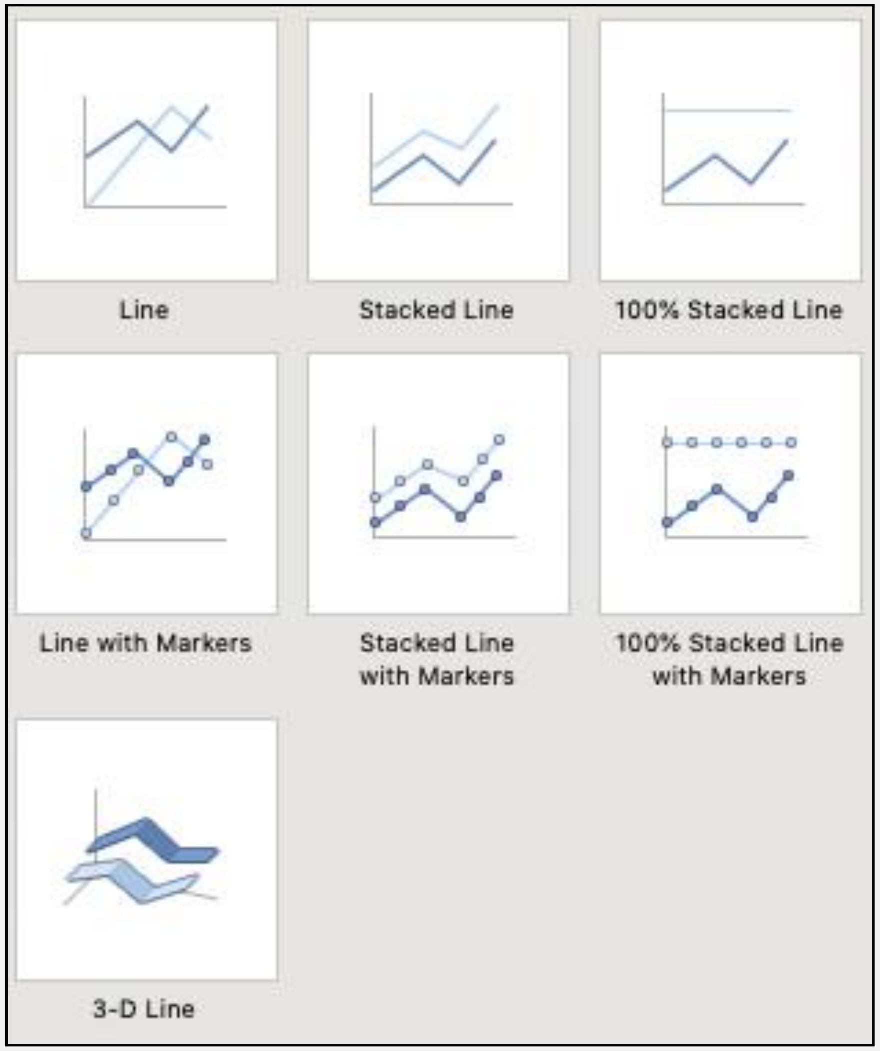
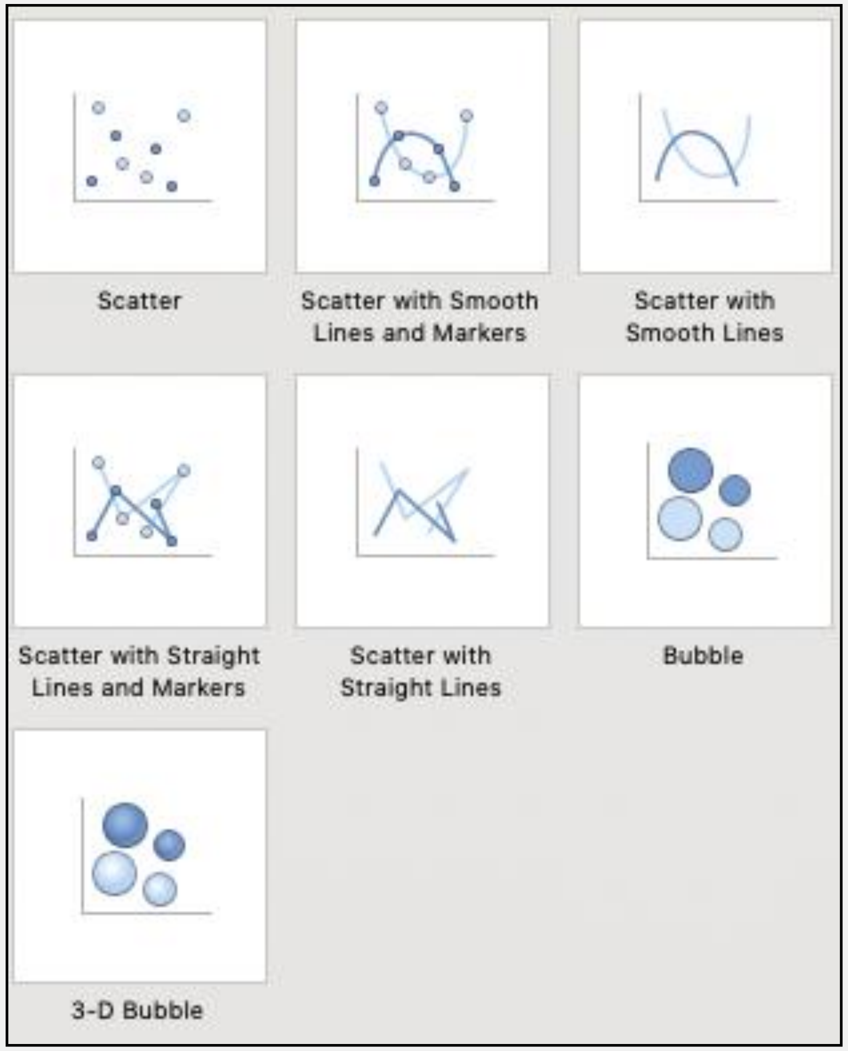
Visualization Design

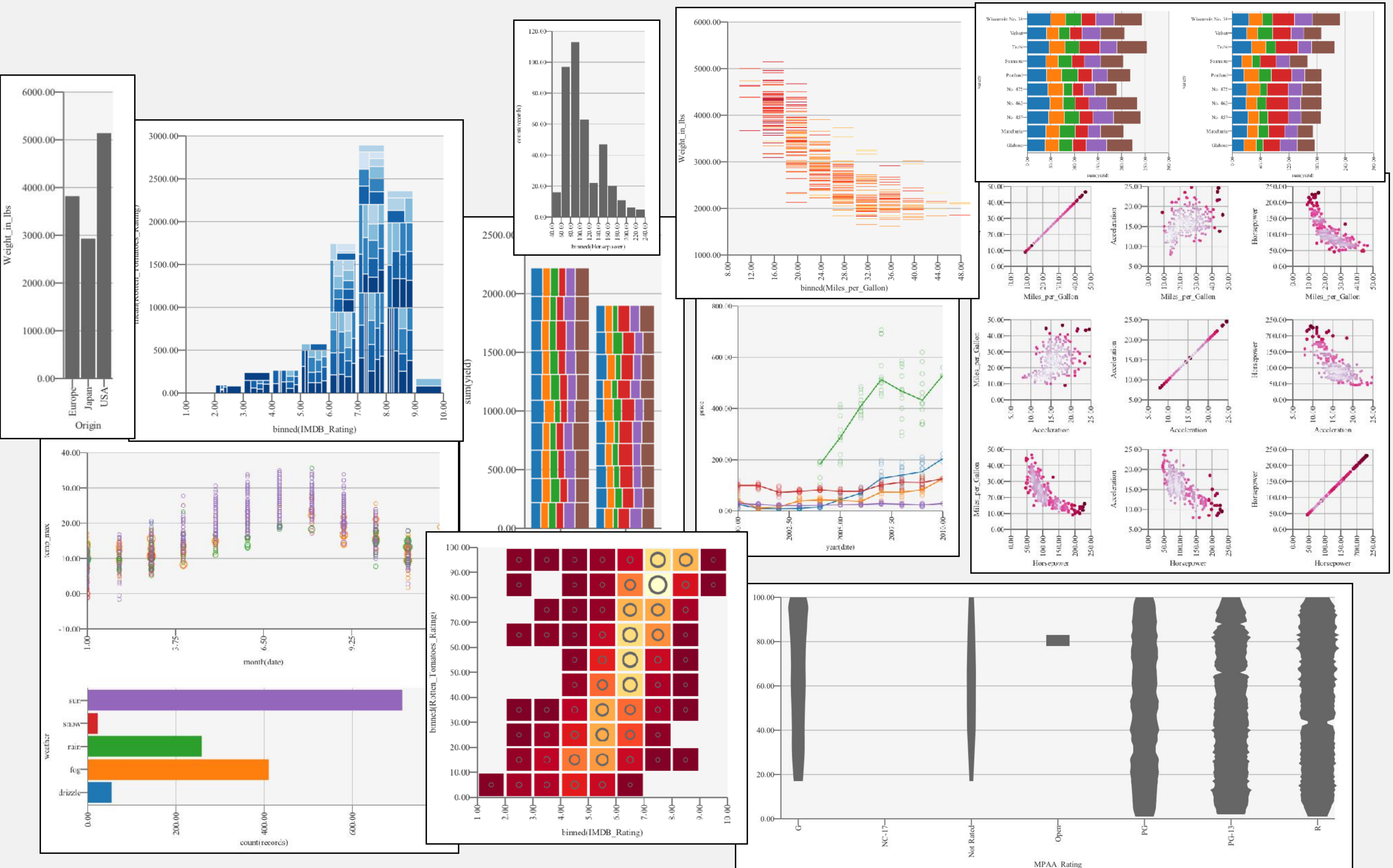
Charts as Design





- Visualization design: more than just a typology of visualizations

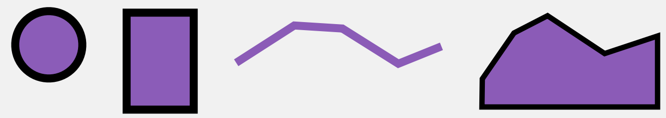
Space of Visualizations

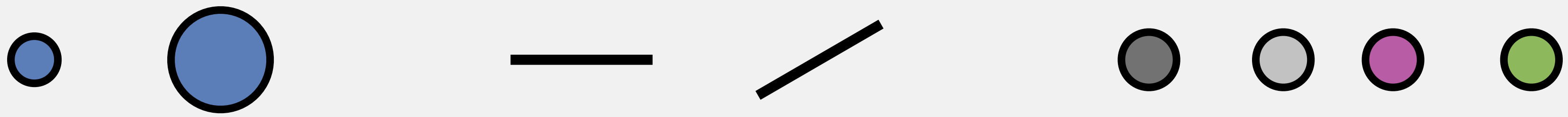


Designing with Data and Graphics

- We will take a more granular approach to authoring visualizations
- Each **data item** is encoded by a graphical mark
- How we draw the mark is dependent on the data item's **attributes**, or fields.
- A structured approach to graphics creation helps us understand the design space

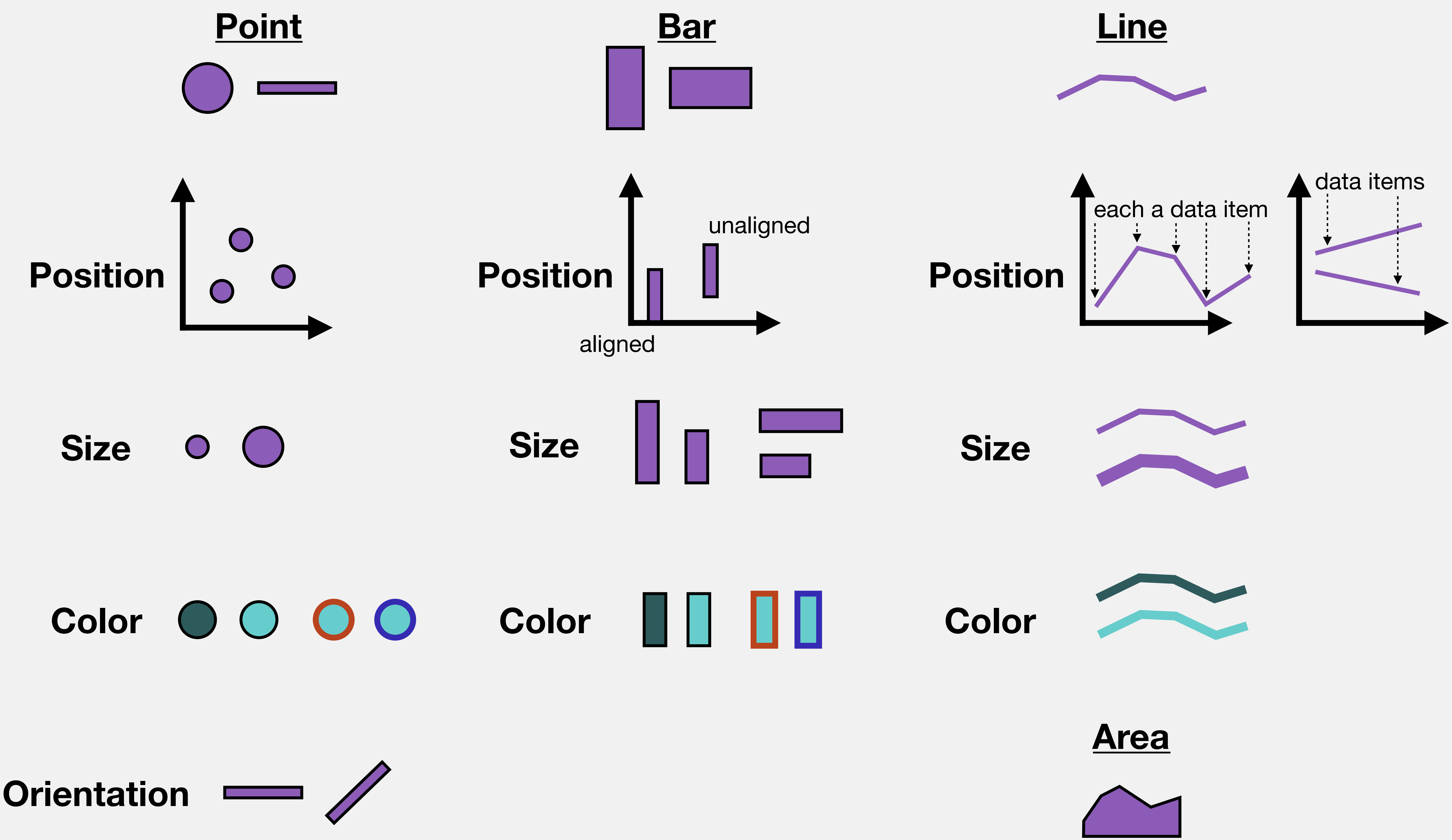
Declarative Approach to Design

- Choose a **mark** 
- For each attribute:
  - Determine the attribute's **type**
  - Choose a **visual channel**



