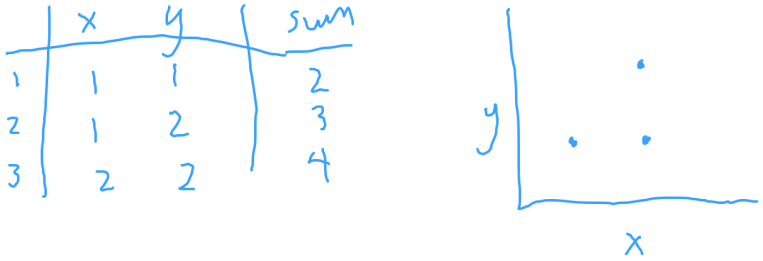
Learning objectives: introduction to data visualization theory
- establish a common vocabulary for discuss data visualizations
-> communicate with each other about properties of data visualizations
- begin programming in R

Survey: What is data visualization?
- translating data into something visual to portray information - Michael



- communicating data in summarized manner -> give a point and take away

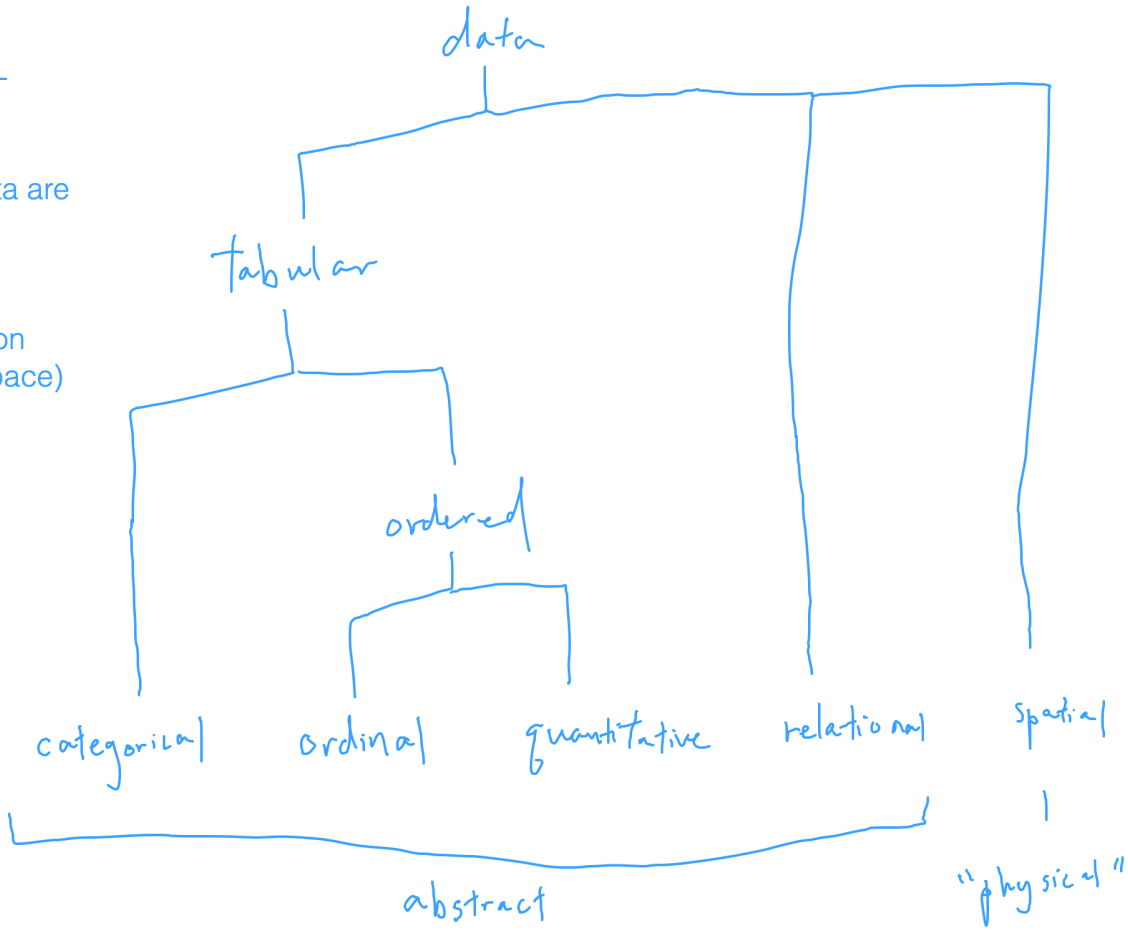
Motivation for data visualization?
- Let someone who doesn't know the data understand something about the data - Yuqi
-> helping people perceive information in the data better/faster

