

Data Visualization & Design

Week 4

1. Fundamentals of **Graphical Perception**
2. **Data Abstraction**
3. Studio (Introduction to **Tableau**)

1. Fundamentals of **Graphical Perception**
2. **Data Abstraction**
3. Studio (Introduction to **Tableau**)

Question:
Is data visualization a **science** or a
language?

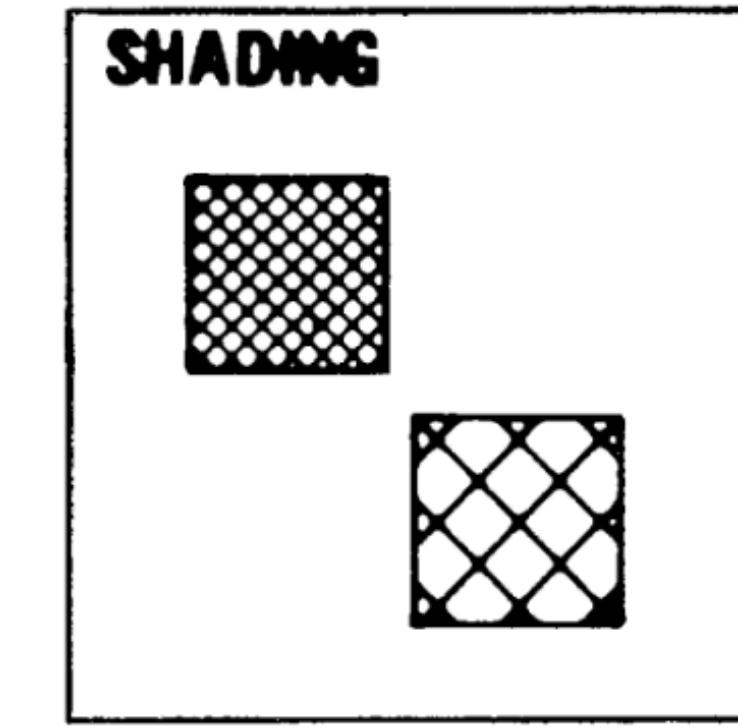
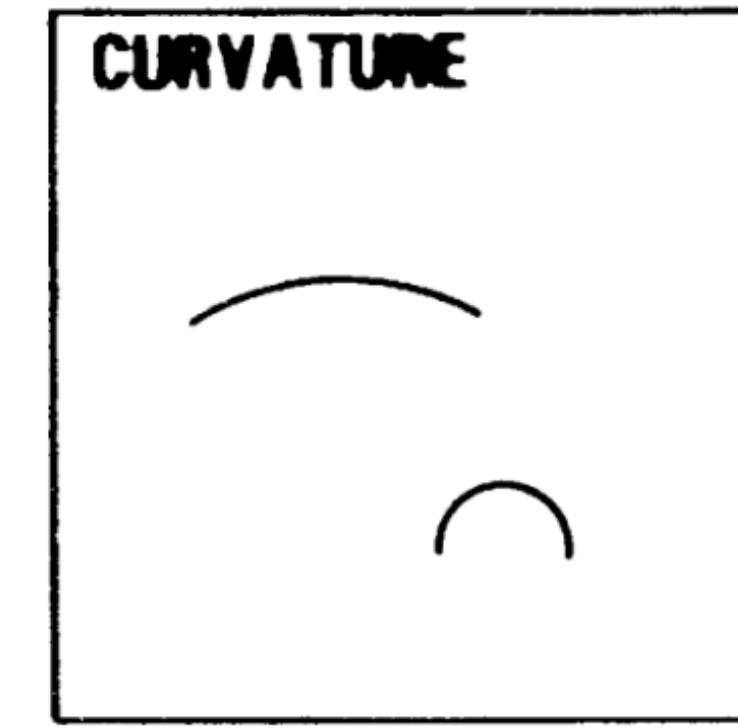
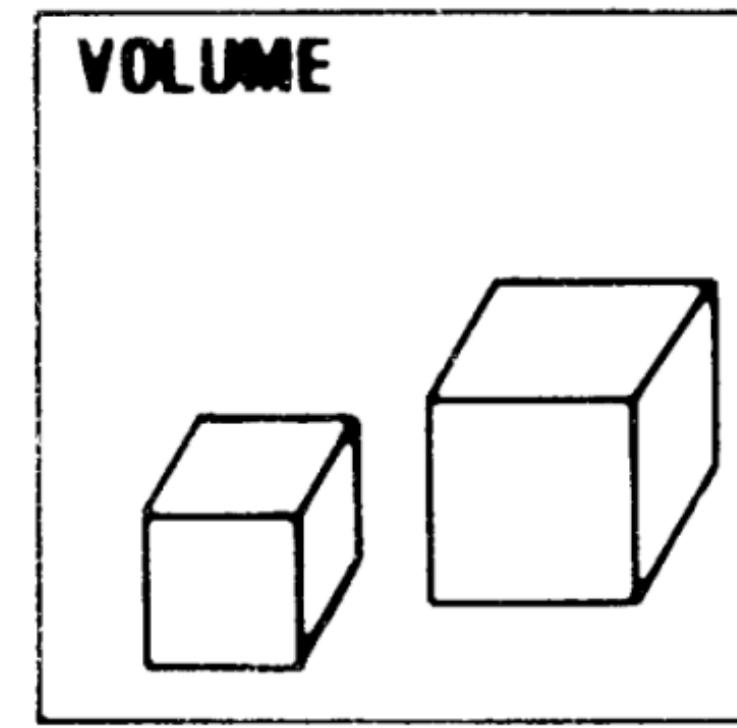
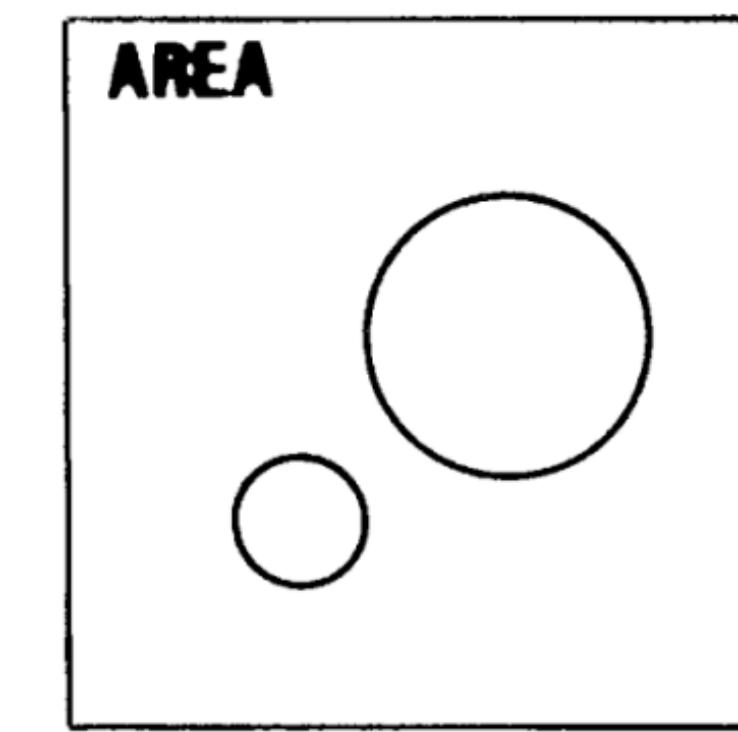
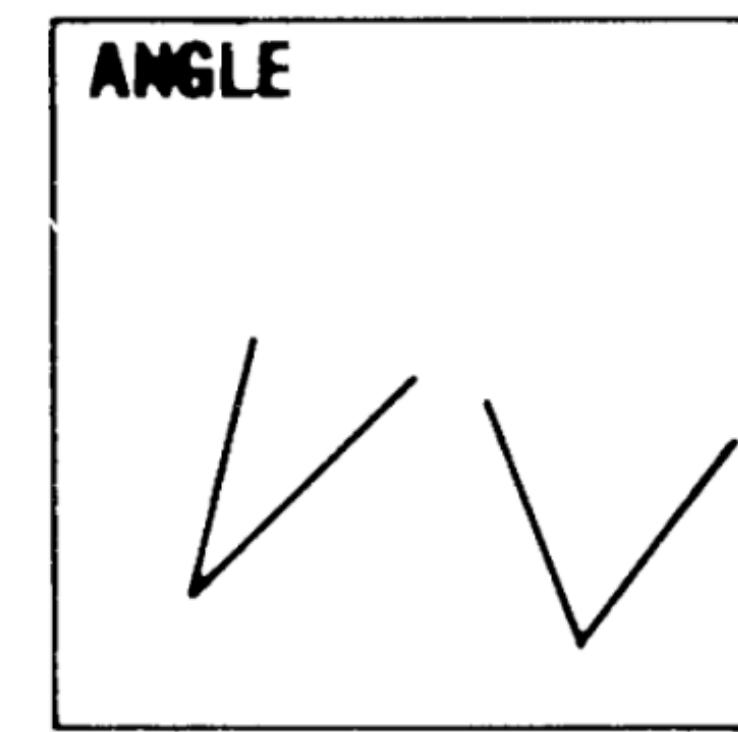
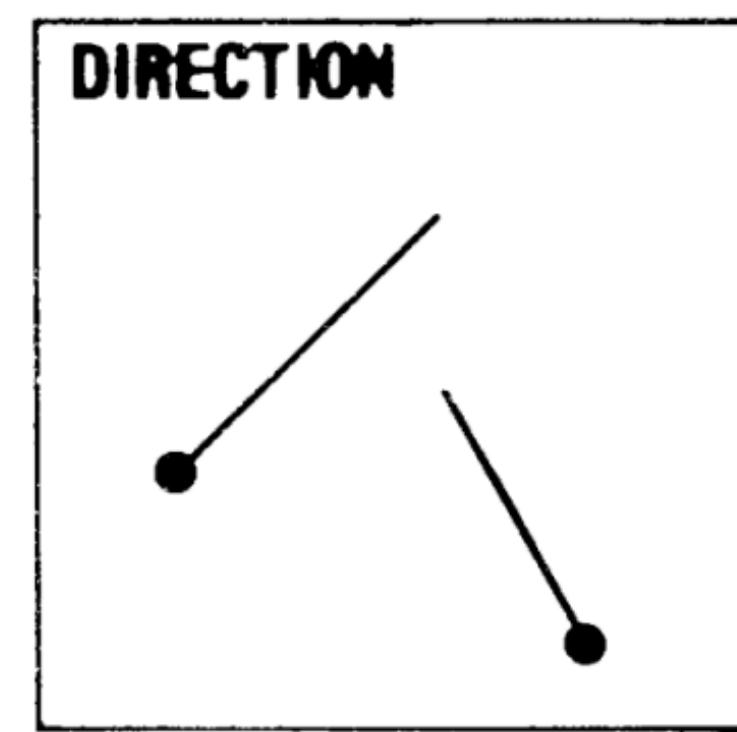
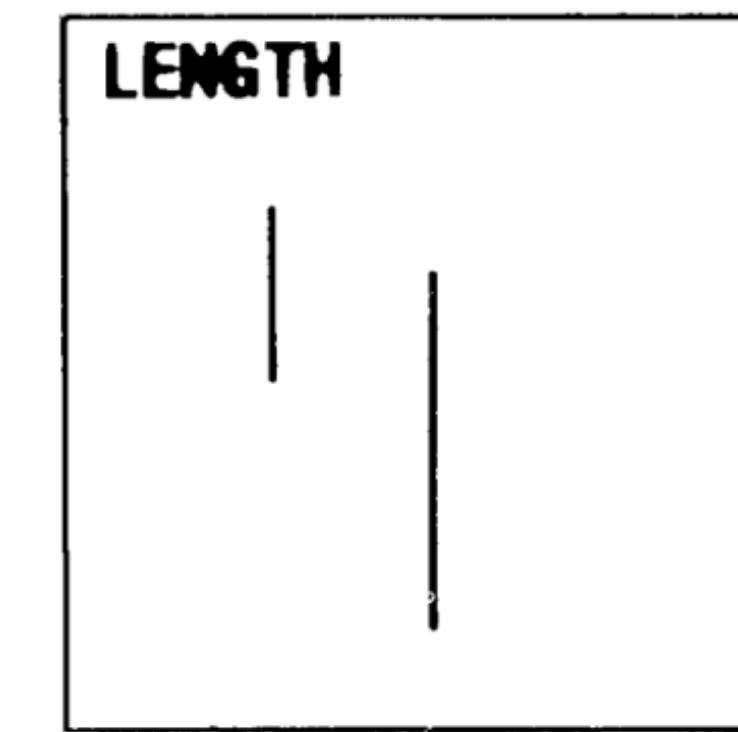
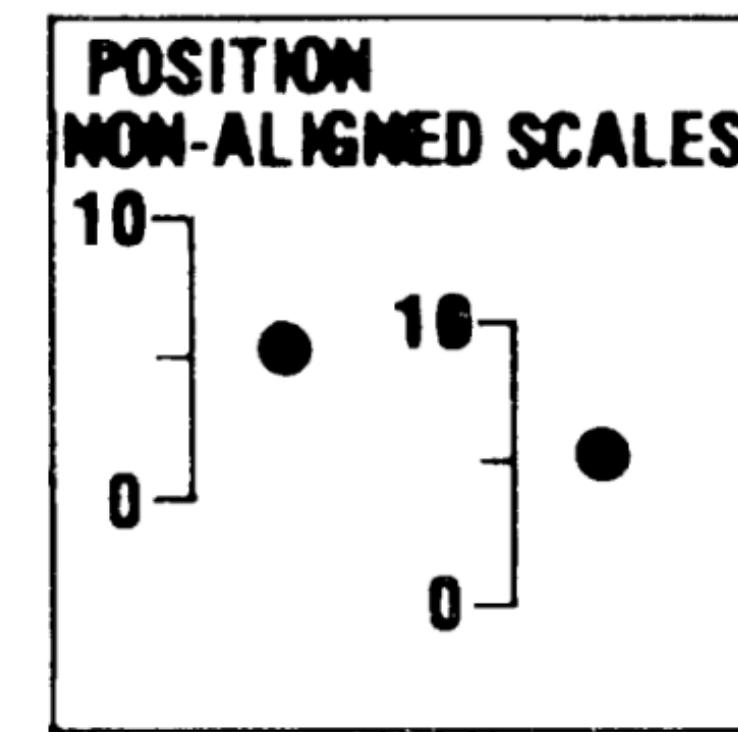
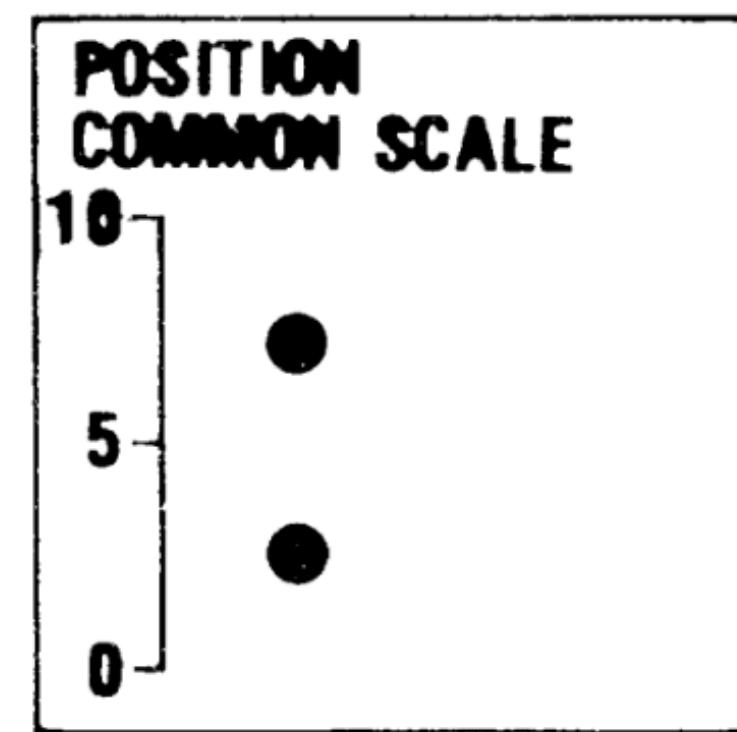
Graphical perception refers to the visual decoding of information encoded in graphs.

Elementary perceptual tasks are the tasks people carry out when extracting quantitative information from graphs.

A series of **studies in perception**:

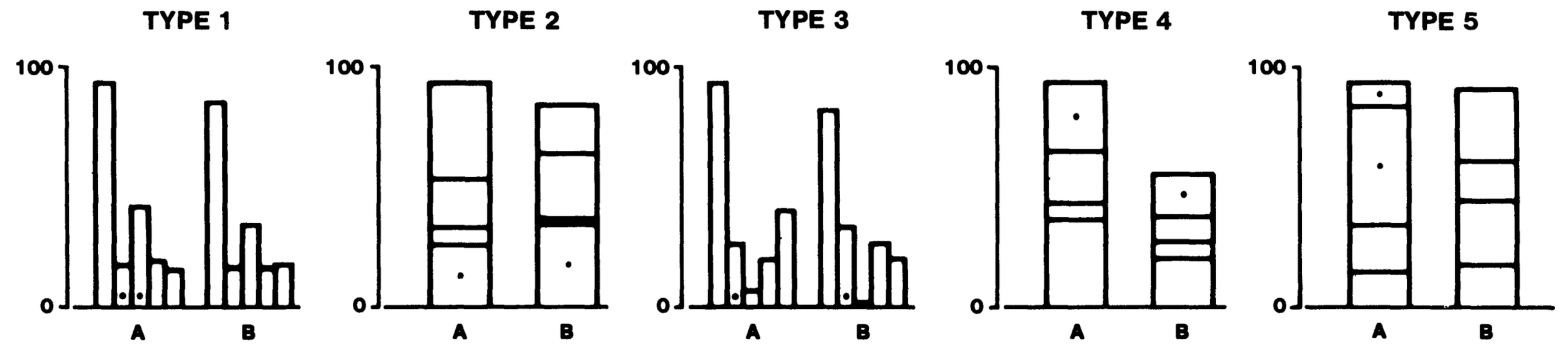
- Cleveland & McGill
- Bostock & Heer
- Croxton
- Zacks & Tversky
- Newman & Scholl
- Talbot et al.
- Fischer, Dewulf, & Hill
- Eells
- Simkin & Hastie
- Kong et al.
- Zacks et al.

