

Paul Rosen

paul.rosen@utah.edu
@paulrosenphd
<https://cspaul.com>



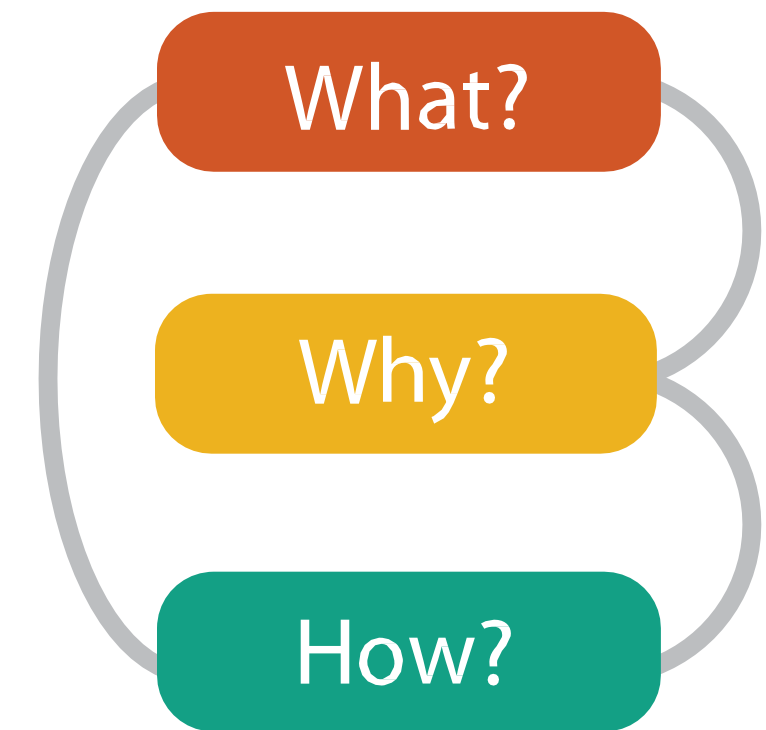
Visualization for Data Science

DS-4630 / CS-5630 / CS-6630

DATA ABSTRACTION

analysis: what, why, and how

- **what** is shown?
 - **why** is the user looking at it?
 - **how** is it shown?
-
- abstract vocabulary avoids domain-specific terms
 - what-why-how analysis framework as scaffold to think systematically about design space



What?

Datasets

Attributes

→ Data Types

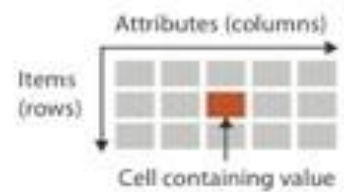
→ Items → Attributes → Links → Positions → Grids

→ Data and Dataset Types

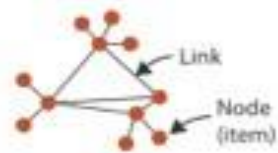
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

→ Dataset Types

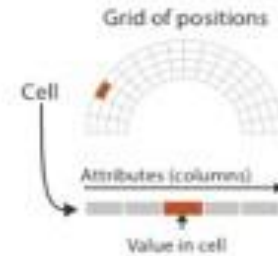
→ Tables



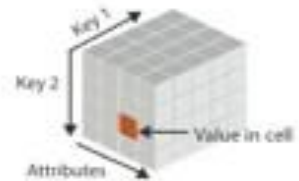
→ Networks



→ Fields (Continuous)



→ Multidimensional Table



→ Trees



→ Geometry (Spatial)



→ Dataset Availability

→ Static



→ Dynamic



→ Attribute Types

→ Categorical



→ Ordered

→ Ordinal



→ Quantitative



→ Ordering Direction

→ Sequential



→ Diverging



→ Cyclic



data abstraction

- the *what* part of an analysis that pertains to the data
- translation of domain-specific terms into words that are as generic as possible

What?

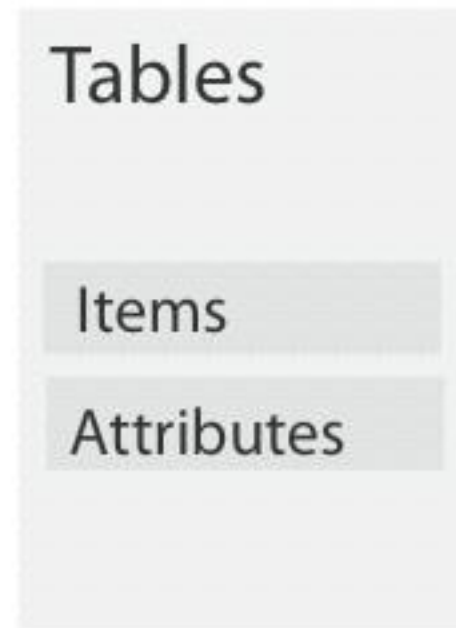
Why?

How?

data types

- Items
- Attributes
- Links
- Positions
- Grids

dataset types



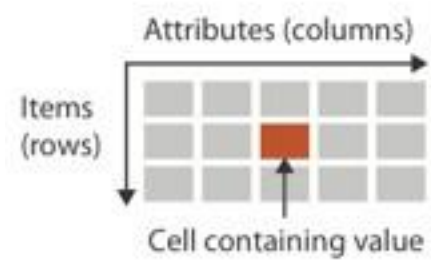
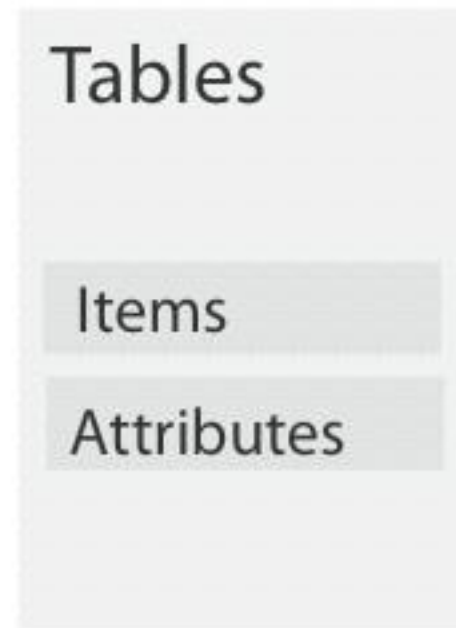
A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box		7/17/07
32	7/16/07	2-High	Medium Box		7/18/07
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	5	4-Not Specified	Small Pack	0.44	6/6/05
69	5	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	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

attribute

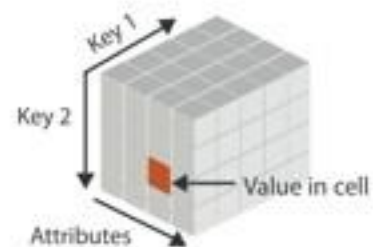
cell

item

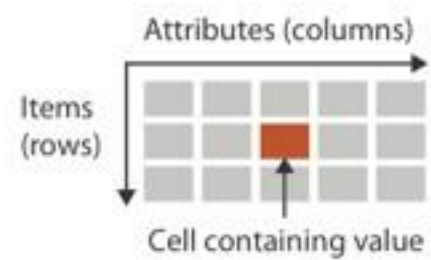
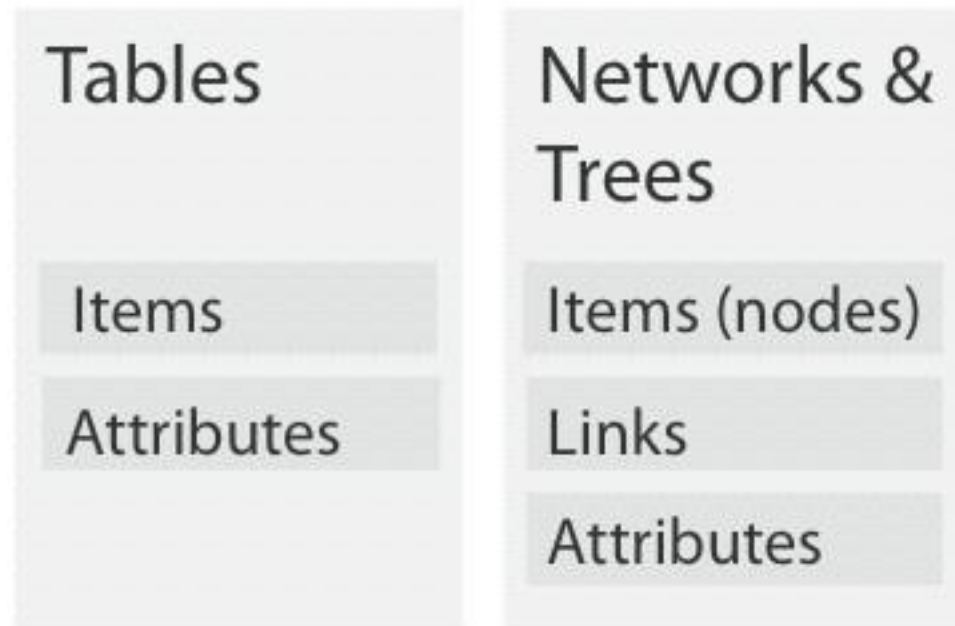
dataset types



