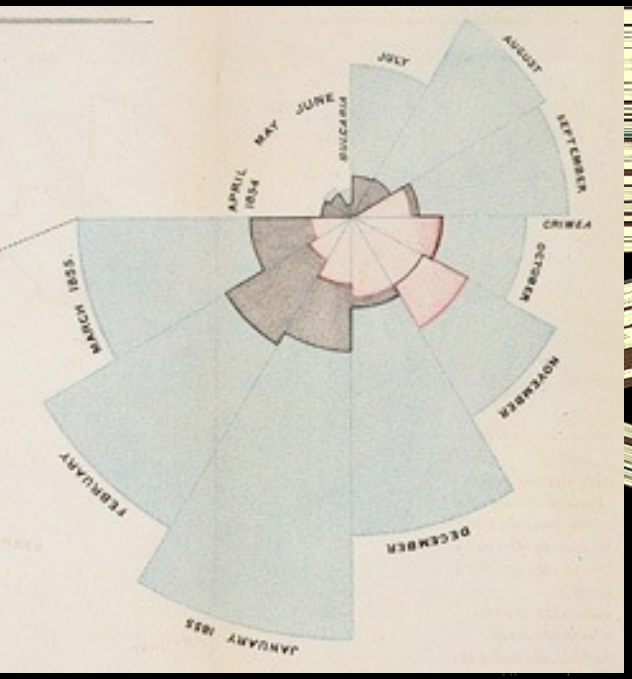
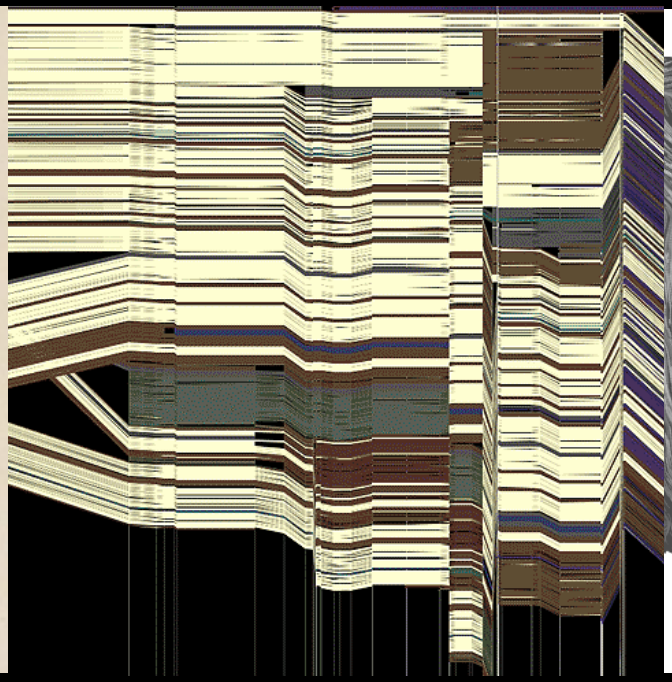
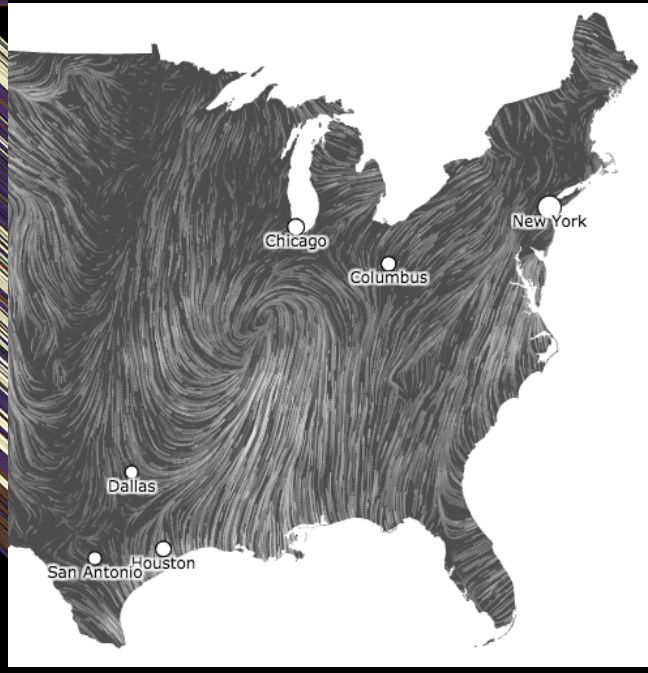
CSE 412 - Intro to Data Visualization

Course Summary







Jane Hoffswell University of Washington

Administrivia

Final Project Deliverables

Demonstration Video (<= 2 min)
Link due on Canvas by midnight **tonight** Wed 3/10

Communicate topics and project goals
Do: Show what viewers can learn from your page
Don't: Enumerate every feature of the page

Video should include: project name, team members' names, link to your website

For other tips, see the video production guide!

Final Project Deliverables

Demonstration Video (<= 2 min)
Link due on Canvas by midnight **tonight** Wed 3/10

Final Project Showcase
We will show demo videos in class, Fri 3/12.

Interactive Web Page & GitHub Repo
All materials online by midnight Mon 3/15.

Read assignment description for more!

Course Evaluation

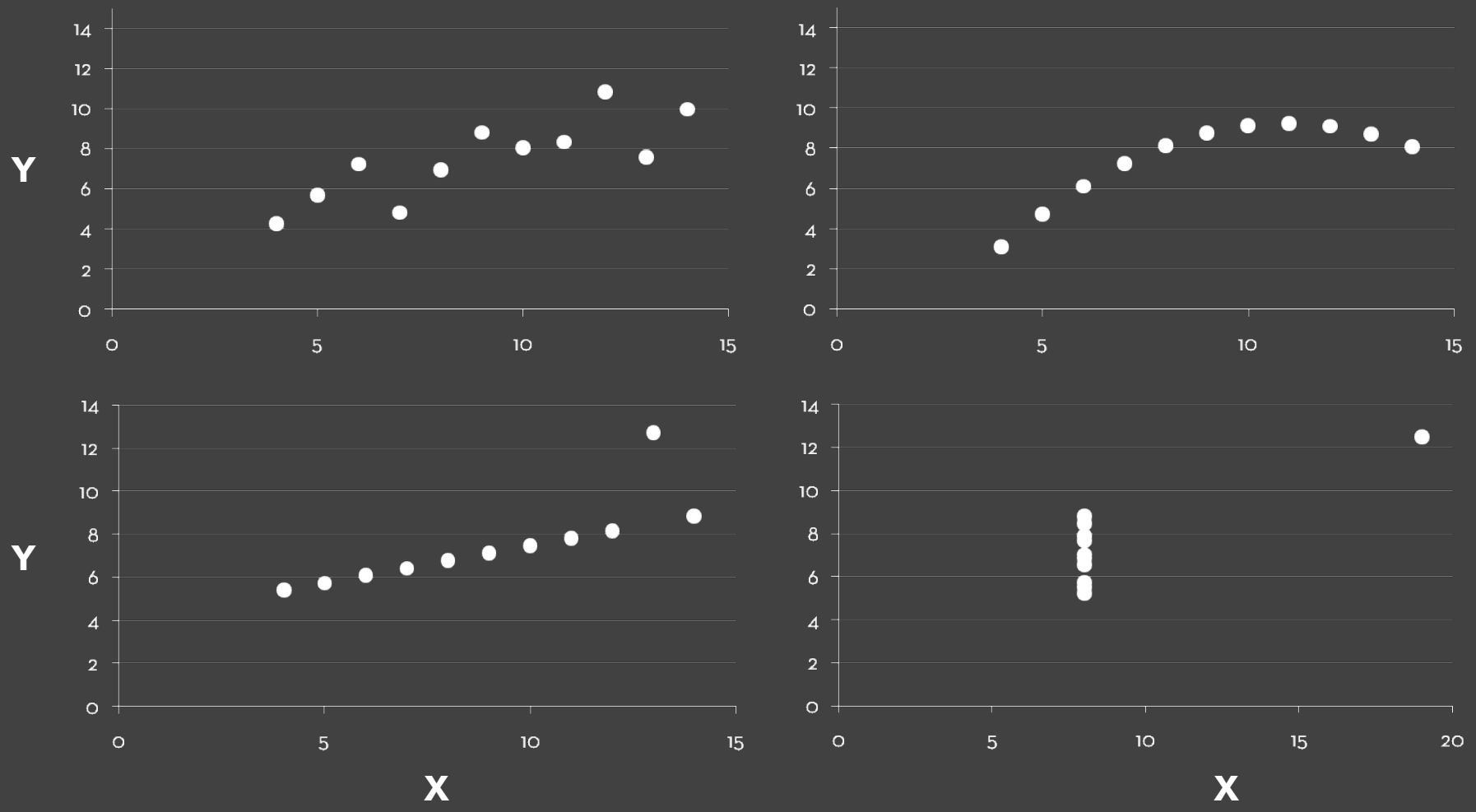
Official course evaluation, due by 3/14

Your opinion is valued!

https://uw.iasystem.org/survey/236202

Course Summary

Value of Visualization



Anscombe's Quartet [Anscombe 73]

Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Summary Statistics

$u_X = 9.0$ $\sigma_X = 3.317$

$u_Y = 7.5$ $\sigma_Y = 2.03$

Linear Regression

$Y = 3 + 0.5\ X$

$R^2 = 0.67$

[Anscombe 73]

The Value of Visualization

Record information

- Blueprints, photographs, seismographs, …

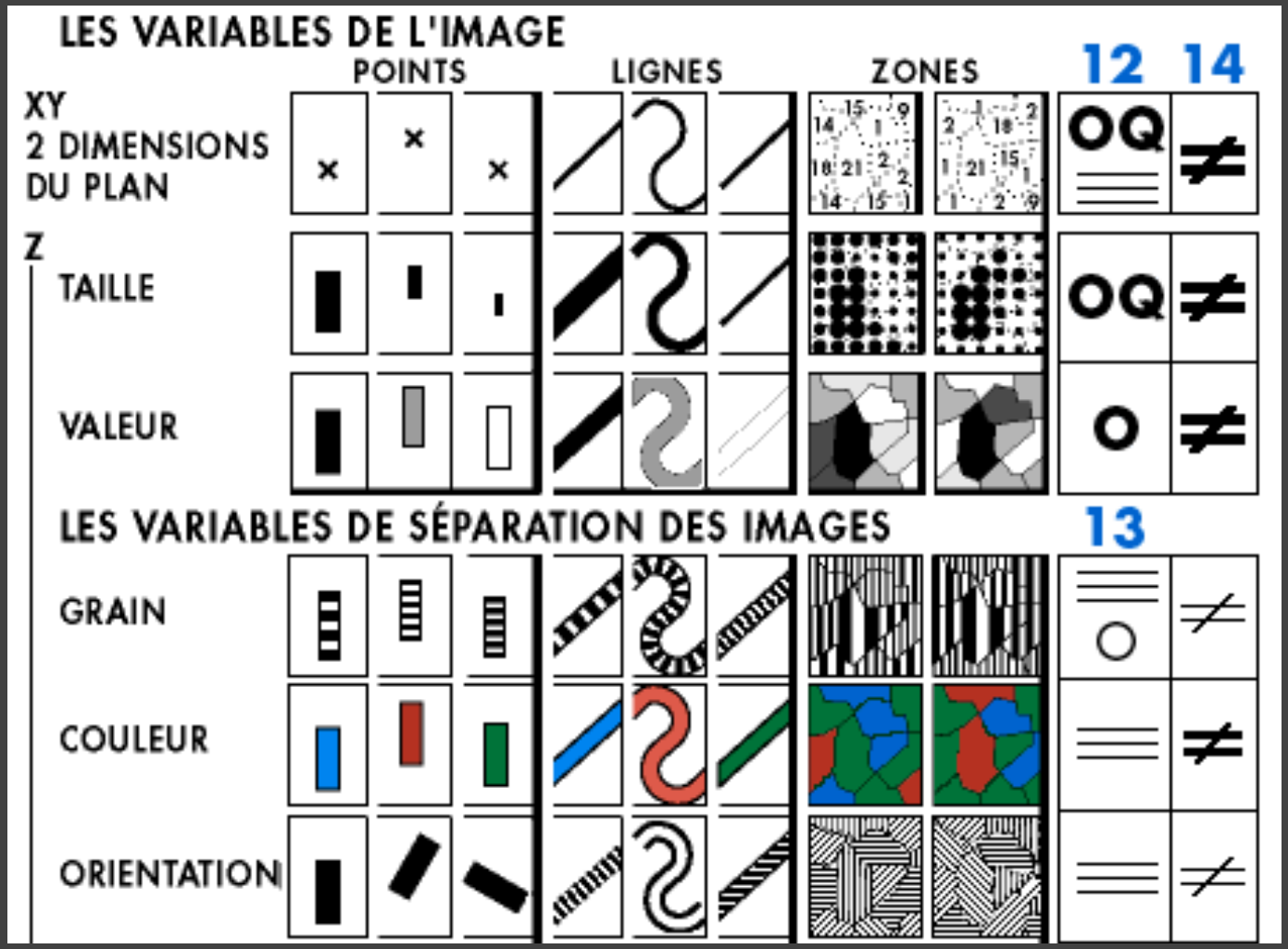
Analyze data to support reasoning

- Develop and assess hypotheses
- Find patterns / Discover errors in data
- Expand memory

Convey information

- Communicate, inform, inspire
- Collaborate and revise

Data and Image Models



Sémiologie Graphique [Bertin 67]

Nominal, Ordinal & Quantitative

N - Nominal (labels or categories) → =, ≠

- Fruits: apples, oranges, …

O - Ordered → =, ≠, <, >

- Quality of meat: Grade A, AA, AAA

Q - Interval (location of zero arbitrary) → =, ≠, <, >, –

- Dates: Jan, 19, 2006; Location: (LAT 33.98, LONG -118.45)
- Only differences (i.e., intervals) may be compared
- Can measure distances or spans

Q - Ratio (zero fixed) → =, ≠, <, >, –, %

- Physical measurement: Length, Mass, Time duration, …
- Counts and amounts
- Can measure ratios or proportions

Dimensions & Measures

Dimensions (~ independent variables)
Often discrete variables describing data (N, O)
Categories, dates, binned quantities

Measures (~ dependent variables)
Data values that can be aggregated (Q)
Numbers to be analyzed
Aggregate as sum, count, avg, std. dev…

Not a strict distinction. The same variable may be treated either way depending on the task.

Design Criteria [Mackinlay 86]

Expressiveness
A set of facts is *expressible* in a visual language if the sentences (i.e. the visualizations) in the language express all the facts in the set of data, and only the facts in the data.

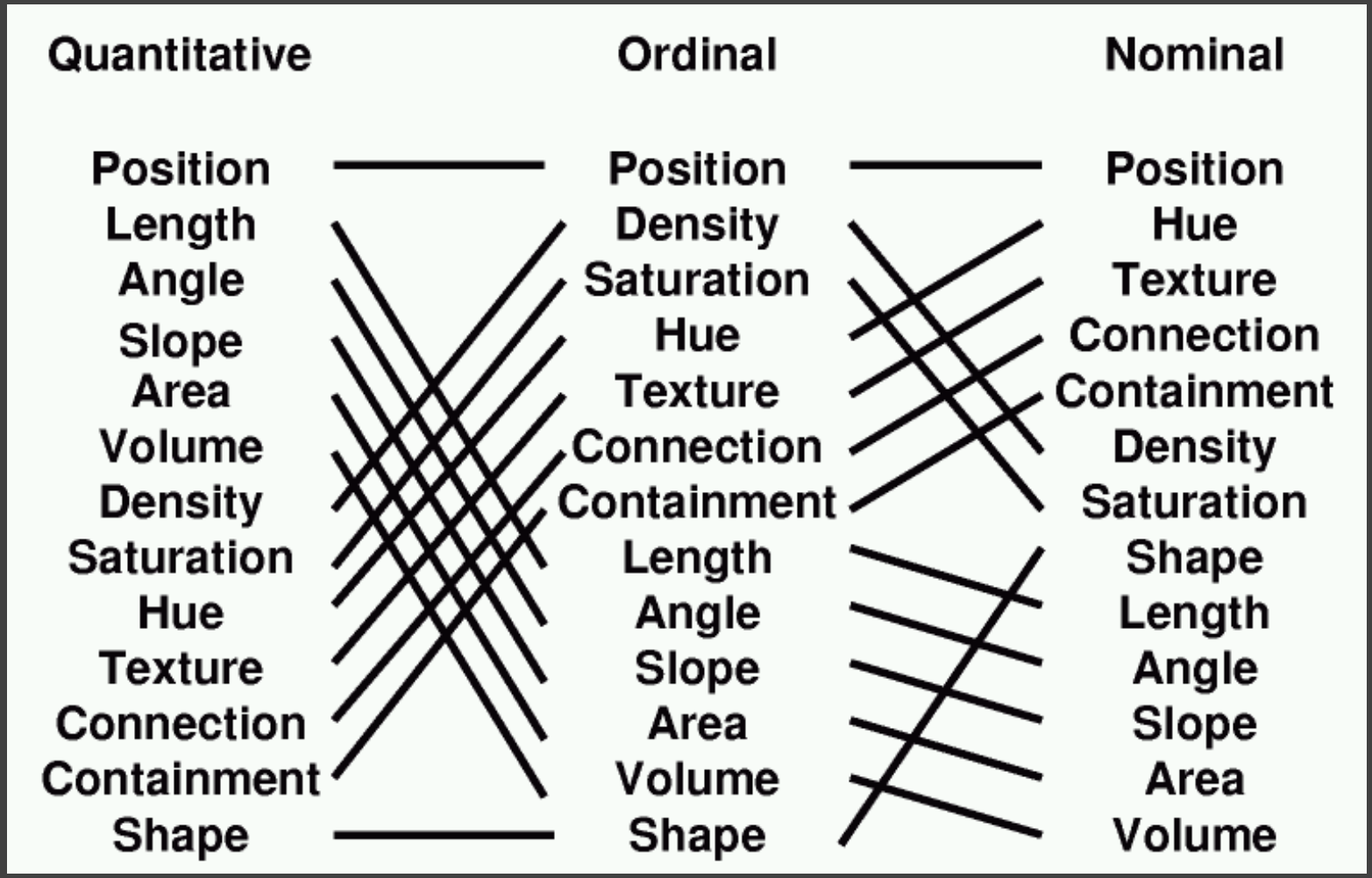
Effectiveness
A visualization is more *effective* than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.

