

# Data Visualization

---

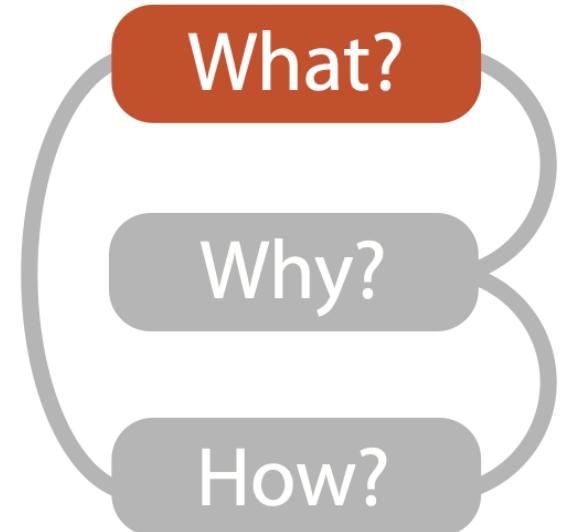
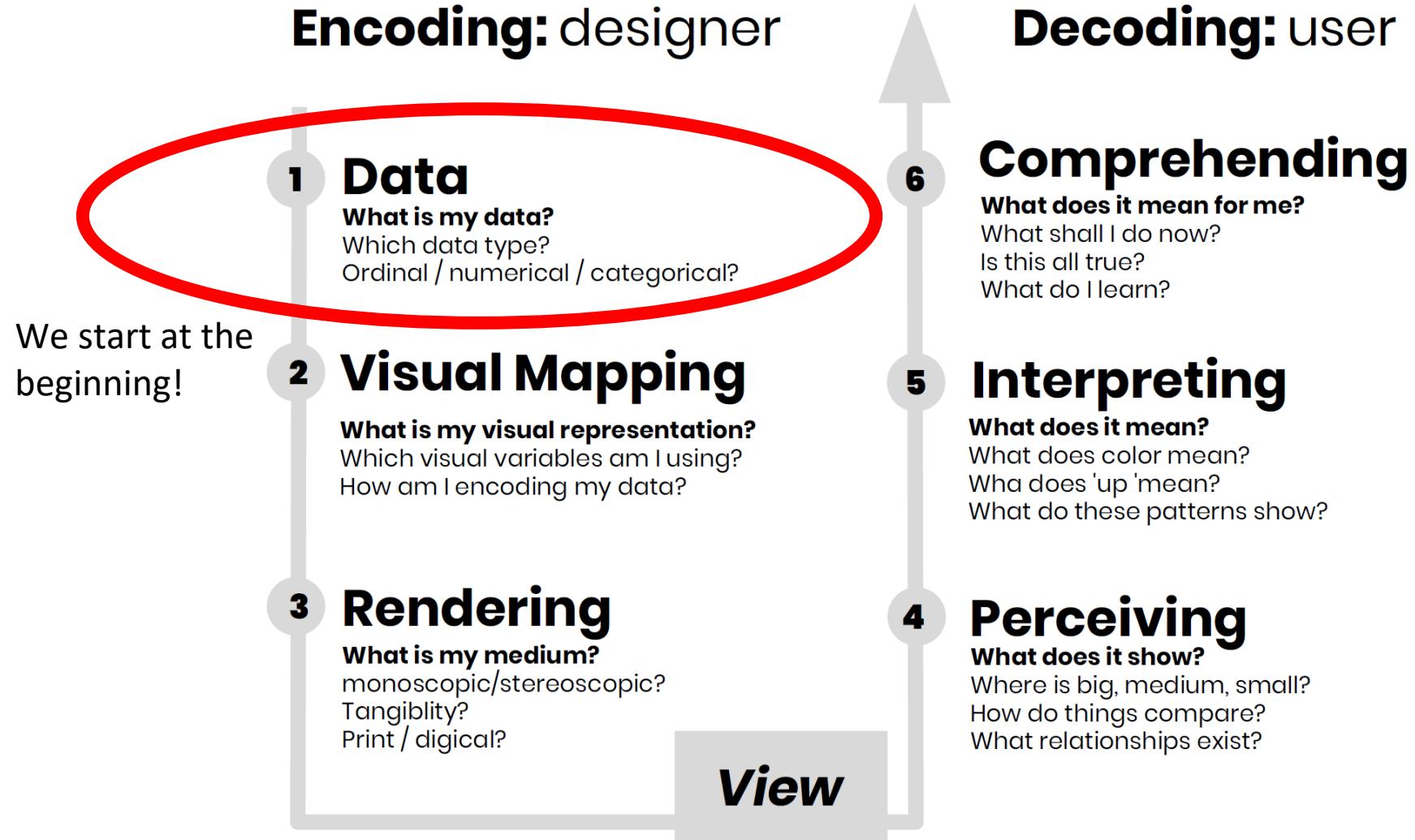
Data

Dr. Claudius Zelenka

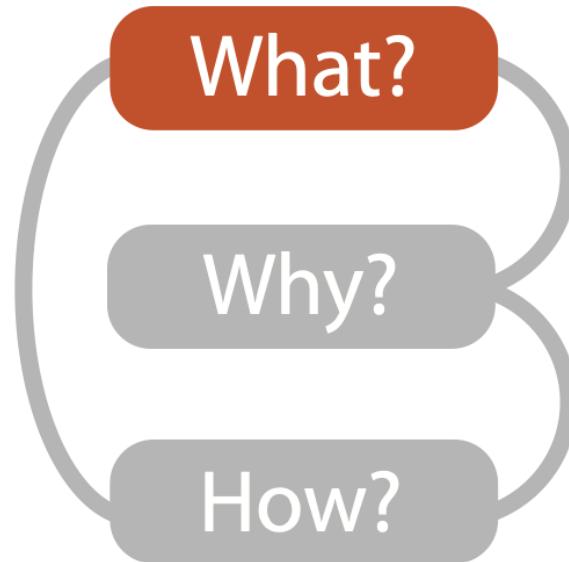
Kiel University

cze@informatik.uni-kiel.de

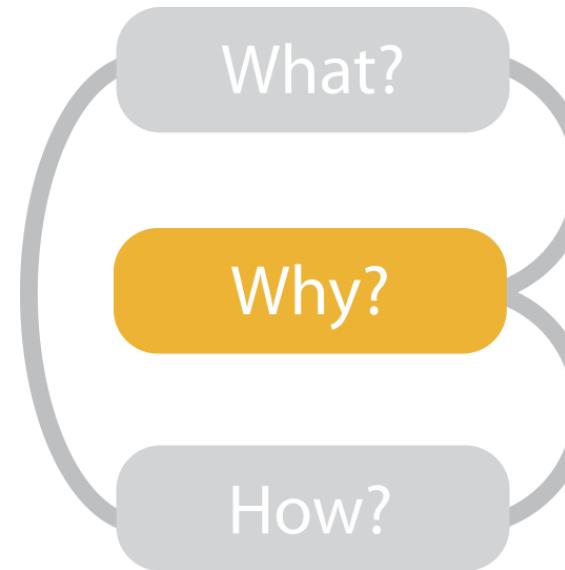
# Visualization Pipeline



# Todays Agenda



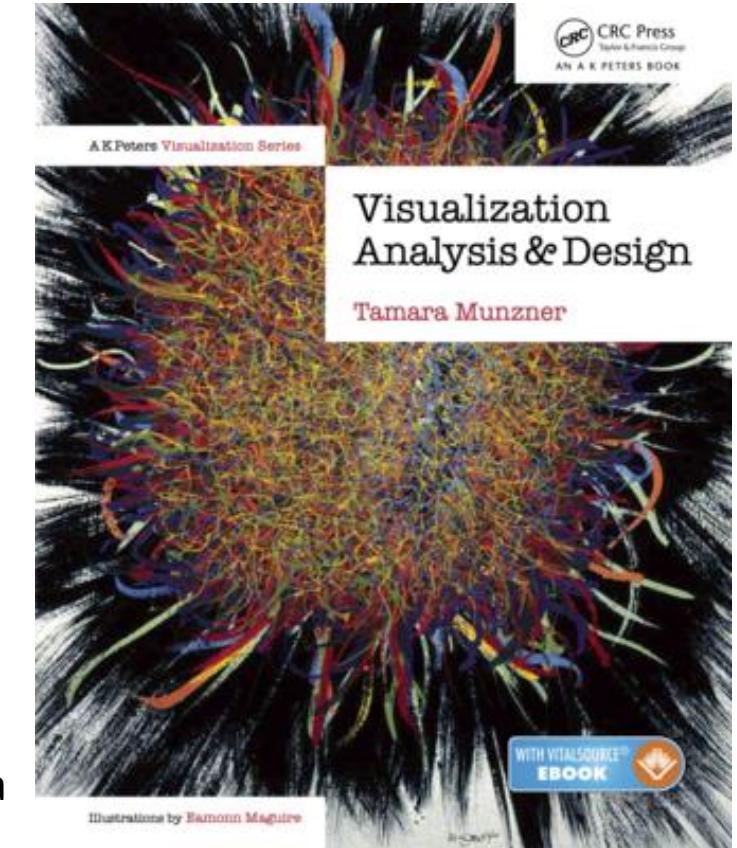
Part 1



Part 2

These part of the lecture is based on  
Chapter 2 and 3 in **Visualization Analysis & Design** and slides by  
**Tamara Munzner**, Department of Computer Science, University of British Columbia

Highly recommended!

