# VISUALIZATION

## Visual Encoding

# Data

Item: individual entity, discrete

e.g., Patient, Car, Stock, City

Attribute: measured, observed, logged property

e.g., Patient: height, blood pressure; Car: horsepower, make

Item: Person  Attributes

| ID | Name | Age | Shirt Size | Favorite Fruit |
|---|---|---|---|---|
| 1 | Amy | 8 | S | Apple |
| 2 | Basil | 7 | S | Pear |
| 3 | Clara | 9 | M | Durian |
| 4 | Desmond | 13 | L | Elderberry |
| 5 | Ernest | 12 | L | Peach |
| 6 | Fanny | 10 | S | Lychee |
| 7 | George | 9 | M | Orange |
| 8 | Hector | 8 | L | Loquat |
| 9 | Ida | 10 | M | Pear |
| 10 | Amy | 12 | M | Orange |

Cell

- Links/Relations
  - Express relationship between two items
  - Friendship on Facebook, Interaction between proteins

# Visual Language is a Sign System

- Image perceived as a set of signs
- Sender encodes information in signs
- Receiver decodes information from signs

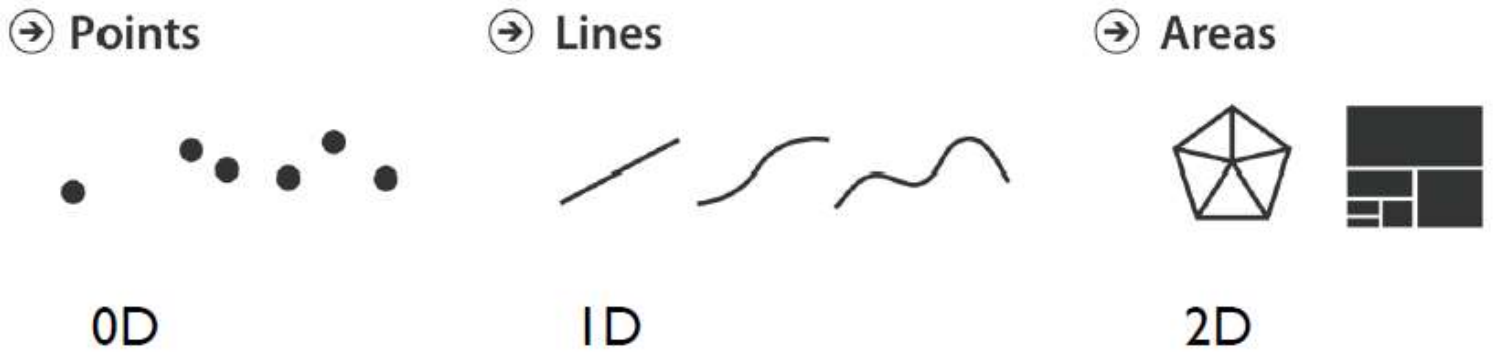
Semiology of Graphics [J. Bertin, 83]

# Marks & Channels

- **Marks**: represent **items** or **links**
- **Channels**: change **appearance** based on **attribute**
- **Channel = Visual Variable**