  - Choose a **mapping**: from data domain to visual range
  - … draw it!
  - (there are some variations to this approach)

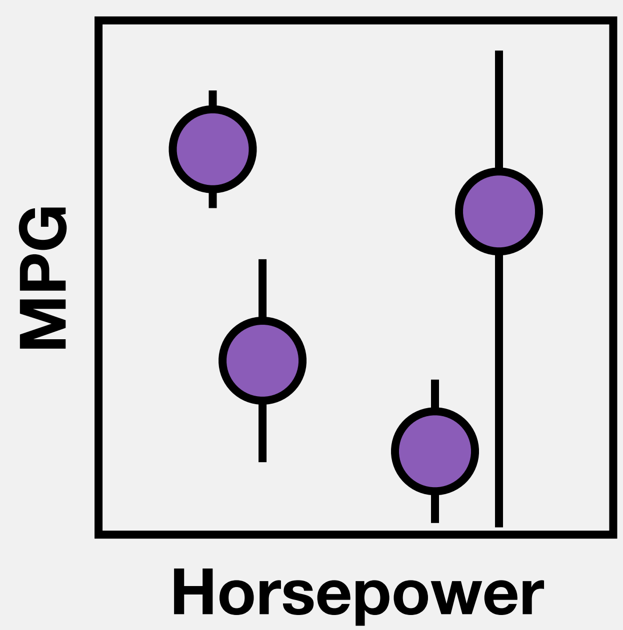
Graphical Marks, Visual Channels



Composite Marks

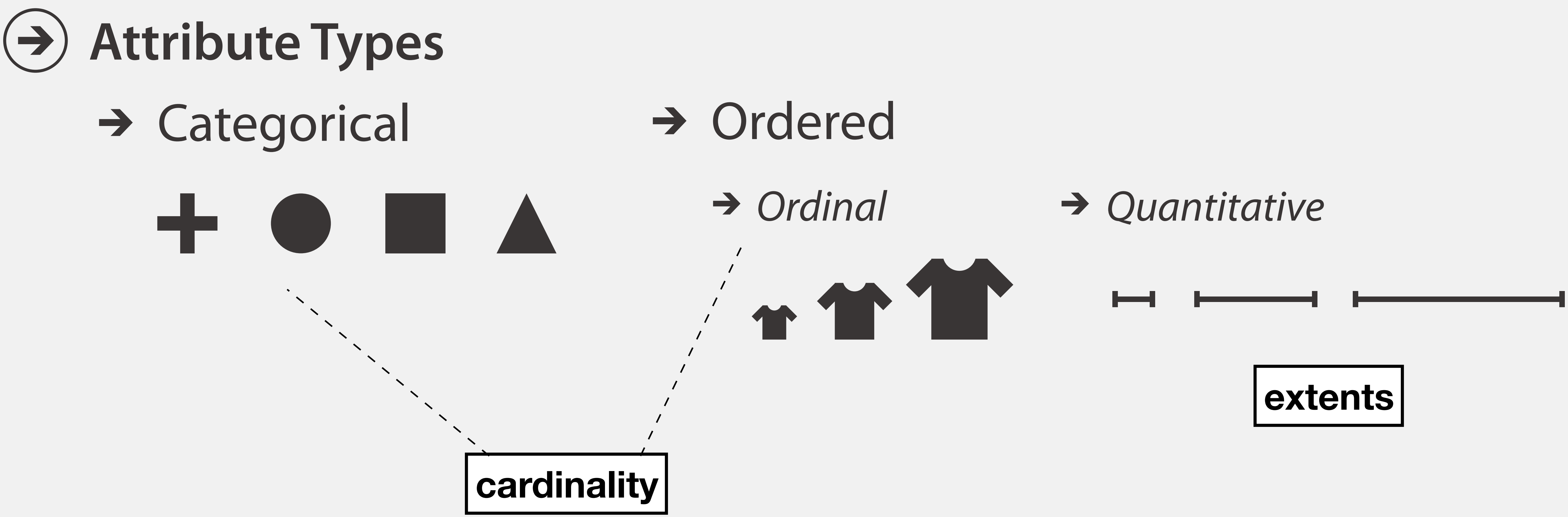
Car	Horsepower	min-MPG	MPG	max-MPG
Car 1	60	23.2	25.6	27.6
Car 2	86	25.4	30.1	31.1
Car 3	55	22.1	32.3	35.4
Car 4	50	32.2	33.8	34.5

The Mark:



Data: Attributes

- Starting from data, how do we go to graphics?
- First, need to determine a data item's **attribute**.



Quantitative ⊂ Ordinal ⊂ Categorical

Scales

- Next: we choose a mapping from data domain, to visual range. A **scale**.

D **data domain**

R **visual range**

$f : D \to R$ **scale**

- We distinguish scales via **data attributes**. Data domain *or* visual range can be categorical, ordinal, or quantitative.

Quantitative : Quantitative Linear Scale

- Domain and range are both quantitative. We first consider a **linear scale**.
- Mathematical preliminaries:
  - Assume data domain has a minimum and maximum value.
    $$d_{min} \longrightarrow d_{max}$$
  - Assume visual range has minimum and maximum value.
    $$r_{min} \longrightarrow r_{max}$$
- We seek a linear function such that:

$$f(d_{min}) = r_{min} \qquad\qquad f(d_{max}) = r_{max}$$

Quantitative : Quantitative Linear Scale

- Two step process:
  - Normalize data in domain. $d \in D \quad \alpha(d) = \dfrac{d - d_{min}}{d_{max} - d_{min}}$
  - Linearly interpolate in the range.

$$f(\alpha(d)) = (1 - \alpha(d)) \cdot r_{min} + \alpha(d) \cdot r_{max}$$

Linear Scale Example

$$d_{min} = 0$$

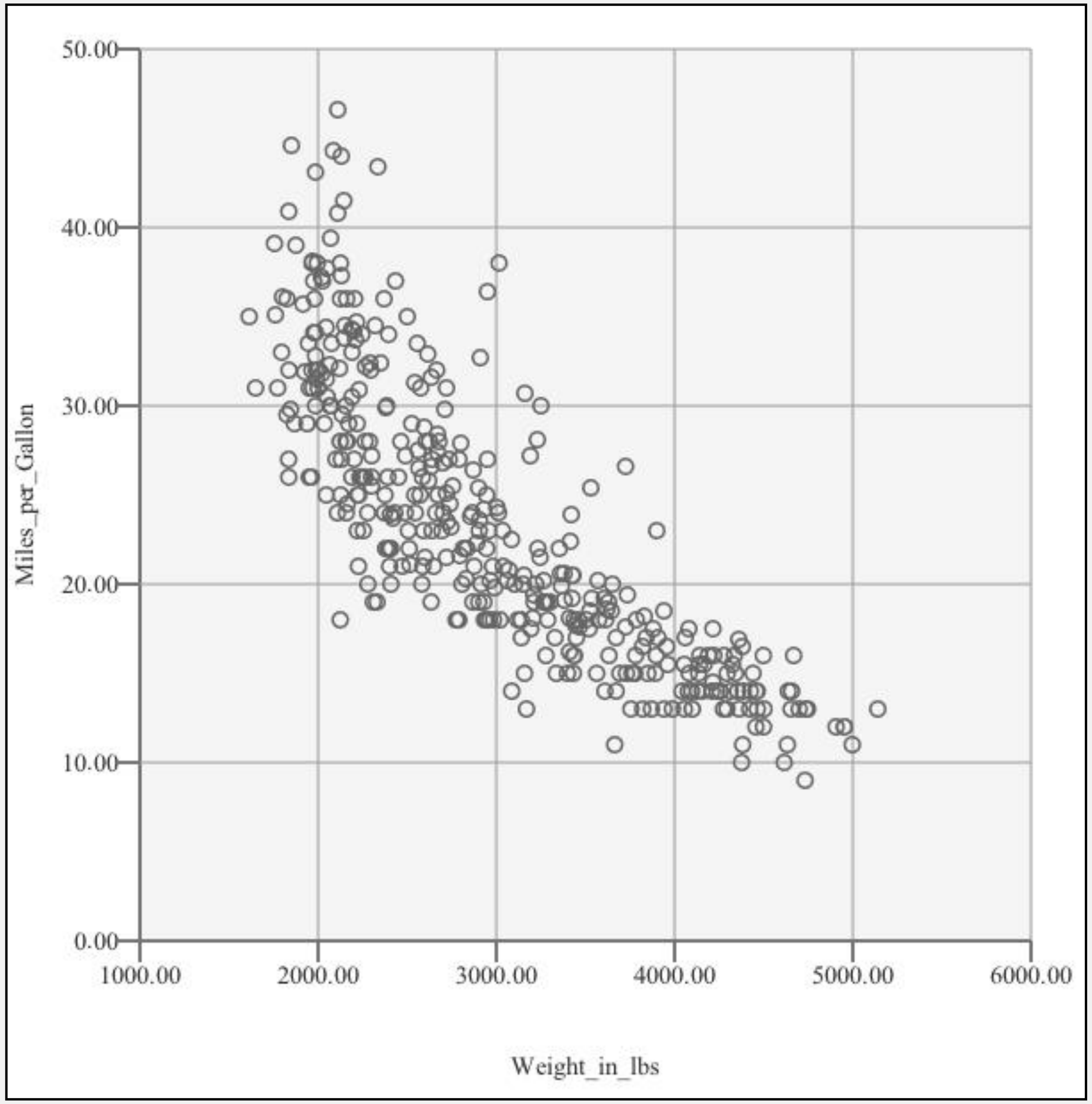
$$d_{max} = 50$$

$$r_{min} = 500\text{px}$$

$$r_{max} = 70\text{px}$$

$$f_y$$

scale for y channel



$$d_{min} = 1000 \quad r_{min} = 70\text{px}$$

$$d_{max} = 6000 \quad r_{max} = 500\text{px}$$

$$f_x$$

scale for x channel

Other Quantitative Scales

- Follow a very similar form:
  - Need to satisfy $f(d_{min}) = r_{min} \quad f(d_{max}) = r_{max}$
  - Otherwise, take on characteristic of prescribed function, e.g. <u>quadratic</u>, <u>square root</u>, etc..
- Special scale: **log**

$$\alpha(d) = \frac{\log(d) - \log(d_{min})}{\log(d_{max}) - \log(d_{min})}$$

$$f(\alpha(d)) = (1 - \alpha(d)) \cdot r_{min} + \alpha(d) \cdot r_{max}$$

Quantitative : Ordinal Quantized Scale

- Our data domain is quantitative, our visual range is discrete - and in particular, ordered.

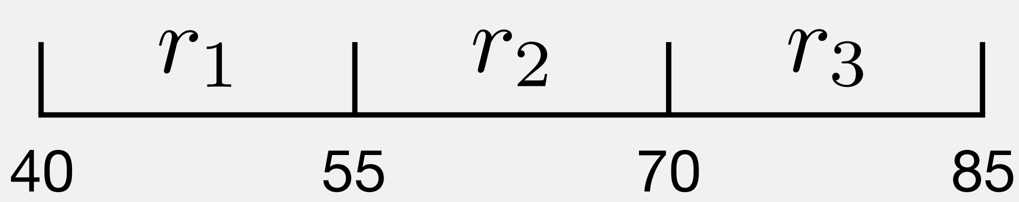
$$d_{min} \longrightarrow d_{max} \qquad\qquad R = [r_1, r_2, \cdots, r_n]$$

- Common assumption: range is uniformly divided by the domain.

$$f(\alpha(d)) = r_i \quad \text{s.t.} \quad i = 1 + \lfloor n \cdot \alpha_d \rfloor$$

Special Case: Binning Transformation

Car	Horsepower	MPG
Car 1	60	25.6
Car 2	85	30.1
Car 3	54	32.3
Car 4	66	33.8
Car 5	68	31.2
Car 6	82	30.4
Car 7	52	22.3
Car 8	45	25.4
Car 9	40	28.4



bin(**Horsepower**) : {
r_1 : [**Car 3, Car 7, Car 8, Car 9**]
r_2 : [**Car 1, Car 4, Car 5**]
r_3 : [**Car 2, Car 6**]
}

Aggregate items in each list

Car	count
r_1	4
r_2	3
r_3	2

Data Aggregation

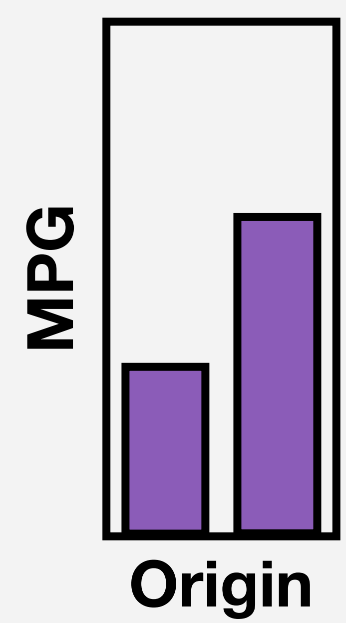
Car	Horsepower	MPG	Year	Weight	Origin
Car 1	60	25.6	2012	3000	USA
Car 2	86	30.1	2013	3024	Europe
Car 3	55	32.3	2014	3100	USA
Car 4	50	33.8	2015	3084	Europe

- **Aggregation**. The *group-by* operation.

group-by(**Origin**) : { USA : [**Car 1**, **Car 3**]
Europe : [**Car 2**, **Car 4**] }

Car	mean(MPG)	Origin
Car Group 1	28.95	USA
Car Group 2	31.95	Europe

Aggregate items in each list

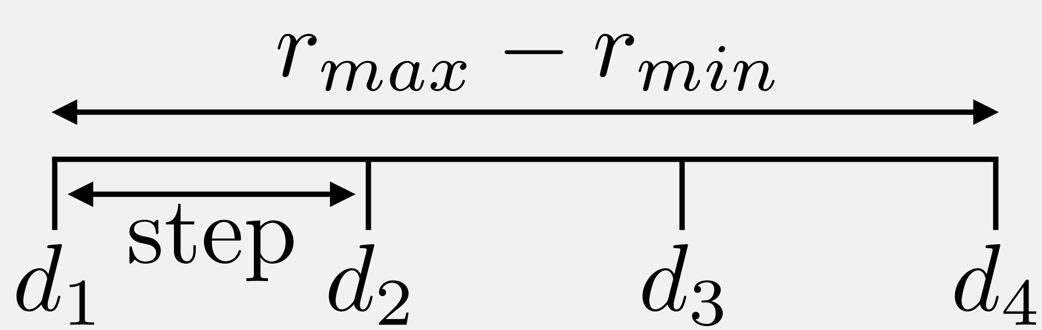


Can group *repeatedly* as well. We will revisit this many times.