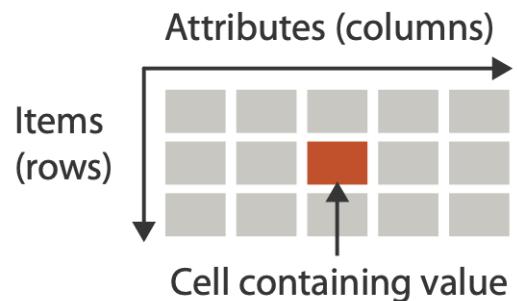


# Three major datatypes

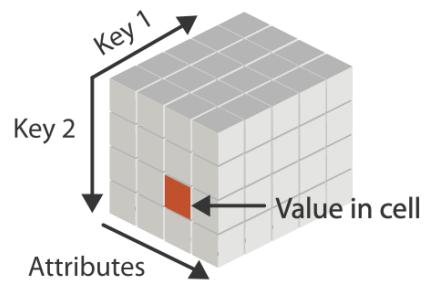
# Three major datatypes

## → Dataset Types

### → Tables

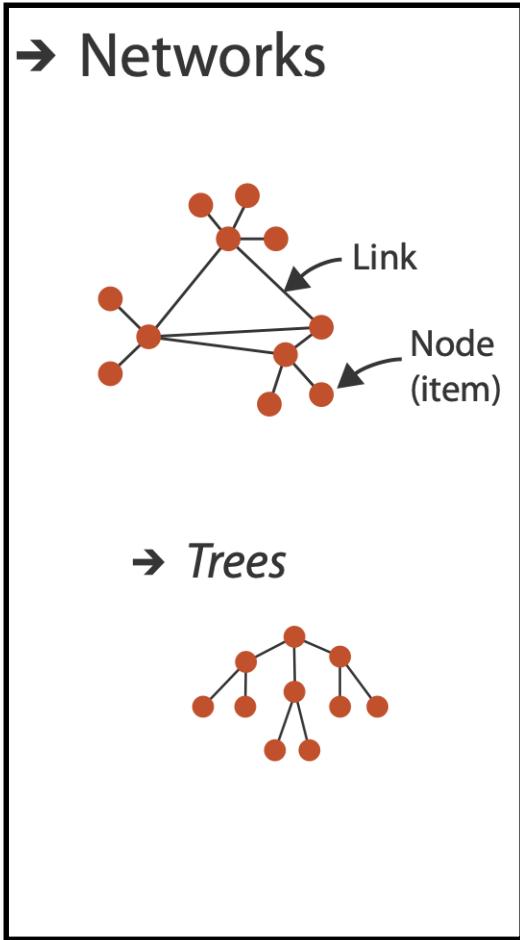
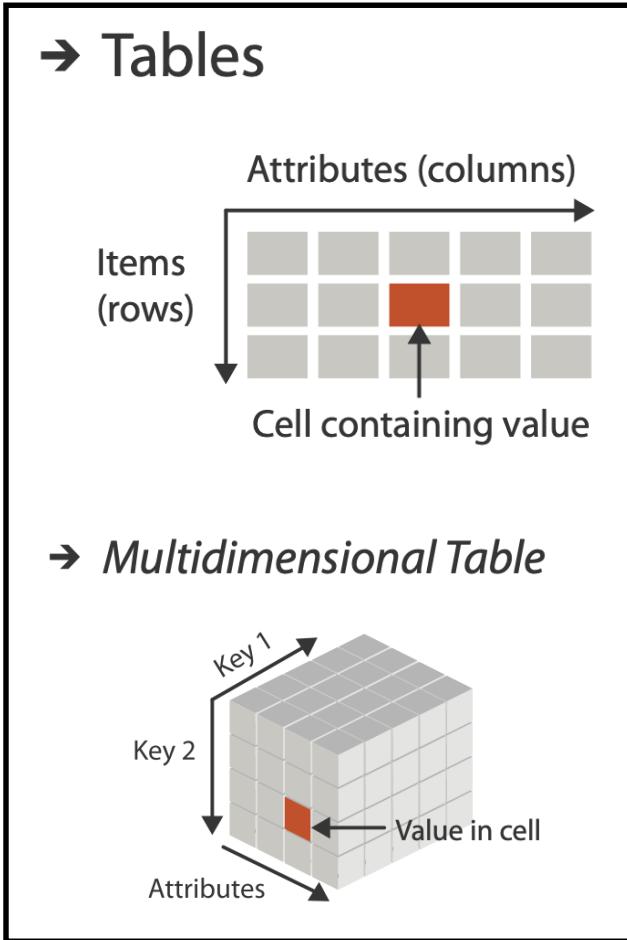


### → *Multidimensional Table*



# Three major datatypes

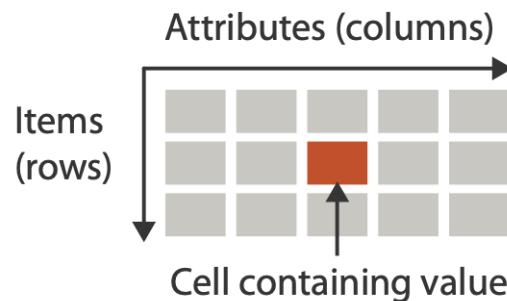
## → Dataset Types



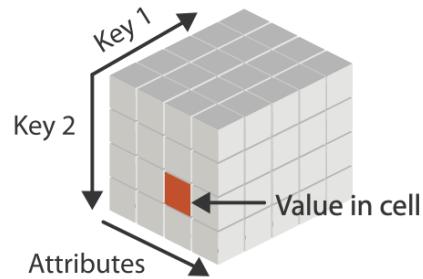
# Three major datatypes

## → Dataset Types

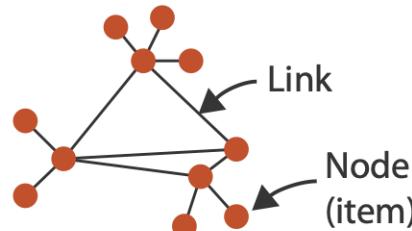
### → Tables



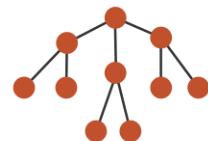
### → Multidimensional Table



### → Networks

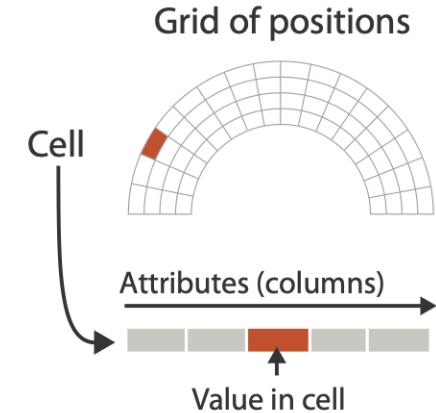


### → Trees

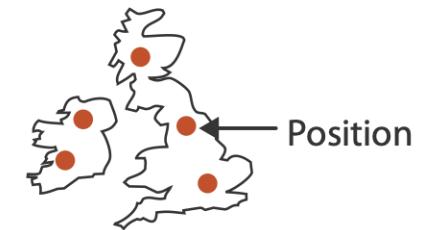


### → Spatial

#### → Fields (Continuous)



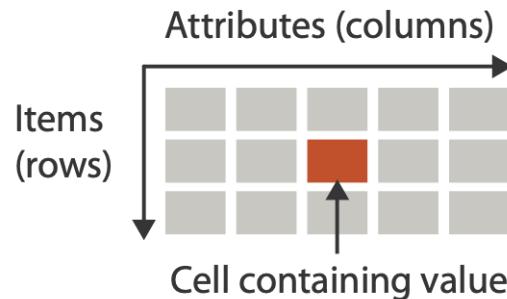
#### → Geometry (Spatial)



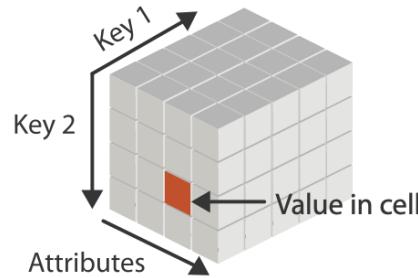
# Three major datatypes

## → Dataset Types

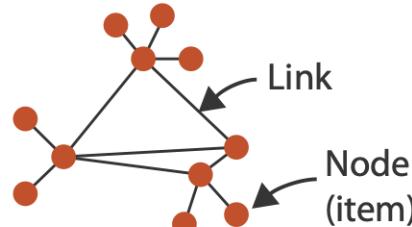
### → Tables



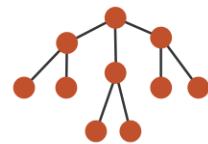
### → Multidimensional Table



### → Networks

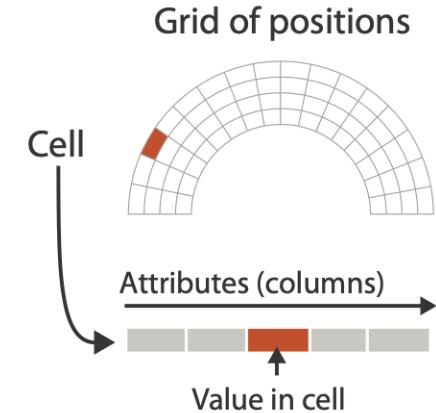


### → Trees

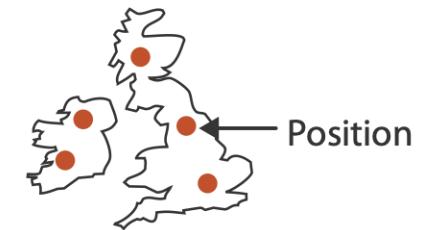


### → Spatial

### → Fields (Continuous)



### → Geometry (Spatial)



- visualization vs computer graphics
  - geometry is design decision

# Attribute types

## → Attribute Types

→ Categorical

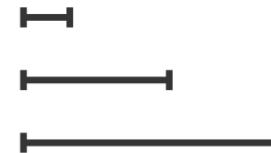


→ Ordered

→ *Ordinal*



→ *Quantitative*



## → Ordering Direction

→ Sequential



→ Diverging



→ Cyclic



# 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	Items

## Attributes

### → Attribute Types

→ Categorical



→ Ordered

→ Ordinal

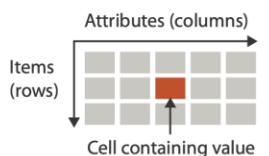


→ Quantitative

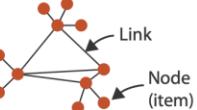


### → Dataset Types

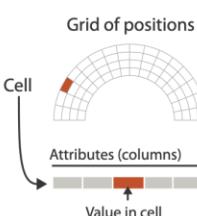
→ Tables



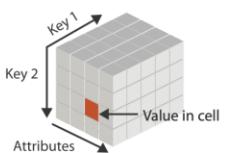
→ Networks



→ Fields (Continuous)



→ Multidimensional Table



→ Trees



→ Geometry (Spatial)



### → Ordering Direction

→ Sequential



→ Diverging



→ Cyclic



### → Dataset Availability

→ Static



→ Dynamic



# What does data mean?

# What does data mean?

14, 2.6, 30, 30, 15, 100001

- What does this sequence of six numbers mean?

# 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?

# 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?

# 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??

# 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 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?

# What does data mean?

14, 2.6, 30, 30, 15, 100001

condition on 7th day of month

- 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

# 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

# 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

# Now what?

- semantics: real-world meaning

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

# Now what?

- semantics: real-world meaning

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

# Now what?

- semantics: real-world meaning
- data types: structural or mathematical interpretation of data
  - item, link, attribute, position, (grid)
  - different from data types in programming!

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

# Items & Attributes

- item: individual entity, discrete
  - eg patient, car, stock, city
  - "independent variable"

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

# Items & Attributes

- item: individual entity, discrete
  - eg patient, car, stock, city
  - "independent variable"

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

item: person

# Items & Attributes

- item: individual entity, discrete
  - eg patient, car, stock, city
  - "independent variable"
- attribute: property that is measured, observed, logged...
  - eg height, blood pressure for patient
  - eg horsepower, make for car
  - "dependent variable"

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

item: person

# Items & Attributes

- item: individual entity, discrete
  - eg patient, car, stock, city
  - "independent variable"
- attribute: property that is measured, observed, logged...
  - eg height, blood pressure for patient
  - eg horsepower, make for car
  - "dependent variable"

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

item: person

# 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

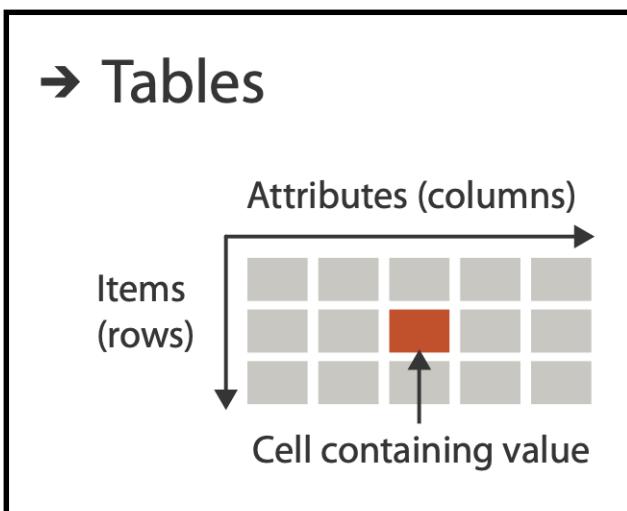
# Dataset types

## Tables

Items

Attributes

- flat table
  - one item per row
  - each column is attribute
  - cell holds value for item-attribute pair



