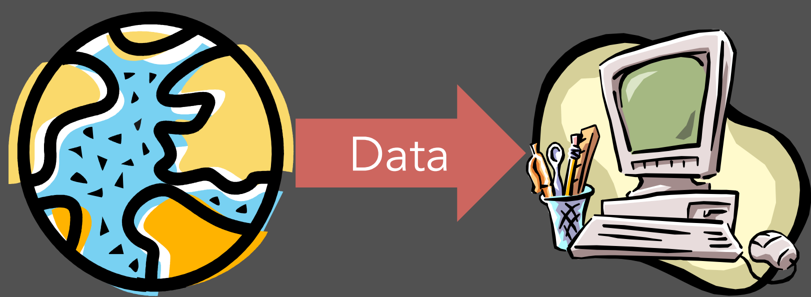
# CSE 512 – Data Visualization
# Uncertainty

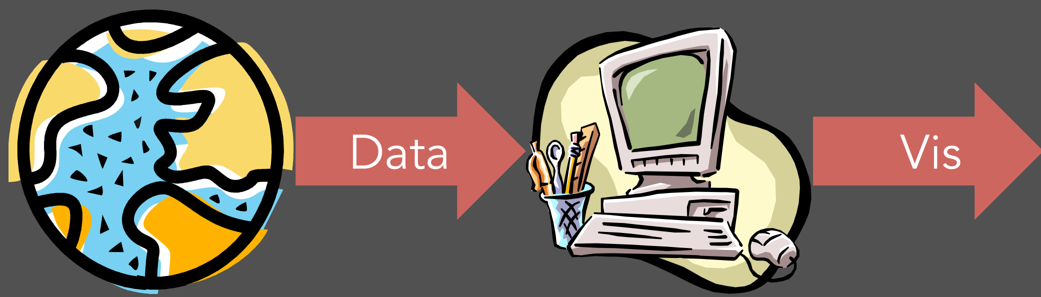

Michael Correll University of Washington

# The Visualization Pipeline




Collection/Curation

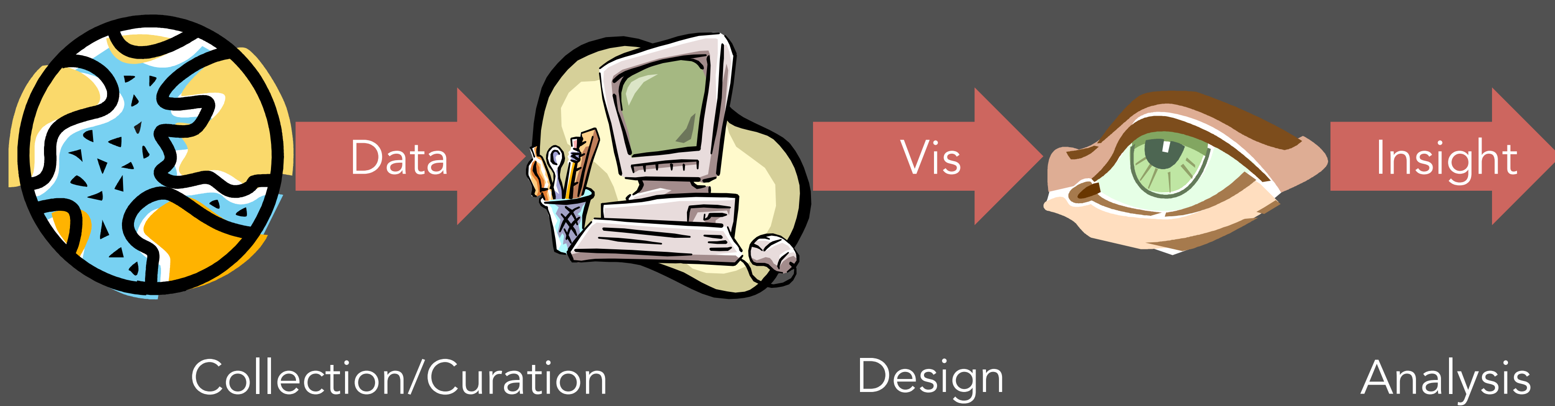
# The Visualization Pipeline



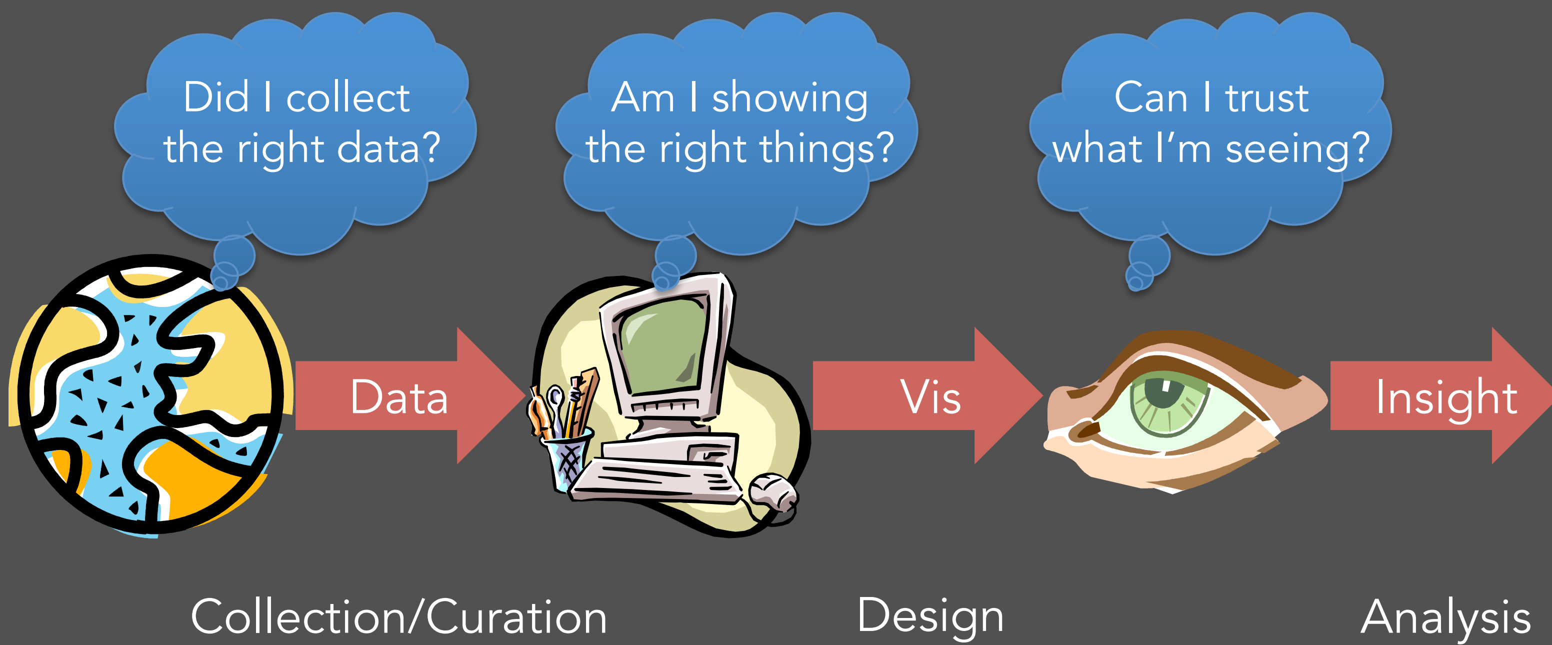

Collection/Curation

Design

# The Visualization Pipeline



# The Visualization Pipeline?



# Unknown Unknowns

# Things “Uncertainty” Can Mean

Doubt

Risk

Variability

Error

Lack of Knowledge

Hedging

…

# 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.

# Terminology

# Terminology

Aleatory Uncertainty

Epistemic Uncertainty

Type I error

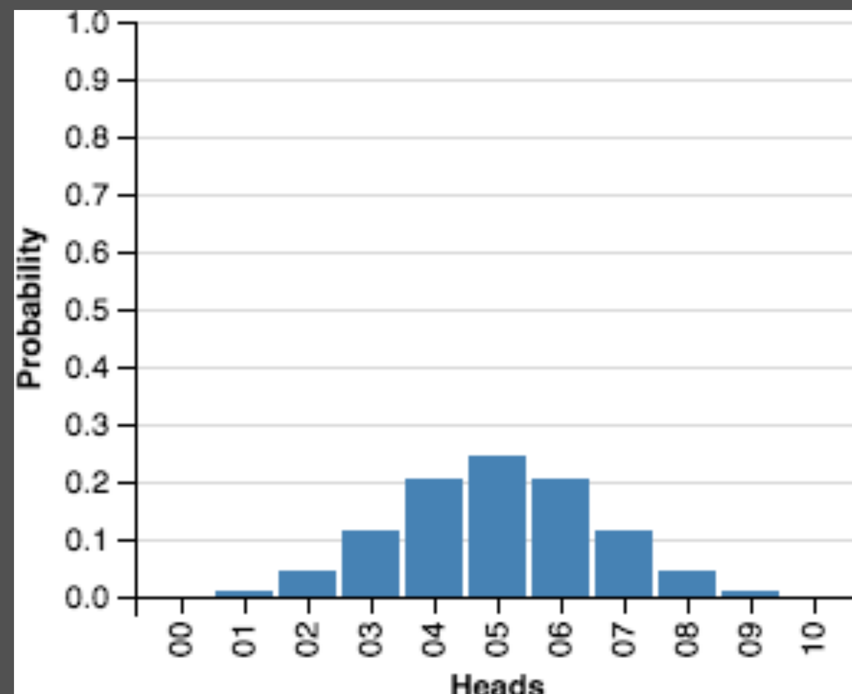
Type II error

Precision

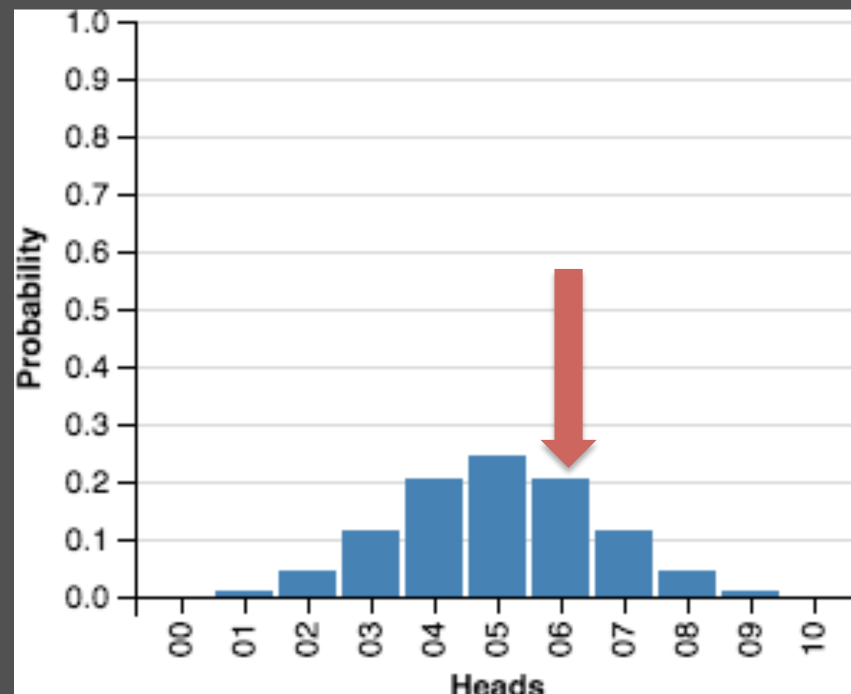
Bias

# Aleatory Uncertainty

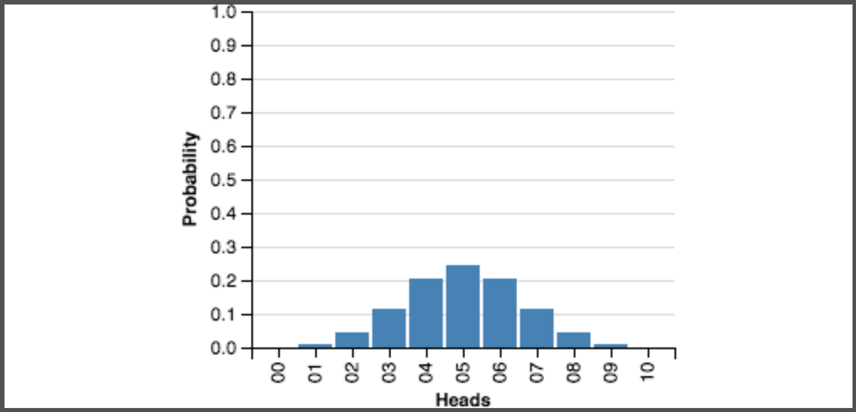
# Aleatory Uncertainty

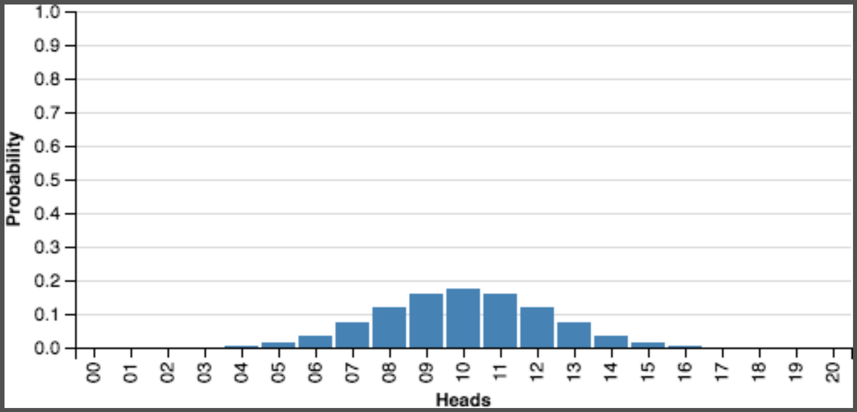


# Aleatory Uncertainty

# Aleatory Uncertainty



# John Edmund Kerrich

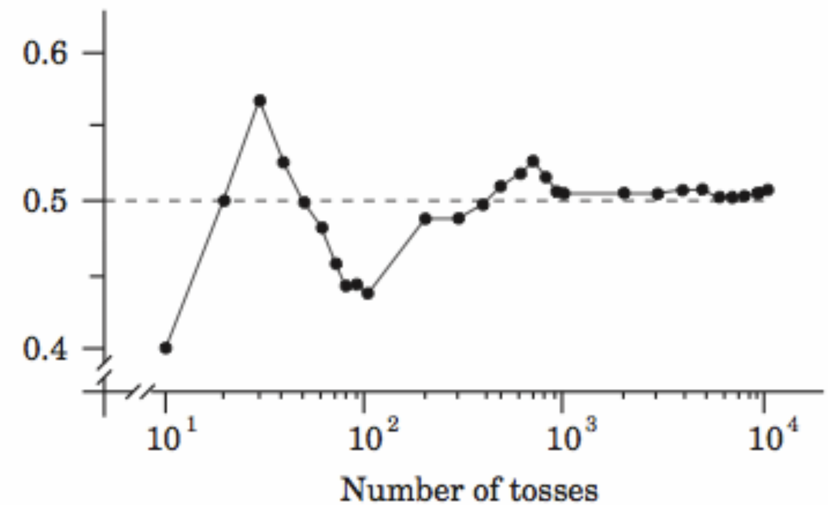
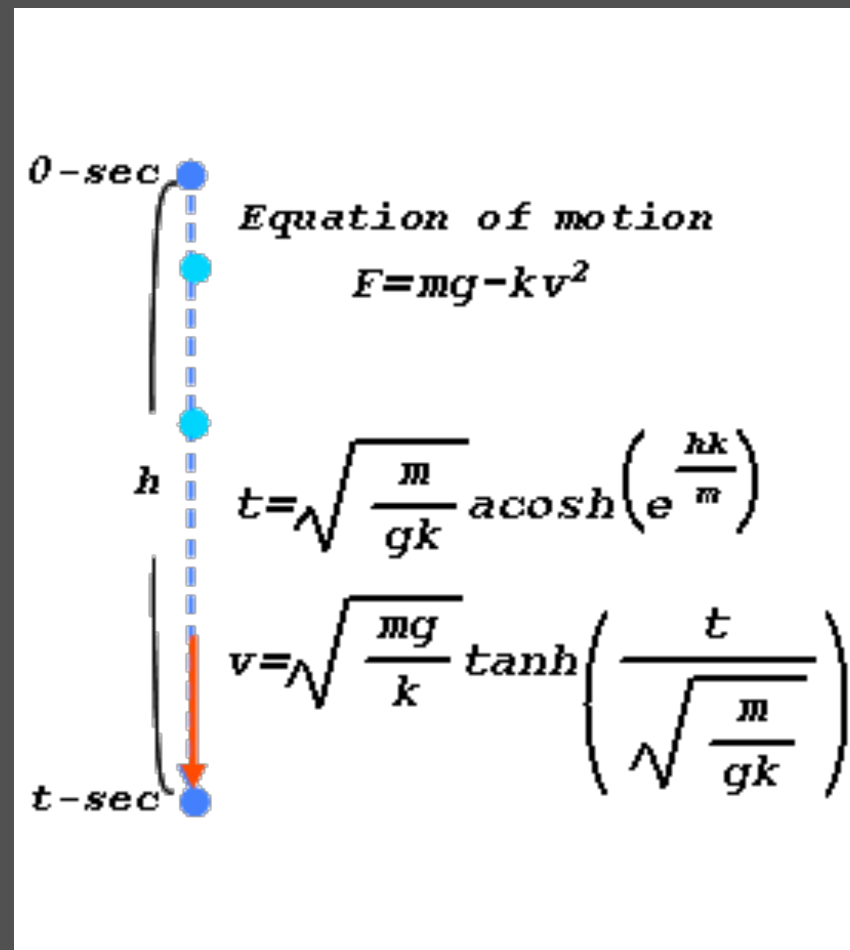



**FIGURE 4.1.1** Proportion of heads versus number of tosses for John Kerrich's coin-tossing experiment.

# Epistemic Uncertainty



# Uncertainty Types

**Aleatory**

Variability: things that we don't know (but can reason about the likelihood of).

**Epistemic**

Things we could in principle know for certain, but have not measured.

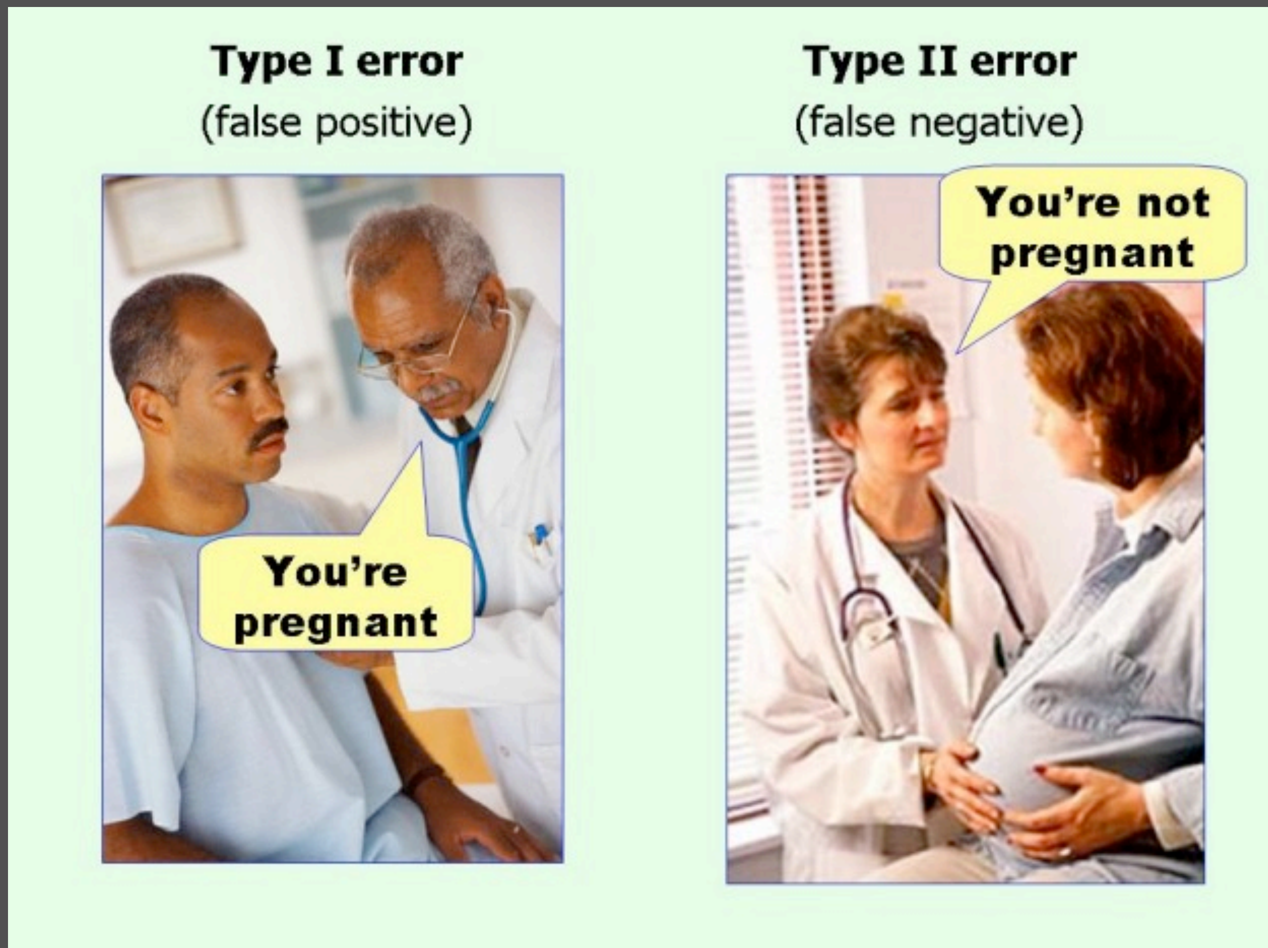
# Should I Bring an Umbrella?

# Decision Uncertainty

"50% Chance of Rain"

# Risk and Error

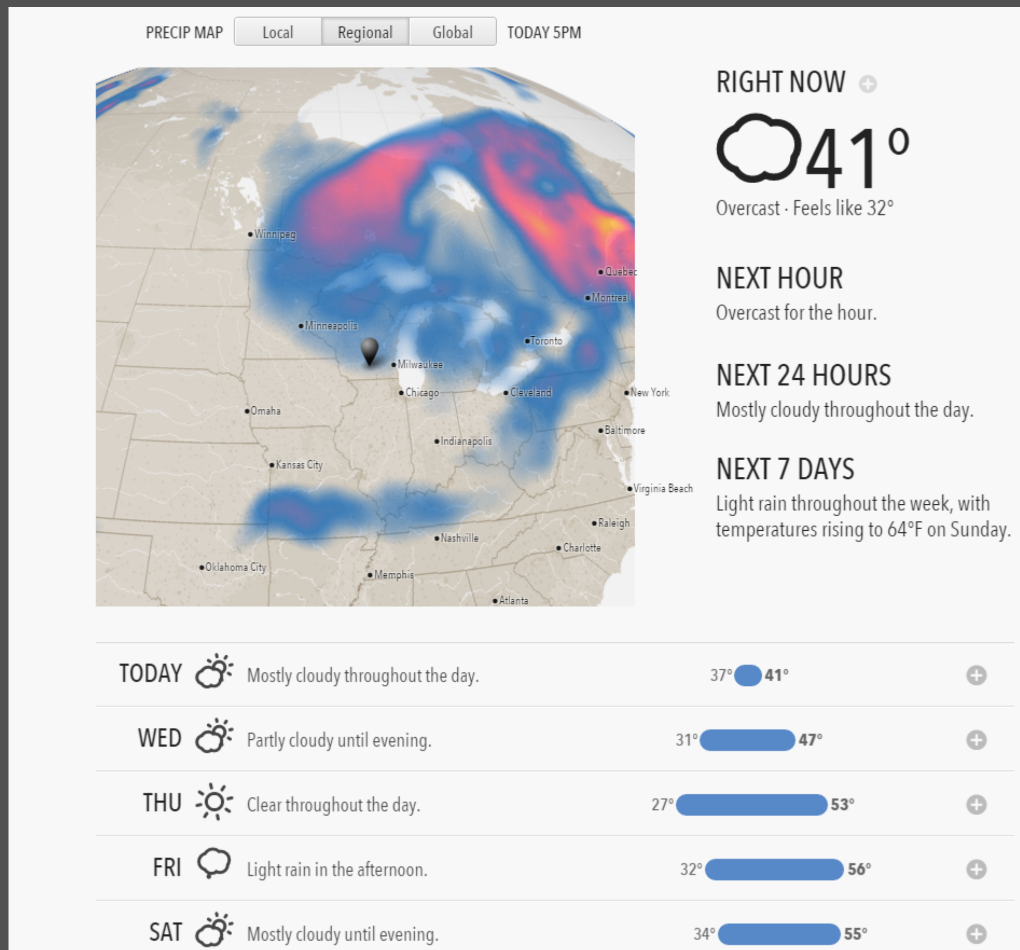
# Type I and Type II



# Model Uncertainty

"50% Chance of Rain"

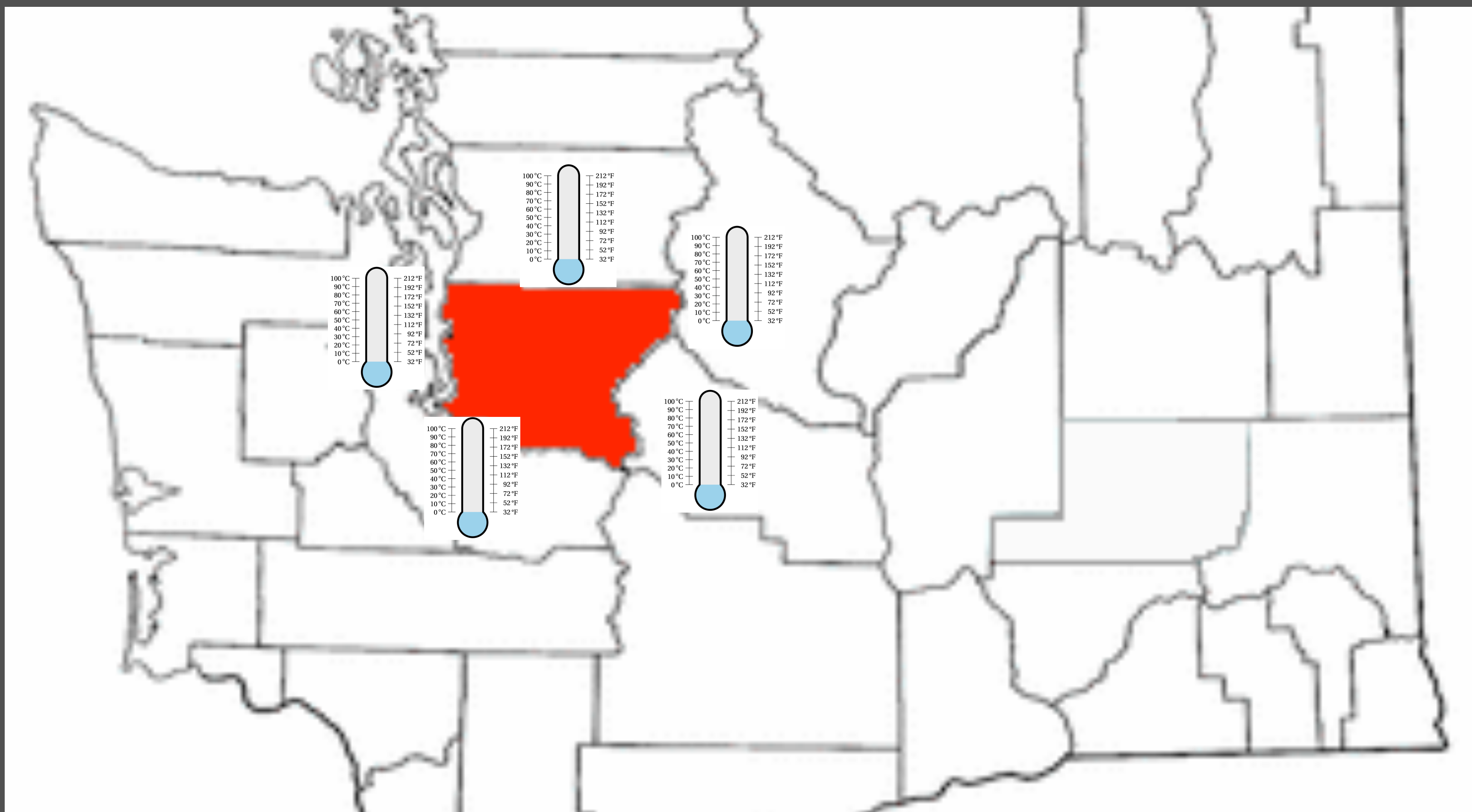
# Model Uncertainty



# Model Uncertainty

# Measurement Uncertainty

# Measurement Uncertainty



# Measurement Uncertainty

**Accuracy**

# Measurement Uncertainty

**Accuracy**

# Measurement Uncertainty

**Accuracy**

**Precision**

# Measurement Uncertainty

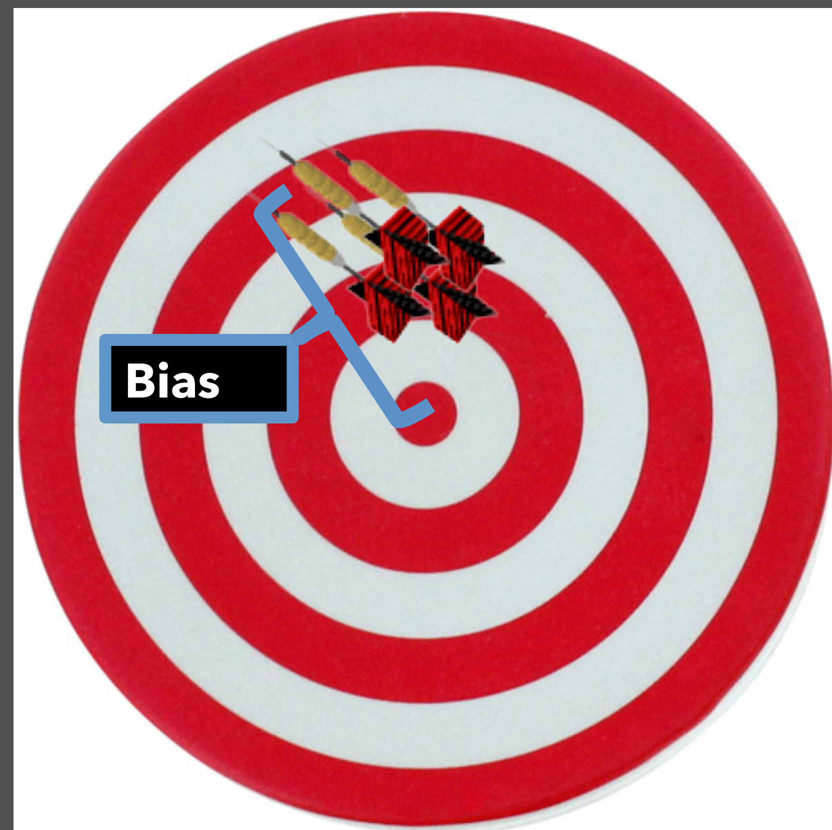
**Accuracy**

**Precision**

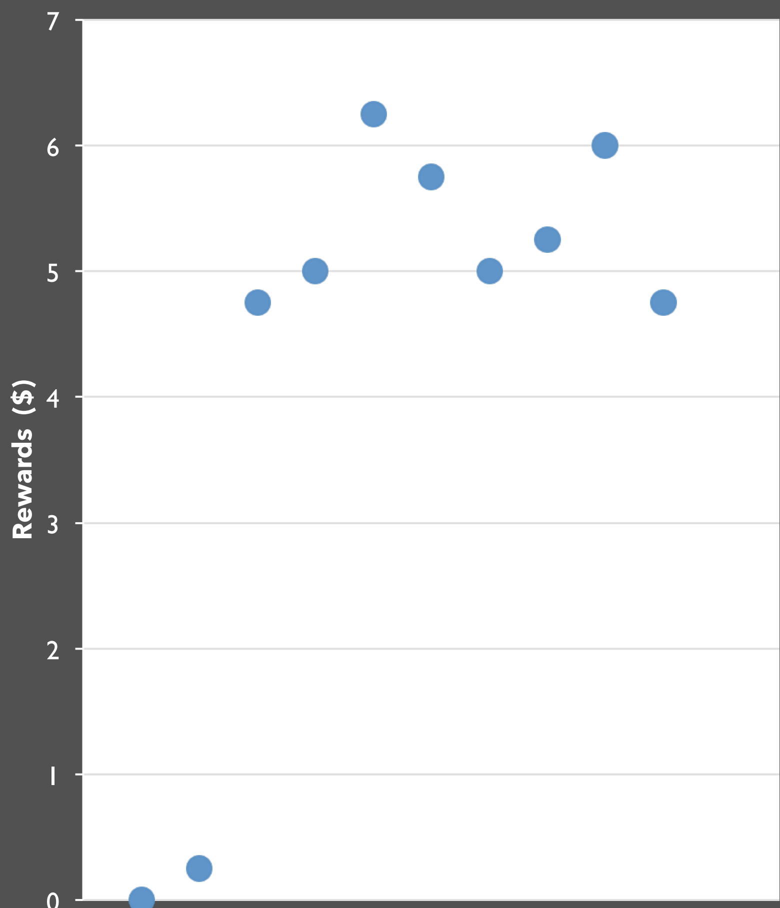
# Measurement Uncertainty

**Accuracy**

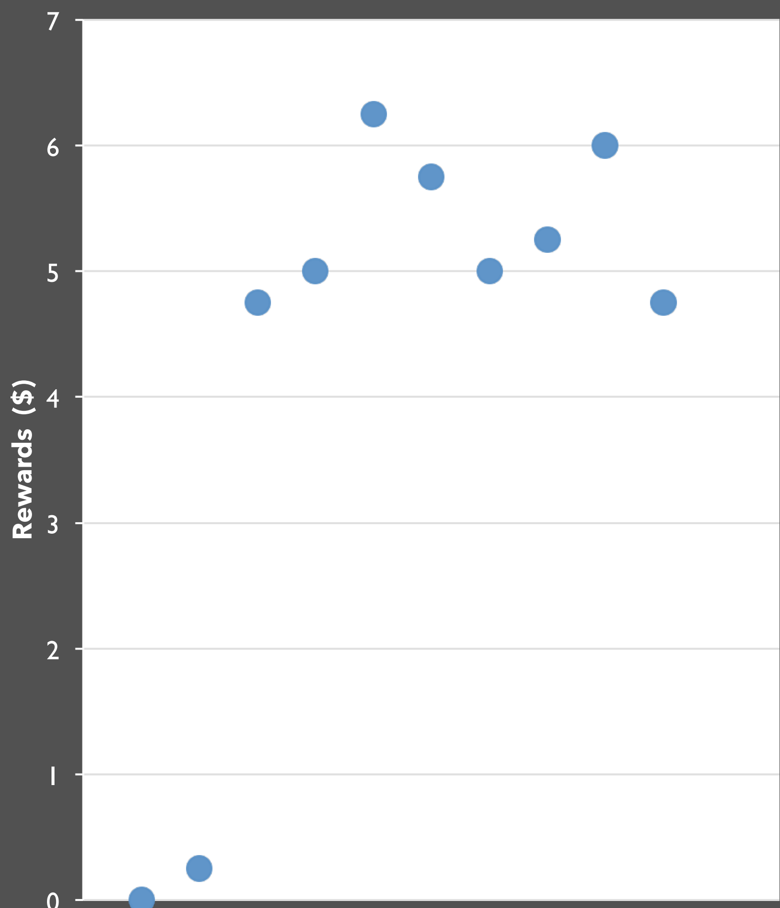
**Precision**



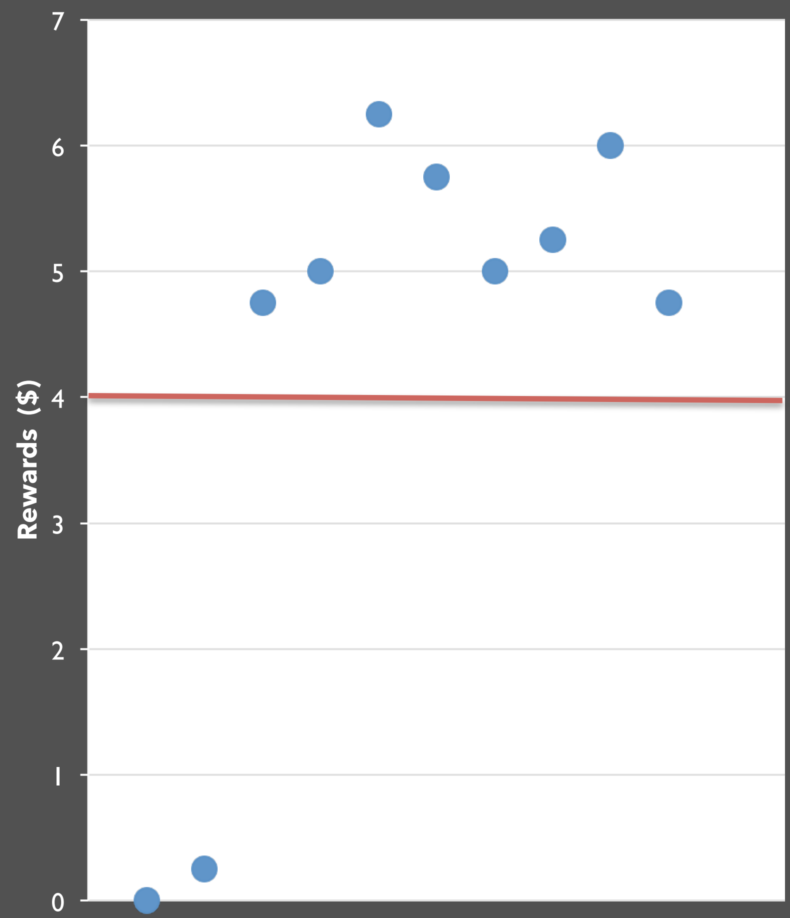
# Should you take this $4 bet?