Ordinal : Quantitative Point Scale

- Our data domain is ordinal, our visual range is quantitative.

$$D = [d_1, d_2, \cdots, d_m] \qquad\qquad r_{min} \longrightarrow r_{max}$$



$$\text{step} = \frac{r_{max} - r_{min}}{m - 1}$$

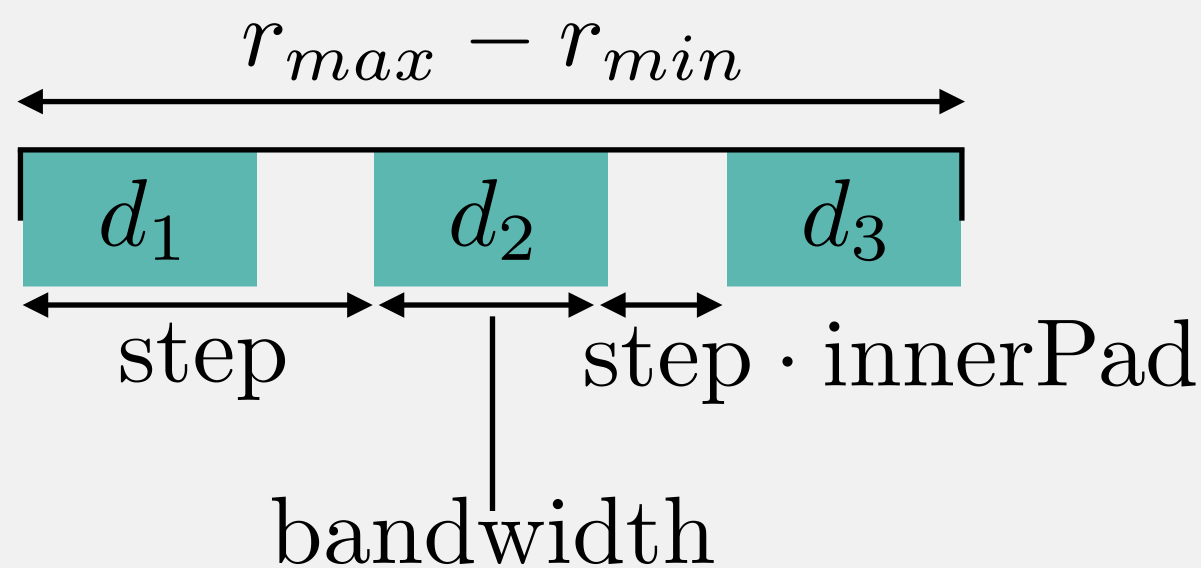
$$f(d_i) = r_{min} + (i - 1) \cdot \text{step}$$

- It is also common to introduce <u>padding</u> at the beginning and end.

Ordinal : Quantitative
Band Scale

- Specific to position visual channel.

$$D = [d_1, d_2, \cdots, d_m] \qquad r_{min} \longrightarrow r_{max}$$

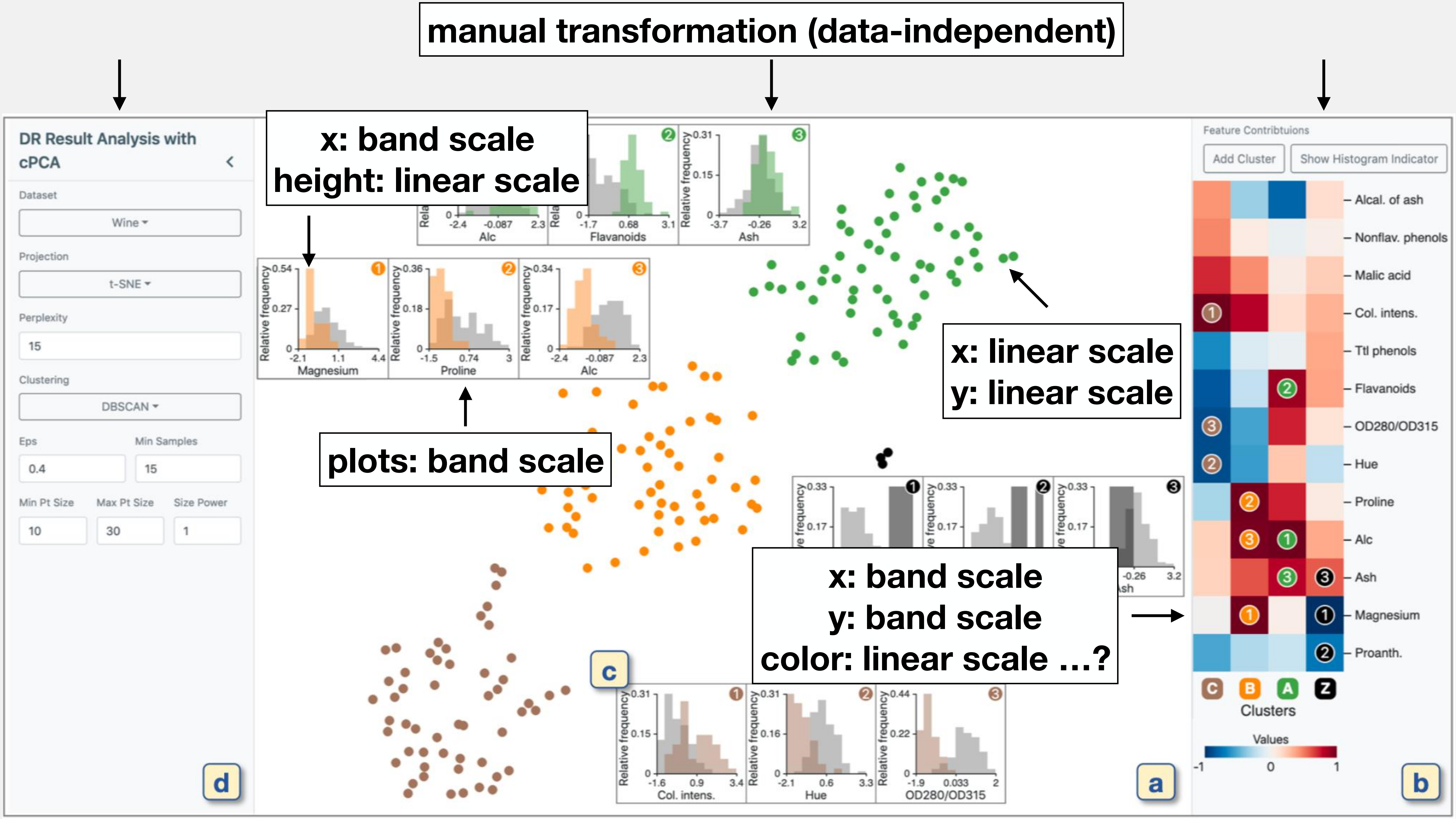


$$f(d_i) = r_{min} + (i - 1) \cdot \text{step}$$

- Used for: bar marks, group transformations

- Also common to introduce padding at the beginning and end.

Example



[Fujiwara et al. 2019]

Recap: A Recipe for Authoring Visualizations

- Ingredients:
  - Identify data items.
  - Select data attributes.
  - Determine a type for each attribute.
  - Determine a graphical mark for an item.
  - Select a visual channel for each attribute.
  - Select a scale for each channel.

A Grammar of Graphics

[Wickham 2010, Wilkinson 2012, Satyanarayan et al. 2014, Satyanarayan et al. 2016]

Vega-Lite: A Grammar of Interactive Graphics

https://observablehq.com/d/3e841f9a6578e3ea

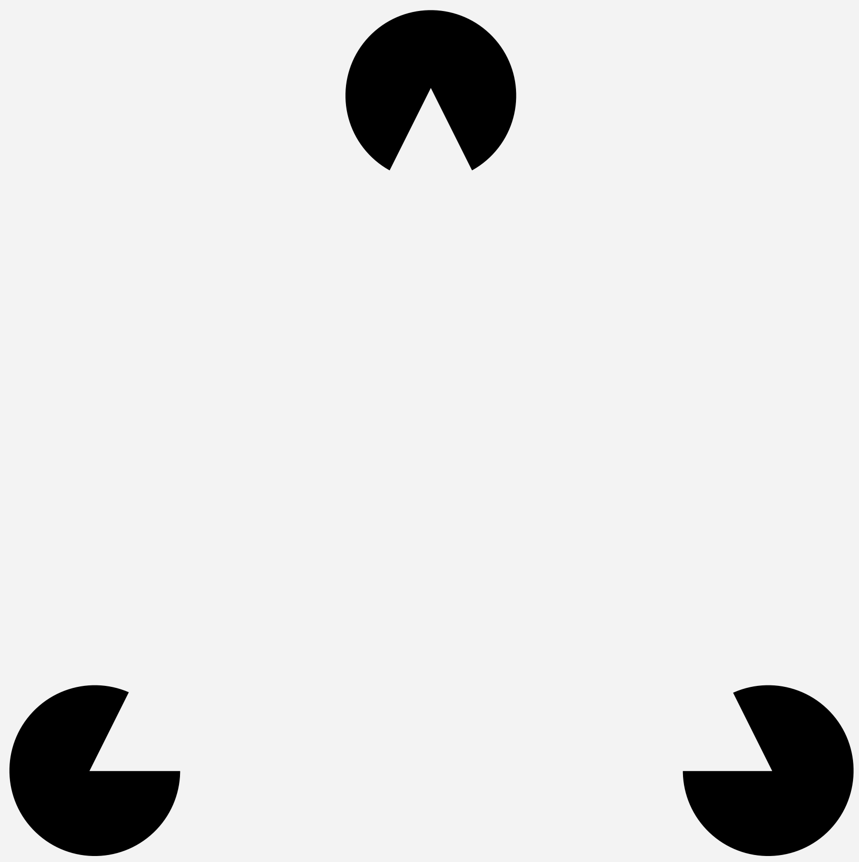
Vega-Lite Presets

- Still lots of design choices that Vega-Lite automates for us - color, spacing, axes, labels, etc..
- Some design choices are aesthetic.
- Others are driven by human perception: how the user *decodes a visualization* should be aligned with their **perceptual abilities**.
  - (In other words: don’t make the user exert a large amount of effort to understand your visualization!)

Gestalt Laws

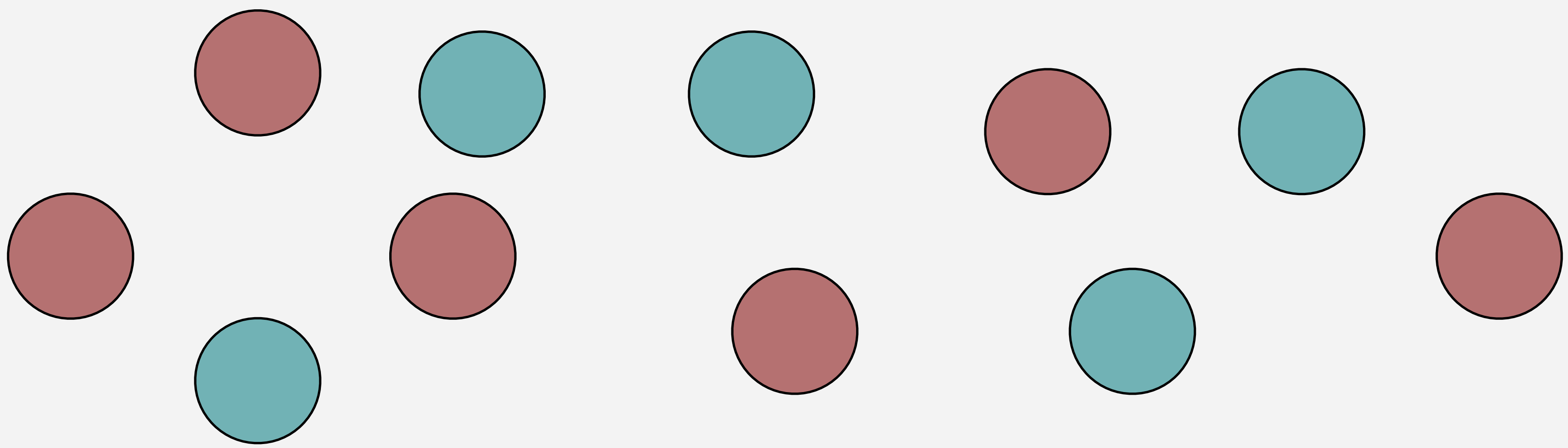
- How does one organize visual information?

- **Gestalt**: shape/form, how we assemble visual objects into a more unified whole



Similarity

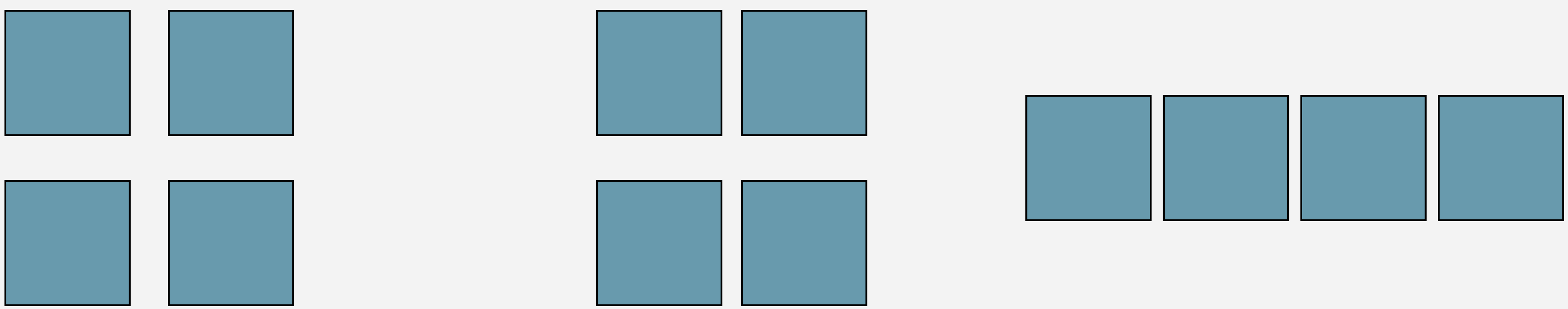
- Objects that *look alike* are perceived as being related.

