

Data Visualization & Design

Week 4

This week in **visualization**...

LONDON

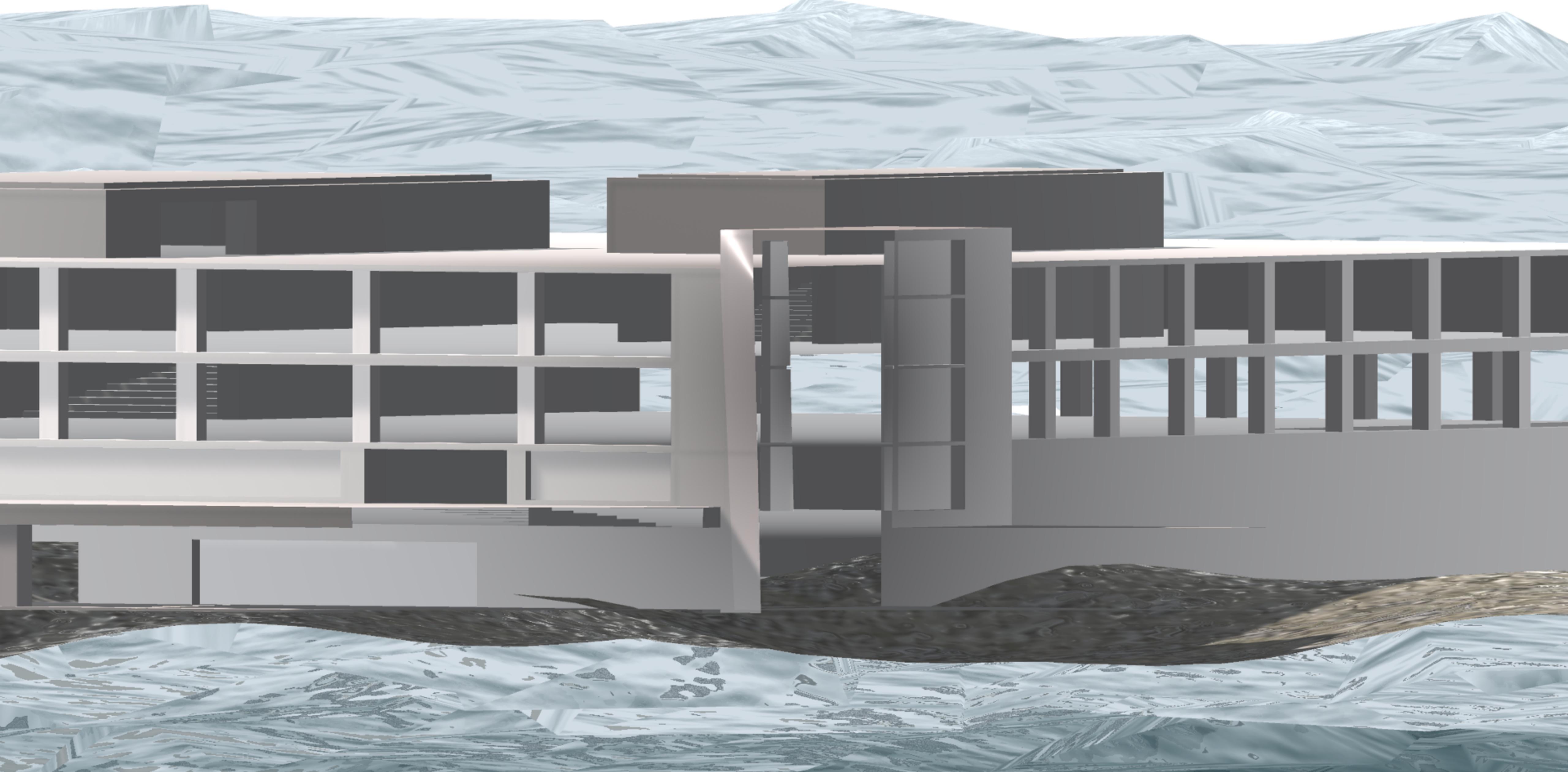
SOCIAL & FUNCTIONAL ANALYSIS

A SIMPLIFICATION OF THE COMMUNITIES & OPEN SPACE SURVEY SHOWING THE EXISTING MAIN ELEMENTS OF LONDON AND THE SUGGESTED GROUPINGS INTO PORT CITY & WEST END ARE GROUPED THE RESIDENTIAL COMMUNITIES WHICH ARE DIVIDED INTO A. THE CENTRAL COMMUNITIES AROUND THE WEST END. B. THE EAST END & SOUTH BANK COMMUNITIES WHICH HAVE A HIGH PROPORTION OF OBSOLETE PROPERTY & IN THEM ARE ADJACENT TO OR MIXED WITH INDUSTRY. C. THE SUBURBAN COMMUNITIES. THE MAJOR OPEN SPACES & INDUSTRIAL CONCENTRATIONS ARE ALSO SHOWN.

CENTRAL COMMUNITIES AROUND WEST END
MAIN INDUSTRIES WHARVES & WAREHOUSES & RAILWAYS
CENTRAL COMMUNITIES WITH HIGH PROPORTION OF OBSOLETE PROPERTY
OPEN SPACES & LARGE INSTITUTIONS WITH SUBSTANTIAL OPEN SPACE
SUBURBAN COMMUNITIES
TOWN HALLS
WATER WAYS RESERVOIRS ETC.
MAIN SHOPPING CENTRES



[Source](#)



[**Source**](#)

Integrity

A violation of integrity may be **accidental** or **intentional**.

Accidental

- Sloppy, lazy (not careful about data cleaning)
- Un-knowing (blindly trusting a default or template)

Intentional

- Removing or selecting data
- Using others' work as your own

What is the line between **storytelling** and **lying**?

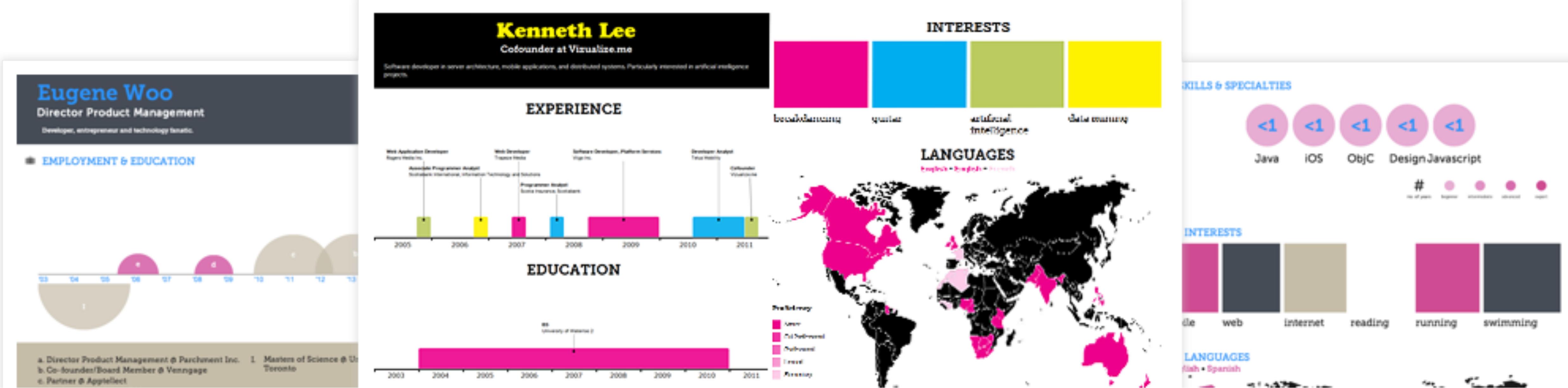
What is the line between **storytelling** and **lying**?

...what about **advocacy** and **propaganda**?

Using templates – ResumUP.com



Using templates — Visualize.me



If you use a component of a visualization generator,
credit the source and use it thoughtfully in your
context.

Guidelines

- ***Do*** credit all resources
- ***Do*** use open-source libraries and tools
- ***Do*** gather inspiration from others' projects and studies
- ***Do not*** use a prepackaged template wholesale, as though it was your own creation
- ***Do not*** copy and paste titles or charts into your projects
- ***Do not*** pass off others' work as your own

“Is this 100% my work?”

— When in doubt, **cite.**

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

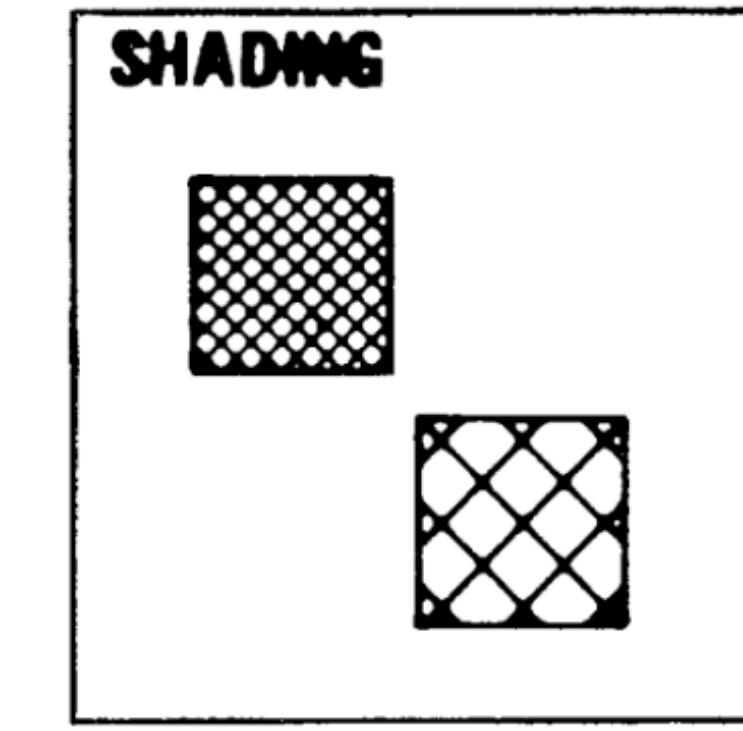
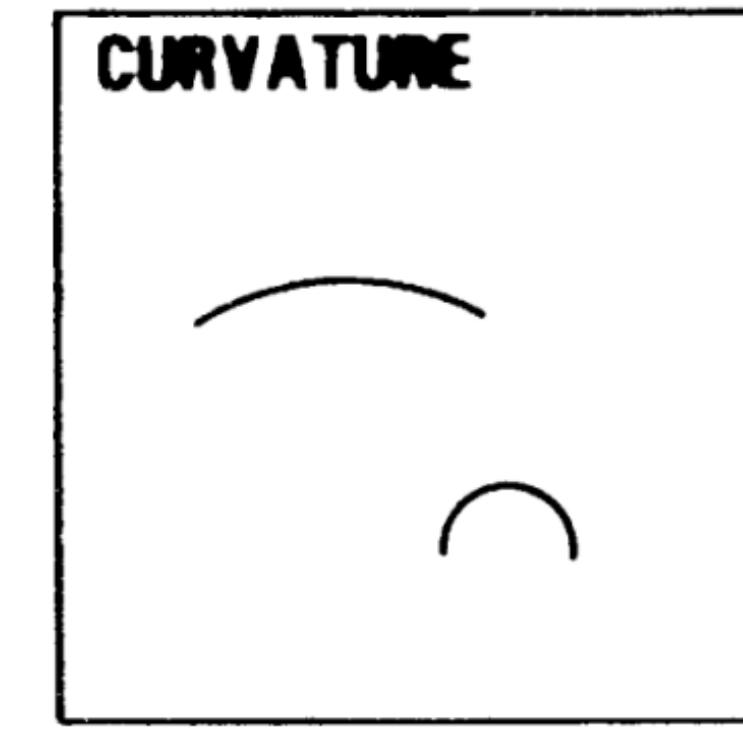
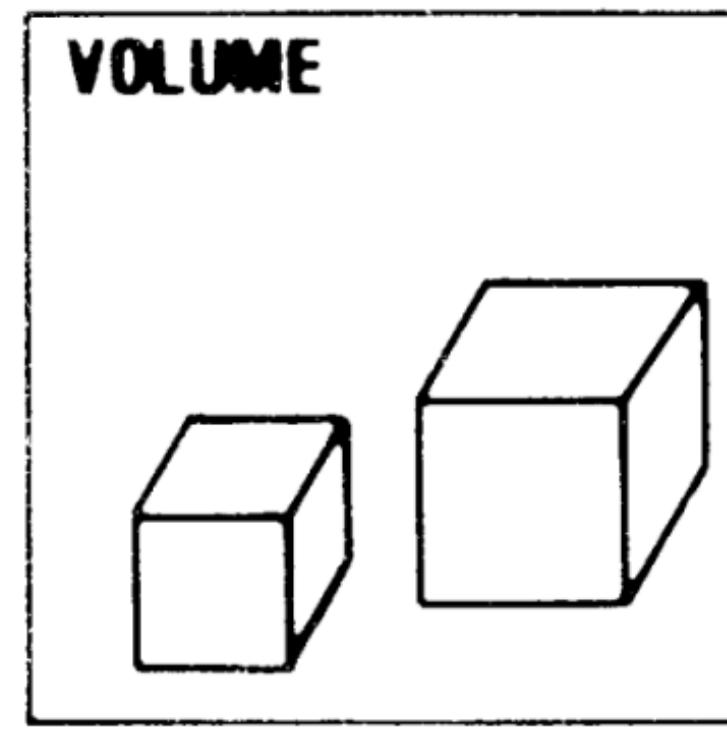
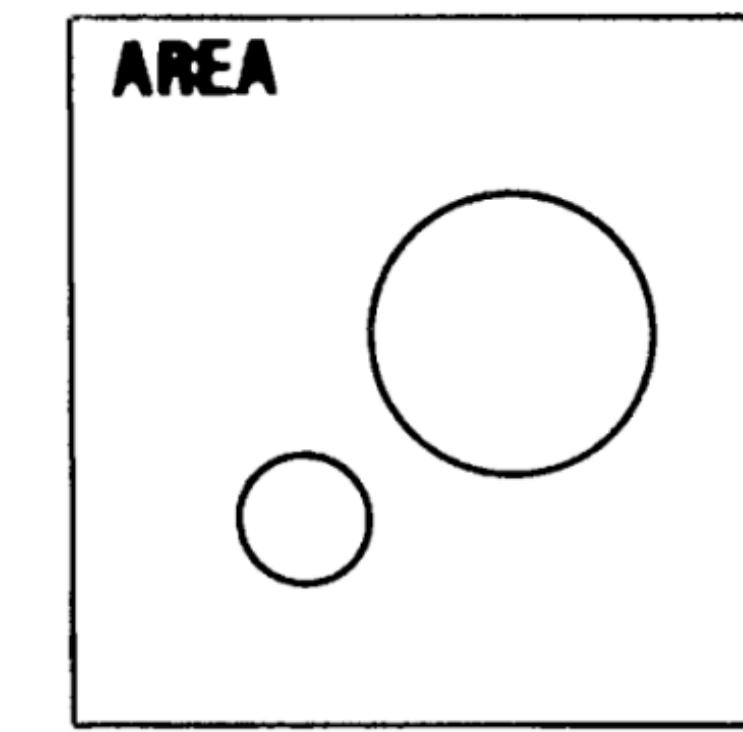
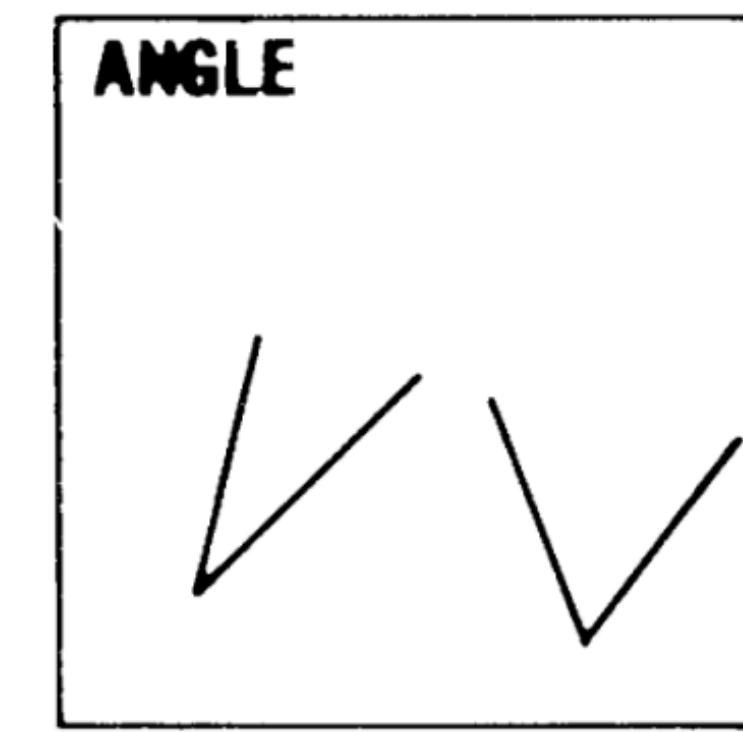
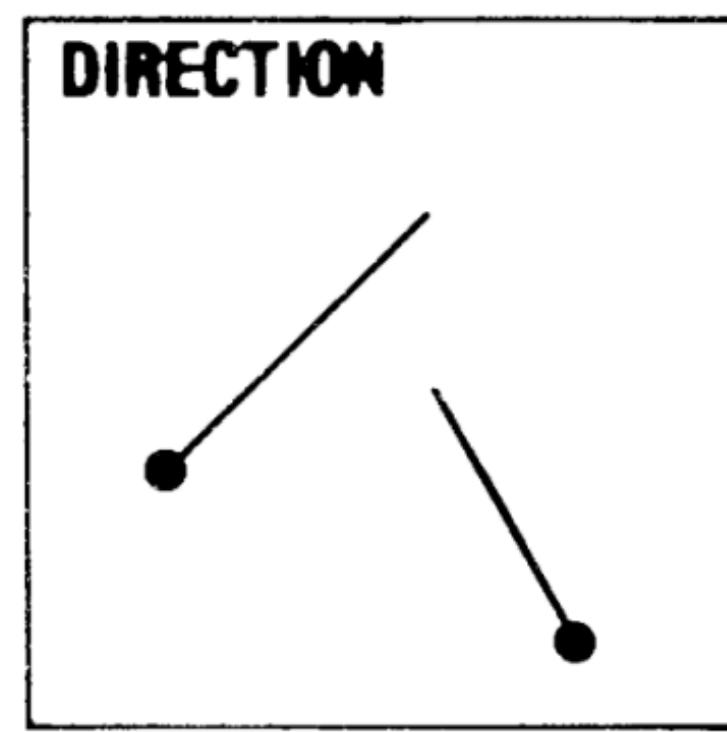
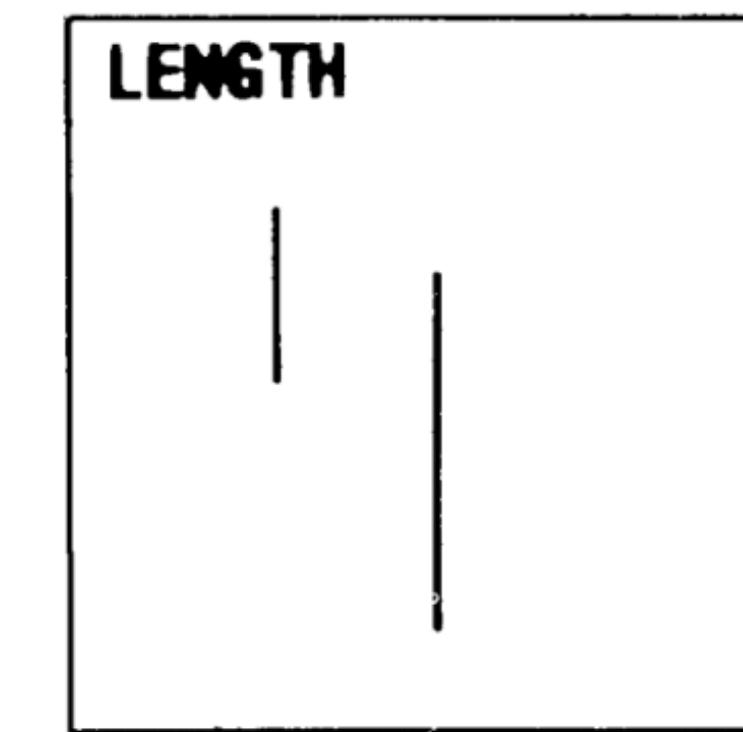
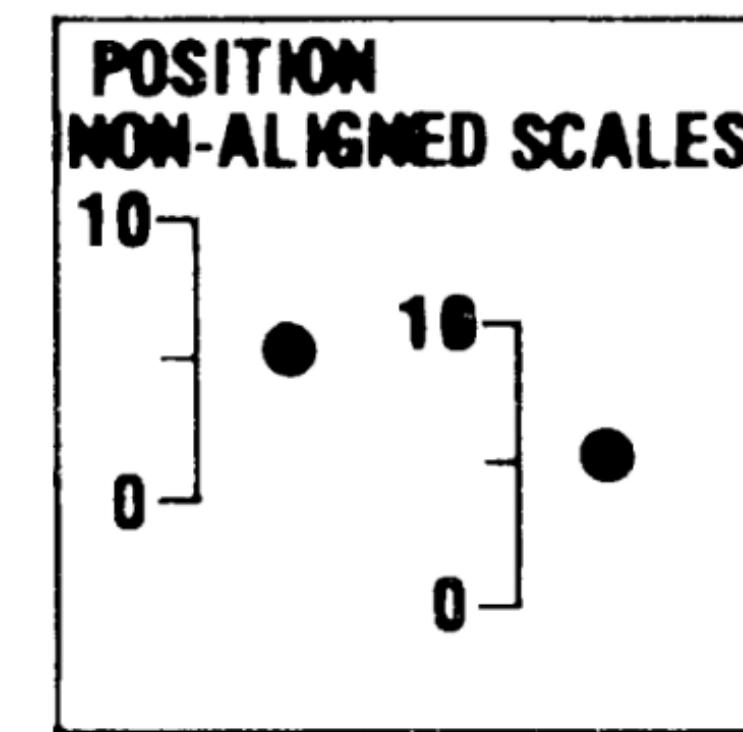
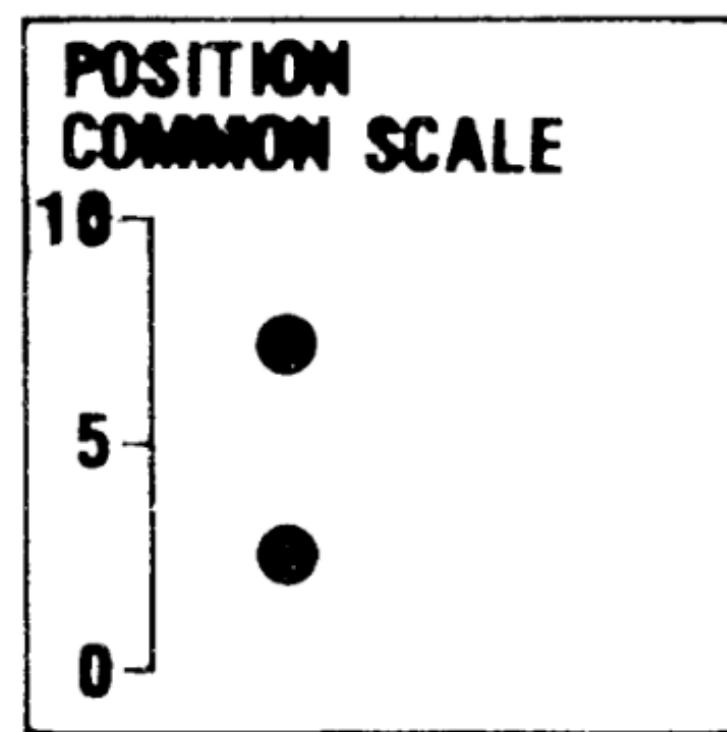
1. Fundamentals of **Graphical Perception**
2. **Data Abstraction & Dataset Types**
3. Studio (Introduction to **Tableau**)

