DATA

TERMS

Dataset types: to be visualized
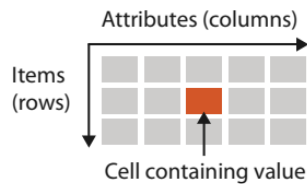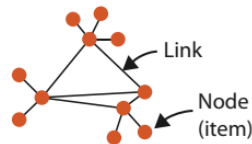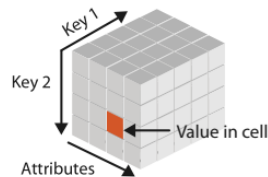
Data types: fundamental units

STRUCTURE

Structured

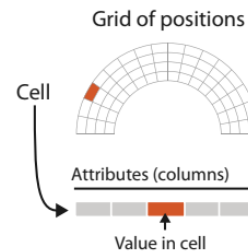- Know data types, semantics

Dataset Types

Tables
Attributes (columns)
Items (rows)
Cell containing value

Networks
Link
Node (item)

Fields (Continuous)
Grid of positions
Cell
Attributes (columns)
Value in cell
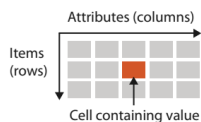
Geometry (Spatial)
Position
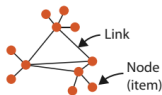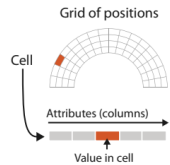
Multidimensional Table
Key 1
Key 2
Value in cell
Attributes
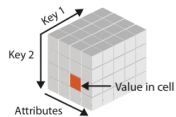
Trees

Unstructured

- No predefined data model

- Text, interspersed with facts (dates, times, locations)

- Video, images

Translate into structured data

- Natural Language Processing, text mining

- Object recognition, tracking

TEXT EXAMPLE: PHRASE NET

Pattern in text: "X's Y" in novels

[van Ham, InfoVis 2009]

DATA SEMANTICS

Basil, 7, S, Pear

What does it mean?

Semantics: real world meaning
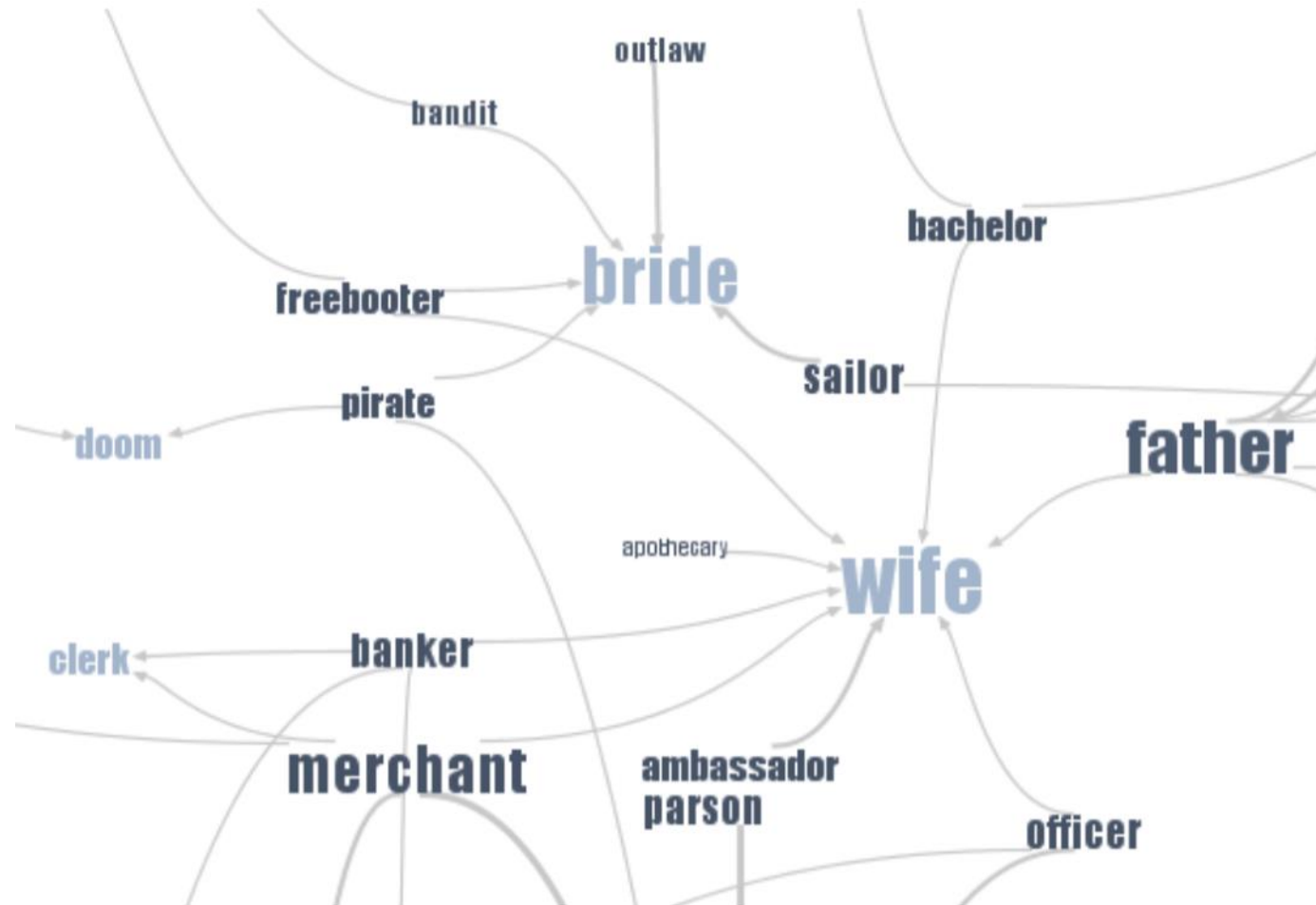
Name? city? Fruit? Height? Age? Day of month?

→ Metadata

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

DATA TYPES

Structural or mathematical interpretation of data

Item, link, attribute, position, grid

Different from data types in programming !

