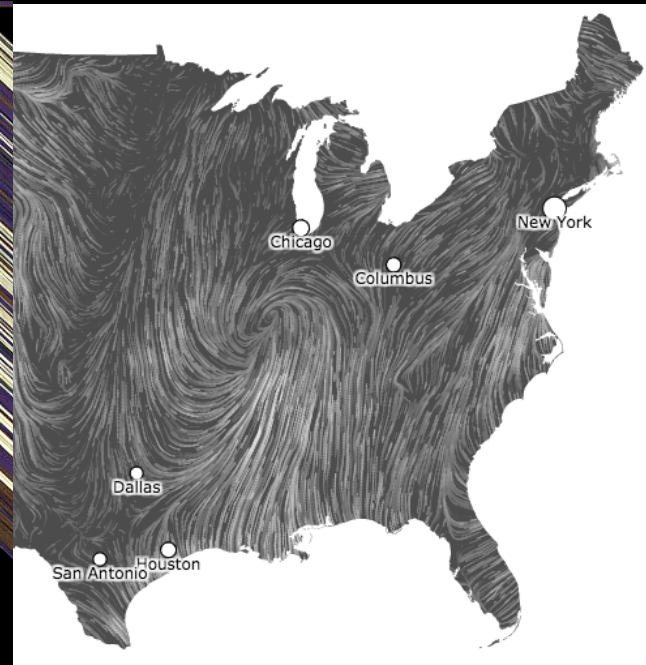
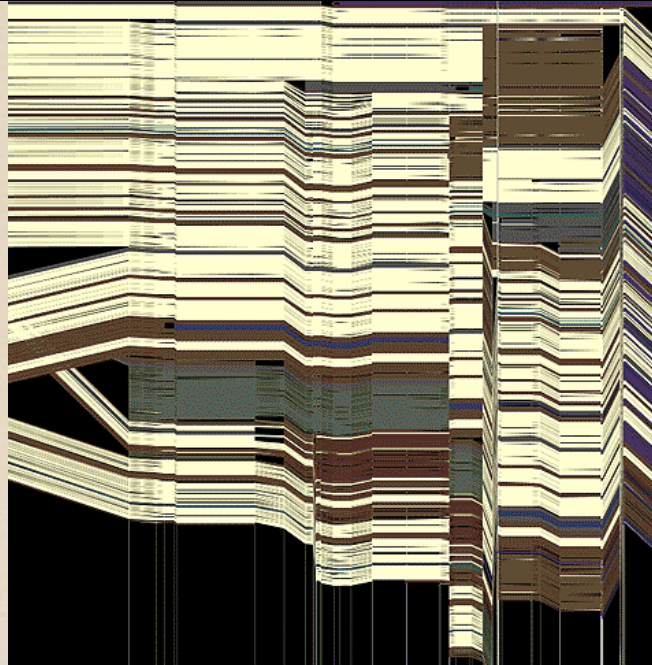
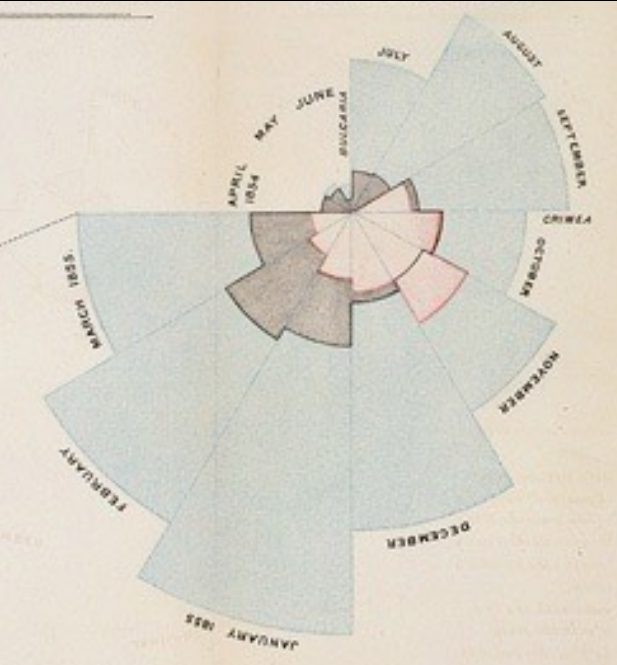


# CSE 512 - Data Visualization

# Uncertainty



Jeffrey Heer University of Washington

(with significant material from Michael Correll)

# Questions To Answer

What Does Uncertainty Mean?

How Should I Visualize It?

What Can Go Wrong?

**What we talk about when we  
talk about “uncertainty”...**

# Things “Uncertainty” Can Mean

Doubt

Risk

Variability

Error

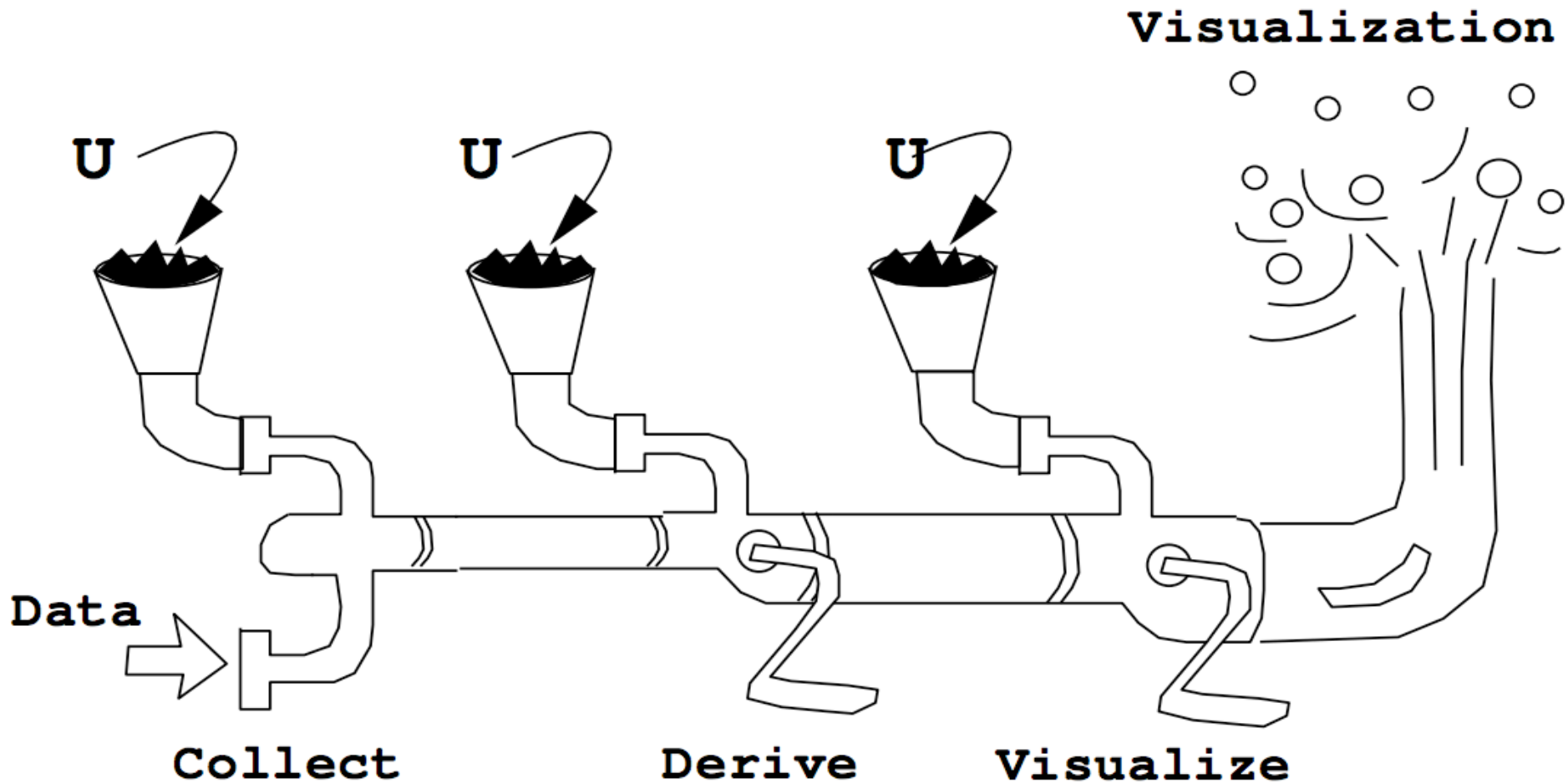
Lack of Knowledge

Hedging

...

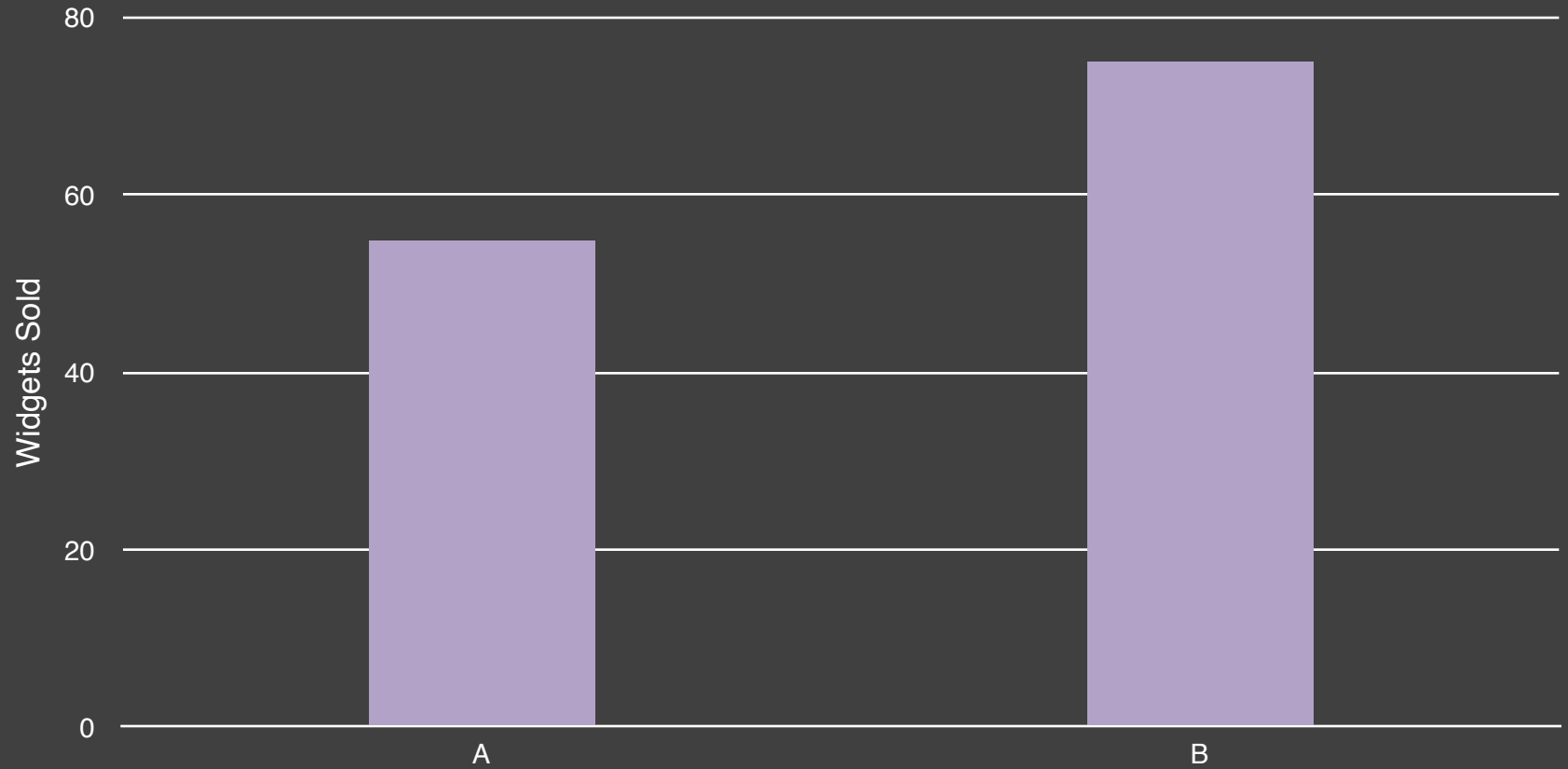


# Uncertainty Vis Pipeline



# A Bar Chart

Sales of Widgets for Stores A and B

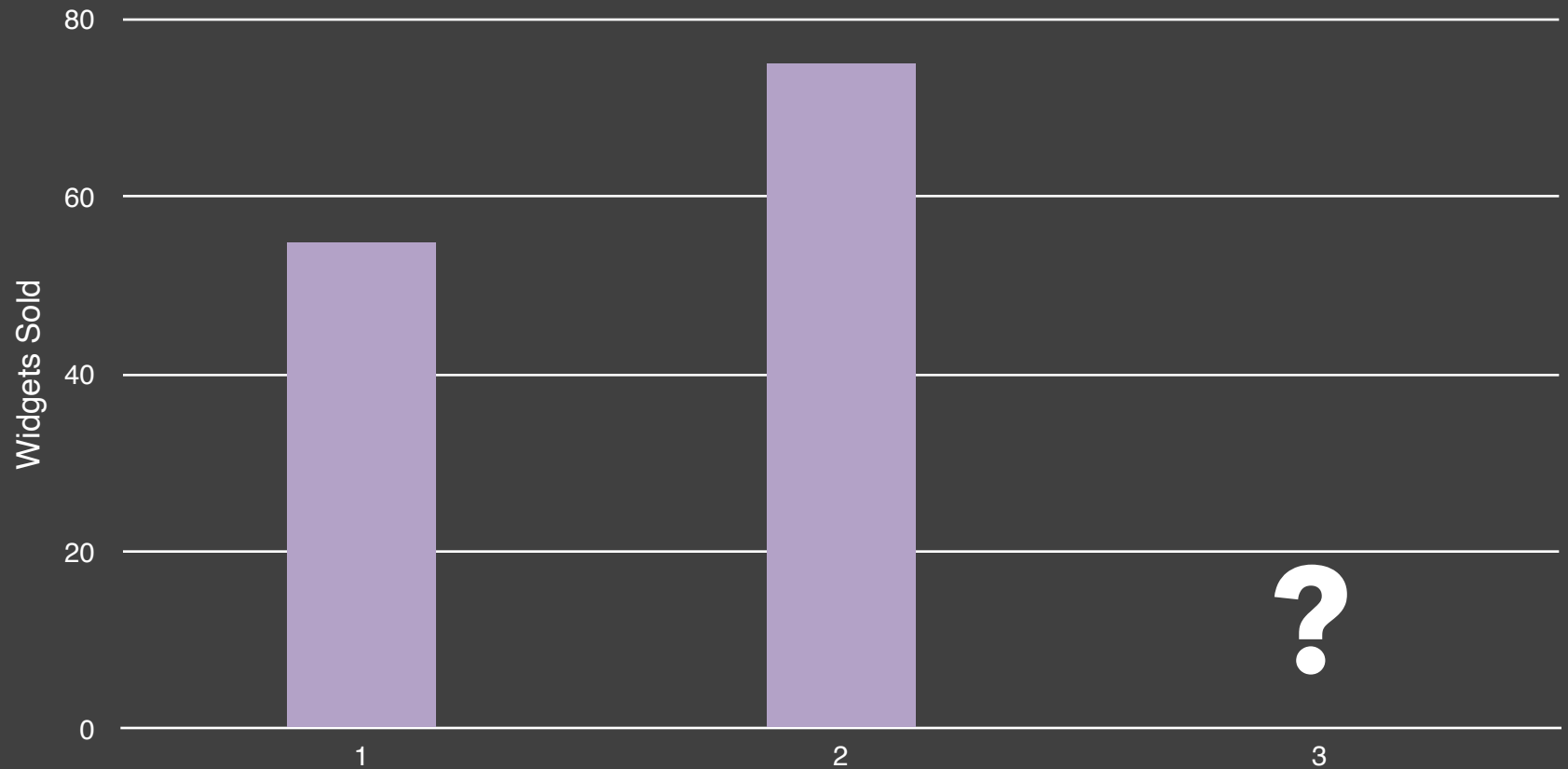


# Measurement Uncertainty



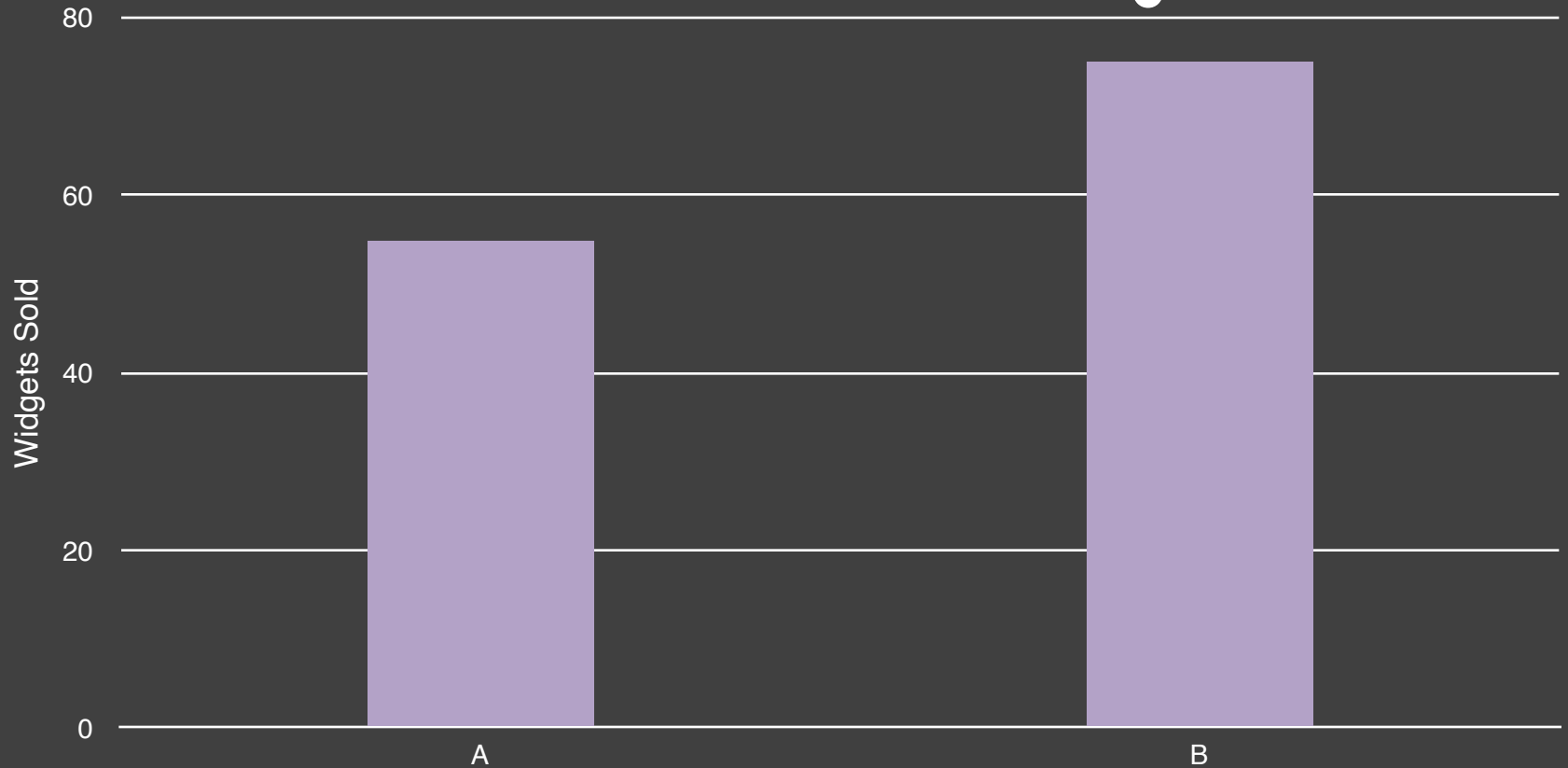
# Forecast Uncertainty

Sales of Widgets for Quarters 1 and 2



# Decision Uncertainty

We Should Close Store A ?



# Uncertainty Sources

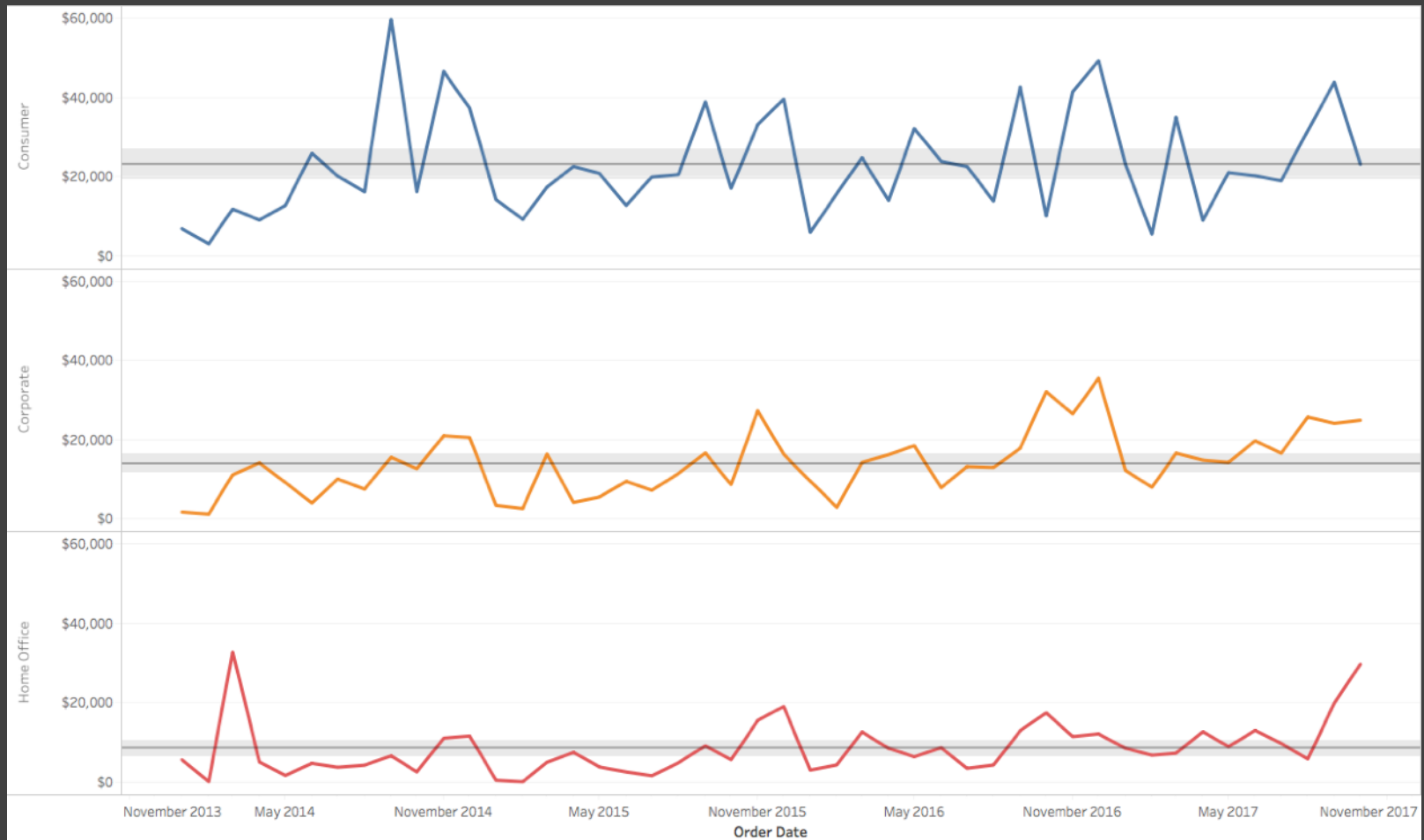
**Measurement Uncertainty:** "We're not sure what the data are"

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

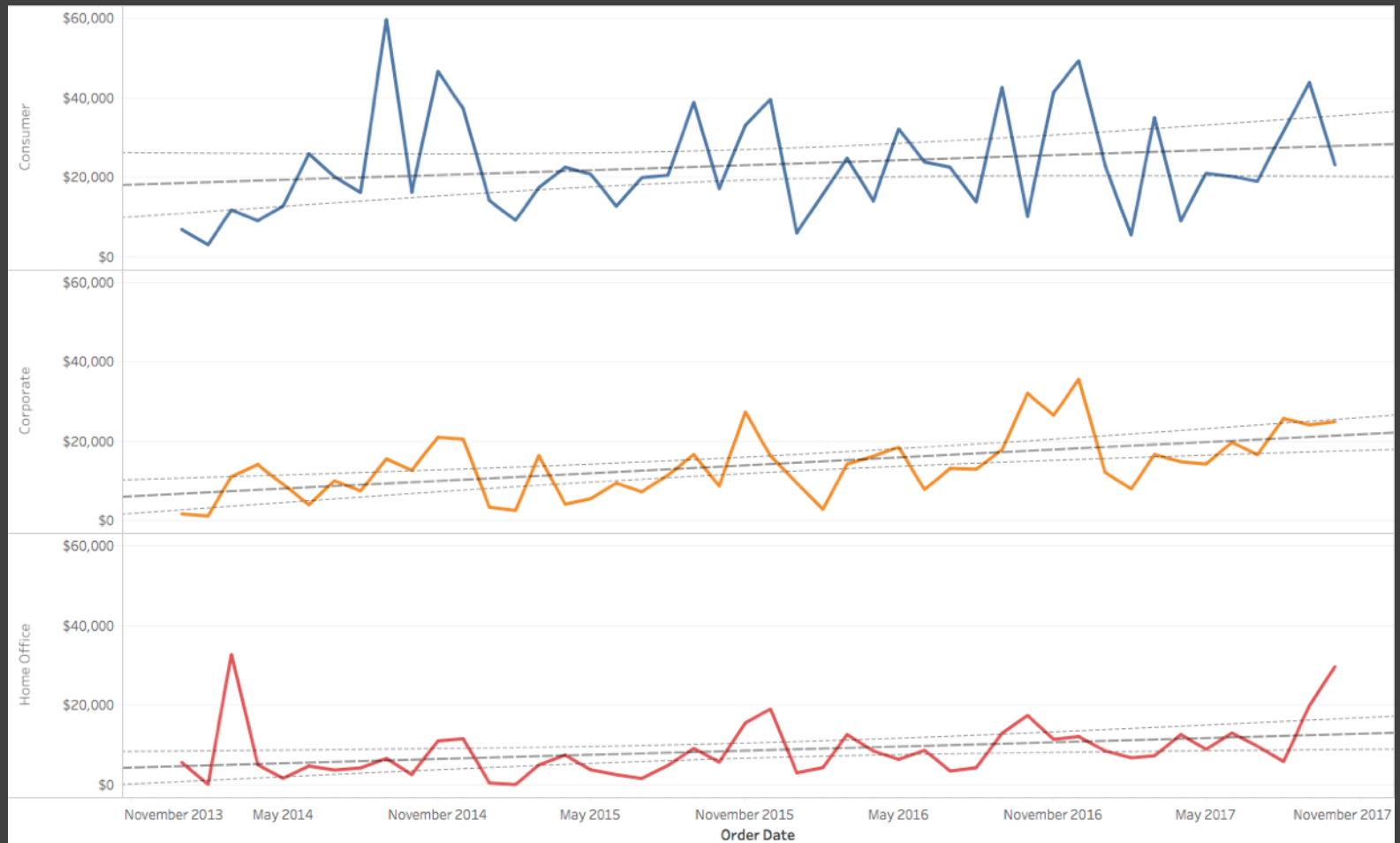
**Forecast Uncertainty:** "We're not sure what will happen to the data next"

**Decision Uncertainty:** "We're not sure what to do with the data"

# Measurement Uncertainty

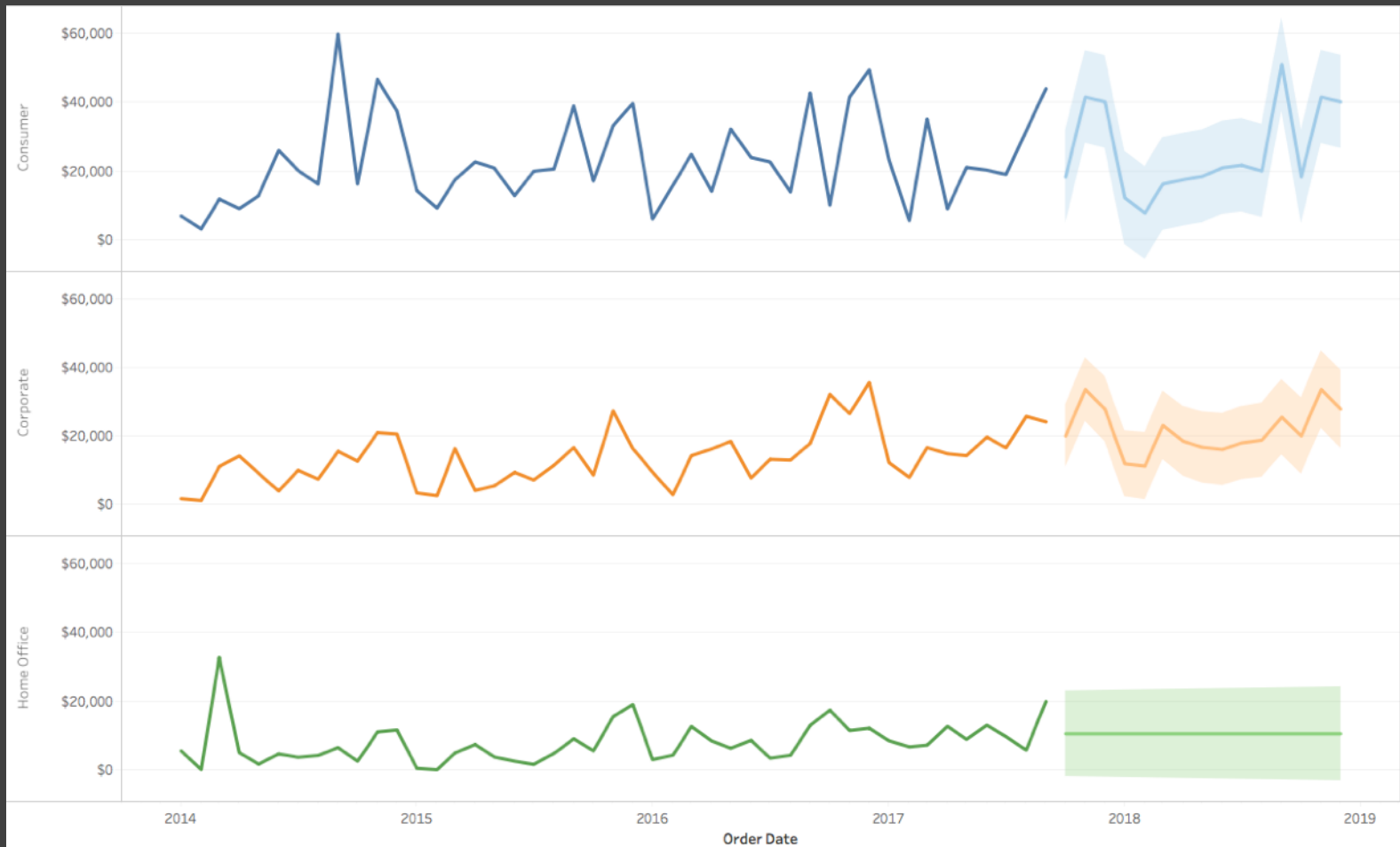


# Model Uncertainty





# Forecast Uncertainty



# Uncertainty Visualization

There are different **types** and **sources** of uncertainty.