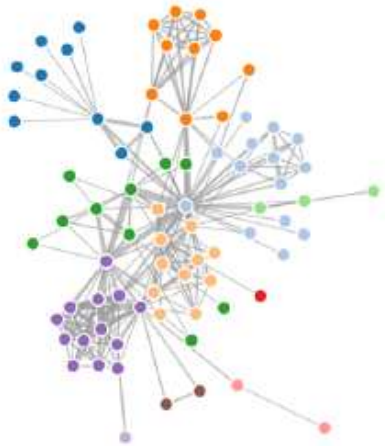
# Marks for Items

Basic geometric elements



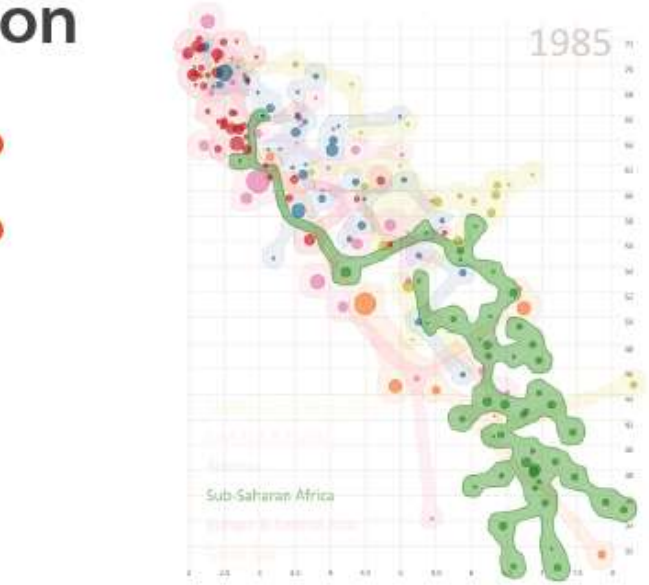

3D mark: Volume, but rarely used

# Marks for Links

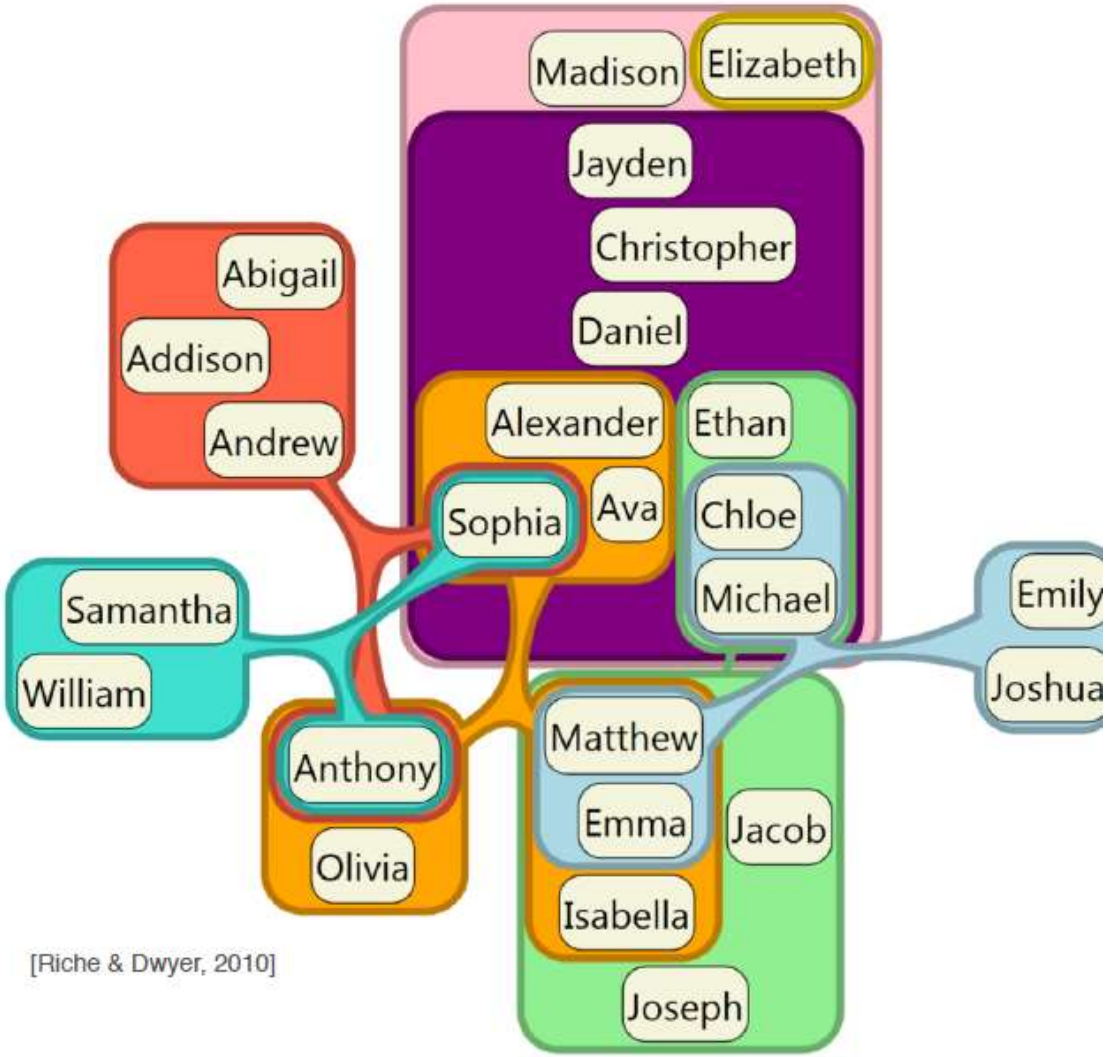


- Containment

- Connection



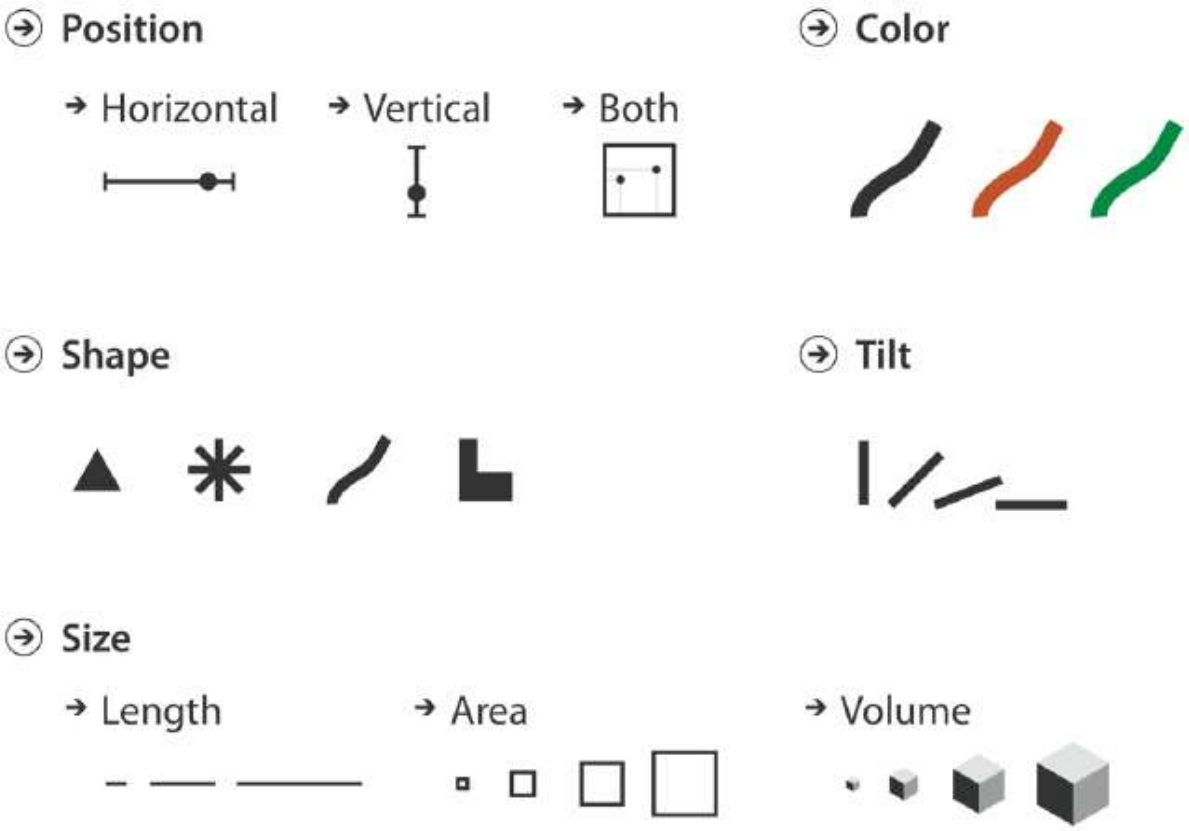
# Containment can be nested




[Riche & Dwyer, 2010]

# Channels (aka Visual Variables)

Control appearance proportional to or based on attributes

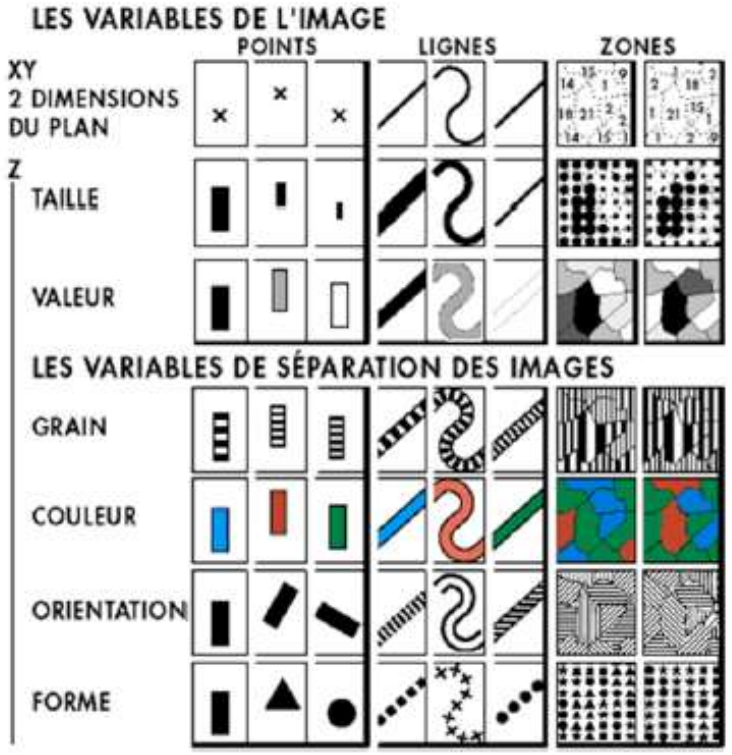


# Bertin's Visual Variables

Marks: Points Lines Areas

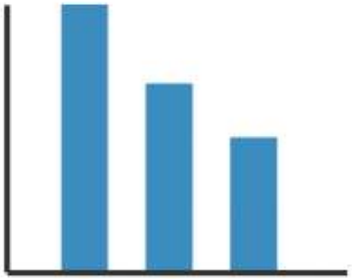
Position
Size
(Grey)Value

Texture
Color
Orientation
Shape




Semiology of Graphics [J. Bertin, 67]