ITEMS & ATTRIBUTES

Item: individual entity, discrete

e.g., Patient, Car, Stock, city

"independent variable"

Attribute: measured, observed, logged property

e.g., Patient: height, blood pressure

Car: horsepower, make

"dependent variable"

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

OTHER DATA TYPES

Links

- Express relationship between two items

- Friendship on Facebook

Positions

- Spatial data -> location in 2D or 3D

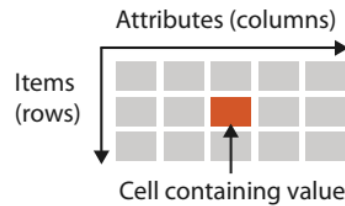- Pixels in photo, Voxels in MRI scan, latitude / longtitude

Grids (Not grids in MS Excel)

- Sampling strategy for continuous data

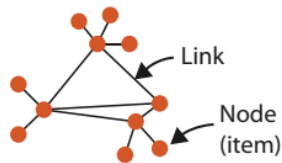- How many Voxels in MRI scan, positions of weather stations in a city
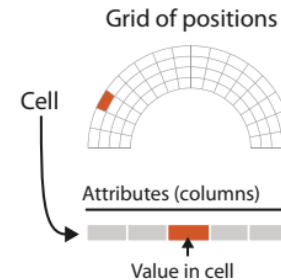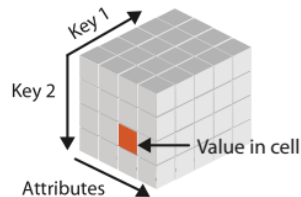
DATASET TYPES

→ **Dataset Types**

→ **Tables**

Attributes (columns)

Items (rows)

Cell containing value

→ **Networks**

Link

Node (item)

→ **Fields** (Continuous)

Grid of positions

Cell

Attributes (columns)

Value in cell

→ **Geometry** (Spatial)

Position

→ *Multidimensional Table*

Key 1

Key 2

Value in cell

Attributes

→ *Trees*

TABLES

Flat table

- one item per row

- Each column is attribute

- Unique (implicit) key

- **No duplicates**

Multidimensional table

- Indexing based on multiple keys
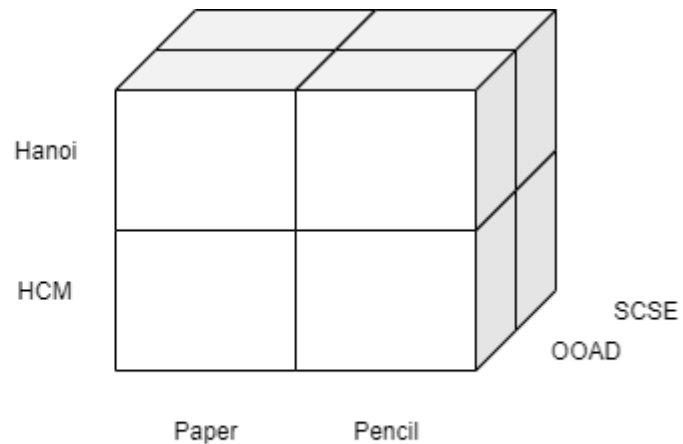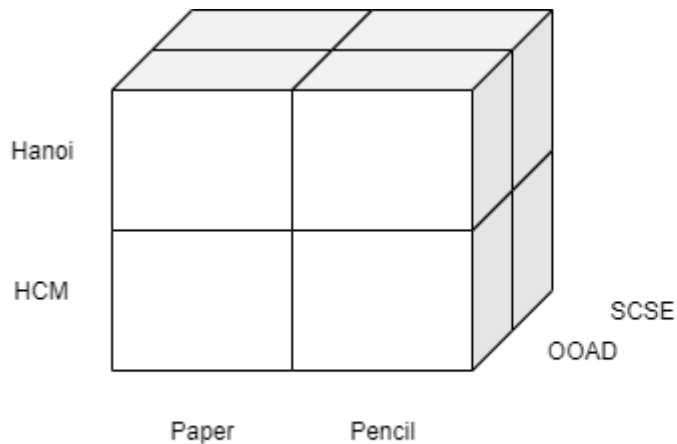
MULTIDIMENSIONAL TABLES

Multiple keys

- Item, Store location, Customer, Period

Attribute

- Quantity

Item	Store location	Customer	Quantity
Paper	HCM	SCSE	400
Pencil	HCM	OOAD	50
Paper	Hanoi	OOAD	100
Pencil	Hanoi	SCSE	20

VISUALIZE TABLES

Optogenetic

https://bl.ocks.org/jasondavies/1341281

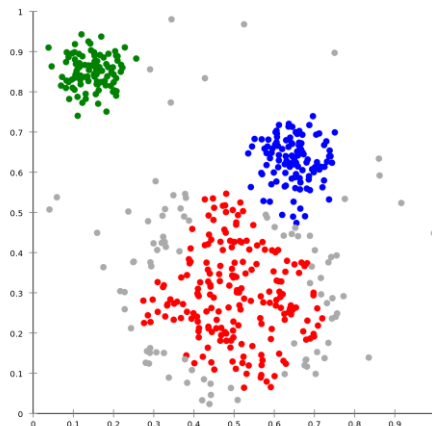# COLLECTIONS

How we groups items

Sets

- Unique items, unordered

List

- Ordered, duplicates allowed

Clusters

- Groups of similar items

Rank	School Name	Academic repu	E	Facult	Citatio	I	I
	Filter: <None>						
1.	Massachusetts Inst						
2.	University of Camb						
3.	Harvard University	100 (1)			100 (1)		
4.	UCL (University Co						
5.	University of Oxfor						
6.	Imperial College L						
7.	Yale University						
8.	University of Chic						