We can **quantify** or **model** our uncertainty.

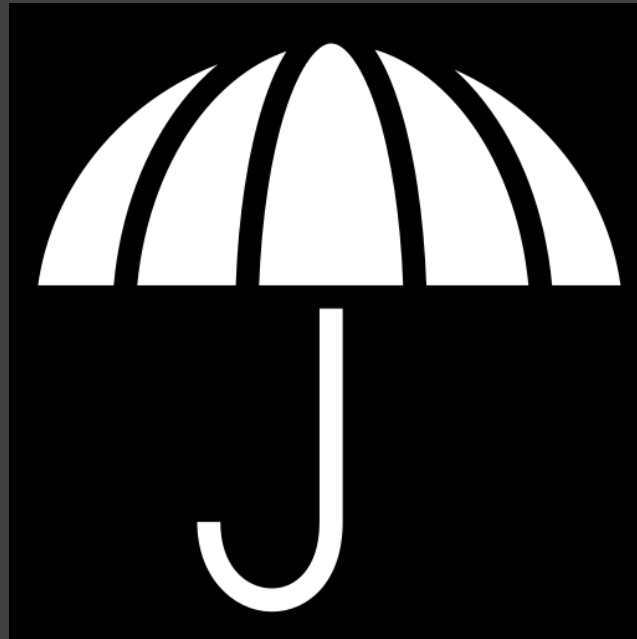
The visual presentation of uncertainty can **clash** with cognitive and perceptual biases.

Should I Bring an Umbrella?







# Decision Uncertainty

"50% Chance of Rain"



# Types of Error



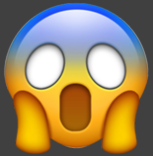
	 False Positive
 False Negative	

# The Boy Who Cried Wolf

**Type I: False Positive**



**Type II: False Negative**





Sean J. Taylor @seanjtaylor

Here's my trick.

TYPE

FALSE POSITIVE

FALSE NEGATIVE



# Model Uncertainty

"50% Chance of Rain"





# Model Uncertainty

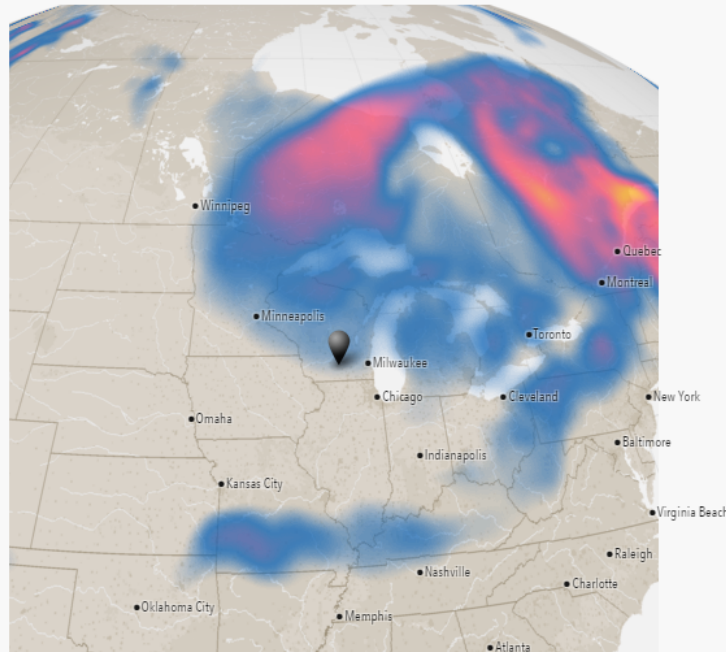
PRECIP MAP

Local

Regional

Global

TODAY 5PM



RIGHT NOW +

☁ 41°

Overcast · Feels like 32°

NEXT HOUR

Overcast for the hour.

NEXT 24 HOURS

Mostly cloudy throughout the day.

NEXT 7 DAYS

Light rain throughout the week, with temperatures rising to 64°F on Sunday.

TODAY ☁ Mostly cloudy throughout the day.

37° 41°



WED ☁ Partly cloudy until evening.

31° 47°



THU ☀ Clear throughout the day.

27° 53°



FRI ☁ Light rain in the afternoon.

32° 56°



SAT ☁ Mostly cloudy until evening.

34° 55°



# Measurement Uncertainty

## Precision



# Measurement Uncertainty

## Precision



# Measurement Uncertainty

## Precision



# Measurement Uncertainty

**Precision**



**Accuracy**





# Measurement Uncertainty

**Precision**



**Accuracy**



# Measurement Uncertainty

**Precision**



**Accuracy**

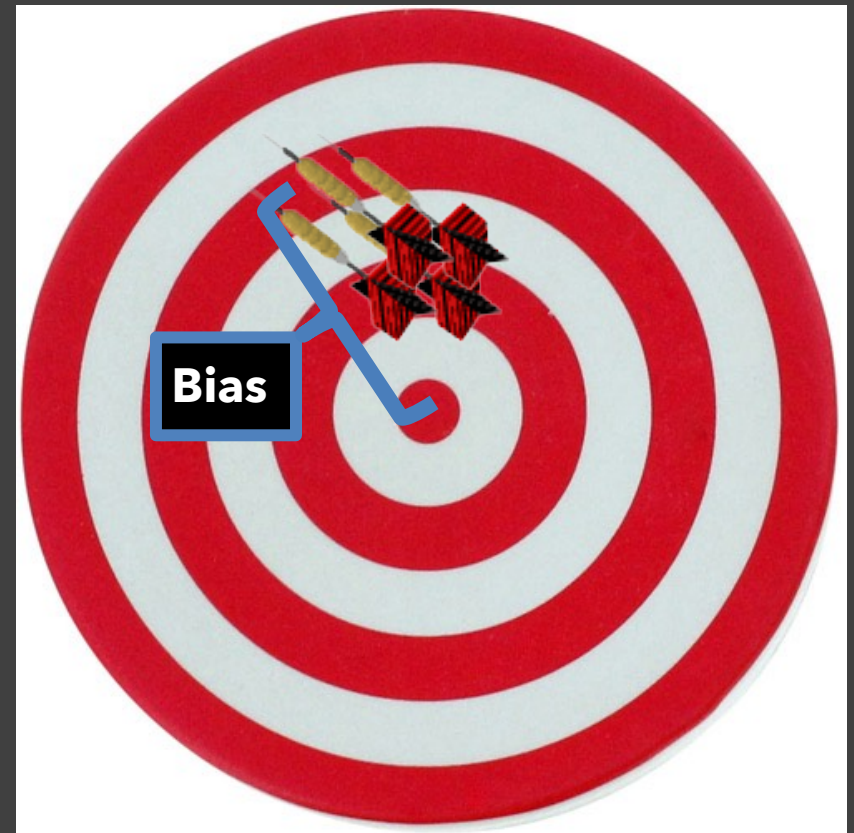


# Measurement Uncertainty

**Precision**



**Accuracy**





# 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?

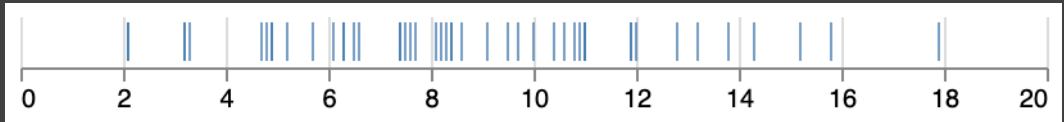
Any one of a number of potentially interconnected quantitative, qualitative, or factors that affect the quality, reliability, or utility of your data or data-driven decisions. Anything that can cause you to be unsure about your data or how to use it.

**LOTS OF  
THINGS**

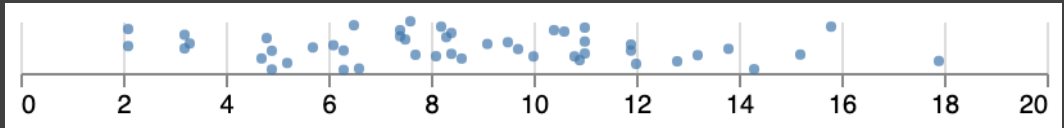
# Visualizing Distributions

# Distribution Visualizations

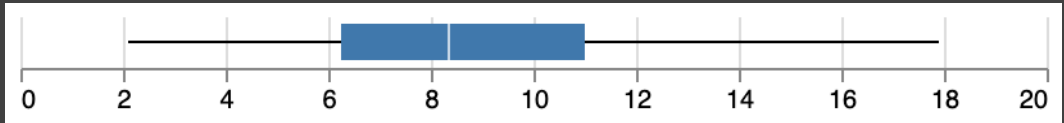
Strip Plot



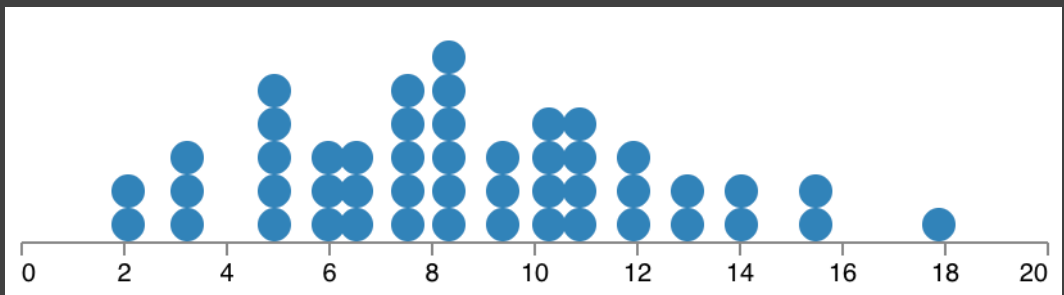
Jittered Plot



Box Plot



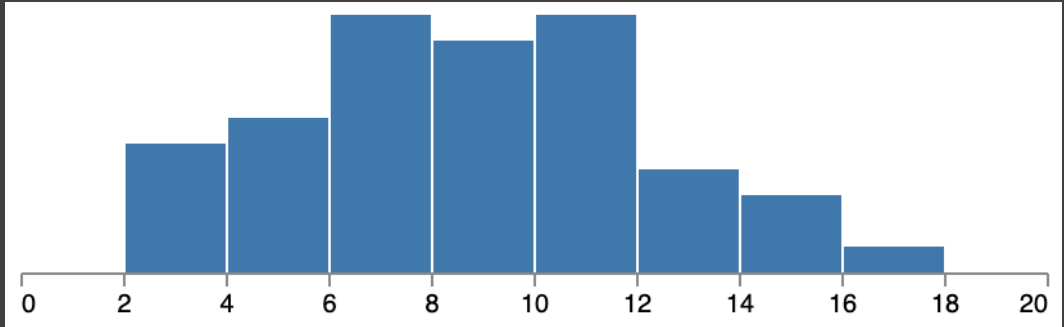
Dot Plot



# Distribution Visualizations

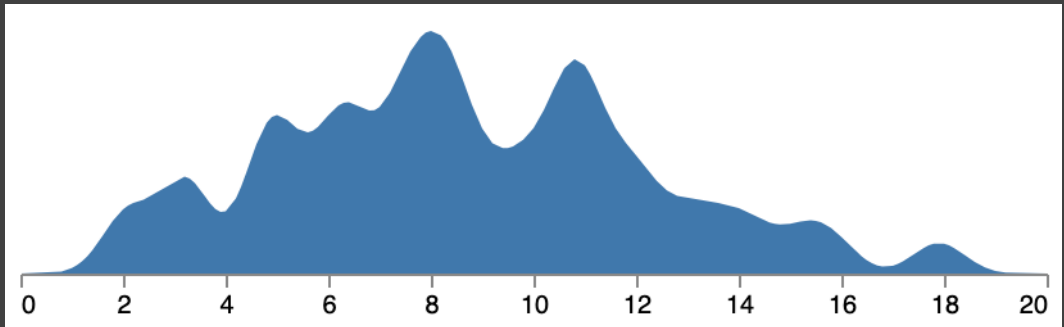
## Histogram

bin size = 2



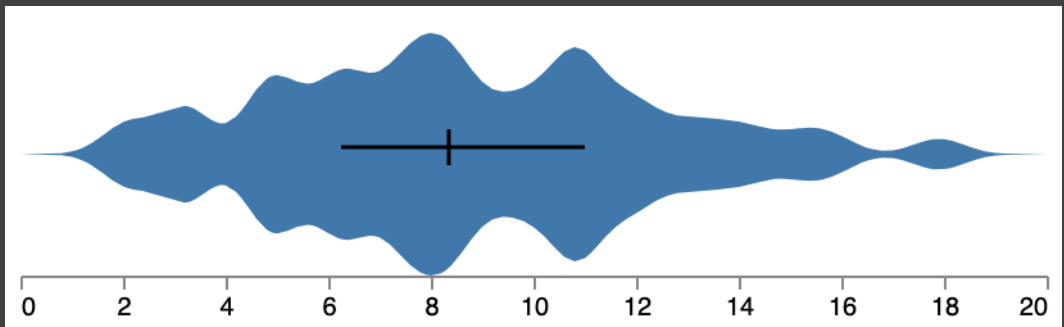
## Density Plot

kde,  $\sigma = 0.5$



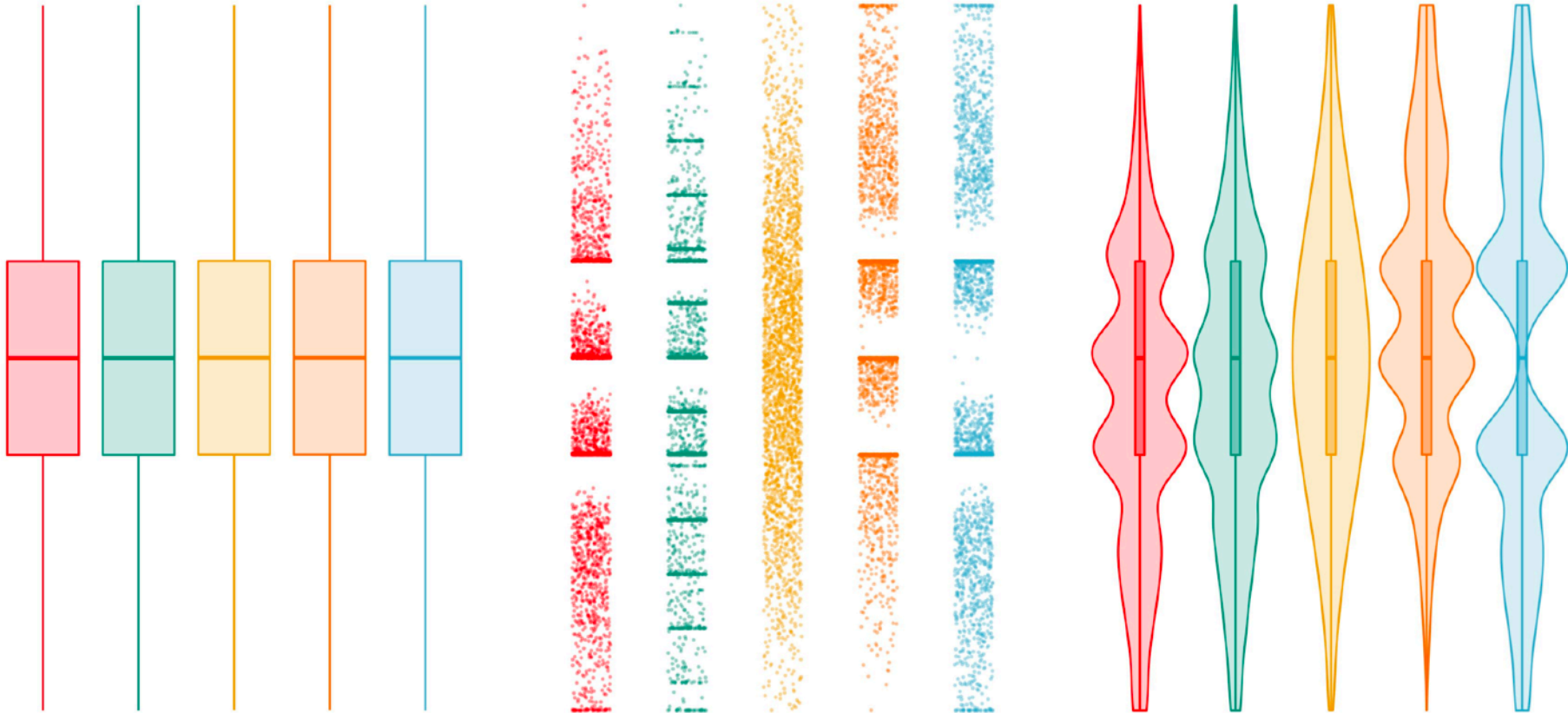
## Violin Plot

kde,  $\sigma = 0.5$

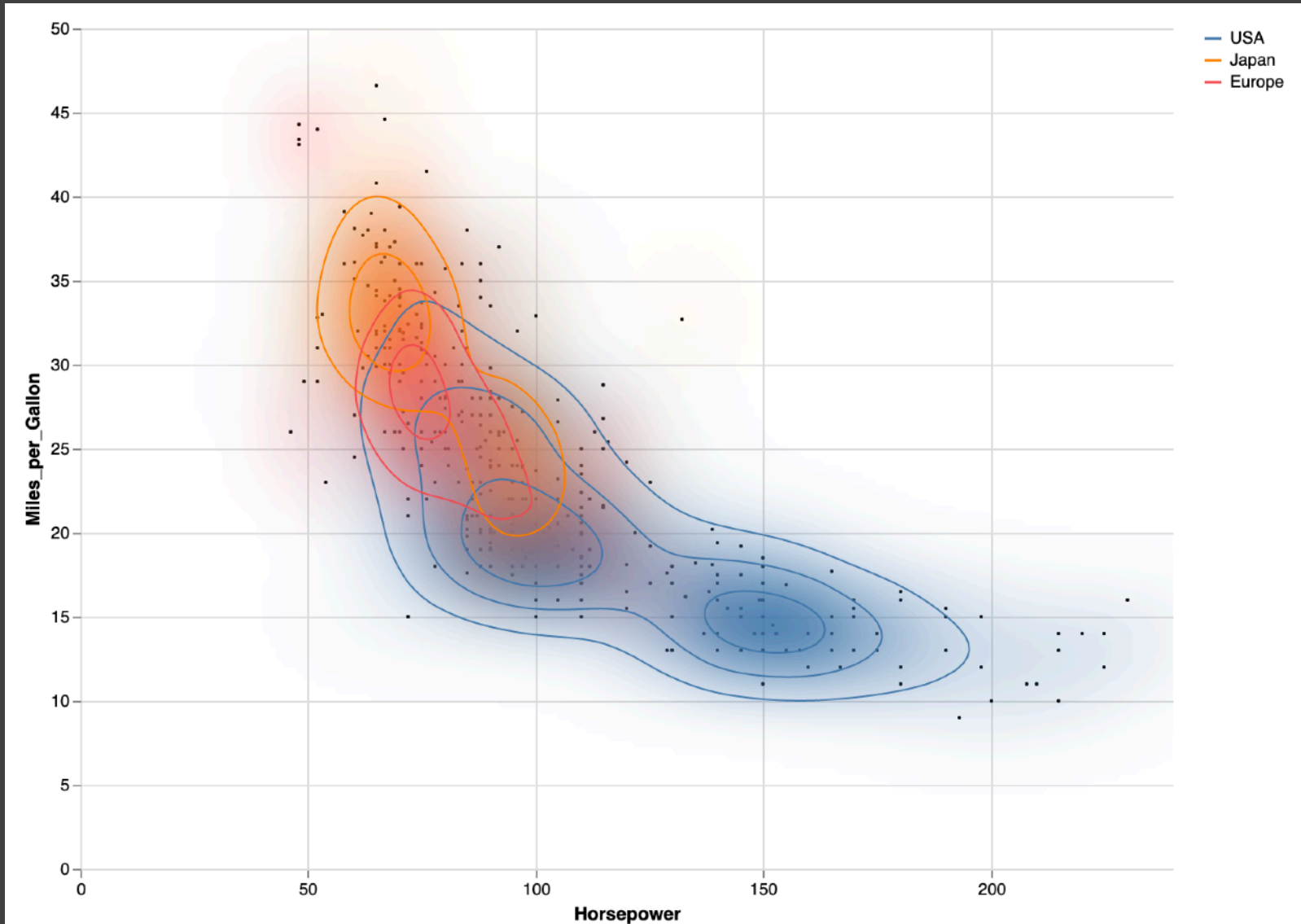


## Identical boxplots, different distributions

Boxplots are great. They show medians and ranges and enable comparison of different groups. However, boxplots can be misleading. Different datasets can have the same descriptive statistics (left), but quite different underlying distributions (middle). Therefore, it is crucial to visualize the distribution in addition to descriptive statistics. Violin plots with integrated boxplots are great for this.



# Now in 2D! Heatmaps, Contours



# Quantified Uncertainty



# Error Bars

Standard Deviation ( $\sigma$ )

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

1.5 \* IQR (Interquartile Range)

Confidence Intervals

*... and so on*

# Confidence Intervals

What does a 95% confidence interval indicate?

One interpretation is: there is a 95% chance that the population mean is within the interval.

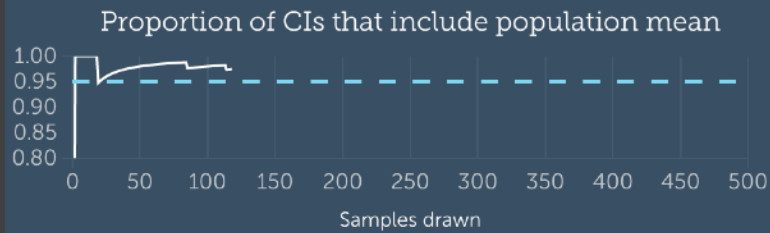
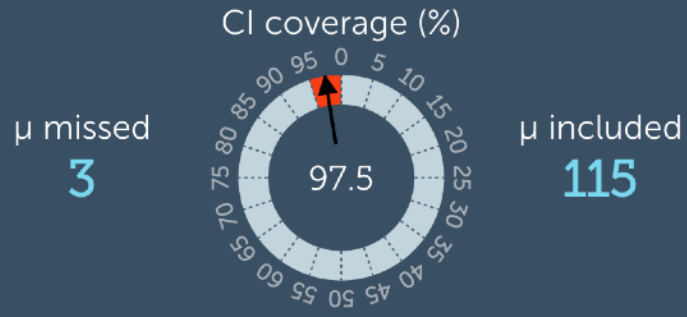
**Wrong!**

Rather, given an infinite number of independent experiments, 95% of the confidence intervals generated will contain the true population mean.

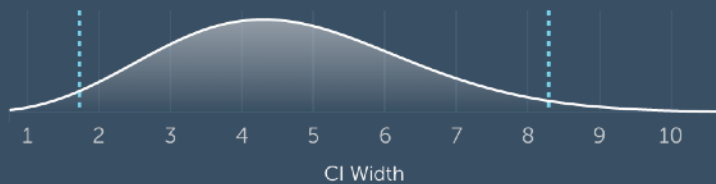
“Confidence” concerns the procedure, not the data. (Though see Bayesian *credible intervals*...)

# Confidence Intervals

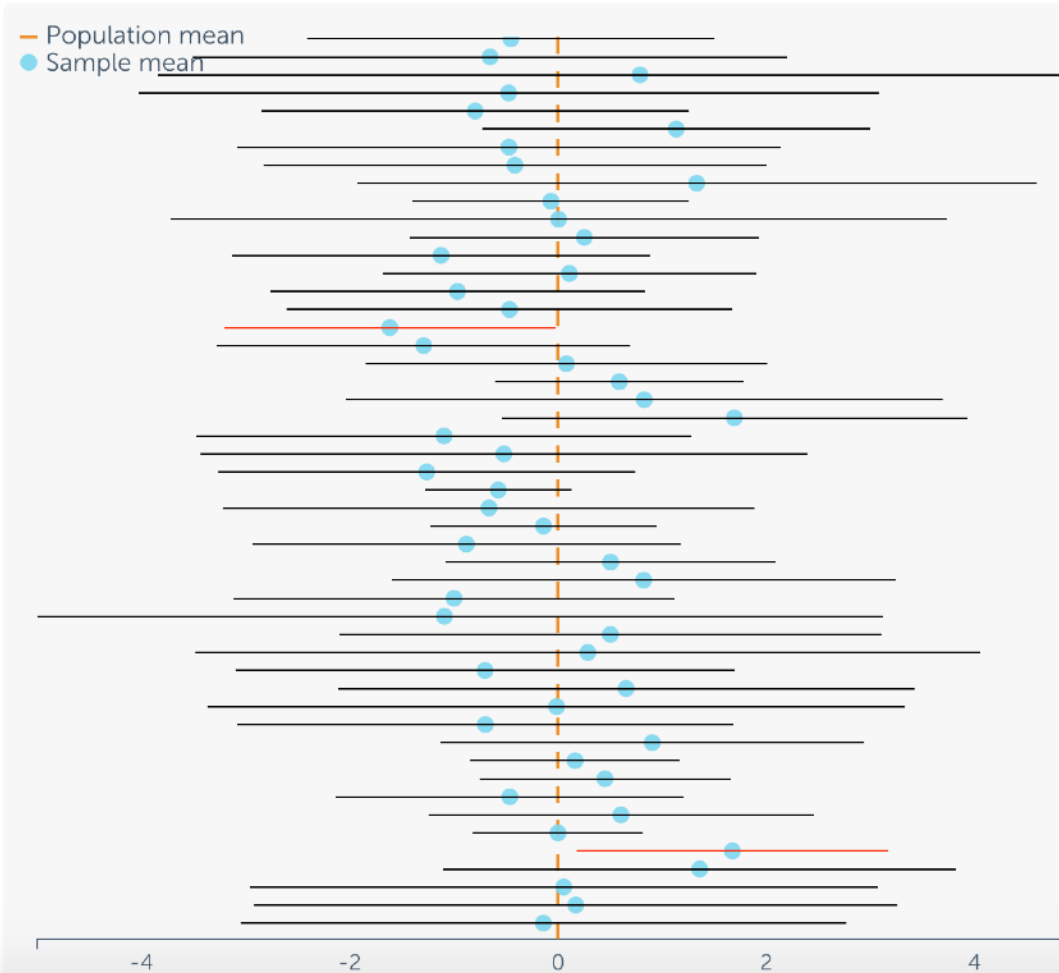
## Simulation statistics



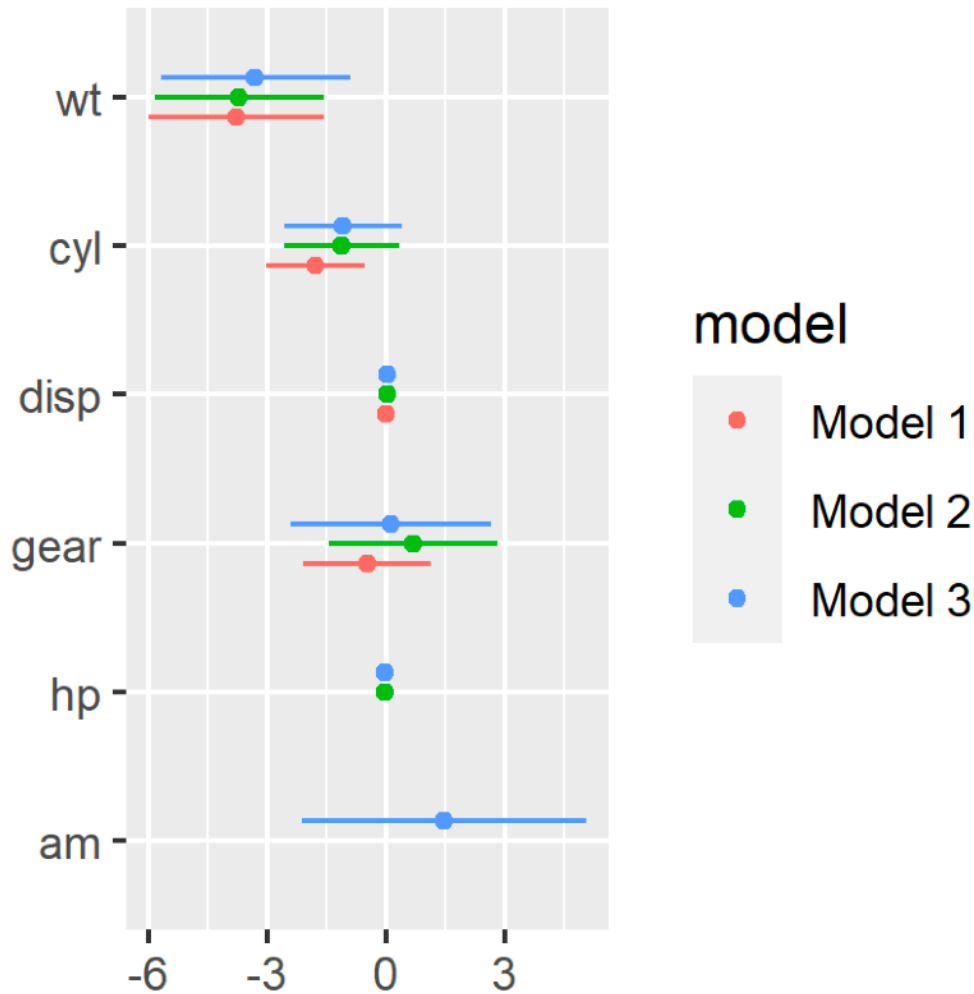
## CIs sampling distribution



## 95% confidence intervals



# Regression Coefficients



95% CIs for regression model parameters.

