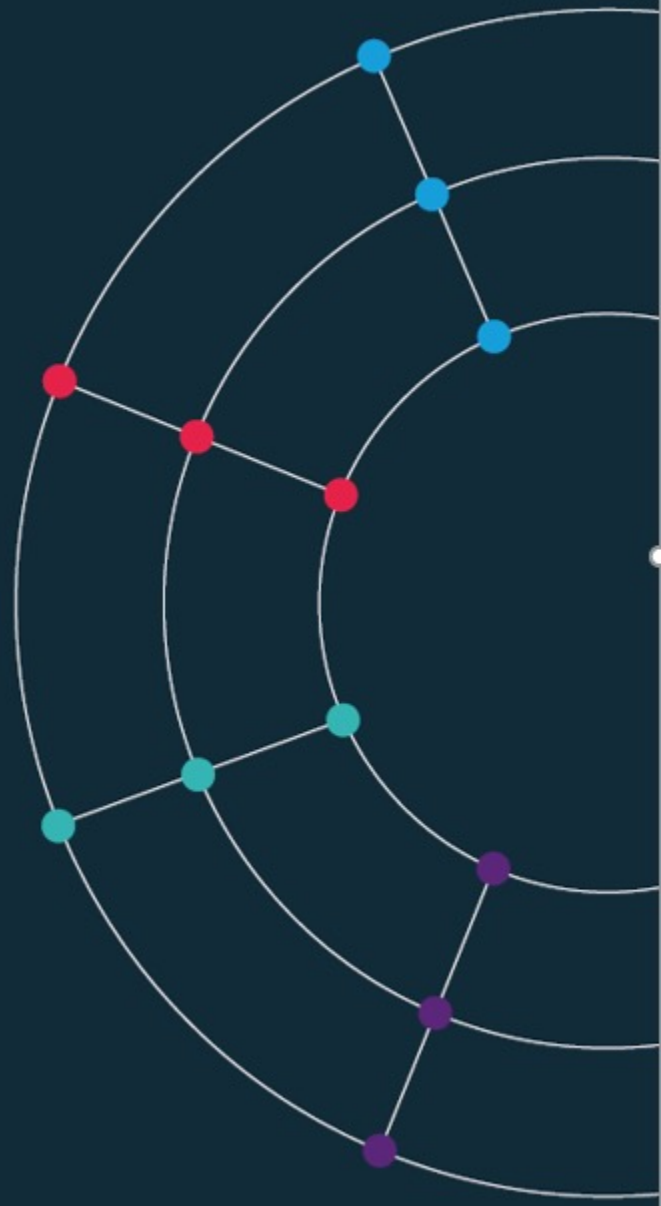
Session 4.

Advanced visualisations

[60 min]



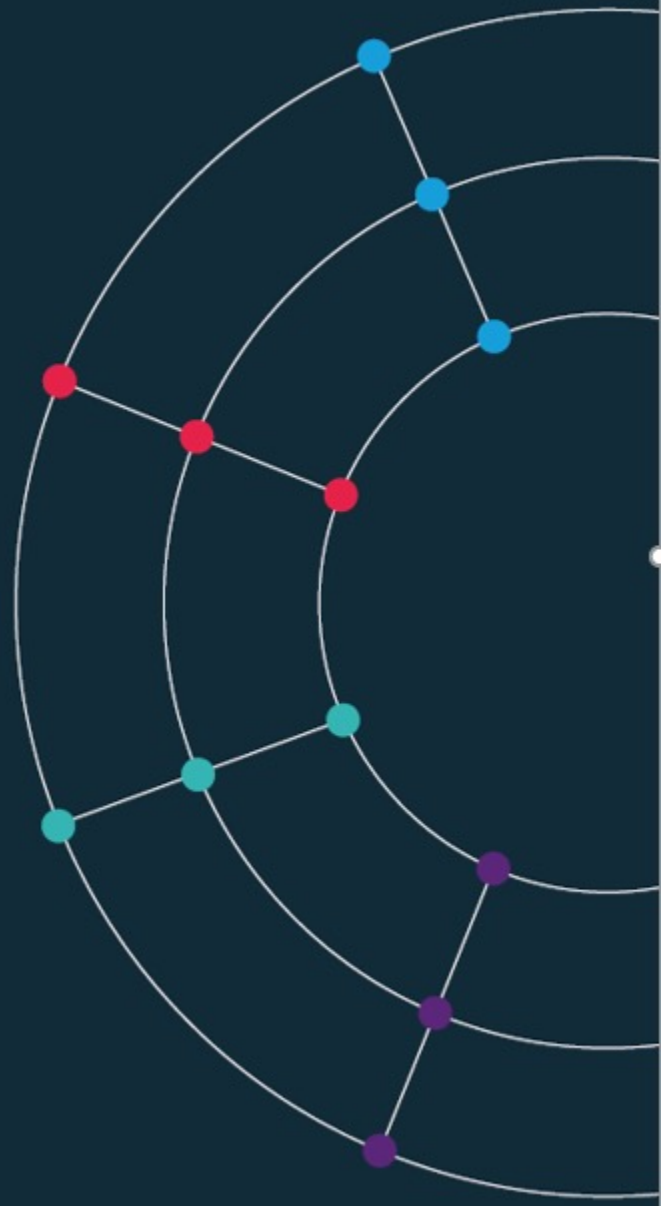
Session 4.

