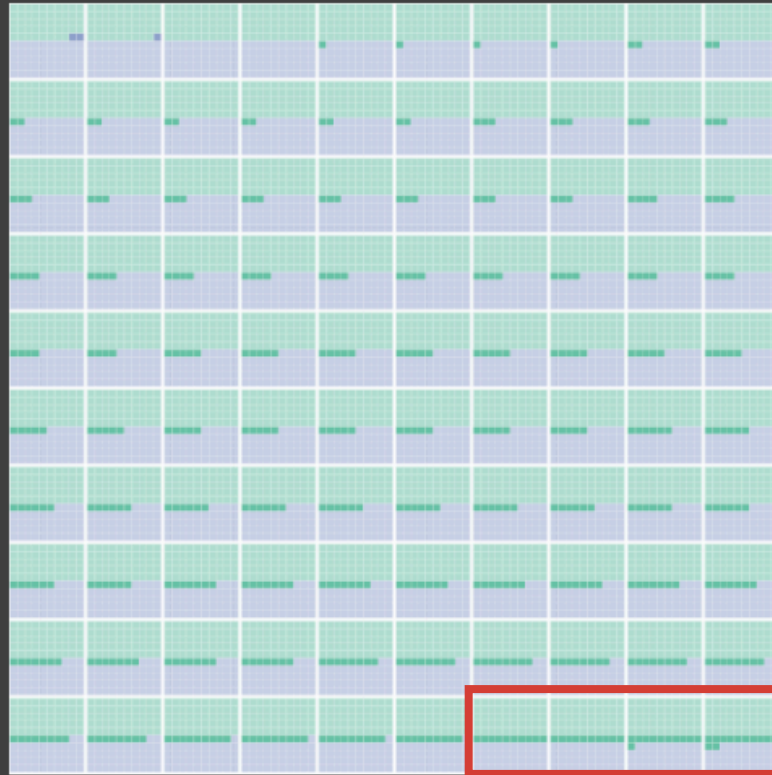
**Candidate A** is ahead of **Candidate B** in the polls,  
with 55% of the likely voters\*



*\*poll of 100 people, margin of error +/-5*

# Pangloss Plot

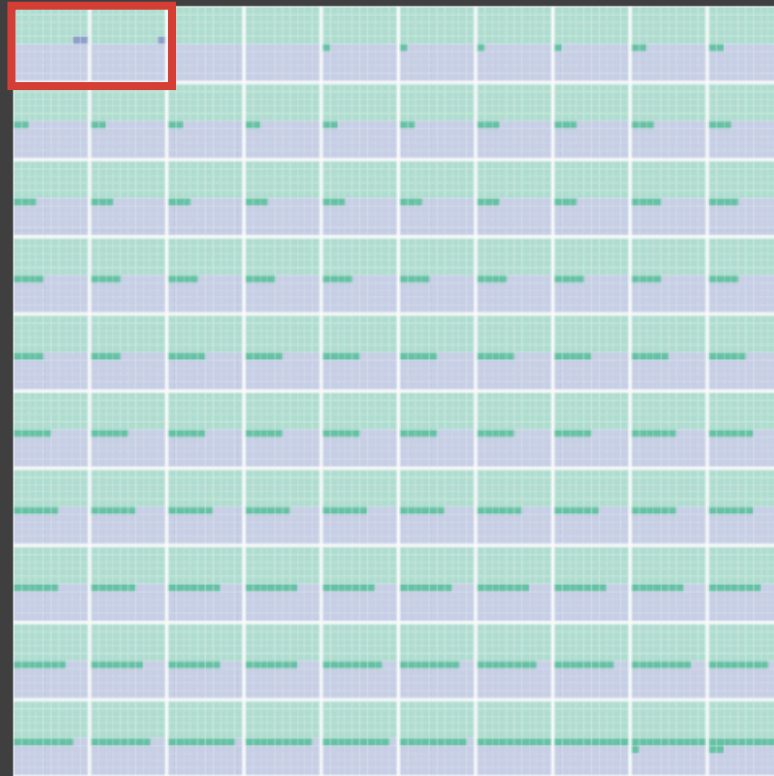
**Candidate A** is ahead of **Candidate B** in the polls,  
with 55% of the likely voters\*



*\*poll of 100 people, margin of error +/-5*

# 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*

# Bubble Swarm

## 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**

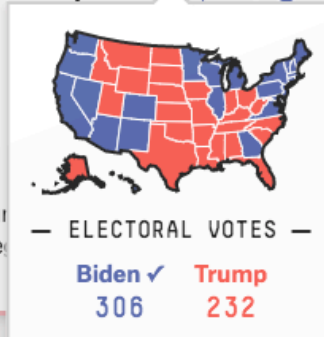
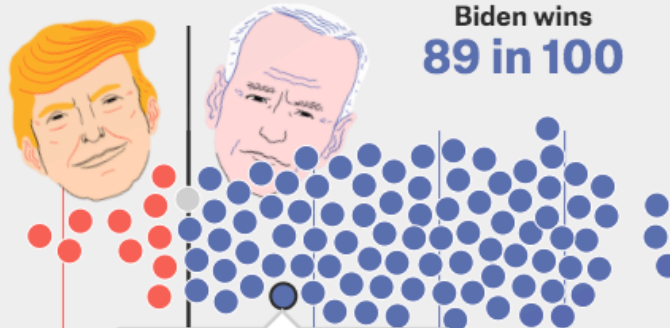
+300  
ELECTORAL VOTE  
MARGIN

+200

+100

0

+300



● Trump  
● No Electoral College

— ELECTORAL VOTES —

Biden ✓ 306  
Trump 232



Don't count the underdog 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 (what results are possible)

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?

