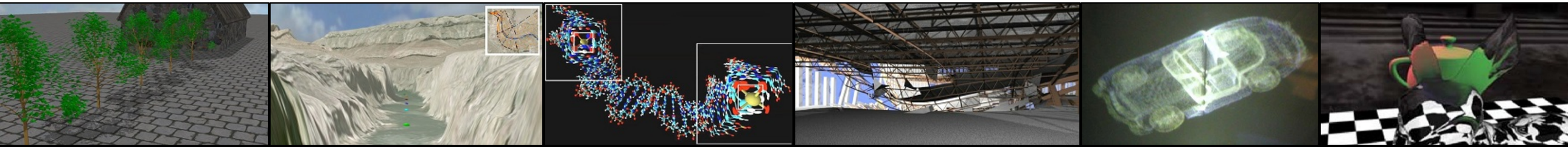
CIS 4930/6930-002
DATA VISUALIZATION

DATA ABSTRACTION

Paul Rosen
Assistant Professor
University of South Florida

slides credits Miriah Meyer (U of Utah)

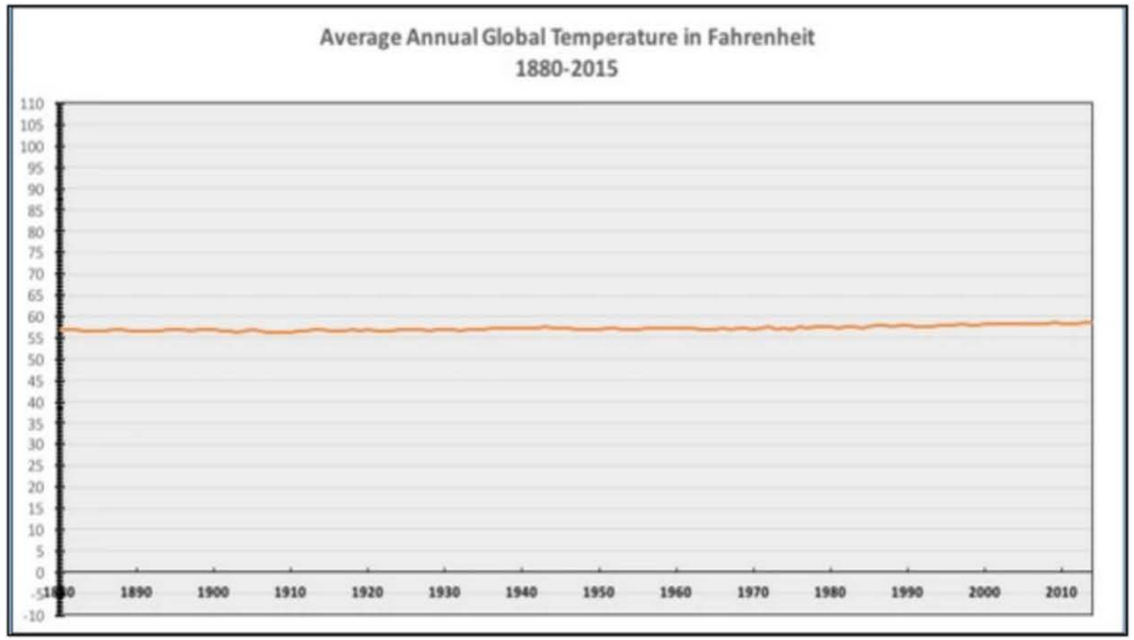


National Review
@NRO

Follow

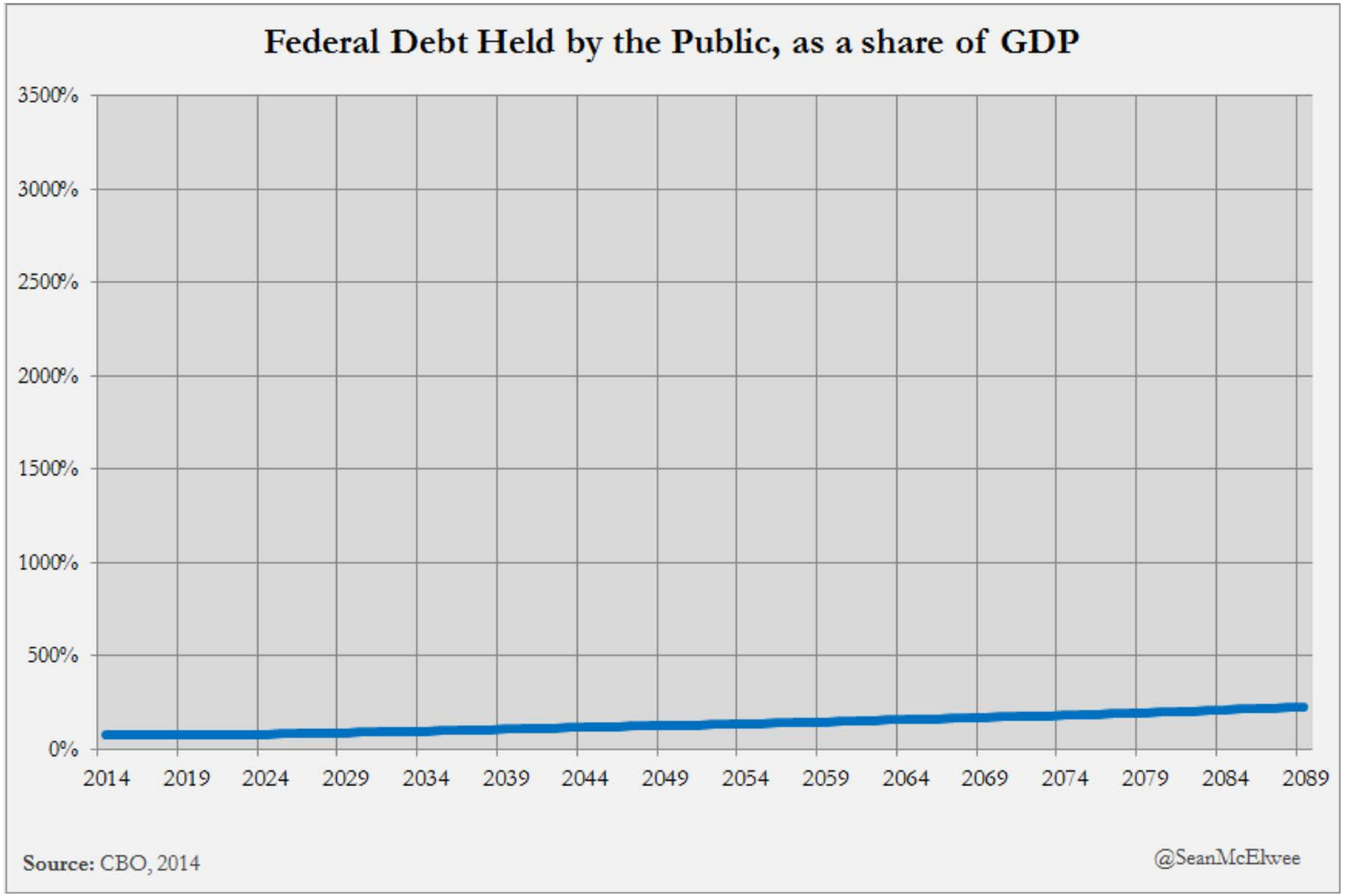
The only #climatechange chart you need to see. natl.re/wPKpro

