



# Uncertainty in Data and Visualization

Wan Fang

Southern University of Science and Technology

---

# Agenda

- What Does Uncertainty Mean?
- Uncertainty Visualization Pipeline
  1. Quantify Uncertainty
  2. Visual Encode Uncertainty
  3. Unify the Data Map and Uncertainty Map
- What Can Go Wrong

---

# When we talk about “uncertainty”

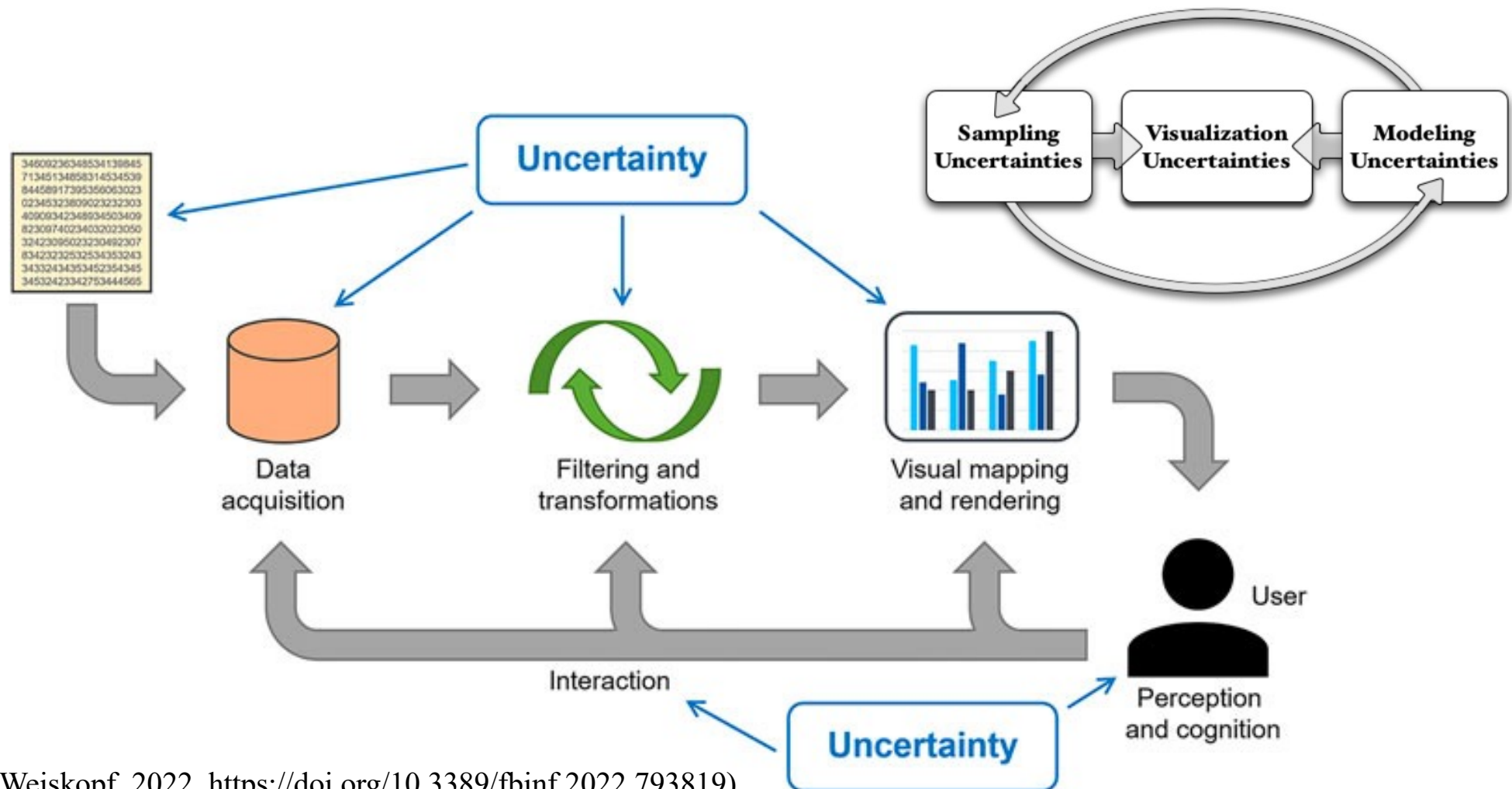
It can mean?

---

# When we talk about “uncertainty”



# The visualization pipeline including uncertainty



# Measurement Uncertainty

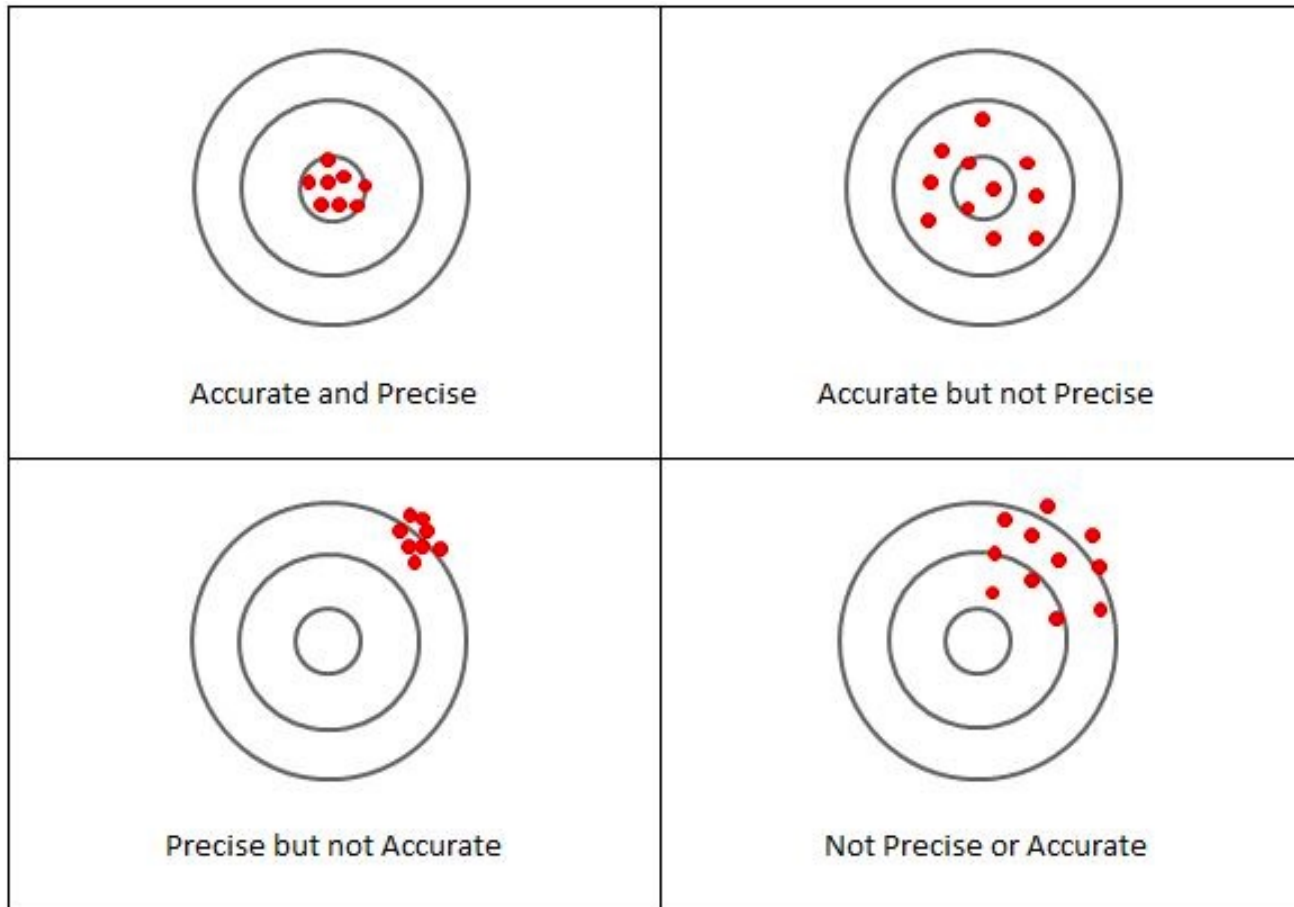
**Accuracy**



**Precision**



# Measurement Uncertainty



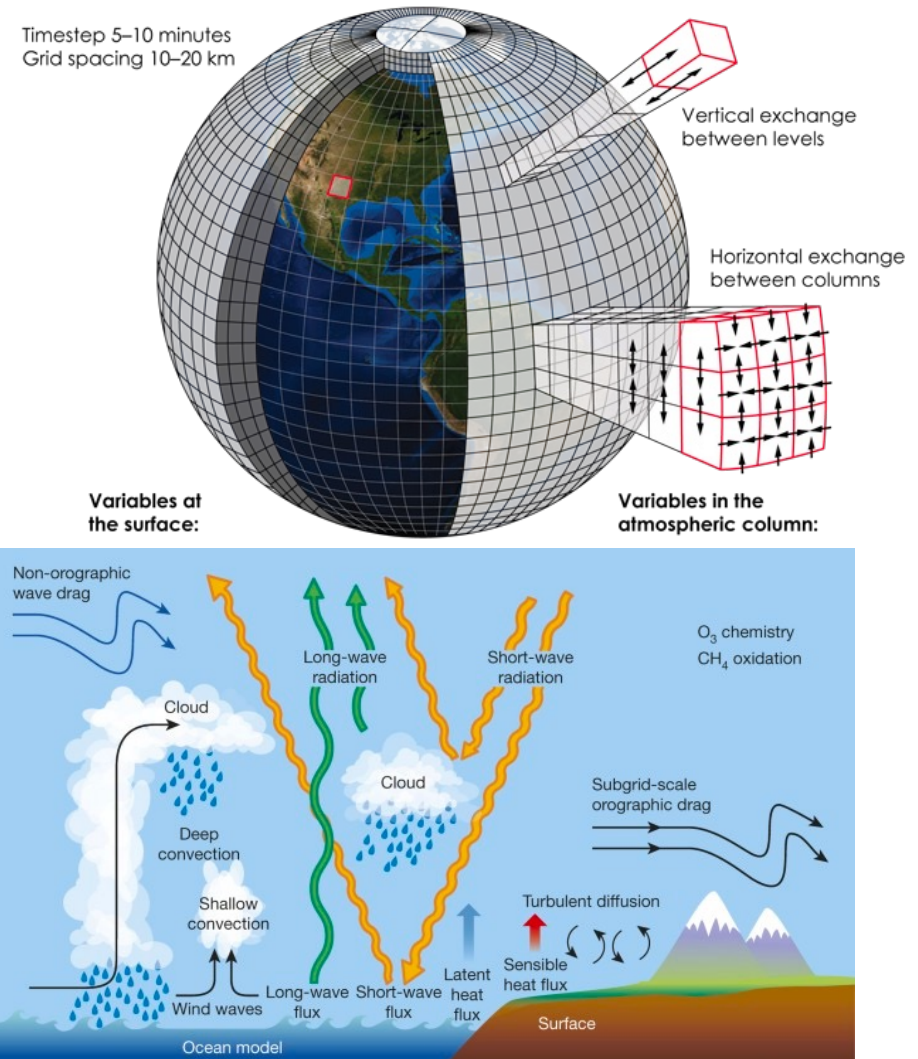


# Model Uncertainty



Weather forecast modeling

Credit: K. Cantner, AGI.