Advanced visualisations

[60 min]

The grammar of graphics

[30 min]

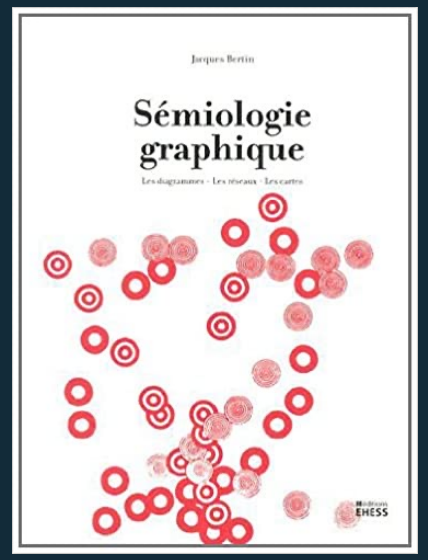


What is Data Science?

- Artificial Intelligence?
- Machine Learning?
- Deep Learning?
- Big Data?

https://twitter.com/TamaraMcCleary/status/1061197523610550272

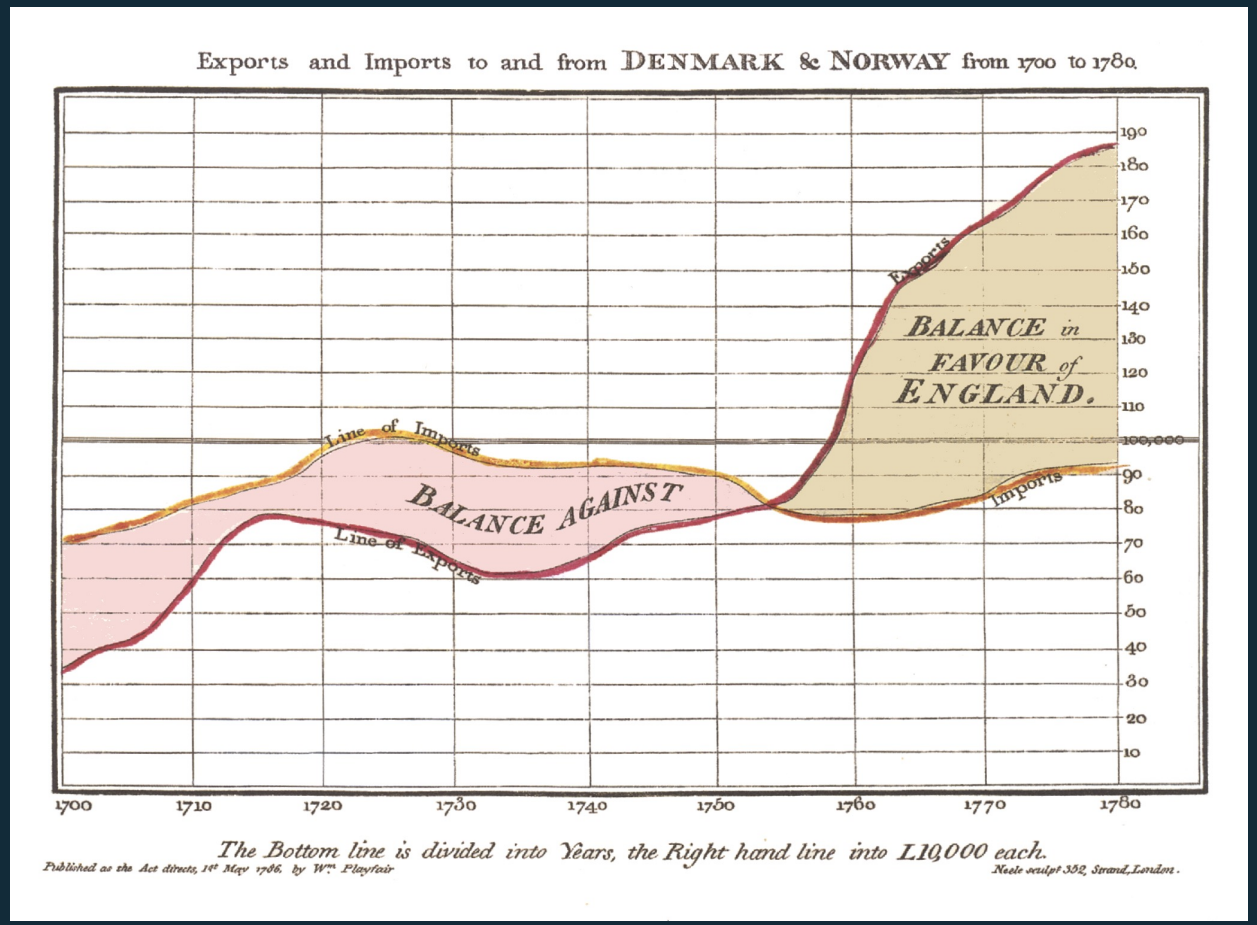
Visual language is a sign system.

