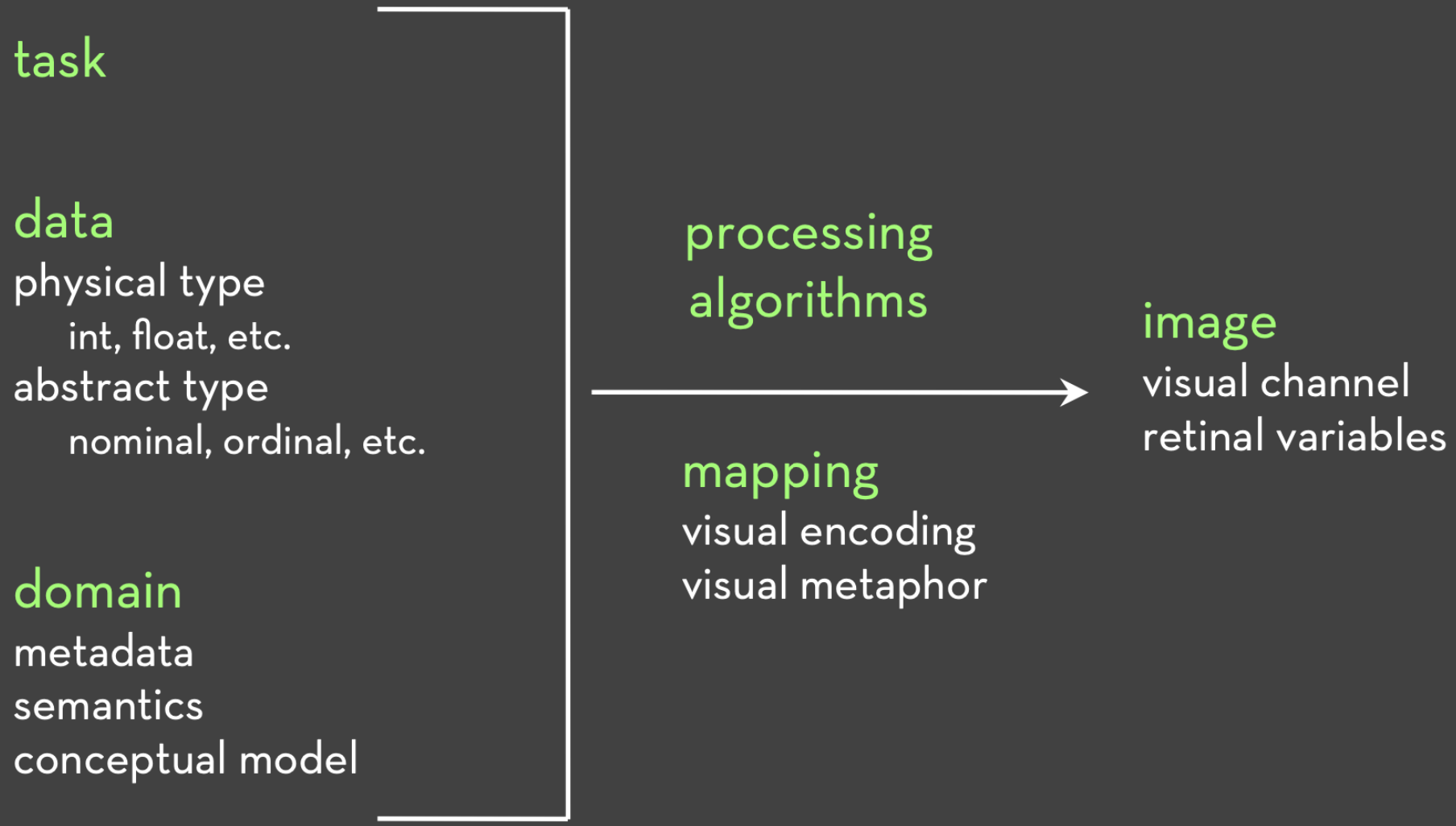
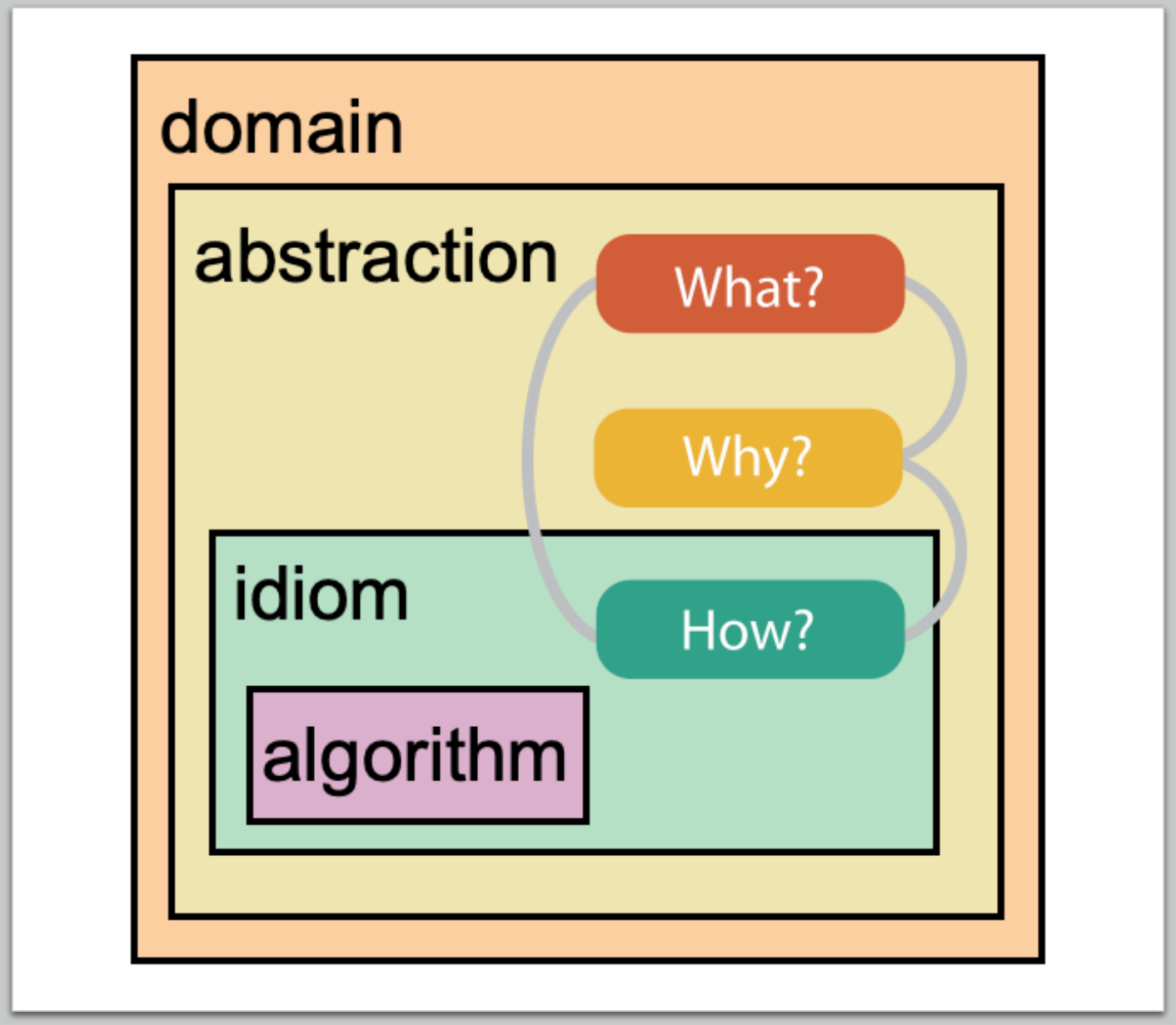

