CS5346

INFORMATION VISUALIZATION

THE DESIGN PROCESS – Stage 4

Ref: Chapters 3-6, Data Visualization by Andy Kirk

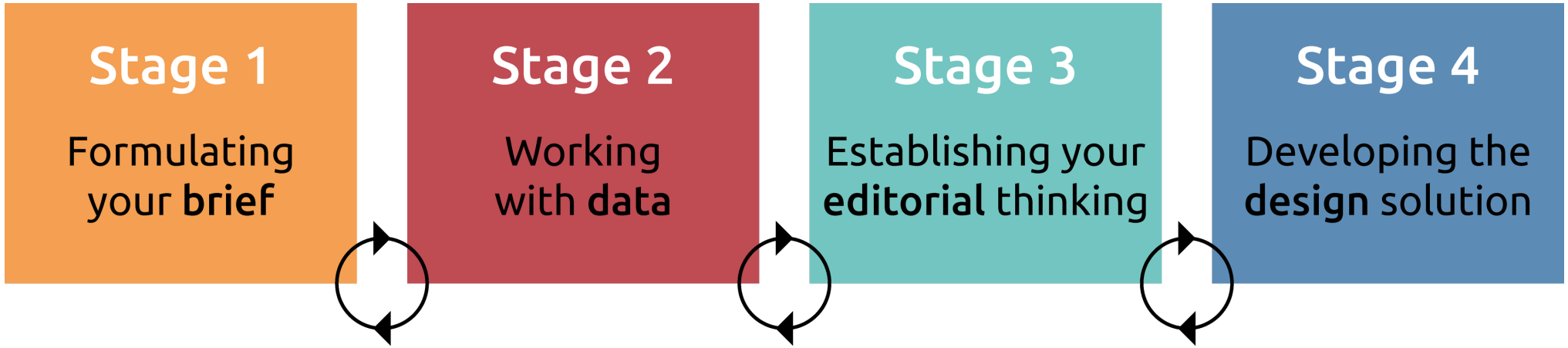
MARKS and CHANNELS

Ref: Chapter 5, Visualization Analysis and Design, by Tamara Munzner

Assoc Prof. Bimlesh Wadhwa | School of Computing | NUS | bimlesh@nus.edu.sg

Visualization Design Process

A 4 Step process



Note: There are good and bad solutions. There are no perfect visualizations.

Stage 4: Visual manifestation of preparatory work

Data representation, interactivity, annotation, colour, composition

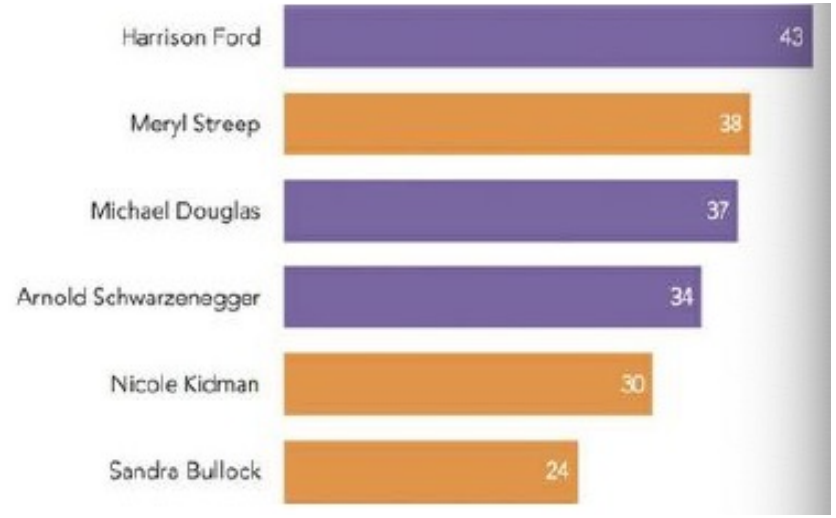
We shall begin with **Data representation**
Other sub-topics will appear(not necessarily in sequence) in the later lectures

Data representation aka Visual representation

Visual Representation of Data

- **Visual Encoding** (aka Data Encoding) : **assigning visual properties to data** values.

Actor	Gender	Years since First Movie
Harrison Ford	Male	43
Maryl Streep	Female	38
Michael Douglas	Male	37
Arnold Schwarzenegger	Male	37
Nicole Kidman	Female	30
Sandra Bullock	Female	24



Important Note: Later we also learn how do we document Visual Encoding (aka Data encoding).

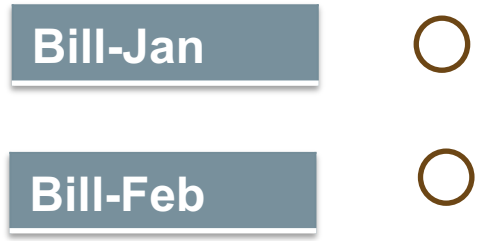
Building blocks for visual encoding : Marks and Channels