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

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

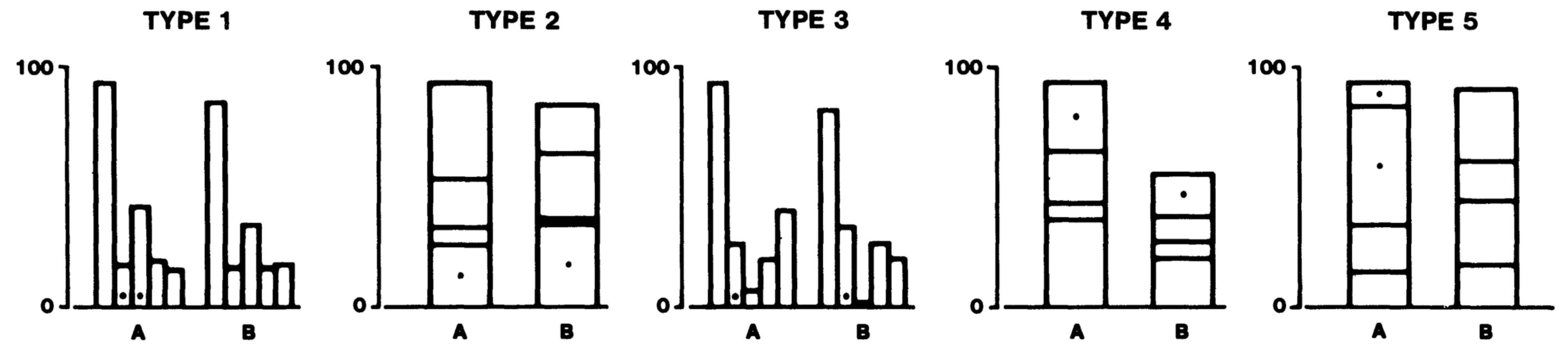
Elementary perceptual tasks are the tasks people carry out when extracting quantitative information from graphs.

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

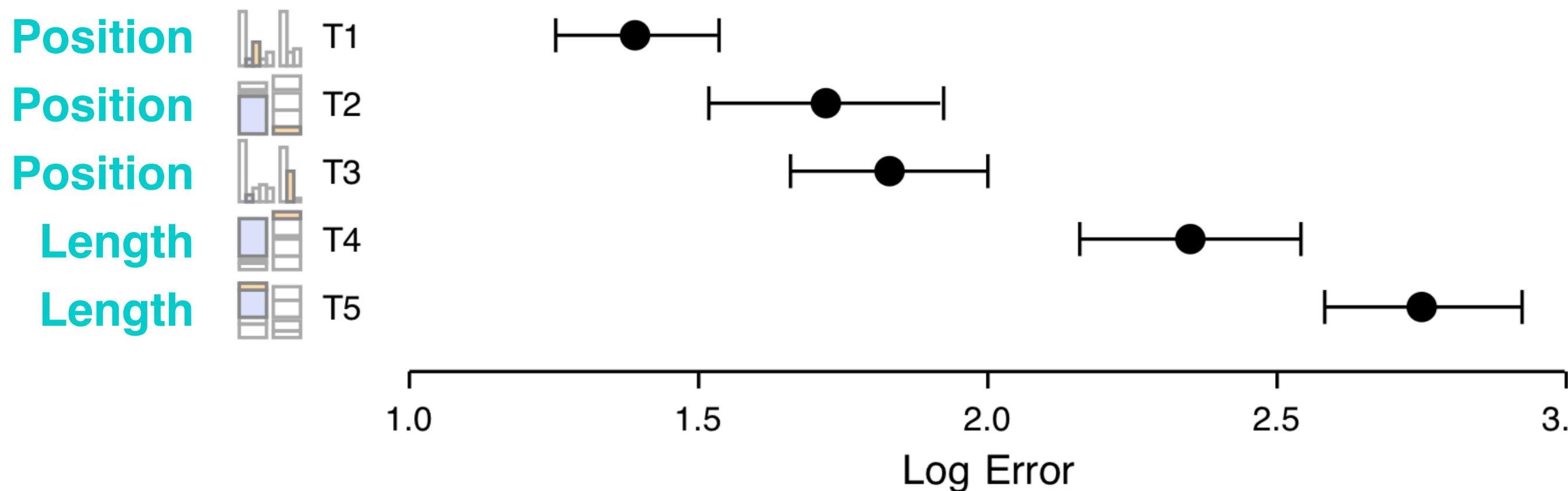
1. **Position** along a *common* scale
2. **Position** on identical but *nonaligned* scales
3. **Length, Angle & Slope** (Tie)
4. **Area**
5. **Volume, Curvature**
6. **Shading, Color Saturation**

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

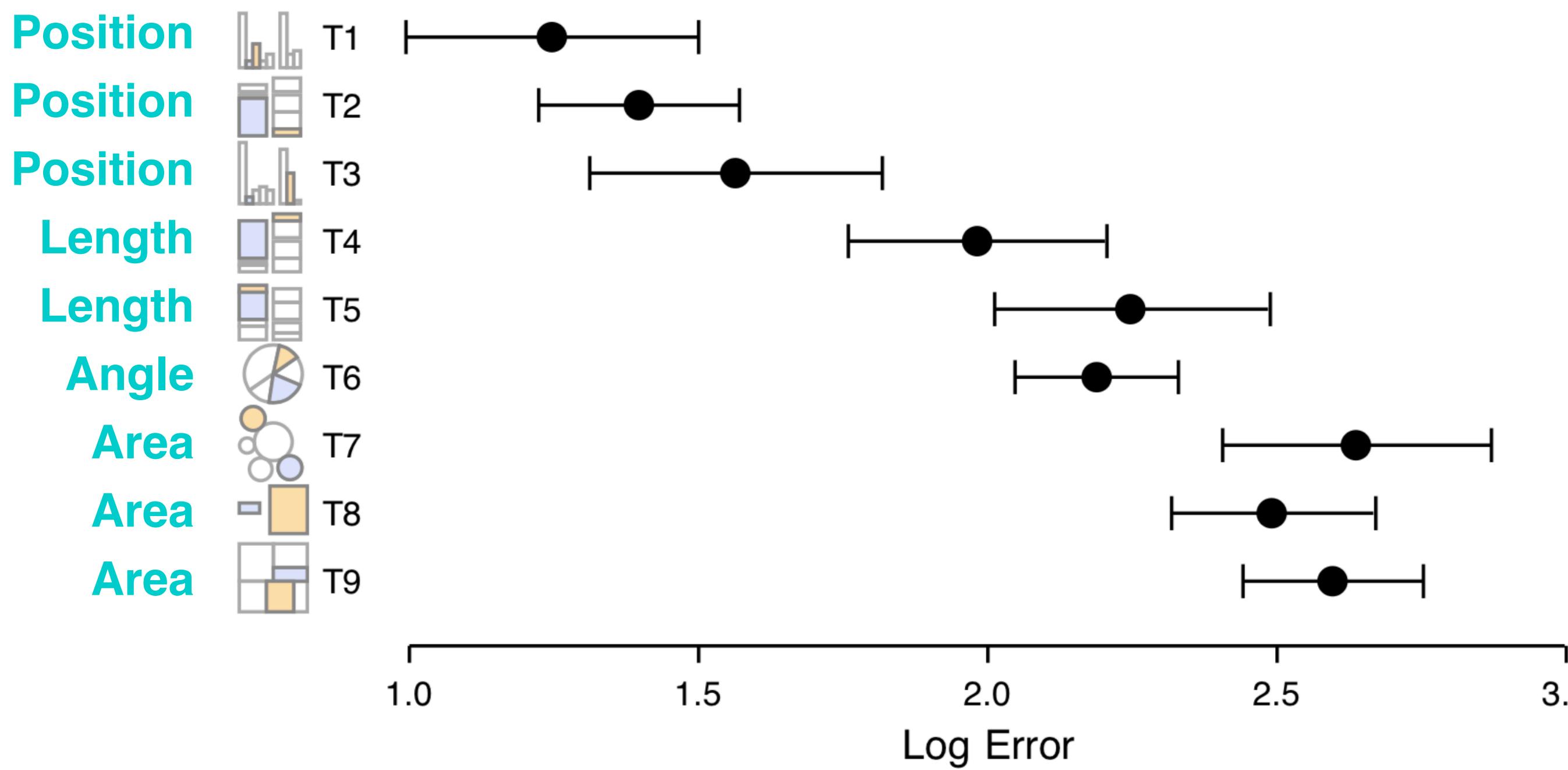
1984 was awhile ago. How do these results compare to viewers today?

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

Cleveland & McGill's Results

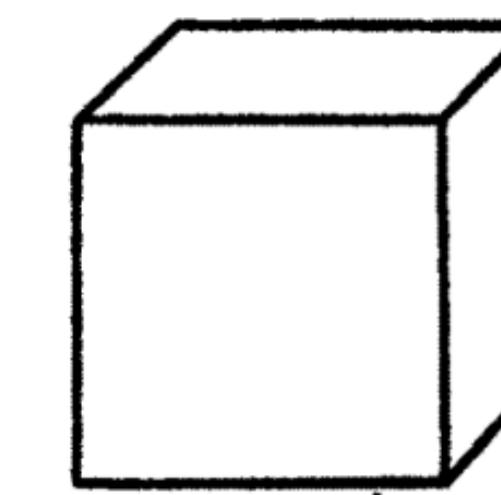
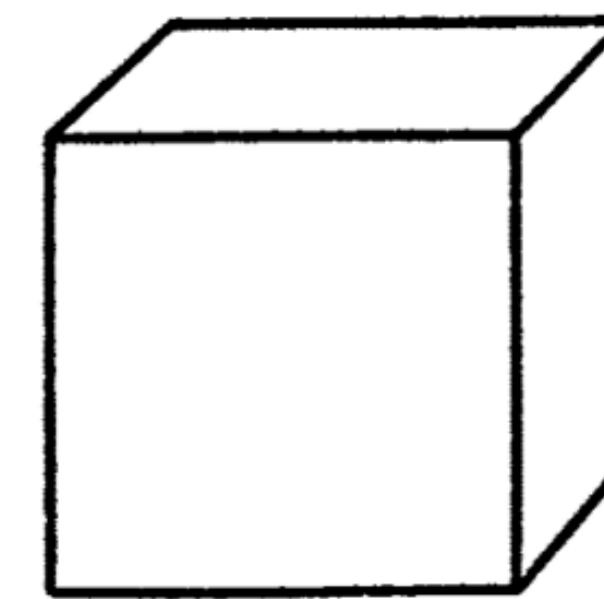
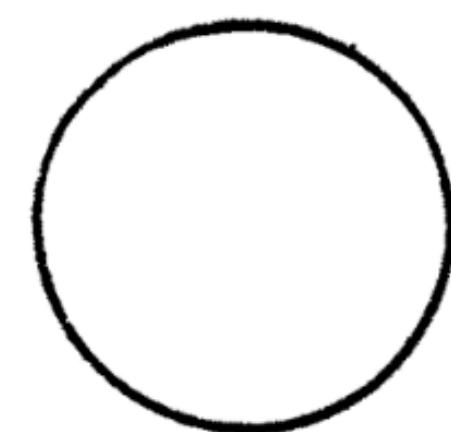
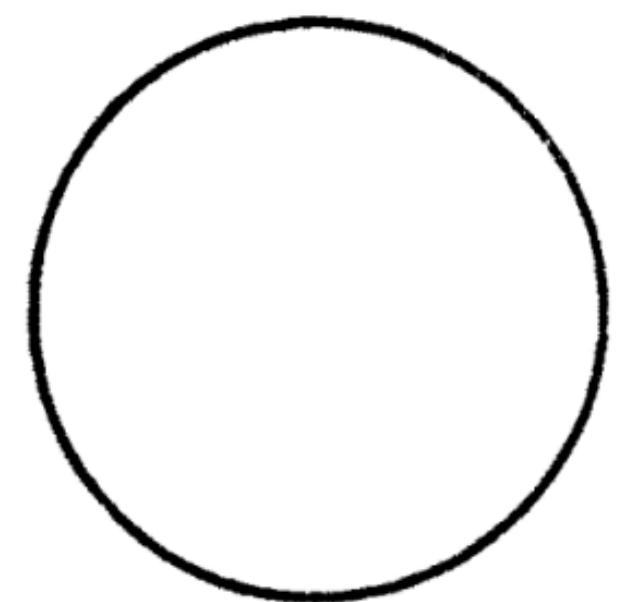
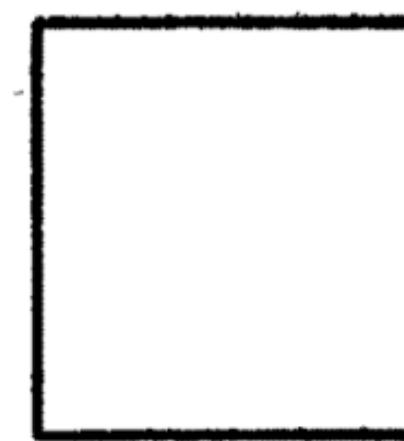
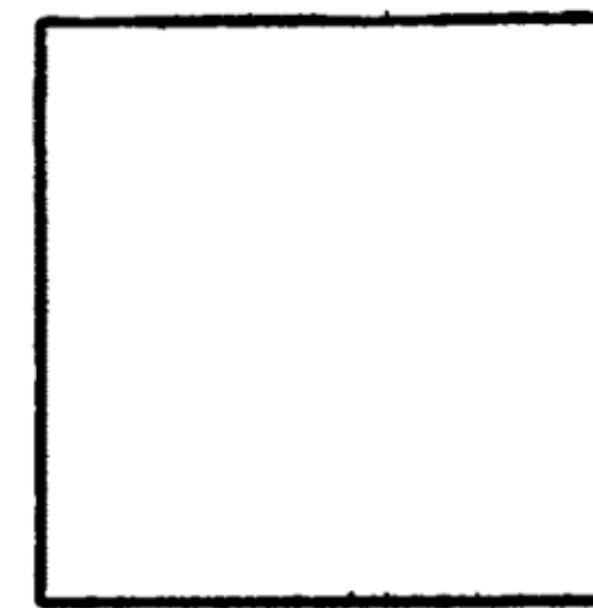


Crowdsourced Results

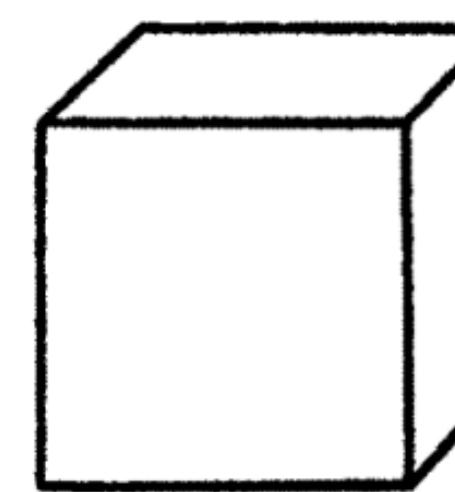
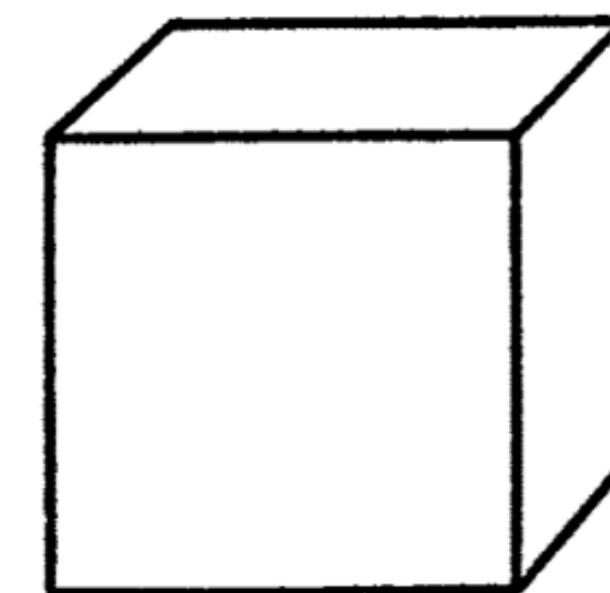
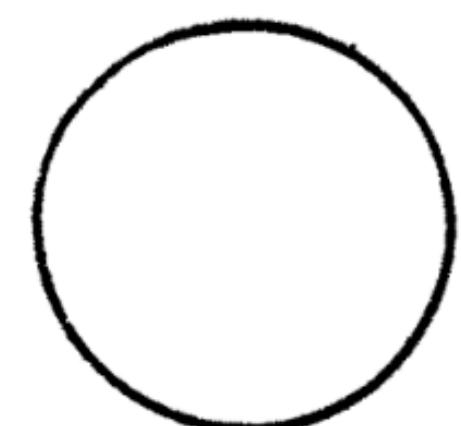
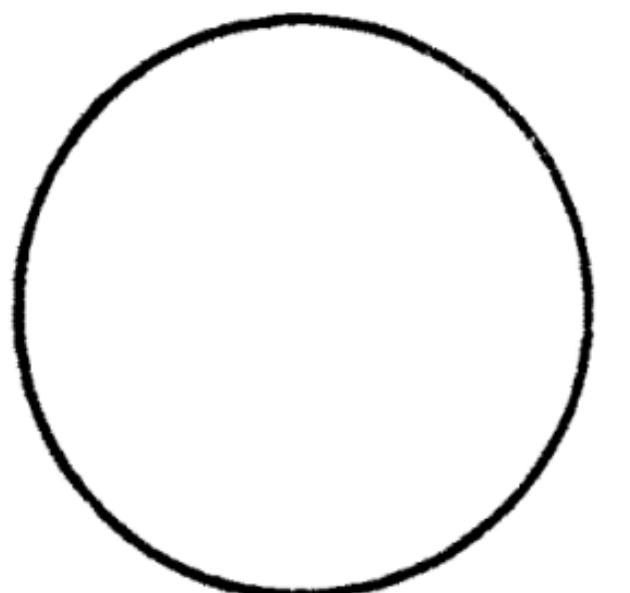
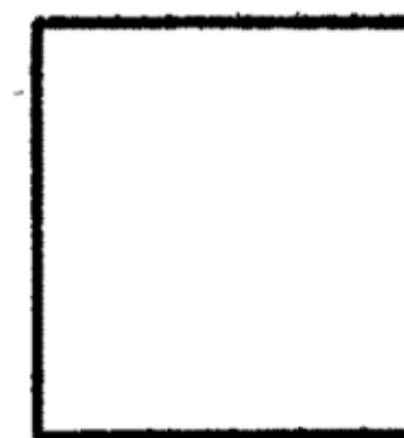
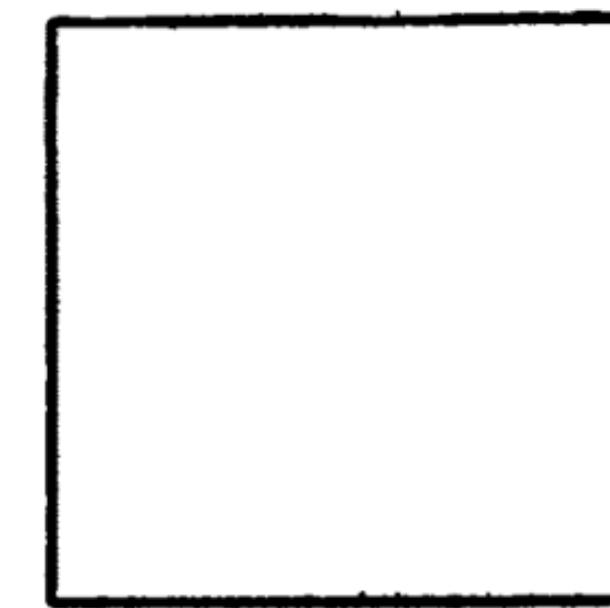


In 1932, **Frederick E. Croxton** also concluded that position is more easily judged than area or volume comparisons.