<https://doi.org/10.1038/nature14956>



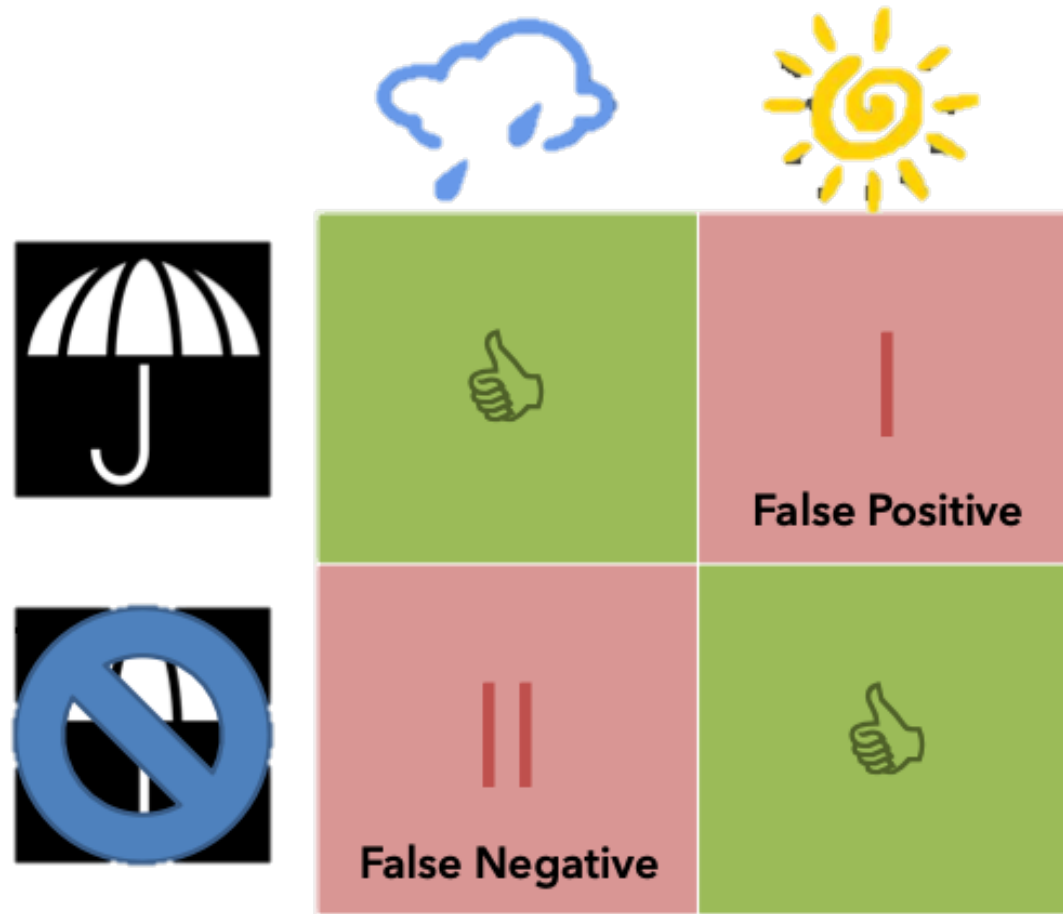
# Decision Uncertainty

Should I Bring an Umbrella?

“50% Chance of Rain”



# Types of Error



# What Does Uncertainty Mean?

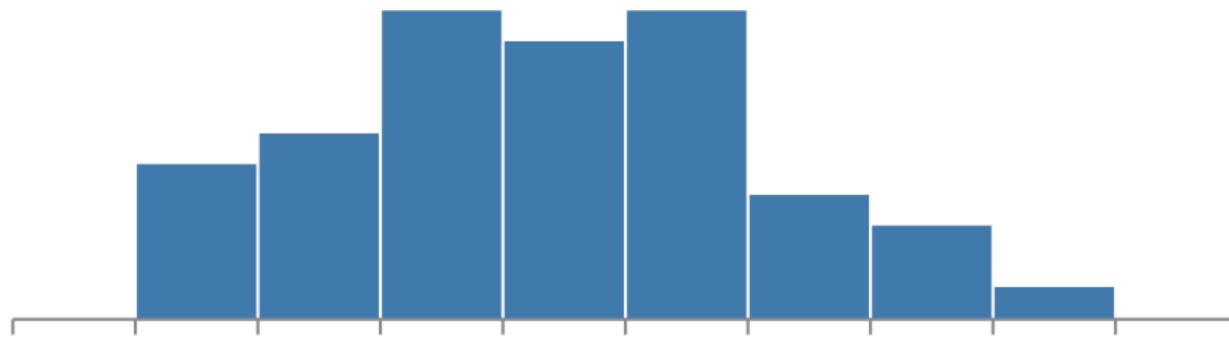
Any one of a number of potentially interconnected quantitative, qualitative, or factors that affect the quality, reliability, or utility of your data or data-driven decisions. Anything that can cause you to be unsure about your data or how to use it.