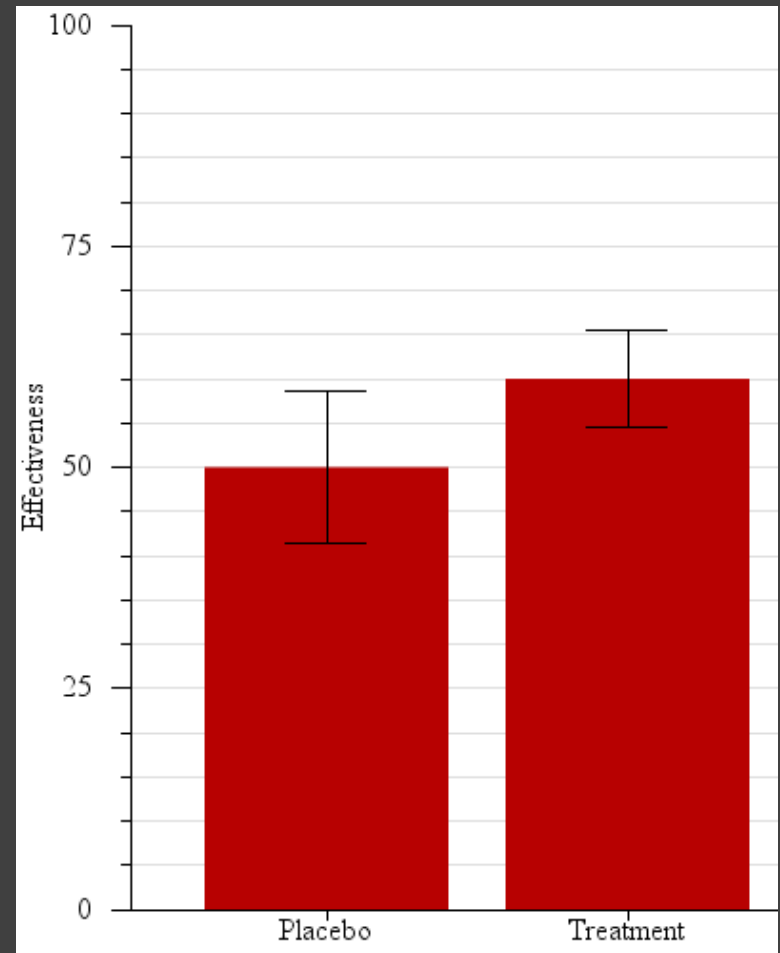
Here, we compare fitted parameters from 3 different models. Not all predictors are included in all models.

Visual comparison: does the CI overlap 0?

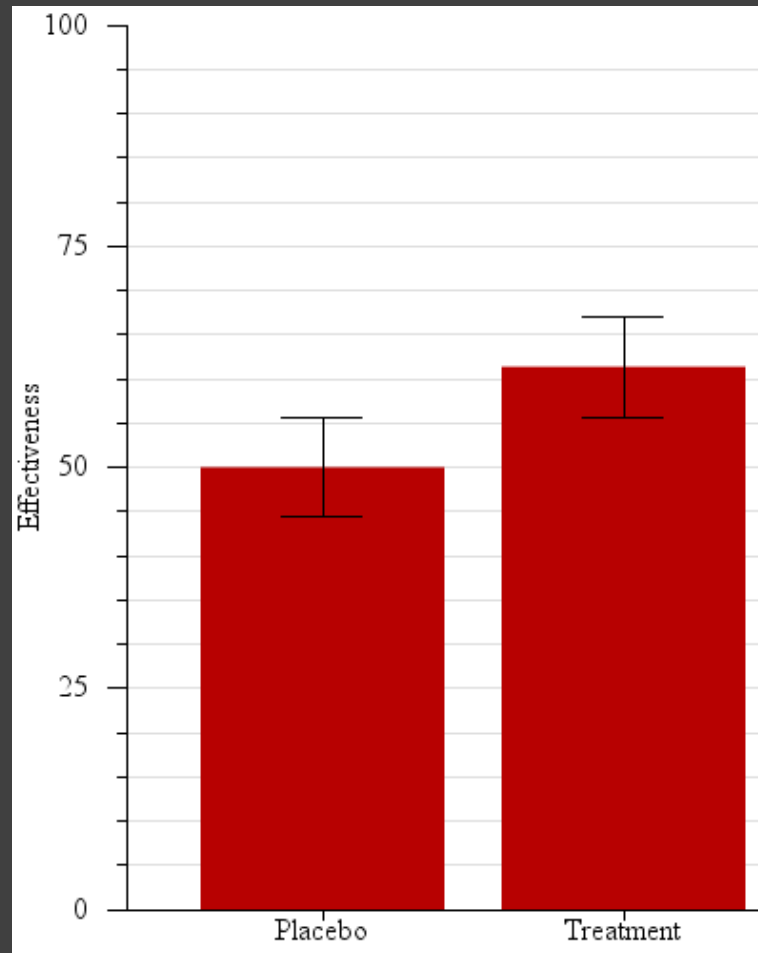
# Error Bars

The mean treatment effect is higher than than the placebo.

Is this difference in means *statistically significant*?

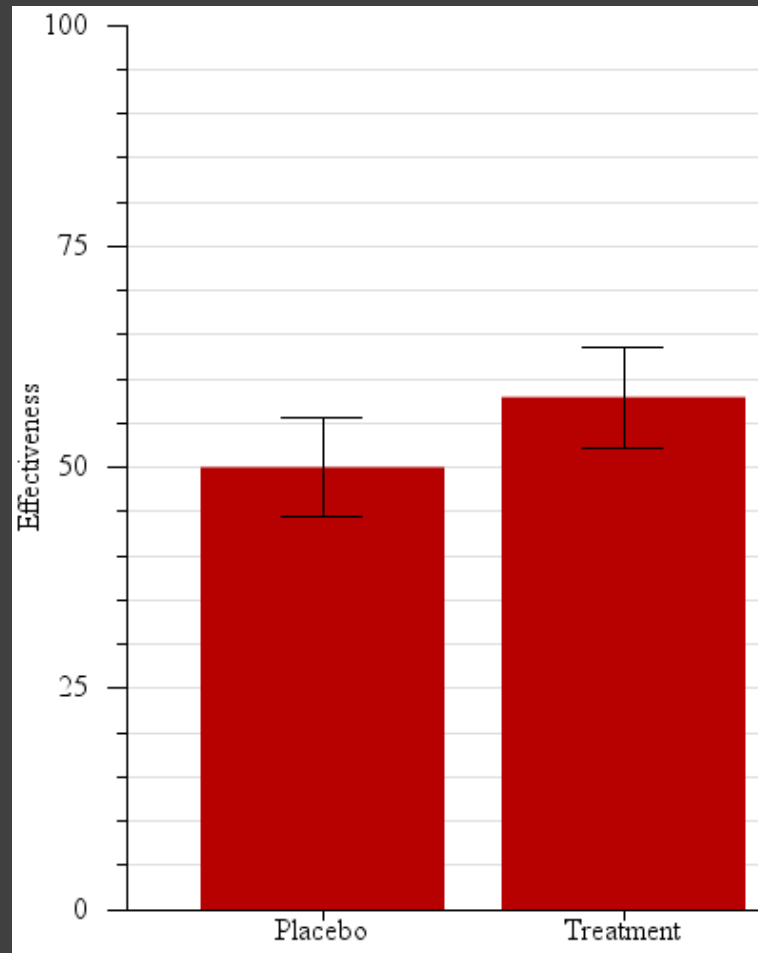


# Guess the p-value...



Error bars depict  
*95% Conf. Interval*

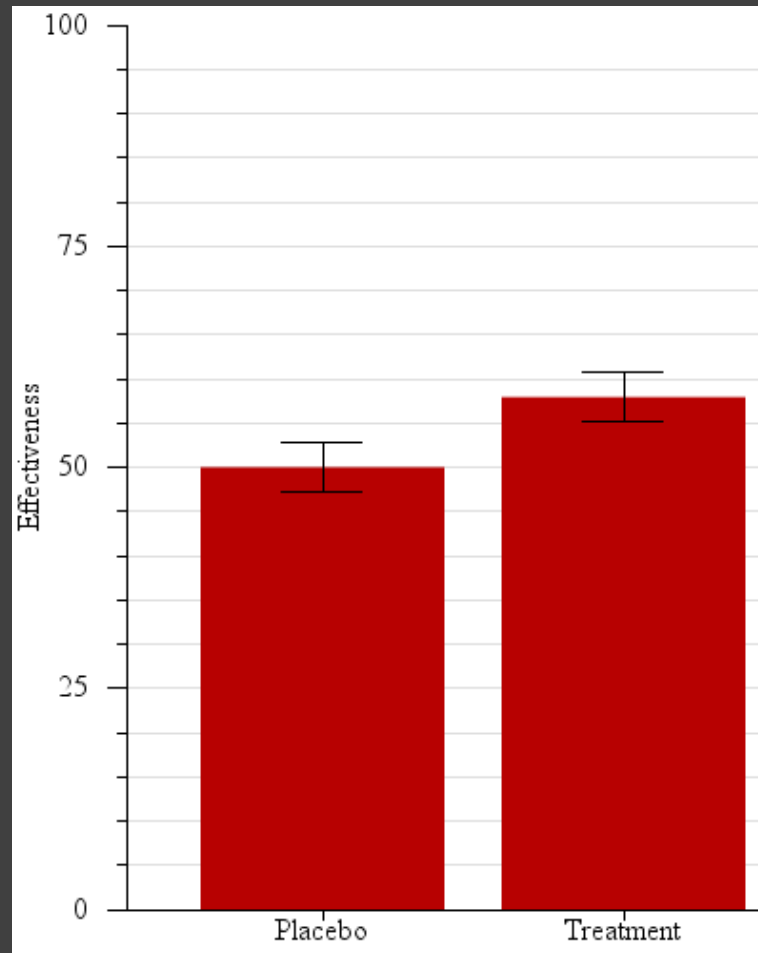
# Guess the p-value...



Error bars depict  
*95% Conf. Interval*

# Guess the p-value...

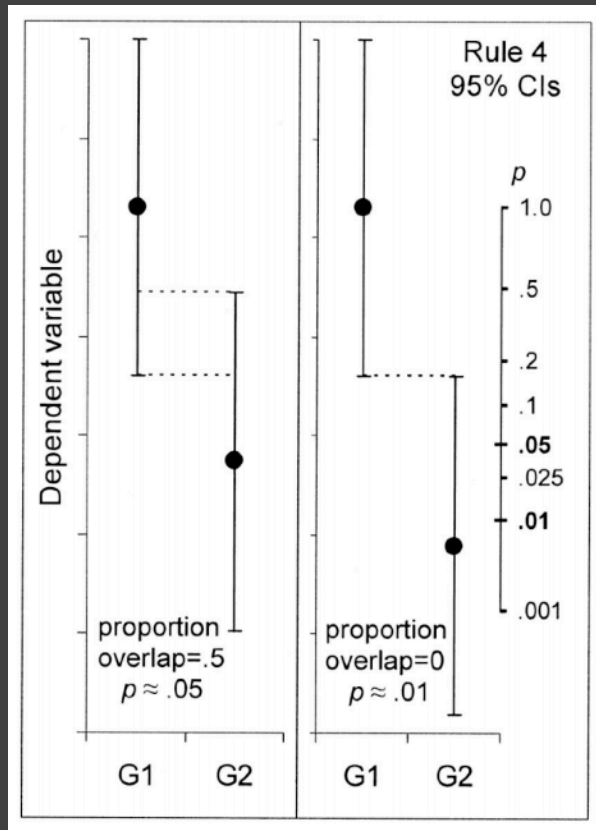
Error bars depict  
*standard error*



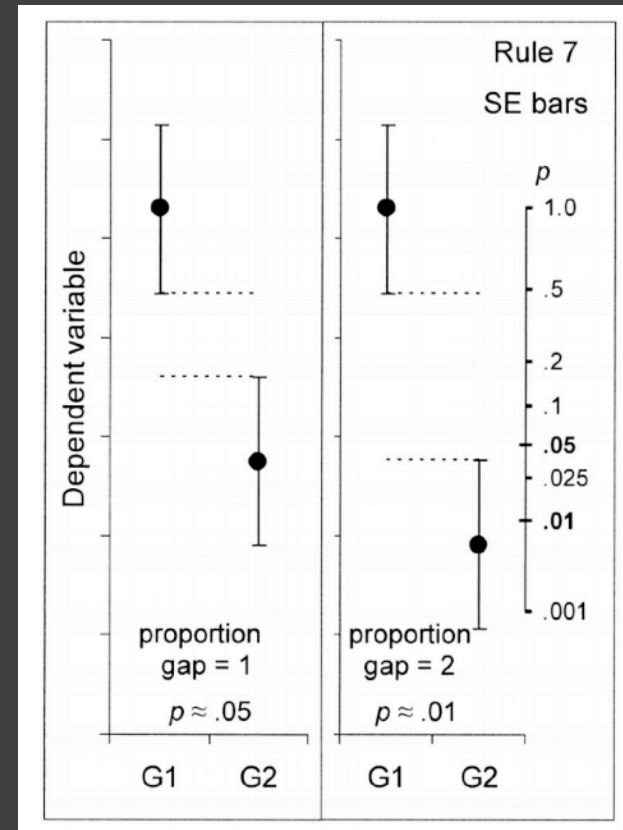


# Inference by Eye

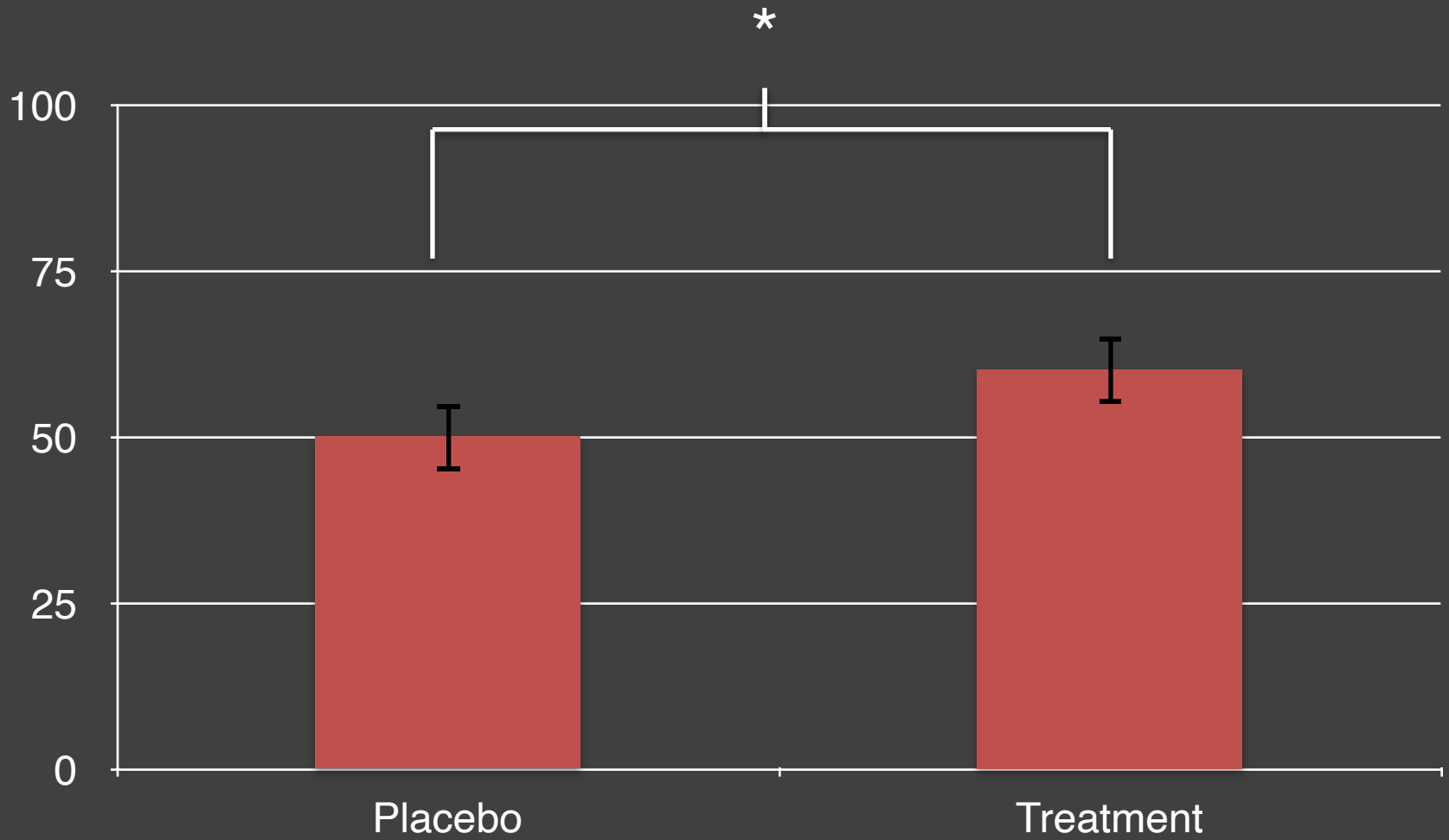
## 95% CIs



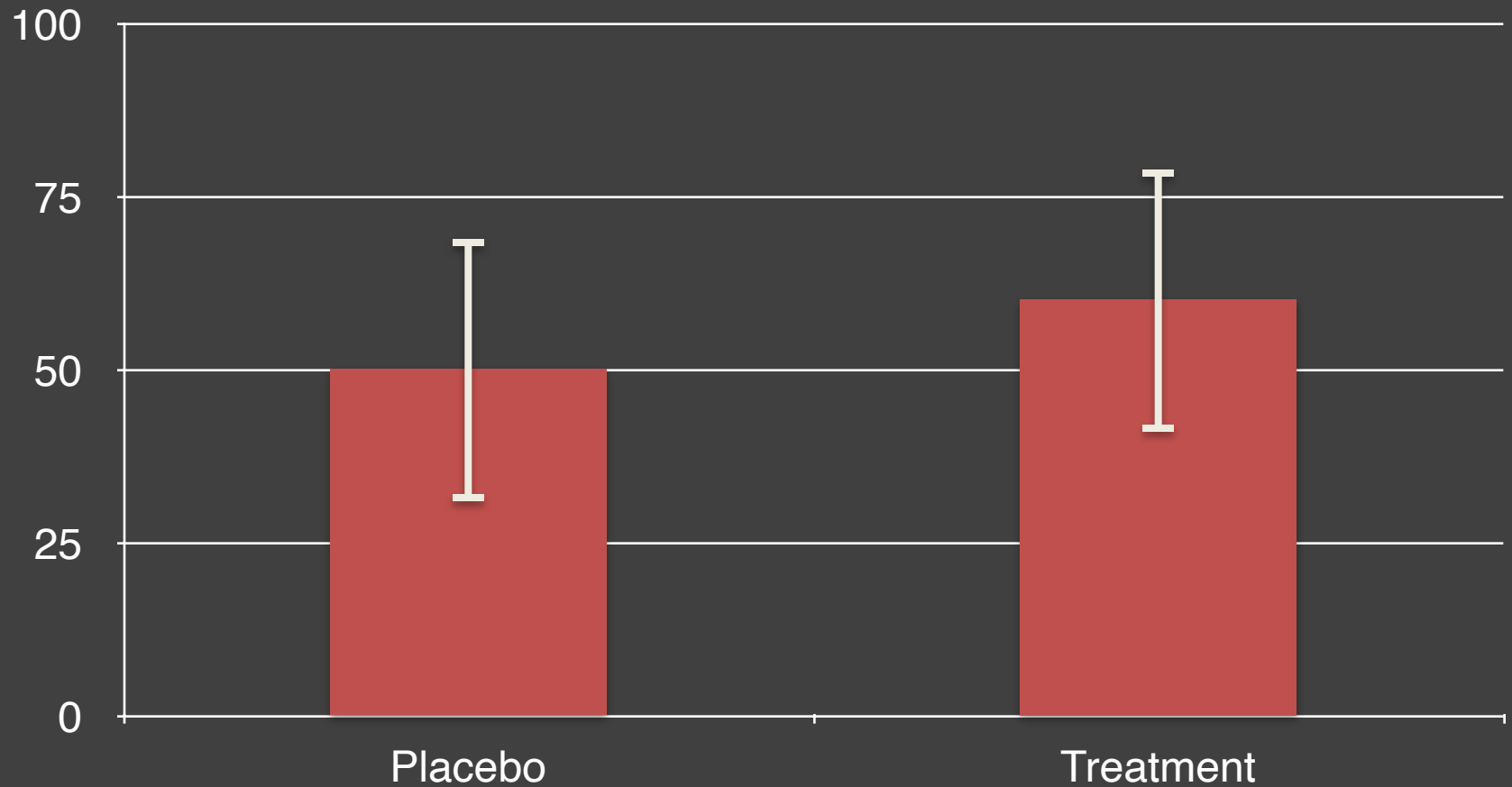
## Standard Error



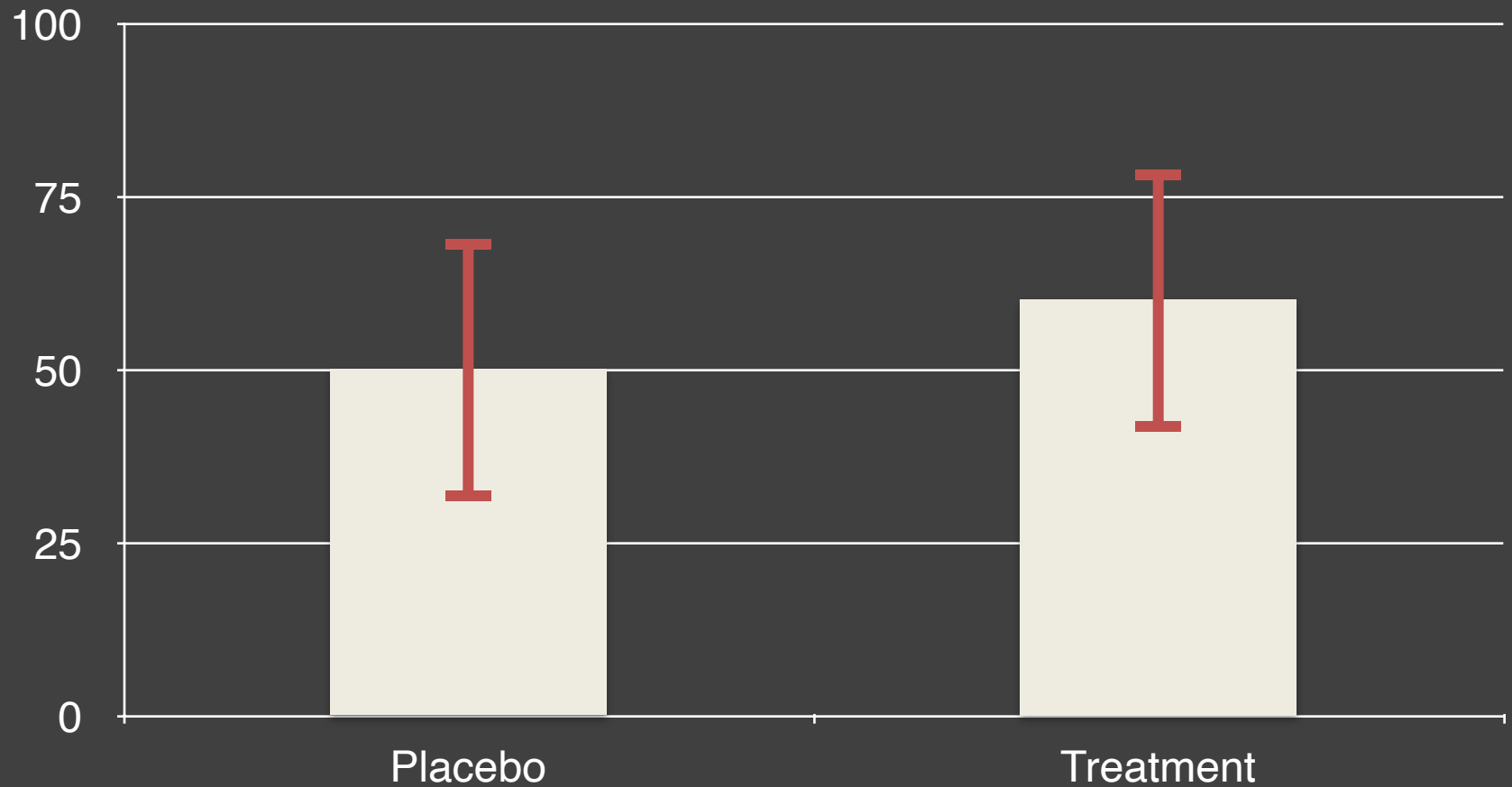
Cumming, Geoff and Finch, Sue. Inference by eye: confidence intervals and how to read pictures of data. American Psychologist, 2005.



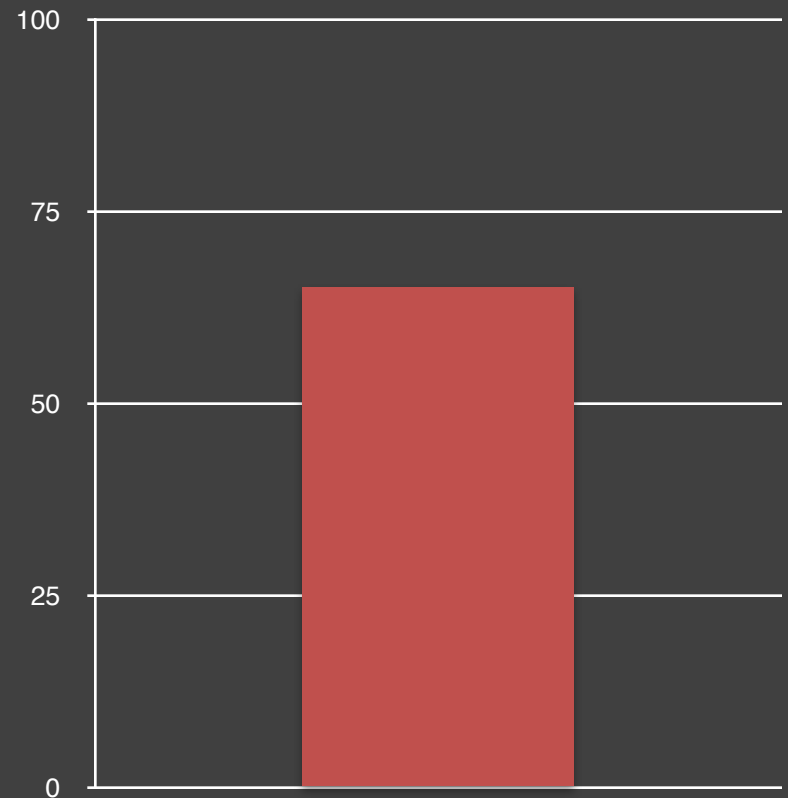
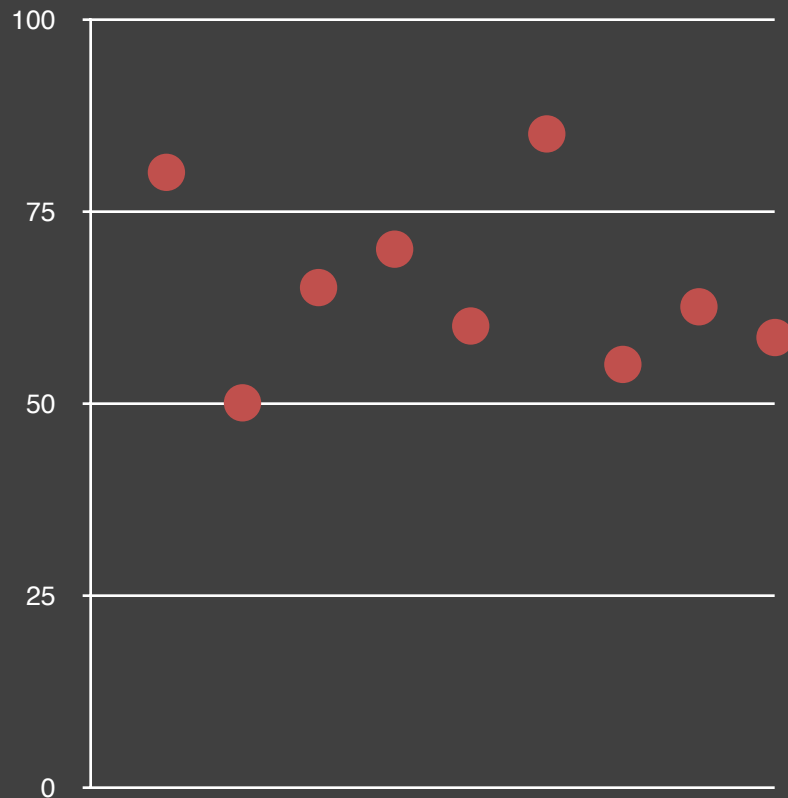
# Misplaced Emphasis?