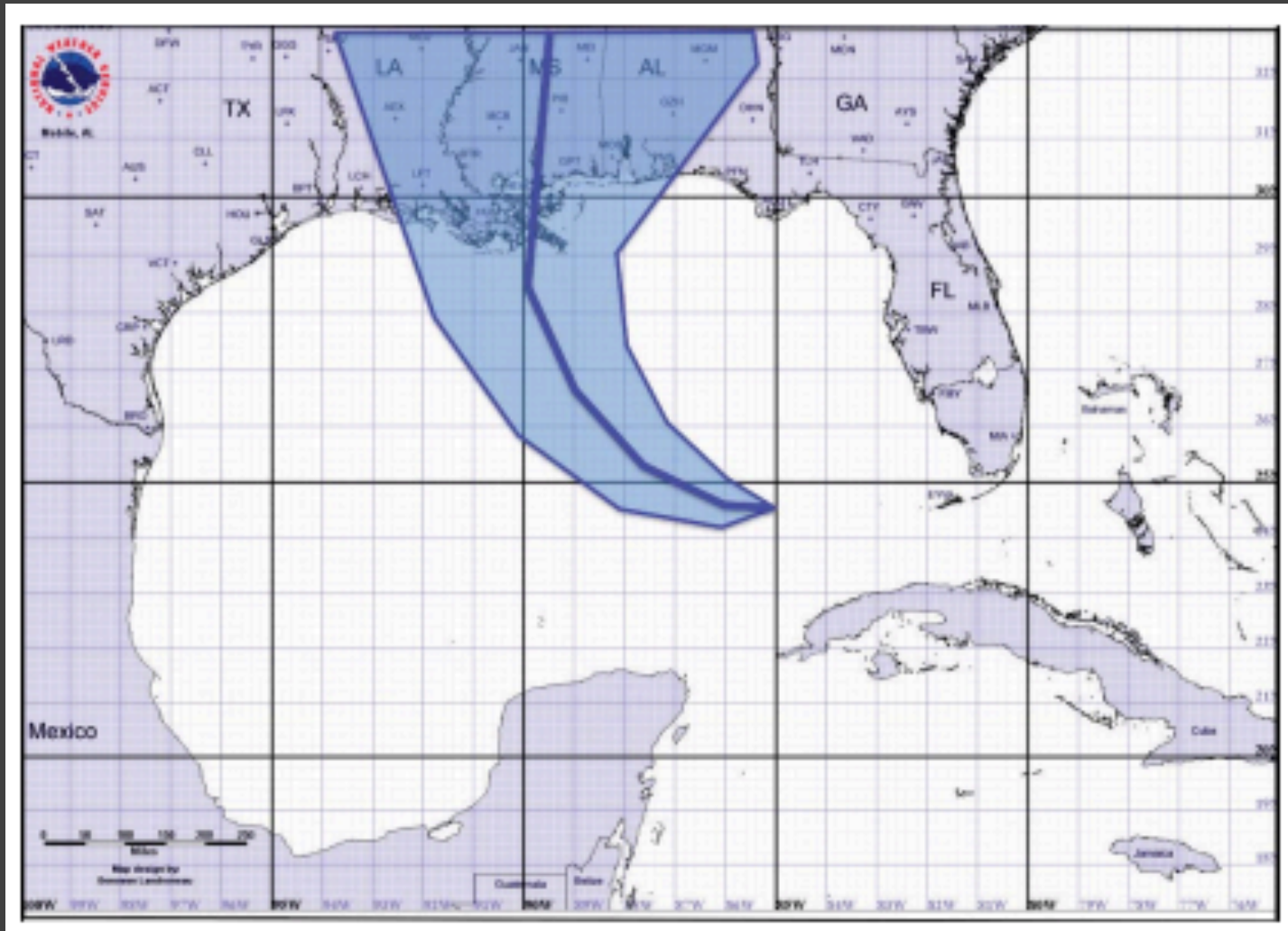
Choose an appropriate visual variable based on the domain, literacy, and expertise of your audience. Be mindful that any display of uncertainty inherently increases the complexity of your visualization, and that there is a preference/performance gap.

