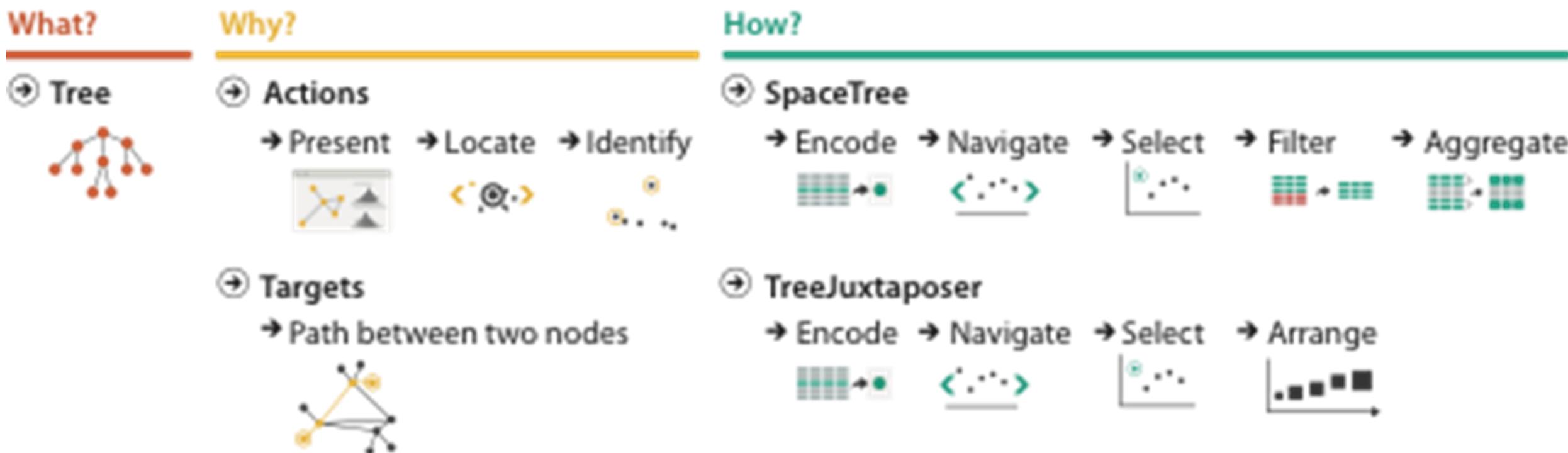
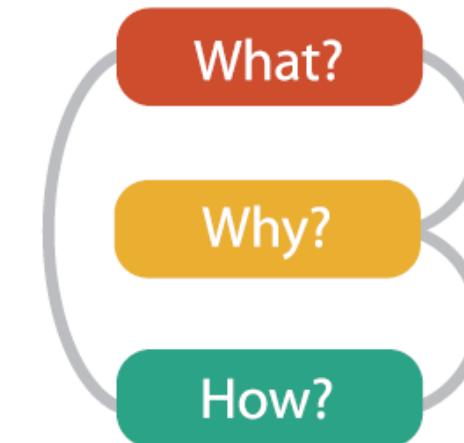


# Information Visualization

## Analysis Framework

Visualization Analysis and Design. Tamara Munzner, with illustrations by Eamonn Maguire. A K Peters Visualization Series, CRC Press, 2014.

- **What** is shown?
  - Data abstraction
- **Why** is the user looking at it?
  - Task abstraction
- **How** is it shown?
  - Idiom: visual encoding and interaction