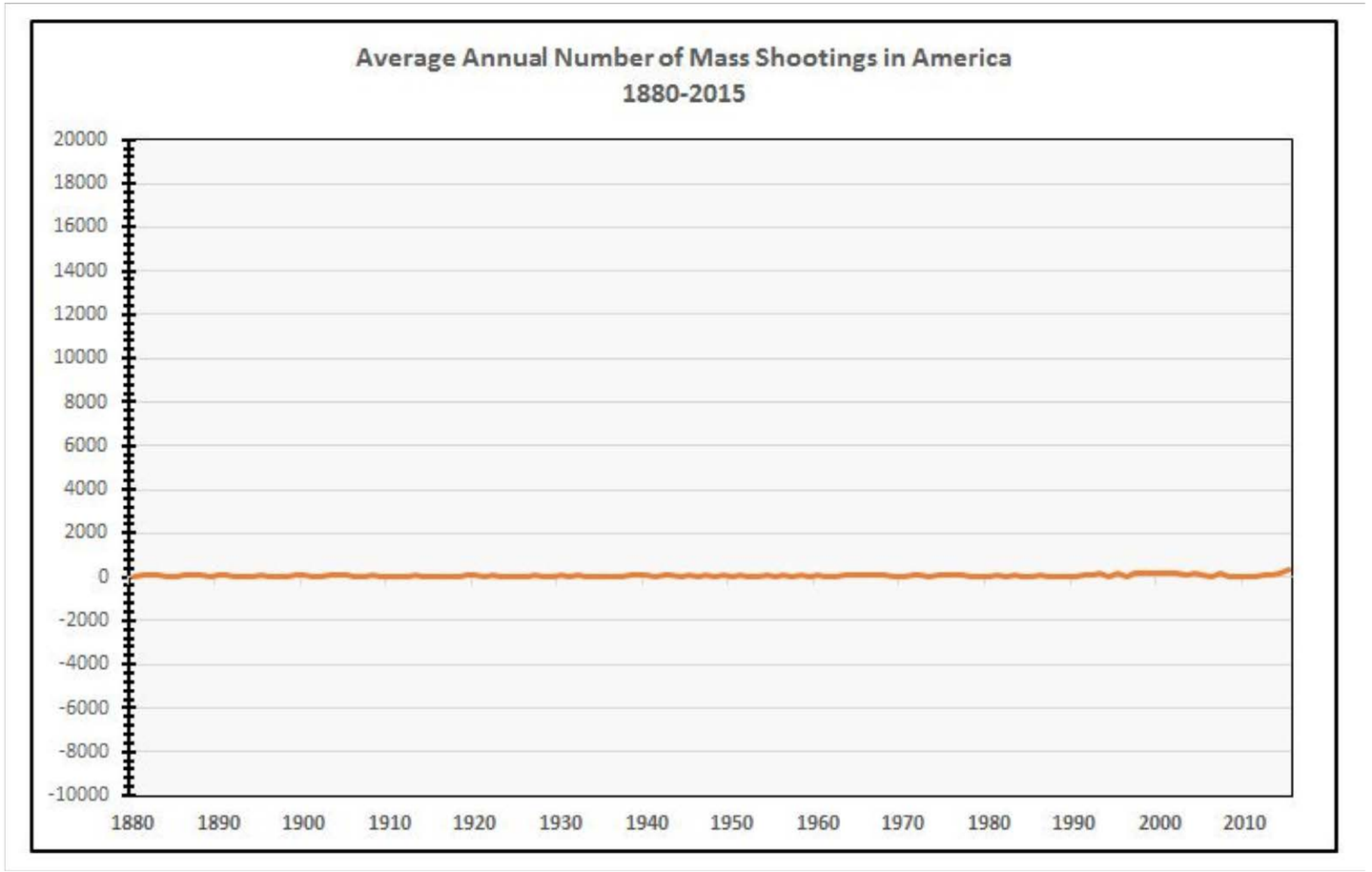
(h/t @powerlineUS)