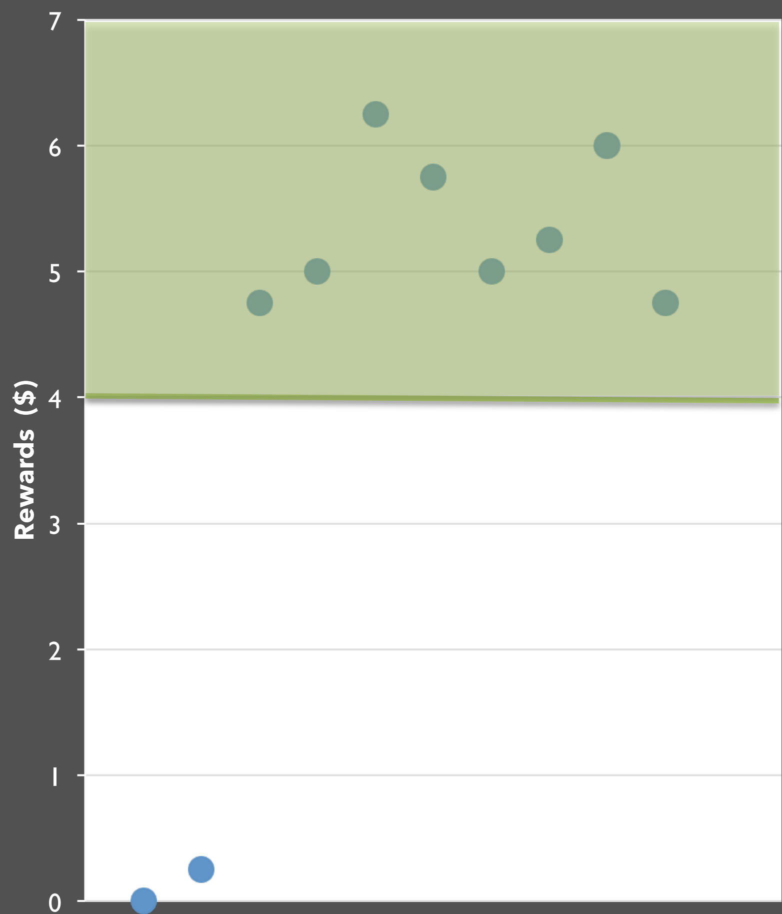
# Samples



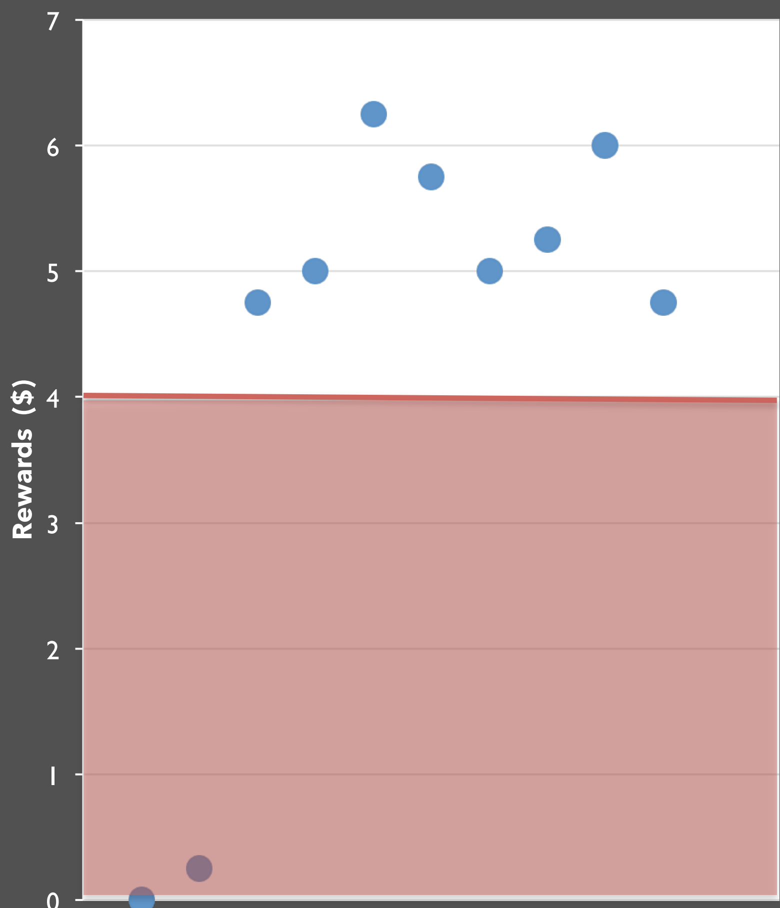
# Should you take this $4 bet?



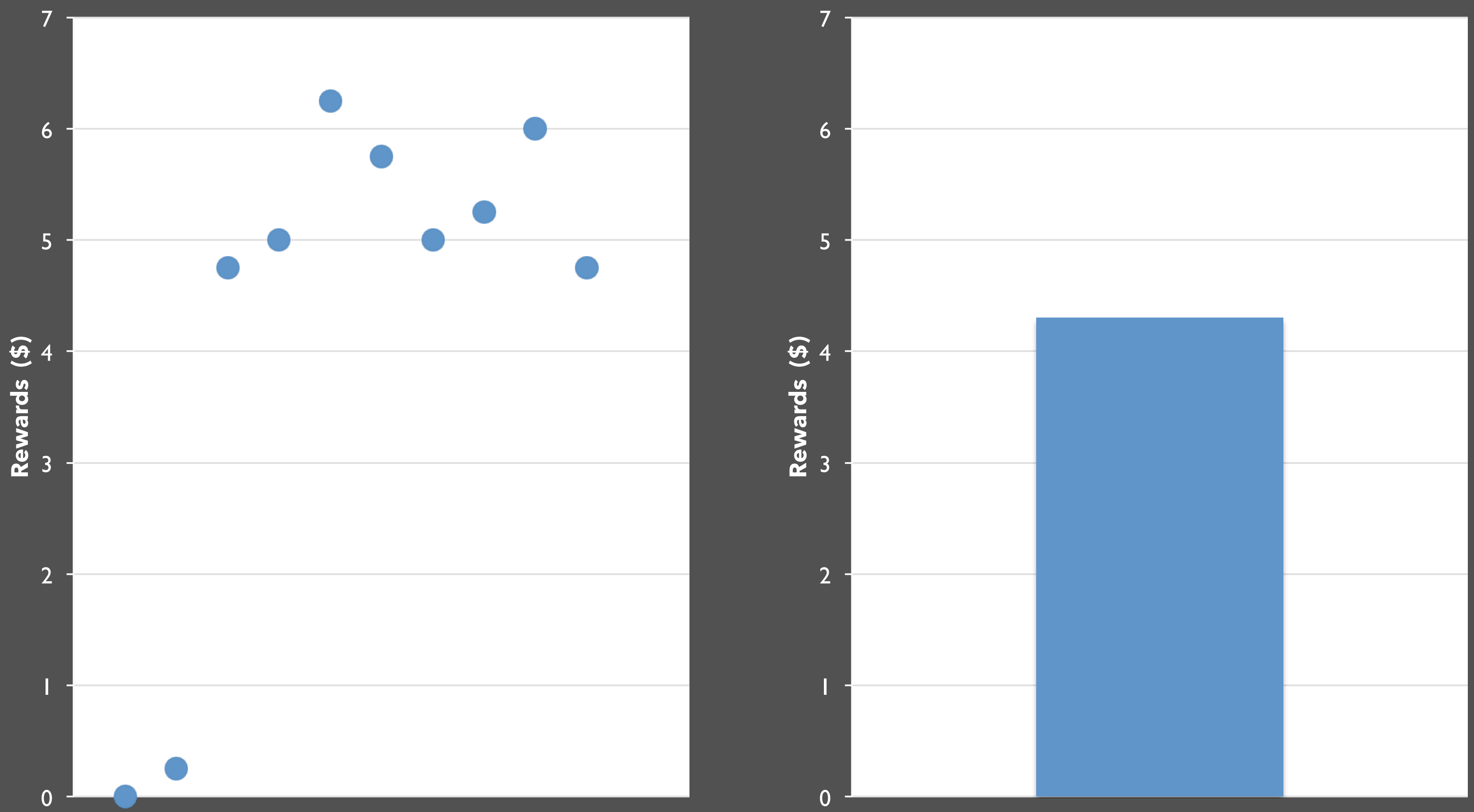
# Should you take this $4 bet?



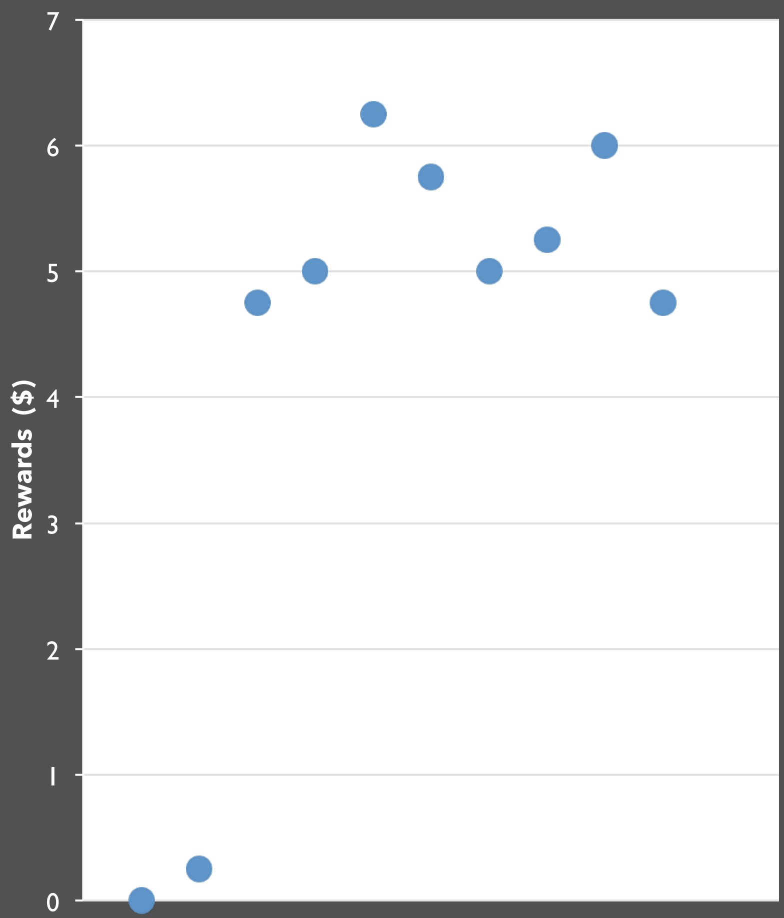
# Should you take this $4 bet?

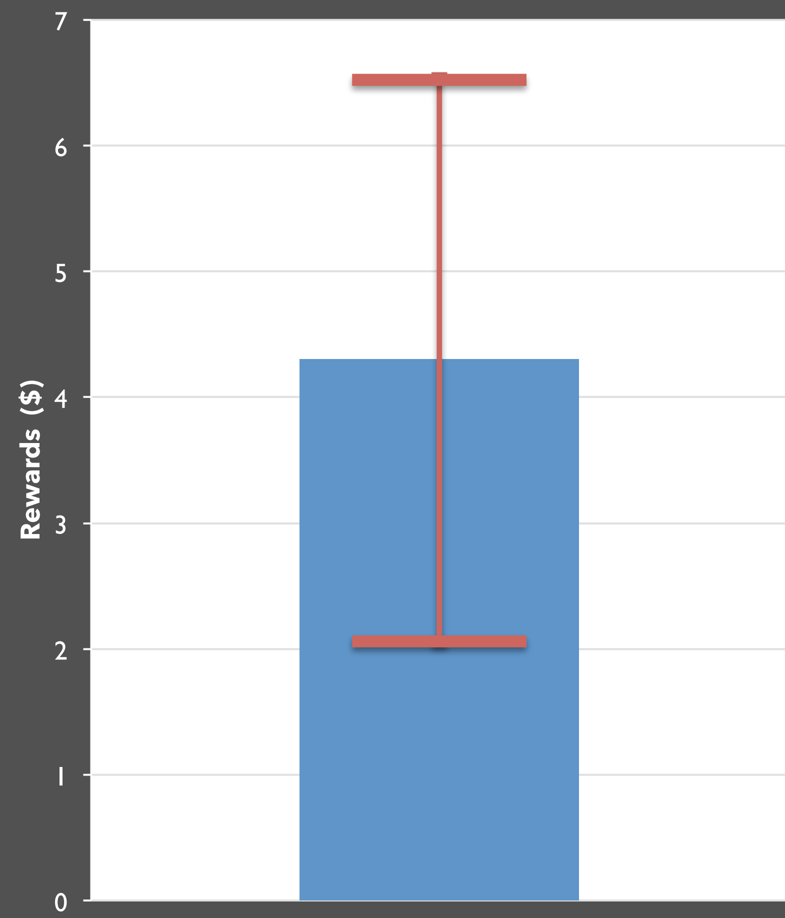


# Expected Value



# Mean And Error



# Statistical Inference

Assuming bet returns are normally distributed.

M = 4.14

SD = 2.33

n = 10

P(μ>4) = **0.95**

❚ Take the bet

# Statistical Inference

Assuming bet returns are normally distributed.

MODEL

M = 4.14

SD = 2.33

n = 10

P(μ>4) = **0.95**

❙ Take the bet

# Statistical Inference

Assuming bet returns are normally distributed. MODEL

M = 4.14

SD = 2.33 MEASUREMENT

n = 10

P(μ>4) = **0.95**

❚ Take the bet

# Statistical Inference

Assuming bet returns are normally distributed.

MODEL

M = 4.14

SD = 2.33

n = 10

MEASUREMENT

P(μ>4) = **0.95**

❚ Take the bet

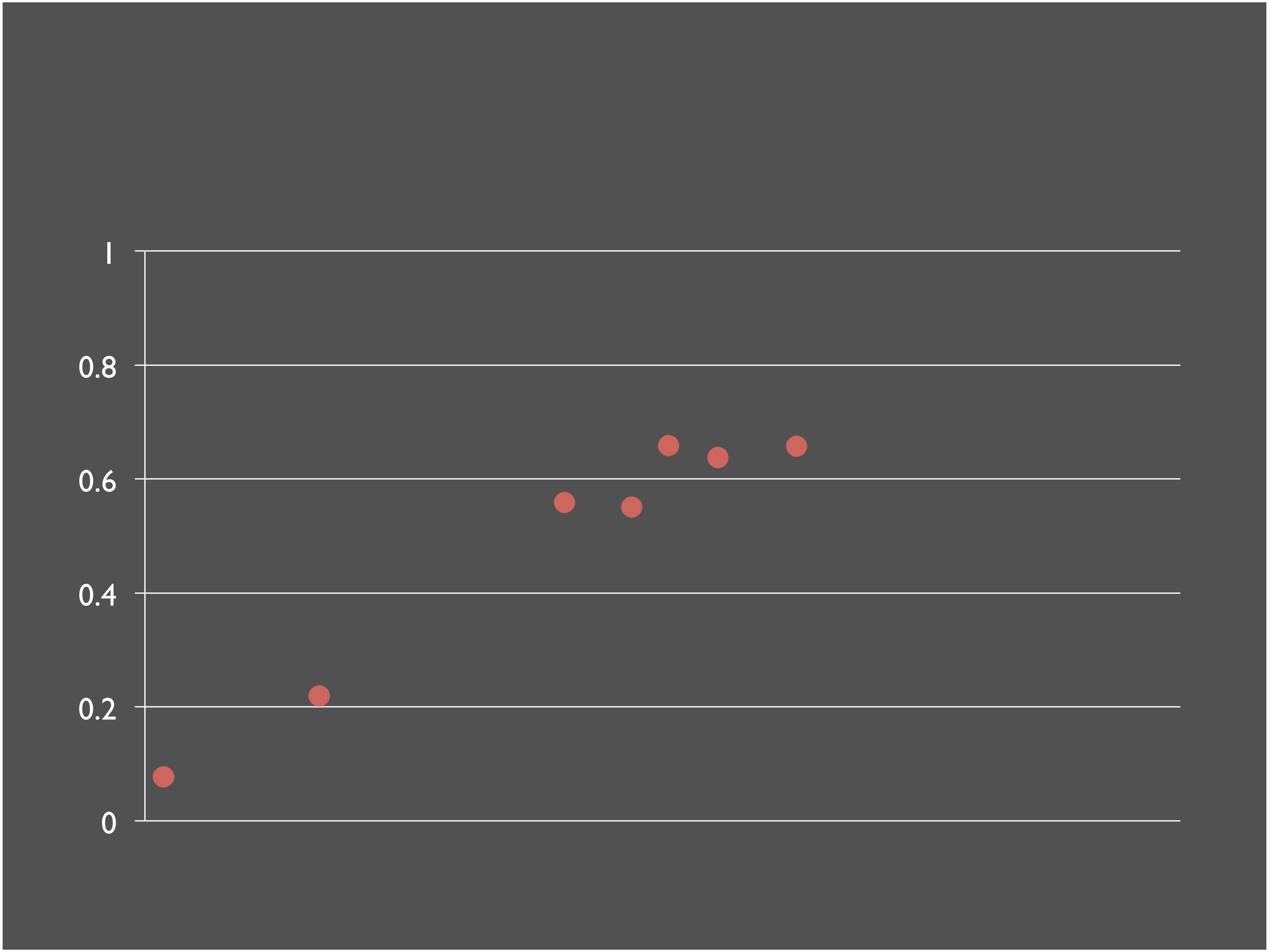
DECISION

# Uncertainty Sources

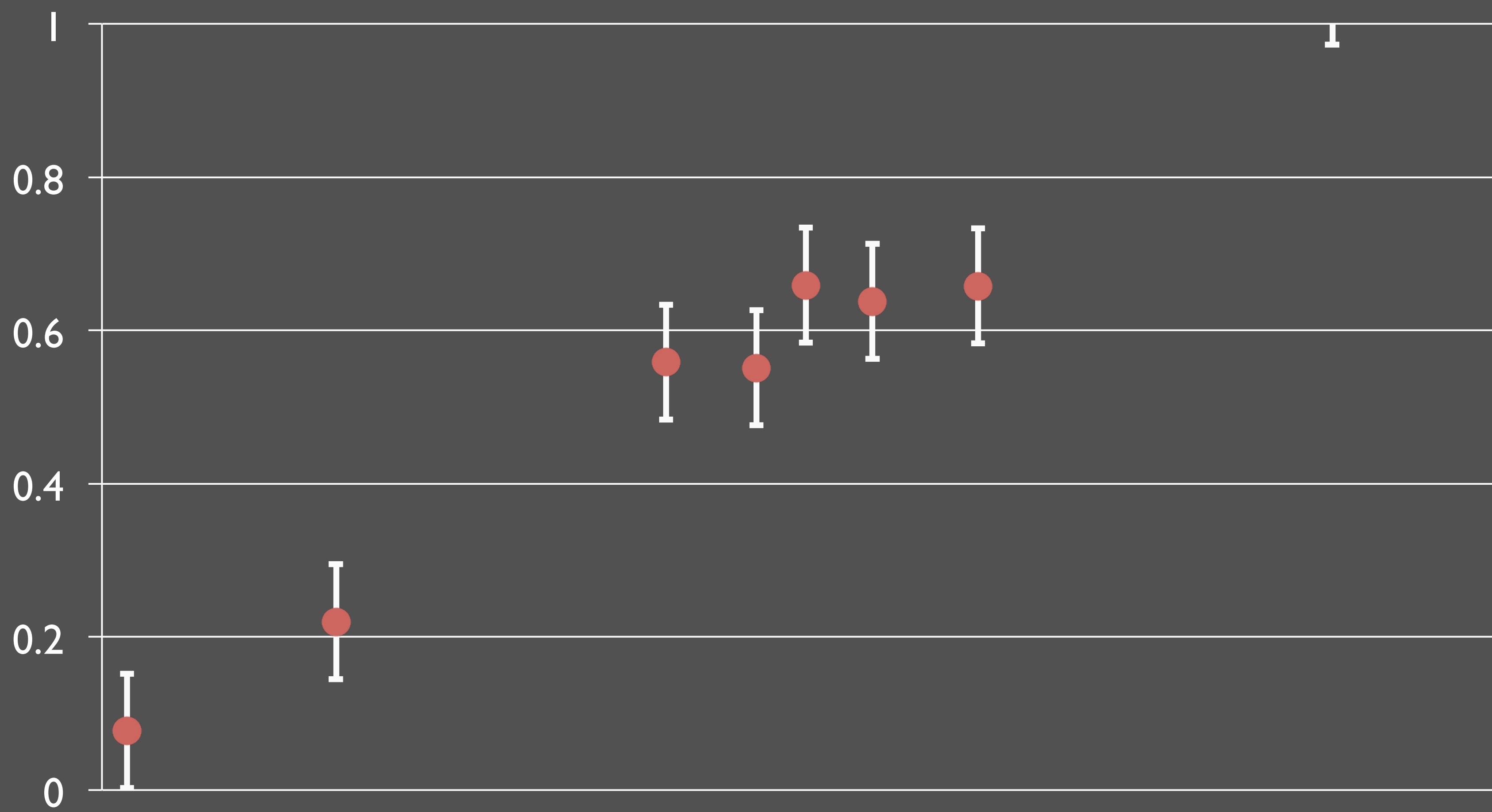
**Measurement Uncertainty**: "We're not sure what the data are"

**Model Uncertainty**: "We're not sure how the data fit together"

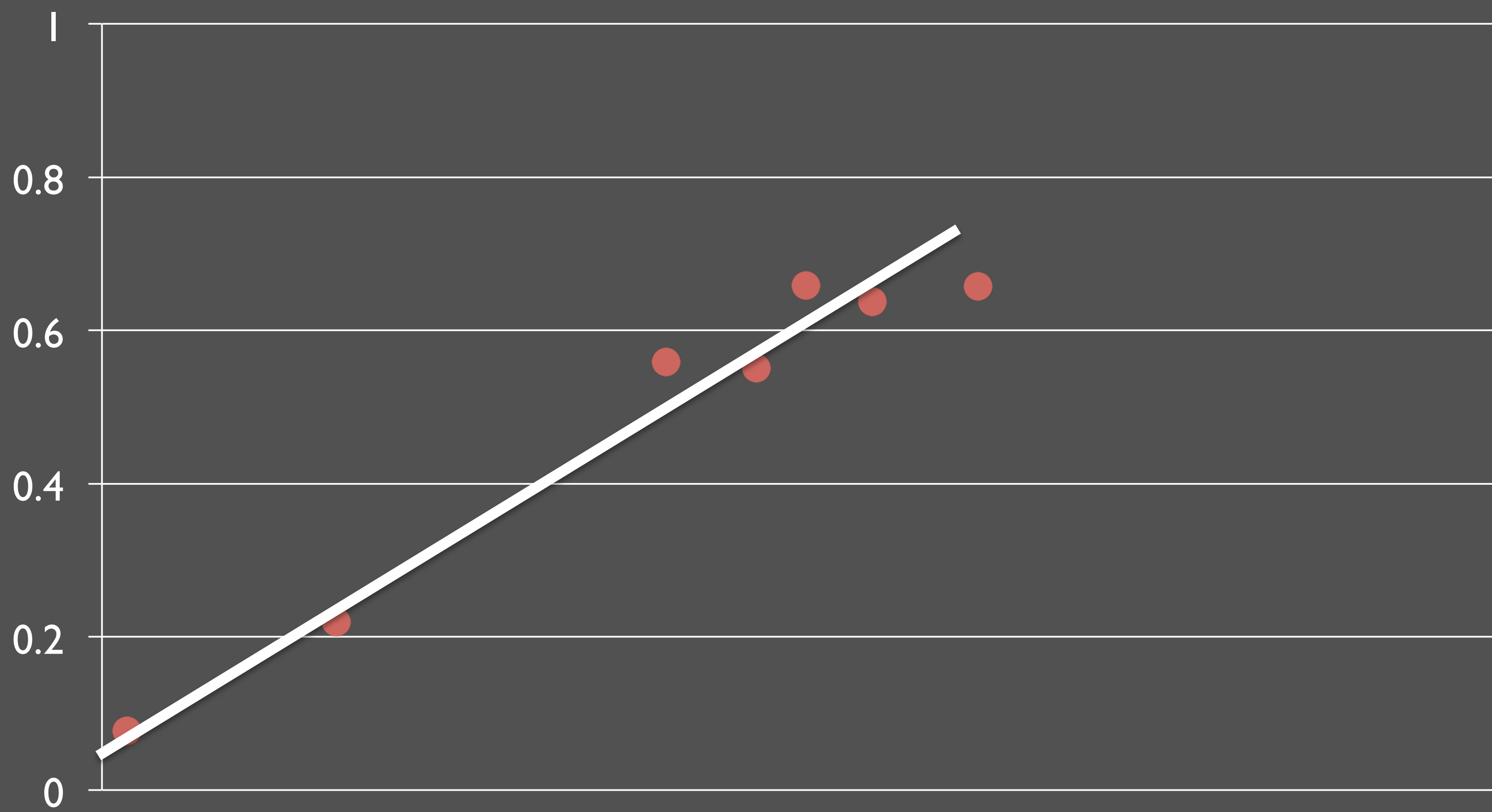
**Decision Uncertainty**: "We're not sure what to do now that we have the data"


1
0.8
0.6
0.4
0.2
0

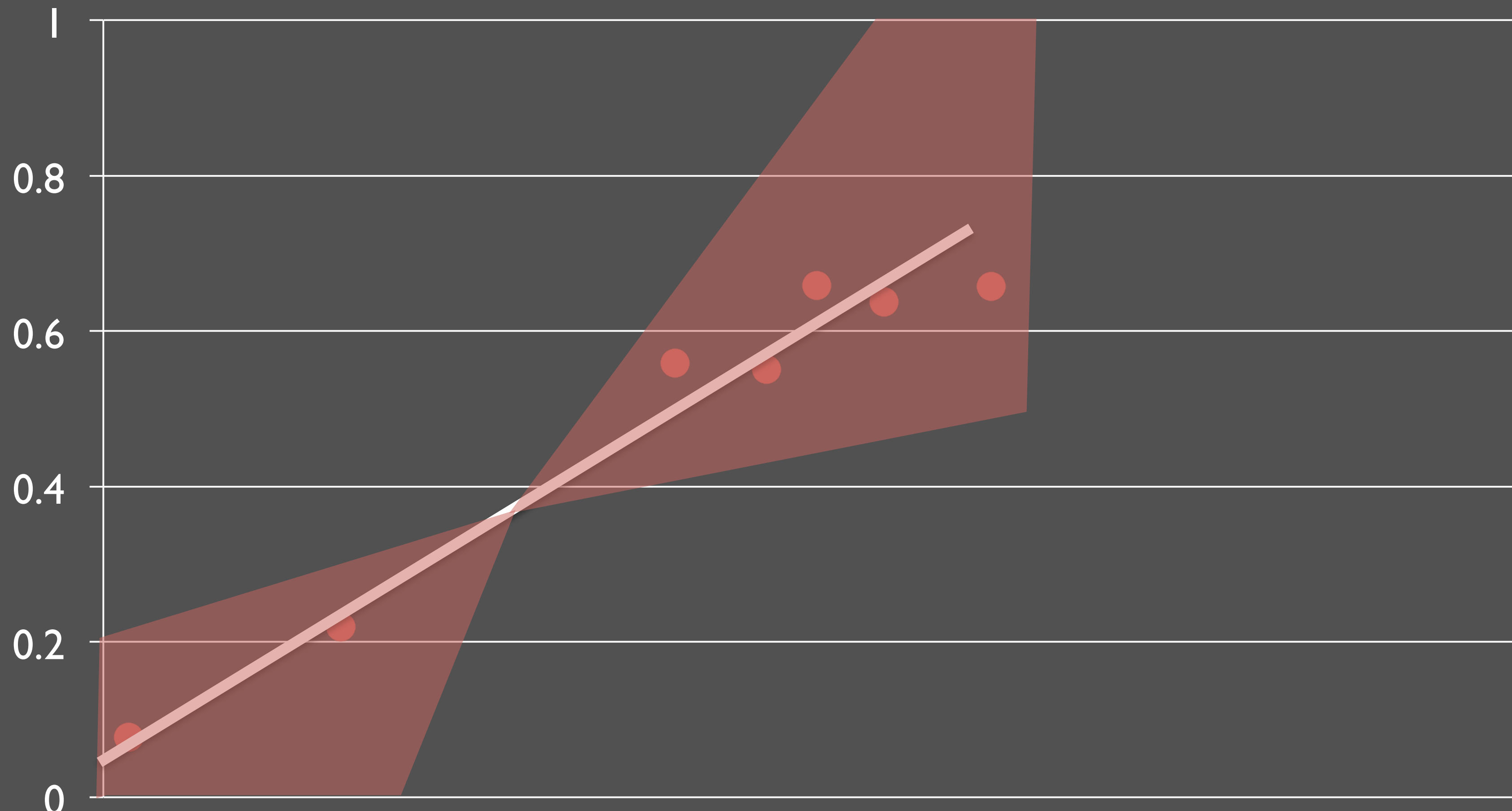
# Measurement Uncertainty



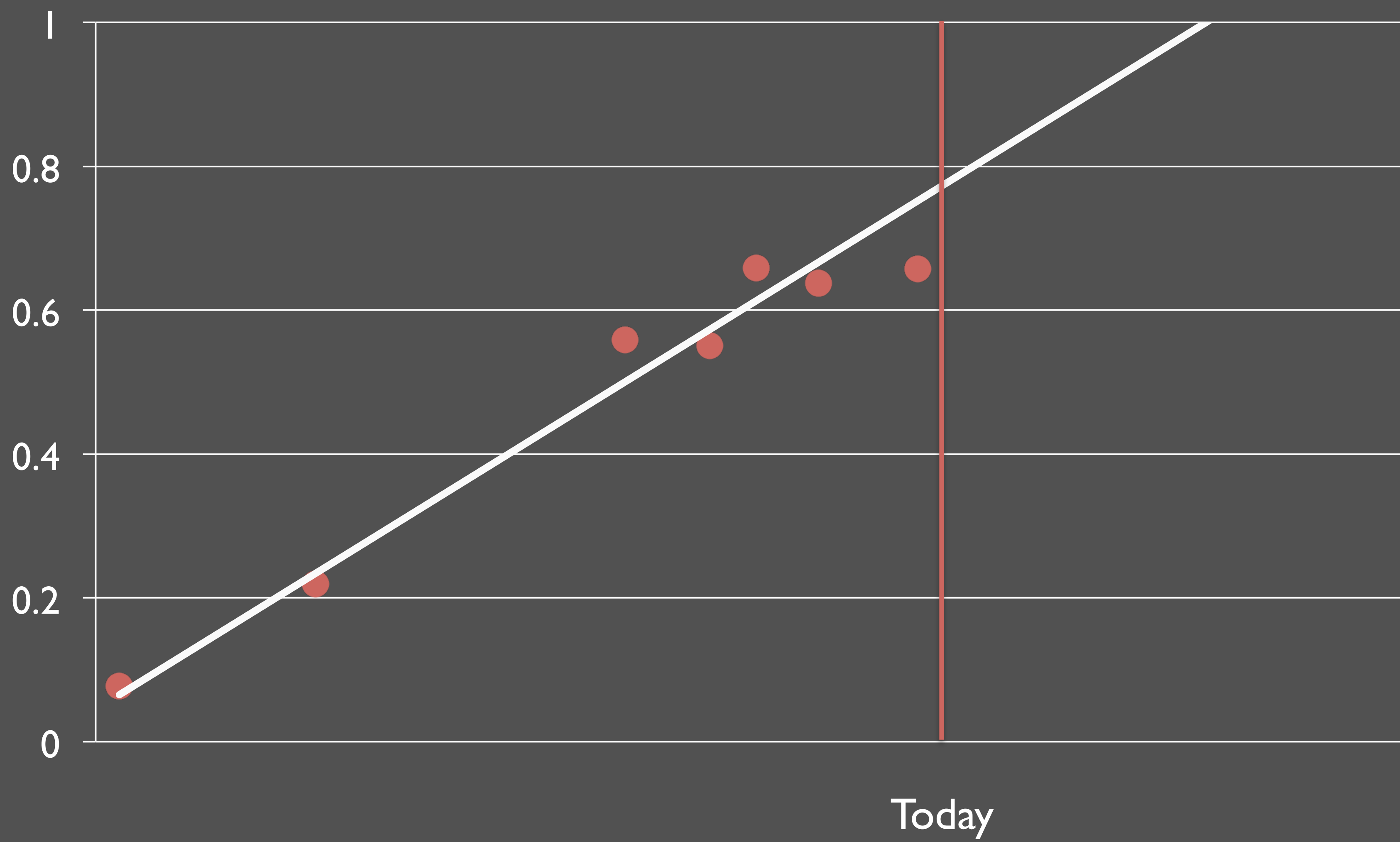
# Model Uncertainty



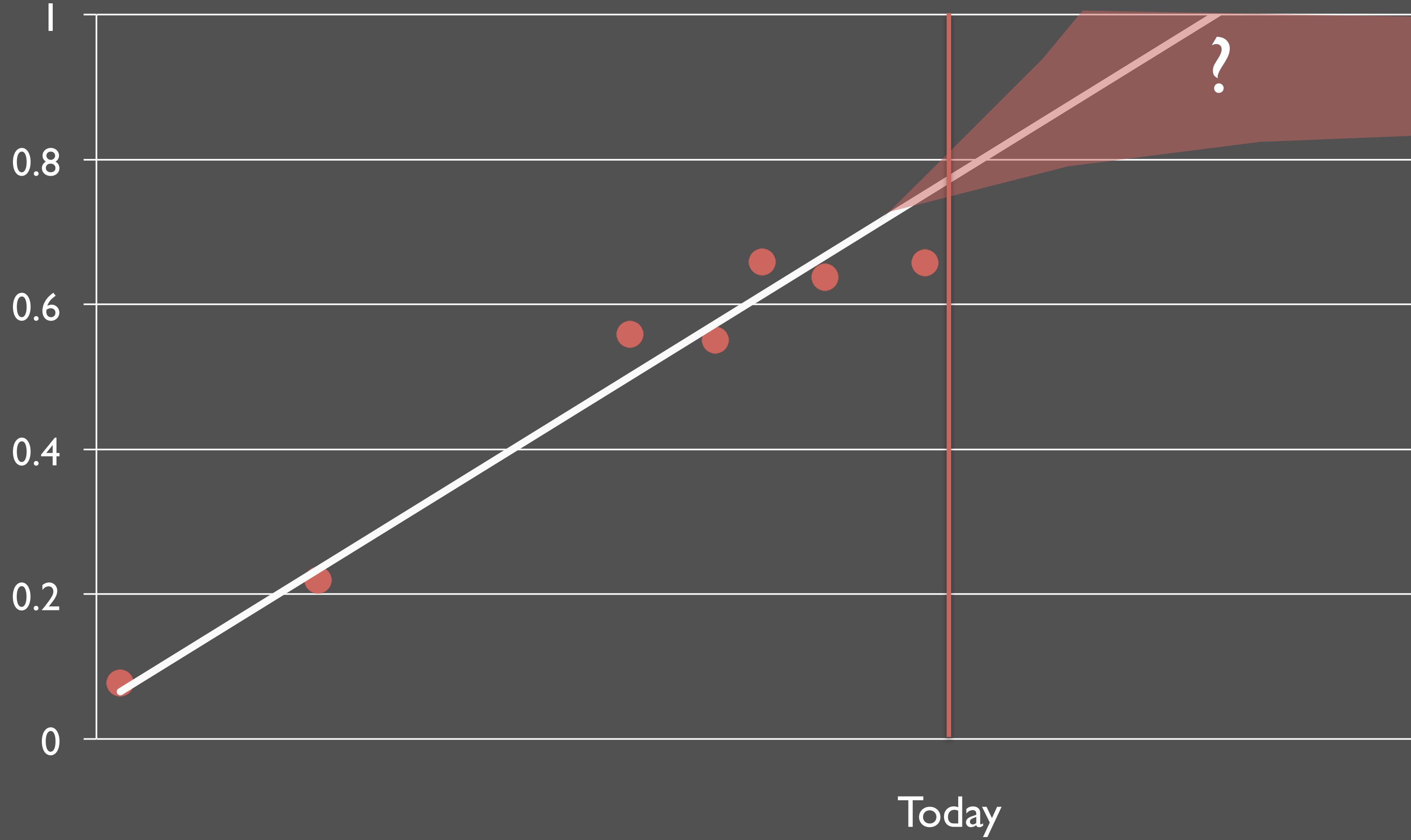
# Model Uncertainty



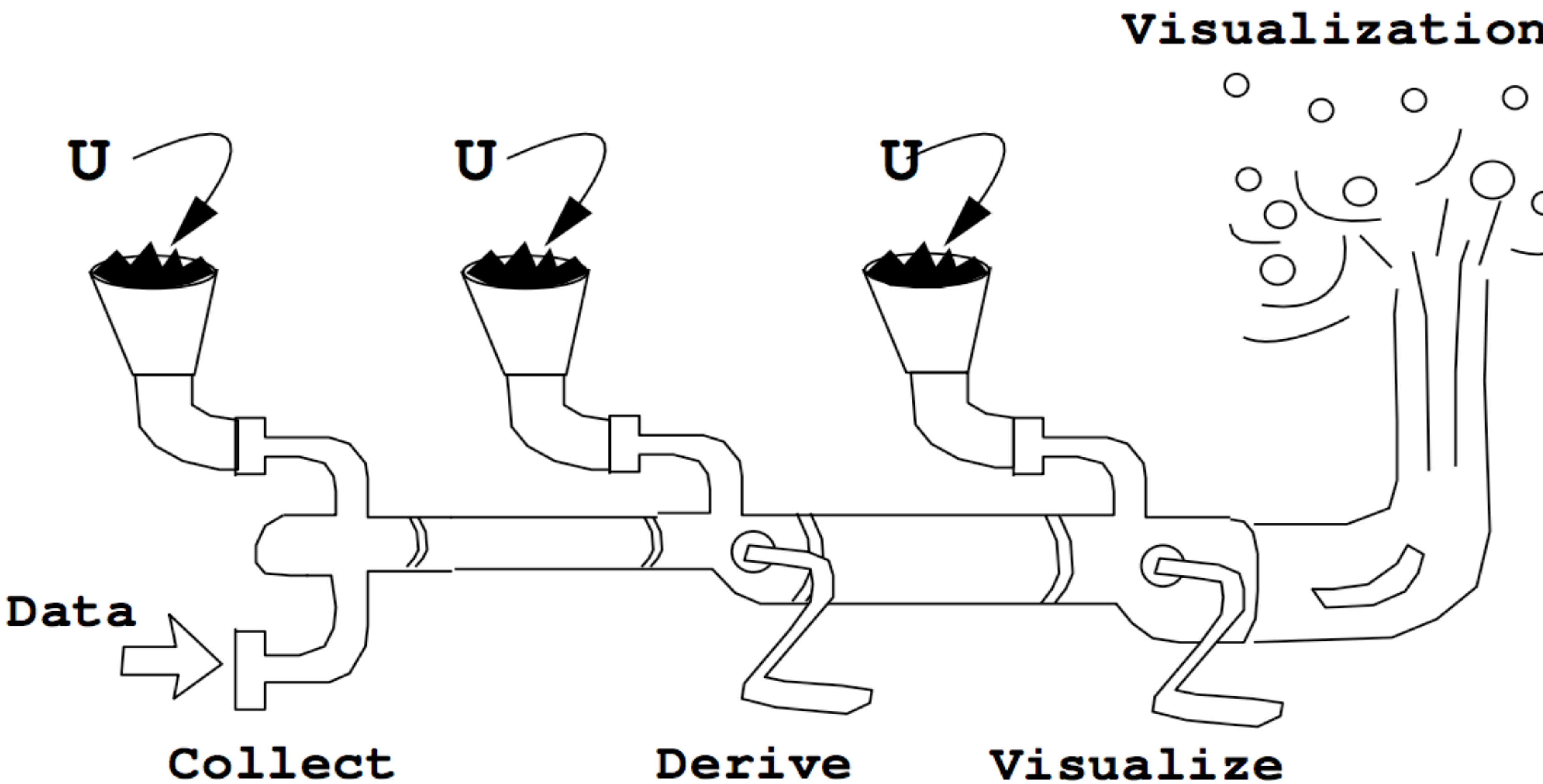
# Decision Uncertainty



# Decision Uncertainty



# Uncertainty Vis Pipeline

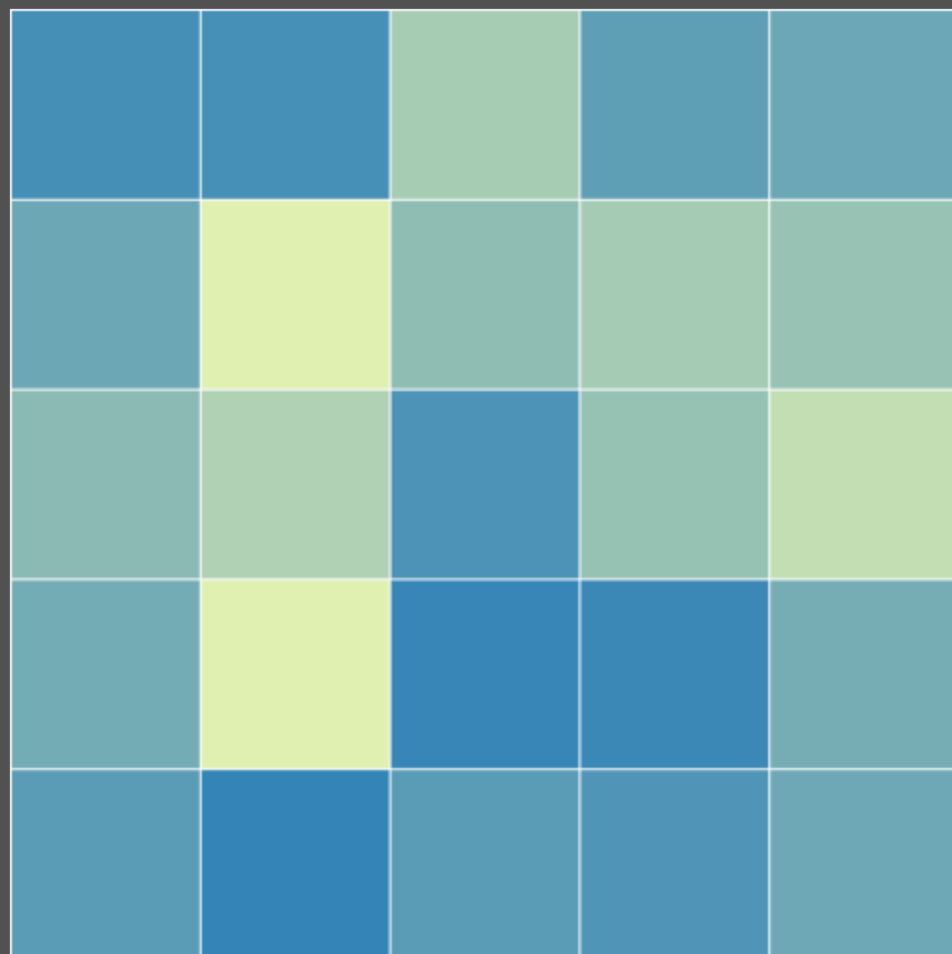



Pang et al. Approaches to Uncertainty Visualization. The Visual Computer, 1997.

