Data Visualization (CIS 490/680)

Channels & Tables

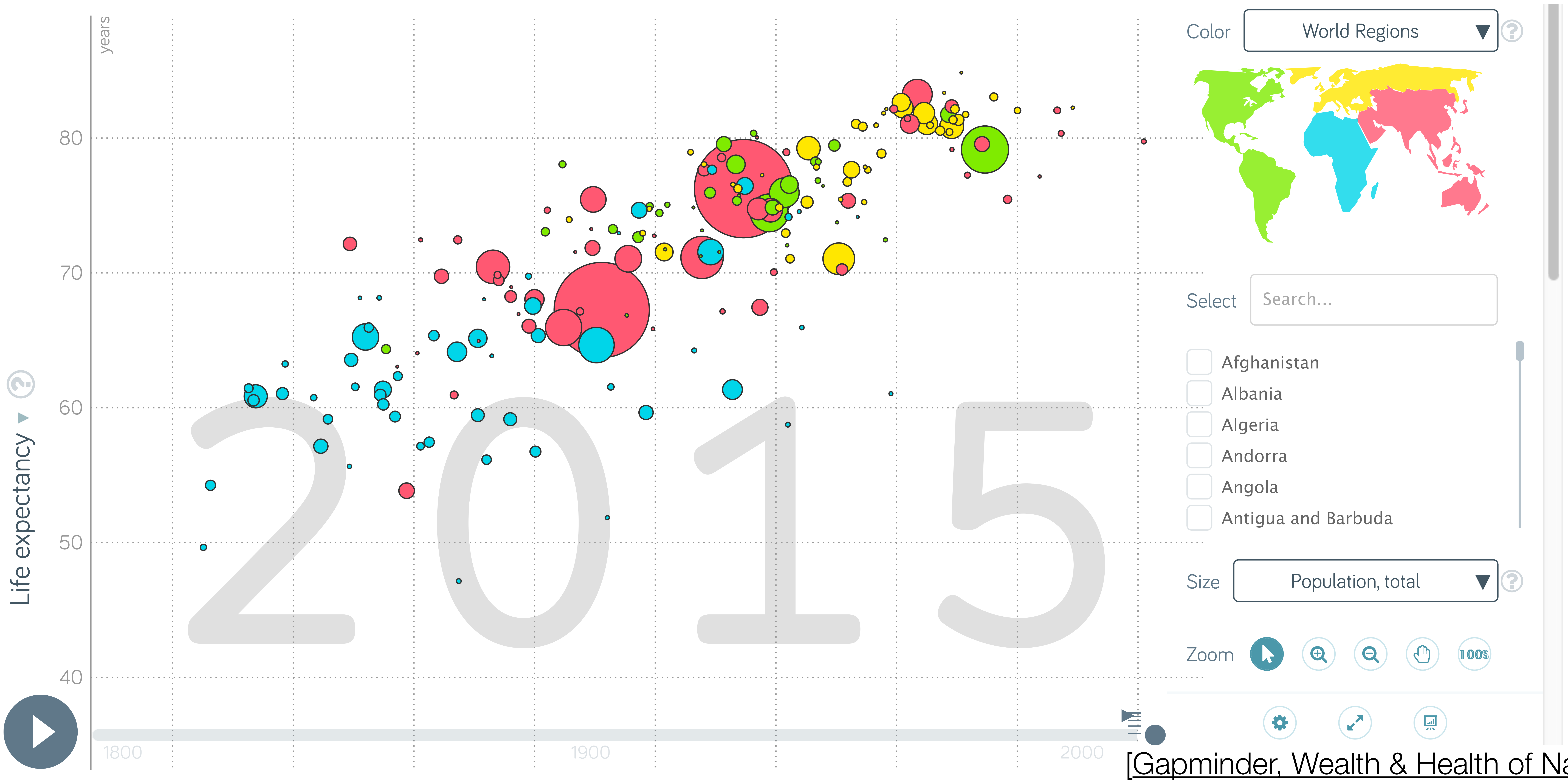
Dr. David Koop

Visual Encoding

- How should we visualize this data?

Name	Region	Population	Life Expectancy	Income
China	East Asia & Pacific	1335029250	73.28	7226.07
India	South Asia	1140340245	64.01	2731
United States	America	306509345	79.43	41256.08
Indonesia	East Asia & Pacific	228721000	71.17	3818.08
Brazil	America	193806549	72.68	9569.78
Pakistan	South Asia	176191165	66.84	2603
Bangladesh	South Asia	156645463	66.56	1492
Nigeria	Sub-Saharan Africa	141535316	48.17	2158.98
Japan	East Asia & Pacific	127383472	82.98	29680.68
Mexico	America	111209909	76.47	11250.37
Philippines	East Asia & Pacific	94285619	72.1	3203.97
Vietnam	East Asia & Pacific	86970762	74.7	2679.34
Germany	Europe & Central Asia	82338100	80.08	31191.15
Ethiopia	Sub-Saharan Africa	79996293	55.69	812.16
Turkey	Europe & Central Asia	72626967	72.06	8040.78

Potential Solution



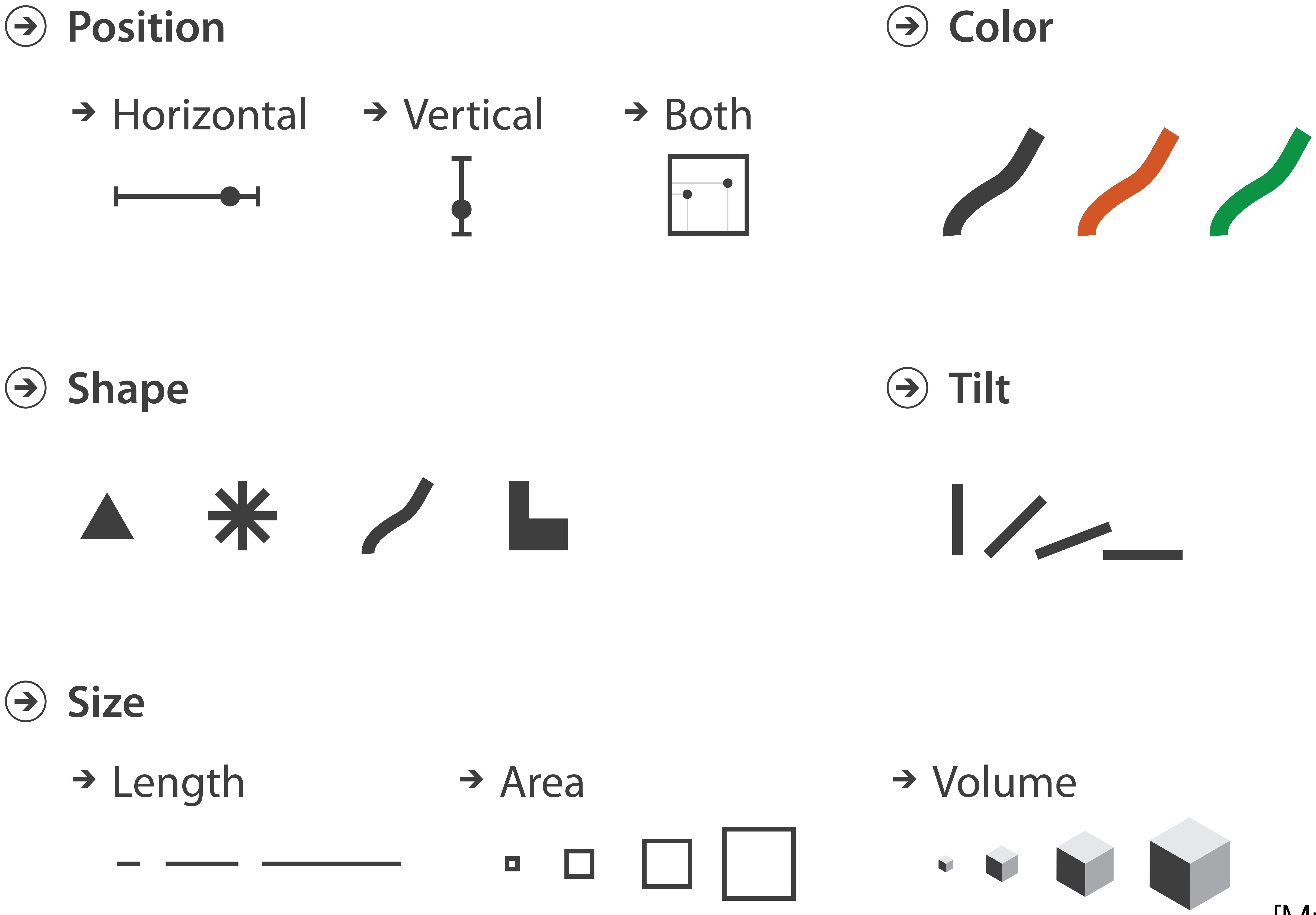
[Gapminder, Wealth & Health of Nations]

Visual Encoding

- How do we encode data visually?
  - **Marks** are the basic graphical elements in a visualization
  - **Channels** are ways to control the appearance of the marks
- Marks classified by dimensionality:

- Also can have surfaces, volumes
- Think of marks as a mathematical definition, or if familiar with tools like Adobe Illustrator or Inkscape, the path & point definitions

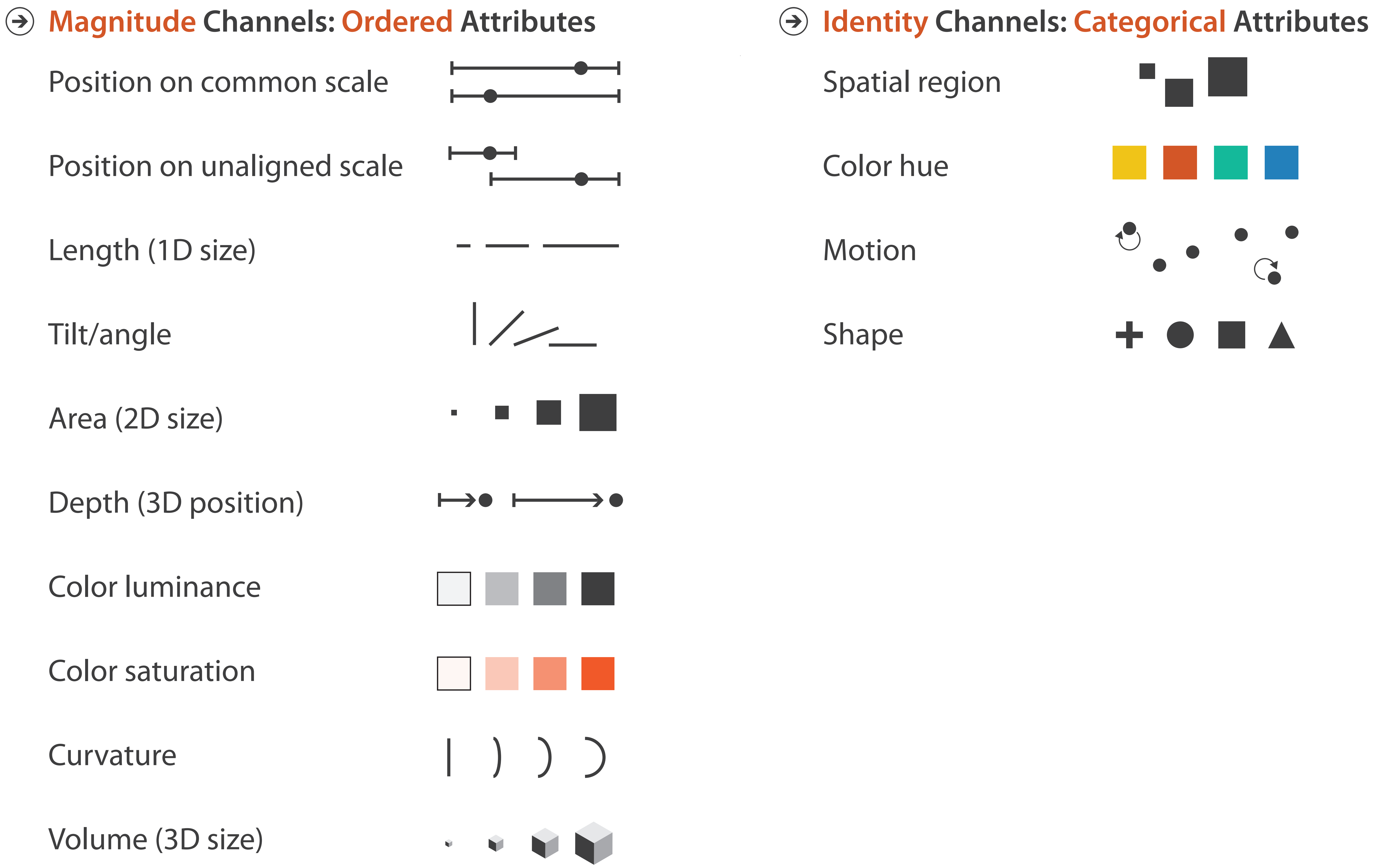
Visual Channels



[Munzner (ill. Maguire), 2014]

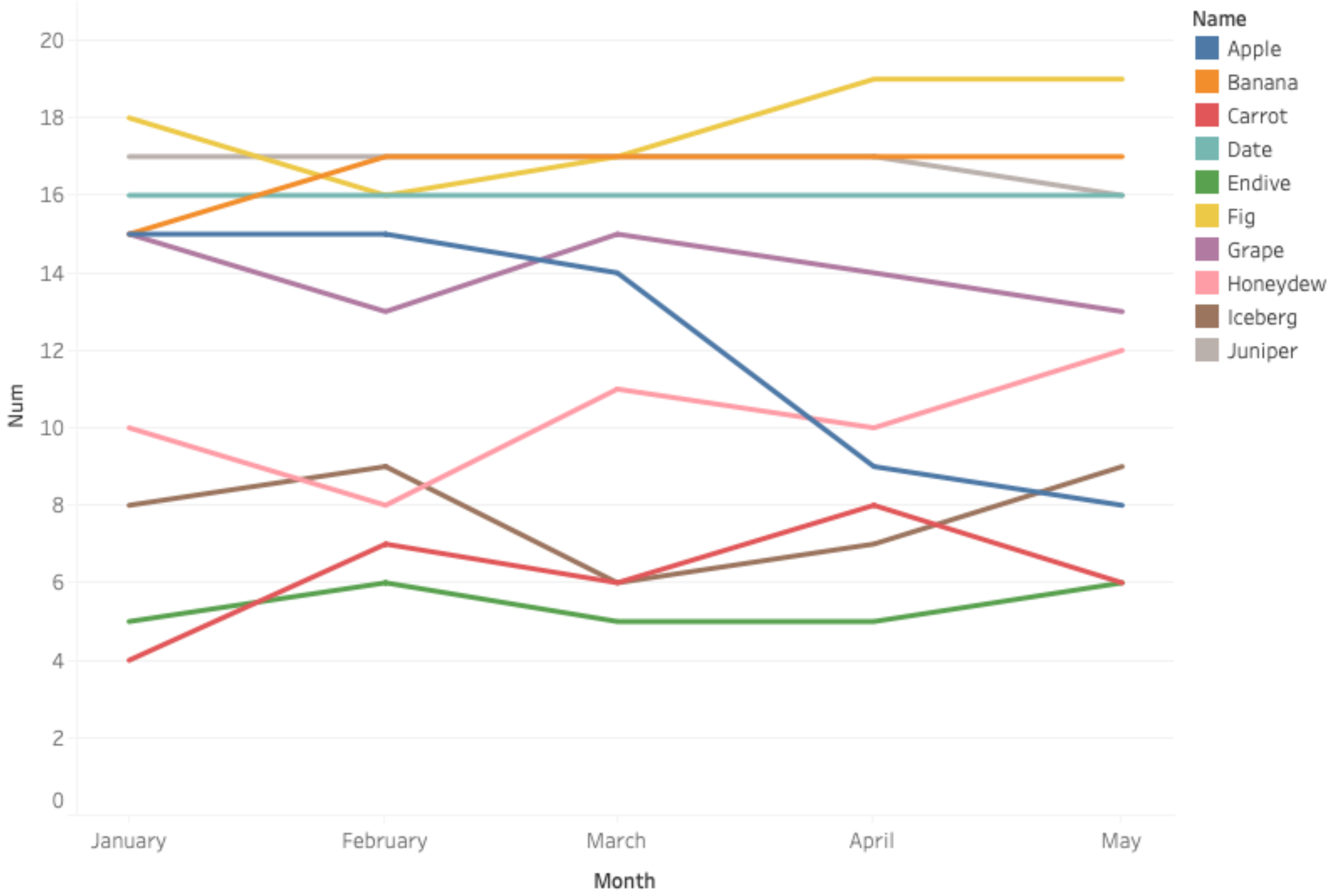
Channel Types

- Identity => what or where, Magnitude => how much

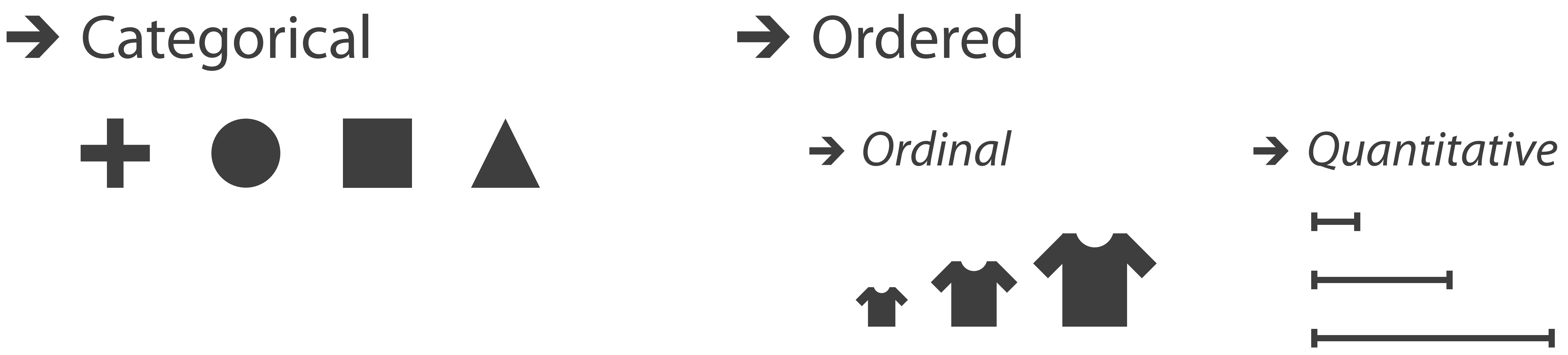


[Munzner (ill. Maguire), 2014]

Tableau Example

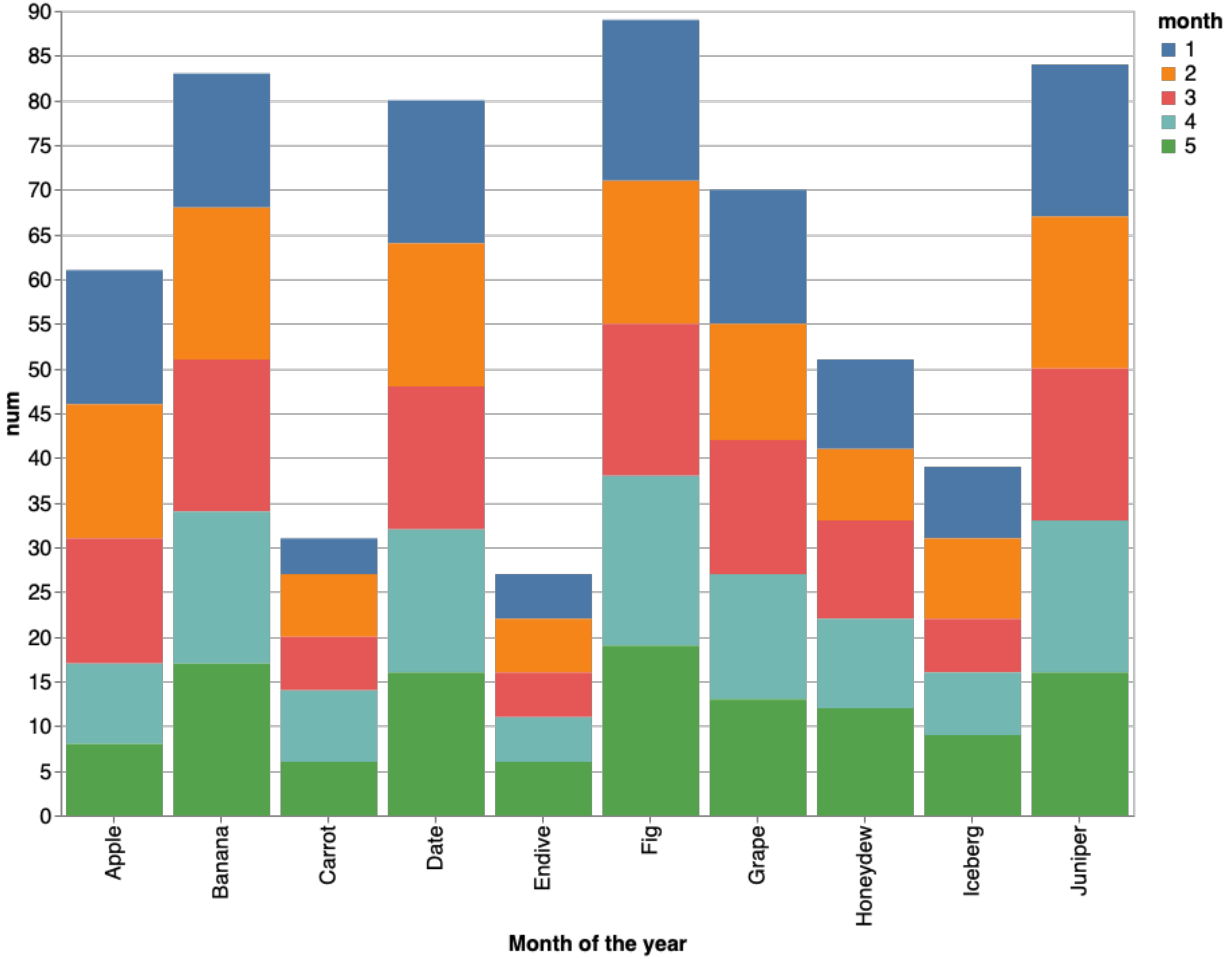


Data In Tableau



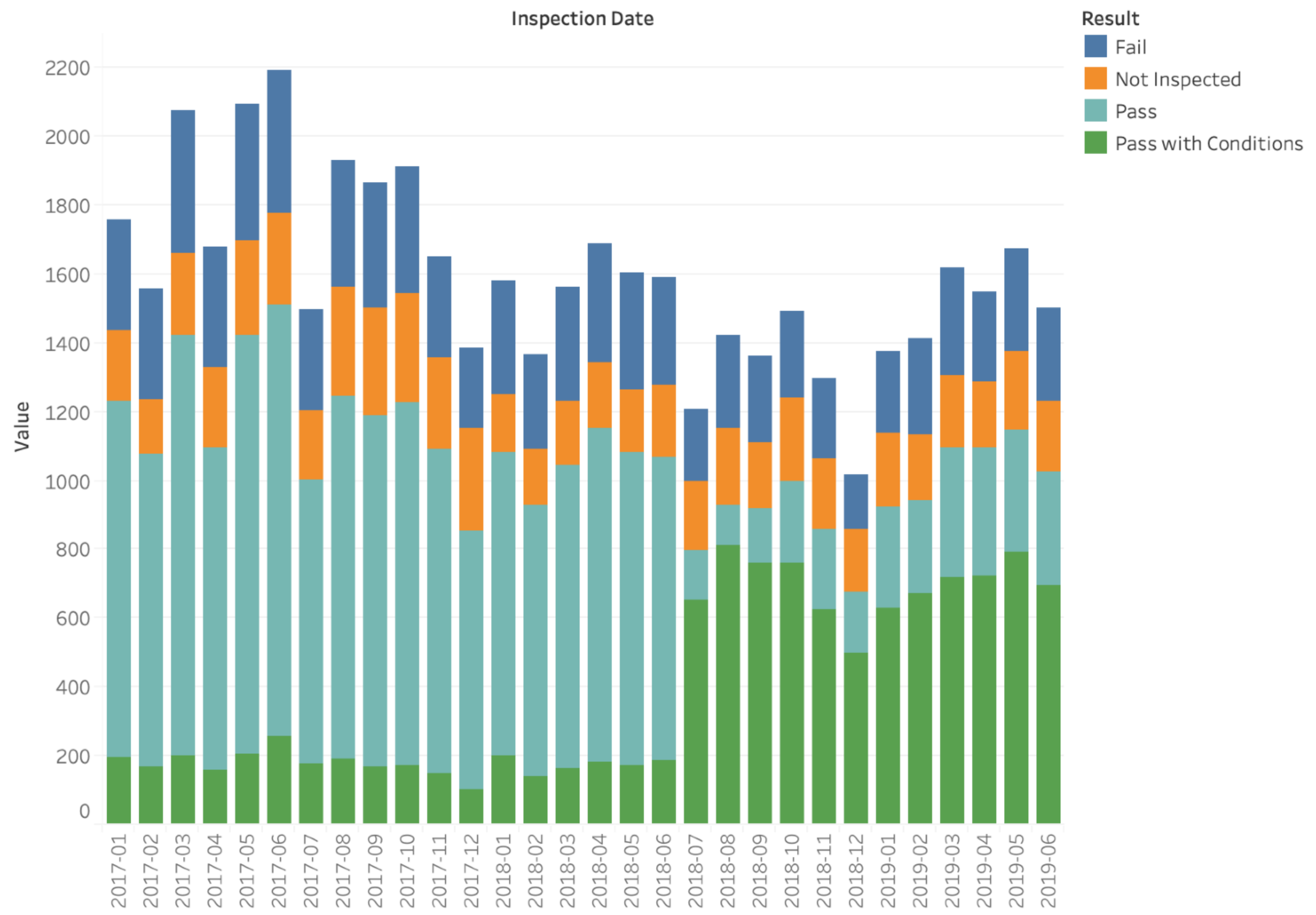
- Categorical data = Dimension
- Quantitative data = Measures