**IT DEPENDS**

**What can go wrong?**

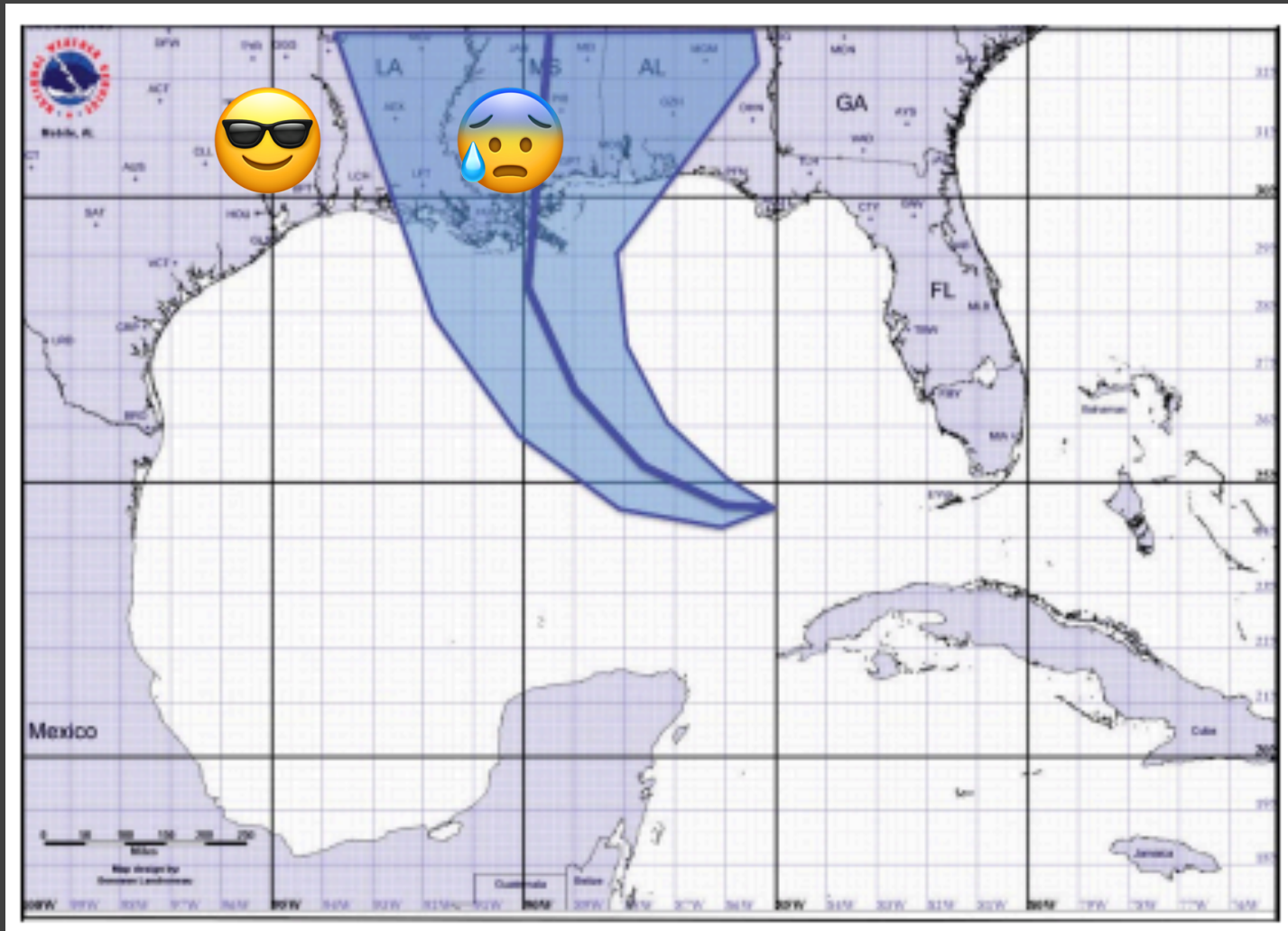
# The Error Cone

[Cox et al. '13]



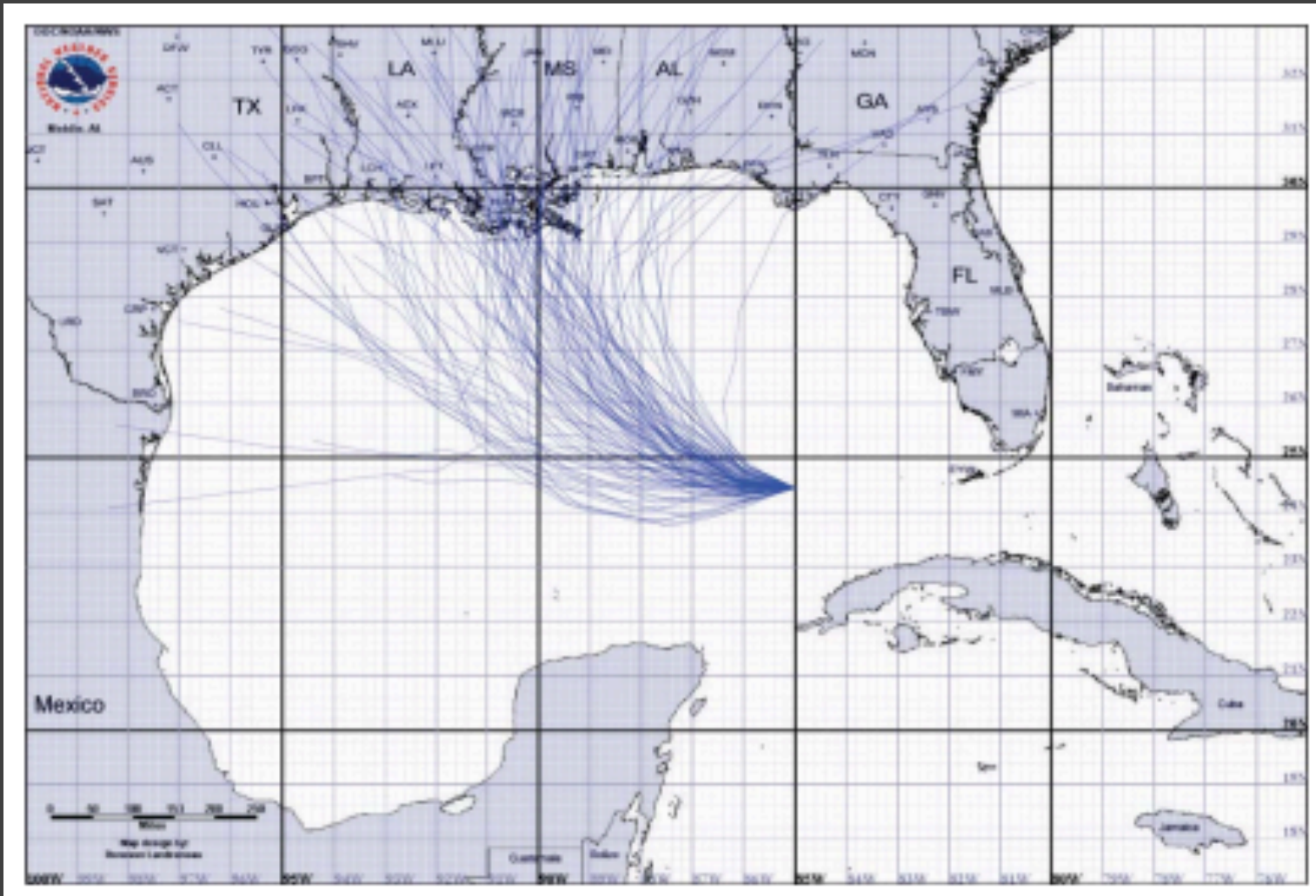
# The Error Cone

[Cox et al. '13]