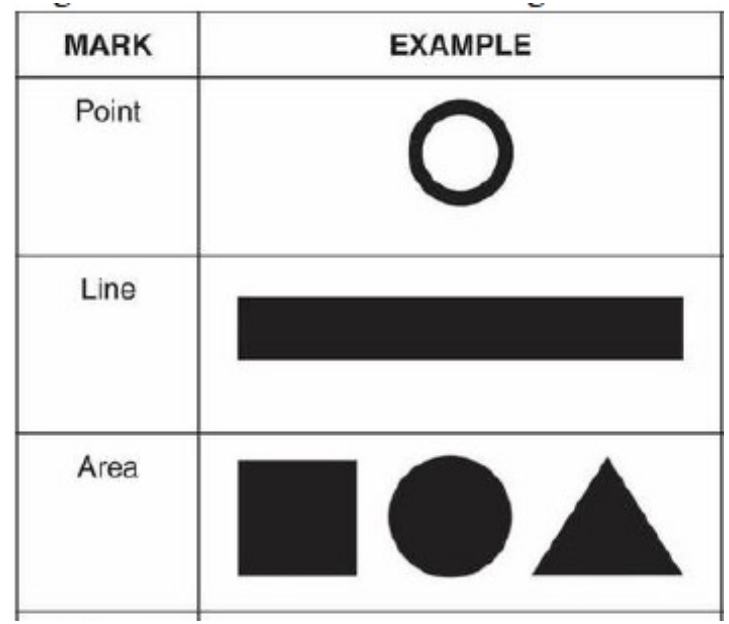
Marks

- **Marks** are visible features like dots, lines and areas

- A mark can represent **a record or instance of data**

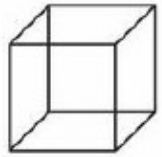
- e.g. if you received a phone bill for a given month.





A **set of marks** would represent **a set of records or instances** (e.g. the monthly phone bills for Jan-Oct 2015).

Different types of Marks

MARK	EXAMPLE	DESCRIPTION
Point		The *point* mark has no variation ('constant') in the spatial dimension. It is largely a placeholder commonly used to represent a quantity through position on a scale, forming the basis of, for example, scatter plots.
Line		The *line* mark has one ('linear') spatial dimension. It is commonly used to represent quantitative value through variation in size, forming the basis of, for example, the bar chart.
Area		The *area* mark has two ('quadratic') spatial dimensions. It is commonly used to represent quantitative values through variation in size and position, forming the basis of, for example, bubble plots.
Form		The *form* mark has three ('cubic') spatial dimensions. It might be used to represent quantitative values through variation in size (specifically, through volume), forming the basis of, for example, a 3D proportional shape chart.

Marks as Links

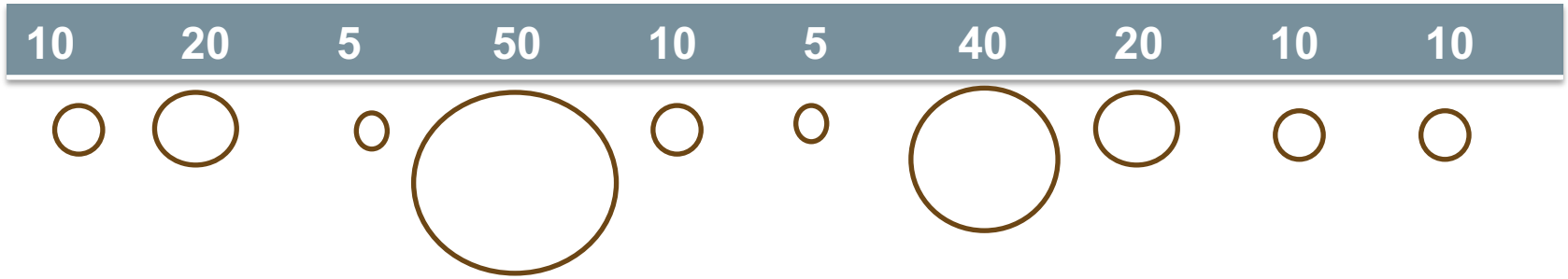
Containment

Connection

Channels

- **Channels** are variations applied to the appearance of Marks

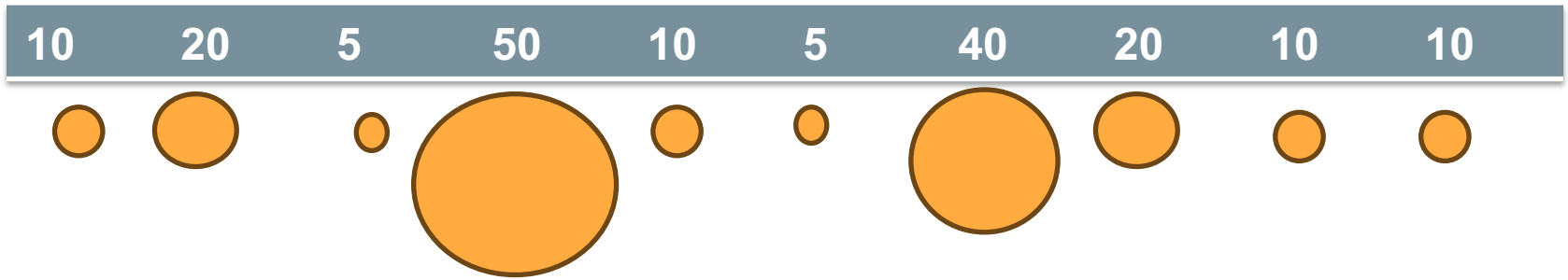
Example- If you had 10 **marks**, one for the amount of each of Jan-Oct phone bill, you could use the **size** of each mark to represent the phone bill totals.



Channels represent the values held by different quantitative or categorical variables.