# Using Marks and Channels



Mark: Line

Channel: Length/Position

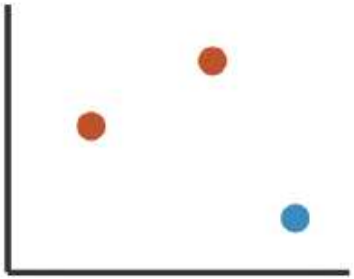
1 quantitative attribute

1 categorical attribute

Mark: Point

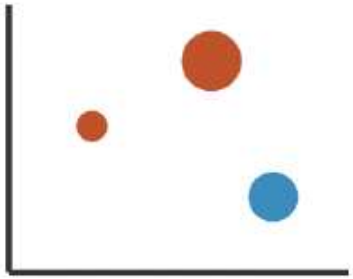
Channel: Position

2 quantitative attr.



Adding Hue

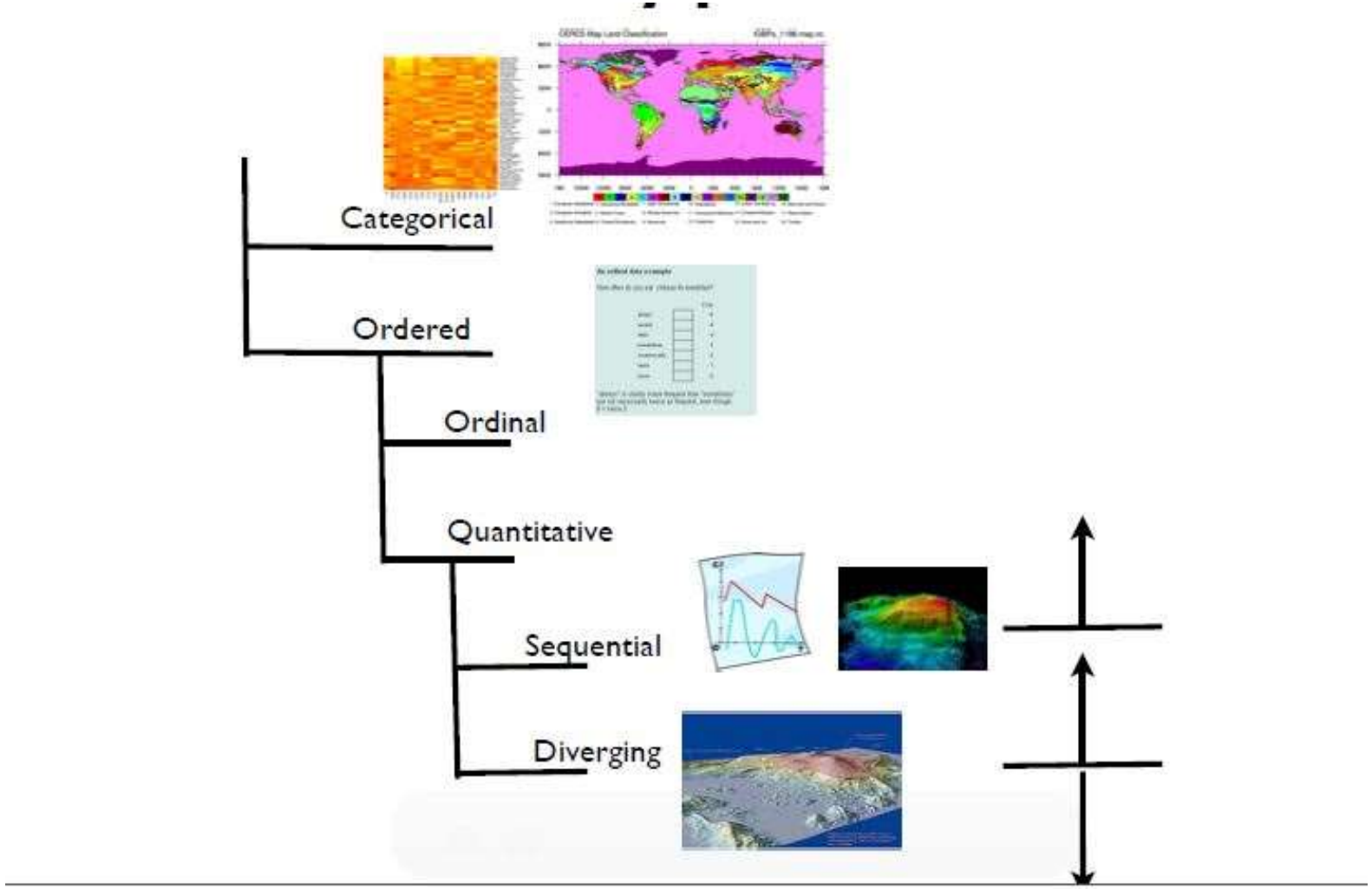
+1 categorical attr.



Adding Size

+1 quantitative attr.

# Attribute Types (Summary)



# Stolte / Hanrahan

| property | marks | ordinal/nominal mapping | quantitative mapping |
|---|---|---|---|
| shape | glyph | ○ □ + △ S U | |
| size | rectangle, circle, glyph, text | | |
| orientation | rectangle, line, text | | |
| color | rectangle, circle, line, glyph, y-bar, x-bar, text, gantt bar | | min … max |

"Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases"
Chris Stolte and Pat Hanrahan

# Types of Channels

**Magnitude Channels**

How much?

Position

Length

Saturation …

**Identity Channels**

What? Where?

Shape

Color (hue)

Spatial region …

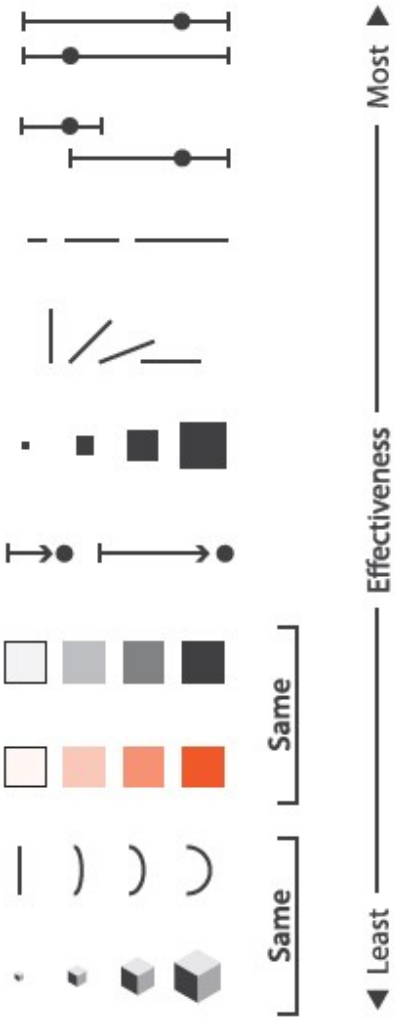
**Ordinal & Quantitative Data**

**Categorical Data**

**Channels:** Expressiveness Types and Effectiveness Ranks

## 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)




## Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

# Characteristics of Channels

Selective

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

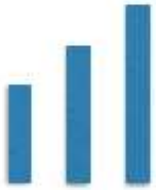
Associative

- Does it support grouping?

