CSE 442 – Data Visualization

Uncertainty

Michael Correll Tableau Research

Questions To Answer

What Does Uncertainty Mean?

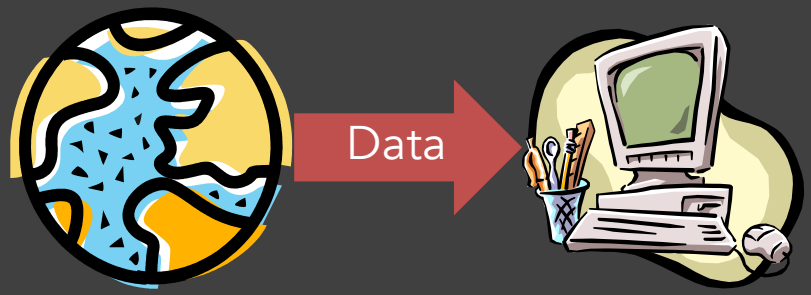
How Should I Visualize It?

What Can Go Wrong?

Definitions and Bookkeeping

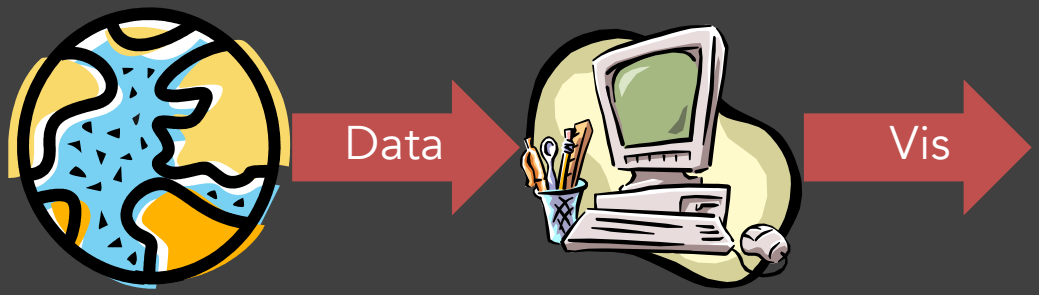
WHAT DOES UNCERTAINTY MEAN, ANYWAY?

The Visualization Pipeline



Collection/Curation

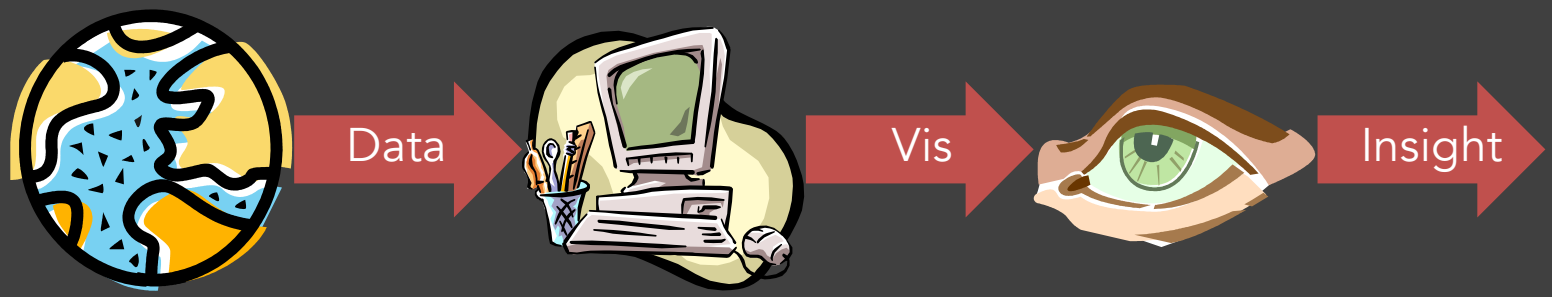
The Visualization Pipeline



Collection/Curation

Design

The Visualization Pipeline

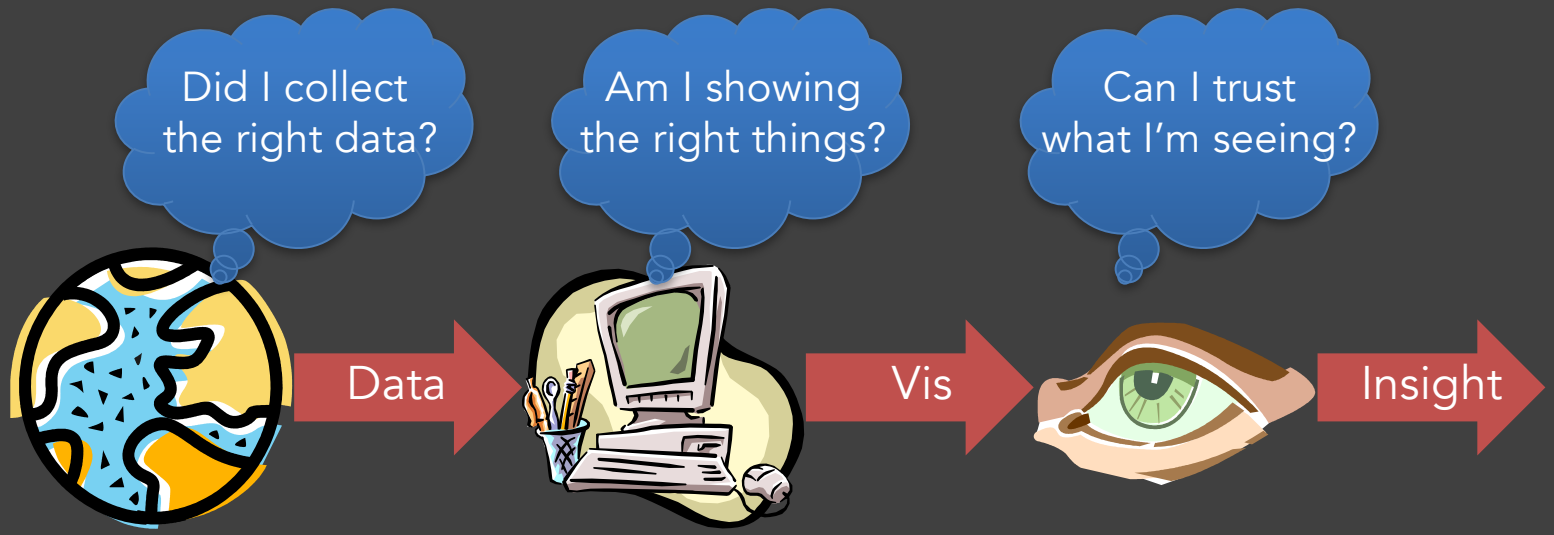


Collection/Curation

Design

Analysis

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

Aleatory Uncertainty

Epistemic Uncertainty

Type I error

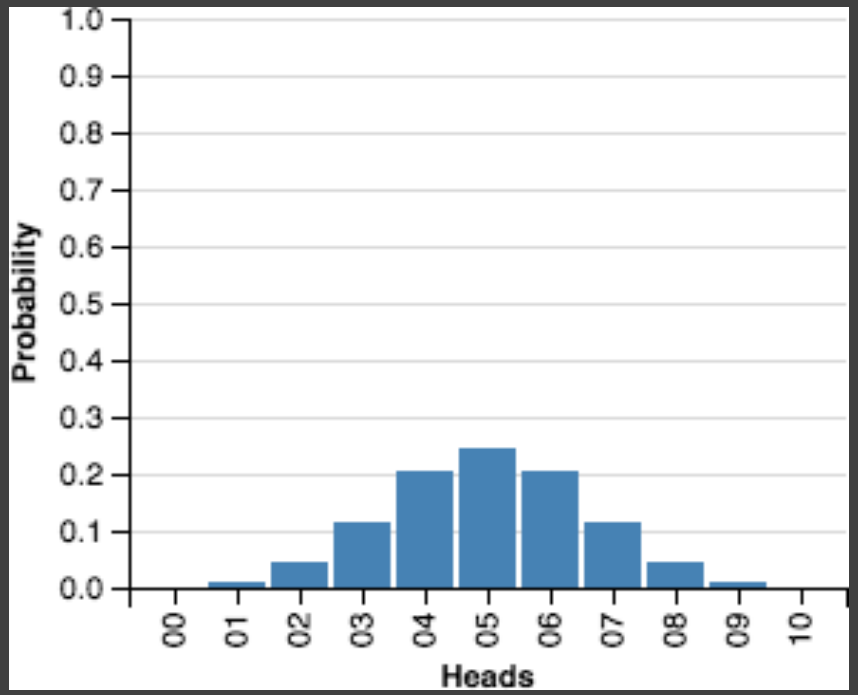
Type II error

Precision

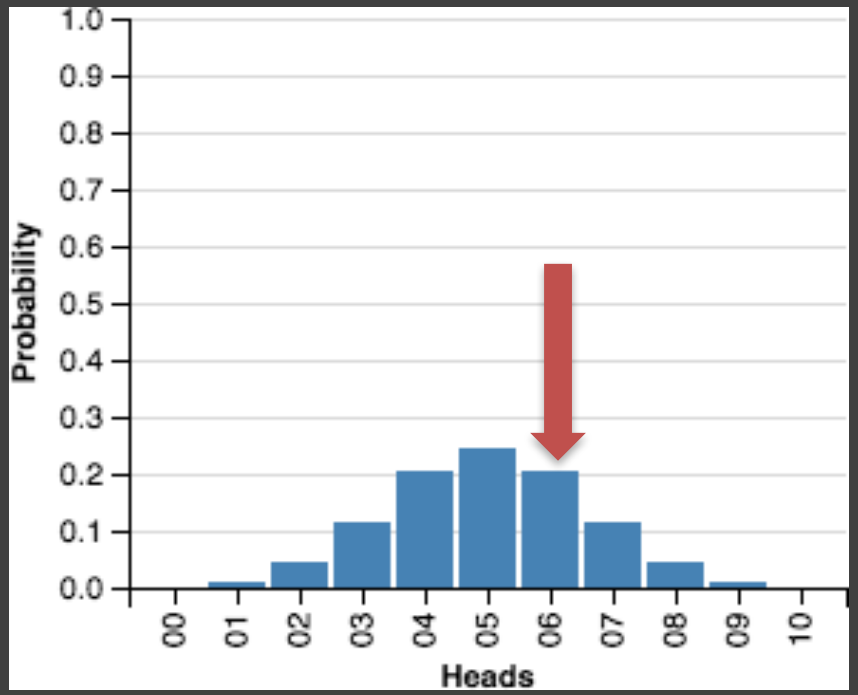
Bias

Aleatory Uncertainty

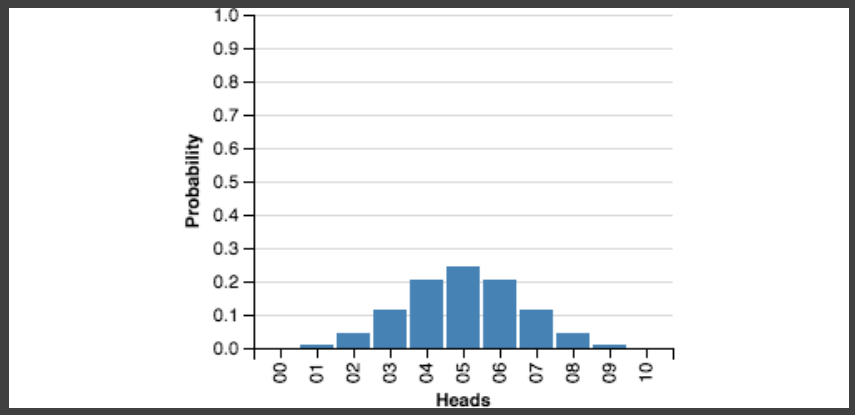
Aleatory Uncertainty

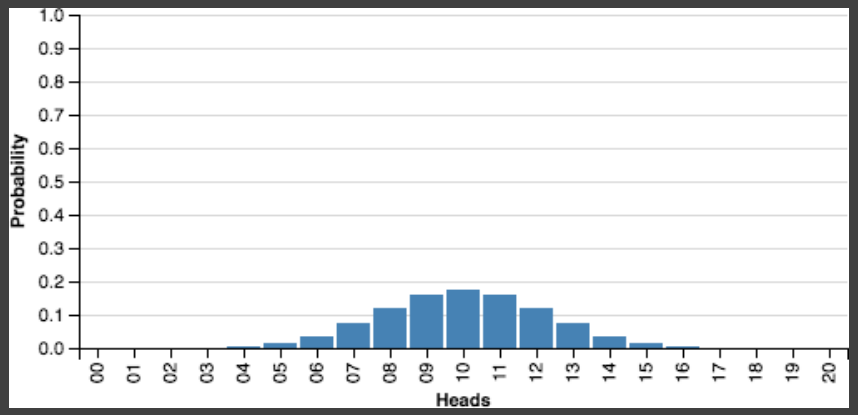


Aleatory Uncertainty



Aleatory Uncertainty





John Edmund Kerrich

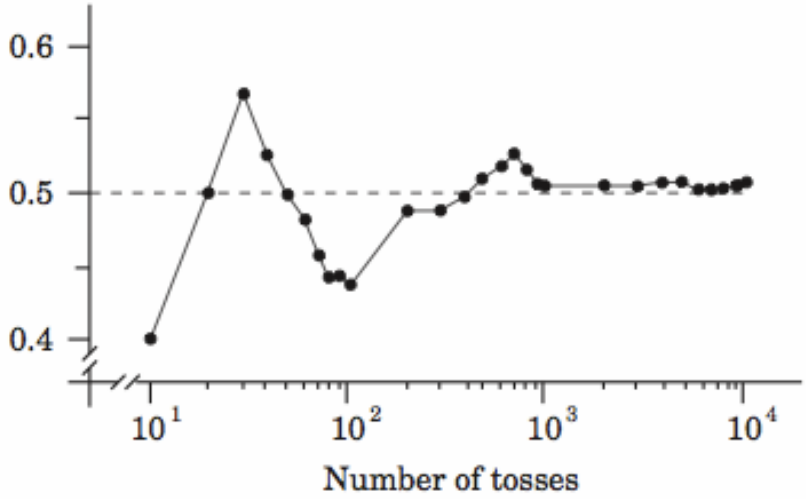
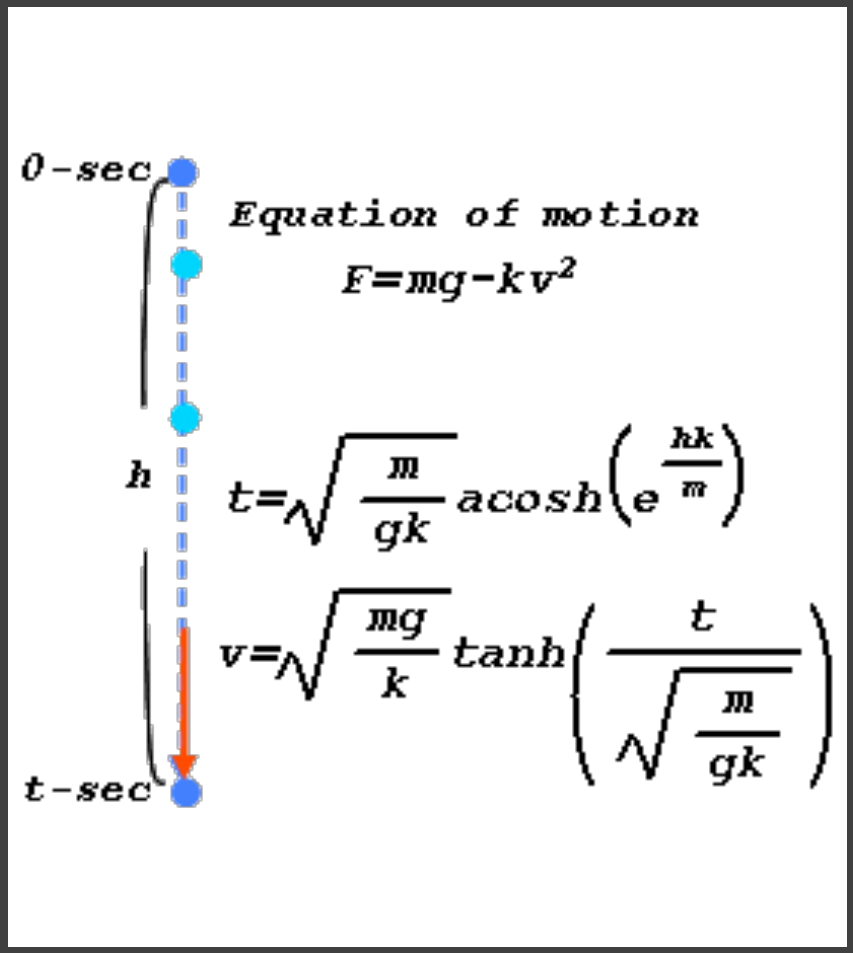


FIGURE 4.1.1 Proportion of heads versus number of tosses for John Kerrich's coin-tossing experiment.

Epistemic Uncertainty



Uncertainty Types

Aleatory

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

Epistemic

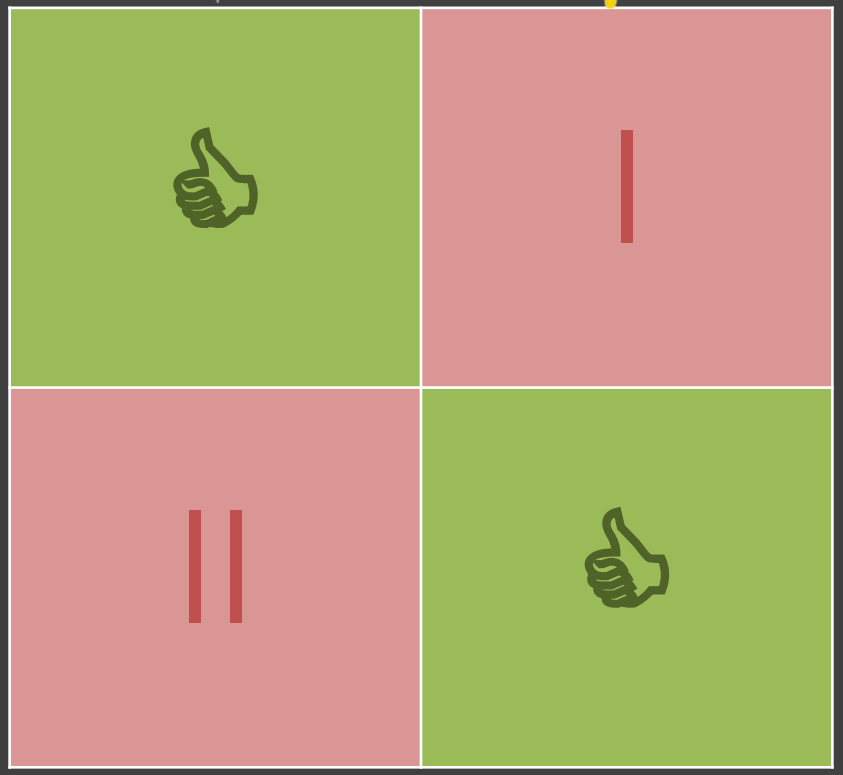
Things we could in principle know for certain, but have not measured.

Should I Bring an Umbrella?

Decision Uncertainty

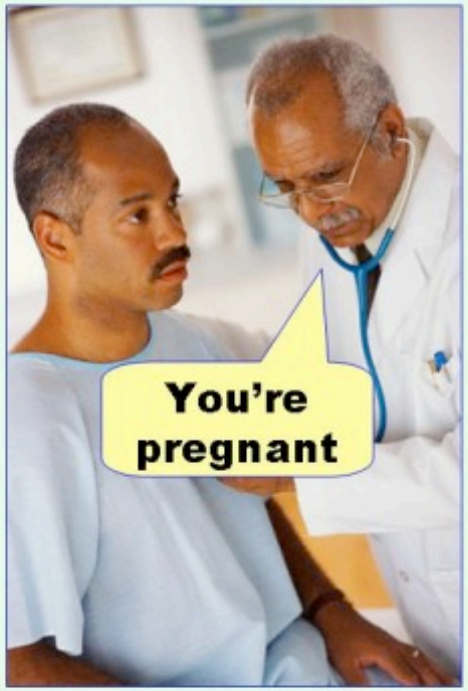
"50% Chance of Rain"

Risk and Error



Type I and Type II

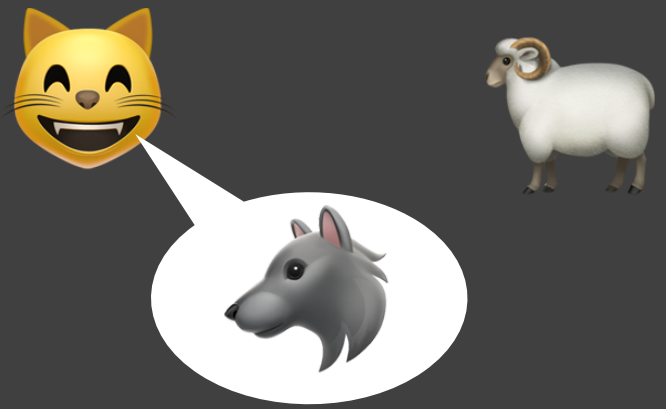
Type I error
(false positive)



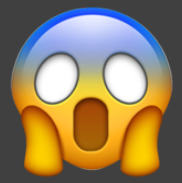
Type II error
(false negative)

The Boy Who Cried Wolf

Type I



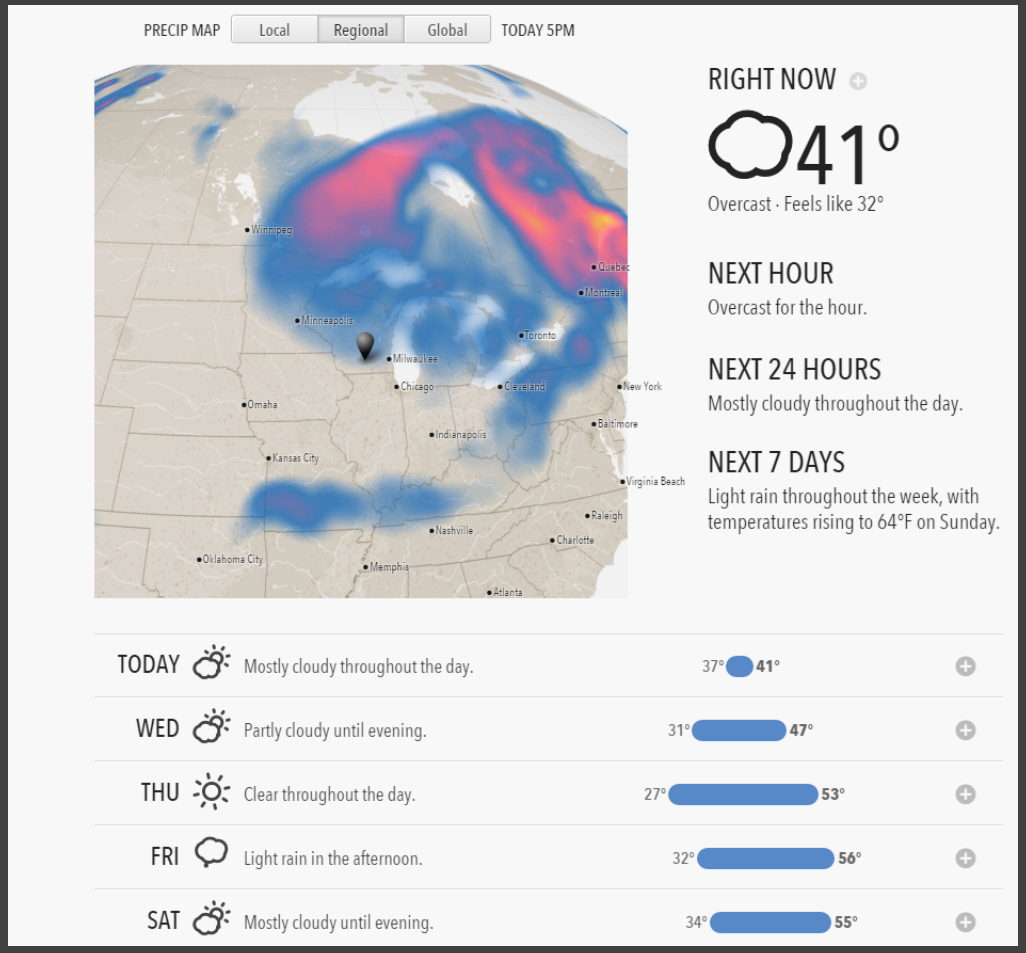
Type II



Model Uncertainty

"50% Chance of Rain"

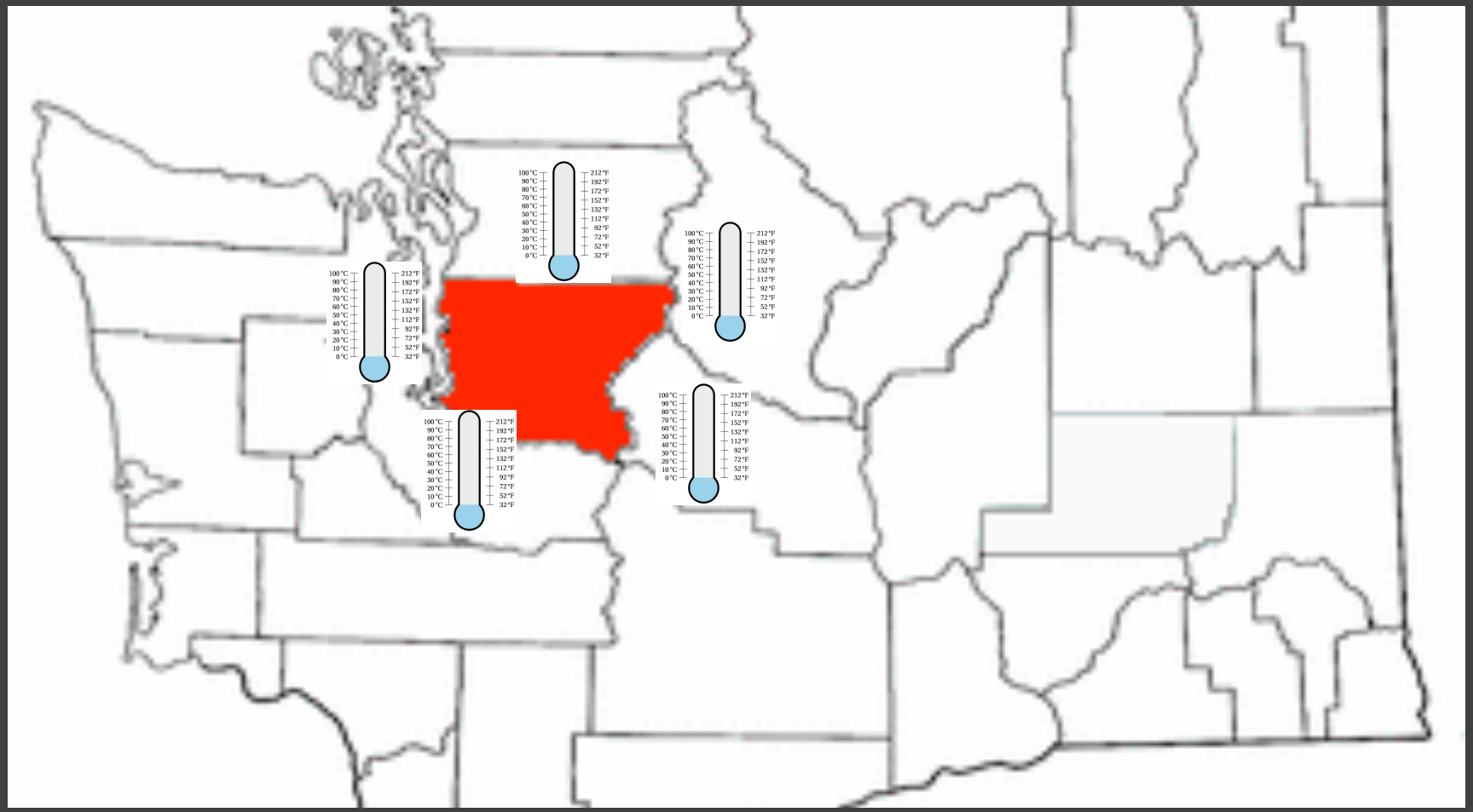
Model Uncertainty



Model Uncertainty

Measurement Uncertainty

Measurement Uncertainty



Measurement Uncertainty

Accuracy

Measurement Uncertainty

Accuracy

Measurement Uncertainty

Accuracy

Precision

Measurement Uncertainty

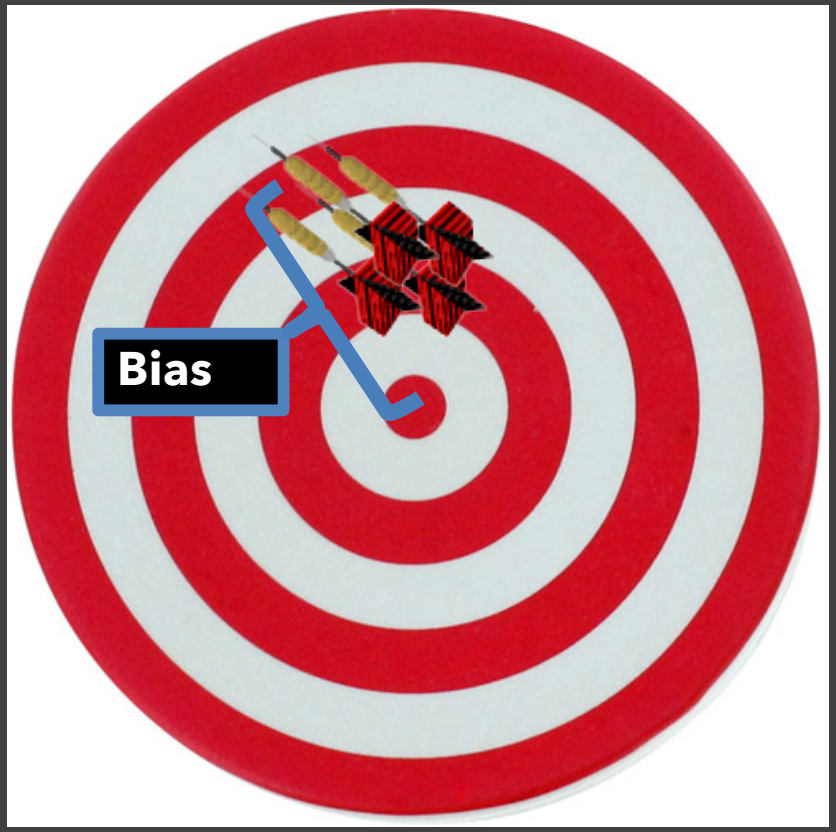
Accuracy

Precision

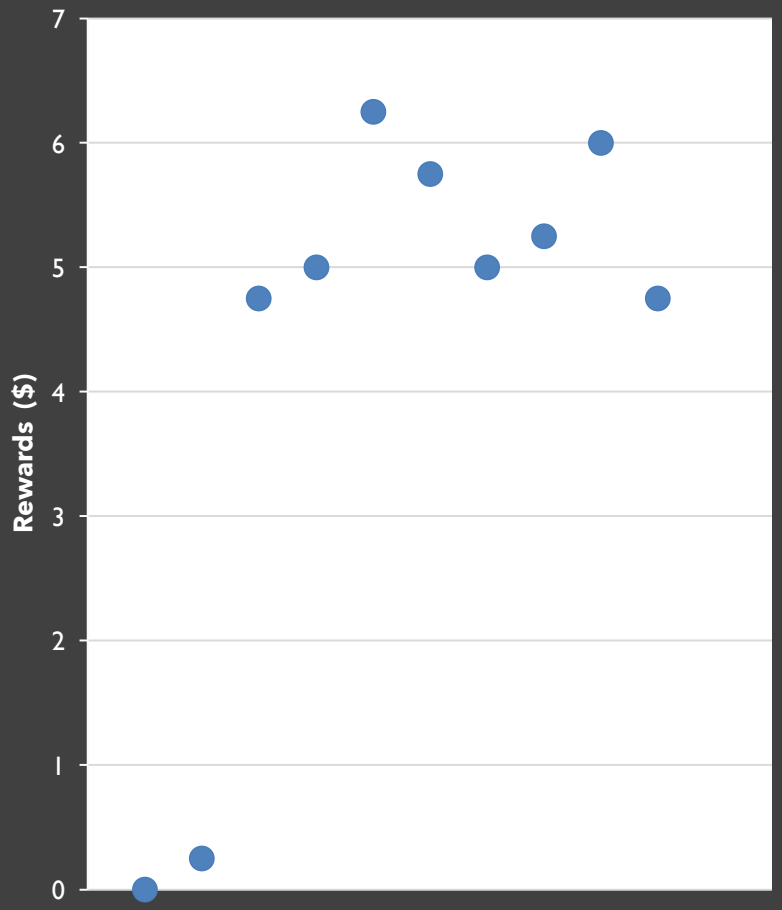
Measurement Uncertainty

Accuracy

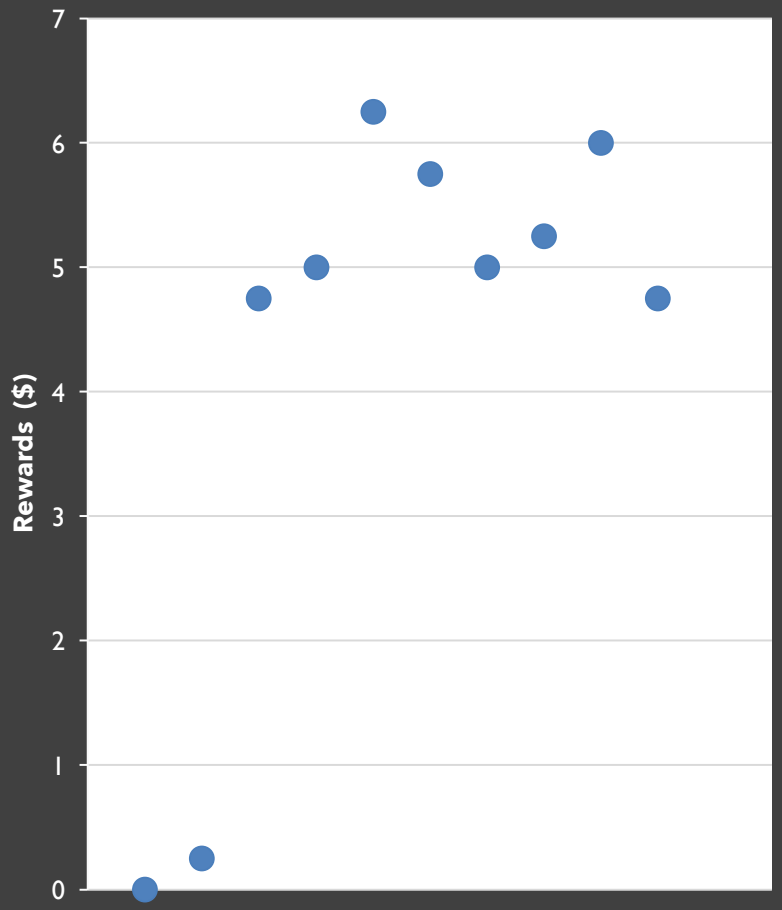
Precision



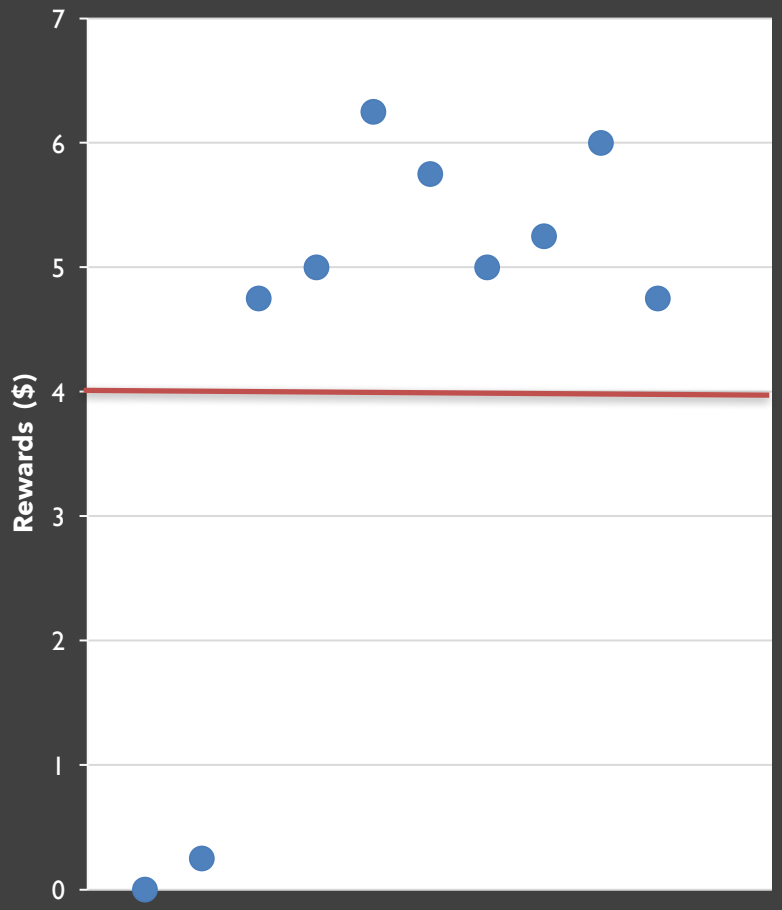
Should you take this $4 bet?



