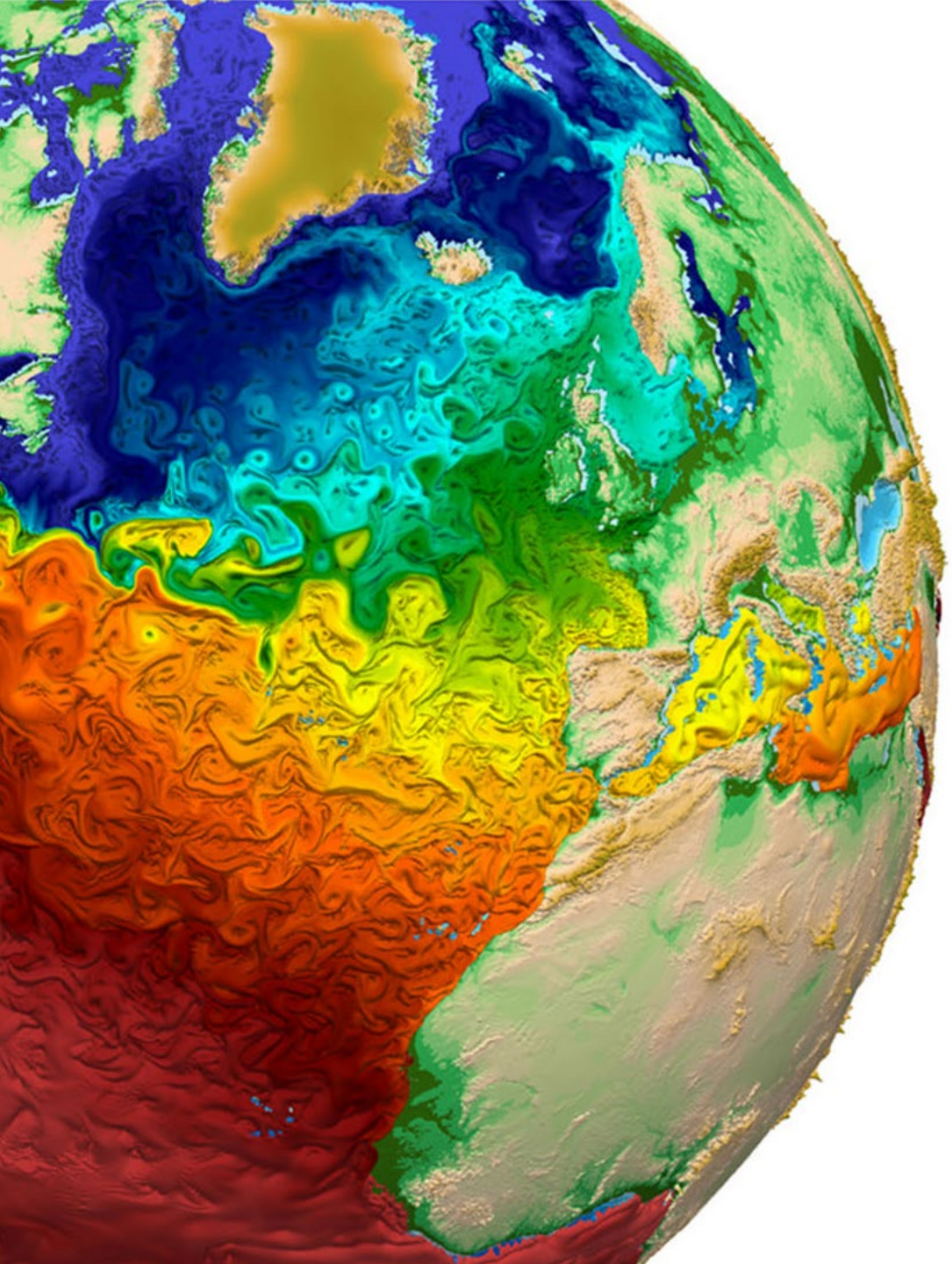


Data Science for People in a Hurry

A Data Taxonomy

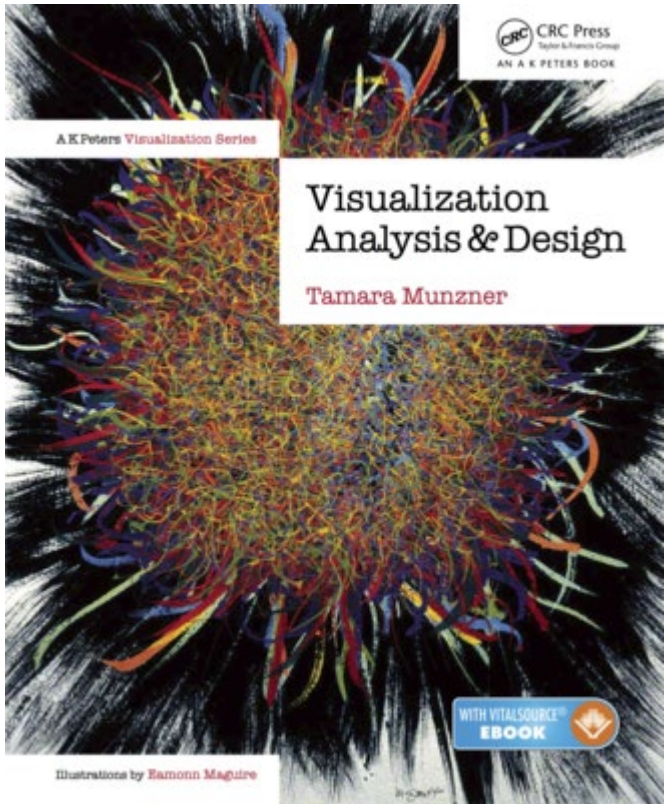
Scientific Visualization

Professor Eric Shaffer

ILLINOIS

Acknowledgment

Material for this lecture from Professor Tamara Munzner



https://www.cs.ubc.ca/~tmm/

ILLINOIS

Some Definitions

Semantics

- real-world meaning

Data Types:

- structural or mathematical interpretation of data
  - different from data types in programming!

```
shaffer1@BOROS:/mnt/c/Users/shaff/Downloads$ tail ClusterDataOriginal.txt
2015-04-30T23:50:00      6.3200e+01      0.0000e+00      0.0000e+00      0.0000e+00
   nan
2015-04-30T23:51:00      6.7714e+01      0.0000e+00      0.0000e+00      0.0000e+00
   nan
2015-04-30T23:52:00      1.2640e+02      0.0000e+00      0.0000e+00      1.0333e+00
   nan
2015-04-30T23:53:00      8.4643e+01      0.0000e+00      0.0000e+00      0.0000e+00
   nan
2015-04-30T23:54:00      1.1060e+02      0.0000e+00      0.0000e+00      0.0000e+00
   nan
2015-04-30T23:55:00      8.4643e+01      0.0000e+00      0.0000e+00      0.0000e+00
   nan
2015-04-30T23:56:00      7.9000e+01      0.0000e+00      0.0000e+00      0.0000e+00
   nan
2015-04-30T23:57:00      8.4643e+01      0.0000e+00      0.0000e+00      0.0000e+00
   nan
2015-04-30T23:58:00      1.7380e+02      0.0000e+00      0.0000e+00      0.0000e+00
   nan
2015-04-30T23:59:00      1.5082e+02      0.0000e+00      0.0000e+00      0.0000e+00
   nan
```

