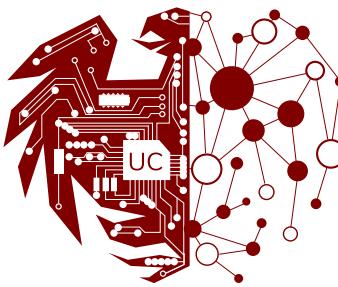


# Surfacing Visualization Mirages



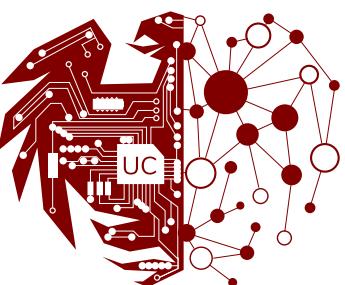
**Andrew McNutt**  
Gordon Kindlmann



Michael Correll



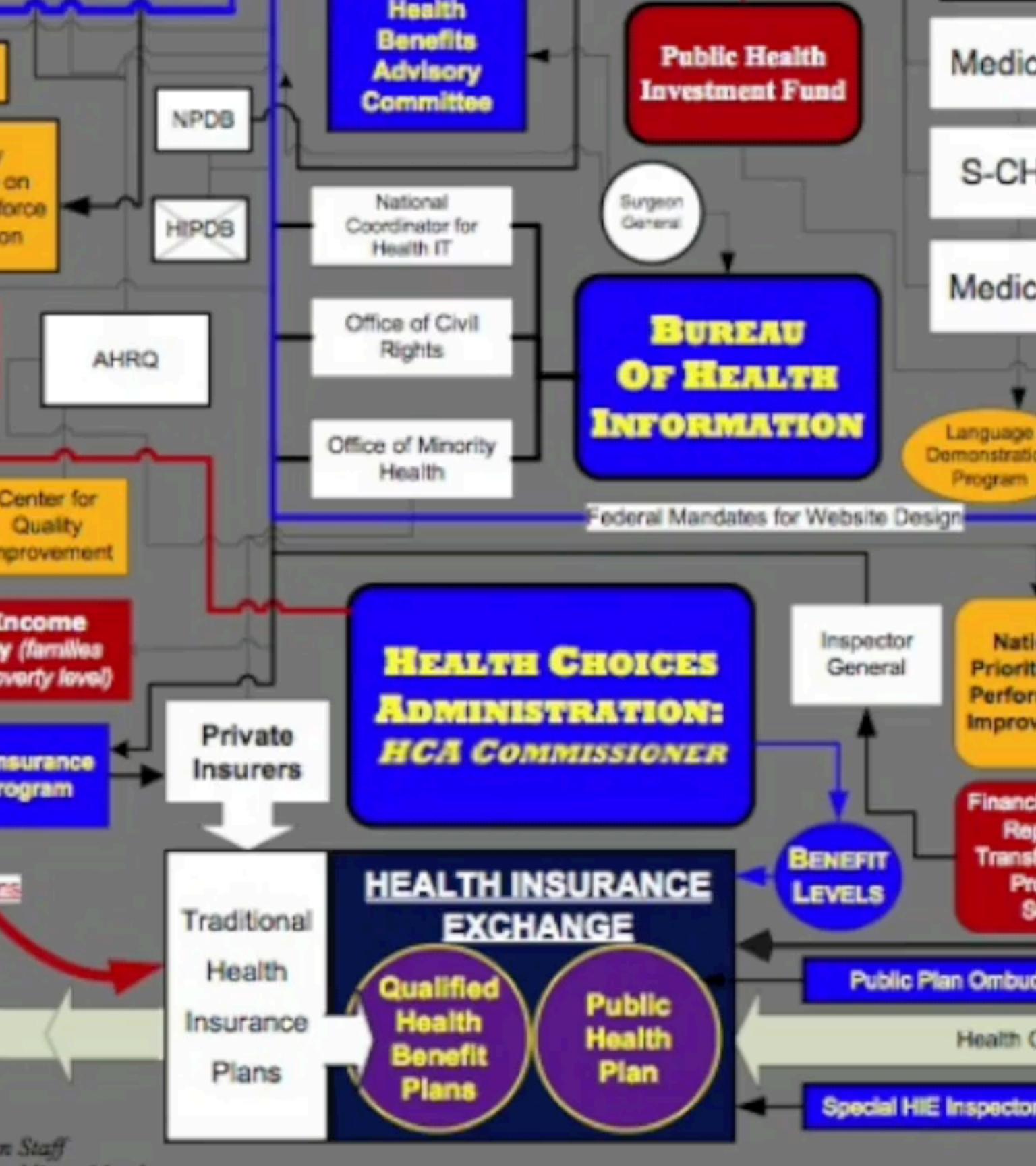
# Surfacing Visualization Mirages



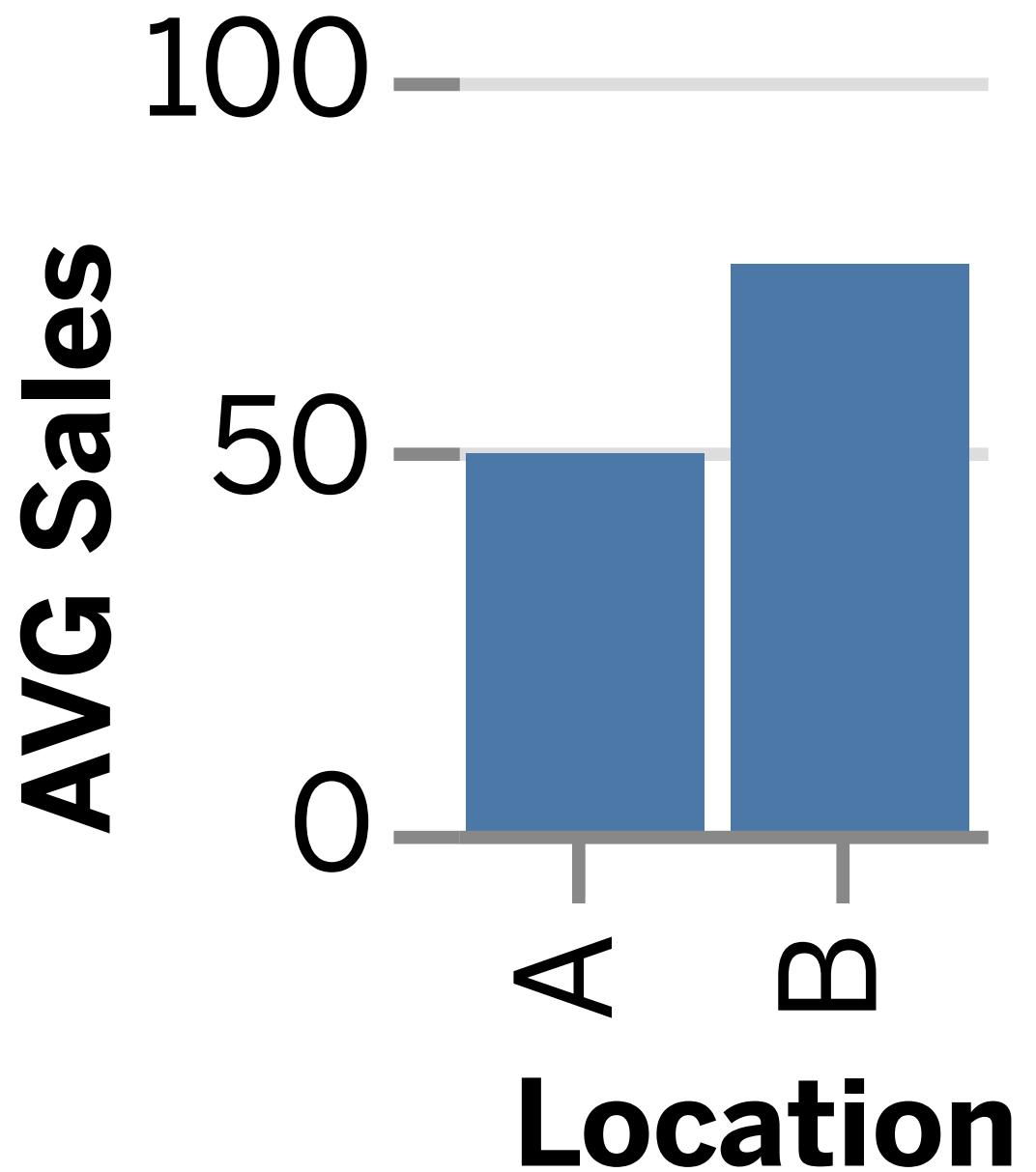
Andrew McNutt  
Gordon Kindlmann



Michael Correll

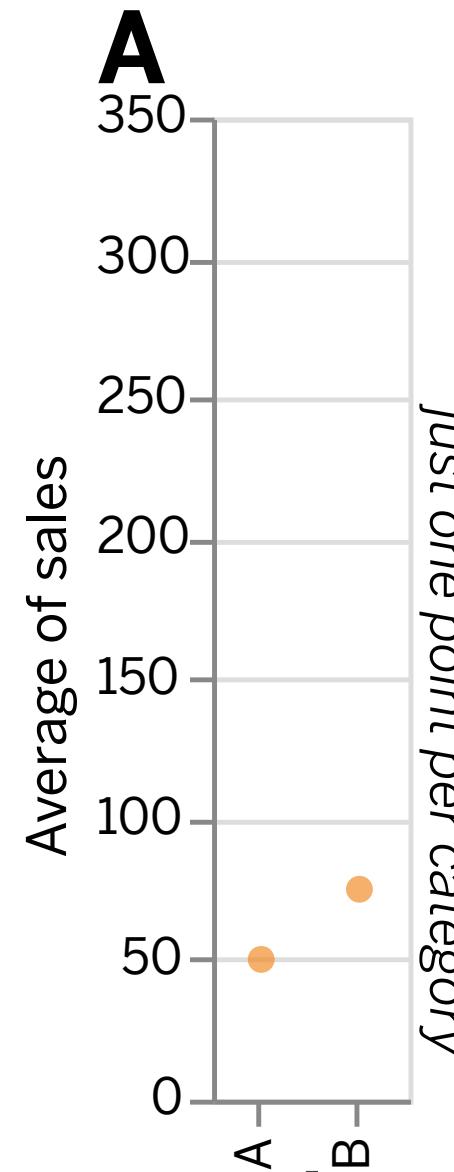


What can be said  
about this chart?

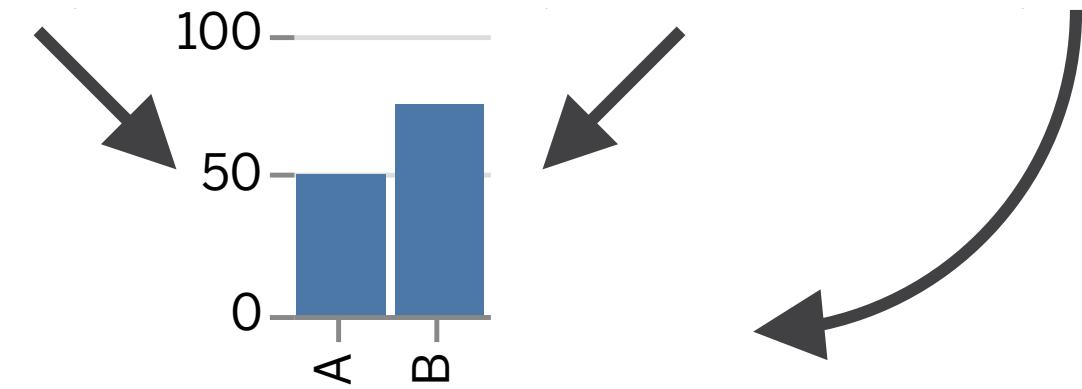


*Aggregations can  
mask  
substantially  
different datasets*

Different Data Sets

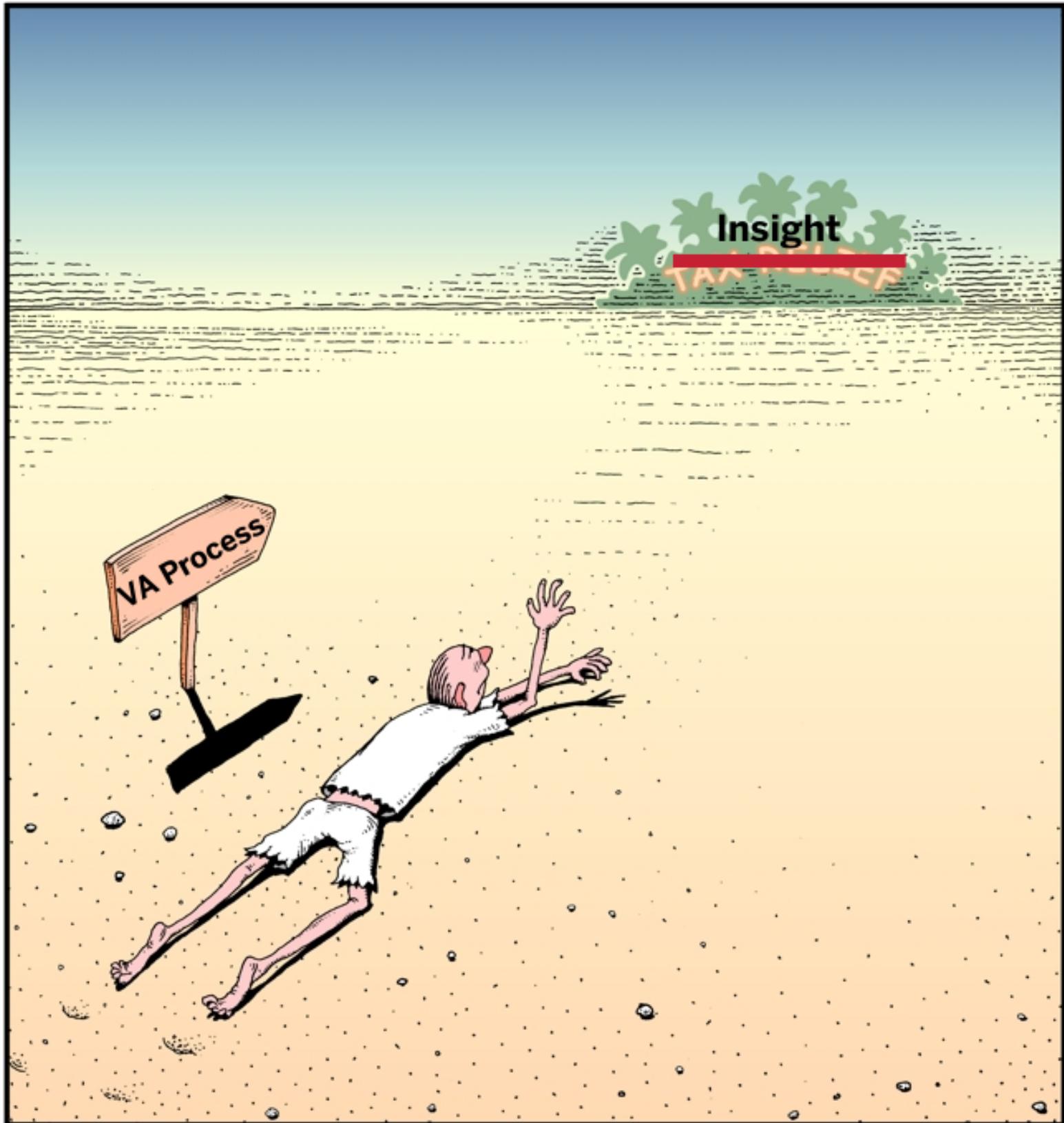


Same Chart



# What is a visualization mirage?

*“A visualization mirage is any visualization where the cursory reading of the visualization would appear to support a particular message arising from the data, but where a closer re-examination would remove or cast significant doubt on this support.”*



## Outliers

Many forms of analysis assume data have similar magnitudes and were generated by similar processes. Outliers, whether in the form of erroneous or unexpectedly extreme values, can greatly impact aggregation and discredit the assumptions behind many statistical tests and summaries. [61]

## Overplotting

We typically expect to be able to clearly identify individual marks, and expect that one visual mark corresponds to a single value or aggregated value. Yet overlapping marks can hide internal structures in the distribution or disguise potential data quality issues. [24, 77, 81]

## Availability Heuristic

Examples that are easier to recall are perceived as more typical than they actually are. In a visual analytics context, this could be reflected in analysts recalling outlying instances more easily than values that match the trend, or assuming that the data patterns they encounter most frequently (for instance, in the default or home view of their tool) are more common than they really are in the dataset as a whole. [28, 29, 35]

## Inappropriate/Missing Aggregation

The size of the dataset is often far larger than what can fit in a particular chart. Aggregation at a particular level of detail is a common technique to reduce the size of the data. However, the choice of aggregation function can lead to differing conclusions based on the underlying distribution of the data. Furthermore, these statistical summaries may fail to capture important features of distribution, such as second-order statistics. Conversely, when a designer fails to apply an aggregation function (or applies one at too low a level of detail), the overplotting, access visual complexity, or reduced discoverability can likewise hide important patterns in the data. [3, 35, 76, 100, 122]

(See the appendix of our paper for this full table)

CURATING ERRORS	Error	Mirage	Singularities	In chart types, such as line series or parallel coordinates plots, many data series can converge into a single point in visual space. Without intervention, viewers can have issues discriminating between which series takes which path after such a singularity. [62]
			Inappropriate Semantic Color Scale	Colors have different effects and semantic associations depending on context (for instance the cultural context of green being associated with money in the United States). Color encodings in charts that violate these assumptions can result in viewers misinterpreting the data: for instance, a viewer might be confused by a map in which the oceans are colored green, and the land colored blue. [70]

Missing or Repeated Records	We often assume that we have one and only one entry for each datum. However, errors in data entry or integration can result in missing or repeated values that may result in inaccurate aggregates or groupings. [61]
Outliers	Many forms of analysis assume data have similar magnitudes and were generated by similar processes. Outliers, whether in the form of erroneous or unexpectedly extreme values, can greatly impact aggregation and discredit the assumptions behind many statistical tests and summaries. [61]