1985
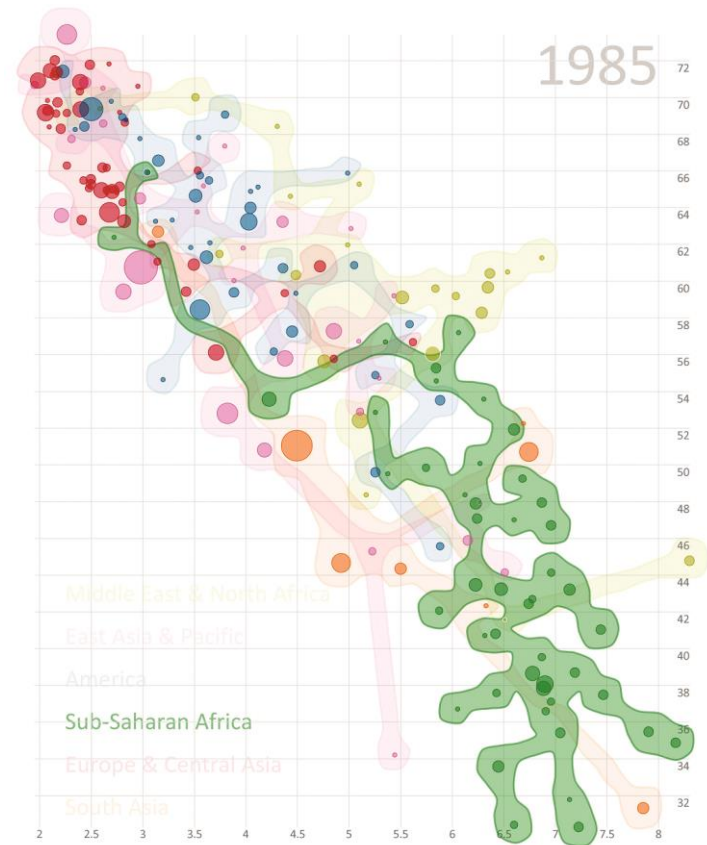
GRAPHS/NETWORKS

Set of nodes, set of edges

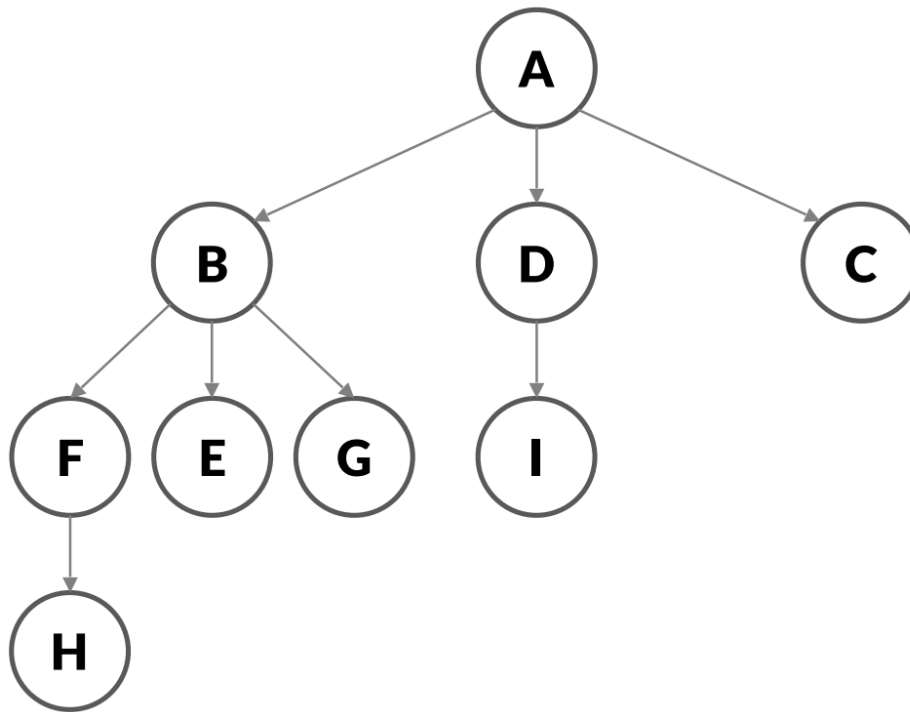**Connecting these vertices**

Diagrammatic Example

TREES
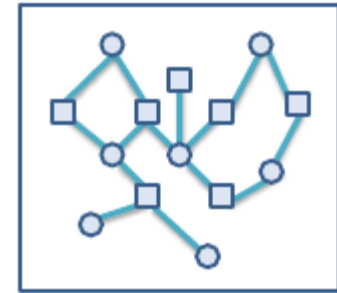
A tree is a graph with no cycles

Tree

SPECIAL GRAPHS
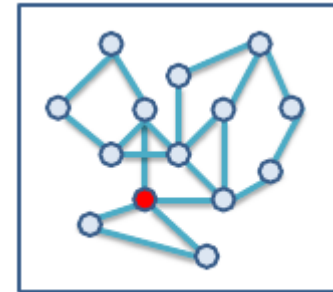
A bipartite graph

- Vertices can be partitioned into two independent sets