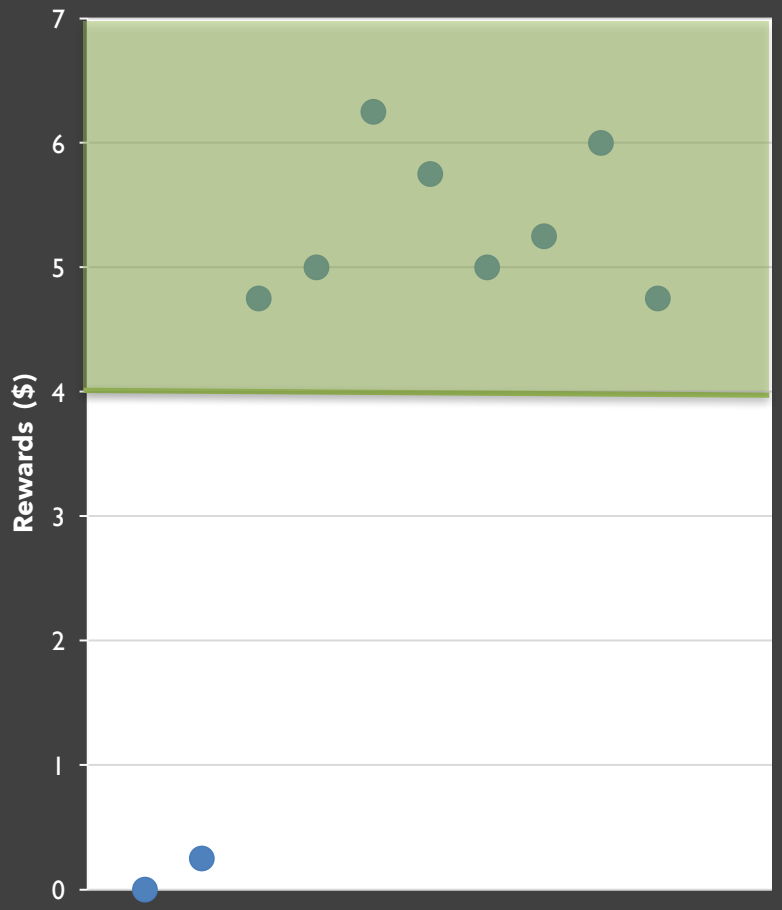
Samples



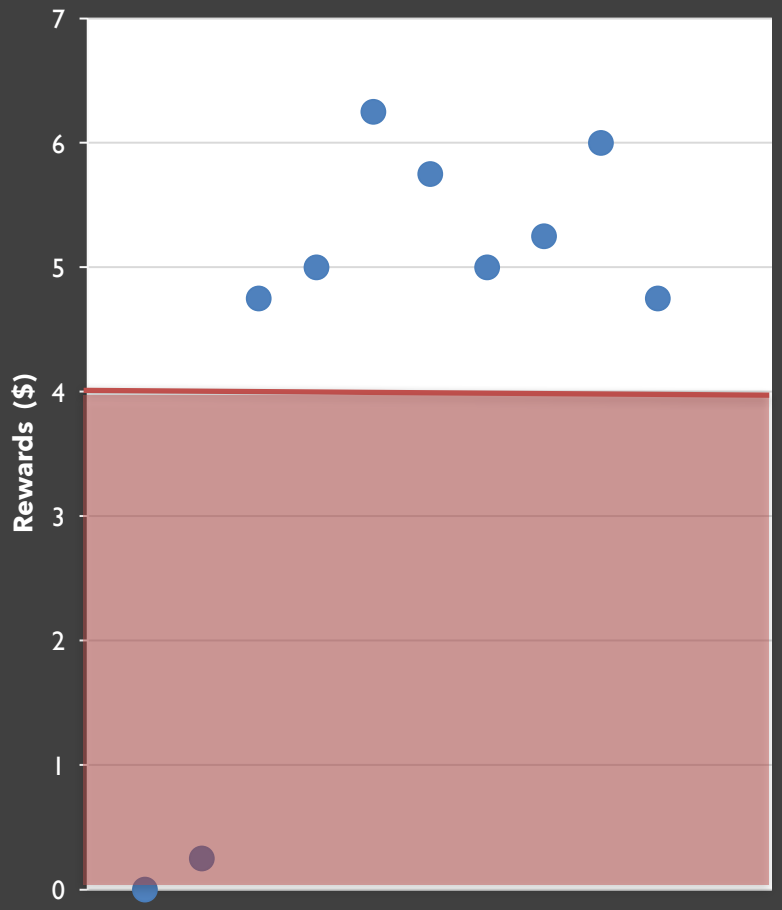
Should you take this $4 bet?



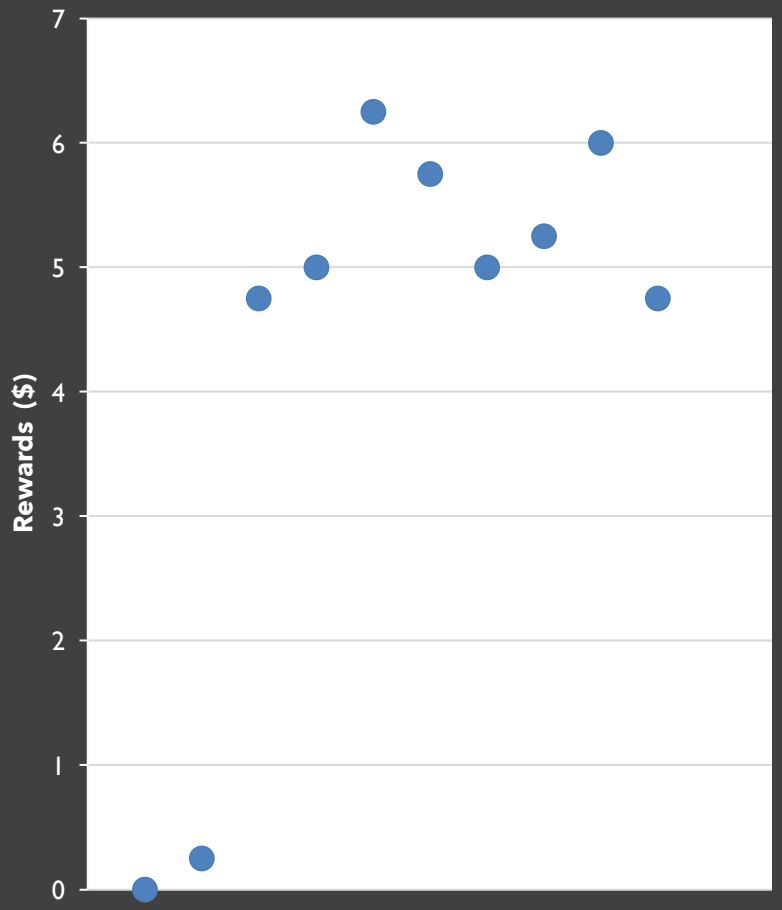
Should you take this $4 bet?

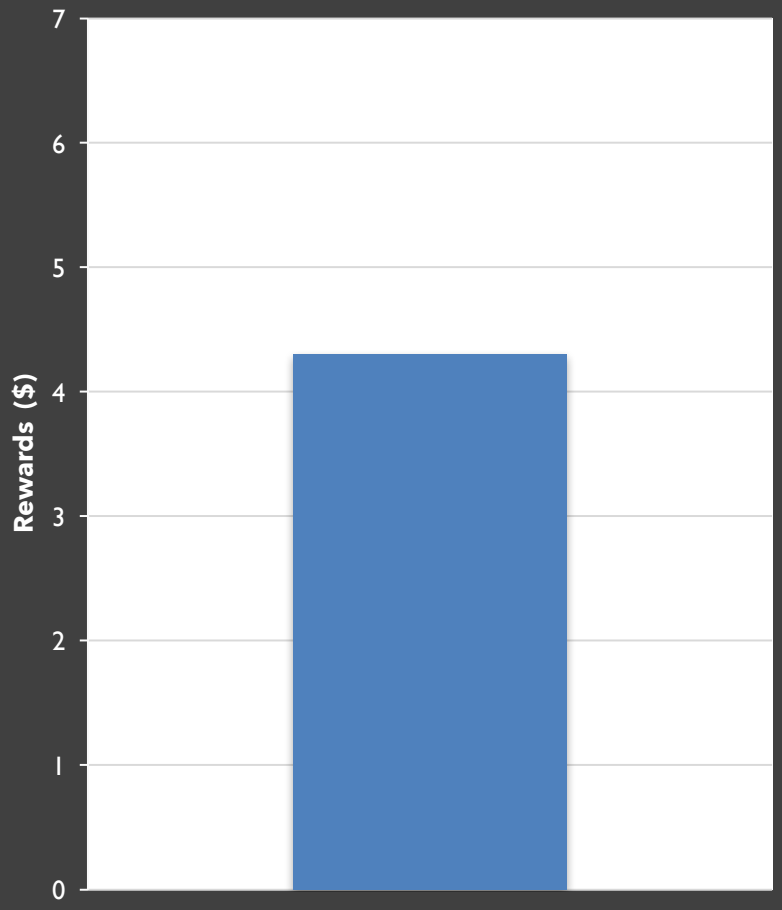


Should you take this $4 bet?

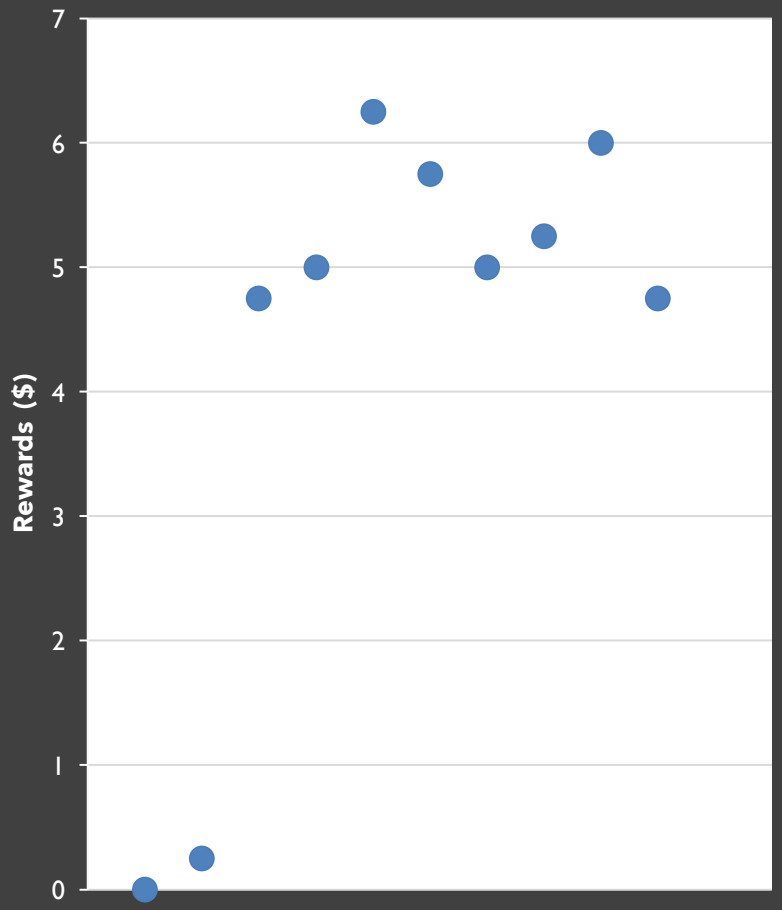


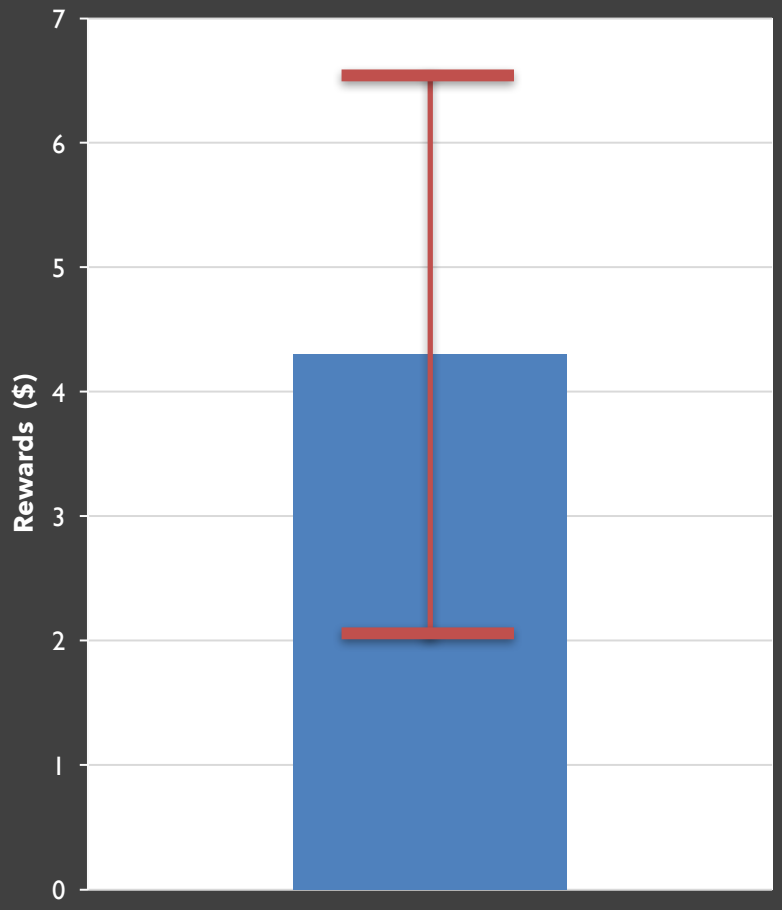
Expected Value





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.

M = 4.14

SD = 2.33

n = 10

P(μ>4) = **0.95**

❙ Take the bet

MODEL

MEASUREMENT

Statistical Inference

Assuming bet returns are normally distributed.

M = 4.14

SD = 2.33

n = 10

P(μ>4) = **0.95**

❙ Take the bet

MODEL

MEASUREMENT

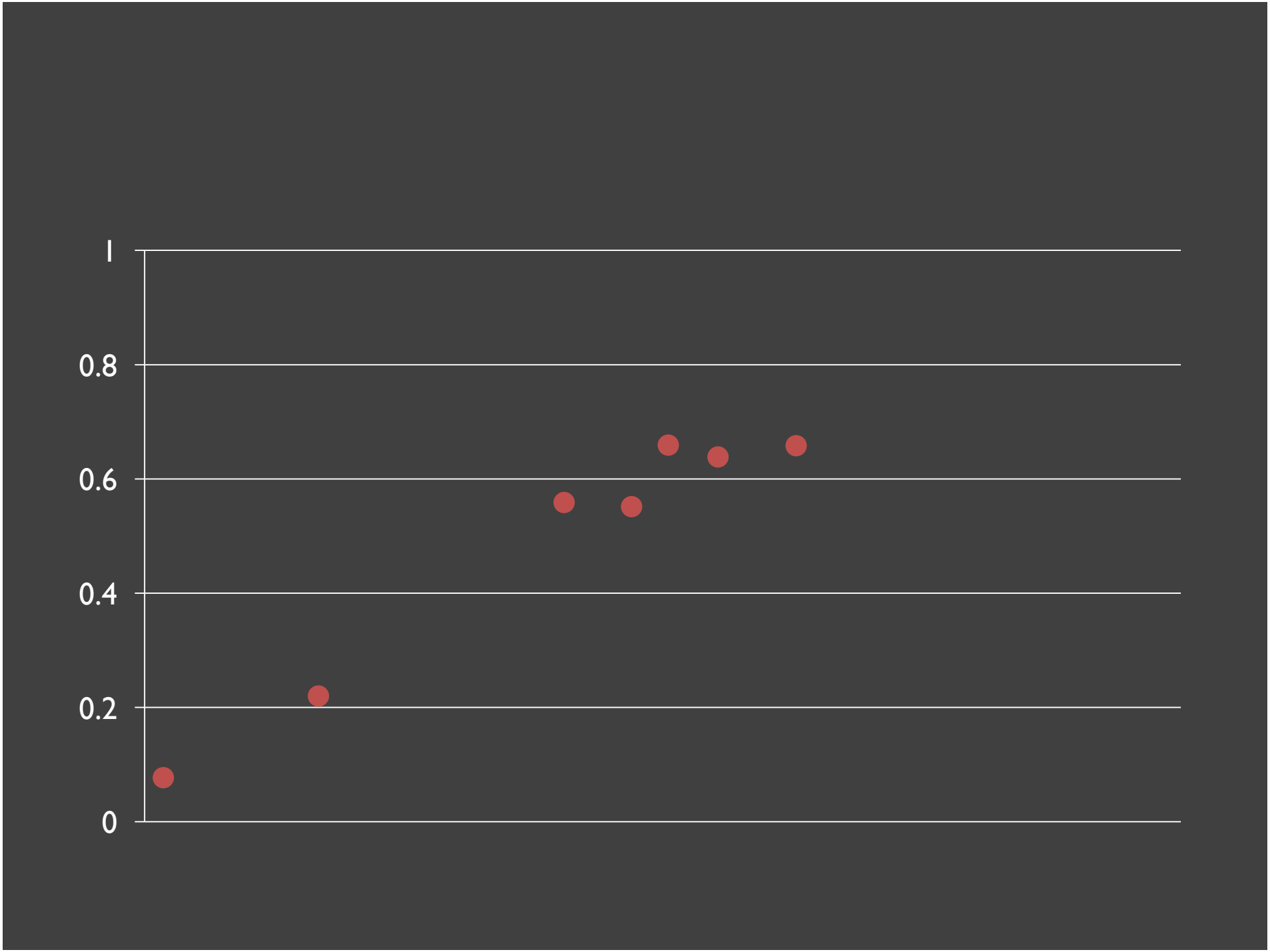
DECISION

Uncertainty Sources

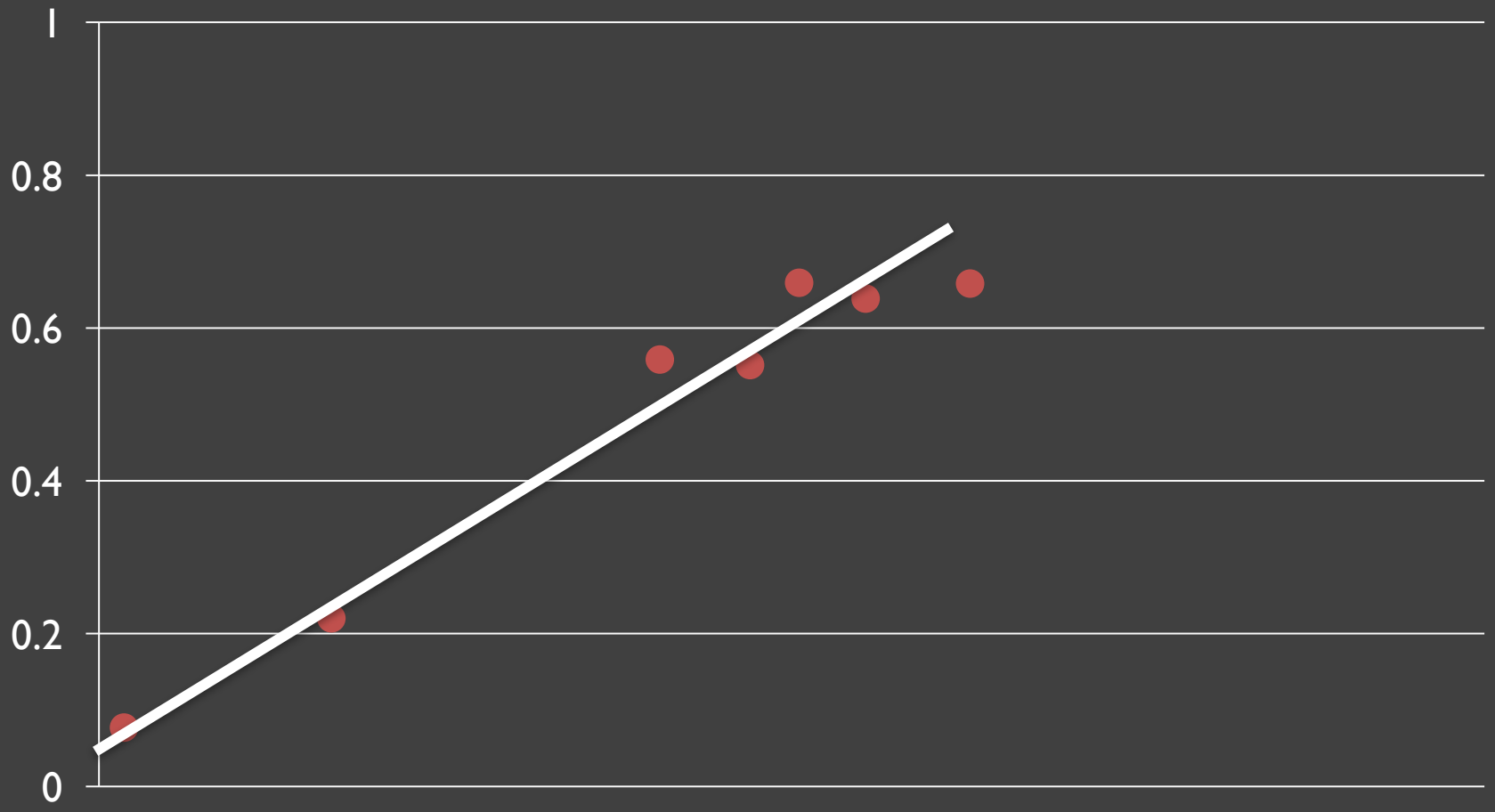
Model Uncertainty: “We’re not sure how the data fit together”

Measurement Uncertainty: “We’re not sure what the data are”

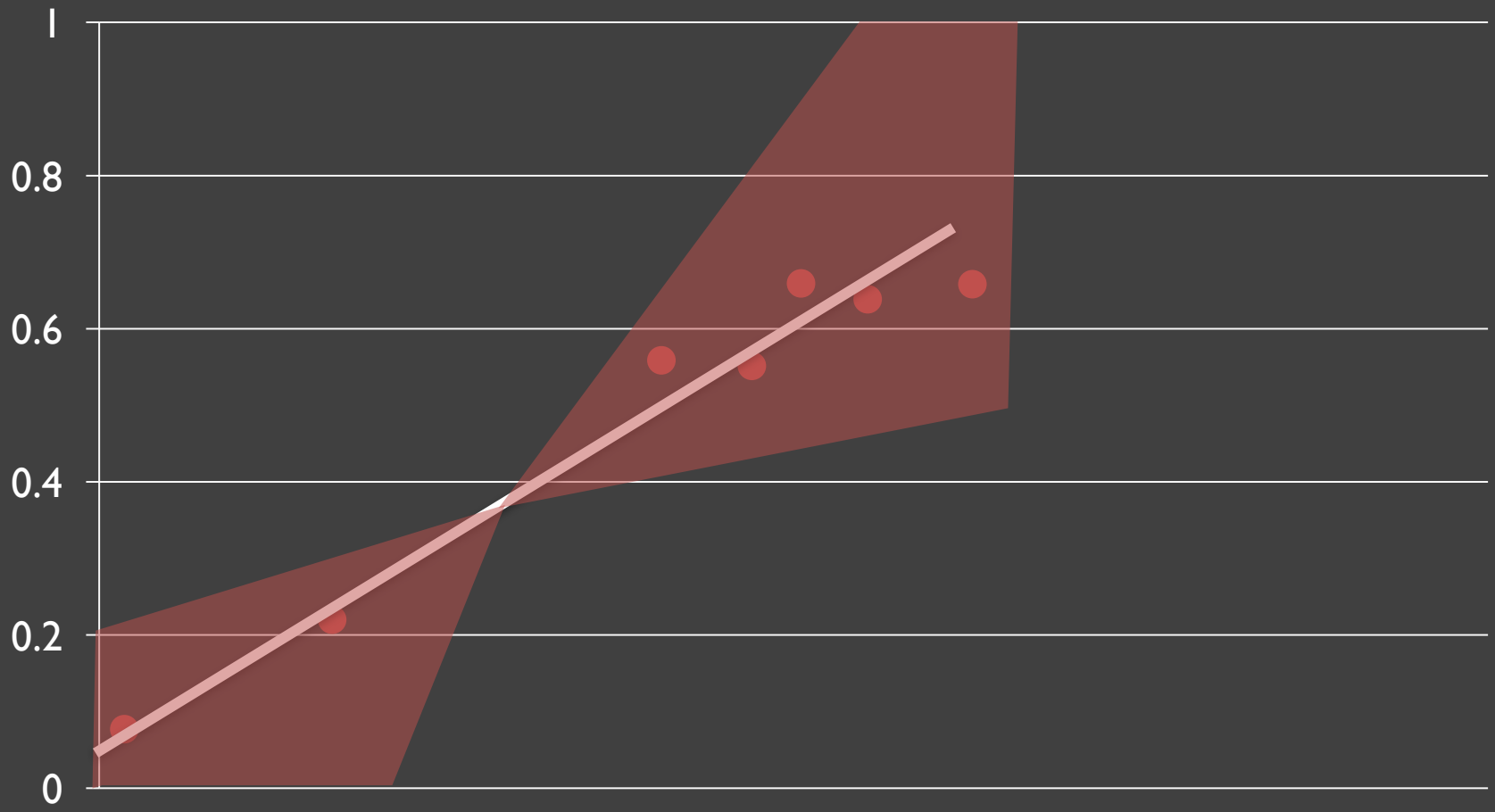
Decision Uncertainty: “We’re not sure what to do now that we have the data”


1
0.8
0.6
0.4
0.2
0

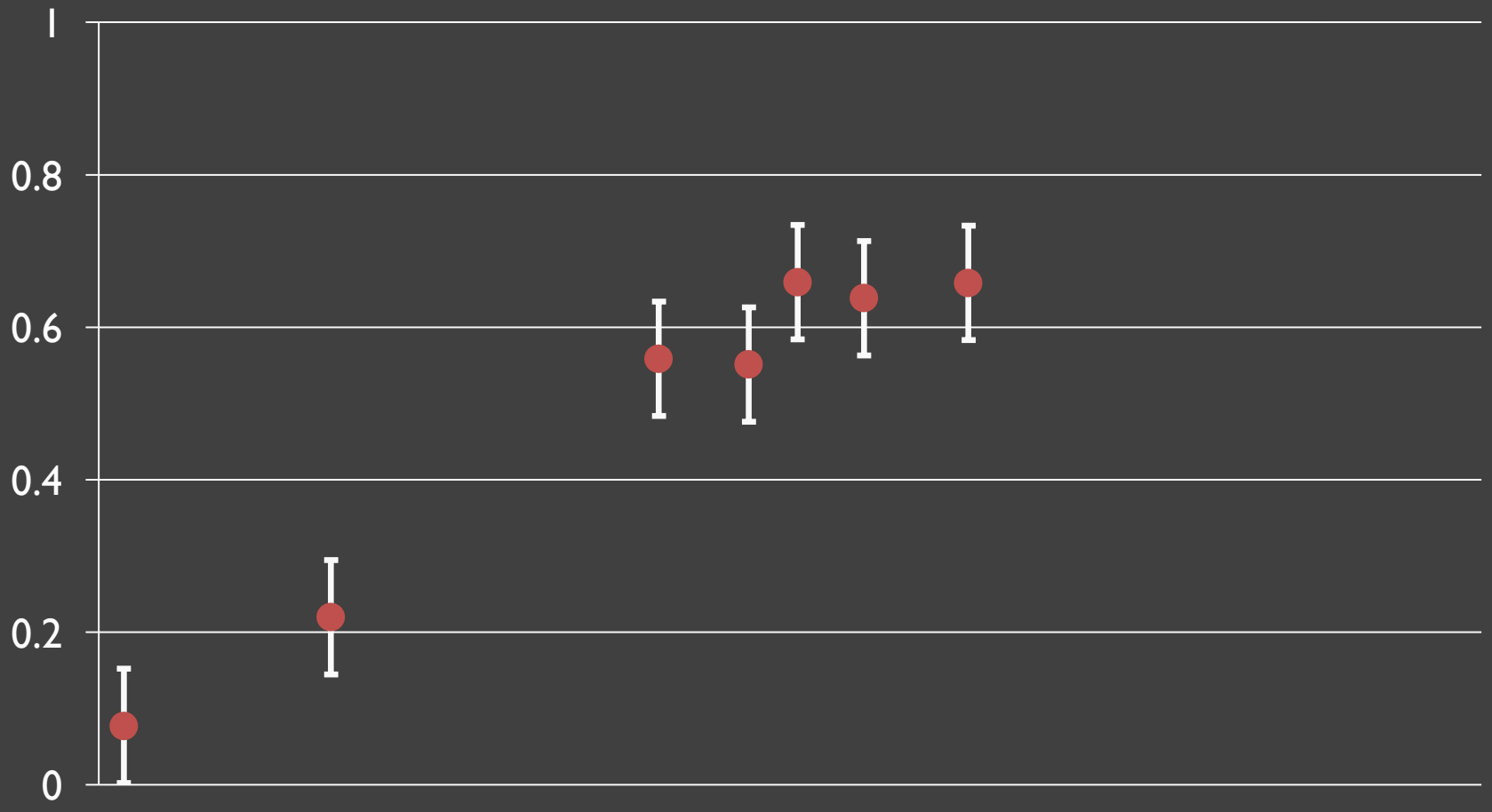
Model Uncertainty



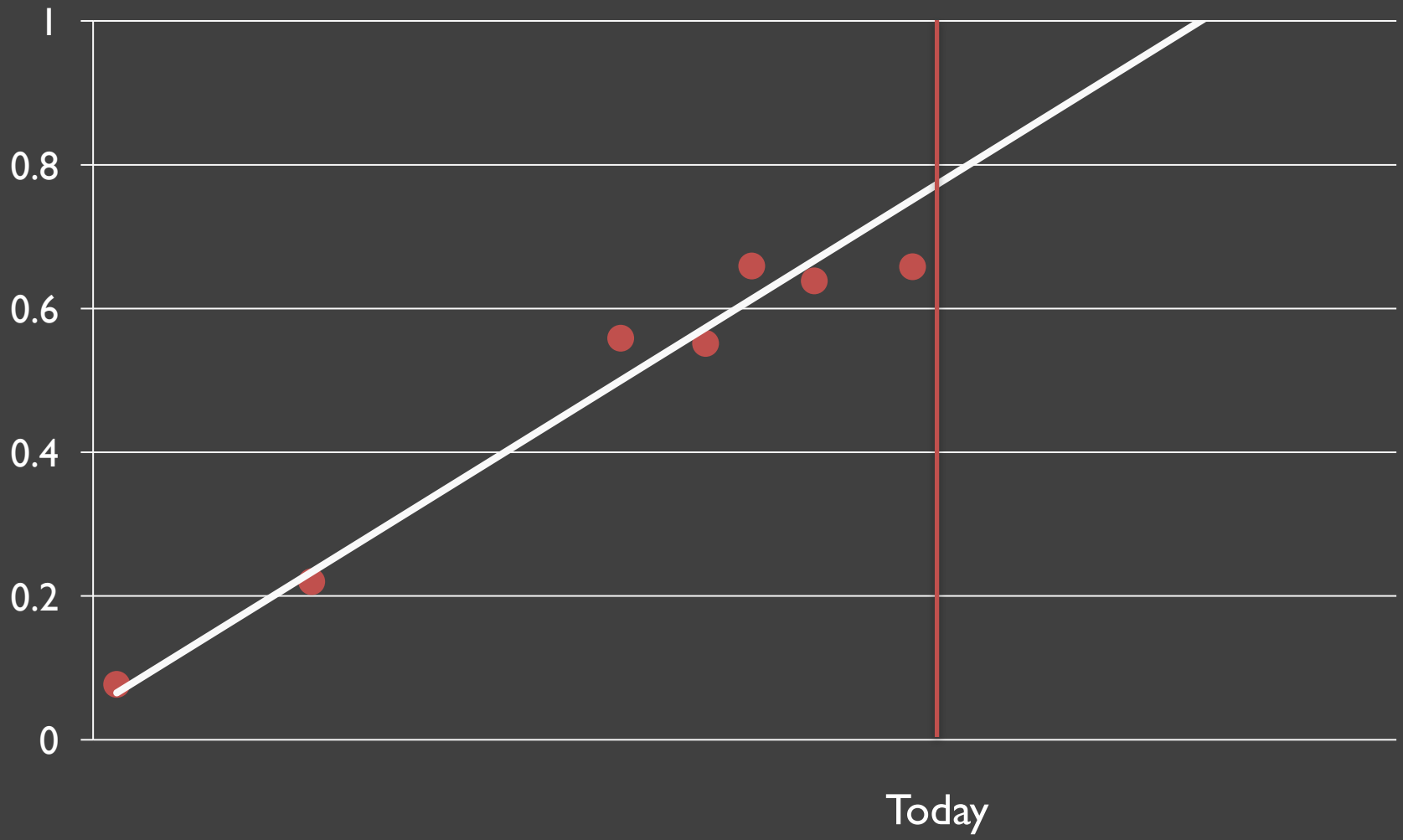
Model Uncertainty



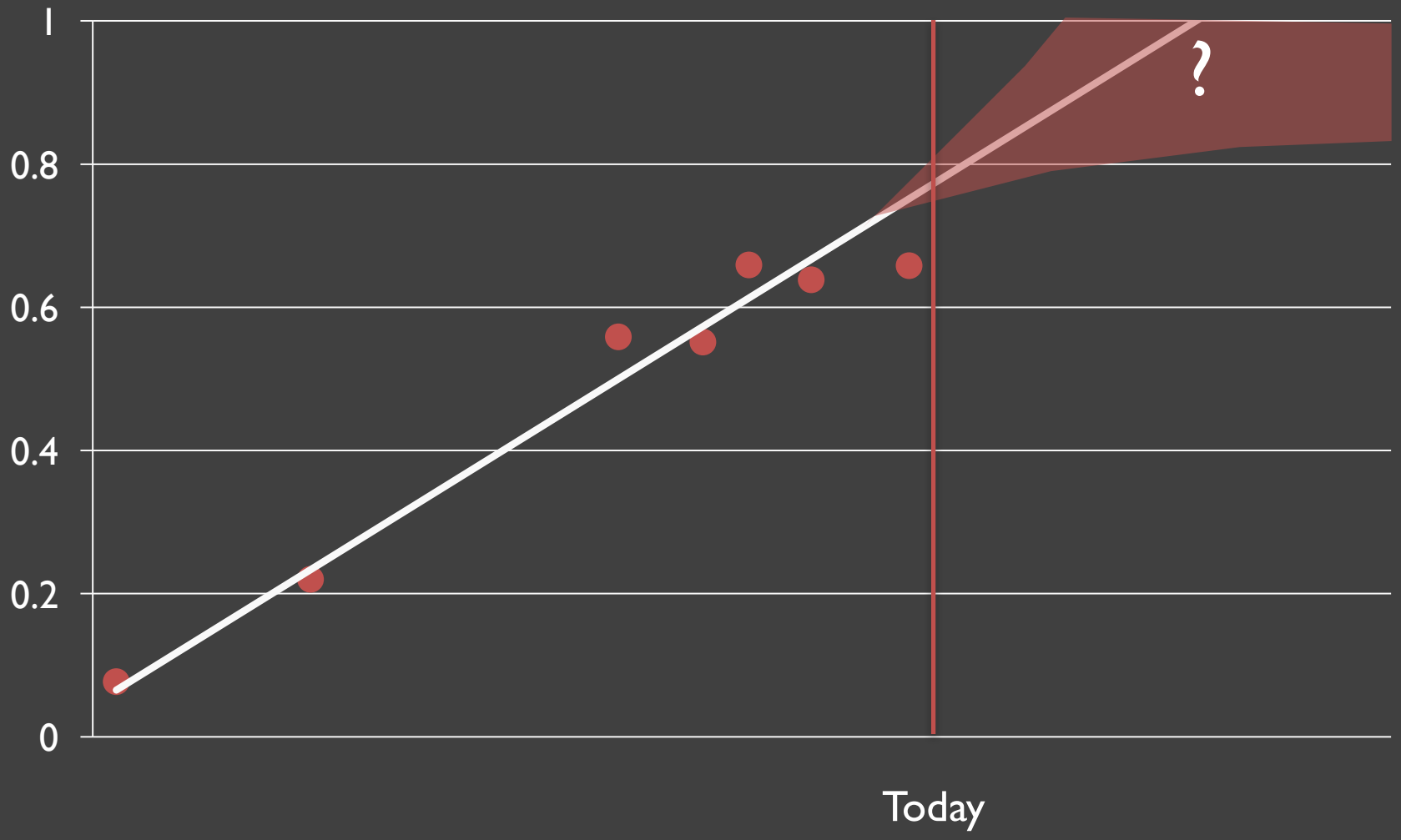
Measurement Uncertainty



Decision Uncertainty



Decision Uncertainty



What Does Uncertainty Mean?

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

What Does Uncertainty Mean?

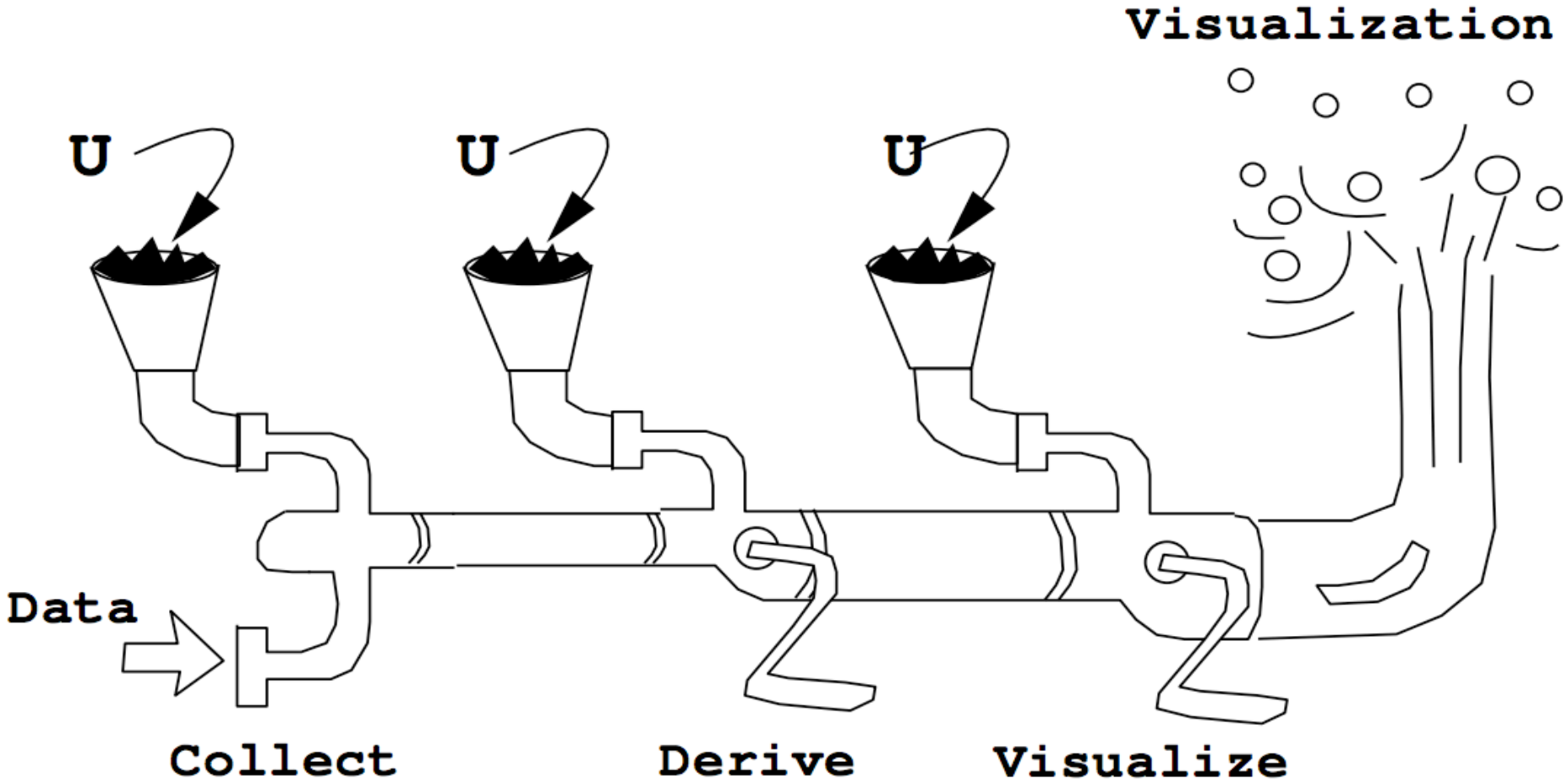
Any one of a number of potentially interconnected quantitative, qualitative, or factors that affect the quality, reliability, or utility of your data or data-driven decisions. Anything that can cause you to be unsure about your data or how to use it.