Manipulation of Scales	the trends present in the data. [38, 62] The axes and scales of a chart are presumed to straightforwardly represent quantitative information. However, manipulation of these scales (for instance, by flipping them from their commonly assumed directions, truncating or expanding them with respect to the range of the data [88, 23, 17, 95, 20], using non-linear transforms, or employing dual axes [63, 13]) can cause viewers to misinterpret
------------------------	--

Differing Number of Records by Group	Certain summary statistics, including aggregates, are sensitive to sample size. However, the number of records aggregated into a single mark can very dramatically. This mismatch can mask this sensitivity and problematic per-mark comparisons; when combined with differing levels of aggregation, it can result in counter-intuitive results such as Simpson's Paradox. [45]
--------------------------------------	--

Visual Complexity	Important trends can be obscured by too much visual complexity, such as when the number of nodes and links in a graph visualization becomes overwhelming. [51, 42]
-------------------	--

READING ERRORS	
----------------	--

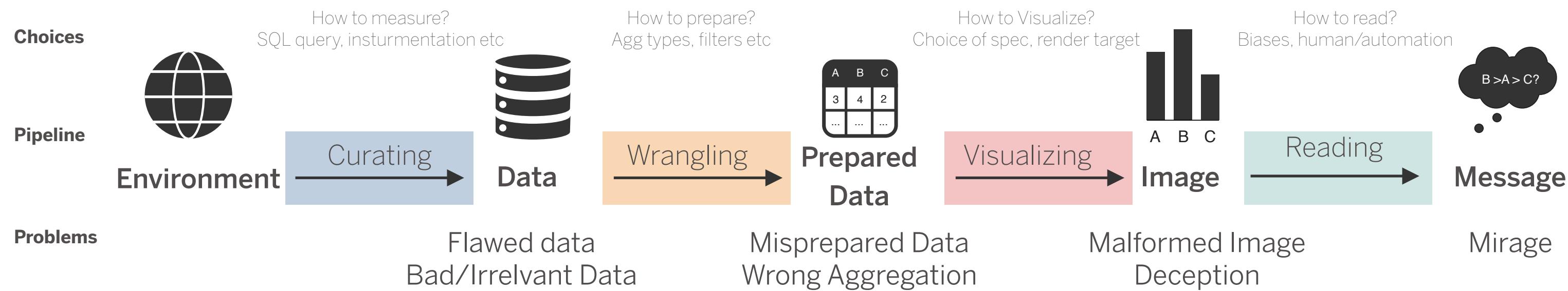
Error	Mirage	VISUALIZING + WRANGLING ERRORS
Outliers Dominate	Numeric and color scales are often automatically bound to the extent of the data. If there are a few	

Default Effect	identical when the designer intended them to be separate or a viewer with dyslexia might mistake similarity named points in an annotated scatter plot as denoting the same entity. [71, 90, 129]
	While default settings in visualization systems are often selected to guide users towards best practices,

Unconventional Scale Directions	which marks belong to which color classes. [109] Viewers have certain prior expectations on the direction of scales. For instance, in languages with left-to-right reading orders, time is likewise assumed to move left to right in graphs. Depending on context, dark or opaque colors are perceived as having higher magnitude values than brighter or more transparent colors. Violating these assumptions can cause slower reading times or even the reversal of perceived trends. [23, 88, 115, 102]
Overplotting	We typically expect to be able to clearly identify individual marks, and expect that one visual mark corresponds to a single value or aggregated value. Yet overlapping marks can hide internal structures in the distribution or disguise potential data quality issues. [24, 77, 81]

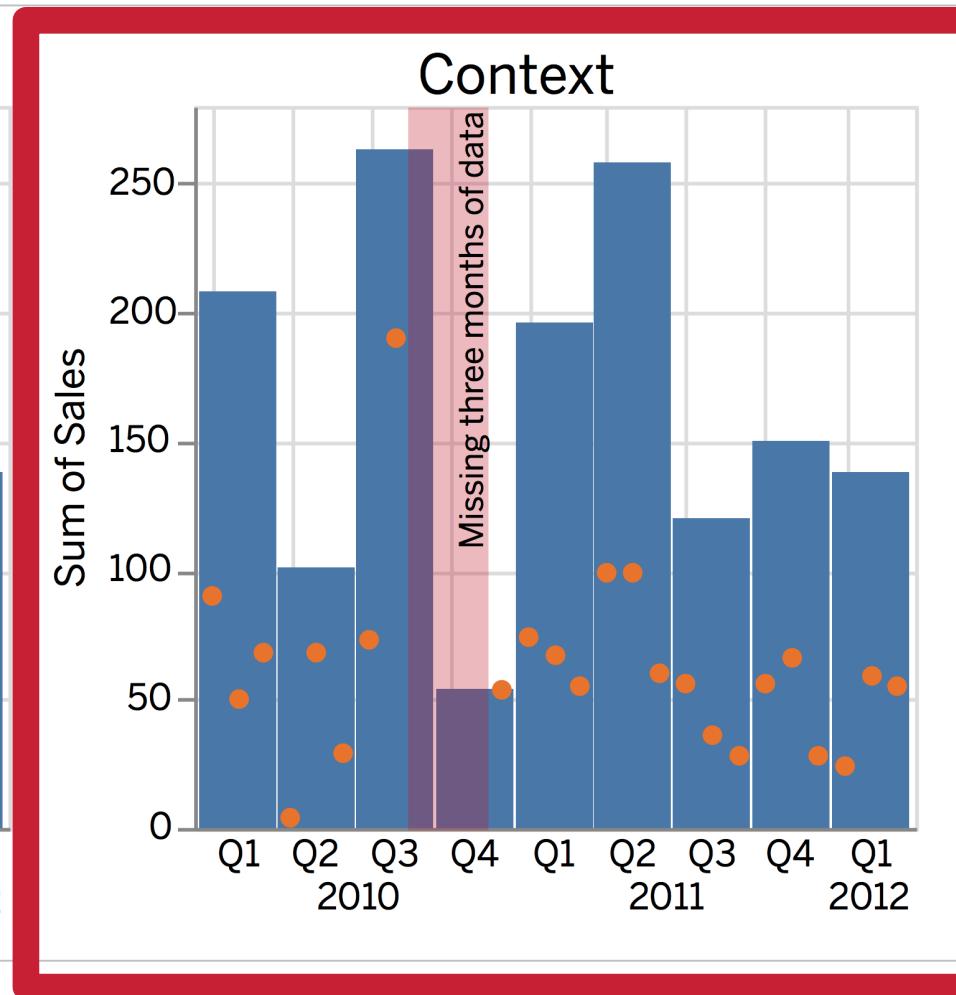
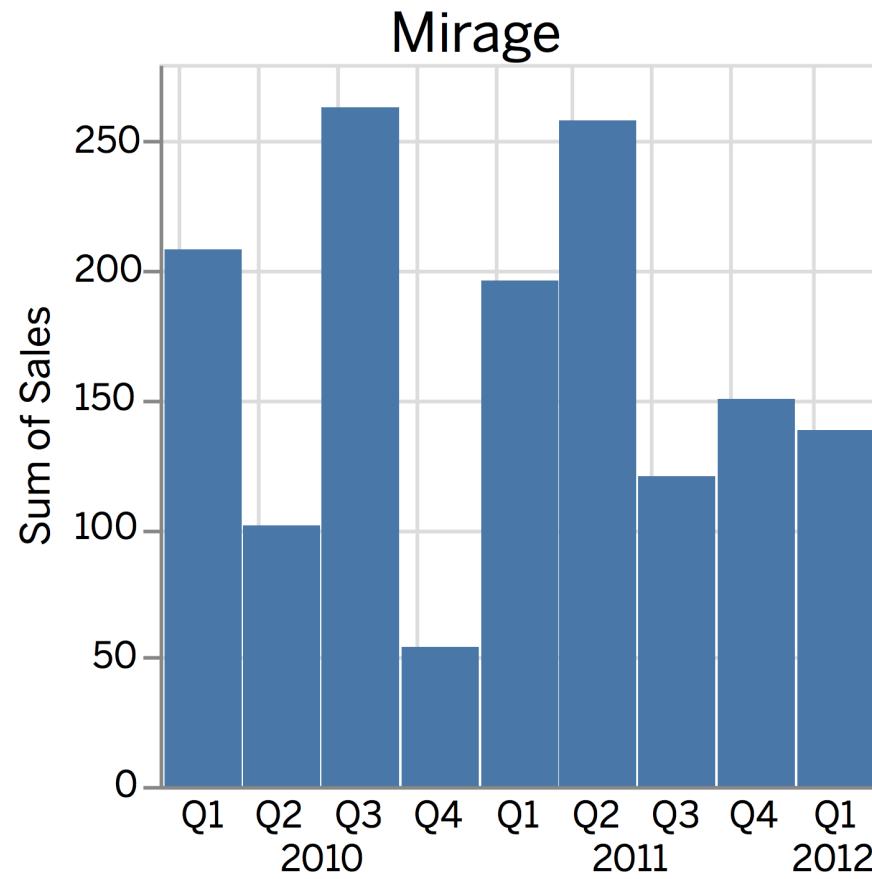
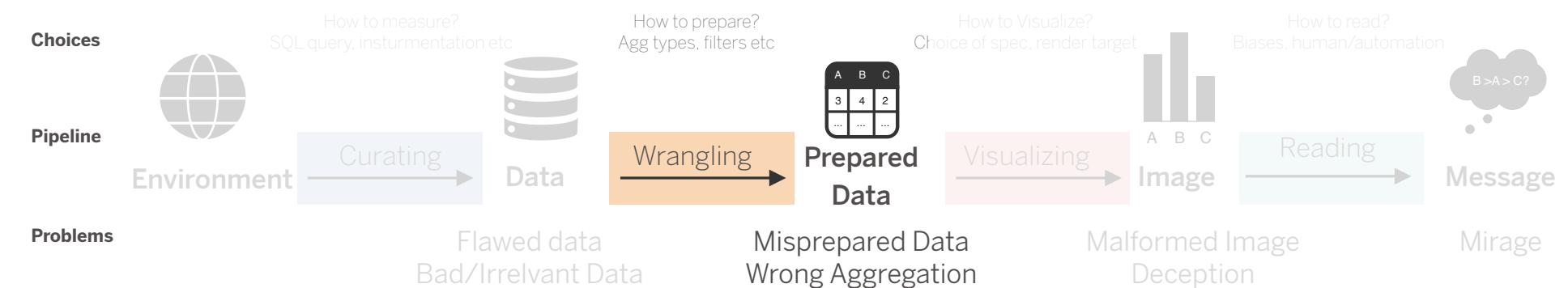
	analytics context, this could be reflected in analysts recalling outlying instances more easily than values that match the trend, or assuming that the data patterns they encounter most frequently (for instance, in the default or home view of their tool) are more common than they really are in the dataset as a whole. [28, 29, 35]
--	--

# The Visual Analytics Pipeline and Agency



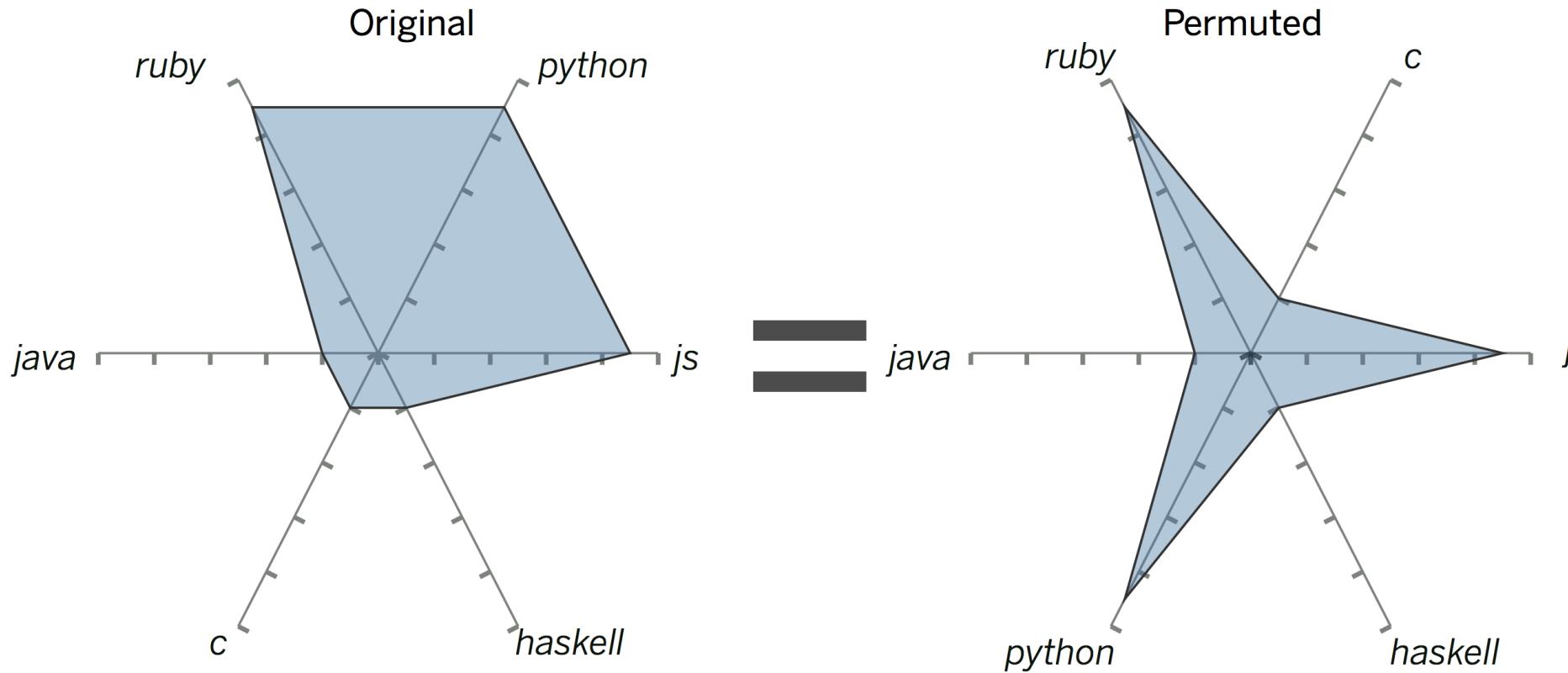
Each stage in the pipeline **offers user agency** but also presents an **opportunity for errors**

# MIRAGES: Dirty Data



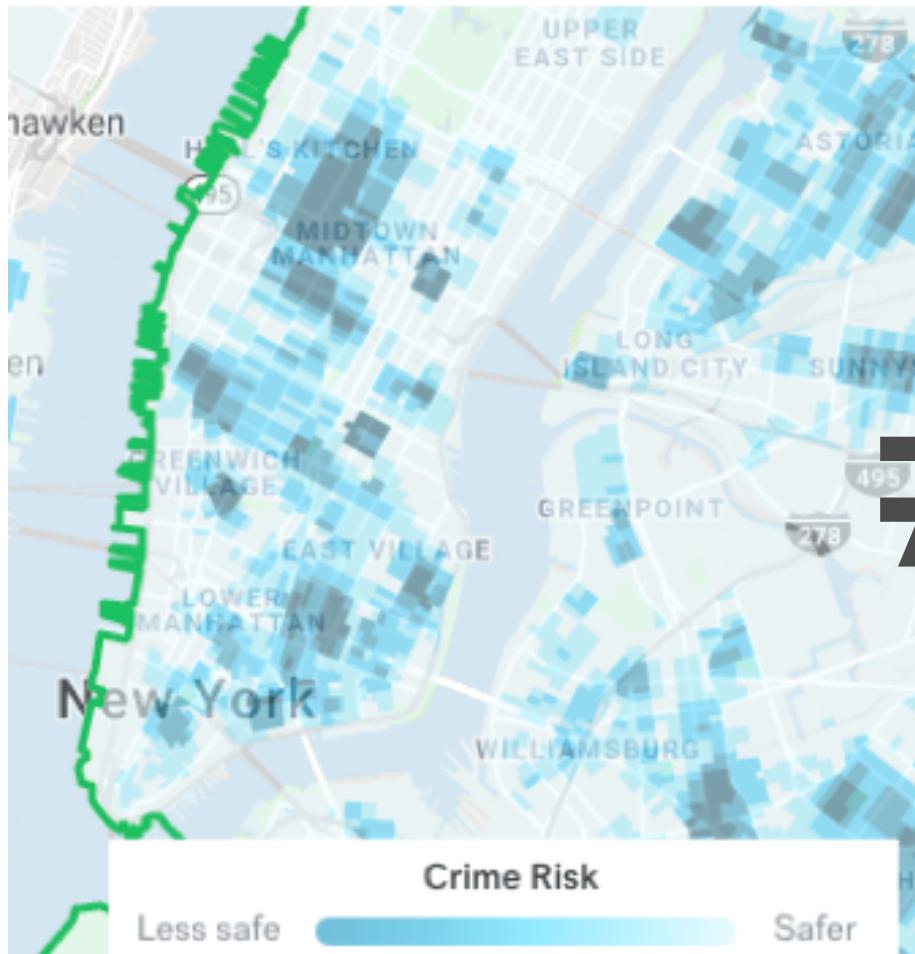
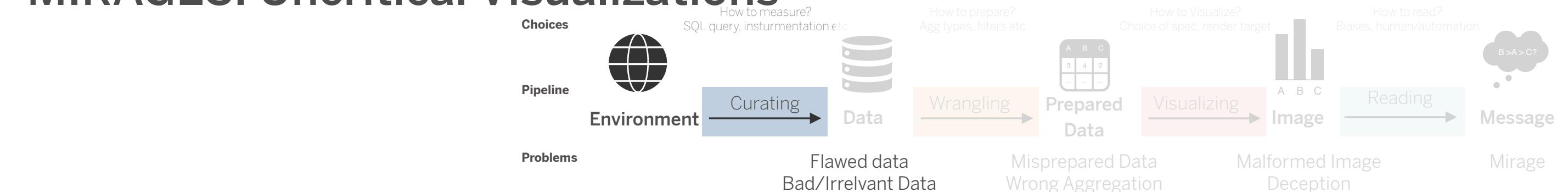
Visualizations that fall prey to **statistical and data entry errors** can cause downstream difficulties