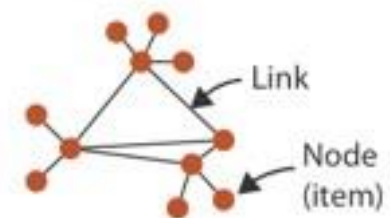
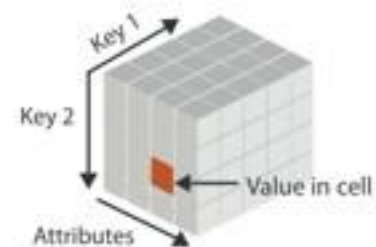
→ *Multidimensional Table*



dataset types



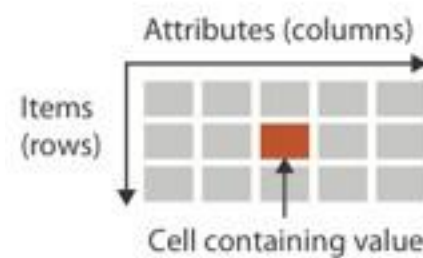
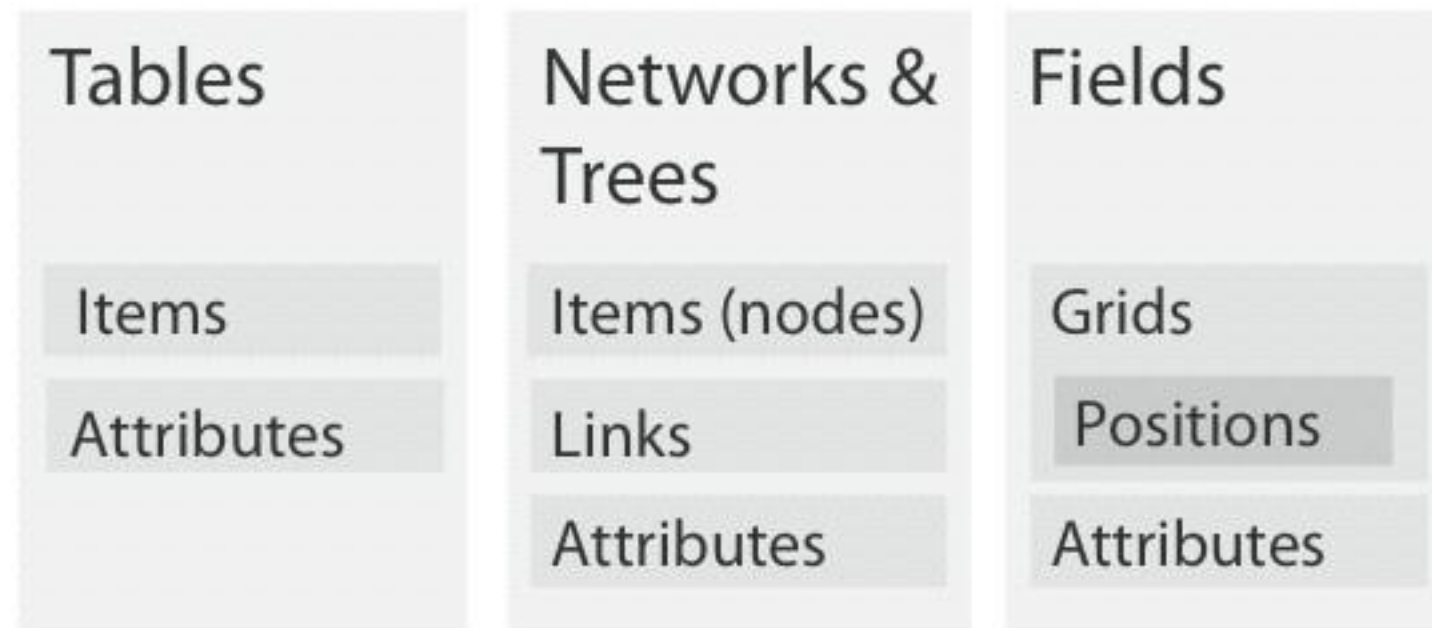
→ *Multidimensional Table*



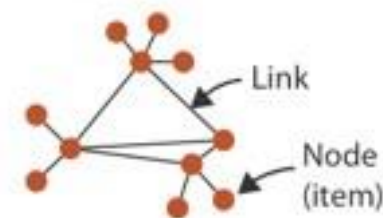
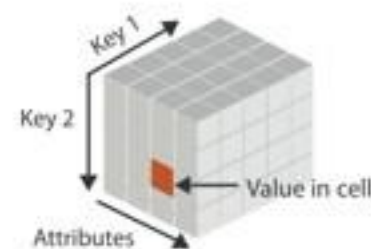
→ *Trees*



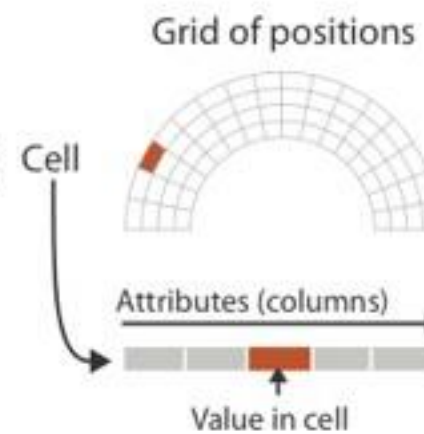
dataset types



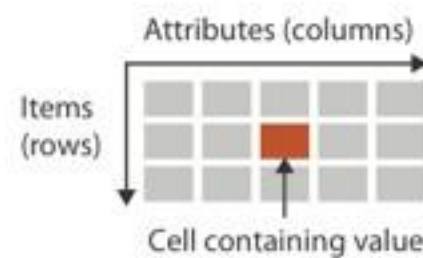
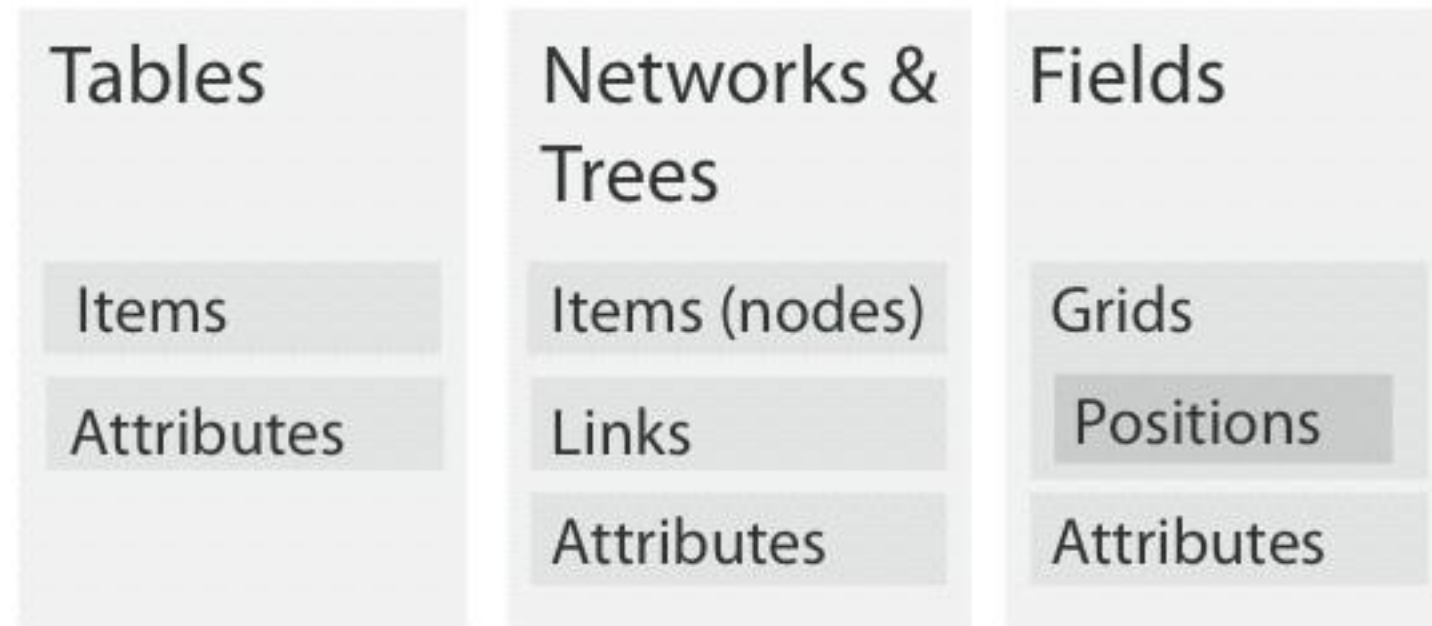
→ *Multidimensional Table*



