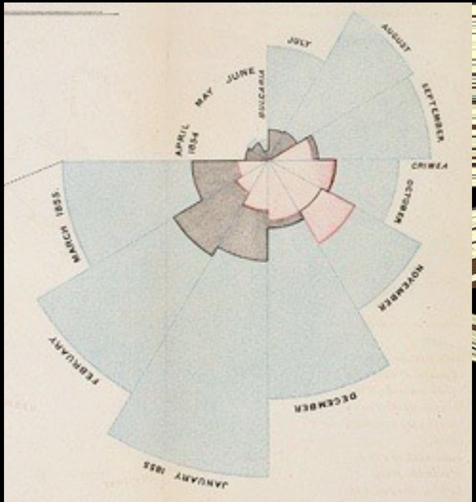
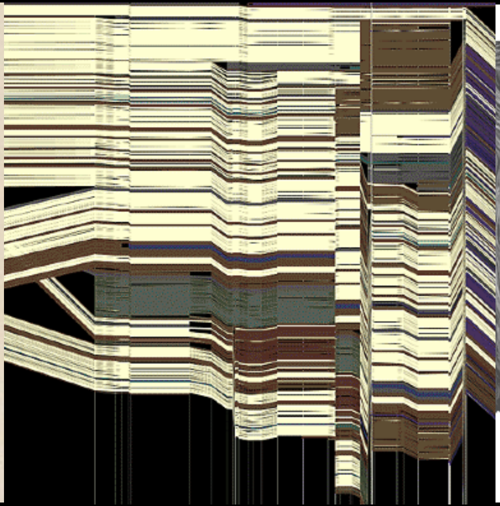
CSE 512 - Data Visualization

Uncertainty





Jeffrey Heer University of Washington

(with significant material from Michael Correll)

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

Anonymous Data Science Team Manager

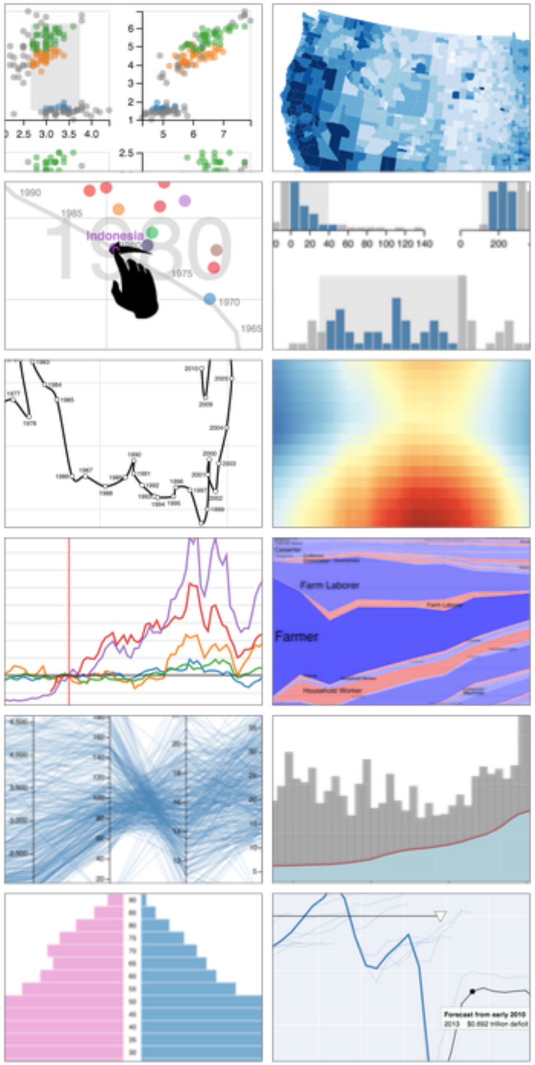
Topics

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

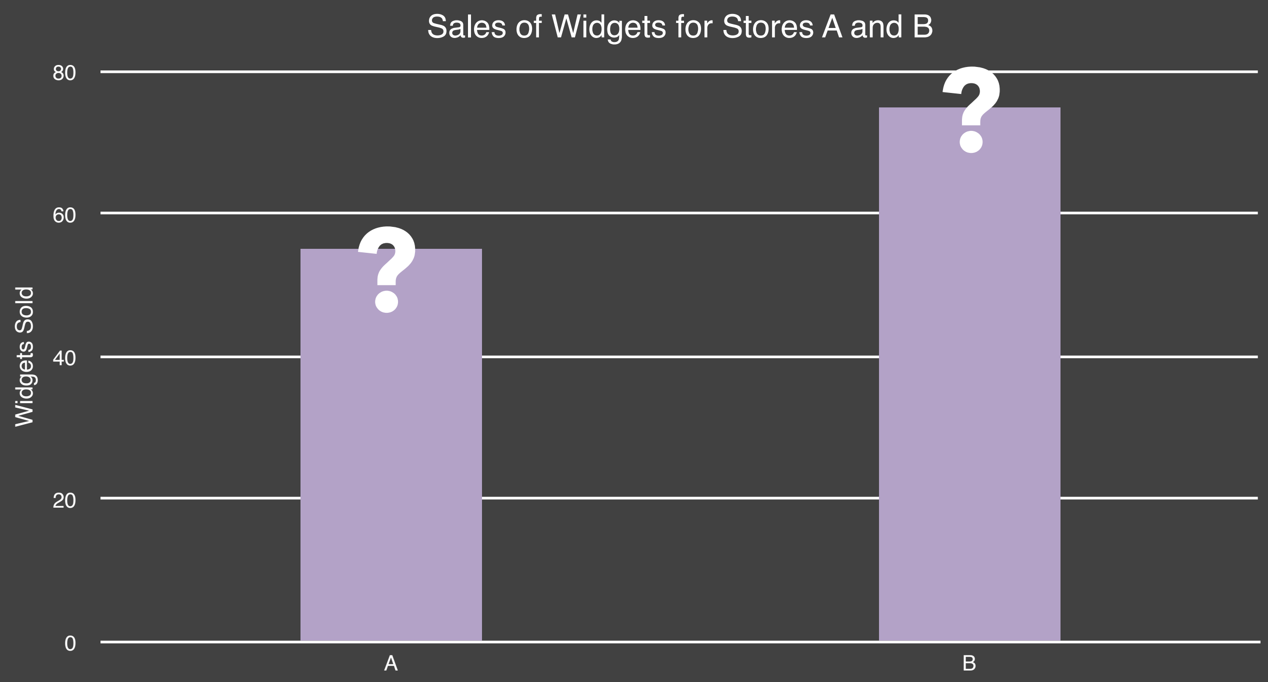
Hedging

…

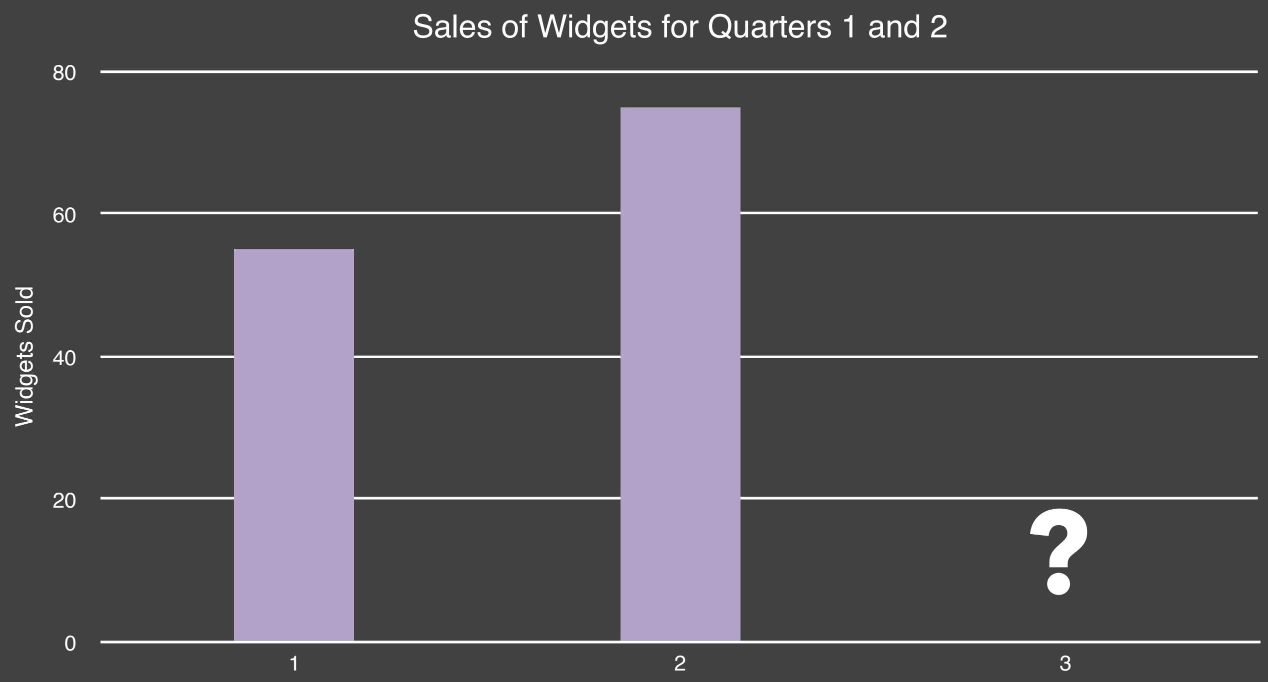
A Bar Chart


Sales of Widgets for Stores A and B
80
60
40
20
0
Widgets Sold
A
B

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”

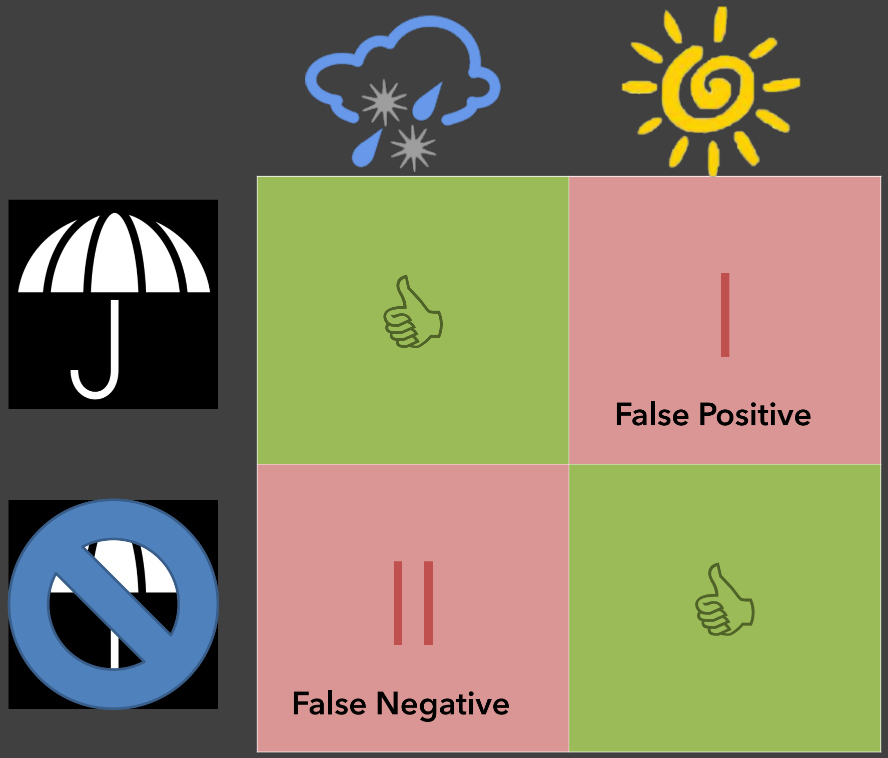
Decision Uncertainty
“We’re not sure what to do with the data”

Should I Bring an Umbrella?

Decision Uncertainty

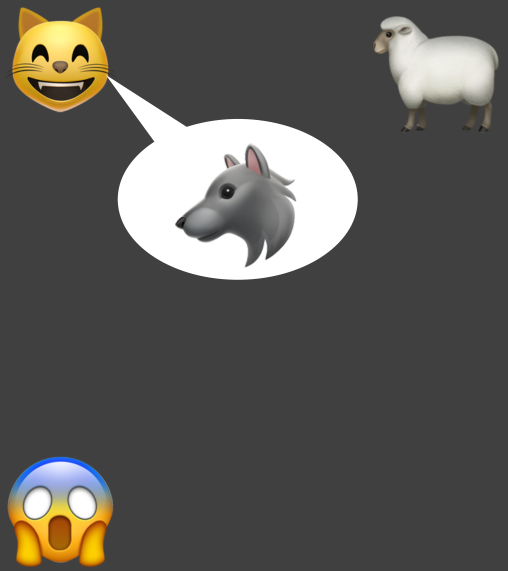
"50% Chance of Rain"

Types of Error



The Boy Who Cried Wolf

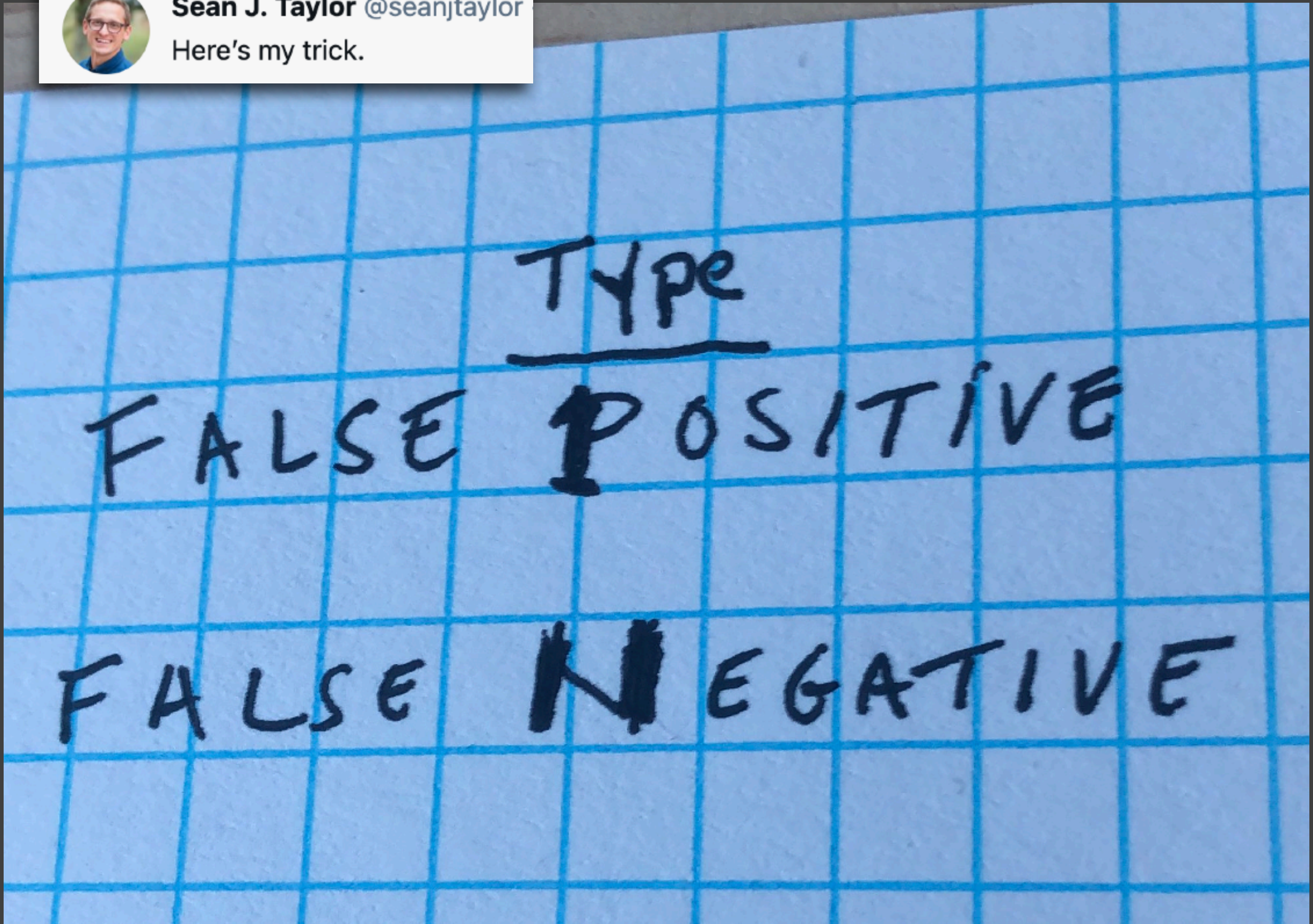
Type I: False Positive



Type II: False Negative

Sean J. Taylor @seanjtaylor

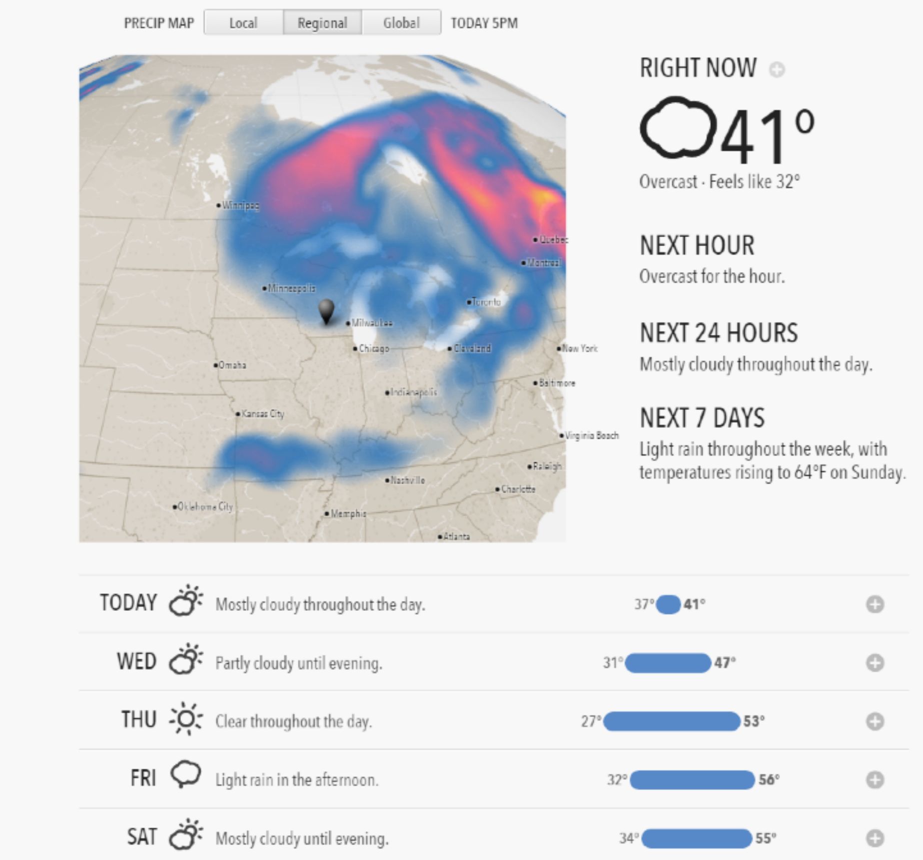
Here's my trick.



Model Uncertainty

"50% Chance of Rain"

Model Uncertainty



What Does Uncertainty Mean?

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

What Does Uncertainty Mean?

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

LOTS OF THINGS

Uncertainty Visualization

Uncertainty Visualization

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

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

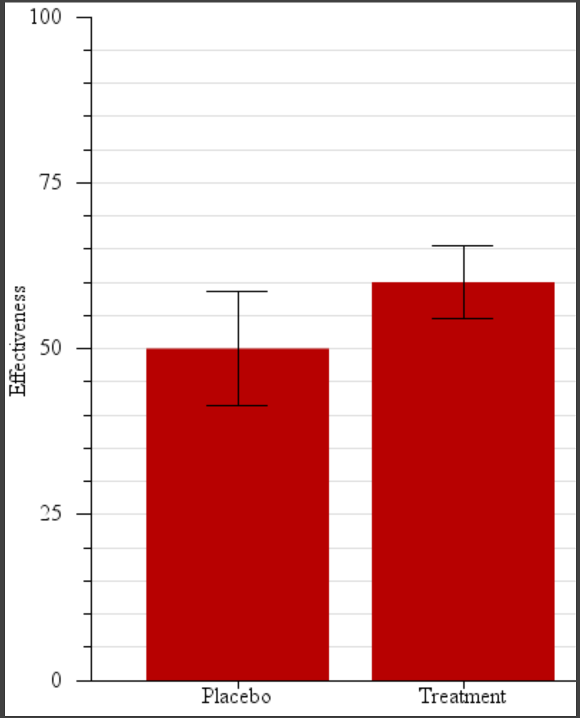
The visual presentation of uncertainty can **clash** with cognitive and perceptual biases.

Avoid Prematurely Suppressing Uncertainty

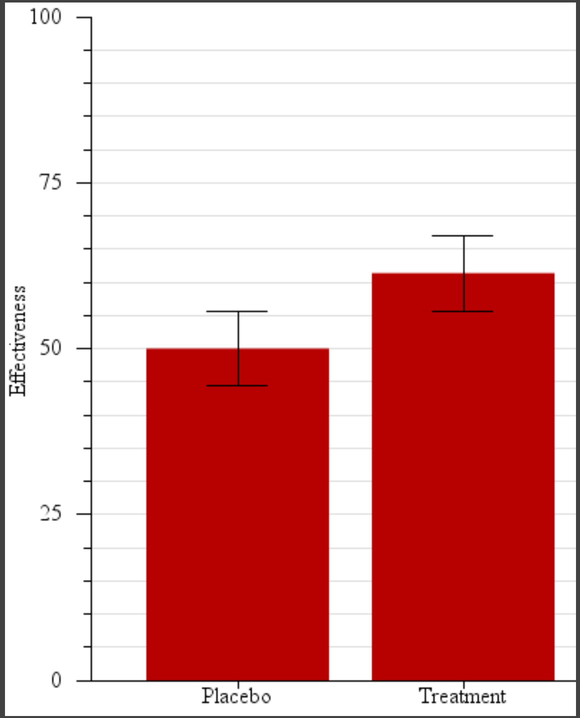
Error Bars

The mean treatment effect is higher than than the placebo.

Is this difference in means *statistically significant?*

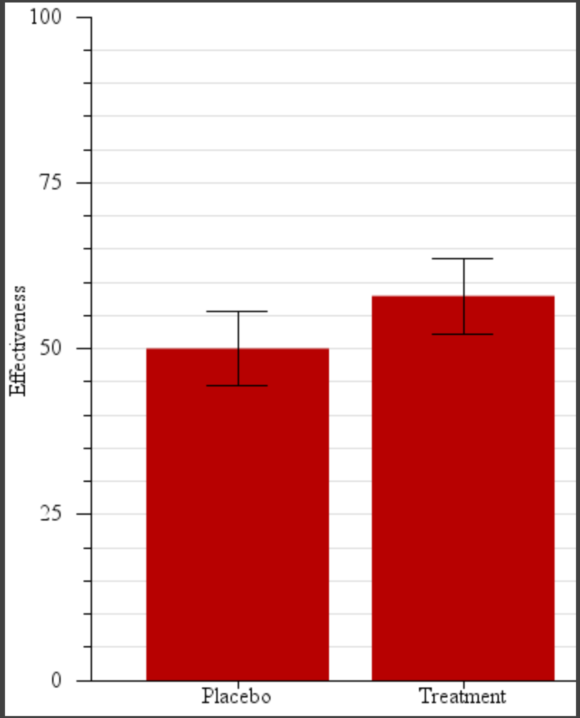


Guess the p-value…



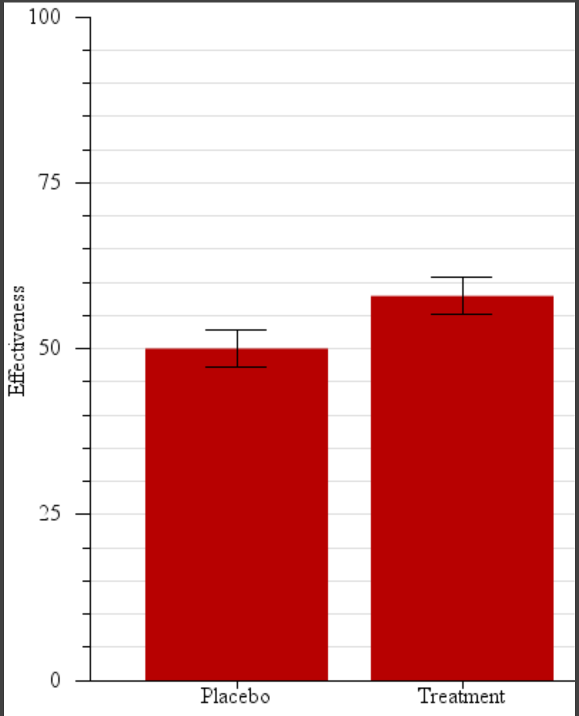
Error bars depict
95% Conf. Interval

Guess the p-value…



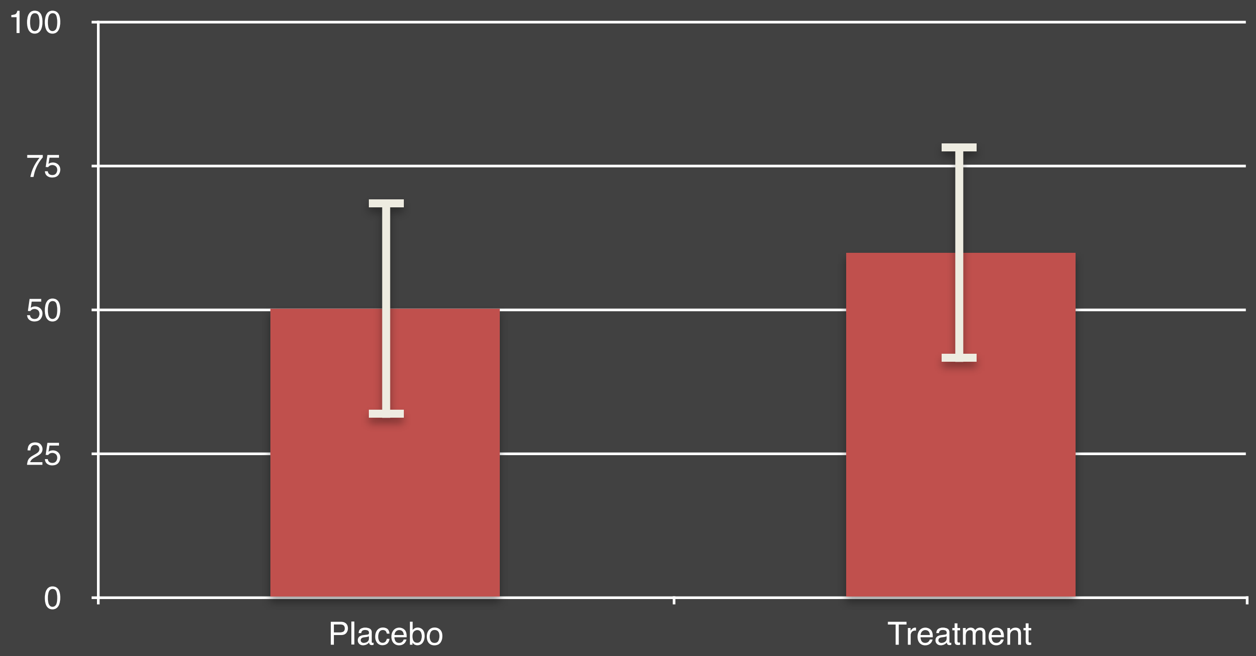
Error bars depict
95% Conf. Interval

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?

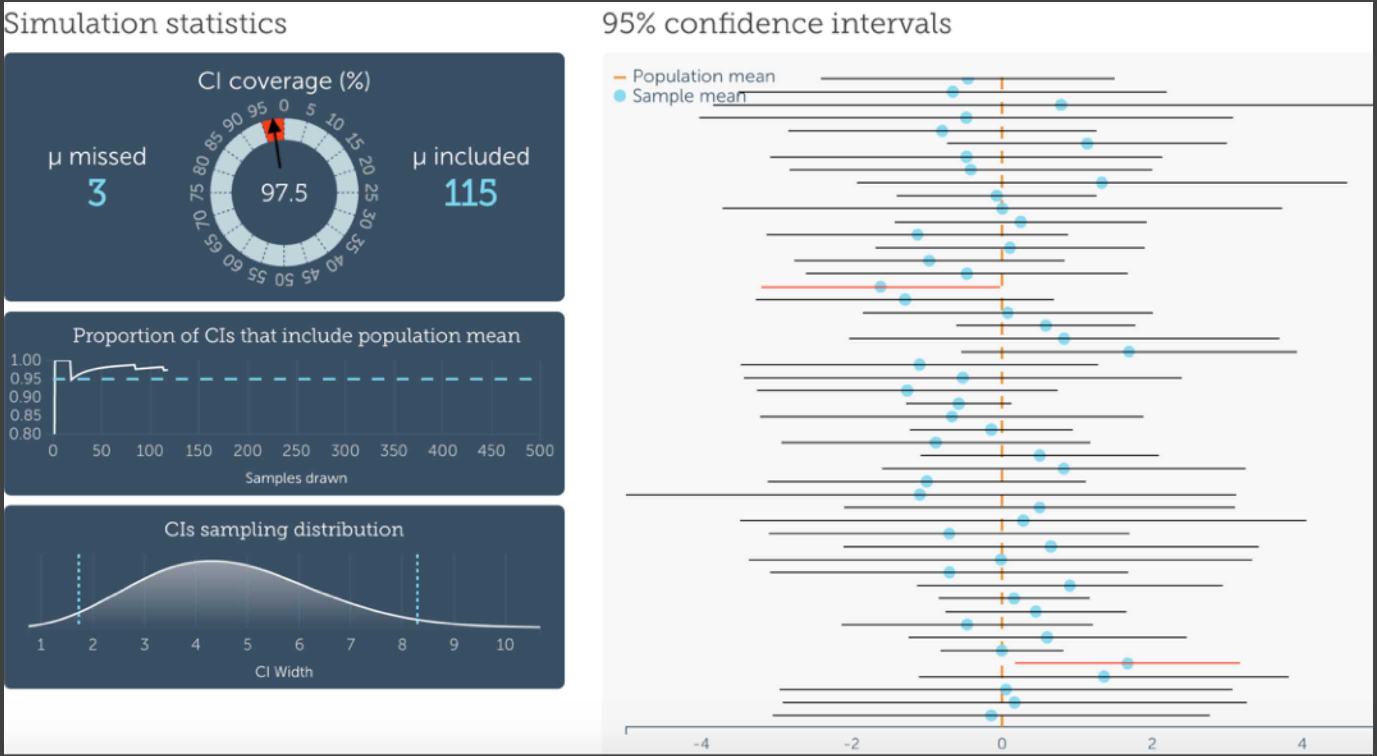
One interpretation is: there is a 95% chance that the population mean is within the interval.