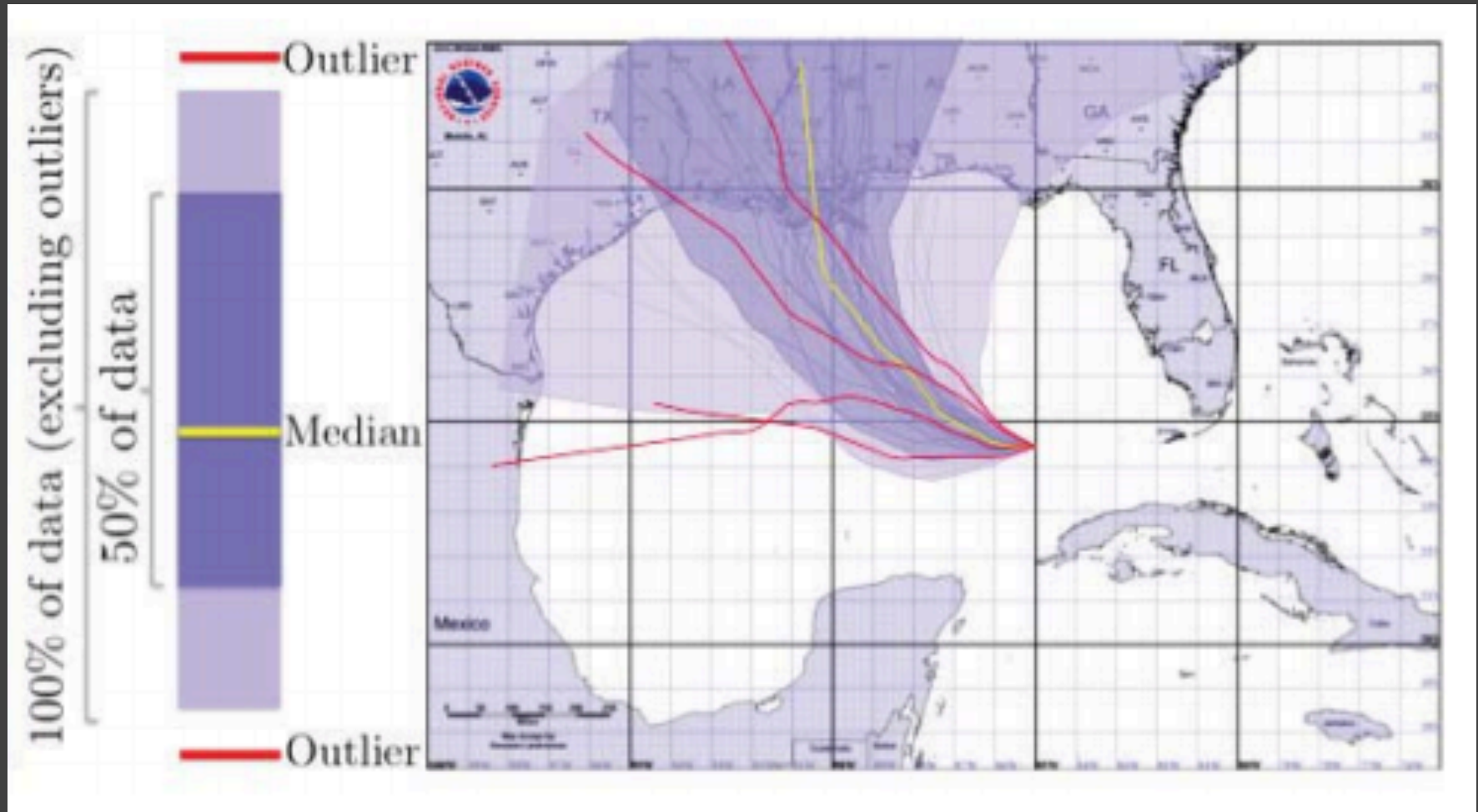
# Ensemble Plots

[Cox et al. '13]



# Curve Boxplot

[Mirzargar et al. '14]