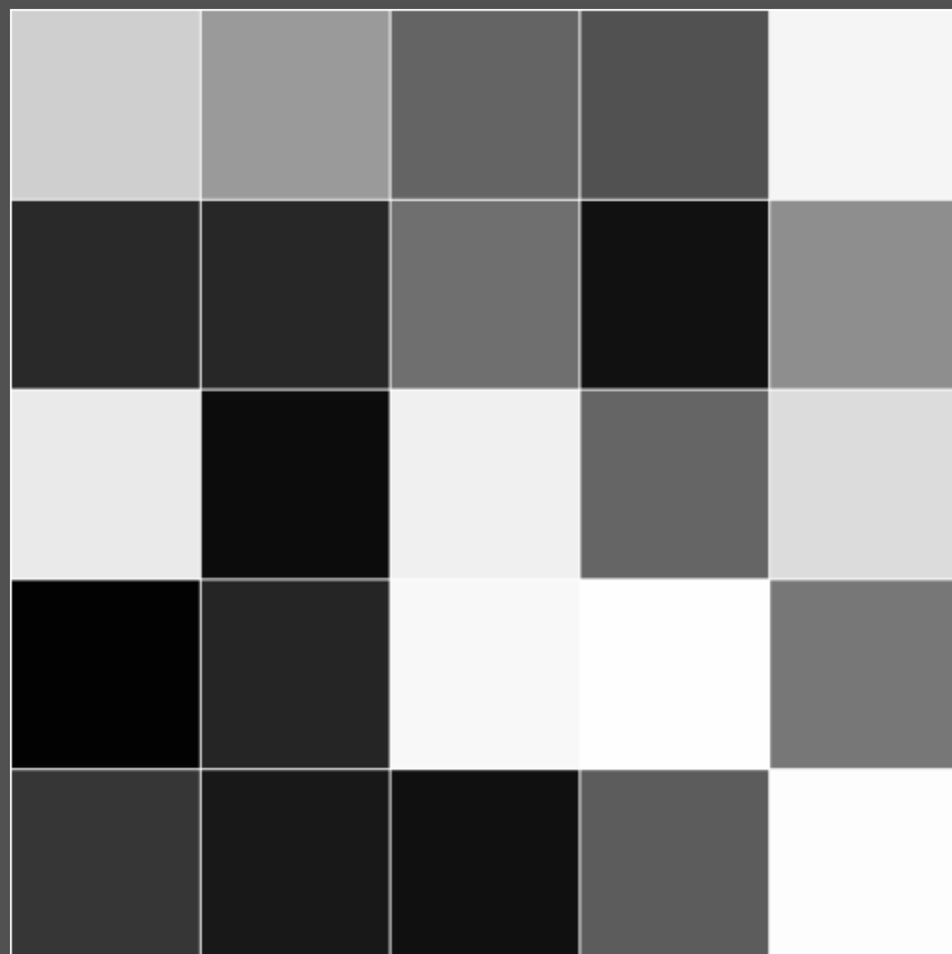
# Uncertainty Vis Pipeline

1) Quantify Uncertainty
2) Choose a free visual variable
3) Encode uncertainty with the variable

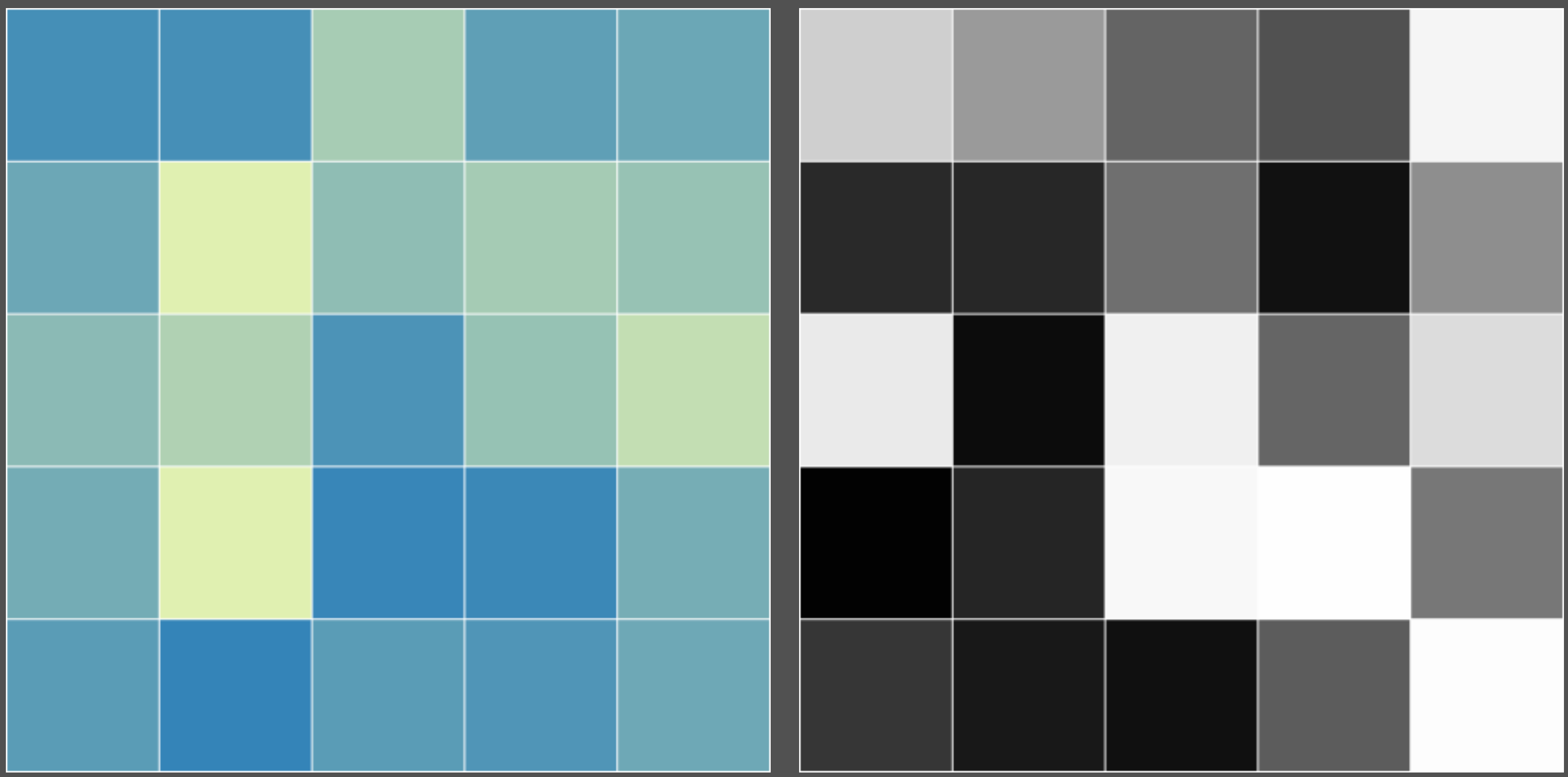
# Data Map



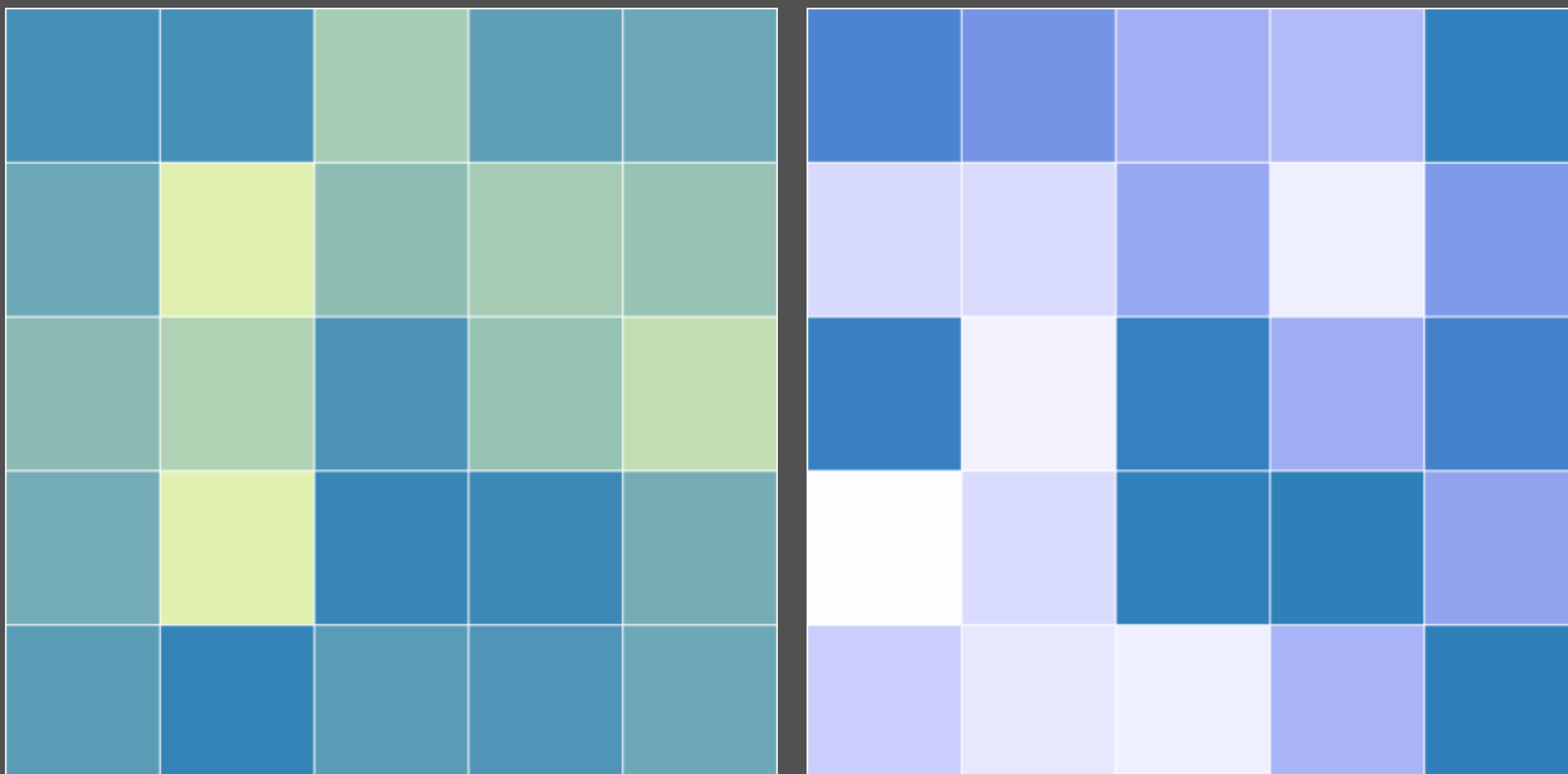
# Uncertainty Map



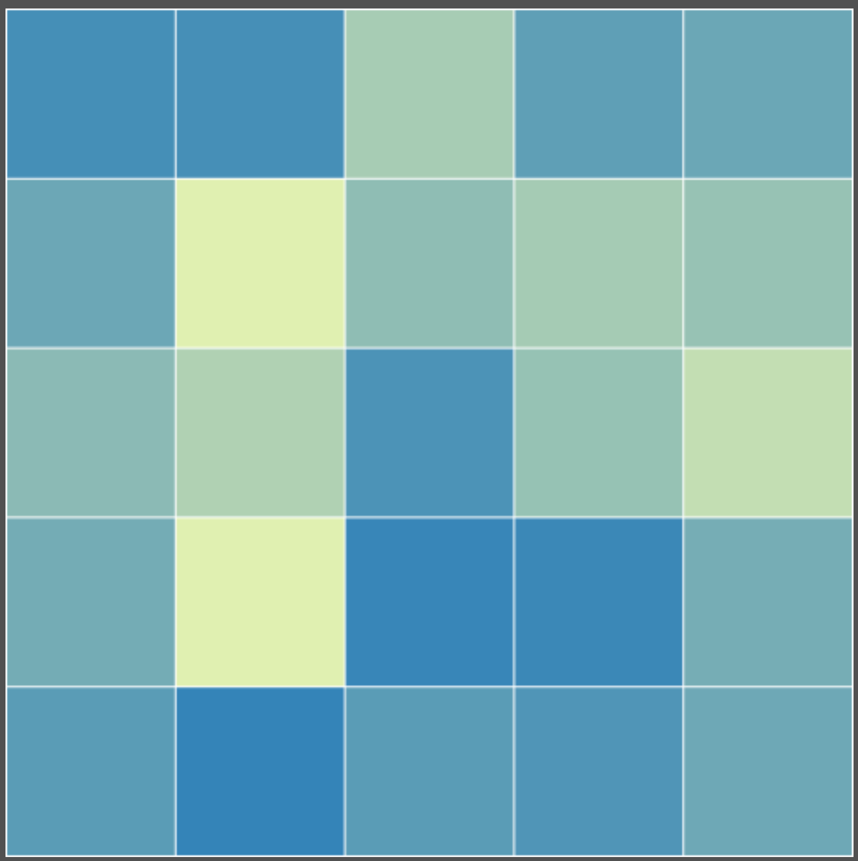
# Juxtaposition

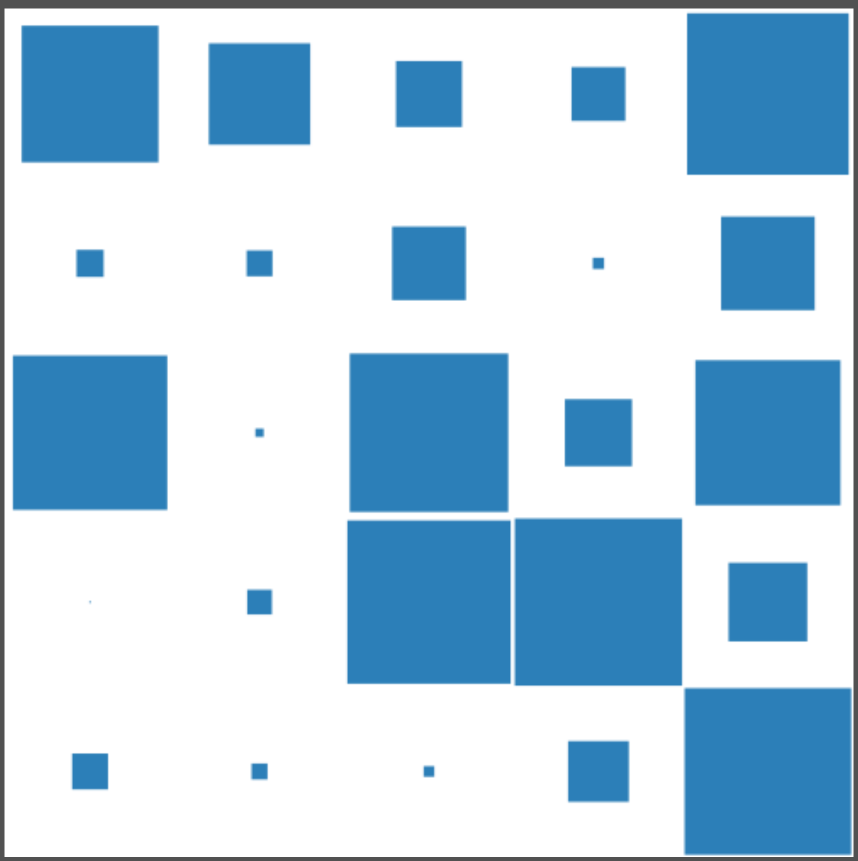


# Juxtaposition

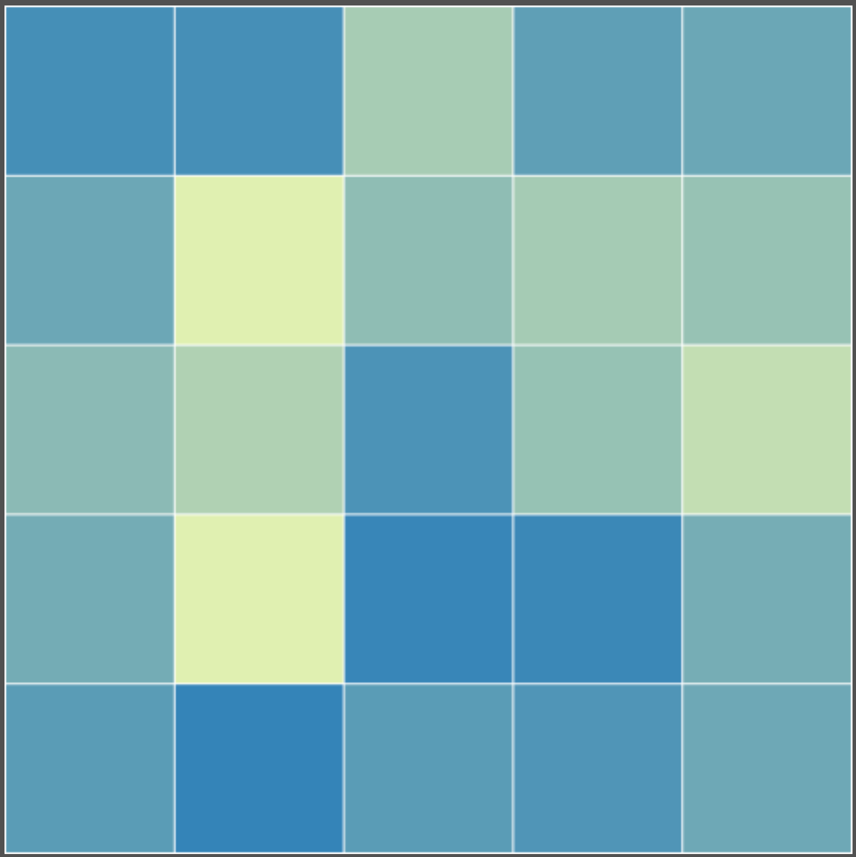


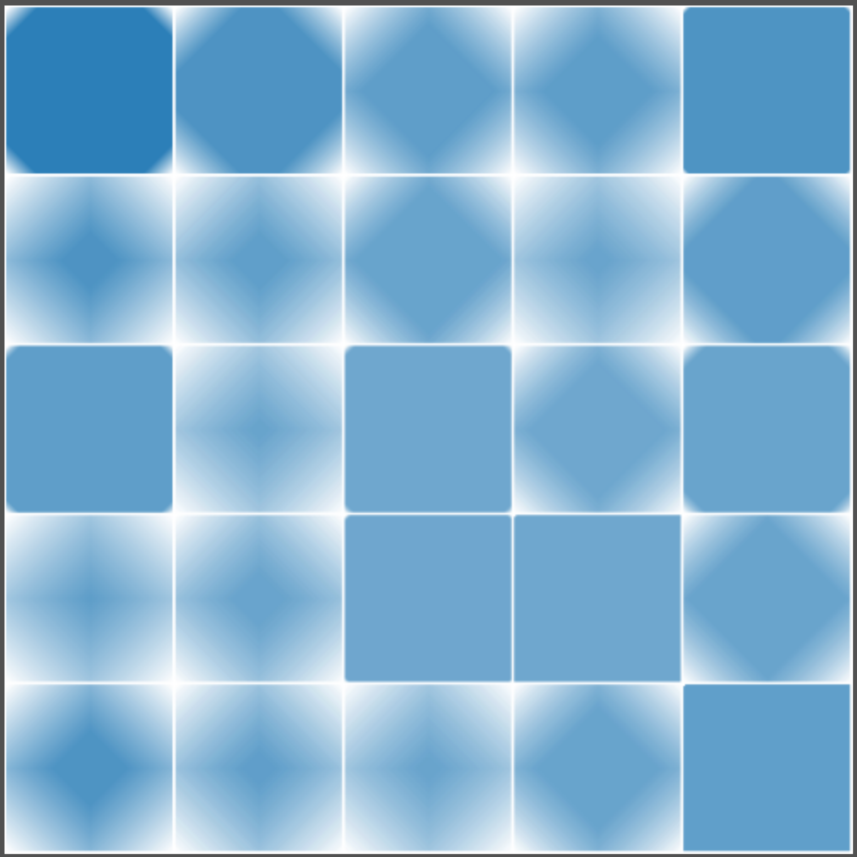
# Juxtaposition



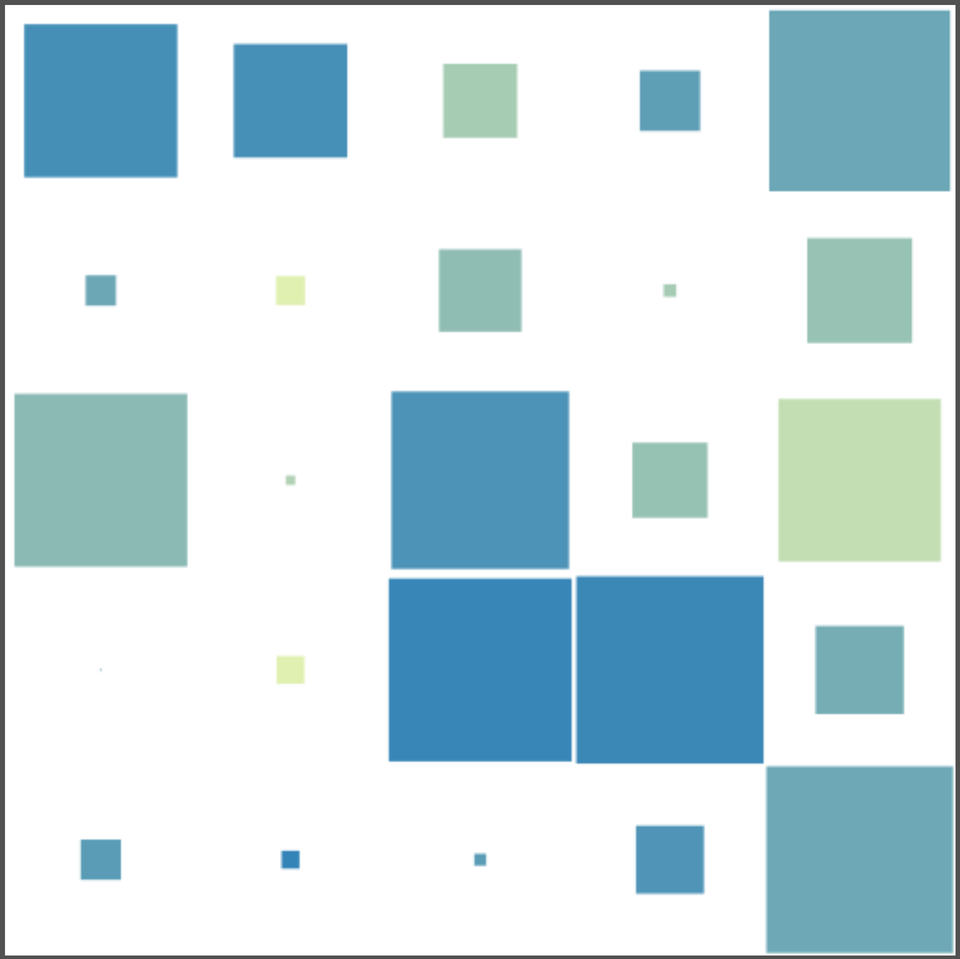


# Juxtaposition

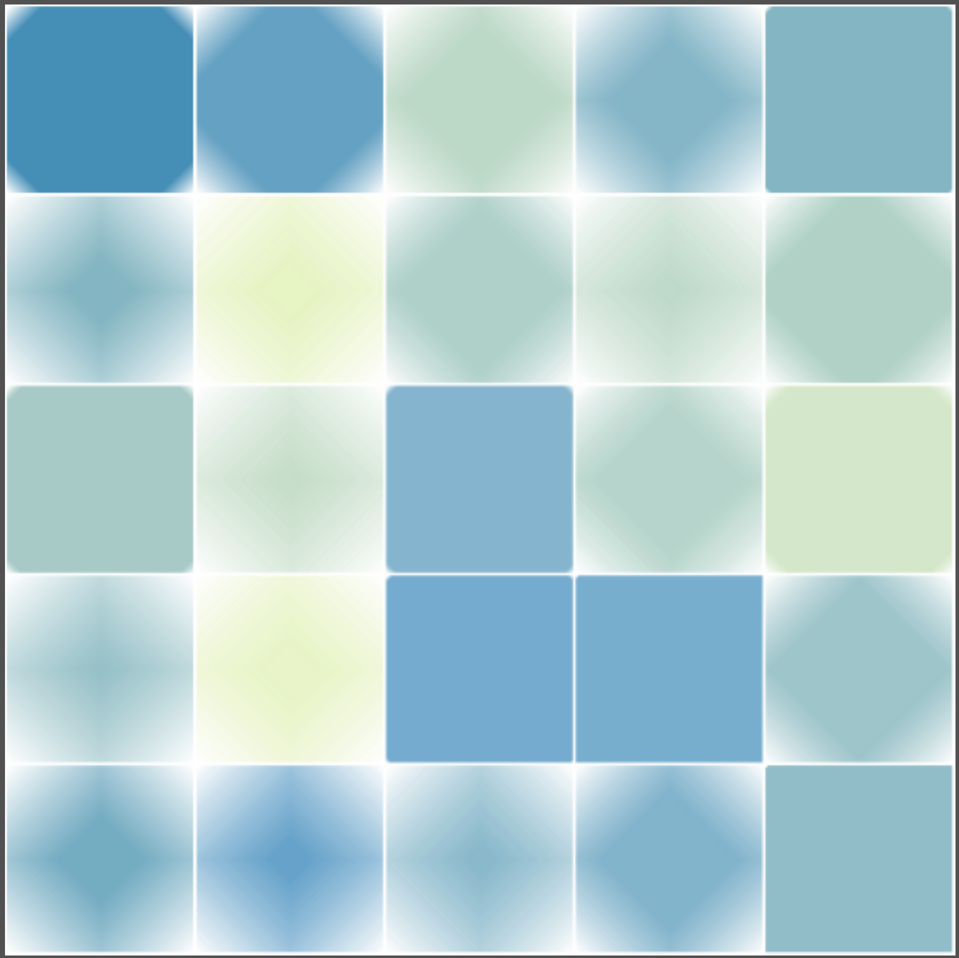




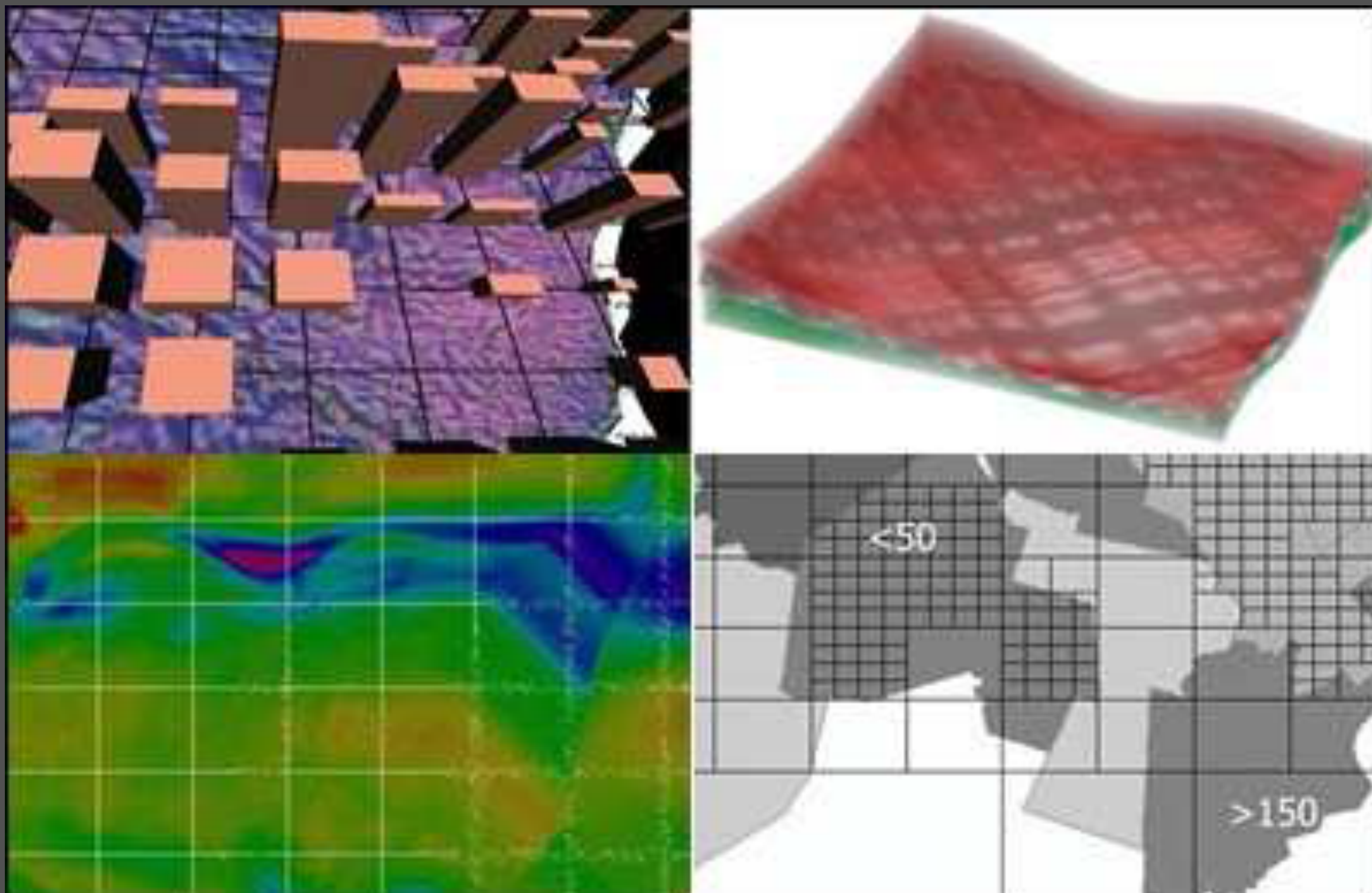
# Superposition



# Superposition



# Superposition




Griethe, Henning and Schumann, Heidrun. The Visualization of Uncertain Data: Methods and Problems. SimVis, 2006.

# Uncertainty Vis Pipeline?

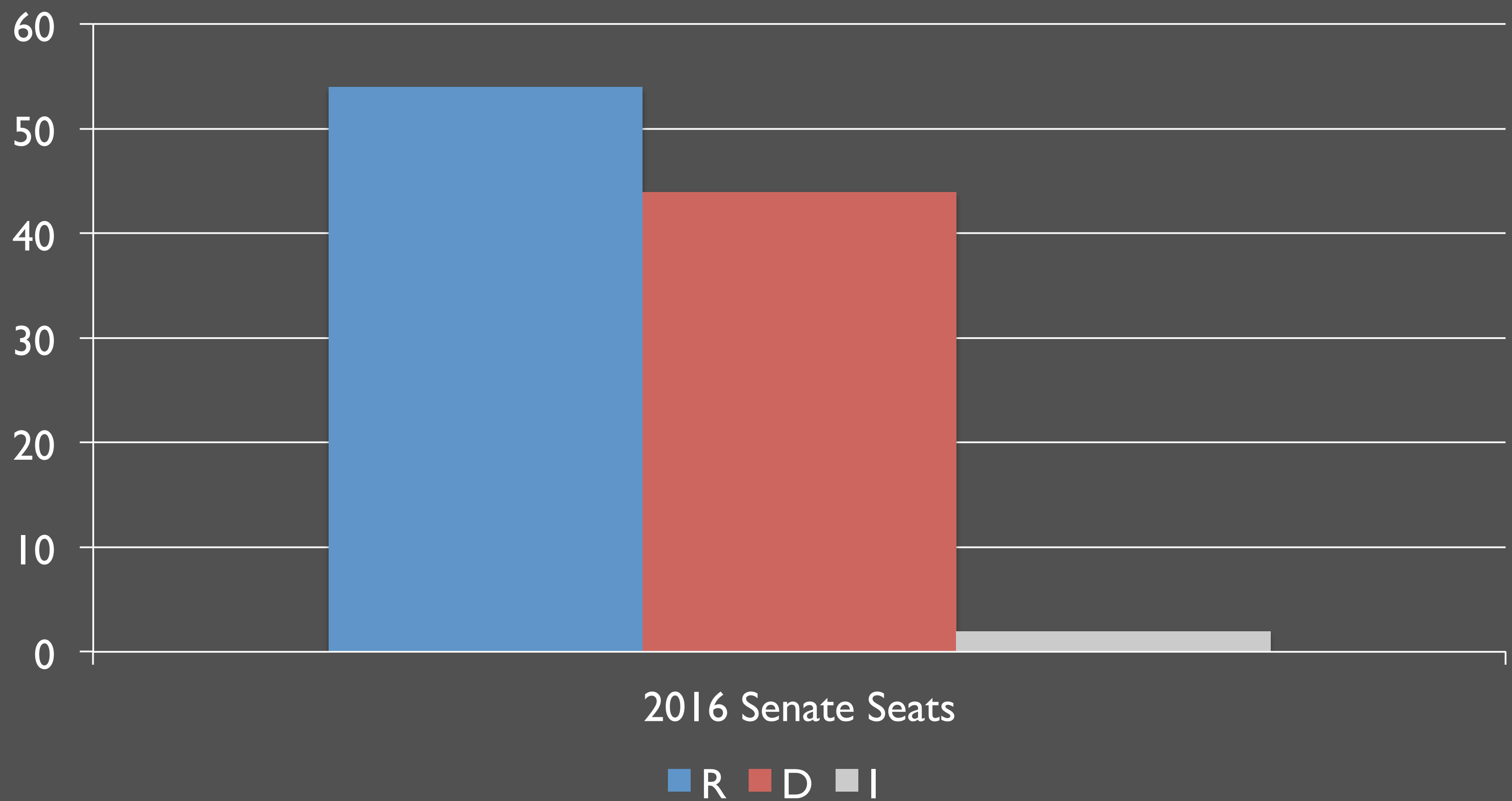
1) Quantify Uncertainty
2) Choose a free visual variable
3) Encode uncertainty with the variable

Design Decisions:

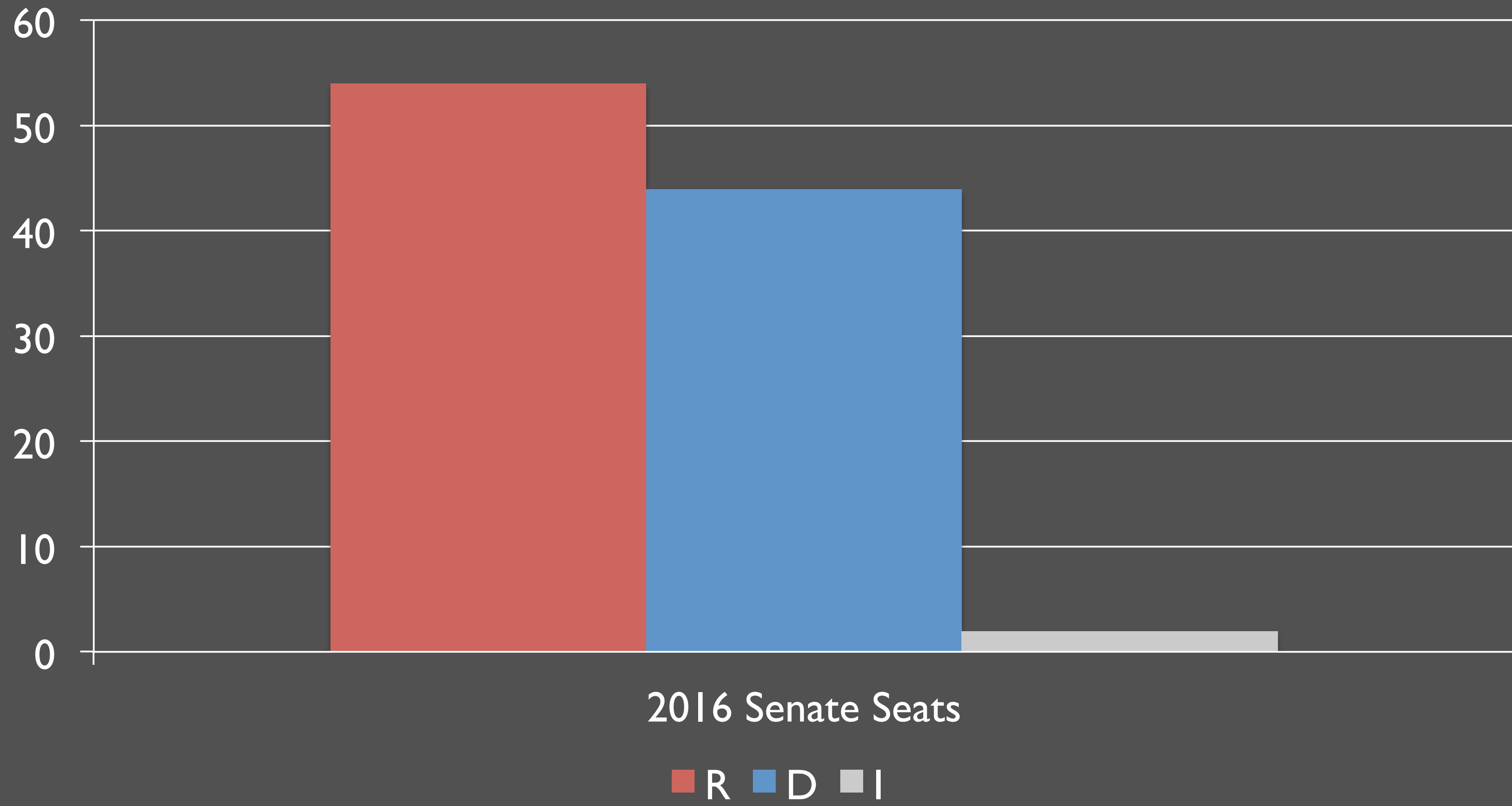
How to unify data and uncertainty map(s)?