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

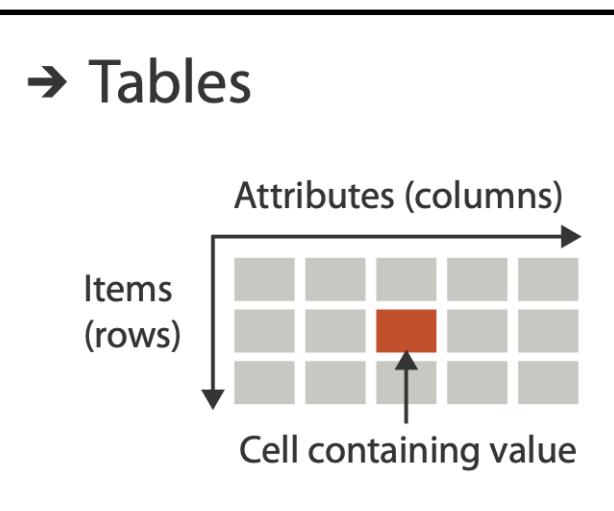
item: person

# Dataset types

## Tables

Items

Attributes



- flat table
  - one item per row
  - each column is attribute
  - cell holds value for item-attribute pair
  - unique key (could be implicit)

attributes: name, age, shirt size, fave fruit

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

item: person

Table

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	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

Table

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
item	32	7/16/07	Medium Box	0.65	7/18/07
	35	10/23/07	Wrap Bag	0.52	10/24/07
	35	10/23/07	Small Box	0.58	10/25/07
	36	11/3/07	Small Box	0.55	11/3/07
	65	3/18/07	Small Pack	0.49	3/19/07
	66	1/20/05	Wrap Bag	0.56	1/20/05
	69	6/4/05	Small Pack	0.44	6/6/05
	69	6/4/05	Wrap Bag	0.6	6/6/05
	70	12/18/06	Small Box	0.59	12/23/06
	70	12/18/06	Wrap Bag	0.82	12/23/06
	96	4/17/05	Small Box	0.55	4/19/05
	97	1/29/06	Small Box	0.38	1/30/06
	129	11/19/08	Small Box	0.37	11/28/08
	130	5/8/08	Small Box	0.37	5/9/08
	130	5/8/08	Medium Box	0.38	5/10/08
	130	5/8/08	Small Box	0.6	5/11/08
	132	6/11/06	Medium Box	0.6	6/12/06
	132	6/11/06	Jumbo Box	0.69	6/14/06
	134	5/1/08	Large Box	0.82	5/3/08
	135	10/21/07	Small Pack	0.64	10/23/07
	166	9/12/07	Small Box	0.55	9/14/07
	193	8/8/06	Medium Box	0.57	8/10/06
	194	4/5/08	Wrap Bag	0.42	4/7/08

Table

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	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

item

attribute

Table

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	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

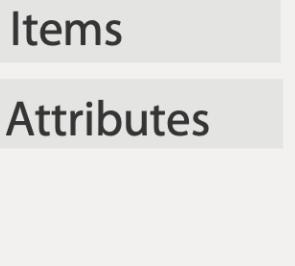
item

cell

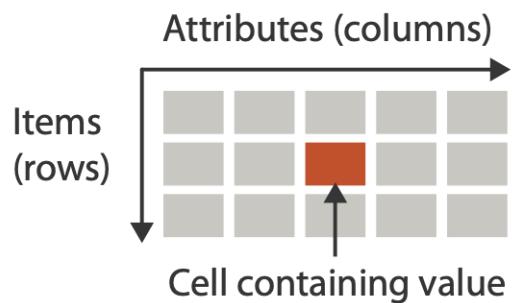
attribute

# Dataset types

## Tables

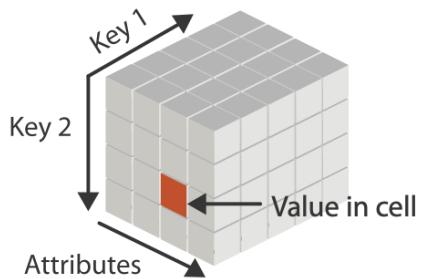


## → Tables



- multidimensional tables
    - indexing based on multiple keys

### → *Multidimensional Table*



	A	B	C	D	E			
1	#	A	B	C	D	E		
2	1	#	A	B	C	D	E	
3	2	1	#	A	B	C	D	E
4	3	2	1	#1.2				
5	4	3	G	2	1500	529		
6	T	5	F	4	L	3	GeneName	DESCRIPTION
7	6	15	P	4	T	3	TCGA-02-0001-01C-01R-0177-01	TCGA-02-0003-01A-01R-0177-01
8	F	7	H	6	LTF	LTF	-1.265728057	2.377012066
9	S	8	R	7	POSTN	POSTN	2.662411805	3.932400324
10	I	9	H	6	TMSL8	TMSL8	-3.082217838	-2.243148513
11	A	10	R	7	HLA-DQA1	HLA-DQA1	-1.739664398	4.577962344
12	I	11	D	9	RP11-35N6.1	RP11-35N6.1	-3.346352968	-2.895400157
13	S	12	STMN2	10	DCX	DCX	-2.578511106	-3.051605144
14	I	13	AGXT2L1	11	-2.26078976	-2.529795801	-3.473035067	-1.729892888
15	F	14	IL13RA2	12	AGXT2L1	AGXT2L1	-2.639493611	-2.844966278
16	M	15	SLN	13	IL13RA2	IL13RA2	-2.93596915	-3.113204863
17	F	16	MEOX2	14	SLN	SLN	-2.466718221	-3.999392588
18	N	17	MEOX2	15	MEOX2	MEOX2	-2.395054066	-2.028406749
19	C	18	COL11A1	16	COL11A1	COL11A1	-1.211934832	-1.026276046
20	F	19	NNMT	17	NNMT	NNMT	0.703745164	-1.783235317
21	C	20	F13A1	18	F13A1	F13A1	-0.224094042	0.664082419
22	M	21	CXCL14	19	CXCL14	CXCL14	-3.1309694	0.222197544
23	K	22	MBP	20	MBP	MBP	-1.906390566	-1.395056071
24	T	23	TF	21	TF	TF	-4.334123292	-2.037626447
25	G	24	KCND2	22	KCND2	KCND2	-1.777692395	-4.680680246
26							-2.100362021	-2.975788666
27							-1.996306032	-2.975788666

# Dataset types

Tables

Items

Attributes

Networks &  
Trees

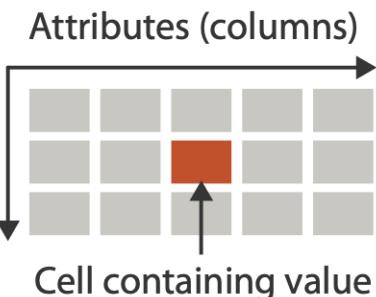
Items (nodes)

Links

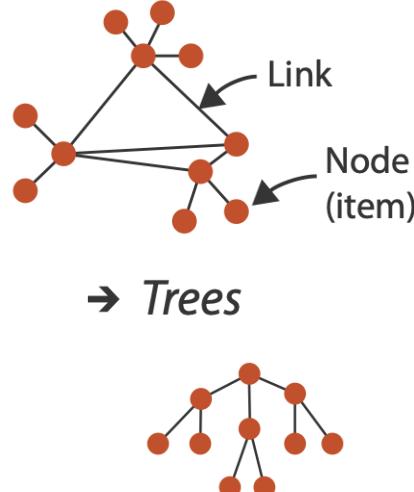
Attributes

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

→ Tables



→ Networks



# Dataset types

## Tables

## Networks & Trees

## Fields

Items

Attributes

Items (nodes)