Wrong!

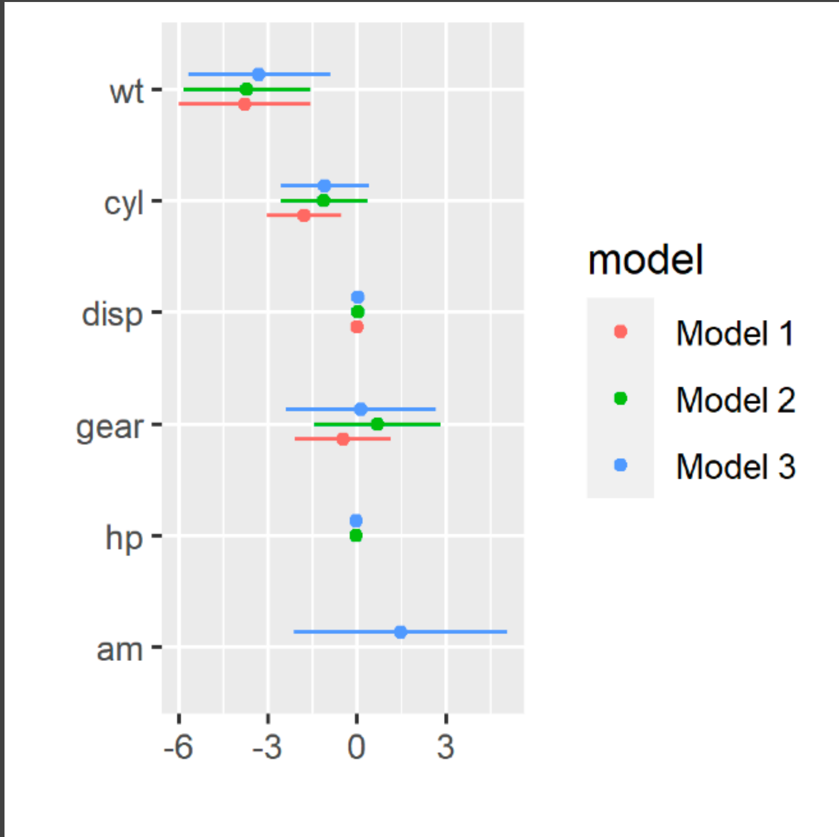
Rather, given an infinite number of independent experiments, 95% of the confidence intervals generated will contain the true population mean.

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

Confidence Intervals



Regression Coefficients



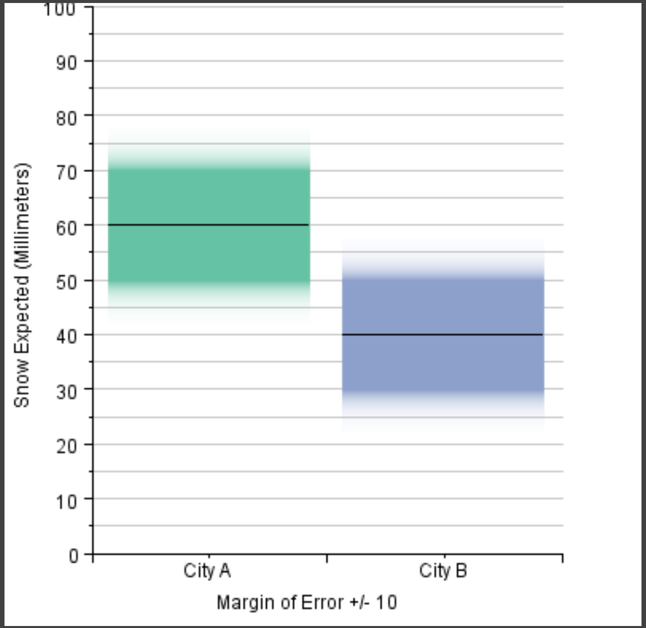
95% CIs for regression model parameters.

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

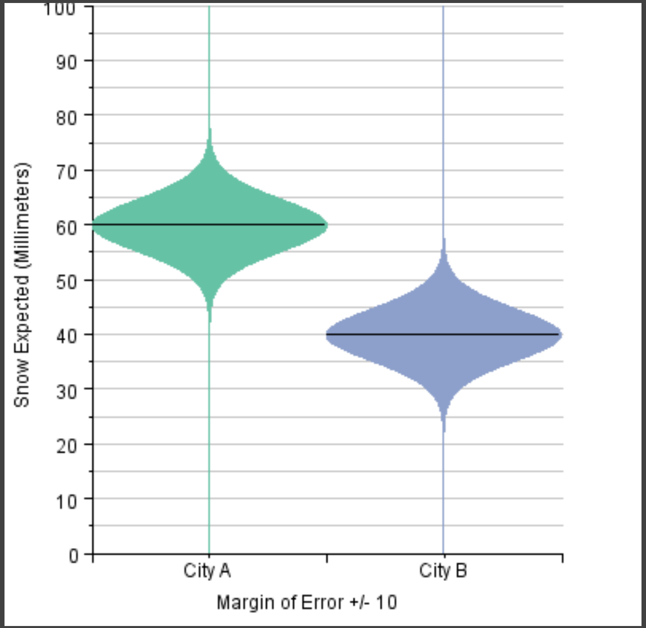
Visual comparison: does the CI overlap 0?

Alternatives to Error Bars

Gradient Plot

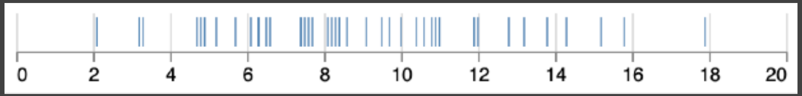


Violin Plot

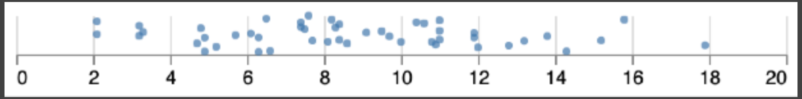


Distribution Visualizations

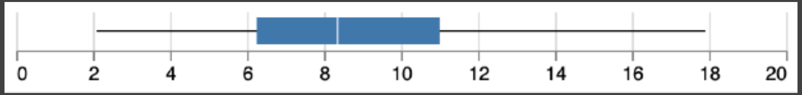
Strip Plot



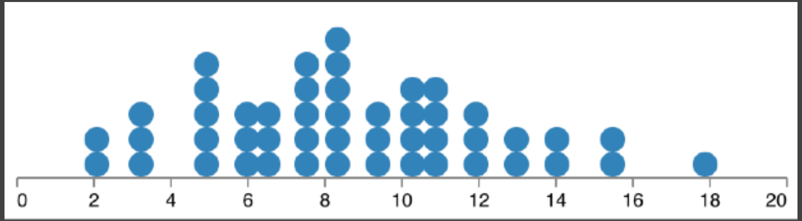
Jittered Plot



Box Plot



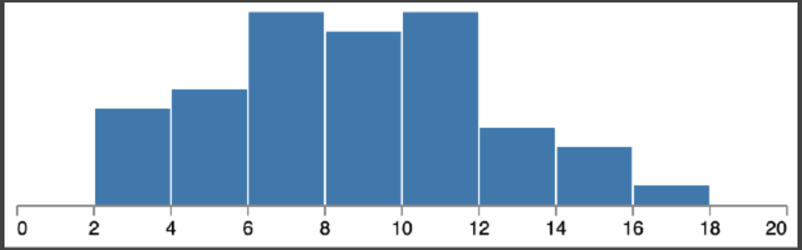
Dot Plot



Distribution Visualizations

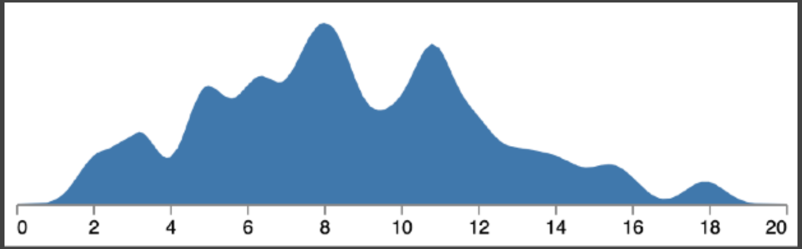
Histogram

bin size = 2



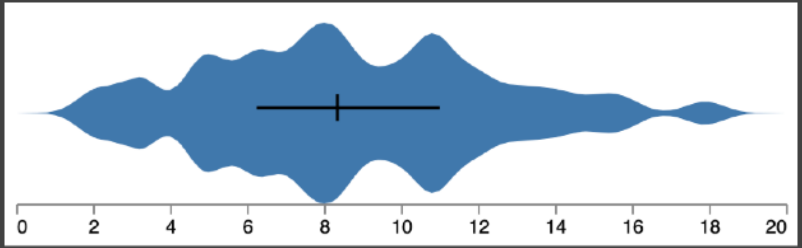
Density Plot

kde, σ = 0.5



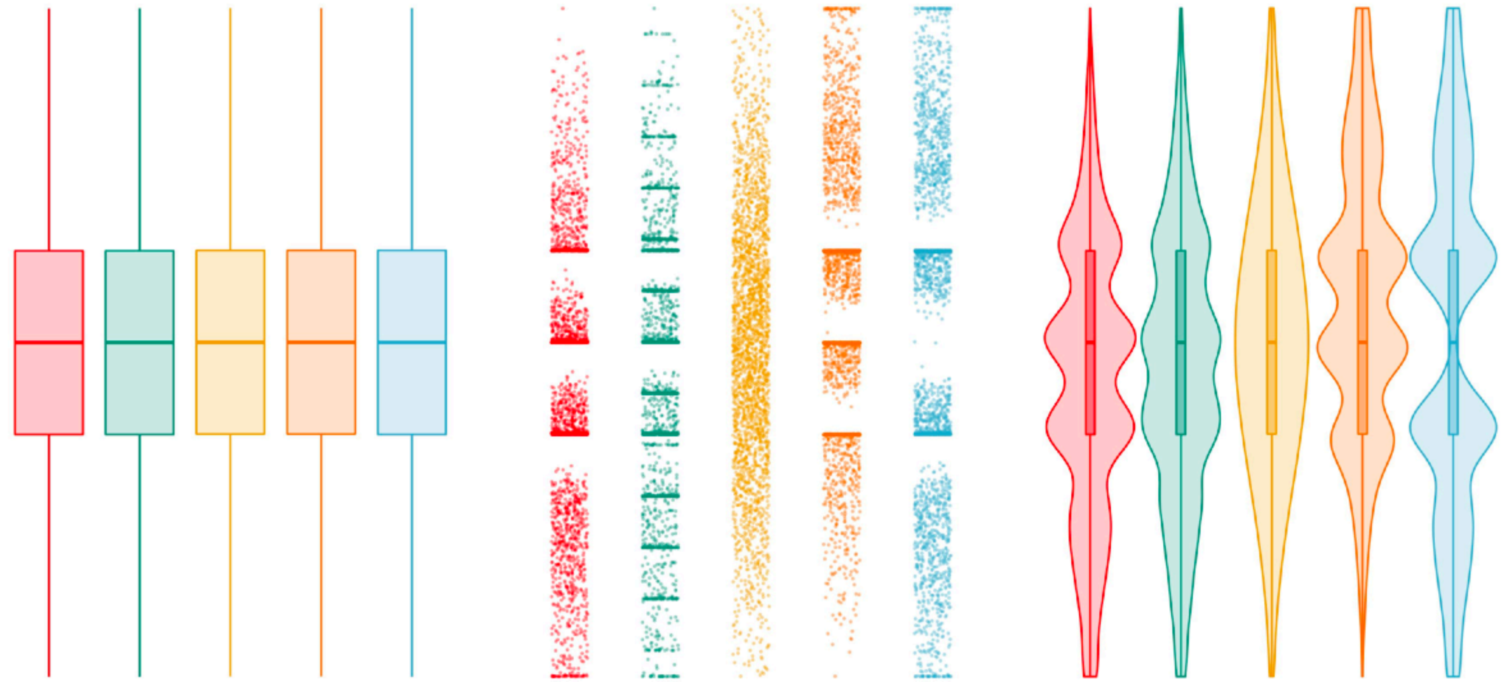
Violin Plot

kde, σ = 0.5



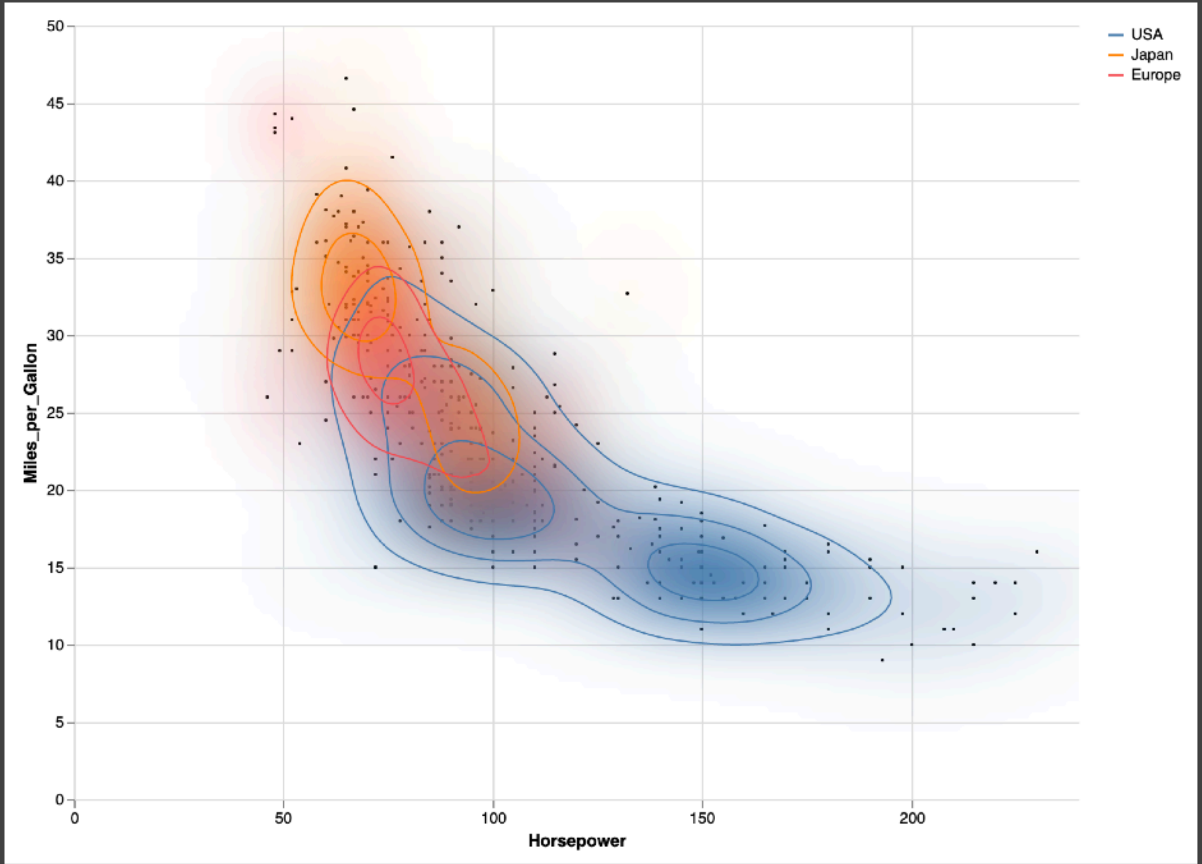
Identical boxplots, different distributions

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.



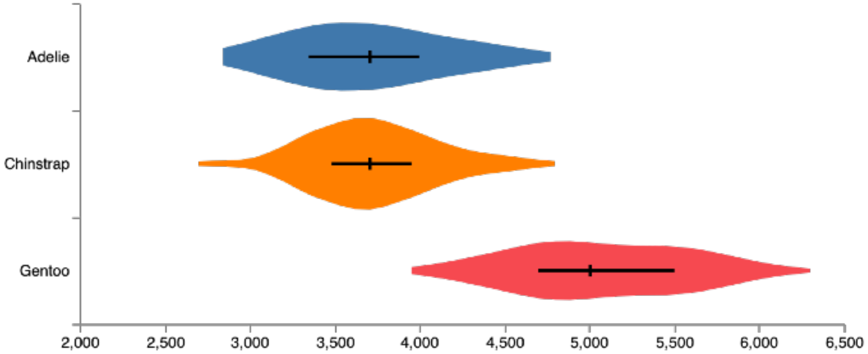
@MBozukova | #30DayChartChallenge | Day 27 - educational
Data source: {datasauRus} R package

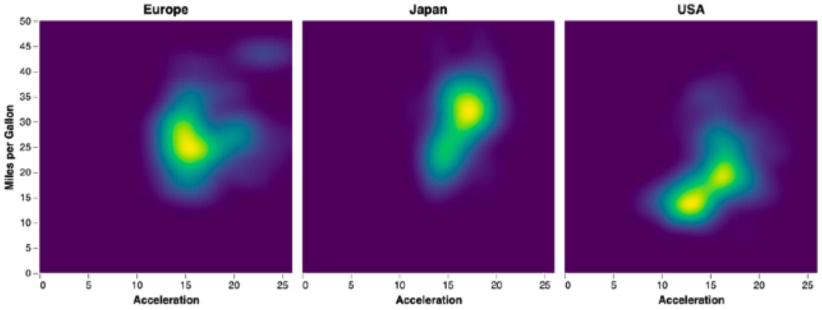
Now in 2D! Heatmaps, Contours

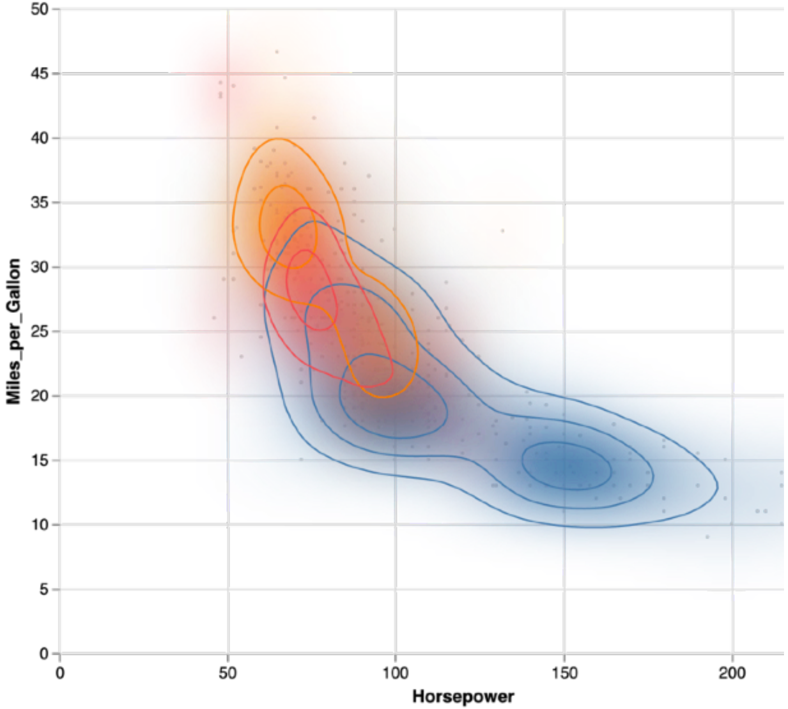


Kernel Density Estimation (KDE)

Enables violin plots, heat maps, contour plots…

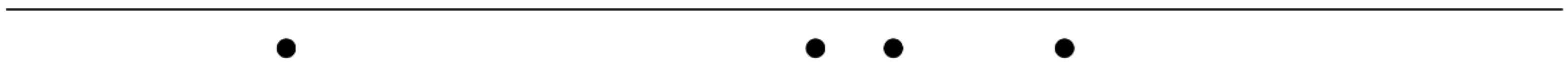






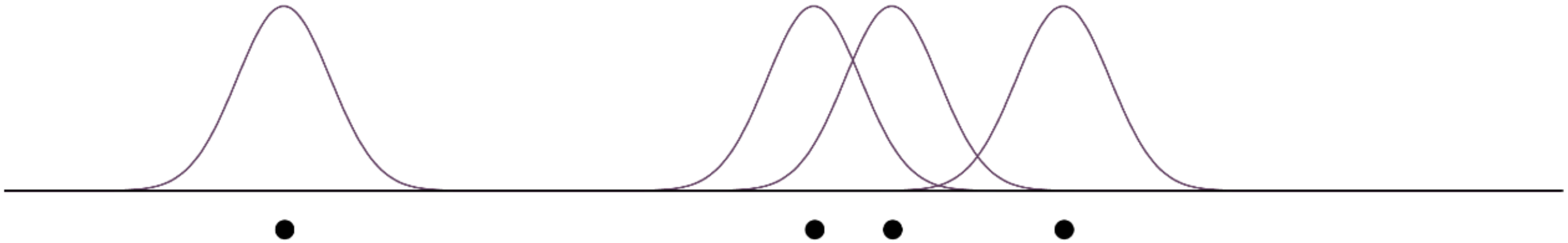
Kernel Density Estimation

For a set of input data points…



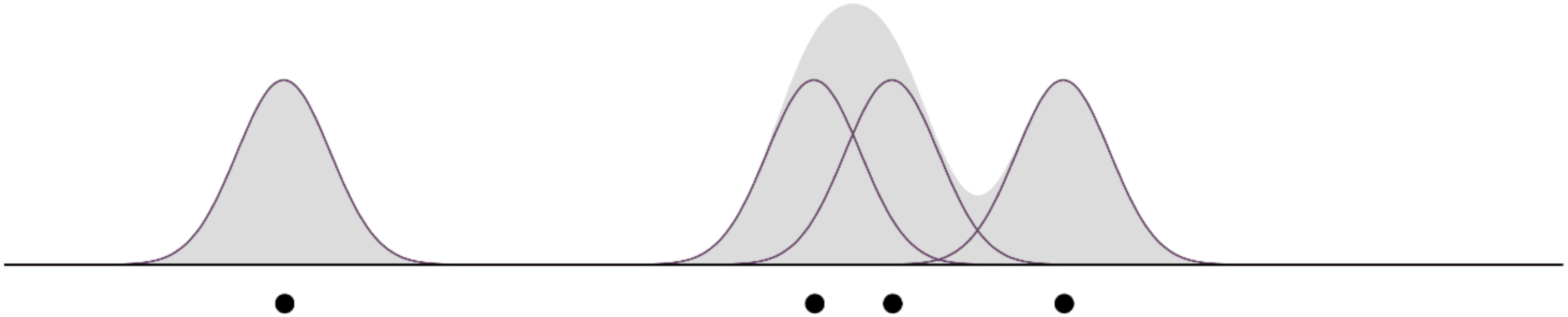
Kernel Density Estimation

Represent each point with a “kernel” distribution



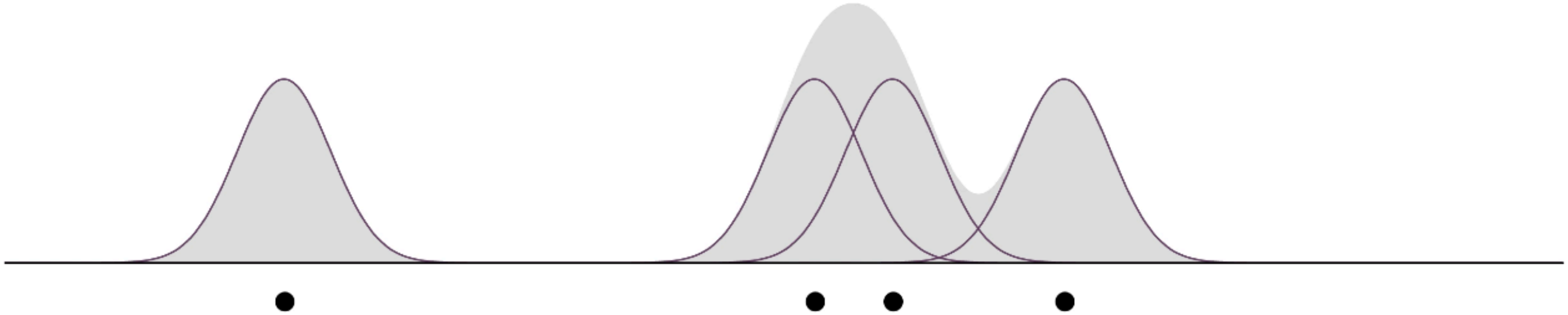
Kernel Density Estimation

Sum the kernels to form a density estimate



Kernel Density Estimation

Sized by bandwidth (standard deviation)



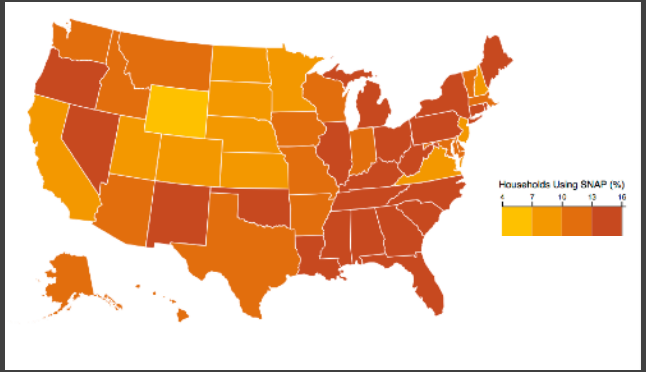
Visual Encodings of Uncertainty

Uncertainty Vis Pipeline

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

SNAP

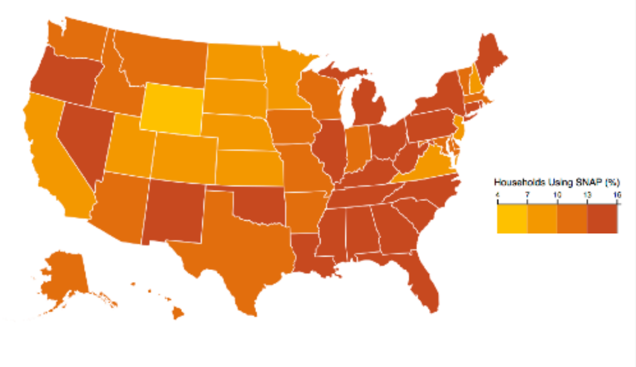
Data Map

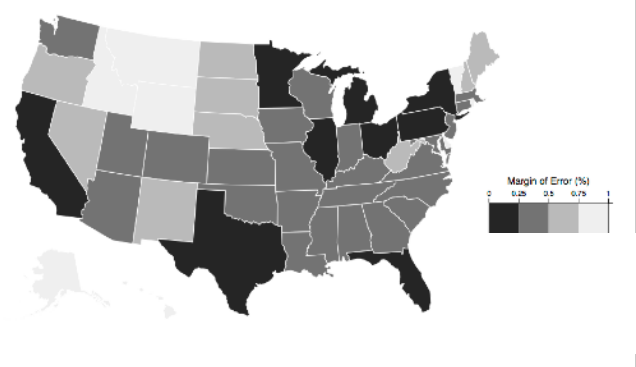


SNAP

Data Map

Uncertainty Map





Uncertainty Vis Pipeline

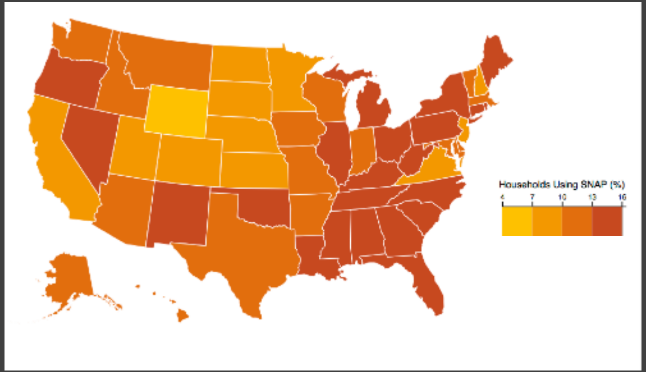
1) Quantify uncertainty
2) Choose a free visual variable
3) Encode uncertainty with the variable

Uncertainty Vis Pipeline

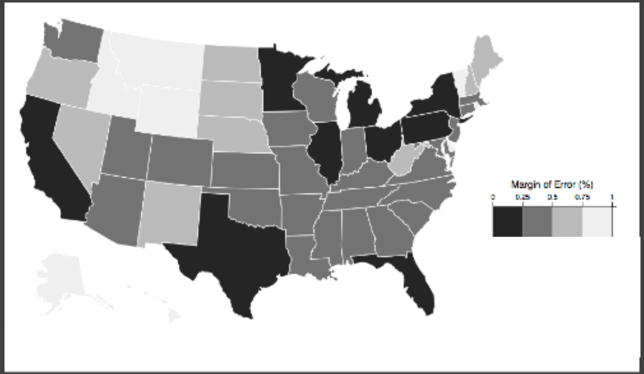
1) Quantify uncertainty
2) Choose a free visual variable
3) Encode uncertainty with the variable
4) Unify the Data Map and Uncertainty Map

How to Unify?