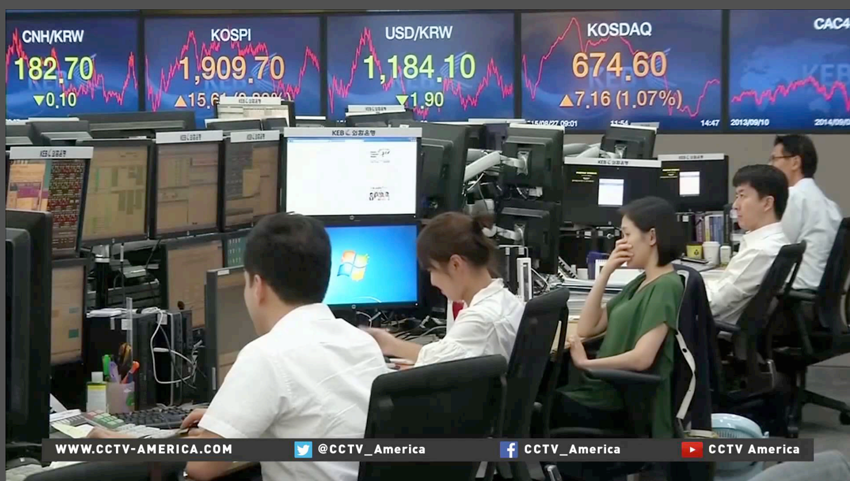
# Semiotics of Uncertainty

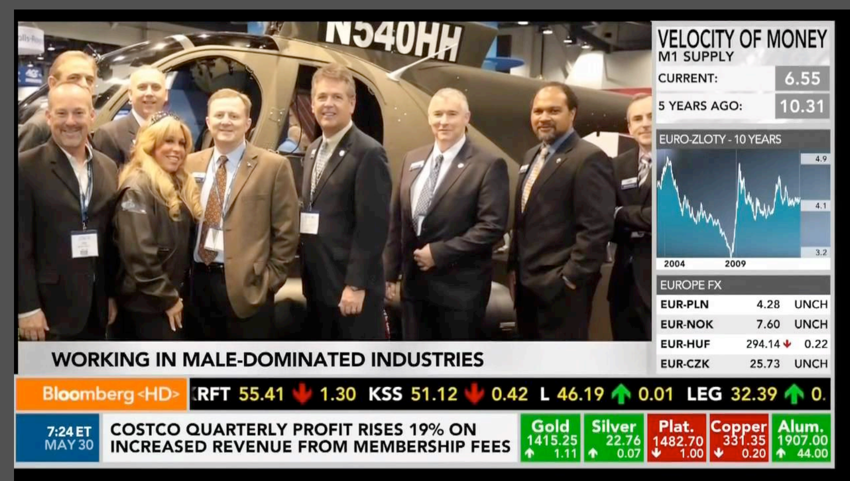
# The Variable Matters!



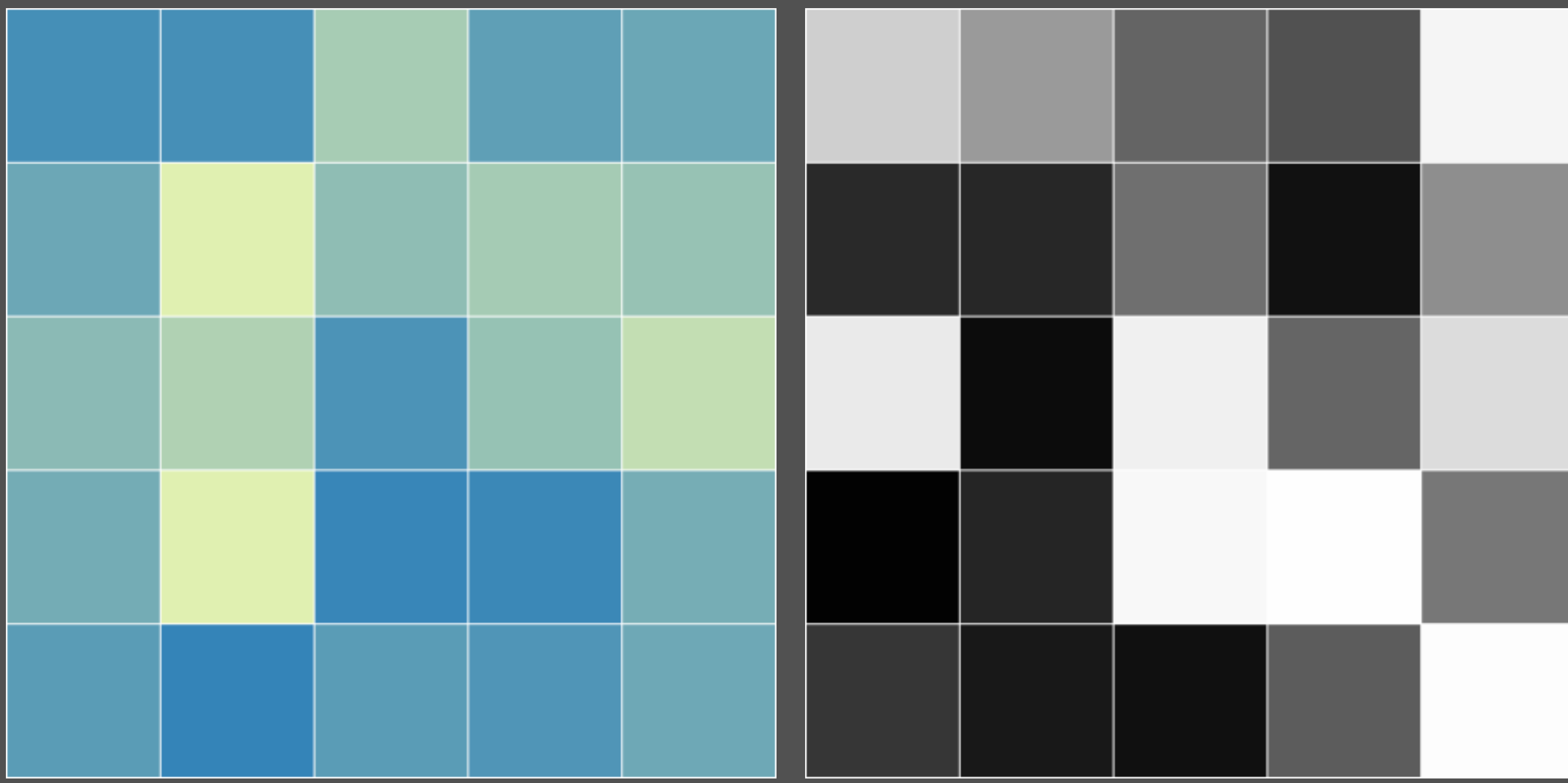
# The Variable Matters!



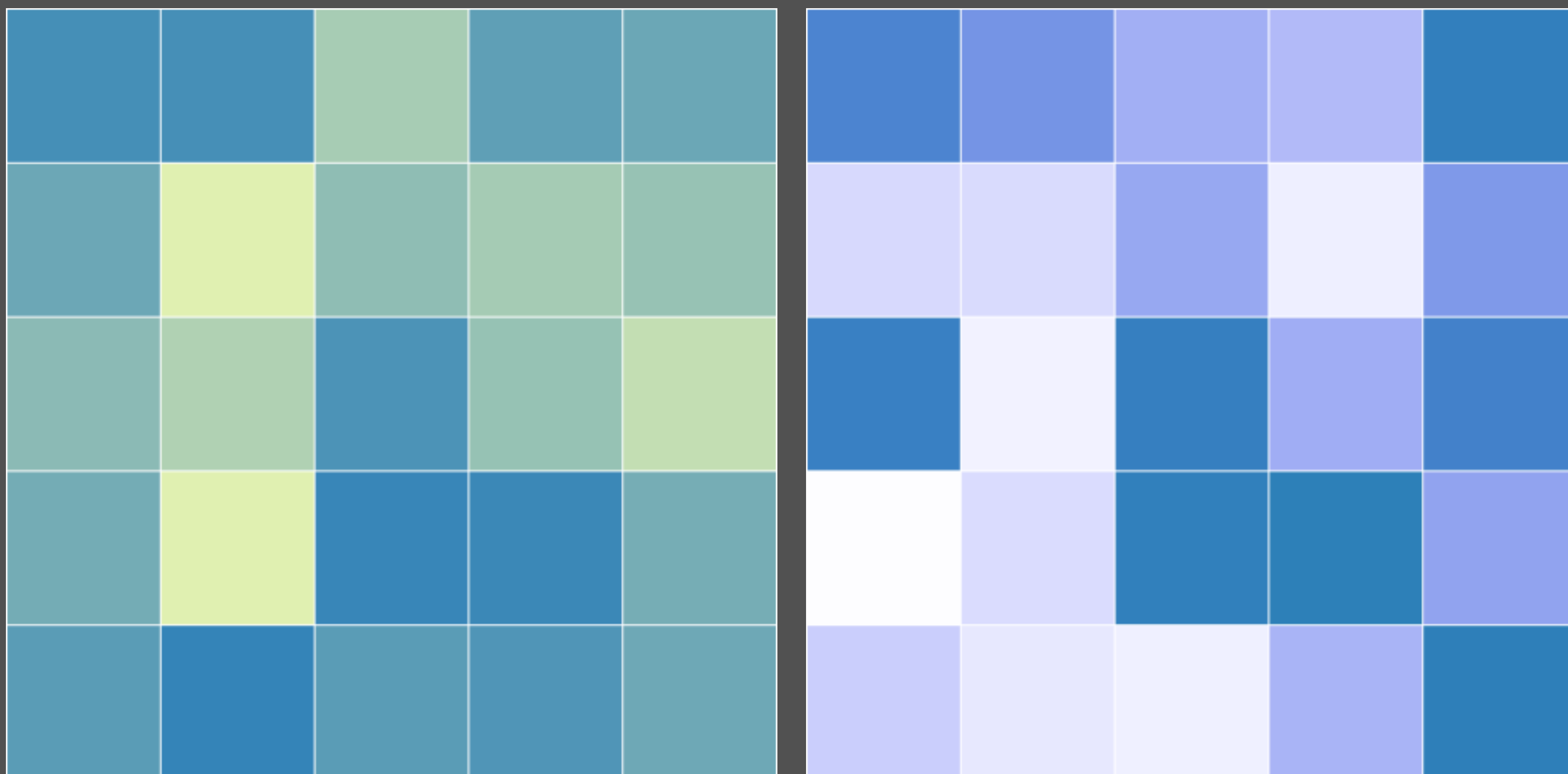

CNH/KRW
182.70
▼0.10
KOSPI
1,909.70
USD/KRW
1,184.10
▼1.90
KOSDAQ
674.60
▲7.16 (1.07%)
WWW.CCTV-AMERICA.COM
@CCTV_America
CCTV_America
CCTV America


N540HH
VELOCITY OF MONEY
M1 SUPPLY
CURRENT: 6.55
5 YEARS AGO: 10.31
EURO-ZLOTY - 10 YEARS
2004 2009
EUROPE FX
EUR-PLN 4.28 UNCH
EUR-NOK 7.60 UNCH
EUR-HUF 294.14 0.22
EUR-CZK 25.73 UNCH
WORKING IN MALE-DOMINATED INDUSTRIES
Bloomberg <HD>
RFT 55.41 1.30 KSS 51.12 0.42 L 46.19 0.01 LEG 32.39
7:24 ET MAY 30
COSTCO QUARTERLY PROFIT RISES 19% ON INCREASED REVENUE FROM MEMBERSHIP FEES
Gold 1415.25 1.11
Silver 22.76 0.07
Plat. 1482.70 1.00
Copper 331.35 0.20
Alum. 1907.00 44.00

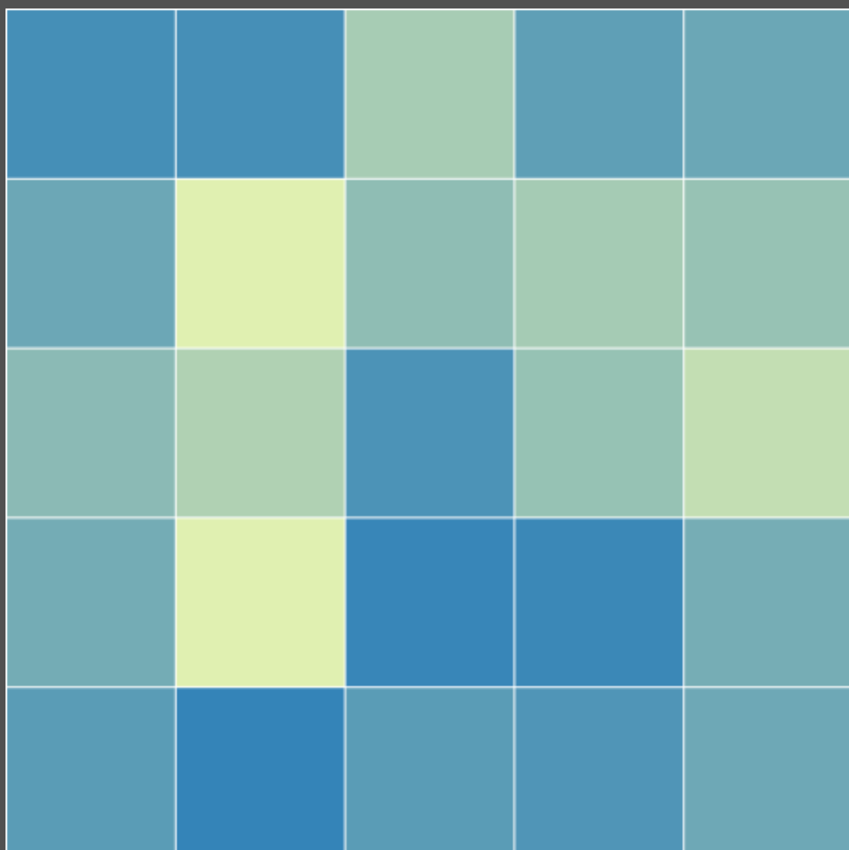
# Visual Variables for Uncertainty

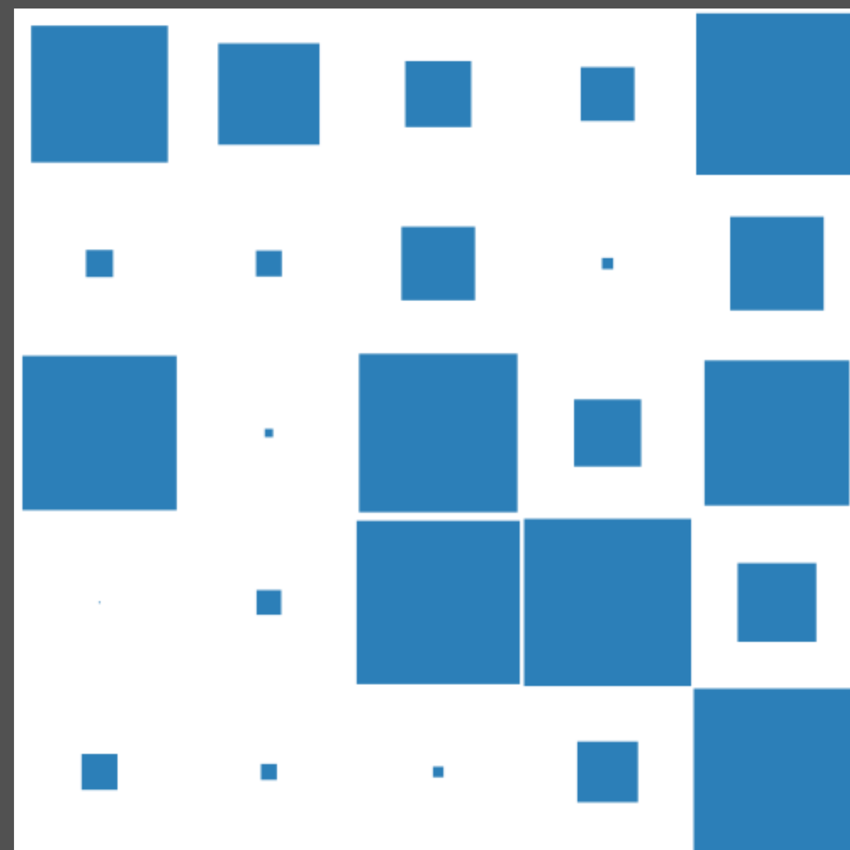


# Value

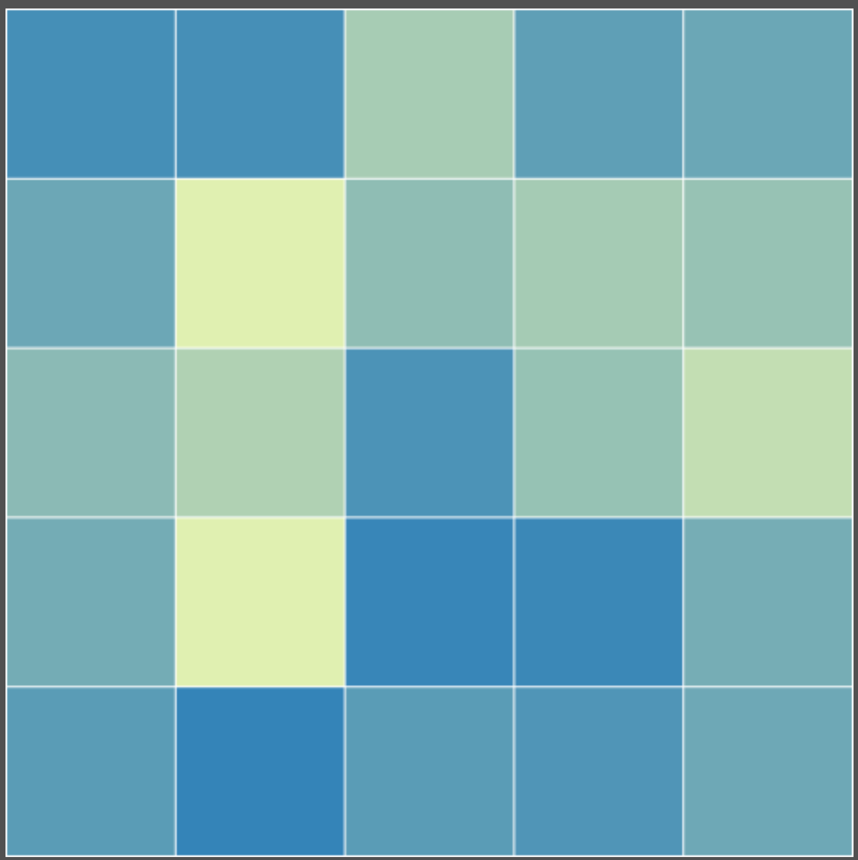


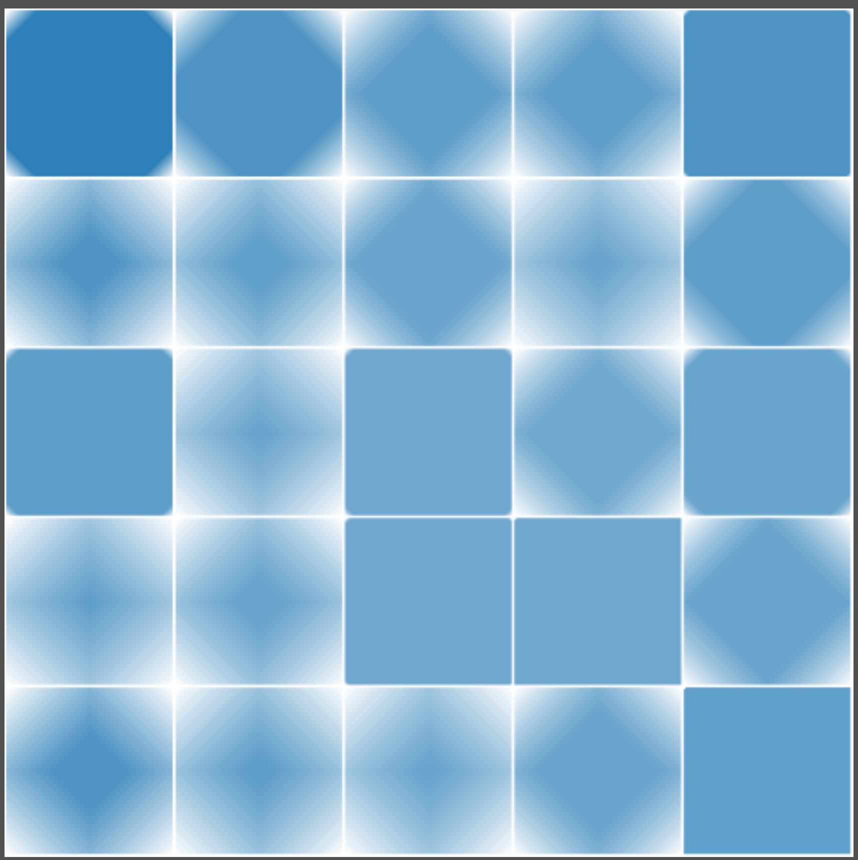
# Size





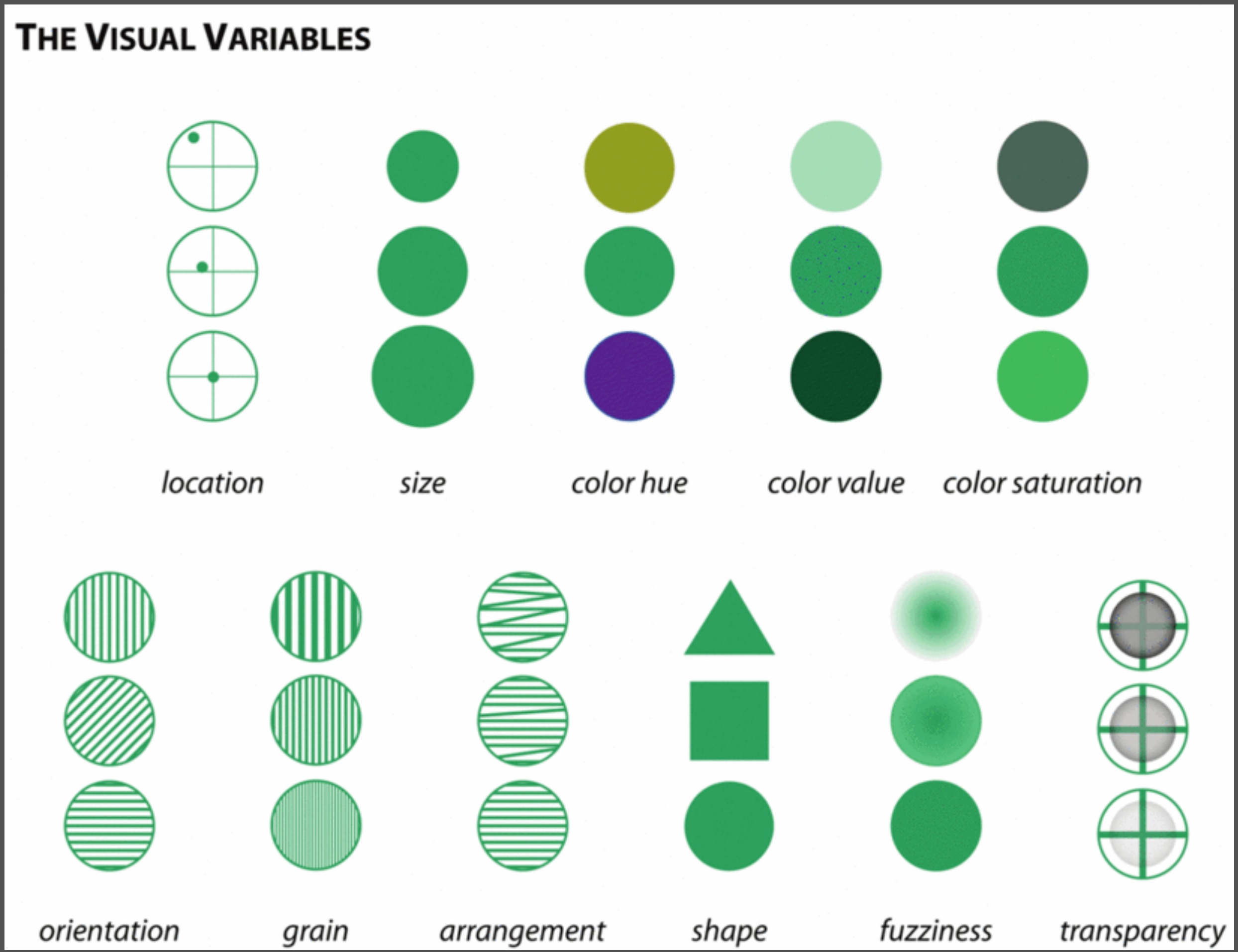
# Fuzziness





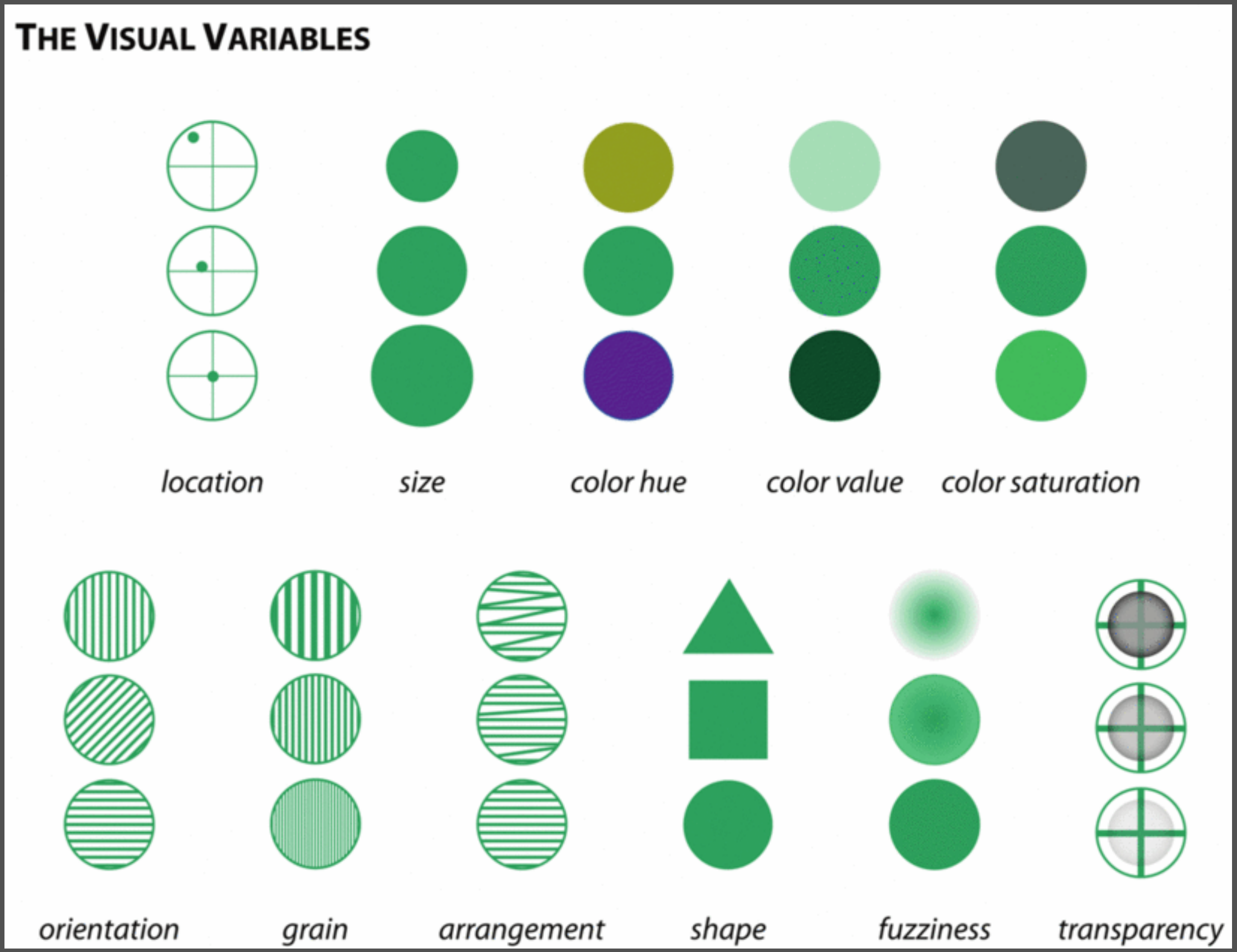
# Semiotics of Uncertainty

# Semiotics of Uncertainty

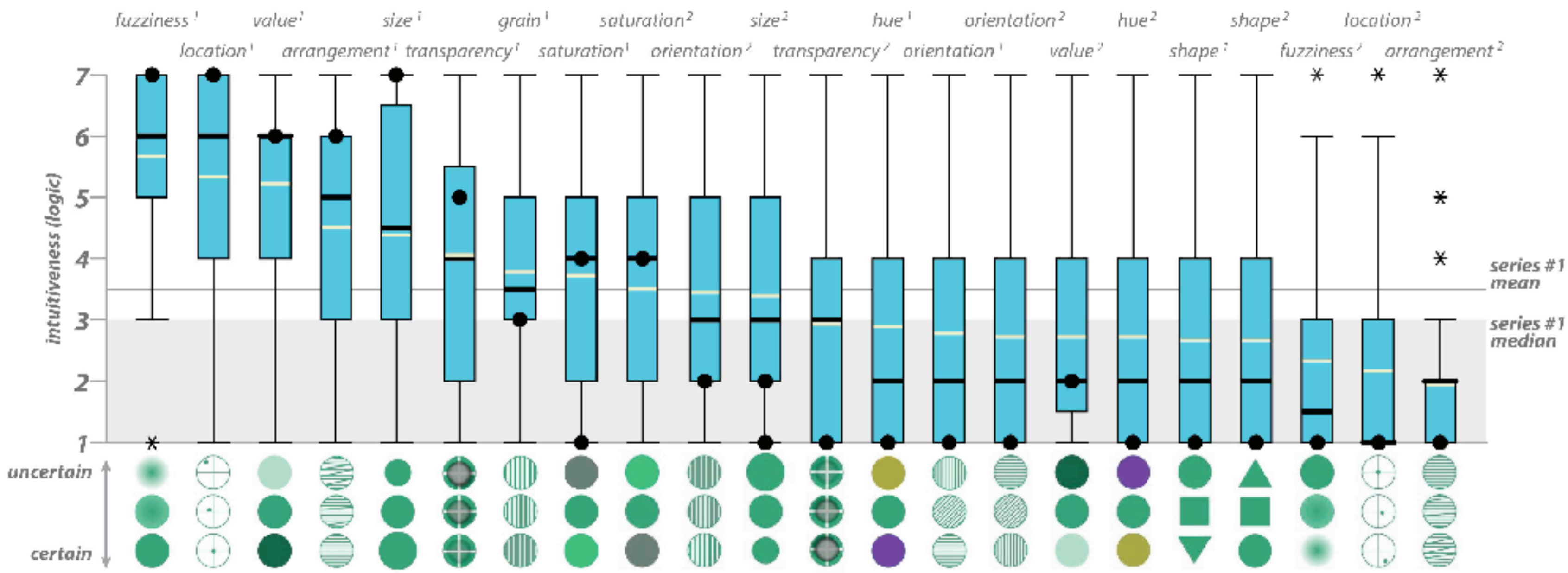



MacEachren, Alan et al. Visual Semiotics & Uncertainty Visualization: An empirical study. IEEE VIS, 2012.

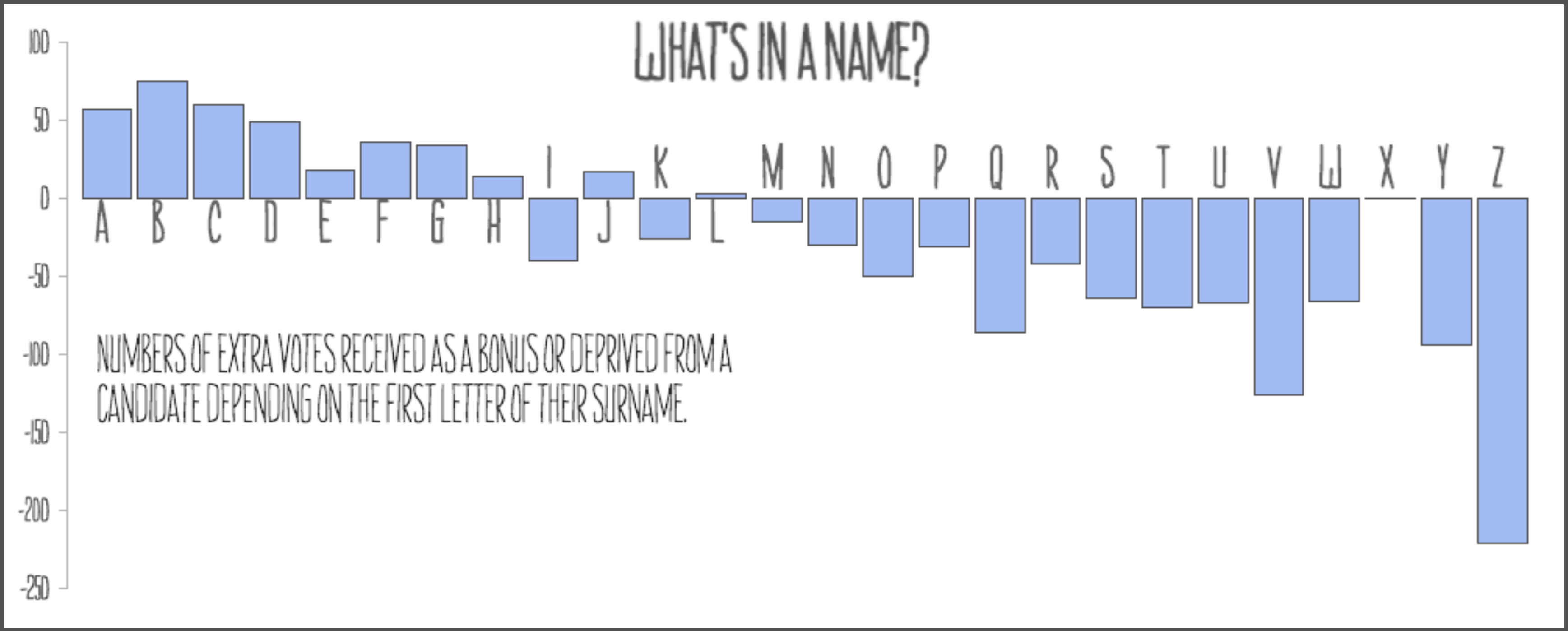
# Semiotics of Uncertainty



## SERIES #1: GENERAL UNCERTAINTY BY VISUAL VARIABLE



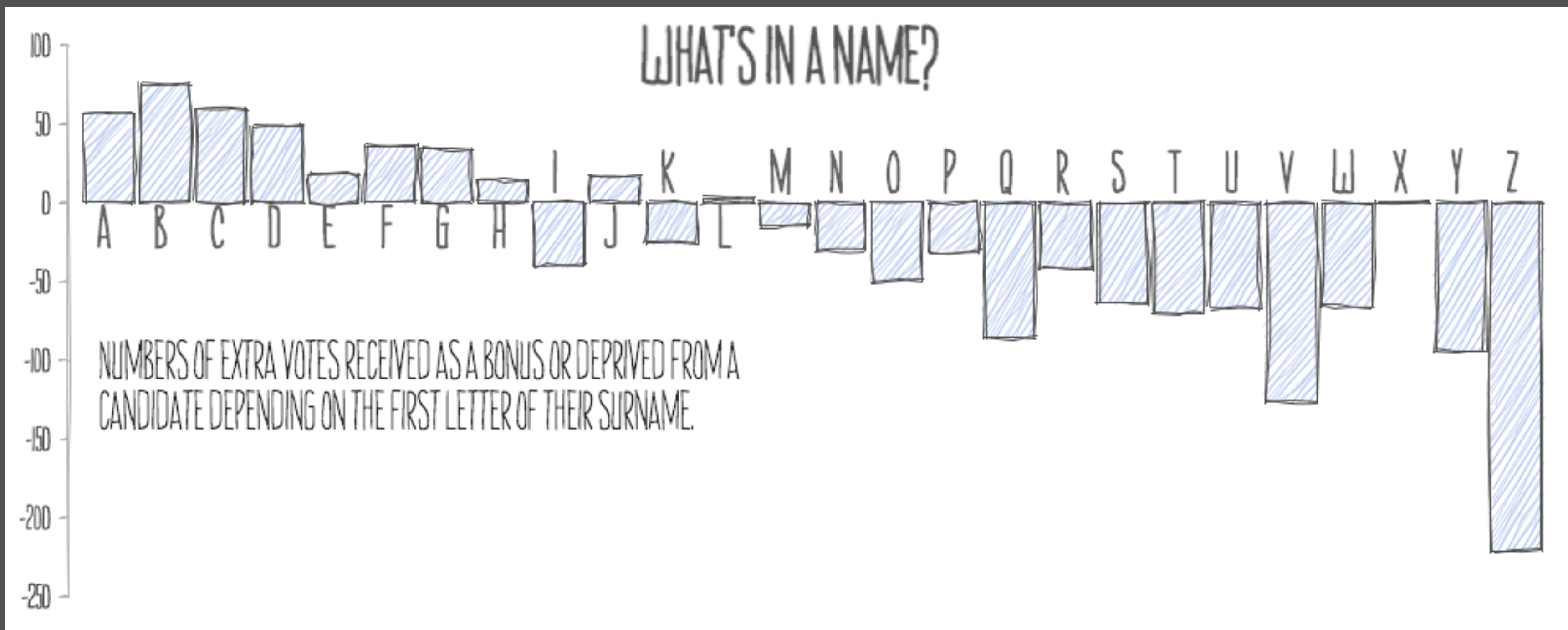
# "Sketchiness"




Wood, Jo et al. Sketchy rendering for information visualization. IEEE VIS, 2012.