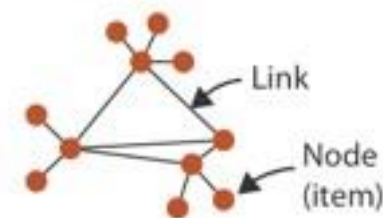
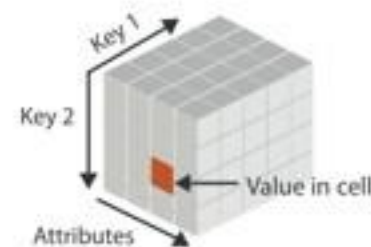
→ *Trees*



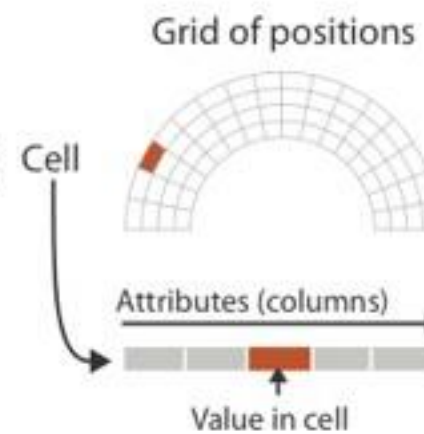
dataset types



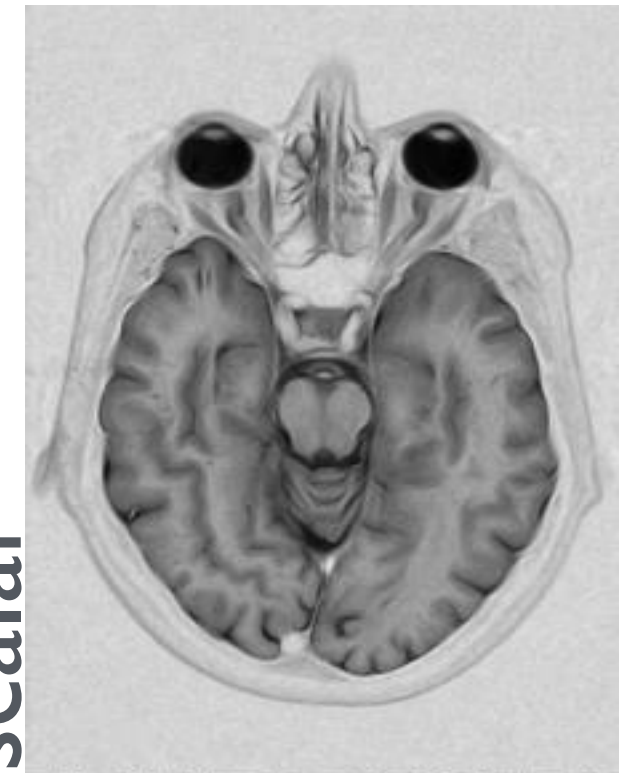
→ *Multidimensional Table*



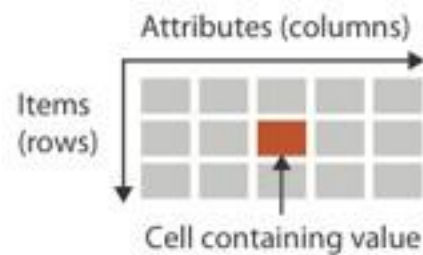
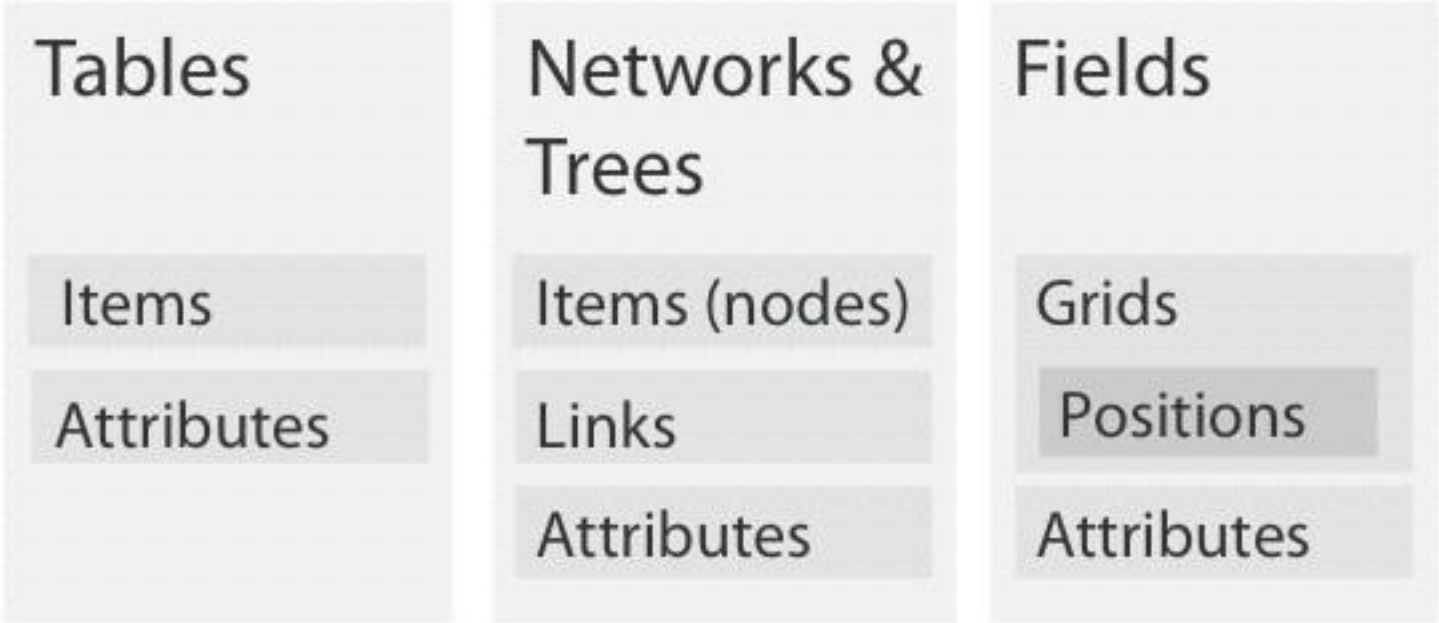
→ *Trees*



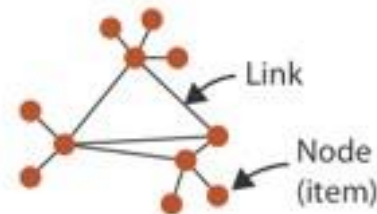
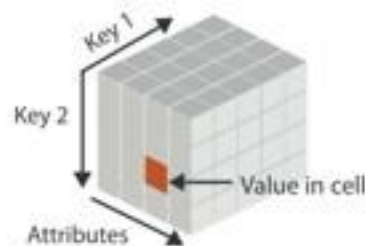
scalar