In 1984, **William S. Cleveland & Robert McGill** conducted a study to identify the “perceptual building blocks” behind visual comprehension.



COLOR SATURATION

Figure 1. Elementary perceptual tasks.



Ranked list of elementary perceptual tasks

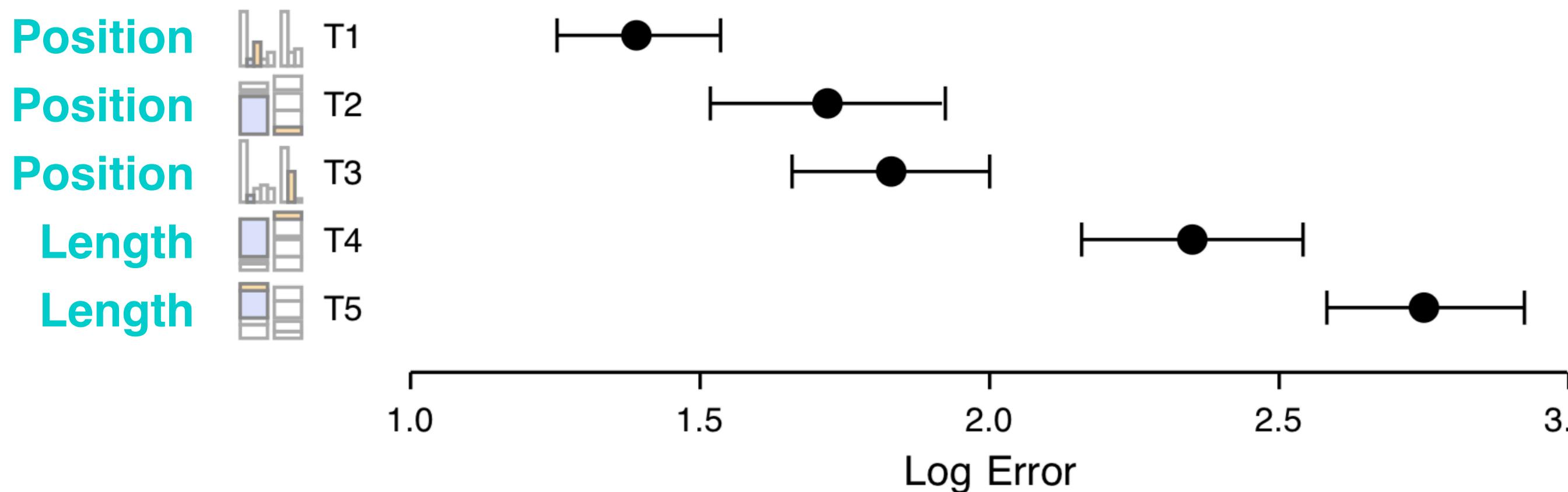
1. **Position** along a common scale
2. **Position** on identical but *nonaligned* scales
3. **Length, Angle & Slope** (Tie)
4. **Area**
5. **Volume, Curvature**
6. **Shading, Color Saturation**

1984 was awhile ago. How do these results compare to viewers today?

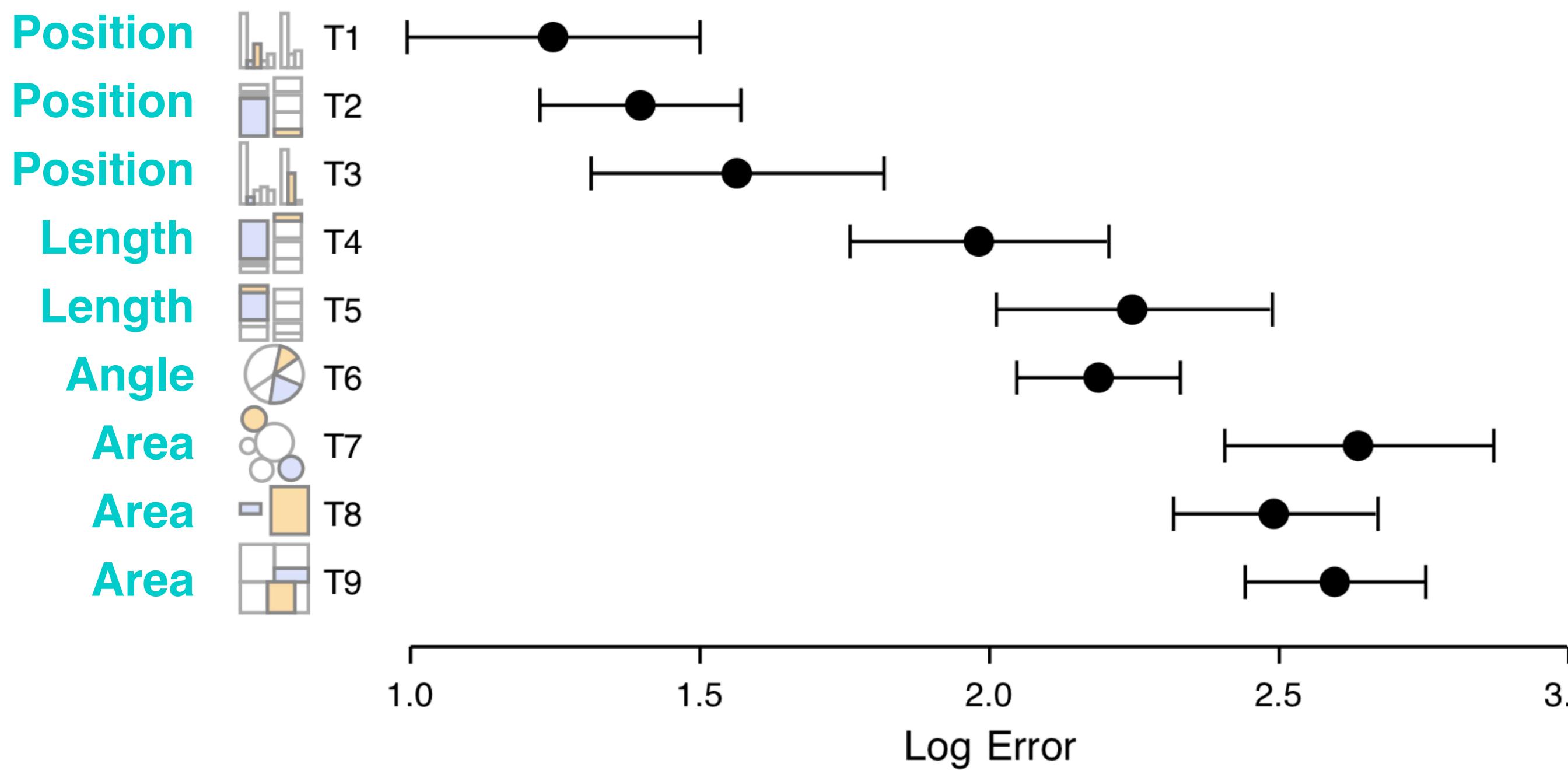
1984 was awhile ago. How do these results compare to viewers today?

...in **2010**, Michael Bostock and Jeffrey Heer re-tested Cleveland & McGill's findings on Mechanical Turk workers.

Cleveland & McGill's Results

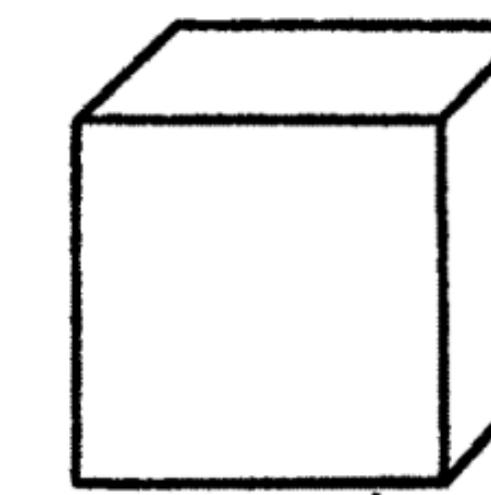
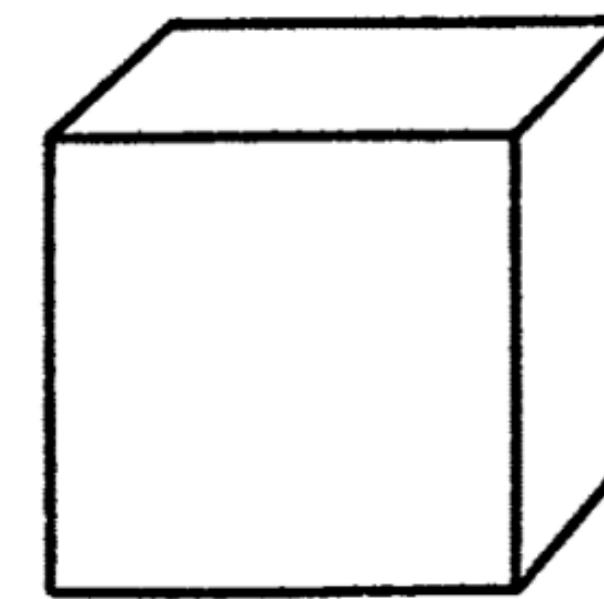
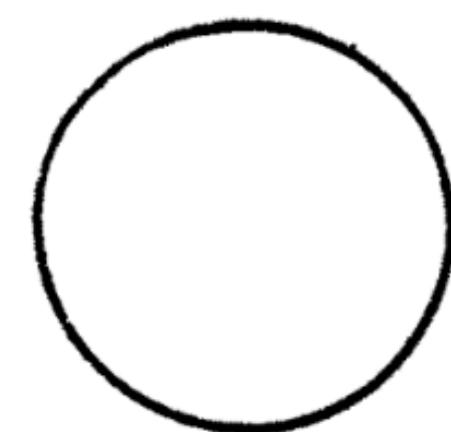
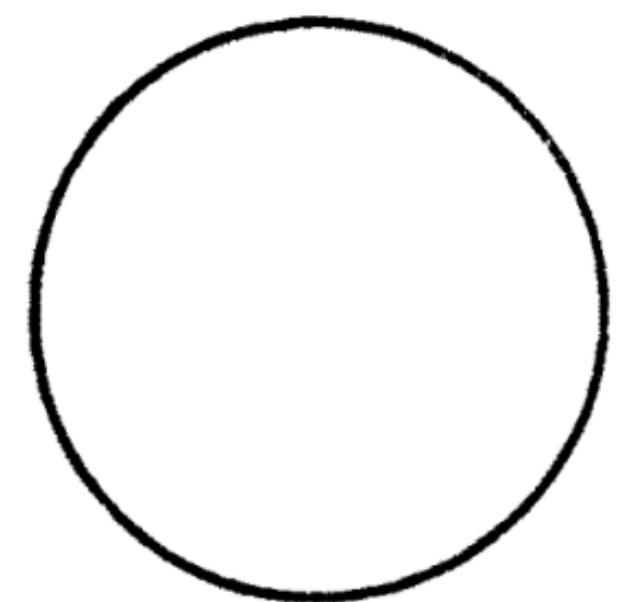
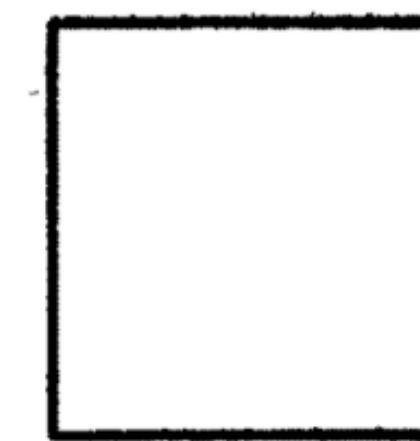
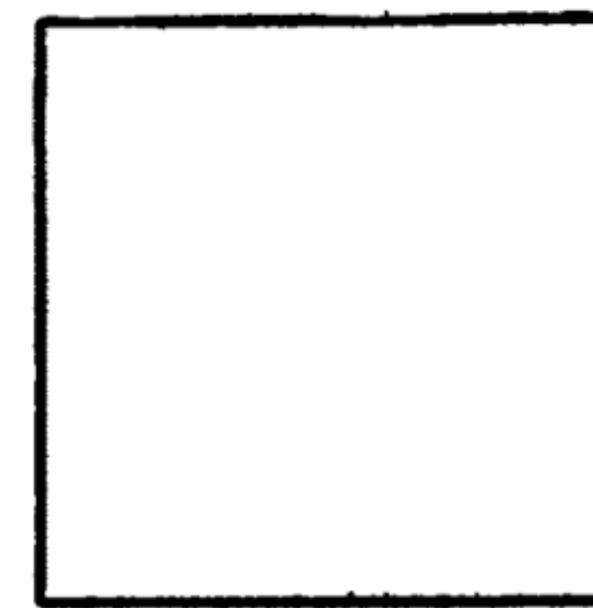


Crowdsourced Results

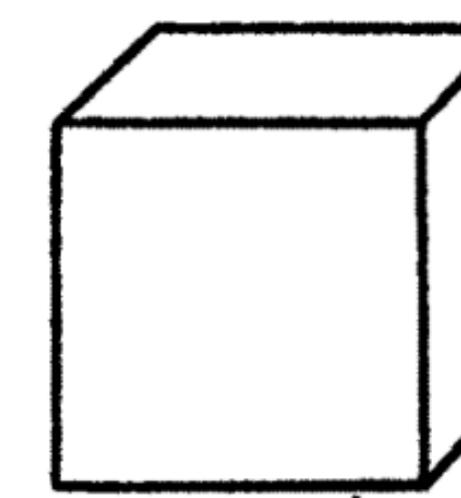
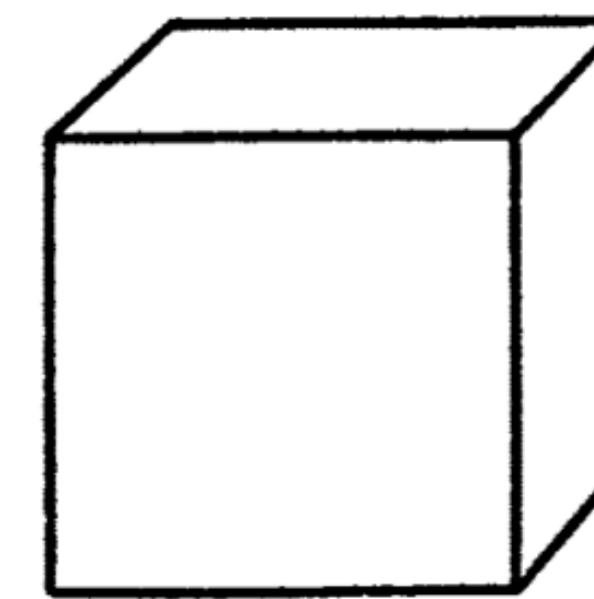
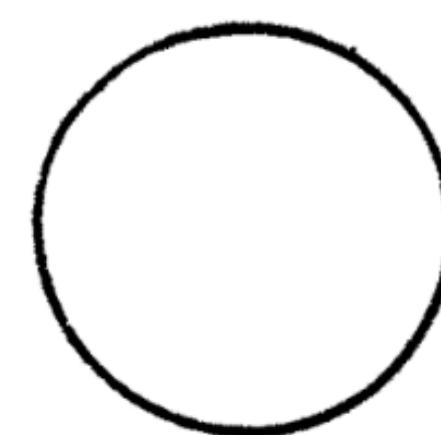
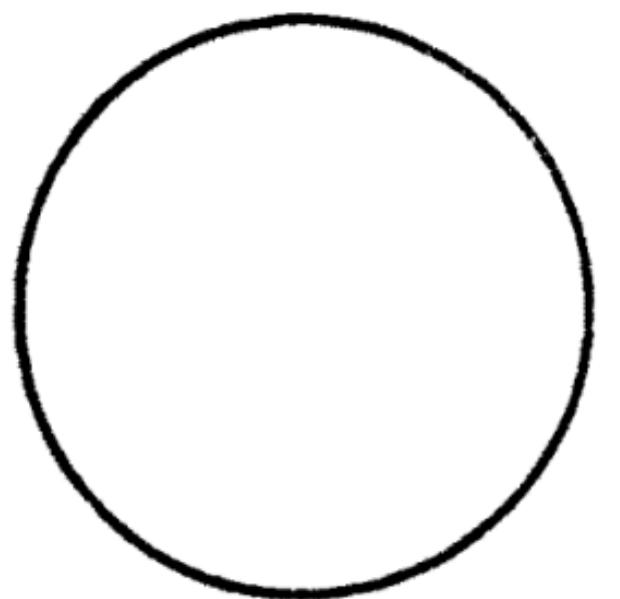
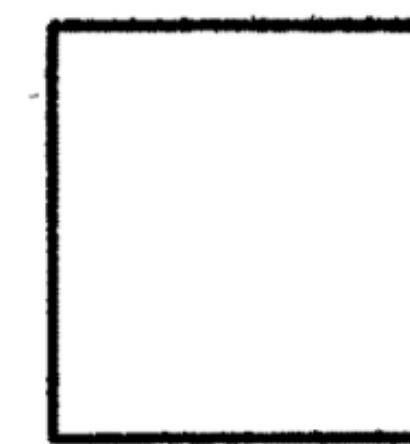
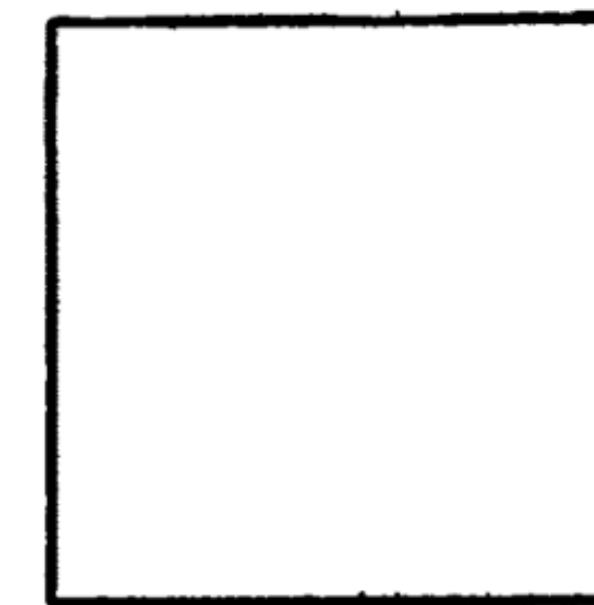


In 1932, **Frederick E. Croxton** also concluded that position is more easily judged than area or volume comparisons.