ILLINOIS

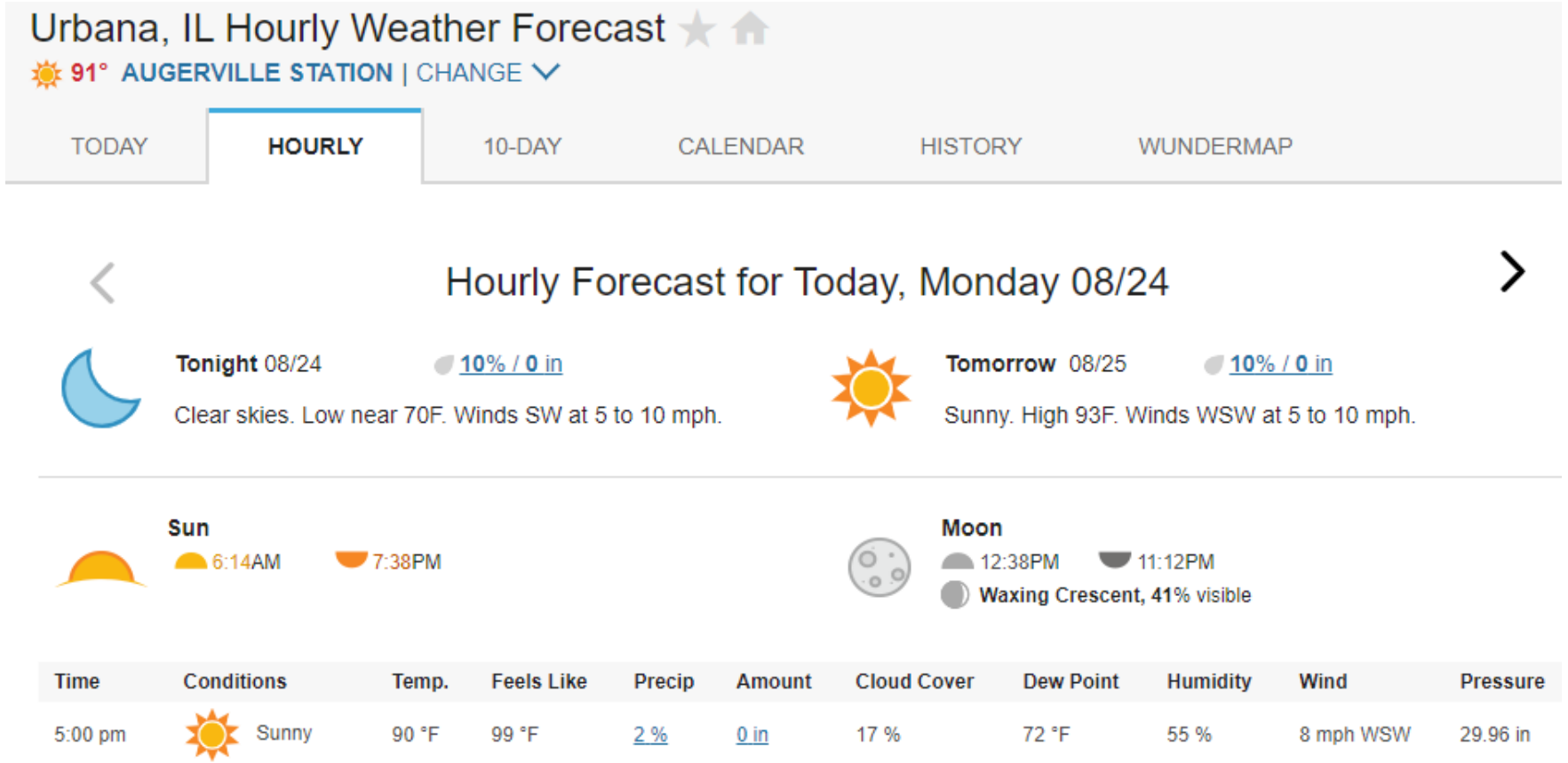
Items and Attributes

Item

- Individual entity

Attribute

- Property of an item
  - Measurement
  - Observation



Other Data Types

- Links
  - express relationship between two items
  - eg friendship on facebook, interaction between proteins
- Positions
  - spatial data: location in 2D or 3D
  - pixels in photo, voxels in MRI scan, latitude/longitude
- Grids
  - sampling strategy for continuous data