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



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



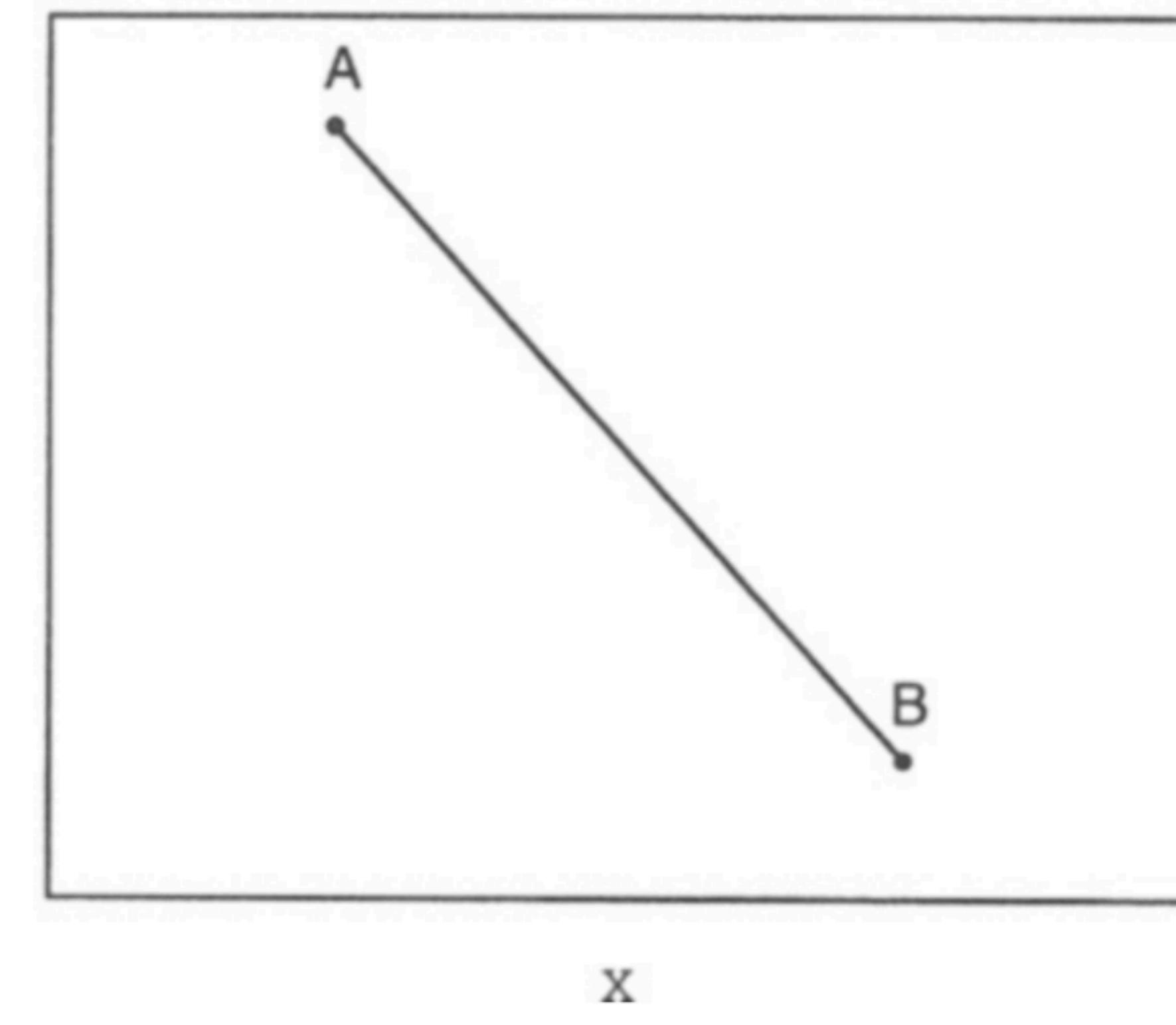
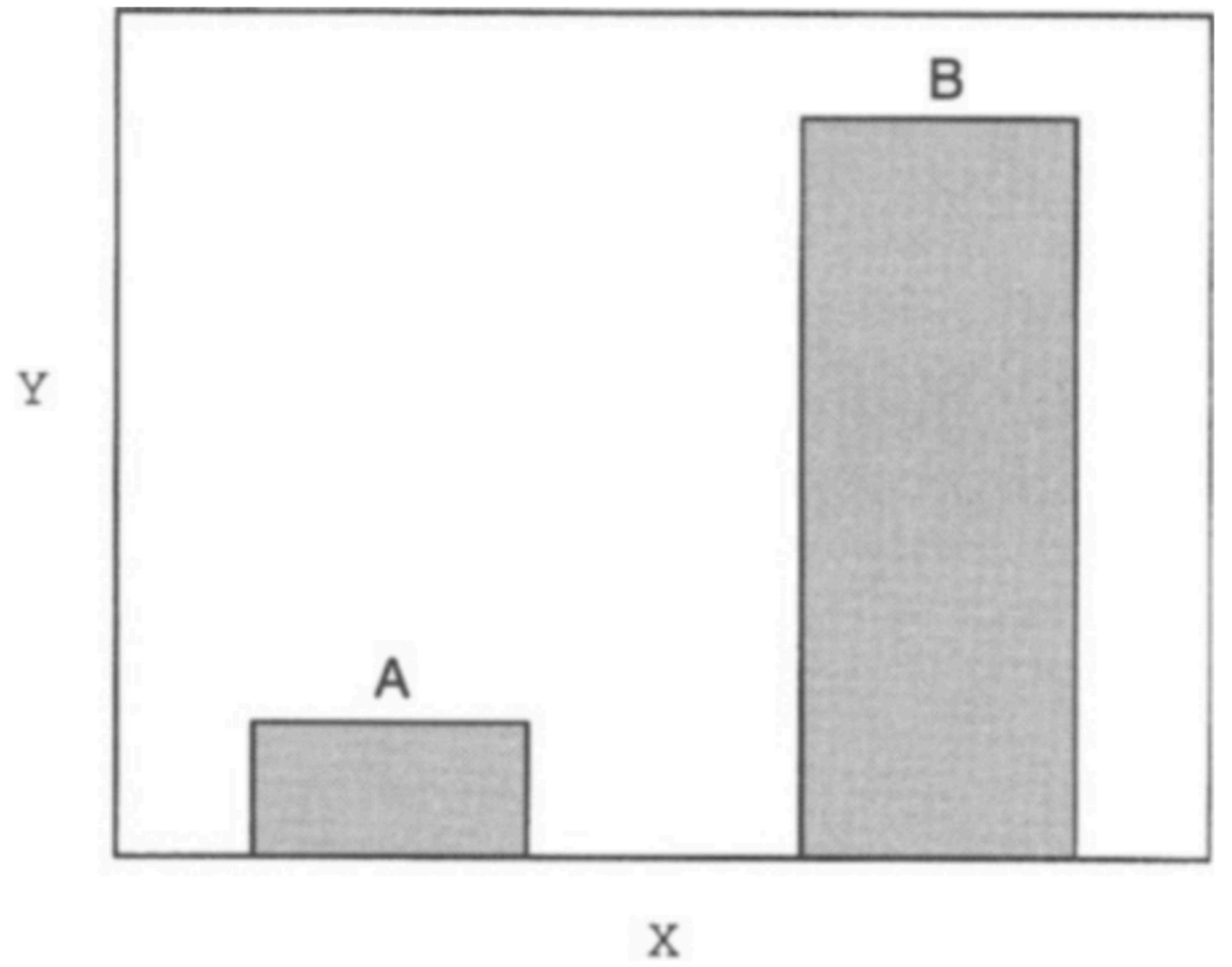
So, **always** use a bar chart.. right?

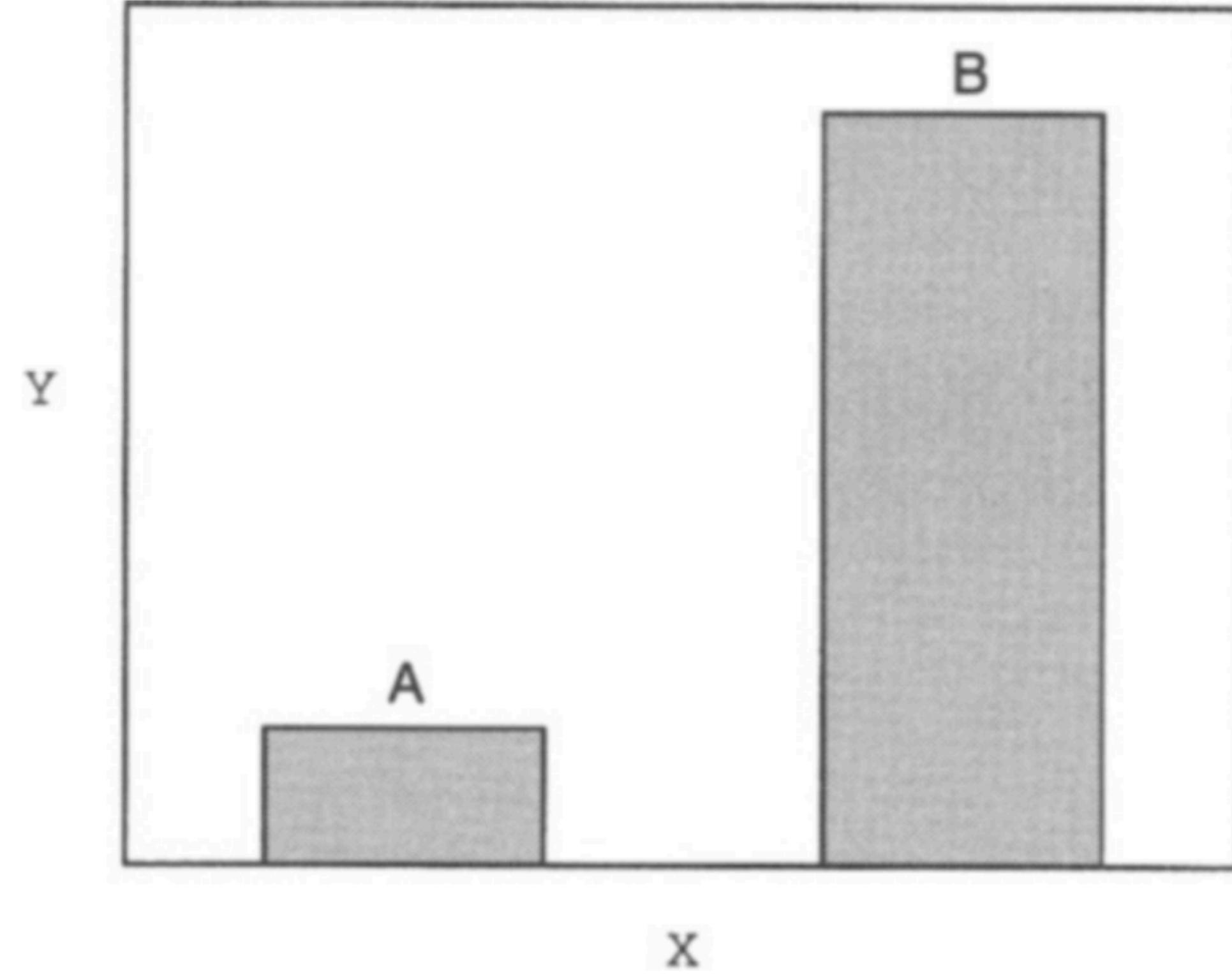
So, ~~always~~ use a bar chart.. right?

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.

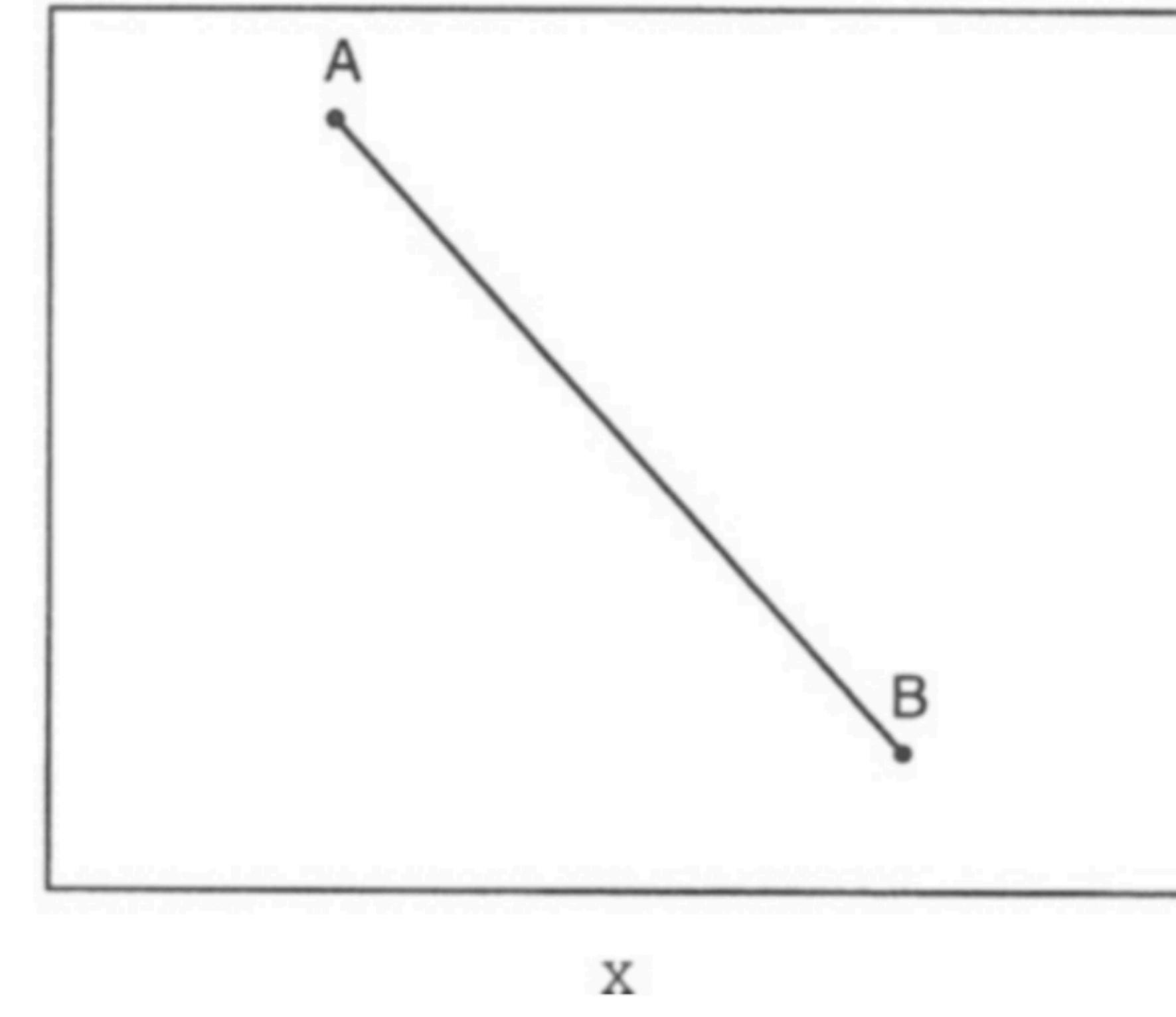
Bar-line message correspondence—

- 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’





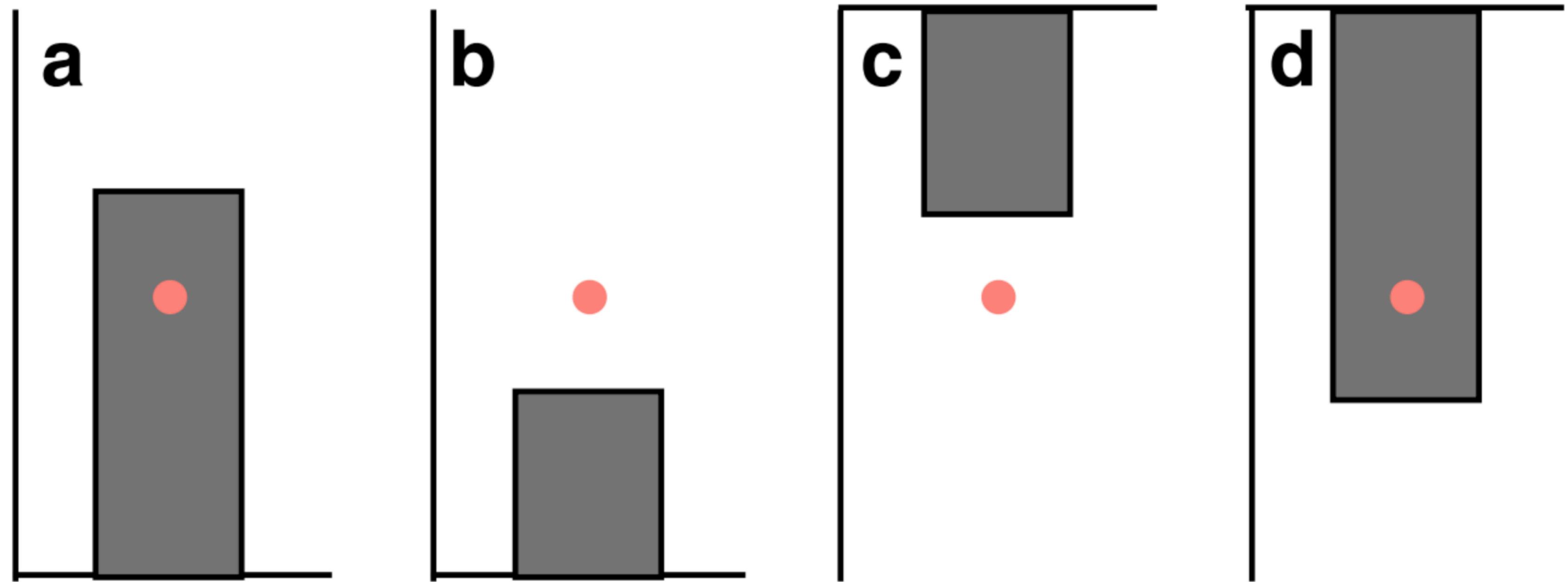
“B is higher than A.”



“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.



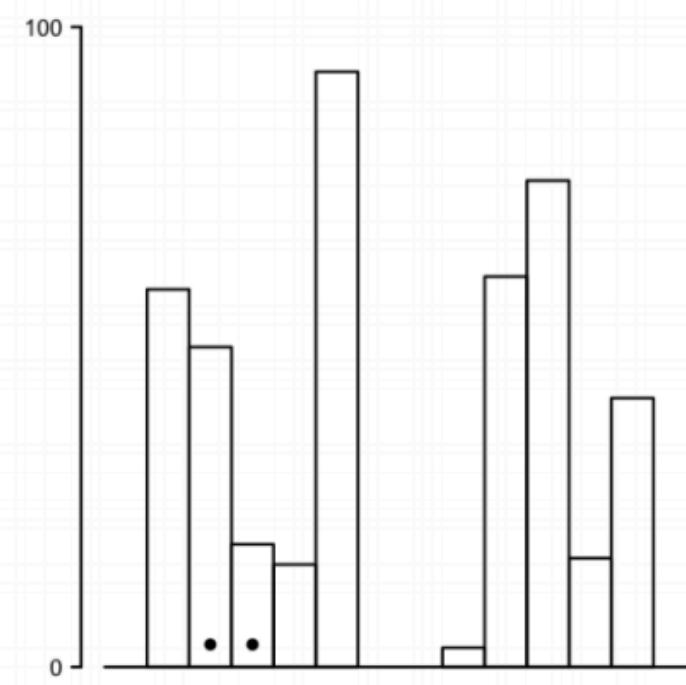
Within-the-bar bias—

- 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

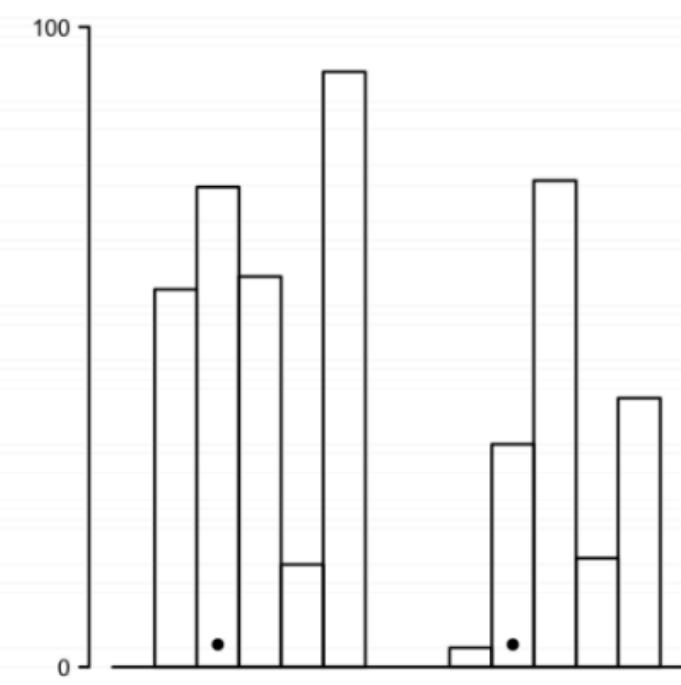
For reports of **central tendencies**, use points instead of bars.