I JUST LOVE DATA.
DATA'S MY FAVORITE.
memegenerator.net

The Big Picture

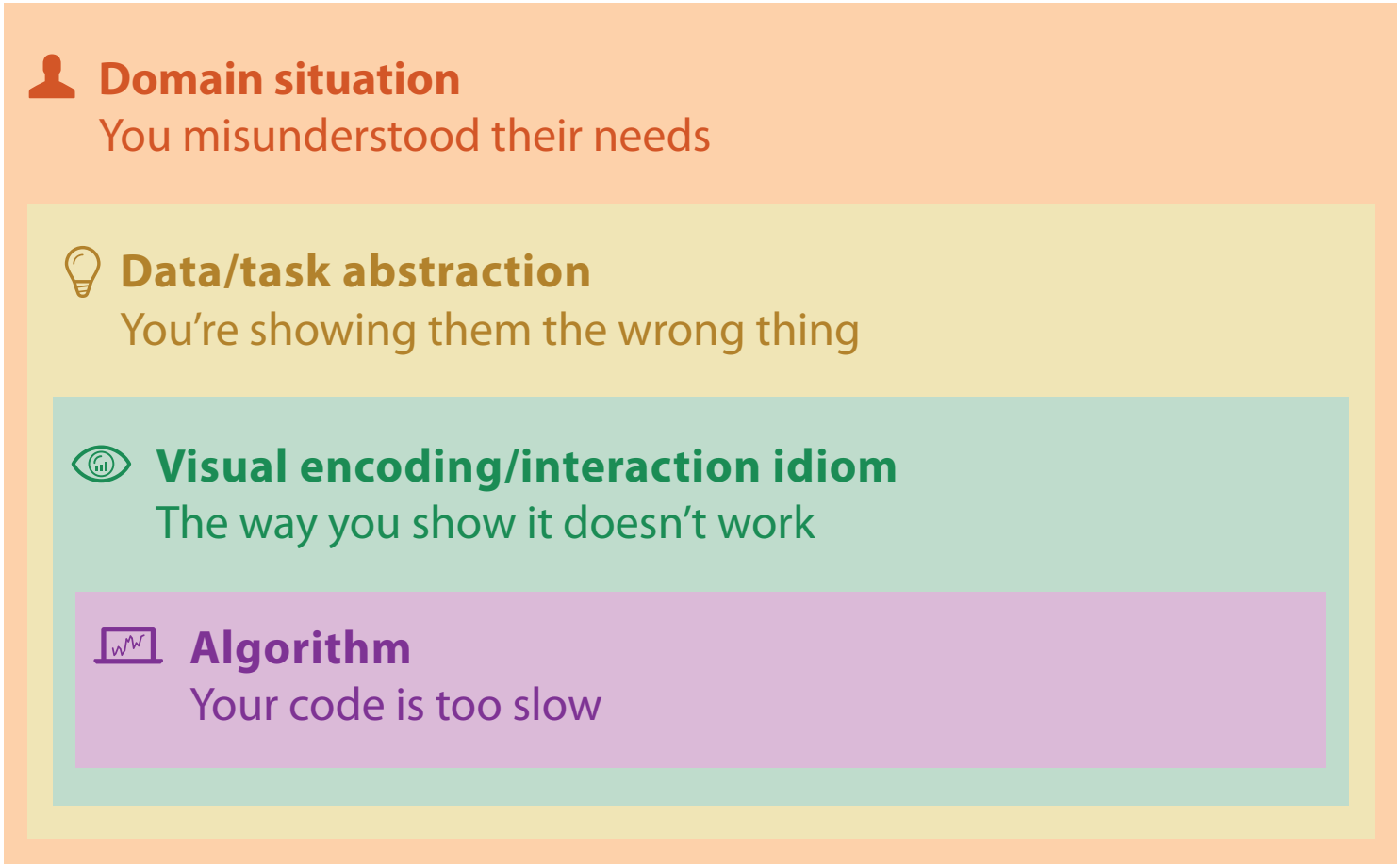


Munzner's Four levels of visualization design

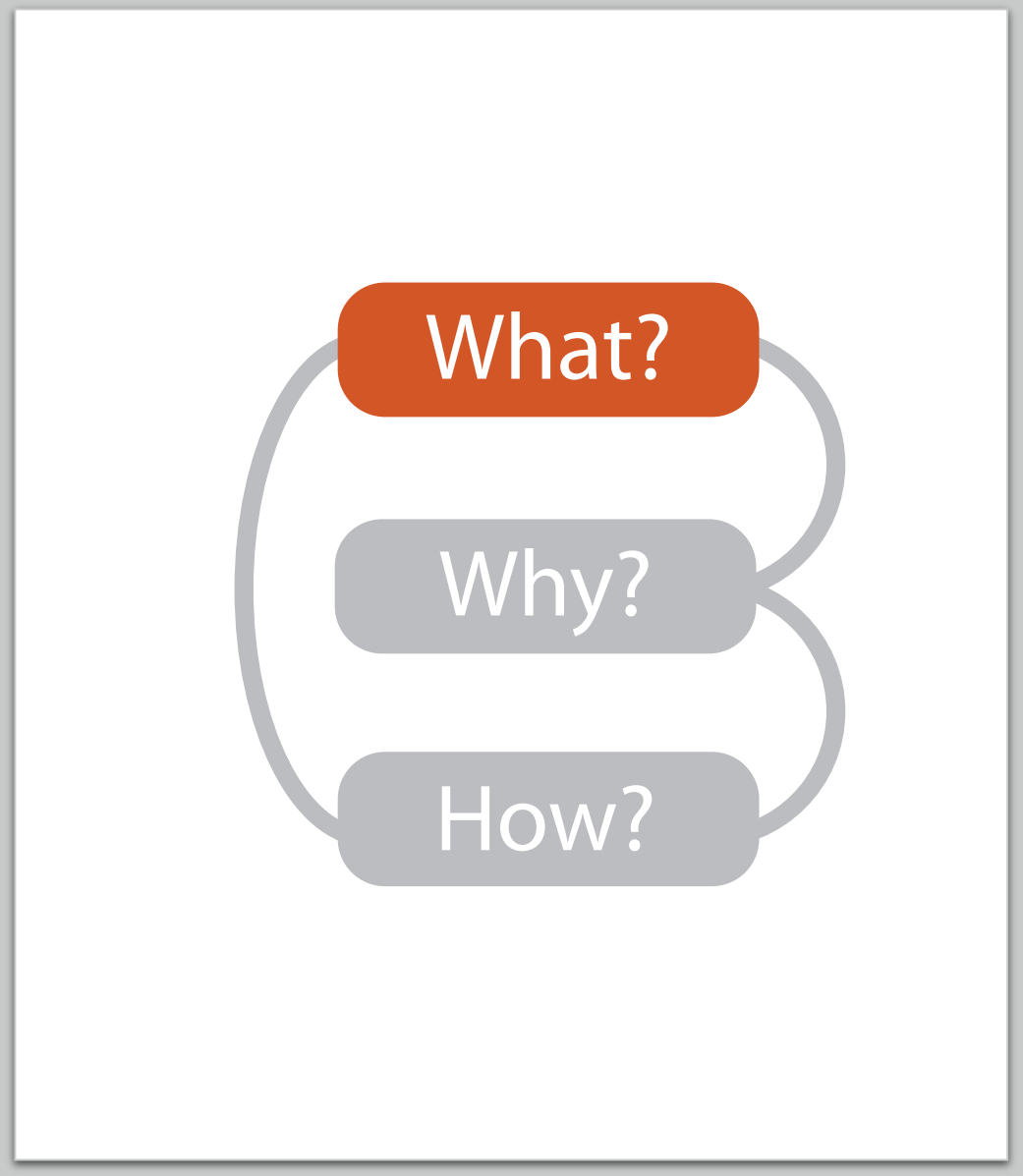
- *domain situation*
  - who are the target users?
- *abstraction*
  - translate from specifics of domain to vocabulary of visualization
    - **what** is shown? **data** abstraction
    - **why** is the user looking at it? **task** abstraction
      - often must transform data, guided by task
- *idiom*
  - **how** is it shown?
    - **visual encoding** idiom: how to draw
    - **interaction** idiom: how to manipulate
- *algorithm*
  - efficient computation



Common Pitfalls Along the Path



Data Abstraction



What Does Data Mean?