Channels

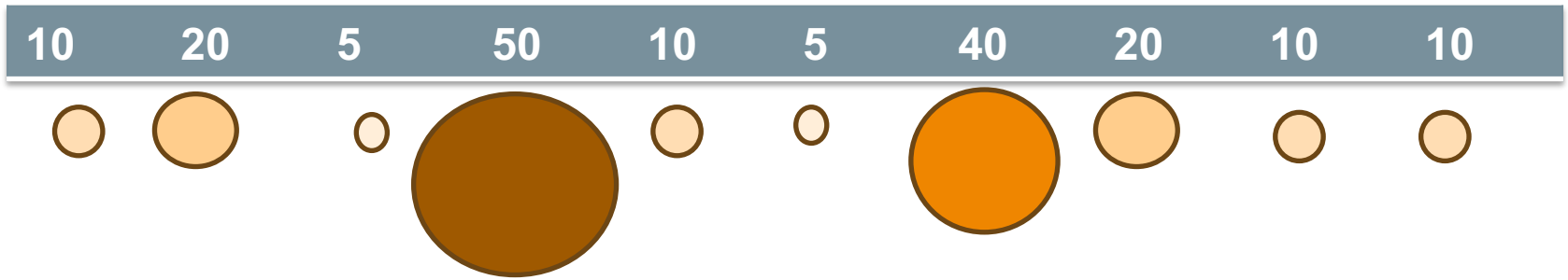
- One can encode an attribute with a single channel or multiple channels.



Channel	Description
Size	Size is used to represent **quantitative values** where larger the size of the mark, larger the value
Hue	Hue is typically used for distinguishing different **categorical data** values but can also be used in conjunction with other color properties to separate certain quantitative scales

Channels

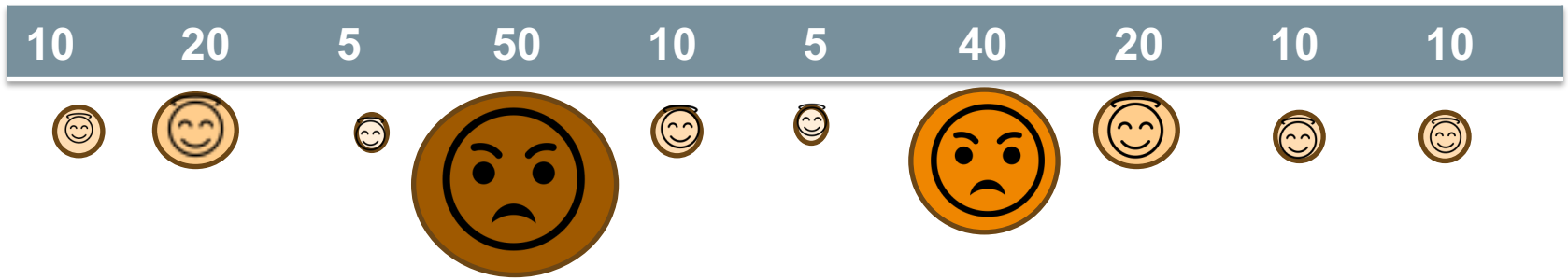
- One can encode an attribute with a single channel or multiple channels.



Channel	Description
Size	Size is used to represent **quantitative values** where larger the size of the mark, larger the value
Hue	Hue is typically used for distinguishing different **categorical data** values but can also be used in conjunction with other color properties to separate certain quantitative scales

Channels

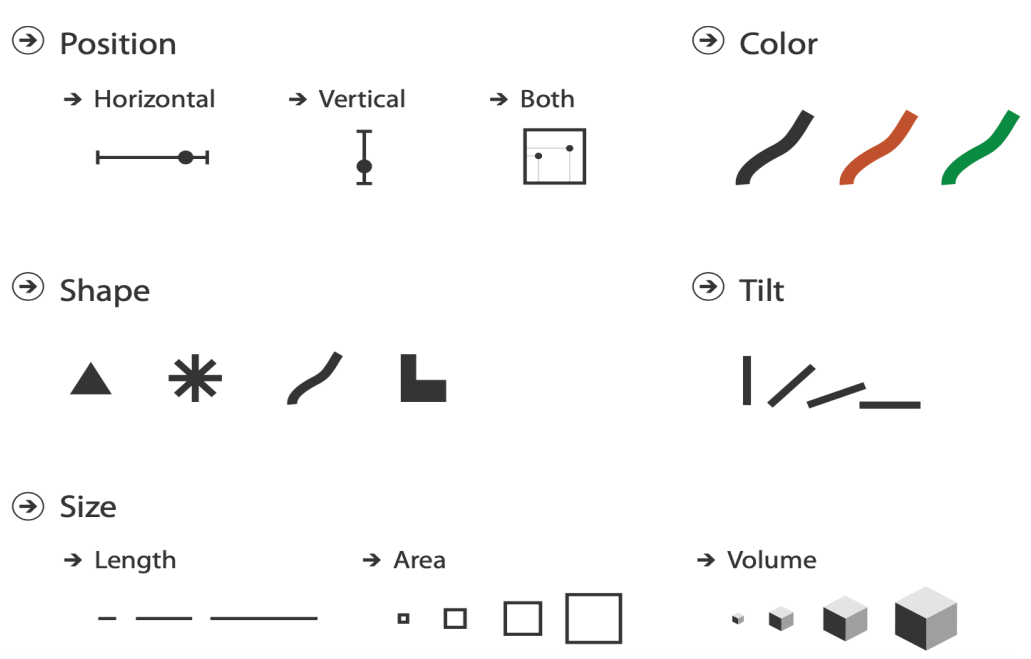
- One can encode an attribute with a single channel or multiple channels.