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

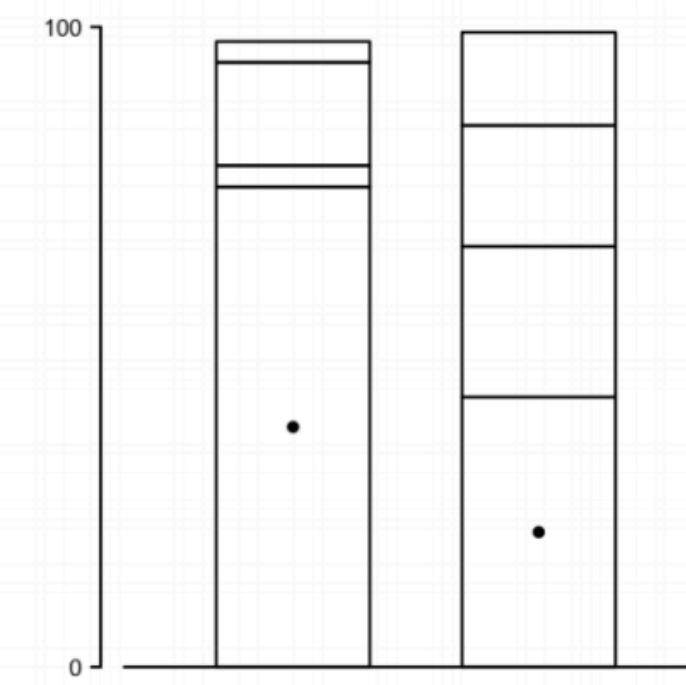
In 2014, Talbot et al. found that **point placement affected readability** in another rendition of Cleveland & McGill's study.



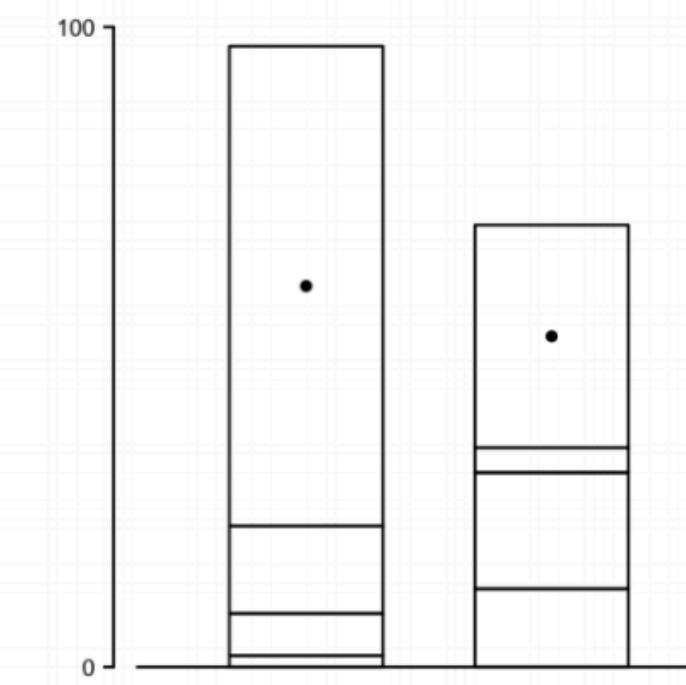
(a) Adjacent Bars
(Type 1)



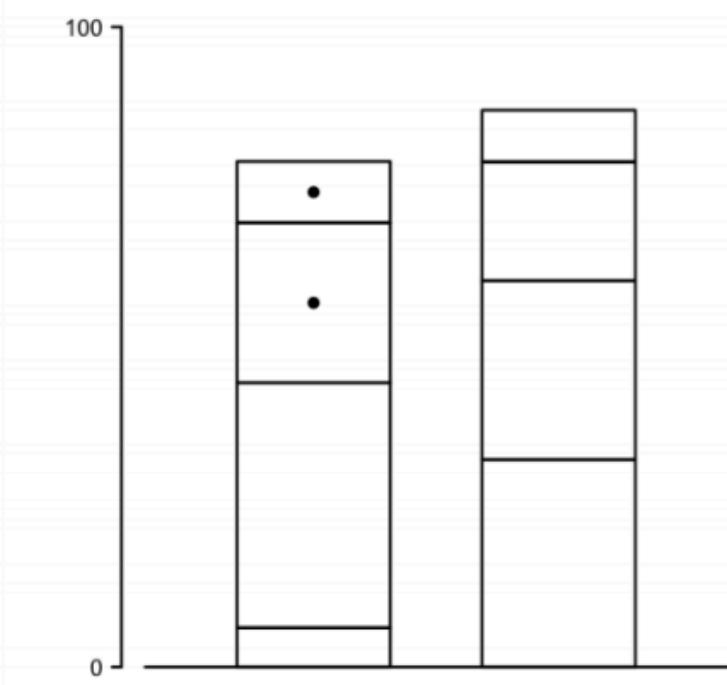
(b) Separated Bars
(Type 3)



(c) Aligned Stacked Bars
(Type 2)



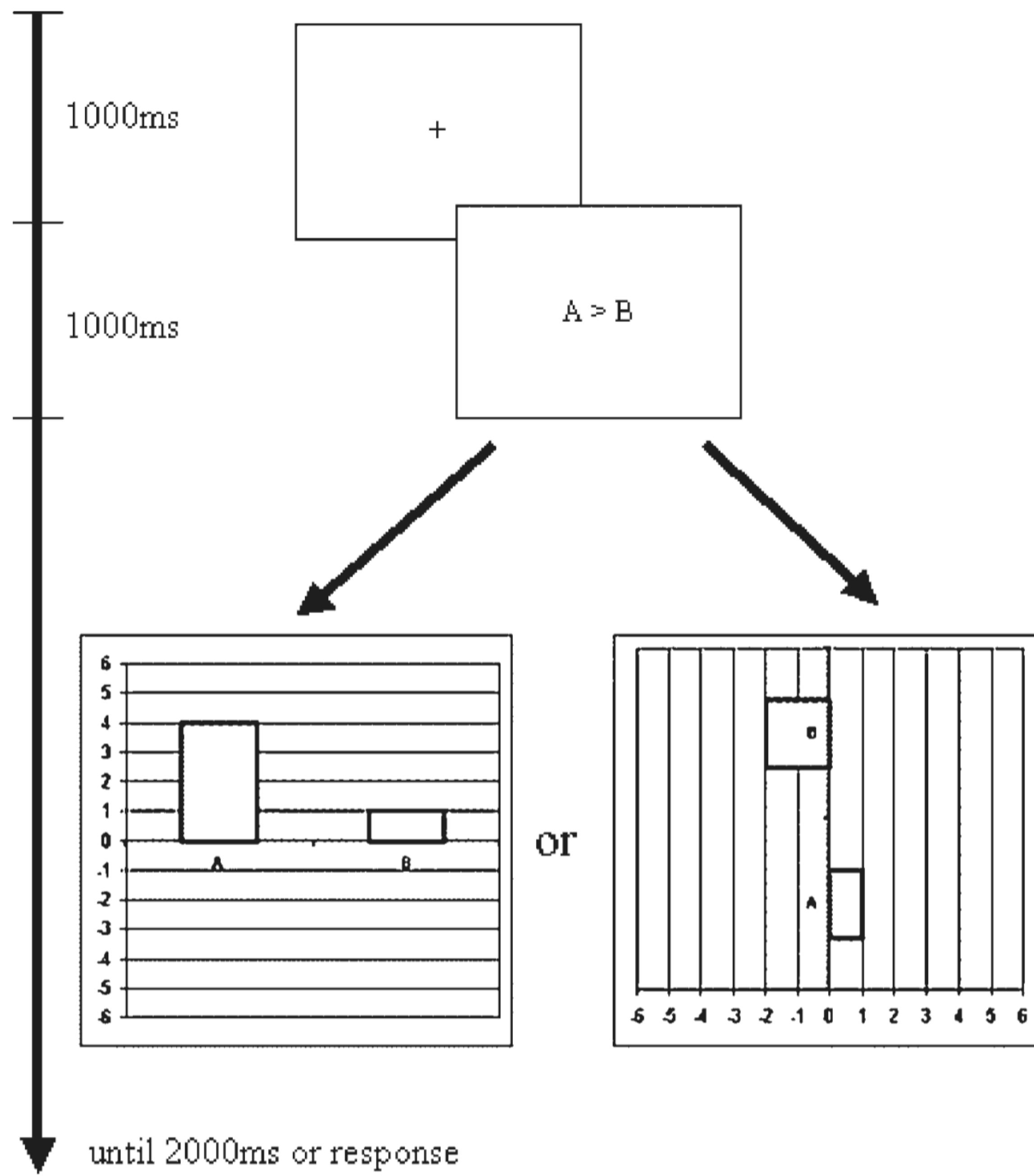
(d) Unaligned Stacked Bars
(Type 4)



(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...

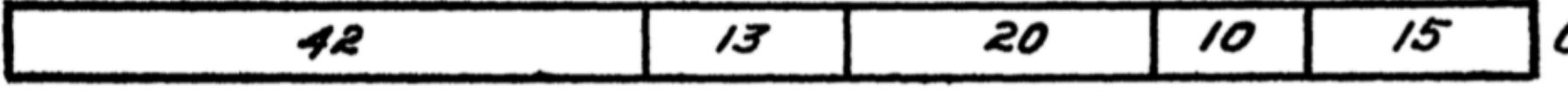
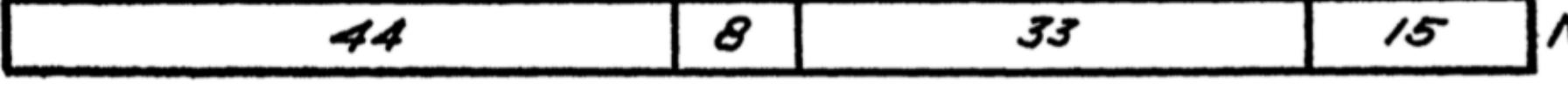
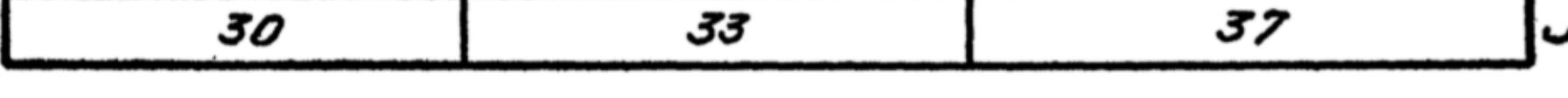
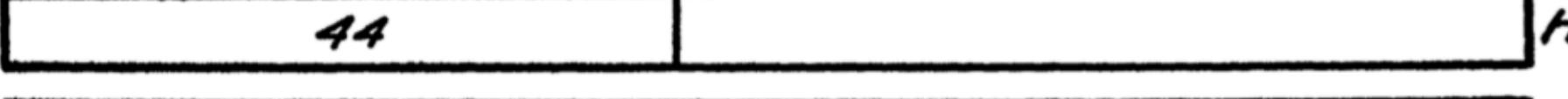
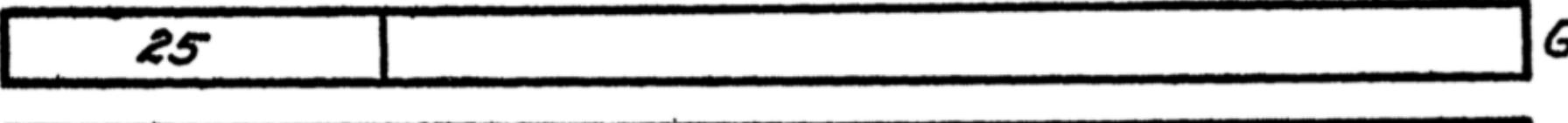
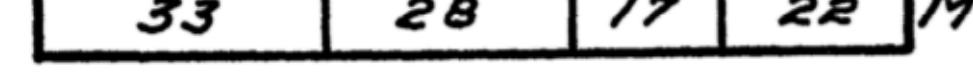
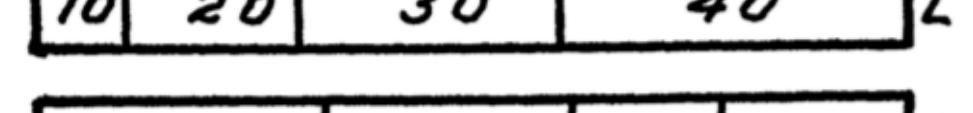
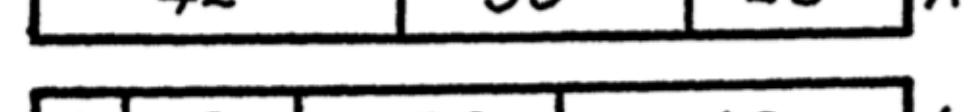
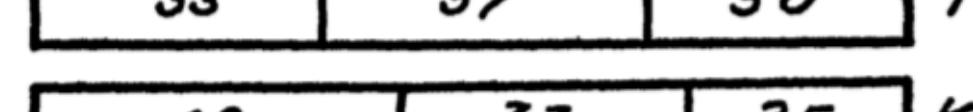
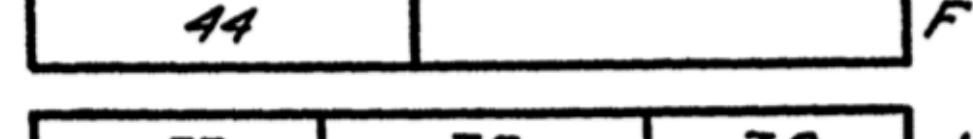
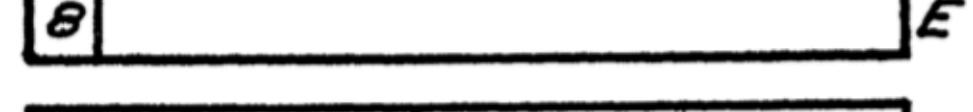
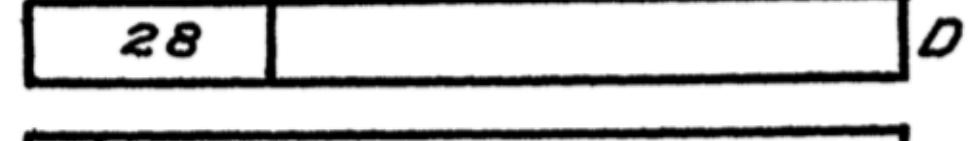
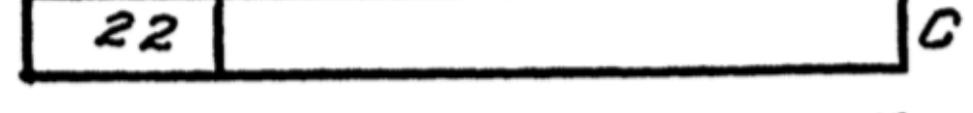
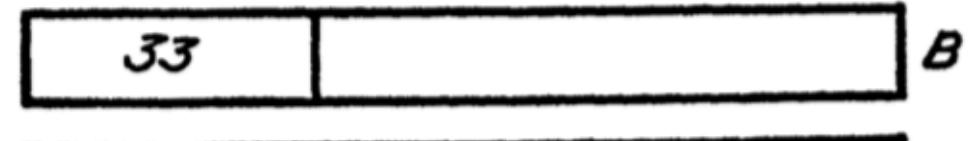
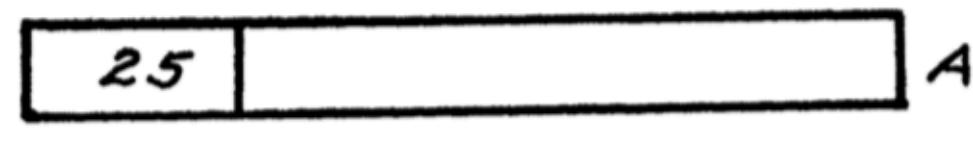
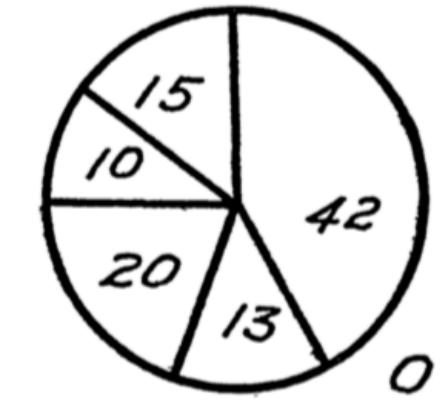
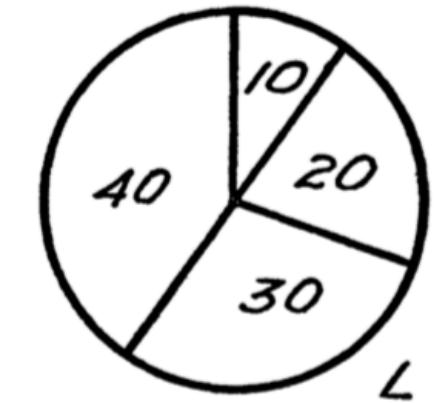
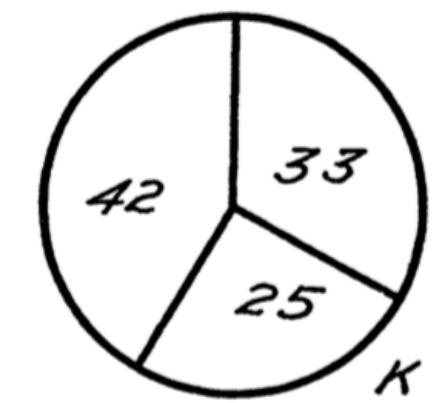
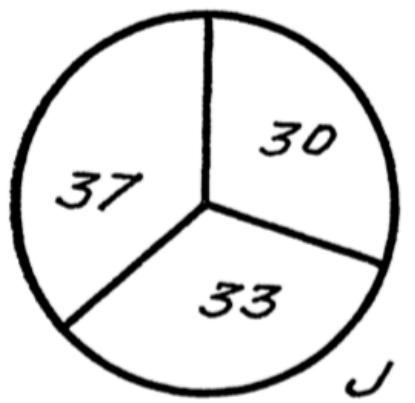
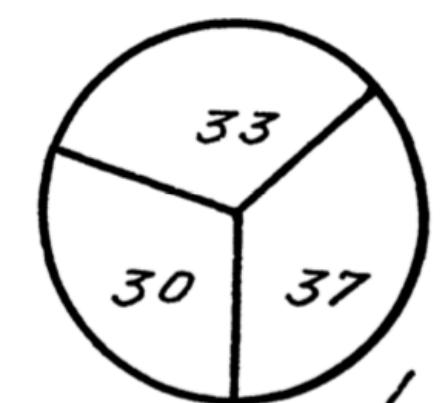
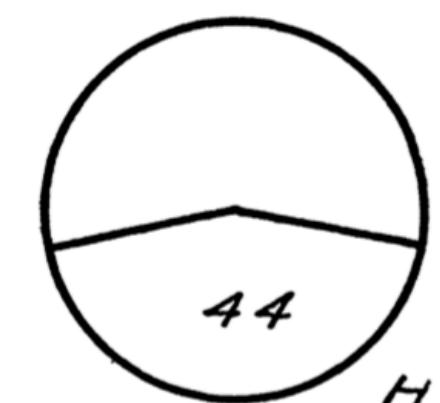
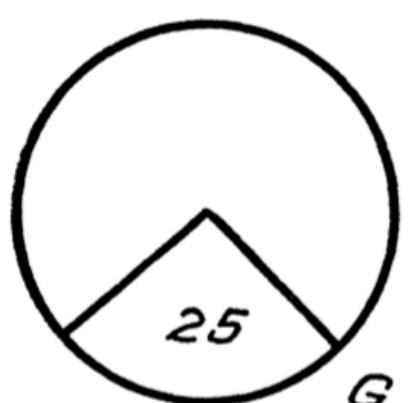
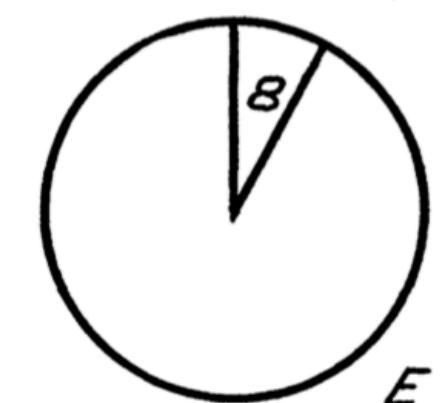
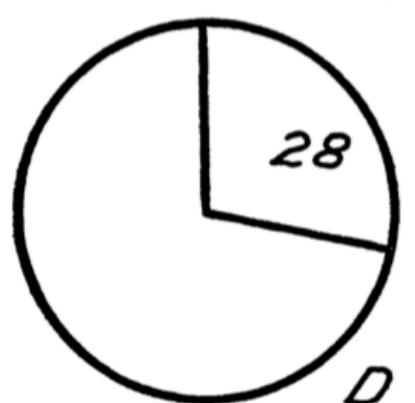
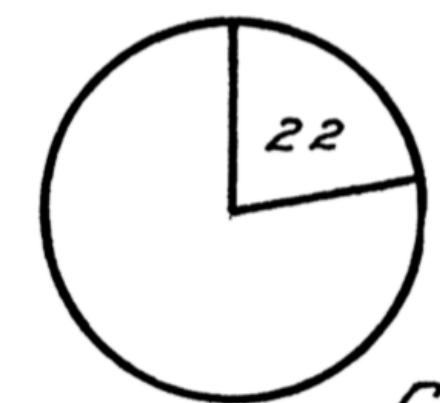
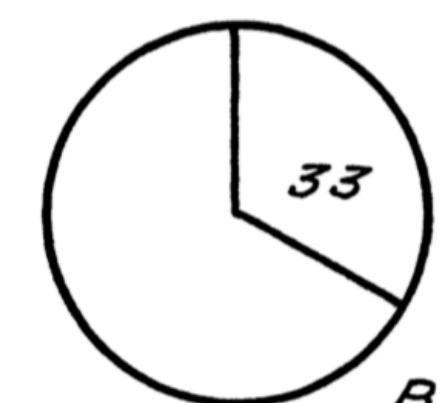
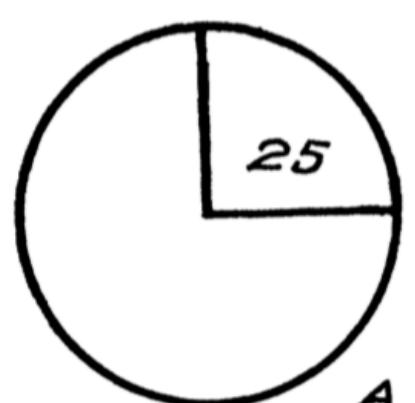


...particularly when dealing with *negative* values. Subjects associated **negativity with a downward movement** more regularly than a “left-ward” movement.