14, 2.6, 30, 30, 15, 100001

- What does this sequence of six numbers mean?
  - two points far from each other in 3D space?
  - two points close to each other in 2D space, with 15 links between them, and a weight of 100001 for the link?
  - something else??

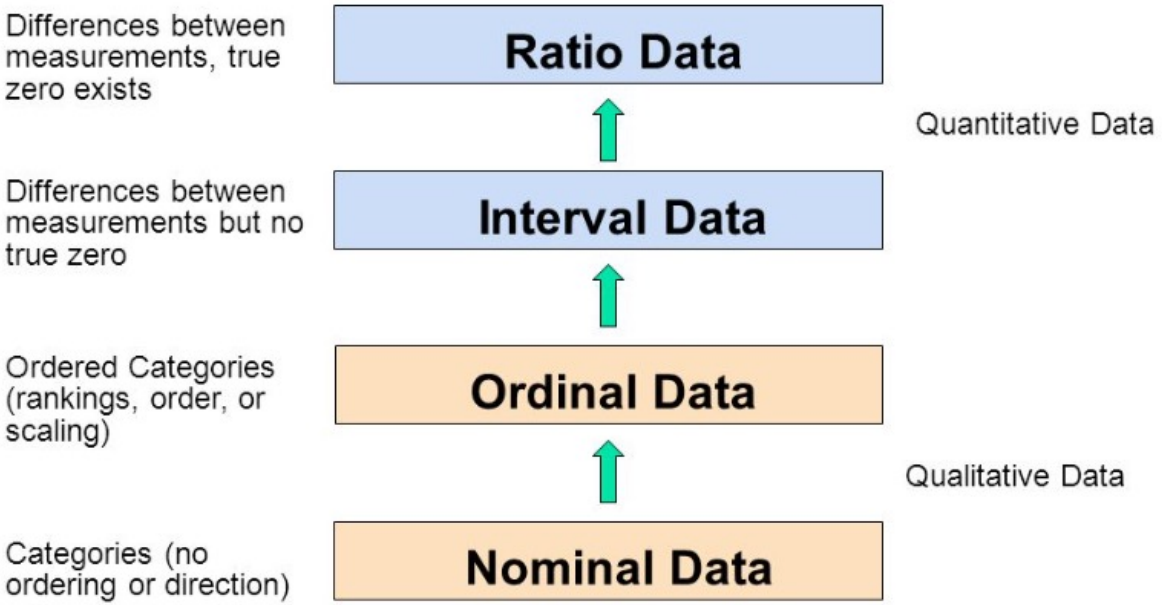
Basil, 7, S, Pear

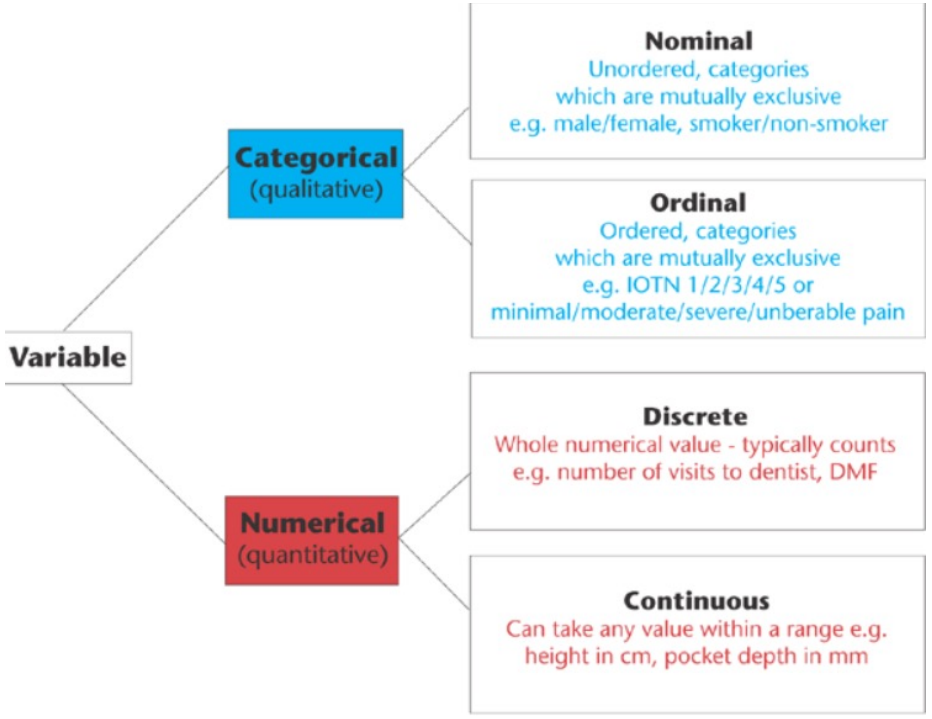
- What about this data?
  - food shipment of produce (basil & pear) arrived in satisfactory condition on 7th day of month
  - Basil Point neighborhood of city had 7 inches of snow cleared by the Pear Creek Limited snow removal service
  - lab rat Basil made 7 attempts to find way through south section of maze, these trials used pear as reward food

- semantics: real-world meaning
- data types: structural or mathematical interpretation of data

What is “Data”

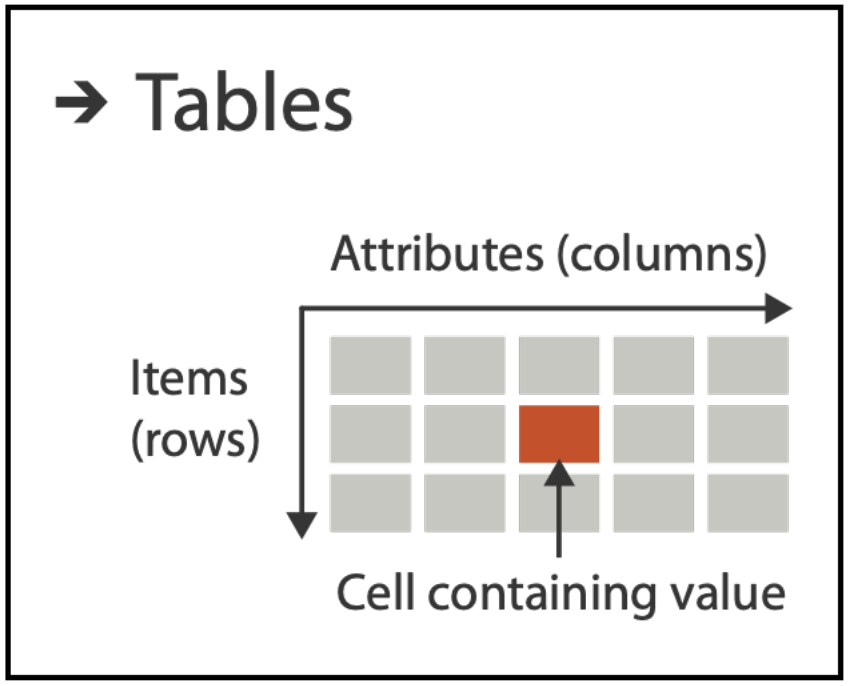
- Physical type in memory
  - float64
  - uint16
  - Etc…
- Abstract type
  - Nominal
  - Ordinal
  - Interval
  - Ratio





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

Dataset Types



- Tables in Python are well represented using the pandas package.
- A flat table has
  - One item/record per row
  - Each column is an attribute
  - Cell holds value for item-attribute pair

