Visualization

CS 347
Maneesh Agrawala

Last time

Low-level cognitive models create **computational proxies of human mental behavior**, to help us characterize and understand how we will engage with a piece of technology

Model human processor, GOMS, KLM

Thinking in the world requires an understanding of cognition as well: **embodied cognition** emphasizes how we think with our bodies, whereas **distributed cognition** emphasizes how we think with the environment and other people in it

When our cognition is overloaded, performance decreases

Today

Choosing the right representation

Data, marks, visual attributes and encodings

Graphical perception

Frontiers of visualization research

Choosing the right representation

Cognitive amplification

Visualization can help, but ultimately this power comes from better **representation.** By better understanding human cognition, we can design technology that makes us smarter.

"The powers of cognition come from abstraction and representation: the ability to represent perceptions, experiences, and thoughts in some medium other than that in which they have occurred, abstracted away from irrelevant details." [Norman '94, Simon '81]

Example: Number scrabble

YOU READ THIS

[Simon 1988]

Take turns picking numbers in 1,2,3,4,5,6,7,8,9 without replacement

Win if any **three** of your numbers add up to **15**.

It's OK if you have extra numbers in your hand, as long as three of them add up to exactly 15.

Ready, set, go!

YOU READ THIS

I will show the series of moves from players A and B so far. Raise your hand when you know what B's best next move should be.

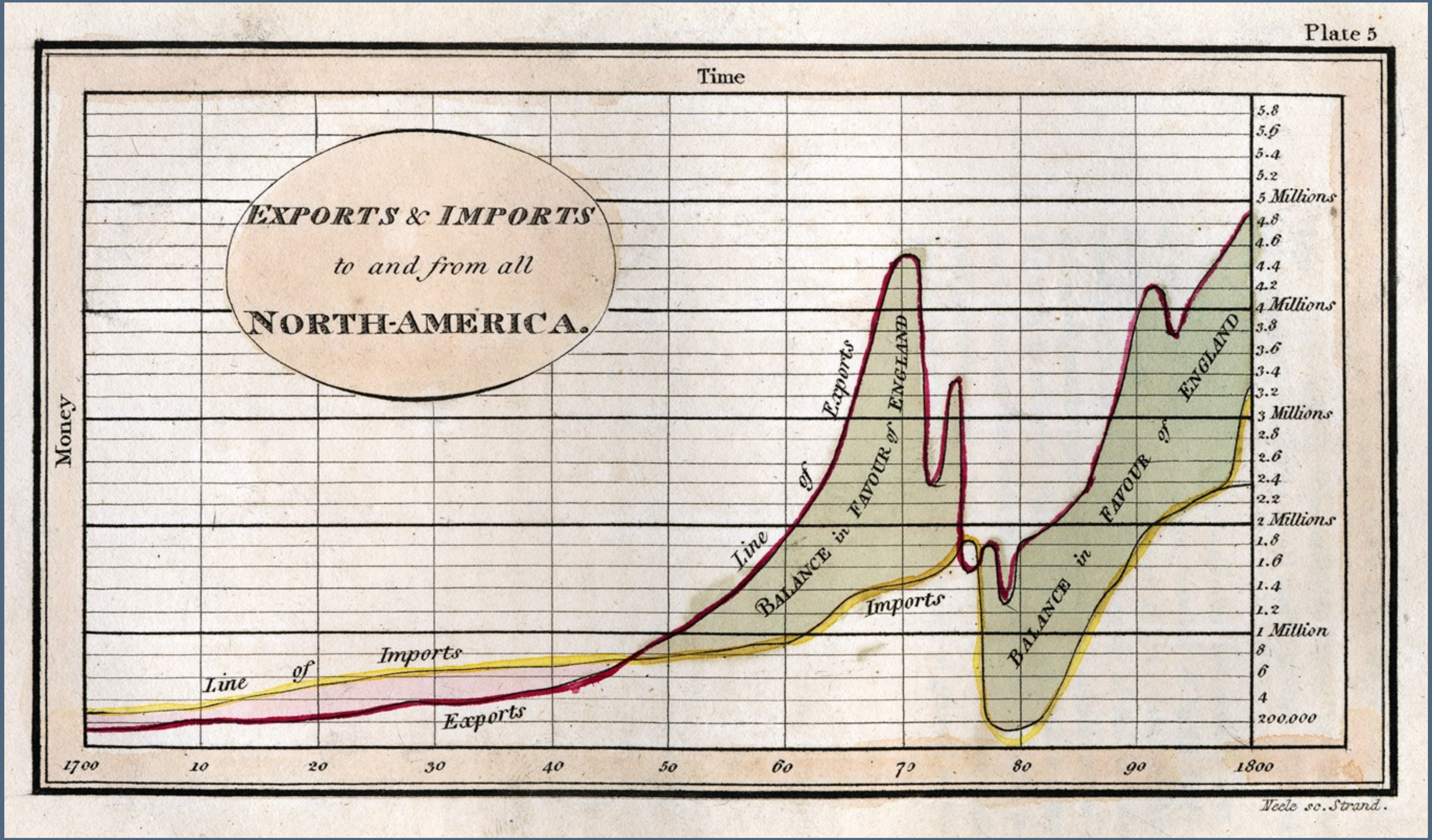
A takes 4
B takes 9
A takes 2
B takes 8
A takes 5
What should B do?

Re-encoding number scrabble

YOU READ THIS

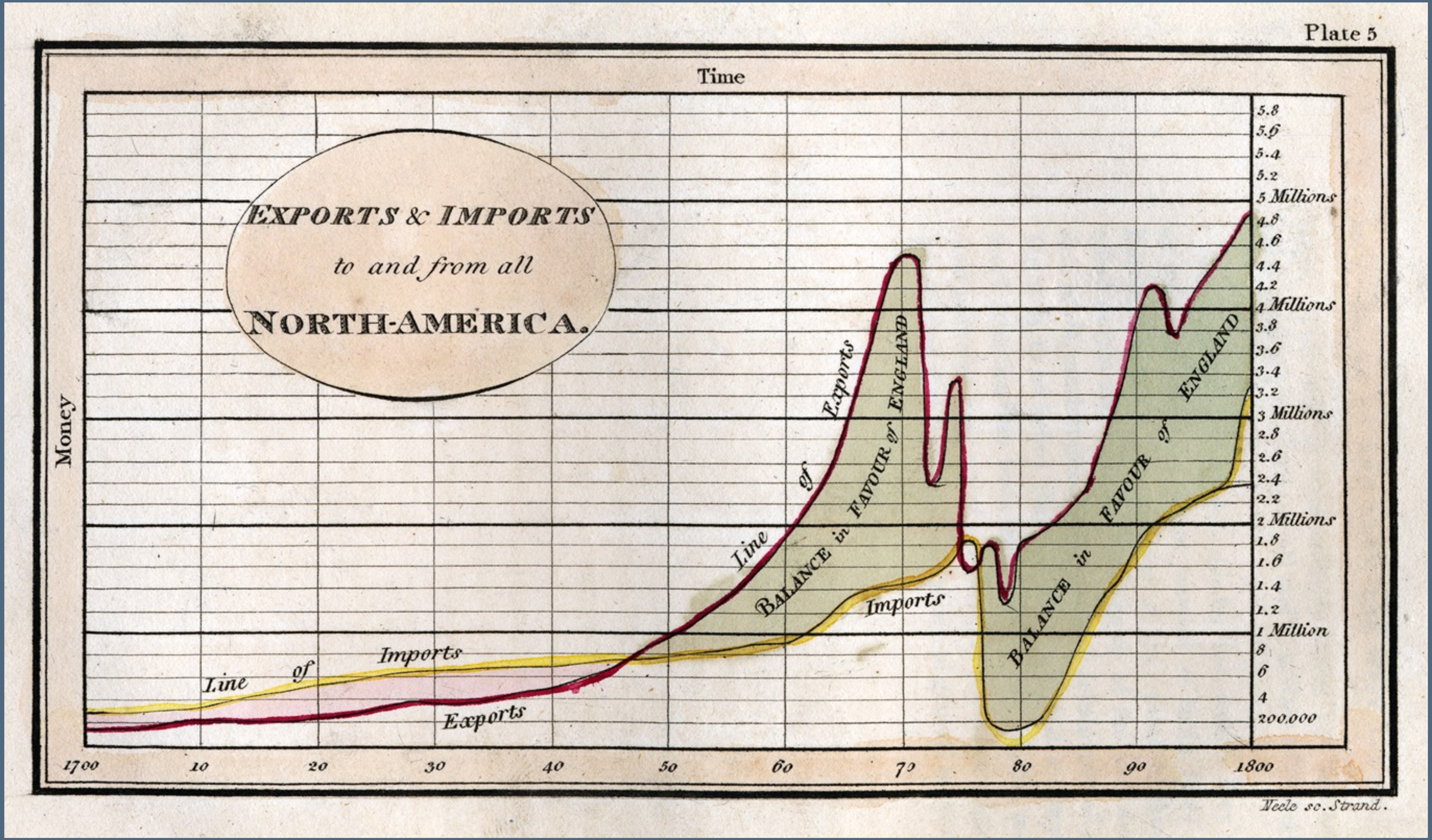
A4	B9	A2
3	A5	7
B8	1	6

Representation: Changing representation to spatial tic-tac-toe board facilitates choice



Exports and Imports to and from all North America [Playfair 1786-1801]

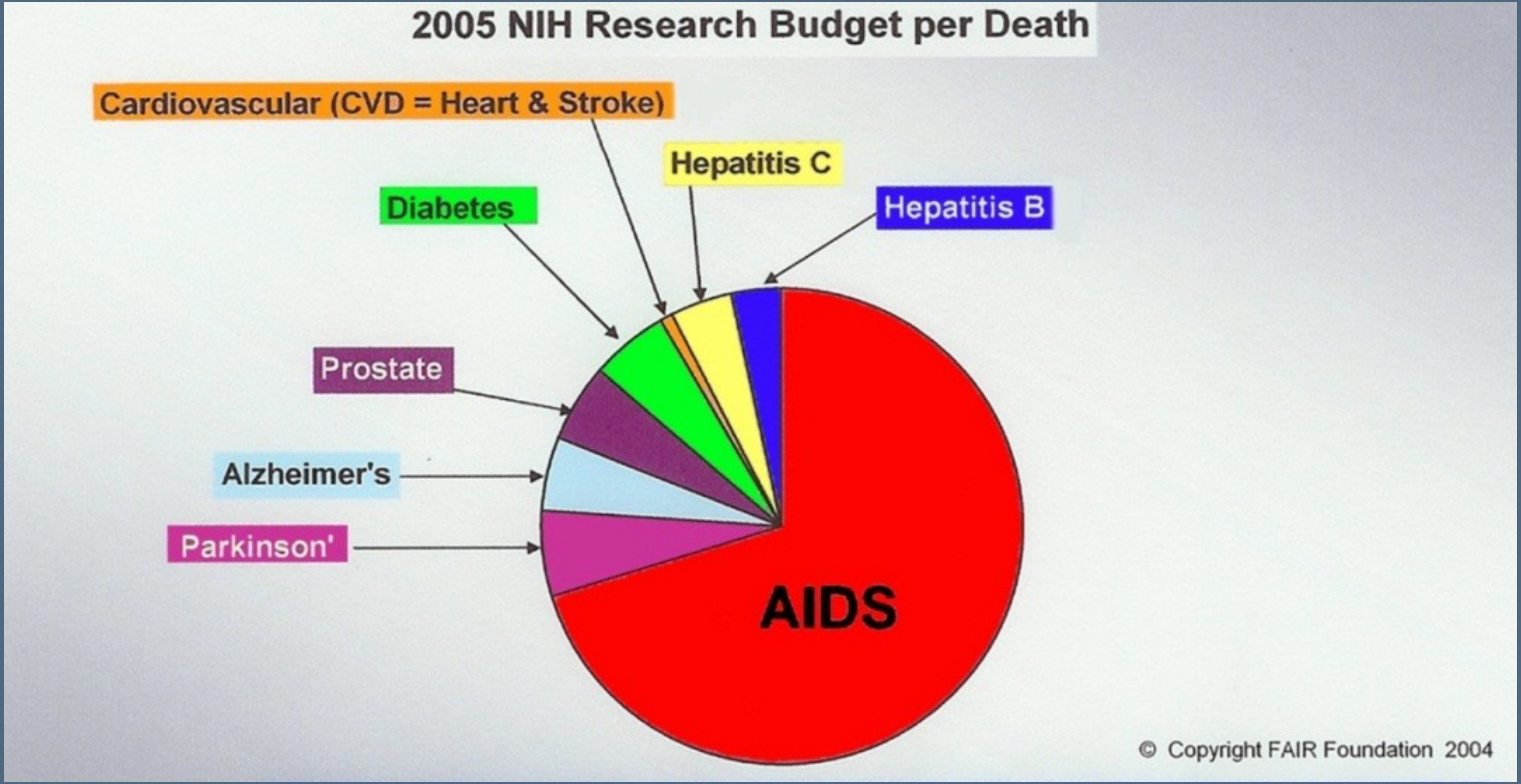
User's task: Understand balance of trade between England and North America over time



Exports and Imports to and from all North America [Playfair 1786-1801]

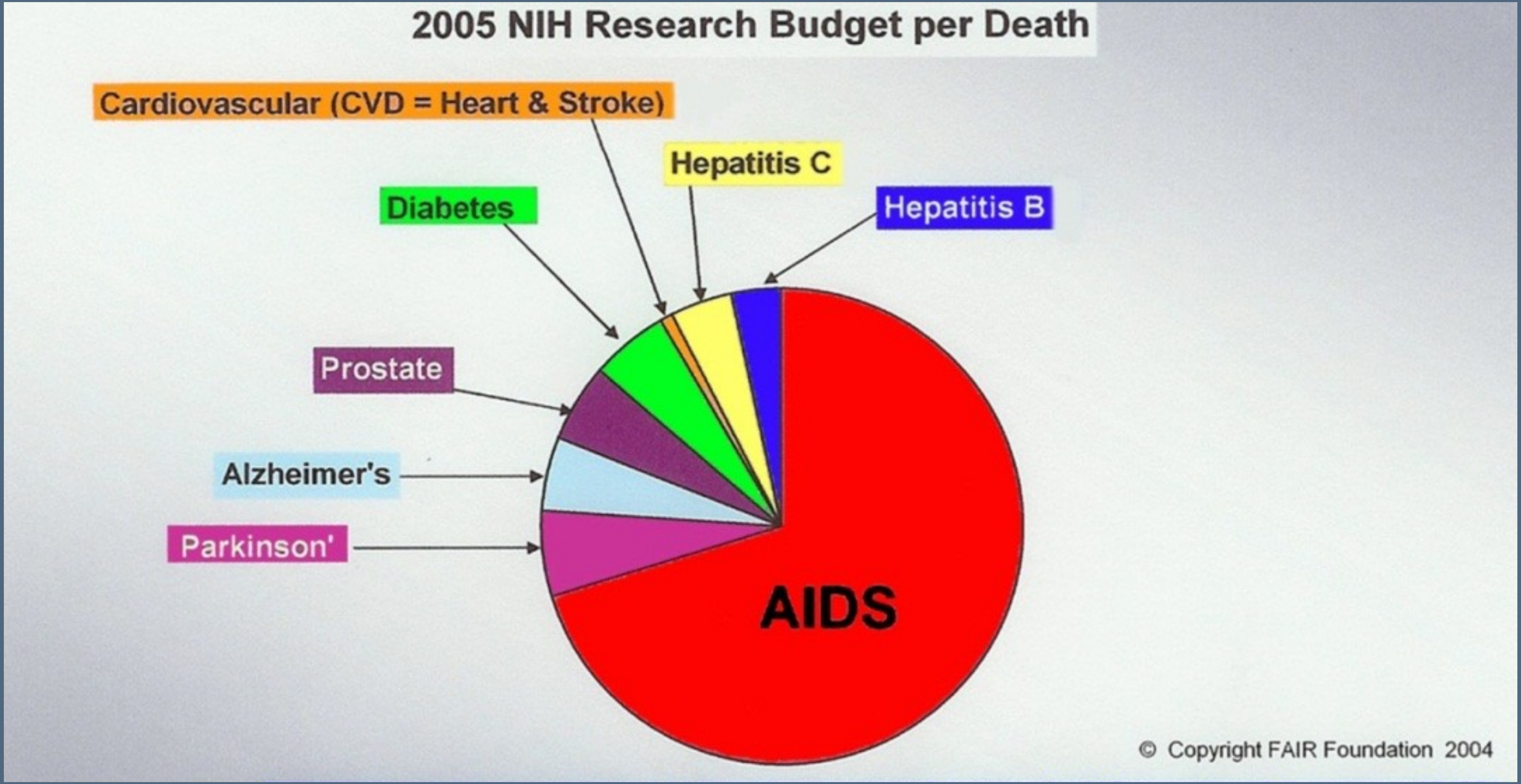
Important Information: Historical differences between exports and imports

Representation: Superimpose line charts of exports and imports to show historical pattern
Shade differences between lines to highlight balance against/in favor



Estimated 2005 NIH Research Budget per Death [FAIR Foundation 04]

User's task: Understand overall proportion of budget allocated to each disease
Compare proportion of budget allocated to each disease