# Quantify Uncertainty

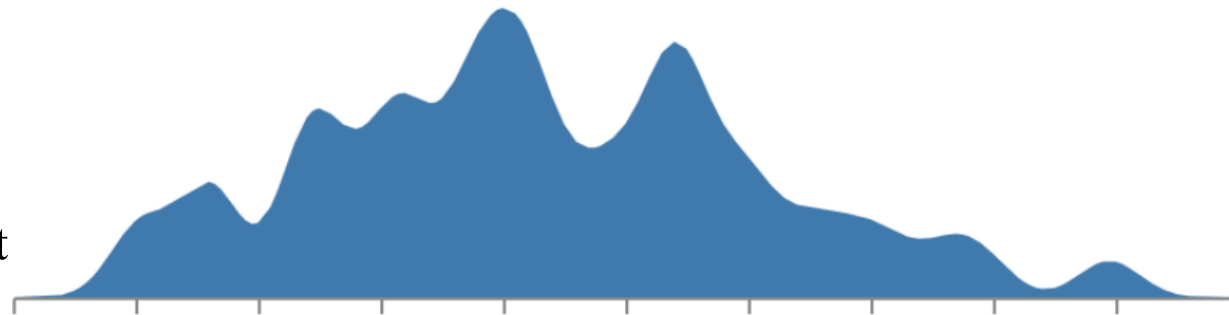
## Distributions

*probabilistic modeling of data uncertainty*

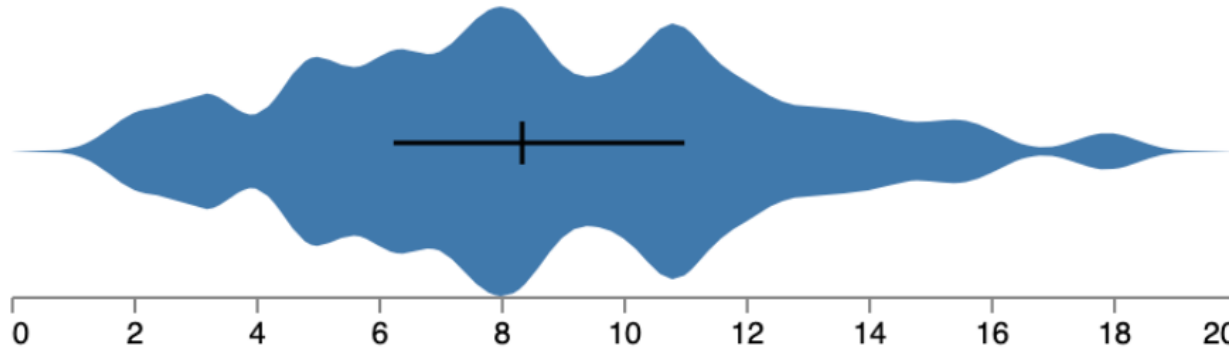
Histogram



Density Plot

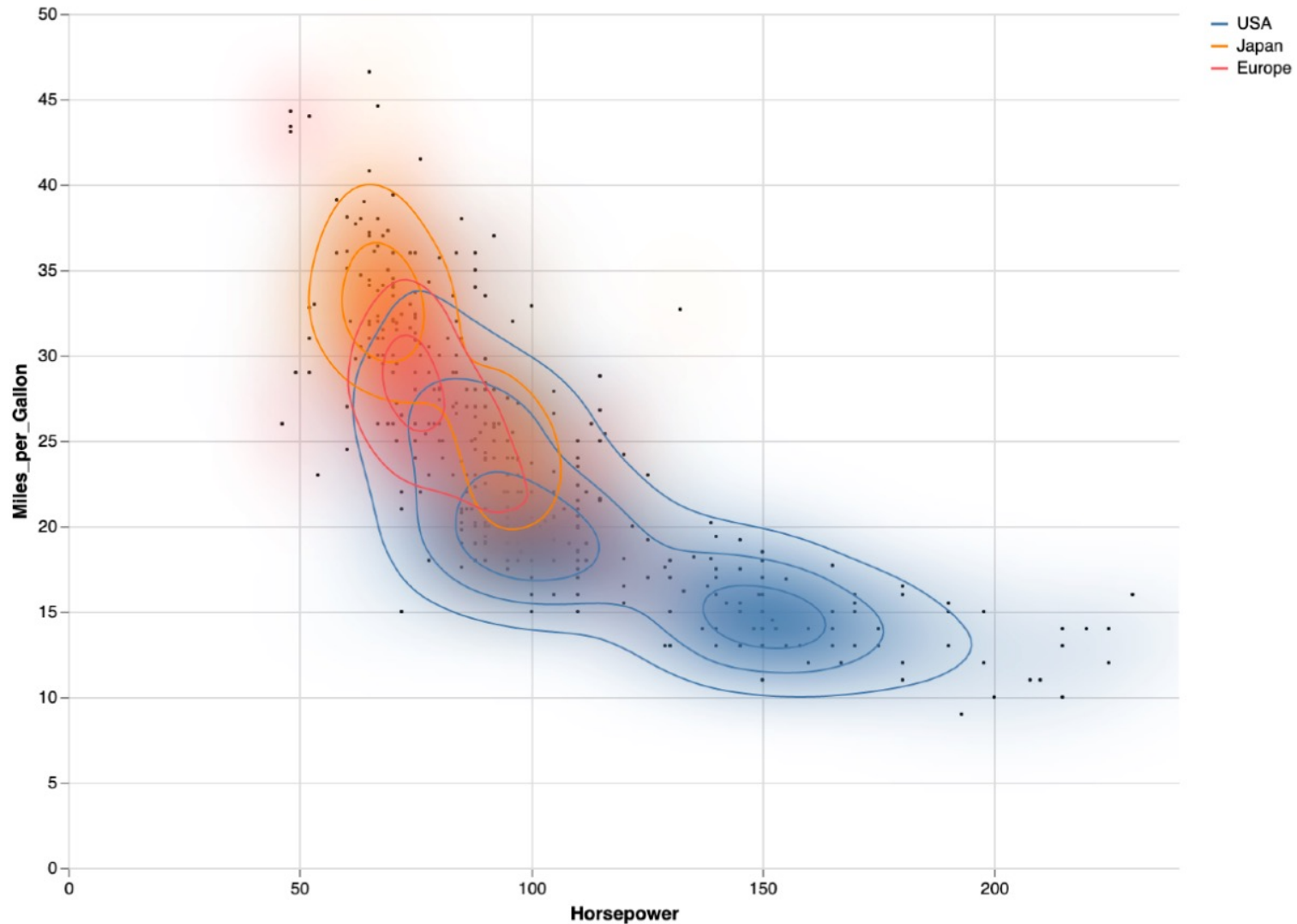


Violin Plot



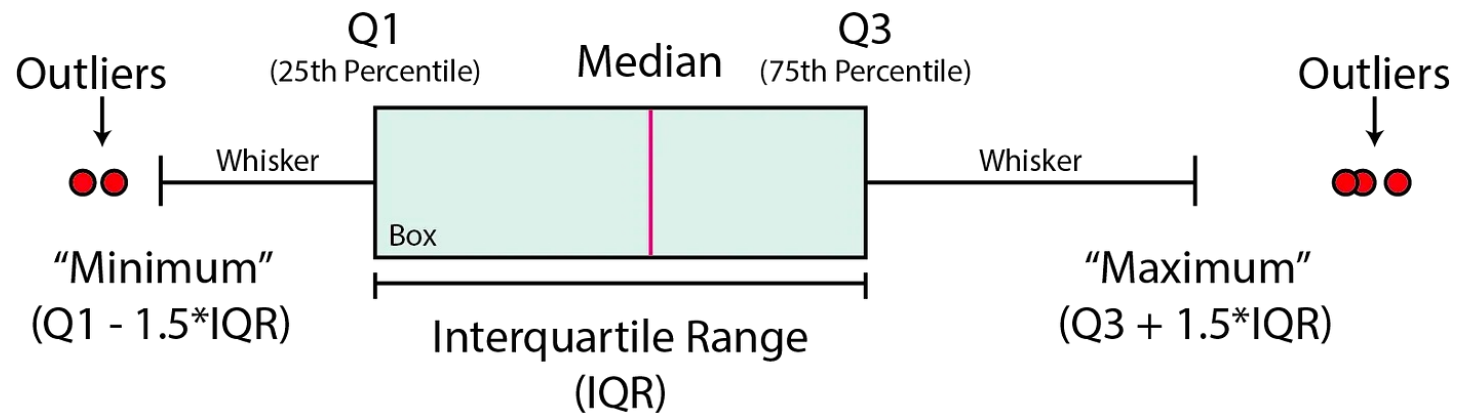
## Distributions

*probabilistic modeling of data uncertainty*



# Error Bars

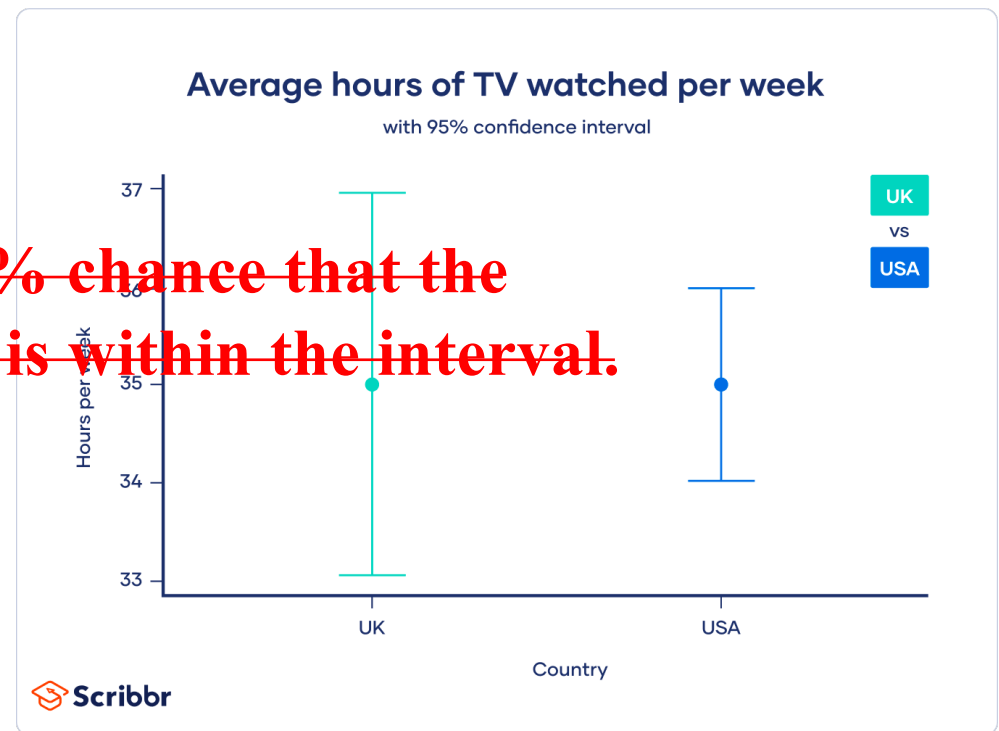
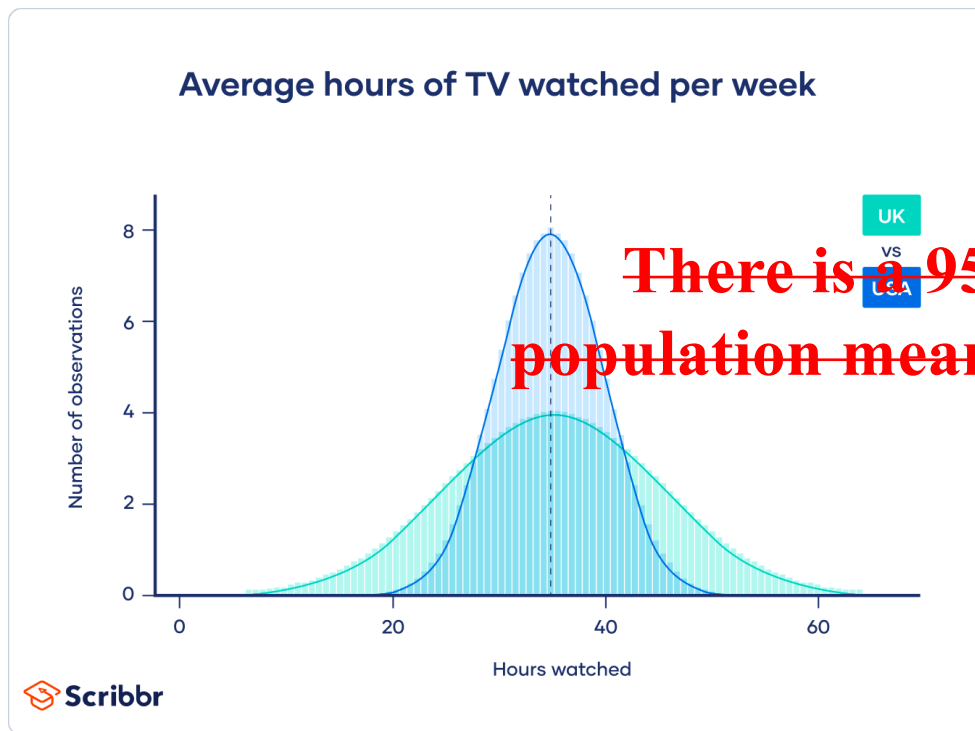
- Standard Deviation ( $\sigma$ ) Standard Error ( $\sigma / \sqrt{n}$ )
- $1.5 * \text{IQR}$  (Interquartile Range)
- Confidence Intervals
- ... and so on





# Confidence Intervals

- You survey 100 Brits and 100 Americans about their television-watching habits, and find that both groups watch an average of 35 hours of television per week.



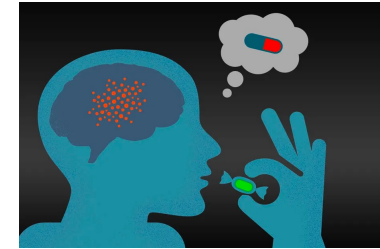
~~There is a 95% chance that the population mean is within the interval.~~

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

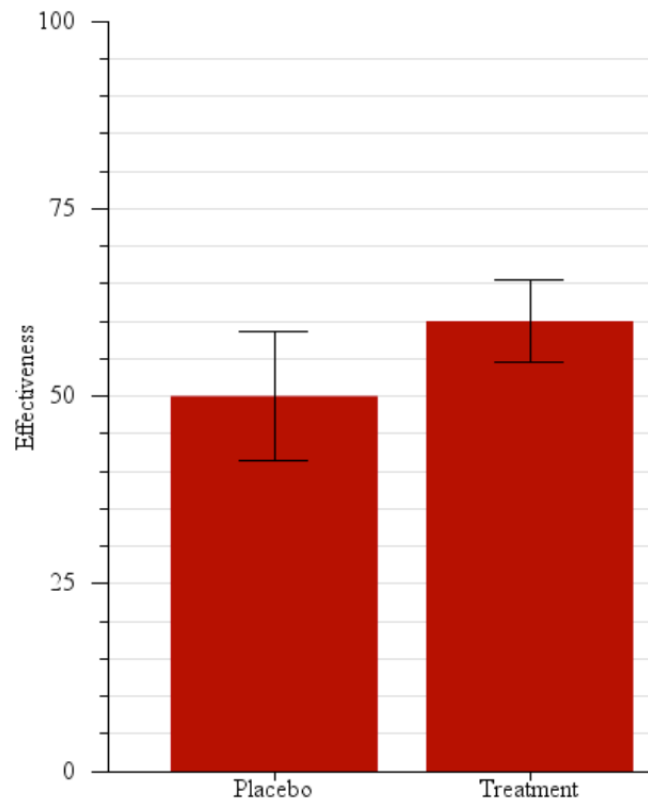
CI = confidence interval  
 $\bar{x}$  = sample mean  
 $z$  = confidence level value

$s$  = sample standard deviation  
 $n$  = sample size

## Error Bars

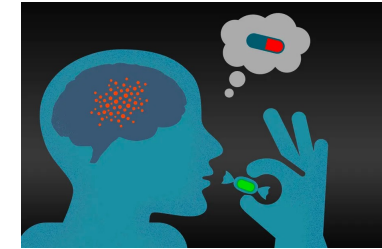


- Double-blind placebo control trials.
- Is this difference in means statistically significant?

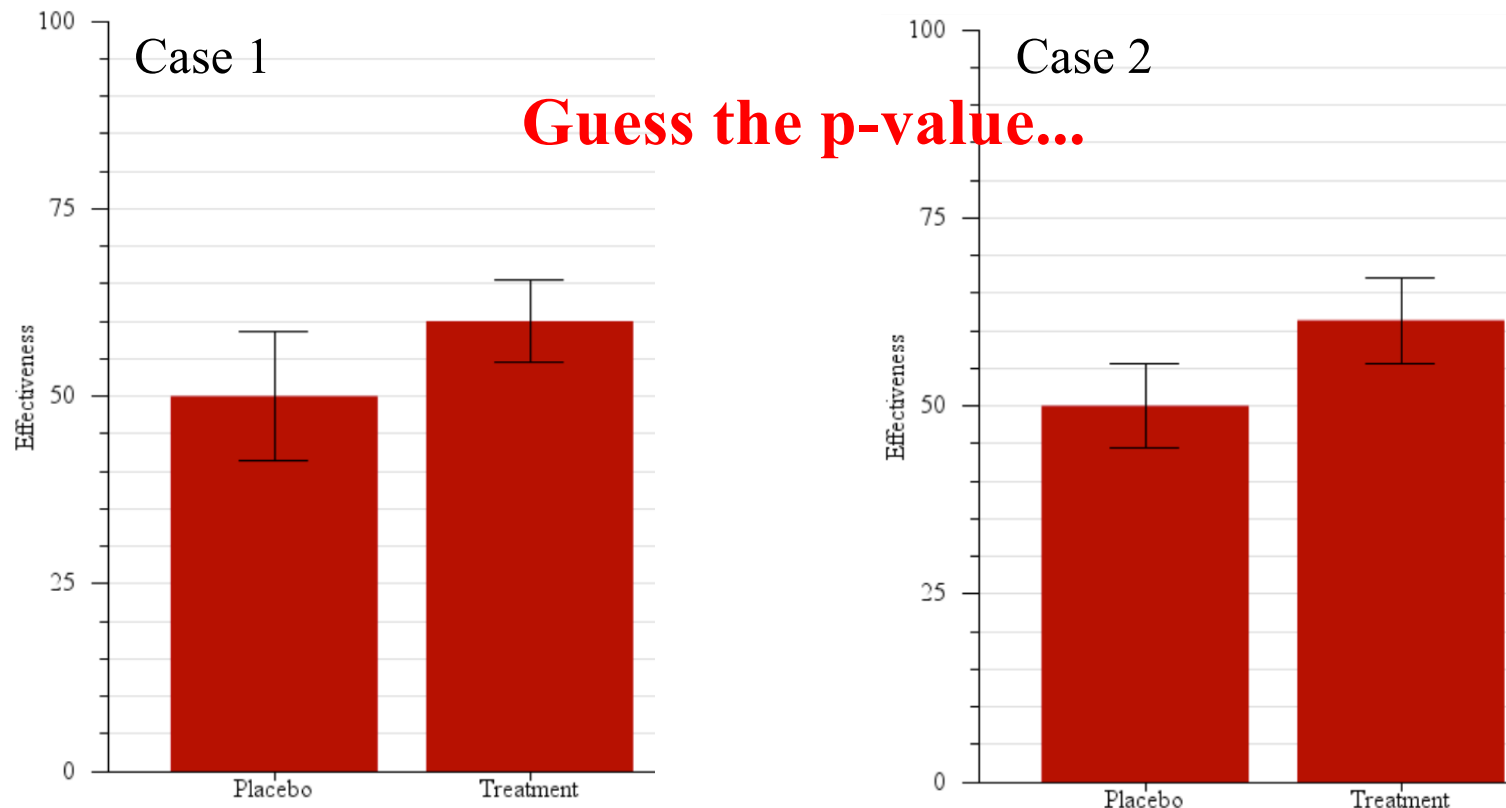


Error bars depict 95% Confidence Interval

# Error Bars

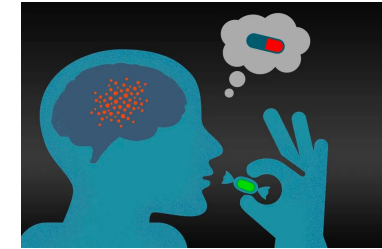


- Double-blind placebo control trials.
- Is this difference in means statistically significant?