Attributes / Features / Variables

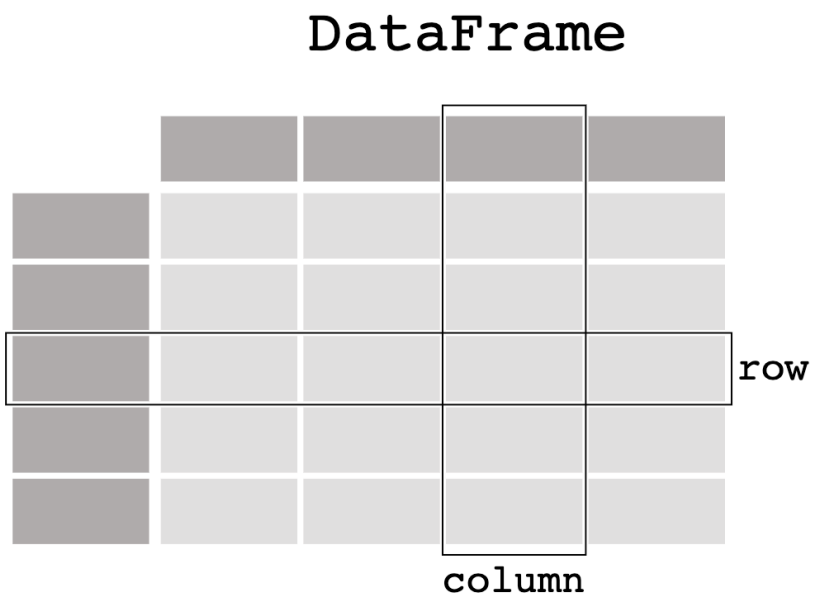
Items / Objects / Records / Observations

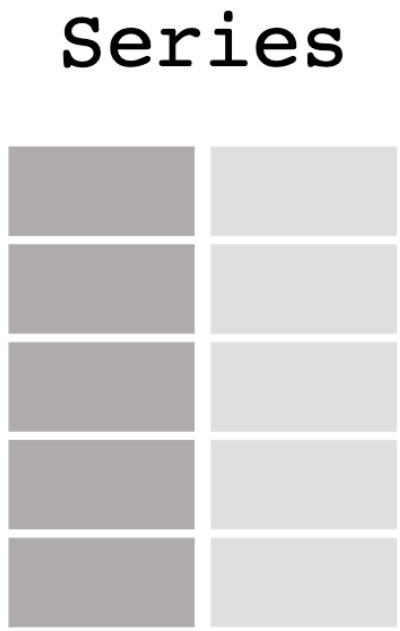
Data Points

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Pandas DataFrame

- 2-dimensional data structure that can store data of different types (including characters, integers, floating point values, categorical data and more) in columns. It is similar to a spreadsheet, a SQL table or the data.frame in R.
- Each column in a DataFrame is a Series
- If the data is numeric, it uses numpy under the hood and can be used for matrix math





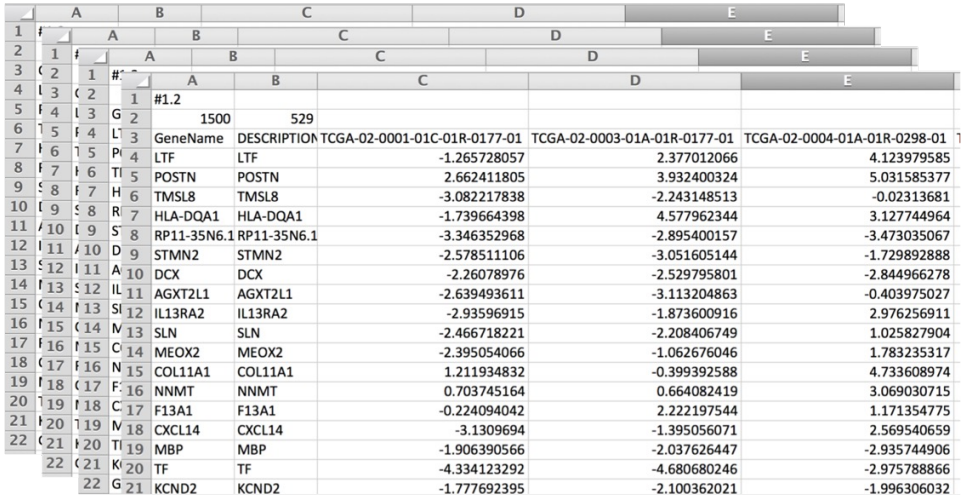
Multidimensional Tables

Tables

Items

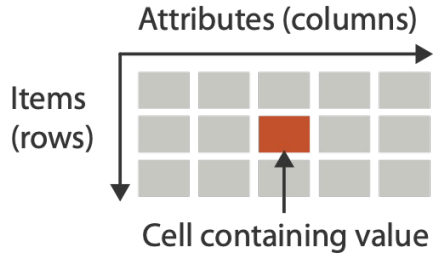
Attributes

- multidimensional tables
  - –indexing based on multiple keys
    - •eg genes, patients

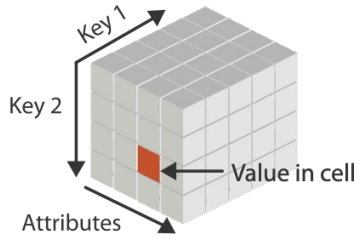


	A	B	C	D	E
1	#1.2				
2	1500	529			
3	GeneName	DESCRIPTION	TCGA-02-0001-01C-01R-0177-01	TCGA-02-0003-01A-01R-0177-01	TCGA-02-0004-01A-01R-0298-01
4	LTF	LTF	-1.265728057	2.377012066	4.123979585
5	POSTN	POSTN	2.662411805	3.932400324	5.031585377
6	TMSL8	TMSL8	-3.082217838	-2.243148513	-0.02313681
7	HLA-DQA1	HLA-DQA1	-1.739664398	4.577962344	3.127744964
8	RP11-35N6.1	RP11-35N6.1	-3.346352968	-2.895400157	-3.473035067
9	STMN2	STMN2	-2.578511106	-3.051605144	-1.729892888
10	DCX	DCX	-2.26078976	-2.529795801	-2.844966278
11	AGXT2L1	AGXT2L1	-2.639493611	-3.113204863	-0.403975027
12	IL13RA2	IL13RA2	-2.93596915	-1.873600916	2.976256911
13	SLN	SLN	-2.466718221	-2.208406749	1.025827904
14	MEOX2	MEOX2	-2.395054066	-1.062676046	1.783235317
15	COL11A1	COL11A1	1.211934832	-0.399392588	4.733608974
16	NNMT	NNMT	0.703745164	0.664082419	3.069030715
17	F13A1	F13A1	-0.224094042	2.222197544	1.171354775
18	CXCL14	CXCL14	-3.1309694	-1.395056071	2.569540659
19	MBP	MBP	-1.906390566	-2.037626447	-2.935744906
20	TF	TF	-4.334123292	-4.680680246	-2.975788866
21	KCND2	KCND2	-1.777692395	-2.100362021	-1.996306032