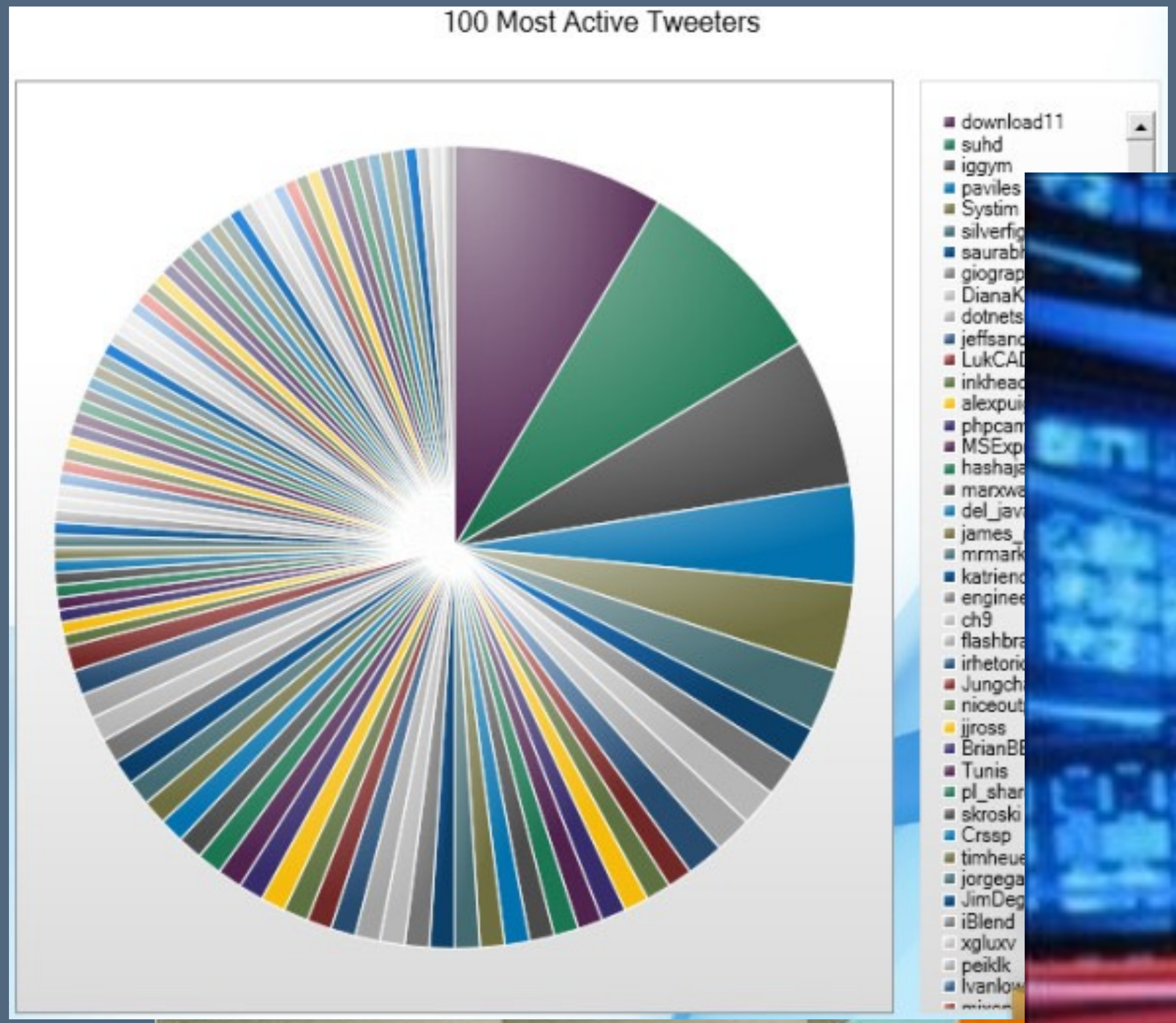


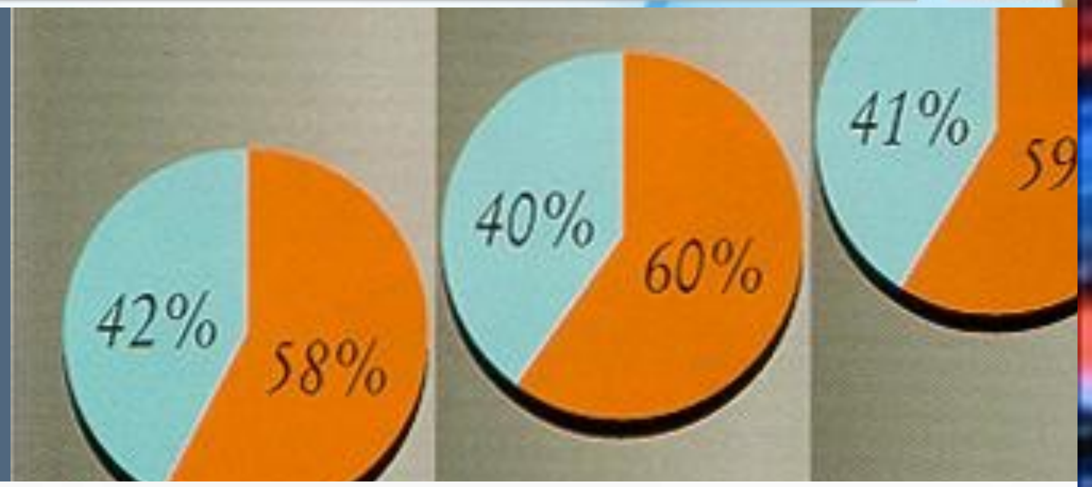
Estimated 2005 NIH Research Budget per Death [FAIR Foundation 04]

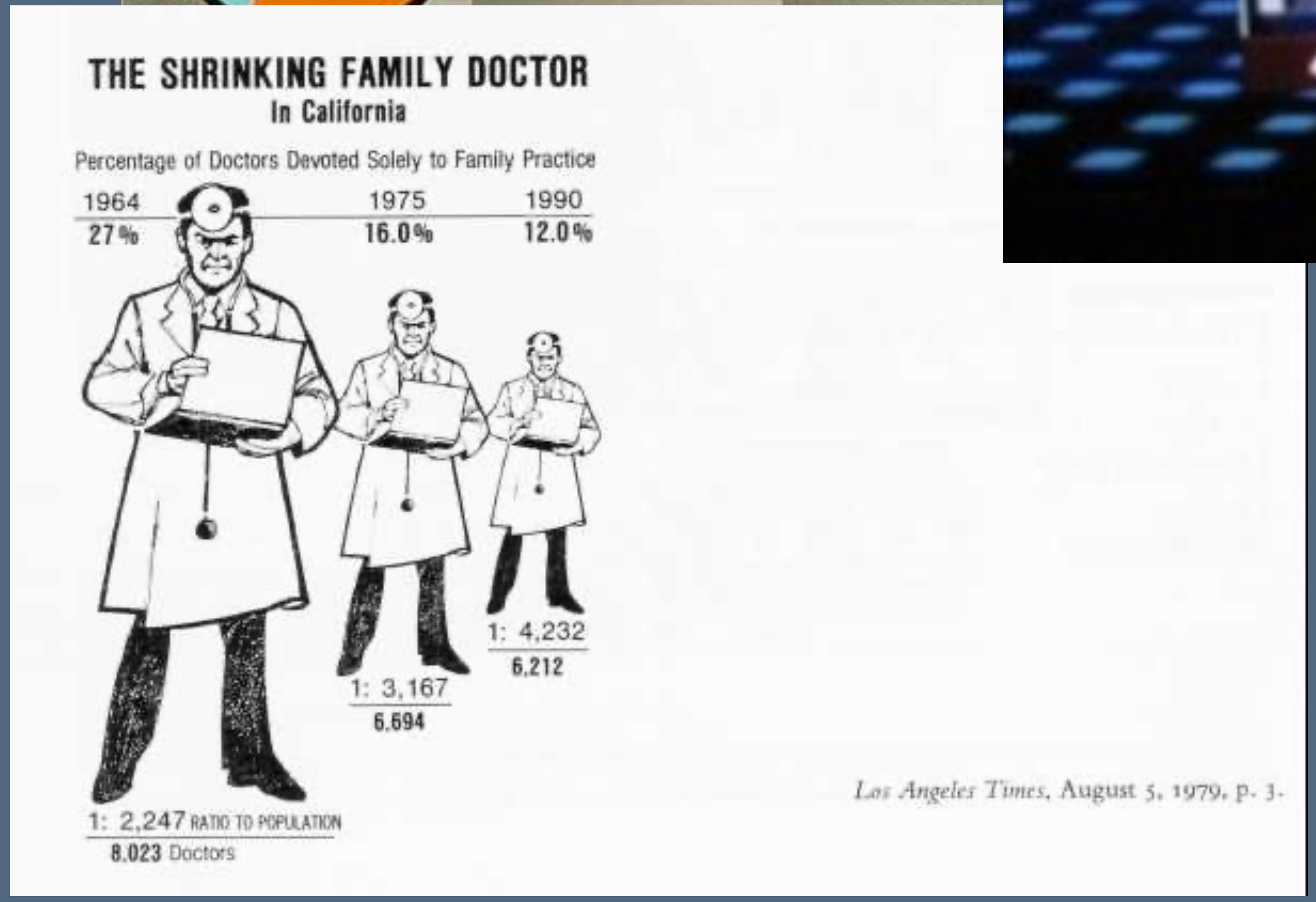
Important Information: Percentage of budget allocated to each disease

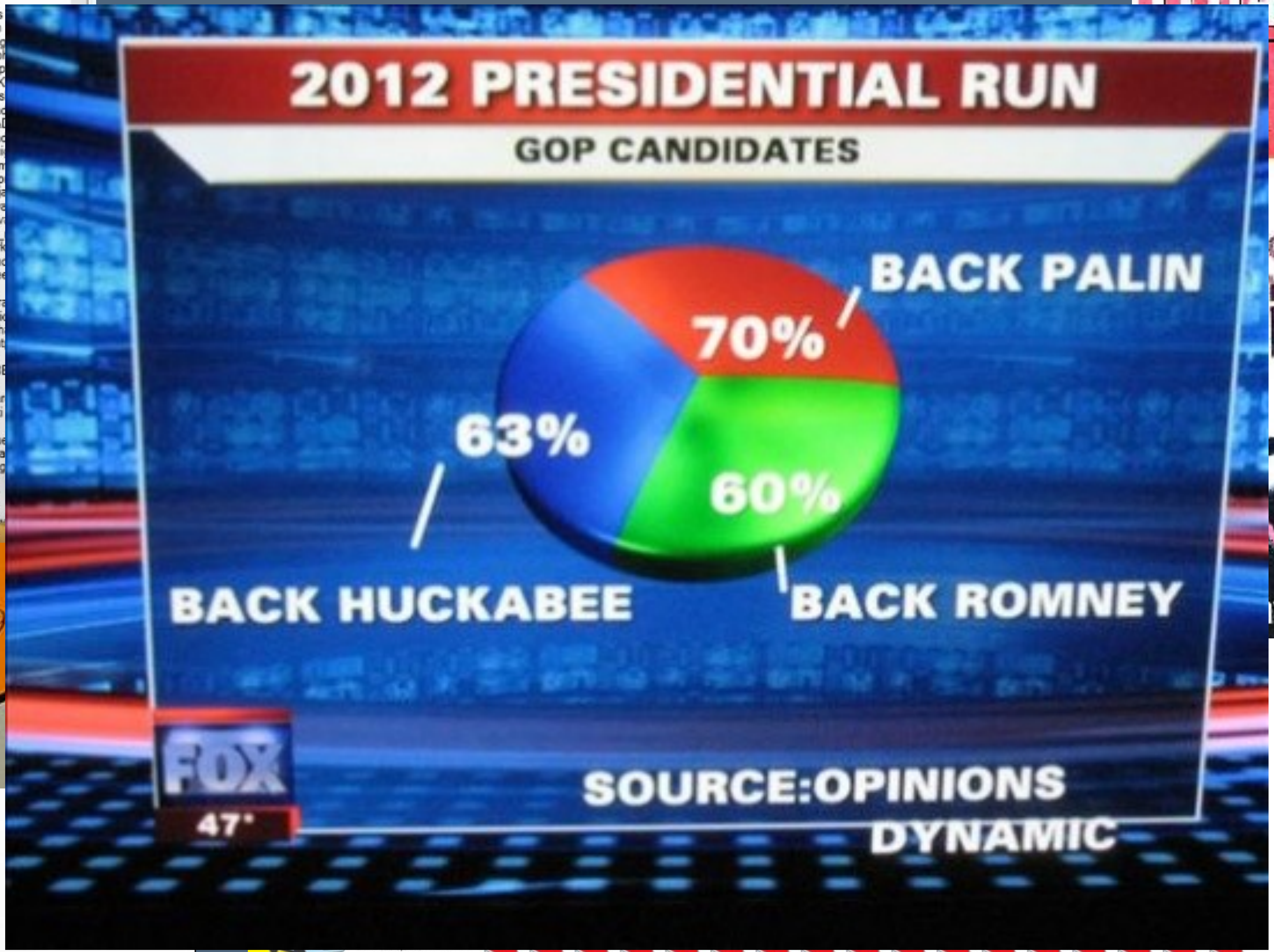
Representation: Encode budget using pie chart to emphasize part-to-whole relationship

~~Allow comparison between diseases by comparing pie slices (angles or areas)~~

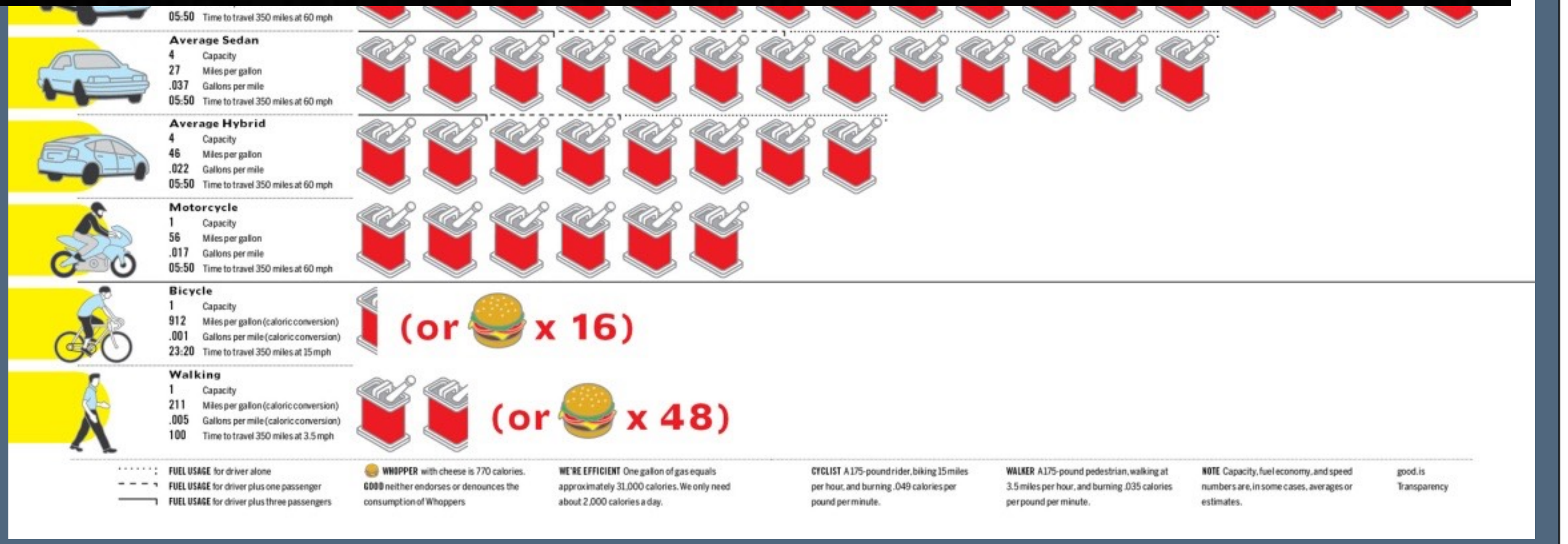

100 Most Active Tweeters
download11
suhd
iggym

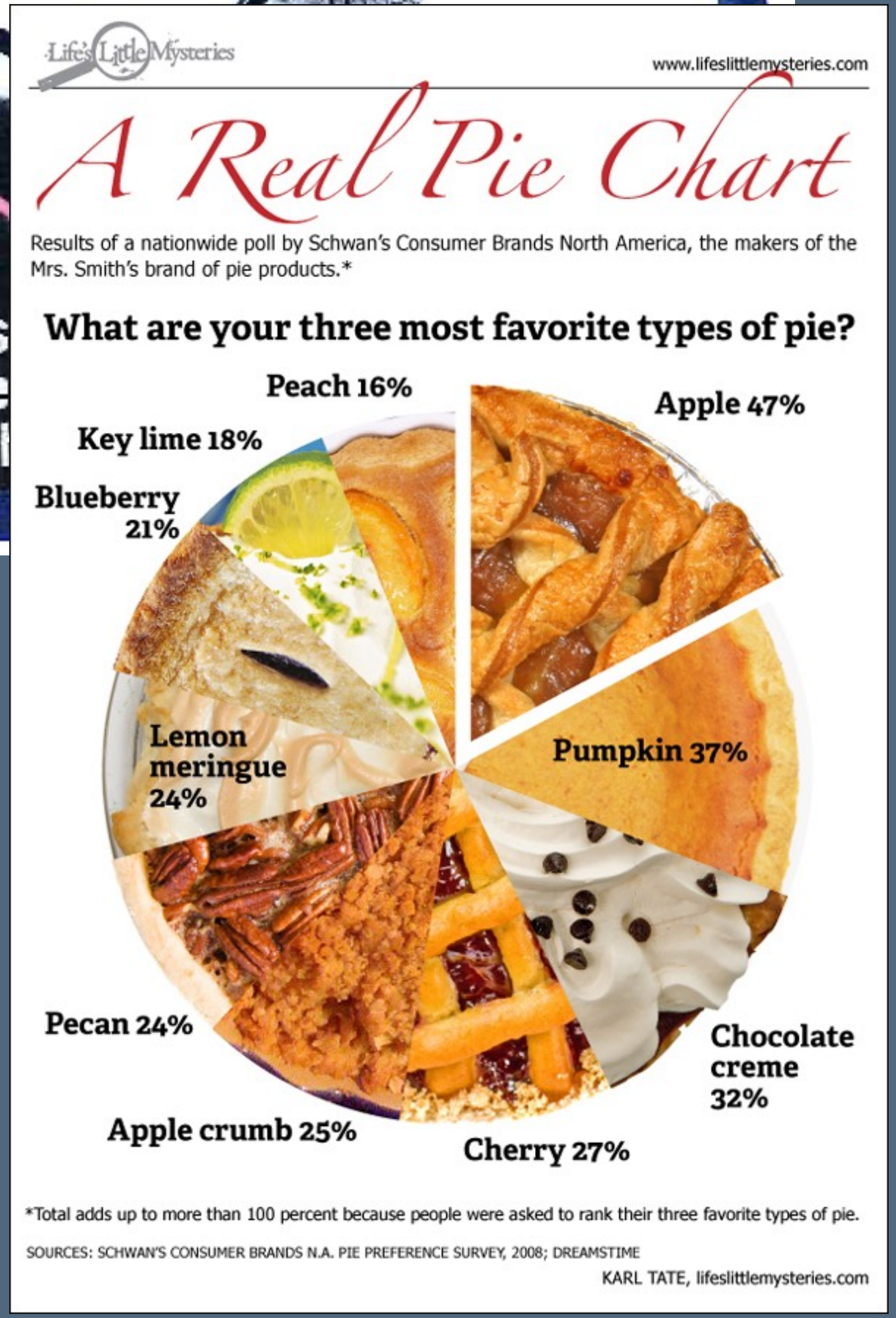

42%
58%
40%
60%
41%


THE SHRINKING FAMILY DOCTOR
In California
Percentage of Doctors Devoted Solely to Family Practice
1964
27%
1975
16.0%
1990
12.0%
1: 2,247 RATIO TO POPULATION
8,023 Doctors
1: 3,167
5,694
1: 4,232
6,212
Los Angeles Times, August 5, 1979, p. 3.