Boukhelifa, Nadia et al. Evaluating skrtchiness as a visual variable for the depiction of qualitative uncertainty. IEEE VIS, 2012.

# “Sketchiness”




Wood, Jo et al. Sketchy rendering for information visualization. IEEE VIS, 2012.

Boukhelifa, Nadia et al. Evaluating skrtchiness as a visual variable for the depiction of qualitative uncertainty. IEEE VIS, 2012.

# Encoding Uncertainty

Some visual variables (like fuzziness and value) have a **semiotic connection** to uncertainty.

However, intuitive variables may not always be accurately interpreted!

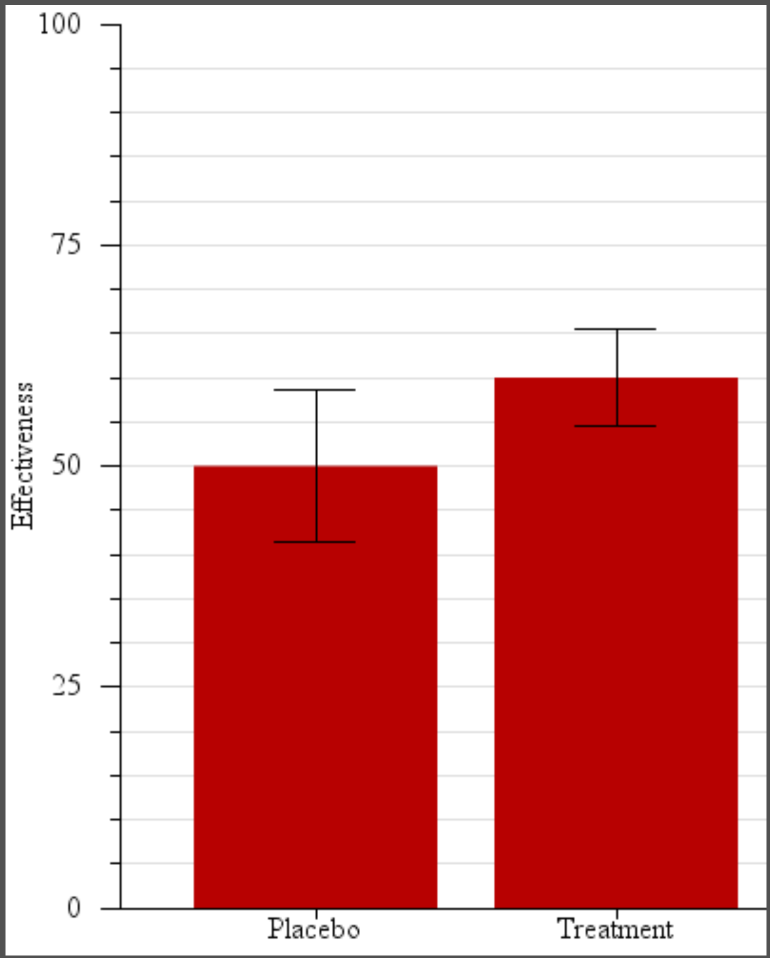
# p-value

The probability of results at least as extreme as the observed results, given some null hypothesis.

If $p<\alpha$ (usually 0.05), then the result is considered to be *statistically significant*.

# Error Bars

Is the treatment *statistically significantly* better than the placebo?



# Error Bars

Standard Deviation?

Standard Error ($\sigma/\sqrt{n}$)

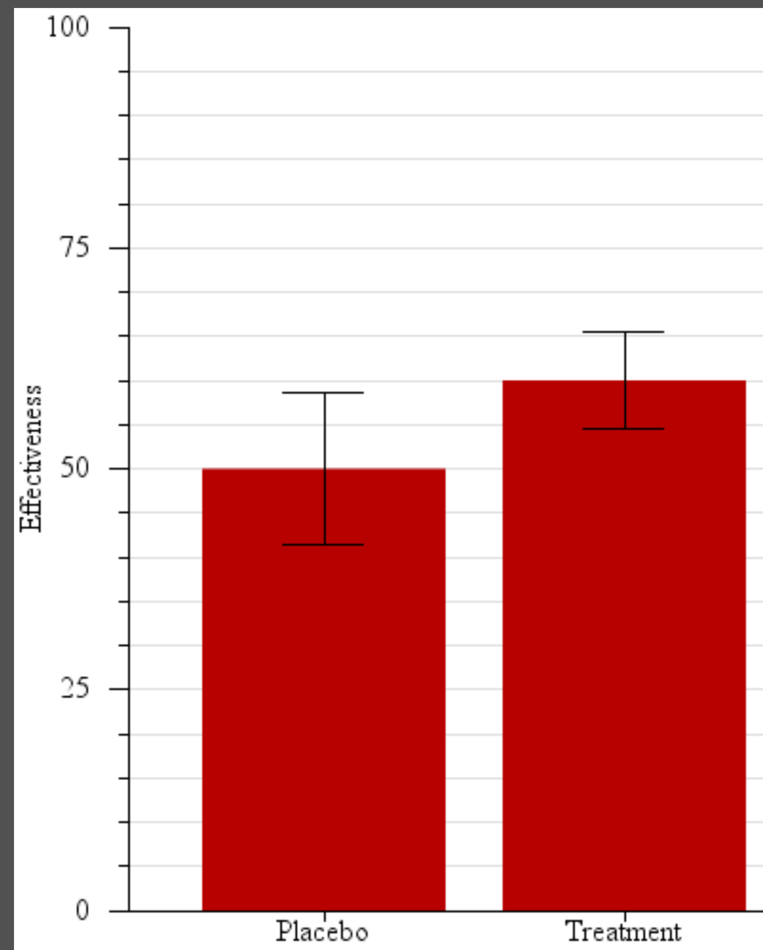
T-Confidence Interval?

Z-Confidence Interval?

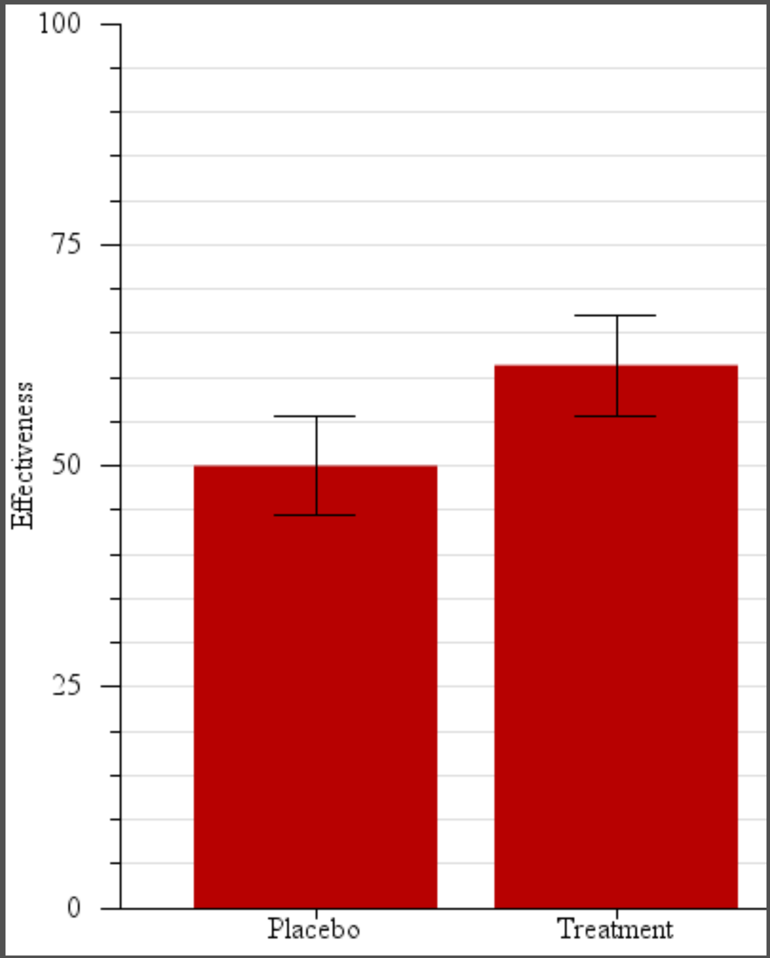
Bootstrapped Interval?

Min/Max?

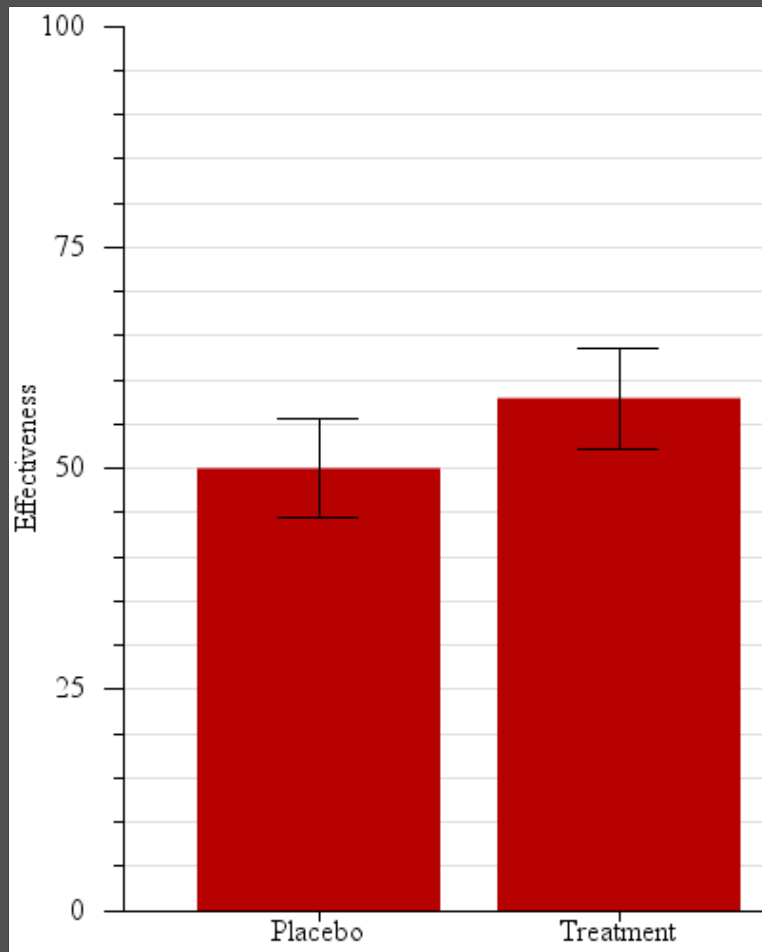
1.5*IQR (Q3-Q1)?



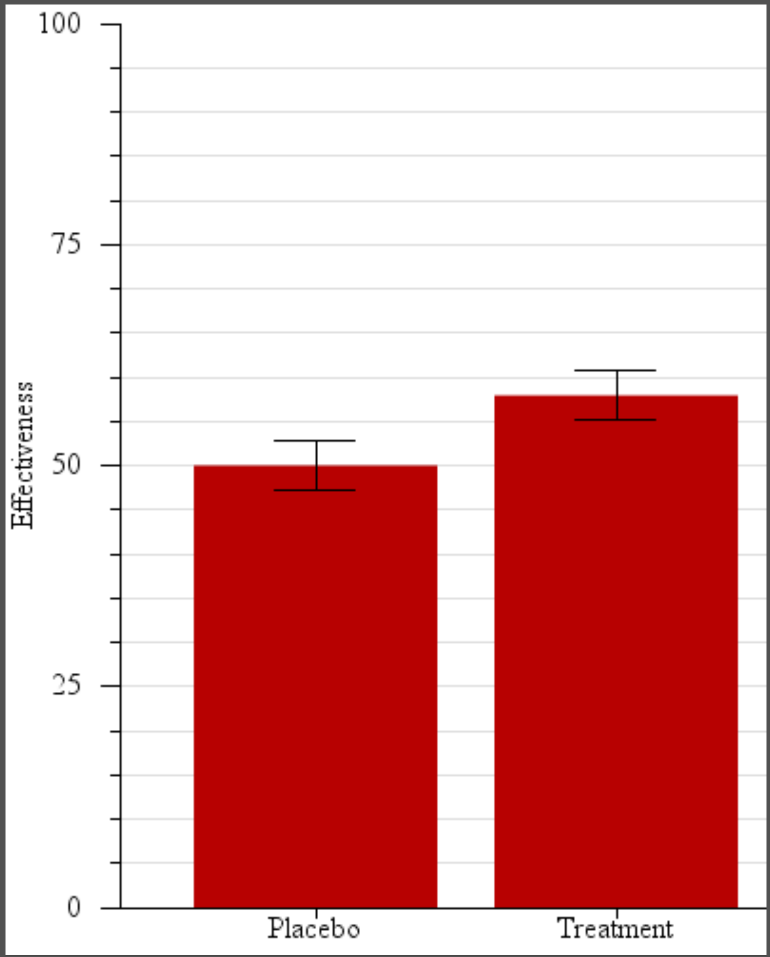
# Guess the p-value



# Guess the p-value

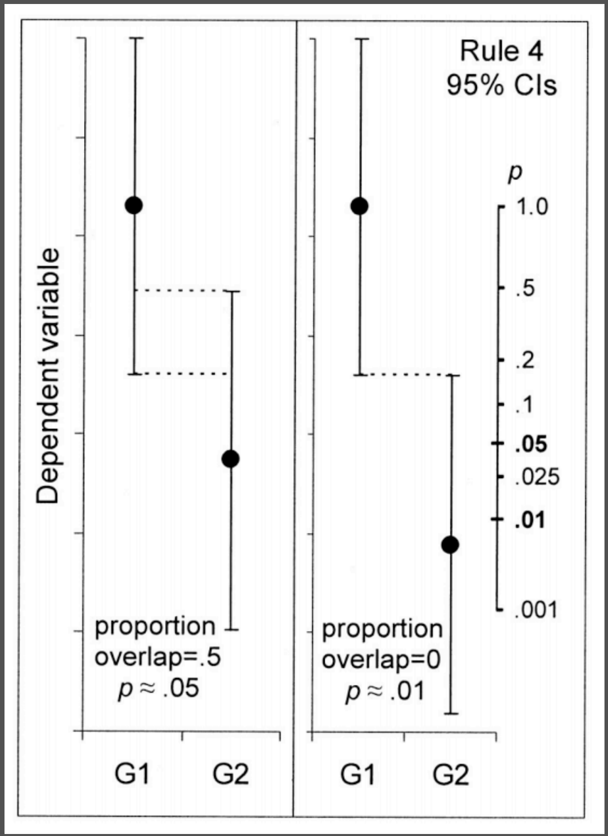


# Guess the p-value

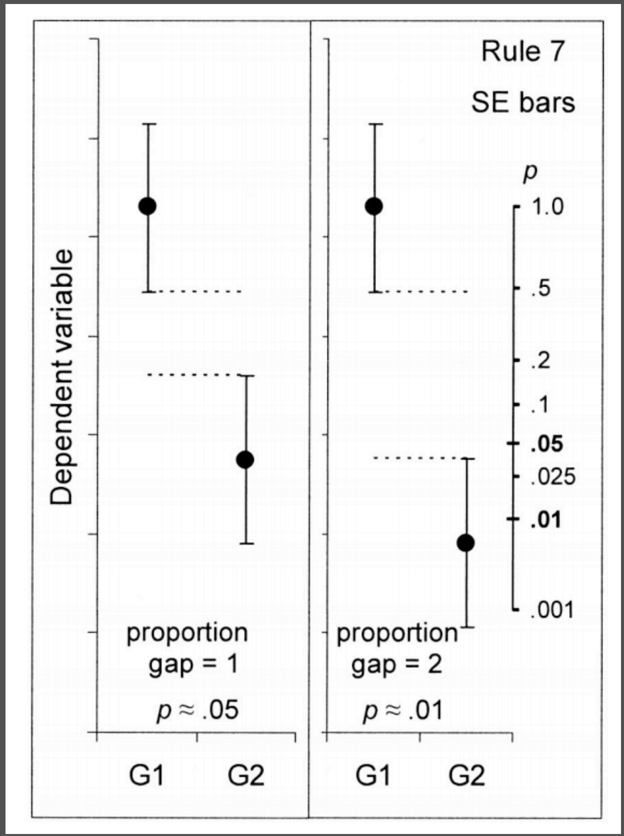


# Inference by Eye

## 95% CIs

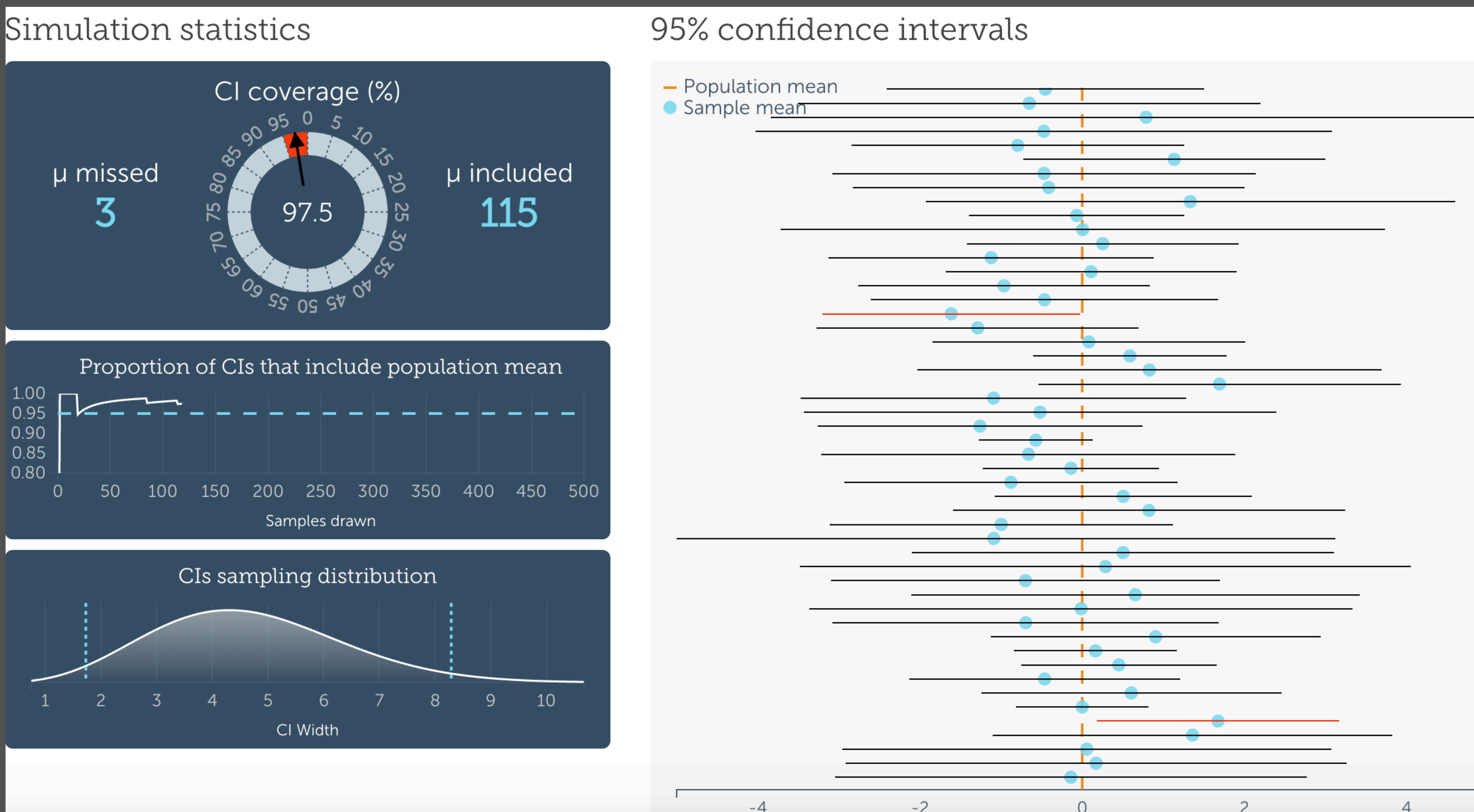


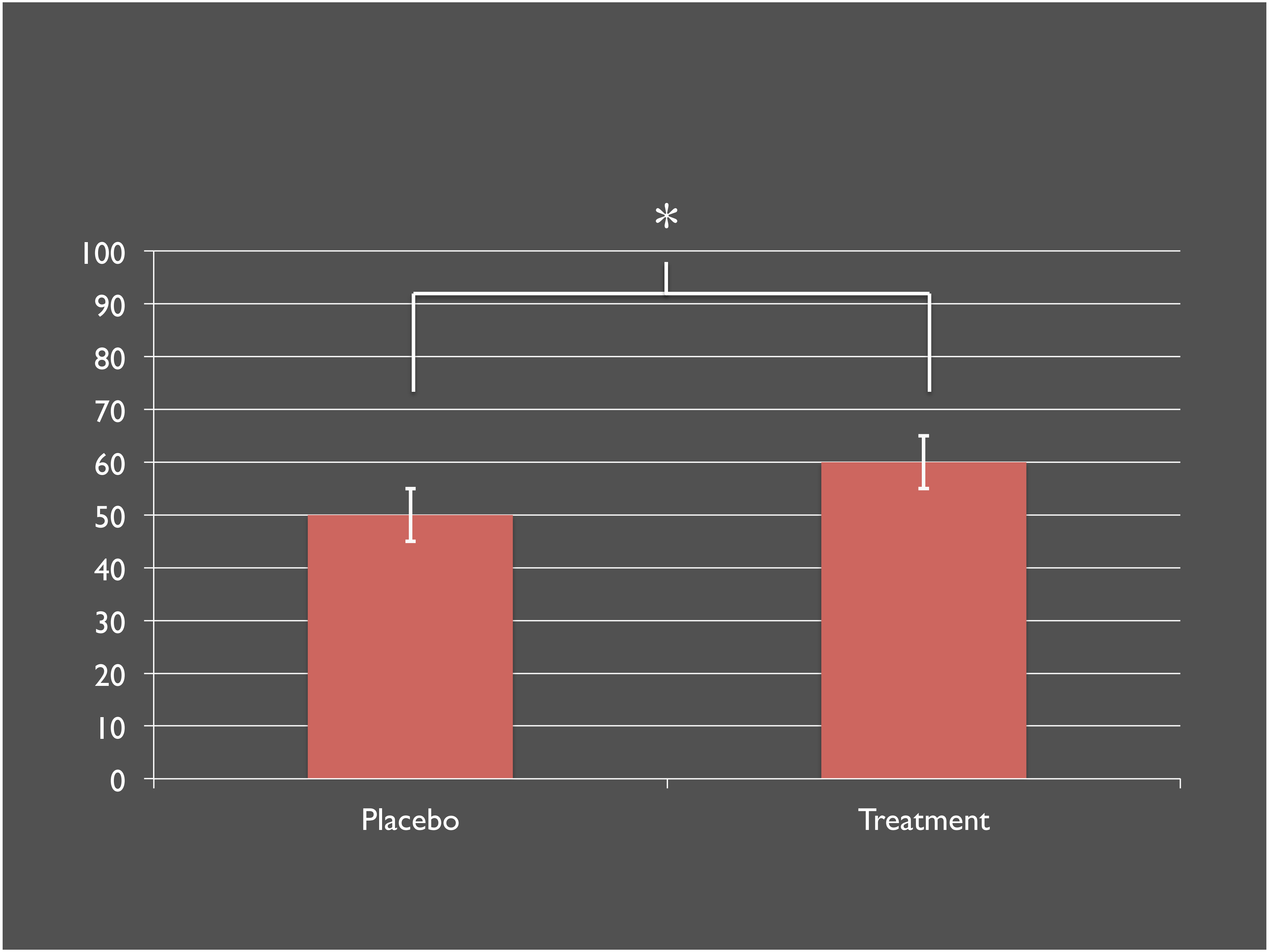

## Standard Error



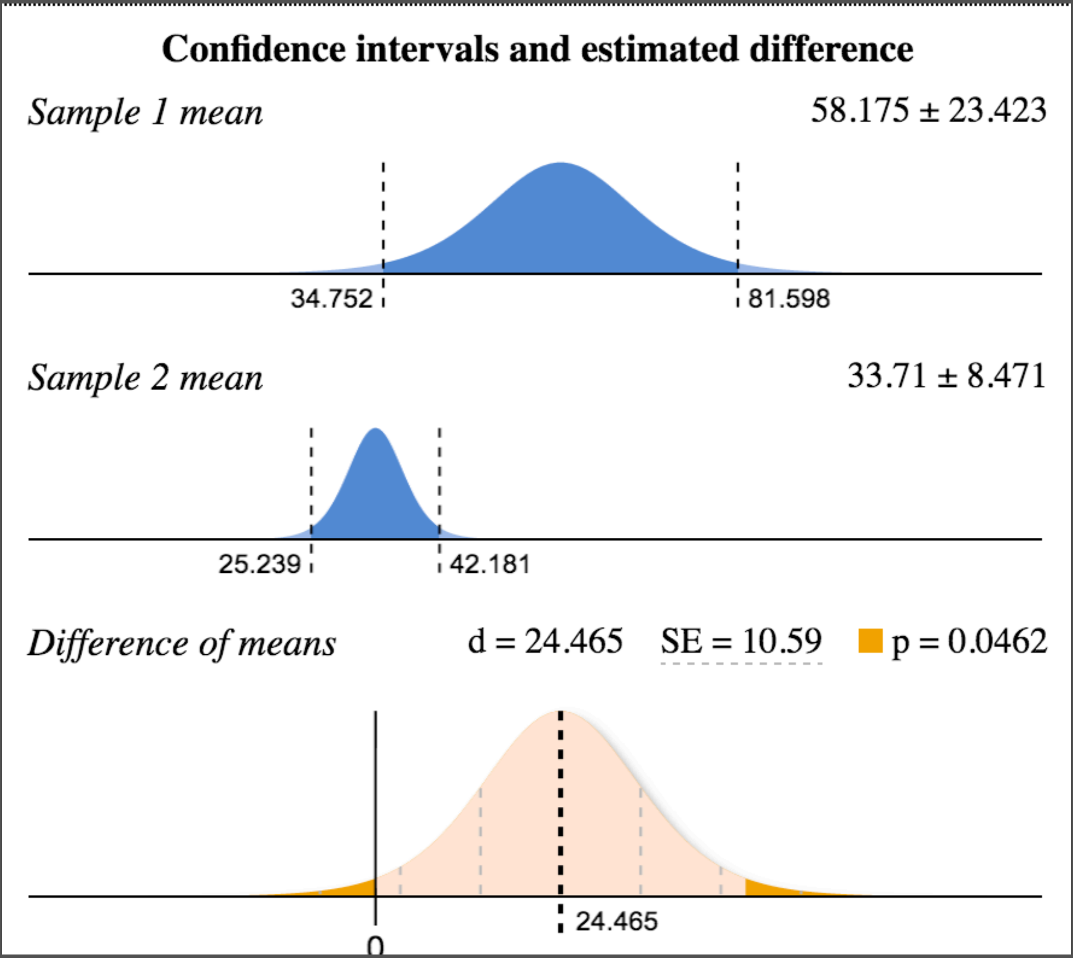

Cumming, Geoff and Finch, Sue. Inference by eye: confidence intervals and how to read pictures of data. American Psychologist, 2005.