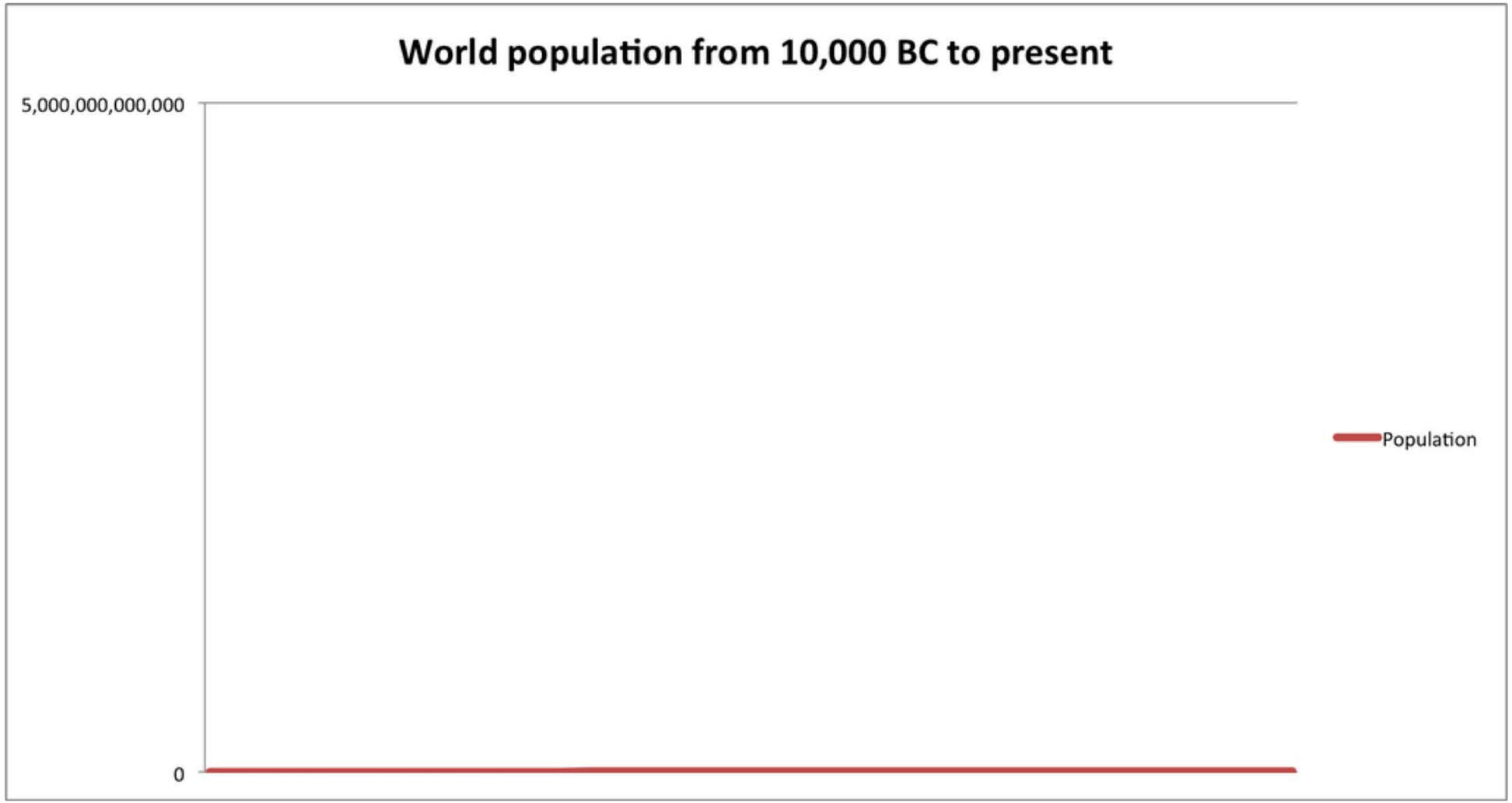
RETWEETS 408 LIKES 322

1:36 PM - 14 Dec 2015


Federal Debt Held by the Public, as a share of GDP
3500%
3000%
2500%
2000%
1500%
1000%
500%
0%
2014
2019
2024
2029
2034
2039
2044
2049
2054
2059
2064
2069
2074
2079
2084
2089
Source: CBO, 2014
@SeanMcElwee



HTTP://WWW.VISLIES.ORG/


World population from 10,000 BC to present
5,000,000,000,000
0
Population

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

TYPE vs SEMANTIC

DATA TYPES

Items

Attributes

Links

Positions

Grids

DATASET TYPES

Tables

Items

Attributes

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		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	6/6/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/6/05	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

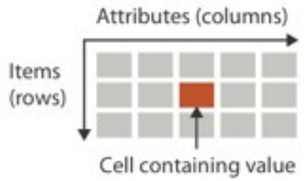
item

DATASET TYPES

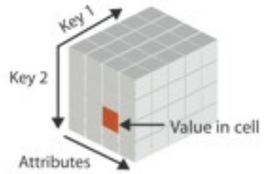
Tables

Items

Attributes

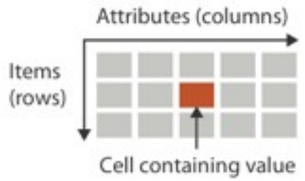


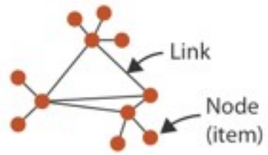
➔ *Multidimensional Table*



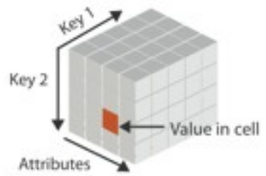
DATASET TYPES

Tables	Networks & Trees
Items	Items (nodes)
Attributes	Links
	Attributes





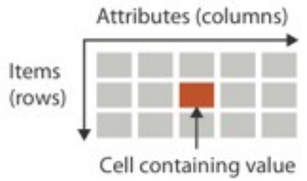
➔ *Multidimensional Table*

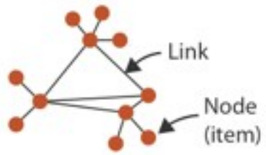


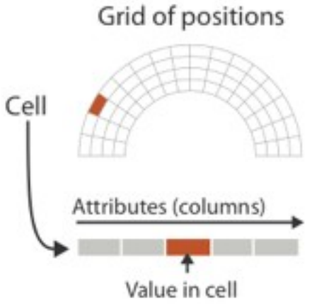
➔ *Trees*

DATASET TYPES

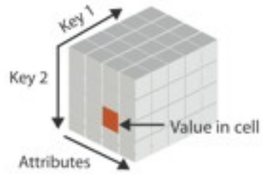
Tables	Networks & Trees	Fields
Items	Items (nodes)	Grids
Attributes	Links	Positions
	Attributes	Attributes