Design Criteria *Translated*

Tell the truth and nothing but the truth
(don't lie, and don't lie by omission)

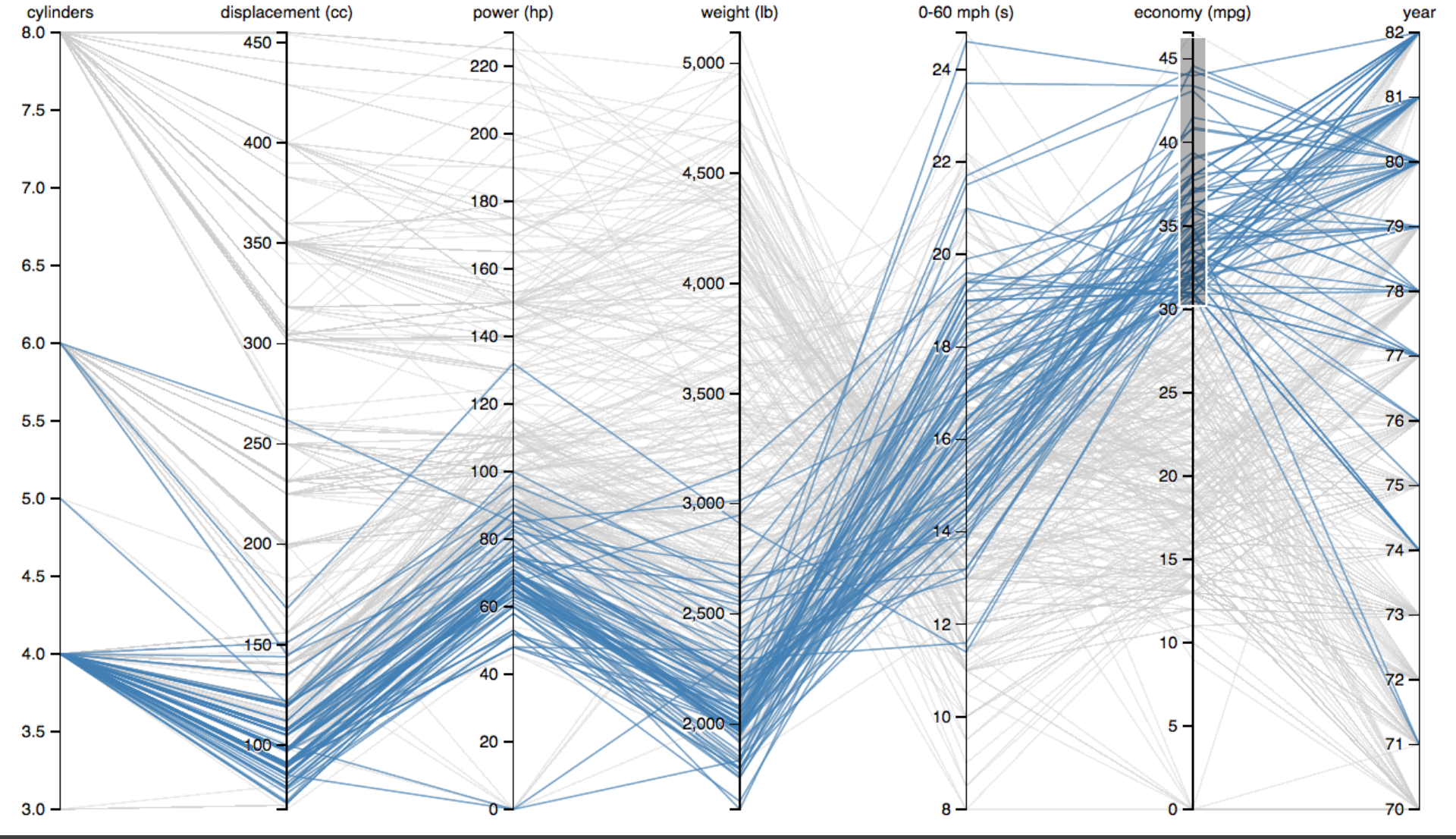
Use encodings that people decode better
(where better = faster and/or more accurate)

Mackinlay's Ranking



Conjectured *effectiveness* of encodings by data type

Exploratory Data Analysis



Data Quality

“The first sign that a visualization is good is that it shows you a problem in your data…

…every successful visualization that I've been involved with has had this stage where you realize, "Oh my God, this data is not what I thought it would be!" So already, you've discovered something.”

Martin Wattenberg

Exploratory Data Analysis Lessons

Check **data quality** and your **assumptions**.

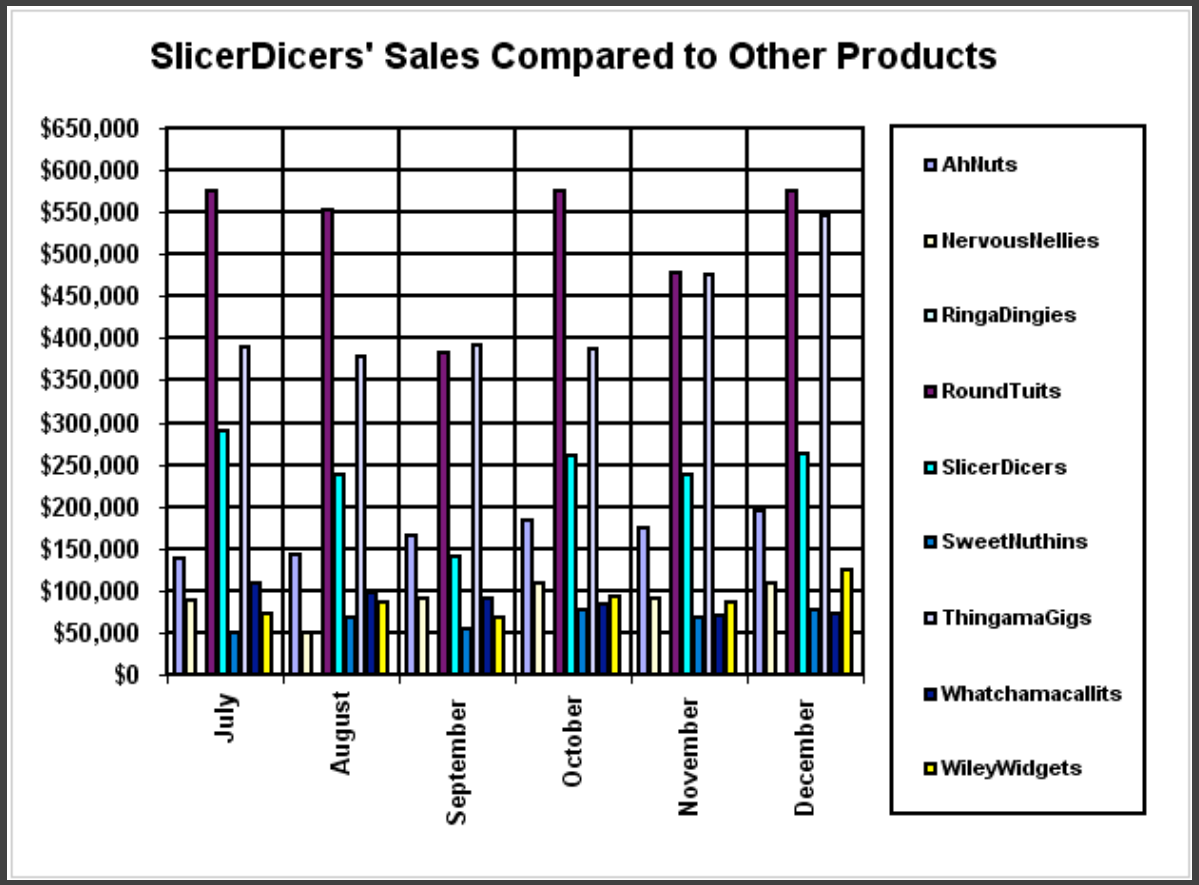
Start with **univariate summaries**, then start to consider **relationships among variables**.

Avoid premature fixation!

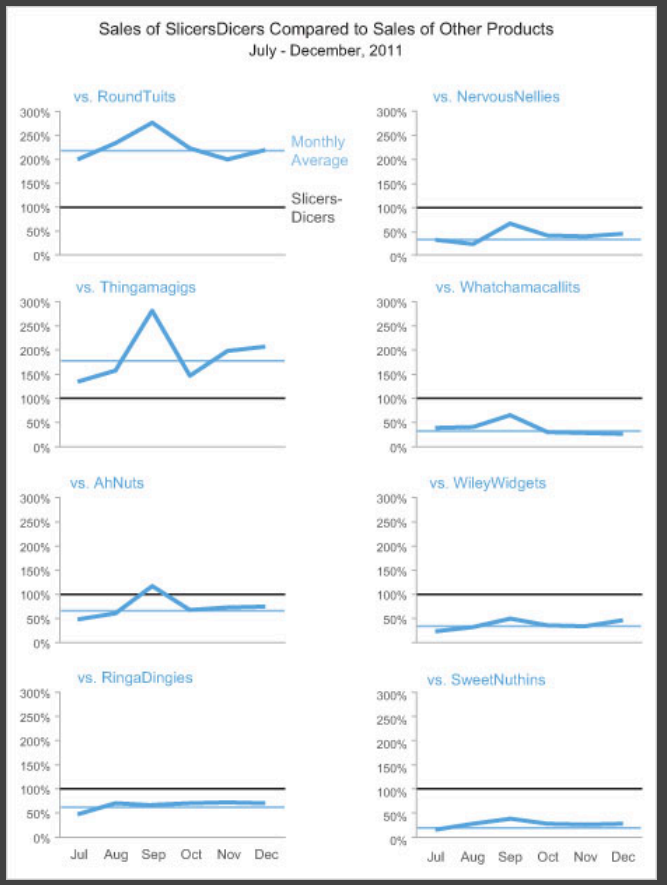
Transform data appropriately (e.g., invert, log)

Show data variation, not design variation [Tufte]

Visual Encoding and Design



Problematic design



Redesign

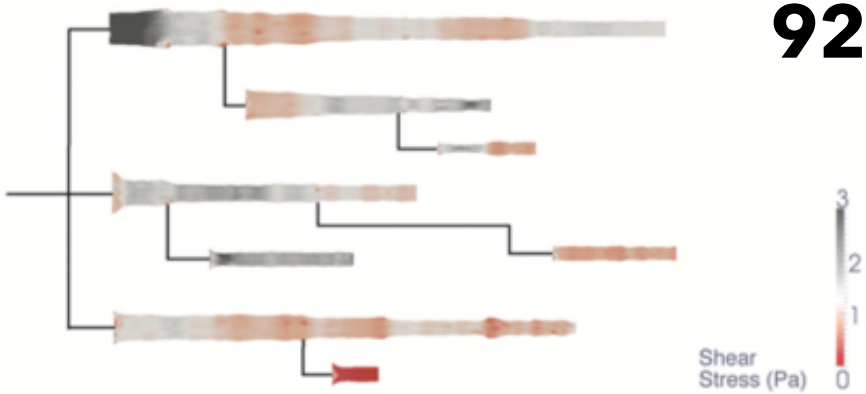
Artery Visualization [Borkin et al '11]

Rainbow Palette

Diverging Palette

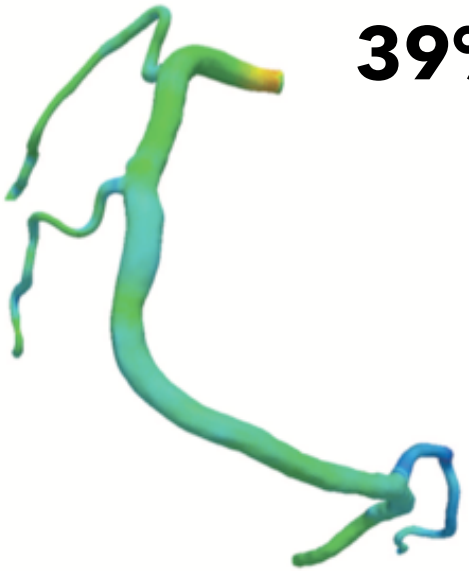
2D

62%

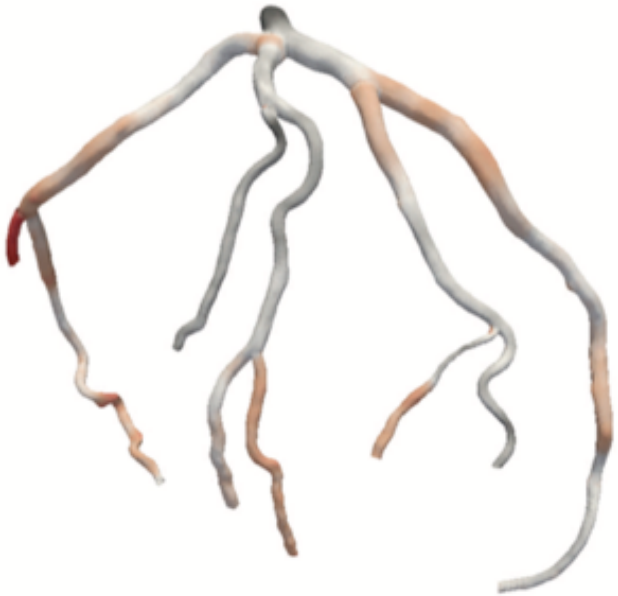


92%

3D



39%



71%

About the design process…

Visualization draws upon both science and art!

Principles like expressiveness & effectiveness are not hard-and-fast rules, but can assist us to guide the process and articulate alternatives.

They can lead us to think more deeply about our design rationale and prompt us to reflect.

It helps to know "the rules" in order to wisely bend (*or break*) them at the right times!

Narrative

Recent elections have placed a heavy emphasis on "swing states" — Ohio, Florida and the other competitive states. Y[illegible] the past, many more states shifted between the Democratic and Republican parties. A look at how the states st[illegible] they have shifted over past elections.

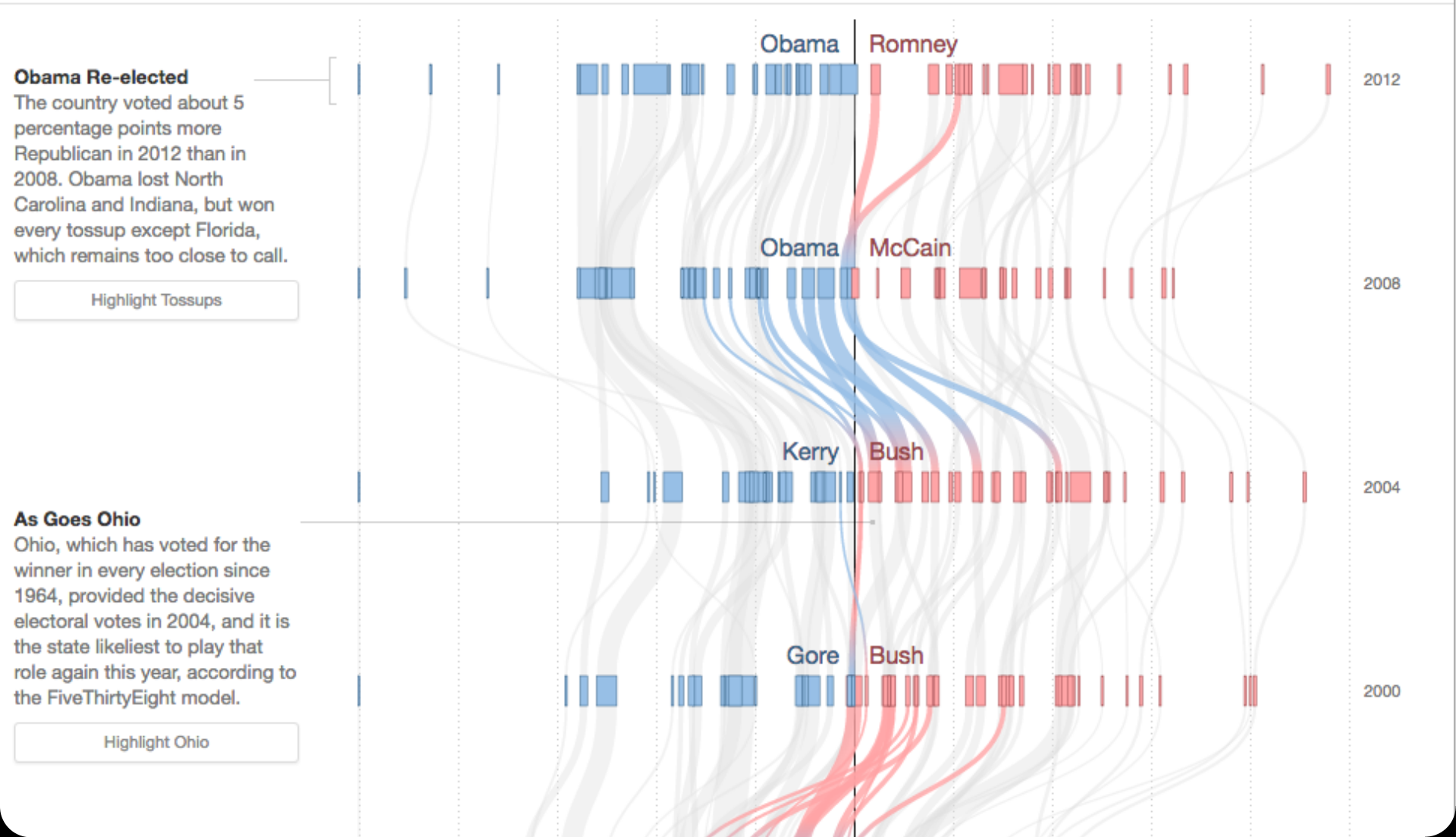
Each box represents a state sized by number of electoral votes.