What about **pie charts**?

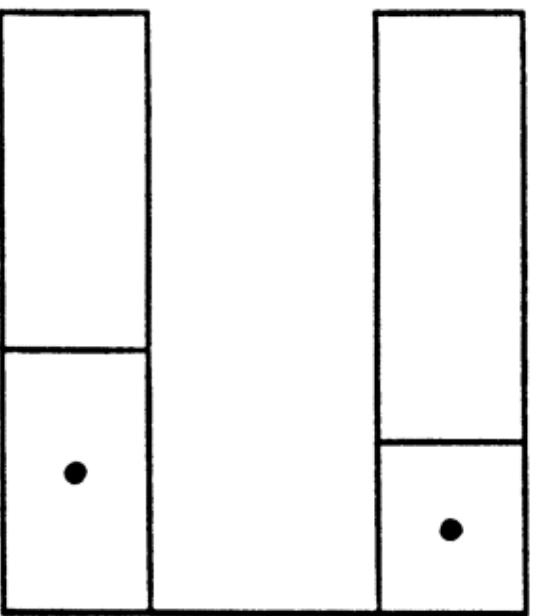
Actually, they're **OK.**

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

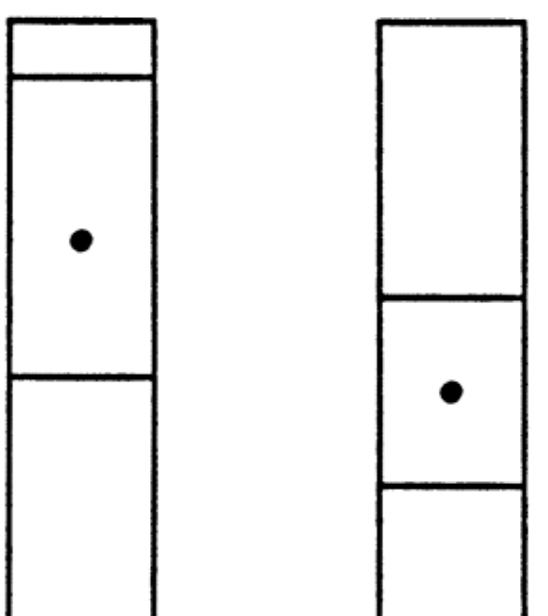


- **51%** of Ell's subjects judged the size of a wedge using the outer arc
- **25%** used areas
- **23%** used angles
- (1% didn't know how to read the pies and used chords)

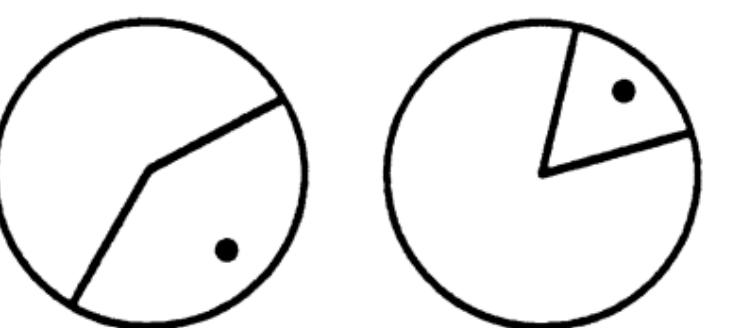
Simple bar chart



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*.

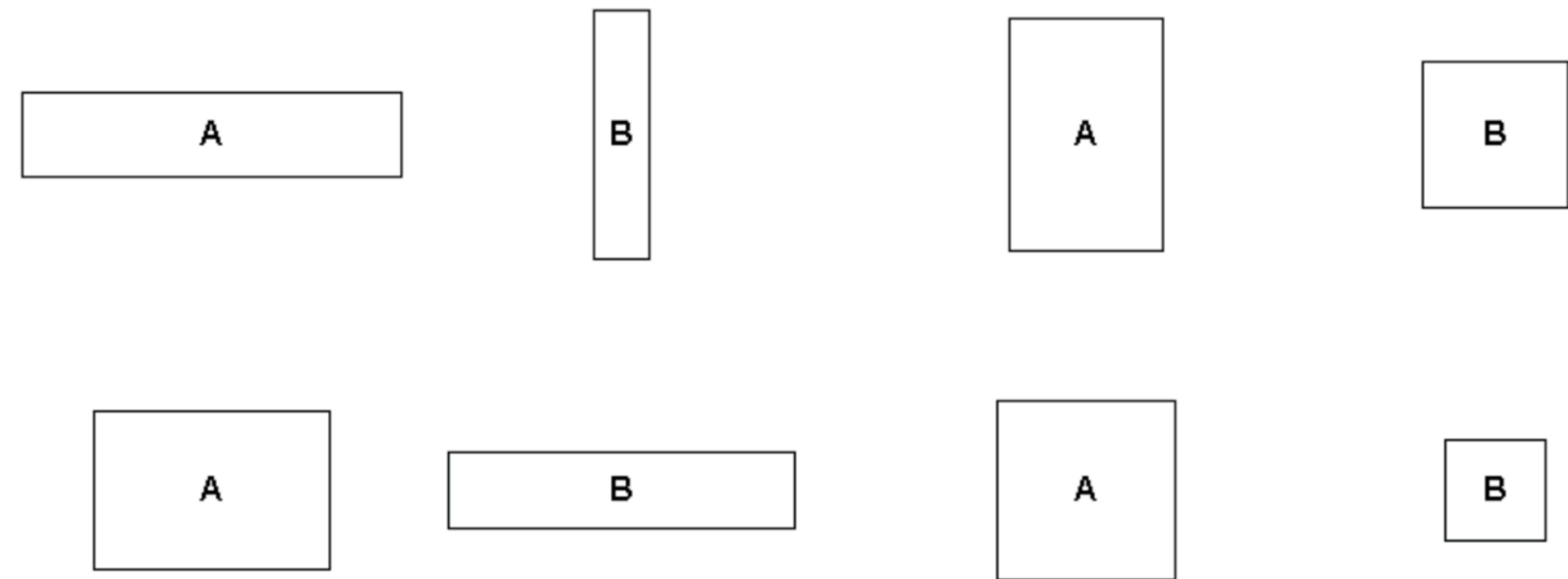
Several studies have since tackled the same issue, and have concluded that **pie charts do not perform as badly** as their reputation would suggest.

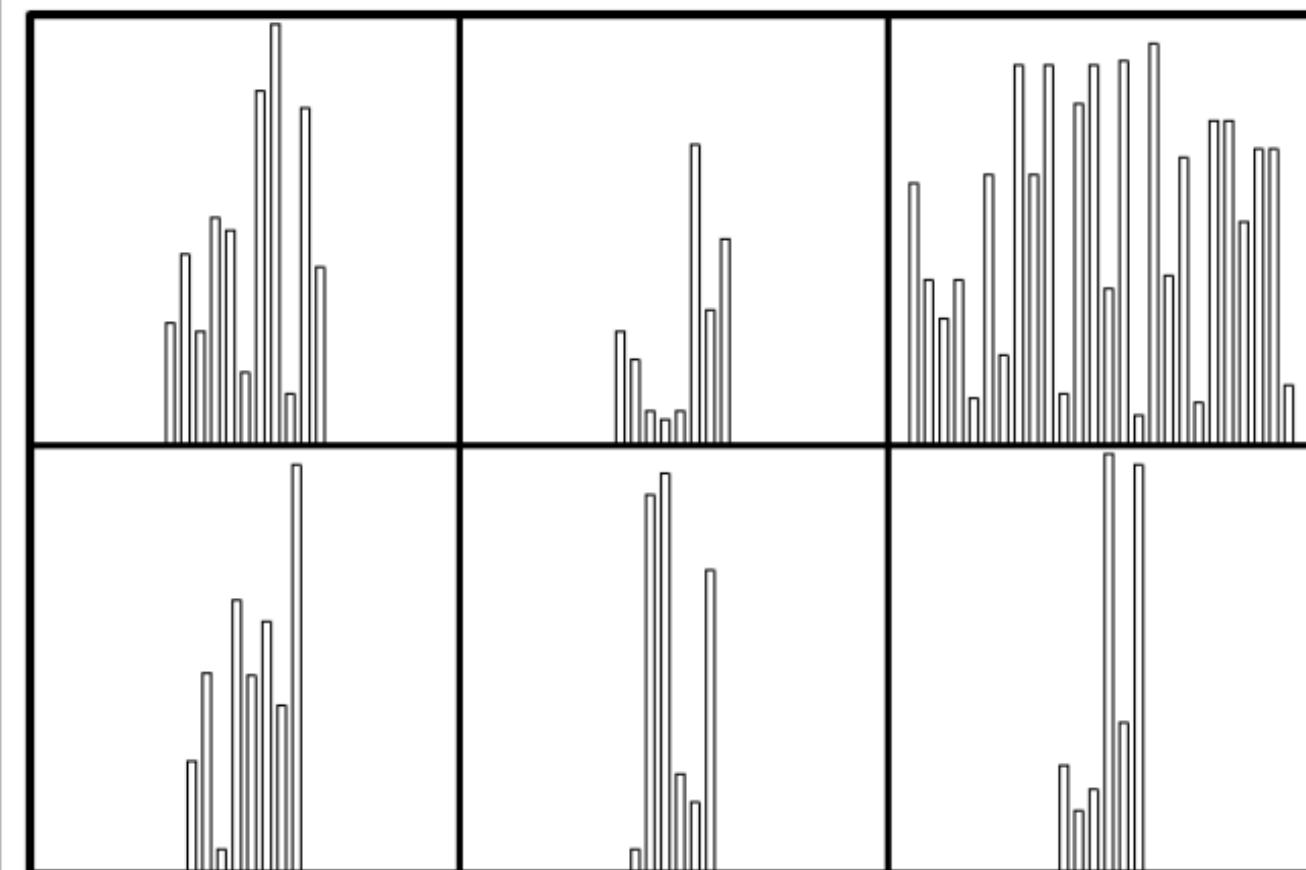
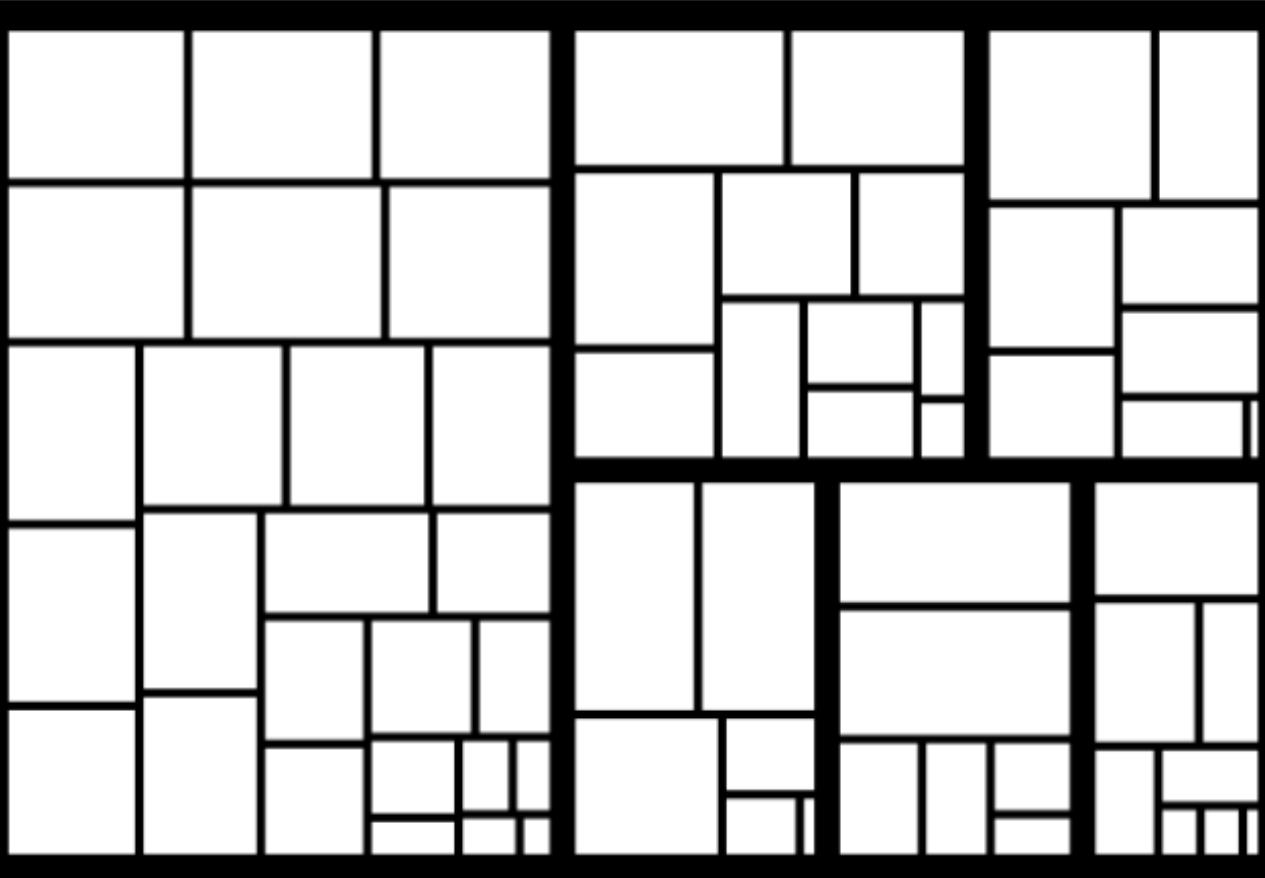
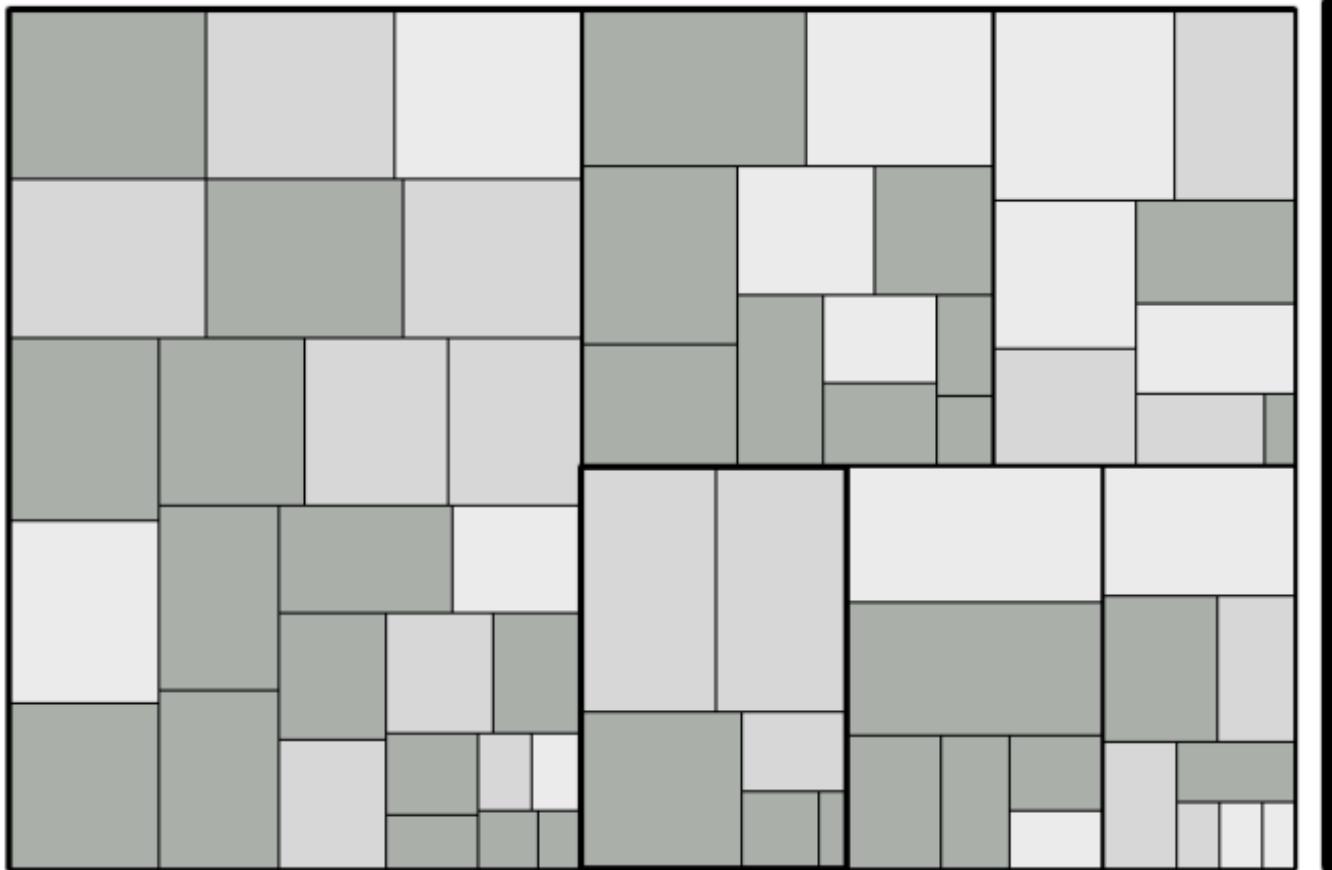
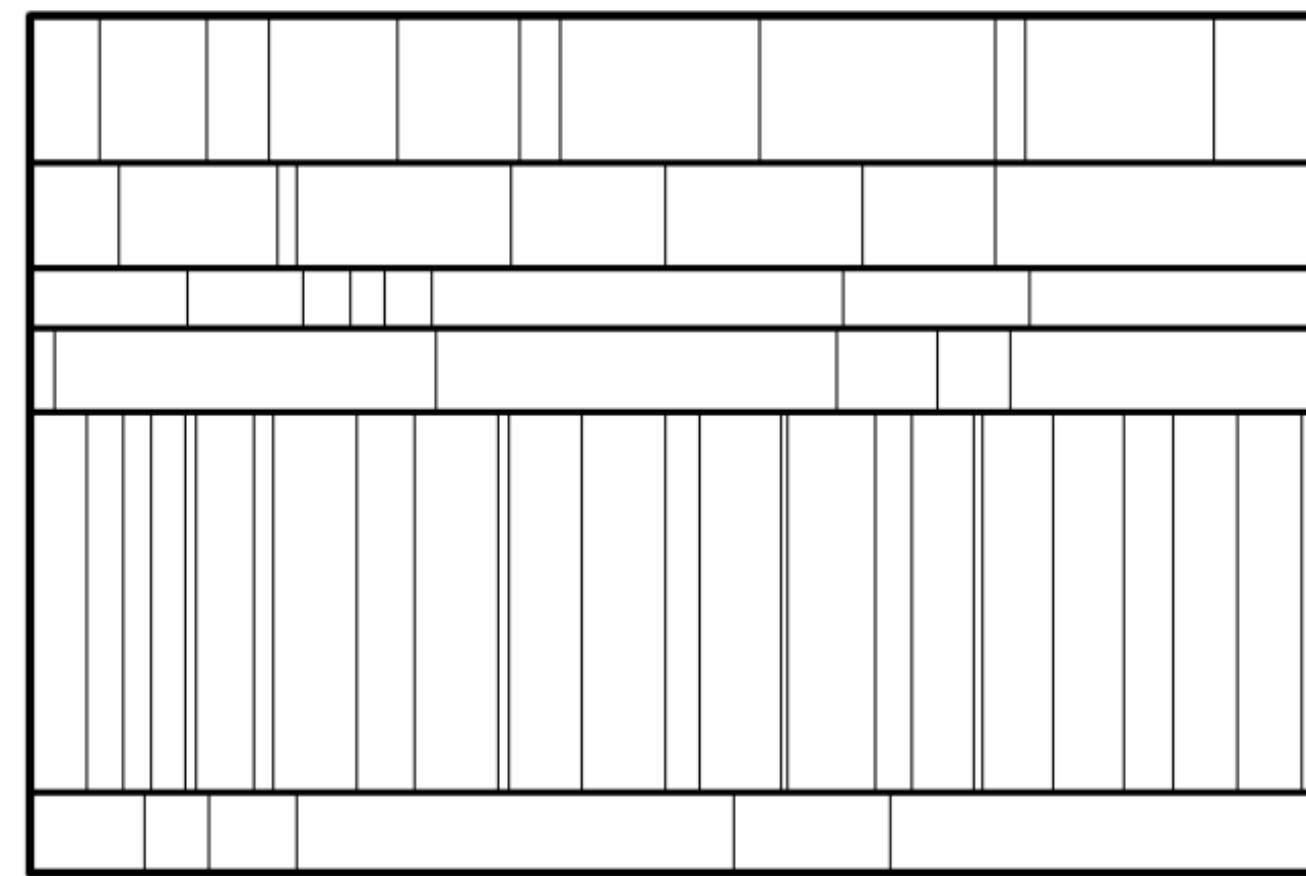
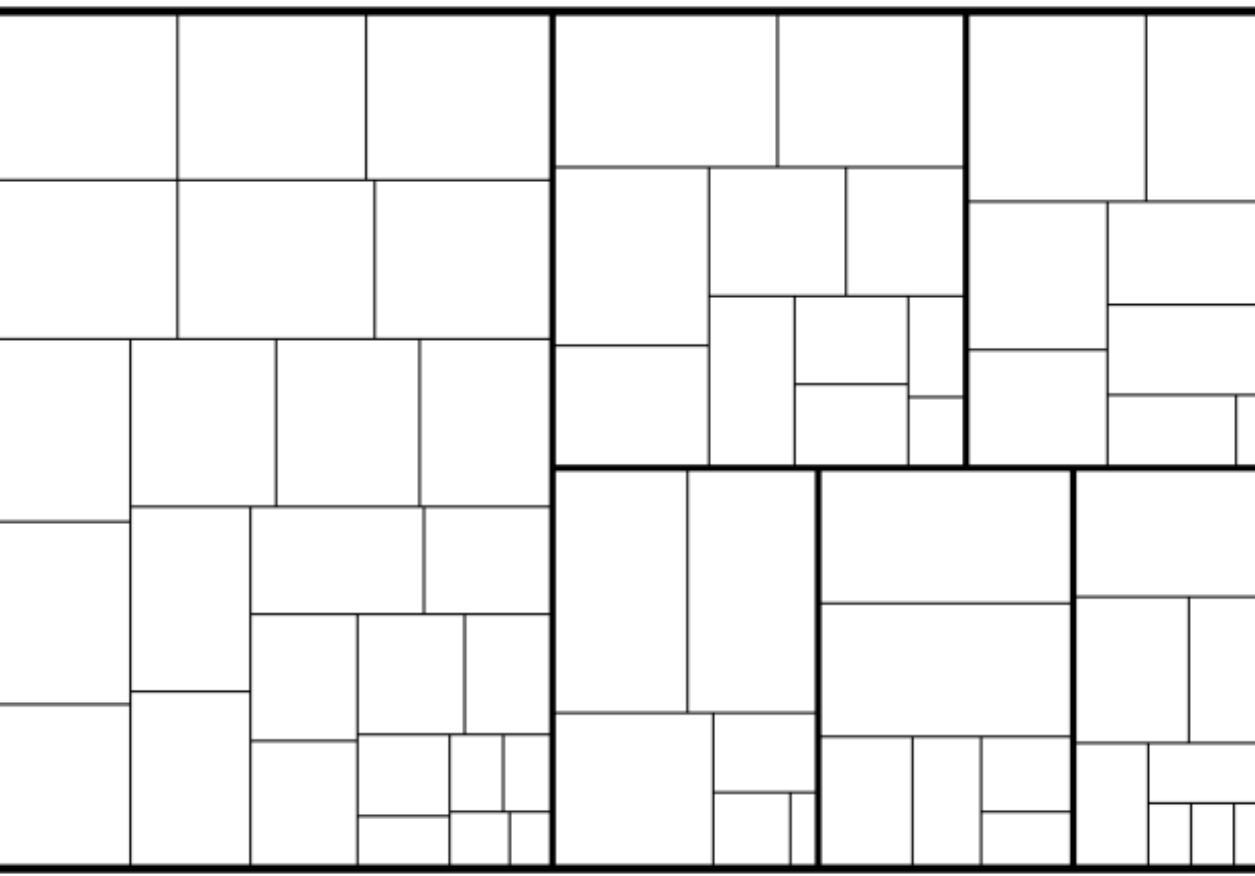
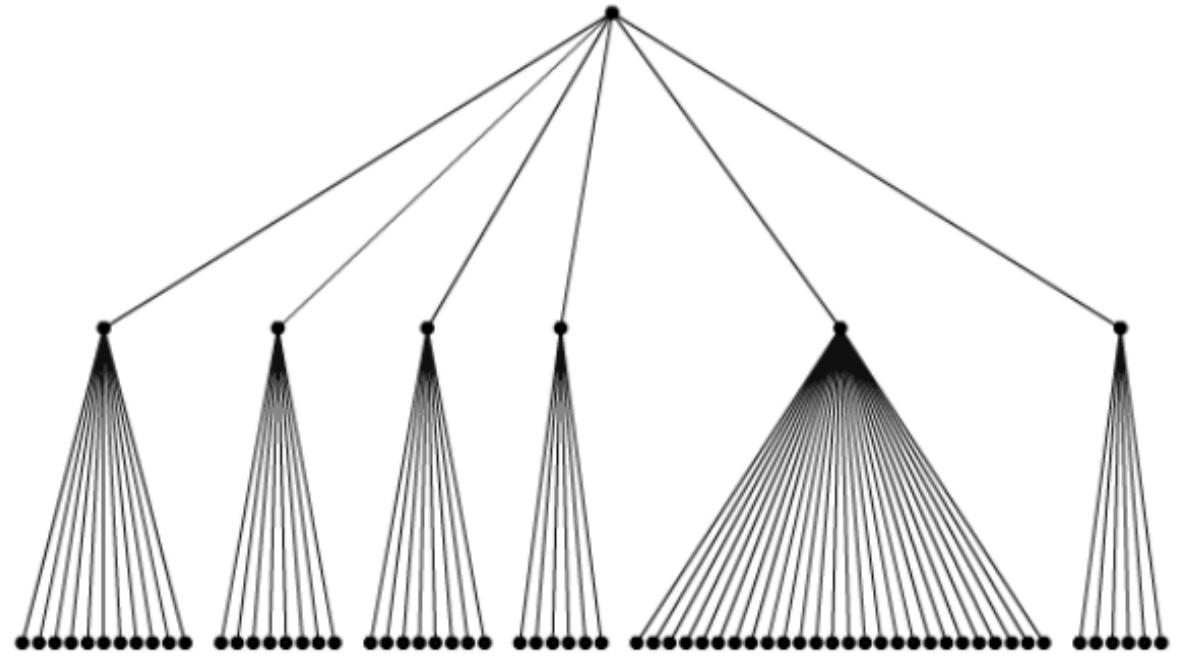
Speaking of **part-to-whole relationships...**