LOTS OF THINGS

Uncertainty Maps and Model Visualization

HOW SHOULD I VISUALIZE UNCERTAINTY?

Uncertainty Vis Pipeline

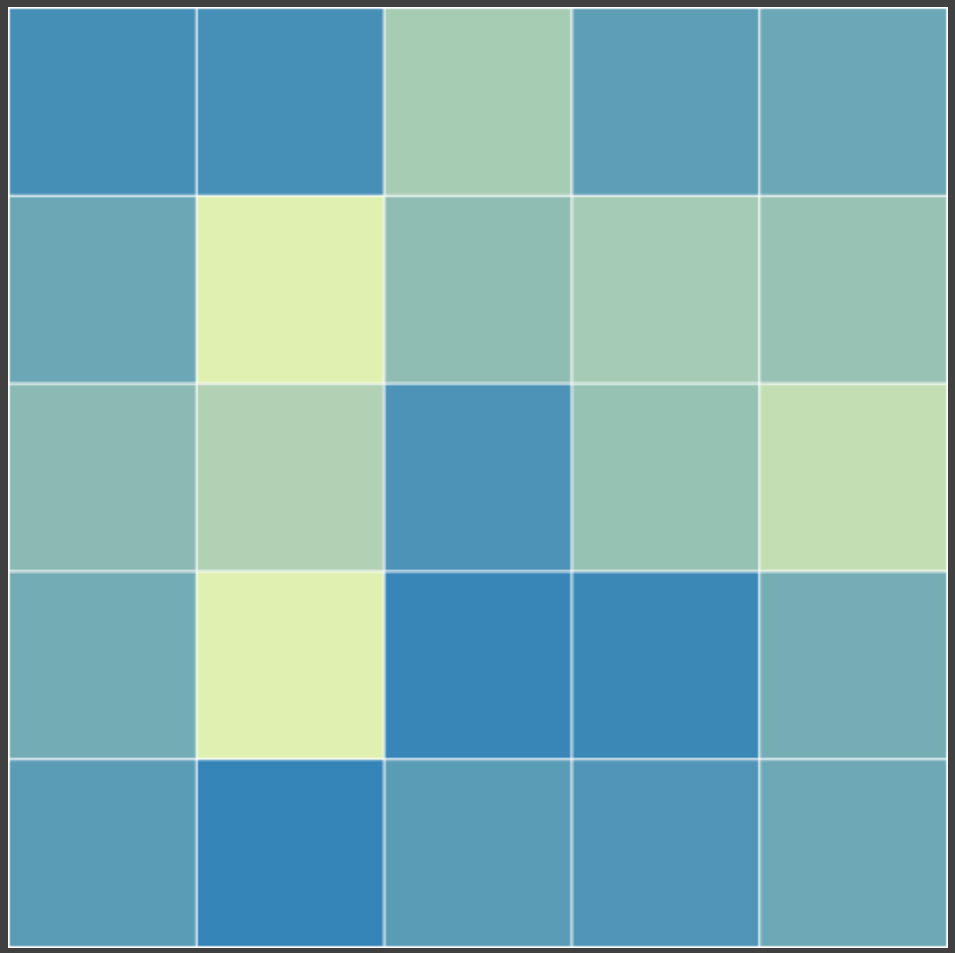


Pang et al. Approaches to Uncertainty Visualization. The Visual Computer, 1997.

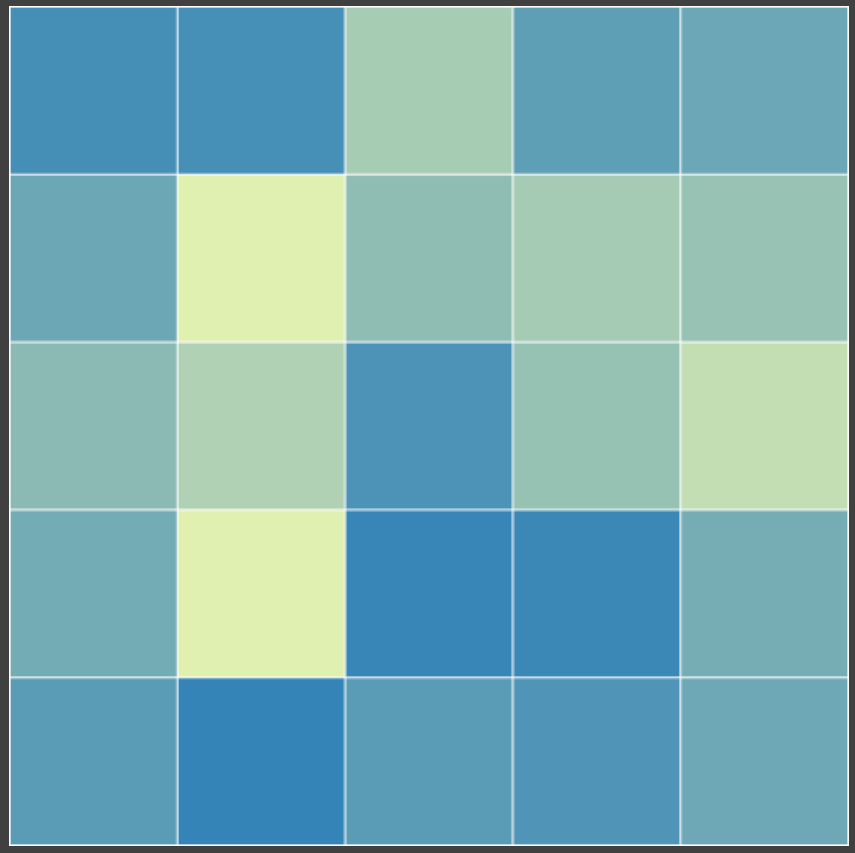
Uncertainty Vis Pipeline

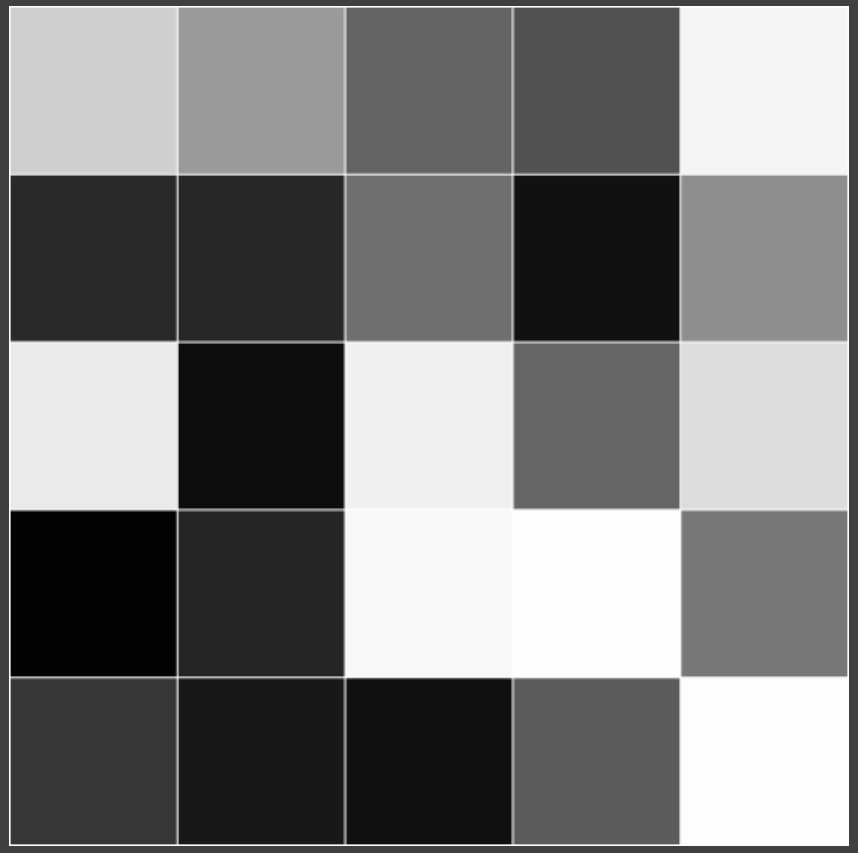
1) Quantify Uncertainty
2) Choose a free visual variable
3) Encode uncertainty with the variable

Data Map

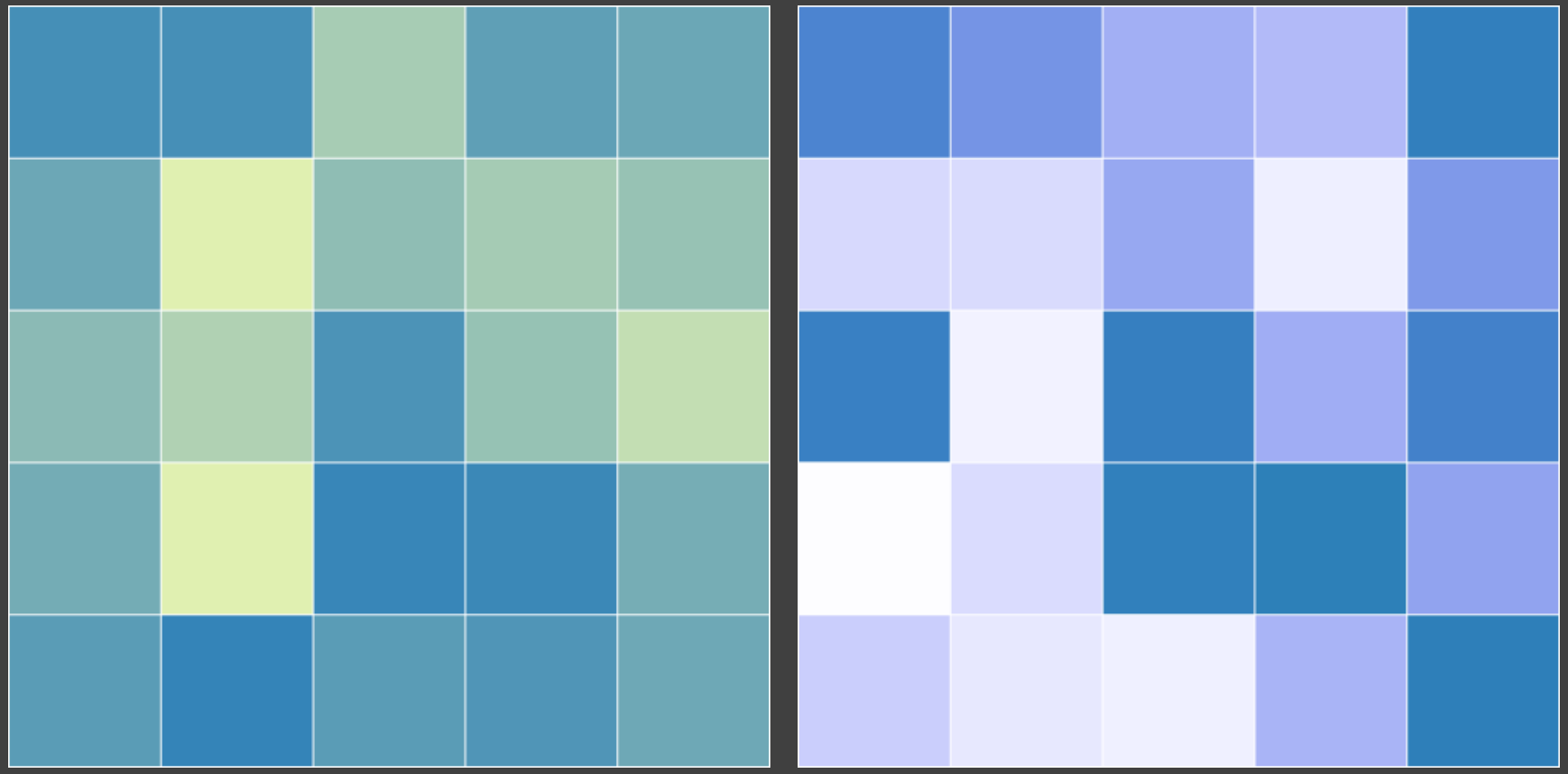


Juxtaposition

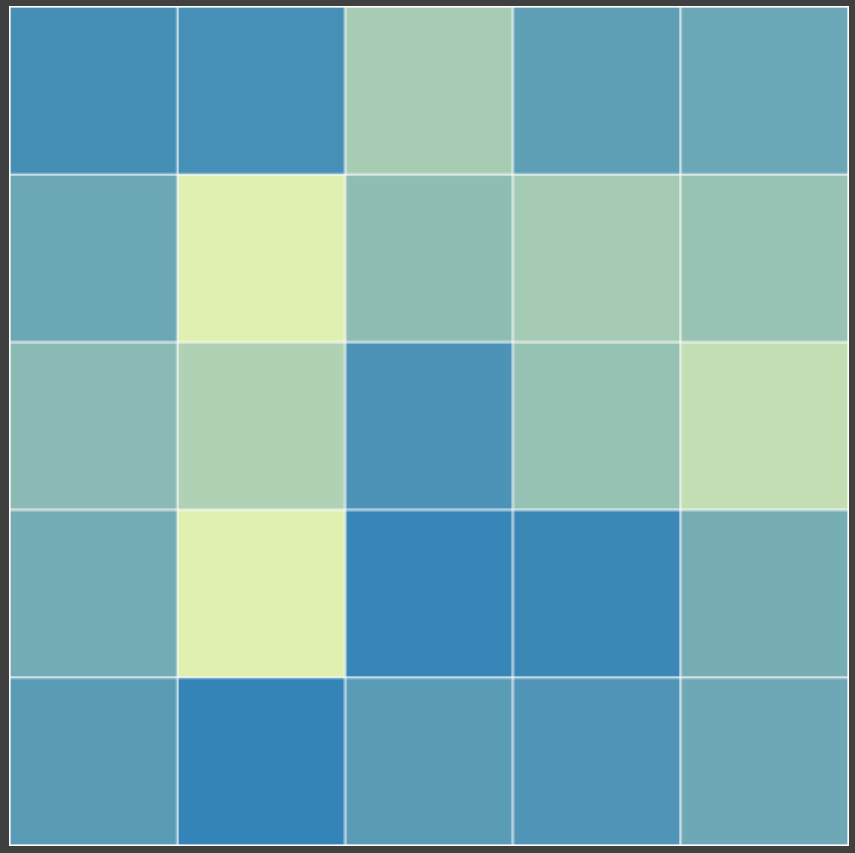


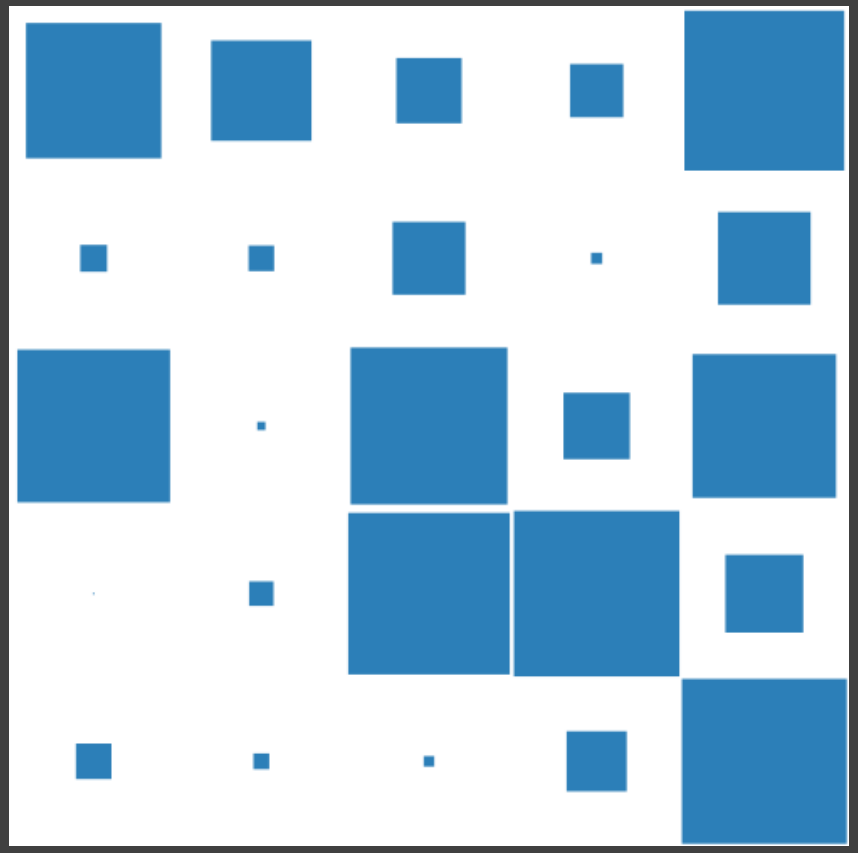


Juxtaposition

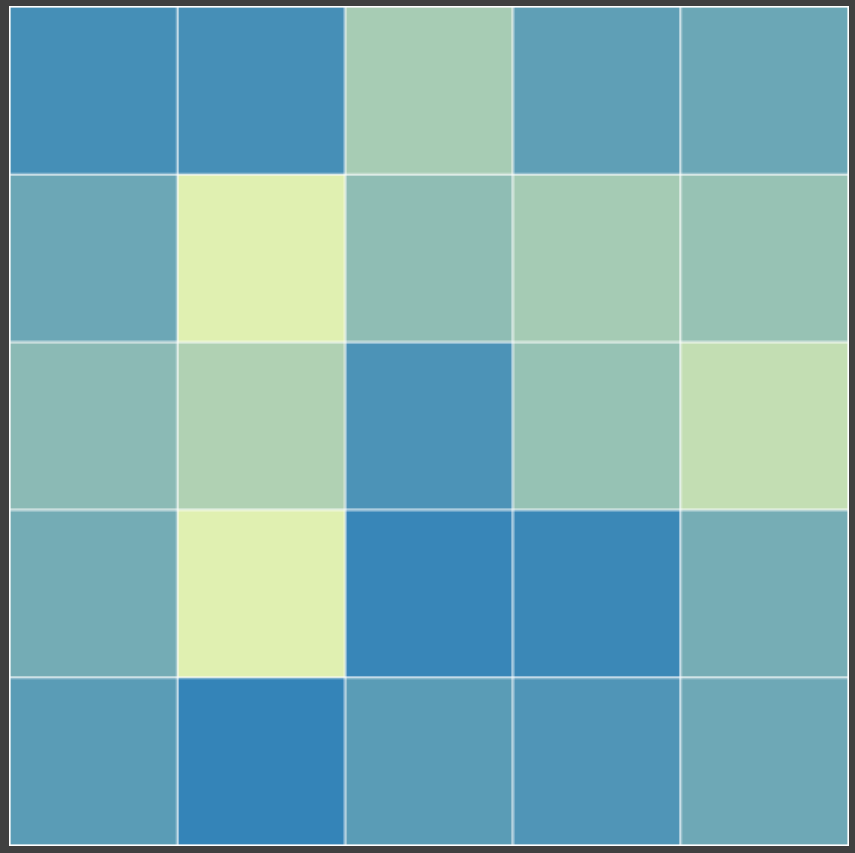


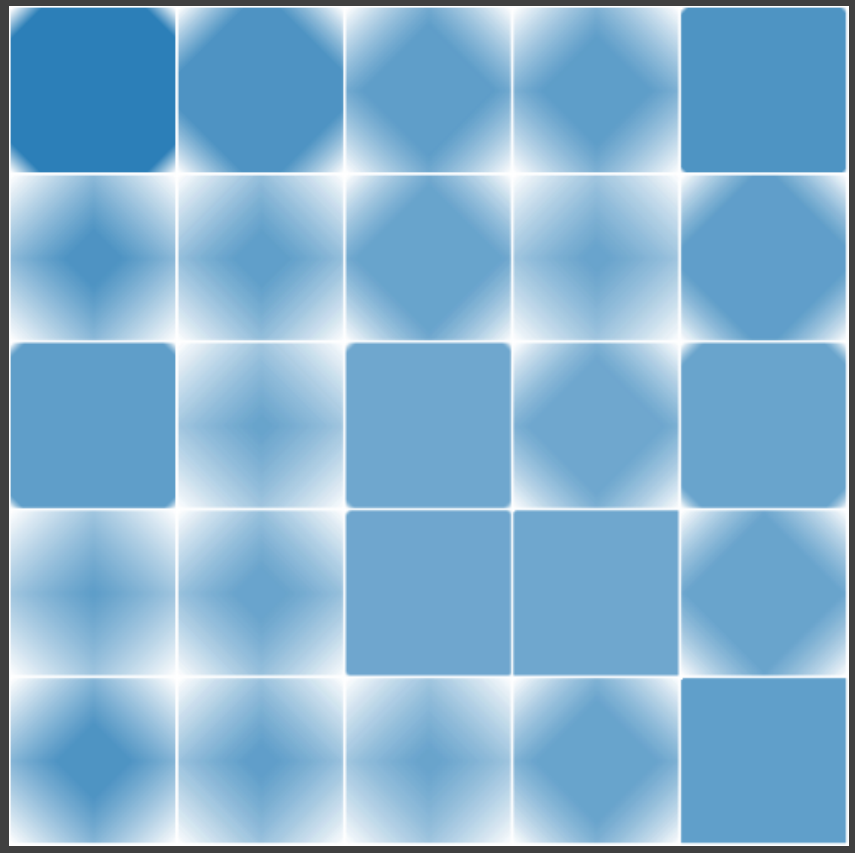
Juxtaposition



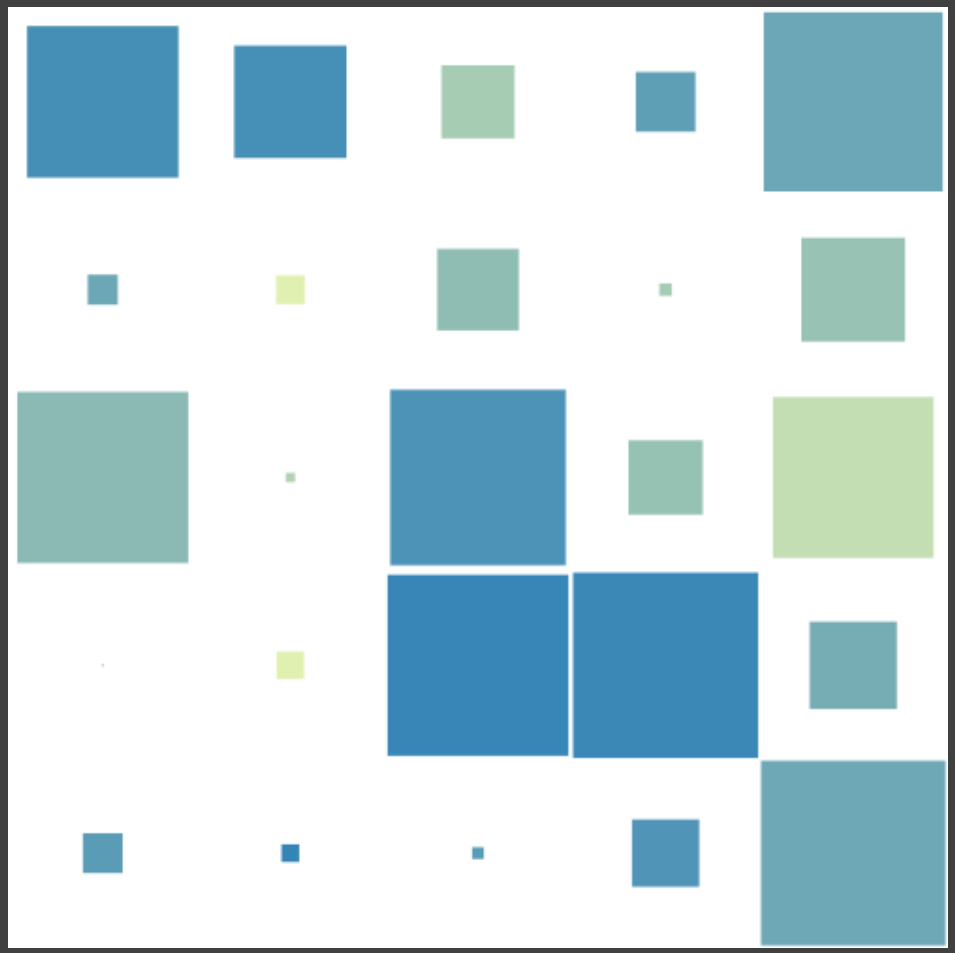


Juxtaposition

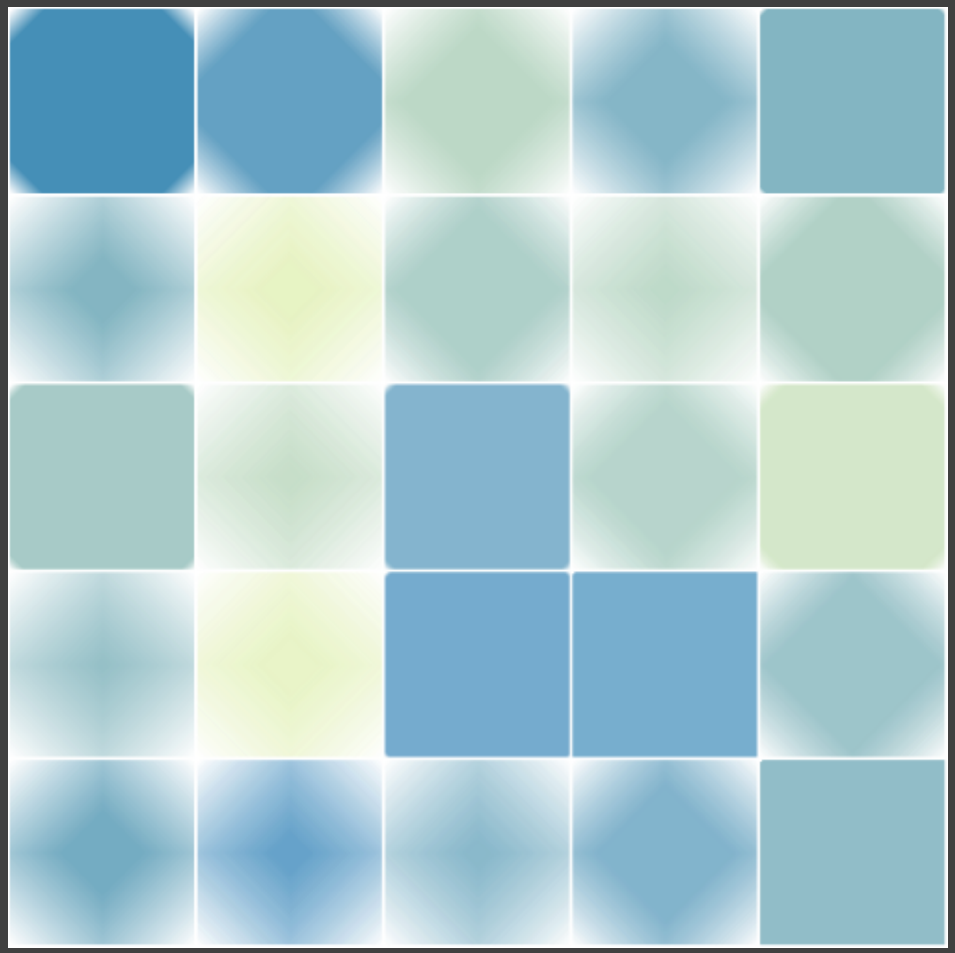




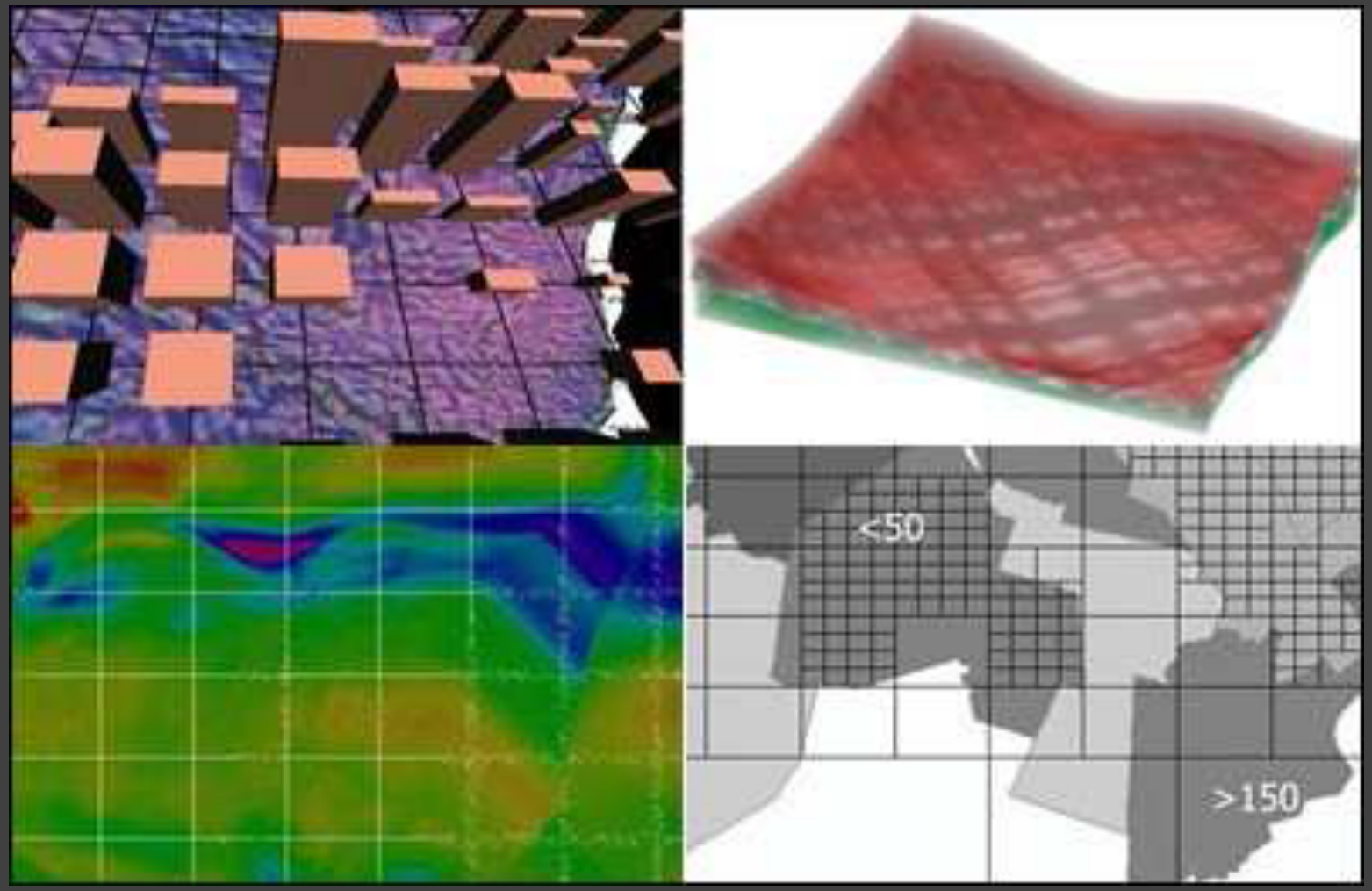
Superposition



Superposition



Superposition



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

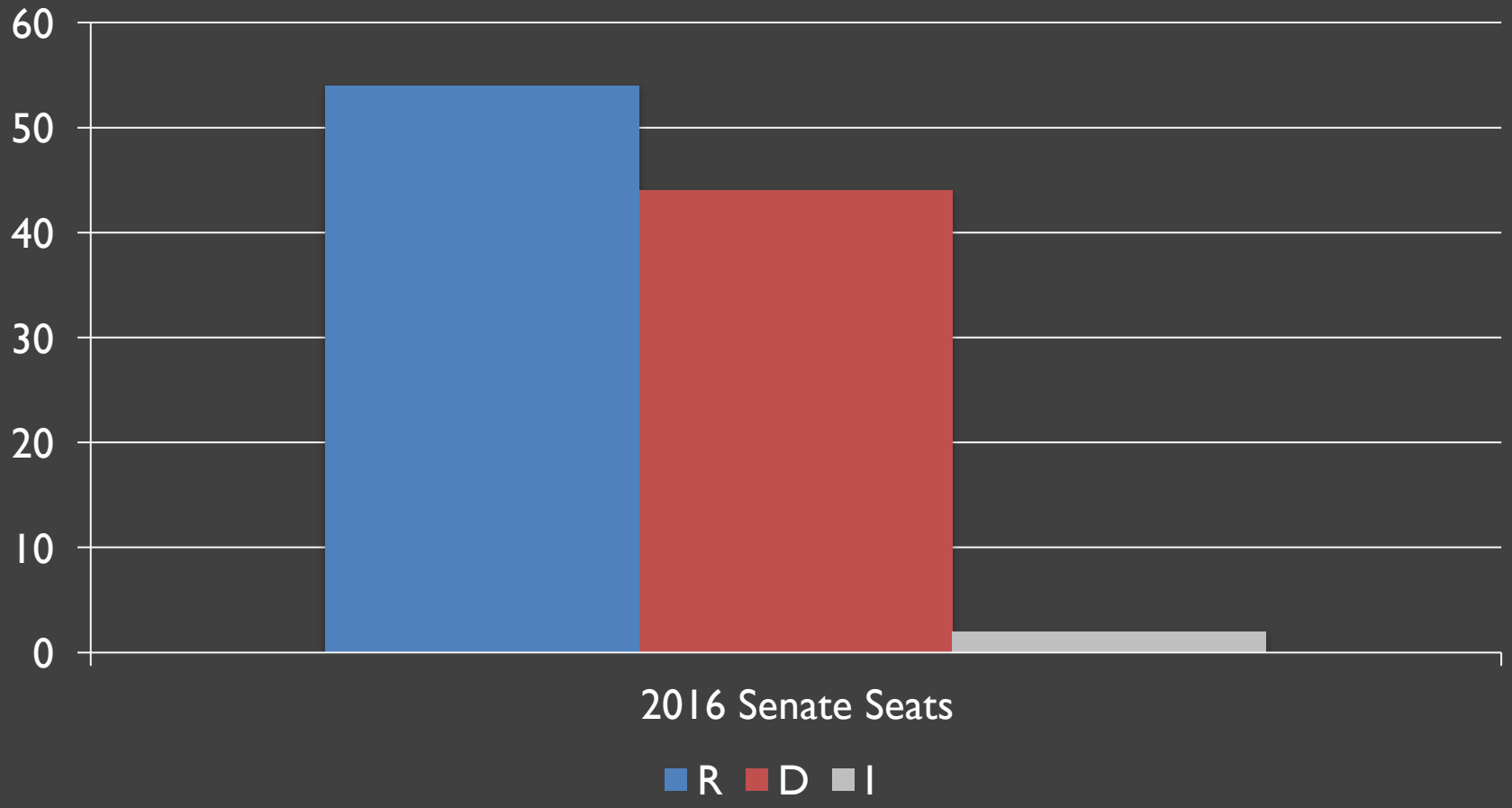
Uncertainty Vis Pipeline?