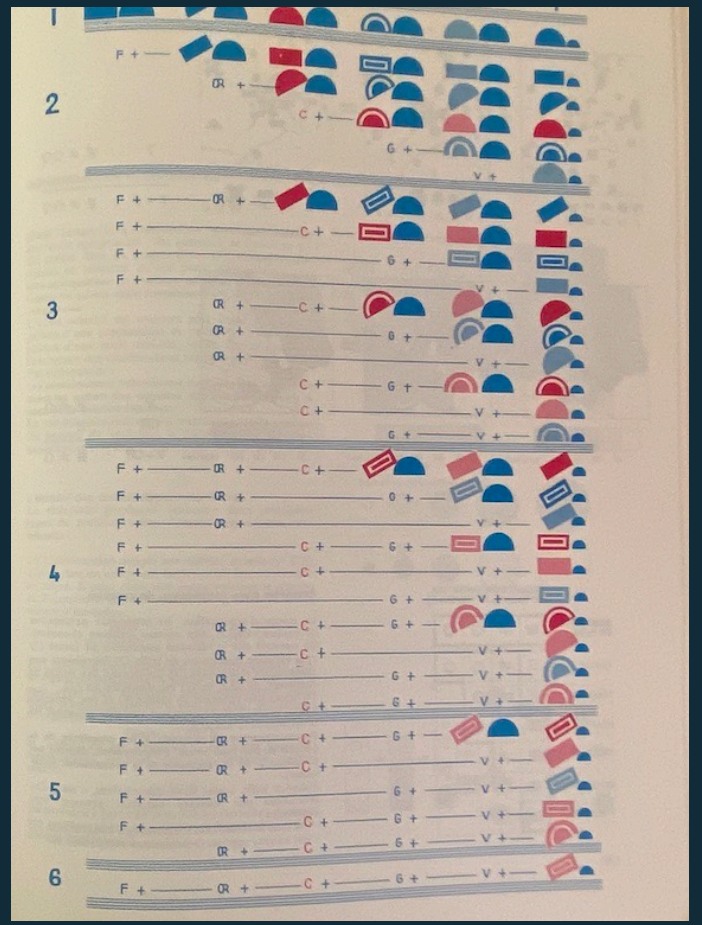


- Images perceived as a set of signs.
- Sender encodes information in signs.
- Receiver decodes information from signs.
- In his foreword to the 1983 English translation Howard Wainer called Bertin's work, the most important work on graphics since the publication of Playfair's Atlas

Jacques Bertin, Semiologie Graphique, 1967

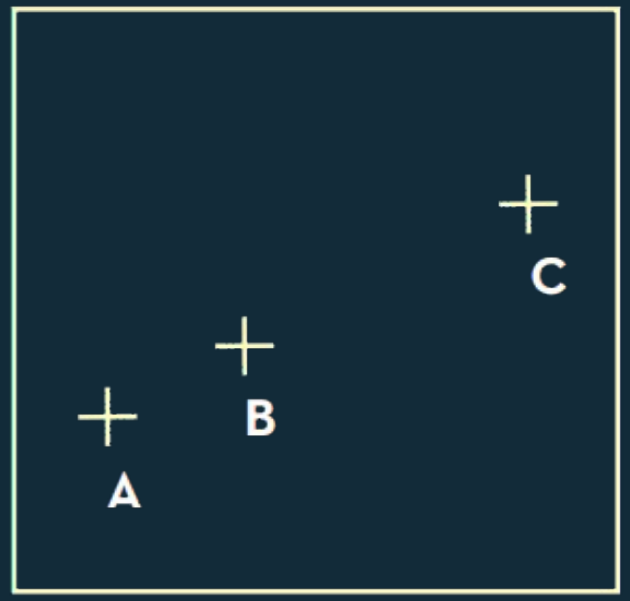
Visual language is a sign system.





William Playfair, The Commerical and Political Atlas, 1785 and Jacques Bertin, Semiologie Graphique, 1967
both collected from Jemery Norman's HistoryOfIformation.com, 2021

Bertin's semiology of graphics.



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

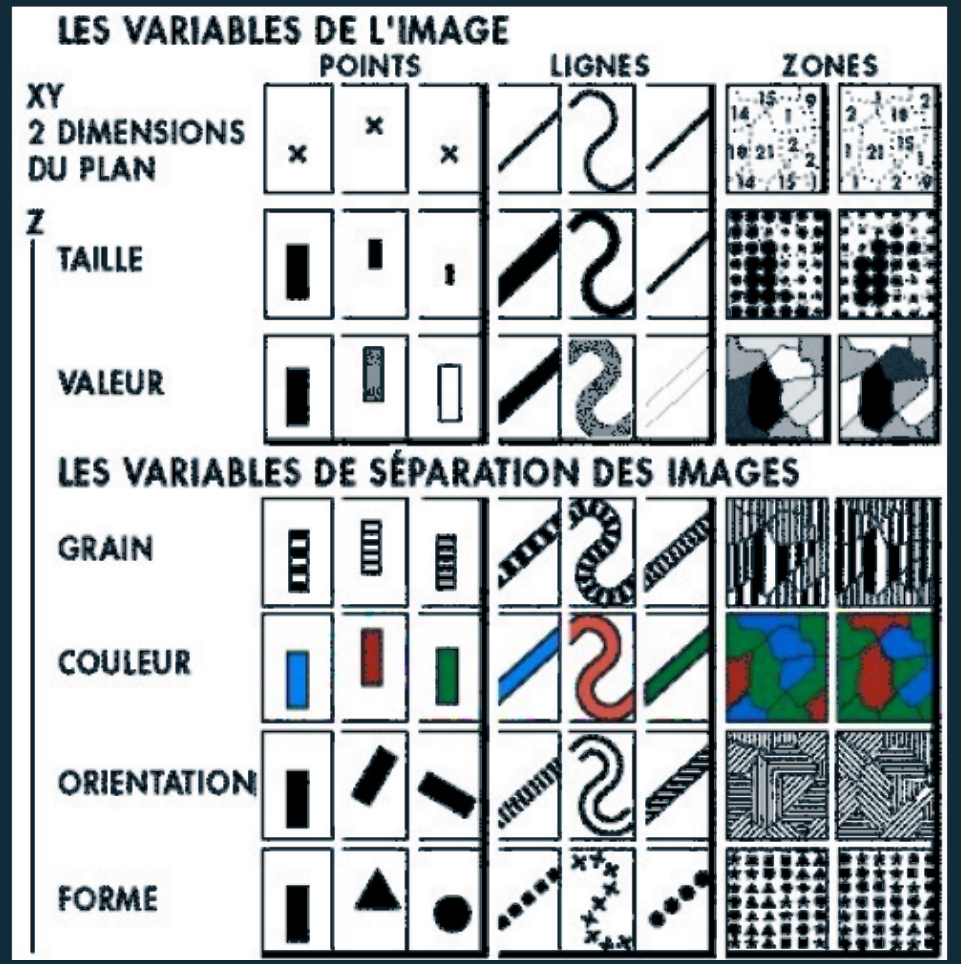
∴ Encode quantitative variables

"Resemblance, order and proportion are the three signifieds in graphics." - Bertin

Adapted from Jeffrey Heer, 2018, UW CSE442, after Jacques Bertin, Semiologie Graphique, 1967

Visual encoding variables. Visual Grammar.