Vega-Lite Example



Assignment 3

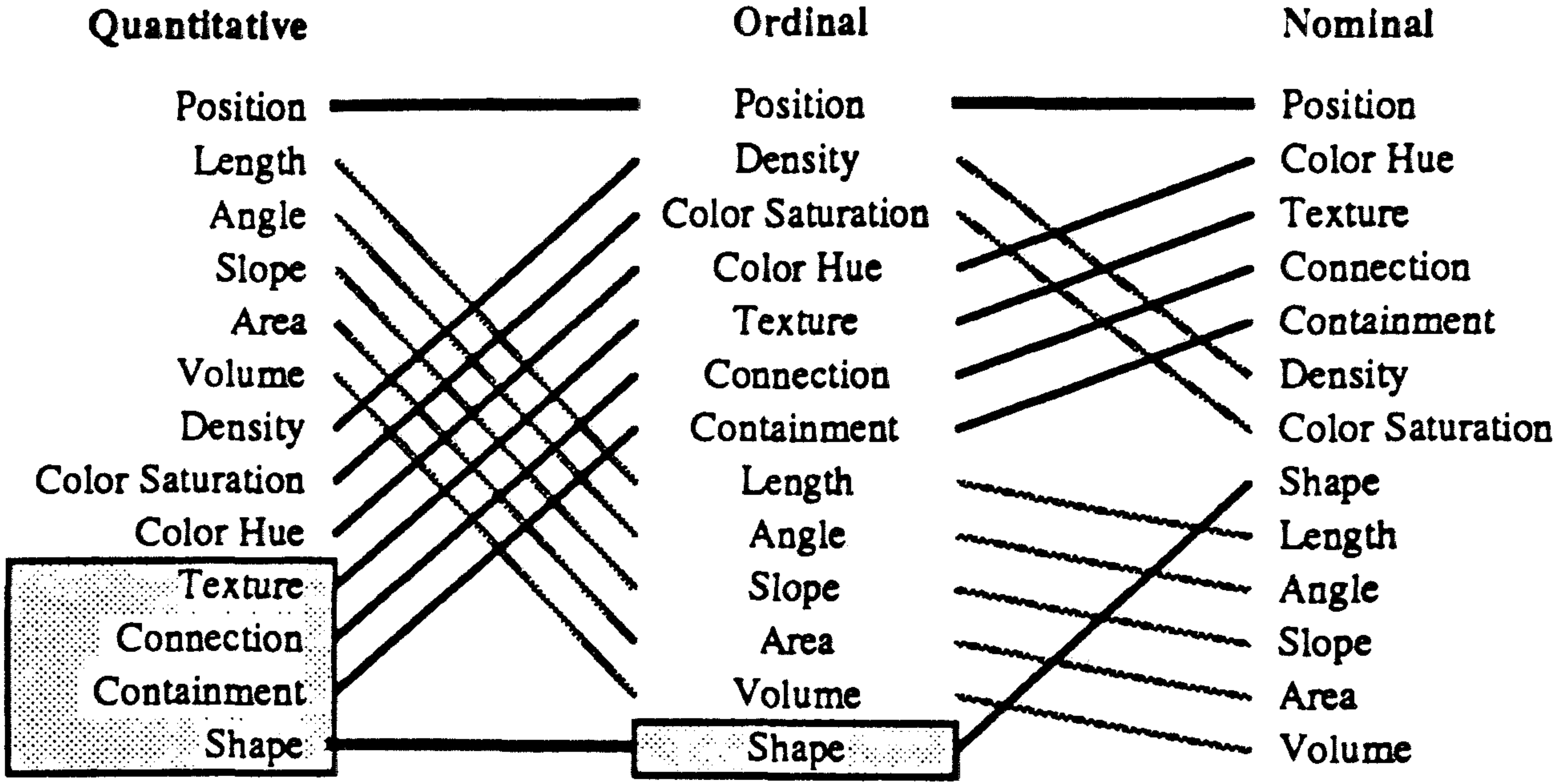
- Same stacked bar chart visualization
- Three tools
  - Tableau (free academic license)
  - Vega-Lite
  - D3
- For Vega-Lite, use the online editor
- For D3, use the template files so the data is properly loaded
- [CS 490] Only need to do a standard bar chart in D3



Expressiveness and Effectiveness

- Expressiveness Principle: all data from the dataset and nothing more should be shown
  - Do encode ordered data in an ordered fashion
  - Don't encode categorical data in a way that implies an ordering
- Effectiveness Principle: the most important attributes should be the most **salient**
  - Saliency: how noticeable something is
  - How do the channels we have discussed measure up?

Mackinlay's Ranking of Perceptual Tasks



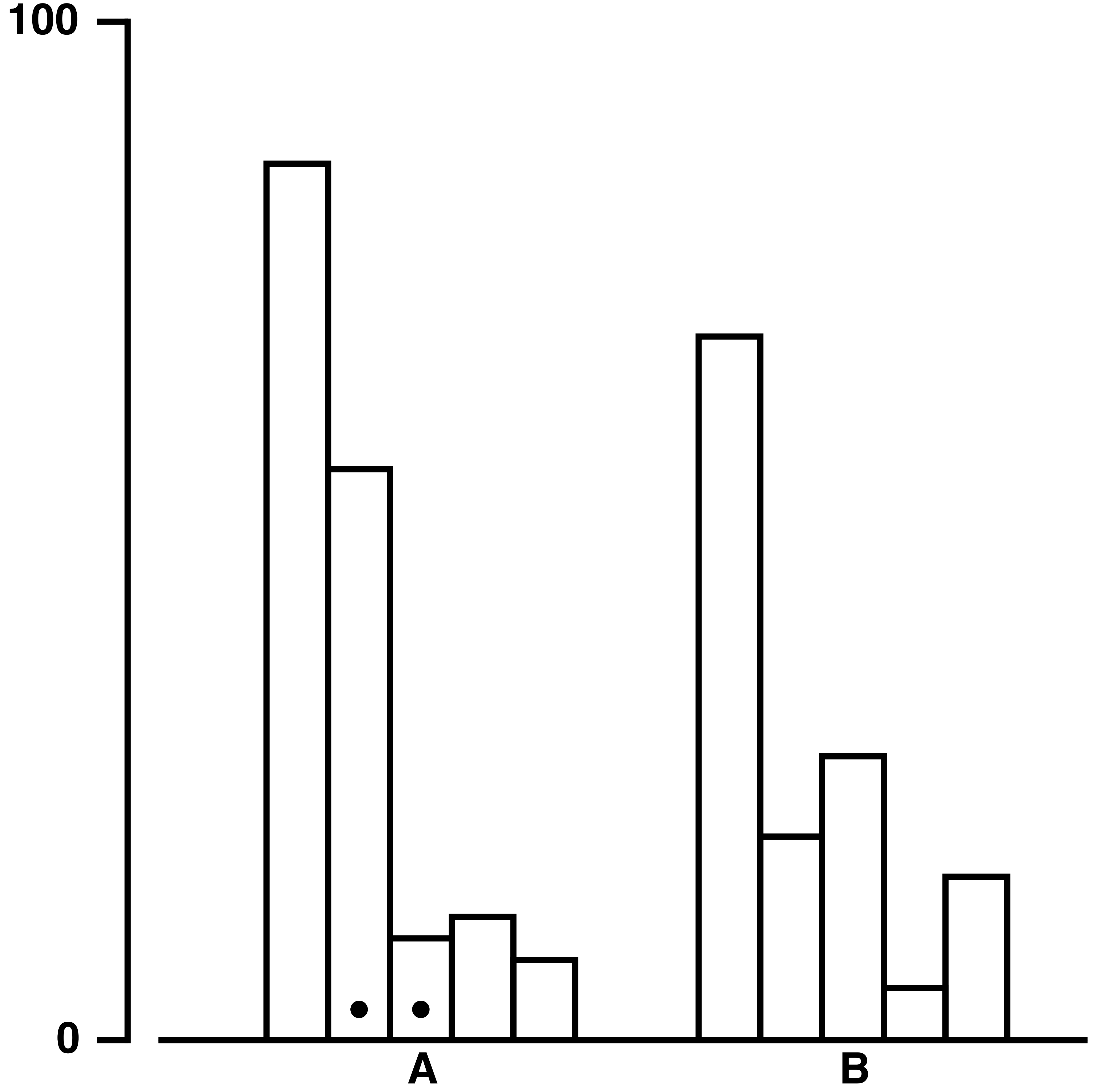
[Mackinlay,1986]

Iliinsky's Best Uses, +Ordering, +NumValues

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	**yes**	**infinite**	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	**optional** (alphabetical or numbered)	**infinite**	Good	Good	Good	Good
	length	**yes**	**many**	Good	Good		
	size, area	**yes**	**many**	Good	Good		
	angle	**yes**	medium/few	Good	Good		
	pattern density	**yes**	few	Good	Good		
	weight, boldness	**yes**	few		Good		
	saturation, brightness	**yes**	few		Good		
	color	no	few (< 20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	**infinite**			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	**yes**	few		Good		

How do we get these rankings?

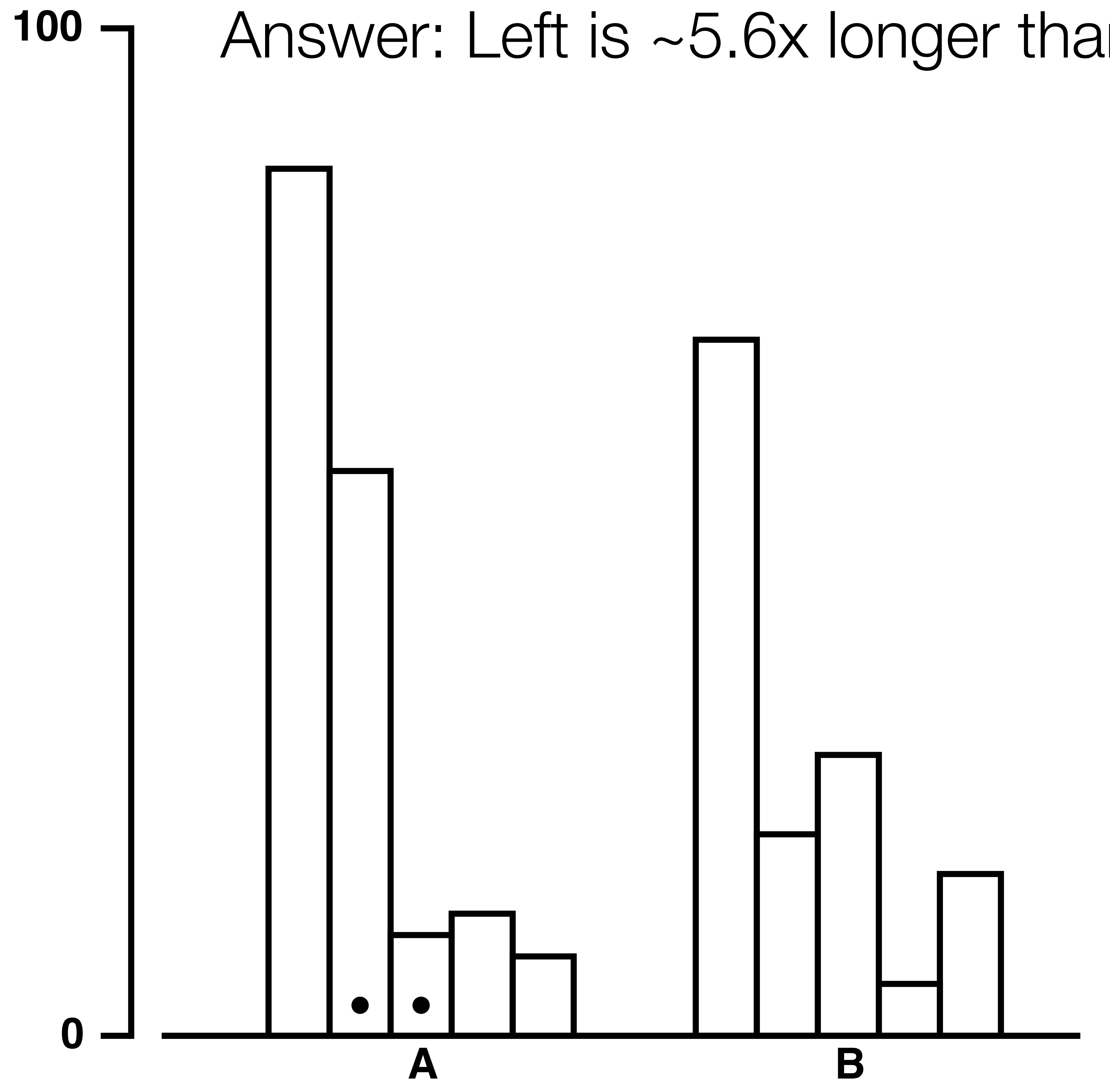
Test % difference in **length** between elements



[Heer & Bostock, 2010]

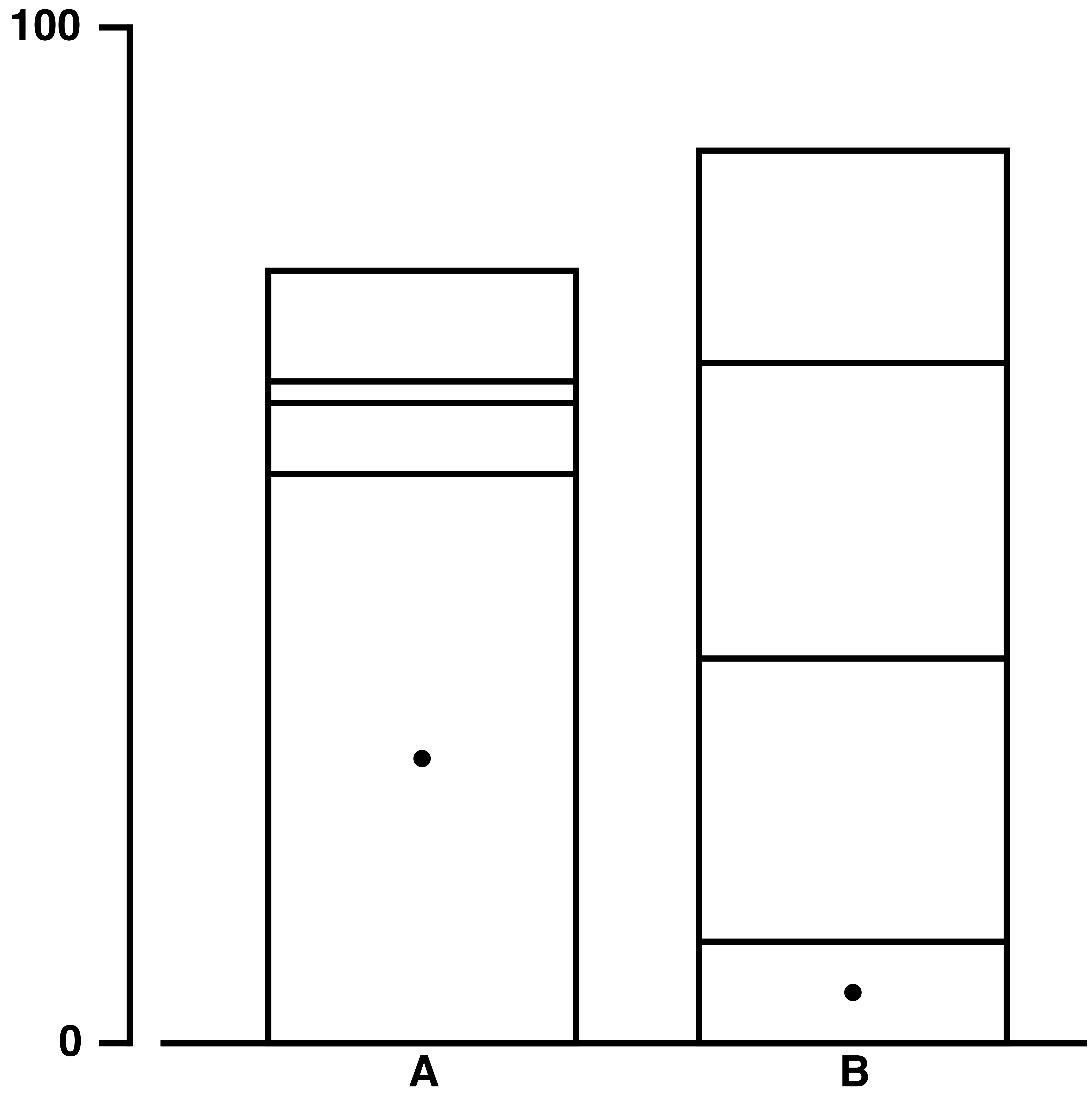
Test % difference in **length** between elements

Answer: Left is ~5.6x longer than Right



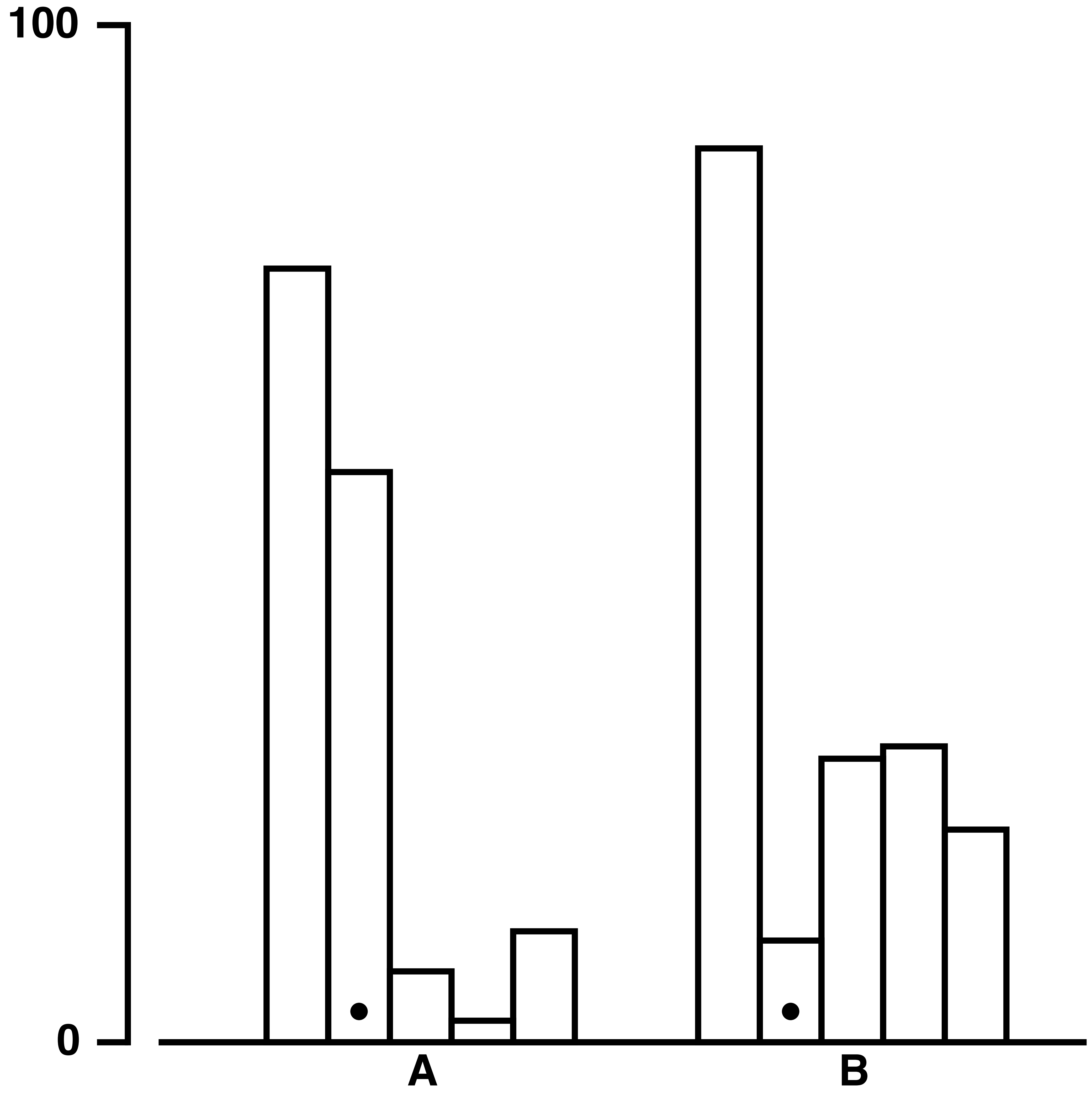
[Heer & Bostock, 2010]

