S. Rinzivillo – rinzivillo@isti.cnr.it

DATA VISUALIZATION AND VISUAL ANALYTICS

Perception and Cognition

VS

Game #4 – How many 3s?

1258965168765132168943213
5463479654321320354968413
2068798417184529529287149
2174953178195293926546831
3546516509898554684982984

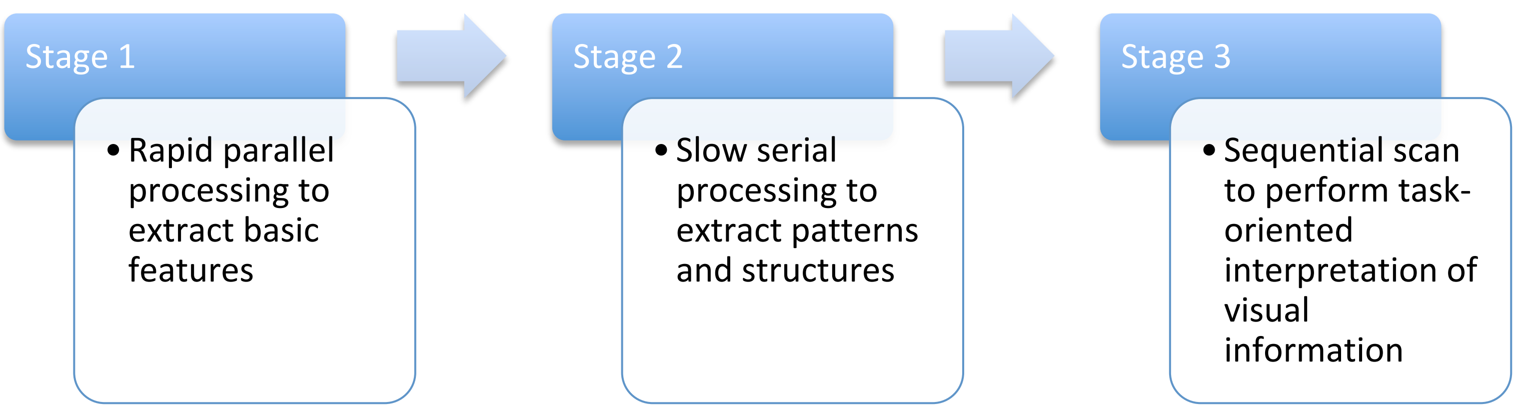
1258965168765132168943213
5463479654321320354968413
2068798417184529529287149
2174953178195293926546831
3546516509898554684982984

How many "3"?

1258965168765132168943213
5463479654321320354968413
2068798417184529529287149
2174953178195293926546831
3546516509898554684982984

Visual Perception

- Early visual processing takes places without our conscious intervention
- Graphs that convey information at this level allow the observer to be more efficient in decoding



Takeaway Messages

- Limitations of human vision system
- Exploits message brodcast at early stage of perception: preattemptive perception
- Avoid possible causes of biases

VISUAL VARIABLES

SCIENCE

Vol. 103, No. 2684 Friday, June 7, 1946

On the Theory of Scales of Measurement

S. S. Stevens

Director, Psycho-Acoustic Laboratory, Harvard University

FOR SEVEN YEARS A COMMITTEE of the British Association for the Advancement of Science debated the problem of measurement. Appointed in 1932 to represent Section A (Mathematical and Physical Sciences) and Section J (Psychology), the committee was instructed to consider and report upon the possibility of "quantitative estimates of sensory events"—meaning simply: Is it possible to measure human sensation? Deliberation led only to disagreement, mainly about what is meant by the term measurement. An interim report in 1938 found one member complaining that his colleagues

by the formal (mathematical) properties of the scales. Furthermore—and this is of great concern to several of the sciences—the statistical manipulations that can legitimately be applied to empirical data depend upon the type of scale against which the data are ordered.

A Classification of Scales of Measurement

Paraphrasing N. R. Campbell (Final Report, p. 340), we may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules. The fact that numerals can be assigned under different rules leads

Data Types

- Nominal (N)
  - Equality relation
  - Apples, bananas, pears,…
- Ordinal (O)
  - Ordering relation
  - Small, medium, large, darker, dark, light,…
- Quantitative (Q)
  - Arithmetic relations
  - 10m, 32 degree, 2 bars,…
- Q-Interval (no reference point)
  - Dates, Location
  - Not directly comparable
  - Distances: A is 3 degree hotter than B
- Q-Ratio (reference point)
  - Length, mass
  - Proportions: A is twice as large as B

Data Types Operators

- Nominal
  - ≠, =
- Ordinal
  - ≠, =, >, <
- Quantitative Interval
  - ≠, =, >, <, +, -
- Quantitative Ratio
  - ≠, =, >, <, +, -, ×, ÷

From Data to Conceptual Model

- Data Model: low-level representation of data and operations
- Conceptual Model: mental and semantic construction