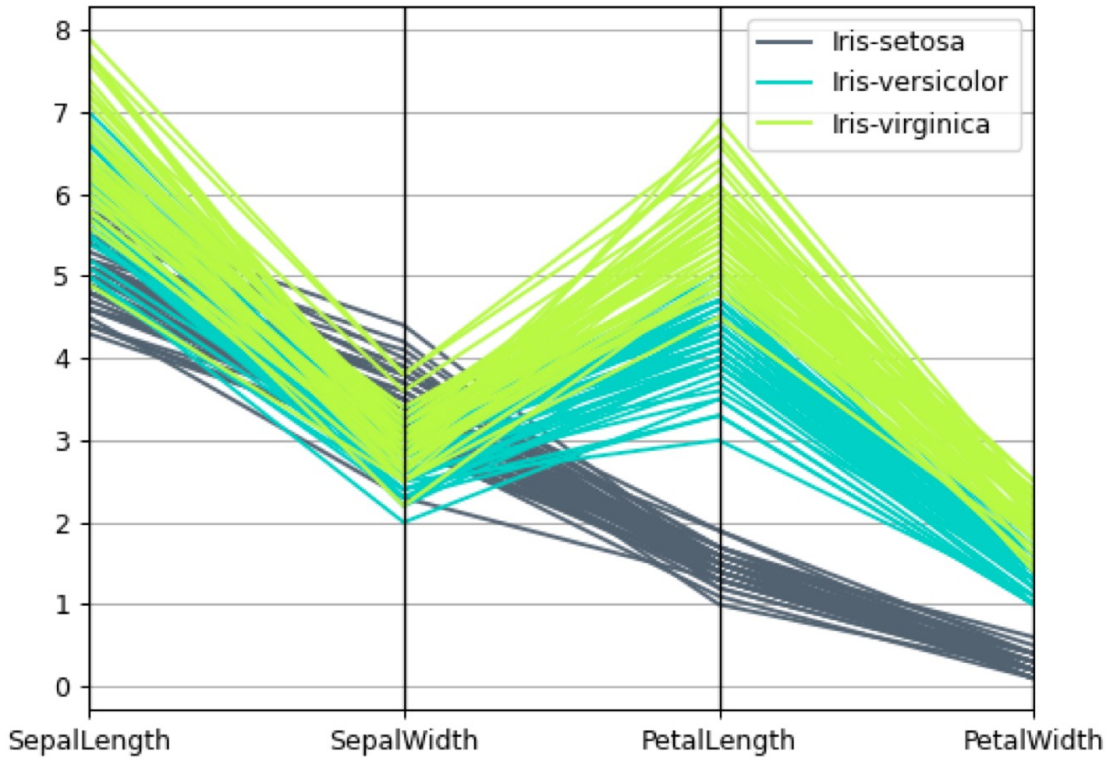
➔ Tables



➔ *Multidimensional Table*



Visualizing a Table (all at once)



https://pandas.pydata.org/docs/reference/api/pandas.plotting.parallel_coordinates.html

Networks / Graphs

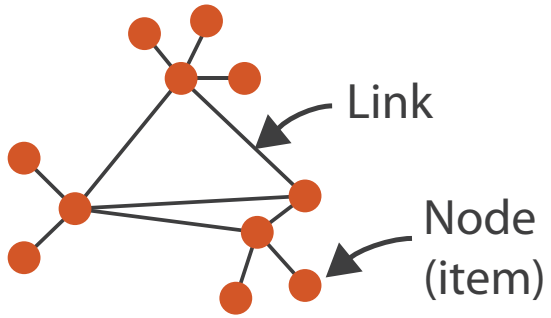
Networks & Trees

Items (nodes)

Links

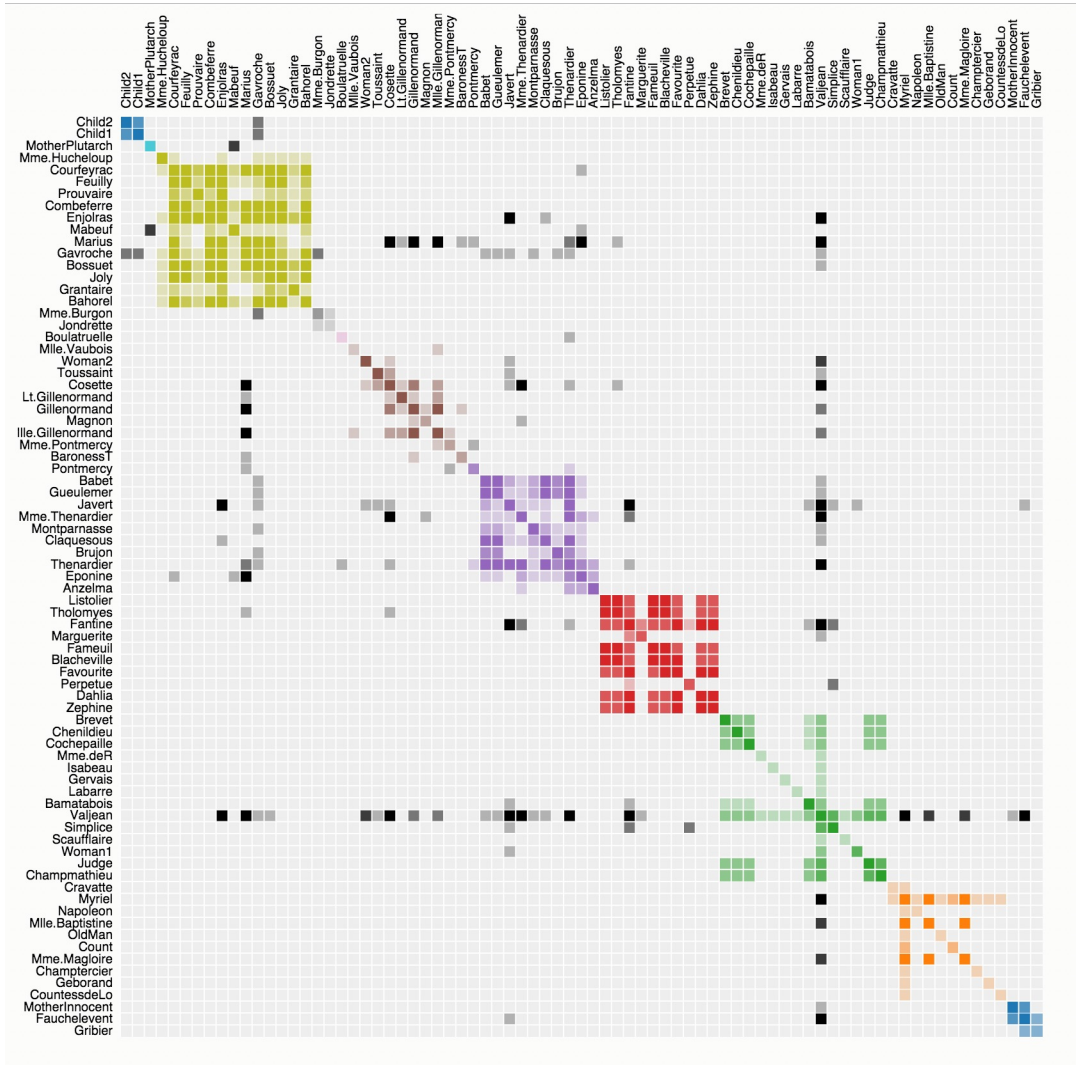
Attributes

➔ Networks



- Nodes (vertices) connected by links (edges)
- Tree - is special case: no cycles
  - often have roots and are directed