Test % difference in **length** between elements



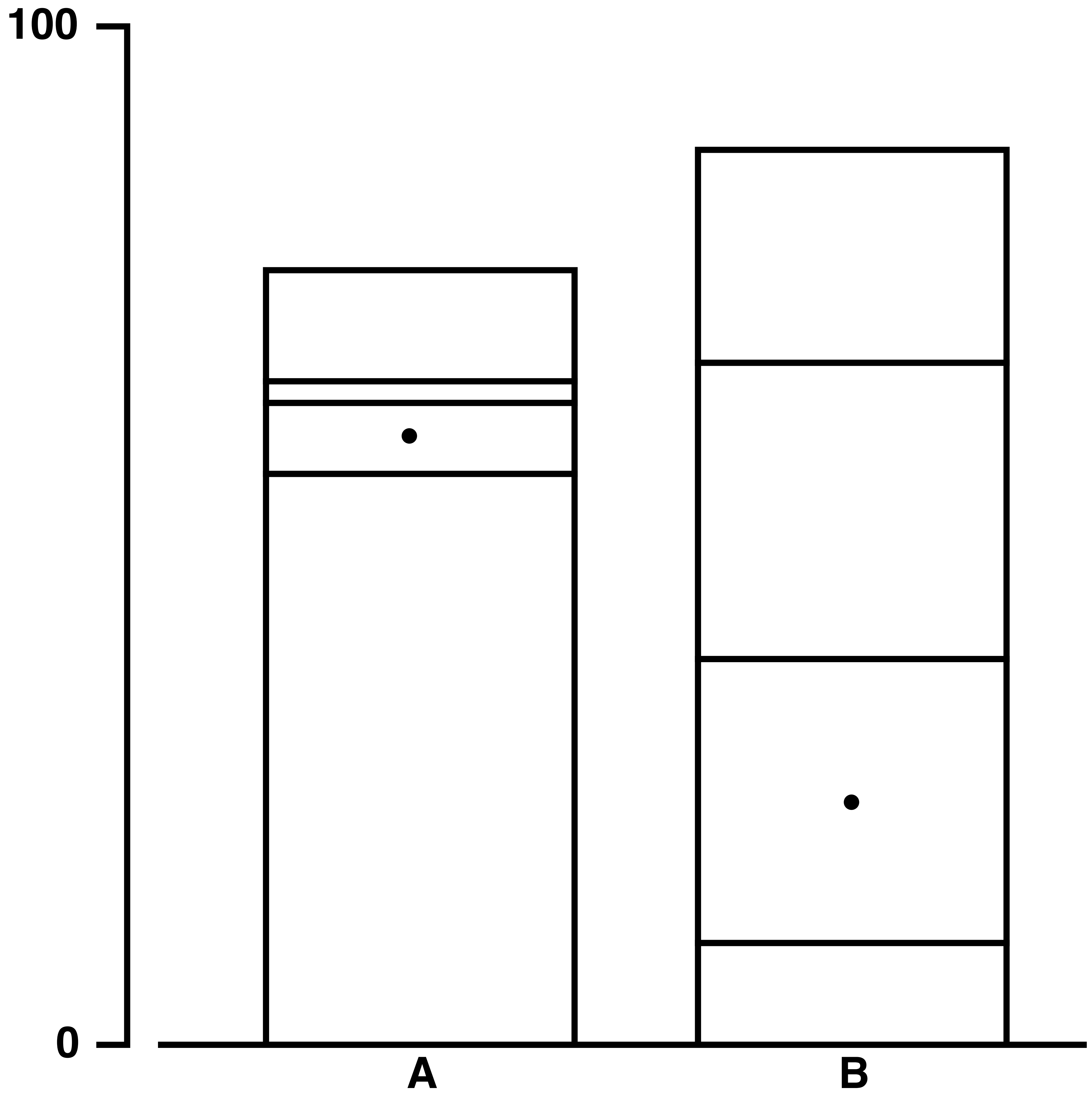
[Heer & Bostock, 2010]

Test % difference in **length** between elements



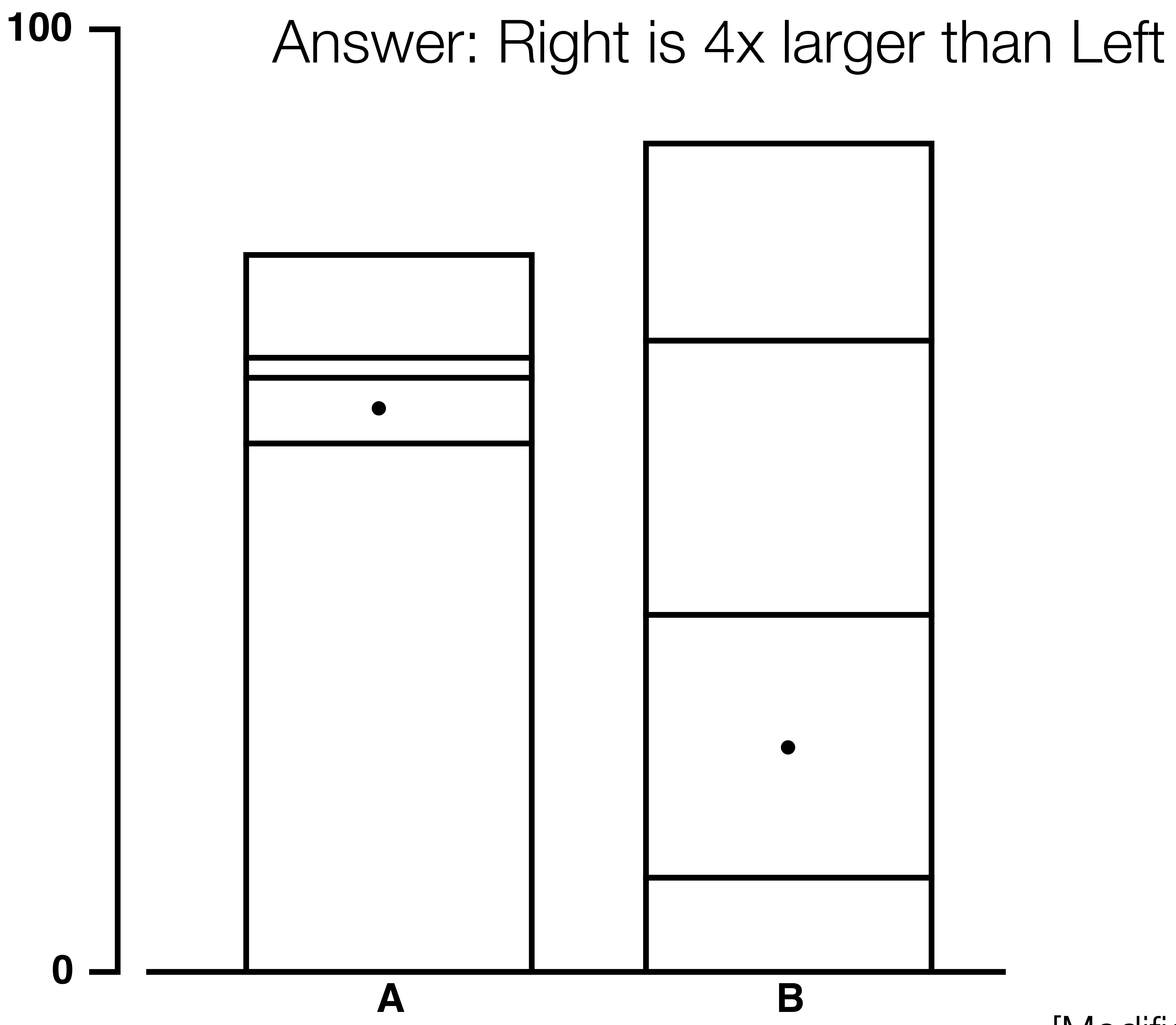
[Heer & Bostock, 2010]

Test % difference in **length** between elements



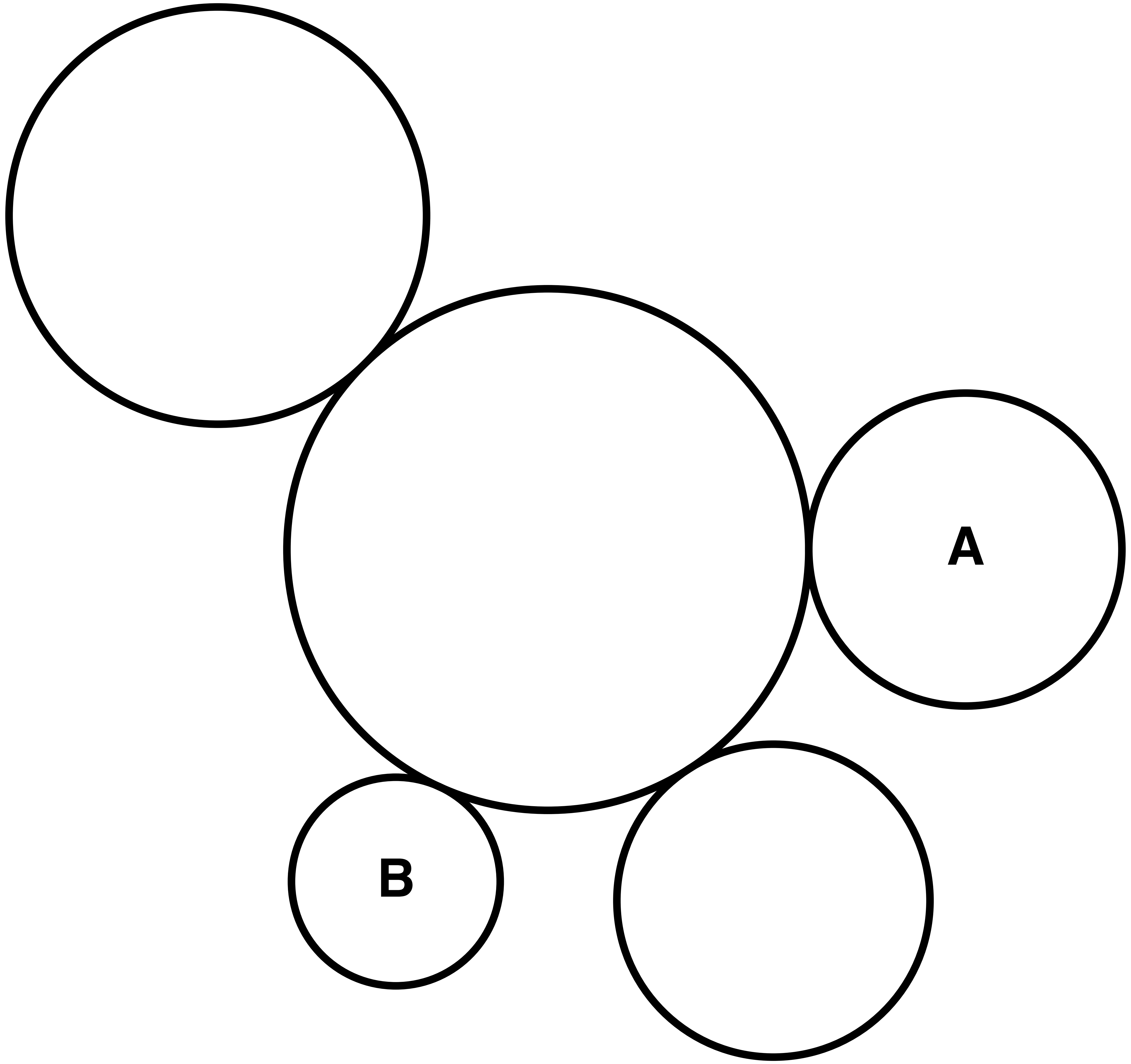
[Modified from Heer & Bostock, 2010]

Test % difference in **length** between elements



[Modified from Heer & Bostock, 2010]

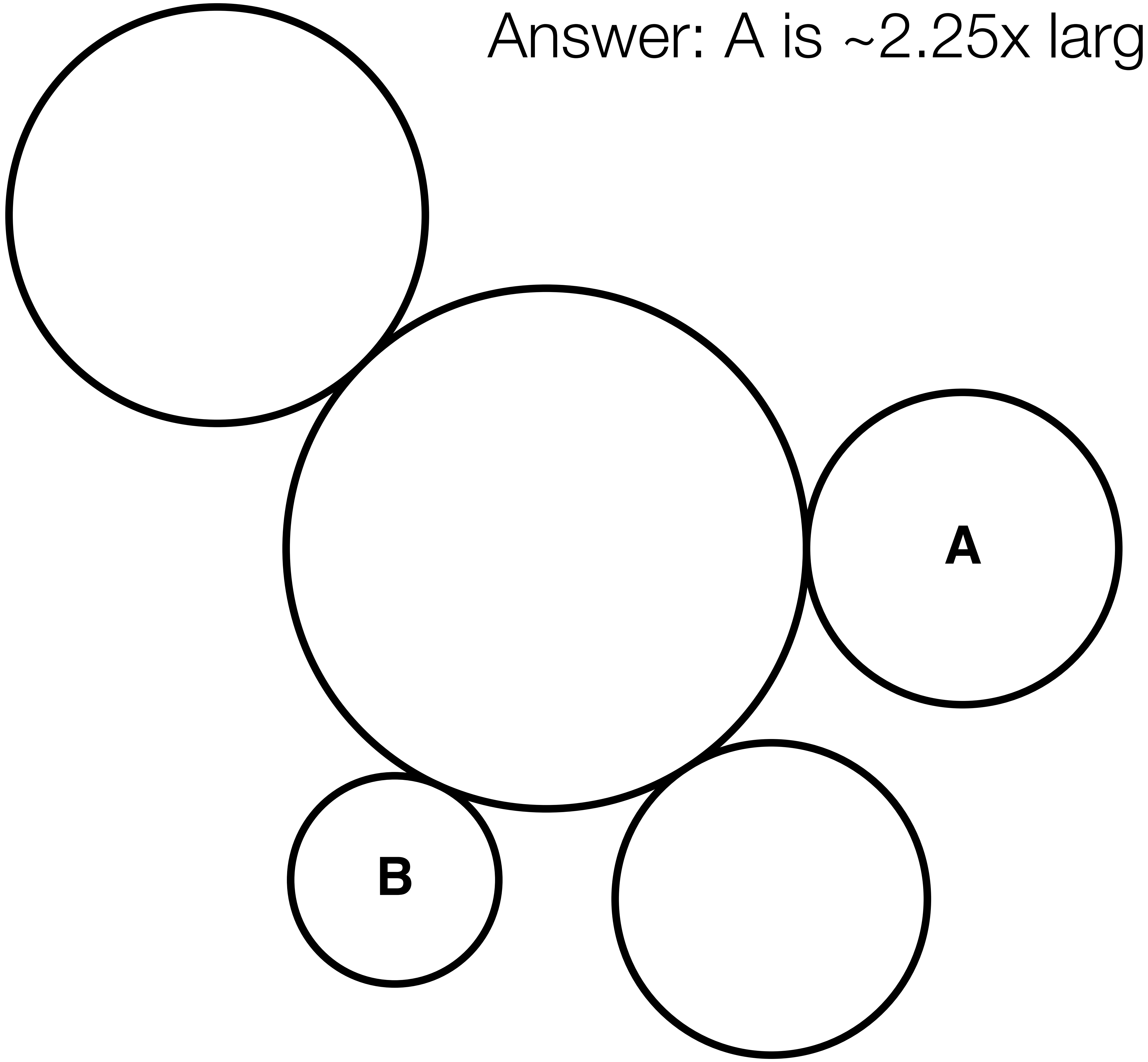
Test % difference in **area** between elements



[Heer & Bostock, 2010]

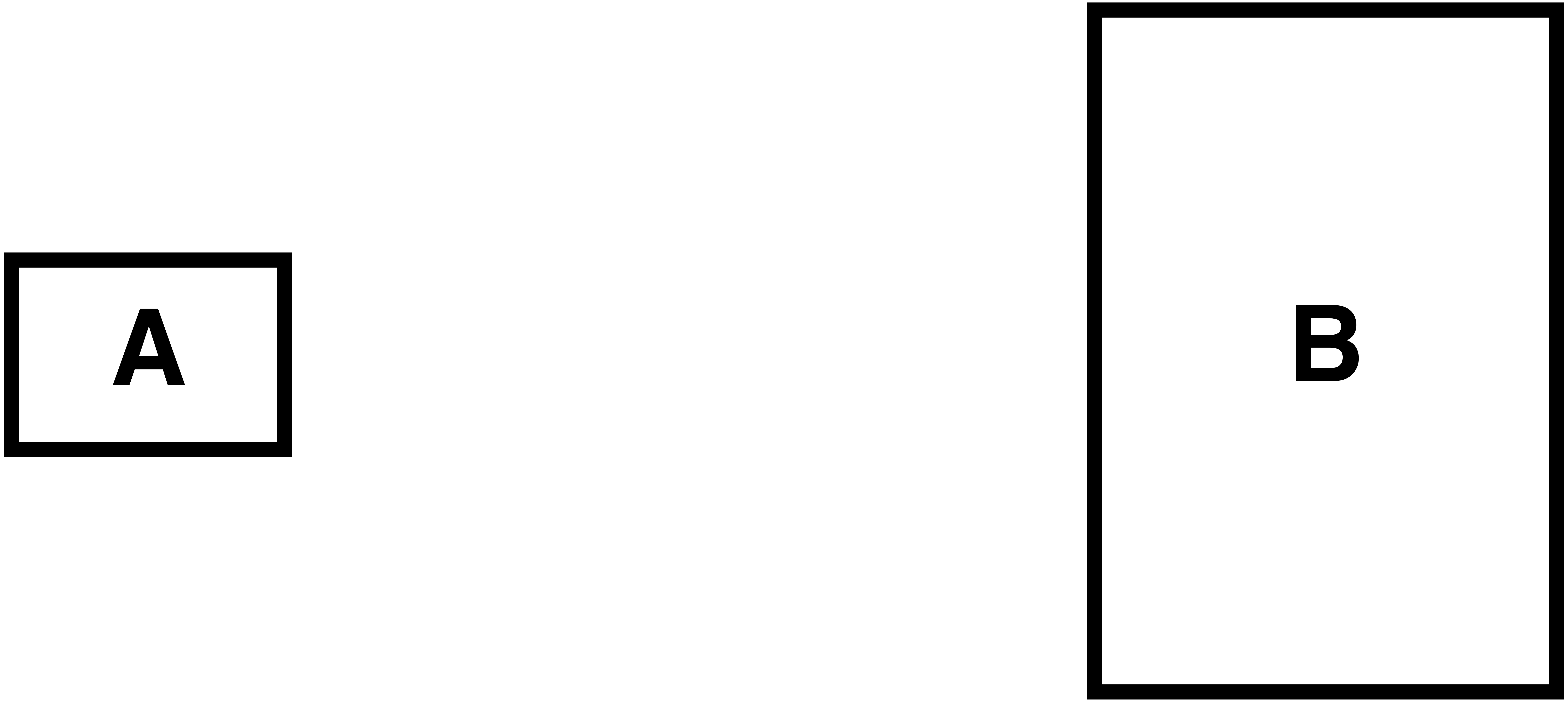
Test % difference in **area** between elements

Answer: A is ~2.25x larger (in area) than B



[Heer & Bostock, 2010]

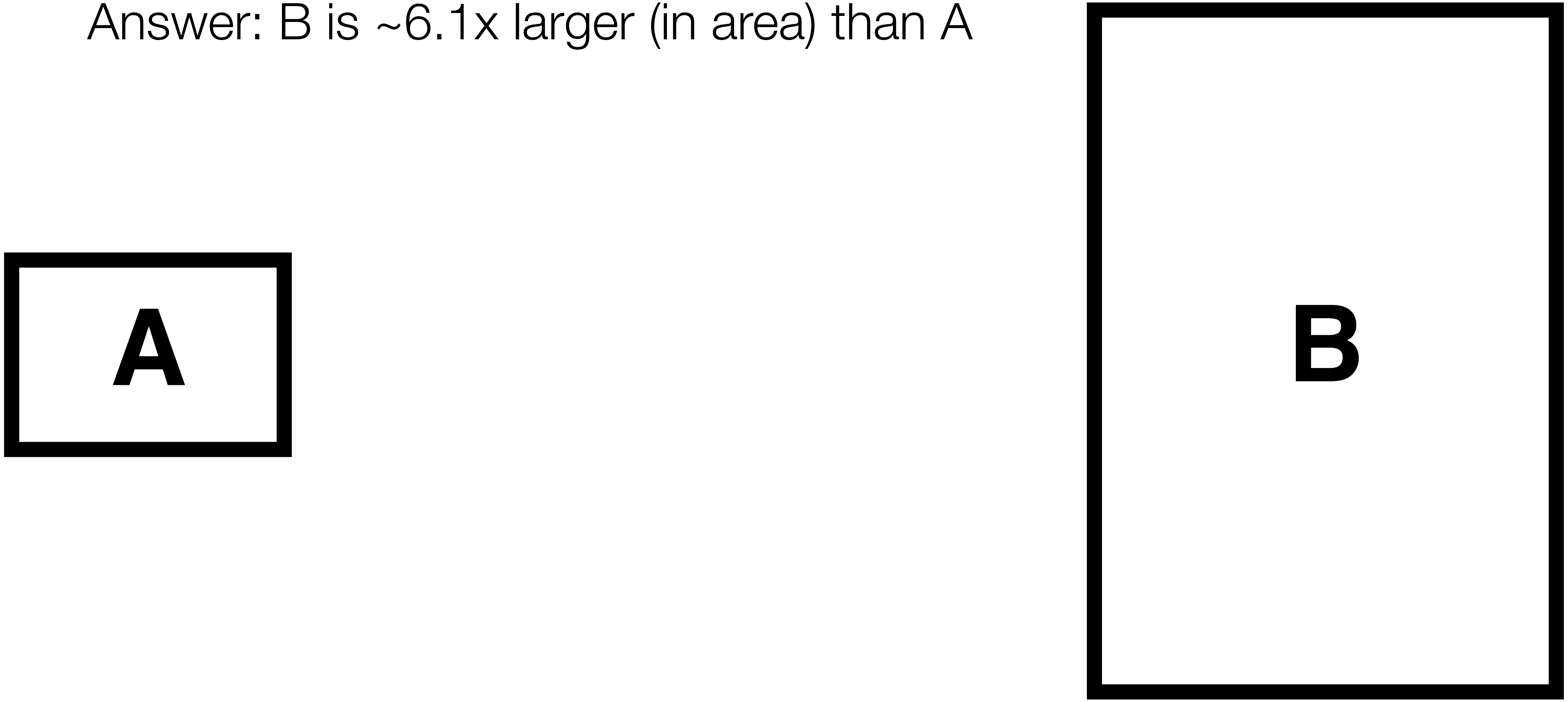
Test % difference in **area** between elements



