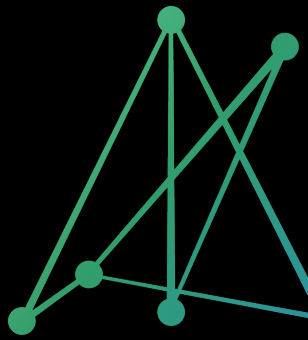


DATA SCIENCE
& ANALYTICS

University of Missouri

Data Visualization

Elements of Visualization

Data Sets

- 1D (sets, sequences)
- 2D (maps)
- Temporal data
- 3D (volume rendering)
- nD (multidimensional data)
- Graphs (relational data)
- Text

Data Types

- Data models are mathematical abstractions.
  - Integer, float, boolean, etc.

- Data types:

  - **Nominal**: they have no ordering; they are mere labels (“apples”, “oranges” or “)

  - **Ordinal**: They can represent ordered entities; not necessarily numerical. (“small”, “medium”, “large”)

  - **Quantitative**: numbers; can be done arithmetic on them.
    - Interval : dates, locations; distance is meaningful.
    - Ratio : measurements such as length, mass; proportions are meaningful.

Data Model vs. Conceptual Model

Data model: it's a low level description (numbers)

Conceptual model: semantics, reasoning, what the data model represents (temperature, salary, etc.)

Data model: 77.5, 32.4, -4.3 (floats)
Conceptual model: temperature

Data type for temperature can be:
Quantitative : continuous measurement in some unit (e.g. Fahrenheit).
Ordinal : "hot", "warm", "cold".
Nominal : "burned", "not burned".

Data Model vs. Conceptual Model

Data type:

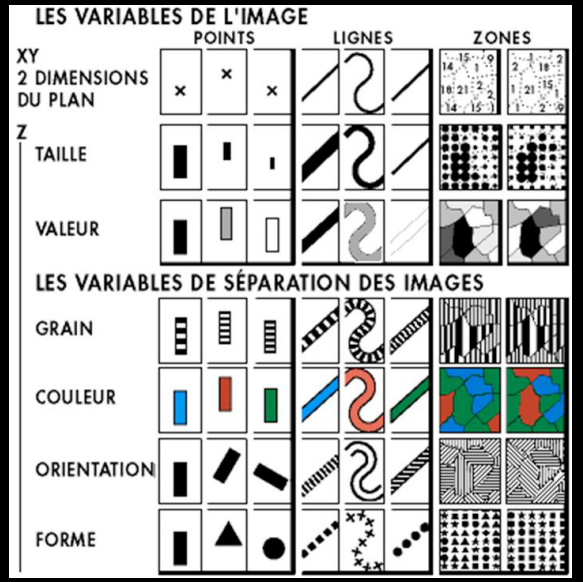
Why is “Order ID” nominal ?

Because the “Order ID” is just a label, it does not represent any ordering, measurement, date, etc.

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Speci	1 = Quantitative	0.6	6/6/05
70	12/18/06	5-Low	2 = Nominal	0.59	12/23/06
70	12/18/06	5-Low	3 = Ordinal	0.82	12/23/06
96	4/17/05	2-High		0.55	4/19/05
97	1/29/06	3-Medium		0.38	1/30/06
129	11/19/08	5-Low		0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Visual Variables

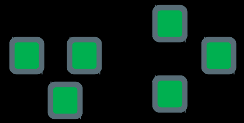
- Used for visual encoding of data.
- J. Bertin (French Cartographer, 1967)
- Position
- Size
- Value (gray)
- Color
- Texture
- Orientation
- Shape



Characteristics of Visual Variables

Selective: is it distinct from others? (shape, color)

Associative: does it support grouping? (positions)

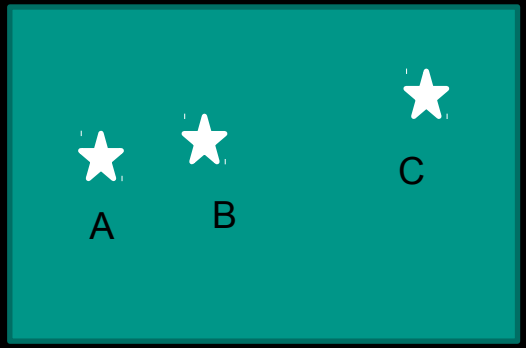


Quantitative: Can the difference between two of them be quantified? (length of bars)

Order: does it support ordering (size, area, how about color?)