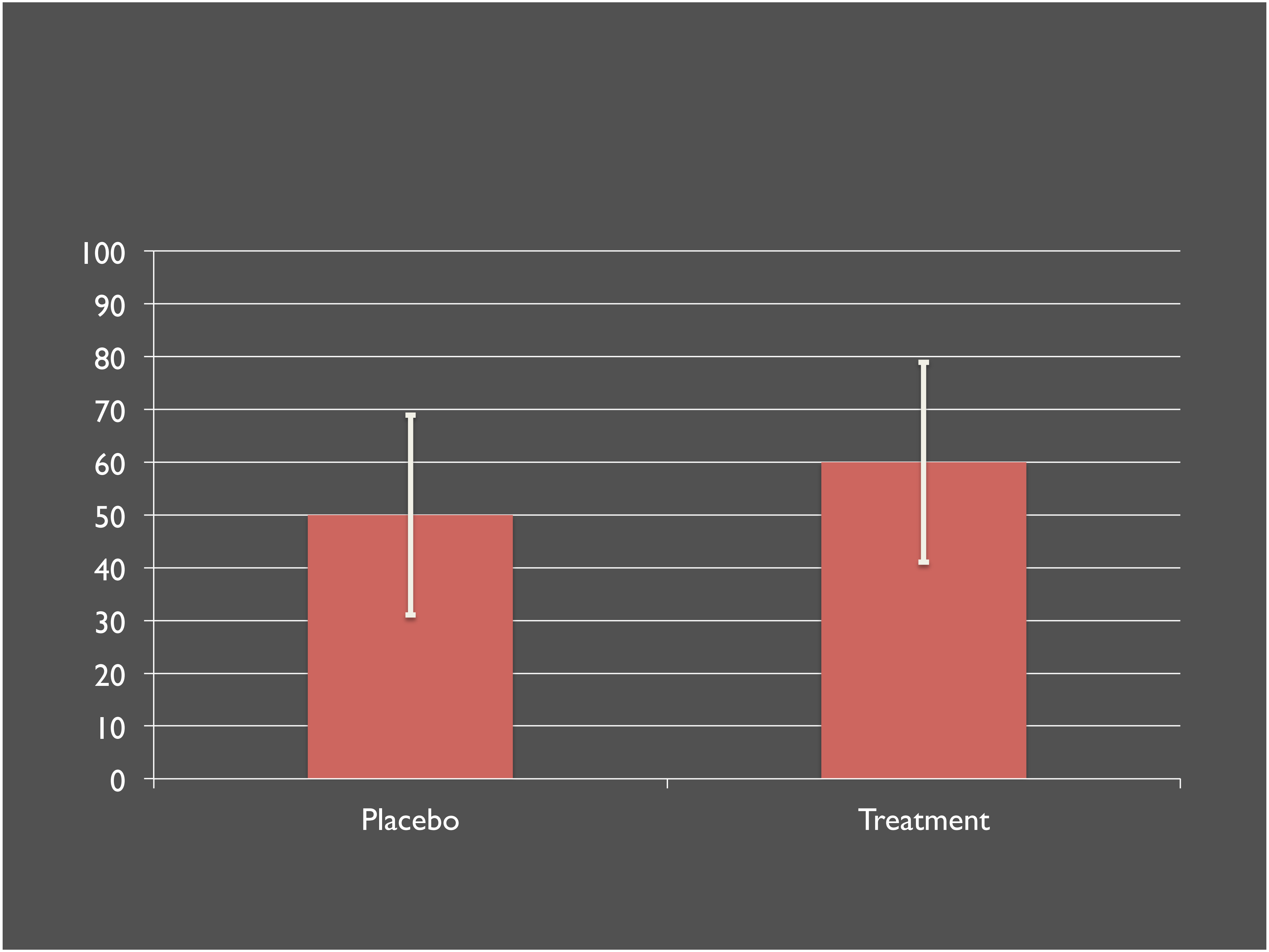
# Confidence Intervals

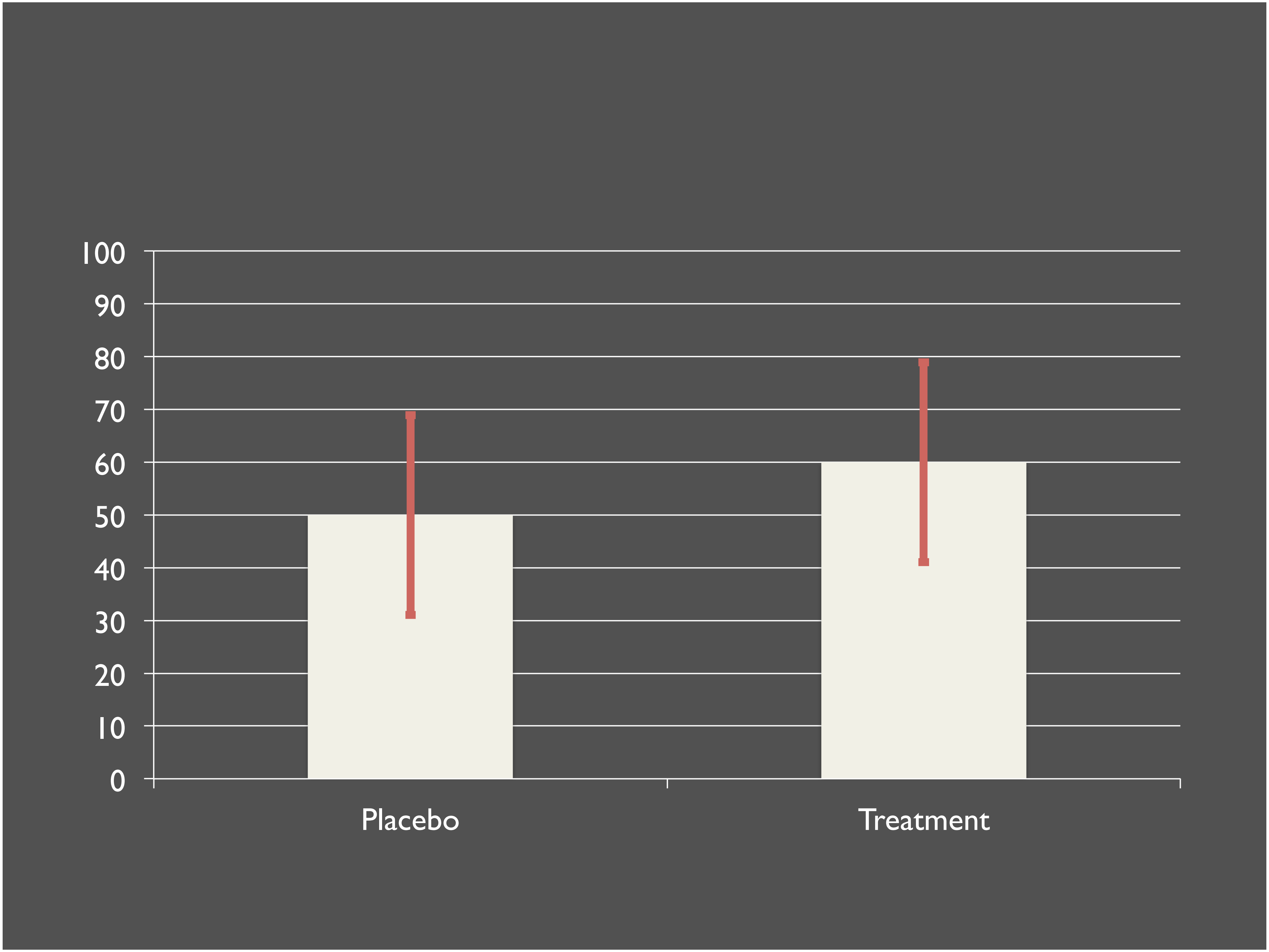



*
100
90
80
70
60
50
40
30
20
10
0
Placebo
Treatment

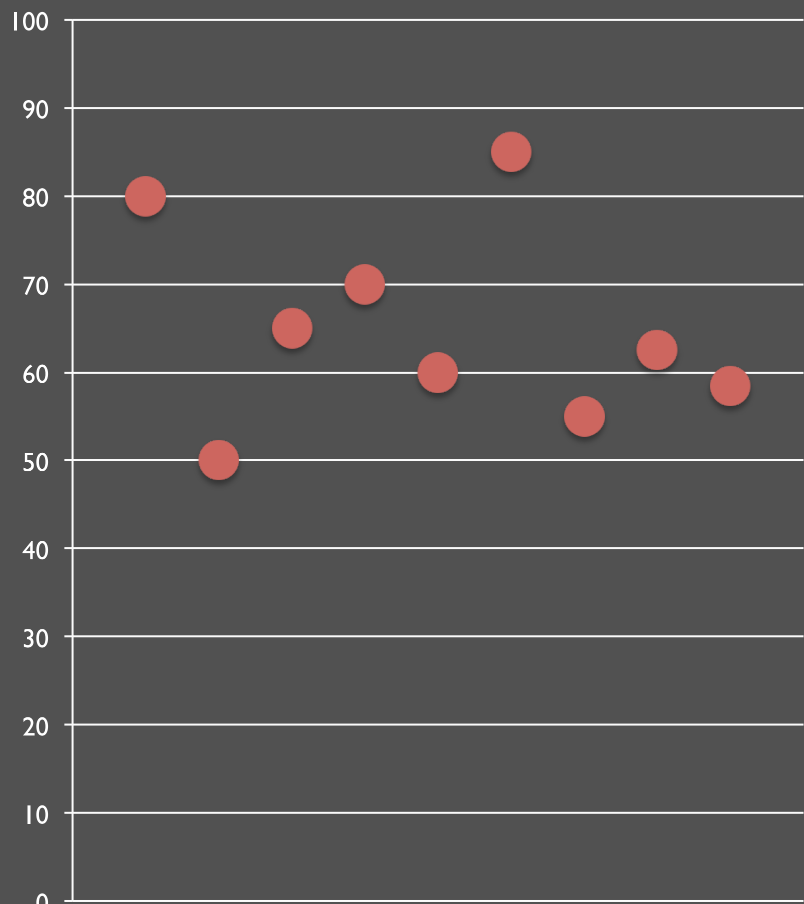
# T-Tests and Confidence Intervals

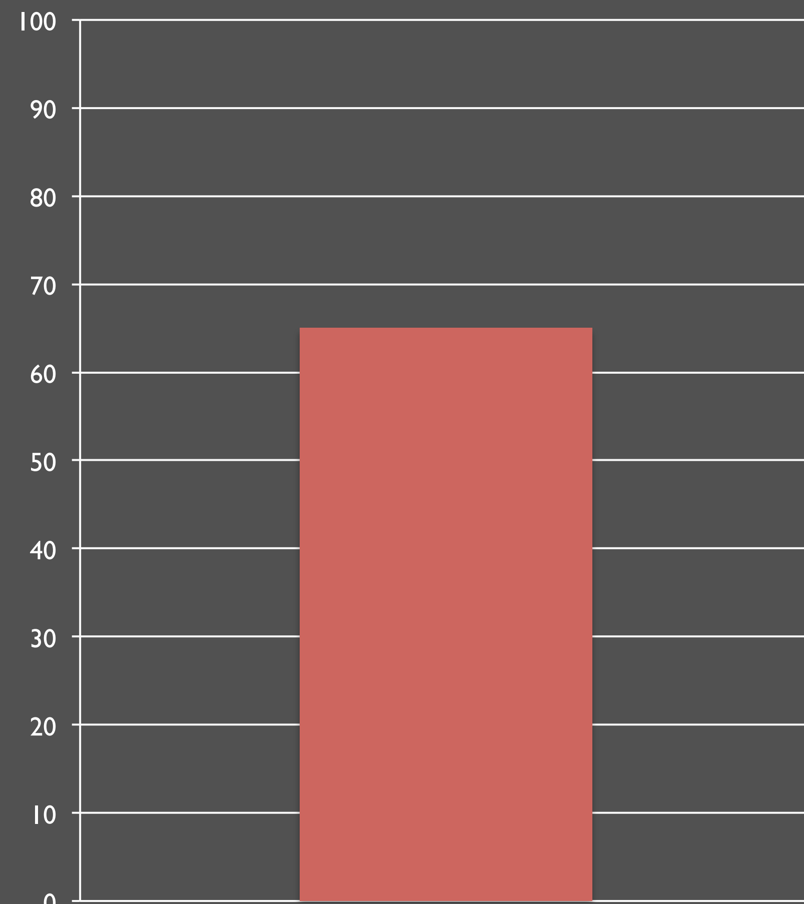



100
90
80
70
60
50
40
30
20
10
0
Placebo
Treatment


100
90
80
70
60
50
40
30
20
10
0
Placebo
Treatment

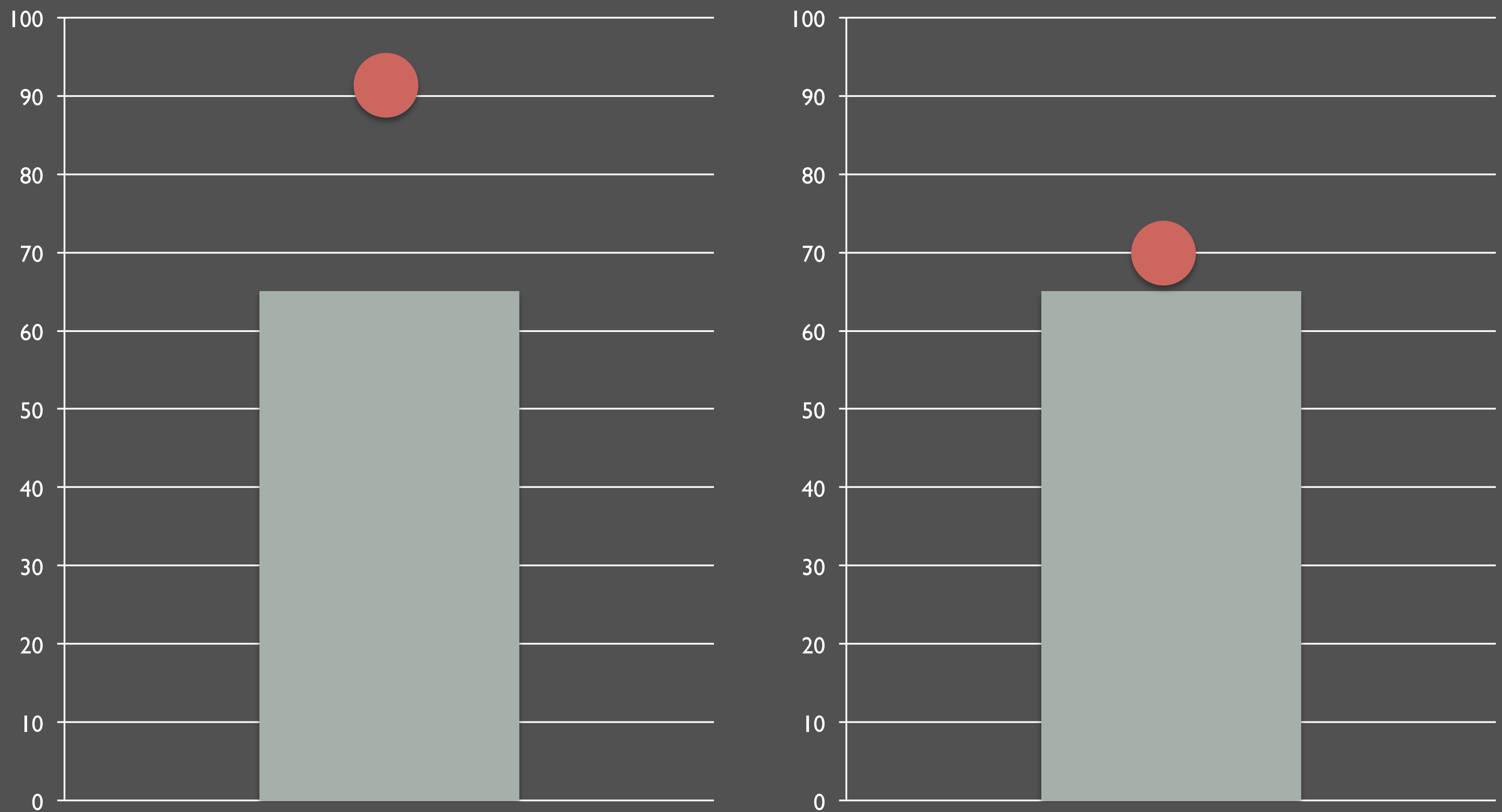
# Within-the-bar bias



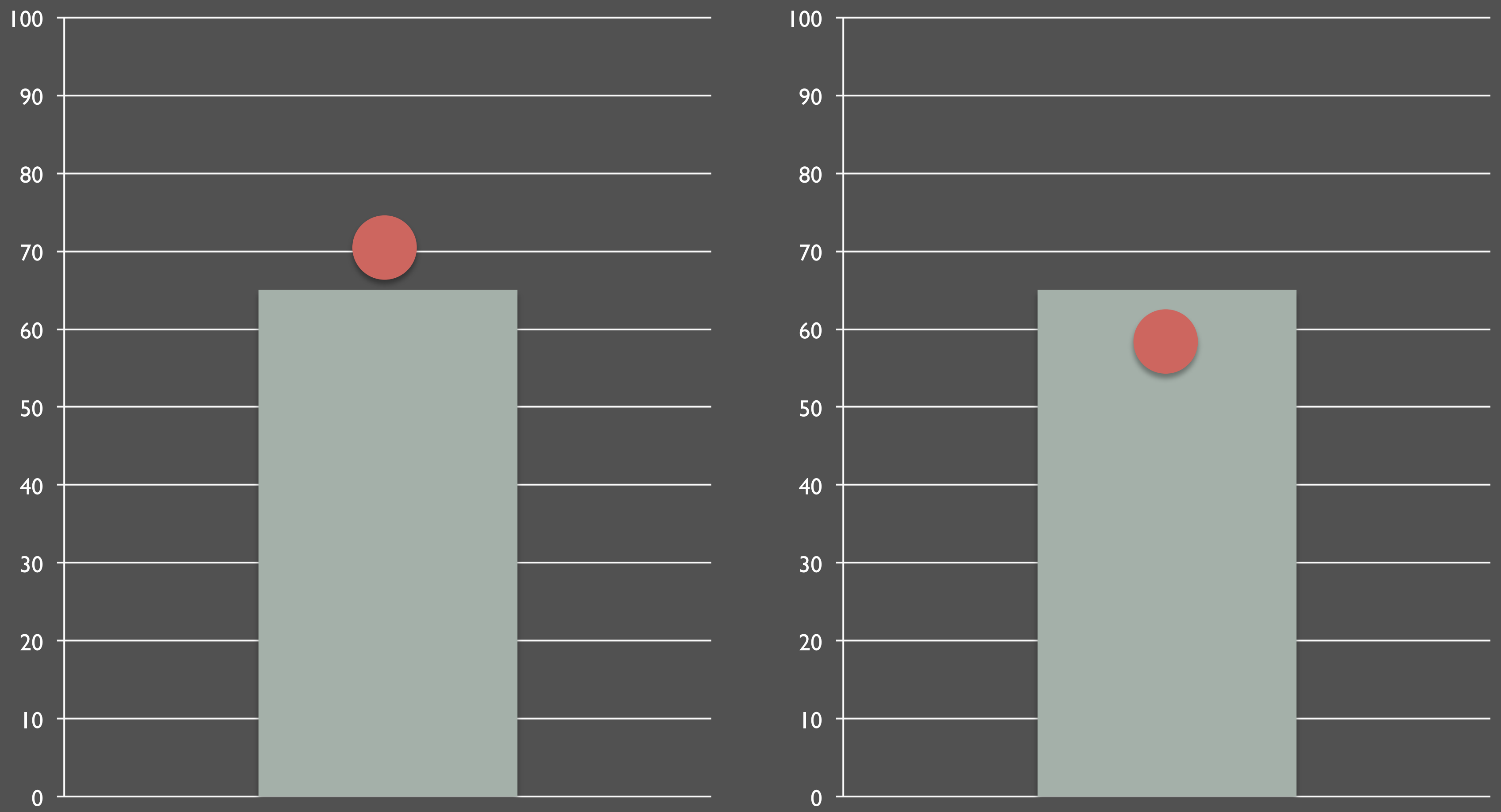




Newman, George E, and Brian J Scholl. "Bar graphs depicting averages are perceptually misinterpreted: the within-the-bar bias." Psychonomic bulletin & review 19.4 (2012): 601–7.

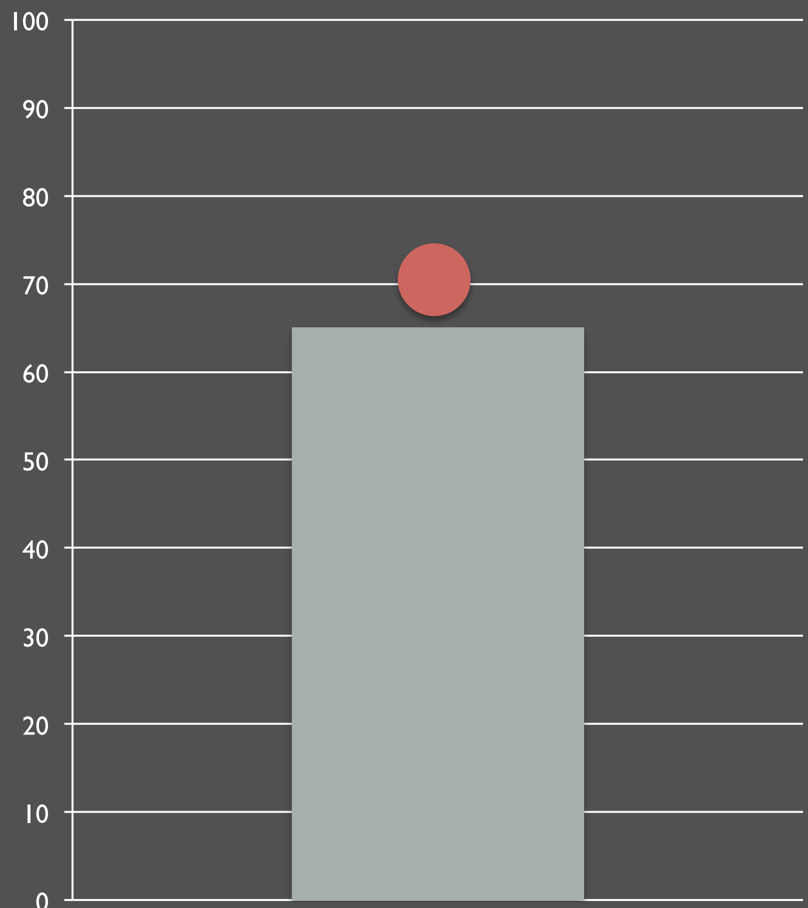
# Within-the-bar bias

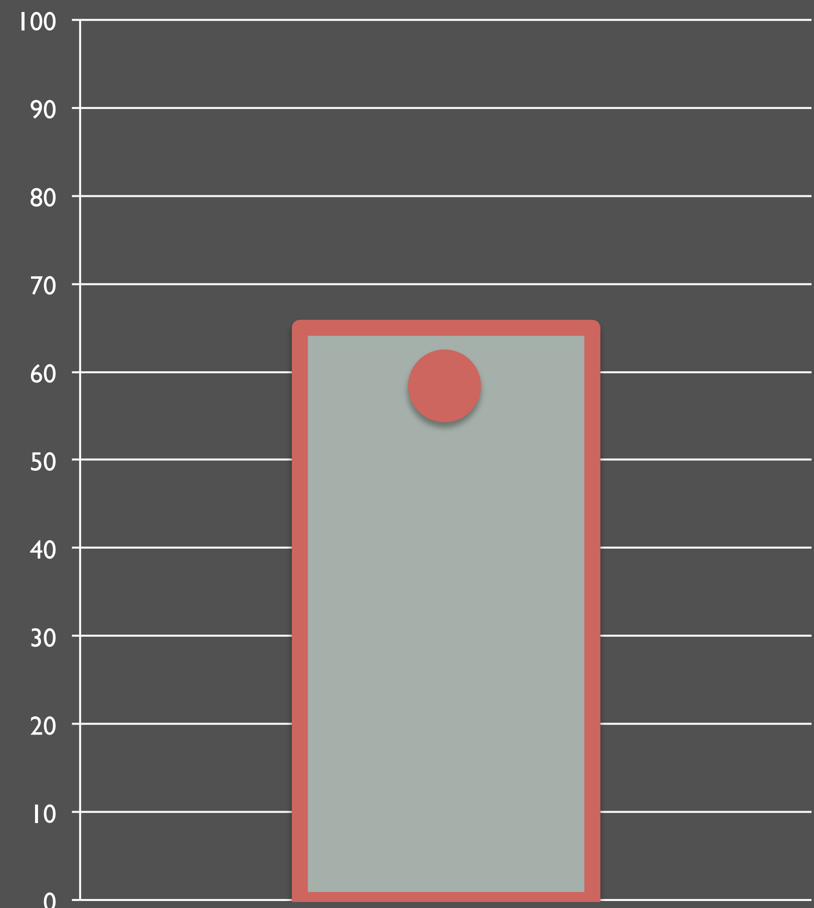


# Within-the-bar bias

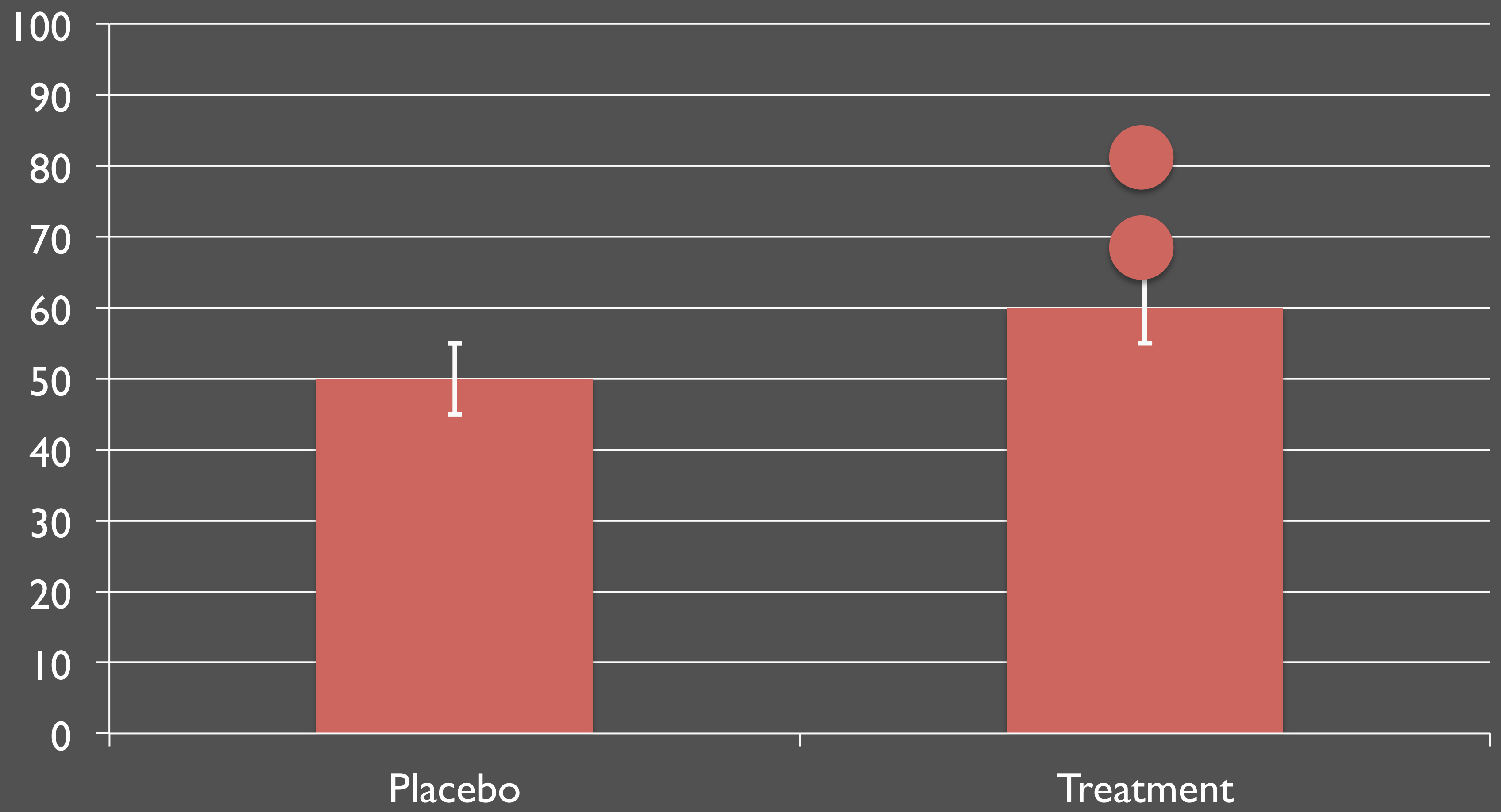


# Within-the-bar bias



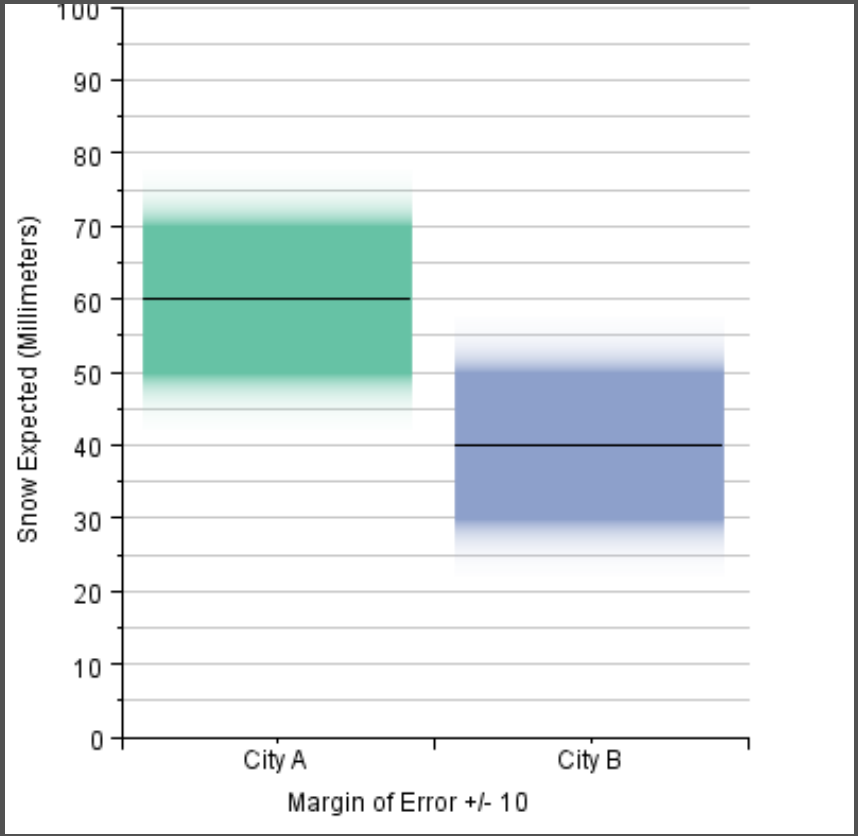



# Binary Bias

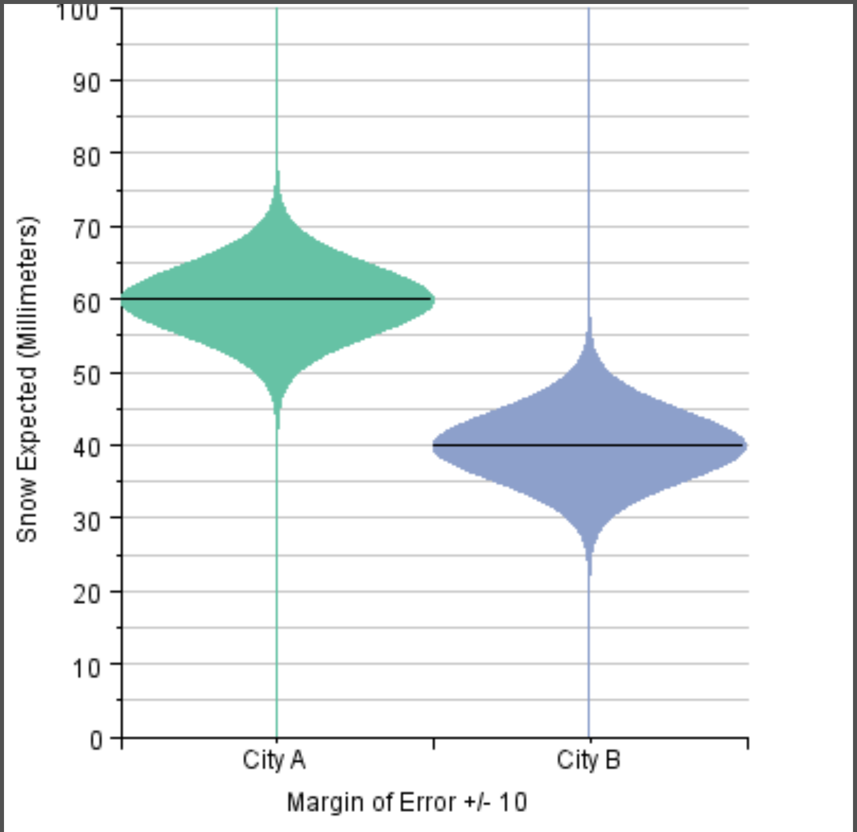


# Alternatives

## Gradient Plot




## Violin Plot



# Model Visualization

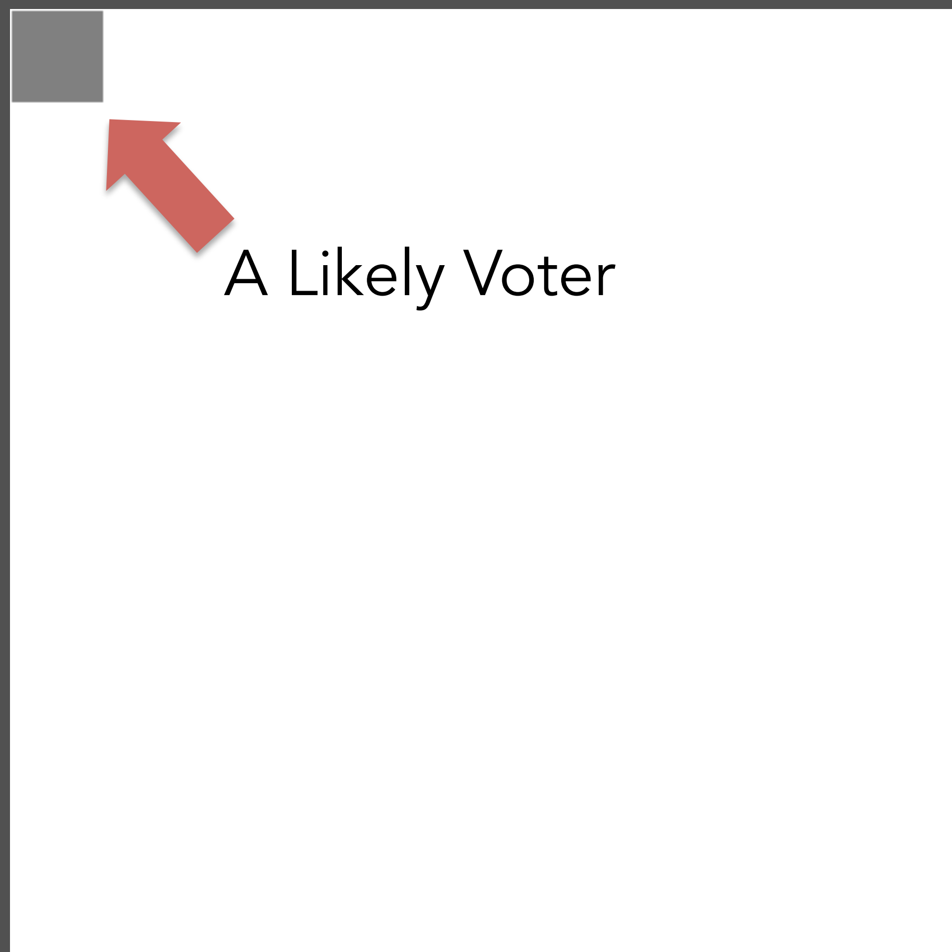
# Polling Data

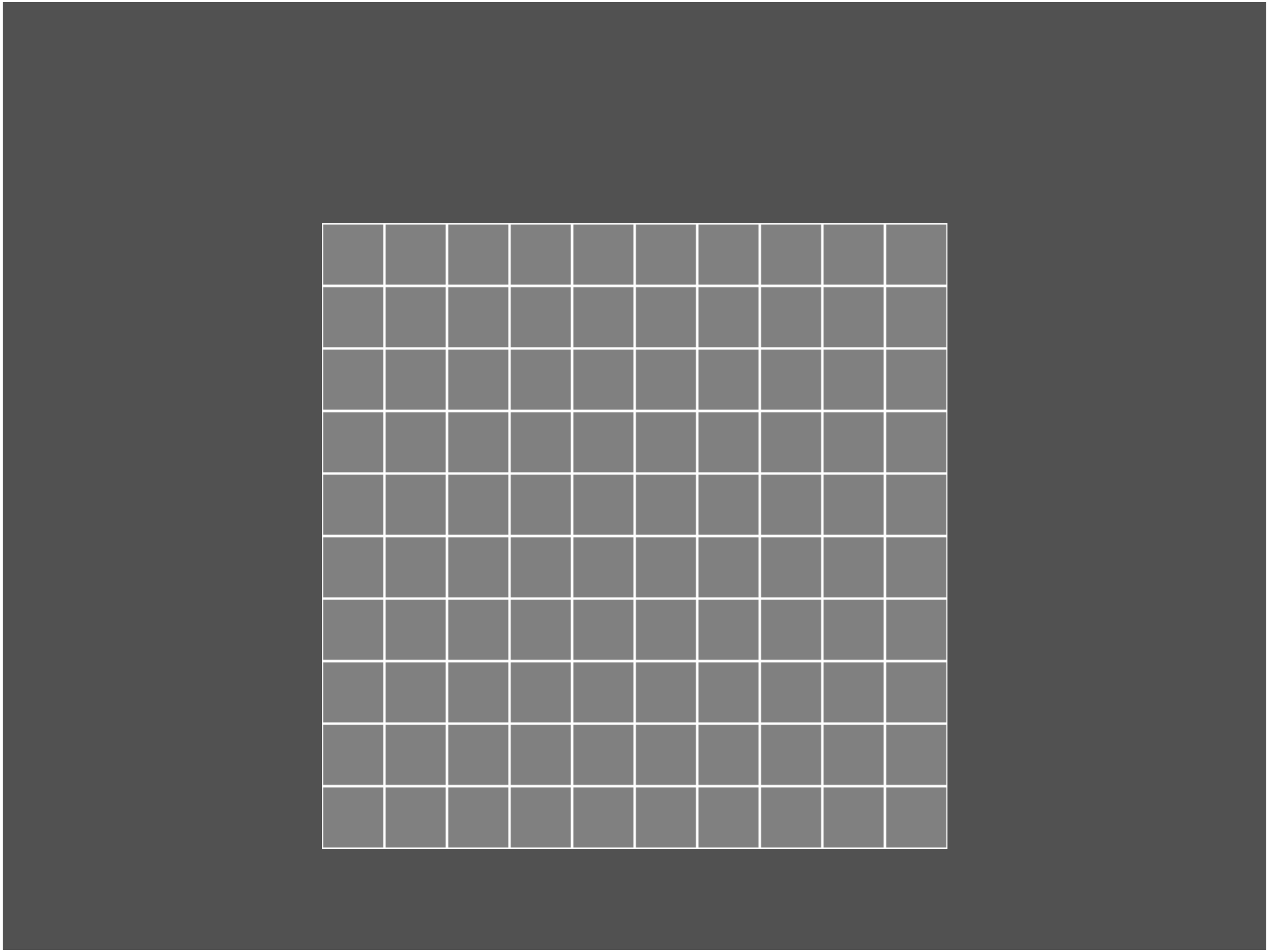
Candidate A is ahead of Candidate B in the polls, with 55% of the likely voters*

# Polling Data

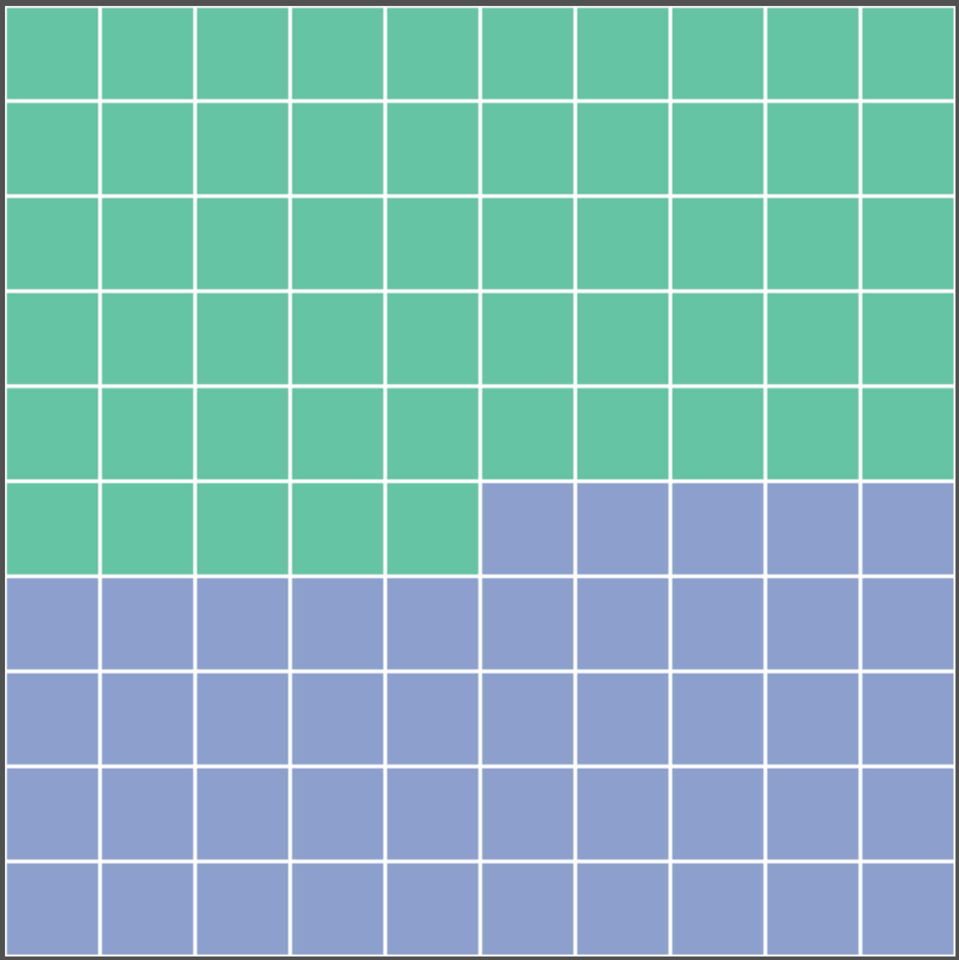
Candidate A is ahead of Candidate B in the polls, with 55% of the likely voters*