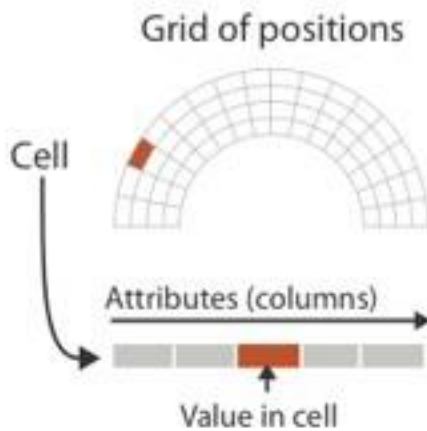
dataset types



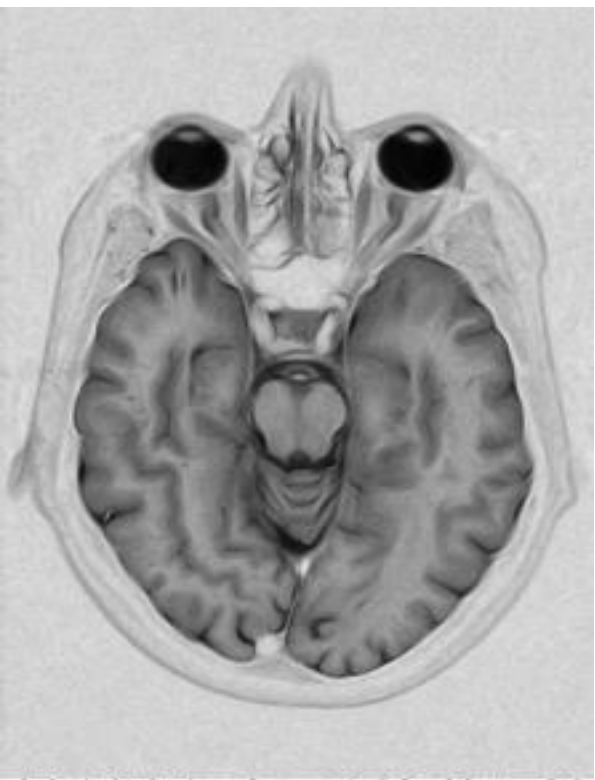
→ Multidimensional Table



→ Trees



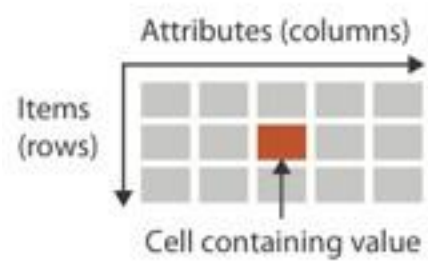
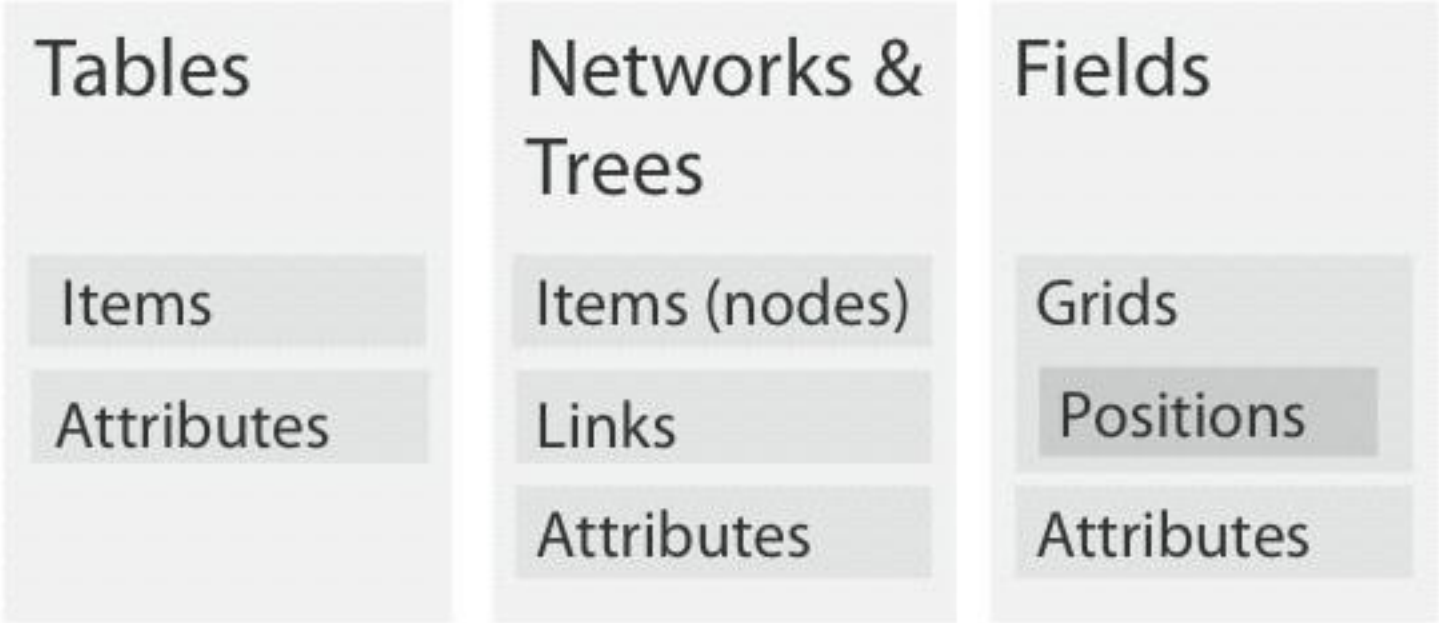
scalar



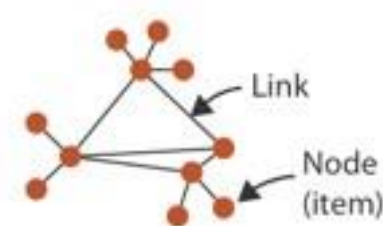
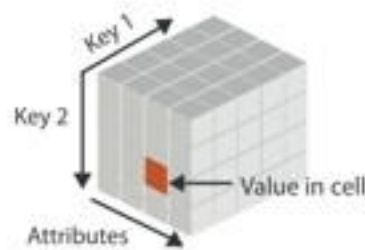
vector