Each curve shows how much it shifted left or right between elections

Chart **Size of Lead** | Chart **Electoral Votes**

← MORE DEMOCRATIC | MORE REPUBLICAN →

≥50% +40% +30% +20% +10% | +10% +20% +30% +40% ≥50%



Obama Re-elected
The country voted about 5 percentage points more Republican in 2012 than in 2008. Obama lost North Carolina and Indiana, but won every tossup except Florida, which remains too close to call.

Highlight Tossups

As Goes Ohio
Ohio, which has voted for the winner in every election since 1964, provided the decisive electoral votes in 2004, and it is the state likeliest to play that role again this year, according to the FiveThirtyEight model.

Highlight Ohio

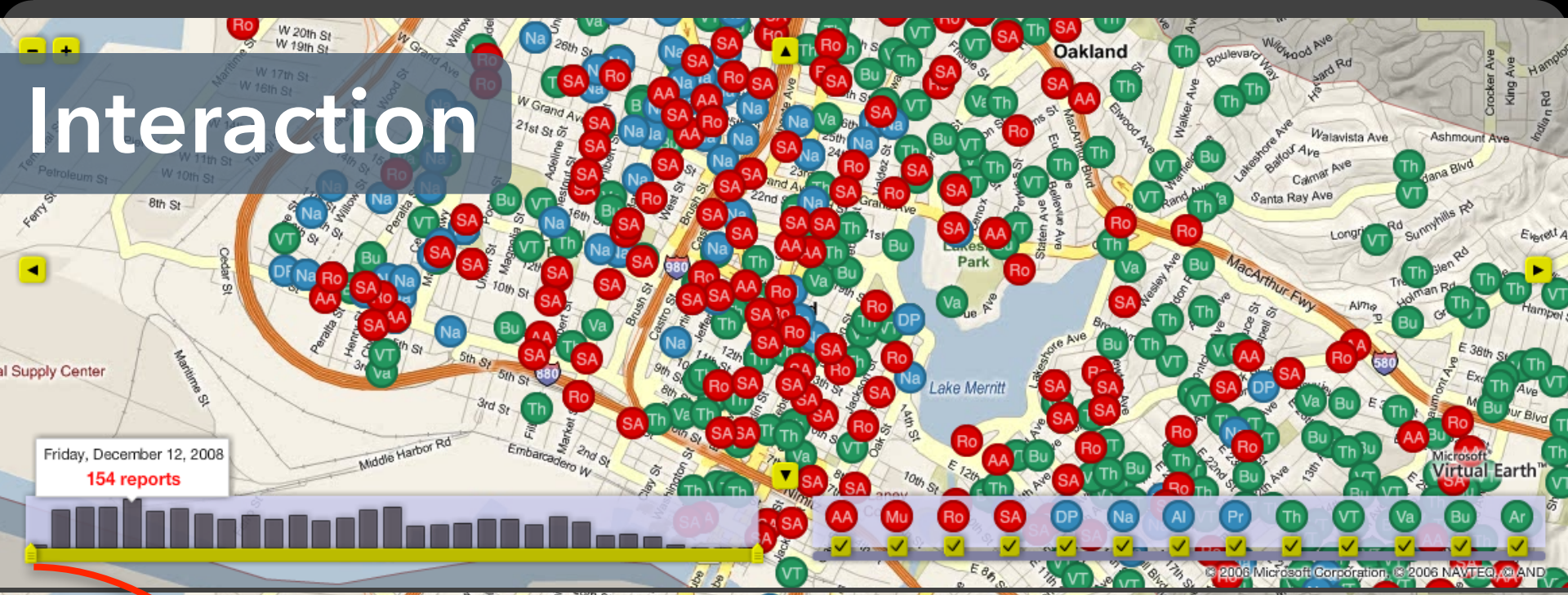
Narrative Storytelling

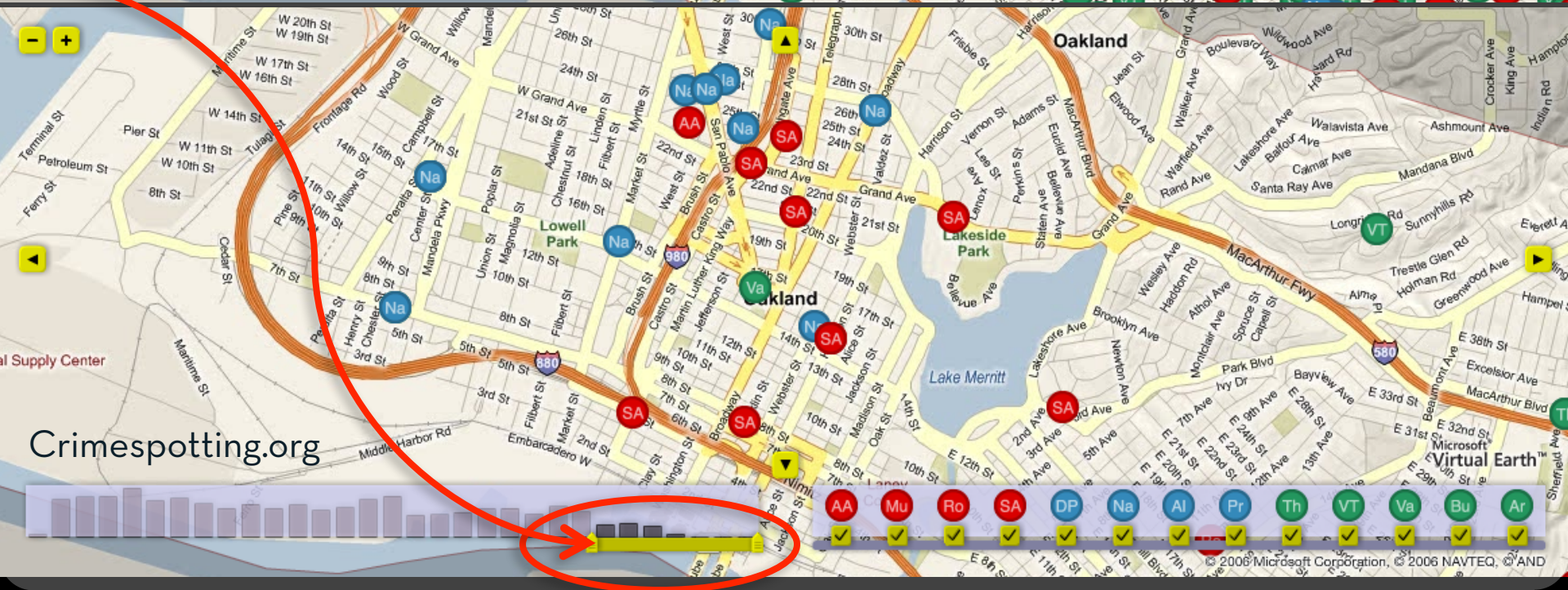
narrative (n): An account of a series of events, facts, etc., given in order and with the establishing of connections between them.

Effective storytelling “require[s] skills like those familiar to movie directors, beyond a technical expert’s knowledge of computer engineering and science.”

- Gershon & Page ‘01

Interaction





Crimespotting.org

Gulf of Execution

The difference between the user's intentions and the allowable actions.

Gulf of Evaluation

The amount of effort that the person must exert to interpret the state of the system and to determine how well the expectations and intentions have been met.

[Norman 1986]

Taxonomy of Interactions

Data and View Specification
Visualize, Filter, Sort, Derive

View Manipulation
Select, Navigate, Coordinate, Organize

Process and Provenance
Record, Annotate, Share, Guide

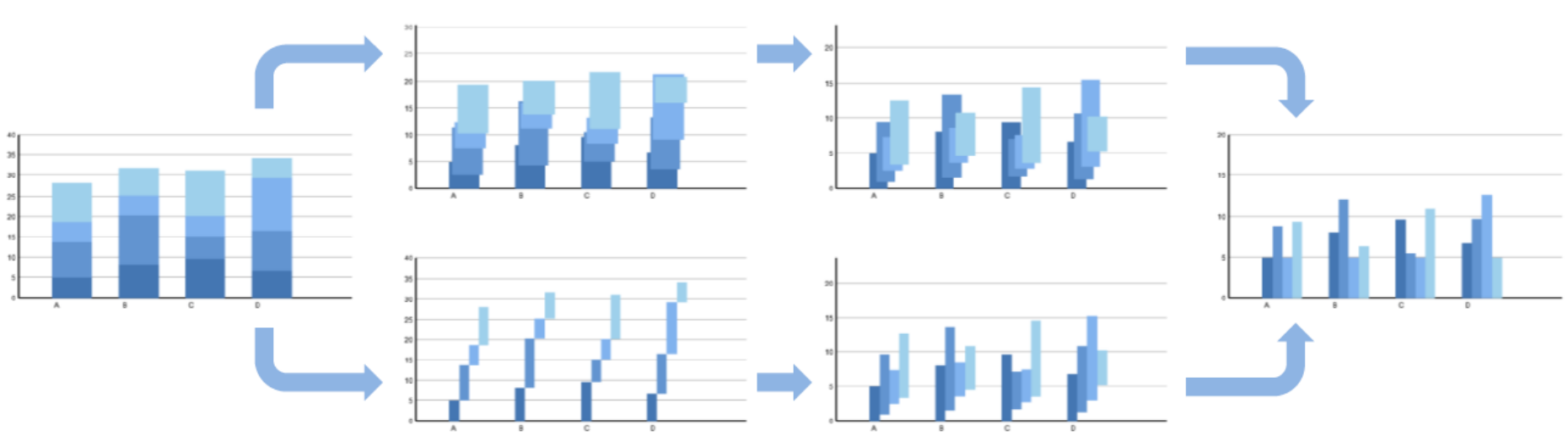
Interaction Takeaways

Most visualizations are interactive
Even passive media elicit interactions

Good visualizations are task dependent
Pick the right interaction technique
Consider the semantics of the data domain

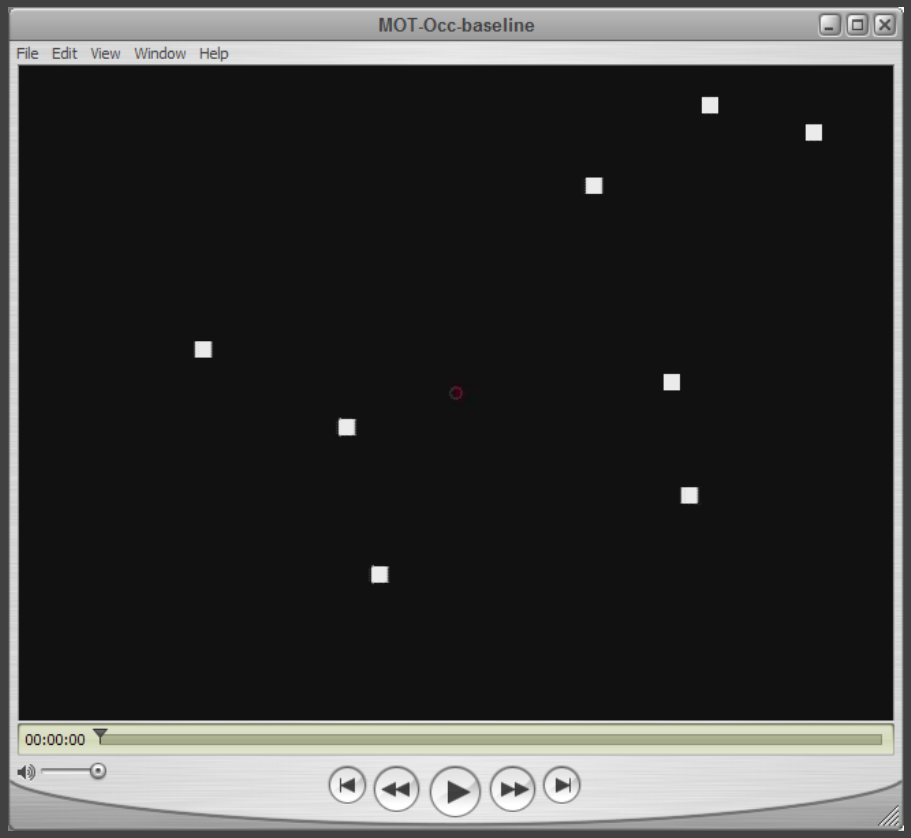
Fundamental interaction techniques
Selection / Annotation, Sorting, Navigation,
Brushing & Linking, Dynamic Queries

Animation



Animated transitions in statistical data graphics [Heer & Robertson 07]

Tracking Multiple Targets



How many dots can we simultaneously track?

~4-6. Difficulty increases sig. at 6. [Yantis 92, Pylyshn 88, Cavanagh 05]

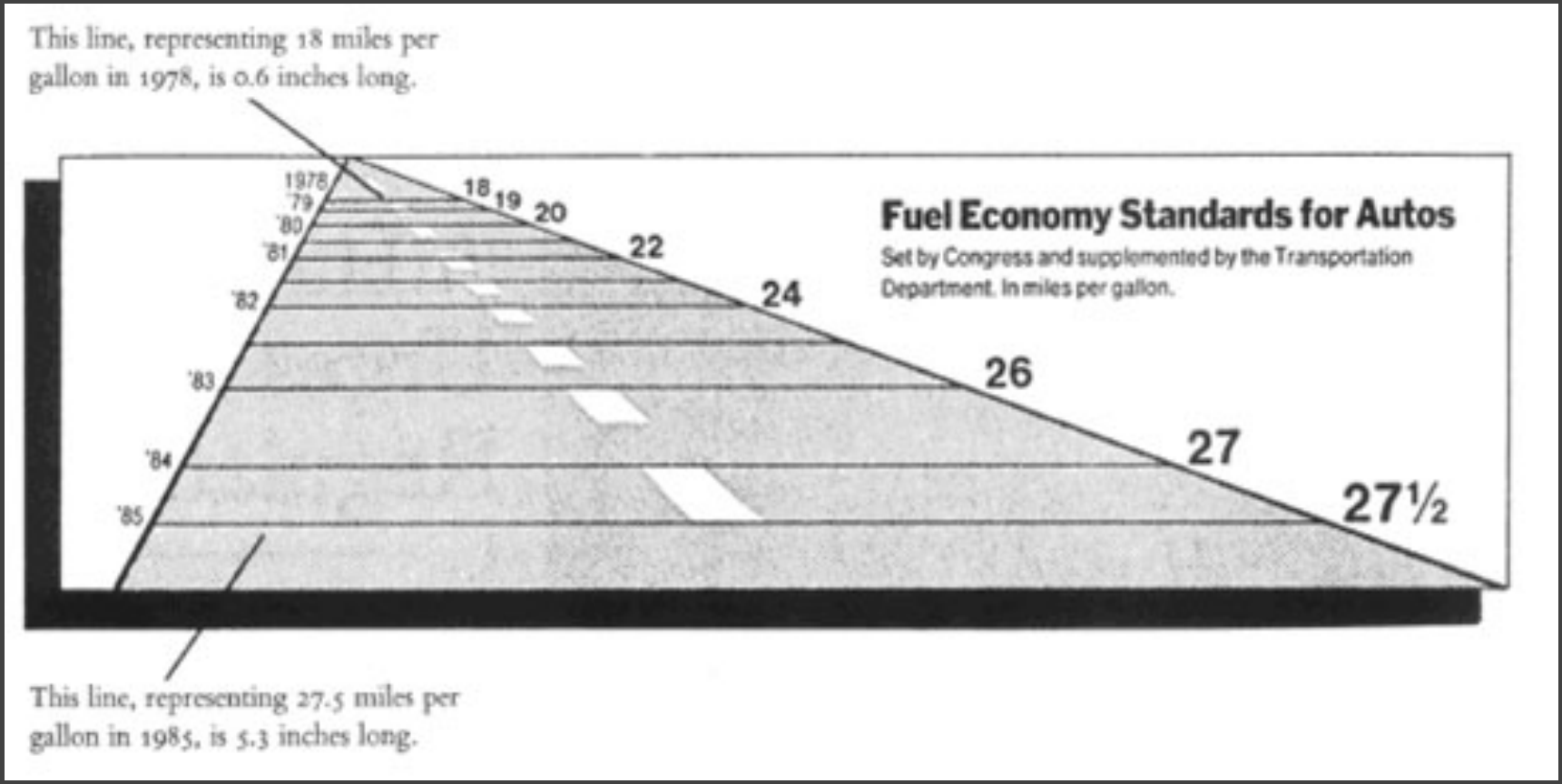
Animation Takeaways

Animation is a salient visual phenomenon
Attention, object constancy, causality, timing
Design with care: congruence & apprehension

For processes, **static images** may be preferable

For transitions, animation has demonstrated benefits, but **consider task and timing**

Ethical & Deceptive Visualization



Tufte's Lie Factor, original visualization from the NYT [Tufte 01]

Deceptive Visualization

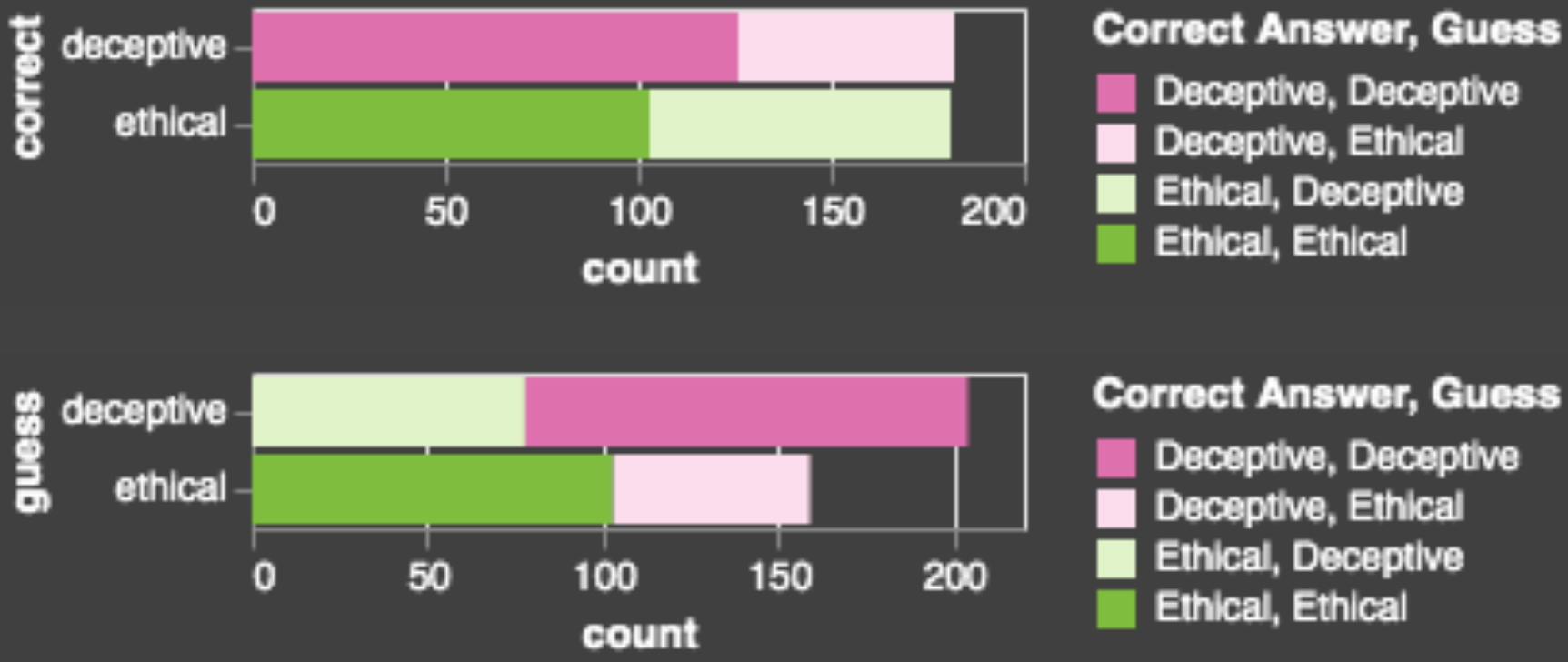
Lie Factor: Distorting the apparent size of the effect in your data, often through choosing ambiguous or non-standard encodings.