Data Map

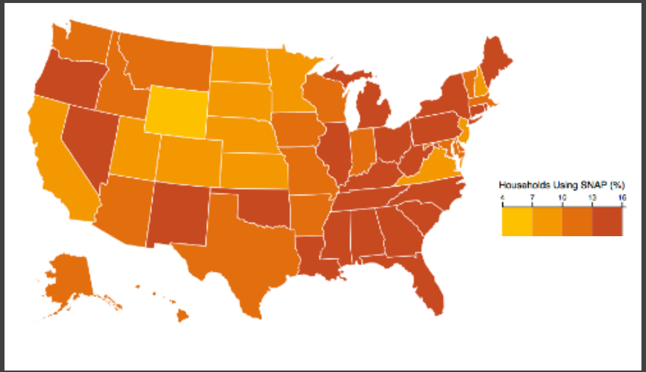


Uncertainty Map

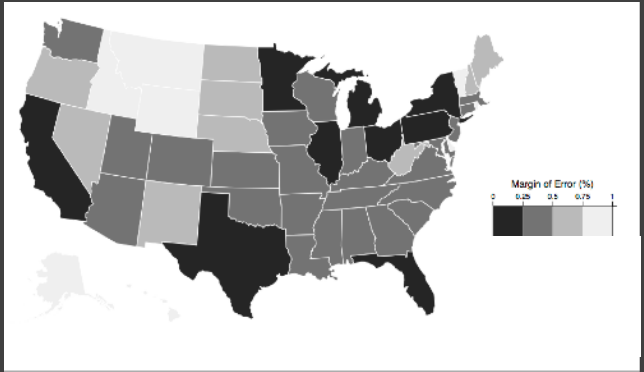


Juxtaposition

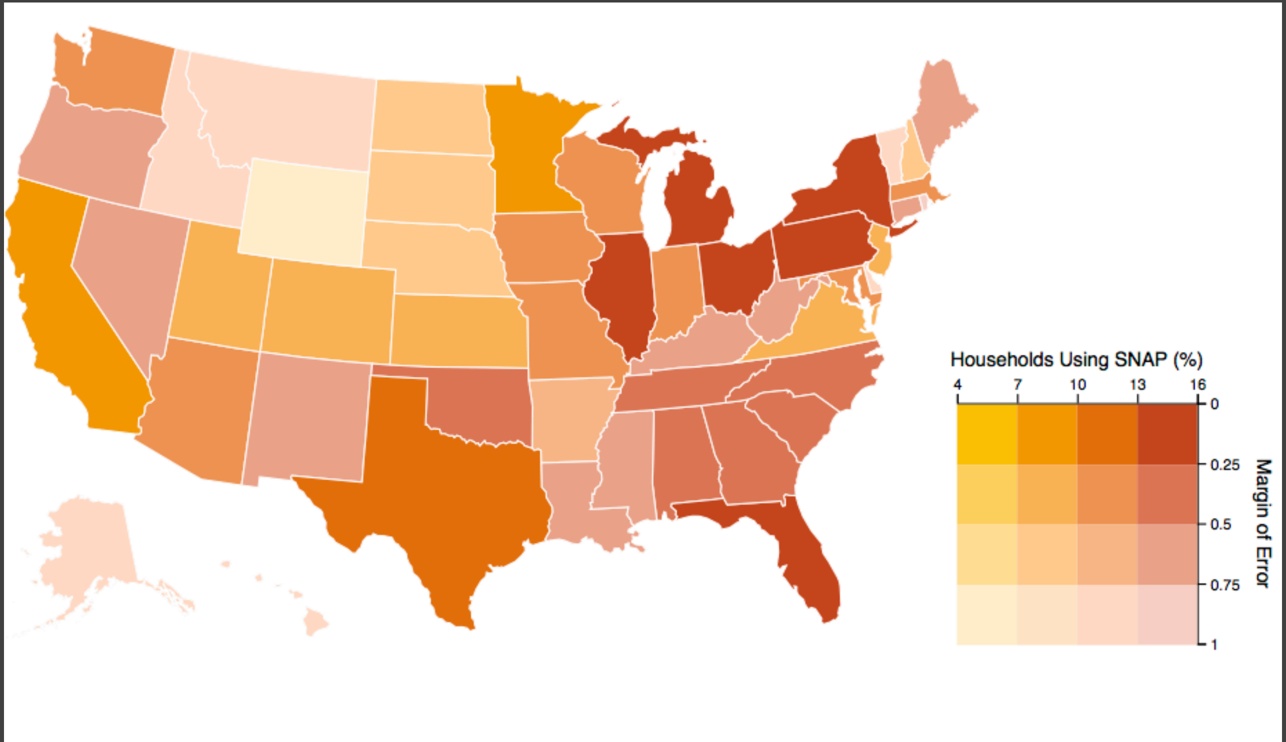
Data Map



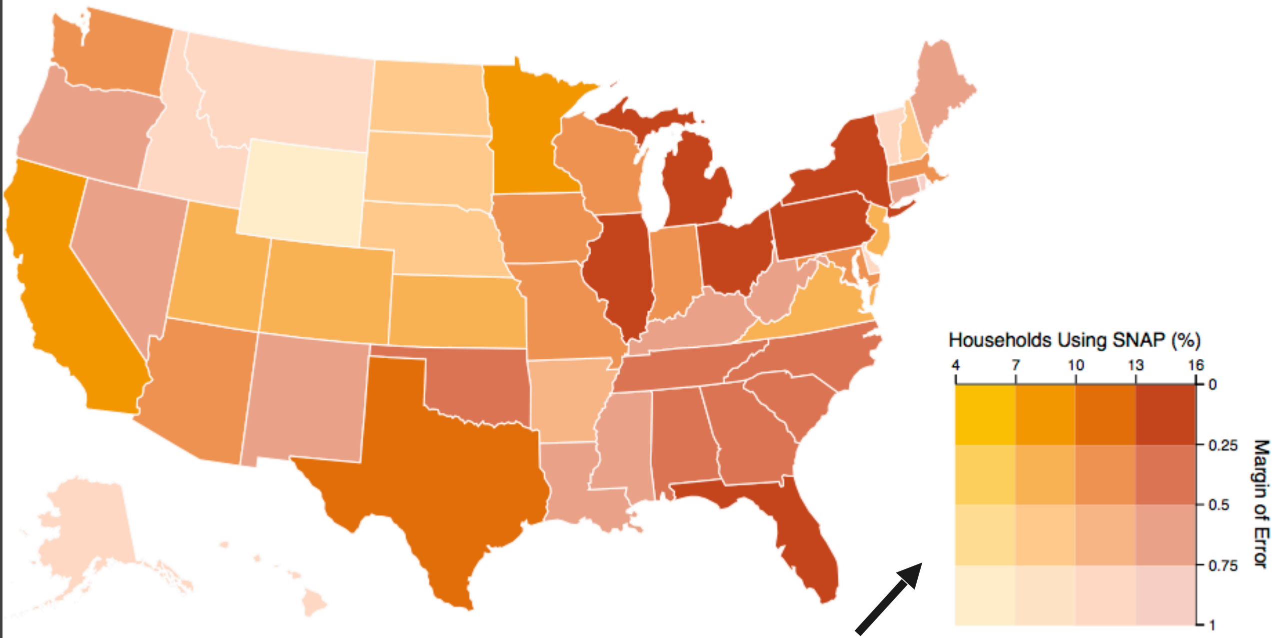
Uncertainty Map



Superposition



Superposition



Bivariate 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

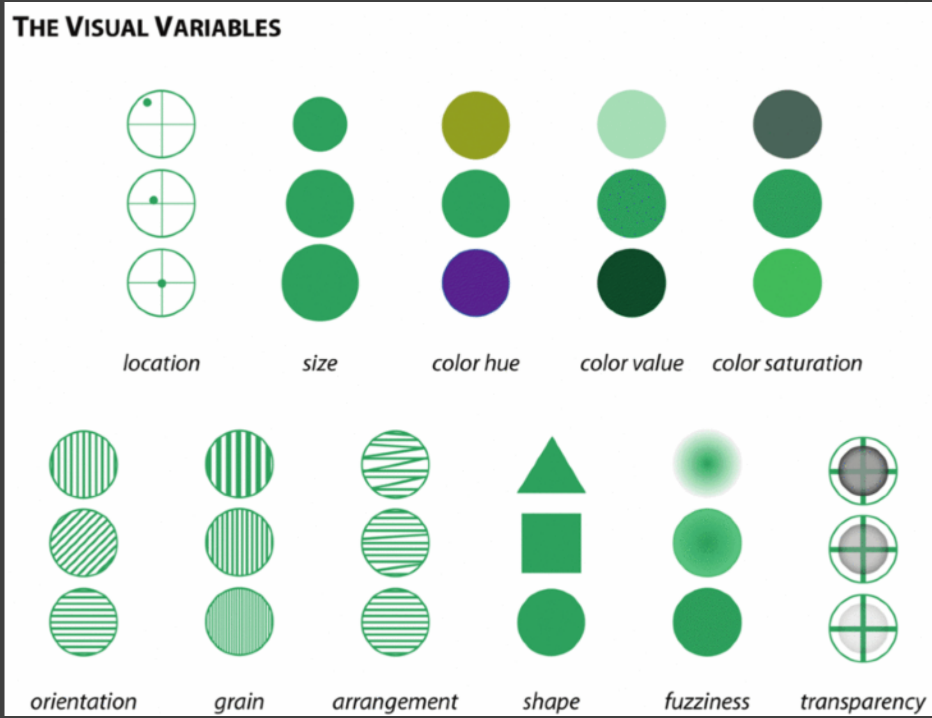
Uncertainty Vis Pipeline

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

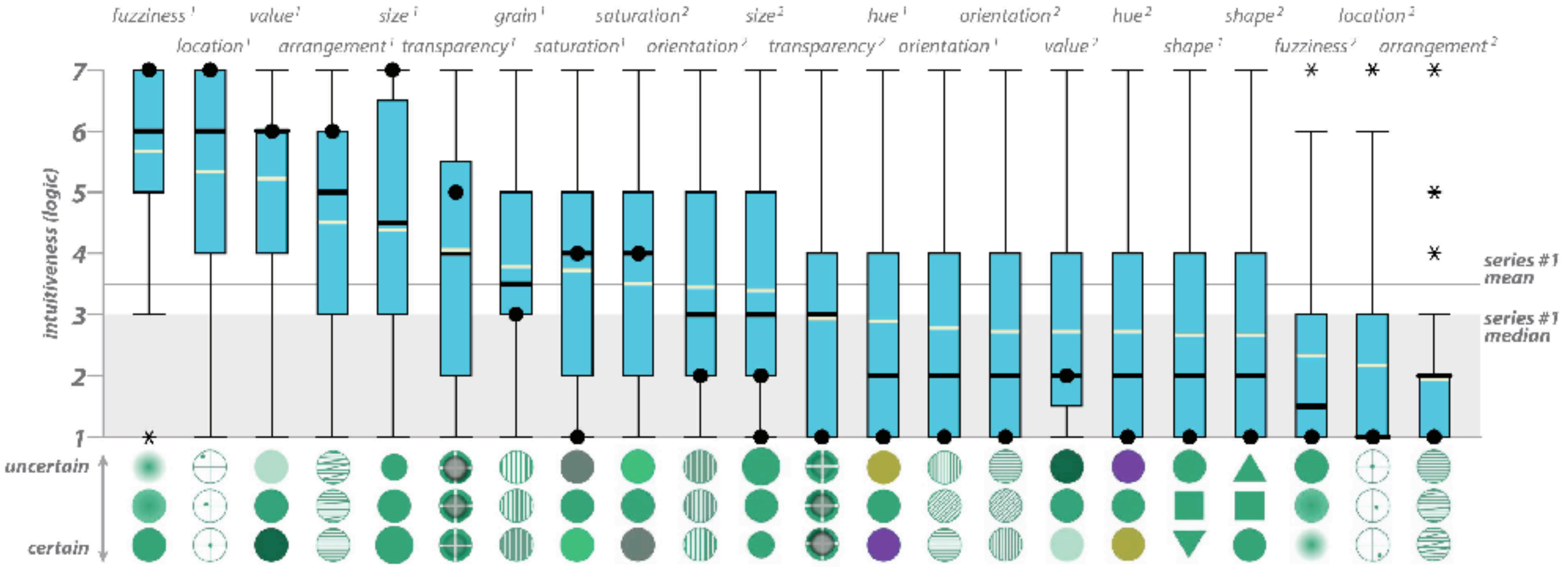
Semiotics of Uncertainty

Semiotics of Uncertainty

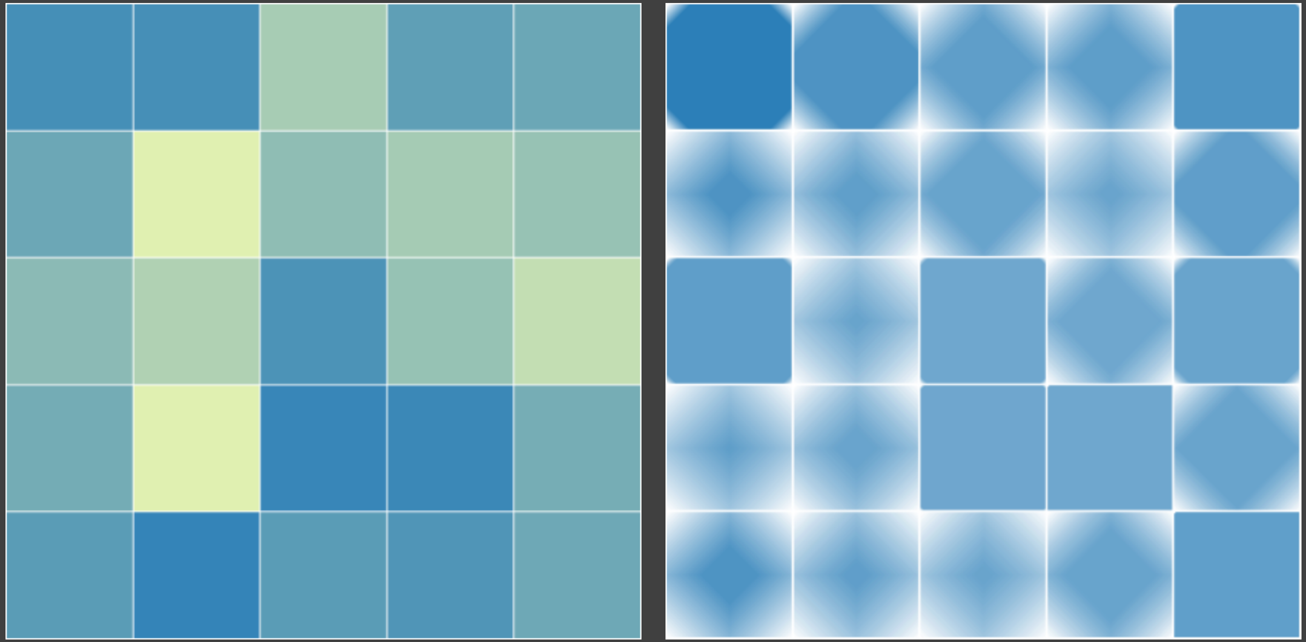


MacEachren et al. Visual Semiotics & Uncertainty Visualization: An empirical study. IEEE VIS, 2012.

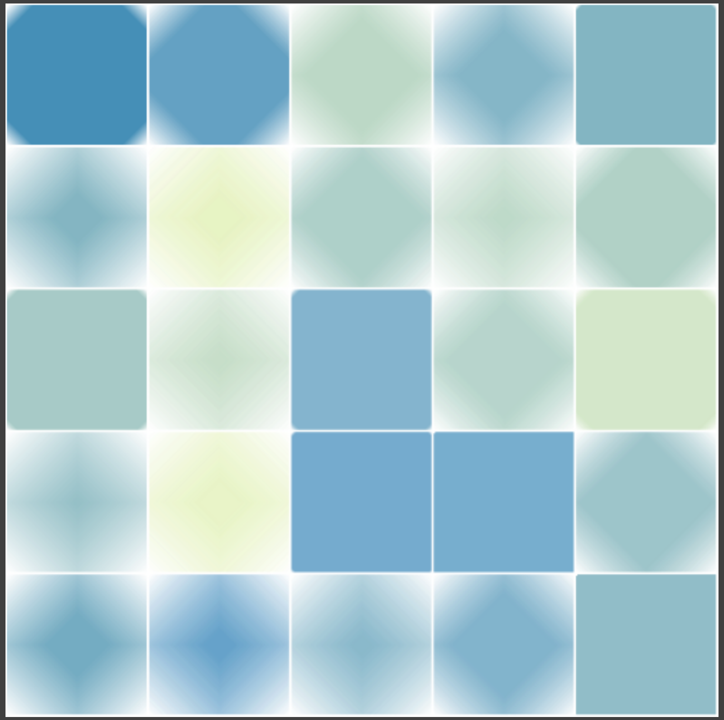
Series #1: General Uncertainty by Visual Variable