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

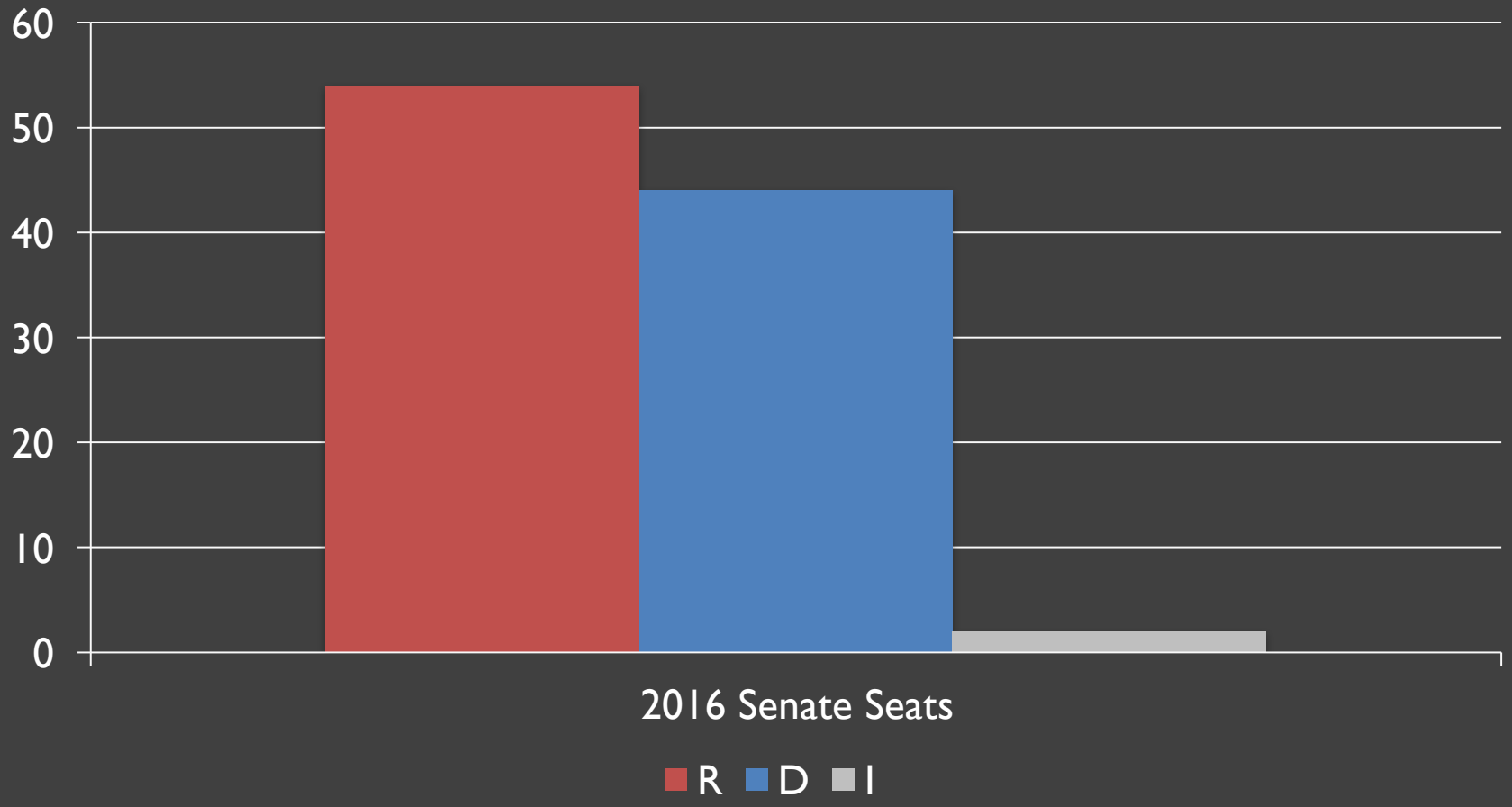
Design Decisions:

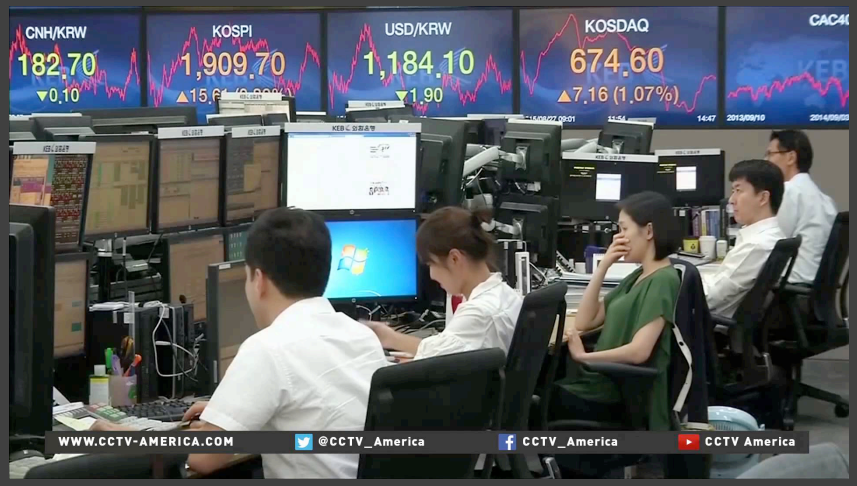
How to unify data and uncertainty map(s)?

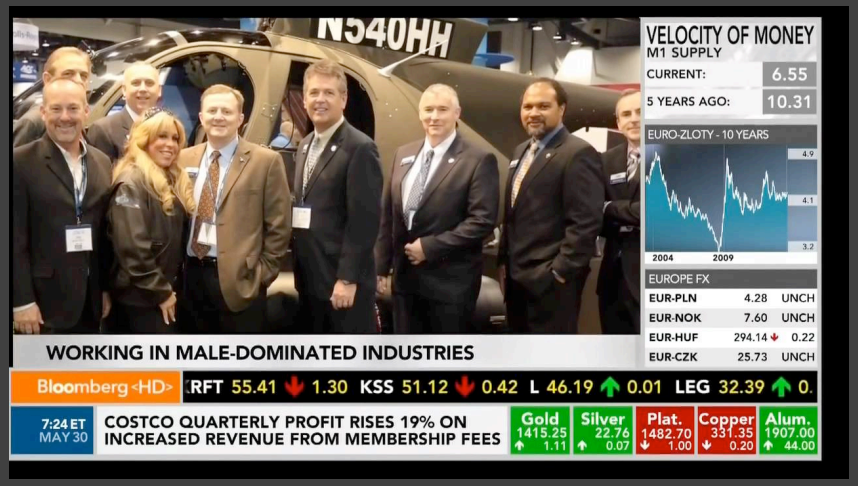
The Variable Matters!



The Variable Matters!

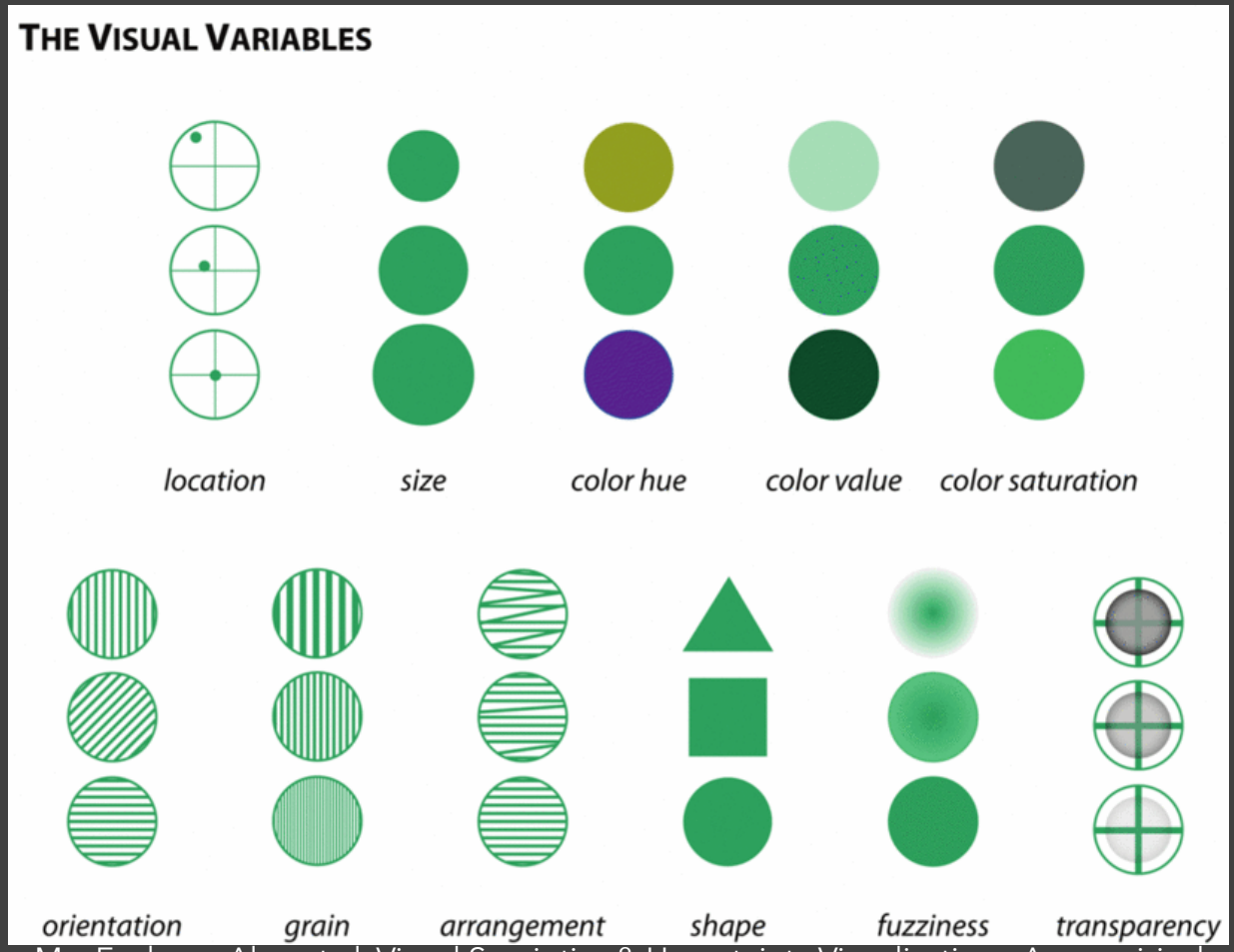



CNH/KRW
182.70
▼0.10
KOSPI
1,909.70
USD/KRW
1,184.10
▼1.90
KOSDAQ
674.60
▲7.16 (1.07%)
WWW.CCTV-AMERICA.COM
@CCTV_America
CCTV_America
CCTV America

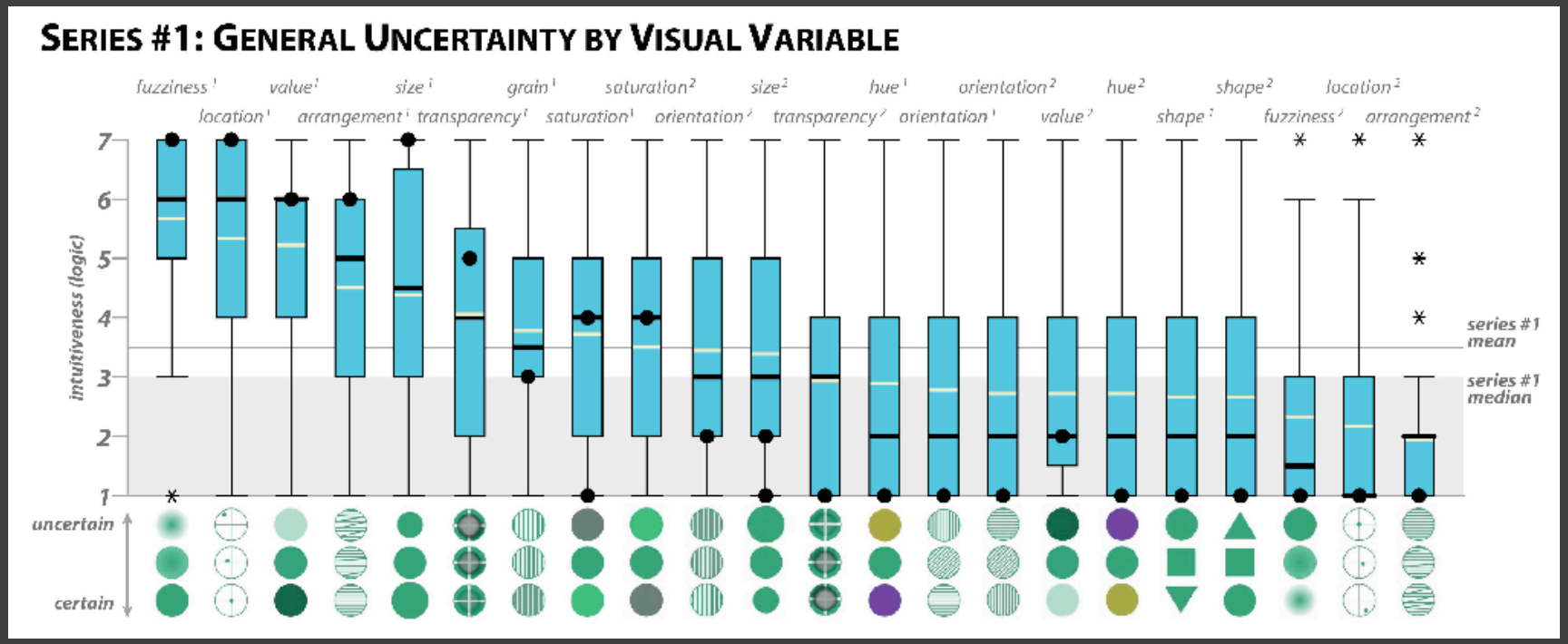

N540HH
VELOCITY OF MONEY
M1 SUPPLY
CURRENT: 6.55
5 YEARS AGO: 10.31
EURO-ZLOTY - 10 YEARS
4.9
4.1
3.2
2004
2009
EUROPE FX
EUR-PLN 4.28 UNCH
EUR-NOK 7.60 UNCH
EUR-HUF 294.14 0.22
EUR-CZK 25.73 UNCH
WORKING IN MALE-DOMINATED INDUSTRIES
Bloomberg <HD>
RFT 55.41 1.30 KSS 51.12 0.42 L 46.19 0.01 LEG 32.39 0.
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

Semiotics of Uncertainty

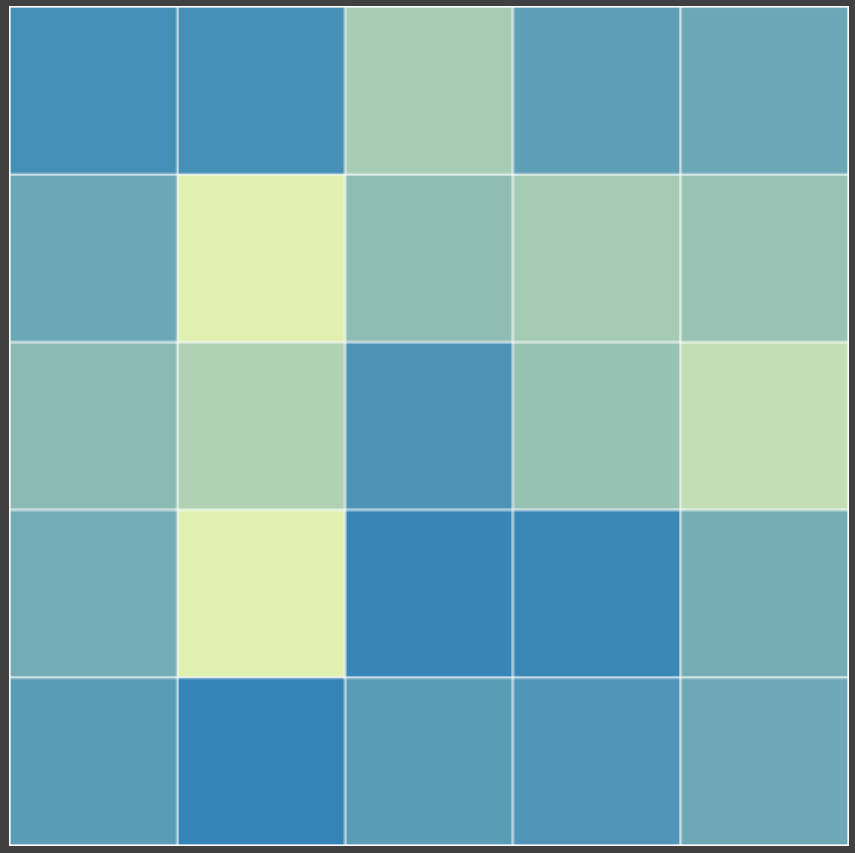
Semiotics of Uncertainty

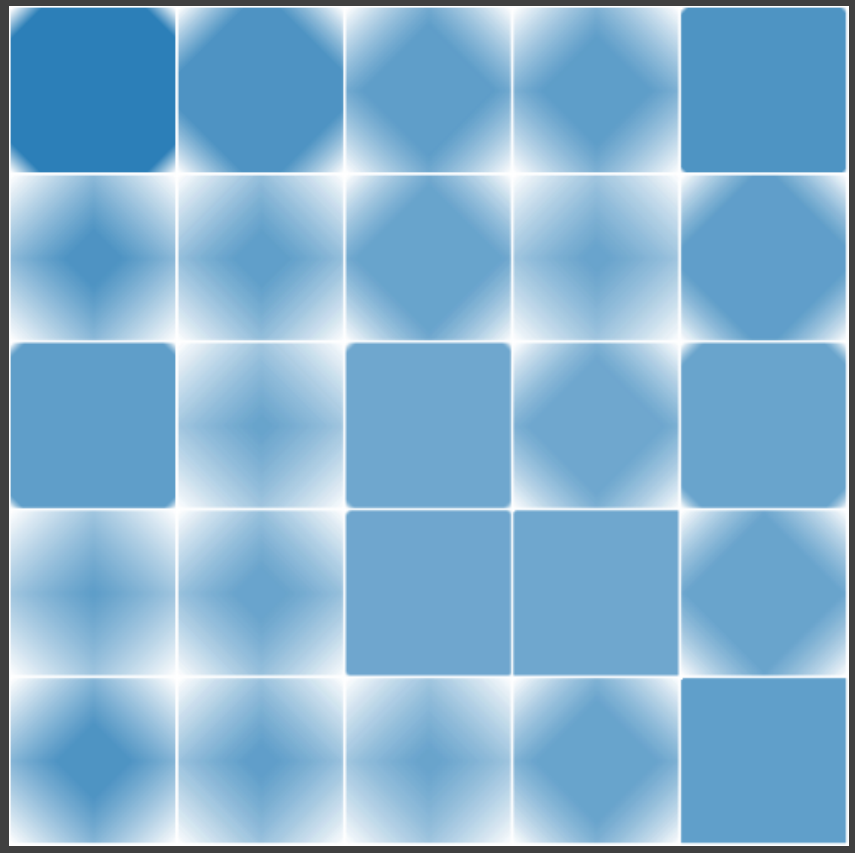


MacEachren, Alan et al. Visual Semiotics & Uncertainty Visualization: An empirical study. IEEE VIS, 2012.

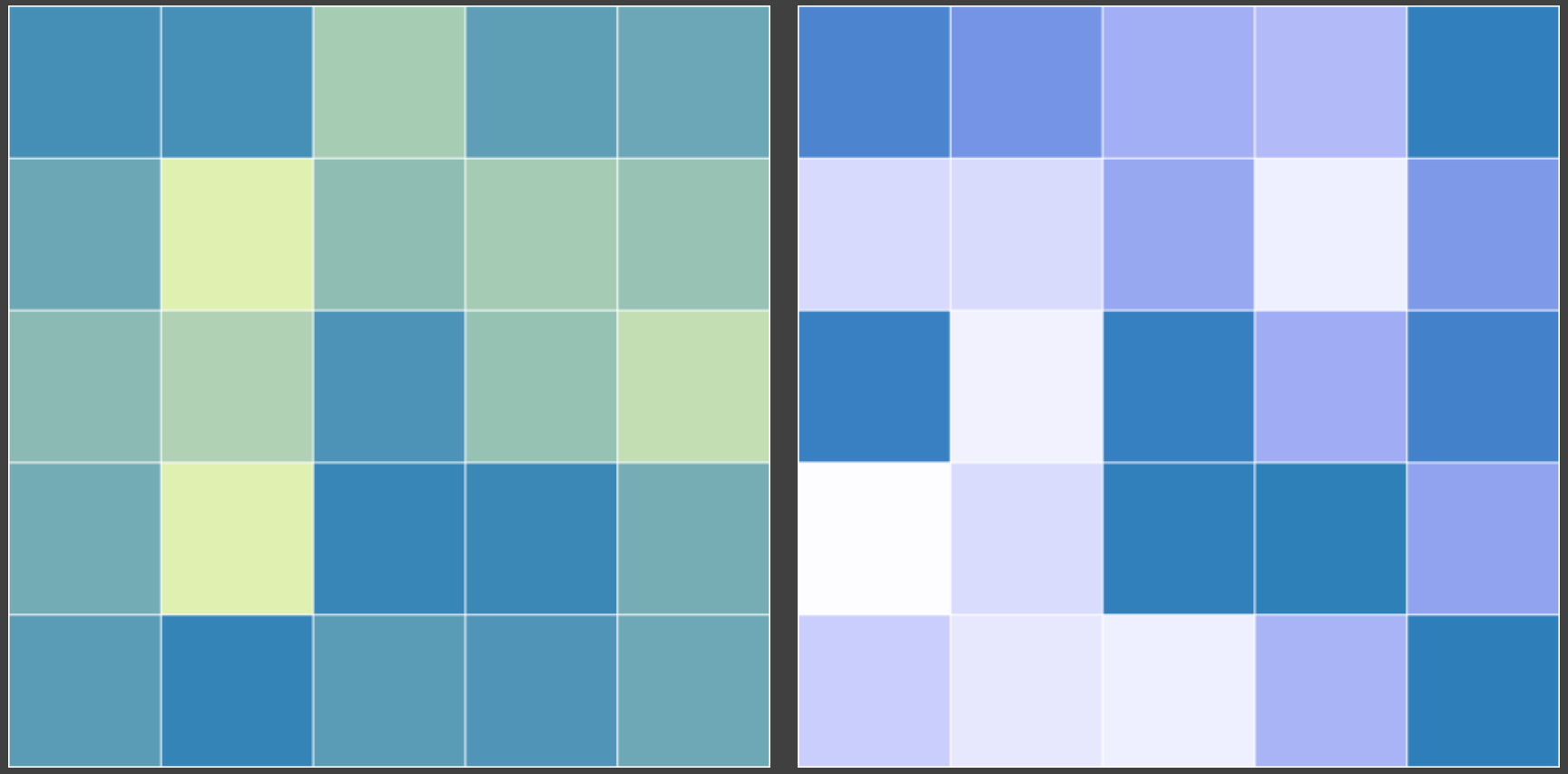

SERIES #1: GENERAL UNCERTAINTY BY VISUAL VARIABLE
fuzziness¹
location¹
value¹
arrangement¹
size¹
transparency¹
grain¹
saturation¹
saturation²
orientation²
size²
transparency²
hue¹
orientation¹
orientation²
value²
hue²
shape¹
shape²
fuzziness²
location²
arrangement²
intuitiveness (logic)
7
6
5
4
3
2
1
series #1 mean
series #1 median
uncertain
certain

Fuzziness

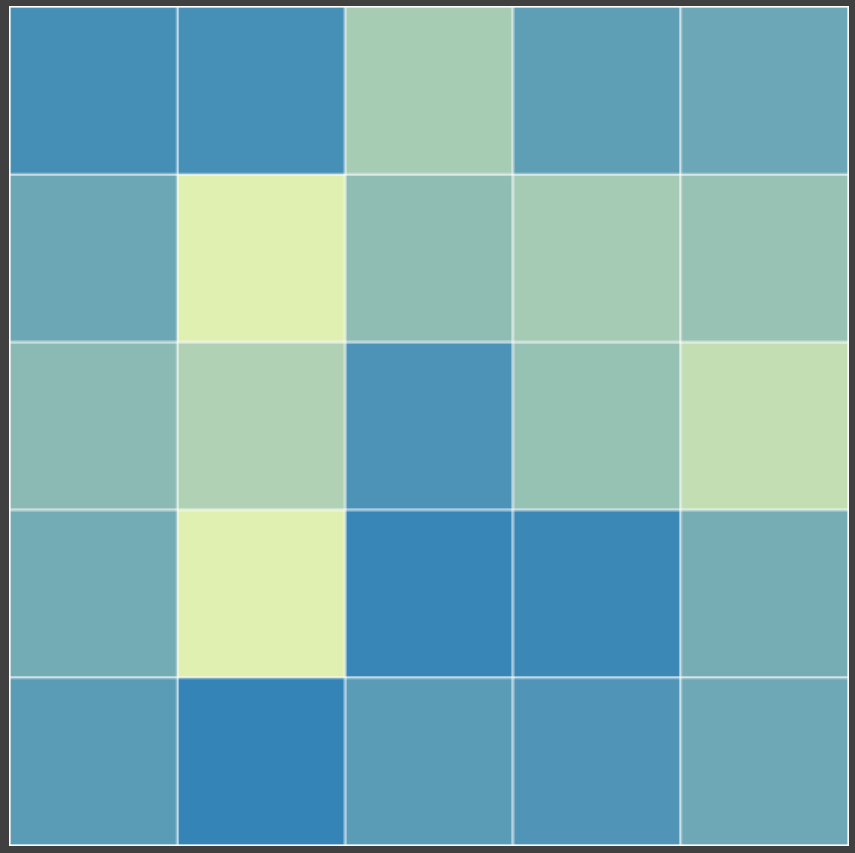


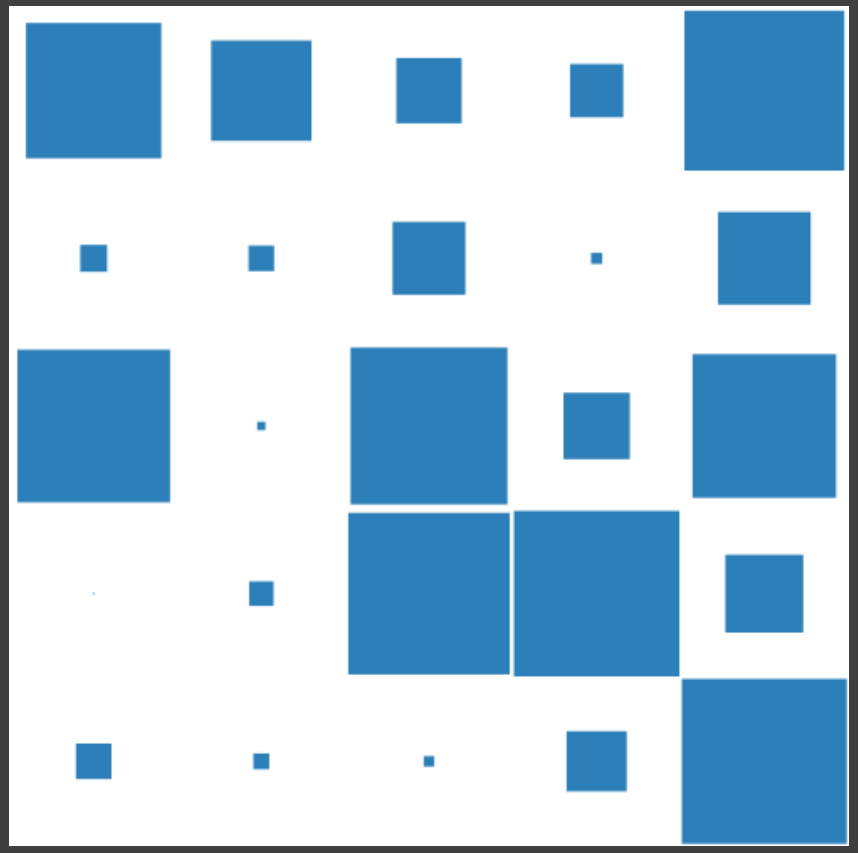


Value

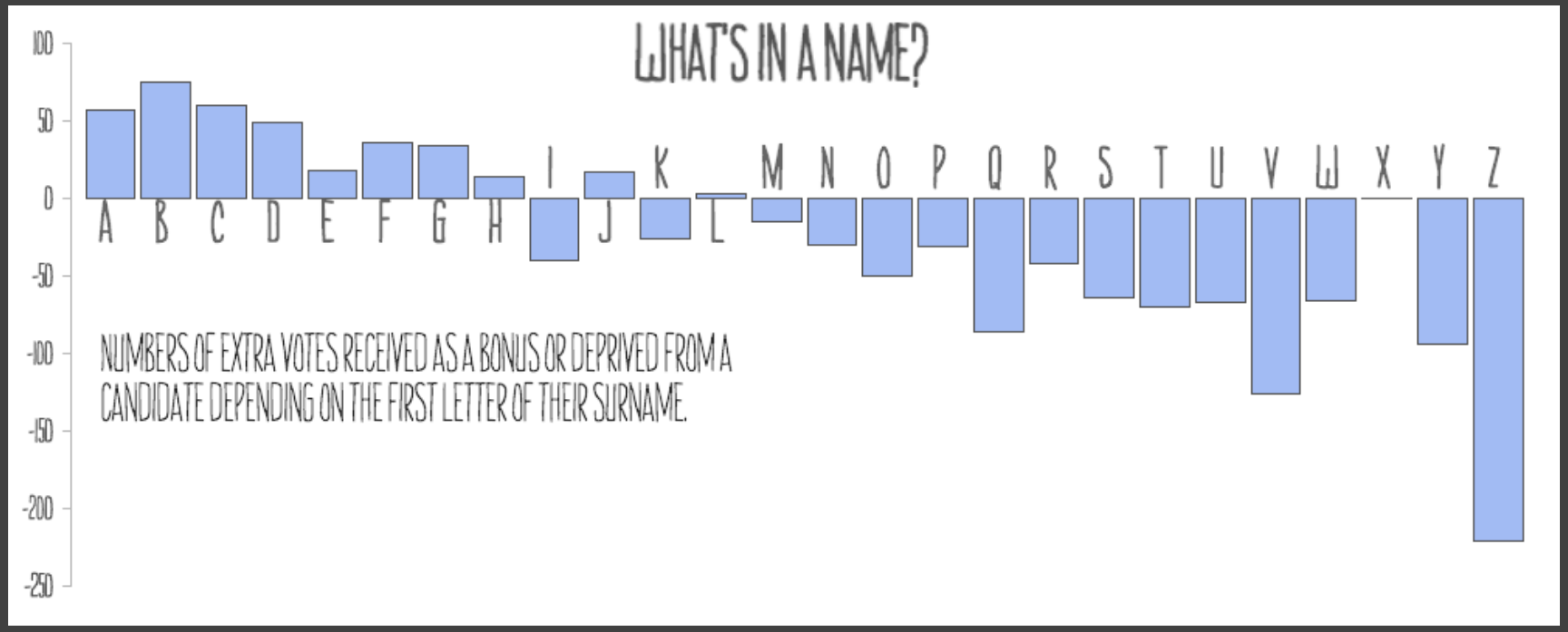


Size





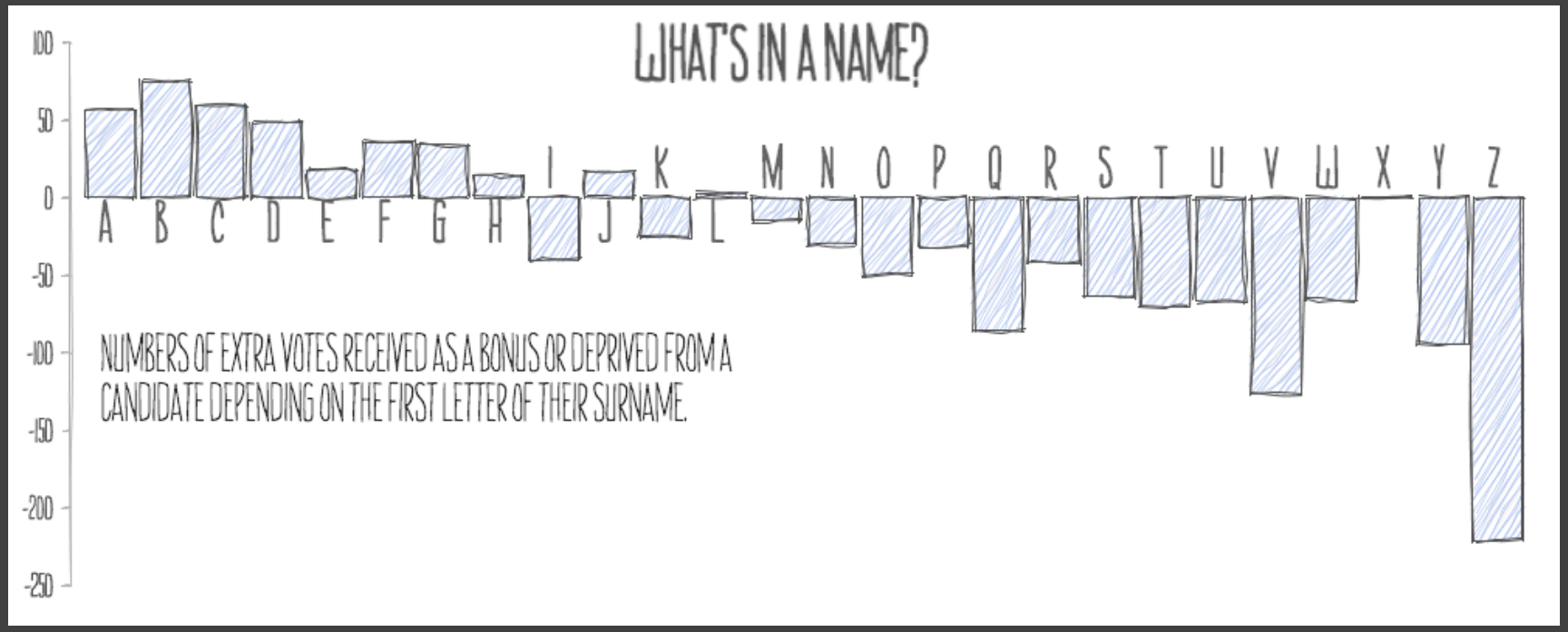
"Sketchiness"



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.

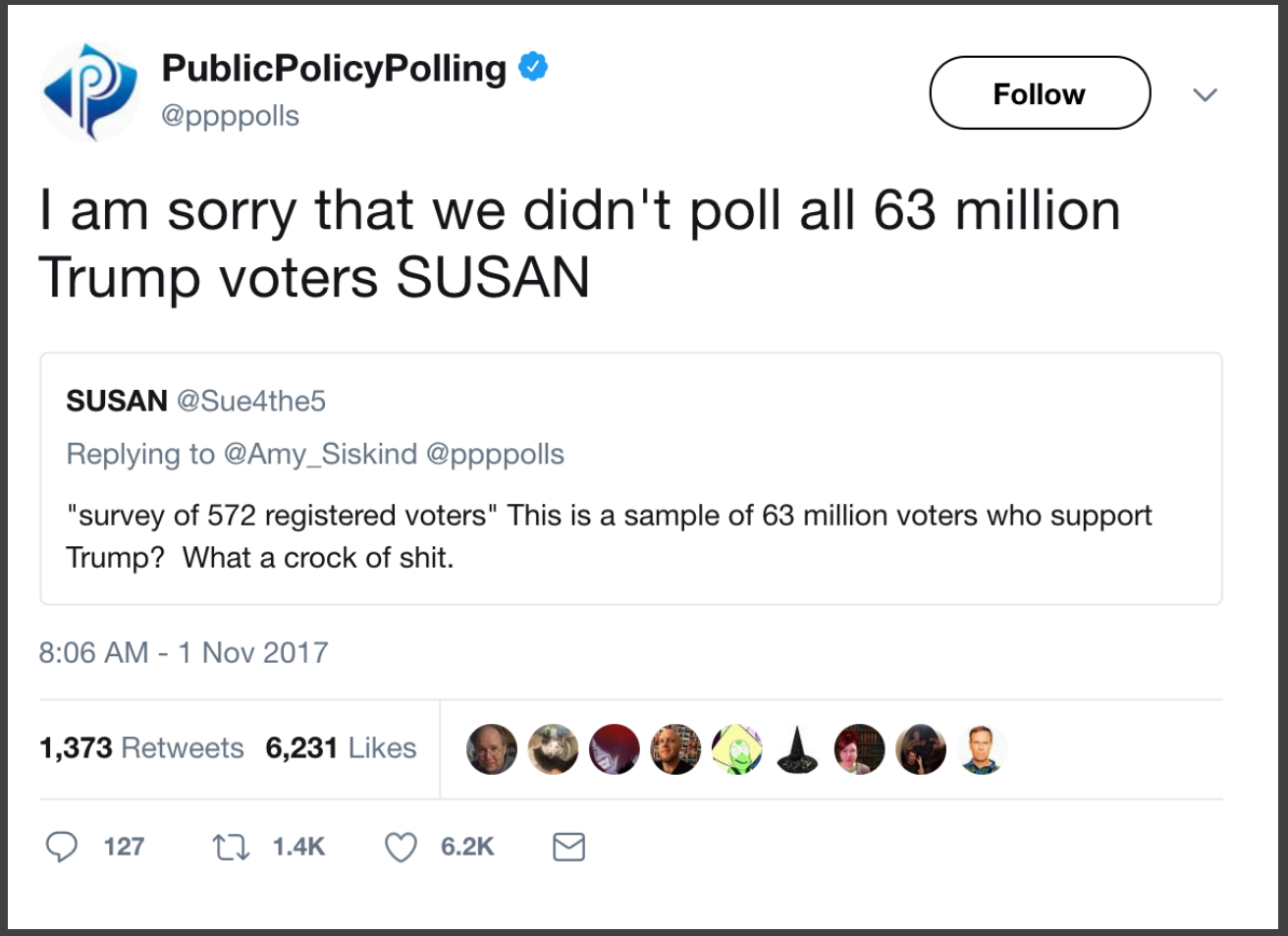
Encoding Uncertainty

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

However, intuitive variables may not always be accurately interpreted!

Model Visualization

Polling Data



PublicPolicyPolling
@ppppolls

Follow

I am sorry that we didn't poll all 63 million Trump voters SUSAN

SUSAN @Sue4the5

Replying to @Amy_Siskind @ppppolls

"survey of 572 registered voters" This is a sample of 63 million voters who support Trump? What a crock of shit.