Scale Manipulation: Changing with the scales of your chart to minimize, magnify, or invert the change in the data.

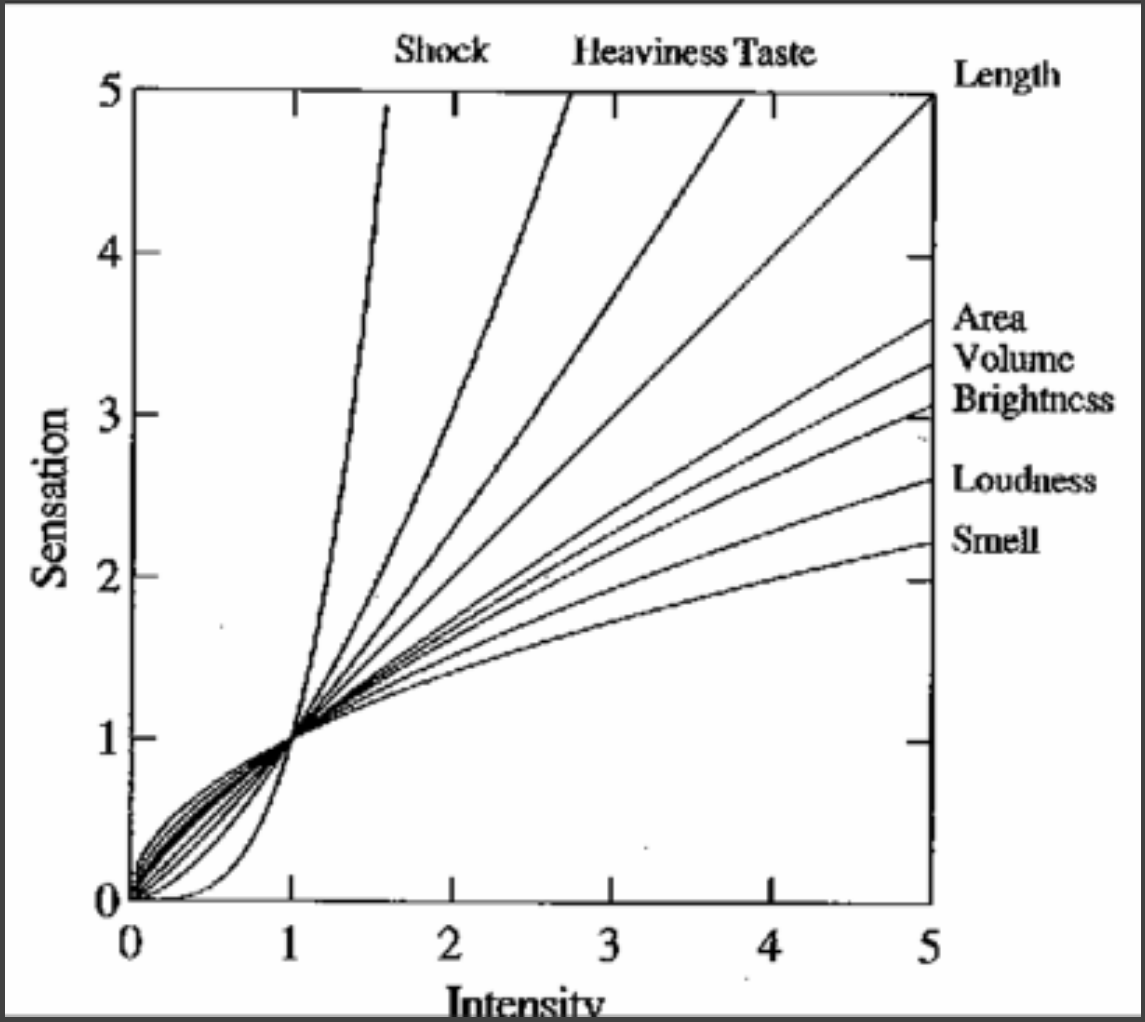
Metric Manipulation: Choosing how data are counted or normalized in order to hide or exaggerate effects in your data.

A3 Review: Ethical or Deceptive?

"You will be assigned at least one ethical and one deceptive visualization; the other two visualizations will be randomly assigned."



Graphical Perception

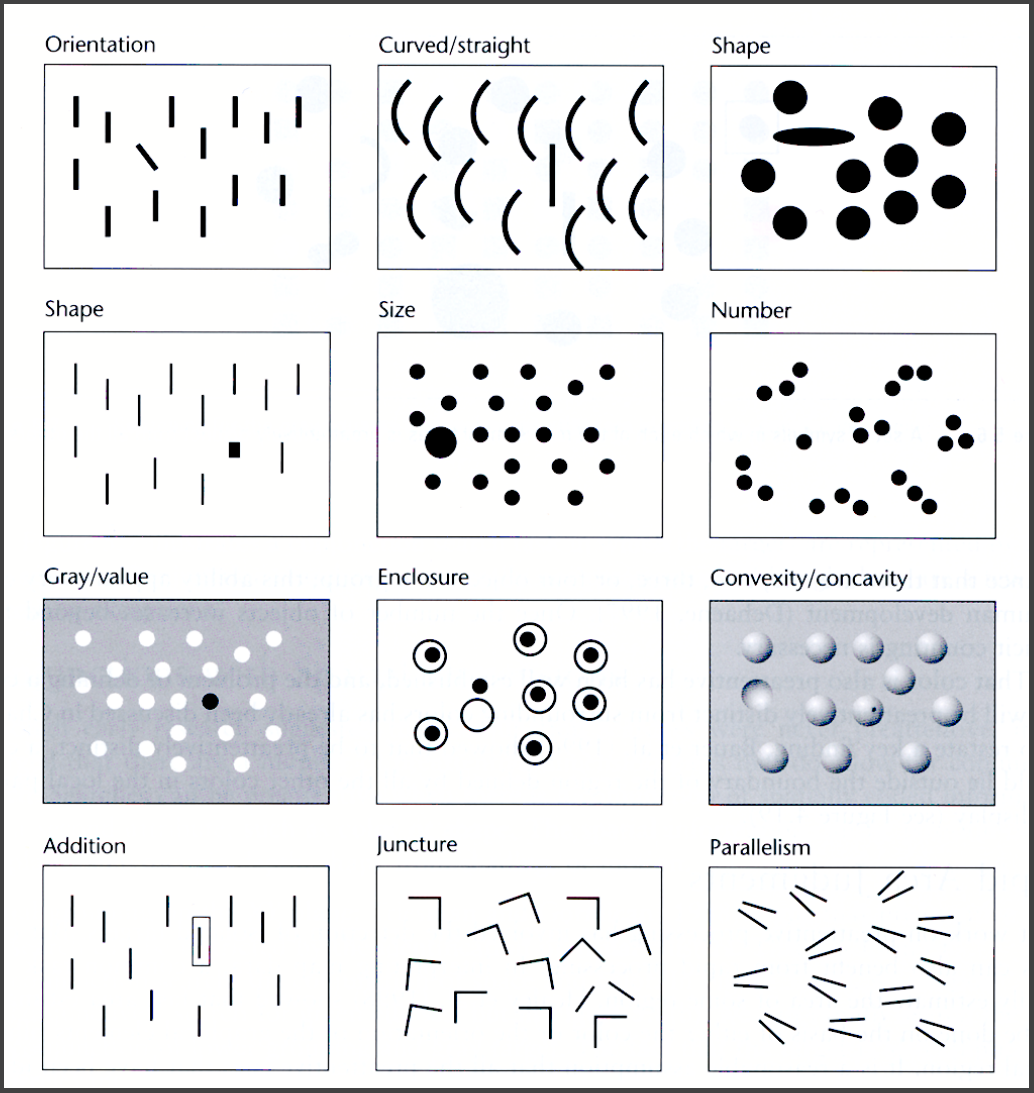


The psychophysics of sensory function [Stevens 61]

Graphical Perception

The ability of viewers to interpret visual (graphical) encodings of information and thereby decode information in graphs.

Pre-Attentive Features

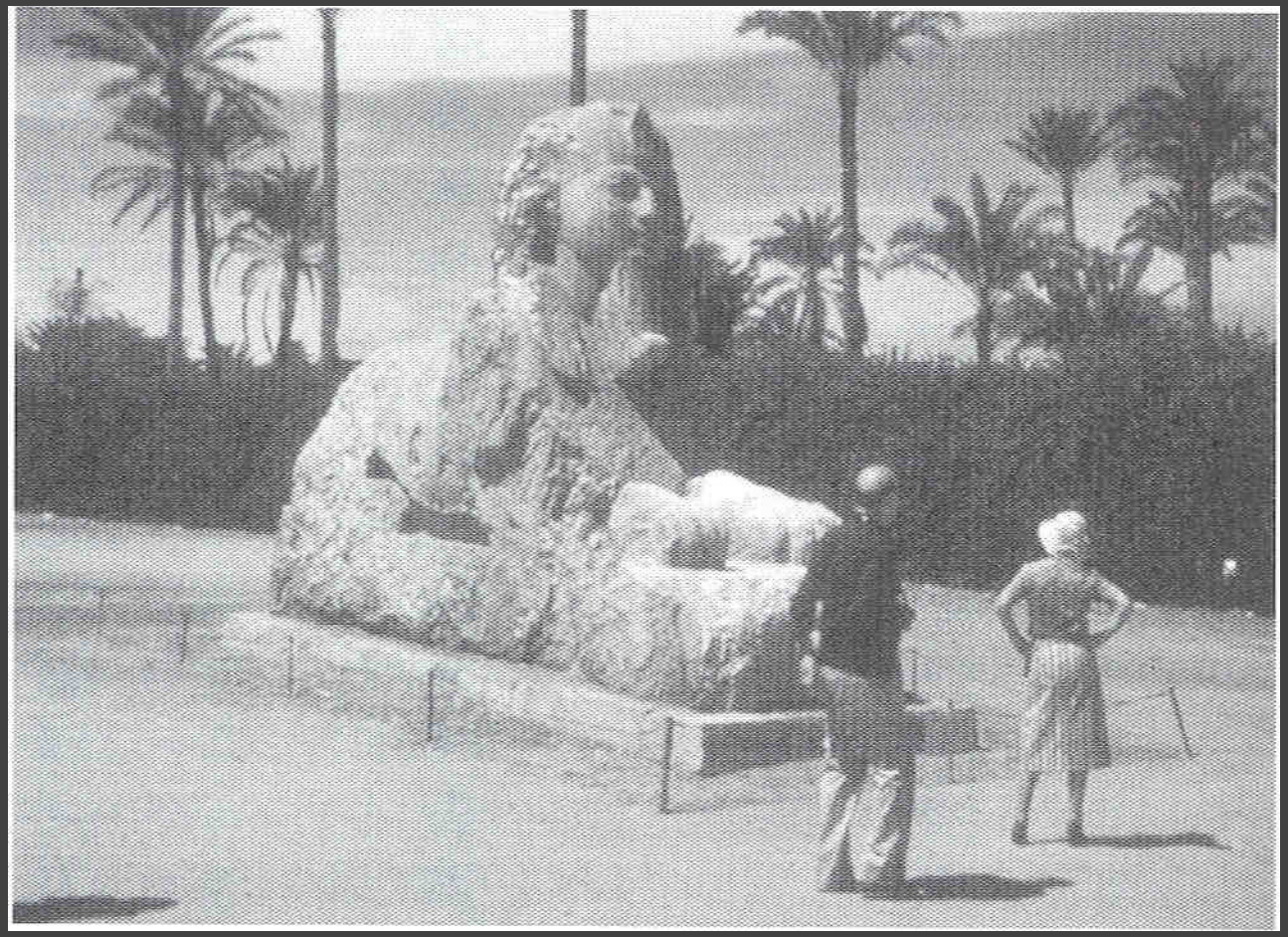


[Information Visualization. Figure 5. 5 Ware 04]

Change Blindness

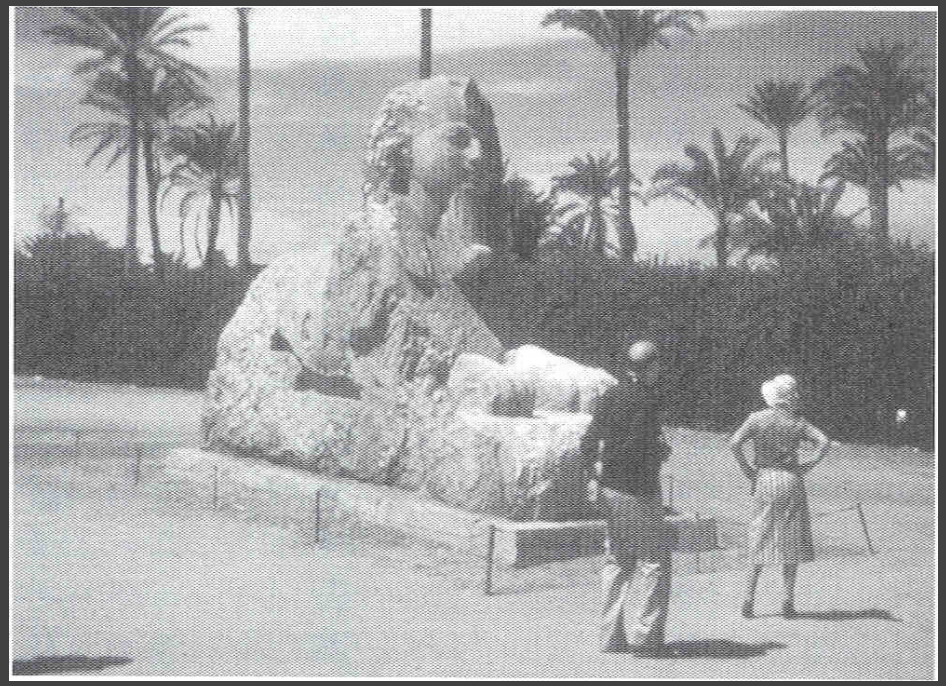
Change Blindness

Change Blindness



Change Blindness

Change Blindness



[Example from Palmer 99, originally due to Rock]

Just Noticeable Difference (JND)

JND (Weber's Law)

Scale Factor
(Empirically Determined)

Perceived Change → $\Delta S = k \frac{\Delta I}{I}$ ← Change of Intensity

← Physical Intensity

Ratios more important than magnitude

Most continuous variation in stimuli are perceived in discrete steps

Summary

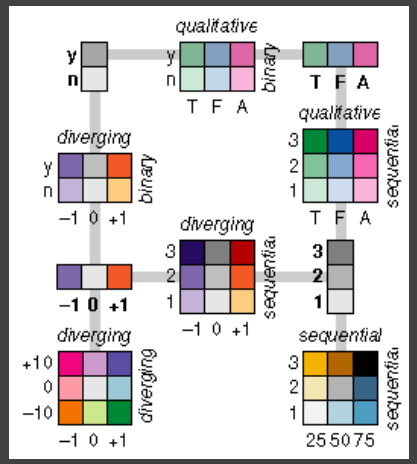
Choosing effective visual encodings requires knowledge of visual perception.

