

Uncertainty and Review

DSC 106: Data Visualization

Sam Lau

UC San Diego

Announcements

Lab 8 due this Friday.

Final Project Demo Video due next week Tuesday.

Lecture on Friday is Feedback Session for final projects.

Sam holding extra OH today 4-5pm in HDSI 455.

No discussion next week – come to OH instead!

FAQs:

1. **Why is the demo video due next week (before we've even finished our final projects)?** To make your demo video, your final project doesn't need to be 100% working. Just show the 20% that is working well! We want you to start thinking seriously about your project story.

A peek behind the curtain: demo videos

FileHomeInsertFormulasDataReviewViewAutomateHelpTweakItOpen in Desktop AppTell me what you want to doEditingShareComments

Toggle ScratchpadSave Output to Sheet

A3fxUUAA

	A	B	C	D	E	F	G
1	CUCG						
2	AAAU						
3	UUAA						
4	GGGG						
5	CCCC						
6	UCAC						
7	AAUU						
8	RRRR						
9							
10							
11							
12							
13							
14							
15							
16							

TweakIt

Available Python variables:

A = A3

```
1 # Example 1
2 i = "AUGC"
3 j = "UACG"
4 tbl = str.maketrans(i, j)
5 my_str = "GUAC"
6 my_str.translate(tbl)
7
8 # Example 2
9 from Bio.Seq import Seq
10 my_dna = Seq("CCGATAG")
11 my_dna.complement()
```



The screenshot shows a web-based spreadsheet application named 'TweakIt'. The interface includes a top menu bar with options like File, Home, Insert, Formulas, Data, Review, View, Automate, Help, and TweakIt. Below the menu bar, there are buttons for 'Toggle Scratchpad' and 'Save Output to Sheet'. The spreadsheet grid has columns labeled A through G and rows numbered 1 through 10. Column A contains the following DNA sequences: CUCG, AAAU, UUAA, GGGG, CCCC, UCAC, AAUU, and RRRR. A blue dashed selection box is drawn around the cell containing 'CCCC' in row 5, column B. A mouse cursor is pointing at the bottom-right corner of this selection box. On the right side of the spreadsheet, there is a 'TweakIt' panel with the text 'Available Python variables:' and 'A = A4'. Below this, there is a code editor with the following Python code:

```
1 from Bio.Seq import Seq
2 my_dna = Seq(A)
3 my_dna.complement()
```

	A	B	C	D	E	F	G
1	CUCG						
2	AAAU						
3	UUAA						
4	GGGG						
5	CCCC	CCCC					
6	UCAC						
7	AAUU						
8	RRRR						
9							
10							

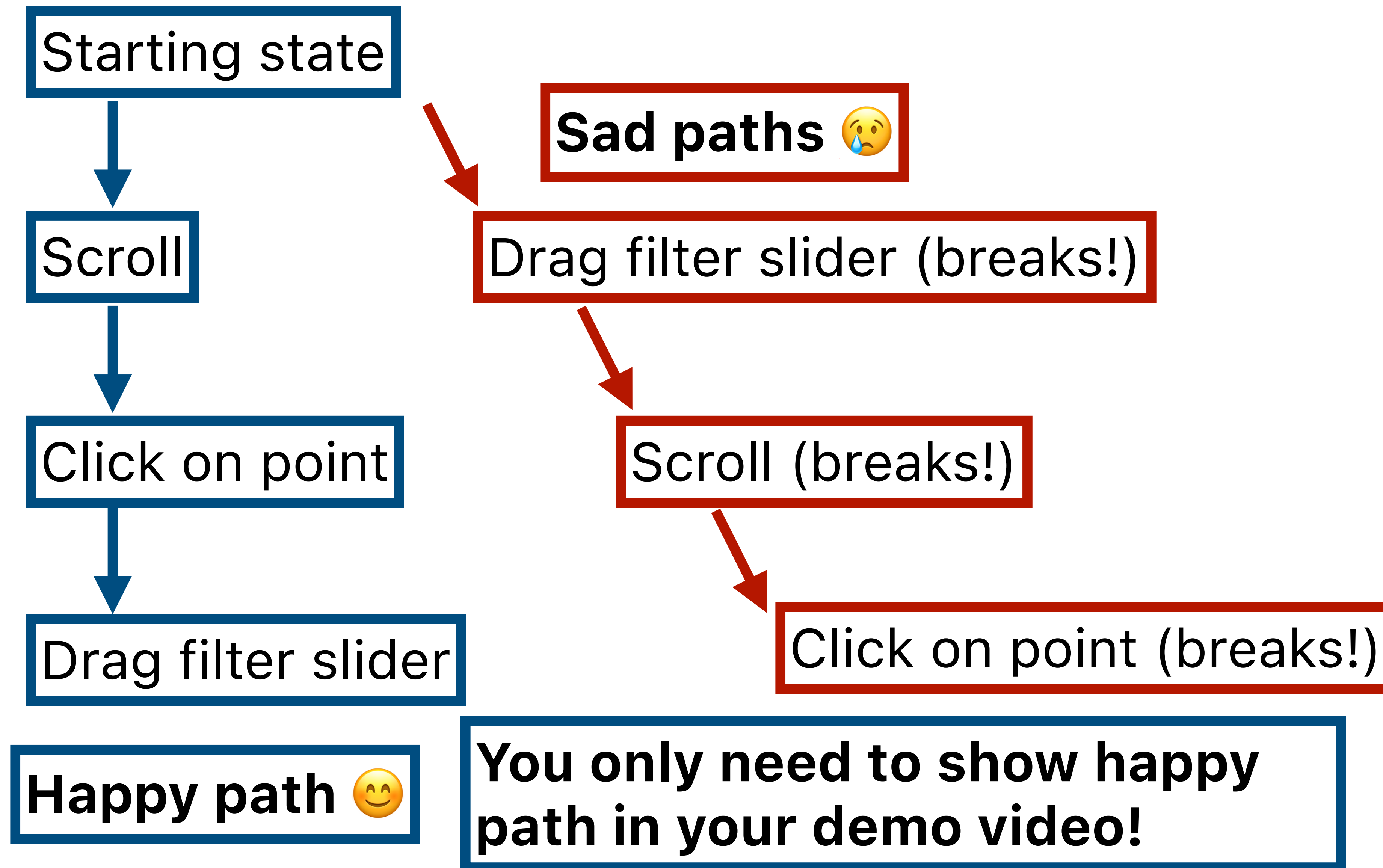
Looks sophisticated, but really I'm just erasing ALL cells in column B and then writing into the cell to the right of the selection.

If there was other data in the spreadsheet, it would be wiped out!

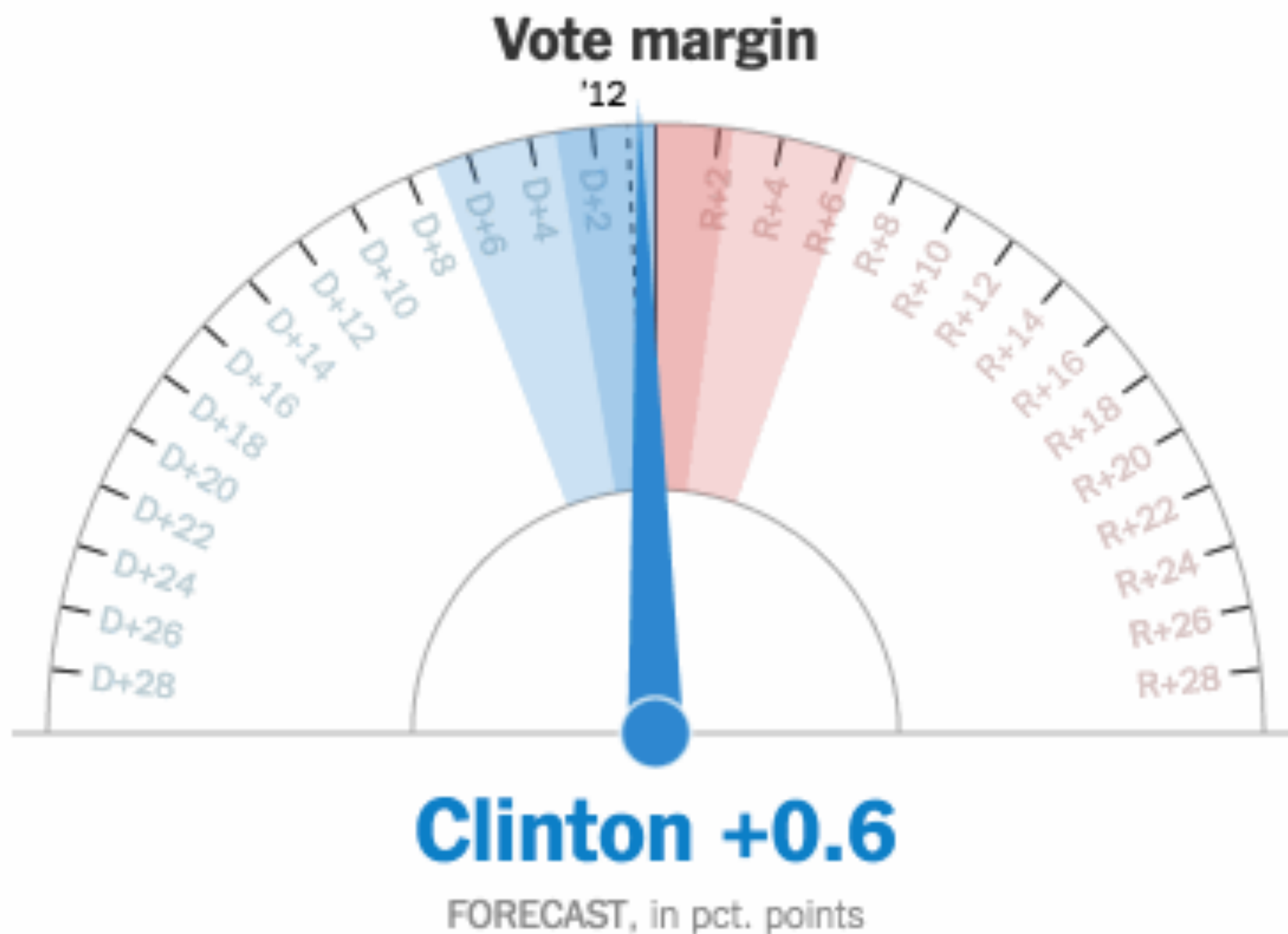
	A	B	C
1	CUCG	GAGC	
2	AAAU	UUUA	
3	UUAA	AAUU	
4	GGGG	CCCC	
5	CCCC	GGGG	
6	UCAC	AGUG	
7	AAUU	UUAA	
8	RRRR	YYYY	
9			

Clicking "Save Output" just removes the blue background from column B. (It doesn't actually work!)

Making a cool demo video



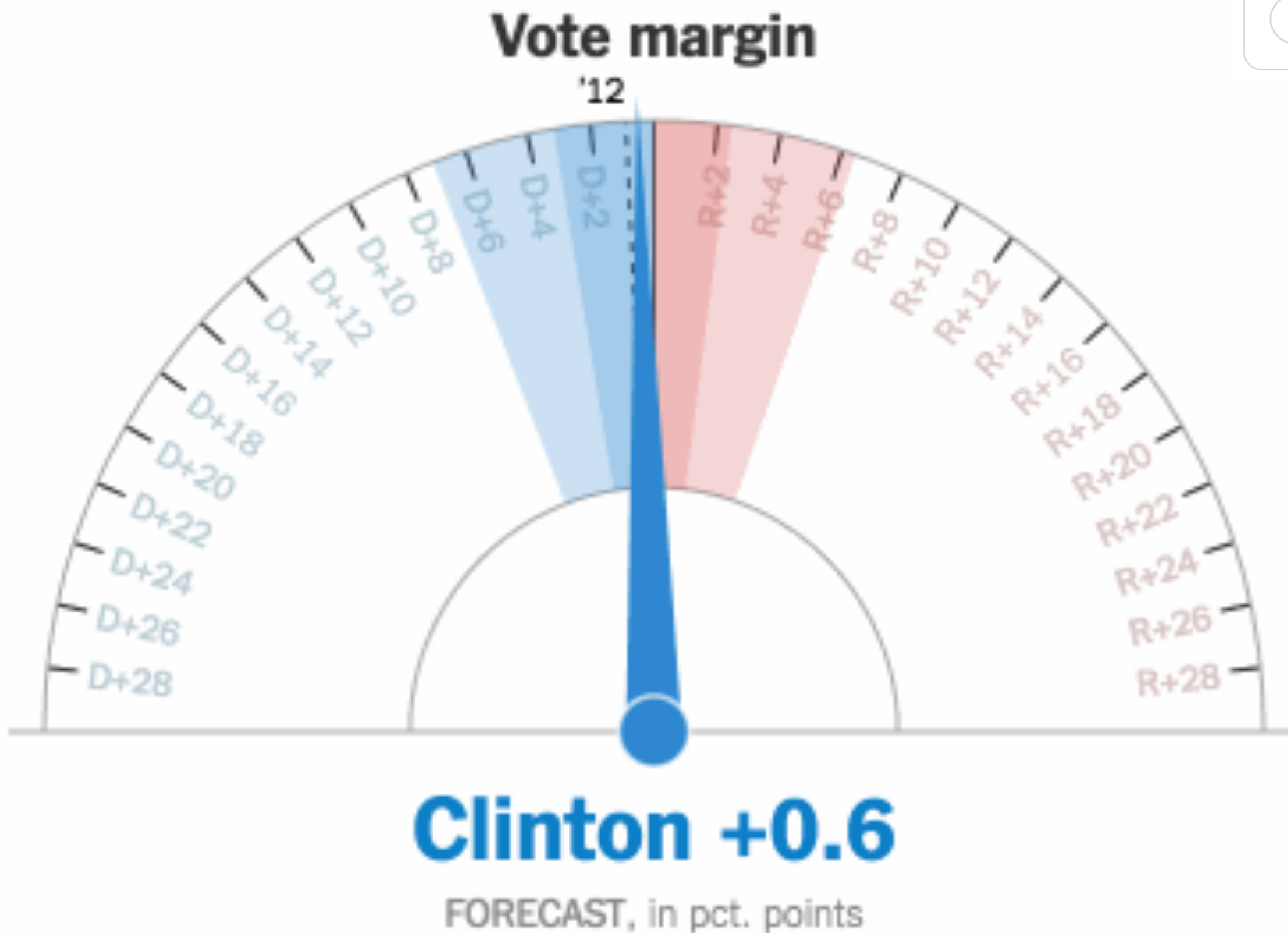
Uncertainty



What is being visualized?

What are the strengths and weaknesses of this visualization?

tryclassbuzz.com
Code: **needle**



(gobsmacked)
@nytimes

I'm taking
damn NY

10:08 PM ·

2

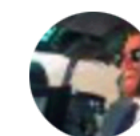


Danielle Rosen
@Daniel_Rosen · Follow

The NYT needle jittering is going to be in my nightmares no matter what happens tonight.

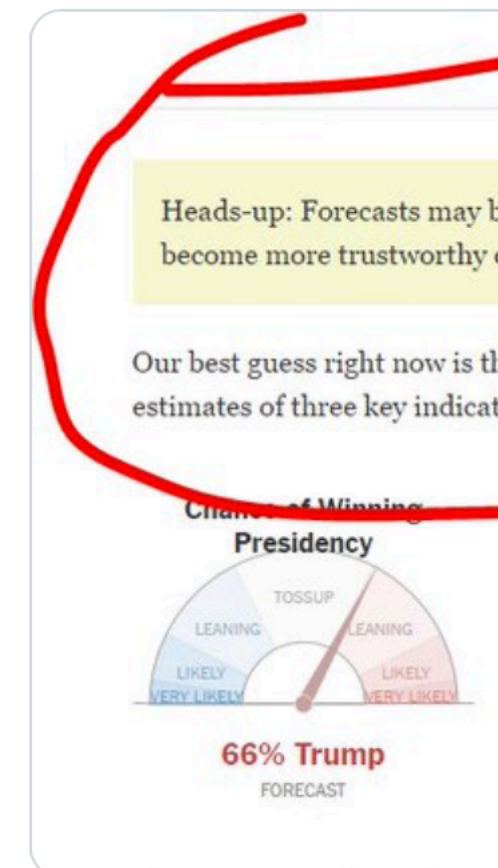
10:00 PM · Nov 8, 2016

2



Paul D. Shinkman ✓
@PDSrinkman · Follow

Has @nytimes caught on to Dems' election-night hair-greying freakout?



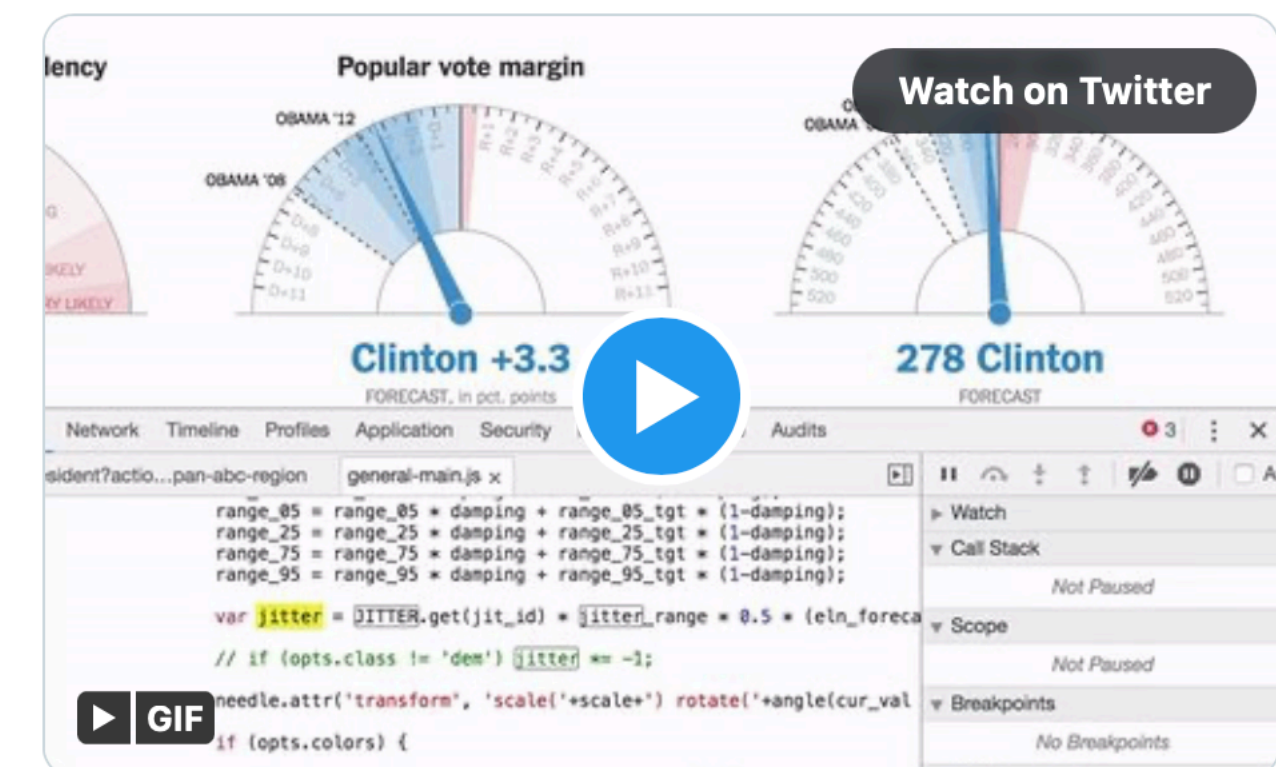
10:03 PM · Nov 8, 2016

Reply



Alp Toker ✓
@atoker · Follow

Looking for trends in @nytimes's presidential forecast needle? Don't look too hard - the bounce is random jitter from your PC, not live data



@gka@vis.social
@driven_by_data · Follow

not just random. this noise is conveying the uncertainty in our forecast (jitter range is from 25th–75th pctl in sims.)

10:01 PM · Nov 8, 2016

33

Reply

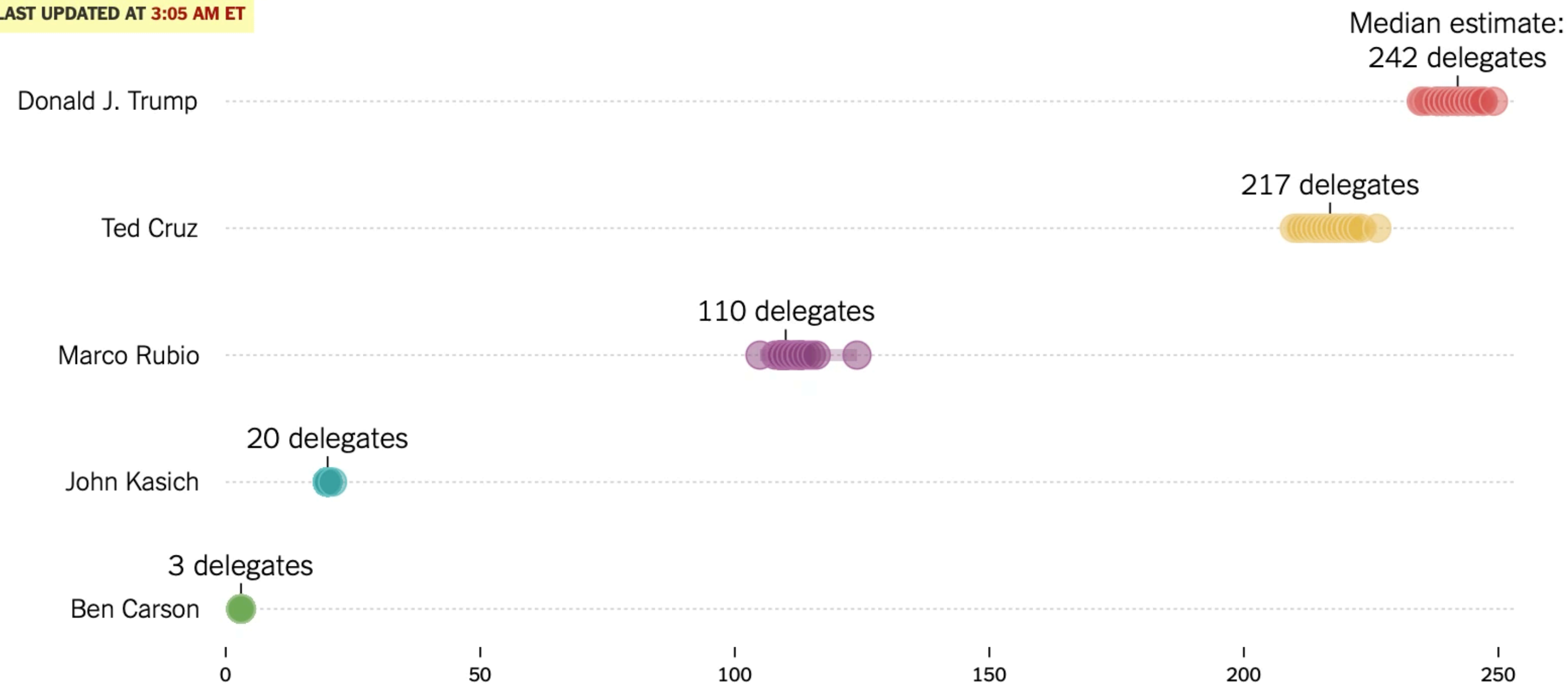
Share

Read 5 replies

Who Will Win Super Tuesday? Live Estimates of Tonight's Final Republican Delegate Count

By AMANDA COX, JOSH KATZ and KEVIN QUEALY MARCH 1, 2016

LAST UPDATED AT 3:05 AM ET



We're simulating the number of delegates each candidate will pick up on Super Tuesday. The dots above represent a range of possible outcomes.

What is being visualized?

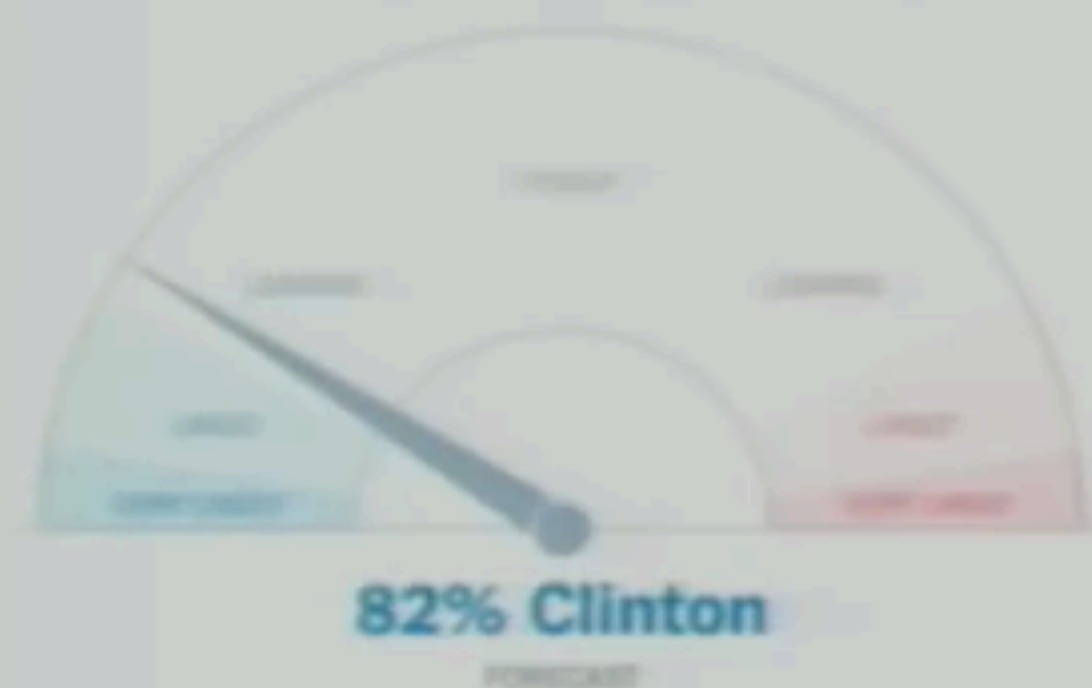
What are the strengths and weaknesses of this visualization?

How does it compare to the needle?

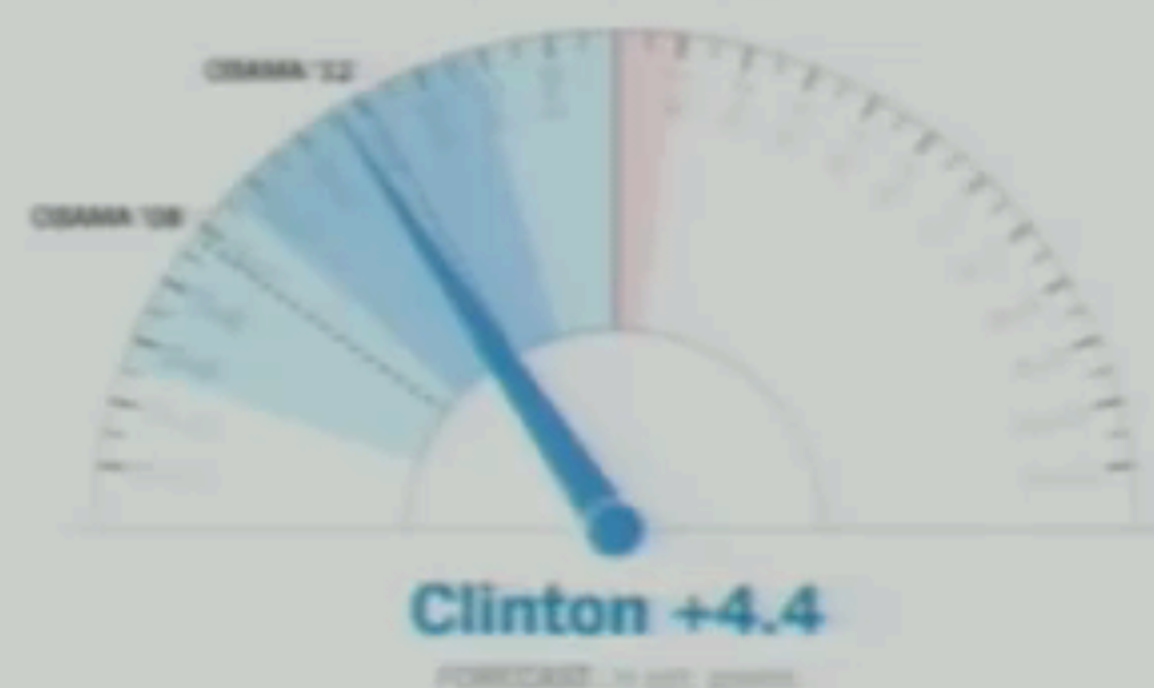
Live Presidential Forecast

Updated 7:55:14 PM ET

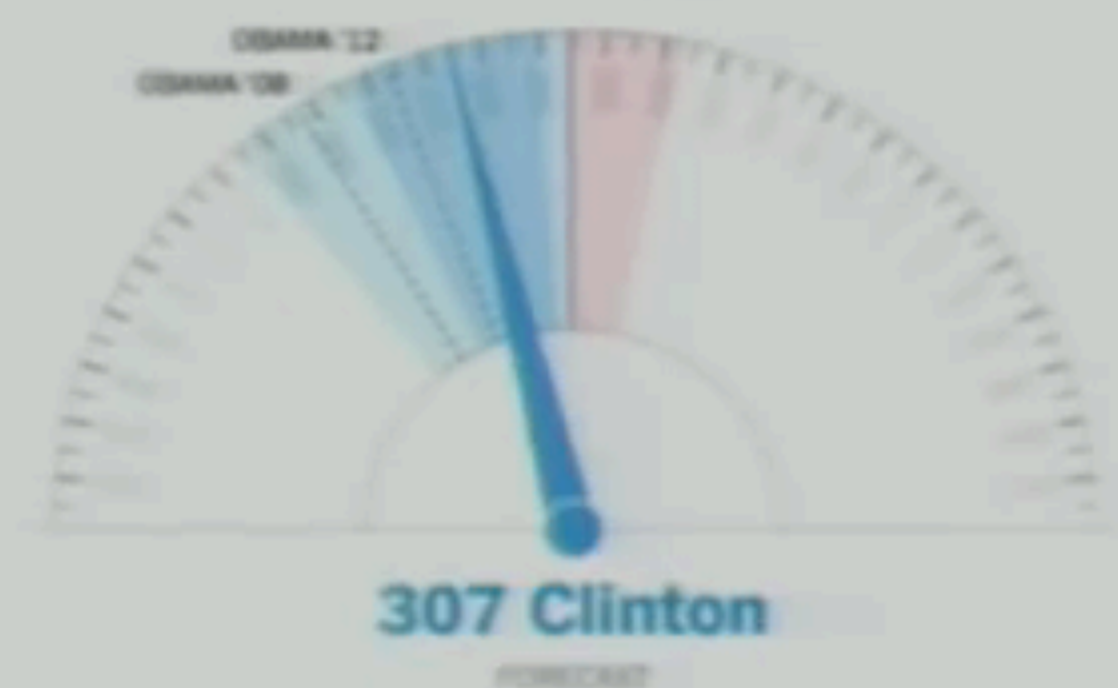
Chance of Winning Presidency



Popular vote margin



Electoral votes



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](#)

State	Est. pct. of voters	Reported margin	NYT projection	NYT win prob.	10%	20%	30%	Even	+10%	+20%	+30%	+40%
Michigan	0%		Clinton +6.5	75% Dem.								
New Mexico	0%		Clinton +6.4	76% Dem.								
Wisconsin	0%		Clinton +5.4	75% Dem.								