# MIRAGES: Volatile Visualization

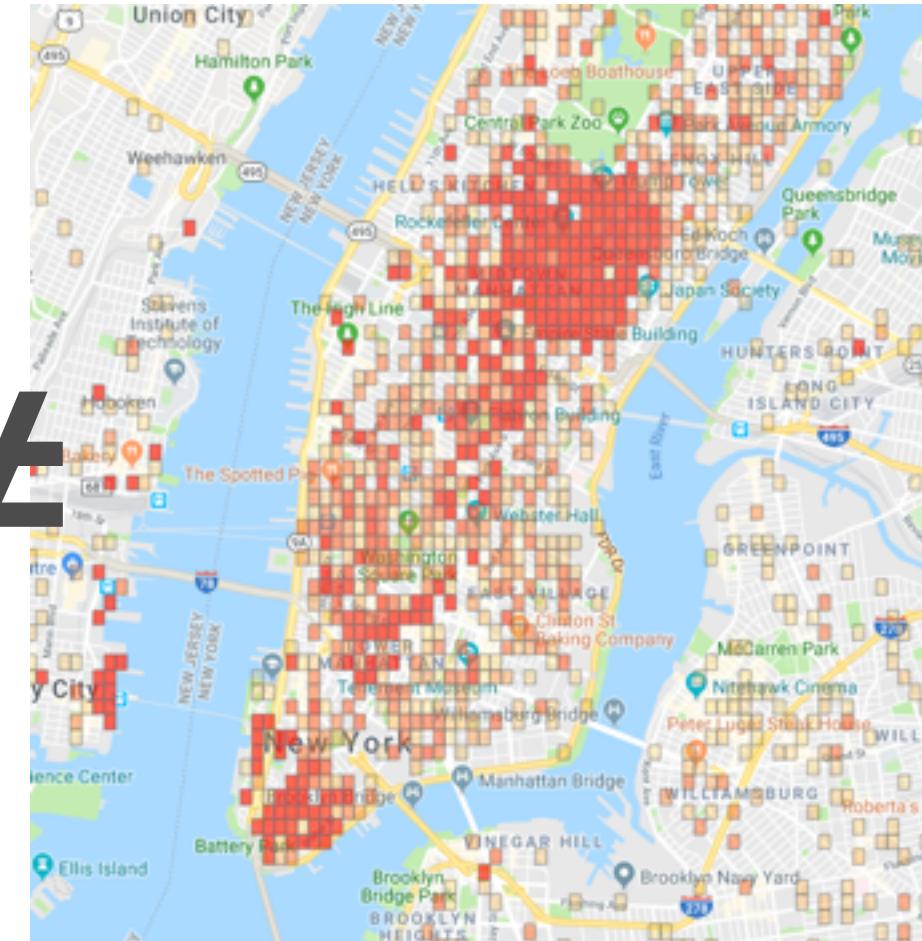


Visualizations that don't reflect **significant changes** to their data in a **significant way**

# MIRAGES: Uncritical Visualizations



Crime by Trulia



White collar crime

Visualizations constructed in an **uncritical or unreflective** way, in such a way that bias and subjectivity are obscured

# Do mirages really happen in practice?

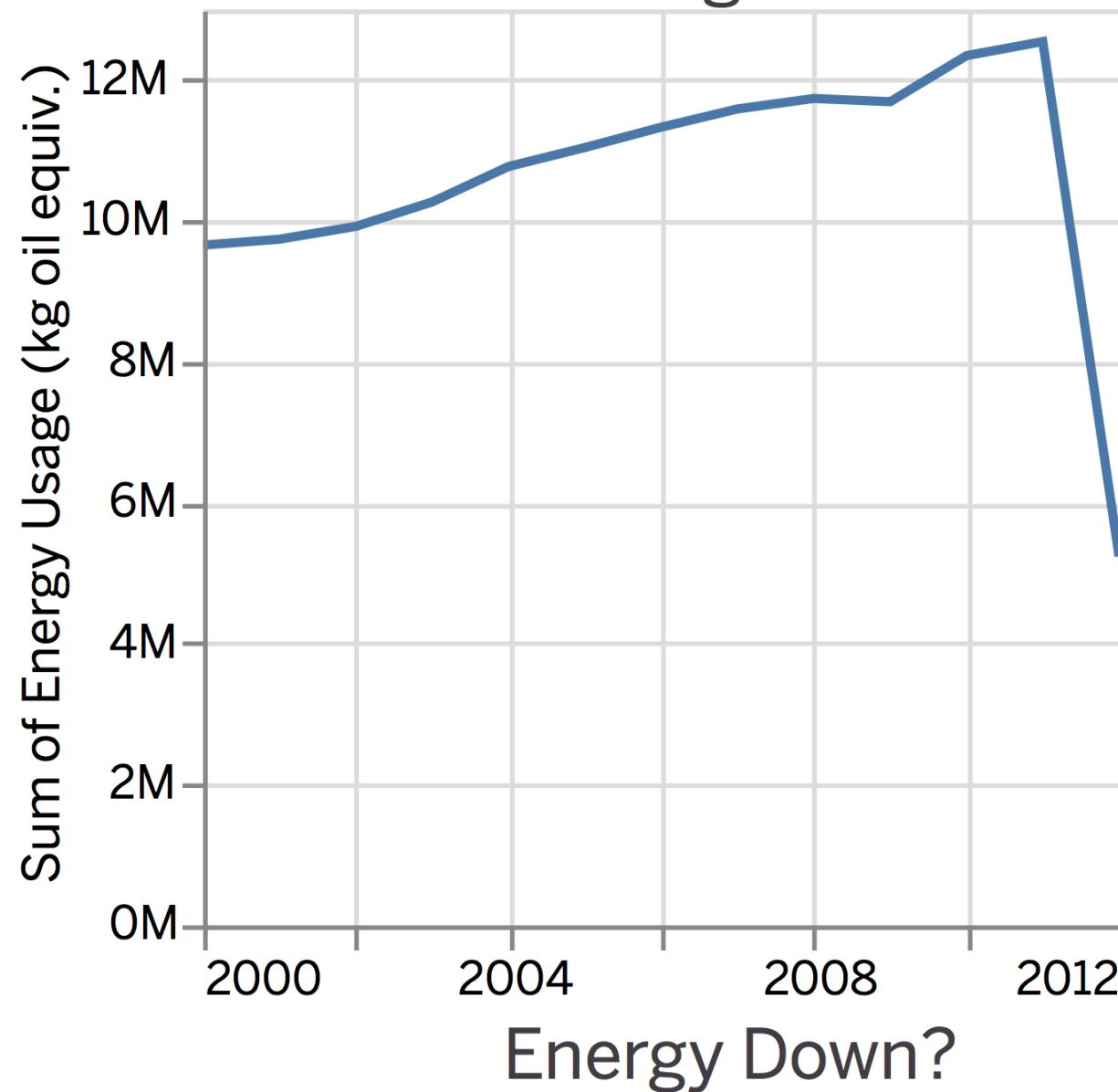
## Example:

What is the trend of global energy usage over time?

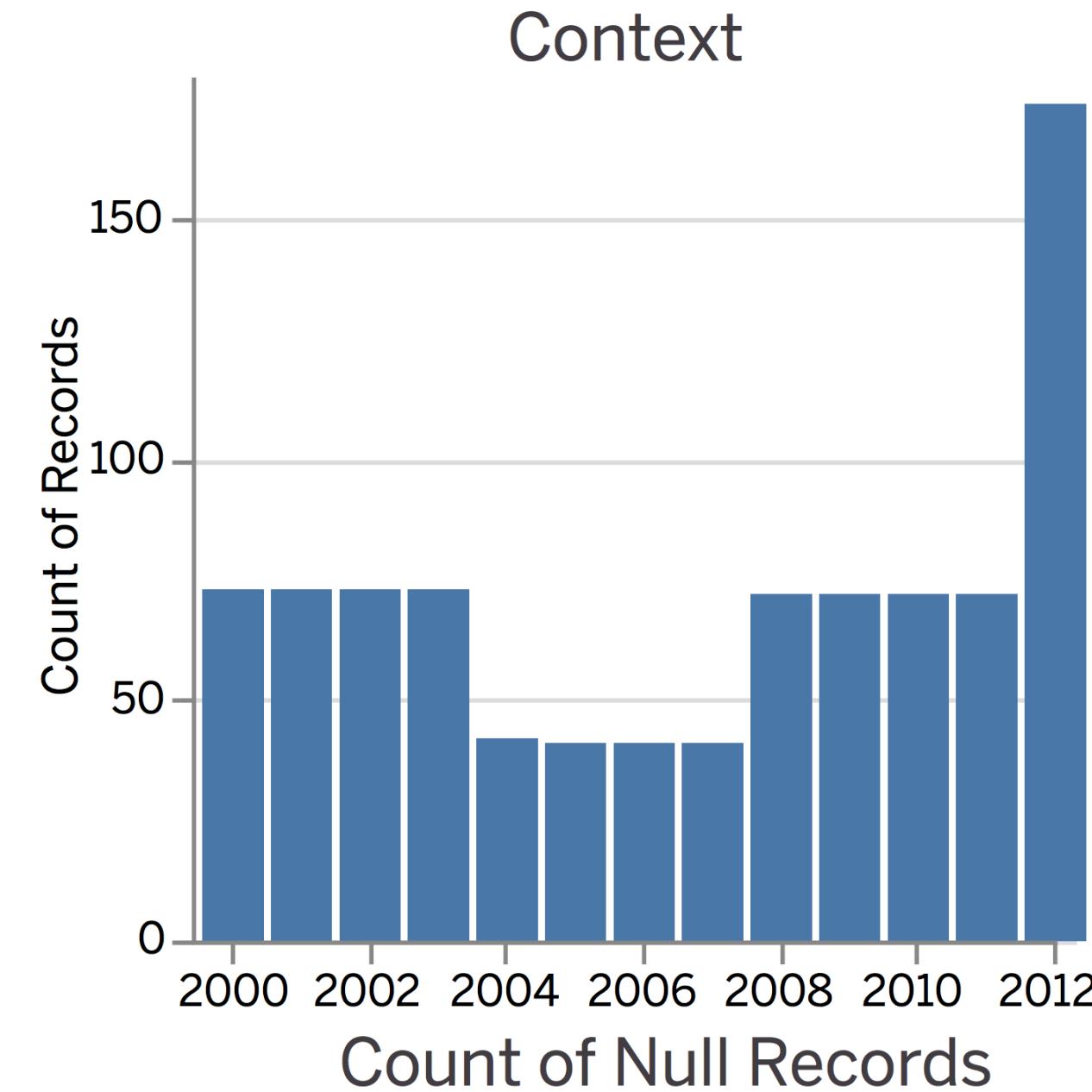
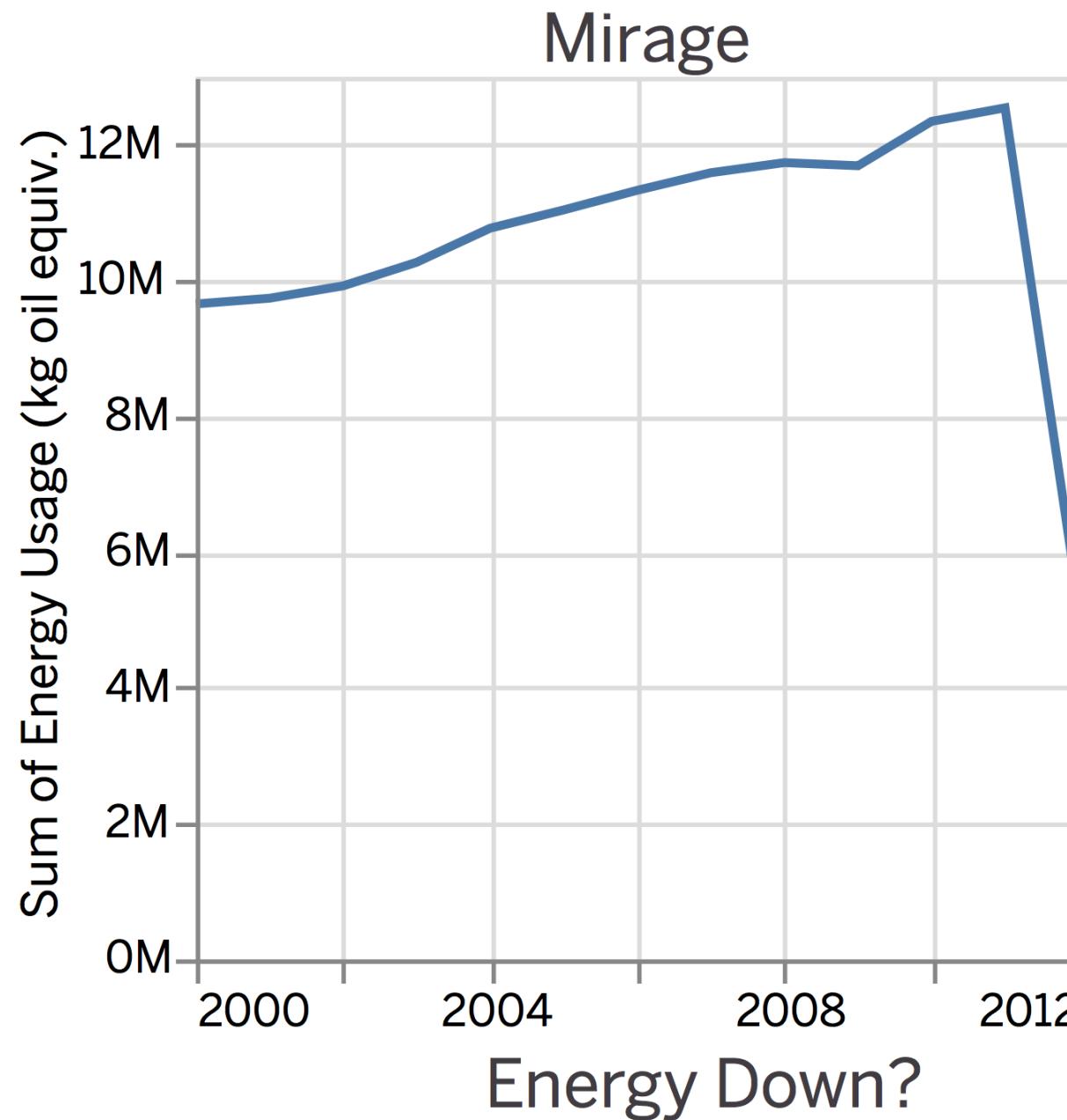


Do these things really happen?