Visual features/attributes
Individual attributes often pre-attentive
Multiple attributes may be separable or integral

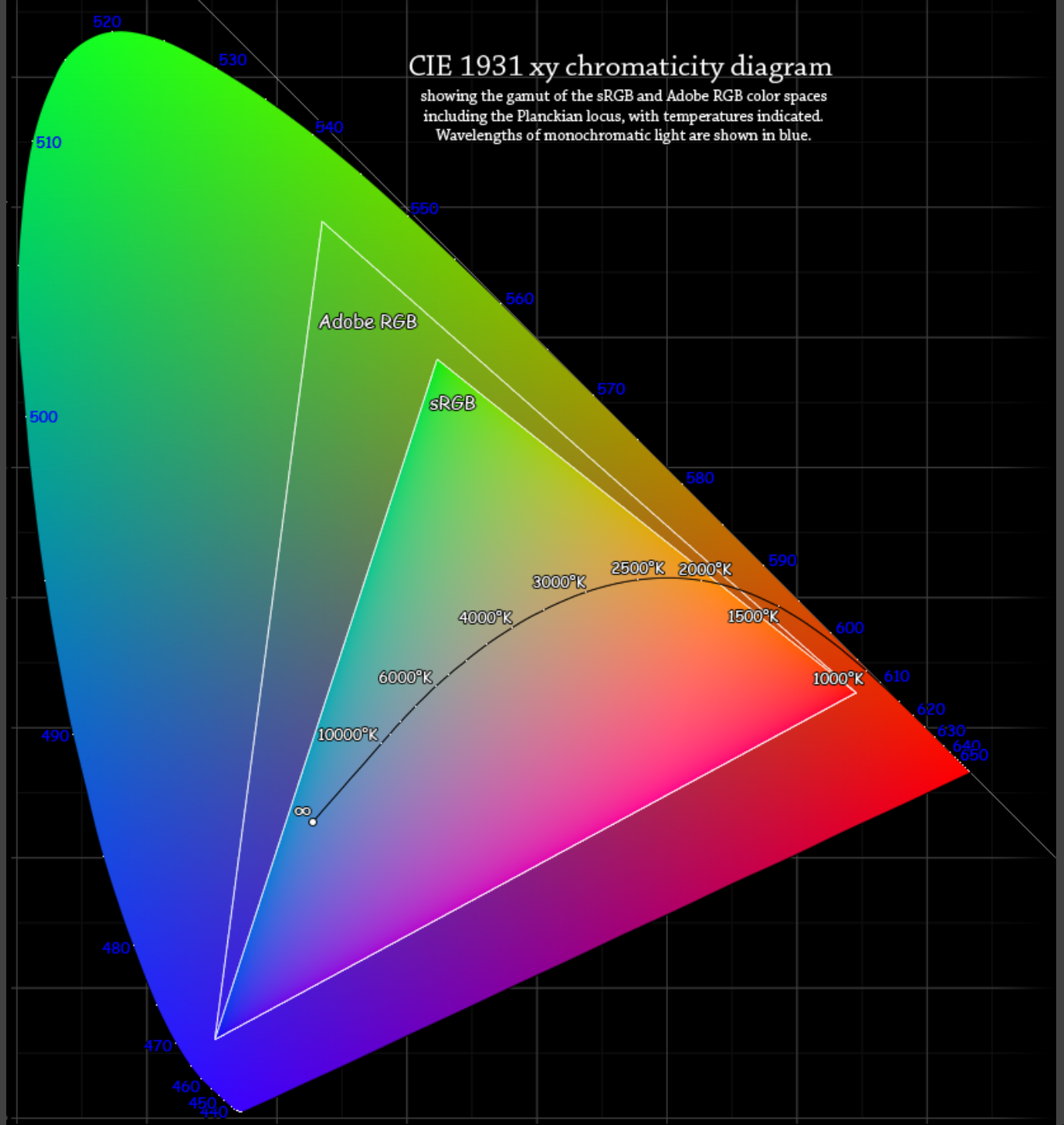
Gestalt principles provide high-level guidelines

We don’t always see everything that is there!

Color

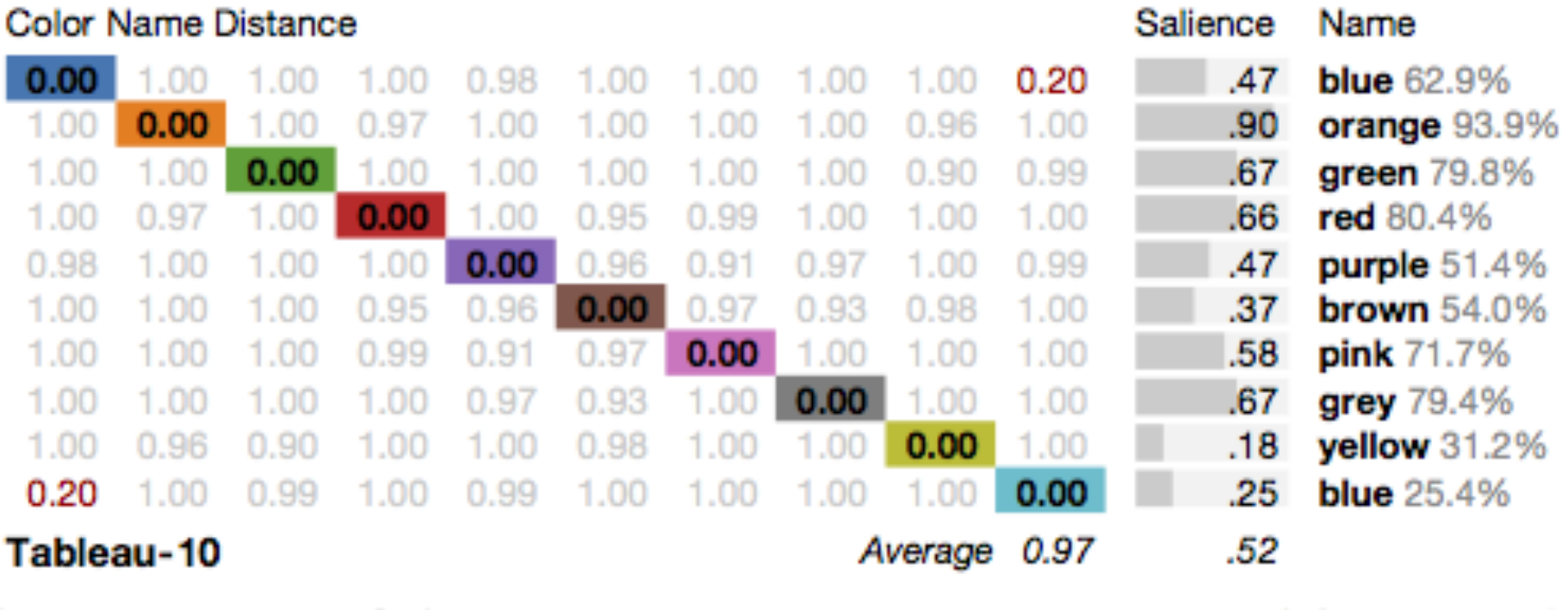


Color Brewer



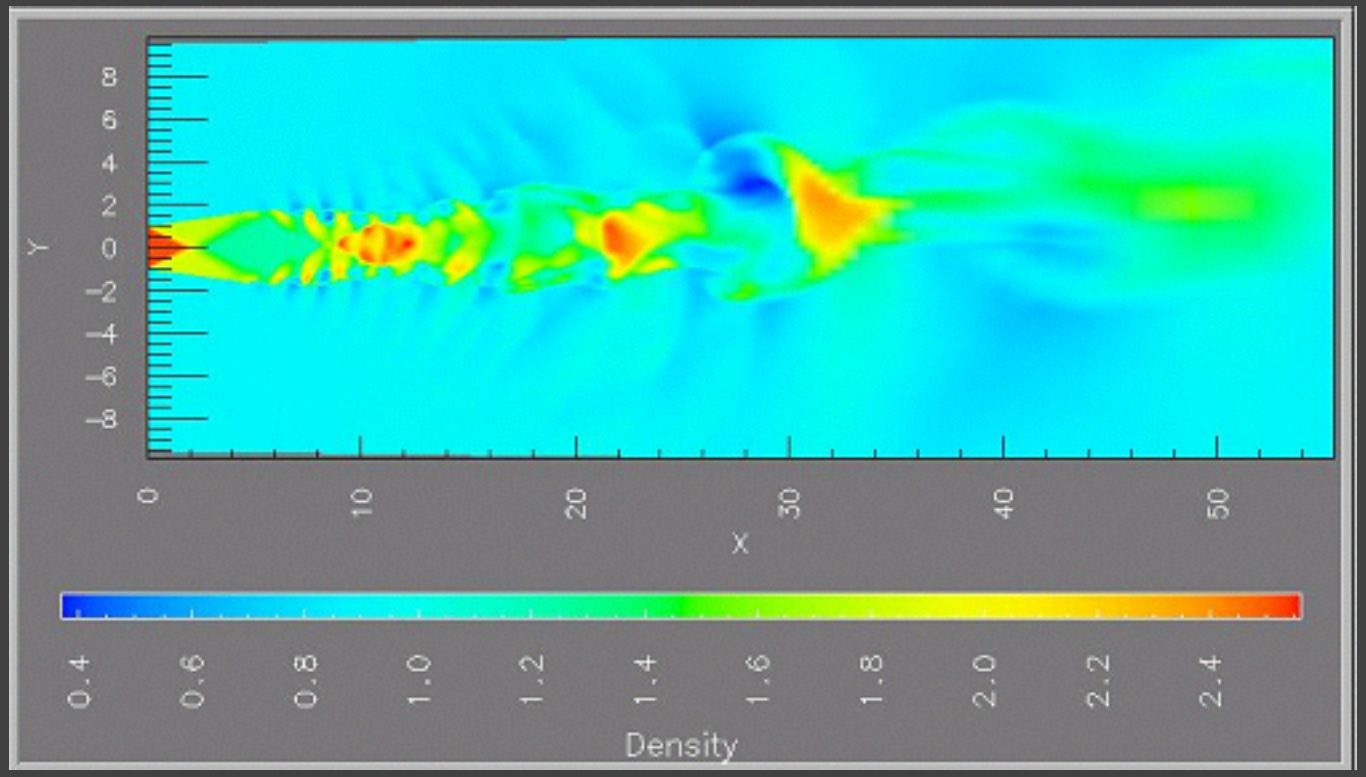
Palette Design & Color Names

Minimize overlap and ambiguity of colors.



http://vis.stanford.edu/color-names

Be Wary of Naïve Rainbows!



1. Hues are not naturally ordered
2. People segment colors into classes, perceptual banding
3. Naive rainbows are unfriendly to color blind viewers
4. Some colors are less effective at high spatial frequencies

Quantitative Color Encoding

Sequential color scale

Ramp in luminance, possibly also hue

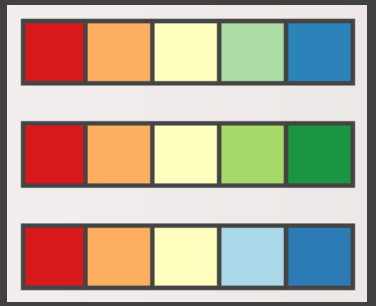
Higher value -> darker color (or vice versa)

Diverging color scale

Useful when data has meaningful “midpoint”

Use neutral color (e.g., grey) for midpoint

Use saturated colors for endpoints

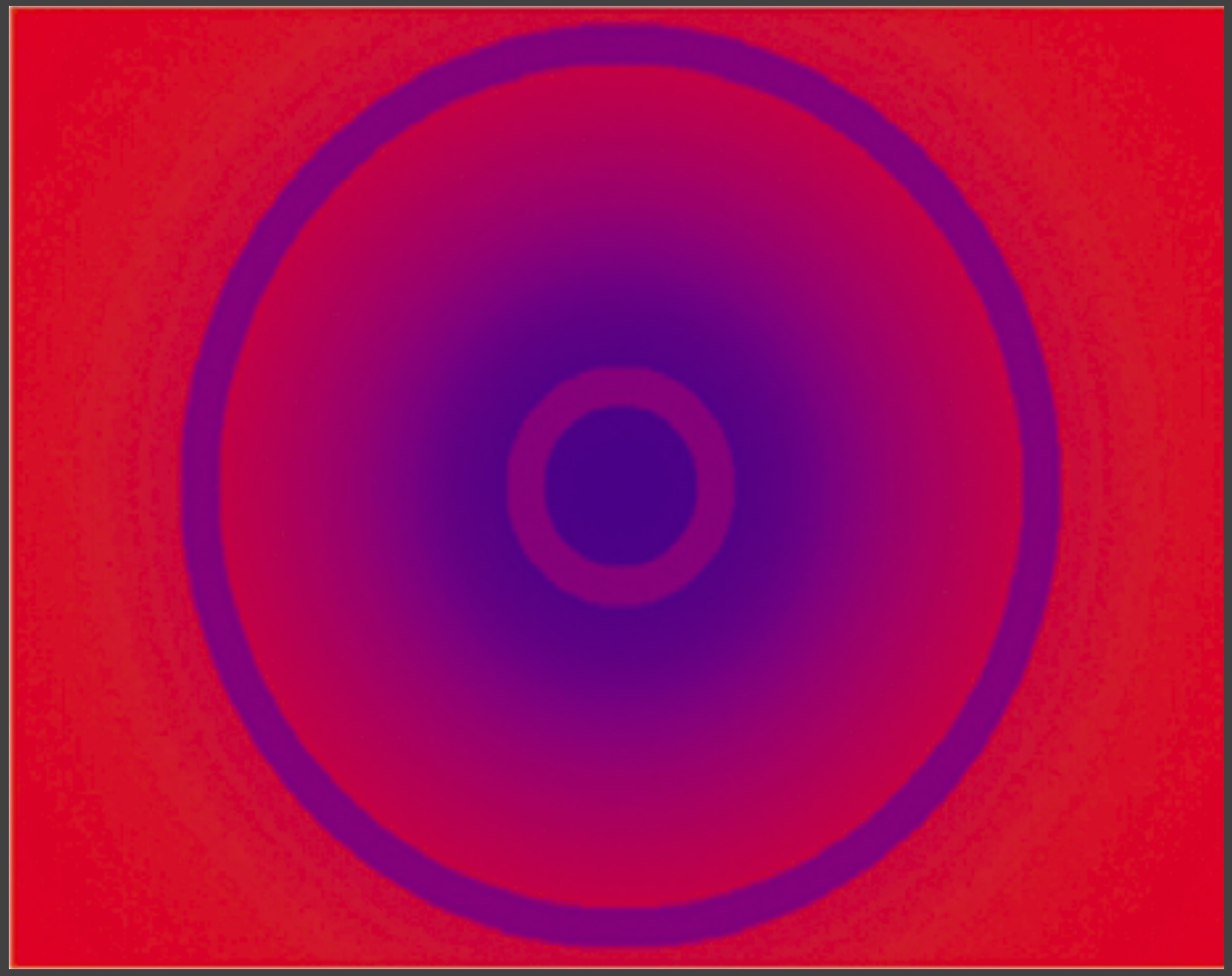


Limit number of steps in color to 3-9

Avoid simultaneous contrast, hold mappings in memory

Simultaneous Contrast

Inner & outer rings are the same physical purple.



Donald
MacLeod

Hints for the Colorist

Use **only a few** colors (~6 ideal)

Colors should be **distinctive** and **named**

Strive for color **harmony** (natural colors?)

Use **cultural conventions**; appreciate symbolism

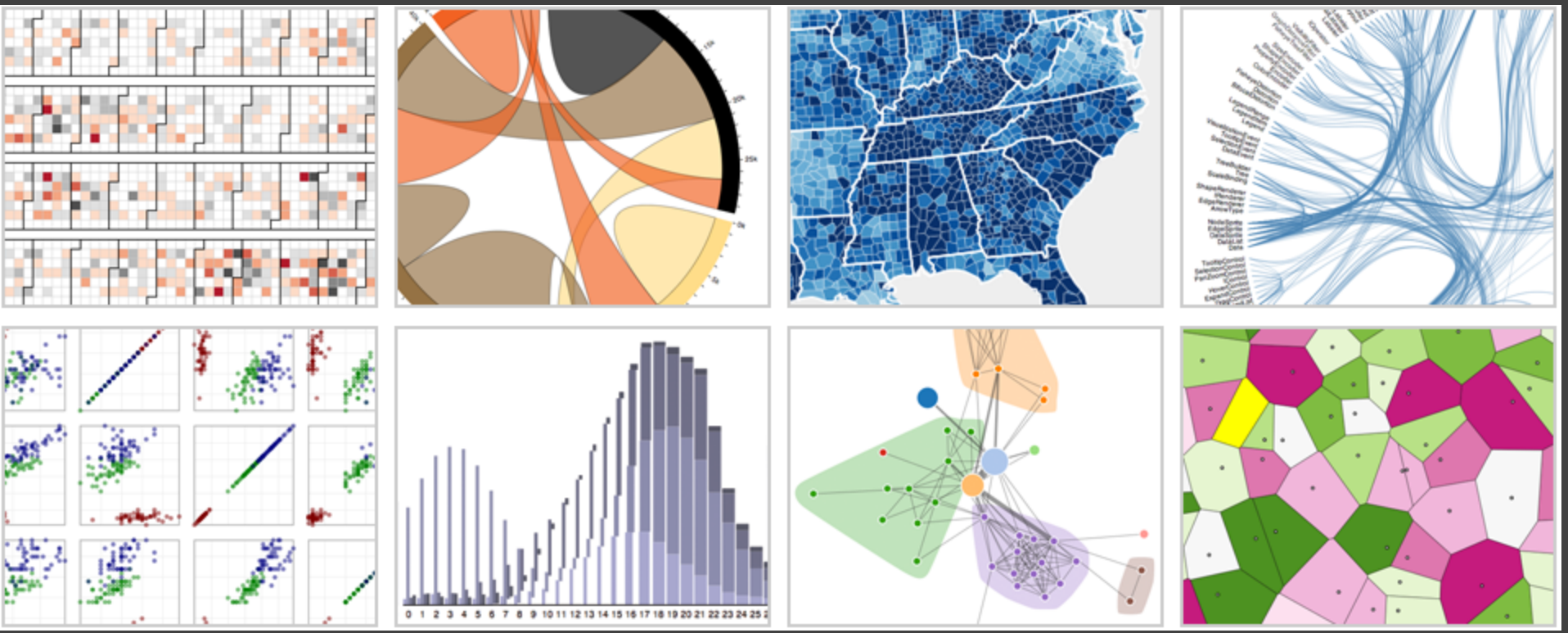
Get it right in **black and white**

Respect the **color blind**

Take advantage of **perceptual color spaces**

Color is cultural and a matter of taste!