- Position (x2)
- Size
- Value (Saturation)
- Texture
- Colour
- Orientation
- Shape
- (Time/Animation, Focus, Opacity)



Adapted from Jeffrey Heer, 2018, UW CSE442, after Jacques Bertin, Semiologie Graphique, 1967

Position	N	O	Q
Size	N	O	Q
Value	N	O	Q
Texture	N	O	
Color	N		
Orientation	N		
Shape	N		

Nominal
Ordered
Quantitative
Note: Q < O < N

- Bertin's list: based on theoretical considerations of semiology

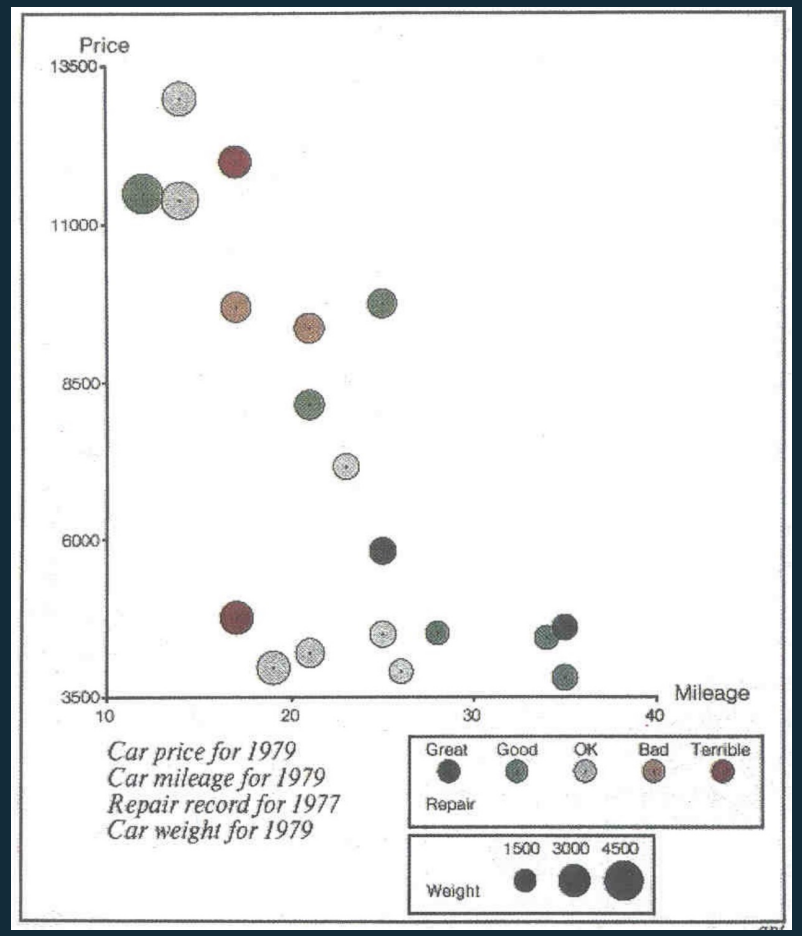
Bertin's

"Levels of Organisation"

Adapted from Jeffrey Heer, 2018, UW CSE442, after Jacques Bertin, Semiologie Graphique, 1967

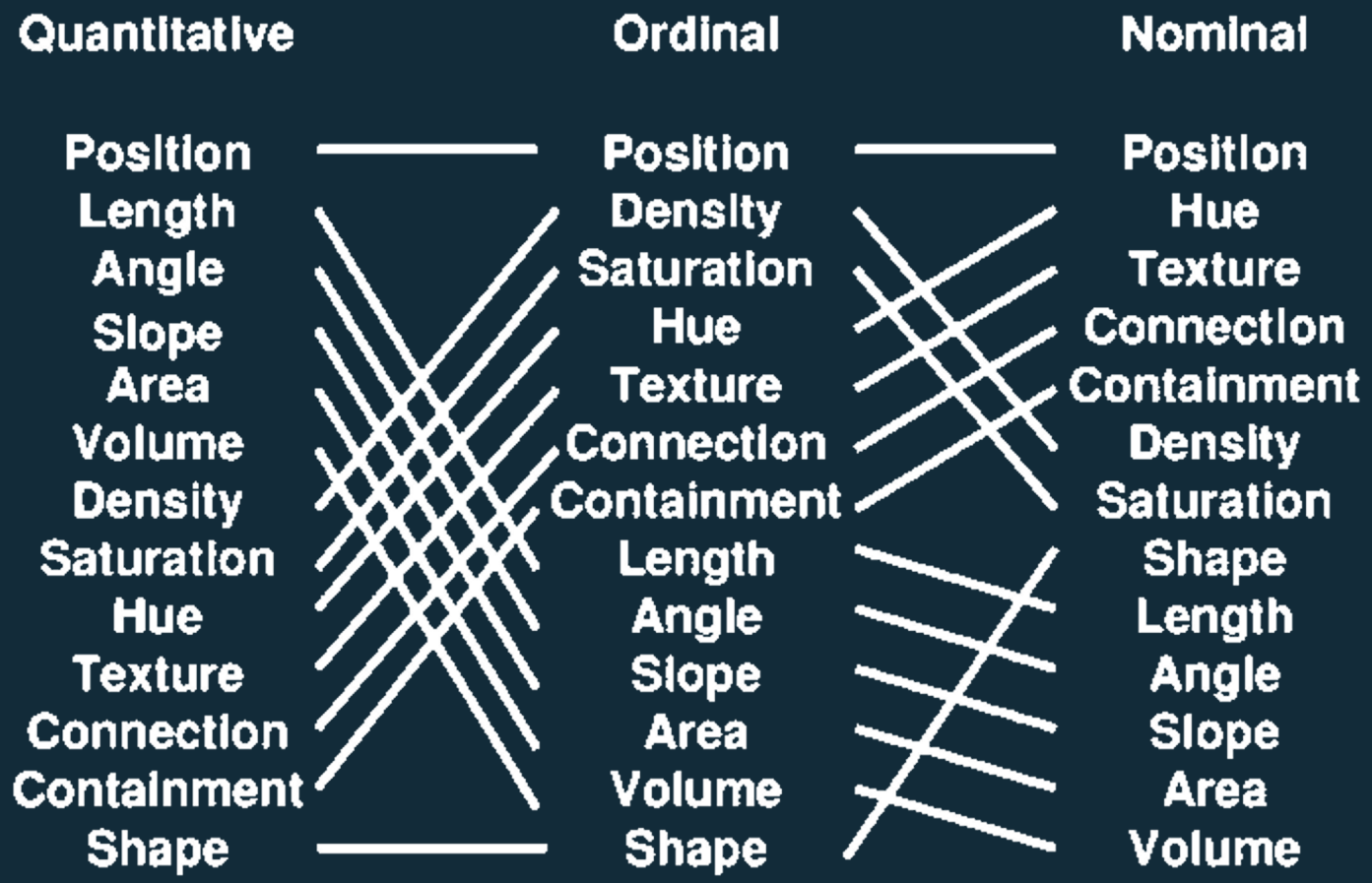
Mackinlay design criteria.