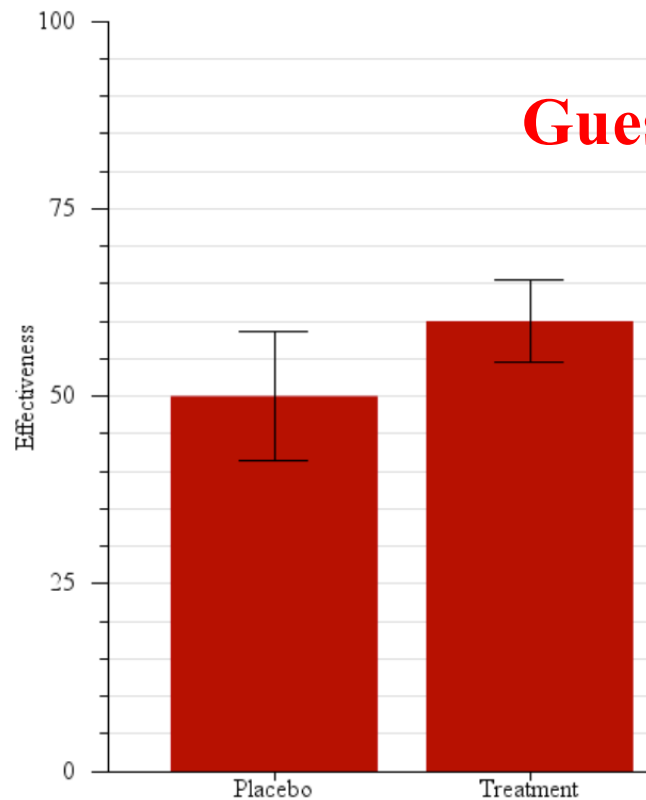


Error bars depict 95% Confidence Interval

## Error Bars

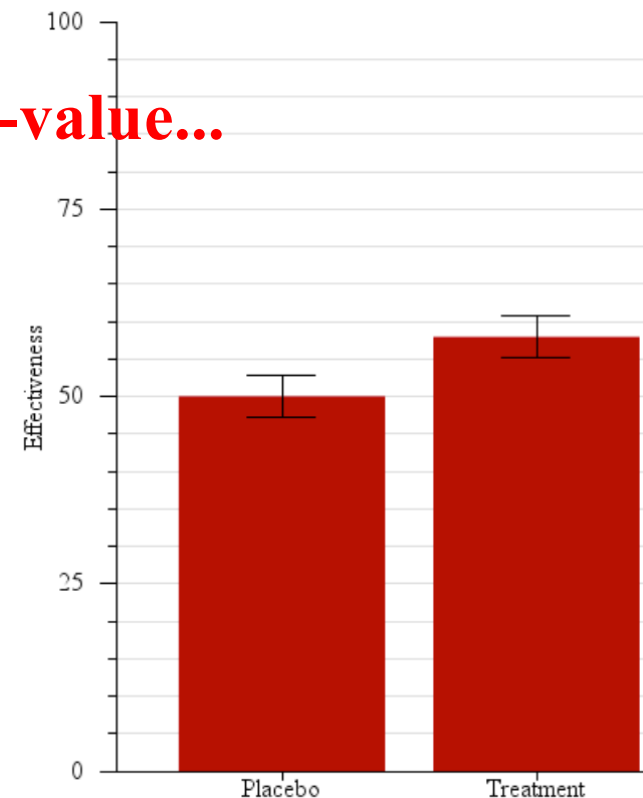


- Double-blind placebo control trials.
- Is this difference in means statistically significant?



Error bars depict 95% Confidence Interval

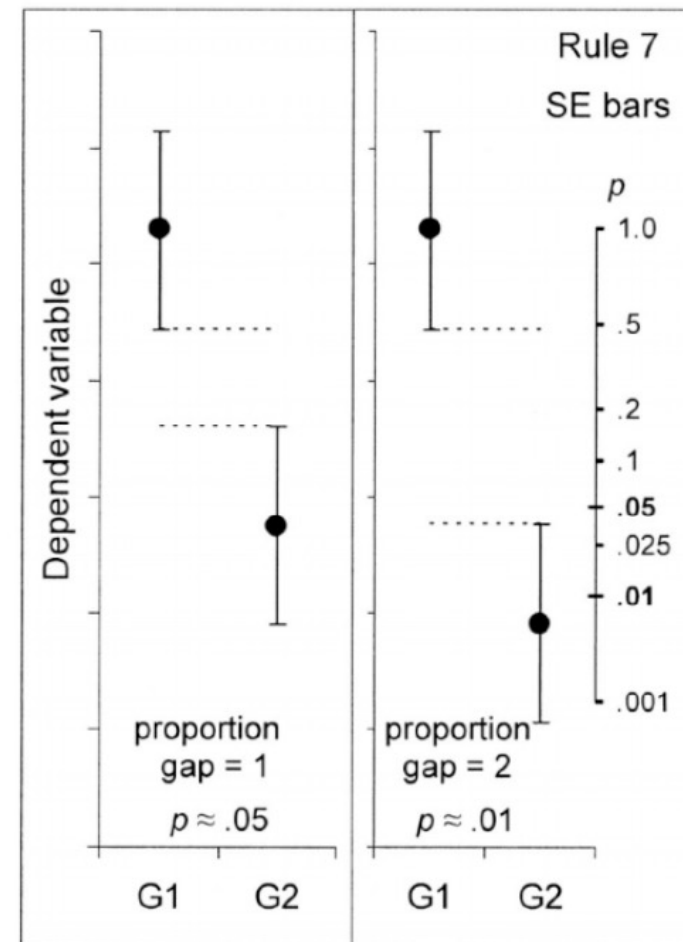
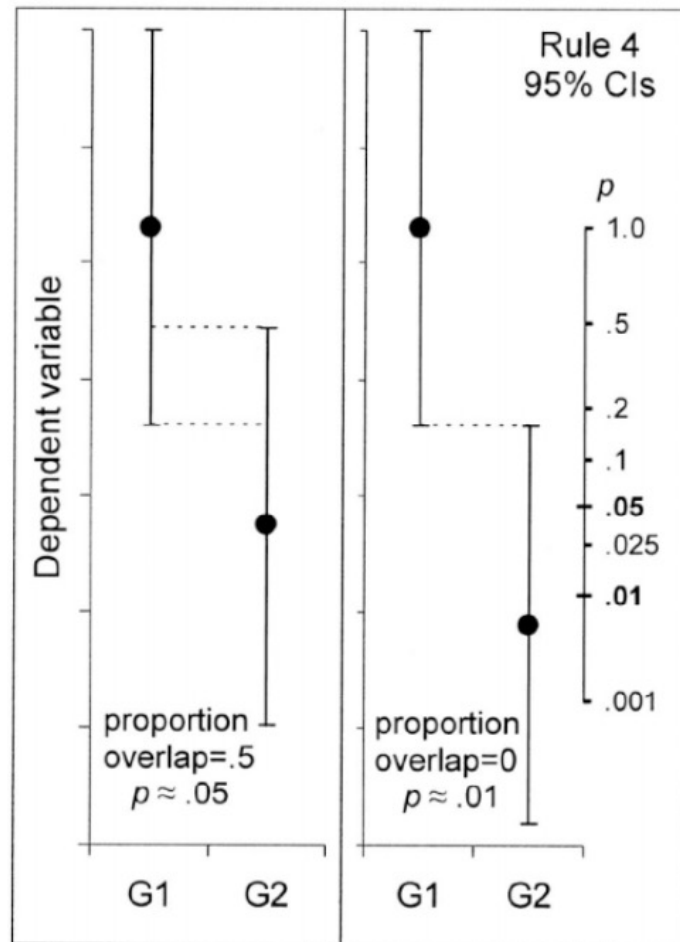
**Guess the p-value...**