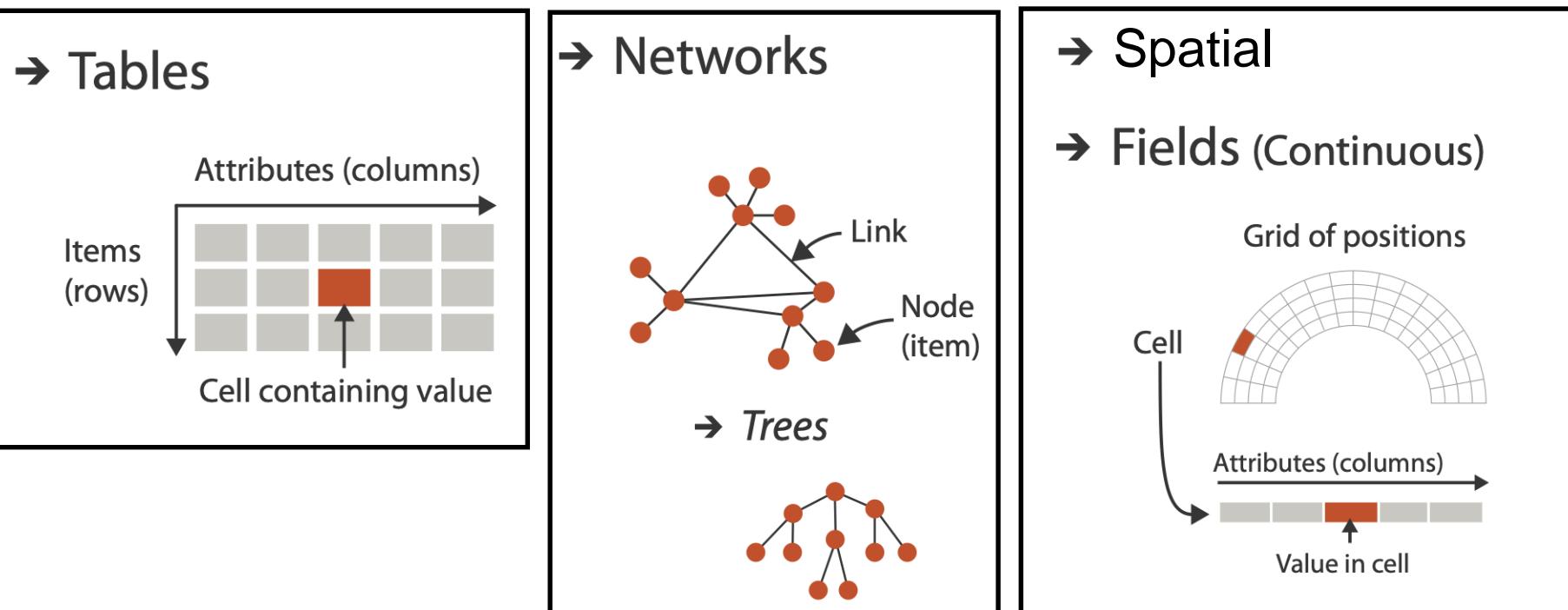
Links

Attributes

Grids

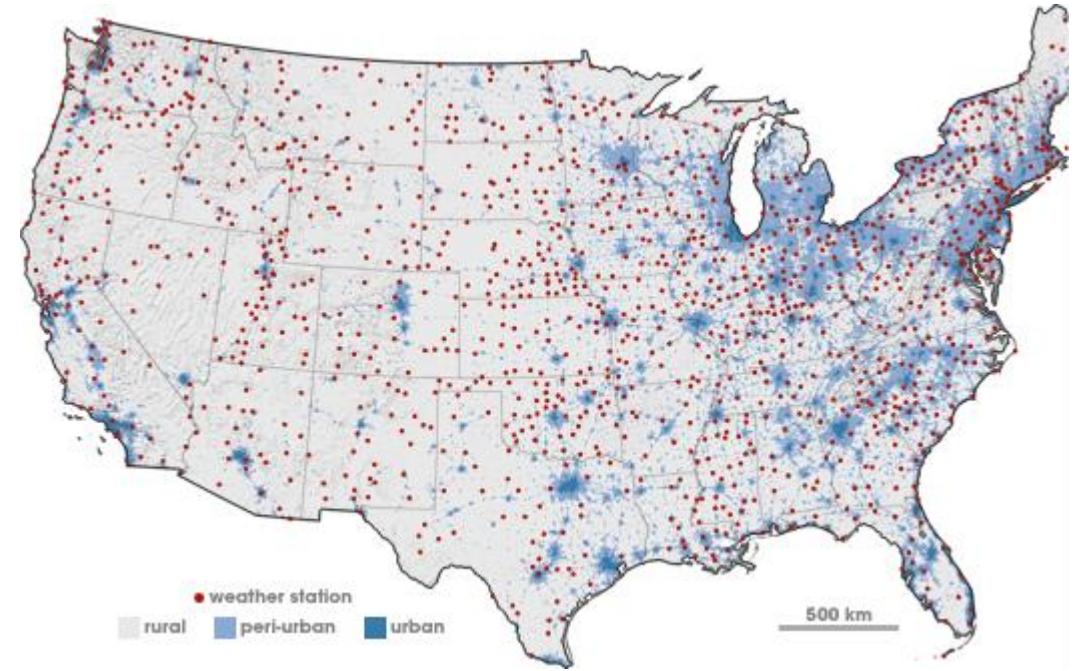
Positions

Attributes

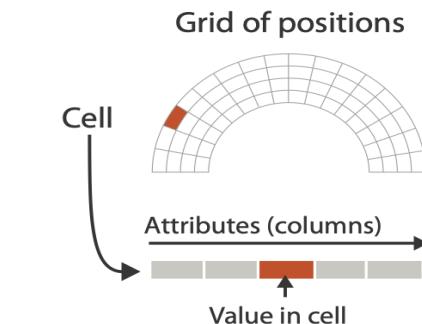


# Spatial fields

- attribute values associated w/ cells
- cell contains value from continuous domain
  - Eg. temperature, pressure, wind velocity
- measured or simulated

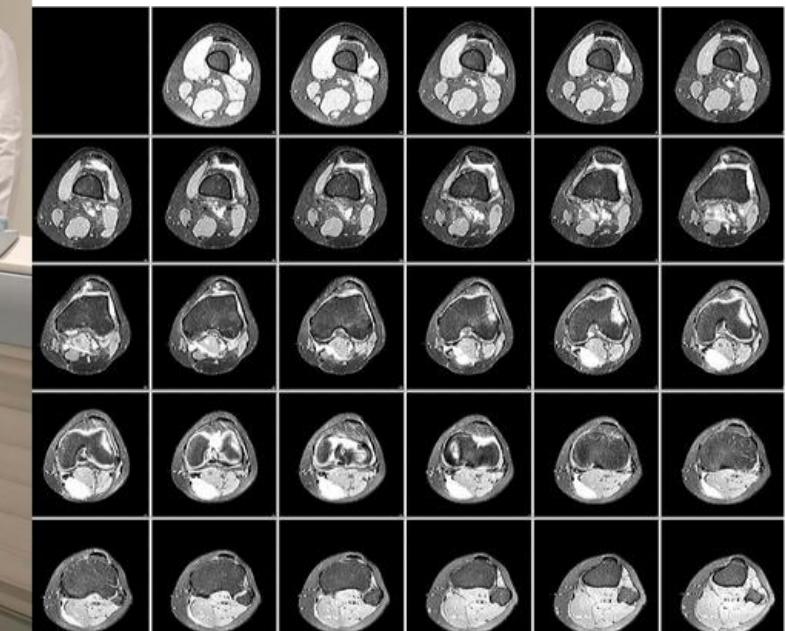
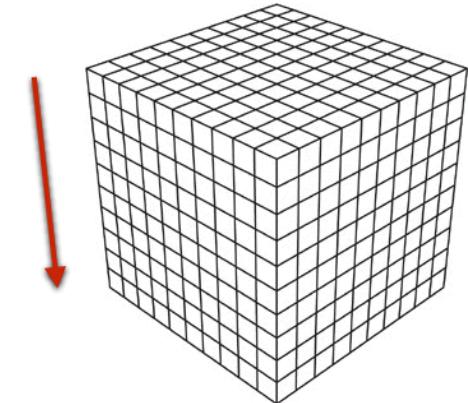


→ Spatial  
→ Fields (Continuous)



# Spatial fields

- attribute values associated w/ cells
- cell contains value from continuous domain
  - eg temperature, pressure, wind velocity
- measured or simulated
- major concerns
  - sampling: where attributes are measured
  - interpolation: how to model attributes elsewhere
  - grid types



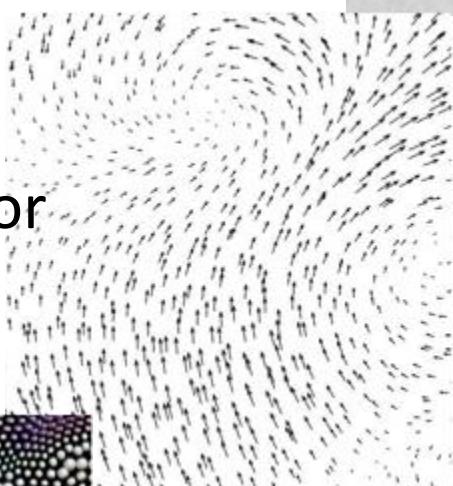
# Spatial fields

- attribute values associated w/ cells
- cell contains value from continuous domain
  - eg temperature, pressure, wind velocity
- measured or simulated
- major concerns
  - sampling:  
where attributes are measured
  - interpolation:  
how to model attributes elsewhere
  - grid types
- major divisions
  - attributes per cell:  
scalar (1), vector (2), tensor (many)

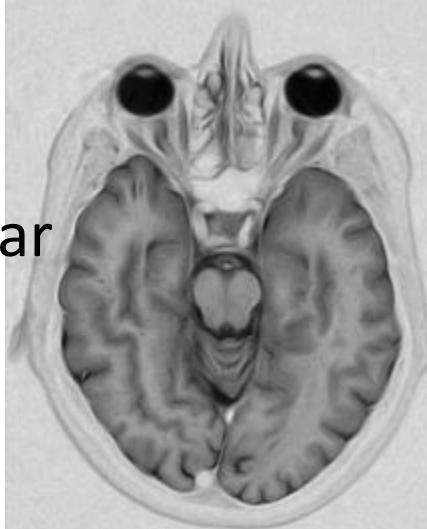
tensor



vector



scalar



# Dataset types

## Tables

Items

Attributes

## Networks & Trees

Items (nodes)

Links

Attributes

## Fields

Grids

Positions

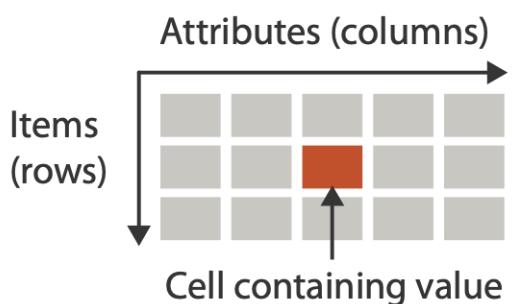
Attributes

## Geometry

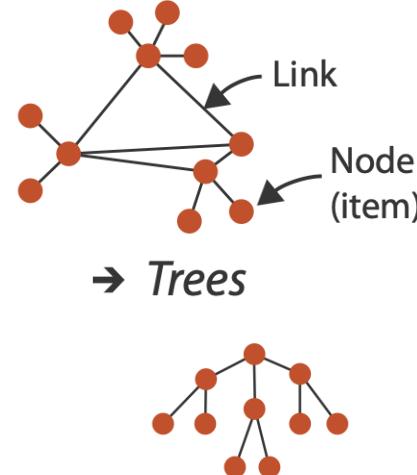
Items

Positions

### → Tables

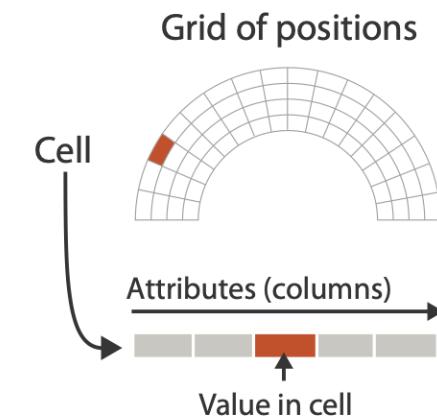


### → Networks

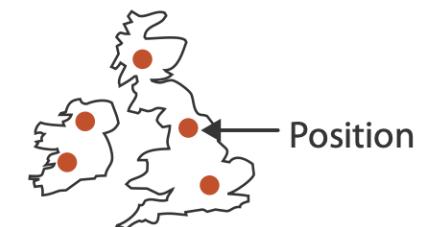


### → Spatial

### → Fields (Continuous)



### → Geometry (Spatial)



# Geometry

- shape of items
- explicit spatial positions / regions
  - points, lines, curves, surfaces, volumes
- boundary between computer graphics and visualization
  - graphics: geometry taken as given
  - vis: geometry is result of a design decision



# Dataset types

## Tables

Items

Attributes

## Networks & Trees

Items (nodes)