[Heer & Bostock, 2010]

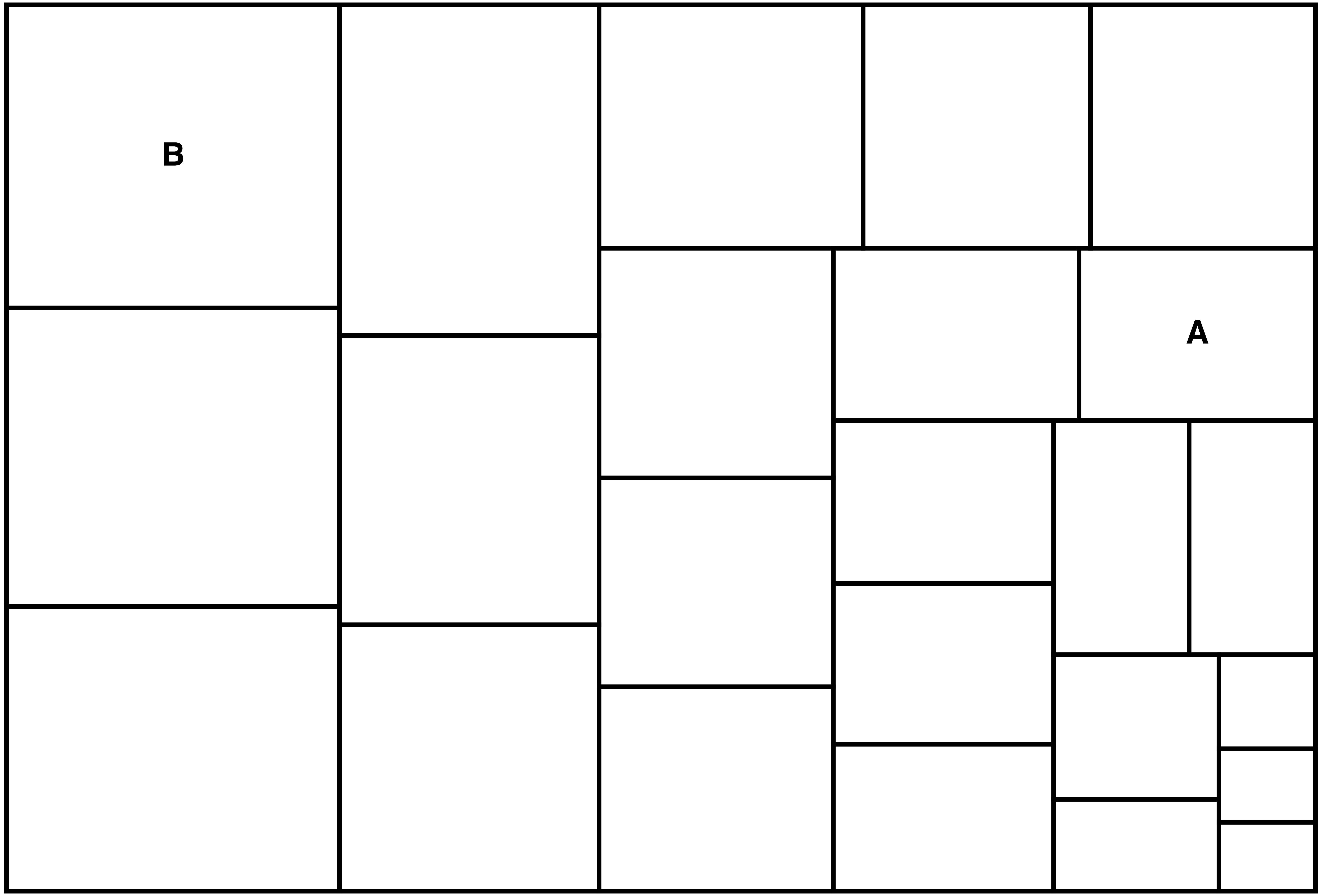
Test % difference in **area** between elements

Answer: B is ~6.1x larger (in area) than A



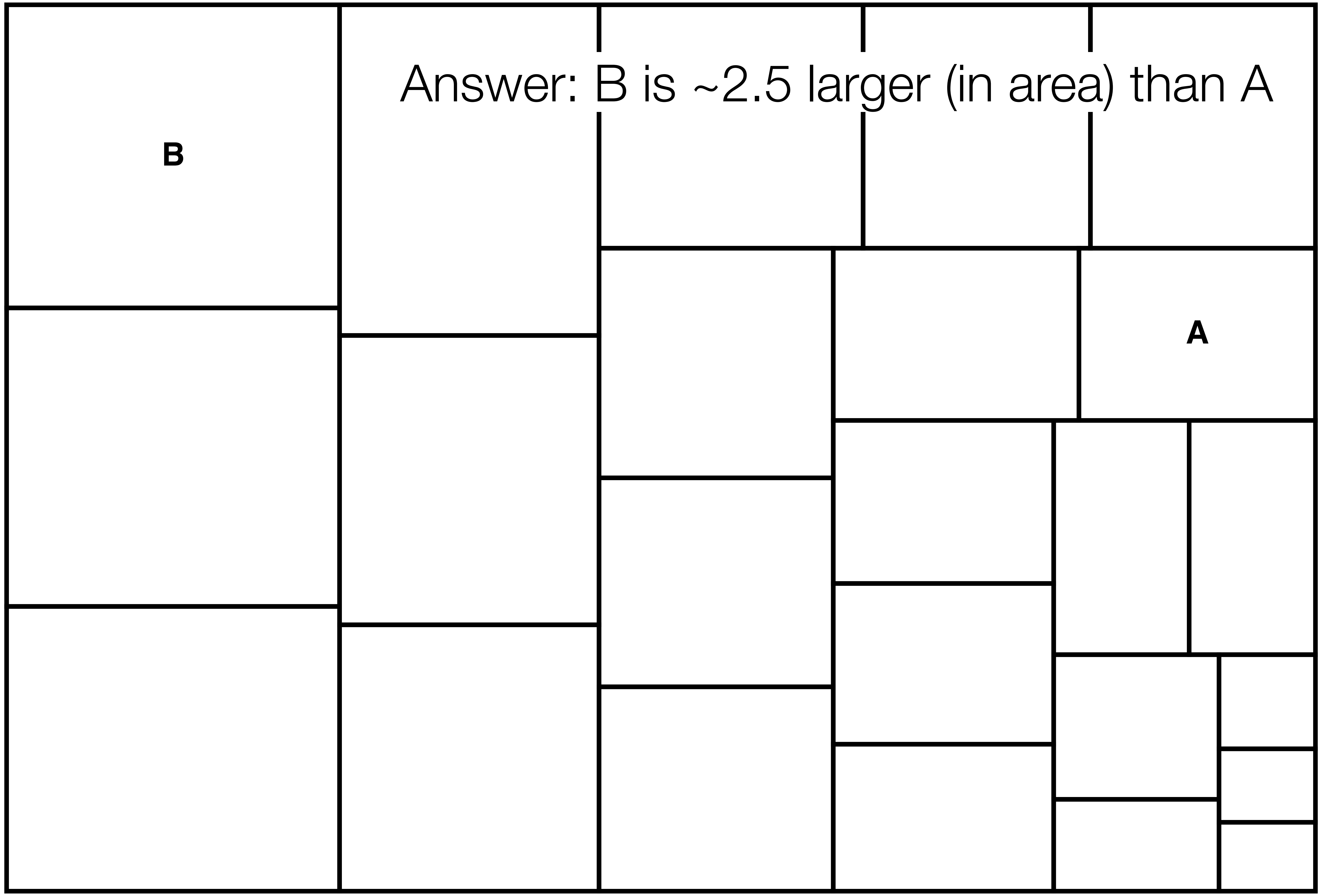
[Heer & Bostock, 2010]

Test % difference in **area** between elements



[Heer & Bostock, 2010]

Test % difference in **area** between elements



Answer: B is ~2.5 larger (in area) than A

[Heer & Bostock, 2010]

Cleveland & McGill Experiments

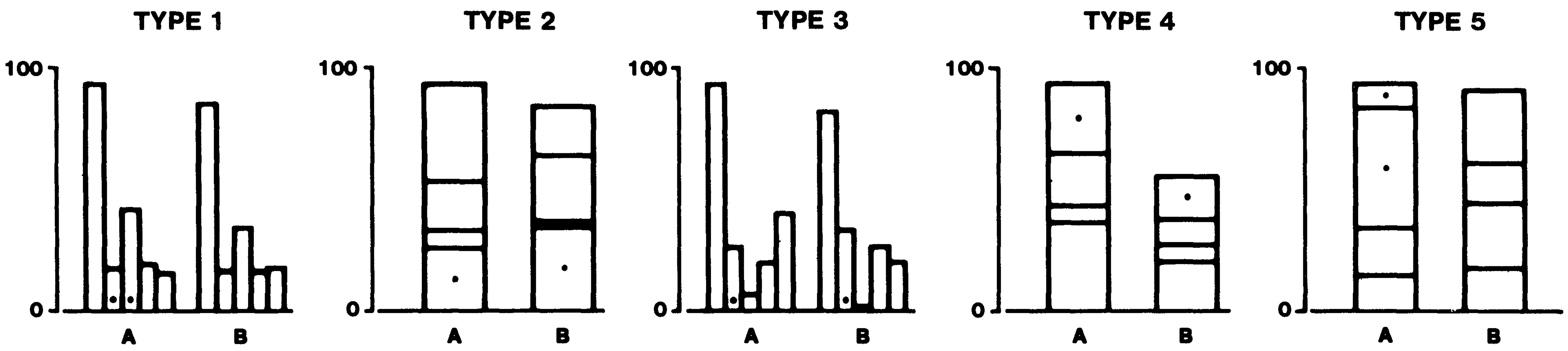


Figure 4. Graphs from position–length experiment.

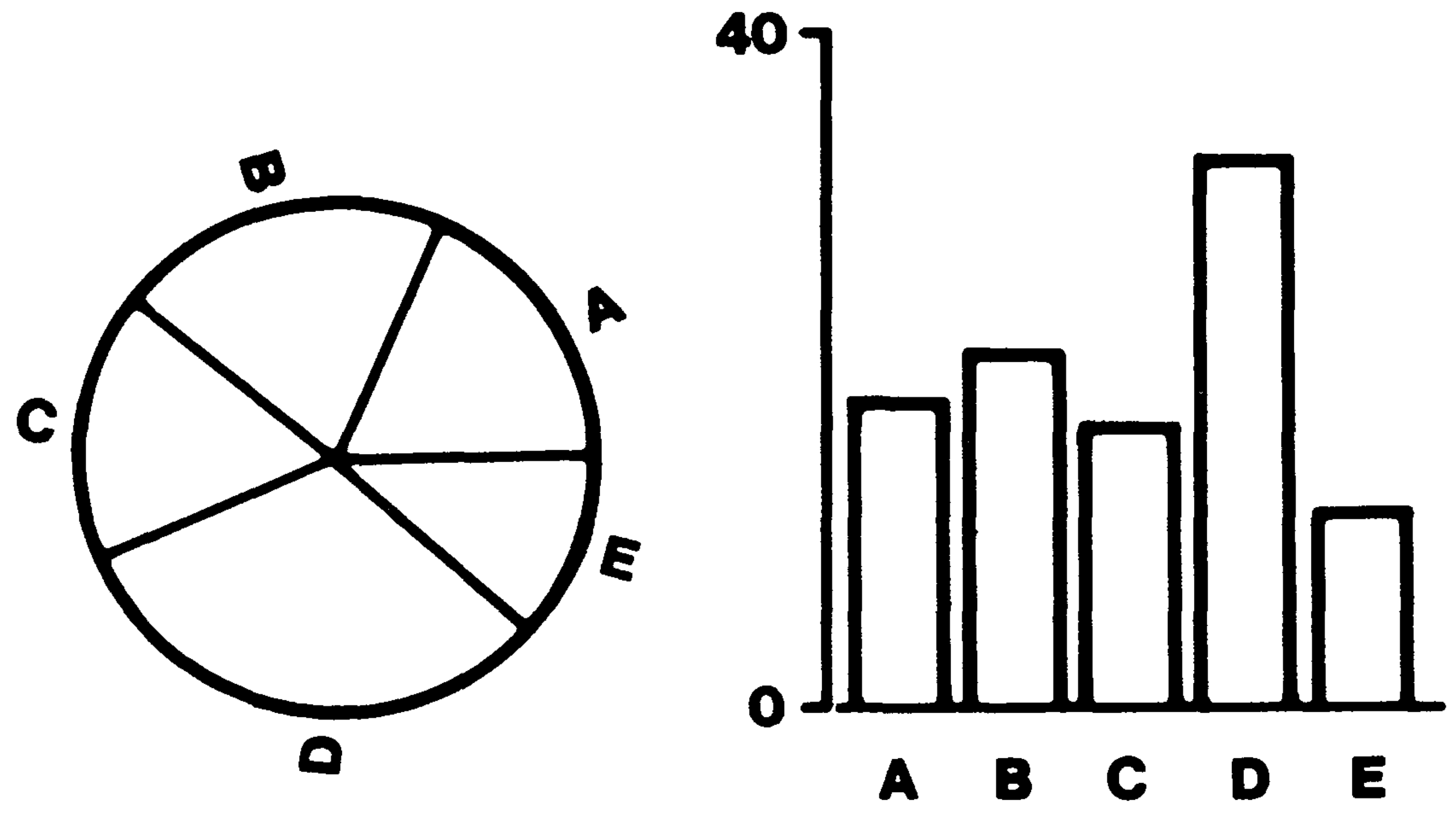


Figure 3. Graphs from position–angle experiment.

[Cleveland & McGill, 1984]

Heer & Bostock Experiments

- Rerun Cleveland & McGill's experiment using Mechanical Turk
- … with more tests

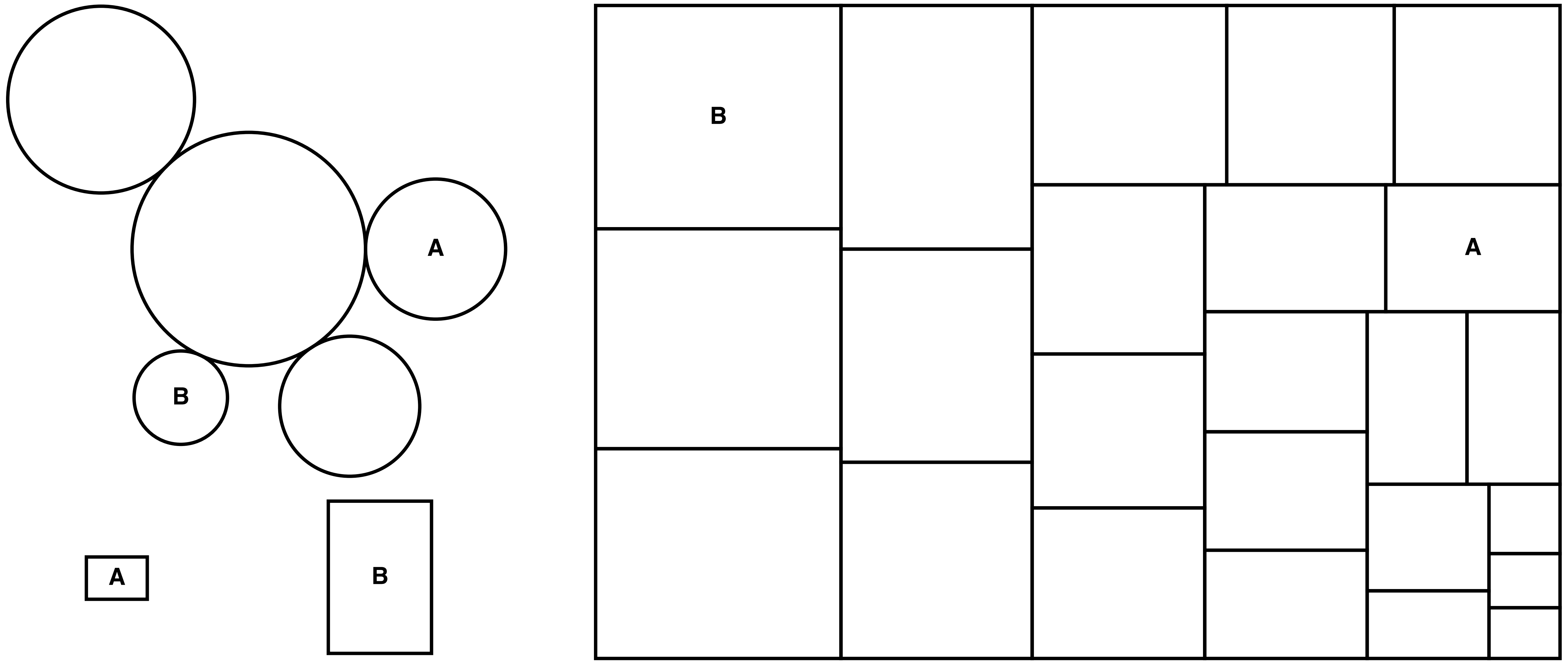
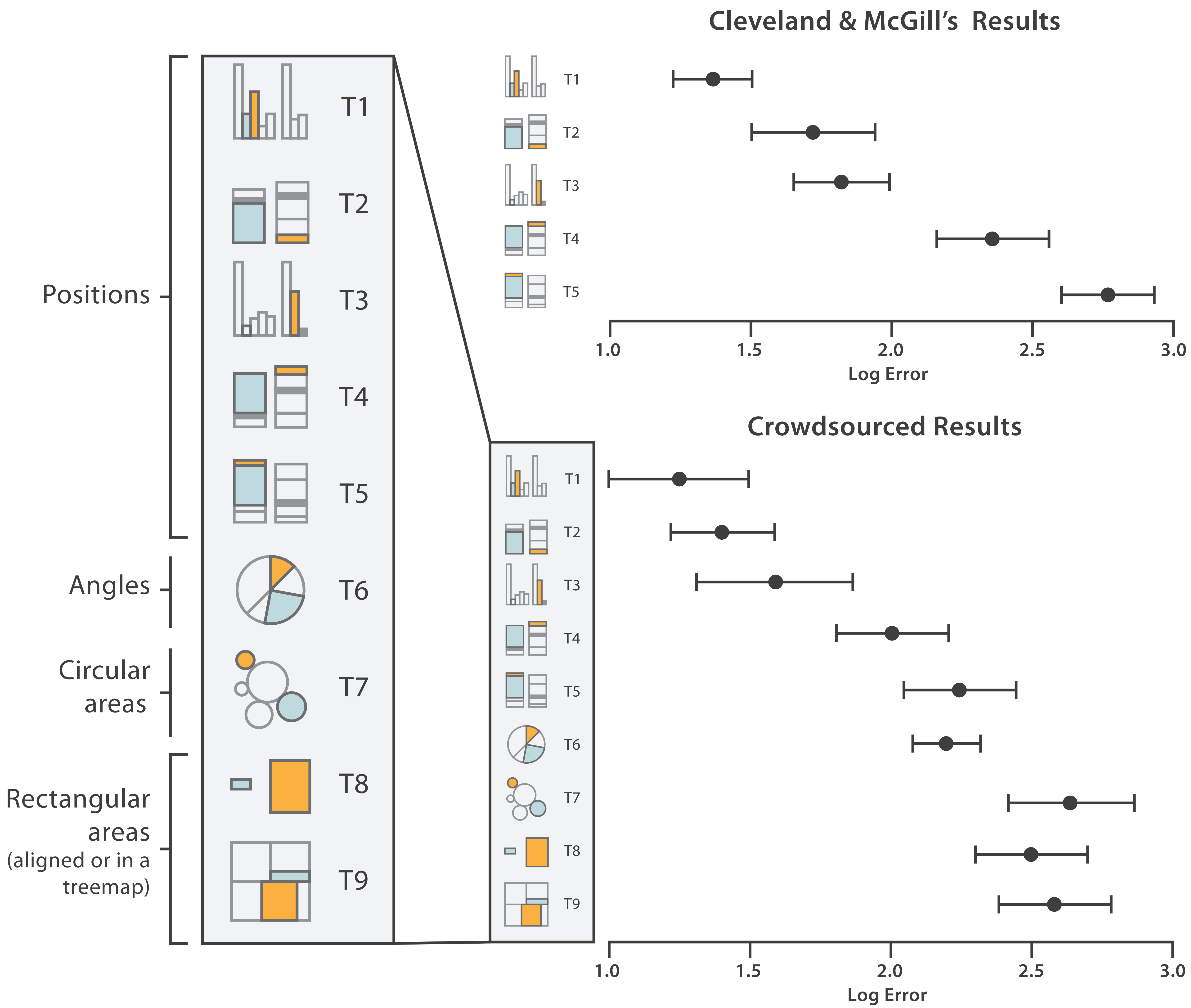


Figure 2: Area judgment stimuli. Top left: Bubble chart (T7), Bottom left: Center-aligned rectangles (T8), Right: Treemap (T9).