Bipartite Graph

An articulation point

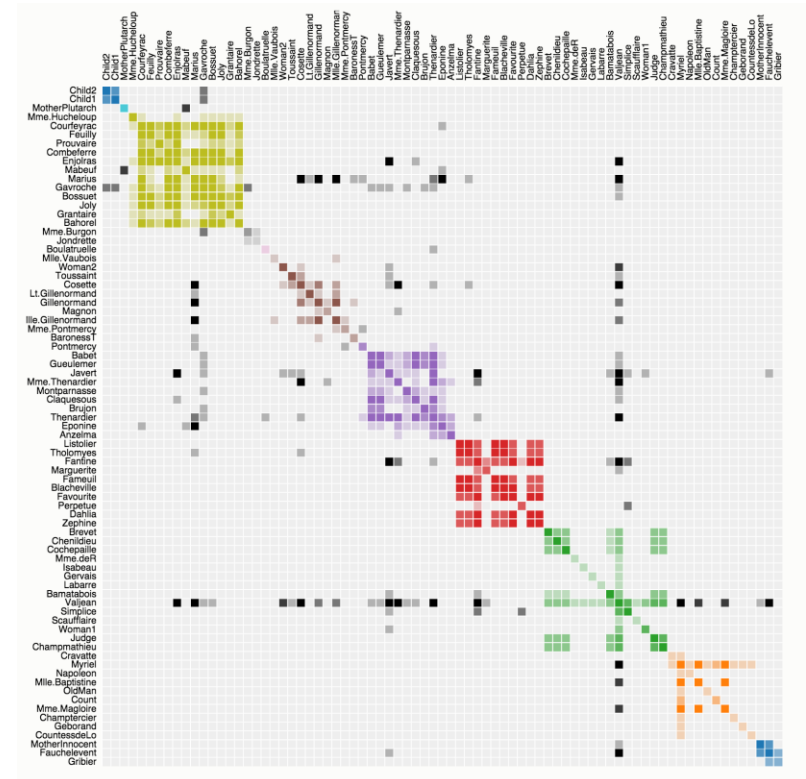- Is a vertex which if deleted would break up a connected graph

Articulation Point (red)
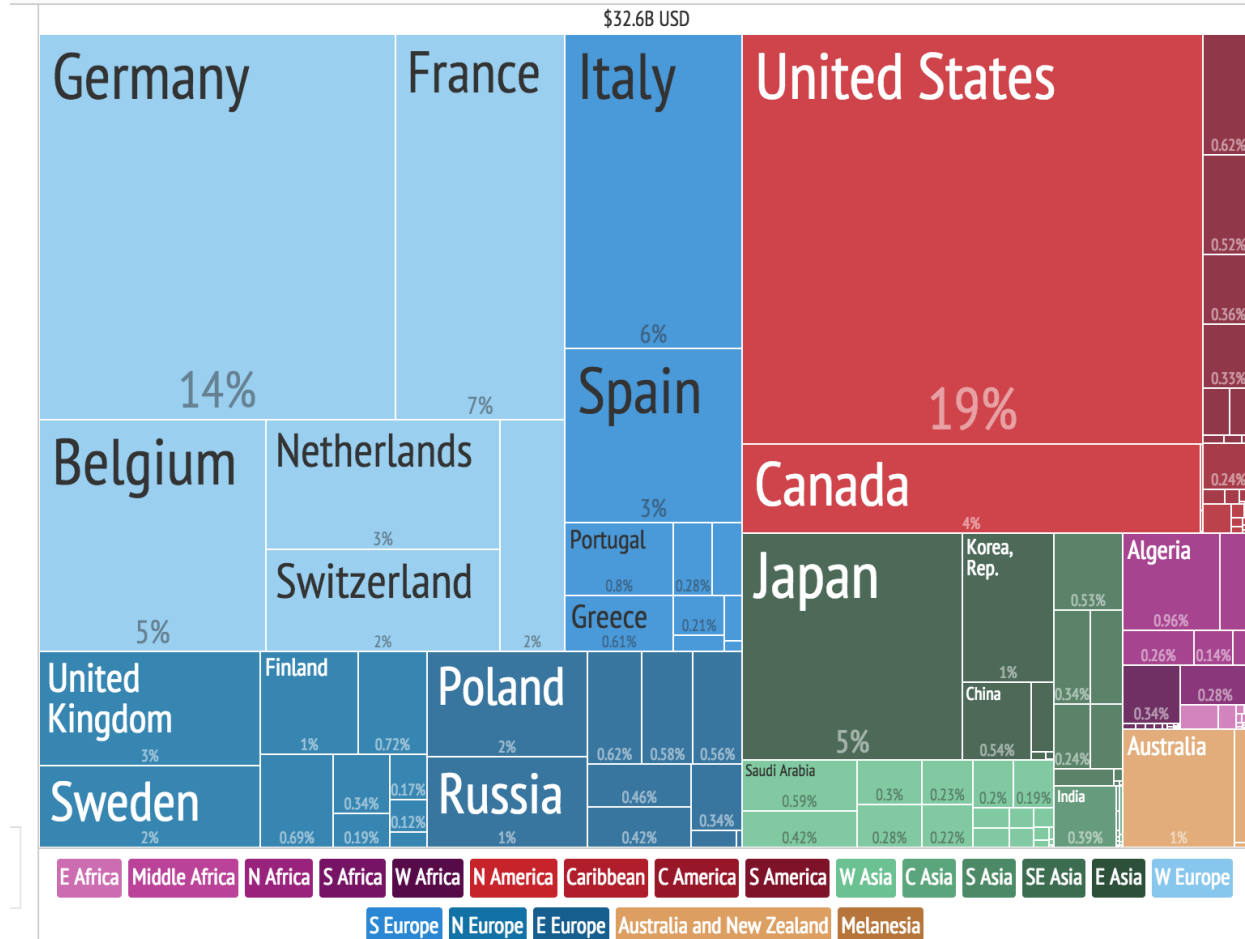
VISUALIZING GRAPHS

Node-Link diagram

https://observablehq.com/@d3/force-directed-graph

Matrix

https://bost.ocks.org/mike/miserables/

VISUALIZING GRAPHS

Treemap (Implicit Tree Visualization)