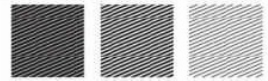
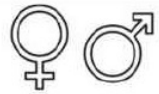
Channel	Description
Size	Size is used to represent **quantitative values** where larger the size of the mark, larger the value
Hue 	Hue is typically used for distinguishing different **categorical data** values but can also be used in conjunction with other color properties to separate certain quantitative scales
Symbol/Shape	Symbols or shapes are generally used with point marks to indicate categorical association

Examples of Channels

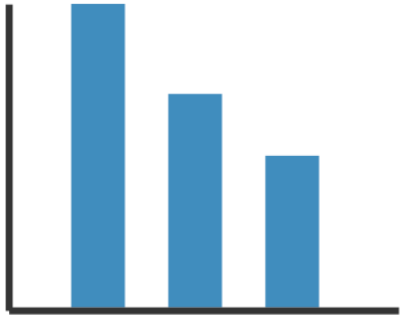
Position (Horizontal, Vertical)
Color (Hue, Saturation, Luminance)
Shape (or symbols)
Tilt (or Angle)
Size (1D-Line, 2D- Area, 3D- Volume)



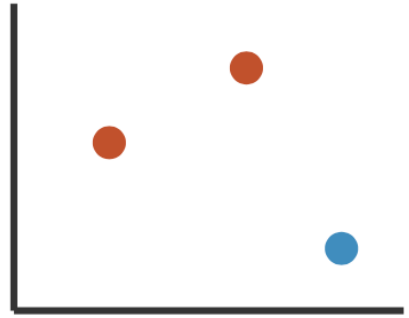
Examples of Channels

Channel	Example	Description
QUANTITATIVE ATTRIBUTES		
Position		Position along a scale is used to indicate a quantitative value.
Size		Size (length, area, volume) is used to represent quantitative values based on proportional scales where the larger the size of the mark, the larger the quantity.
Angle/Slope		Variation in the size of angle forms the basis of pie chart sectors representing parts-of-a-whole quantitative values; the larger the angle, the larger the proportion. The slope of an incline formed by angle variation can also be used to encode values.
Quantity		The quantity of a repeated set of point marks can be used to represent a one-to-one or a one-to-many unit count.
Colour: Saturation		Colour saturation can be used (often in conjunction with other colour properties) to represent quantitative scales; typically, the greater the saturation, the higher the quantity.
Colour: Lightness		Colour lightness can be used (often in conjunction with other colour properties) to represent quantitative scales; typically, the darker the colour, the higher the quantity.
Pattern		Variation in pattern density or difference in pattern texture can be used to represent quantitative scales or distinguish between categorical ordinal states.
Motion		Motion is more rarely seen but it could be used as a binary indicator to draw focus (motion vs no motion) or by incorporating movement through speed and direction to represent a quantitative scale ramp.
CATEGORICAL ATTRIBUTES		
Symbol/shape		Symbols or shapes are generally used with point markers to indicate categorical association.
Colour: Hue		Colour hue is typically used for distinguishing different categorical data values but can also be used in conjunction with other colour properties to represent certain quantitative scales.
RELATIONAL ATTRIBUTES		
Connection/Edge		A connection or edge indicates a relationship between two nodes. Sometimes arrows may be added to indicate direction of relationship, but largely it is just about the presence or absence of a connection.
Containment		Containment is a way of indicating a grouping relationship between categories that belong to a related hierarchical 'parent' category.