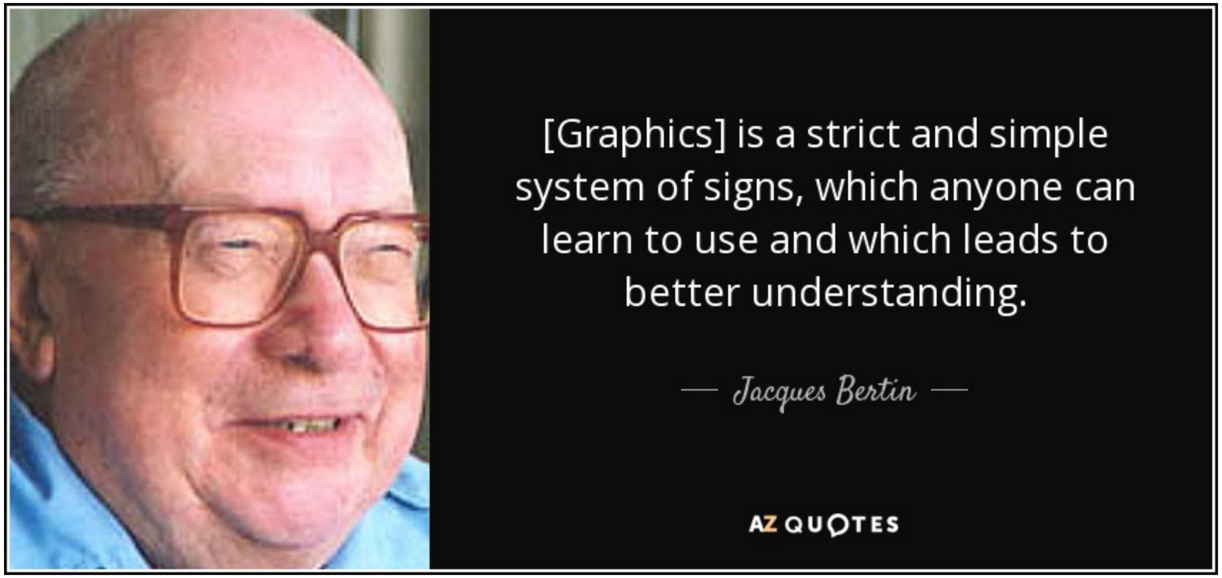
Data	Concept
1D number	Temperature
2D numbers	Geographic Coordinate
3D numbers	Spatio-temporal position

From Data to Conceptual Model

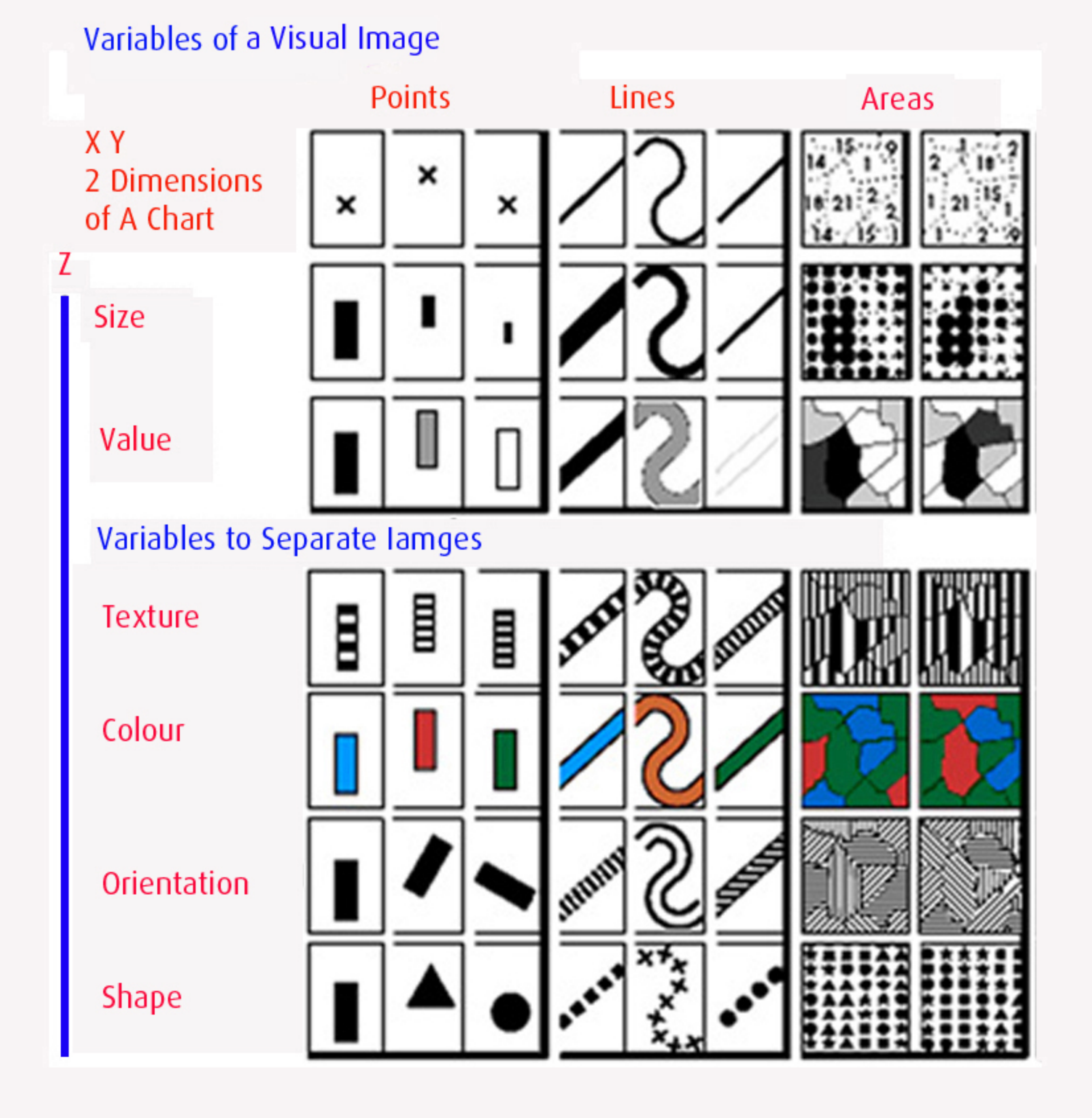
- From data model…
  - 70.8, 27.2, -10.2,…
- … using conceptual model …
  - Temperature
- … to data type
  - Continuous variation
  - Warm, hot, cold
  - Burned vs not burned

Visual Variables

- Jacques Bertin (1918-2010), cartographer
- Theoretical principles of visual encodings
- Semiology of Graphics (1967)

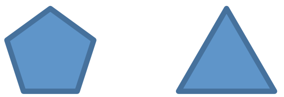


Bertin’s Visual Variables

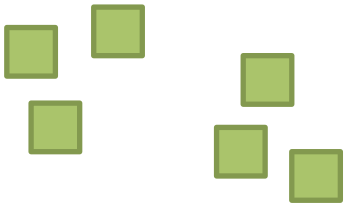


Characteristics of Visual Variables

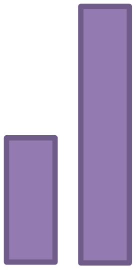
- Selective
  - May I distinguish a symbol from the others



- Associative
  - May I identify groups?