https://atlas.cid.harvard.edu/explore/?tradeDirection=import&year=2012&

product=726&country=undefined&redirected=true

FIELDS

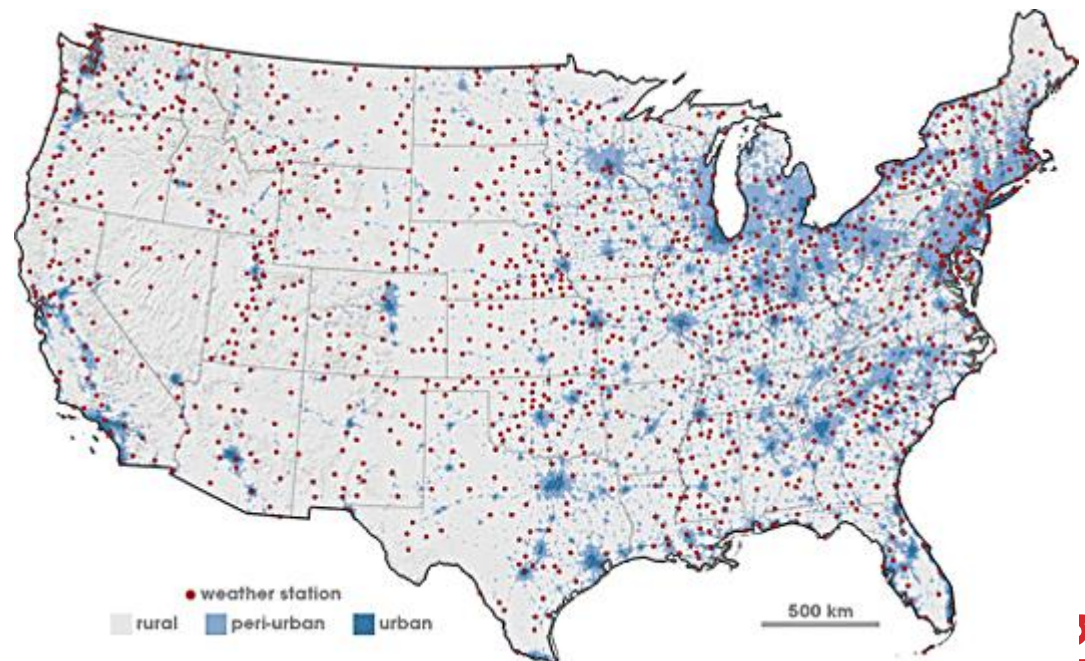Sets of attributes values associated with cells

Cell contains data from **continuous domain**

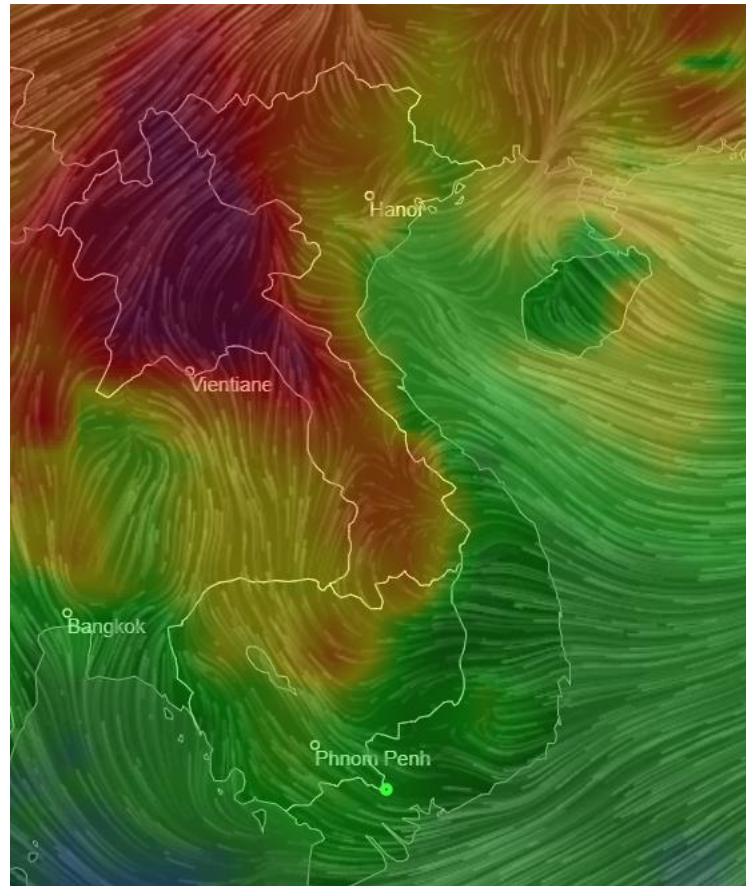- Temperature, pressure, wind velocity

Measured or simulated

Sampling & **Interpolation**

- Signal processing & stats

- weather station
- rural peri-urban urban

500 km

FIELD EXAMPLES: AIR QUALITY

Air visual
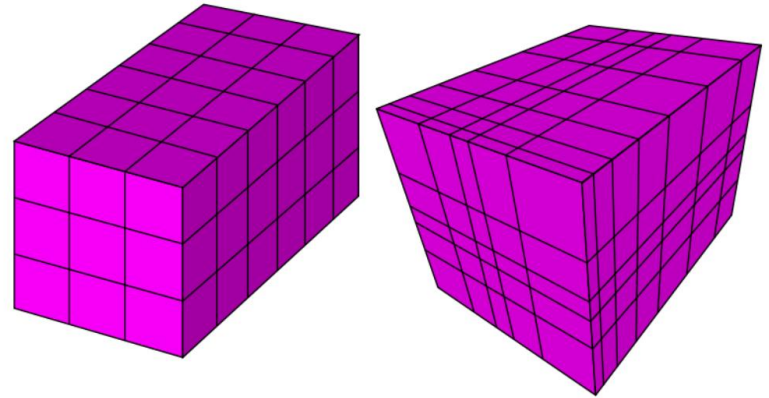
https://www.iqair.com/earth?nav