BARS, SQUARES, CIRCLES, AND CUBES SHOWING 50 TO 100 RELATIONSHIP



BARS, SQUARES, CIRCLES, AND CUBES SHOWING 50 TO 100 RELATIONSHIP



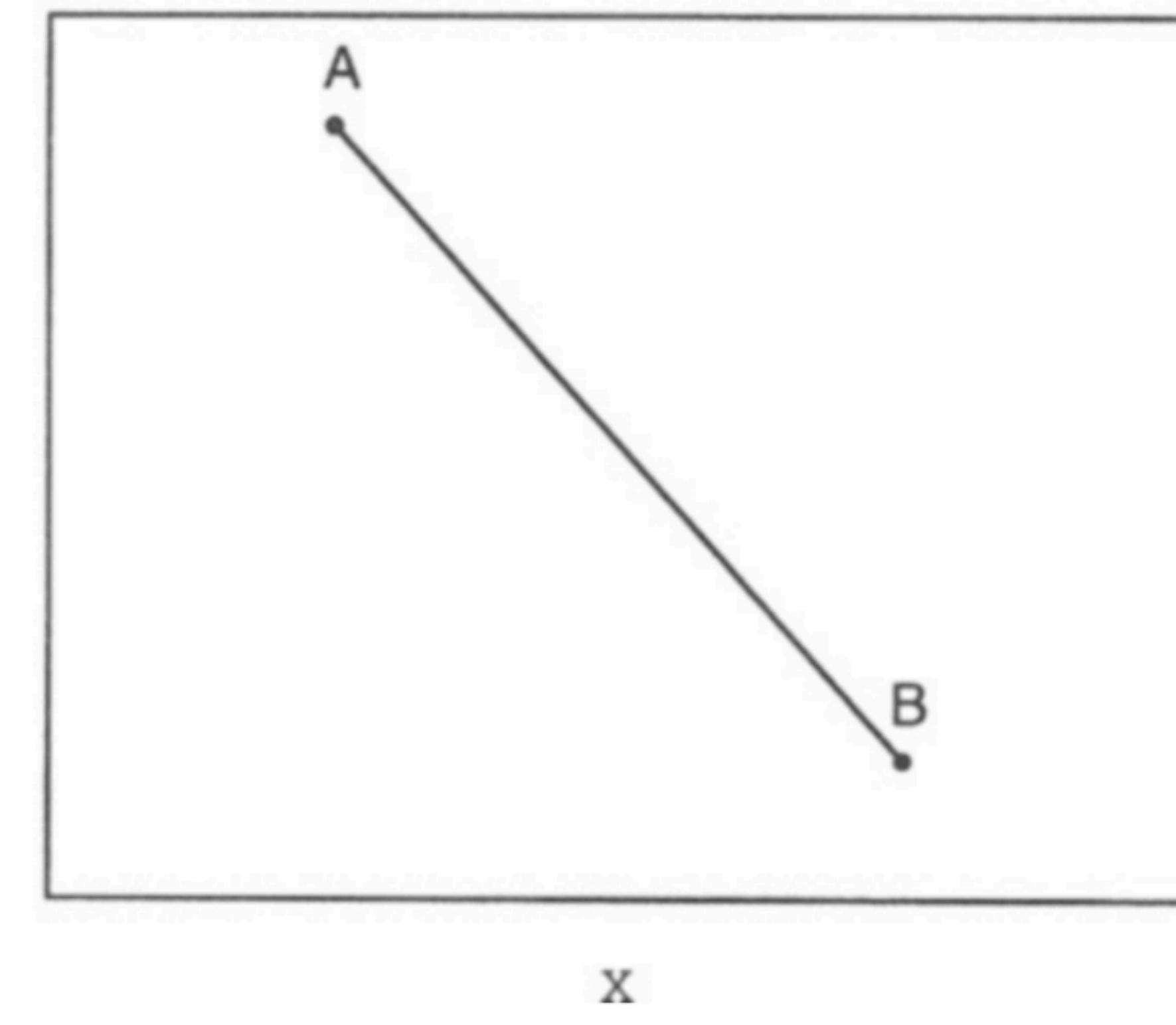
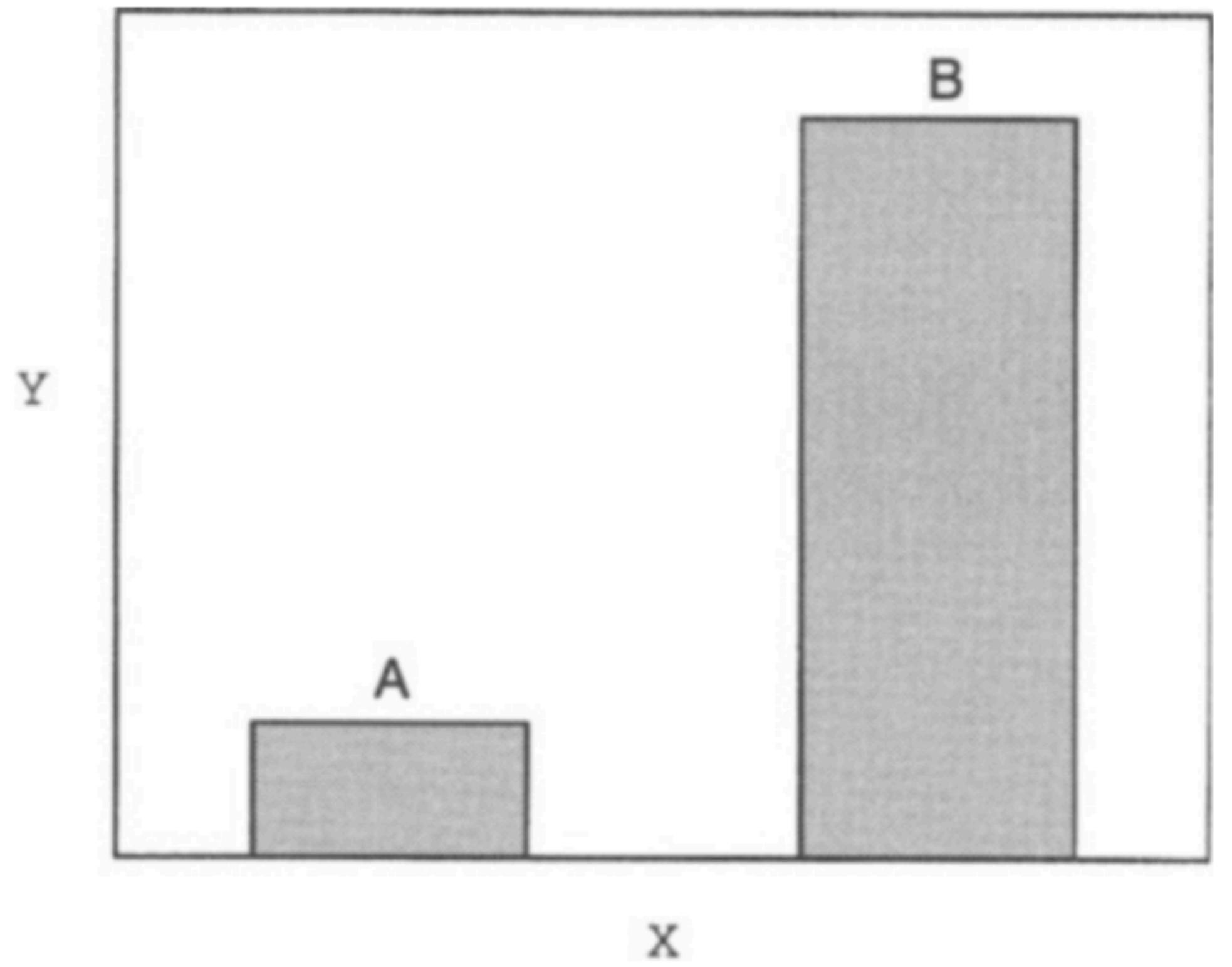
So, **always** use a bar chart.. right?

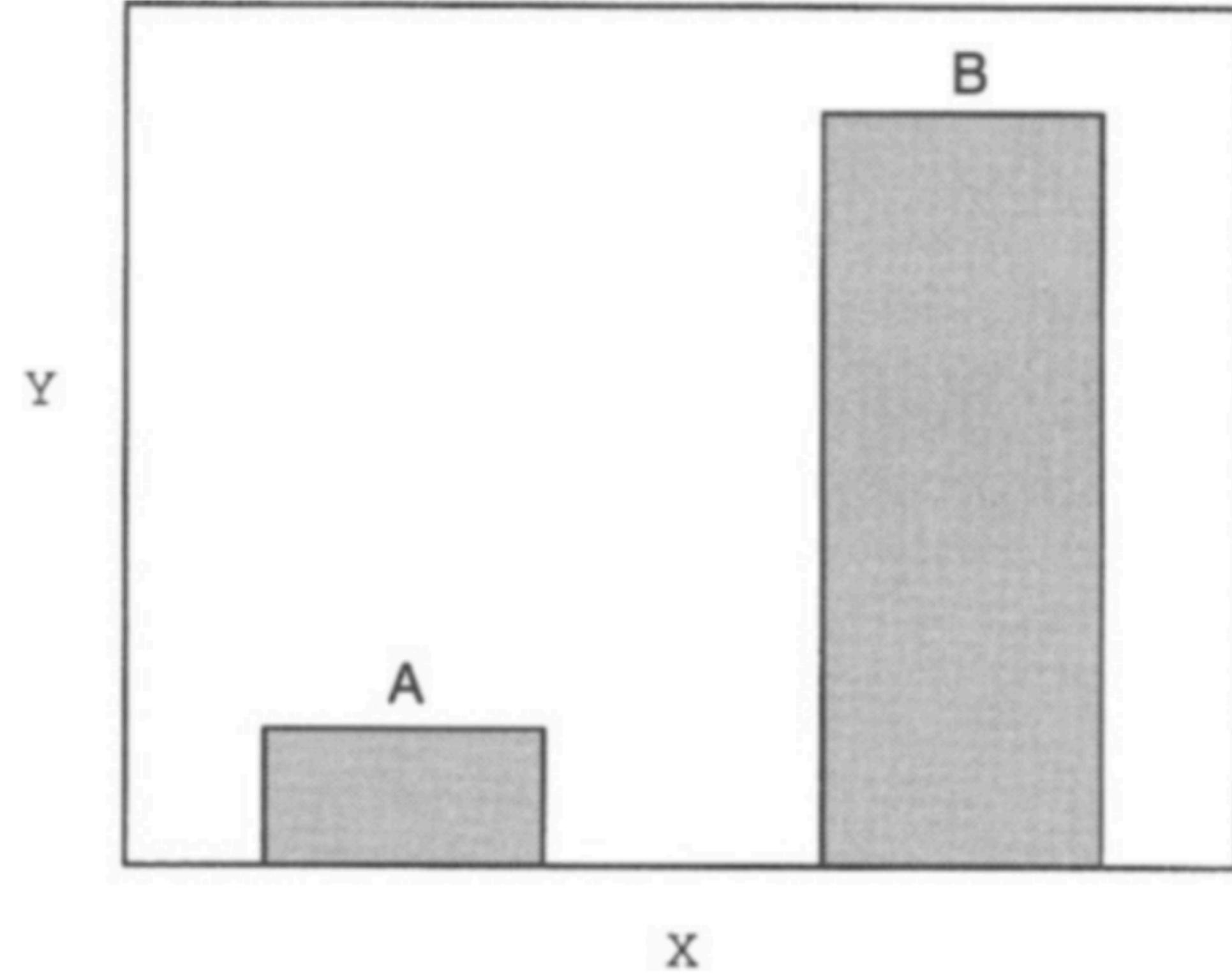
So, ~~always~~ use a bar chart.. right?

A study by Jeff Zacks and Barbara Tversky (1999) supports the idea that **visualization interpretation seems rooted in *cognitive naturalness*** as opposed to arbitrary correspondence.

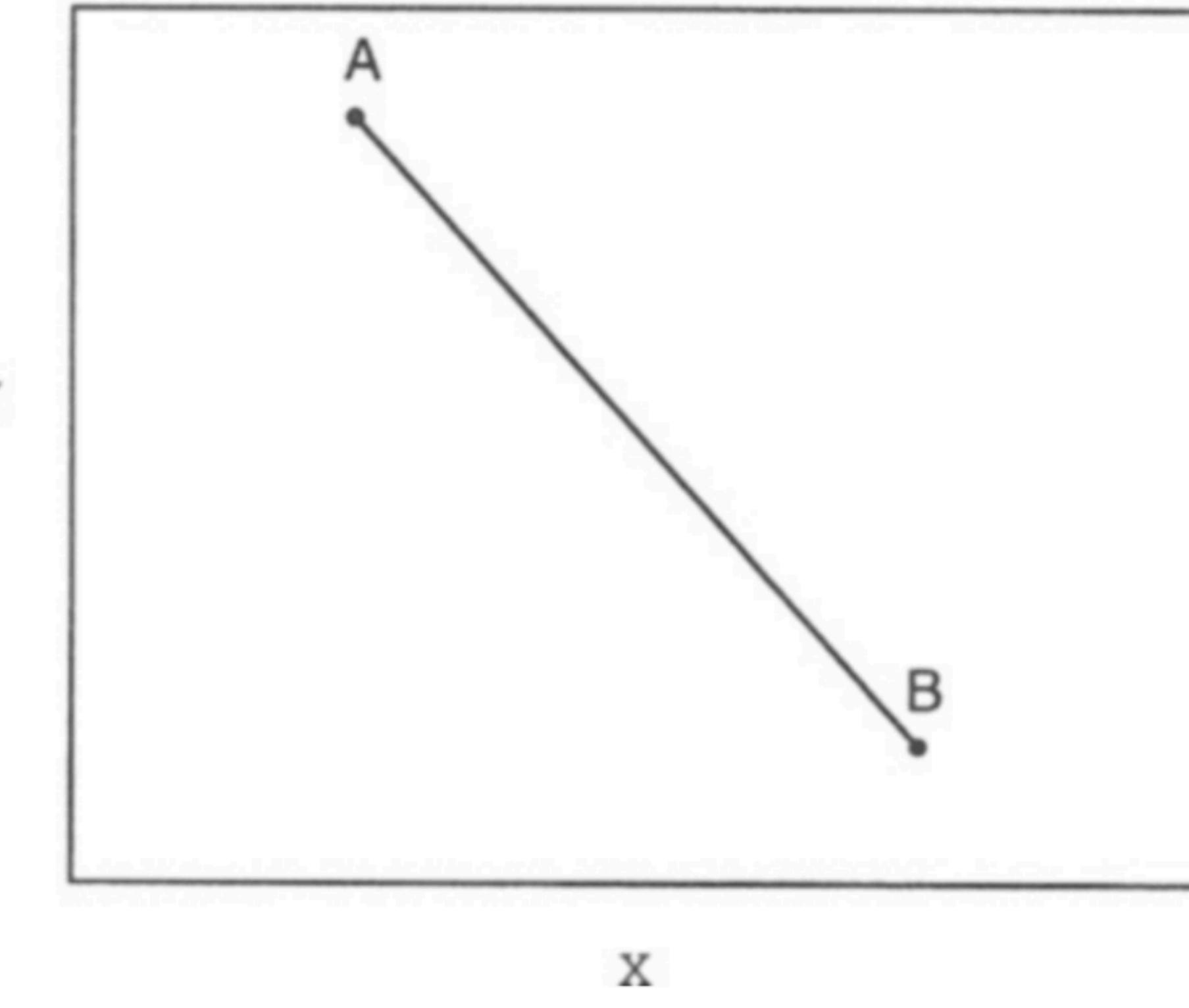
Bar-line message correspondence—

- People more readily associate **bars with discrete comparisons** between data points because bars are discrete entities and facilitate point estimates
- They more readily associate **lines with trends** because lines connect discrete entities and directly represent slope
- This correspondence does *not* seem to depend on knowledge of ‘rules’