- Formalizes Bertin for machines.
- Expresiveness A set of facts is expressible in a visual language if the sentences (i.e. the visualisations) express all the facts in the set of data, and only the facts.
- Effectiveness A visualisation is more effective than another if the information conveyed by one visualisation is more readily perceived than the information in the other visualization.



Adapted from Jeffrey Heer, 2018, UW CSE442 and Jock Mackinlay's presentation at Ohio State University, 2005, both based on Jock Mackinlay, Tableau/Xerox PARC/Stanford 1986

Mackinlay design criteria.



Jock Mackinlay, Tableau/Xerox PARC/Stanford 1986

- Mackinlay's list: based on his experiments with computer graphics, trying to automate and formalize the creation of charts
- The Vega visual language/grammar is built on Mackinlay's work

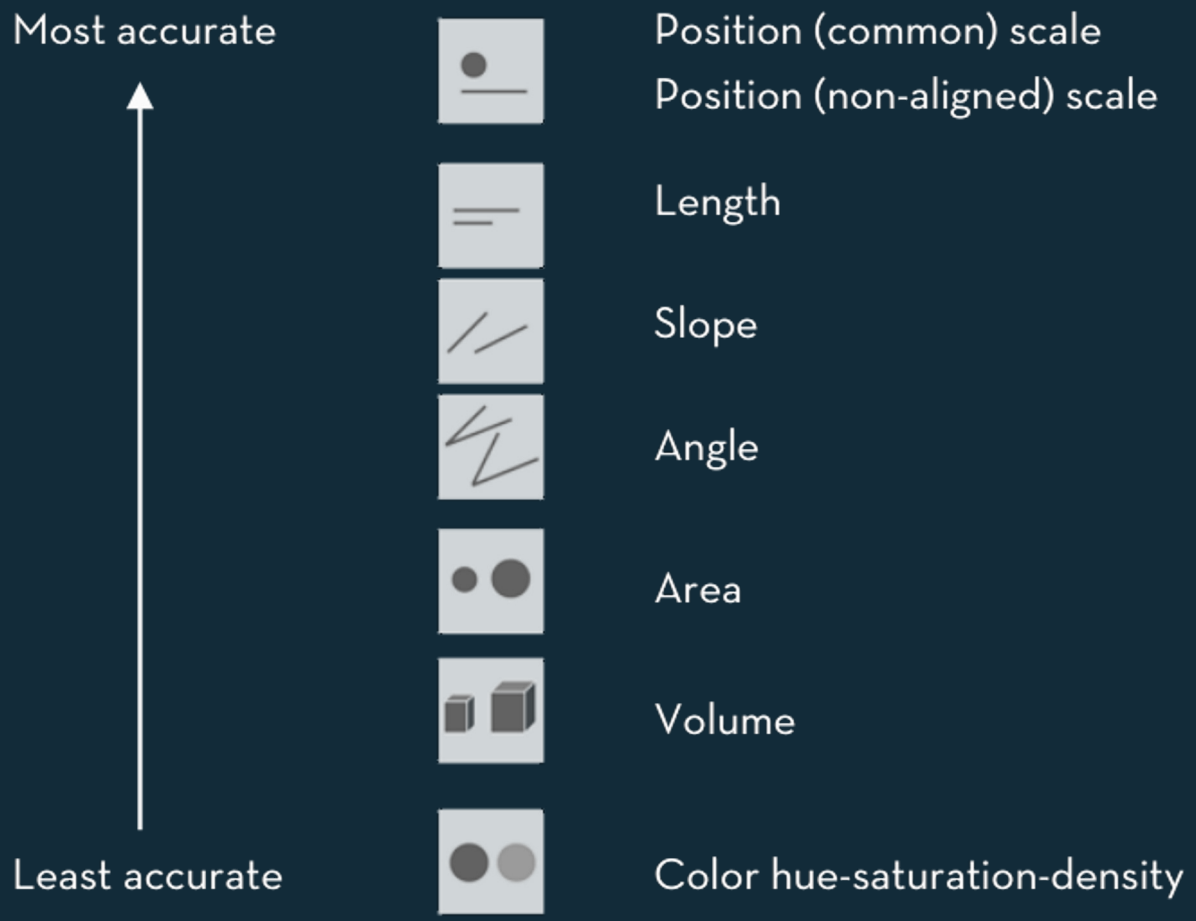
Stevens' power law.

sensation = intensity^{exponent}

Our senses are not linear!