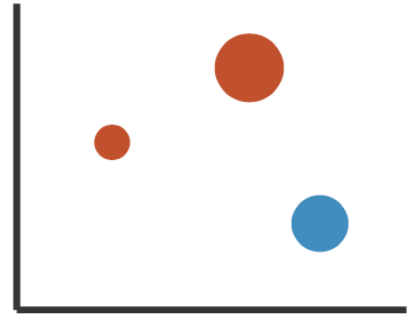
Using Marks and Channels in Charts



Mark: Line
Channel: Vertical position

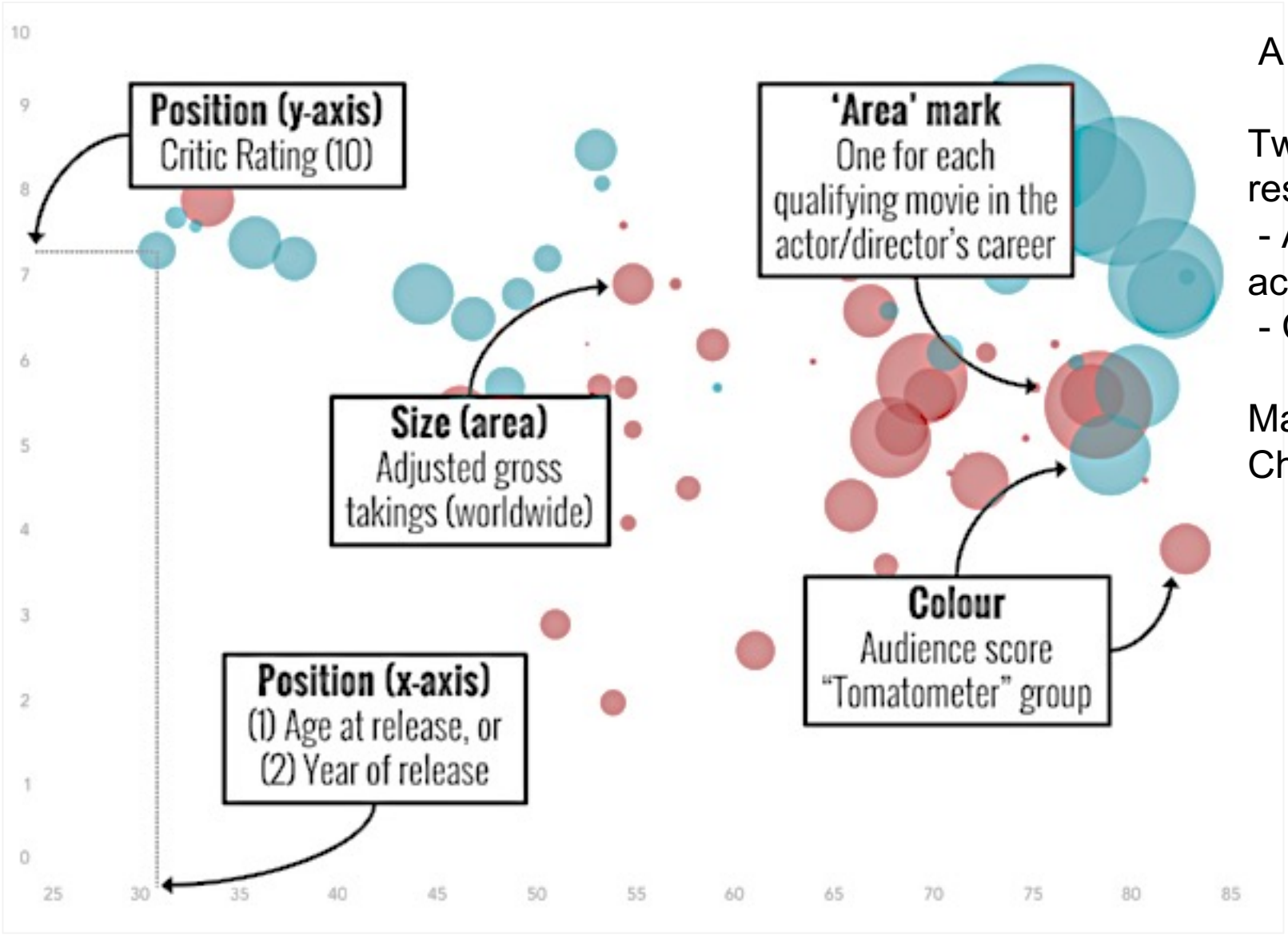


Mark: point
Channels: Vertical position, Horizontal position, Hue



Mark: point
Channels: Vertical position, Horizontal position, Hue, size

Using Marks and Channels in Charts



A bubble plot.

Two quantitative variables along the respective axes.
- Age at release (or year of release) across the x-axis
- Critic Rating across y-axis

Mark – Area for each movie
Channels – Color for A.Score
Size for Gross takings

Visual Encoding : How do we communicate it ?

Example

Canvas→ Modules→ Articles→ Visual encoding(aka Data-Encoding) in charts .pdf

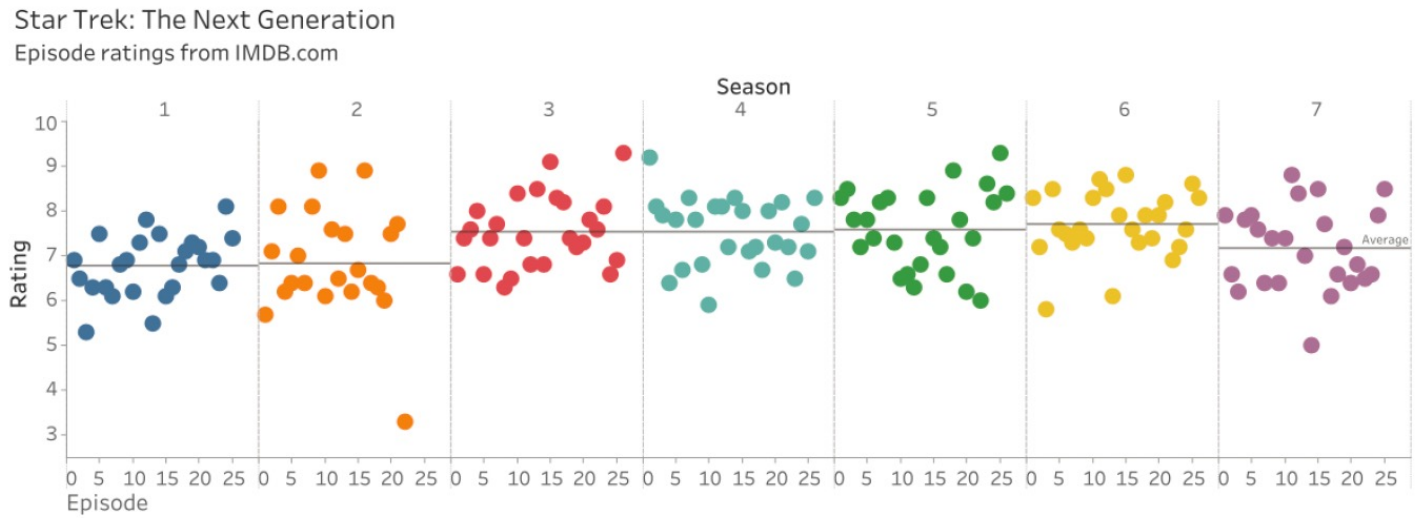


FIGURE 1.12 Every episode of *Star Trek: The Next Generation* rated.

Source: IMDB.com

TABLE 1.3 Data used in Figure 1.12.

Data	Data Type	Encoding	Note
Episode	Categorical	Position	Each episode is represented by a dot. Each dot has its own position on the canvas.
Episode Number	Ordinal	Position	The x-axis shows the number of each episode in each season.
Season	Ordinal	Color Position	Each season is represented by a different color (hue). Each season also has its own section on the chart.
IMDB rating	Ordinal	Position	The better the episode, the higher it is on the y-axis.
Average season rating	Quantitative	Position	The horizontal bar in each pane shows the average rating of the episodes in each season. There is some controversy over whether you should average ordinal ratings. We believe that the practice is so common with ratings it is acceptable.

Characteristics of Channels

Selective

Is a mark distinct from other marks? Can we make out the difference between two marks?

Associative

Does it support grouping?

Quantitative

Can we quantify the difference between two marks?

Order

Can we see a change in order?