2012 PRESIDENTIAL RUN
GOP CANDIDATES
BACK PALIN
70%
63%
60%
BACK HUCKABEE
BACK ROMNEY
FOX
47°
SOURCE:OPINIONS
DYNAMIC


ITHACA
ES
Planning Board approves
Widewaters development
IC students occupy Job Hall
Cayuga Vocal Ensemble
ushers in the holidays with
"Judas Maccabaeus"
sing
igns
Cornell
University
Tuition


Average Sedan
Average Hybrid
Motorcycle
Bicycle
Walking
(or x 16)
(or x 48)


Life's Little Mysteries
www.lifeslittlemysteries.com
A Real Pie Chart
Results of a nationwide poll by Schwan's Consumer Brands North America, the makers of the Mrs. Smith's brand of pie products.*
What are your three most favorite types of pie?
Peach 16%
Apple 47%
Key lime 18%
Blueberry 21%
Lemon meringue 24%
Pumpkin 37%
Pecan 24%
Chocolate creme 32%
Apple crumb 25%
Cherry 27%
*Total adds up to more than 100 percent because people were asked to rank their three favorite types of pie.
SOURCES: SCHWAN'S CONSUMER BRANDS N.A. PIE PREFERENCE SURVEY, 2008; DREAMSTIME
KARL TATE, lifeslittlemysteries.com

Data, marks, visual attributes and encodings

On the theory of
scales of measurements
S. S. Stevens, 1946

Data Types

N - Nominal (labels)

Fruits: Apples, oranges, …

Operations: **=, ≠**

O - Ordered

Quality of eggs: Grade AA, A, B

Operations: **=, ≠, <, >**

Q - Interval (location of zero arbitrary)

Dates: Jan, 19, 2016; Loc.: (LAT 33.98, LON -118.45)

Like a geometric point. Cannot compare directly

Only differences (i.e. intervals) may be compared

Operations: **=, ≠, <, >, -**

Q - Ratio (location of zero fixed)

Physical measurement: Length, Mass, …

Counts and amounts

Like a geometric vector, origin is meaningful

Operations: **=, ≠, <, >, -, ÷**

U.S. Census Data

People Count: # of people in group

Year: 1850 – 2000 (every decade)

Age: 0 – 90+

Sex: Male, Female

Marital Status: Single, Married, Divorced, …

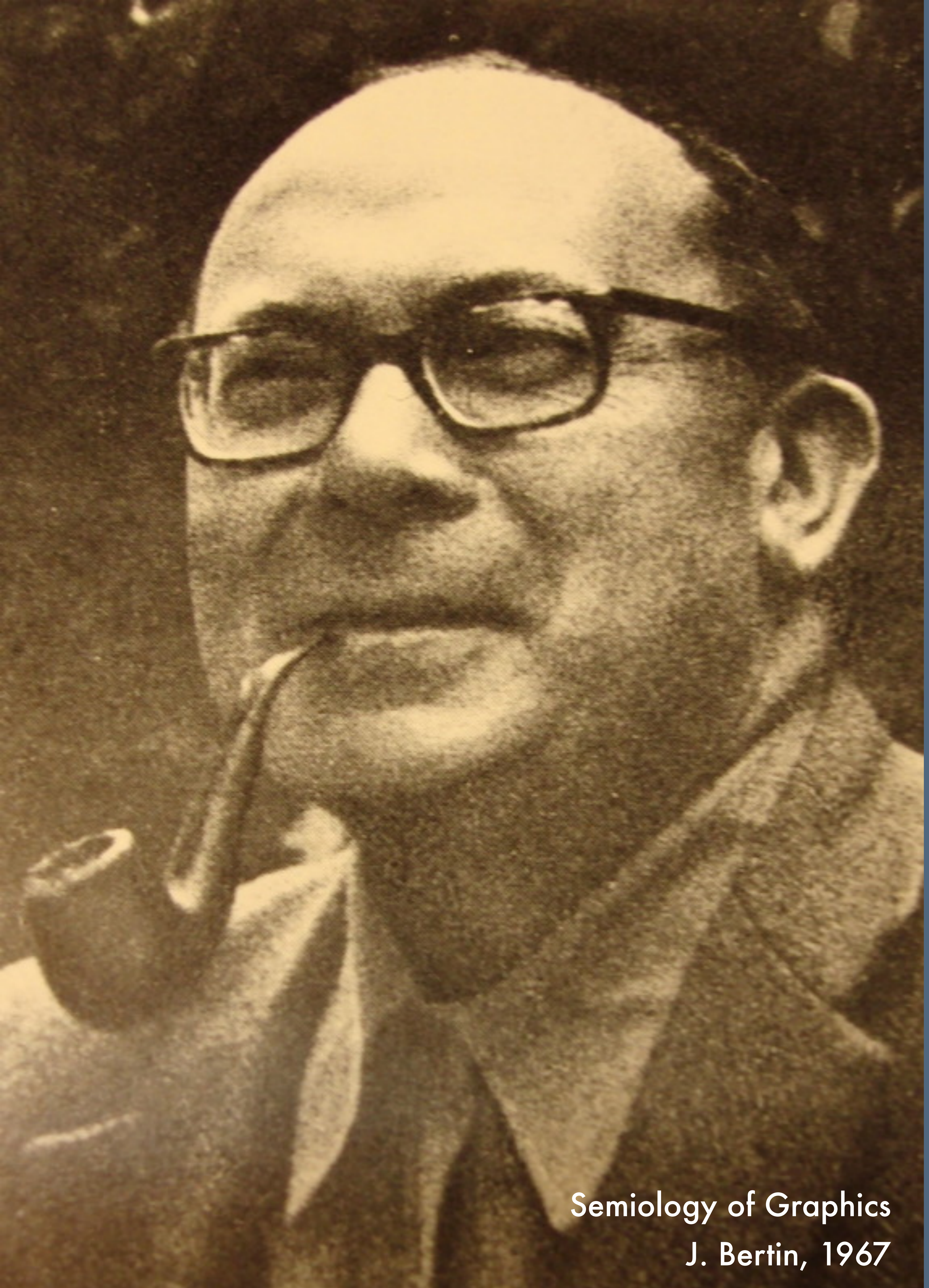
2348 data points

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343
23	1850	50	0	2	286580
24	1850	55	0	1	194080
25	1850	55	0	2	187208
26	1850	60	0	1	174976

Census N, O, Q

- People Count: Q-Ratio
- Year: Q-Interval (maybe O)
- Age: Q-Ratio (maybe O)
- Sex: N
- Marital Status: N

	A	B	C	D	E
1	year	age	marst	sex	people
2	1850	0	0	1	1483789
3	1850	0	0	2	1450376
4	1850	5	0	1	1411067
5	1850	5	0	2	1359668
6	1850	10	0	1	1260099
7	1850	10	0	2	1216114
8	1850	15	0	1	1077133
9	1850	15	0	2	1110619
10	1850	20	0	1	1017281
11	1850	20	0	2	1003841
12	1850	25	0	1	862547
13	1850	25	0	2	799482
14	1850	30	0	1	730638
15	1850	30	0	2	639636
16	1850	35	0	1	588487
17	1850	35	0	2	505012
18	1850	40	0	1	475911
19	1850	40	0	2	428185
20	1850	45	0	1	384211
21	1850	45	0	2	341254
22	1850	50	0	1	321343
23	1850	50	0	2	286580
24	1850	55	0	1	194080
25	1850	55	0	2	187208
26	1850	60	0	1	174976



Semiology of Graphics
J. Bertin, 1967

Marks & Visual Attributes

Marks: geometric primitives

points lines areas

Visual Attributes: control mark appearance

Position (2x)

Size

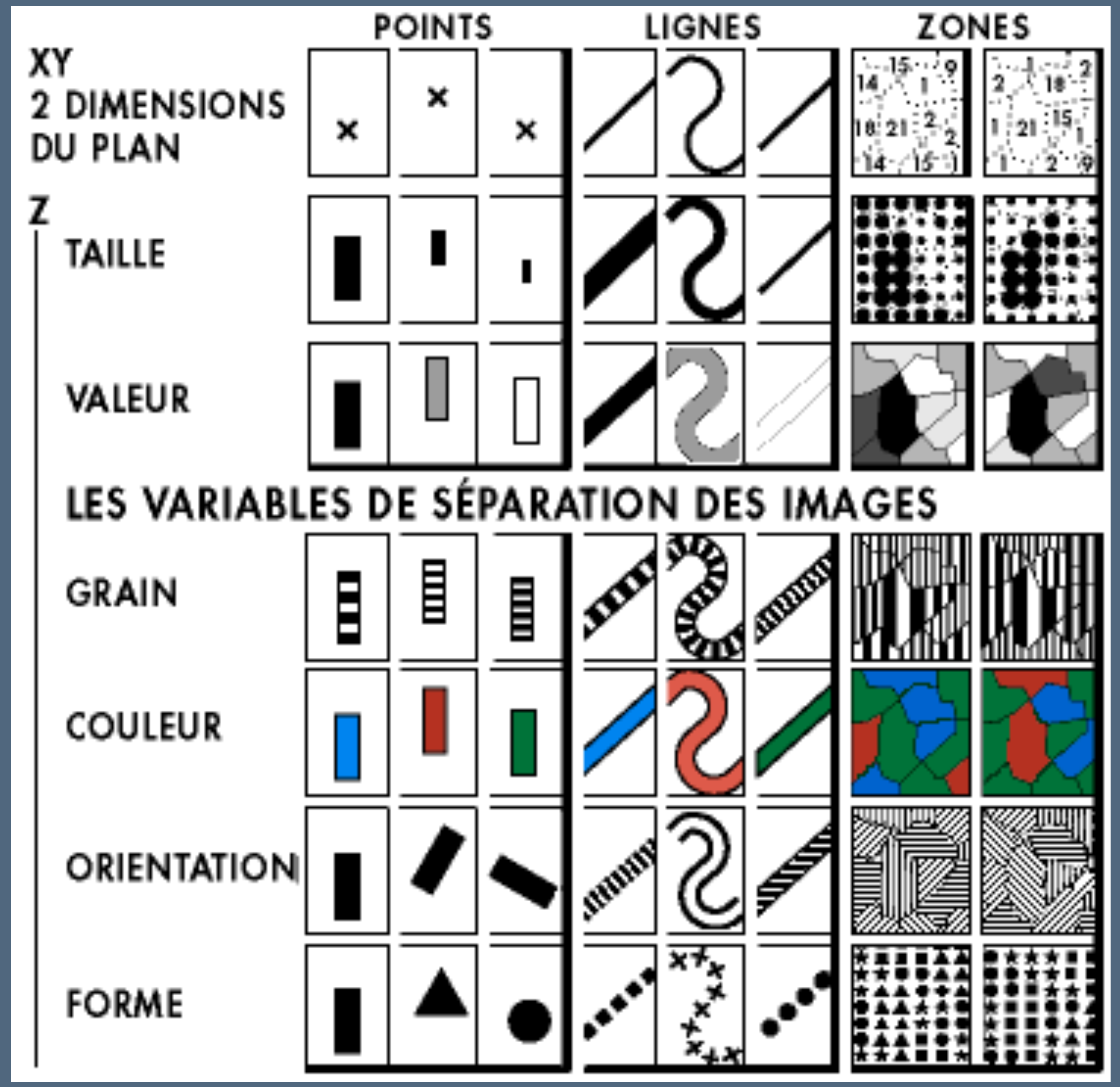
Value

Texture

Color

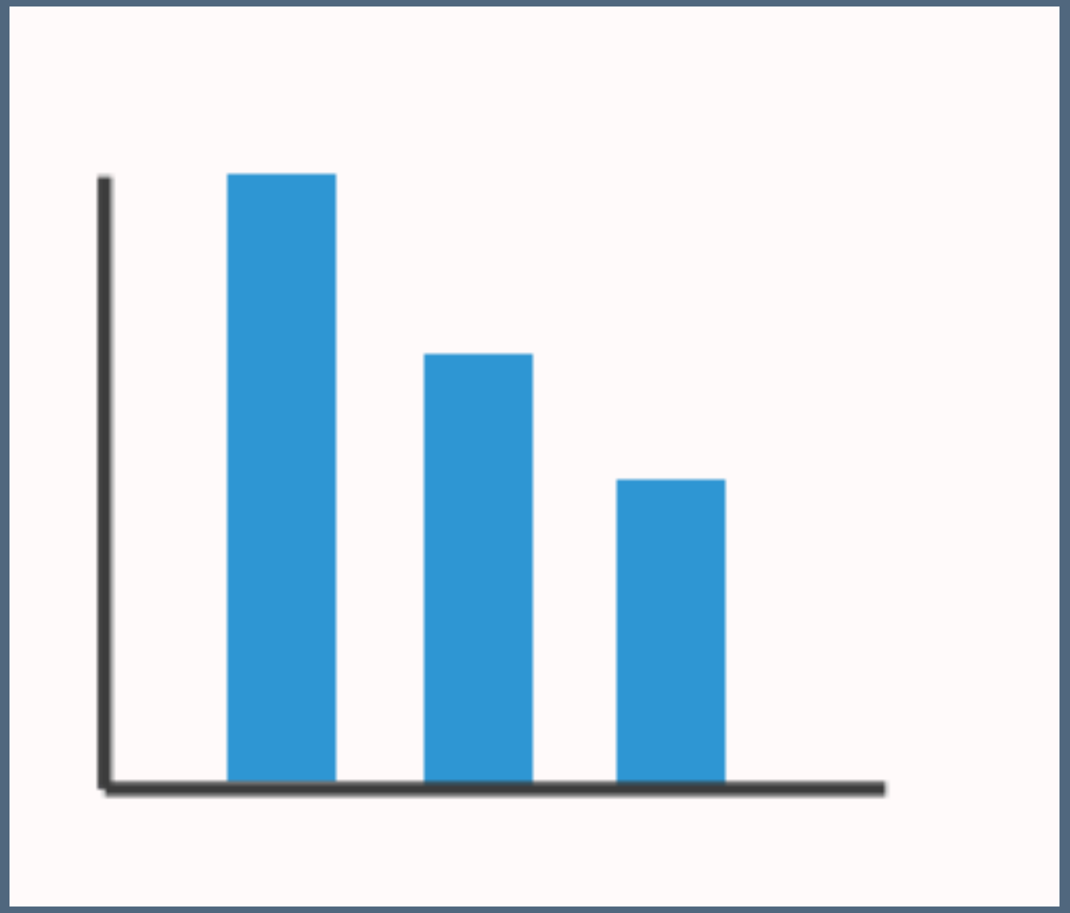
Orientation

Shape

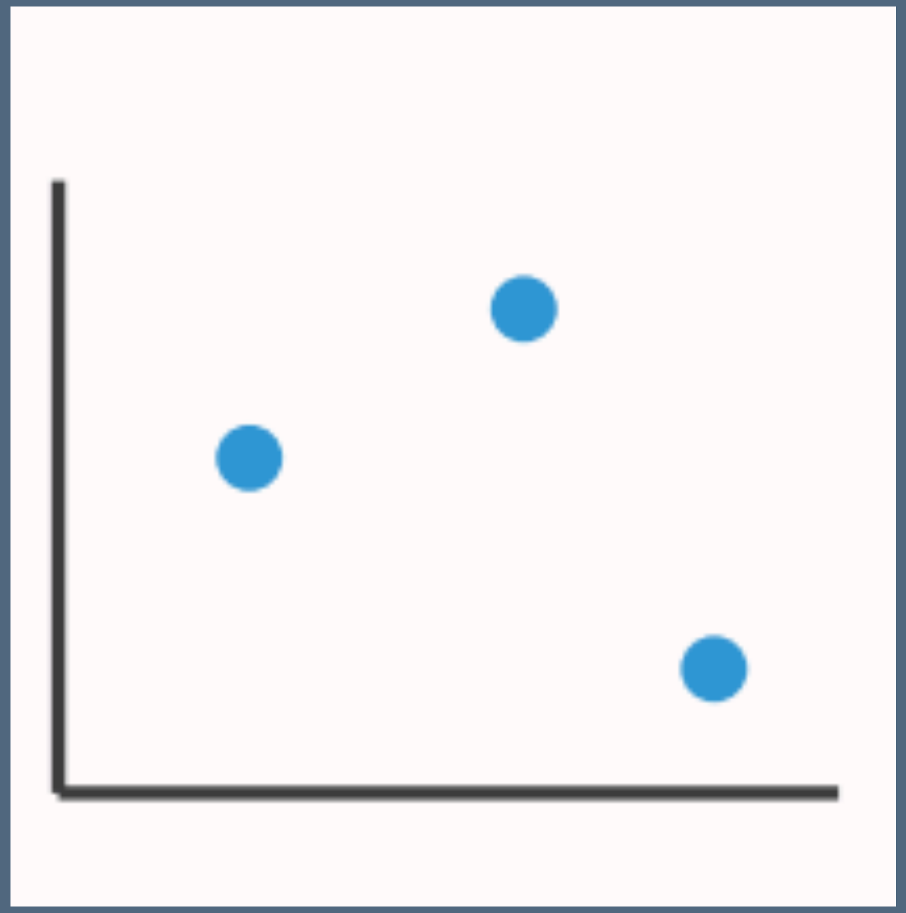


Encodings

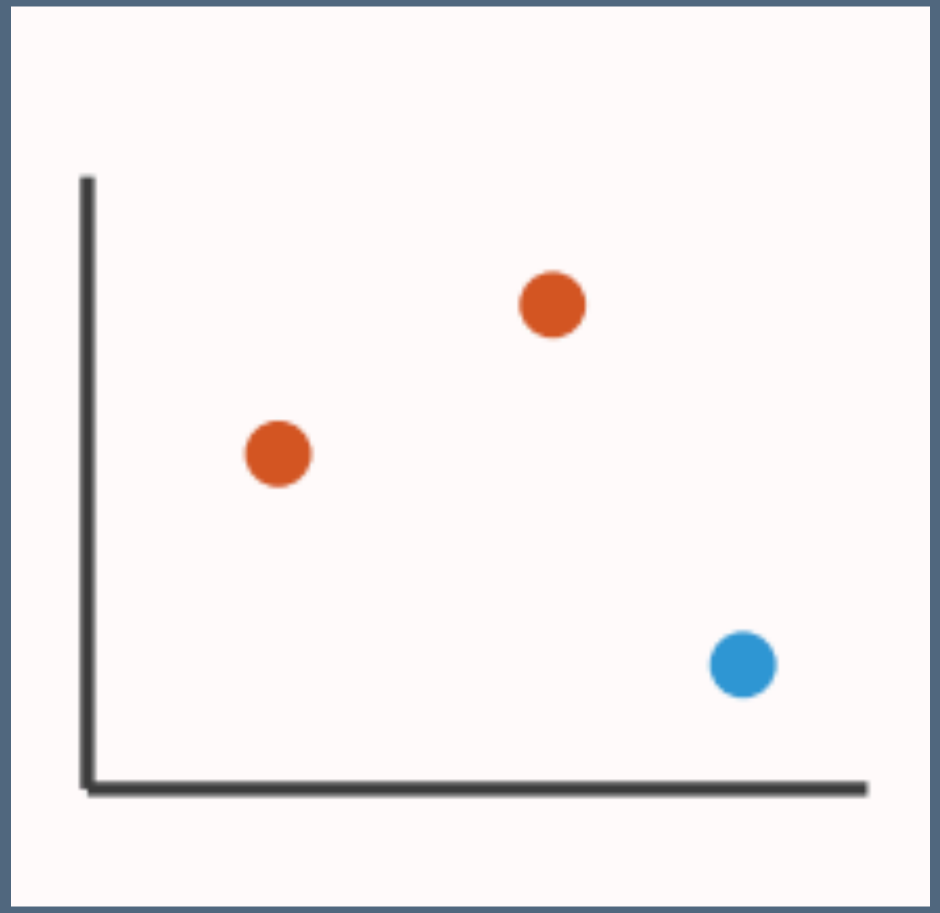
A map from data to visual attributes of marks