Error bars depict standard error

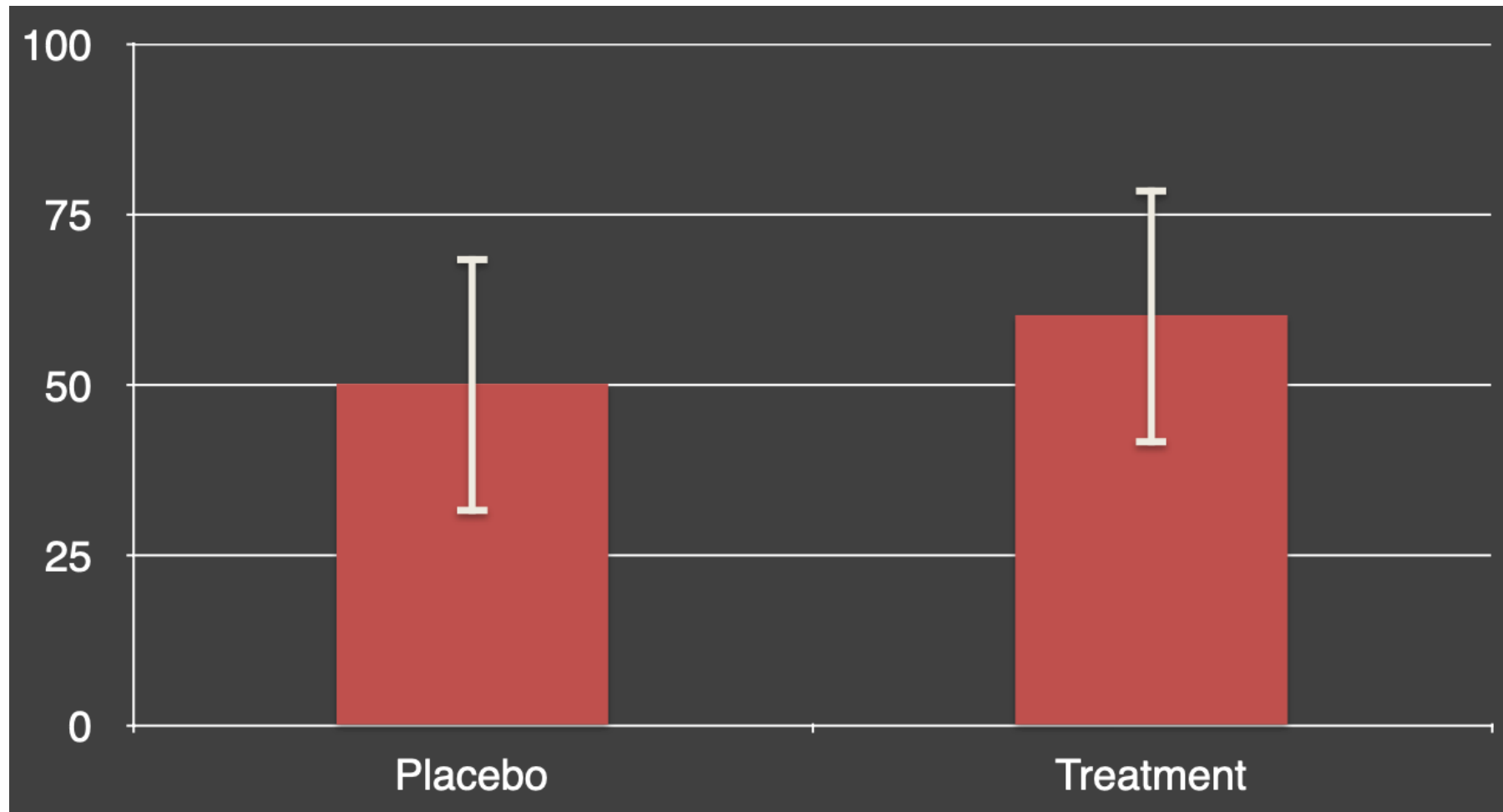
# Inference by Eye

95% CIs

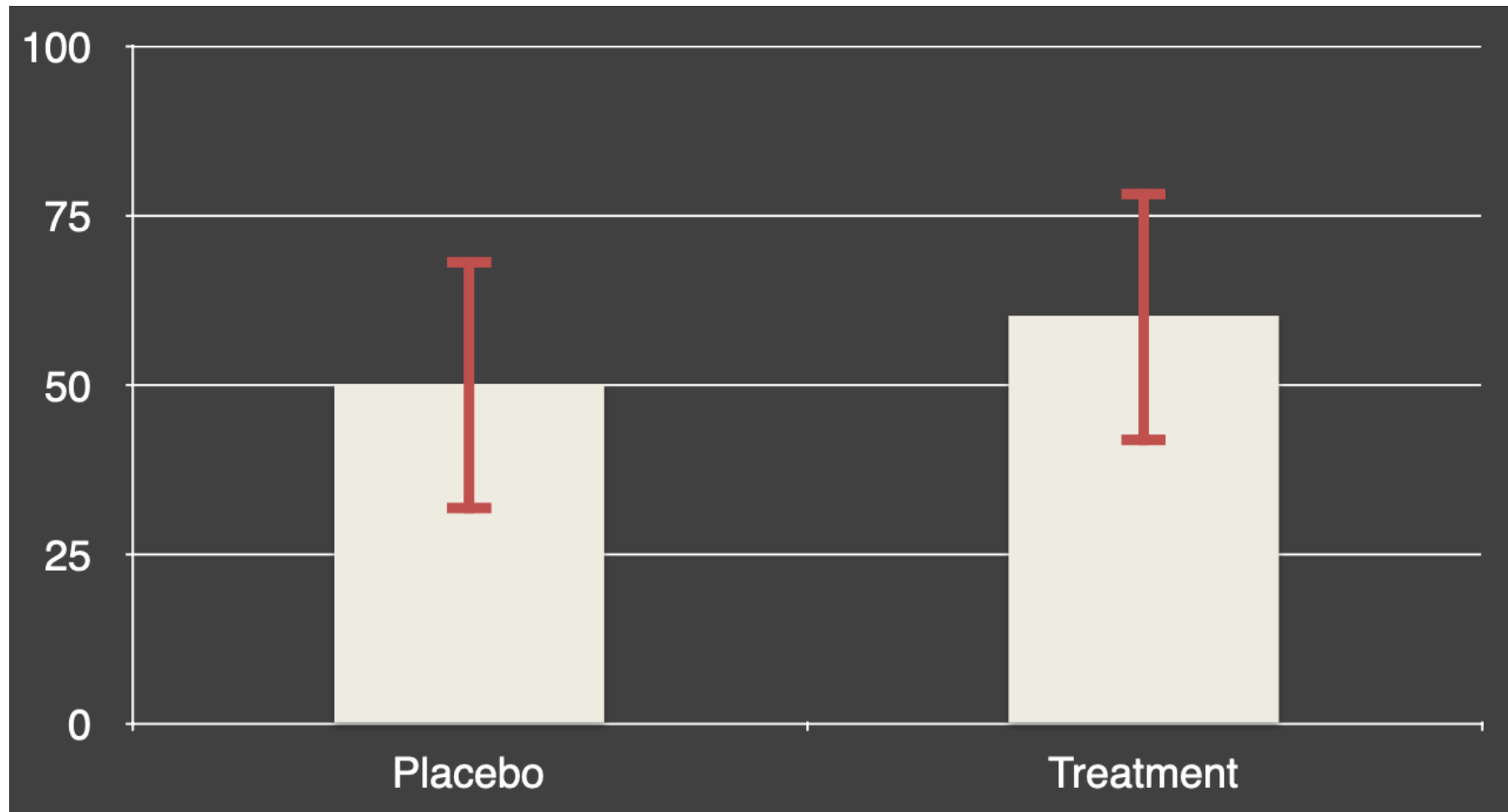


Cumming, Geoff and Finch, Sue. Inference by eye: confidence intervals and how to read pictures of data. American Psychologist, 2005.

# Misplaced Emphasis?

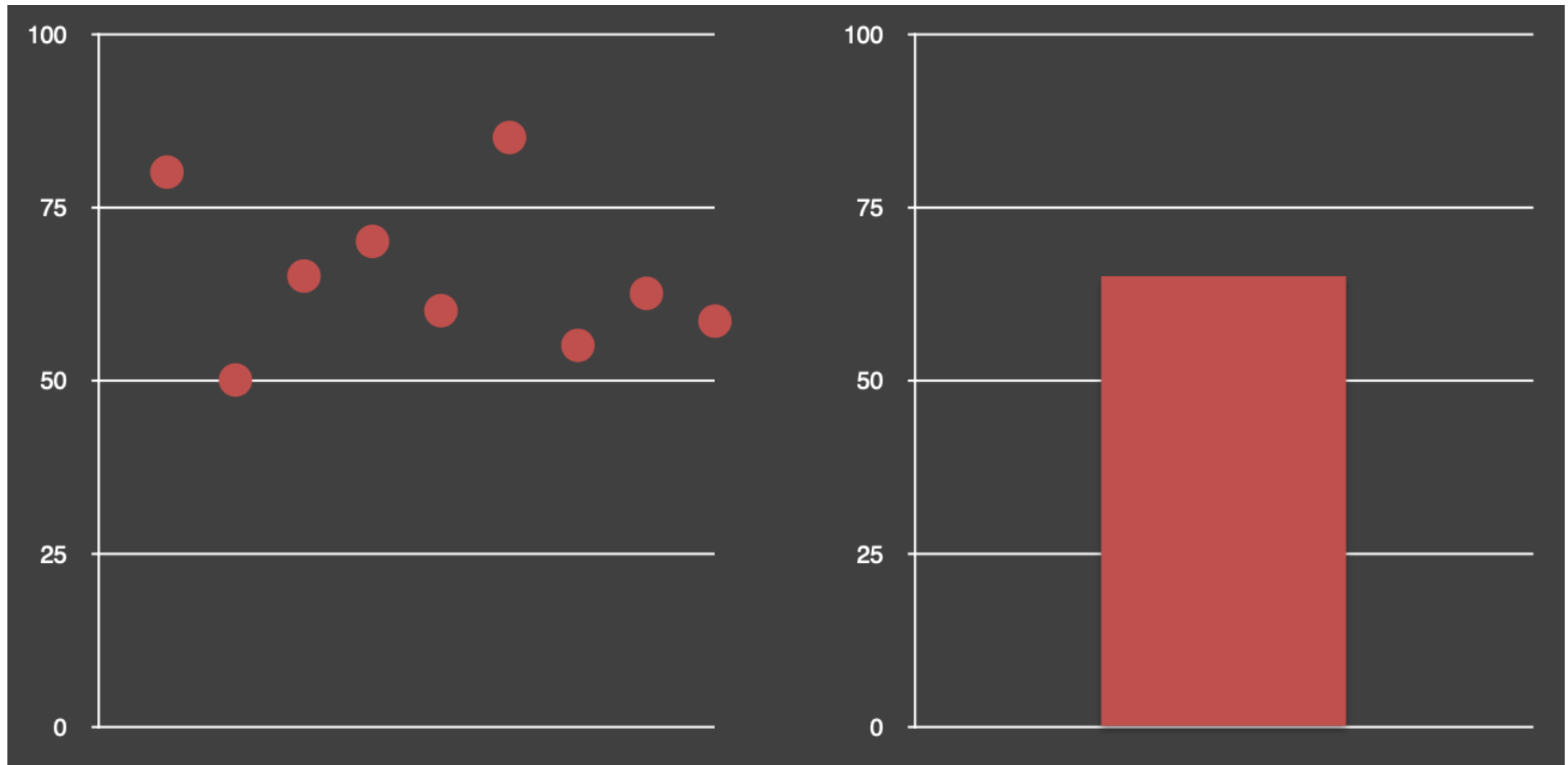


# Misplaced Emphasis?



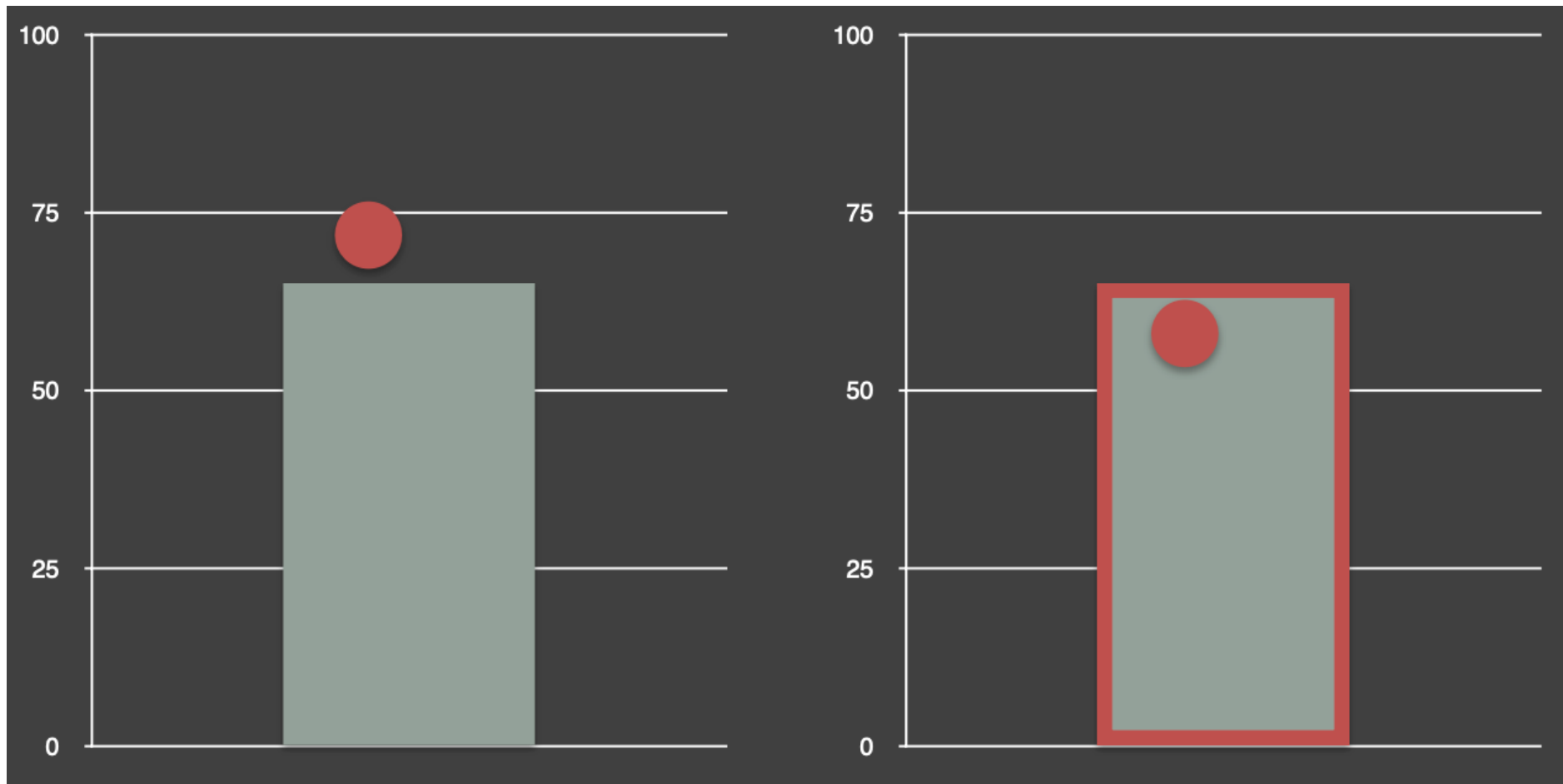


# Within-the-Bar Bias



Newman & Scholl. (2012) “Bar graphs depicting averages are perceptually misinterpreted: the within-the-bar bias.”