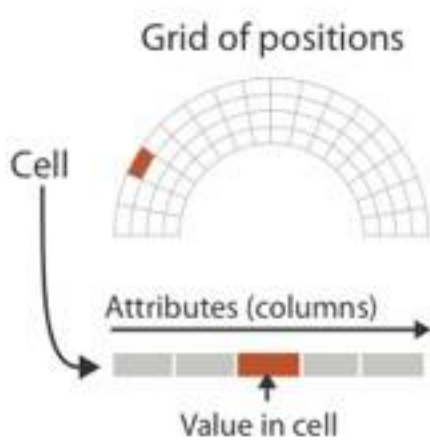
dataset types



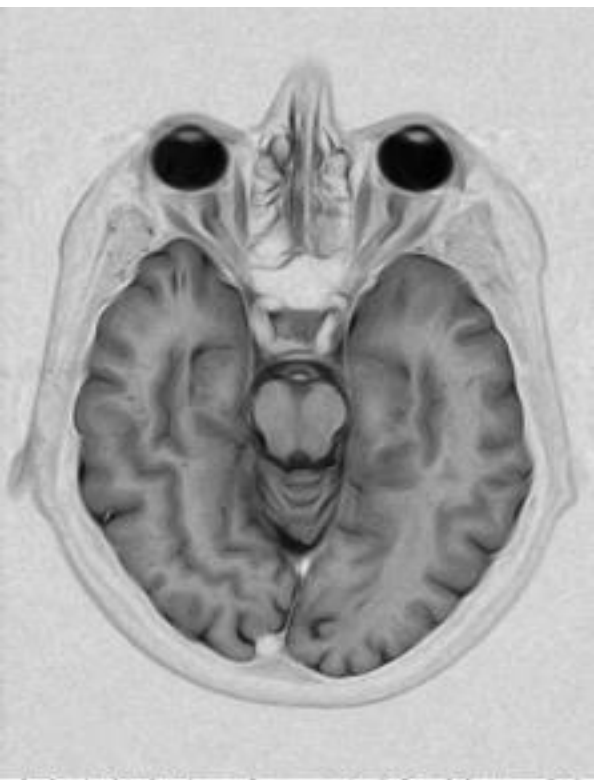
→ Multidimensional Table



→ Trees



scalar



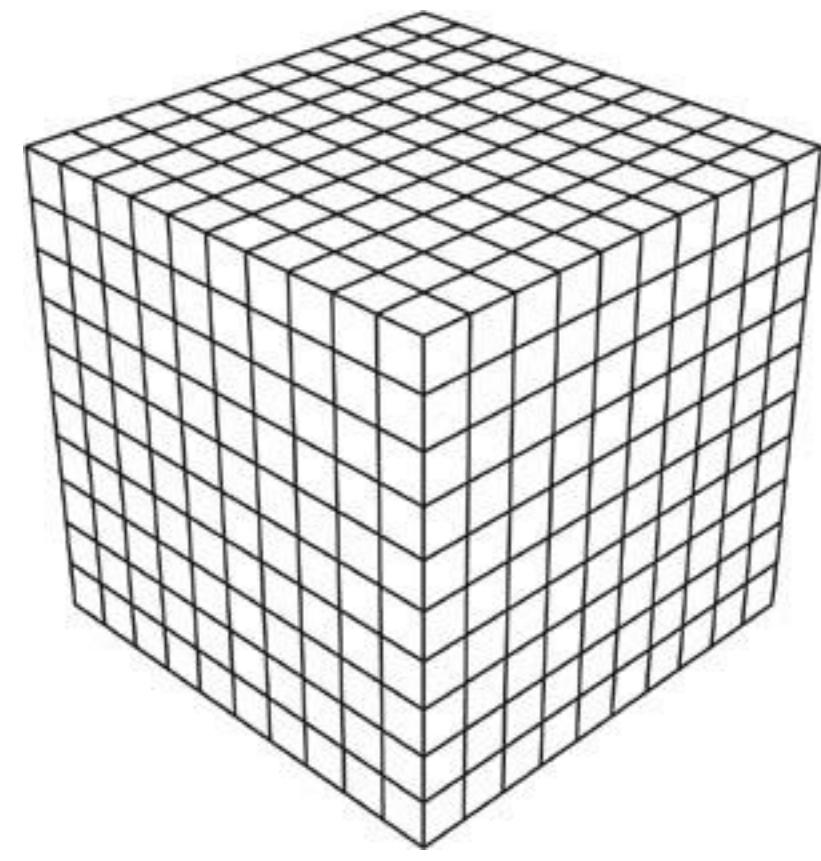
vector

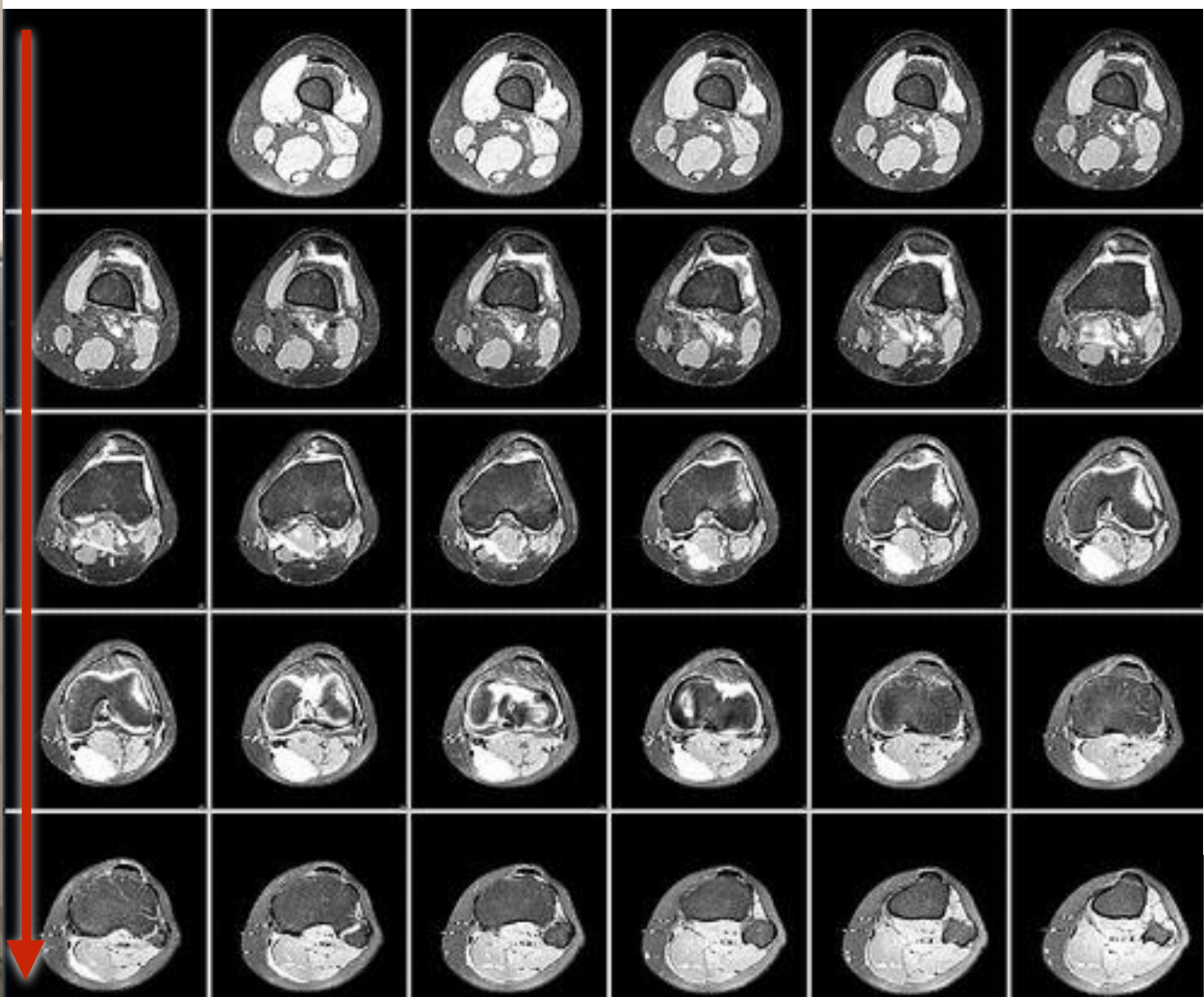
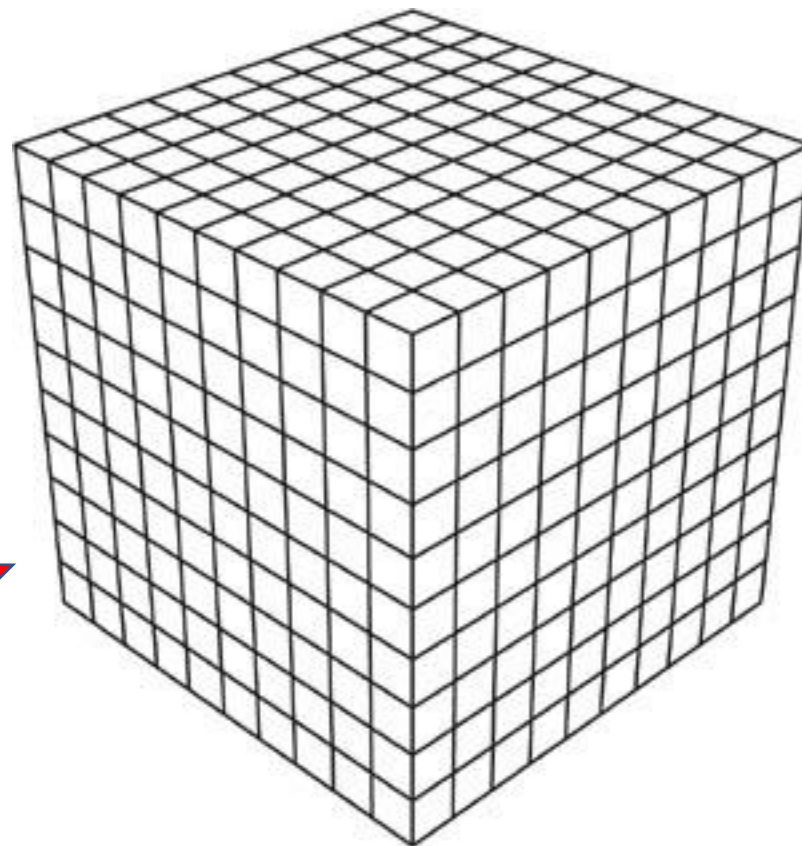
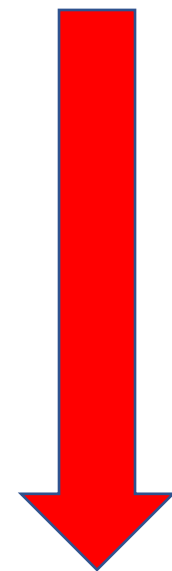