Links

Attributes

## Fields

Grids

Positions

Attributes

## Geometry

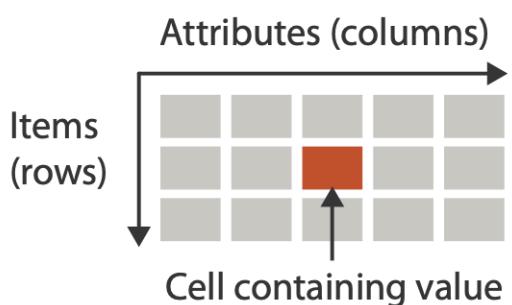
Items

Positions

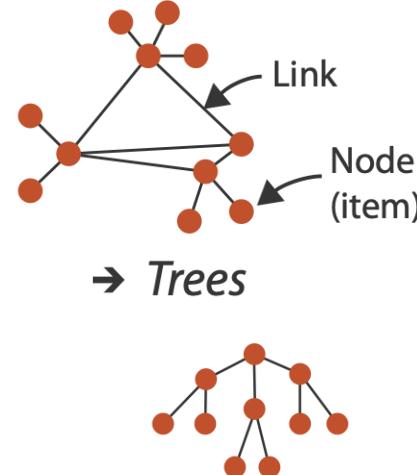
## Clusters, Sets, Lists

Items

### → Tables

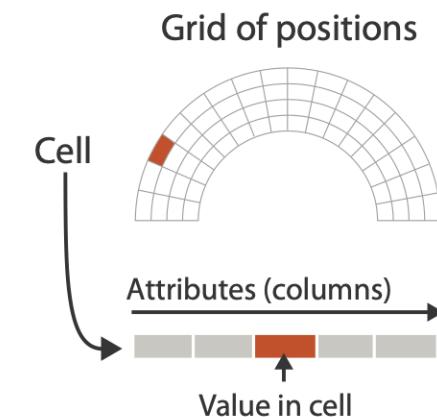


### → Networks

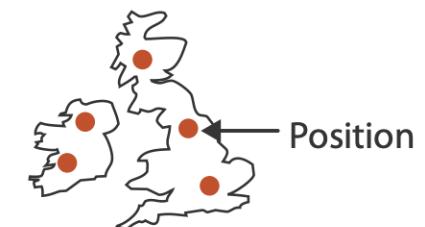


### → Spatial

### → Fields (Continuous)



### → Geometry (Spatial)

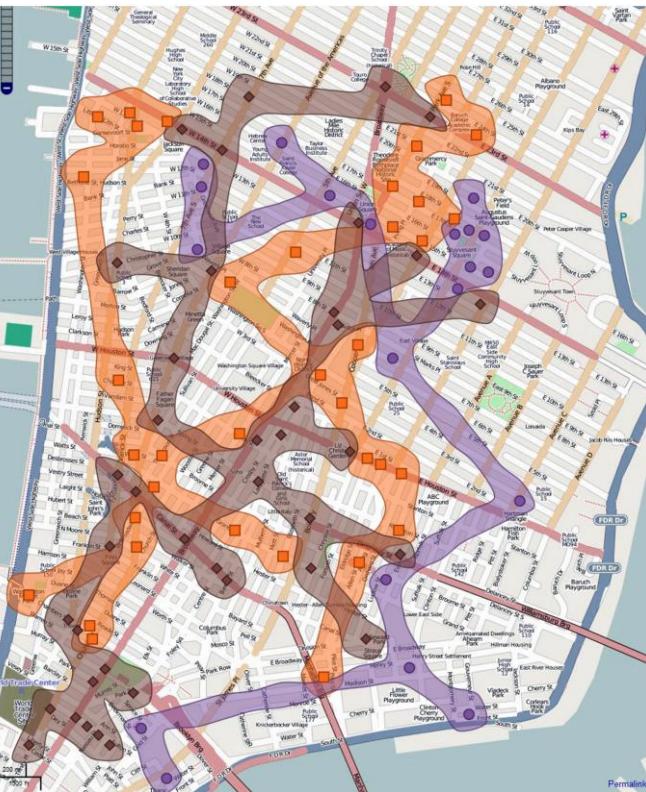


# Collections

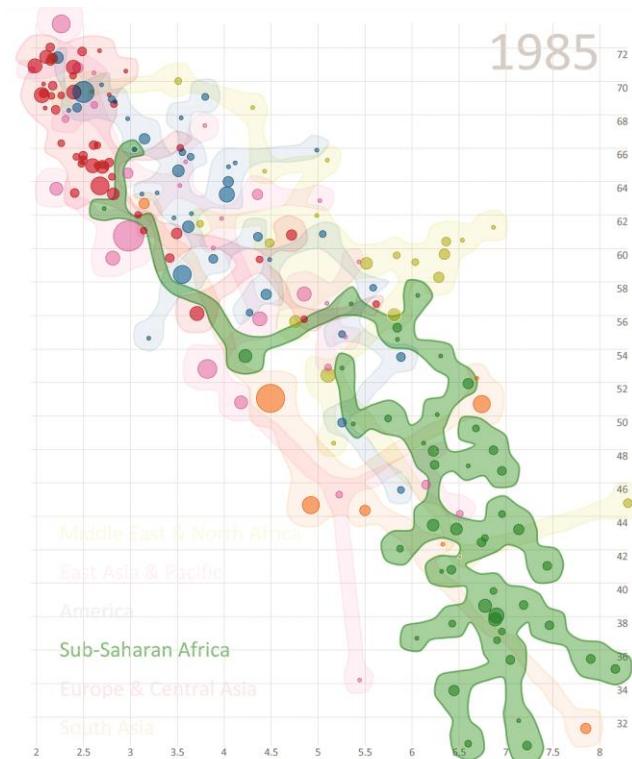
- how we group items

# Collections

- how we group items
- sets
  - unique items, unordered

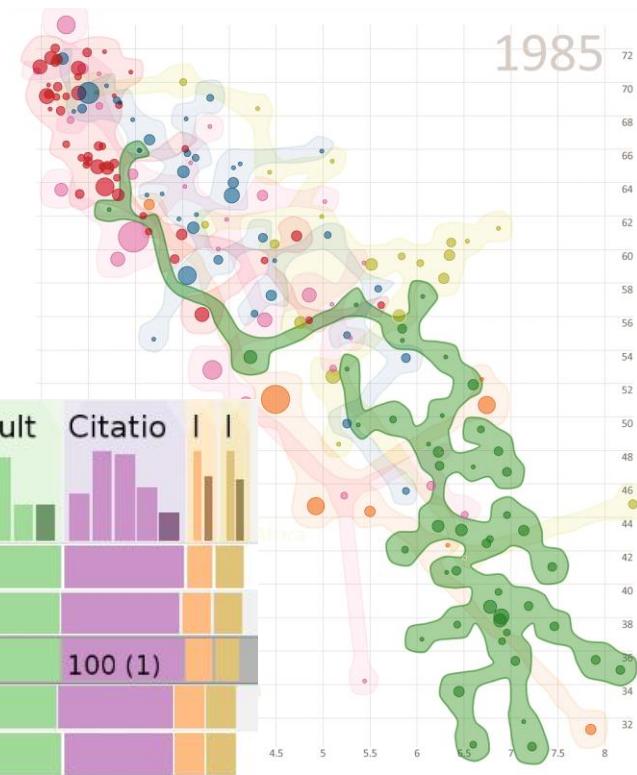
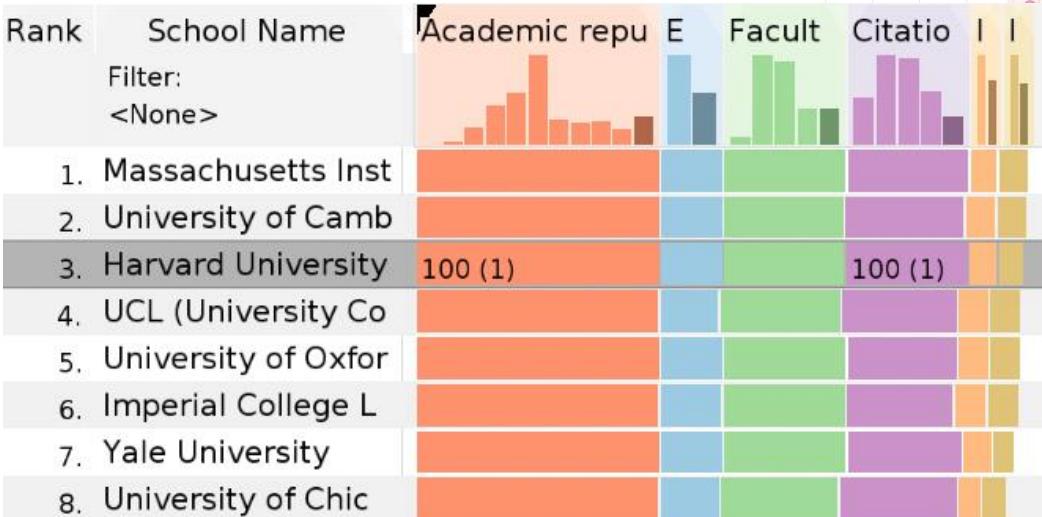


*Bubble Sets: Revealing Set Relations with Isocontours over Existing Visualizations Christopher Collins et. al.*



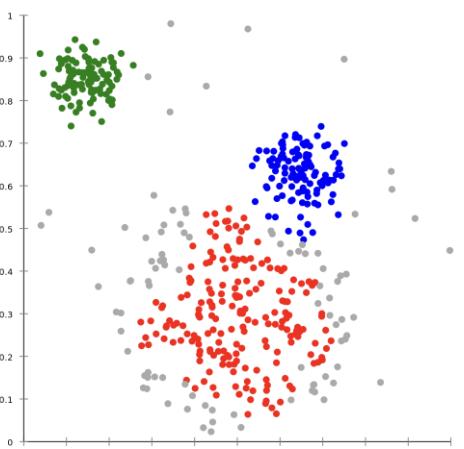
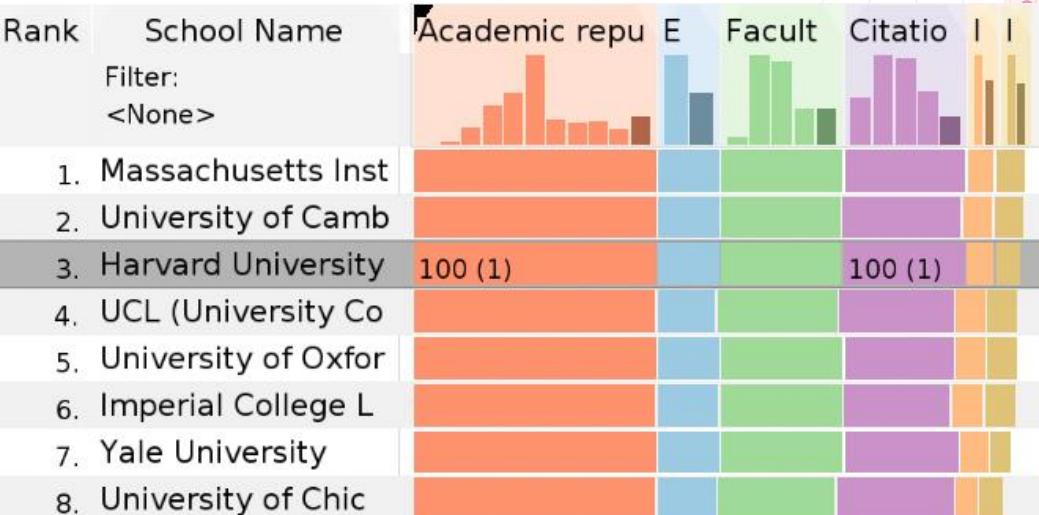
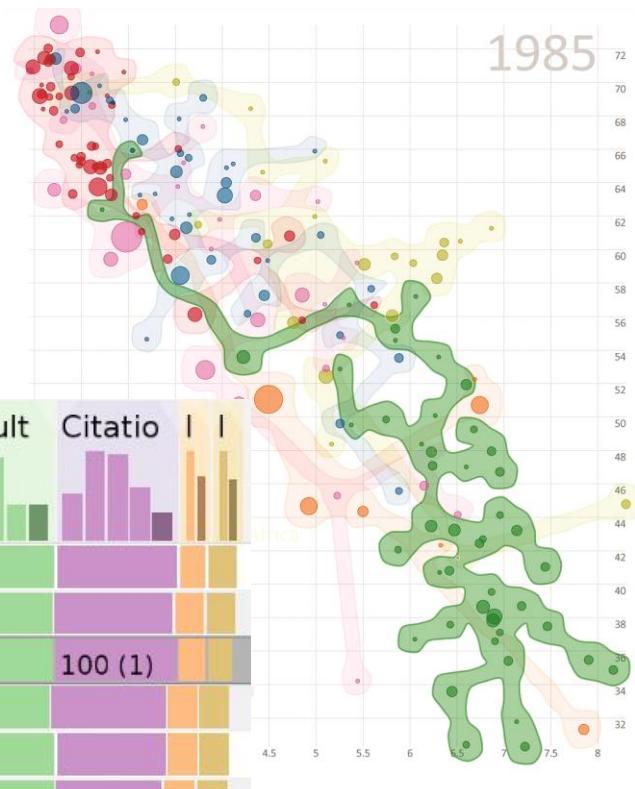
# Collections

