



# CS-GY 6313 B: Information Visualization

10/31/2024

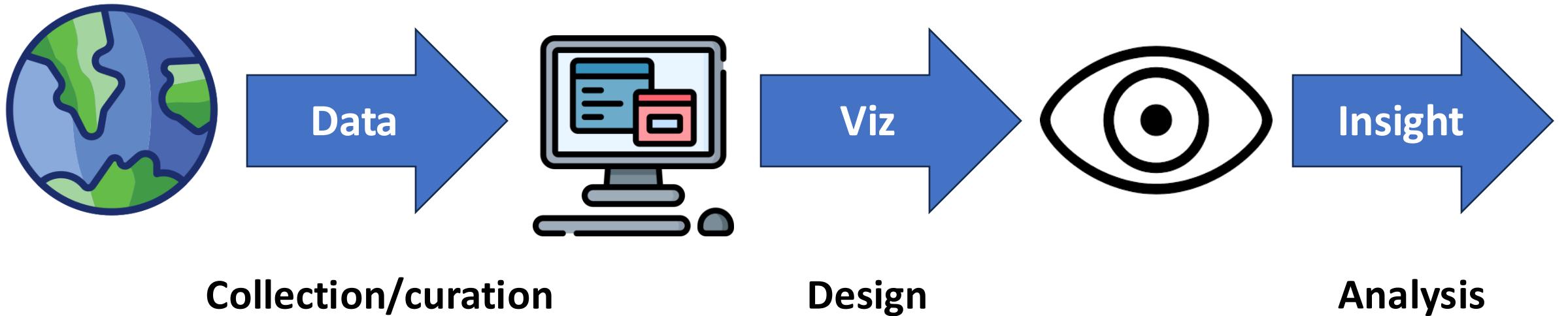
# Logistics

- Assignment 3 due Sunday
- Guest lecture Nov 21:
  - Dr. David Saffo @ JPMC
  - 3D + immersive viz
- Final project proposals next week!
  - Prepare a google slide deck
  - **MAKE SURE IT CAN BE VIEWED WITHOUT LOGGING IN**
  - Submit a link to your google slide deck

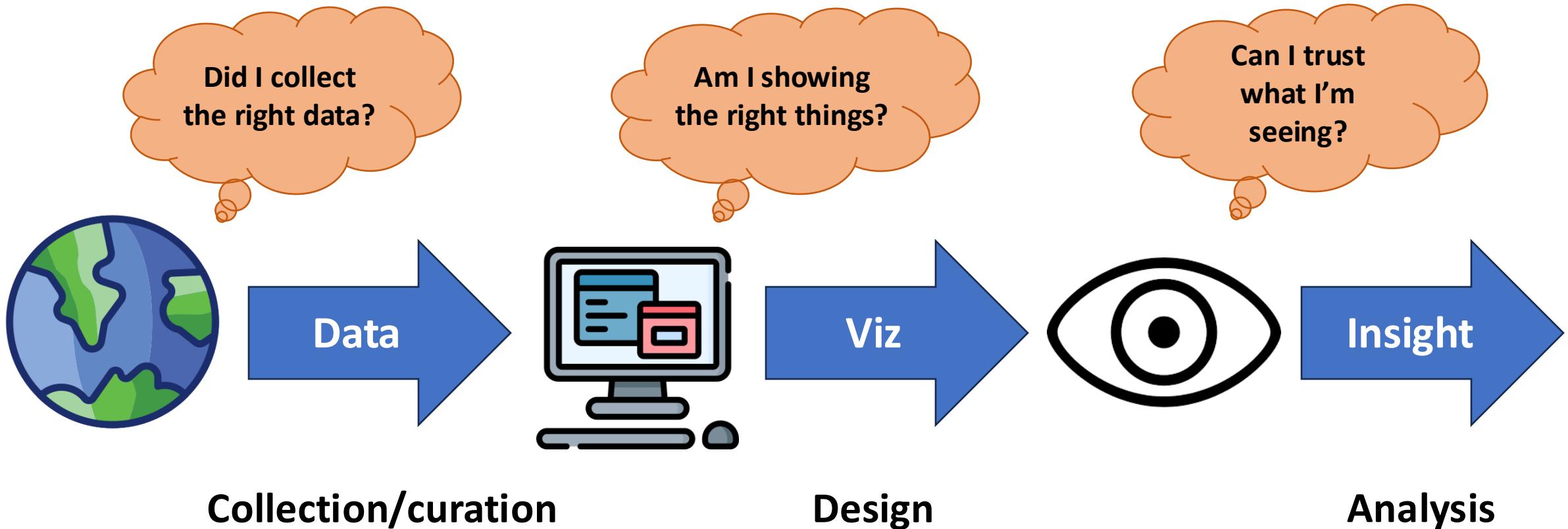


# Uncertainty

# The visualization pipeline



# The visualization pipeline?



# Unknown unknowns



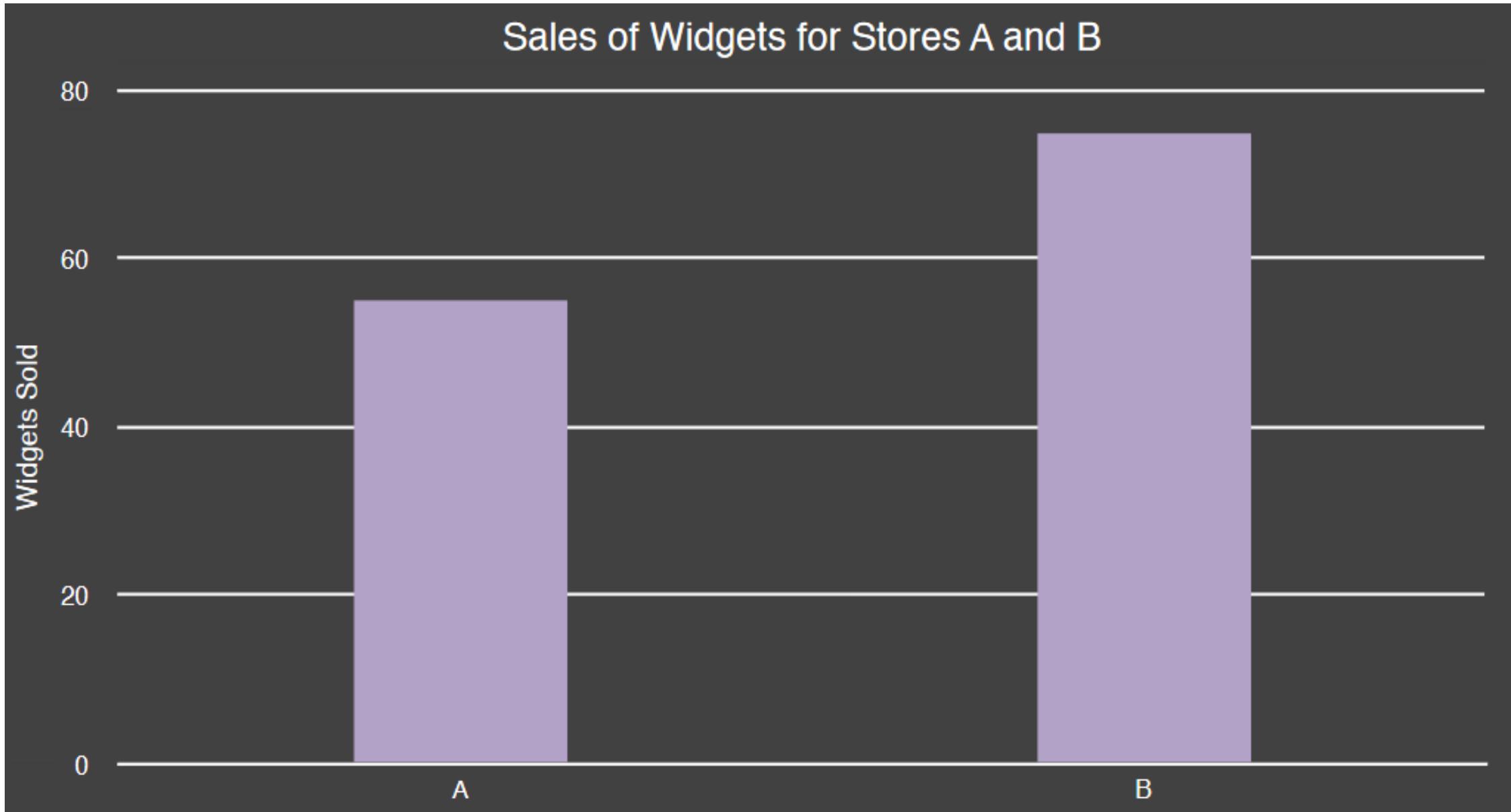
# What does “uncertainty” mean?

- Ideas?
  - Doubt
  - Risk
  - Variability
  - Error
  - Lack of knowledge
  - Hedging
  - ...

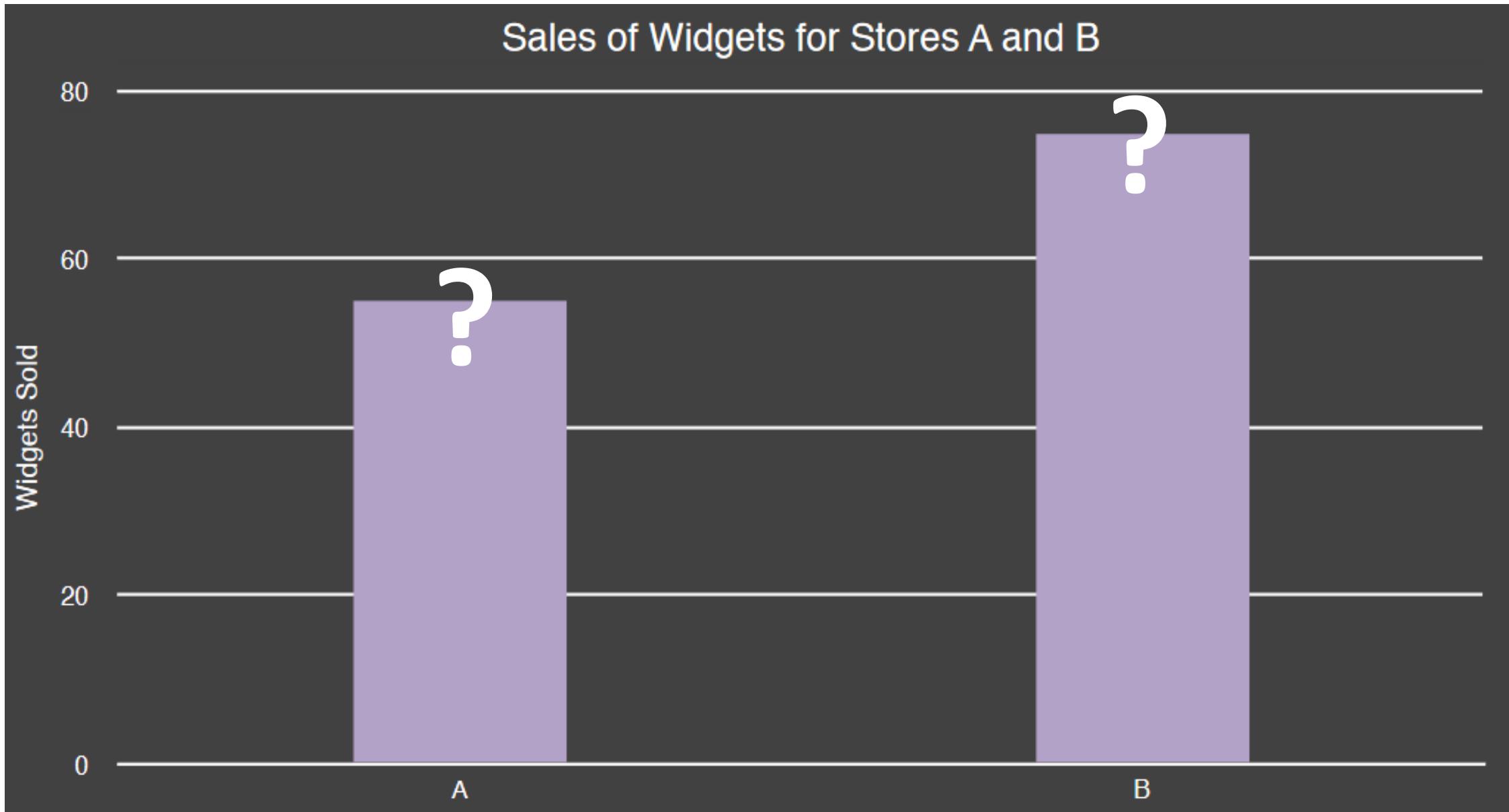
# Uncertainty sources

- **Measurement Uncertainty**
  - “We’re not sure what the data are”
- **Model Uncertainty**
  - “We’re not sure how the data fit together”
- **Forecast Uncertainty**
  - “We’re not sure what will happen to the data next”
- **Decision Uncertainty**
  - “We’re not sure what to do with the data”

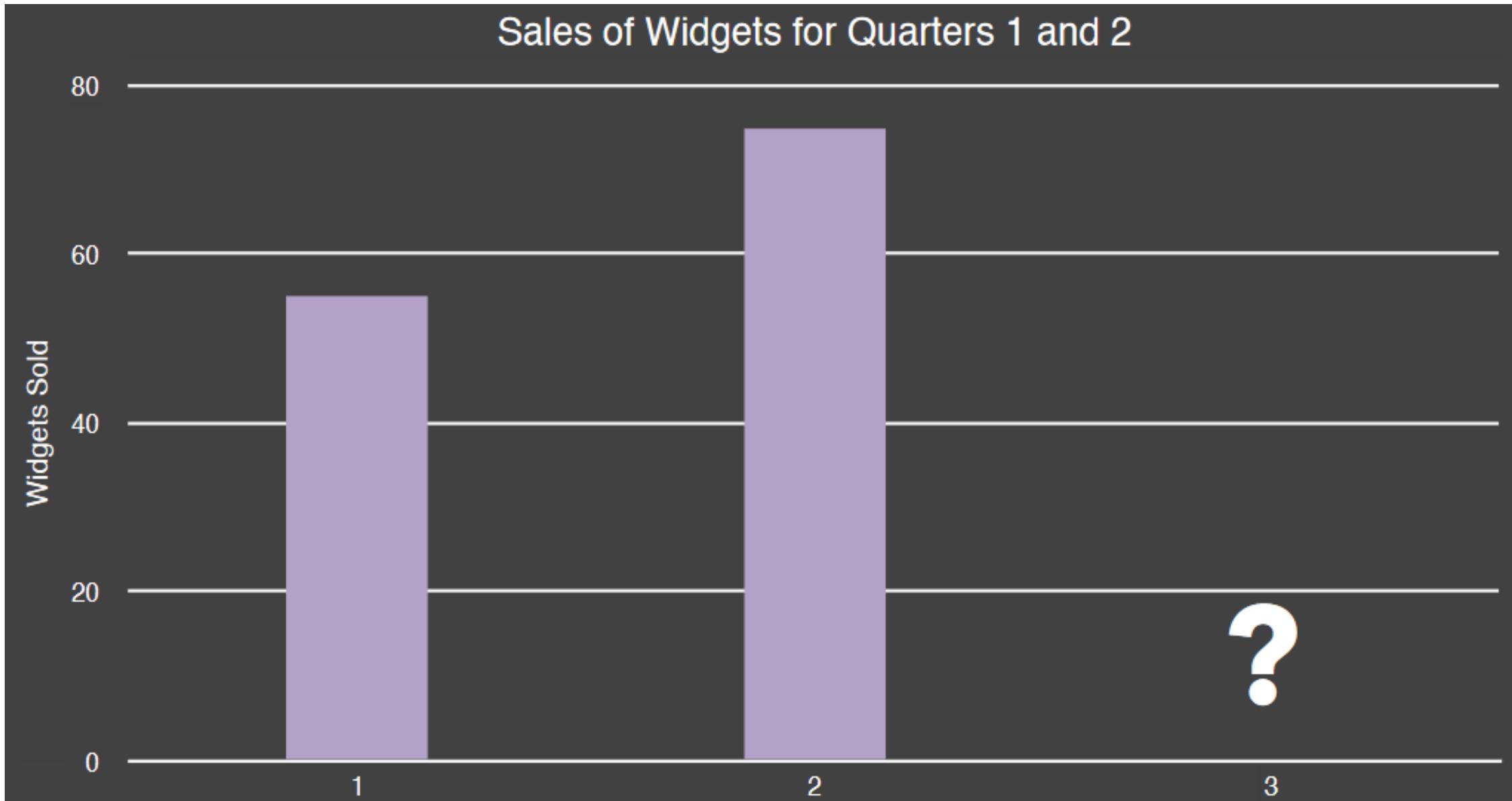
# A bar chart



# Measurement uncertainty



# Forecast uncertainty



# Decision uncertainty



# Types of error

- Uncertainty can cause errors

Should I bring an umbrella?



# Decision uncertainty

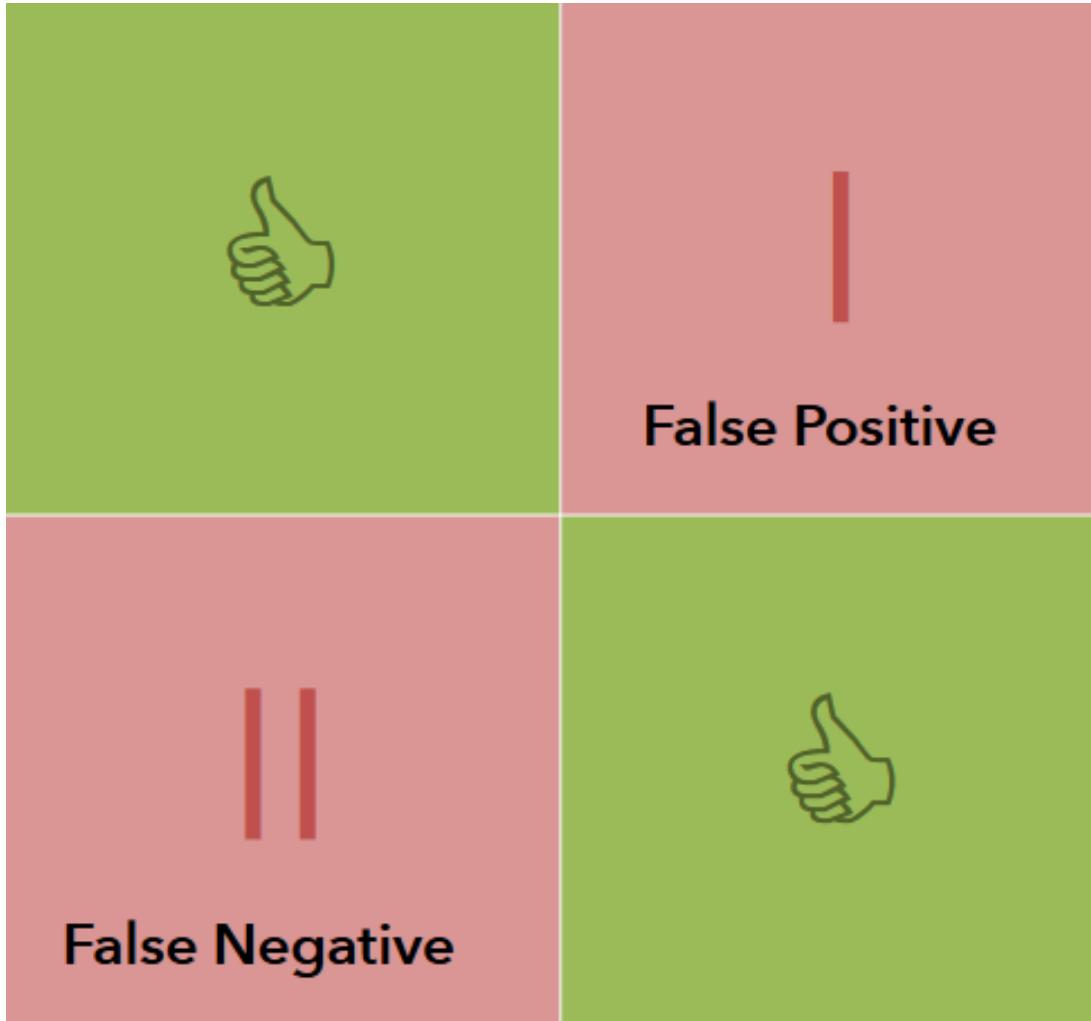
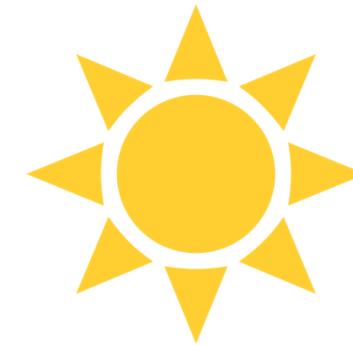
- Uncertainty can cause errors

“50% chance of rain”



# Types of error

- Uncertainty can cause errors



# The boy who cried wolf

Type I: False Positive



Type II: False Negative

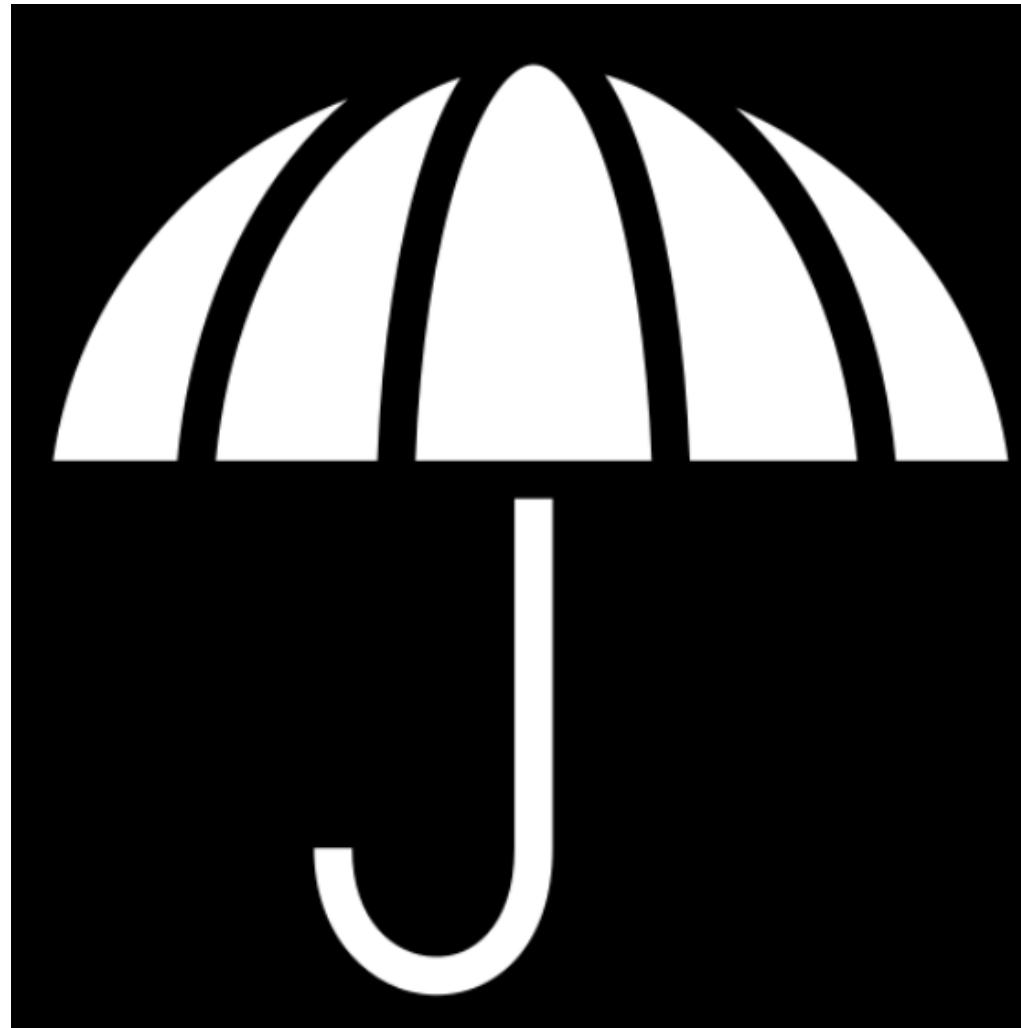


# Types of error

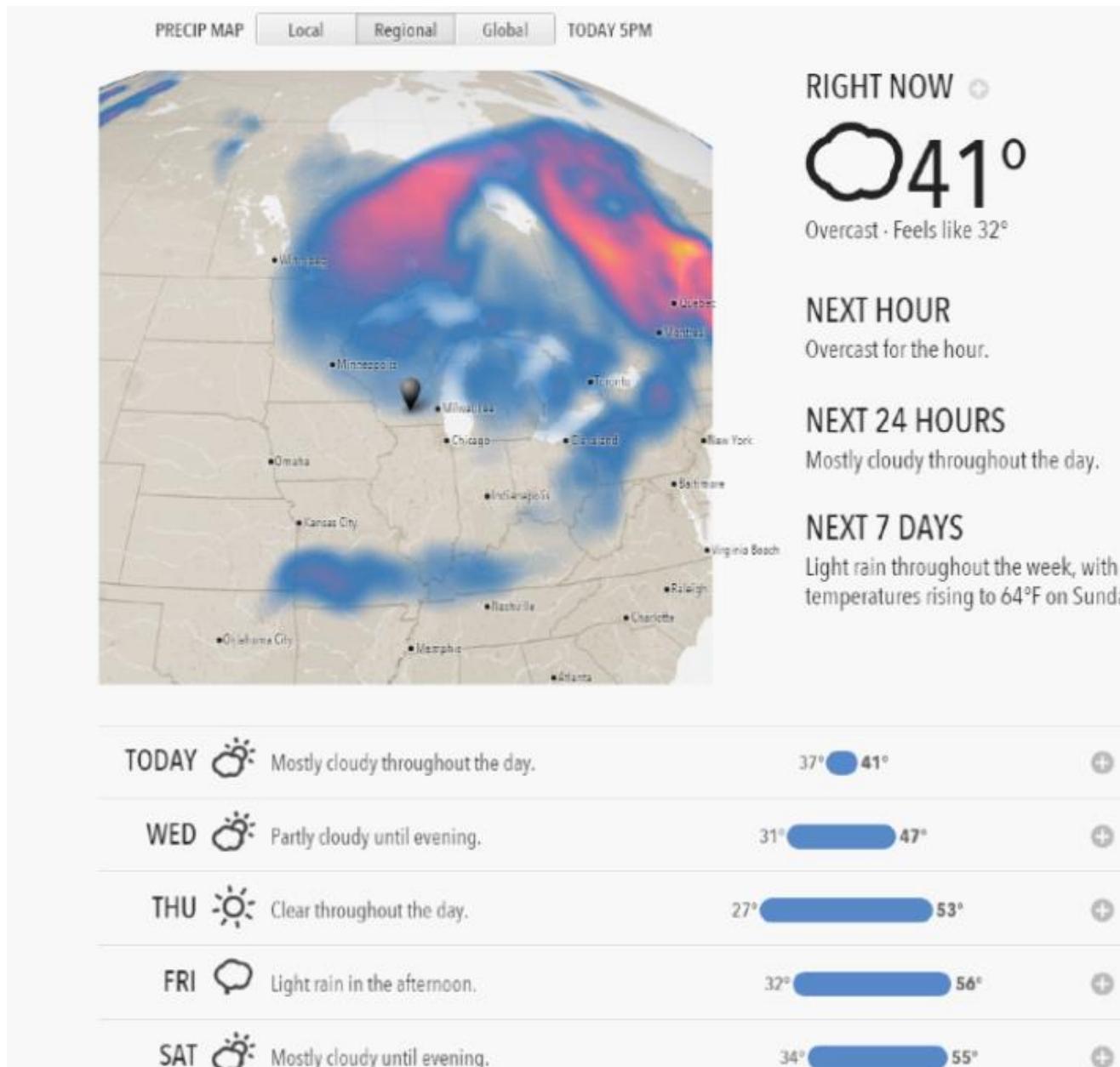
<u>TYPE</u>
FALSE POSITIVE
FALSE NEGATIVE

# Model uncertainty

“50% chance of rain”



# Model uncertainty



# What does uncertainty mean?

- Any one of a number of potentially interconnected quantitative or qualitative 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.