# Misplaced Emphasis?

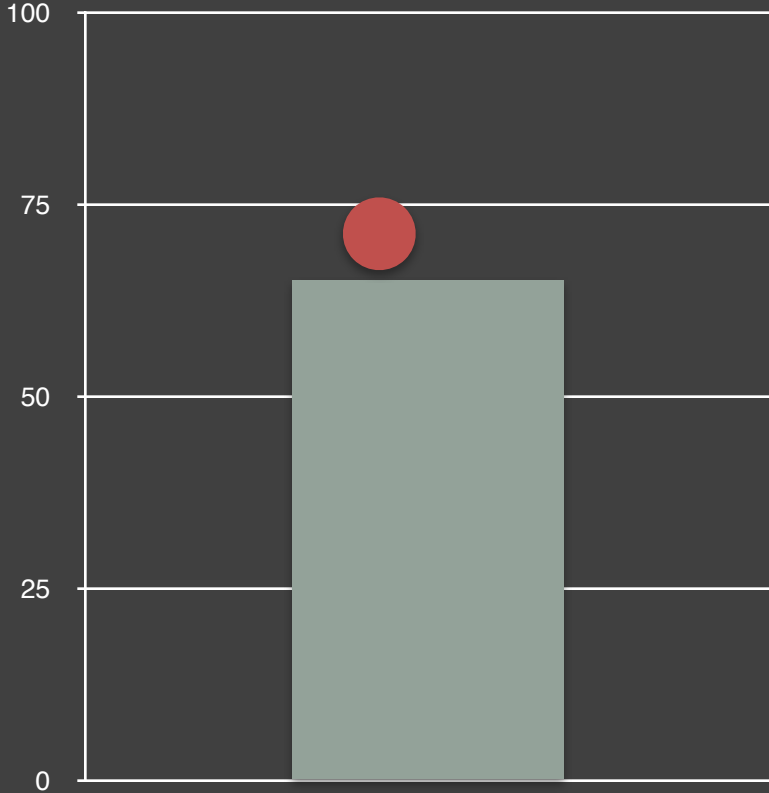
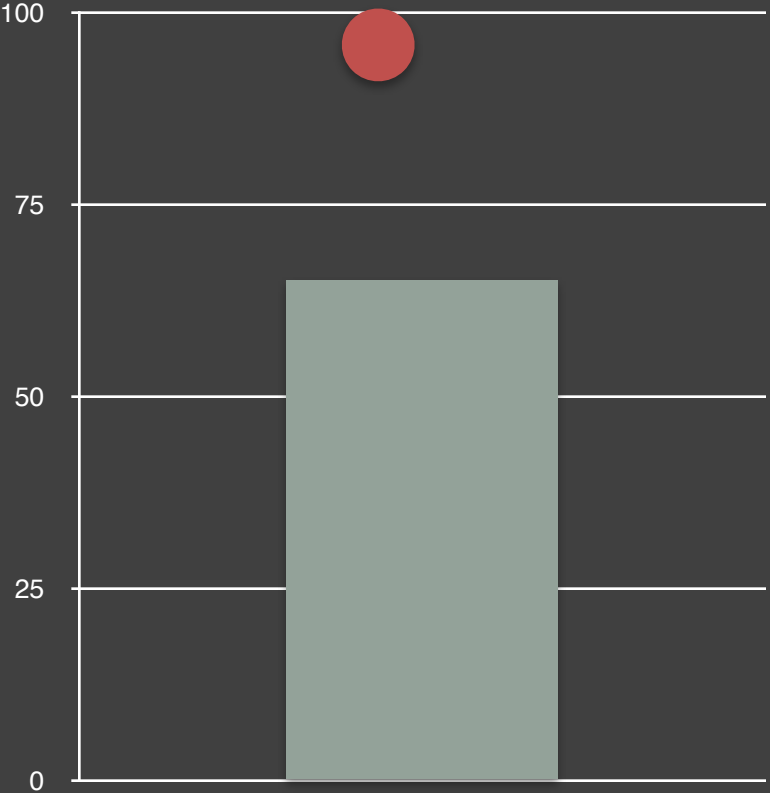


# Within-the-Bar Bias

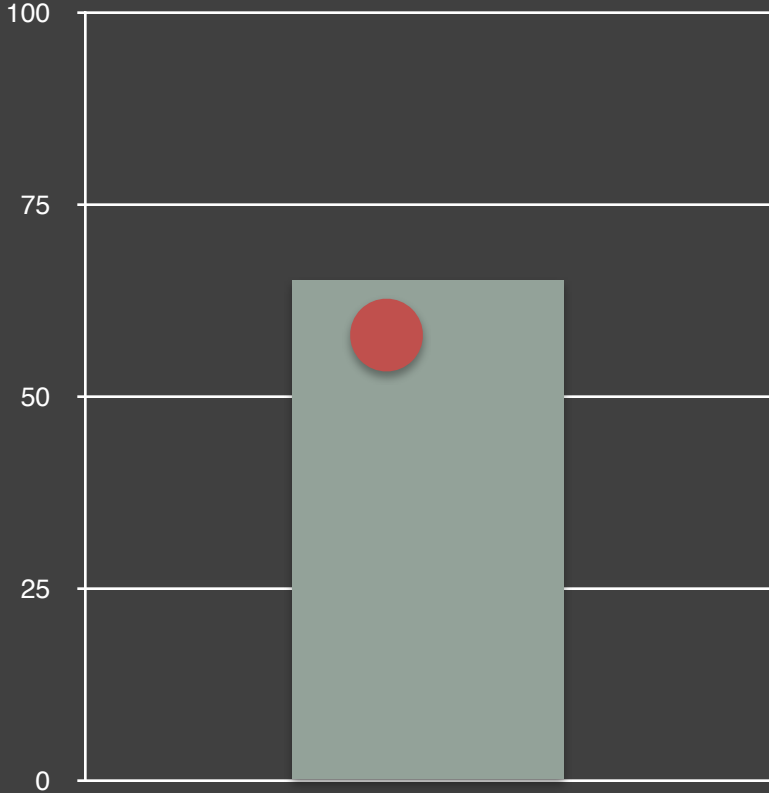
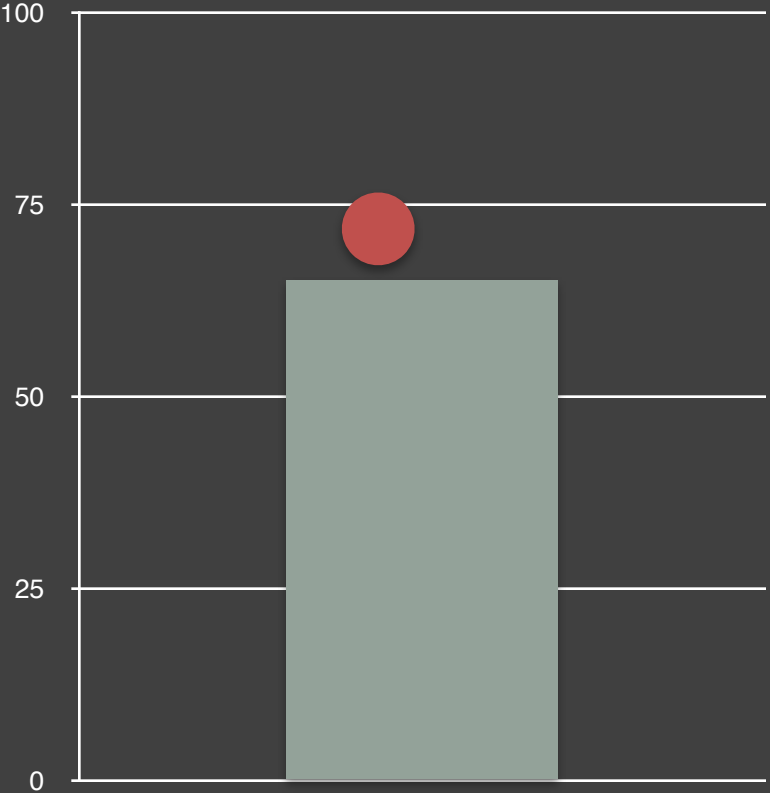


Newman & Scholl. (2012) "Bar graphs depicting averages are perceptually misinterpreted: the within-the-bar bias."

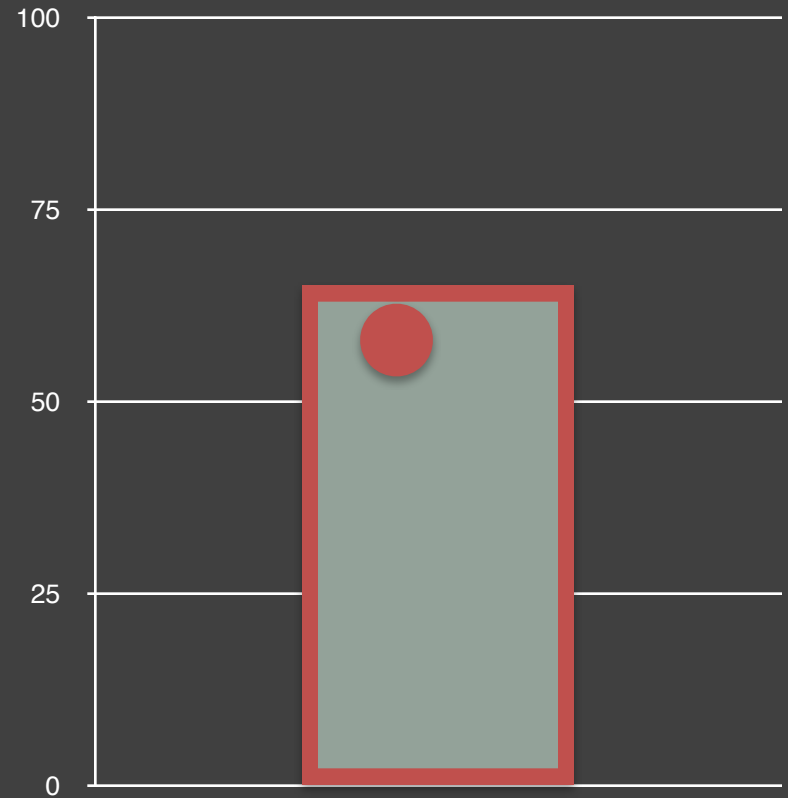
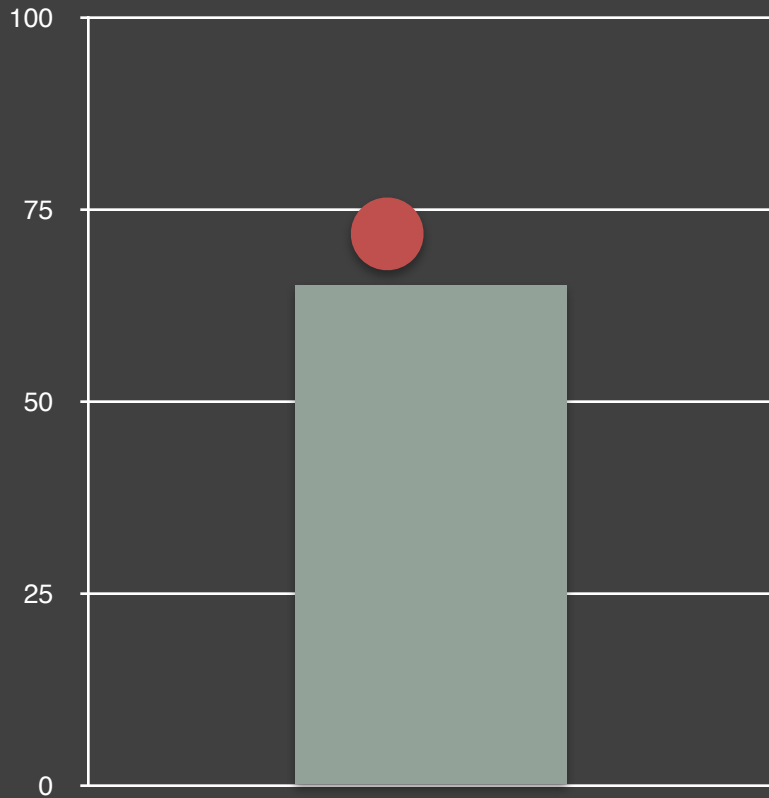
# Within-the-Bar Bias



# Within-the-Bar Bias



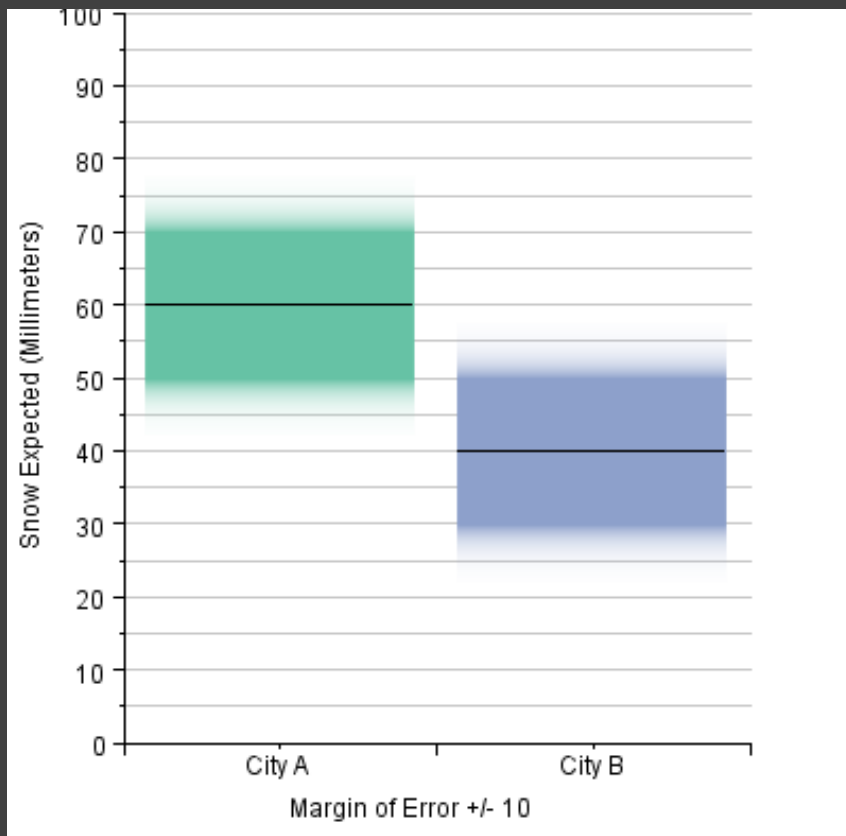
# Within-the-Bar Bias



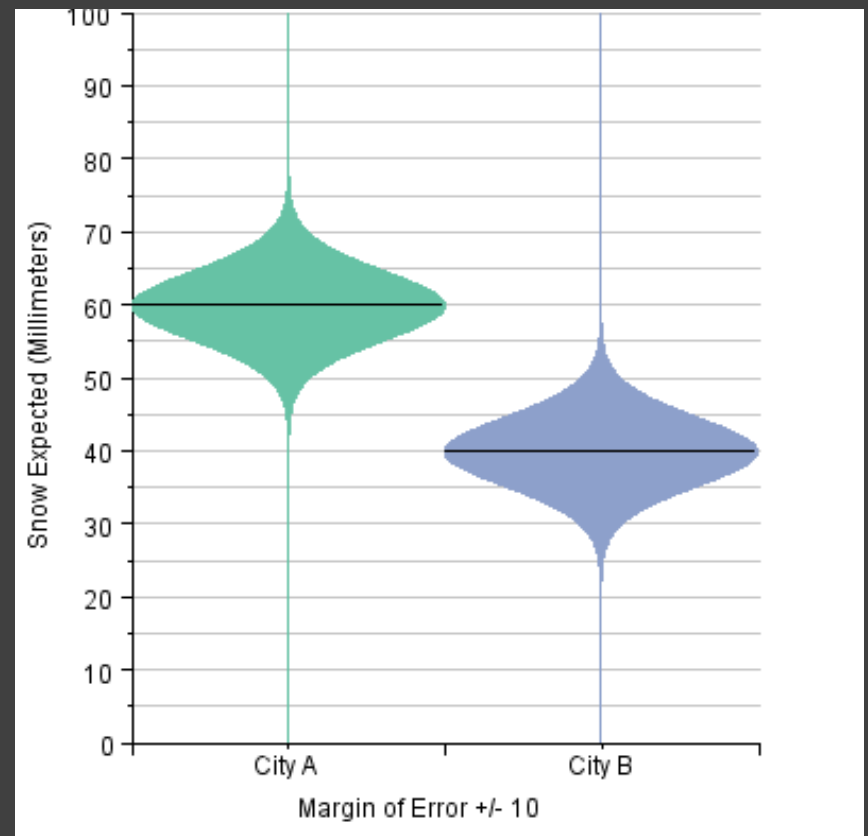


# Alternatives to Error Bars

## Gradient Plot



## Violin Plot



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

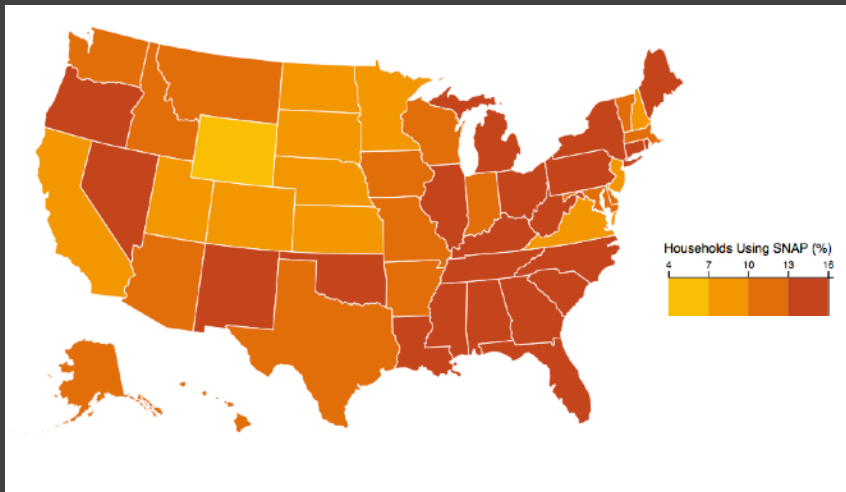
# Encoding Uncertainty

# Uncertainty Vis Pipeline

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

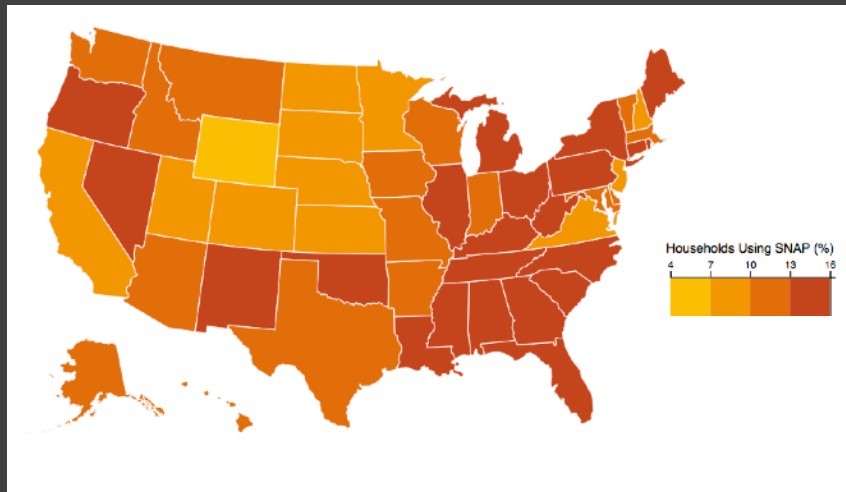
# SNAP

## Data Map

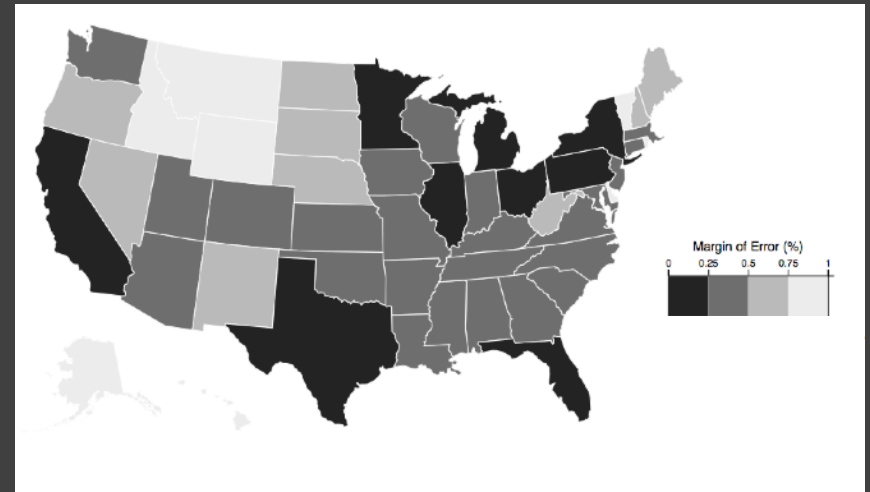


# SNAP

Data Map



Uncertainty Map



# Uncertainty Vis Pipeline

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

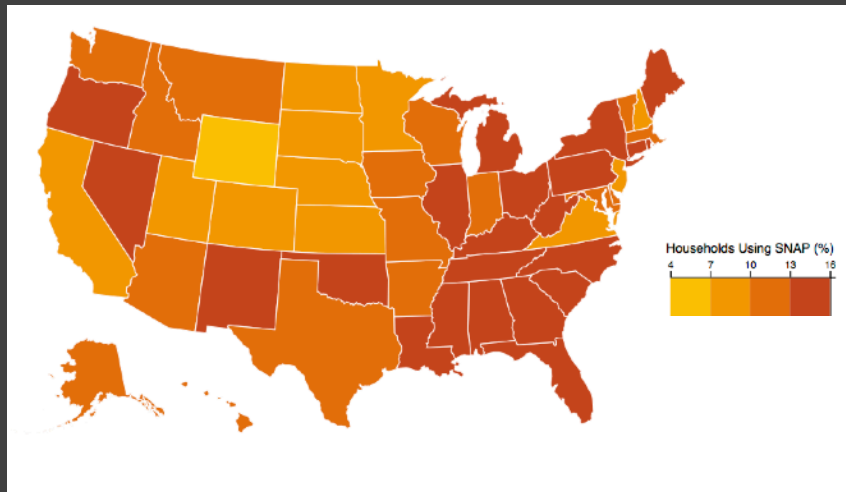
# Uncertainty Vis Pipeline

- 1) Quantify uncertainty
- 2) Choose a free visual variable
- 3) Encode uncertainty with the variable
- 4) Unify the Data Map and Uncertainty Map

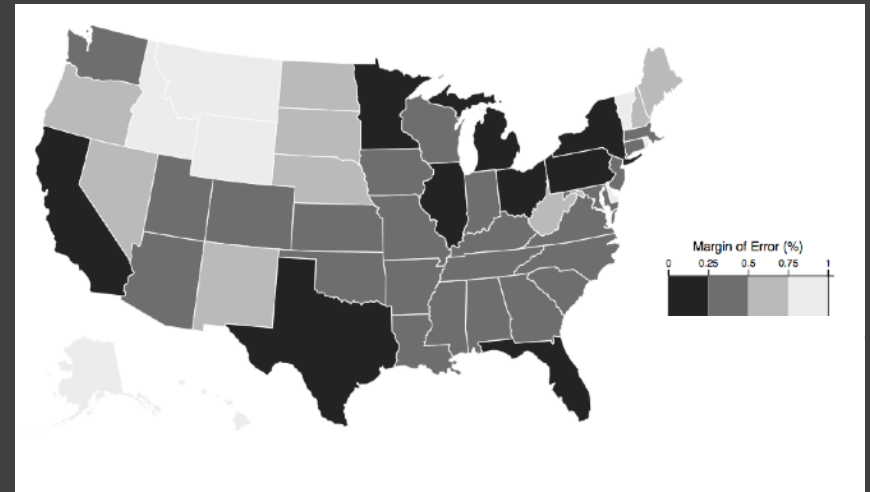


# How to Unify?

Data Map

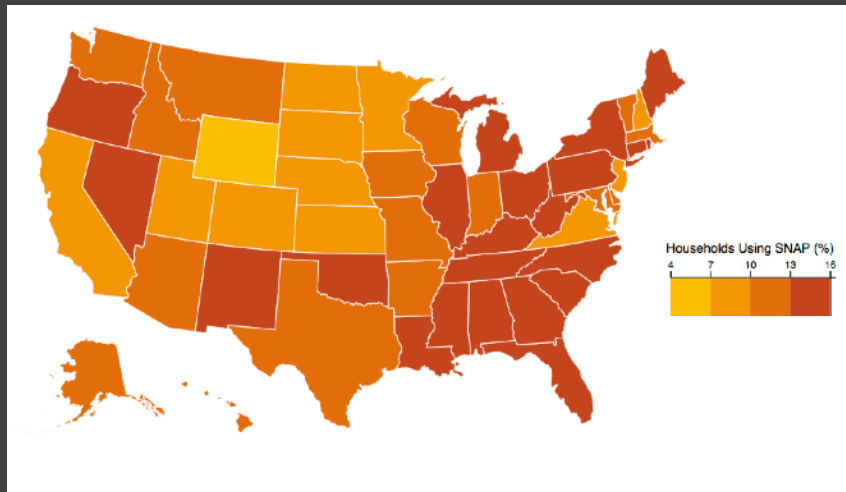


Uncertainty Map

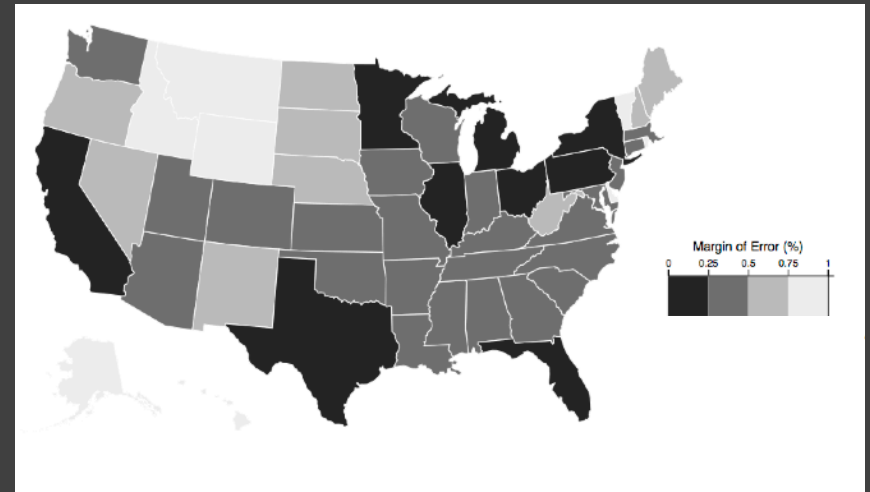


# Juxtaposition

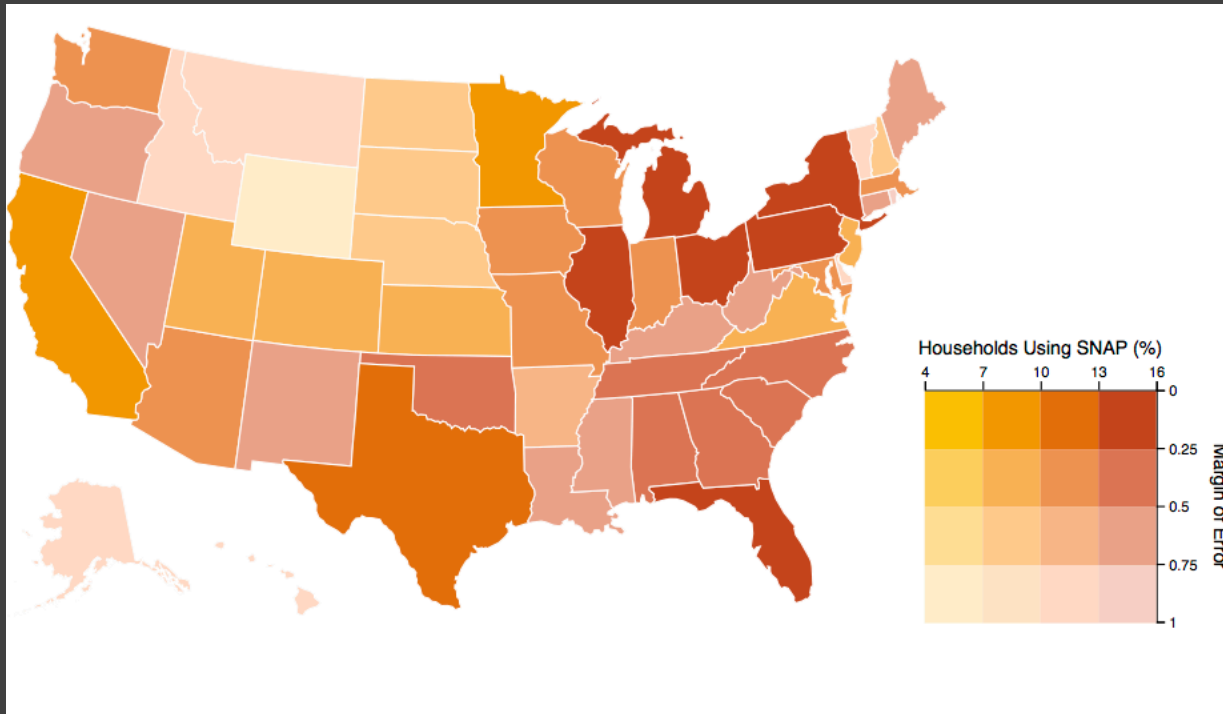
Data Map



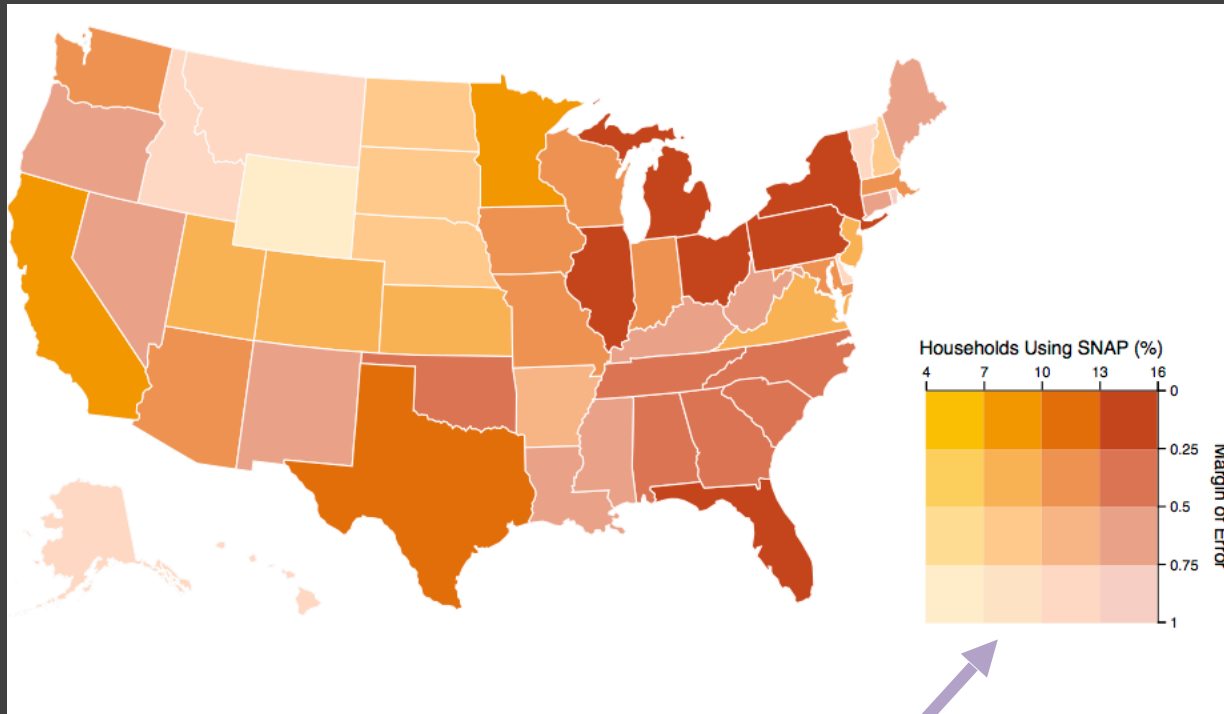
Uncertainty Map



# Superposition

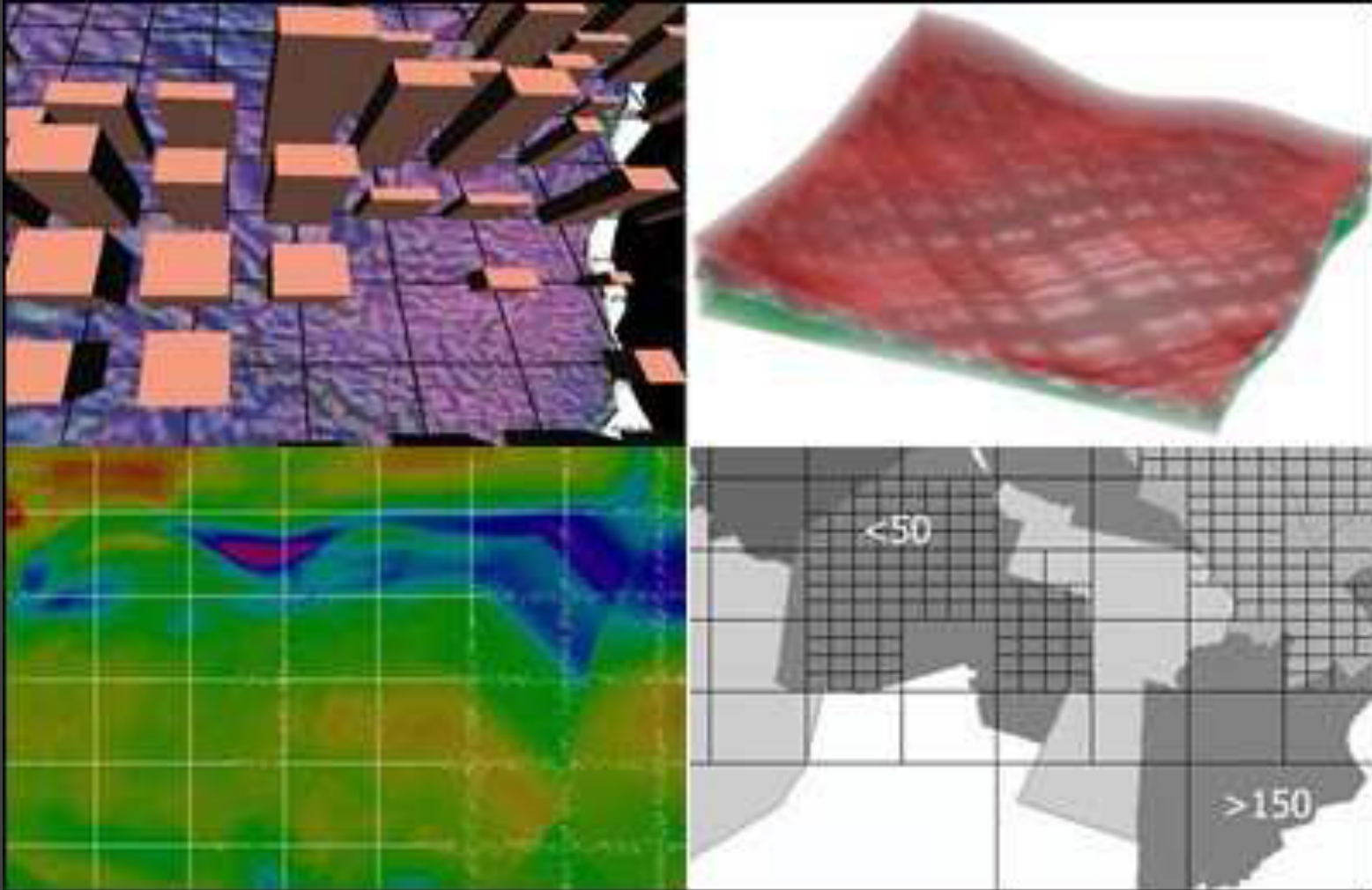


# Superposition



Bivariate Map

# 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
- 4) Unify the Data Map and Uncertainty Map

# Uncertainty Vis Pipeline

- 1) Quantify uncertainty
- 2) Choose a free **visual variable**
- 3) Encode uncertainty with the variable
- 4) Unify the Data Map and Uncertainty Map