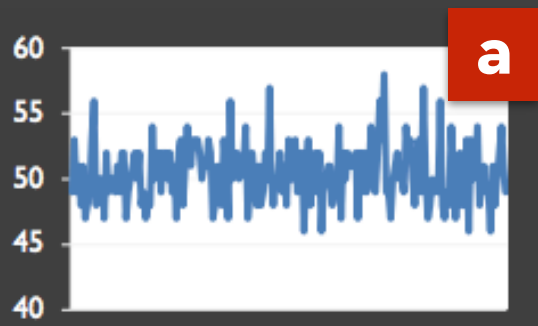
# Things that can go wrong

1. People confuse uncertainty with certainty

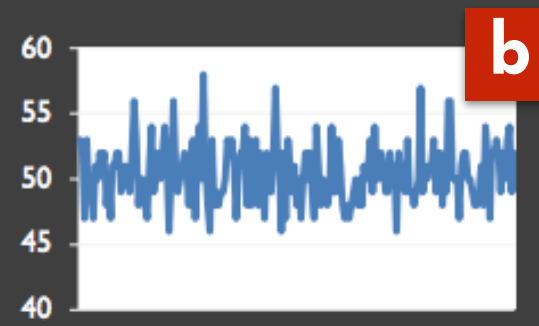


# Obama's Approval Rating

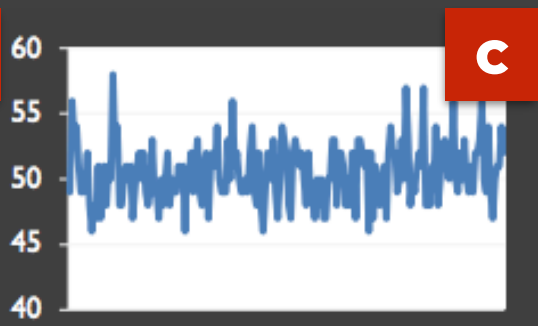




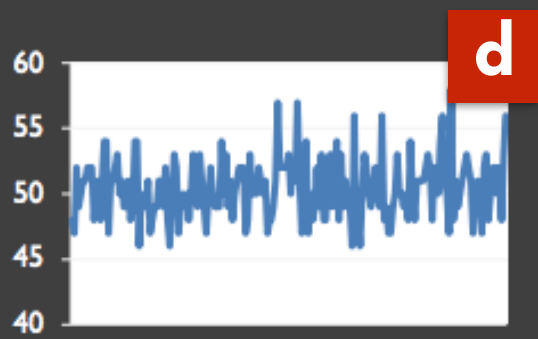
**a**



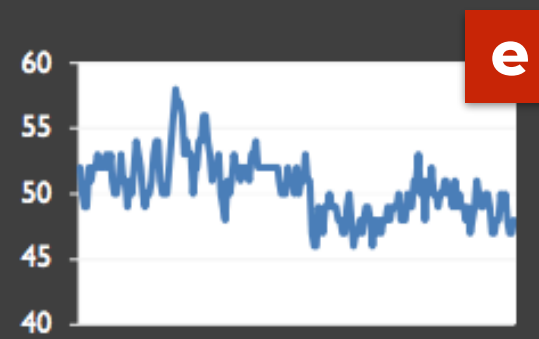
**b**



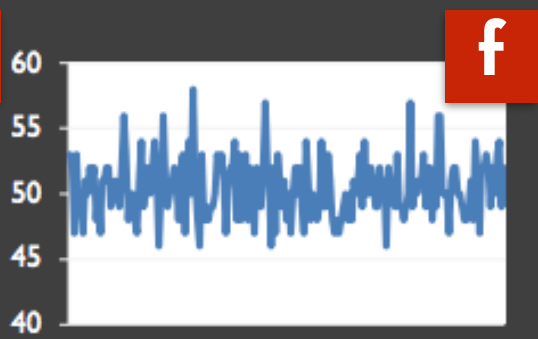
**c**



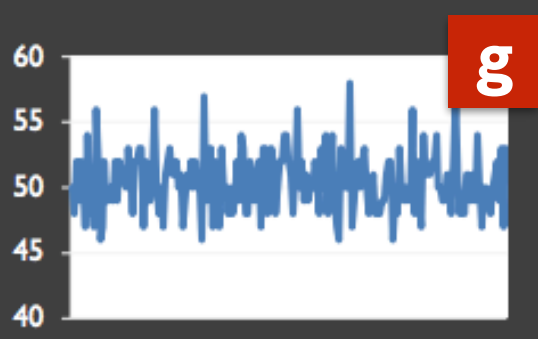
**d**



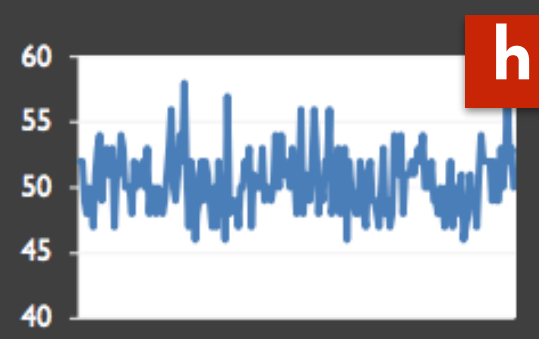
**e**



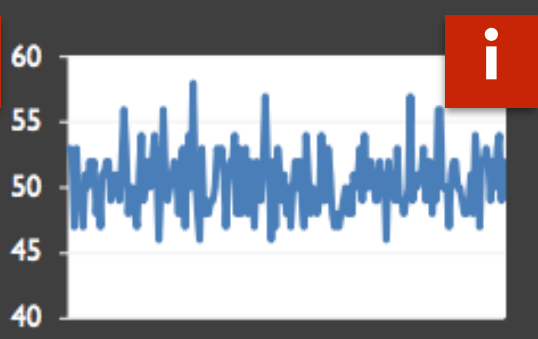
**f**



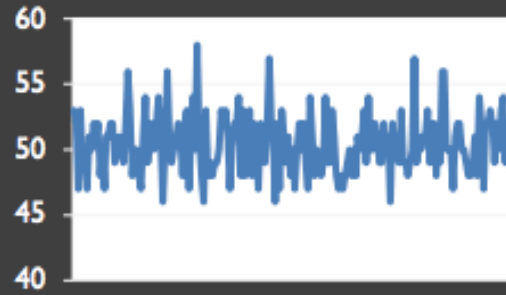
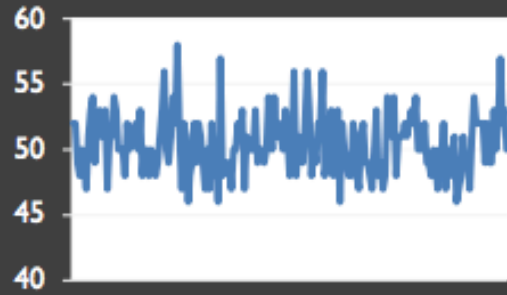
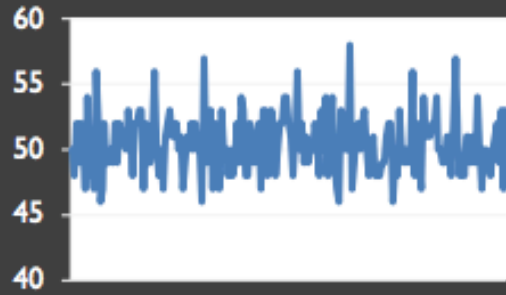
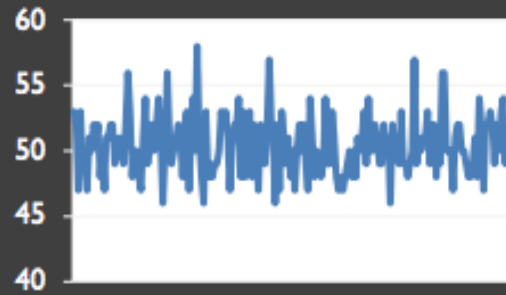
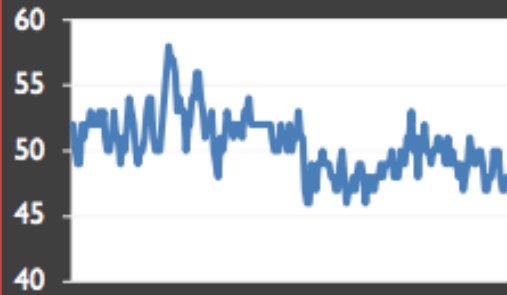
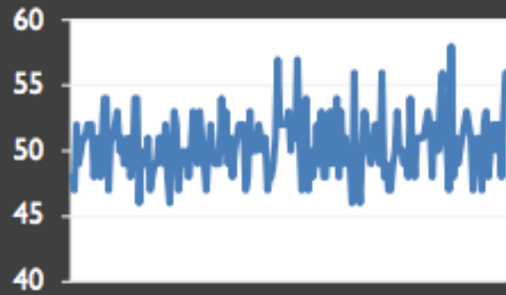
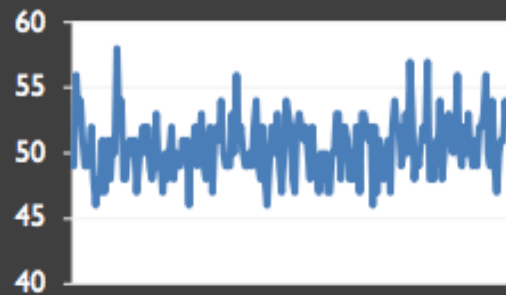
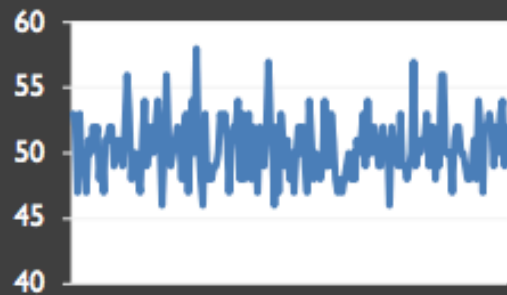
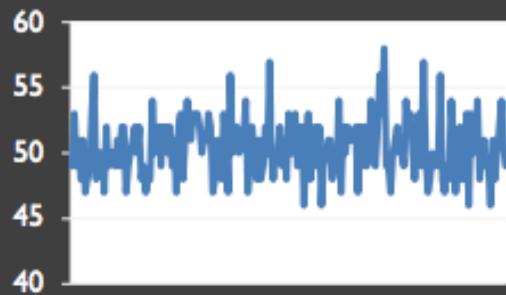
**g**



