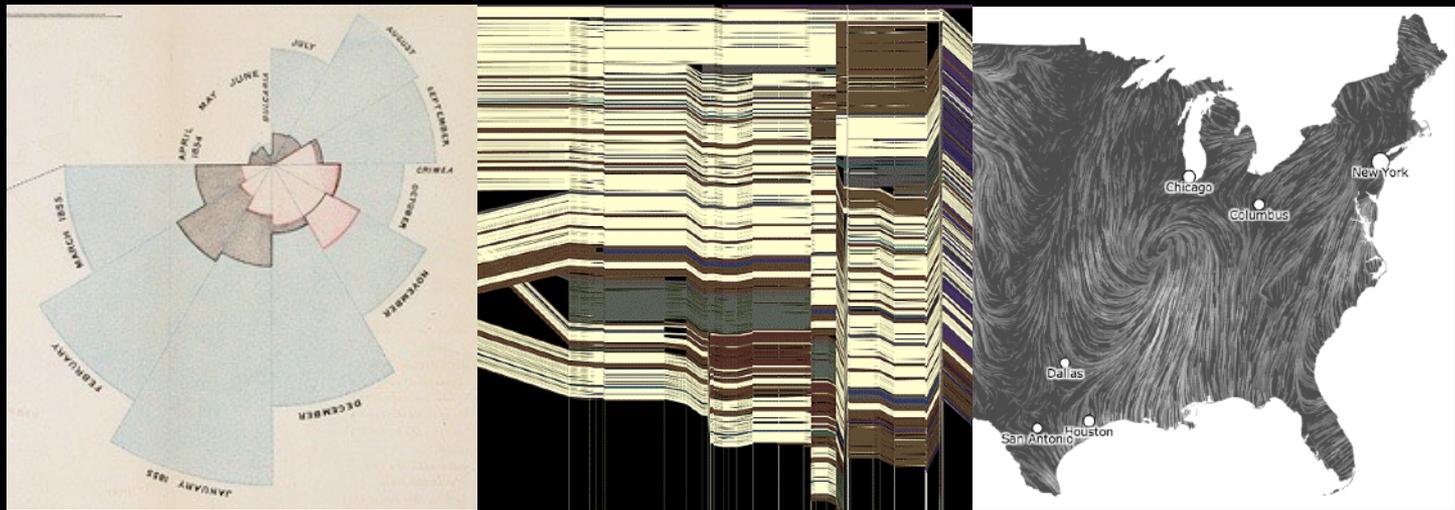


CSE 442 - Data Visualization

# Uncertainty



Jeffrey Heer University of Washington

**"I estimate that we catch 25% of our 100x errors, and 5% of our 5x errors."**

Anonymous Data Science Team Manager

# Session Outline

## What Does Uncertainty Mean?

### Uncertainty Visualization

Avoid Prematurely Suppressing Uncertainty

Visual Encodings of Uncertainty

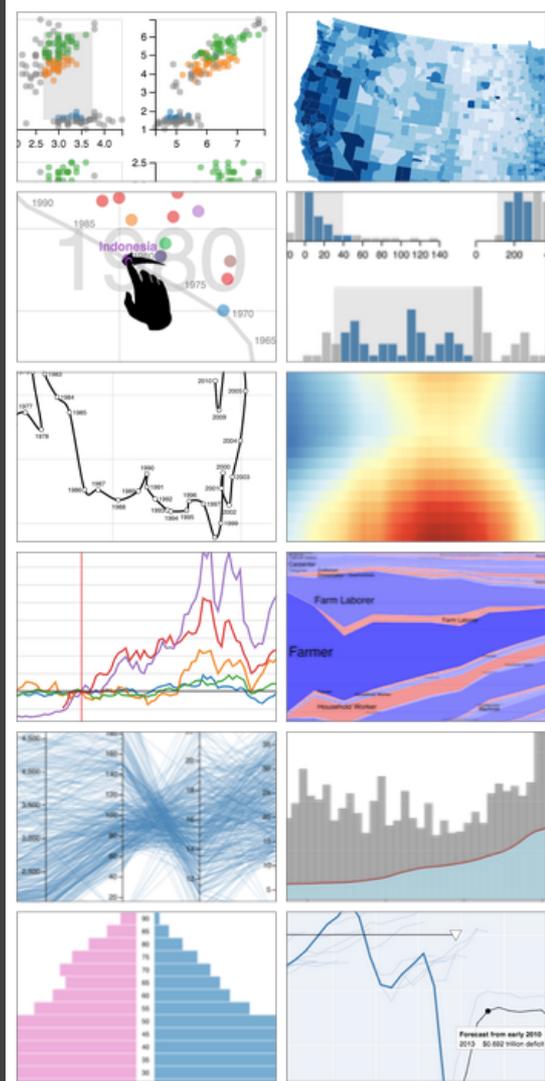
Frequency Framing & Hypothetical Outcomes

## What Can Go Wrong?

Inferential Integrity

Graphical Inference & Model Checks

**GOAL:** Try not to fool yourself!



**What Does Uncertainty Mean?**

# Things “Uncertainty” Can Mean

Doubt

Risk

Variability

Error

Lack of Knowledge

Hedging

...

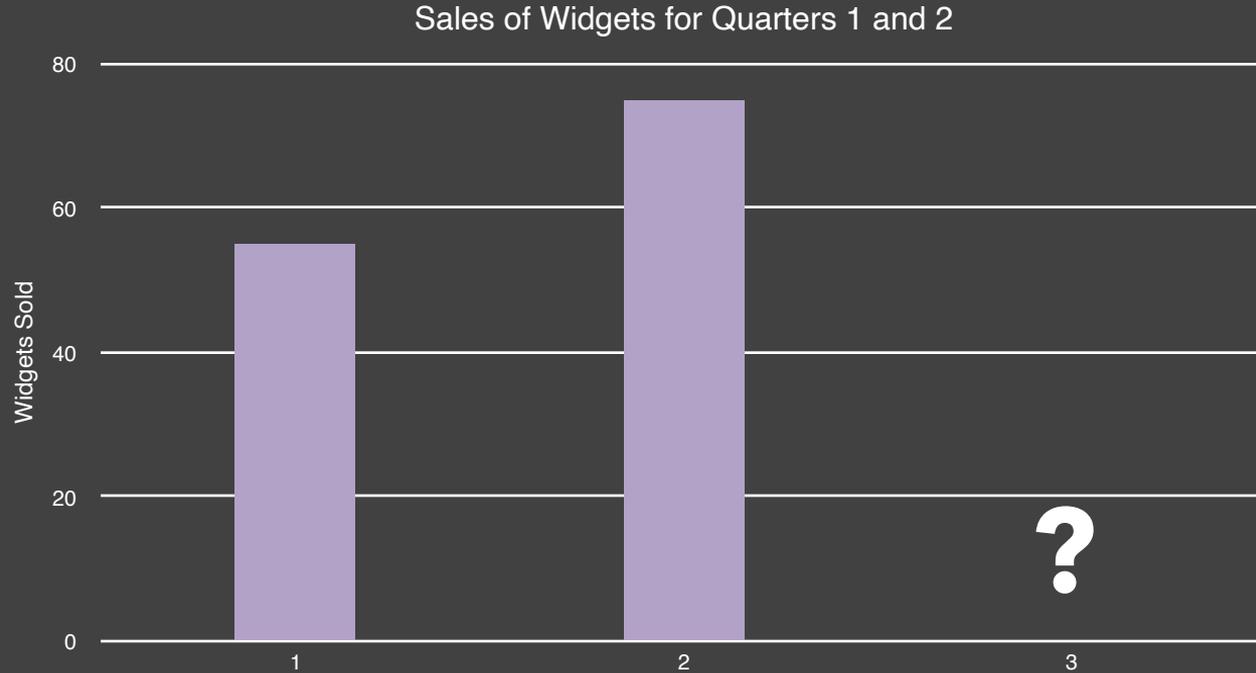
# A Bar Chart



# Measurement Uncertainty



# Forecast Uncertainty



# Decision Uncertainty



# 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"

Should I Bring an Umbrella?

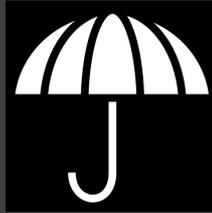


# Decision Uncertainty

"50% Chance of Rain"



# Types of Error




# The Boy Who Cried Wolf

**Type I: False Positive**



**Type II: False Negative**





Sean J. Taylor @seanjtaylor

Here's my trick.

TYPE

FALSE POSITIVE

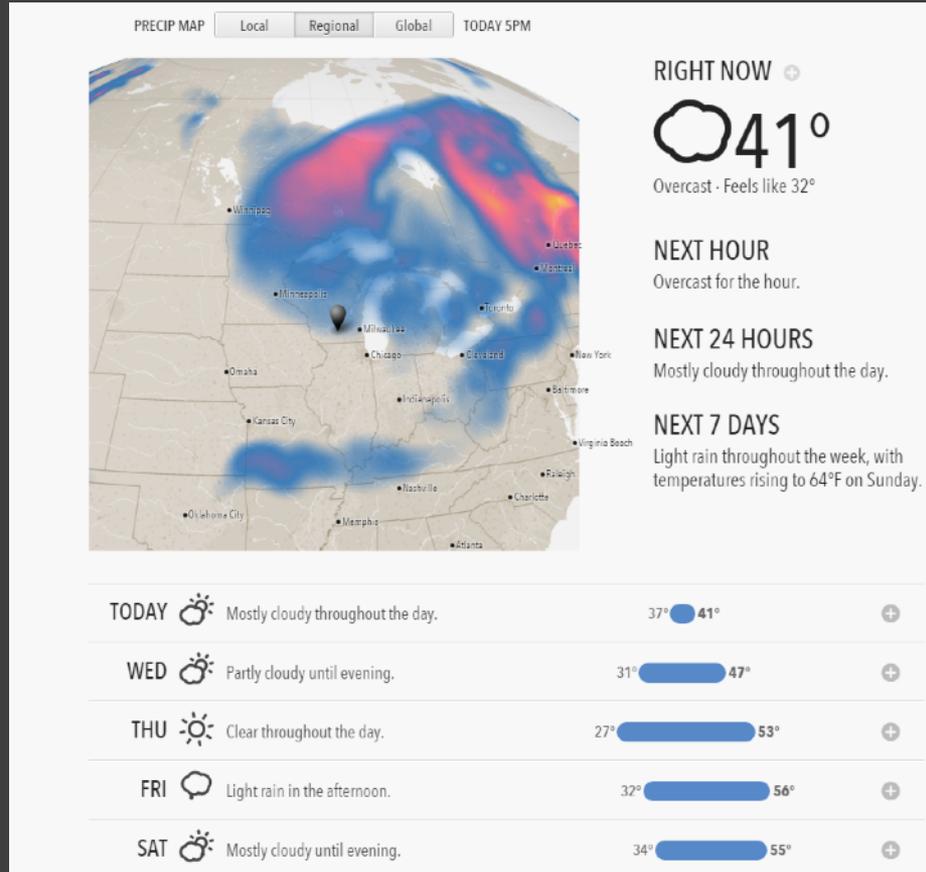
FALSE NEGATIVE

# Model Uncertainty

"50% Chance of Rain"



# Model Uncertainty



# Measurement Uncertainty

**Precision**



# Measurement Uncertainty

## Precision



# Measurement Uncertainty

**Precision**



# Measurement Uncertainty

**Precision**



**Accuracy**



# Measurement Uncertainty

**Precision**



**Accuracy**



# Measurement Uncertainty

**Precision**

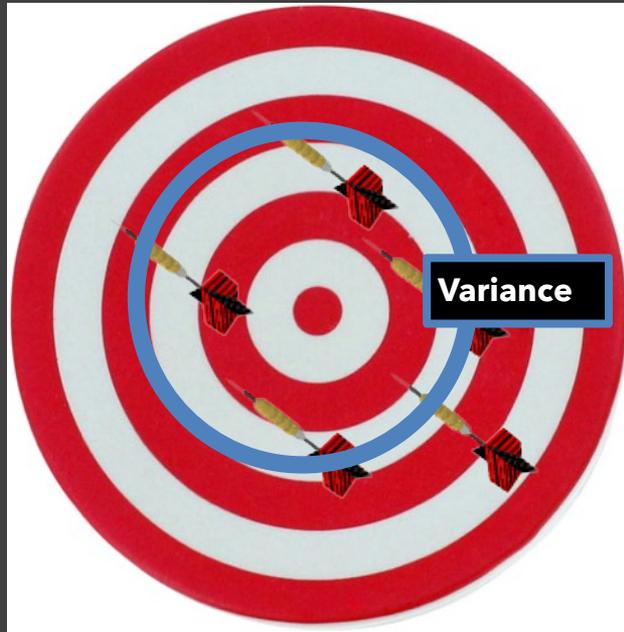


**Accuracy**

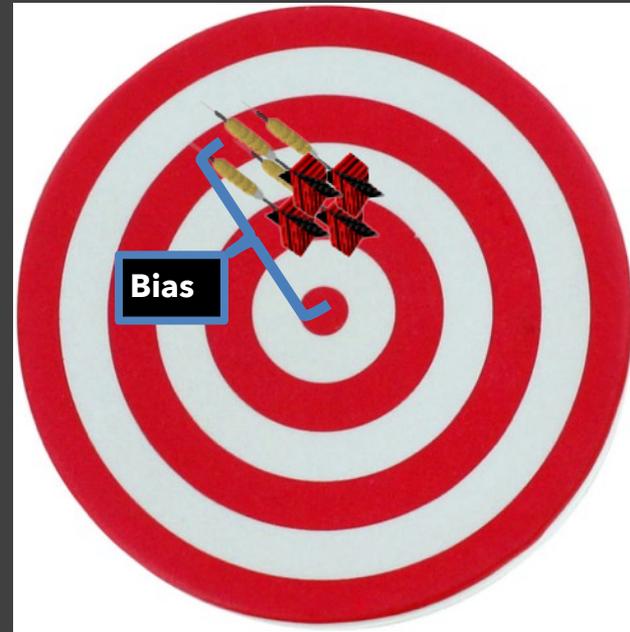


# Measurement Uncertainty

**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**

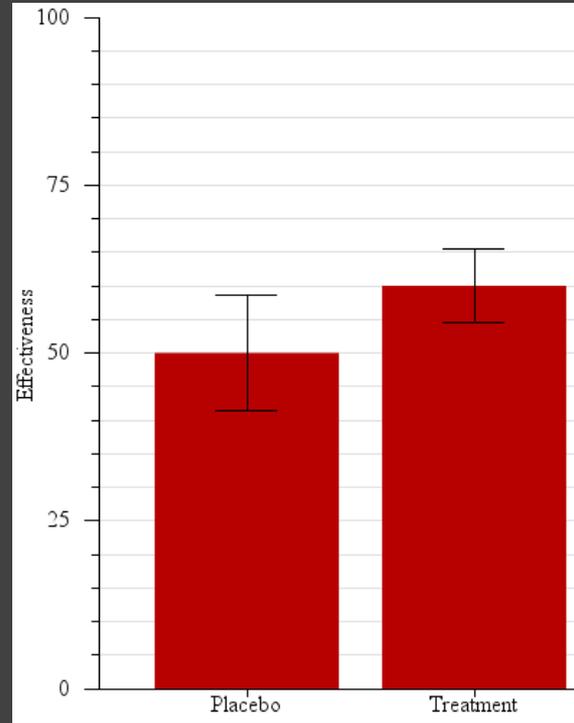
# Uncertainty Visualization

**Avoid Prematurely  
Suppressing Uncertainty**

# Error Bars

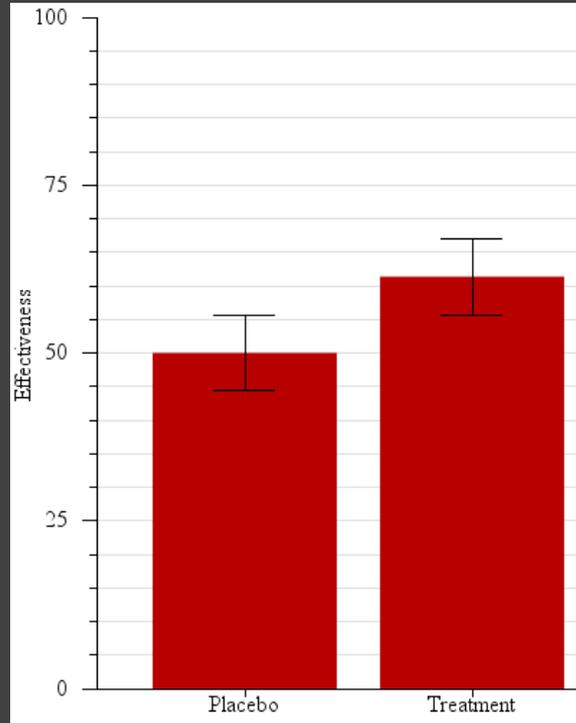
The mean treatment effect is higher than than the placebo.

Is this difference in means *statistically significant*?



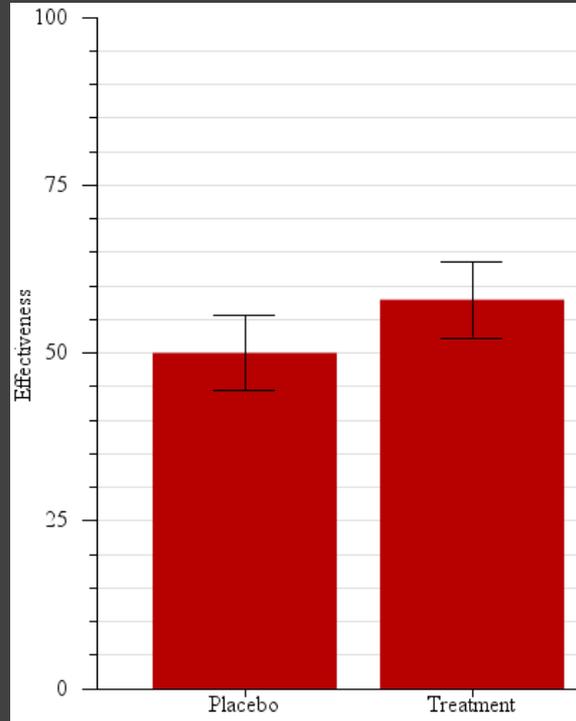
# Guess the p-value...

Error bars depict  
*95% Conf. Interval*



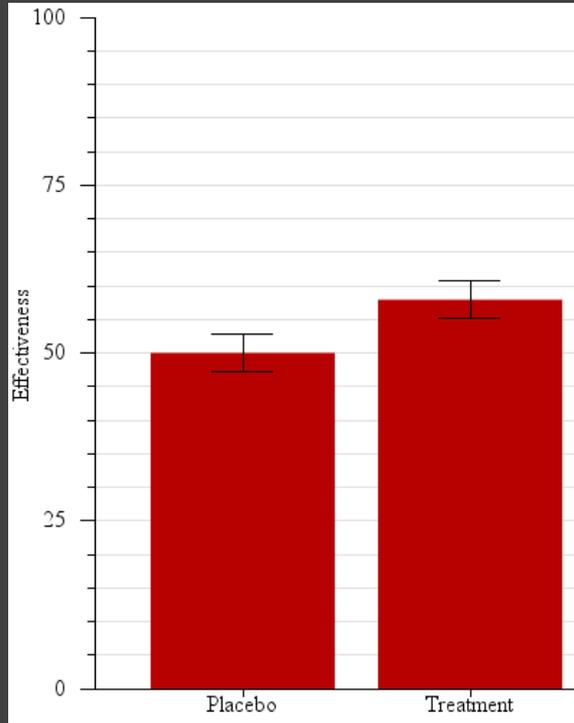
# Guess the p-value...

Error bars depict  
*95% Conf. Interval*

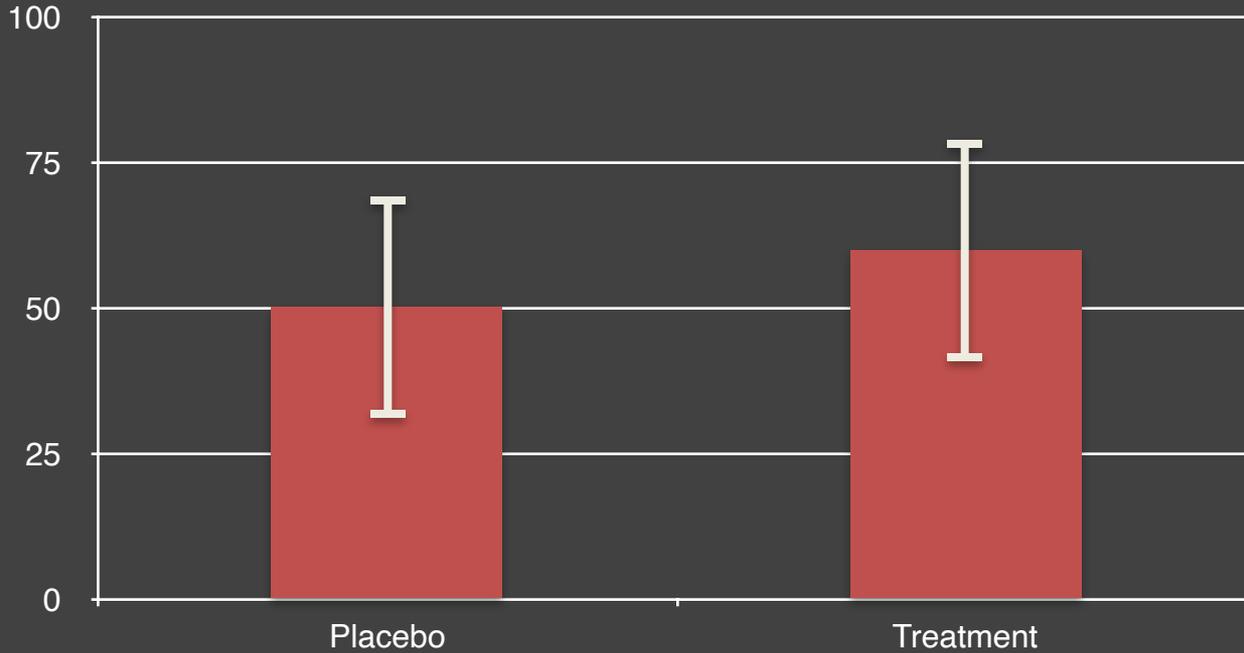


# Guess the p-value...

Error bars depict  
*standard error*



# Misplaced Emphasis?



# Misplaced Emphasis?



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

# 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.

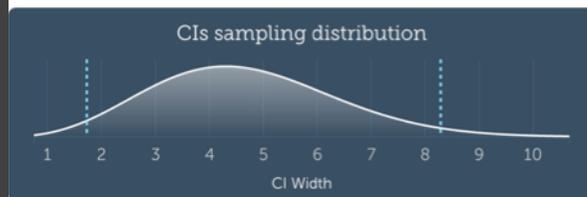
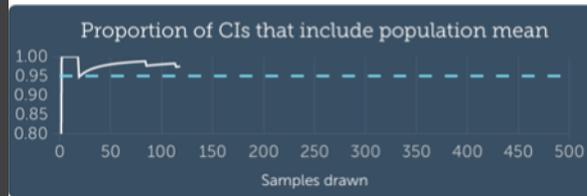
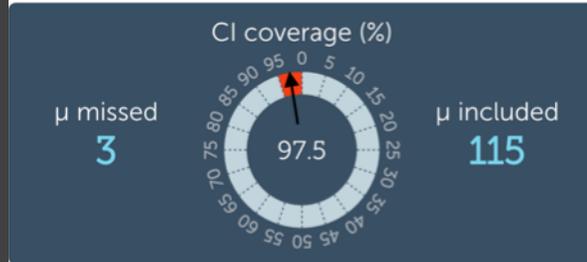
**Wrong!**

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

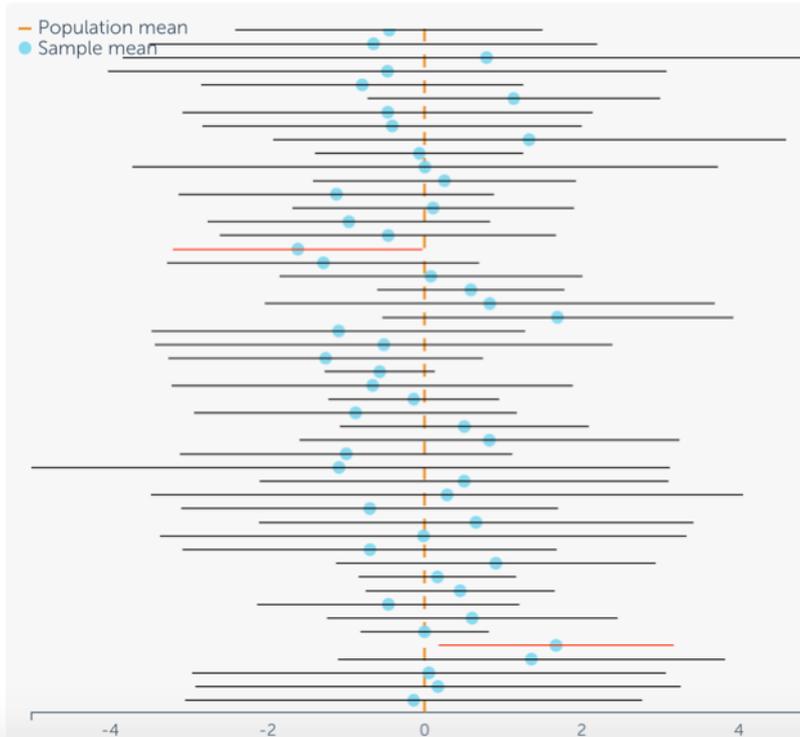
“Confidence” concerns the procedure, not the data. (Though see Bayesian *credible intervals*...)