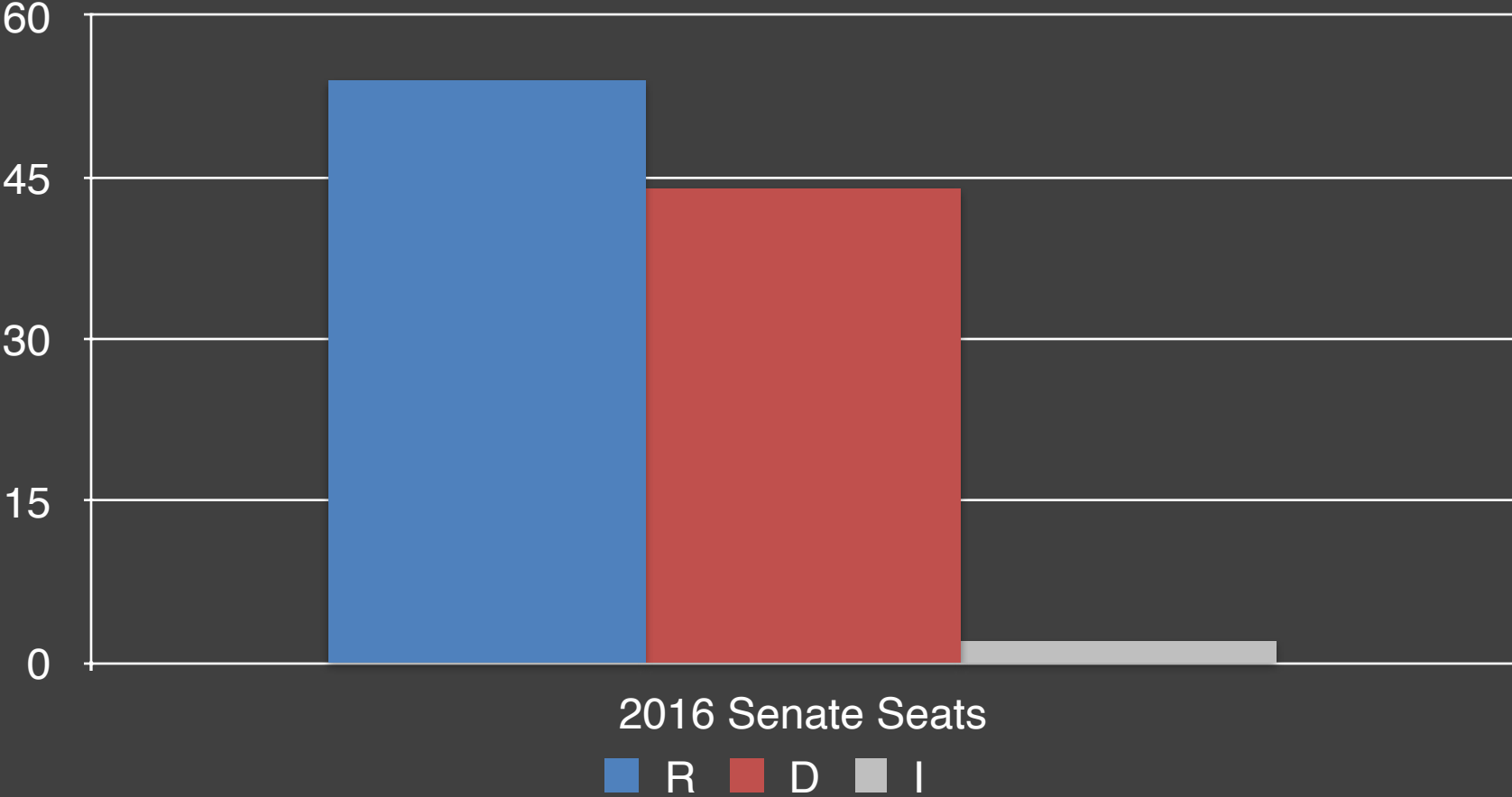
# Semiotics of Uncertainty



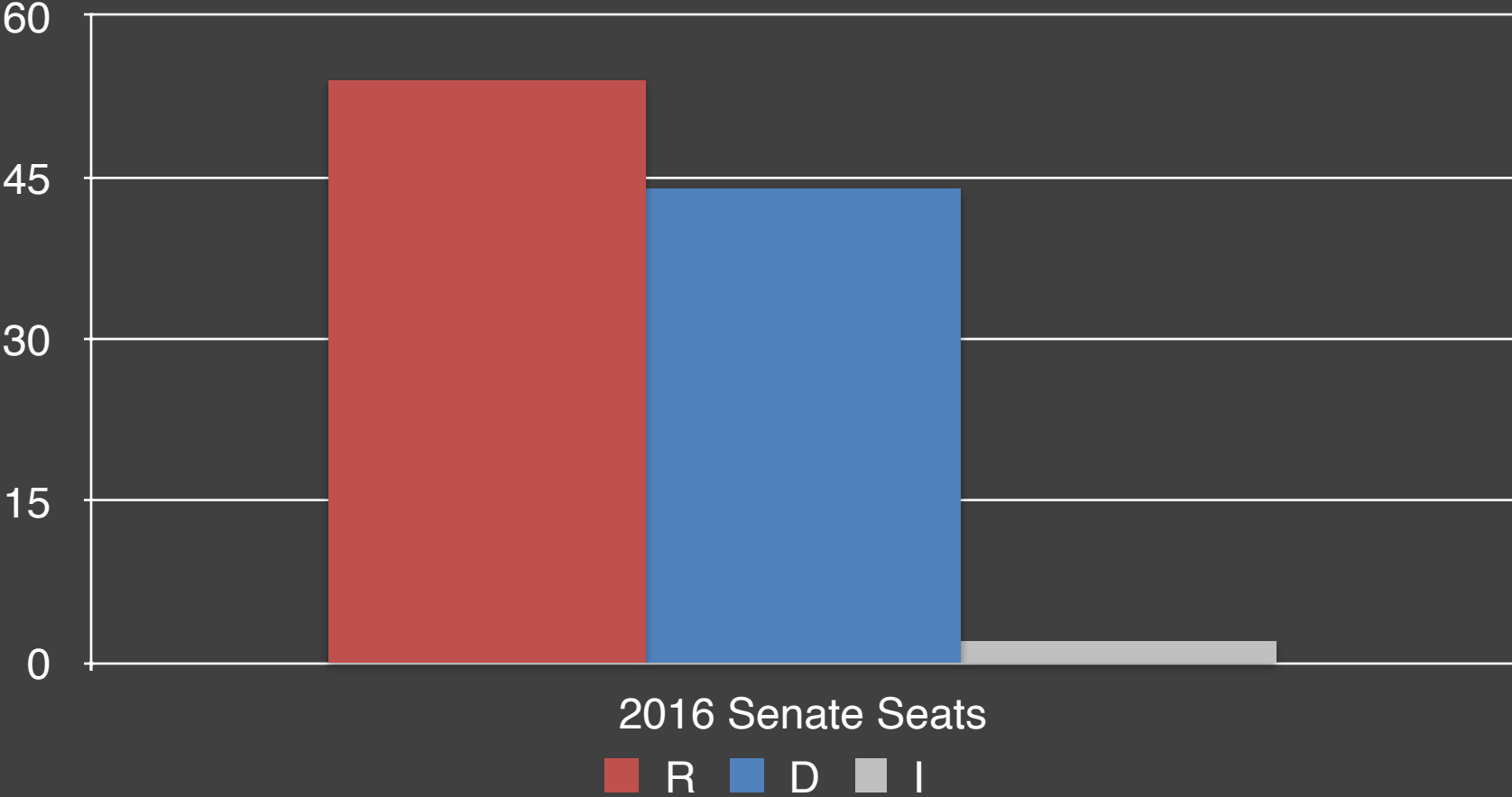
*Ceci n'est pas une pipe.*



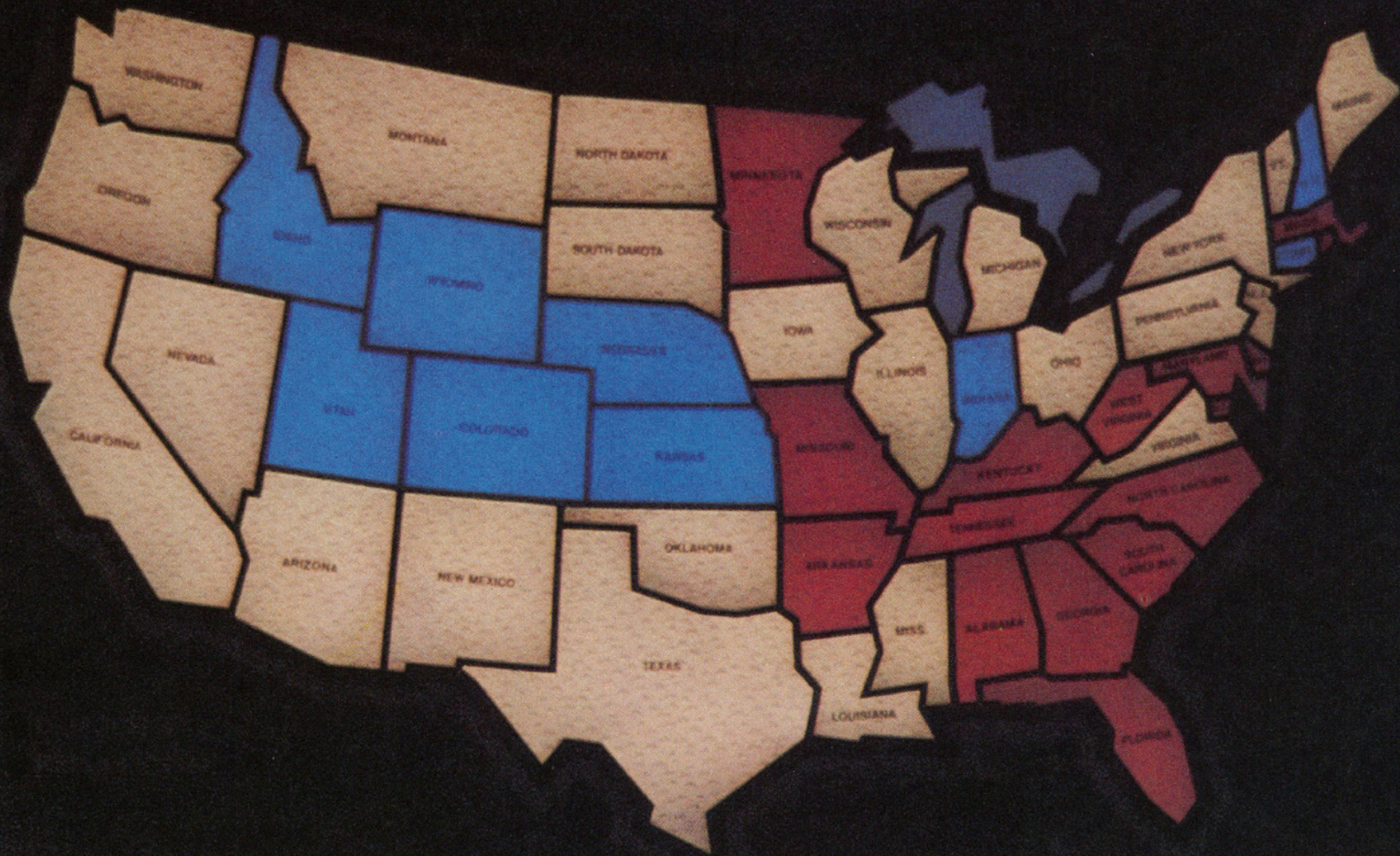
# The Variable Matters!



# The Variable Matters!











**VELOCITY OF MONEY**  
 M1 SUPPLY  
 CURRENT: 6.55  
 5 YEARS AGO: 10.31

EURO-ZLOTY - 10 YEARS  
 2004 2009  
 4.9  
 4.1  
 3.2

**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

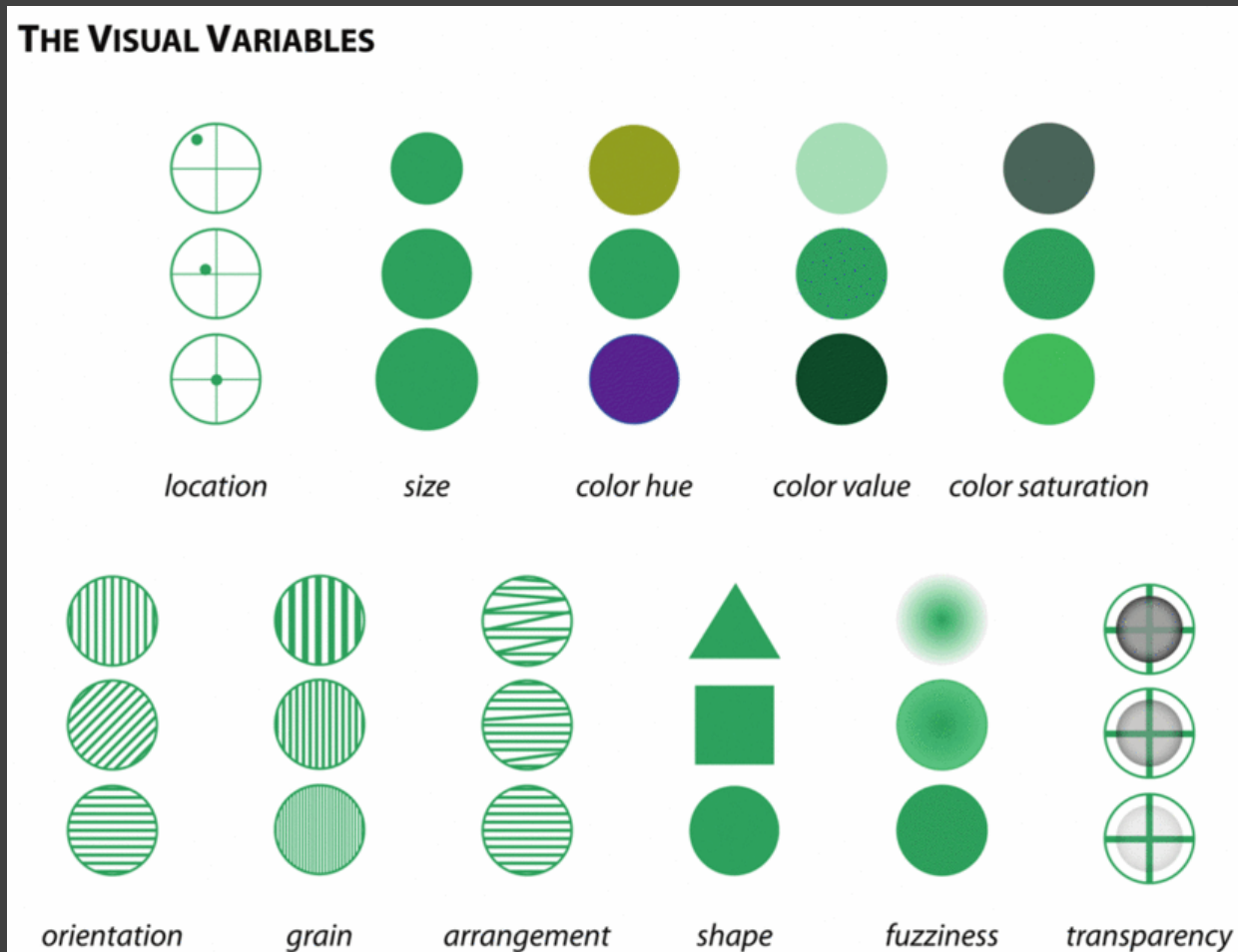
7:24 ET MAY 30 **COSTCO QUARTERLY PROFIT RISES 19% ON INCREASED REVENUE FROM MEMBERSHIP FEES**

Gold	Silver	Plat.	Copper	Alum.
1415.25	22.76	1482.70	351.35	1907.00
↑ 1.11	↑ 0.07	↓ 1.00	↓ 0.20	↑ 44.00

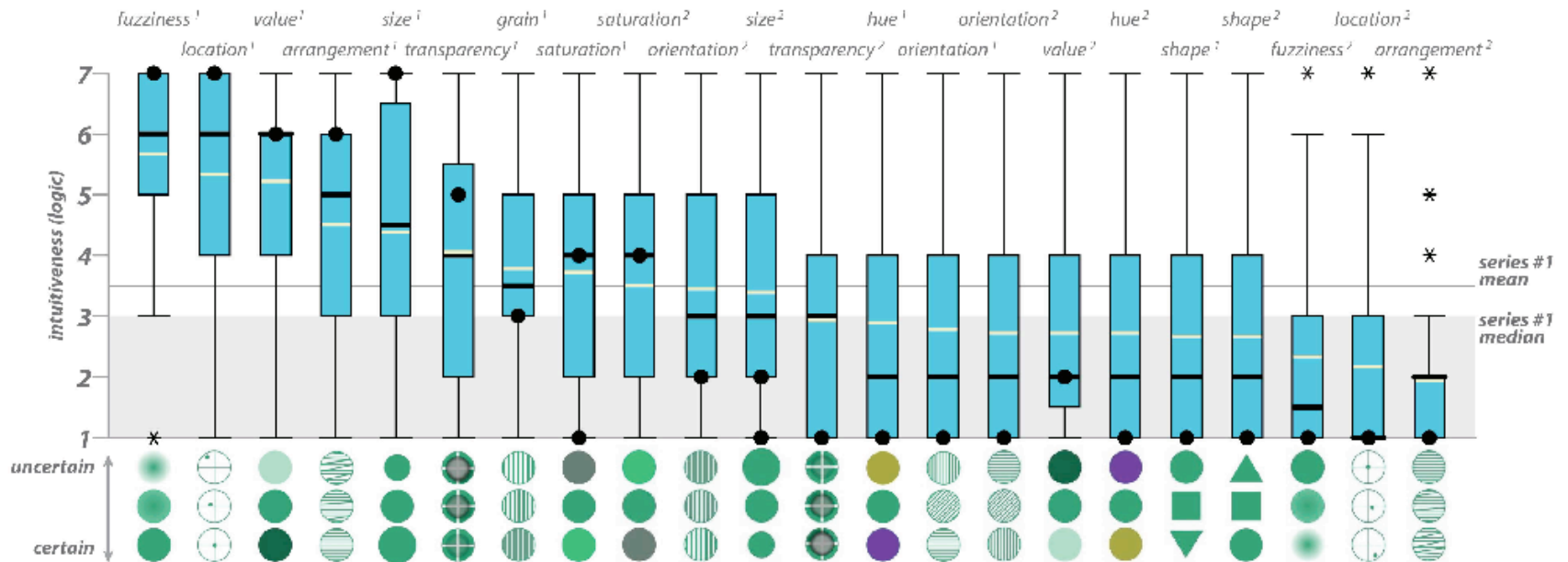
# Semiotics of Uncertainty



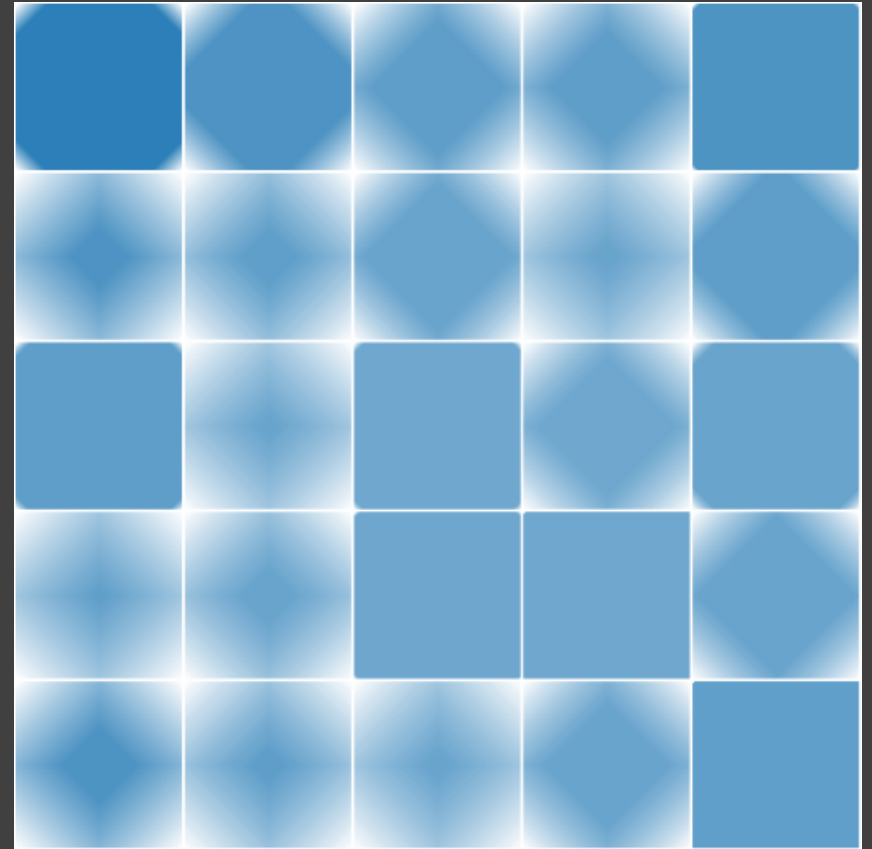
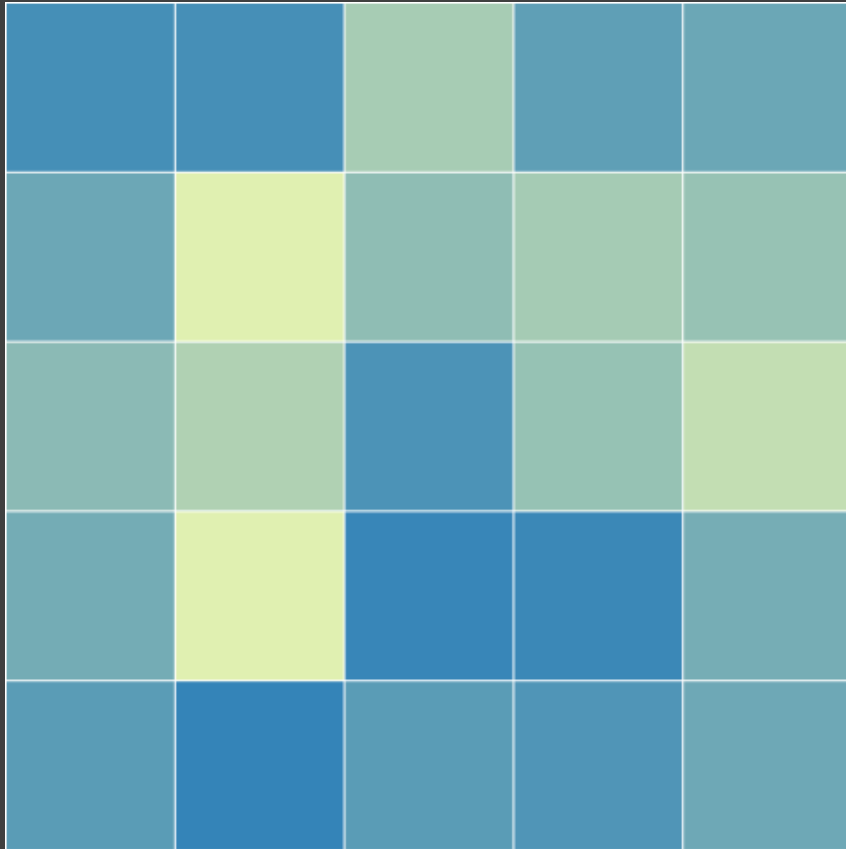
# Semiotics of Uncertainty



# SERIES #1: GENERAL UNCERTAINTY BY VISUAL VARIABLE

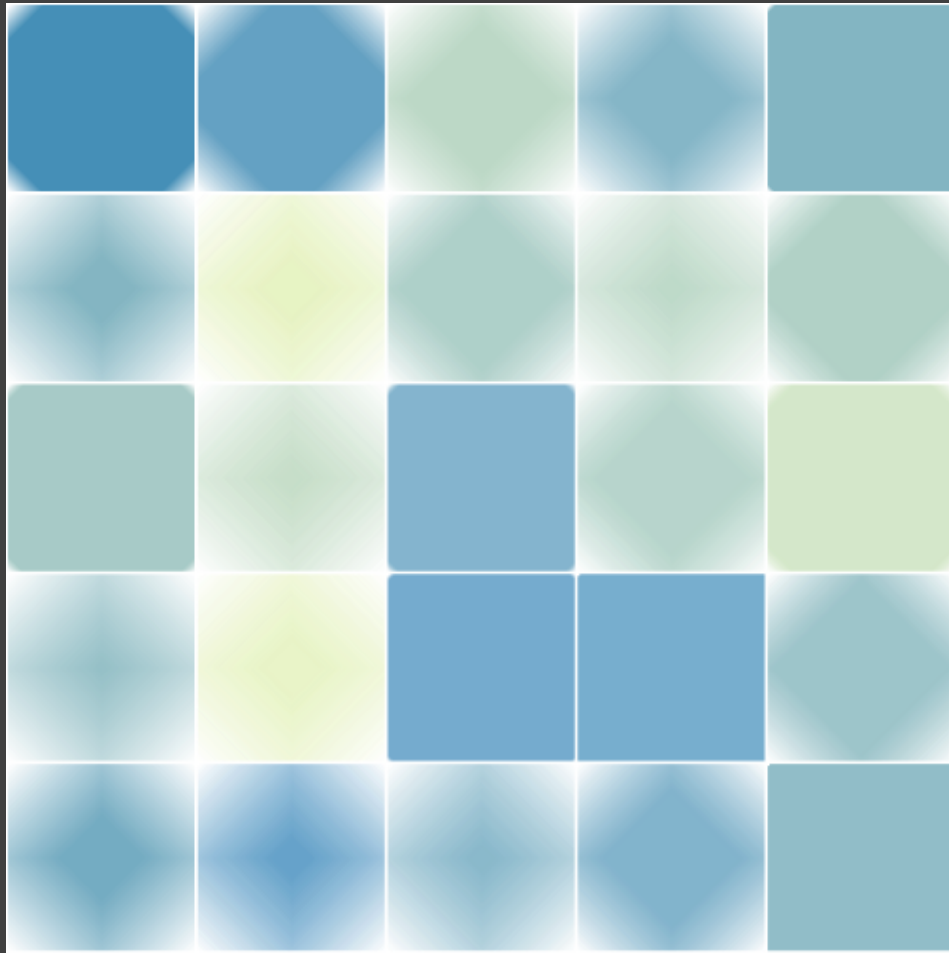


# Fuzziness Juxtaposition

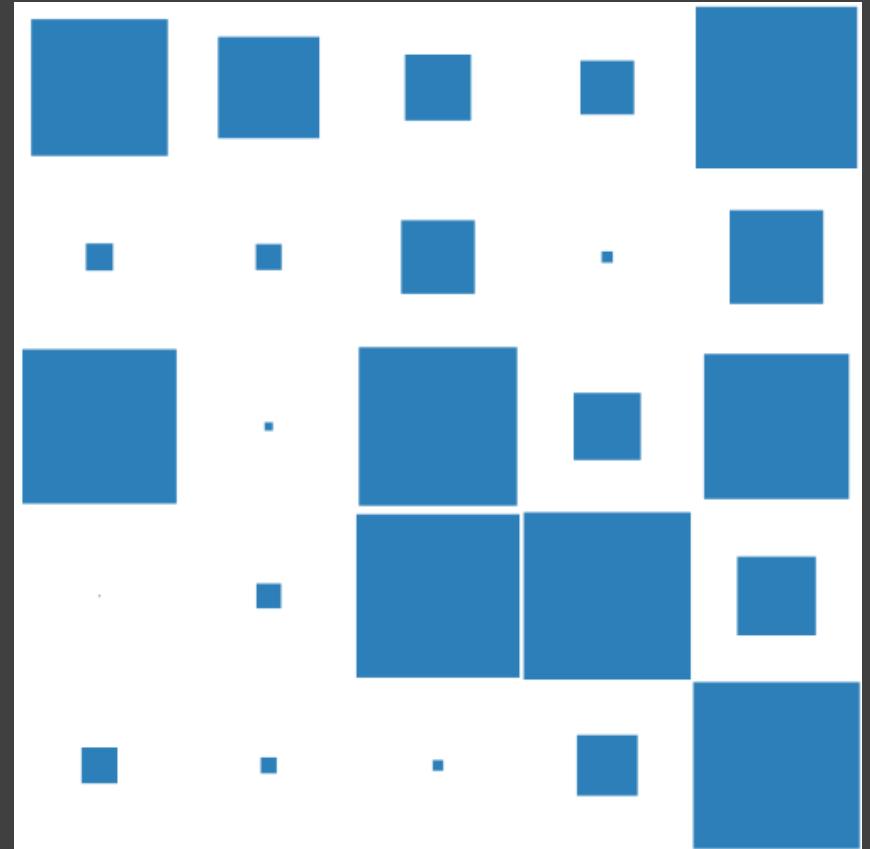
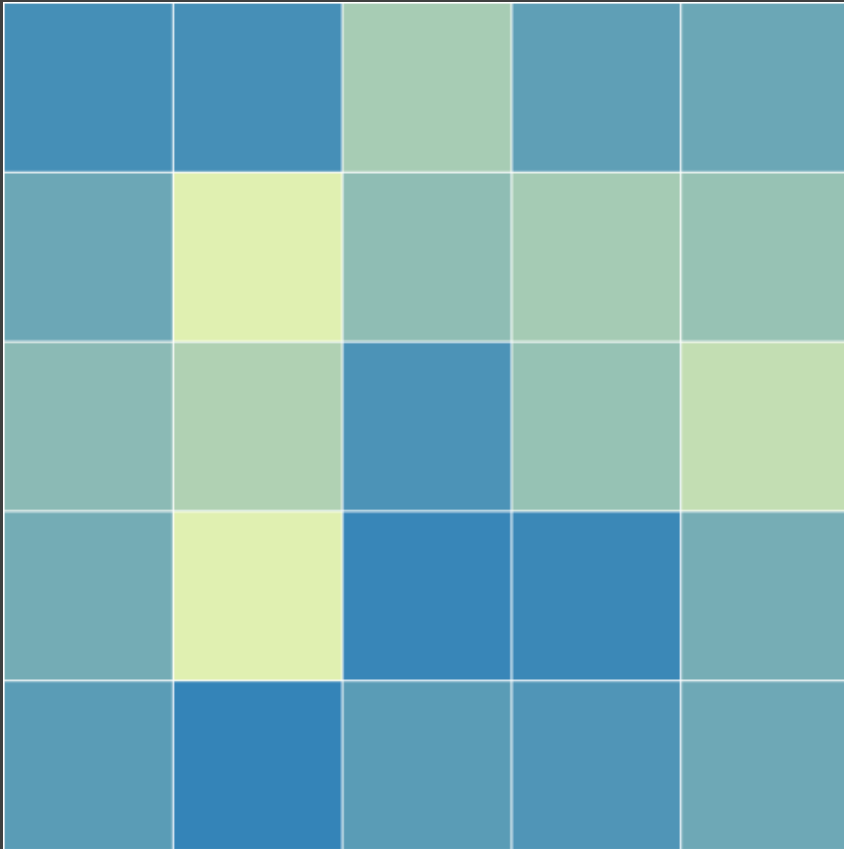