FIELDS: GRID TYPES

Uniform grid

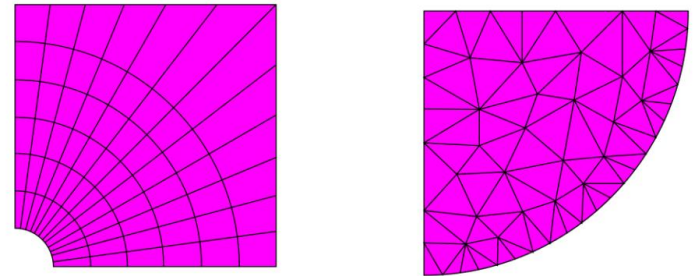Geometry & topology can be computed

Rectilinear Grid

Nonuniform sampling
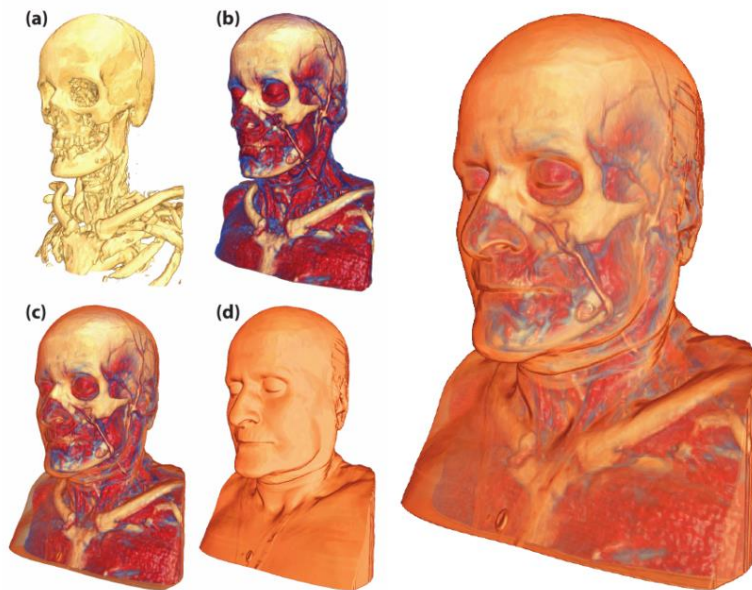
Structured grid

Allow curvilinear grids

Unstructured grid
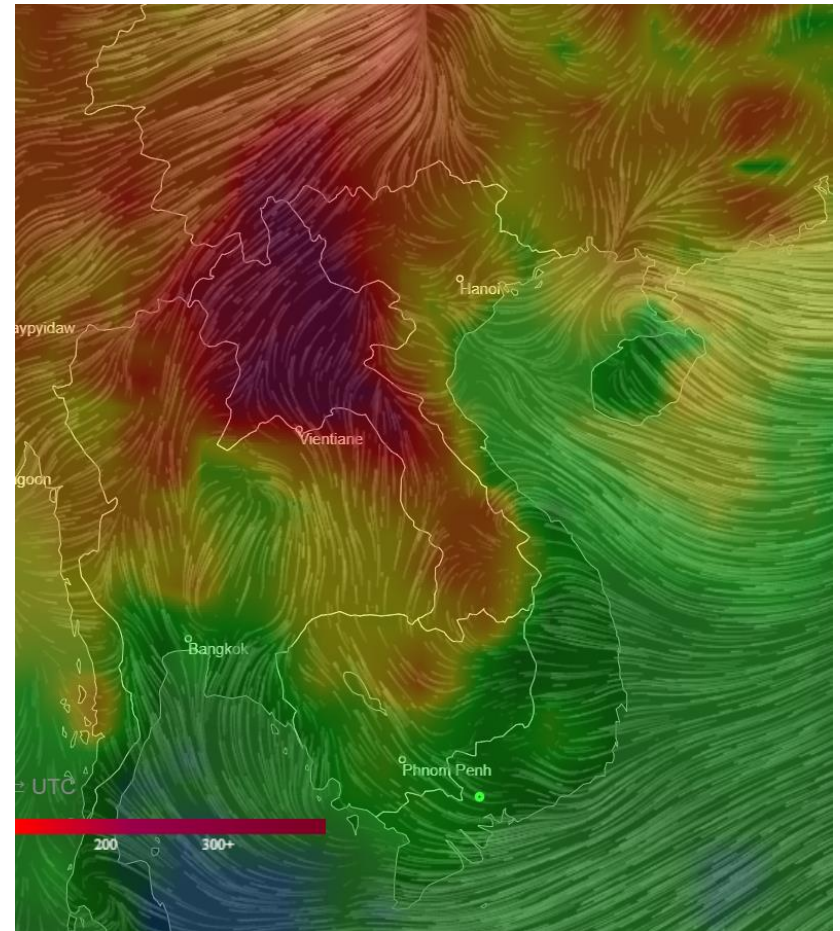
full flexibility, store position and connection

https://en.wikipedia.org/wiki/Regular_grid

VISUALIZING FIELDS

[Bruckner 2007]

INFOVIS VS SCIVIS

ACADEMIC SUBFIELDS

Information Vis

"Abstract Data"

Tables, Graphs, Maps

Free to choose spatial layout

Perception Research

Visual Analytics

InfoVis + Stats + Machine learning

Applied Work

Systems

Funding buzzword

Scientific Vis

"Spatial Data" (Fields)