[Heer & Bostock, 2010]

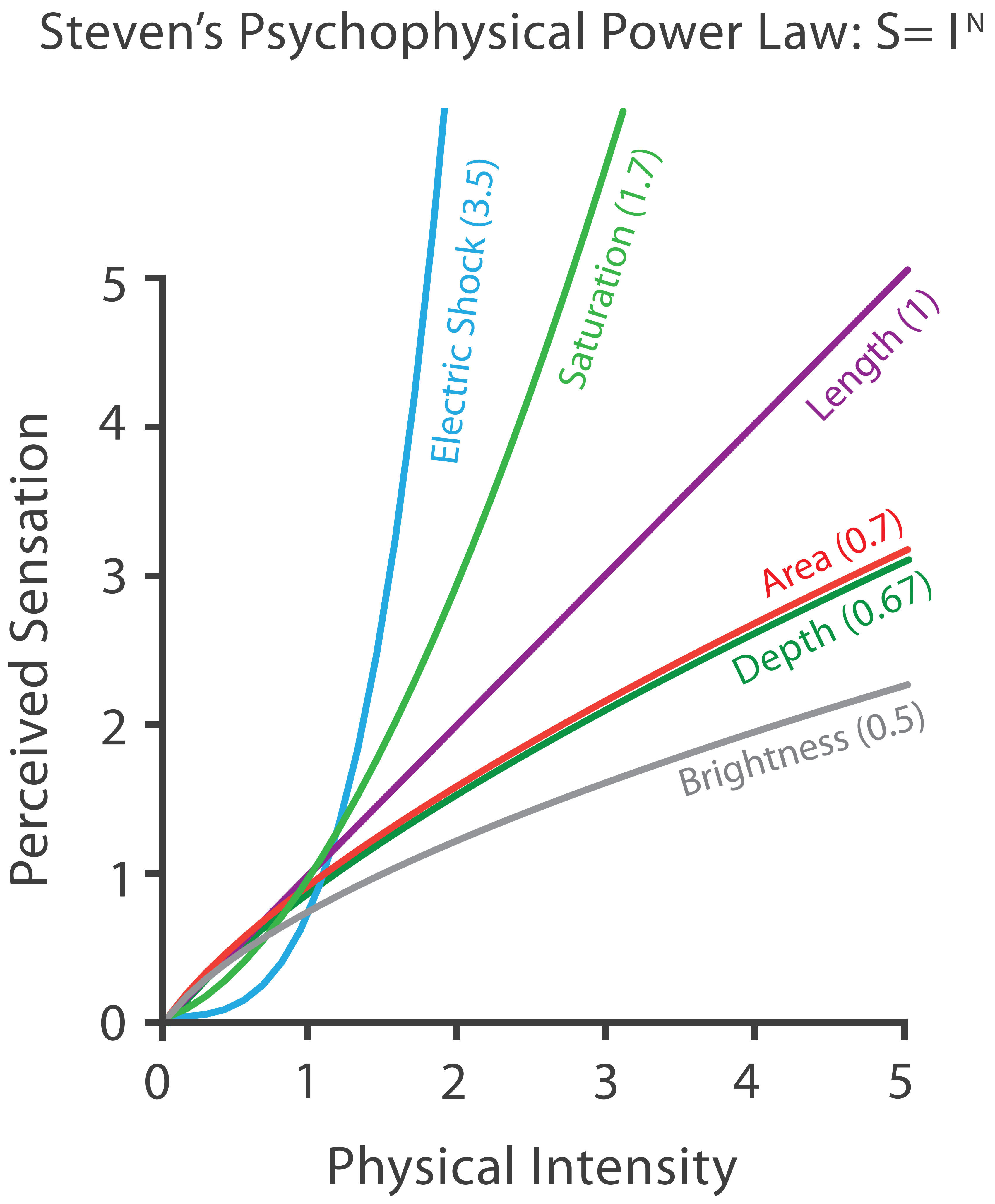
Results Summary



[Munzner (ill. Maguire) based on Heer & Bostock, 2014]

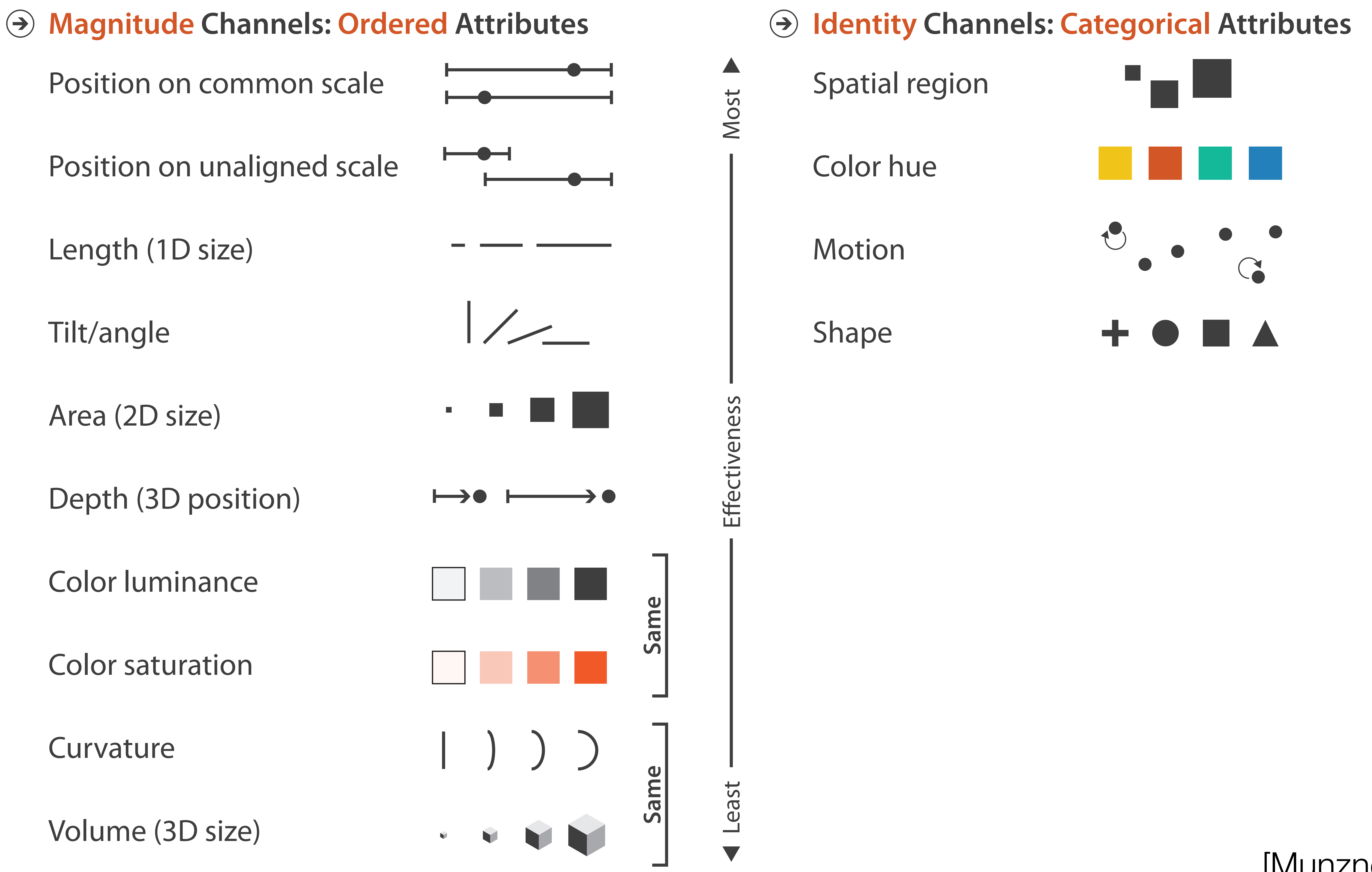
Psychophysics

- How do we perceive changes in stimuli
- The Psychophysical Power Law [Stevens, 1975]: All sensory channels follow a power function based on stimulus intensity ($S = I^n$)
- Length is fairly accurate
- Magnified vs. compressed sensations



[Munzner (ill. Maguire), 2014]

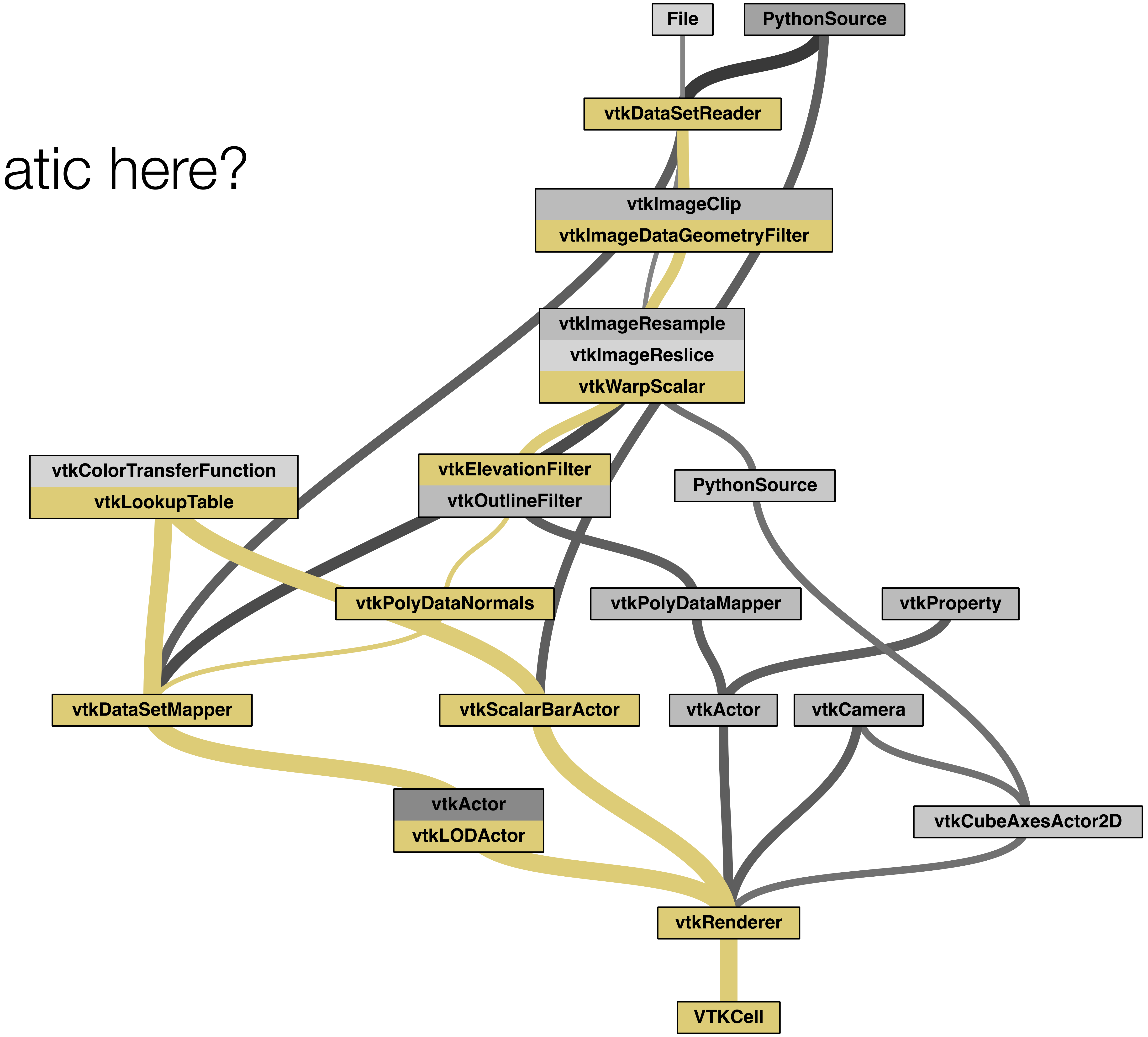
Ranking Channels by Effectiveness



[Munzner (ill. Maguire), 2014]

Discriminability

What is problematic here?



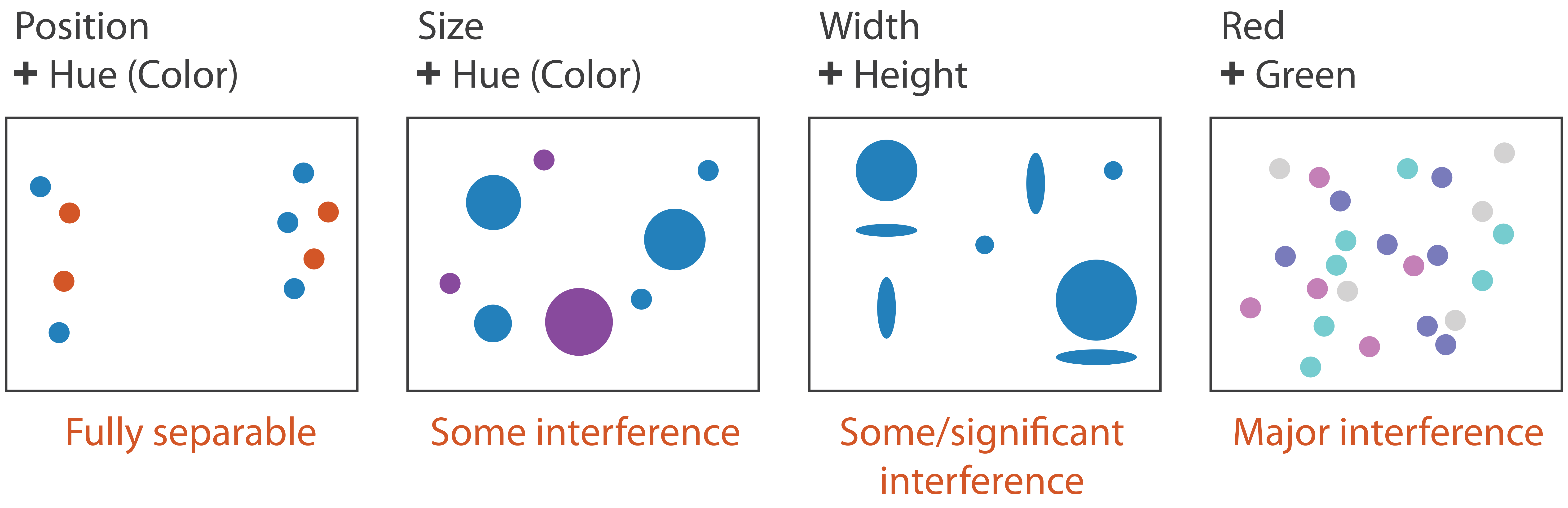
[Koop et al., 2013]

Discriminability

- Can someone tell the difference?
- How many values (bins) can be used so that a person can tell the difference?
- Example: Line width
  - Matching a particular width with a legend
  - Comparing two widths

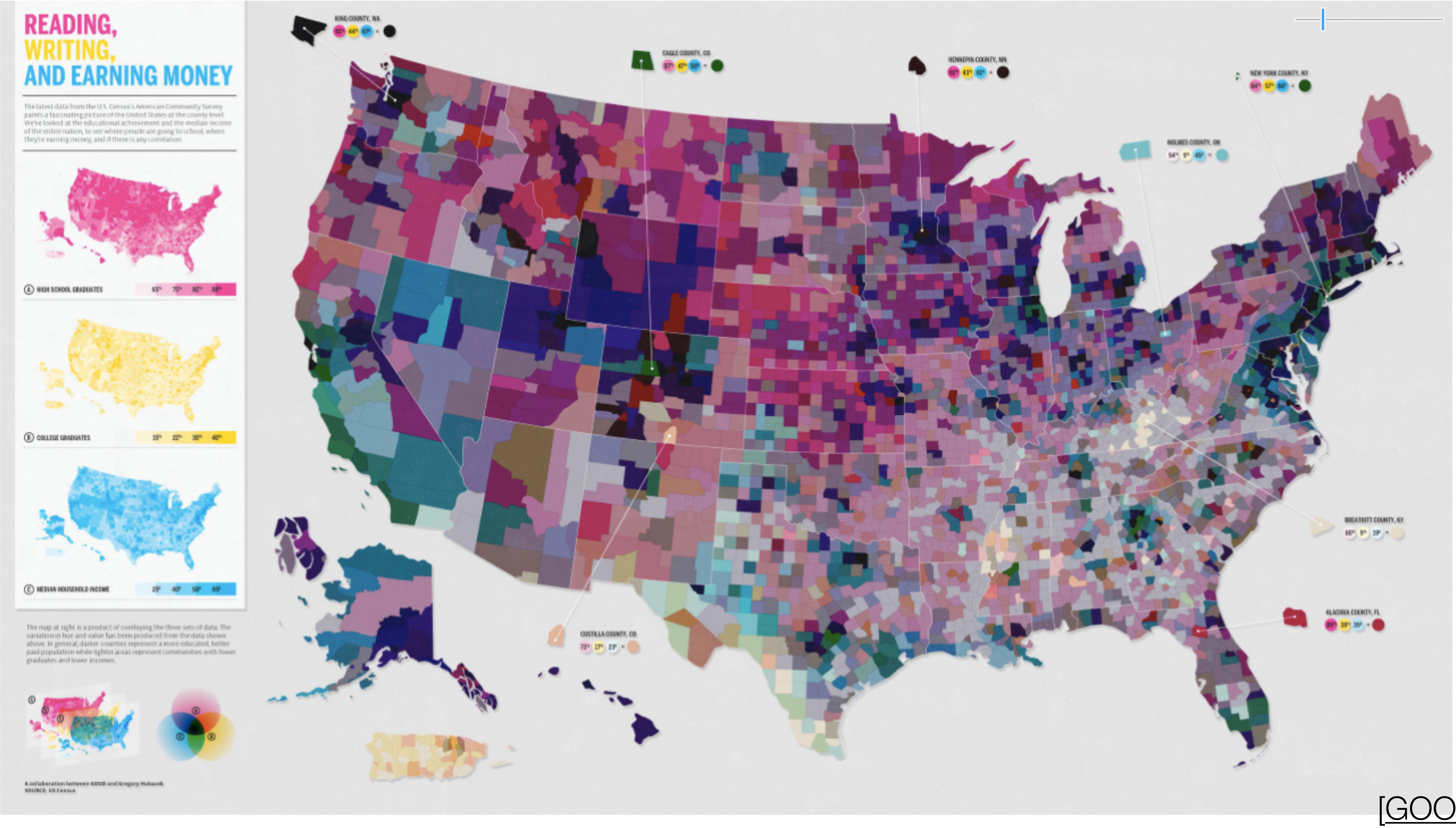
Separability

- Cannot treat all channels as independent!
- **Separable** means each individual channel can be distinguished
- **Integral** means the channels are perceived together



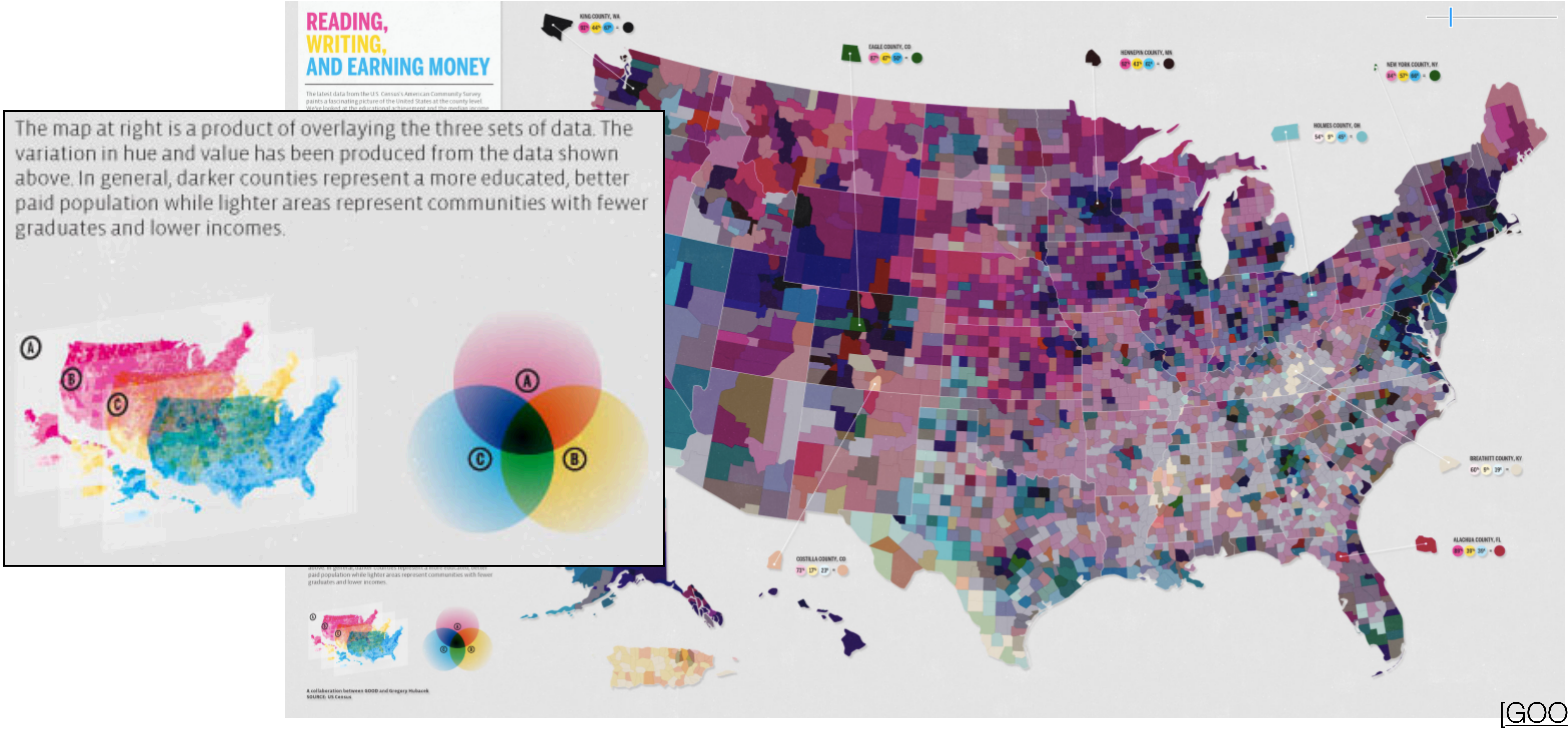
[Munzner (ill. Maguire) based on Ware, 2014]

Separable or Integral?