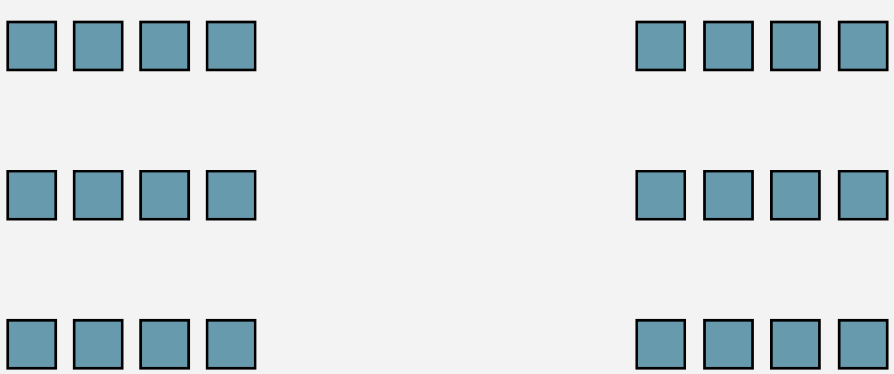


- **Implications**:
  - How should we visually encode nominal data?

Proximity

- Objects that are in close *spatial proximity* are perceived as belonging together.
- How do you group?

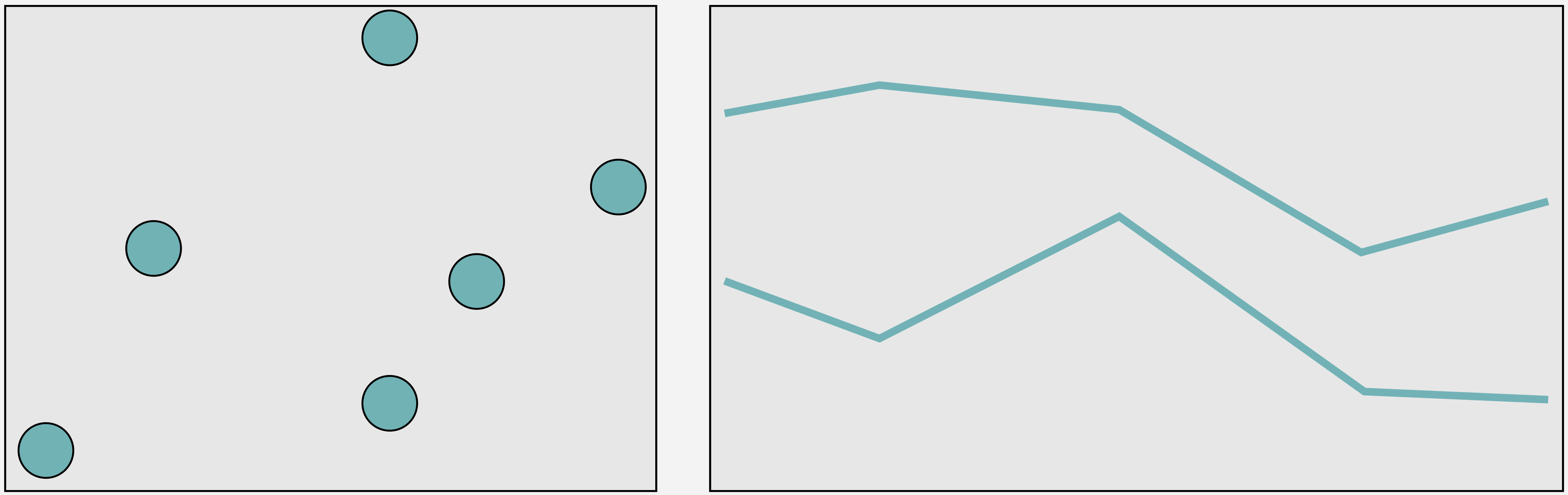




Spatial visual channel: *think in hierarchies* - other channels do not provide such opportunities!

Enclosure

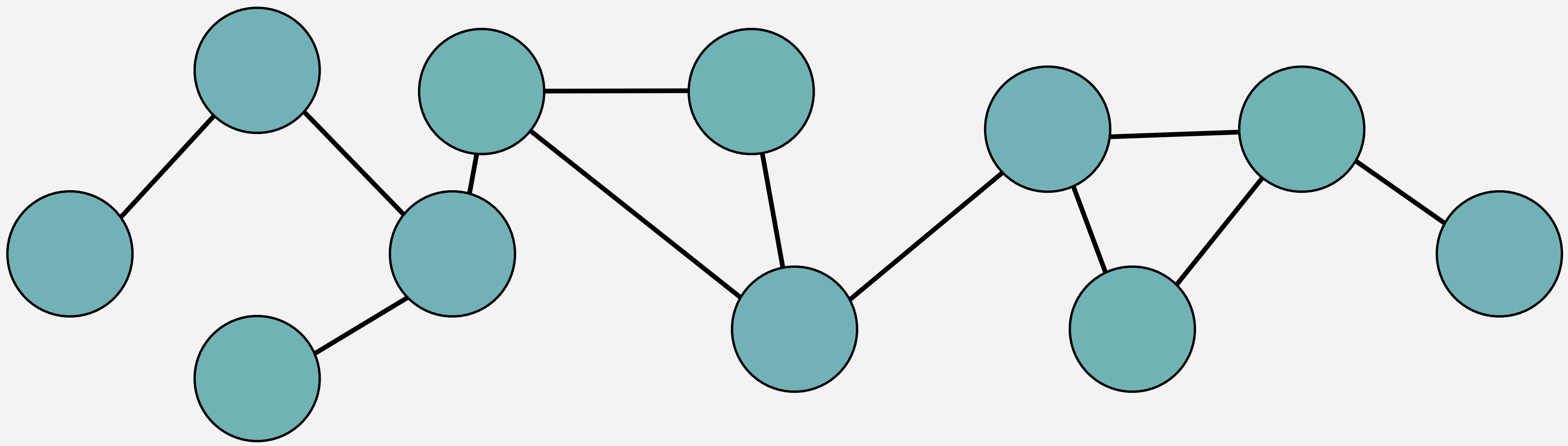
- Objects that are bound by a common region perceived as belonging together.



- **Implication**: ensure sufficient discrimination between your plots. Don’t let marks “float in space”.

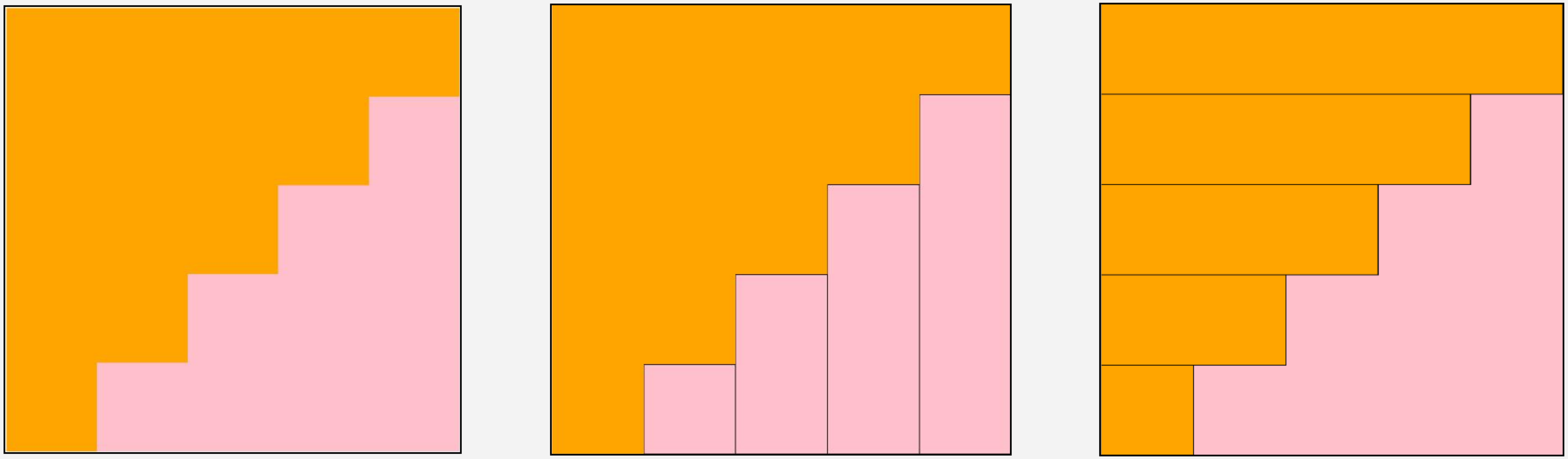
Connectedness

- Objects that are visually connected in some form are perceived as related.
- **Implications**: graphs and networks.



Figure/Ground

- Ensure marks are distinguished from background.



https://emeeks.github.io/gestaltdataviz/section4.html

Visualization for Perception

- We want our visualizations to yield visual queries that are *rapid* and *effective*.

ehklhfdiyaioryweklblkhockxlyhirhu**p**werlkhlkuyxoiasusifdhlk
sajdhflkihqdaklljerlajesljselusdslfjsalsuslcjlsdsjaf;ljdulafjlujou
fojrto**p**jhklghqlkshlkfhlkdshflymcvciwo**p**zlsifhrmckreieui

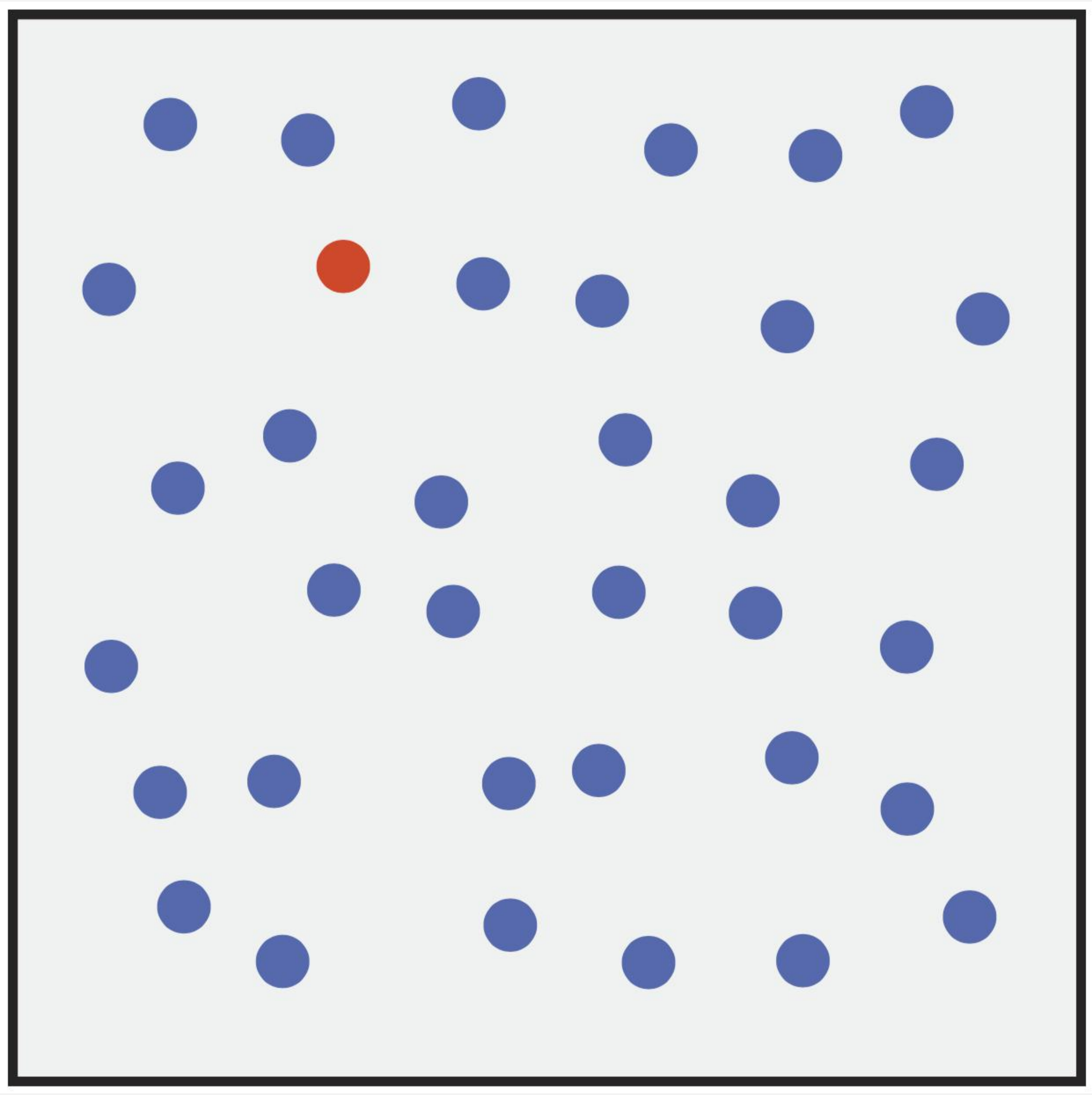
Find the p's **Find the q's**

- Find the p's: our brain operates *in parallel*.

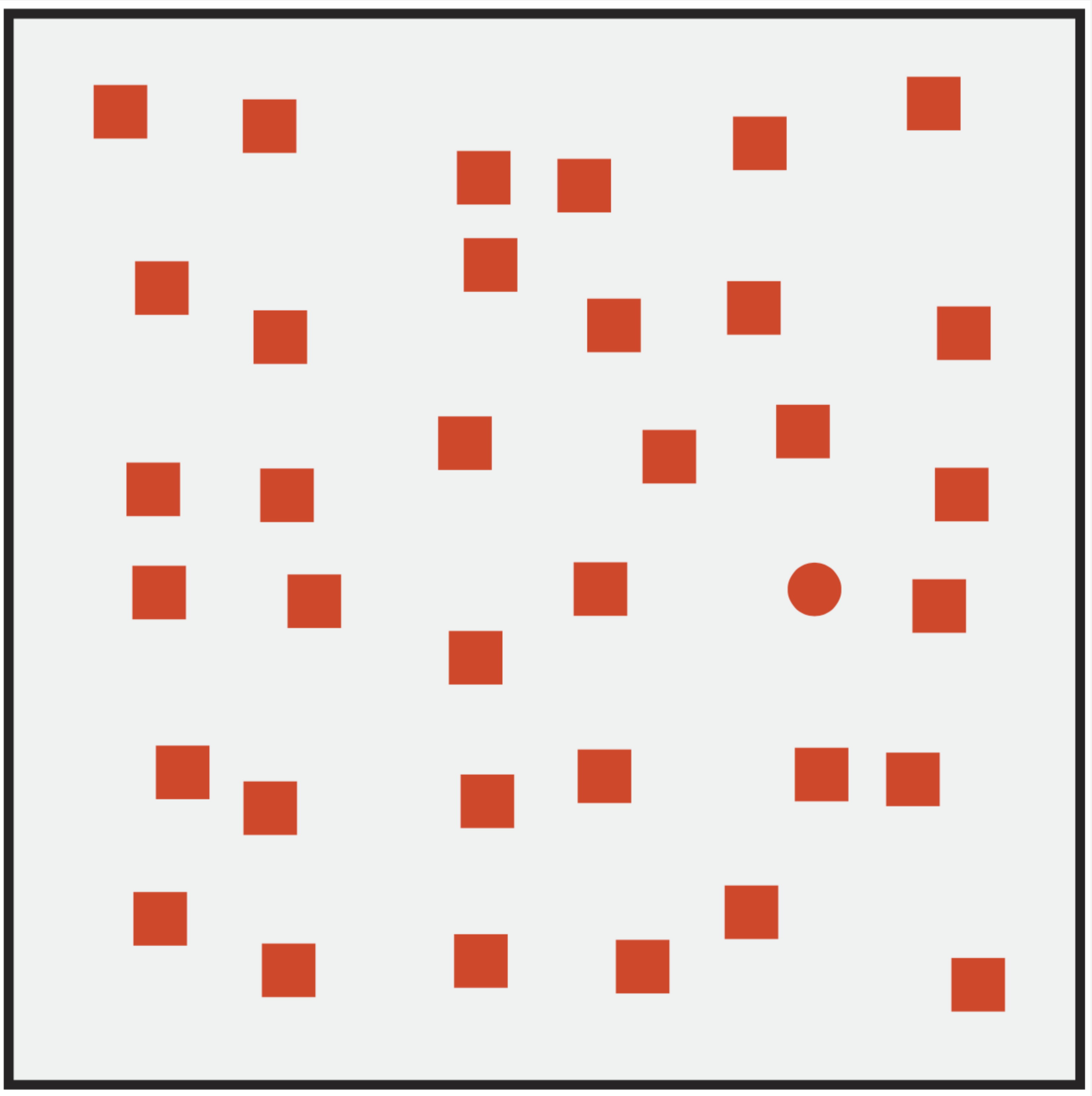
- Find the q's: our brain operates *in serial*.