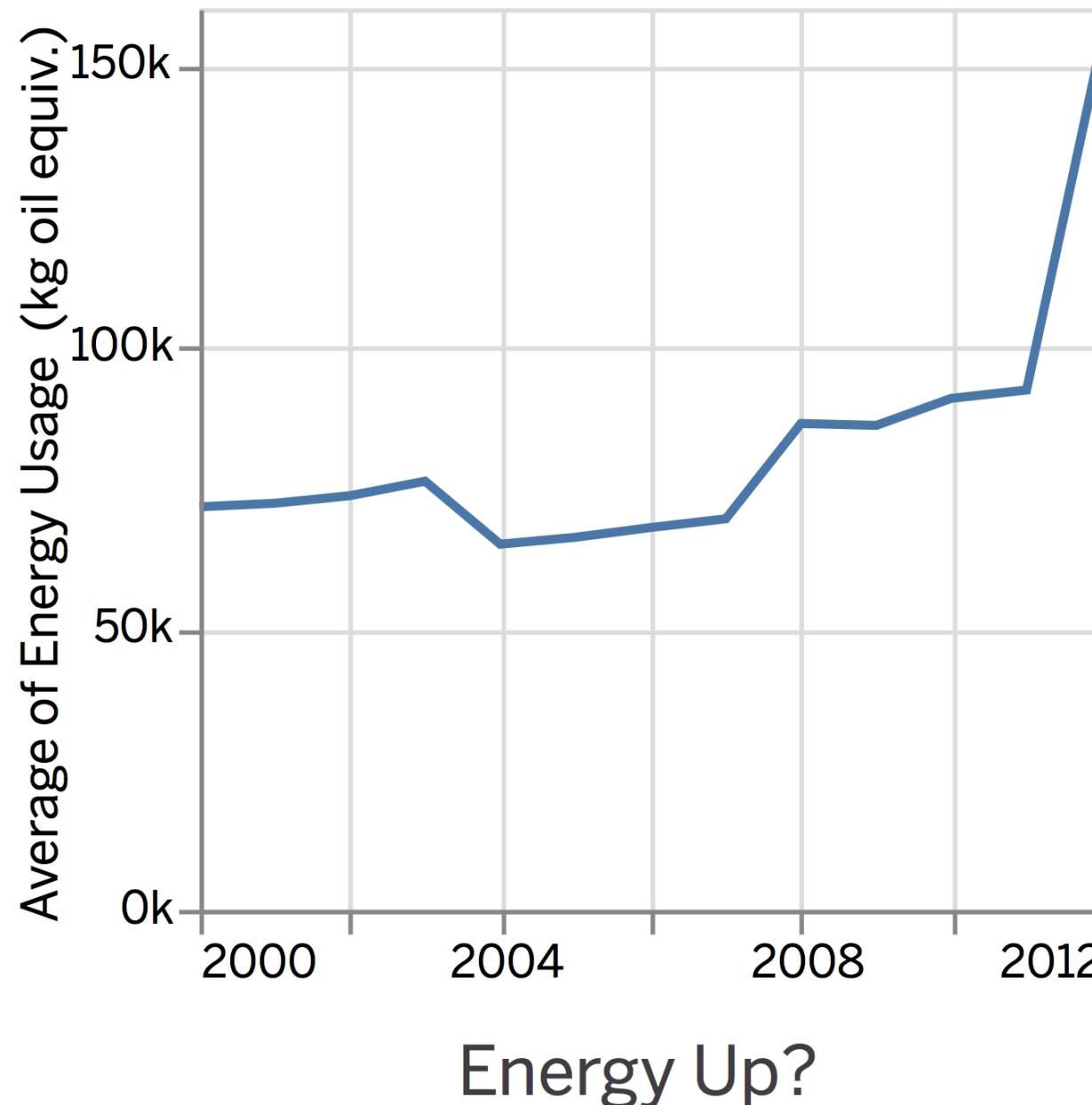
Mirage



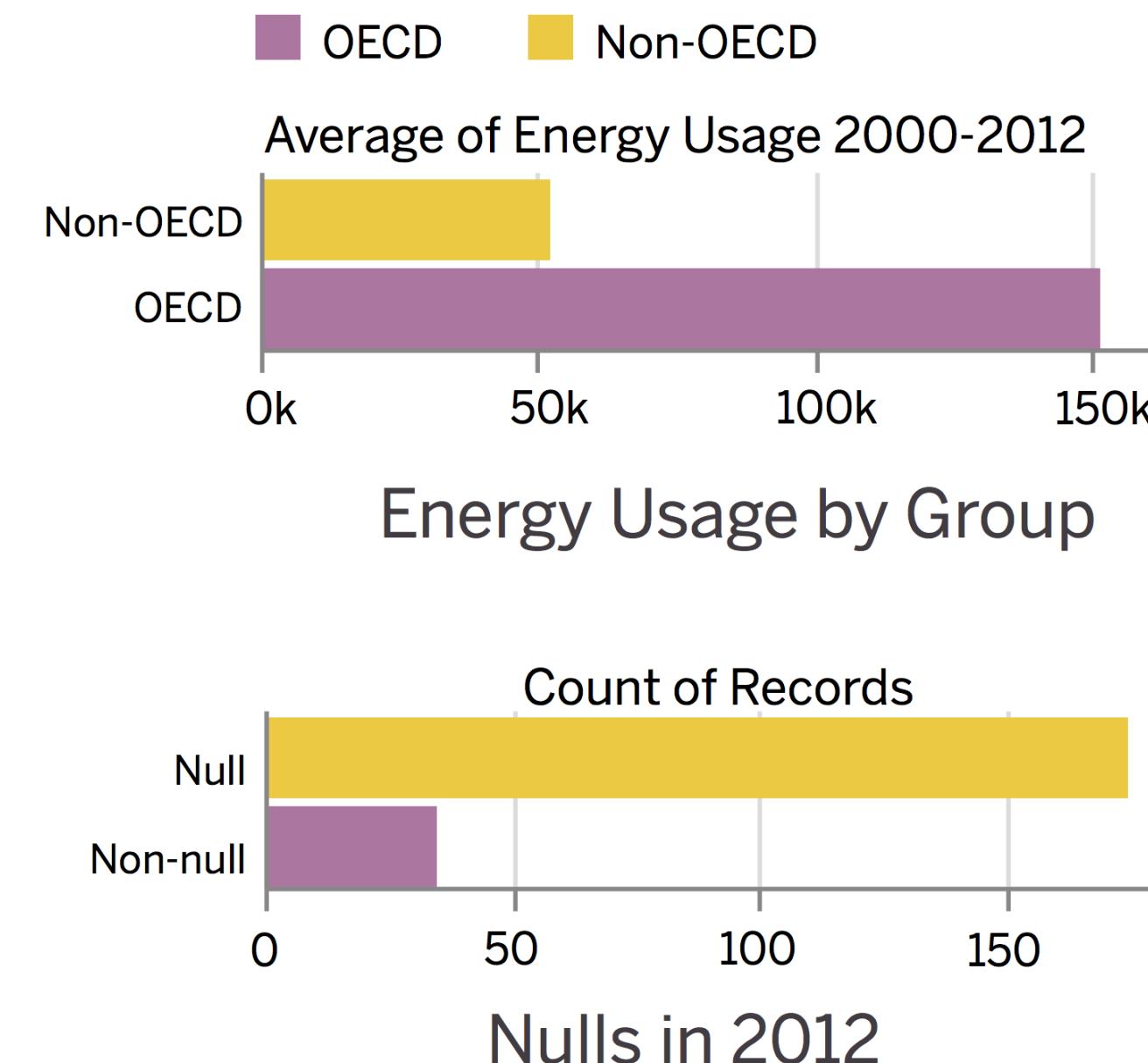
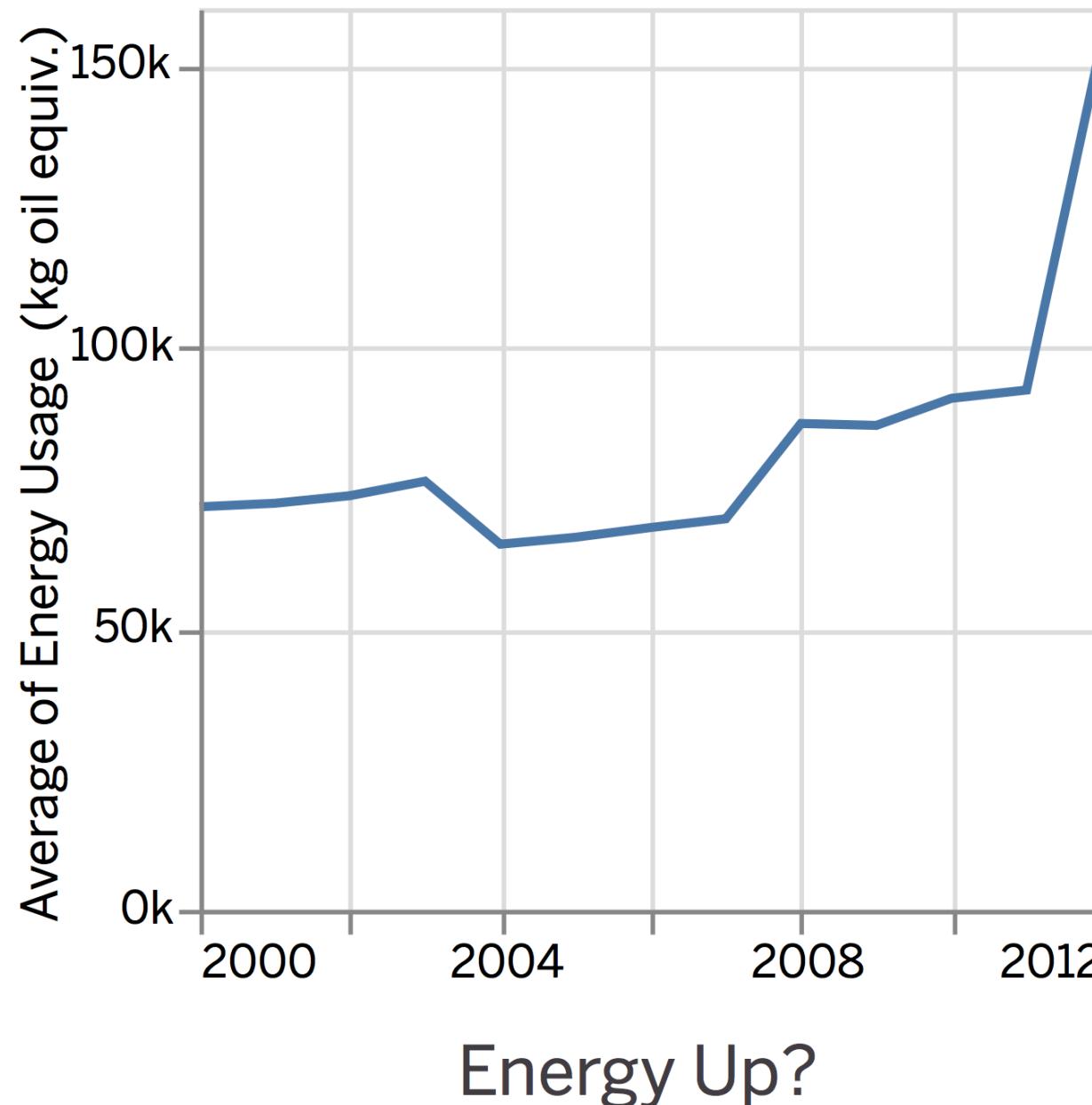
# Do these things really happen?



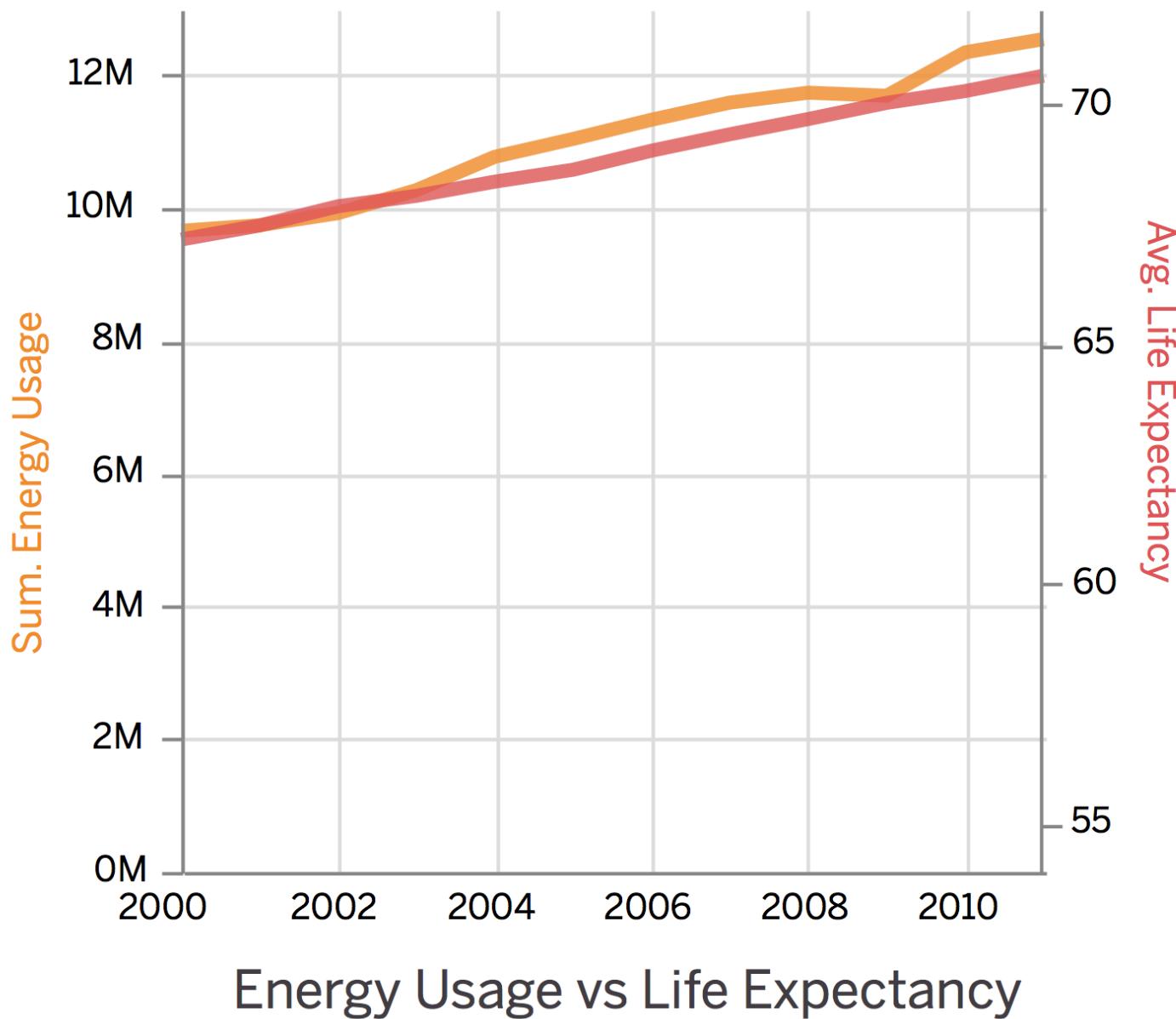
# Do these things really happen?



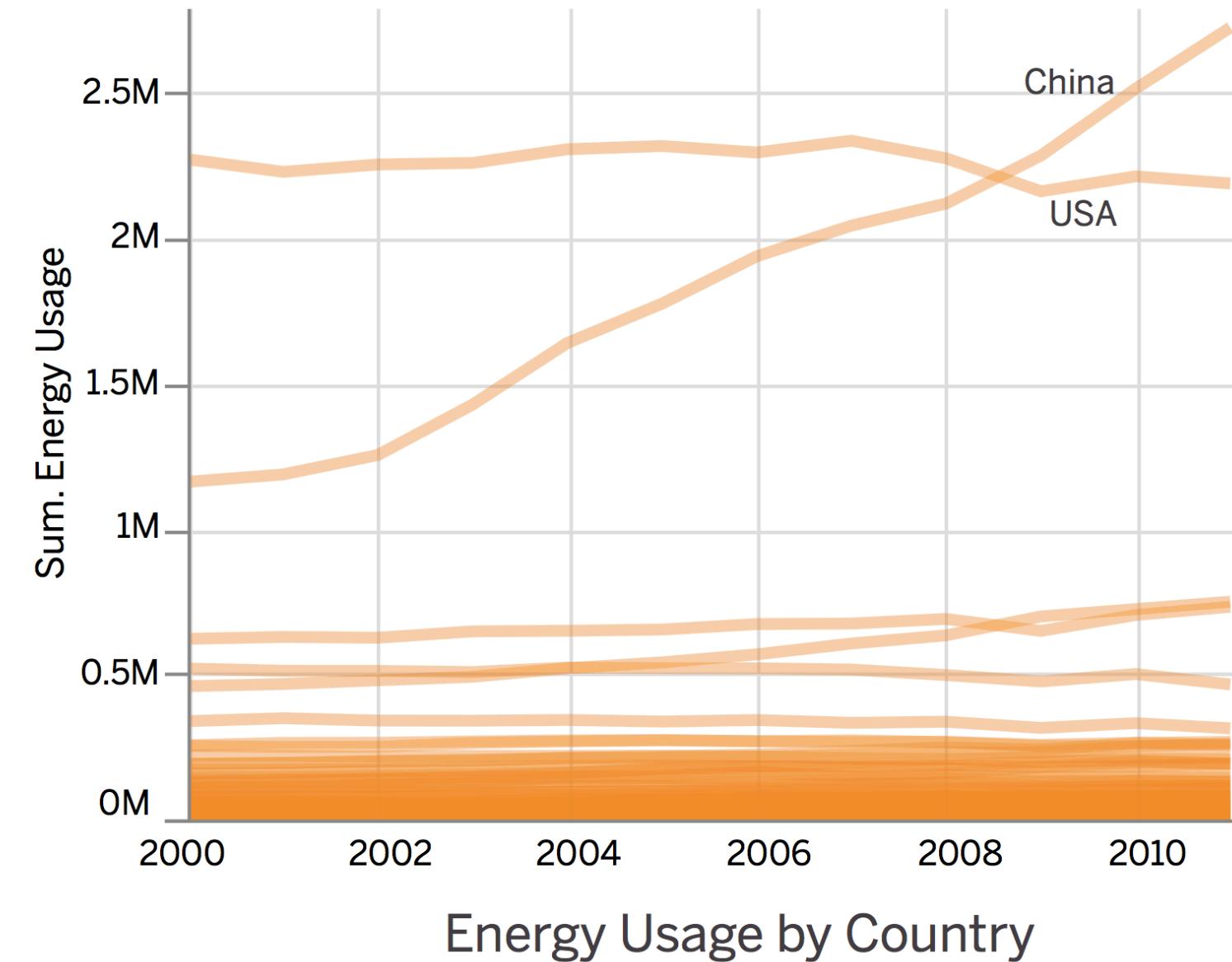
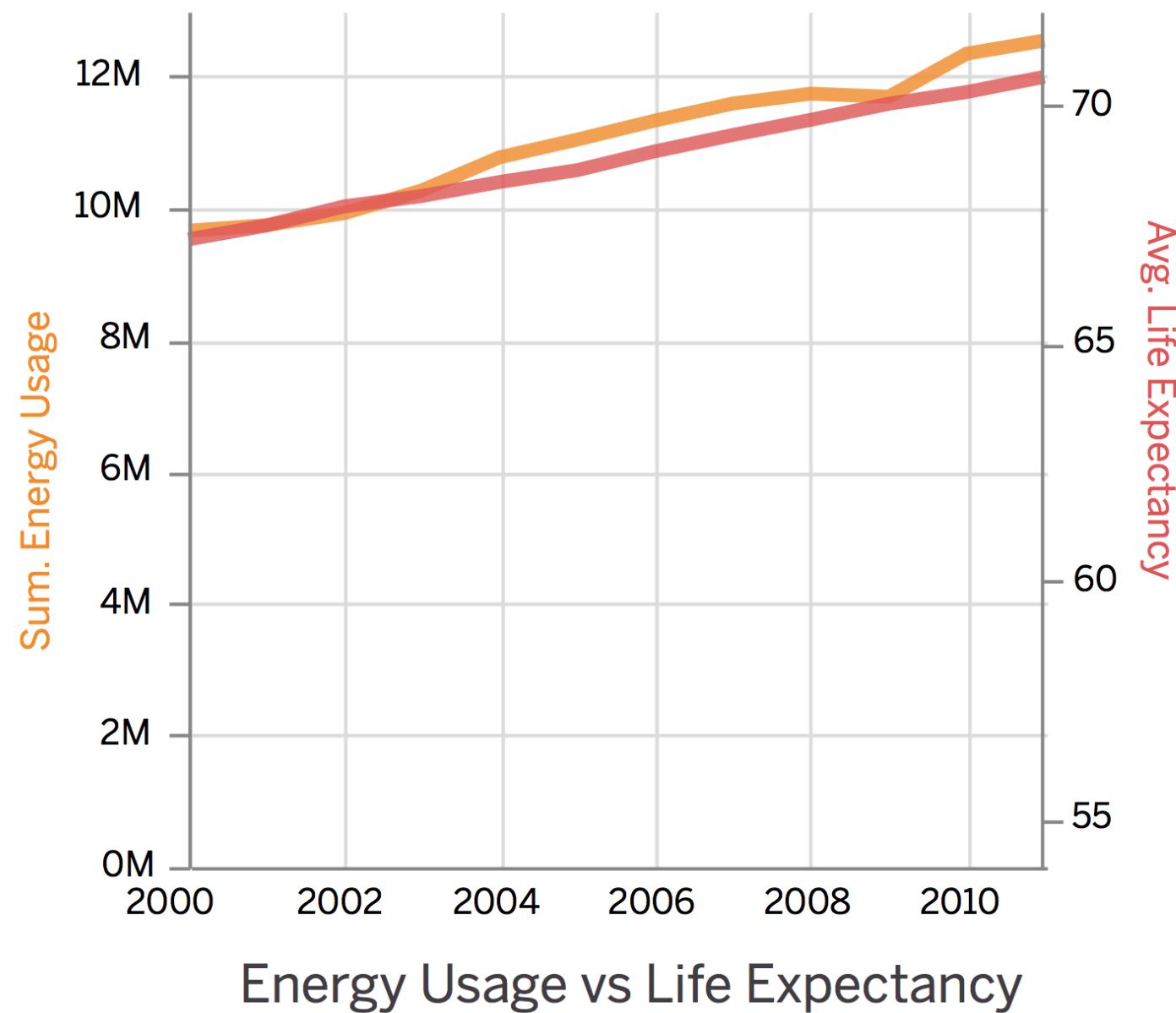
# Do these things really happen?