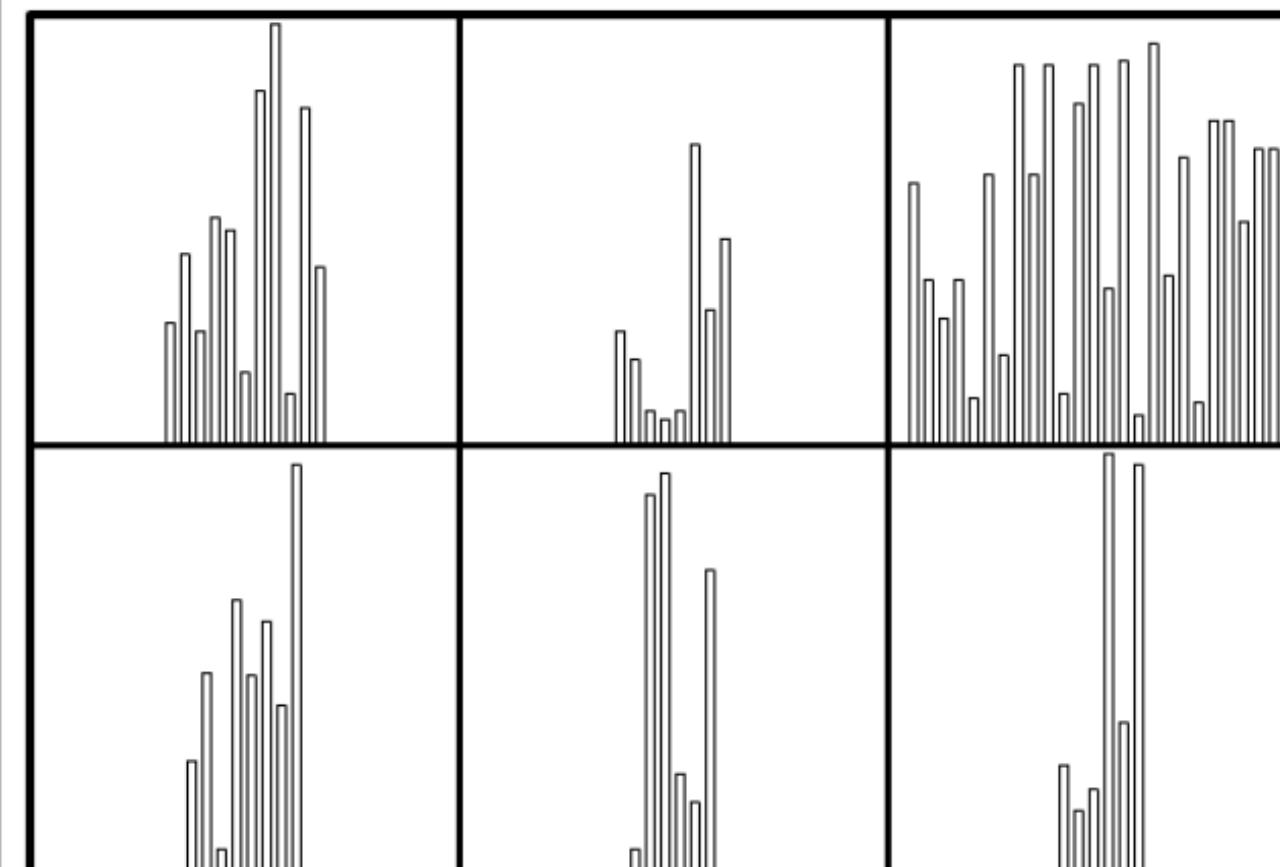
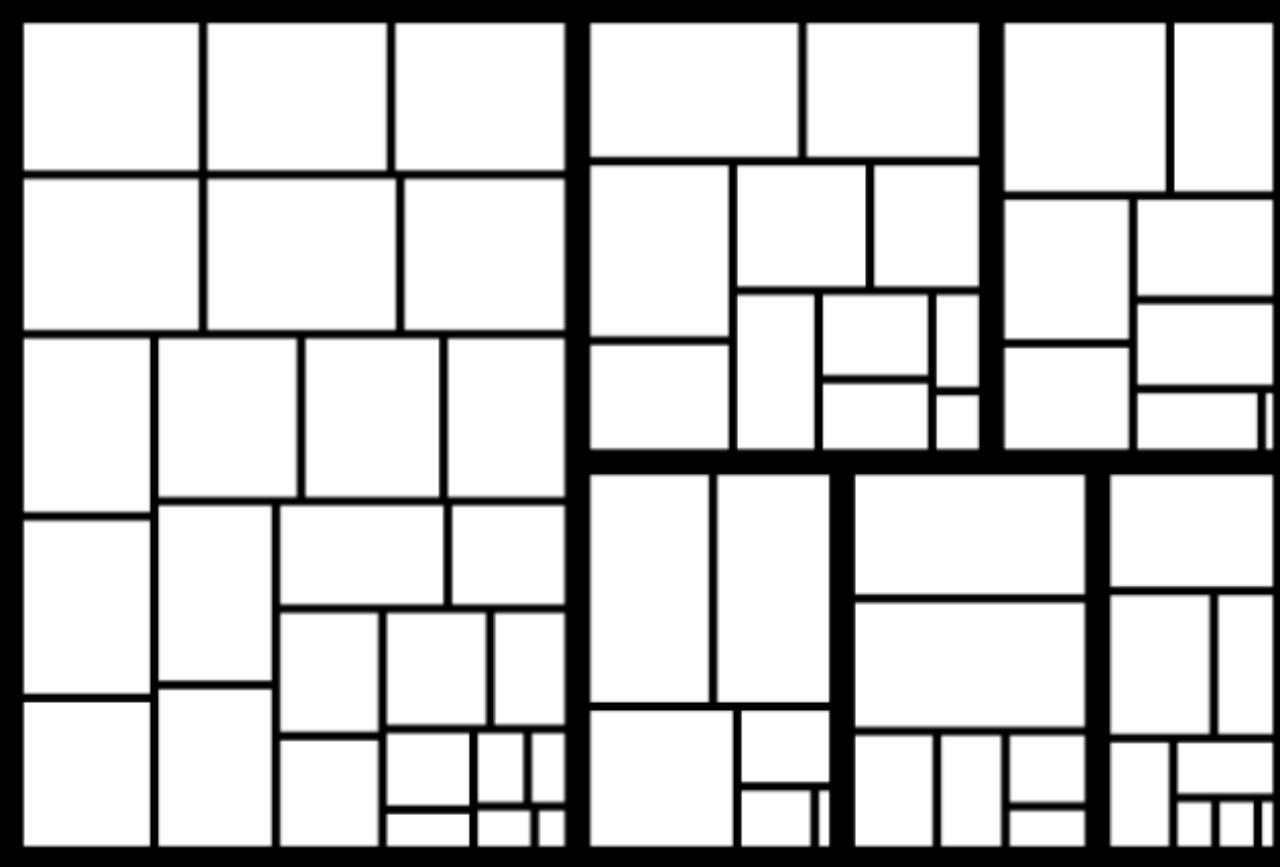
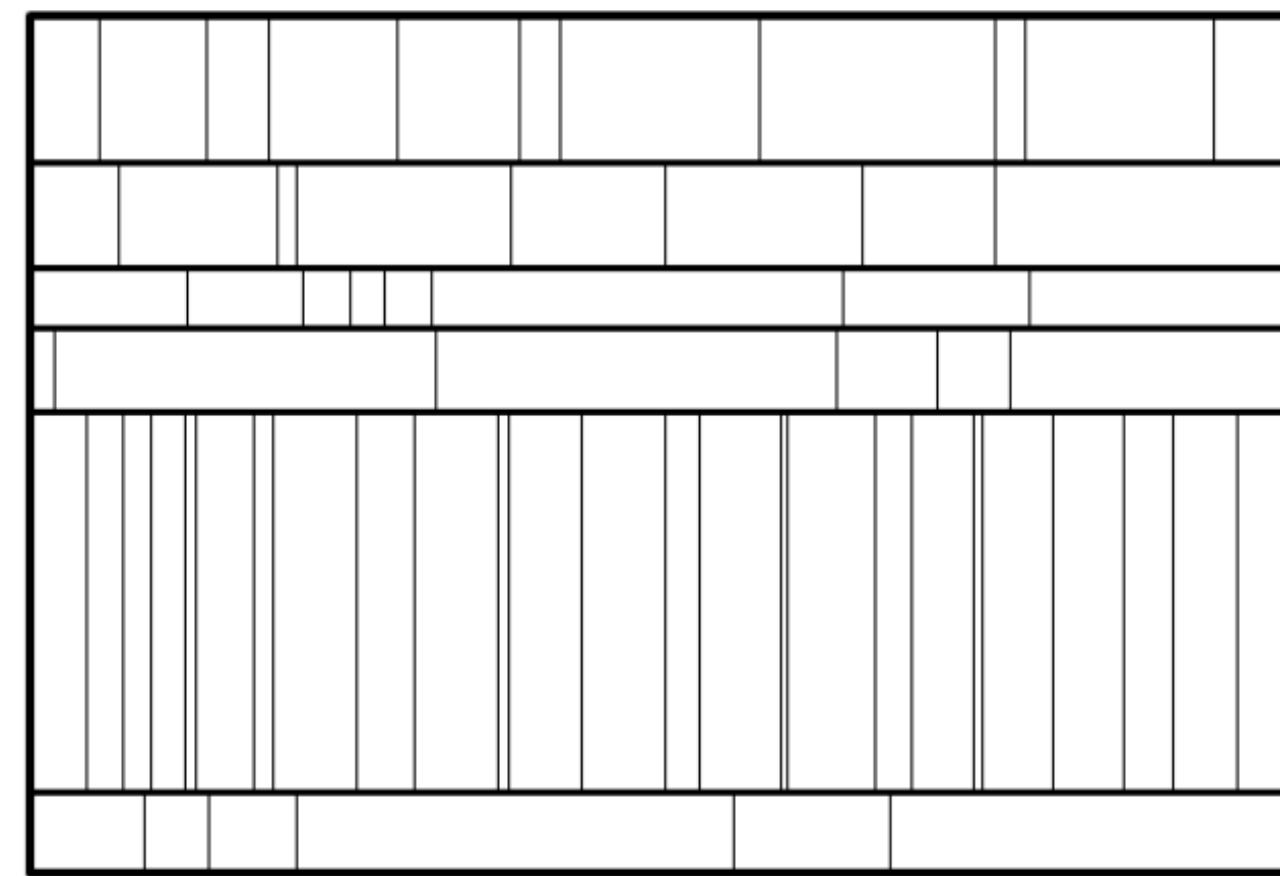
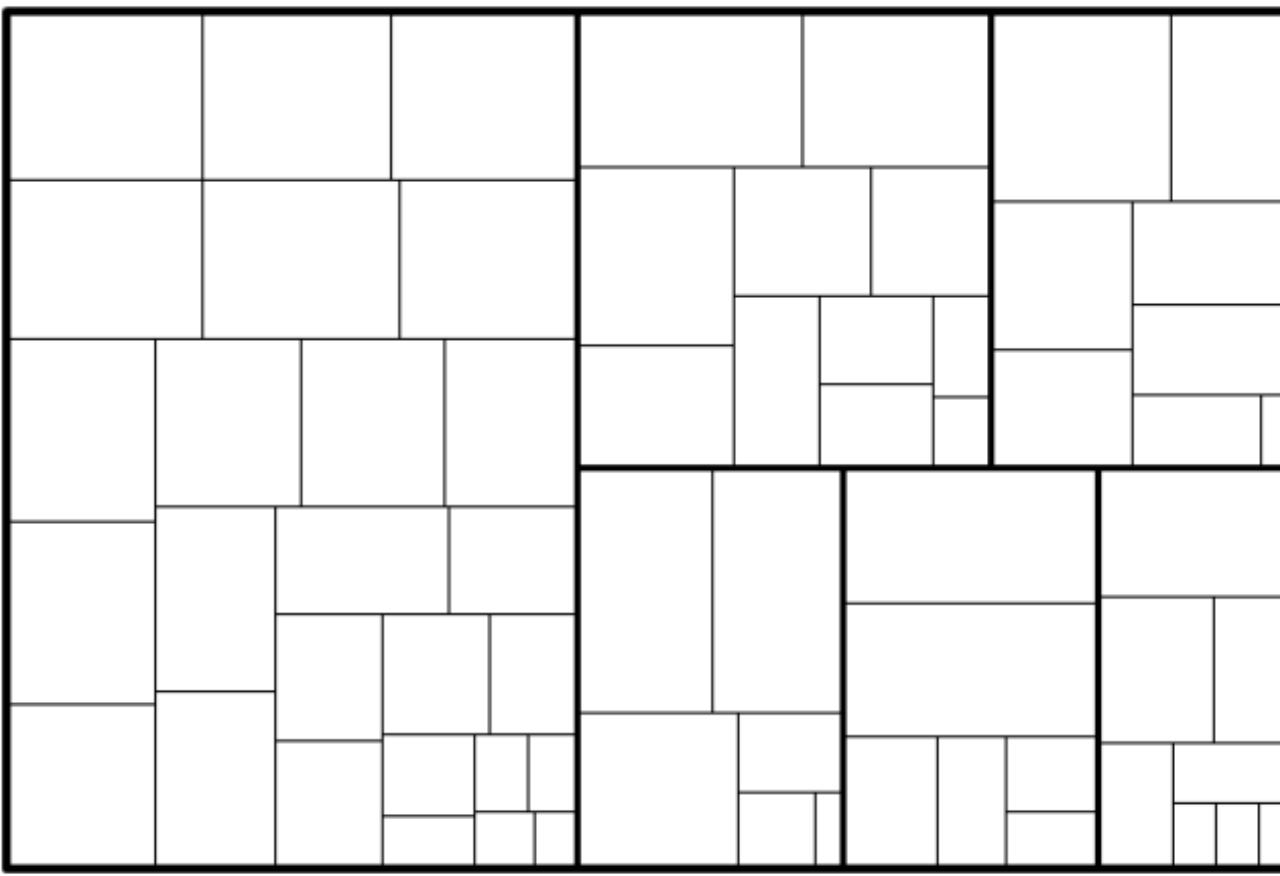
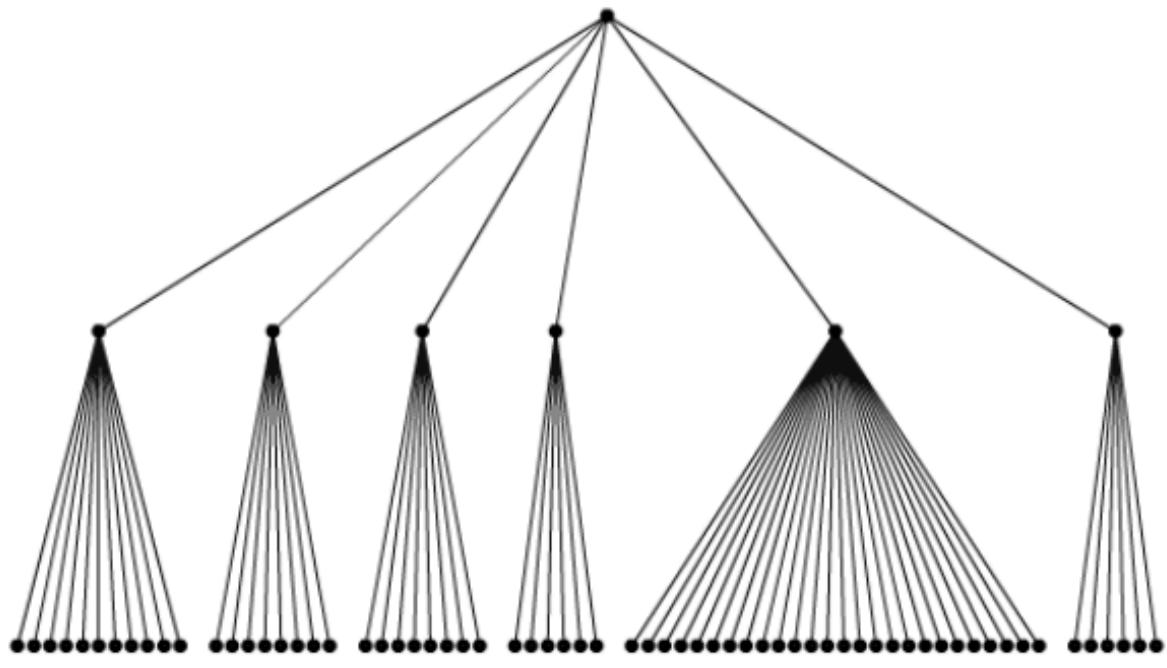
In 2010, Kong et al. isolated the **treemap design parameters** that can affect perception of rectangle area.