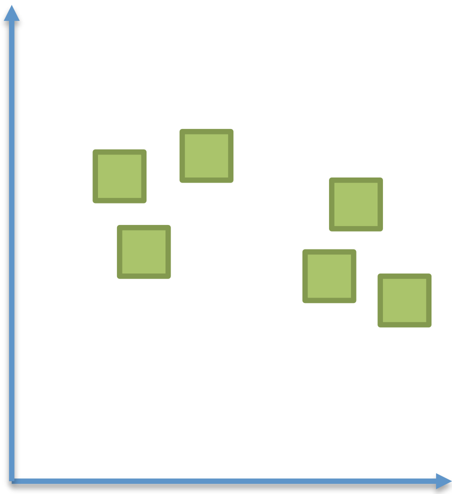


- Quantitative
  - May I quantify the difference of two values?
- Order
  - May I idenfiy an ordering?
- Length
  - How many values?



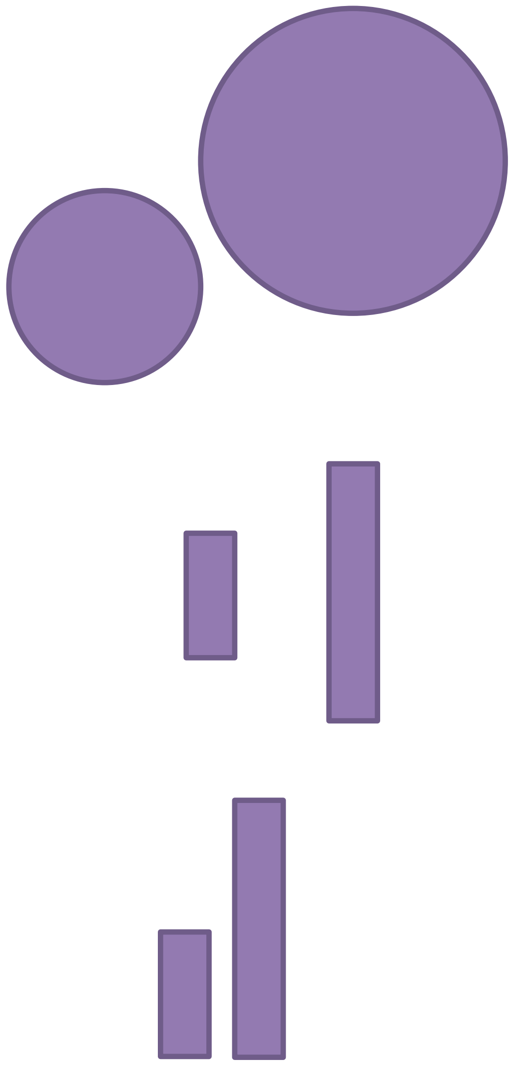
VV: Position

- Strongest visual variable
- Compatible for all data types
- Cons:
  - Not always applicable (e.g. nD data)
  - Cluttering

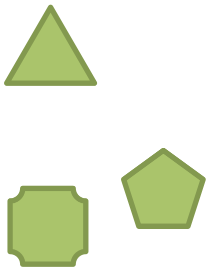


VV: Size and Length

- Easy to compare dimensions
- Grouping
- Estimate differences
  - Quantitative encoding
  - Changes in lengths
  - Worse for change in area



VV: Shapes



- Strong for nominal encoding
- No ordering
- No grouping

VV: Value (intensity)

- Quantitative representation (when size and length are used)
- Limited number of shades
- Support grouping

VV: Color (Tint)

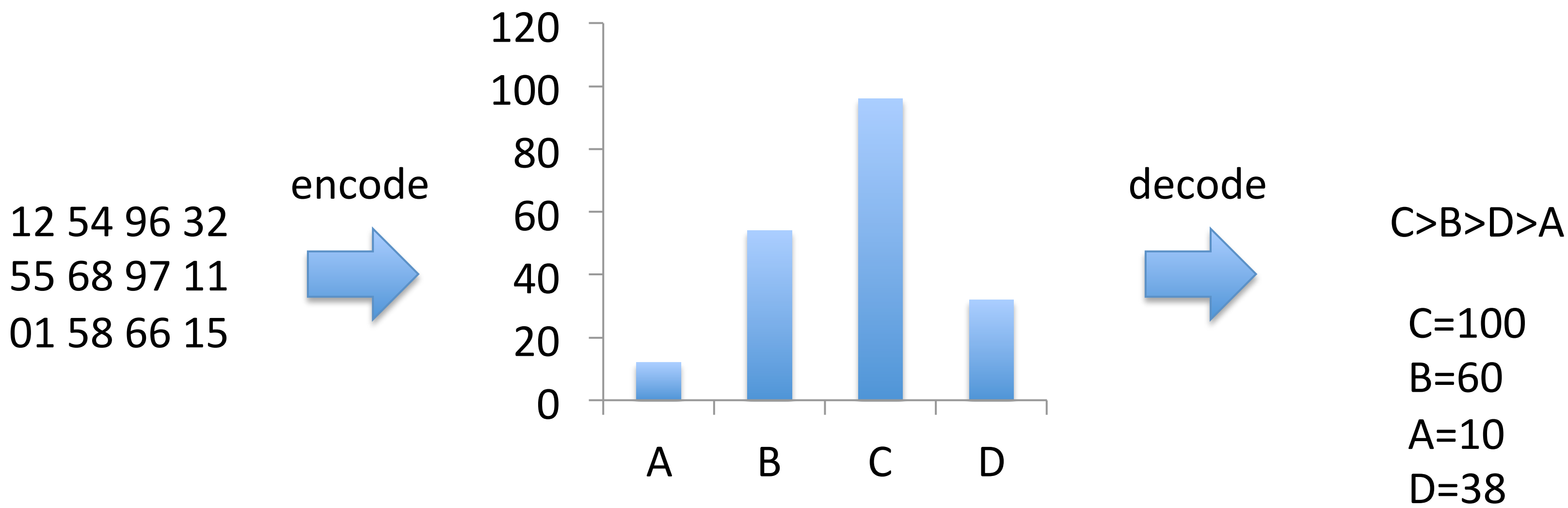
- Good for qualitative data
- Limited number of classes (!!!)
- Not good for quantitative data
- Be careful!!