8:06 AM - 1 Nov 2017

1,373 Retweets **6,231** Likes

127 1.4K 6.2K

Polling Data

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

Polling Data

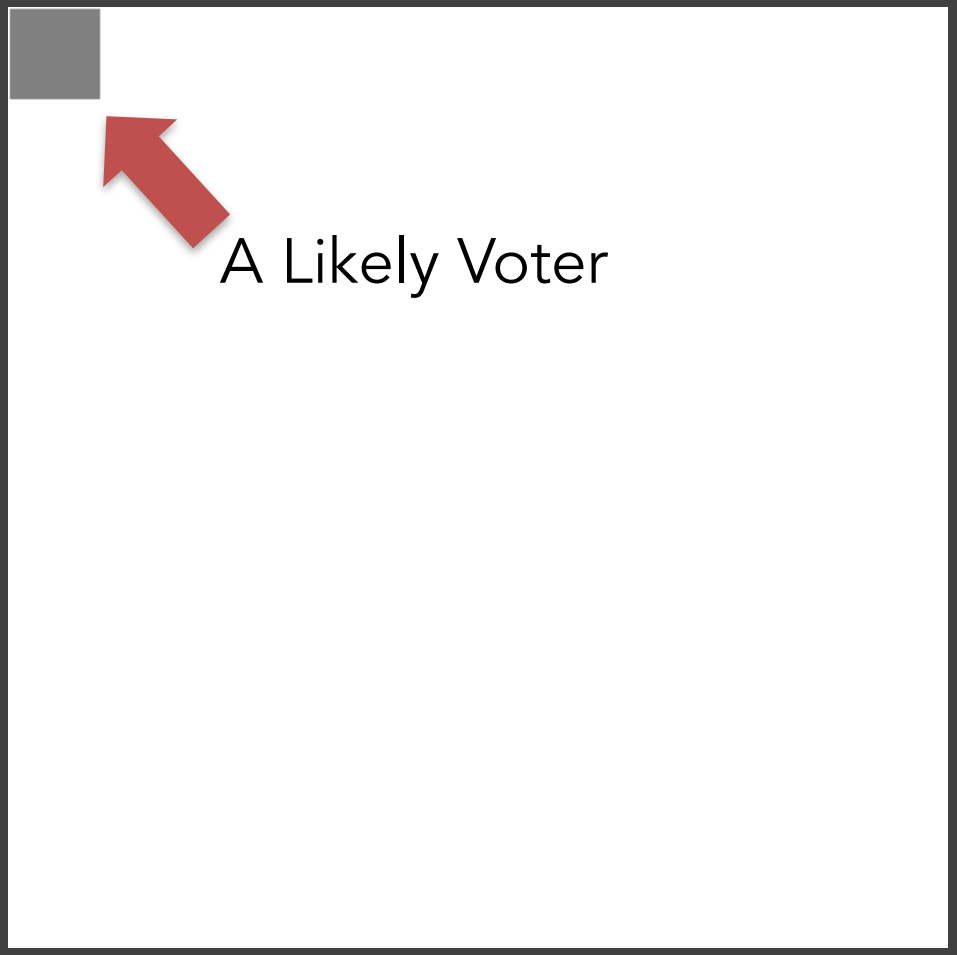
Candidate A is ahead of Candidate B in the polls, with 55% of the likely voters*

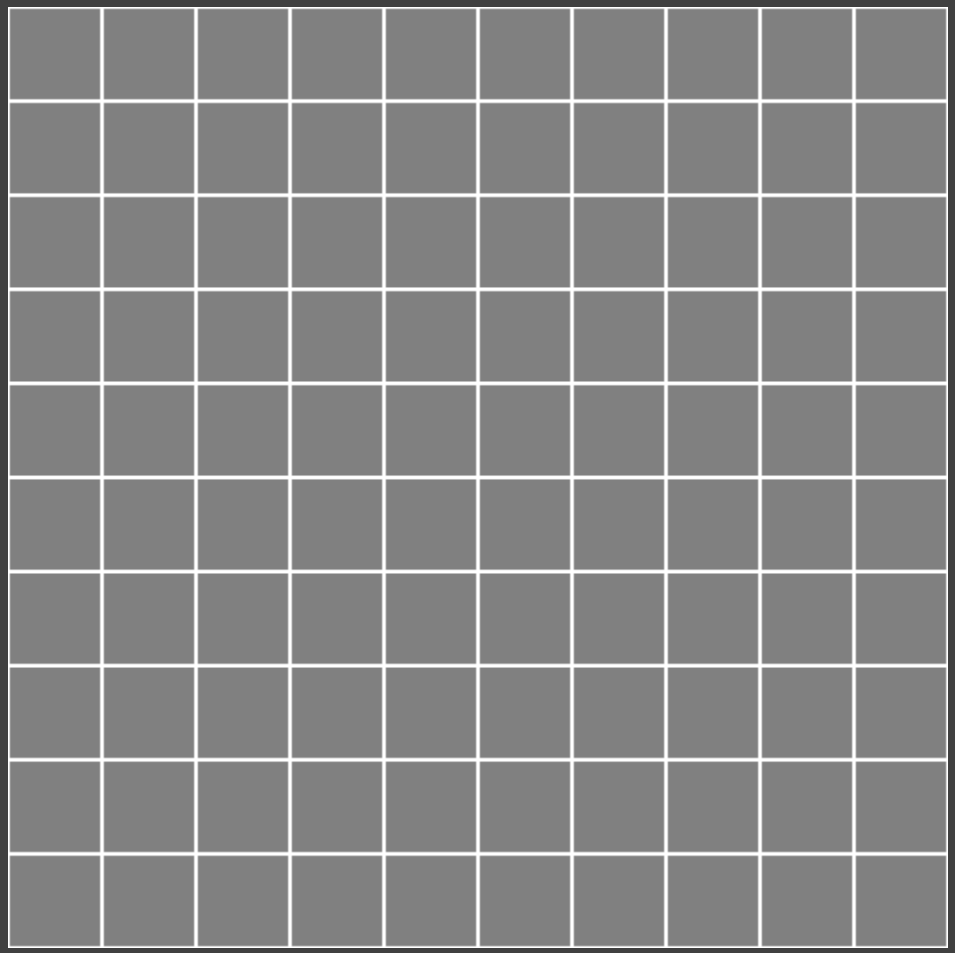
*poll of 100 people, margin of error +/-5

Polling Data

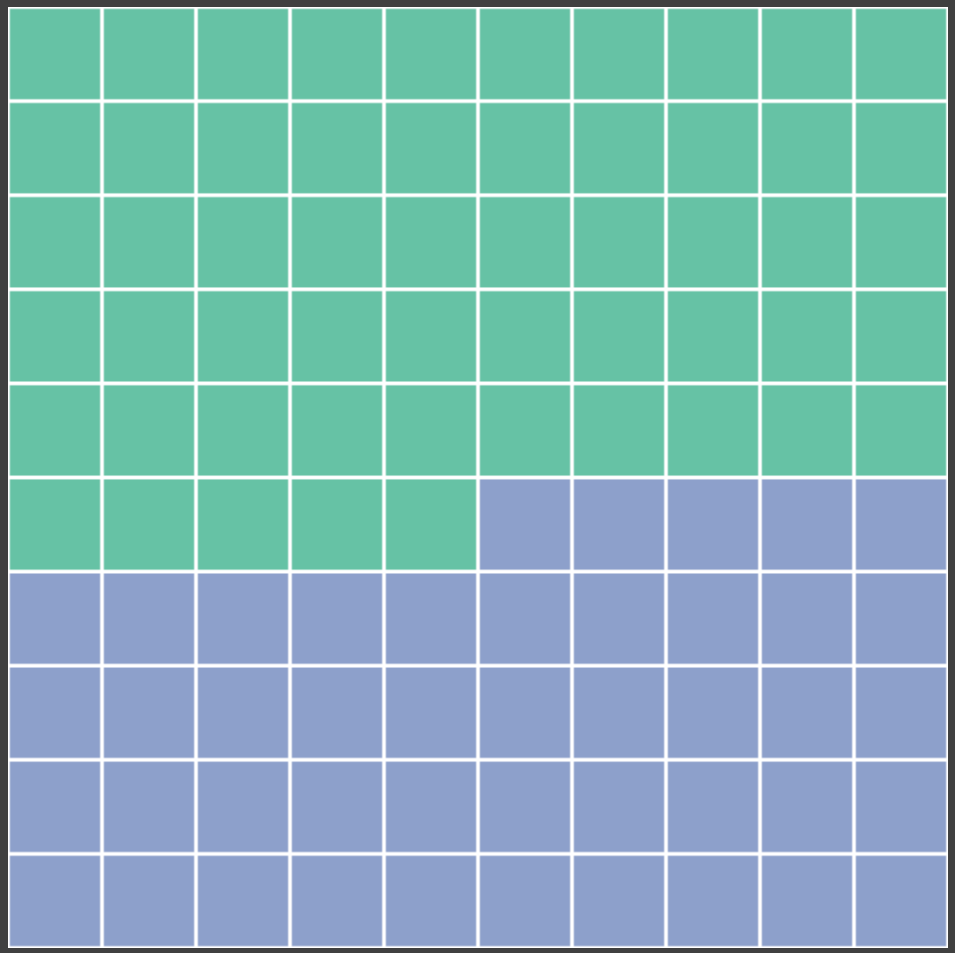
Candidate A is ahead of Candidate B in the polls, with 55% of the likely voters*

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

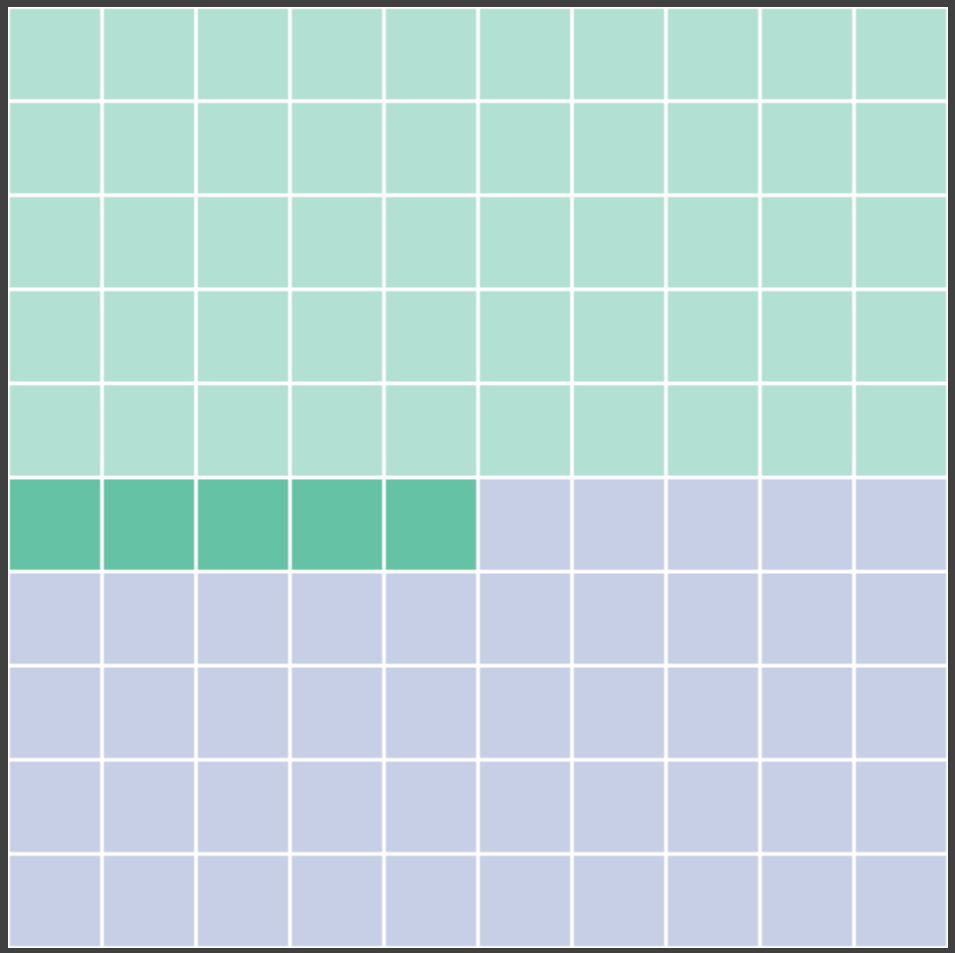

A Likely Voter



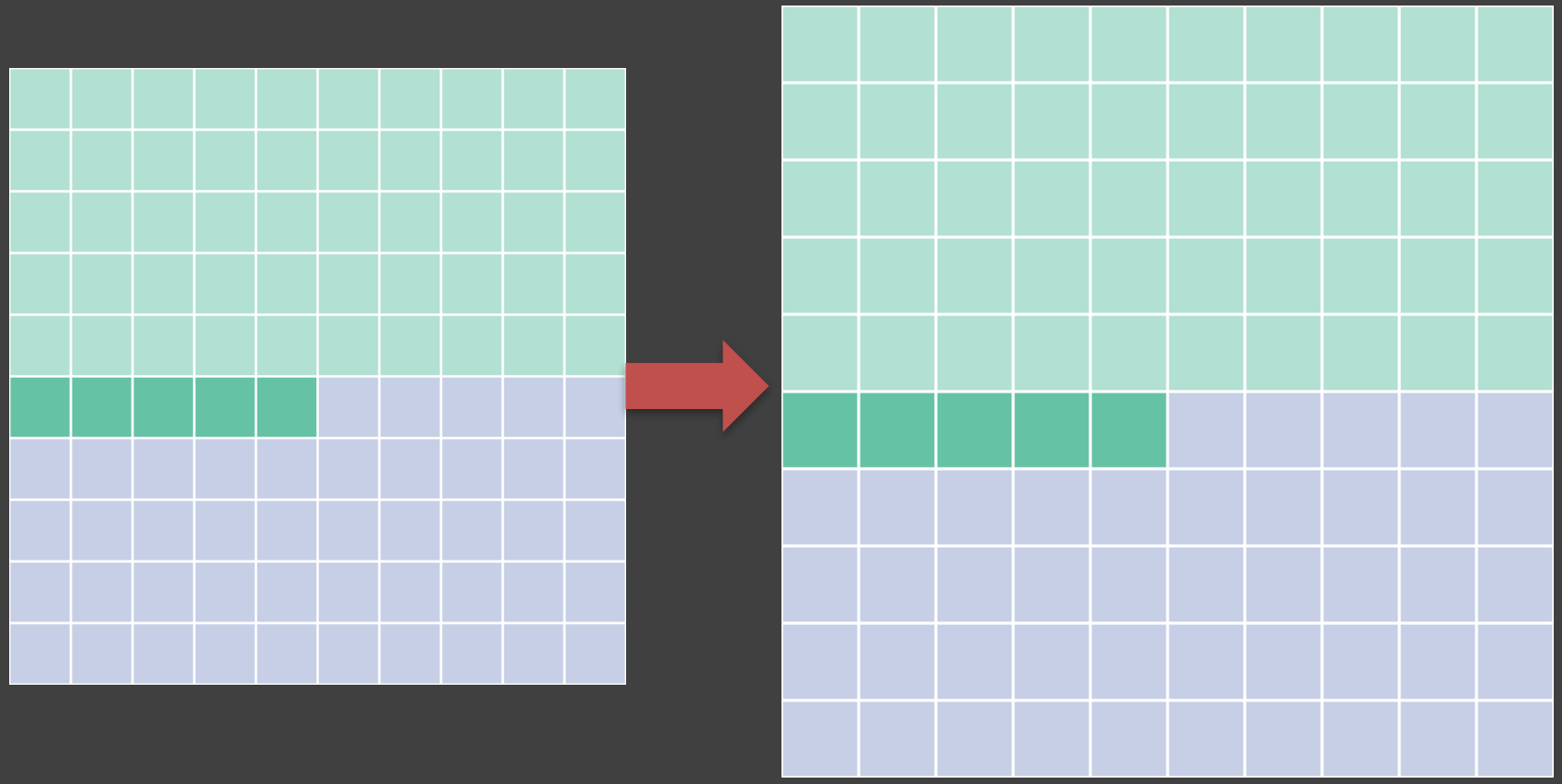
Poll



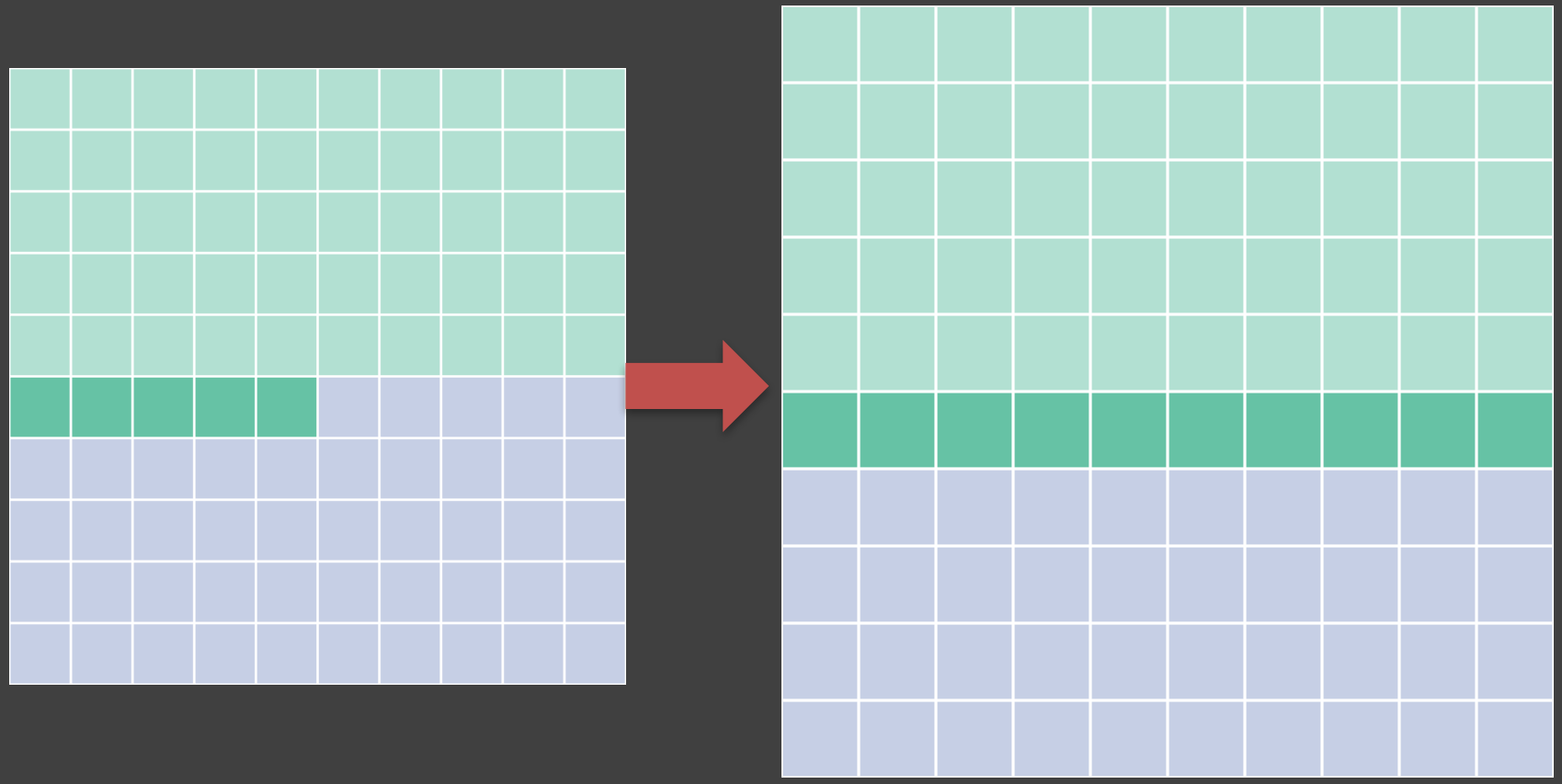
Poll



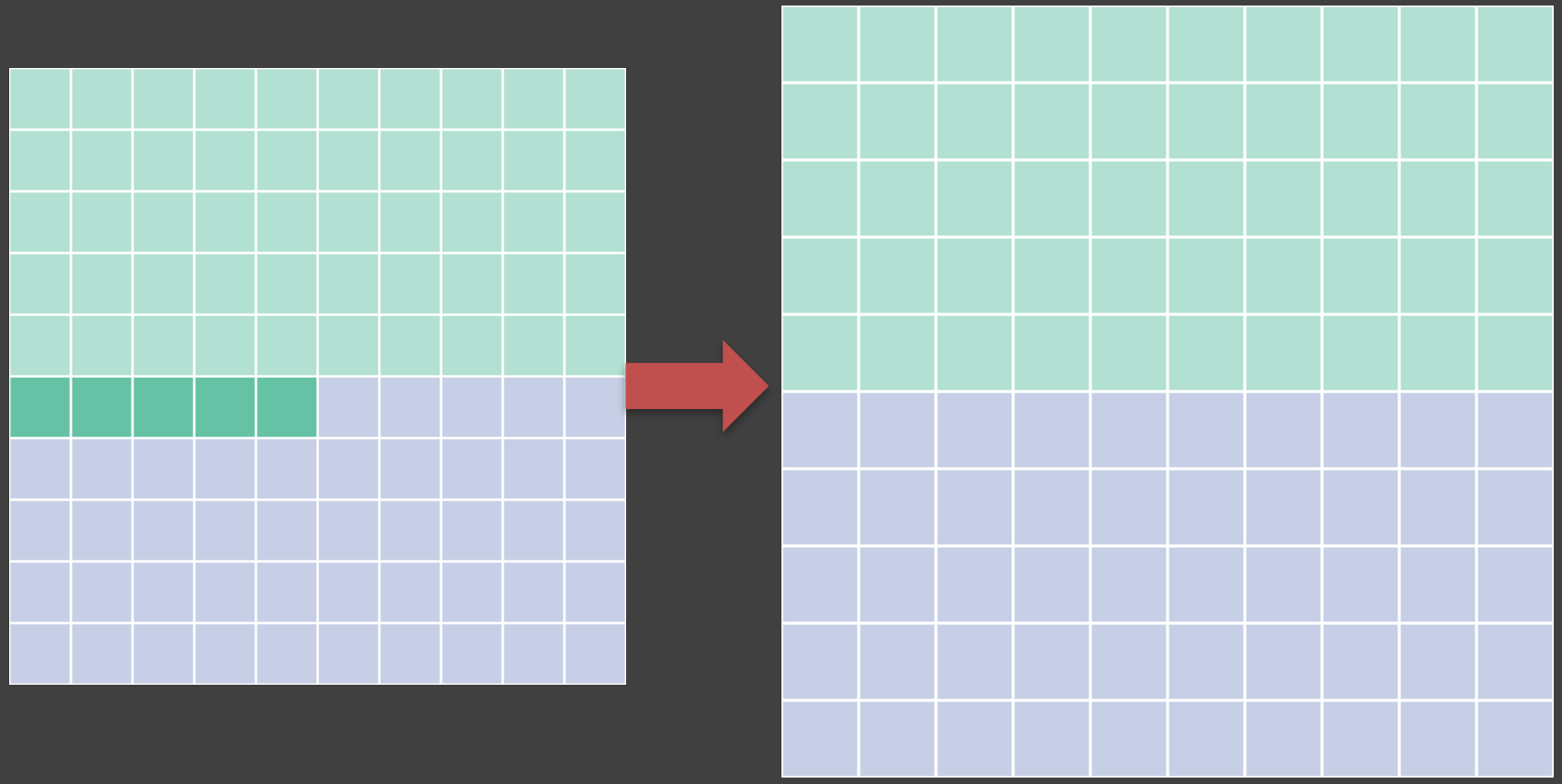
Actual Election?



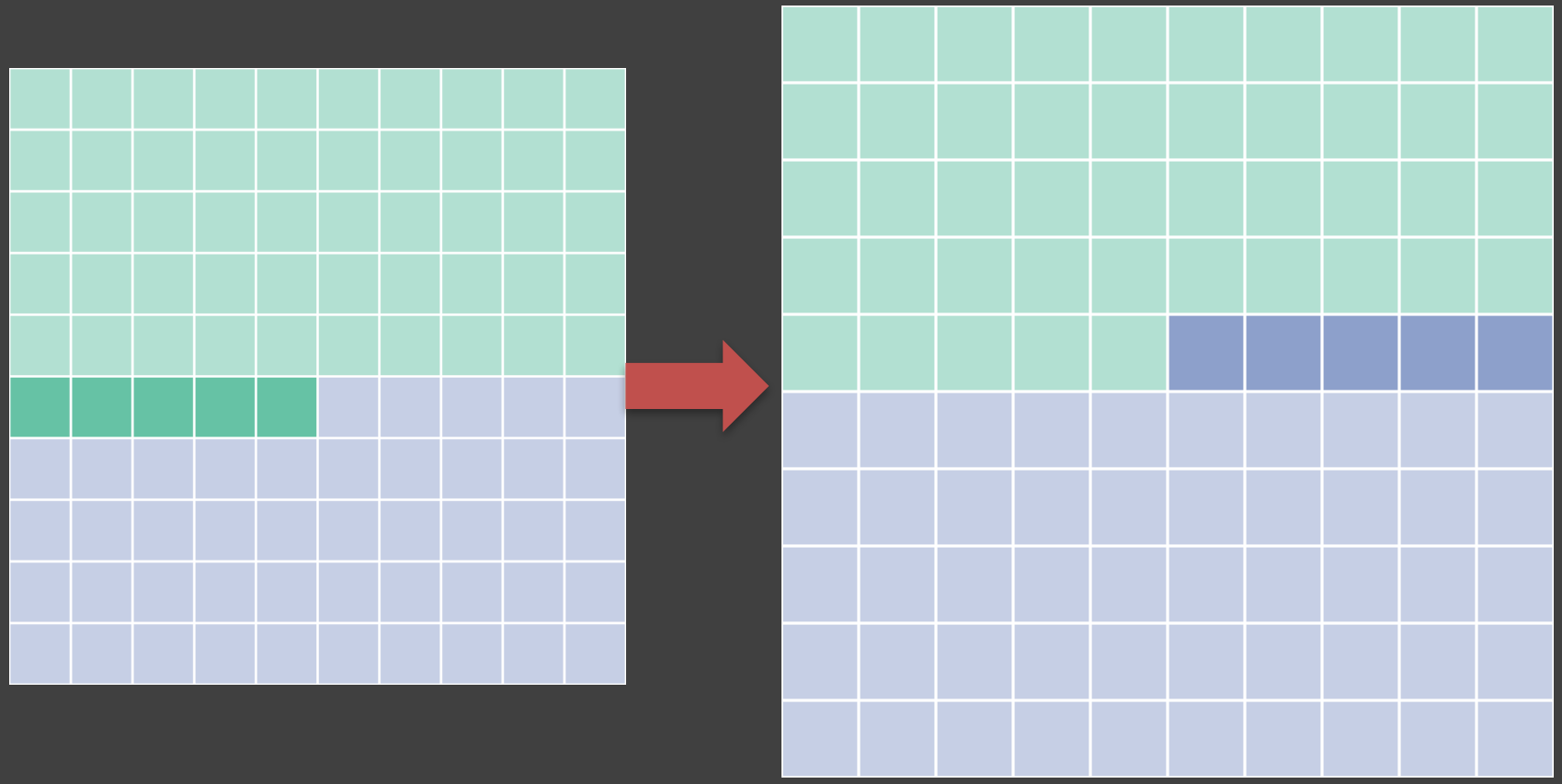
Actual Election?

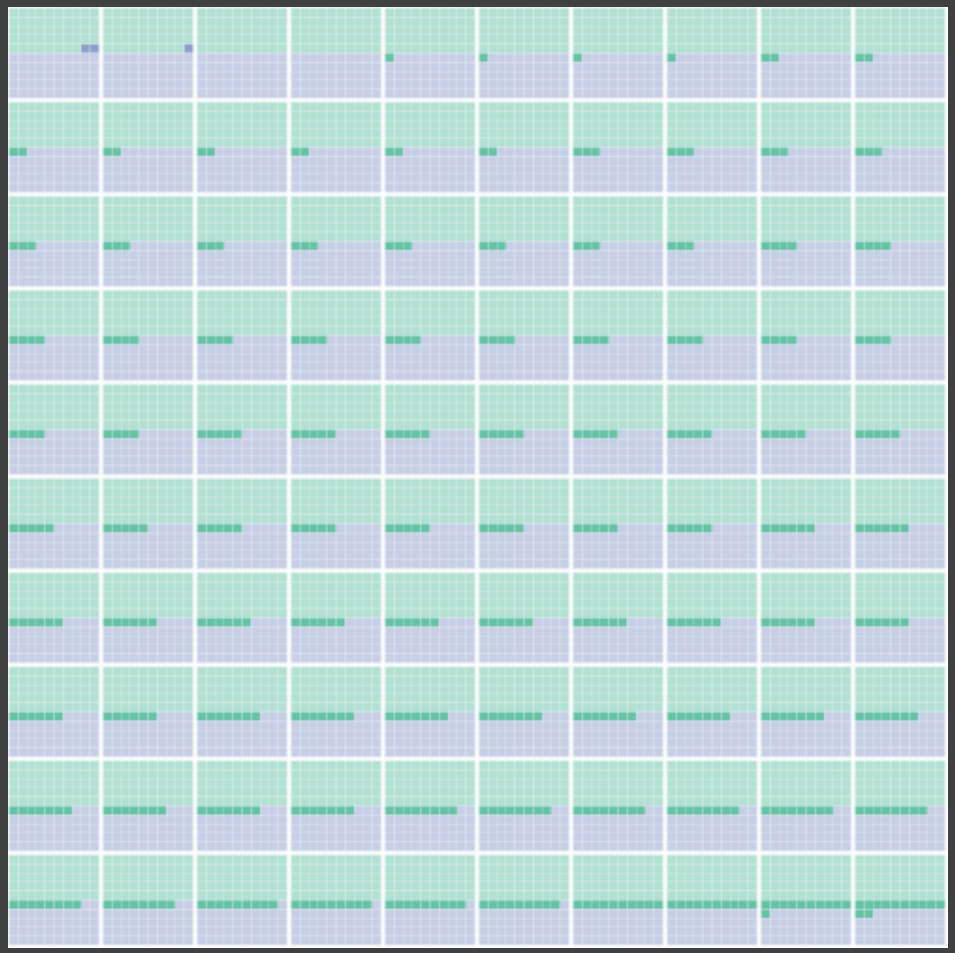


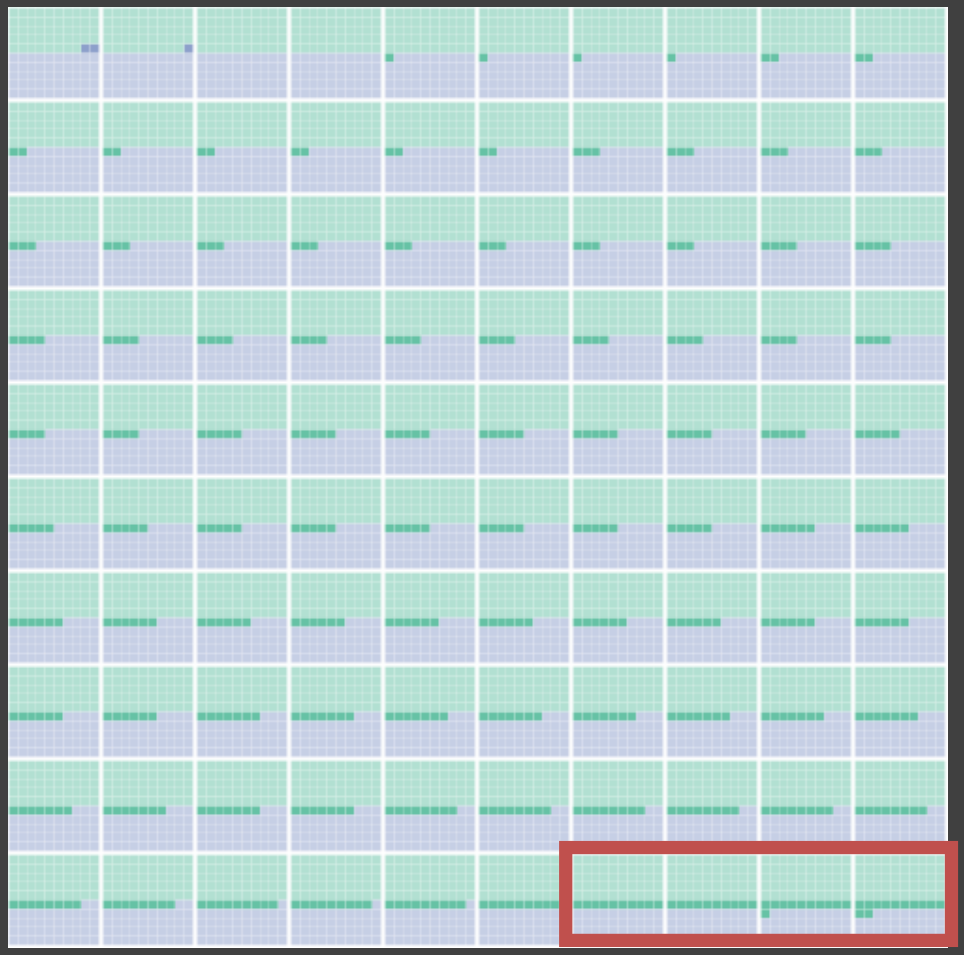
Actual Election?

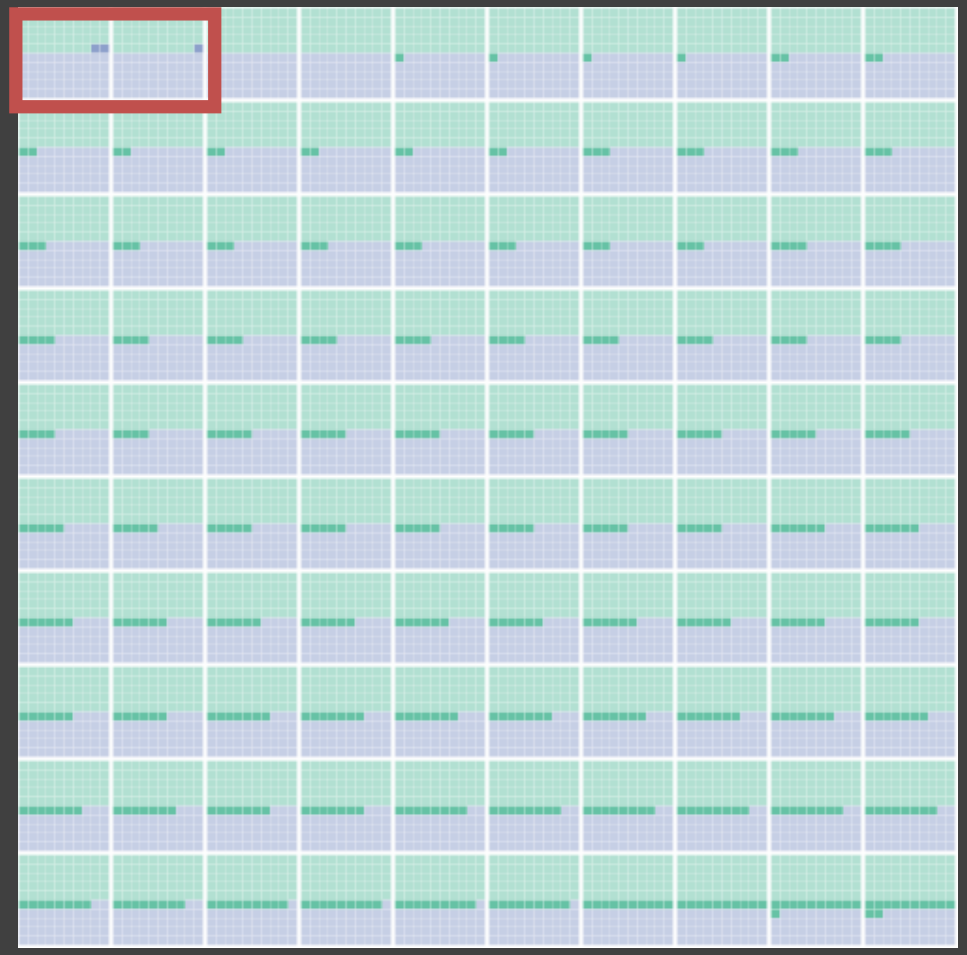


Actual Election?