1. exploratory data visualizations -> assist in identifying patterns in the data
2. explanatory data visualizations -> assist in communicate about the data

Data types

relational data - represent how numeric and categorical data are related

spatial data - physical location (and shape, space)



categorical data - representing qualitative characteristics; can take on numeric value but these values don't have any meaning
- ex. tall, short -> ordered?
- ex. round, square, triangular
- ex. credit, auditing
- ex. cat, dog, fish
- ex. neuron vs. glia (cell-types)

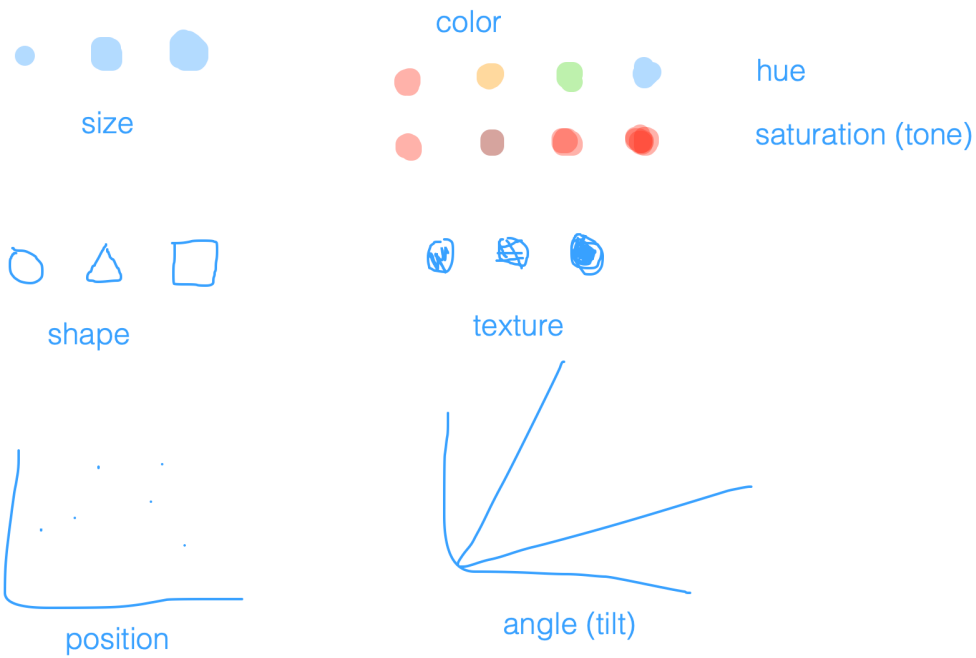
ordinal data - data fall into categories and can be numeric but in numbers have meanings
- ex. short, tall
- ex. freshman, sophomore, junior, senior
- ex. mild, severe, hospitalized, dead
- ex. gene being lowly expressed, highly expressed

quantitative data - represents numeric values; these values can be interpreted mathematically
- O2 stats: 100% to 70%
- ex. genes numeric expression magnitudes

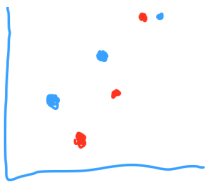
Types of encodings for different data types

Def: ~~graphical primitives~~ geometric primitives

Def: visual properties (control appearance of geometric primitives)