Example

Channel	Selective	Associative	Quantitative	Order
Position	Y	Y	Y	Y
Hue	Y	Y	No	No

Channel Rankings

Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

Same

Same

< Least ----------Effectiveness----------Best >

Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

[Munzner, 2014]

[Munzner, 2014]

MAGNITUDE and IDENTITY channels

- **MAGNITUDE** channels

  Tell us HOW much something is

  e.g. use Position, Length, Color(Saturation) channels for Ordinal and Quantitative Data

- **IDENTITY** channels

  Tell us information about WHAT something is or WHERE it is

  e.g. use Shape, Color(hue) channels for Categorical data.

[Bertin, Semiology of Graphics, 1967]

Bertin's channels organisation

	Categorical	Ordinal	Quantitative
Position	Y	Y	Y
Size	Y	Y	Y
Shape	Y	N	N
Value (lightness)	Y	Y	~
Hue	Y	N	N
Orientation	Y	N	N
Texture	Y	~	N

[Jock Mackinlay, 1986. ACM Trans. Graph. 5(2)]

Mackinlay's Channels organisation

Qualitative Nominal	Qualitative Ordinal	Quantitative Interval, Ratio
Position	Position	Position
Colour (Hue)	Pattern (Density)	Size (Length)
Pattern (Texture)	Colour (Lightness)	Angle
Connection	Colour (Hue)	Size (Area)
Containment	Pattern (Texture)	Size (Volume)
Pattern (Density)	Connection	Pattern (Density)
Colour (Lightness)	Containment	Colour (Lightness)
Symbol	Size (Length)	Colour (Hue)
Size (Length)	Angle	Pattern (Texture)
Angle	Size (Area)	Connection
Size (Area)	Size (Volume)	Containment
Size (Volume)	Symbol	Symbol

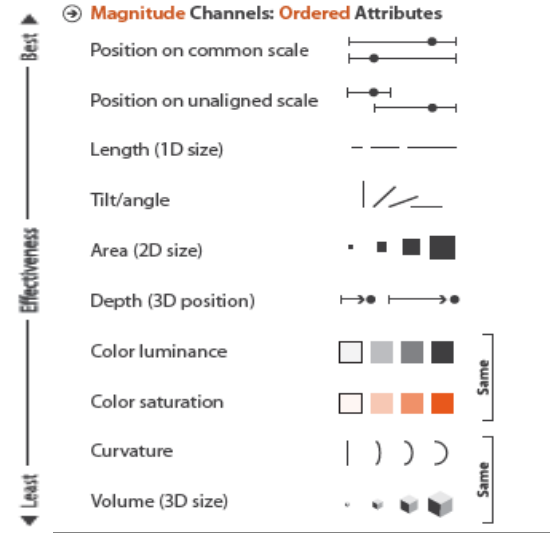
Principles of **expressiveness**

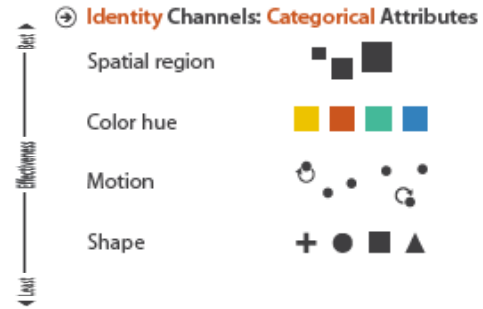
- Visual encoding should **express all of**, and **only**, the information in the dataset attributes

→ Ordered data should be shown in a way that our perceptual system intrinsically senses as ordered

→ Unordered data should **not** be shown in a way that perceptually implies an ordering that does not exist

  - Magnitude channel for ordered data
  - Identity channels for categorical data



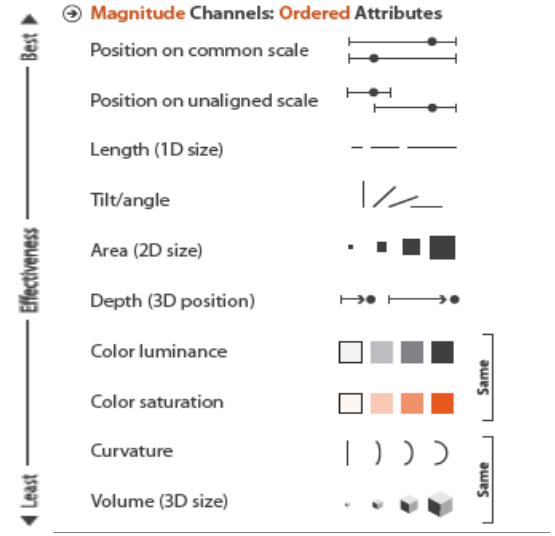


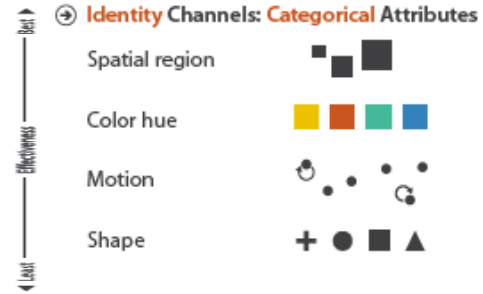
***It is possible in theory to use a magnitude channel for categorical data or an identity channel for ordered data, that choice would be a poor one and the expressiveness principle would be violated*

Principles of **effectiveness**

- The **importance of attribute** should match with the salience of channel i.e. its **noticeability**.

→ The most important attributes should be encoded with the most effective channels in order to be most noticeable.