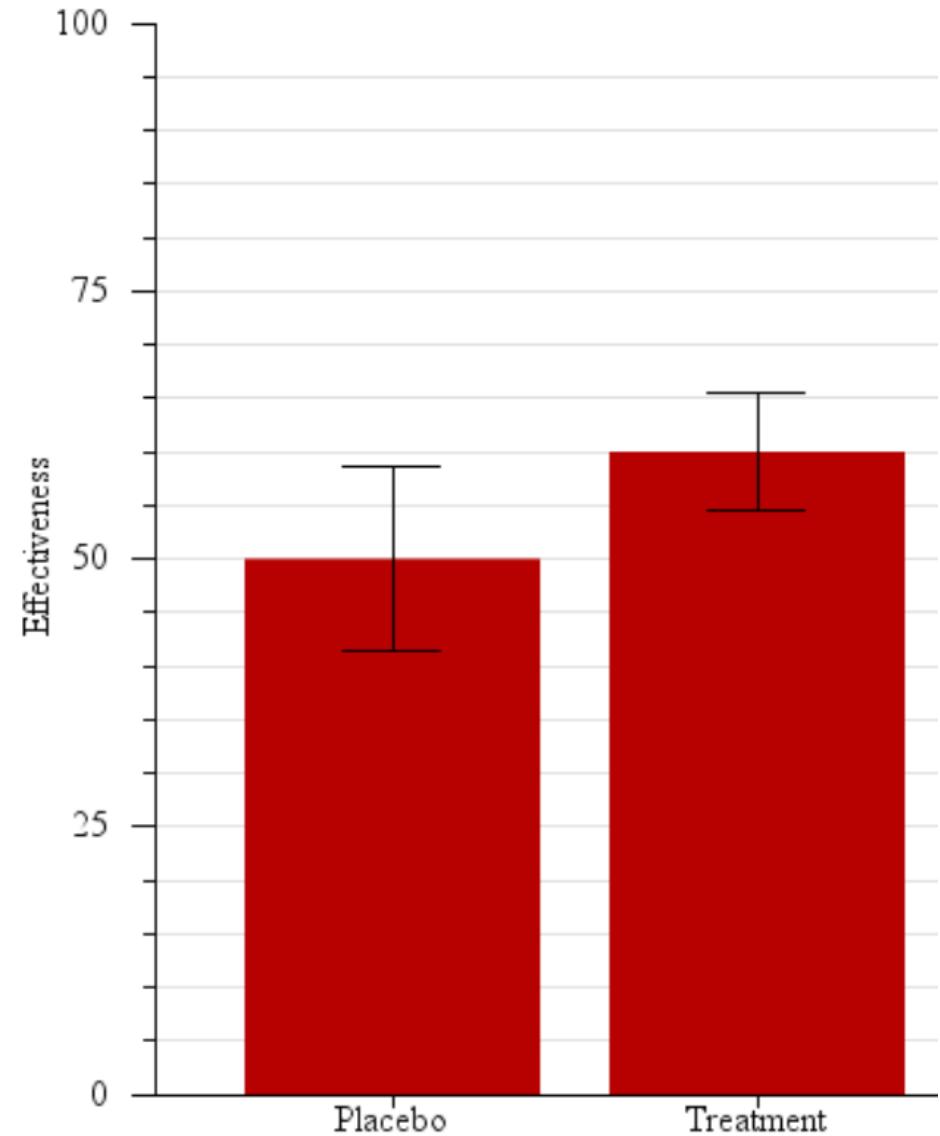
**LOTS OF THINGS**

# 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.
  
- Make sure: don't prematurely suppress uncertainty!

# Error bars

- Represents the variability in the data.
- Can represent any of the following:
  - Standard deviation
  - Standard error
  - Confidence interval
- Must specify!



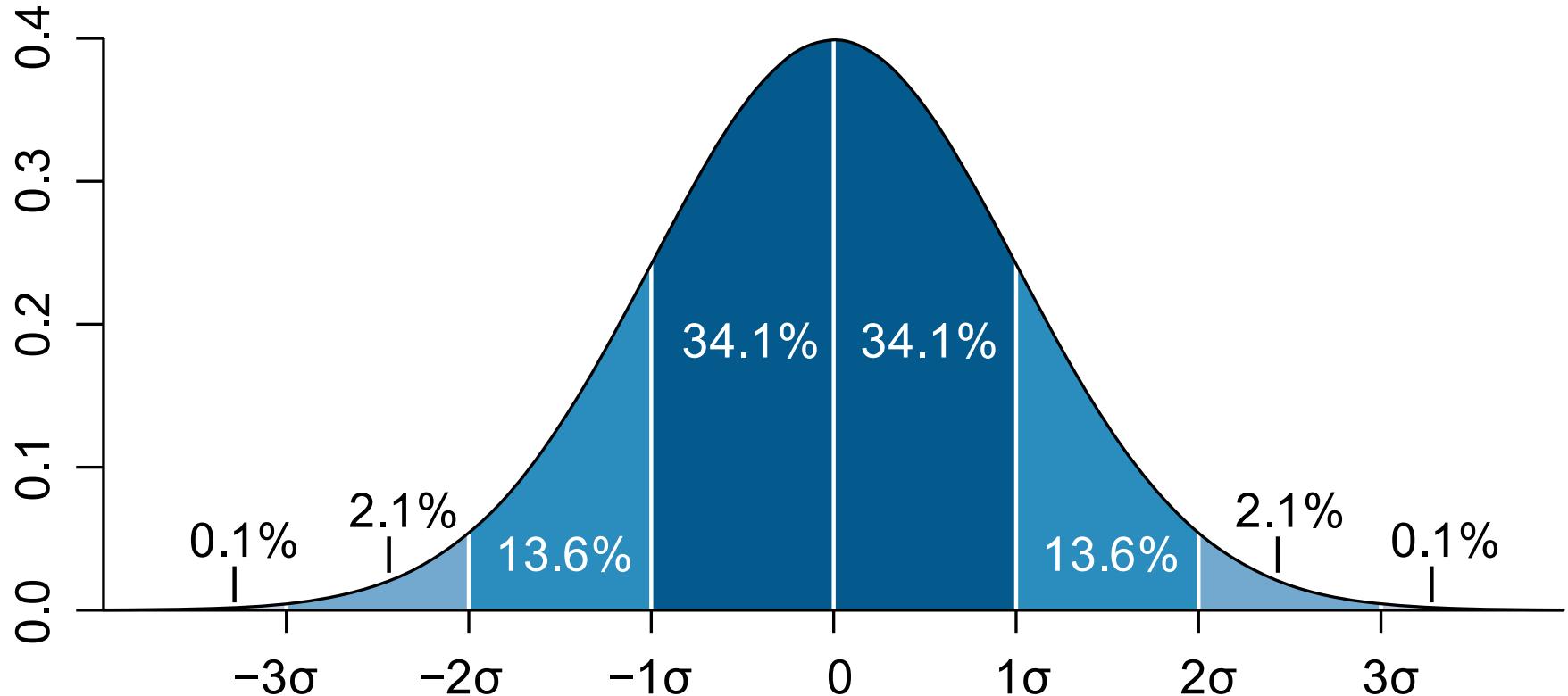
# Error bars: standard deviation

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

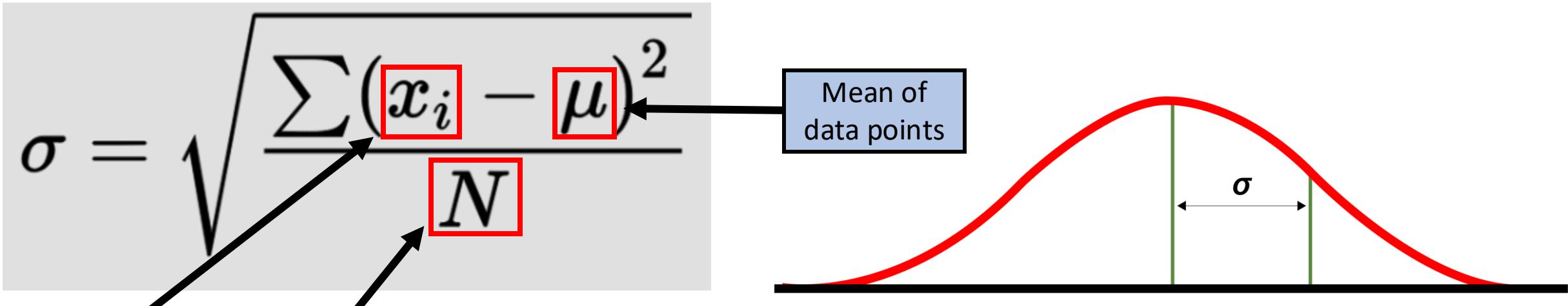
Mean of data points

One data point

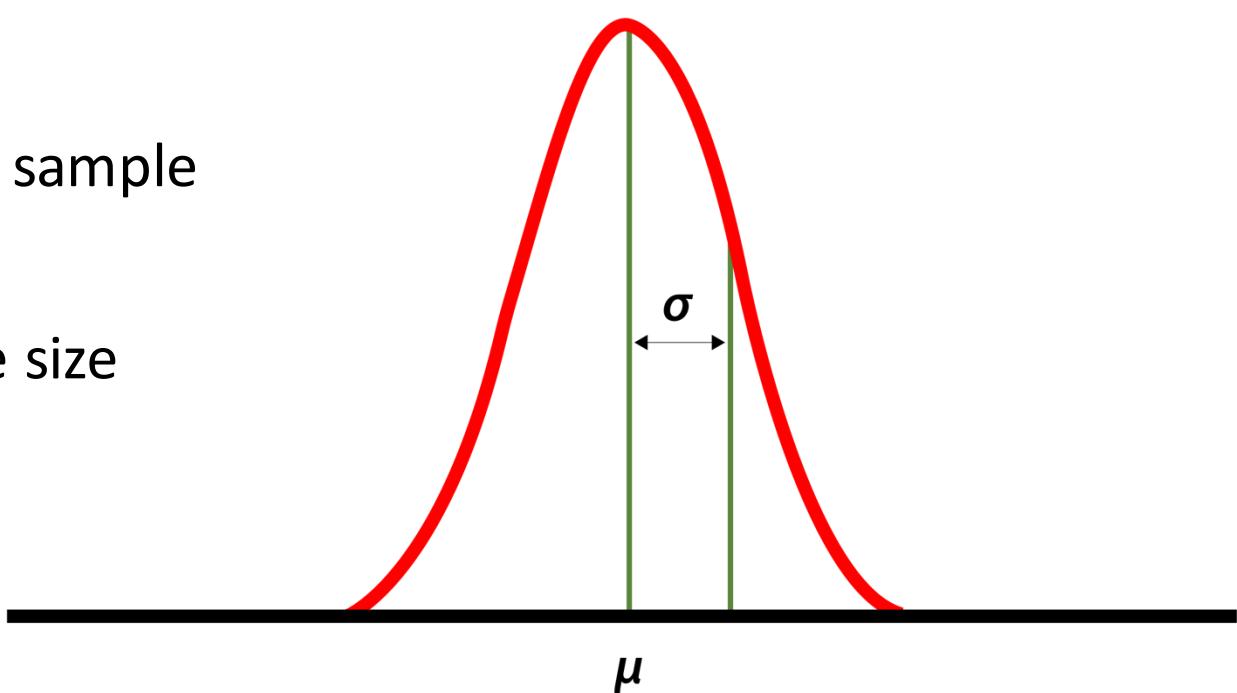
Total # of data points



# Error bars: standard deviation



- Variation of individual data points within a sample
- Shows spread of data from the mean
- Doesn't change systematically with sample size



# Error bars: standard error

$$SE = \frac{\sigma}{\sqrt{n}}$$

Standard deviation

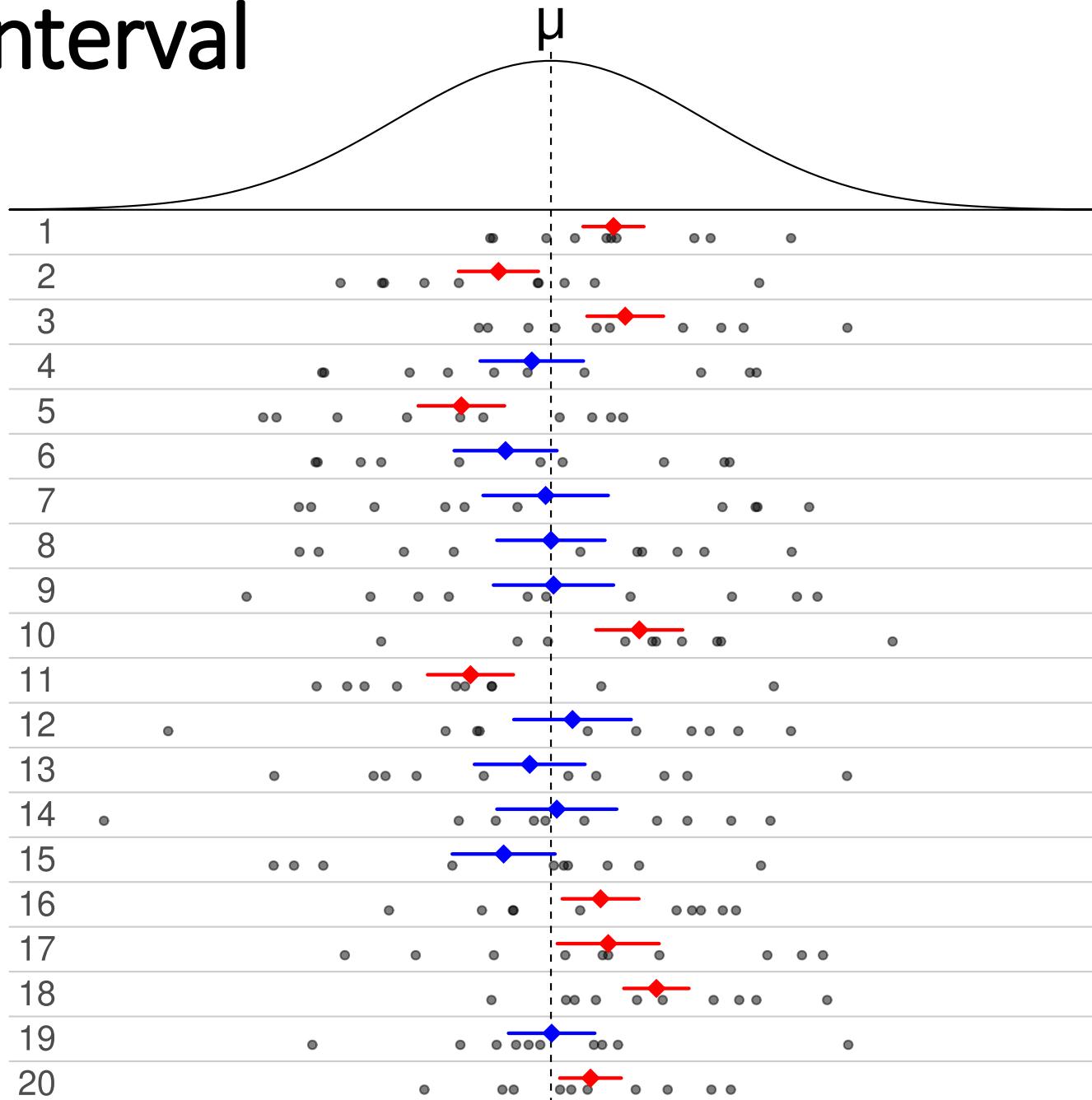
Number of samples

The diagram illustrates the formula for Standard Error (SE). It shows  $SE = \frac{\sigma}{\sqrt{n}}$ . A red box highlights the standard deviation symbol ( $\sigma$ ). A blue box highlights the square root of  $n$  term ( $\sqrt{n}$ ). Arrows point from a 'Standard deviation' callout to the  $\sigma$  box and from a 'Number of samples' callout to the  $\sqrt{n}$  box.

- Variation of sample means
- Shows how accurately your sample represents the population
- Decreases as sample size increases

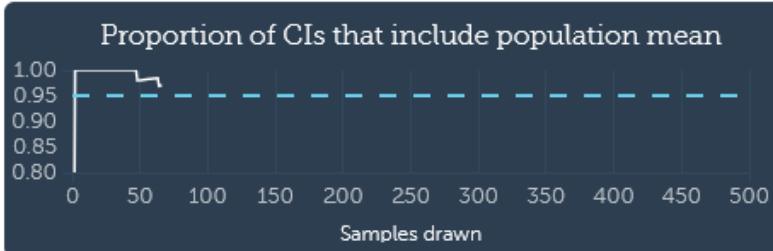
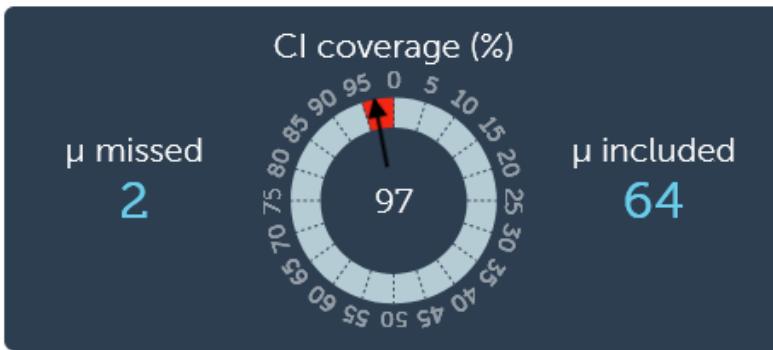
# Error bars: confidence interval

- We want to estimate some parameter of the whole population
- Can't collect data on **everyone**, so we take a sample of the population
  - Estimate the population parameter from our sample
- But what if we got (un)lucky with our sample?
- Confidence interval tells us how likely it is that our sample's parameter estimate is close to the true population parameter
  - E.g.  $p < 0.001$  \*\*\*

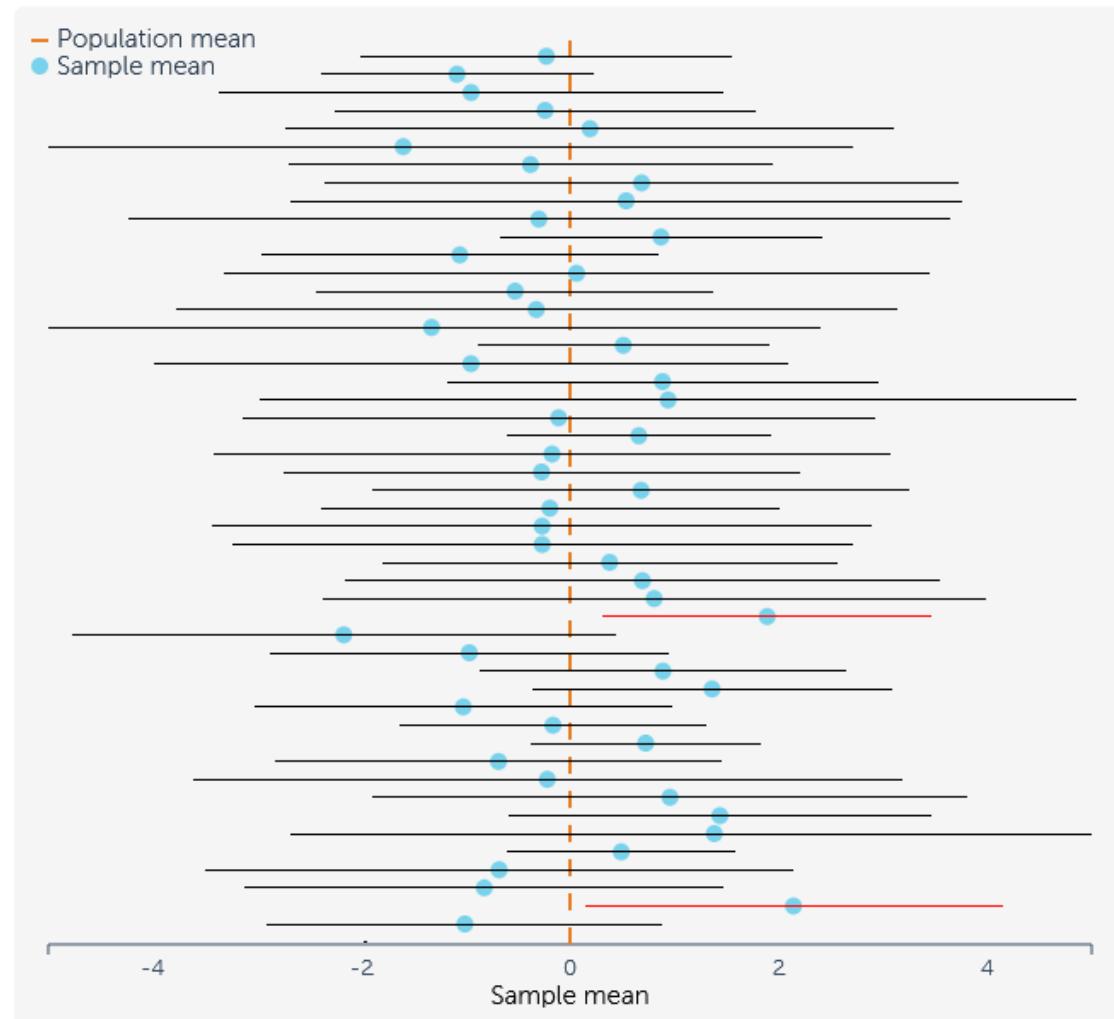


# Error bars: confidence interval

Simulation statistics



95% confidence intervals



Sample size

- | +

Speed

- | +

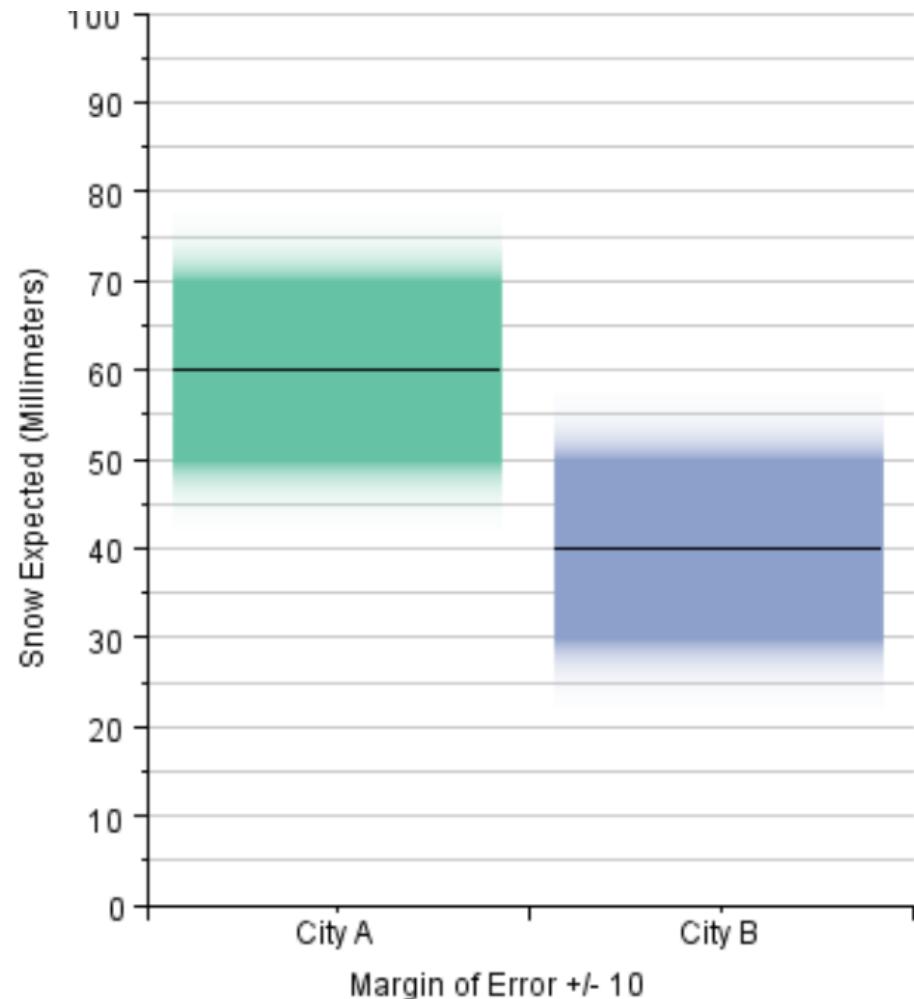
# Error bars: saliency

- Be careful about over-emphasizing the error bars
  - Importance of error bars depends on the task!
- For inference tasks, focus on the **uncertainty** (error bars), not the point estimates!
  - Uncertainty will tell us if the differences in point estimates are meaningful

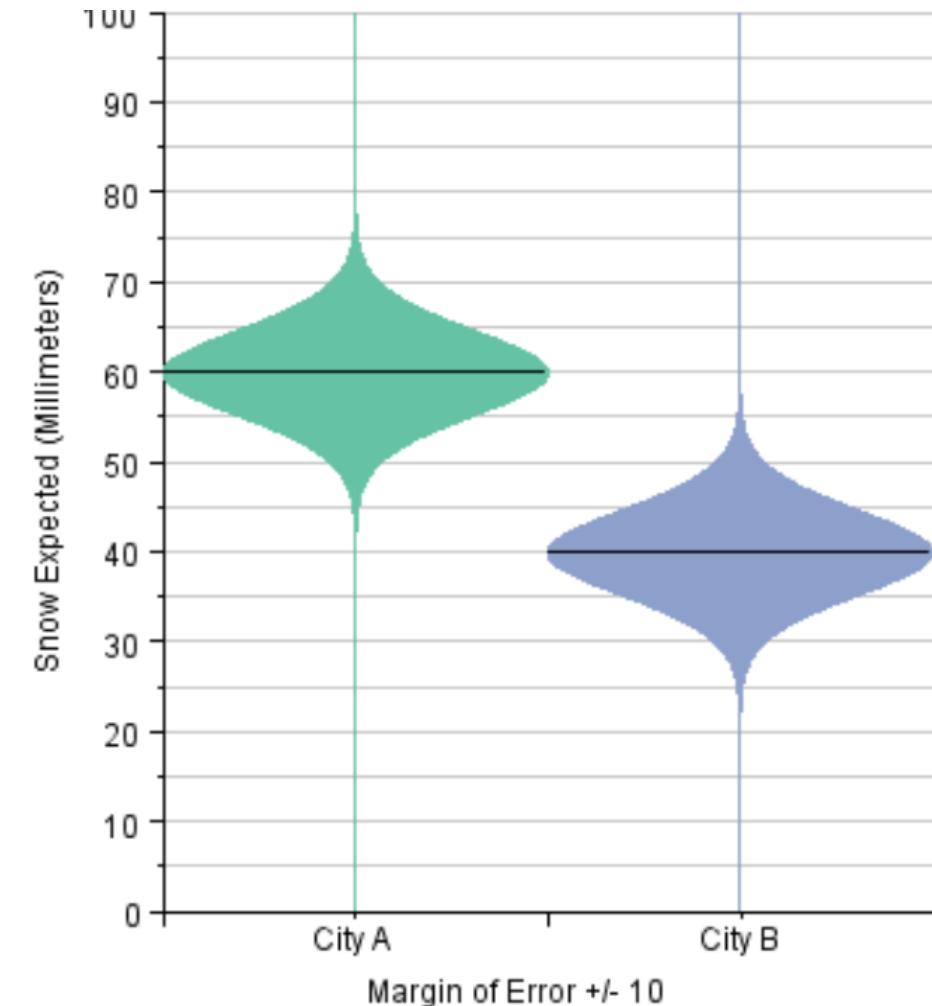


# Error bar alternatives

- Gradient plot



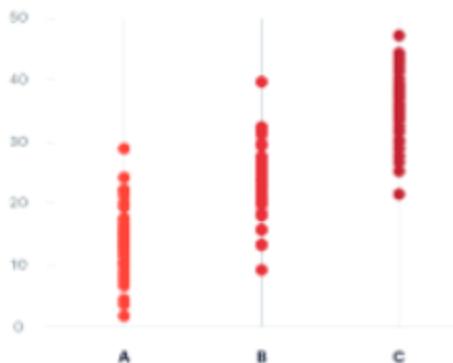
- Violin plot



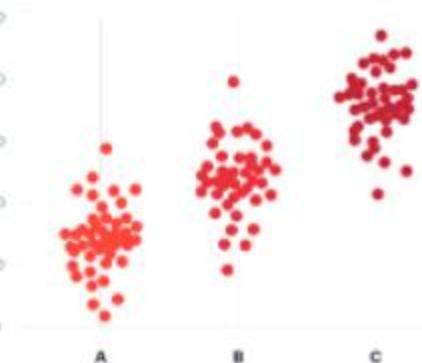
# Visualizing distributions of data

- Your dataset is a distribution of data
  - Many measurement samples
- These samples span some range of values
- How to visualize this range?

Strip plot



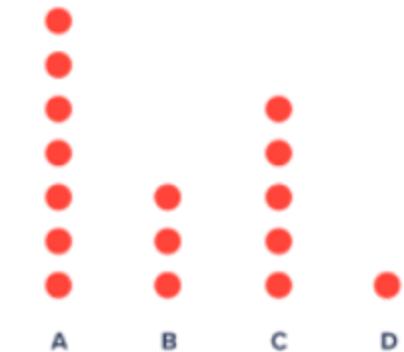
Jitter plot



Box plot



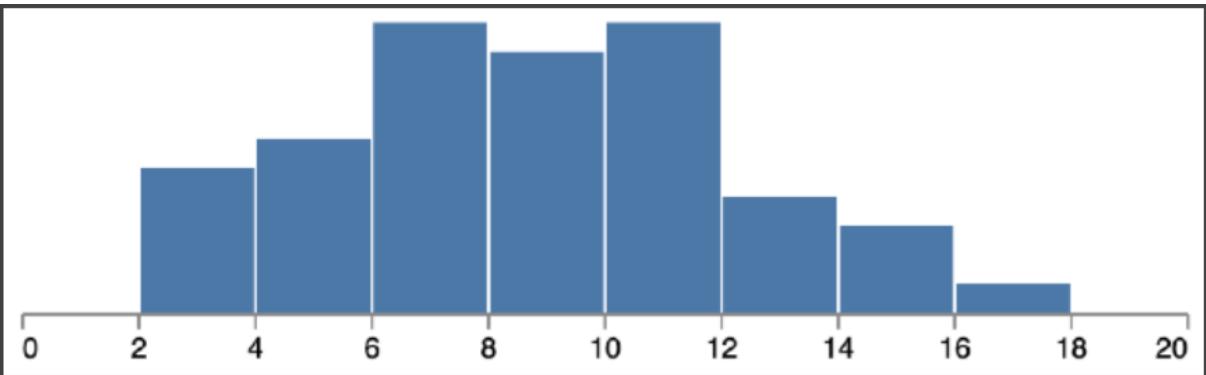
Dot plot



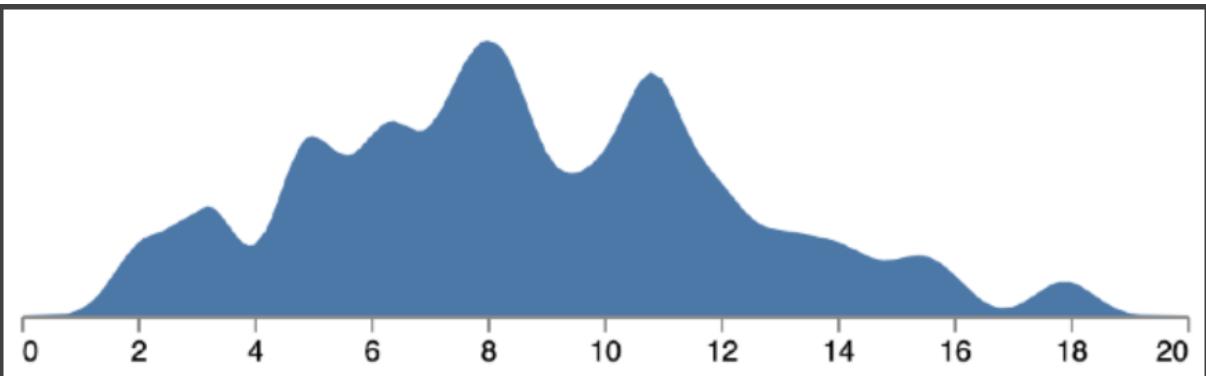
# Visualizing distributions of data

- More...