Stanley Smith Stevens, Harvard, 1957

Stevens' power law.



Stanley Smith Stevens, Harvard, 1957

- Stevens' list: based on psychological experiments with human senses

Visual language is a sign system.

- When designing visual information use correct encodings
- data → information correct data model
- information → knowledge correct visual representation
  - Bertin's semiology of graphics
  - Mackinlay design criteria
  - Stevens' power law

Data visualization zoo.

- The actual “A Tour through the Visualisation Zoo”
- **Economics Observatory**  **Visualisation Guidelines**
- Financial Times Visual Vocabulary
- Vega Edition of the Visual Vocabulary
- The D3 Graph Gallery
- Andy Kirks's The Chartmaker Directory

Heer, Bostock, Ogievetsky, *A Tour through the Visualization Zoo*, ACM, 2010

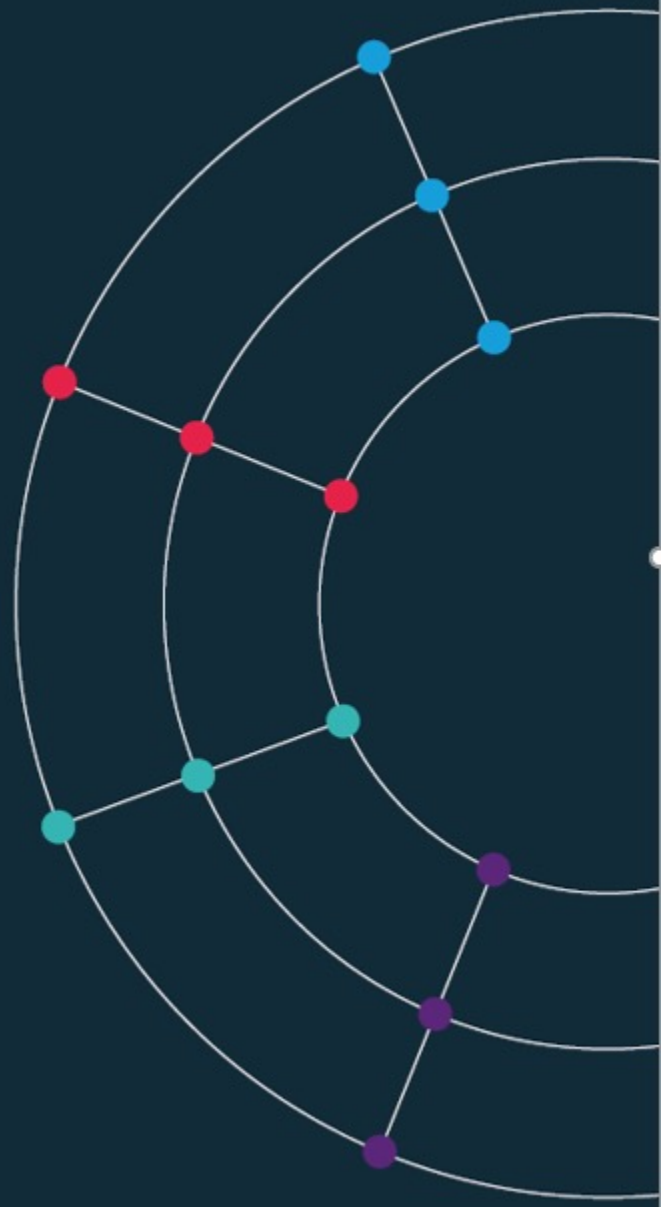
Session 4.

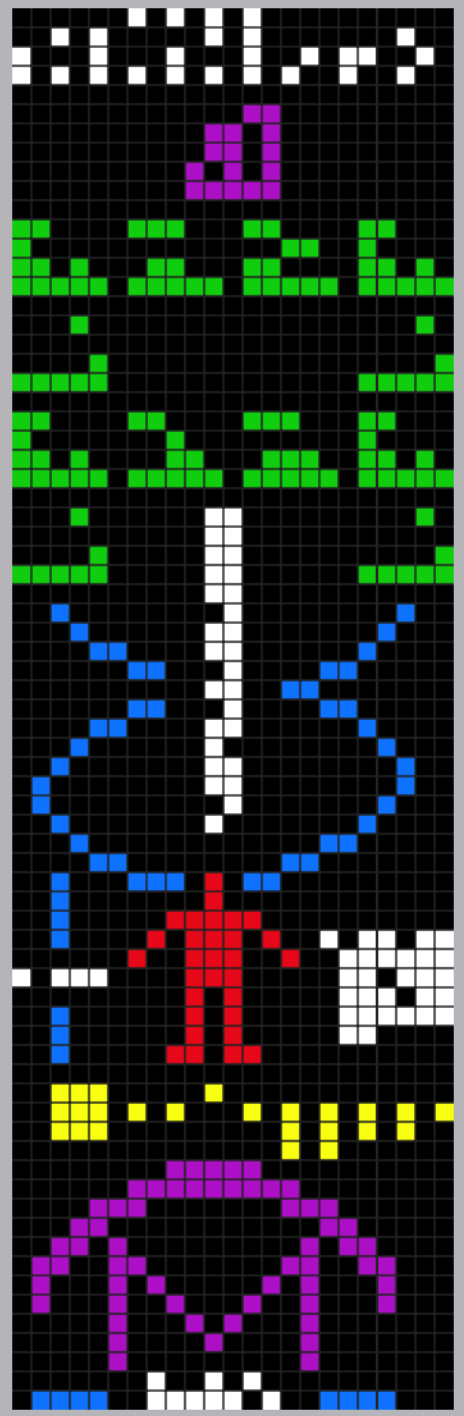
Advanced visualisations

[60 min]