Bertin Visual Variables

	Nominal	Ordinal	Quantitative
Position	✔	✔	✔
Size	✔	✔	~
Value (intensity)	✔	✔	~
Texture	✔	~	✗
Color	✔	✗	✗
Orientation	✔	✗	✗
Shape	✔	✗	✗

Visual Encoding/Decoding

- A graph encode a set of information as a set of graphical attributes
- The observer have to decode the graphical attributes to extract the original information



TAXONOMY OF VISUAL VARIABLES

Cleveland McGill [1984]

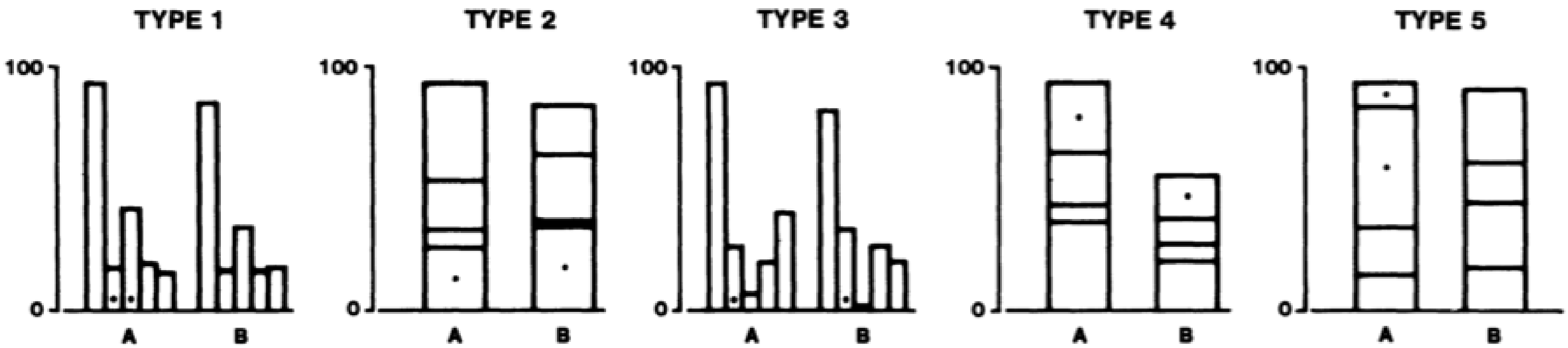


Figure 4. Graphs from position–length experiment.

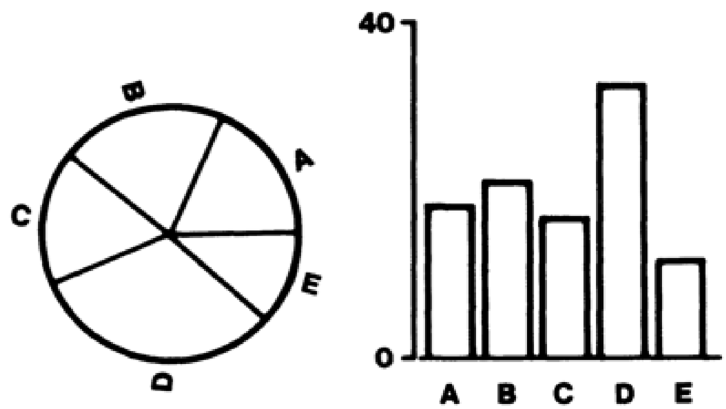
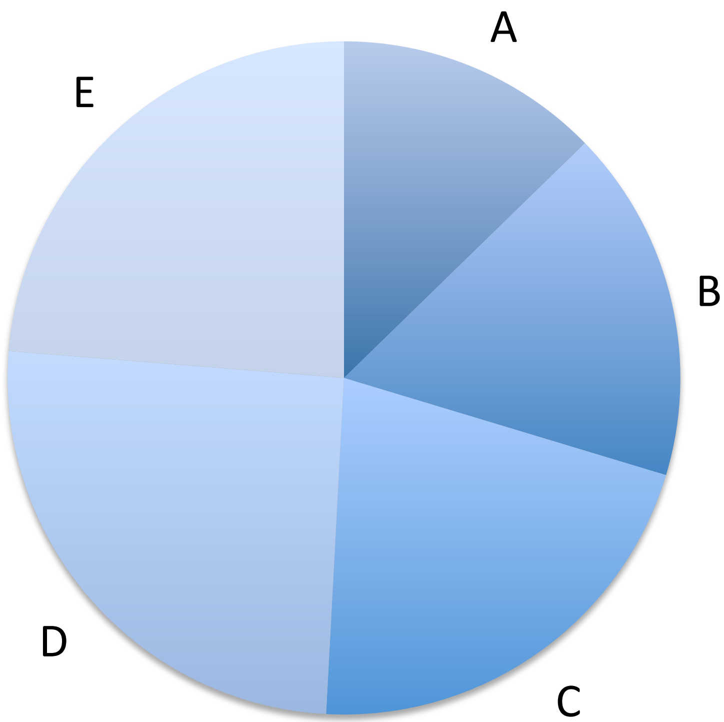


Figure 3. Graphs from position–angle experiment.