# Within-the-Bar Bias



Newman & Scholl. (2012) “Bar graphs depicting averages are perceptually misinterpreted: the within-the-bar bias.”

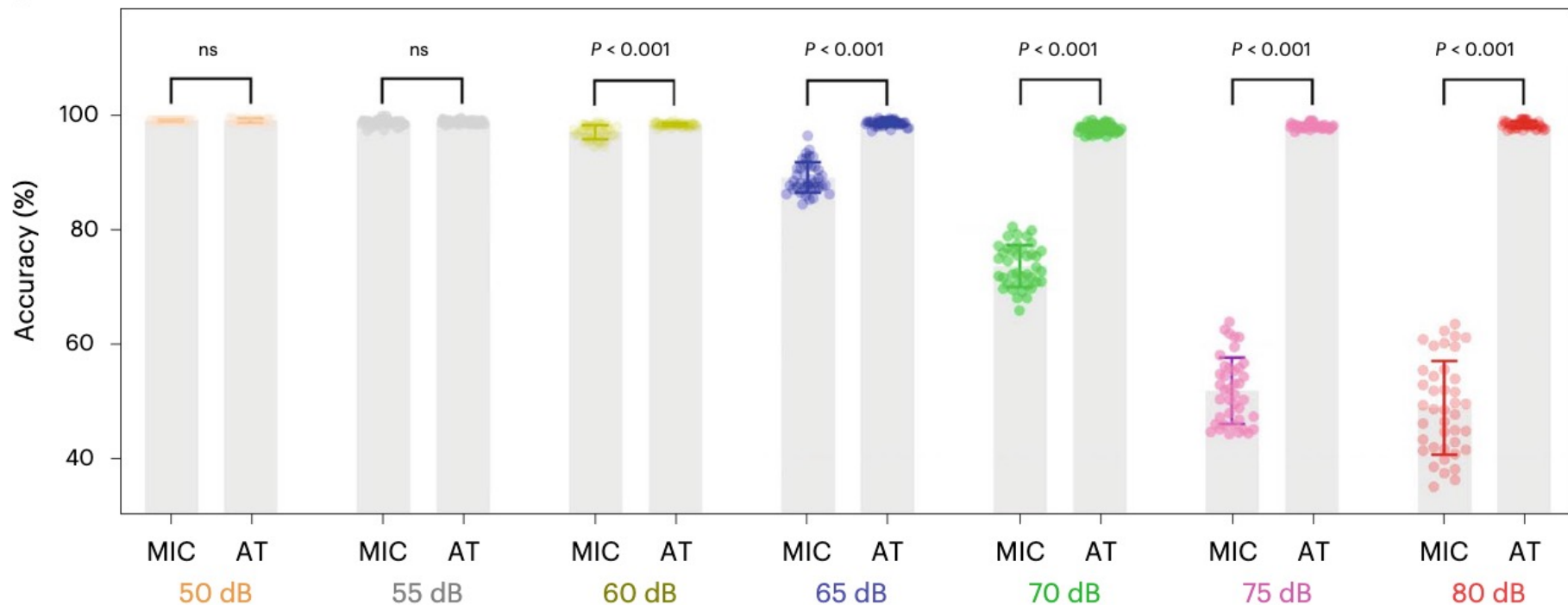
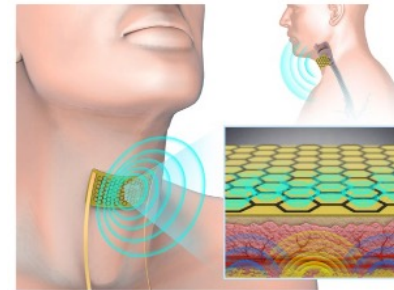
# Alternatives to Error Bars

nature machine intelligence

Article

<https://doi.org/10.1038/s42256-023-00616-6>

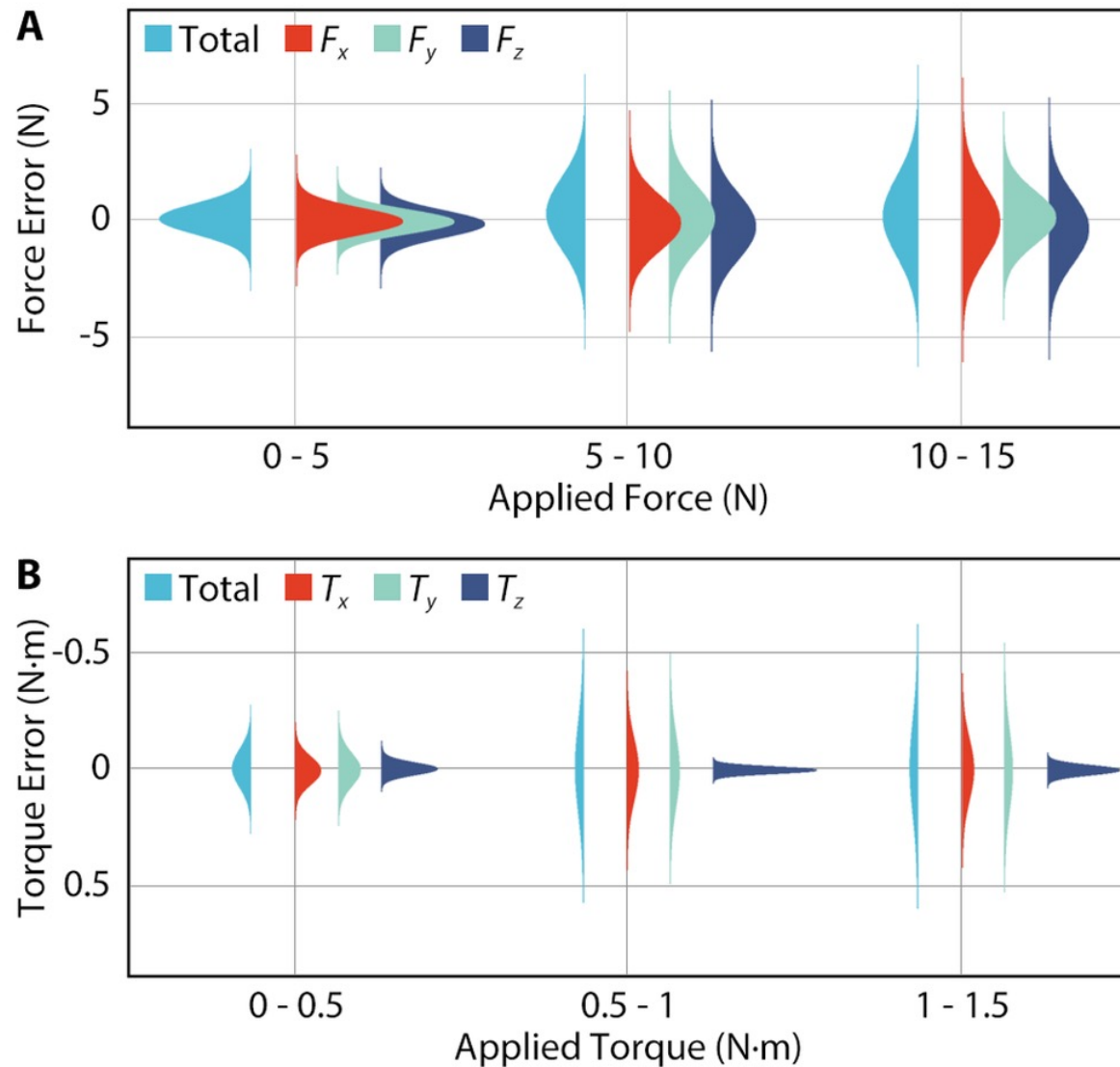
## Mixed-modality speech recognition and interaction using a wearable artificial throat



Comparisons of vowel recognition accuracy between the AT and a commercial microphone under different noise levels ( $n = 40$  independent experiments; error bars indicate s.d.).

# Alternatives to Error Bars

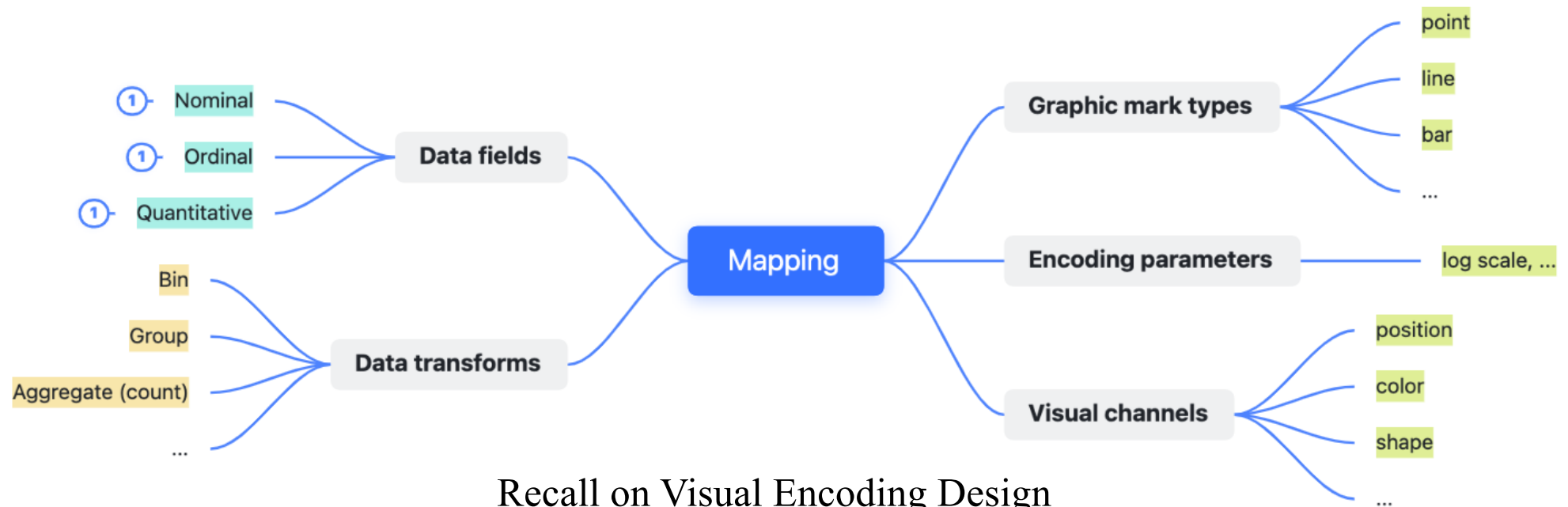
Violin Plot



For inference tasks,  
focus on the *uncertainty*  
not the point estimate!