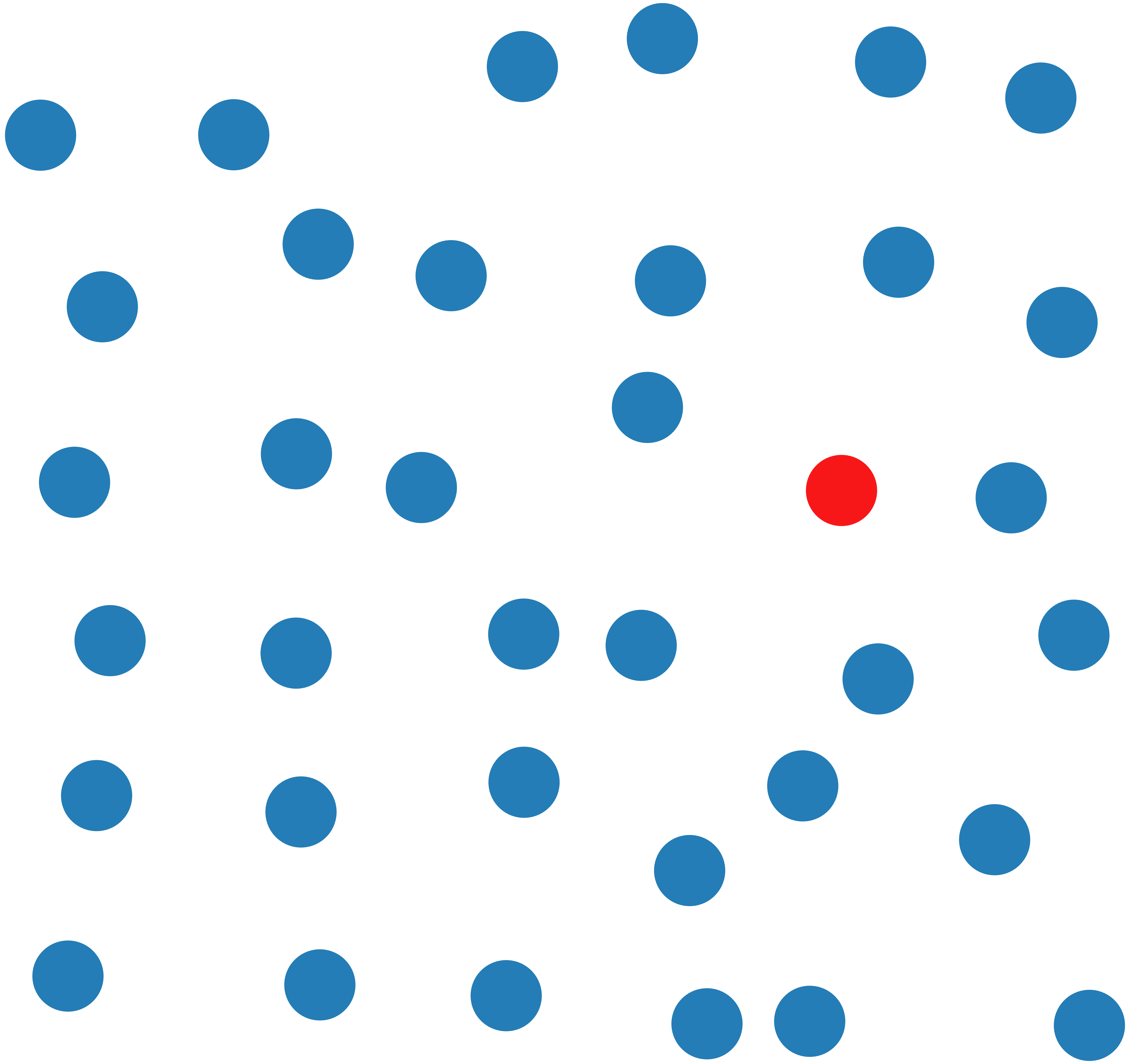
[GOOD]

Separable or Integral?



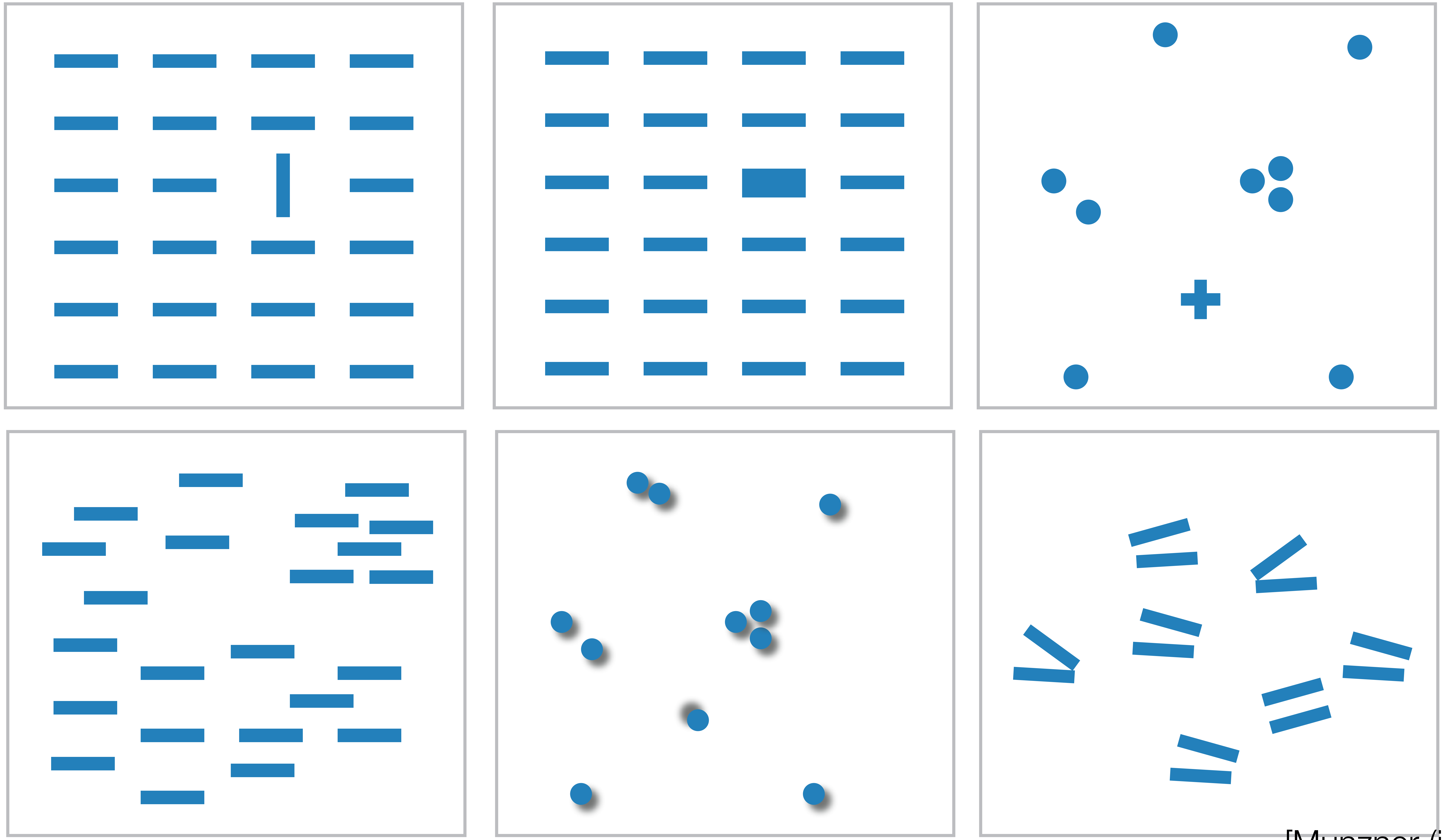
[GOOD]

Visual Popout



[C. G. Healey]

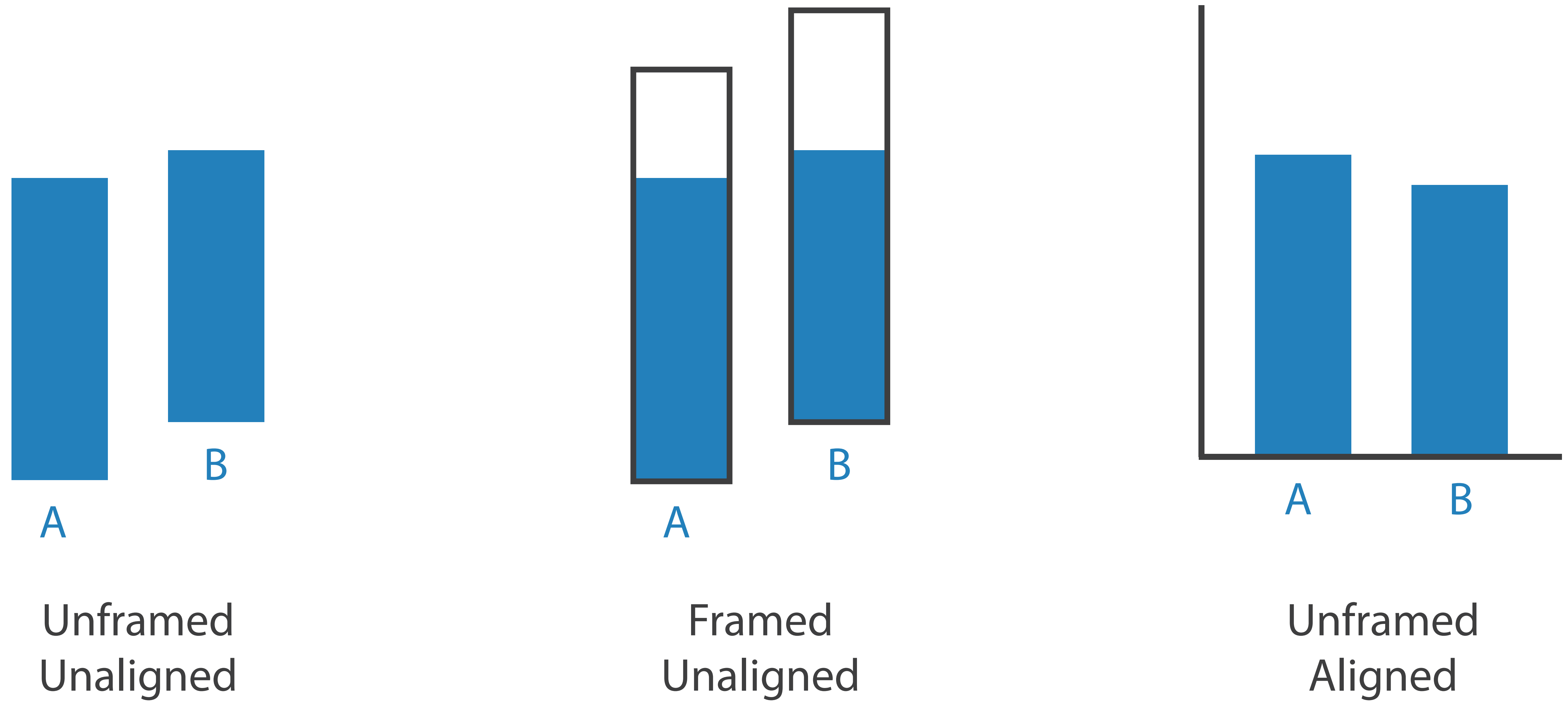
Visual Popout: Parallel Lines Require Search…



[Munzner (ill. Maguire), 2014]

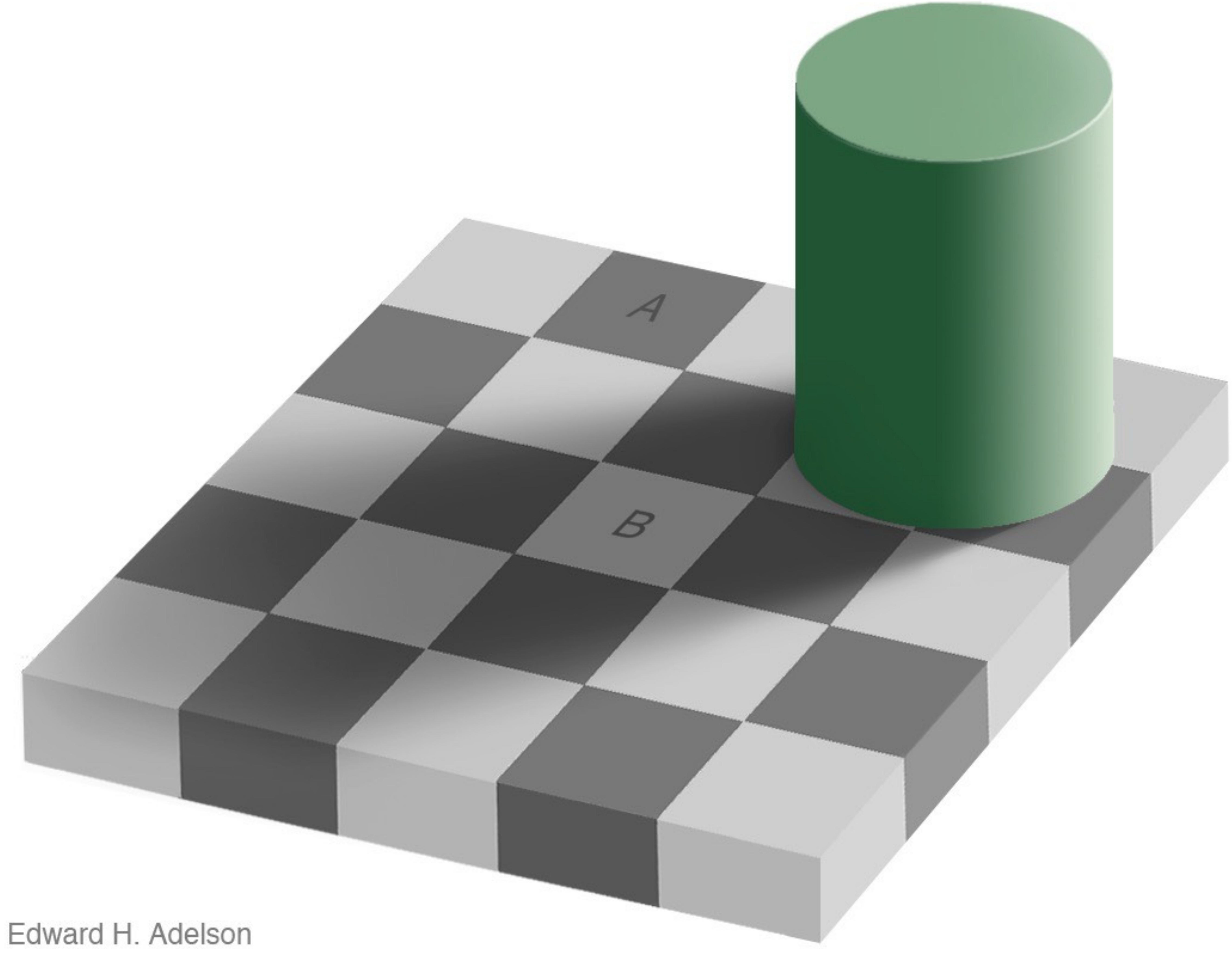
Relative vs. Absolute Judgments

- Weber's Law:
  - We judge based on relative not absolute differences
  - The amount of perceived difference is relative to the object's magnitude!



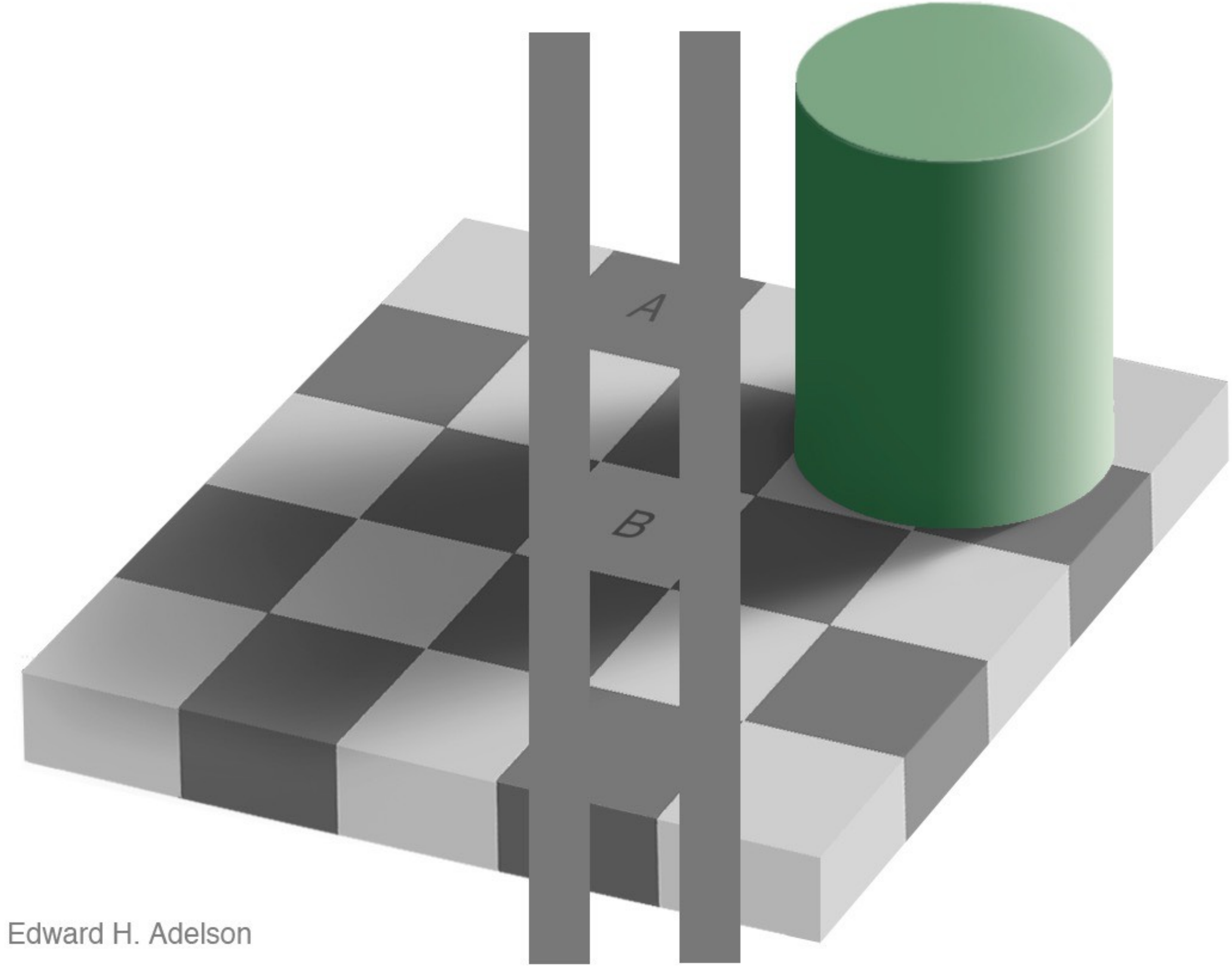
[Munzner (ill. Maguire), 2014]

Luminance Perception



[E. H. Adelson, 1995]

Luminance Perception



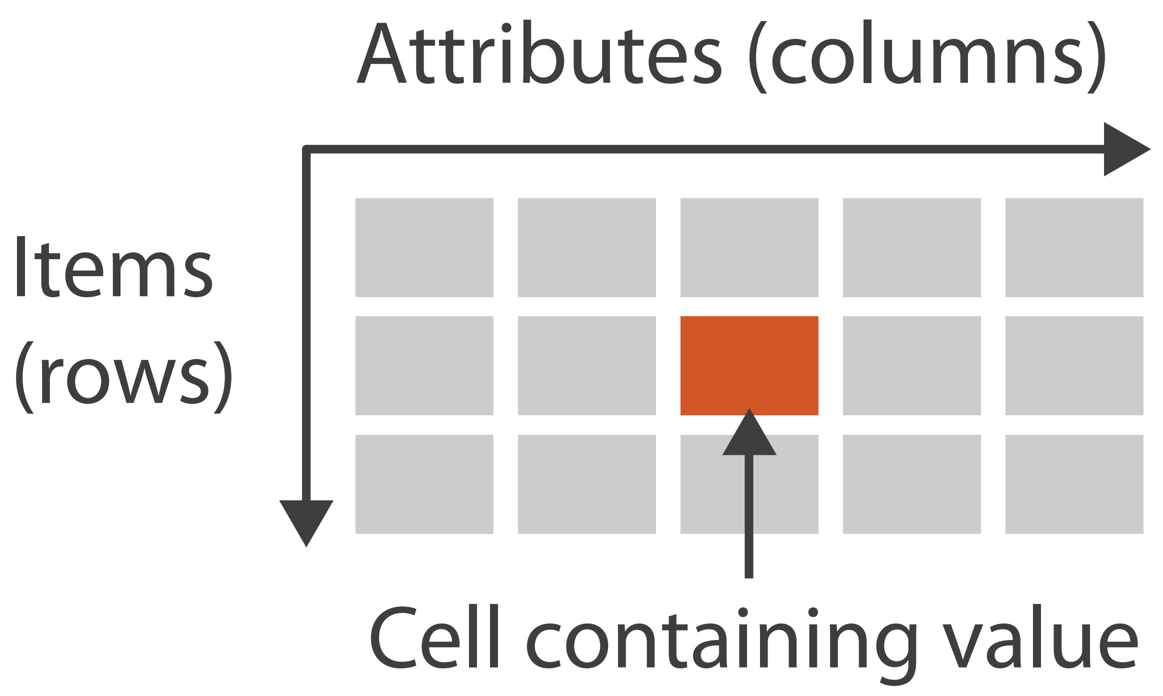
[E. H. Adelson, 1995]

Tables

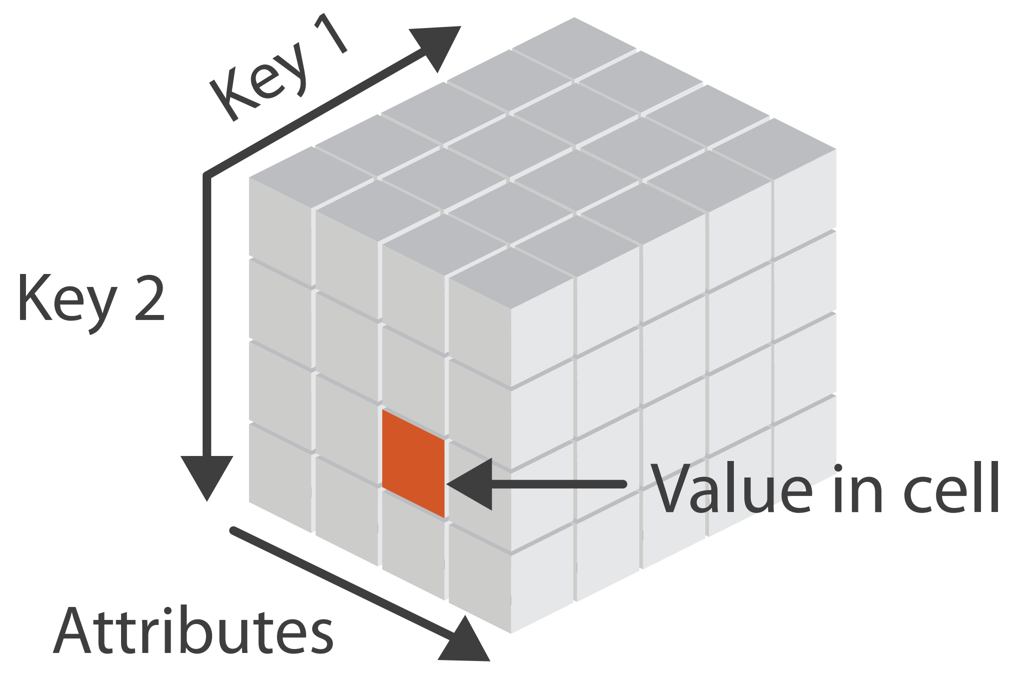
	REMOTE	STATION	FF ▼	SEN/DIS	7-D AFAS UNL	D AFAS/RMF I	JOINT RR TKT	7-D UNL	30-D UNL
1	R011	42ND STREET & 8TH AVENUE	00228985	00008471	00000441	00001455	00000134	00033341	00071255
2	R170	14TH STREET-UNION SQUARE	00224603	00011051	00000827	00003026	00000660	00089367	00199841
3	R046	42ND STREET & GRAND CENTRAL	00207758	00007908	00000323	00001183	00003001	00040759	00096613
4	R012	34TH STREET & 8TH AVENUE	00188311	00006490	00000498	00001279	00003622	00035527	00067483
5	R293	34TH STREET - PENN STATION	00168768	00006155	00000523	00001065	00005031	00030645	00054376
6	R033	42ND STREET/TIMES SQUARE	00159382	00005945	00000378	00001205	00000690	00058931	00078644
7	R022	34TH STREET & 6TH AVENUE	00156008	00006276	00000487	00001543	00000712	00058910	00110466
8	R084	59TH STREET/COLUMBUS CIRCLE	00155262	00009484	00000589	00002071	00000542	00053397	00113966
9	R020	47-50 STREETS/ROCKEFELLER	00143500	00006402	00000384	00001159	00000723	00037978	00090745
10	R179	86TH STREET-LEXINGTON AVE	00142169	00010367	00000470	00001839	00000271	00050328	00125250
11	R023	34TH STREET & 6TH AVENUE	00134052	00005005	00000348	00001112	00000649	00031531	00075040
12	R029	PARK PLACE	00121614	00004311	00000287	00000931	00000792	00025404	00065362
13	R047	42ND STREET & GRAND CENTRAL	00100742	00004273	00000185	00000704	00001241	00022808	00068216
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Visualization of Tables