Visualizing Networks



Fields

Fields

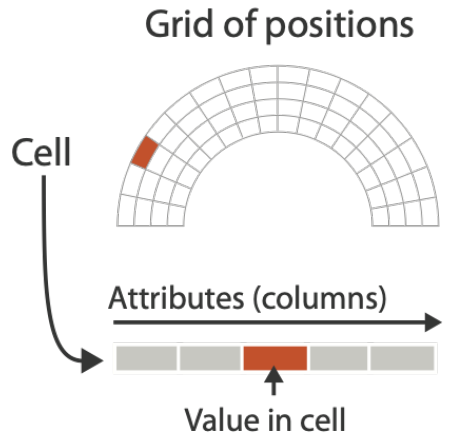
Grids

Positions

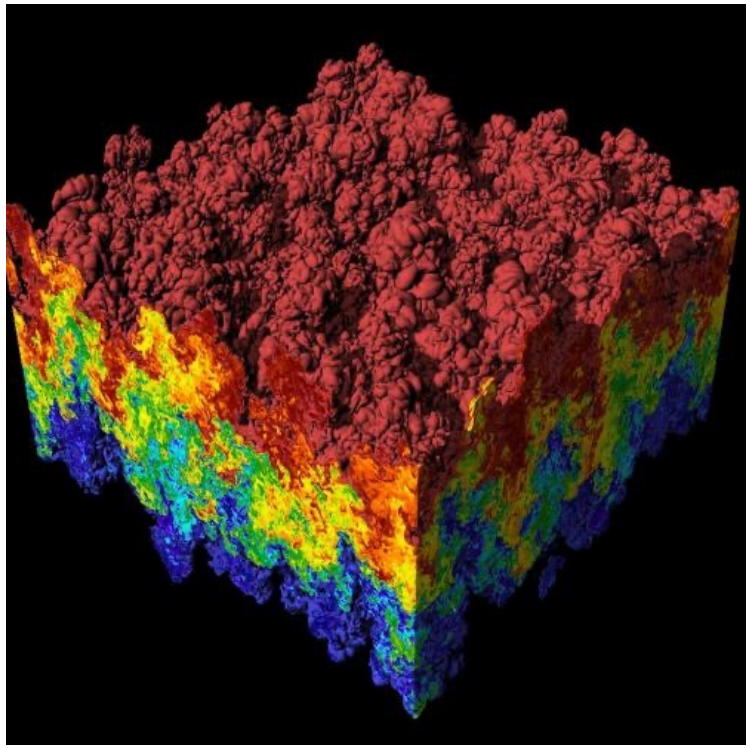
Attributes

➔ Spatial

➔ Fields (Continuous)



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



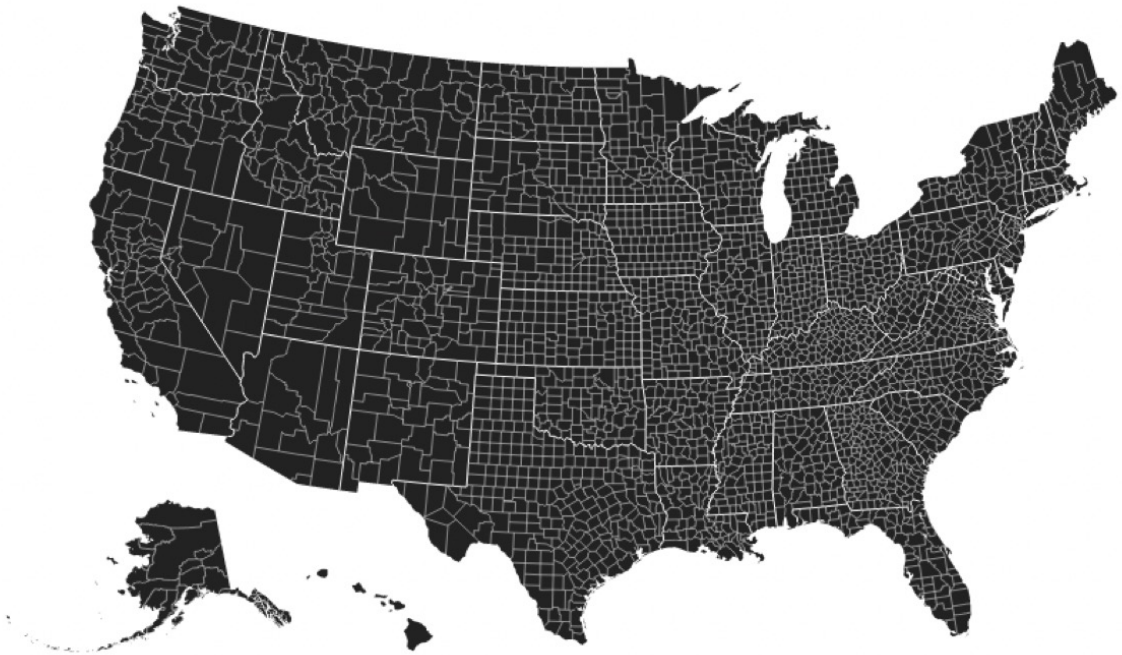
Geometry

Geometry

Items

Positions

➔ Geometry (Spatial)