OPEN
VIS 2017
CONF

Uncertainty

What does it mean?

How should I visualize it?

Uncertainty

What does it mean?

How should I visualize it?

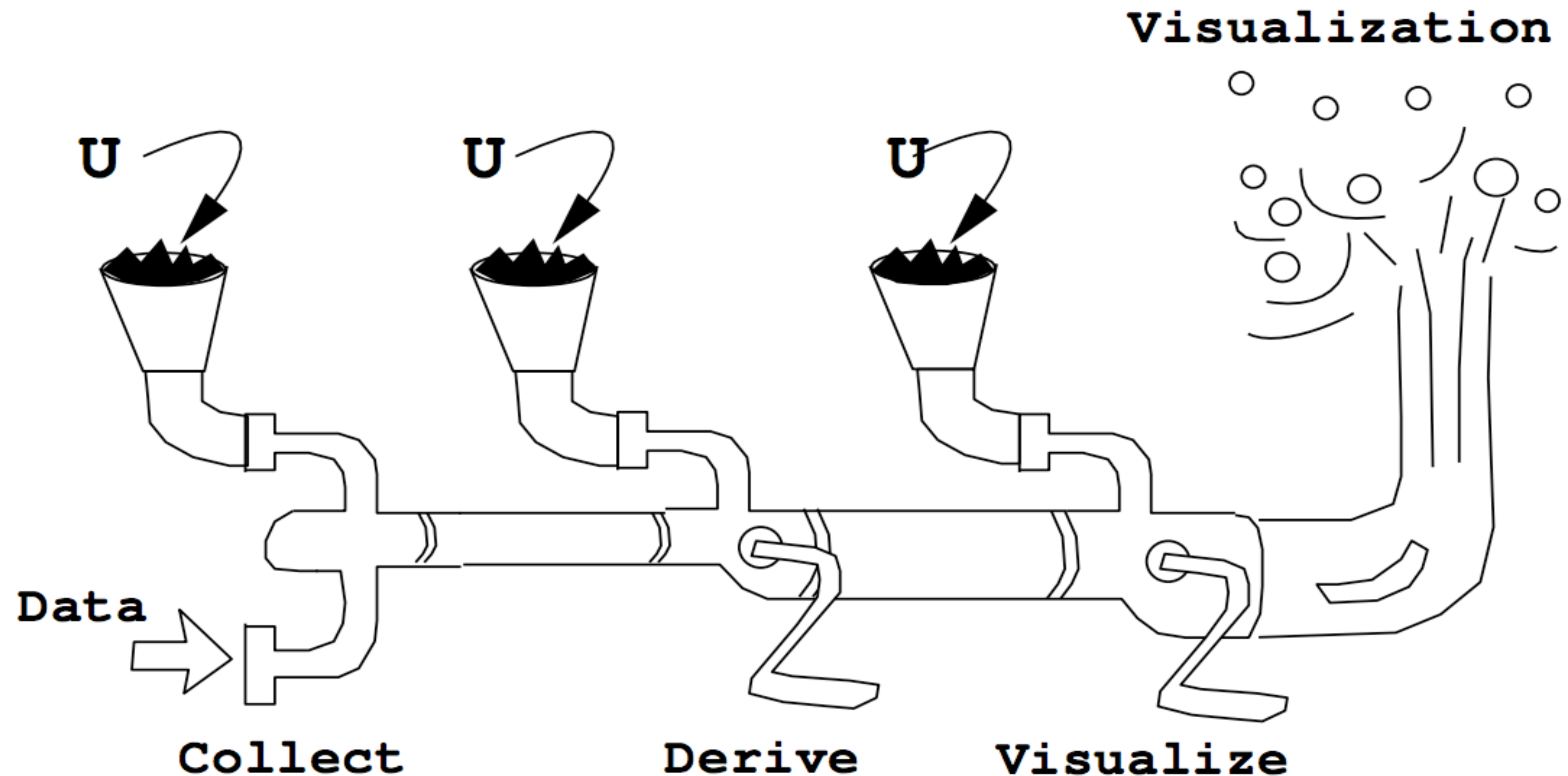
Uncertainty

What does it mean?

How should I visualize it?

Doubt
Risk
Variability
Error
Lack of knowledge
Hedging
etc...

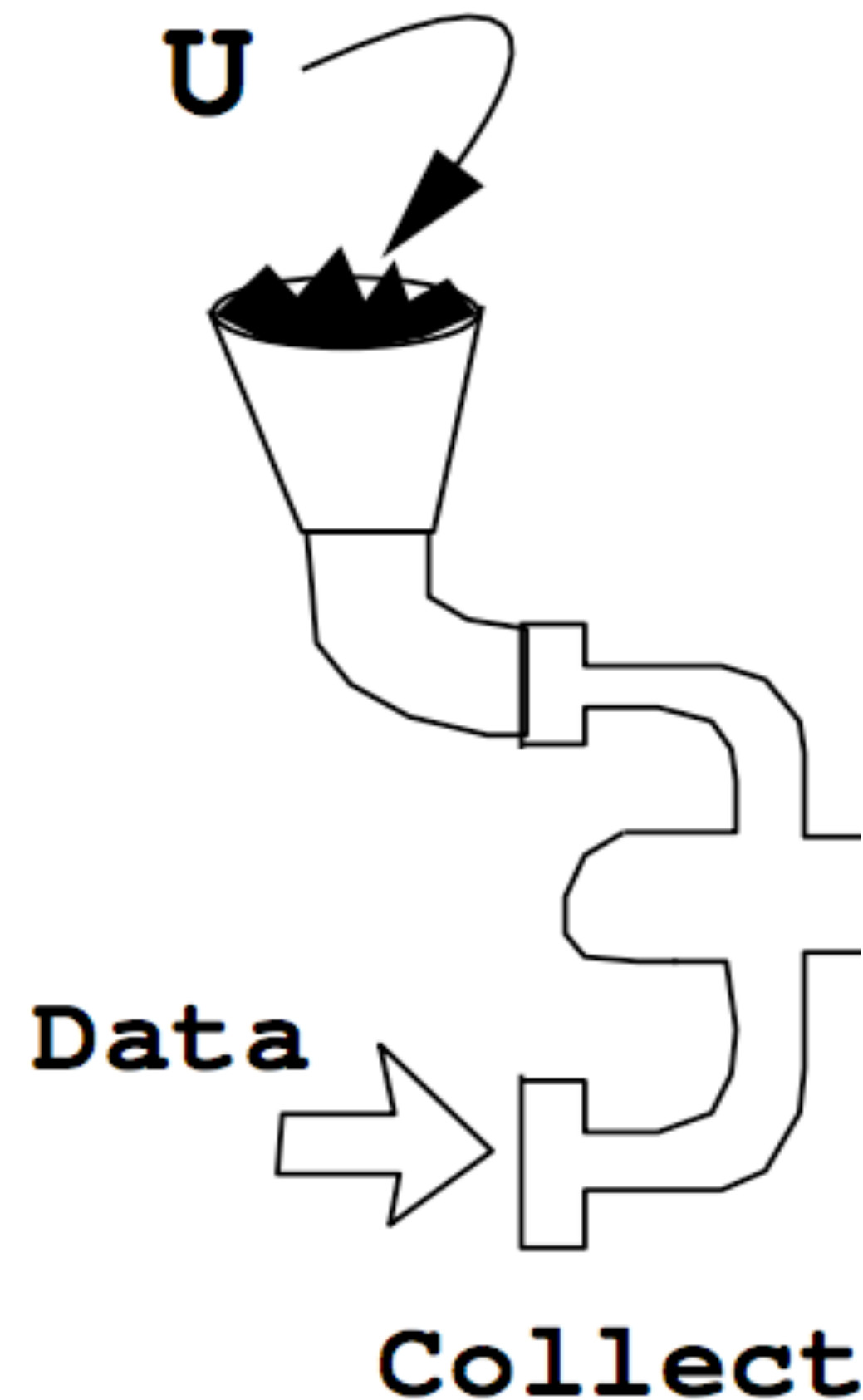
Sources and Types of Uncertainty



Sources and Types of Uncertainty

Measurement Uncertainty

How and how much should we sample the data?



Precision

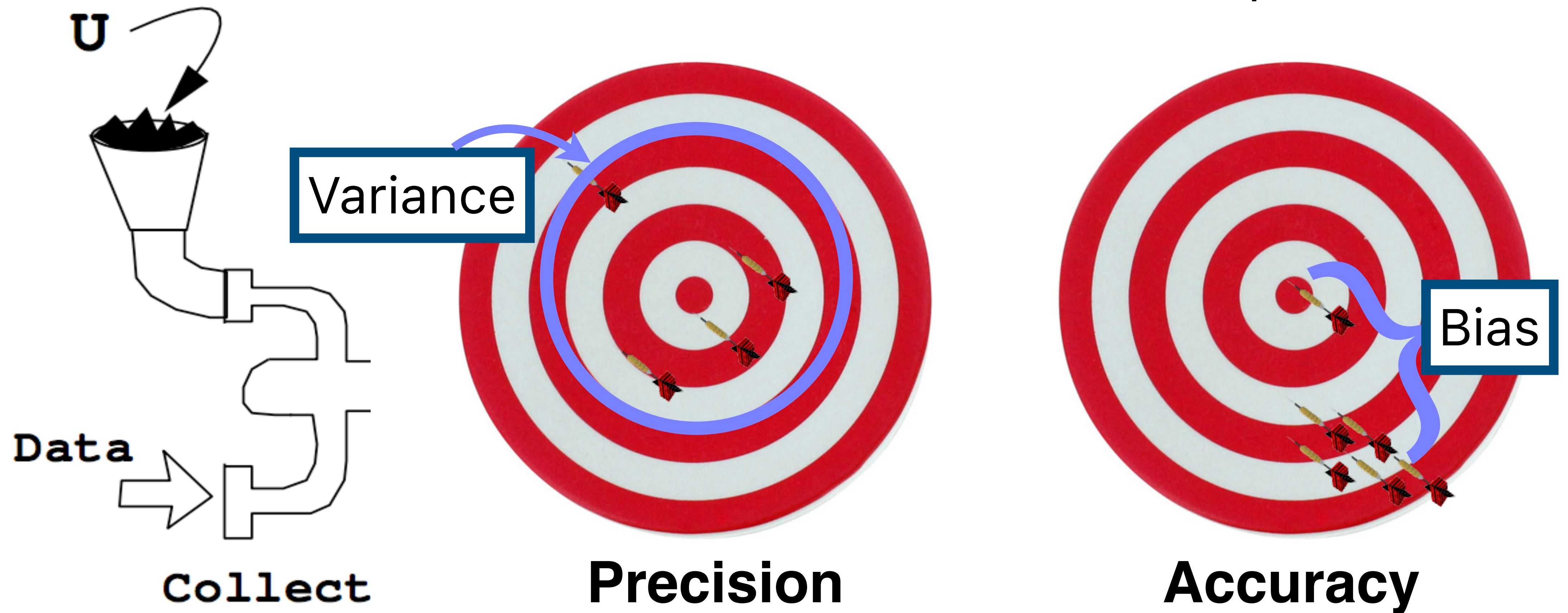


Accuracy

Sources and Types of Uncertainty

Measurement Uncertainty

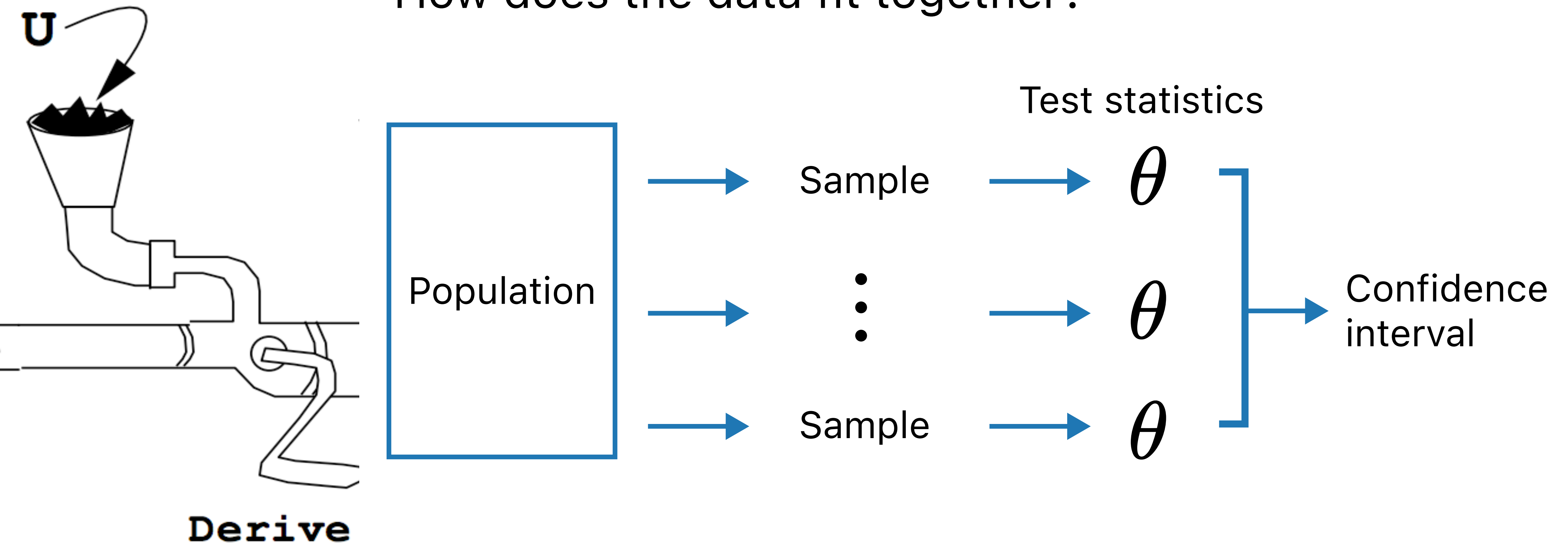
How and how much should we sample the data?



Sources and Types of Uncertainty

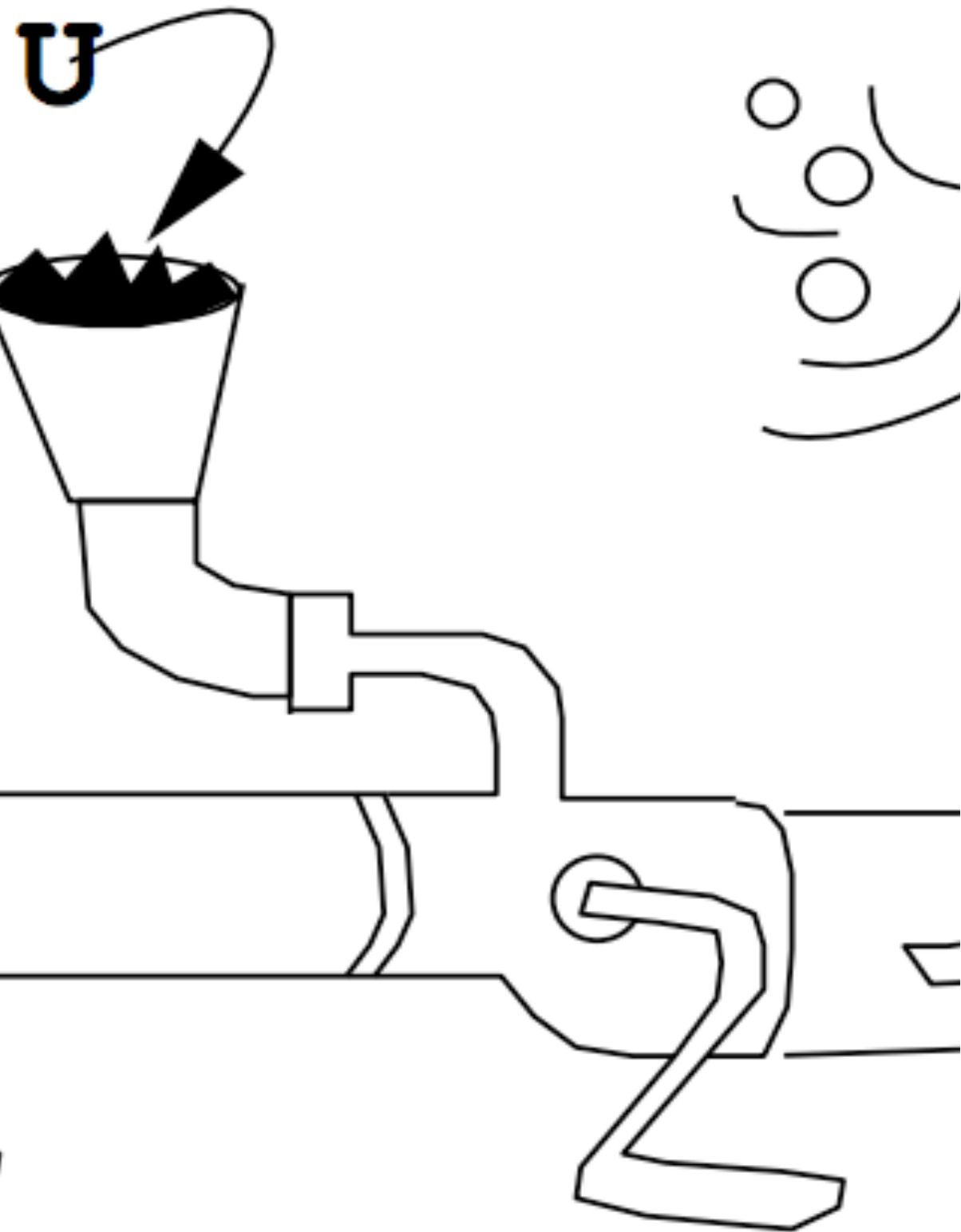
Model Uncertainty

How does the data fit together?



Sources and Types of Uncertainty

Visuali



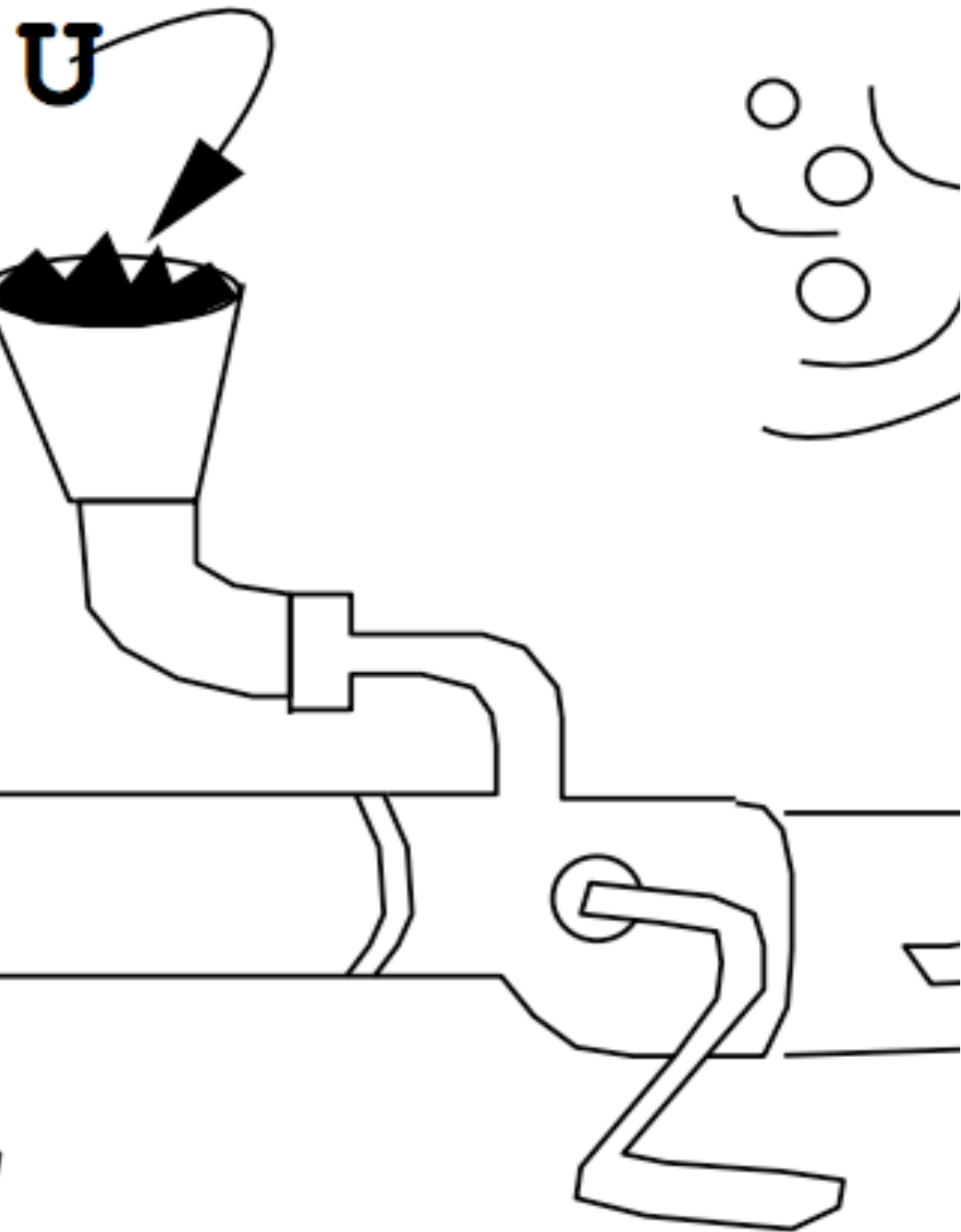
Visualize

Decision/Forecast Uncertainty
How do I assess the risk or error?



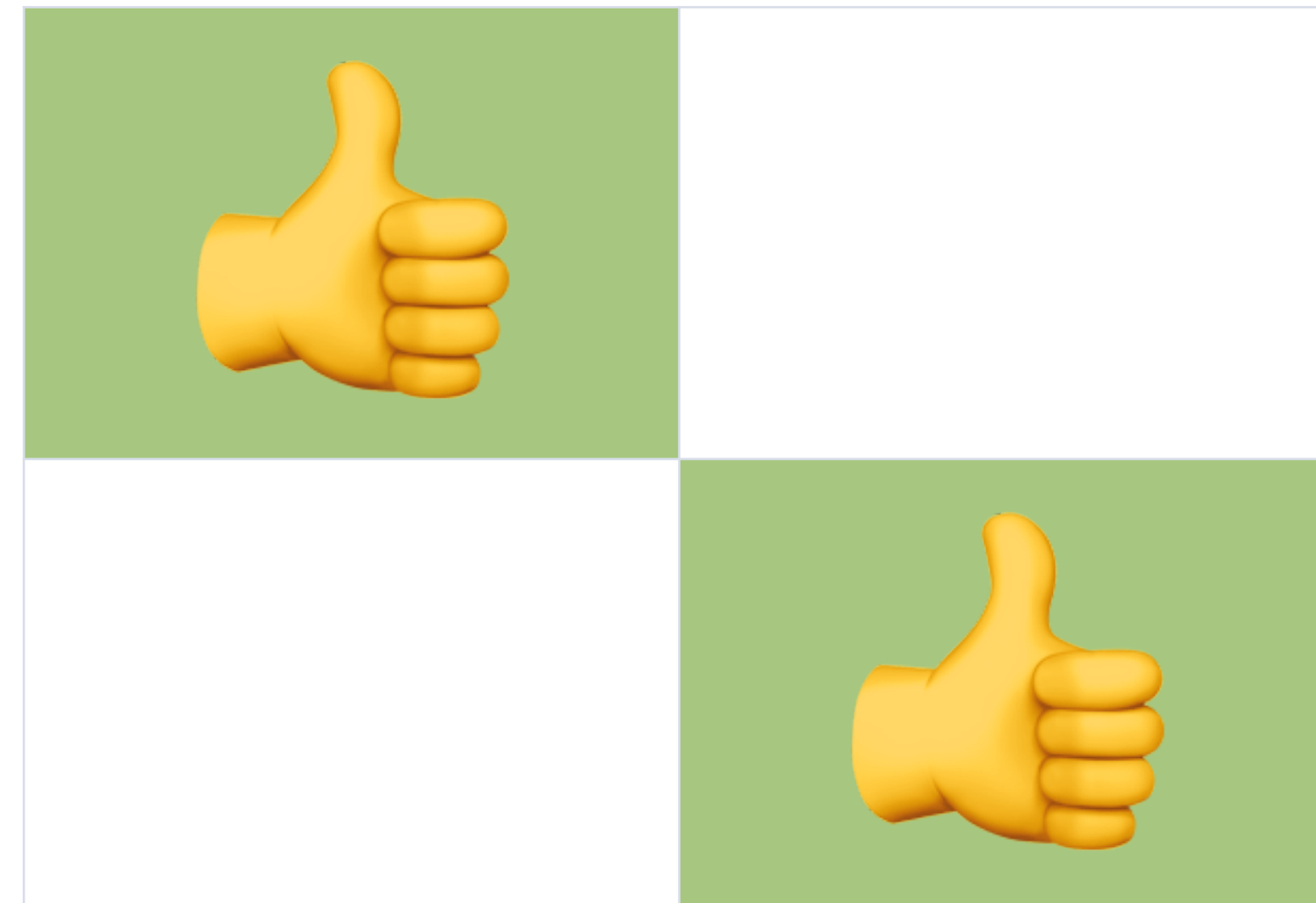
Sources and Types of Uncertainty

Visuali



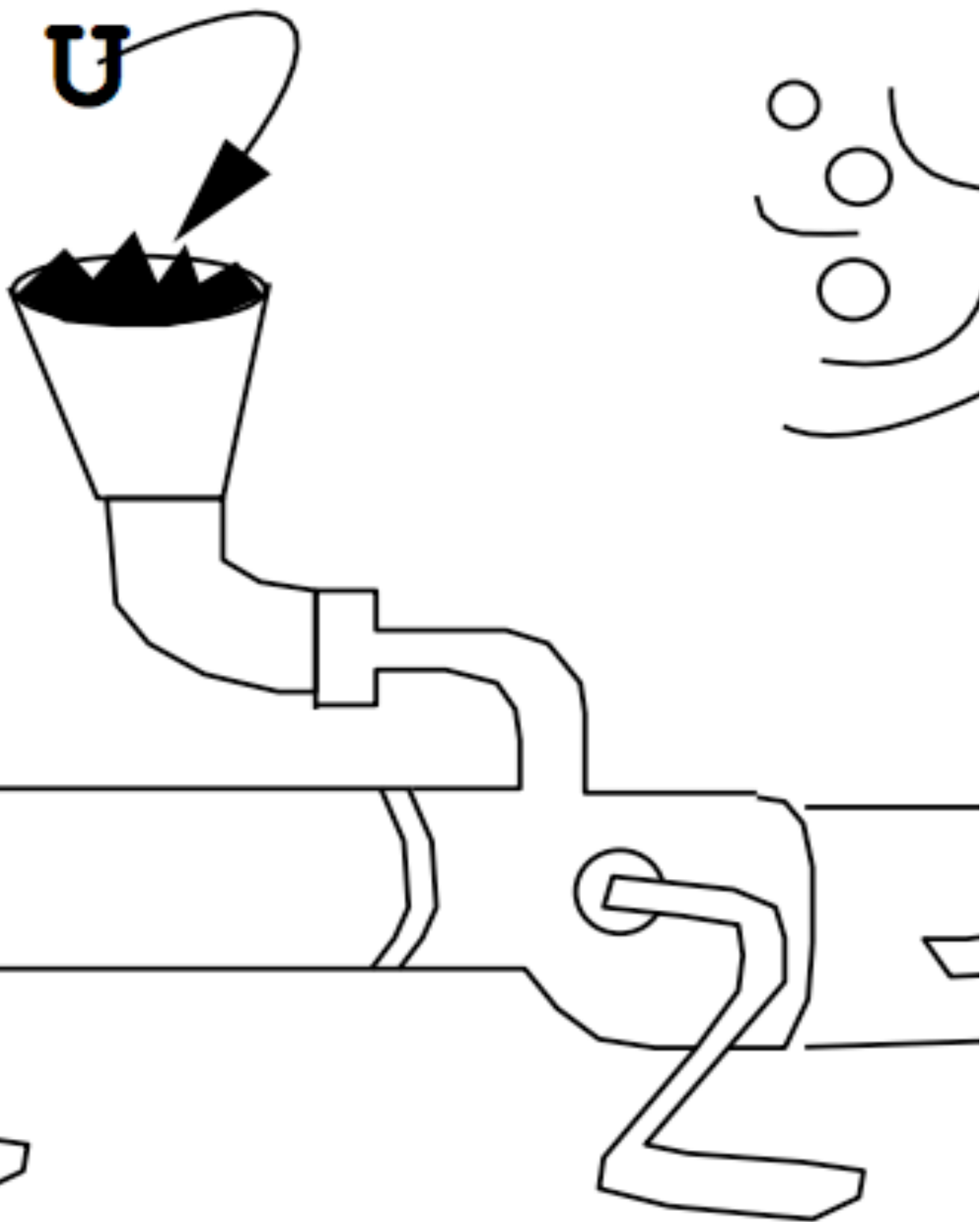
Visualize

Decision/Forecast Uncertainty
How do I assess the risk or error?



Sources and Types of Uncertainty

Visuali



Visualize

Decision/Forecast Uncertainty
How do I assess the risk or error?



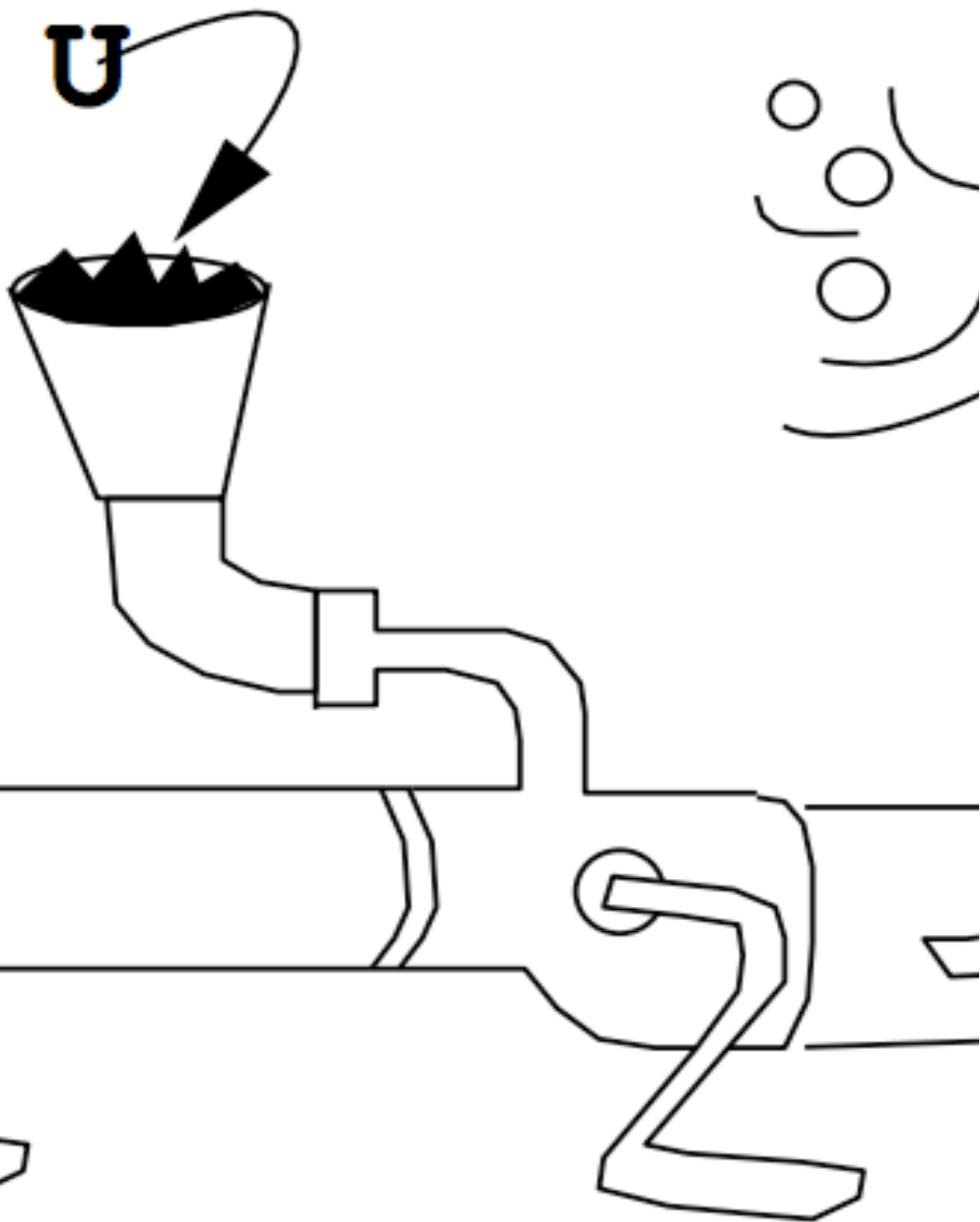
	Type I False Positive

Sources and Types of Uncertainty

Visuali



Decision/Forecast Uncertainty
How do I assess the risk or error?