Preattentive Processing

Detect the target object!



Detect the target object!



Preattentive Processing

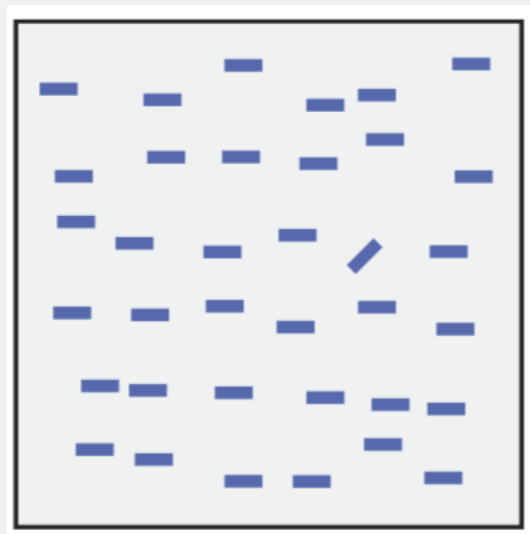
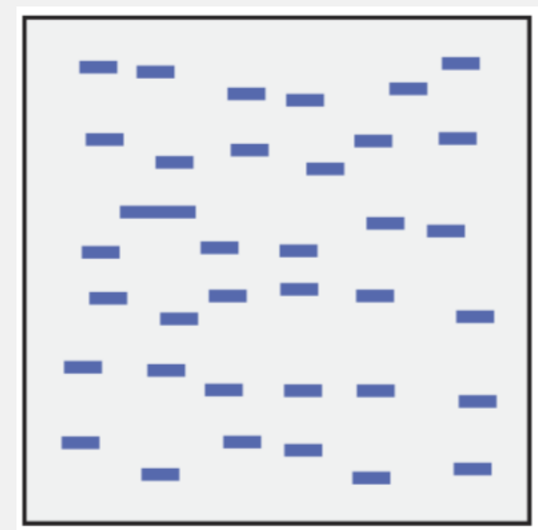
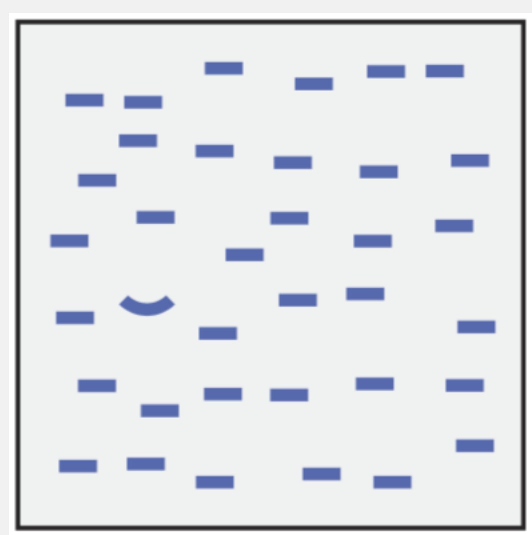
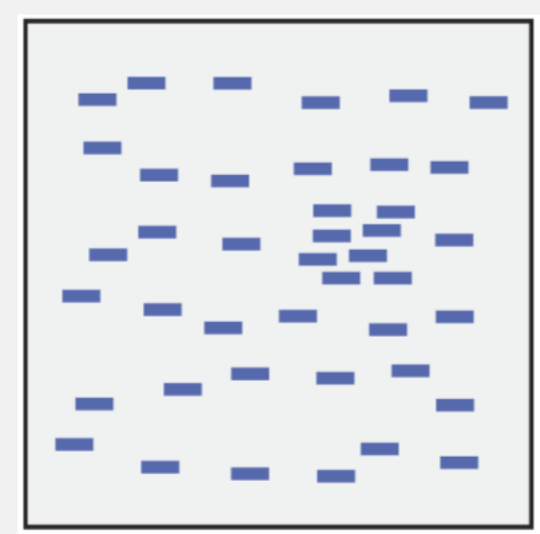
- Eye movements

POP OUT

- Typically of the order of 200-250 ms

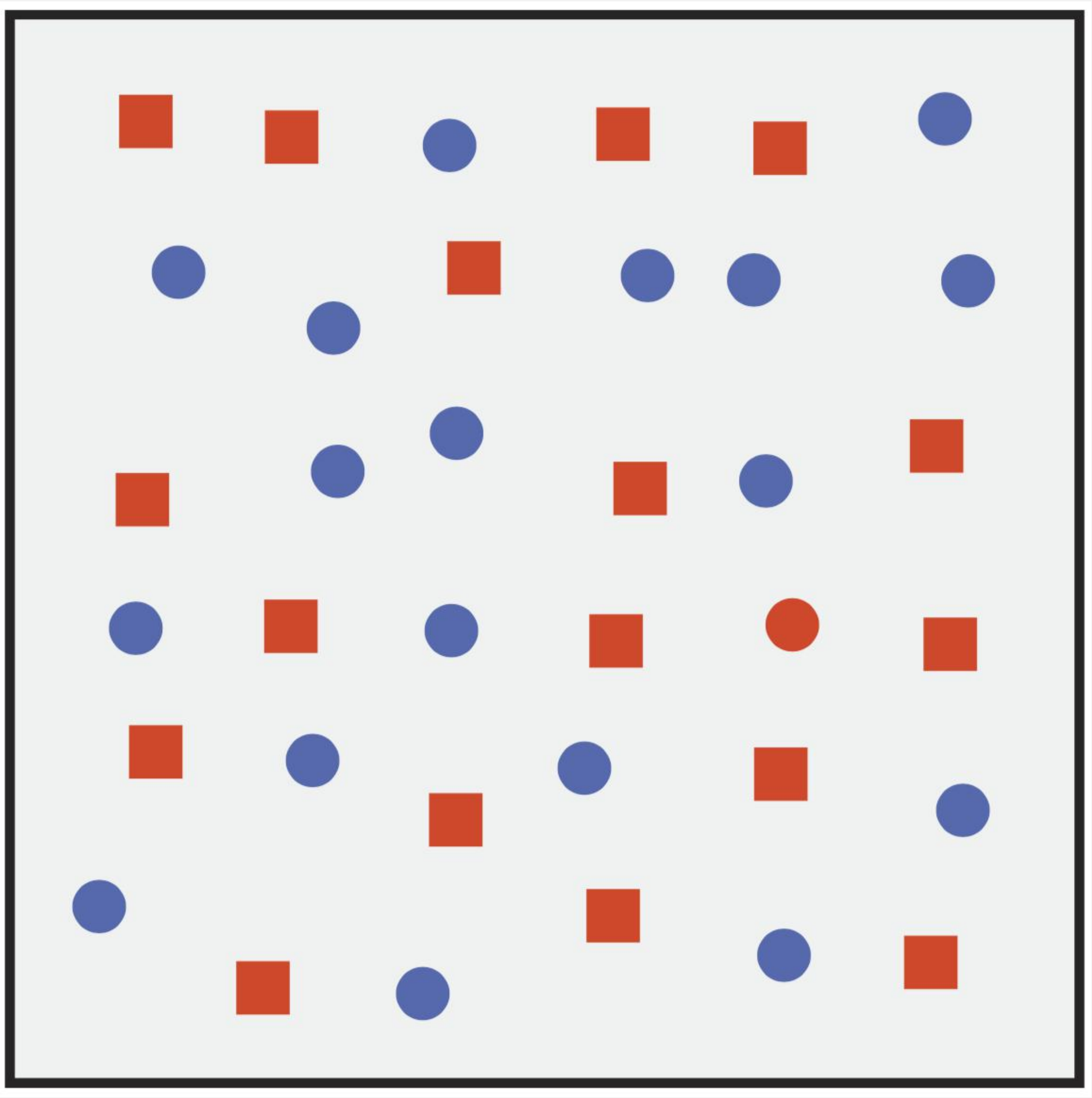
Preattentive Visual Channels

- hue
- shape
- length
- orientation
- curvature
- size
- density
- depth
- etc…

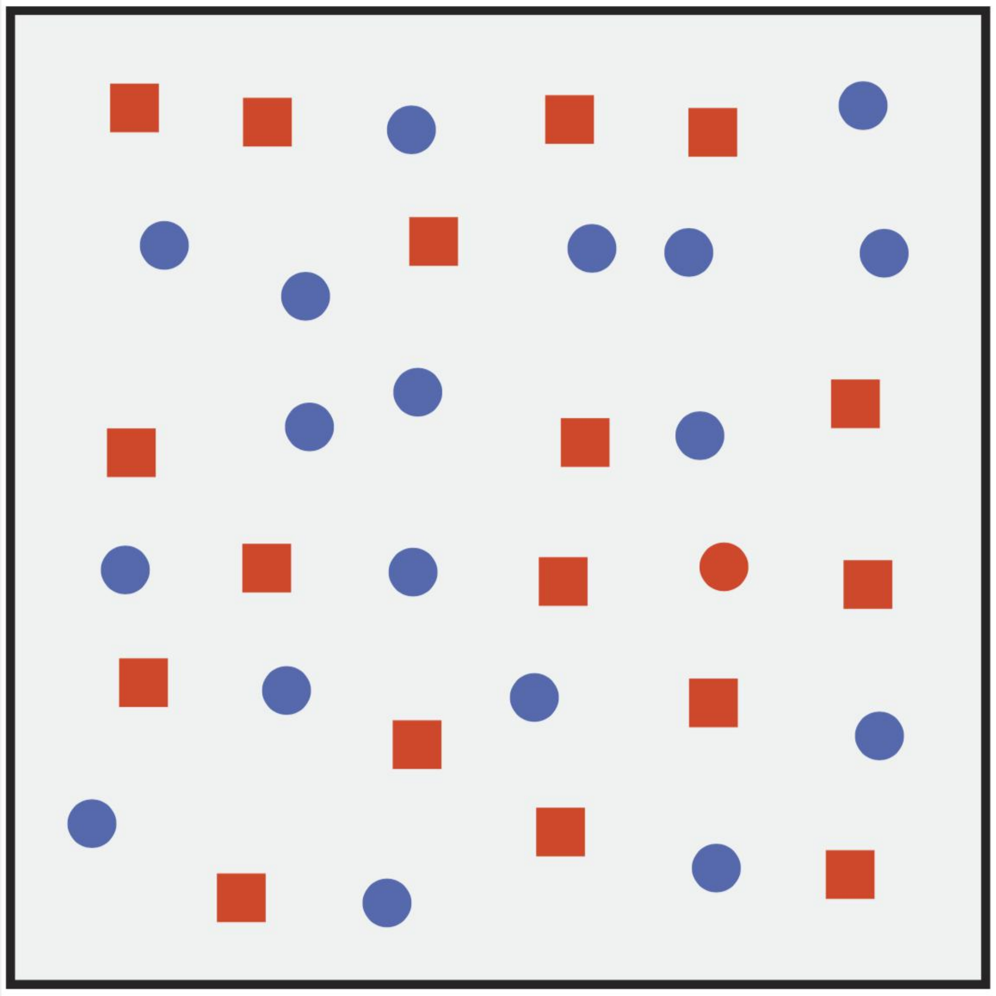
[Healey & Enns, 2012]

Detect the target object!



What isn't preattentive?

- **Most combinations of visual channels.**
- We resort to serial search.