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**—

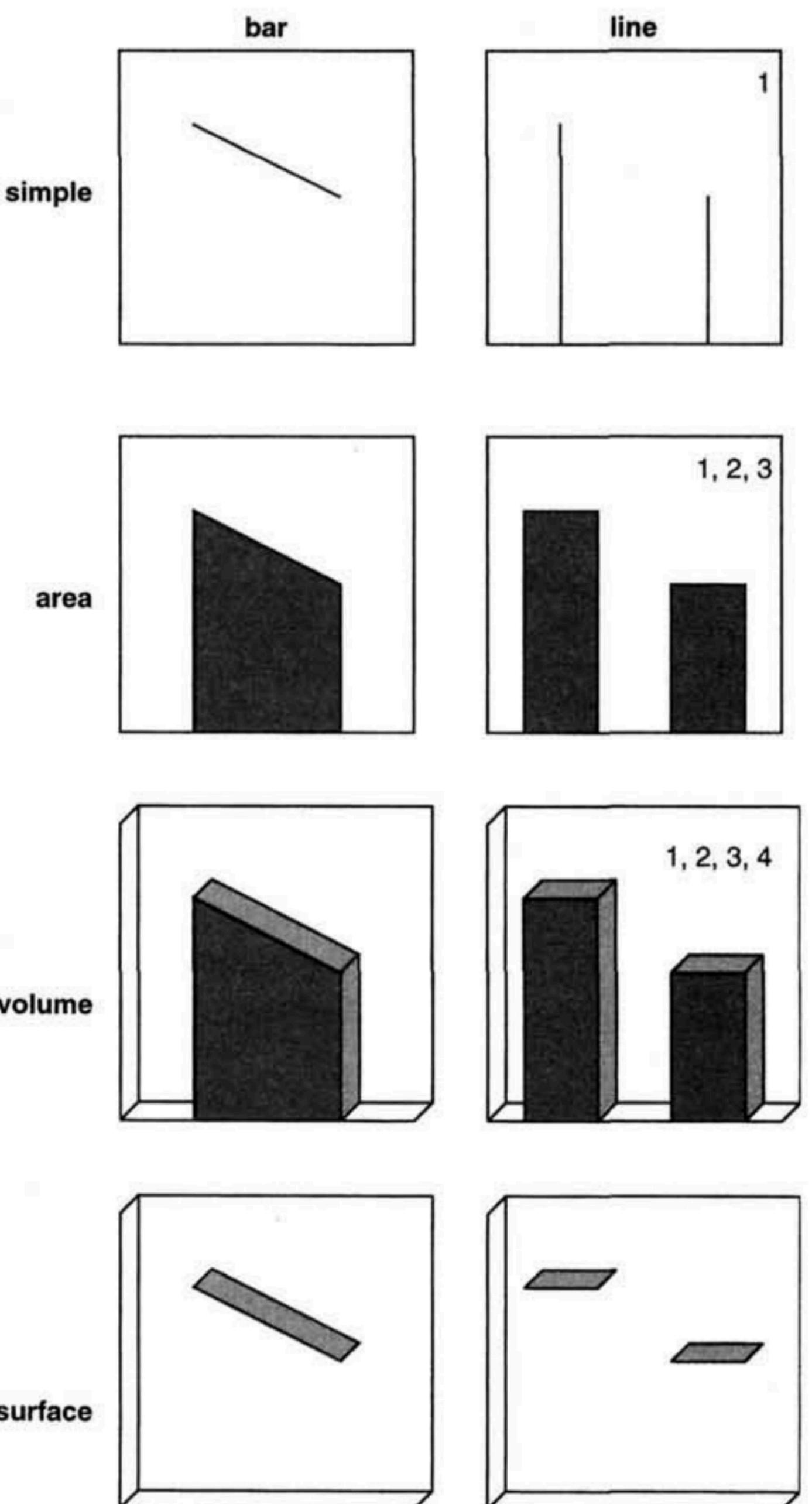
- Use treemap layouts that **avoid extreme aspect ratios** (+ squares)
- Use bar charts for low-density data, treemaps for high-density data
- Use tree maps when **comparing non-leaf nodes**
- Use **luminance** to encode secondary values

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 & Dataset Types**
3. Studio (Introduction to **Tableau**)

Data abstraction

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*



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



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 – Human name

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

Dataset **types**

Dataset types

- Table
- Network, tree
- Field
- Geometry

Dataset type — **Table**

- Each data **item** in a new row
- Each column contains an **attribute**

Dataset type – Table

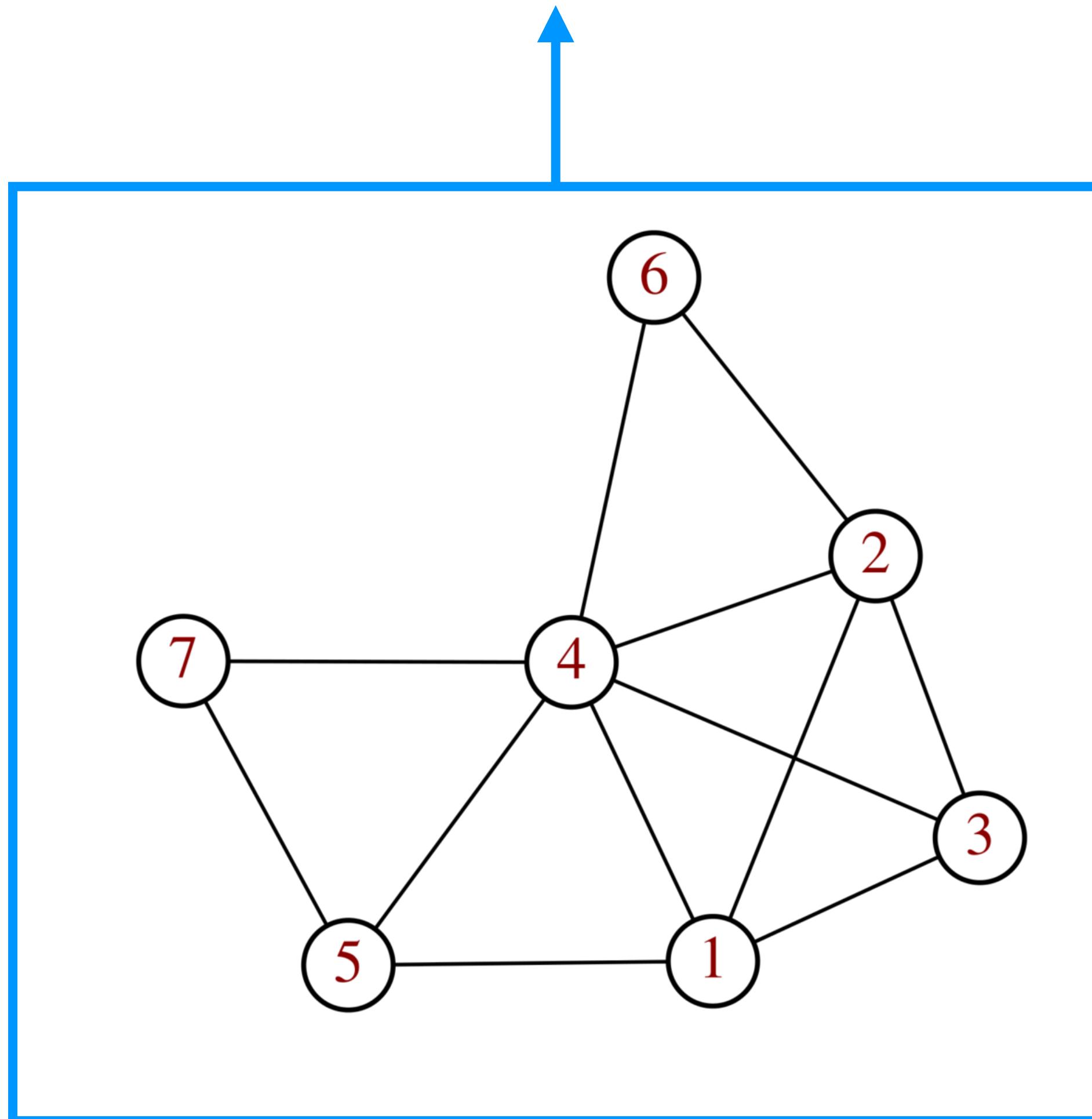
Item

Attribute

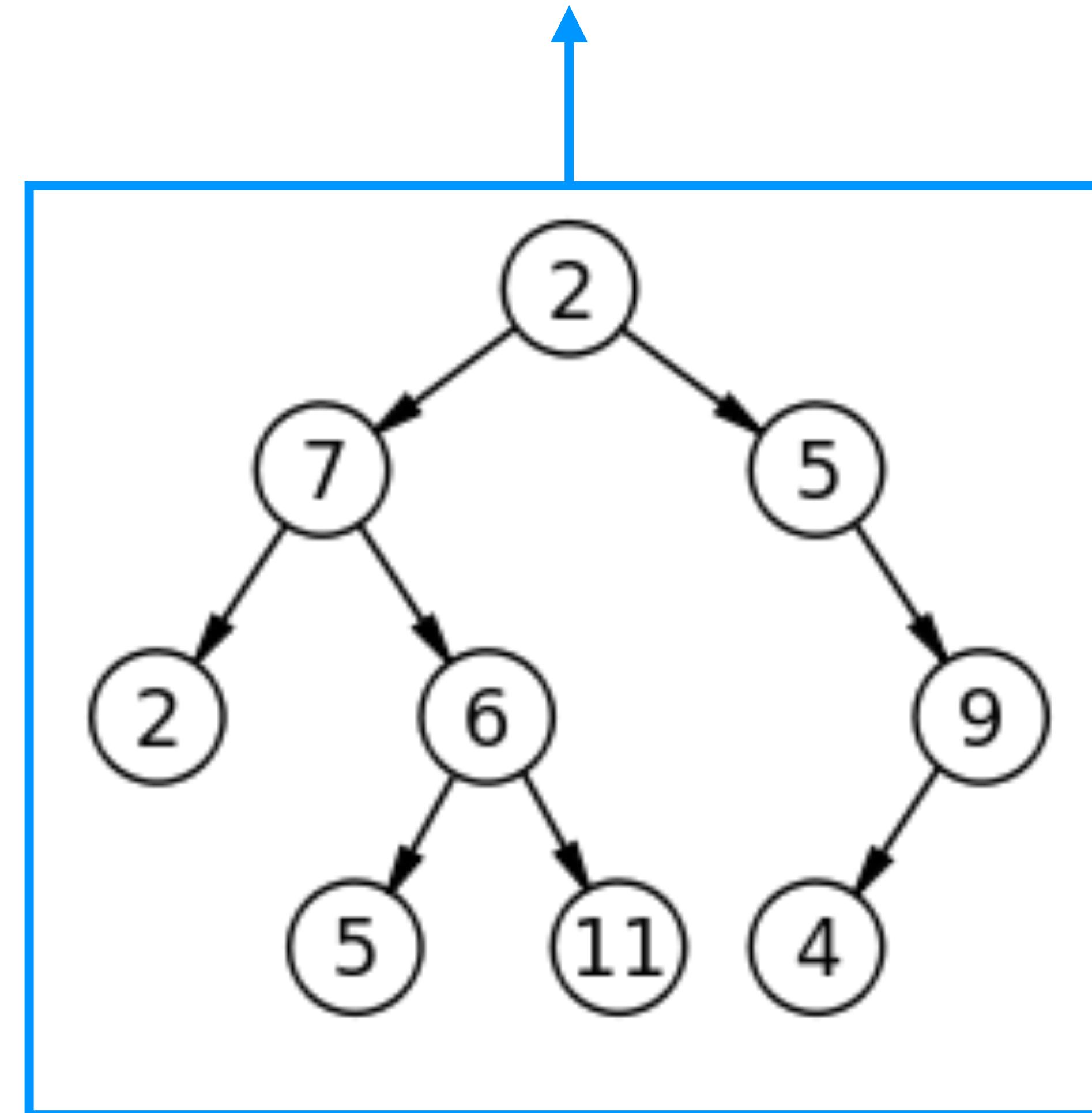
Dataset type — **Network**

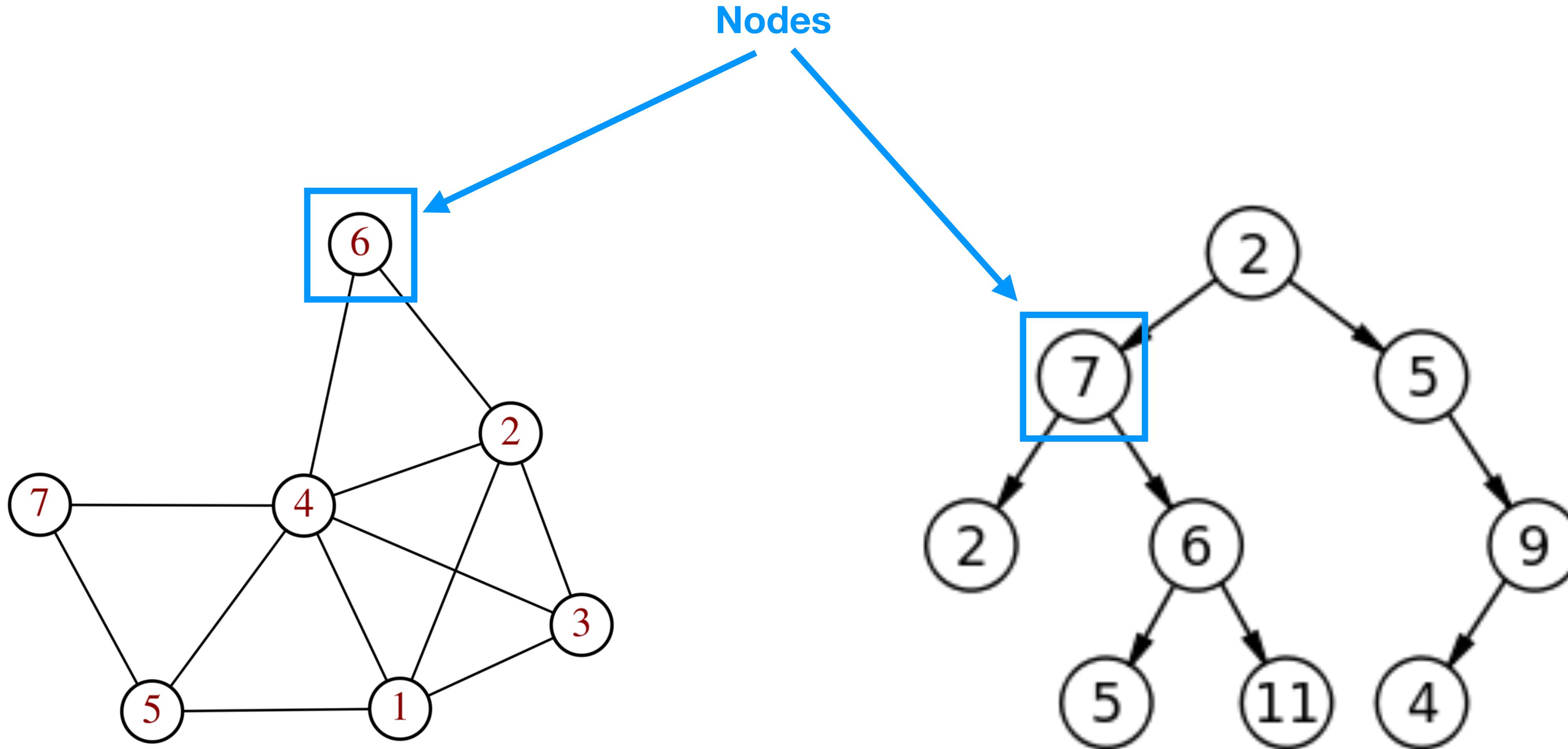
- Each item in a network is a **node**
- The relation between two items is a **link**
- Specifies that there is some kind of relationship between two or more items