Visualize



**Type I
False
Positive**

**Type II
False
Negative**



Uncertainty

What does it mean?

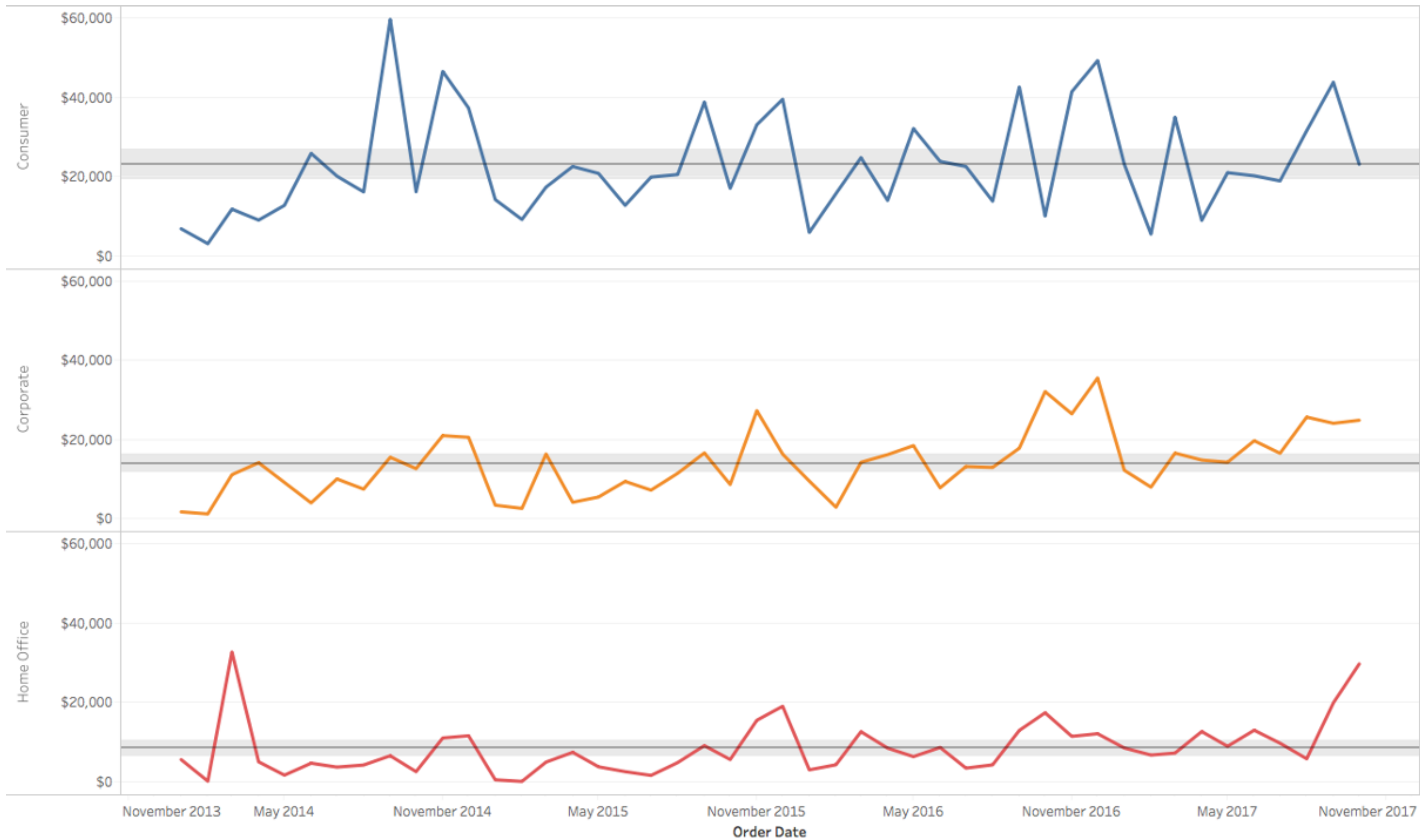
Lots of things!

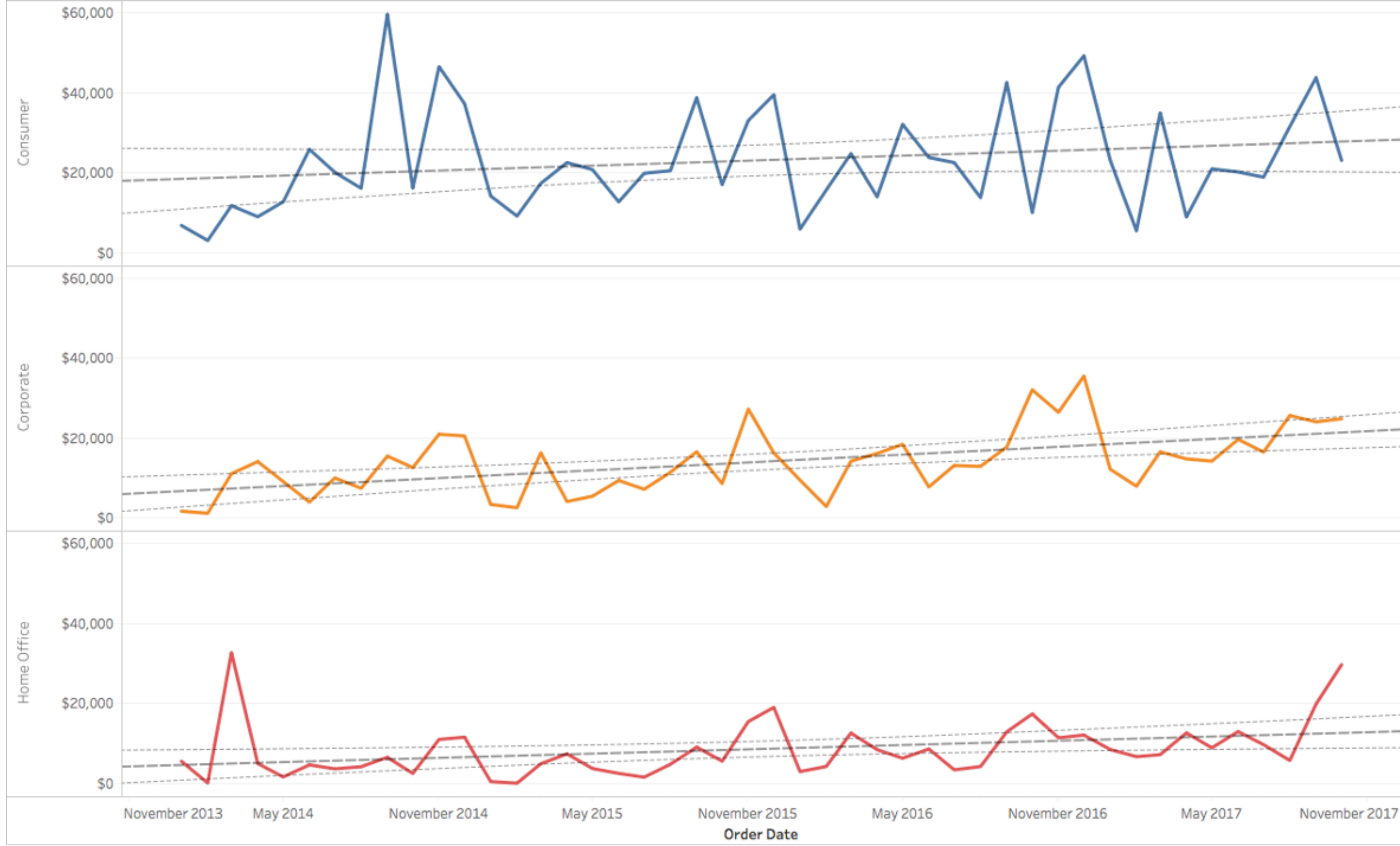
How should I visualize it?

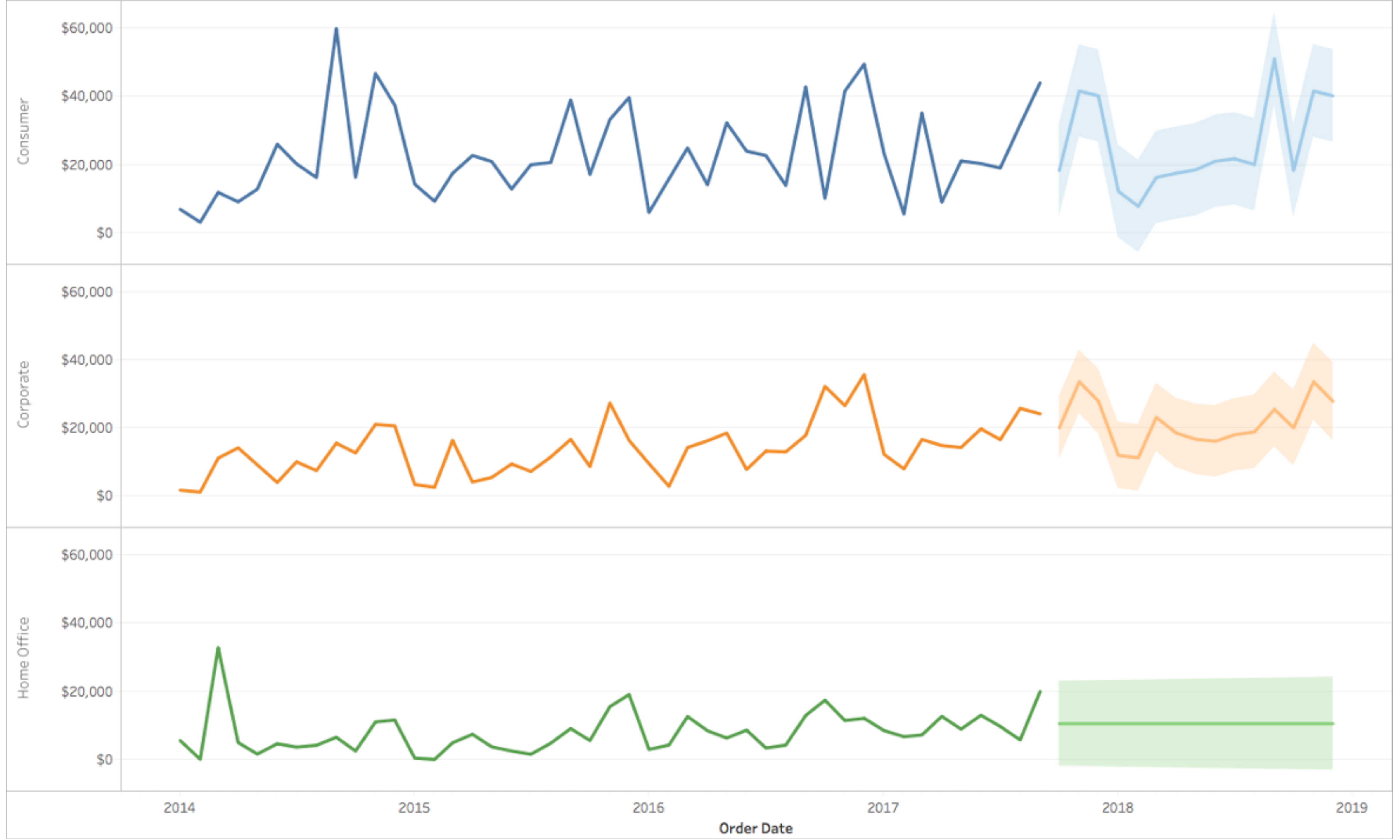
Uncertainty

What does it mean?

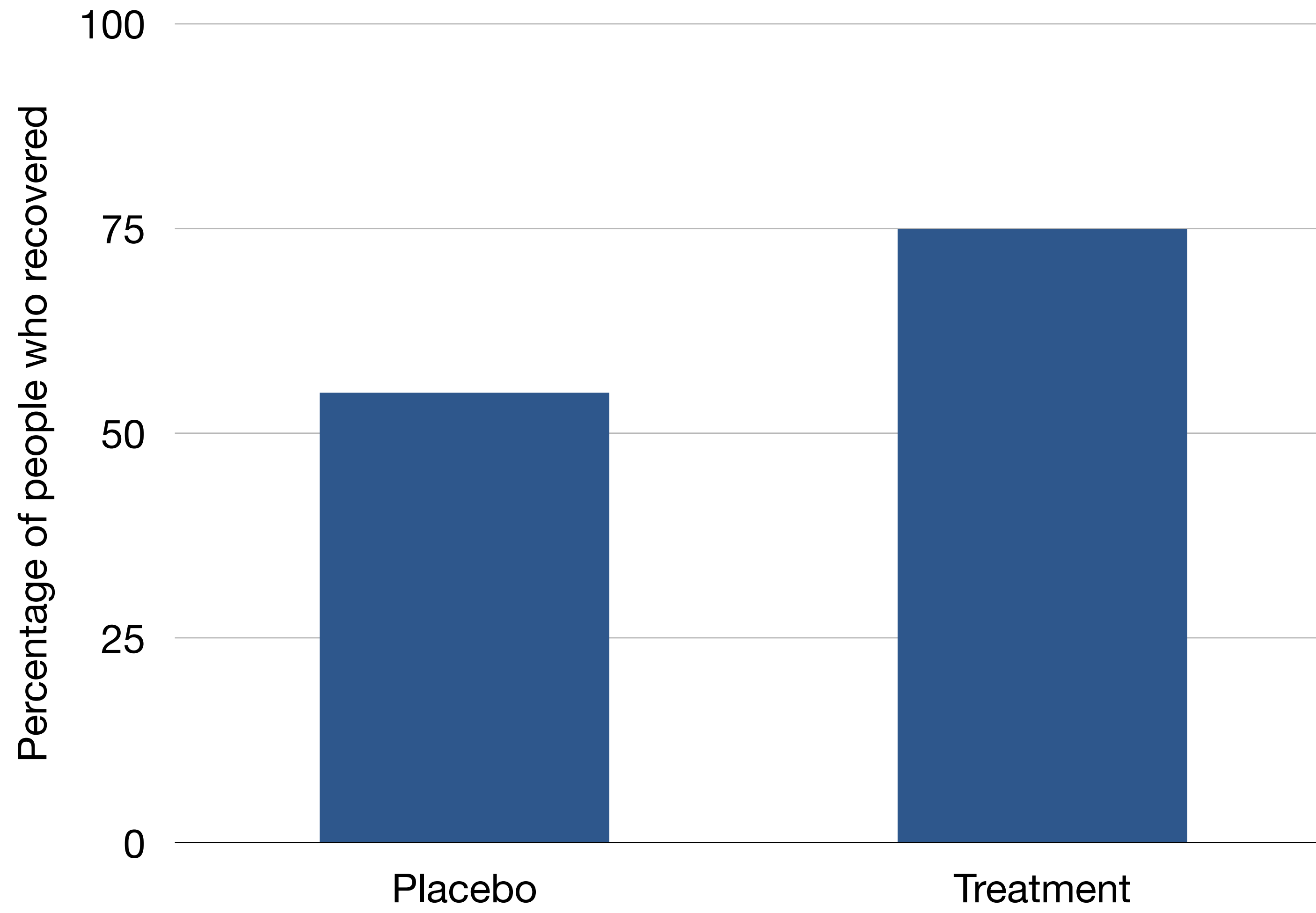
How should I visualize it?



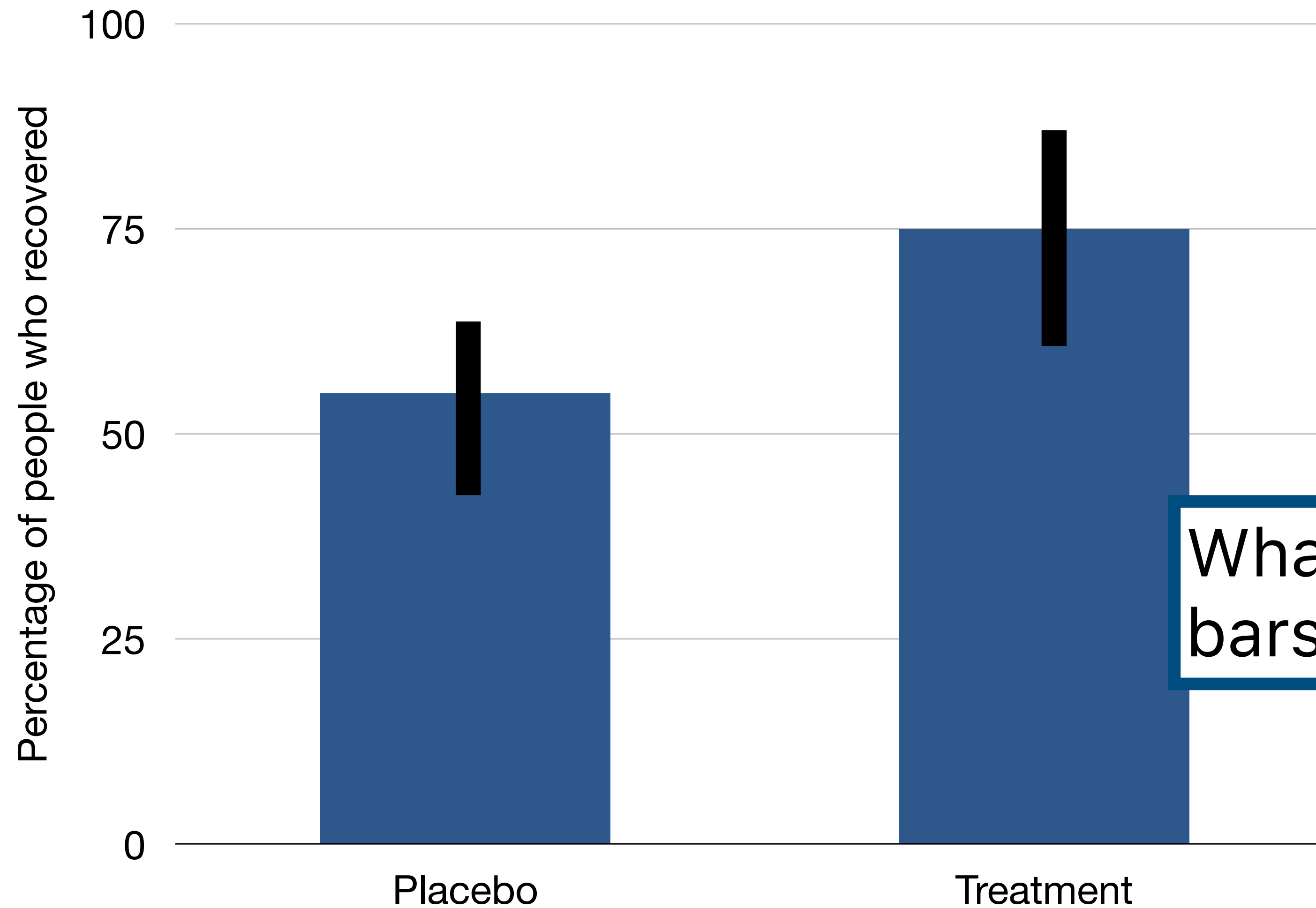




Trial of new medicine



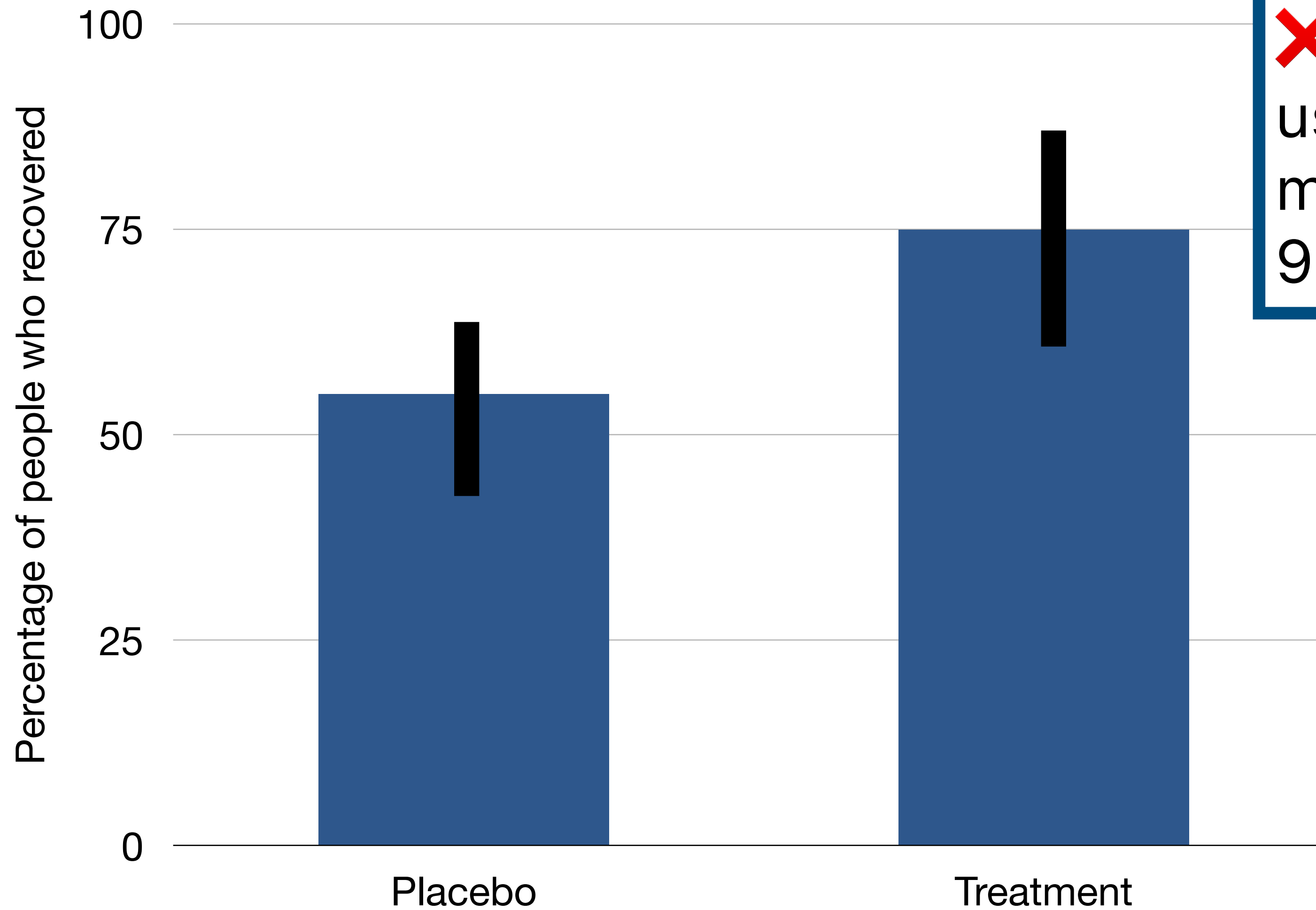
Trial of new medicine



What do you think these error bars are implying?

tryclassbuzz.com
Code: **errorbars**

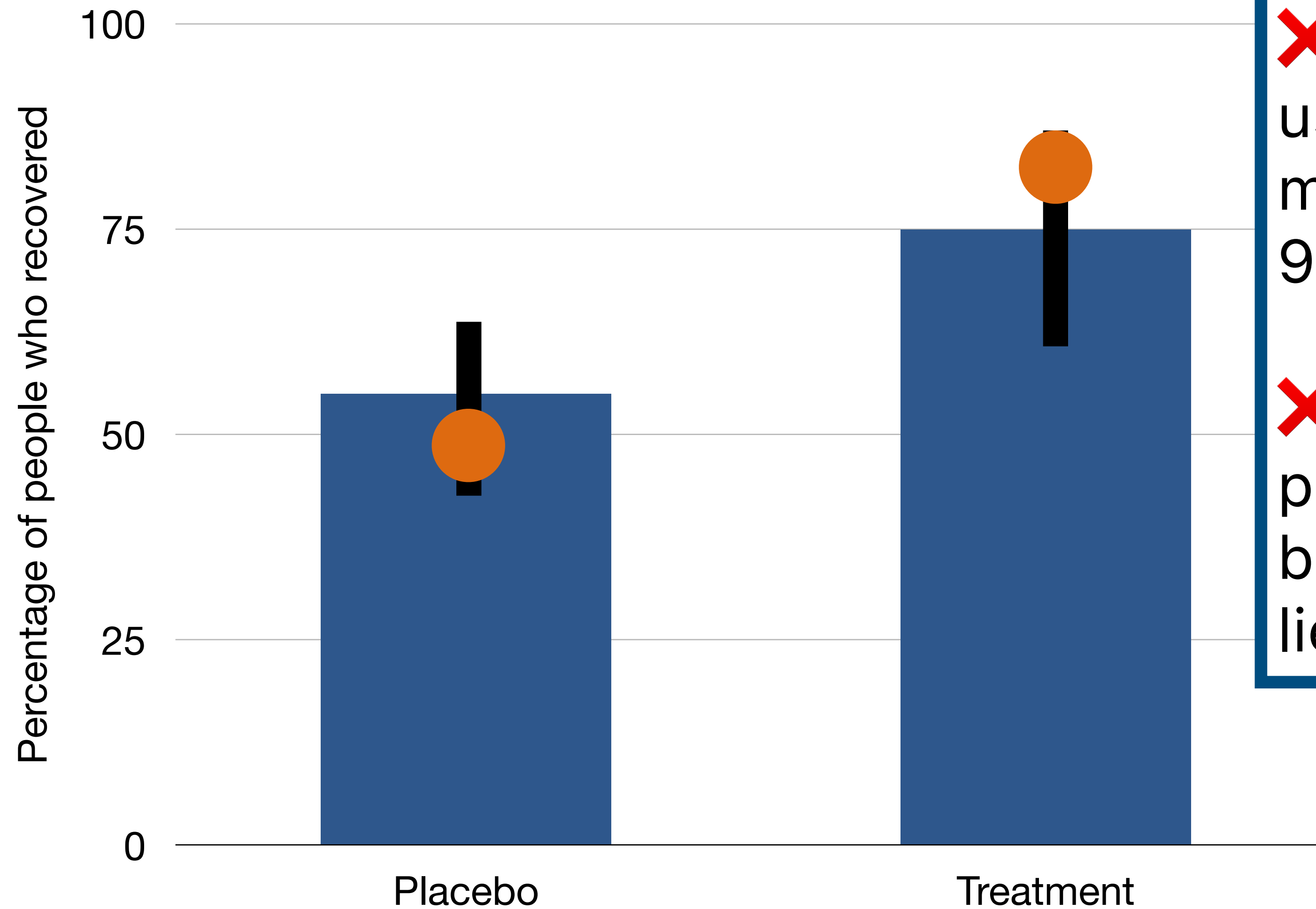
Trial of new medicine



Expressiveness?

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

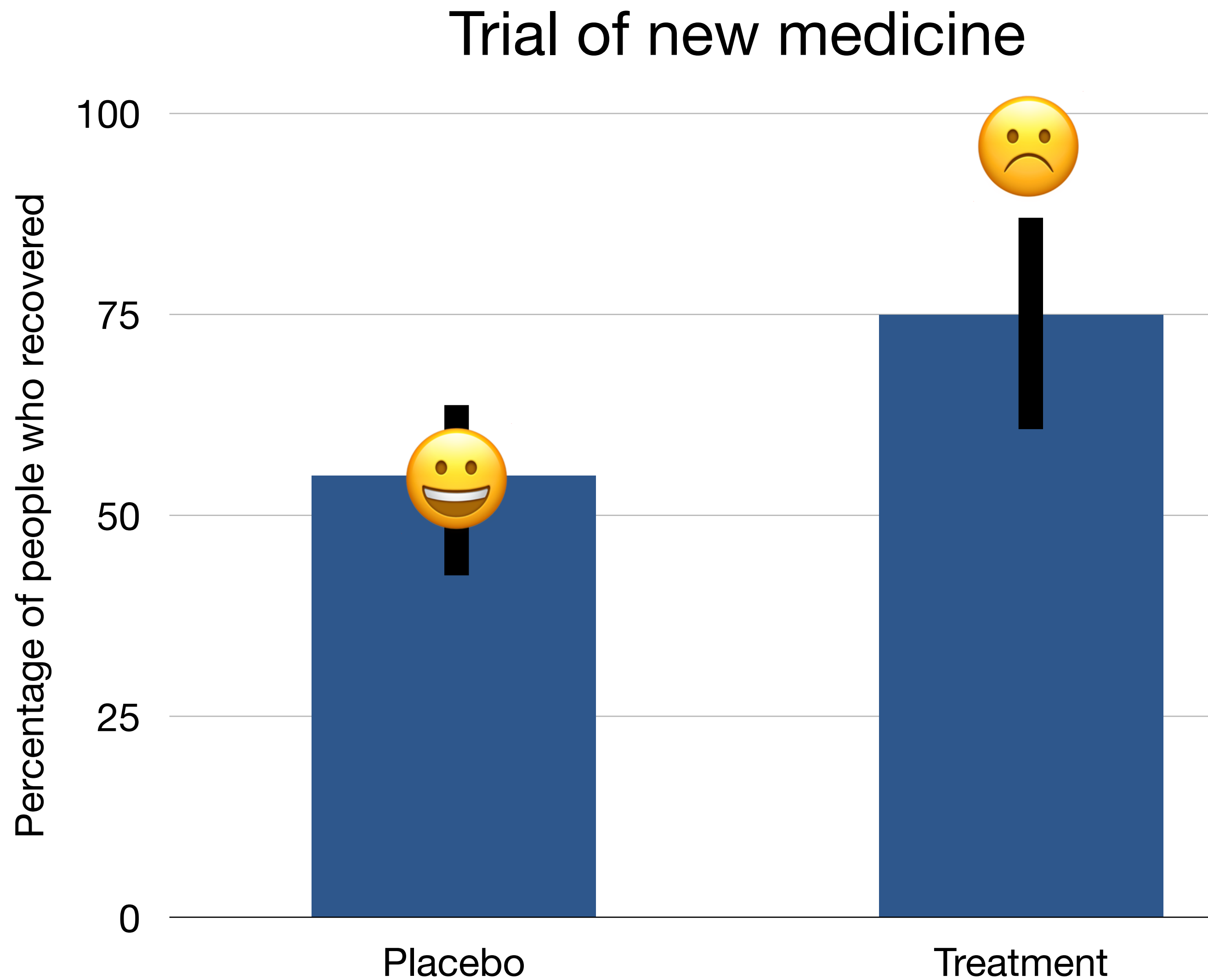
Trial of new medicine



Expressiveness?