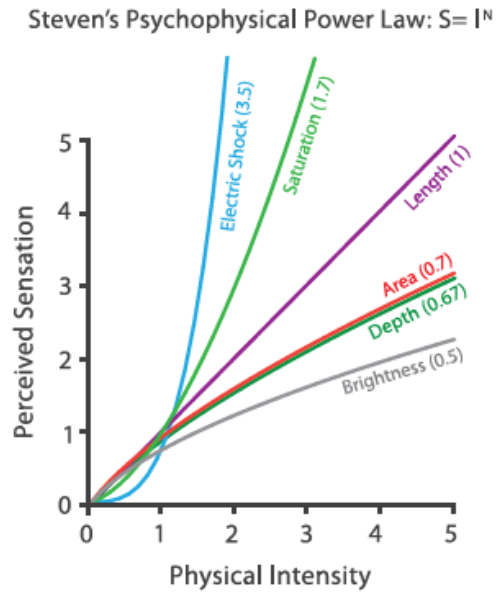


Channel Effectiveness

Accuracy : how close is human perceptual judgement to some objective measurement of the stimulus?

- [many studies] Our perception of **Position** and **Length** are a close match to the true value compared to other channels e.g. **Area, Brightness**, or **Saturation**.

FYI : Steven's Power Law
The apparent magnitude of all sensory channels follows a power function based on the stimulus intensity:
$S = I^N$
where S is the perceived sensation and I is the physical intensity.

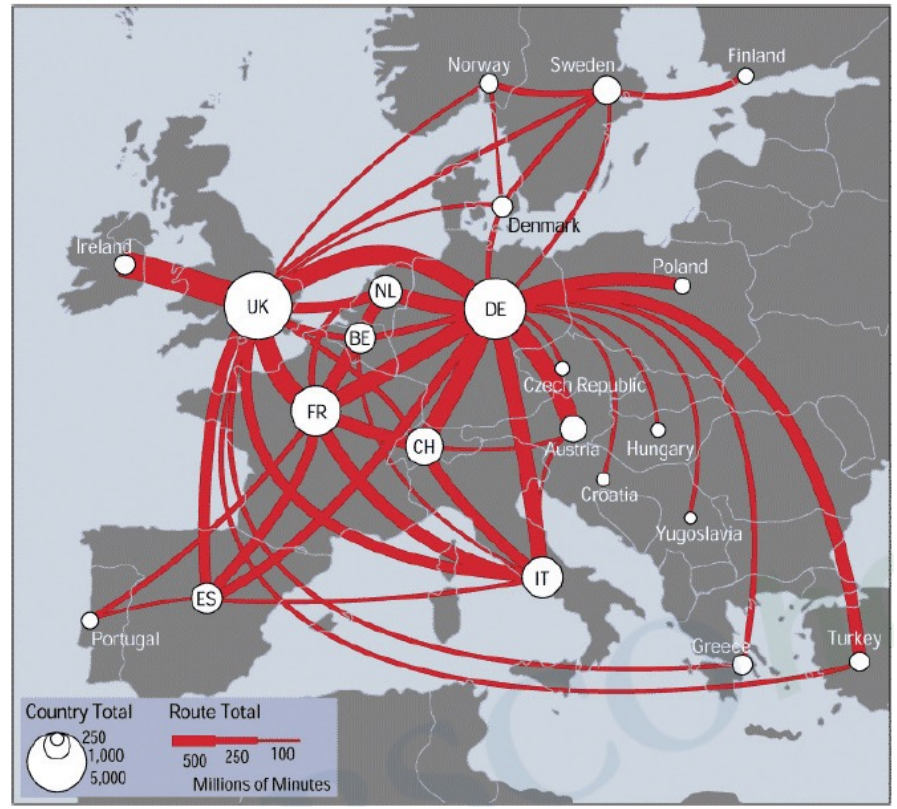


Channel Effectiveness

Discriminability: Are the difference between encoded items perceptible to the human as intended?

- #Bins in a visual channel : each bin is a distinguishable step or level from the other
  - e.g. Line width: Changing line width works for limited number of steps.
    Increasing line width past that limit → mark perceived as a polygon area than a line mark

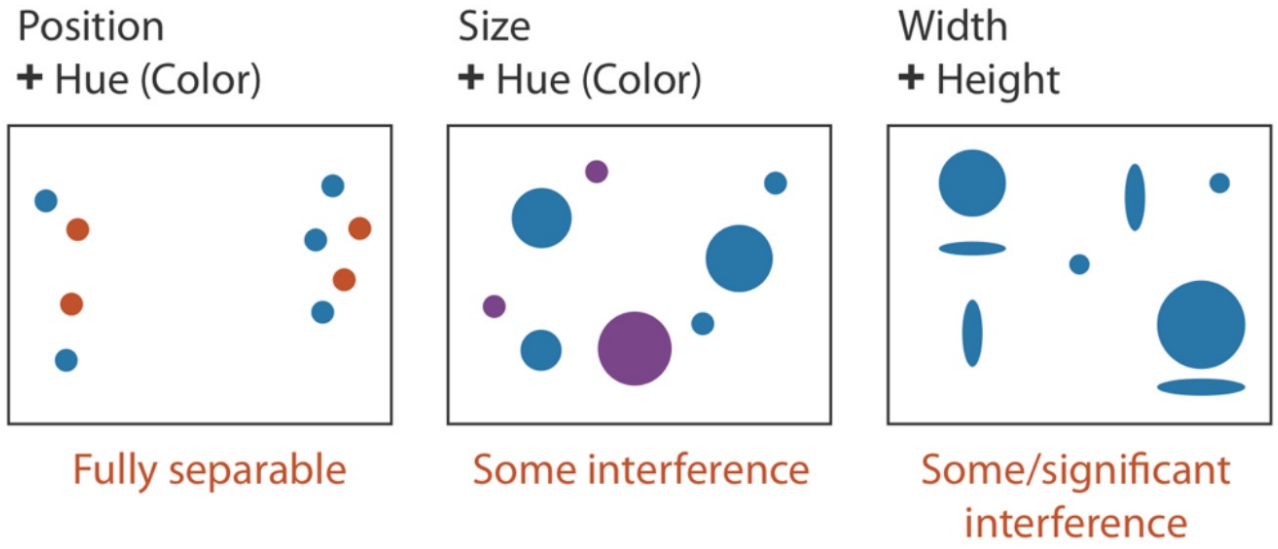
→ A small number of bins is ok if #values to encode is also small
→ the #values that need to be shown for the attribute being encoded =< #bins for the visual channel



Channel Effectiveness

Separability : Can we combine multiple visual channels ? How easy is it to focus attention on one channel when encoding with multiple channels ?