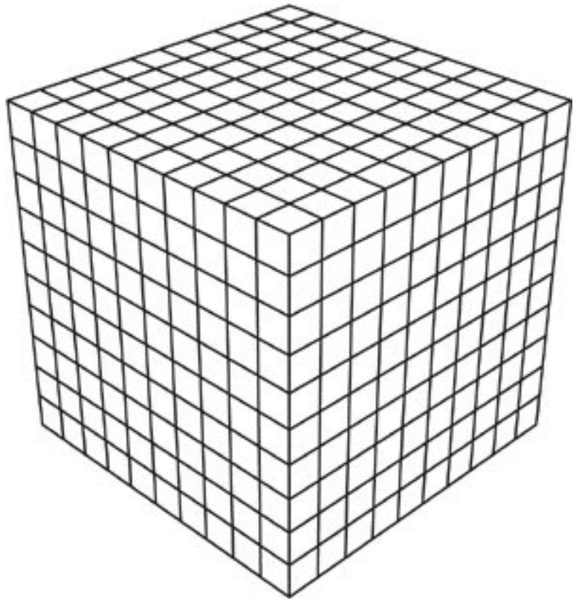
→ *Multidimensional Table*



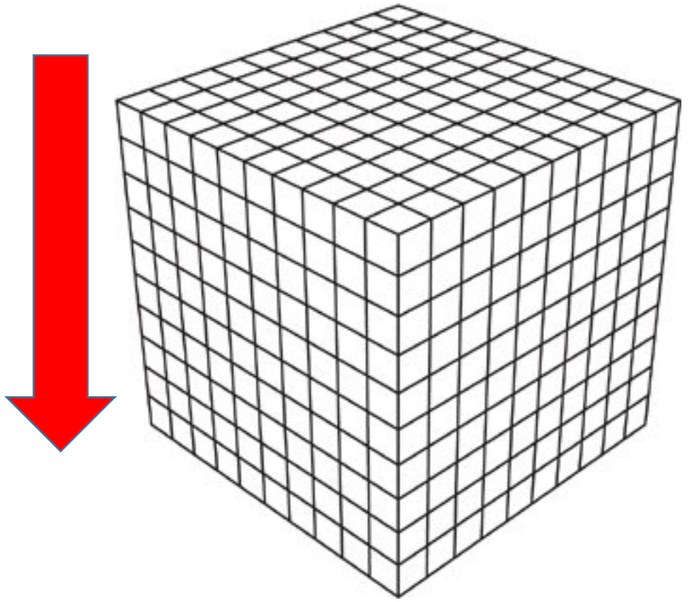
→ *Trees*

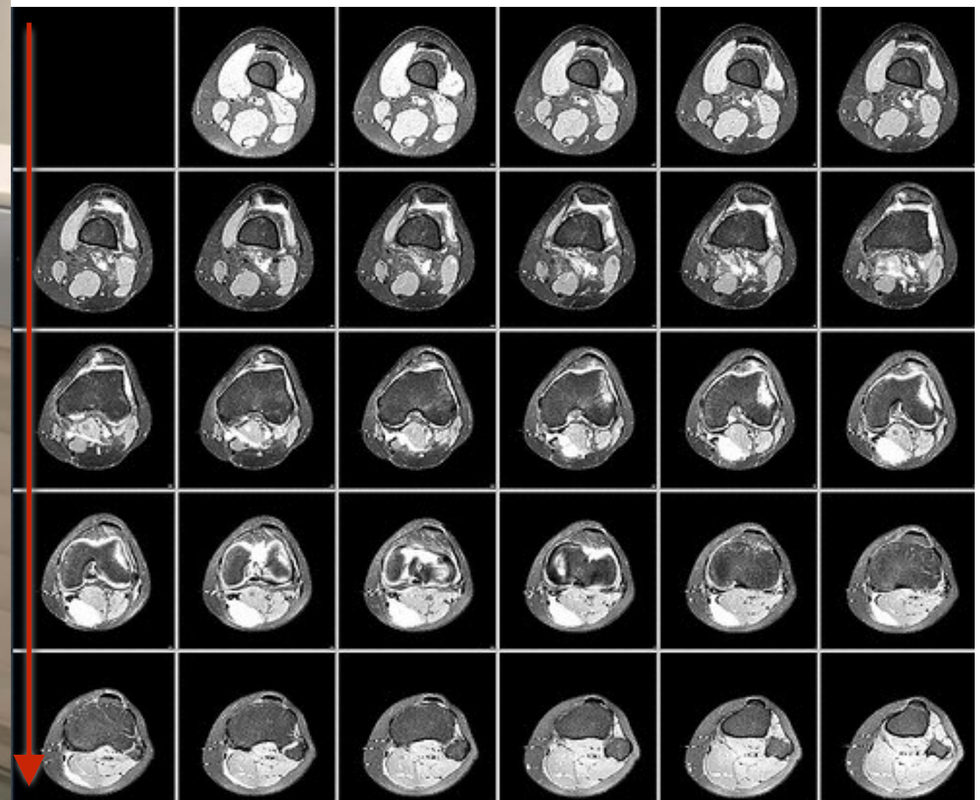

TOSHIBA


TOSHIBA

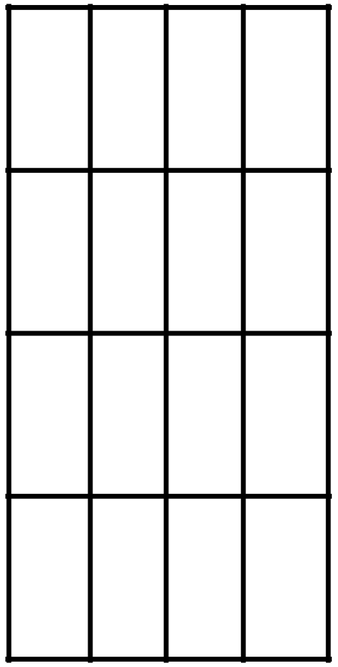



TOSHIBA

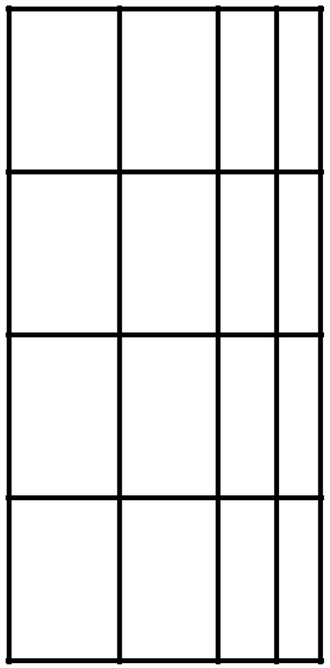




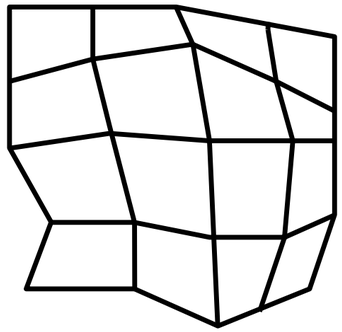
GRID TYPES



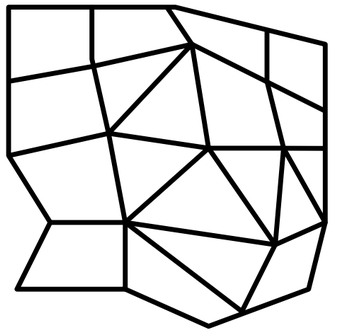
uniform



rectilinear



structured



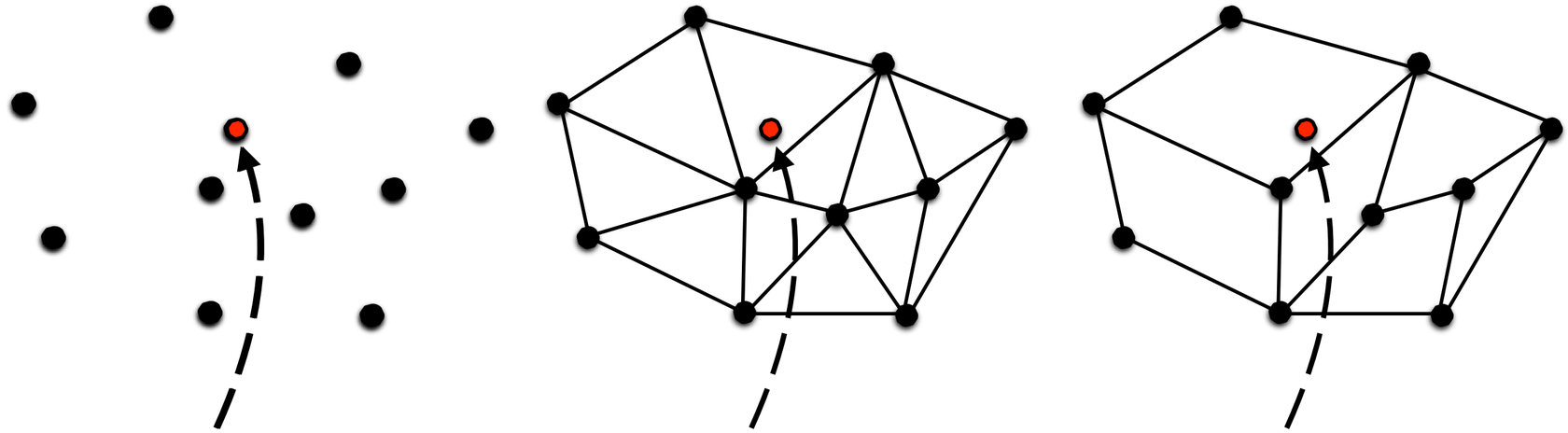
unstructured

GRID CHOICES IMPACT HOW CONTINUOUS DATA IS INTERPRETED

two key considerations:

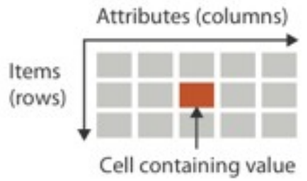
Sampling – the choice of where attributes are measured

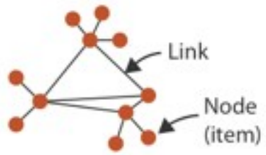
Interpolation – how to model the attributes in the rest of space

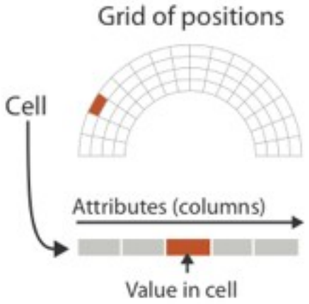