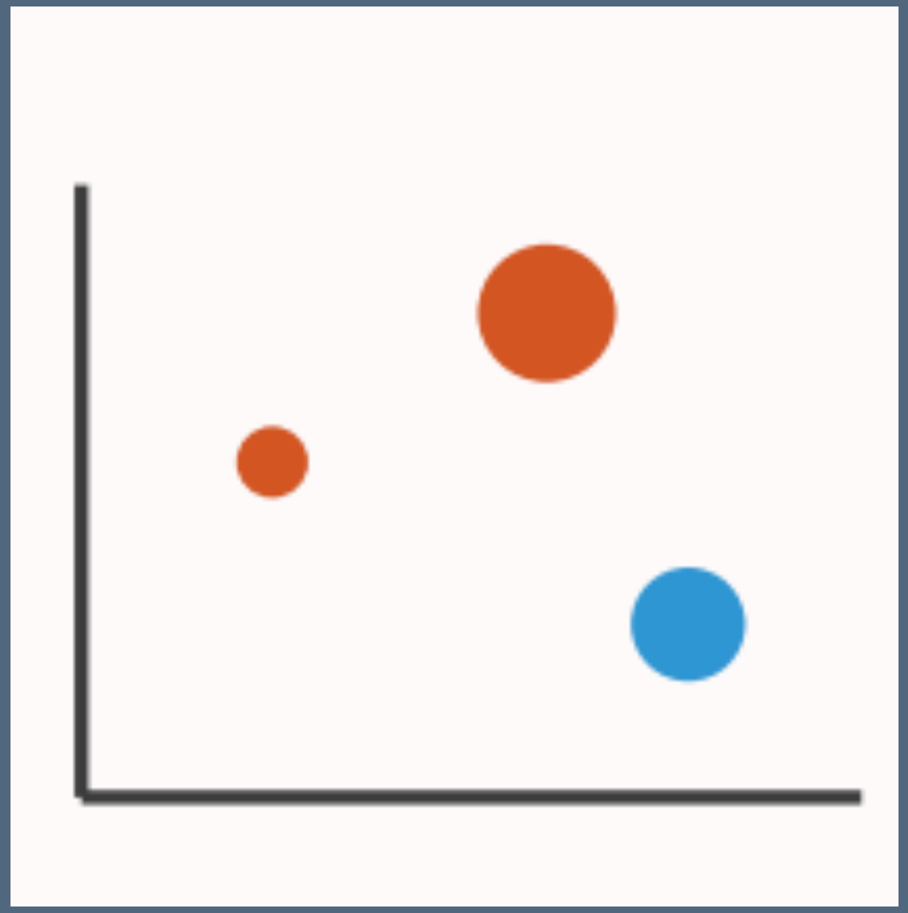
Mark: lines
county(N) ➞ x
population(Q) ➞ size or length



Mark: point
acreage(Q) ➞ x
population(Q) ➞ y



Mark: point
acreage(Q) ➞ x
population(Q) ➞ y
county(N) ➞ color



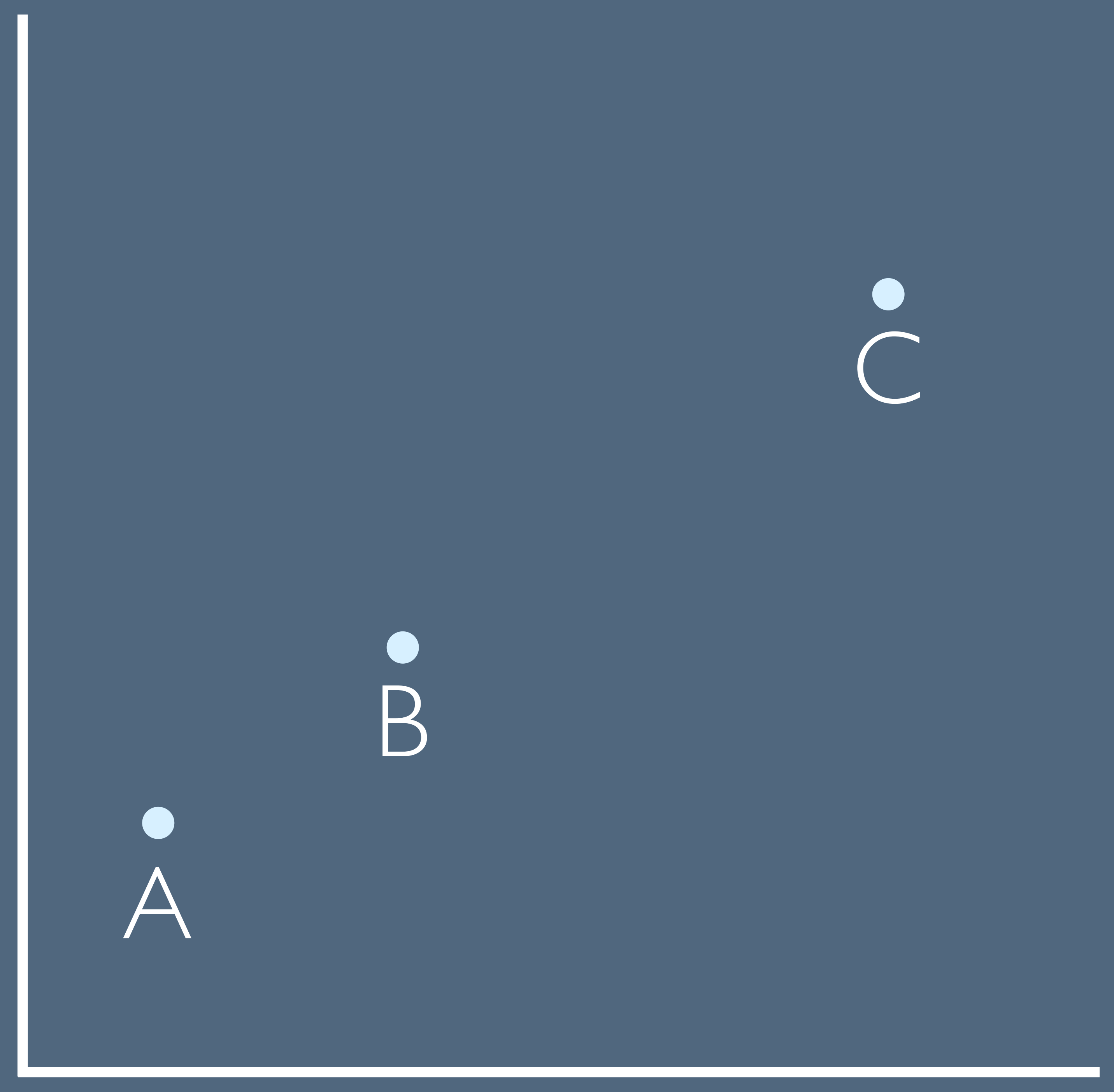
Mark: point
acreage(Q) ➞ x
population(Q) ➞ y
county(N) ➞ color
avg_income(Q) ➞ size

"Best" encoding based on **perceptual effectiveness** of visual attribute for data type

Graphical perception

How we look at graphs

Position



We immediately notice that:

- A, B, C are distinguishable
- Points are collinear. B is between A and C
- BC is twice as long as AB

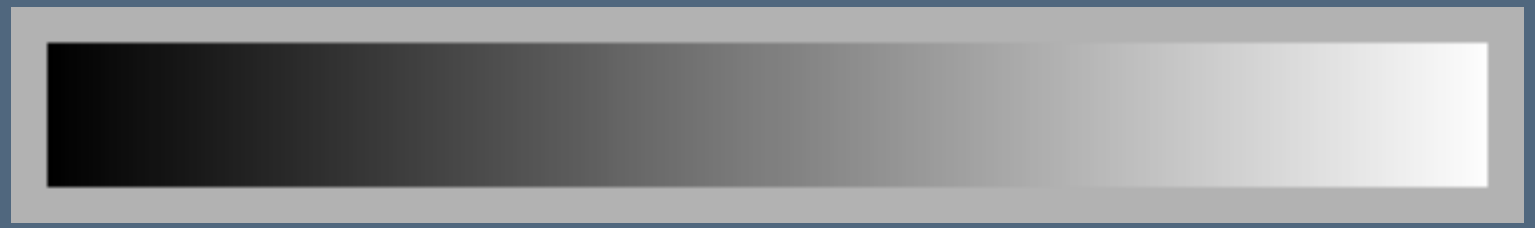
Position encodes quantitative data well

Color

Value or gray level is perceived as ordered

So, it encodes ordinal data well

But, fine differences hard to perceived, so encodes quantitative data less well



Hue is typically perceived as unordered

So, hue encodes nominal data well

Bertin's "Levels of Organization"

Position	N	O	Q
Size	N	O	Q
Value	N	O	Q
Texture	N	O	
Color	N		
Orientation	N		
Shape	N		

N	Nominal
O	Ordered
Q	Quantitative

Which is brighter?

(128, 128, 128) (130, 130, 130)

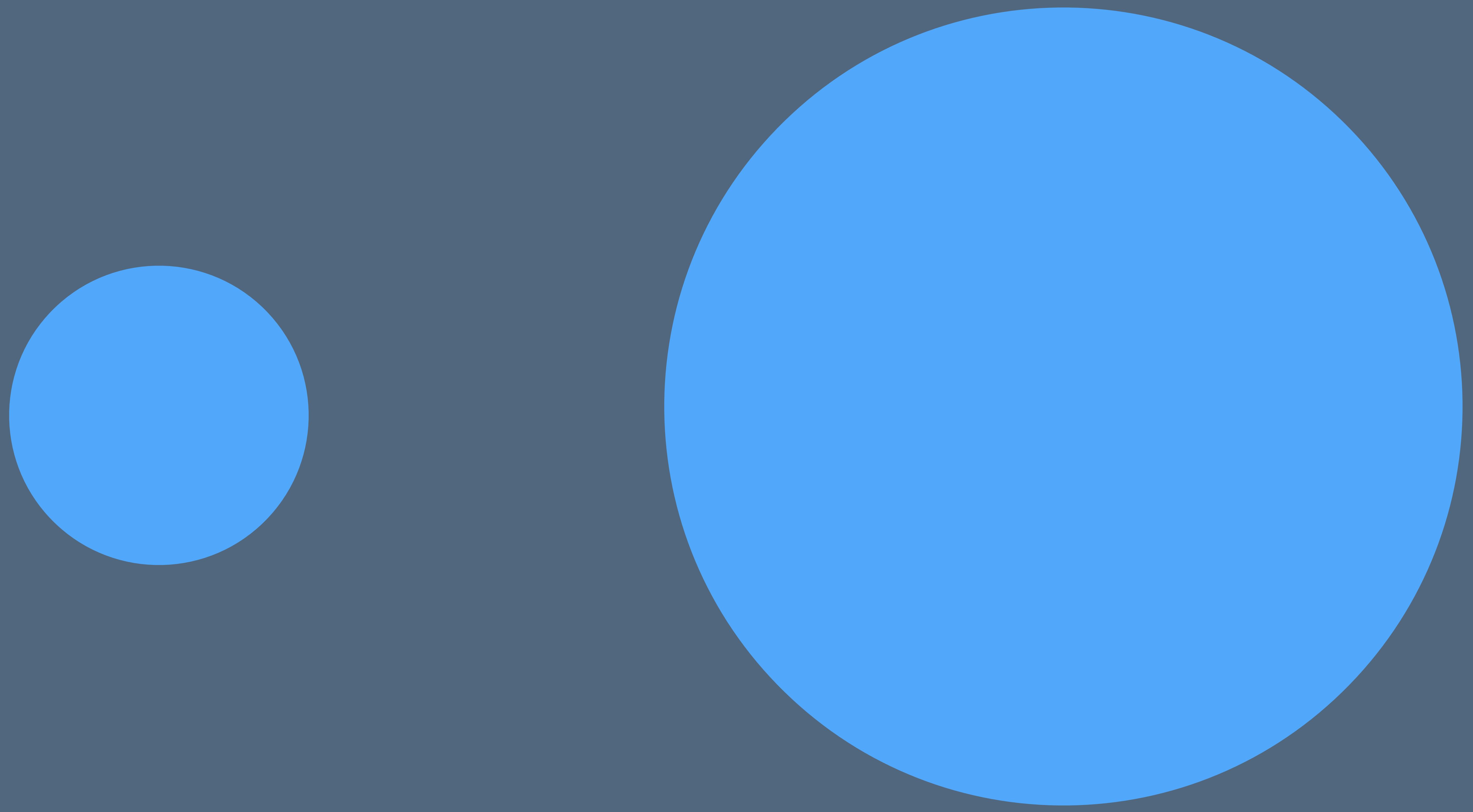
Which is brighter?

Just Noticeable Differences

JND (Weber's Law)

$$\Delta S = k \frac{\Delta I}{I}$$

- Ratios more important than magnitude
- Most continuous variations in stimuli are perceived in discrete steps