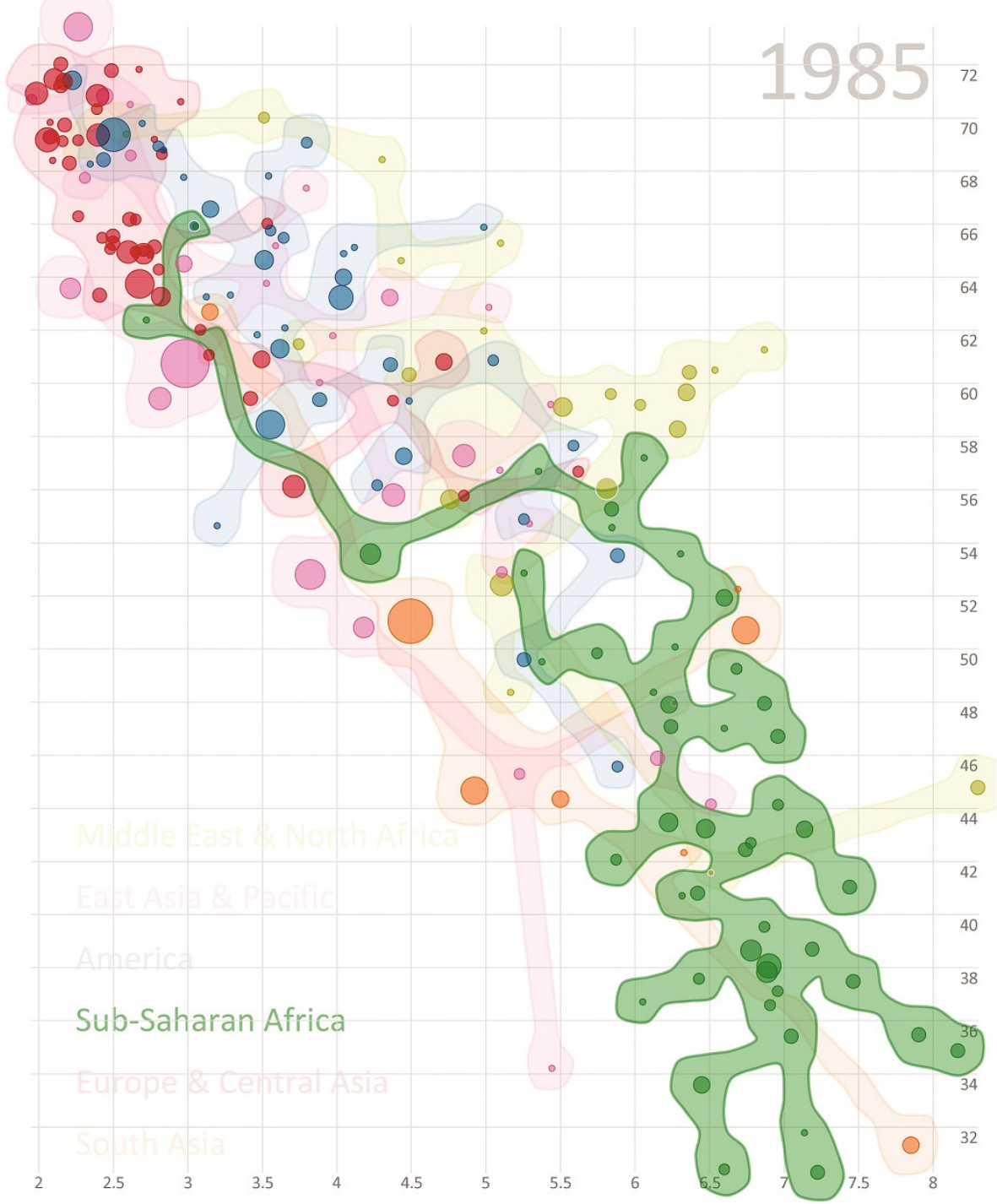


Clusters

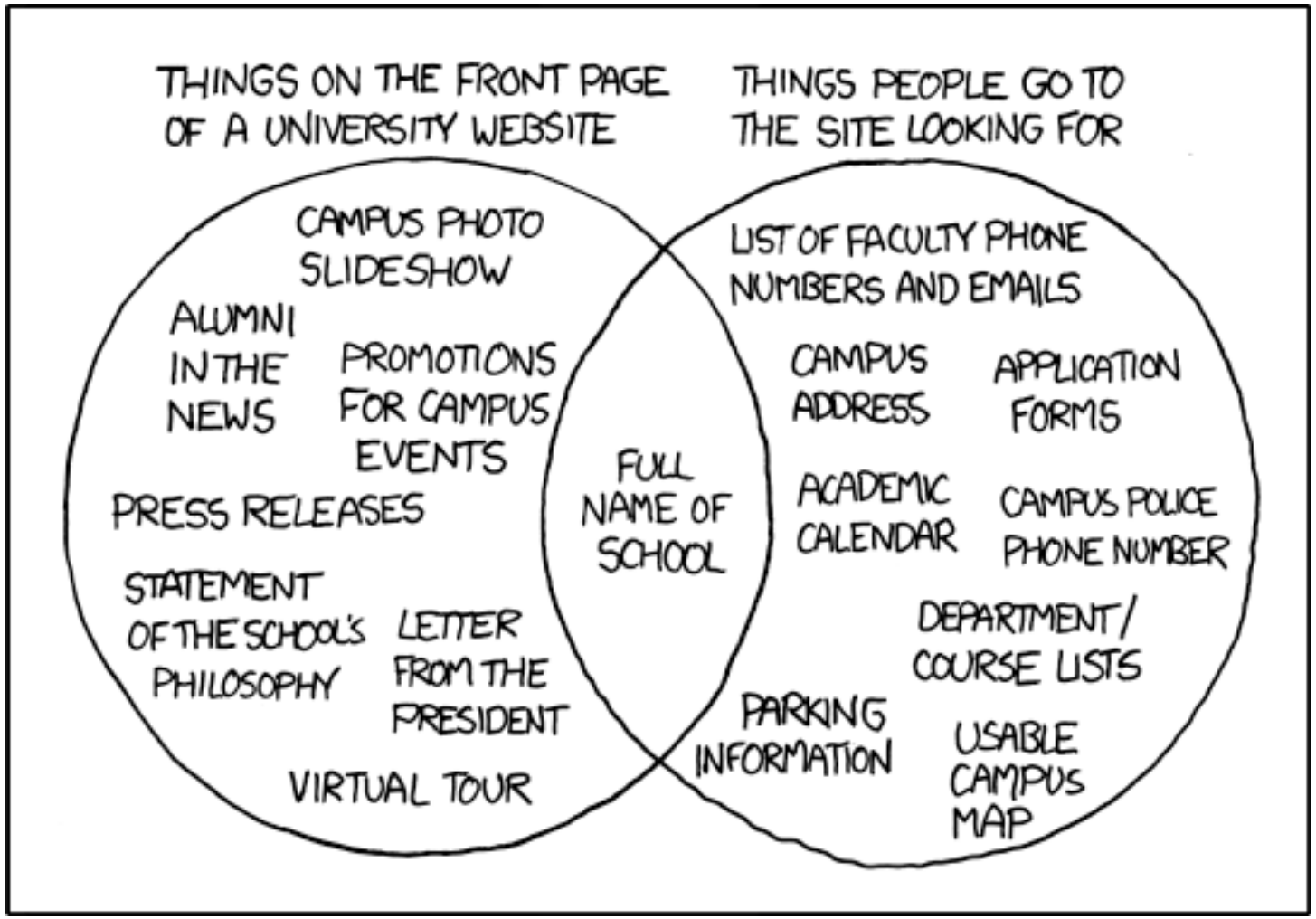
Clusters, Sets, Lists

Items

- how we group items
- sets
  - unique items, unordered
- lists
  - ordered, duplicates possible
- clusters
  - groups of similar items



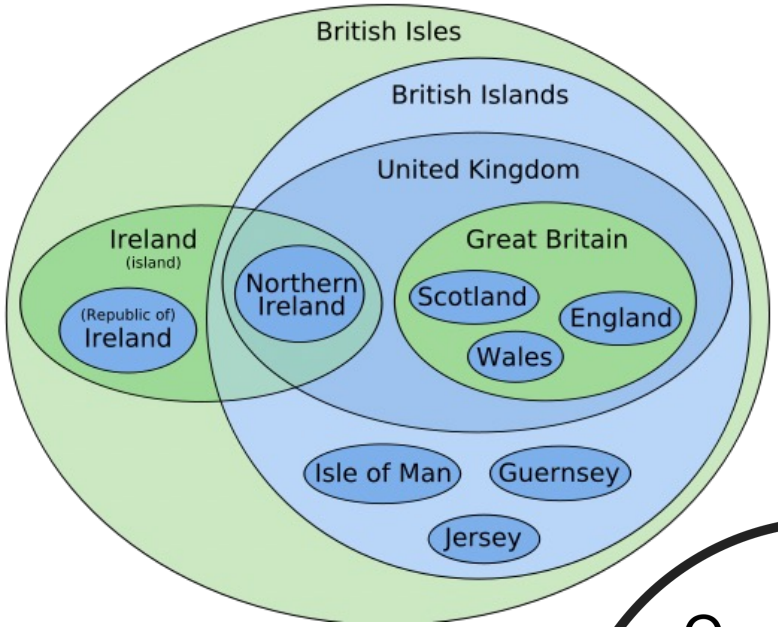
Visualizing Sets



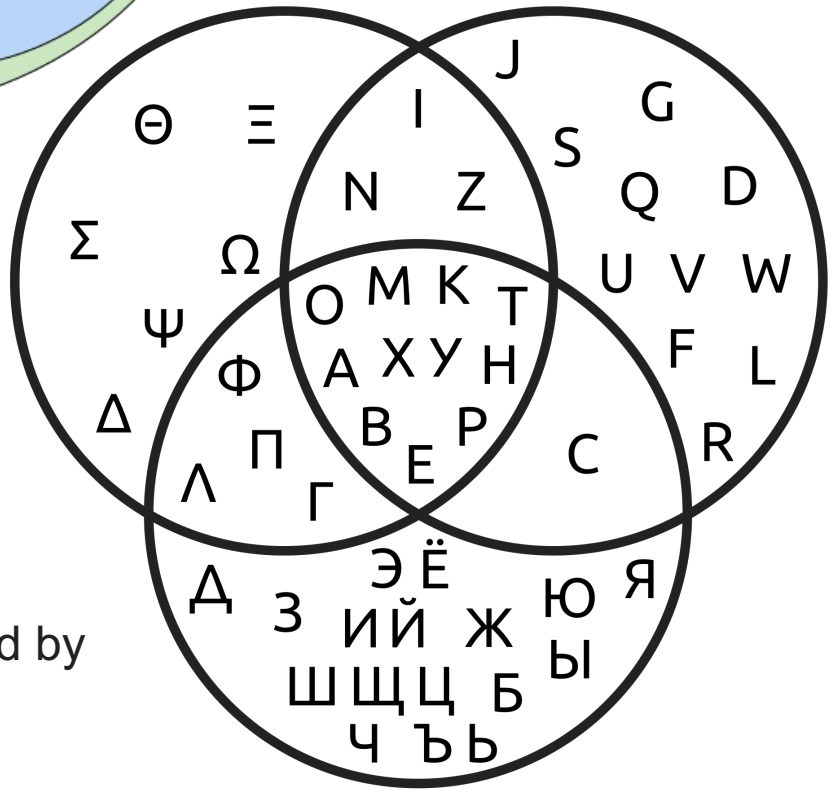
https://xkcd.com/773/

Visualizing Sets

- **Euler Diagram**
  - Shows Logical relations
  - May omit empty intersections



- **Venn Diagram**
  - Shows all possible logical relations between sets



Venn diagram showing the uppercase glyphs shared by the Greek, Latin, and Cyrillic alphabets