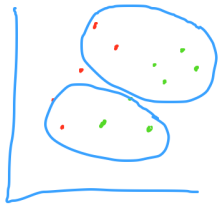


data visualizations as combinations of geometric primitives and visual properties



Gestalt principles
- principles of visual design (not just for data visualization)
-> logo designers
- we will focus on a few of them that are applicable to data visualization (in general there are > 10 principles)
->based on perception and cognitive psychological research

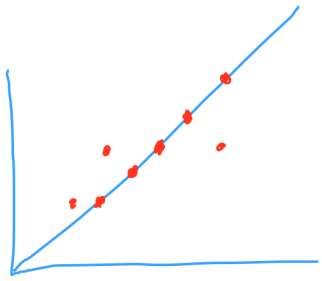
1. Similarity - items alike in terms of geometric primitives or visual channel tend to be perceived as being in a related group



2. proximity - items near to each other tend to be perceived as being in a related group

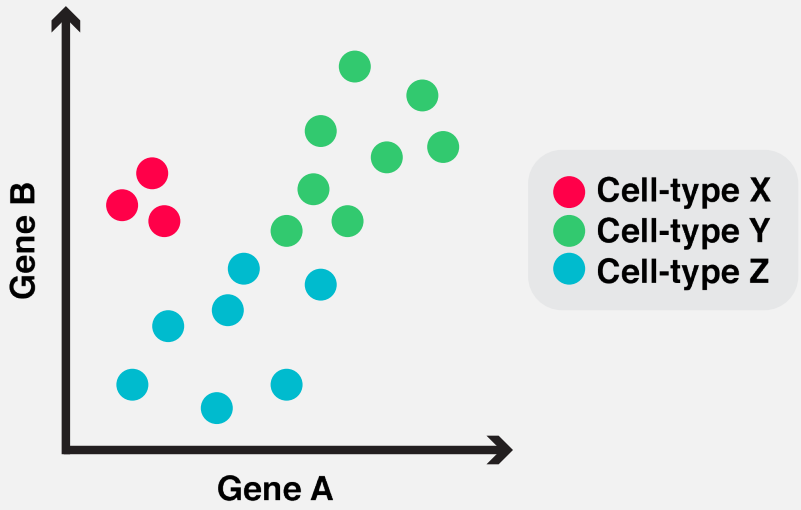
3. enclosure - items surrounded by something like a line or an object tend to be perceived as a group

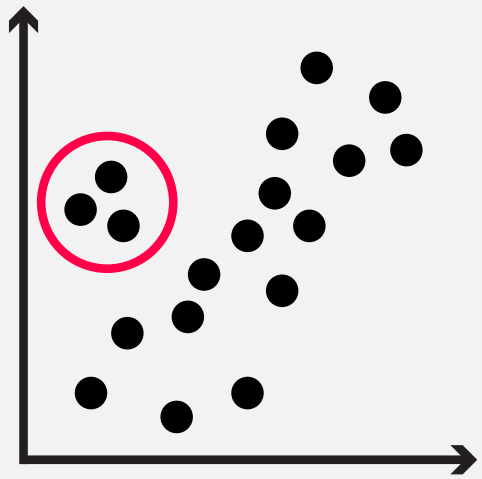
4. continuity - items tend to be perceived as part of a smooth group or continuous lines rather than sharp broken lines

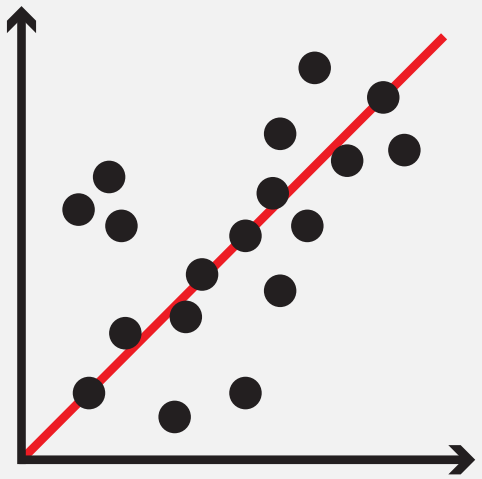


Salience
- defined: how the viewer's eye is drawn to what is important in the data visualization
-> data visualization (good one) is meant to enhance saliency
    -> improve viewer's understanding of the data
        -> from your visualization, different groups can be readily distinguished
        -> different groups can be separated, processed preattentively
        -> encode more important information more effective