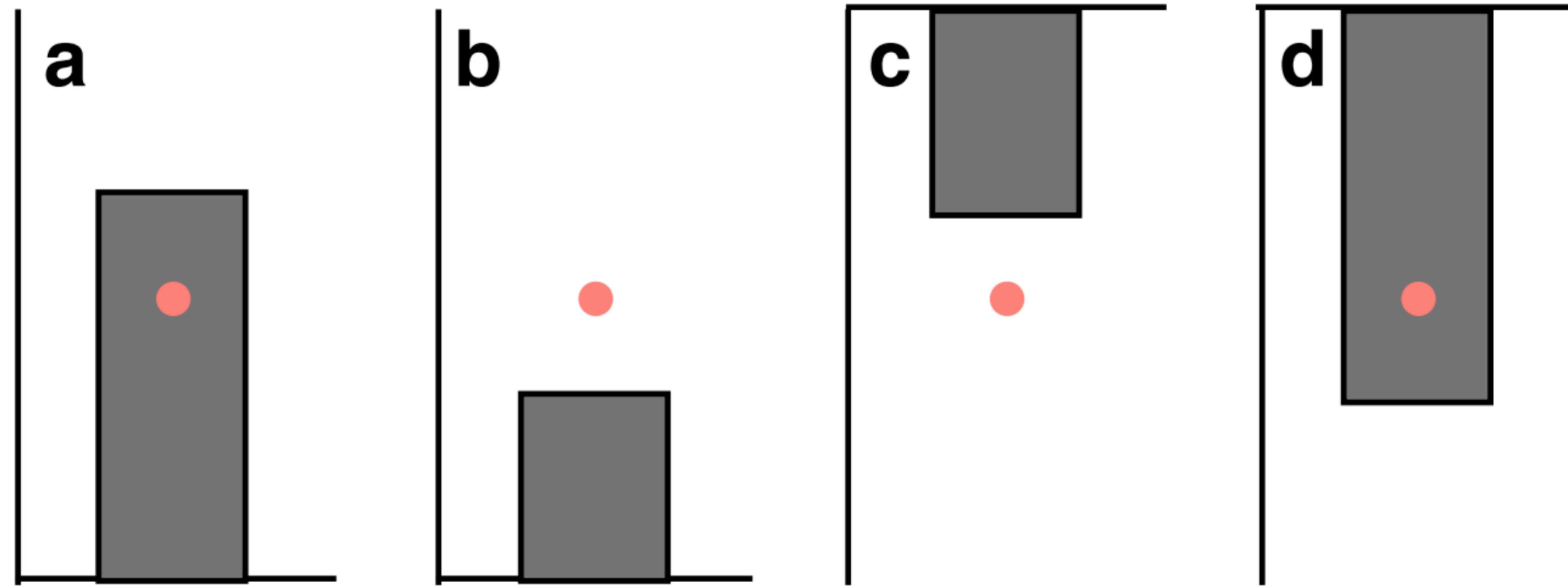
“B is higher than A.”



“A is decreasing.”

Multiple attributes further complicate bar charts.

In a study conducted in 2012, Newman and Scholl determined that the **placement of points within bar charts** affected readability.



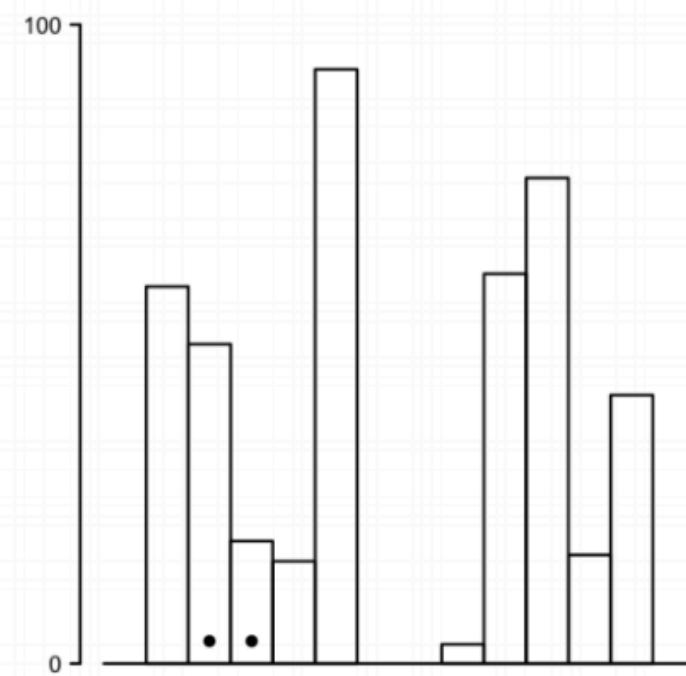
Within-the-bar bias—

- Bar charts are often used to depict measures of central tendency, but do so *asymmetrically* (i.e. the edge represents a value)
- Mean is often depicted as a point
- **Graphical asymmetry → cognitive asymmetry**
- Viewers judge points *within the bar* as more likely to fall into the underlying distribution than those that fall *outside the bar*, even if the two points are equidistant from the mean

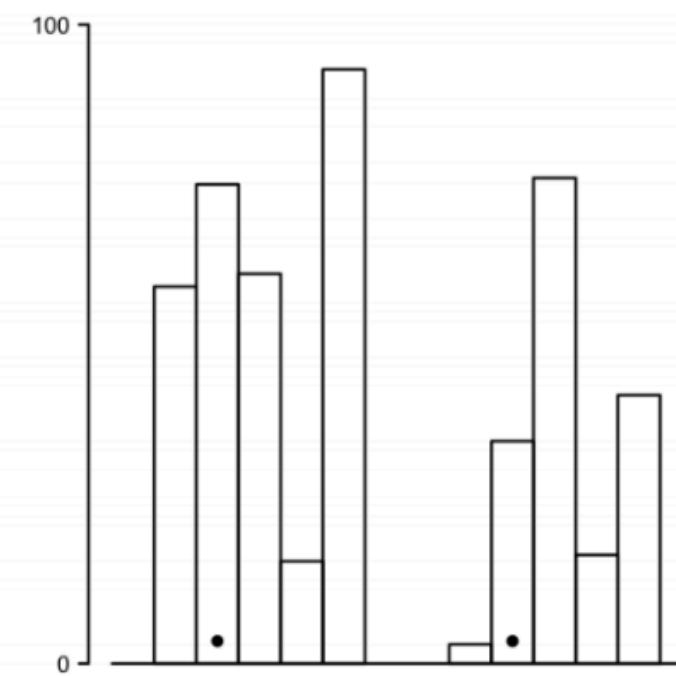
For reports of **central tendencies**, use points instead of bars.

For representing values that are **inherently asymmetric** (counts, ranges, measures of extremity), bars are okay.

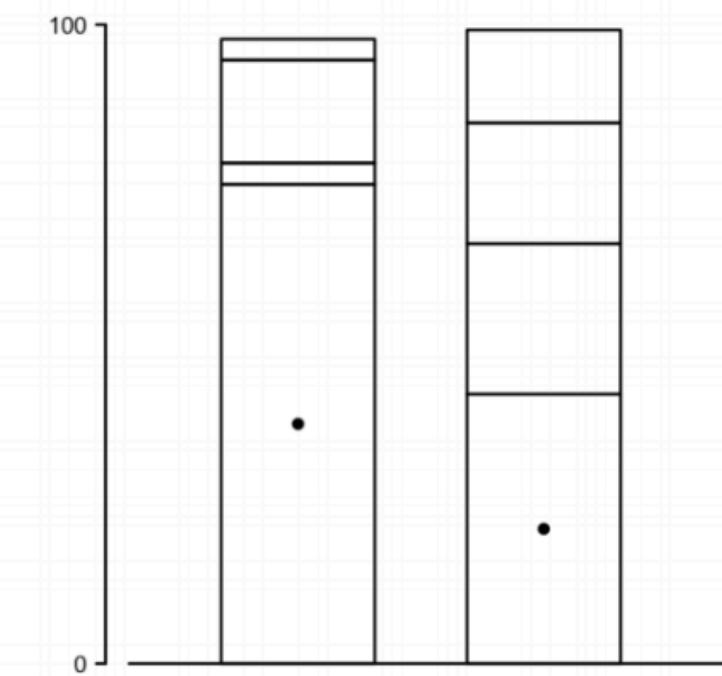
In 2014, Talbot et al. found that **point placement affected readability** in another rendition of Cleveland & McGill's study.



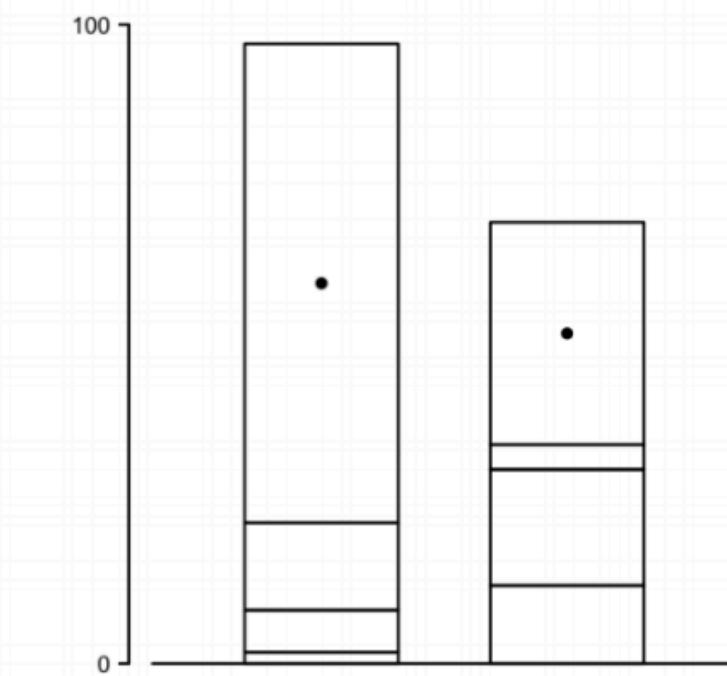
(a) Adjacent Bars
(Type 1)



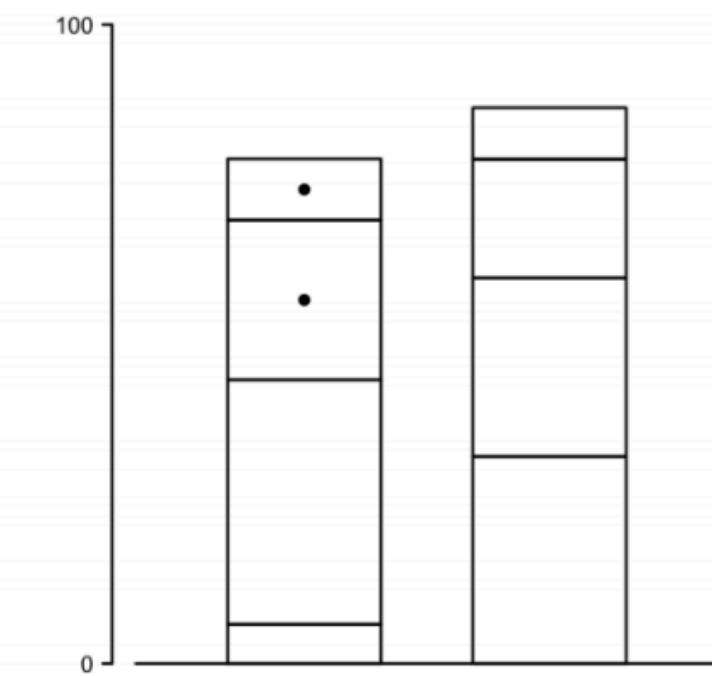
(b) Separated Bars
(Type 3)



(c) Aligned Stacked Bars
(Type 2)



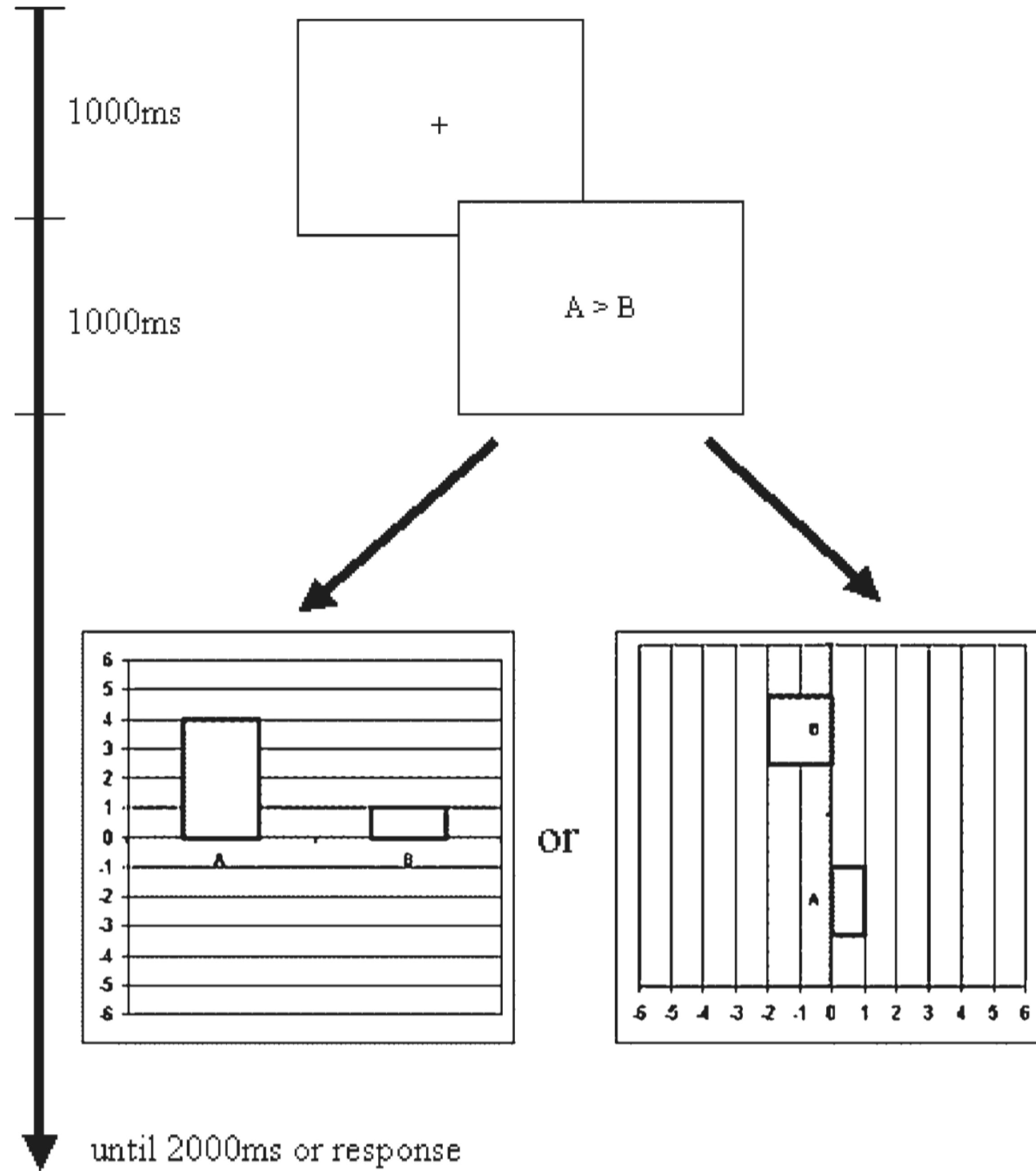
(d) Unaligned Stacked Bars
(Type 4)



(e) Divided Bar
(Type 5)

Bar orientation makes a difference, too.

Fischer, Dewulf, and Hill determined that **vertical bar charts are slightly more effective** than horizontal bar charts...

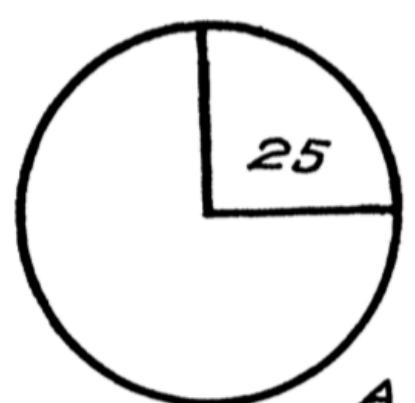


...particularly when dealing with *negative* values. Subjects associated **negativity with a downward movement** more regularly than a “left-ward” movement.

What about **pie charts**?

Actually, they're **OK.**

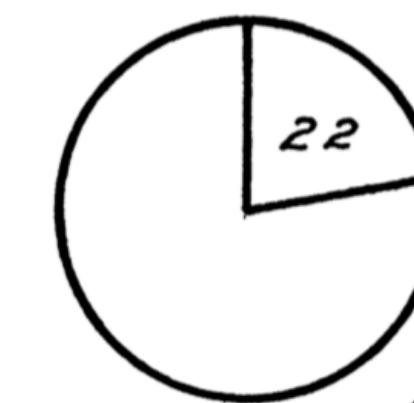
In a 1926 study, Walter Crosby Eells determined that pie charts can be read **“fully as rapidly and easily”** as stacked bars.