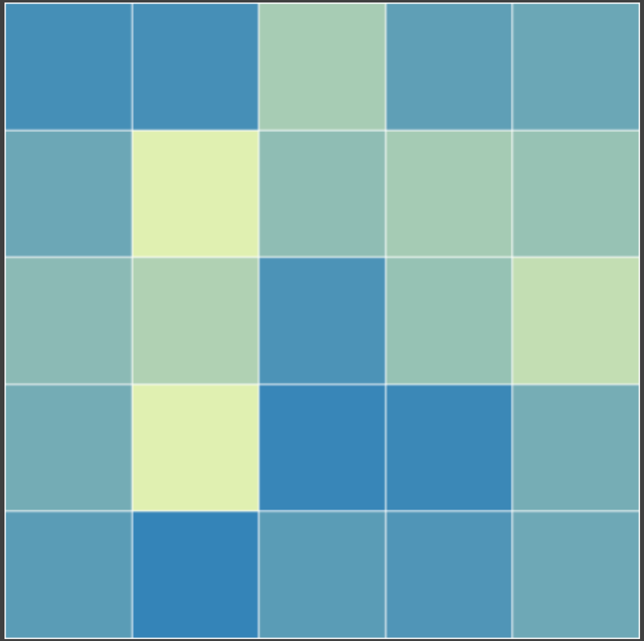
Fuzziness Juxtaposition

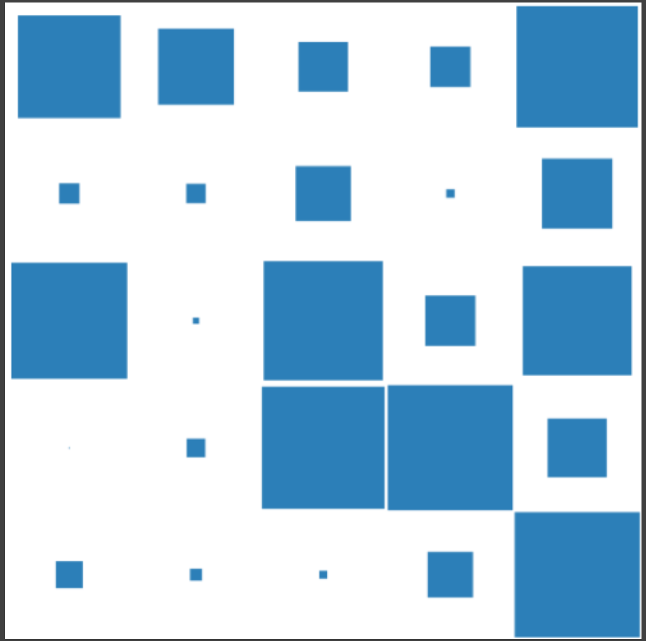


Fuzziness Superposition

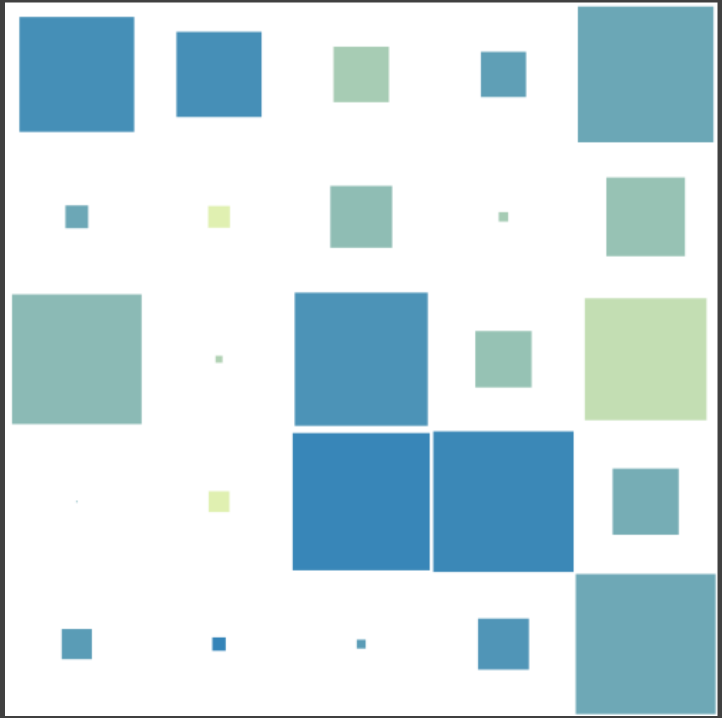


Size Juxtaposition

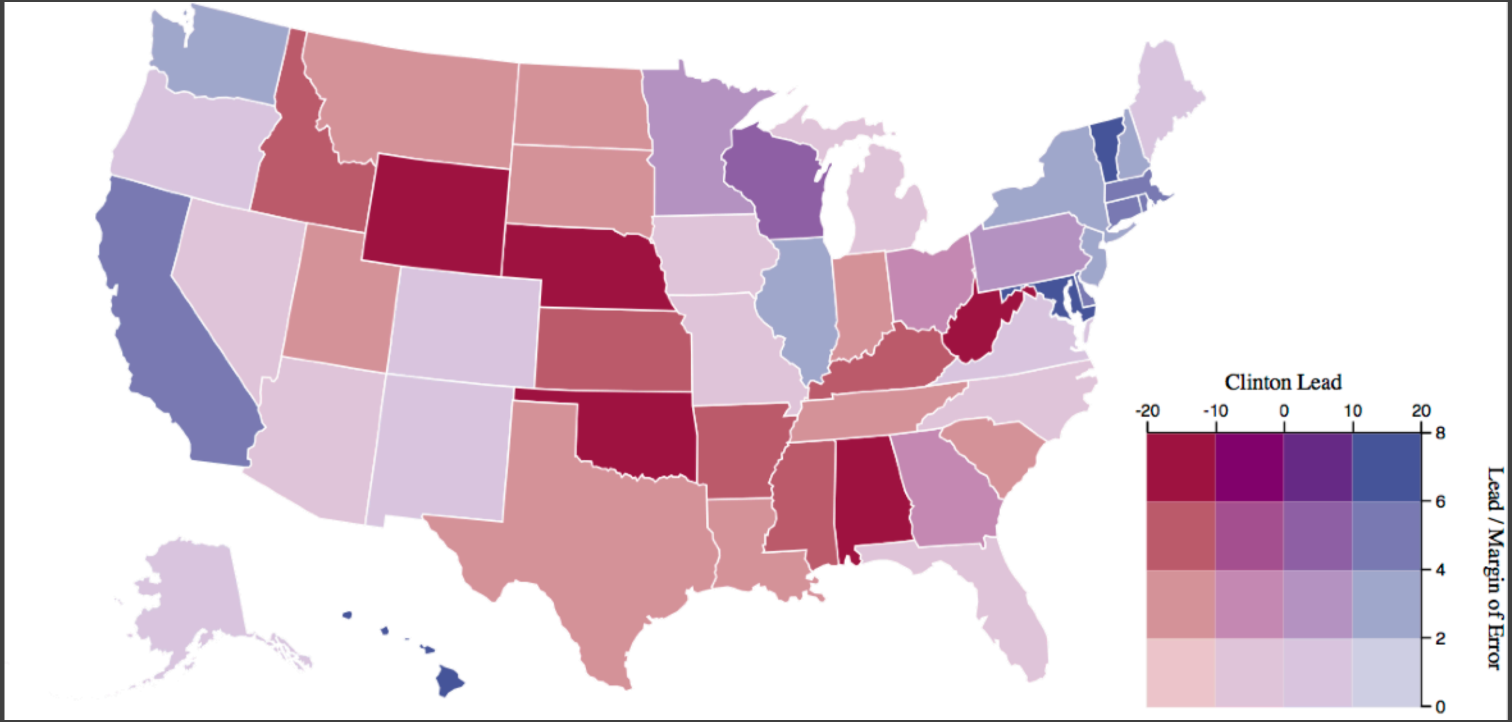




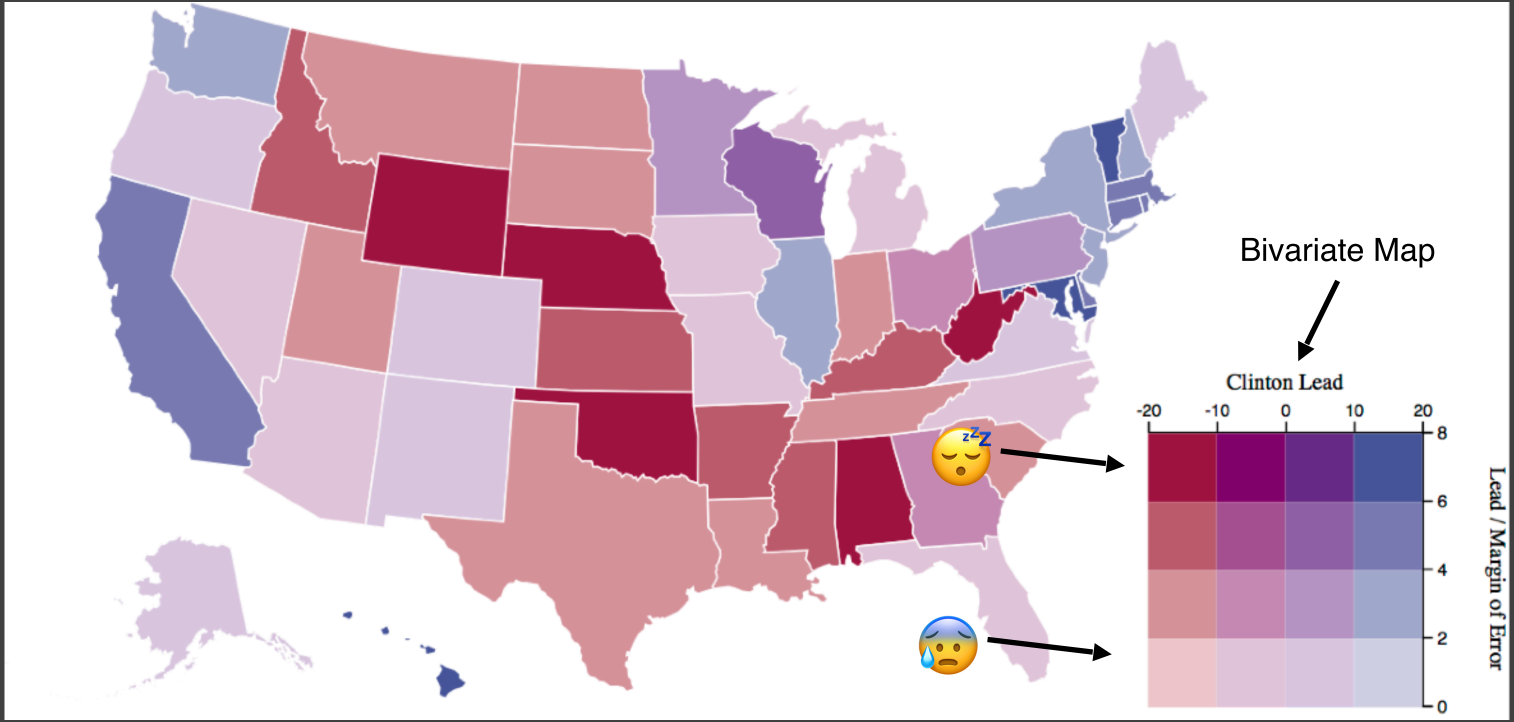
Size Superposition



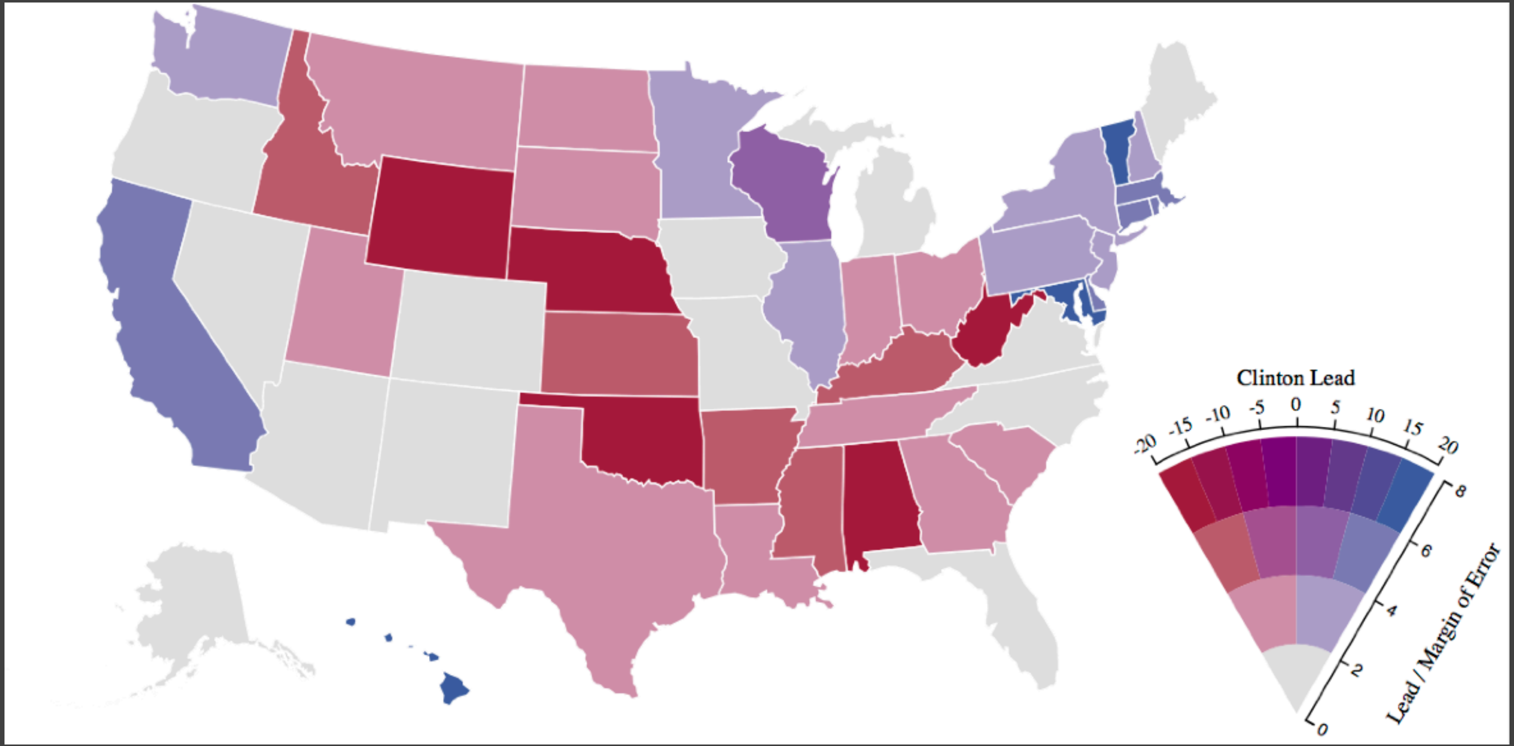
Value Suppressing Uncertainty Palettes



Bivariate Map

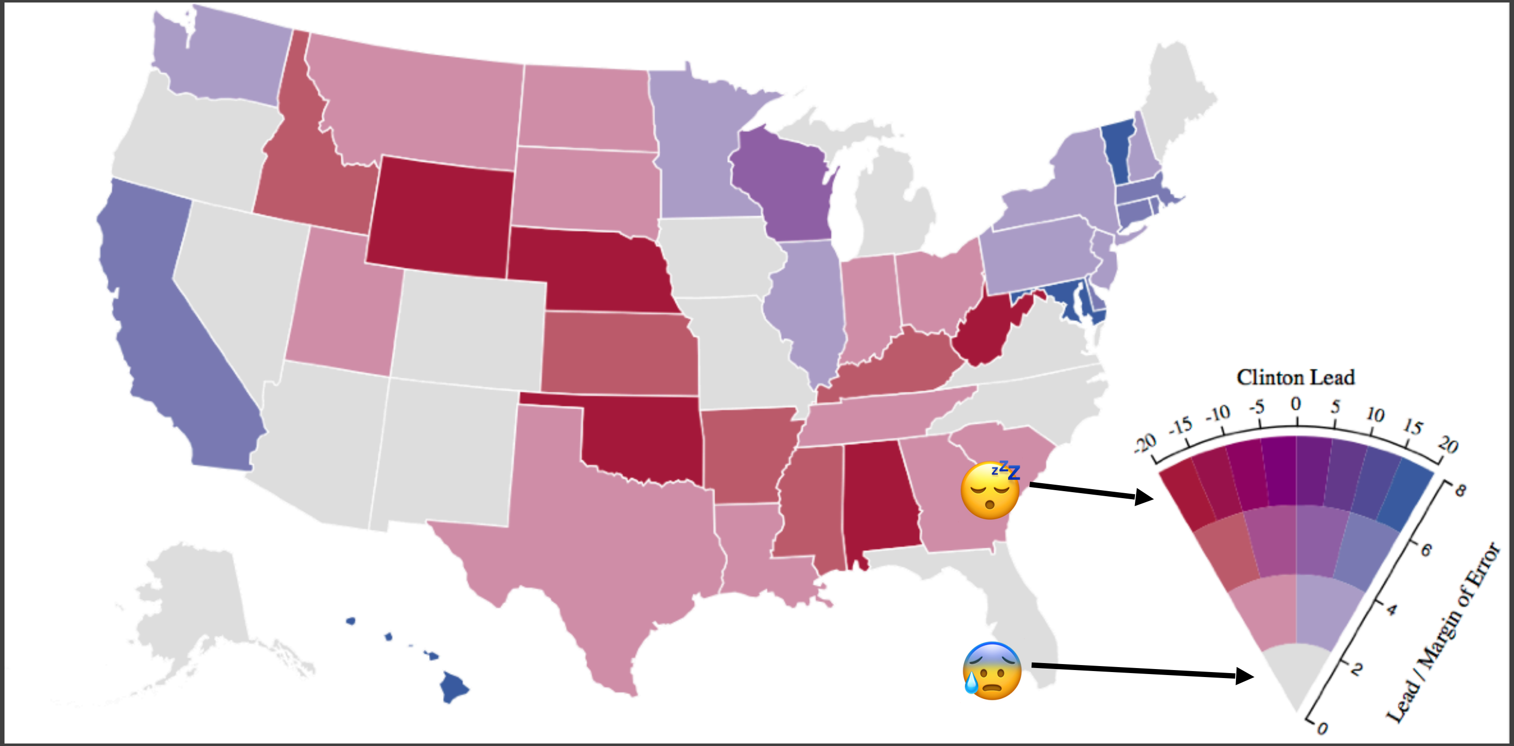


Value Suppressing Uncertainty Palettes



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

Value Suppressing Uncertainty Palettes



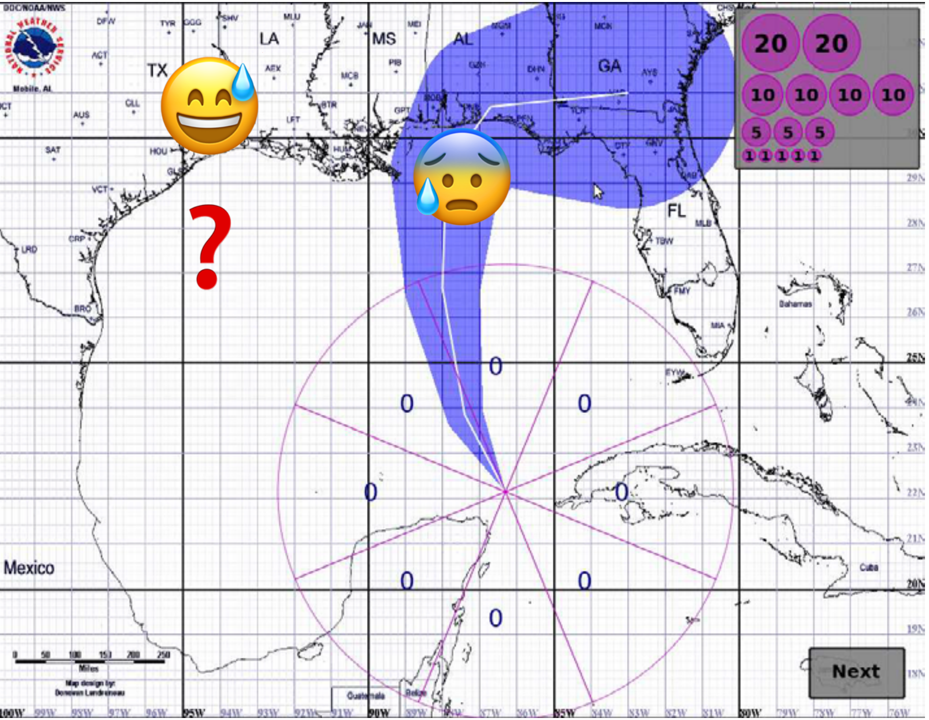
Correll, Moritz & Heer. “Value-Suppressing Uncertainty Palettes.” CHI 2018.

Encoding Uncertainty

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

However, intuitive variables may not always be accurately interpreted!

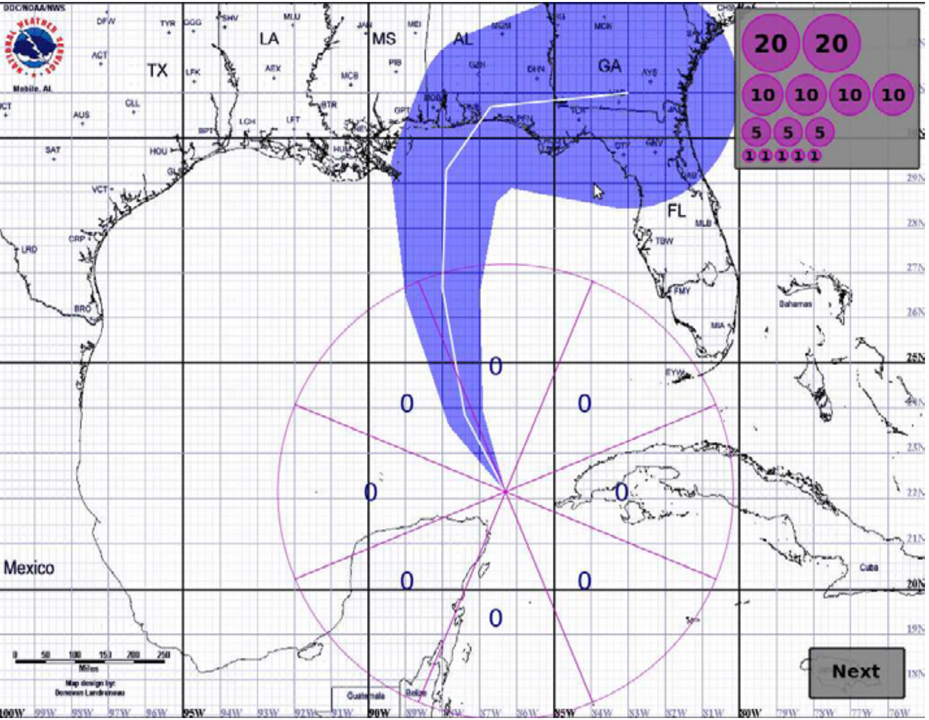
Frequency Framing & Hypothetical Outcomes

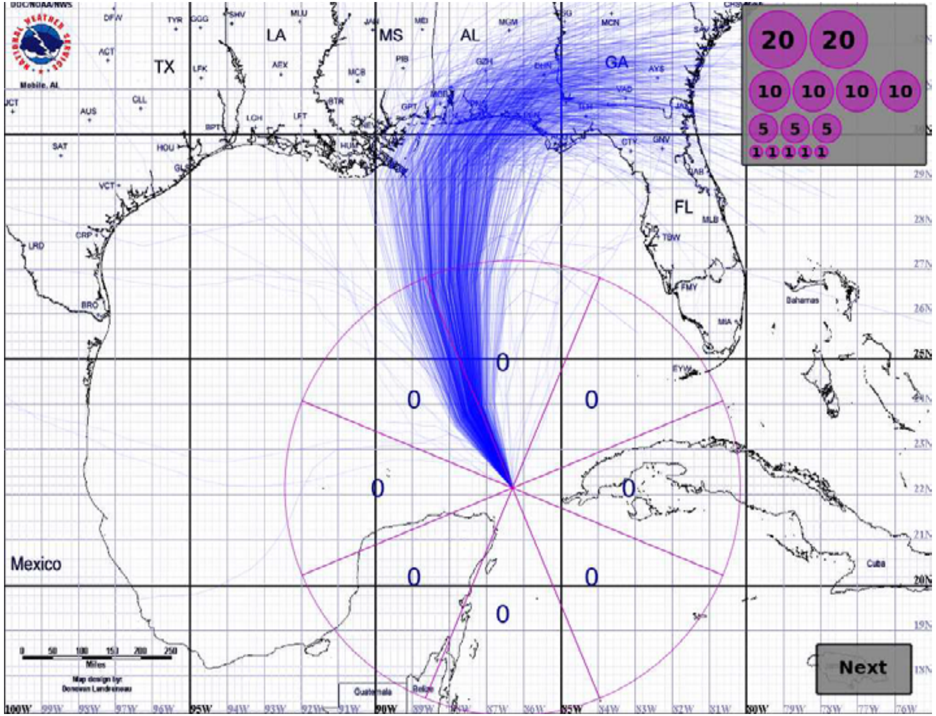


Size or likelihood of hurricane?

Is New Orleans safe?

[Cox et al. 2013]

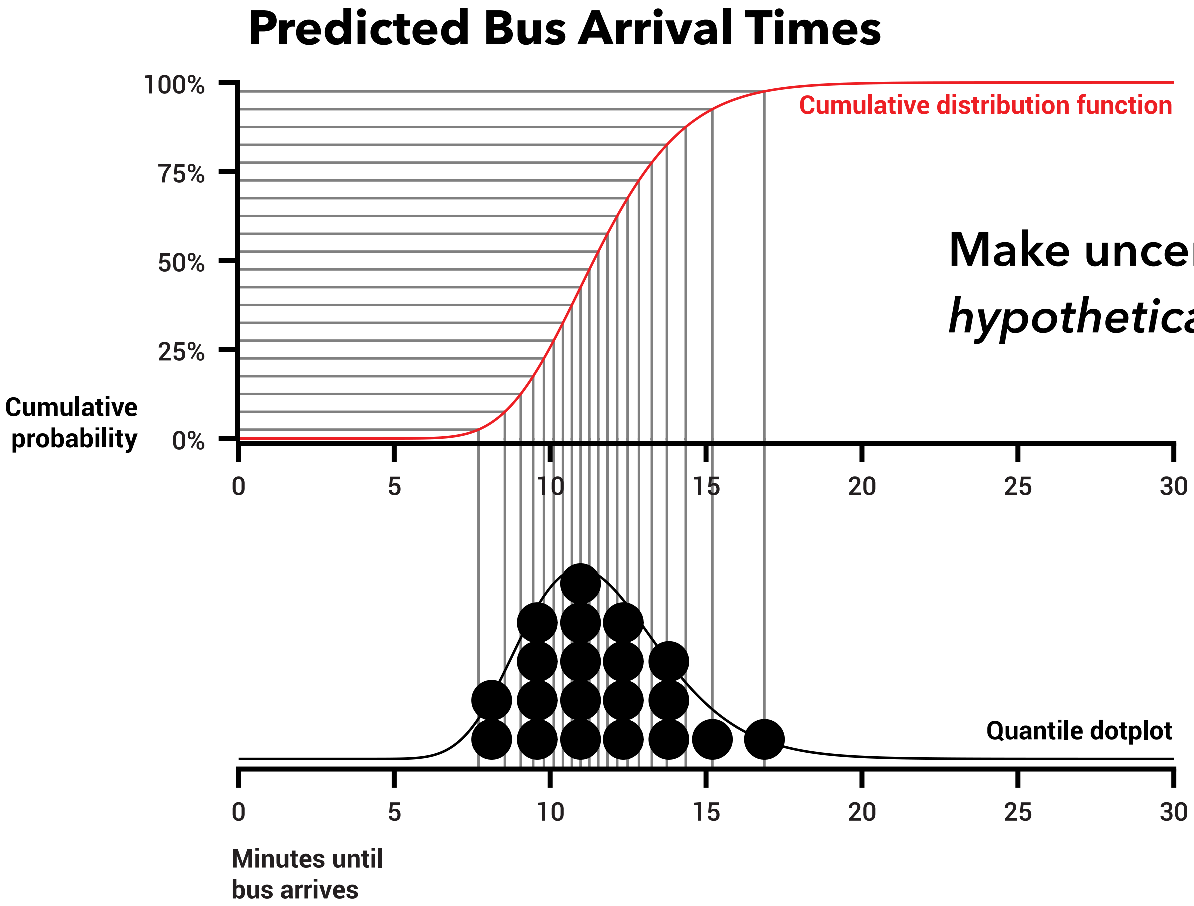




Size or likelihood of hurricane?
Is New Orleans safe?

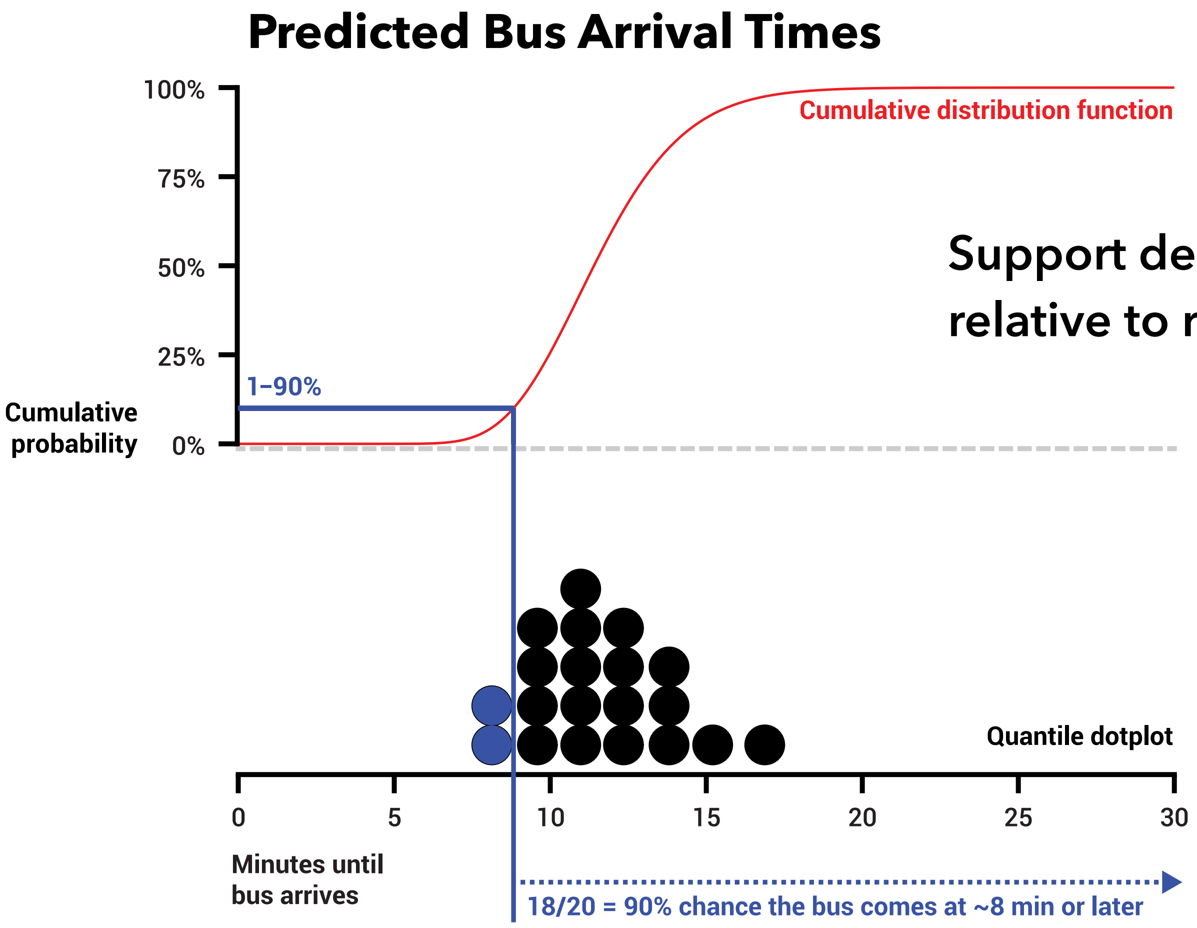
Make uncertainty more concrete
via *hypothetical outcomes*.

[Cox et al. 2013]



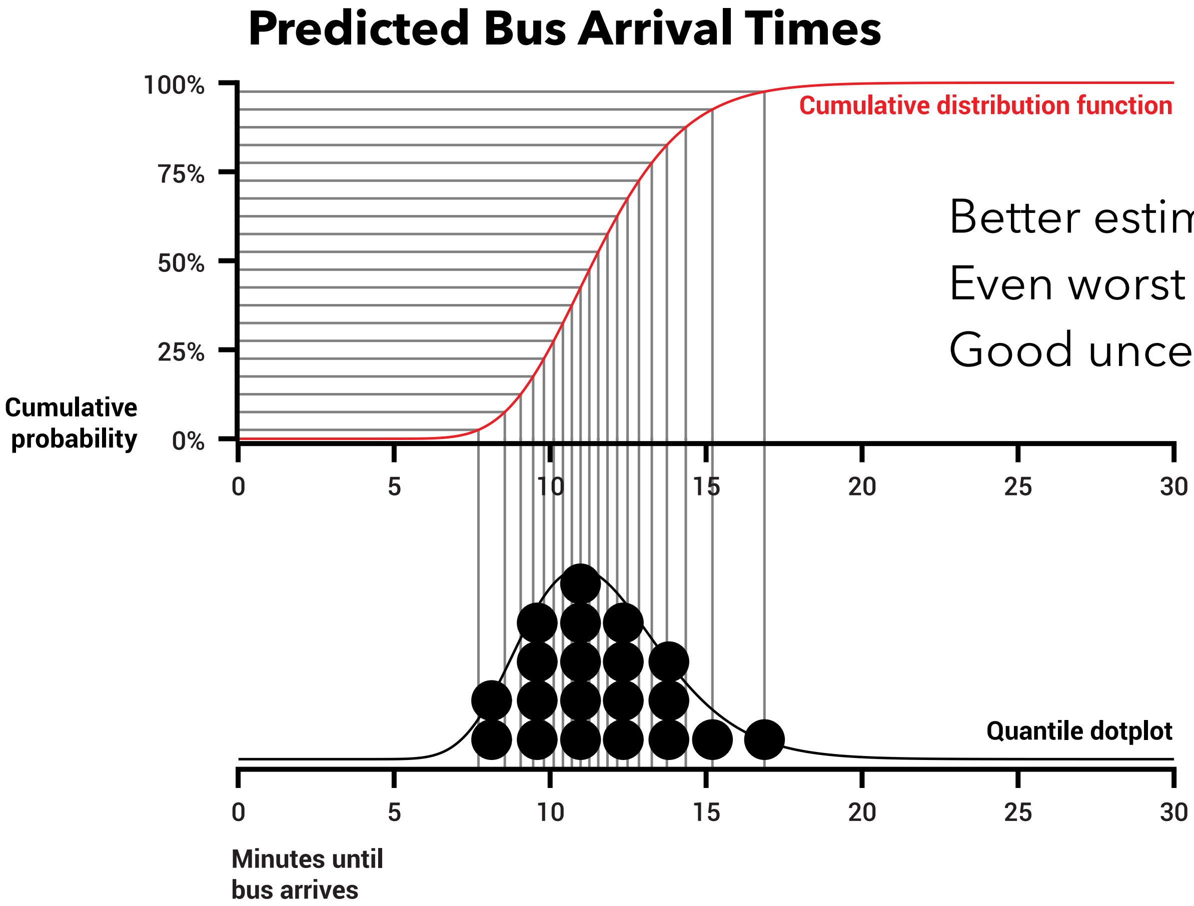
Make uncertainty concrete via *hypothetical outcomes*.

[Kay et al. 2016]



Support decision making relative to risk tolerance.

[Kay et al. 2016]



Better estimates, decisions with time.

Even worst performers improve.

Good uncertainty displays possible!

[Kay et al. 2016]

How to Present Probabilities

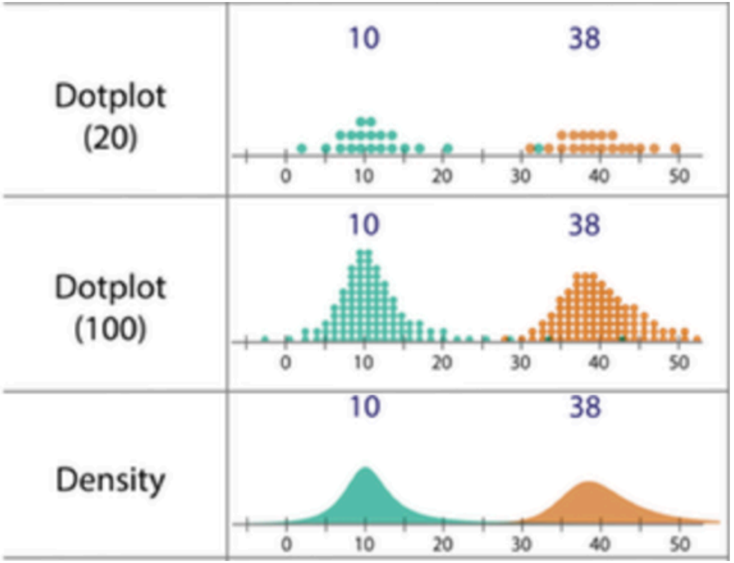
Less Intuitive	**Probability**	$P(A) = 0.6$
	Percentage	60% chance of A
More Intuitive	**Natural Frequency**	3 out of 5 times, A happens.