Cleveland & McGill: graphical encodings

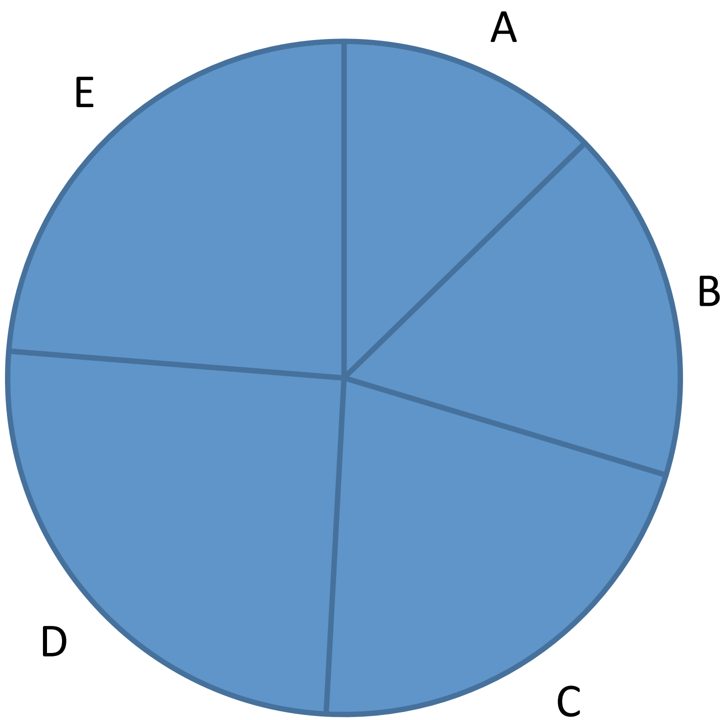
- Angle
- Area
- Color Hue
- Color Saturation
- Density
- Length
- Position on a common scale
- Position on non aligned scale
- Slope
- Volume

Angle decoding



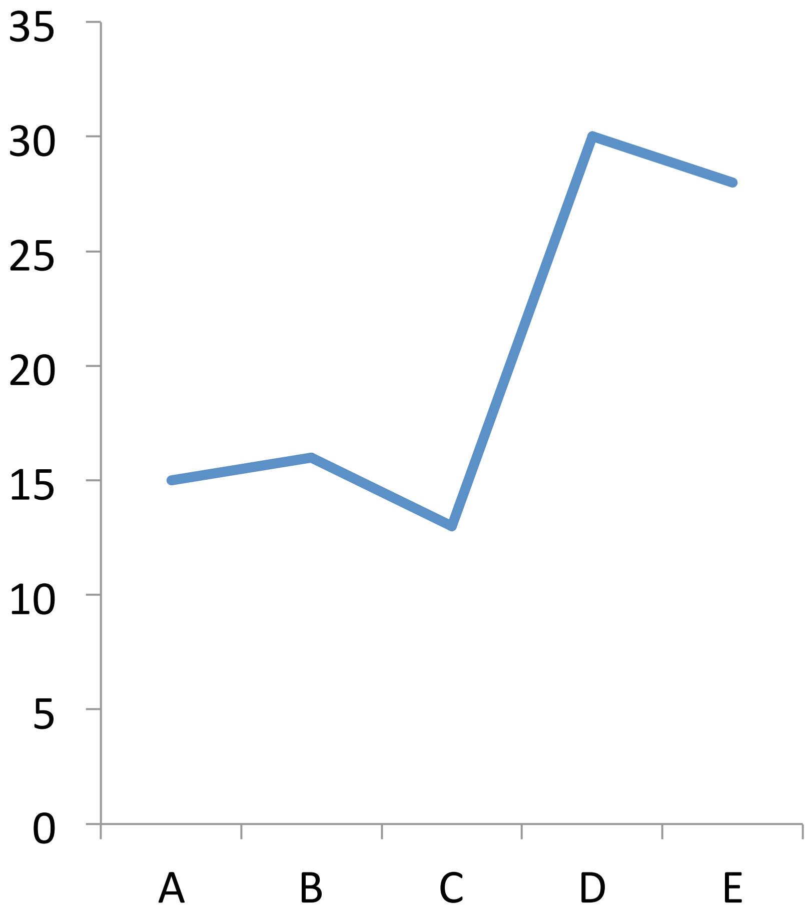
- It is difficult to compare angles
  - Underestimation of acute angles
  - Overestimation of obtuse angles
  - Easier if bisectors are aligned
- Area estimation helps

Angle decoding



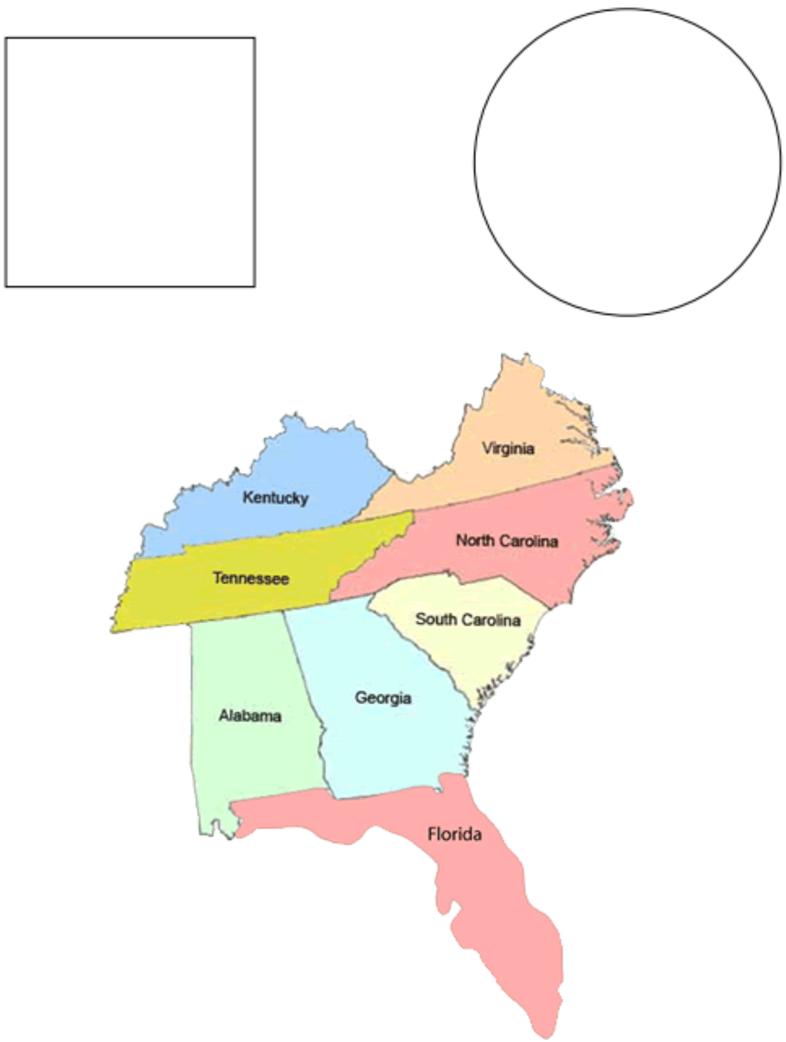
- It is difficult to compare angles
  - Underestimation of acute angles
  - Overestimation of obtuse angles
  - Easier if bisectors are aligned