Quantitative (Magnitude vs Identity Channels)

- Can we quantify the difference between two marks?

# Characteristics of Channels



Order (Magnitude vs Identity)

- Can we see a change in order?

Length

- How many unique marks can we make?

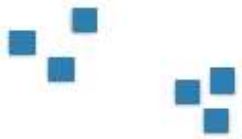
# Position

Strongest visual variable

Suitable for all data types

Problems:

- Sometimes not available (spatial data)
- Cluttering



Selective: yes

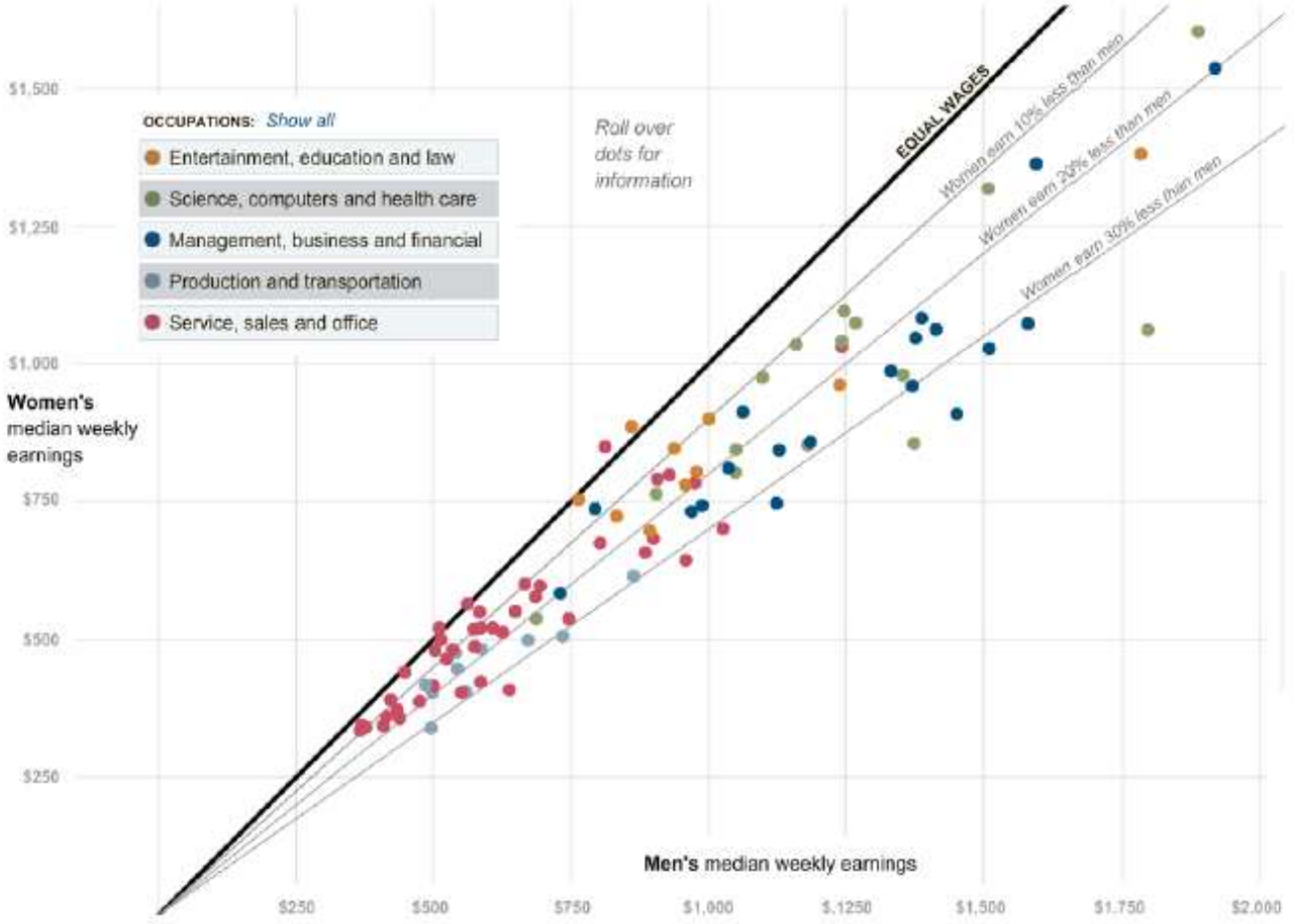
Associative: yes

Quantitative: yes