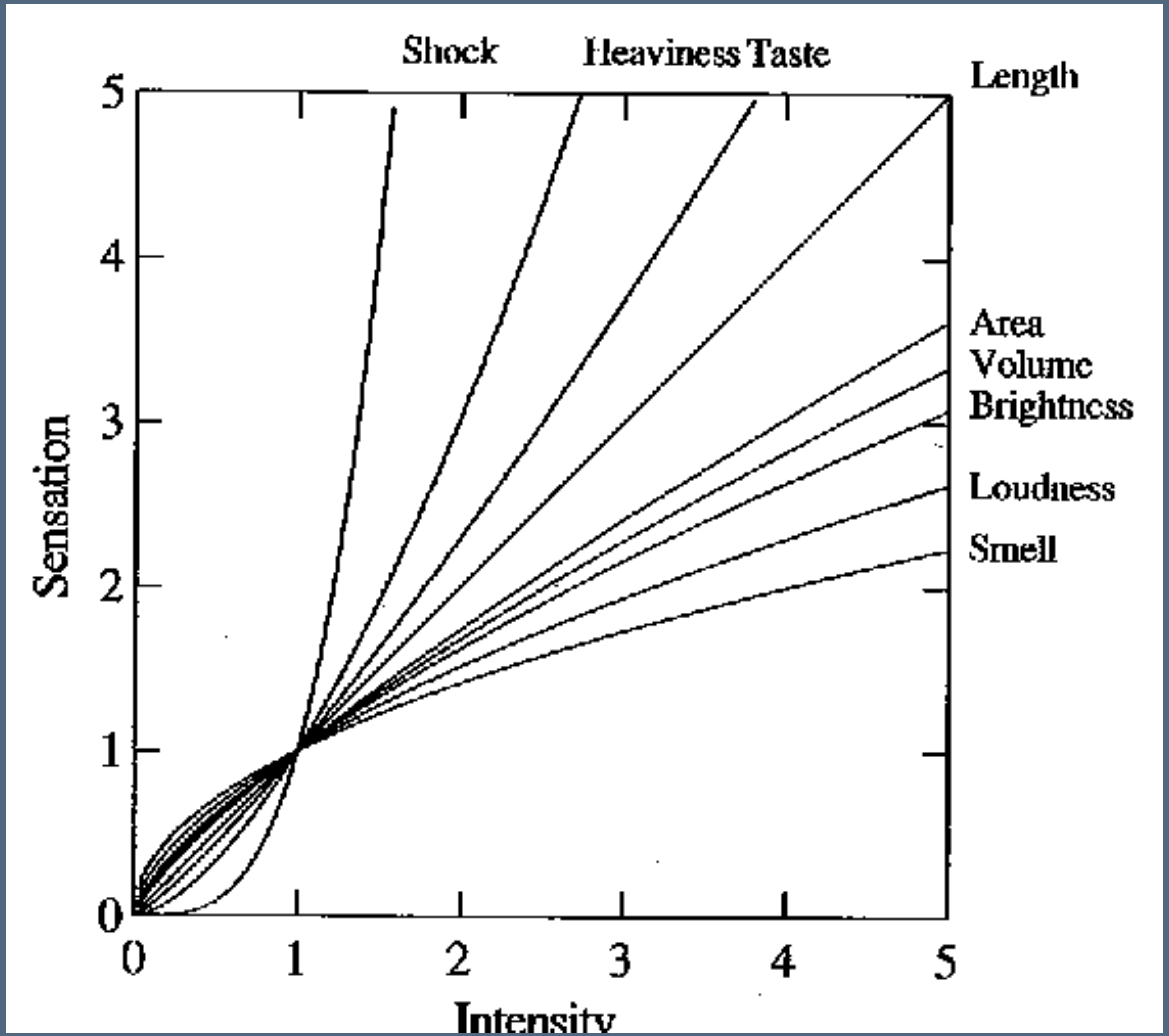
Compare areas of circles

Compare lengths of bars

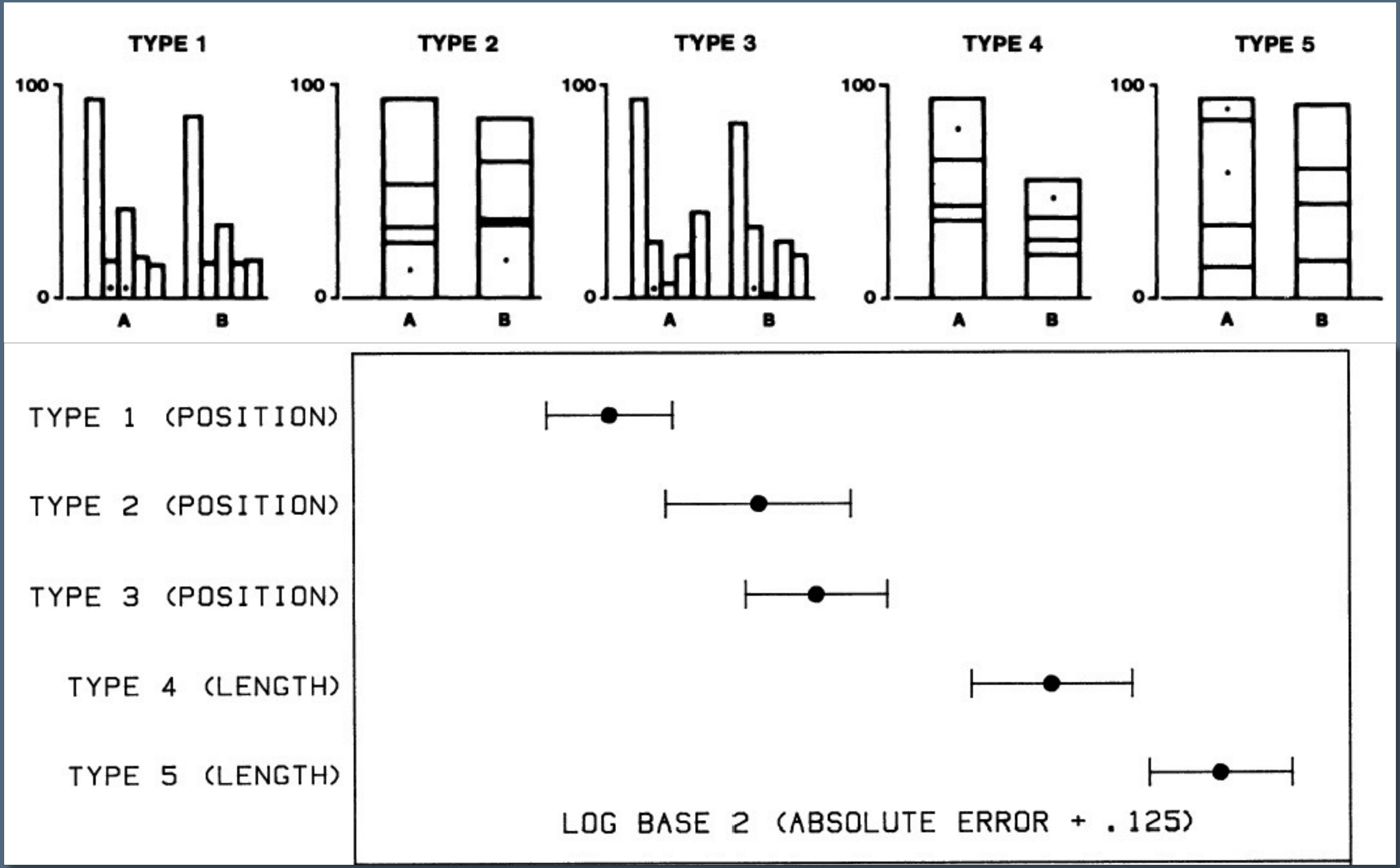
Steven's Power Law

$$S = kI^p$$

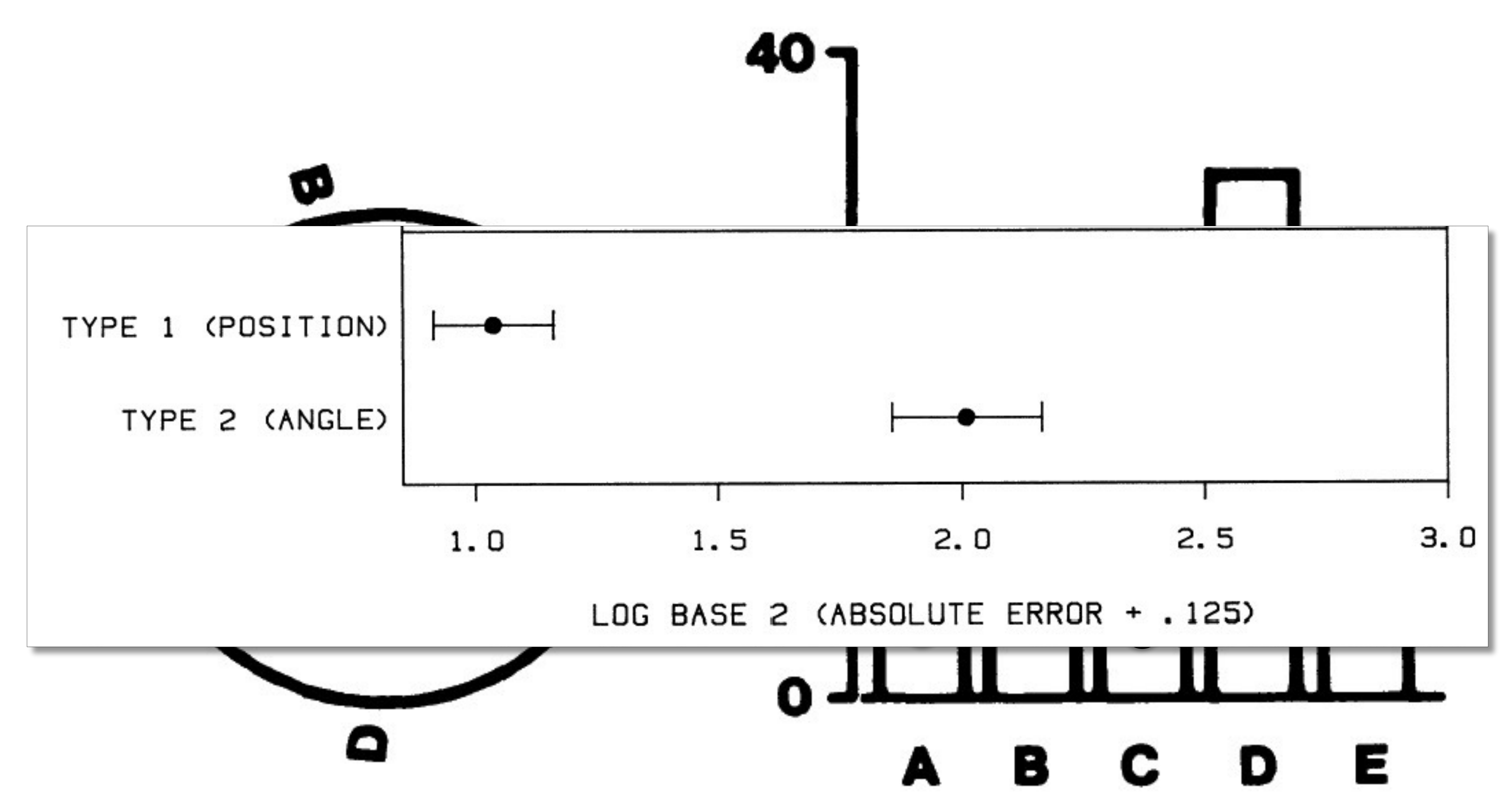
p < 1 : underestimate
p > 1 : overestimate



[graph from Wilkinson 99, based on Stevens 61]

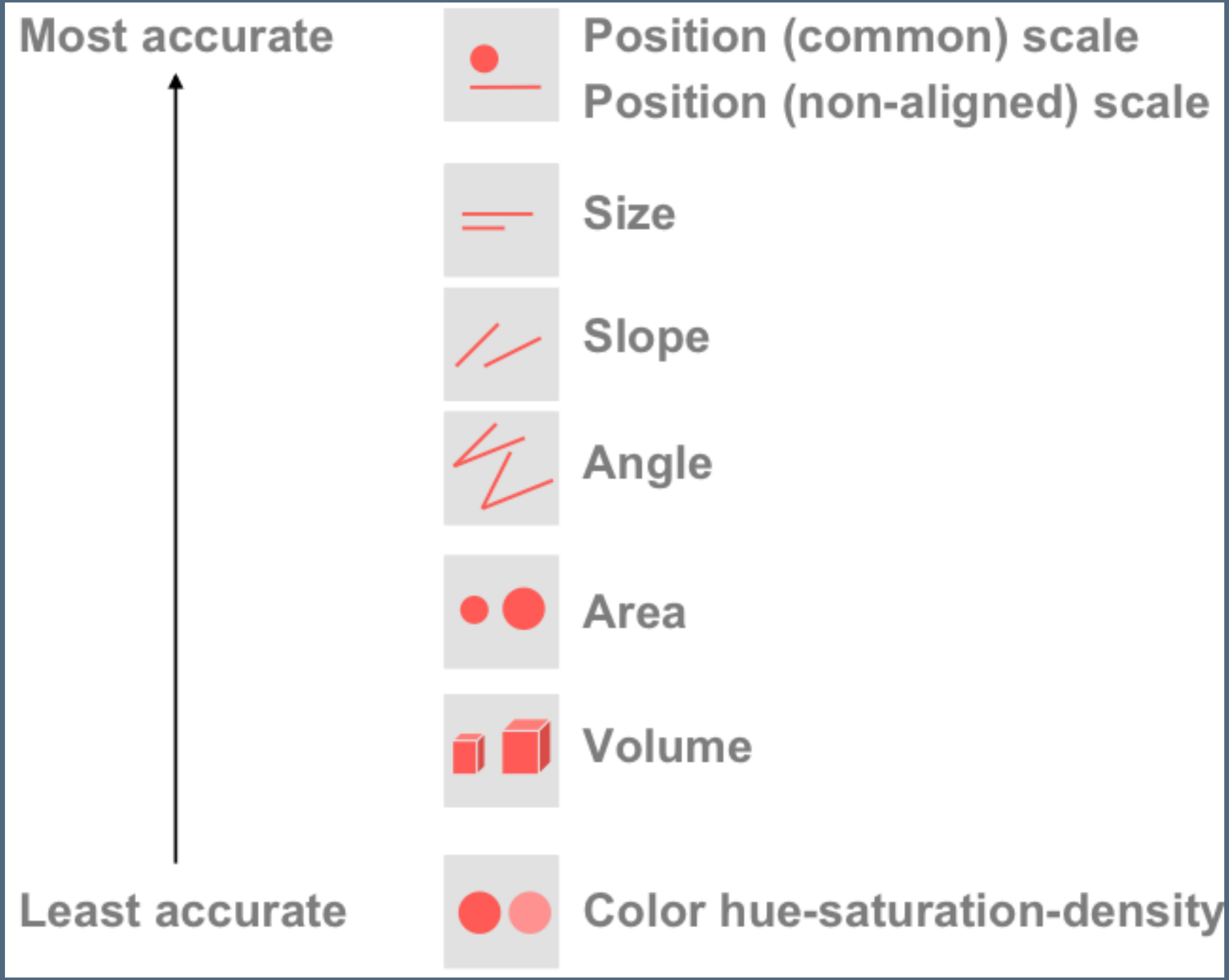


[Cleveland and McGill 84]



[Cleveland and McGill 84]

Relative Magnitude Estimation



Mackinlay's ranking [1986]

Quantitative	Ordinal	Nominal
Position	Position	Position
Length	Density	Color hue
Angle	Color saturation	Texture
Slope	Color hue	Connection
Area	Texture	Containment
Volume	Connection	Density
Density	Containment	Color saturation
Color saturation	Length	Shape
Color hue	Angle	Length
Texture	Slope	Angle
Connection	Area	Slope
	Volume	Area
	Shape	Volume

Automating the design of graphical presentation of relational information
J. Mackinlay, 1986

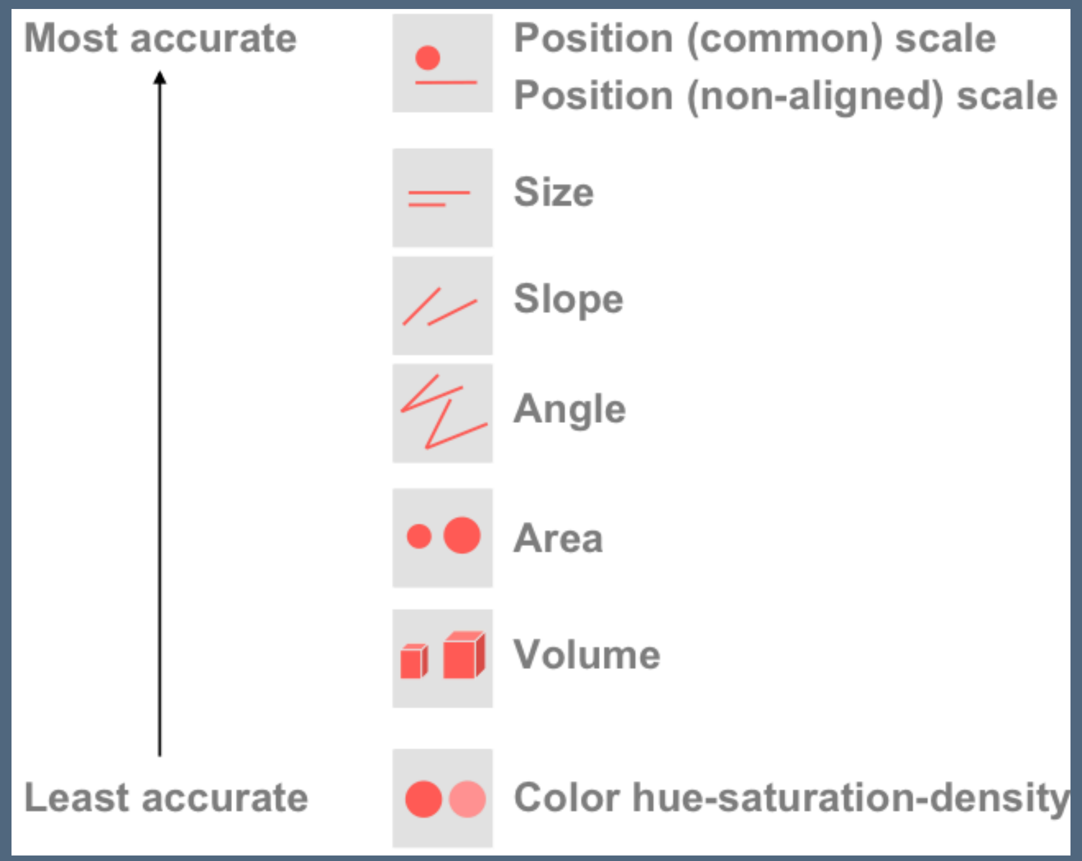
Algorithm for Chart Construction

Encode most important data using highest ranking visual variable for the data type

Year	Exports	Imports
1700	170,000	300,000
1701	171,000	302,000
1702	176,000	303,000
…	…	…

→

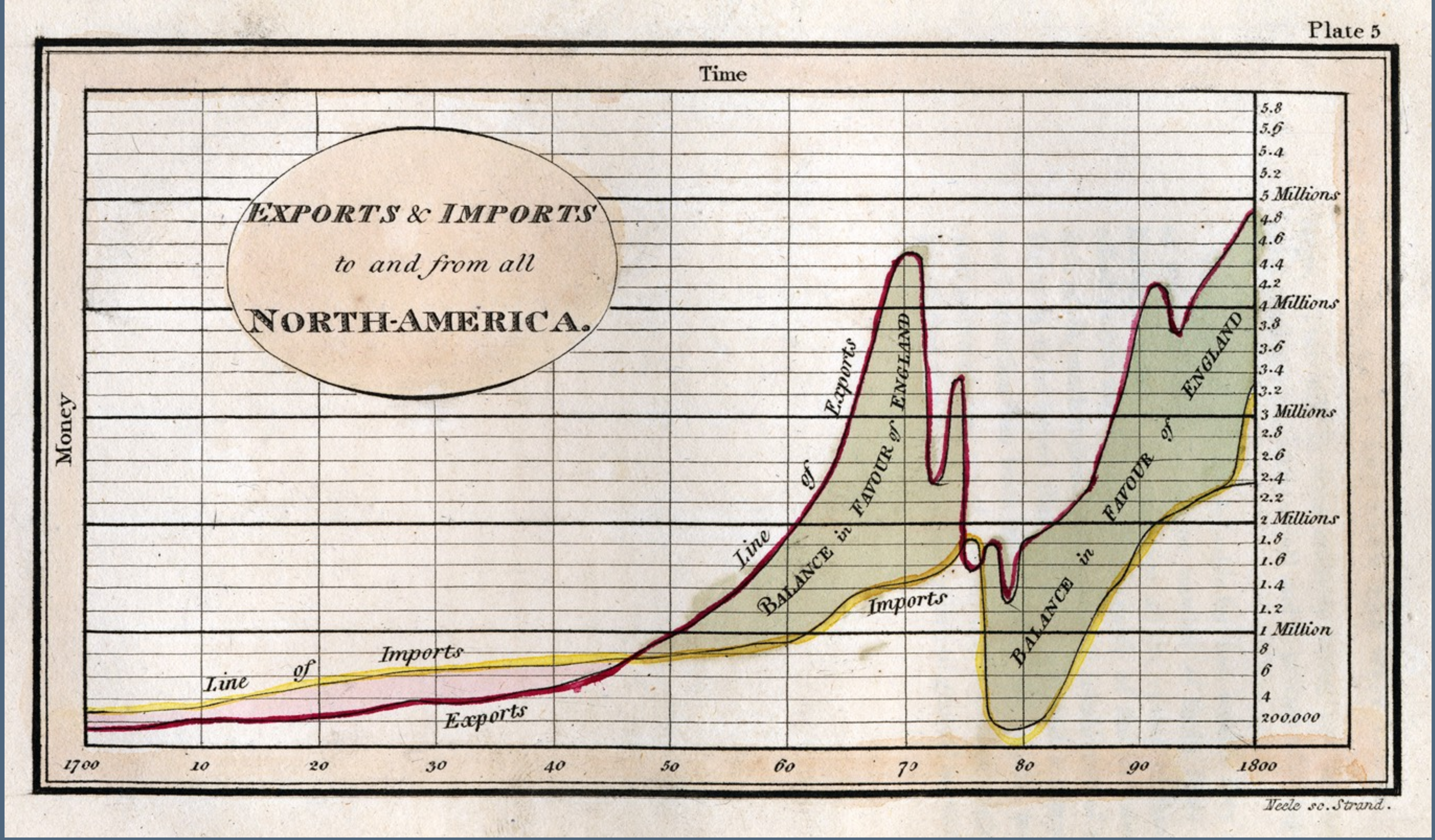
1. Year (Q)
2. Exports (Q)
3. Imports (Q)



→

mark: lines

Year (Q) → x-pos
Exports (Q) → y-pos
Imports (Q) → y-pos



Data

Year	Exports	Imports
1700	170,000	300,000
1701	171,000	302,000
1702	176,000	303,000
1703	180,000	312,000
1704	187,000	319,000
…	…	…