DATASET TYPES

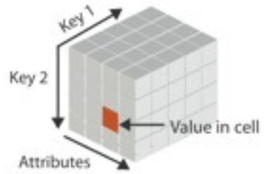
Tables	Networks & Trees	Fields
Items	Items (nodes)	Grids
Attributes	Links	Positions
	Attributes	Attributes







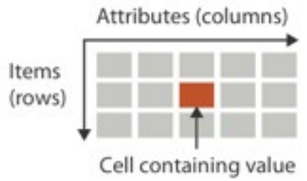
→ *Multidimensional Table*



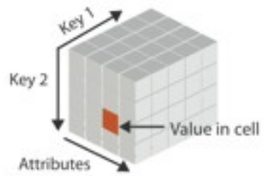
→ *Trees*

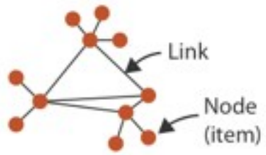
DATASET TYPES

Tables	Networks & Trees	Fields
Items	Items (nodes)	Grids
Attributes	Links	Positions
	Attributes	Attributes

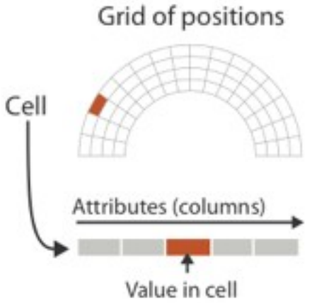


→ *Multidimensional Table*



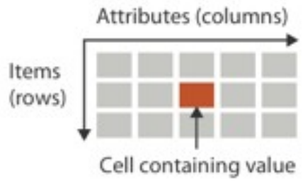


→ *Trees*

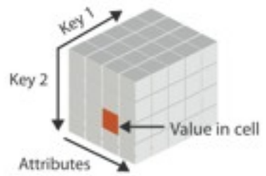


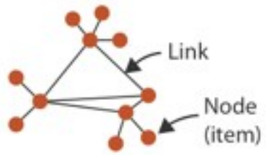
DATASET TYPES

Tables	Networks & Trees	Fields
Items	Items (nodes)	Grids
Attributes	Links	Positions
	Attributes	Attributes