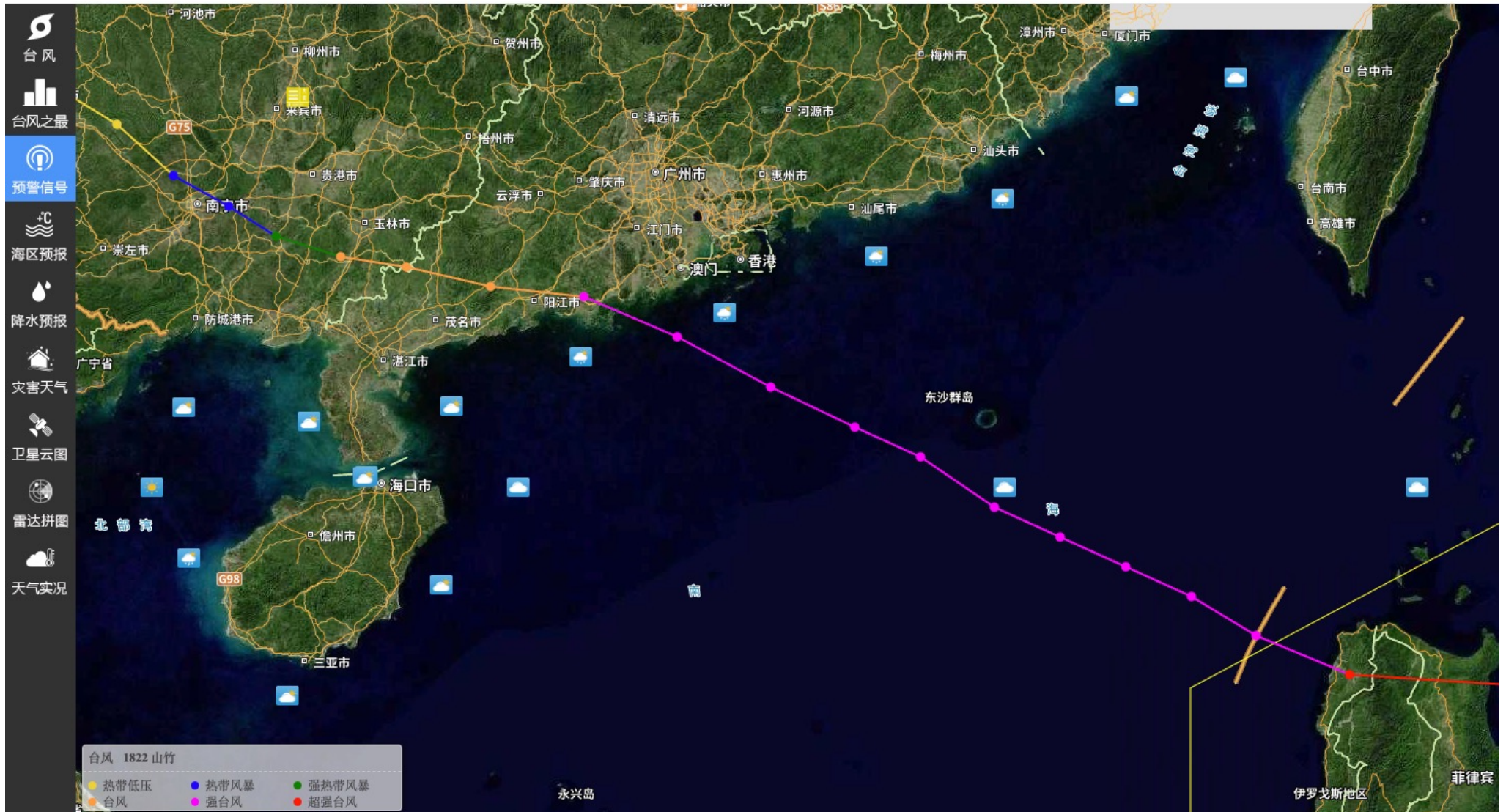
# Uncertainty Visualization Pipeline

1. Quantify Uncertainty
2. Visual Encode Uncertainty
3. Unify the Data Map and Uncertainty Map



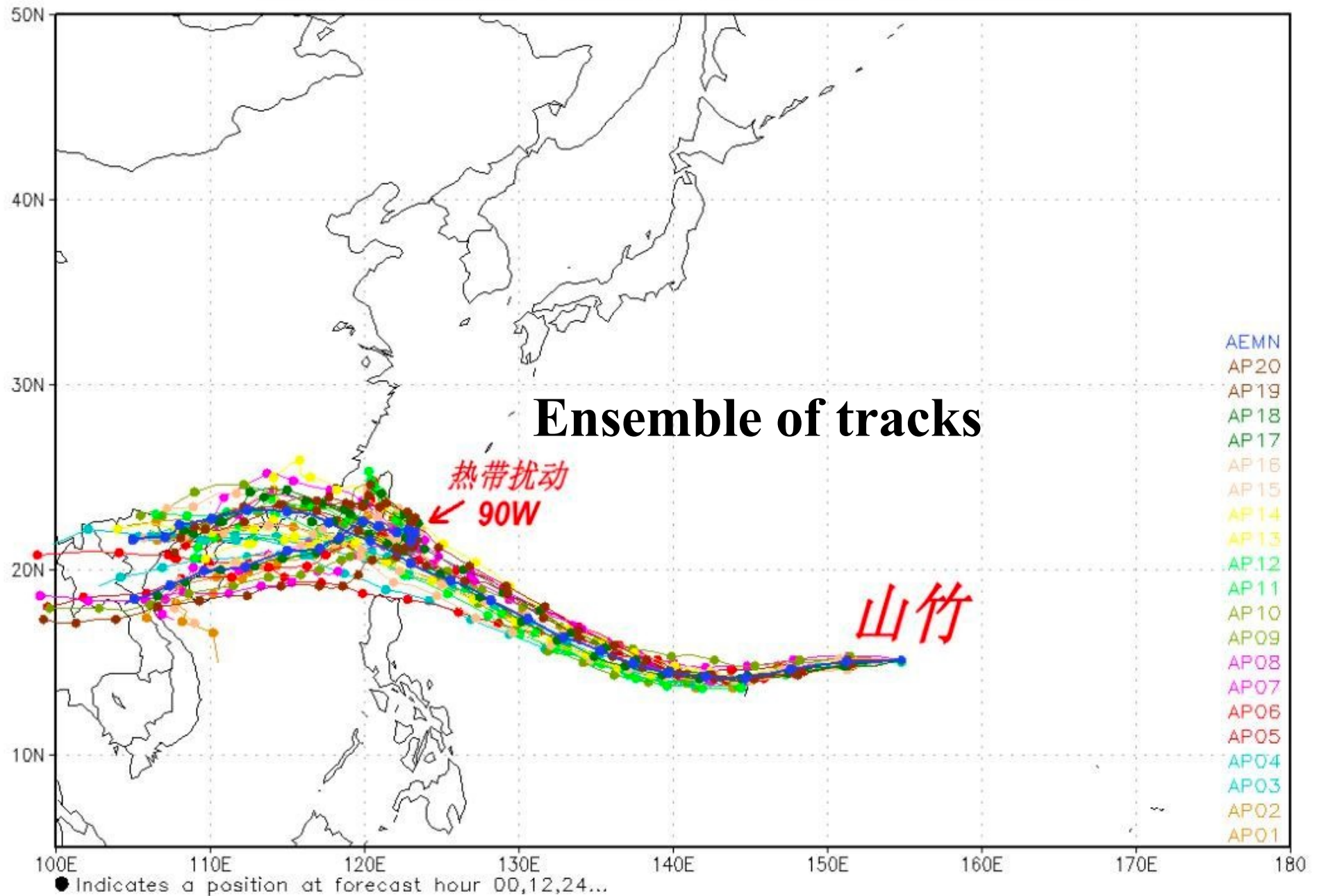
# Super Typhoon Mangkhut (山竹)

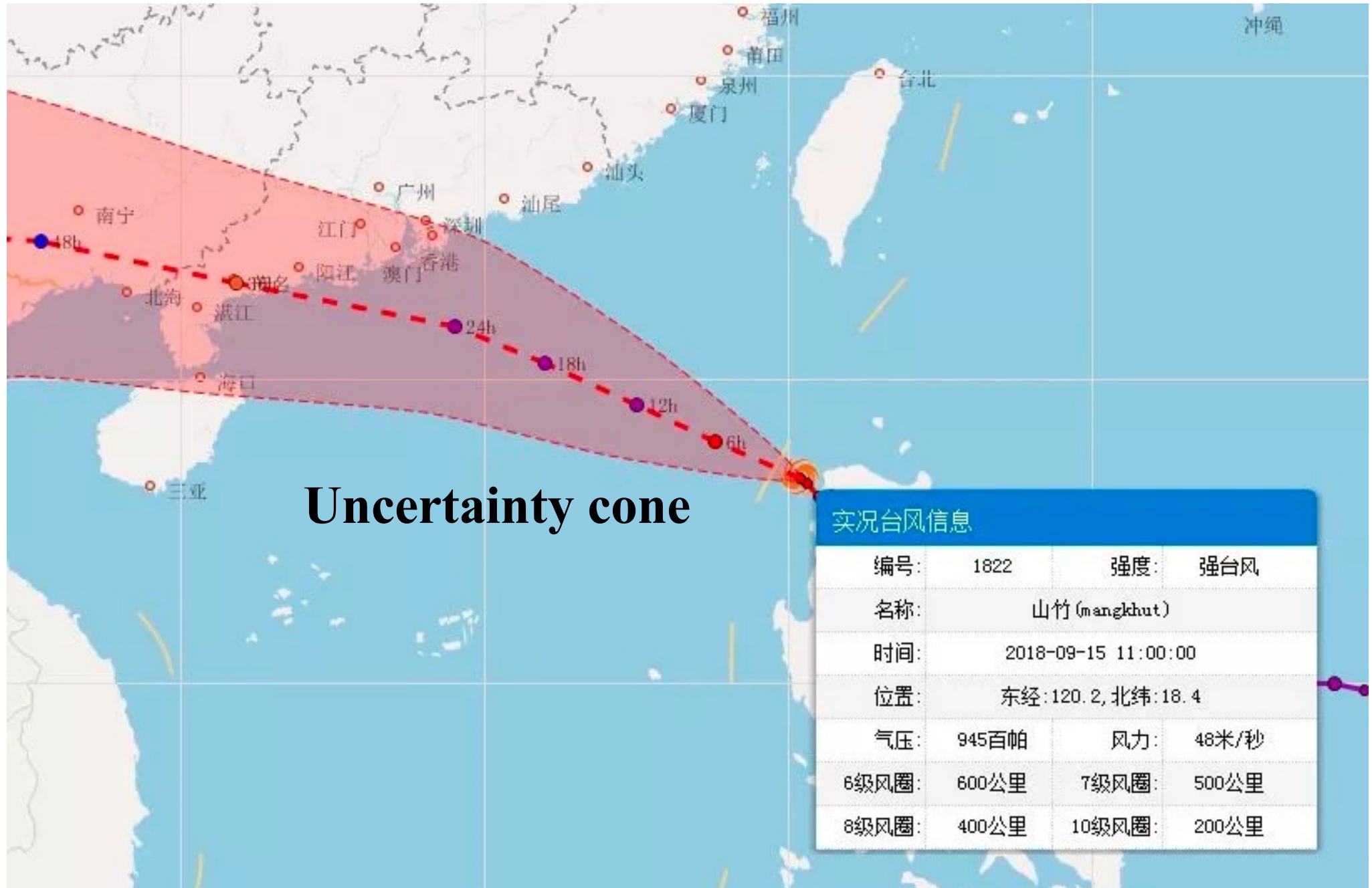
中央气象台·台风网 v3 <http://typhoon.nmc.cn/web.html>