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

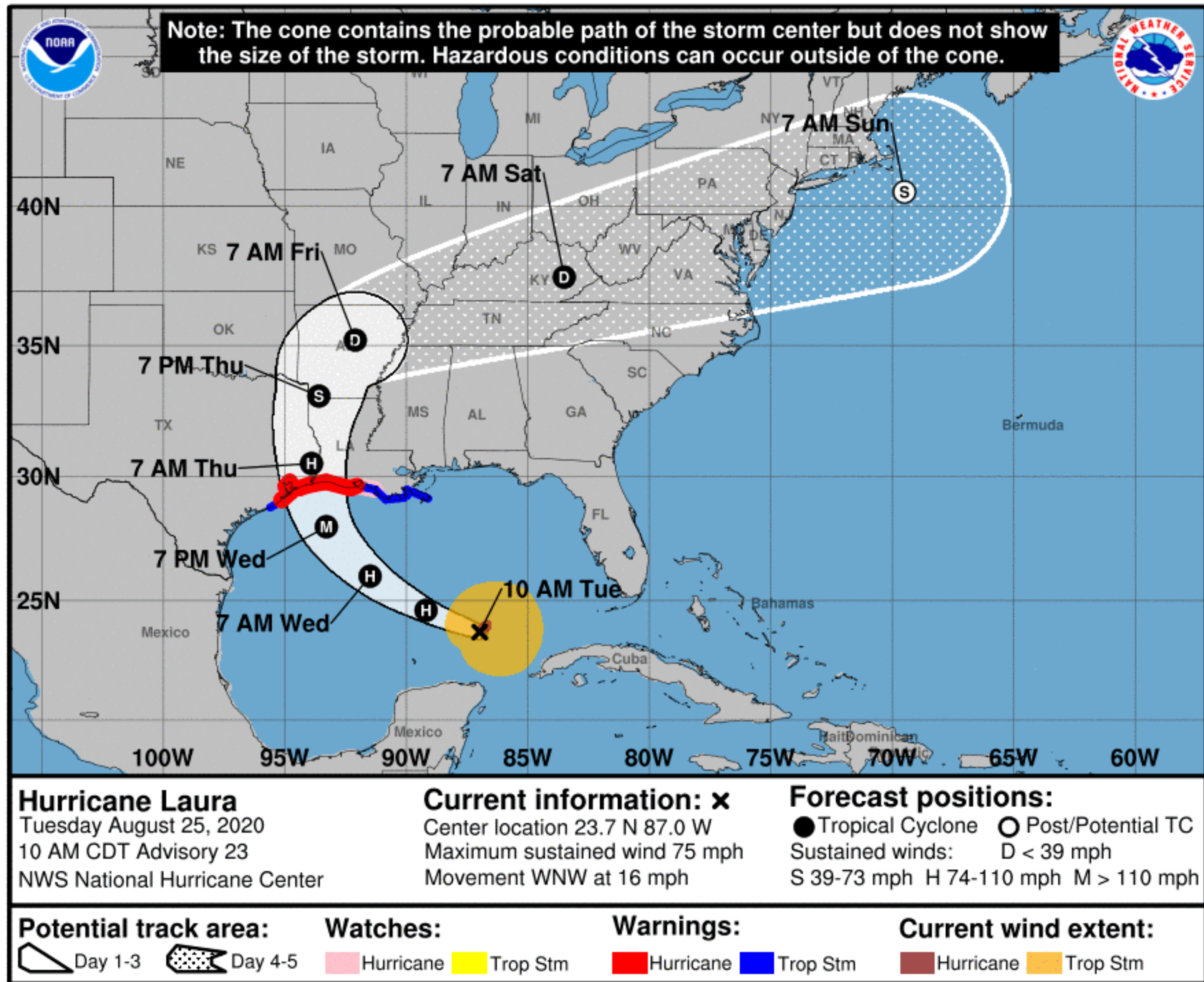
✗ Within-the-bar bias: people perceive points falling within the bar as more likely than those that lie outside.



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

Expressiveness?

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

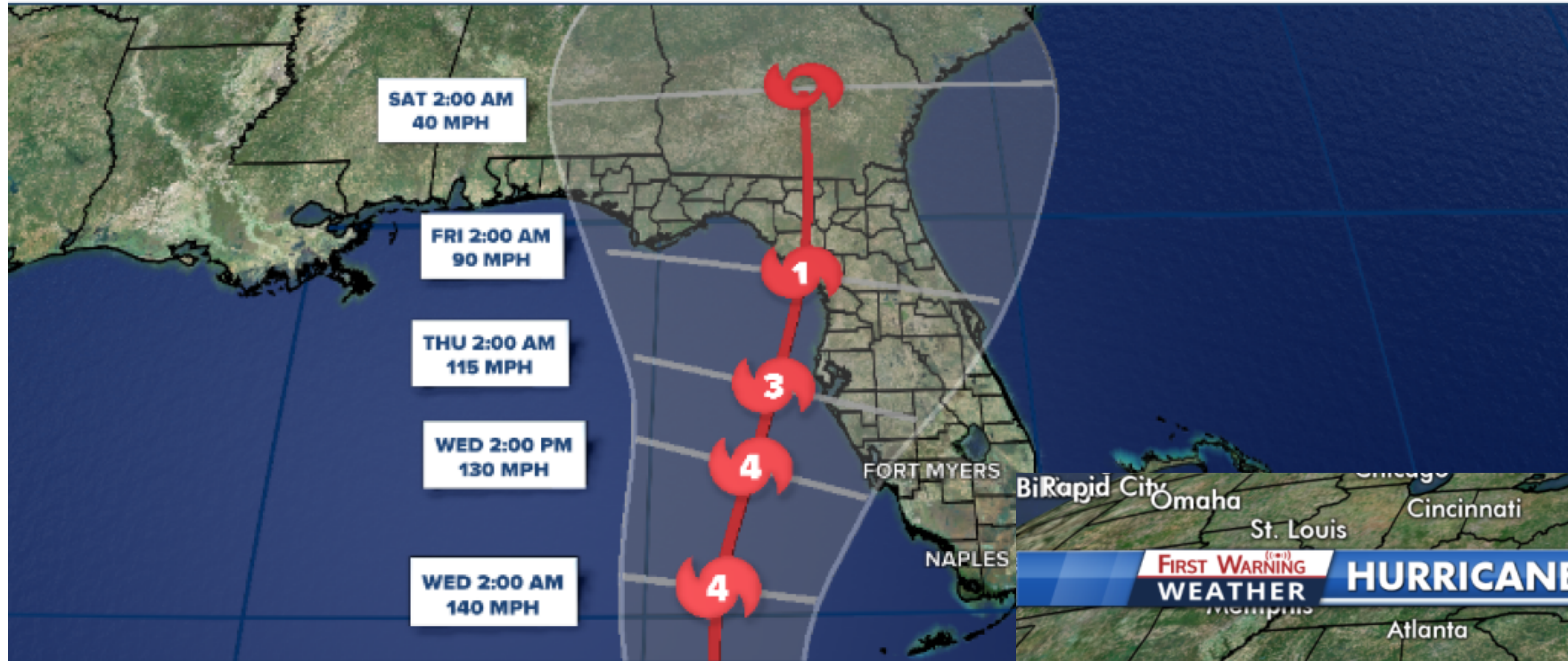


What is being visualized?

What are the strengths and weaknesses of this visualization?

FOX4 HURRICANE IAN

5:00 AM ADVISORY

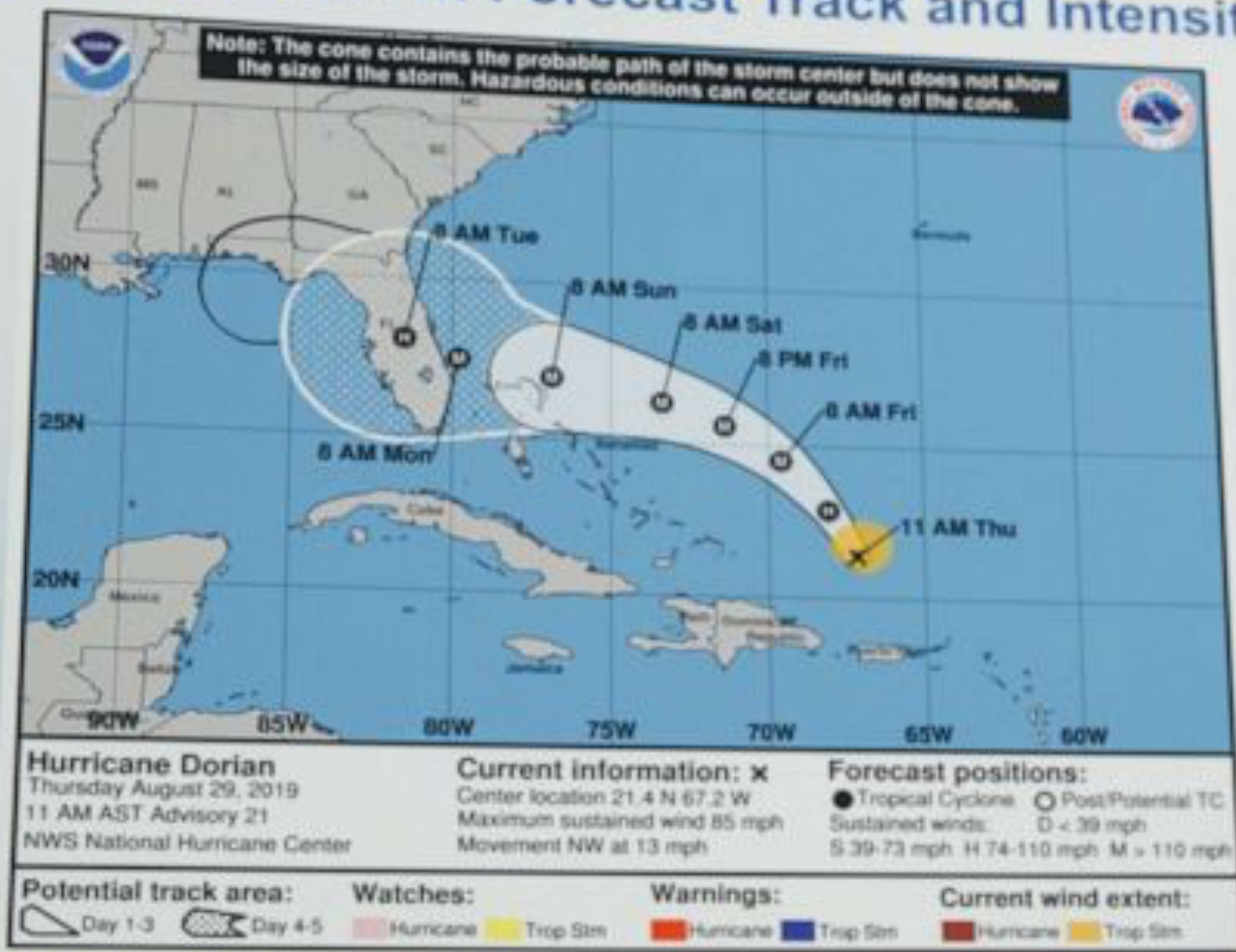


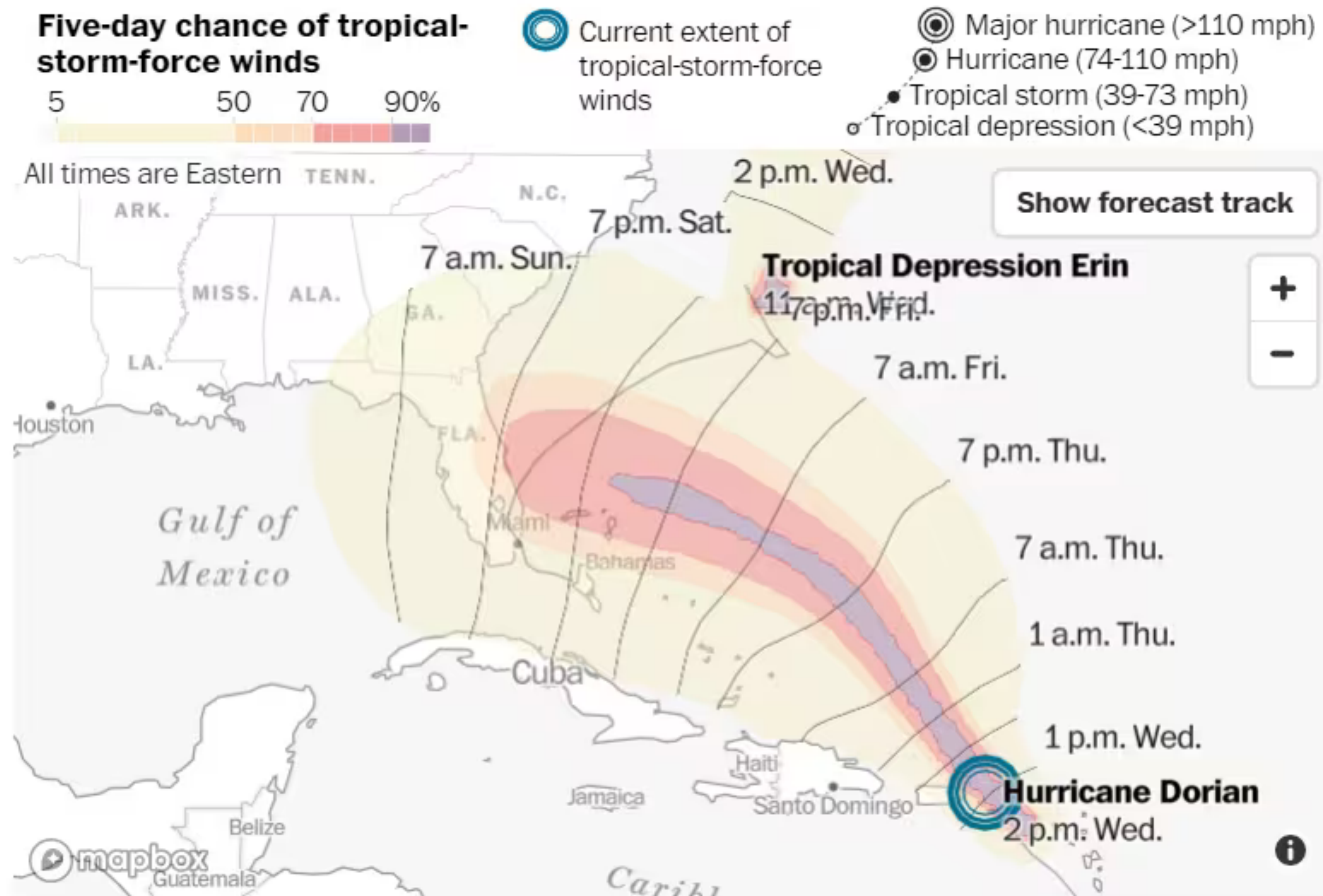
CURRENT LOCATION: 90 MI SW OF G





Hurricane Dorian Forecast Track and Intensity

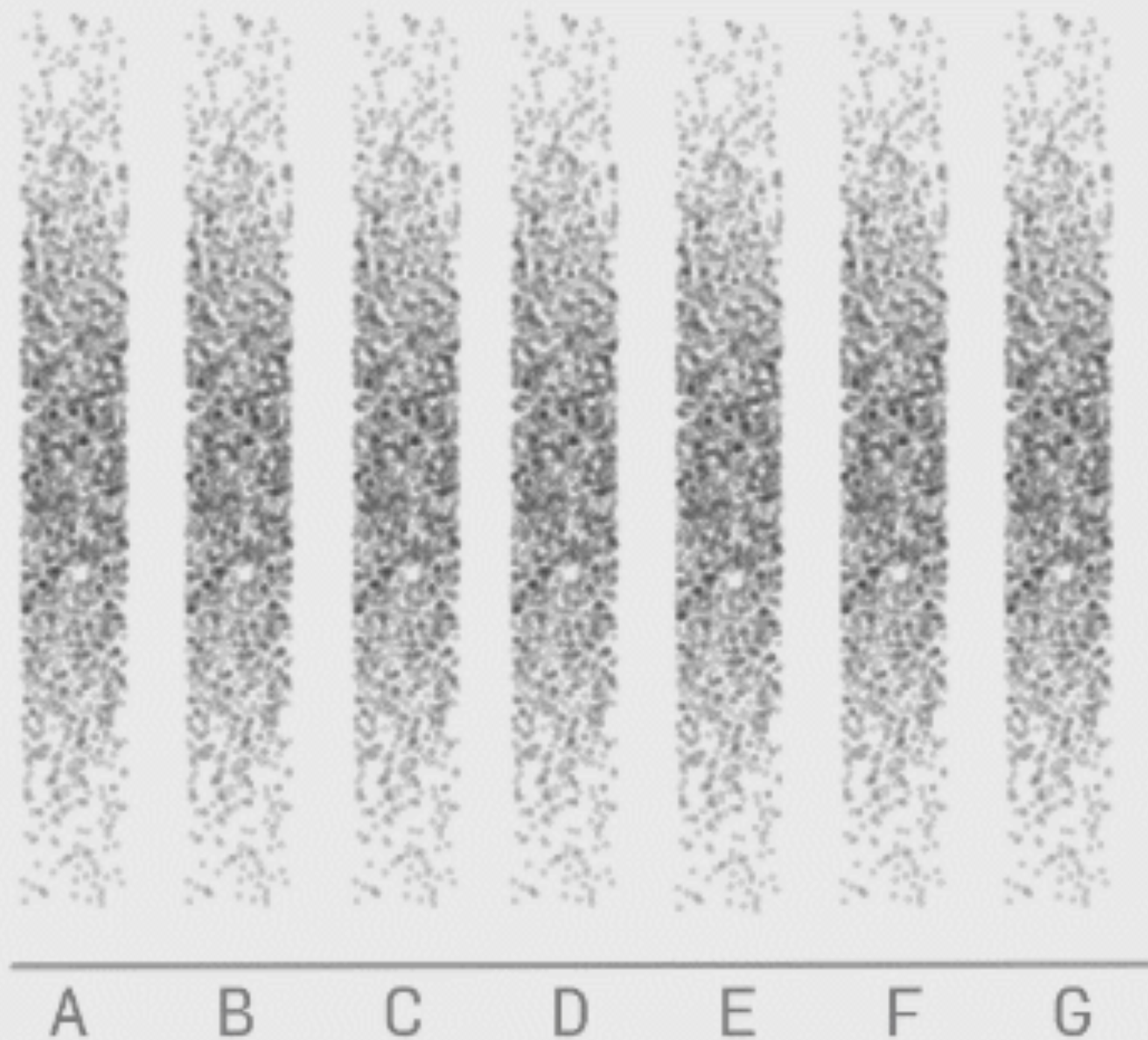




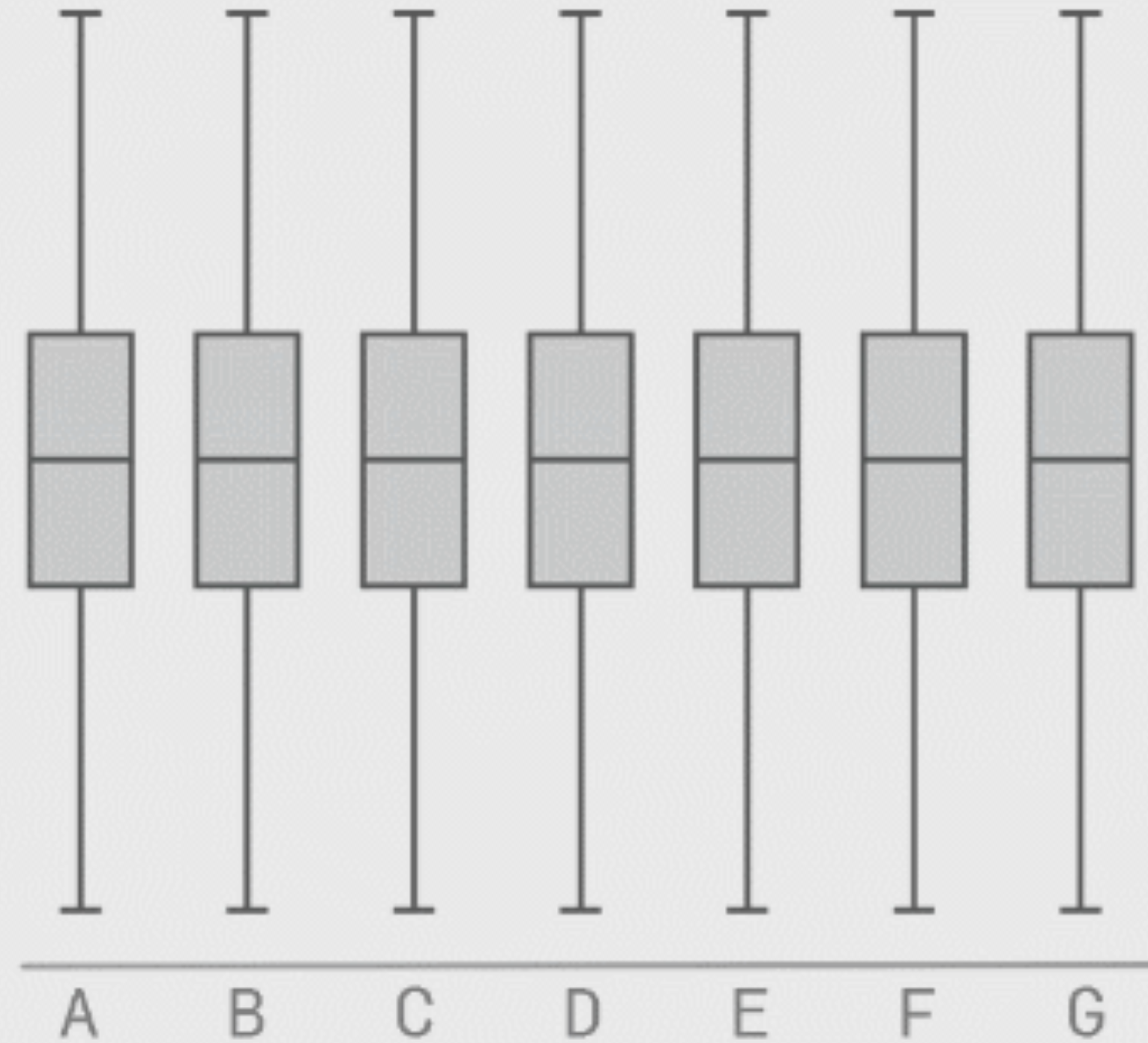
Source: National Weather Service. Note: Impact lines represent the earliest reasonable arrival time of tropical-storm-force winds.