# Fuzziness Superposition



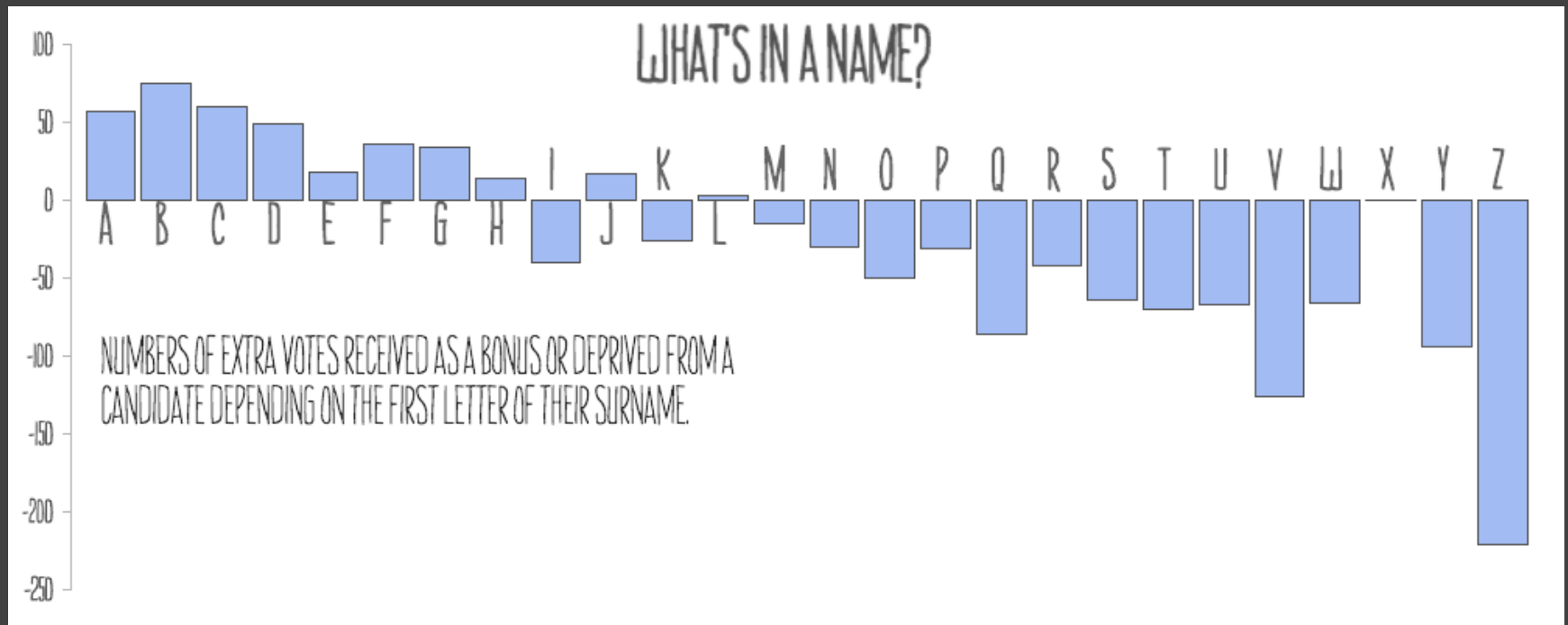
# Size Juxtaposition



# Size Superposition



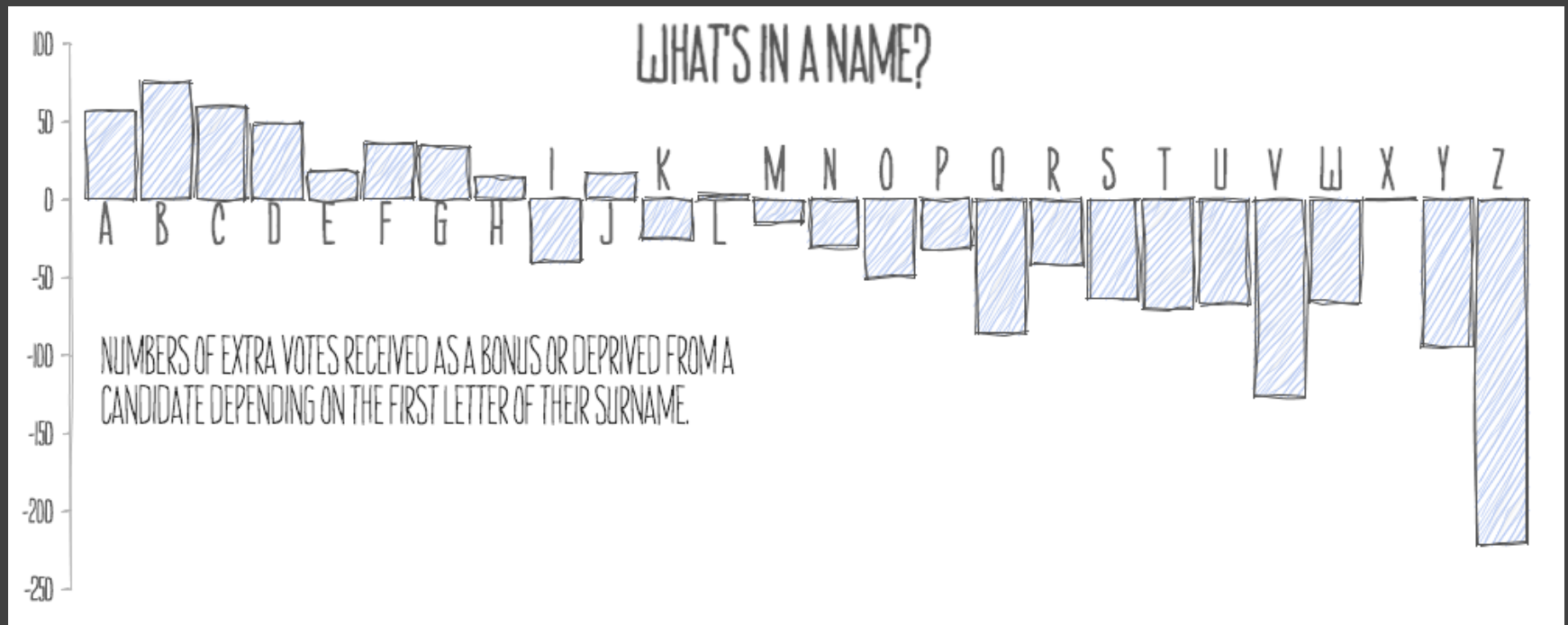
# "Sketchiness"



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

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

# "Sketchiness"

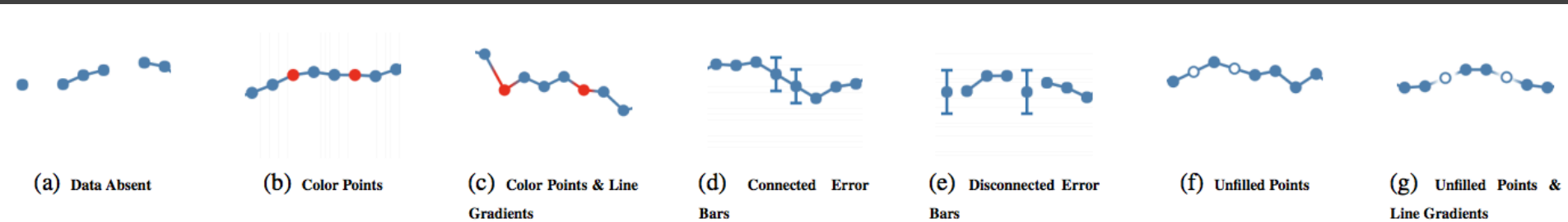


Wood, Jo et al. Sketchy rendering for information visualization. IEEE VIS, 2012.

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



# Perceived Data Quality



Error & perceived quality decrease with more missing data

Highlighting missing data increased perceived quality

Linear interpolation led to highest perceived quality

Absent data led to lower perceived quality, credibility, confidence

Limited evidence for accuracy bias from imputation methods

# 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

# KRAFTWEAK



# THE MODEL

# Polling Data



**PublicPolicyPolling** ✓

@ppppolls

Follow



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

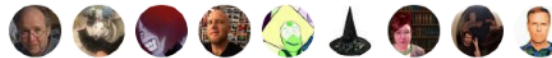
**SUSAN** @Sue4the5

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



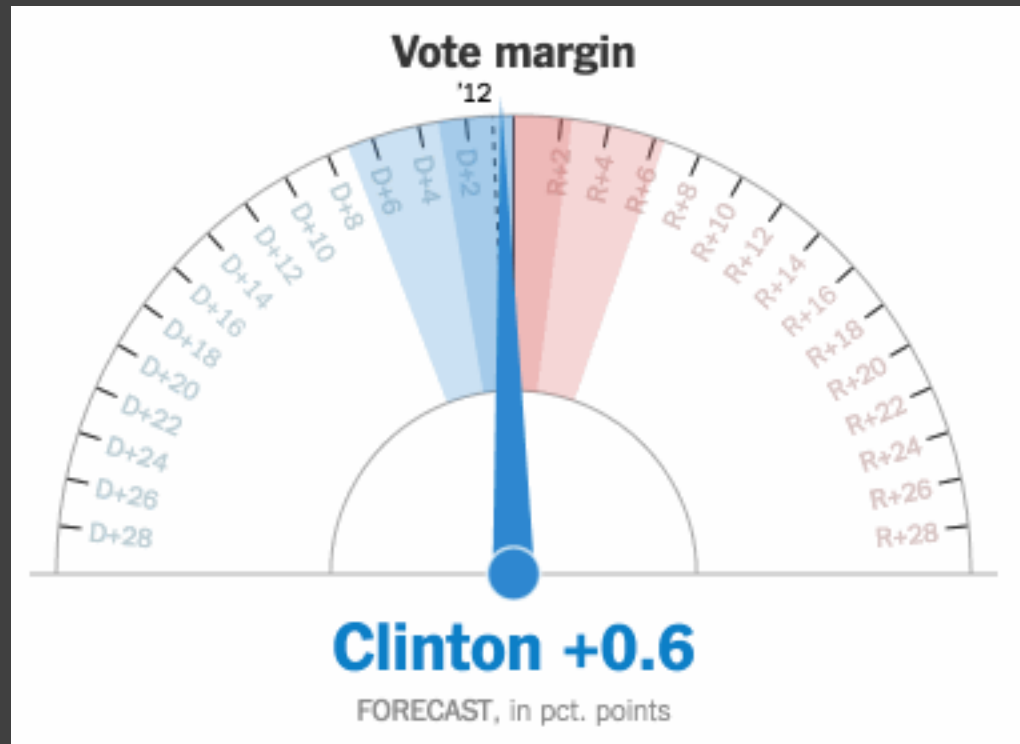
127

1.4K

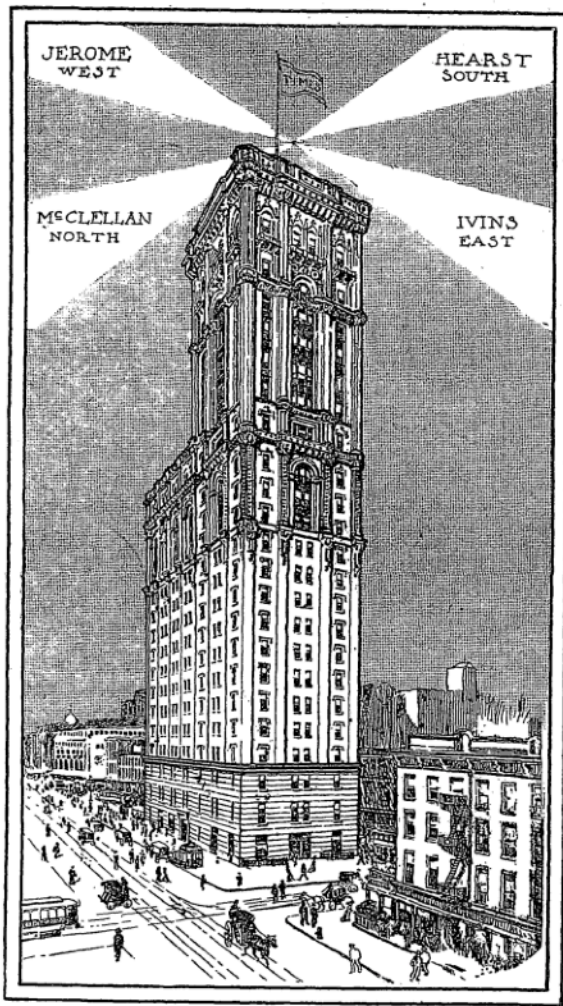
6.2K



# The NYT Needle



# ELECTION RESULTS BY SEARCHLIGHT.



The Times Election Searchlight Code.

## News Will Be Flashed from the Tower of The Times Building on Tuesday Night.

The results of the election next Tuesday night will be flashed by electric light from the tower of the Times Building, so that for miles around people will be able to tell which of the candidates has won.

This will be entirely separate and distinct from the elaborate bulletin service which THE TIMES will also maintain. To display the detailed bulletins so that the crowds can see them easily and comfortably, a stereopticon machine will be set up in the triangle north of the Times Building and the bulletins displayed on canvas stretched from the north side of the building. There will be a similar

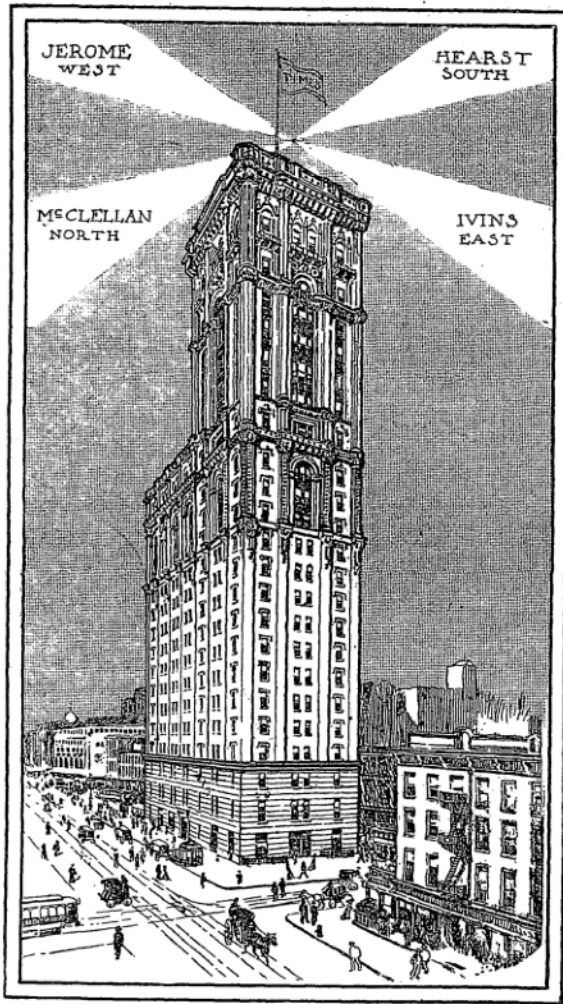
service at the Harlem office of THE TIMES, 129 West 125th Street.

The electric signals from the tower of the Times Building will be flashed from a point 365 feet above the street level. A steady light to the north will show that McClellan has been elected; a steady light to the east will indicate Ivins's election, and a steady light to the south will indicate that Hearst has won.

Jerome's election will be indicated by a steady light to the west. A light to the north, waving from east to west, will indicate Osborne's election. A light to the south, waving from east to west, will indicate Shearn's election.



## ELECTION RESULTS BY SEARCHLIGHT.



The Times Election Searchlight Code.

### News Will Be Flashed from the Tower of The Times Building on Tuesday Night.

The results of the election next Tuesday night will be flashed by electric light from the tower of the Times Building, so that for miles around people will be able to tell which of the candidates has won.

This will be entirely separate and distinct from the elaborate bulletin service which THE TIMES will also maintain. To display the detailed bulletins so that the crowds can see them easily and comfortably, a stereopticon machine will be set up in the triangle north of the Times Building and the bulletins displayed on canvas stretched from the north side of the building. There will be a similar

service at the Harlem office of THE TIMES, 129 West 125th Street.

The electric signals from the tower of the Times Building will be flashed from a point 365 feet above the street level. A steady light to the north will show that McClellan has been elected; a steady light to the east will indicate Ivins's election, and a steady light to the south will indicate that Hearst has won.

Jerome's election will be indicated by a steady light to the west. A light to the north, waving from east to west, will indicate Osborne's election. A light to the south, waving from east to west, will indicate Shearn's election.

# Election Bulletins

## BY BOMBS.

### TUESDAY NIGHT

# THE TRIBUNE

will send up from the roof of the

## GREAT NORTHERN HOTEL

hourly, shells containing blue and red stars—exactly on the hour—at 7, 8, 9, 10, 11 p. m. 12 midnight, 1 and 2 a. m. Wednesday morning, unless election is decided earlier, in which case twelve bombs will be sent up in rapid succession. Blue to indicate McKinley's election. Red to indicate Bryan's election.

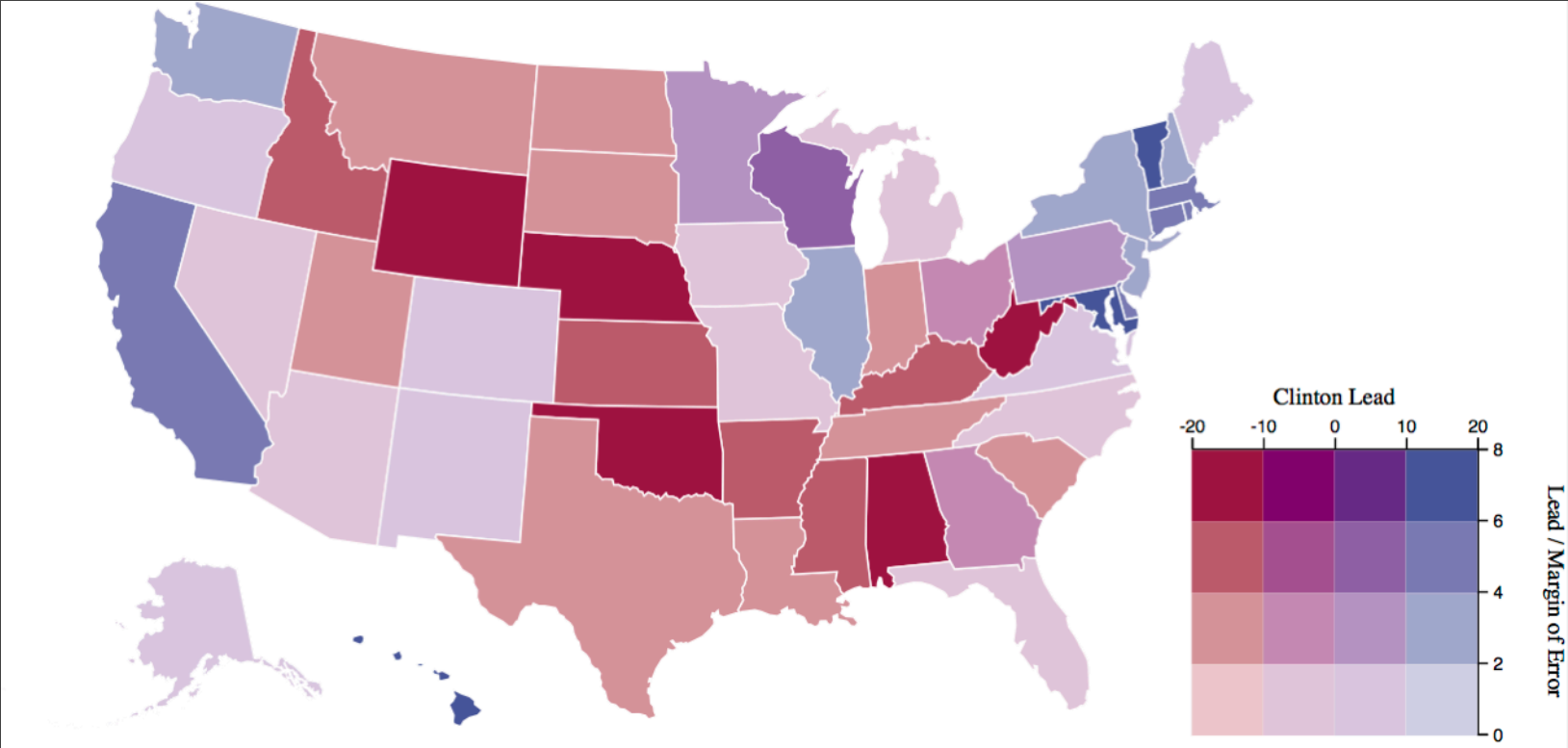
## SIX BOMBS EVERY HOUR.

The first bomb sent up, if blue, indicates the returns in **COOK COUNTY** at that hour are favorable to McKinley; if red, favorable to Bryan.

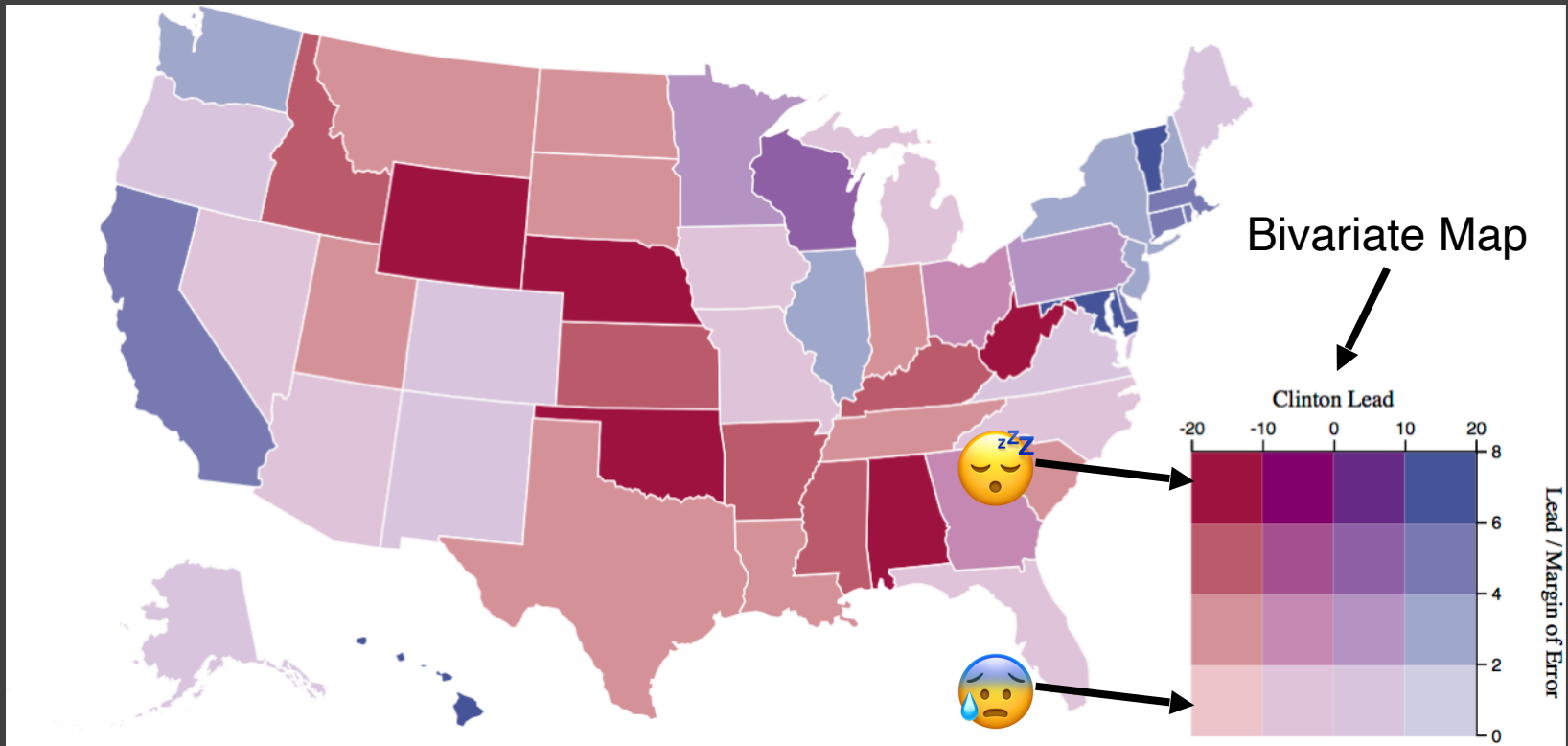
After sixty seconds two bombs will be sent up in rapid succession, and will indicate, if blue, that returns from **ILLINOIS** favor McKinley; if red, Bryan.

After sixty seconds more three bombs will be sent up in rapid succession, and if blue will indicate that at that hour returns from the **entire country** favor McKinley; if red, Bryan. Each bomb bursts high in the air, scattering a shower of stars.