- Items and attributes
- For now, attributes are not known to be positions
- Keys and values
  - **key** is an independent attribute that is unique and identifies item
  - **value** tells some aspect of an item
- Keys: categorical/ordinal
- Values: +quantitative
- Levels: unique *values* of categorical or ordered attributes

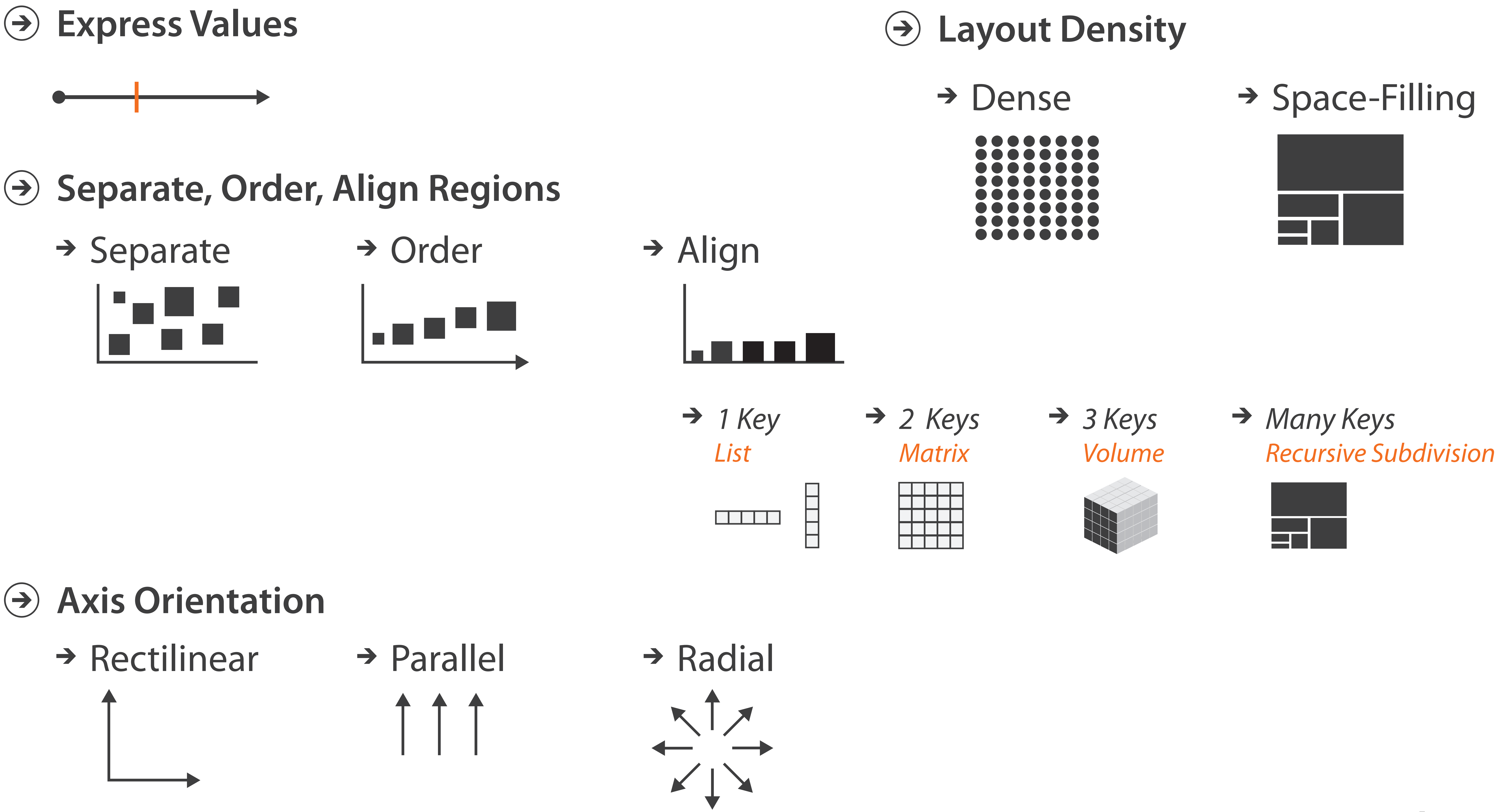


➔ *Multidimensional Table*



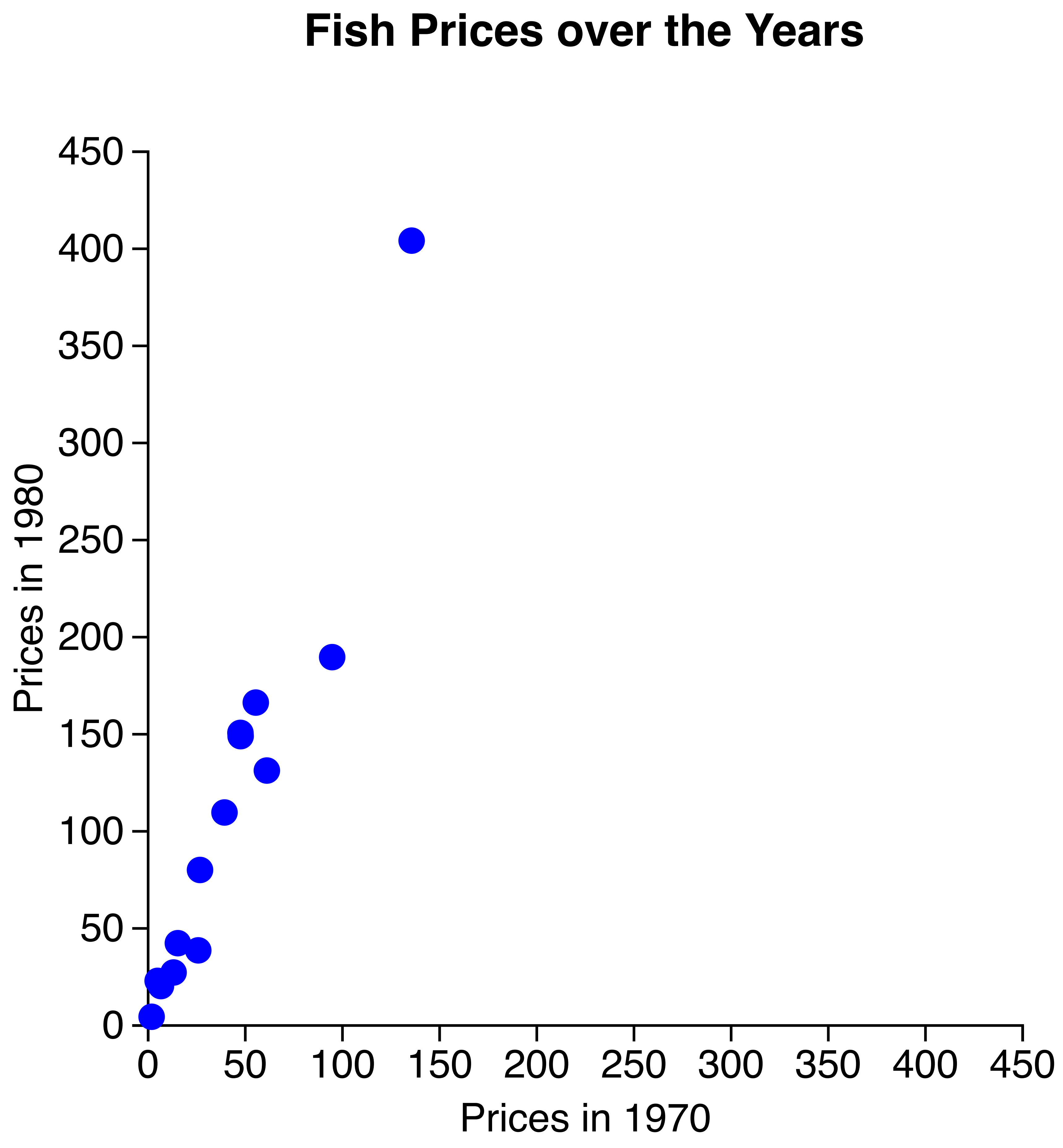
[Munzner (ill. Maguire), 2014]

Arrange Tables



[Munzner (ill. Maguire), 2014]

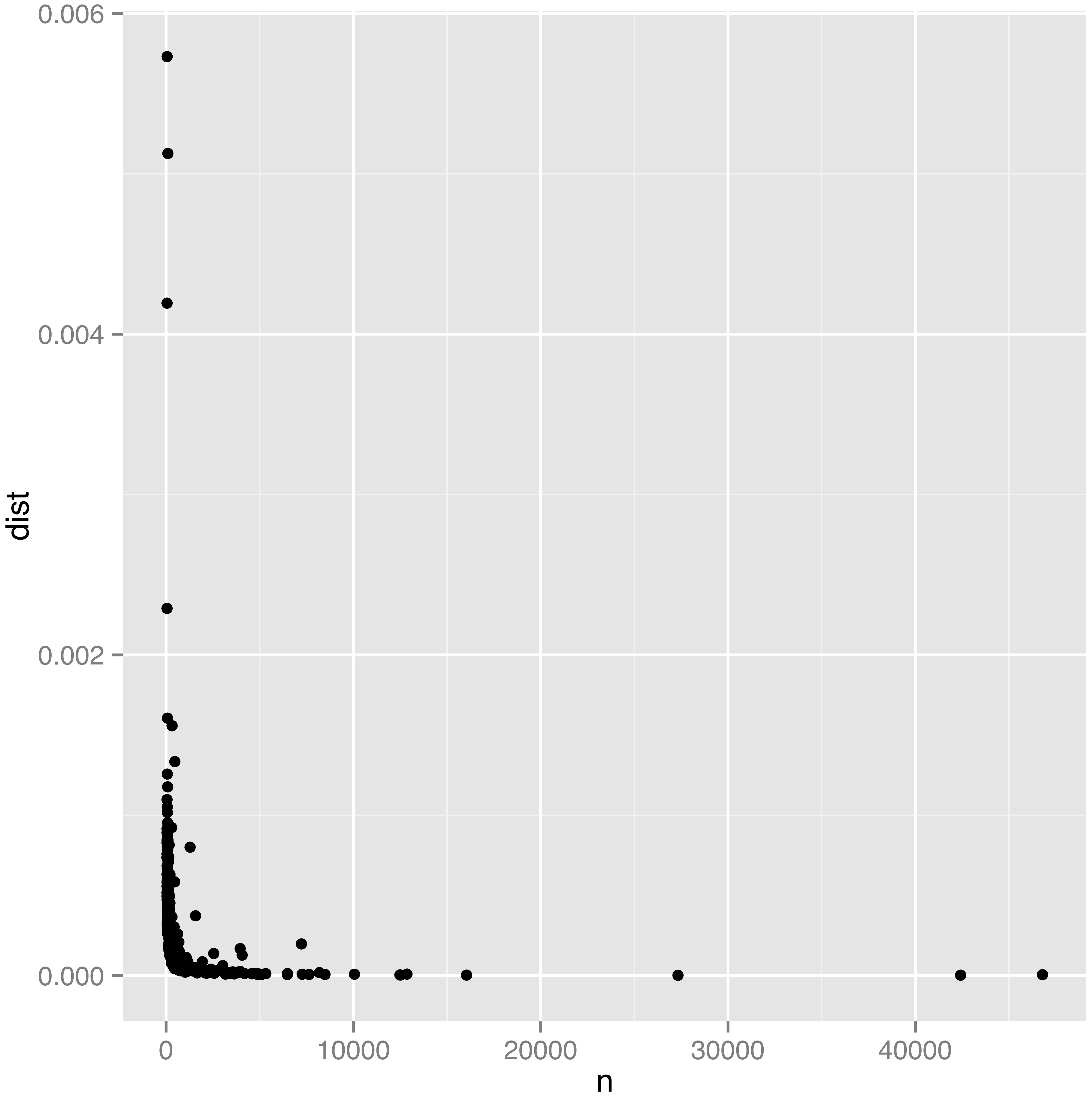
Express Values: Scatterplots



- Data: two quantitative values
- Task: find trends, clusters, outliers
- How: marks at spatial position in horizontal and vertical directions

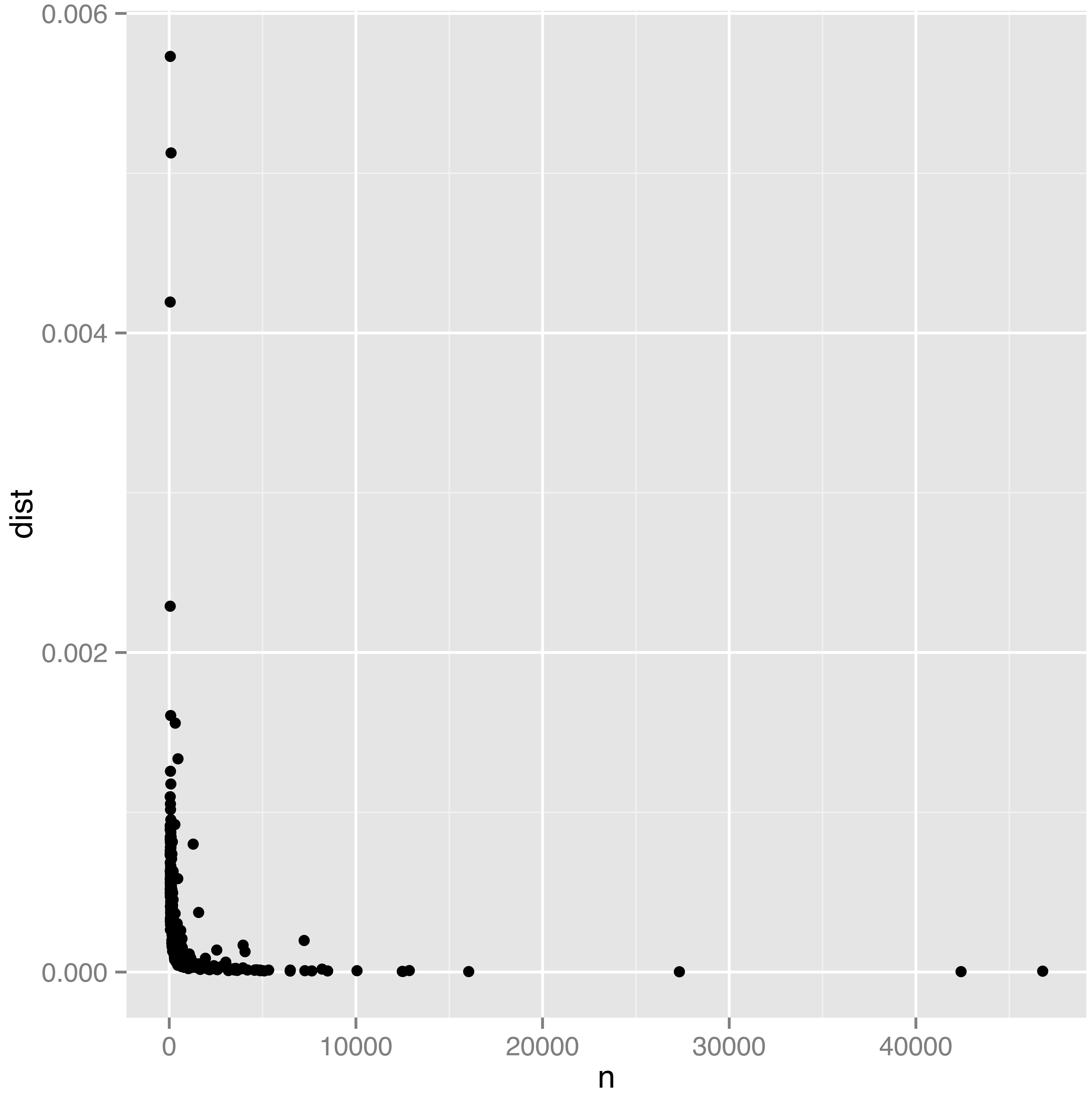
- Correlation: dependence between two attributes
  - Positive and negative correlation
  - Indicated by lines
- Coordinate system (axes) and labels are important!

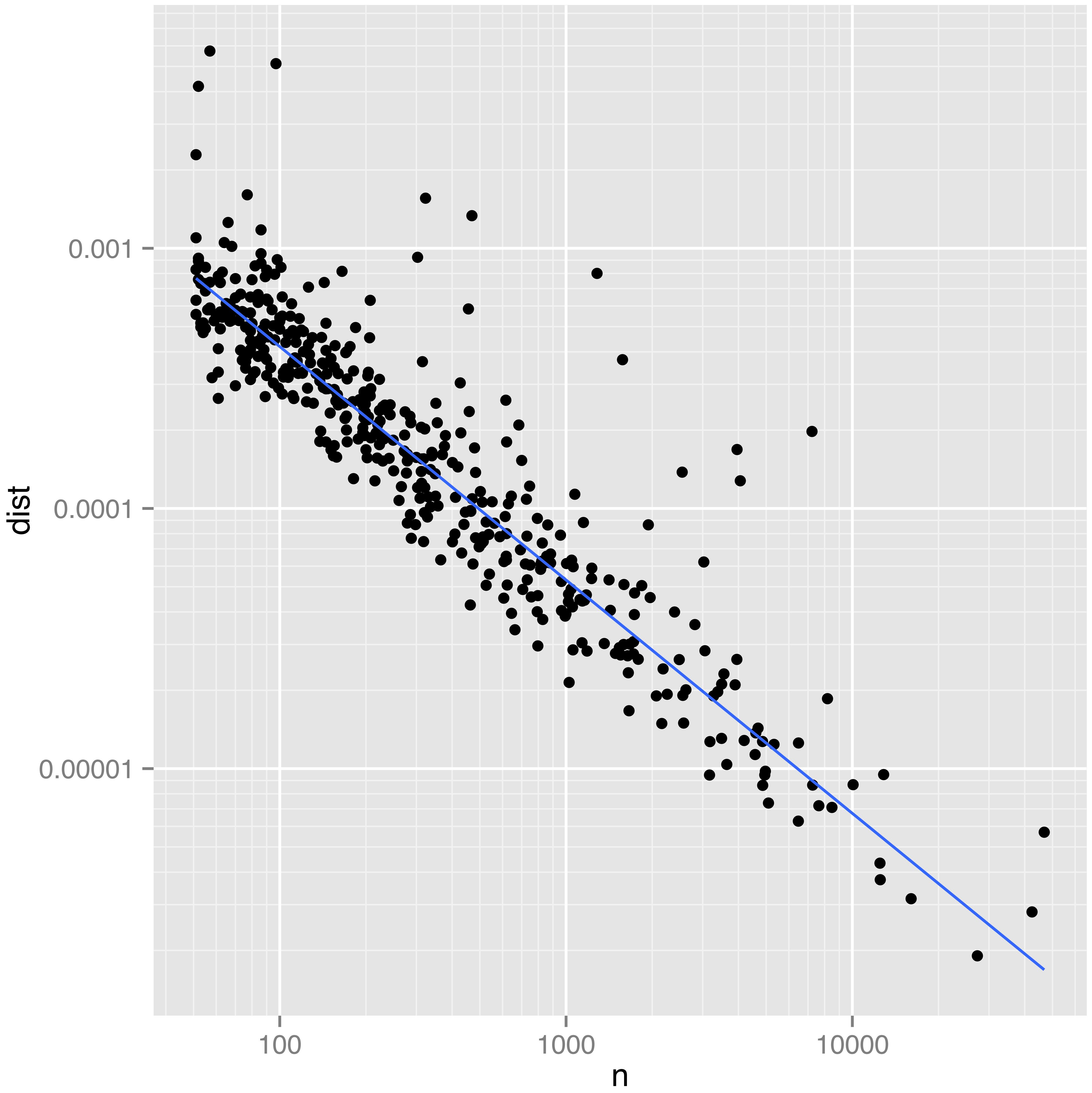
Coordinate Systems



[Wickham, 2014]

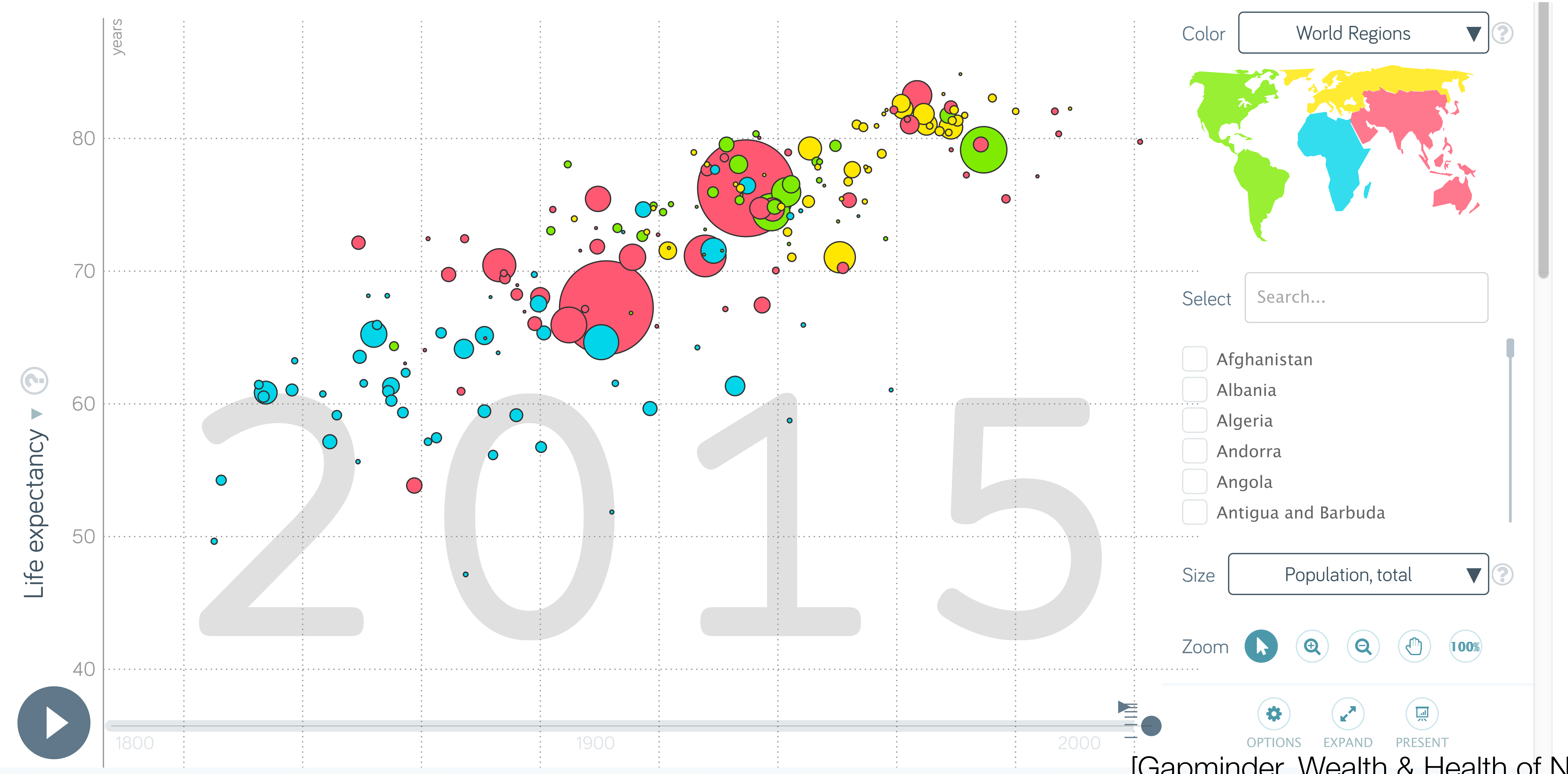
Coordinate Systems





[Wickham, 2014]