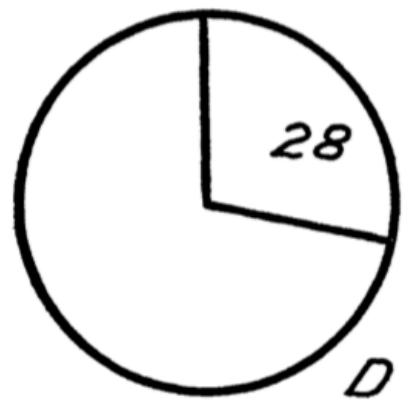
A



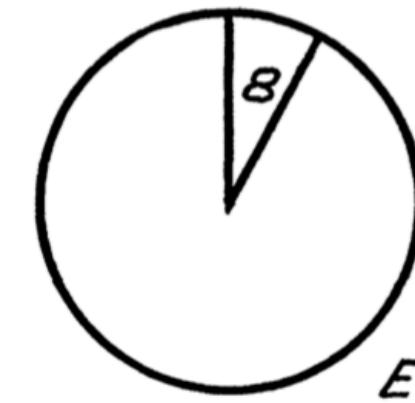
B



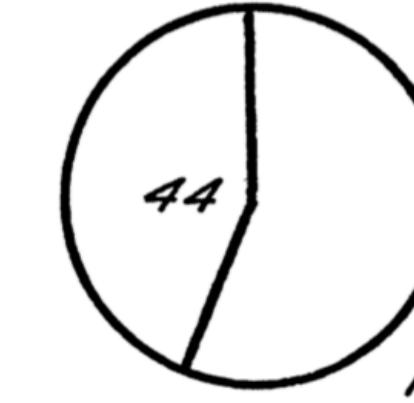
C



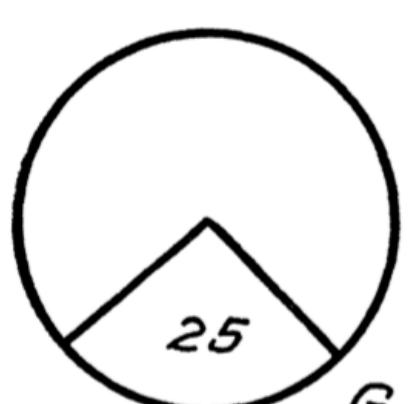
D



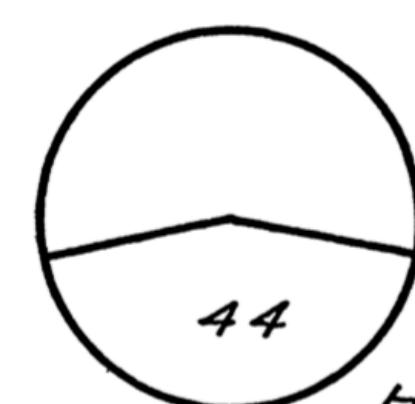
E



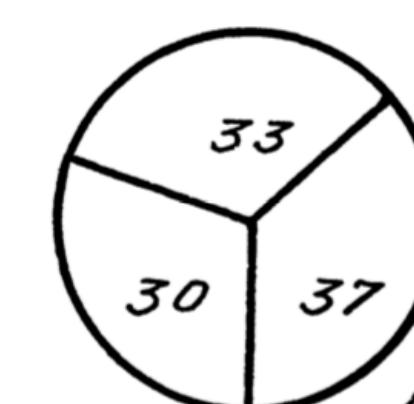
F



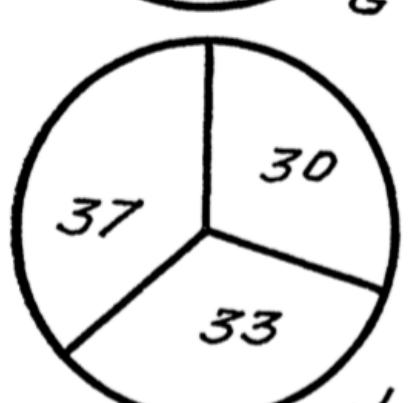
G



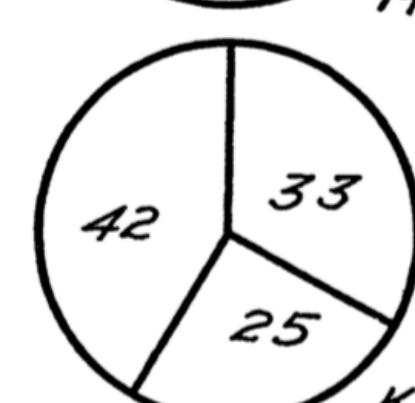
H



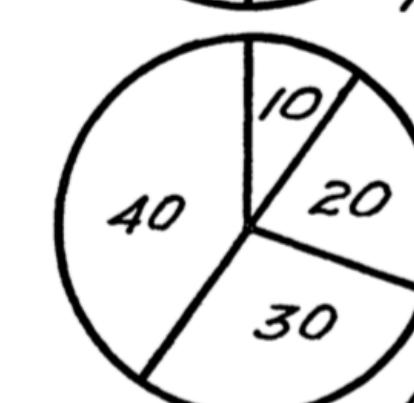
I



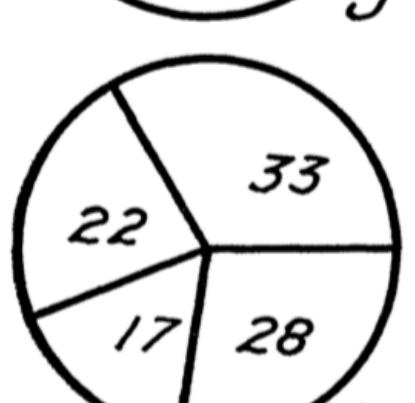
J



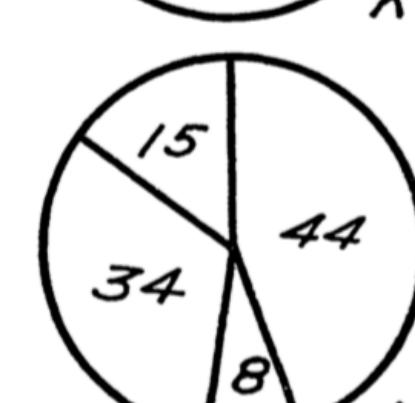
K



L



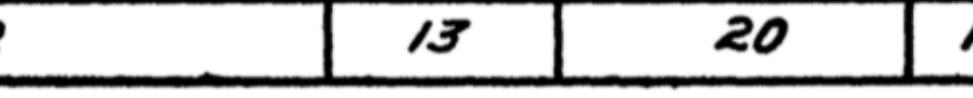
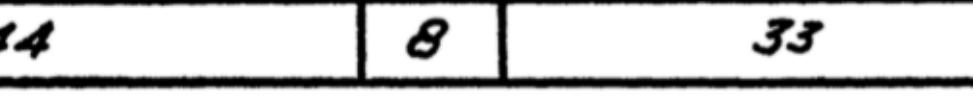
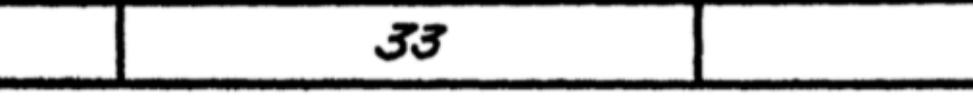
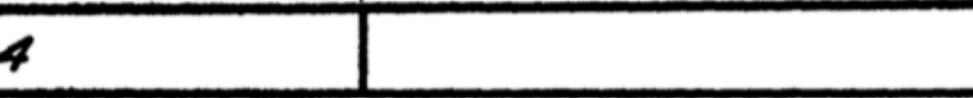
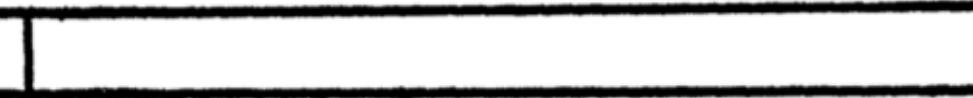
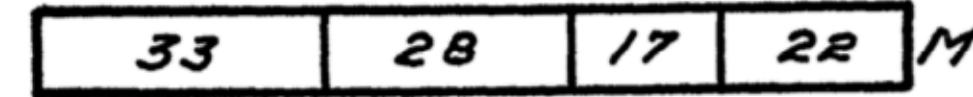
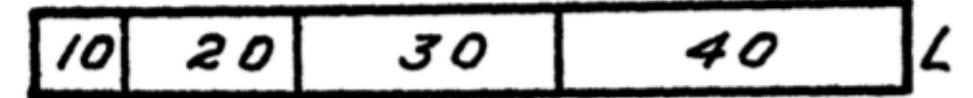
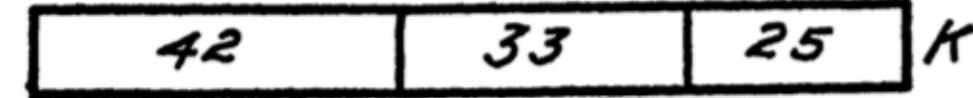
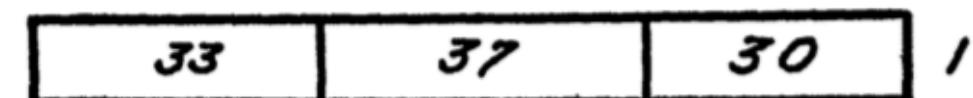
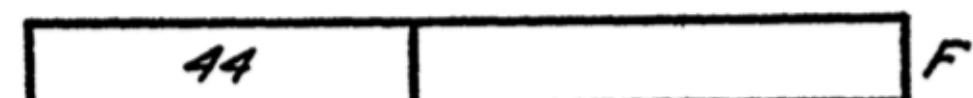
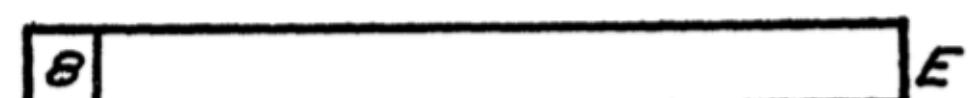
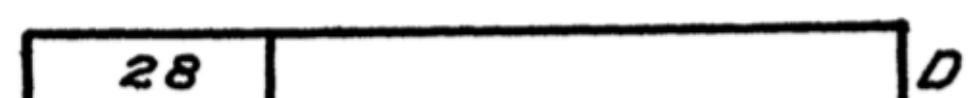
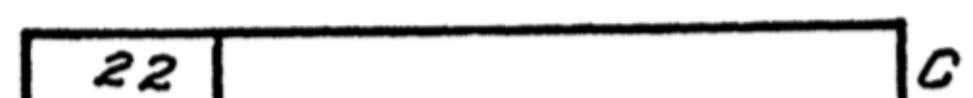
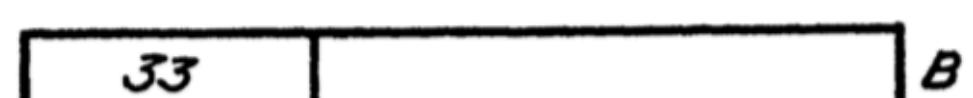
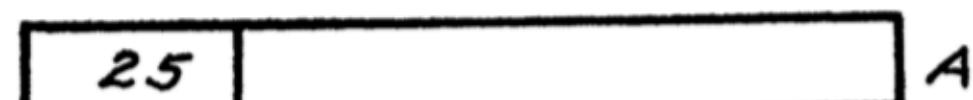
M



N

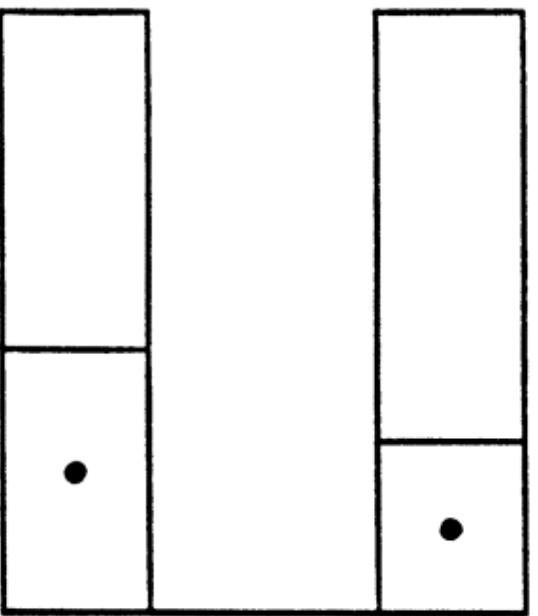


O

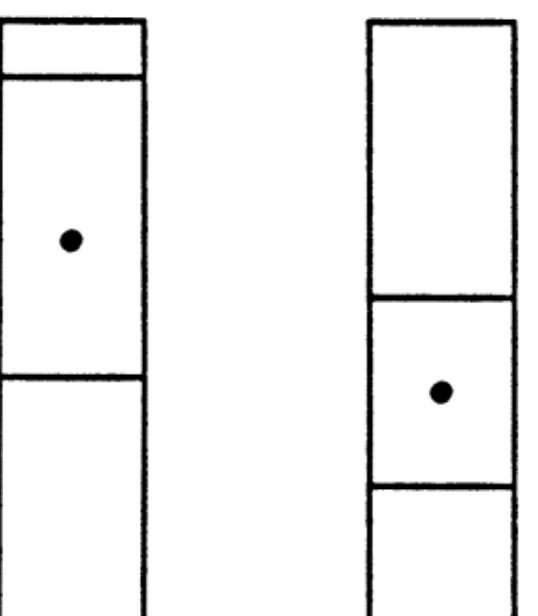


- **51%** of Ell's subjects judged the size of a wedge using the outer arc
- **25%** used areas
- **23%** used angles
- (1% didn't know how to read the pies and used chords)

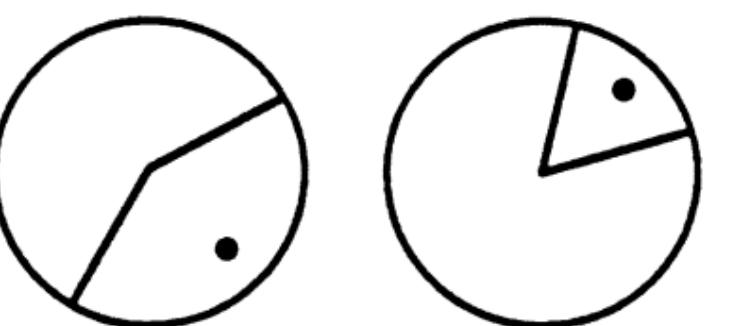
Simple bar chart



Divided bar chart



Pie Chart



In a later study (1987), Simkin and Hastie concluded that for **segment-to-segment comparisons**, simple bar charts worked best.

However, for **part-to-whole** judgments, *position* and *angle* produced more accurate assessments than *length*.

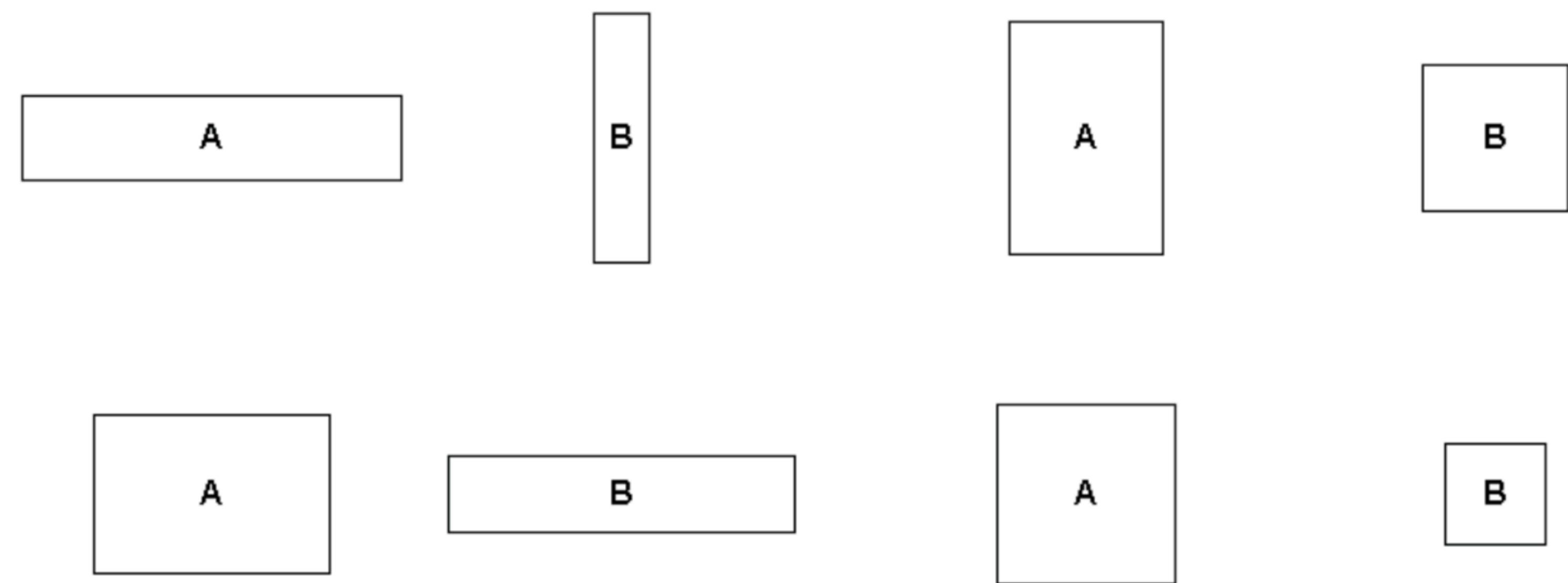
Several studies have since tackled the same issue, and have concluded that **pie charts do not perform as badly** as their reputation would suggest.