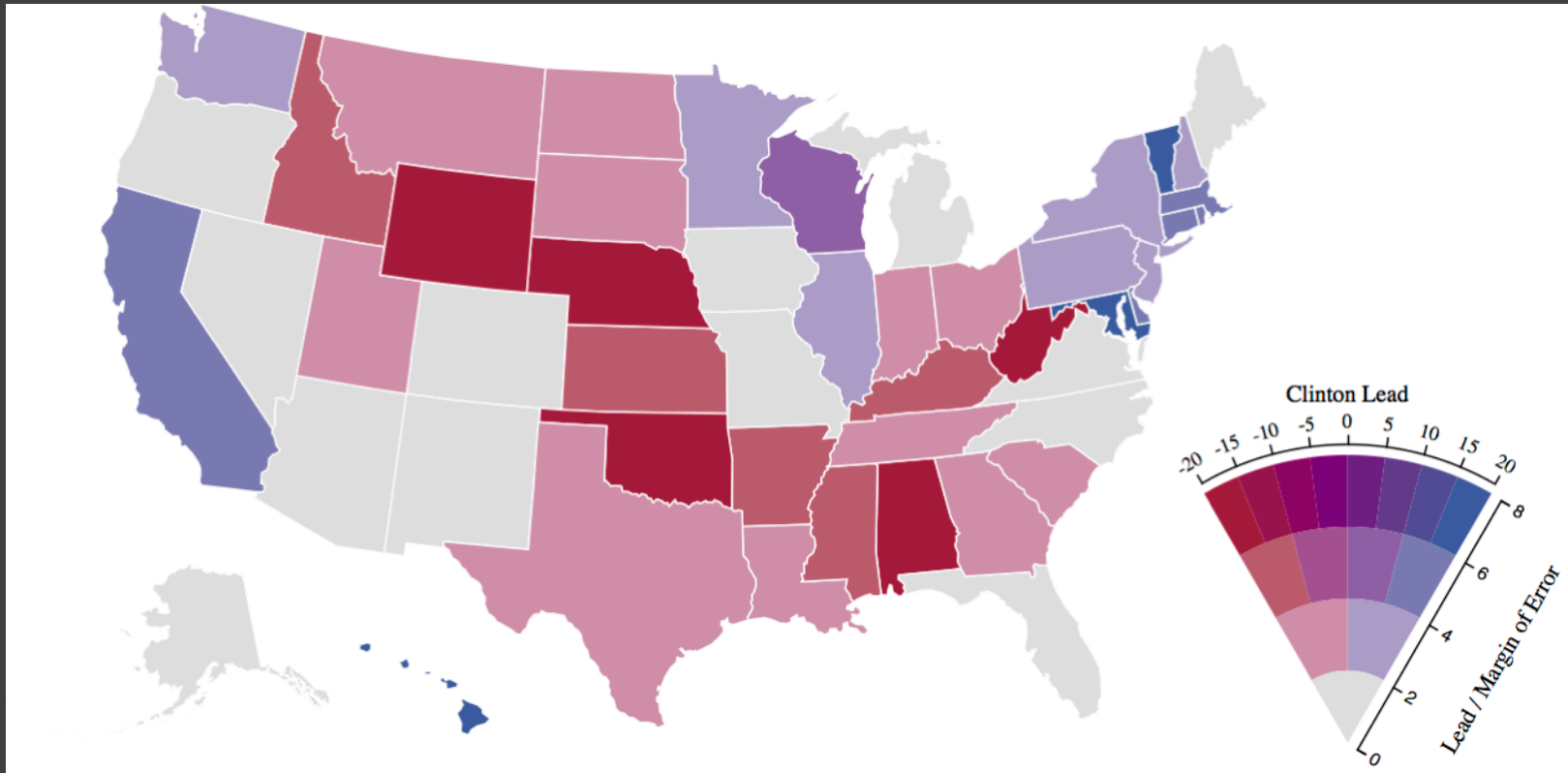
# Value-Suppressing Uncertainty Palette



# Bivariate Map



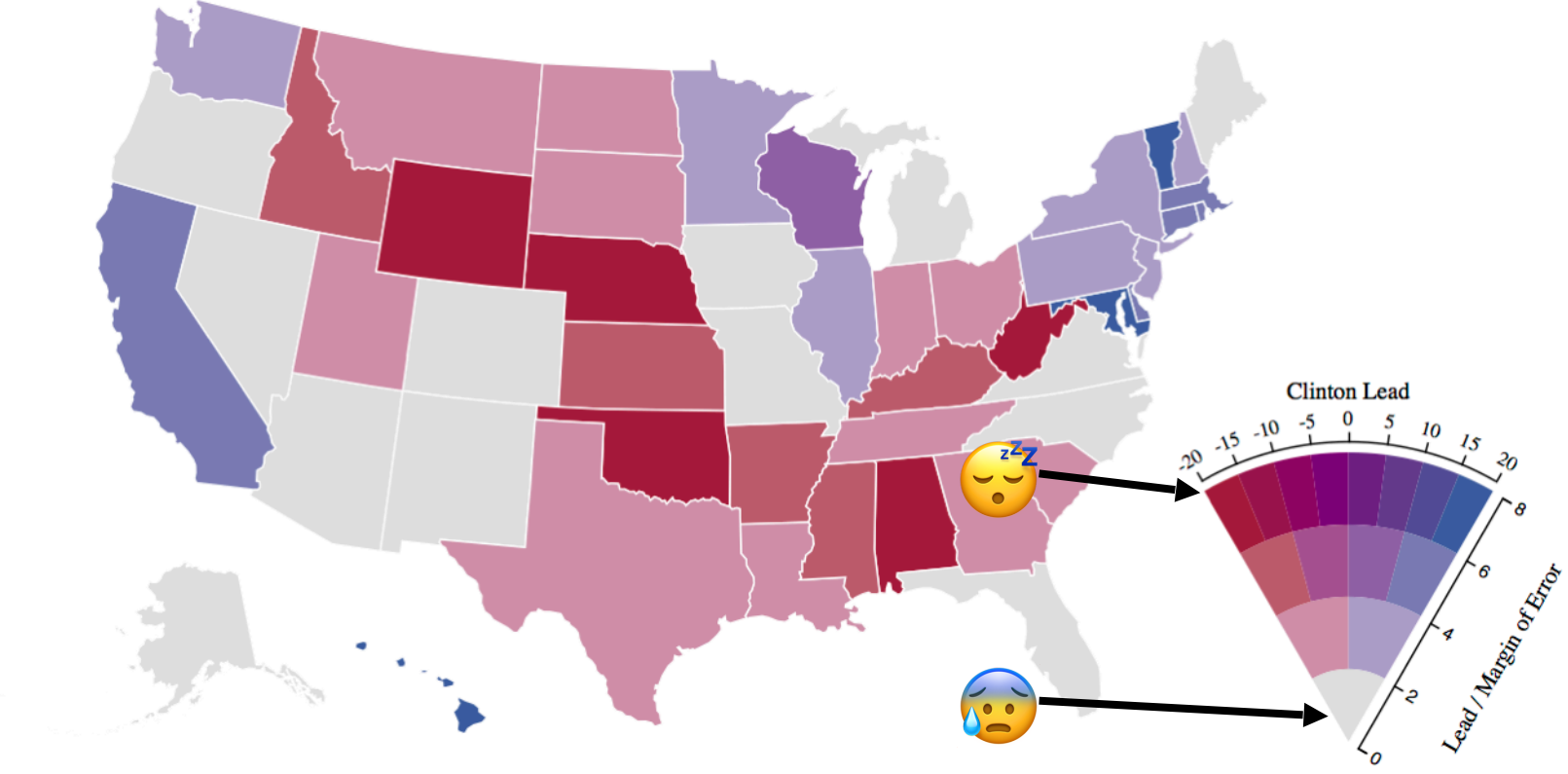
# Value-Suppressing Uncertainty Palette



Correll, Moritz & Heer. "Value-Suppressing Uncertainty Palettes." CHI 2018.

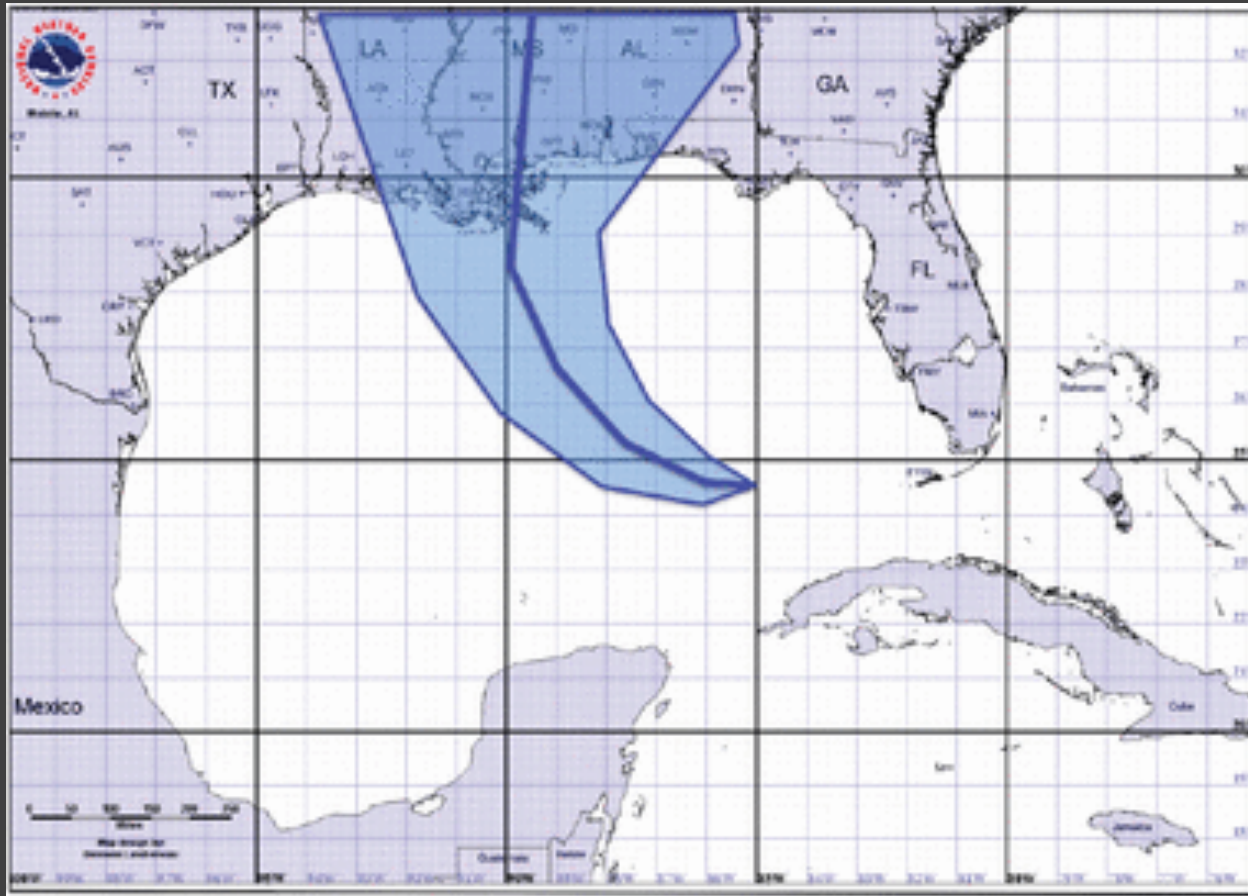


# Value-Suppressing Uncertainty Palette



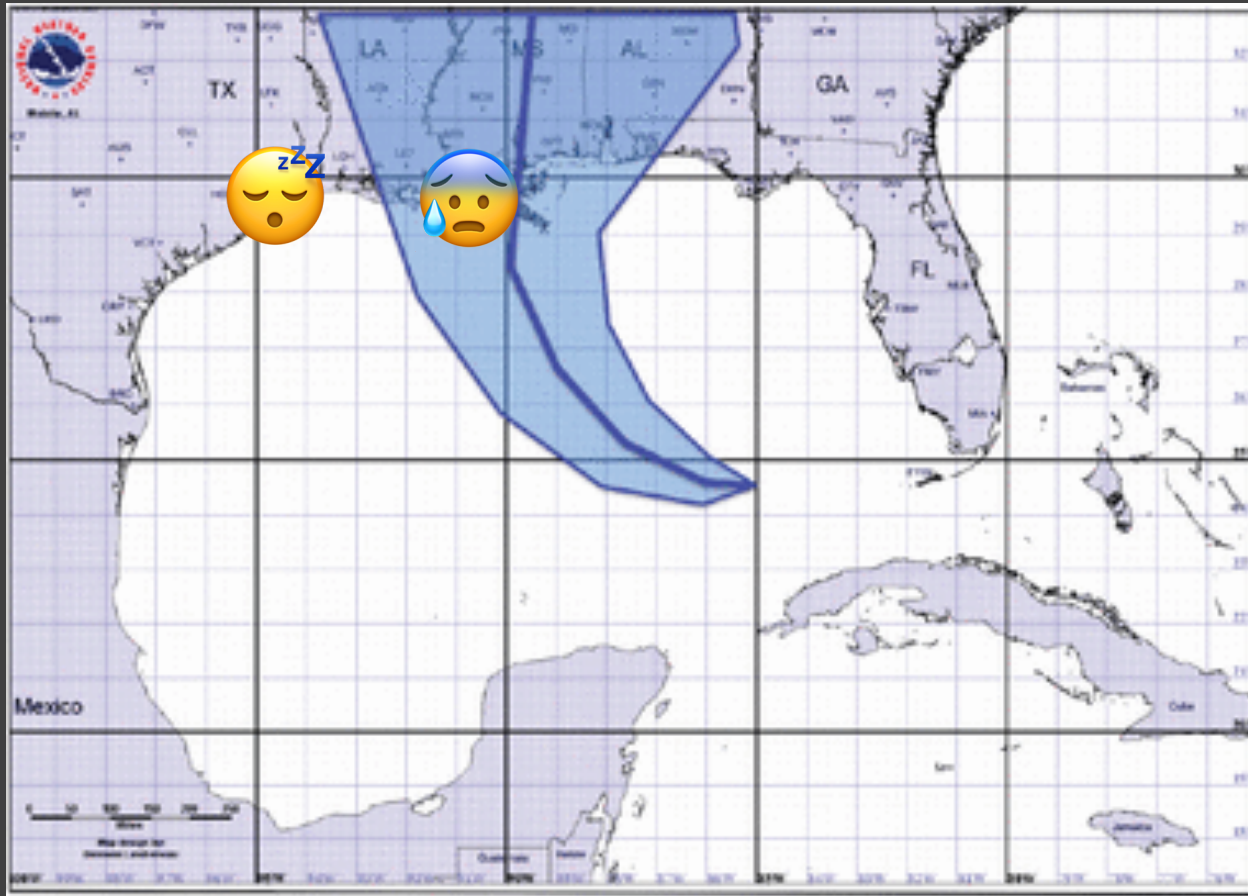
Correll, Moritz & Heer. "Value-Suppressing Uncertainty Palettes." CHI 2018.

# Model Visualization



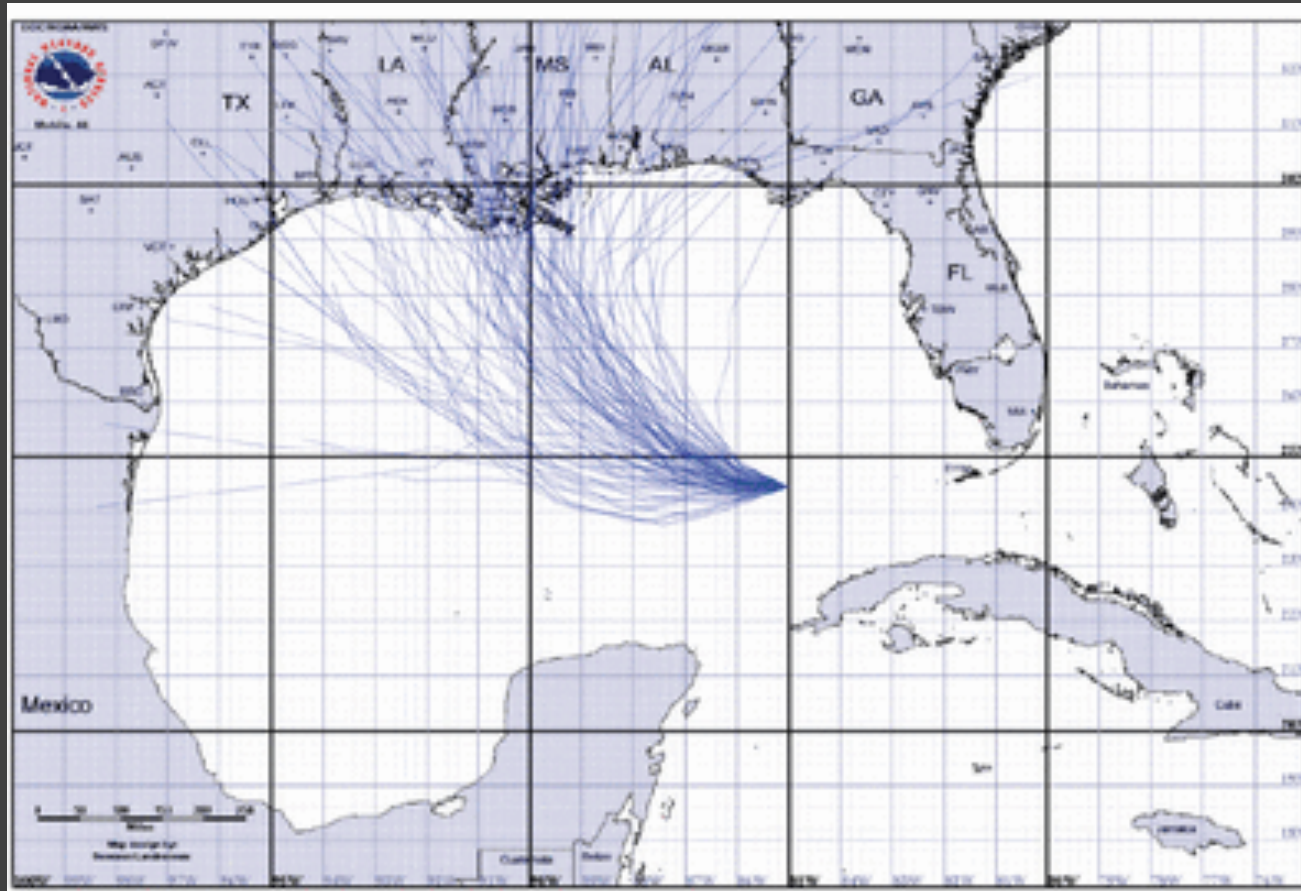
Cox, Jonathan and House, Donald and Lindell, Michael. Visualising uncertainty in predicted hurricane tracks. *International Journal for Uncertainty Quantification*, 2013.

# Model Visualization

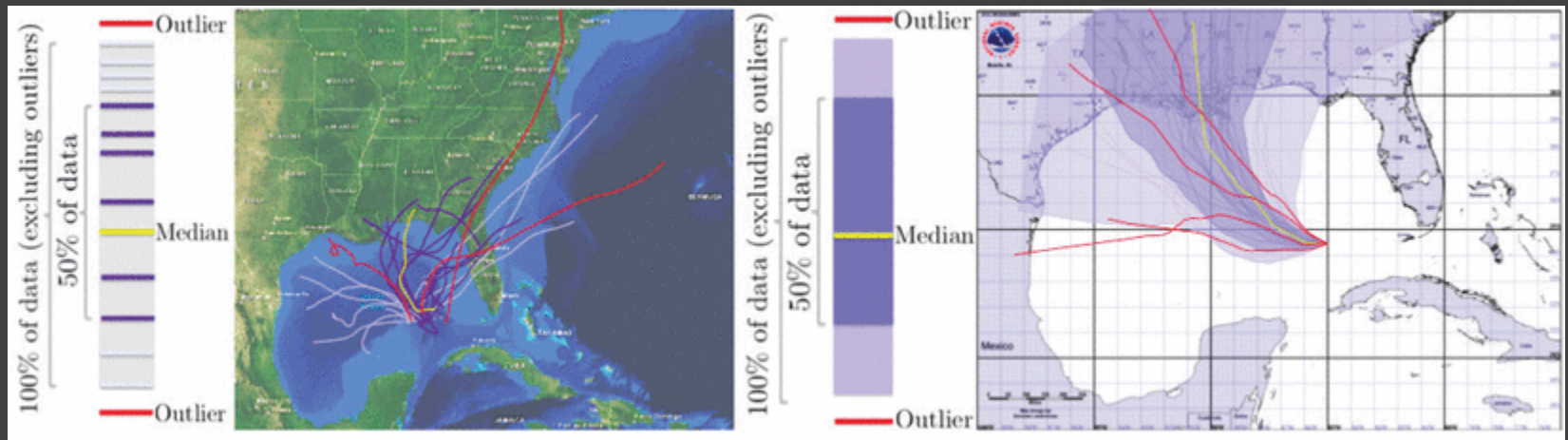


Cox, Jonathan and House, Donald and Lindell, Michael. Visualising uncertainty in predicted hurricane tracks. International Journal for Uncertainty Quantification, 2013.

# Model Visualization



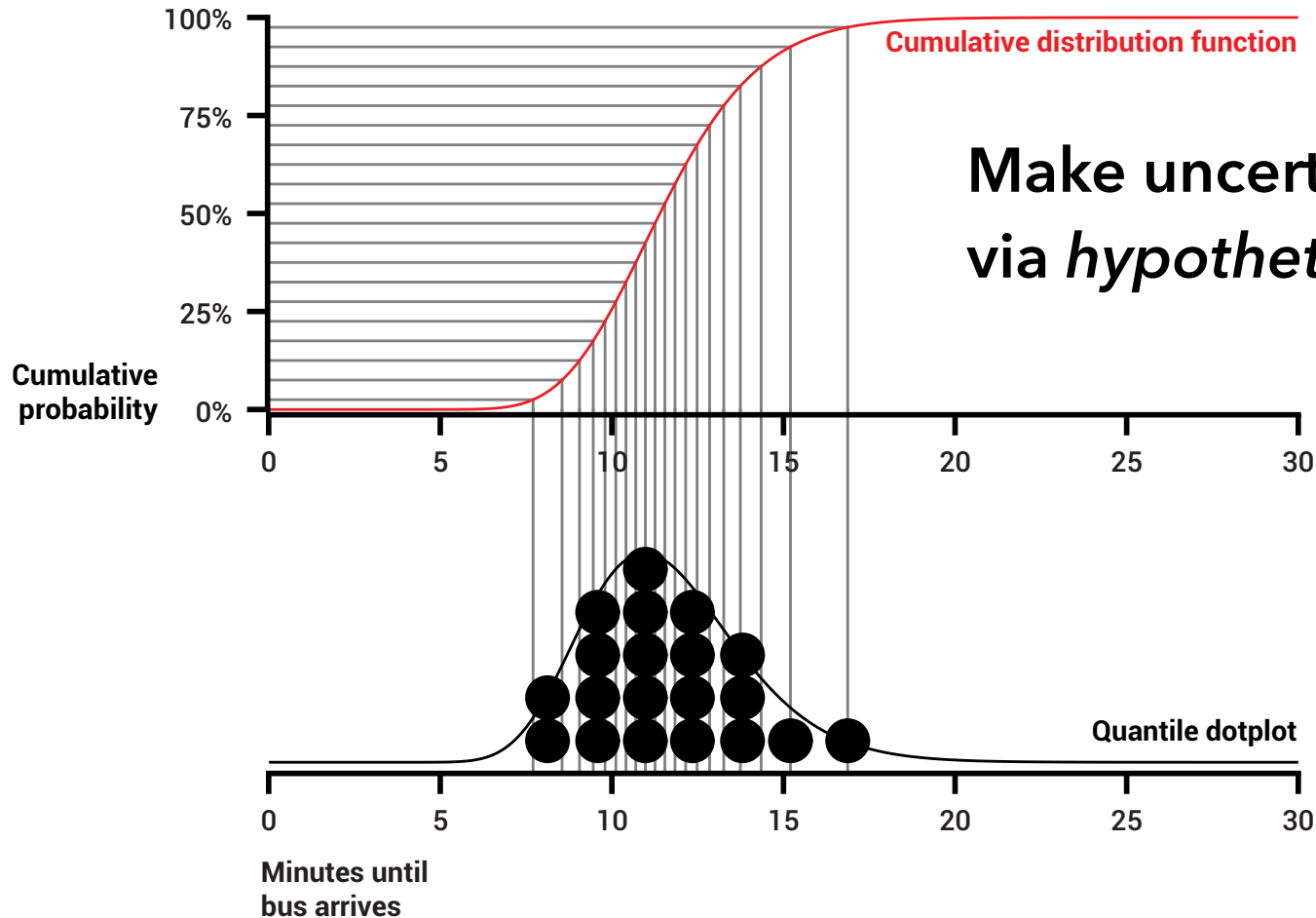
# Model Visualization



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

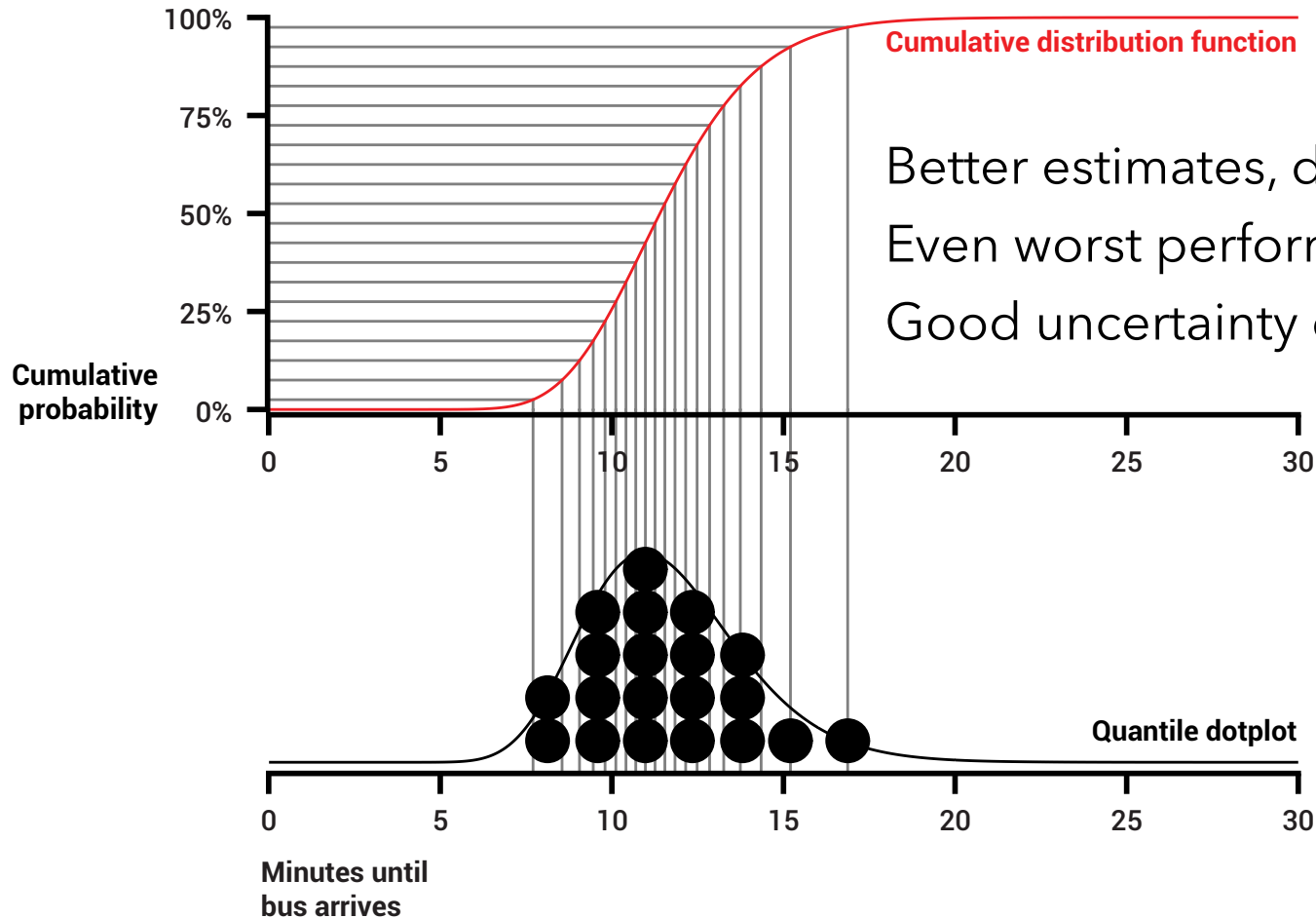


# Predicted Bus Arrival Times

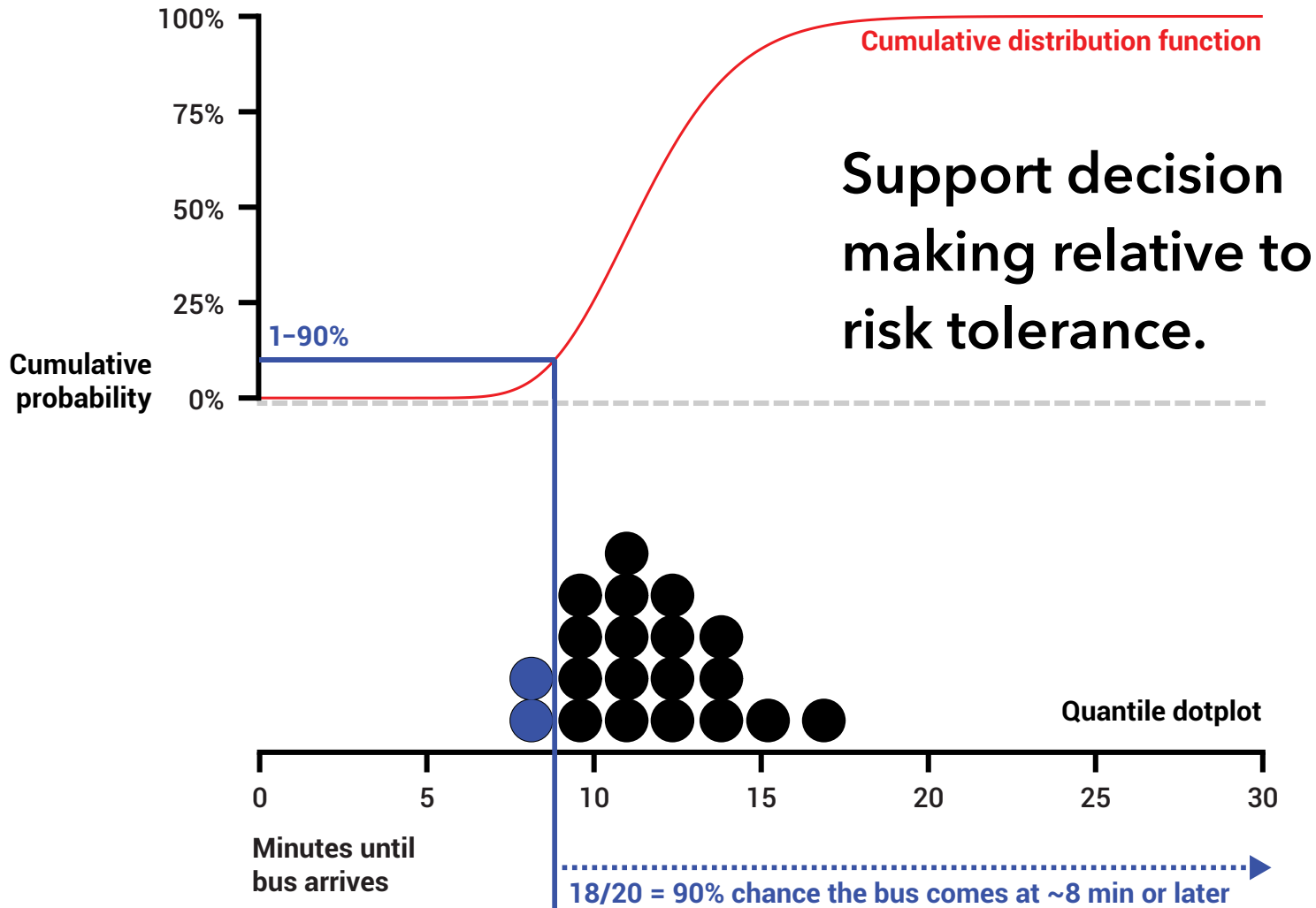


Make uncertainty concrete  
via *hypothetical outcomes*.