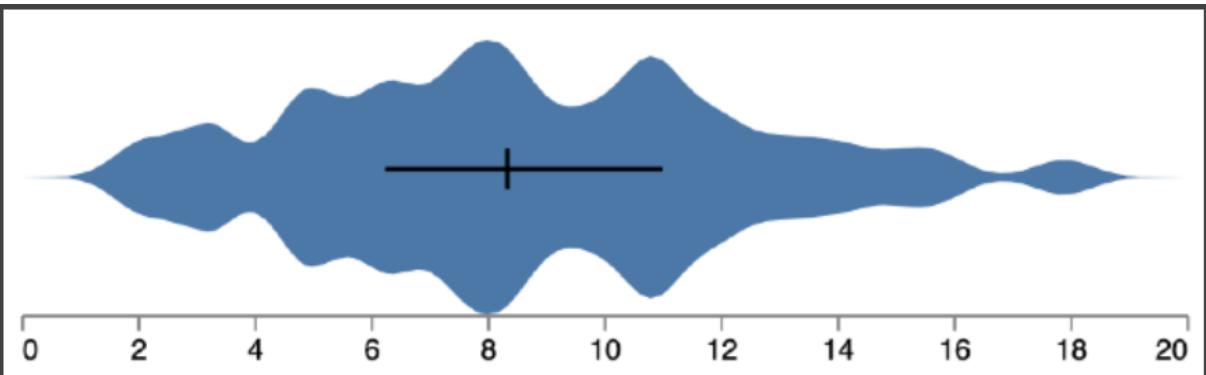
Histogram



Density plot

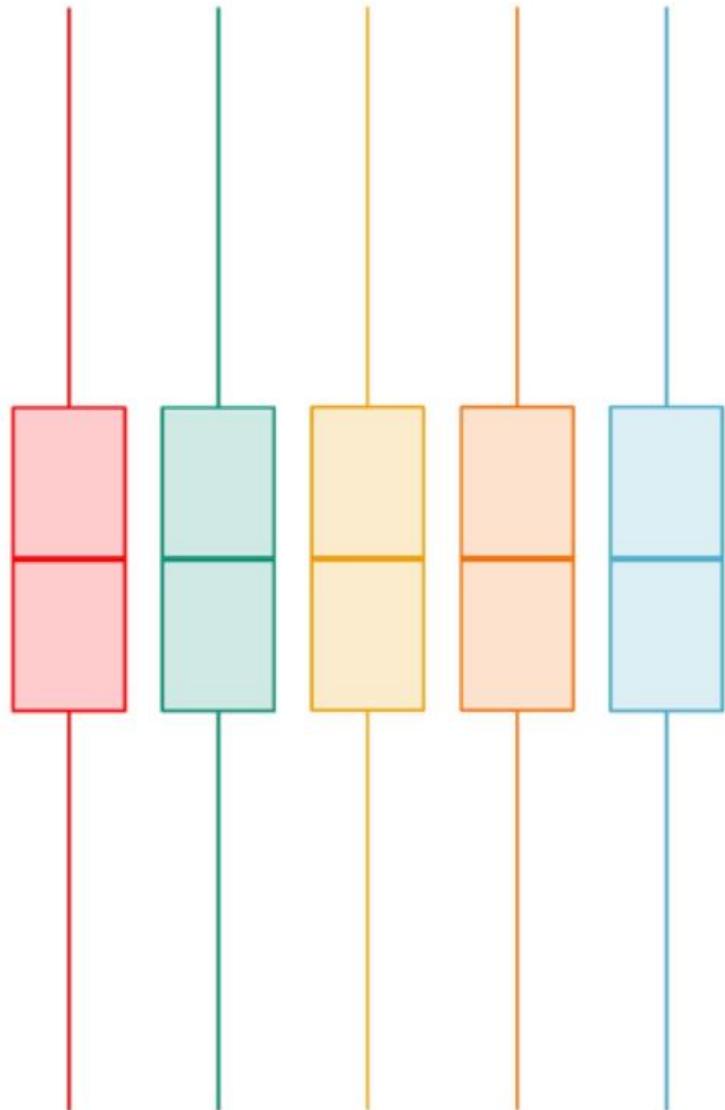


Violin plot



# Visualizing distributions of data

Box plot



Jitter plot

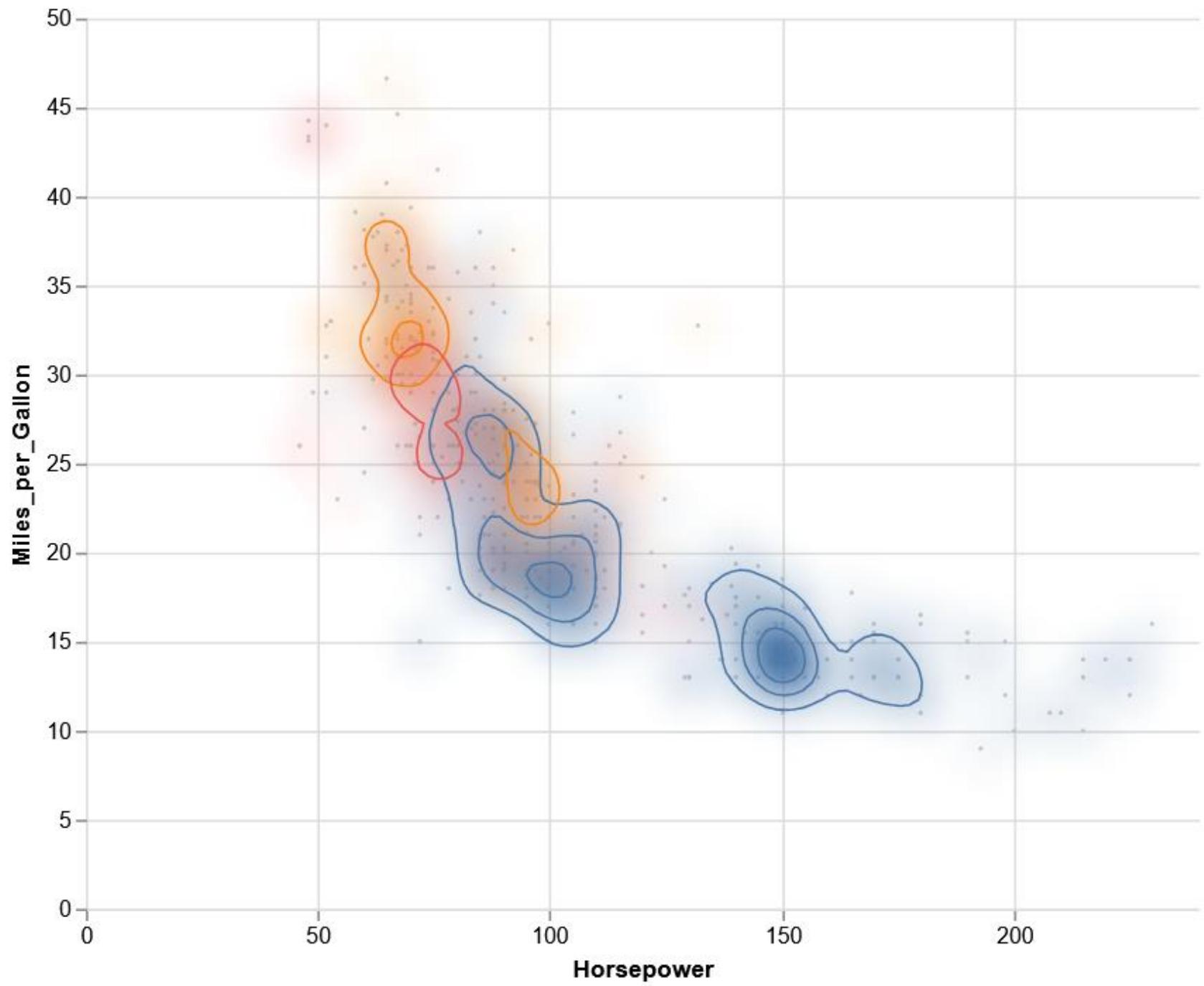


Violin plot



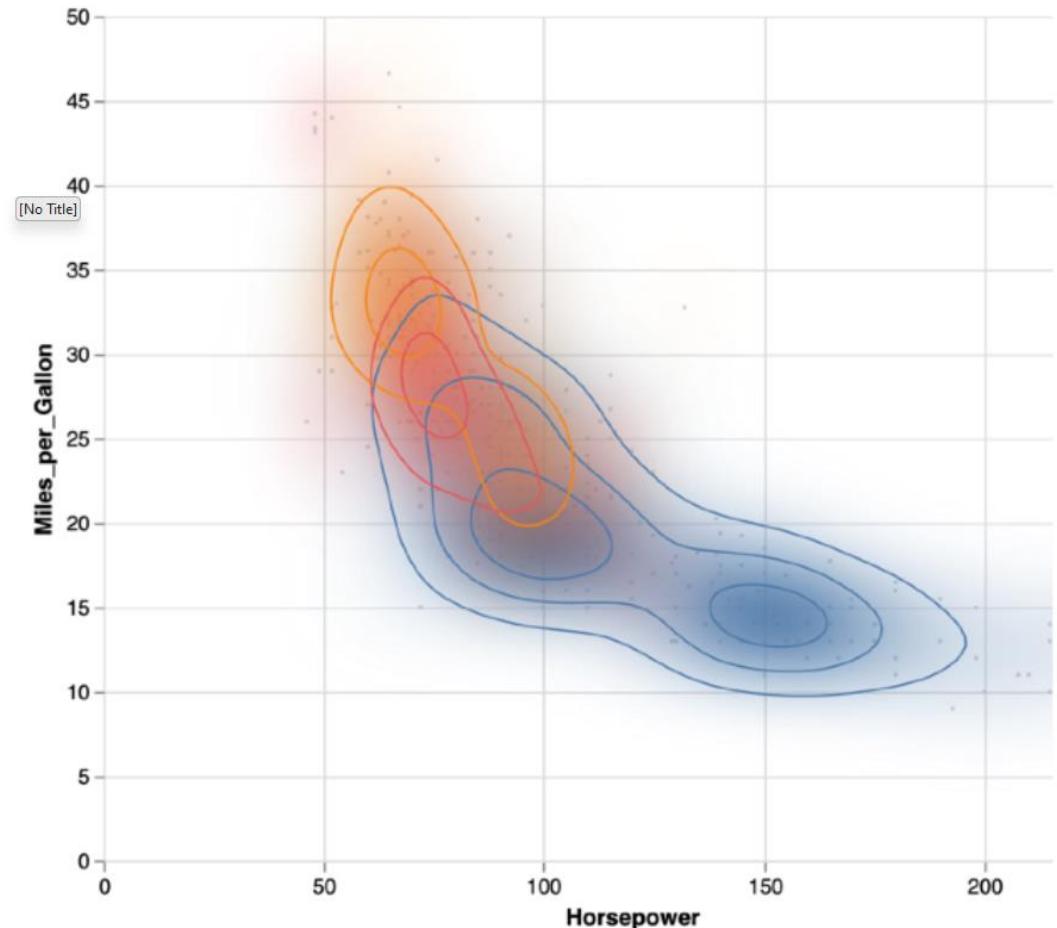
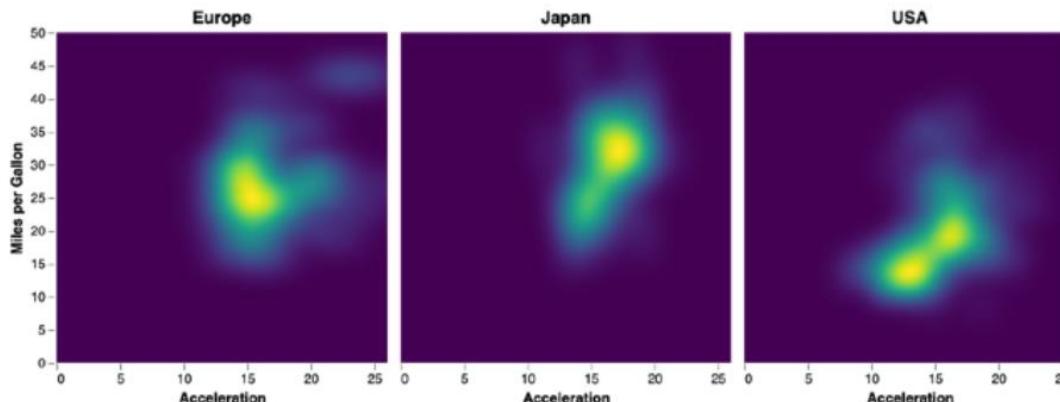
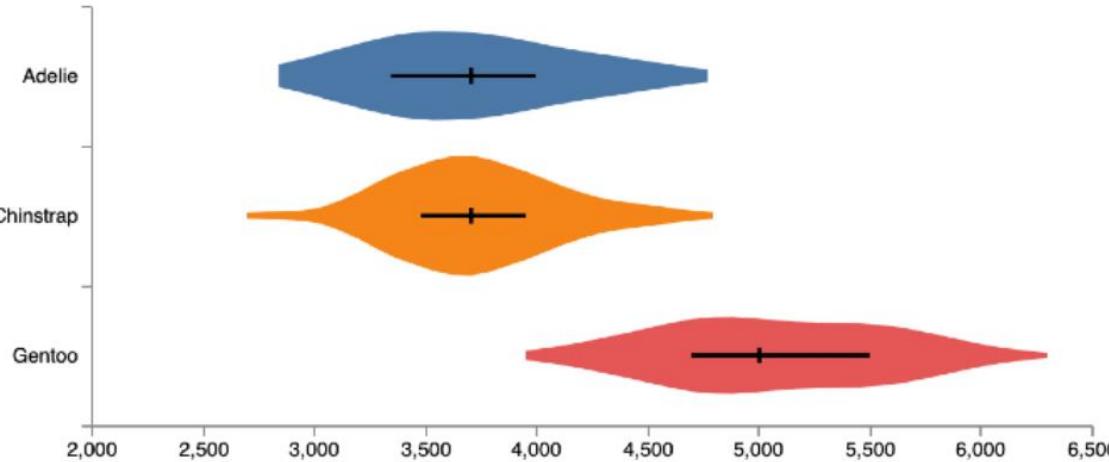
# Now in 2D

- <https://vega.github.io/vega/examples/contour-plot/>



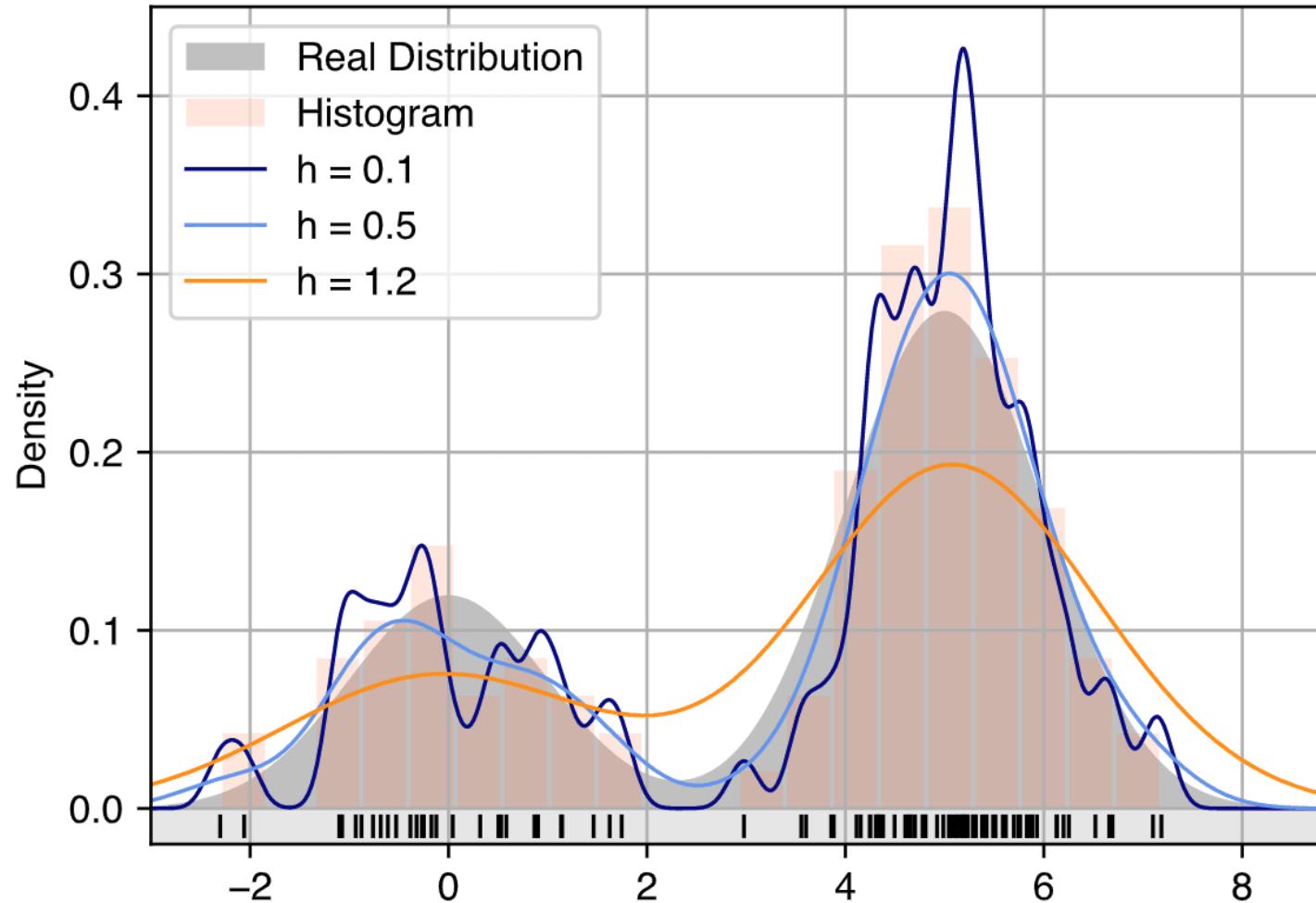
# Kernel Density Estimation

- Given discrete samples, how to estimate continuous probability density?
  - Kernel density estimation!
- Backbone of density plots, violin plots, contour plots...



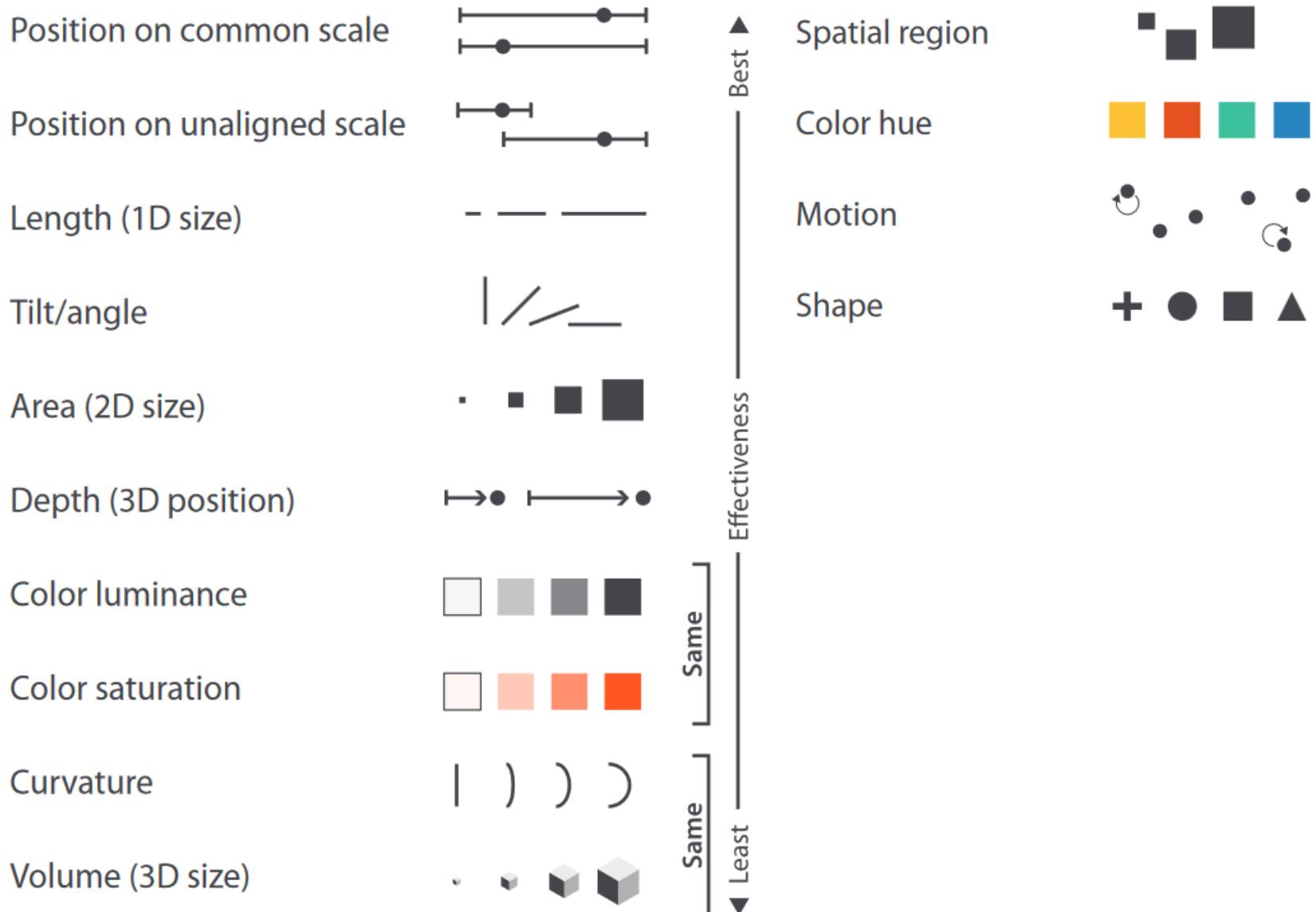
# Kernel Density Estimation

- Given discrete samples, how to estimate continuous probability density?
  - Kernel density estimation!
- Backbone of density plots, violin plots, contour plots...



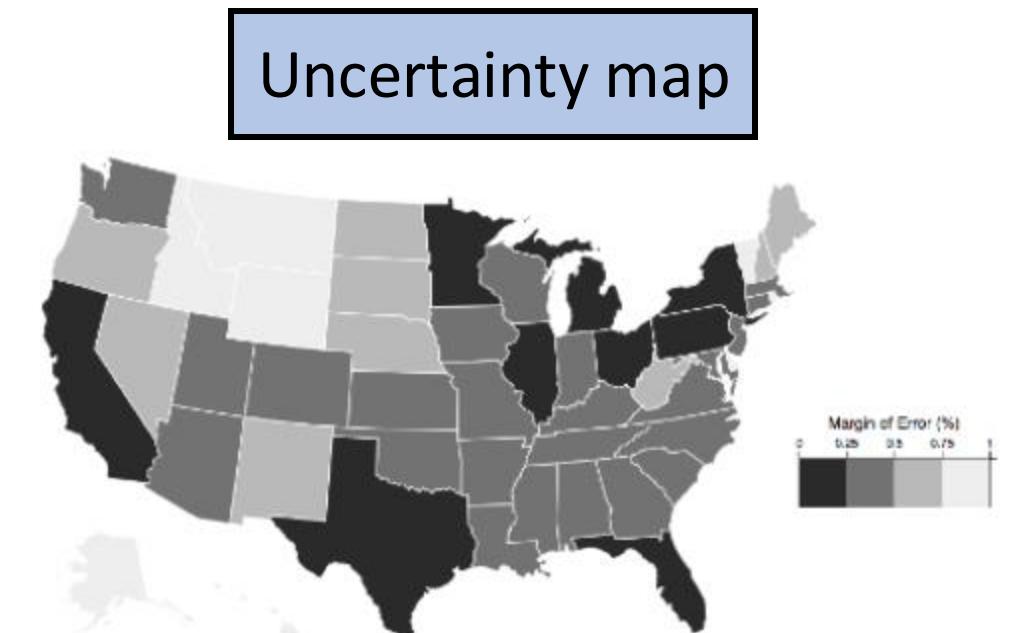
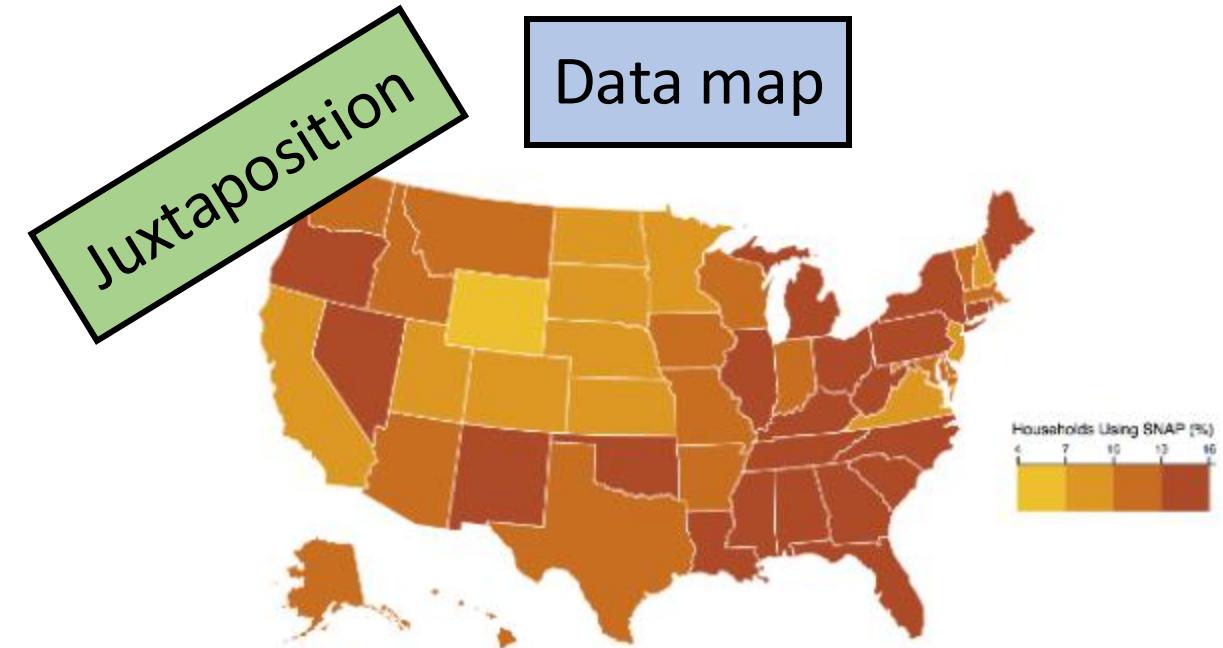
# Visual encodings of uncertainty

- What's the best way to visually encode uncertainty?