Marks

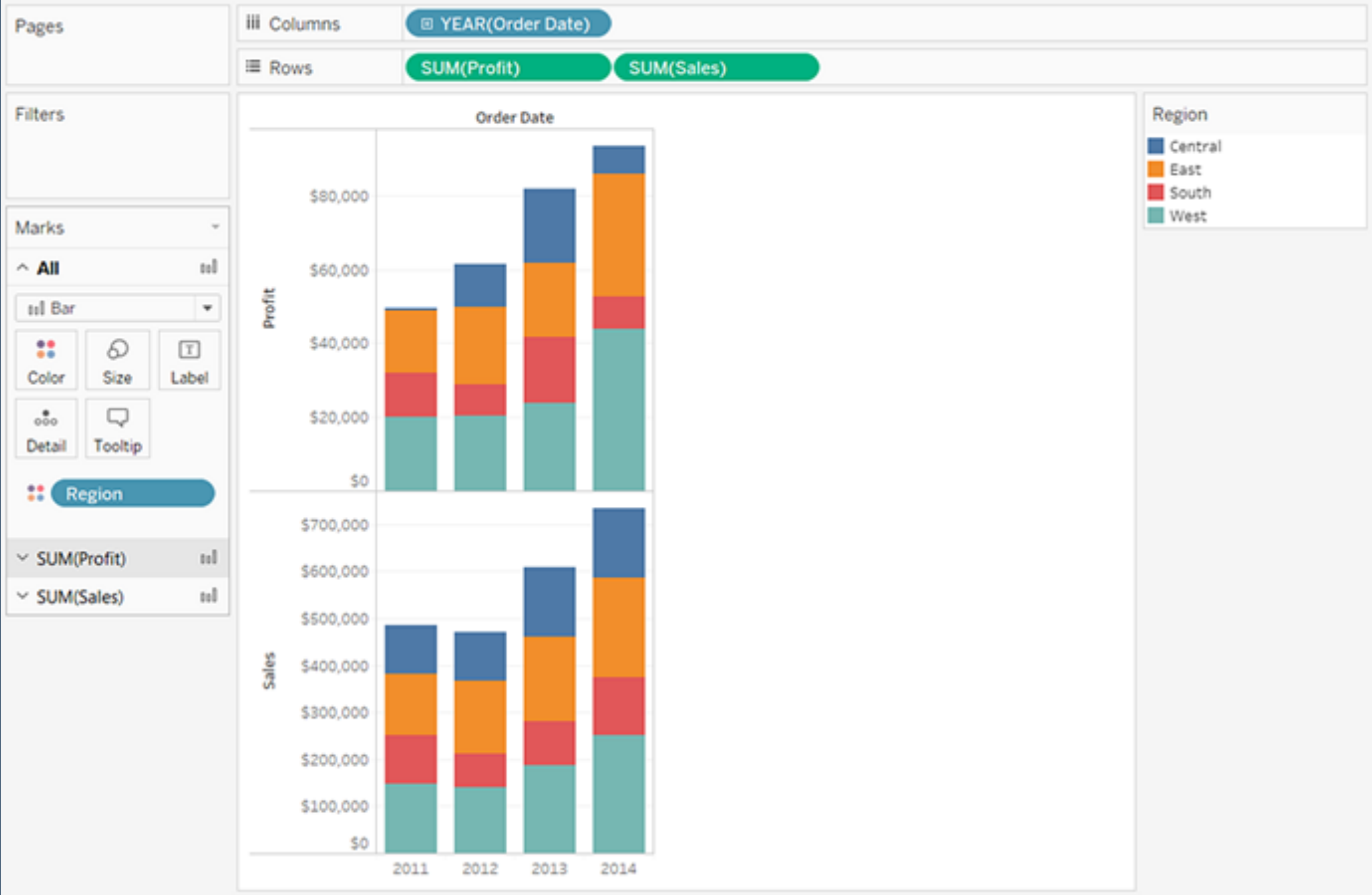
Mappings

mark: lines

Year → x-pos (Q)
Exports → y-pos (Q)
Imports → y-pos (Q)
Exports → color (N)
Imports → color (N)

Impact

Mackinlay's approach gets extended by Chris Stolte and Pat Hanrahan into VizQL, which then becomes…



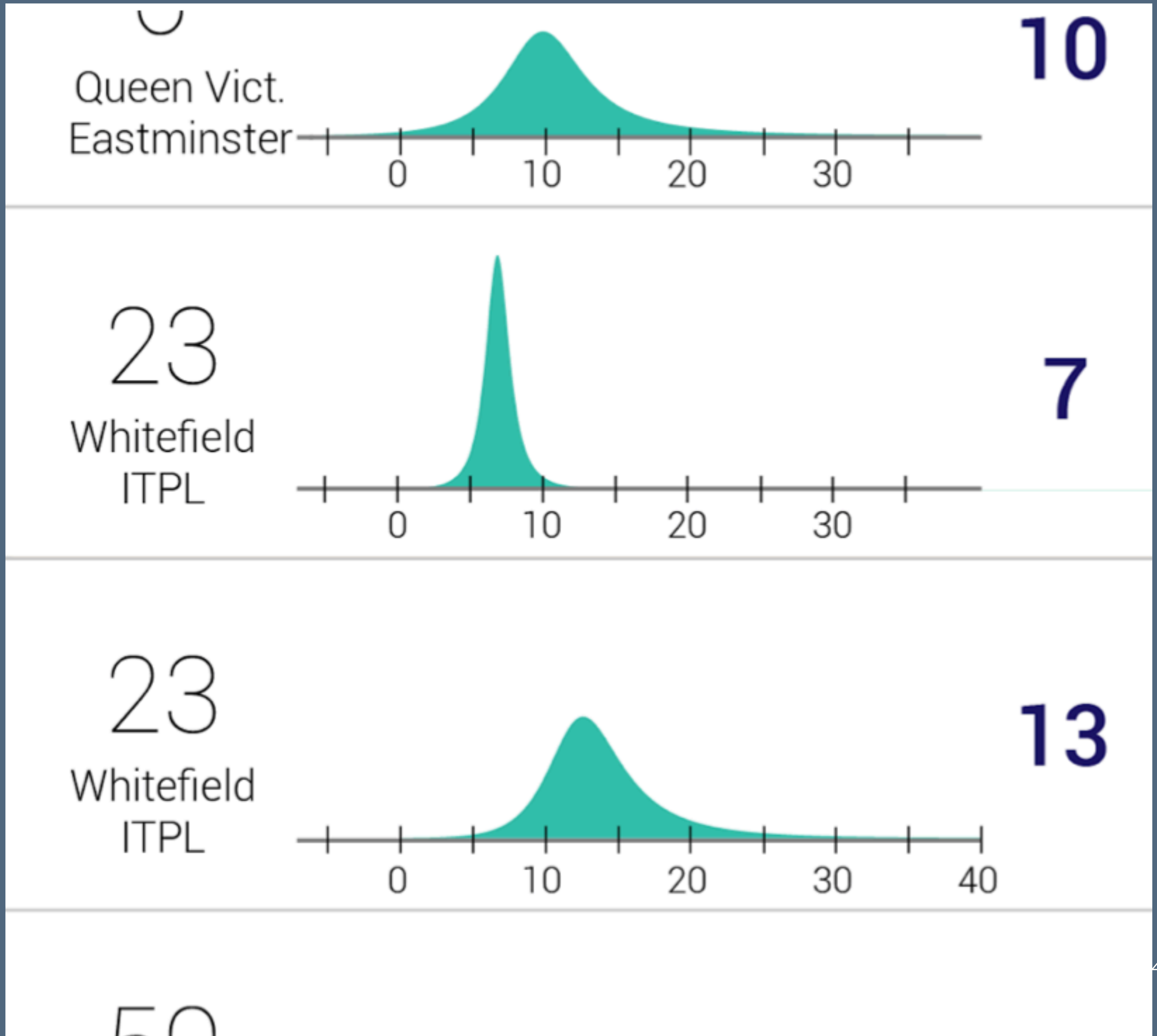
tableau® SOFTWARE

Frontiers of Visualization Research

Conveying uncertainty

[Kay et al. 2016]

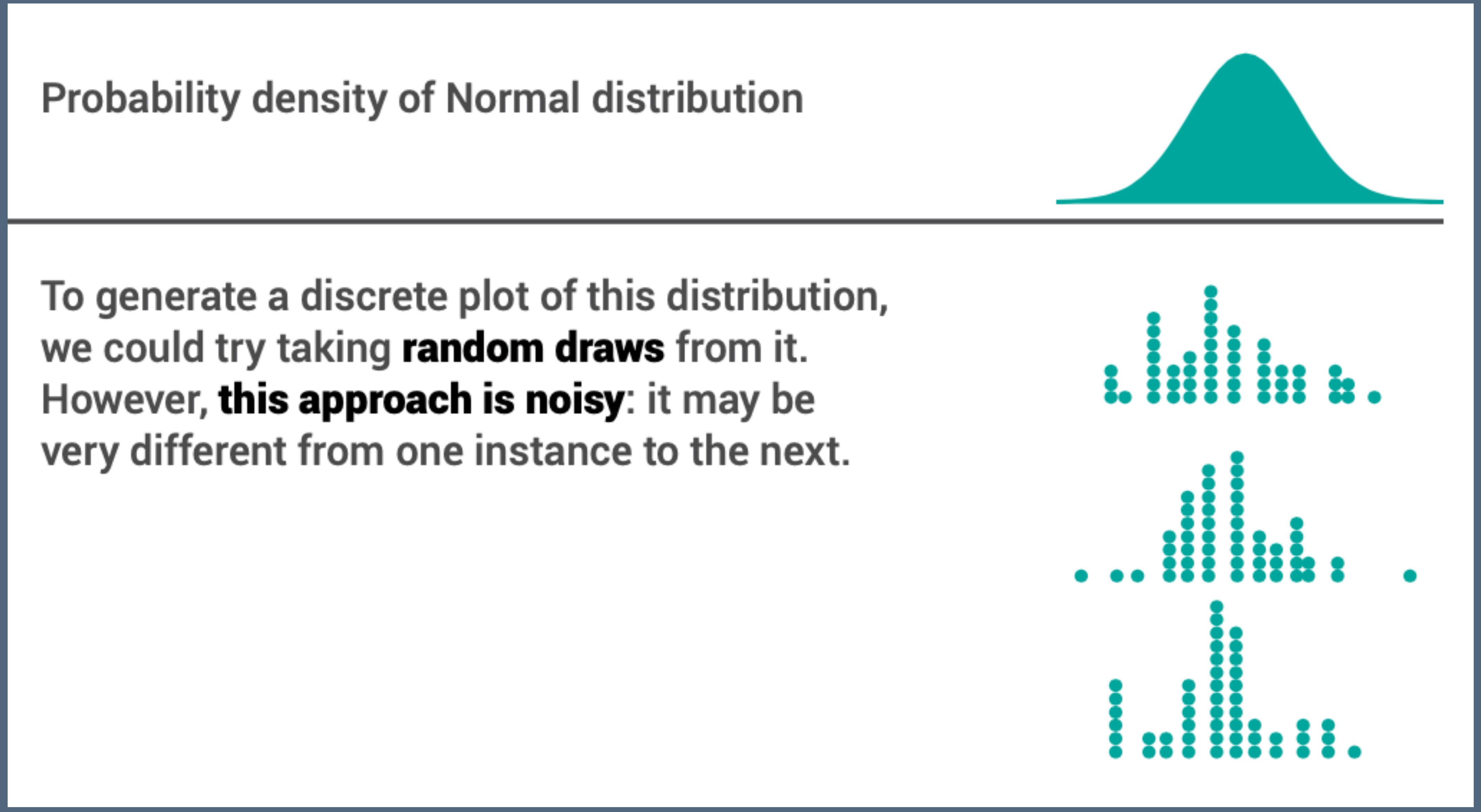
We over-rely on point estimates. People simplify distributions and attend to point estimates on the right (10min until the bus comes)

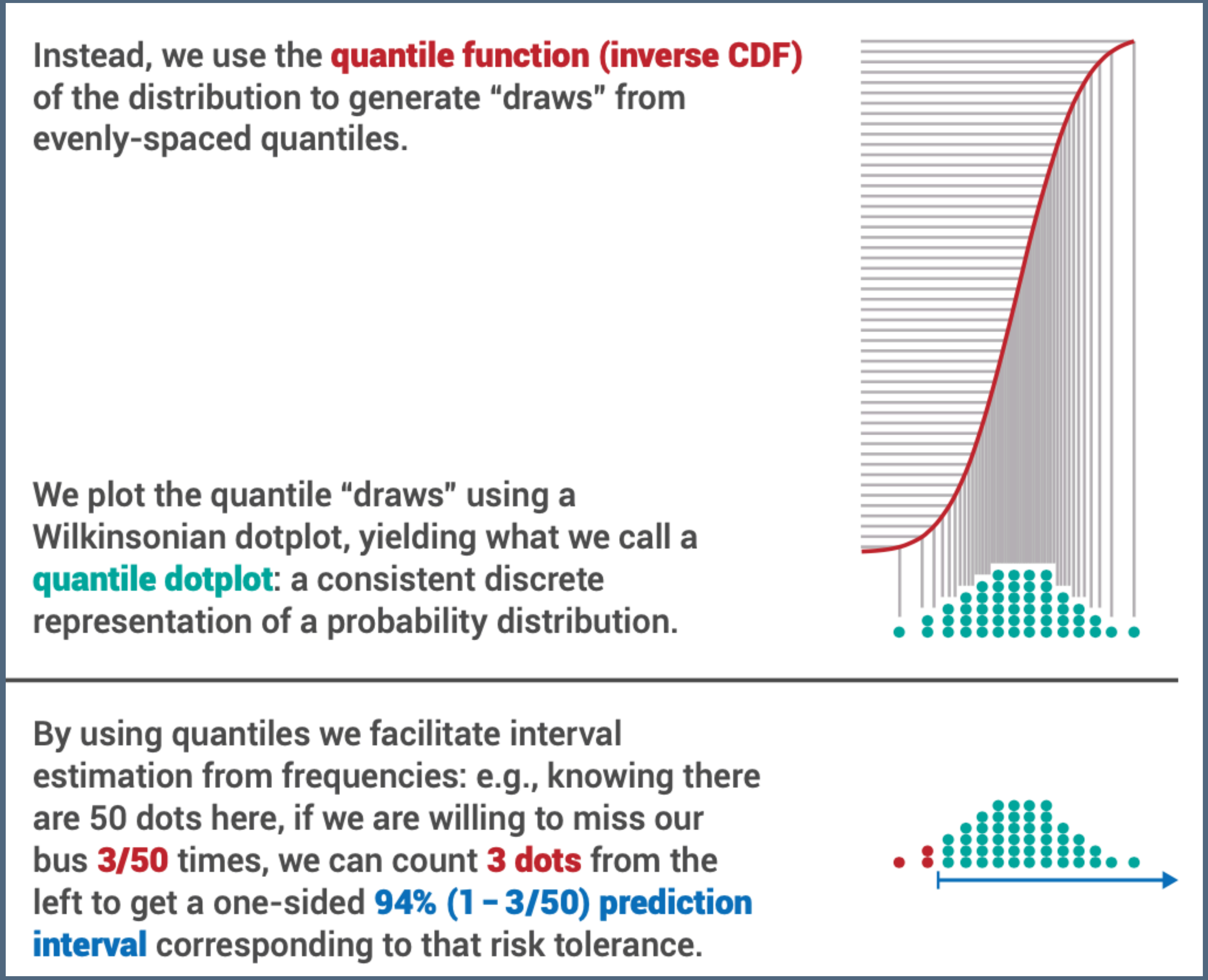


Conveying uncertainty

[Kay et al. 2016]

Suggestion: quantile dot plots

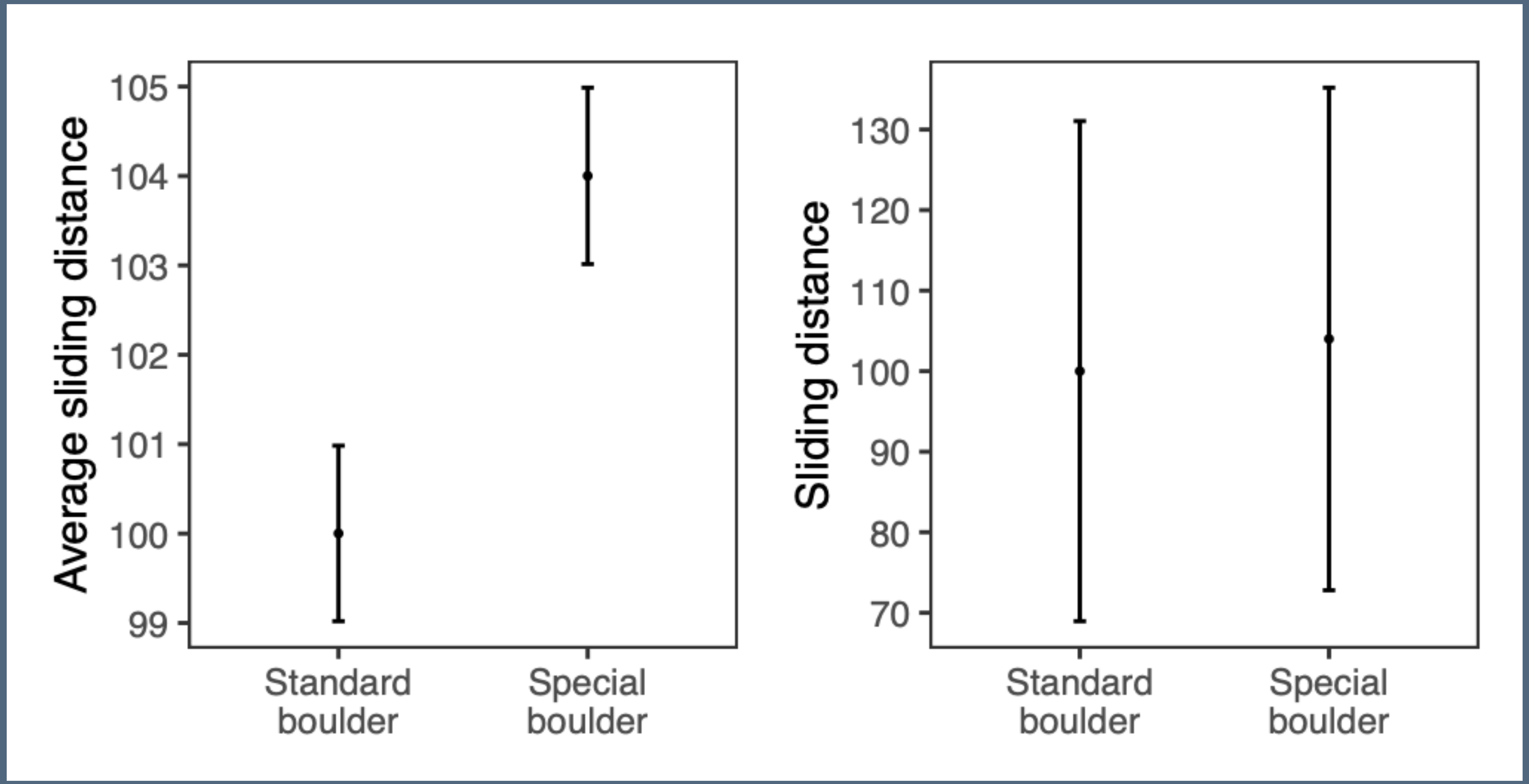




Interpretation errors

[Hofman, Goldstein, and Hullman 2020]

Two common visualizations of uncertainty:



Std. Error: uncertainty in the population mean

Std. Deviation: uncertainty in a single sample

Experiment: people overestimate treatment effects when shown standard errors instead of standard deviation

Exploratory analysis

[Wongsuphasawat et al. 2015]

User inputs dataset and variables of interest, and recommender automatically generates visualizations of relevant other variables

Intentionally difficult?