- how we group items
- sets
  - unique items, unordered
- lists
  - ordered, duplicates possible



# Collections

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



# Dataset and data types

## → Data and Dataset Types

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

## → Data Types

→ Items

→ Attributes

→ Links

→ Positions

→ Grids

# 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

## ➔ Attribute Types

➔ Categorical

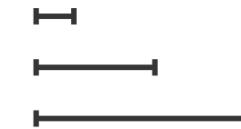


➔ Ordered

➔ *Ordinal*



➔ *Quantitative*



Table

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	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

	A	B	C	S	T	U
	Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
categorical	3	10/14/06	5-Low	Large Box	0.8	10/21/06
ordinal	6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
quantitative	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	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

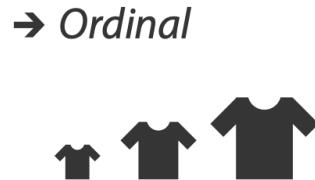
# Other data concerns

## → Attribute Types

→ Categorical

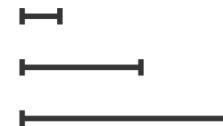


→ Ordered



→ Ordinal

→ Quantitative



## → Ordering Direction

→ Sequential



→ Diverging



→ Cyclic



## → Dataset Availability

→ Static



→ Dynamic



# Data abstraction: Three operations

- translate from domain-specific language to generic visualization language
- identify dataset type(s), attribute types
- identify cardinality
  - how many items in the dataset?
  - what is cardinality of each attribute?
    - number of levels for categorical data
    - range for quantitative data
- consider whether to transform data
  - guided by understanding of task

# Data vs conceptual models

- data model
  - mathematical abstraction
    - sets with operations, eg floats with \* / - +
    - variable data types in programming languages
- conceptual model
  - mental construction (semantics)
  - supports reasoning
  - typically based on understanding of tasks [stay tuned!]
- data abstraction process relies on conceptual model
  - for transforming data if needed

# Data vs conceptual model, example

# Data vs conceptual model, example

- data model: floats
  - 32.52, 54.06, -14.35, ...

## Data vs conceptual model, example

- data model: floats
  - 32.52, 54.06, -14.35, ...
- conceptual model
  - temperature

# Data vs conceptual model, example

- data model: floats
  - 32.52, 54.06, -14.35, ...
- conceptual model
  - temperature
- multiple possible data abstractions

# Data vs conceptual model, example

- data model: floats
  - 32.52, 54.06, -14.35, ...
- conceptual model
  - temperature
- multiple possible data abstractions
  - continuous to 2 significant figures: quantitative
    - task: forecasting the weather

# Data vs conceptual model, example

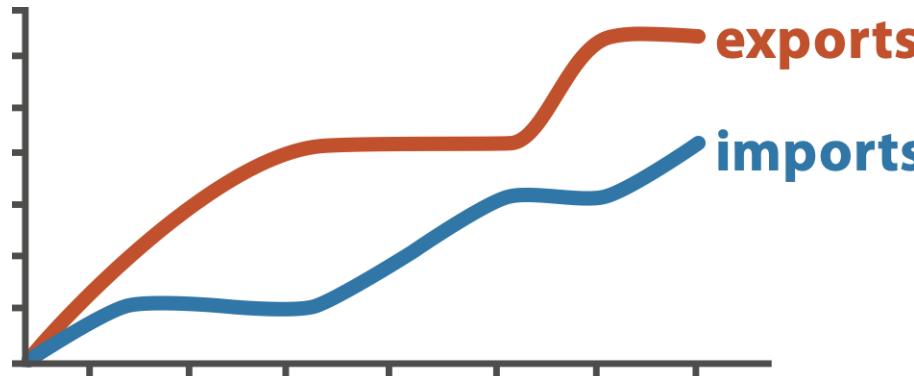
- data model: floats
  - 32.52, 54.06, -14.35, ...
- conceptual model
  - temperature
- multiple possible data abstractions
  - continuous to 2 significant figures: quantitative
    - task: forecasting the weather
  - hot, warm, cold: ordinal
    - task: deciding if bath water is ready

# Data vs conceptual model, example

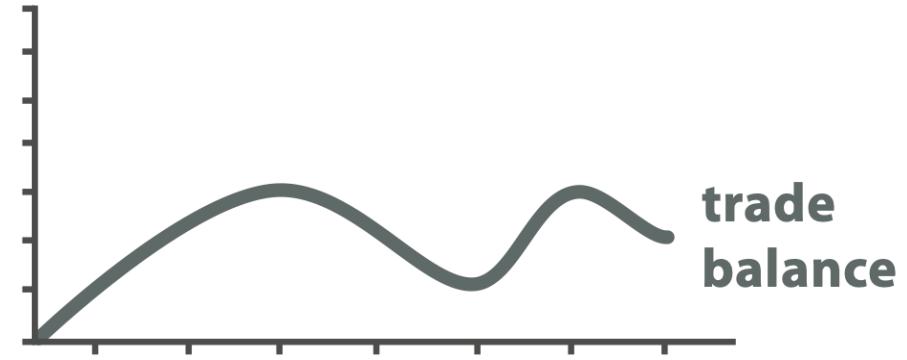
- data model: floats
  - 32.52, 54.06, -14.35, ...
- conceptual model
  - temperature
- multiple possible data abstractions
  - continuous to 2 significant figures: quantitative
    - task: forecasting the weather
  - hot, warm, cold: ordinal
    - task: deciding if bath water is ready
  - above freezing, below freezing: categorical
    - task: decide if I should leave the house today

# Derived attributes

- derived attribute: compute from originals
  - simple change of type
  - acquire additional data
  - complex transformation



Original Data



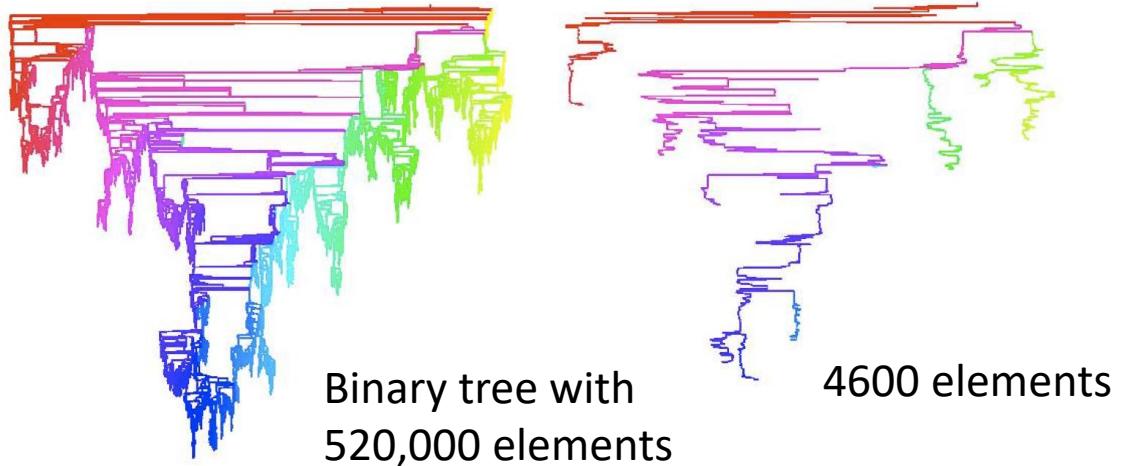
$$\text{trade balance} = \text{exports} - \text{imports}$$

Derived Data

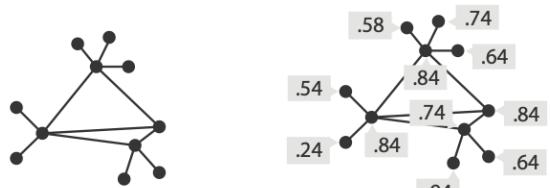
# Analysis example: Derive one attribute

- Strahler number
  - centrality metric for trees/networks
  - derived quantitative attribute
- draw top 5K of 500K for good skeletonization

[Using Strahler numbers for real time visual exploration of huge graphs. Auber. Proc. Intl. Conf. Computer Vision and Graphics, pp. 56–69, 2002.]



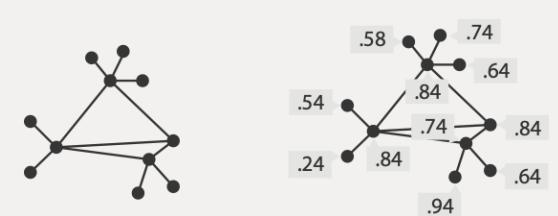
## Task 1



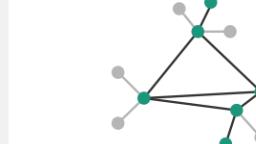
In Tree → Out Quantitative attribute on nodes

What?	Why?
→ In Tree	→ Derive
→ Out Quantitative attribute on nodes	

## Task 2



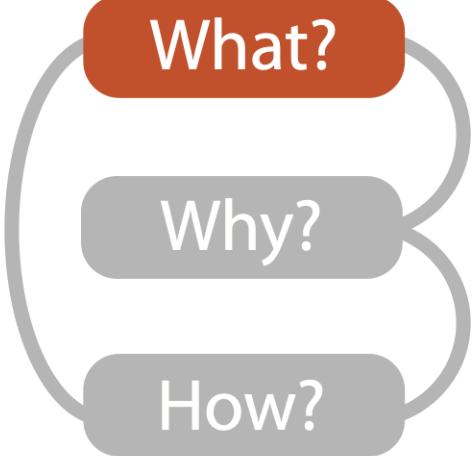
In Tree + In Quantitative attribute on nodes



→ Out Filtered Tree  
Removed unimportant parts

What?	Why?	How?
→ In Tree	→ Summarize	→ Reduce
→ In Quantitative attribute on nodes	→ Topology	→ Filter
→ Out Filtered Tree		

## 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	Clusters, Sets, Lists
Attributes	Links	Positions	Positions	Items

### Attributes

#### → Attribute Types

→ Categorical



→ Ordered

→ Ordinal

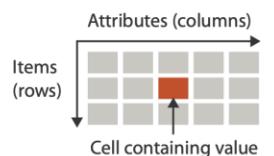


→ Quantitative

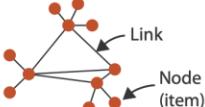


#### → Dataset Types

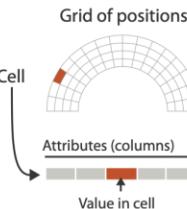
→ Tables



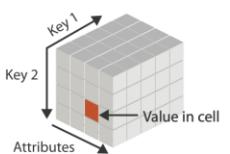
→ Networks



→ Fields (Continuous)



→ Multidimensional Table



→ Trees



→ Geometry (Spatial)