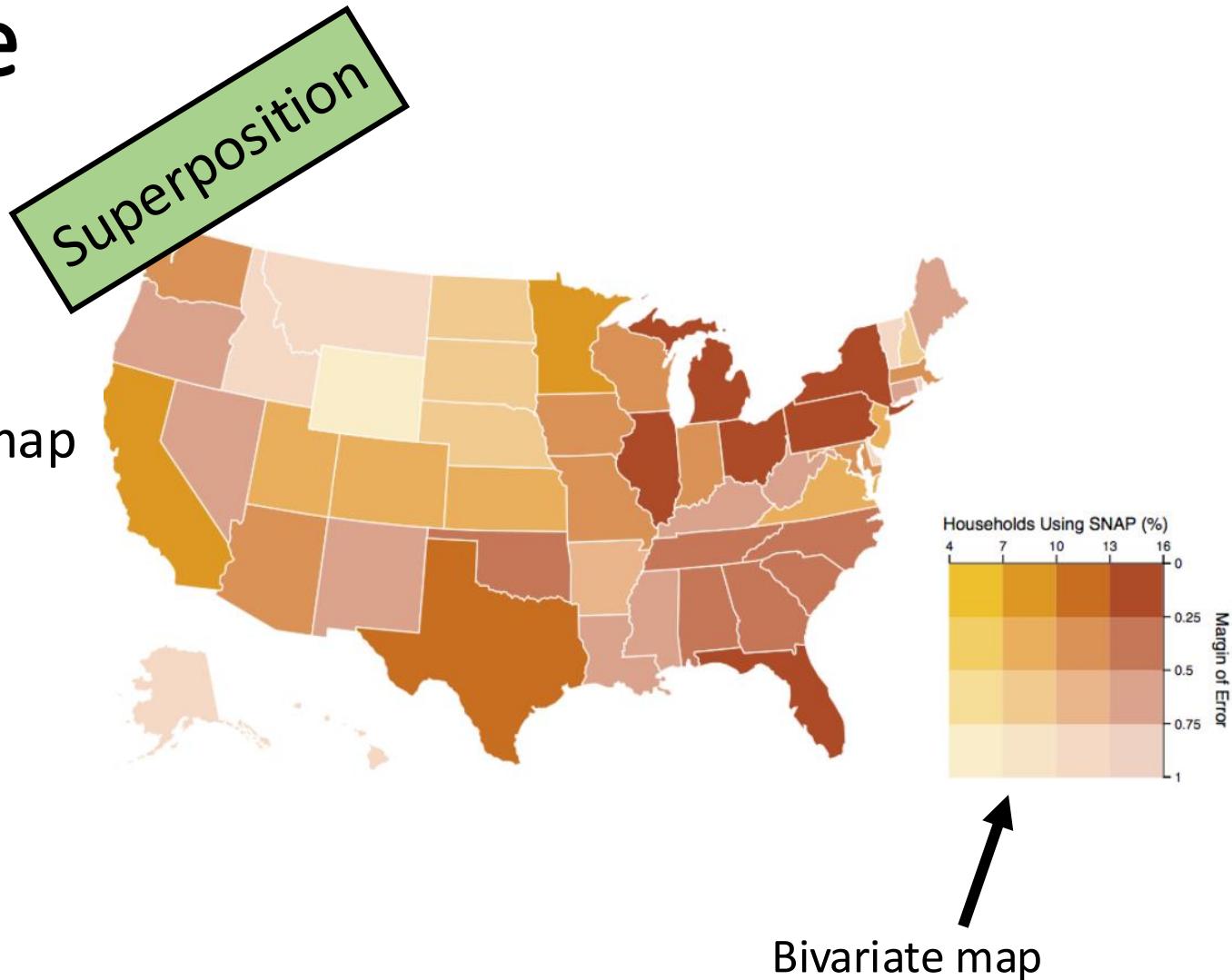
# Uncertainty viz pipeline

1. Quantify uncertainty
2. Choose a free visual channel
3. Encode uncertainty with that channel
4. Unify the data map and uncertainty map



# Uncertainty viz pipeline

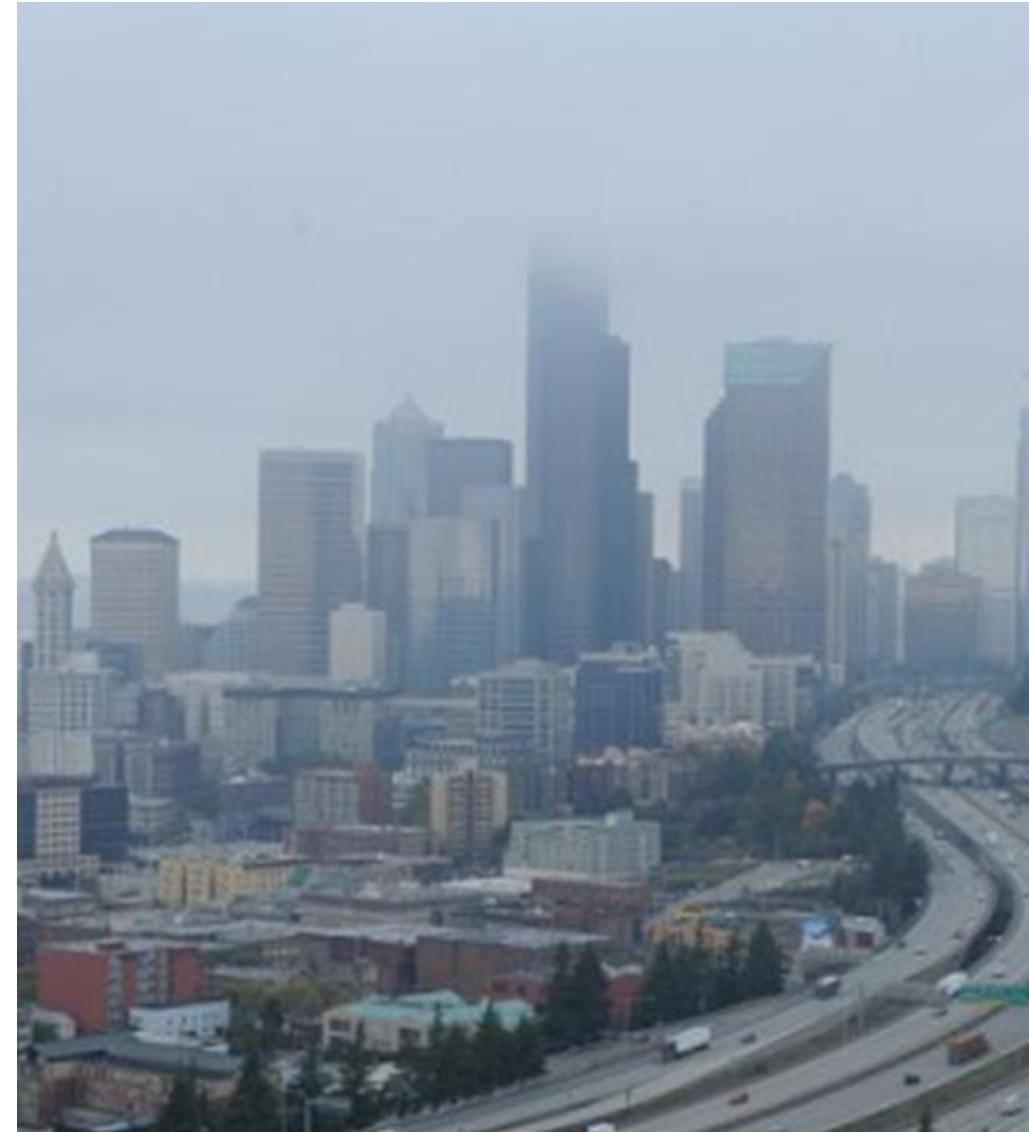
1. Quantify uncertainty
2. Choose a free visual channel
3. Encode uncertainty with that channel
4. Unify the data map and uncertainty map



# Semiotics of uncertainty



# Semiotics of uncertainty



# Semiotics of uncertainty

- Which ones are effective encodings?
  - MacEachren, Alan M., et al "Visual semiotics & uncertainty visualization: An empirical study." IEEE transactions on visualization and computer graphics 18.12 (2012): 2496-2505.



*location*



*size*



*color hue*



*color value*



*color saturation*



*orientation*



*grain*



*arrangement*



*shape*

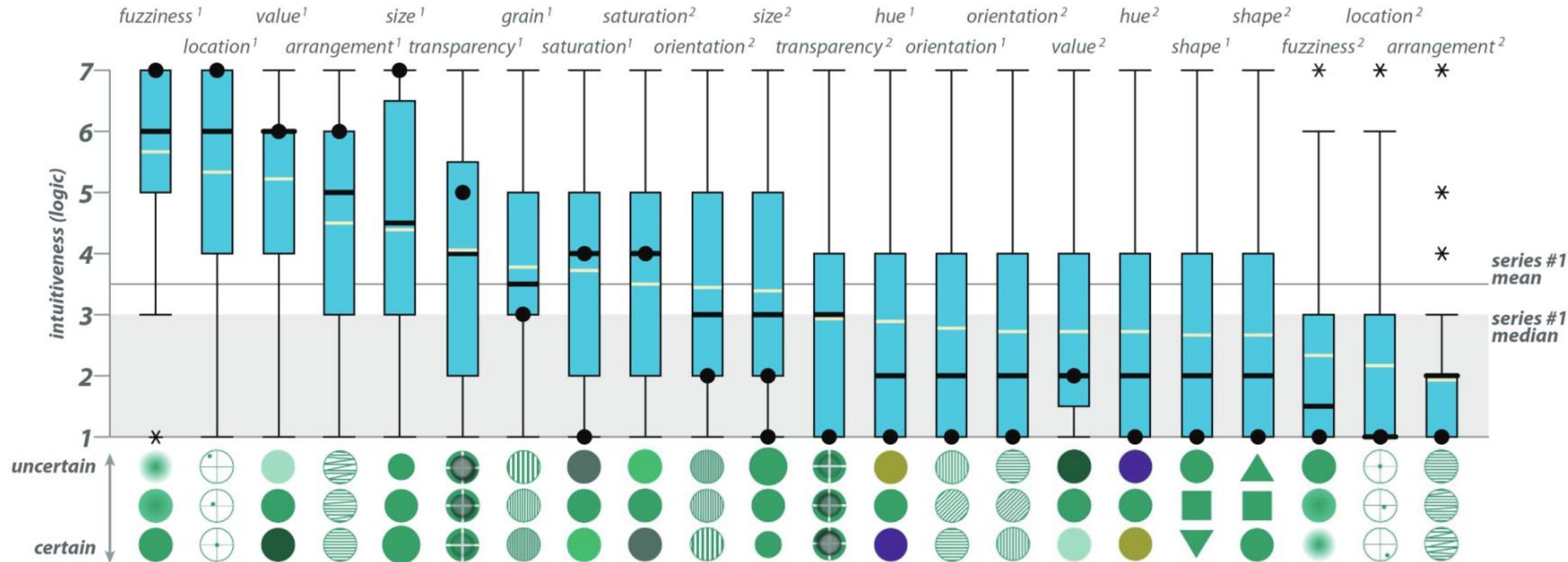


*fuzziness*



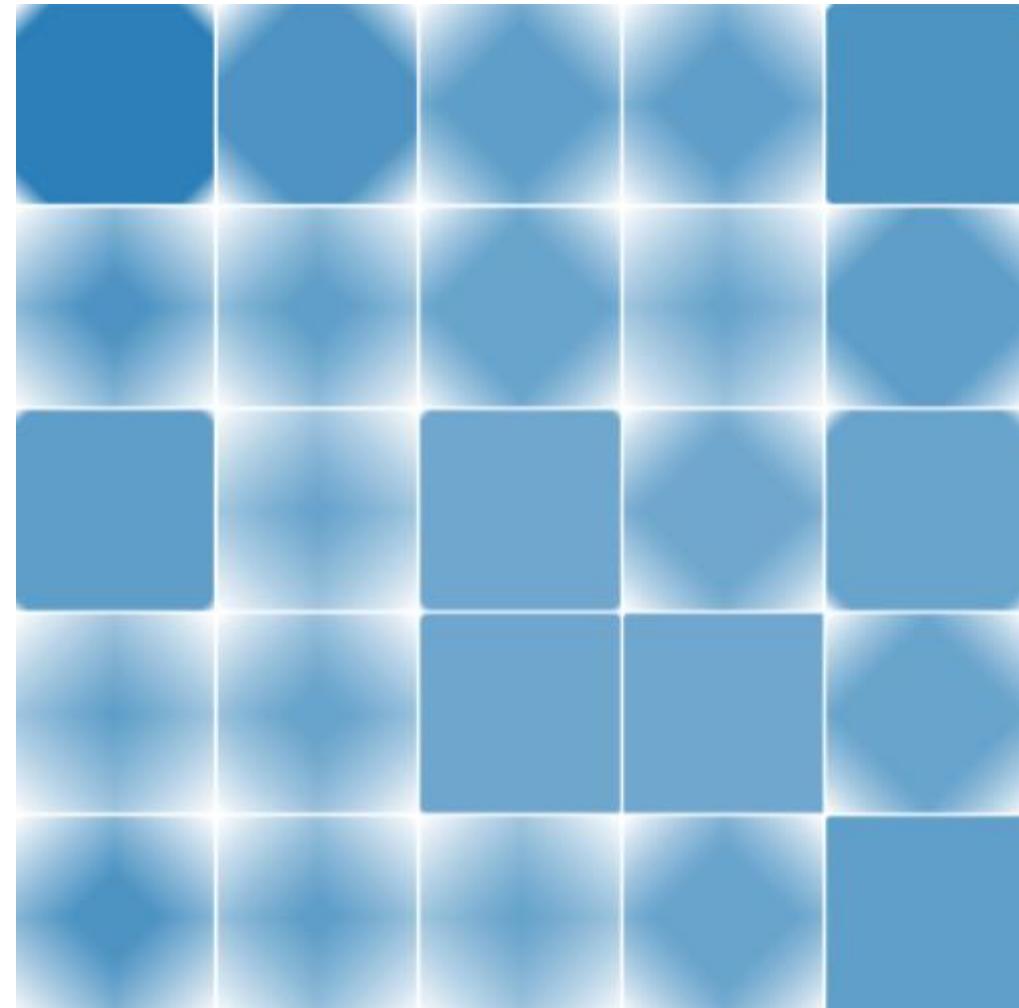
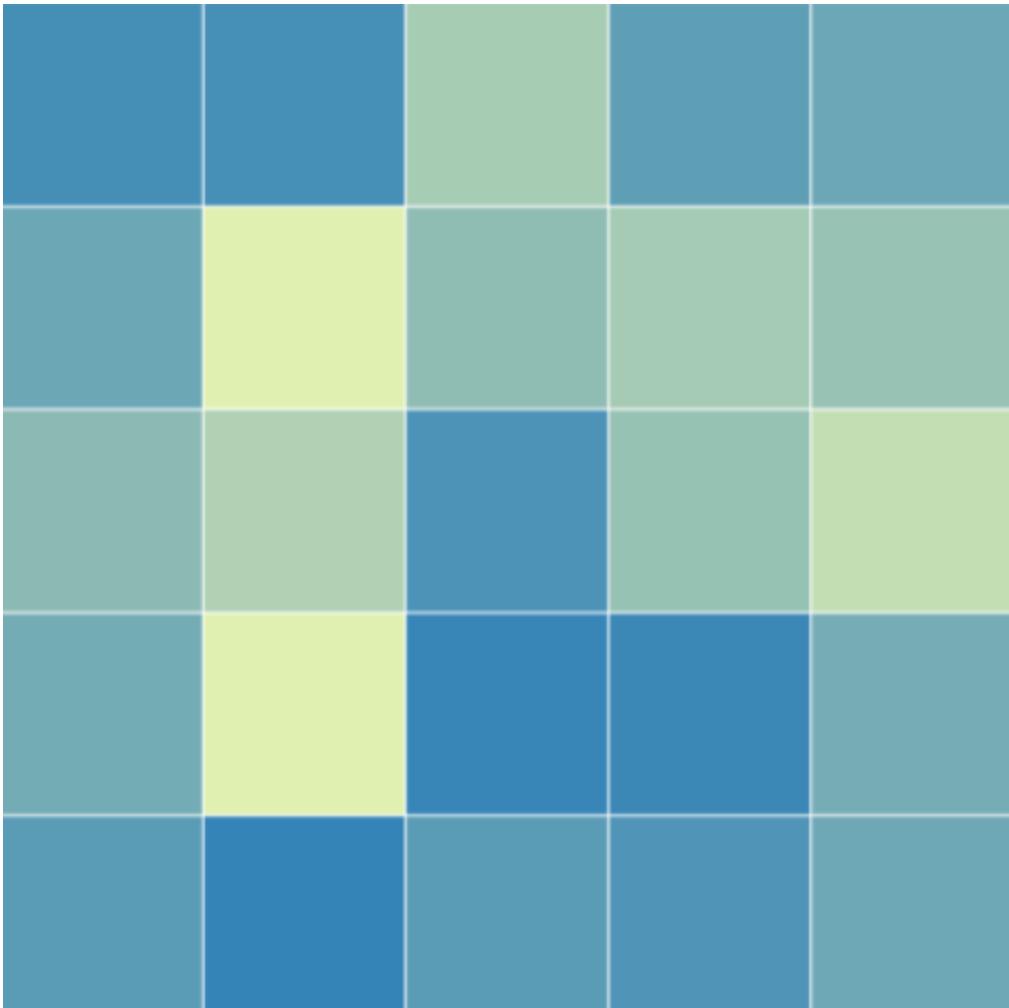
*transparency*

# Semiotics of uncertainty



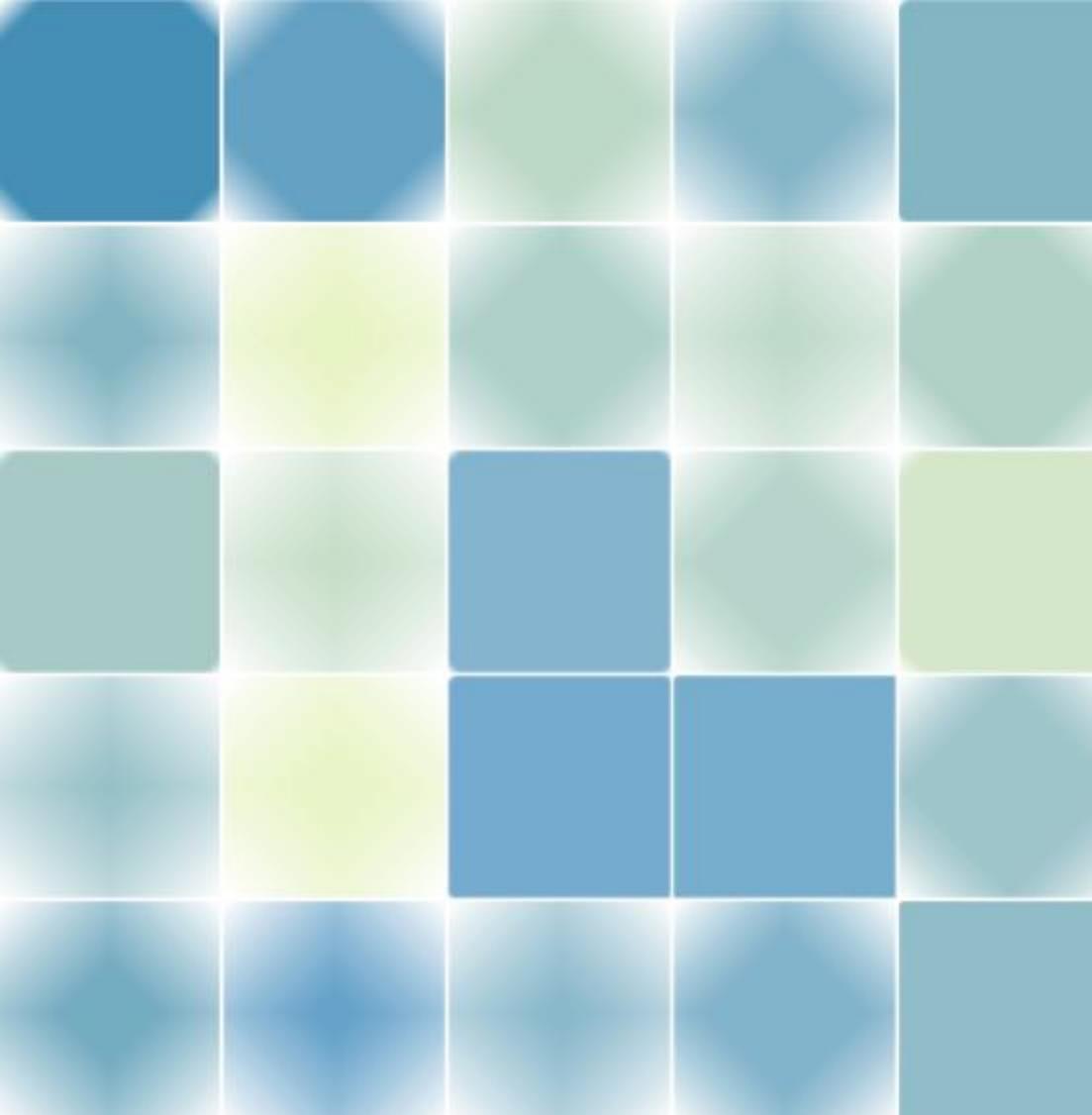
# Semiotics of uncertainty

- Fuzziness juxtaposition



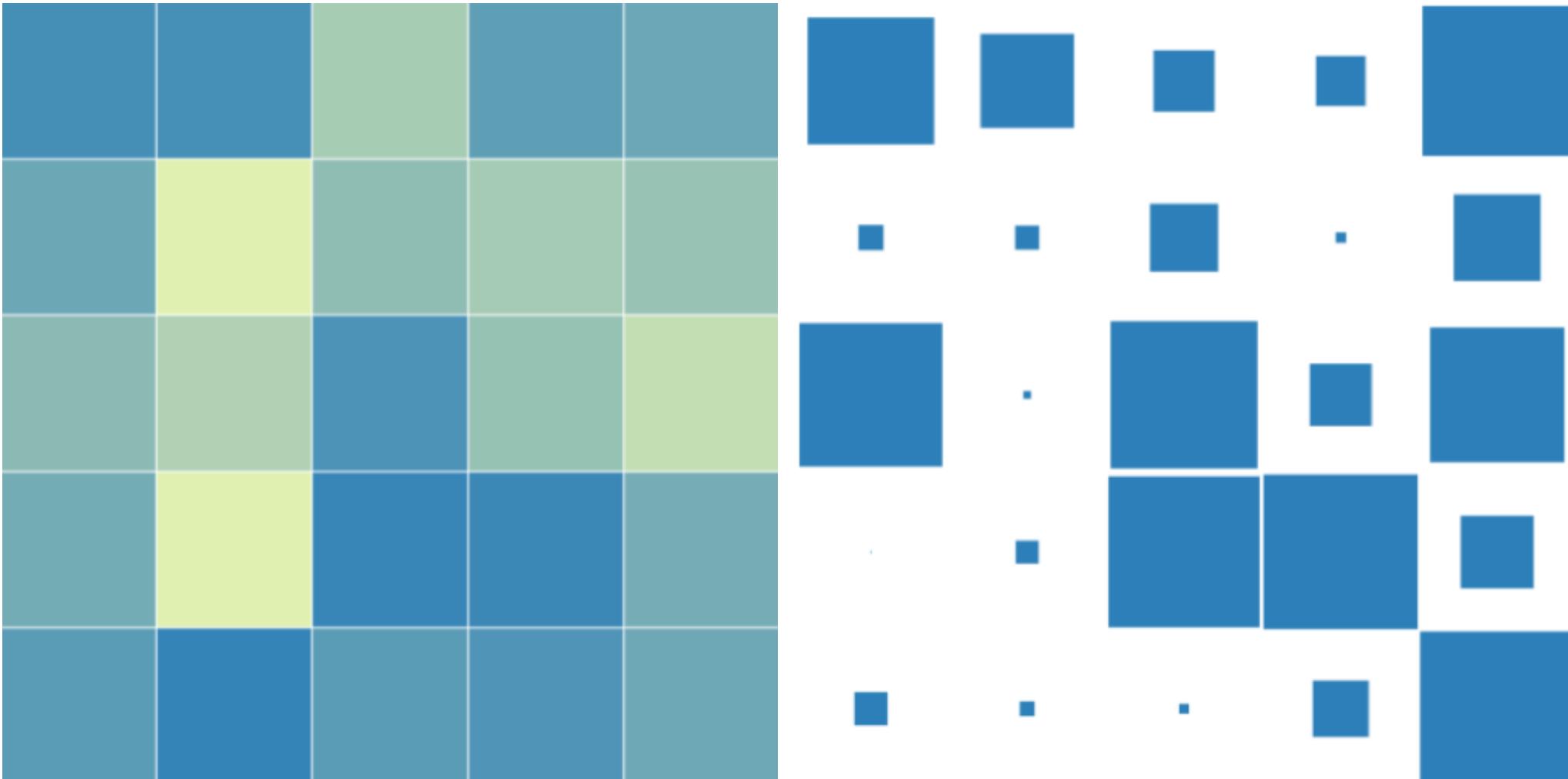
# Semiotics of uncertainty

- Fuzziness superposition



# Semiotics of uncertainty

- Size juxtaposition



# Semiotics of uncertainty

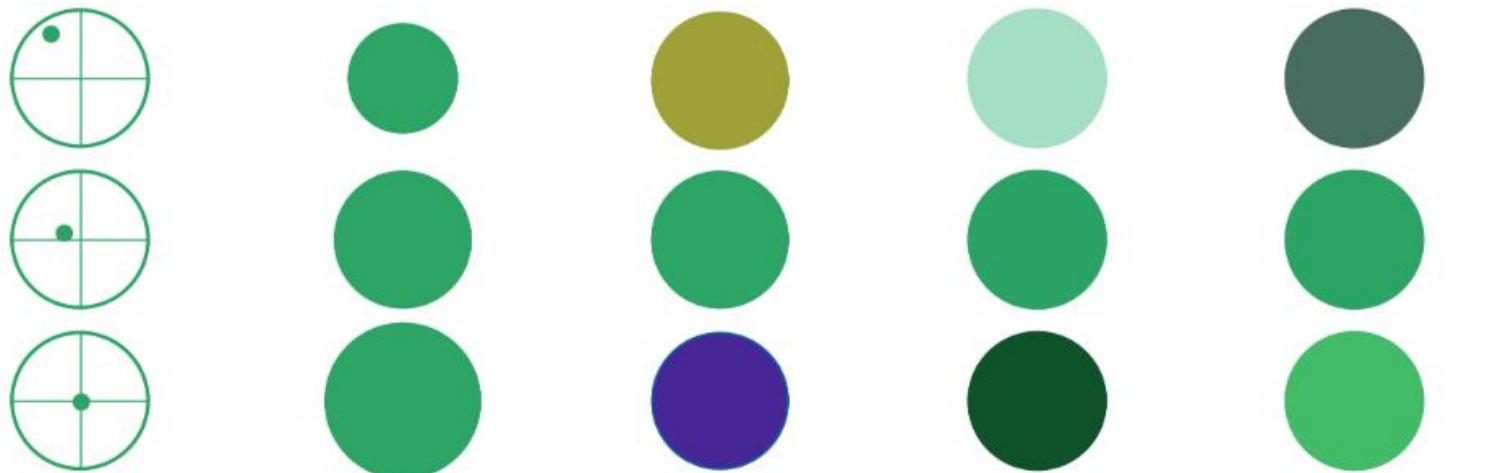
- Size superposition



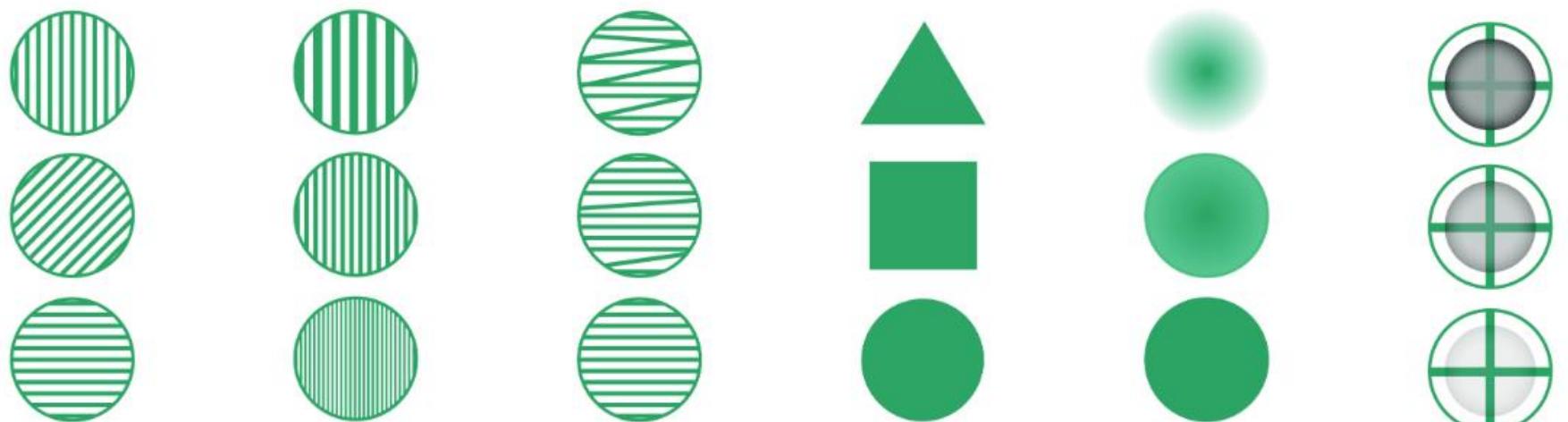
# Semiotics of uncertainty

- Problems?