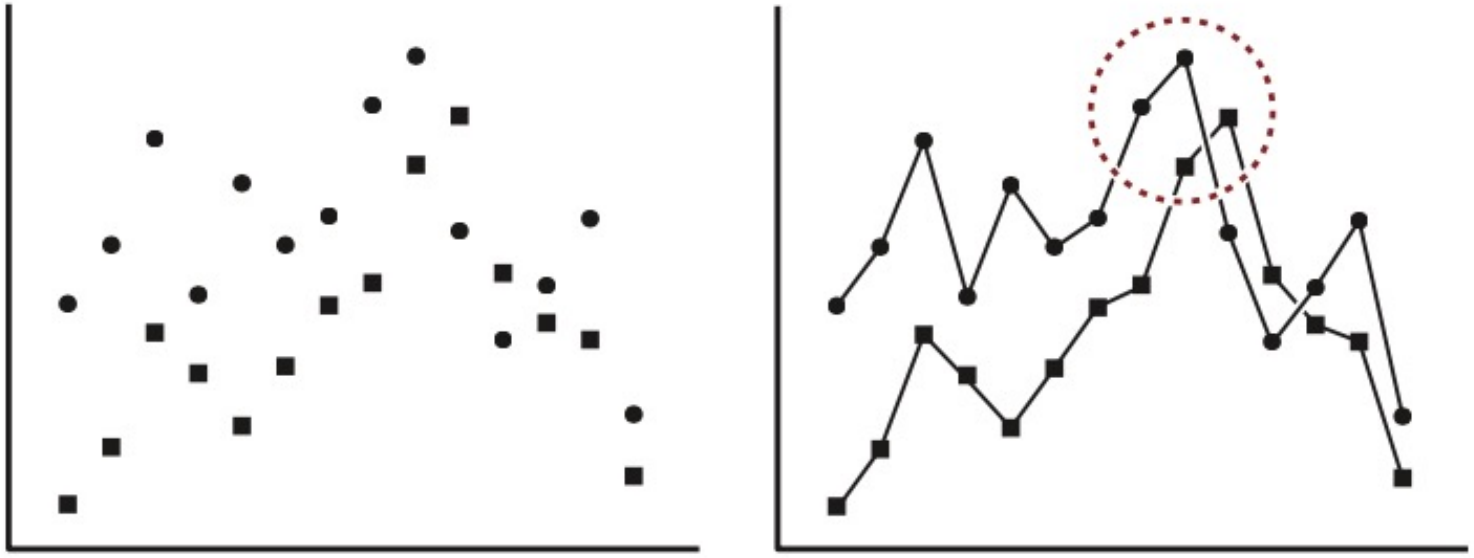
WHAT DIFFERENT STORIES DO EACH OF THESE DIFFERENT DATA VISUALIZATIONS OF THE SAME DATA TELL?