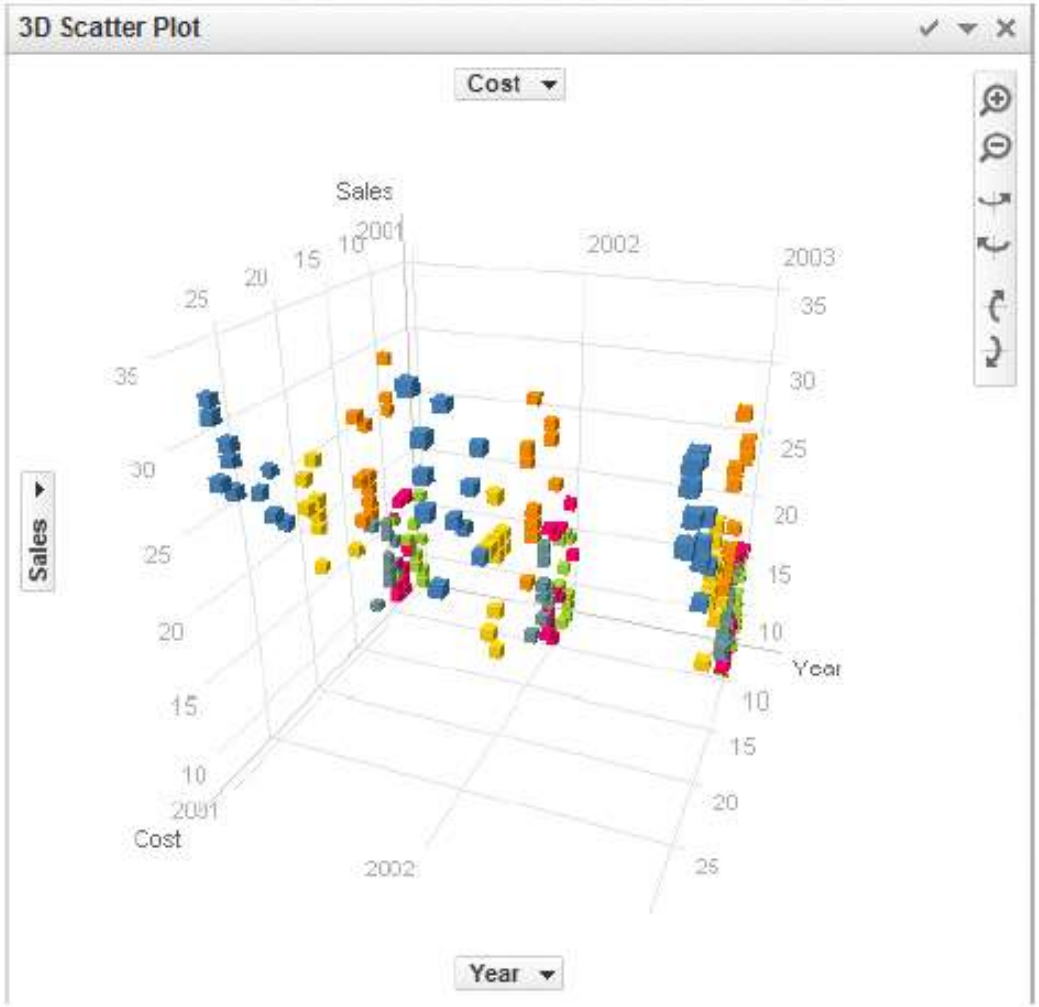
Order: yes

Length: fairly big

# Example: Scatterplot



# Position in 3D?



# Length & Size

Good for 1D, OK for 2D, Bad for 3D

Easy to see whether one is bigger

Aligned bars use position redundantly

For 1D length:

Selective: yes

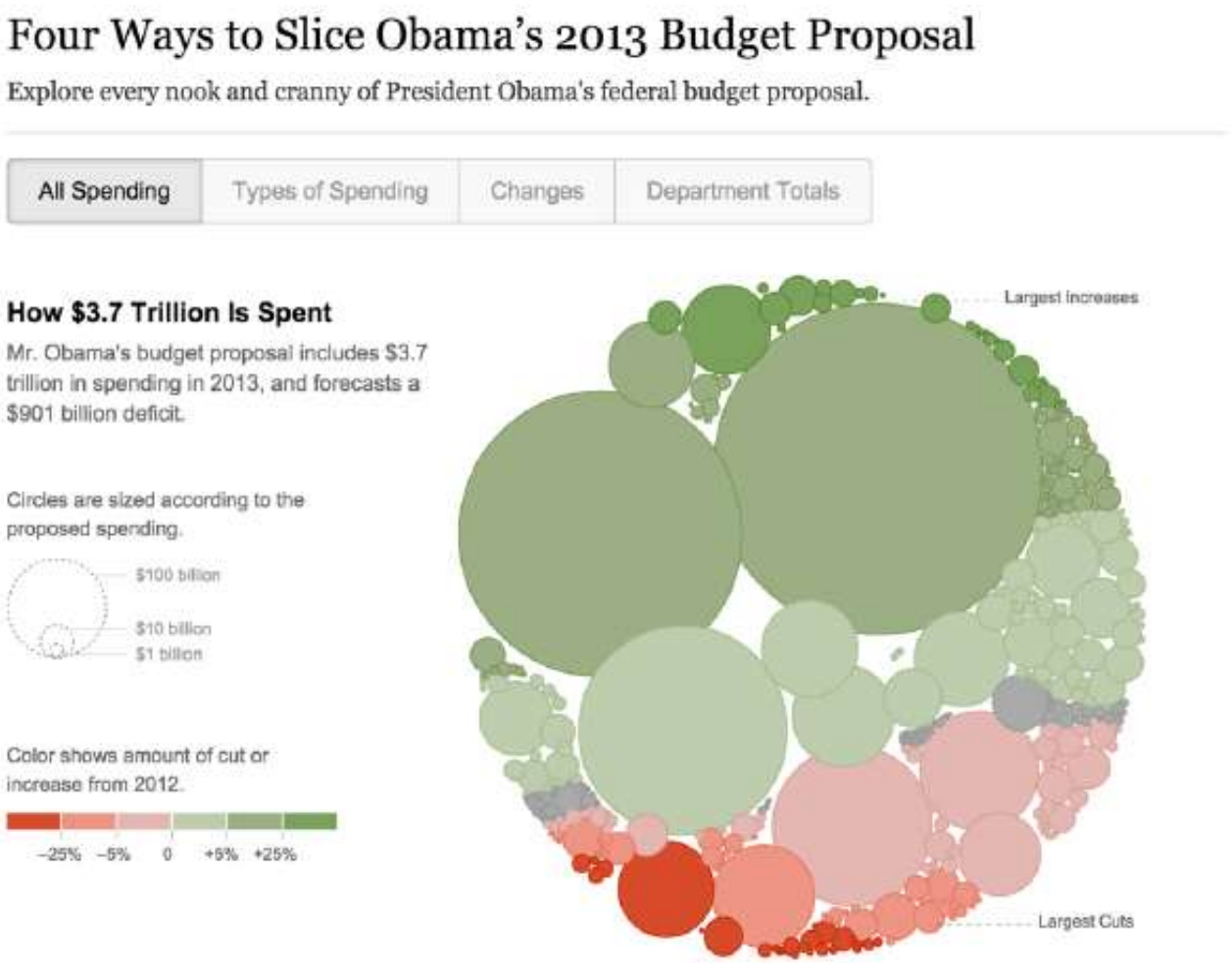
Associative: yes

Quantitative: yes

Order: yes

Length: high

# Example 2D Size: Bubbles



# Information in Color

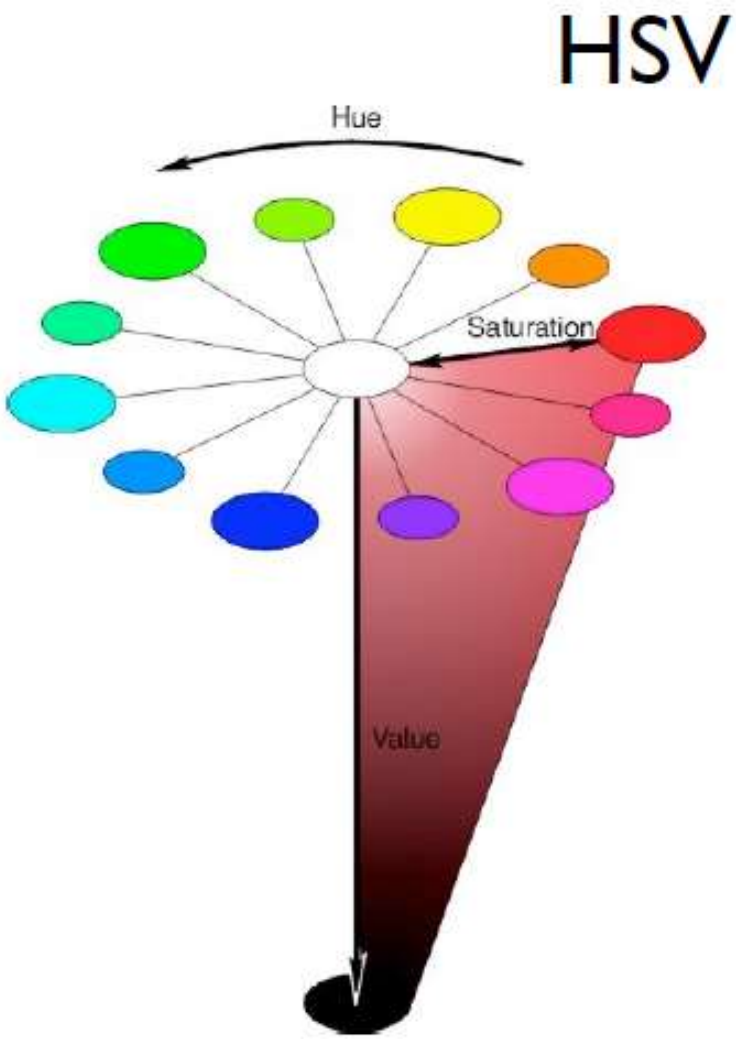
- (Grey)Value is perceived as ordered (O)