Position

Position is the strongest visual variable.
Suitable for all data types.
Problem: it can become cluttered.

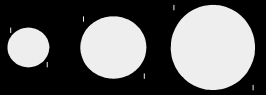


A, B, C are distinguishable.
B is between A and C.
BC is twice as long as AB.

Size, Length, Area

- Good visual variable.
- Easy to compare big vs. small.
- Supports grouping.

- Ability to judge differences:
  - Aligned bars: good
  - Changes in length: OK
  - Changes in area: harder to judge (especially for circles).



Other Visual Variables

Shape:

- Good for representing classes (nominal data type).
- Does not support ordering, grouping.

Texture:

- Good in black and white.
- Useful for maps.
- Can be confusing.

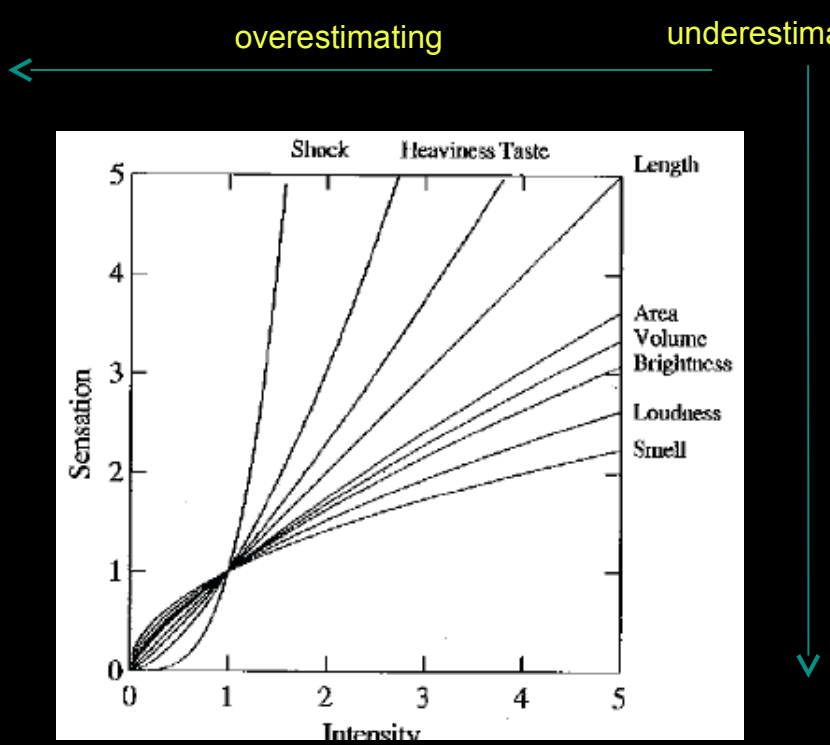
Value (grayscale)

- Good for quantitative data type.
- It's preattentive (stands out) if sufficiently different.
- Supports grouping.
- Too many shades are not recognizable.

Color

- Good for qualitative data.
- Supports limited number of classes for nominal type.
- Several pitfalls to be aware of! (check the "color" slides).