**h**

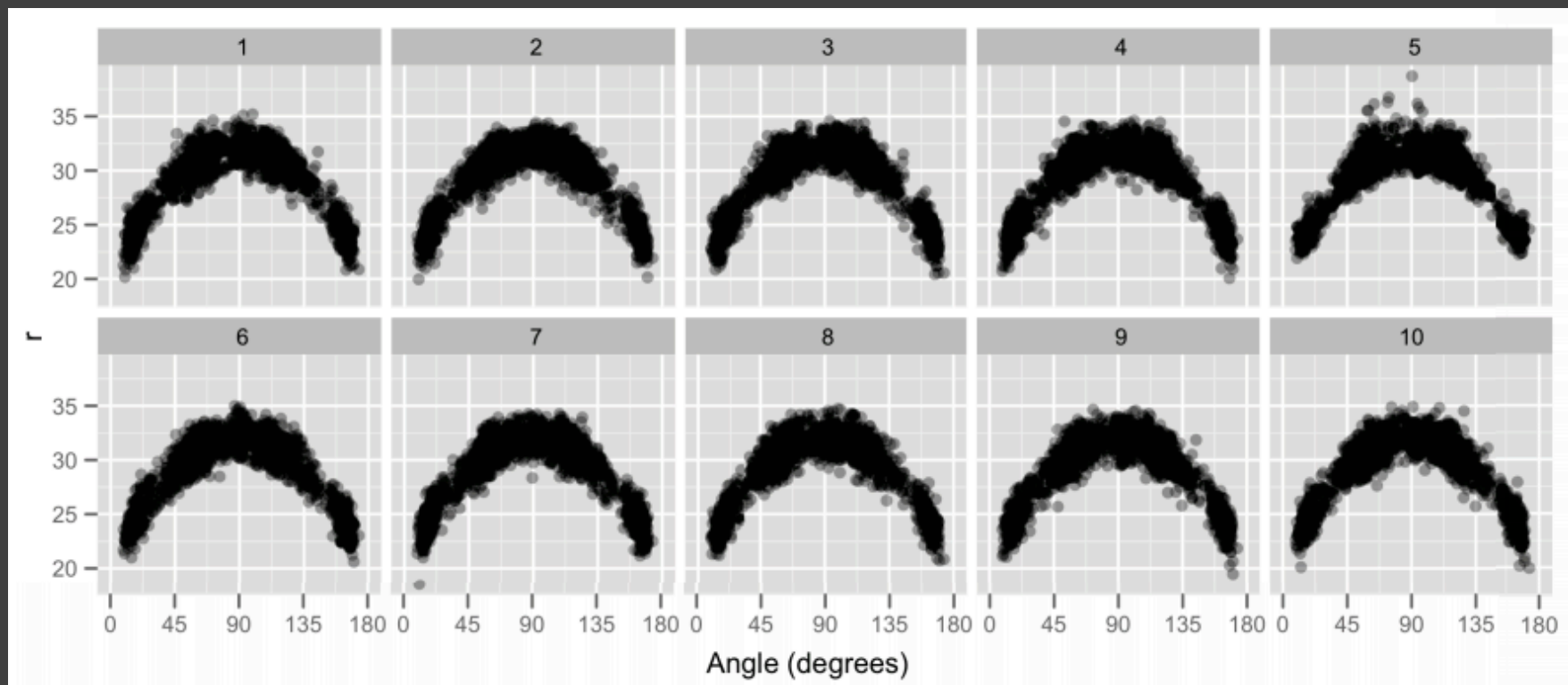


**i**



# Graphical Inference

[Wickham et al. '10]