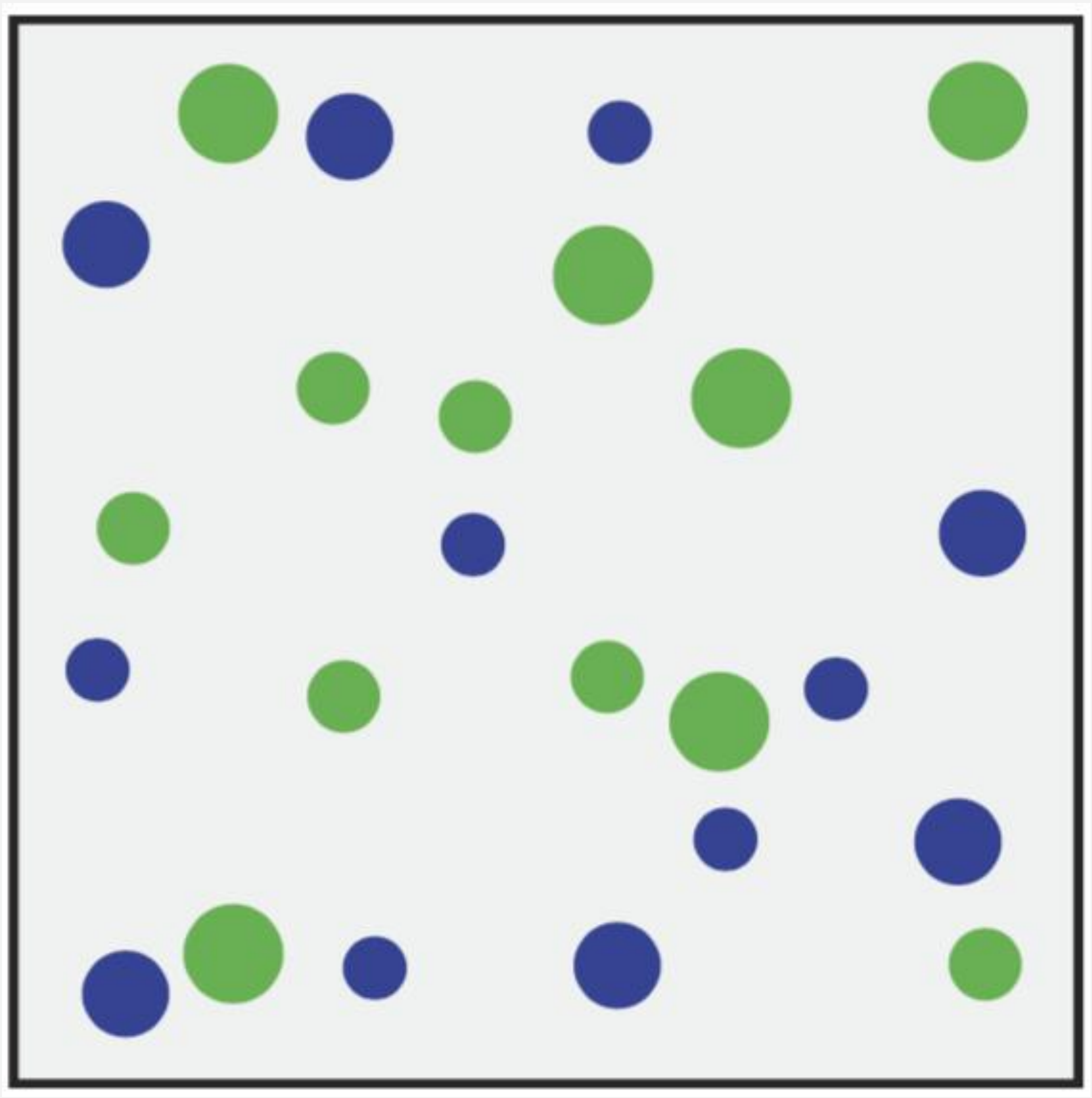
- Target: red circle
- Red: in all distractors - squares
- Circles: in all distractors - blue hue

Implications in Visual Analytics

- Recall certain analytical activities: *detect* **extrema**, **outliers**, **clusters**, etc…
- Visualization design should promote preattentive processing, so that users can quickly perform these low-level analytical tasks
- By the same token: discourage designs that could inhibit preattentive processing
  - **Visual channel interference**

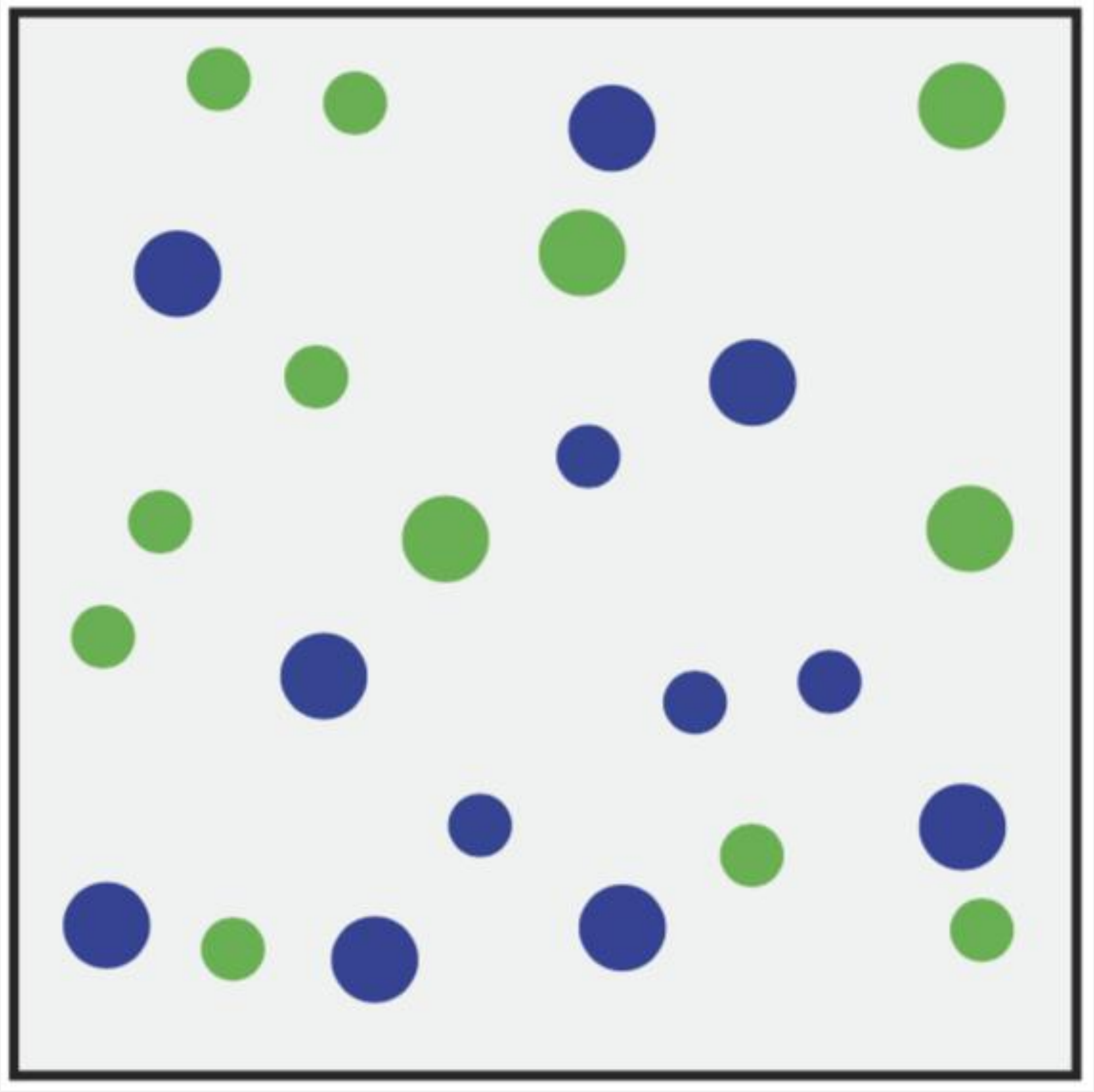
Ensemble Coding

- Humans are effective in determining **summaries.**

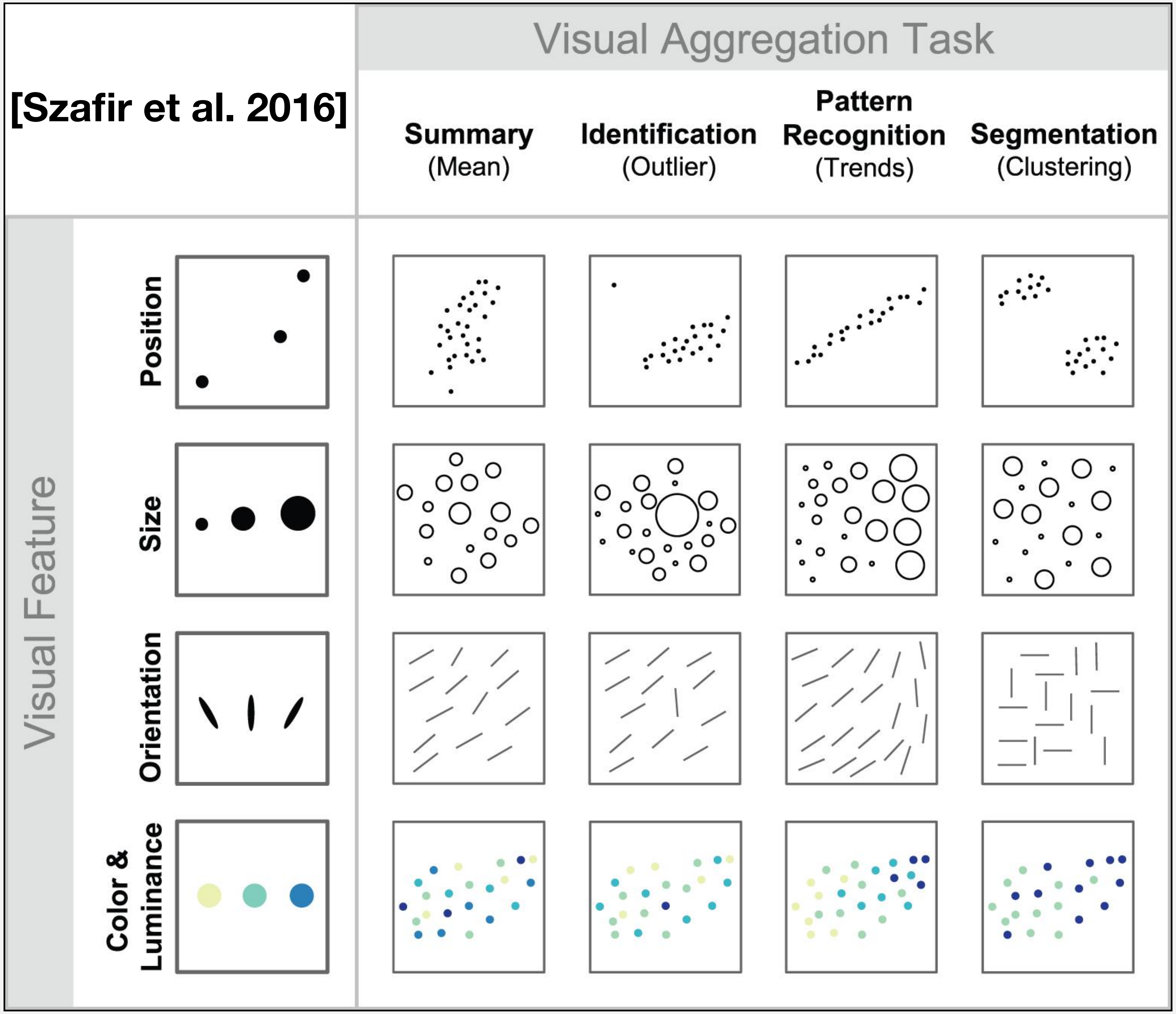


Ensemble Coding

- Humans are effective in determining **summaries.**



Types of Ensemble Coding

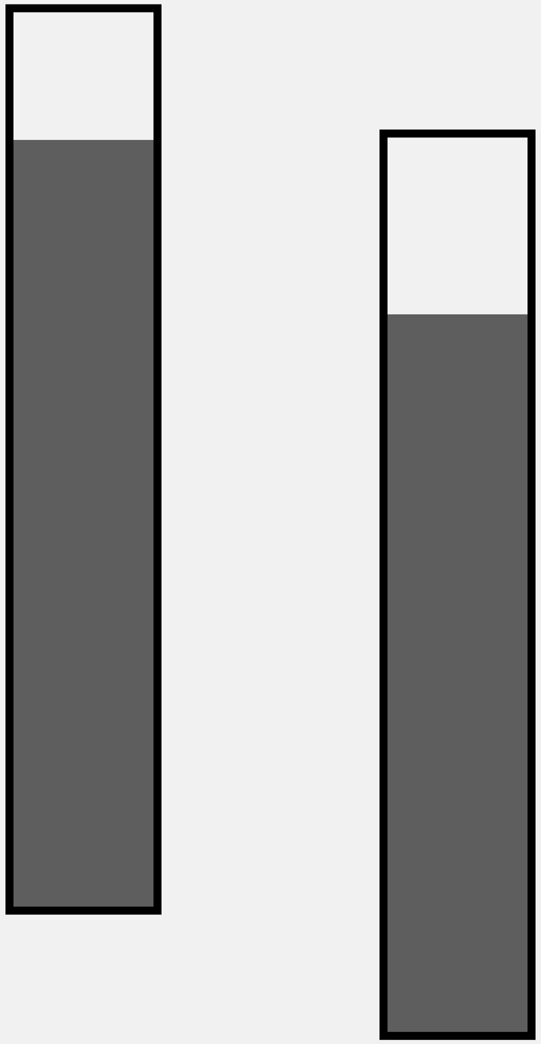


Visual Tasks

- Other types of tasks that we perform when reading a visualization:
  - Identification: *read a value* corresponding to a mark
  - *Compare values* represented by two marks