For uncertainty, use **visual variables** instead of visualizing point estimates

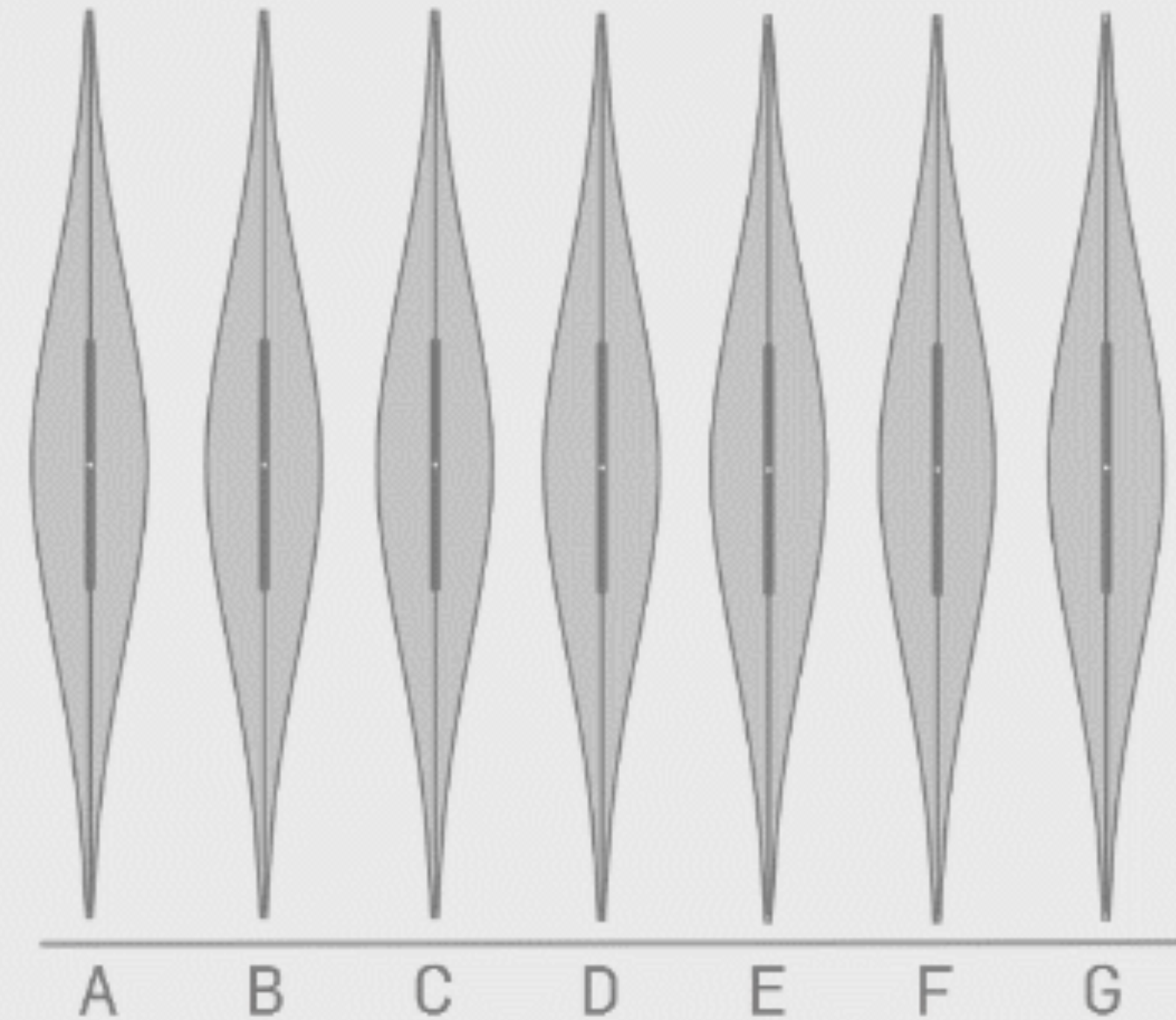
Raw Data



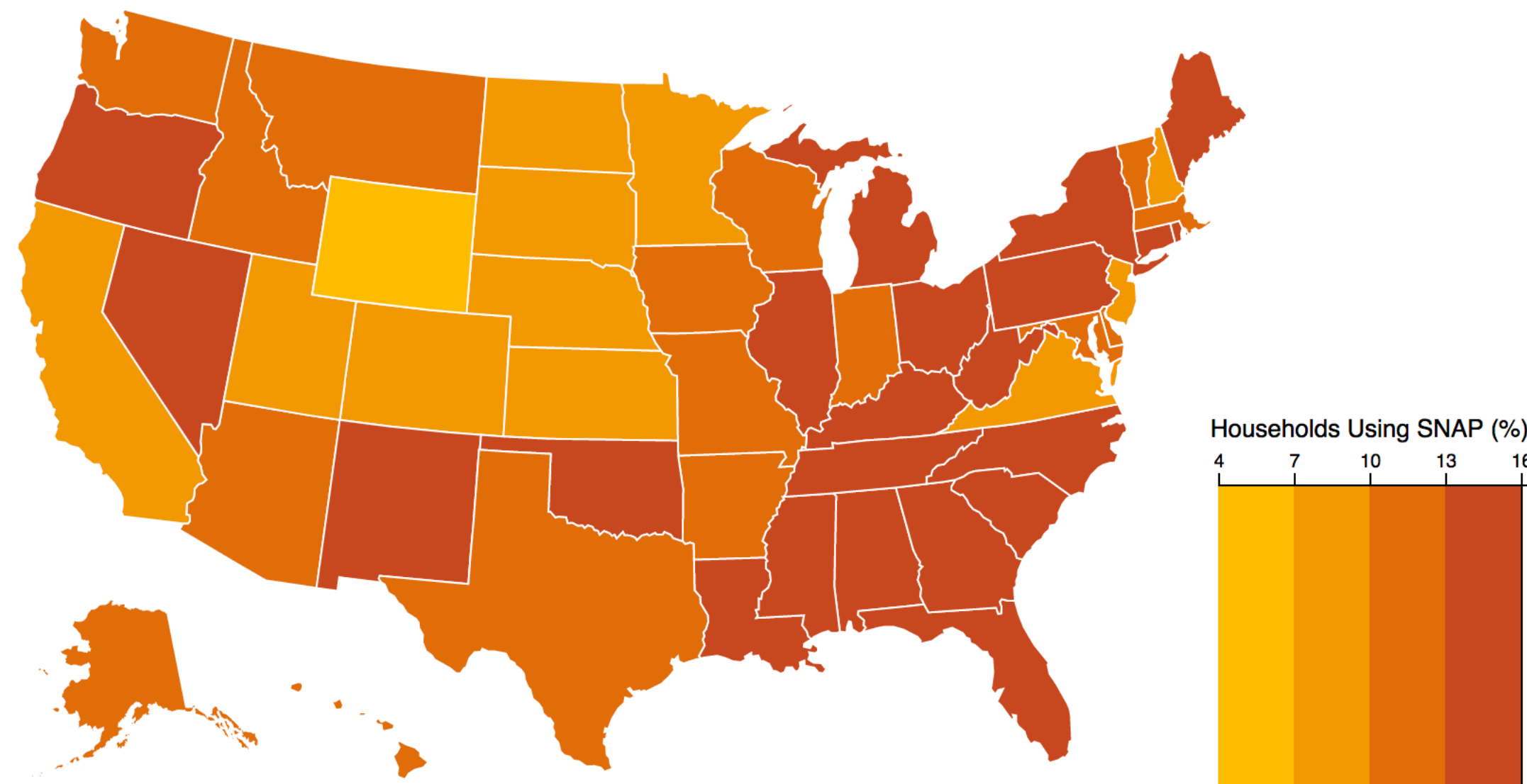
Box-plot of the Data



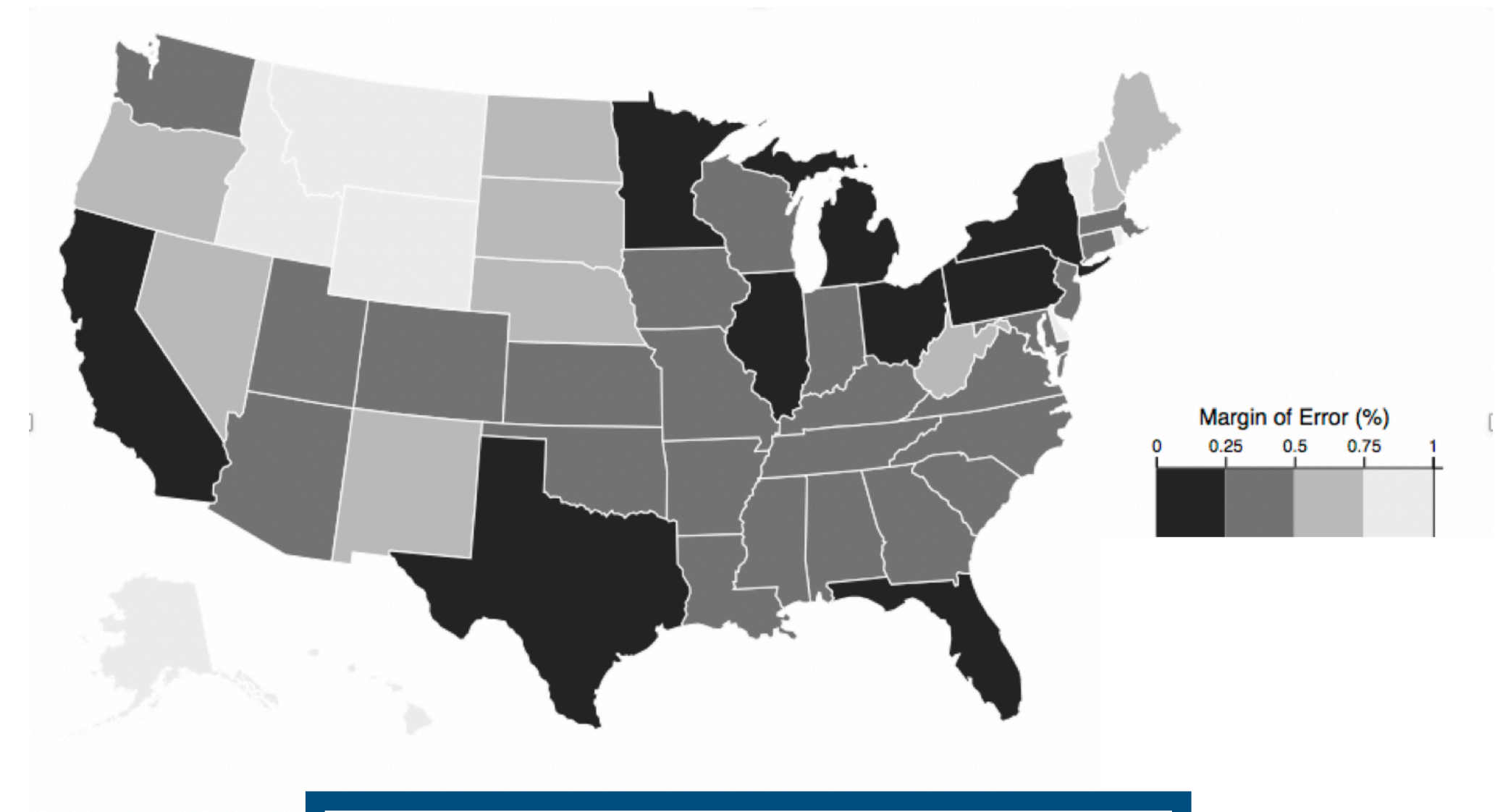
Violin-plot of the Data



For uncertainty, use **visual variables** instead of visualizing point estimates

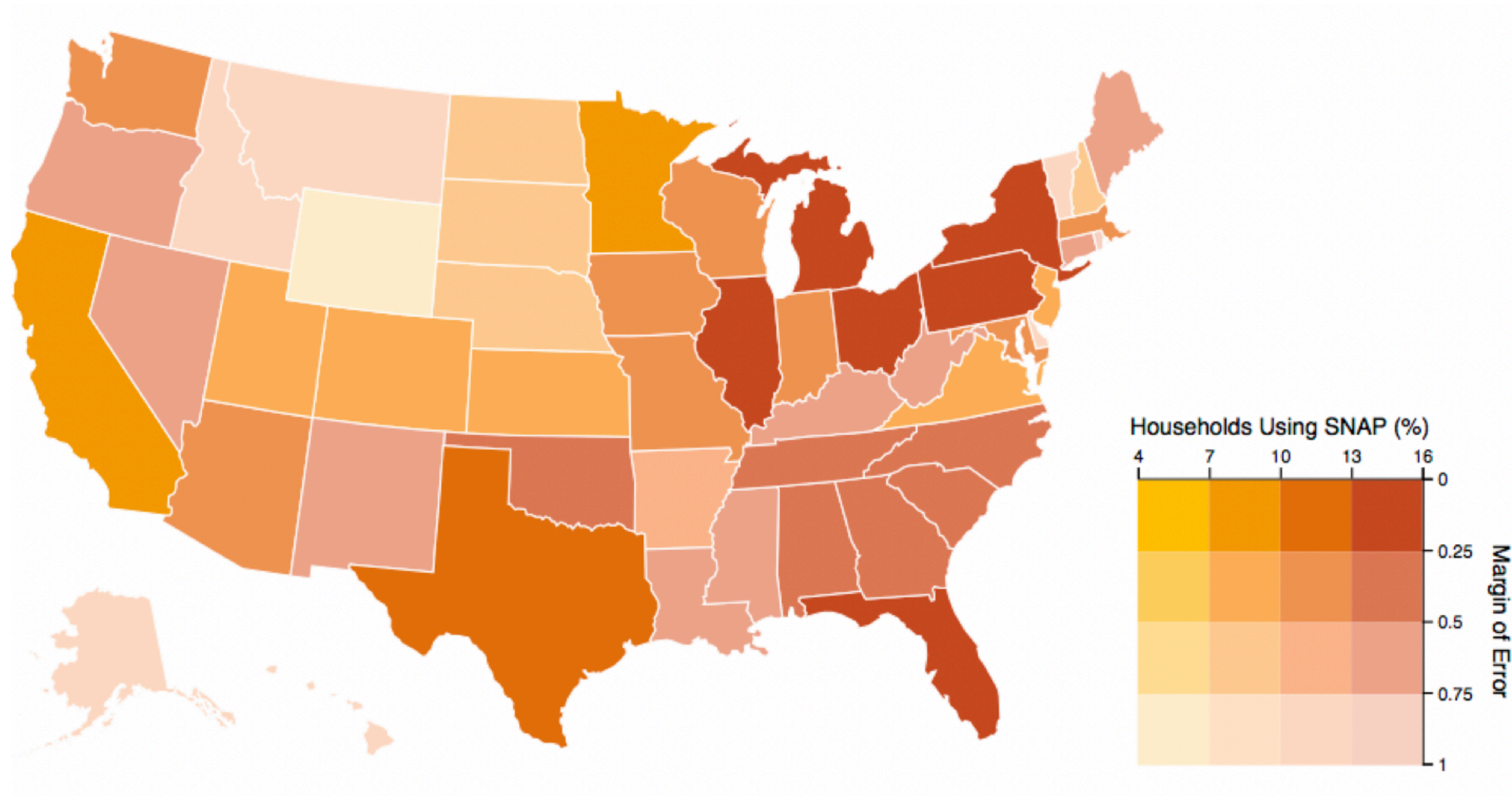


Data Map

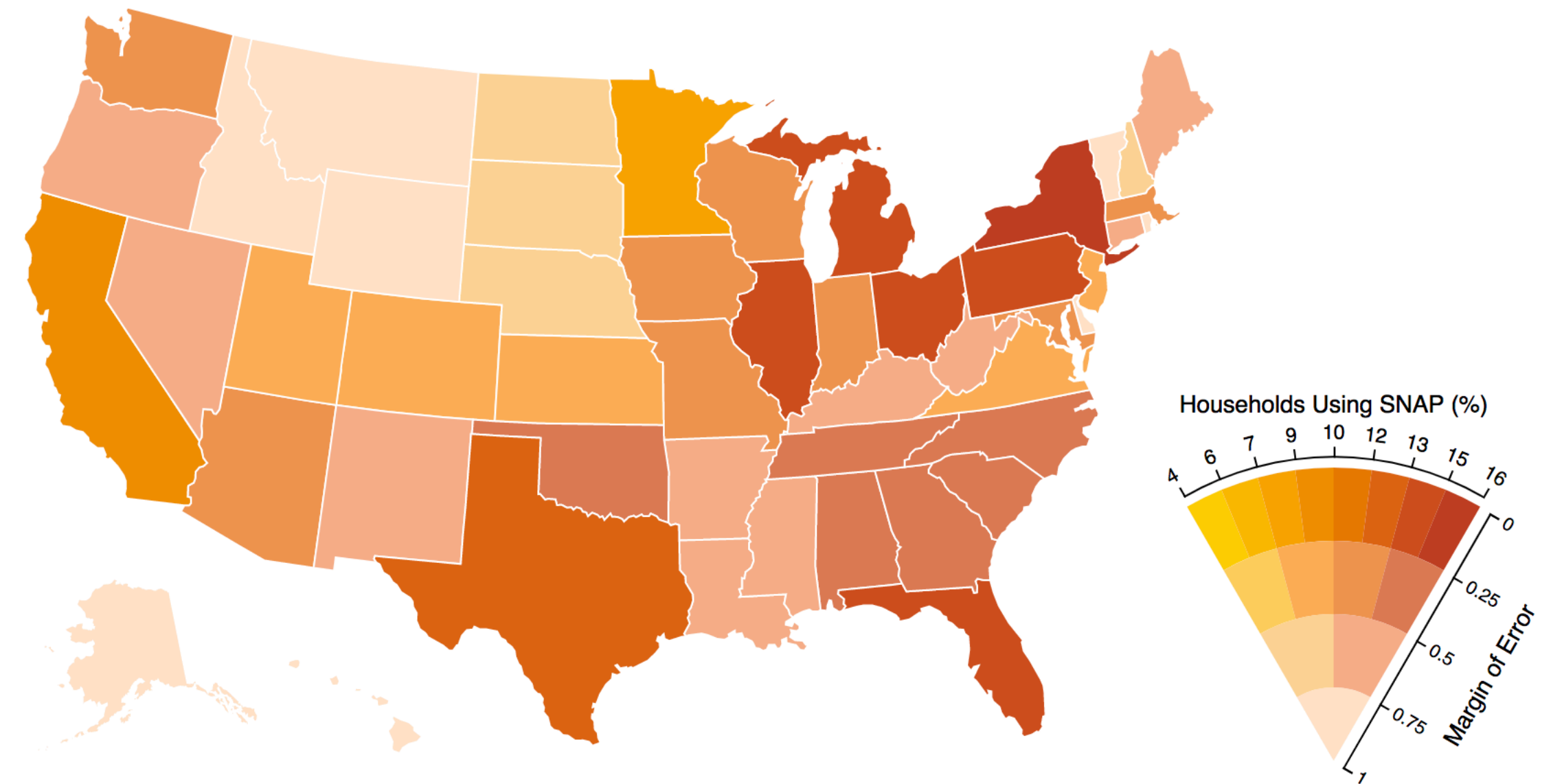


Uncertainty Map

For uncertainty, use **visual variables** instead of visualizing point estimates

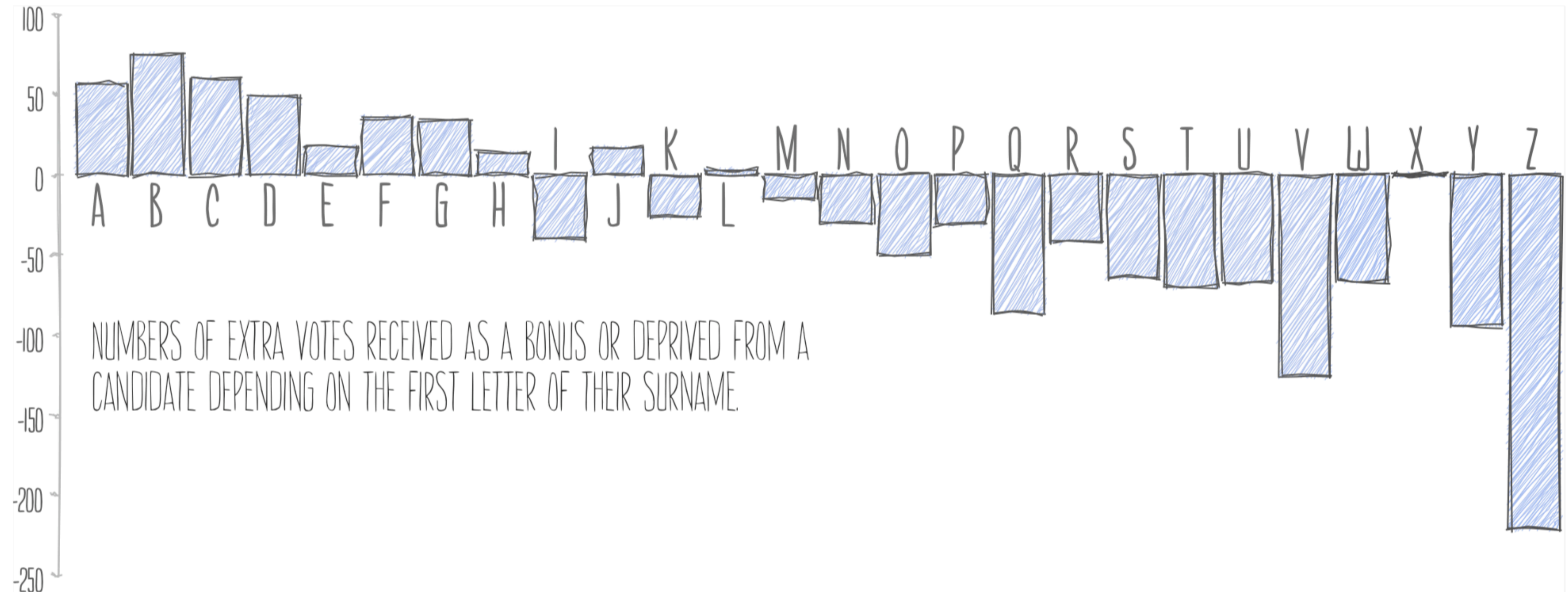


Bivariate Map (Data + Uncertainty)

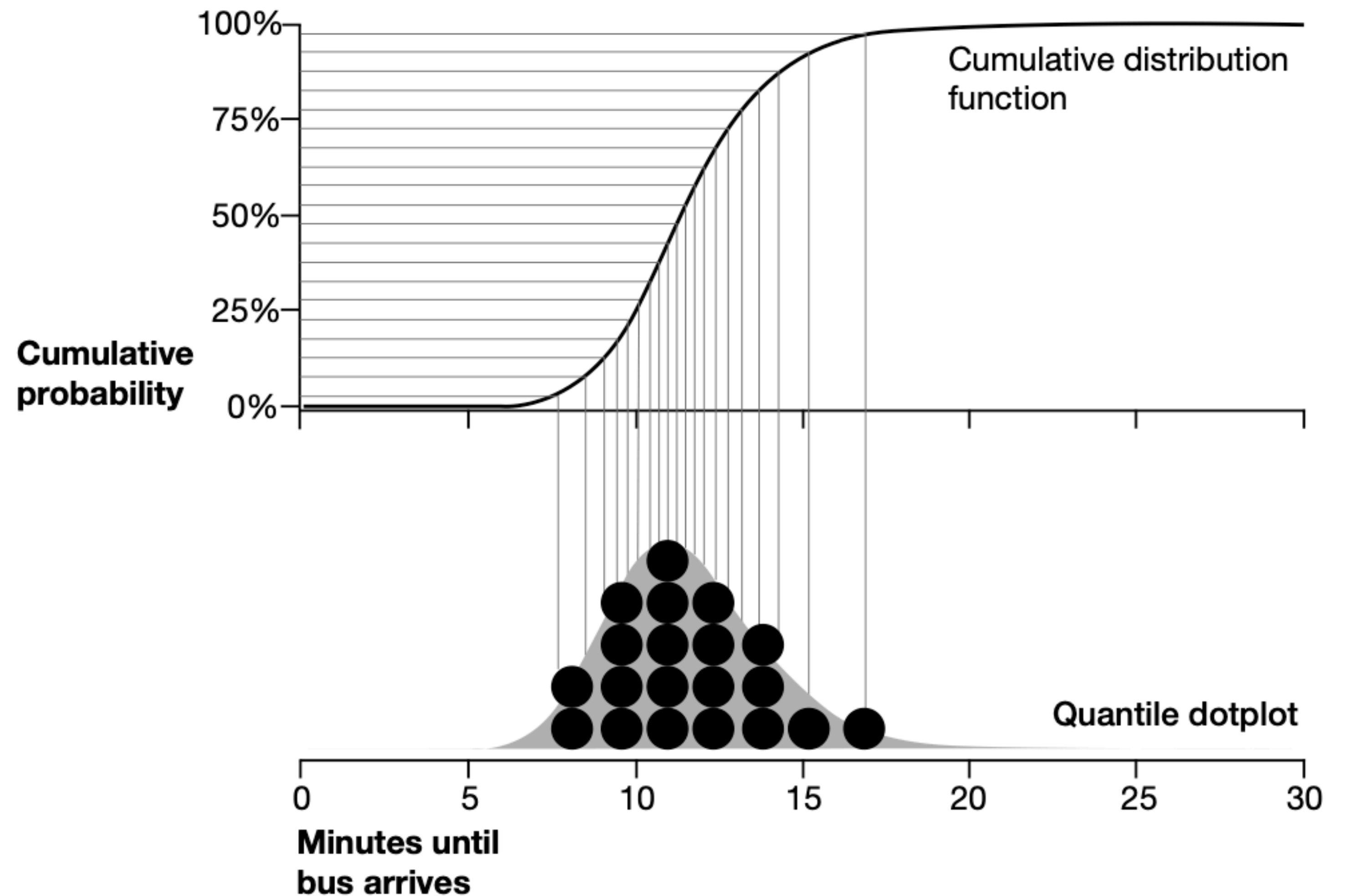
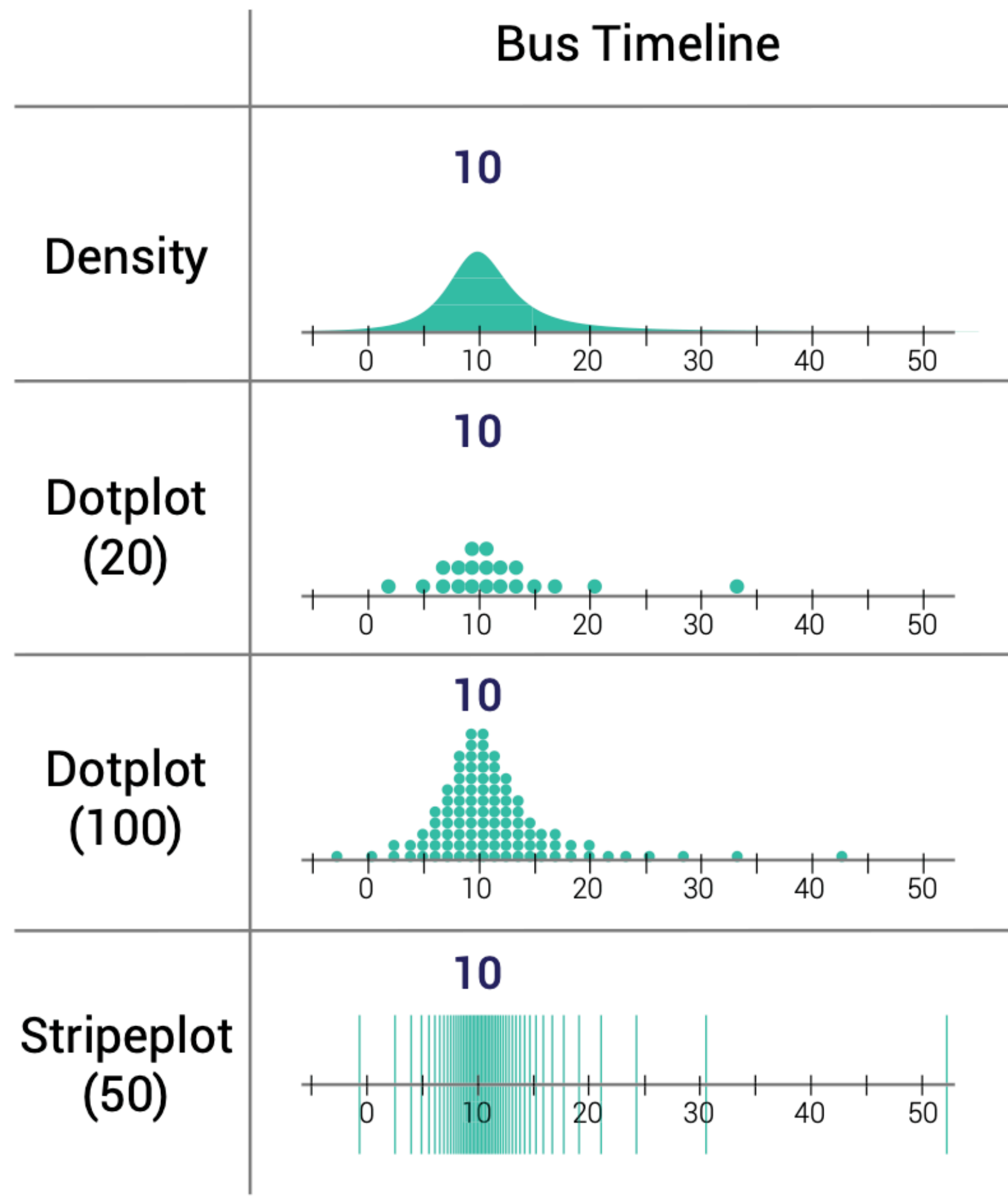


Value-Suppressing Uncertainty Map

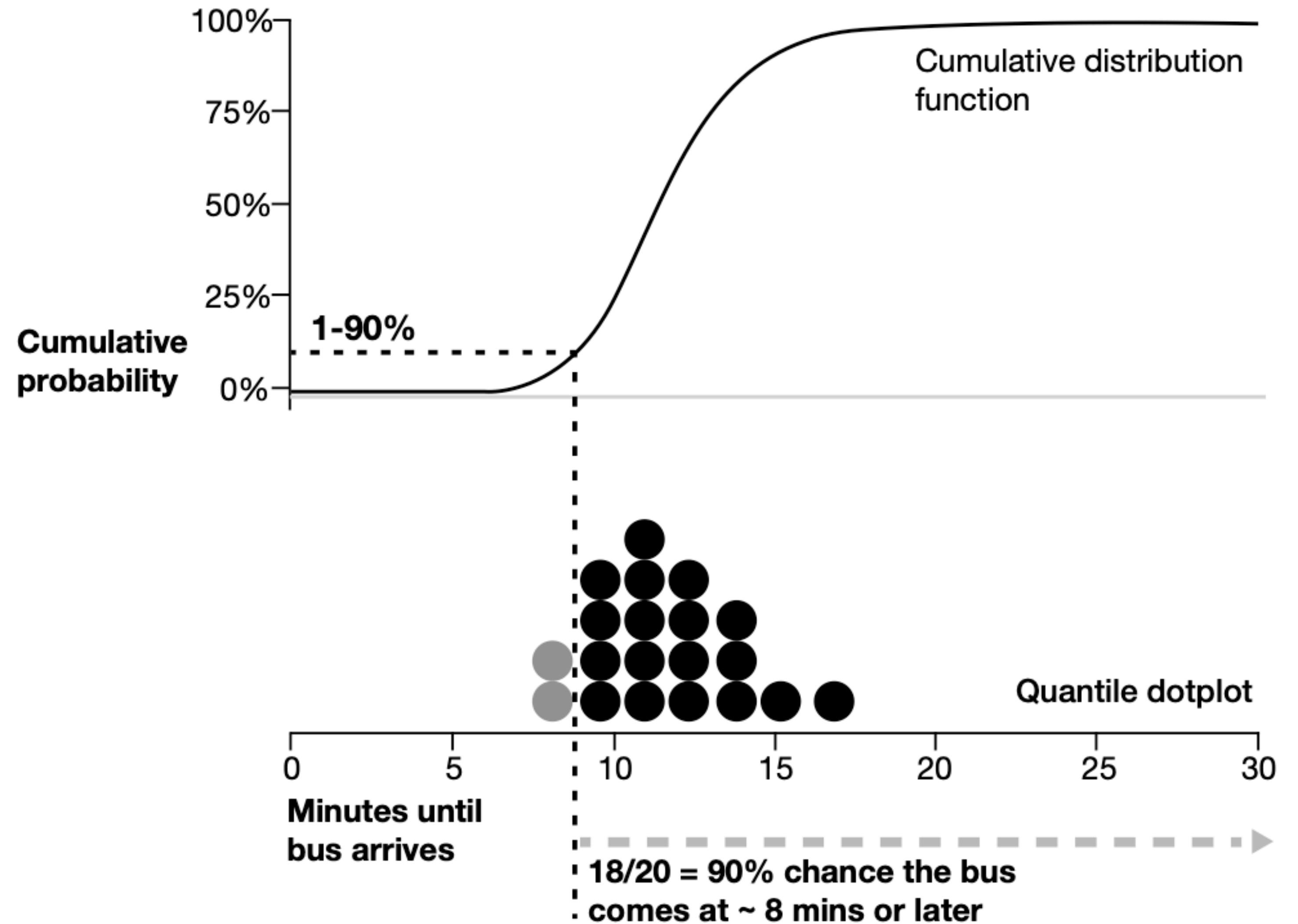
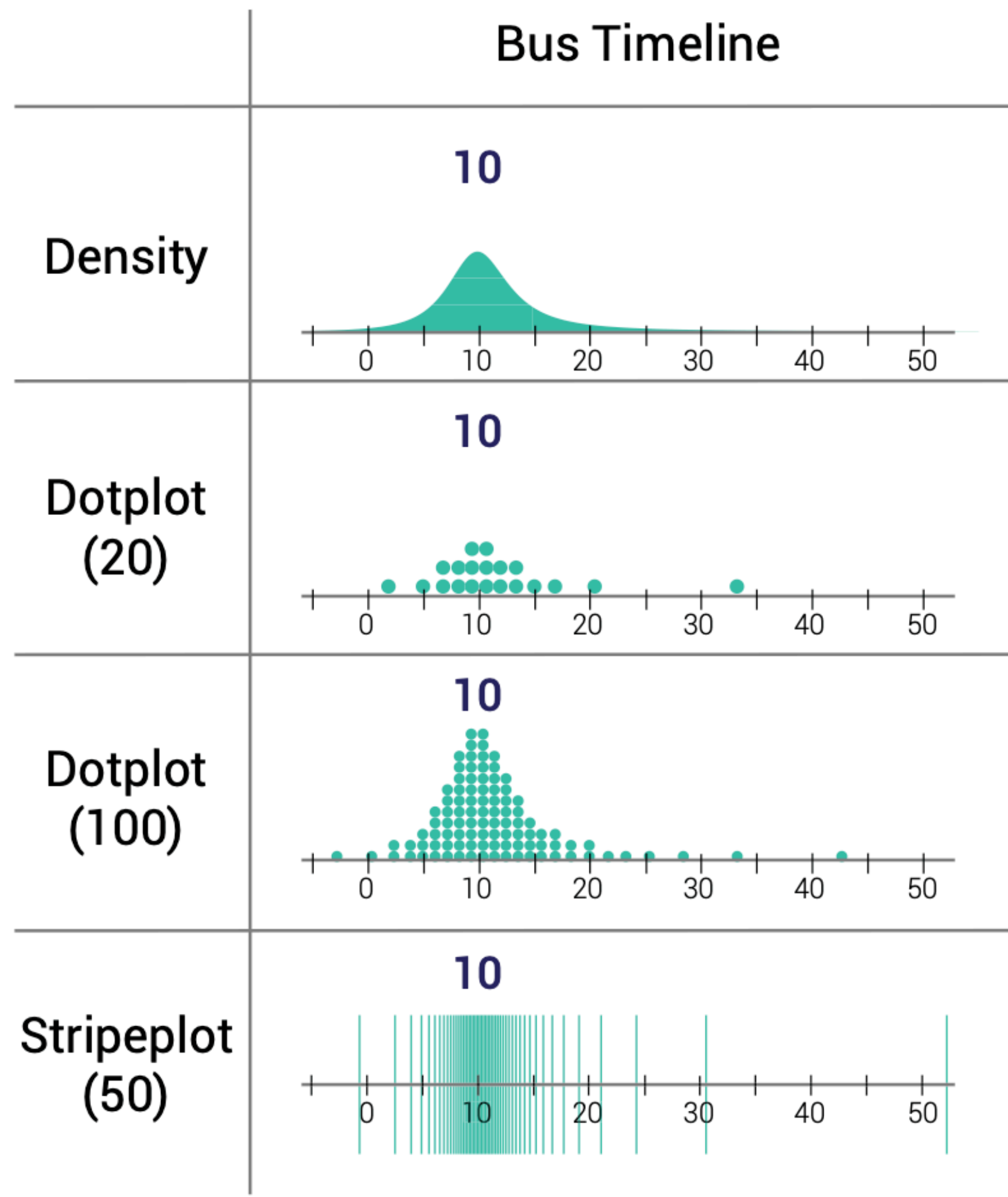
For uncertainty, use **visual variables** instead of visualizing point estimates



"Set of draws" technique

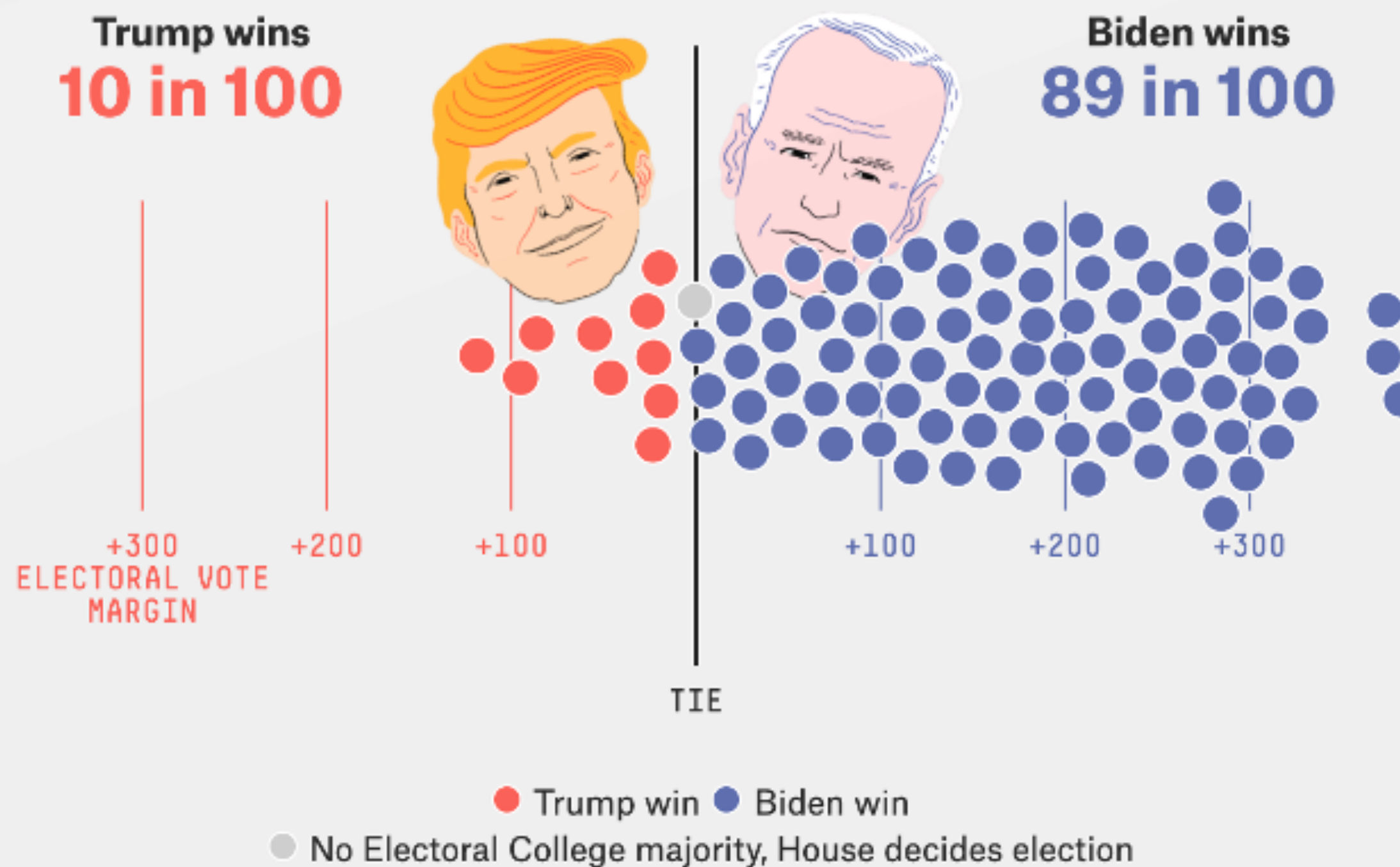


"Set of draws" technique



Biden is *avored* 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.