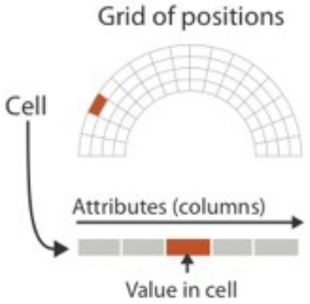


→ *Multidimensional Table*

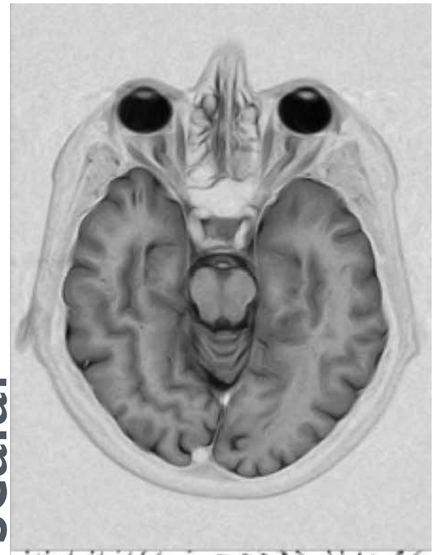




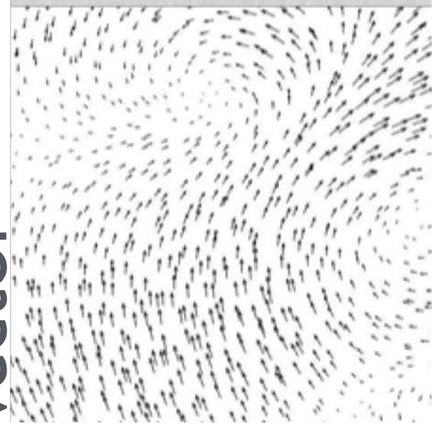
→ *Trees*



scalar

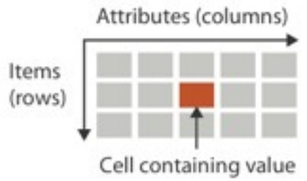


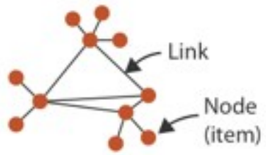
vector

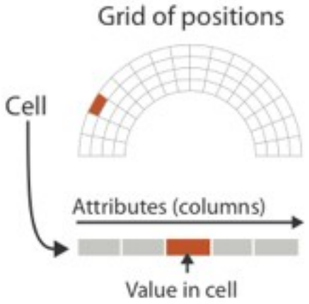


DATASET TYPES

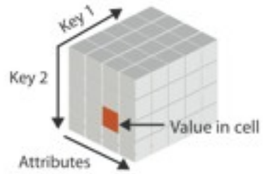
Tables	Networks & Trees	Fields
Items	Items (nodes)	Grids
Attributes	Links	Positions
	Attributes	Attributes





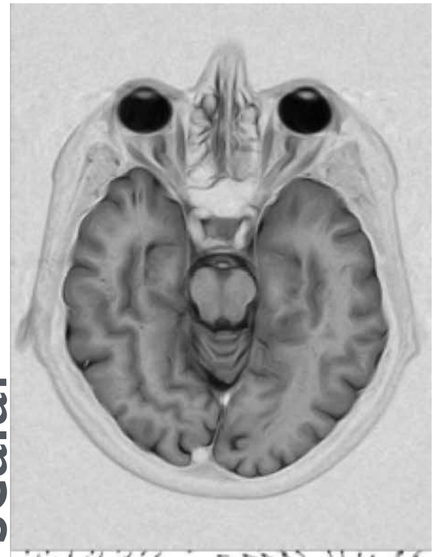


→ *Multidimensional Table*

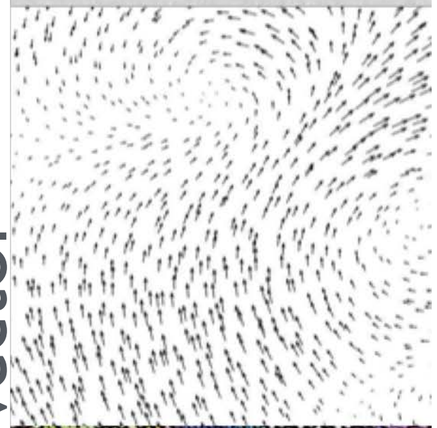


→ *Trees*

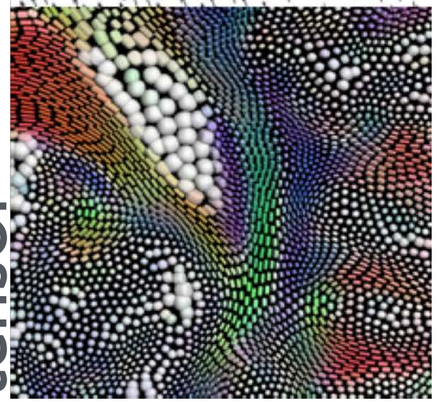
scalar



vector

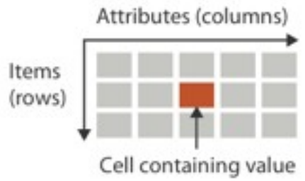


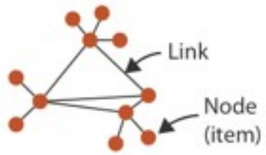
tensor

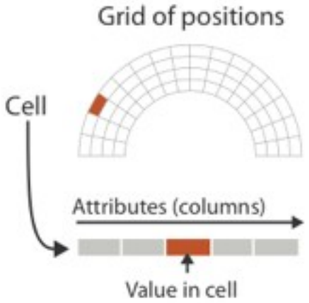


DATASET TYPES

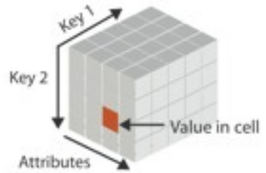
Tables	Networks & Trees	Fields	Geometry
Items	Items (nodes)	Grids	Items
Attributes	Links	Positions	Positions
	Attributes	Attributes	