- Interference between channels
- Difficult to interpret (multiple comparisons)



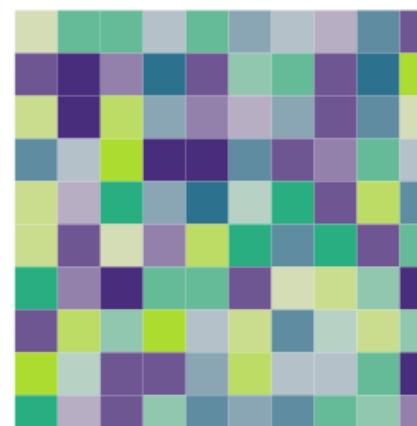
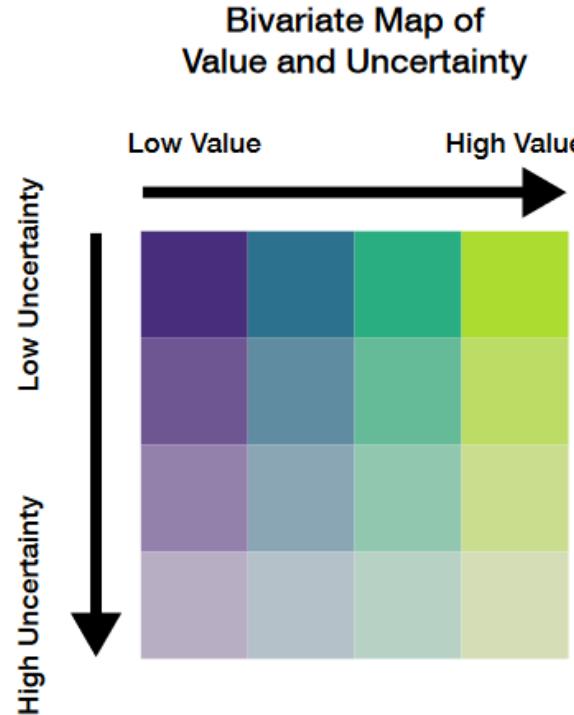
*location*      *size*      *color hue*      *color value*      *color saturation*



*orientation*      *grain*      *arrangement*      *shape*      *fuzziness*      *transparency*

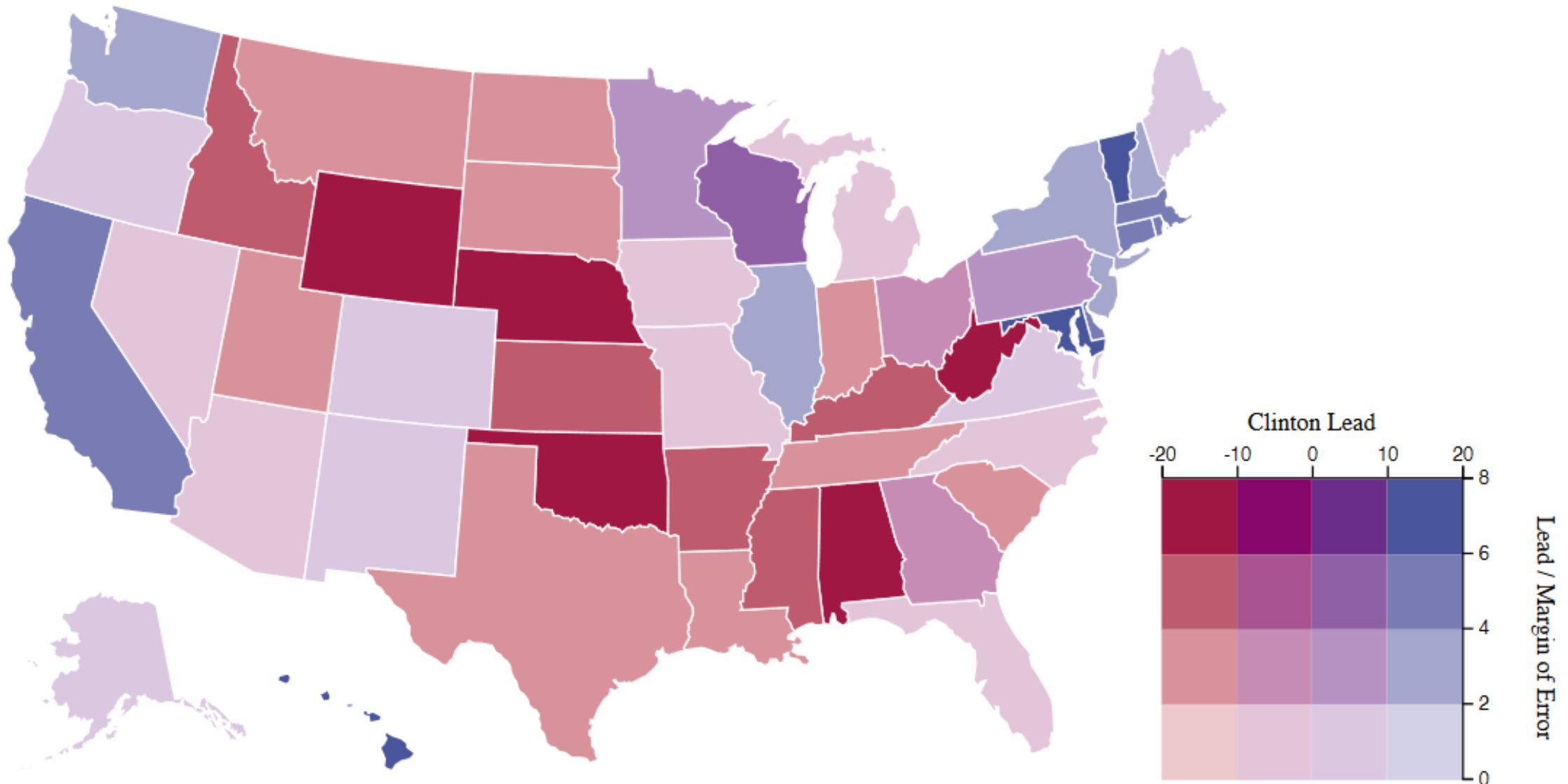
# Value-suppressing uncertainty palettes

- Bivariate maps have issues
- Value-suppressing uncertainty palettes (VSUP)
  - As uncertainty increases, the number of color options decreases
- Correll, Michael, Dominik Moritz, and Jeffrey Heer. "Value-suppressing uncertainty palettes." *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018.



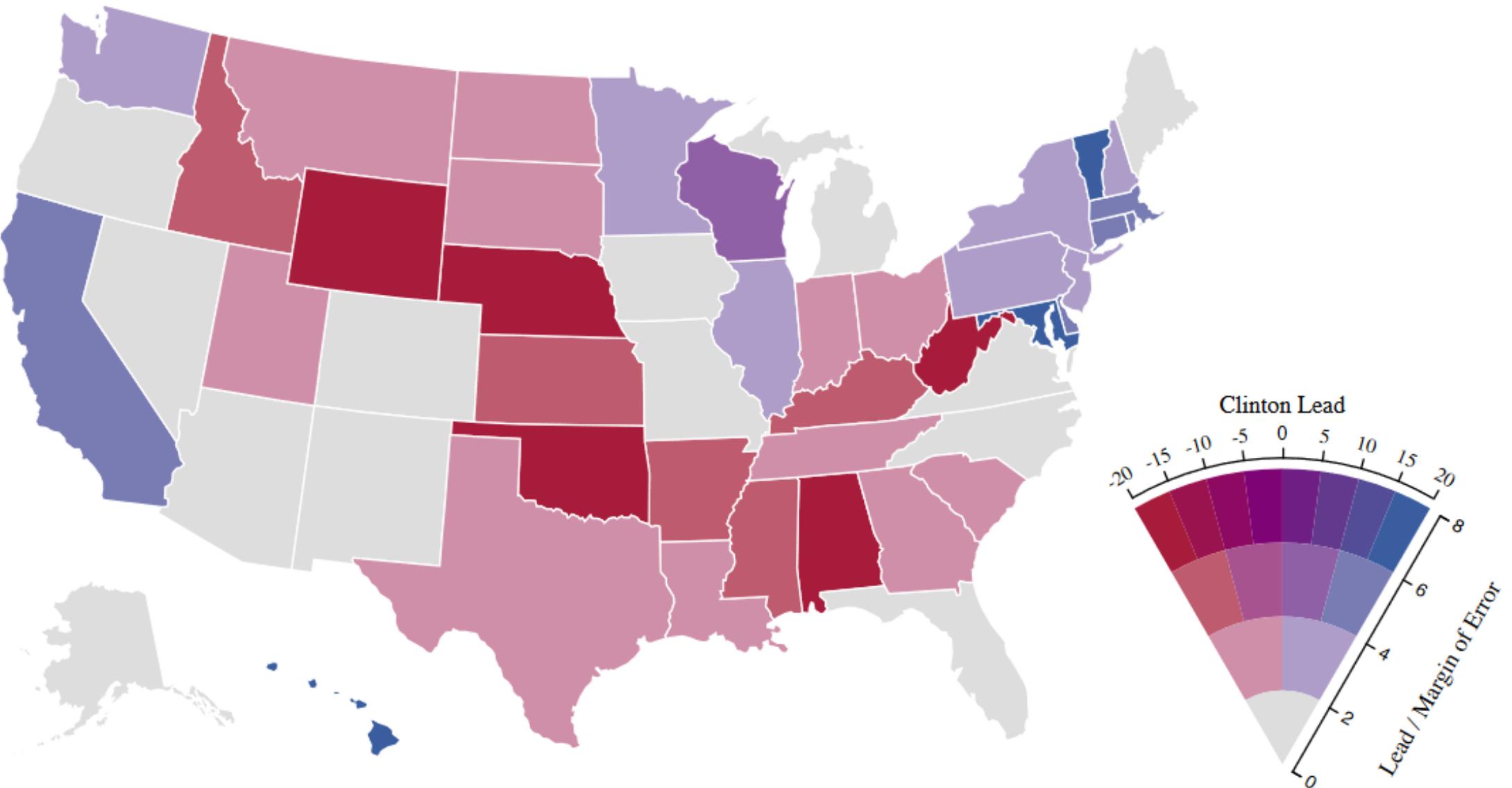
# Value-suppressing uncertainty palettes

- 2016 US election



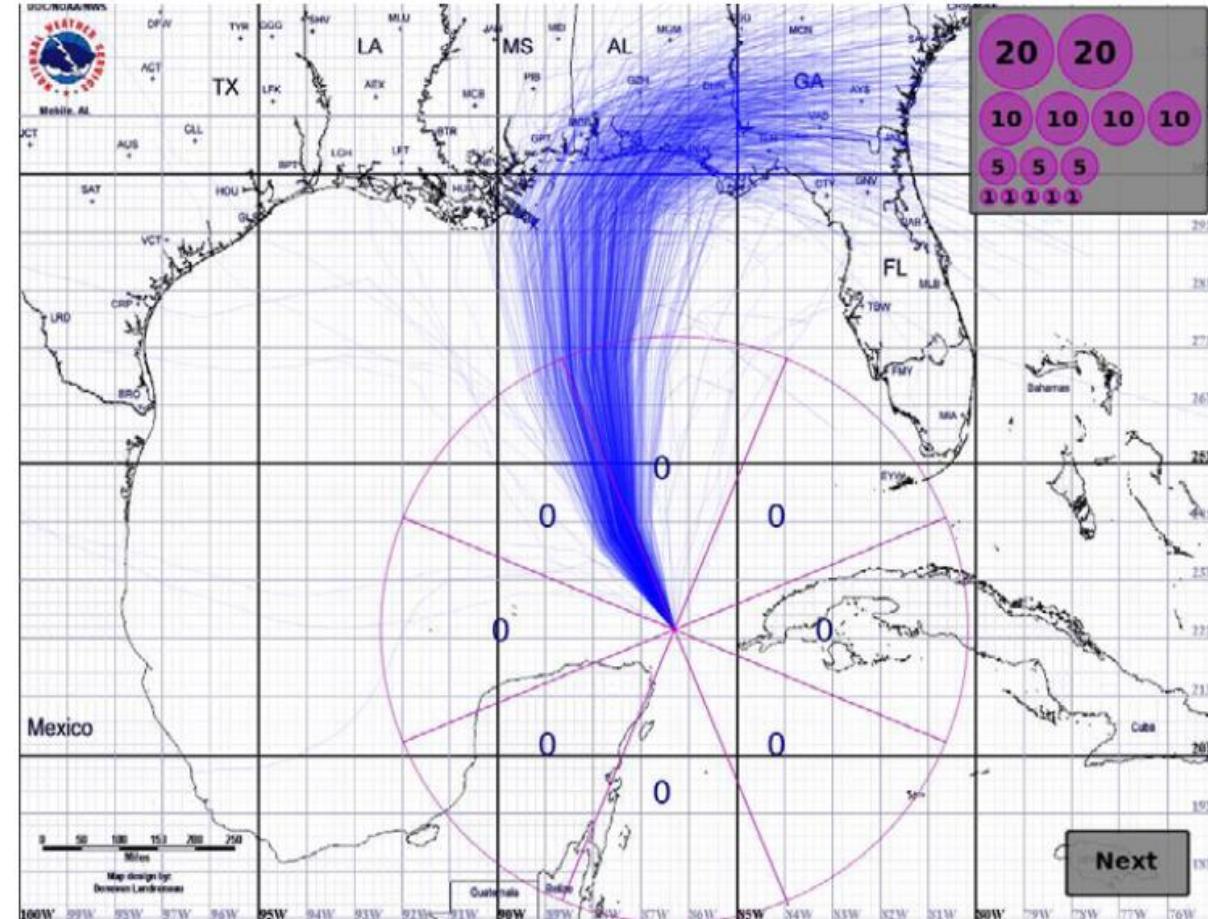
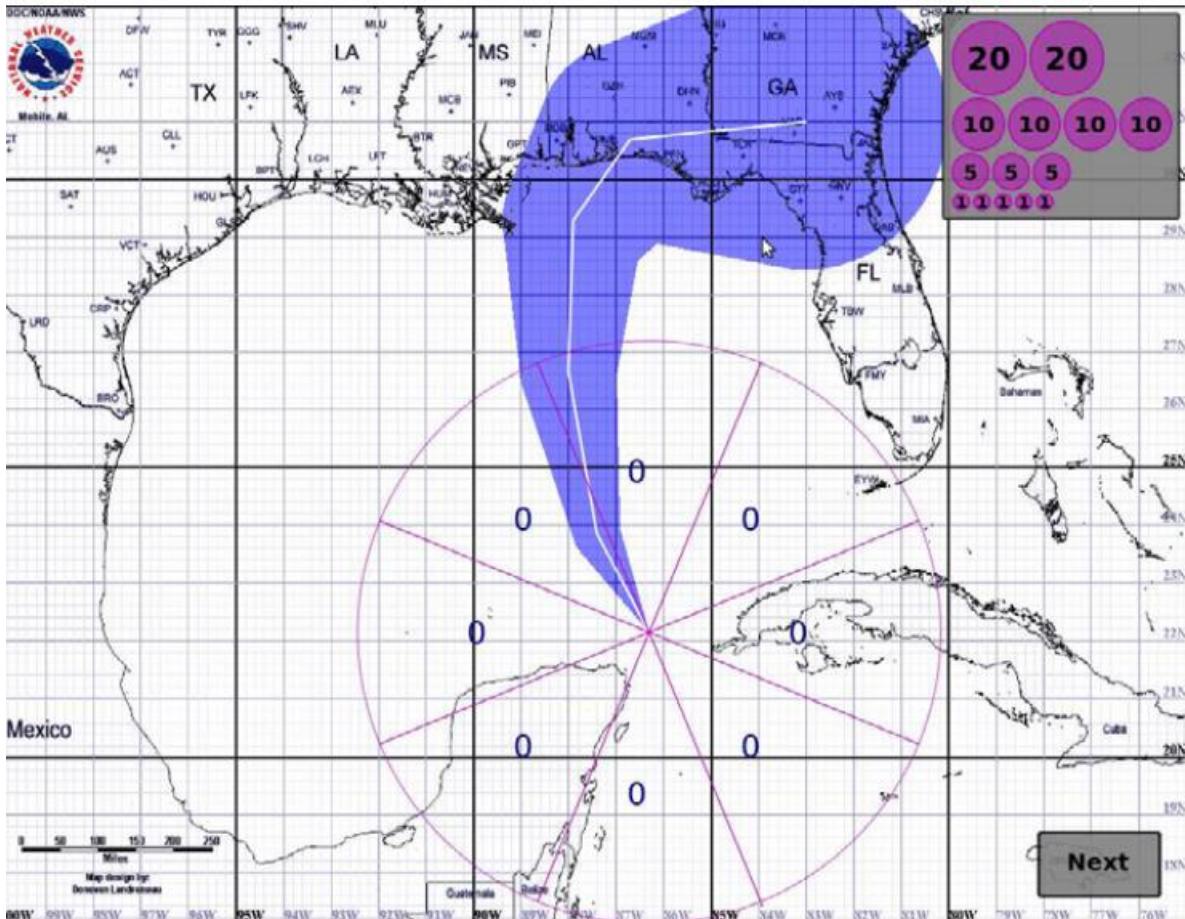
# Value-suppressing uncertainty palettes

- 2016 US election



# Frequency framing and hypothetical outcomes

- Size of a hurricane? Or likelihood? Is New Orleans safe?
- Make uncertainty more concrete via hypothetical outcomes

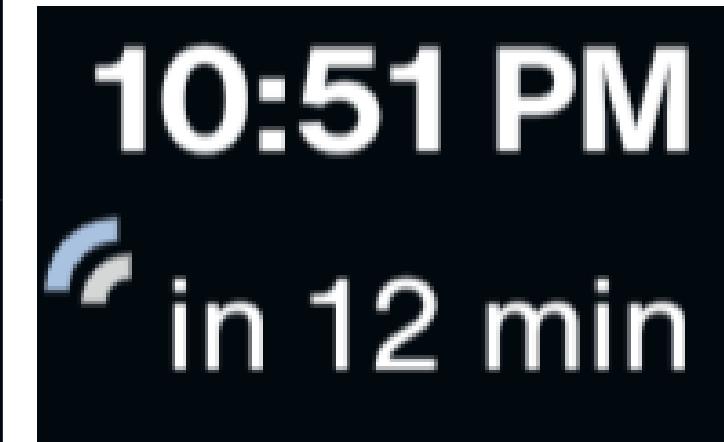
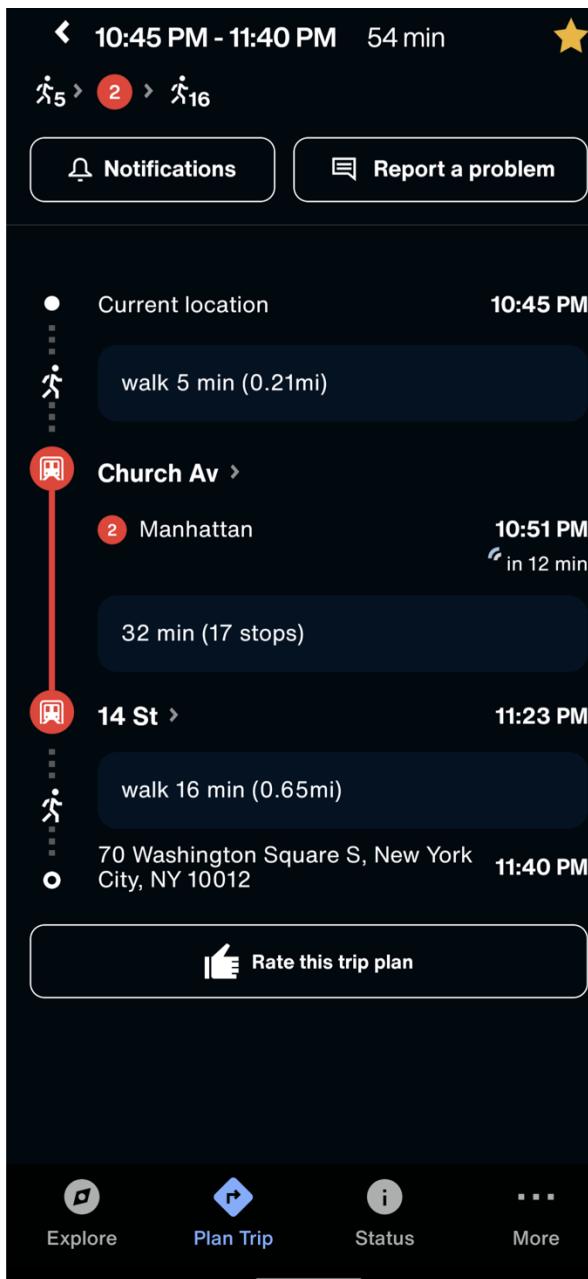




Break  
back at 12:32 pm

# When does my bus arrive?

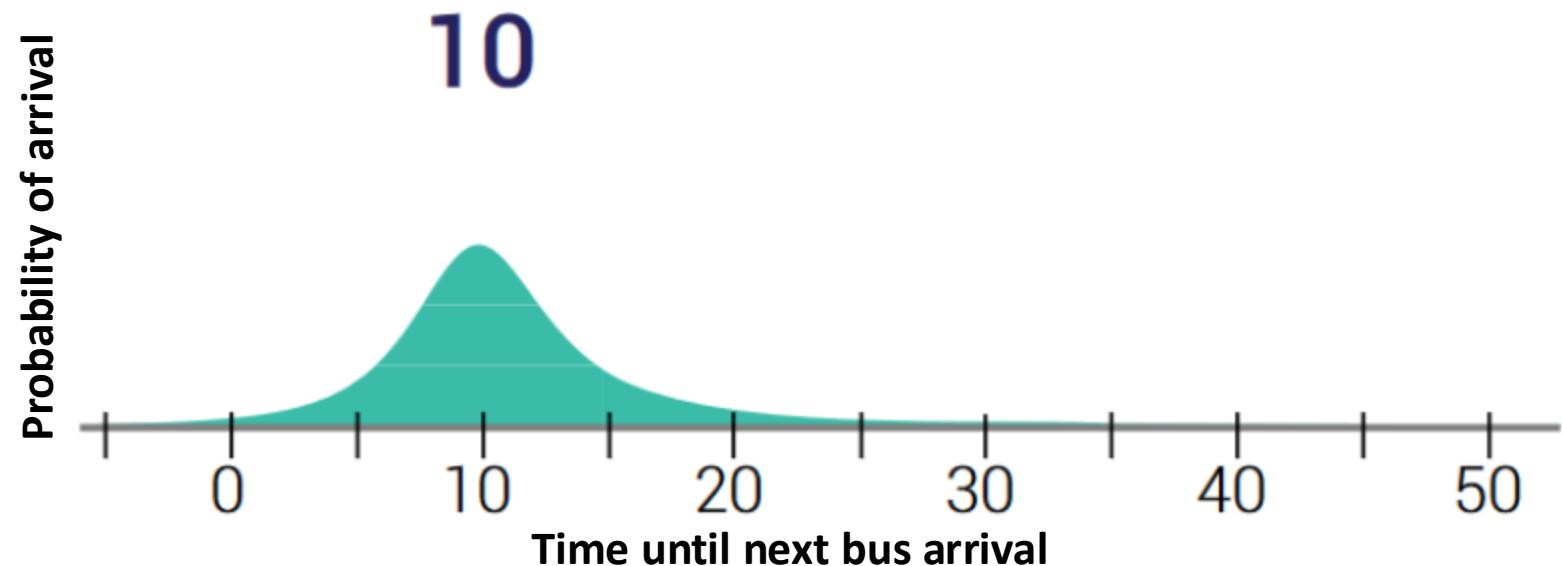
- How can we show the likelihood of bus arrival times in an effective way?



Delays?

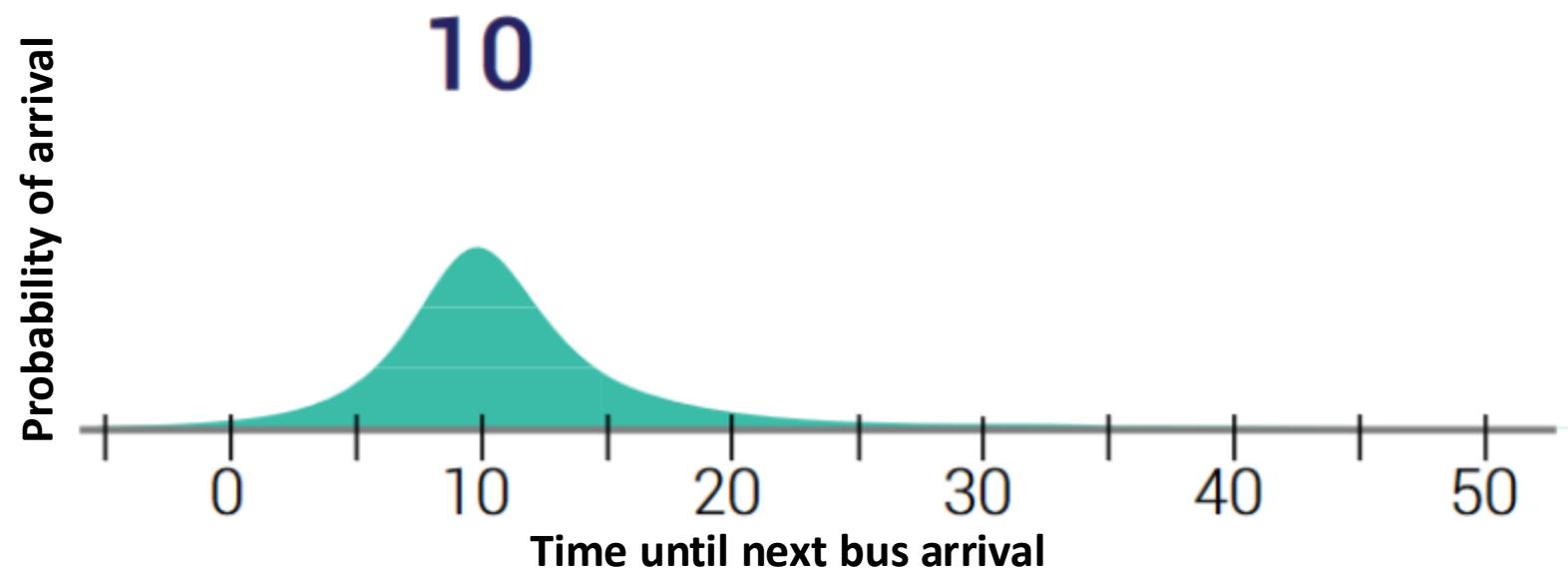
# When does my bus arrive?

- Things to convey:
  - When to leave for the bus stop?
  - How long will I have to wait?
  - If I miss the bus, when does the next one arrive?
  - Will I get to my destination in spite of bus delays?
  - Can I do \_\_\_ before the bus arrives?



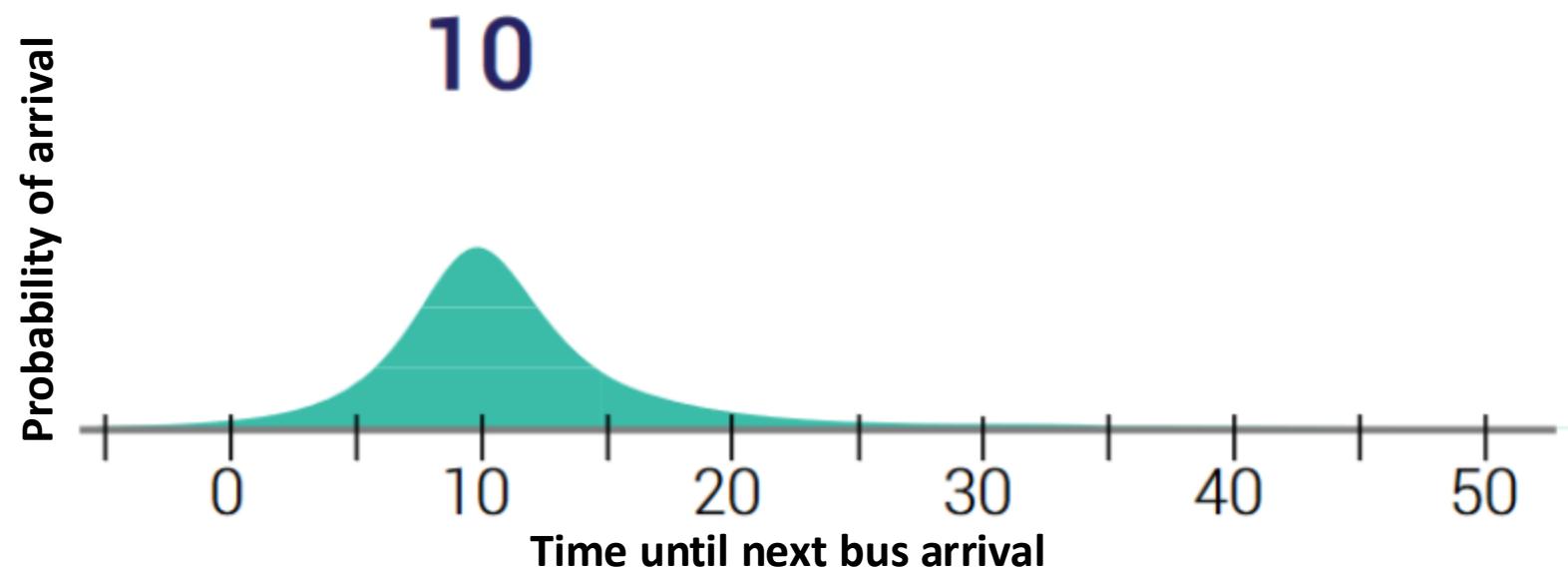
# When does my bus arrive?