Data models

[30 min]



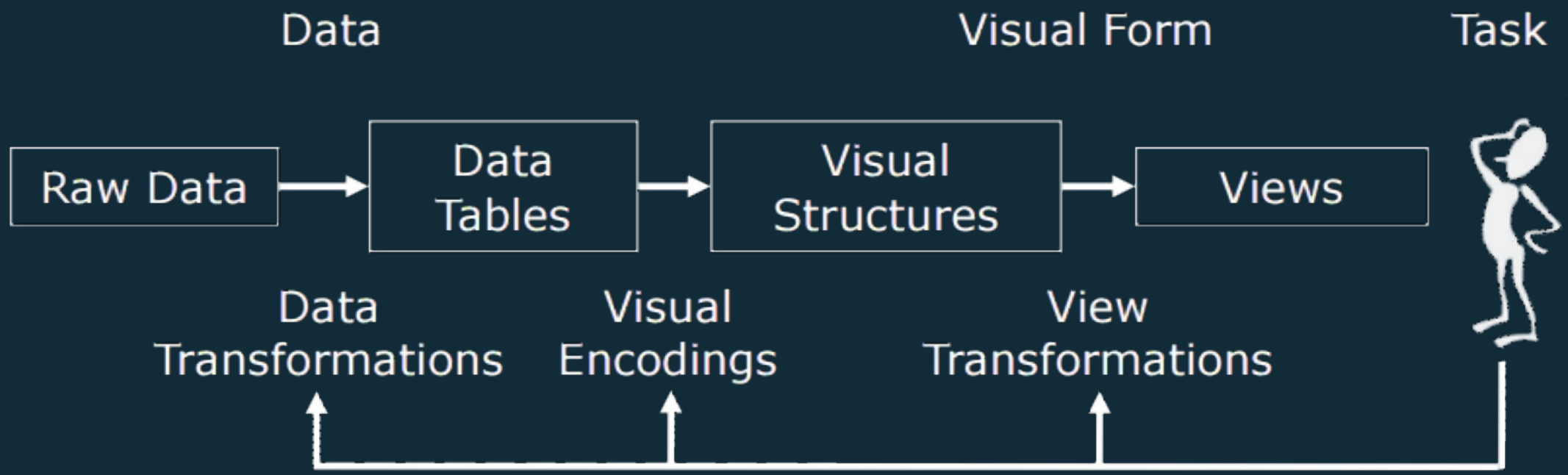


Data.

Information.

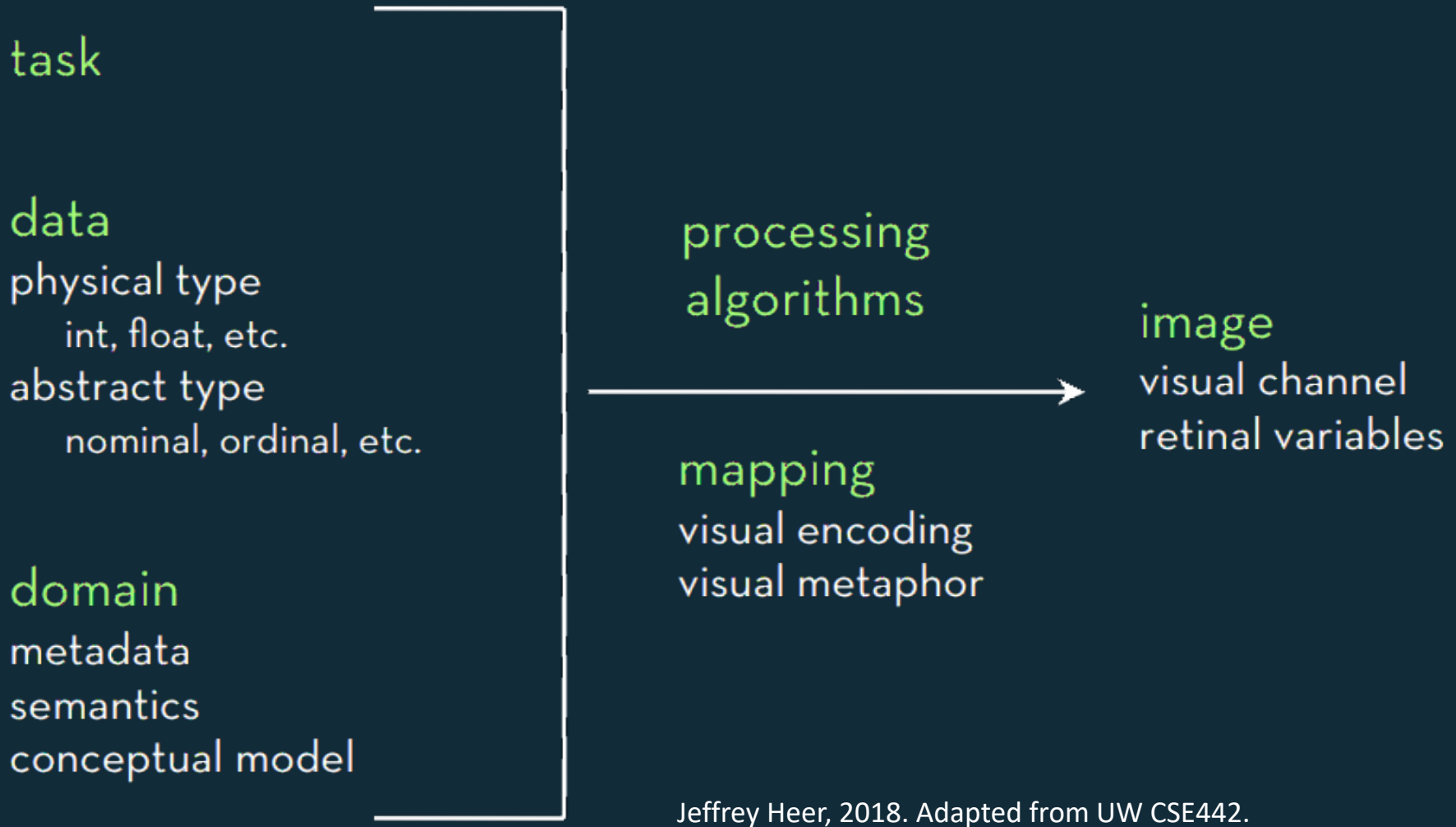
Knowledge.