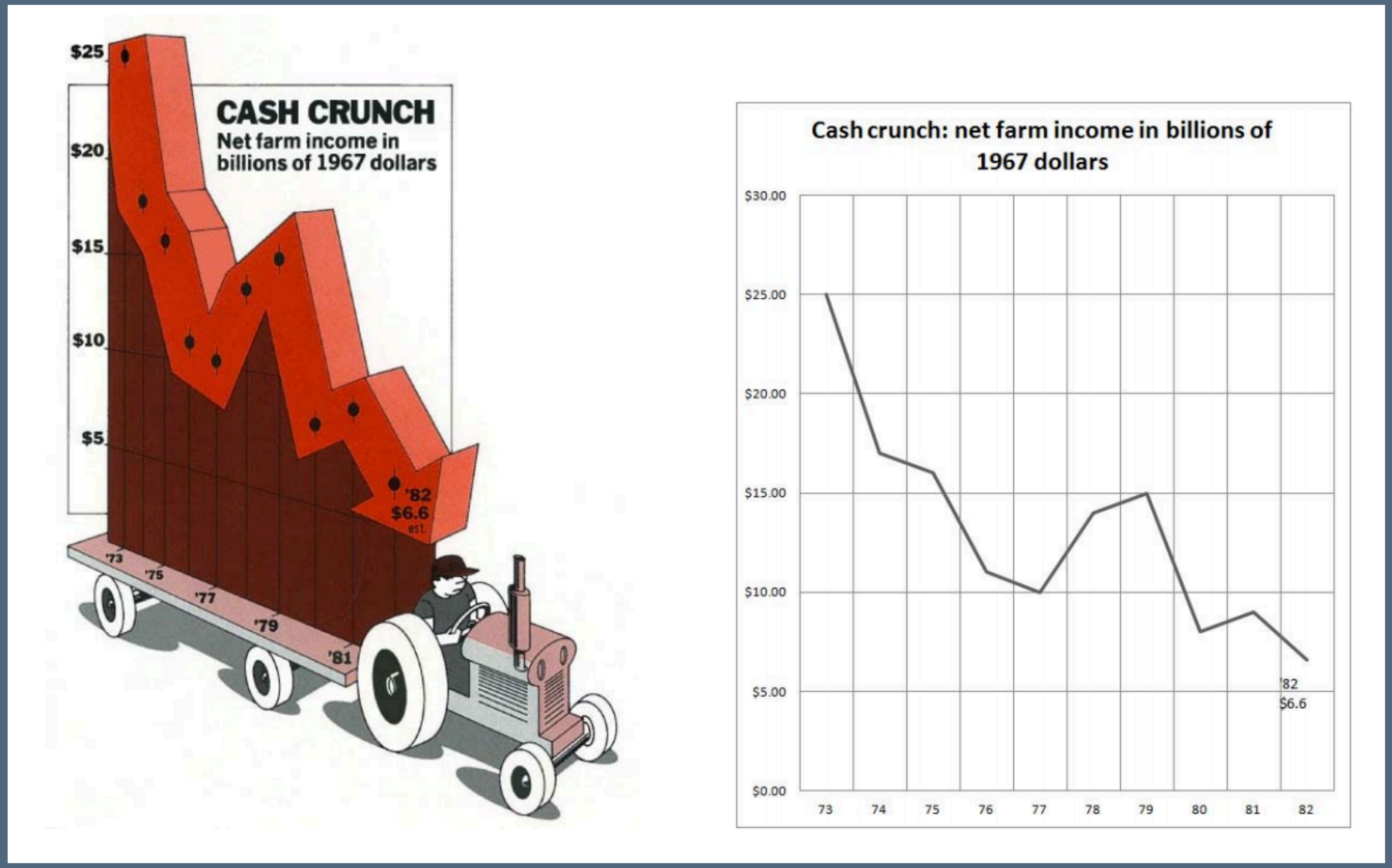
[Hullman and Adar 2011]

Generally, visualization (and HCI more broadly) argue optimizing for clear and correct interpretation

Yet difficult visualizations may support better comprehension and recall

Why? It induces active processing:

- Forcing active construction of meaning
- Disfluent learning experiences avoid heuristics and superficial reasoning



Difficult with chartjunk — Easy

Towards Understanding How Readers Integrate Charts and Captions: A Case Study with Line Charts [Kim et al. 2021]

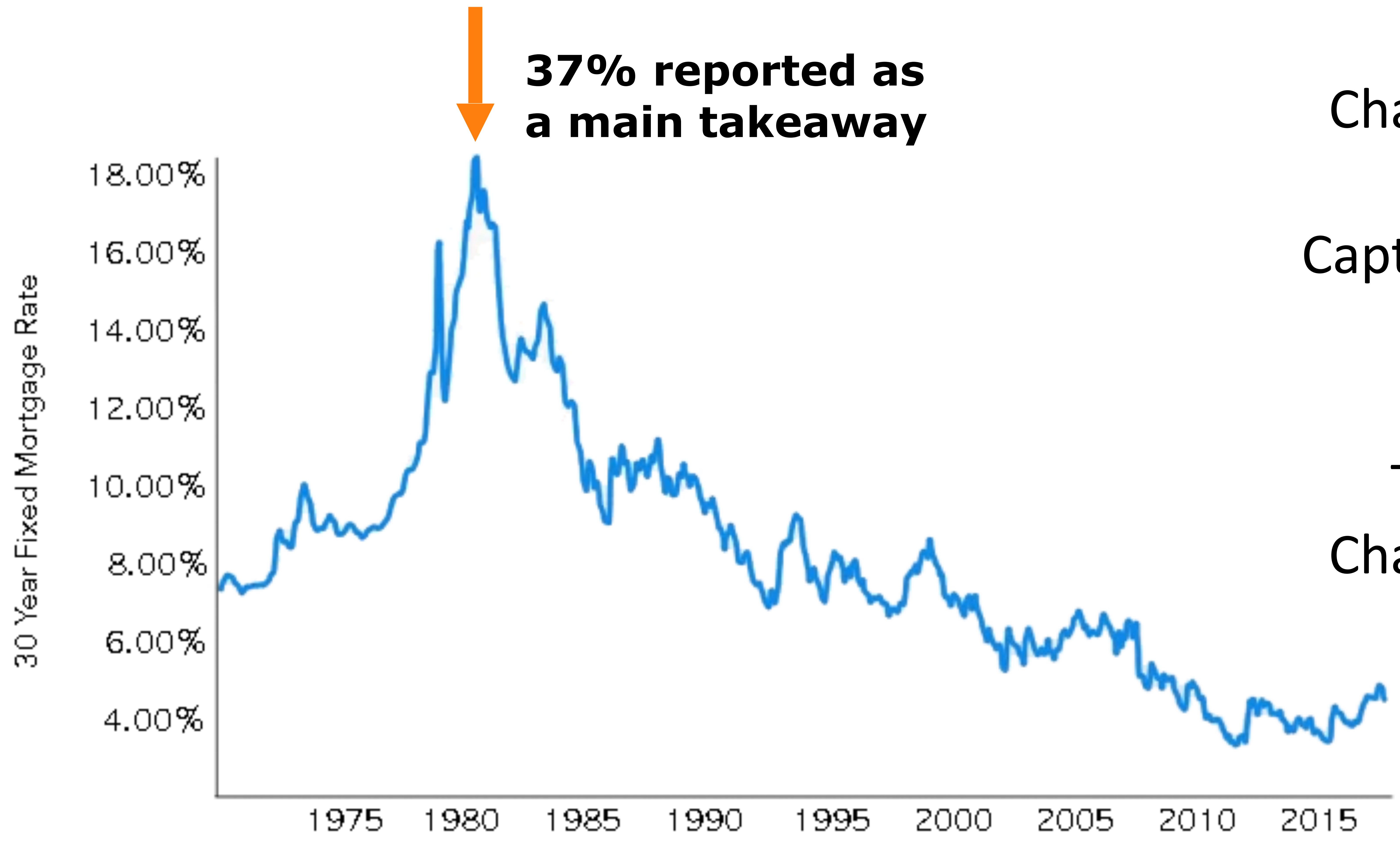
Towards Understanding How Readers Integrate Charts and Captions: A Case Study with Line Charts [Kim et al. 2021]

The 30-year fixed mortgage rate increased slightly from 1997 to 1999.

Towards Understanding How Readers Integrate Charts and Captions: A Case Study with Line Charts [Kim et al. 2021]

The 30-year fixed mortgage rate increased slightly from 1997 to 1999.

Towards Understanding How Readers Integrate Charts and Captions: A Case Study with Line Charts [Kim et al. 2021]



The 30-year fixed mortgage rate increased slightly from 1997 to 1999.

Chart Emphasis
≠
Caption Emphasis
↓
Takeaway:
Chart Emphasis

Towards Understanding How Readers Integrate Charts and Captions: A Case Study with Line Charts [Kim et al. 2021]

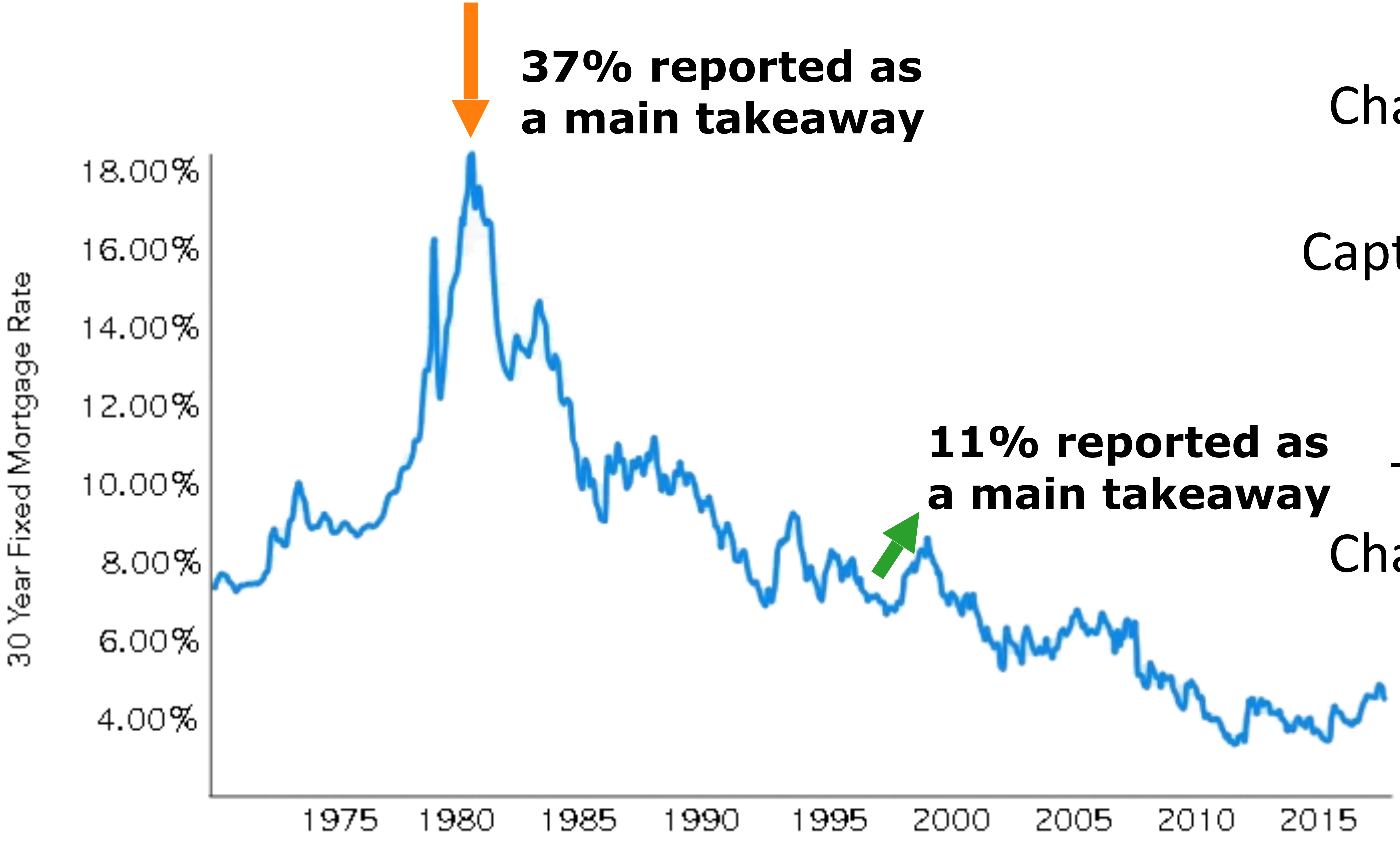


Chart Emphasis
≠
Caption Emphasis
↓
Takeaway:
Chart Emphasis

The 30-year fixed mortgage rate increased slightly from 1997 to 1999.

Towards Understanding How Readers Integrate Charts and Captions: A Case Study with Line Charts [Kim et al. 2021]

Chart Emphasis
=
Caption Emphasis
↓
Takeaway:
Chart Emphasis

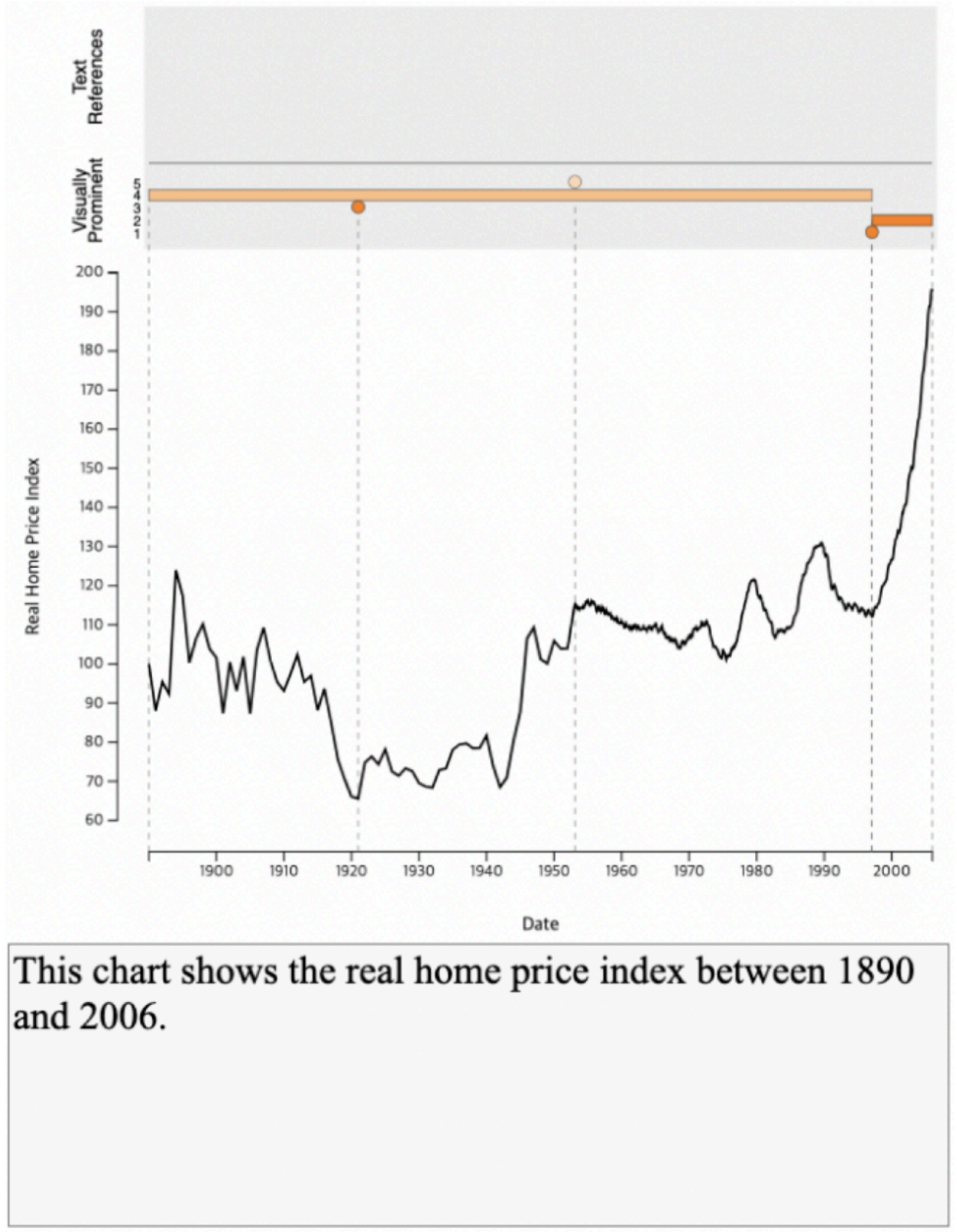
The 30-year fixed mortgage rate reached its peak of 18.5% in 1981.