Visualization Software



D3: Data-Driven Documents

Chart Typologies

Excel, Many Eyes, Google Charts

Visual Analysis Grammars

VizQL, ggplot2, Vega-Lite

Visualization Grammars

Protovis, D3.js, Vega

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Chart Typologies

Excel, Many Eyes, Google Charts

Charting Tools

Visual Analysis Grammars

VizQL, ggplot2, Vega-Lite

Visualization Grammars

Protovis, D3.js, Vega

Declarative Languages

Component Architectures

Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

Programming Toolkits

Interactive Data Exploration

Tableau, Lyra, Polestar, Voyager

Graphical Interfaces

Visual Analysis Grammars

VizQL, ggplot2, Vega-Lite

Visualization Grammars

Protovis, D3.js, Vega

Declarative Languages

Component Architectures

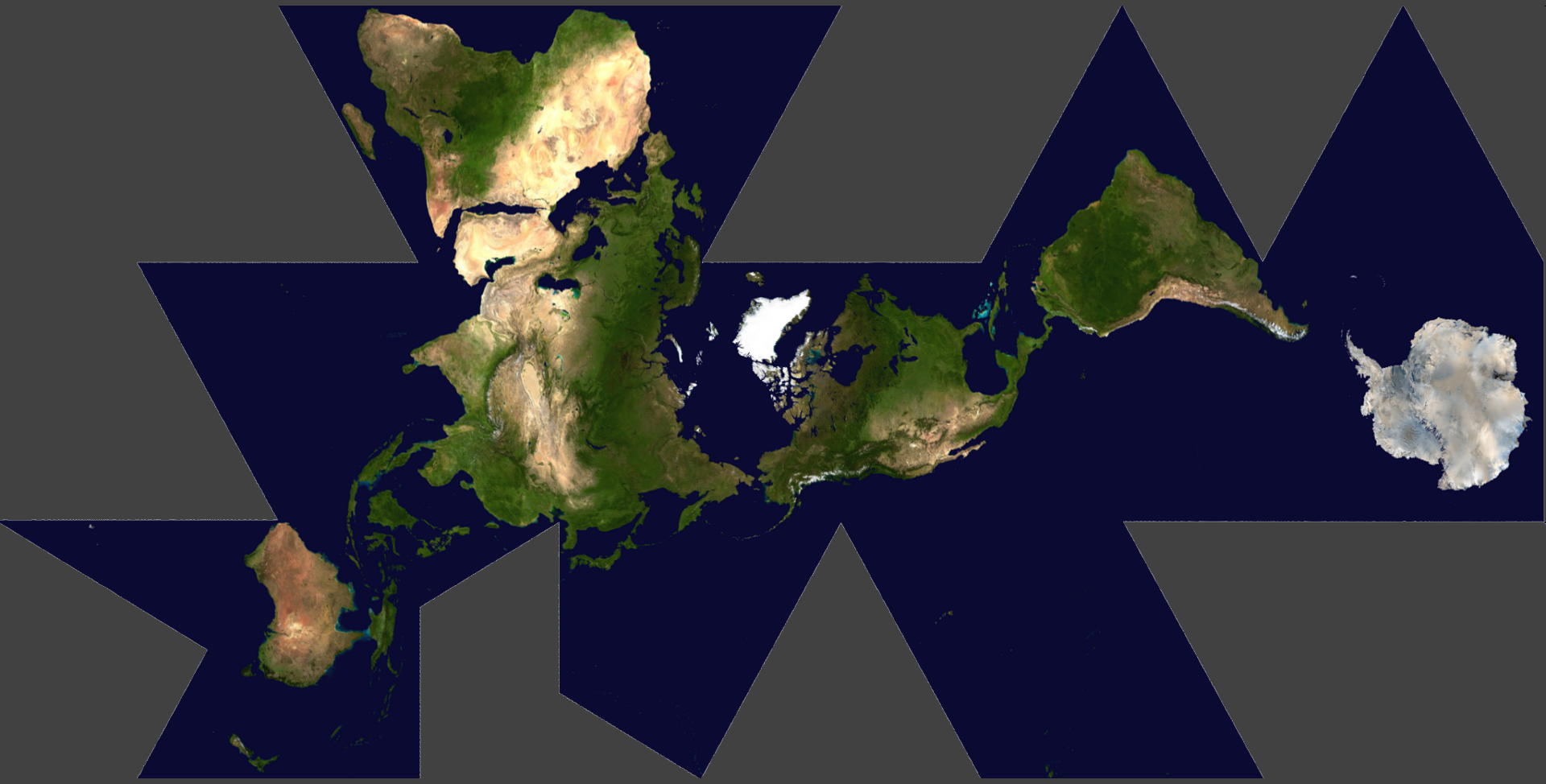
Prefuse, Flare, Improvise, VTK

Graphics APIs

Processing, OpenGL, Java2D

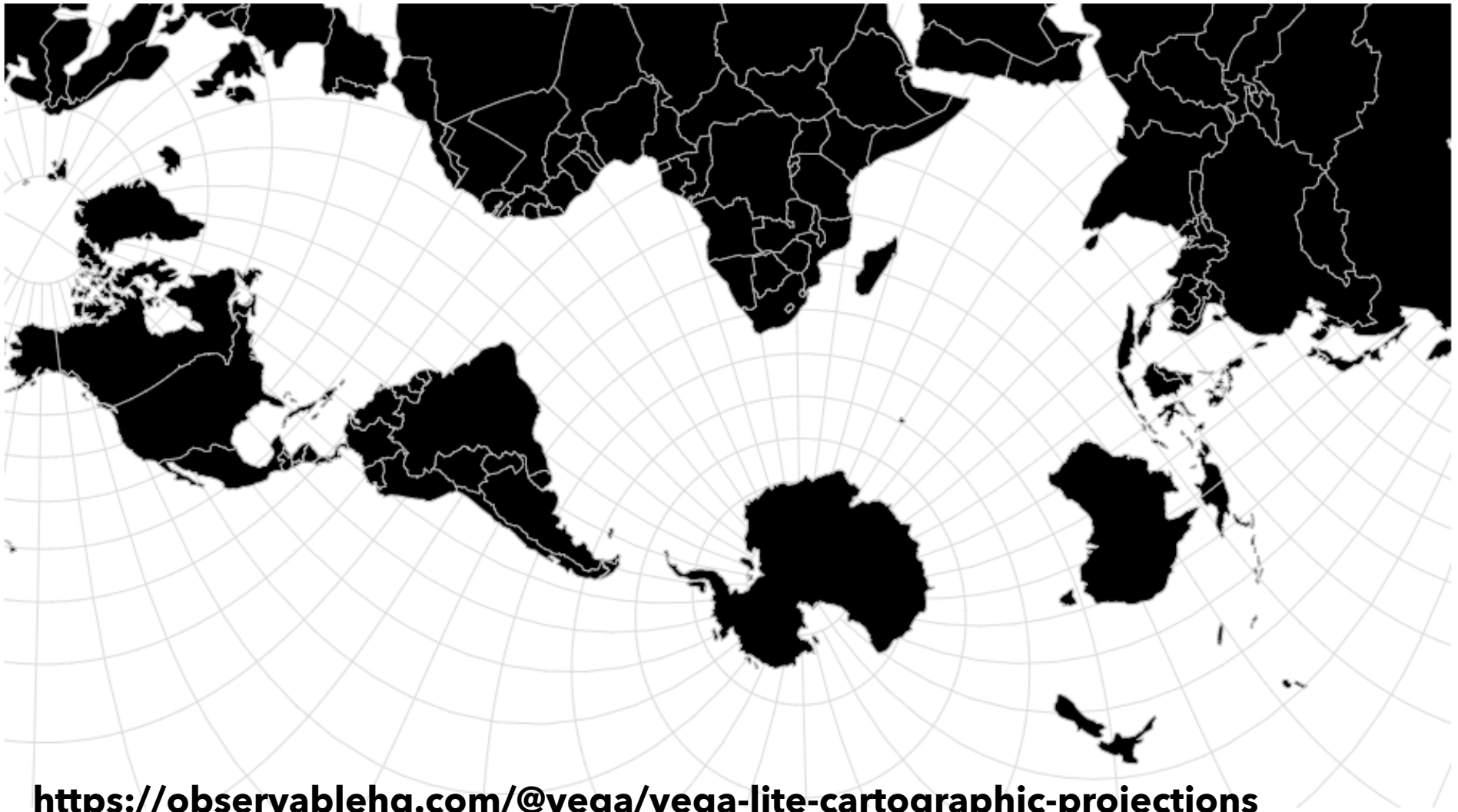
Programming Toolkits

Maps



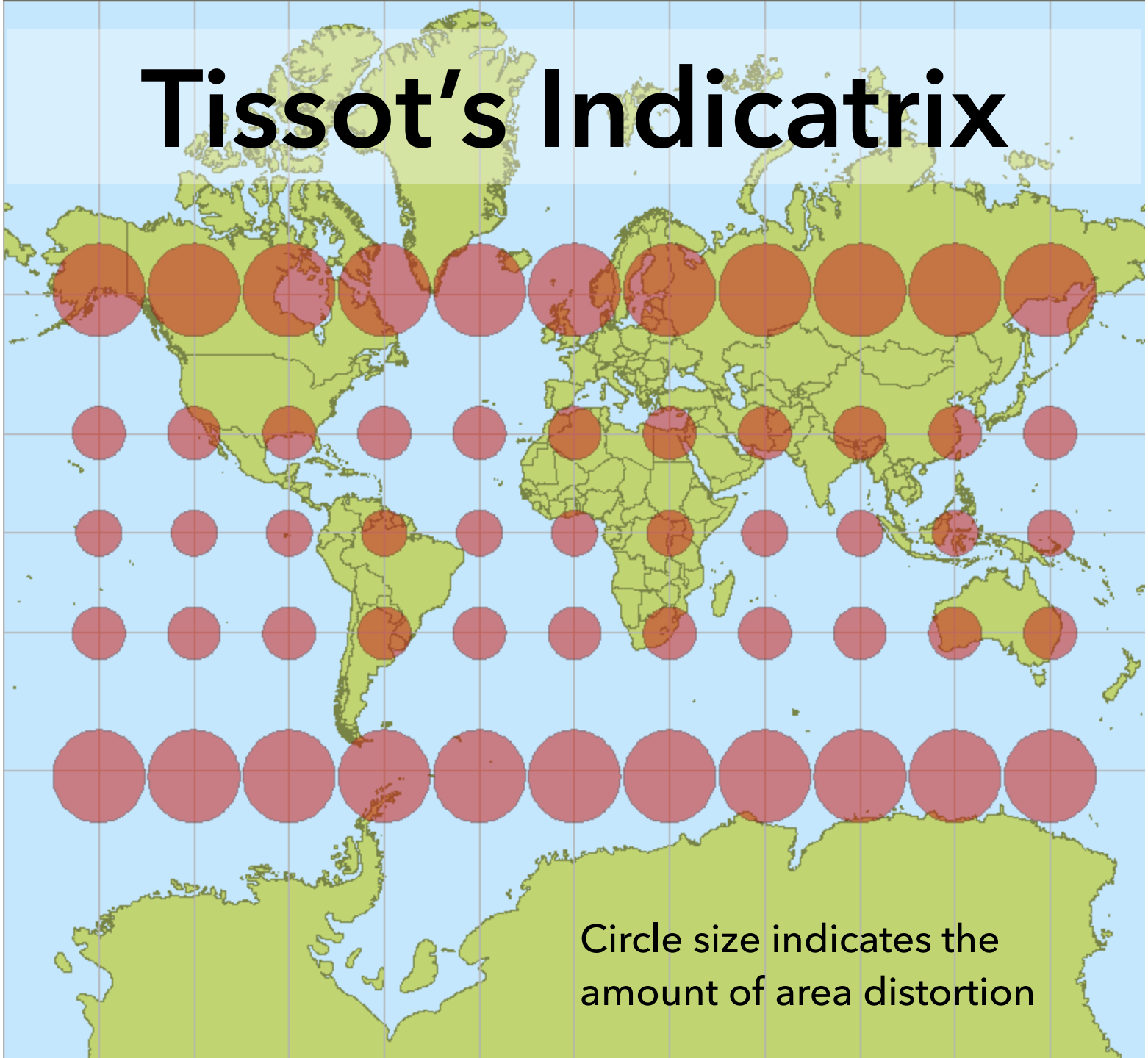
Dymaxion Maps [Fuller 46]