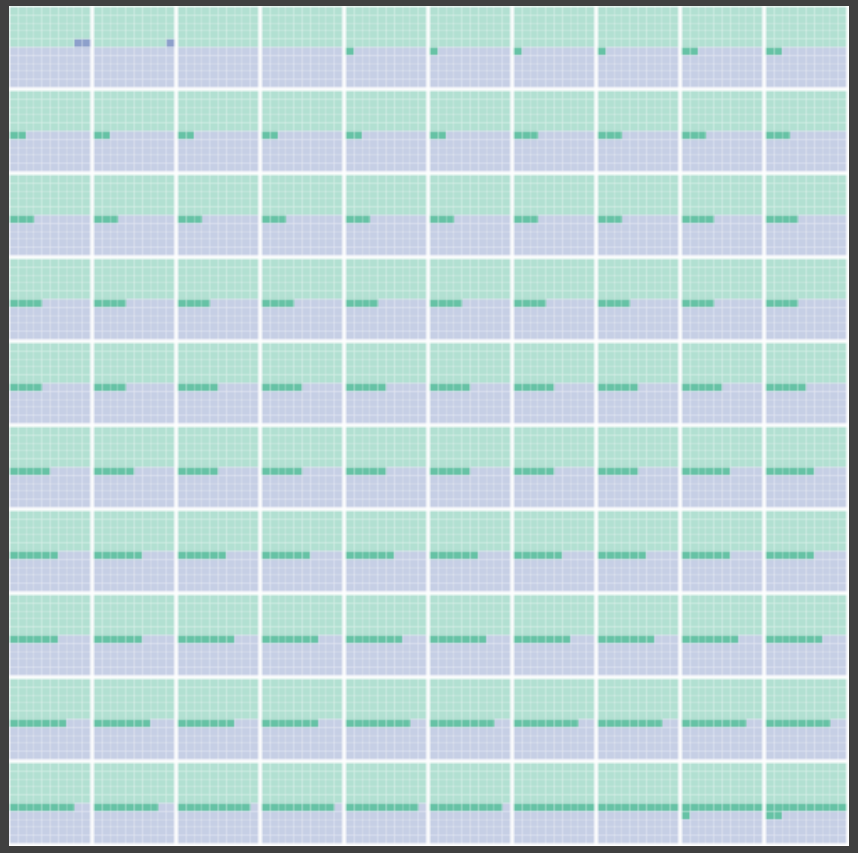






Pangloss Plot

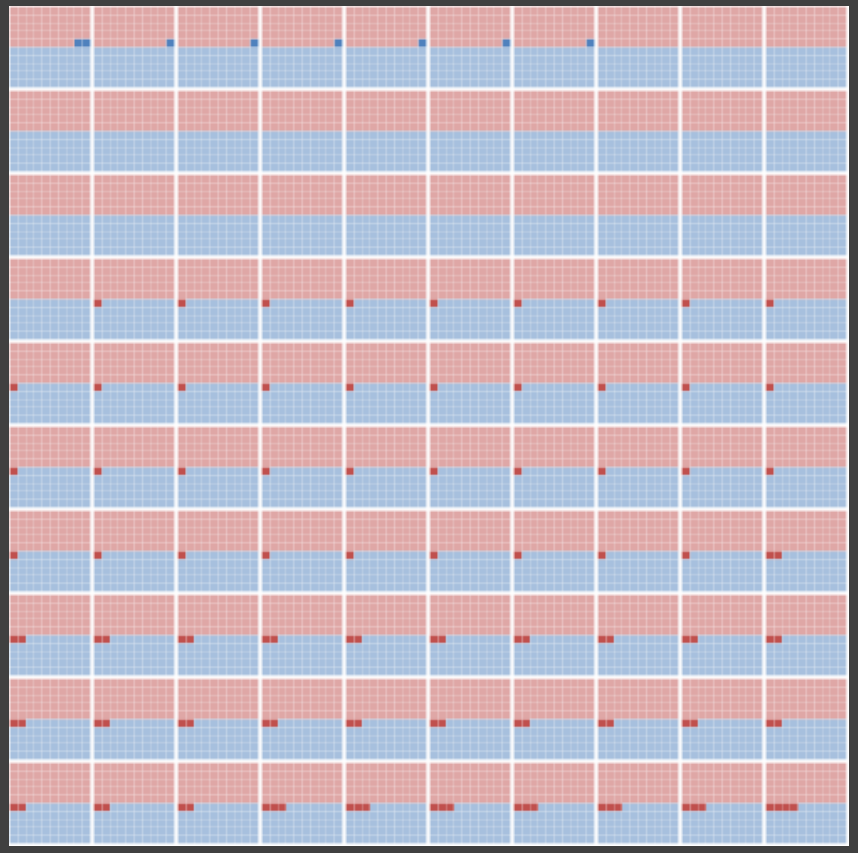
Candidate A is ahead of Candidate B in the polls, with 55% of the likely voters*



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

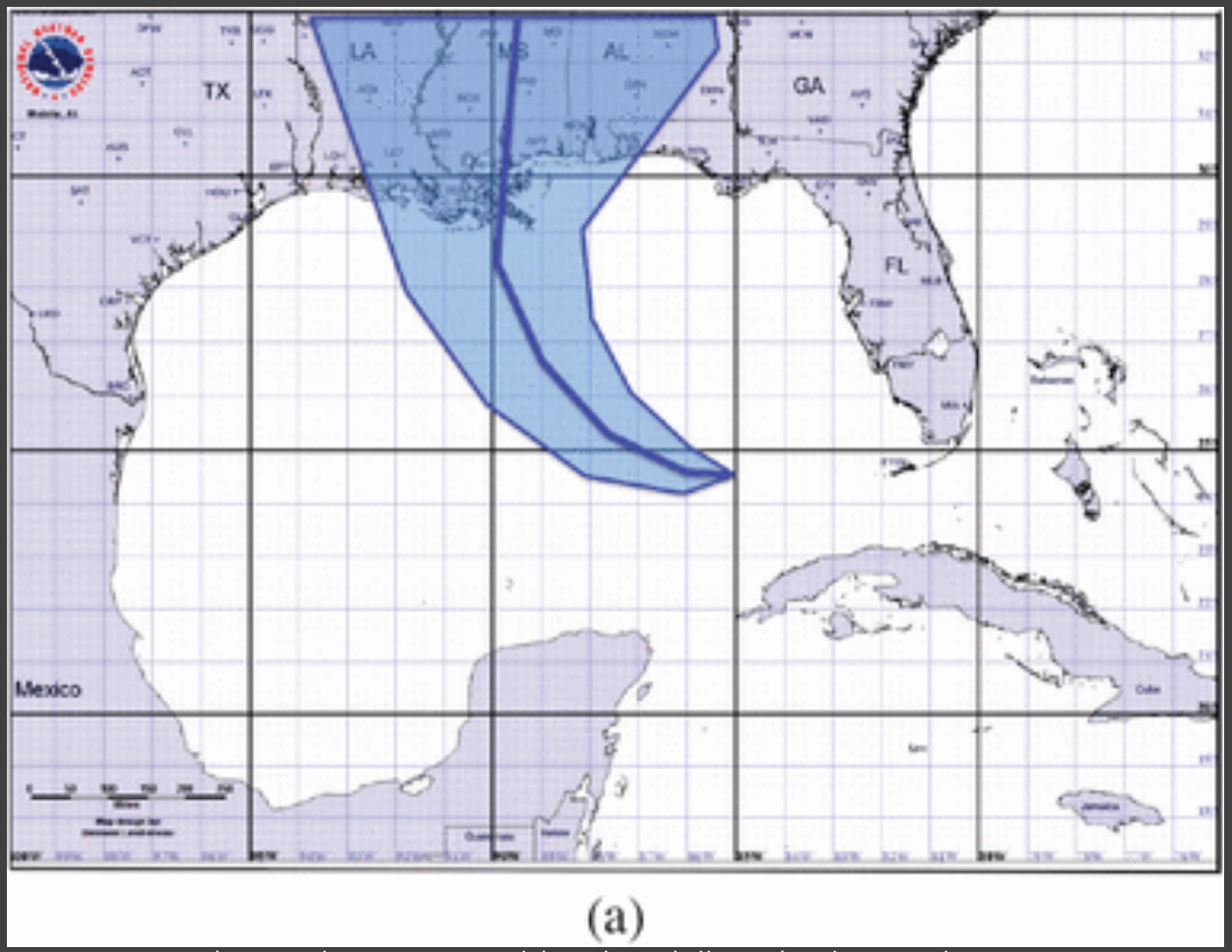
Pangloss Plot

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



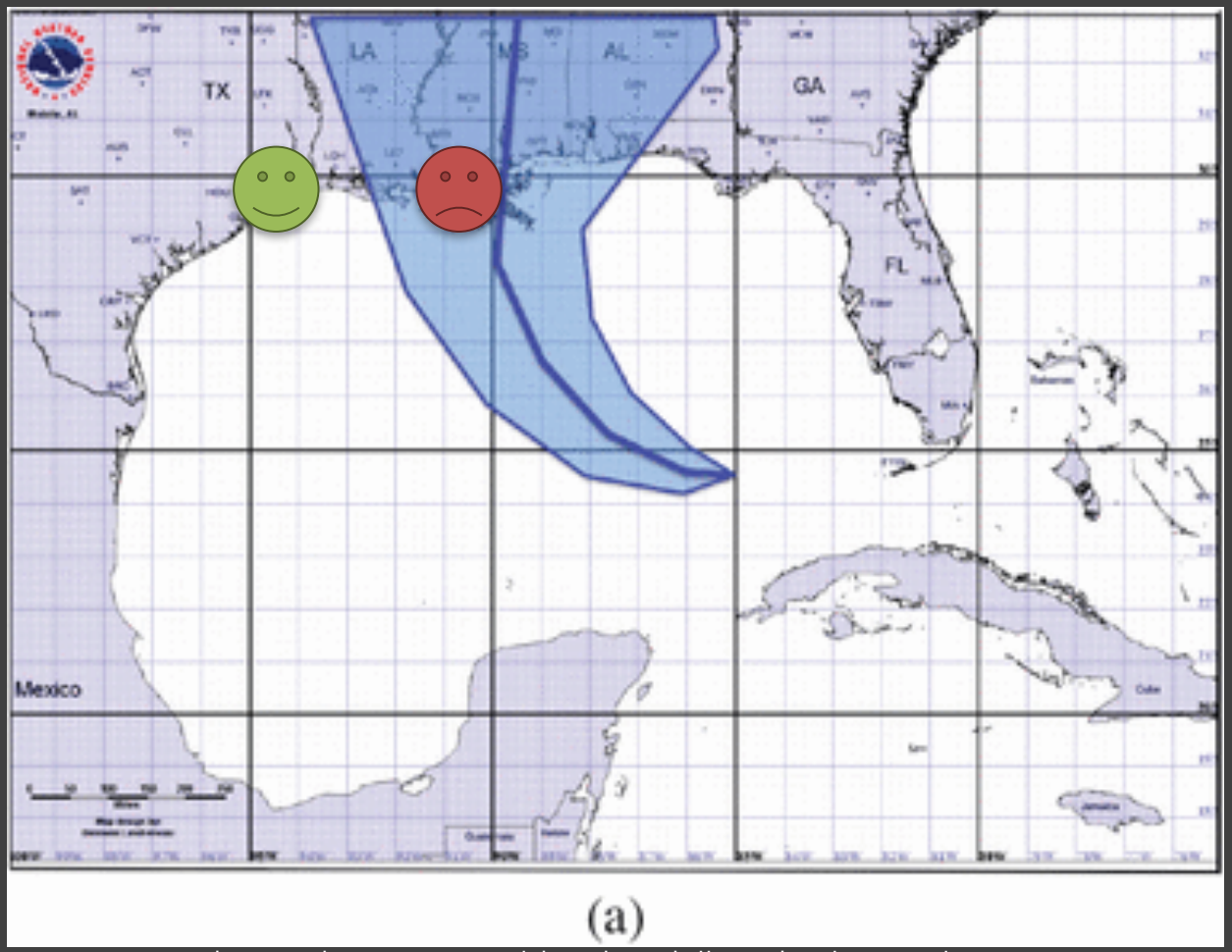
*poll of 3,117 people, margin of error +/-2

Model Visualization



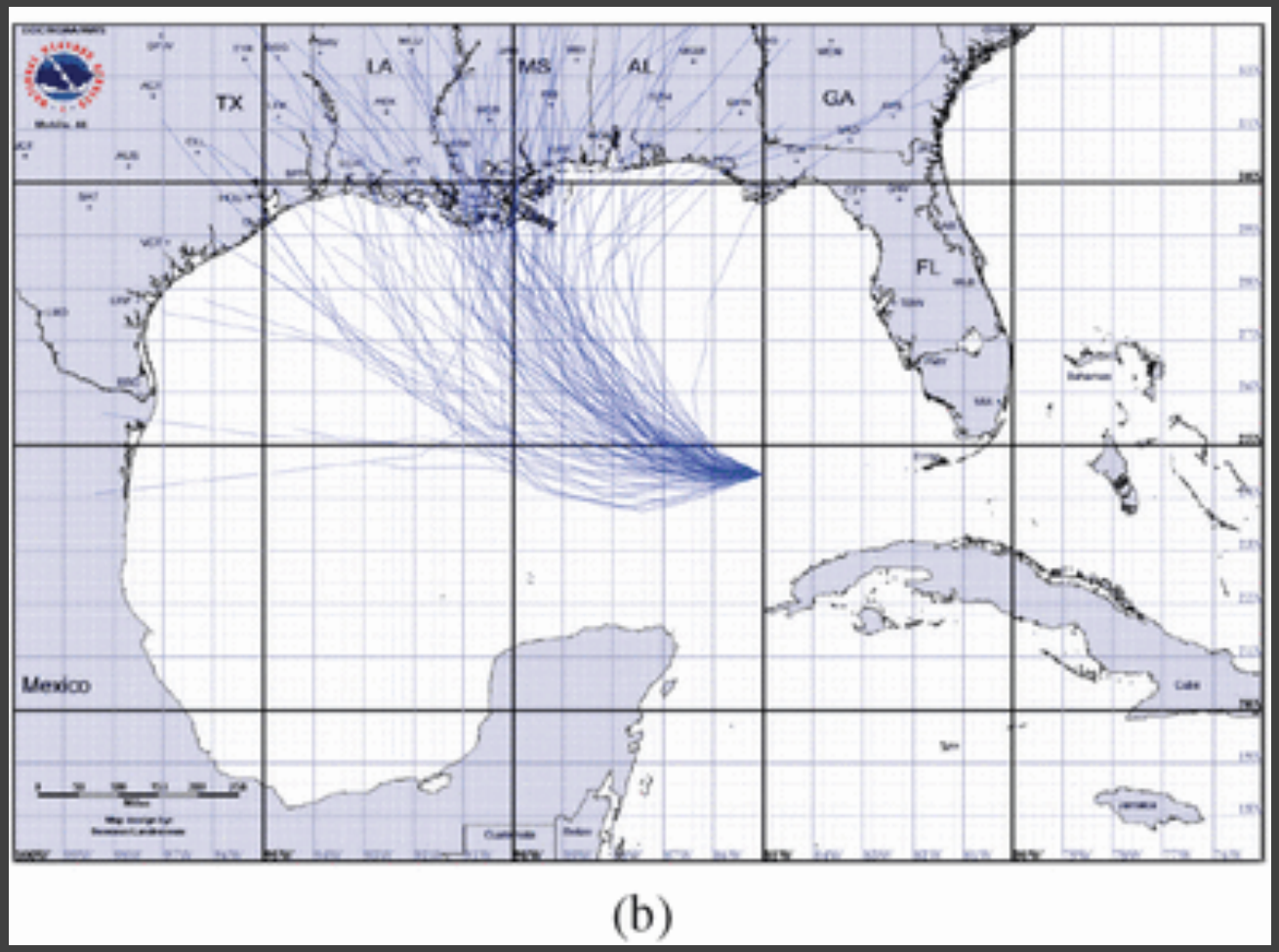
Cox, Jonathan and House, Donald and Lindell, Michael. Visuazlising uncertainty in predicted hurricane tracks. International Journal for Uncertainty Quantification, 2013.

Model Visualization



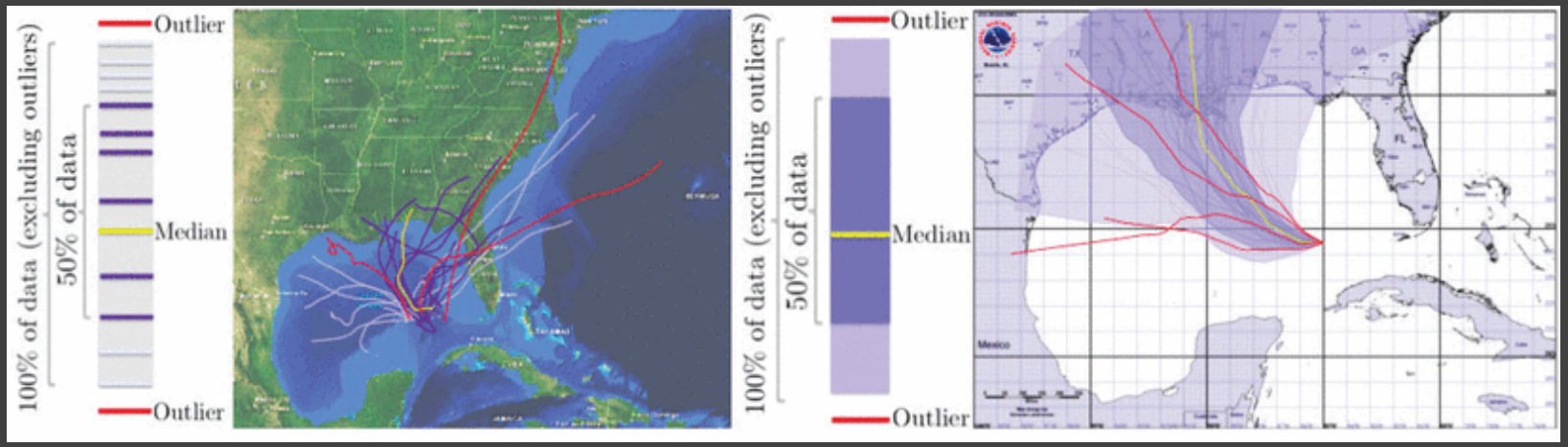
Cox, Jonathan and House, Donald and Lindell, Michael. Visuazlising uncertainty in predicted hurricane tracks. International Journal for Uncertainty Quantification, 2013.

Model Visualization



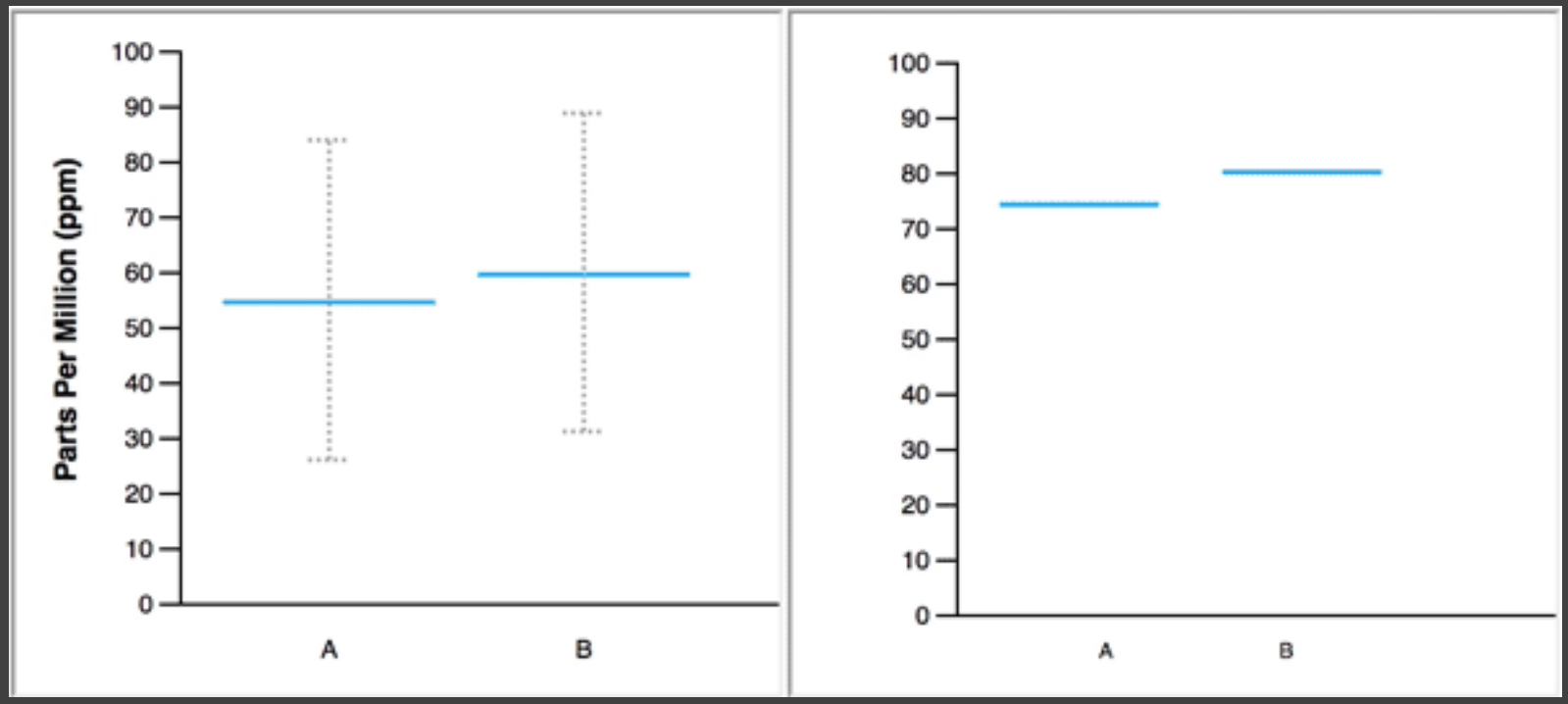
(b)

Model Visualization

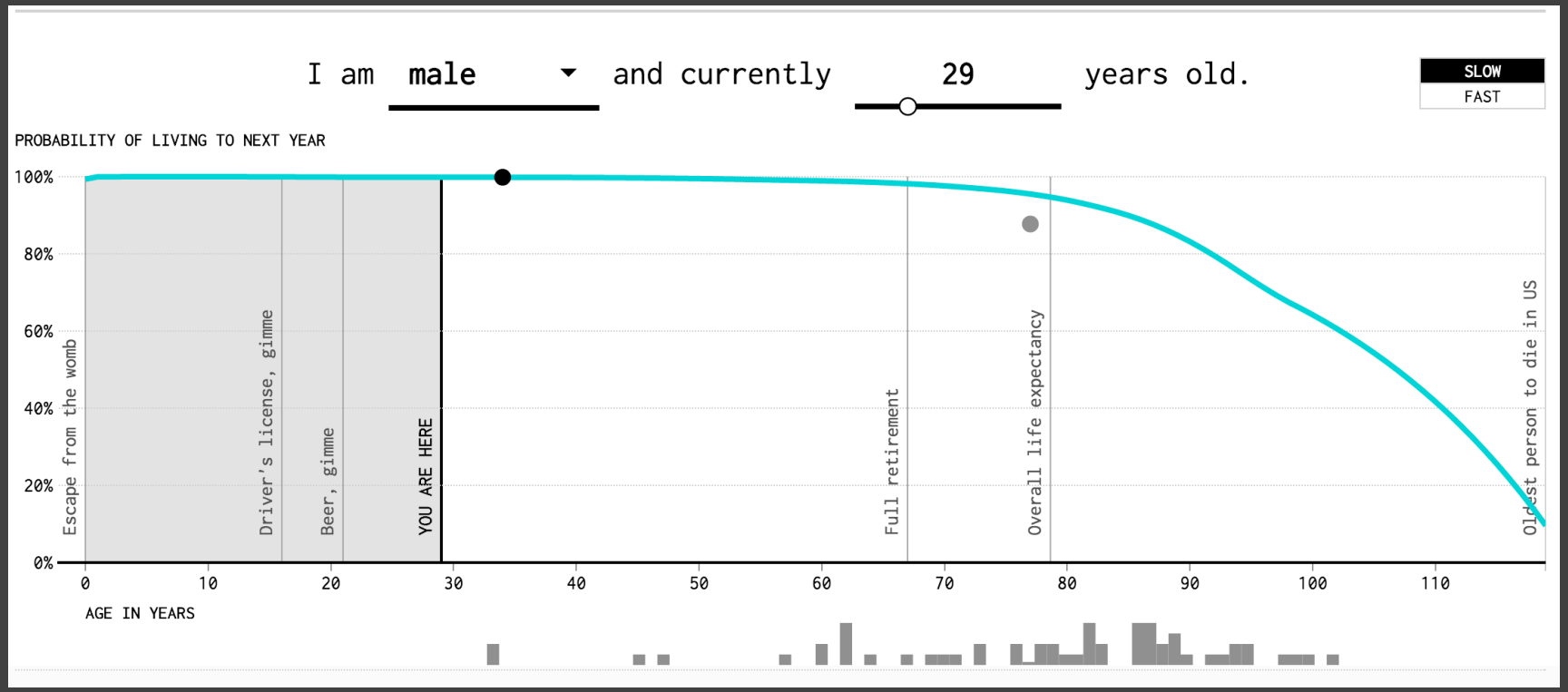


M. Mirzargar, R. Whitaker and R. Kirby. Curve Boxplot: Generalization of Boxplot for Ensembles of Curves. IEEE VIS 2014.

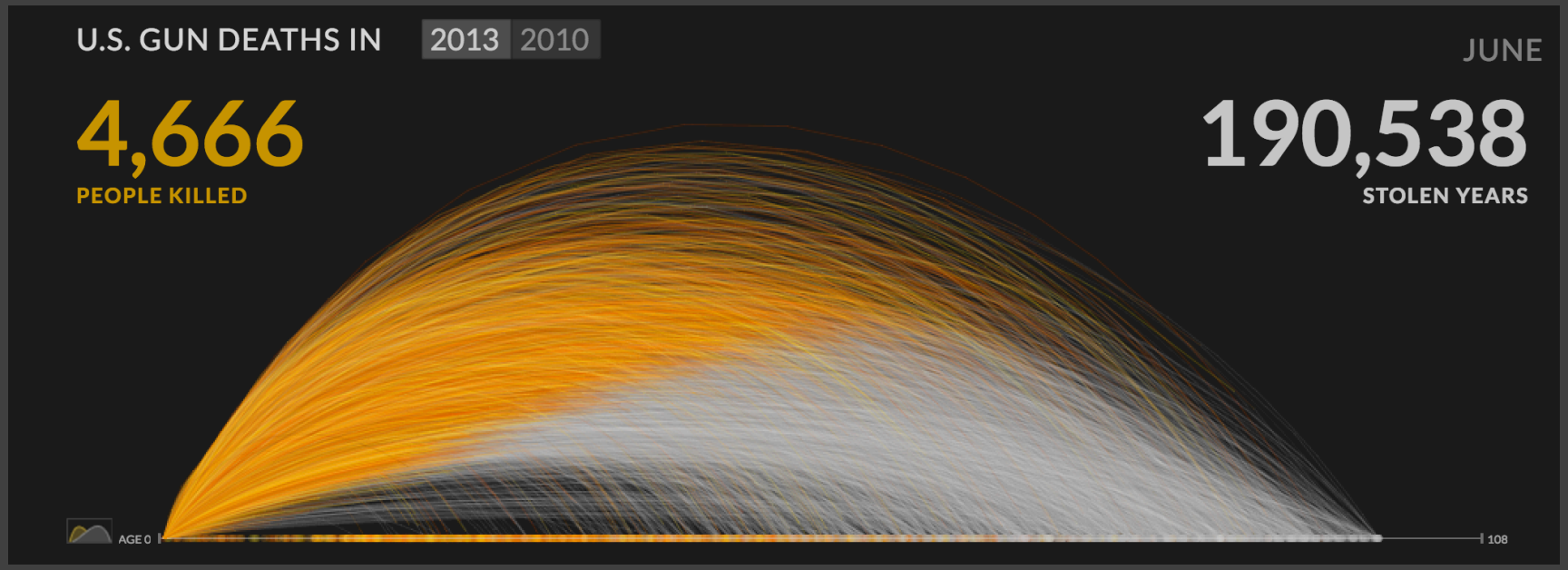
Hypothetical Outcome Plots



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

How Should I Visualize Uncertainty?

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

How Should I Visualize Uncertainty?

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

IT DEPENDS

Cognitive and Perceptual Biases and Disfluencies

WHAT CAN GO WRONG WHEN VISUALIZING UNCERTAINTY?

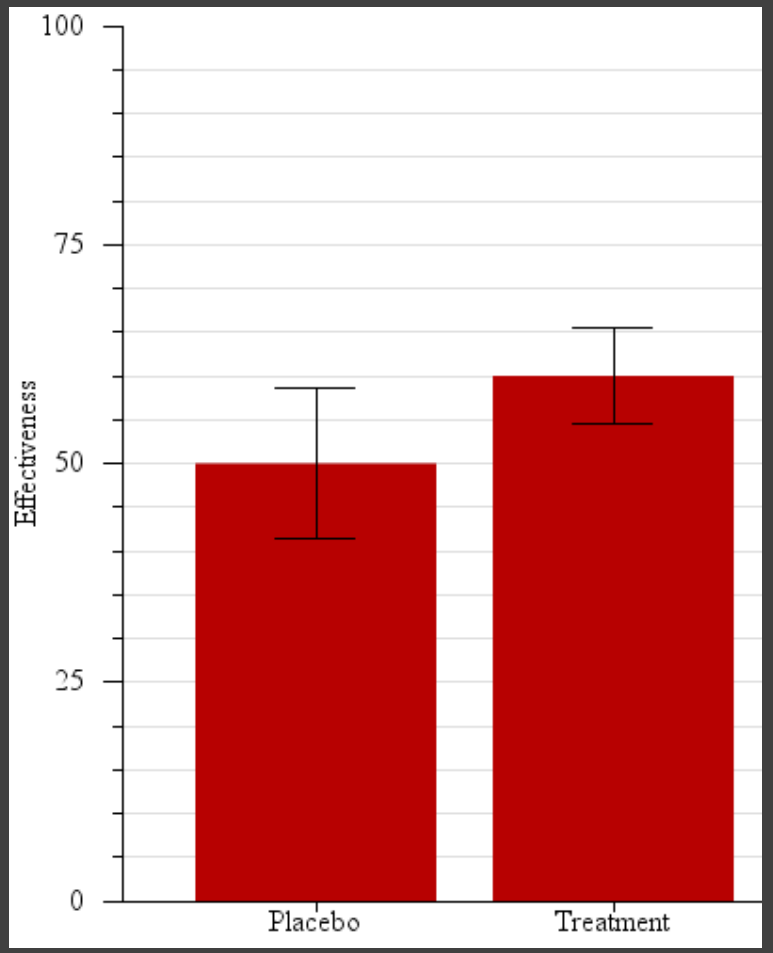
p-value

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

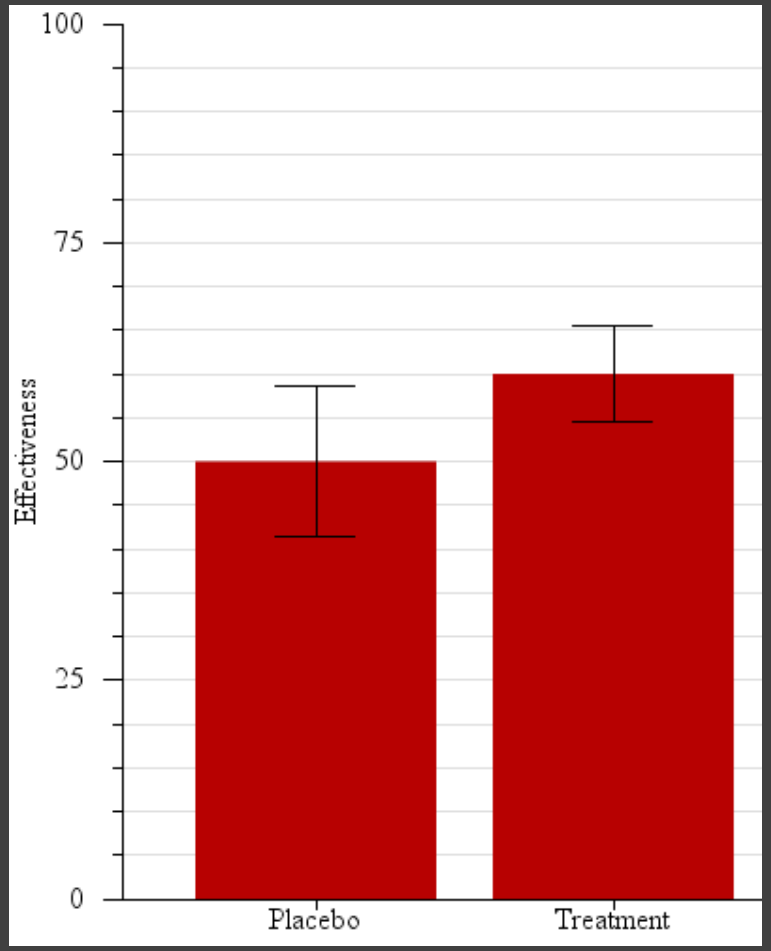
T-Confidence Interval?

Z-Confidence Interval?

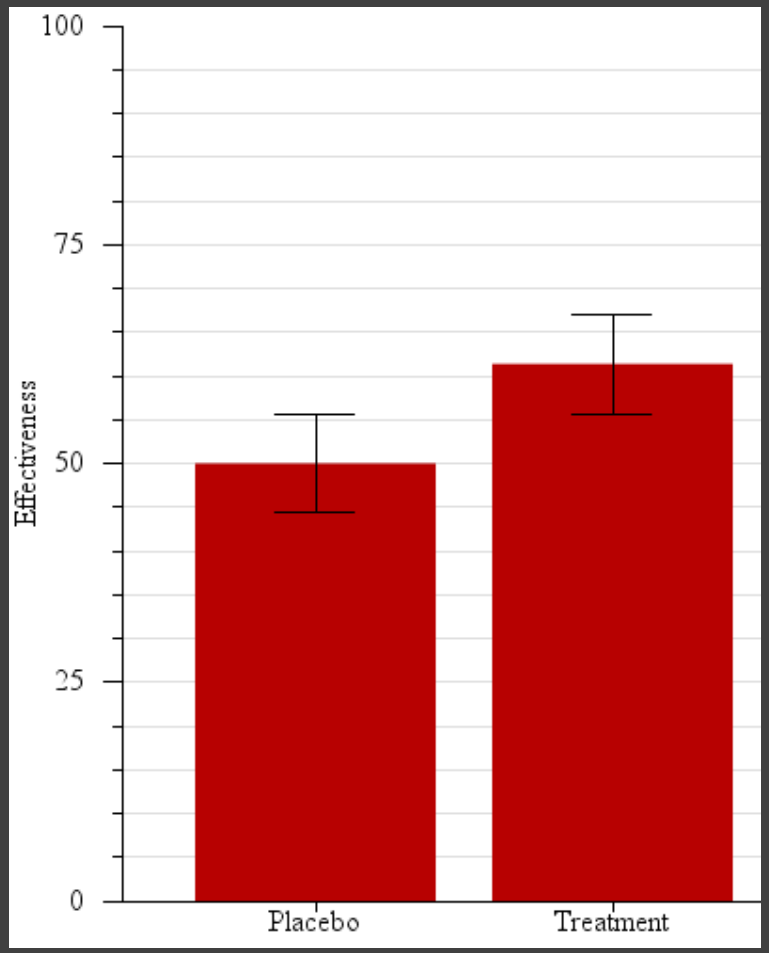
Bootstrapped Interval?

Min/Max?

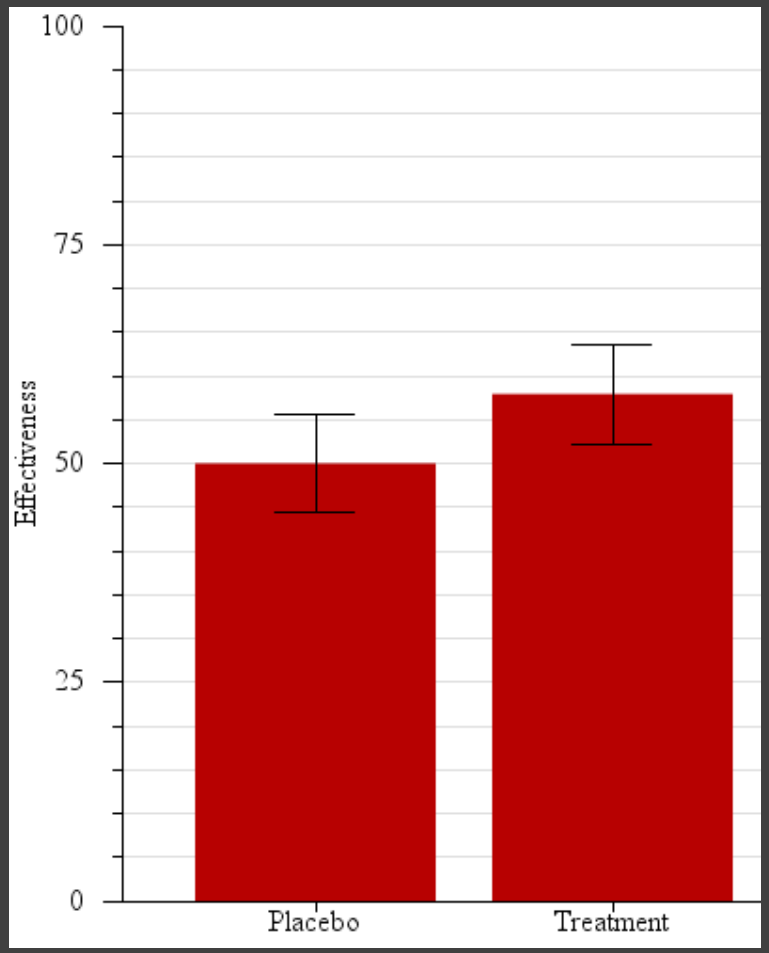
1.5*IQR (Q3-Q1)?