Not free to choose spatial layout

Find best way to depict reality

INFOVIS VS SCIVIS

InfoVis: White Background
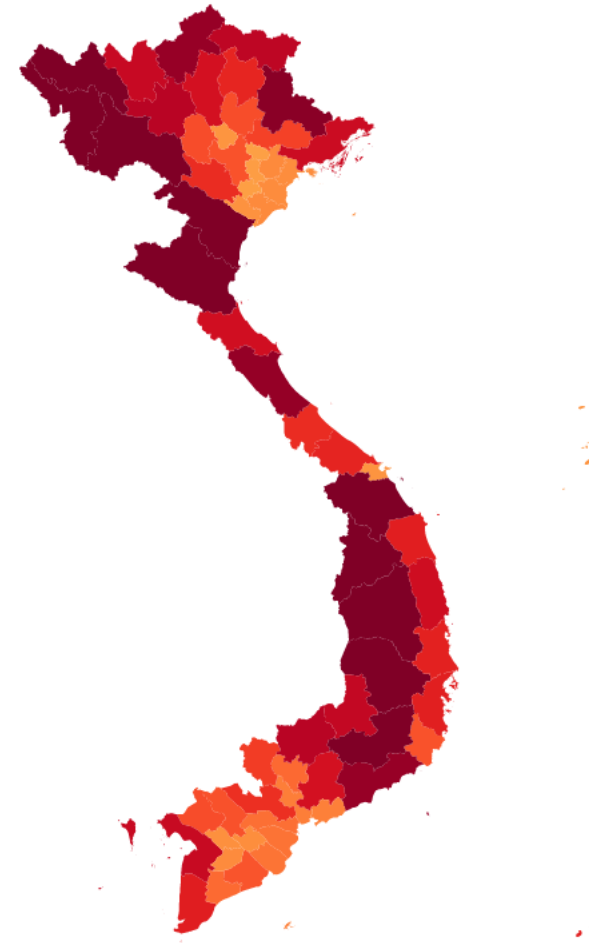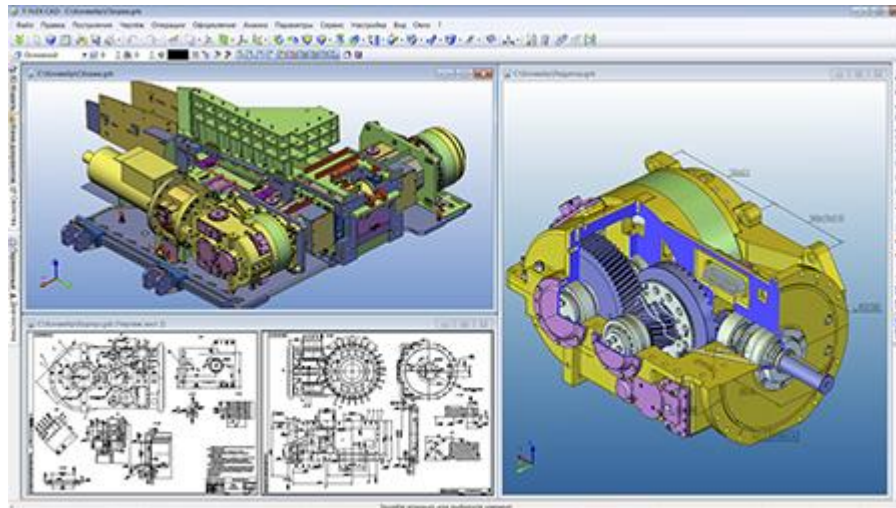
SciVis: Black Background

GEOMETRY

Shape of items

Explicit spatial positions

Points, lines, curves, surfaces, regions, volumnes

Important in Computer Graphics, CAD

Not a core visualization topic

ATTRIBUTE TYPES

Categorical (nominal)

- Compare equality

Ordered

- Ordinal

    - Great/Less than defined

Quantitative

- Arithmetic possible

    - e.g., length, weight

QUANTITATIVE: INTERVAL

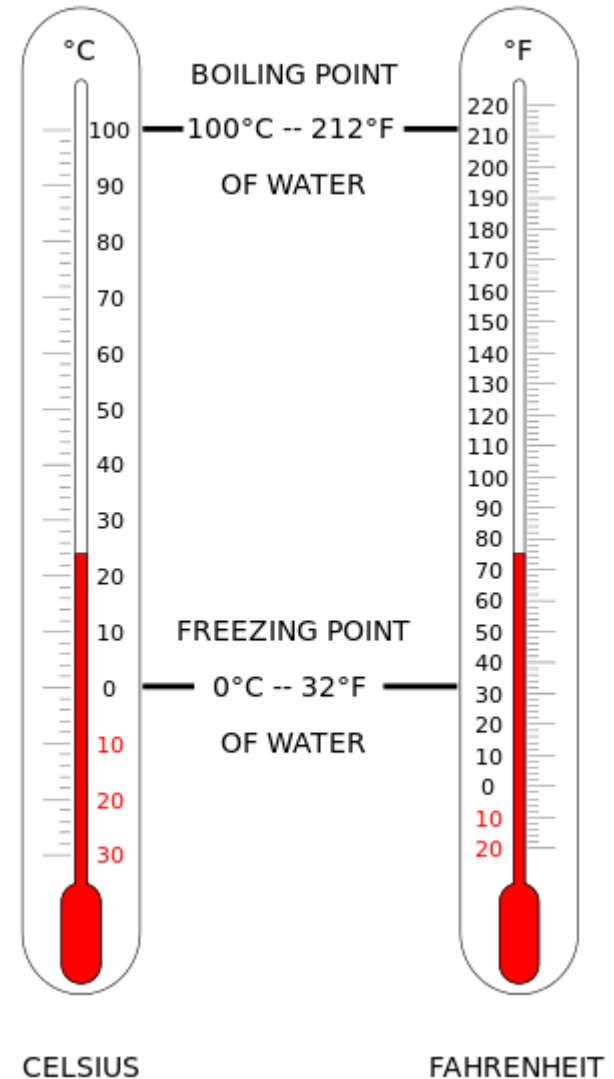May have **equal differences** between successive points on the scale

But the position of zero is arbitrary

Does zero mean none?

- Dates; Location;

Cannot compare directly (Fahrenheit vs Celsius)

Only differences can be compared (but ratios)

QUANTITATIVE: RATIO

The relative magnitudes of scores and the differences between them matter. The position of zero is fixed

Zero: there is nothing of the measured entity observed

Measurements: length, mass, age, weight, speed

Can measure ratios & proportions

e.g., Can say 2 times faster than the other

DATA TYPES

Nominal (categories, labels)

      Operations: =, ≠

Ordinal (ordered)

      Operations: =, ≠, >, <

Interval (location of zero arbitrary)

      Operations: =, ≠, >, <, +, - (distance)

Ratio (zero fixed)

      Operations: =, ≠, >, <, +, -, ×, ÷ (proportions)

QUIZ

What type of variable (nominal, ordinal, interval, or ratio) are the following:

1. Marathon race times

2. Major

3. Product rating
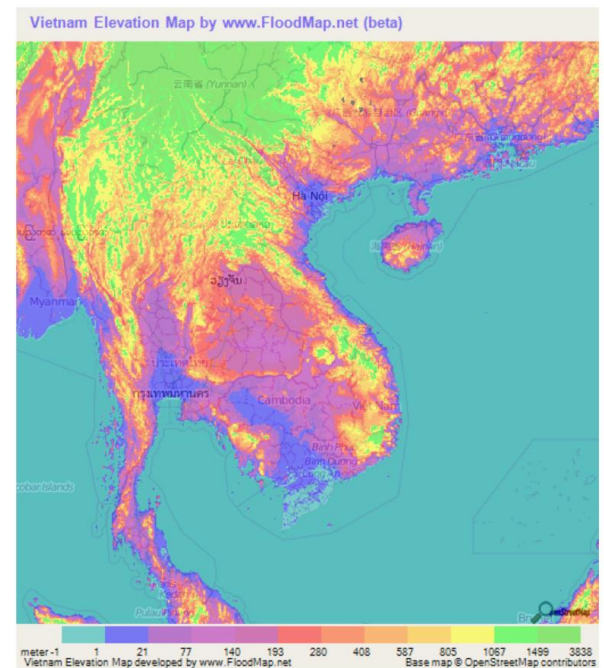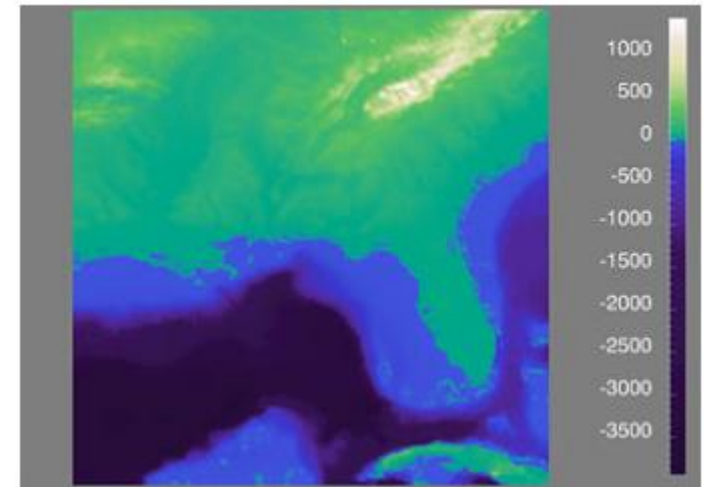
4. IQ score

5. Product name

SEQUENTIAL & DIVERGING DATA

Sequential

- Homogeneous from min to max

Diverging

- Two or multiple sequences that meet

- Elevation dataset: above sea level & below sea level

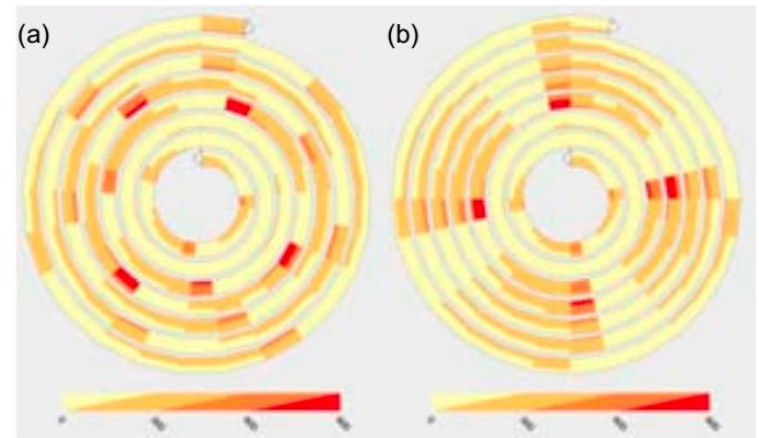- Temperature of water: below or above freezing / boiling

Vietnam Elevation Map by www.FloodMap.net (beta)
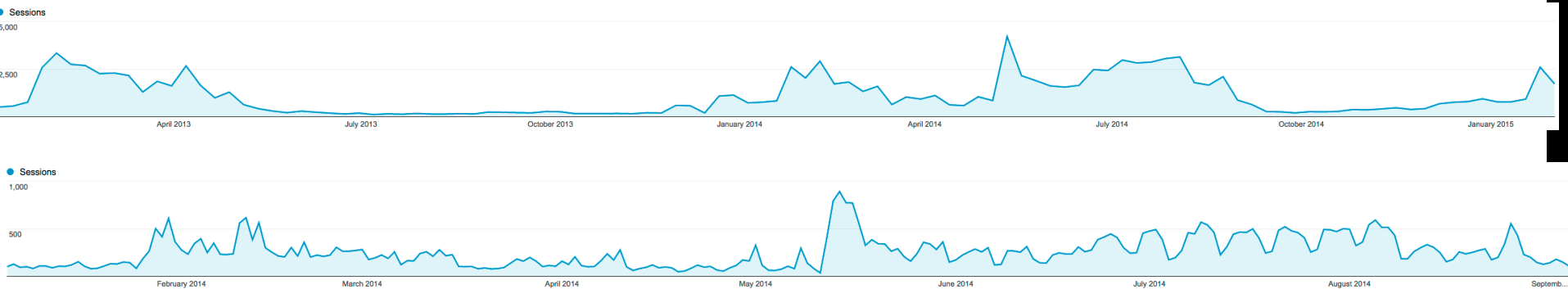
OTHER STRUCTURE

Cyclic data

- Time (hours, week, month, year)

Aggregation

- Might be patterns on multiple levels

Respiratory disease cases.
Left: 25 day pattern
Right: 28 day pattern
[Tominski 2008]

DATA VS CONCEPTUAL MODEL

Data model: low-level description of the data

Set with operations

Conceptual model: mental construction

Includes semantics, support reasoning

Data	Conceptual
1D floats	temperature
3D vector of floats	space

DATA VS CONCEPTUAL MODEL

From data model

  19.5, 29.0, -1 (floats)

Using conceptual model

  Temperature

To data type

  Continuous to 4 significant digits (Q)

  Hot, warm, cold (O)

  Burned vs Not burned (N)

COMBINATIONS, DERIVED DATA

Networks can have attributes

Attributes have hierarchies

Data types can be transformed