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

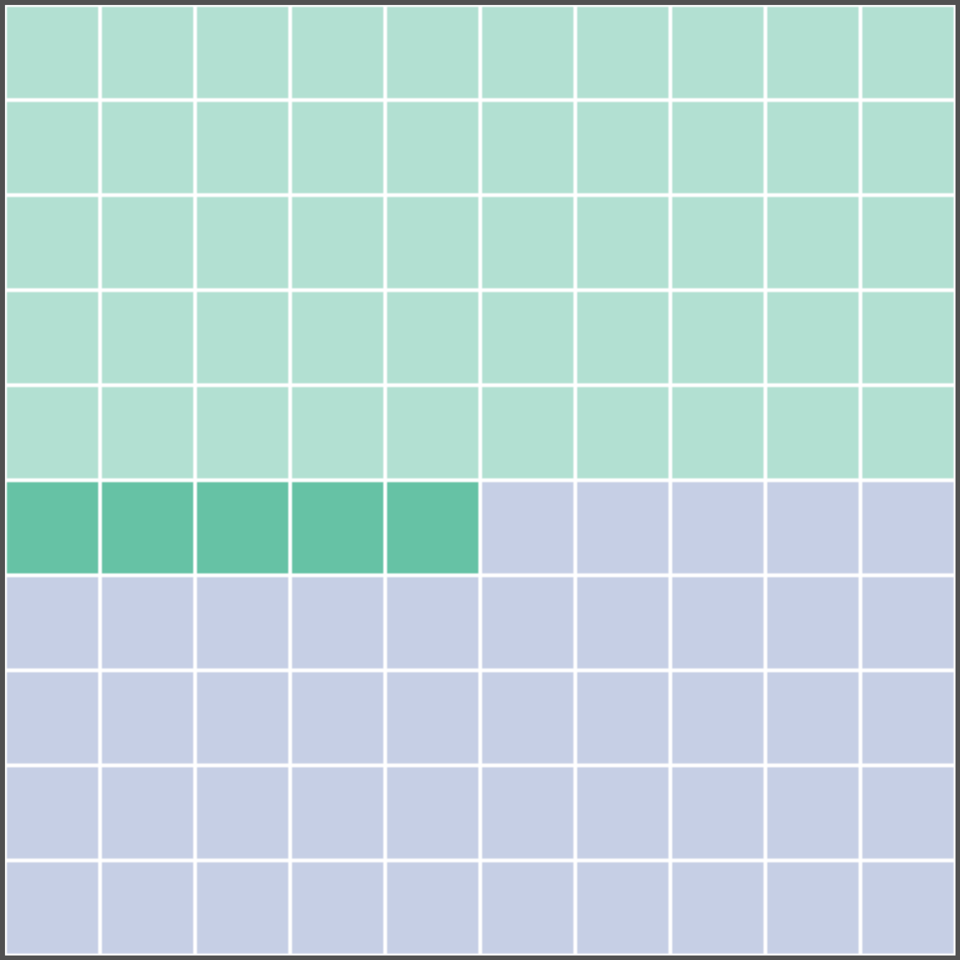

A Likely Voter



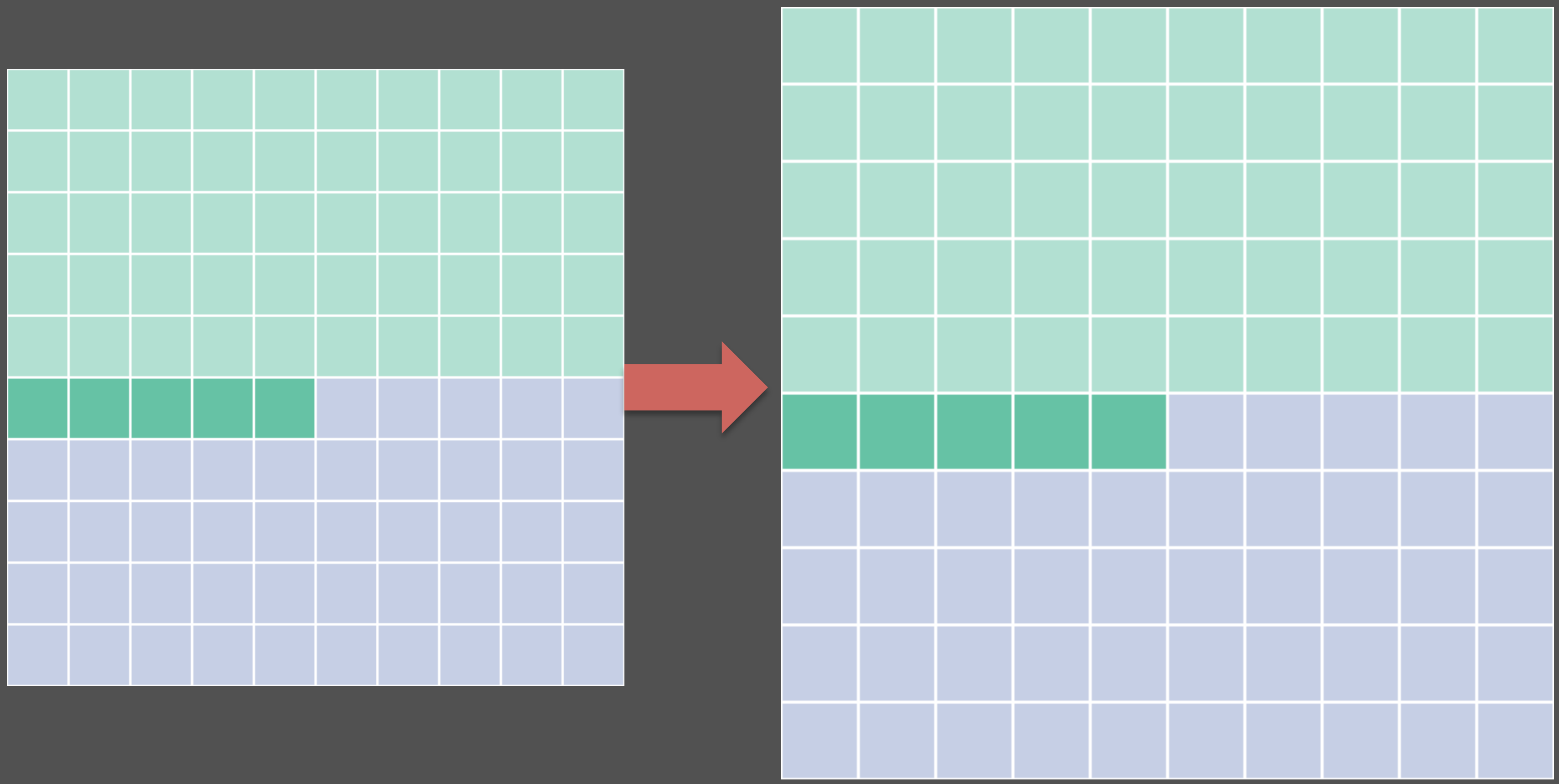
# Poll



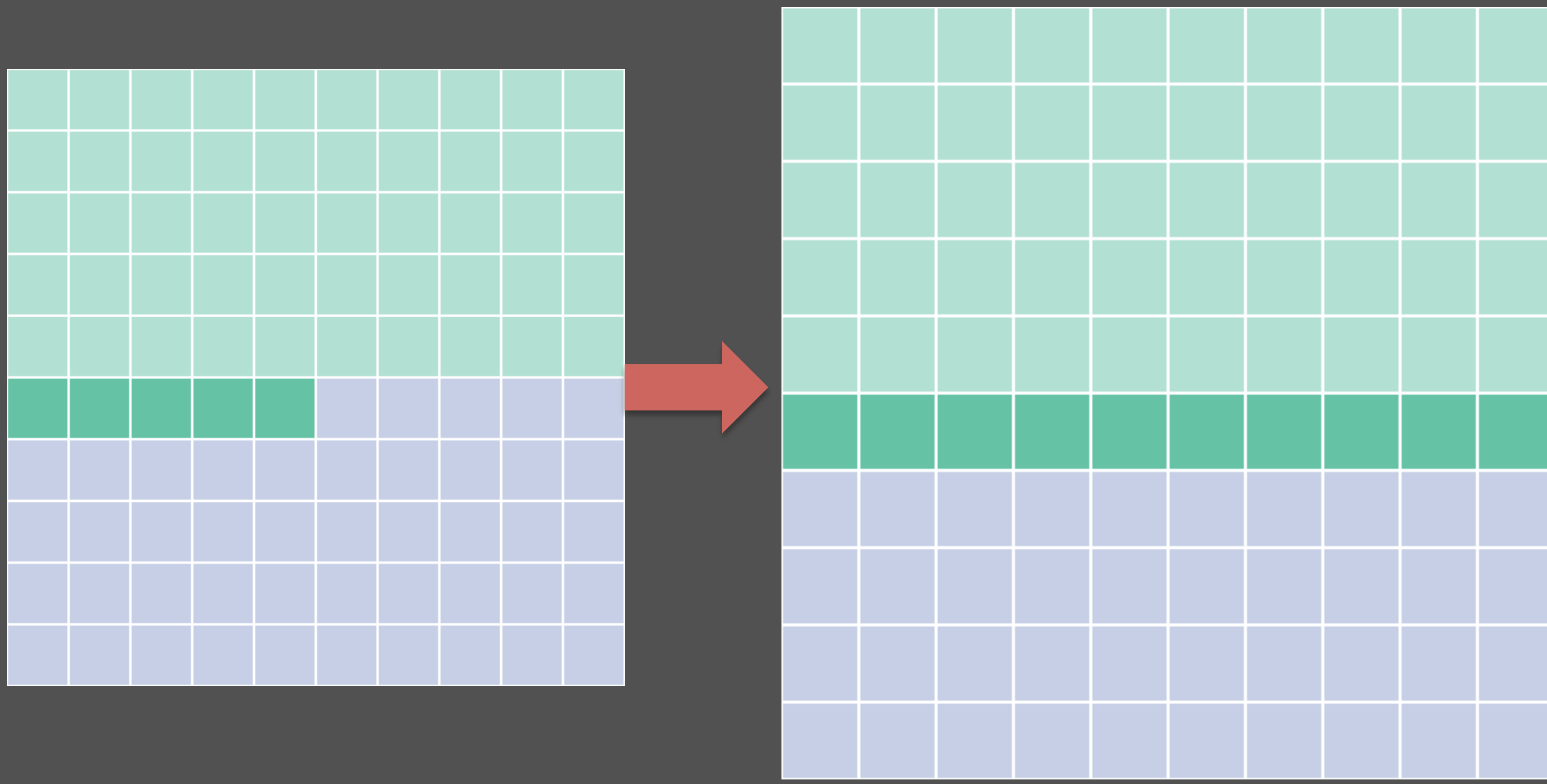
# Poll



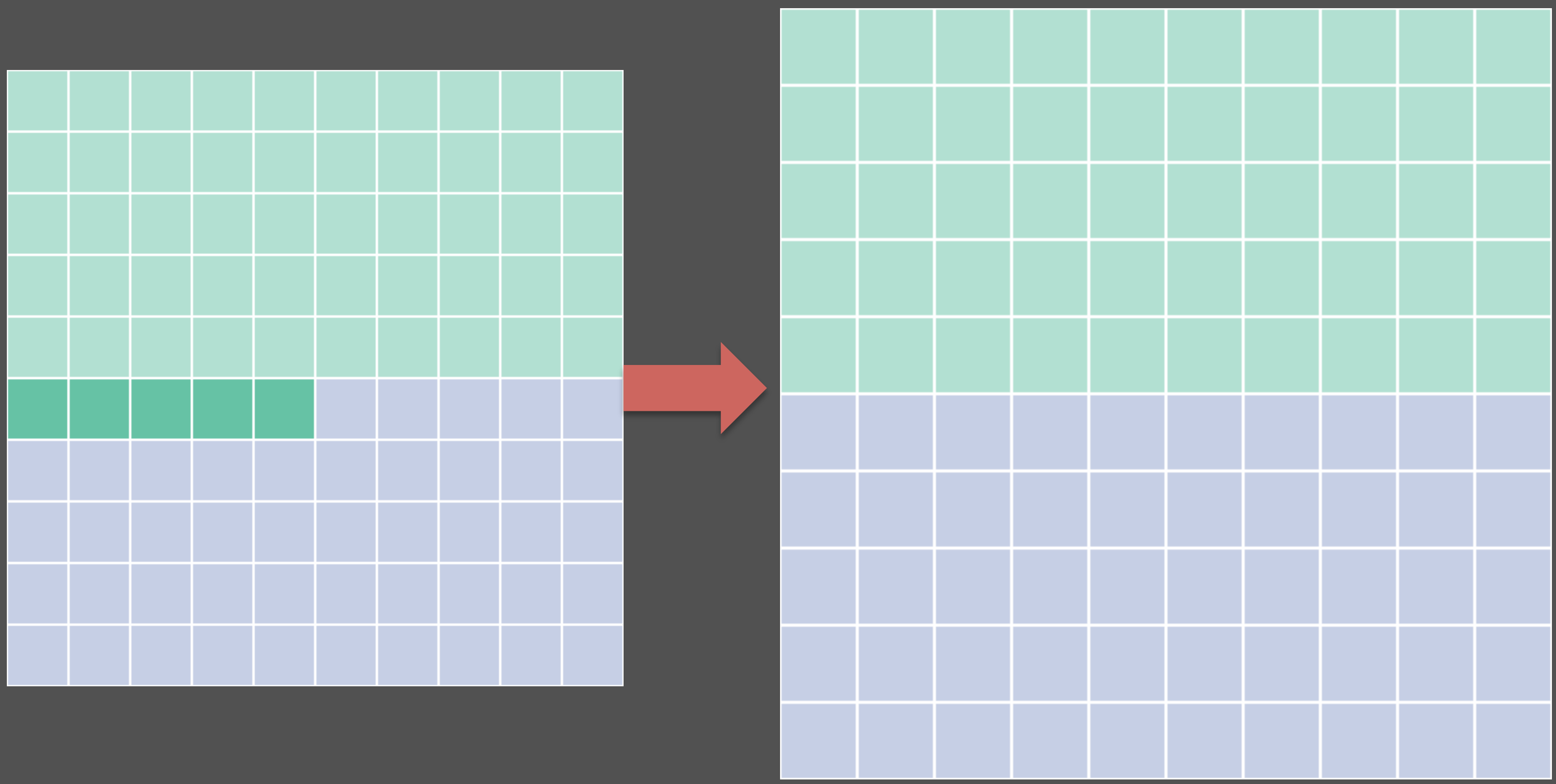
# Actual Election?



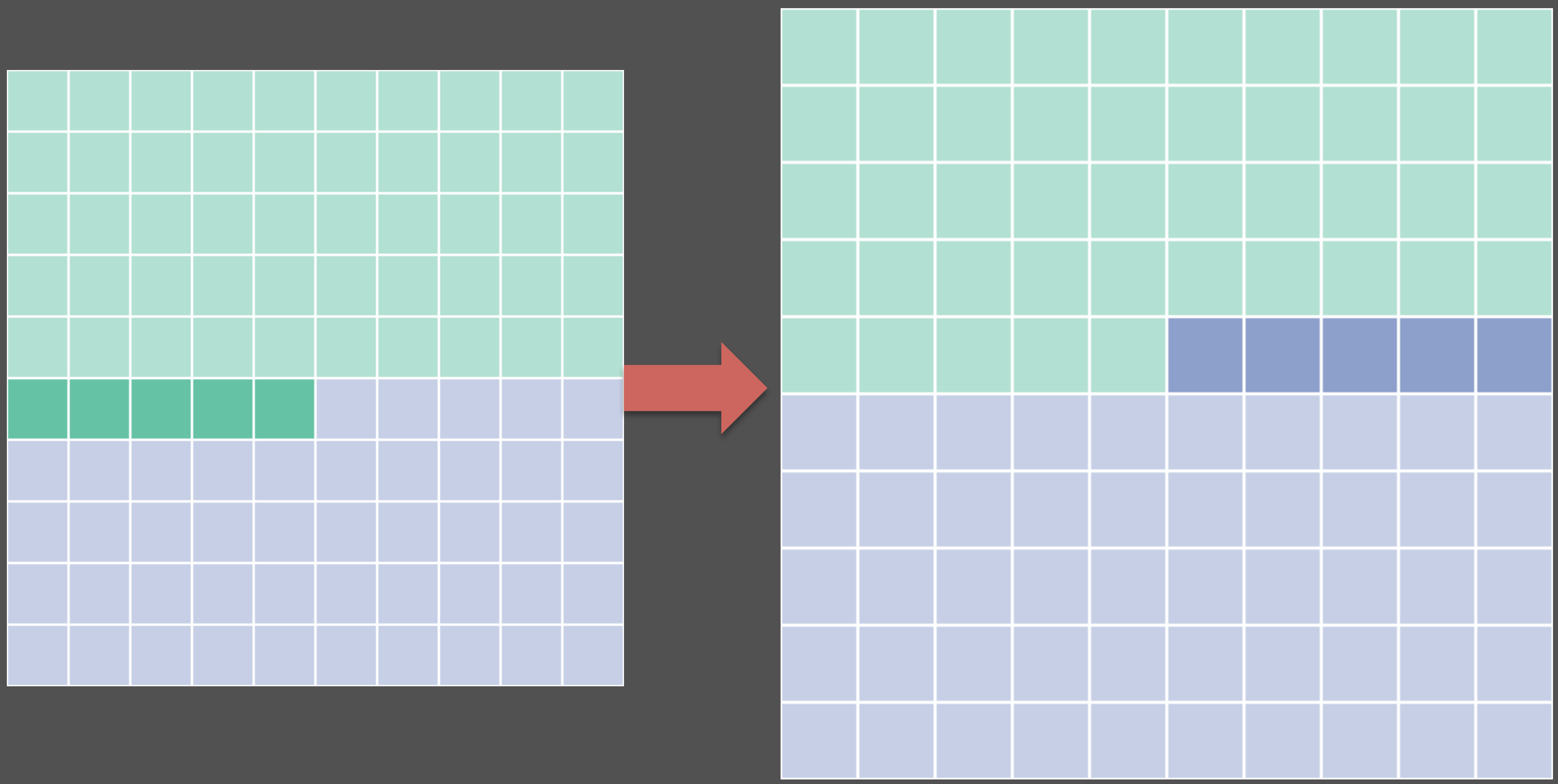
# Actual Election?

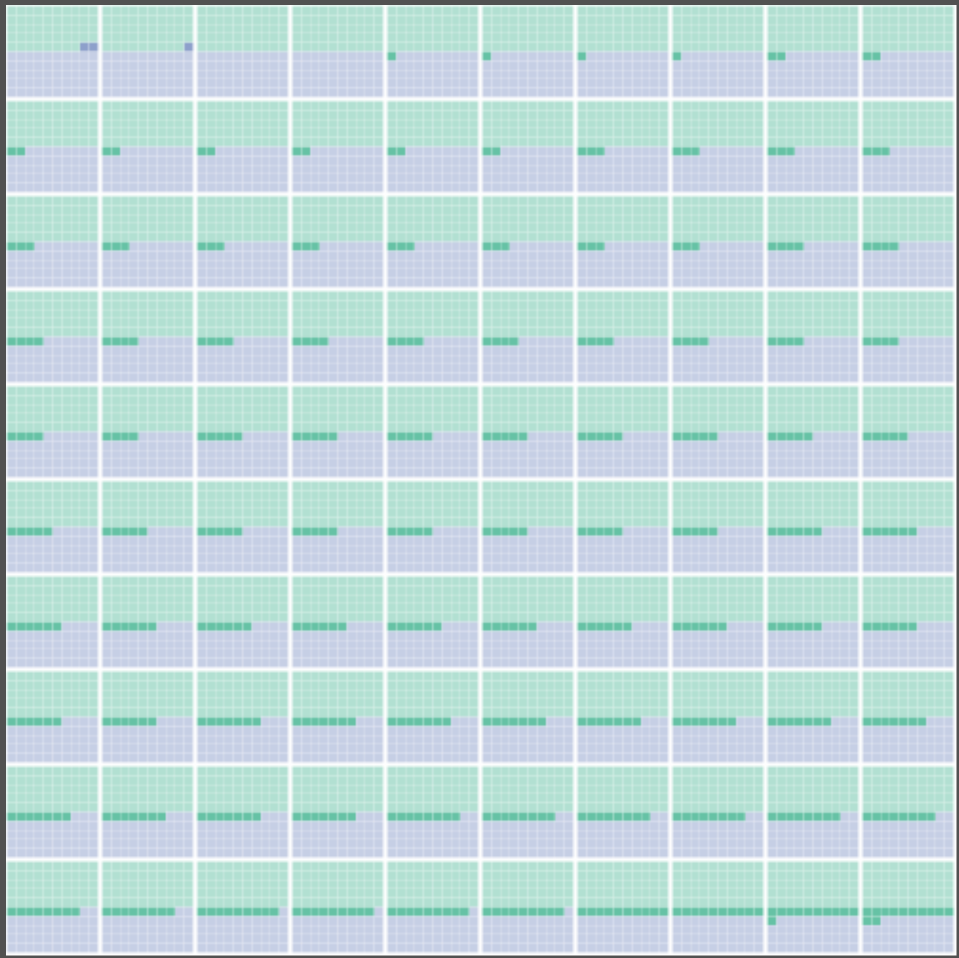


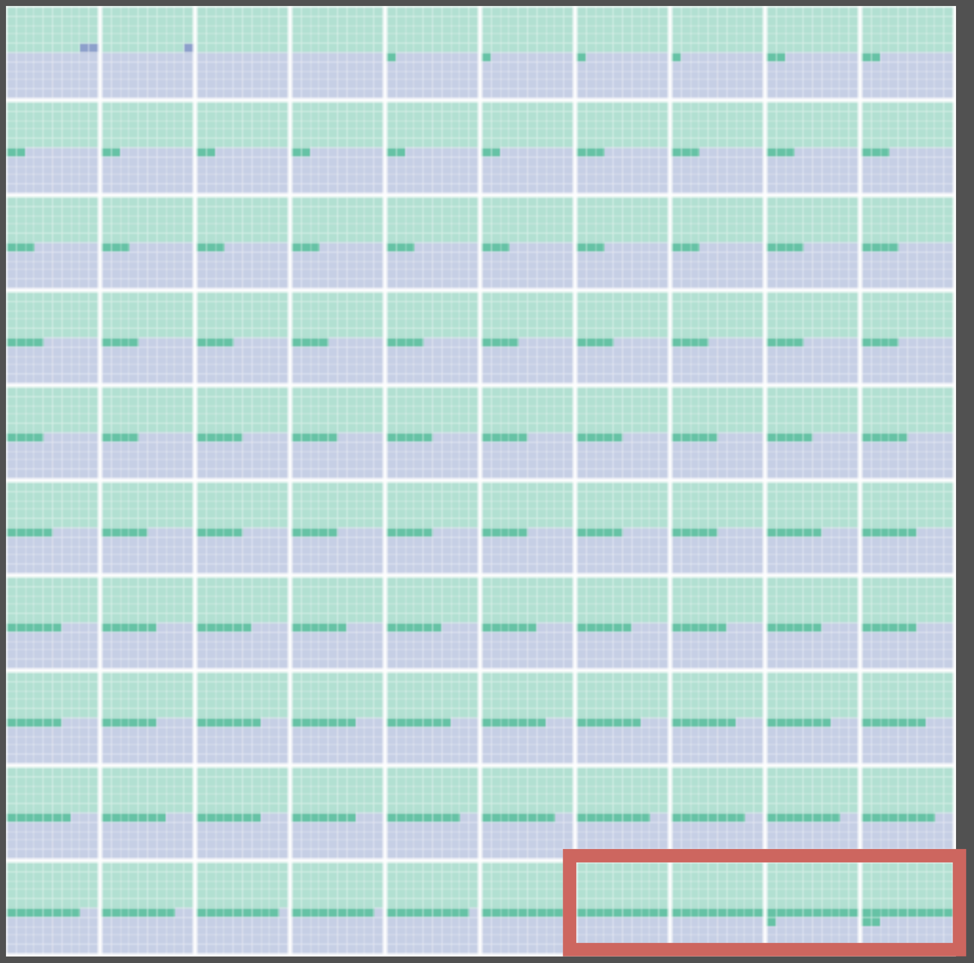
# Actual Election?

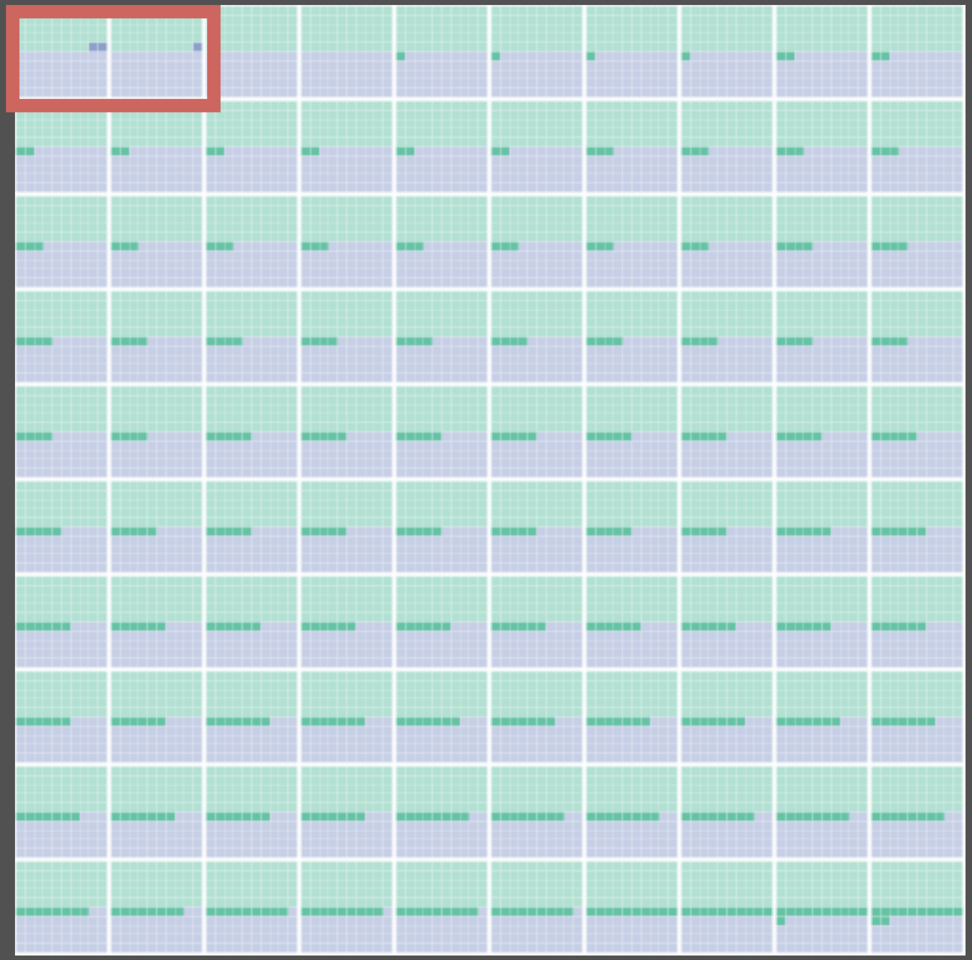


# Actual Election?



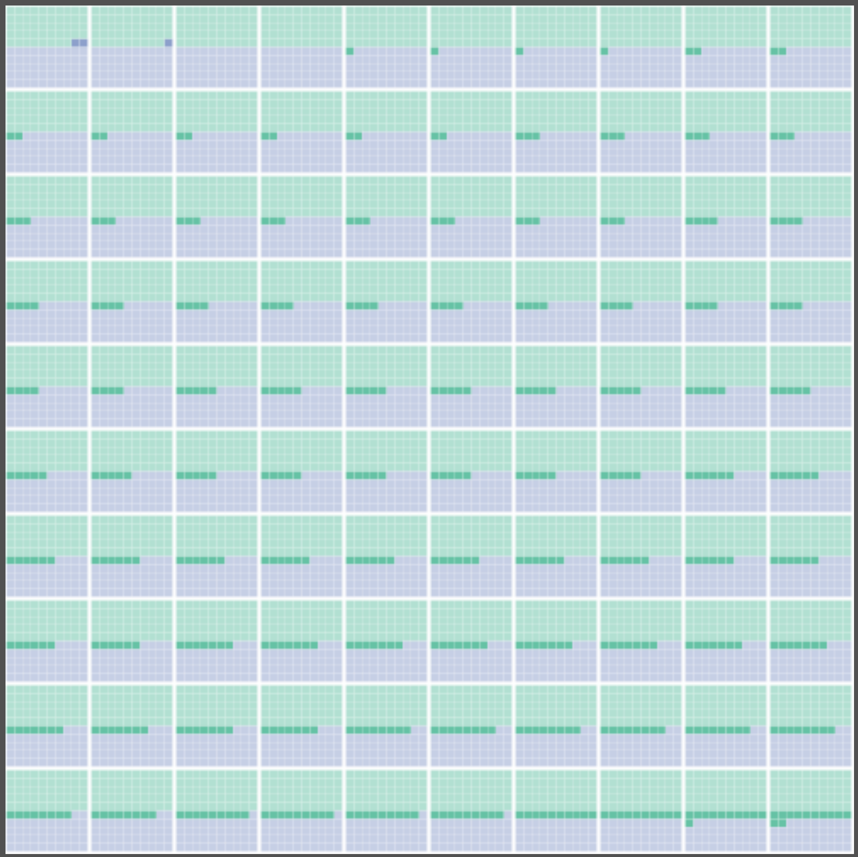






# Pangloss Plot

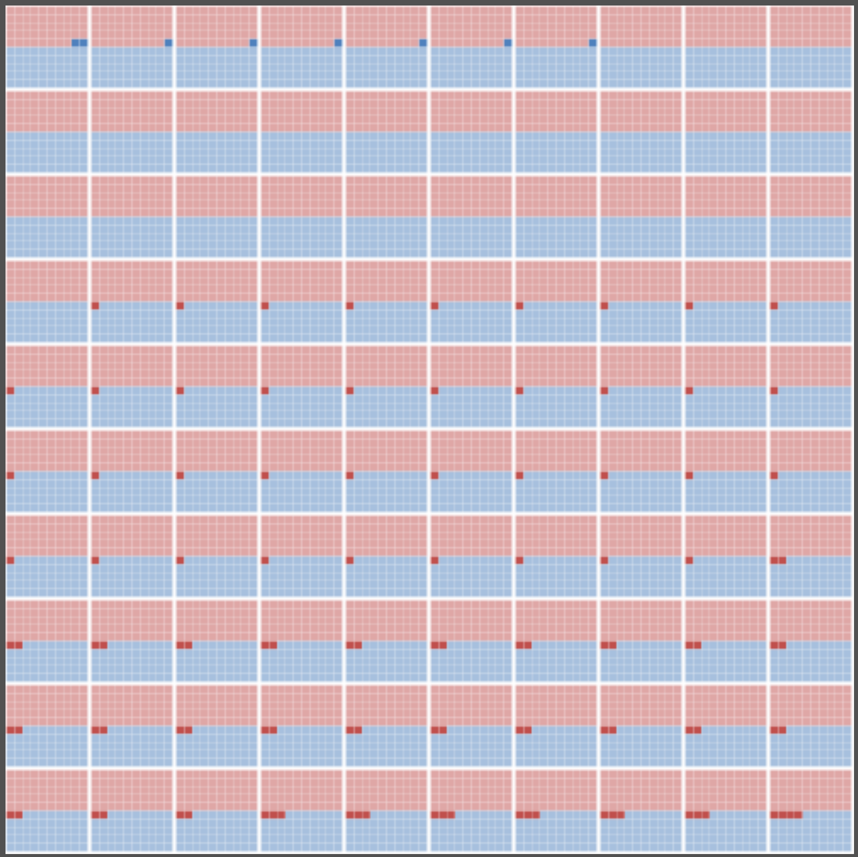
Candidate A is ahead of Candidate B in the polls, with 55% of the likely voters*



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

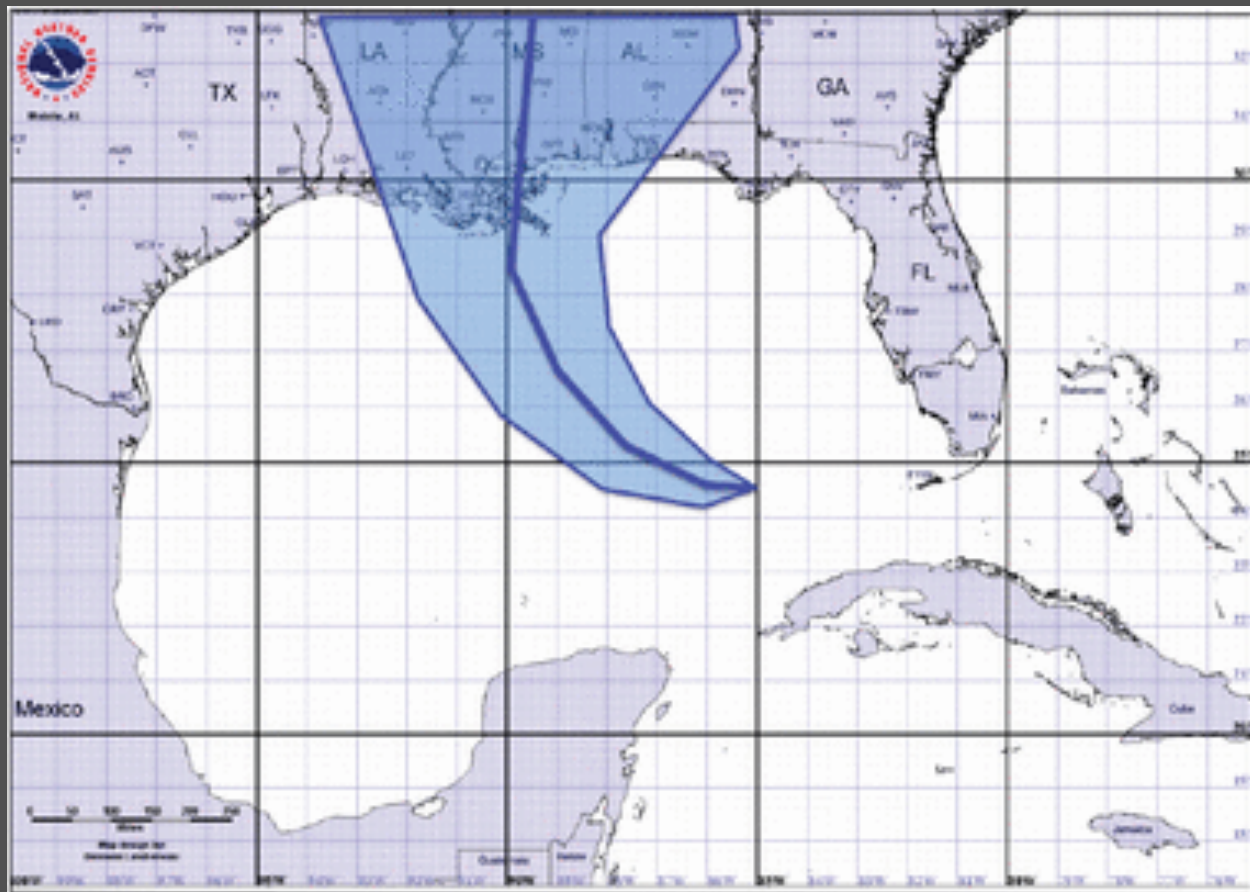
# Pangloss Plot

Romney is ahead of Obama in the polls, with 51% of the likely voters*



*poll of 3,117 people, margin of error +/-2

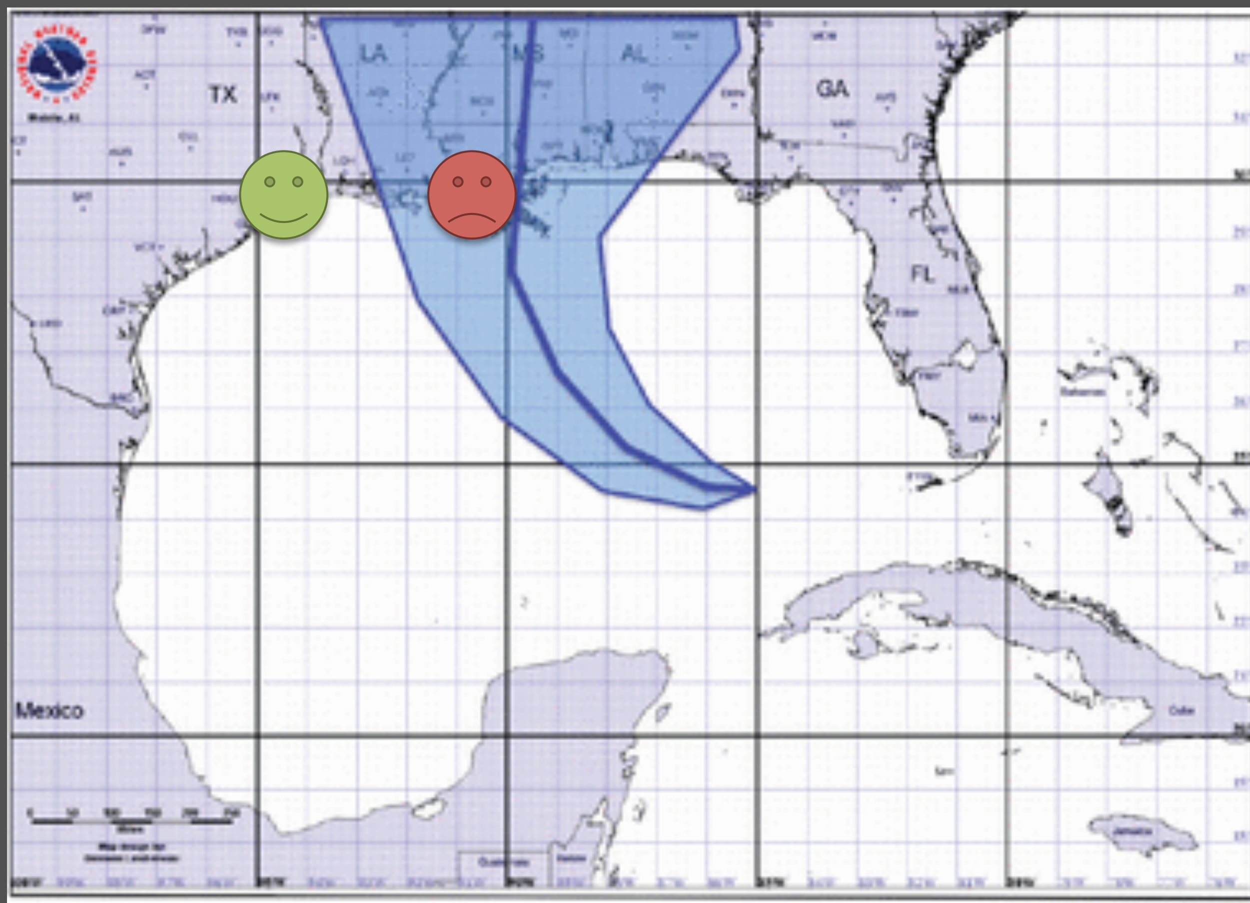
# Model Visualization




(a)

Cox, Jonathan and House, Donald and Lindell, Michael. Visuazlising uncertainty in predicted hurricane tracks. International Journal for Uncertainty Quantification, 2013.

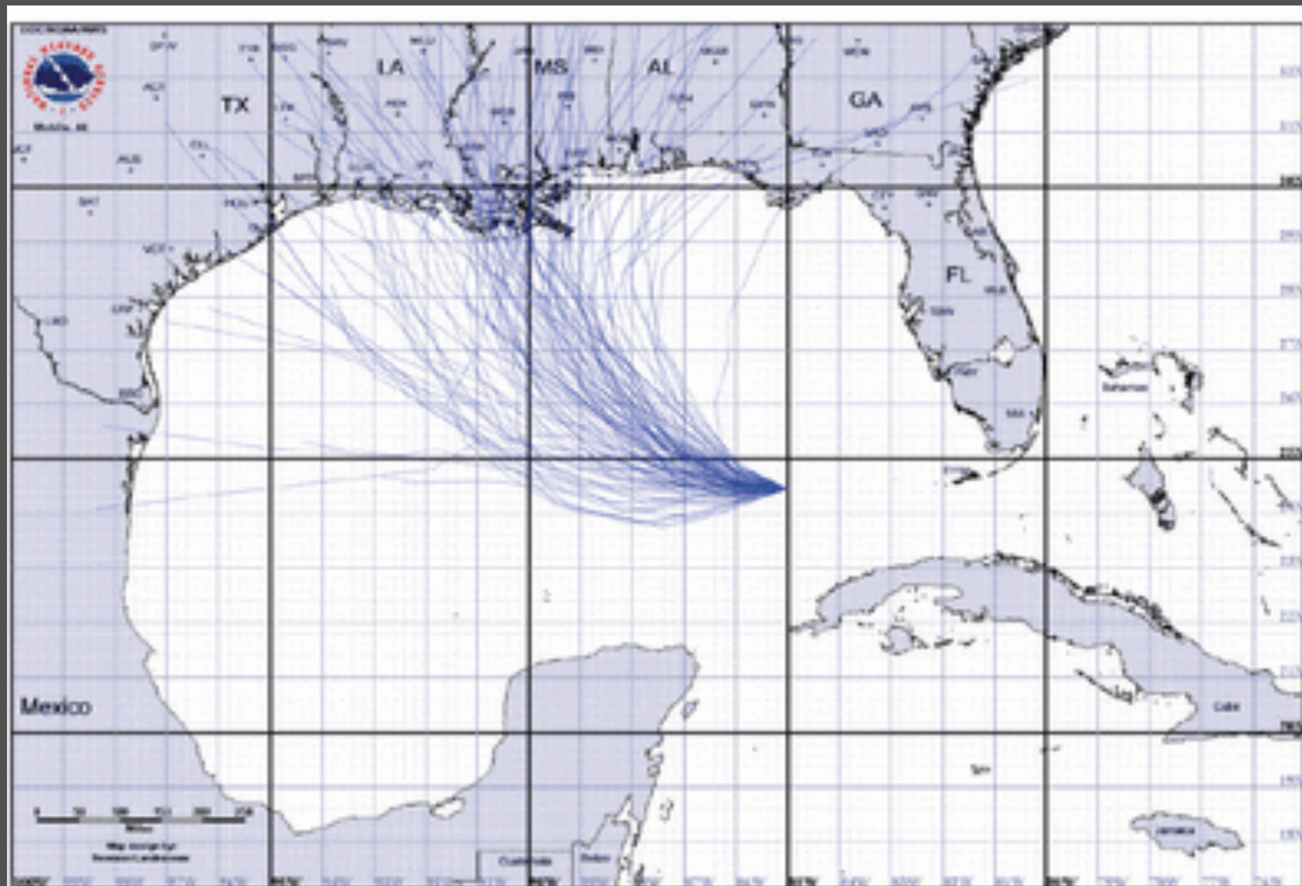
# Model Visualization




(a)

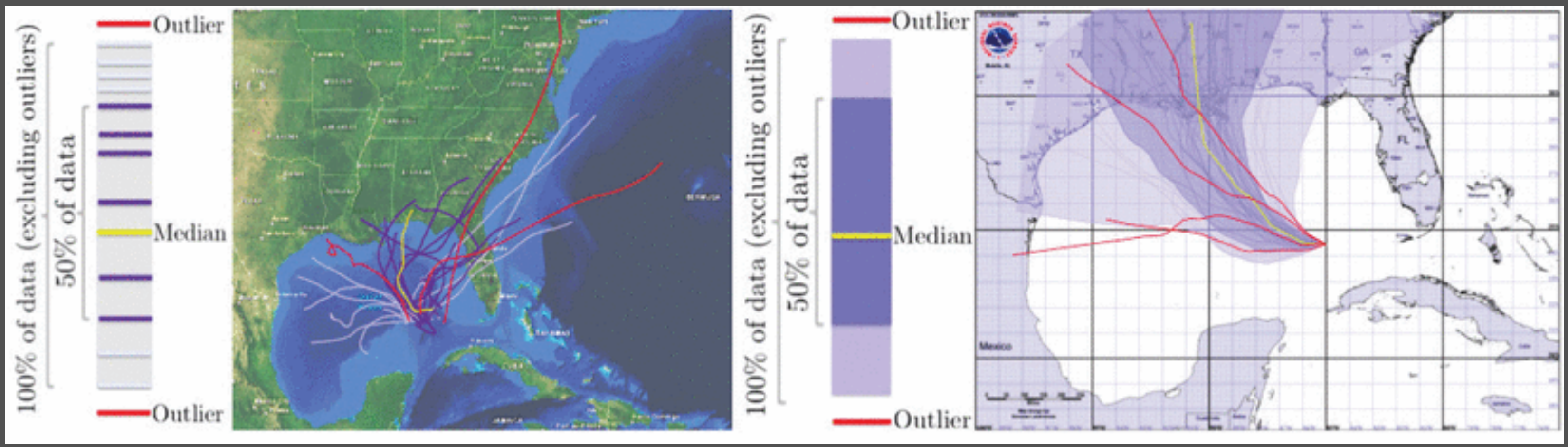
Cox, Jonathan and House, Donald and Lindell, Michael. Visuazlising uncertainty in predicted hurricane tracks. International Journal for Uncertainty Quantification, 2013.