# Confidence Intervals

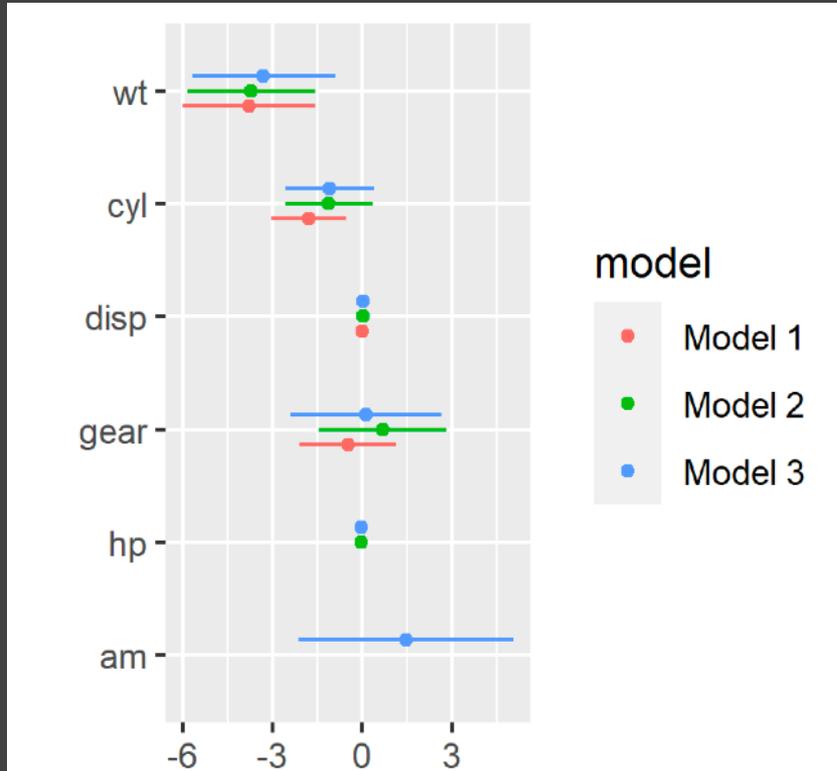
## Simulation statistics



## 95% confidence intervals



# Regression Coefficients



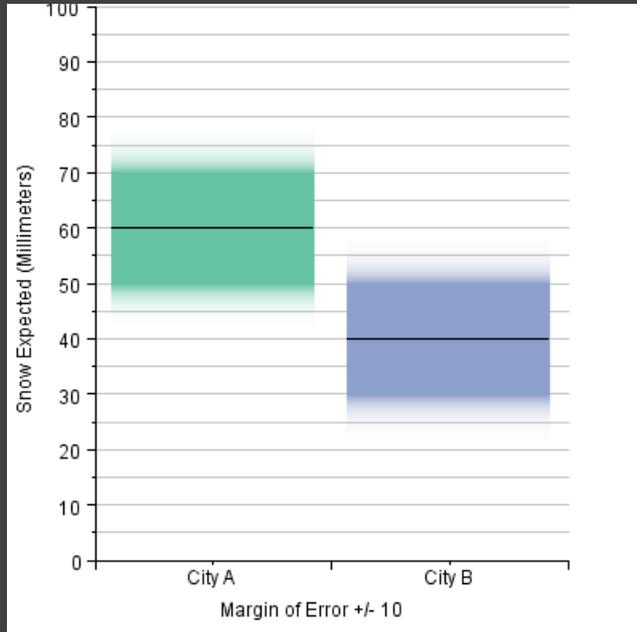
95% CIs for regression model parameters.

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

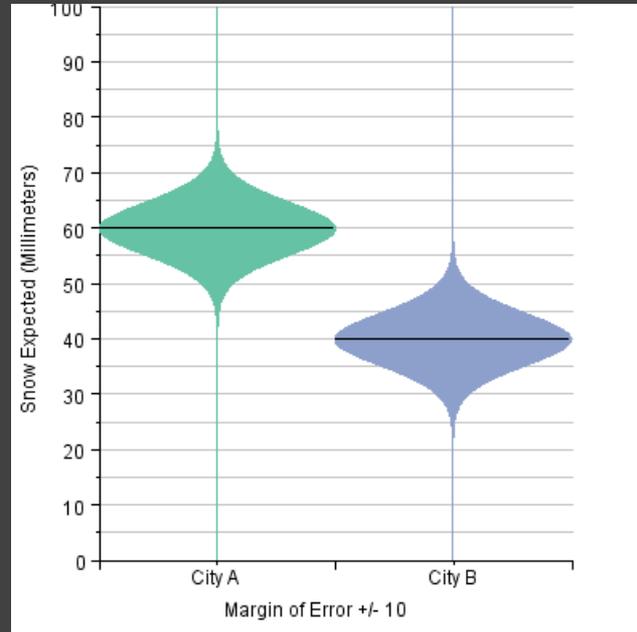
Visual comparison: does the CI overlap 0?

# Alternatives to Error Bars

## Gradient Plot



## Violin Plot



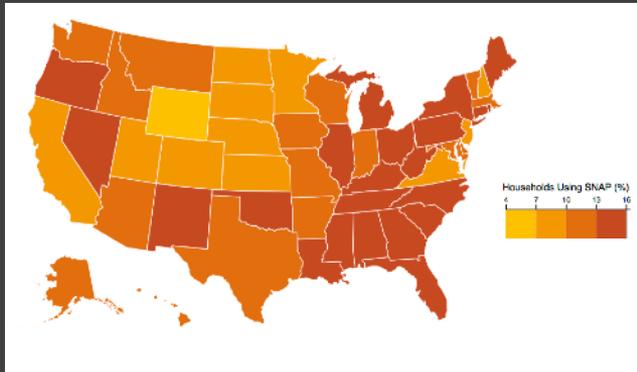
# Visual Encodings of Uncertainty

# Uncertainty Vis Pipeline

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

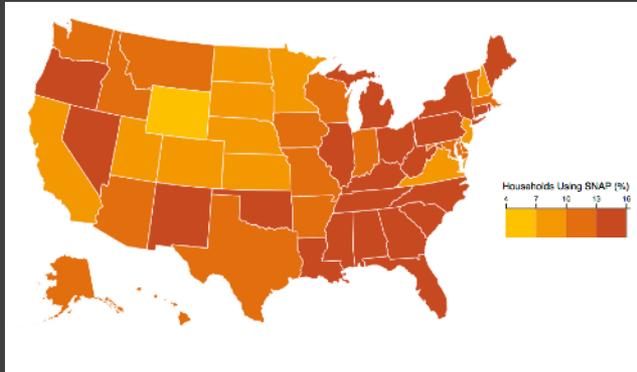
# SNAP

## Data Map

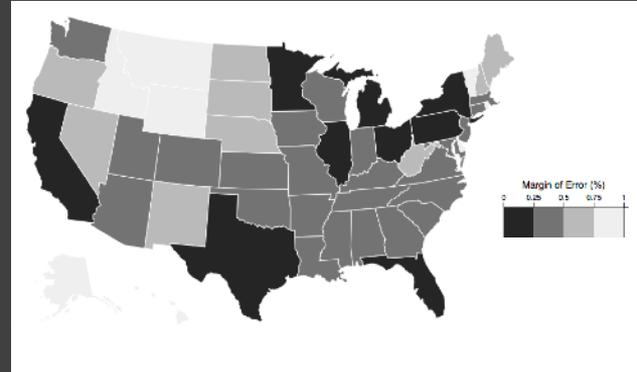


# SNAP

Data Map



Uncertainty Map



# Uncertainty Vis Pipeline

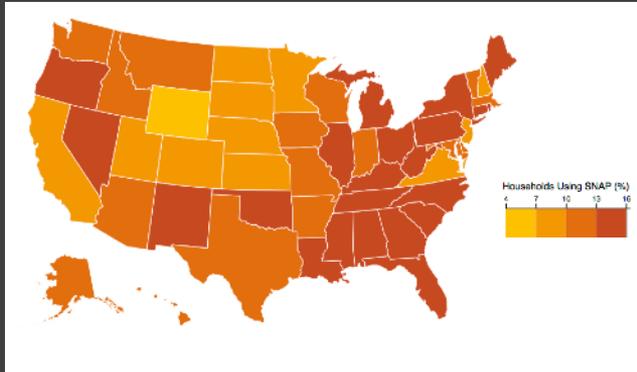
- 1) Quantify uncertainty
- 2) Choose a free visual variable
- 3) Encode uncertainty with the variable

# Uncertainty Vis Pipeline

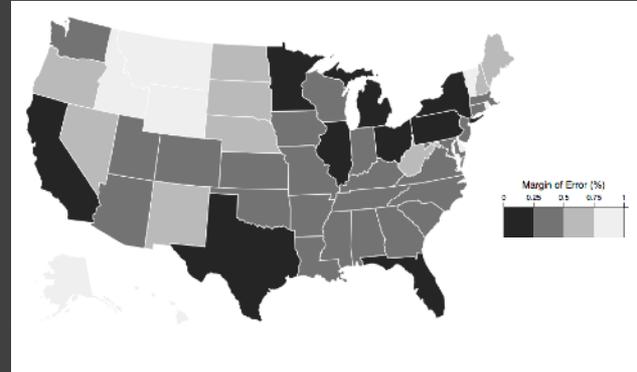
- 1) Quantify uncertainty
- 2) Choose a free visual variable
- 3) Encode uncertainty with the variable
- 4) Unify the Data Map and Uncertainty Map

# How to Unify?

Data Map

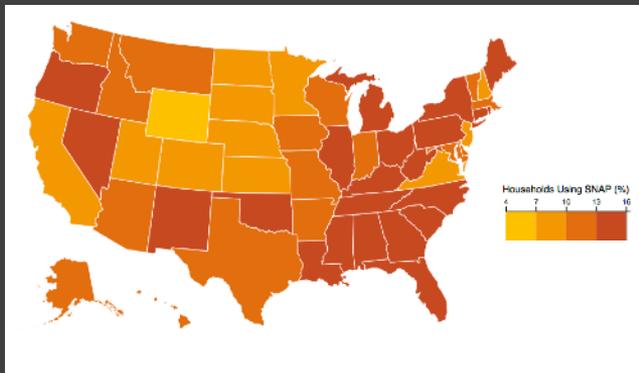


Uncertainty Map

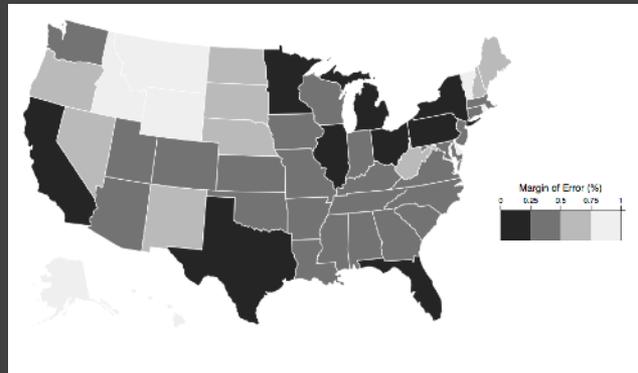


# Juxtaposition

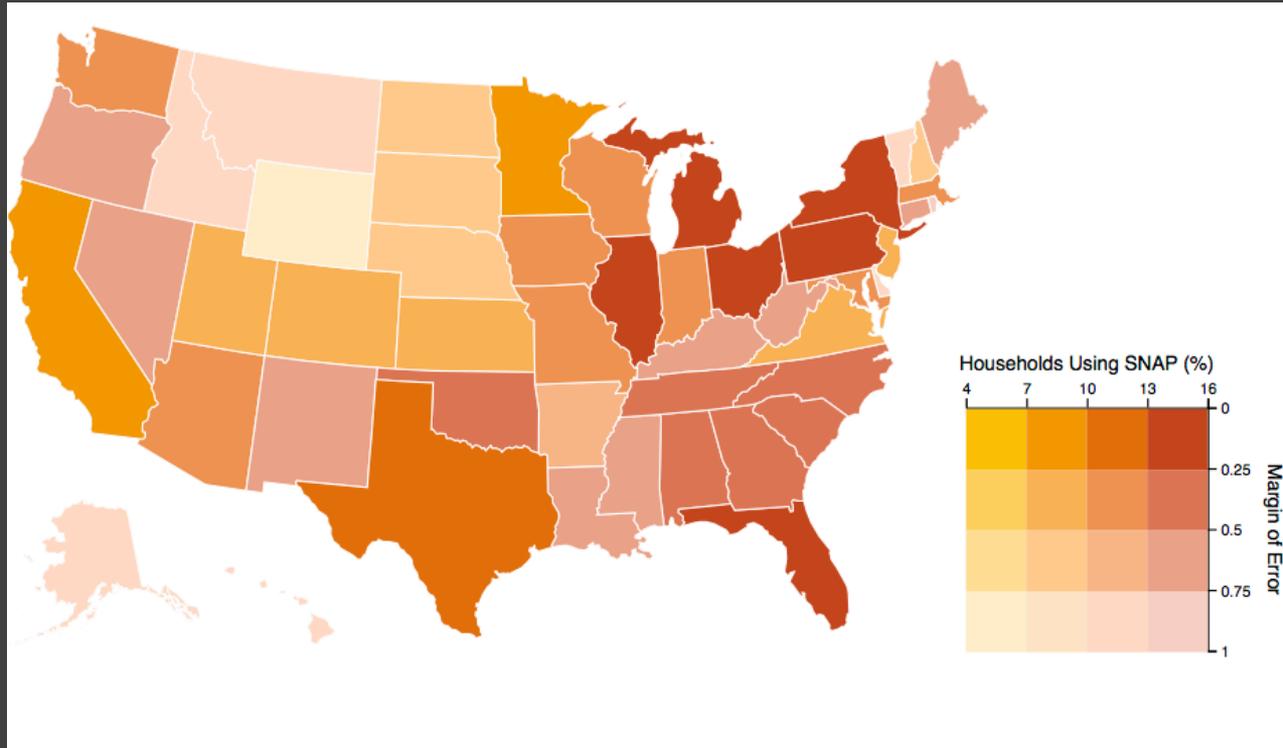
Data Map



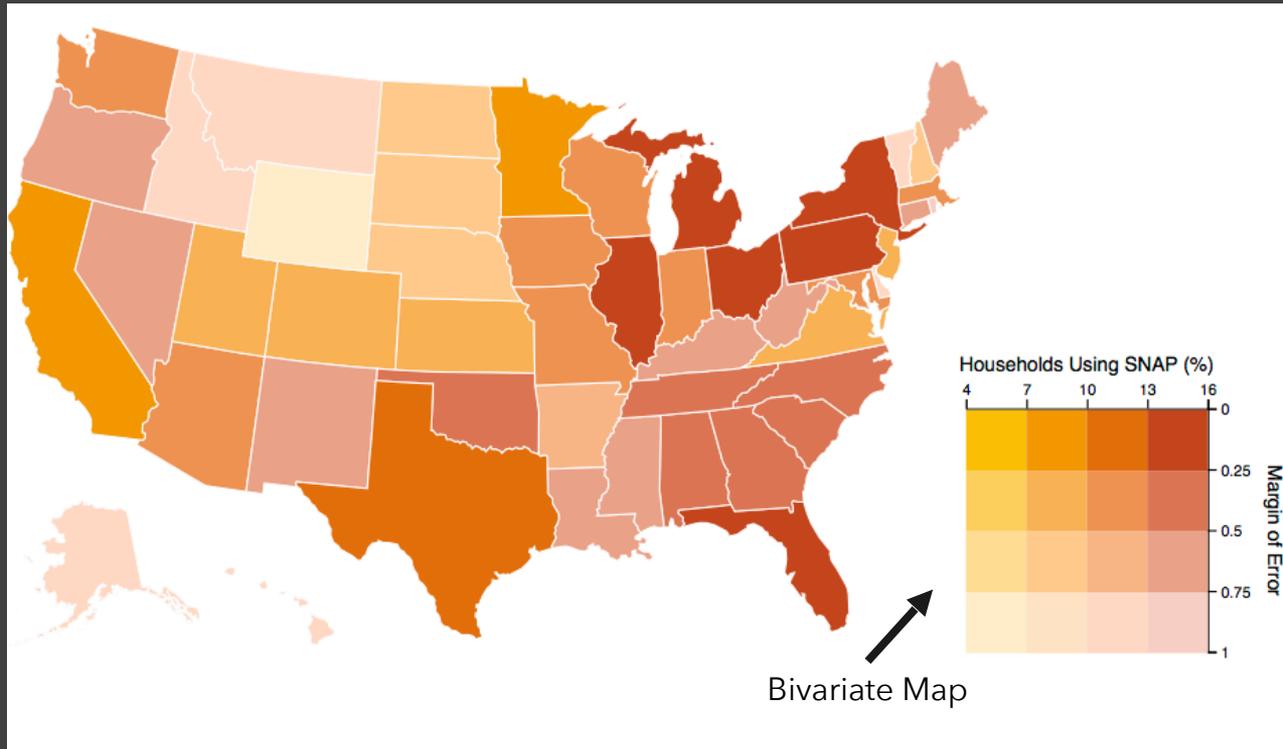
Uncertainty Map



# Superposition



# Superposition



# 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

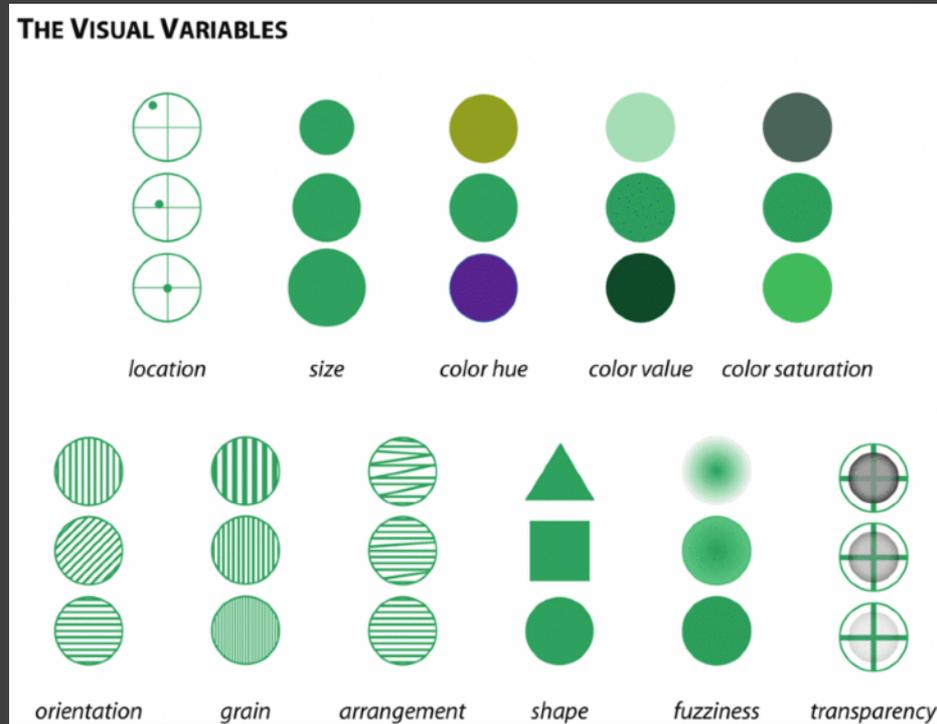


*Ceci n'est pas une pipe.*

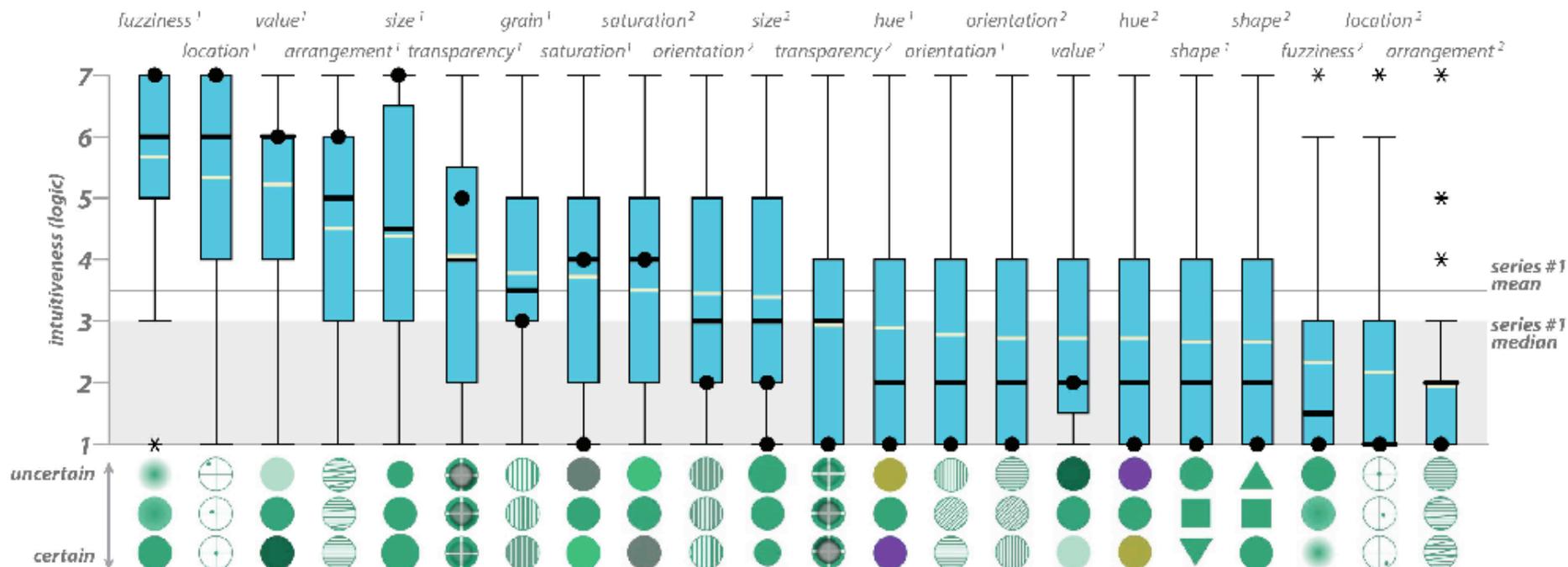
# Semiotics of Uncertainty



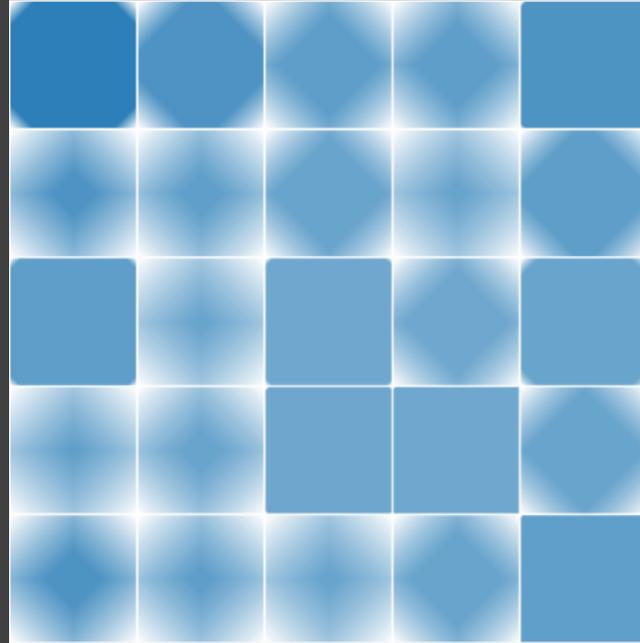
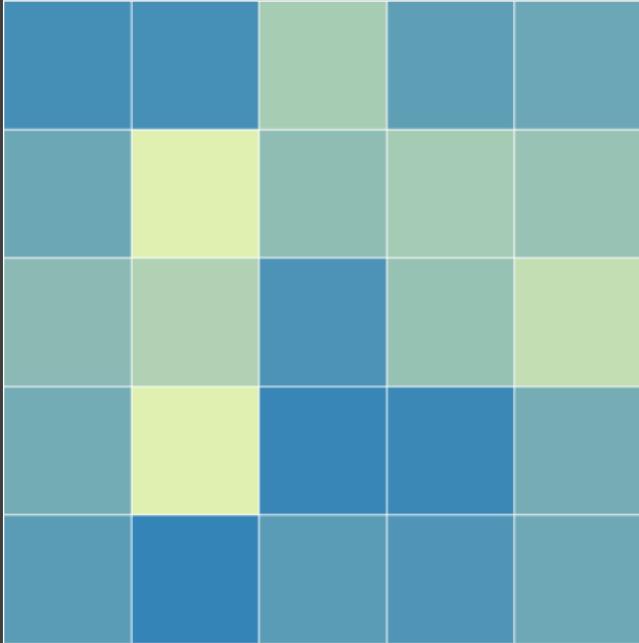
# Semiotics of Uncertainty



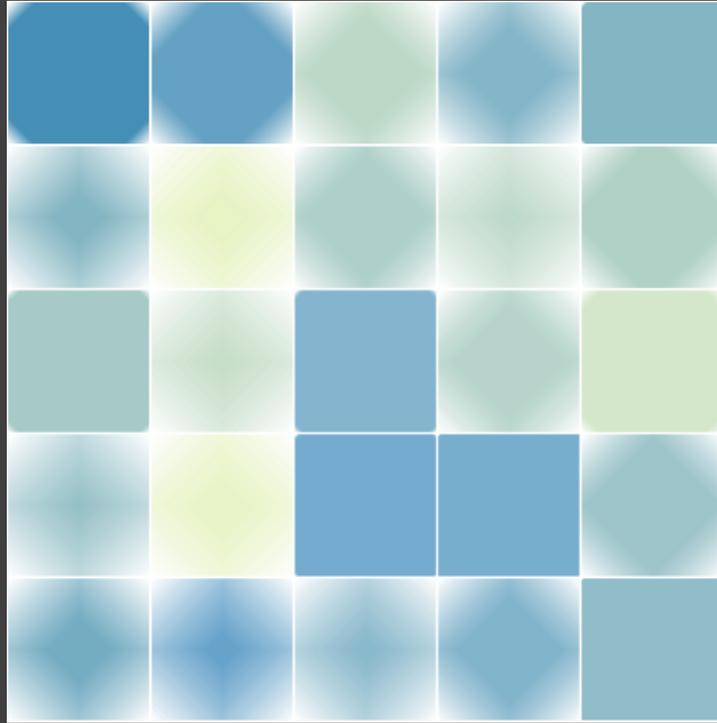
# SERIES #1: GENERAL UNCERTAINTY BY VISUAL VARIABLE



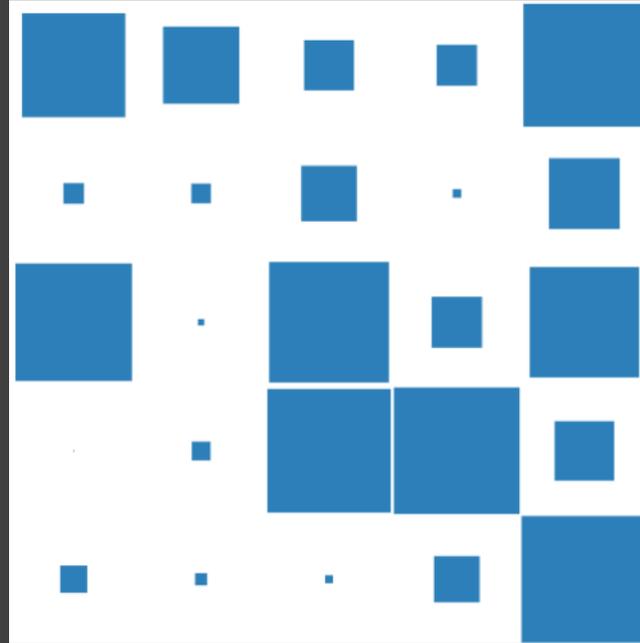
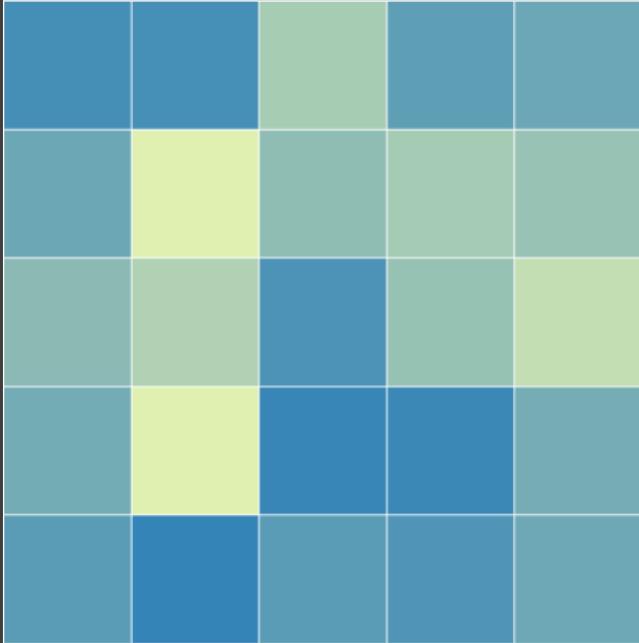
# Fuzziness Juxtaposition