#### → Ordering Direction

→ Sequential



→ Diverging



→ Cyclic



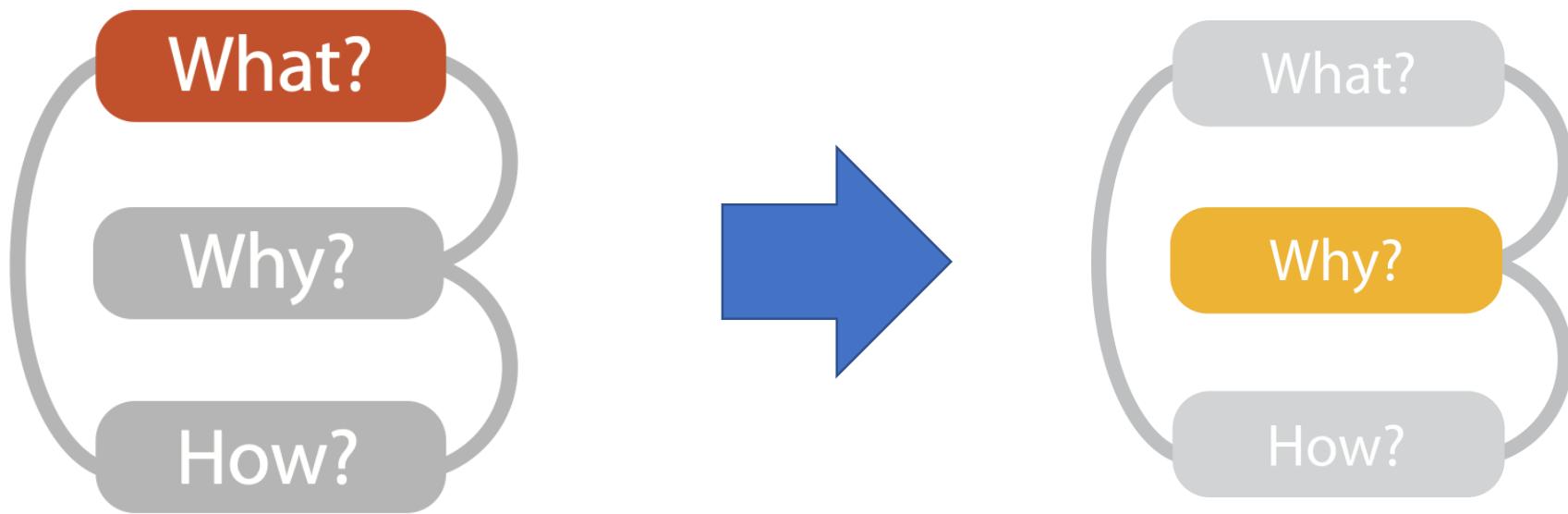
#### → Dataset Availability

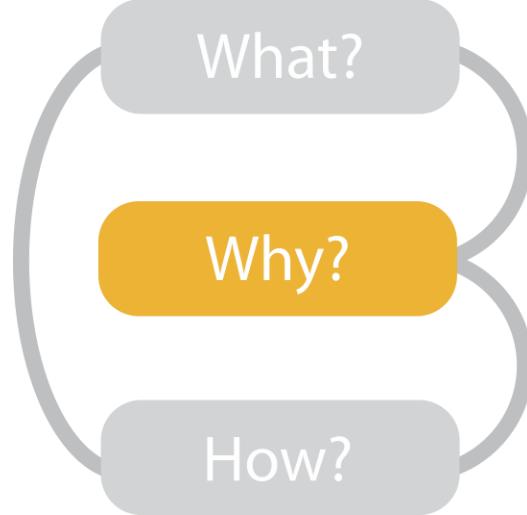
→ Static



→ Dynamic







## Actions

### → Analyze

→ Consume

→ Discover



→ Present



→ Enjoy



→ Produce

→ Annotate



→ Record



→ Derive



### → Search

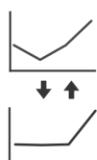
	Target known	Target unknown
Location known	••• <i>Lookup</i>	••• <i>Browse</i>
Location unknown	←•••→ <i>Locate</i>	←•••→ <i>Explore</i>

### → Query

→ Identify



→ Compare



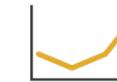
→ Summarize



## Targets

### → All Data

→ Trends



→ Outliers



→ Features



### → Attributes

→ One



→ Many



→ Distribution

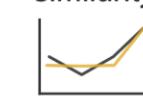
→ Dependency



→ Correlation



→ Similarity



→ Extremes



### → Network Data

→ Topology



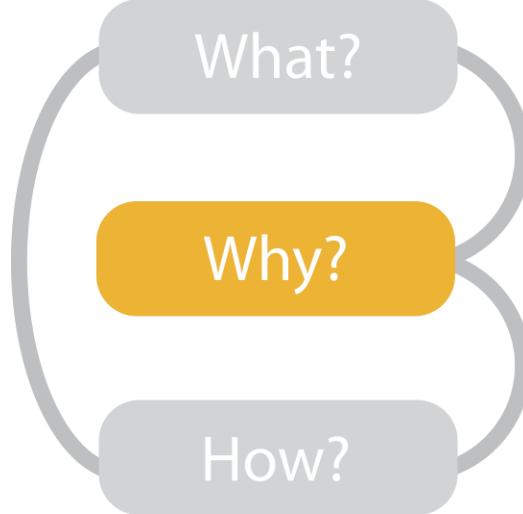
→ Paths



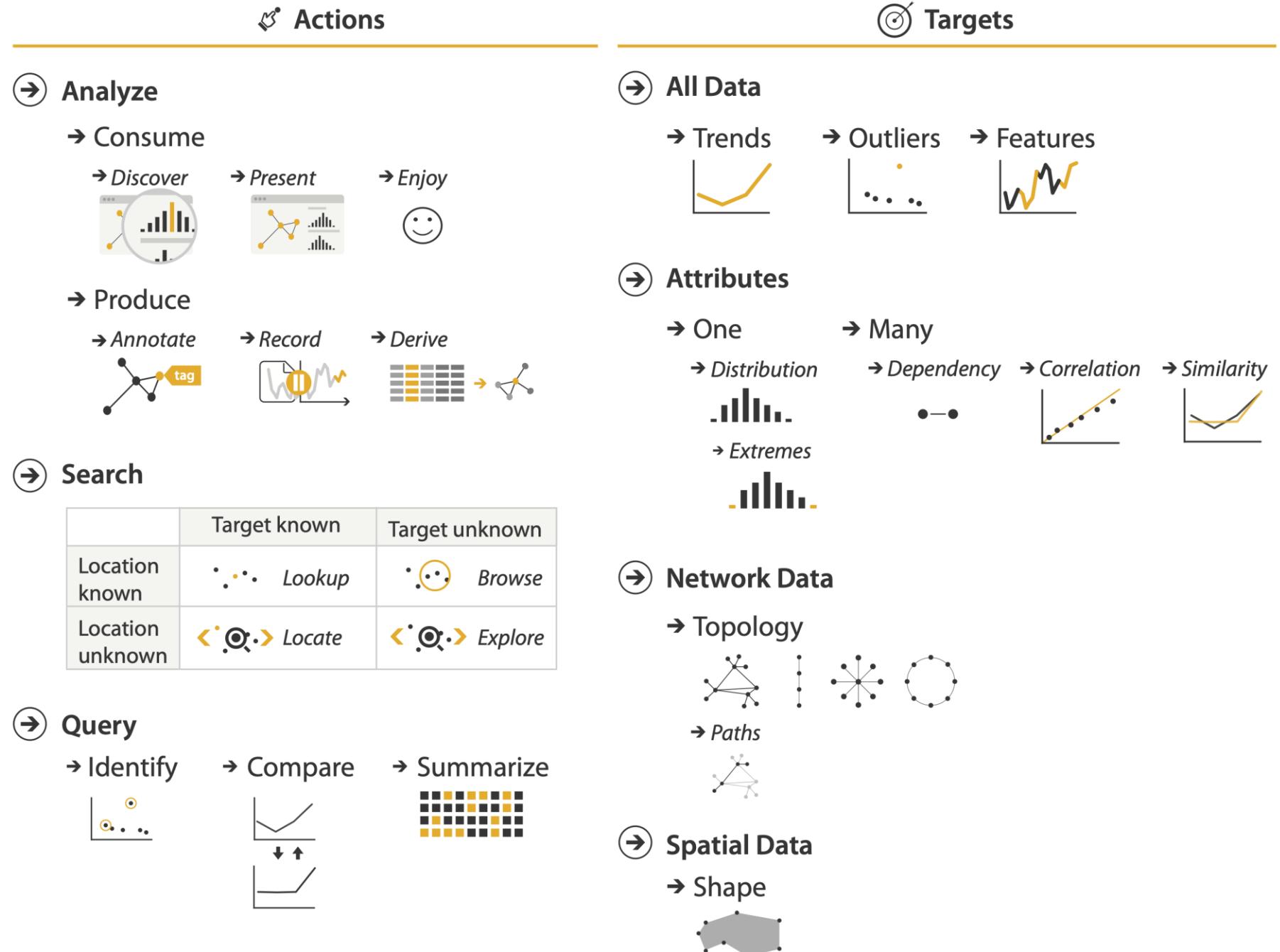
### → Spatial Data

→ Shape



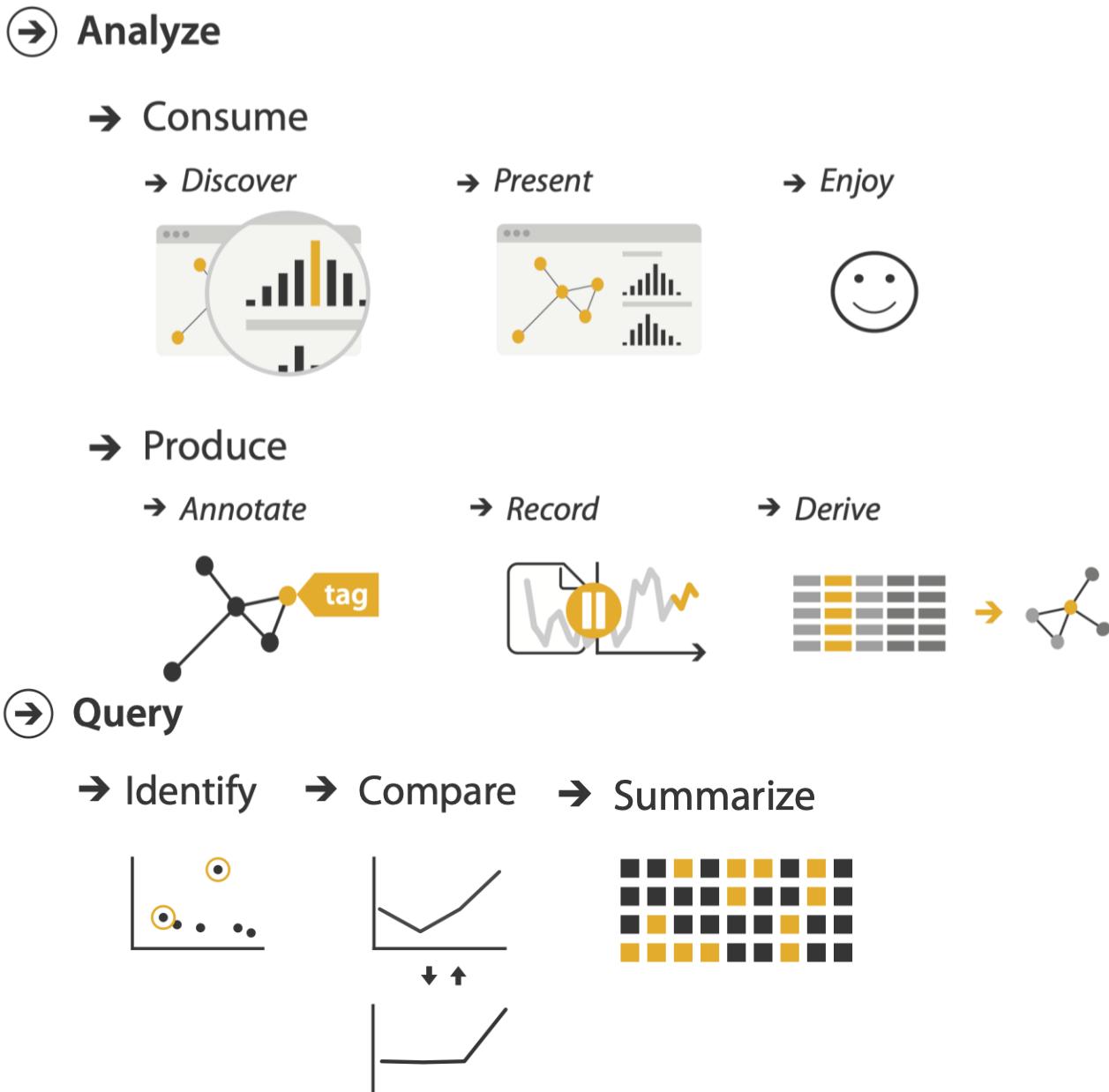


- {action, target} pairs
  - *discover distribution*
  - *compare trends*
  - *locate outliers*
  - *browse topology*