# Model Visualization



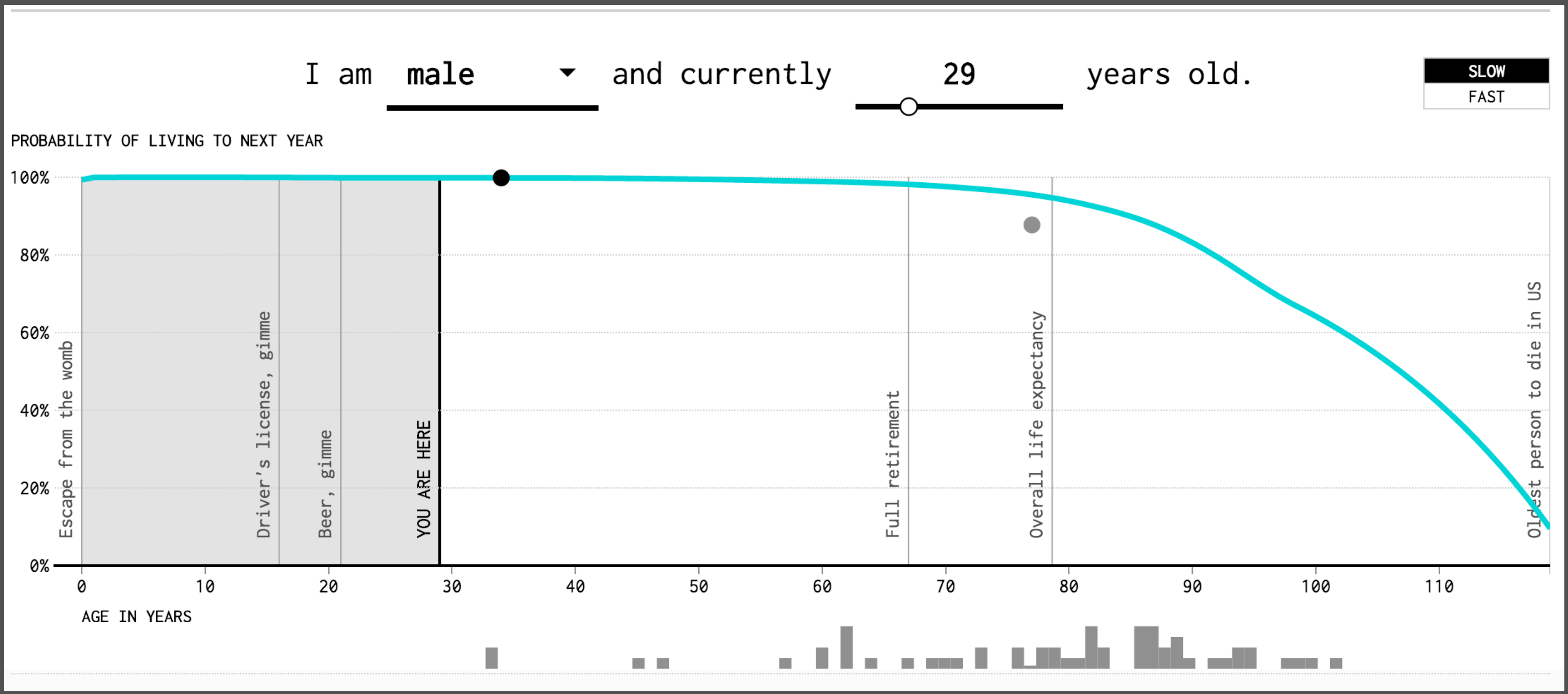

(b)

# Model Visualization

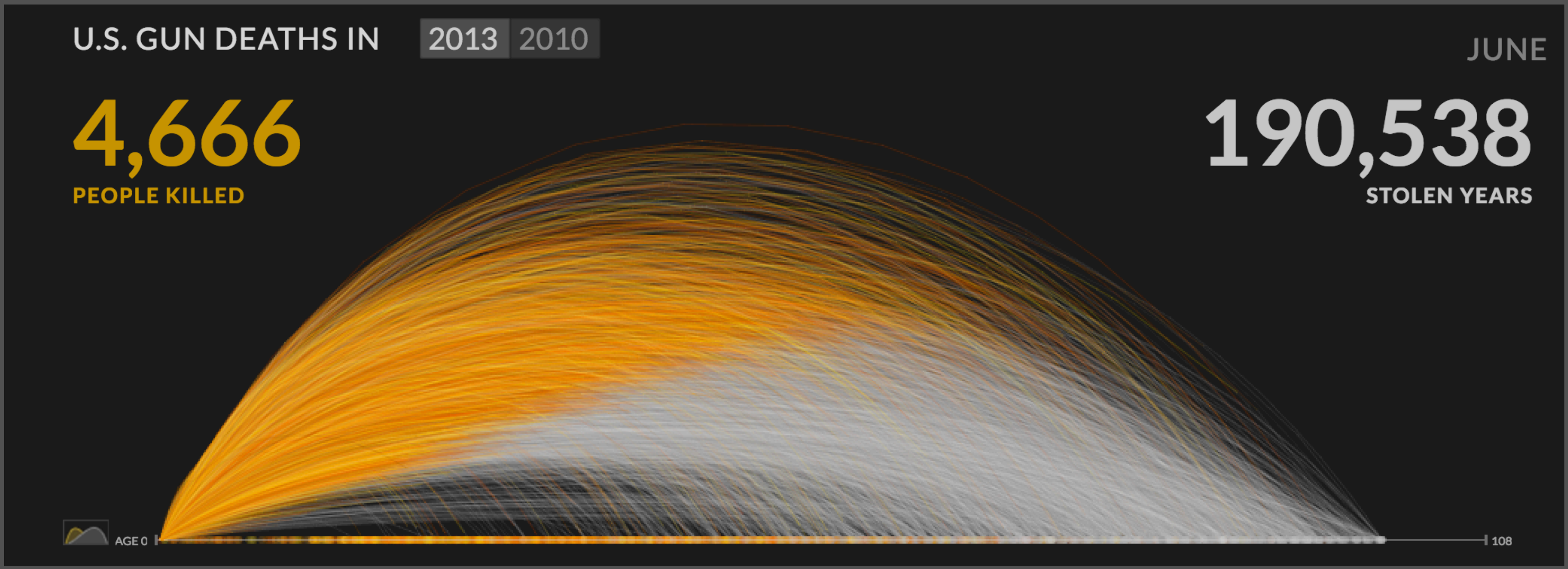



M. Mirzargar, R. Whitaker and R. Kirby. Curve Boxplot: Generalization of Boxplot for Ensembles of Curves. IEEE VIS 2014.

# Life Expectancy



# Gun Deaths



U.S. GUN DEATHS IN
2013
2010
JUNE
4,666
PEOPLE KILLED
190,538
STOLEN YEARS
AGE 0
108

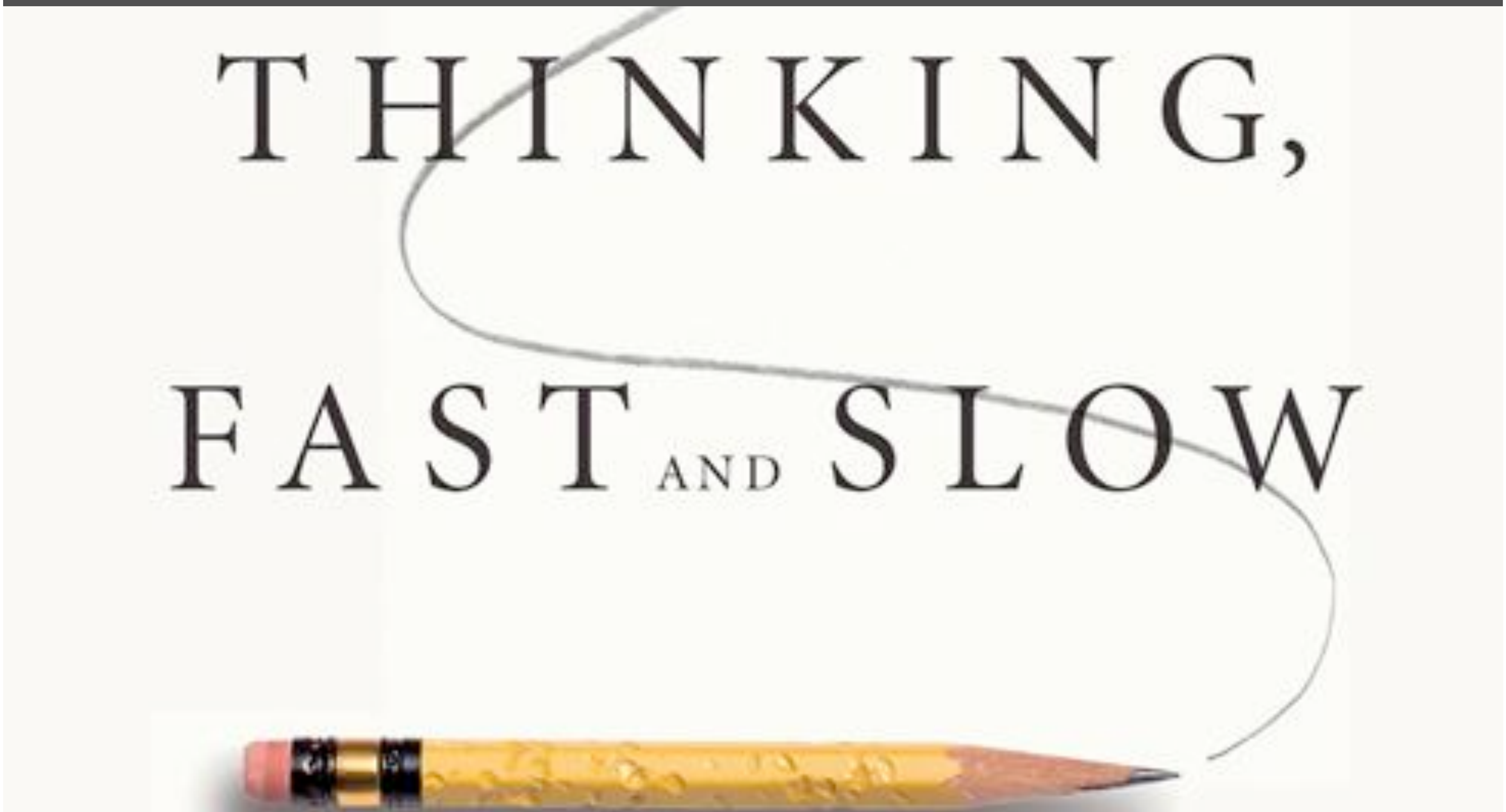
# Model Visualization

Building models is necessary to quantify uncertainty

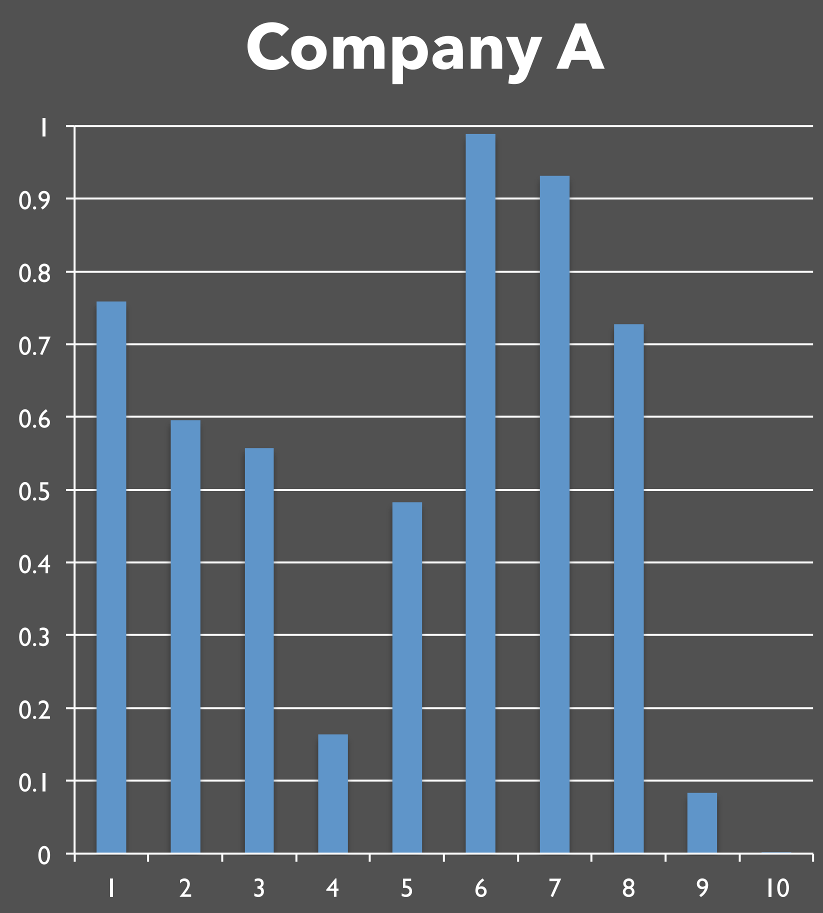
It is important to communicate the variability in model outcomes

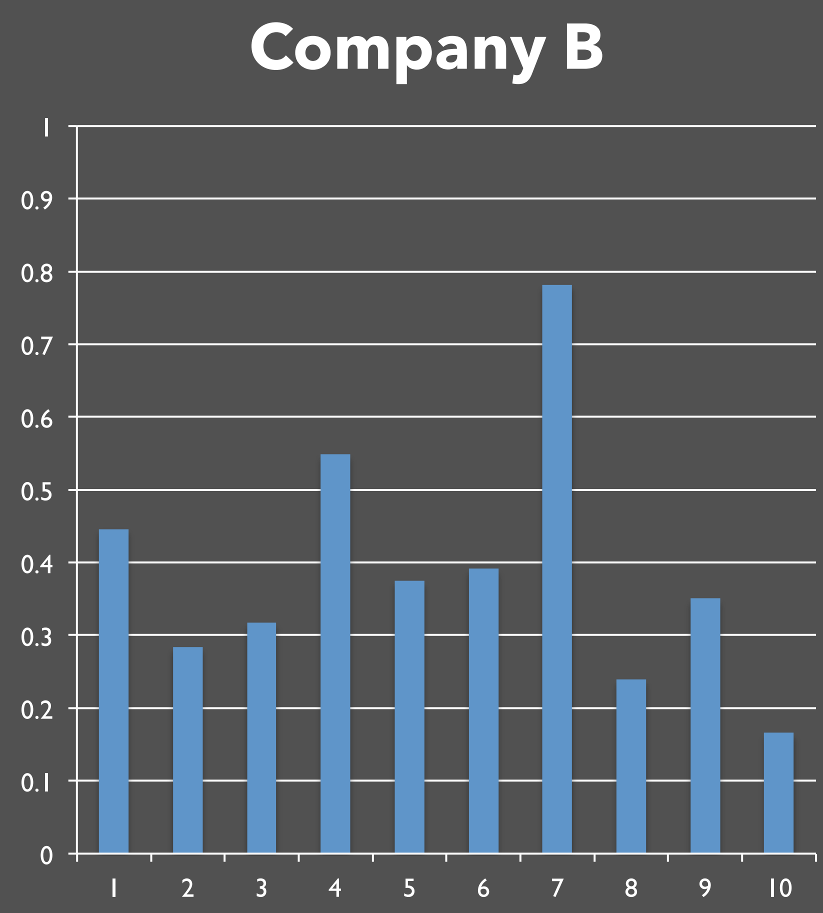
Dynamic displays can help communicate complex models

# Cognitive Biases

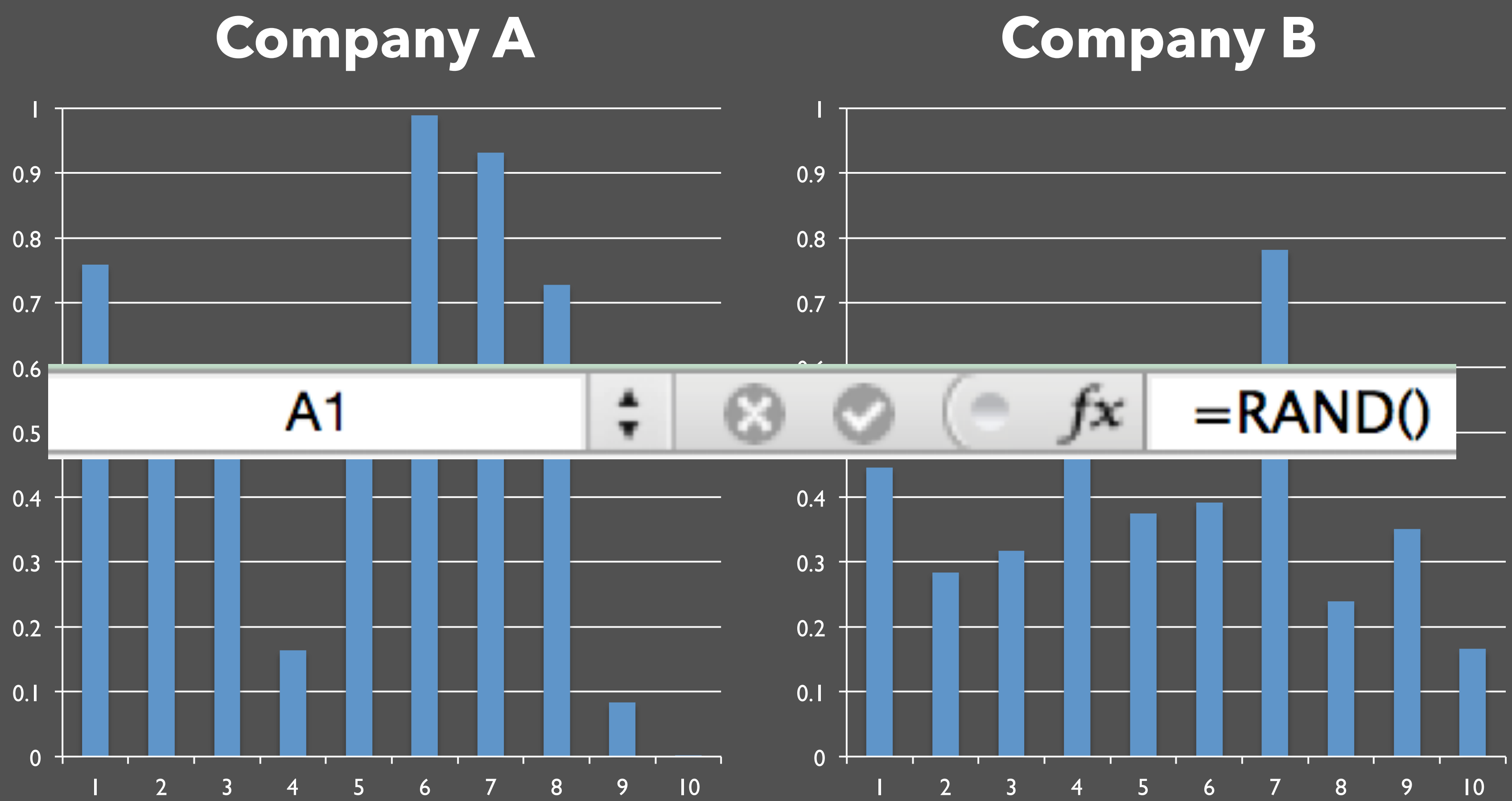


# Which Stock To Buy?



# Neither!


Company A
Company B
A1
=RAND()
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
2
3
4
5
6
7
8
9
10

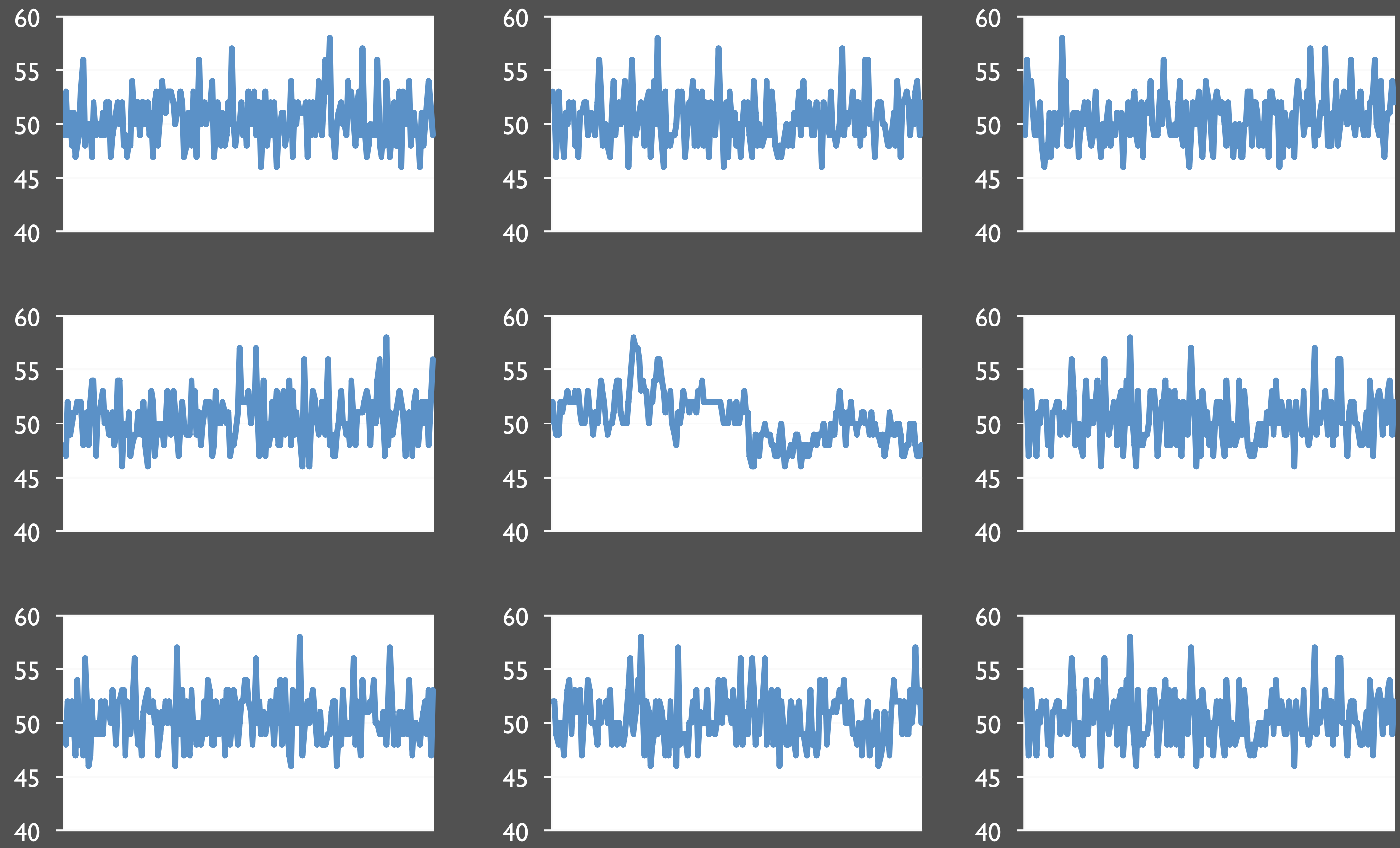
# Wu Wei

無為

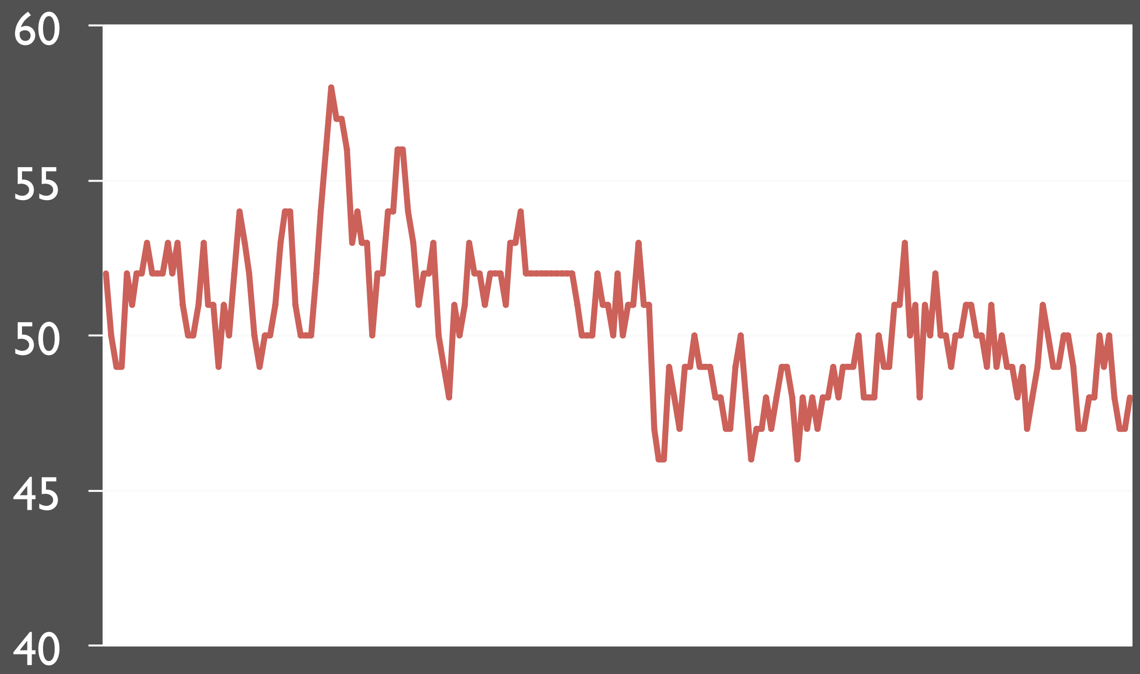
# Pareidolia

# Jobs Reports

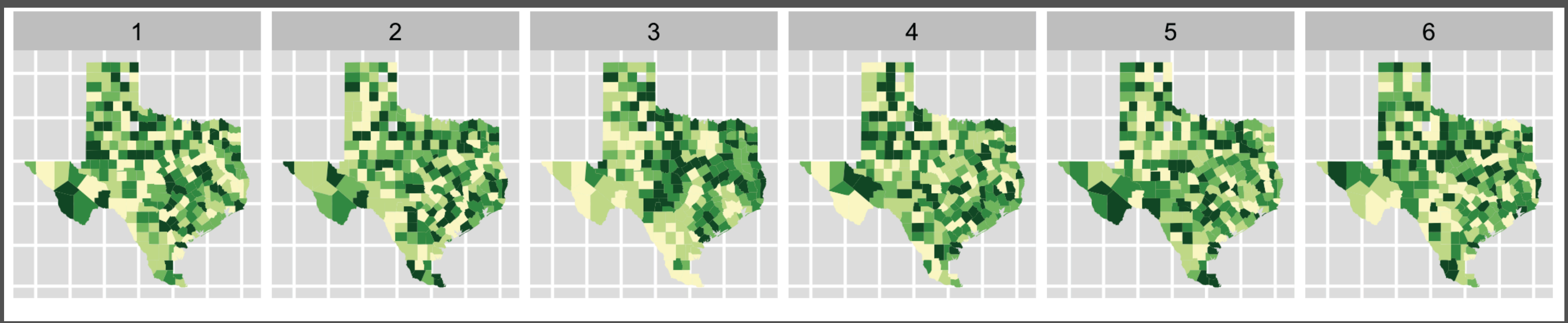
# Have People Made Up Their Mind About Obama?


60
55
50
45
40

# Visual Lineups

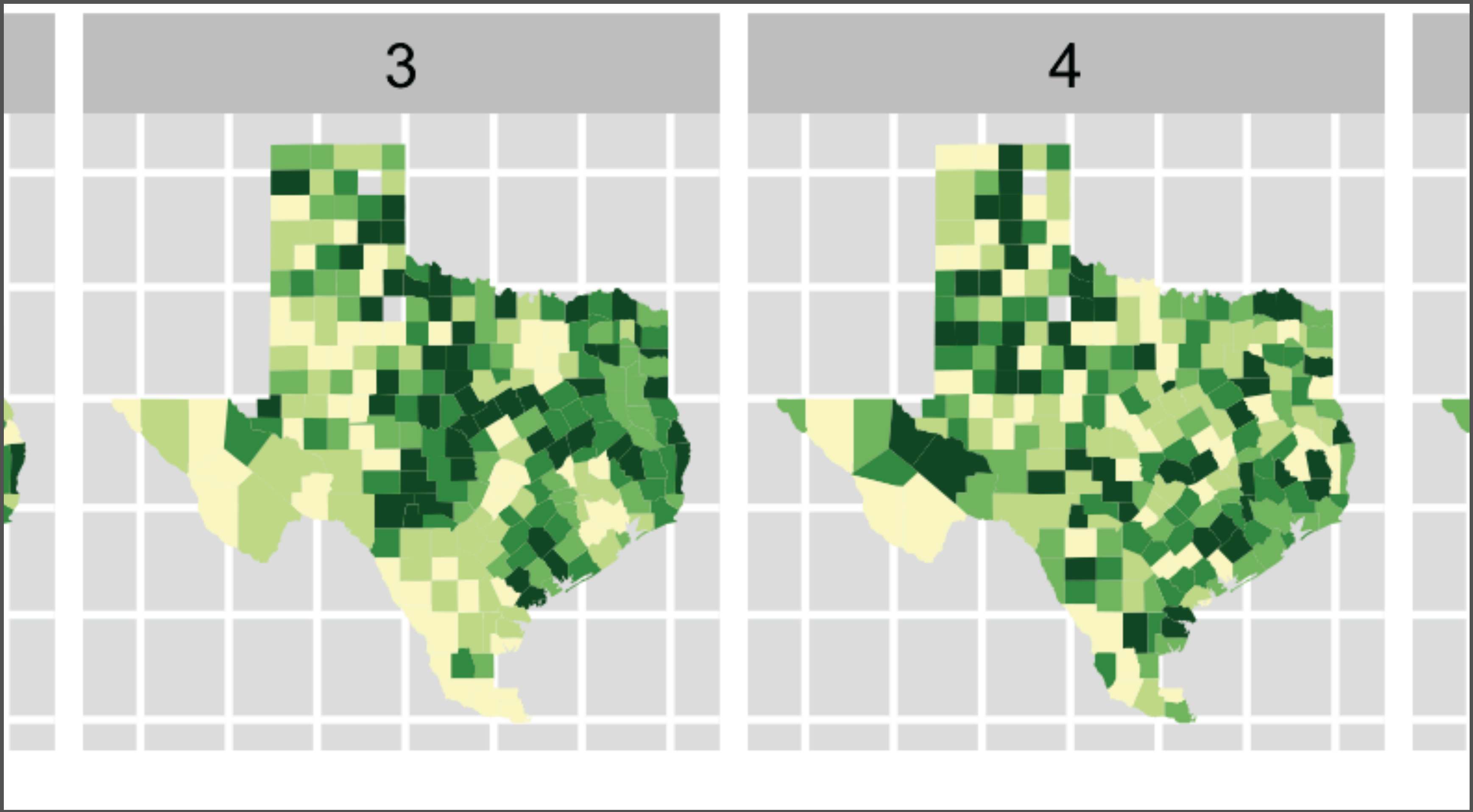


# Visual Lineups




Wickham, Hadley et al. "Graphical inference for Infovis." IEEE transactions on visualization and computer graphics 16.6 (2010): 973–9.

# Visual Lineups



# 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.

# Base Rate Fallacy

1% of 40 year old women have breast cancer

The probability a mammogram will detect breast cancer is 80%

The probability of a false positive is 10%.

If a 40 year old woman gets a positive result, what is the probability she has breast cancer?

# Bayes' Law

$P(A|B) = P(B|A)P(A) / P(B)$

# Bayes' Law

$P(A|B) = P(B|A)P(A) / P(B)$

$P(\text{Cancer} \mid +\text{Test}) = P(+\text{Test}|\text{Cancer})P(\text{Cancer})/P(+\text{Test})$

# Bayes' Law

$P(A|B) = P(B|A)P(A) / P(B)$

$P(Cancer \mid +Test) = P(+Test|Cancer)P(Cancer)/P(+Test)$

$P(+) = P(+ \wedge C)P(C) + P(+ \wedge \sim C)P(\sim C)$

# Bayes' Law

$P(A|B) = P(B|A)P(A) / P(B)$

$P(\text{Cancer} \mid +\text{Test}) = P(+\text{Test}|\text{Cancer})P(\text{Cancer})/P(+\text{Test})$

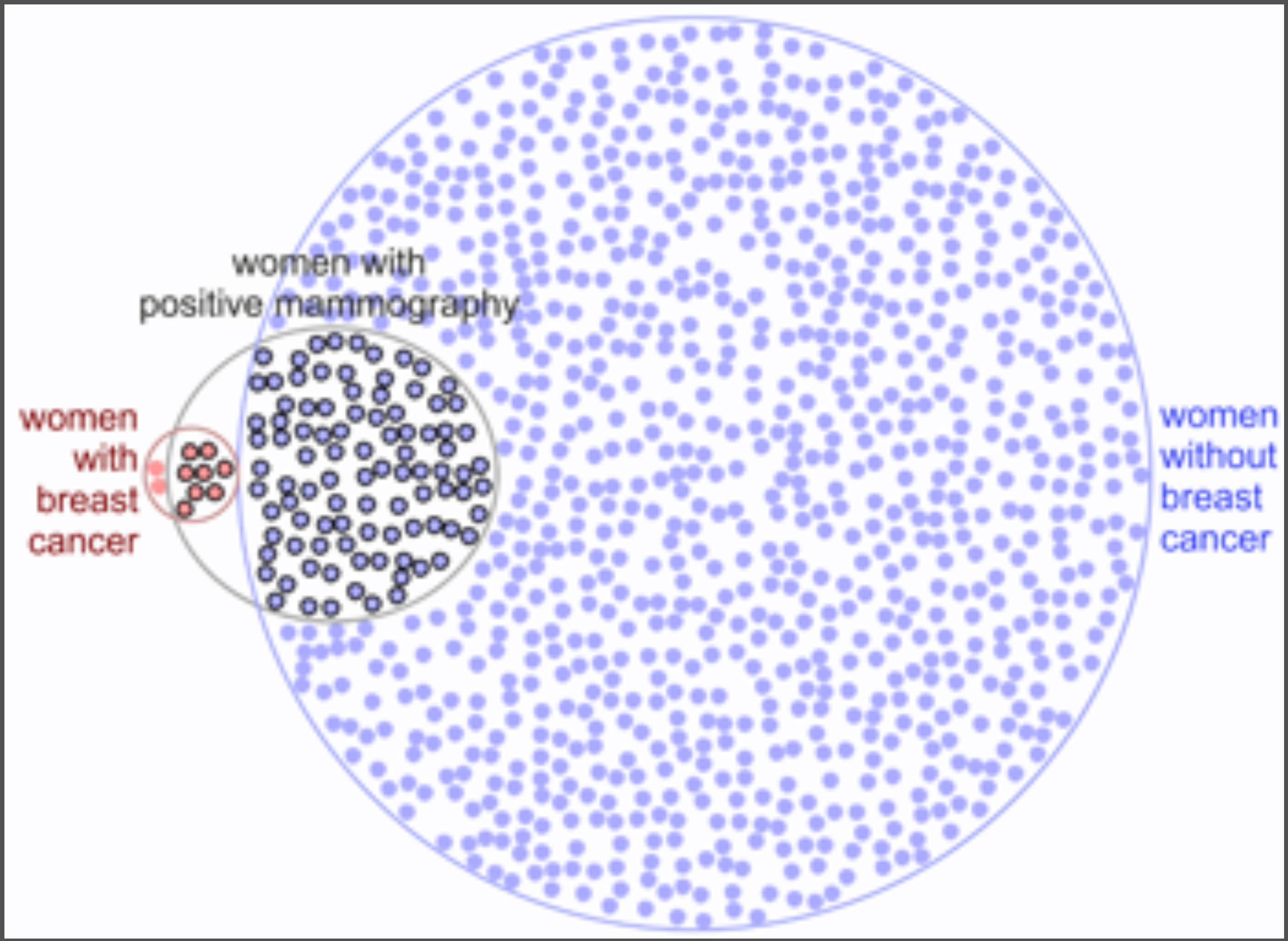
$P(+) = P(+ \wedge C)P(C) + P(+\wedge \sim C)P(\sim C)$

$P(+) = 0.01*0.8 + 0.99*0.1$

$P(+) = 0.107$

$P(C \mid +) = 0.8 * 0.01 / 0.107 \approx$ **0.075**

# Base Rate Fallacy



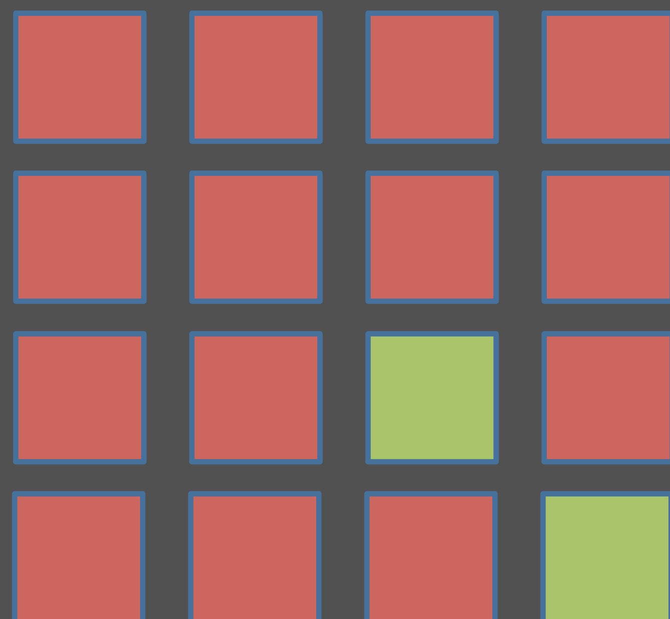

Micallef, Luana, Pierre Dragicevic, and Jean-Daniel Fekete. "Assessing the Effect of Visualizations on Bayesian Reasoning Through Crowdsourcing." Visualization and … October (2012).

# Risk

"1 out of every 8 people with small cell lung cancer survive for at least 5 years"

I. Lipkus and J. Hollands. The Visual Communication of Risk. Journal of the National Cancer Institute, 1999.

# Risk

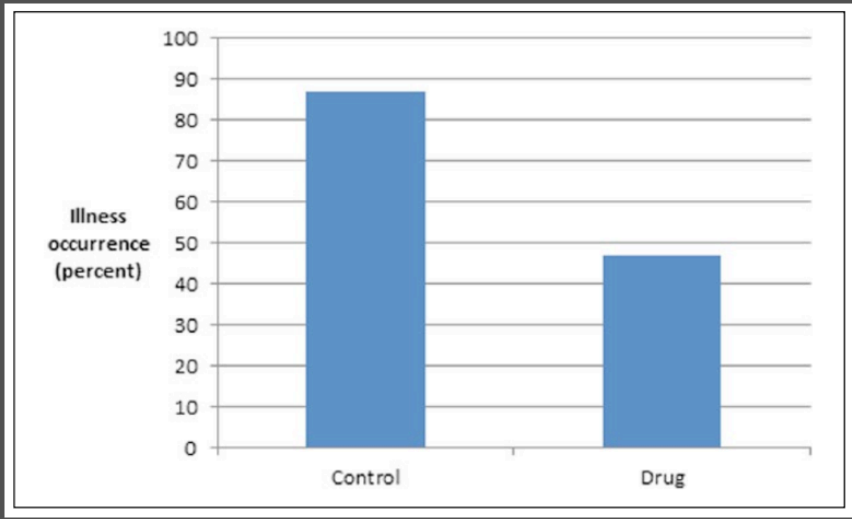


# Risk

“A large pharmaceutical company has recently developed a new drug to boost peoples’ immune function. It reports that trials it conducted demonstrated a drop of forty percent (from eighty seven to forty seven percent) in occurrence of the common cold. It intends to market the new drug as soon as next winter, following FDA approval.”

# Persuaded by Nothing

"A large pharmaceutical company has recently developed a new drug to boost peoples' immune function. It reports that trials it conducted demonstrated a drop of forty percent (from eighty seven to forty seven percent) in occurrence of the common cold. It intends to market the new drug as soon as next winter, following FDA approval."




Tal, Aner and Wansink, Brian. Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy. Public Understanding of Science, 2016.

# Cognitive Biases

Humans can be quite poor at reasoning about uncertain values.

Minor changes in visual design can influence decision-making for better or worse.

# Conclusion

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.