tensor

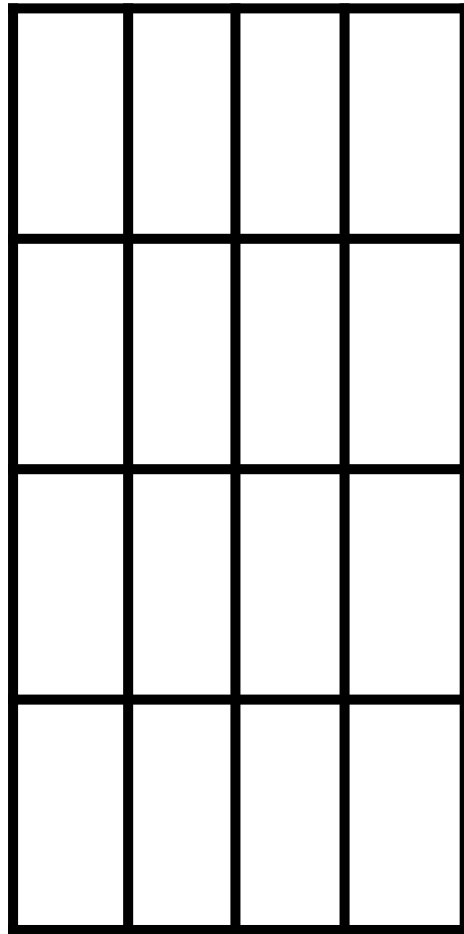




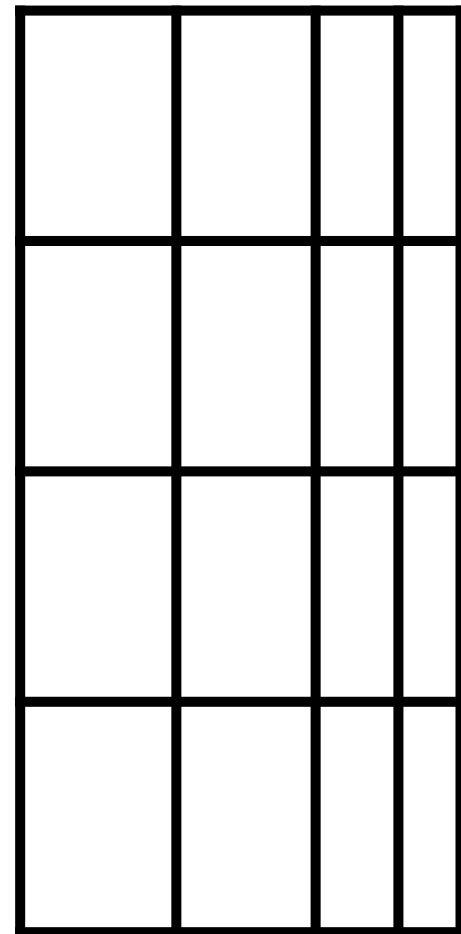




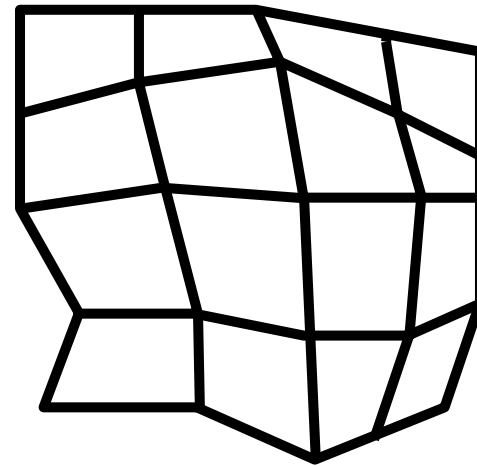
grid types



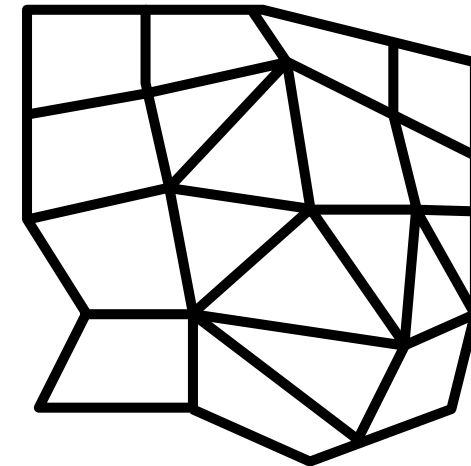
uniform



rectilinear



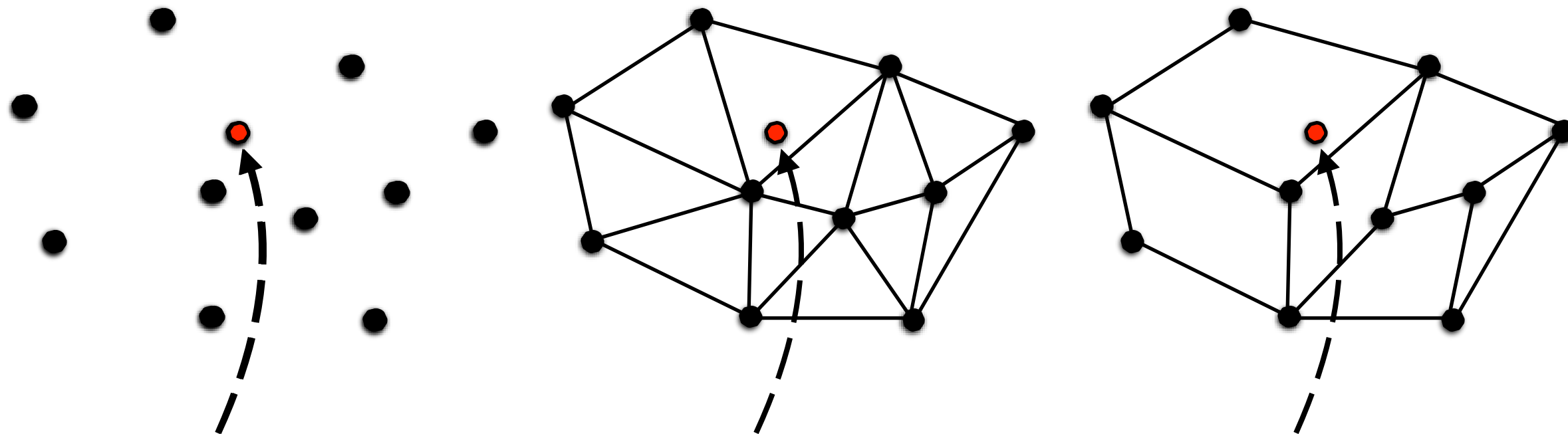
structured



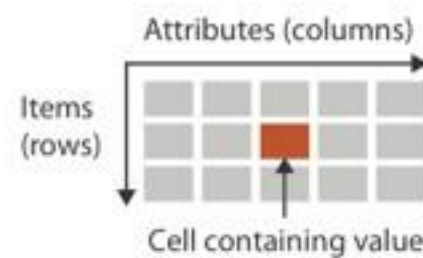
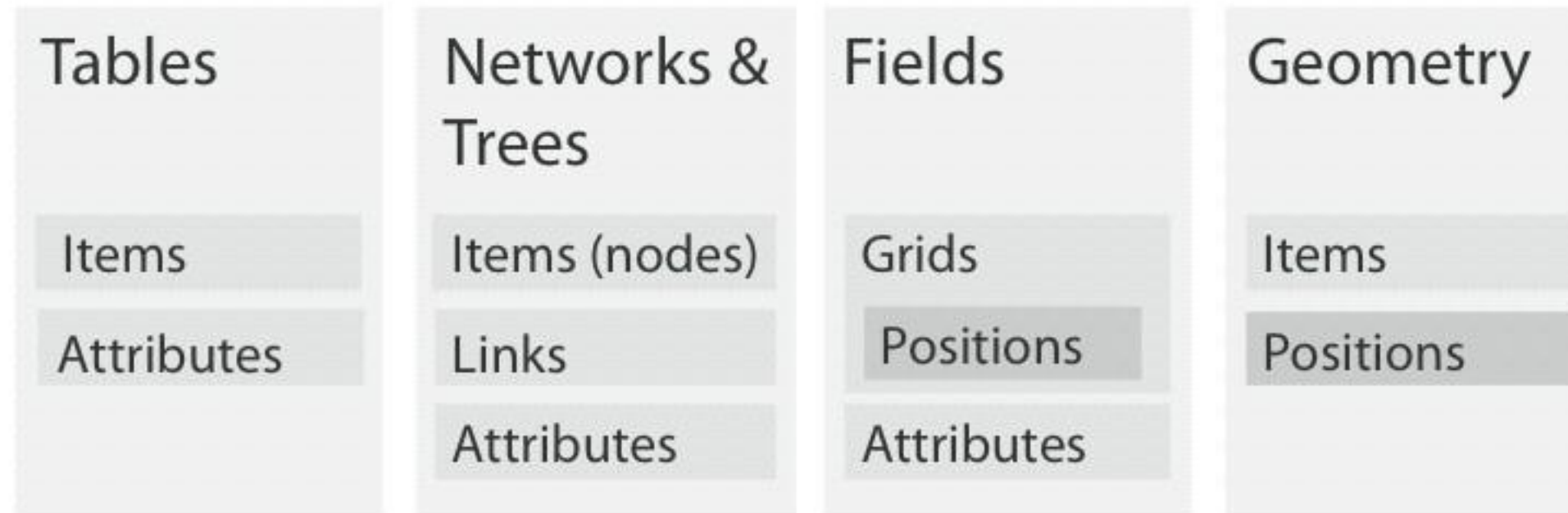
unstructured

grid choices impact how continuous data is interpreted

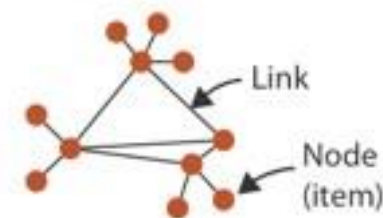
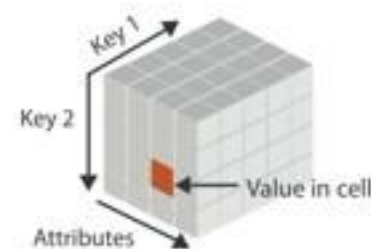
- two key considerations:
 - Sampling – the choice of where attributes are measured
 - Interpolation – how to model the attributes in the rest of space