[Ottley et al. 2016]

Quantile Dot Plots

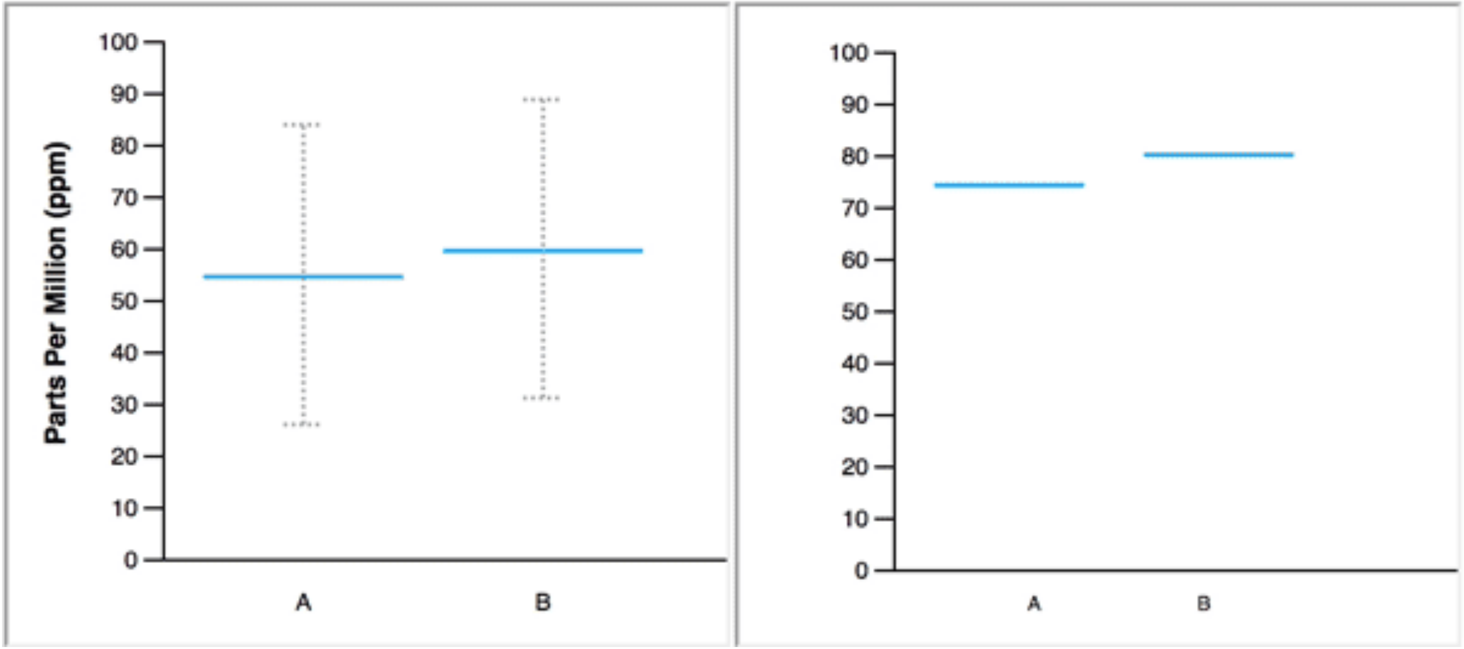
Less Error

More Error



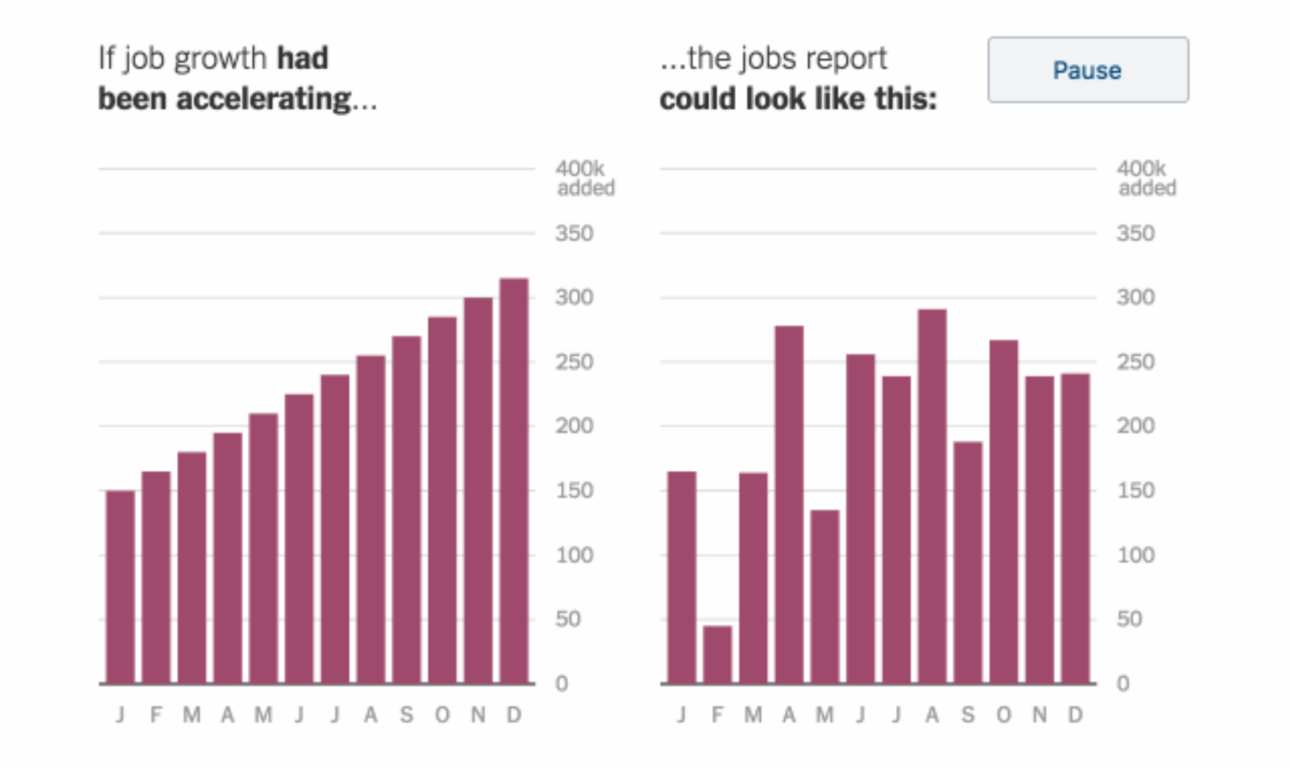
[Kay et al. 2016]

Hypothetical Outcome Plots



[Hullman et al.]

Hypothetical Outcome Plots



[NYTimes, 2014]

Hypothetical Outcomes

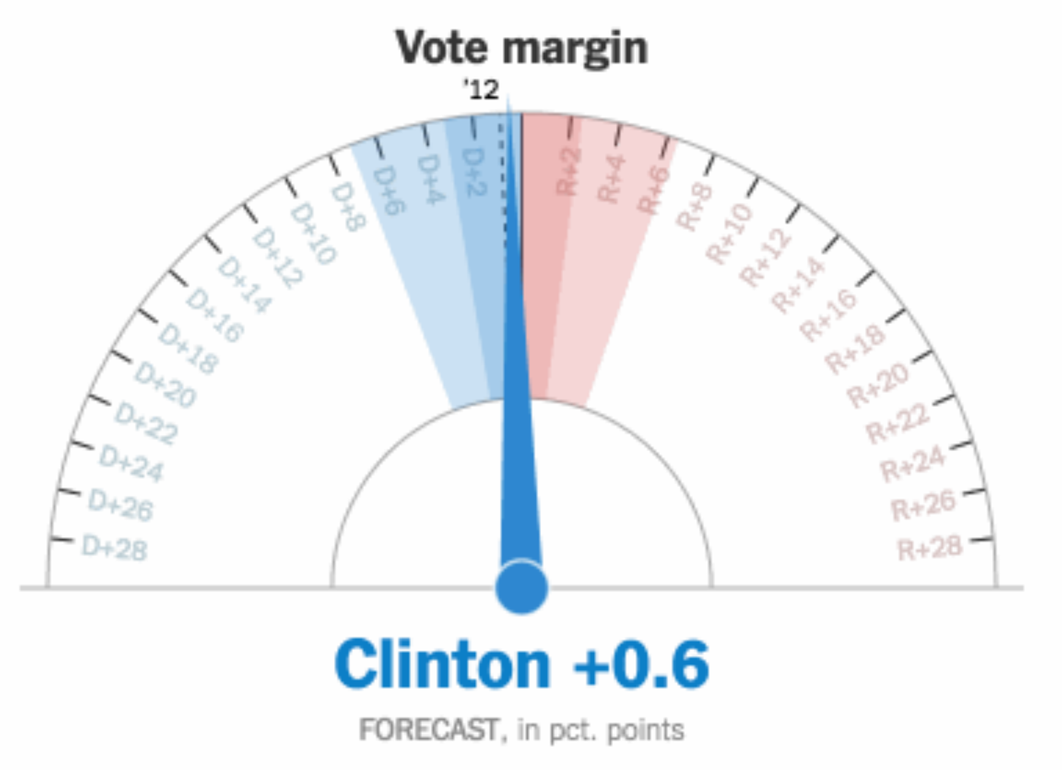
If the economy actually added 150,000 jobs last month, it would be possible to see any of these headlines:

The jobs number is just an estimate, and it comes with uncertainty.

Under 55,000 jobs	55,000 to 110,000	110,000 to 140,000	160,000 to 190,000	190,000 to 245,000	245,000+
4% chance	19% chance	19% chance	19% chance	19% chance	4% chance

[NYTimes, 2014]

The NY Times Needle



[NYTimes, 2016]

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?

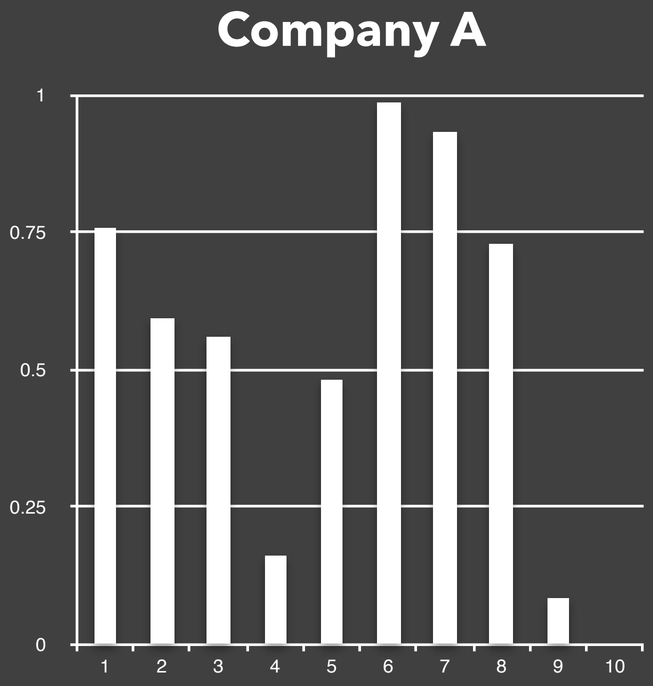
Choose an appropriate visual variable based on the domain, literacy, and expertise of your audience. Be mindful that any display of uncertainty inherently increases the complexity of your visualization, and that there is a preference/performance gap.

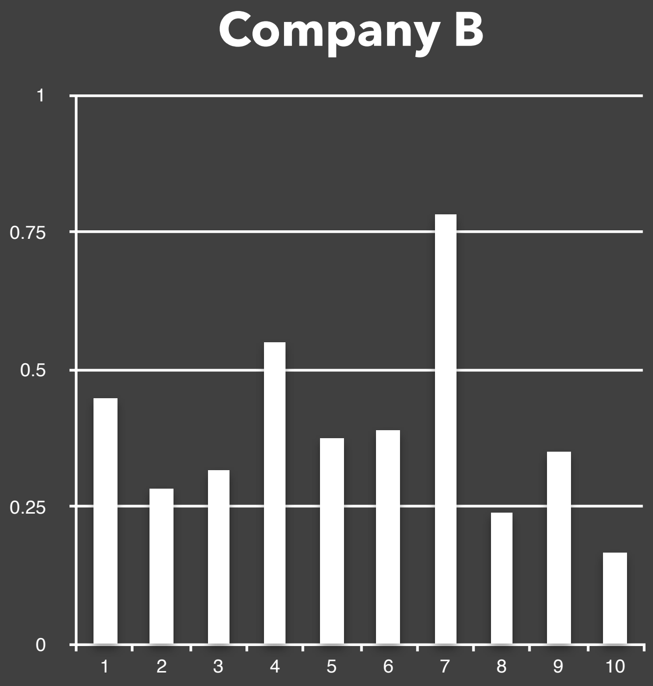
IT DEPENDS

What Can Go Wrong?

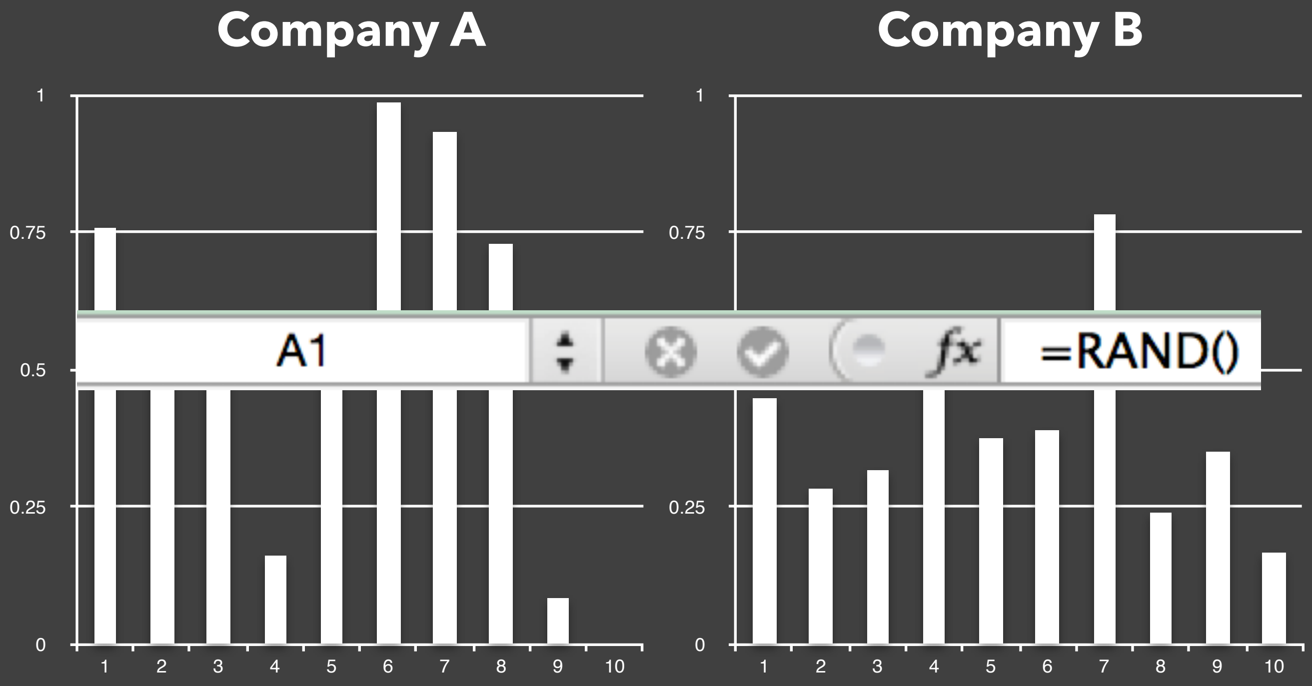
Inferential Integrity

Which Stock To Buy?



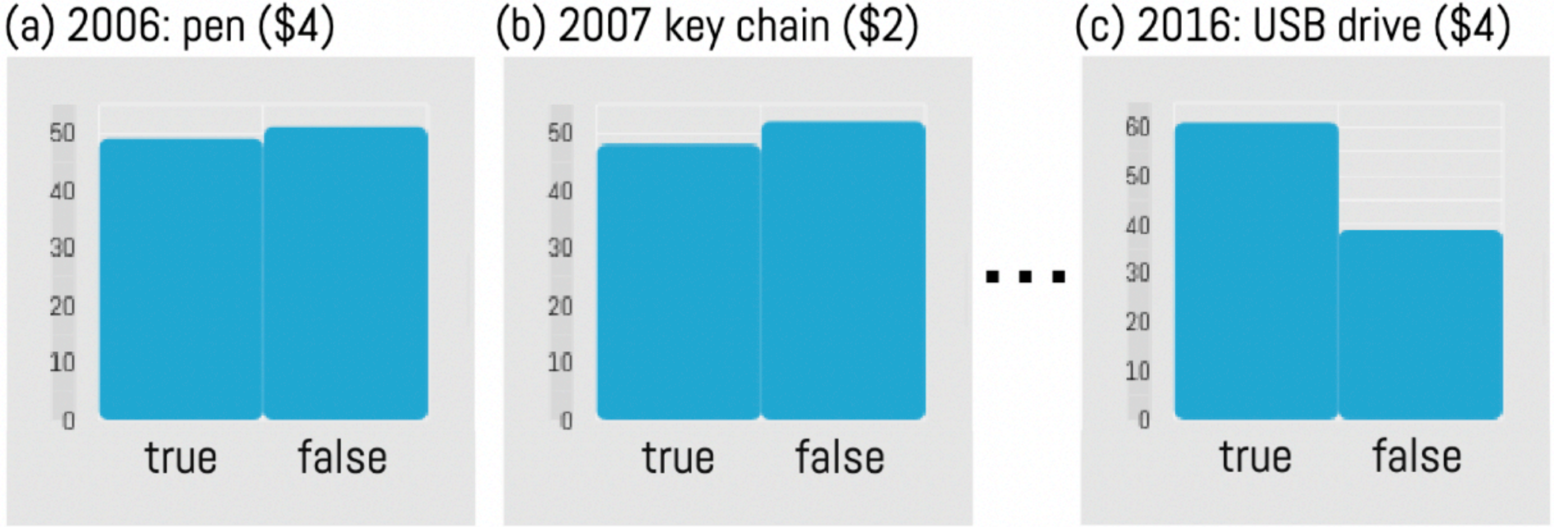


Neither!



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

Recurring Donor?



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

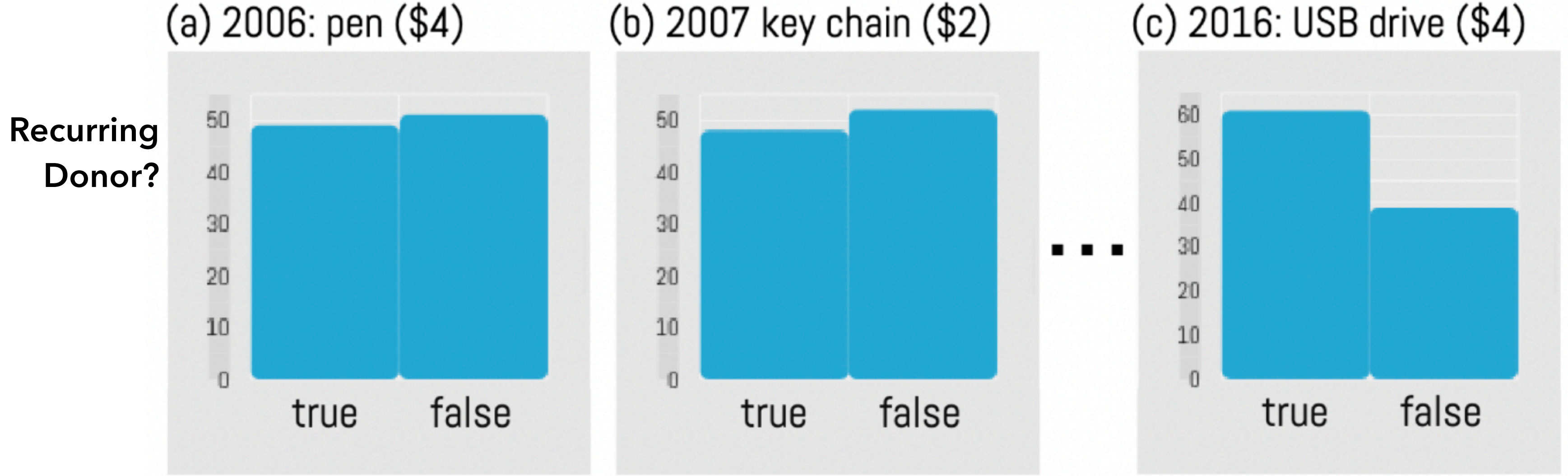
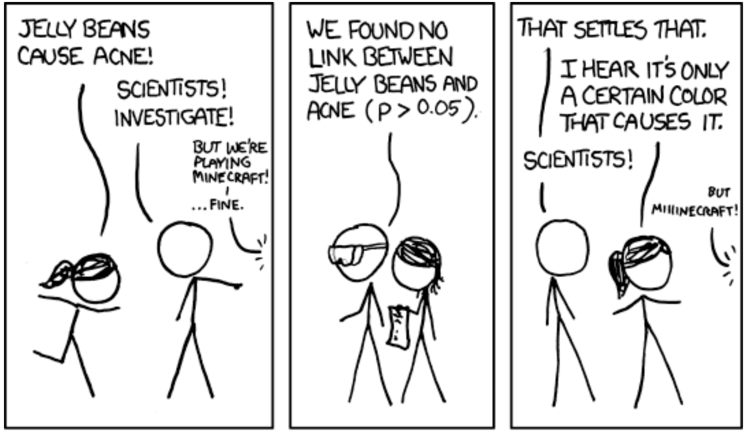
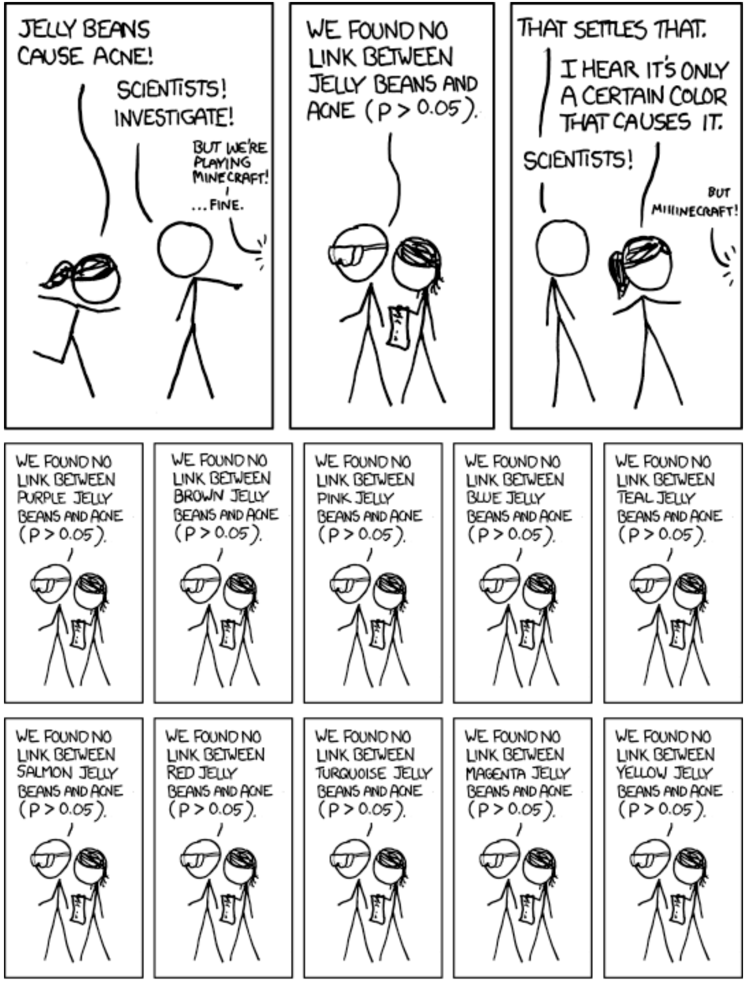
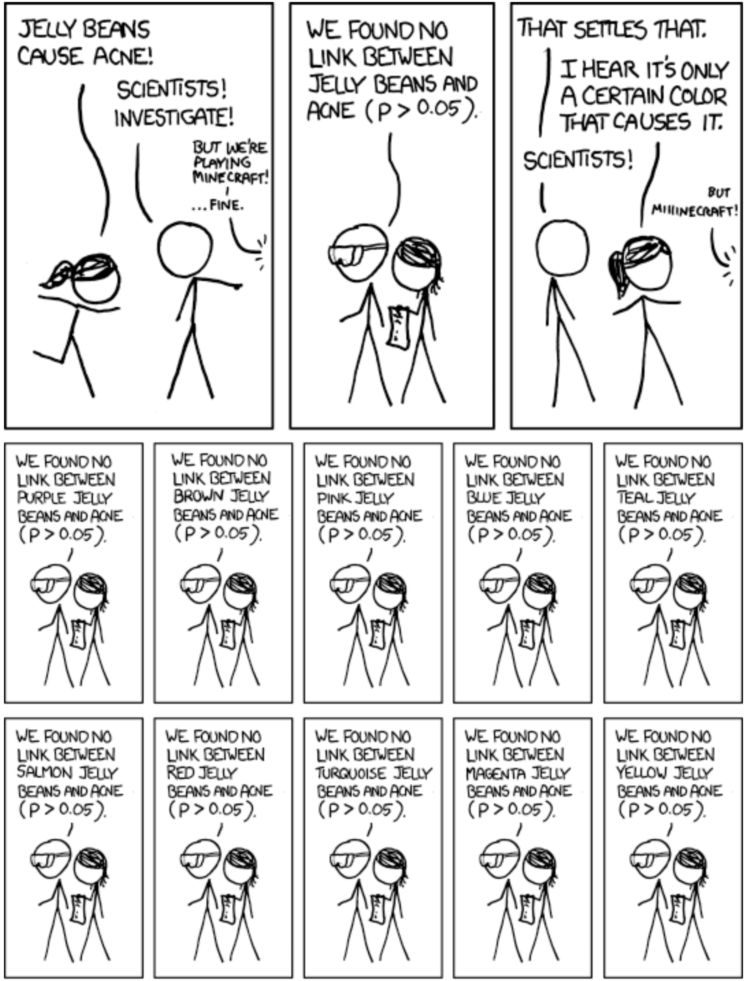


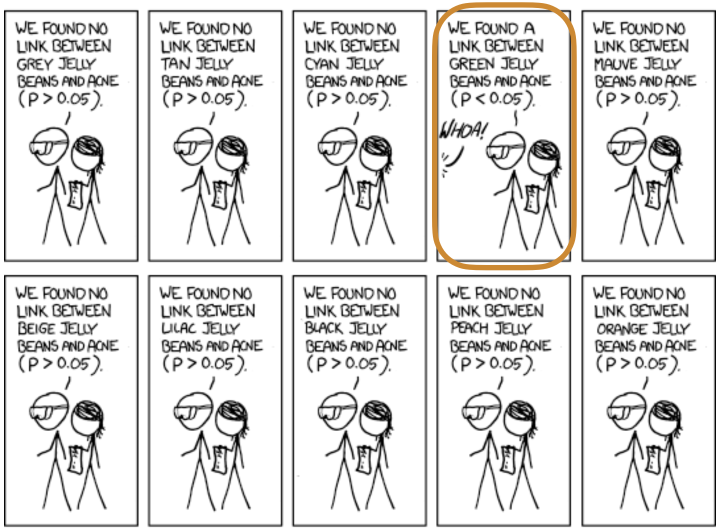
Figure 1. A user inspects several graphs and wrongly flags (c) as an insight because it looks different than (a) and (b). All were generated from the same uniform distribution and are the "same". By viewing lots of visualizations, the chances increase of seeing an apparent insight that is actually the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

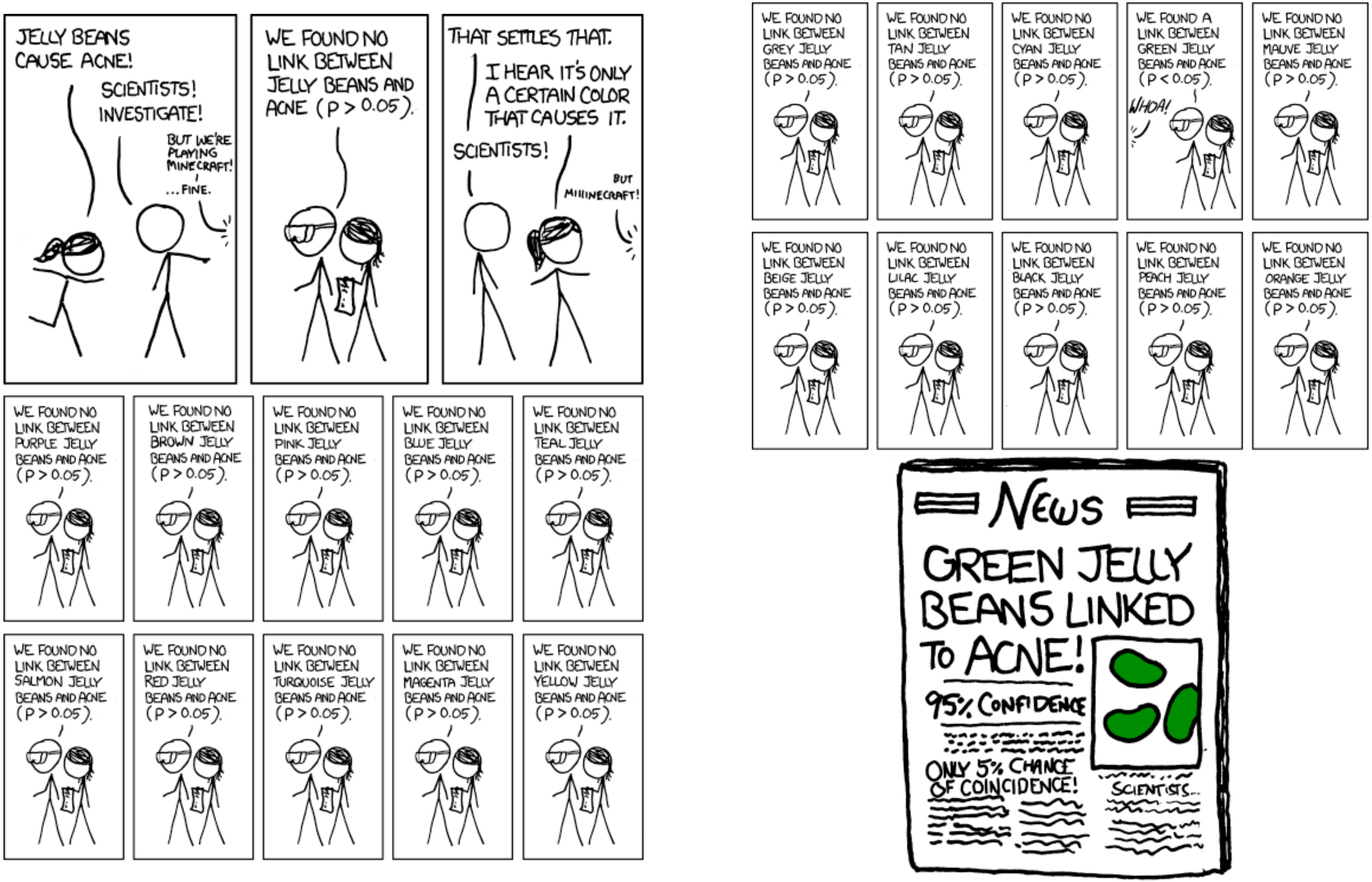

JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
...FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!


JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
...FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).