Dataset Types: Tables

Tables

Items

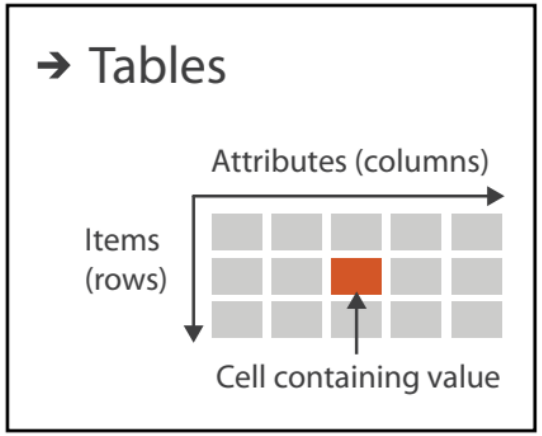
Attributes

flat table

- one item per row
- each column is attribute
- cell holds value

attributes: name, age, shirt size, fave fruit

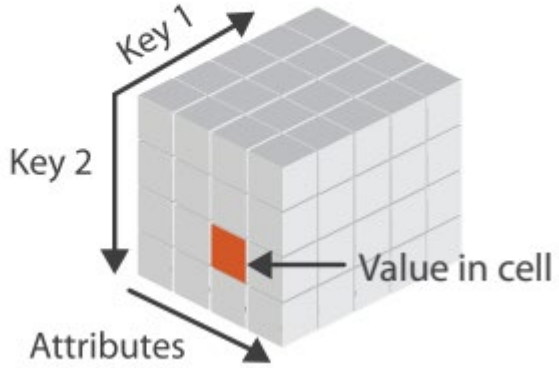
Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange