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

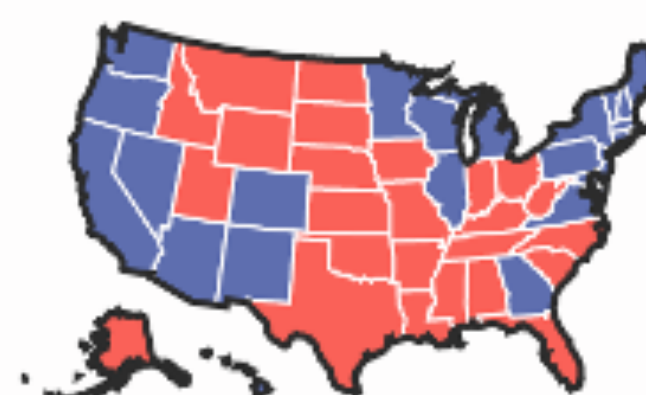
Biden wins
89 in 100

+300
ELECTORAL VOTE
MARGIN

+200

+100

0 +300



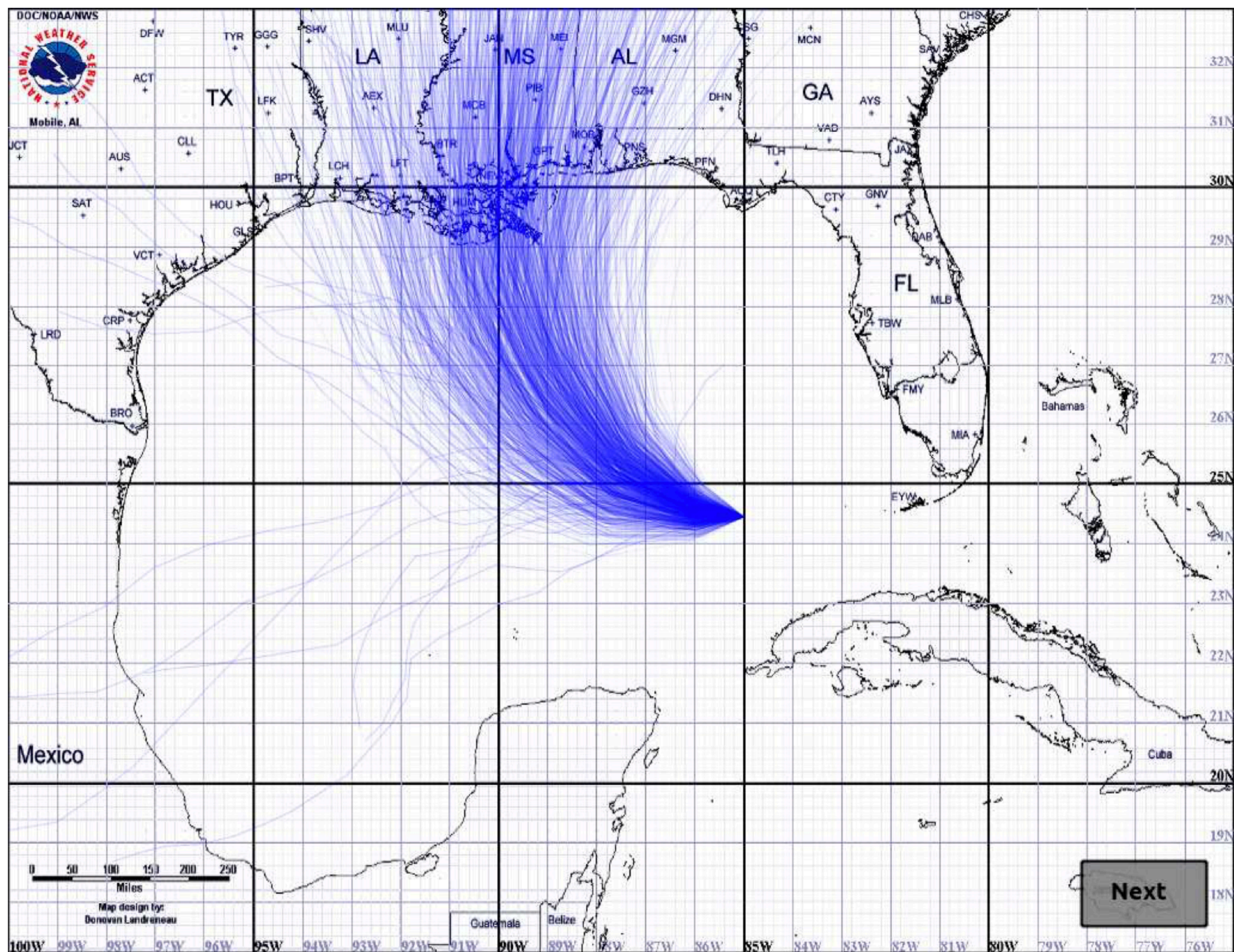
— ELECTORAL VOTES —

Biden ✓ 306
Trump 232

Don't count the
underdog out!
Upset wins are
surprising but not
impossible.

● Trump
● No Electoral College

tion



• **TheUpshot**

STATISTICAL NOISE

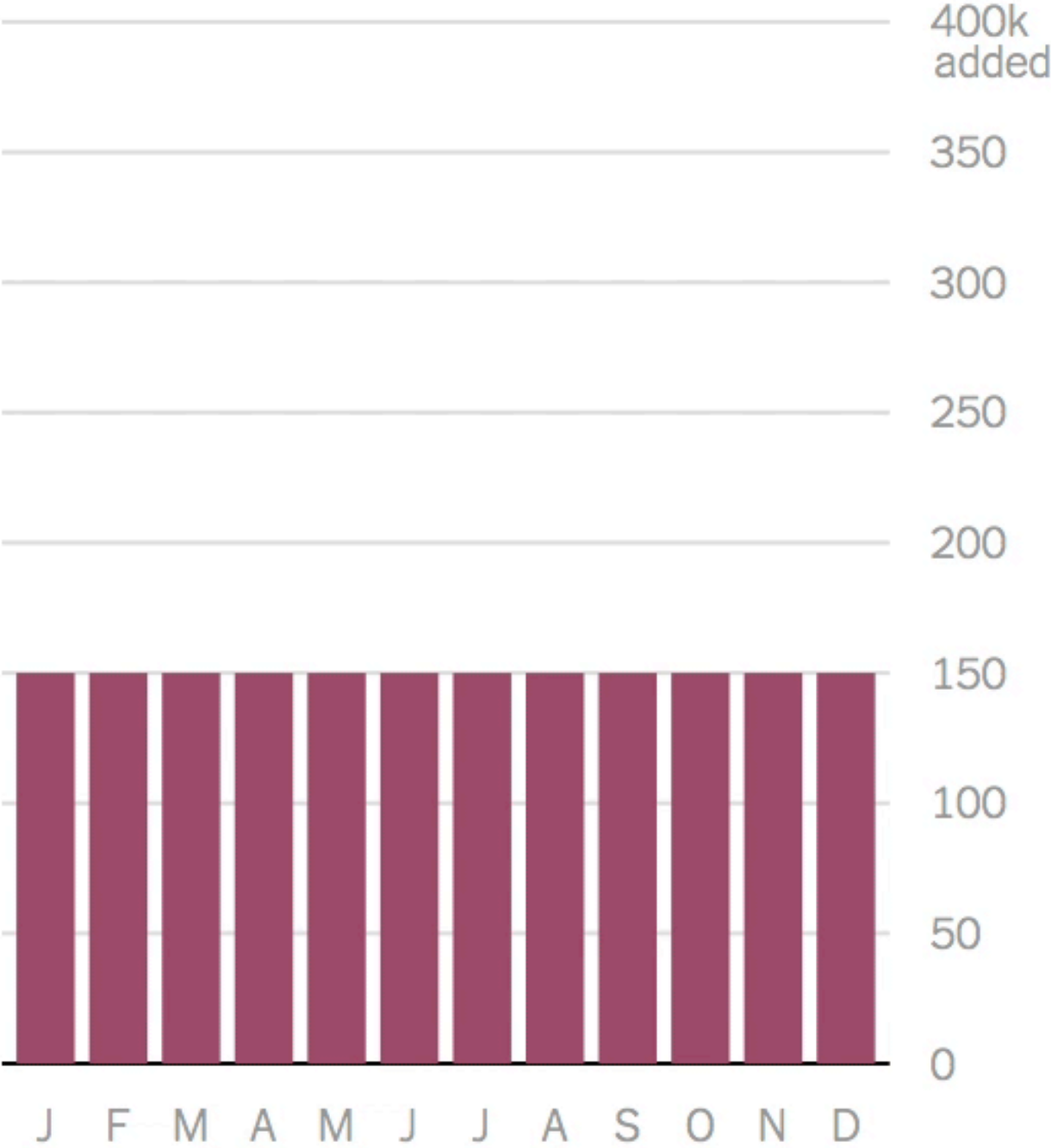
How Not to Be Misled by the Jobs Report

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.

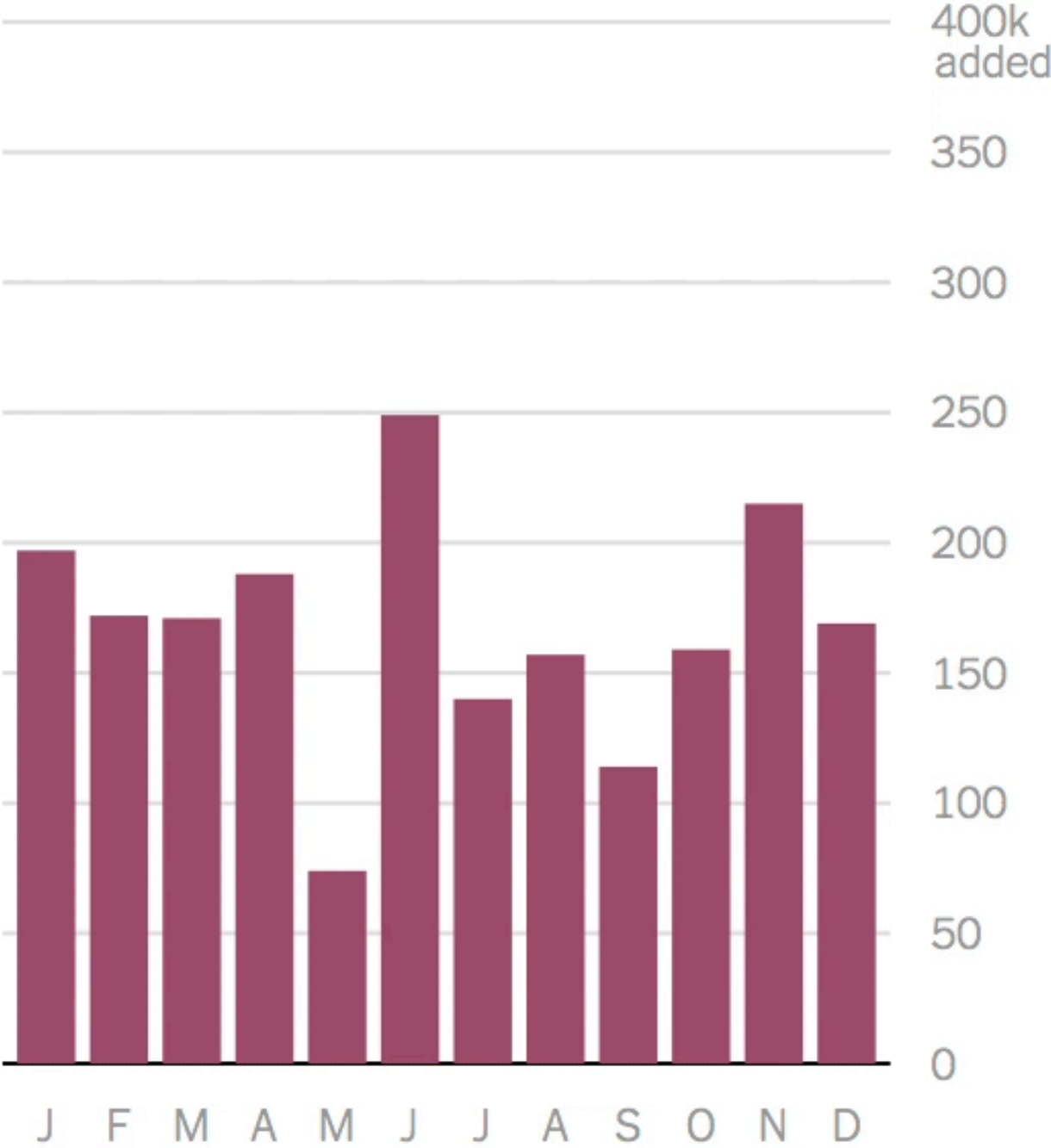


If job growth **were actually steady** over the last 12 months...

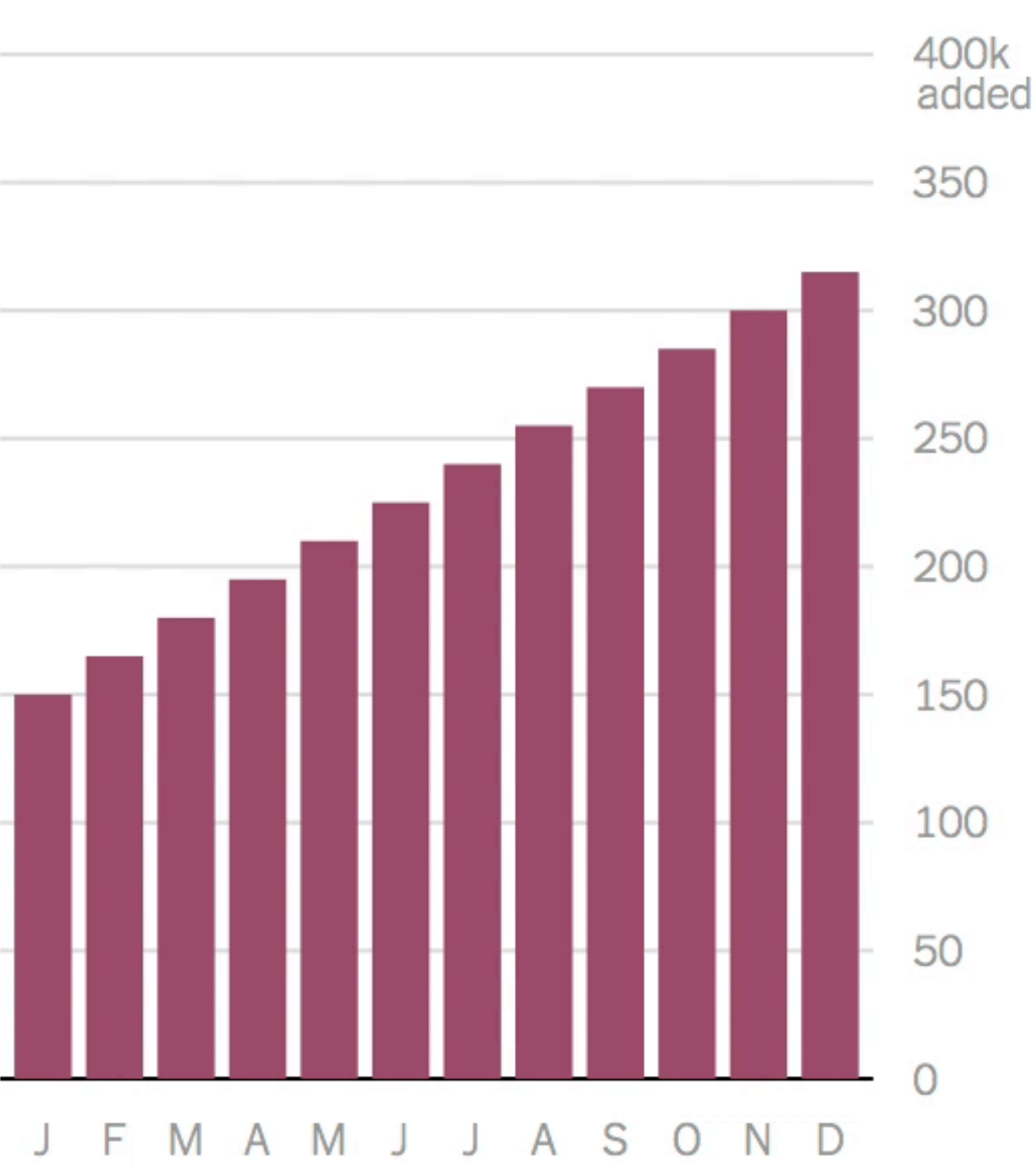


...the jobs report **could look like this:**

Play

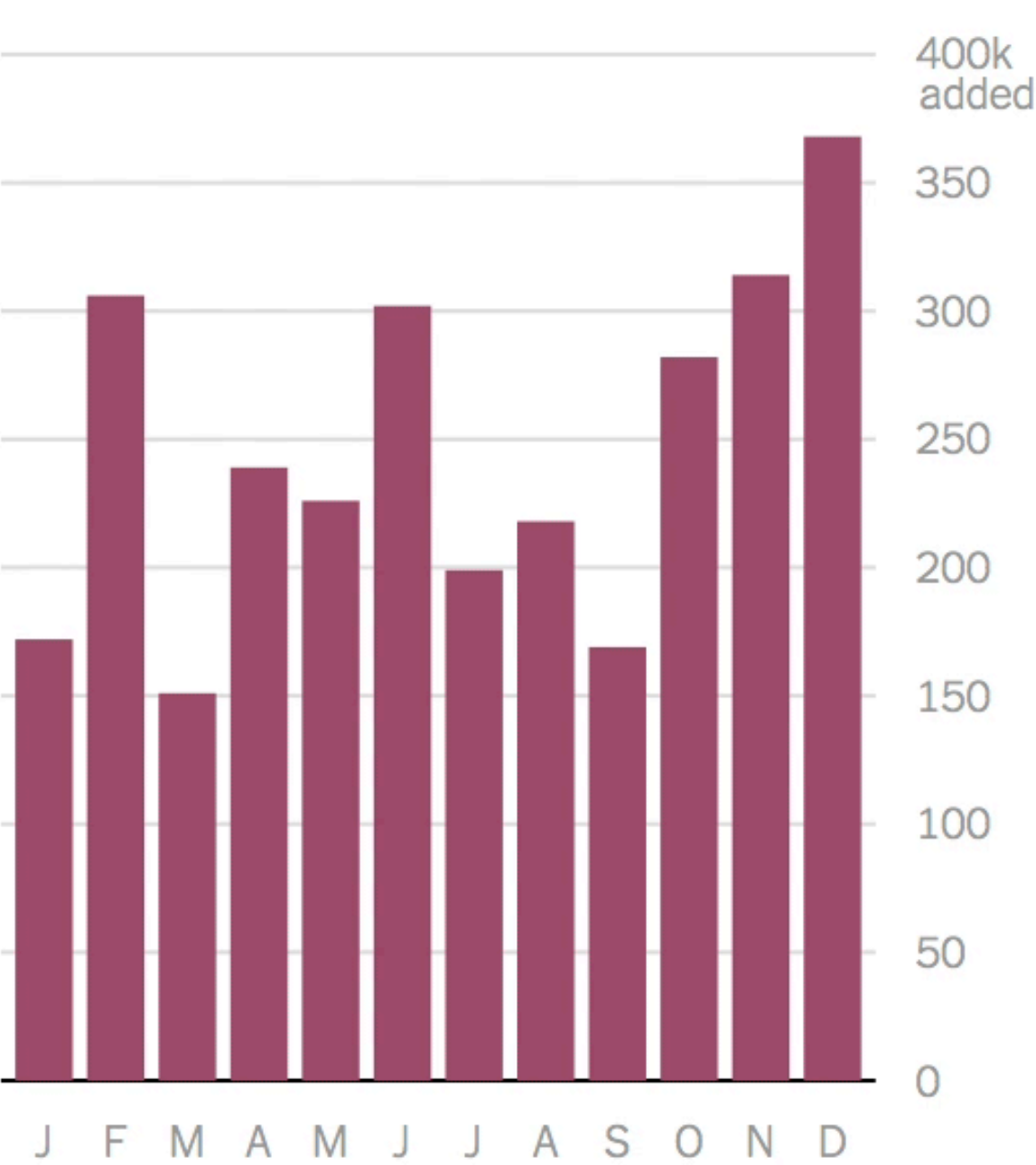


If job growth **had been accelerating...**



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

Play



SPIN

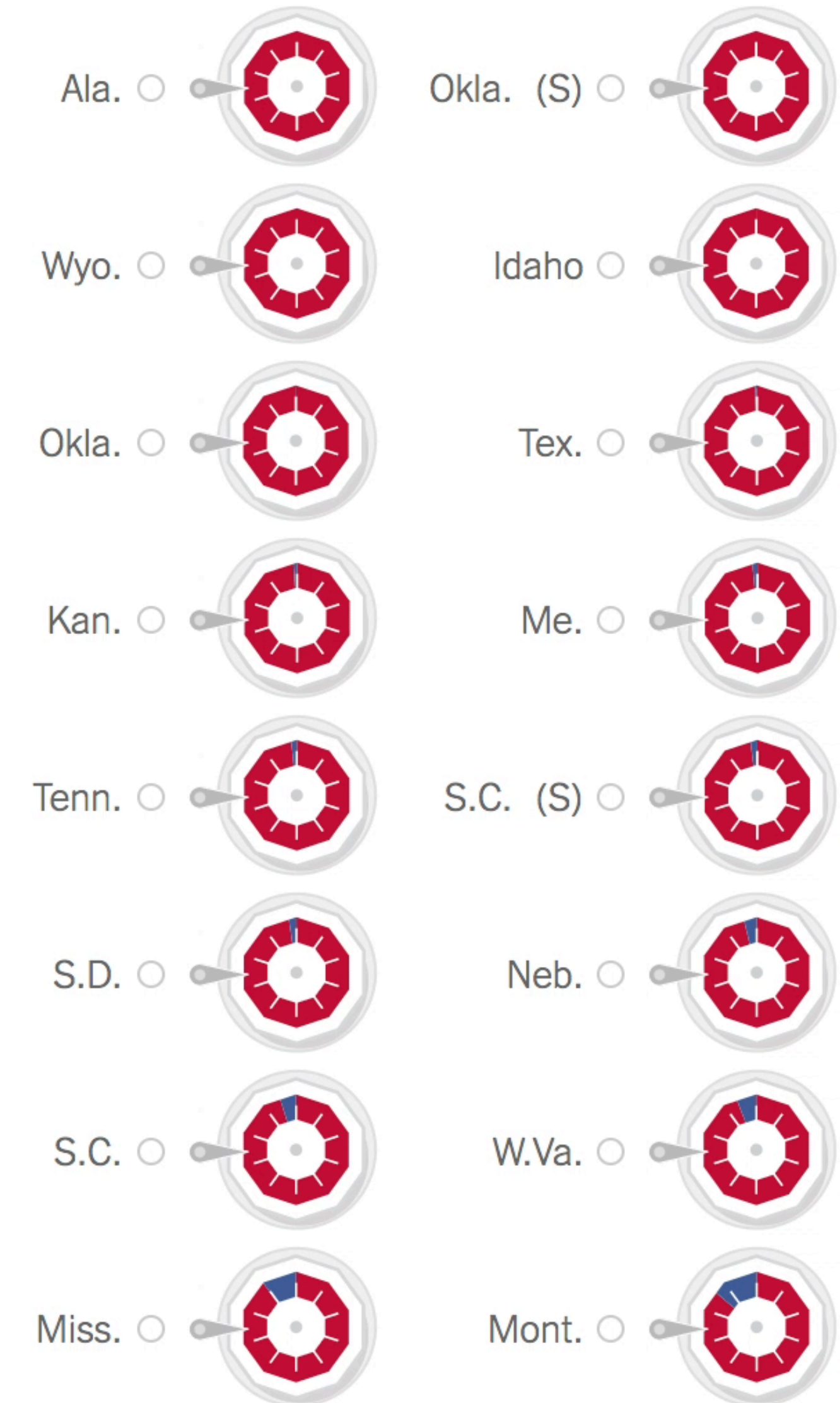
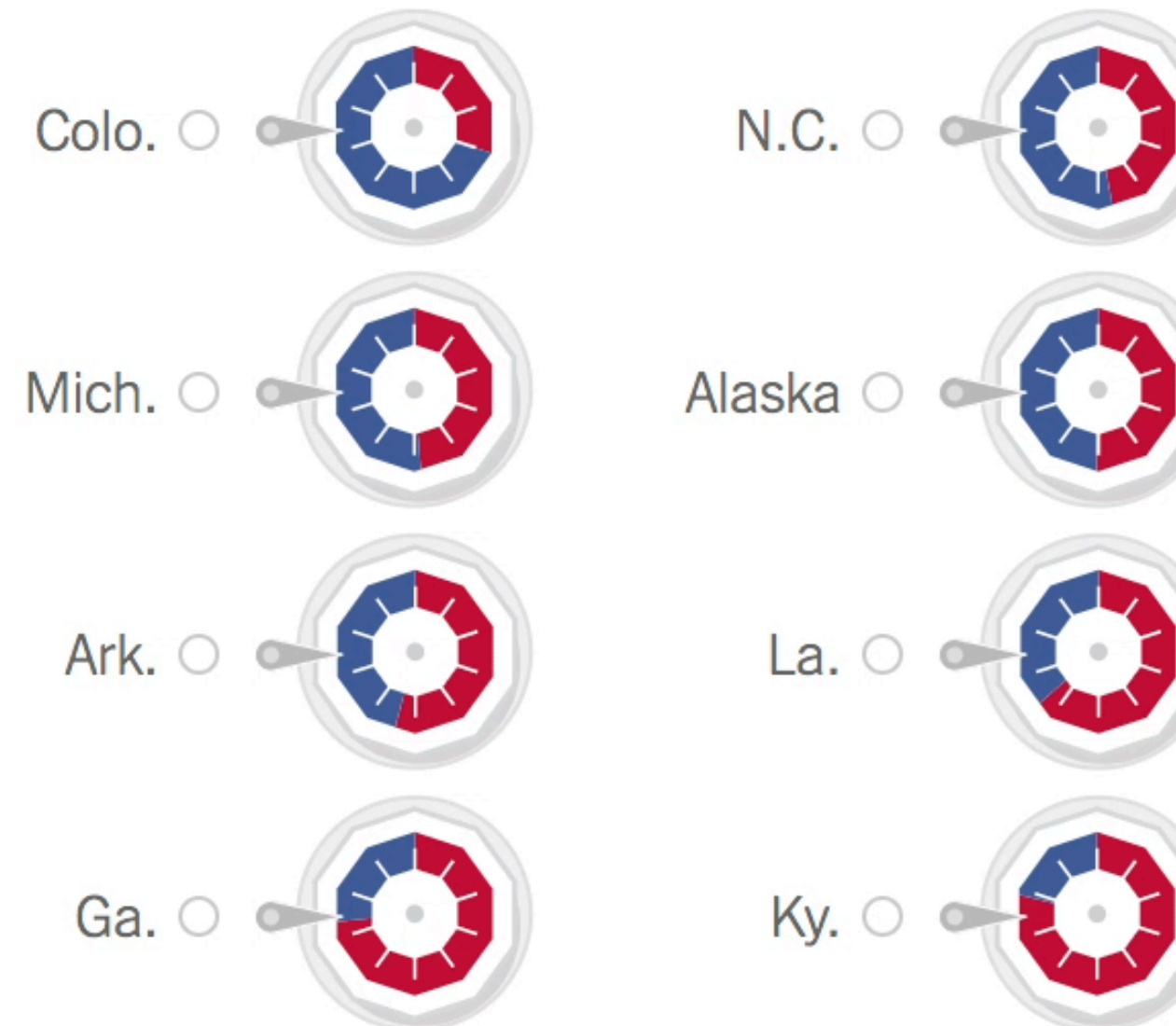
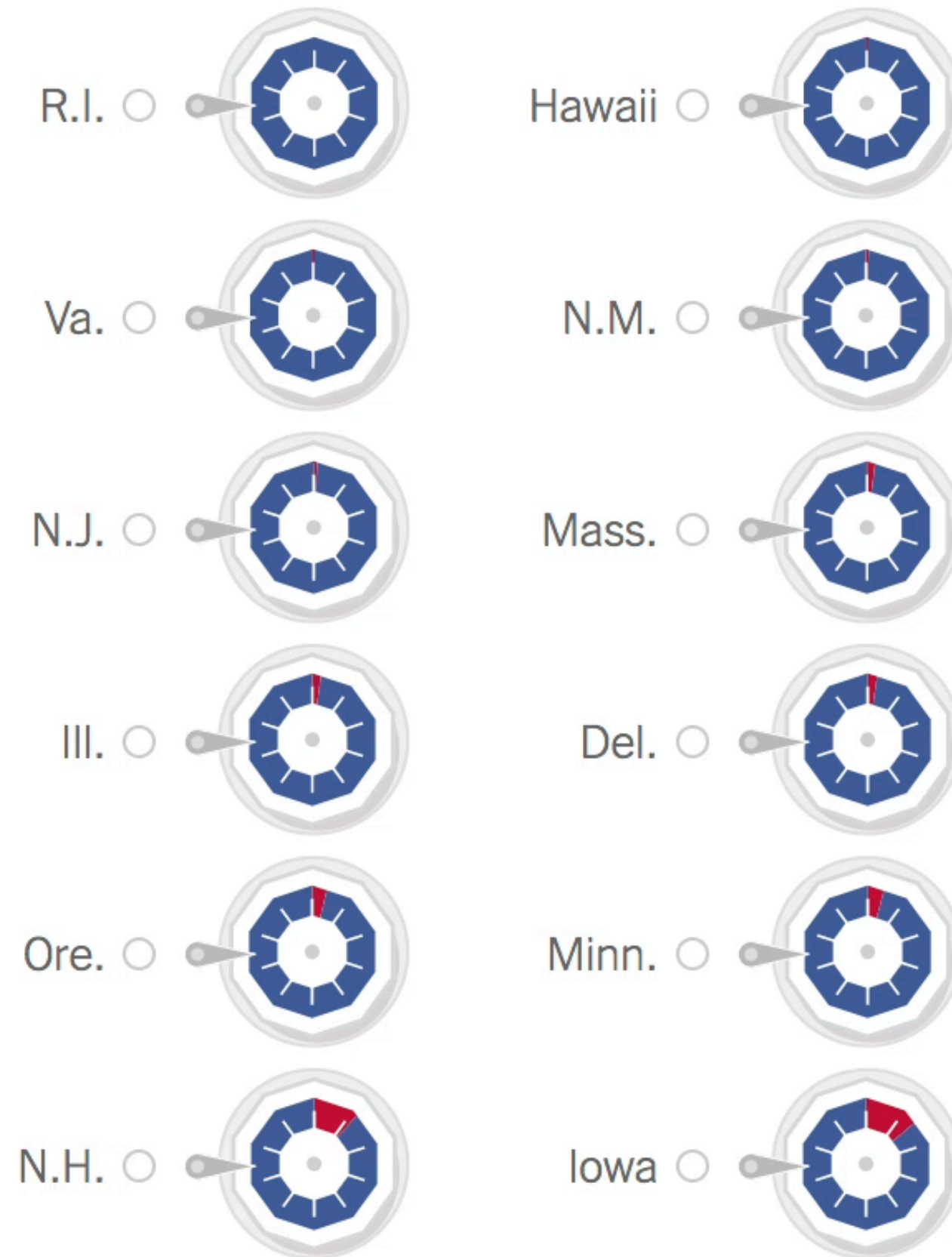
Democrats: ??