# Do these things really happen?



# Do these things really happen?

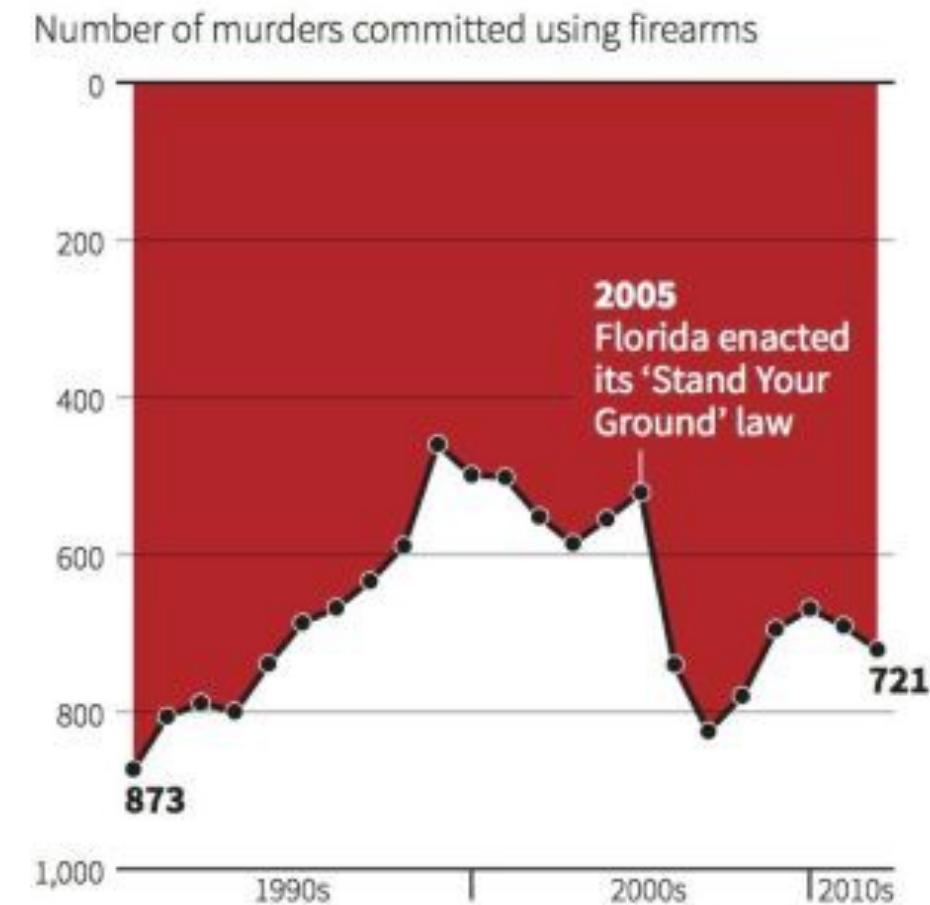


# Visualization Literacy Won't Save You

---

- Despite a push for literacy, people can still be deceived
- Visualizations are rhetorical devices, that are easy to trust too deeply
- Sometimes you are tired, and you miss something “obvious”

## Gun deaths in Florida



Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

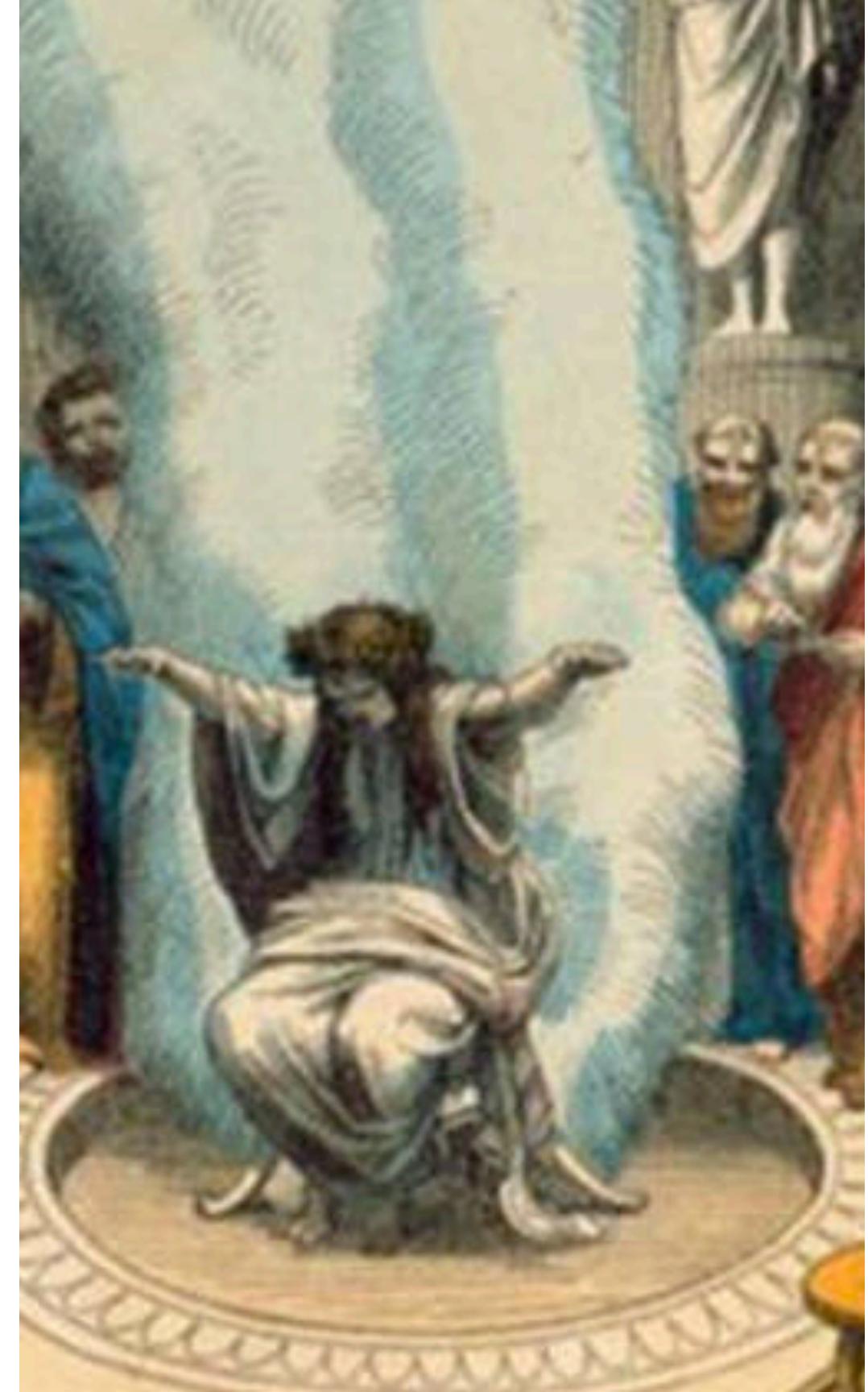
Alerting users to  
mirages is  
essential to  
dispelling them



# Metamorphic Testing

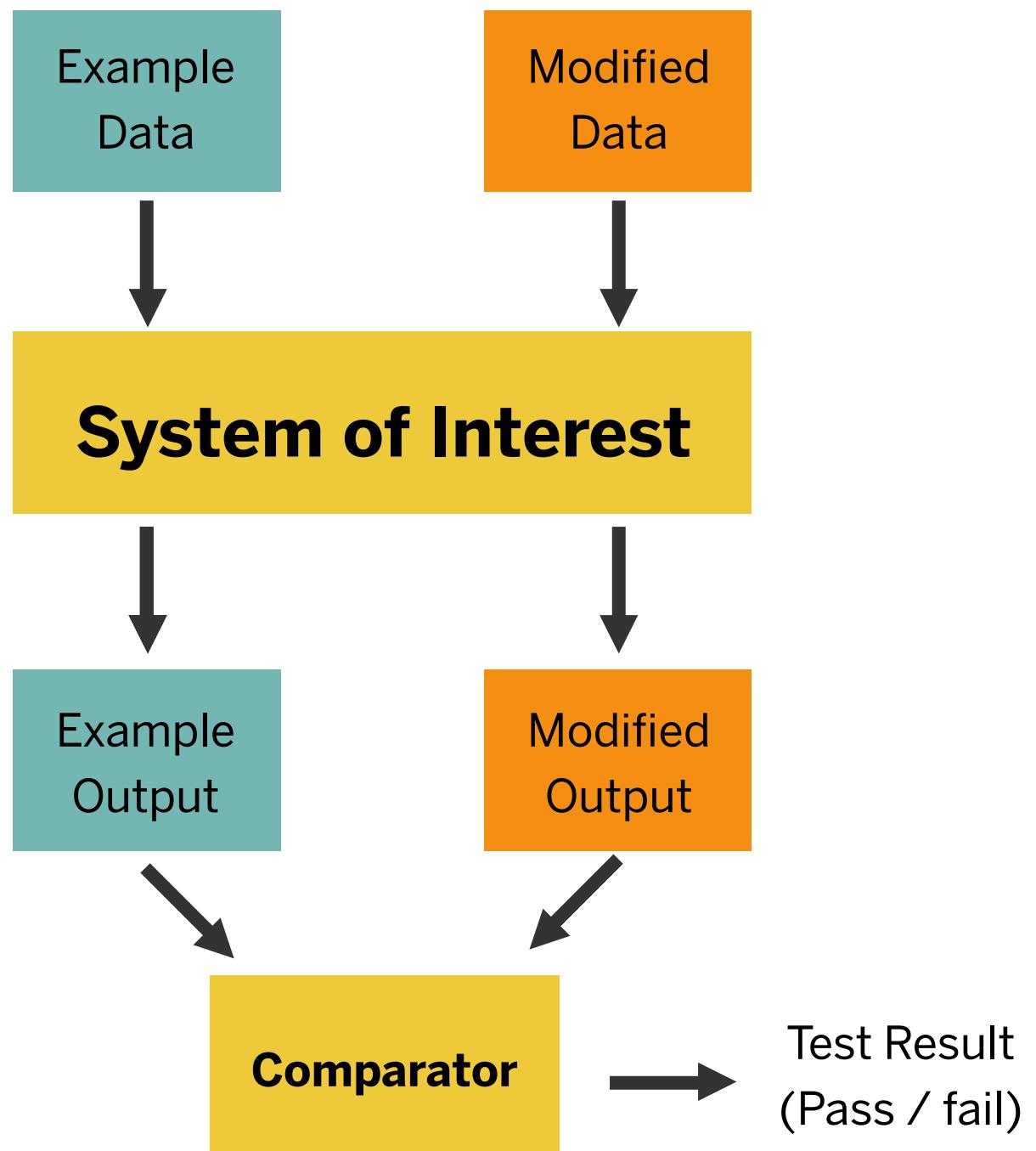
**Test oracles** are not always available

Sometimes it's easier to make assertions about the behavior of outputs **across changes** to the input



# Metamorphic Testing

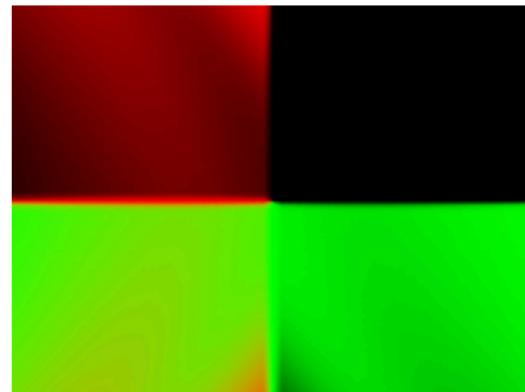
- Find an example of normal behavior
- Mess up the inputs in somehow
- Compare the old and new outputs



# Metamorphic Testing



Original output



Injected Code

Modified Output

## Algebraic Vis Design

“A good visualization, will accurately reflect significant changes in the data with **COMPREHENSIBLE** & significant changes to the visualization”

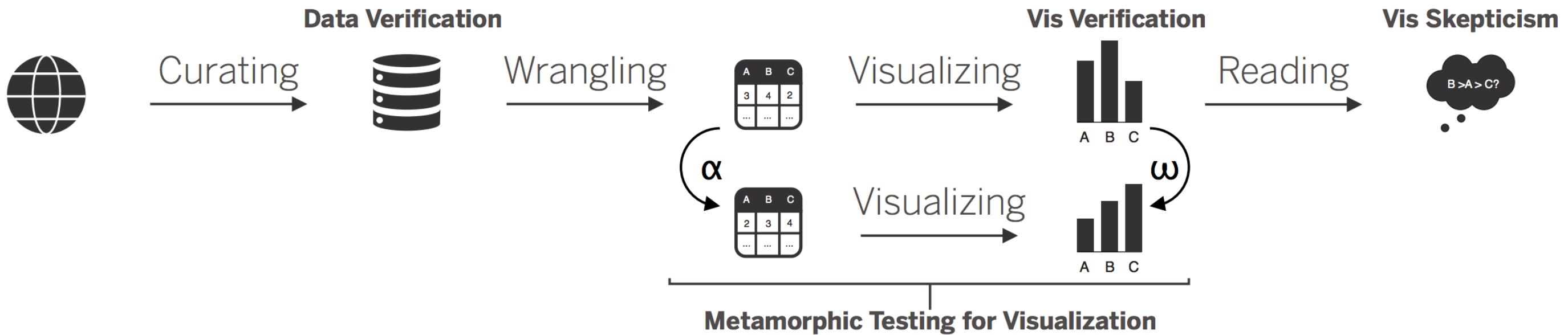
$$\begin{array}{ccccc} D_1 & \xrightarrow{r_1} & R_1 & \xrightarrow{\nu} & V_1 \\ \alpha \downarrow & & & & \downarrow \omega \\ D_2 & \xrightarrow{r_2} & R_2 & \xrightarrow{\nu} & V_2 \end{array}$$

## Metamorphic Testing

$$\forall x : p(f_I(x)) = f_O(p(x))$$

These framings turn out to be isomorphic!

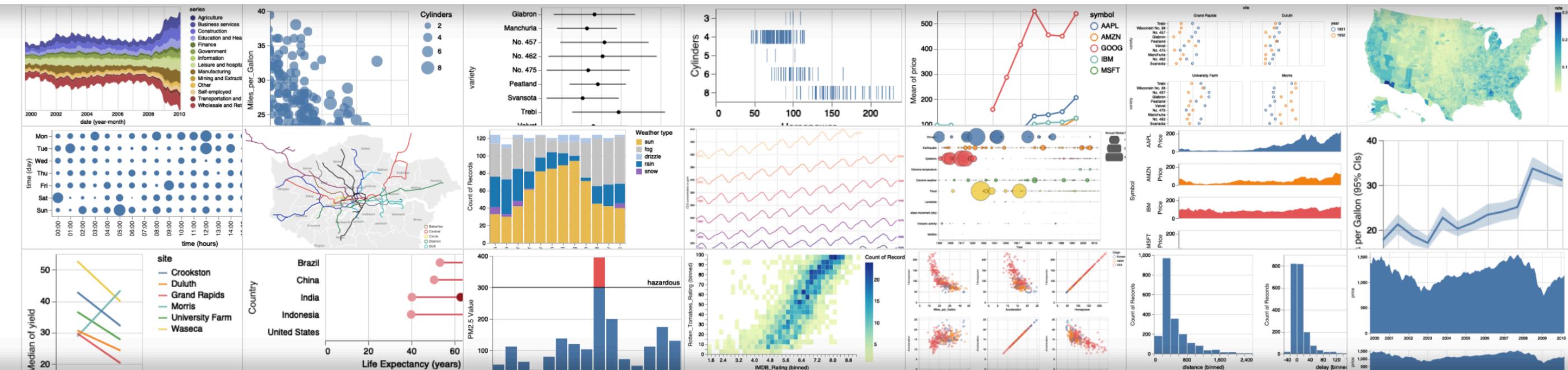
# Metamorphic Testing for Visualization


$$MTV ::= (Eq, \alpha, \omega) \Rightarrow (spec, data) \Rightarrow Boolean$$
$$MTV(Eq, \alpha, \omega)x = Eq(v(\alpha(x)), \omega(v(x)))$$