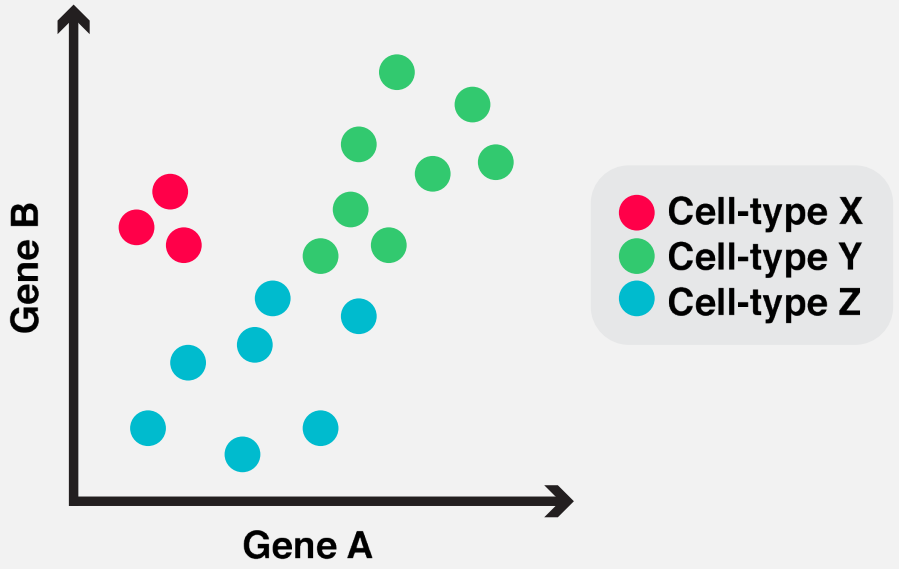


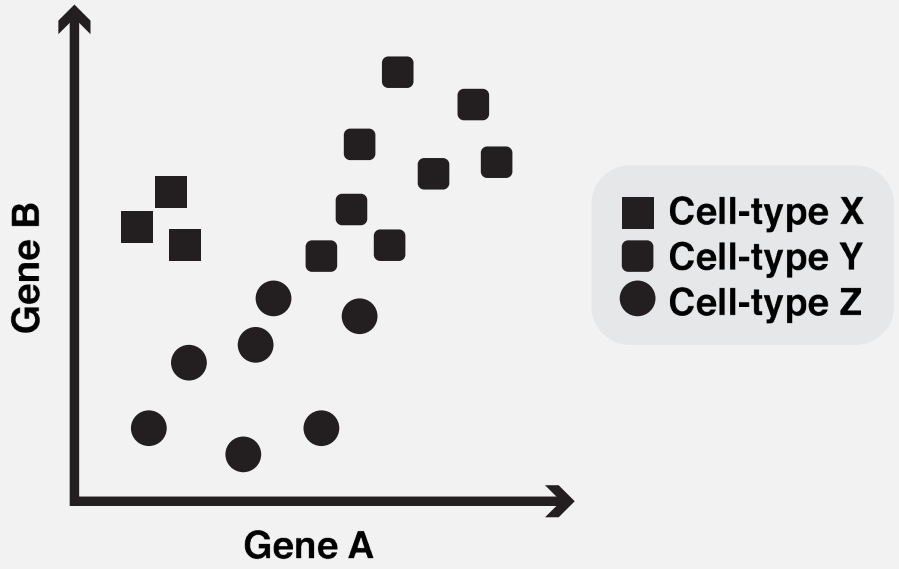
WHAT'S YOUR INTERPRETATION OF THIS DATA VISUALIZATION STORY?



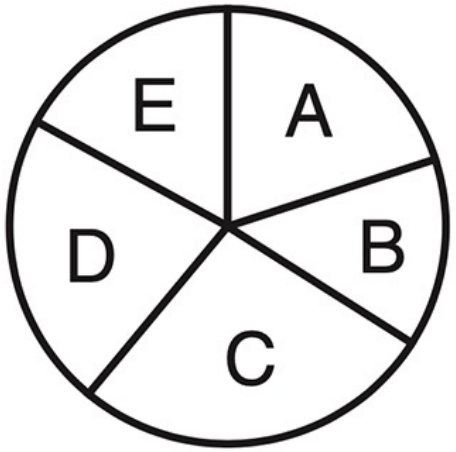
WHAT GESTALT PRINCIPALS ARE BEING APPLIED?