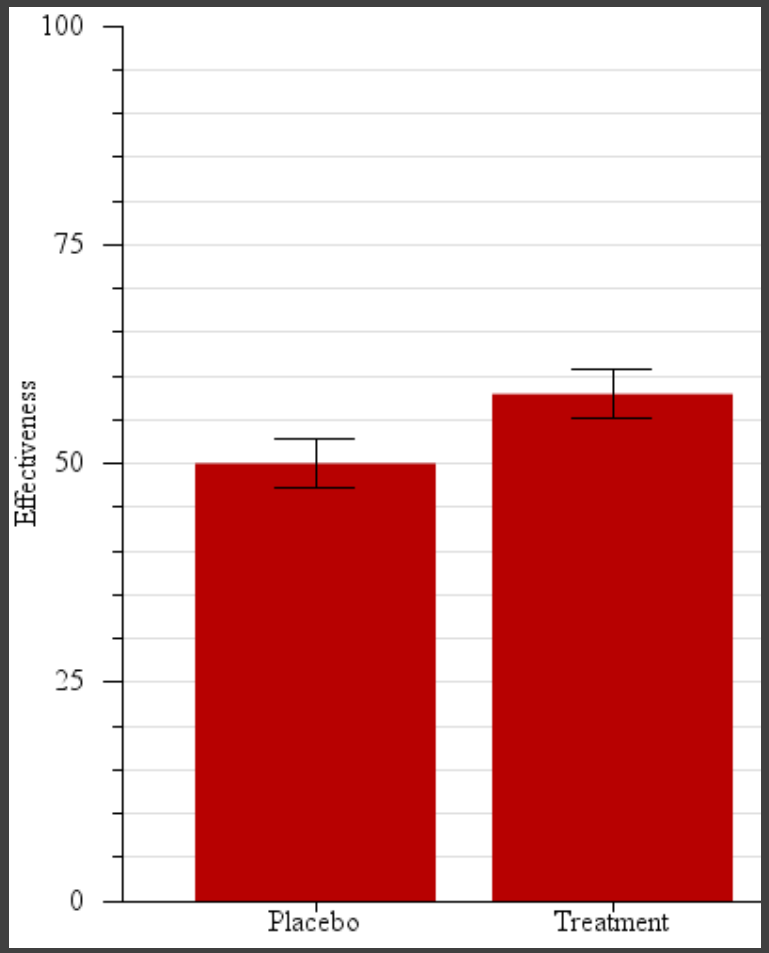
Guess the p-value



Guess the p-value

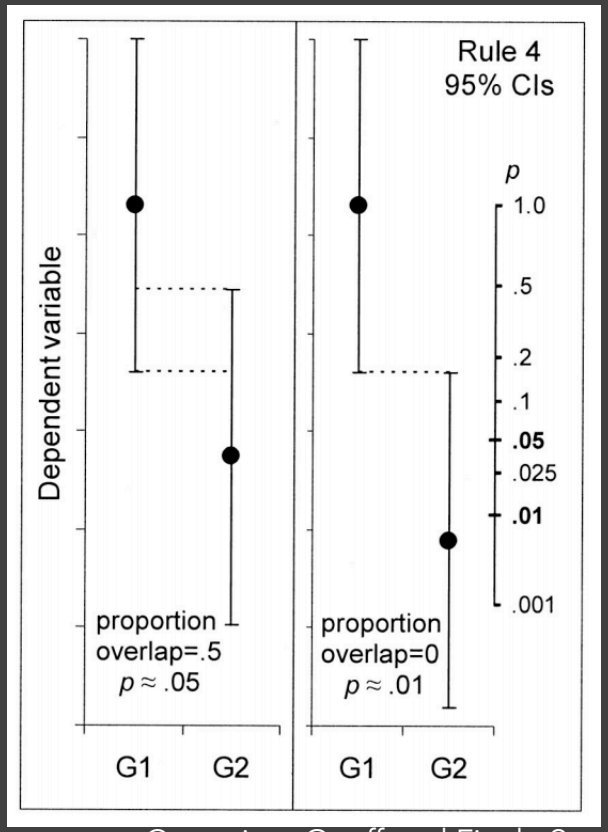


Guess the p-value

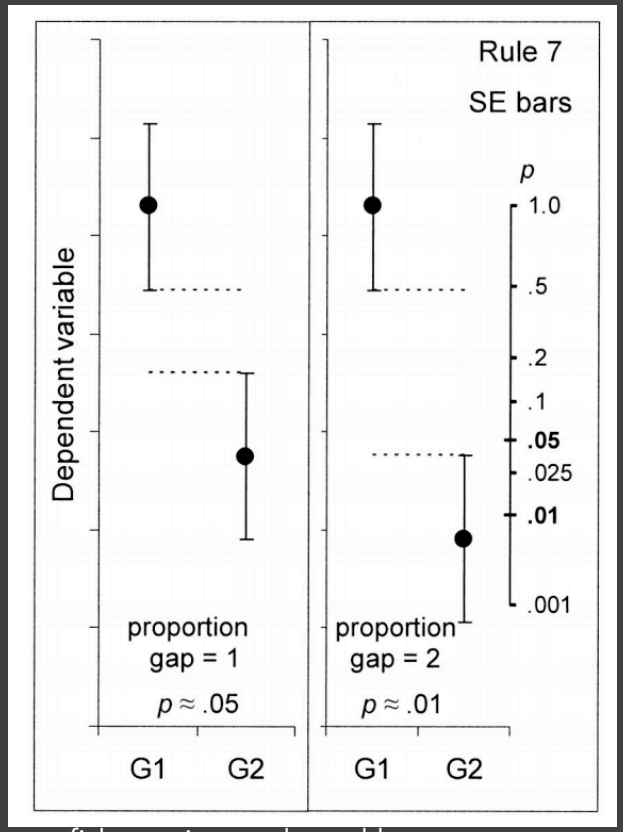


Inference by Eye

95% CIs

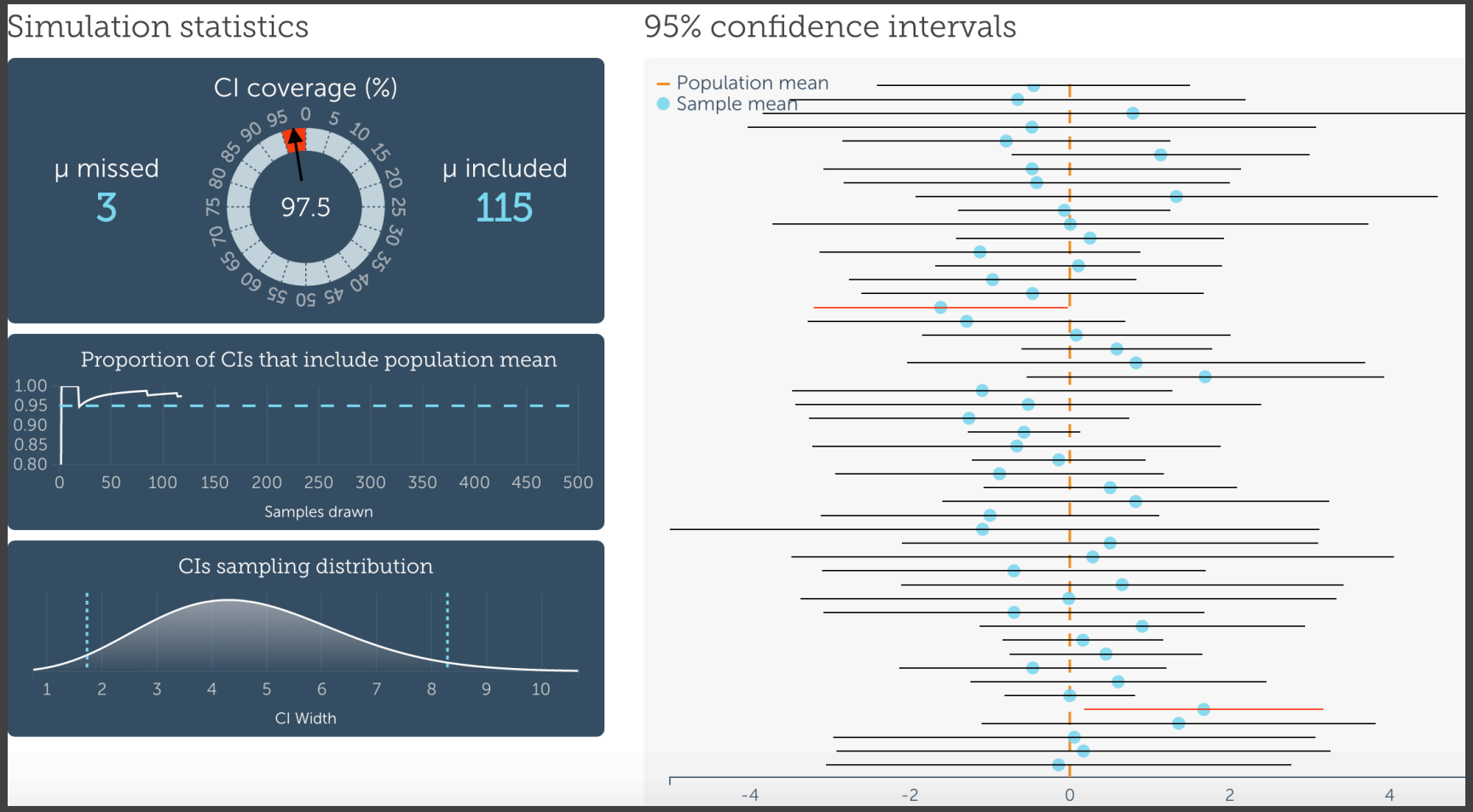


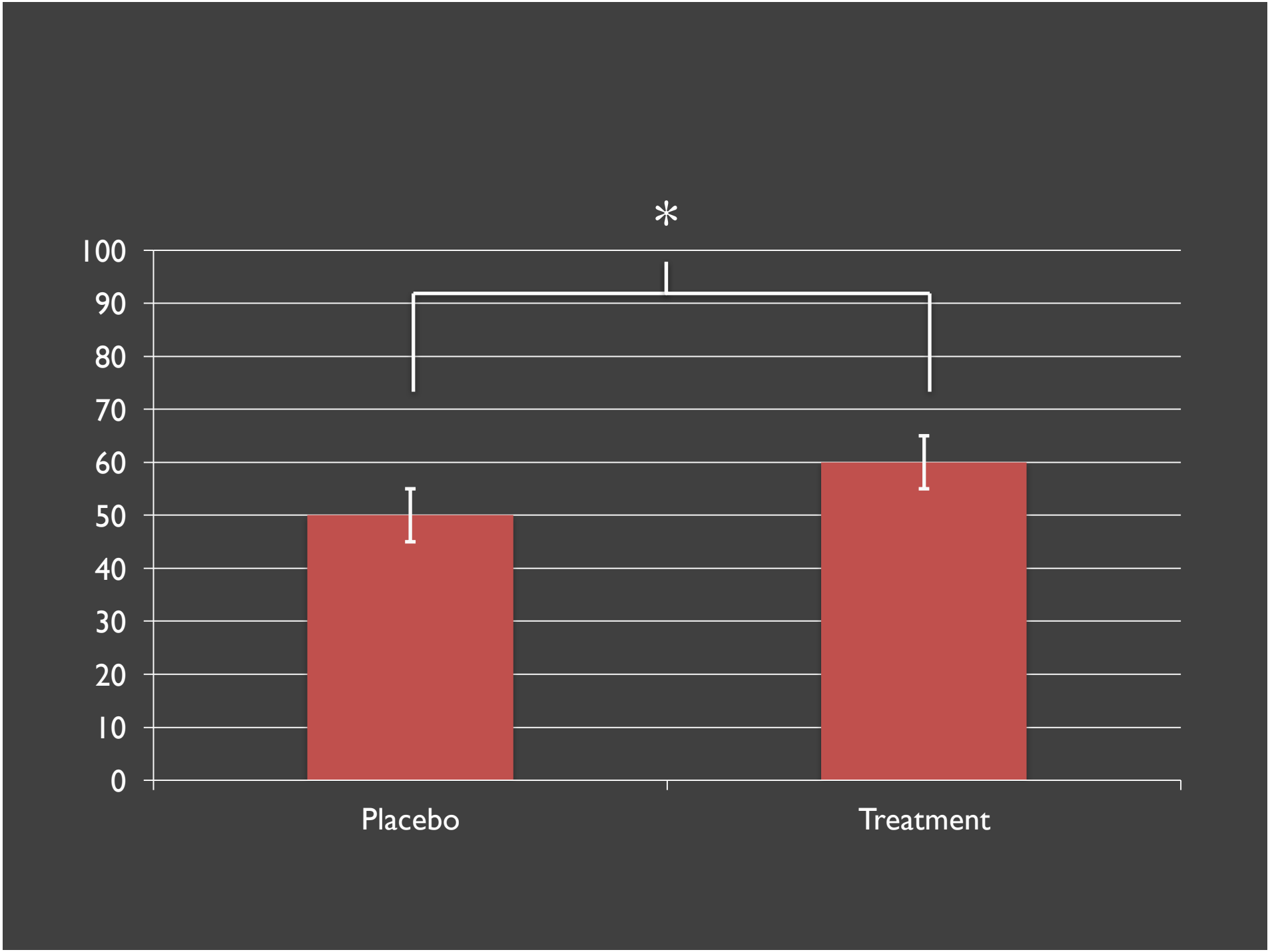
Standard Error



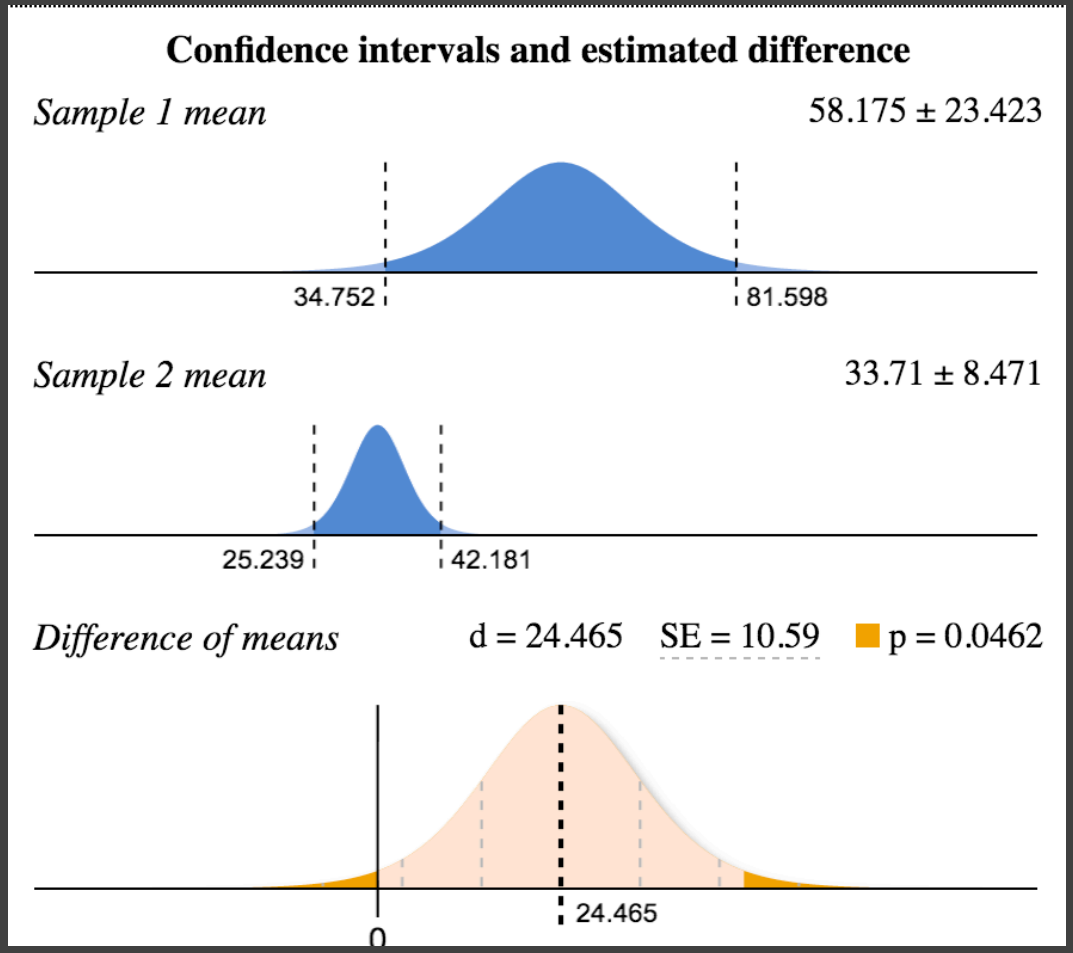
Cumming, Geoff and Finch, Sue. Inference by eye: confidence intervals and how to read pictures of data. American Psychologist, 2005.

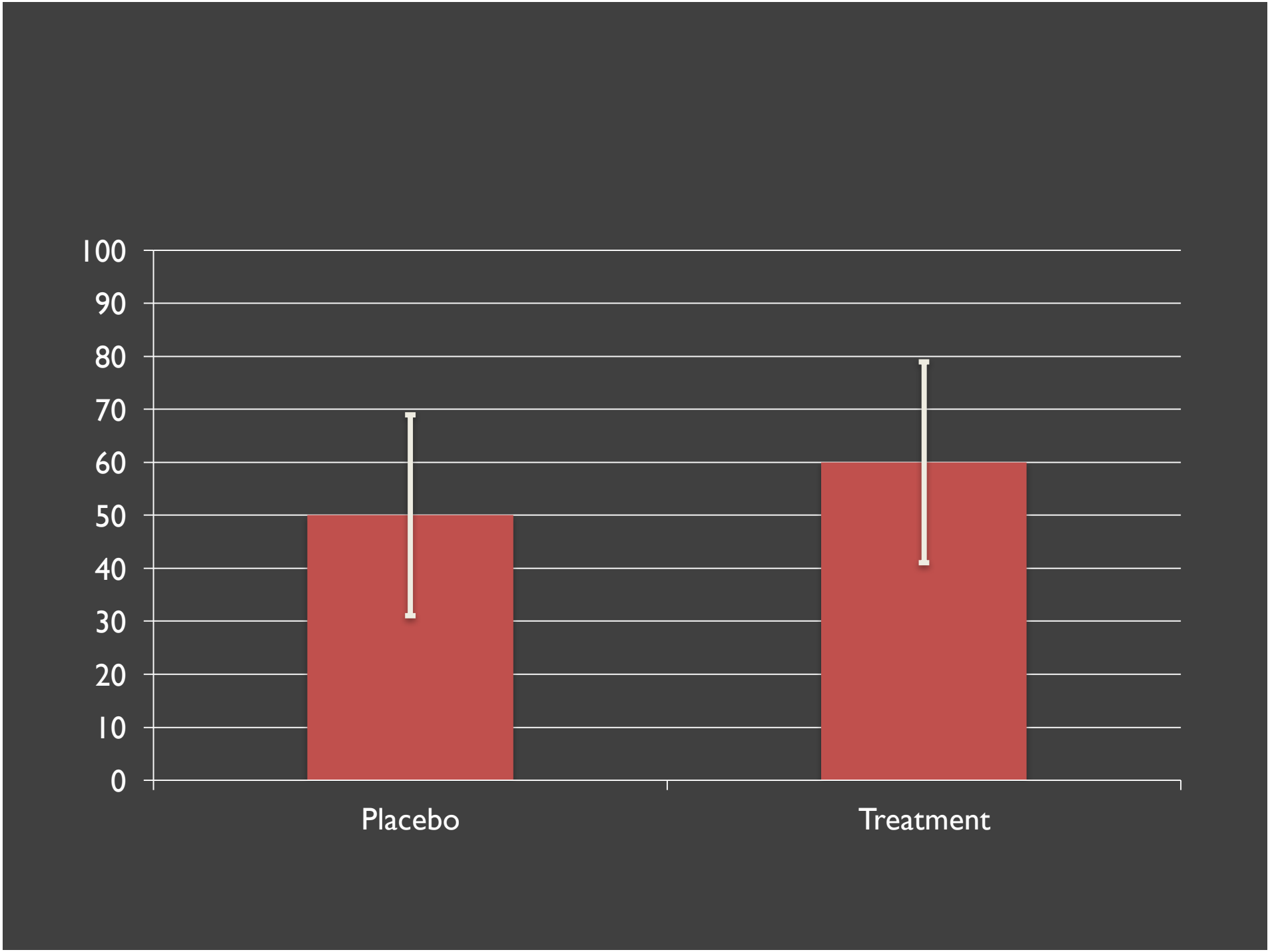
Confidence Intervals

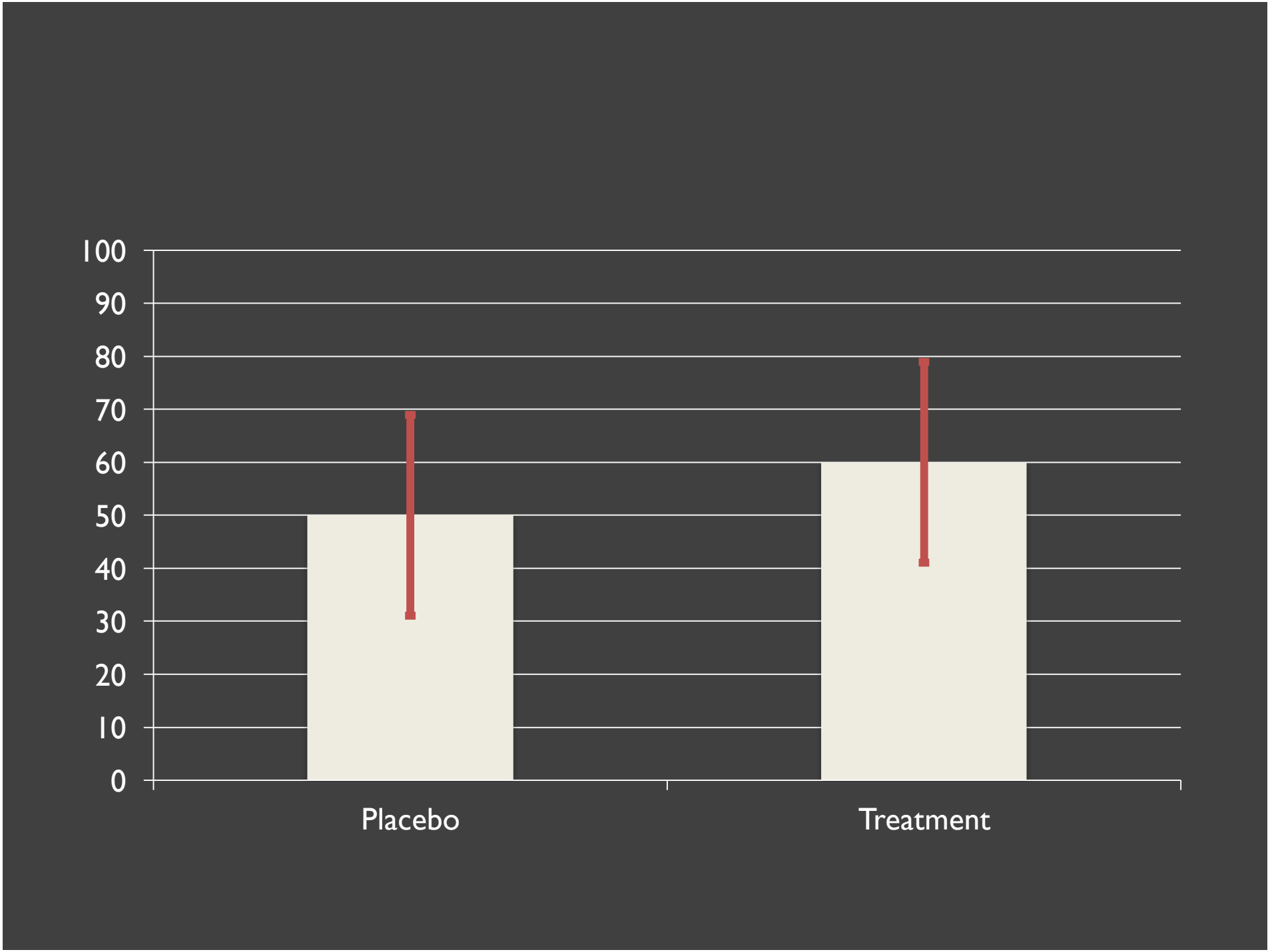



*
100
90
80
70
60
50
40
30
20
10
0
Placebo
Treatment

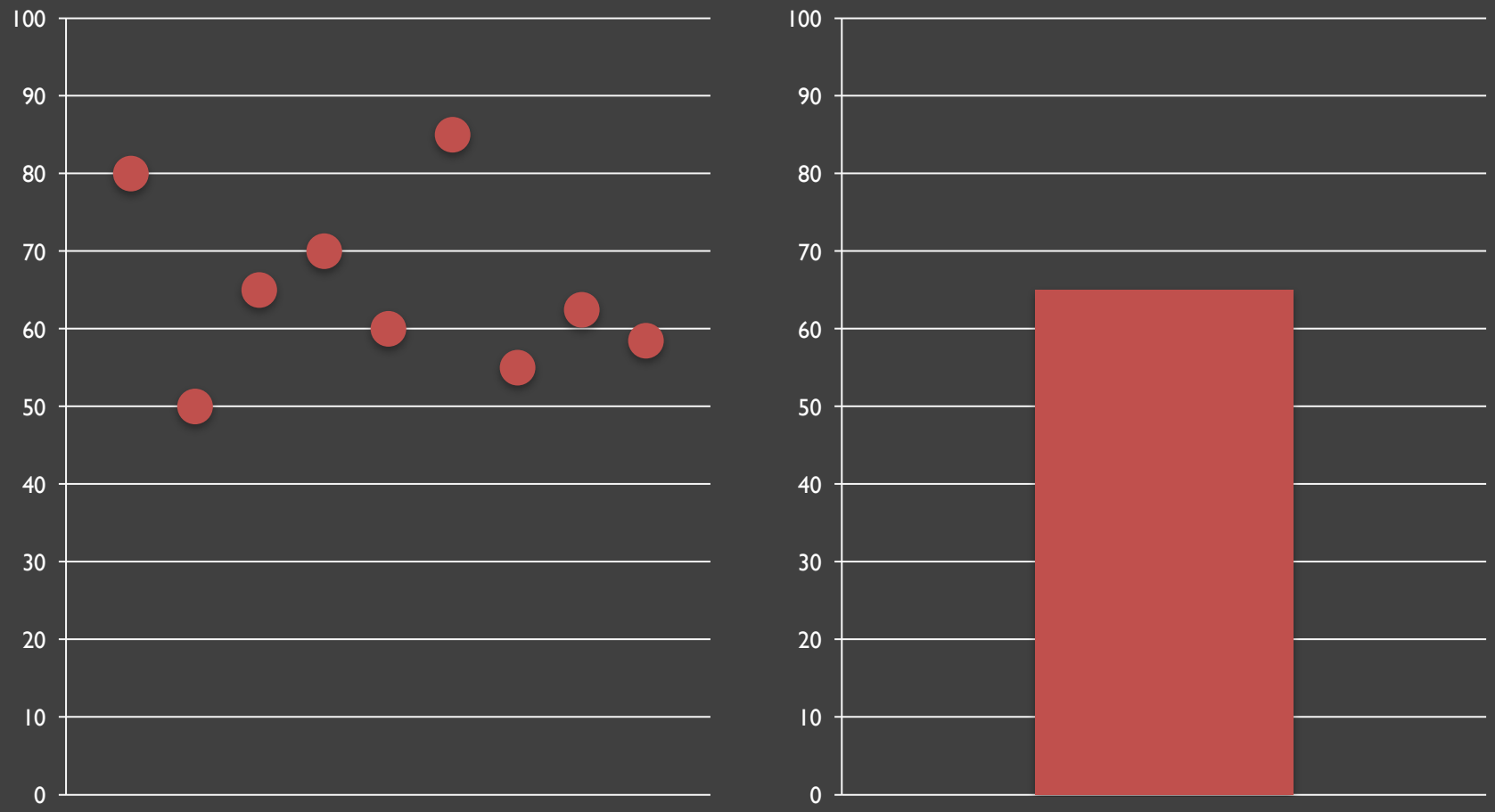
T-Tests and Confidence Intervals




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

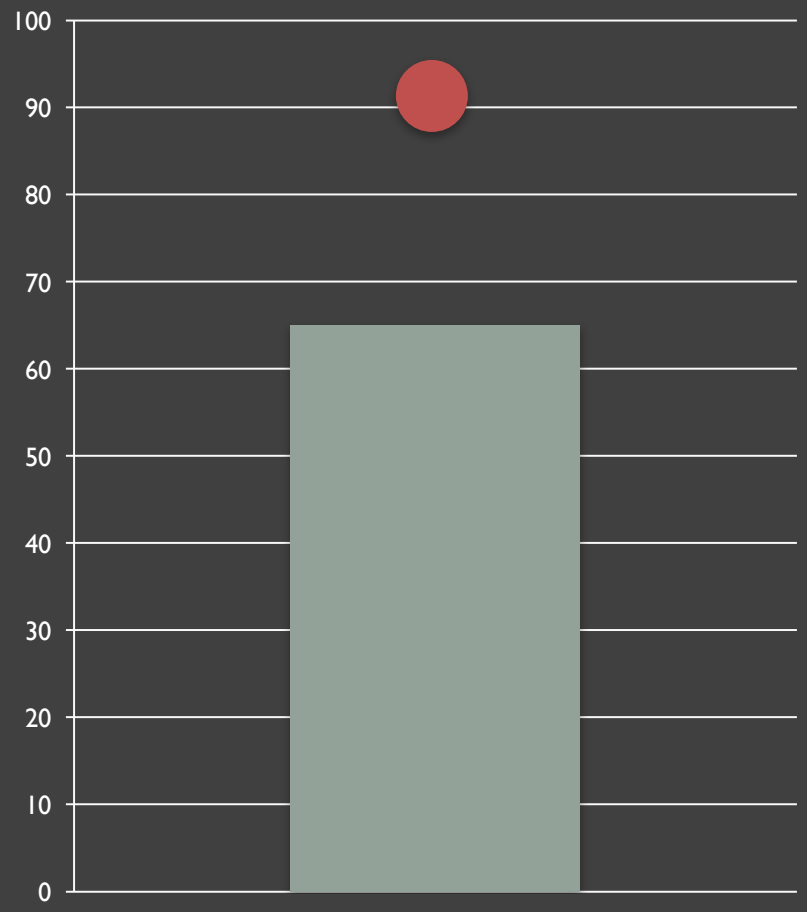

100
90
80
70
60
50
40
30
20
10
0
Placebo
Treatment

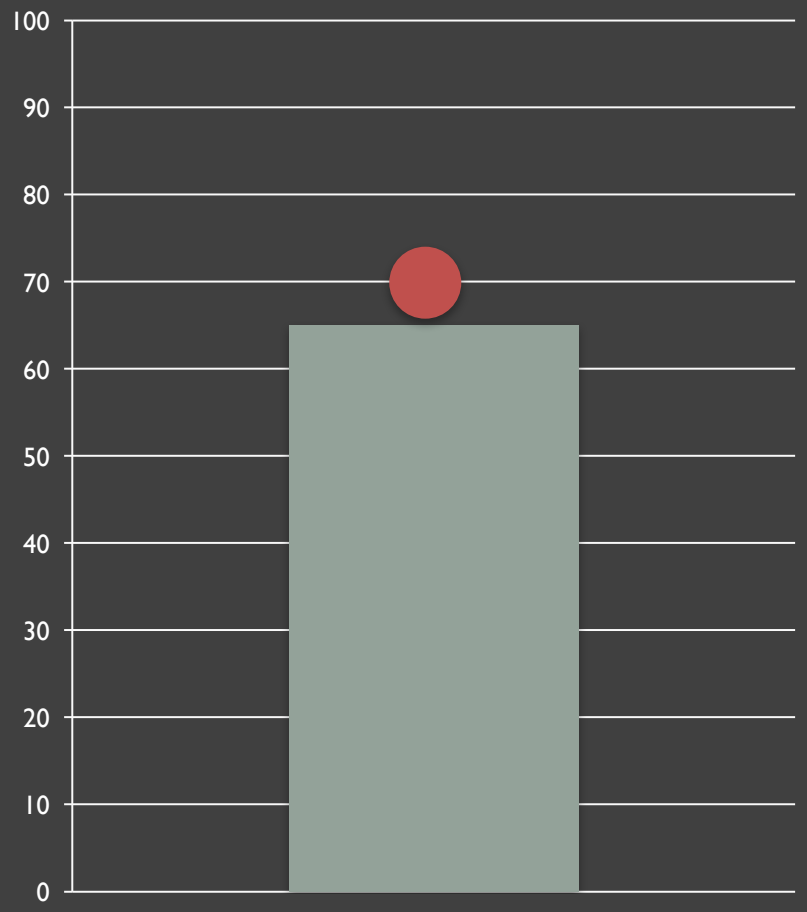
Within-the-bar bias



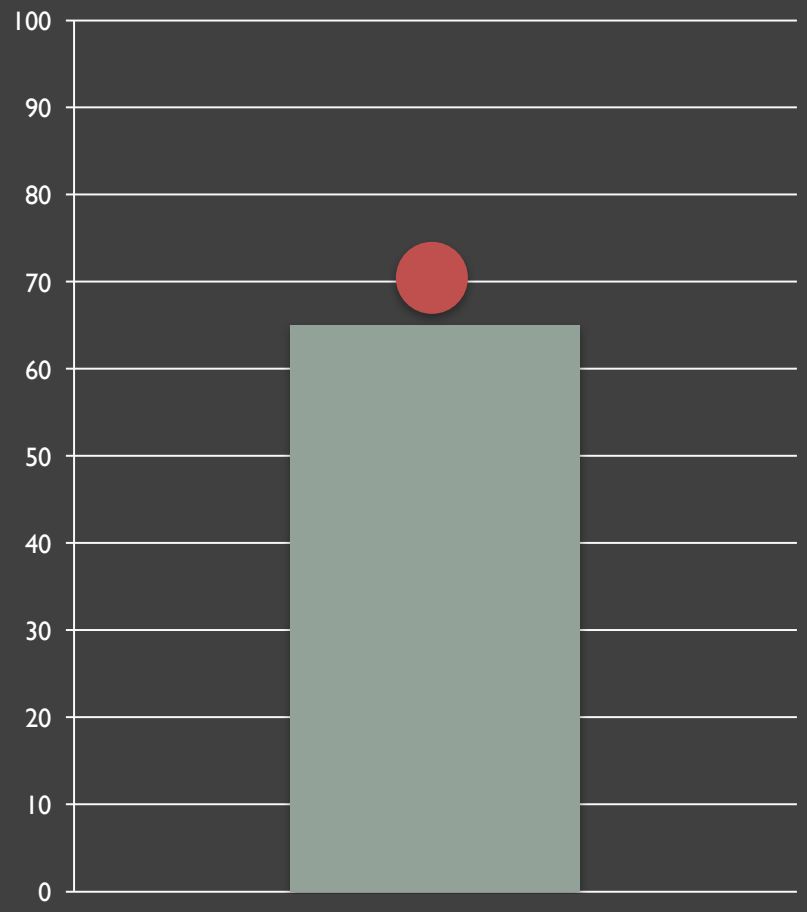
Newman, George E, and Brian J Scholl. "Bar graphs depicting averages are perceptually misinterpreted: the within-the-bar bias." Psychonomic bulletin & review 19.4 (2012): 601–7.