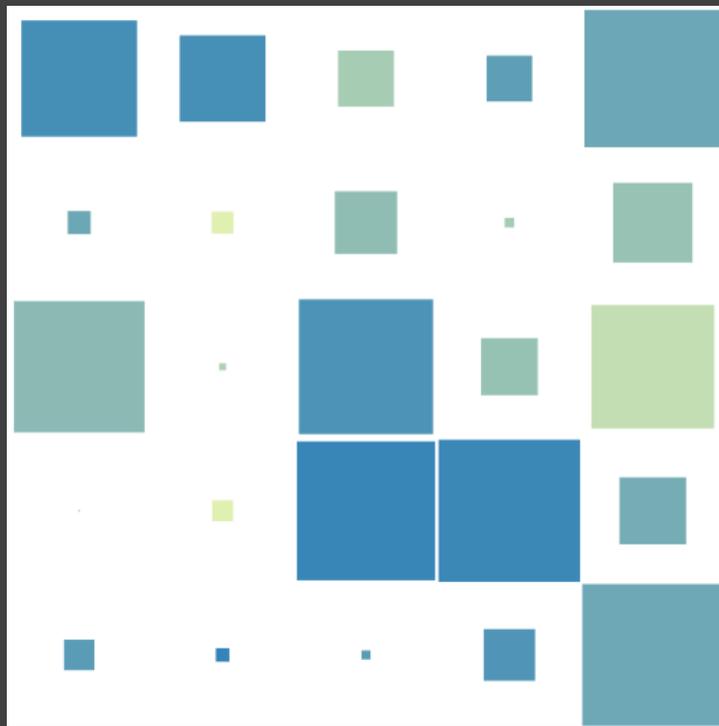
# Fuzziness Superposition



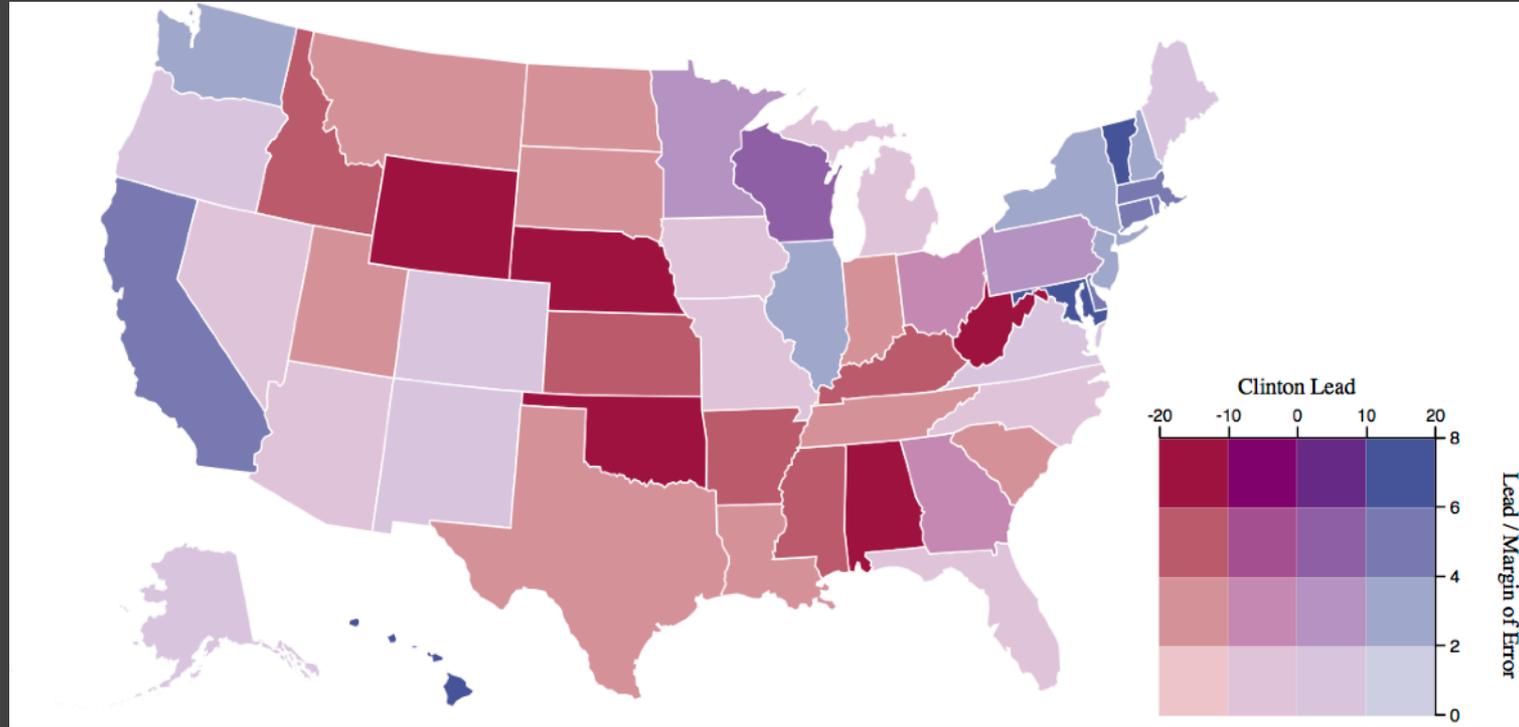
# Size Juxtaposition



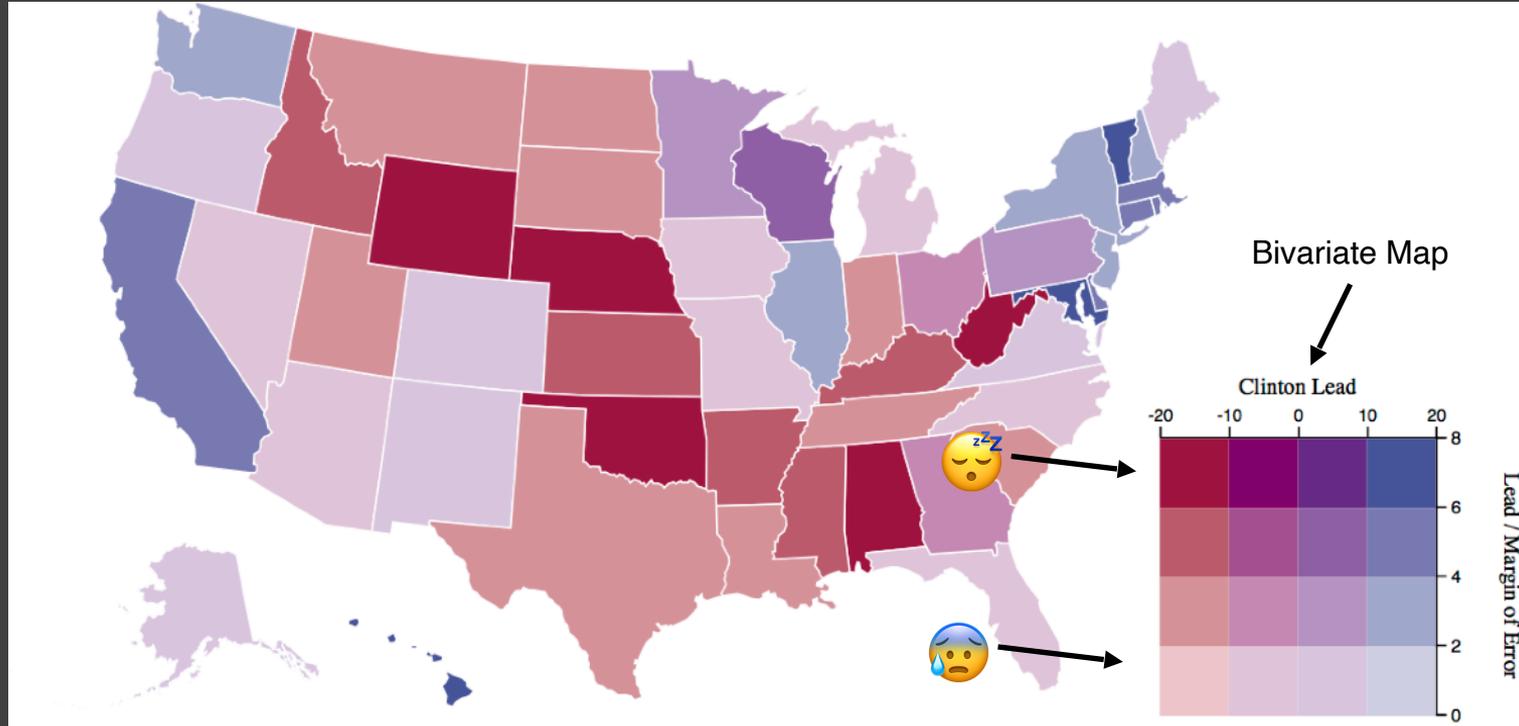
# Size Superposition



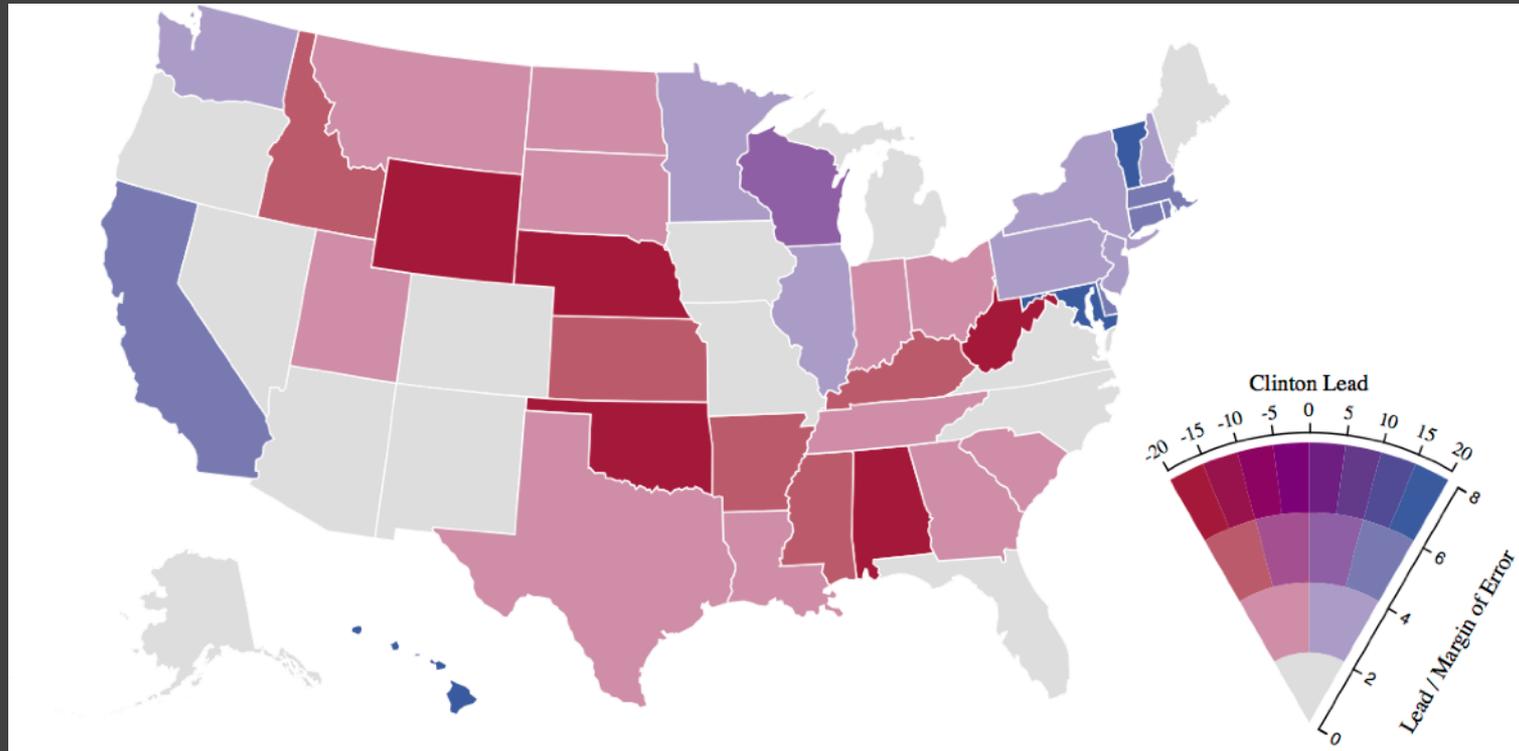
# Value Suppressing Uncertainty Palettes



# Bivariate Map

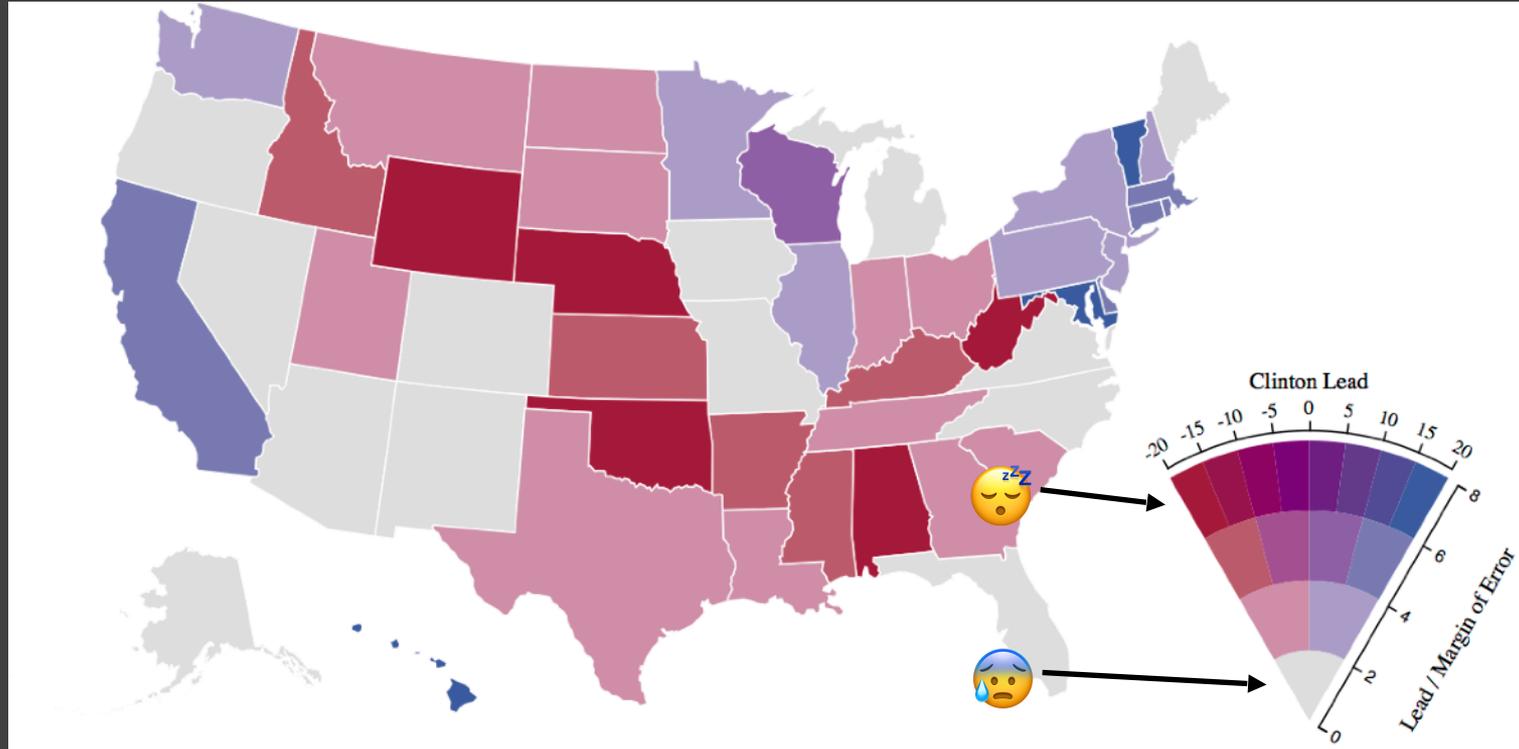


# Value Suppressing Uncertainty Palettes



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

# Value Suppressing Uncertainty Palettes

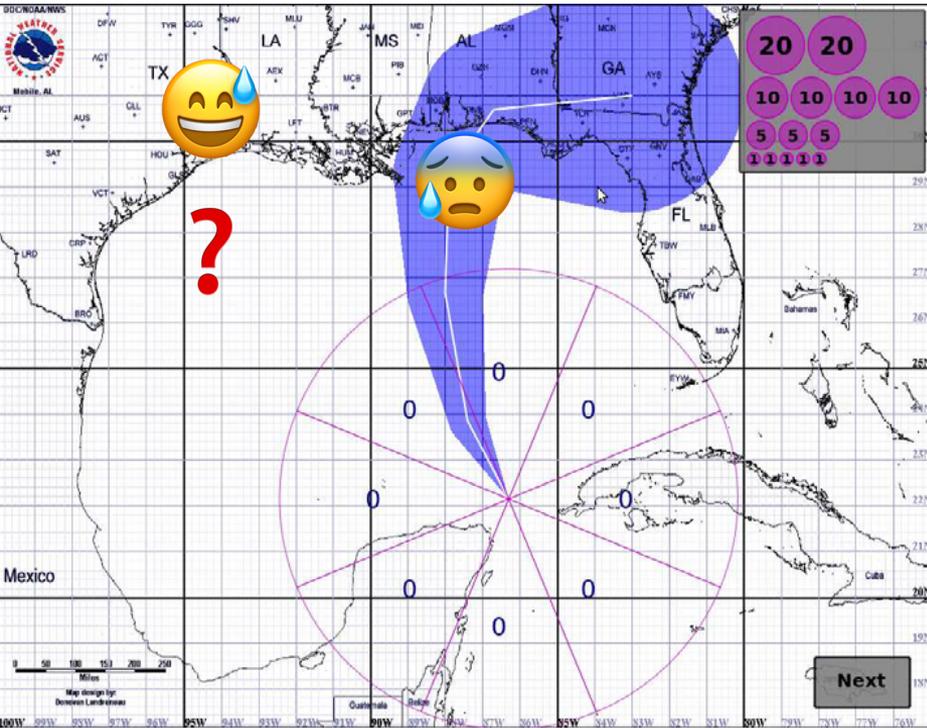


# Encoding Uncertainty

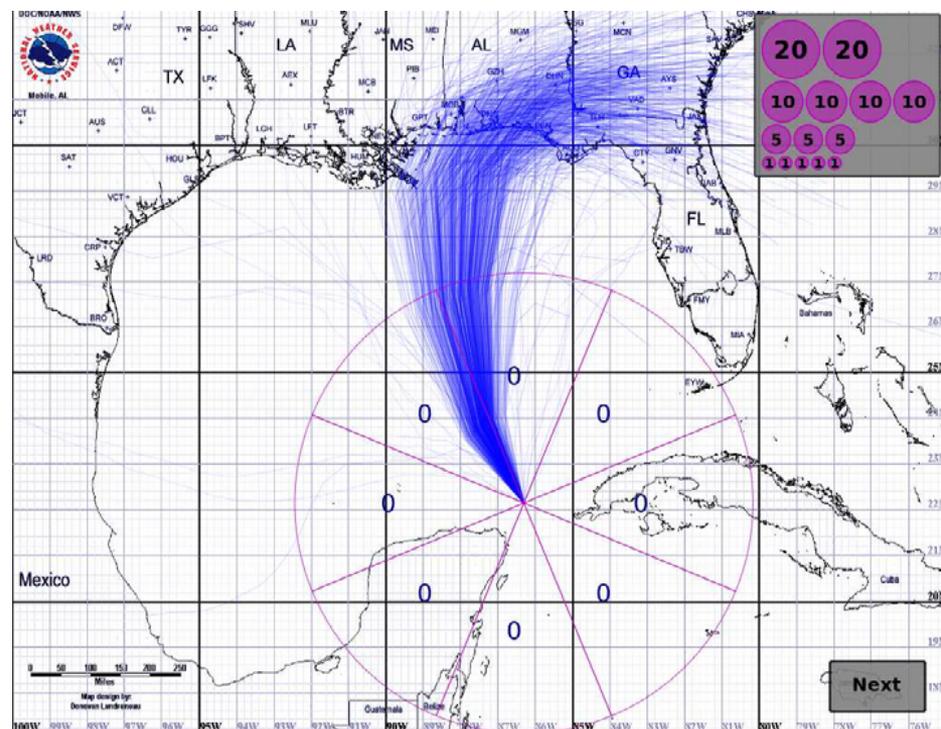
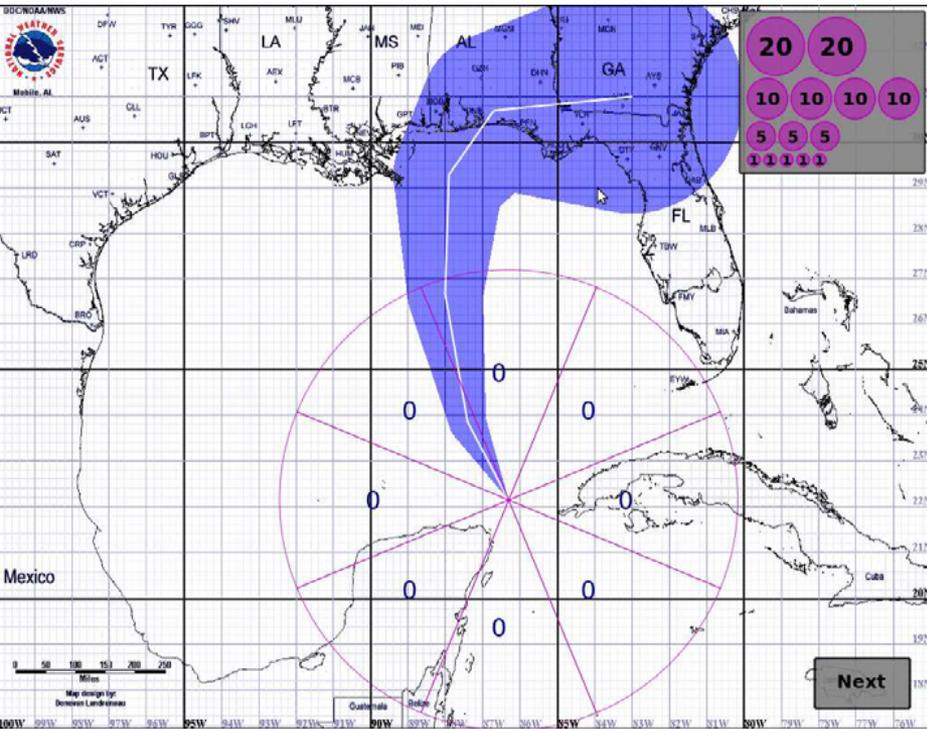
Some visual variables (like fuzziness and value) have a **semiotic connection** to uncertainty.

However, intuitive variables may not always be accurately interpreted!

# Frequency Framing & Hypothetical Outcomes



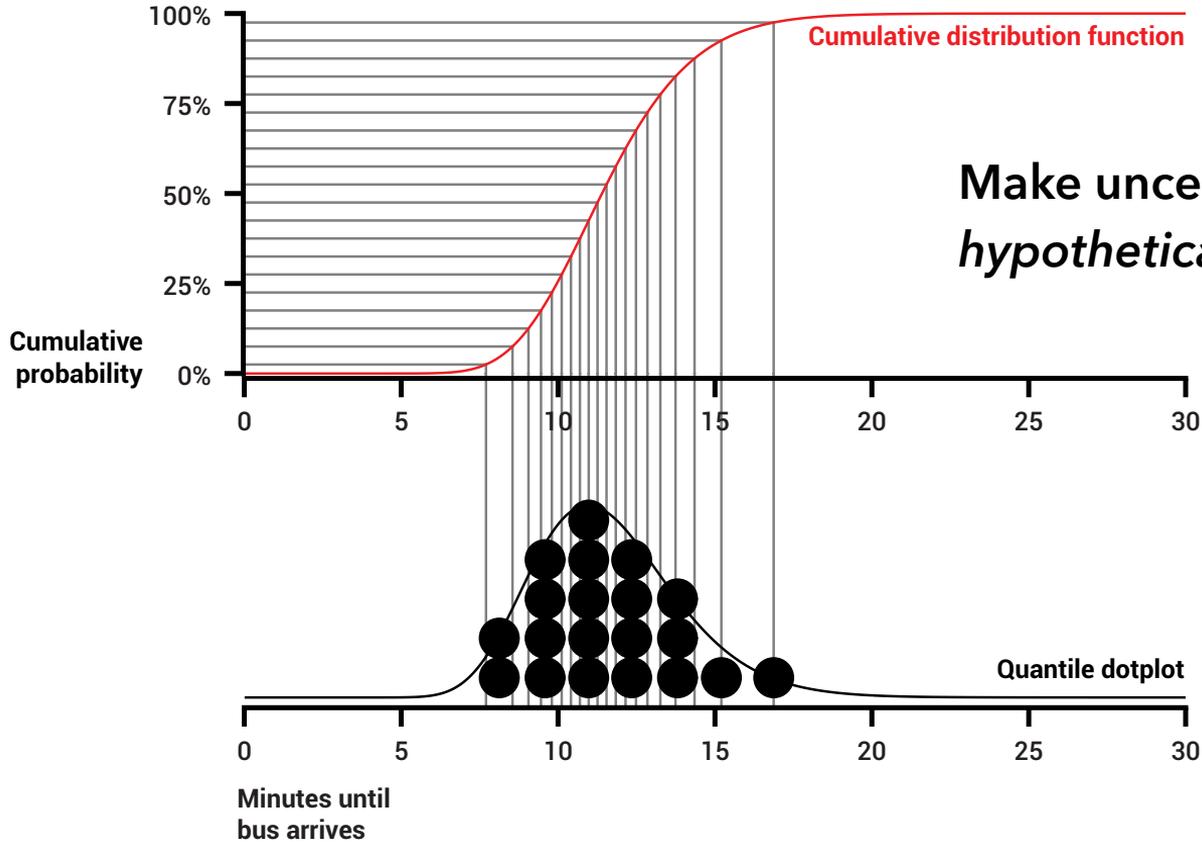
Size or likelihood of hurricane?  
 Is New Orleans safe?



Size or likelihood of hurricane?  
Is New Orleans safe?