Bubble Plot



[Gapminder, Wealth & Health of Nations]

Scatterplot

- Data: two quantitative values
- Task: find trends, clusters, outliers
- How: marks at spatial position in horizontal and vertical directions
- **Scalability**: hundreds of items

- "Ranking Visualizations of Correlation Using Weber's Law", 2014:
  - Correlation perception can be modeled via Weber's Law
  - Scatterplots are one of the best visualizations for both positive and negative correlation
  - Further analysis: M. Kay and J. Heer, "Beyond Weber's Law", 2015

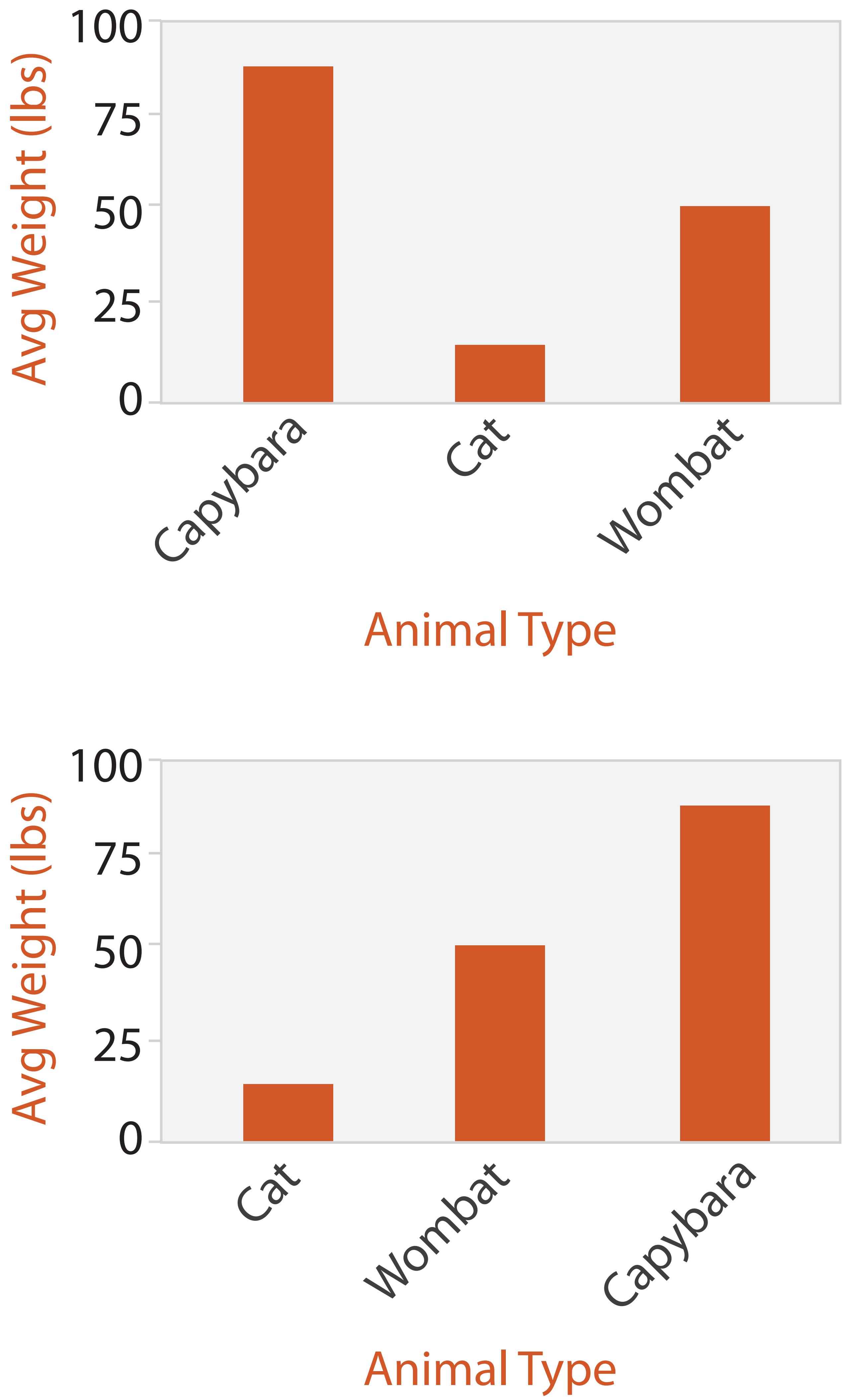
Separate, Order, and Align: Categorical Regions

- Categorical: =, !=
- Spatial position can be used for categorical attributes
- Use **regions**, distinct contiguous bounded areas, to encode categorical attributes
- Three operations on the regions:
  - Separate (use categorical attribute)
  - Align
  - Order

  (use some other ordered attribute)
- Alignment and order can use same or different attribute

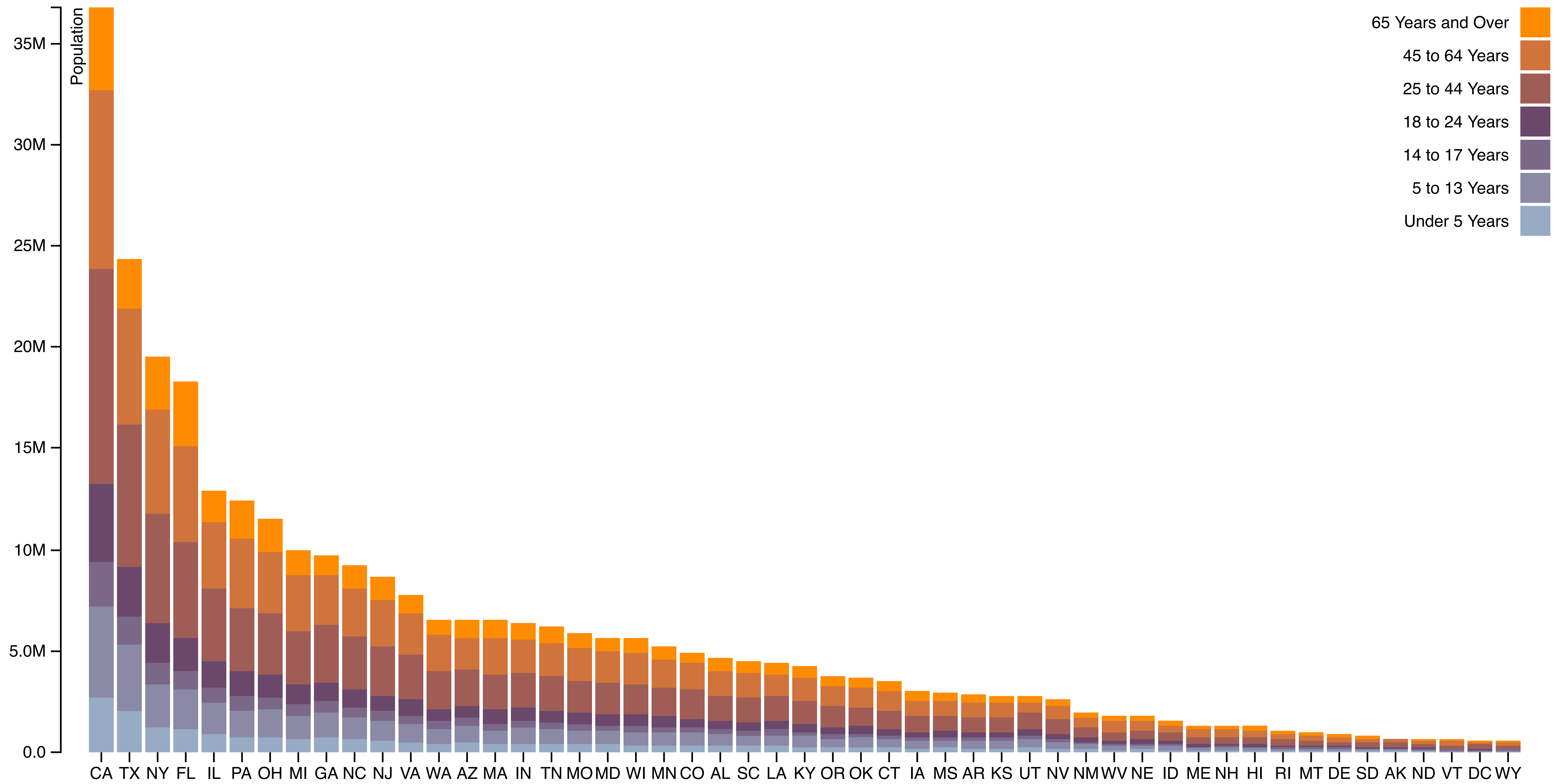
List Alignment: Bar Charts

- Data: one quantitative attribute, one categorical attribute
- Task: lookup & compare values
- How: line marks, vertical position (quantitative), horizontal position (categorical)
- What about **length**?
- Ordering criteria: alphabetical or using quantitative attribute
- Scalability: distinguishability
  - bars at least one pixel wide
  - hundreds



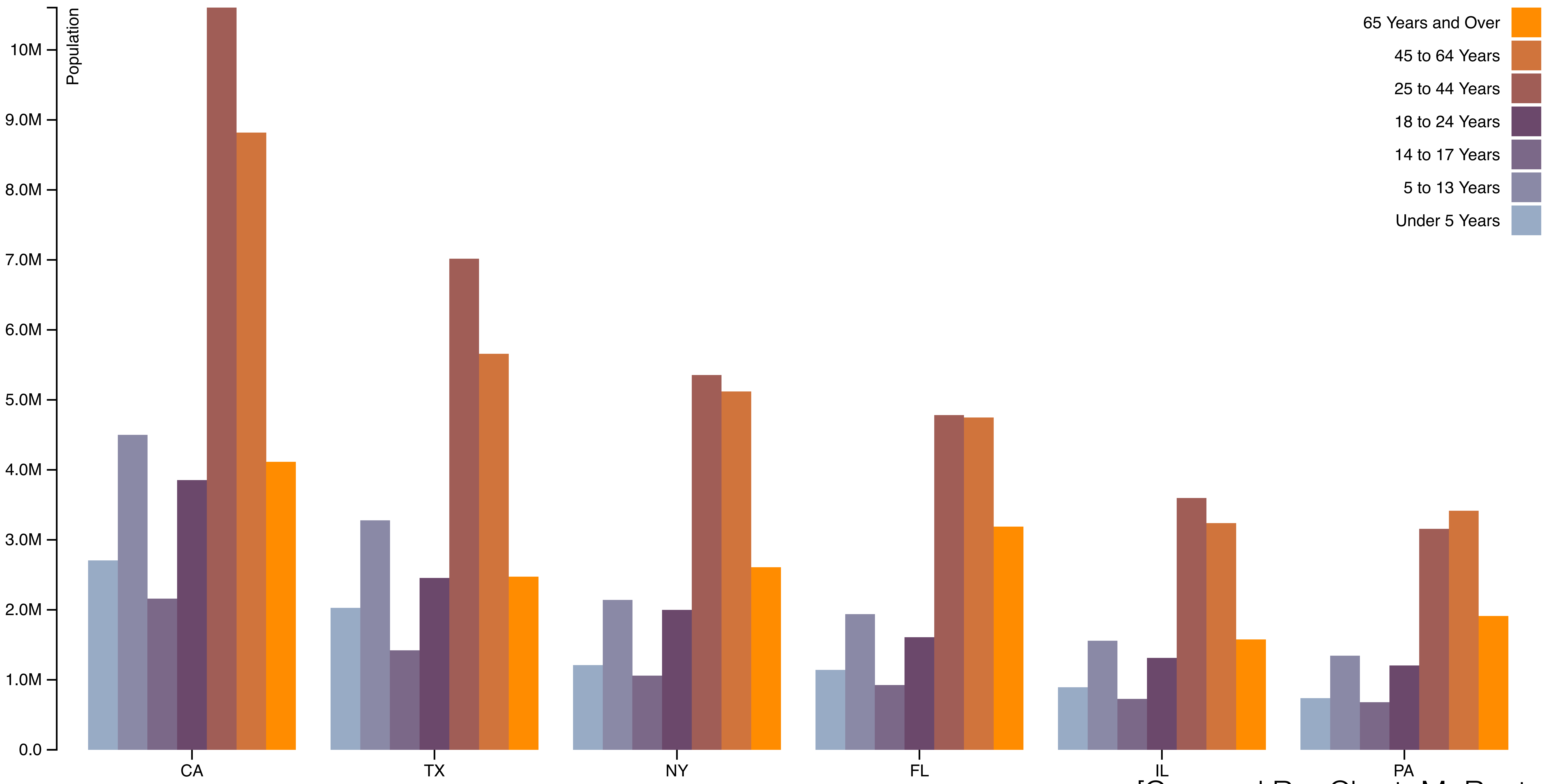
[Munzner (ill. Maguire), 2014]

Stacked Bar Charts



[Stacked Bar Chart, M. Bostock, 2017]

Grouped Bar Chart



[Grouped Bar Chart, M. Bostock, 2017]

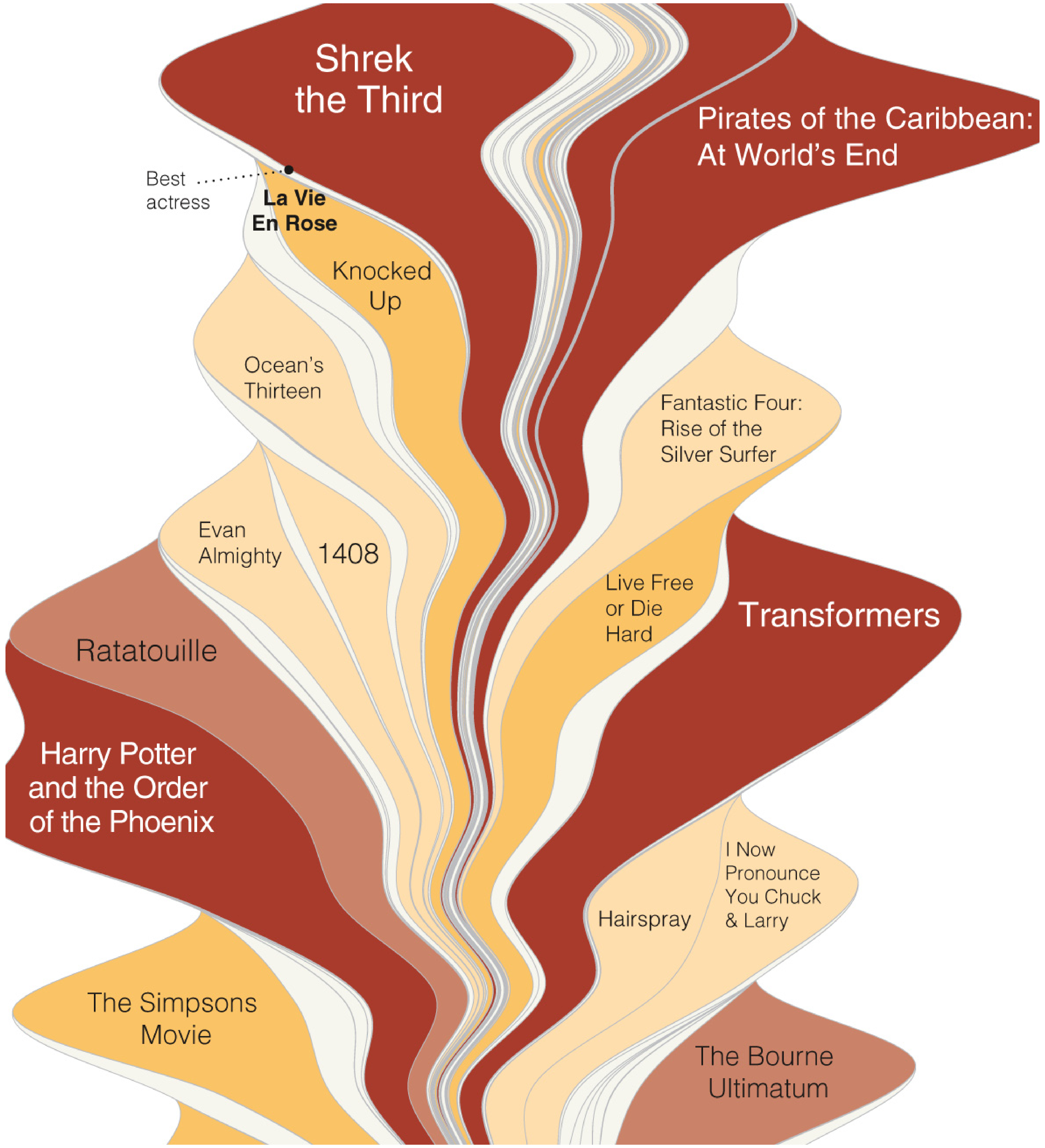
Stacked Bar Charts

- Data: multidimensional table: one quantitative, **two** categorical
- Task: lookup values, part-to-whole relationship, trends
- How: line marks: position (both horizontal & vertical), subcomponent line marks: length, color
- Scalability: main axis (hundreds like bar chart), bar classes (<12)

- Orientation: vertical or horizontal (swap how horizontal and vertical position are used.

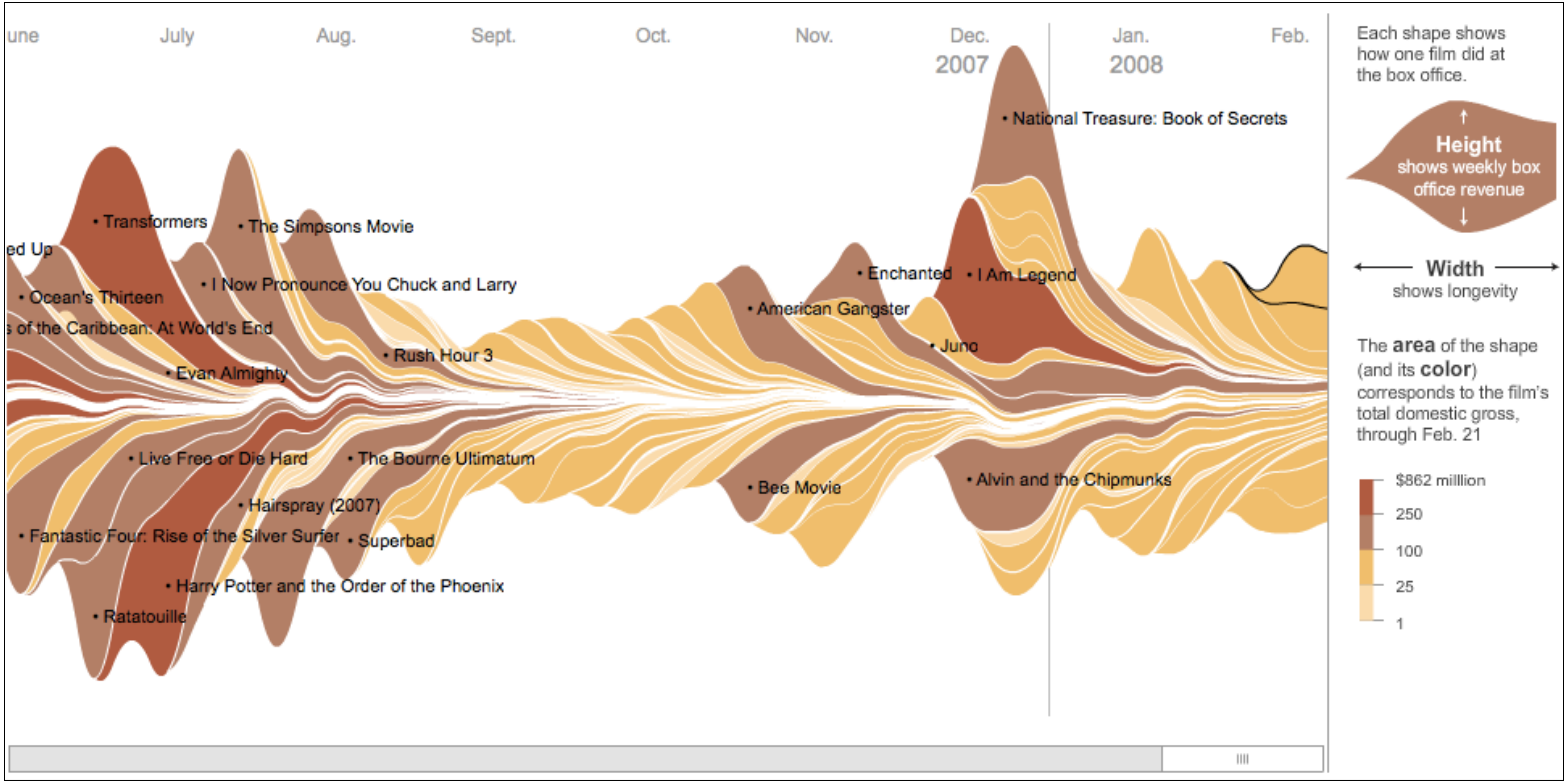
Streamgraphs

- Include a time attribute
- Data: multidimensional table, one quantitative attribute (count), one ordered key attribute (time), one categorical key attribute
- + derived attribute: layer ordering (quantitative)
- Task: analyze trends in time, find (maxmial) outliers
- How: derived position+geometry, length, color



[Byron and Wattenberg, 2012]

Streamgraphs



[Ebb and Flow of Movies, M. Bloch et al., New York Times, 2008]