ILLINOIS

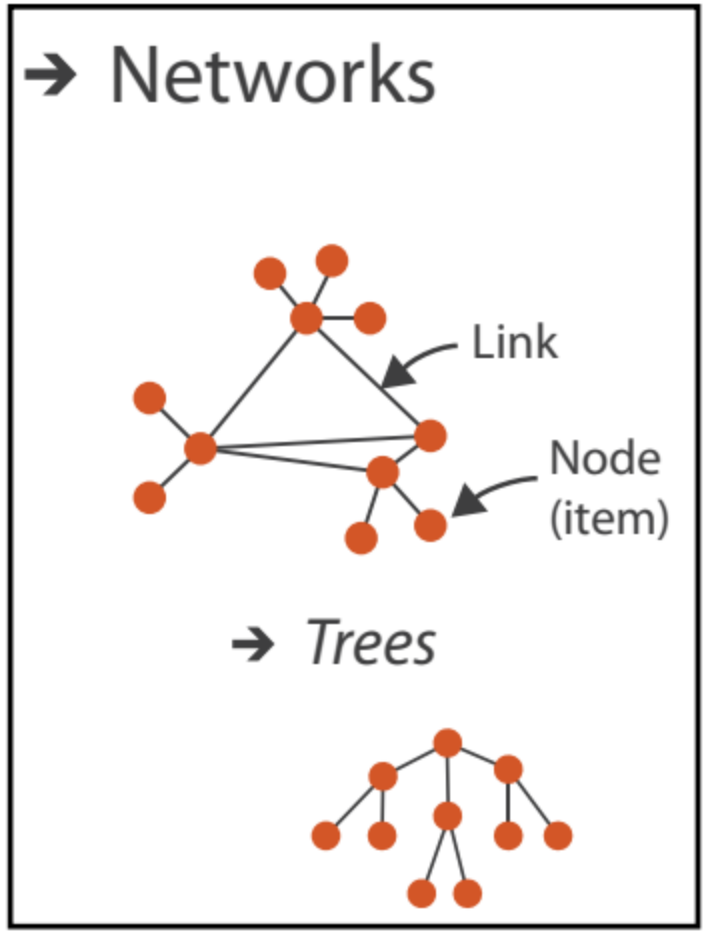
Dataset Types: Tables

➔ *Multidimensional Table*



- multidimensional tables
  - –indexing based on multiple keys

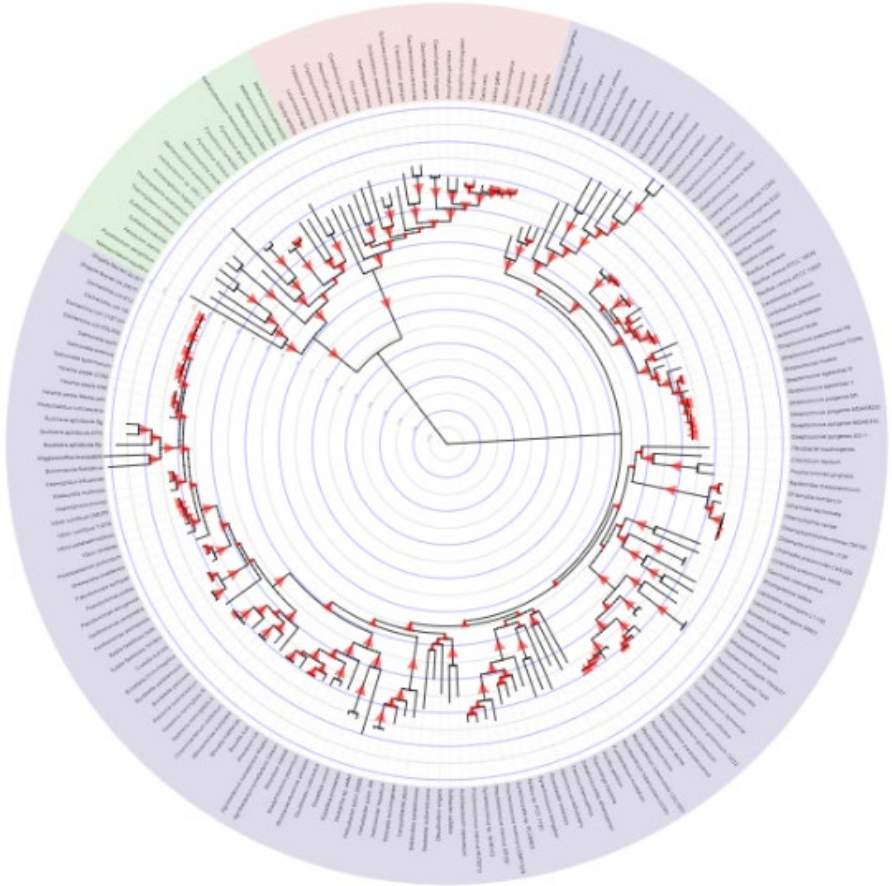
Dataset Types: Networks and Graphs

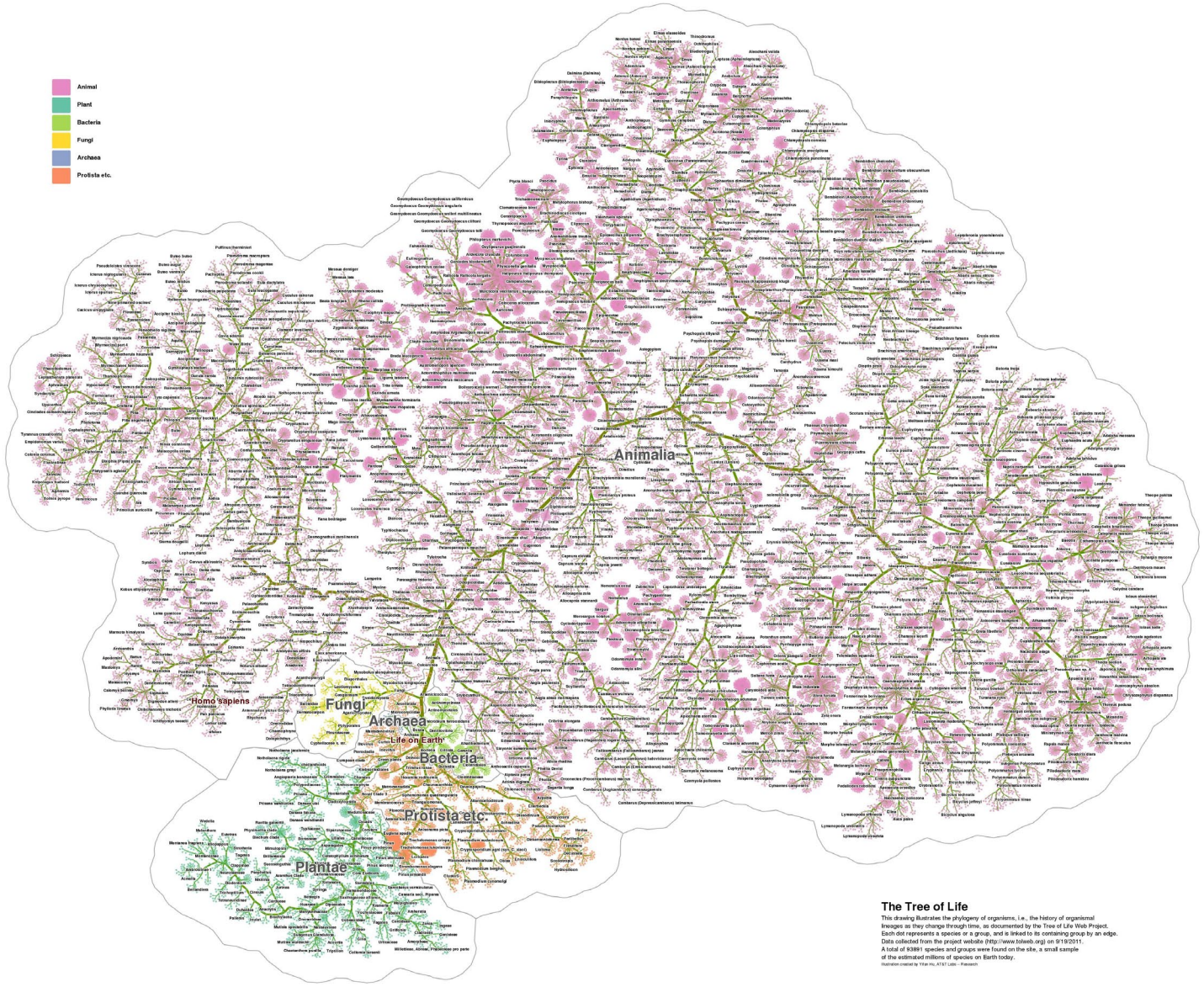


Networks & Trees
Items (nodes)
Links
Attributes