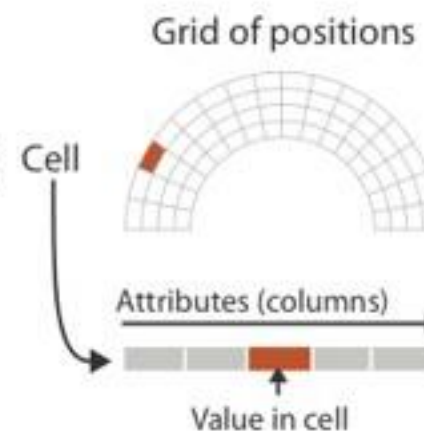
dataset types

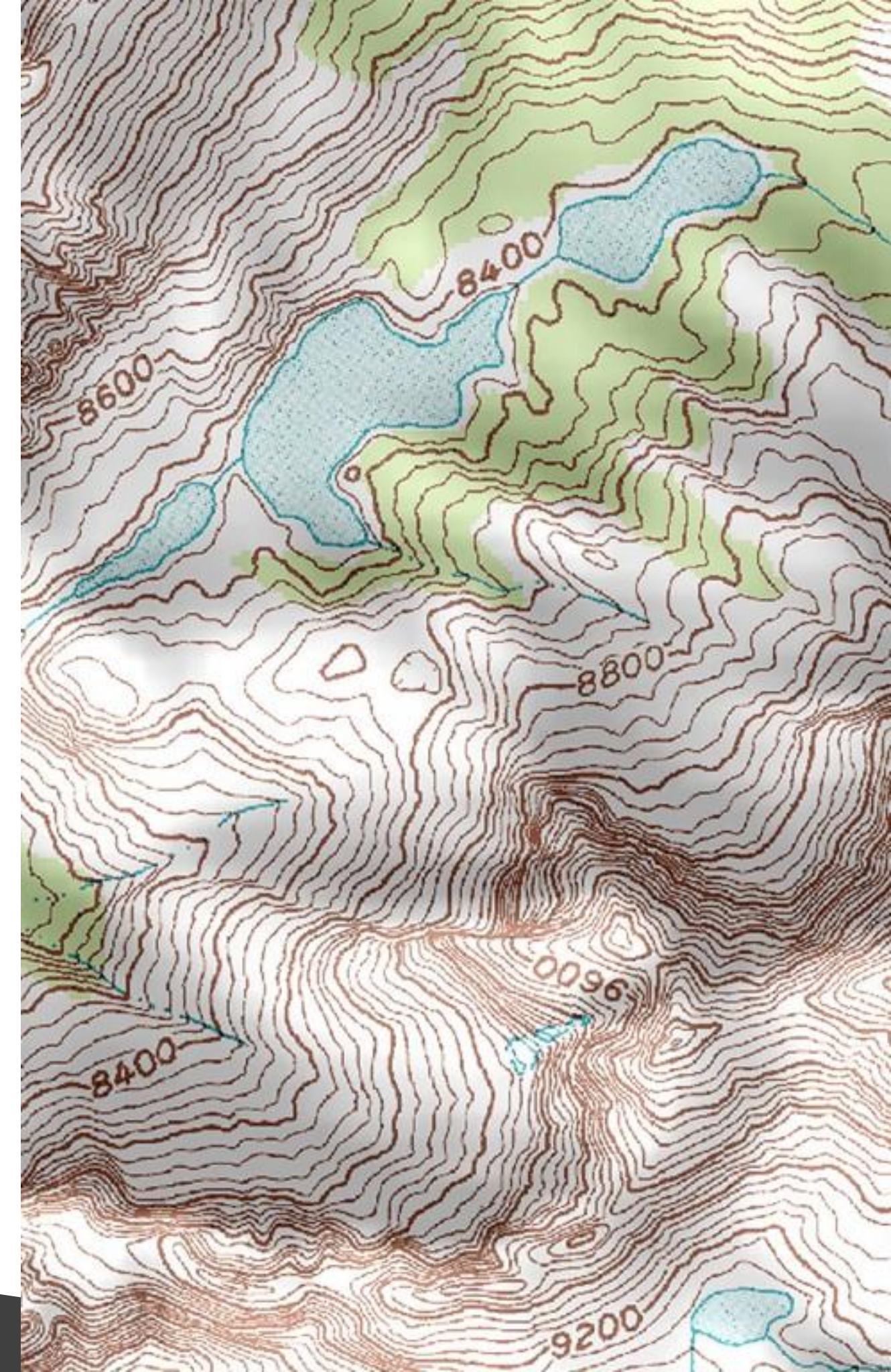
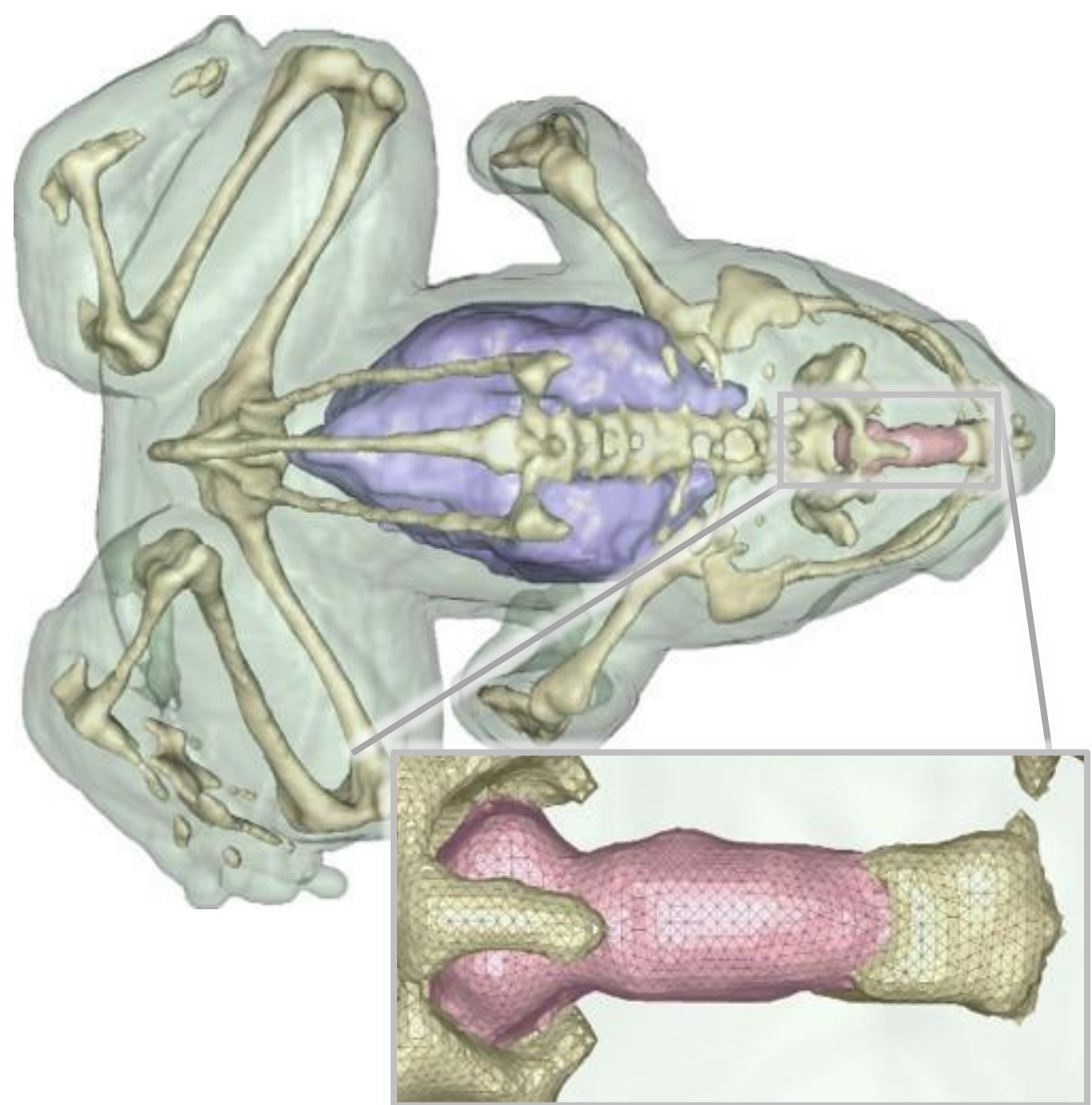