- Want to support:
  - Point estimate of arrival time
  - Probabilistic estimate of arrival time
  - Probabilistic estimate of arrival status
- Solutions:
  - Different layouts for different use cases
  - Point estimates and probabilistic estimates should coincide spatially



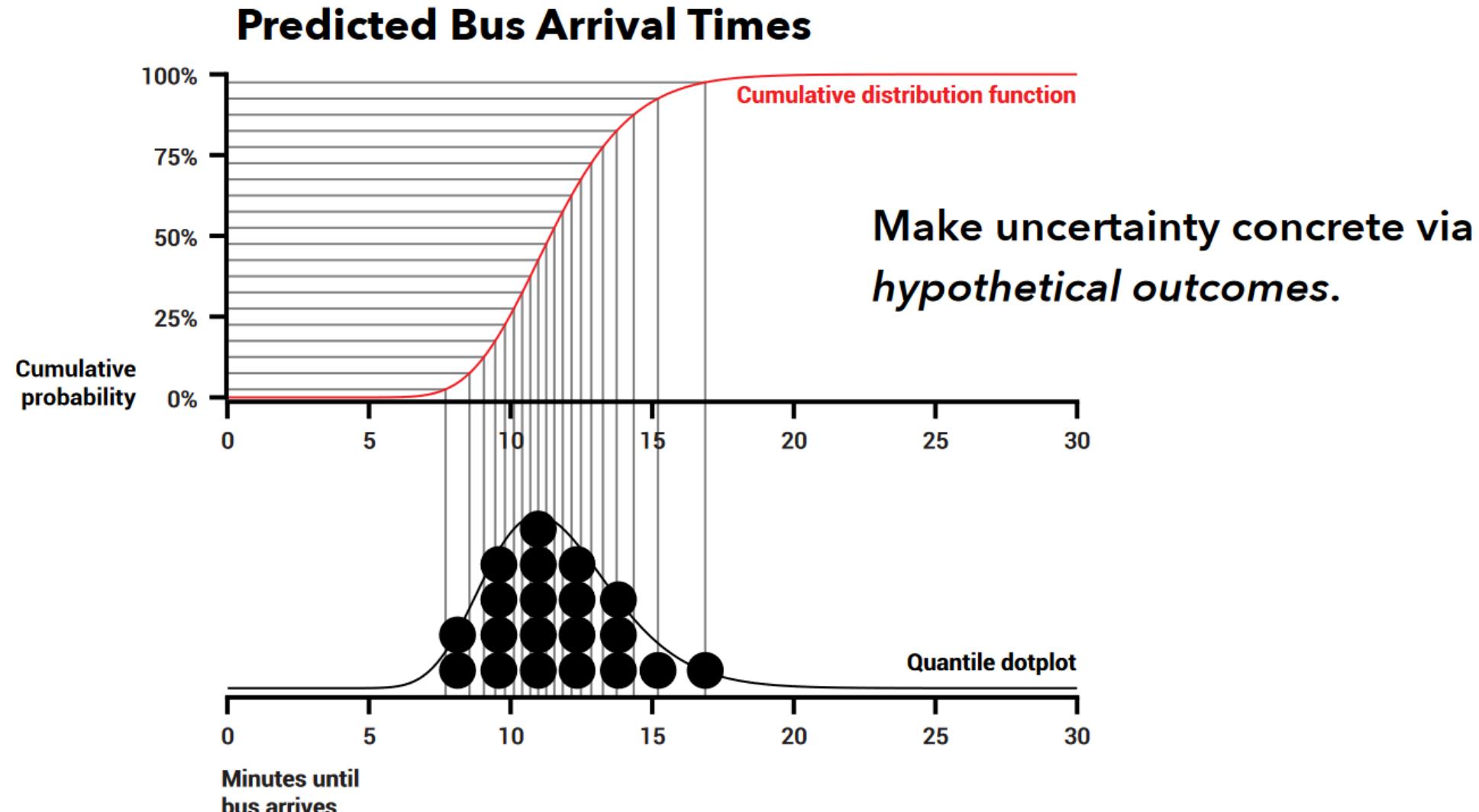
# When does my bus arrive?

- Want to support:
  - Point estimate of arrival time
  - Probabilistic estimate of arrival time
  - Probabilistic estimate of arrival status
- Solutions:
  - Different layouts for different use cases
  - **Point estimates and probabilistic estimates should coincide spatially**



# When does my bus arrive?

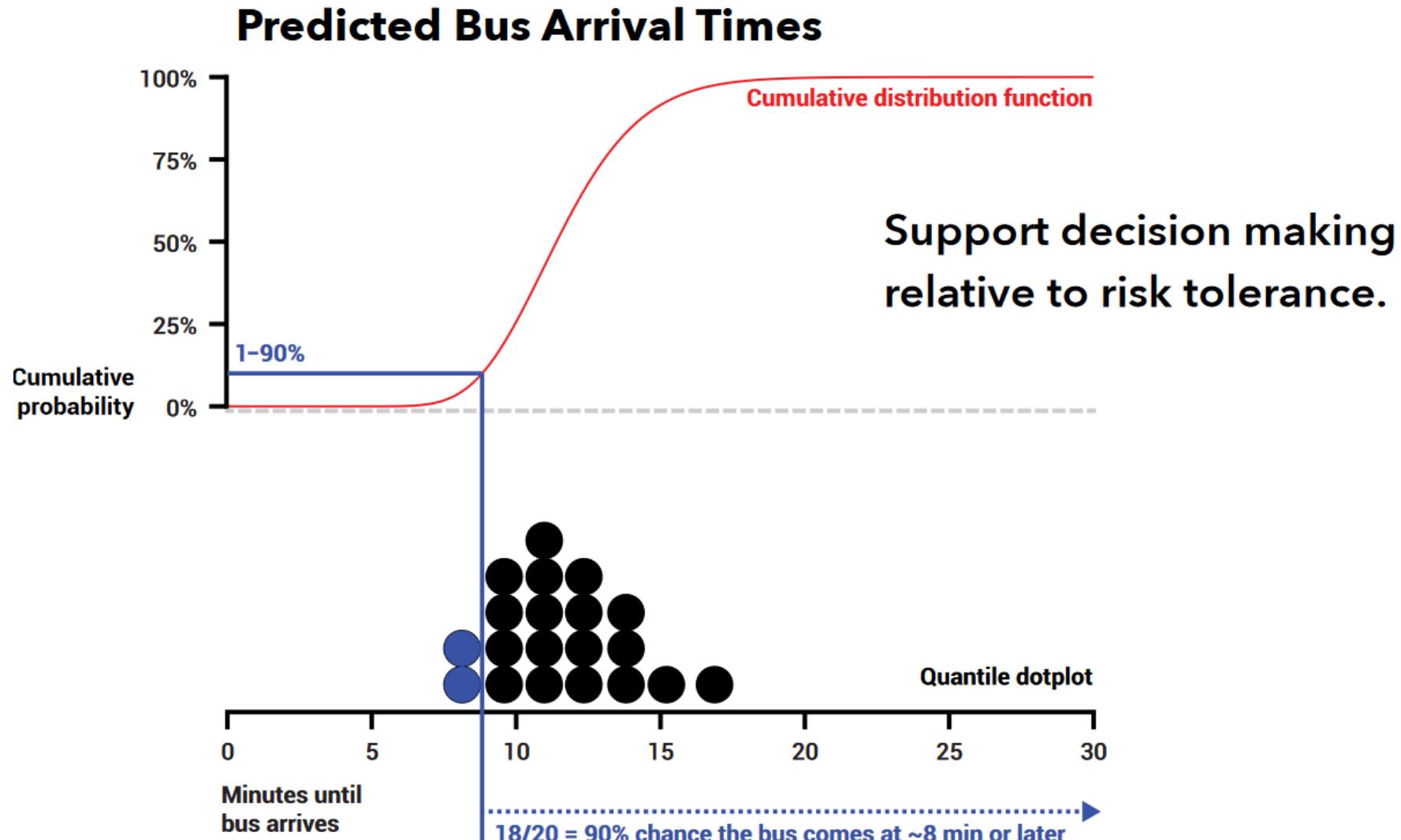
*Point estimates and probabilistic estimates should coincide spatially*



Kay, Matthew, et al. "When (ish) is my bus? user-centered visualizations of uncertainty in everyday, mobile predictive systems." *Proceedings of the 2016 chi conference on human factors in computing systems*. 2016.

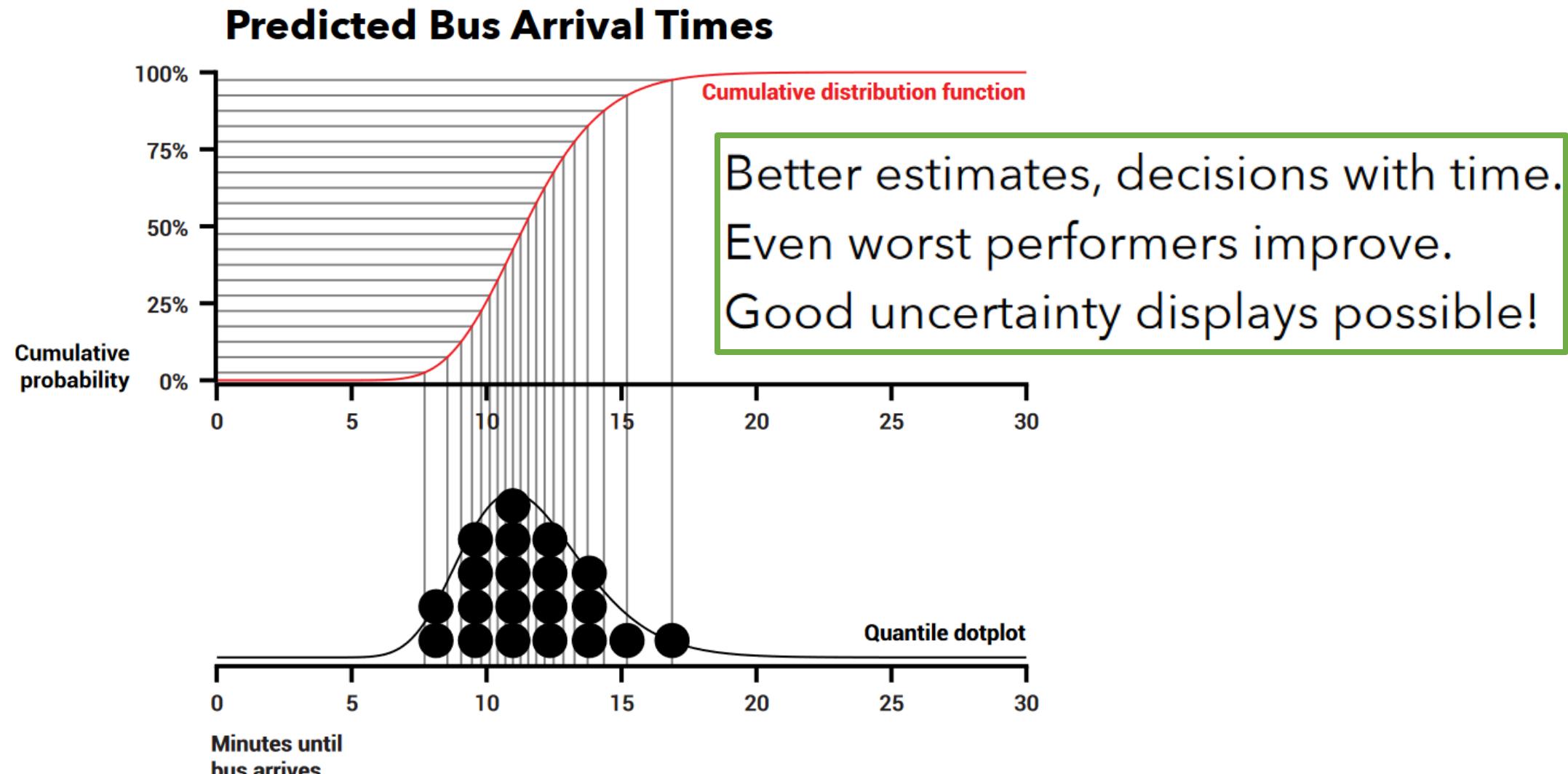
# When does my bus arrive?

*Point estimates and probabilistic estimates should coincide spatially*



# When does my bus arrive?

*Point estimates and probabilistic estimates should coincide spatially*



Kay, Matthew, et al. "When (ish) is my bus? user-centered visualizations of uncertainty in everyday, mobile predictive systems." *Proceedings of the 2016 chi conference on human factors in computing systems*. 2016.

# Takeaway: how to present probabilities

Less Intuitive

**Probability**

$P(A) = 0.6$



**Percentage**

60% chance of A

More Intuitive

**Natural  
Frequency**

3 out of 5 times, A  
happens.

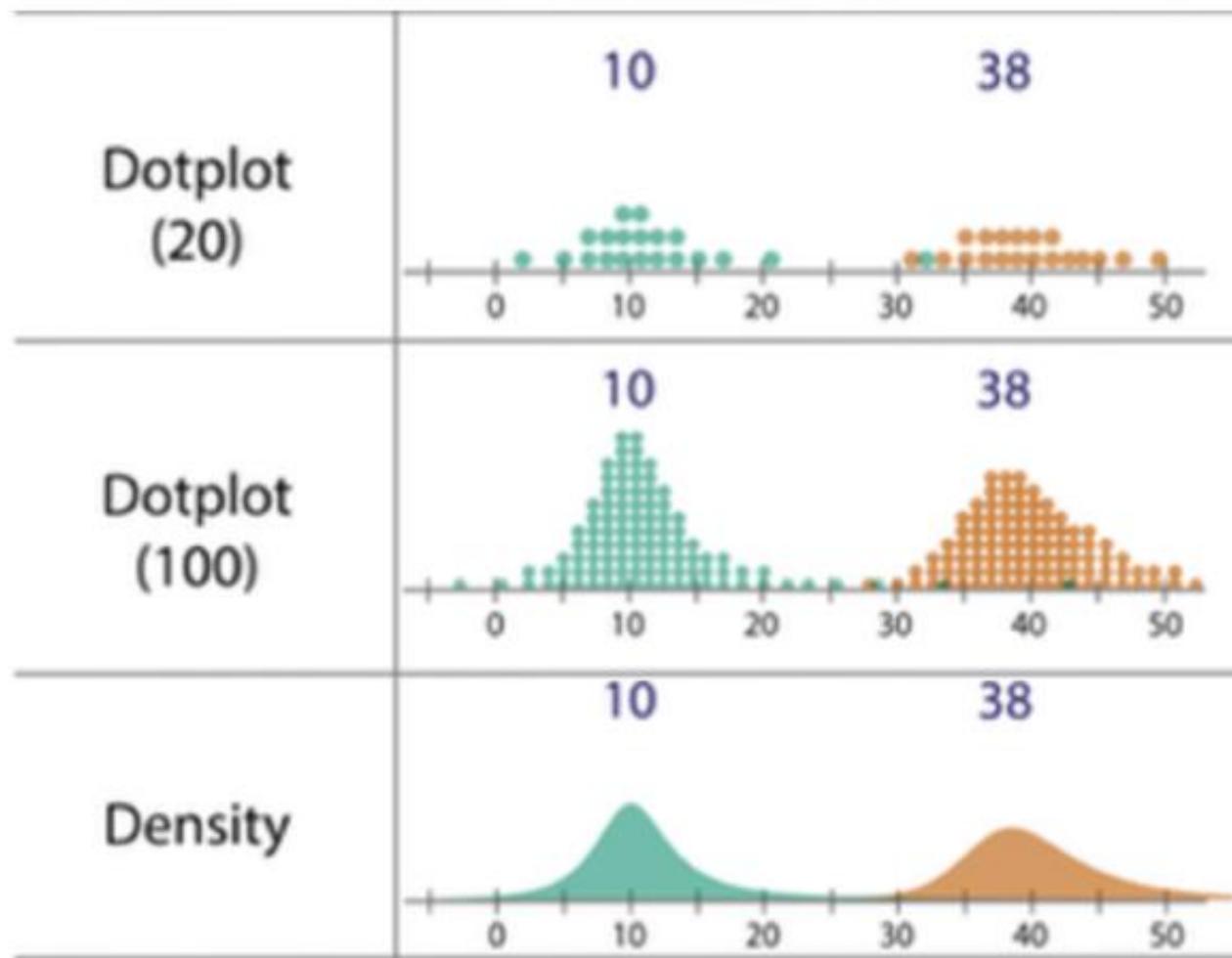
# Quantile dot plots

- Why?

Less Error

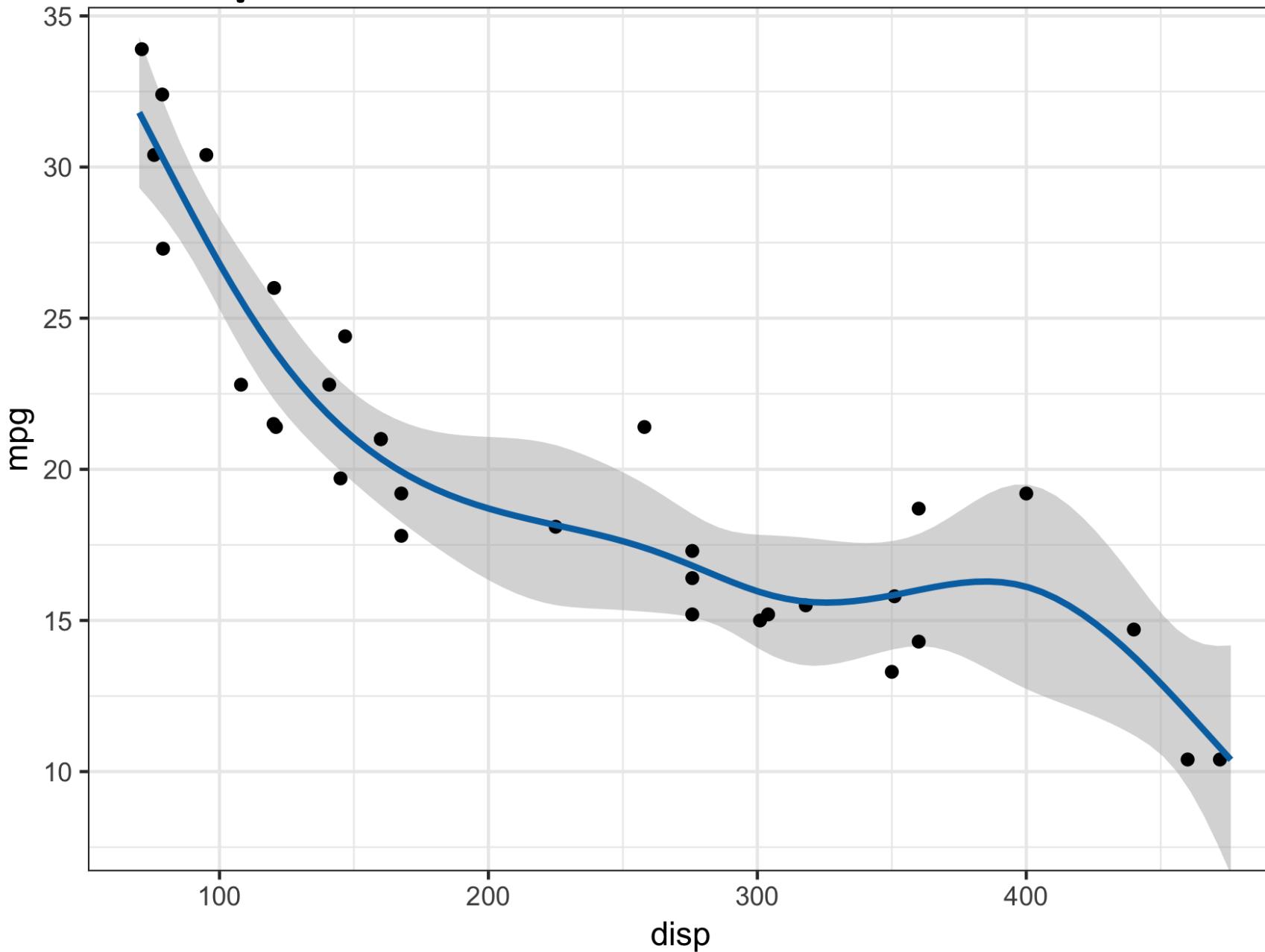


More Error



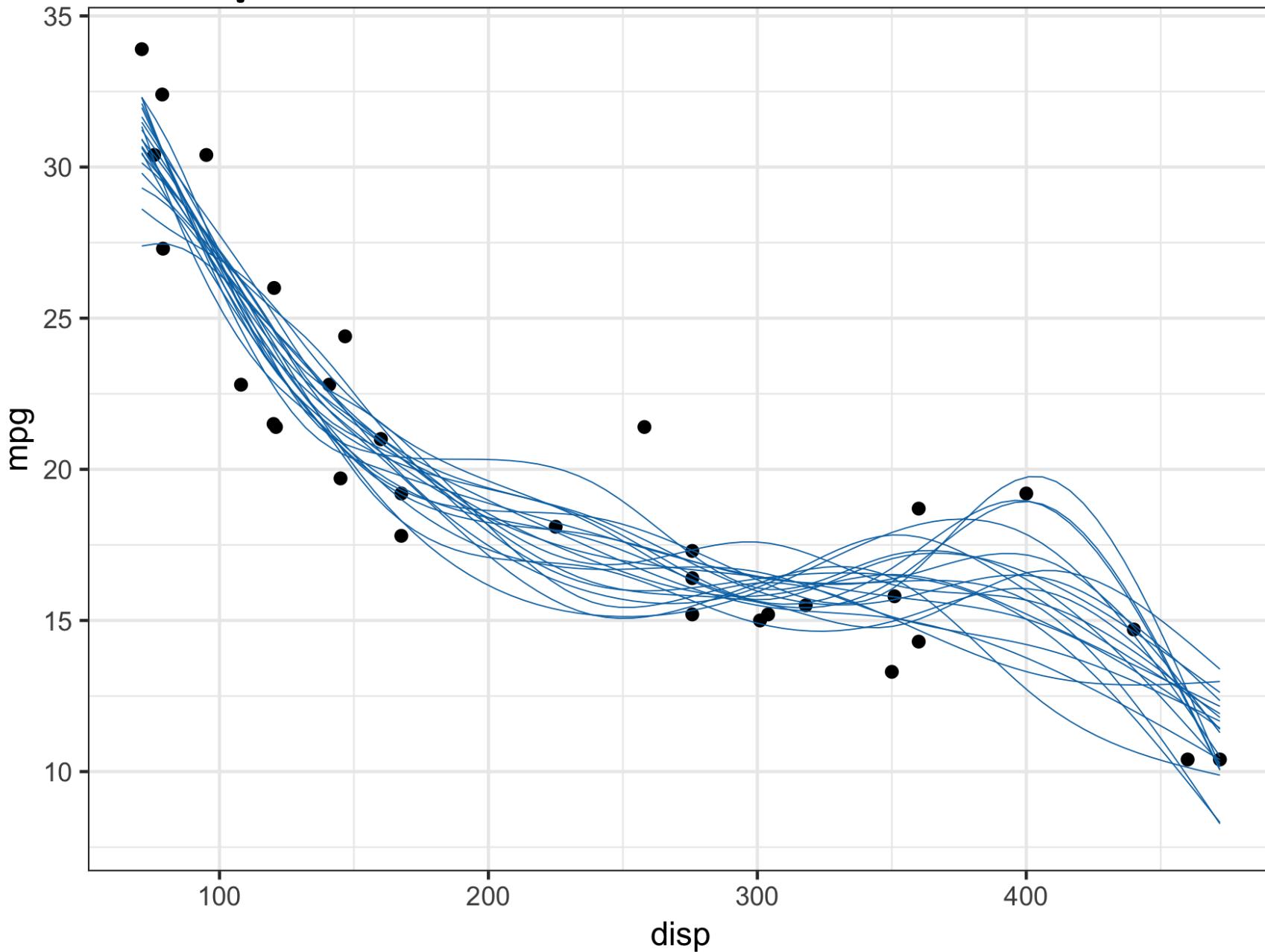
# Hypothetical outcome plots

- Hullman, Jessica, Paul Resnick, and Eytan Adar. "Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering." PloS one 10.11 (2015): e0142444.



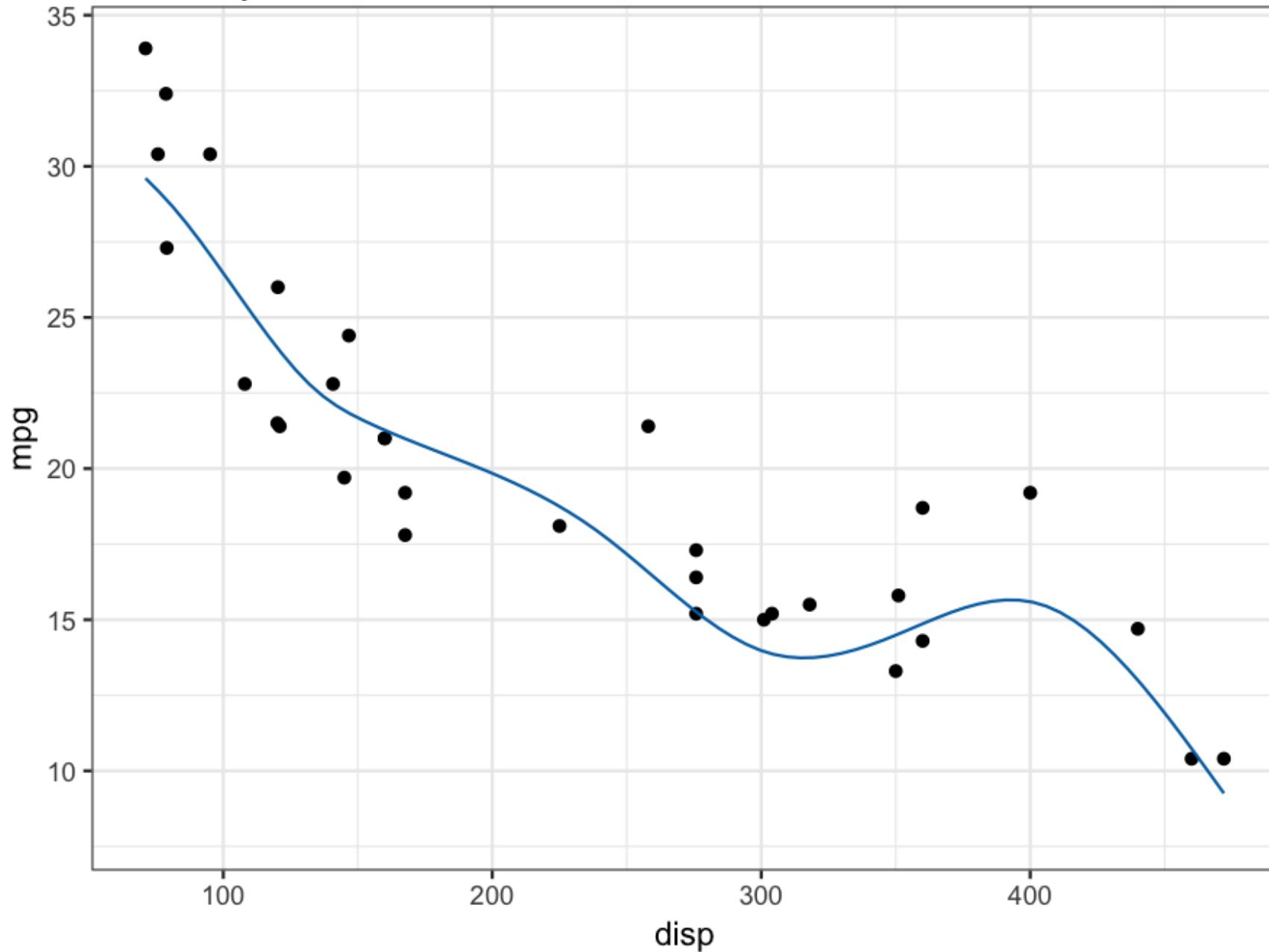
# Hypothetical outcome plots

- Hullman, Jessica, Paul Resnick, and Eytan Adar. "Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering." PloS one 10.11 (2015): e0142444.



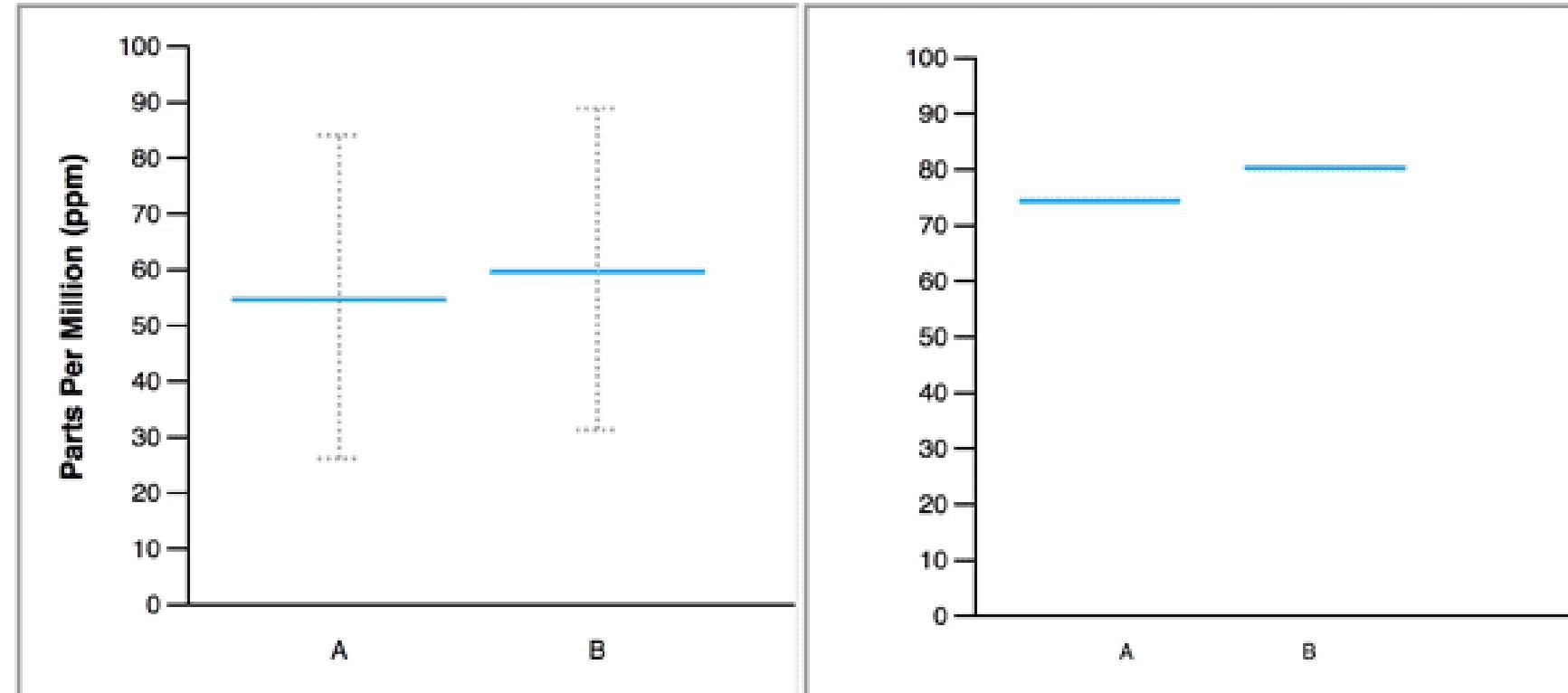
# Hypothetical outcome plots

- Hullman, Jessica, Paul Resnick, and Eytan Adar. "Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering." PloS one 10.11 (2015): e0142444.