Republicans: ??

Likely Democratic

Competitive

Likely Republican



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.

Uncertainty

What does it mean?

How should I visualize it?

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.

Why Authors Sometimes Don't Visualize Uncertainty

A visualization expresses a signal

Authors simplify, crystallize, abstract the complexity of data.

Process validates signal

Authors decide whether process has "low enough" uncertainty.

Uncertainty obfuscates signal

Could distract, or require too much work from the reader.

Uncertainty

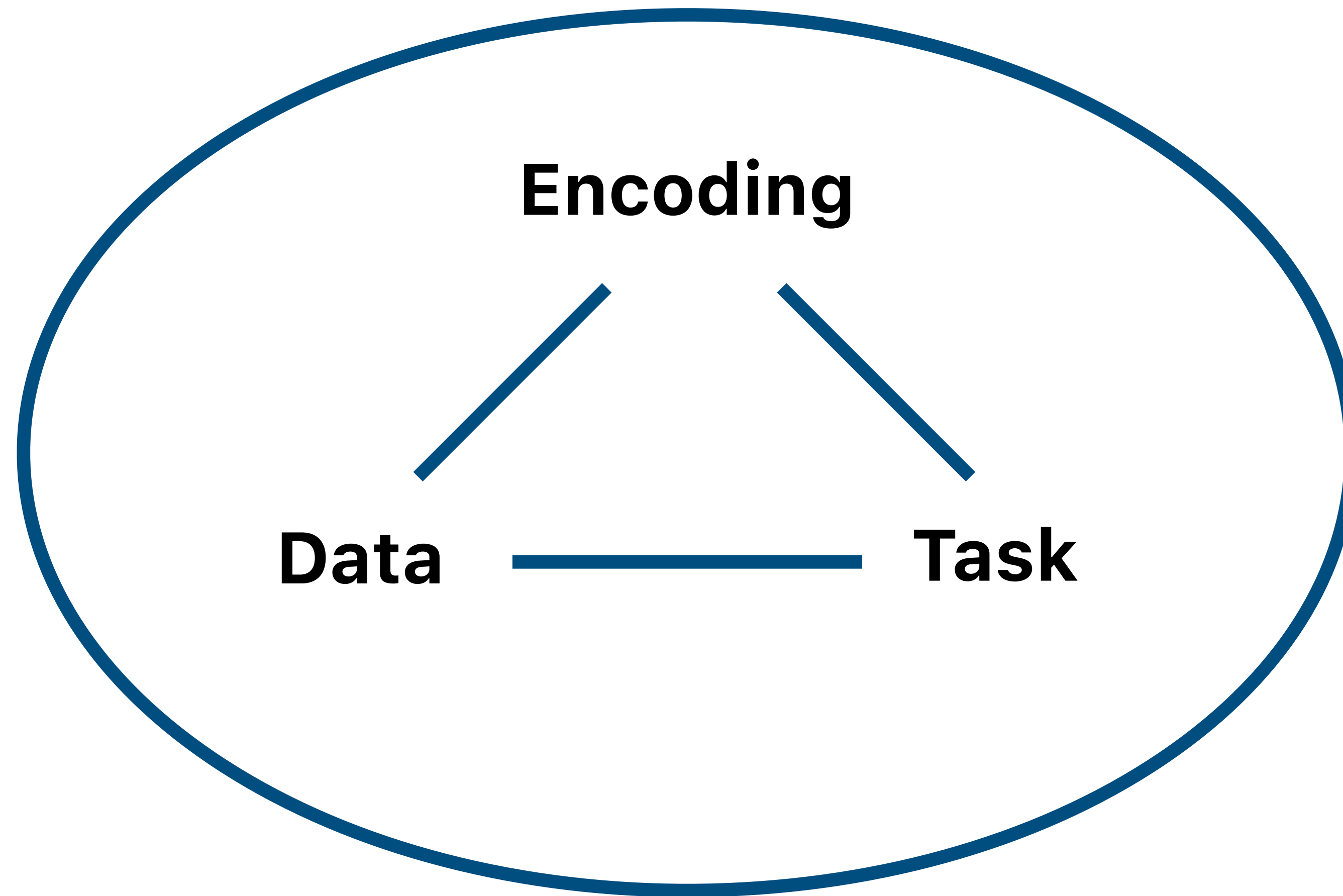
What does it mean?

Lots of things!

How should I visualize it?

It depends!

Conclusion



Users & Domain

Data and Image Models

Channels: Expressiveness Types and Effectiveness Ranks

➔ **Magnitude** Channels: **Ordered** Attributes



➔ **Identity** Channels: **Categorical** Attributes



Tamara Munzner, *Visualization Analysis and Design* (2014).

Visual Encoding and Dark Patterns

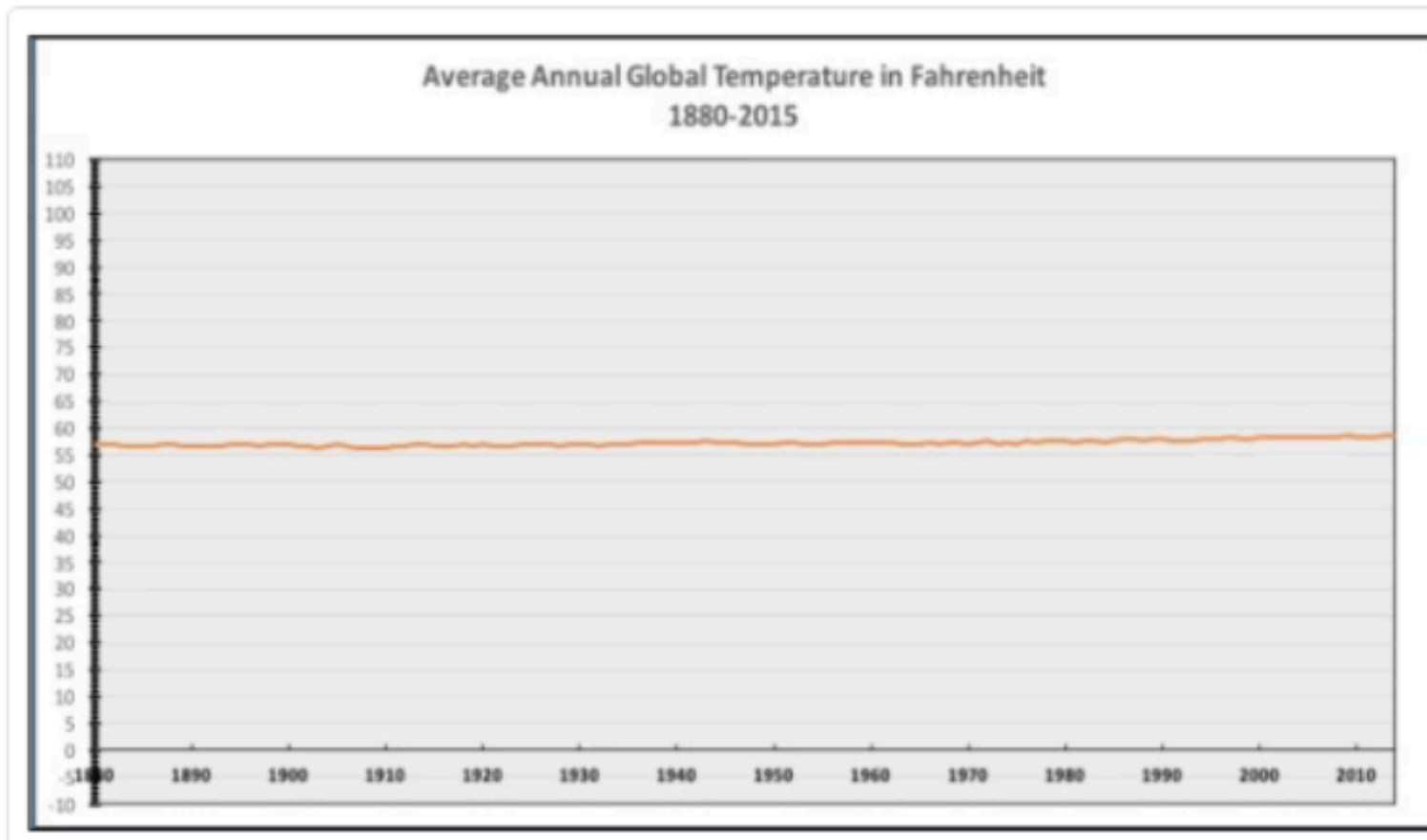
Truncating the y-axis?

To emphasize Q-interval (vs. Q-ratio)
If the zero value doesn't make much sense.
If it is the norm (e.g., stock charts).



The only [#climatechange](#) chart you need to see. natl.re/wPKpro

(h/t [@powerlineUS](#))

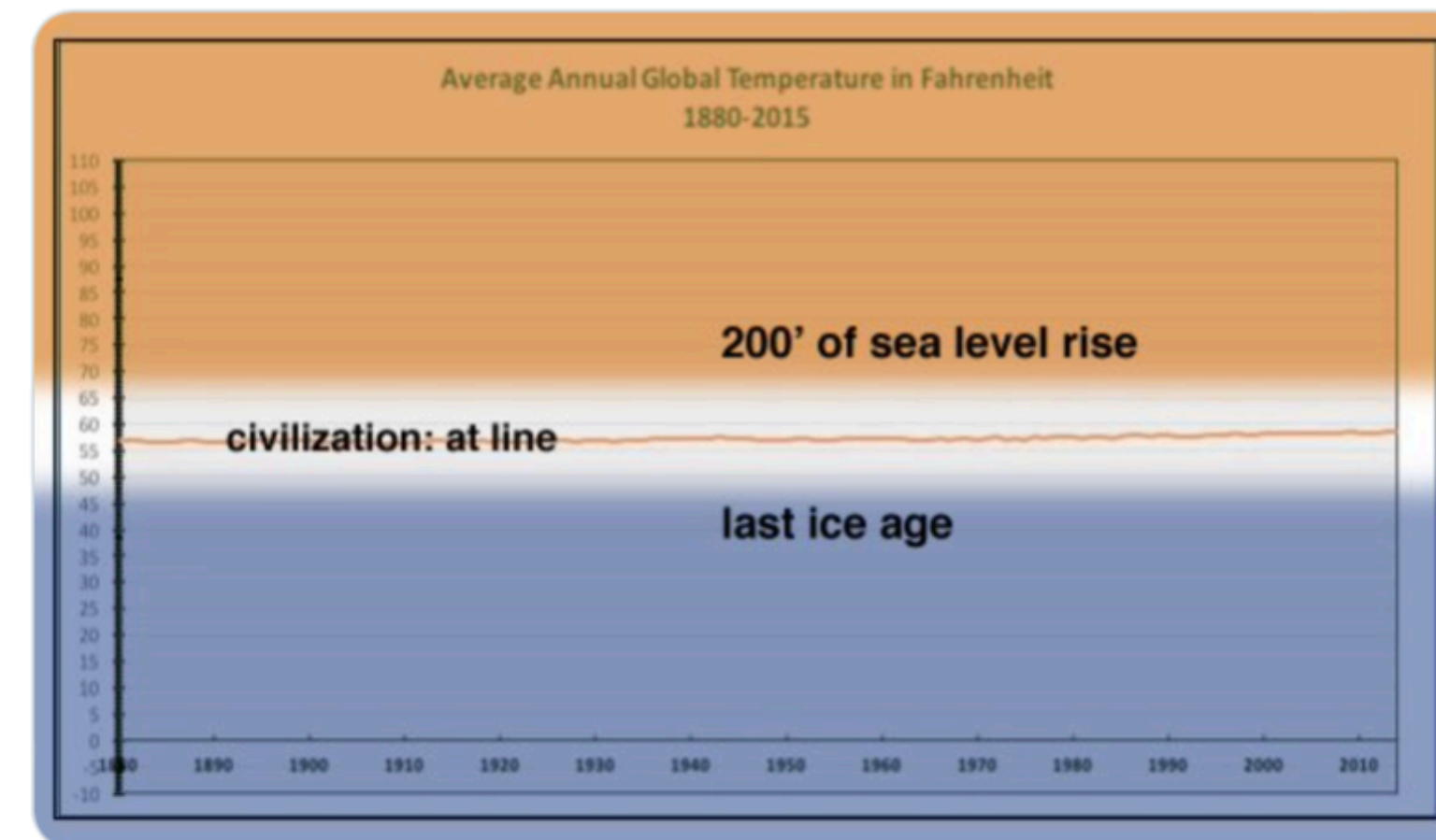


12:36 PM - 14 Dec 2015



Replying to [@NRO](#)

[@NRO](#) [@powerlineUS](#) [@bradplumer](#) I'm sure someone else has fixed this for you, but here you go. Great idea, thx --



5:28 PM · Dec 14, 2015

78 Retweets 1 Quote Tweet 208 Likes

39

Perception

Graphical Perception Studies

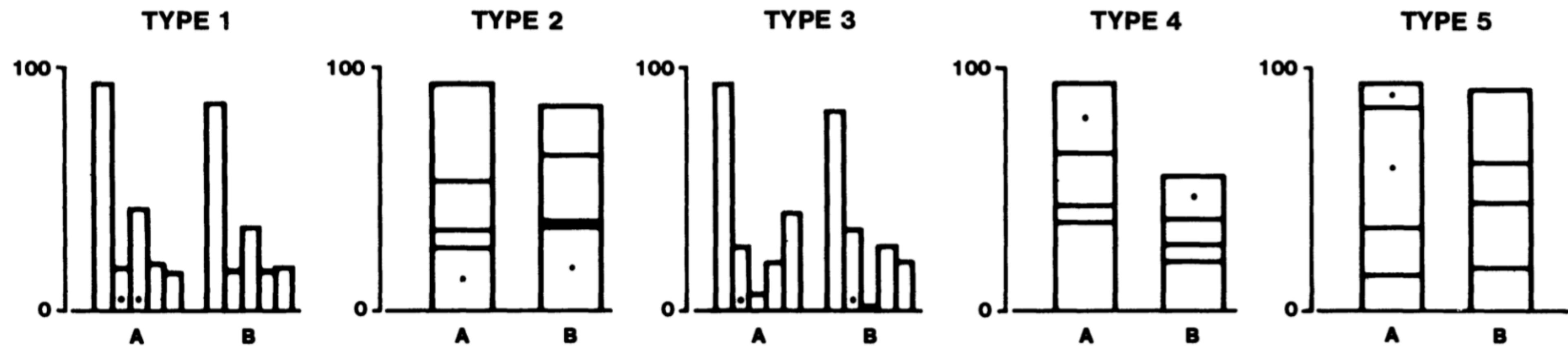


Figure 4. Graphs from position-length experiment.

What proportion is the smaller marked section of the larger?

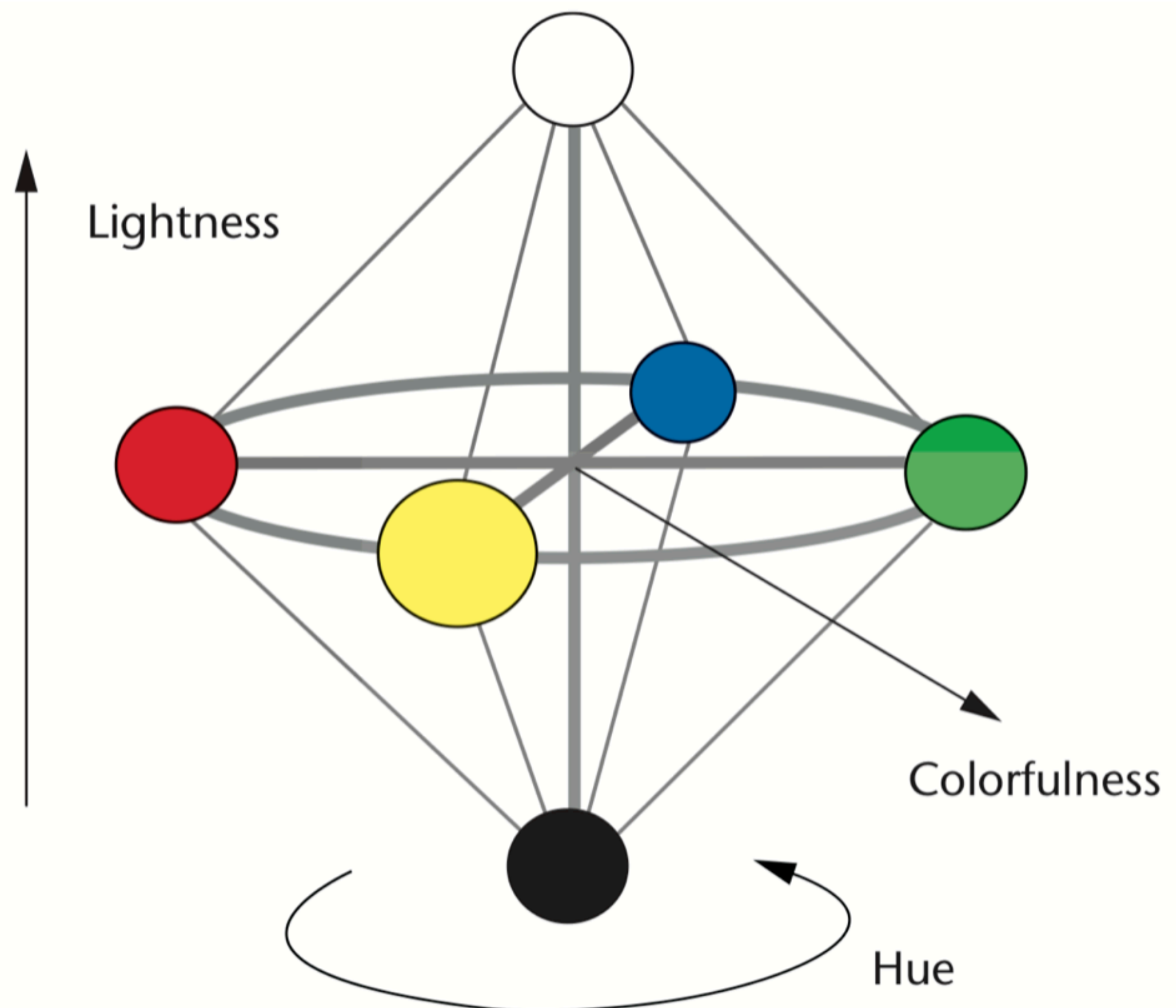
Color

OKLAB Color Space

Oklab is modern version of CIELAB that we recommend

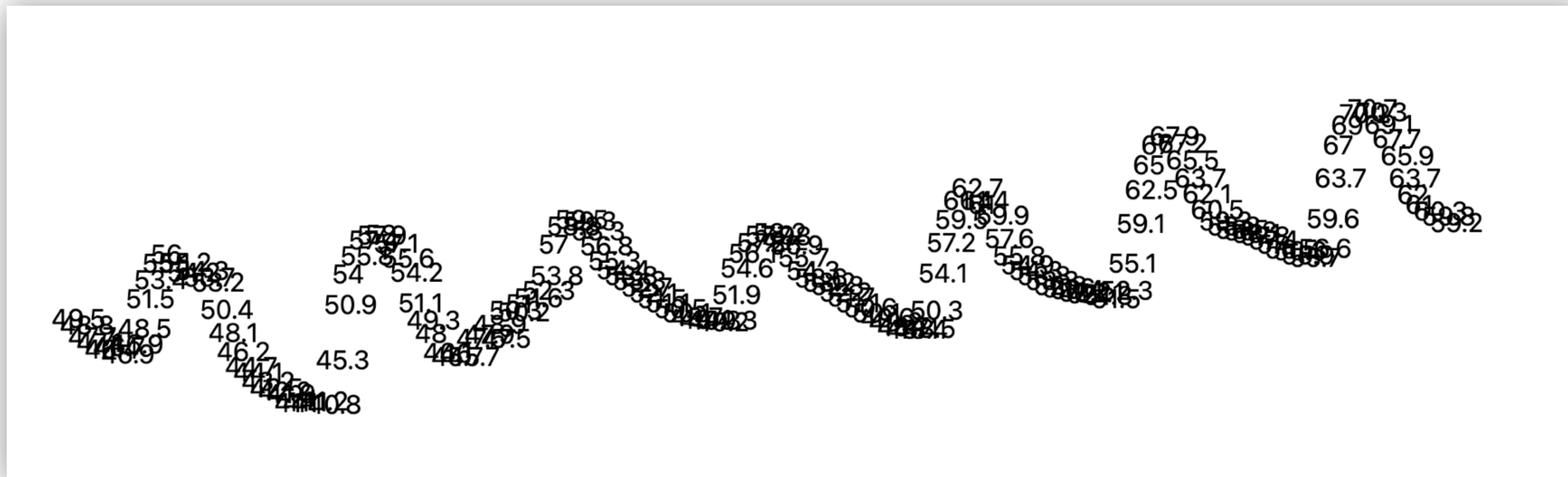
In CSS:

```
oklch(65% 50% 0)
```



JavaScript

Now, let's make our very first data visualization in JS:



[js-lecture/weather02/](#)

(demo)

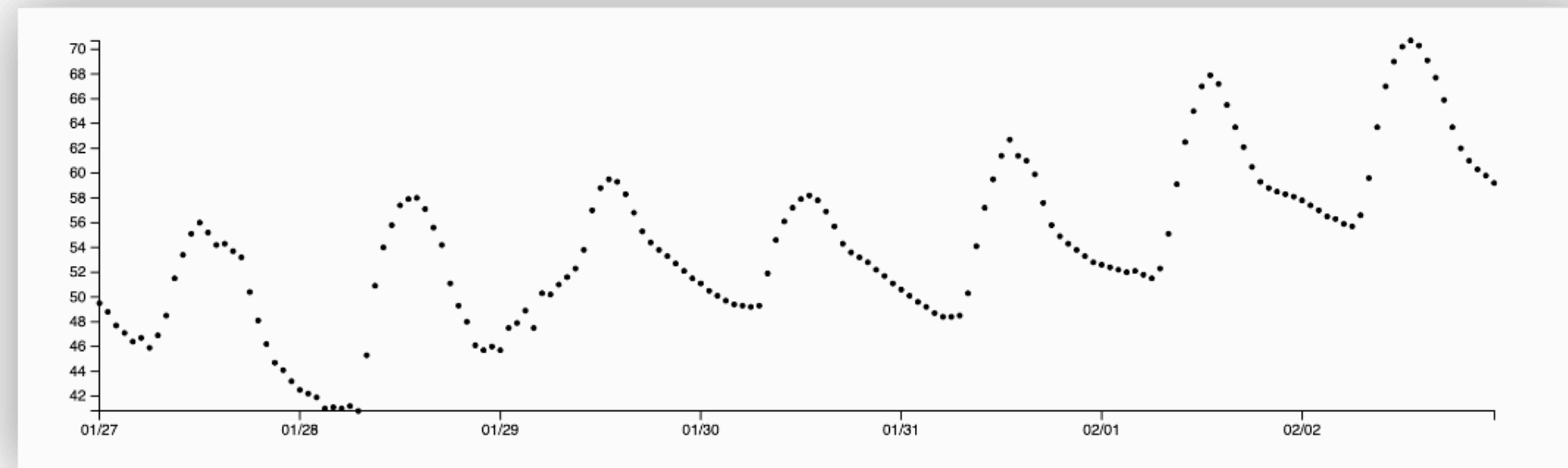
[js-lecture/weather03/](#)

(demo)

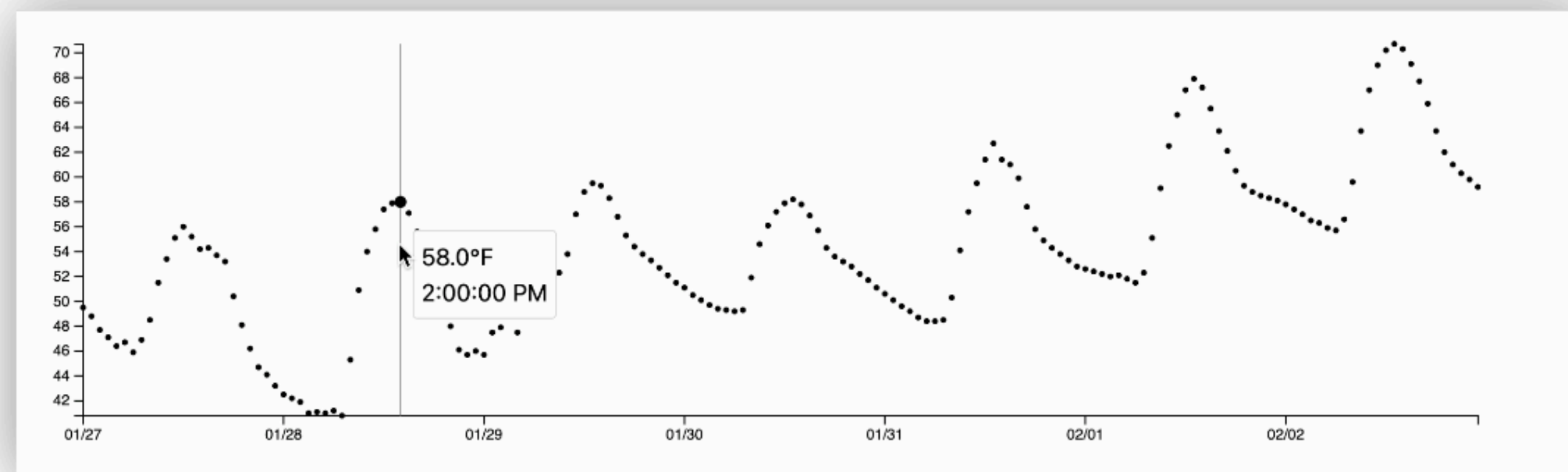
D3

Step 5: Improving our tooltip

Before:

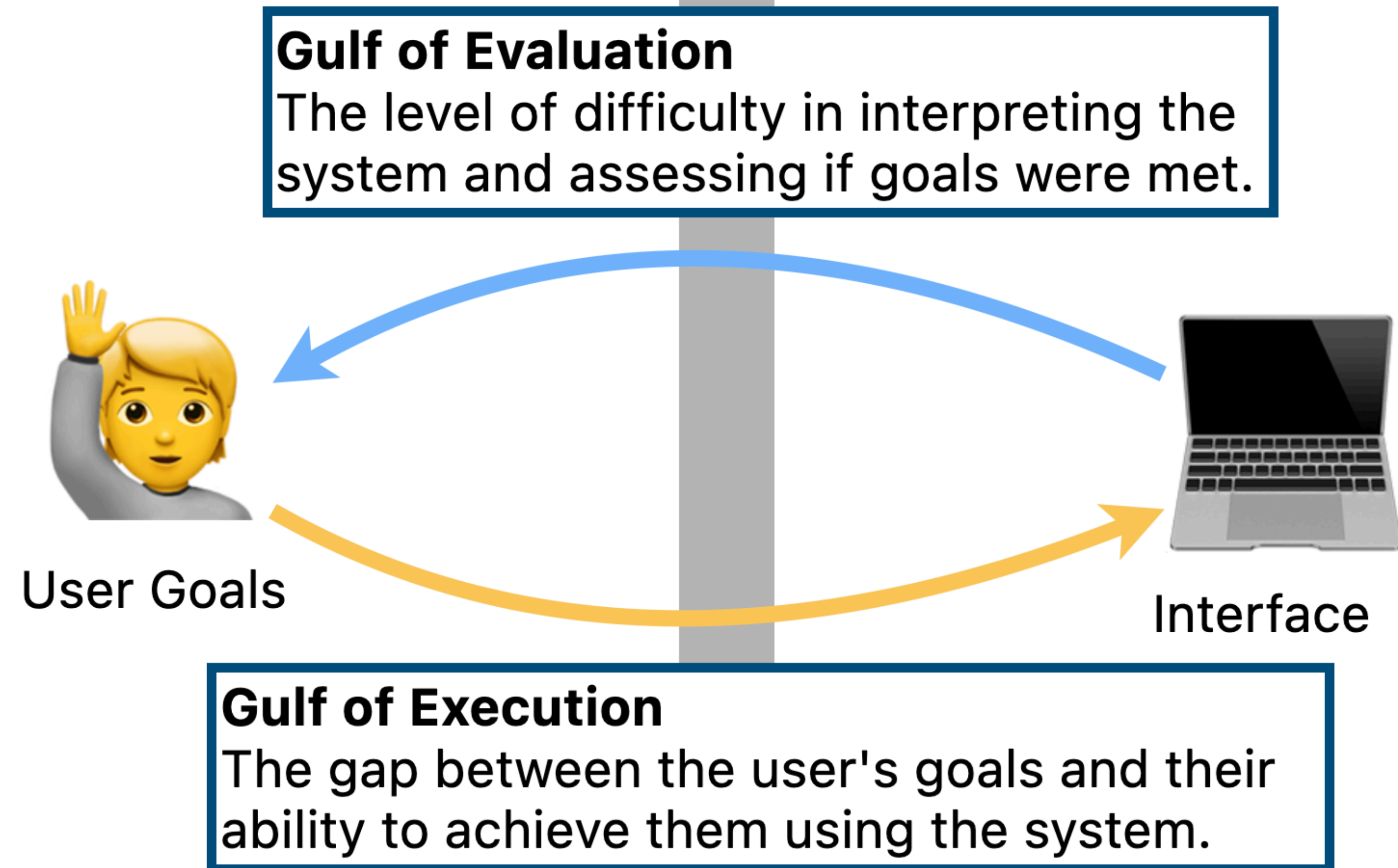


After:



Demo: [d3-lecture/weather05](#)