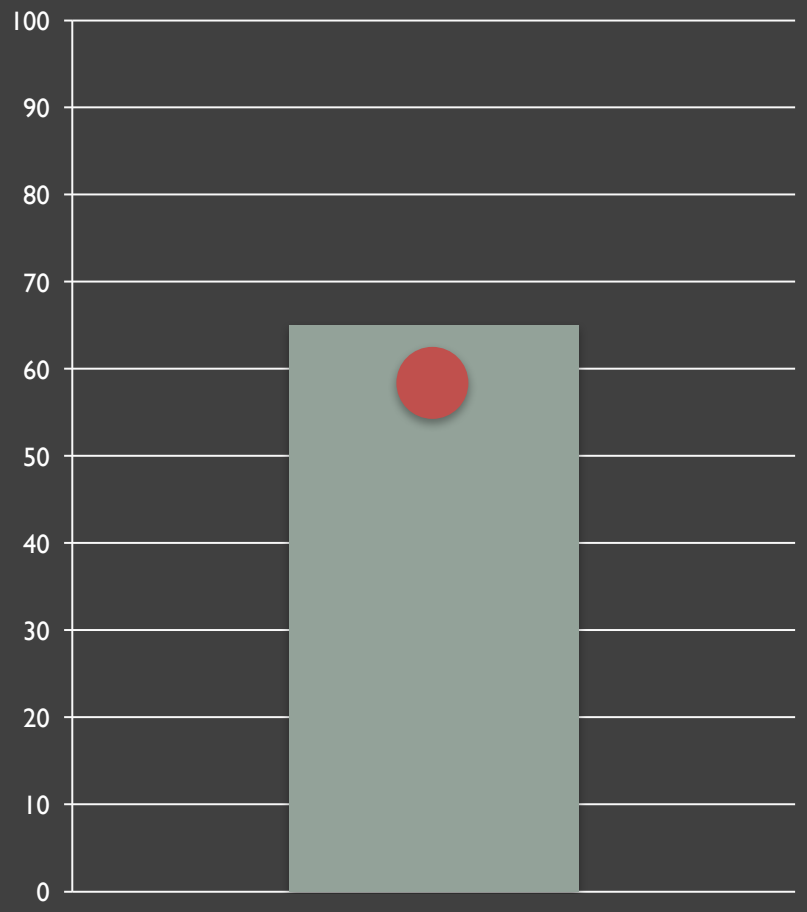
Within-the-bar bias



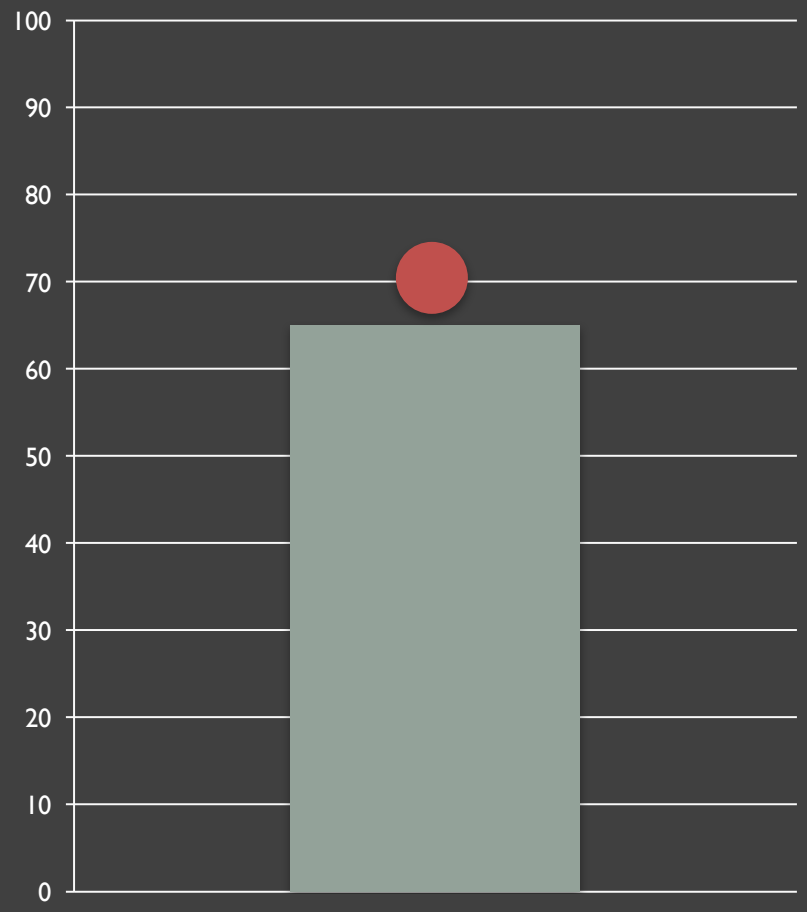


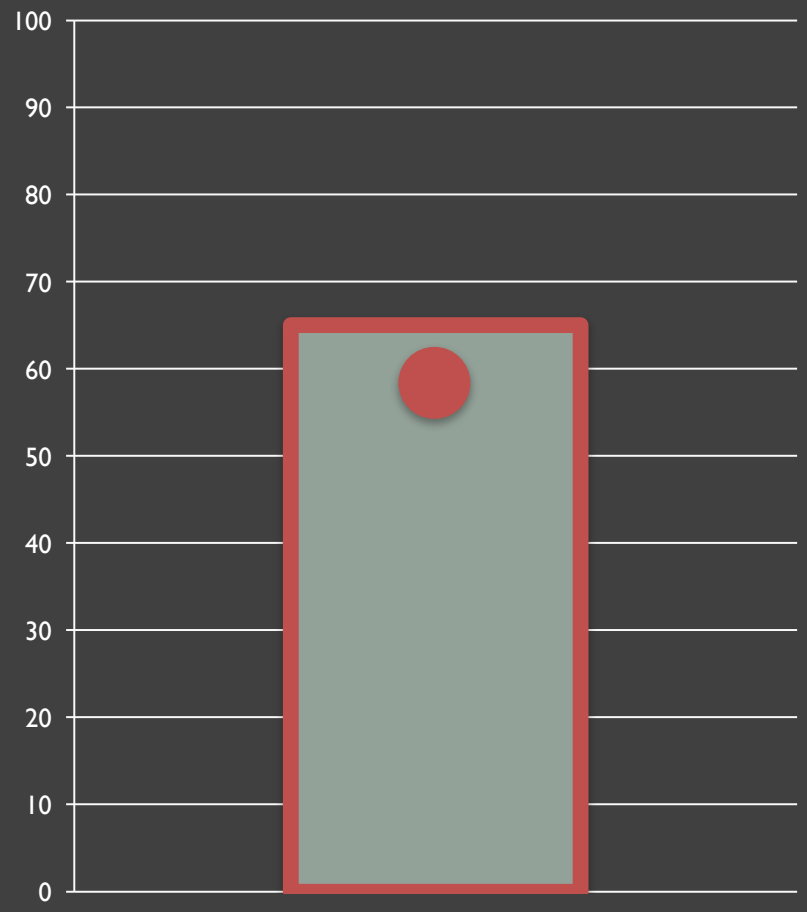
Within-the-bar bias



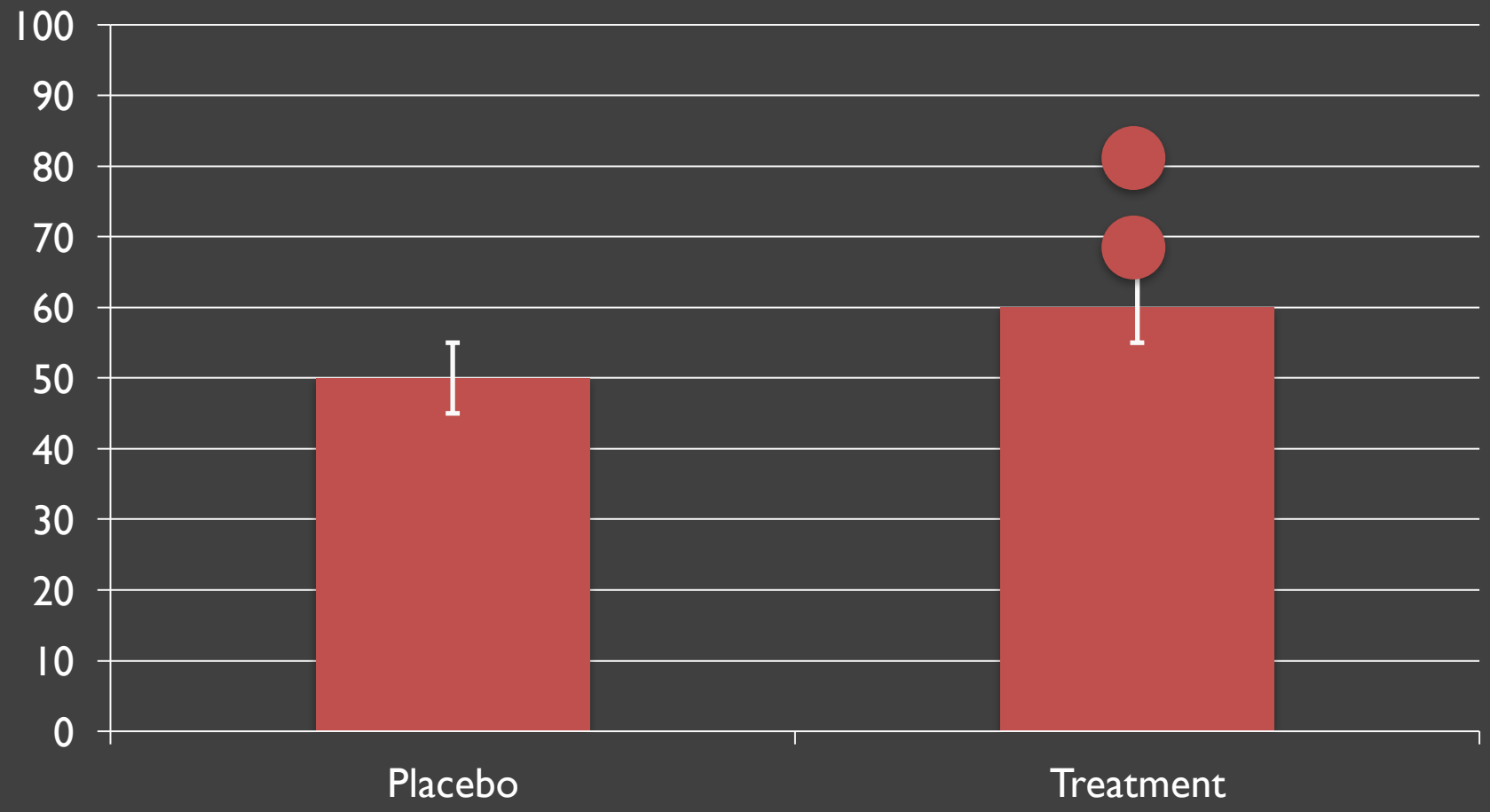


Within-the-bar bias



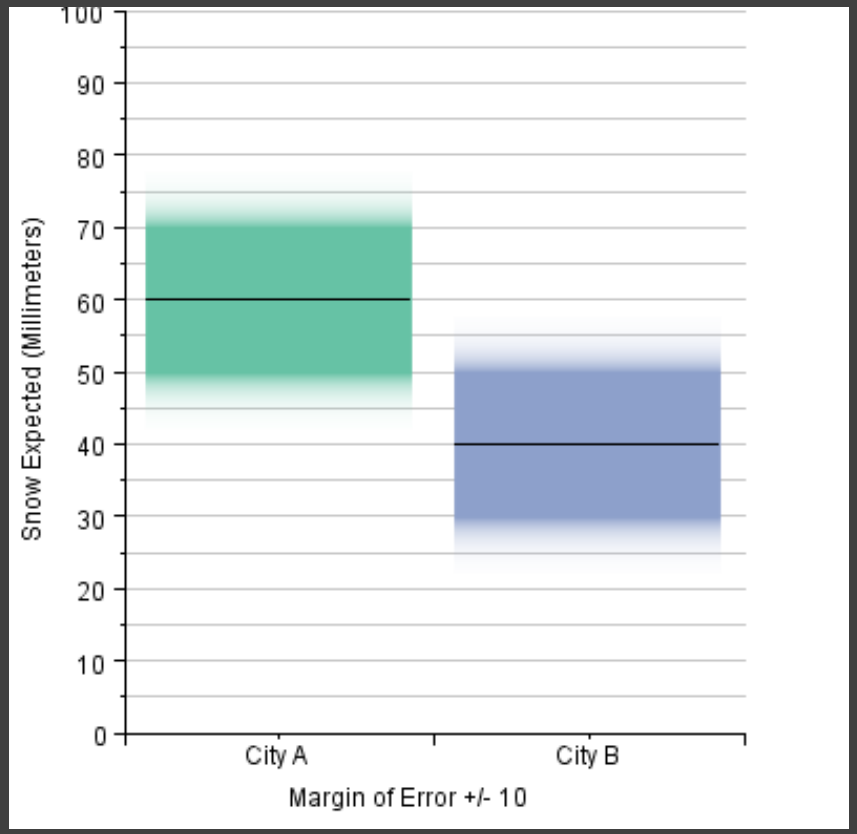


Binary Bias

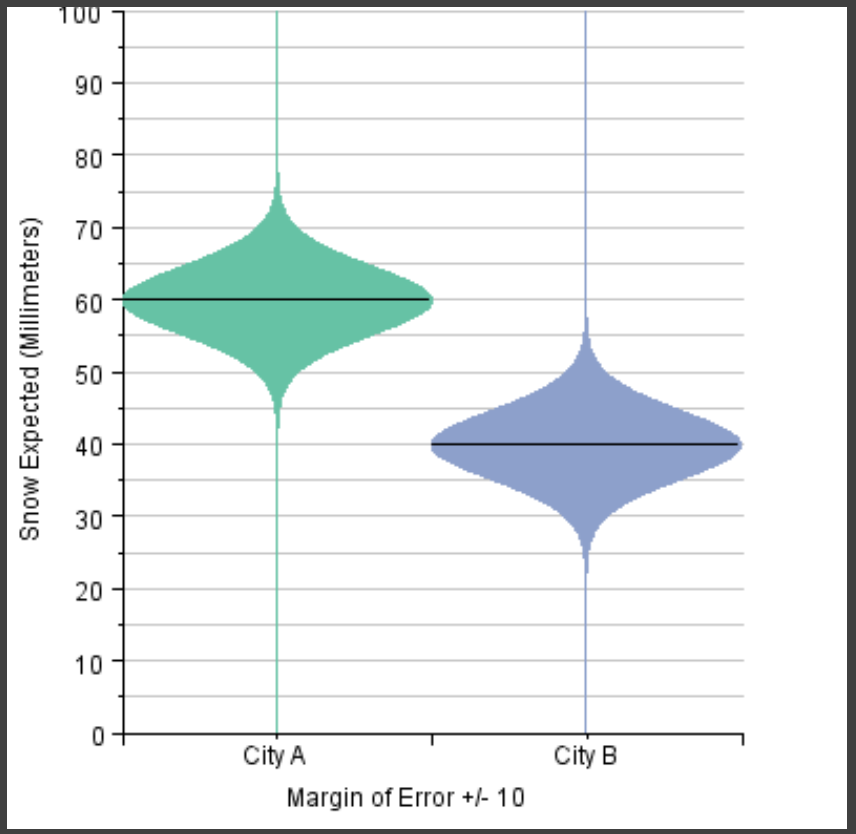


Alternatives

Gradient Plot

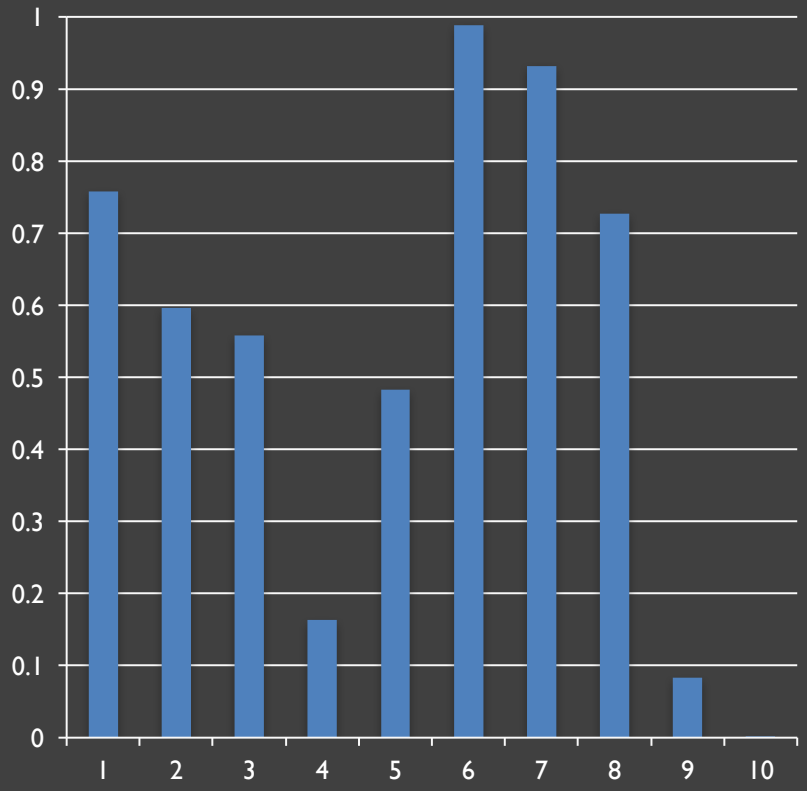


Violin Plot

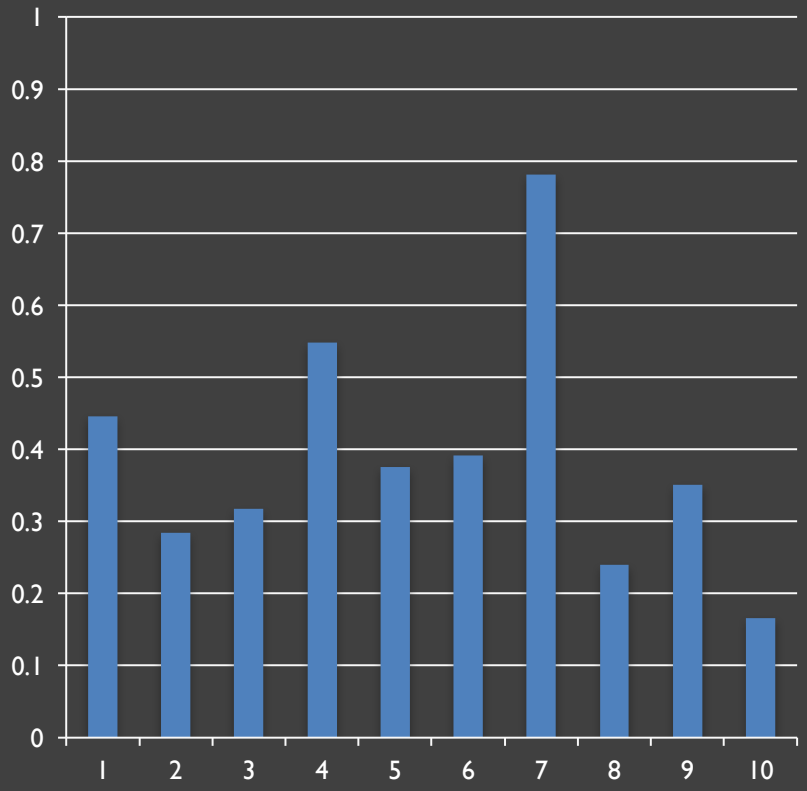


Which Stock To Buy?

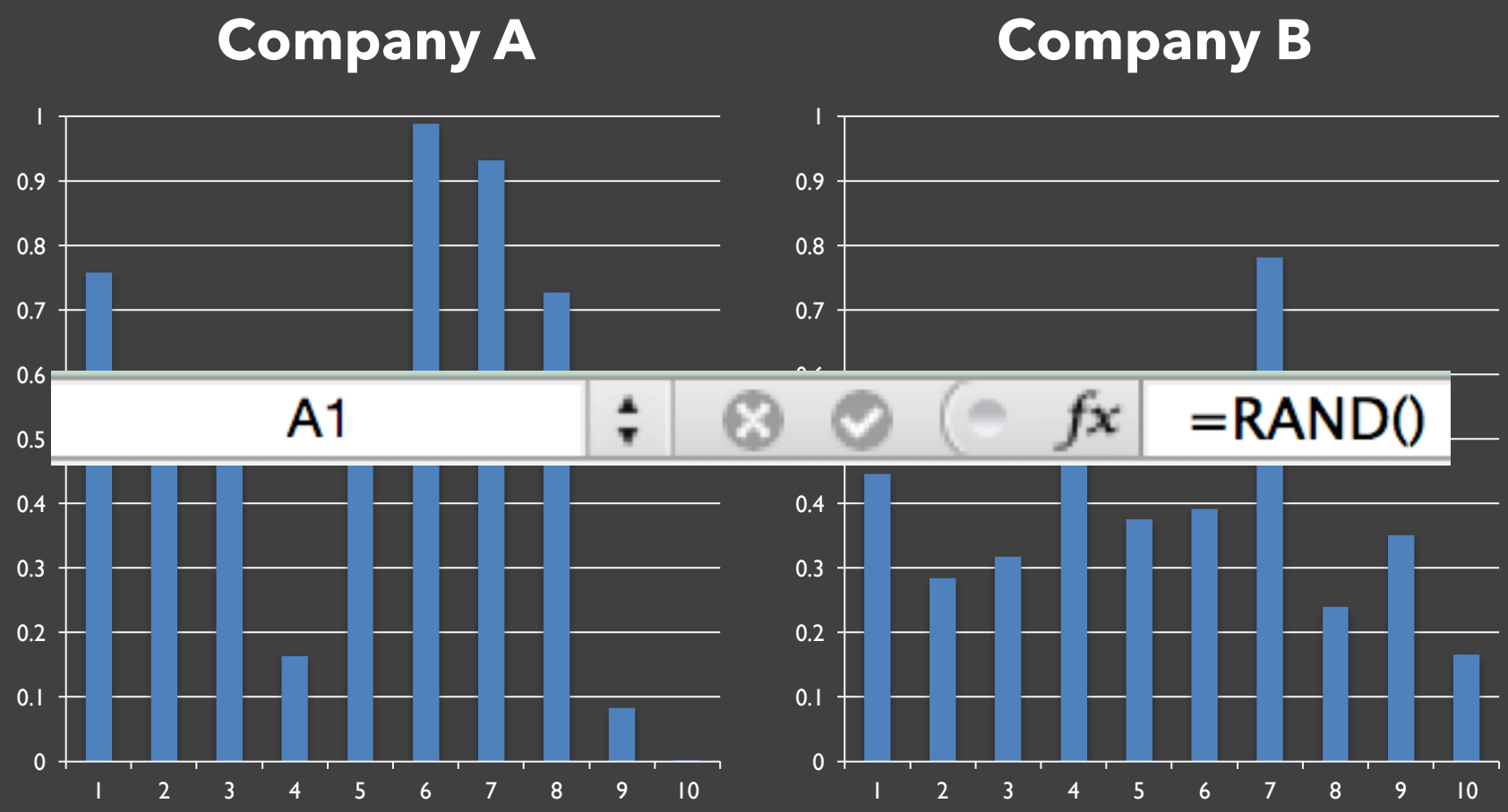
Company A



Company B



Neither!


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

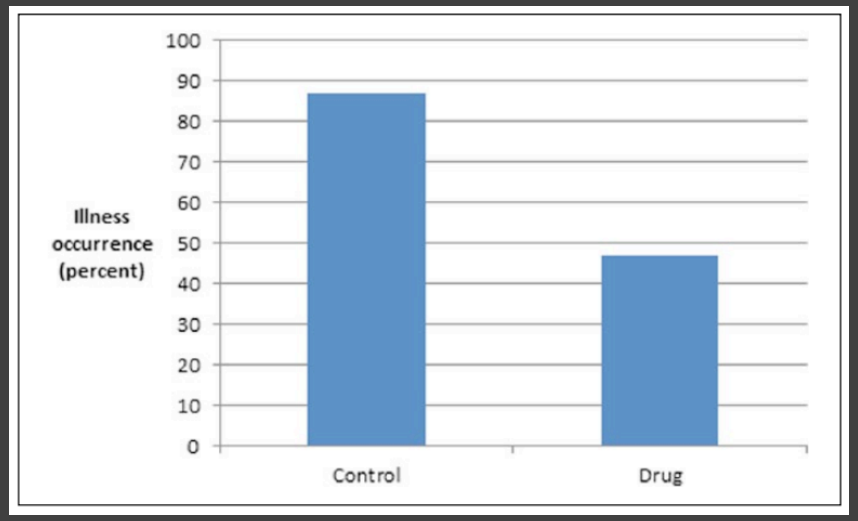
Wu Wei

Pareidolia

“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

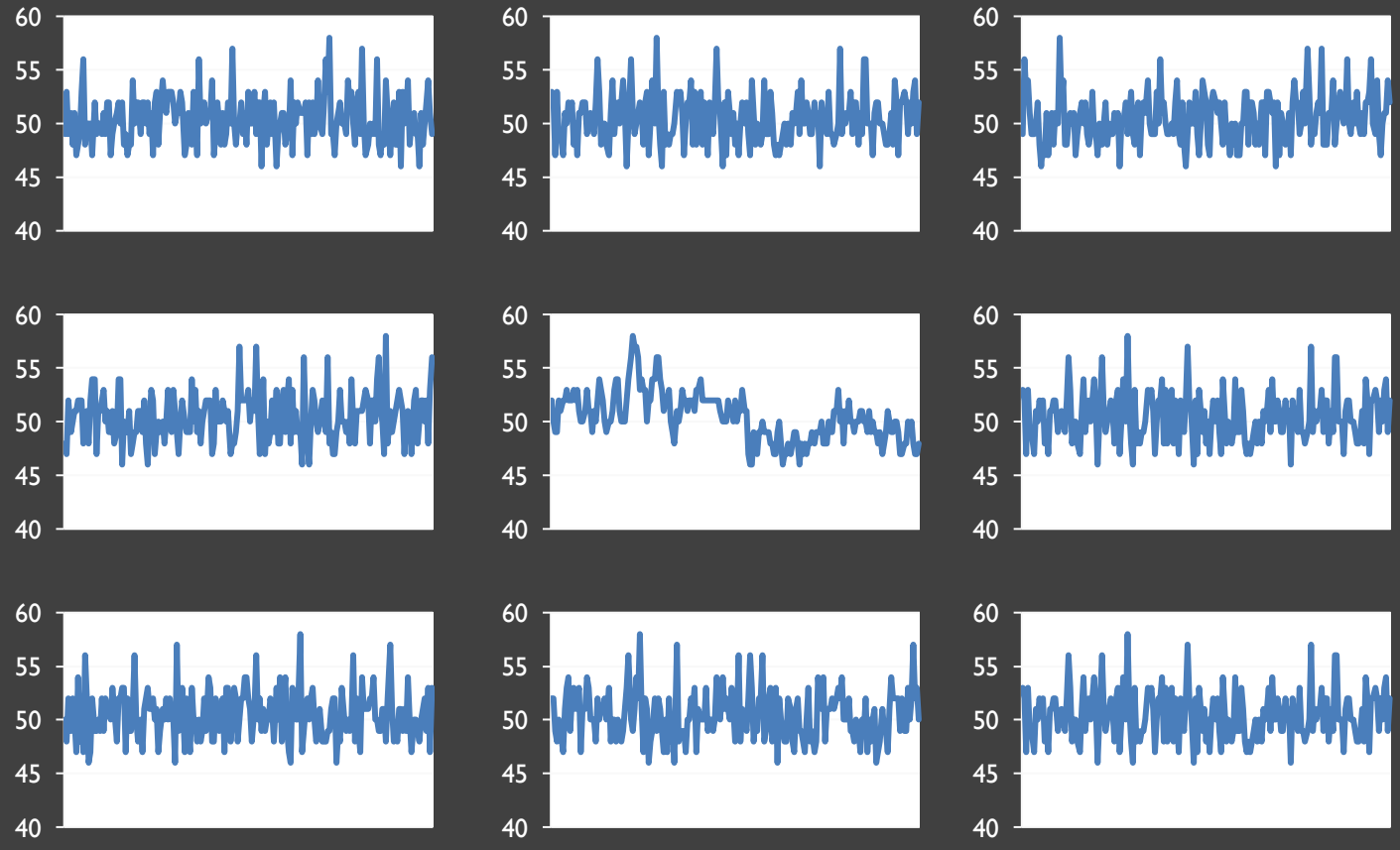
"A large pharmaceutical company has recently developed a new drug to boost peoples' immune function. It reports that trials it conducted demonstrated a drop of forty percent (from eighty seven to forty seven percent) in occurrence of the common cold. It intends to market the new drug as soon as next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



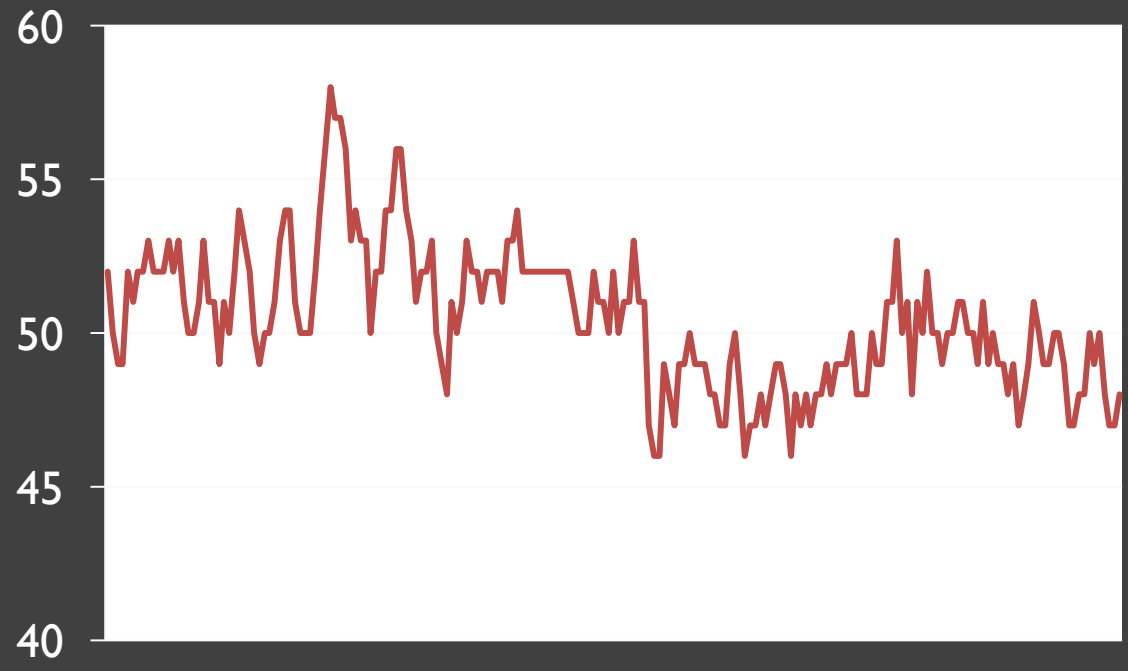
Tal, Aner and Wansink, Brian. Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy. Public Understanding of Science, 2016.

Jobs Reports

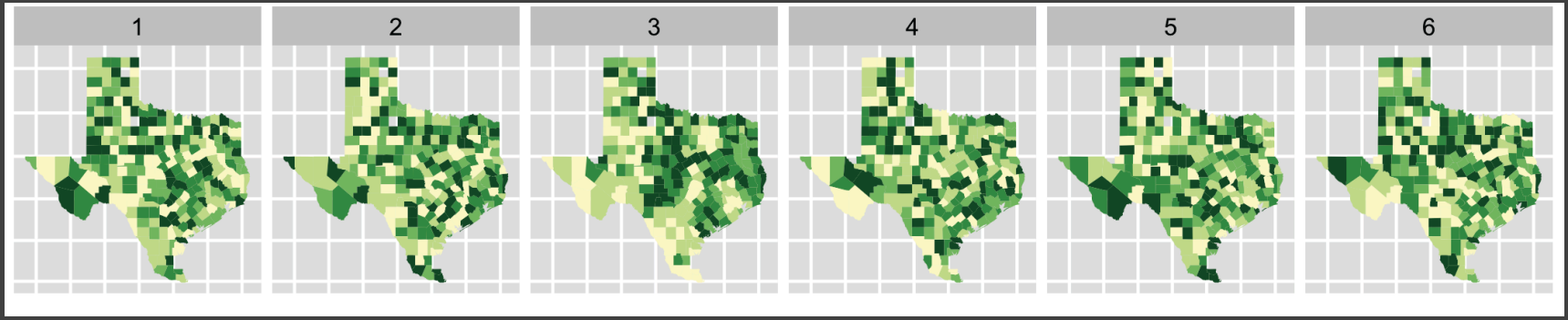
Have People Made Up Their Mind About Obama?


60
55
50
45
40

Visual Lineups

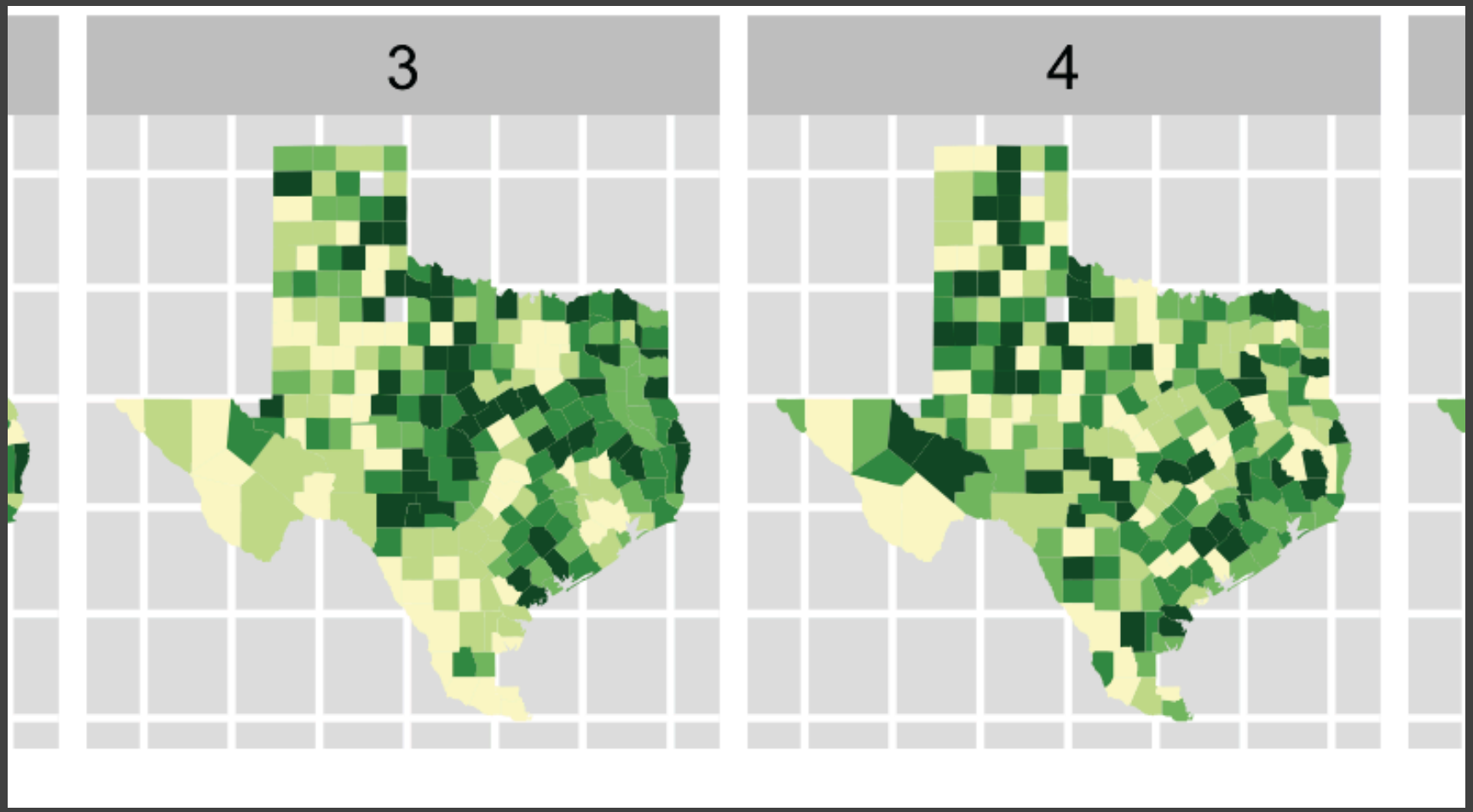


Visual Lineups



Hadley Wickham 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

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

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

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

Bayes' Law

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

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

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

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

$$P(+) = 0.107$$

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

Problems

People are bad at this.

People who should be good at this are bad at it.

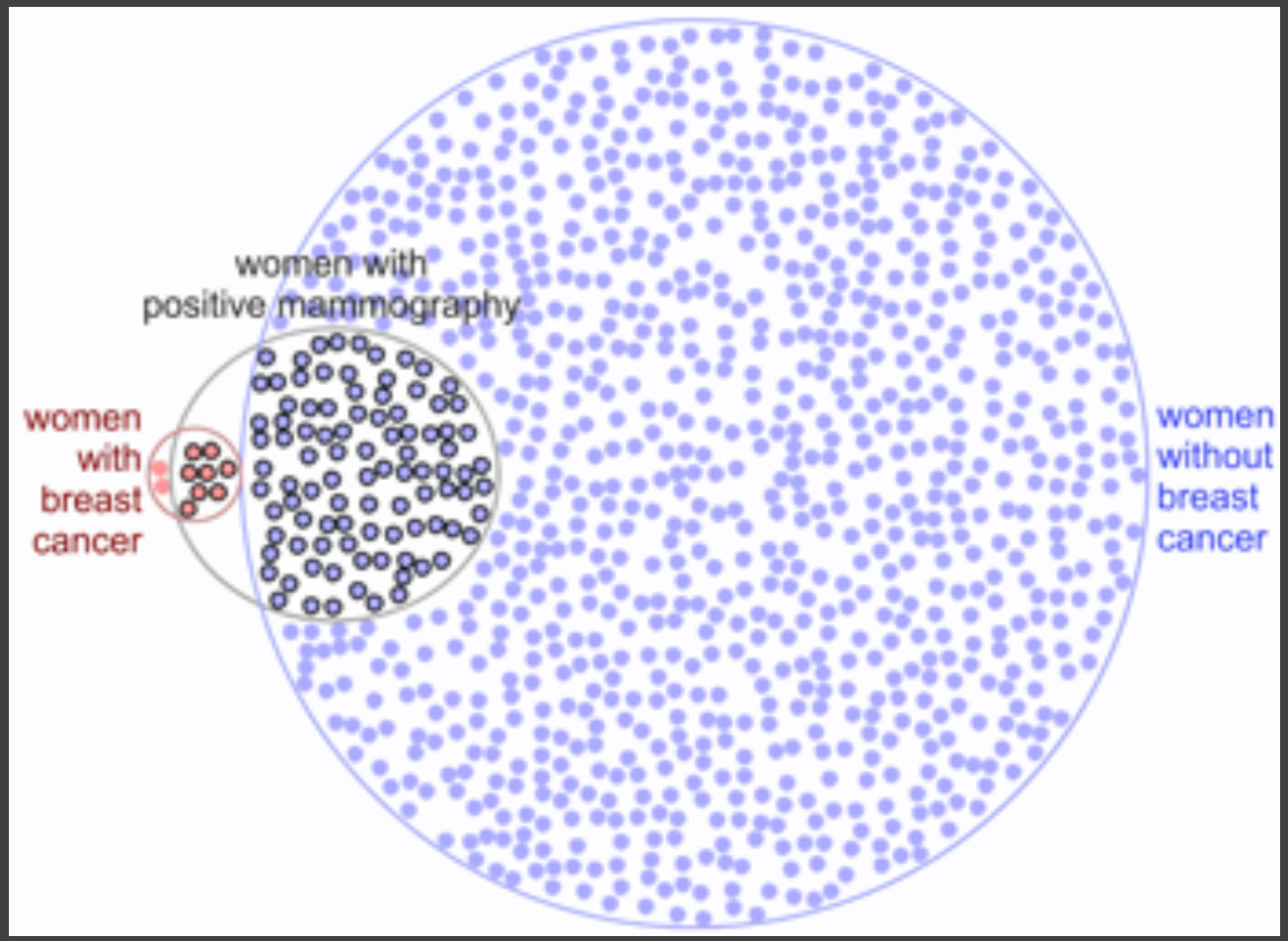
How you present the problem affects how bad people are at it.

How To Present Probabilities

Probability	P(A) = 0.6
Percentage	60% of A
Natural Frequency	3 out of 5 times, A happens.

Alvitta Ottley, et al. “Improving Bayesian reasoning: the effects of phrasing, visualization, and spatial ability.” VIS 2016.

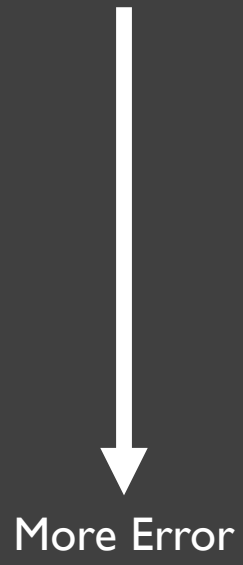
Base Rate Fallacy



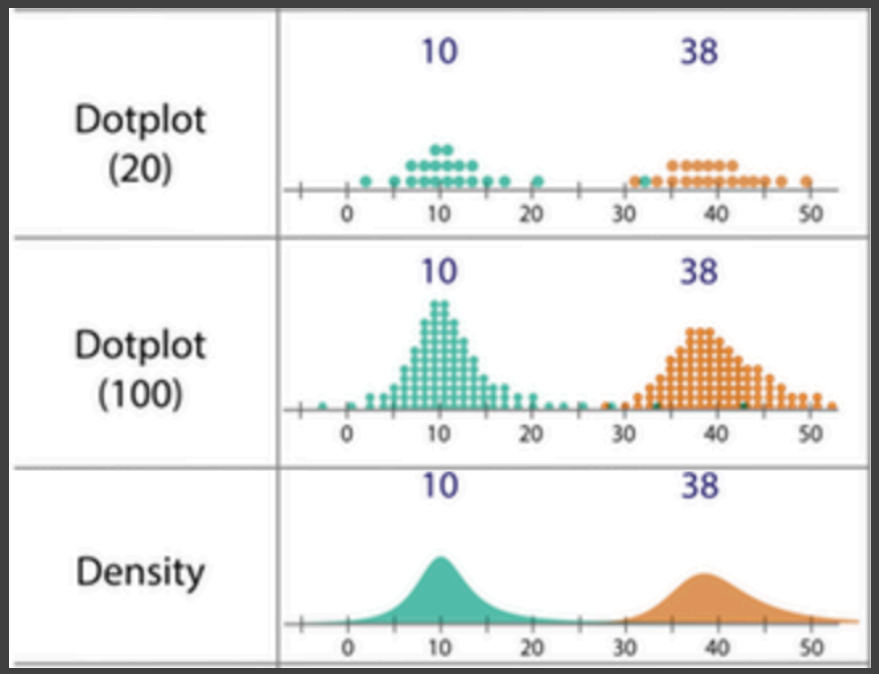
Luana Micallef, Pierre Dragicevic, and Jean-Daniel Fekete. "Assessing the Effect of Visualizations on Bayesian Reasoning Through Crowdsourcing." VIS 2012.

How To Present Probability Distributions

Less Error



More Error



Matthew Kay, Tara Kola, Jessica Hullman, Sean Munson. "When(ish) is My Bus? User-centered Visualizations of Uncertainty in Everyday, Mobile Predictive Systems." CHI 2016.

What Can Go Wrong?

Uncertainty can be difficult to understand, and require a statistical background and high numeracy. Additionally, cognitive and perceptual biases can result in people making poor or error-prone decisions from uncertain data.

What Can Go Wrong?

Uncertainty can be difficult to understand, and require a statistical background and high numeracy. Additionally, cognitive and perceptual biases can result in people making poor or error-prone decisions from uncertain data.

A LOT

Questions To Answer

What Does Uncertainty Mean?

How Should I Visualize It?

What Can Go Wrong?

Questions To Answer

What Does Uncertainty Mean?

LOTS OF THINGS

How Should I Visualize It?

IT DEPENDS

What Can Go Wrong?

A LOT

Stuff I Showed You

http://flowingdata.com/2015/09/23/years-you-have-left-to-live-probably/

http://guns.periscopic.com/?year=2013

http://rpsychologist.com/d3/CI/

http://www.evanmiller.org/ab-testing/t-test.html

https://www.nytimes.com/2014/05/02/upshot/how-not-to-be-misled-by-the-jobs-report.html?_r=0

Thanks!

Michael Correll

mcorrell@tableau.com