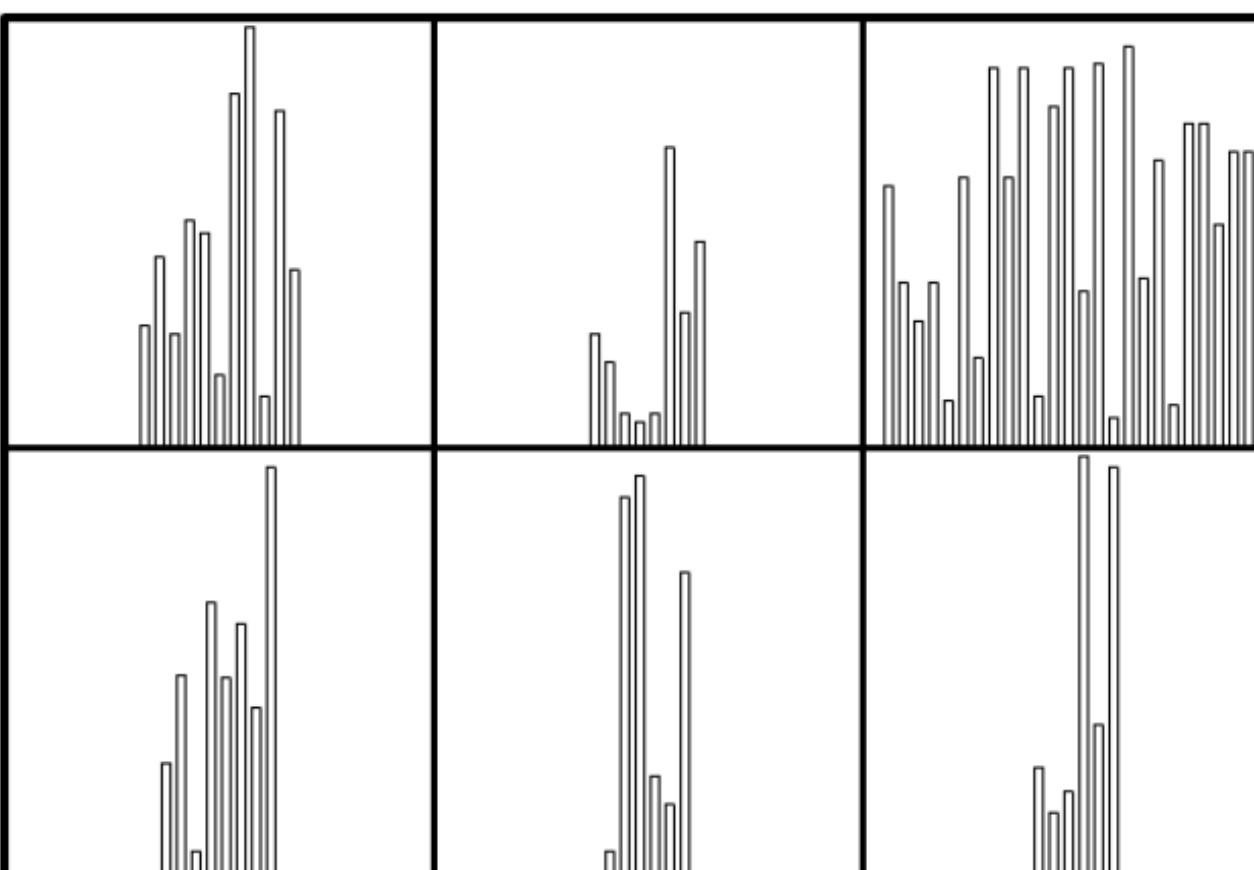
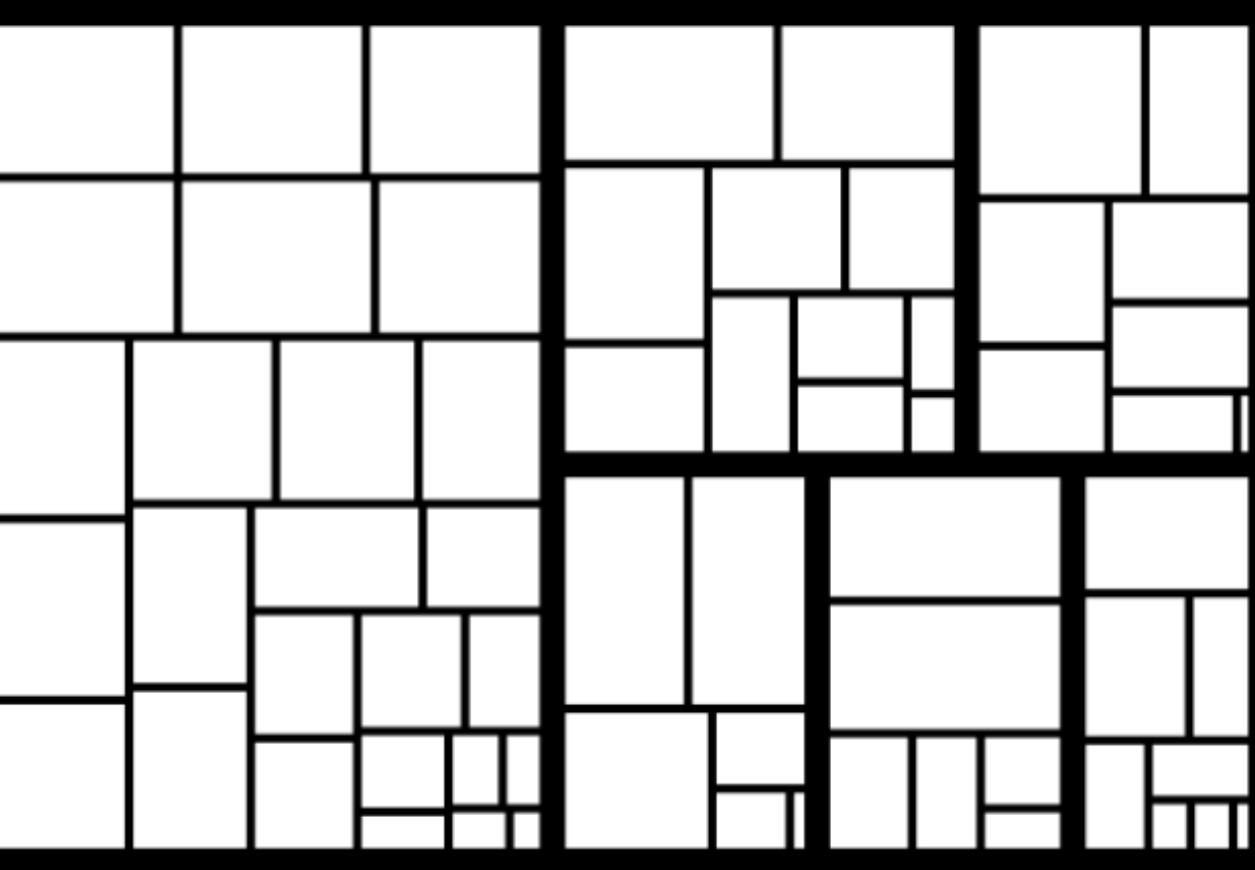
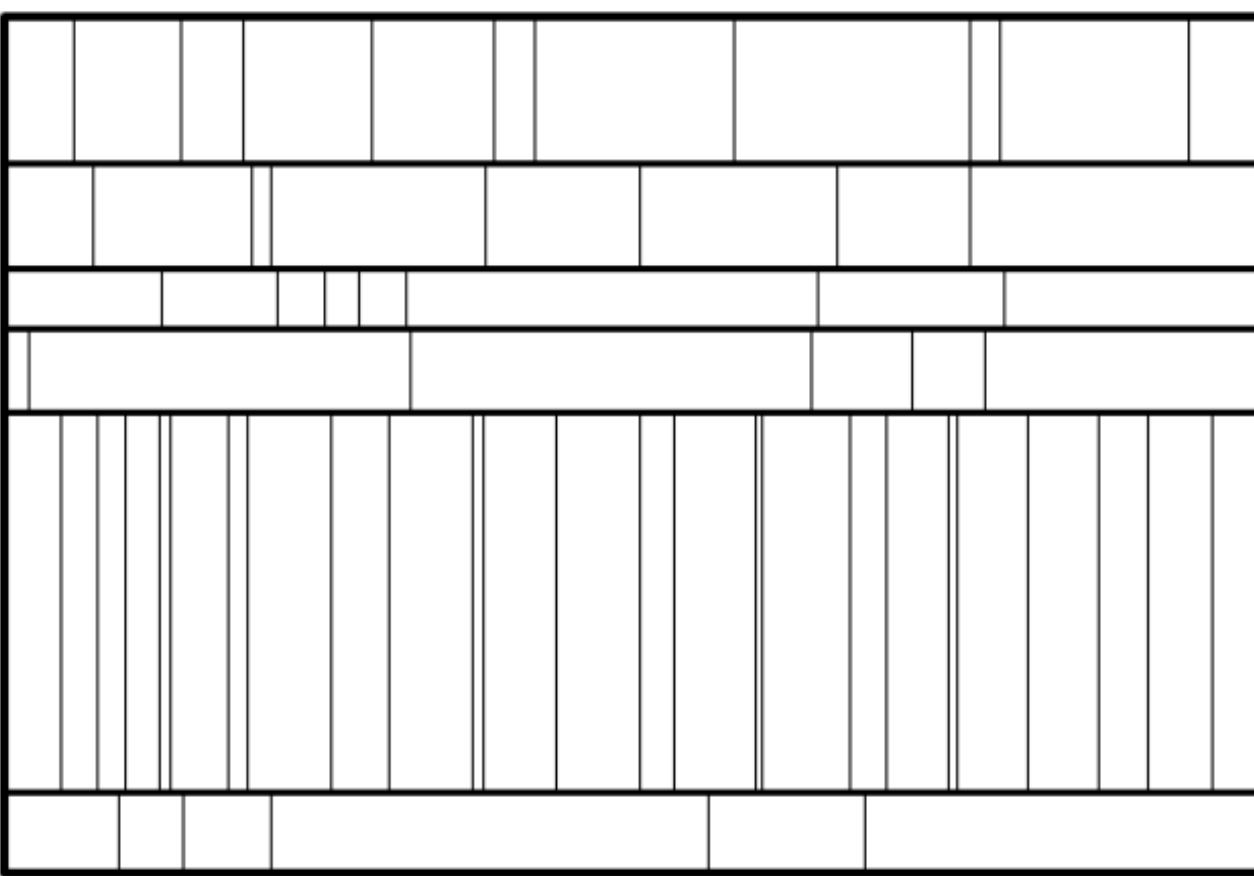
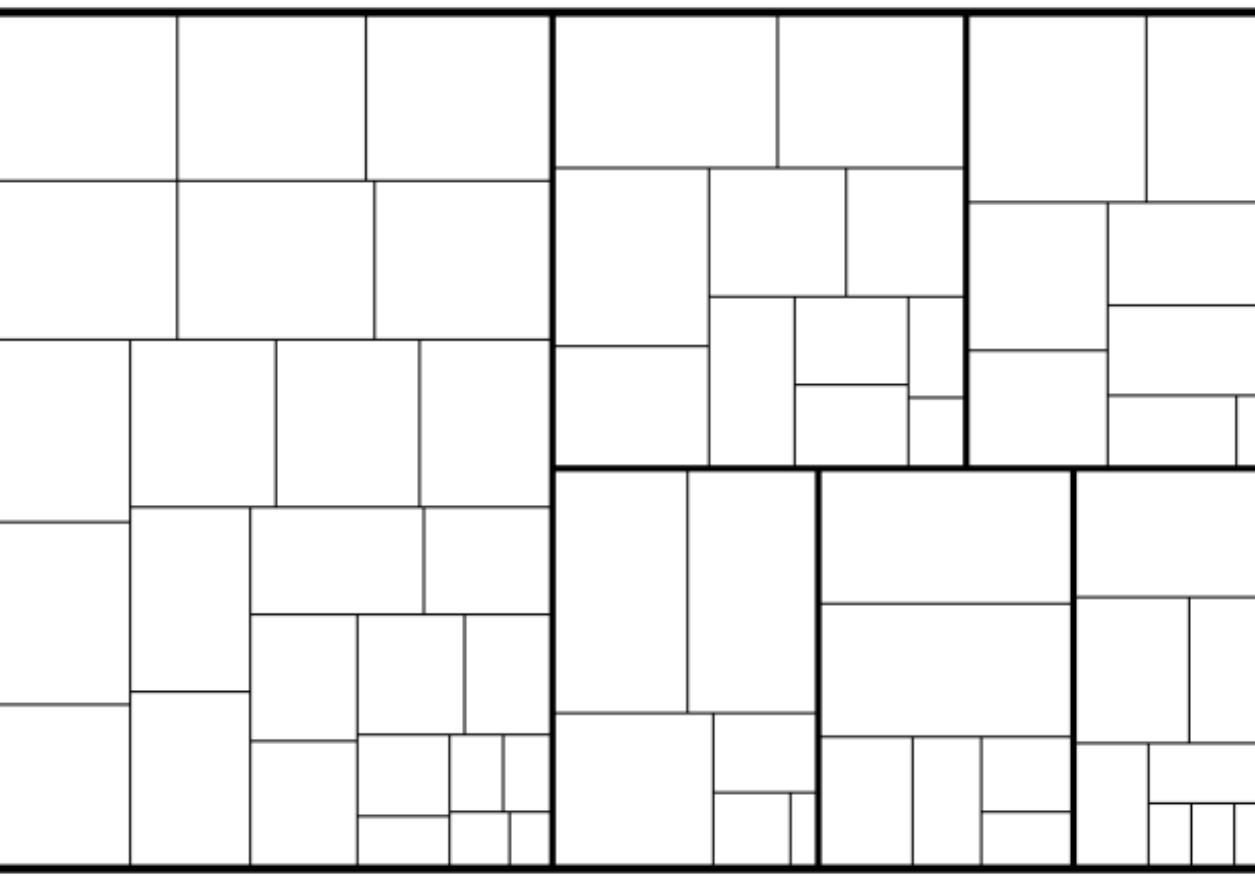
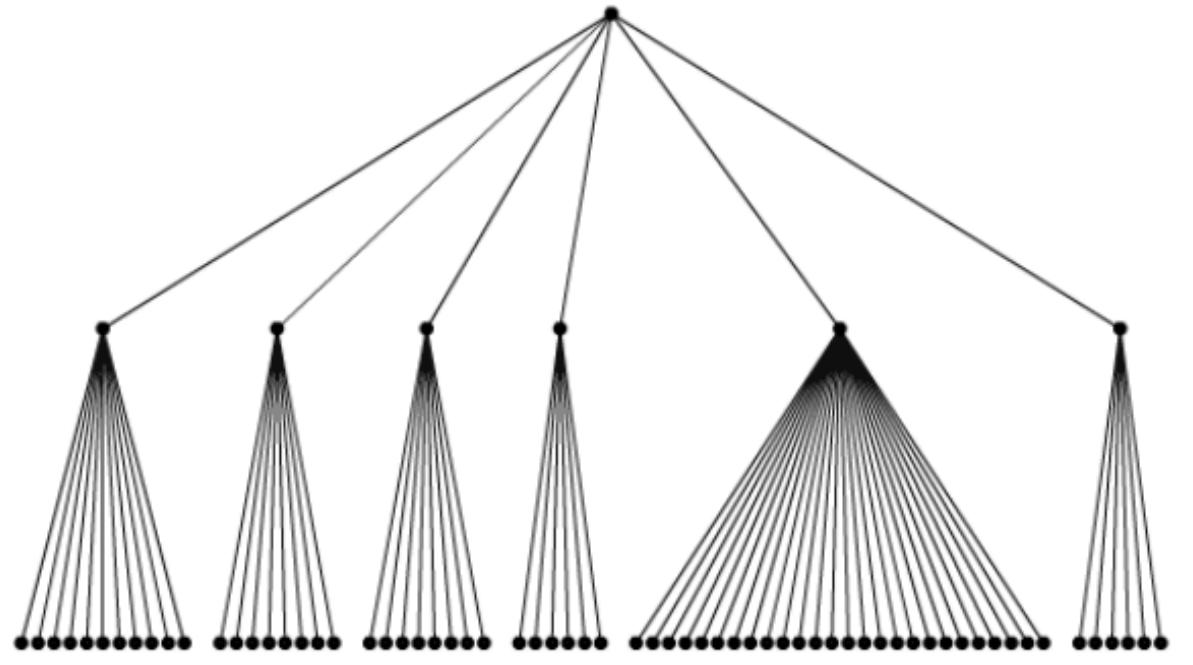
Speaking of **part-to-whole relationships...**