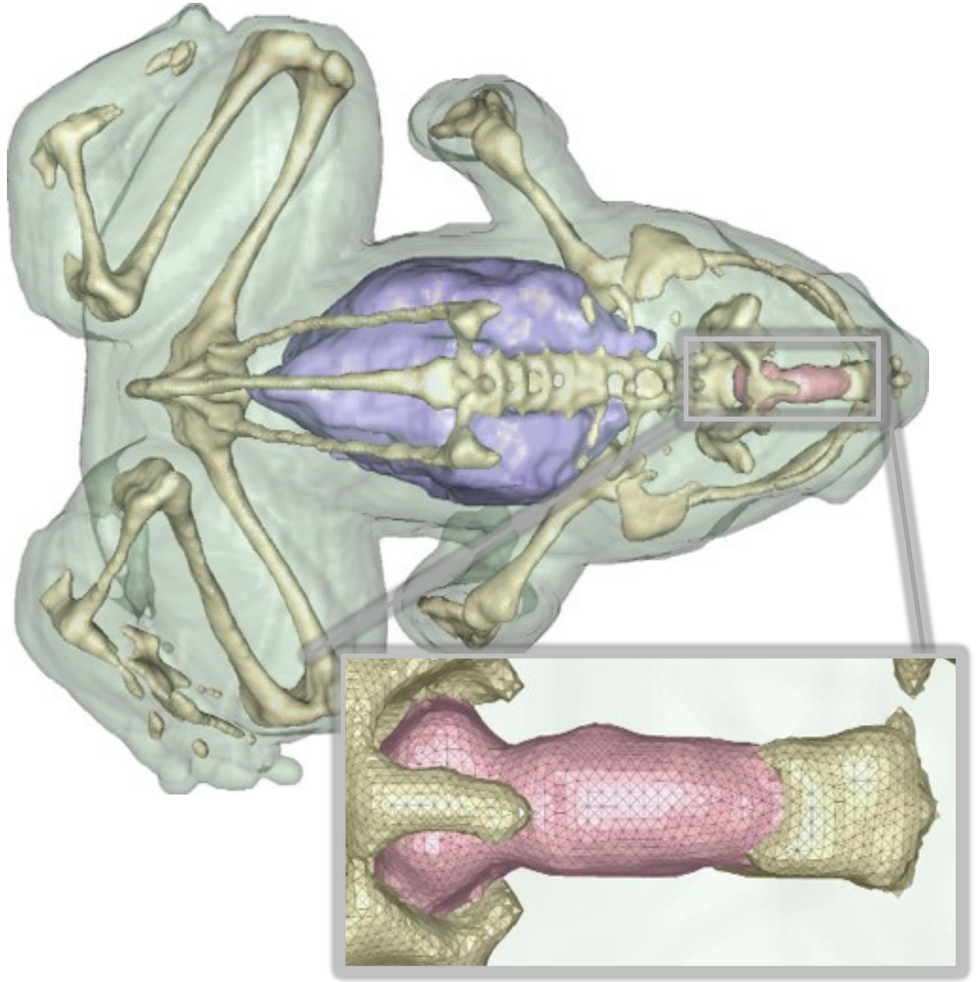


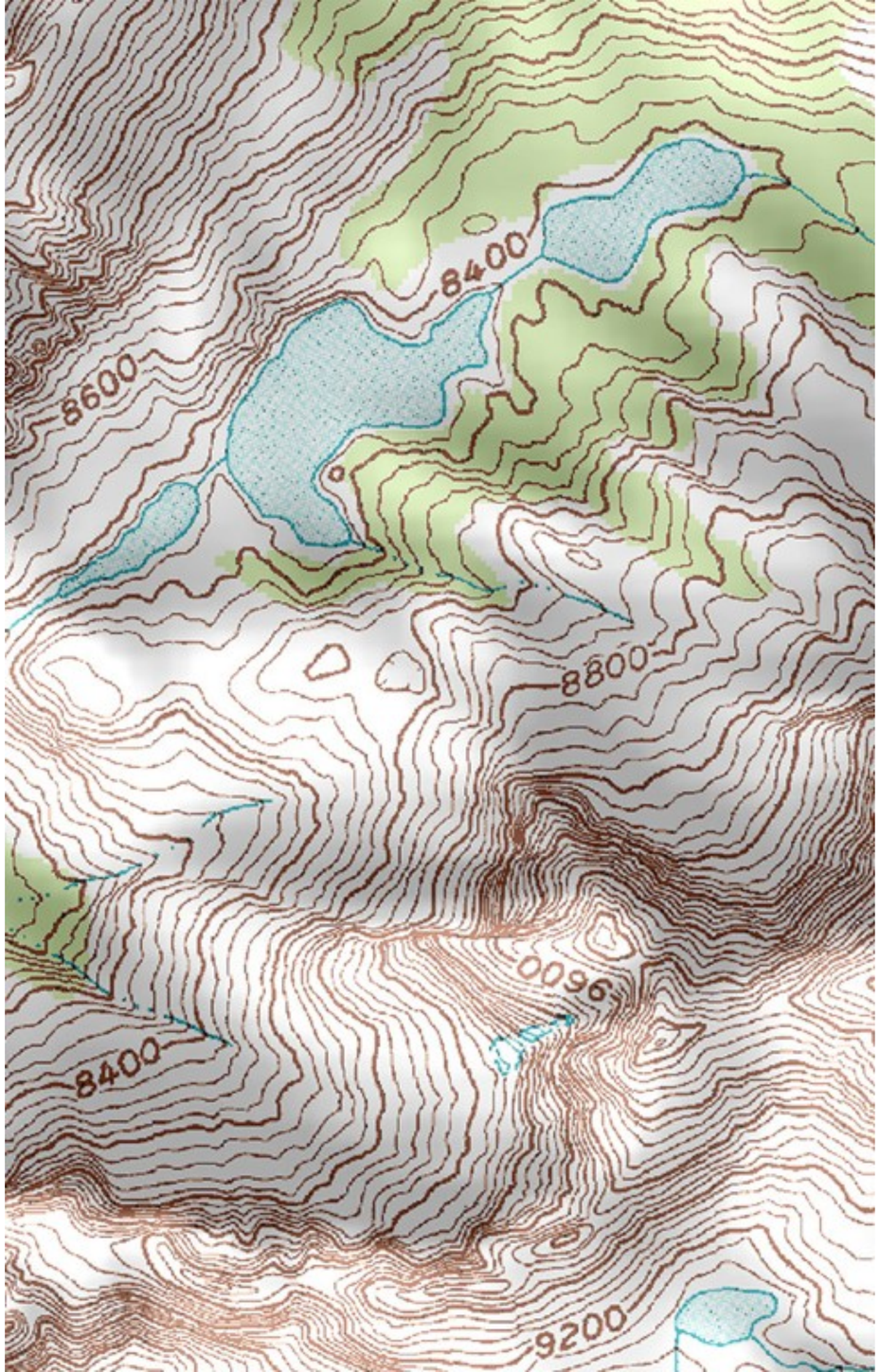


→ *Multidimensional Table*



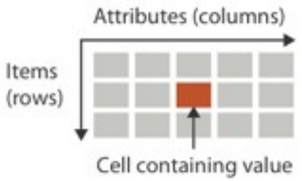
→ *Trees*

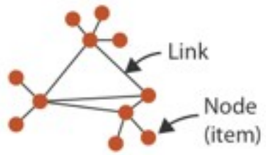


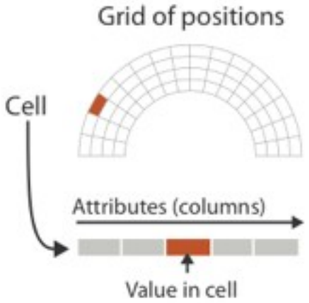

8400
8600
8800
9600
8400
9200

DATASET TYPES

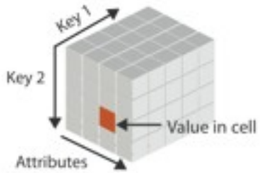
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		