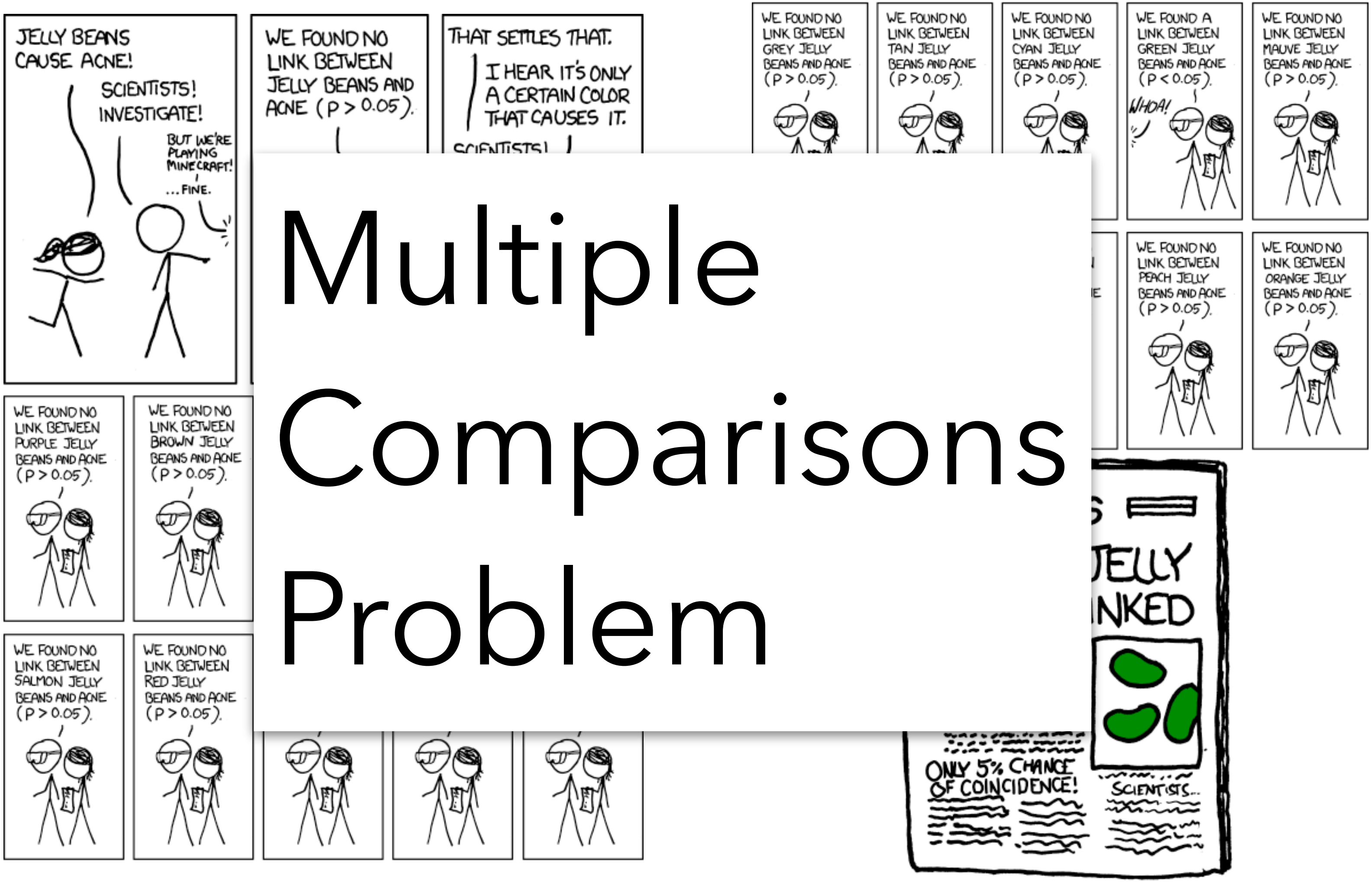

JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
...FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).


JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
...FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).

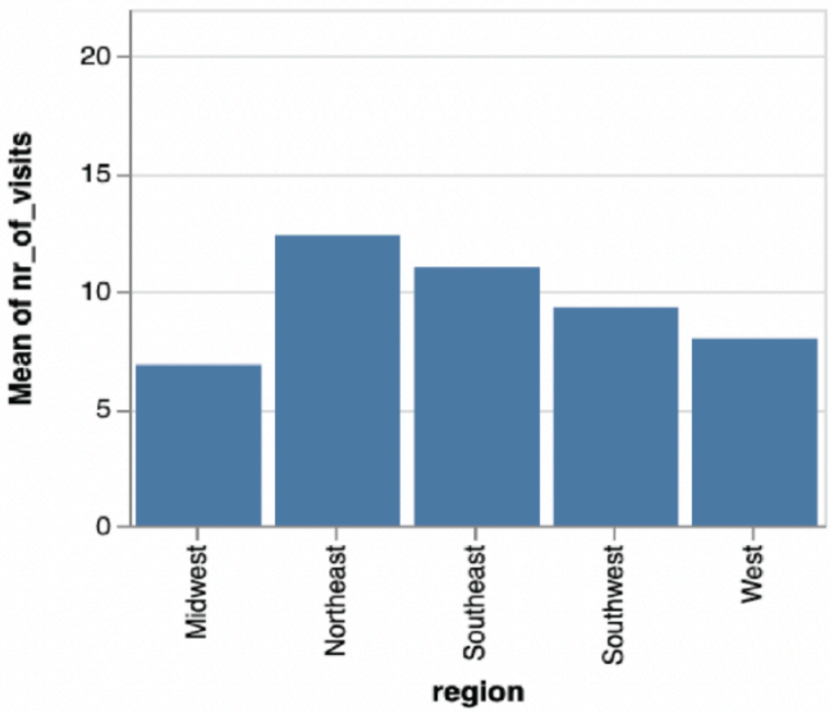

WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).

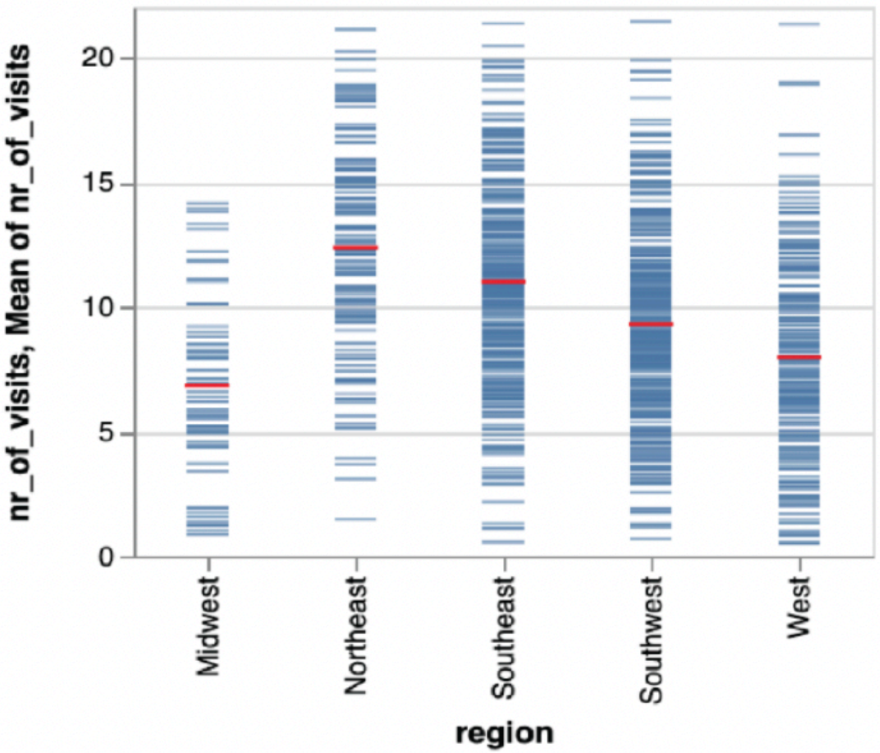

JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
...FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).
News
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

Multiple Comparisons Problem

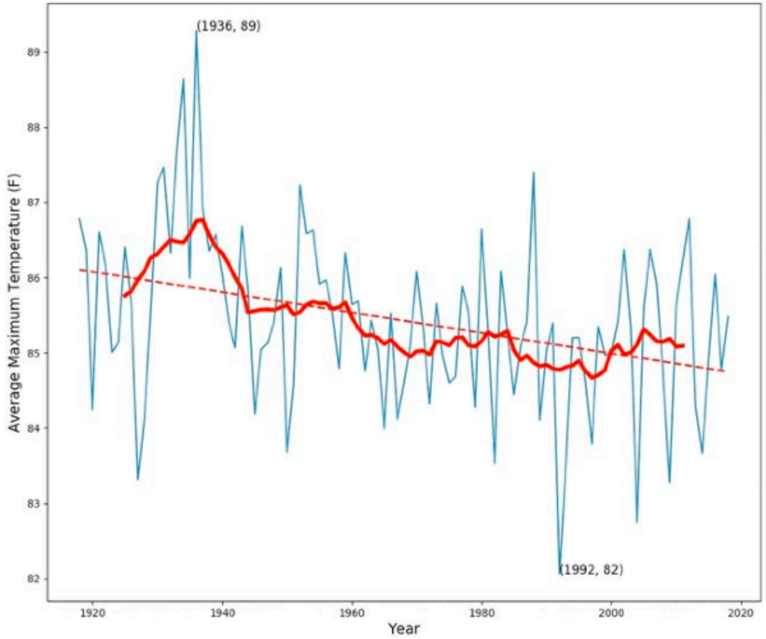


Aggregated vs. Disaggregated Views [Nguyen et al. '20]

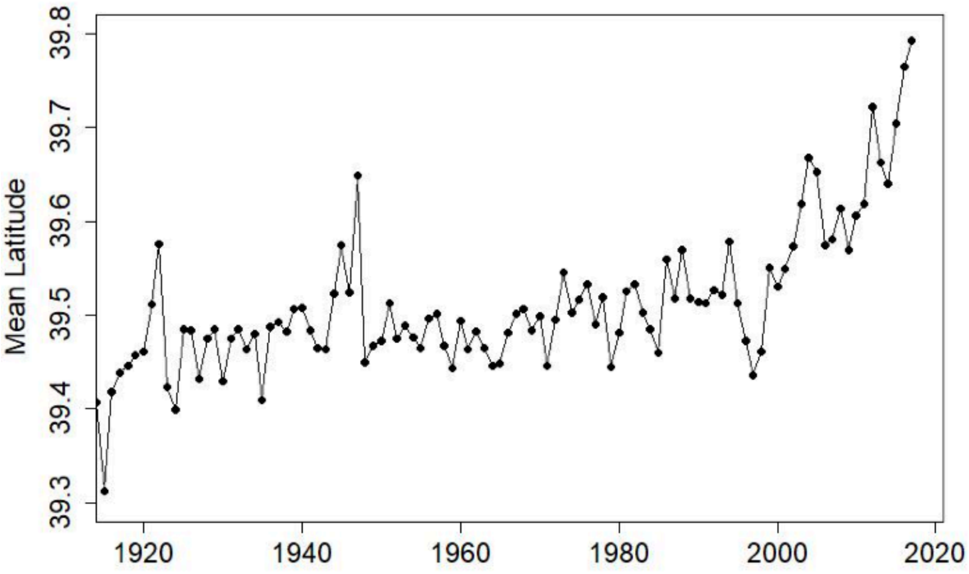




Example: Is the U.S. cooling?



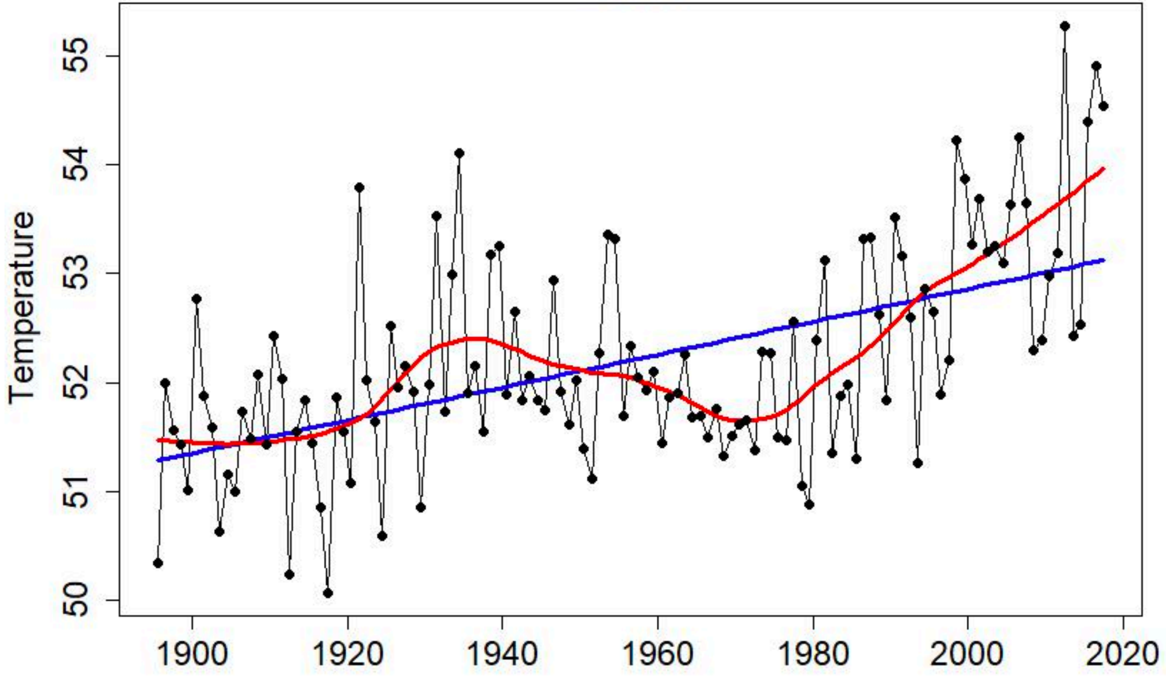
Starts at 1918. Summer temps only.
Raw average over weather stations.



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

tamino.wordpress.com/2018/08/08/usa-temperature-can-i-sucker-you/

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

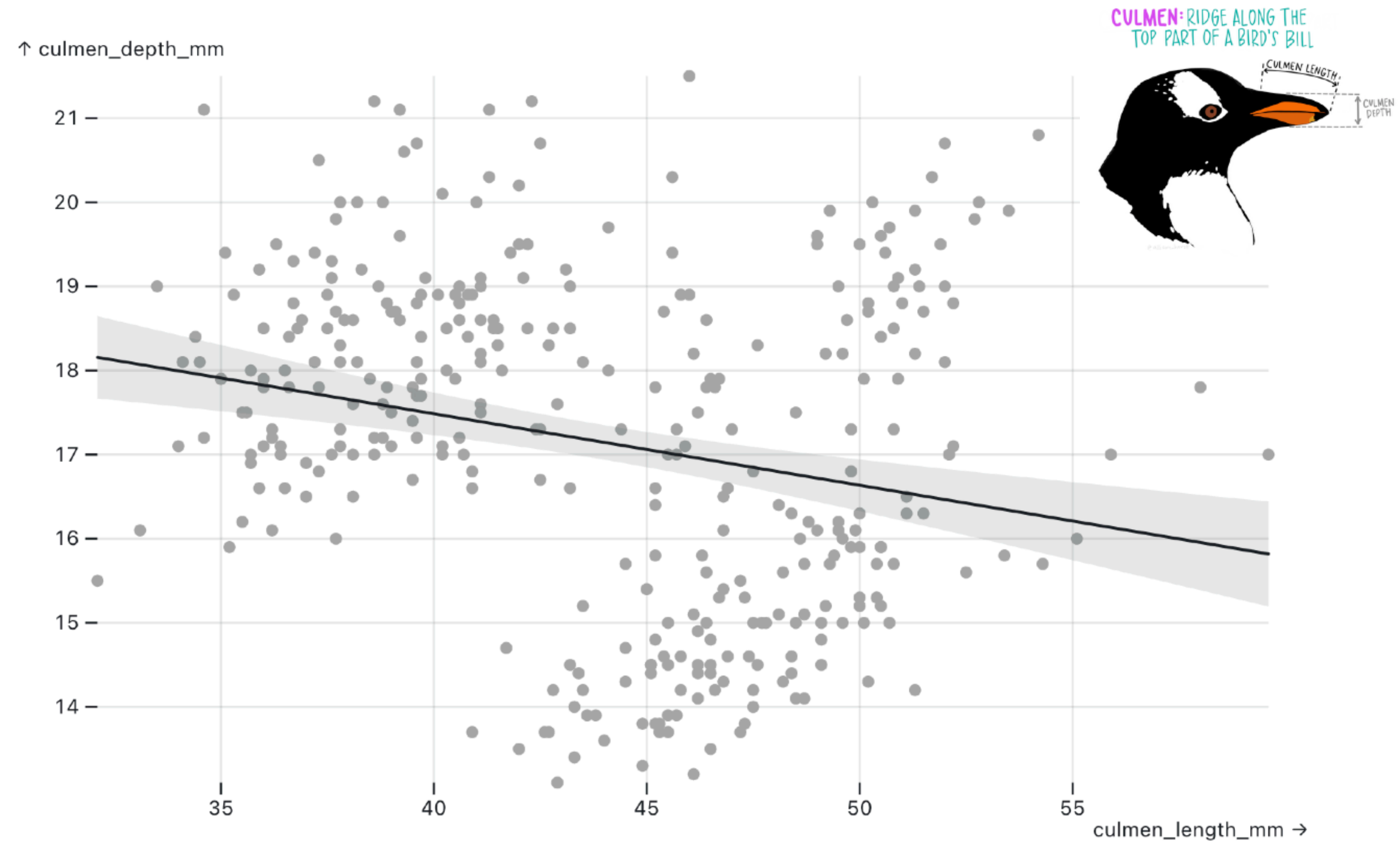


Include more historical data.

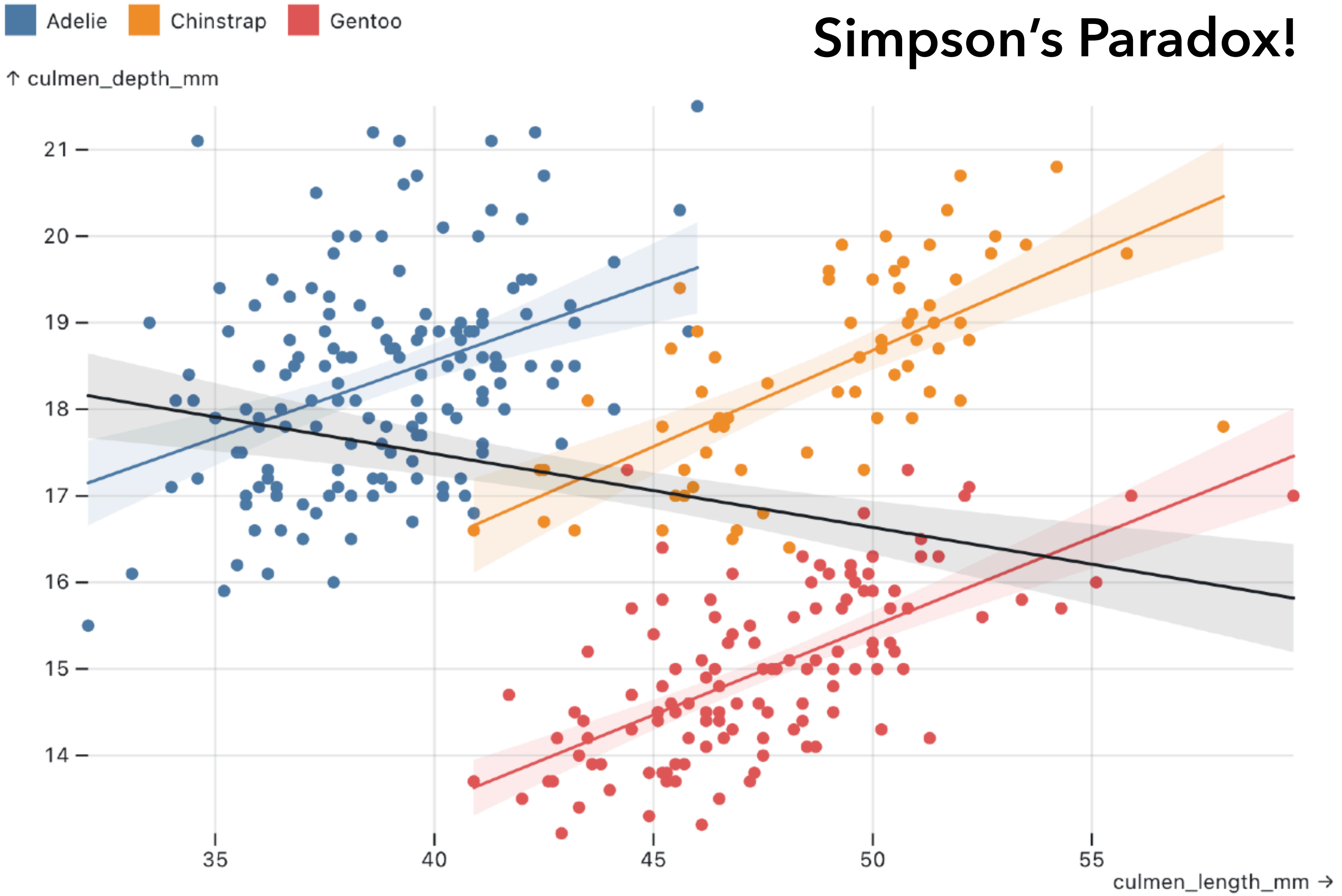
Include all four seasons.

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

tamino.wordpress.com/2018/08/08/usa-temperature-can-i-sucker-you/


↑ culmen_depth_mm
21
20
19
18
17
16
15
14
35
40
45
50
55
culmen_length_mm →
CULMEN: RIDGE ALONG THE TOP PART OF A BIRD'S BILL
CULMEN LENGTH
CULMEN DEPTH

Simpson's Paradox!



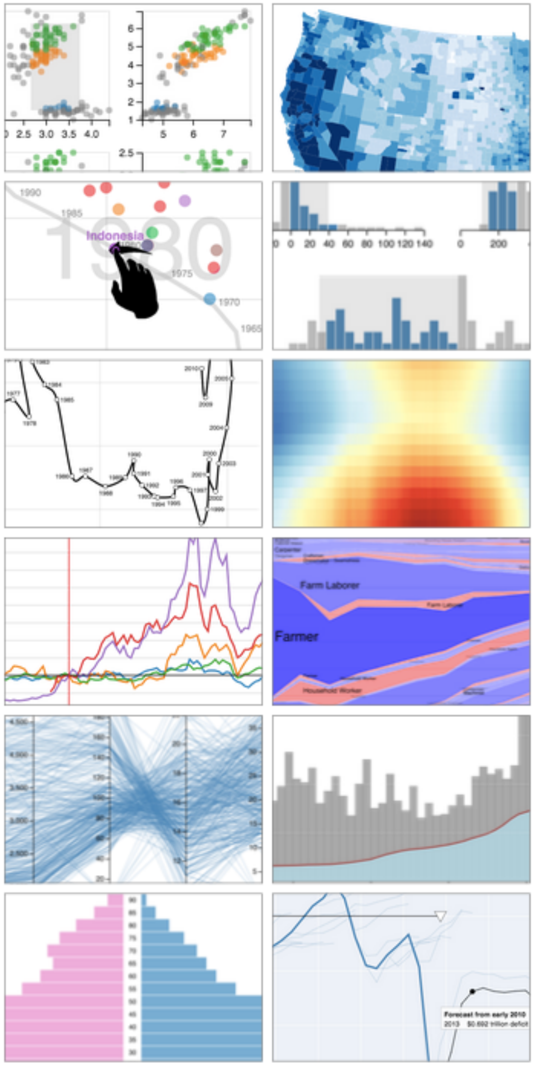
Some Causes of Inferential Failure

Premature Suppression of Uncertainty

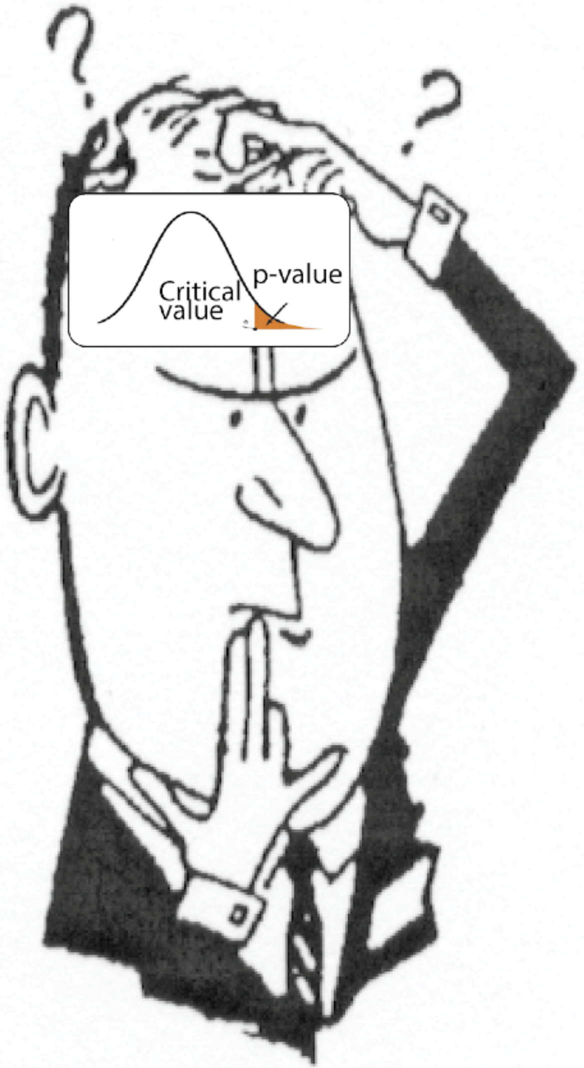
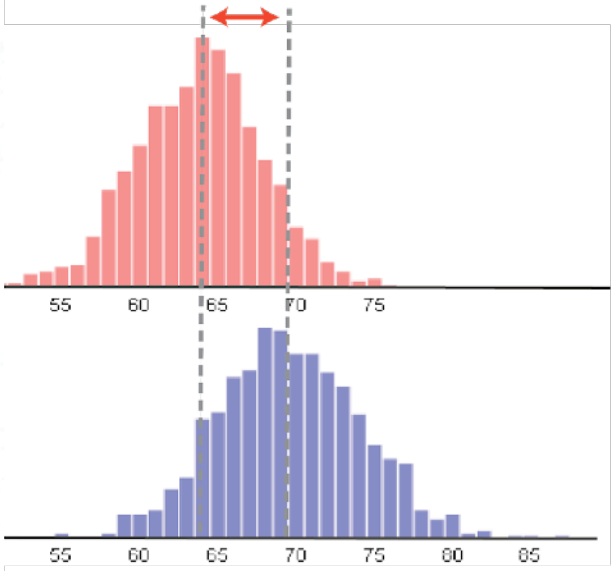
False Discovery due to Random Fluctuation