# Predicted Bus Arrival Times

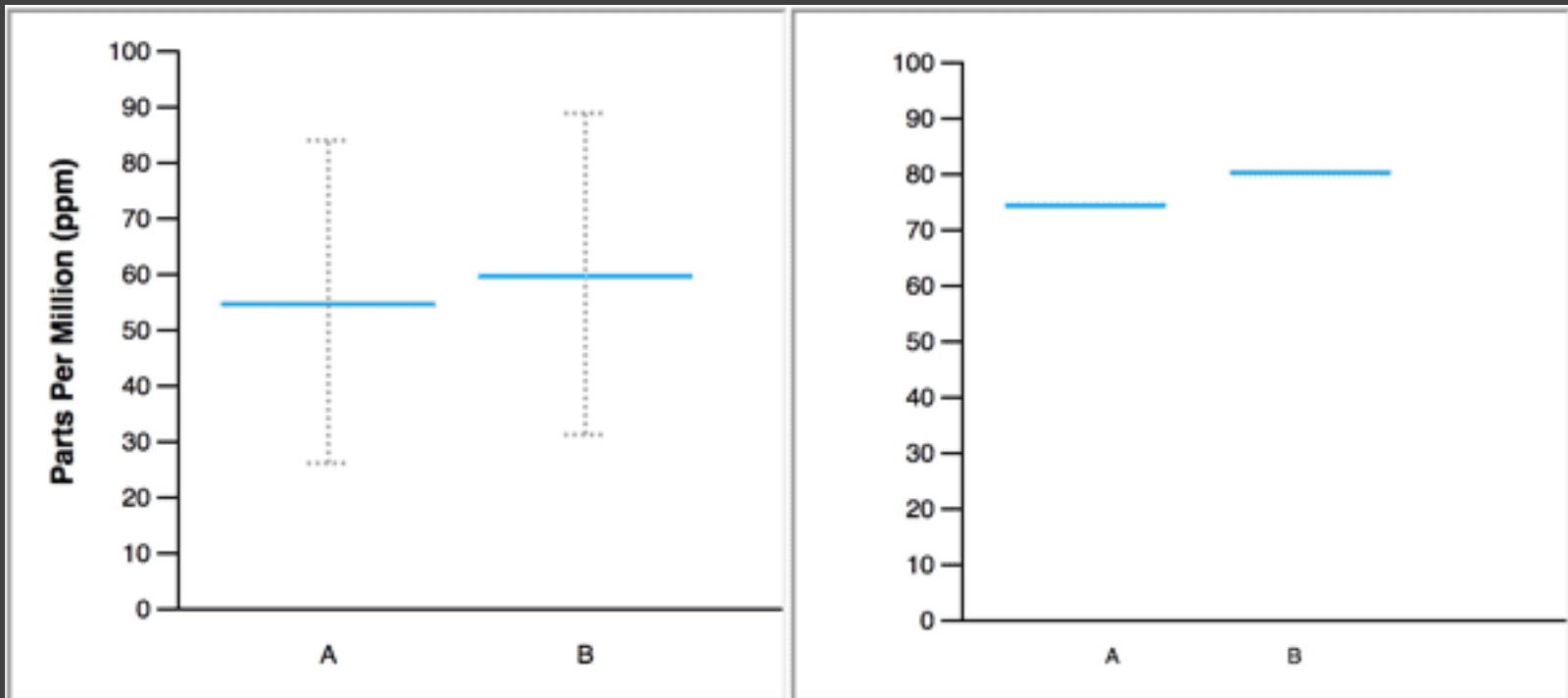


# Predicted Bus Arrival Times





# Hypothetical Outcome Plots



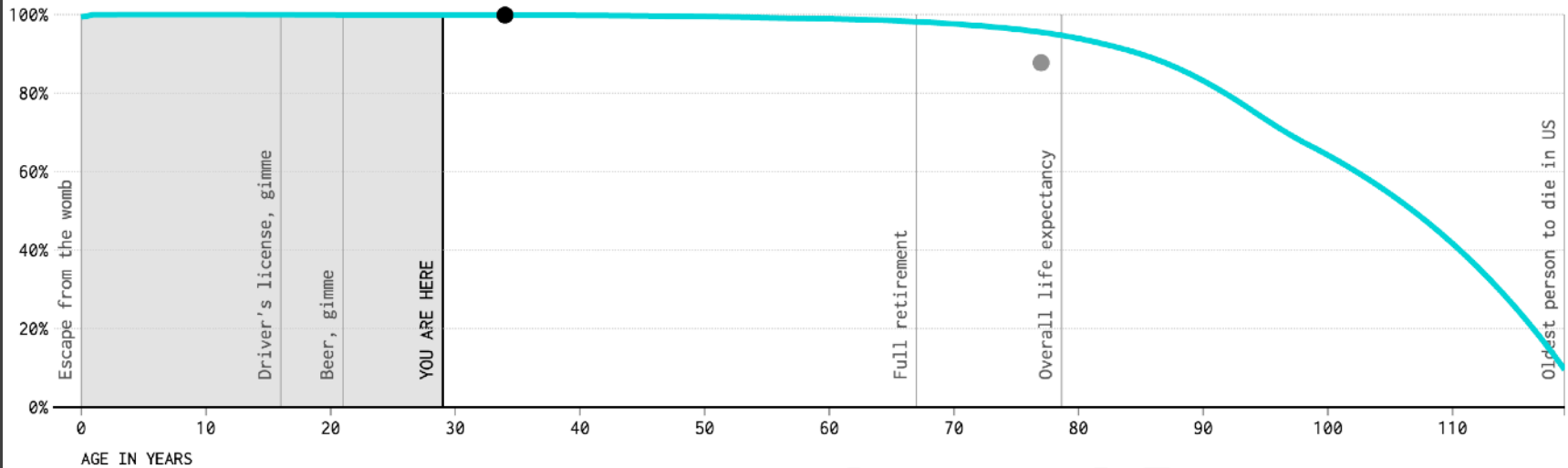
# Life Expectancy

I am **male** and currently **29** years old.

SLOW

FAST

PROBABILITY OF LIVING TO NEXT YEAR



# Model Visualization

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

# 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**

# Administrivia

# Reminders!

Final Project Proposal Due **Wed 5/19, 11:59pm**

<https://courses.cs.washington.edu/courses/cse512/21sp/fp.html>

Three Peer Evaluations Due **Tue 5/18, 11:59pm**

<https://courses.cs.washington.edu/courses/cse512/21sp/a3-peereval.html>

# Final Project Schedule

<i>Proposal</i>	Wed, May 19
<i>Milestone</i>	Thu, May 27
<i>Video</i>	Wed, June 2
<i>Deliverables</i>	Wed, June 9

## **Logistics**

Final project description posted online

Groups of up to 4 people

Select topics and form groups now!

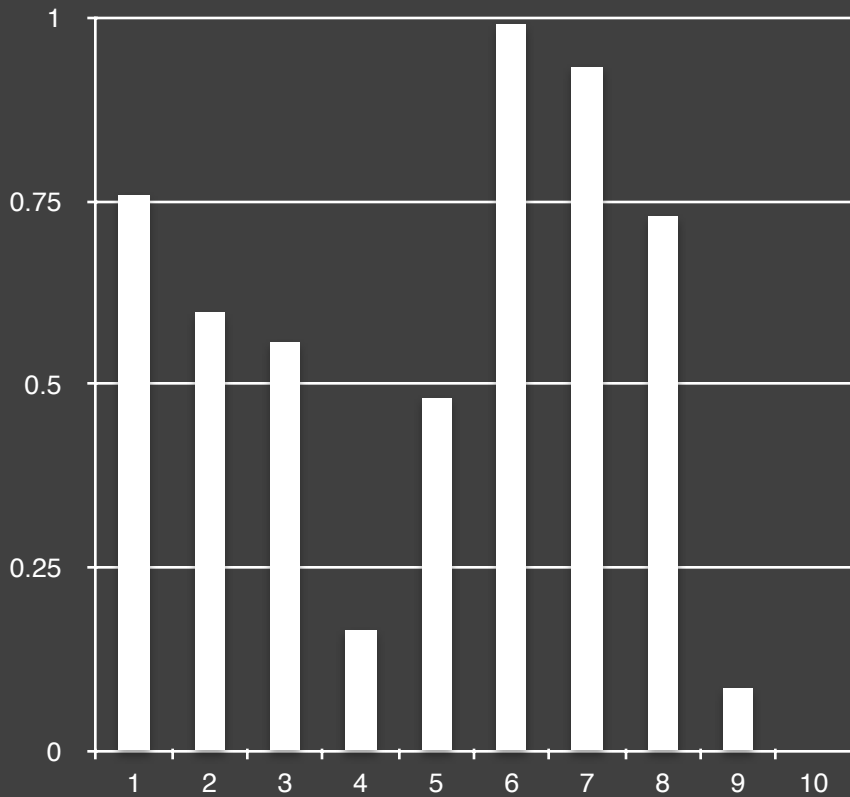


Break Time!

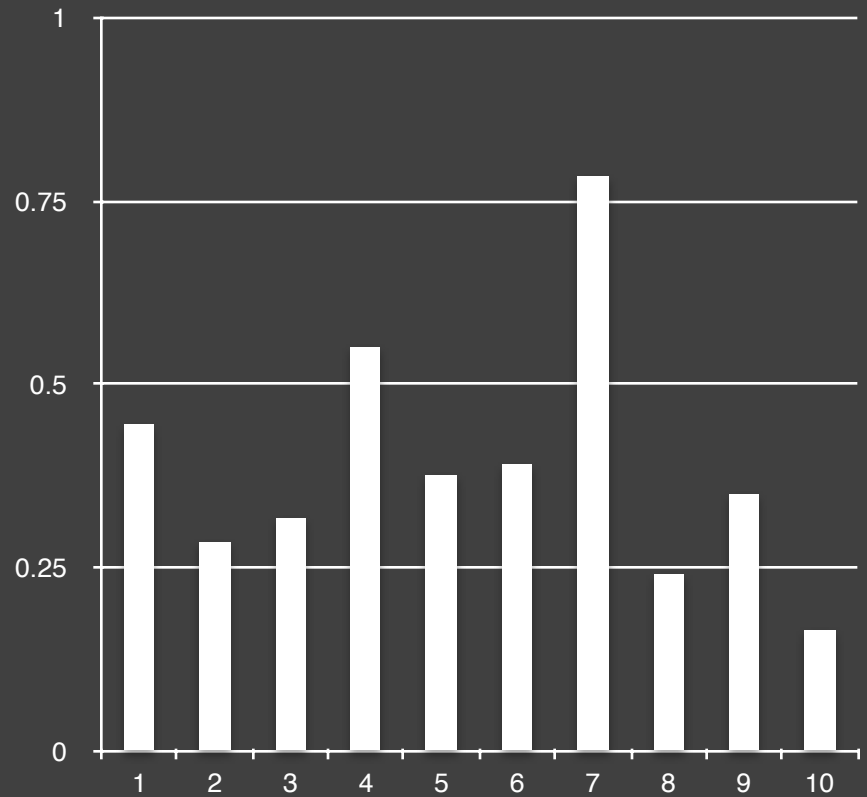
**What Can Go Wrong?**

# Which Stock To Buy?

## Company A

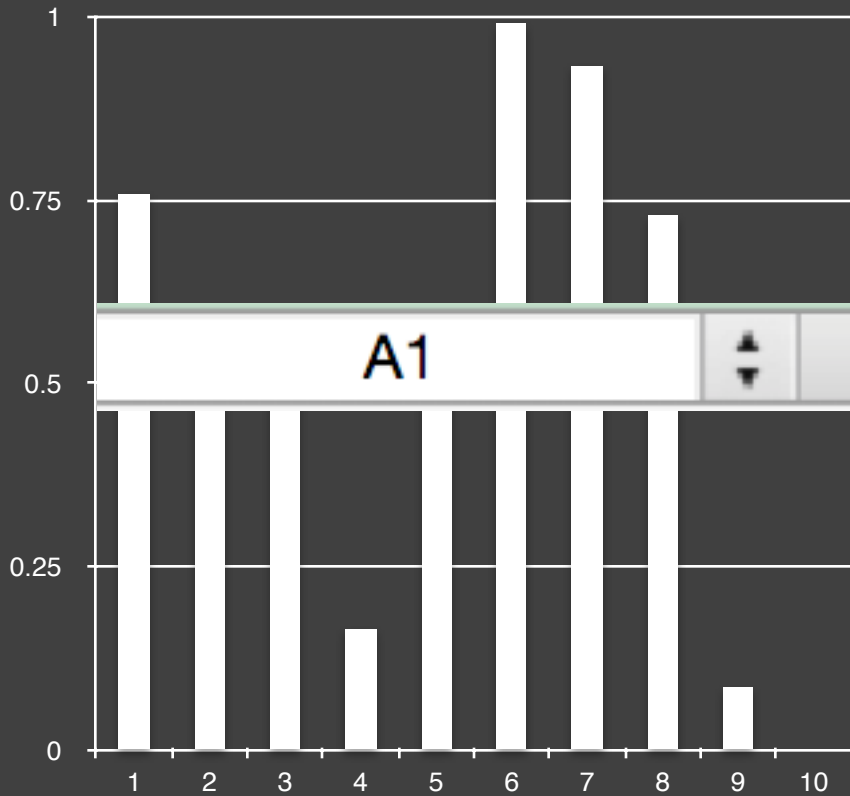


## Company B

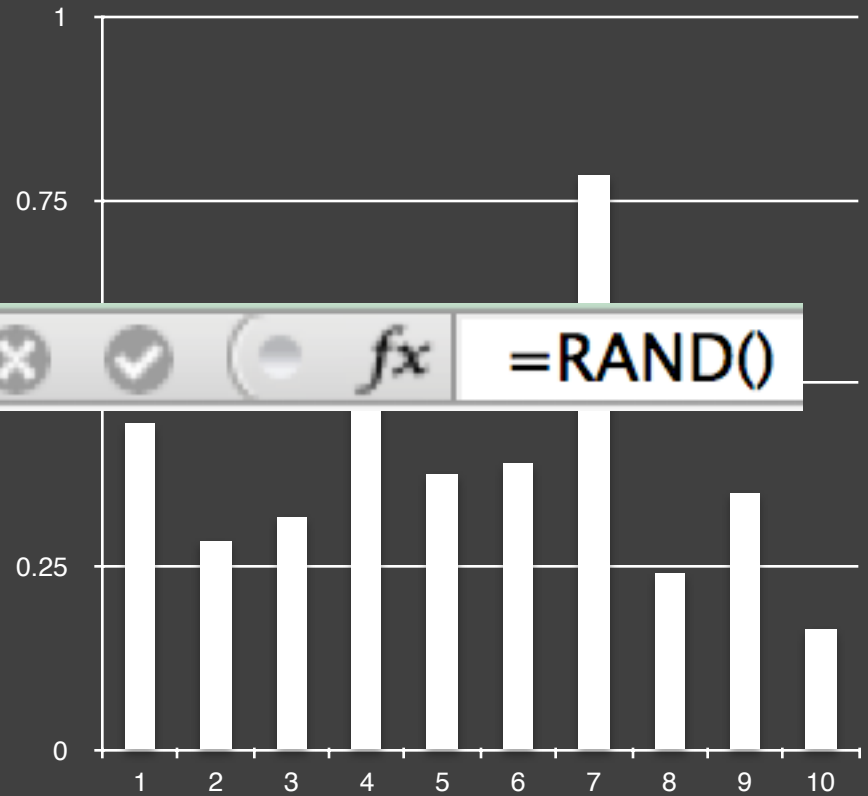


# Neither!

## Company A



## Company B



A1



*fx*

=RAND()

# Pareidolia



# Jobs Reports

**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.

*Job Growth  
Plummets Amid  
Prospect Of  
New Slump*

Under 55,000 jobs  
4% chance

*Disappointing  
Jobs Report  
Raises  
Economic  
Worries*

55,000 to 110,000  
19% chance

*Slower Job  
Creation  
Disappoints  
Economists*

110,000 to 140,000  
19% chance

*Job Growth  
Steady, New  
Report Says*

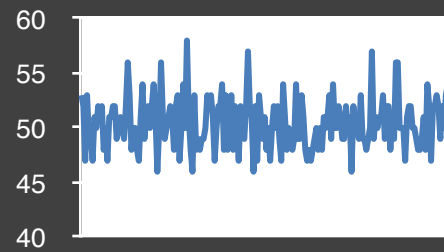
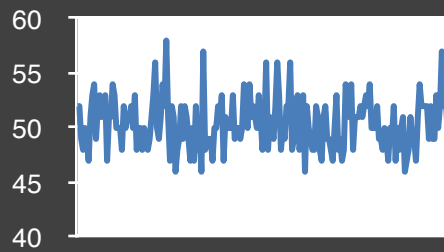
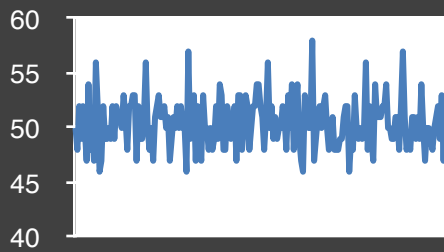
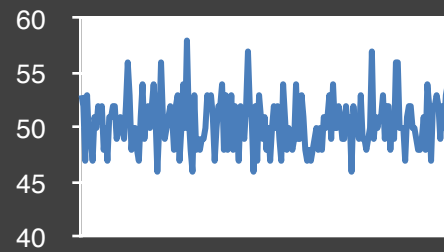
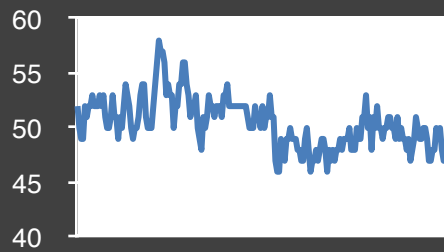
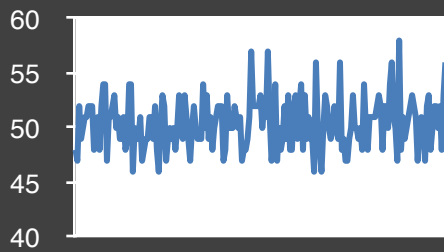
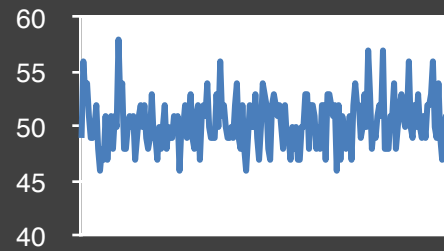
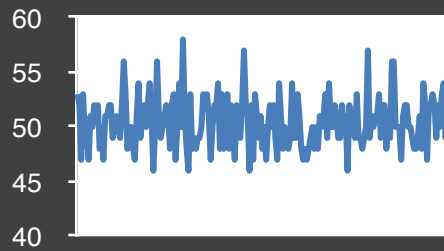
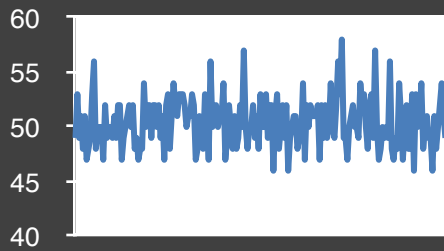
160,000 to 190,000  
19% chance

*Job Creation  
Accelerates In  
Sign Of  
Economy  
Improving*

190,000 to 245,000  
19% chance

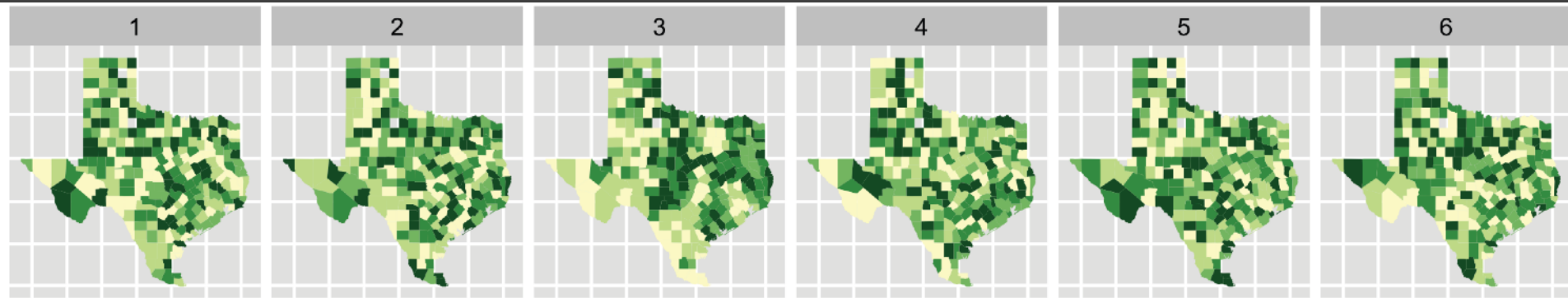
*Job Growth  
Robust,  
Pointing To  
Economy  
Surging*

245,000+  
4% chance









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.

1

2

3

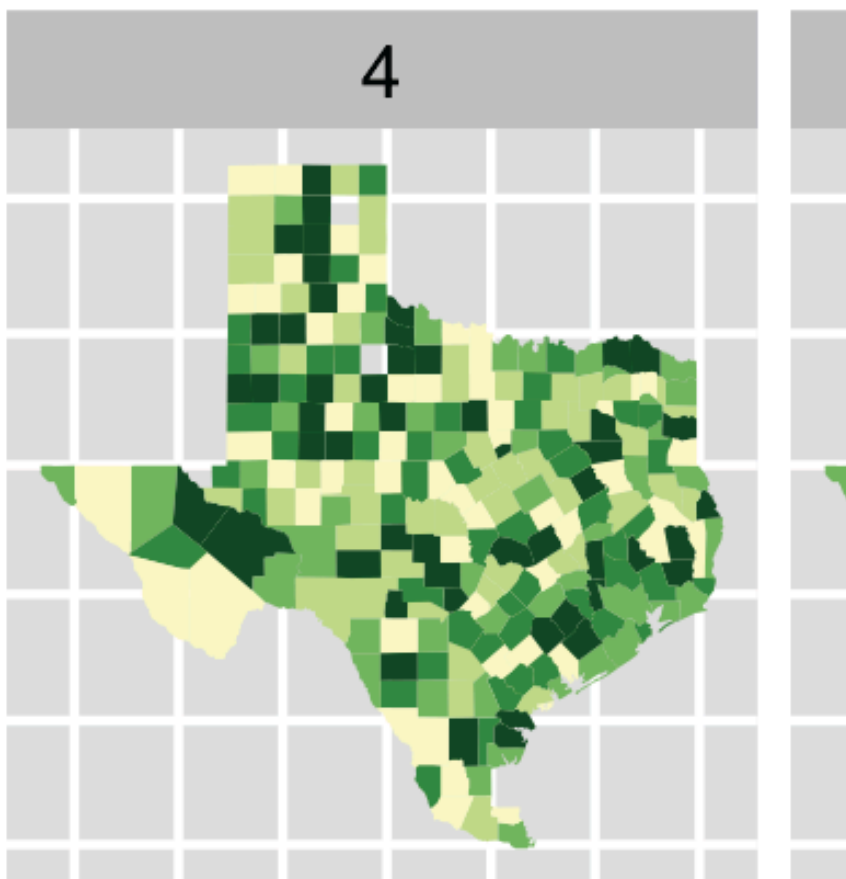
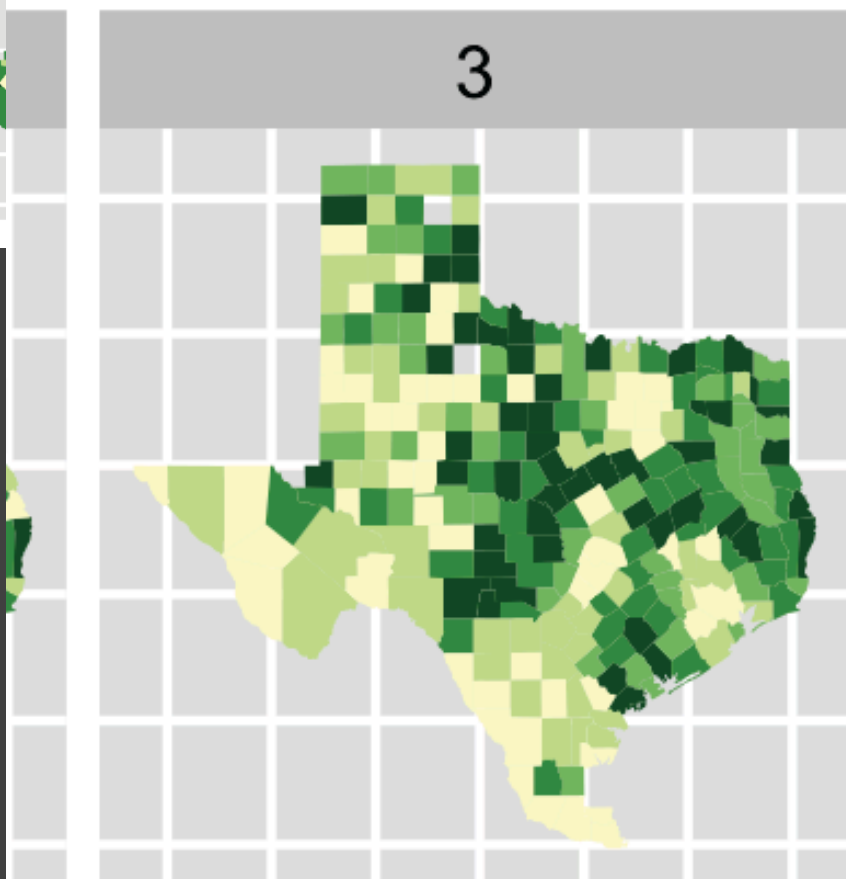
4

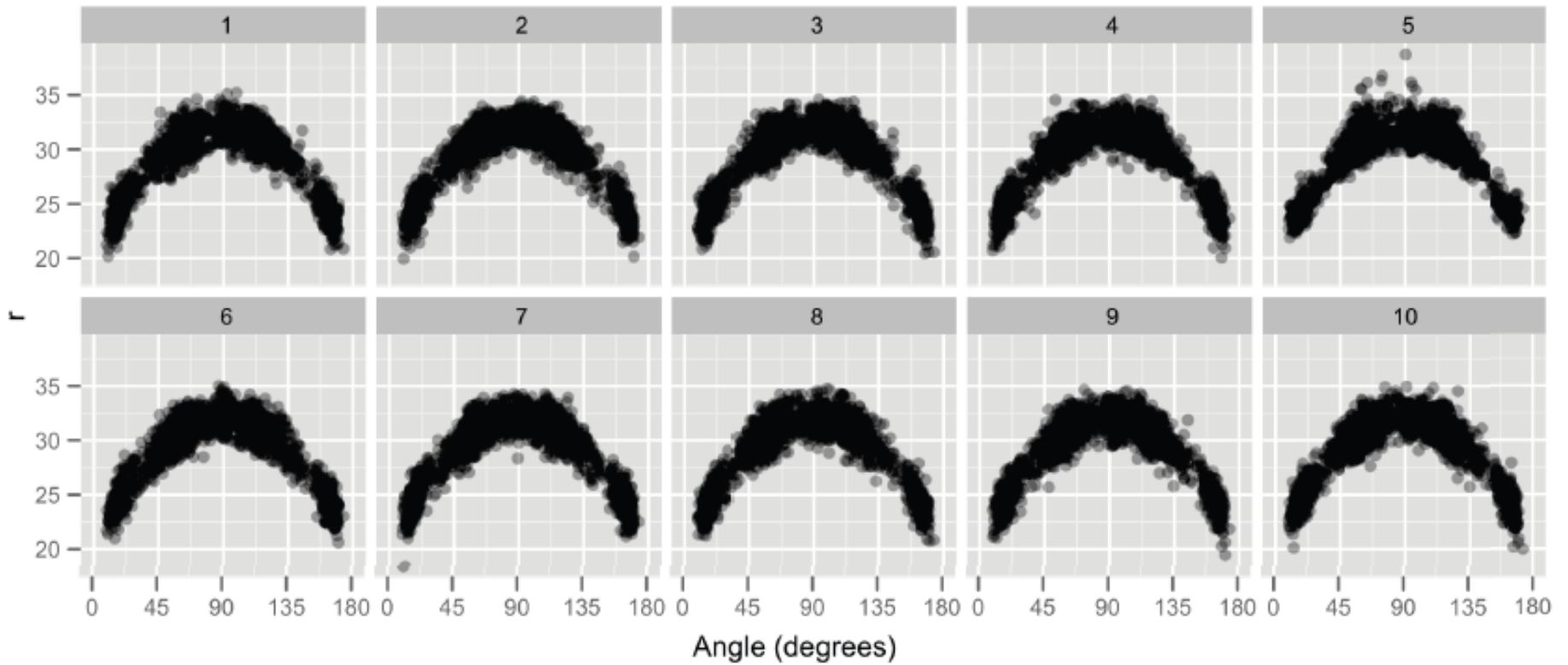
5

6

3

4





Distance vs. angle for 3 point shots by the LA Lakers.  
One plot is the real data. The others are generated according to a null hypothesis of quadratic relationship.

# 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

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

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

# Bayes' Law

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

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

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



# Bayes' Law

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

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

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

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

$$P(+)= 0.107$$

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

# 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.

# 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 Probabilities

Less Intuitive

Probability

$P(A) = 0.6$



Percentage

60% chance of A

More Intuitive

Natural  
Frequency

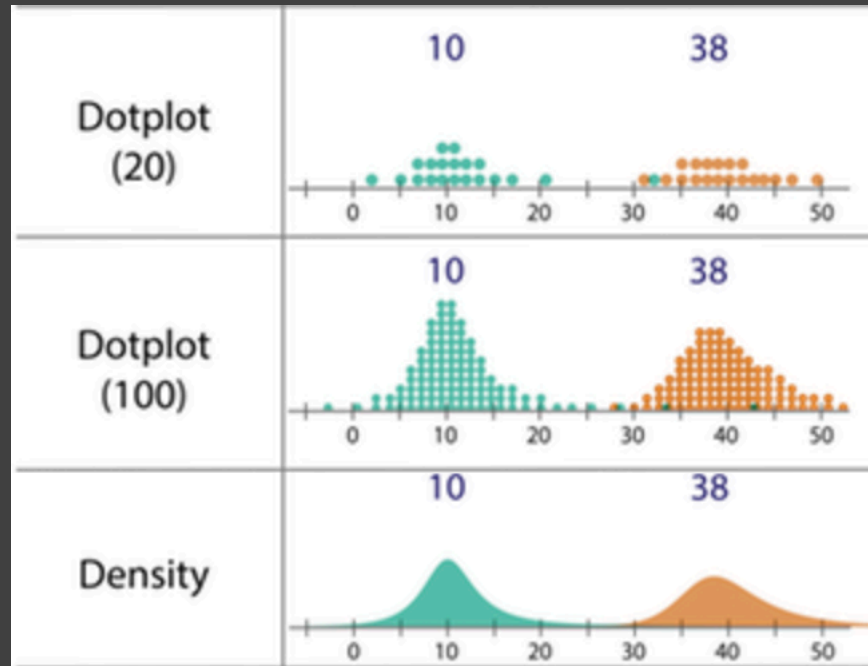
3 out of 5 times,  
A happens.

# Quantile Dot Plots

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**

# 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.