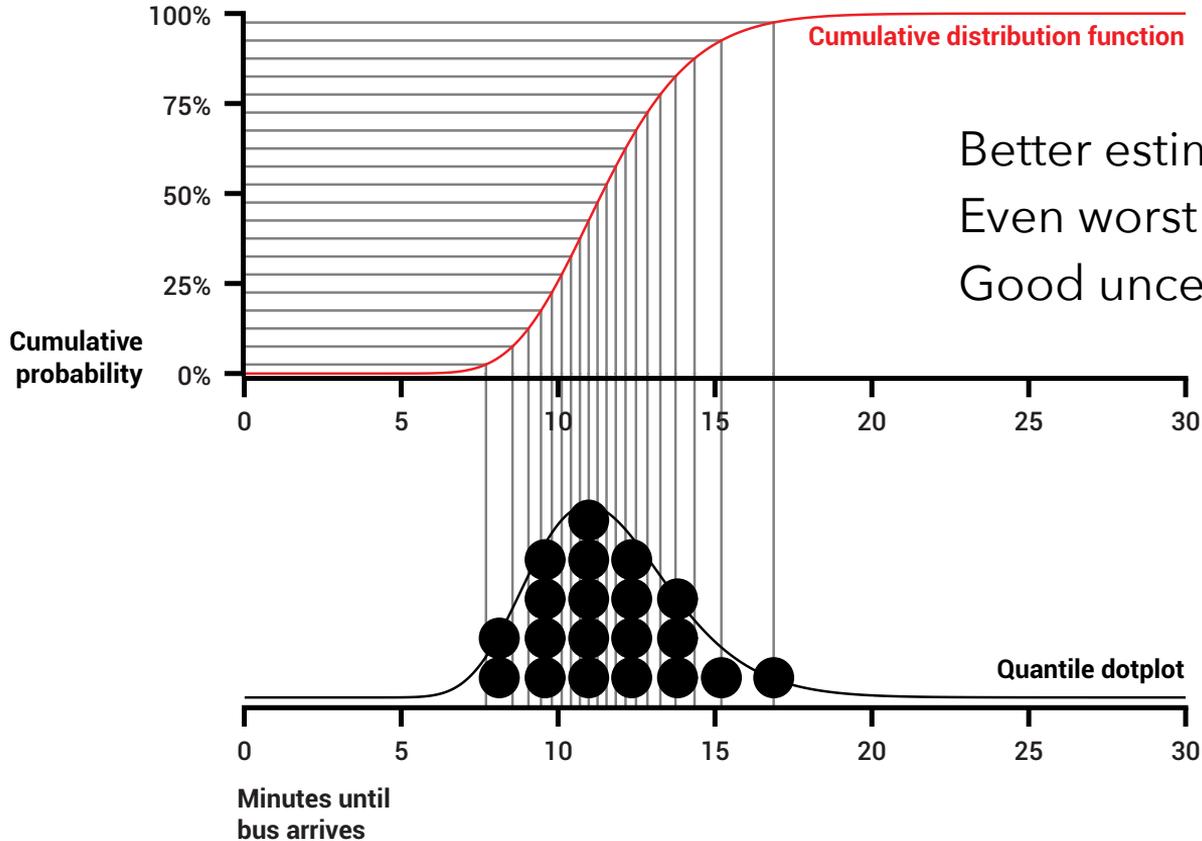
Make uncertainty more concrete  
via *hypothetical outcomes*.

# Predicted Bus Arrival Times



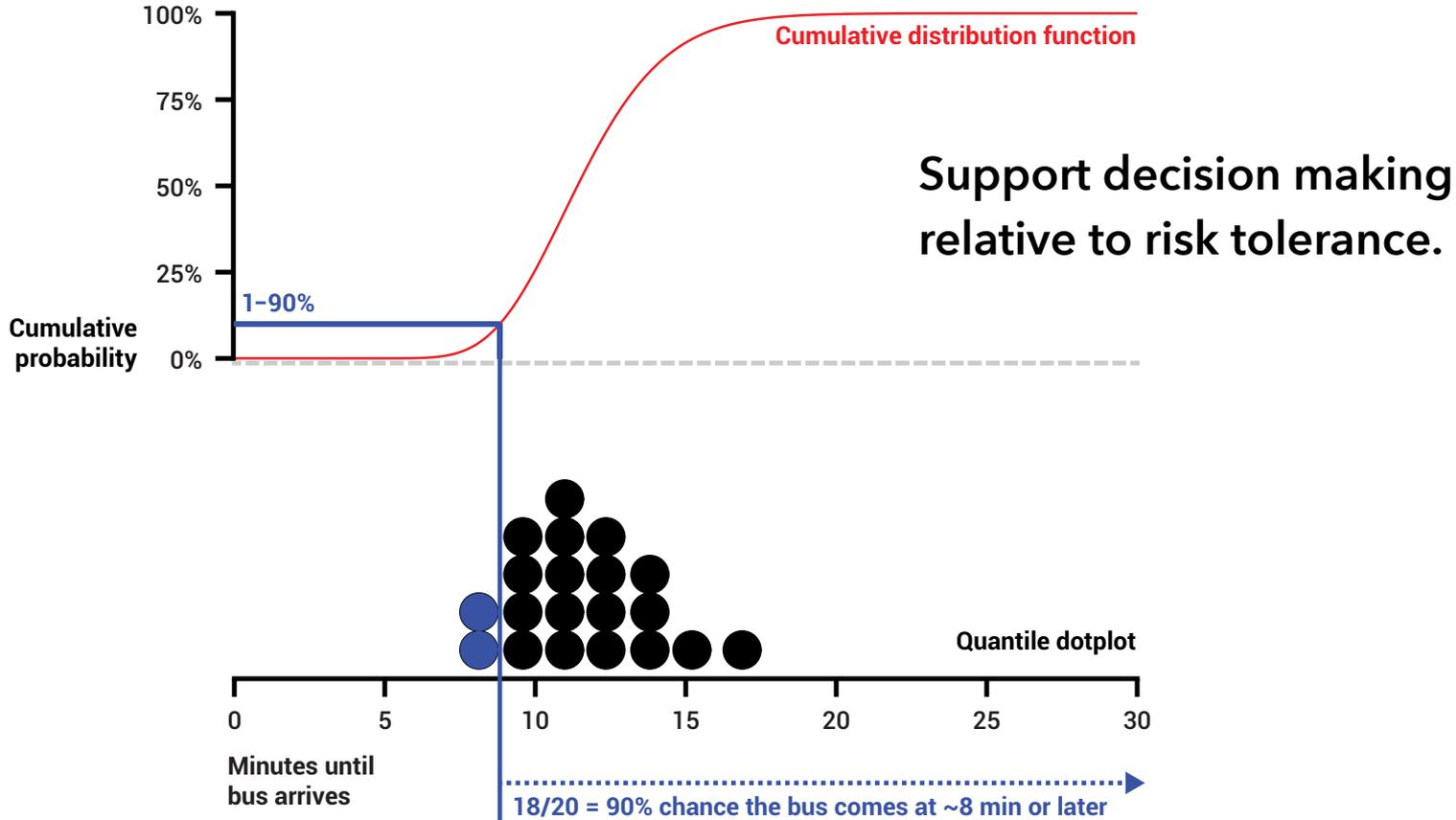
Make uncertainty concrete via  
*hypothetical outcomes.*

# Predicted Bus Arrival Times

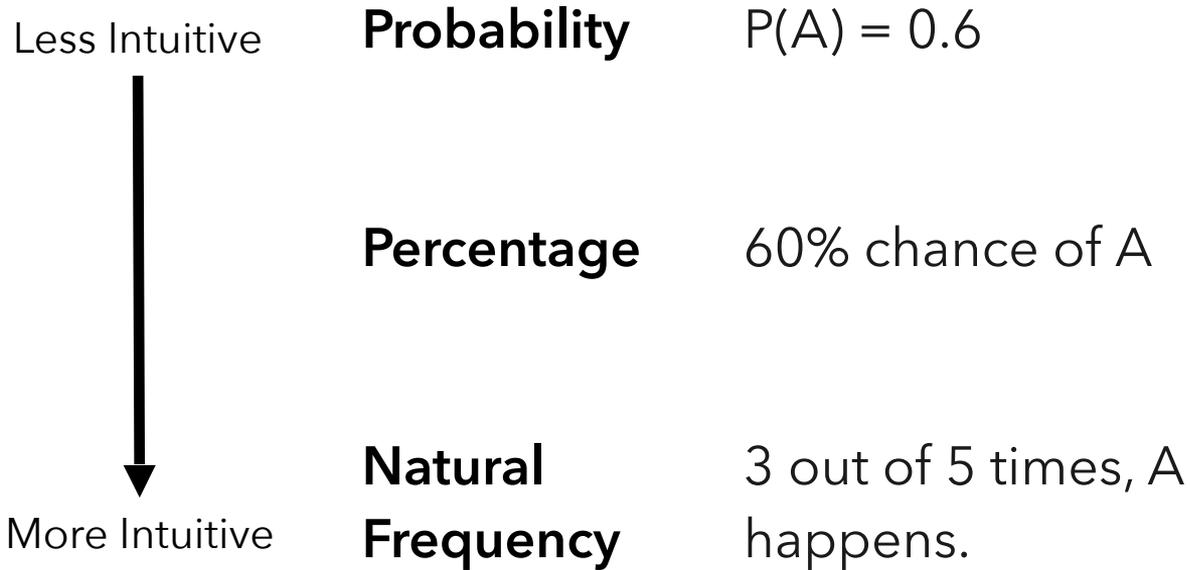


Better estimates, decisions with time.  
Even worst performers improve.  
Good uncertainty displays possible!

# Predicted Bus Arrival Times



# How to Present Probabilities

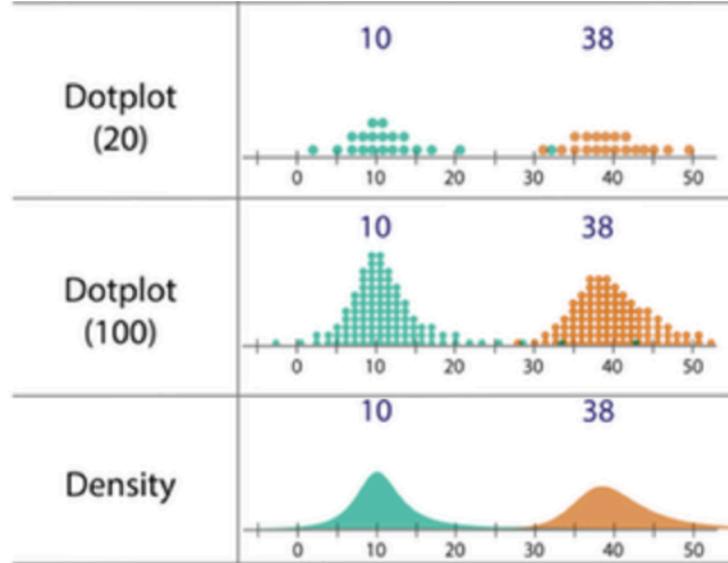


# Quantile Dot Plots

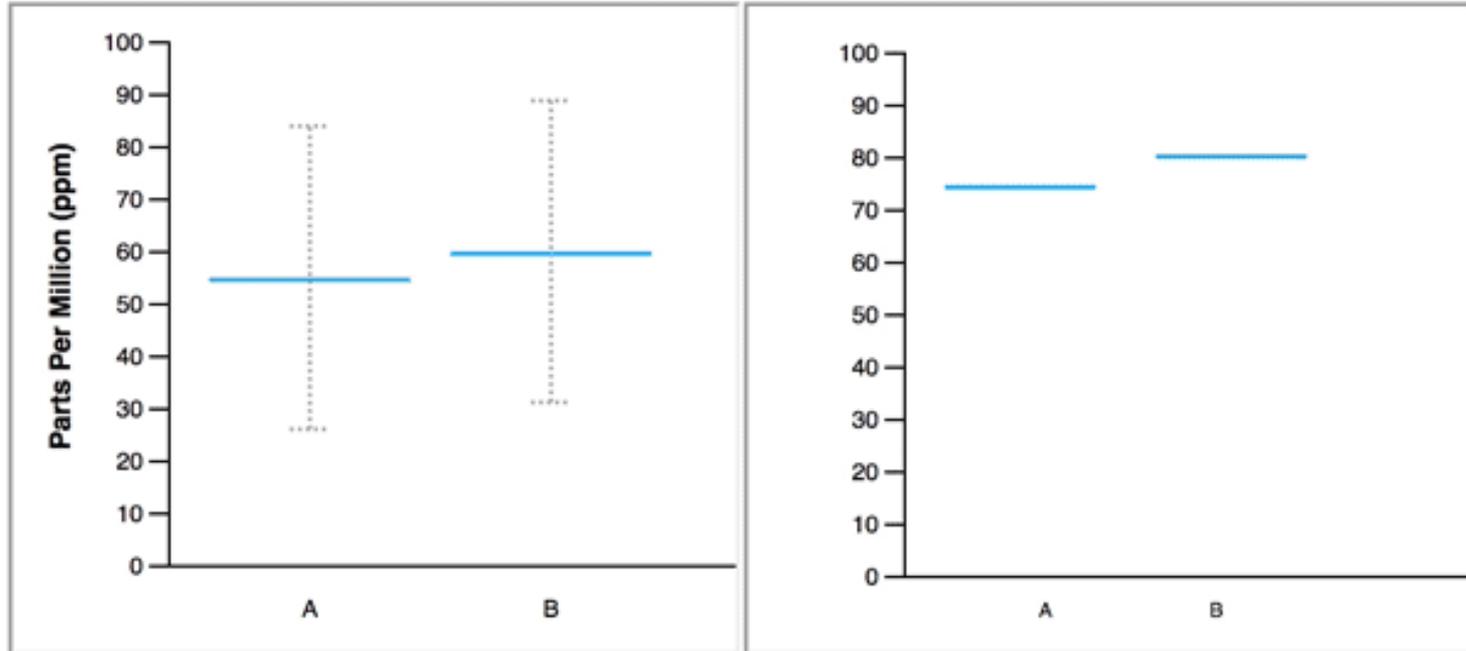
Less Error



More Error

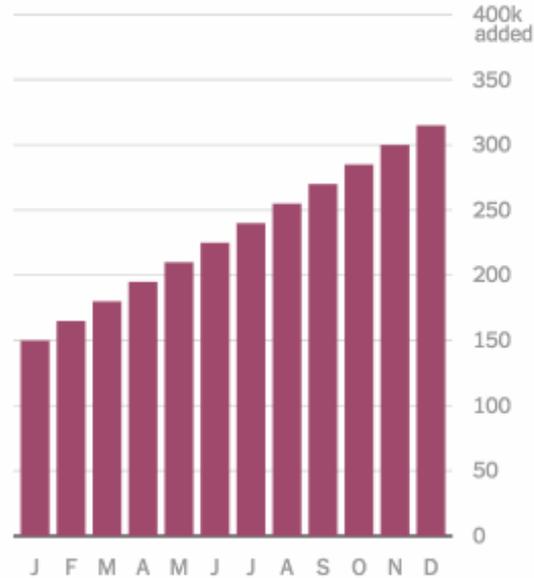


# Hypothetical Outcome Plots



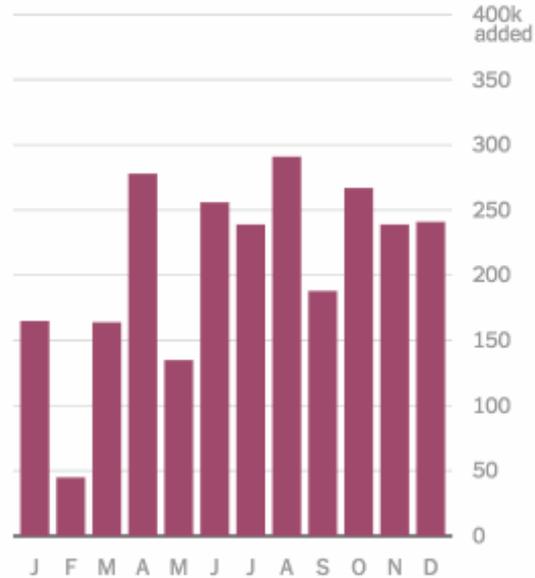
# Hypothetical Outcome Plots

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



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

Pause



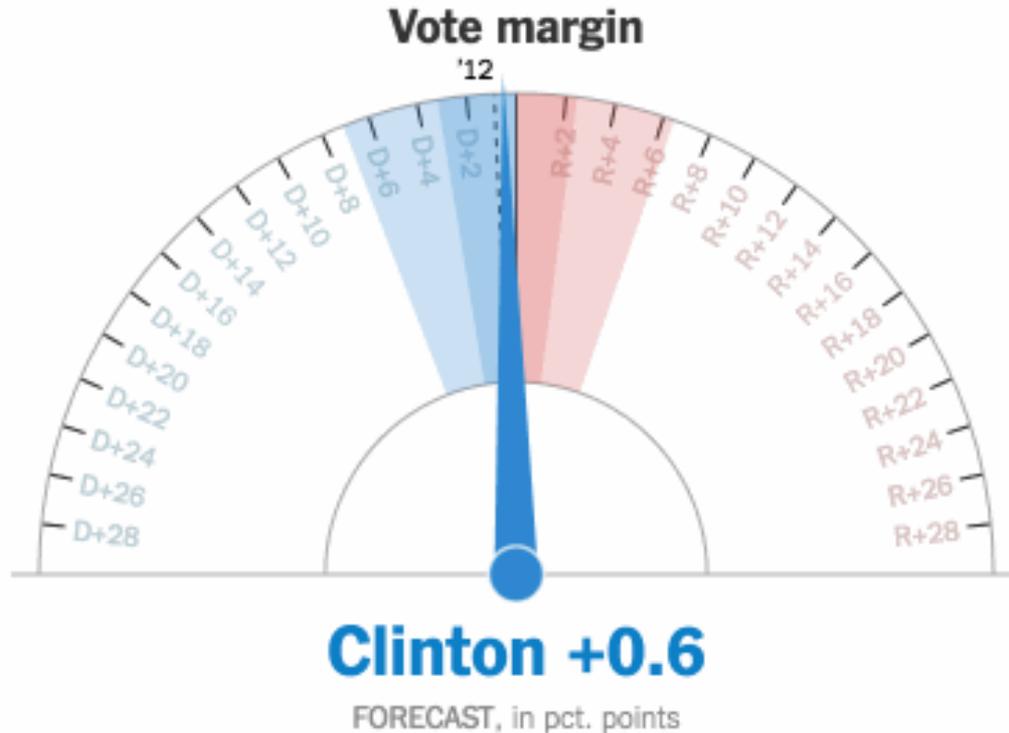
# Hypothetical Outcomes

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.



# The NY Times Needle



# 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.

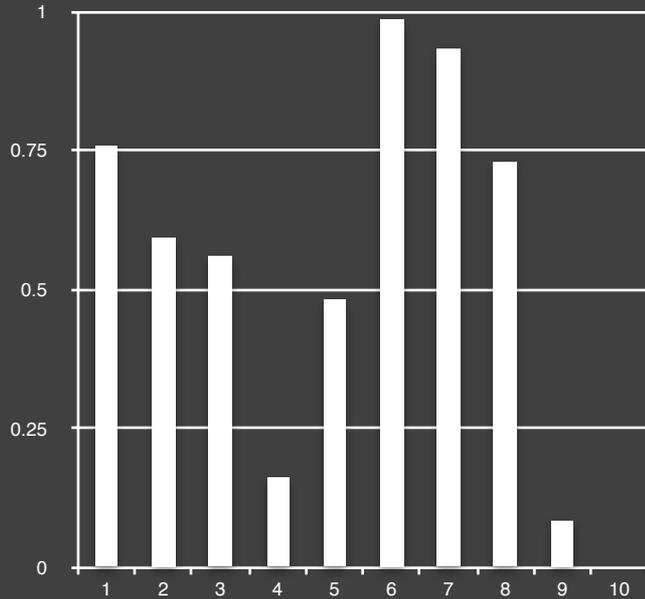
**IT DEPENDS**

**What Can Go Wrong?**

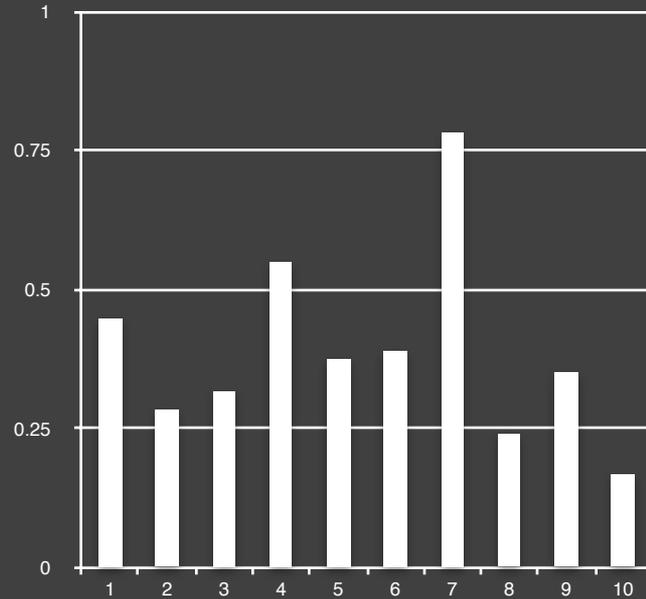
# Inferential Integrity

# Which Stock To Buy?

## Company A

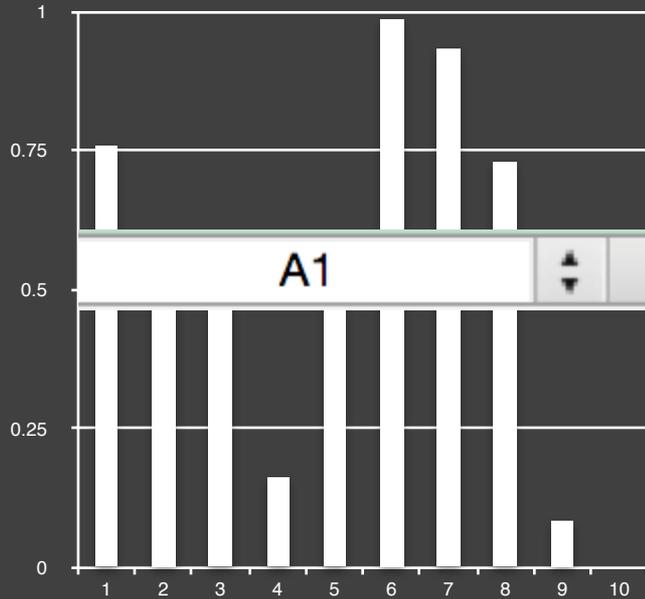


## Company B

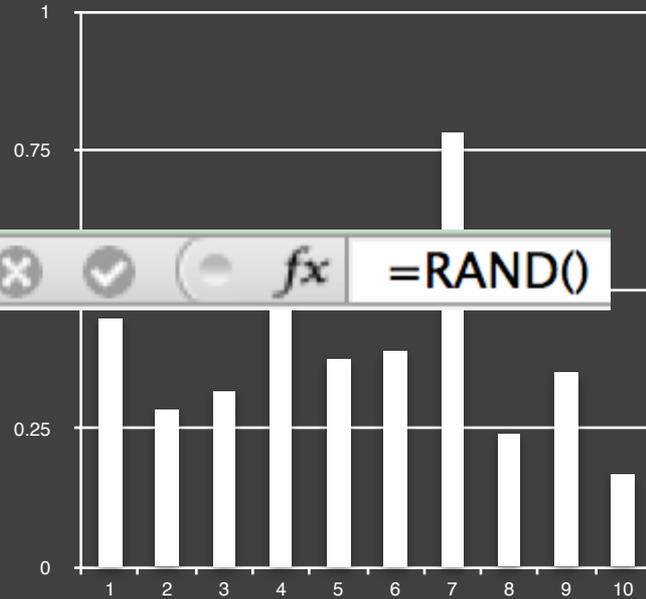


# Neither!

## Company A



## Company B



A1

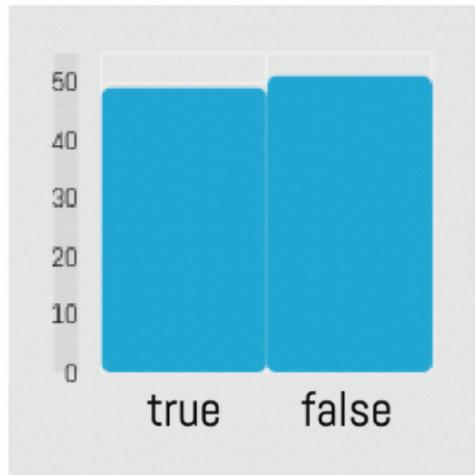


*fx*

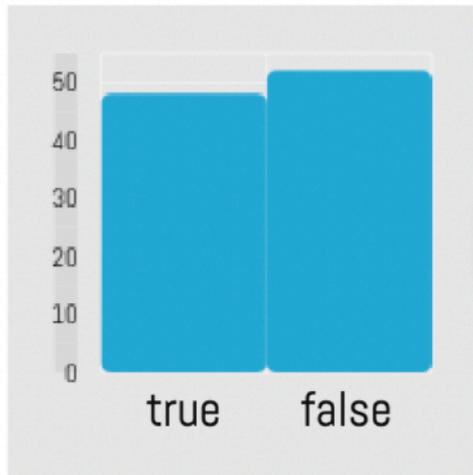
=RAND()

# What Swag Should We Send? [Zgraggen et al. '18]

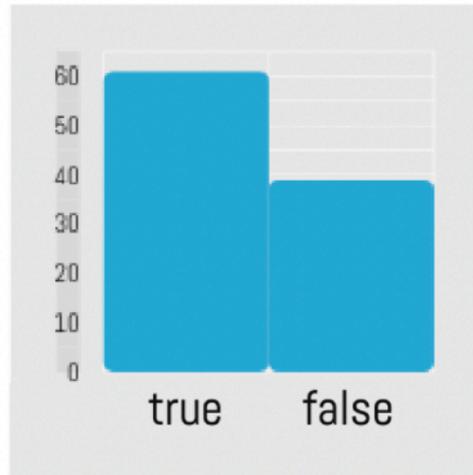
(a) 2006: pen (\$4)



(b) 2007 key chain (\$2)



(c) 2016: USB drive (\$4)

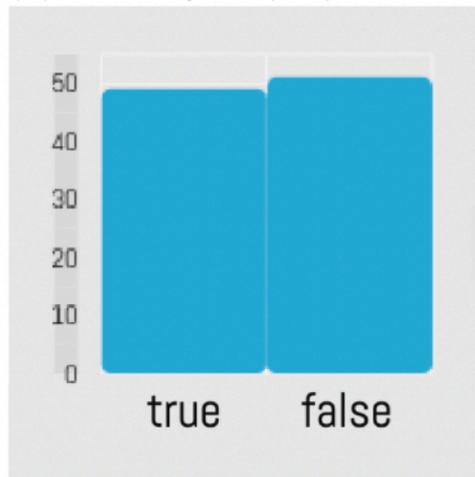


Recurring Donor?

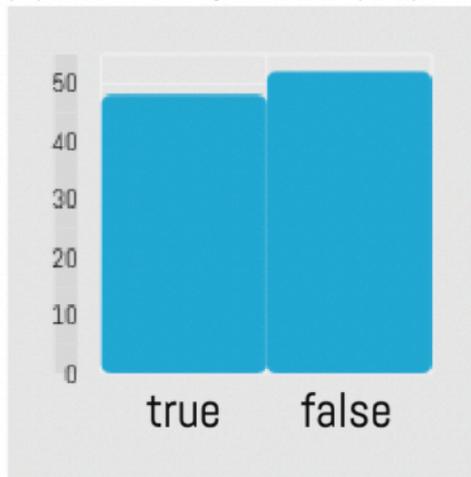
...