Slopes Decoding



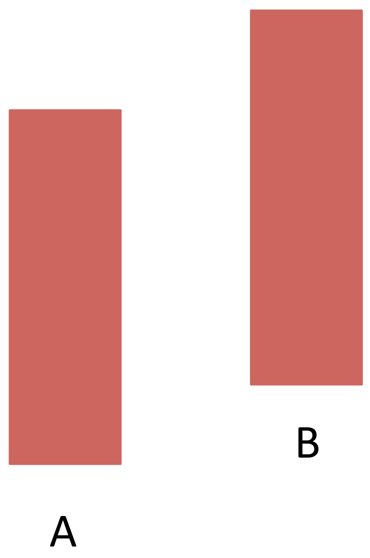
- Same difficulties as angles
- Easier task since one branch is aligned with x-axis

Area Decoding



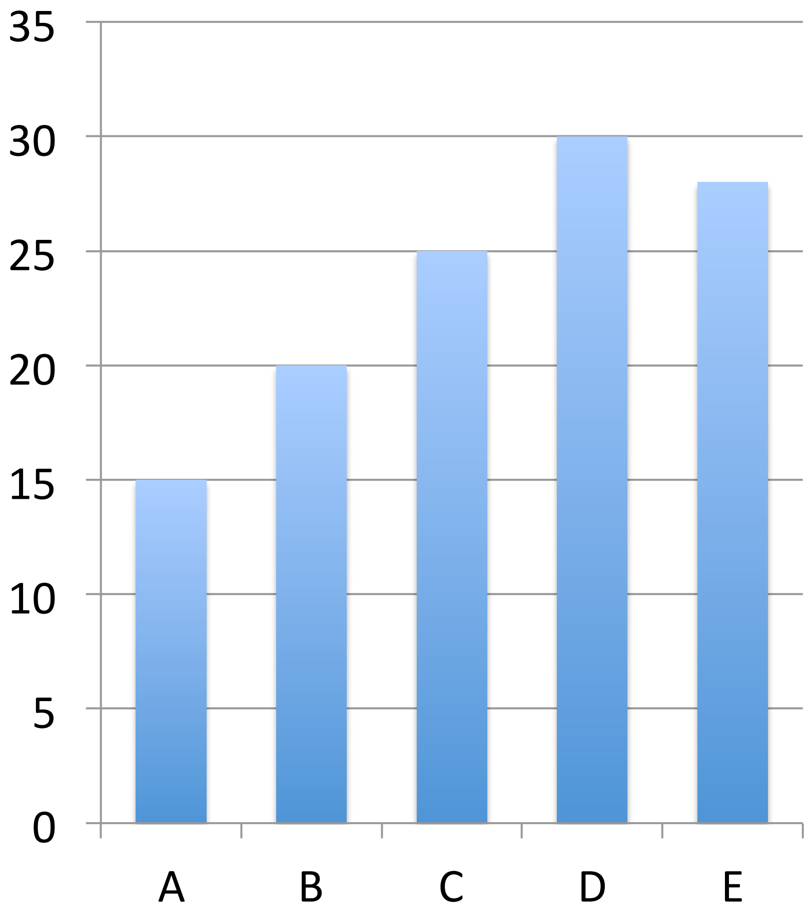
- Area is not well decoded
  - Different regular shapes
  - Irregular shapes
  - Context influences (thin area within compact thick area)

Length Decoding



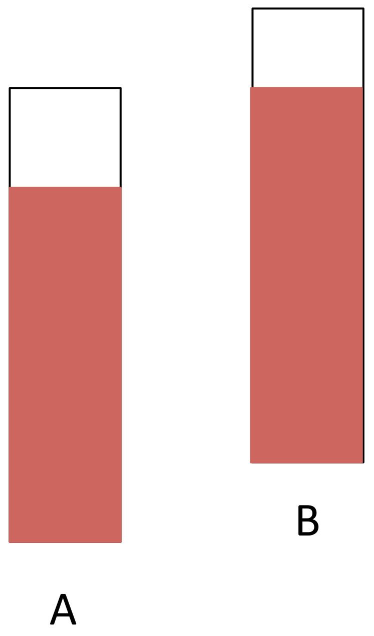
- Straight forward to endoce numerical values
- Difficulties with relative lengths