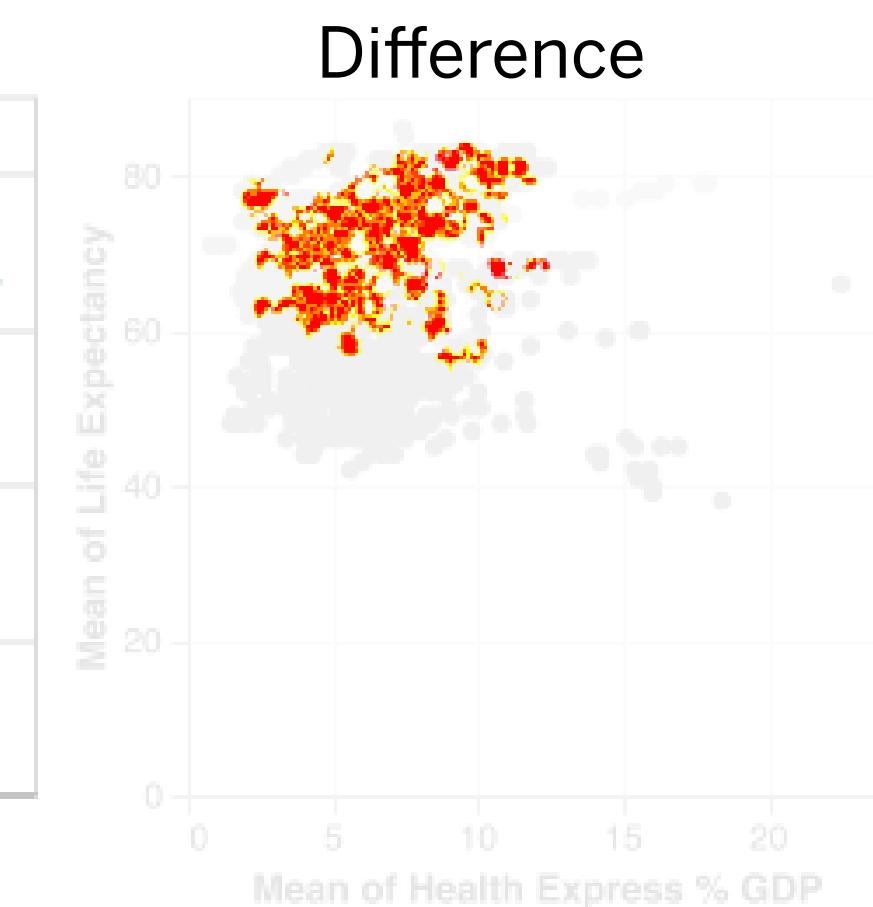
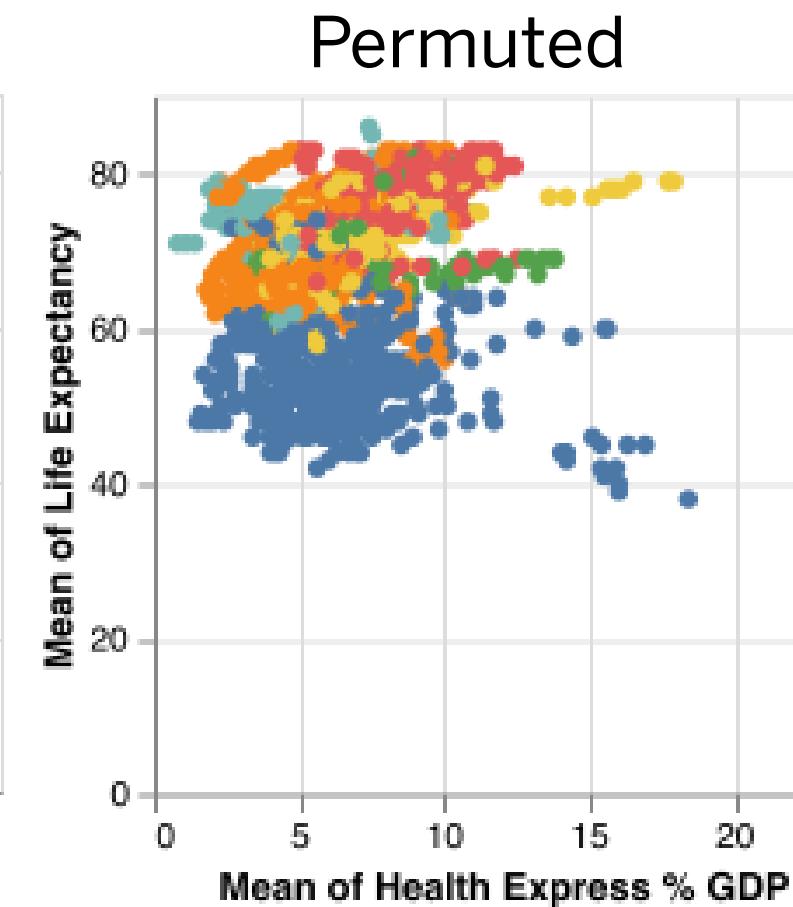
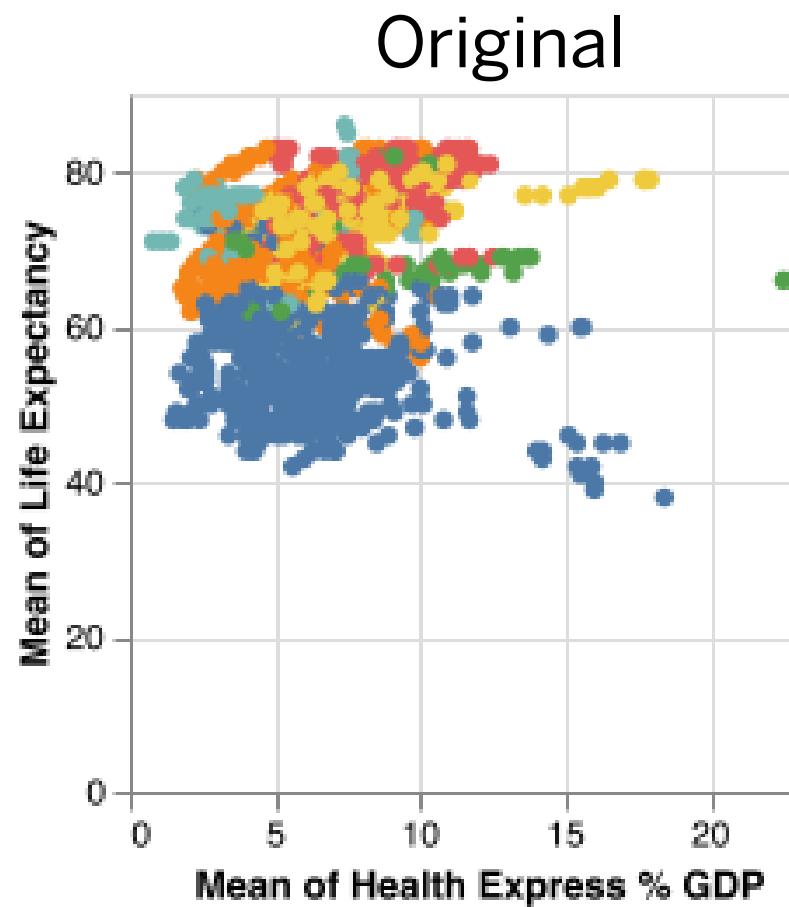
# Metamorphic Testing for Visualization

We build a proof-of-concept test based on vega-lite

## Vega-Lite – A Grammar of Interactive Graphics



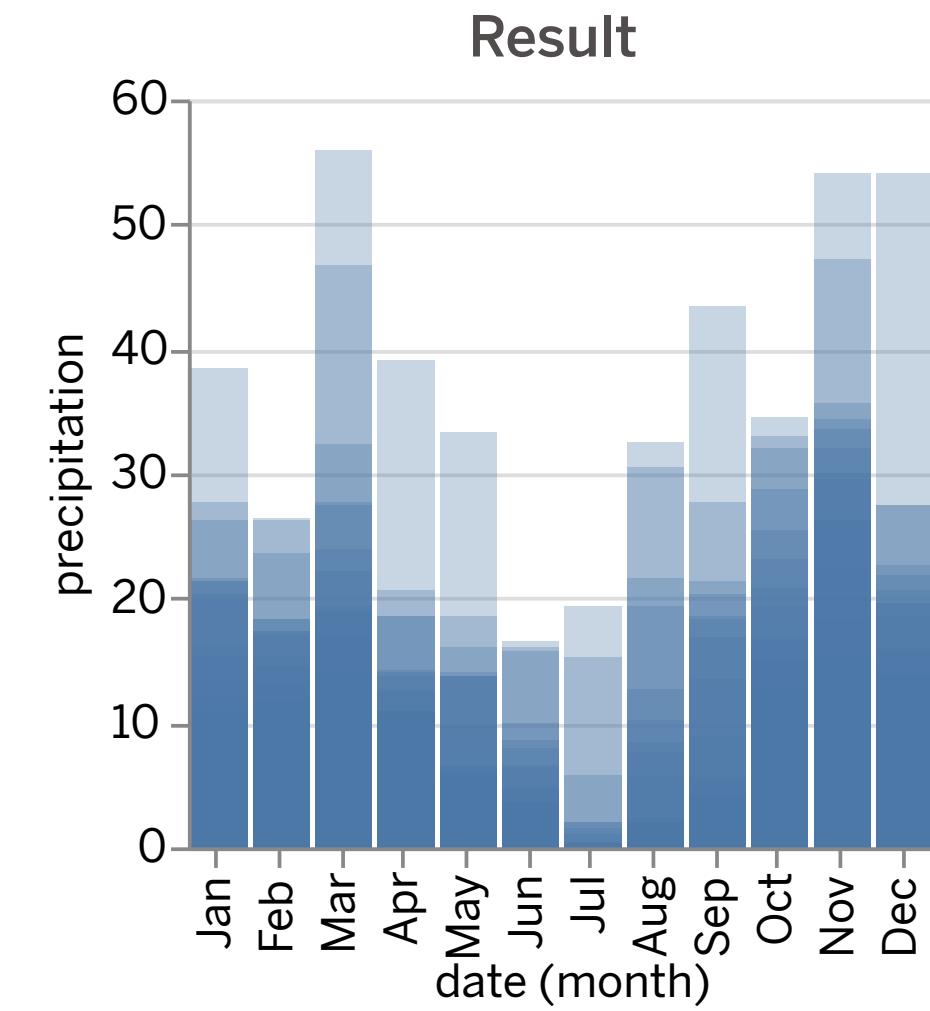
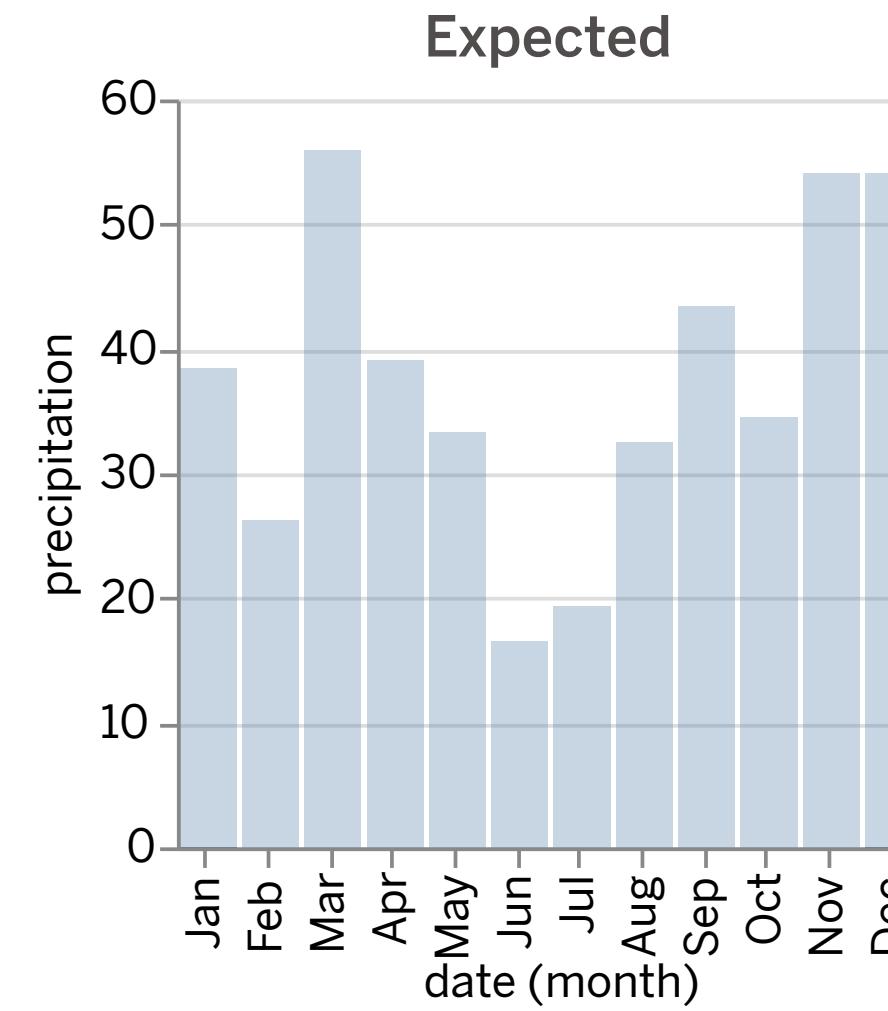
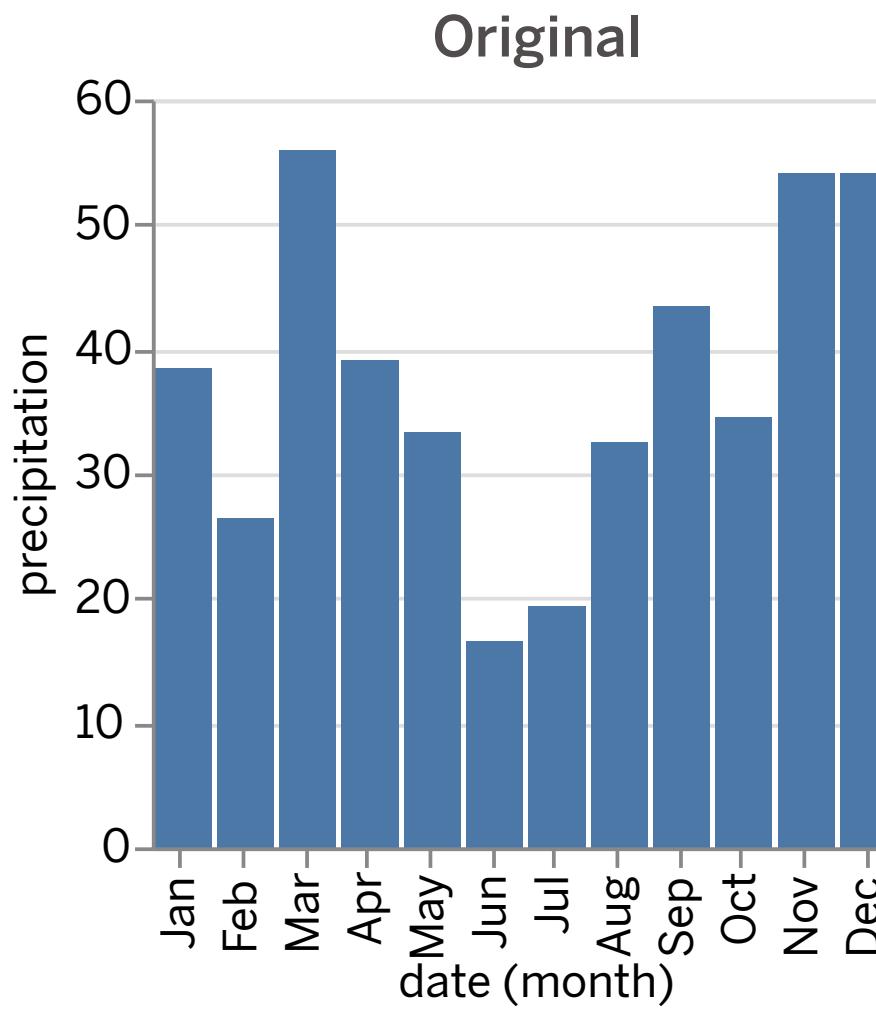
# Proof of Concept: SHUFFLE



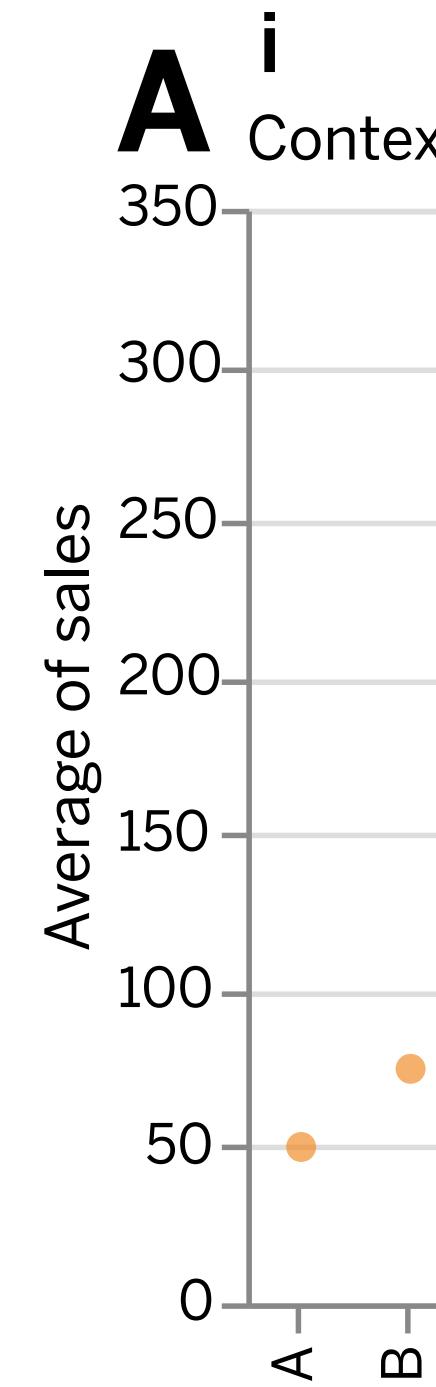
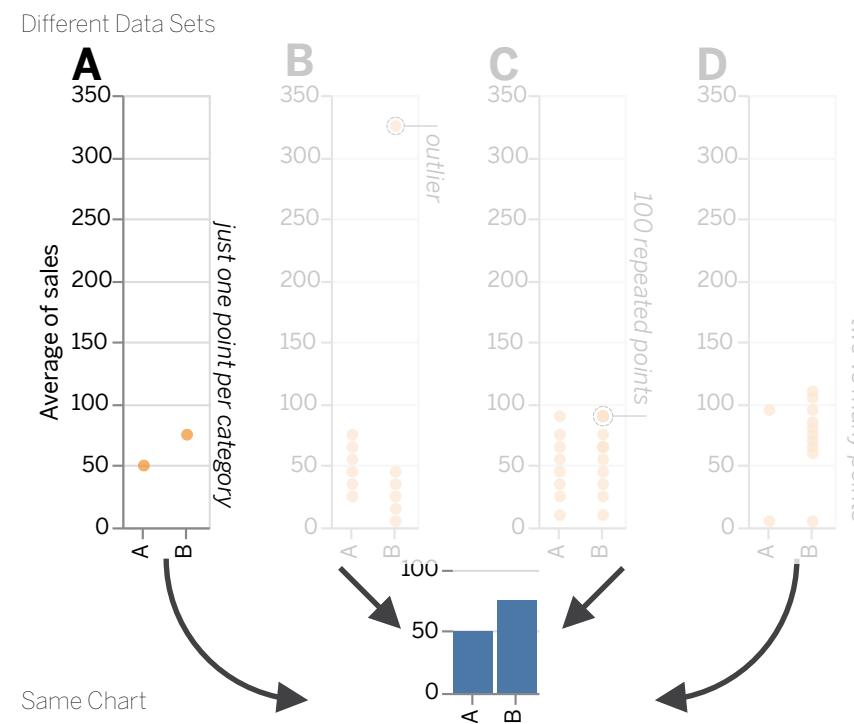
**Region**