## NCEP Ensemble Forecast TC Tracks 2018090900

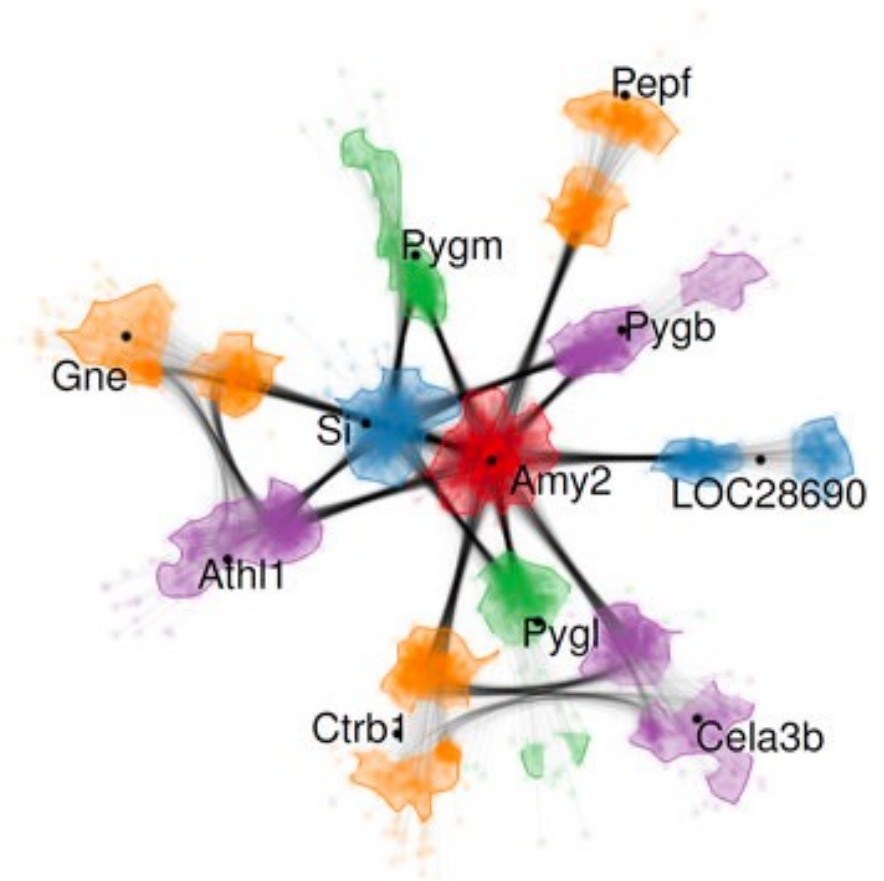
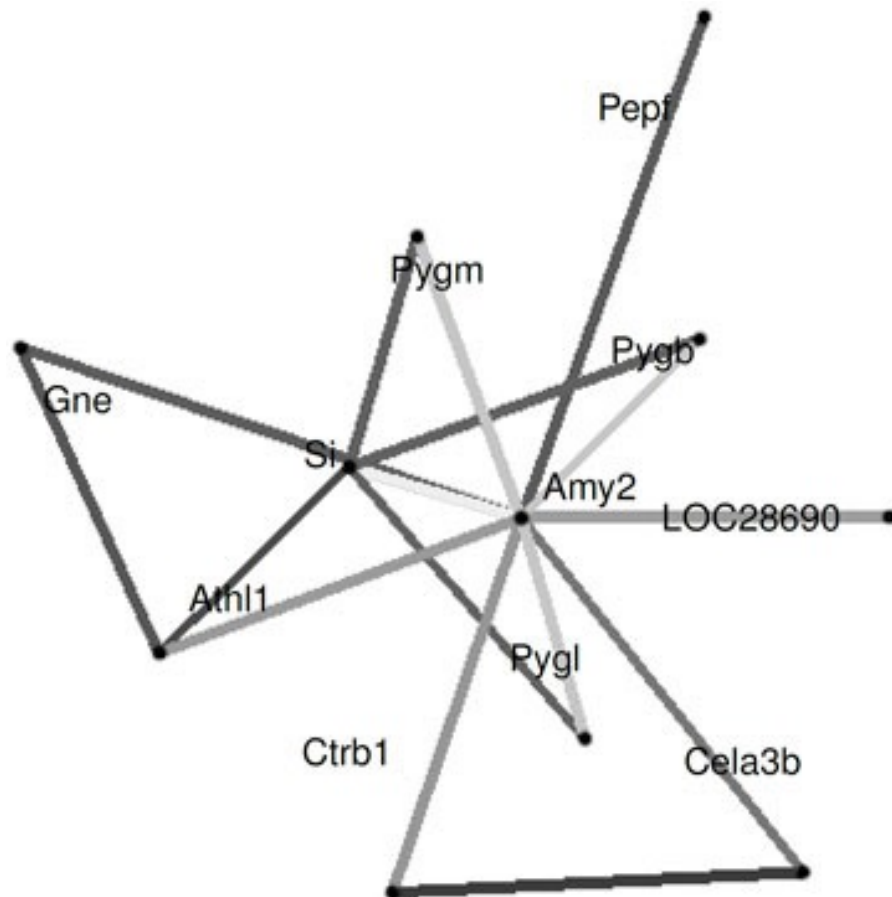




# Visualizing protein–protein interactions

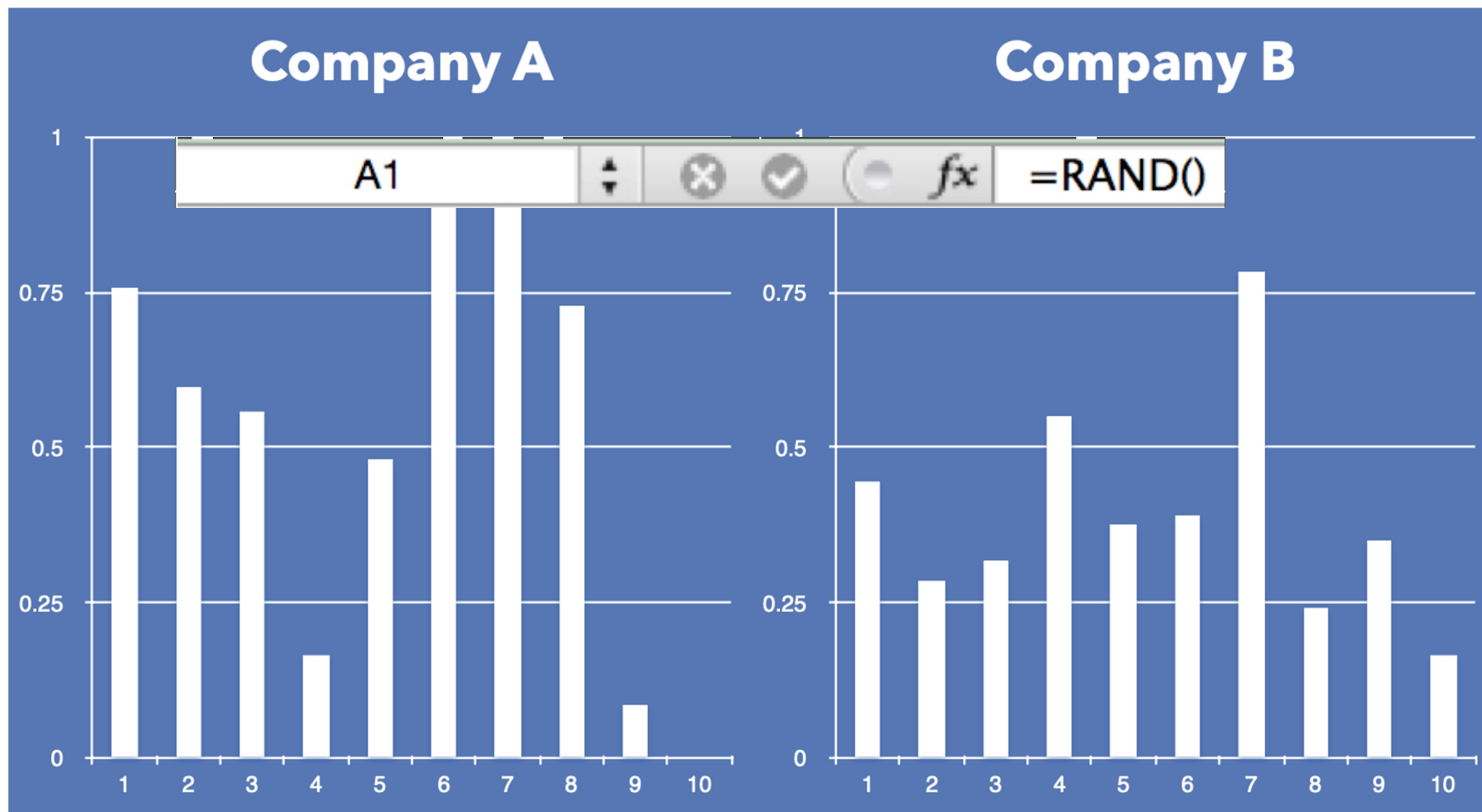
**Traditional non-uncertainty visualization**

**Uncertainty visualization**



# What Can Go Wrong?

- Which Stock To Buy? **Neither**



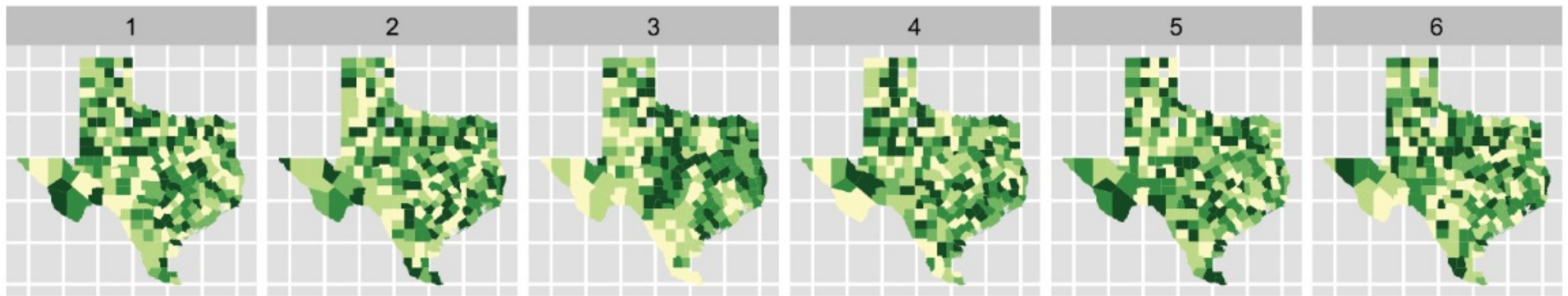


# Pareidolia (幻想性视错觉)



The tendency to perceive a specific, often meaningful image in a random or ambiguous visual pattern

# Choropleth maps of cancer deaths in Texas.



One plot shows a real data sets. The others are simulated under the null hypothesis of spatial independence.

Can you spot the real data? If so, you have some evidence of spatial dependence in the data.

# Negative Results

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

Negative or null results can correspond to weak and non-robust visual patterns across a model space.



---

# 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!**

---

# Conclusion

- There are different types and sources of uncertainty associated with data.
- We can quantify or model our uncertainty.
- The visual presentation of uncertainty can clash with cognitive and perceptual biases.



# DS363: Design and Learning with Data

---

<https://ds363.ancorasir.com/>

**Thank you~**

Wan Fang  
Southern University of Science and Technology