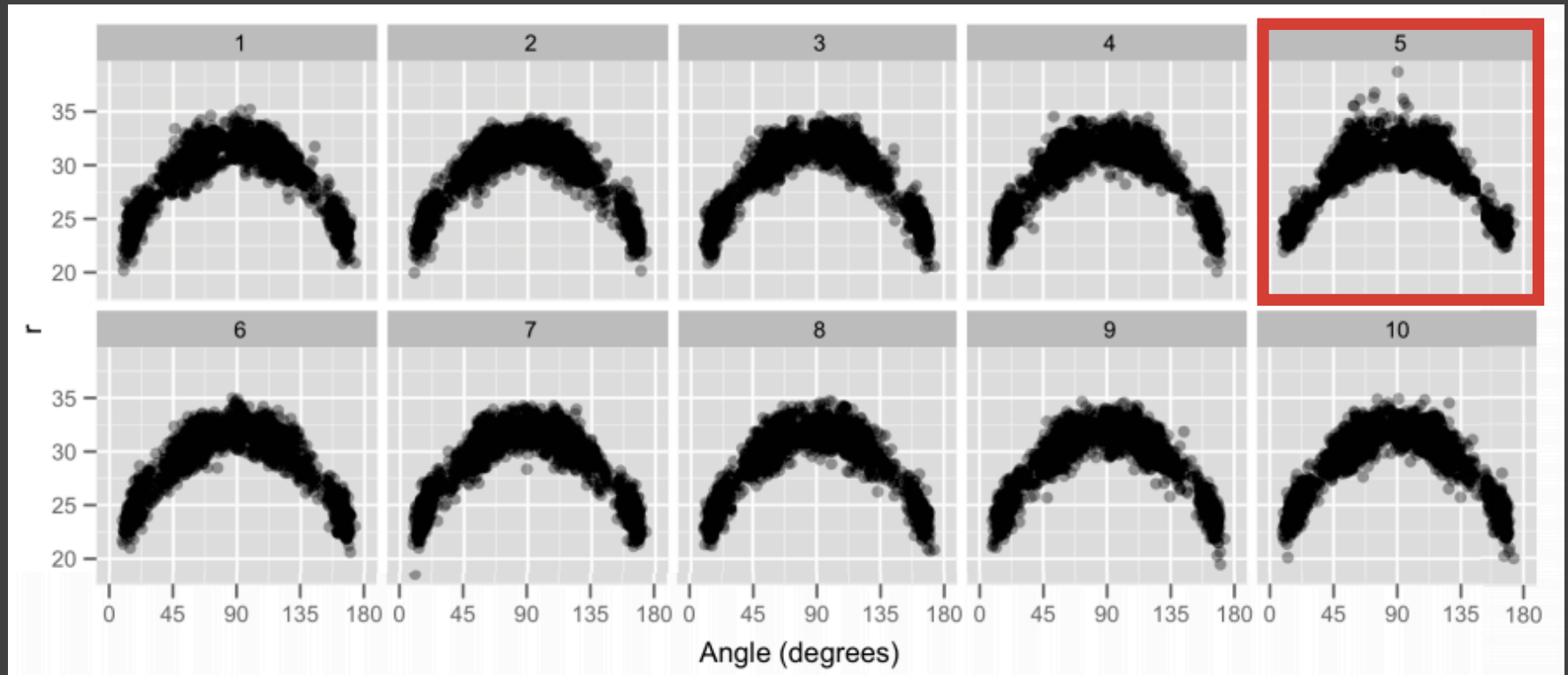


Distance vs. angle for 3 point shots by the LA Lakers.

One plot is the real data. The others are generated from a null hypothesis of quadratic relationship.

# Graphical Inference

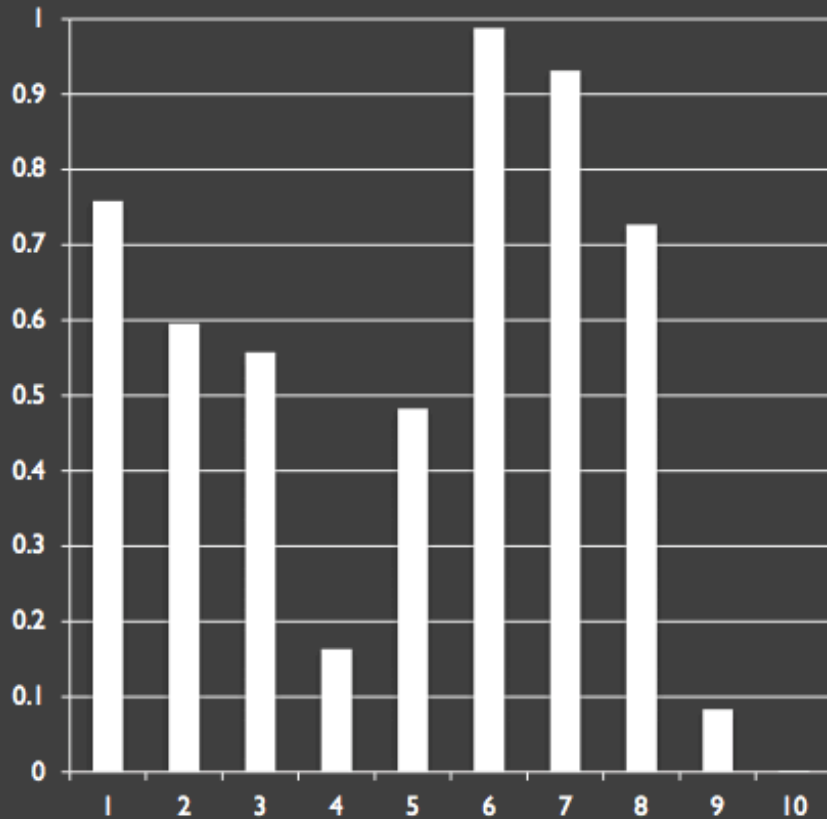
[Wickham et al. '10]



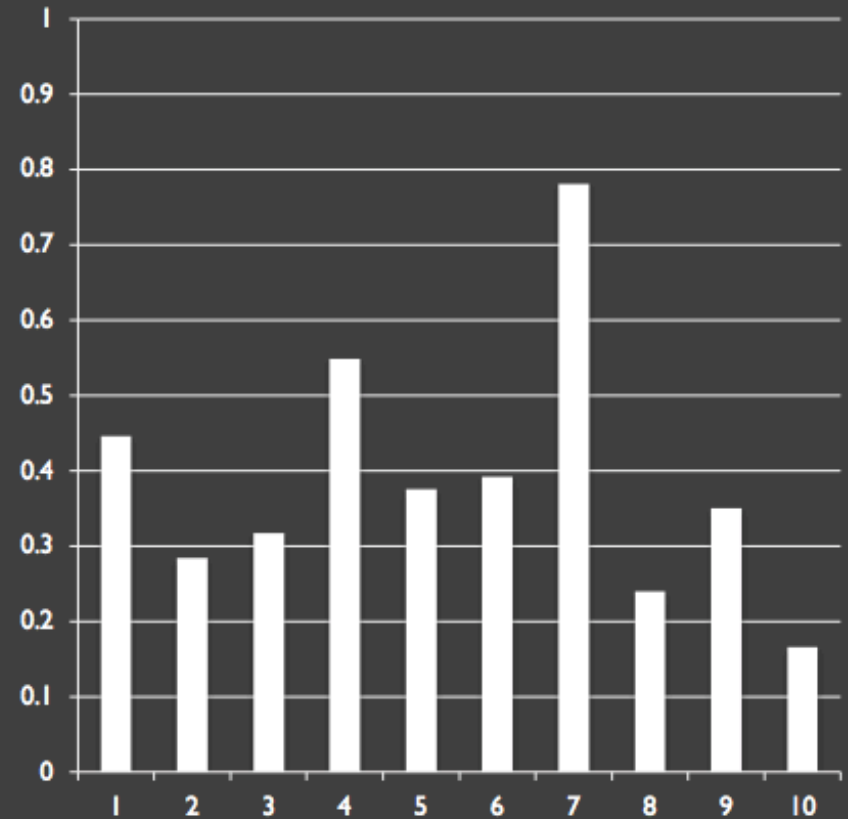
Distance vs. angle for 3 point shots by the LA Lakers. One plot is the real data. The others are generated from a null hypothesis of quadratic relationship.