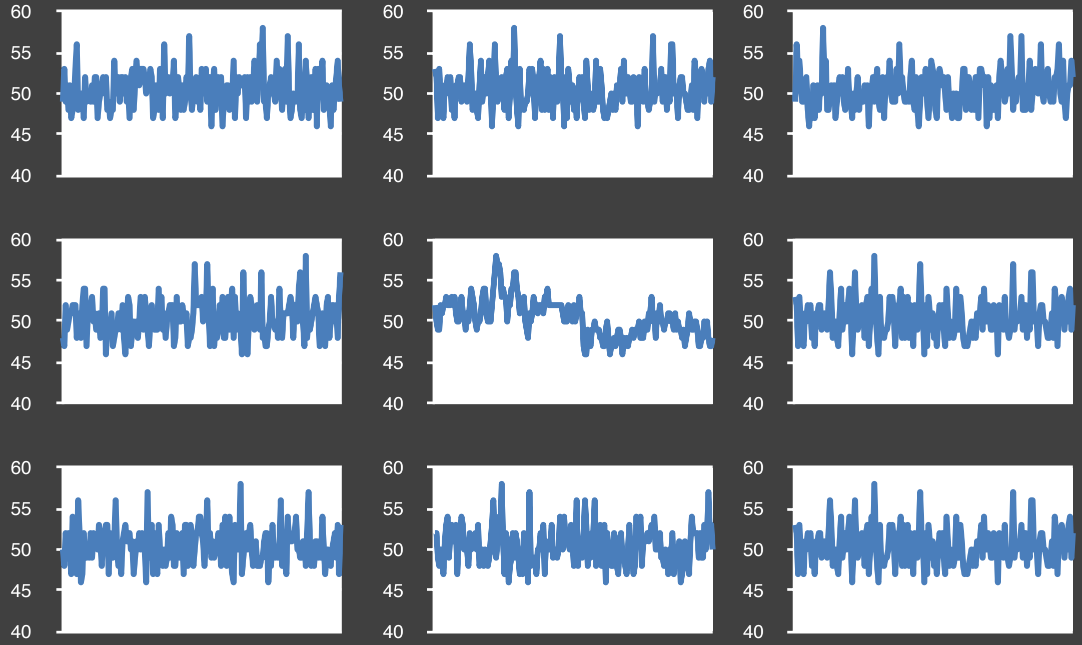
Incomplete or Biased Data

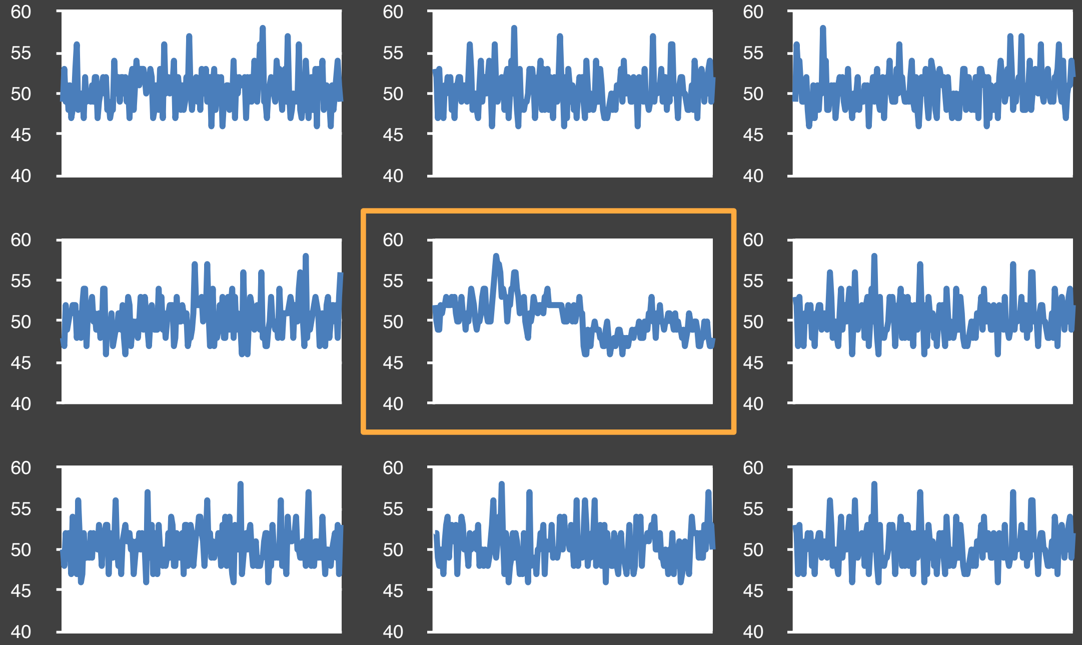
Confounding Variables



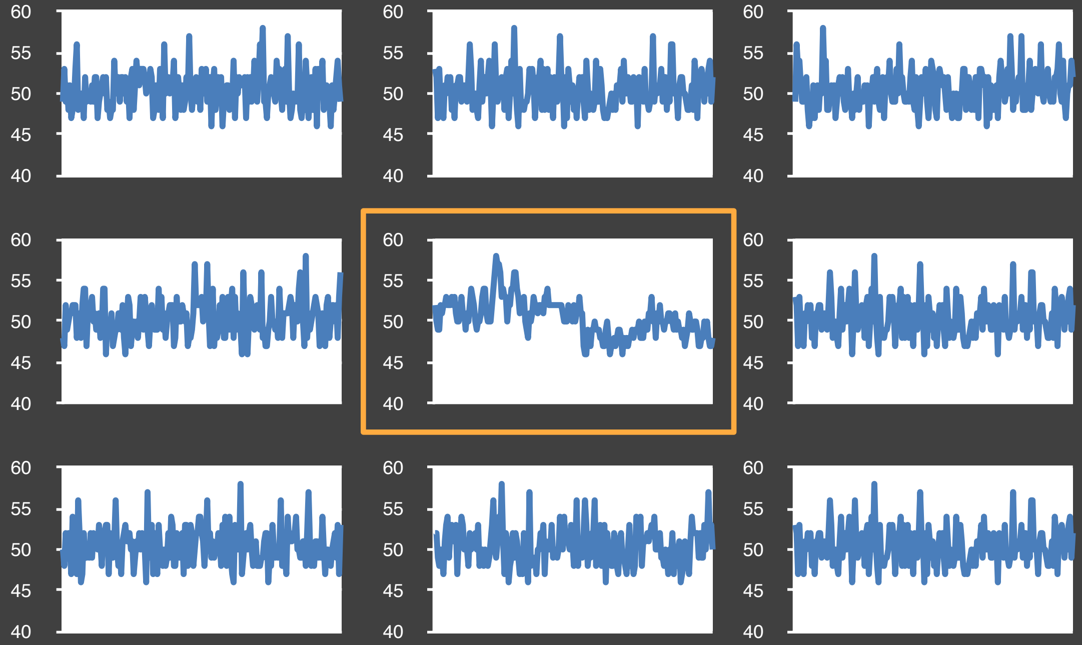
Graphical Inference & Model Checks

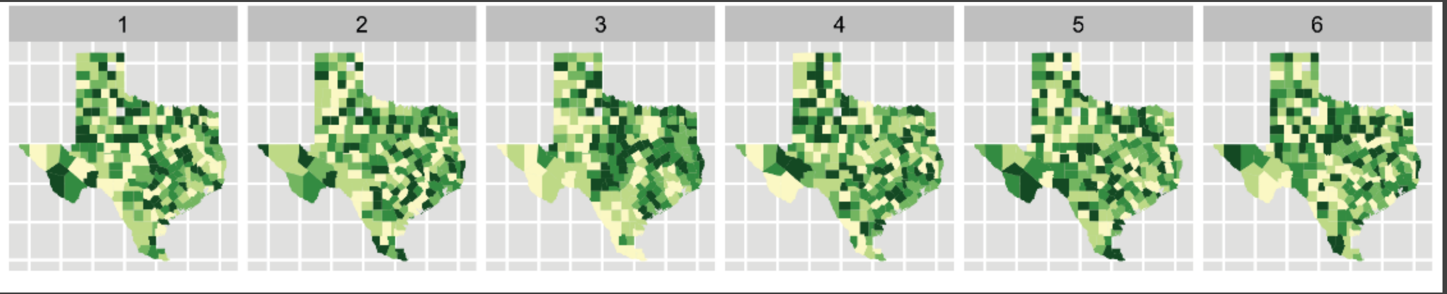

p-value
Critical
value

55
60
65
70
75
55
60
65
70
75
80
85


60
55
50
45
40
60
55
50
45
40
60
55
50
45
40
60
55
50
45
40
60
55
50
45
40
60
55
50
45
40
60
55
50
45
40
60
55
50
45
40
60
55
50
45
40


60
55
50
45
40

Visual Lineups



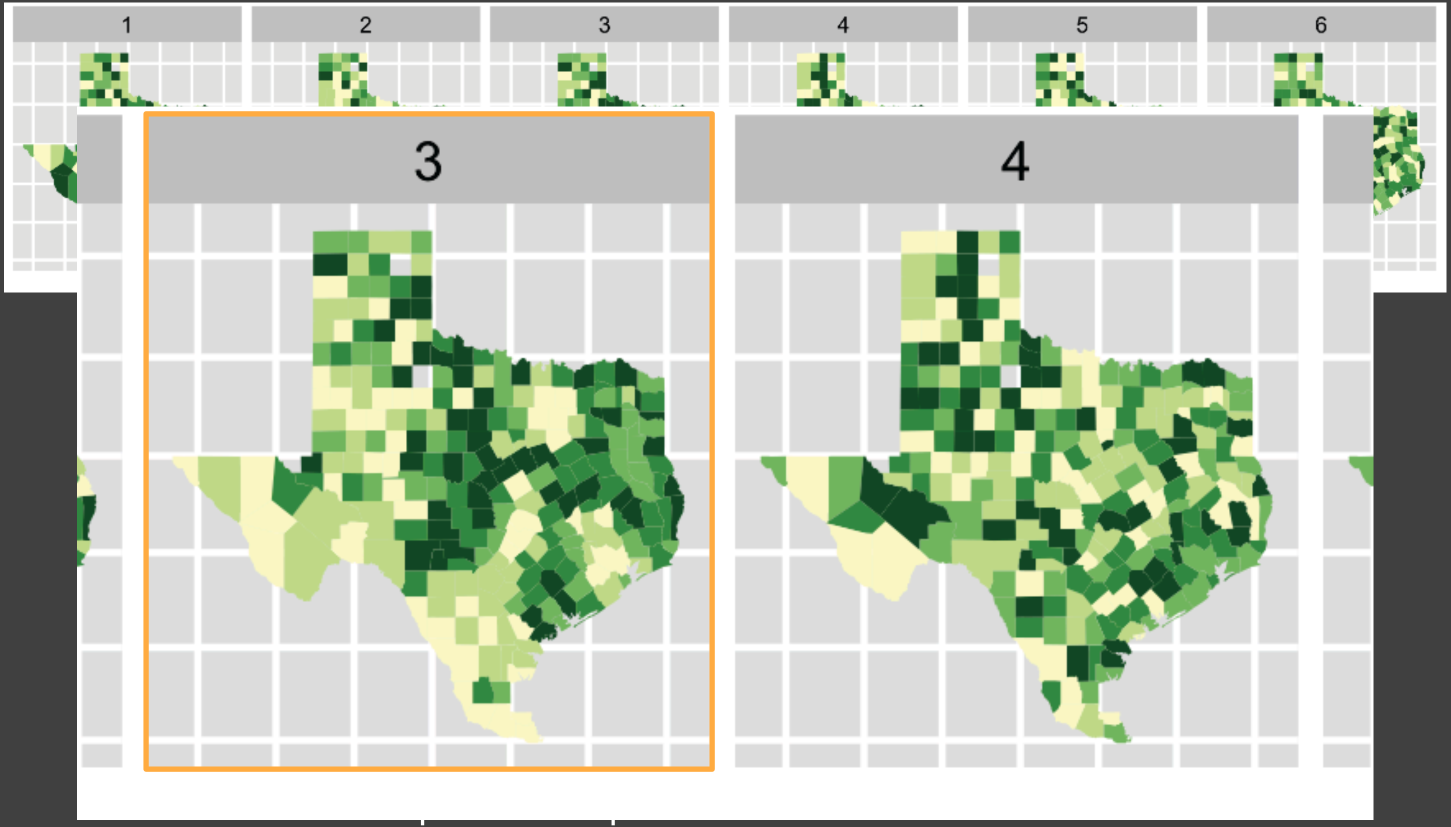


Choropleth maps of cancer deaths in Texas.

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.

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


1
2
3
4
5
6
3
4

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

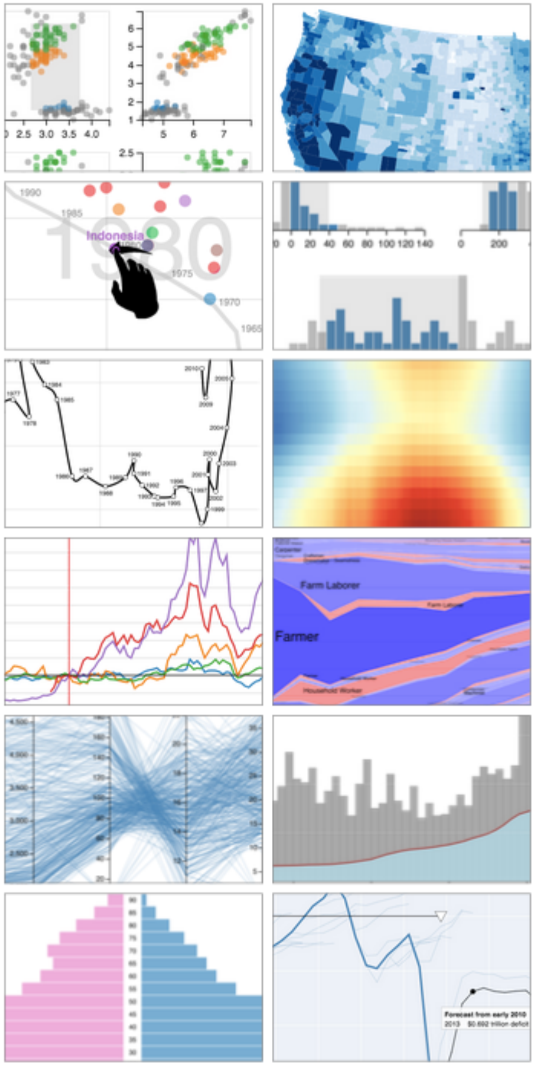
Permute (shuffle) relationship between variables, or

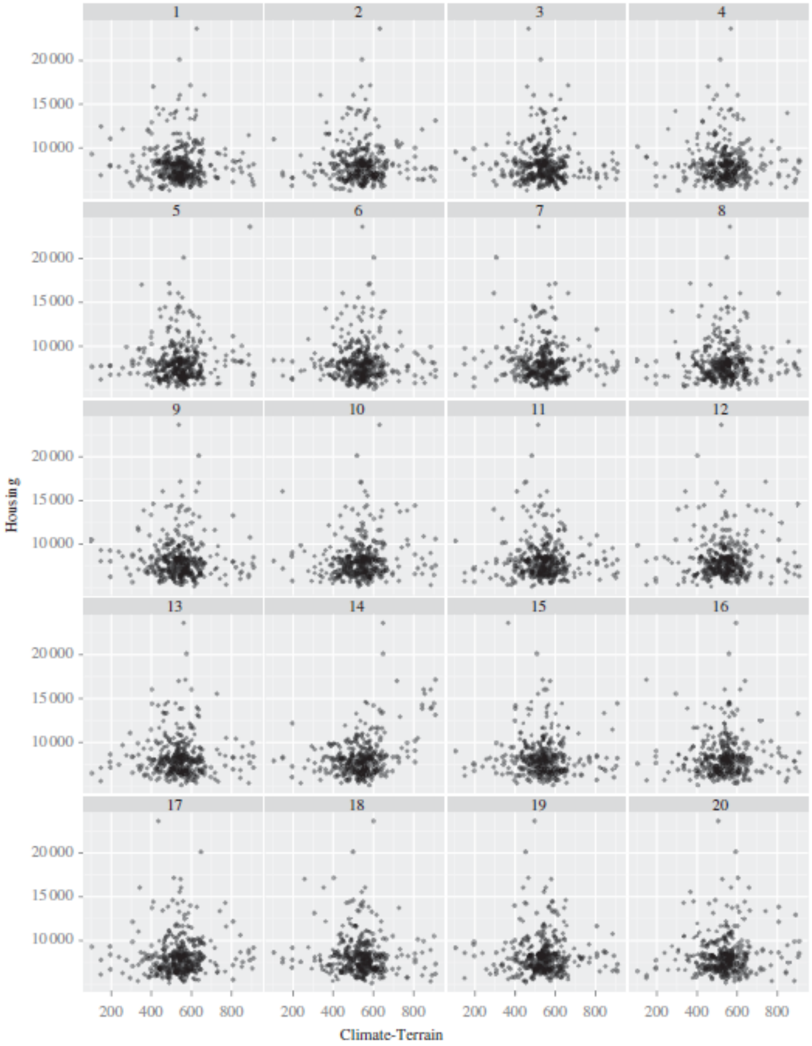
Choose a meaningful “null” model

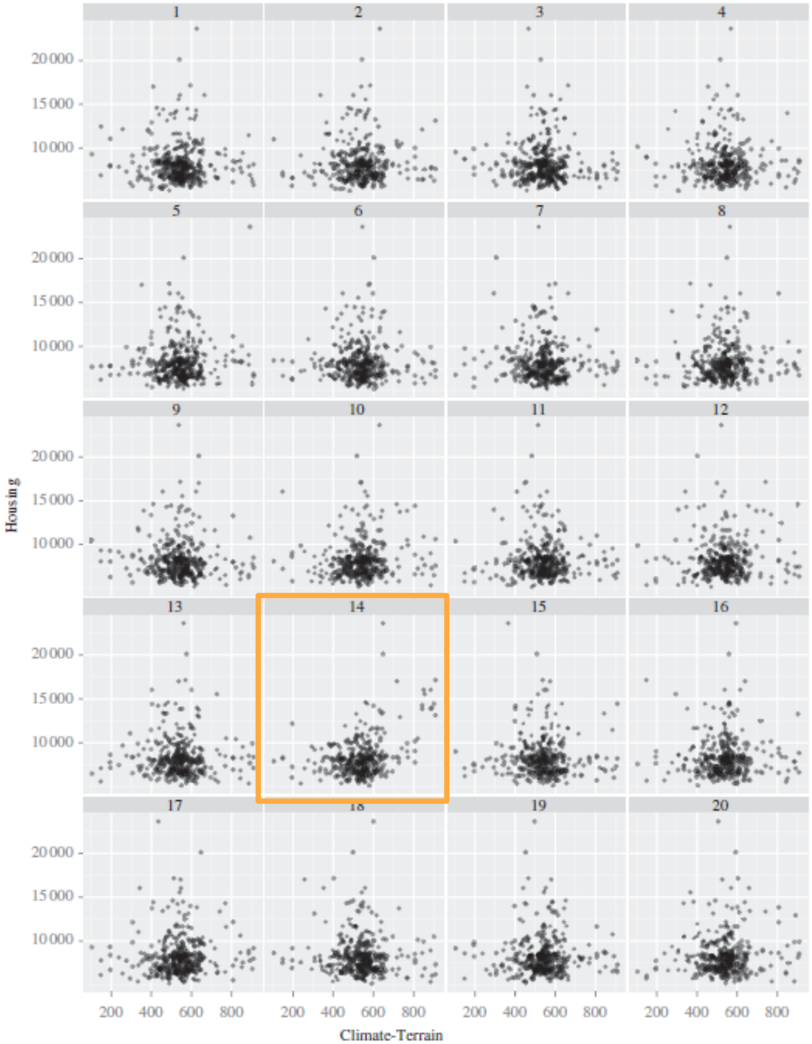
Perform visual comparison

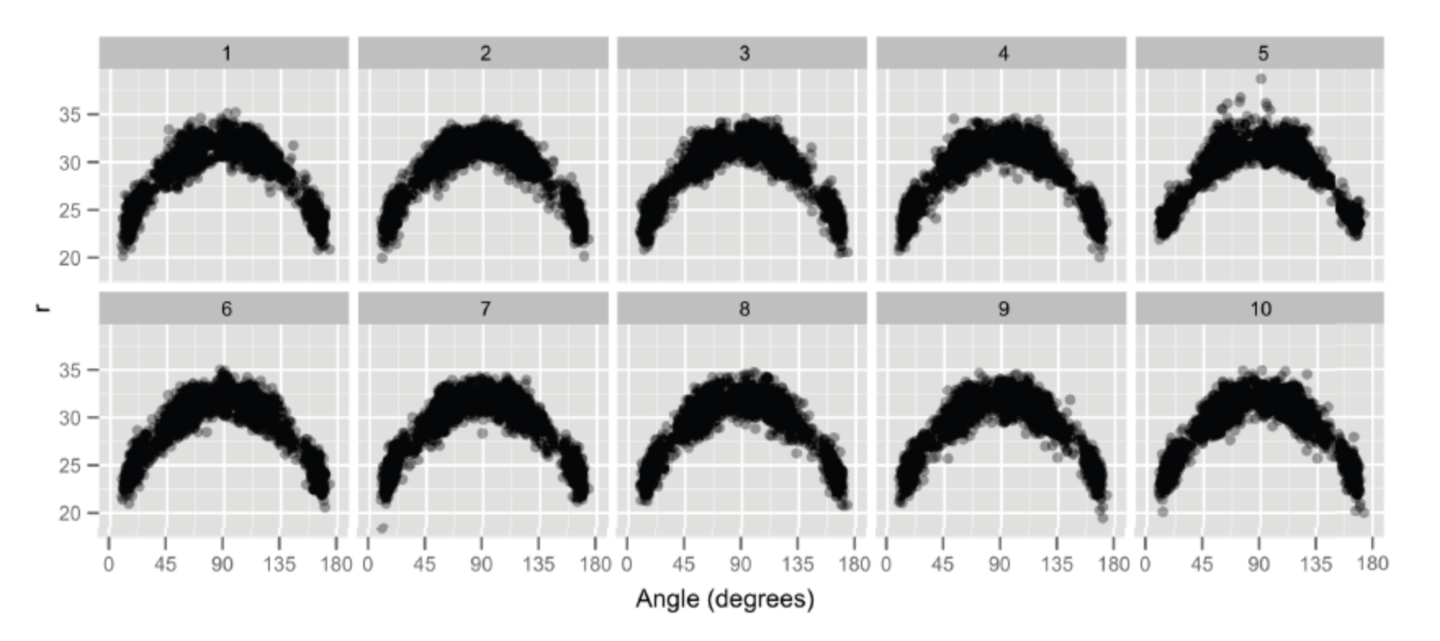
In the “lineup” protocol, we compare the real data against a number of generated variants.

Can we spot the difference?



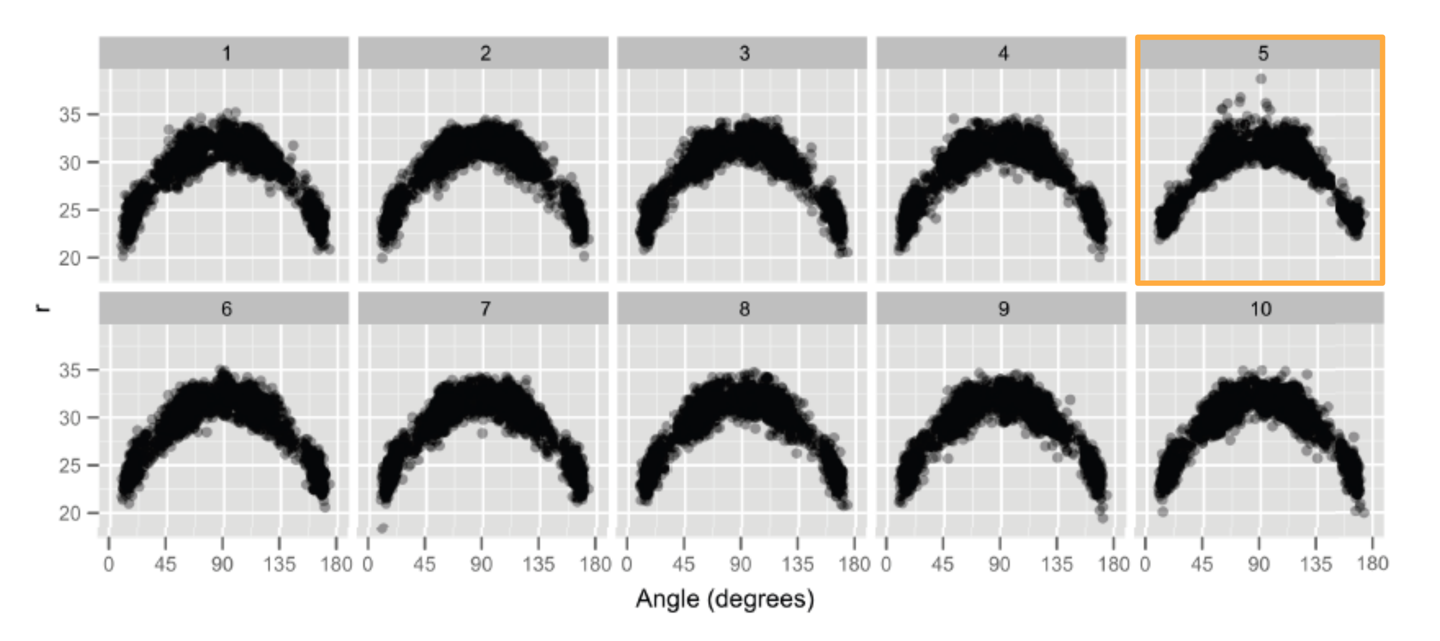

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Housing
20 000
15 000
10 000
200
400
600
800
Climate-Terrain


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
20 000
15 000
10 000
Housing
200
400
600
800
Climate-Terrain



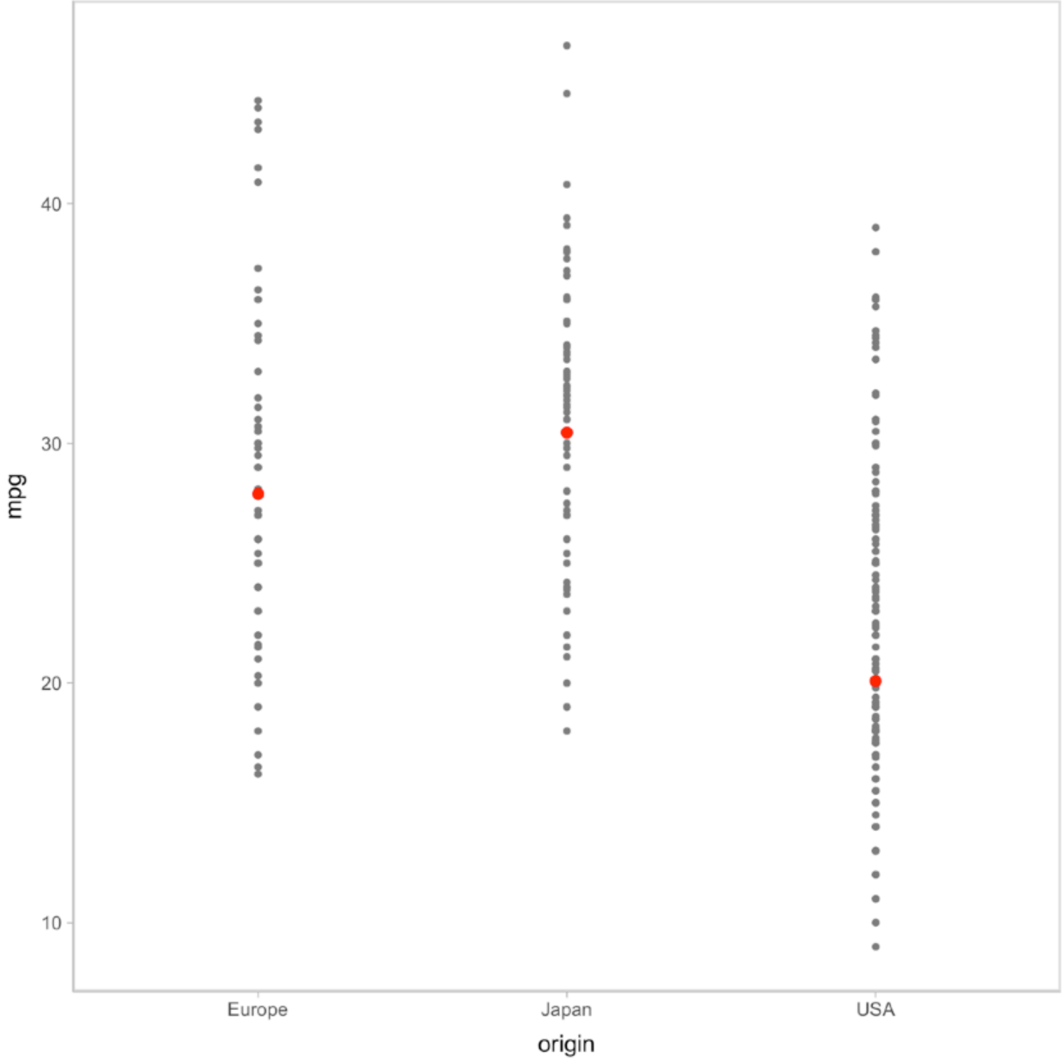
Distance vs. angle for 3 point shots by the LA Lakers.

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.

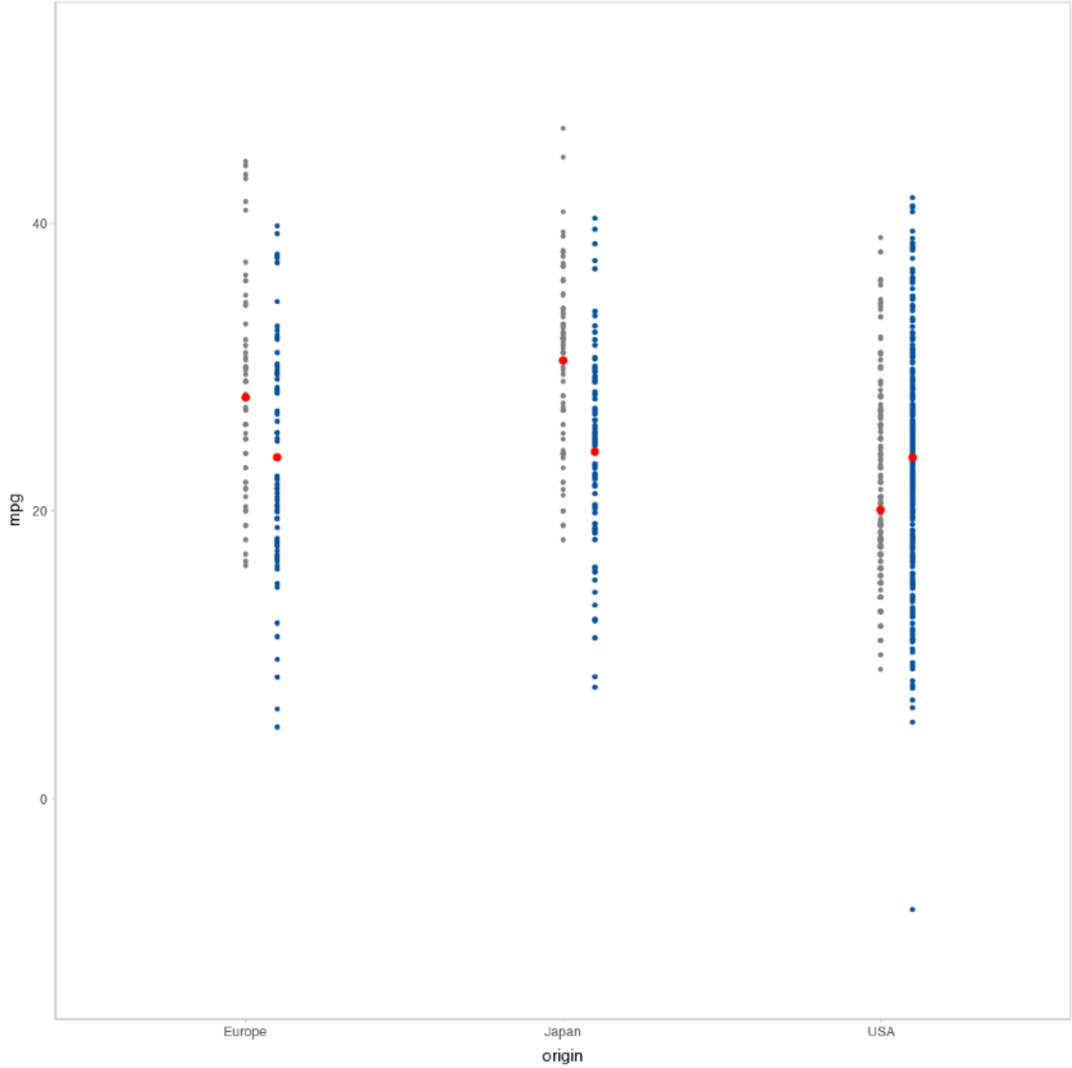
One plot is the real data. The others are generated according to a null hypothesis of quadratic relationship.