# What Swag Should We Send? [Zgraggen et al. '18]

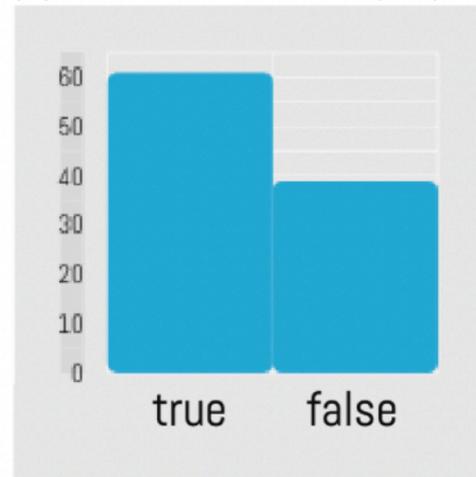
(a) 2006: pen (\$4)



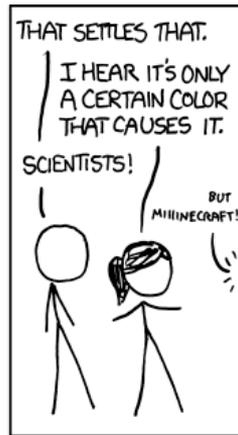
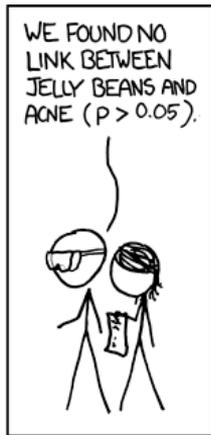
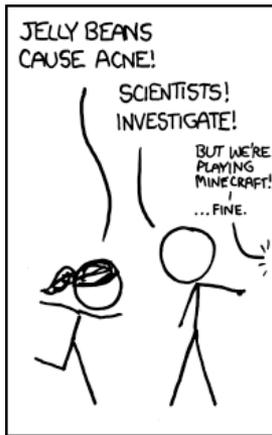
(b) 2007 key chain (\$2)

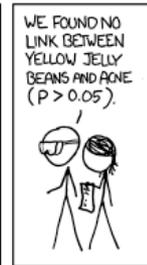
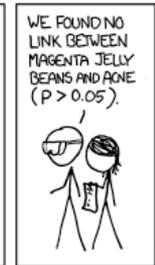
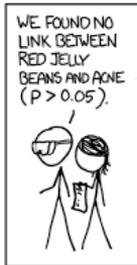
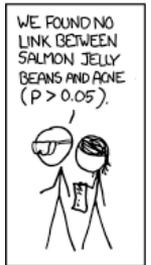
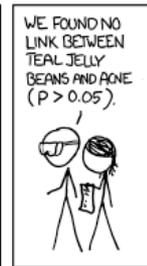
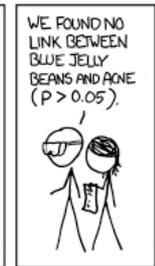
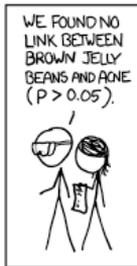
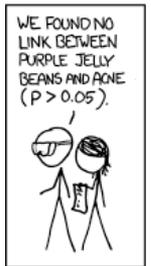


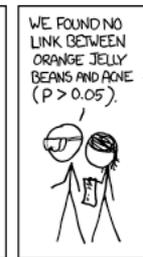
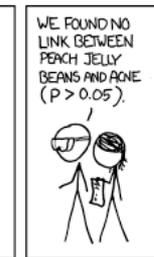
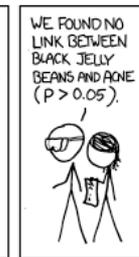
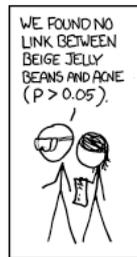
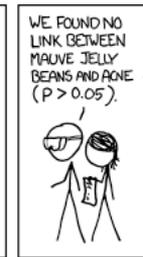
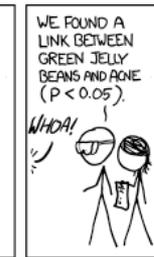
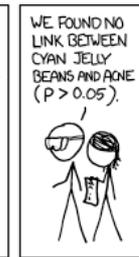
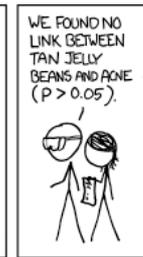
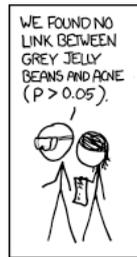
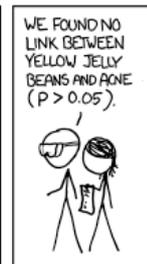
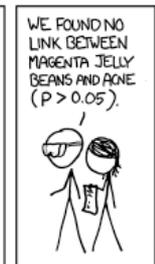
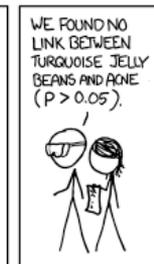
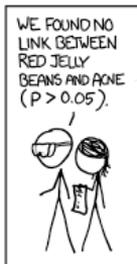
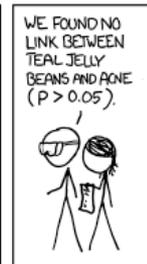
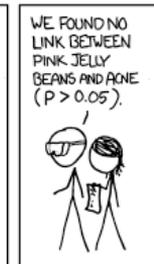
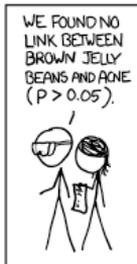
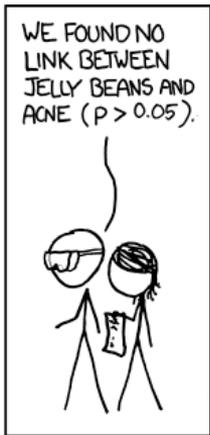
(c) 2016: USB drive (\$4)

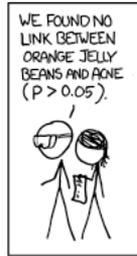
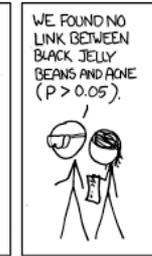
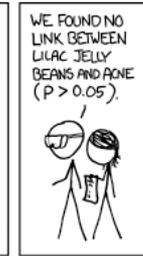
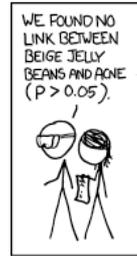
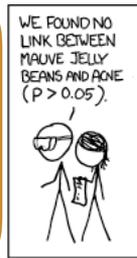
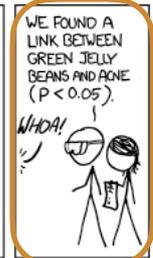
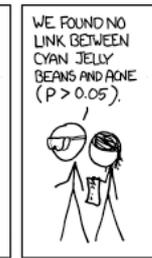
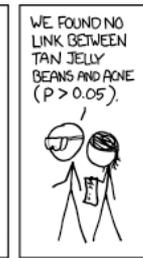
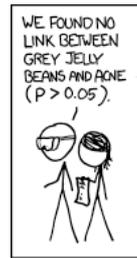
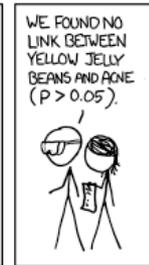
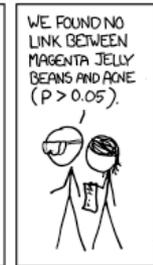
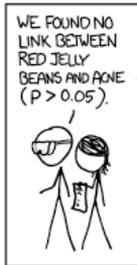
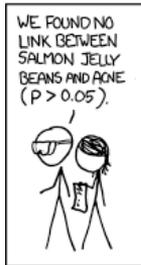
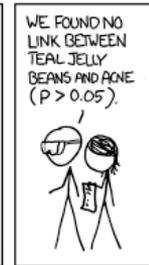
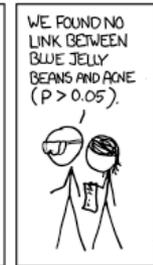
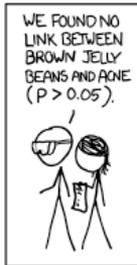
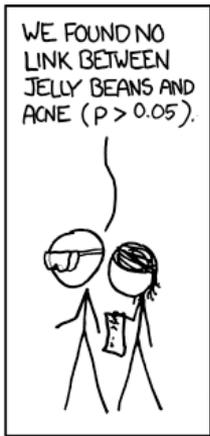


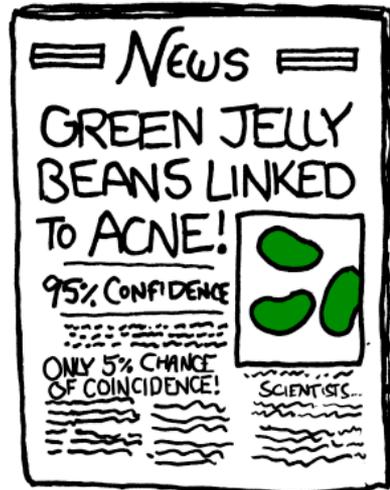
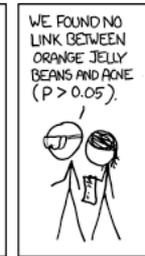
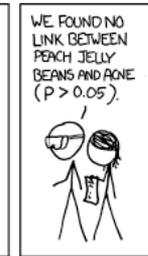
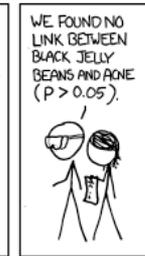
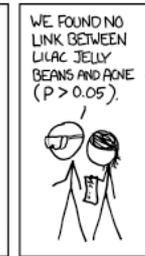
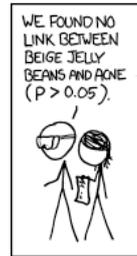
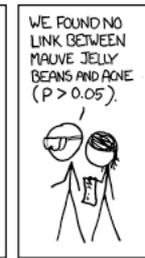
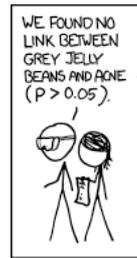
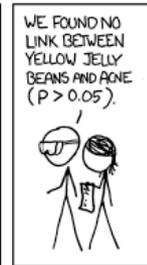
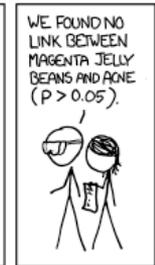
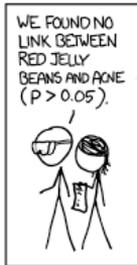
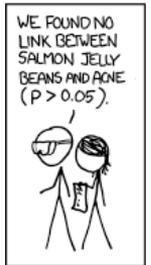
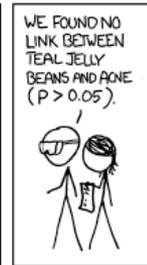
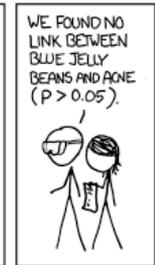
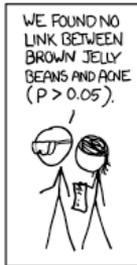
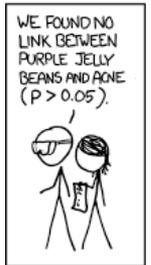
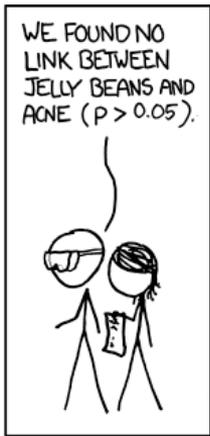
**Figure 1.** A user inspects several graphs and wrongly flags (c) as an insight because it looks different than (a) and (b). All were generated from the same uniform distribution and are the “same”. By viewing lots of visualizations, the chances increase of seeing an apparent insight that is actually the product of random noise.

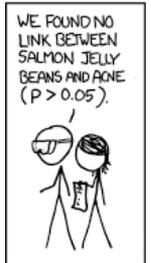
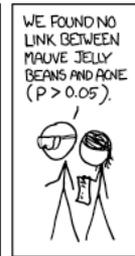
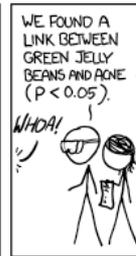
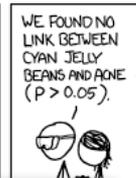
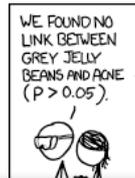
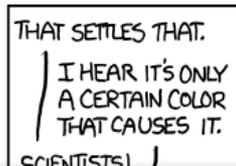
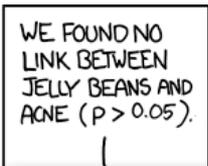








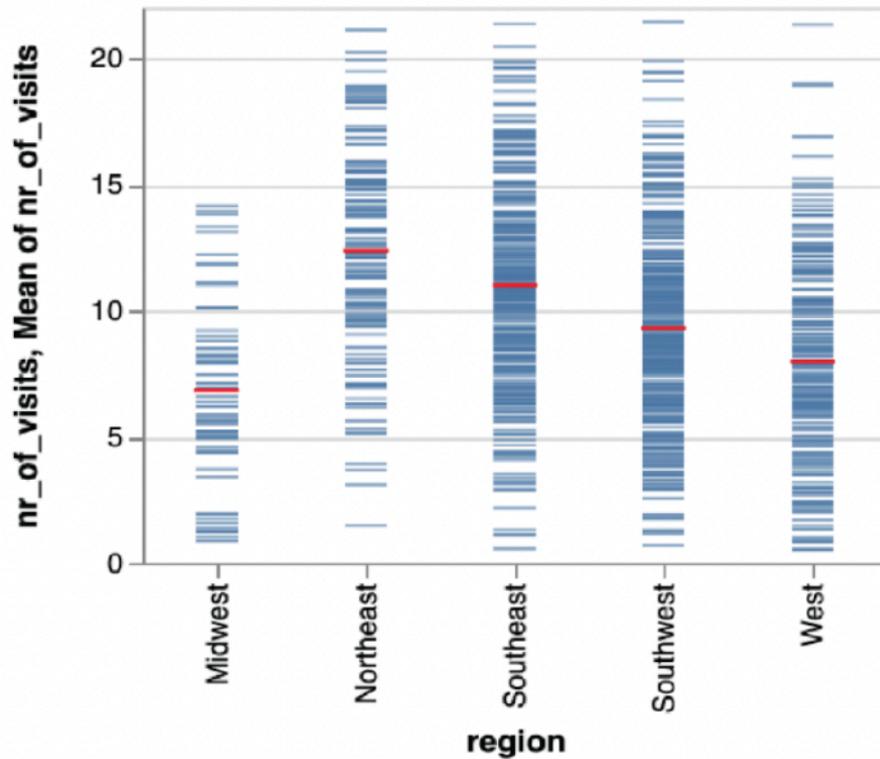
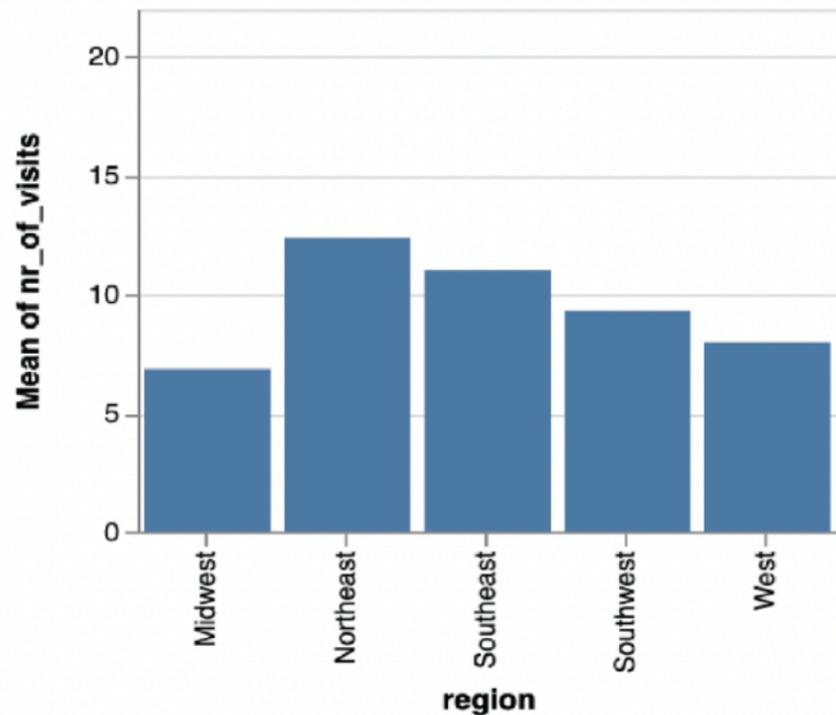




# Multiple Comparisons Problem



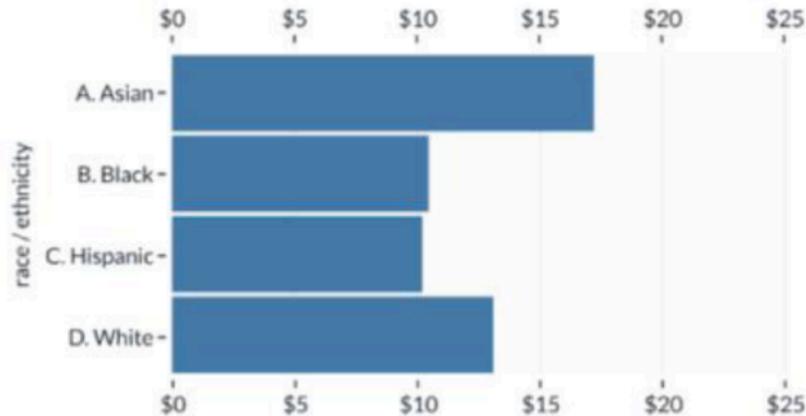
# Aggregated vs. Disaggregated Views [Nguyen et al. '20]



# Hiding Variability Encourages Stereotypes? [Xiong et al. '22]

## Waiter / Waitress Wages

Average hourly wage for waiters and waitresses, including tips.

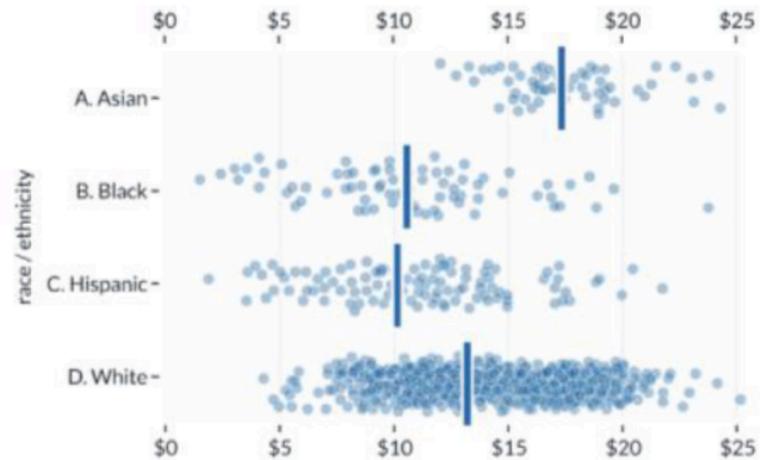


This chart compares earnings across 4 different race / ethnicity groups. Each bar represents at least  $n \geq 50$  people who identify as the corresponding race / ethnicity. The bar's width represents the average of each person's average hourly earnings over a 24 week period.

## Waiter / Waitress Wages

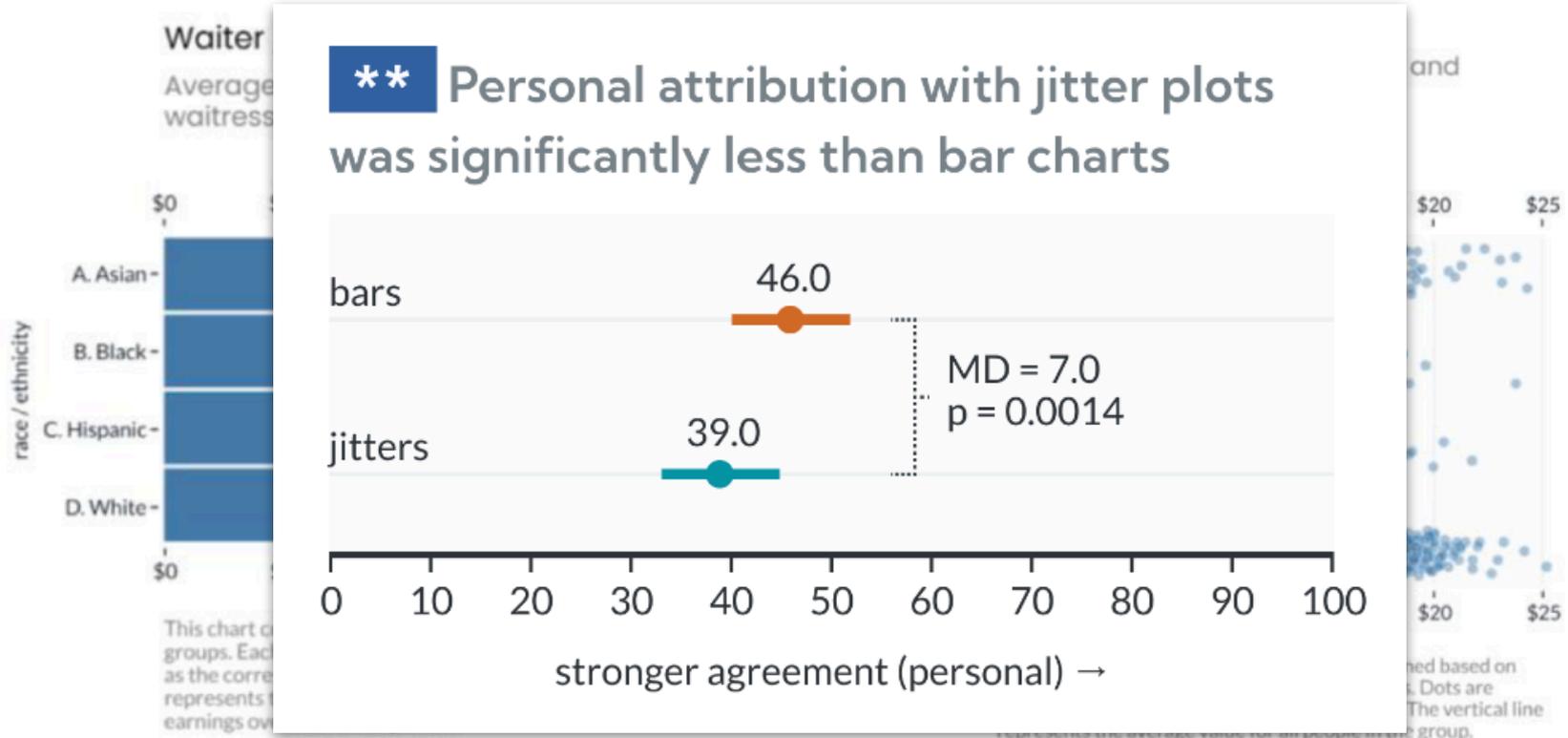
Average hourly wage for waiters and waitresses, including tips.

• = 1 person | = group average

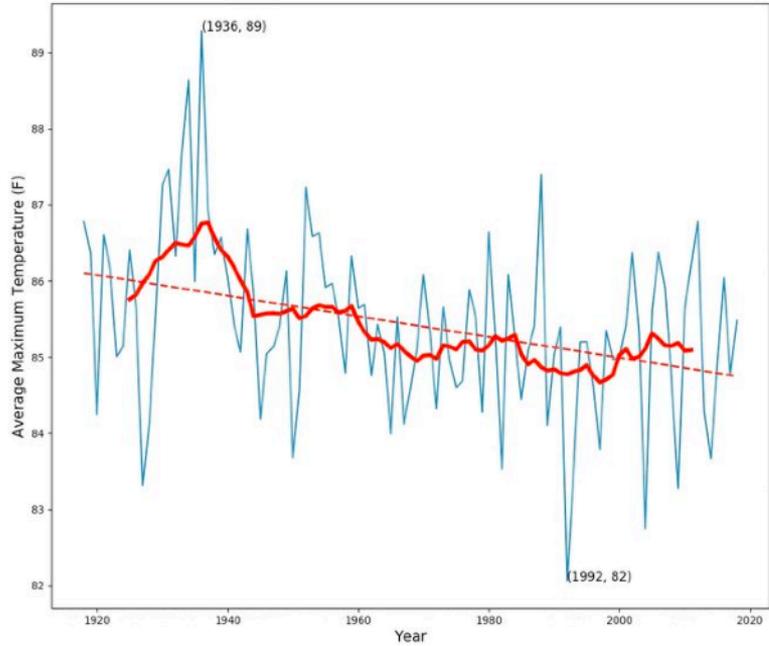


Each dot represents 1 person. The dot is positioned based on the person's average hourly wage over 24 weeks. Dots are grouped into rows based on the person's group. The vertical line represents the average value for all people in the group.

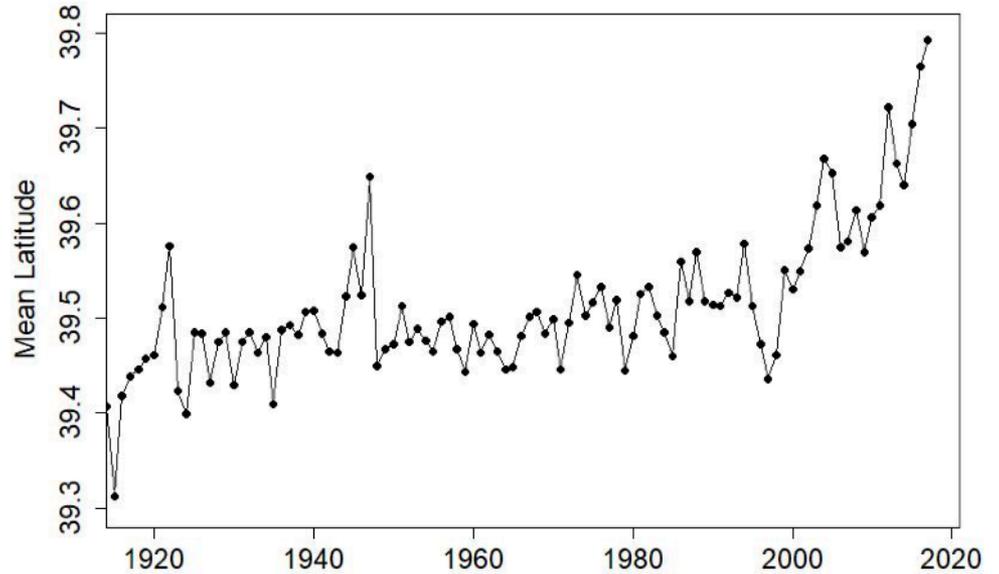
# Hiding Variability Encourages Stereotypes? [Xiong et al. '22]



# Example: Is the U.S. cooling?

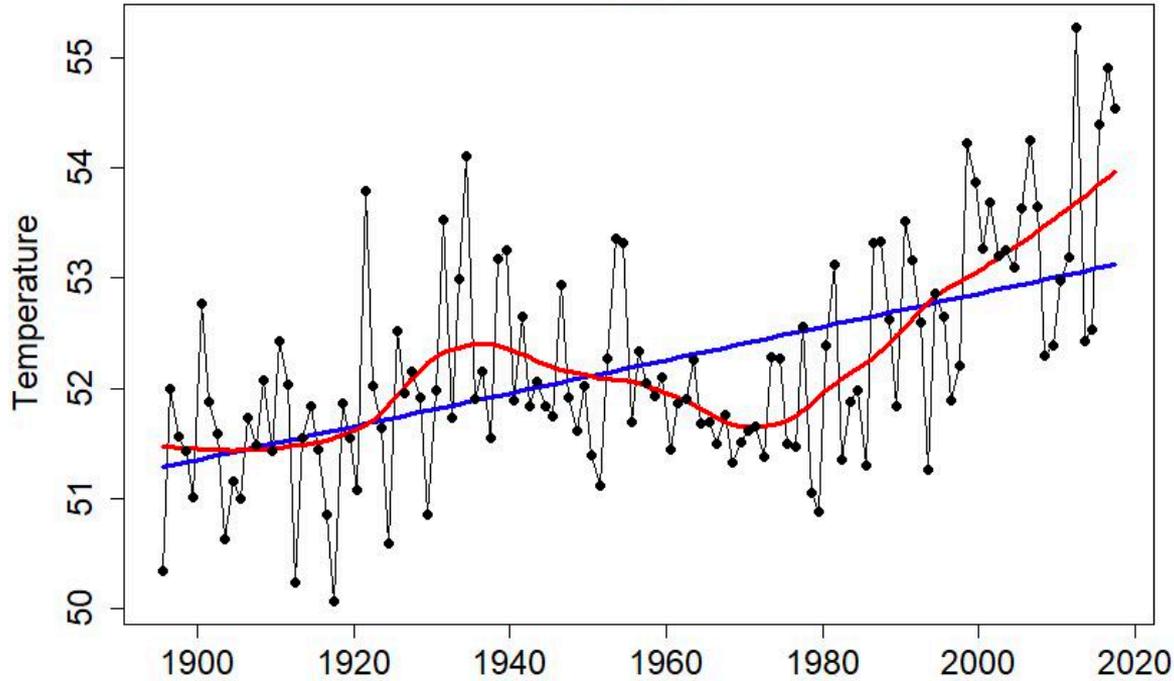


Starts at 1918. Summer temps only.  
Raw average over weather stations.