- Africa
- Asia
- Europe
- Middle East
- Oceania
- The Americas

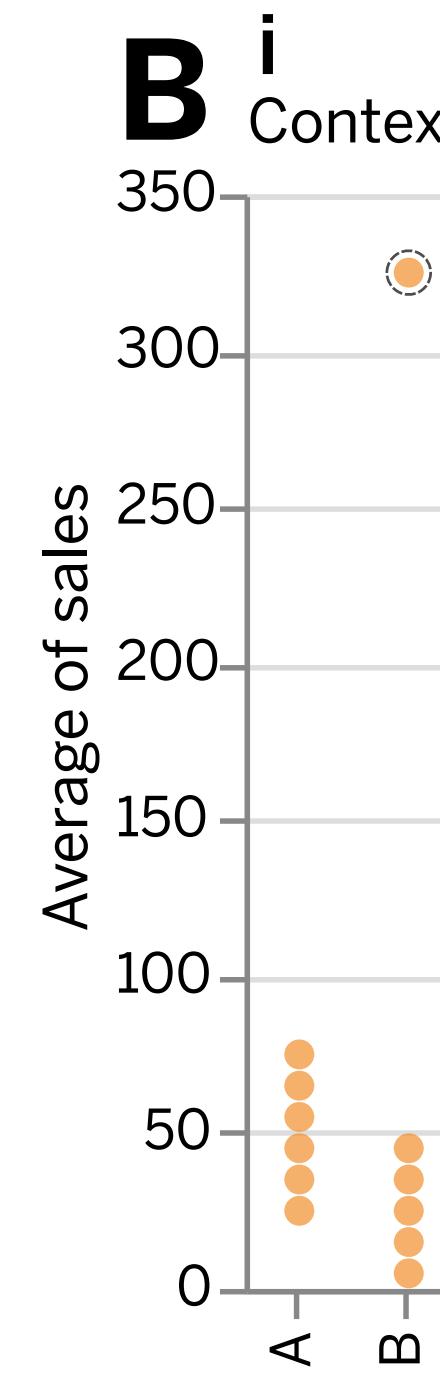
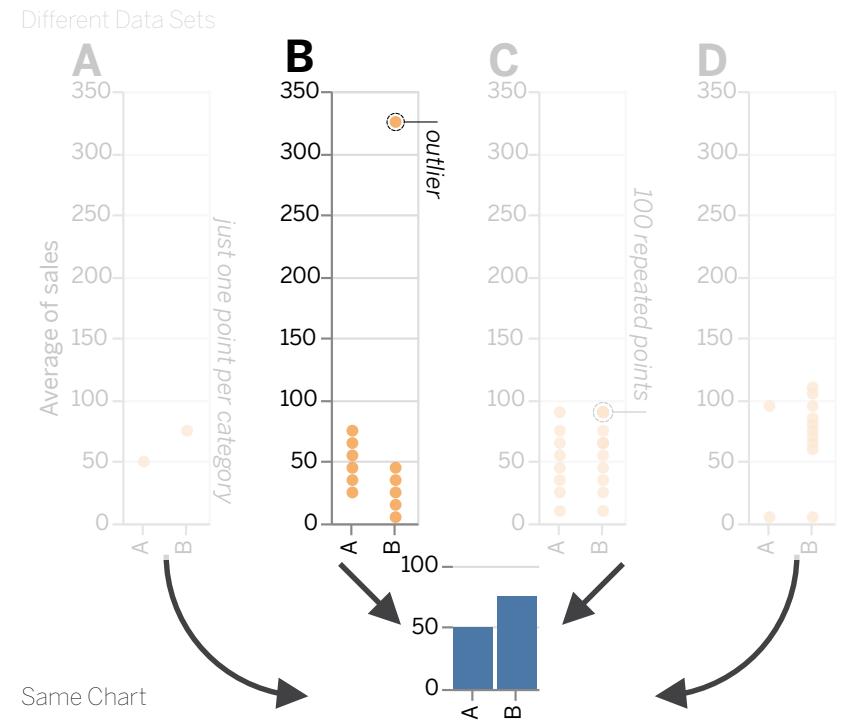
# Proof of Concept: CHANGE OPACITY



# Proof of Concept: BOOTSTRAP WITHIN AGGREGATE MARKS



# Proof of Concept: BOOTSTRAP WITHIN AGGREGATE MARKS



## Limitations

We find that this approach can detect a wide variety of visualization errors

Not all visualization mirages can be tested for analytically.



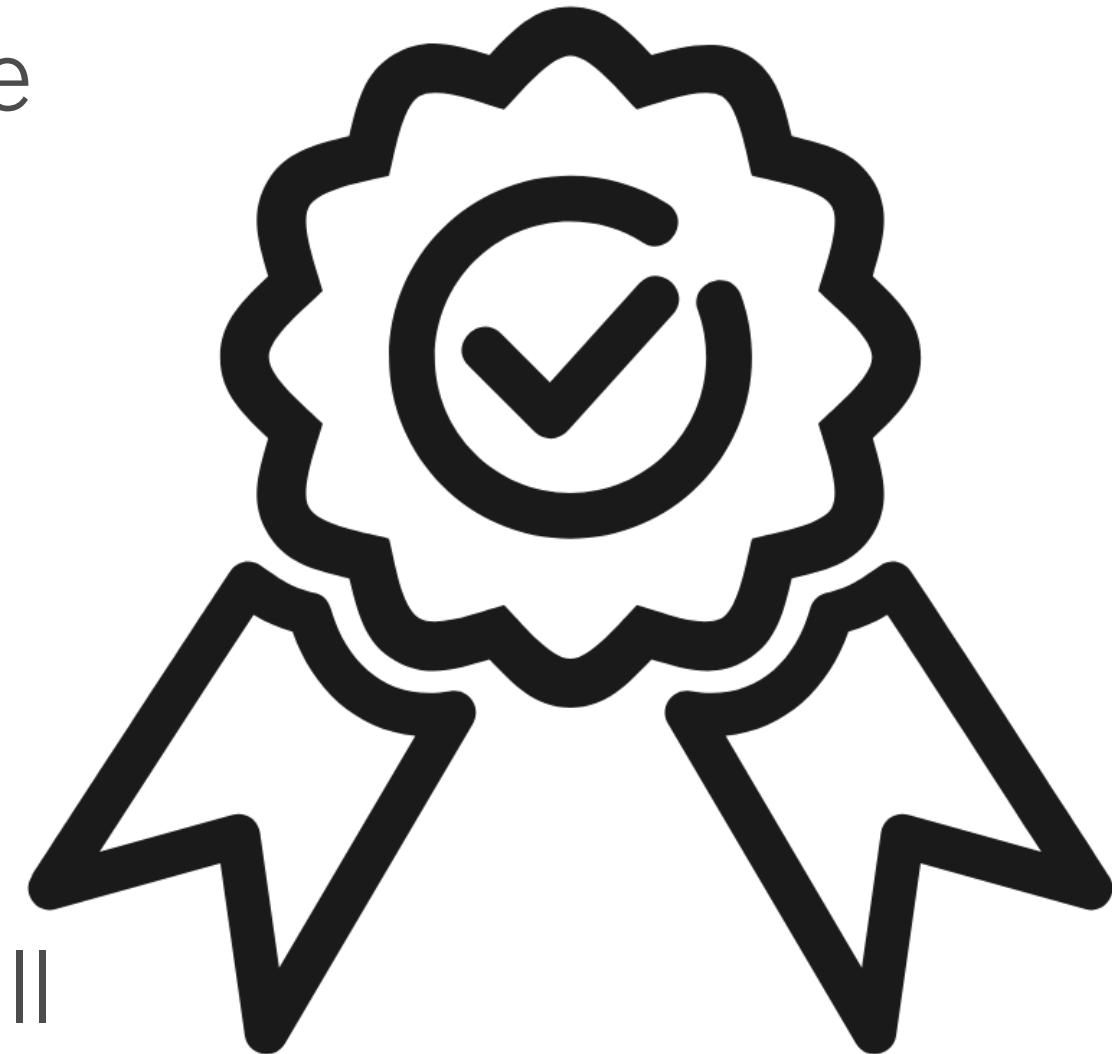
*(e.g. how do you automatically test for inequality)*

## Visualization Validation

Visual analytics systems should do more to protect their readers from noisy insights and themselves

MT4V & Mirages are the first steps towards a **visual analytic validation**

The right interface for this problem is still unknown, but linting for visualization looks promising

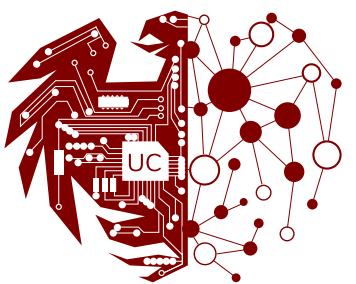


# Thanks!

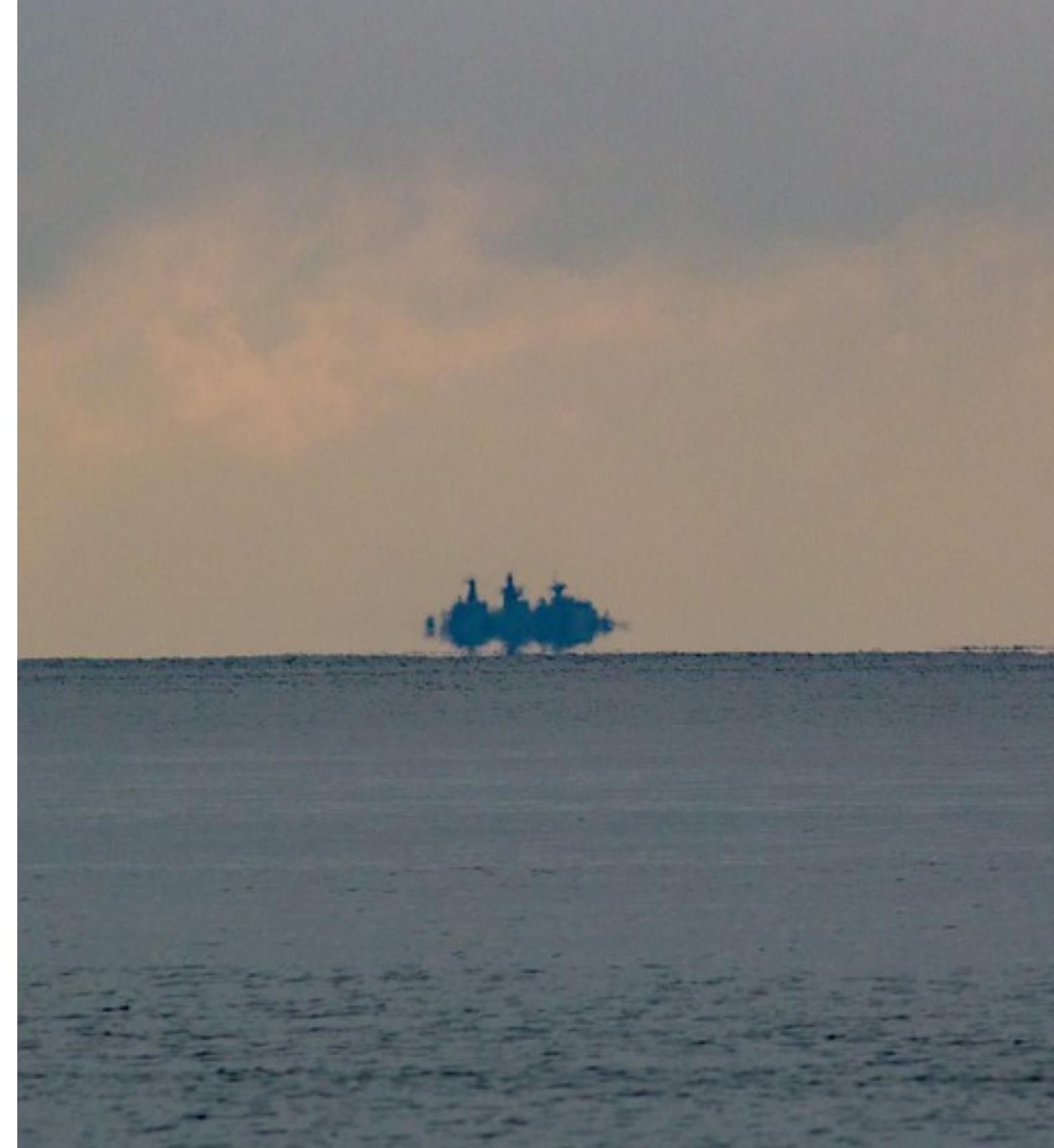
Happy to take any questions

Find out more at

[tinyurl.com/mirage-vis](http://tinyurl.com/mirage-vis)

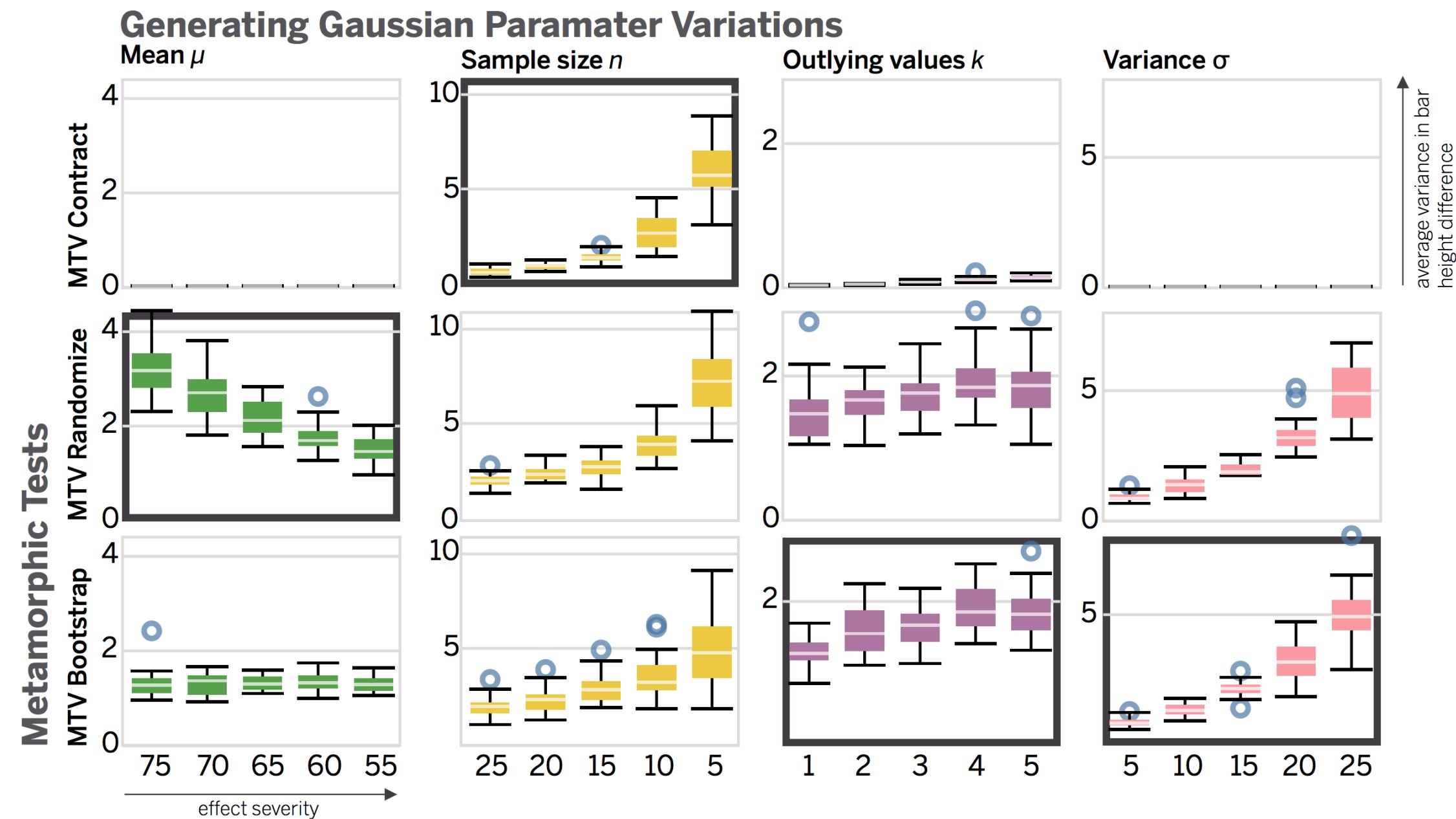
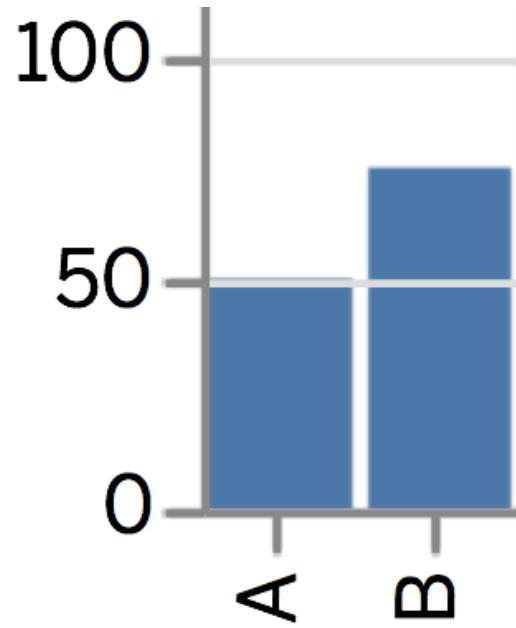