Position on a common scale



- Widely used in statistical charts

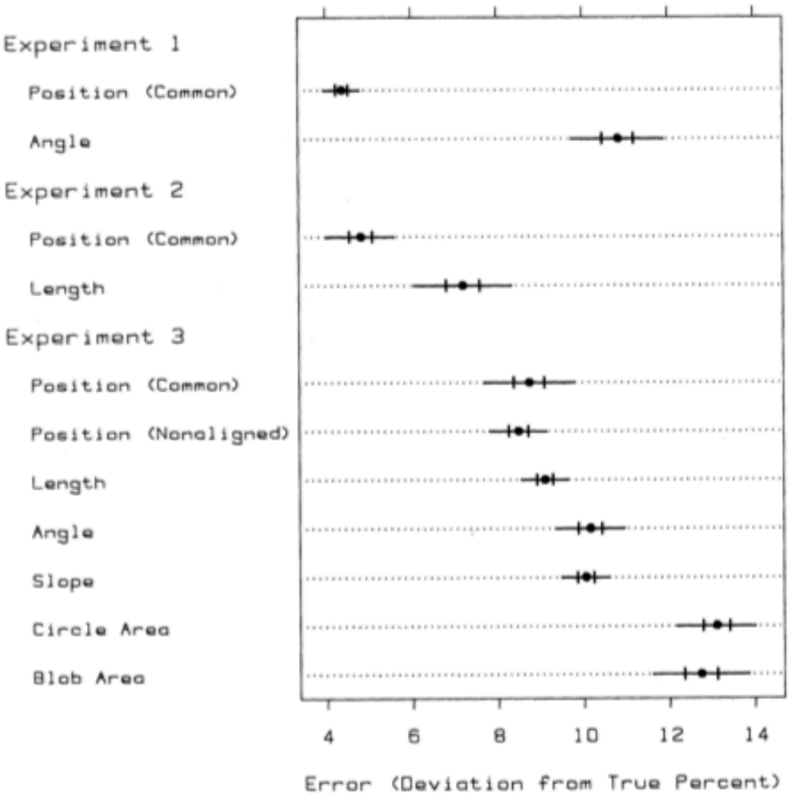
Position on non-aligned scale

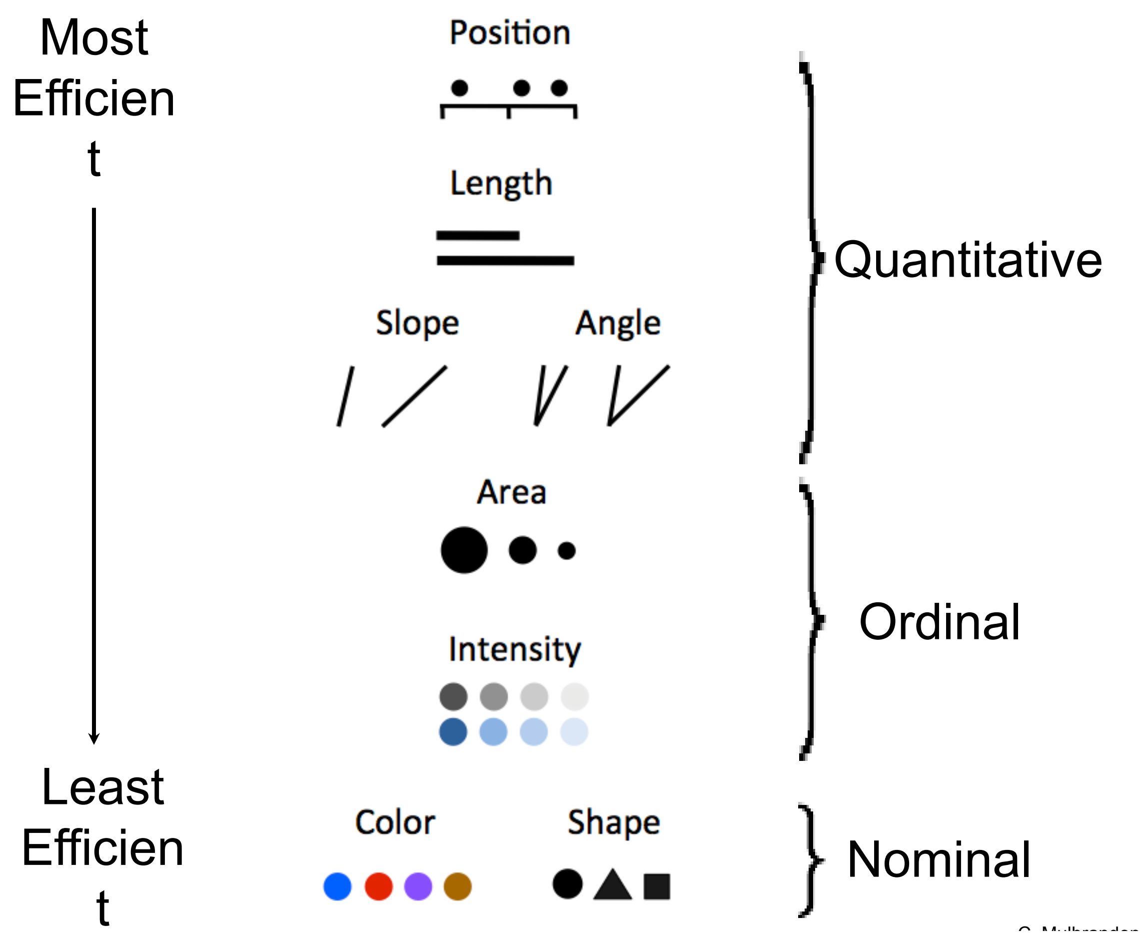


- Not as bas as common scale
- Still acceptable

Designing Effective Visualizations

- If possible, use graphical encoding that are easily decoded
- Graphical Attributes ordered(Cleveland & McGill):
  - Position along a common scale
  - Position on non aligned scales
  - Length
  - Angle and Slope
  - Area
  - Volume, density, color saturation
  - Color Hue




Most
Efficien
t
Position
Length
Slope
Angle
Area
Intensity
Least
Efficien
t
Color
Shape
Quantitative
Ordinal
Nominal

PERCEPTION LAWS

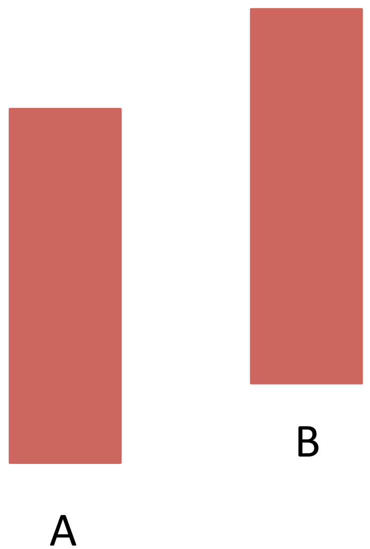
Weber's Law

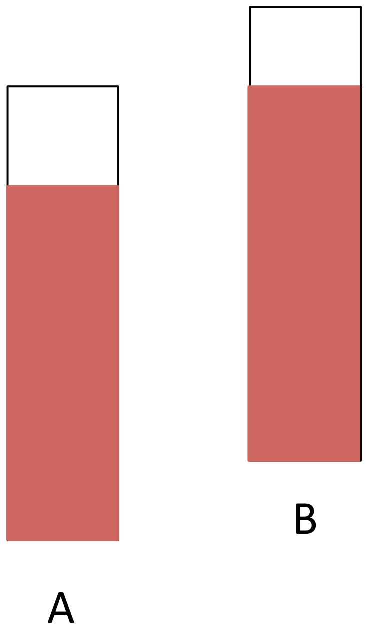
- Just-noticeable difference between two stimuli is proportional to their magnitudes
- Case study on length
  - Given two lines with lengths x and $x+w$
  - If w is small, it is difficult to notice difference between the two lines
  - If w is larger, it is easier to catch the difference
- How large should w be?
  - The probability of detecting the change is proportional to the reltaive value w/x