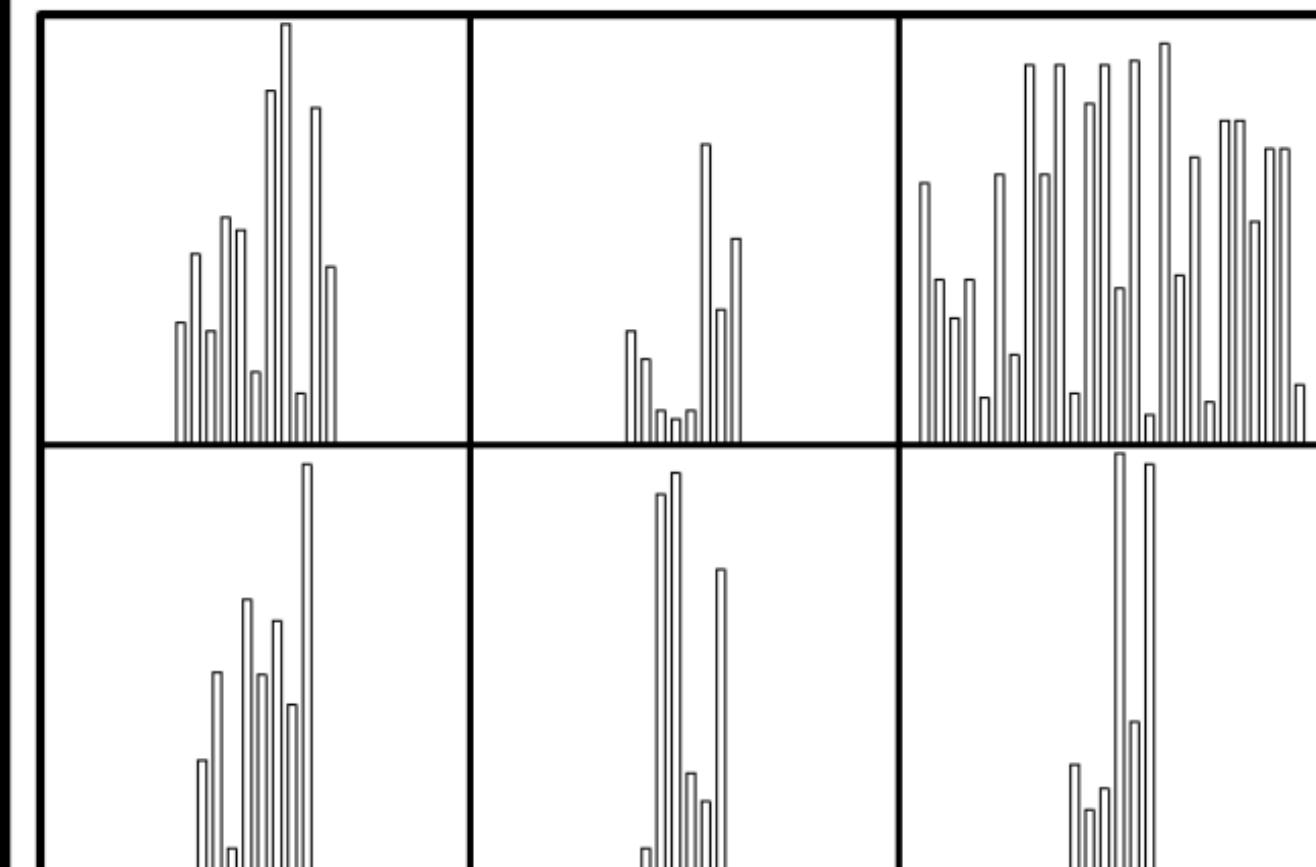
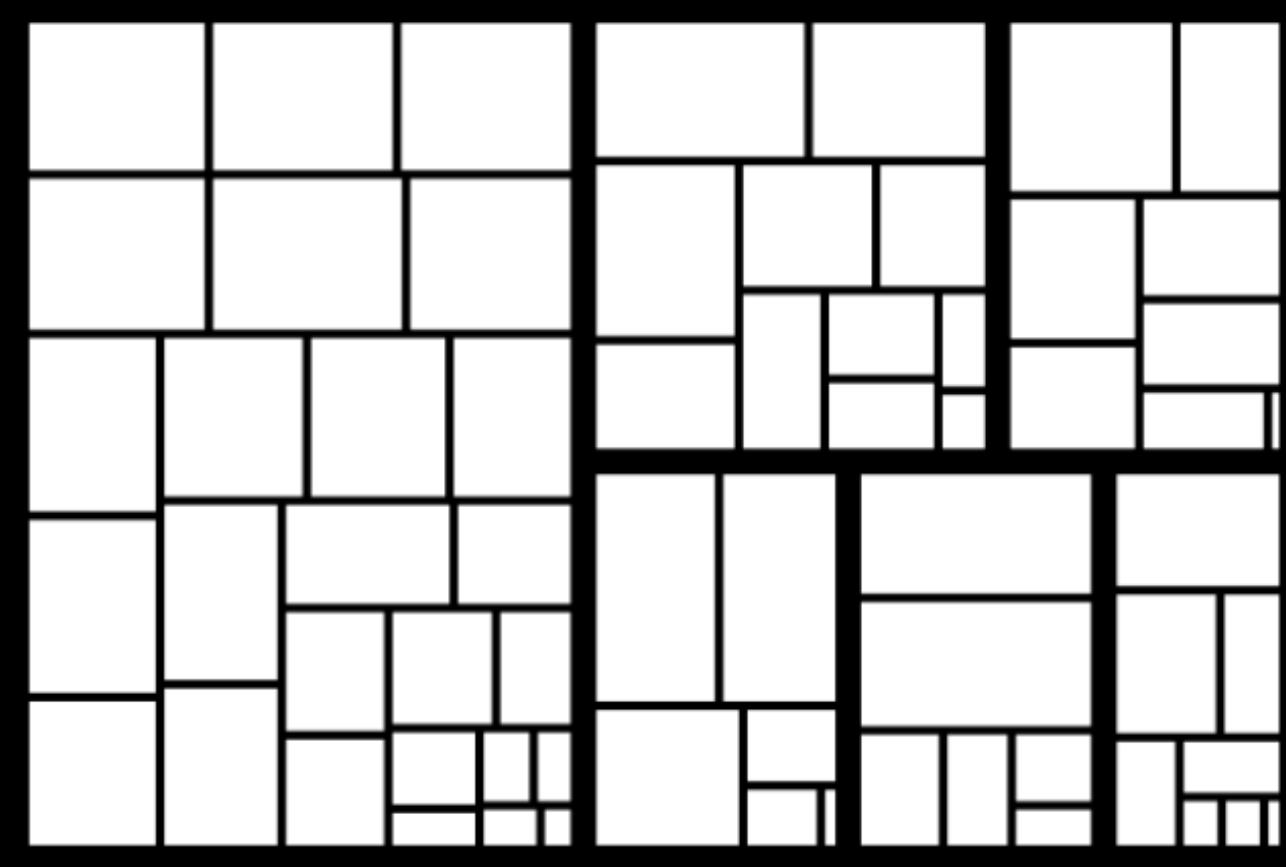
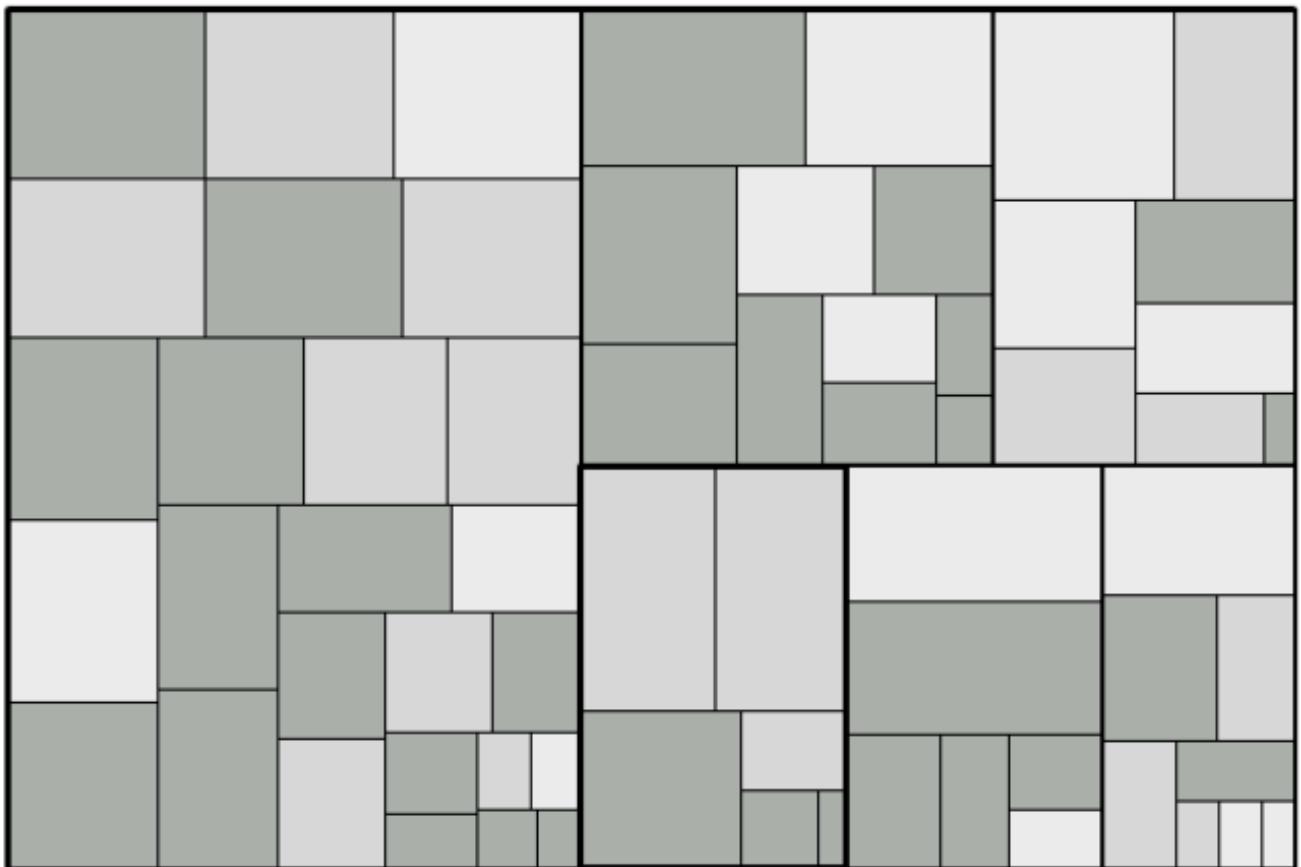
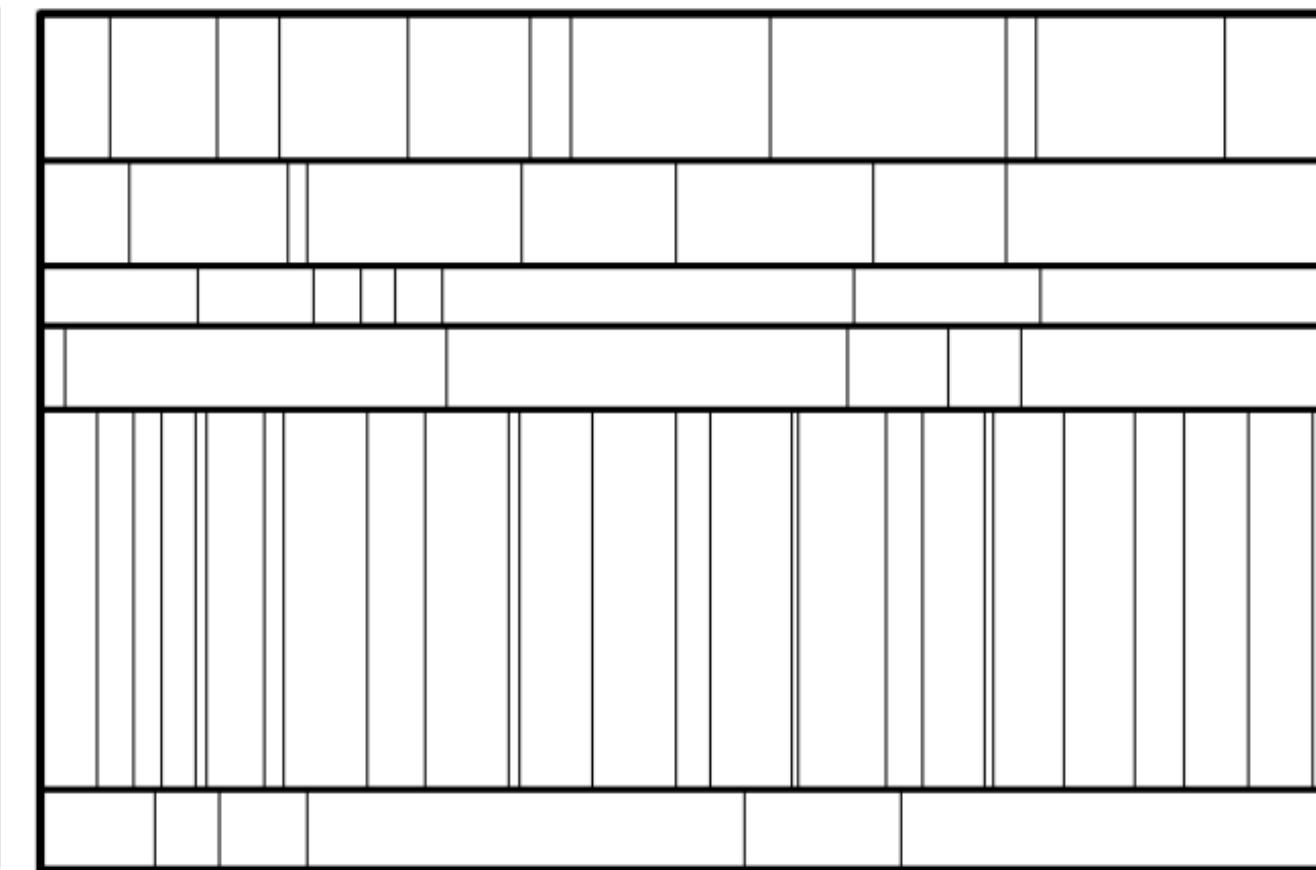
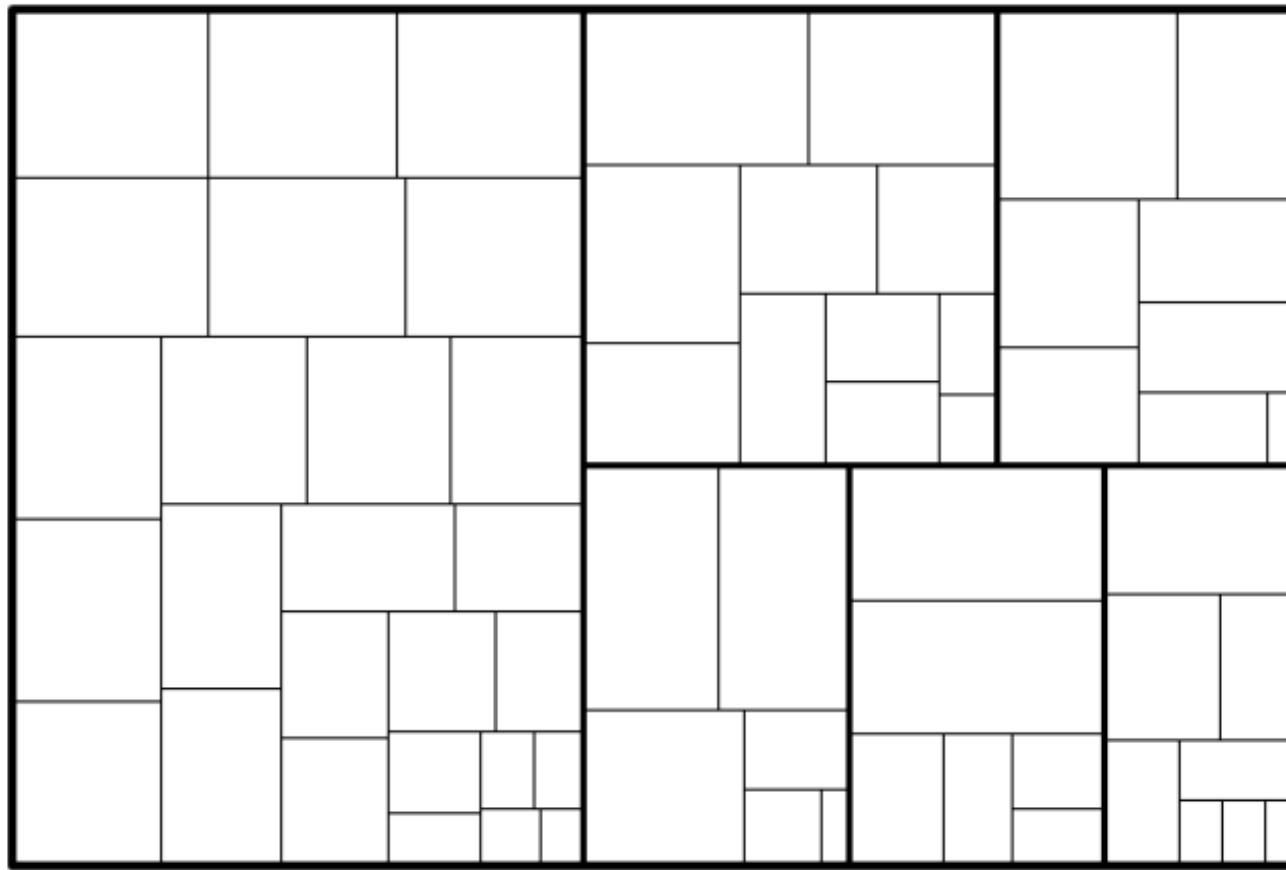
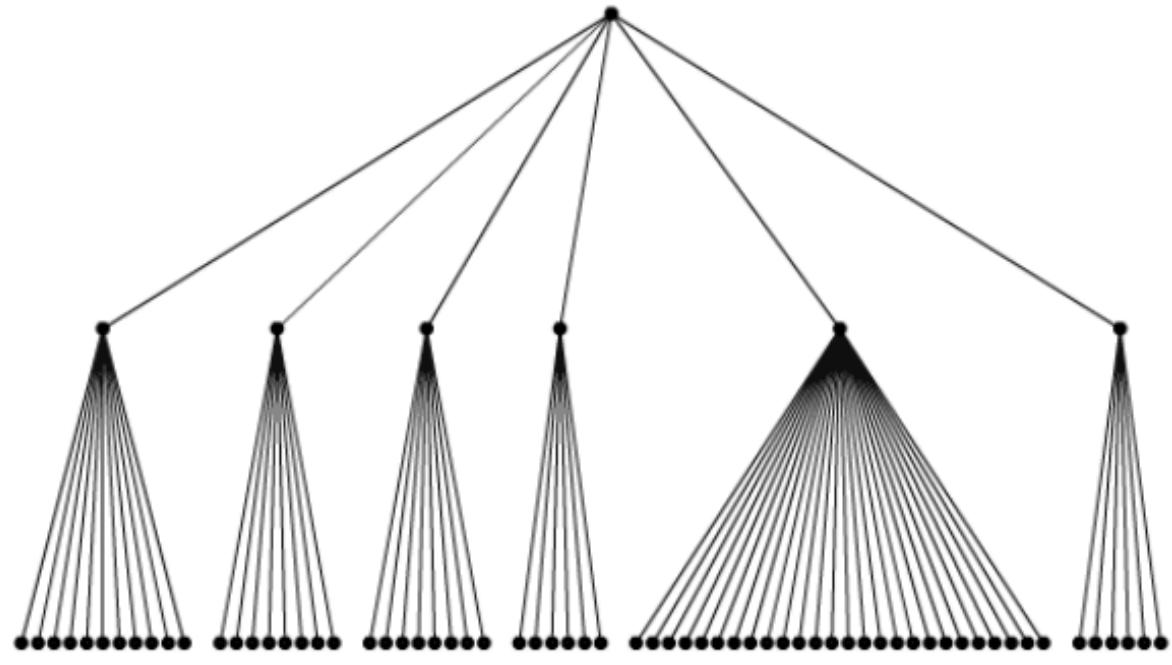
In 2010, Kong et al. isolated the **treemap design parameters** that can affect perception of rectangle area.