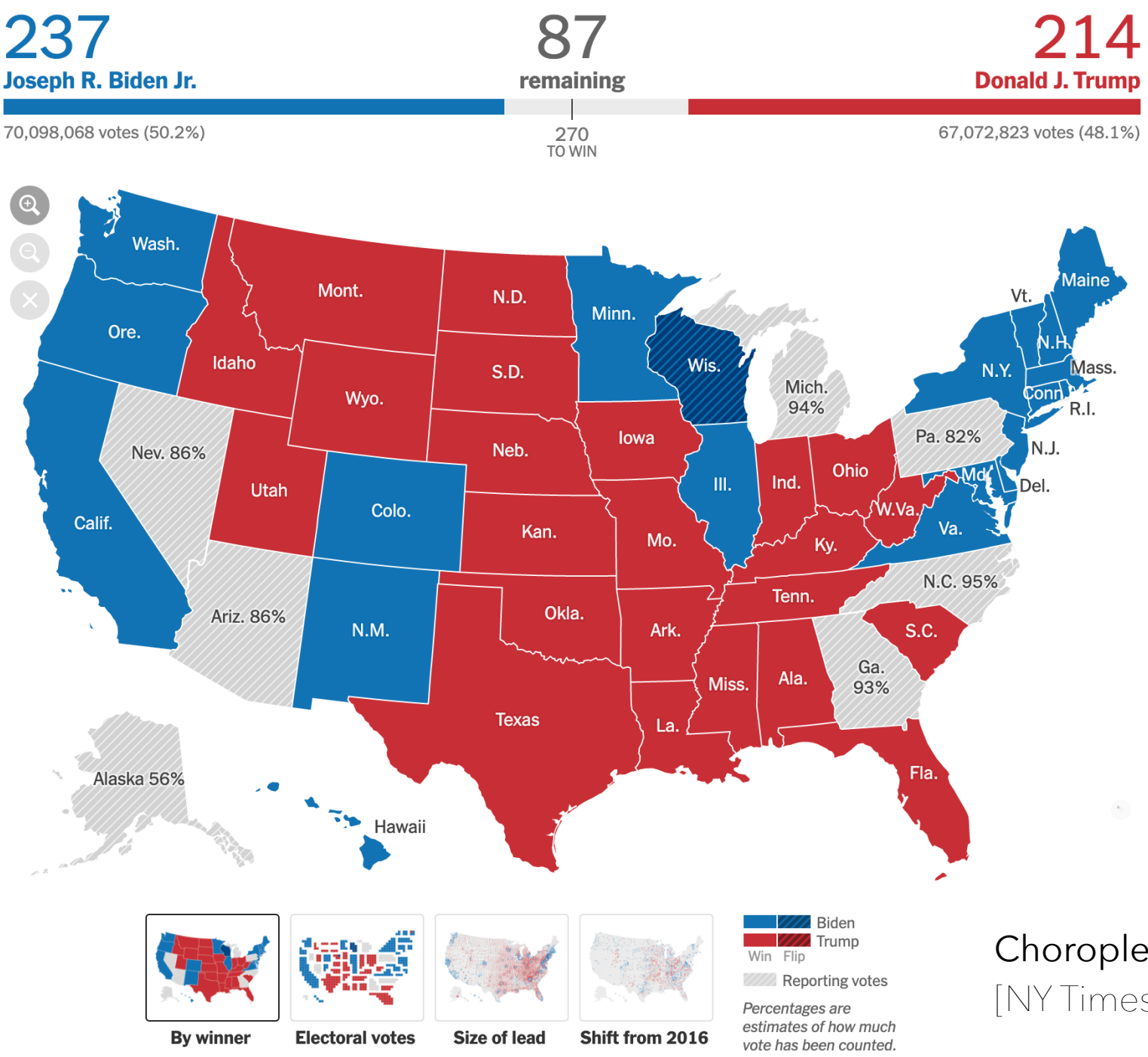
Exploring Projections…



https://observablehq.com/@vega/vega-lite-cartographic-projections

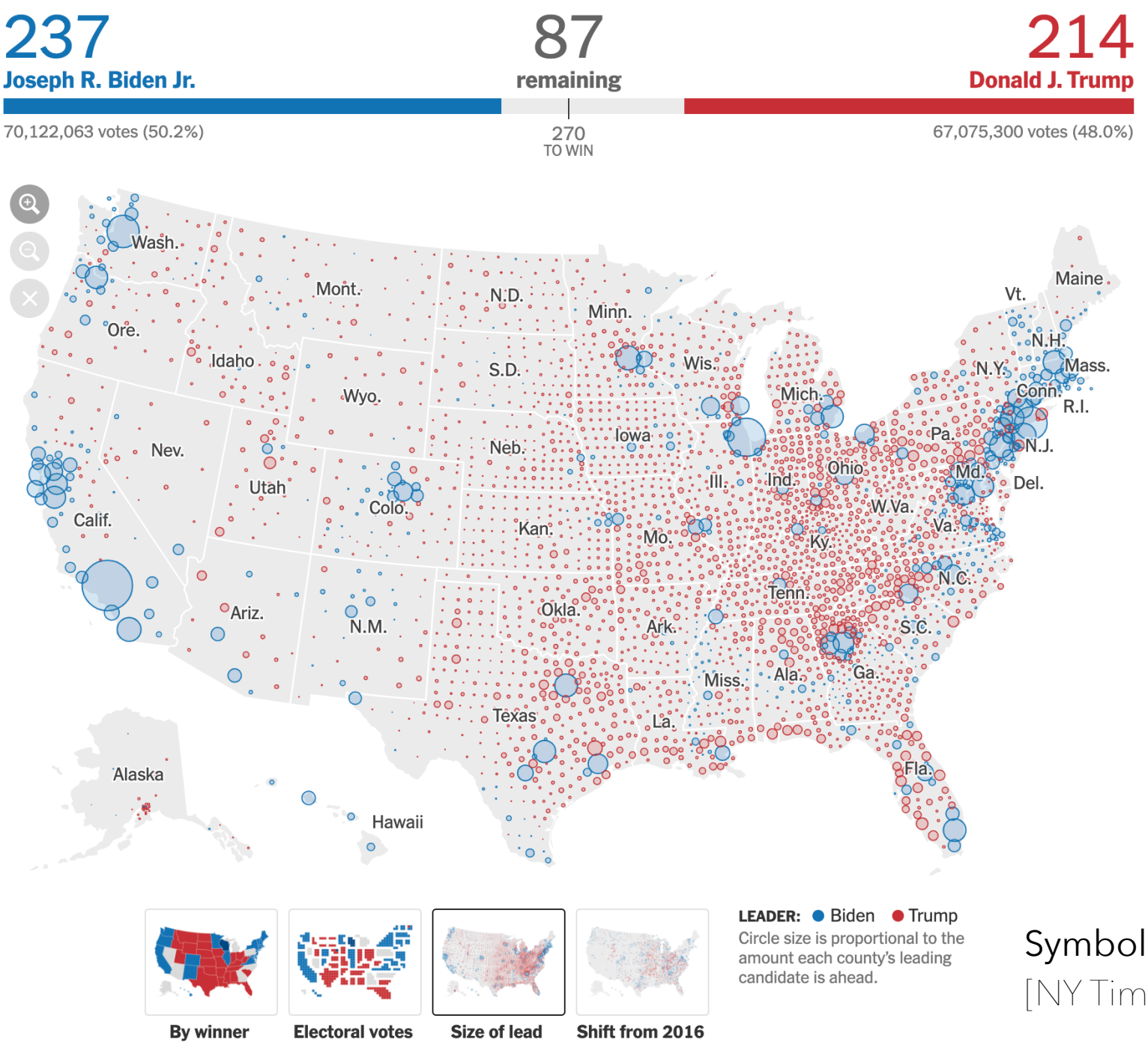
Tissot's Indicatrix



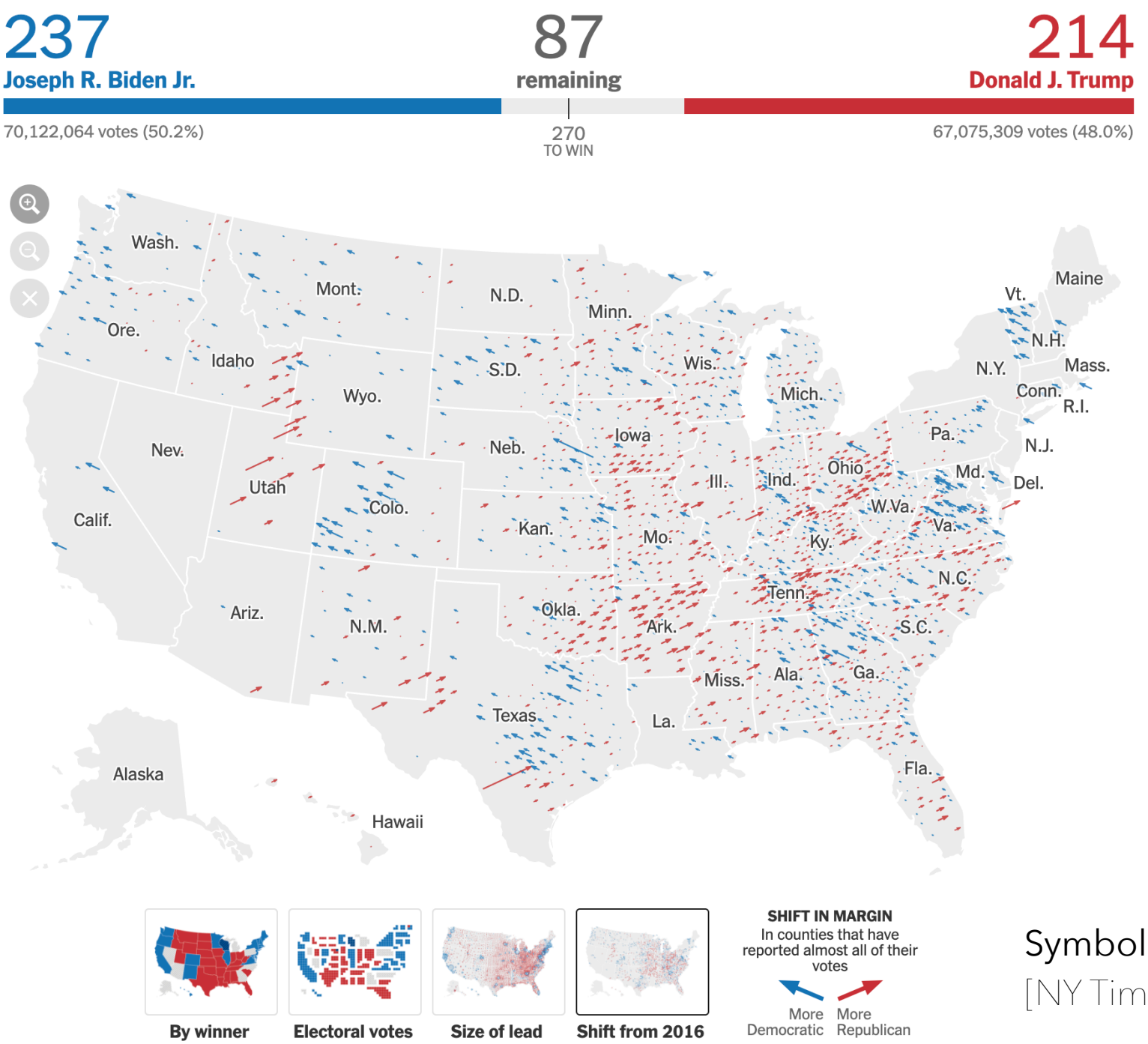


Choropleth Map

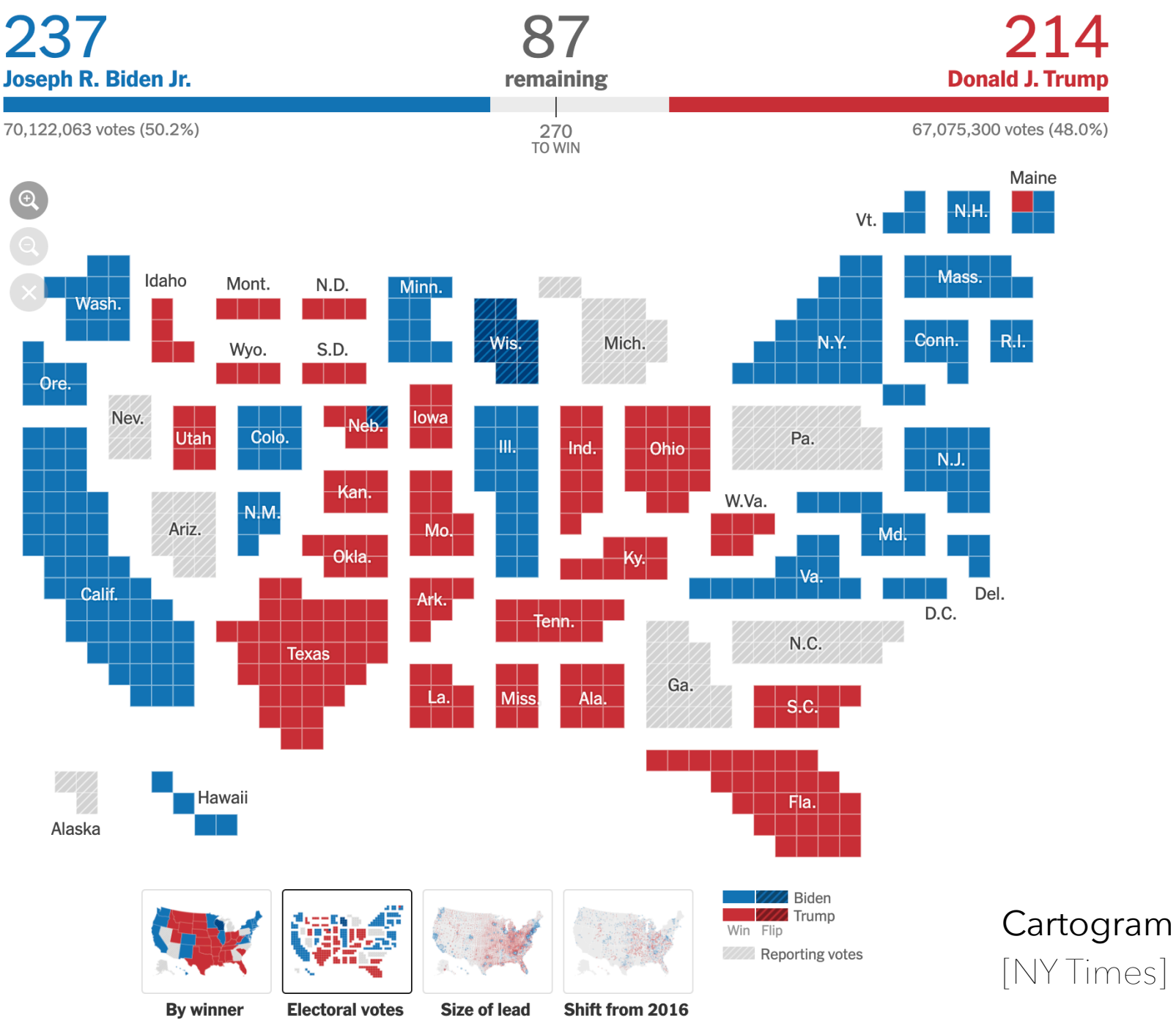
[NY Times]



Symbol Map
[NY Times]

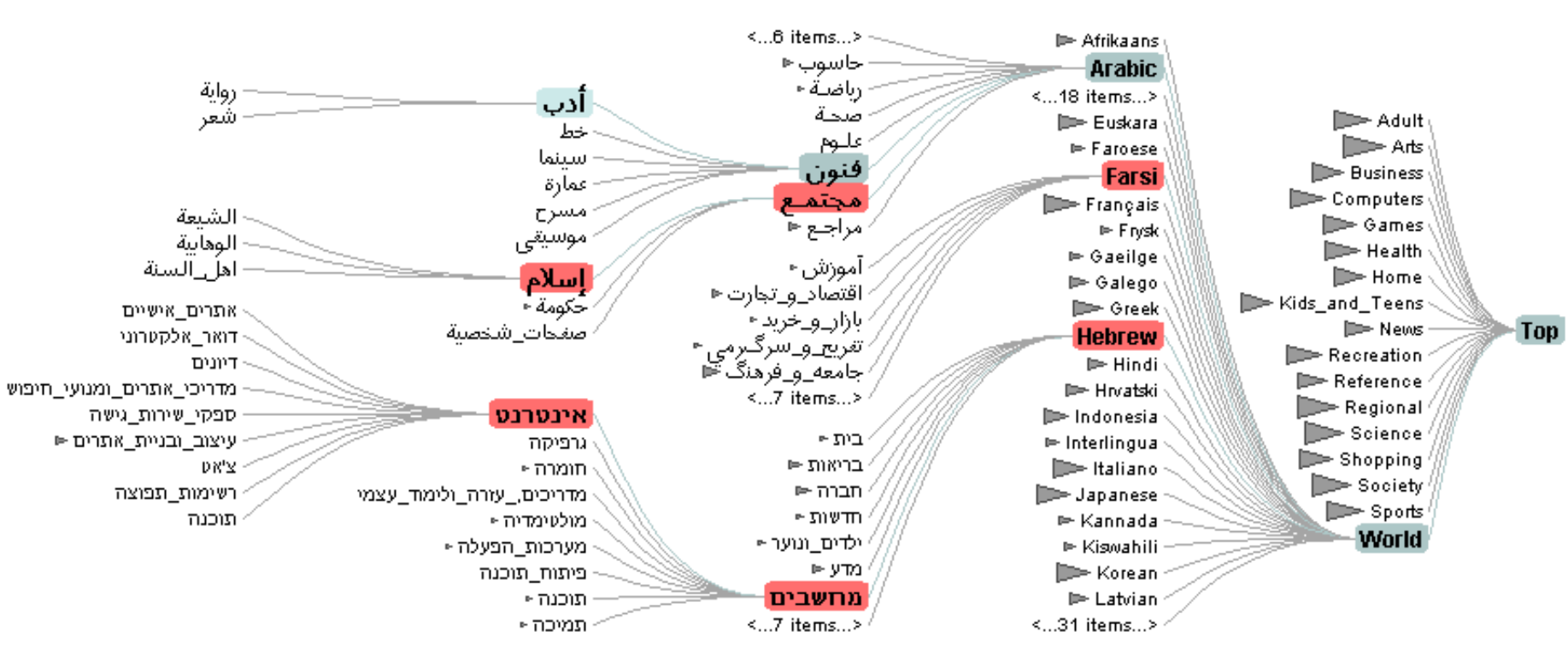


Symbol Map
[NY Times]



Cartogram
[NY Times]

Hierarchies and Networks



Degree-Of-Interest Trees [Heer & Card 04]

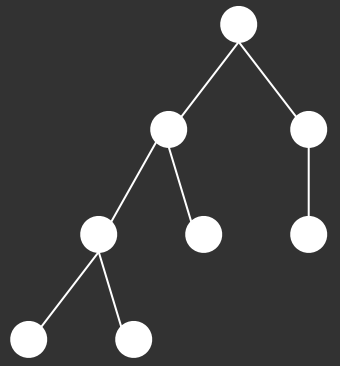
Trees and Graphs

Trees

Graphs with hierarchical structure

Connected graph with N-1 edges

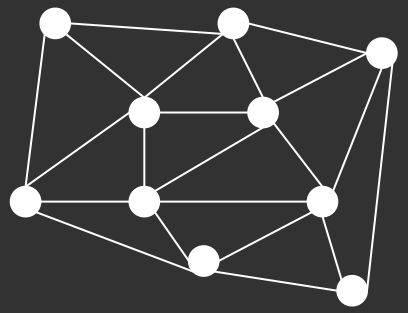
Nodes as *parents* and *children*



Graphs

Model relations among data

Nodes and *edges*



Network Analysis Tasks [Pretorius '13]

Structure-based: relationships and connectivity

Find all of the friends of friends for Taylor.

Find all of the people who are friends with Jordan and Alex.

Six degrees of separation: shortest path between two individuals.

Attribute-based: specific node/link attributes

Find all "students" attending CSE412.

Find all the "friends" and "family" of Alex.

Browsing: understand paths in the data

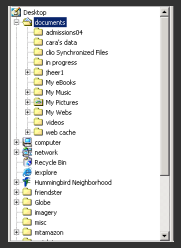
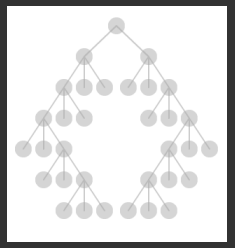
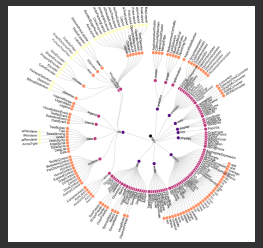
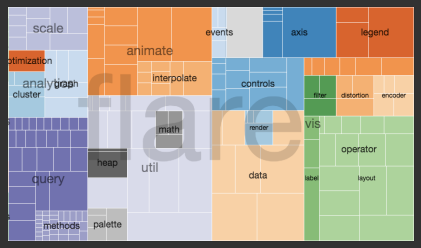
Find Alex's friend Taylor, and then Taylor's friend Jordan.

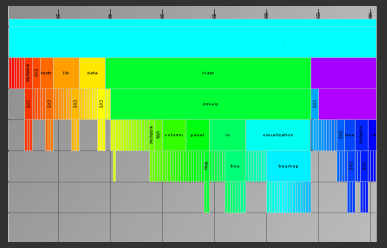
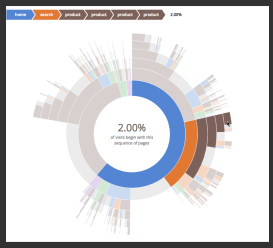
Estimation: summarization and temporal changes

How does Jordan's friend group change over the course of the year?