visualisation reference model

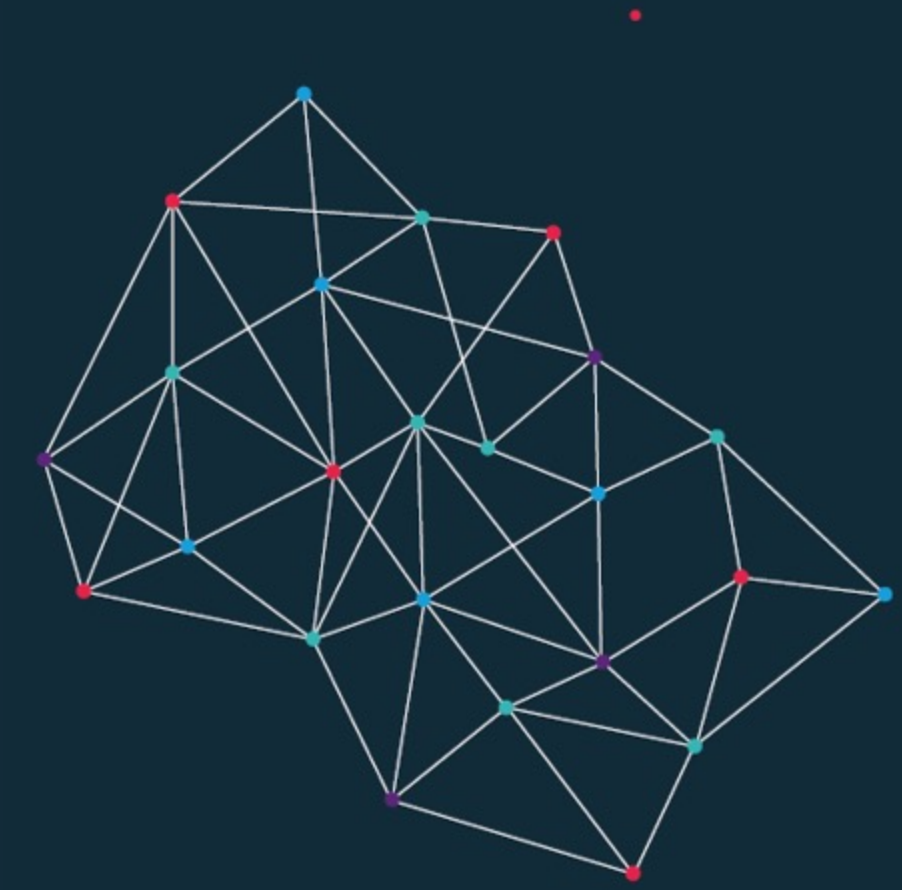


Jeffrey Heer, 2018. Adapted from UW CSE442.

data → visualisation process



Jeffrey Heer, 2018. Adapted from UW CSE442.



data domain.

data models | conceptual models

- **Data models** are a low-level description of the data.
  - Math: sets with operations on them (+ / ·). *How many? / What kind?*
- **Conceptual models** are higher level mental abstractions.
  - Semantics and support reasoning. *What does is mean/describe?*

Examples:

- **1D float** vs Temperature
- **3D float** vs Space

Jeffrey Heer, 2018. Adapted from UW CSE442.

data types

N - Nominal (labels)

- Fruits: Apples, oranges, ...

O - Ordered

- Quality of meat: Grade A, AA, AAA

O/ - Interval (Location of zero arbitrary)

(T)

- Dates: Jan, 19, 2006; Location: (LAT 33.98, LONG -118.45)
- Like a geometric point. Cannot compare directly
- Only differences (i.e. intervals) may be compared

Q - Ratio (zero fixed)

- Physical measurement: Length, Mass, Temp, ...
- Counts and amounts
- Like a geometric vector, origin is meaningful

S. S. Stevens, On the theory of scales of measurements, 1946

data types

N - Nominal (labels)

- Operations: =, ≠

O - Ordered

- Operations: =, ≠, <, >

O/ - Interval (Location of zero arbitrary)

(T)

- Operations: =, ≠, <, >, -
- Can measure distances or spans

Q - Ratio (zero fixed)

- Operations: =, ≠, <, >, -, %
- Can measure ratios or proportions

S. S. Stevens, On the theory of scales of measurements, 1946

TIDY data

this (wide form)

Country	2019	2020	2021
Austria	42	13	69
Belgium	75	12	77

should be converted to this (long form)

Country	Year	Value
Austria	2019	42
Austria	2020	13
Austria	2021	69
Belgium	2019	75
Belgium	2020	12
Belgium	2021	77