- network/graph
  - –nodes (vertices) connected by links (edges)
  - –tree is special case: no cycles
    - •often have roots and are directed

Visualizing Networks





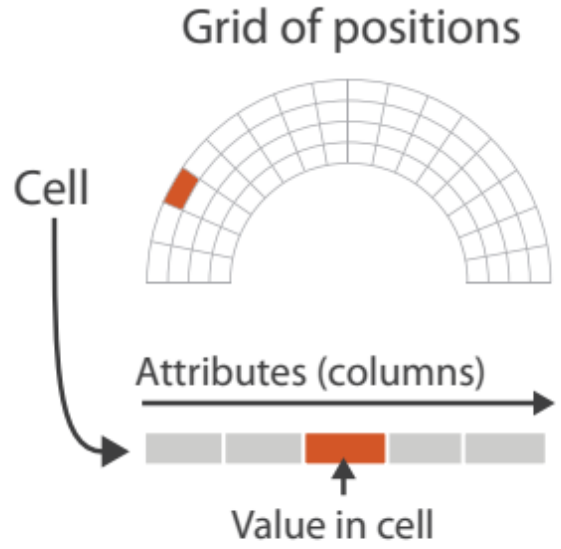
https://itol.embl.de/

A **phylogenetic tree** is a diagram that represents **evolutionary** relationships among organisms.