- Visual channels have dependencies and interactions with each other.



Guideline:
If the goal is to show two different attributes either of which can be attended to selectively → a separable channel pair of position and Hue

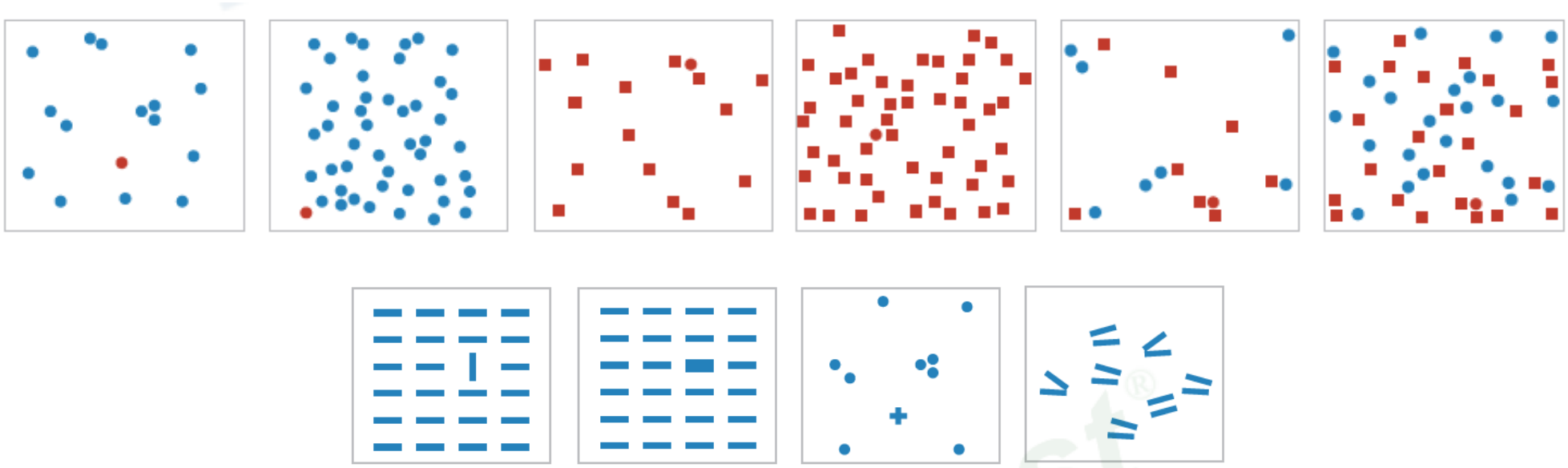
If the goal is to show a single data attribute with three categories → use integral channel pair of width and height .

Channel Effectiveness

Visual Popout : Can we make a distinct item stand out ?

- e.g. Spotting a red object in a sea of ~~grey b~~lue OR spotting a circle from a sea of squares

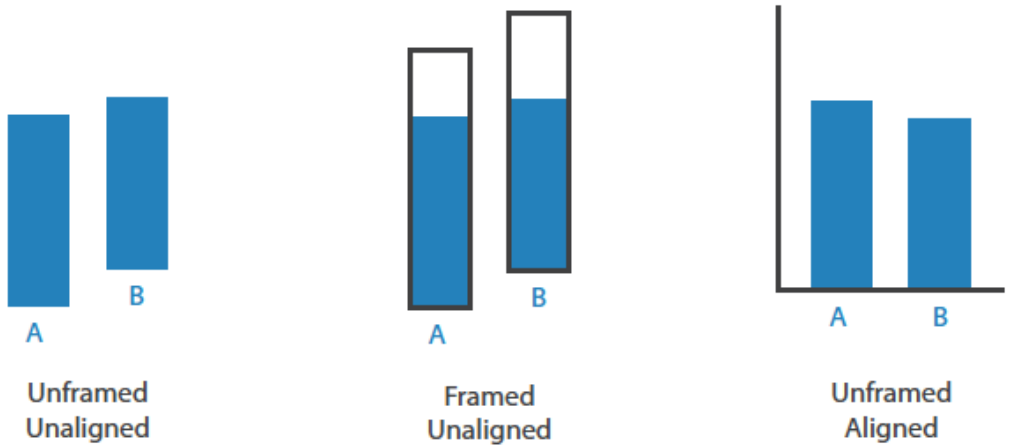
Many channels support visual popout : e.g. Tilt, Size, Shape



Additionally - Relative vs Absolute Judgement

- Amount of length difference we can detect is a percentage of the object's length

- When 2 objects are directly next to each other and aligned
  → much more precise judgements <u>than</u> when they are not aligned



Related - FYI : Weber's law: The detectable difference in stimulus intensity **I** as a fixed percentage **K** of the object magnitude $\frac{\delta I}{I} = K$

Summary

Visual Encoding

Decisions of visual representation of data

We ecode our data using two visual properties – Marks and Channels

Marks and Channels

Marks: visual elements representing data items

Channels: for variations in the visual appearance of Marks to represent the values associated with the data item

Channel ranking and effectiveness

Different ways of encoding data might offer varying degree of accuracy(effectiveness) in the perception of data values.

Channels are ranked based on the perceptual accuracy they allow when used in encoding data.