But here is the mean latitude of  
US weather stations, per year...

# Example: Is the U.S. cooling? (No.)

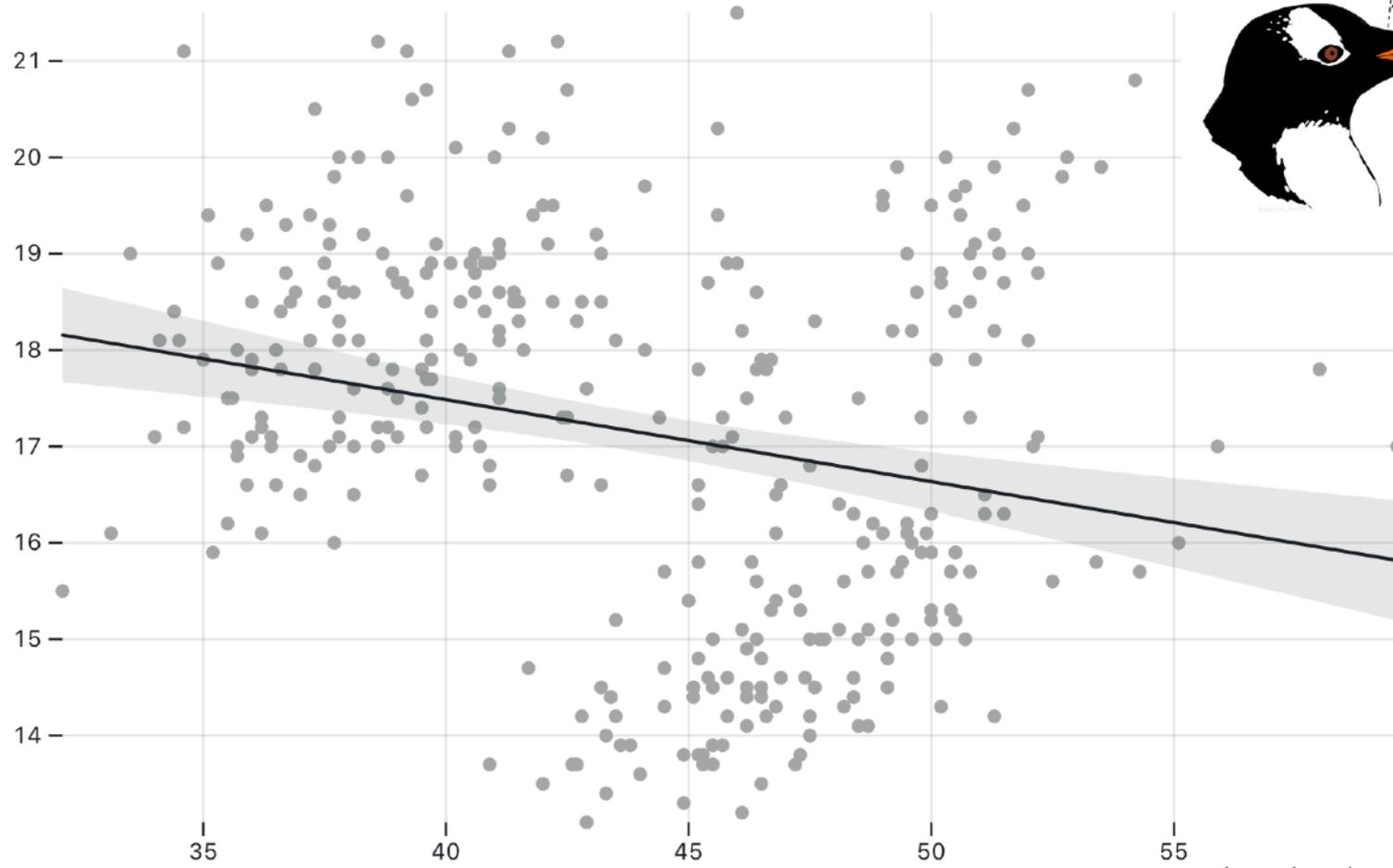


Include more historical data.

Include all four seasons.

Correct spatial averages  
to account for changes in  
weather station locations.

↑ culmen\_depth\_mm

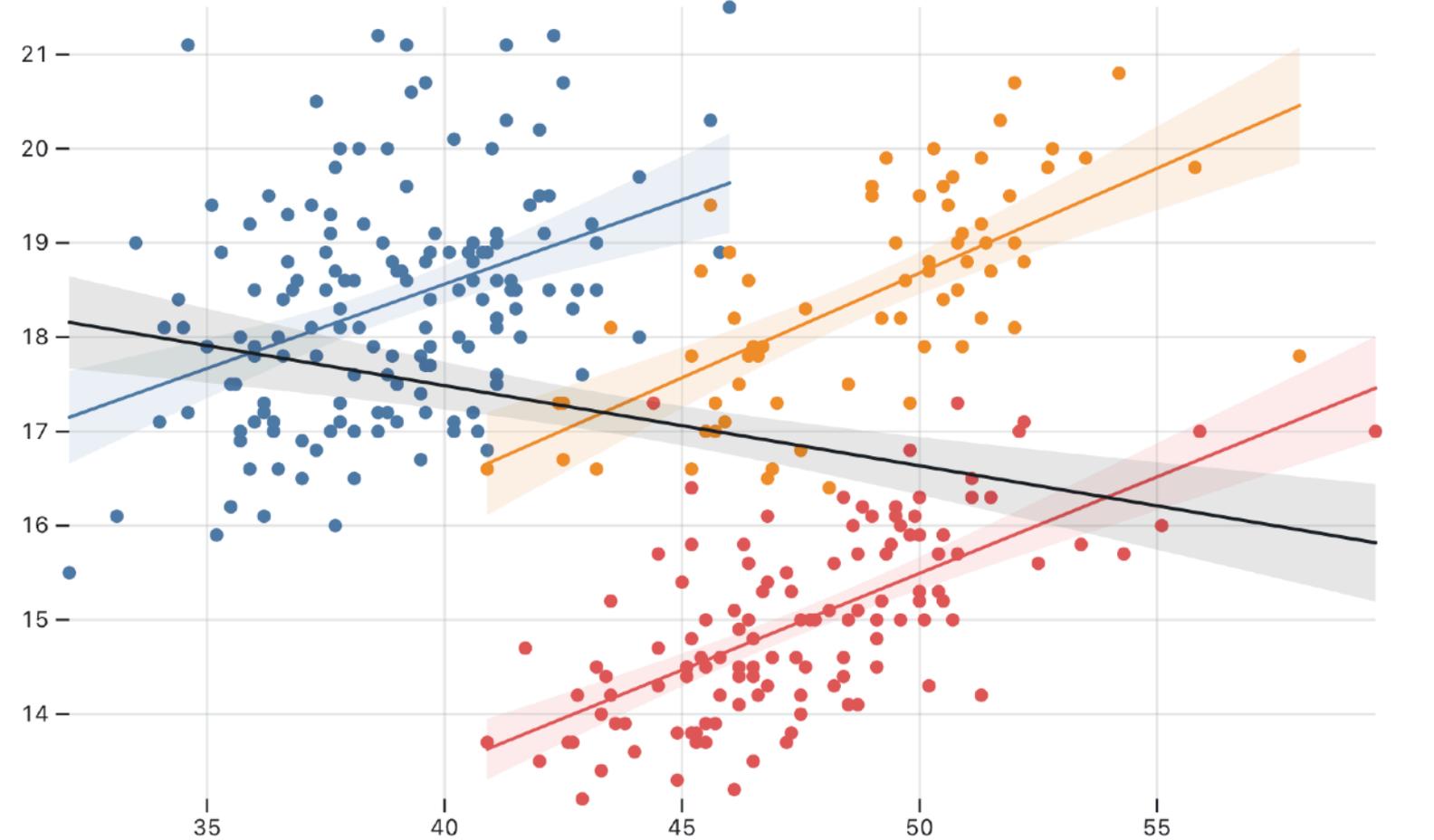


culmen\_length\_mm →

# Simpson's Paradox!

Adelie Chinstrap Gentoo

↑ culmen\_depth\_mm



culmen\_length\_mm →

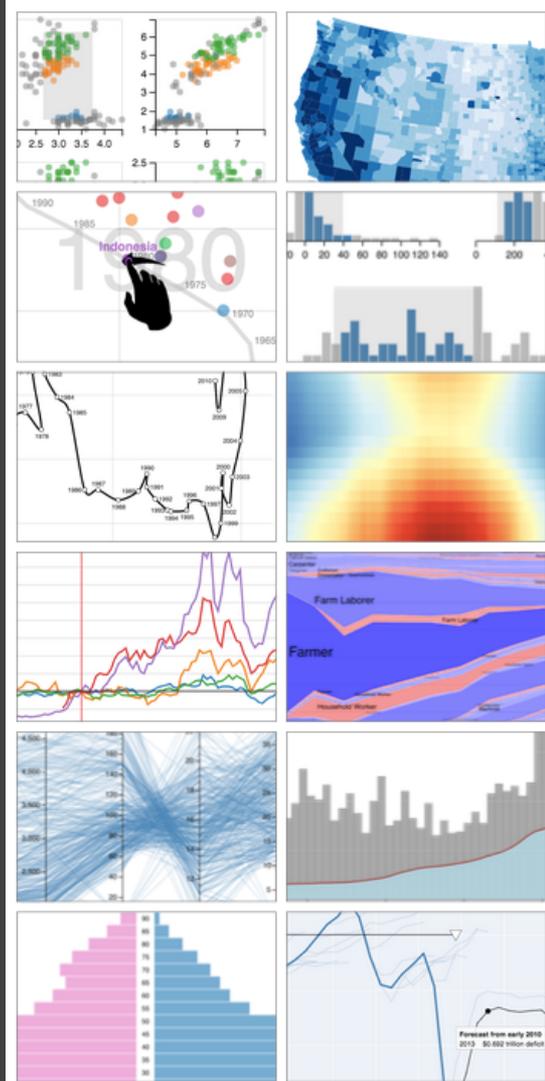
# Some Causes of Inferential Failure

Premature Suppression of Uncertainty

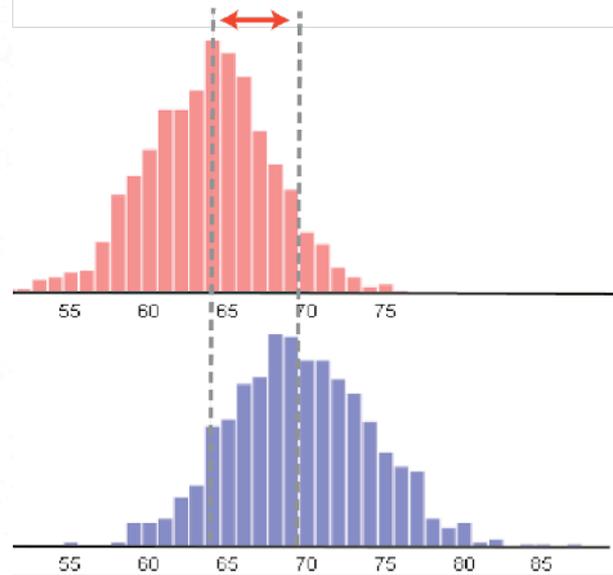
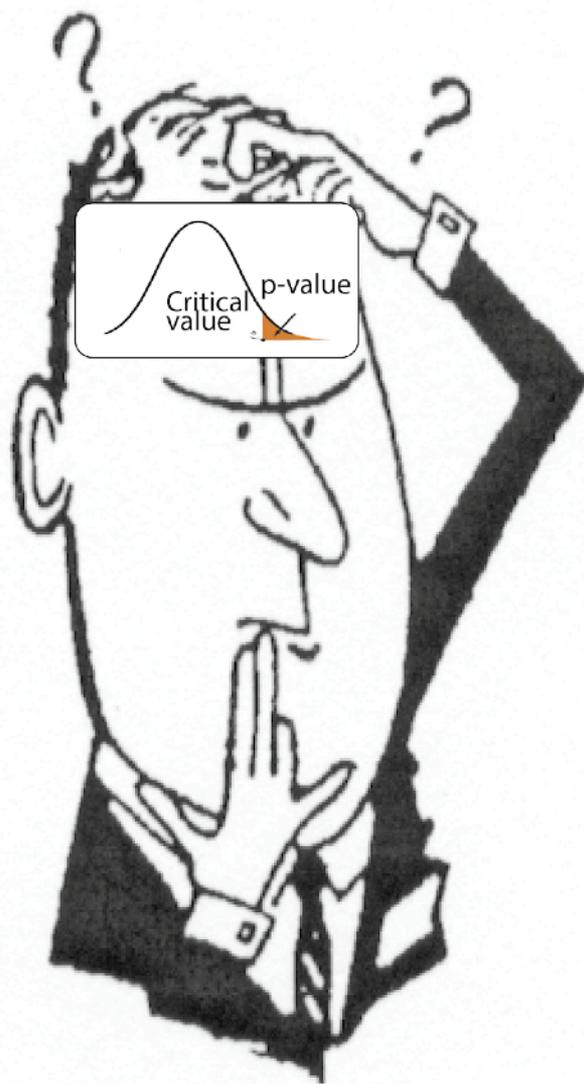
False Discovery due to Random Fluctuation

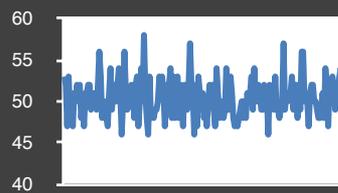
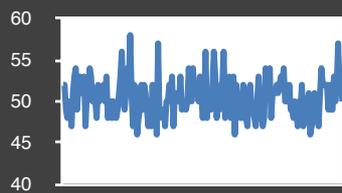
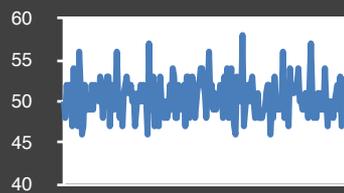
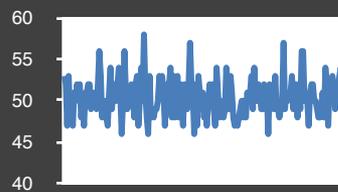
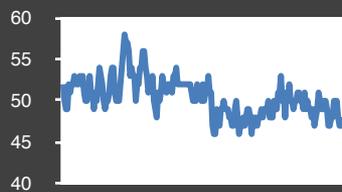
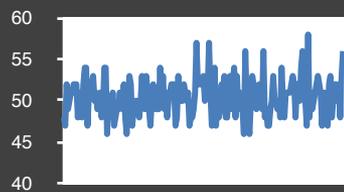
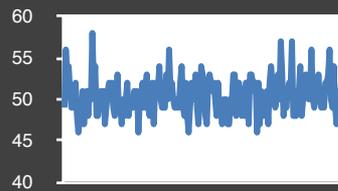
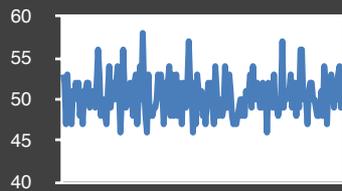
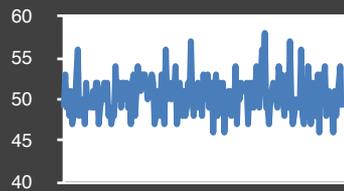
Incomplete or Biased Data

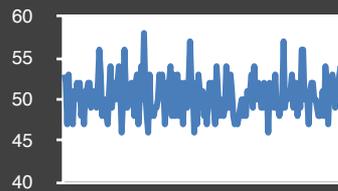
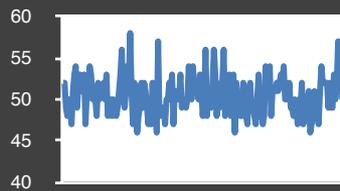
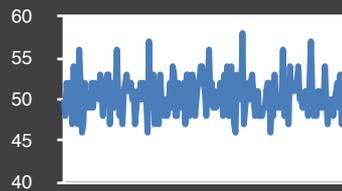
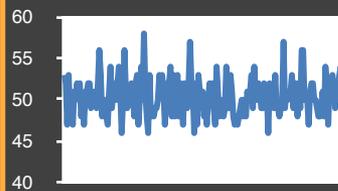
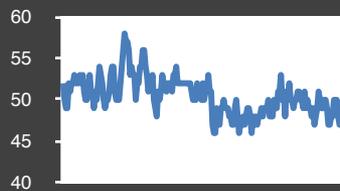
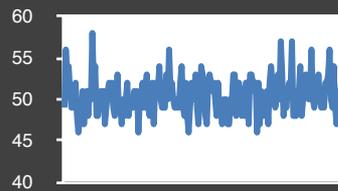
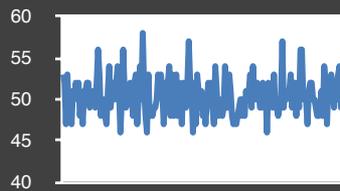
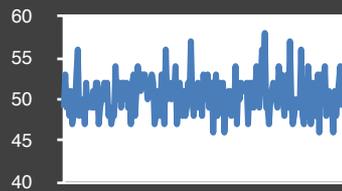
Confounding Variables



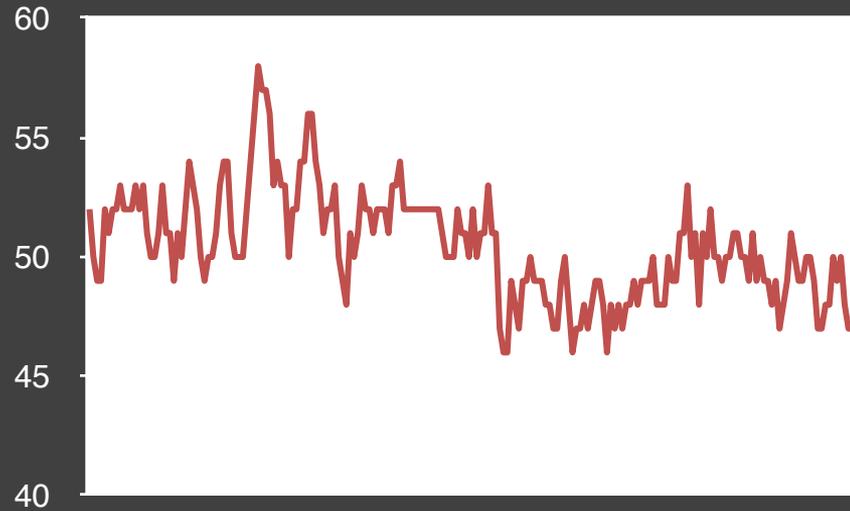
# Graphical Inference & Model Checks

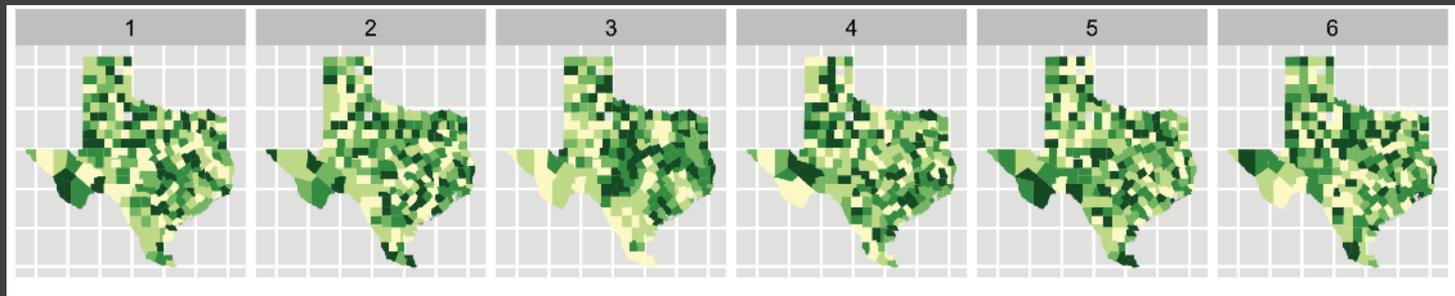






# Visual Lineups

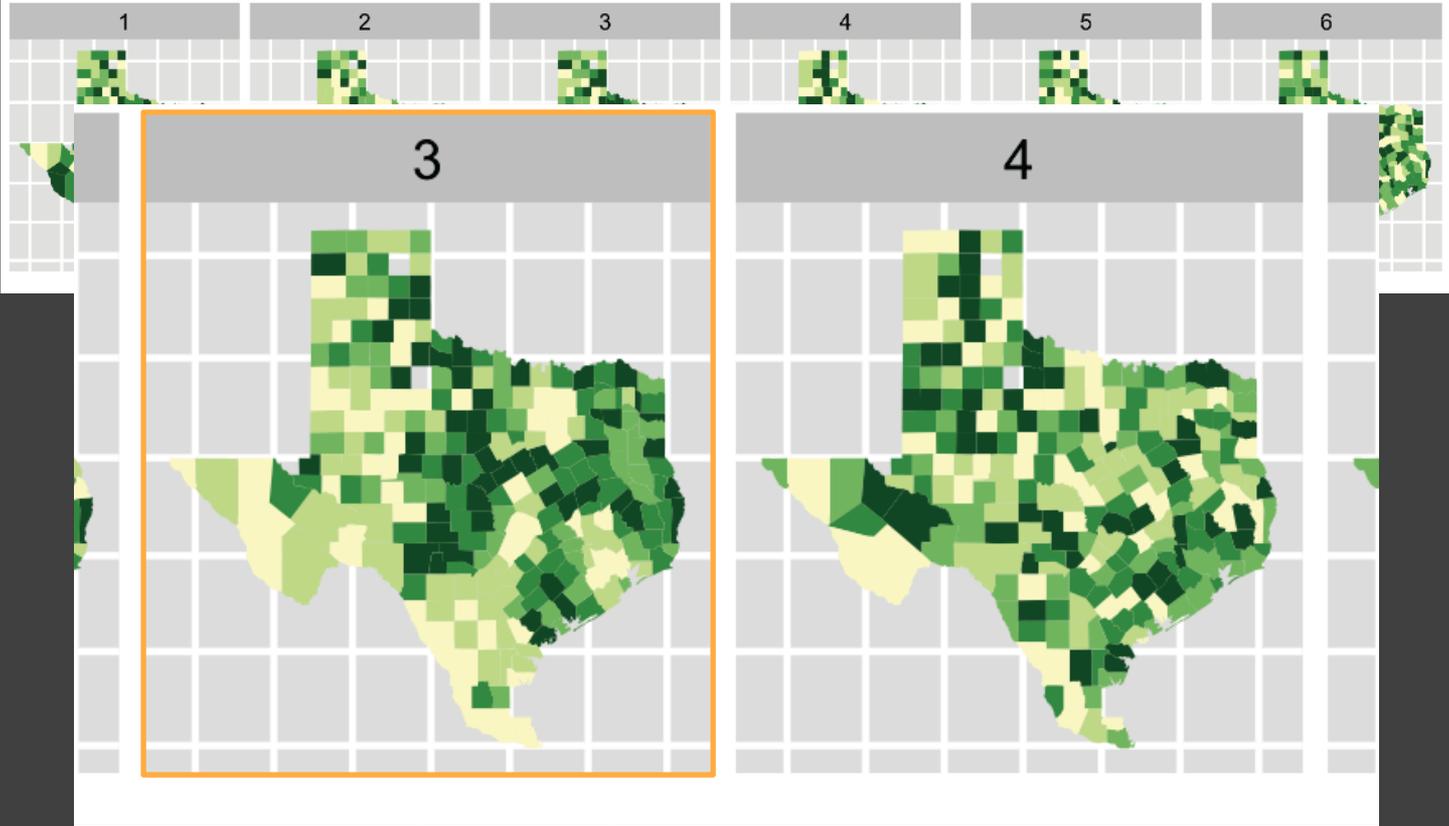




Choropleth maps of cancer deaths in Texas.

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.



# Graphical Inference

## Compare data to replicated data under a model

Can we articulate a possible data generating process?

If we model that, how does it compare to our data?