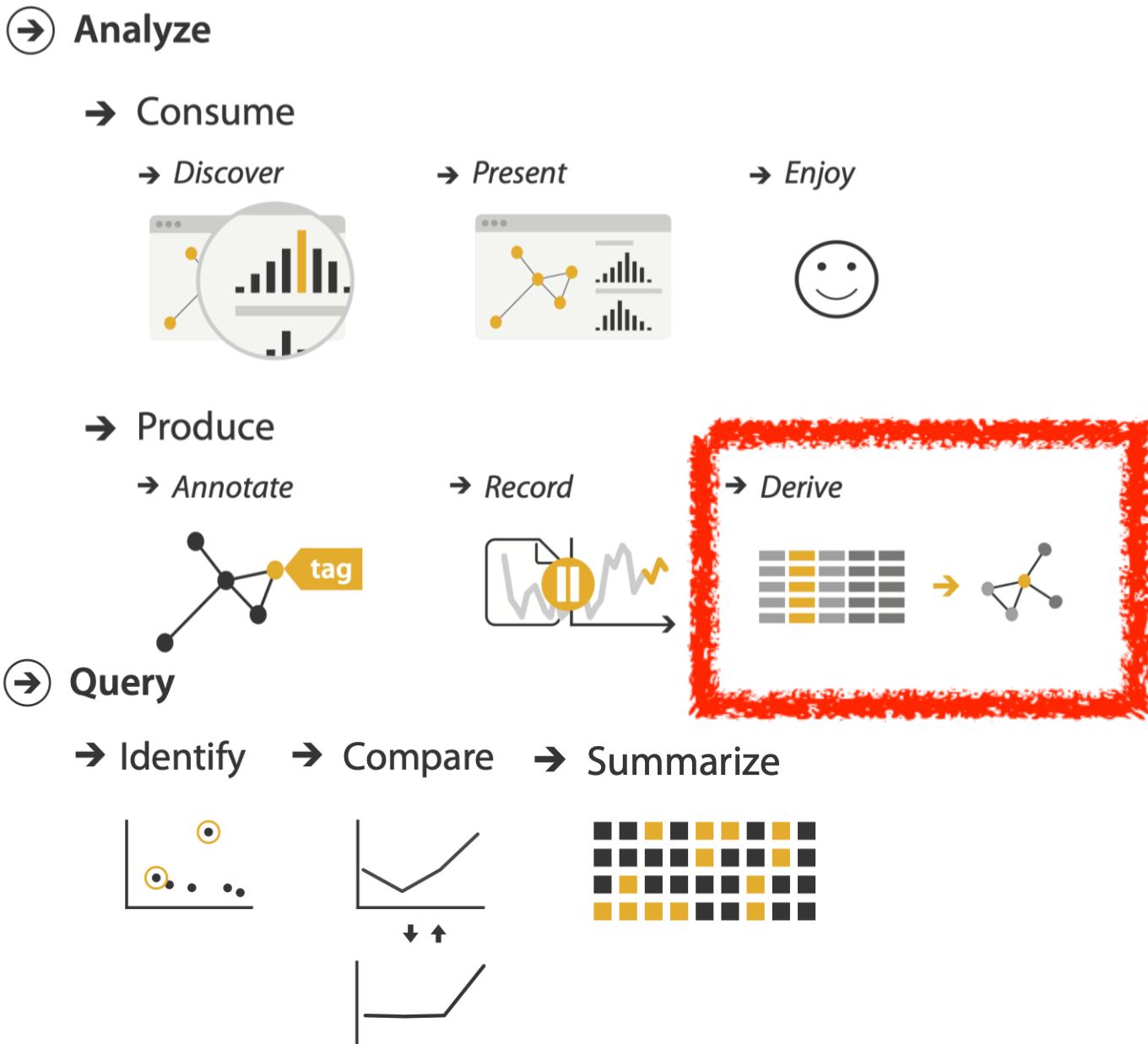
# Actions: Analyze, Query

- analyze
  - consume
    - discover vs present
      - aka explore vs explain
    - enjoy
      - aka casual, social,
  - produce
    - annotate, record, derive
- query
  - how much data matters?
    - one, some, all
- independent choices
  - analyze, query, (search)



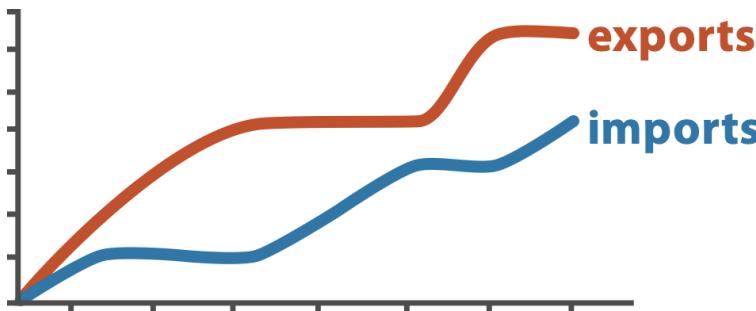
# Actions: Analyze, Query

- analyze
  - consume
    - discover vs present
      - aka explore vs explain
    - enjoy
      - aka casual, social
  - produce
    - annotate, record, derive
- query
  - how much data matters?
    - one, some, all
- independent choices
  - analyze, query, (search)

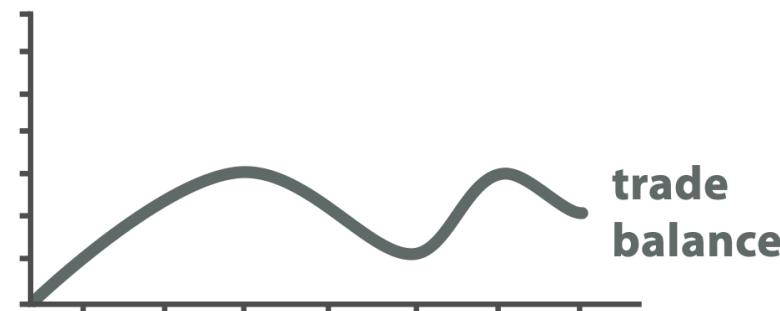


# Derive

- don't necessarily just draw what you're given!
  - decide what the right thing to show is
  - create it with a series of transformations from the original dataset
  - draw that
- one of the four major strategies for handling complexity



Original Data



$$\text{trade balance} = \text{exports} - \text{imports}$$

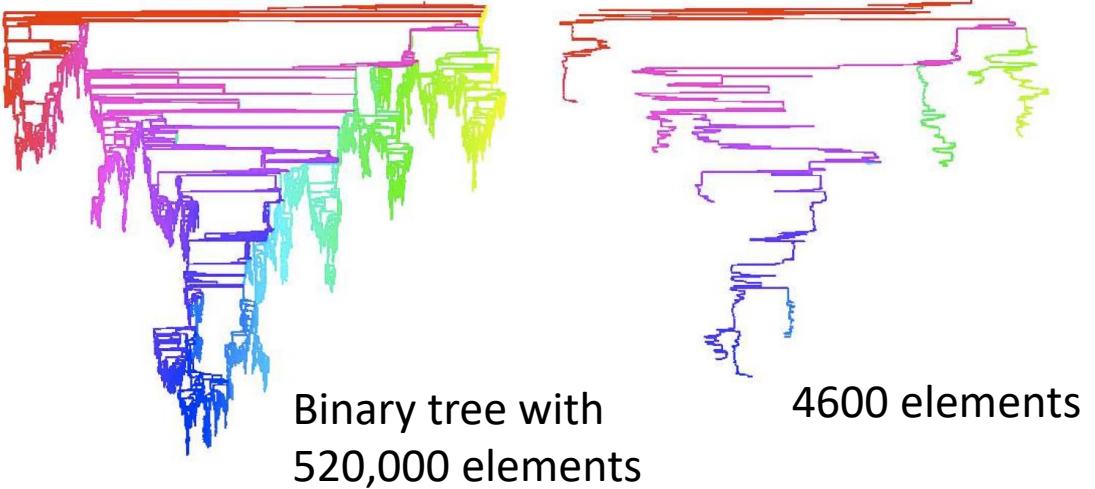
Derived Data

# Analysis example: Derive one attribute

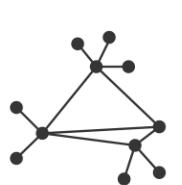
- Strahler number

- centrality metric for trees/networks
- derived quantitative attribute
- draw top 5K of 500K

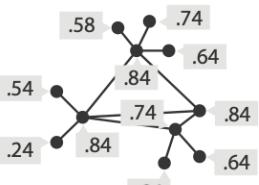
[Using Strahler numbers for real time visual exploration of huge graphs. Auber. Proc. Intl. Conf. Computer Vision and Graphics, pp. 56–69, 2002.]



## Task 1



In Tree



Out Quantitative attribute on nodes

### What?

→ In Tree

→ Out Quantitative attribute on nodes

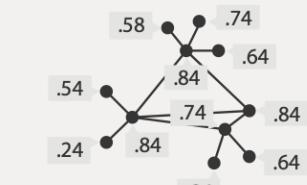
### Why?

→ Derive

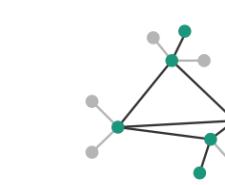
## Task 2



In Tree



In Quantitative attribute on nodes



Out Filtered Tree  
Removed unimportant parts

### What?

### What?

→ In Tree

→ In Quantitative attribute on nodes  
→ Out Filtered Tree

### Why?

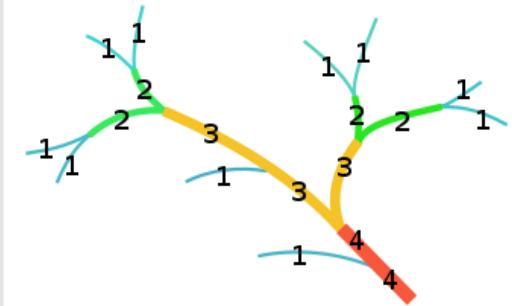
→ Summarize

→ Topology

### How?

→ Reduce

→ Filter



Strahler number

# Why: Targets

## → All Data

→ Trends



→ Outliers



→ Features



## → Attributes

→ One

→ Distribution



→ Extremes

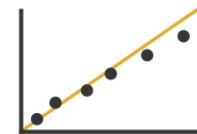


→ Many

→ Dependency



→ Correlation

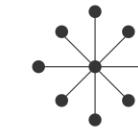


→ Similarity



## → Network Data

→ Topology



→ Paths



## → Spatial Data

→ Shape