ILLINOIS

Dataset Types: Fields

- Spatial
- Fields (Continuous)



Fields

Grids

Positions

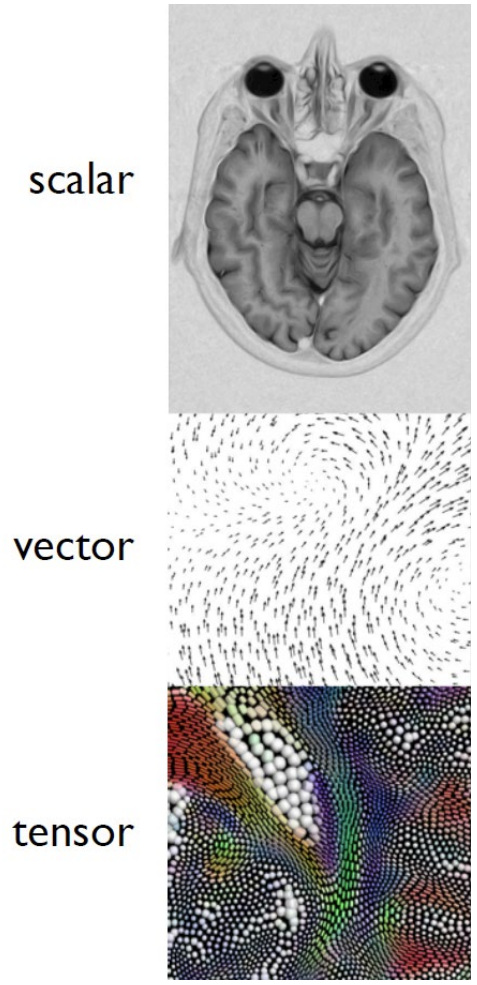
Attributes

- attribute values associated with cells
- cell contains value from continuous domain
  - eg temperature, pressure, wind velocity
- measured or simulated

ILLINOIS

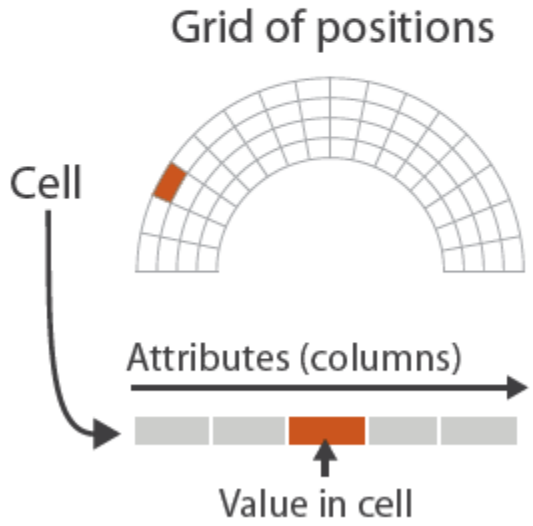
Spatial Fields

Field data



Dataset Types: Geometry

- Spatial
- Fields (Continuous)
- Geometry (Spatial)



29

Geometry

Items

Positions

Attribute Types

- which classes of values & measurements?
- categorical (nominal)
  - compare equality
  - no implicit ordering
- ordered
  - ordinal
    - less/greater than defined
  - quantitative
    - meaningful magnitude
    - arithmetic possible

A Taxonomy of Data

What?

Datasets

Data Types

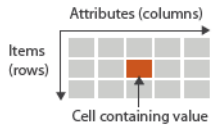
- Items
- Attributes
- Links
- Positions
- Grids

Data and Dataset Types

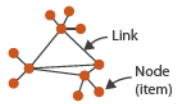
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

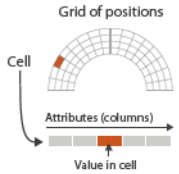
- Tables



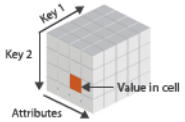
- Networks



- Fields (Continuous)



- *Multidimensional Table*



- *Trees*

- Geometry (Spatial)