Steven's Power Law, 1961

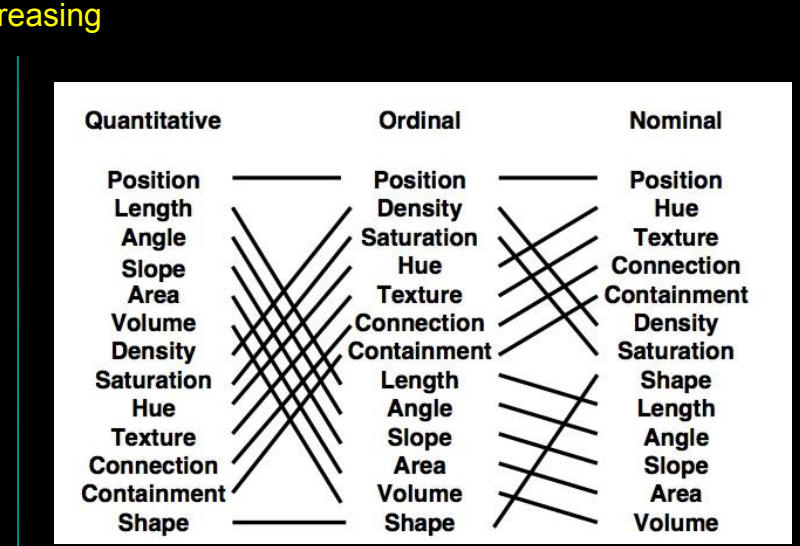


- physical stimulus vs. its perceived intensity or strength.
- Human visual perception underestimates the area and volume.
- Estimating Length is accurate.

- Keep this in mind when choosing visual variables to represent data.

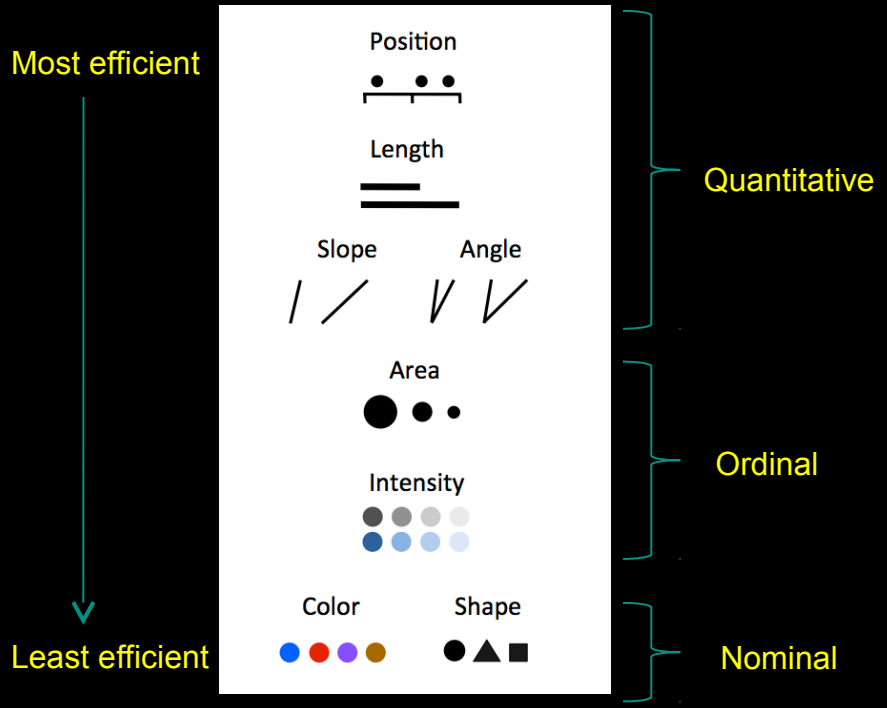
Mackinlay's Visual Variables

- J. Mackinlay extended the list of visual variables (1986).
- Provided sorting for their accuracy depending on the type.
- Example: Length has little use for nominal data type, but is very accurate to represent quantitative data.



Efficient Visual Variables

- Choice of visual variables dependent upon data type.
- Position and length are the most efficient variables.
- Shape is the least efficient variable; should be used for only nominal data type (small number of classes, etc.)



Ineffective Visualizations

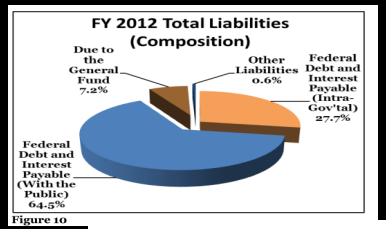


Figure 10

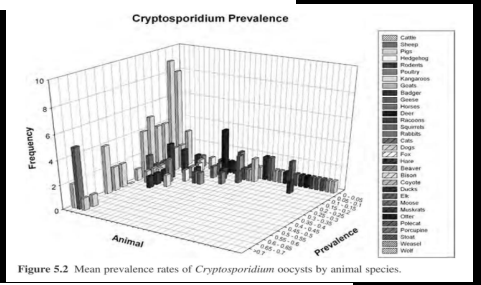


Figure 5.2 Mean prevalence rates of *Cryptosporidium* oocysts by animal species.