Weber's Law

- Given values (90, 92)
- Detect with probability of 2/90

- Given values(90,92)
- Detect with probability of 2/10



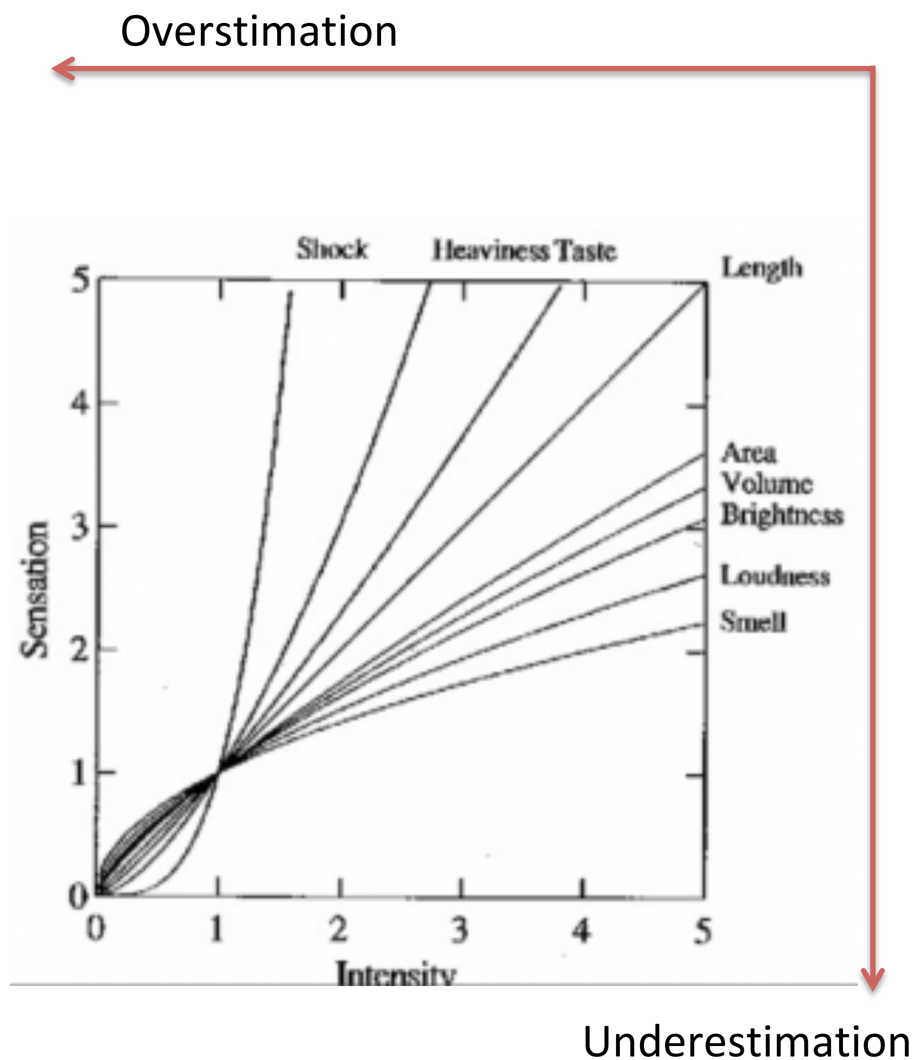


Stevens' Law

- Model the relation between a stimulus and its perceived intensity
- Given a stimulus x encoded with a visual attribute
- An observer decode a perceived value $p(x)$
- Stevens' law states that
  - $p(x) = kx^{\beta}$
  - where k is constant and
  - β is a constant that depends on the nature of stimulus

Stevens' law

- Better effectiveness when $p(x) = kx^{\beta}$ is linear
- Linearity depends only on β
- Different visual encodings yields typical ranges for β
  - Lengths: 0.9 – 1.1
  - Area: 0.6 – 0.9
  - Volume: 0.5 – 0.8



Weber and Stevens' Laws

- Given two values x_1 and x_2
- Let the perceived values be $p(x_1)$ and $p(x_2)$

$$\frac{p(x_1)}{p(x_2)} = \left(\frac{x_1}{x_2}\right)^{\beta}$$

Weber and Stevens' Laws: areas

- For areas β=0.7
- Let x_1=2 and x_2=1
- The perceived difference will be

$$\frac{p(2)}{p(1)} = \left(\frac{2}{1}\right)^{0.7} = 1,6245$$

- For areas β=0.7
- Let x_1=0,5 and x_2=1
- The perceived difference will be

$$\frac{p(\frac{1}{2})}{p(1)} = \left(\frac{1/2}{1}\right)^{0.7} = 0,6155$$

Weber and Stevens' Laws: areas vs lengths

- For areas β=0.7
- Let $x_2 = x_1 + w$
- The perceived difference will be

$$\left(\frac{x+w}{x}\right)^{0.7} \approx 1+\frac{0.7w}{x}$$

- For lengths β=1
- Let $x_2 = x_1 + w$
- The perceived difference will be

$$\left(\frac{x+w}{x}\right)^{1} = 1+\frac{w}{x}$$

Takeaway messages

- Data type for entities and relationships
- Visual variables for representation
- Mapping of types to VVs
- Some VVs are more appropriate for specific data types