→ *Multidimensional Table*

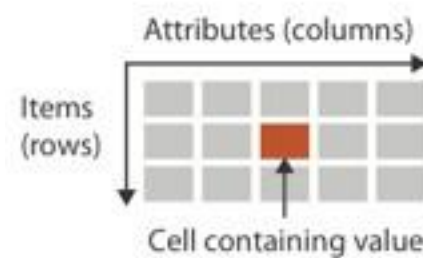


→ *Trees*

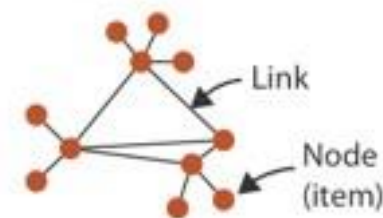
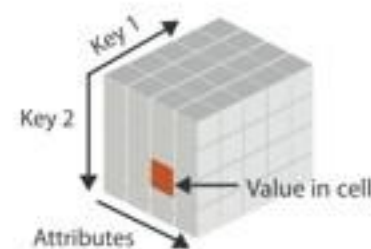




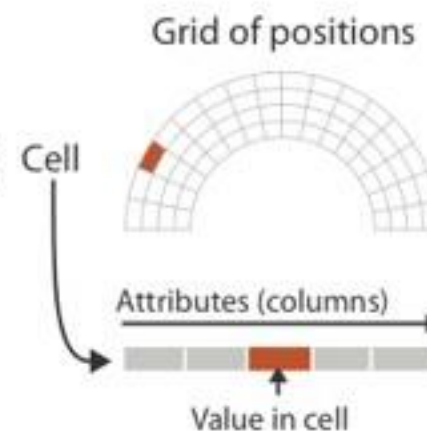
dataset types



→ *Multidimensional Table*



→ *Trees*



attribute types

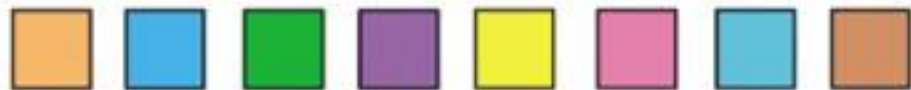
Categorical

no implicit ordering

attribute types

Categorical

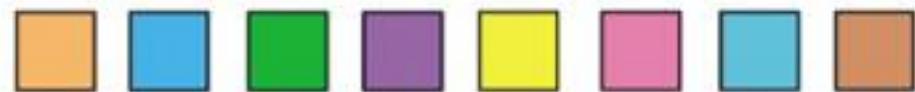
no implicit ordering



attribute types

Categorical

no implicit ordering



Ordered

Ordinal

Quantitative

attribute types

Categorical

no implicit ordering



Ordered

Ordinal

Quantitative

meaningful magnitude
(can do arithmetic)

attribute types

Categorical

no implicit ordering



Ordered

Ordinal



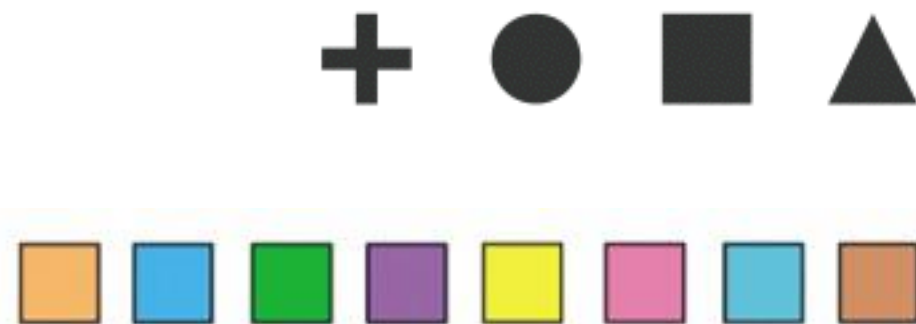
Quantitative

meaningful magnitude
(can do arithmetic)

attribute types

Categorical

no implicit ordering



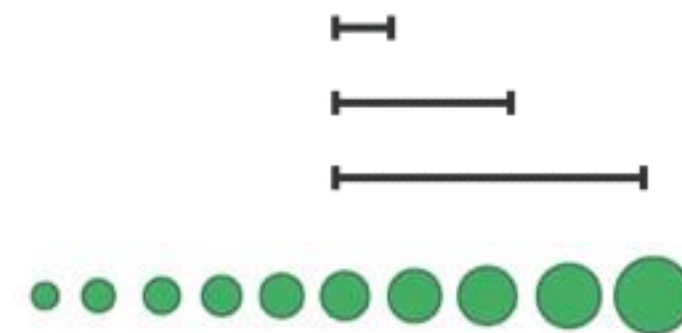
Ordered

Ordinal



Quantitative

meaningful magnitude
(can do arithmetic)



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 Specified		0.6	6/6/05
70	12/18/06	5-Low		0.59	12/23/06
70	12/18/06	5-Low		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

quantitative
ordinal
categorical

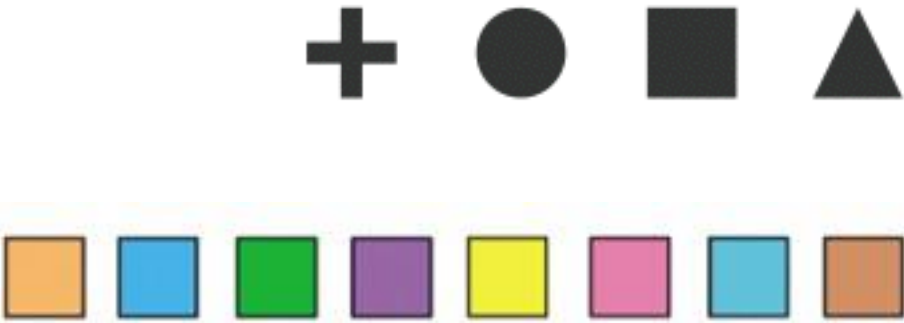
A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	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 Specified		0.6	6/6/05
70	12/18/06	5-Low		0.59	12/23/06
70	12/18/06	5-Low		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

quantitative
ordinal
categorical

attribute types

Categorical

no implicit ordering



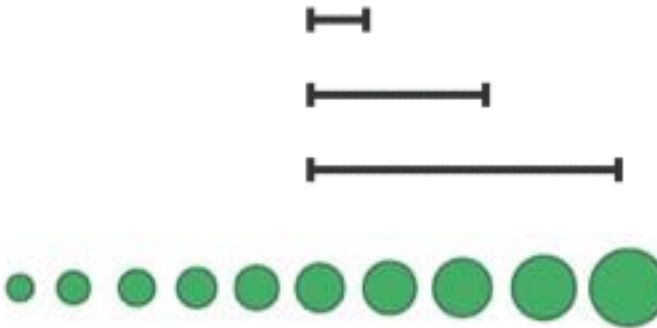
Ordered

Ordinal



Quantitative

meaningful magnitude
(can do arithmetic)



Sequential



Divergent

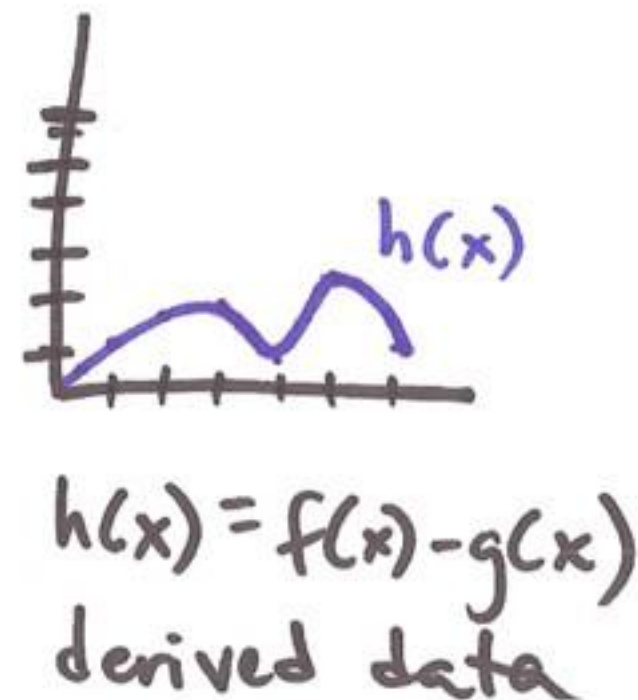


Cyclic



DERIVED ATTRIBUTES

- derived attribute: computed from originals
 - simple change of type
 - acquire additional data
 - complex transformation
 - transformation is abstraction choice



TYPE vs SEMANTIC

- **data model:** mathematical abstraction (data abstraction)
 - set with operations, e.g., floats with $*$ / $-$ $+$
- **conceptual model:** mental construction (semantics)
 - includes semantics, supports reasoning
- conceptual model motivates data abstraction choices

EXAMPLE

- **from data model . . .**
 - - 32.52, 54.06, -17.35, . . . (floats)
- **using conceptual model . . .**
 - temperature
- **to new data abstraction...**
 - continuous to 2 significant figures (quantitative)
 - hot, warm, cold (ordinal)
 - above freezing, below freezing (categorical)