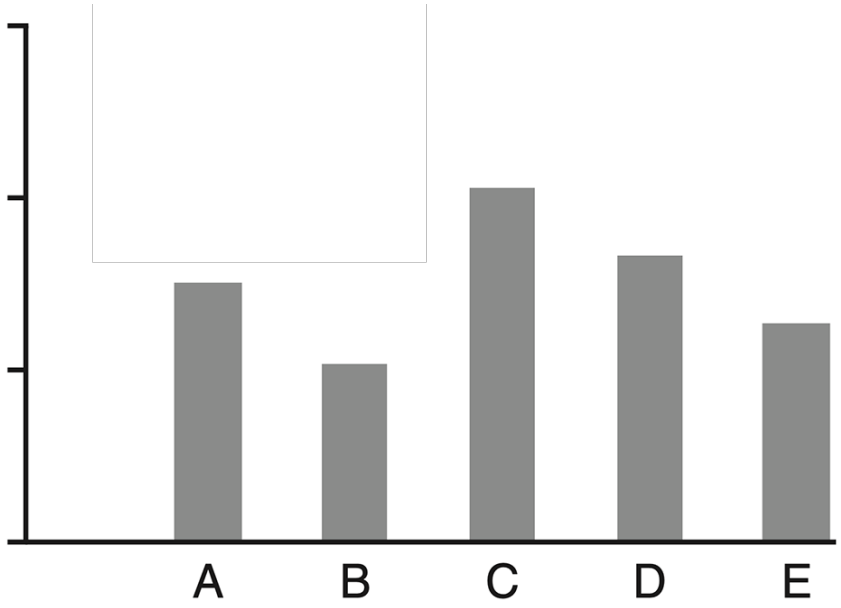
DISCUSSION: WHICH DATA VISUALIZATION DO YOU THINK ENHANCES SALIENCY?



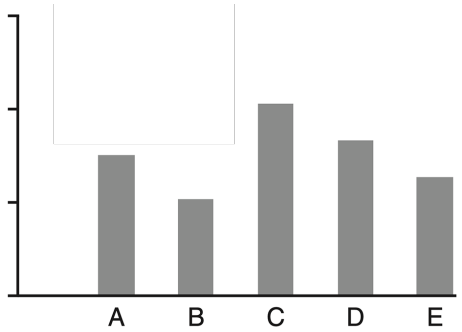


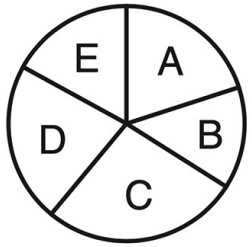
WHICH LETTER IS MORE COMMON?





(Credit: Bang Wong, Nature Methods, 2010)