Plot:

mpg by origin

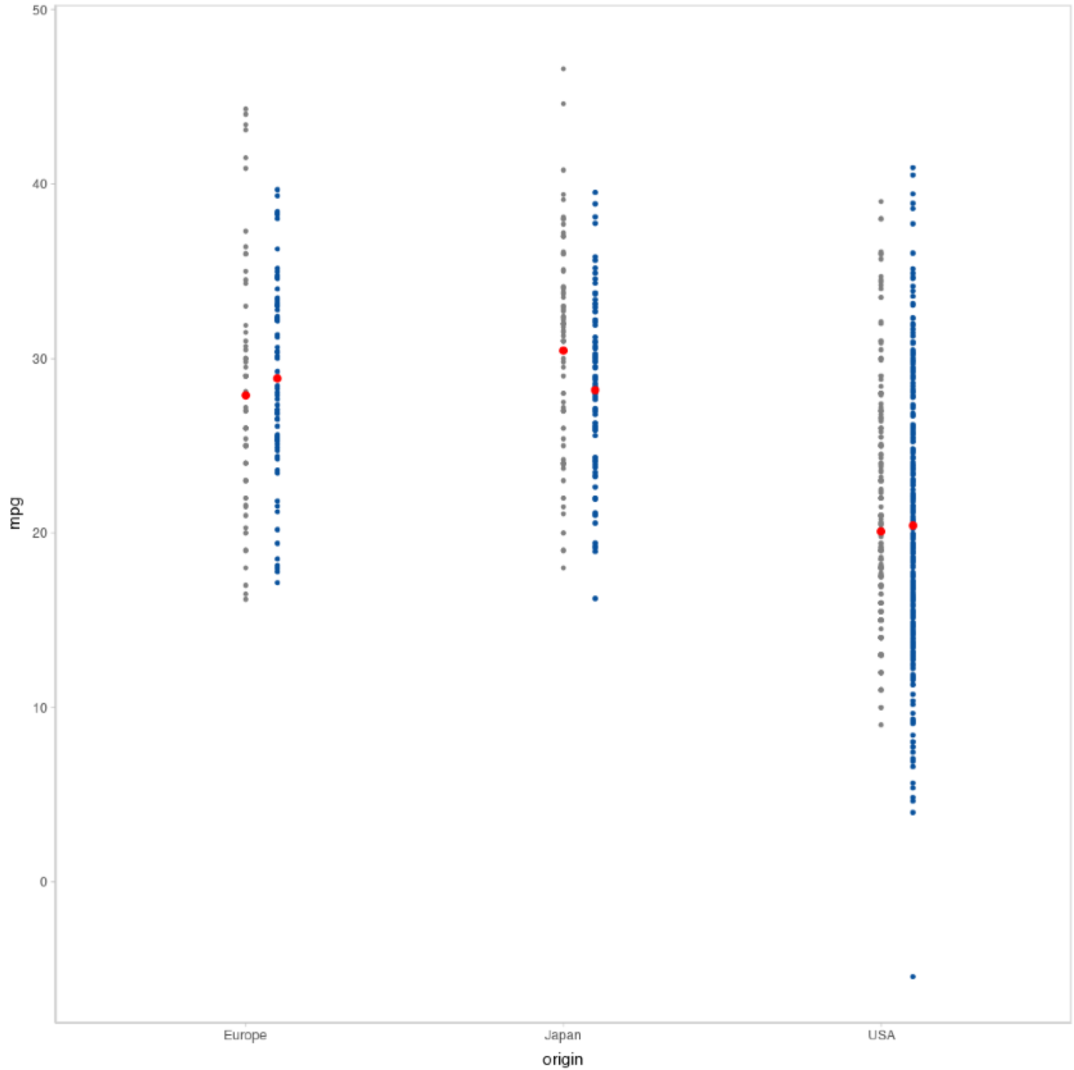
What might our implicit model be?



Model:

mpg ~ 1

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

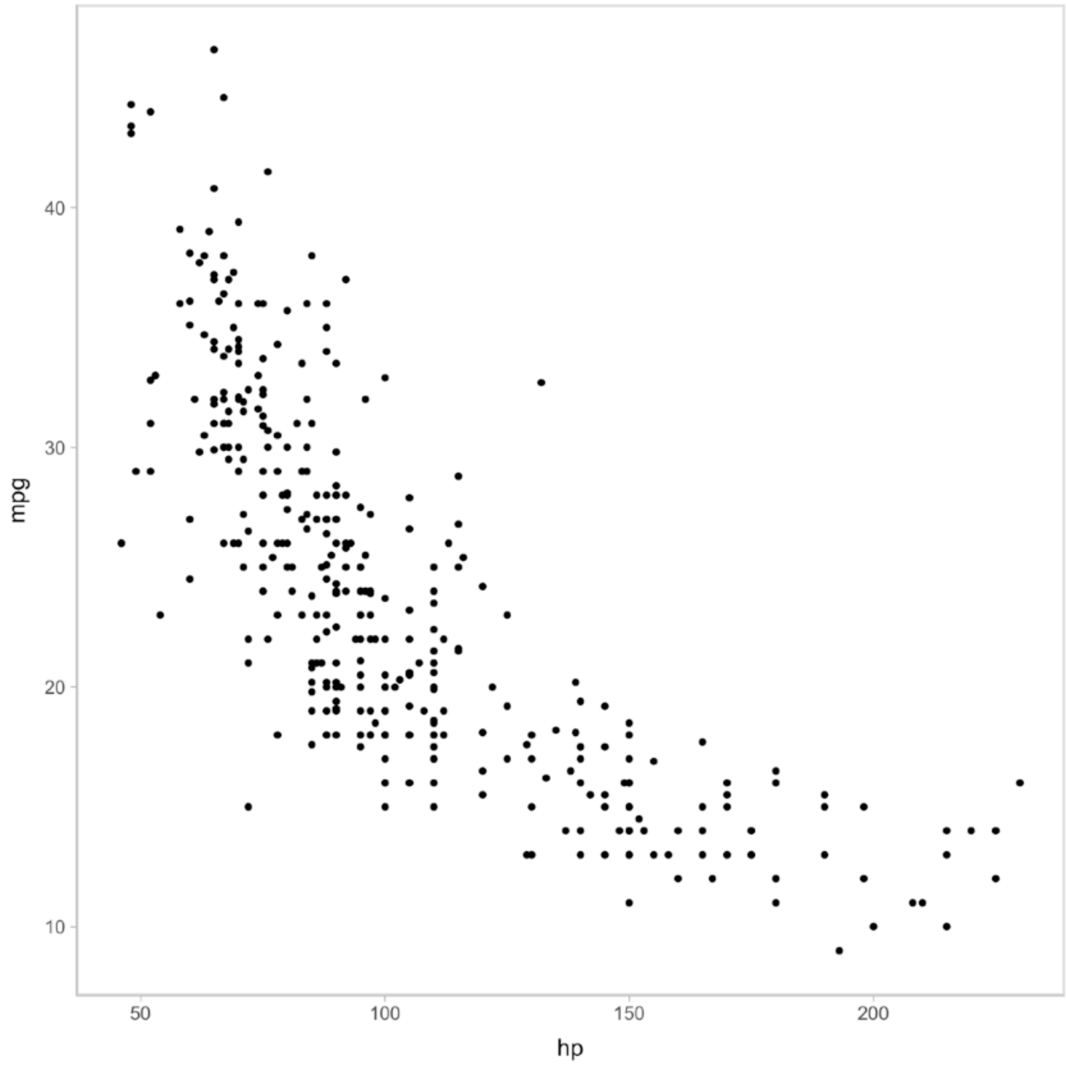


Model:
mpg ~ cyl

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

more cylinders
→ more fuel consumption
→ worse mileage

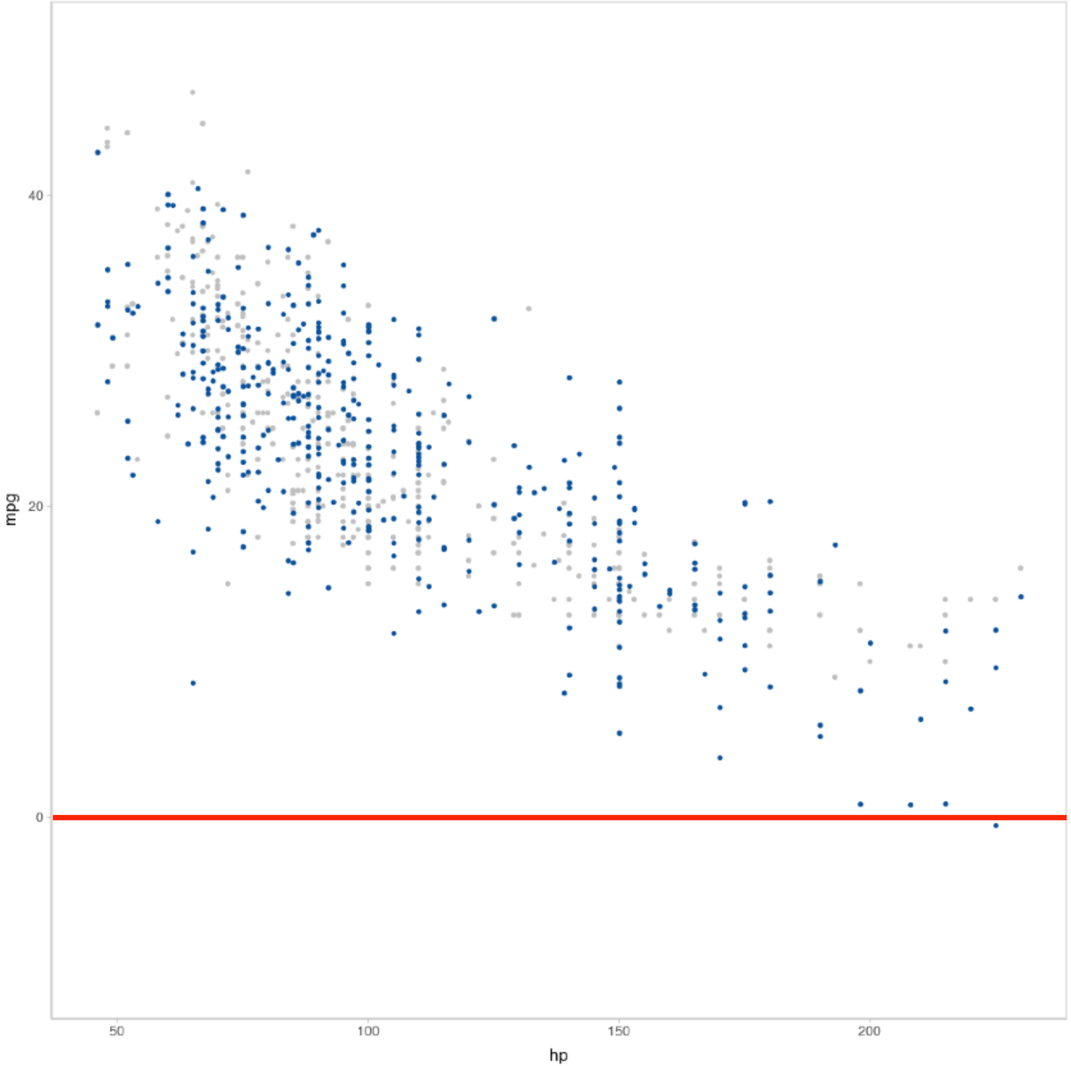
Might this explain the differences across regions?



Plot:

mpg by hp

What might our implicit model be?

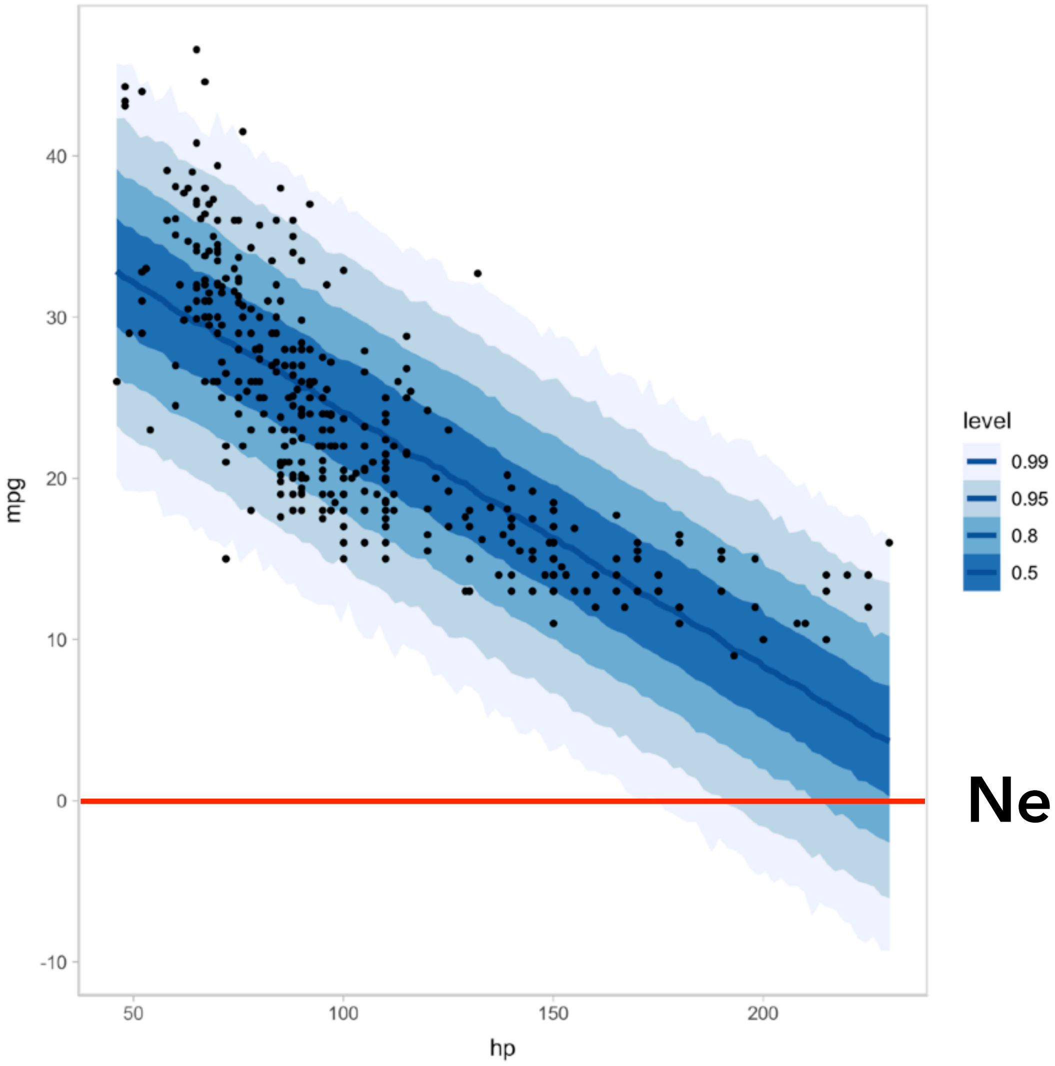


Model:

mpg ~ hp

Linear model, similar to a standard regression. Blue points are model predictions.

Negative mileage?!

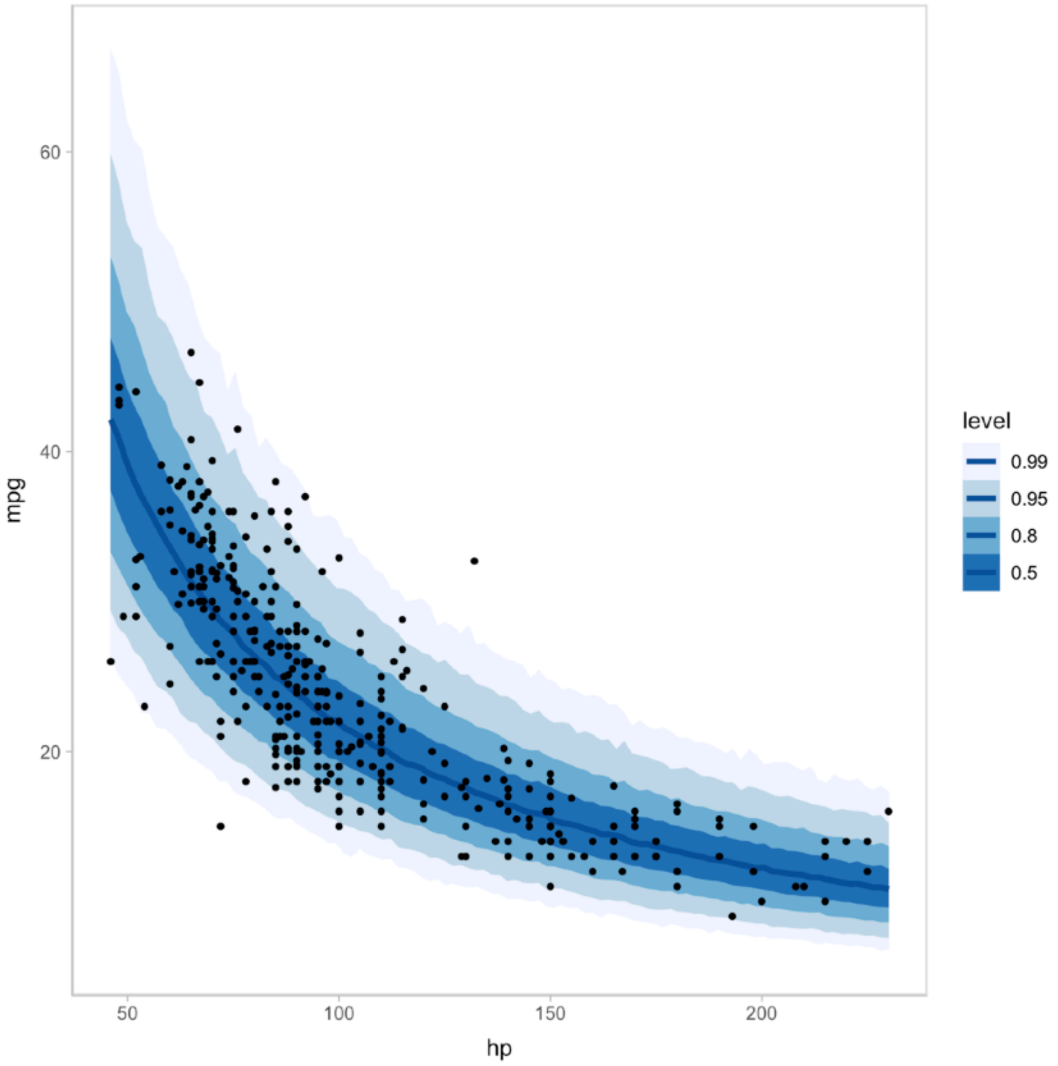


Model:

mpg ~ hp

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

Negative mileage?!

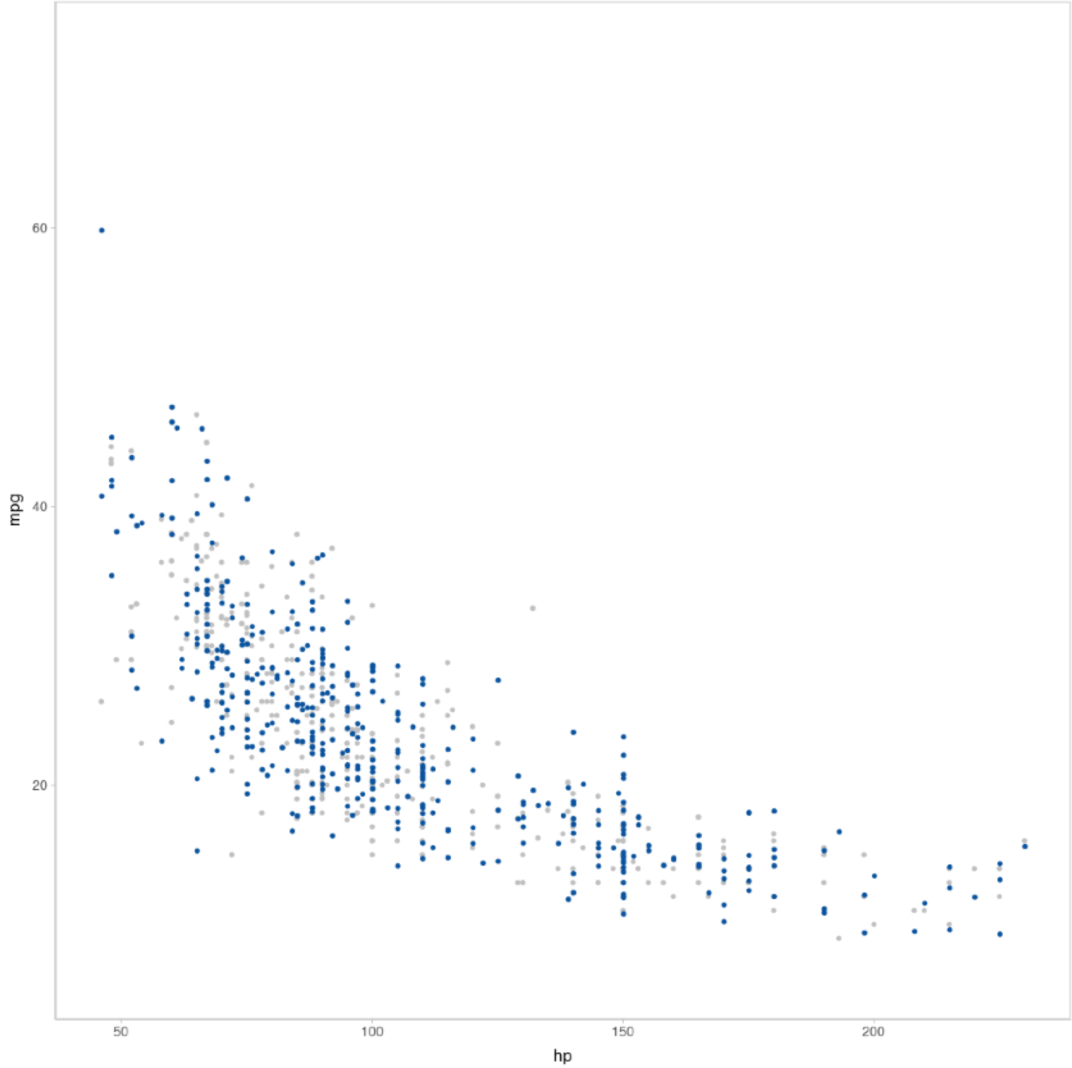


Model:

mpg ~ log(hp)

family = lognormal

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



Model:

mpg ~ log(hp)

family = lognormal

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

Tools for Model Checks

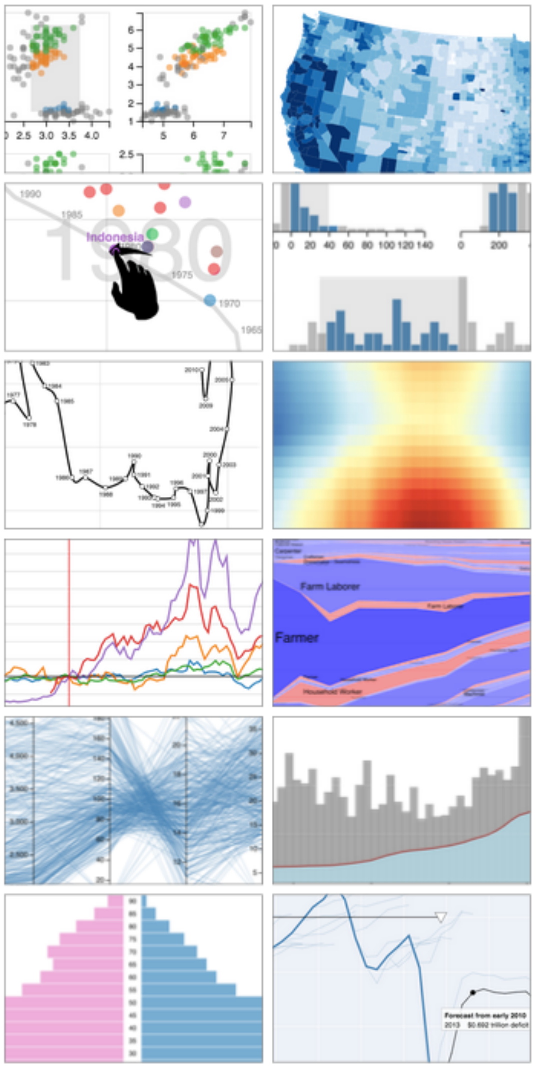
R provides the needed modeling and visualization tools

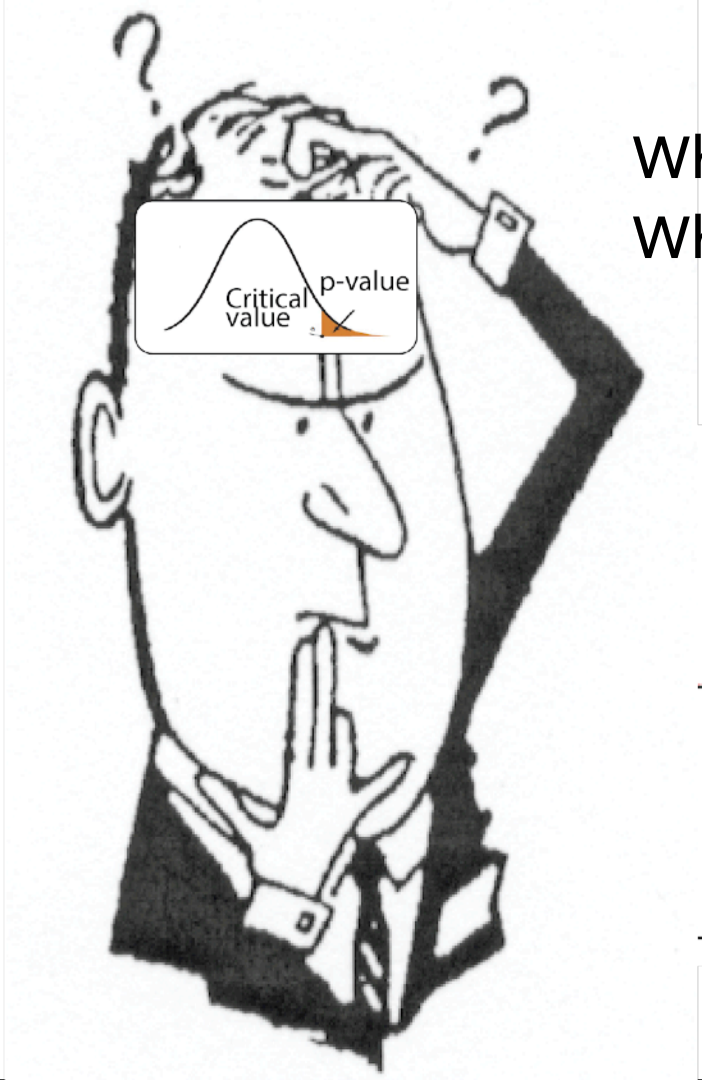
For example:

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

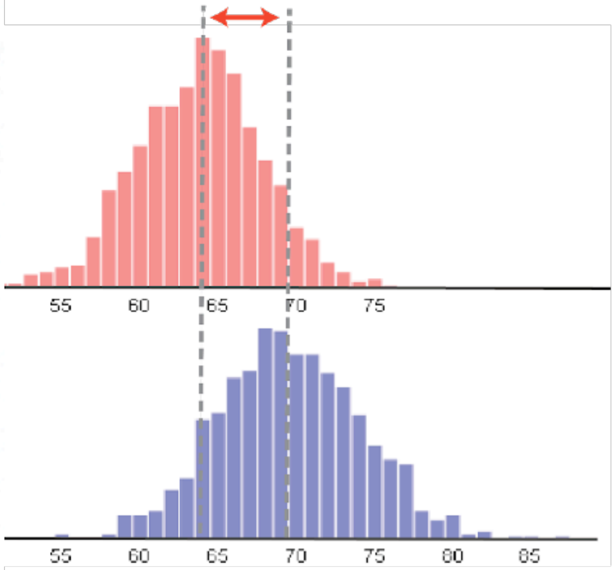
To get started, I recommend the `tidybayes` vignettes:
https://mjskay.github.io/tidybayes/

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?



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