# Hypothetical outcome plots

- Hullman, Jessica, Paul Resnick, and Eytan Adar. "Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering." PloS one 10.11 (2015): e0142444.

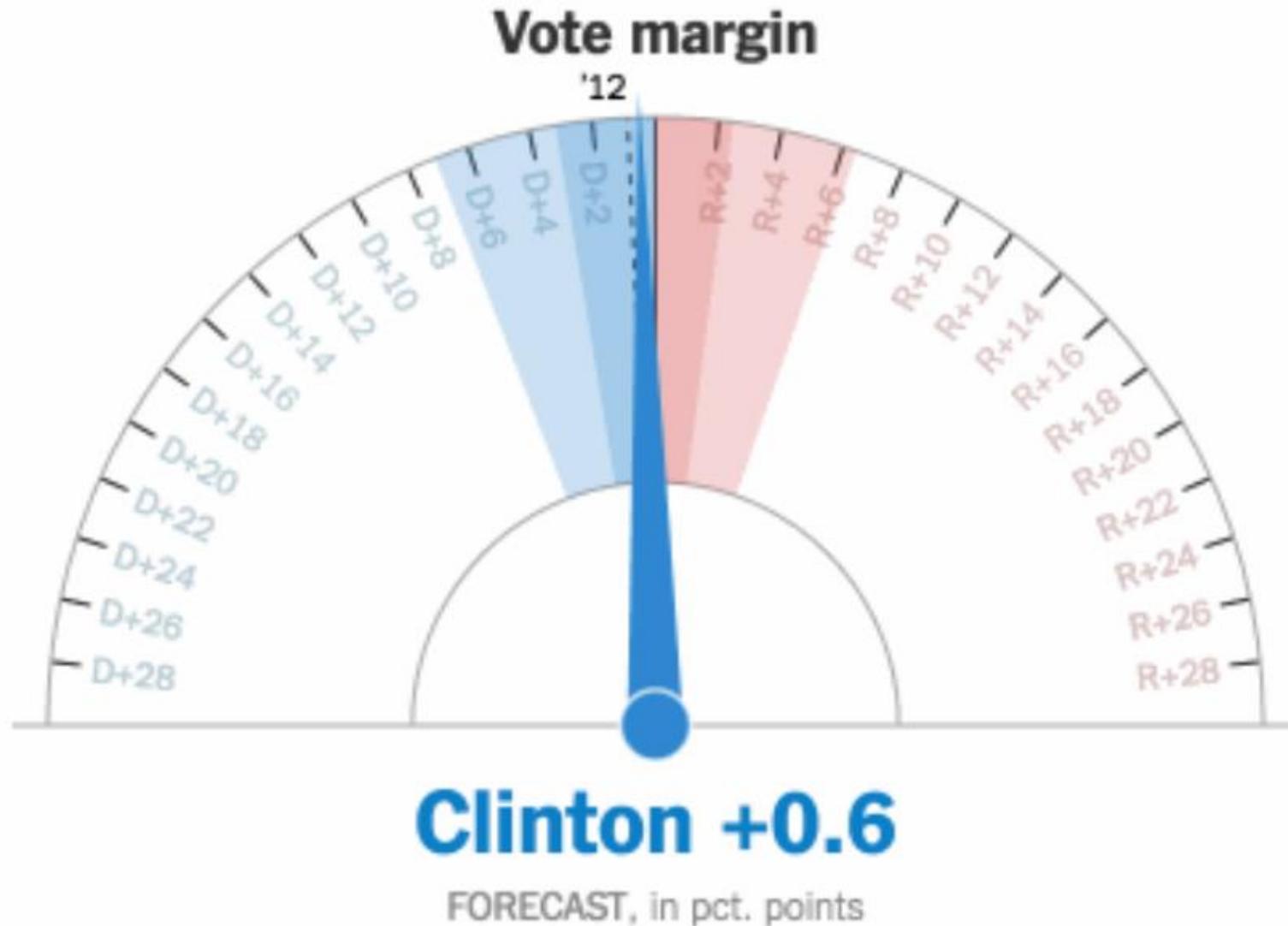


# Hypothetical outcome plots

- Pros:
  - Very intuitive and easy to understand
    - Probability is hard to understand
  - Don't require a new visual encoding variable (e.g. hue, blur, transparency, size, etc.)
- Drawbacks
  - Sampling error: only see a limited number of draws from the distribution
  - Memory: viewer has to combine information across multiple frames
  - Only works for computer displays (not static)
- <https://www.nytimes.com/2014/05/02/upshot/how-not-to-be-misled-by-the-jobs-report.html>

# New York Times needle

- <https://www.youtube.com/watch?v=HuE8e5Nnpok>



# Summary: how do visualize uncertainty?

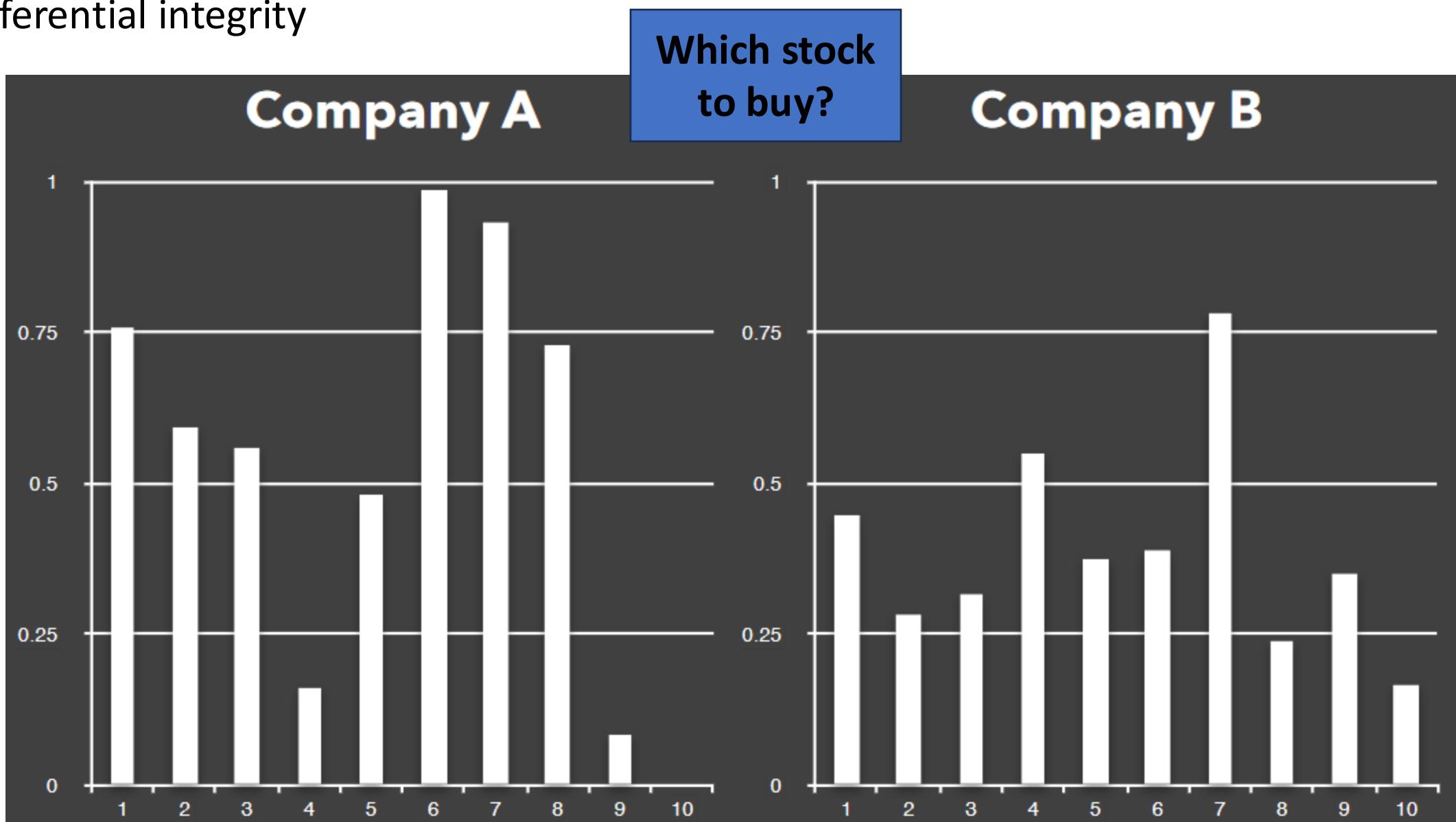
- Choose an appropriate visual variable based on the domain, literacy, and expertise of your audience.
- Be mindful that any display of uncertainty inherently increases the complexity of your visualization, and that there is a preference/performance gap.
- Consider medium of your visualization: does your device support dynamic changes?

**IT DEPENDS!**

# What can go wrong?

# What can go wrong?

- Inferential integrity



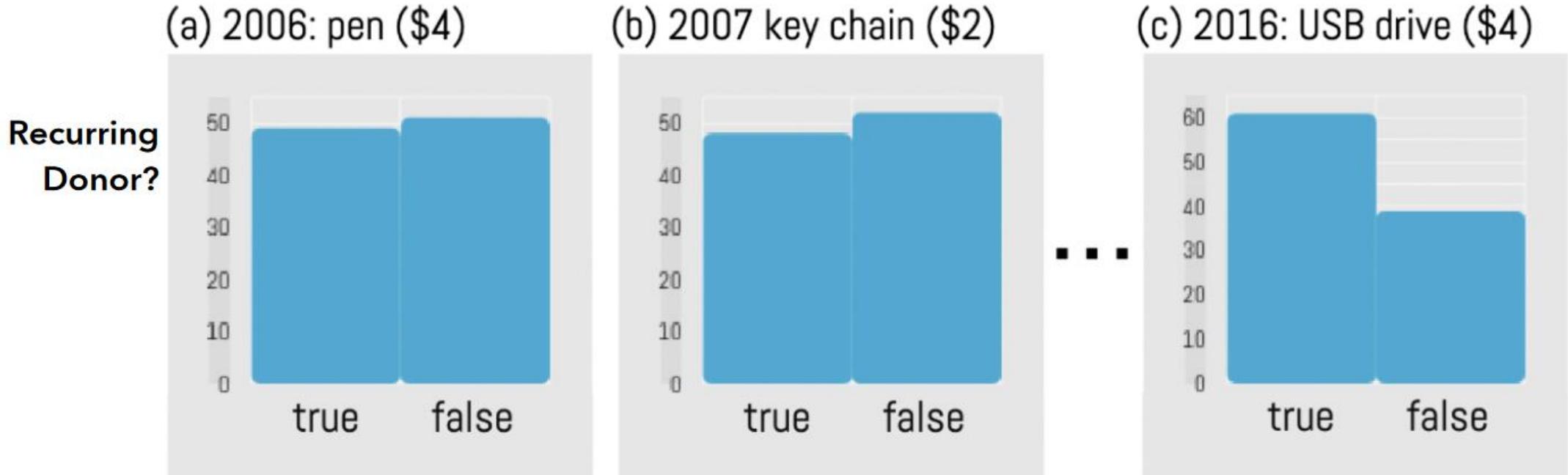
# What can go wrong?

- Inferential integrity



# What can go wrong?

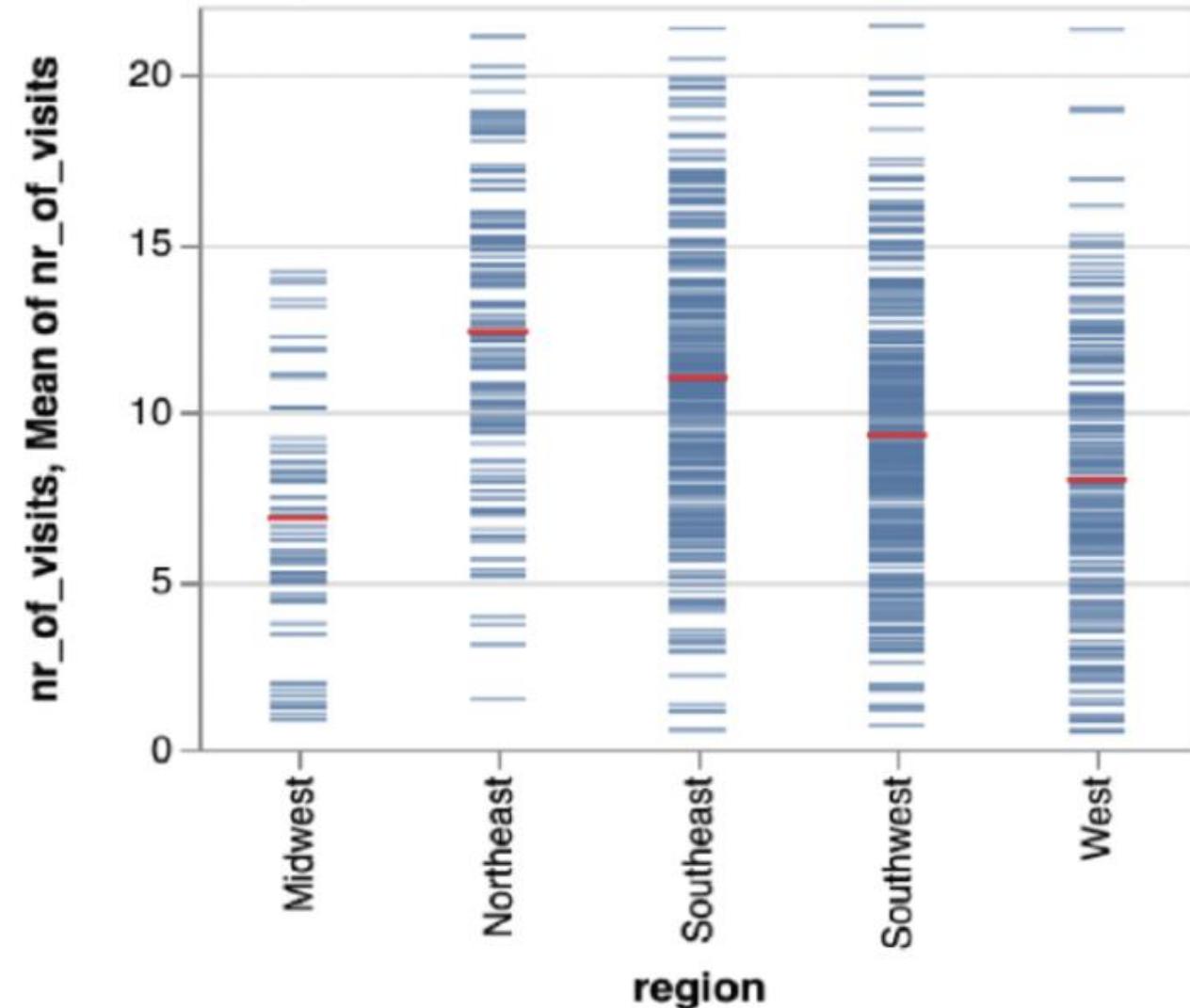
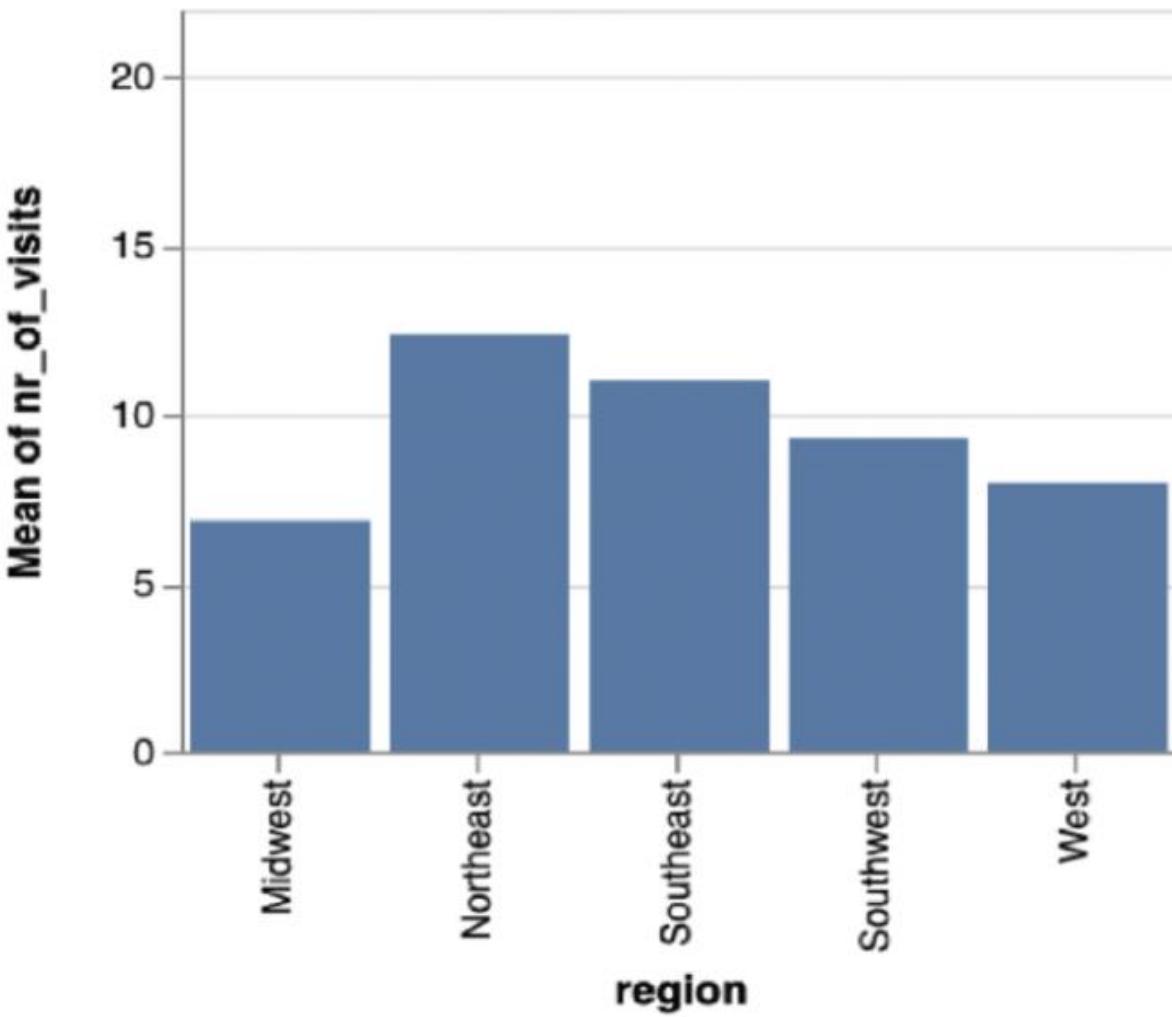
- Multiple comparisons
  - Zgraggen, Emanuel, et al. "Investigating the effect of the multiple comparisons problem in visual analysis." Proceedings of the 2018 chi conference on human factors in computing systems. 2018.



# What can go wrong?

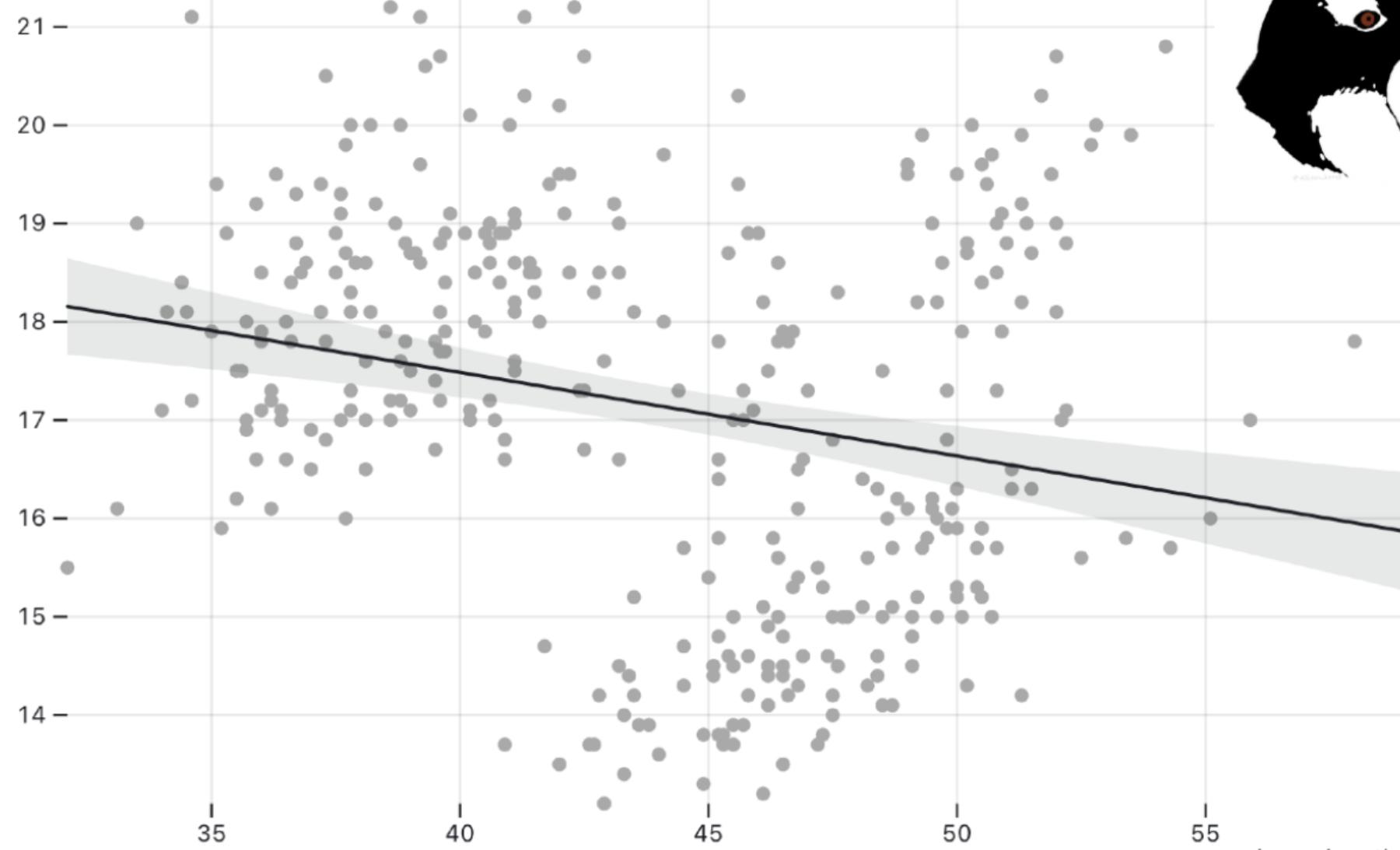
- Aggregated vs. Disaggregated Views

- Nguyen, Francis, et al. "Exploring the effects of aggregation choices on untrained visualization users' generalizations from data." Computer graphics forum. Vol. 39. No. 6. 2020.



# What can go wrong?

↑ culmen\_depth\_mm



CULMEN: RIDGE ALONG THE  
TOP PART OF A BIRD'S BILL

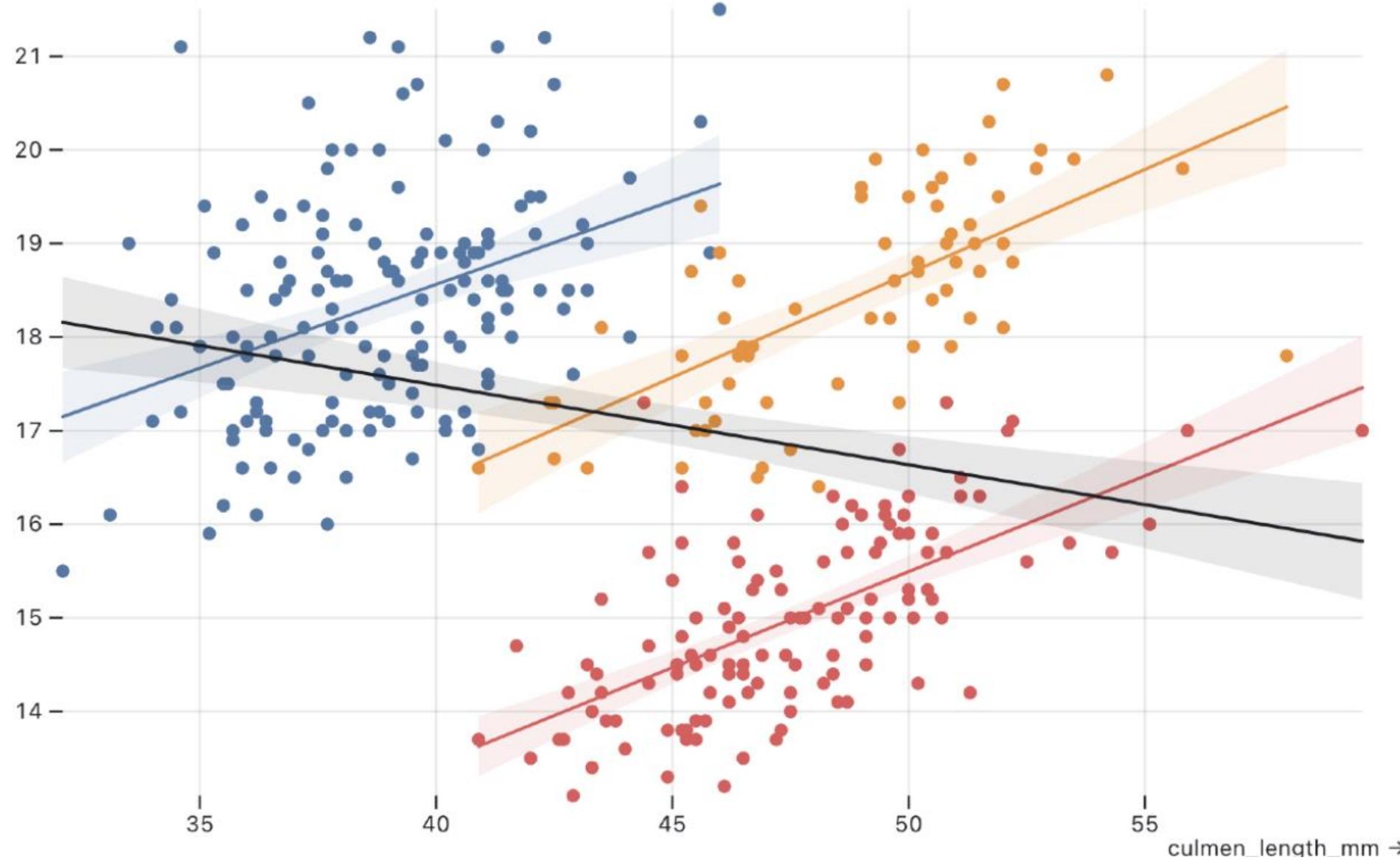


# What can go wrong?

Adelie Chinstrap Gentoo

## Simpson's Paradox!

↑ culmen\_depth\_mm



culmen\_length\_mm →

# What can go wrong?

- Uncertainty can be difficult to understand, and require a statistical background and high numeracy.
- Cognitive and perceptual biases can result in people making poor or error-prone decisions from uncertain data.
- Random fluctuations can create apparent outliers that are actually false flags.