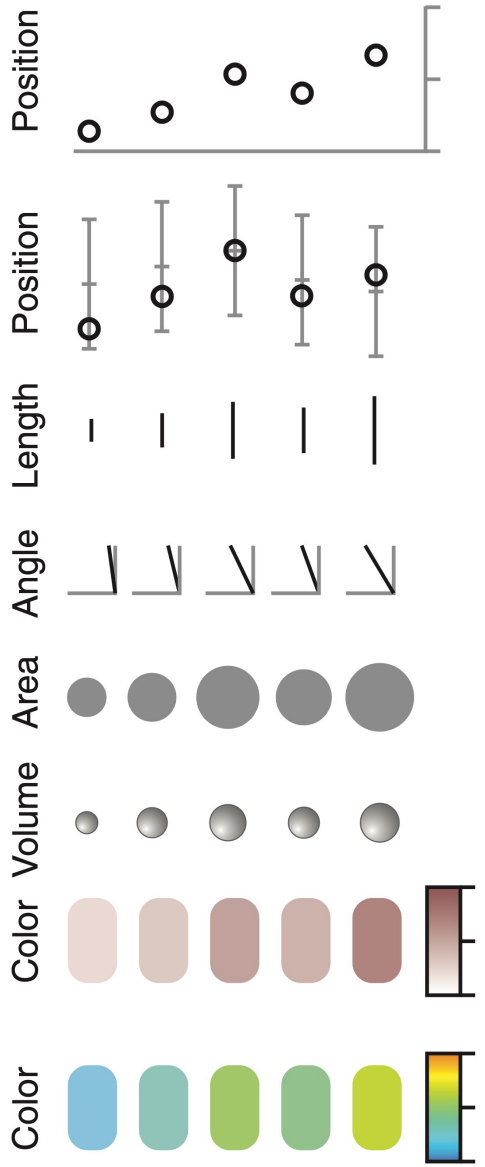
Encoding

Best

Worst

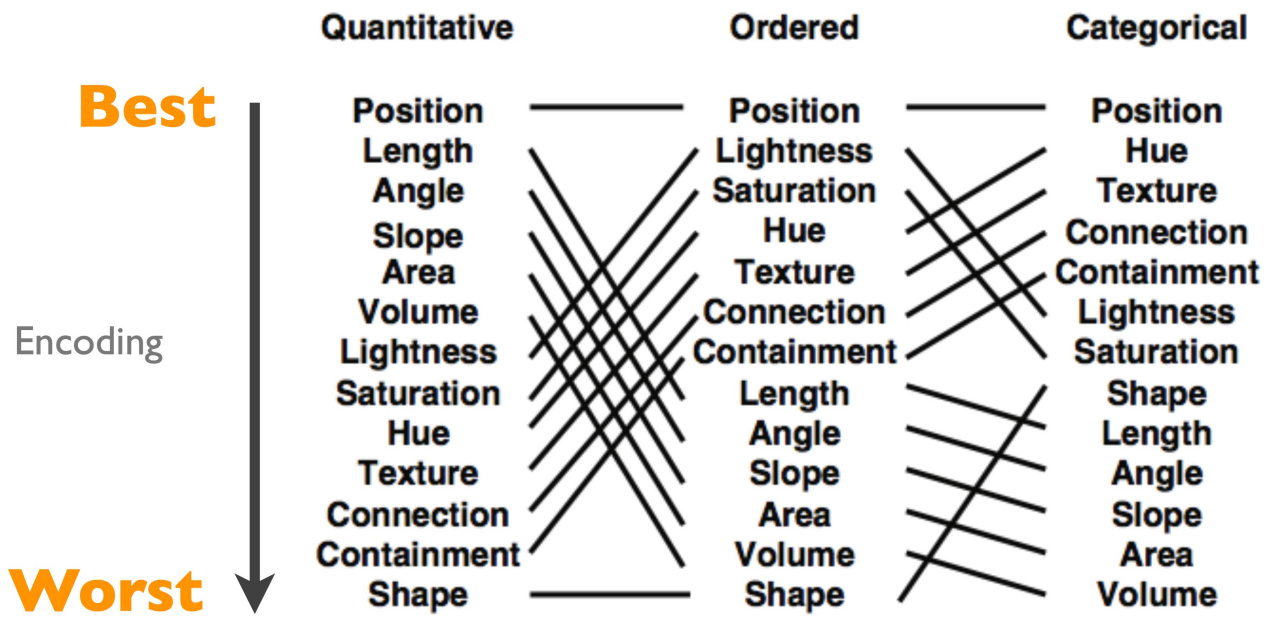
Quantitative

Position
Length
Angle
Slope
Area
Volume
Lightness
Saturation
Hue
Texture
Connection
Containment
Shape



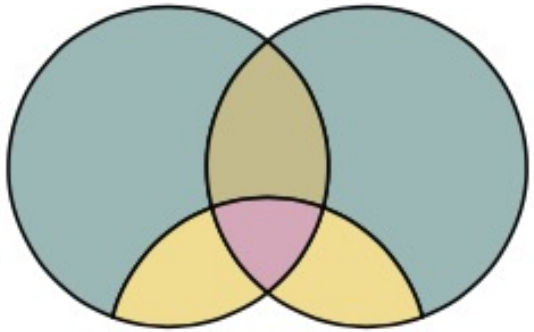
(Credit: Bang Wong, Nature Methods, 2010)

RANKING OF ENCODINGS FOR DIFFERENT DATA TYPES



(Credit: Nils Gehlenborg, ISMB/ECCB 2011)

DISCUSSION: WHICH DATA VISUALIZATION DO YOU PREFER? AND WHY?



(Credit: Martin Krzywinski, Nature Methods 2013)