Compare two bars

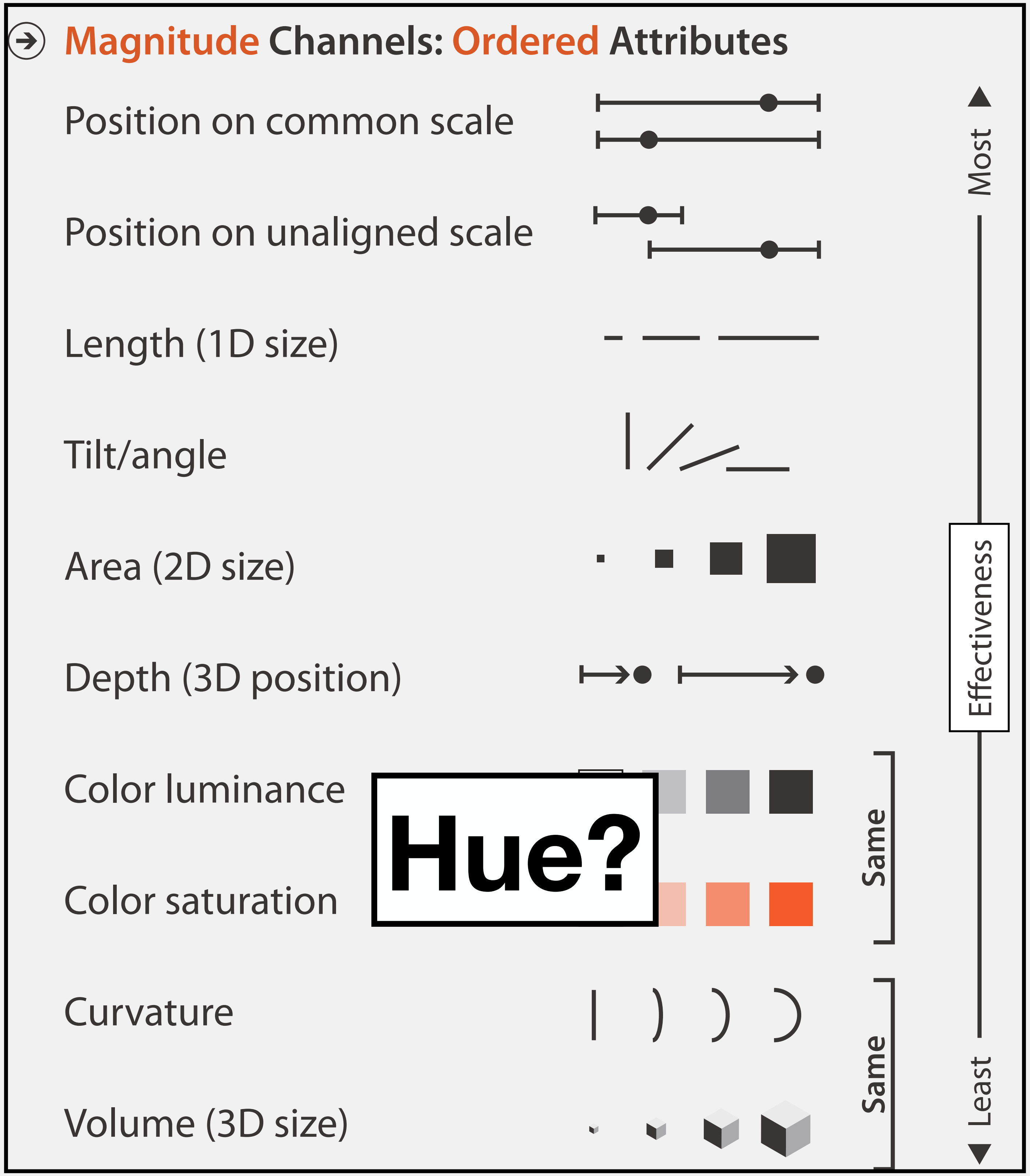
Compare two bars



Compare two bars

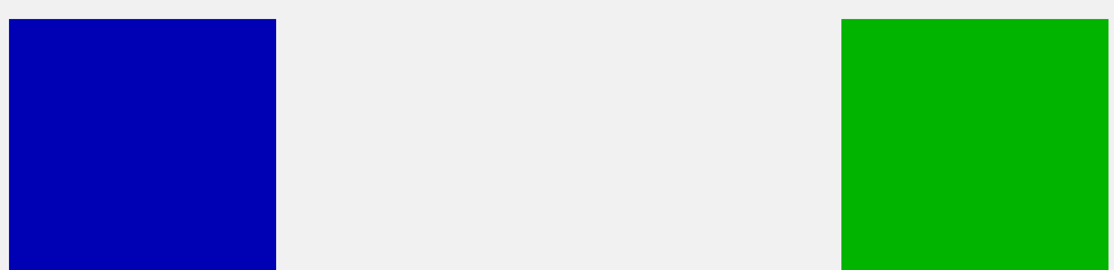
Same difference in length! Easier to perceive the difference when the absolute length is smaller.

Channel Rankings



Color

- We assign color to *everything* we draw.
- Color plays an important role in the effectiveness of a visualization.
- Issue at hand: how do we define a **visual range** for color?
- Red-Green-Blue
- Which color *appears* brighter?



Color Spaces

- RGB: treated as individual visual channels, a poor color space

- We want a color space that best supports the types of data we previously discussed: **ordered** and **categorical**.

- Ordered: colors that we can naturally rank

- Categorical: colors that we can name, uniquely identify

L A B **b:** blue-yellow

luminance: perceived brightness **a:** green-red

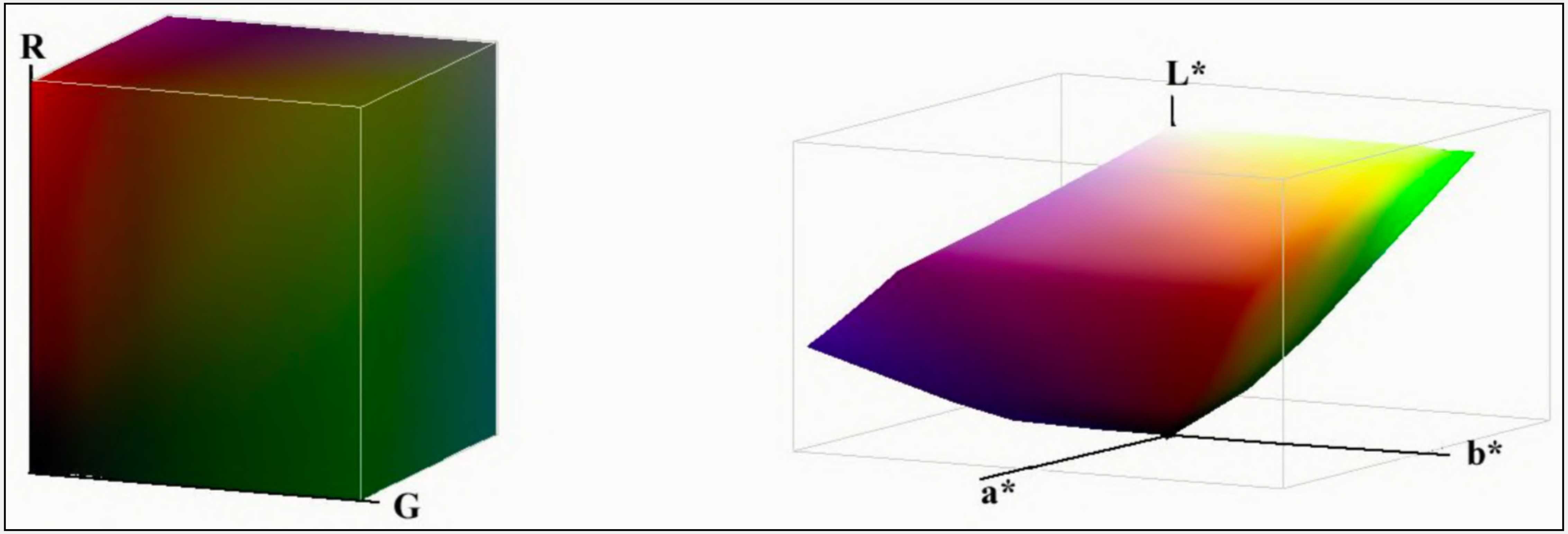
Perceptually uniform: at *any* point in LAB, if we move in *any* direction by at least 1 unit, we perceive the difference.

LAB and HCL

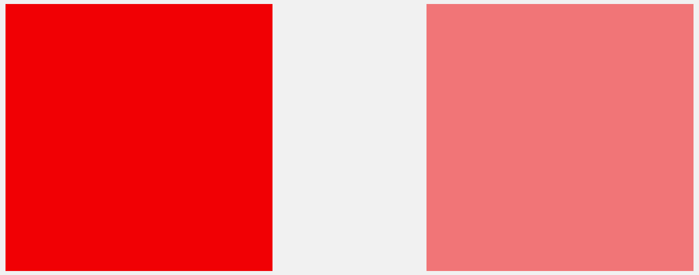
- More convenient color space: **HCL**
    - Hue: amount of red-green, blue-yellow
    - Chroma: color purity, or the saturation of a color
    - Luminance: brightness

One Drawback: Color Gamut

- Color gamut: all possible colors offered by a given color space or monitor display.
- Most monitors: RGB.
- RGB and LAB?



Color Design Considerations



- Choice of colors greatly influences user attention.
- Prioritize the colors of graphics.
    - Less important graphics (axes, guides, legends): color should not draw attention
    - Important graphics (graphical marks, title, label): use color to draw attention
    - Bright, saturated colors: use sparingly, if at all

Color and text

Hi, I am text. Can you read me?

Color and text

Hi, I am text. Can you read me?

Color and text

Hi, I am text. Can you read me?

Color and text

Hi, I am text. Can you read me?

Luminance for Design

- “Get it right in black and white” - Maureen Stone

Hi, I am text. Can you read me?

$$\Delta_L = 10$$

Luminance for Design

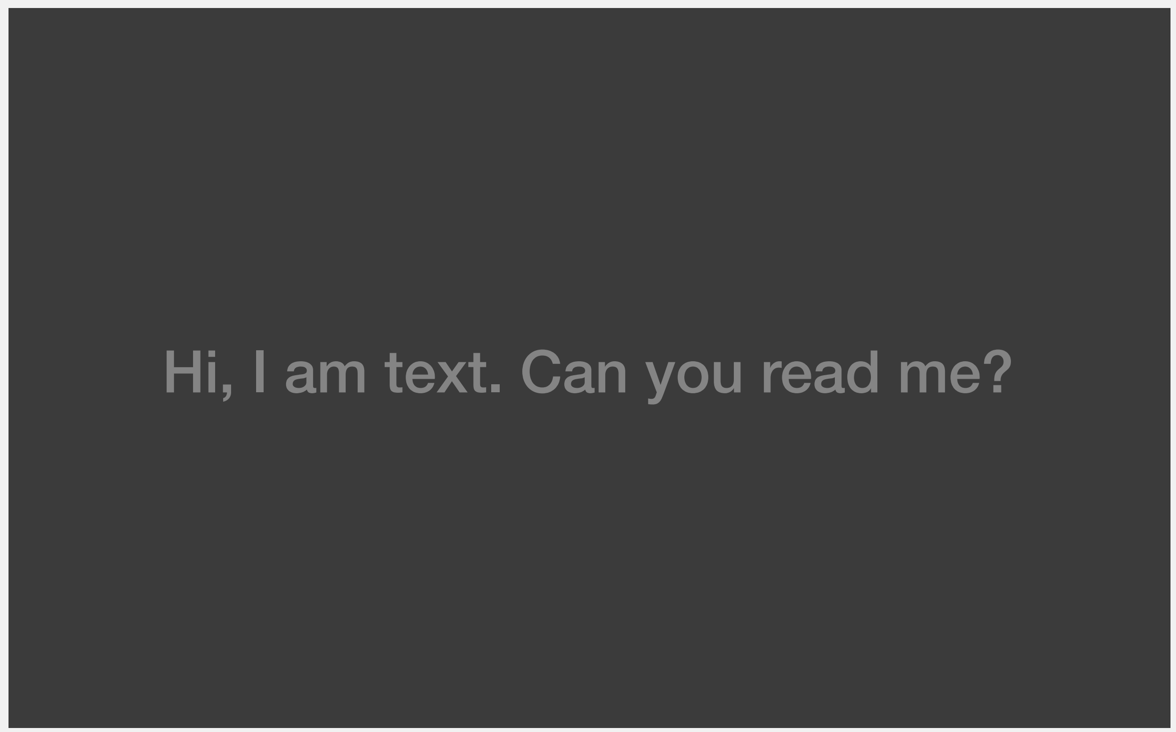
- “Get it right in black and white” - Maureen Stone

Hi, I am text. Can you read me?

$$\Delta_L = 20$$

Luminance for Design

- “Get it right in black and white” - Maureen Stone



$$\Delta_L = 30$$

Luminance for Design

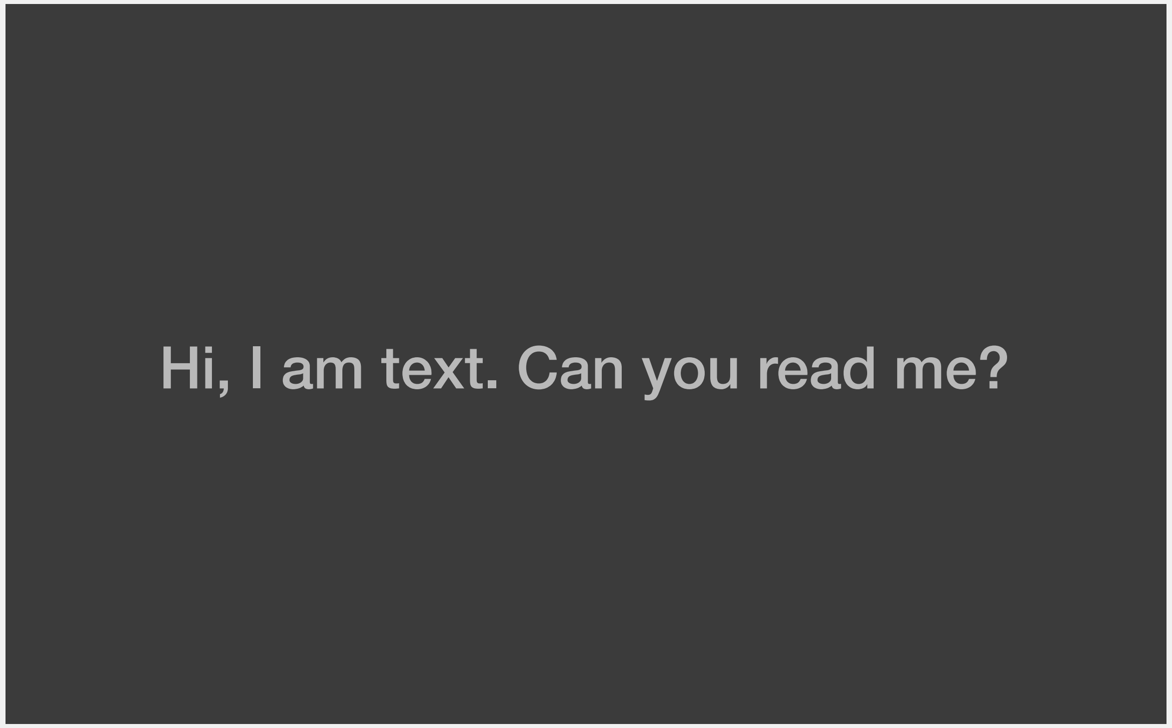
- "Get it right in black and white" - Maureen Stone

Hi, I am text. Can you read me?