- Can encode quantitative values (Q) [not as well]

- Hue is normally perceived as unordered (N)

# Color HSV

- Based on polar coordinates, not Cartesian coordinates.
- HSV is a non-linearly transformed (skewed) version of RGB cube
  - Hue: quantity that distinguishes color family, say red from yellow, green from blue
  - Saturation (Chroma): color intensity (strong to weak). Intensity of distinctive hue, or degree of colour sensation from that of white or grey
  - Value (luminance): light color or dark color


HSV
Hue
Saturation
Value
V

# Value/Luminance/Saturation

OK for quantitative data when length & size are used.

Not very many shades recognizable

Selective: yes

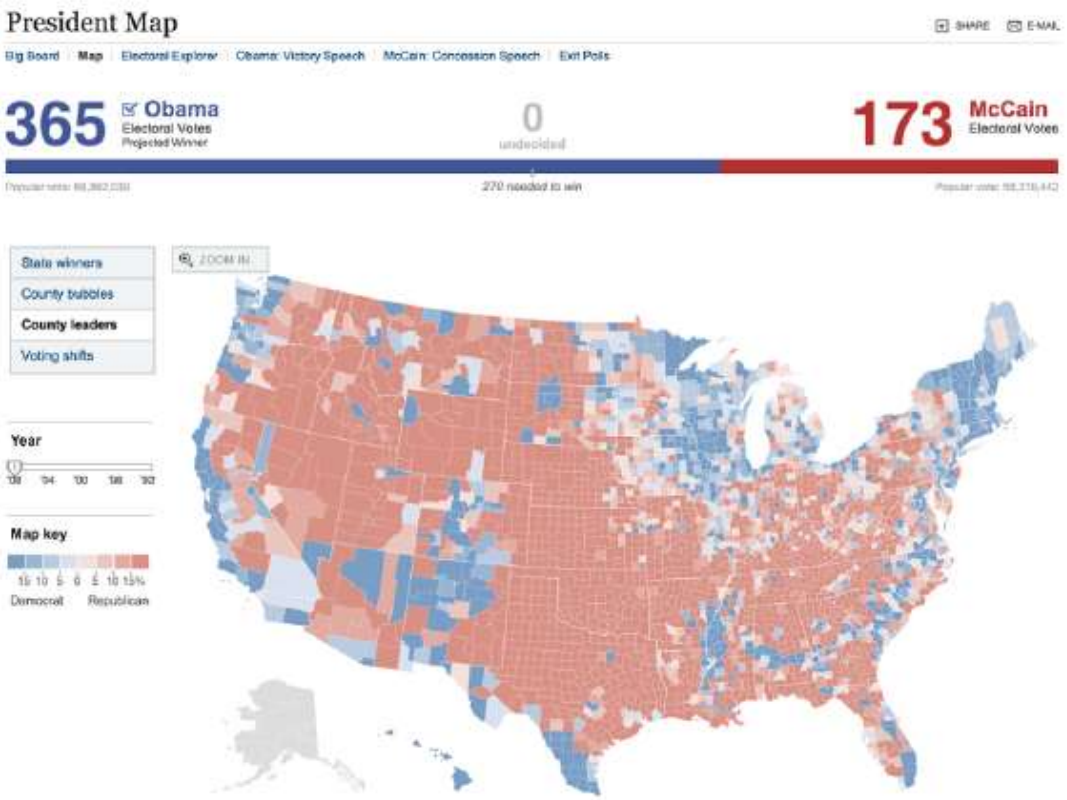
Associative: yes

Quantitative: somewhat (with problems)

Order: yes

Length: limited

# Example: Diverging Value-Scale



# Color Hue

- Good for qualitative data (identity channel)
- Limited number of classes/length (~7-10!)
- Does not work for quantitative data!
  - Selective: yes
  - Associative: yes
  - Quantitative: no
  - Order: no
  - Length: limited

# Color: Bad Example

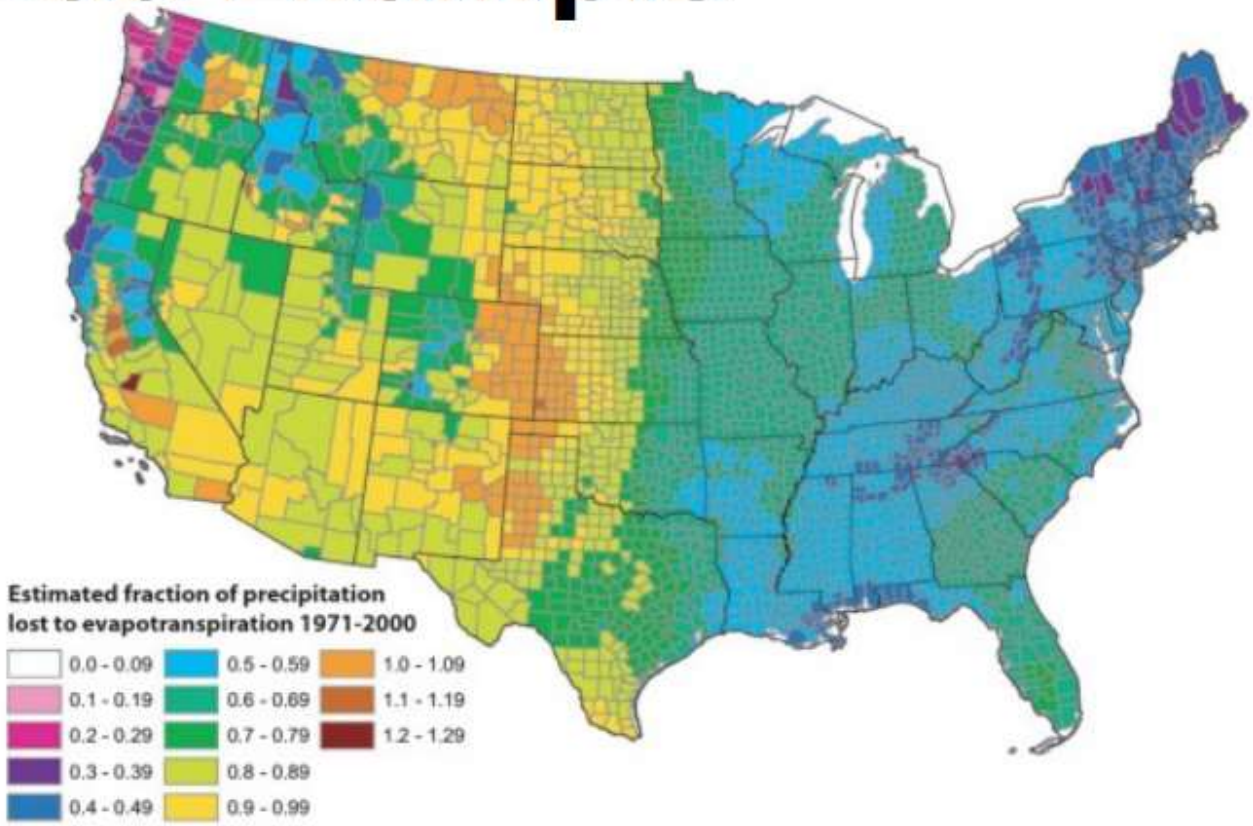



FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation ($P$) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of $ET/P$ were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions >1 are agricultural counties that either import surface water or mine deep groundwater.