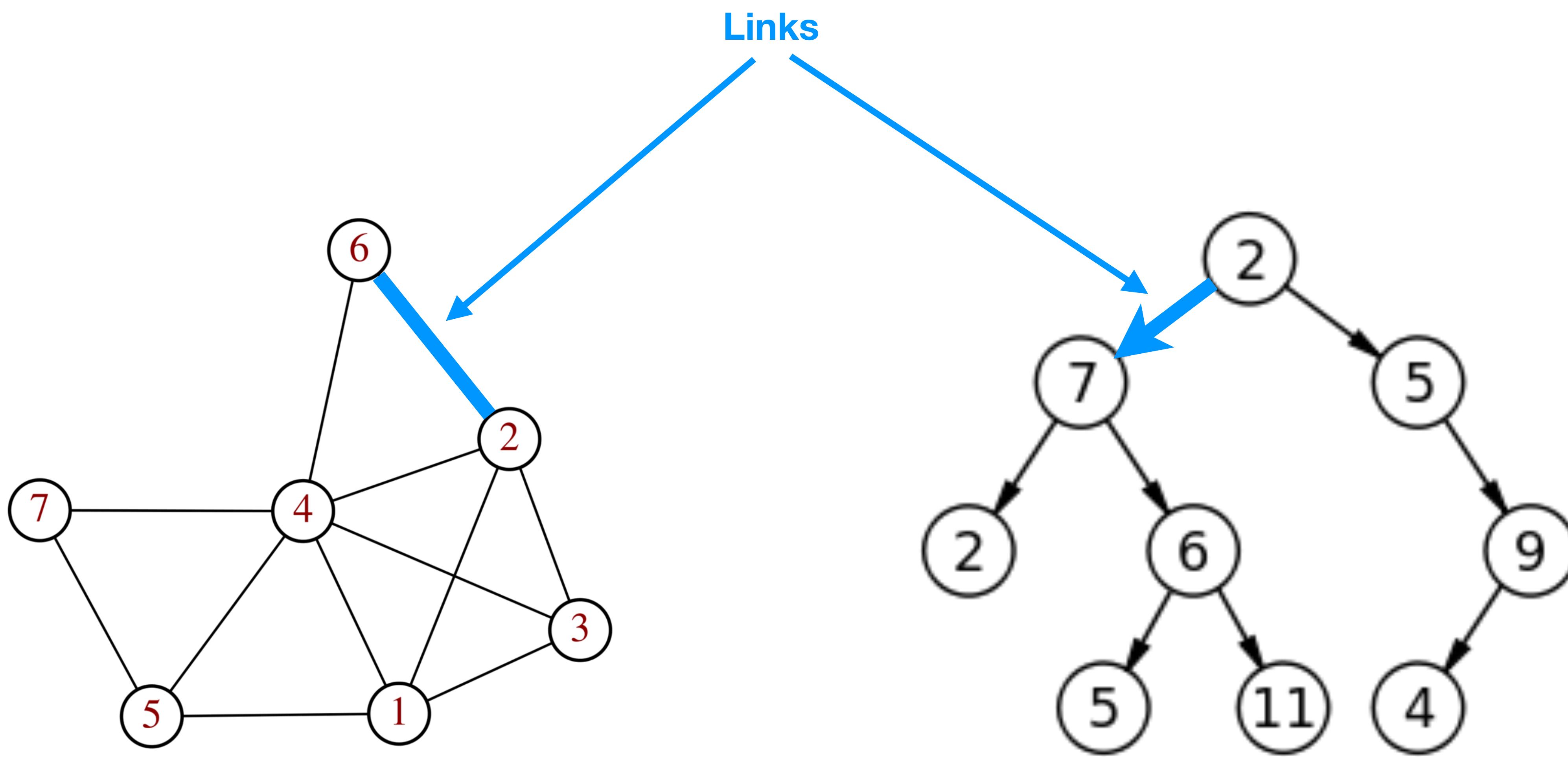
Dataset type — Network



Dataset type — Tree





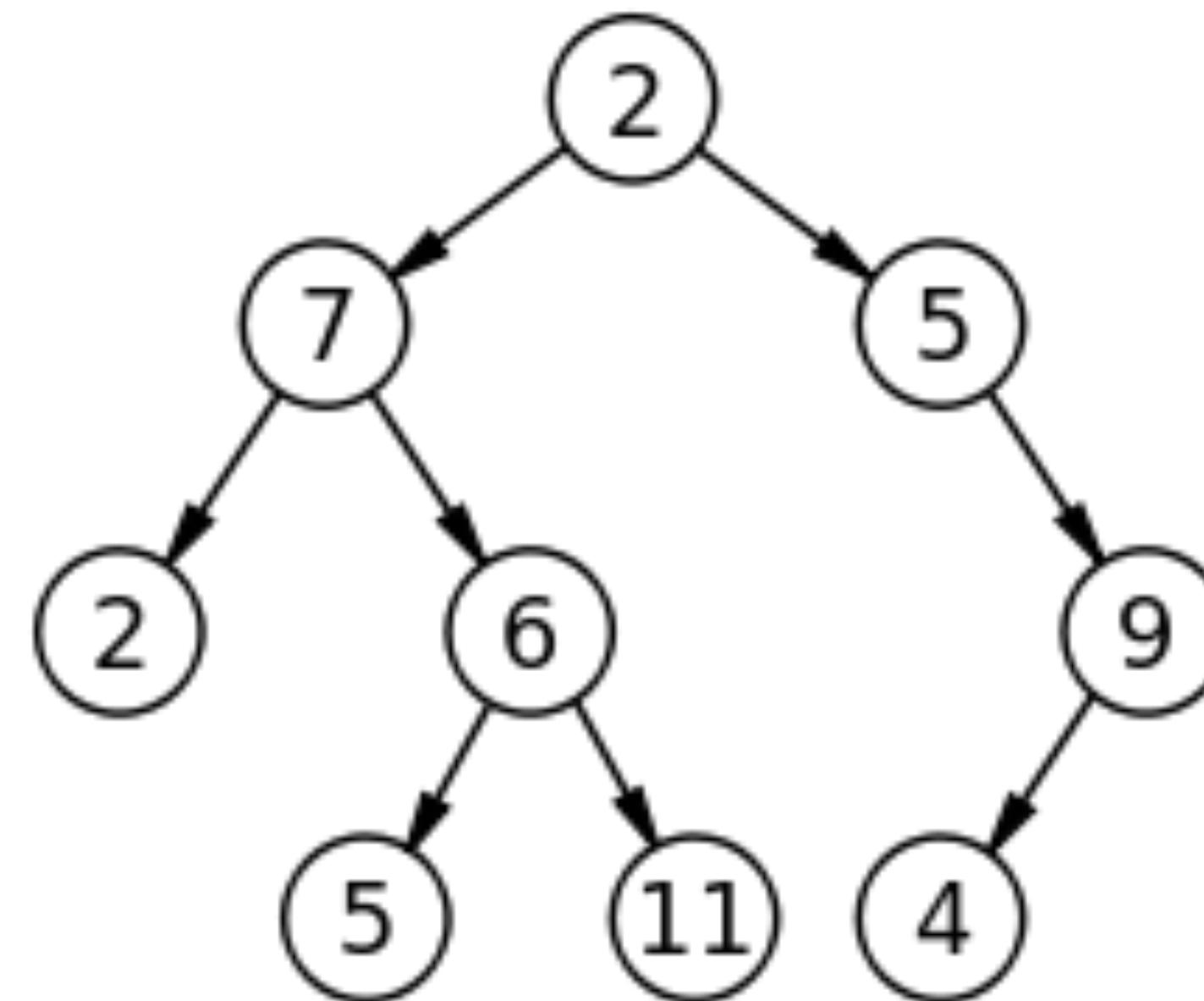


Social network



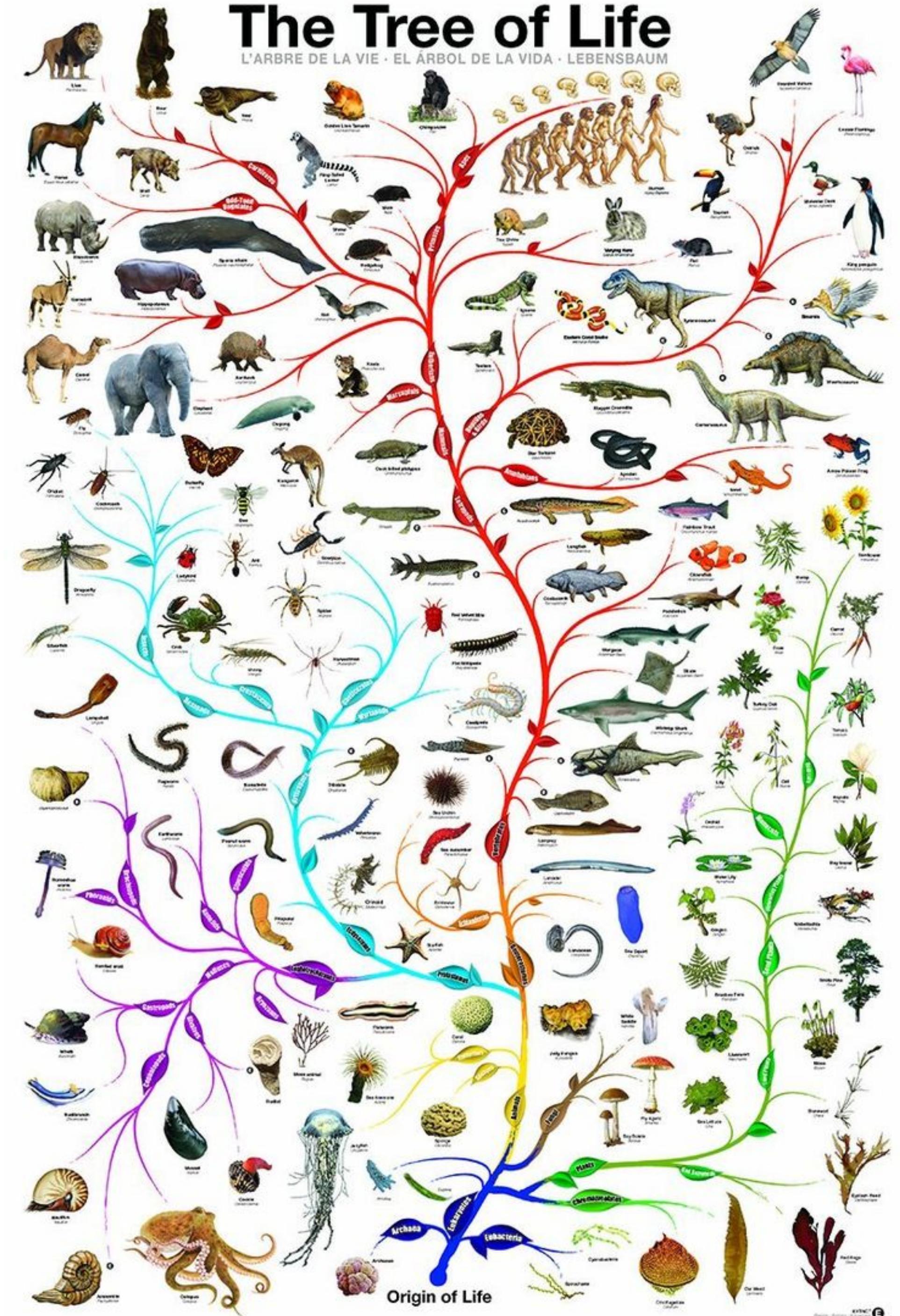
Dataset type — **Network (tree)**

- Networks with a hierarchical structure are **trees**
- Each child node has only one parent node pointing to it



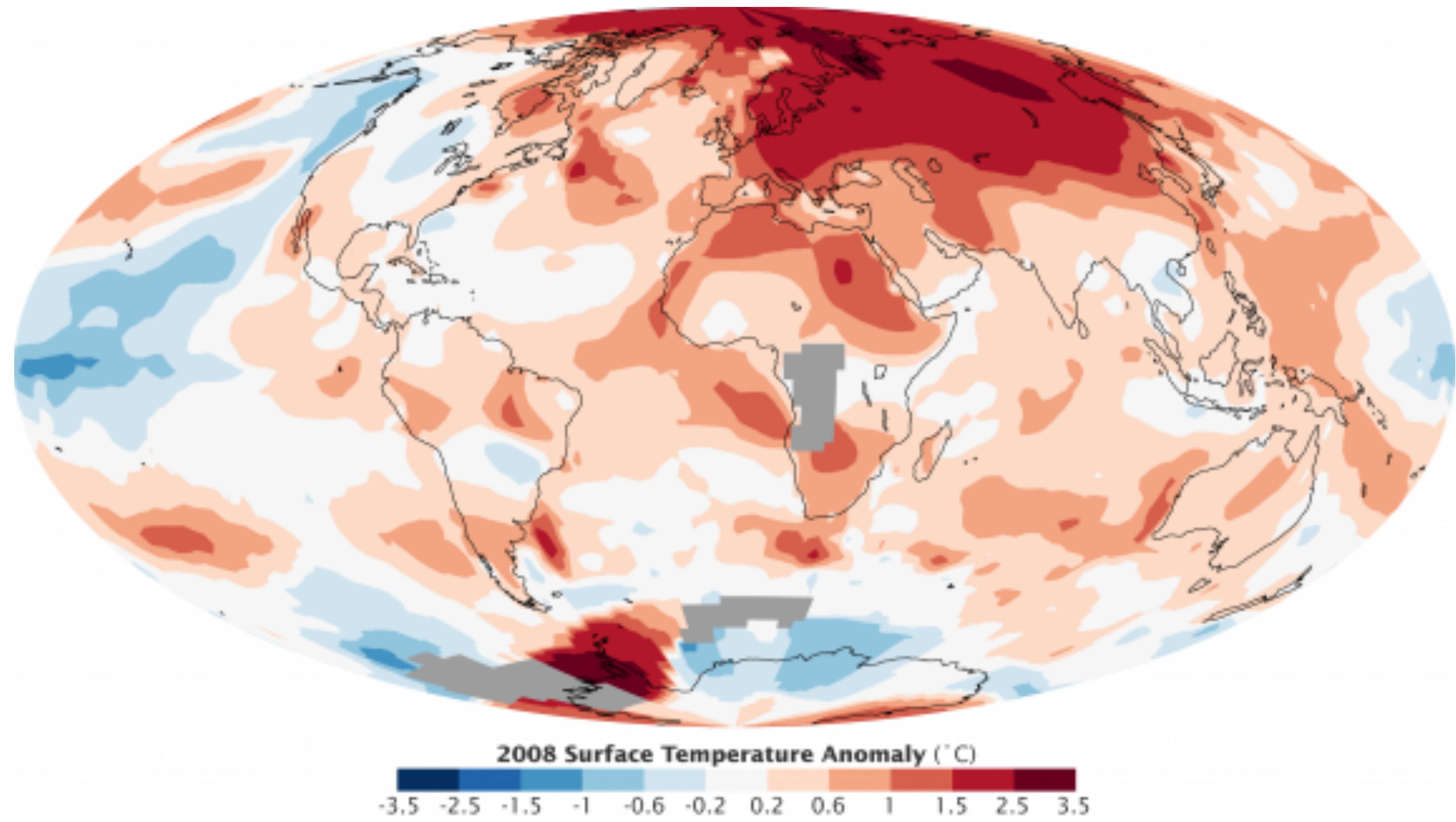
The Tree of Life

L'ARBRE DE LA VIE · EL ÁRBOL DE LA VIDA · LEBENSBAUM



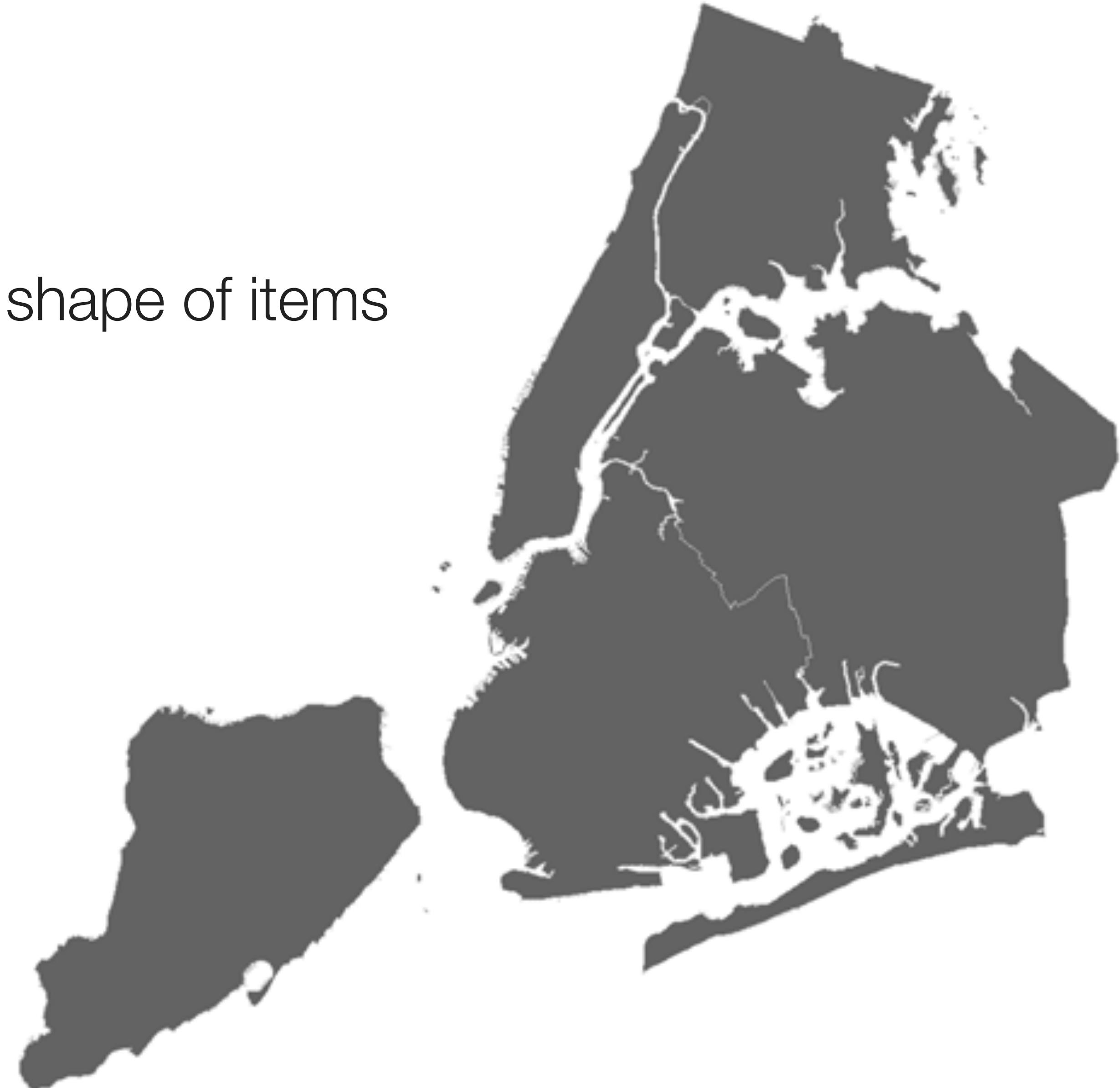
Type — **Field**

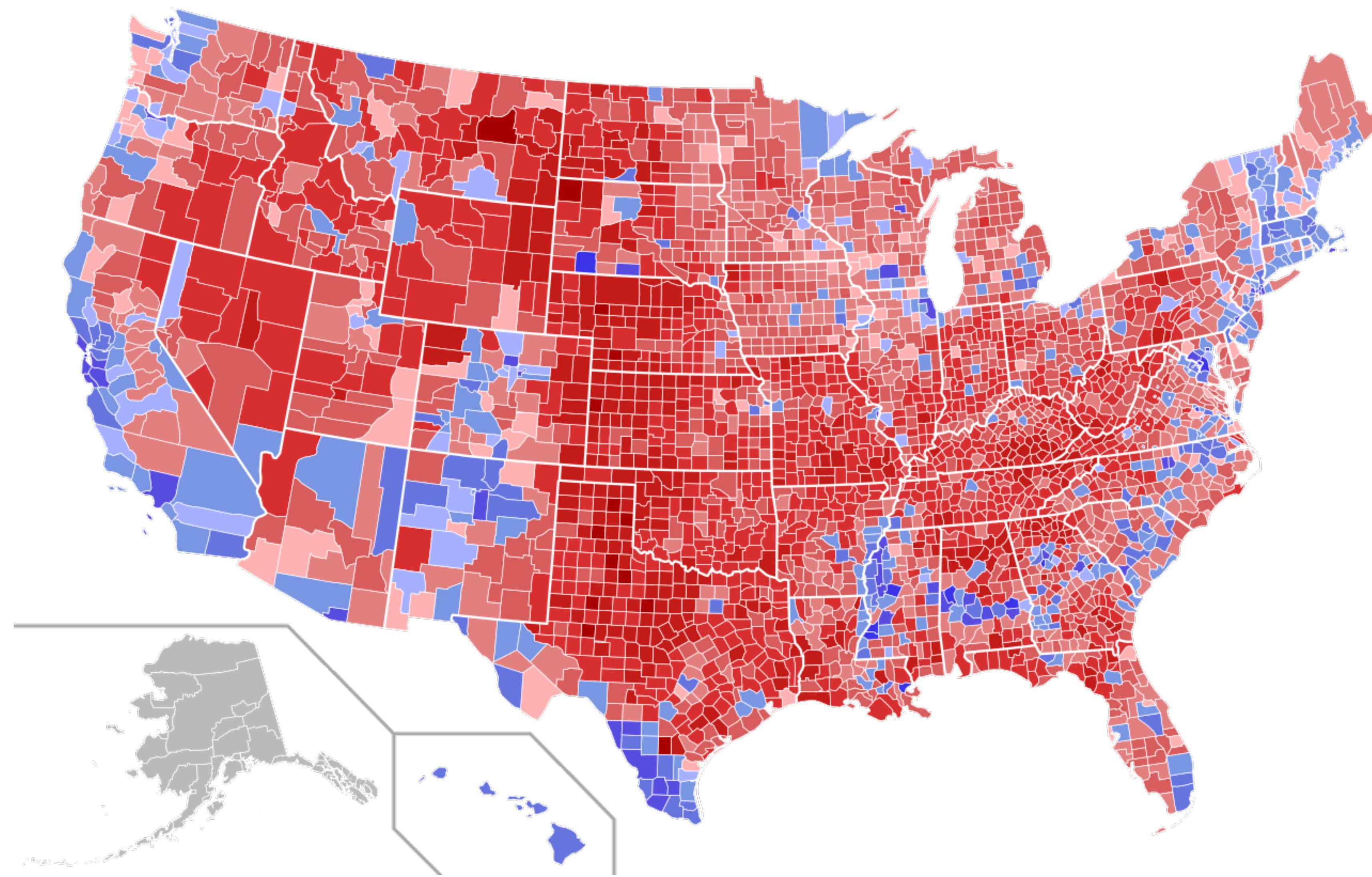
- The **field** dataset type also contains attribute values associated with cells
- Each **cell** in a field contains measurements or calculations from a continuous domain (conceptually infinite values).
- Considerations
 - **Sampling** – How frequently to take measurements
 - **Interpolation** – How to show values between points in a way that does not mislead
- Ex. (in the physical world) temperature, pressure, wind velocity



Type — **Geometry**

- Specifies information about the shape of items with specific spatial positions





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

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

-