→ *Multidimensional Table*



→ *Trees*

ATTRIBUTE TYPES

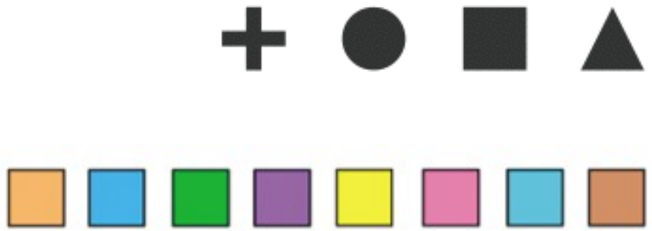
Categorical

no implicit ordering

ATTRIBUTE TYPES

Categorical

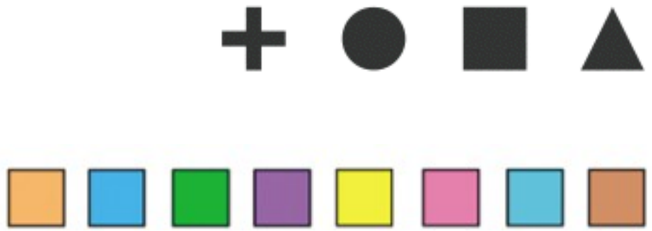
no implicit ordering



ATTRIBUTE TYPES

Categorical

no implicit ordering



Ordered

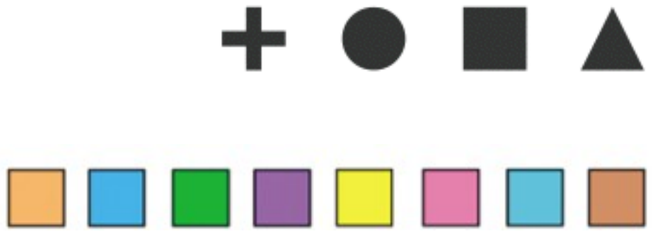
Ordinal

Quantitative

ATTRIBUTE TYPES

Categorical

no implicit ordering



Ordered

Ordinal

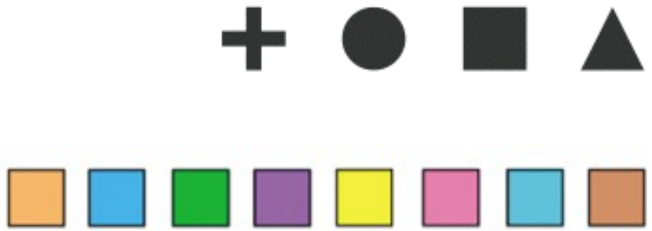
Quantitative

meaningful magnitude
(can do arithmetic)

ATTRIBUTE TYPES

Categorical

no implicit ordering



Ordered

Ordinal

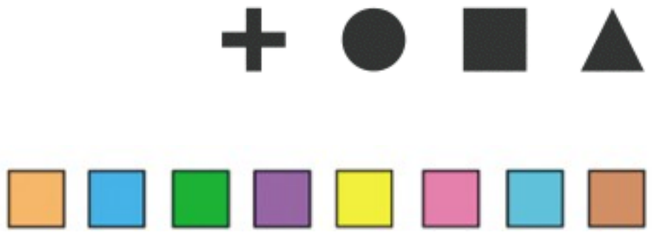
Quantitative

meaningful magnitude
(can do arithmetic)

ATTRIBUTE TYPES

Categorical

no implicit ordering

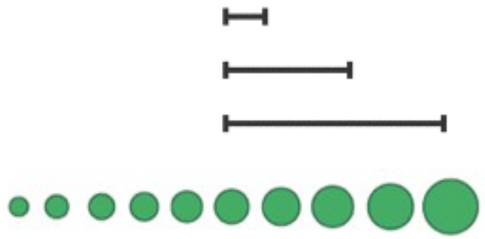


Ordered

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 Spec		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

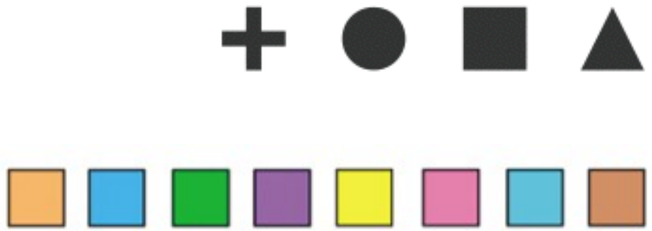
A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	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	[illegible]	0.6	6/6/05
70	12/18/06	5-Low	[illegible]	0.59	12/23/06
70	12/18/06	5-Low	[illegible]	0.82	12/23/06
96	4/17/05	2-High	[illegible]	0.55	4/19/05
97	1/29/06	3-Medium	[illegible]	0.38	1/30/06
129	11/19/08	5-Low	[illegible]	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

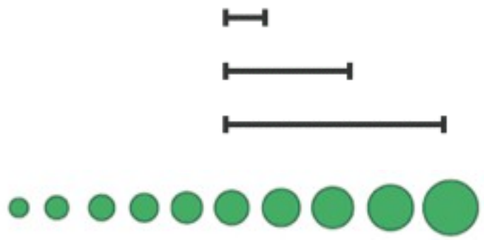


Ordered

Ordinal

Quantitative

meaningful magnitude
(can do arithmetic)



Sequential

Divergent

Cyclic

DERIVED ATTRIBUTES

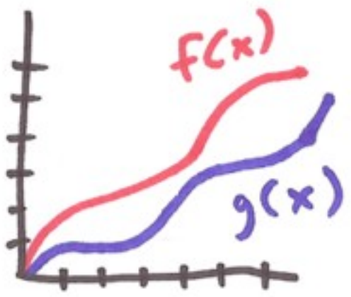
derived attribute: computed from originals

simple change of type

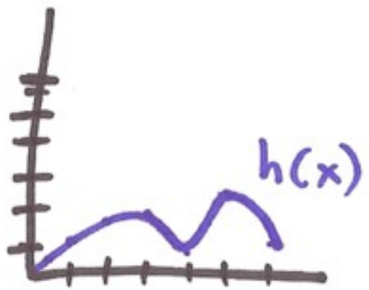
acquire additional data

complex transformation

transformation is abstraction choice



original data



h(x) = f(x) - g(x)
derived data

DATA MODEL vs CONCEPTUAL MODEL

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)