Interaction



Maps

Choropleth

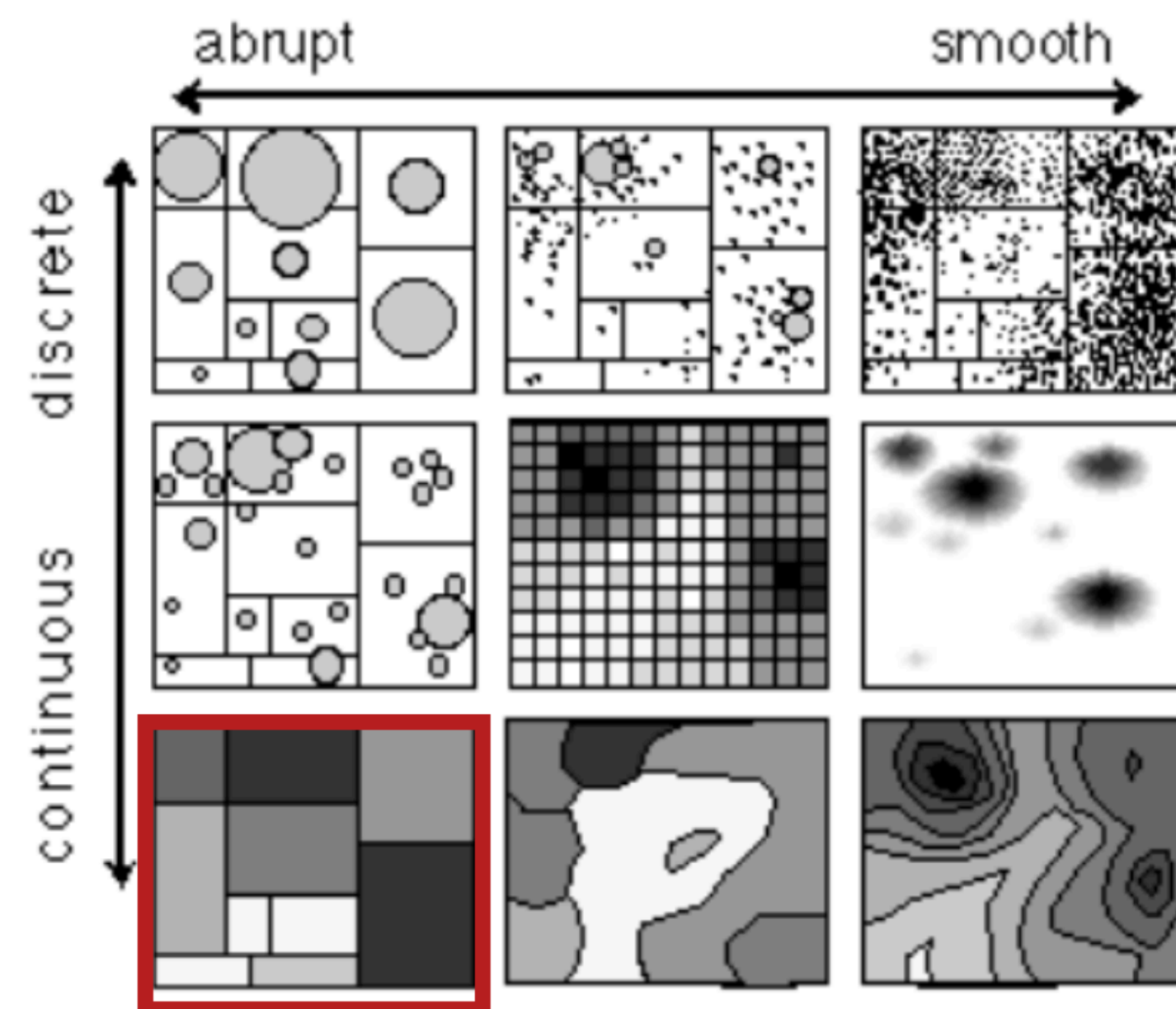
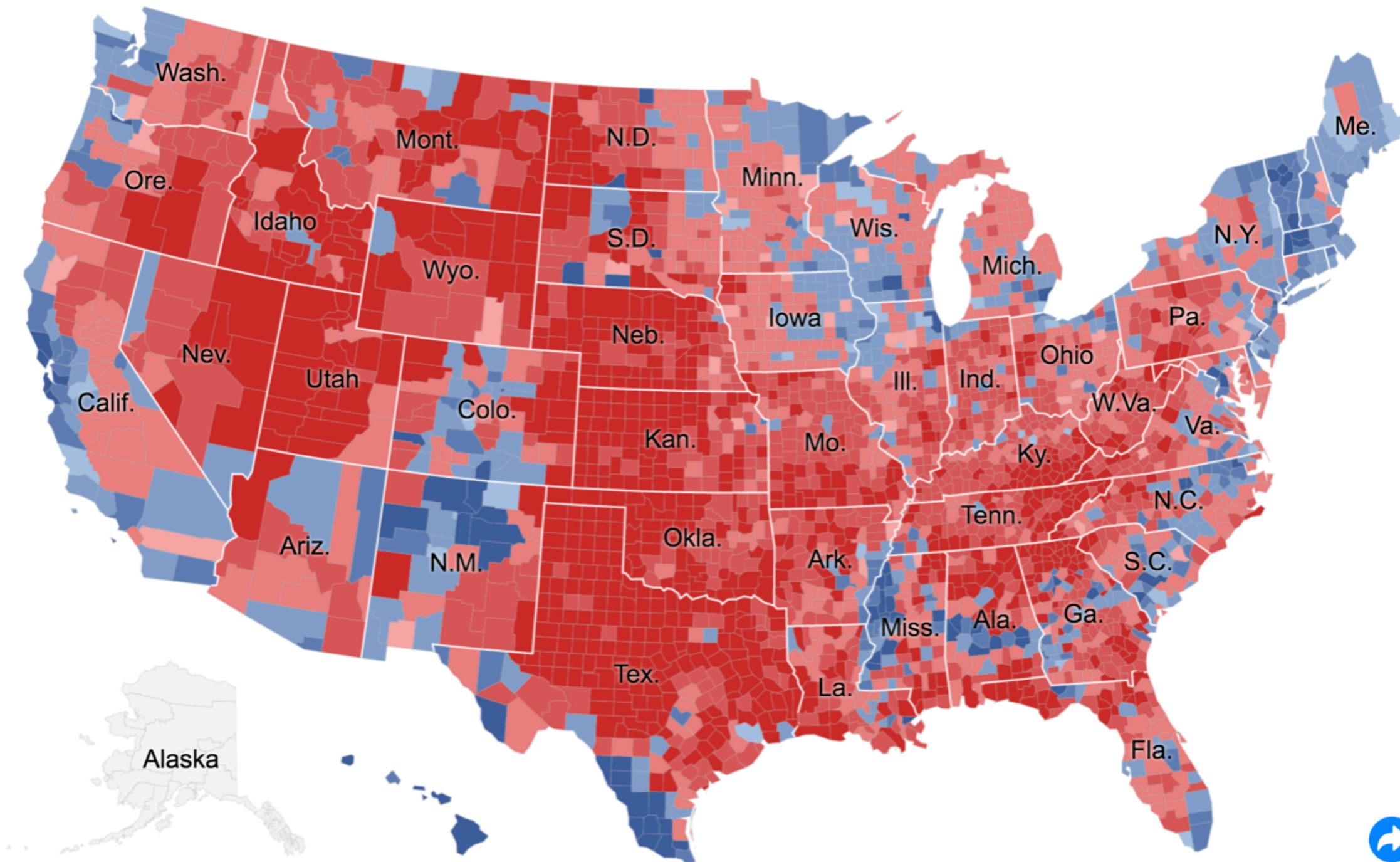


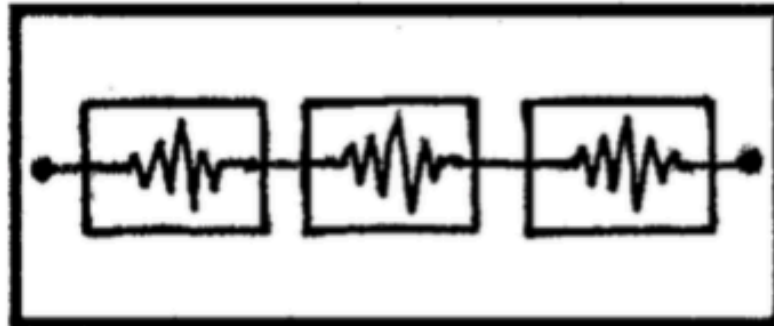
Fig. 9. Possible 2D translations of the 3D data models shown in figure 8.



<https://www.nytimes.com/interactive/2016/11/01/upshot/many-ways-to-map-election-results.html>

Narrative

Interactive Slideshow



R2
D3

A visual introduction to machine learning

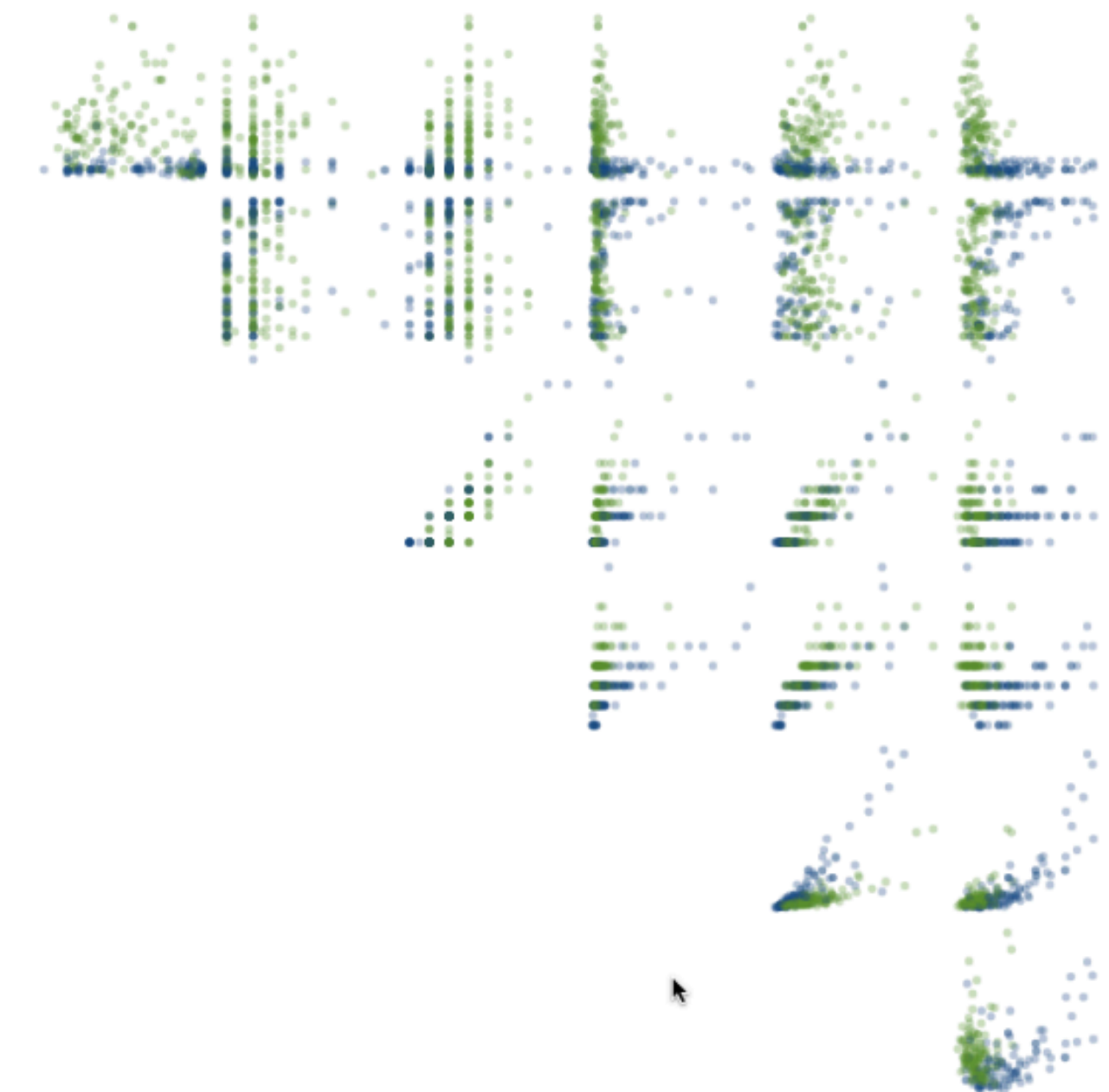
English

In machine learning, computers apply **statistical learning** techniques to automatically identify patterns in data. These techniques can be used to make highly accurate predictions.

Keep scrolling. Using a data set about homes, we will create a machine learning model to distinguish homes in New York from homes in San Francisco.

“Scrolly”-telling

SCROLL

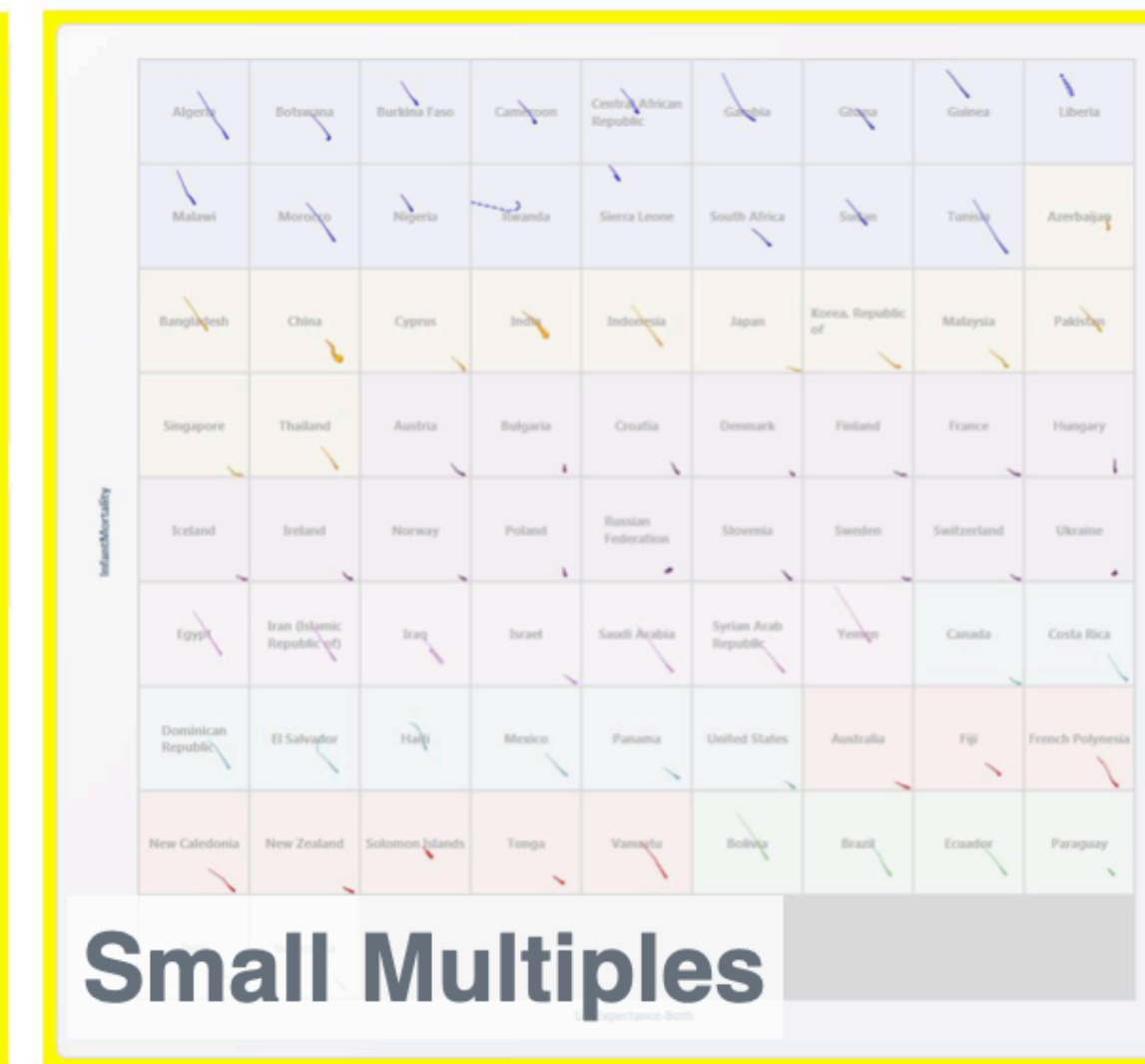
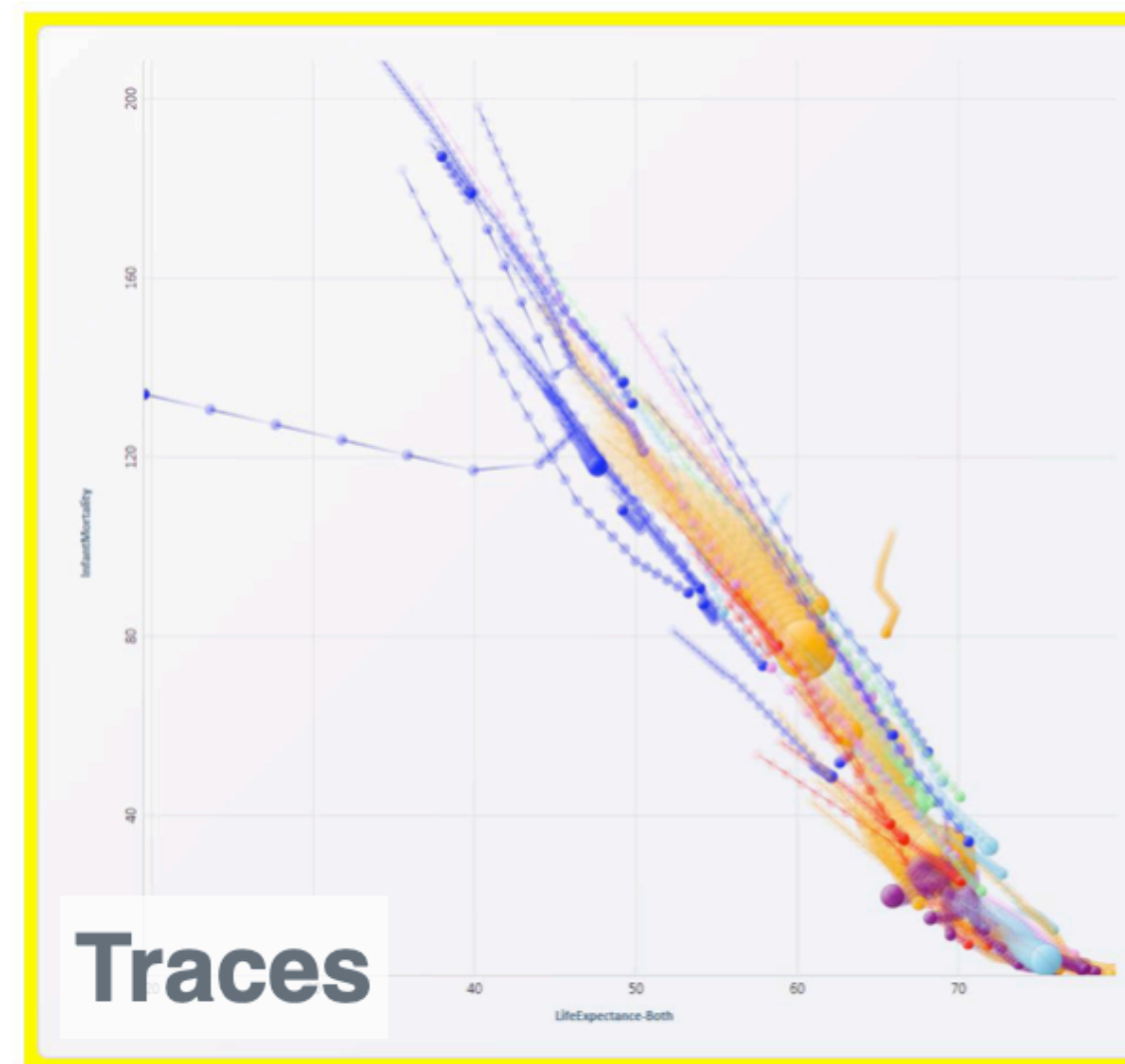
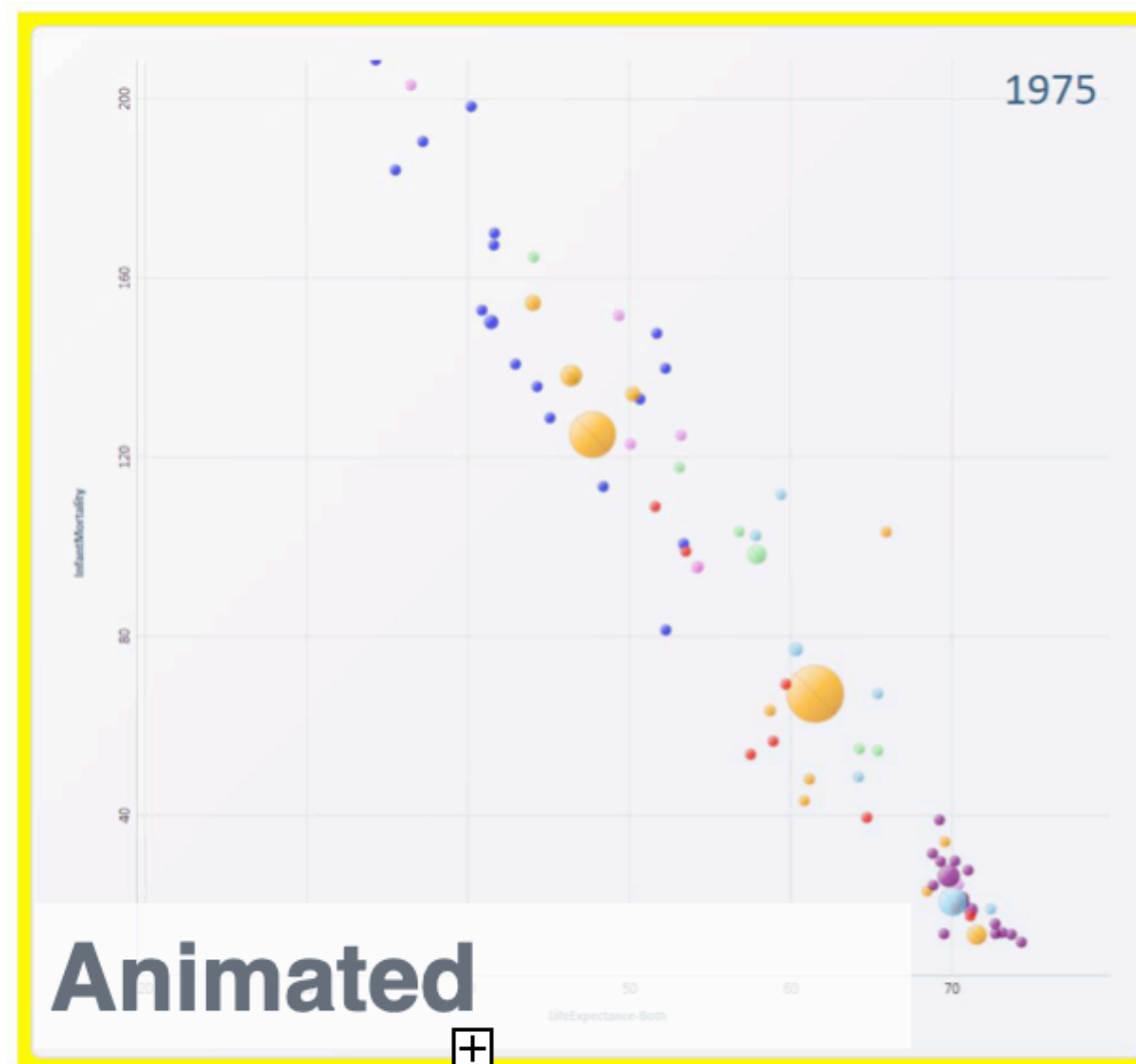


<http://www.r2d3.us/visual-intro-to-machine-learning-part-1/>

Animation

Study Conclusions

Analysis Task and Presentation Task.
Presentation condition included narration.
Subjects asked comprehension questions.

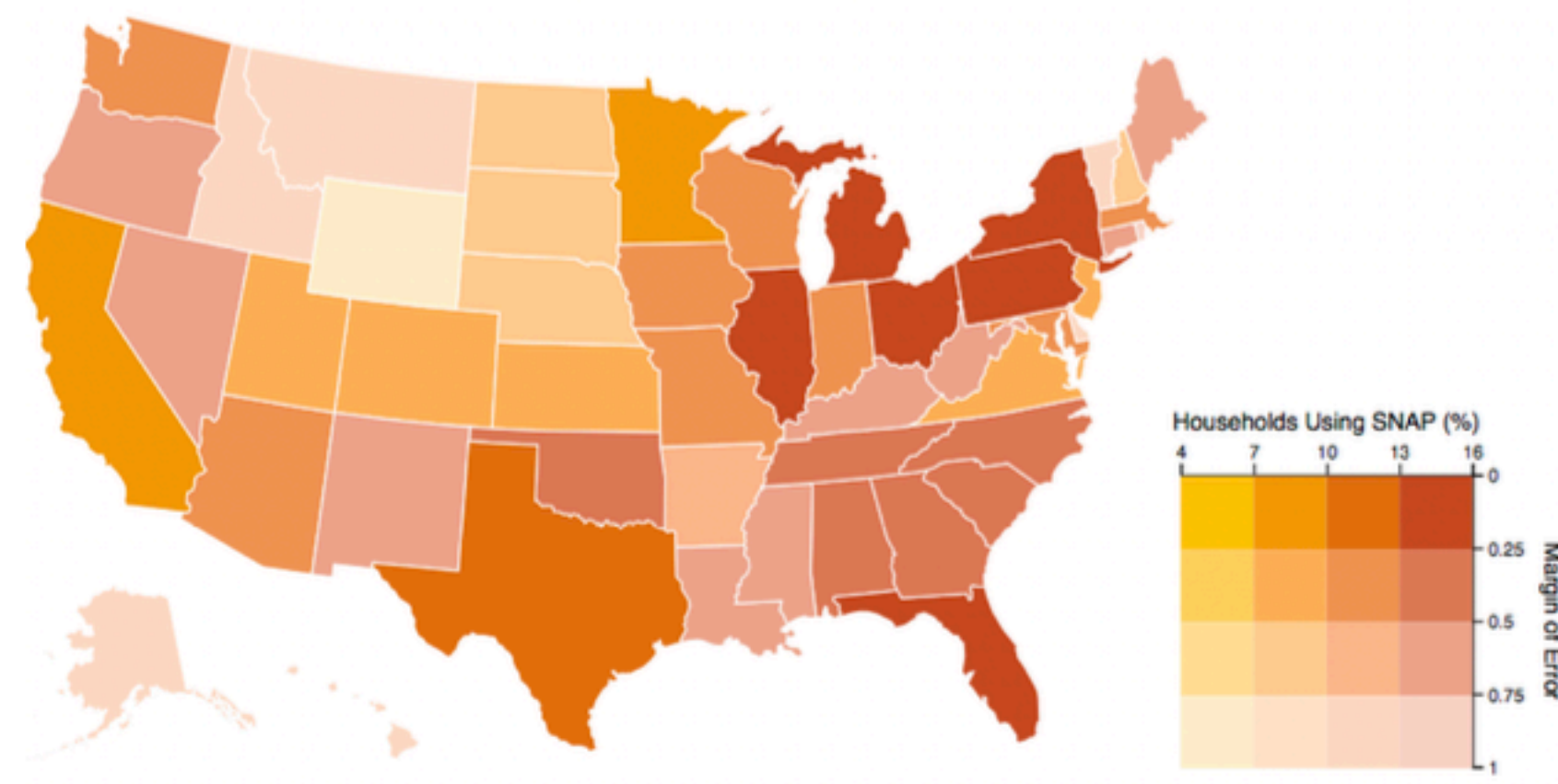


Which condition would participants:
be more **accurate**, be **faster**, and **prefer**?

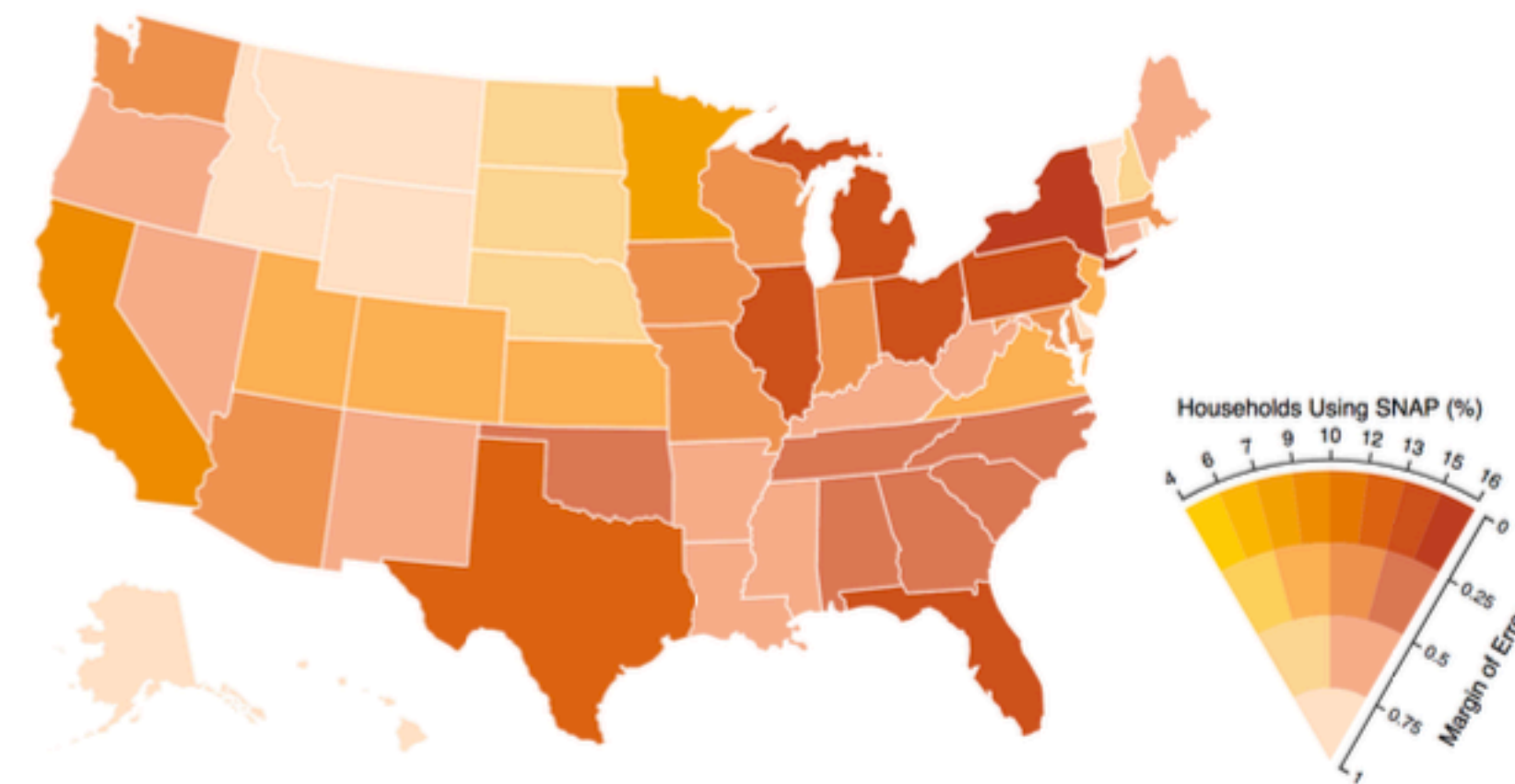
tryclassbuzz.com
Code: **anim**

Uncertainty

For uncertainty, use **visual variables** instead of visualizing point estimates



Bivariate Map (Data + Uncertainty)



Value-Suppressing Uncertainty Map

Thank You!

TAs: Giorgia, Heng

Tutors: Jesse, Lauren

Apply to join our staff in the future!