# Which stock should you buy?

## Company A



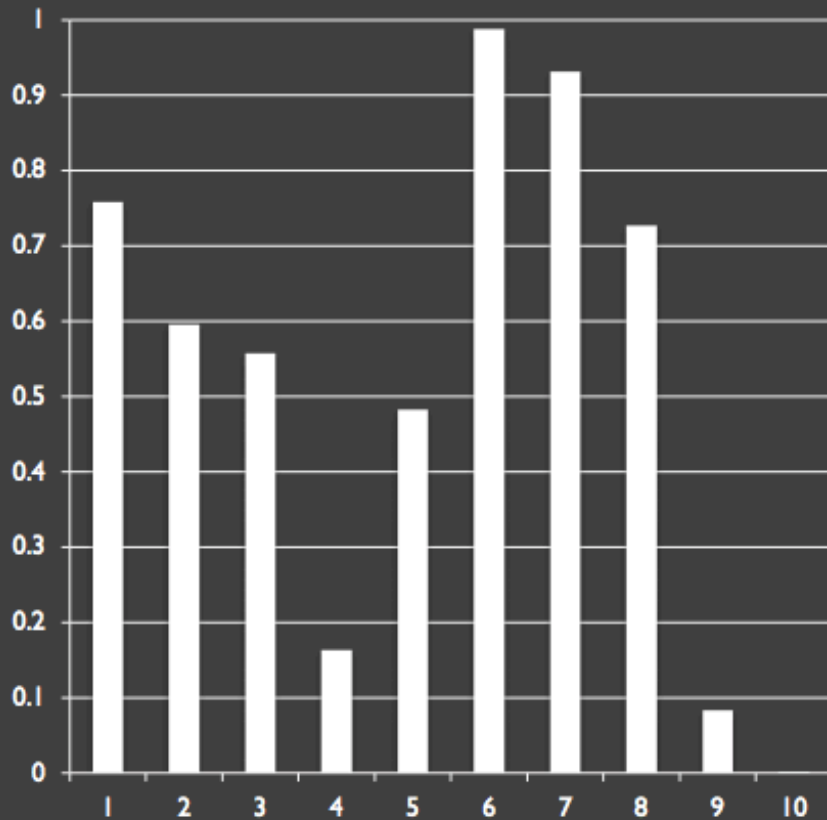
## Company B



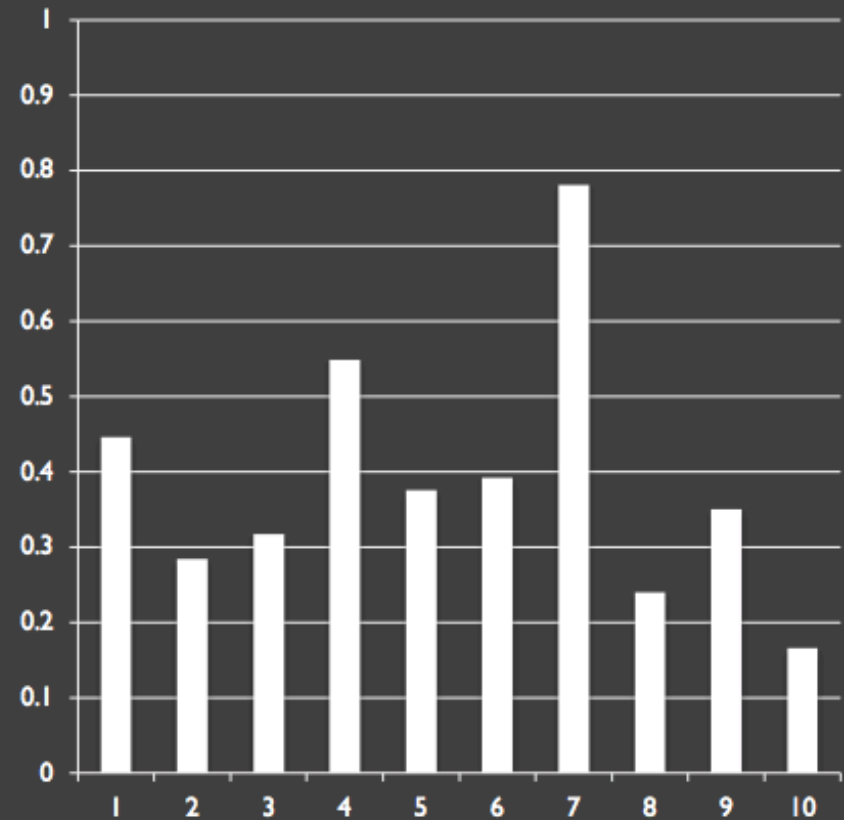
# Neither!



**Company A**



**Company B**



# Pareidolia

Perception of object or pattern as meaningful





# Things that can go wrong

1. People confuse uncertainty with certainty
2. People confuse signal with noise

# 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.

**A LOT**

# 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**

# Summary

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

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

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

Consider when uncertainty is high enough that doing nothing is the right thing to do.