Cliff Mass

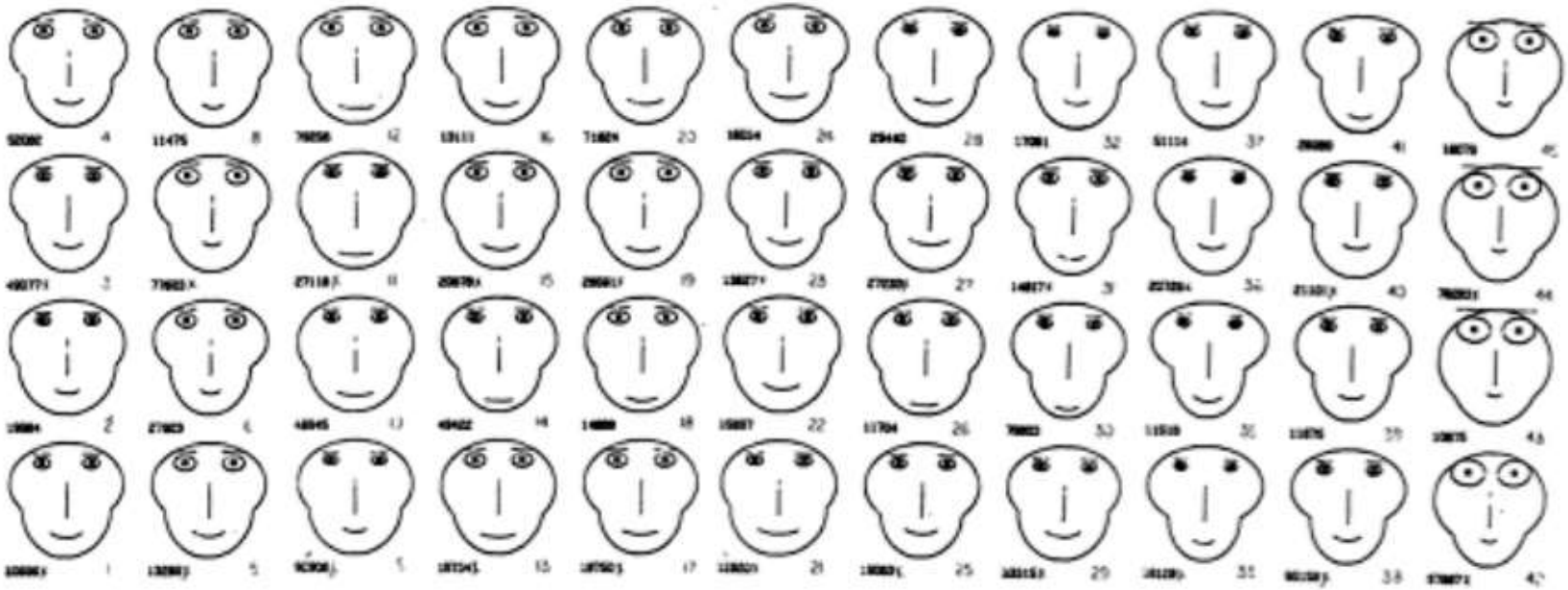
# Shape

- Great to recognize many classes.
  - No grouping, ordering.
  - Selective: yes
  - Associative: limited
  - Quantitative: no
  - Order: no
  - Length: vast


ALFA ROMEO
ACURA
ASTON MARTIN
ASTON MARTIN
Audi
BENTLEY
BMW
MITSUBISHI MOTORS
Ferrari
Photoshop plugins

# Chernoff Faces

Idea: use facial parameters to map quantitative data



Does it work?
Not really!

Critique: https://eagereyes.org/criticism/chernoff-faces

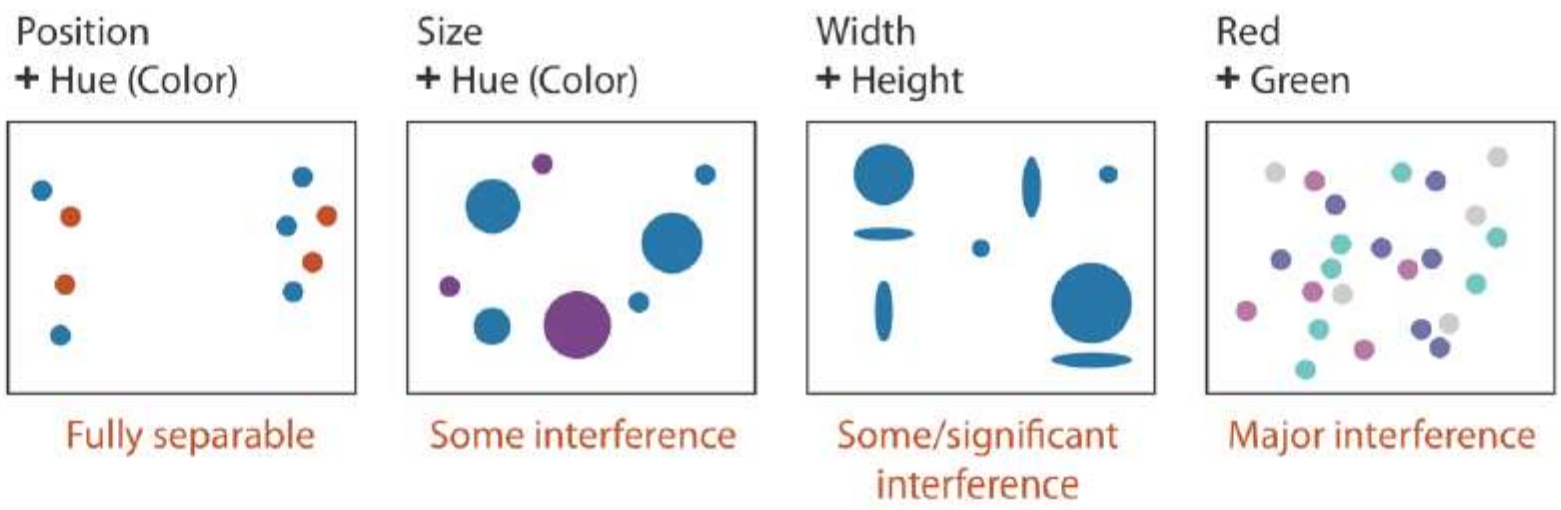
# Mapping to Data Types

| | Nominal | Ordinal | Quantitative |
|---|---|---|---|
| Position | ✔ | ✔ | ✔ |
| Size | ✔ | ✔ | ✔ |
| (Grey)Value | ✔ | ✔ | ~ |
| Texture | ✔ | ~ | ✖ |
| Color | ✔ | ✖ | ✖ |
| Orientation | ✔ | ✖ | ✖ |
| Shape | ✔ | ✖ | ✖ |