## Choose a model for comparison

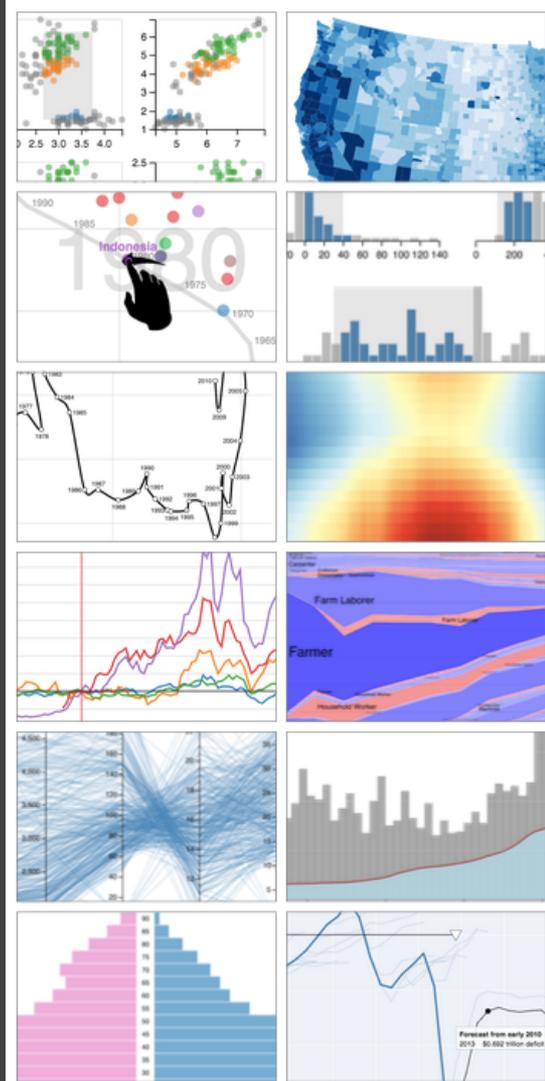
Permute (shuffle) relationship between variables, or

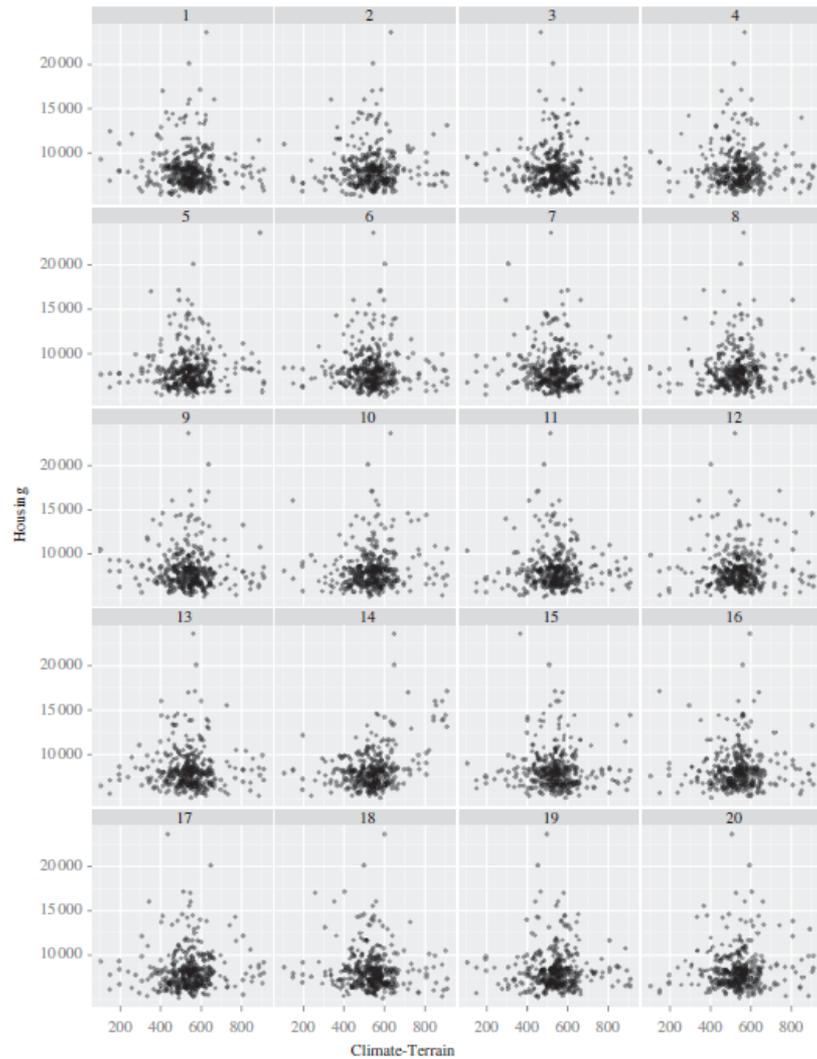
Choose a meaningful “null” model

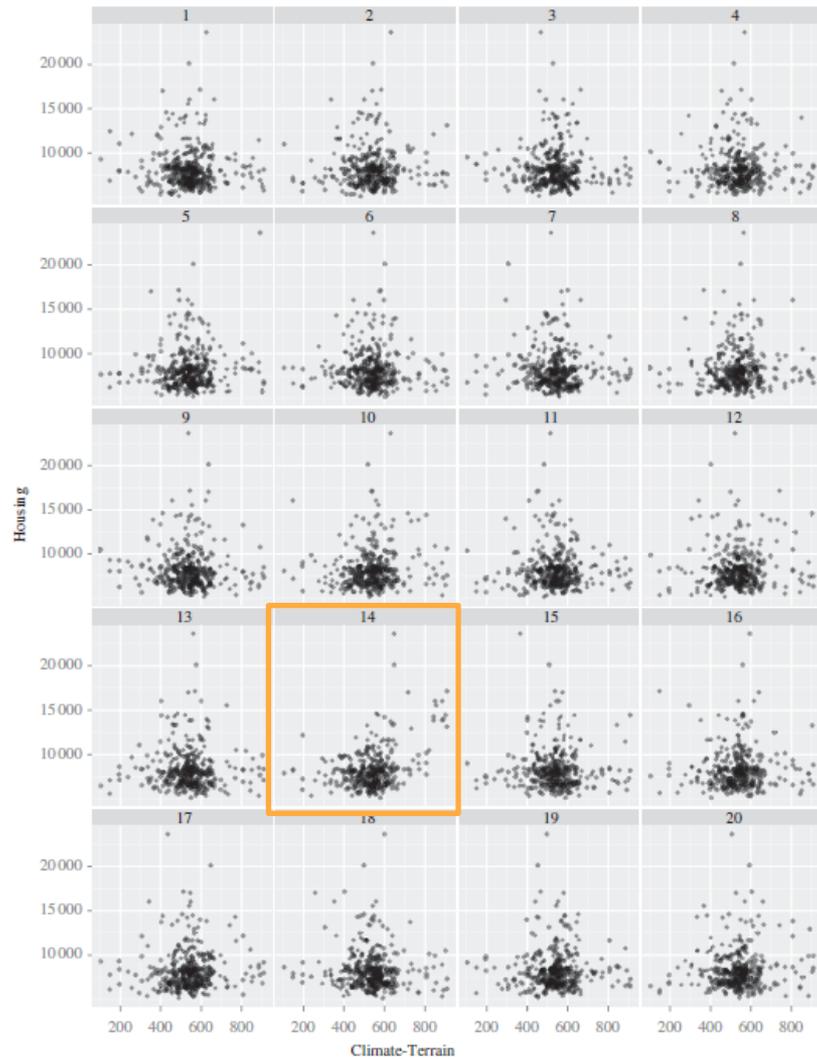
## Perform visual comparison

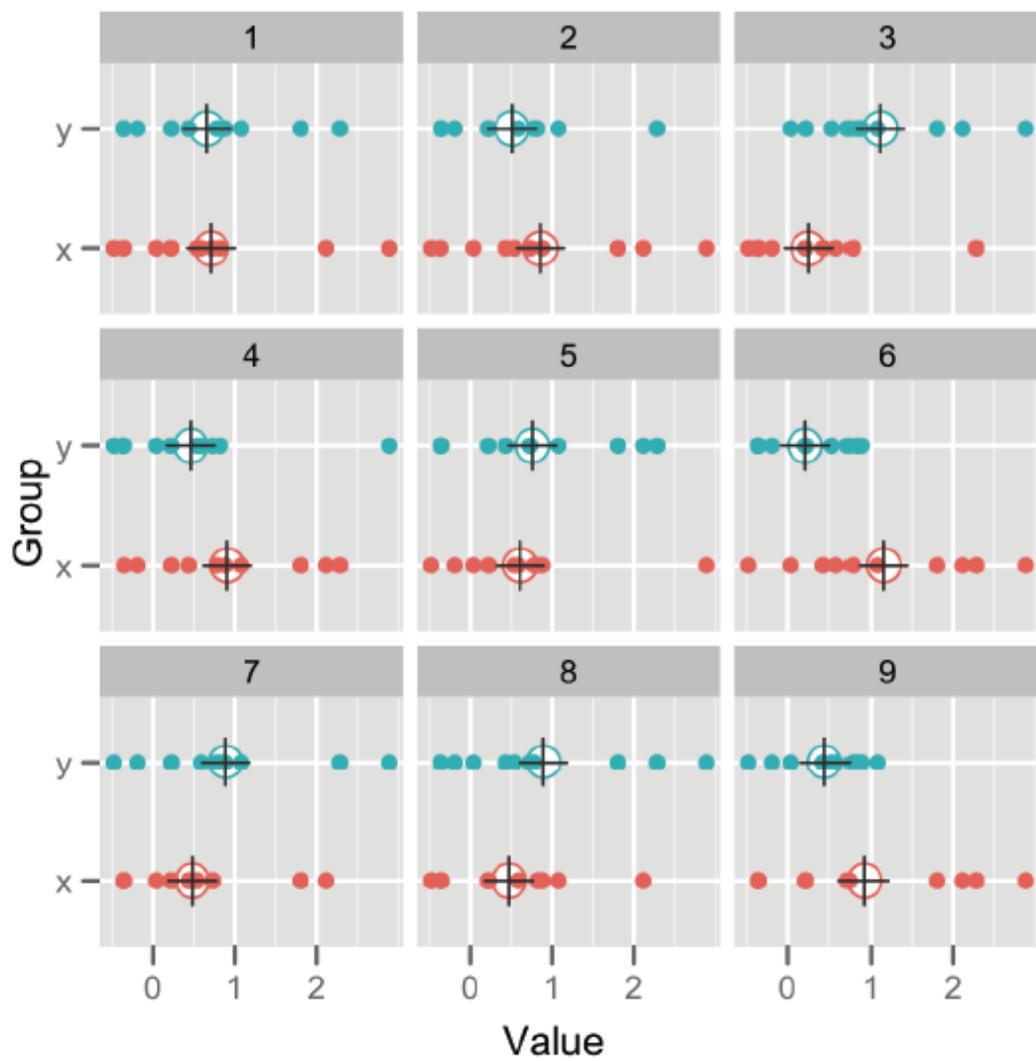
In the “lineup” protocol, we compare the real data against a number of generated variants.

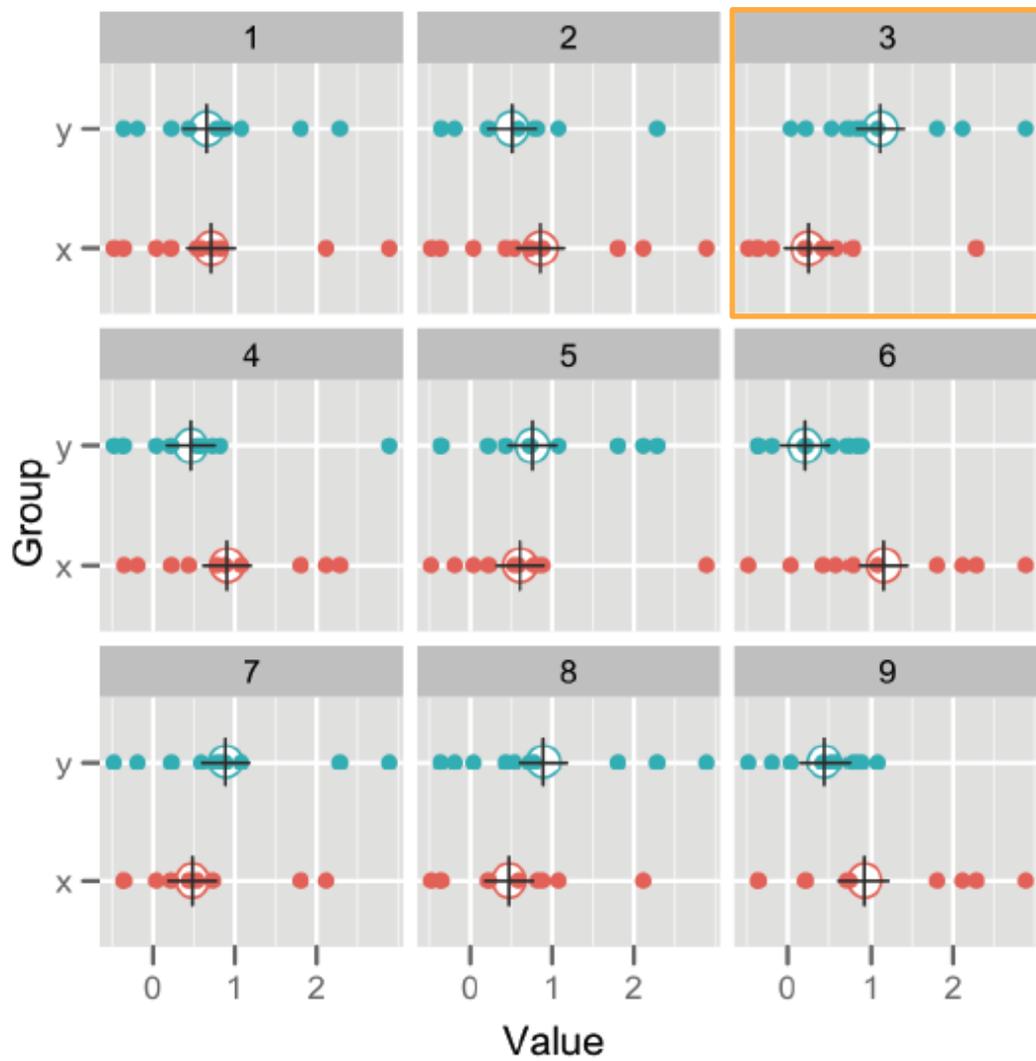
*Can we spot the difference?*

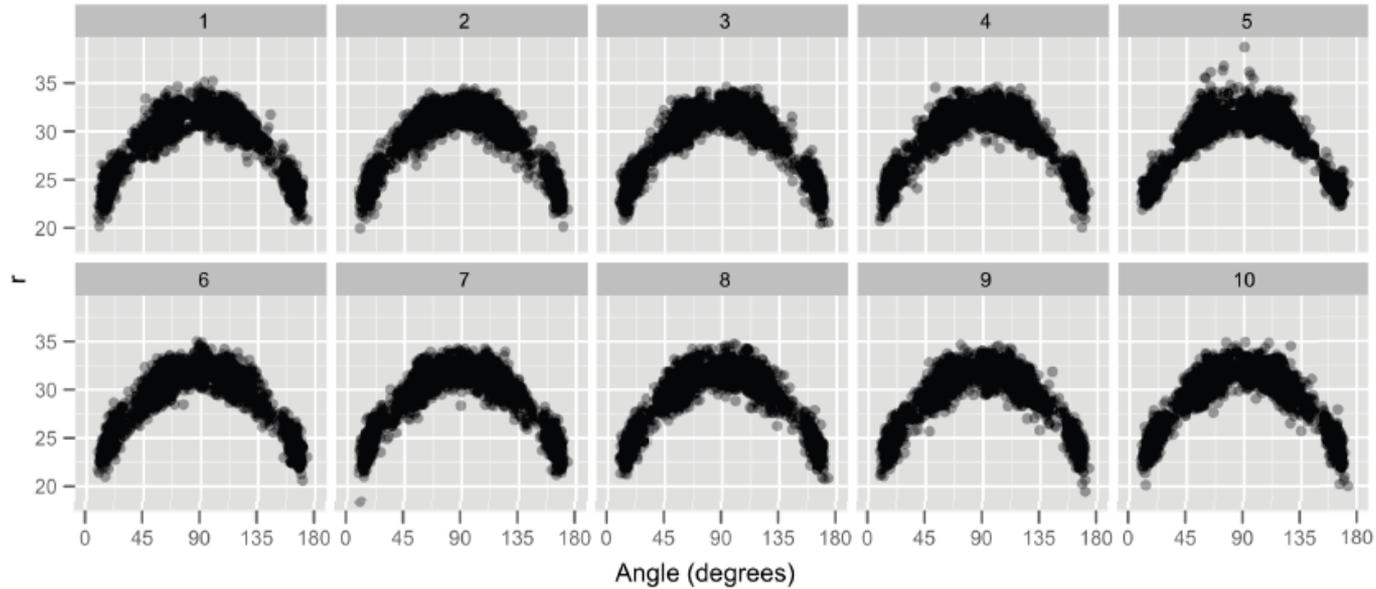






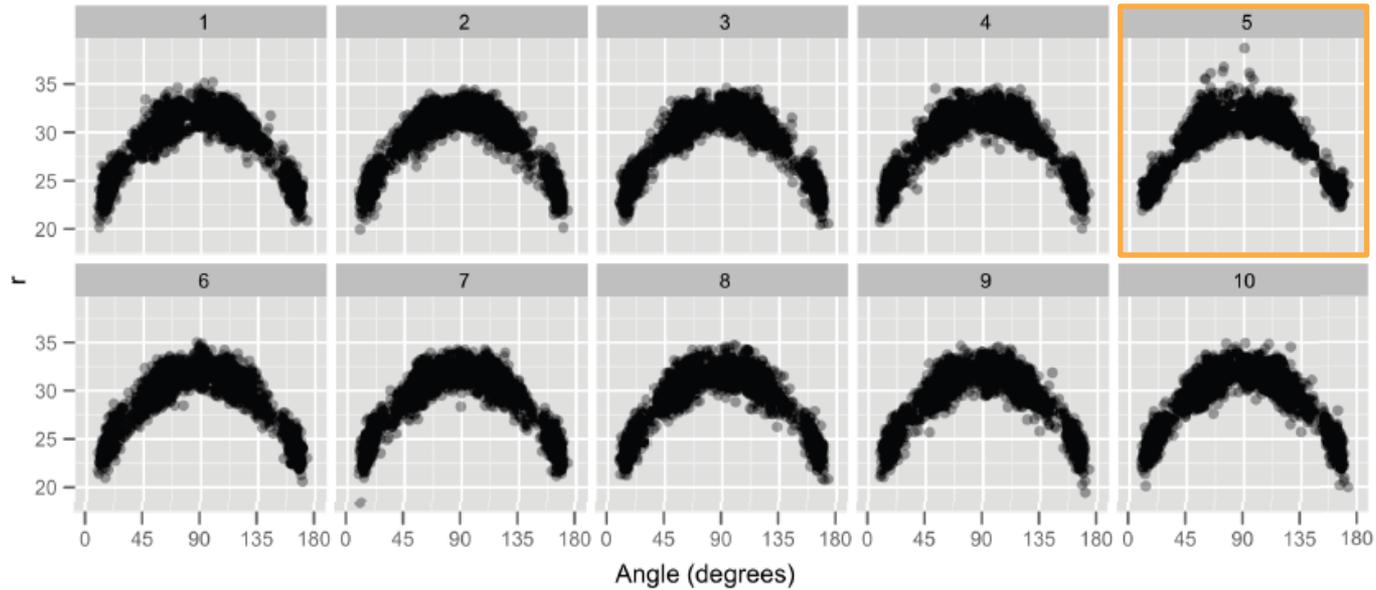






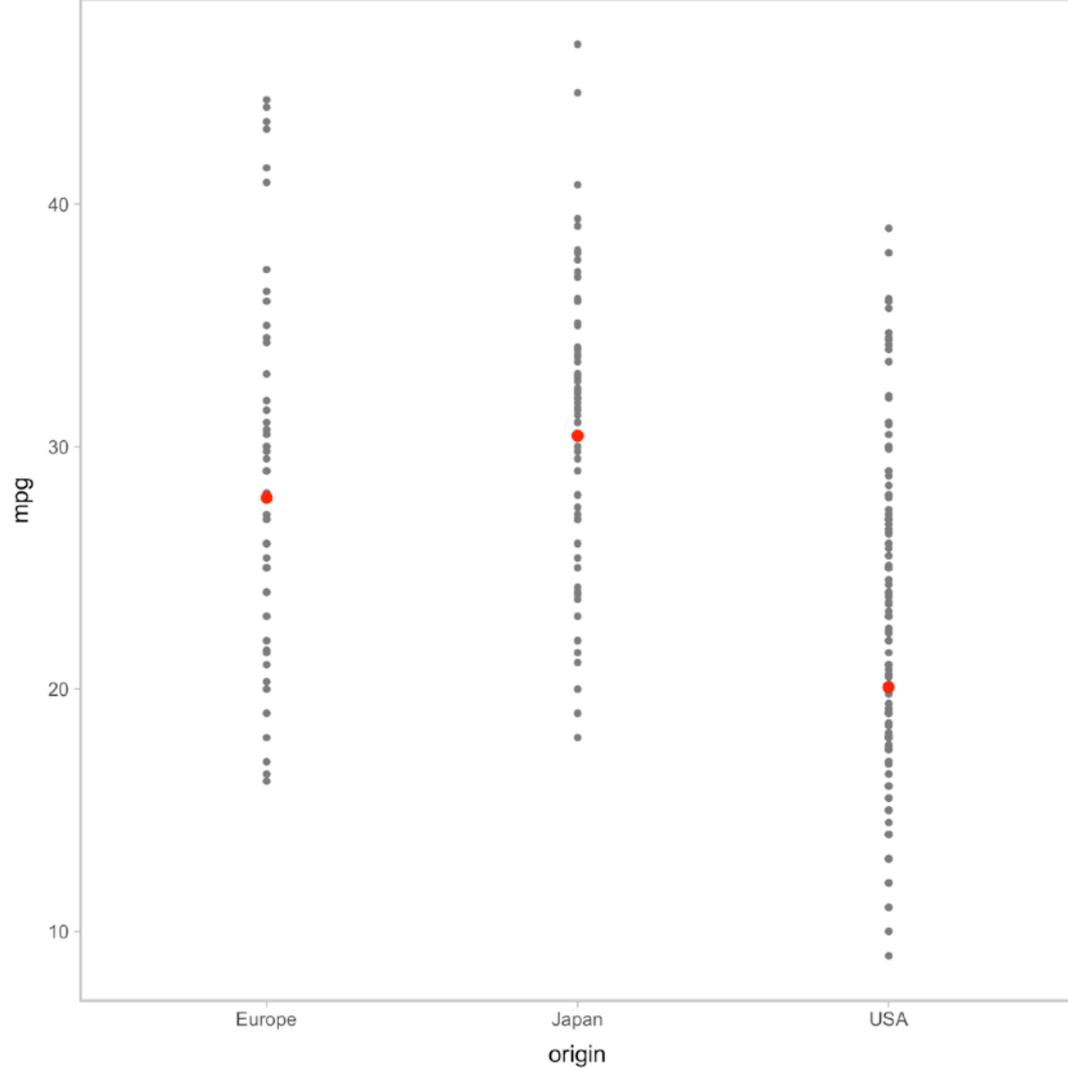
## Distance vs. angle for 3 point shots by the LA Lakers.

One plot is the real data. The others are generated according to a null hypothesis of quadratic relationship.



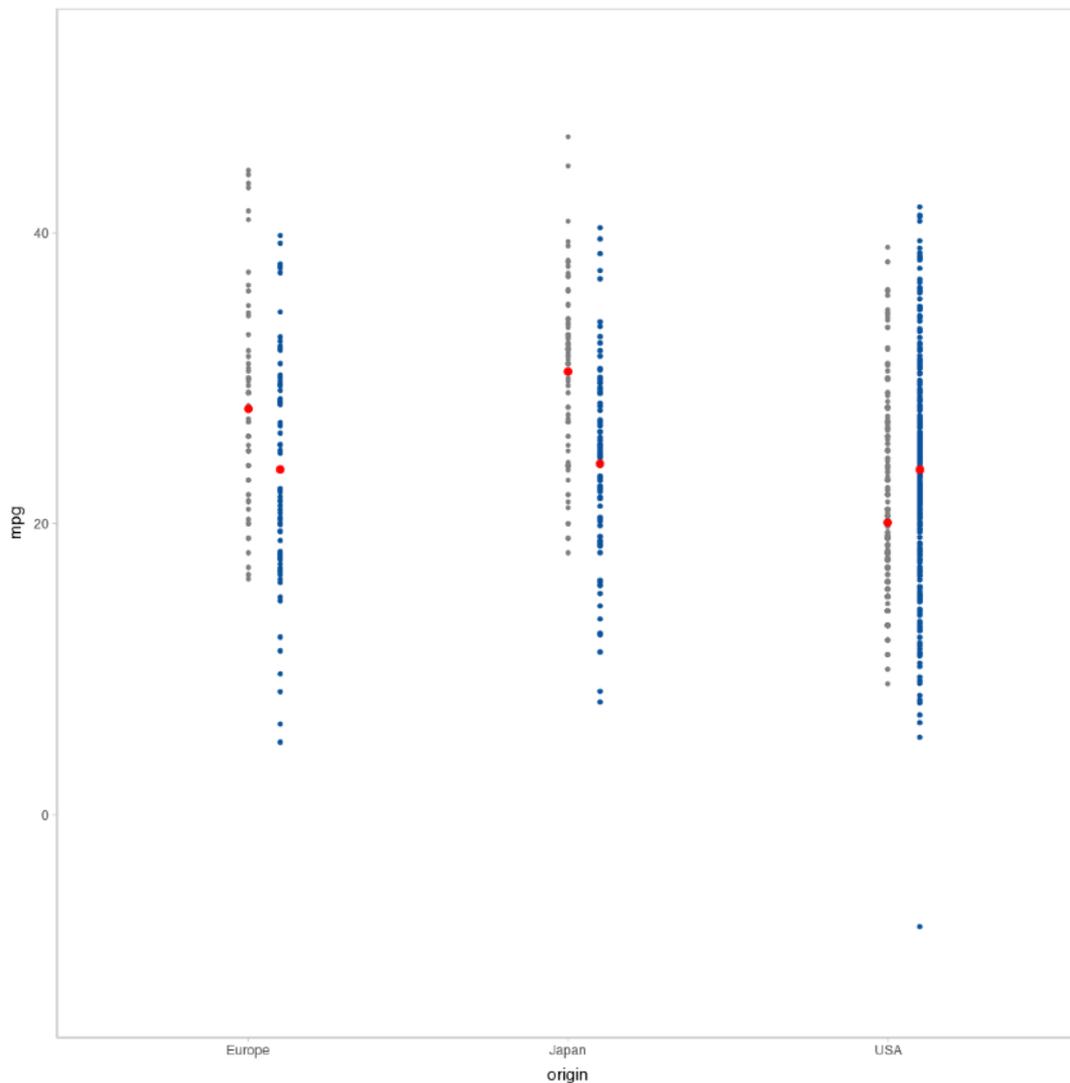
## Distance vs. angle for 3 point shots by the LA Lakers.

One plot is the real data. The others are generated according to a null hypothesis of quadratic relationship.