A LOT

# Uncertainty: Summary

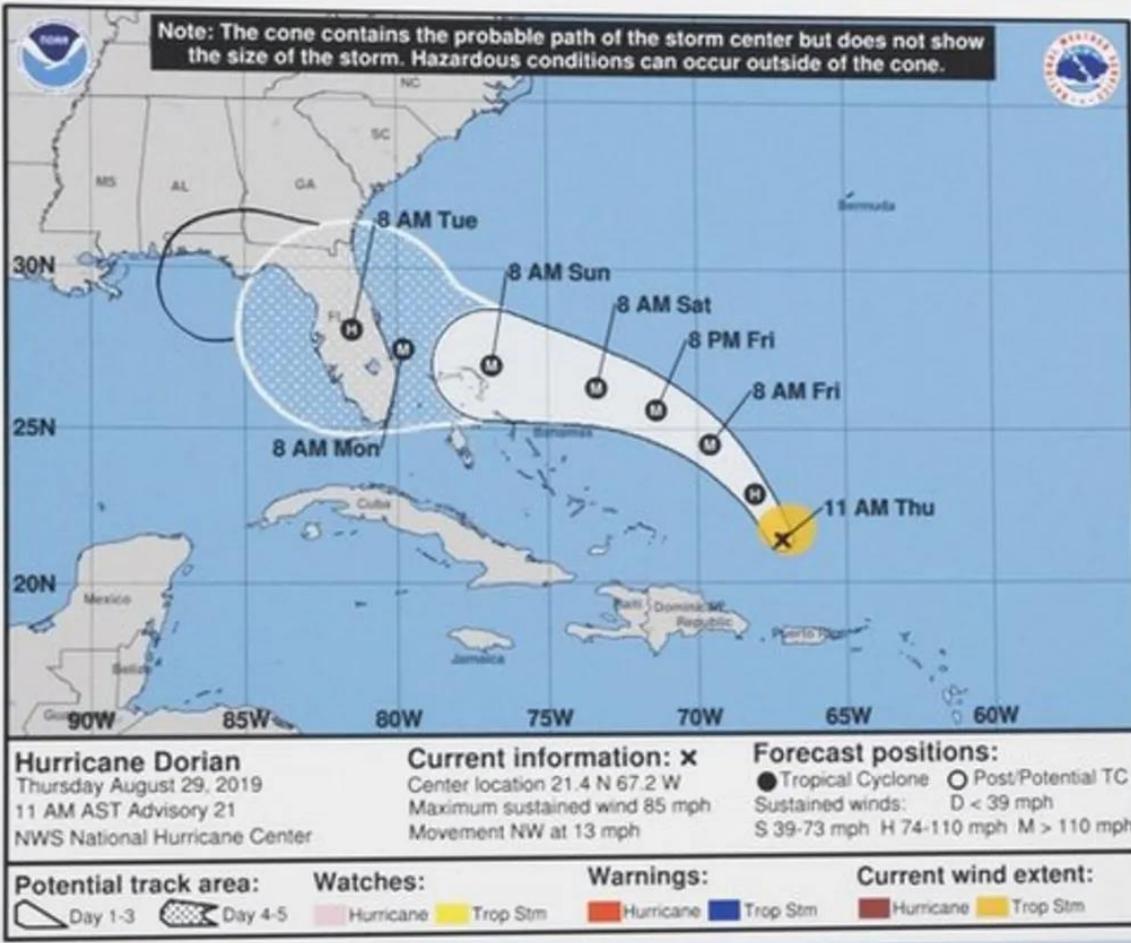
- What Does Uncertainty Mean?
  - LOTS OF THINGS
- How Should I Visualize It?
  - IT DEPENDS!
- What Can Go Wrong?
  - A LOT



# Misinformation pt 1

# Misinformation Part 1

## Hurricane Dorian Forecast Track and Intensity

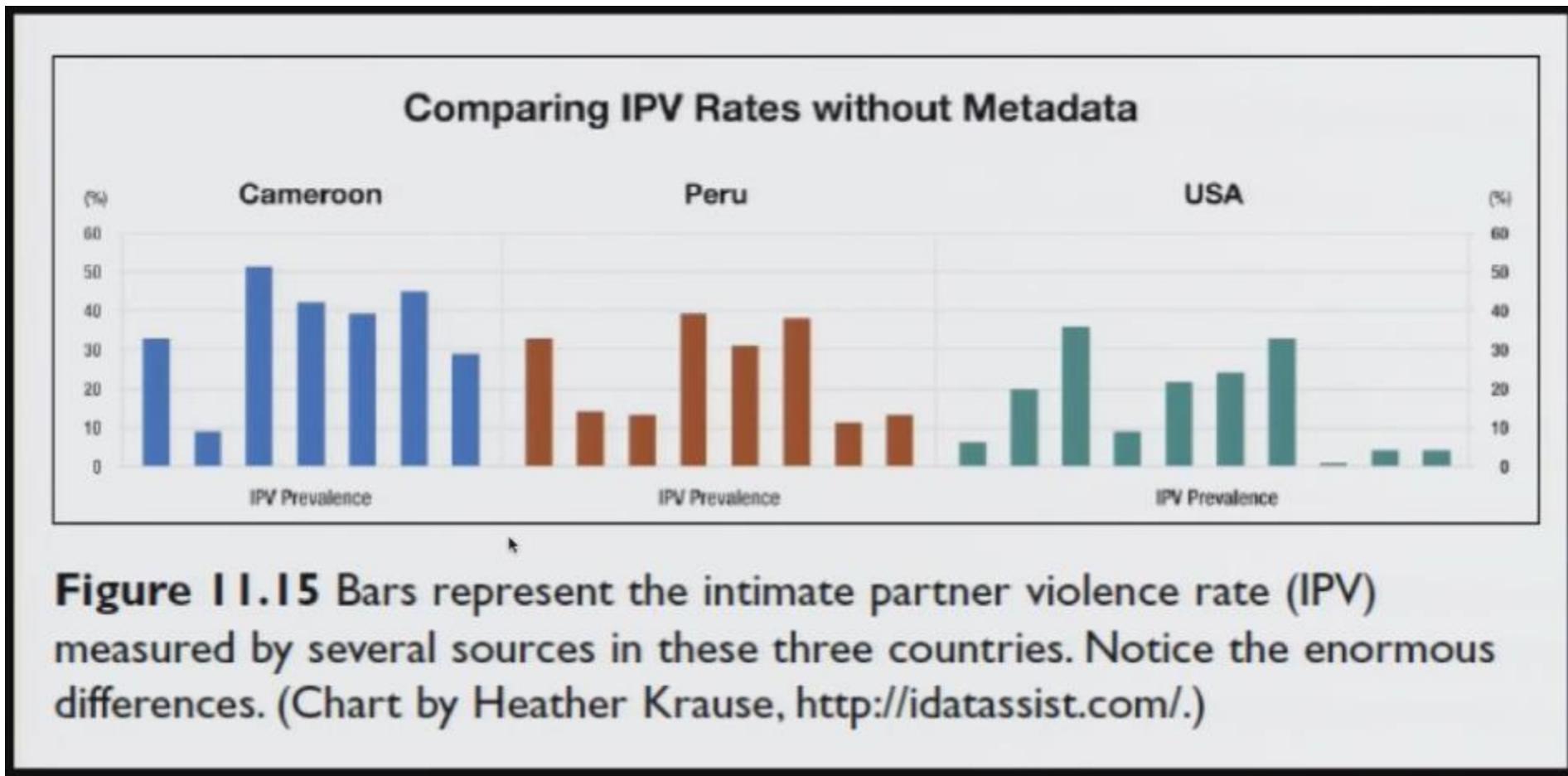


# Misinformation Part 1

- 5 rules to not get mislead:
  - Is the author using the correct data and disclosing the source?
  - Are you reading too much into the graphic?
  - Are the data represented accurately?
  - Is the visualization showing an appropriate amount of data?
  - Is uncertainty relevant? If yes, is it shown?

# Misinformation Part 1

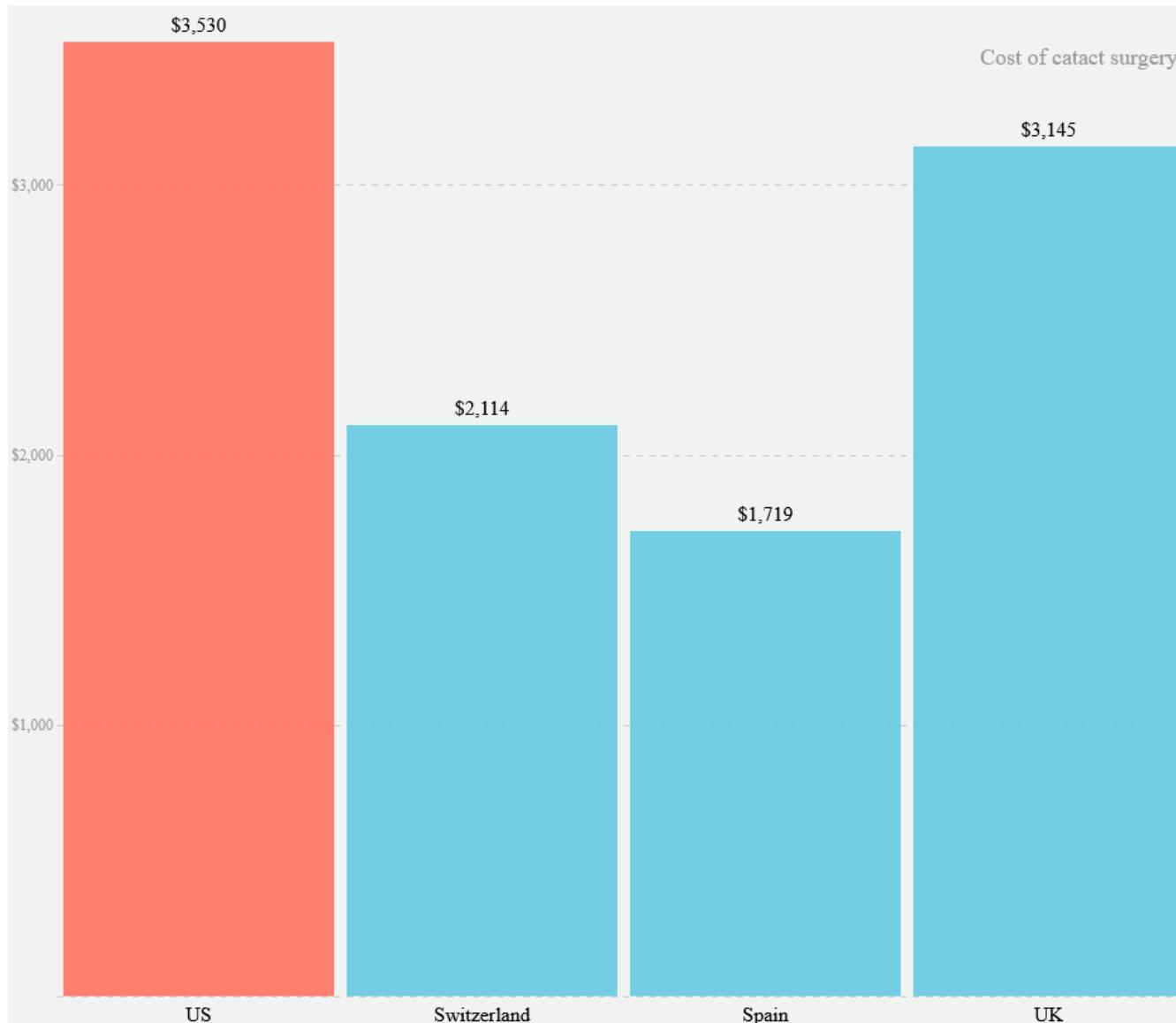
- Is the author using the correct data and disclosing the source?



Example by Heather Krause quoted in *The Truthful Art* (Chapter 11)

# Misinformation Part 1

- Is the author using the correct data and disclosing the source?



<https://www.vox.com/a/health-prices#chart/9>

# Misinformation Part 1

- Is the author using the correct data and disclosing the source?



International Federation  
of Health Plans

2015 Comparative Price Report  
Variation in Medical and Hospital Prices by Country

## 2015 Survey Overview

This year's survey has been conducted in much the same way as the previous study, which includes pricing for several specialty prescription drugs, other prescription drugs and a selection of typical medical procedures. Prices for each country were submitted by participating federation member plans, and are drawn from public or commercial sectors as follows:

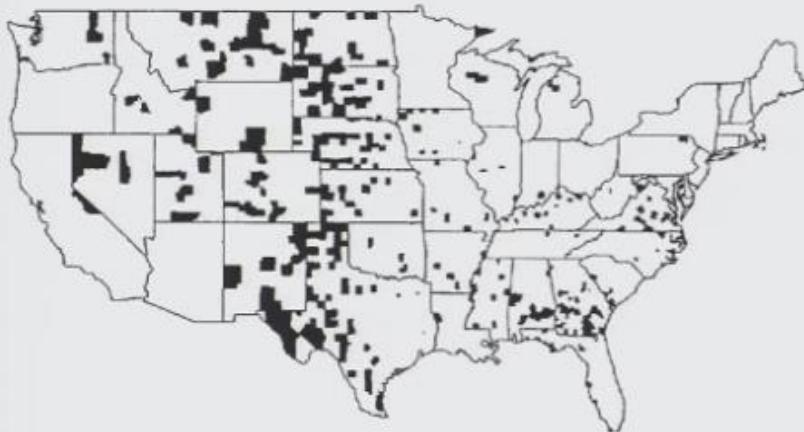
- Prices for the United States were derived from over 370 million medical claims and over 170 million pharmacy claims that reflect prices negotiated and paid to health care providers.
- Prices for Australia, New Zealand, Spain, South Africa, Switzerland and the UK are from the private sector, with data provided by one private health plan in each country.

# Misinformation Part 1

- Are you reading too much into the graphic?

Always tell yourself:

“A chart shows as much as it hides —think about what might be hidden”

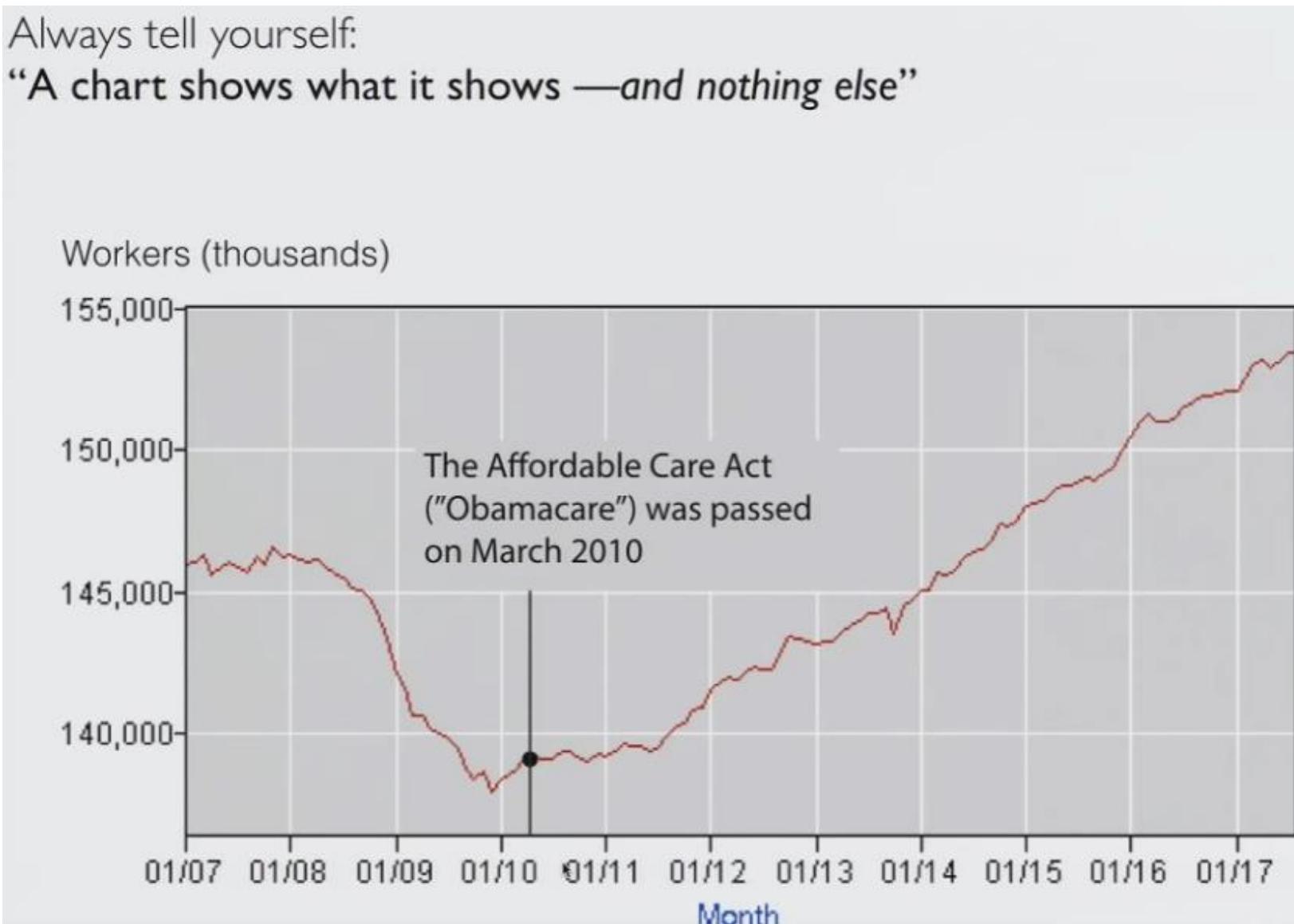


Counties with the LOWEST  
kidney cancer death rates  
(1980-1989)

# Misinformation Part 1

- Are you reading too much into the graphic?

Always tell yourself:  
“A chart shows what it shows —*and nothing else*”

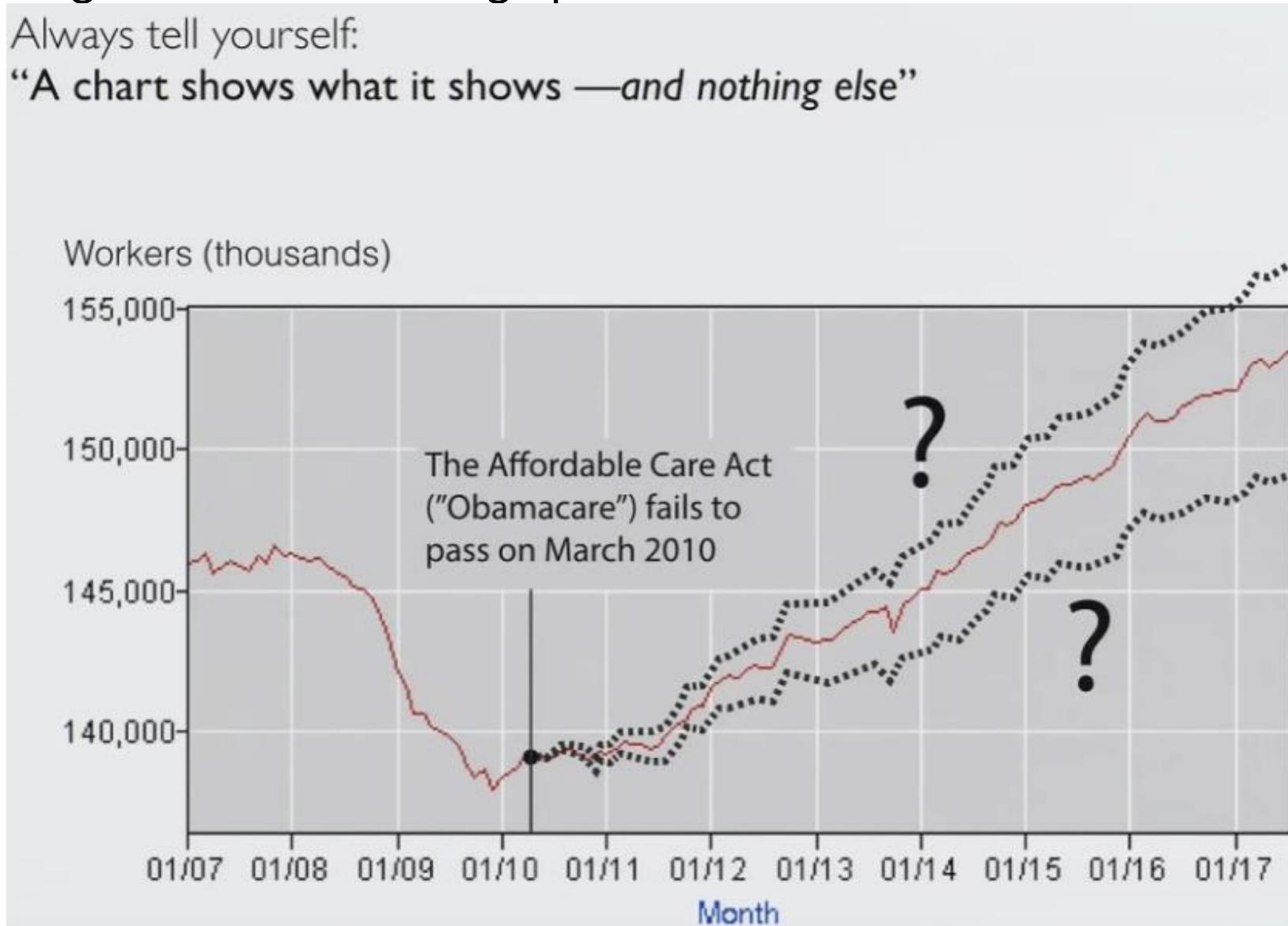


# Misinformation Part 1

- Are you reading too much into the graphic?

Always tell yourself:

“A chart shows what it shows —*and nothing else*”



# Misinformation Part 1

- Are the data represented accurately?

The screenshot shows the homepage of the Venezuelan Television (VTV) website. The top navigation bar includes links for 'INICIO', 'NOTICIAS', 'PROGRAMAS', 'MULTIMEDIA', 'OPINIÓN', 'EL CANAL', 'ELECCIONES', and 'HUGO CHÁVEZ'. The date 'Lunes, 15 de Abril 2013' is displayed at the top right. A banner for the 'ELECCIÓN PRESIDENCIAL 2013' features portraits of Nicolás Maduro and Henrique Capriles, along with a 3D bar chart representing the percentages. The chart shows a red bar at approximately 50.66% and a blue bar at approximately 49.07%. The main headline on the right reads: 'Ganó el hijo de Chávez: Nicolás Maduro es el Presidente Electo de Venezuela (+Video)'. Below it, a quote from Tibisay Lucena states: 'La tendencia es irreversible, señaló Tibisay Lucena, presidenta del ente comicial / Maduro obtuvo 50.66% v Capriles 49.07%'.

**ELECCIÓN PRESIDENCIAL 2013**

**PORCENTAJES**

Nicolás Maduro 50.66%

Henrique Capriles 49.07%

Ganó el hijo de Chávez: Nicolás Maduro es el Presidente Electo de Venezuela (+Video)

La tendencia es irreversible, señaló Tibisay Lucena, presidenta del ente comicial / Maduro obtuvo 50.66% v Capriles 49.07%

Nicolás Maduro: Hoy es un día de Historia, vemos con Cristo Redentor (+Fotos)

# Misinformation Part 1

- Are the data represented accurately?

Lunes, 15 de Abril 2013

Venezolana de Televisión VENEZUELA ¡de verdad!

Buscar en el Sitio

SERIAL EN VIVO RADIONET VTV

INICIO NOTICIAS PROGRAMAS MULTIMEDIA OPINIÓN EL CANAL ELECCIONES HUGO CHÁVEZ

**VOTO 2013** El presidente electo

ELECCIÓN PRESIDENCIAL 2013

**PORCENTAJES**

50,66 % 49,07 %

Nicolás Maduro

Nicolas Capriles Radonski

Ganó el hijo de Chávez: Nicolás Maduro es el Presidente Electo de Venezuela (+Video)

ELECCIONES

La tendencia es irreversible, señaló Tibisay Lucena, presidenta del ente comicial / Maduro obtuvo 50.66% y Capriles 49.07%

Ler más

Ganó el hijo de Chávez: Nicolás Maduro es el Presidente Electo de Venezuela (+Video)

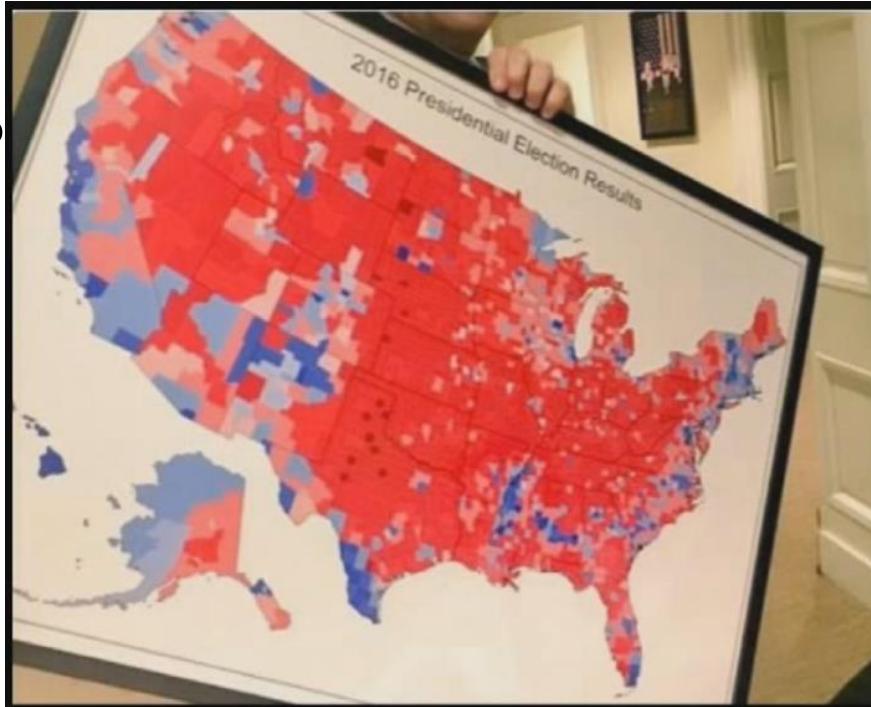
Nicolas Maduro rechazó "Pacto" que le propuso Capriles antes de conocerse resultados oficiales

Hackearon cuentas Twitter del Presidente Nicolás Maduro y del Pdvsa

Nicolas Maduro: Hoy es un día de historia, vamos con Cristo Redentor (+Fotos)

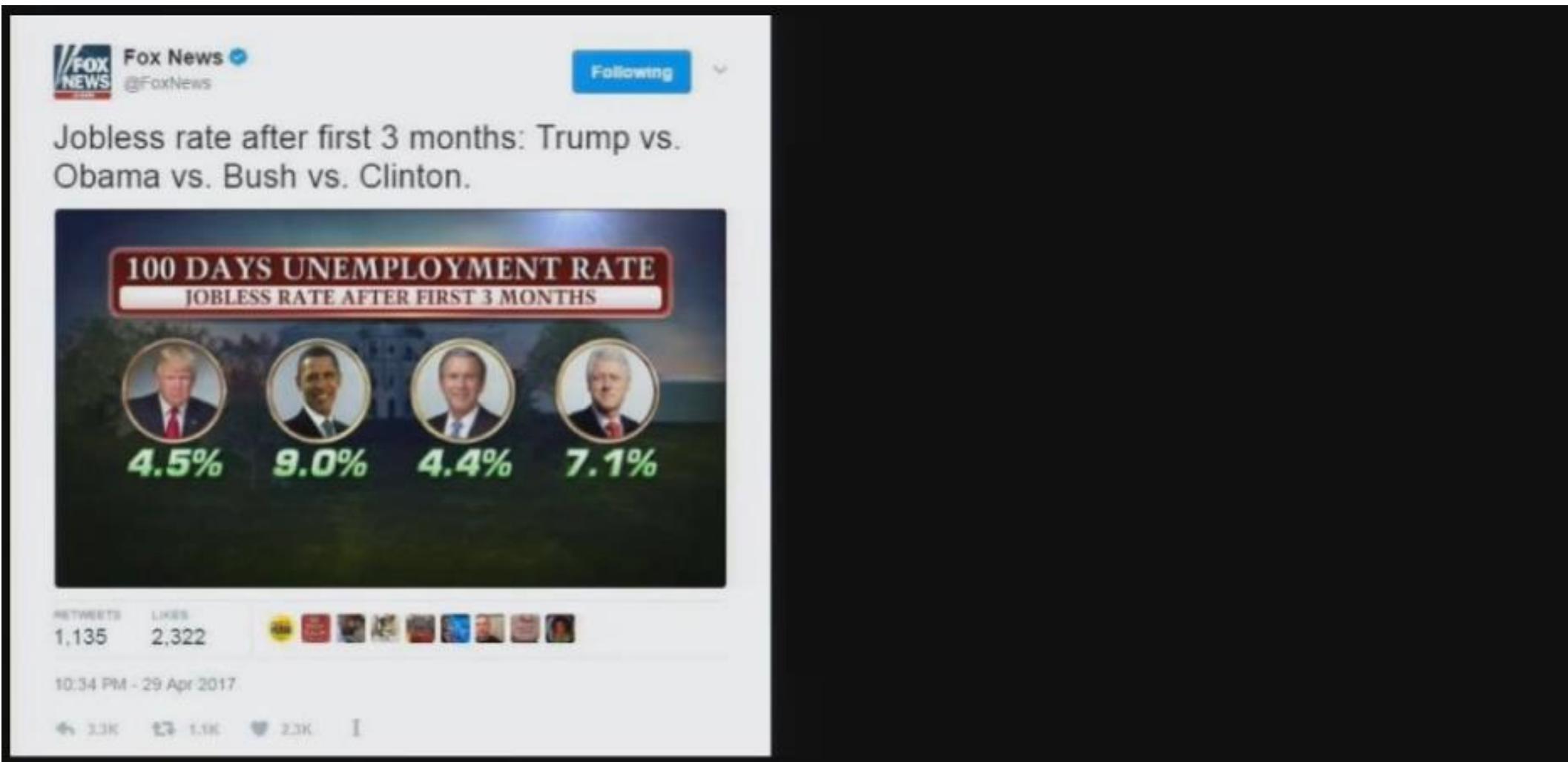
# Misinformation Part 1

- Are the data represented accurately?



# Misinformation Part 1

- Is the visualization showing an appropriate amount of data?



# Misinformation Part 1

- Is uncertainty relevant? If yes, is it shown?

The screenshot shows a news article from EL PAÍS. At the top left is the EL PAÍS logo with three horizontal bars and the word "EL PAÍS". At the top right is the text "IN ENGLISH". Below the logo is a navigation bar with links to "POLITICS", "ARTS", "SPORTS", "OPINION", and "SPANISH WAY OF LIFE". Underneath this is a "REGIONAL POLITICS" section header. The main title of the article is "Catalan public opinion swings toward “no” for independence, says survey". A subtitle below the title reads "Growing numbers of citizens support federal model that falls short of secession from Spain". There are social media sharing icons for Facebook, Twitter, and LinkedIn at the bottom of the article area. The author's name, PERE RÍOS, and the publication date, Barcelona - 19 DIC 2014 - 10:23 EST, are also visible. At the very bottom of the page is a black footer bar containing the URL [http://ccaa.elpais.com/ccaa/2014/12/19/catalunya/1418984873\\_128596.html](http://ccaa.elpais.com/ccaa/2014/12/19/catalunya/1418984873_128596.html).

# Misinformation Part 1

- Is uncertainty relevant? If yes, is it shown?

For the first time since Catalan leader Artur Mas began his ongoing independence drive in 2012, a survey shows that a majority in the region would reject secession if a referendum were held now.

[http://www.thefunctionalart.com/  
2014/12/adventures-in-margin-of-  
error.html](http://www.thefunctionalart.com/2014/12/adventures-in-margin-of-error.html)

The latest poll by the Catalan executive's Opinion Studies Center (CEO) shows that 45.3 percent of citizens would vote no to the question: "Would you like Catalonia to become an independent state?" compared with 44.5 percent who would support the move.

*Do you want Catalonia  
to become an independent state?*



Margin of error: +/-2.95 at 95% confidence level