Ask me internet questions!  
 @\_mcnutt\_

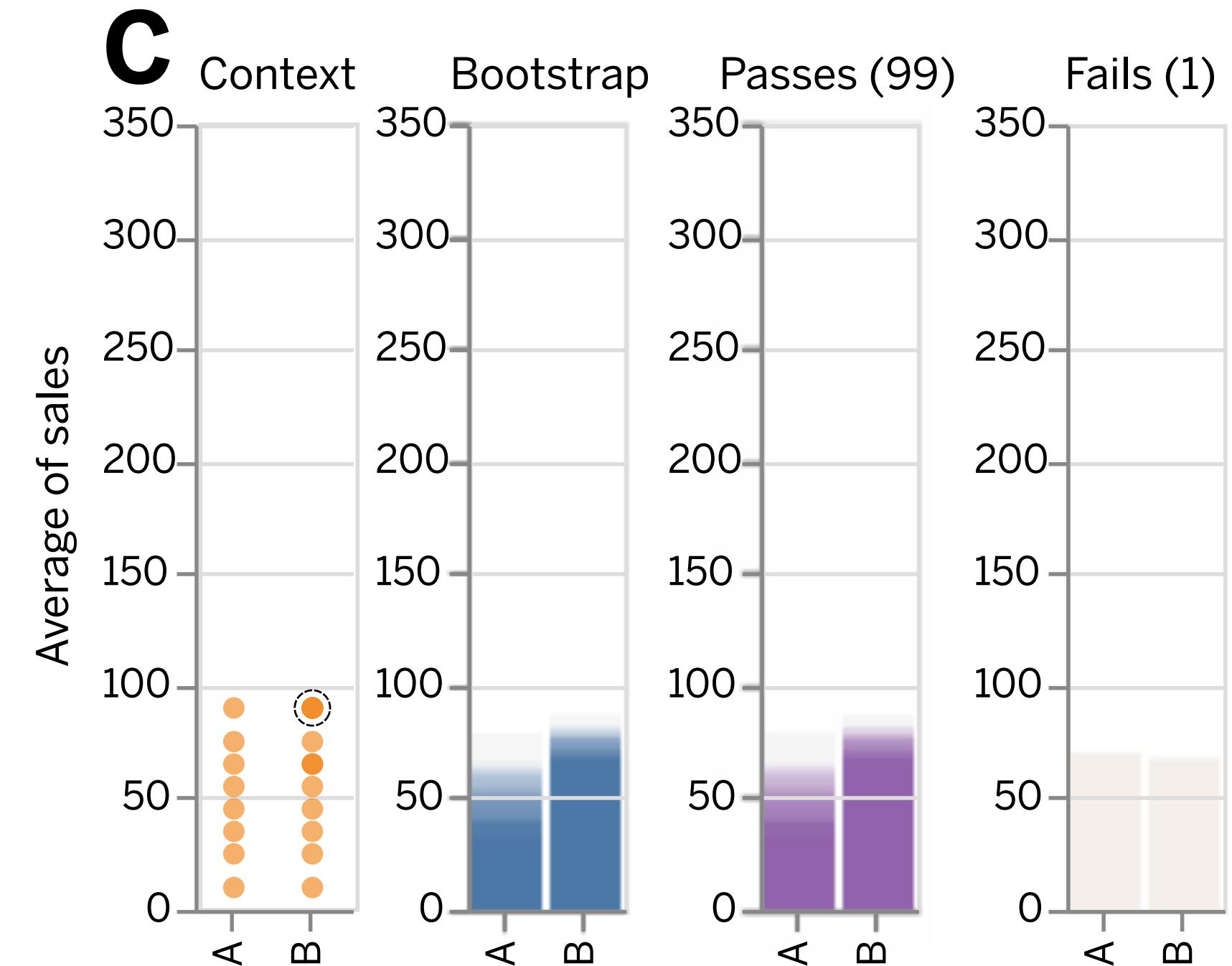
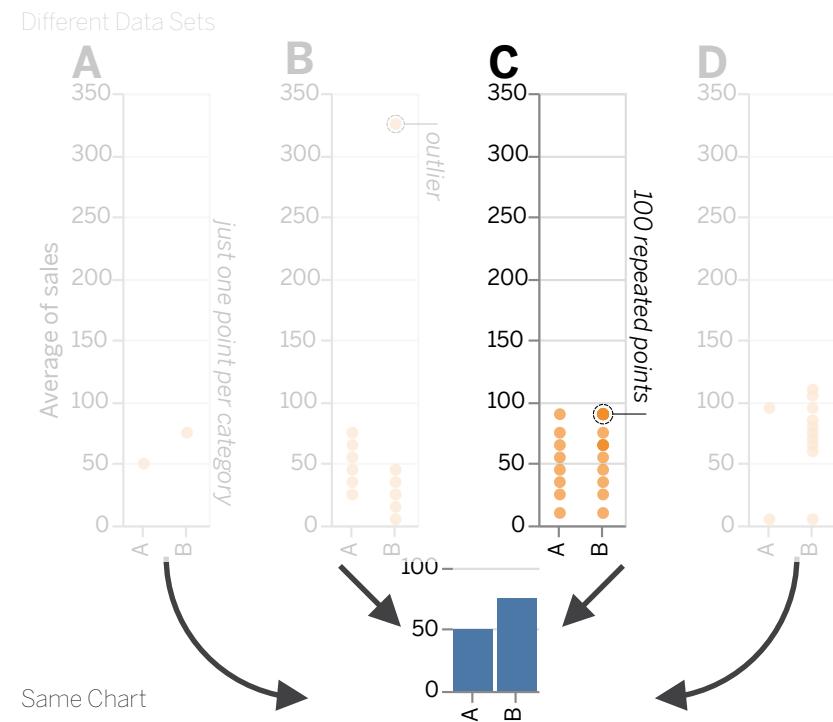


# Validation

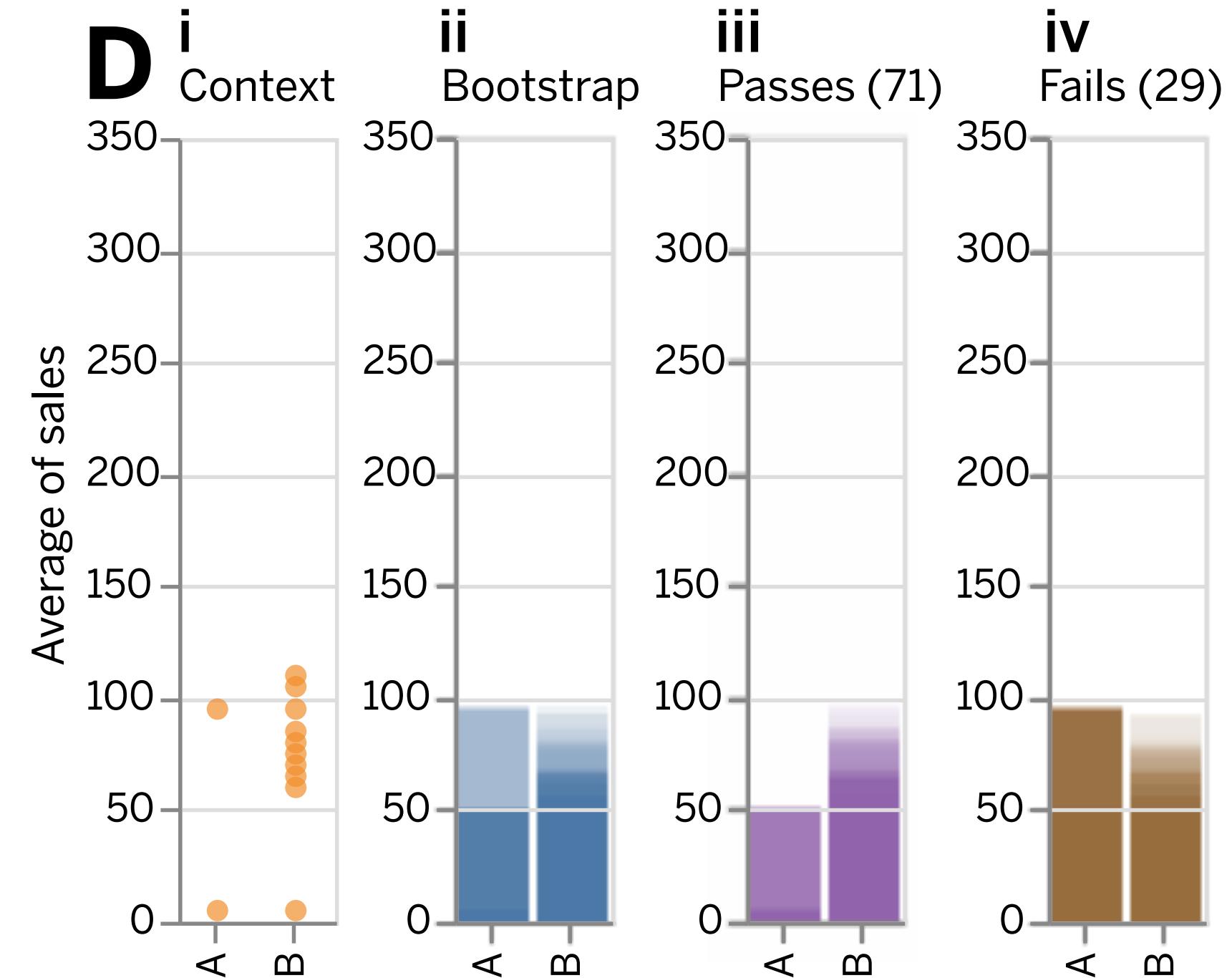
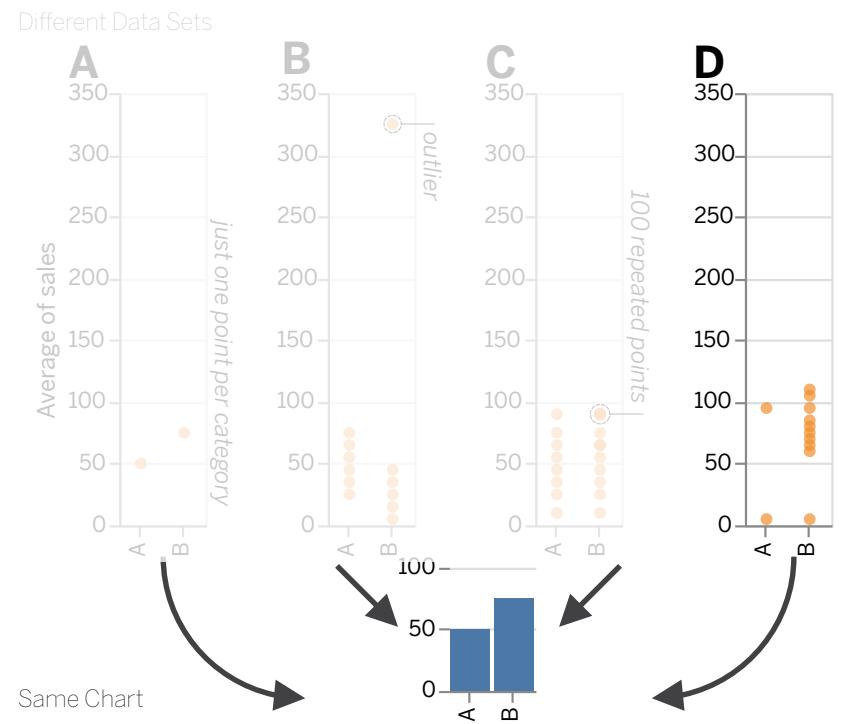
We validate this statistical technique through a simulation on 600 two class Gaussian bar charts



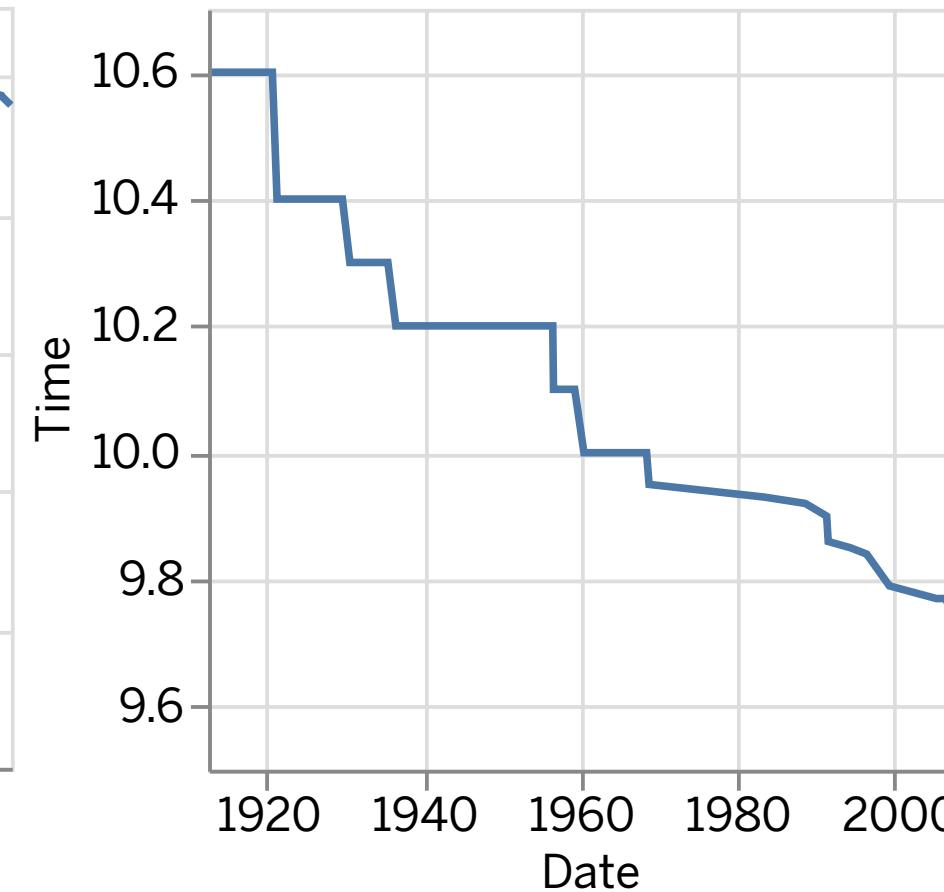
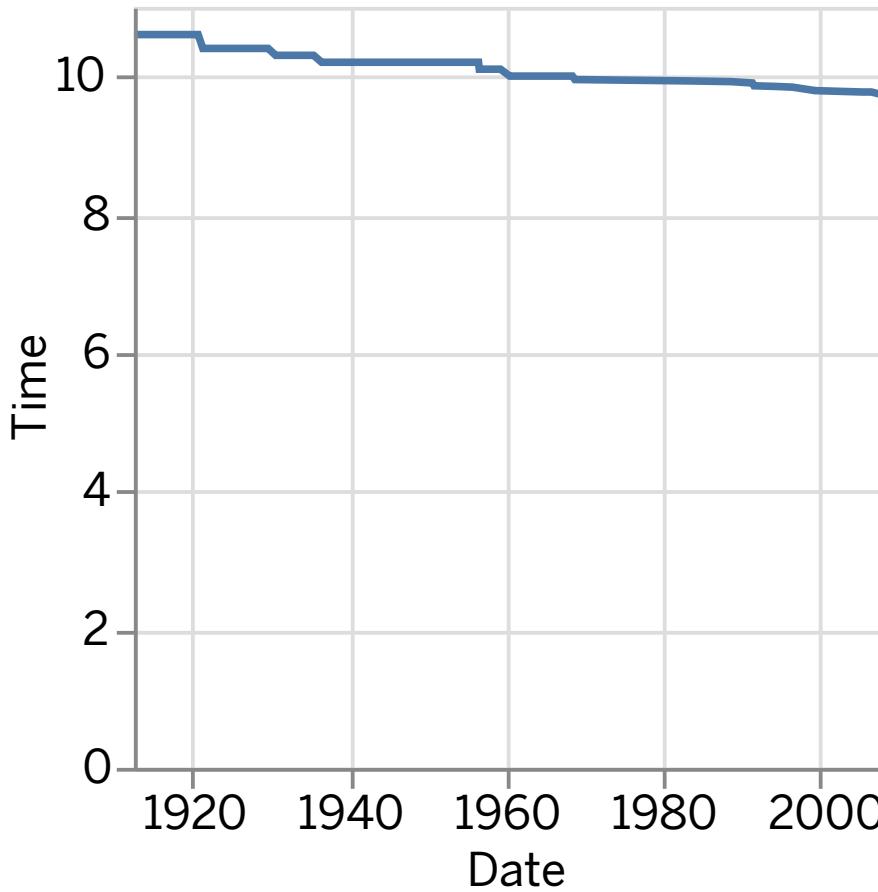
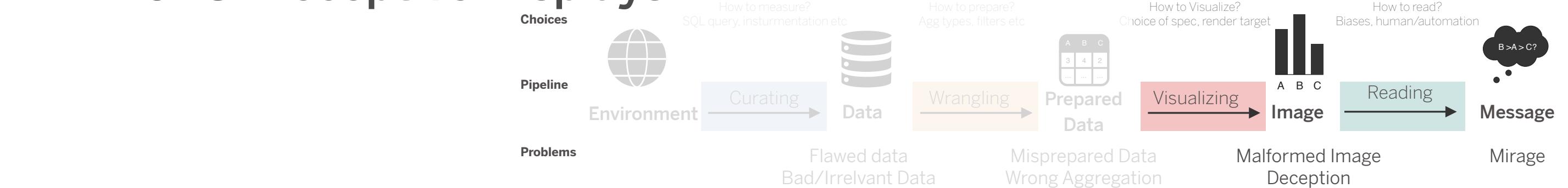
# Proof of Concept: BOOTSTRAP WITHIN AGGREGATE MARKS



# Proof of Concept: BOOTSTRAP WITHIN AGGREGATE MARKS



# MIRAGES: Deceptive Displays



Visualizations that **mis-use**  
**visual encodings** and  
**abuse reader assumption**  
in a way that confuses  
interpretation