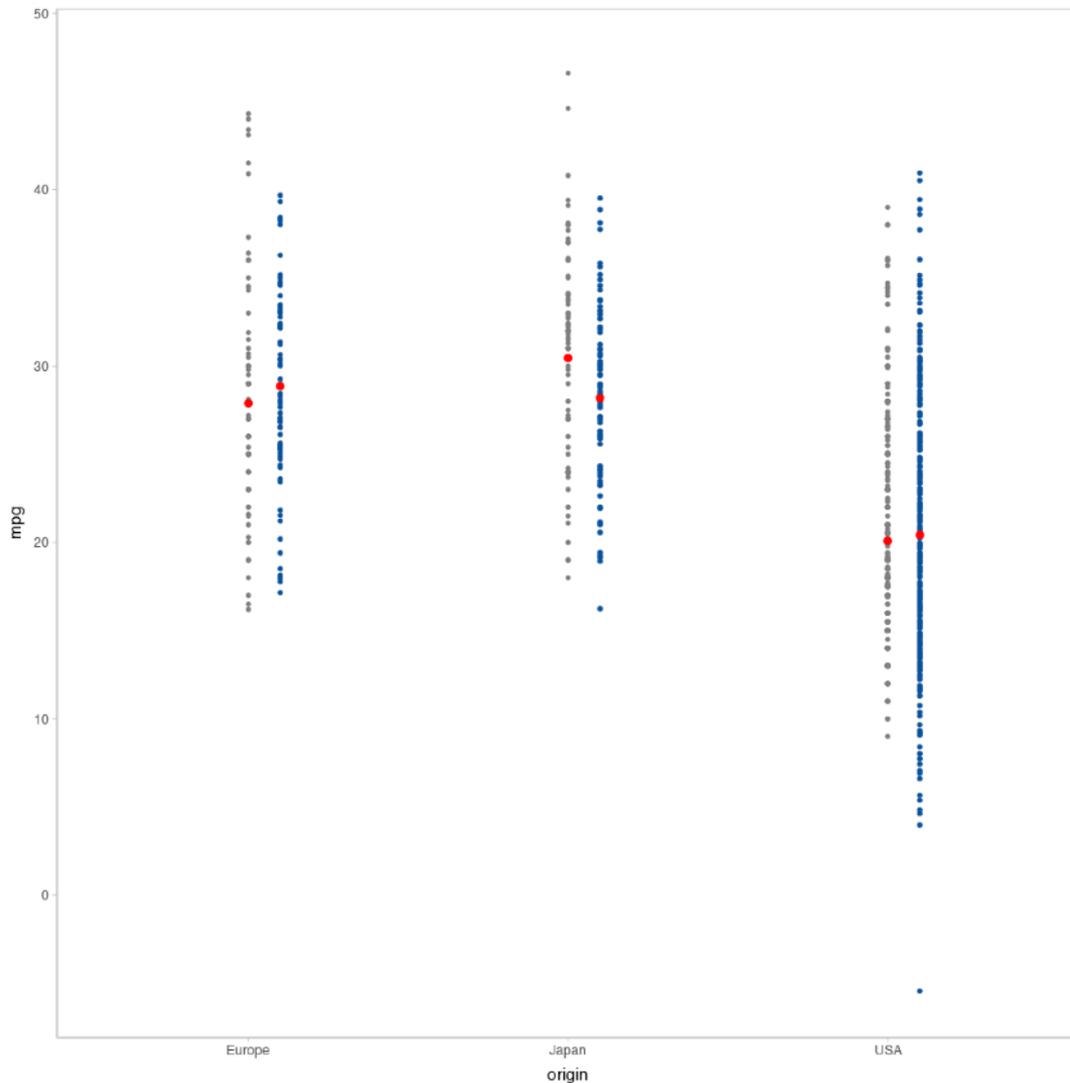
**Plot:**  
mpg by origin

*What might our  
implicit model be?*



**Model:**  
 $\text{mpg} \sim 1$

*Blue points are predictions from a **null model** based on the mean and stdev of the miles per gallon.*



**Model:**  
 $\text{mpg} \sim \text{cyl}$

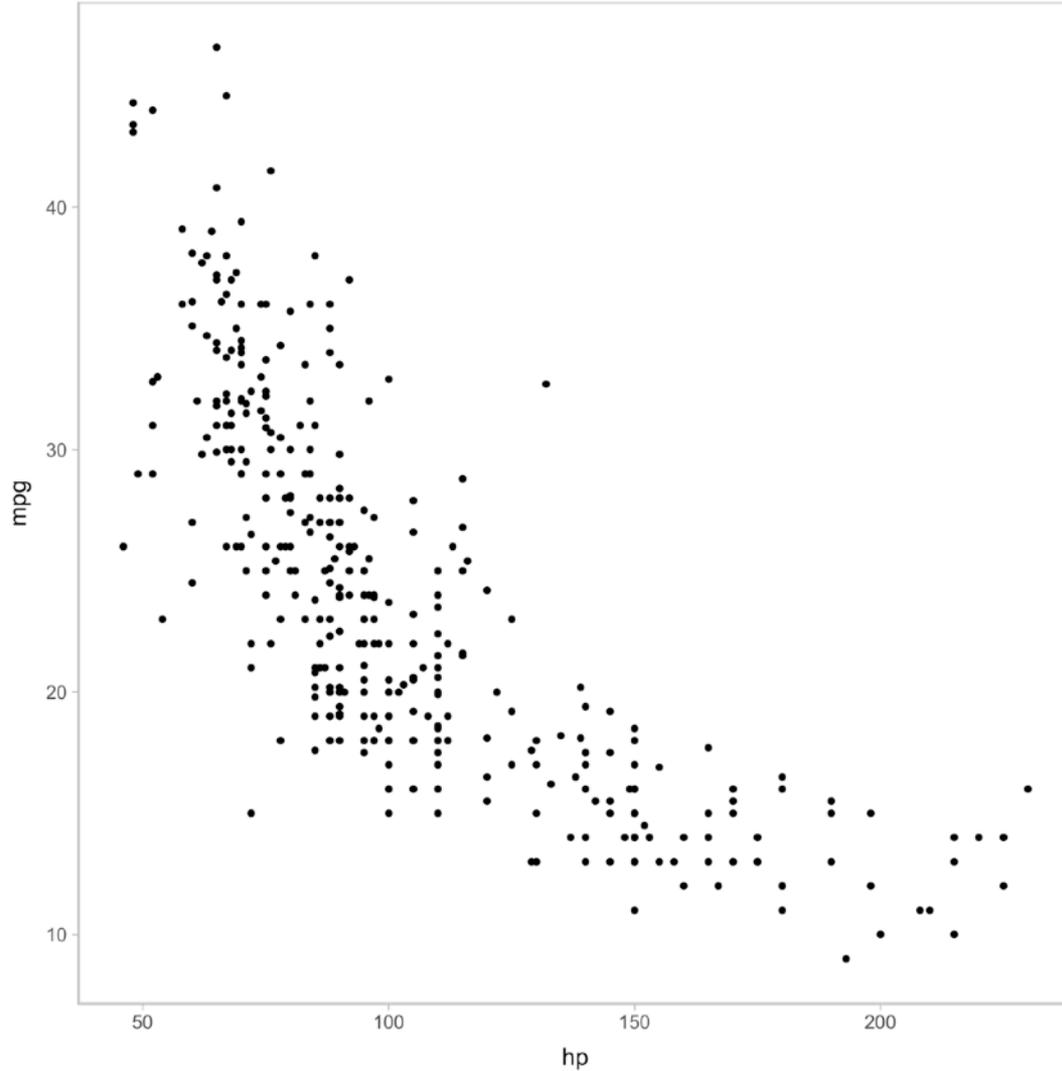
*Blue points are predictions from a model with cylinder count as a predictor.*

more cylinders

→ more fuel consumption

→ lower mileage

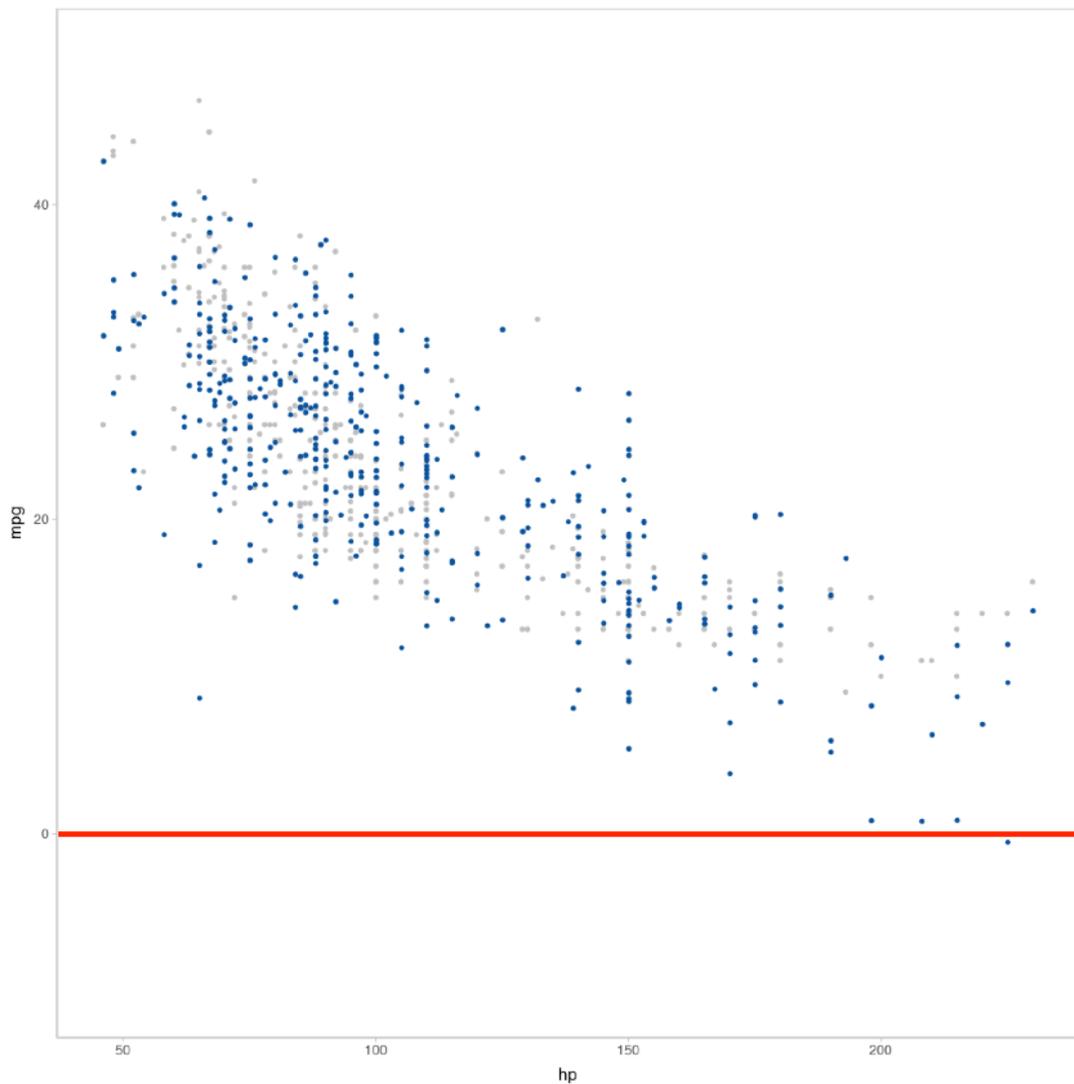
Might this explain the differences across regions?



**Plot:**

mpg by hp

*What might our  
implicit model be?*



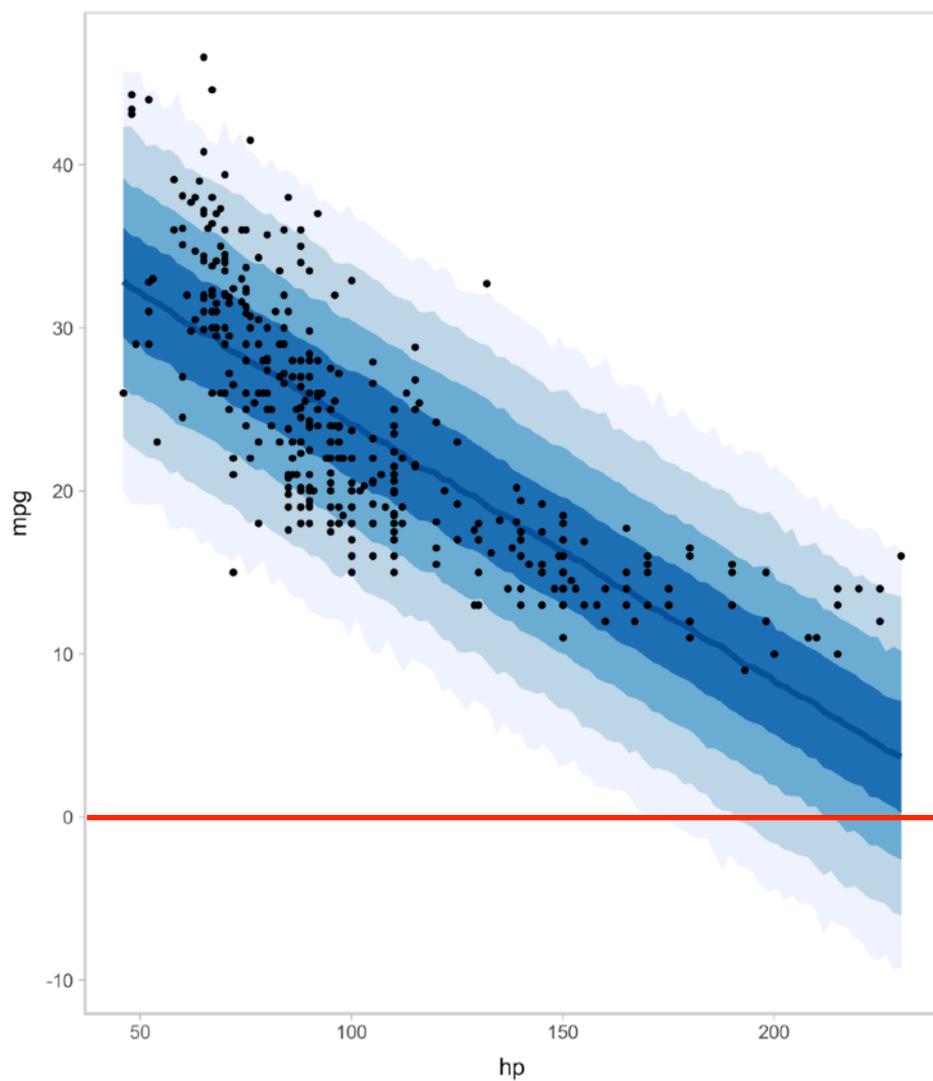
**Model:**

$$\text{mpg} \sim \text{hp}$$

*Linear model, similar to a standard regression.*

*Blue points are model predictions.*

**Negative mileage?!**

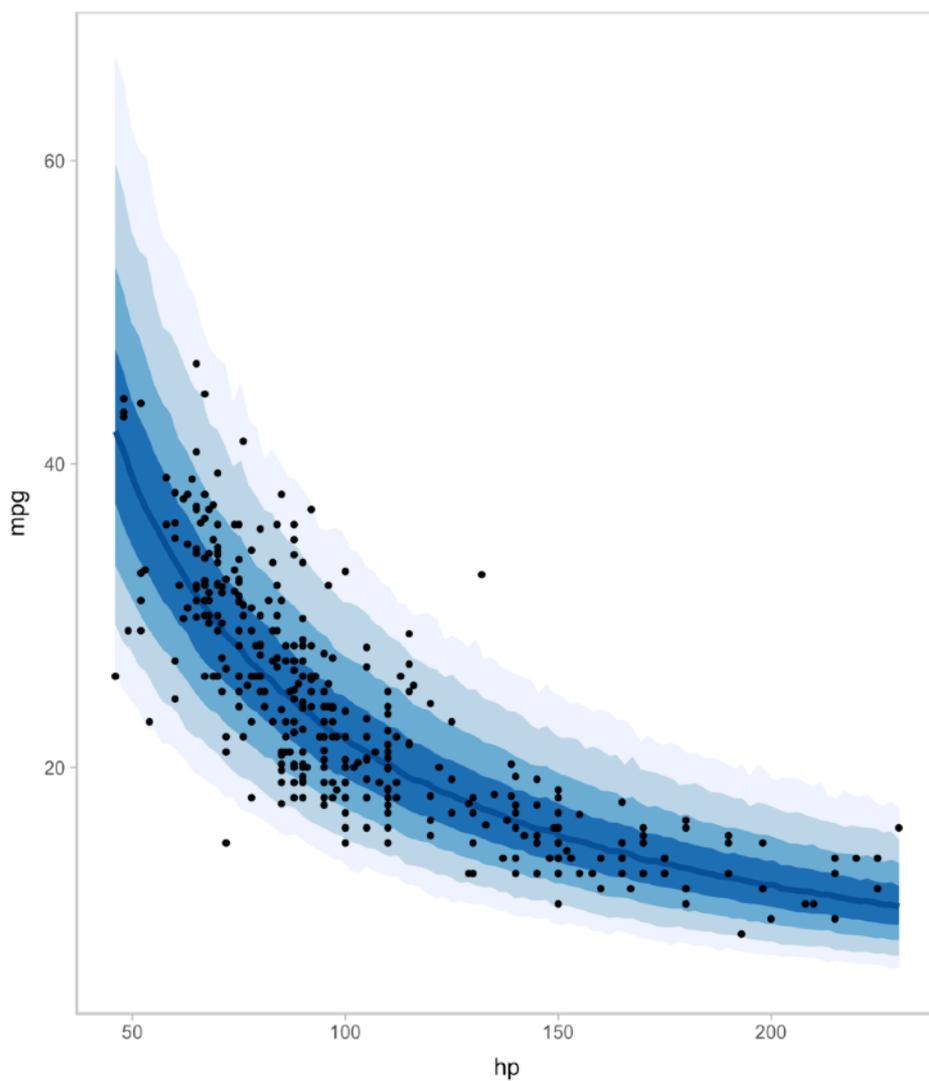


**Model:**  
 $\text{mpg} \sim \text{hp}$

level  
0.99  
0.95  
0.8  
0.5

*Linear model, similar to a standard regression.  
Bands show CI levels.*

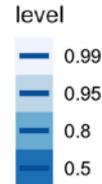
**Negative mileage?!**



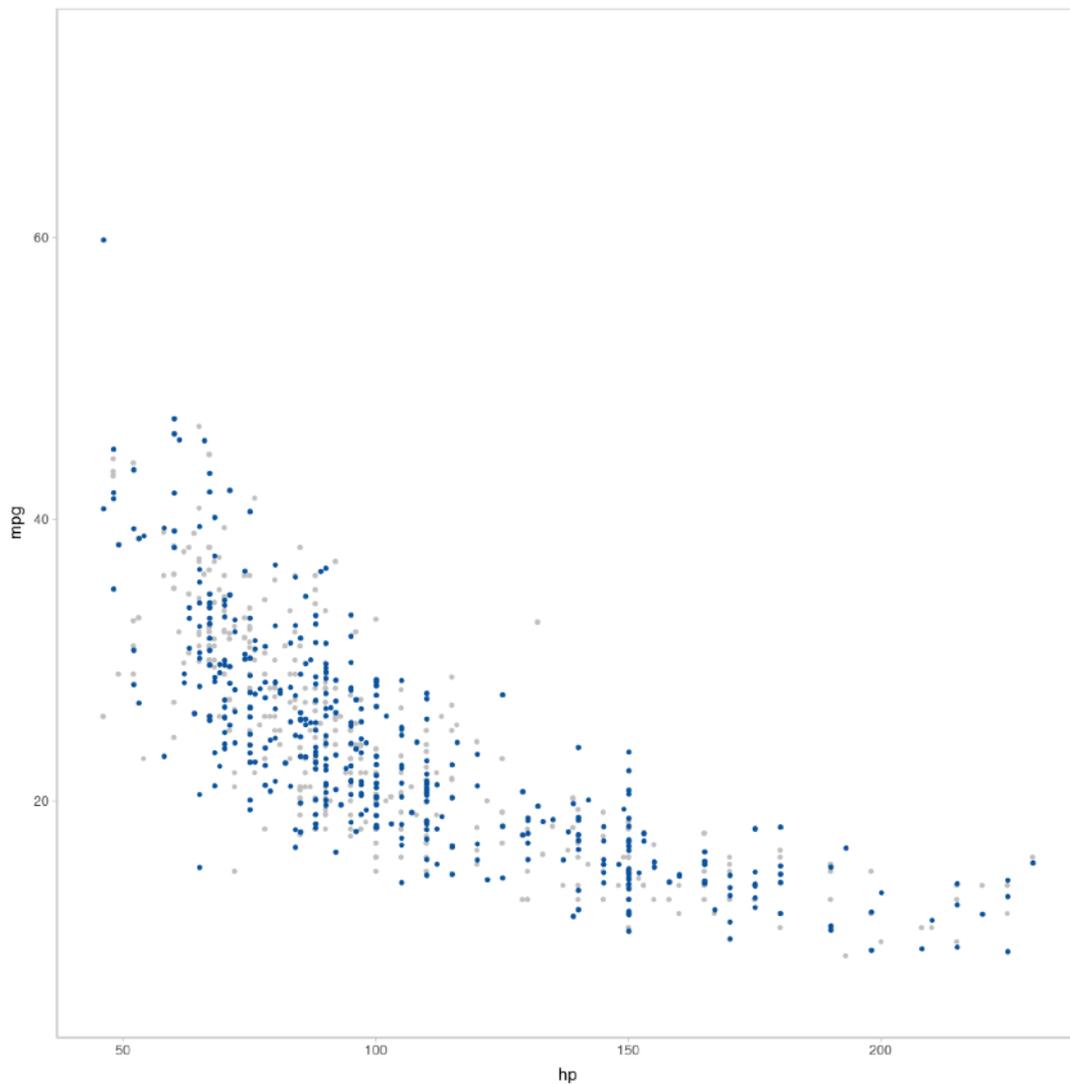
## Model:

$$\text{mpg} \sim \log(\text{hp})$$

family = lognormal



*A log-normal model better fits the data and does not "hallucinate" negative values.*



## Model:

$$\text{mpg} \sim \log(\text{hp})$$

family = lognormal

*A log-normal model better fits the data and does not "hallucinate" negative values.*

# Tools for Model Checks

R provides the needed modeling and visualization tools

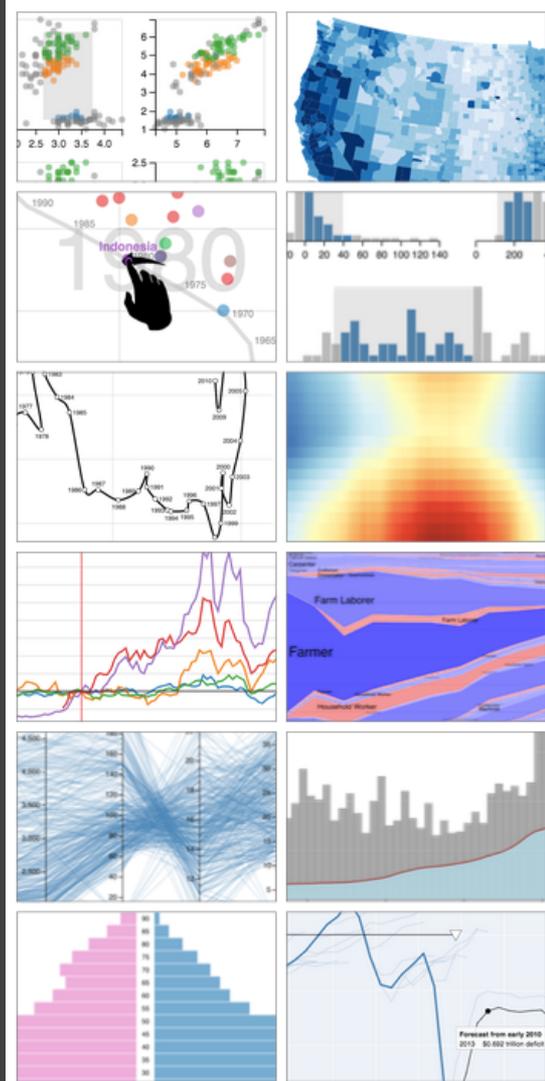
For example:

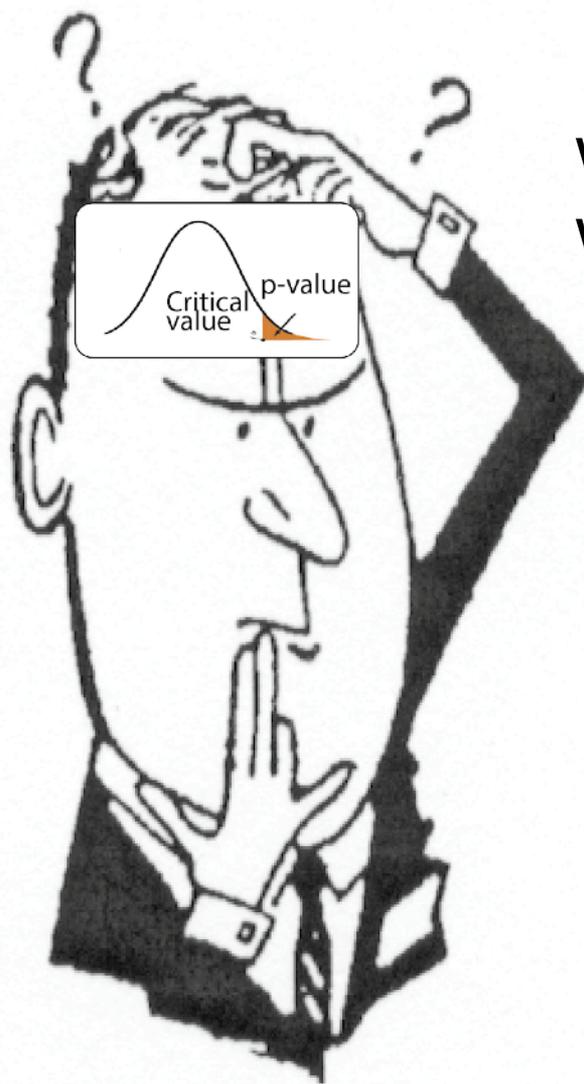
- brms to fit (Bayesian) models
- tidybayes to sample and plot predicted values
- gganimate to create animated HOPs

To get started, I recommend the `tidybayes` vignettes:

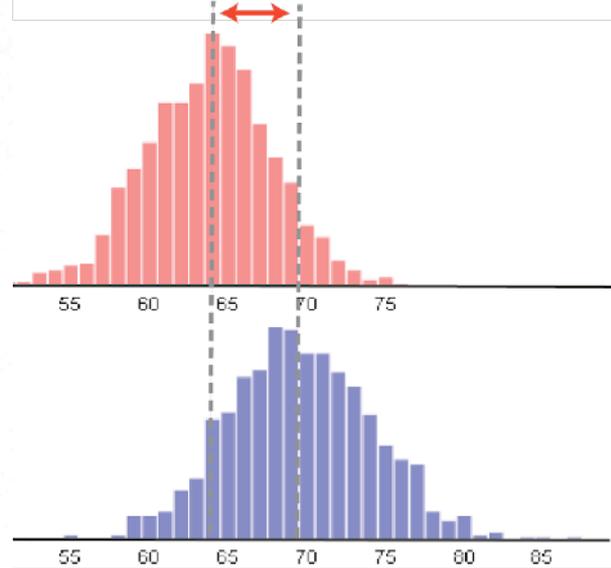
<https://mjskay.github.io/tidybayes/>

Model checks can be complicated to create and interpret. This is a promising area for innovation!





What might "random" look like?  
What process generated the data?



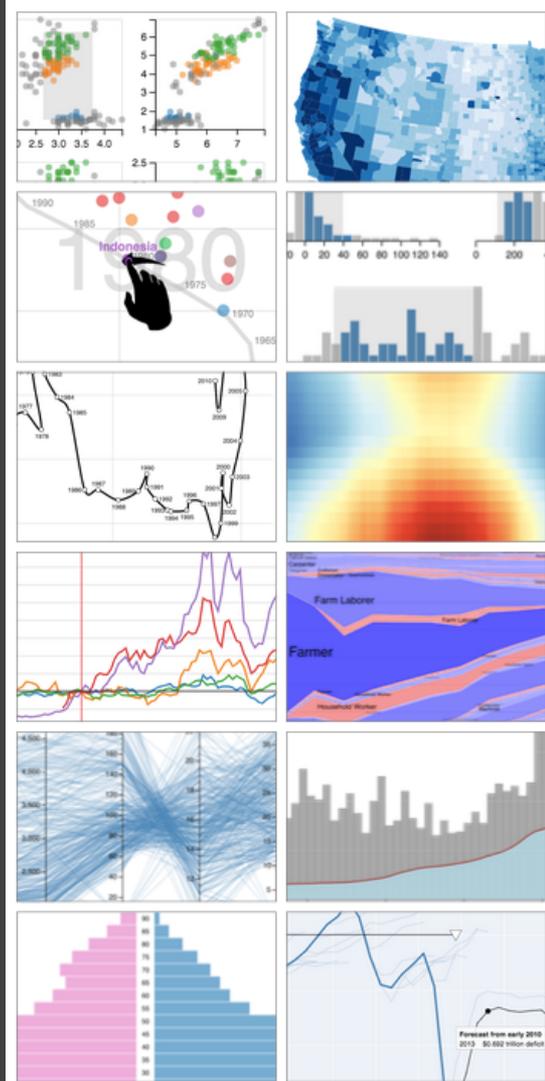
# Summary

Recognize deceptive visualizations

Don't prematurely suppress uncertainty

Frequency framing & hypothetical outcomes

Bridging exploration and confirmation



# 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**

# Summary

What Does Uncertainty Mean?

**LOTS OF THINGS**

How Should I Visualize It?

**IT DEPENDS**

What Can Go Wrong?

**A LOT**