Towards Understanding How Readers Integrate Charts and Captions: A Case Study with Line Charts [Kim et al. 2021]

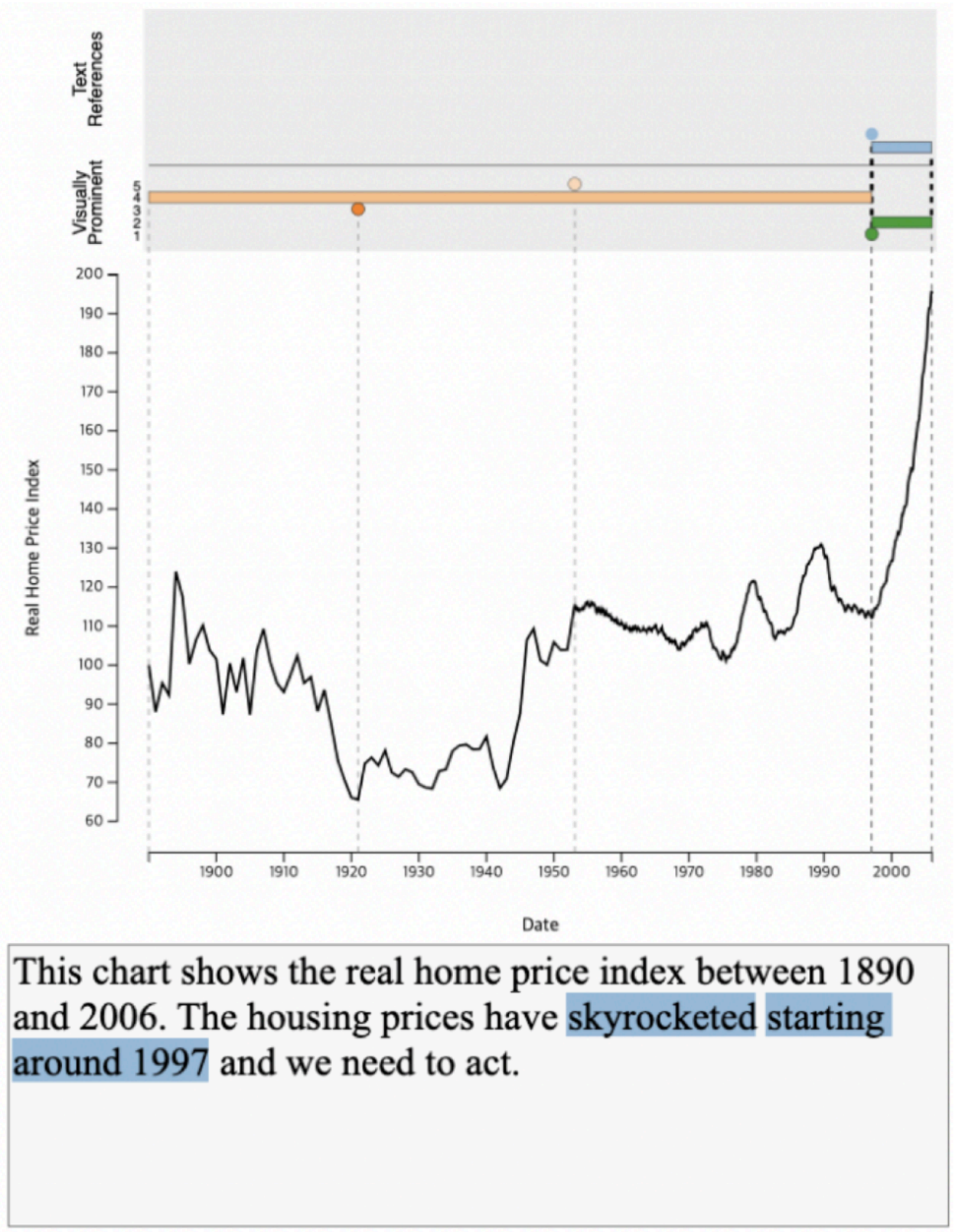
Reading Charts and Captions

[Kim et al. 2021]

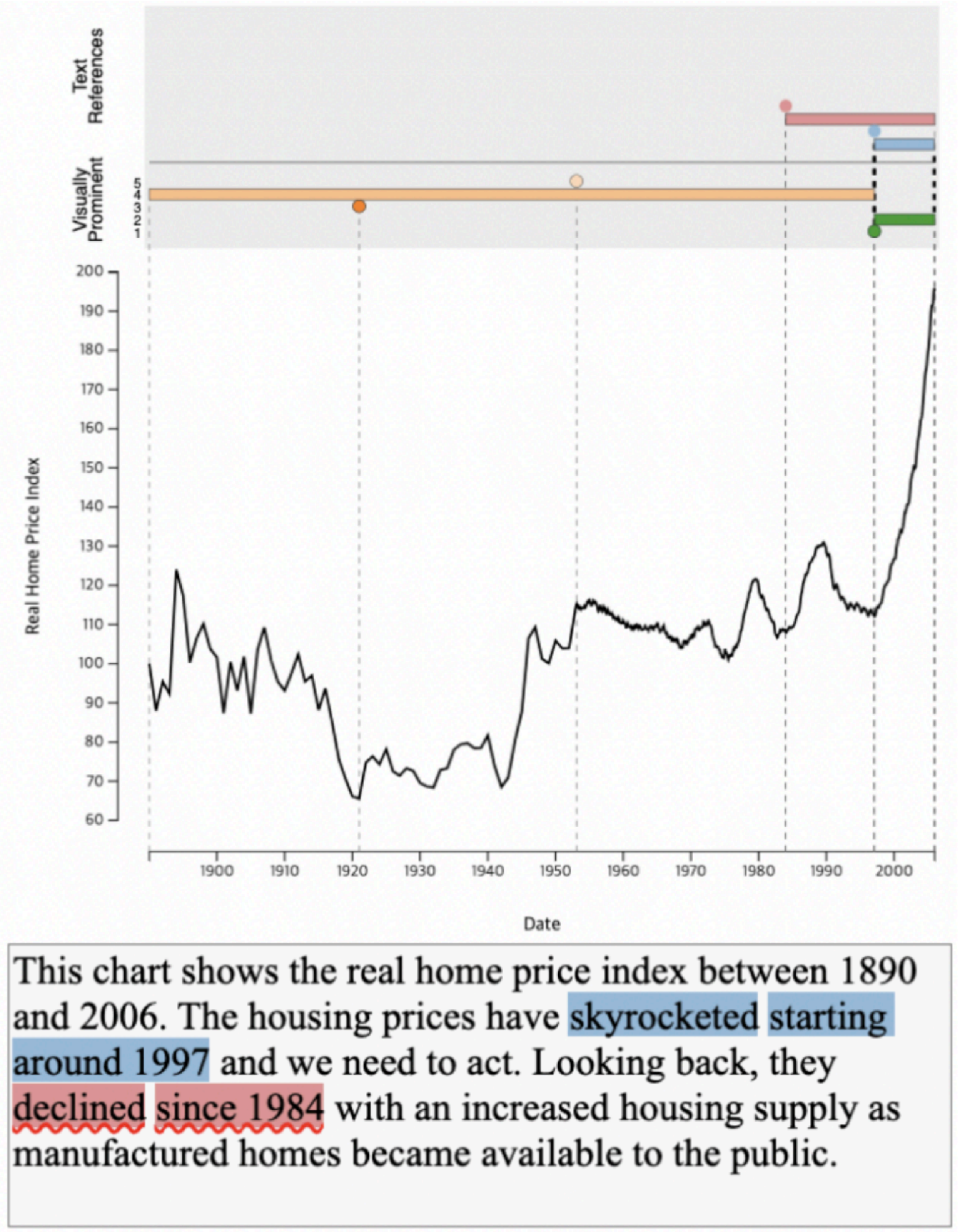
When text and visualization emphasis **mismatch**, readers **rely more on the chart** and can miss information in the caption.



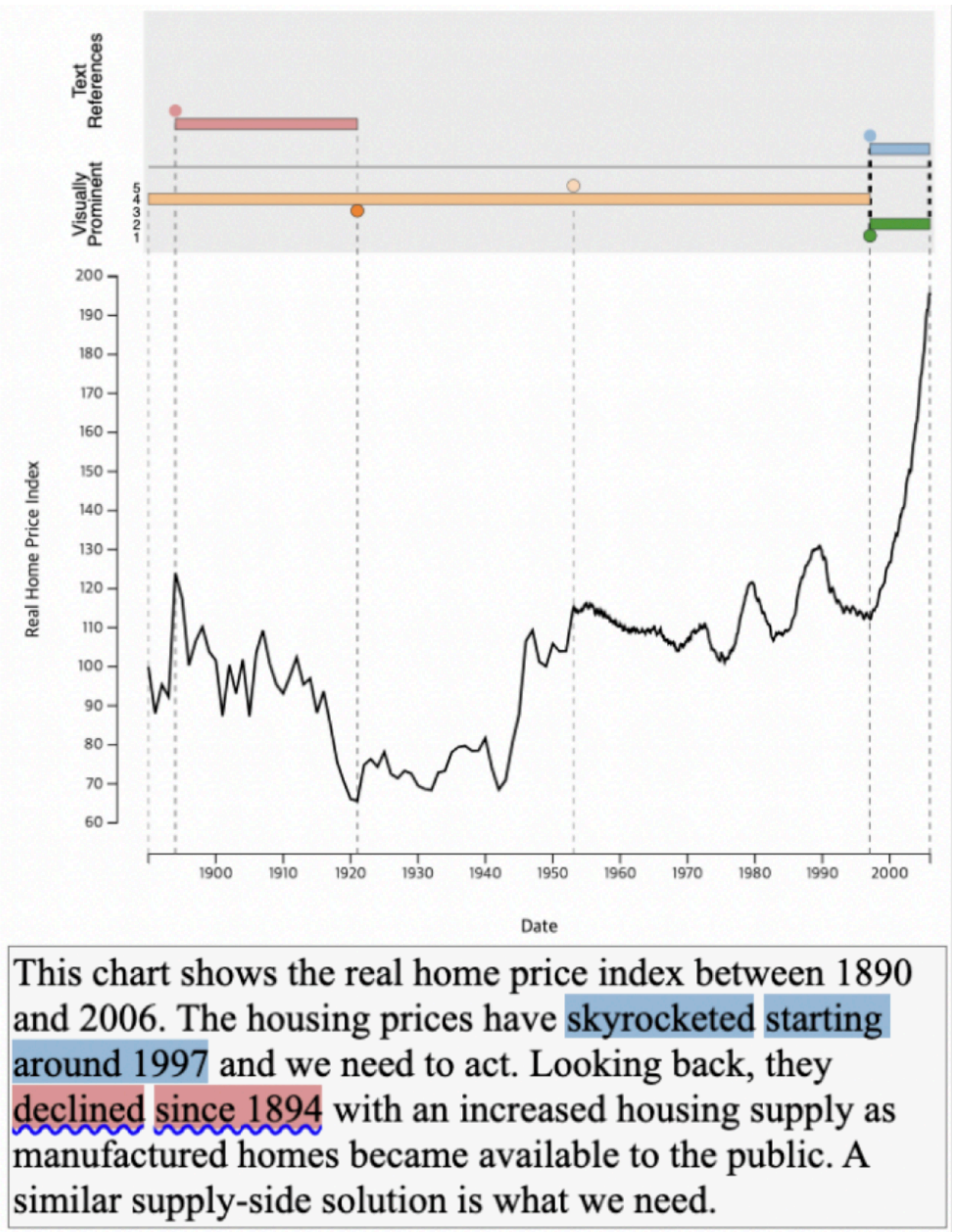
(a) Prominent features & Basic caption



(b) Caption text about prominent feature



(c) Caption including false information



(d) Caption about less prominent feature

EmphasisChecker: A Tool for Guiding Chart and Caption Emphasis [Kim et al. 2023]

Summary

Visualizations can be represented as **encodings** that map from **data to marks & visual attributes** based on **data types**

Our **cognitive and perceptual systems determine which encodings are effective**: we (mis)read data if encoded poorly

Active research at frontiers investigating **how users can create effective visualizations** and **how readers take information away from them**

To learn more about visualization consider taking **CS 448B: Fall 2025**

- An understanding of key visualization techniques and theory, including data models, graphical perception and methods for visual encoding and interaction.
- Exposure to a number of common data domains and corresponding analysis tasks, including exploratory data analysis and network analysis.
- Practical experience building and evaluating visualization systems using Vega-Lite and D3.js.
- The ability to read and discuss research papers from the visualization literature.

CS 448B Visualization

Stanford CS course on data visualization techniques (Fall 2021)

Location: Huang Eng. 18
Time: MW 11:30am-1pm

ABOUT

LEARNING GOALS

TEXTBOOKS/RESOURCES

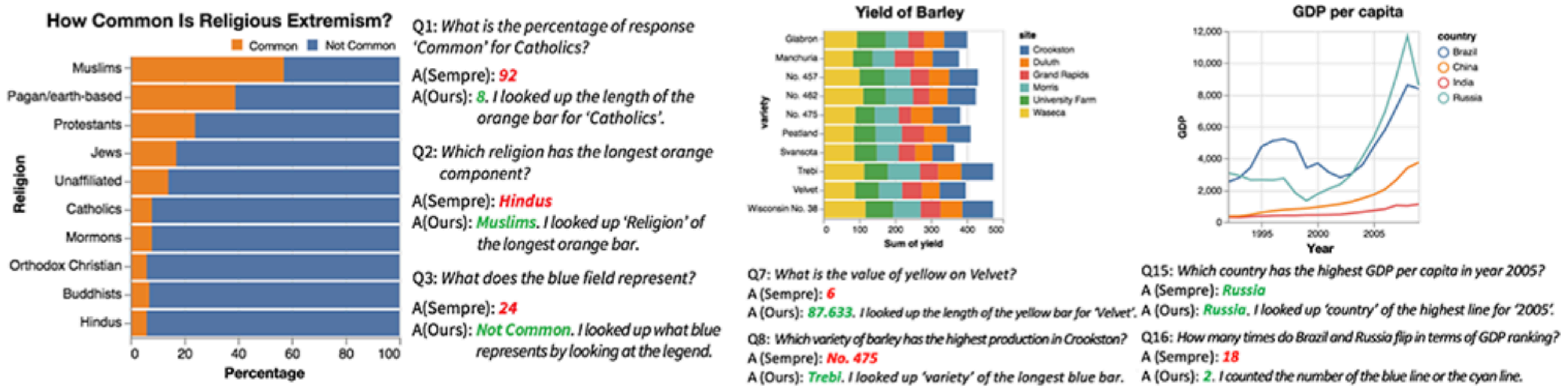
SCHEDULE

- Week 1
- Week 2
- Week 3
- Week 4
- Week 5
- Week 6
- Week 7
- Week 8
- Week 9
- Week 10

TEACHING STAFF

ASSIGNMENTS

- Class Participation
- Assignment 1
- Assignment 2
- Assignment 3
- Final Project



Well designed visualizations capitalize on human facilities for processing visual information and thereby improve comprehension, memory, inference, and decision making. In this course we will study techniques and algorithms for creating effective visualizations based on principles from graphic design, visual art, perceptual psychology and cognitive science. The course is targeted both towards students interested in using visualization in their own work, as well as students interested in building better visualization tools and systems.

There are no official prerequisites for the class, but familiarity with the material in CS147, CS148 and CS142 is especially useful. Most important is a basic working knowledge of, or willingness to learn, web-programming, especially JavaScript, Vega-Lite and D3.js. While we will cover a little bit of Vega-Lite and D3.js in class, we will also expect students learn some introductory material, especially about Javscript on their own, as necessary. Tutorials on Javascript are available on the web and we will help you find the relevant information as you need it.

*Contact us via Slack if you are worried about whether you have the background for the course.

Learning Goals

The goals of this course are to provide students with the foundations necessary for understanding and extending the current state of the art in visualization. By the end of the course, students will have:

- An understanding of key visualization techniques and theory, including data models, graphical perception and methods for visual encoding and interaction.
- Exposure to a number of common data domains and corresponding analysis tasks, including exploratory data analysis and network analysis.
- Practical experience building and evaluating visualization systems using Vega-Lite and D3.js.
- The ability to read and discuss research papers from the visualization literature.

Textbooks/Resources

1. The Visual Display of Quantitative Information (2nd Edition). E. Tufte. Graphics Press.
2. Envisioning Information. E. Tufte. Graphics Press.
3. ***Optional Textbook.*** Visualization Analysis and Design. Tamara Munzner. A K Peters Visualization Series. CRC Press.
4. ***Optional Reference.*** Interactive Data Visualization for the Web (2nd Edition). Scott Murray. O'Reilly Press. [Read Online] [Code Examples on Github]

Your best bet is to order them online. **Please order soon. Readings will be assigned in the first week of class.**

References

Ahlberg, Christopher, and Ben Shneiderman. "Visual information seeking: Tight coupling of dynamic query filters with starfield displays." Proceedings of the SIGCHI conference on Human factors in computing systems. 1994.

Anscombe, Francis J. "Graphs in statistical analysis." The american statistician 27.1 (1973): 17-21.

Bertin, Jacques. Semiology of graphics. 1967.

Buja, Andreas, et al. "Interactive data visualization using focusing and linking." Proceedings of the 2nd conference on Visualization'91. 1991.

Card, Stuart K., Jock D. Mackinlay, and Ben Shneiderman, eds. "Readings in information visualization: using vision to think." (1999).

Cleveland, William S., and Robert McGill. "Graphical perception: Theory, experimentation, and application to the development of graphical methods." Journal of the American statistical association 79.387 (1984): 531-554.

Heer, Jeffrey, Michael Bostock, and Vadim Ogievetsky. "A tour through the visualization zoo." Communications of the ACM 53.6 (2010): 59-67.

Heer, Jeffrey. Data Types, Graphical Marks, and Visual Encoding Channels. 2021. https://observablehq.com/@uwdata/data-types-graphical-marks-and-visual-encoding-channels

Hofman, Jake M., Daniel G. Goldstein, and Jessica Hullman. "How visualizing inferential uncertainty can mislead readers about treatment effects in scientific results." Proceedings of the 2020 chi conference on human factors in computing systems. 2020.

Hullman, Jessica, Eytan Adar, and Priti Shah. "Benefitting infovis with visual difficulties." IEEE Transactions on Visualization and Computer Graphics 17.12 (2011): 2213-2222.

References

Kay, Matthew, et al. "When (ish) is my bus? user-centered visualizations of uncertainty in everyday, mobile predictive systems." Proceedings of the 2016 chi conference on human factors in computing systems. 2016.

Mackinlay, Jock. "Automating the design of graphical presentations of relational information." Acm Transactions On Graphics (Tog) 5.2 (1986): 110-141.

McCormick, Bruce H. "Visualization in scientific computing." Acm Sigbio Newsletter 10.1 (1988): 15-21.

Minard, Charles Joseph. Carte figurative des pertes successives en hommes de l'Armée française dans la campagne de Russie 1812-1813. Graphics Press., 1869.

Moritz, Dominik, et al. "Formalizing visualization design knowledge as constraints: Actionable and extensible models in draco." IEEE transactions on visualization and computer graphics 25.1 (2018): 438-448.

Li, Huiyang, and Nadine Moacdieh. "Is "chart junk" useful? An extended examination of visual embellishment." Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Vol. 58. No. 1. Sage CA: Los Angeles, CA: Sage Publications, 2014.

Playfair, William, and James Corry. The Commercial and Political Atlas: Representing, by Means of Stained Copper-plate Charts, the Exports, Imports, and General Trade of England; the National Debt, And Other Public Accounts; With Observations And Remarks: by William Playfair, (author Of Regulations For The Interest Of Money.) To which are Added, Charts of the Revenue and Debts of Ireland, Done In The Same Manner, by James Corry, Esq. The Commercial Part is Taken from the Custom-House Books, and the Public Accounts from the Journals of the .... J. Debrett, Piccadilly; GG and J. Robinson, Pater-Noster Row; J. Sewell, Cornhill; the engraver, SJ Neele, NO. 352, Strand; W. Creech and C. Elliot, Edinburgh; and L. White, Dublin, 1786.

References

Satyanarayan, Arvind, et al. "Vega-lite: A grammar of interactive graphics." IEEE transactions on visualization and computer graphics 23.1 (2016): 341-350.

Stevens, Stanley Smith. "On the theory of scales of measurement." Science 103.2684 (1946): 677-680.