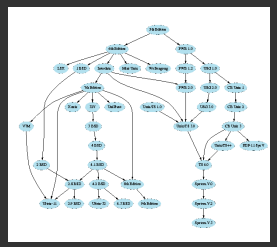
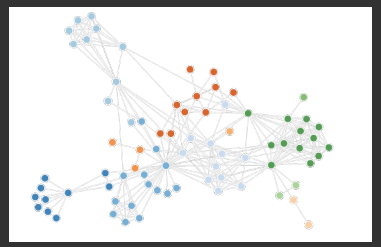
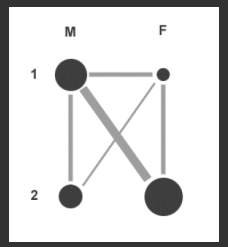
Hierarchies and Networks

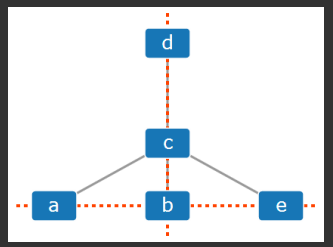
Mon 2/22 - Tree Visualization

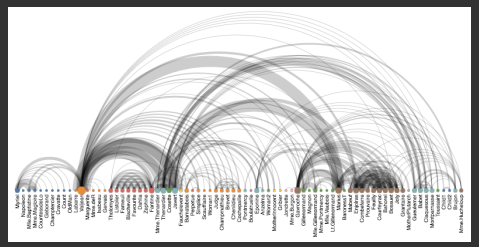





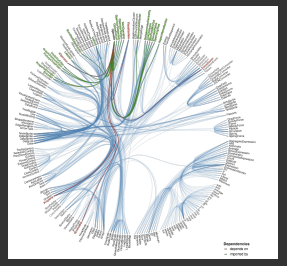
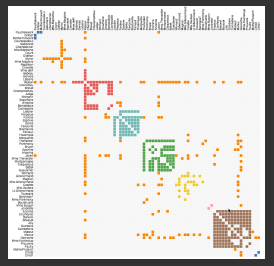
Wed 2/24 - Graph Layout: Node-Link Diagrams

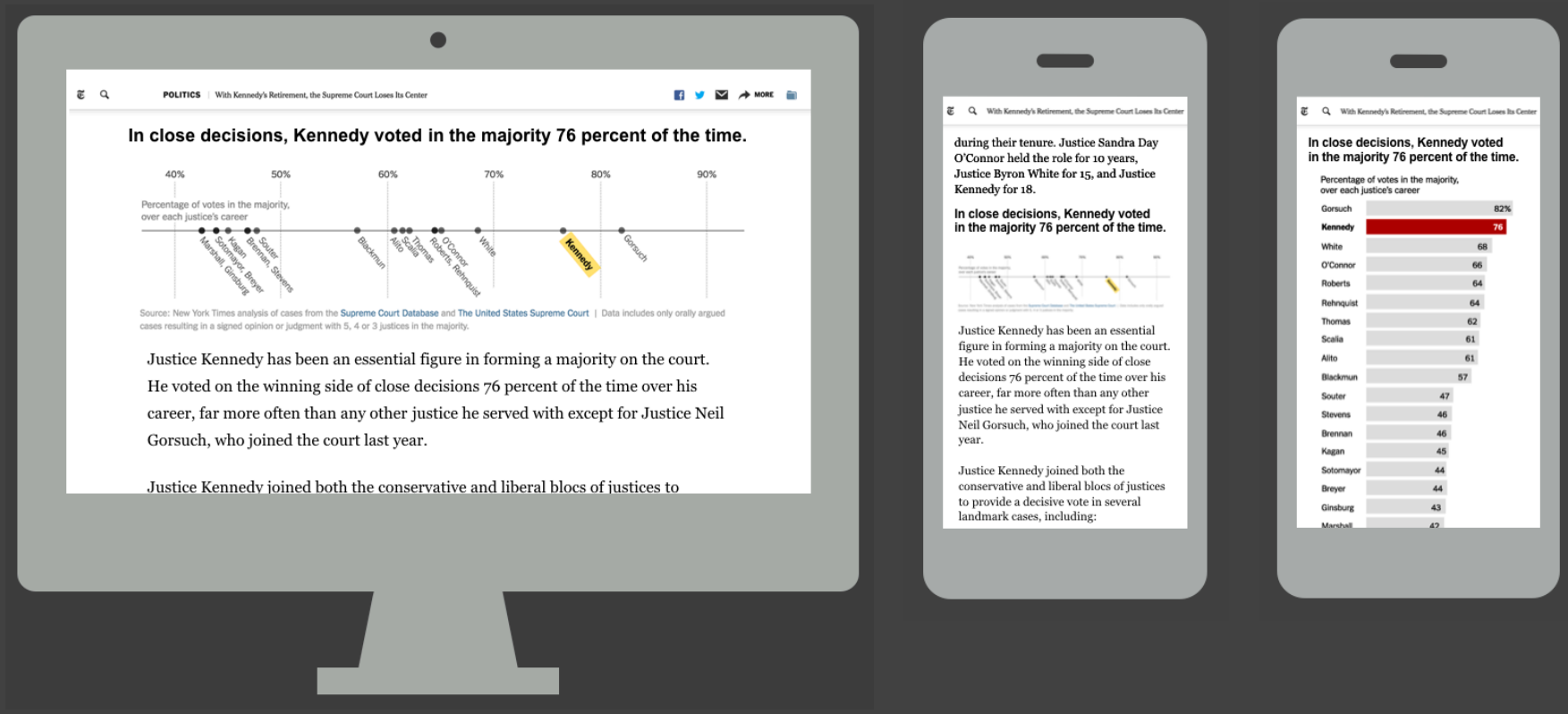




Wed 2/24 - Alternative Visualizations & Techniques

Responsive Visualization



Techniques for Responsive Visualization Design [Hoffswell 20]

Responsive Visualization Summary

Good visualizations are task dependent
Who is the audience and what is the task?
Pick the right interaction technique

Visualizations are not one size fits all
Context might change user goals

Text

explore
visualizations
data sets
comments
topic hubs

participate
create visualization
upload data set
create topic hub
register

learn more
quick start
visualization types
data format & style
about Many Eyes
FAQ
blog

contact Us
contact
report a bug

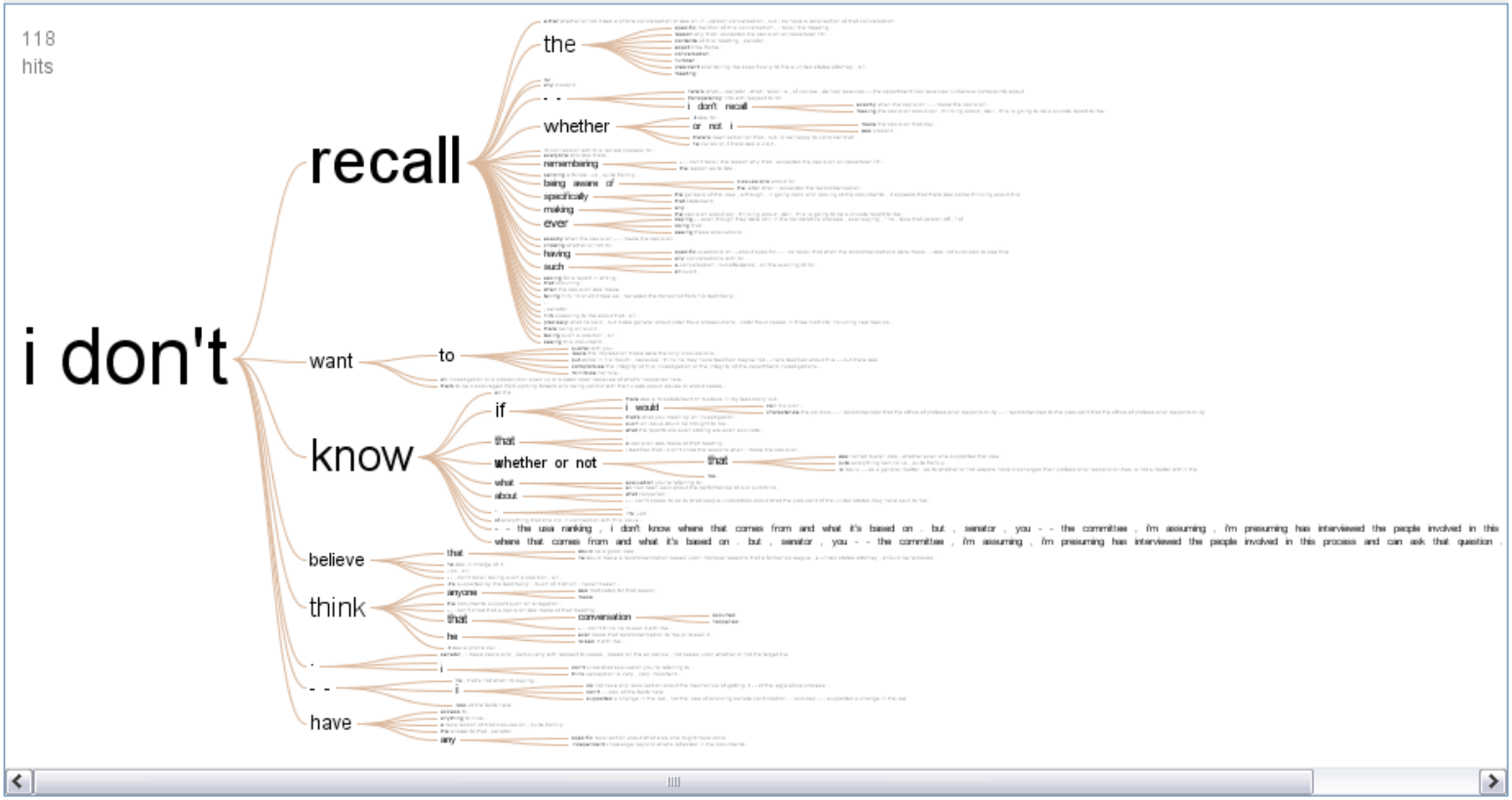
legal
terms of use

Popular Dataset Tags
2007 2008 bible blog books census crime education eharmony election energy food health inauguration internet ireland literature lyrics media music network obama people politics population president prices religion

Visualizations : Word tree / Alberto Gonzales

Creator: Martin Wattenberg
Tags:

Search i don't | Back | Forward | Start | End | Occurrence Order | Clicks Will Zoom



Data file: Word in testimony from Gonzales, 4/19/2007 Data source: CQ Transcript Wire via the Washington Post This data set has not yet been rated

full image | share this | watch this | add to topic hub | rate this

Comments (4)

currently showing

This visualization has 4 positive and 0 negative

Text Processing Pipeline

Tokenization

Segment text into terms.

Remove stop words? *a, an, the, of, to be*

Numbers and symbols? *#huskies, @UW, OMG!!!!!!*

Entities? *Washington State, Seattle, U.S.A*

Stemming

Group together different forms of a word.

Porter stemmer? *visualization(s), visualize(s), visually* → visual

Lemmatization? *goes, went, gone* → go

Ordered list of terms

Text Visualization Takeaways

High Dimensionality

Where possible use text to represent text…
… which terms are the most descriptive?

Context & Semantics

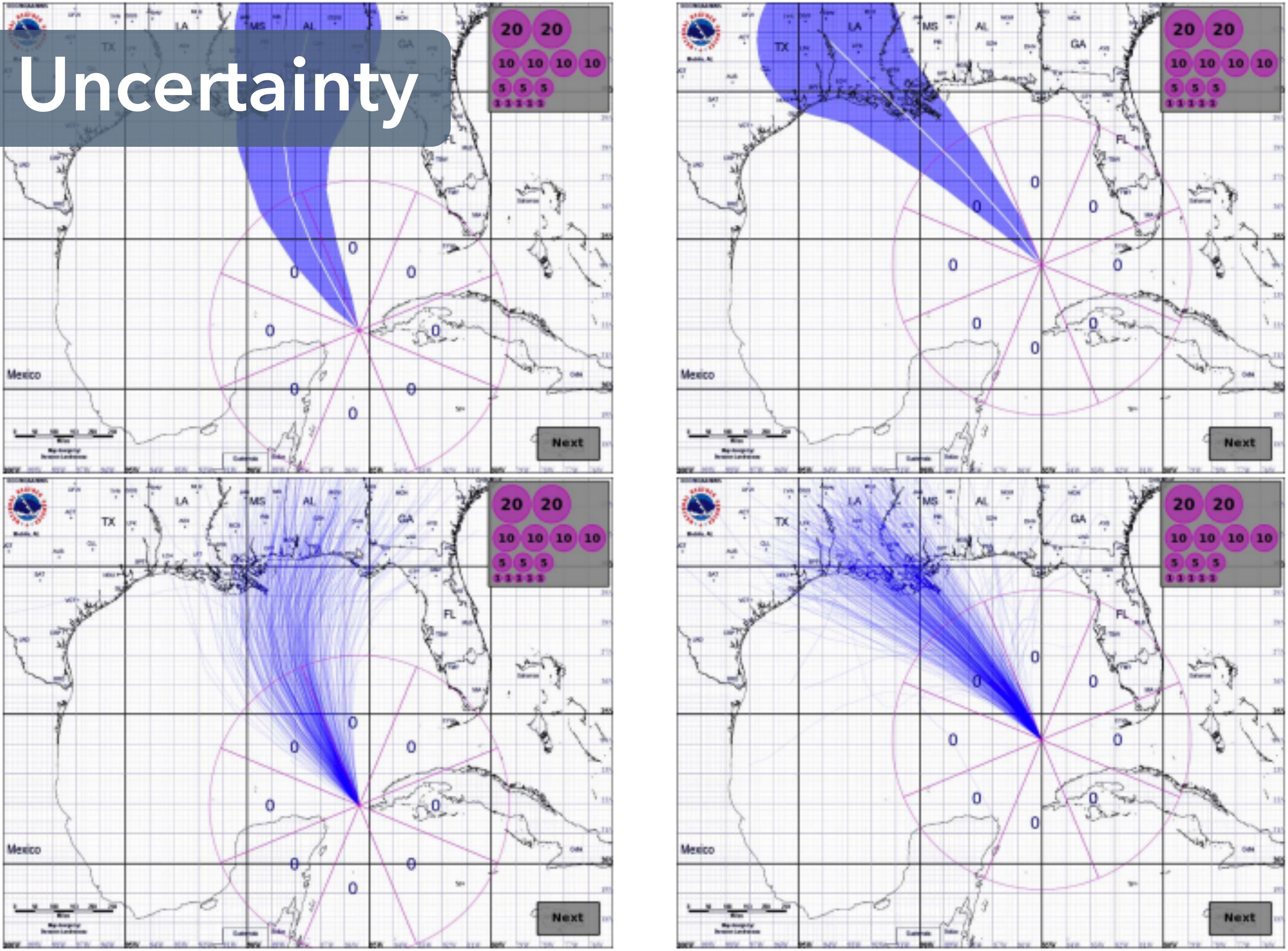
Provide relevant context to aid understanding.
Show (or provide access to) the source text.

Modeling Abstraction

Understand abstraction of your language models.
Match analysis task with appropriate tools and models.

Currently: from bag-of-words to *vector space embeddings*

Uncertainty



What does uncertainty mean?

Measurement Uncertainty:

"We're not exactly sure what the values in the data are."

Forecast Uncertainty:

"We're not exactly sure what will happen to the data next."

Model Uncertainty:

"We're not exactly sure how the data fits together."

Decision Uncertainty:

"We're not exactly sure what to do with the data."

Uncertainty Visualization Summary

Uncertainty can happen at all stages of the analysis process, from data collection to final decision-making

Variables like blur and transparency can be intuitive for showing uncertainty, but hard to decode.

Consider using discrete samples to show variation and uncertainty in a model

Consider when uncertainty is high enough that doing nothing is the right thing to do.

Scalability

1.7 B stars. 1.2 TB of data.
Visualizations running in-browser.

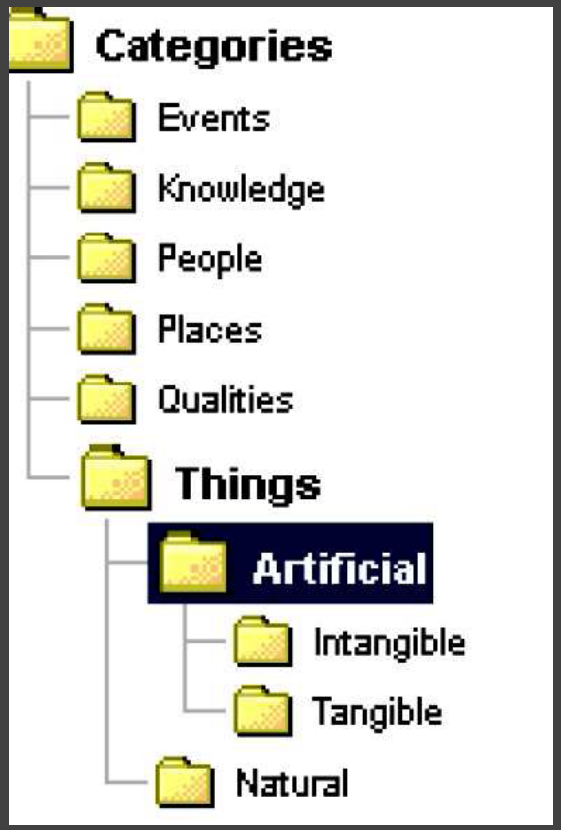
Powered by Falcon 0.13.0

Interactive Scalability Strategies

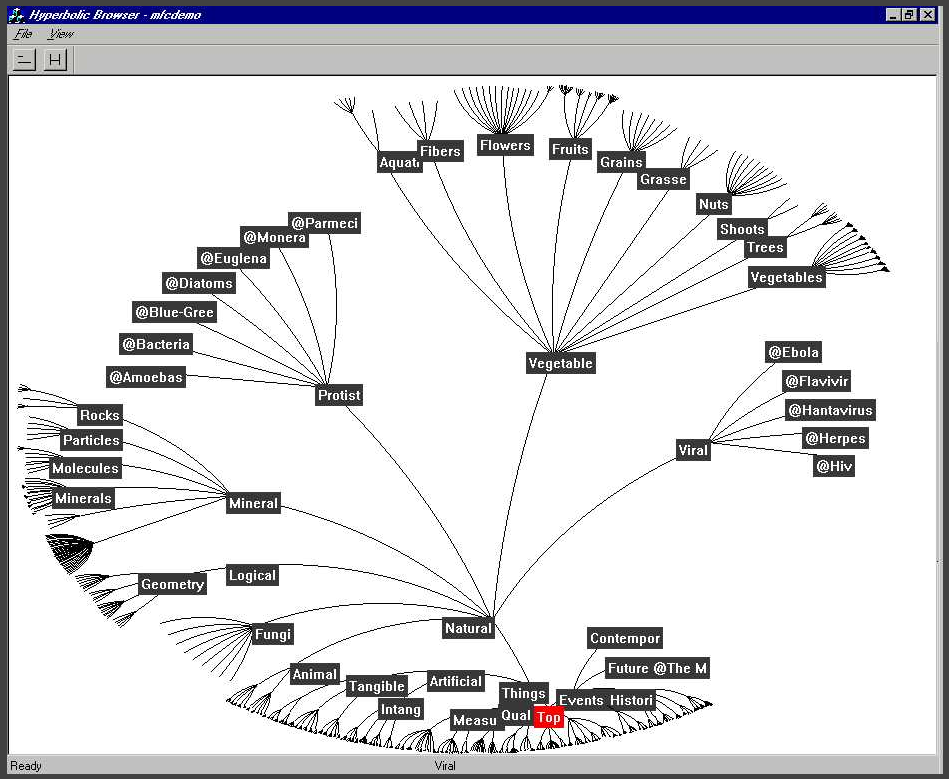
1. Query Database
2. Client-Side Indexing / Data Cubes
3. Prefetching
4. Approximation

These strategies are **not** mutually exclusive!
Systems can apply them in tandem.

Evaluation



vs.



Microsoft File Explorer

Xerox PARC Hyperbolic Tree

Visualization Evaluation Summary

Design and analyze visualization techniques in context of real-world use.

Time/error analyses can be insightful, but they don’t provide a complete picture.

Performance measures may be more suited to serious analysis than casual use?

Zoom Poll:
Top 3 Course Topics

The Future of Visualization

Where is more work required?

What emerging technologies and societal trends will impact visualization design?

What did you find most difficult in creating visualizations and designing techniques?

Thank You!

(Final Project Showcase in lecture on Friday!)