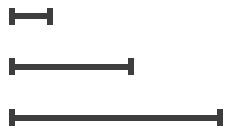
Other Data Concerns

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*



Ordering Direction

- Sequential

- Diverging

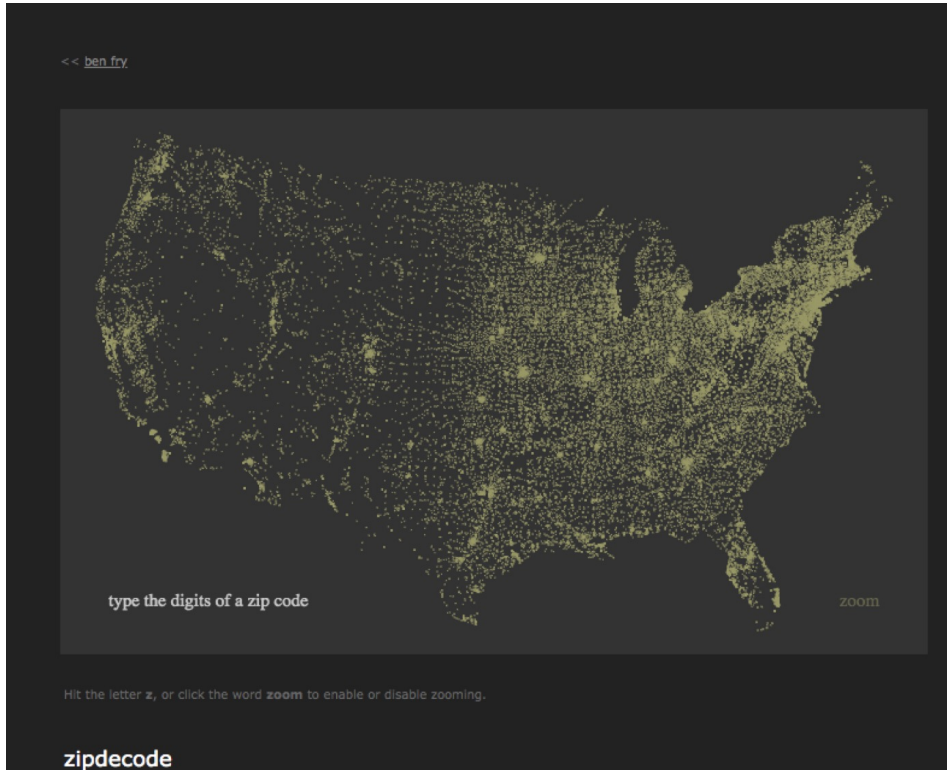
- Cyclic

Hierarchical Data

- Multi-level structure
  - space
  - time
  - Etc.

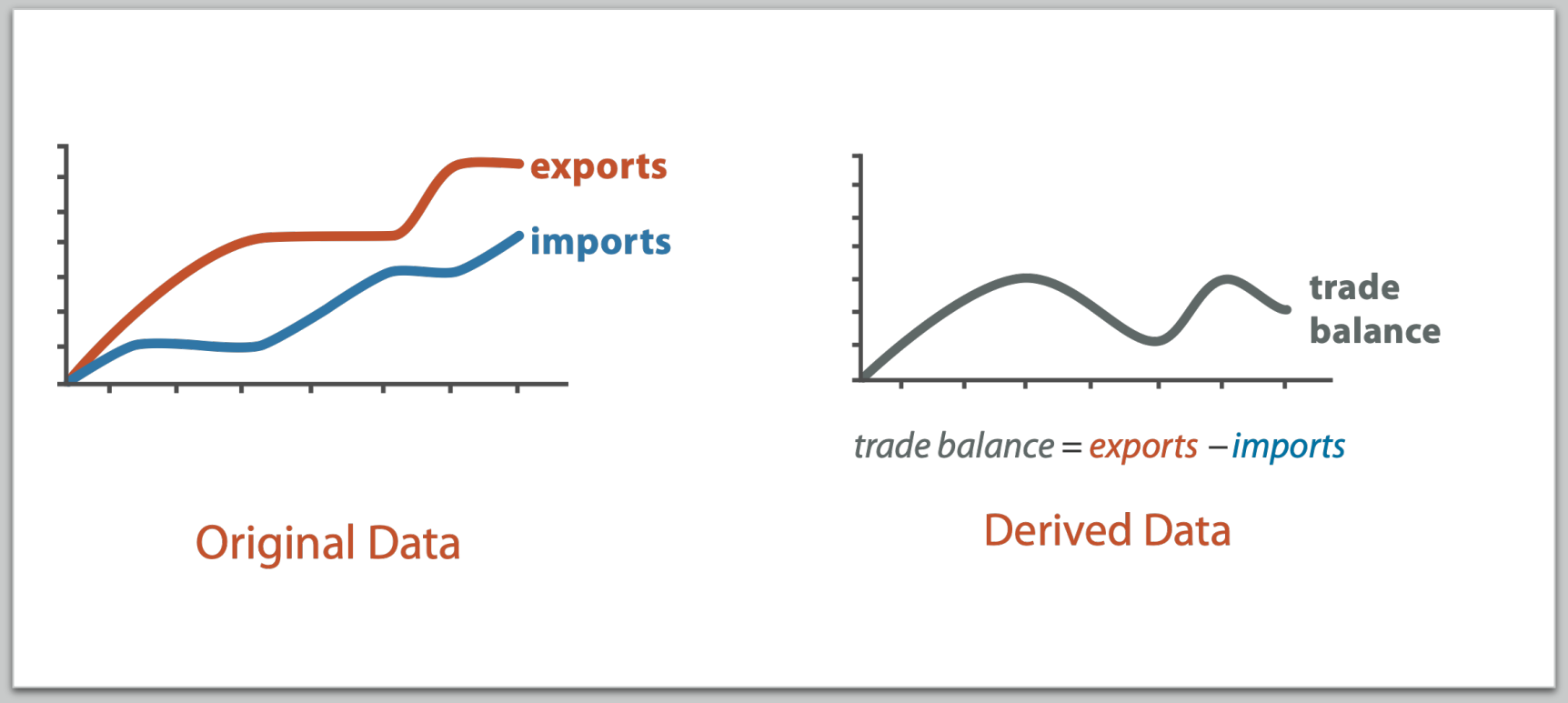
Example:

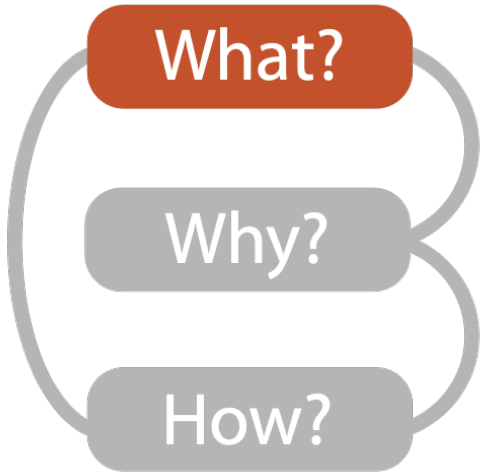
- https://benfry.com/zipdecode/



Derived Data

- Derived attribute: compute from original data
  - simple change of type
  - acquire additional data
  - complex transformation





What?

Datasets

Data Types

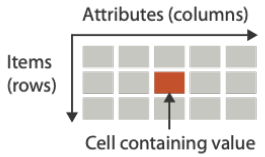
- Items
- Attributes
- Links
- Positions
- Grids

Data and Dataset Types

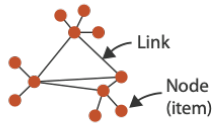
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

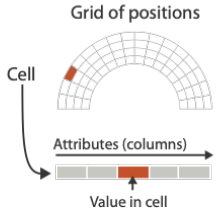
- Tables



- Networks



- Fields (Continuous)



- *Multidimensional Table*

- *Trees*

Attributes

Attribute Types

- Categorical

- Ordered
  - *Ordinal*

  - *Quantitative*

Ordering Direction

- Sequential

- Diverging

- Cyclic