Dataset Availability

- Static

- Dynamic

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

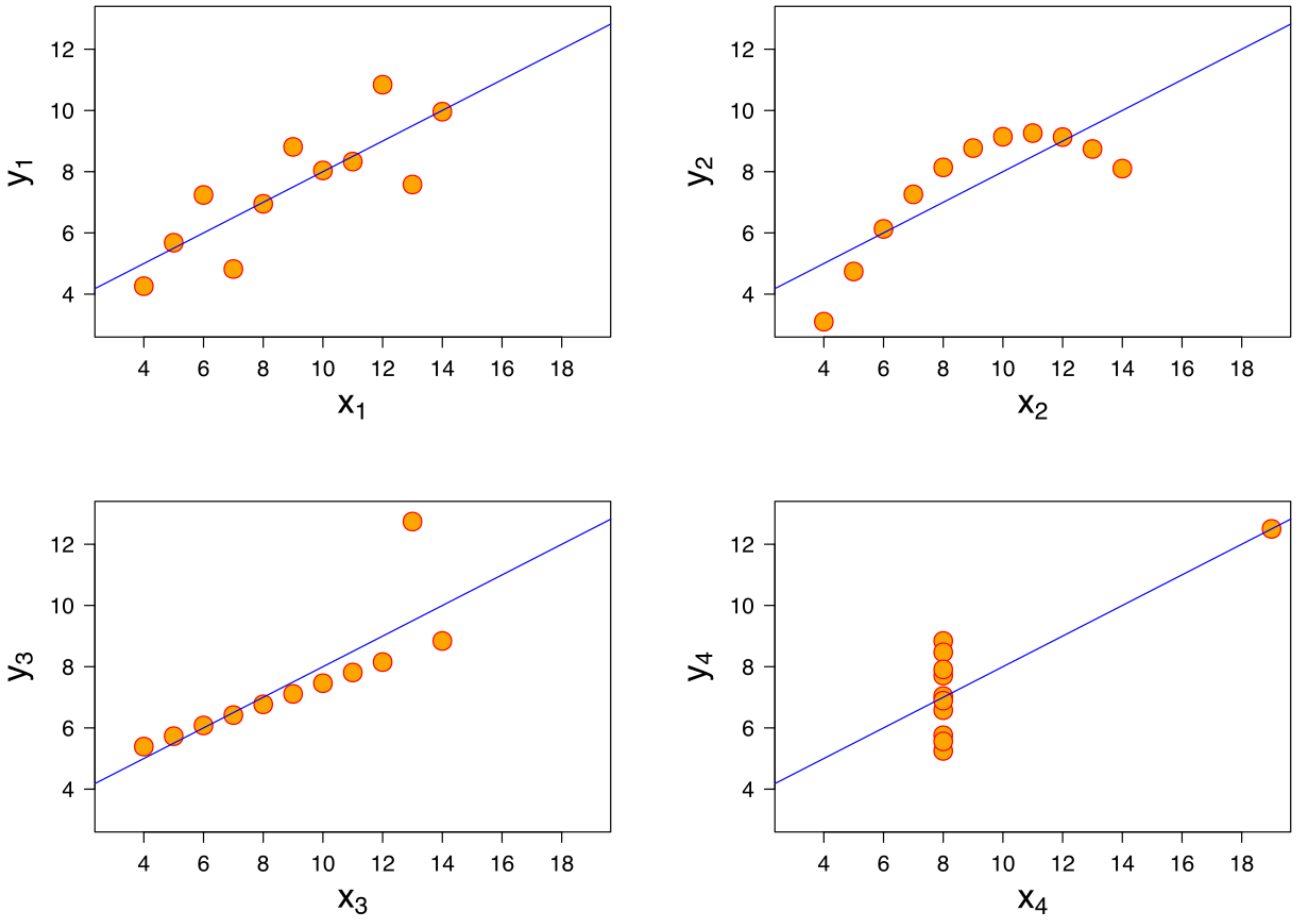
Ordering Direction

- Sequential

- Diverging

- Cyclic

Value of Visualization: Anscombe's Quartet



They were constructed in 1973 by the statistician Francis Anscombe to demonstrate both the importance of graphing data before analyzing it and the effect of outliers and other influential observations on statistical properties. He described the article as being intended to counter the impression among statisticians that "numerical calculations are exact, but graphs are rough."

-- Wikpedia

Mean and variance are the same for all 4 data sets

ILLINOIS