Treemap parameters affecting **readability**—

- Aspect ratios
- Luminance
- Border thickness
- Data density







Weird alternative

Though squares proved to be more difficult to compare than rectangles, **‘squarified’ treemap algorithms proved more successful** because the rectangles could typically not achieve “ideal squarification.”

Resultant **design guidelines**—

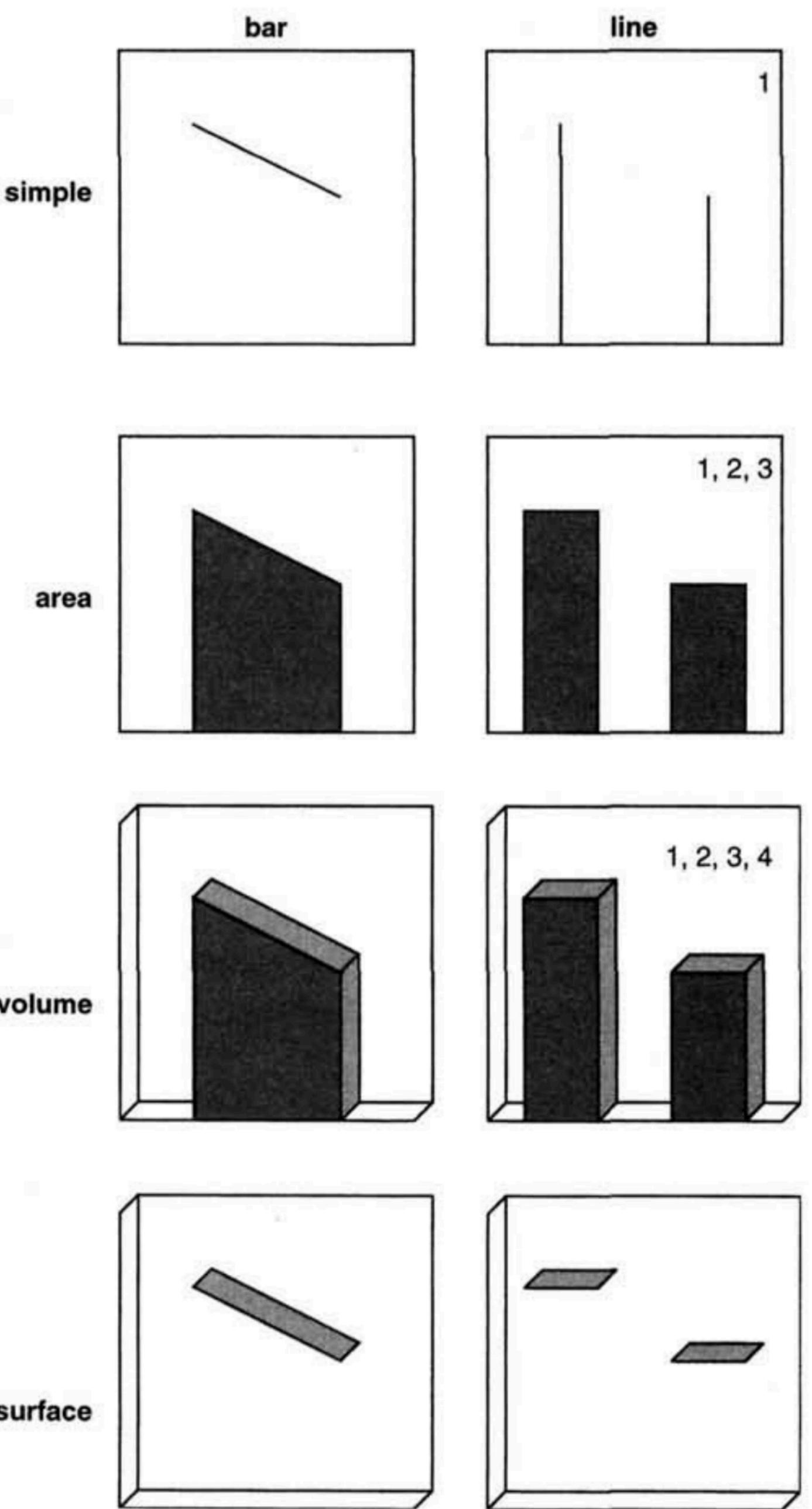
- Use treemap layouts that **avoid extreme aspect ratios** (+ squares)
- Use bar charts for low-density data, treemaps for high-density data
- Use tree maps when **comparing non-leaf nodes**
- Use **luminance** to encode secondary values

Is **3D** ever okay?

Is **3D** ever okay?

...it's not as *bad* as it's made out to be, but is
probably best avoided.

In 1998, Jeff Zacks et al. determined that **adding 3D depth perception cues** to both pie charts and bar charts **lowered accuracy**.



...but, was it because this introduced the idea of 'volume' or
because **the marks were simply extraneous?**

Ultimately, 3D cues were **not the biggest culprit—**

- Depth cues seemed to affect *immediate* judgment of values, but less so judgment from memory
- Distortions due to *neighboring elements* were more of a concern

Language driven by **science**?
Science driven by **art**?

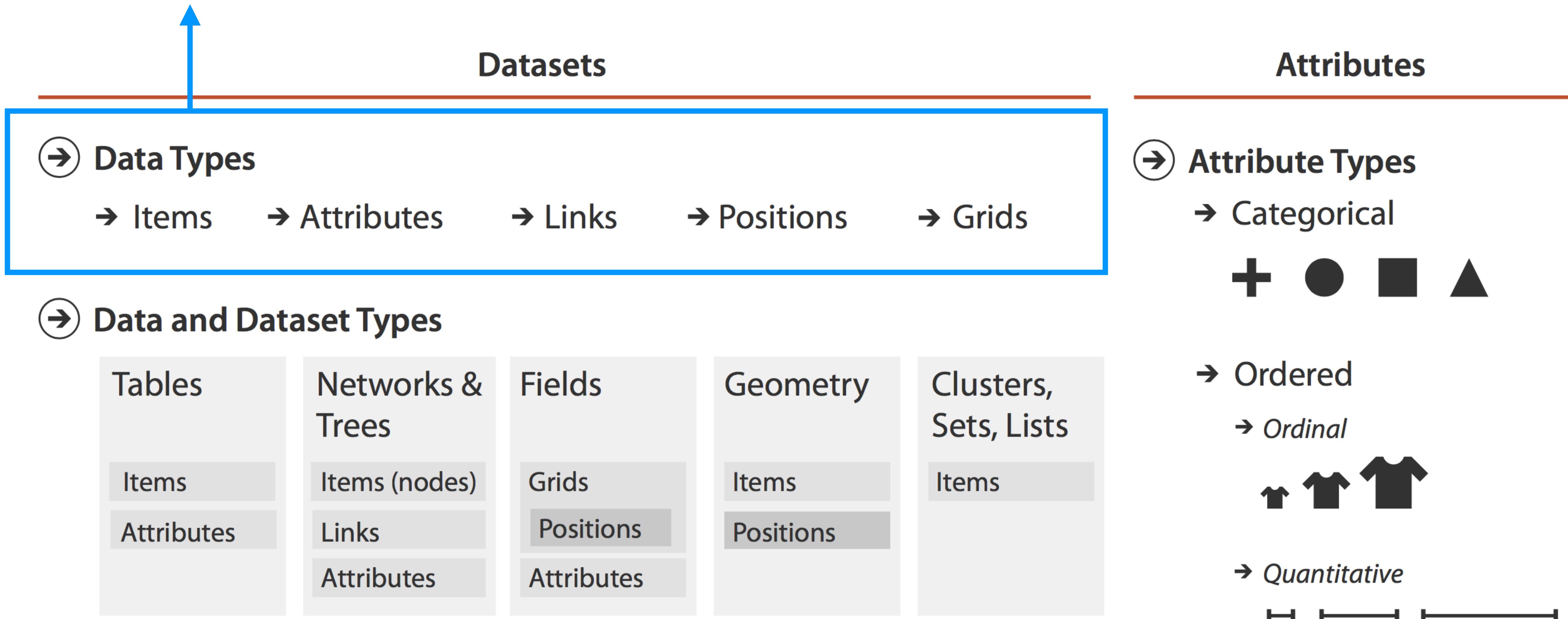
1. Fundamentals of **Graphical Perception**
2. **Data Abstraction**
3. Studio (Introduction to **Tableau**)

Data abstraction

Data types and **semantic meaning** inform the appropriate visualization methods for a given dataset.

Type refers to the *structural or mathematical* interpretation of the data.

What kind of thing is it?



Datasets

→ Data Types

→ Items → Attributes → Links → Positions → Grids

→ Data and Dataset Types

Tables

Items

Attributes

Networks &
Trees

Items (nodes)

Links

Attributes

Fields

Grids

Positions

Attributes

Geometry

Items

Positions

Clusters,
Sets, Lists

Items

Attributes

→ Attribute Types

→ Categorical



→ Ordered

→ Ordinal



→ Quantitative



How do these combine to form a larger structure?

What kinds of mathematical operations are meaningful for it?

Datasets

→ Data Types

→ Items → Attributes → Links → Positions → Grids

→ Data and Dataset Types

Tables

Items

Attributes

Networks & Trees

Items (nodes)

Links

Attributes

Fields

Grids

Positions

Attributes

Geometry

Items

Positions

Clusters, Sets, Lists

Items

Attributes

→ Attribute Types

→ Categorical



→ Ordered

→ Ordinal



→ Quantitative



...and sometimes, type is **not *immediately***
apparent.

Example — “Quantity” vs. “Code”



11222

If this value is a ***count of boxes of detergent*** —

- Type = **quantity**
- Adding two quantities together makes sense
- Informs appropriate visualization method (bar chart, line chart, etc.)

Example — “Quantity” vs. “Code”



11222

If this value is a **zip code** —

- Type = **code**
- Adding two codes together *does not* make sense
- Informs appropriate visualization method (point map, etc.)

Semantics refers to the *real-world meaning* of the data.

- Zip code?
- Detergent box count?
- Human name?
- Company name?
- Abbreviated company name?
- Fruit?
- Age?
- Day of the month?

Semantic meaning – Diameter of this tree's stump



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	tree_id	block_id	created_at	tree_dbh	stump_diam	curb_loc	status	postcode	zip_city	borocode	borough	state	latitude	longitude	
2	572319	331819	6/1/16	0	25	OnCurb	Stump	11422	Rosedale	4	Queens	New York	40.6688111	-73.728427	
3	585056	325177	6/14/16	0	10	OnCurb	Stump	11434	Jamaica	4	Queens	New York	40.6772349	-73.764664	
4	587296	330973	6/15/16	0	7	OnCurb	Stump	11422	Rosedale	4	Queens	New York	40.6598988	-73.734966	
5	572361	300947	6/1/16	0	8	OnCurb	Stump	11103	Astoria	4	Queens	New York	40.7640278	-73.91605	
6	593629	309967	6/19/16	7	0	OnCurb	Dead	11418	Richmond Hi	4	Queens	New York	40.7000675	-73.816498	
7	597109	315660	6/22/16	0	12	OnCurb	Stump	11357	Whitestone	4	Queens	New York	40.7901525	-73.796495	
8	580336	326472	6/9/16	2	0	OnCurb	Dead	11434	Jamaica	4	Queens	New York	40.6900457	-73.774142	
9	597459	315730	6/22/16	0	27	OnCurb	Stump	11357	Whitestone	4	Queens	New York	40.7911871	-73.79757	
10	582323	228767	6/11/16	0	4	OnCurb	Stump	11218	Brooklyn	3	Brooklyn	New York	40.6382914	-73.980829	
11	575903	325006	6/6/16	0	26	OnCurb	Stump	11413	Springfield G	4	Queens	New York	40.6749404	-73.759078	
12	576661	324444	6/6/16	0	37	OnCurb	Stump	11413	Springfield G	4	Queens	New York	40.6693284	-73.762856	
13	576084	324623	6/6/16	0	24	OnCurb	Stump	11413	Springfield G	4	Queens	New York	40.671571	-73.761066	
14	585579	330820	6/14/16	0	17	OnCurb	Stump	11422	Rosedale	4	Queens	New York	40.6581045	-73.730949	
15	576923	323496	6/6/16	0	13	OnCurb	Stump	11413	Springfield G	4	Queens	New York	40.6676041	-73.760091	
16	586656	408083	6/14/16	0	10	OnCurb	Stump	10314	Staten Island	5	Staten Island	New York	40.6025797	-74.163348	
17	581859	325776	6/10/16	4	0	OnCurb	Dead	11434	Jamaica	4	Queens	New York	40.6842424	-73.782754	
18	593623	309967	6/19/16	6	0	OnCurb	Dead	11418	Richmond Hi	4	Queens	New York	40.6995605	-73.816215	
19	572986	326645	6/2/16	4	0	OnCurb	Alive	11412	Saint Albans	4	Queens	New York	40.6907011	-73.760654	
20	588177	330530	6/15/16	0	12	OnCurb	Stump	11422	Rosedale	4	Queens	New York	40.6522019	-73.729556	
21	591515	336208	6/17/16	0	23	OnCurb	Stump	11004	Glen Oaks	4	Queens	New York	40.7426455	-73.705278	
22	577831	324407	6/7/16	0	5	OnCurb	Stump	11434	Jamaica	4	Queens	New York	40.6680682	-73.76812	
23	591020	335919	6/17/16	0	18	OnCurb	Stump	11001	Floral Park	4	Queens	New York	40.7380018	-73.701568	
24	597255	335427	6/22/16	0	17	OnCurb	Stump	11426	Bellerose	4	Queens	New York	40.732777	-73.729001	
25	576421	601868	6/6/16	0	13	OnCurb	Stump	11413	Springfield G	4	Queens	New York	40.6736951	-73.753955	
26	580898	324483	6/10/16	0	17	OnCurb	Stump	11434	Jamaica	4	Queens	New York	40.6679339	-73.77592	
27	577425	324545	6/7/16	0	5	OnCurb	Stump	11434	Jamaica	4	Queens	New York	40.6707688	-73.772322	
28	588107	333868	6/15/16	0	30	OnCurb	Stump	11429	Queens Villa	4	Queens	New York	40.7125499	-73.739211	
29	589300	413339	6/16/16	0	12	OnCurb	Stump	10312	Staten Island	5	Staten Island	New York	40.5539115	-74.187089	
30	585413	325432	6/14/16	0	37	OnCurb	Stump	11434	Jamaica	4	Queens	New York	40.6811769	-73.770083	
31	577870	409952	6/7/16	6	0	OnCurb	Dead	10307	Staten Island	5	Staten Island	New York	40.5092941	-74.243033	
32	584826	400169	6/13/16	0	25	OnCurb	Stump	10314	Staten Island	5	Staten Island	New York	40.6090531	-74.145255	
33	591978	207551	6/17/16	3	0	OnCurb	Dead	11214	Brooklyn	3	Brooklyn	New York	40.5865455	-73.985235	

In a table, the **column name** typically provides the semantic meaning of an attribute.

Semantic meaning can open up possibilities for
metaphor and **visual storytelling**.

1. Fundamentals of **Graphical Perception**
2. **Data Abstraction**
3. Studio (Introduction to **Tableau**)

[https://github.com/emilyfuhrman/
datavis_design](https://github.com/emilyfuhrman/datavis_design)

-