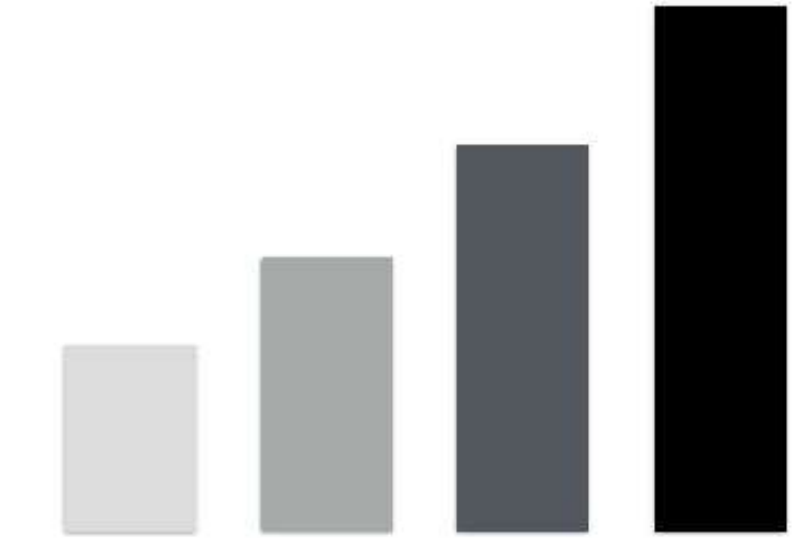
✔ = Good
~ = OK
✖ = Bad

# Separability of Attributes

Can we combine multiple visual variables?



# Redundant encoding



Length, Position and Value

# Gestalt Law of Pattern Perception

- Gestalt Principles are principles/laws of human perception that describe how humans group similar elements, recognize patterns and simplify complex images when we perceive objects.

- Designers should use the principles to organize content on websites and other interfaces so it is aesthetically pleasing and easy to understand.

# Gestalt Law of Pattern Perception (1/2)

- Proximity
  - Things close together are perceptually grouped together
- Similarity
  - Similar elements get grouped together
- Connectedness
  - Connecting different objects by lines unifies them
- Continuity
  - More likely to construct visual entities out of smooth, continuous visual elements

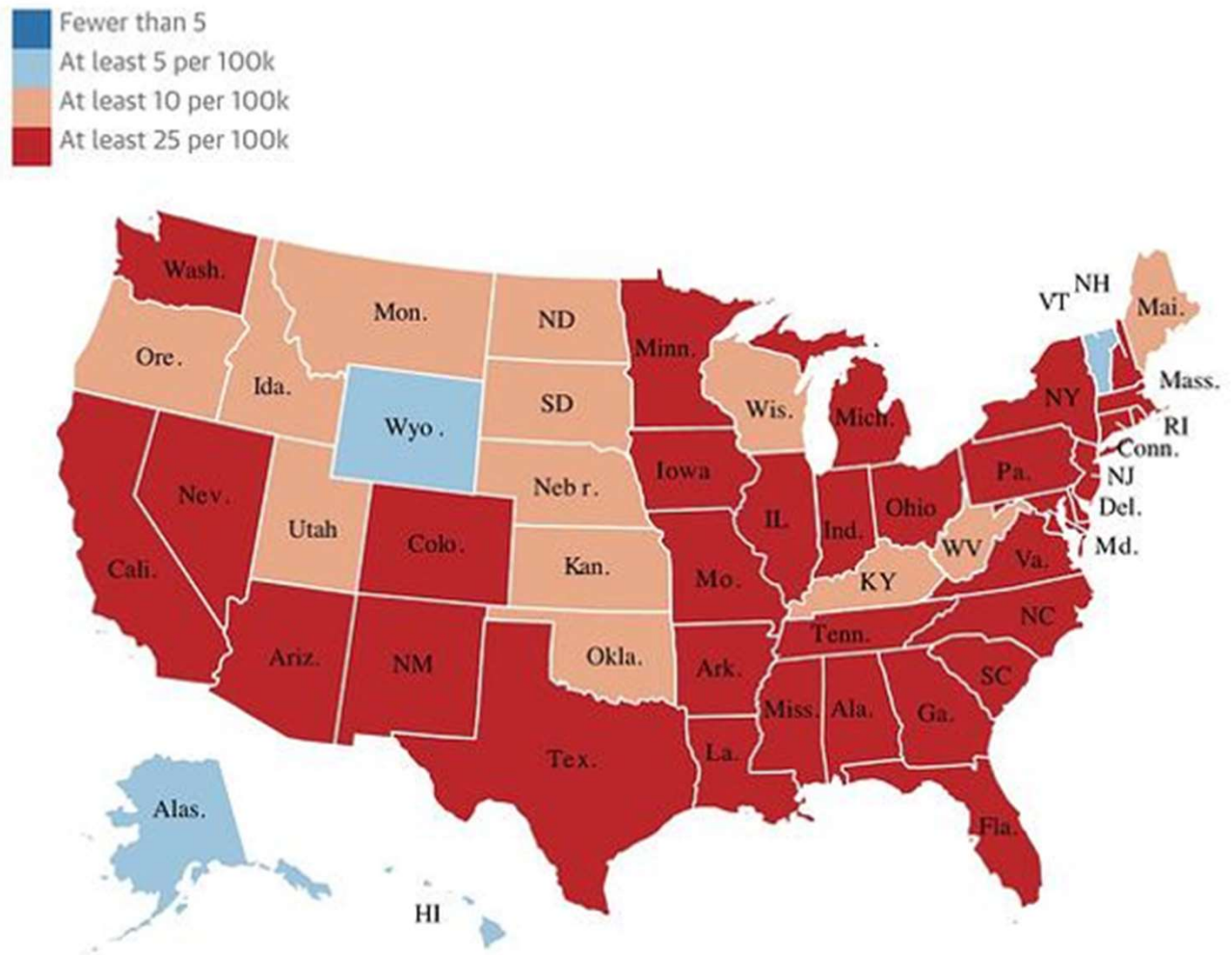
# Gestalt Law of Pattern Perception (2/2)

- Symmetry
  - Symmetrical patterns are perceived more as a whole
- Closure
  - A closed contour is seen as an object
- Relative Size
  - Smaller components of a pattern as perceived as objects
- Figure & Ground
  - Figure is foreground, ground is behind

# Exercise

- Select any visualization
- Determine the data that is represented
  - Semantics
  - Data Item
  - Data Attributes (including type)
  - Dimension
  - Dataset Type
- Determine Visual Encoding of the data
- Any criticism of the visualization?

# Number of Covid Deaths per 100,000 Americans



# Most Valuable Sports Franchises