$$\Delta_L = 40$$

Luminance for Design

- “Get it right in black and white” - Maureen Stone



$$\Delta_L = 50$$

Luminance for Design

- “Get it right in black and white” - Maureen Stone

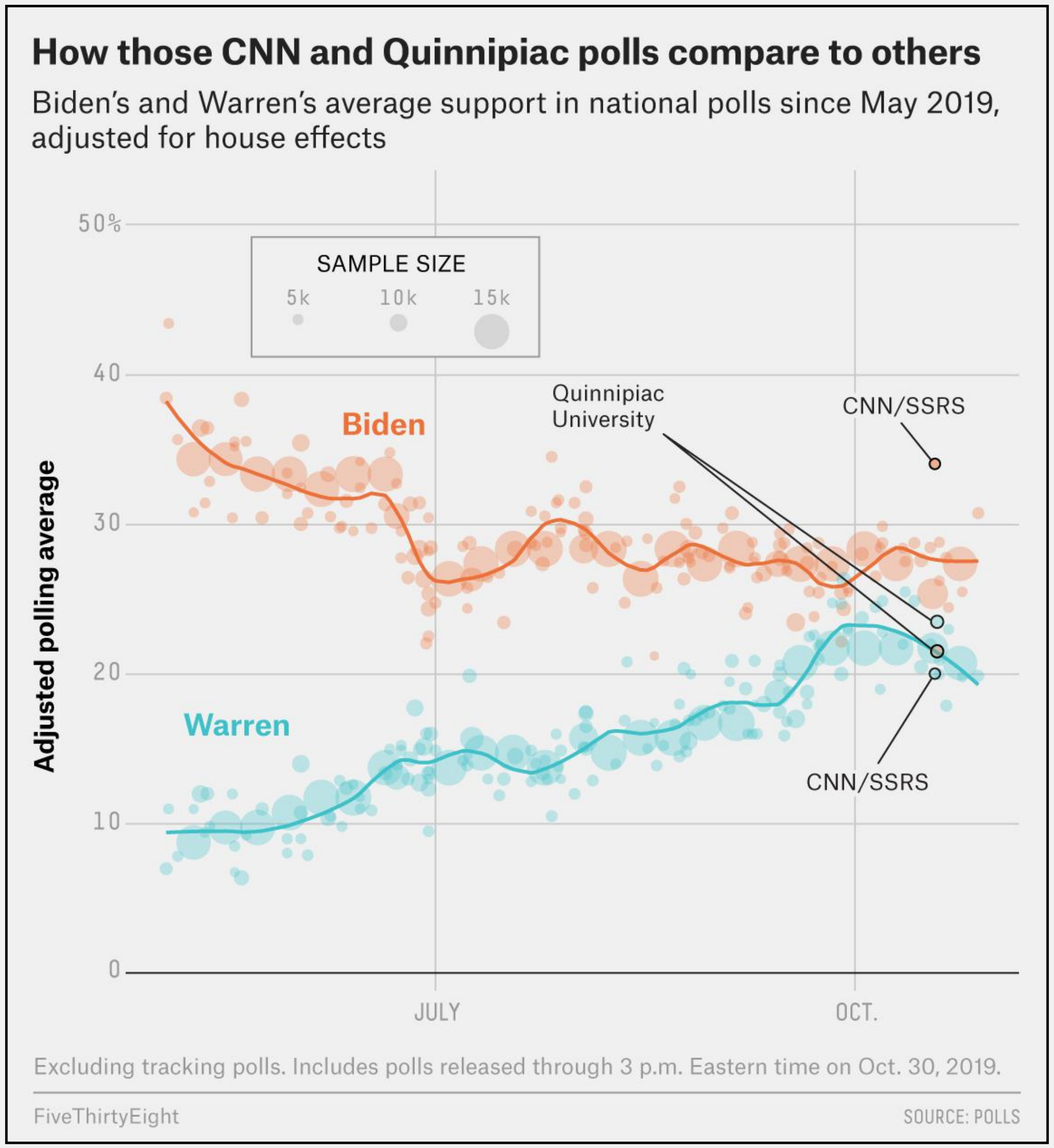
Hi, I am text. Can you read me?

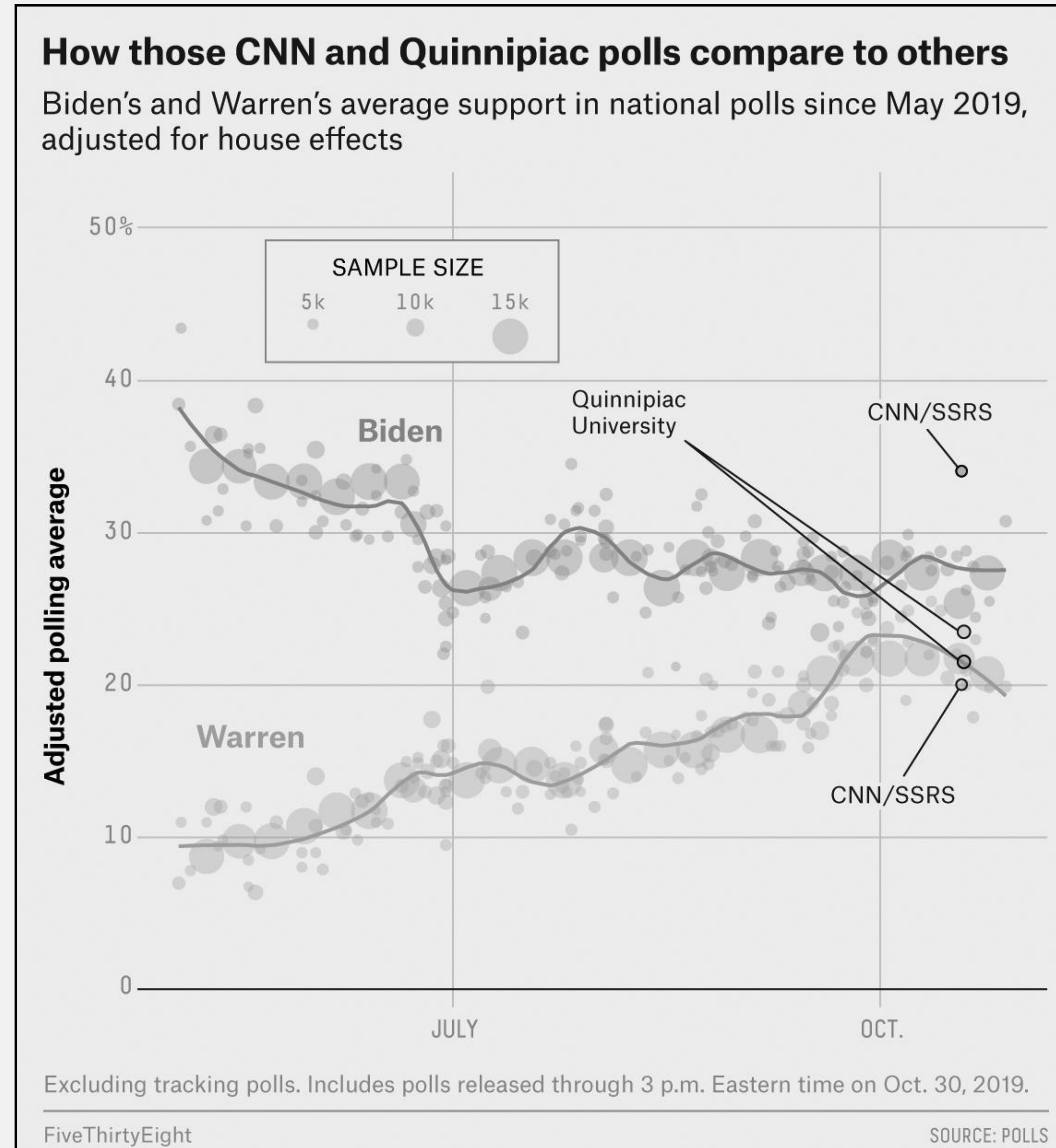
then consider chroma/hue

Luminance Contrast Inverted

Hi, I am text. Can you read me?

Example





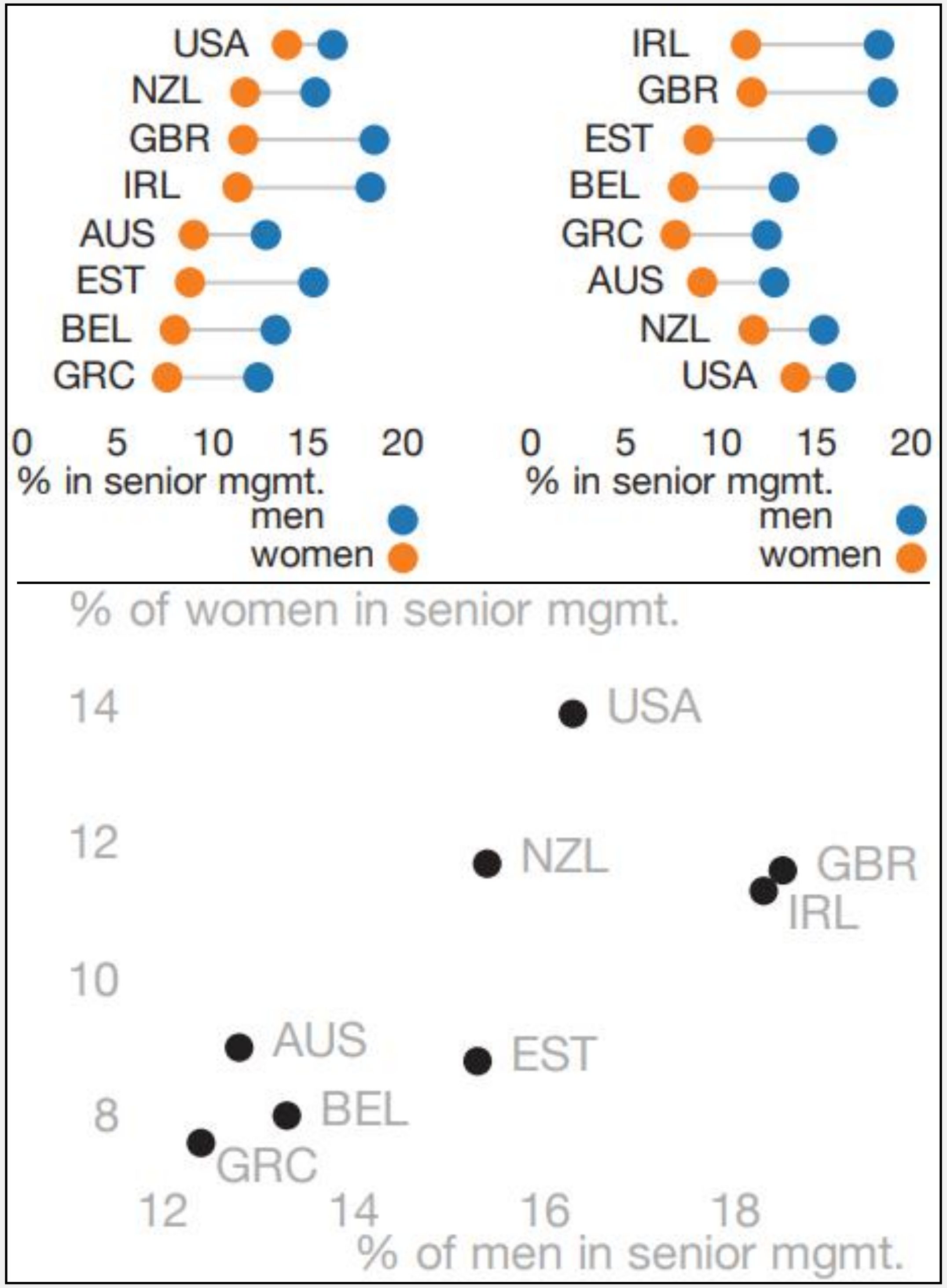
https://fivethirtyeight.com/features/biden-up-15-warren-up-7-are-primary-polls-too-far-apart/

Algebraic Visualization Design

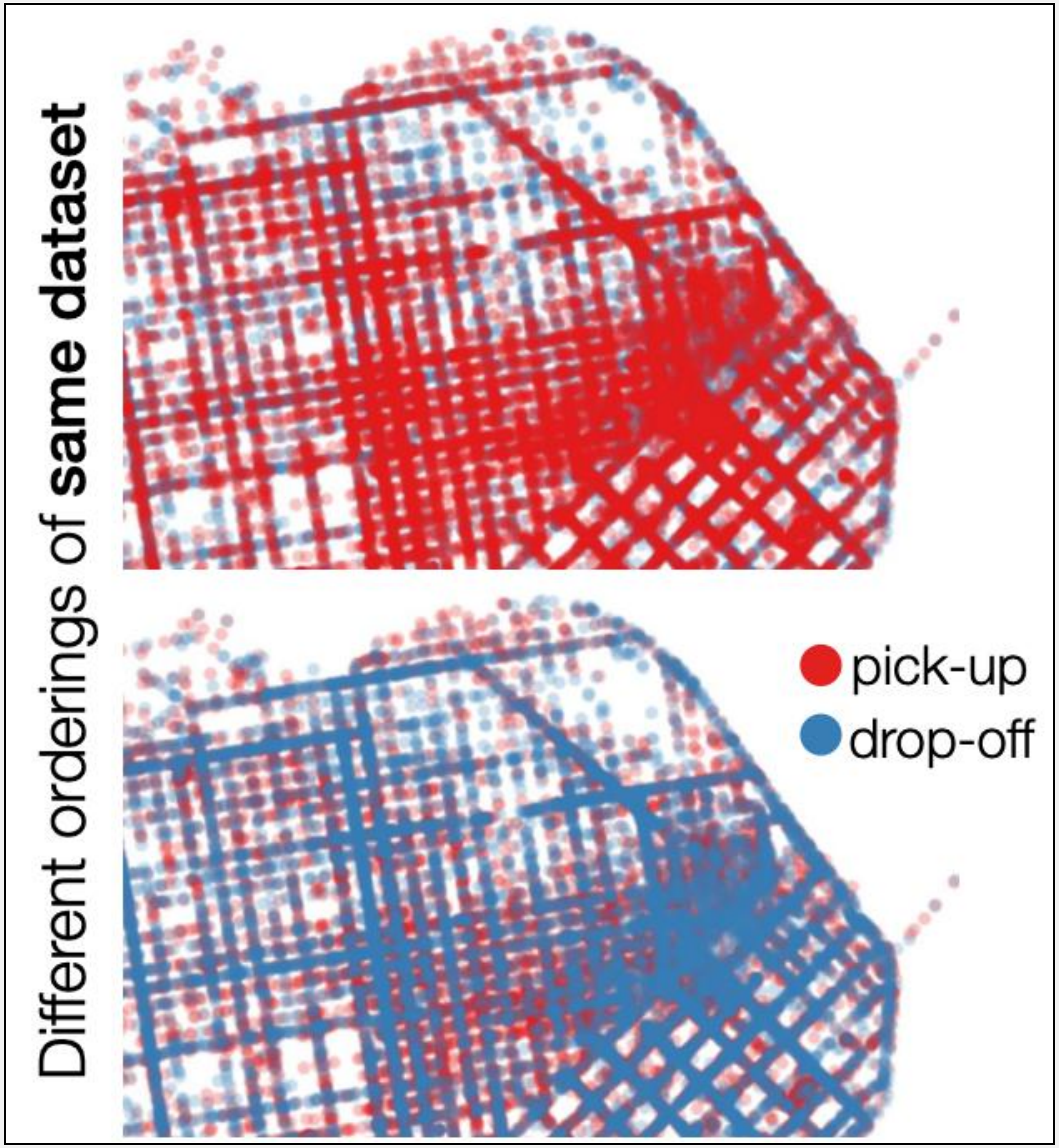
[Kindlmann & Sheidegger 2014]

- A "what-if" approach to validating visualization design.
- What if I change my data representation? (**Representation Invariance**)
- What if two visualizations are the same, does this imply the data is the same as well? (**Unambiguous Data Depiction**)
- What if I change the data, can I change the visualization in a meaningful way? (**Correspondence Principle**) (vice-versa too)

Representation Invariance



Representation Invariance (2)



Unambiguous Data Depiction