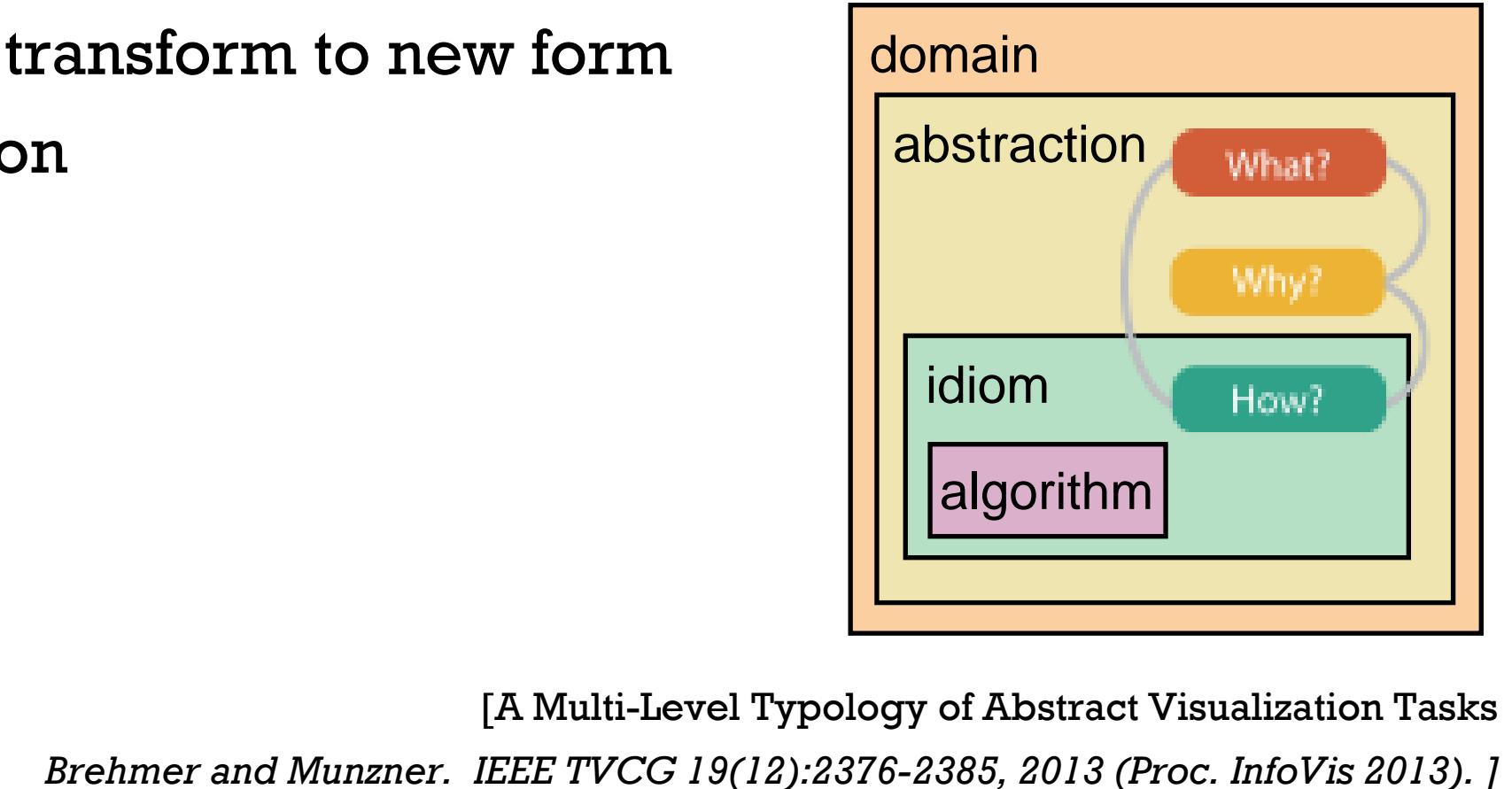
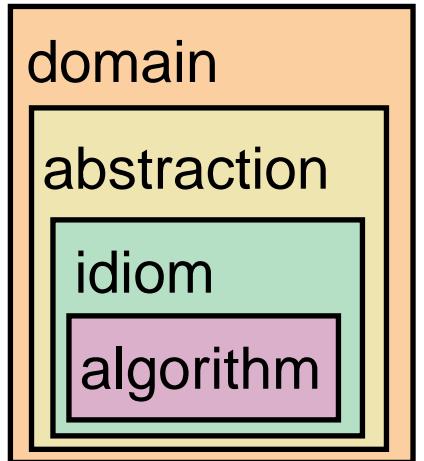


# Analysis framework: Four levels, three questions

- **domain situation**
  - who are the target users?
- **abstraction**
  - translate from specifics of domain to vocabulary of vis
- **what is shown? data abstraction**
  - often don't just draw what you're given: transform to new form
  - **why** is the user looking at it? task abstraction
- **idiom**
- **how** is it shown?
  - visual encoding idiom: how to draw
  - interaction idiom: how to manipulate
- **algorithm**
  - efficient computation

[A Nested Model of Visualization Design and Validation.]

Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009). ]



# Why is validation difficult?

- different ways to get it wrong at each level



## Domain situation

You misunderstood their needs



## Data/task abstraction

You're showing them the wrong thing



## Visual encoding/interaction idiom

The way you show it doesn't work



## Algorithm

Your code is too slow

# Why is validation difficult?

- solution: use methods from different fields at each level

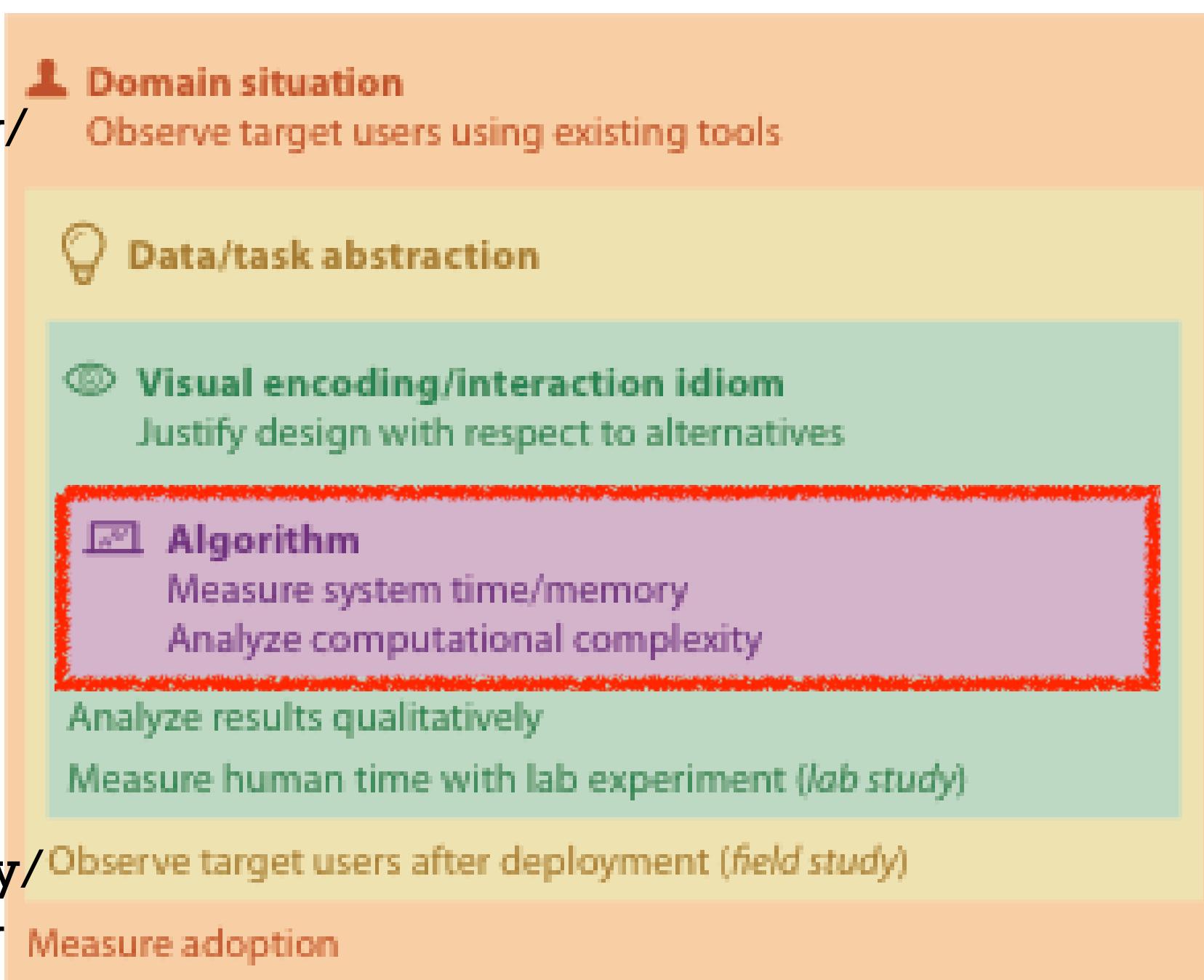
anthropology/  
ethnography

design

computer  
science

cognitive  
psychology

anthropology/  
ethnography



technique-  
driven work

# What?

## Datasets

## Attributes

### ④ Data Types

- Items
- Attributes
- Links
- Positions
- Grids

### ④ Attribute Types

- Categorical



### ④ 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

- Ordered

- Ordinal

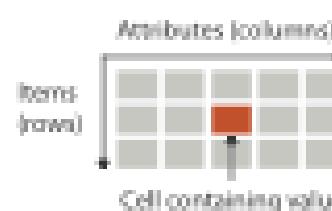


- Quantitative



### ④ Dataset Types

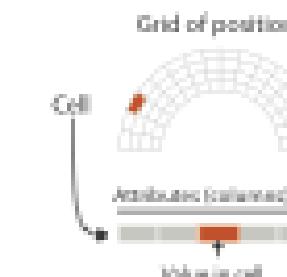
- Tables



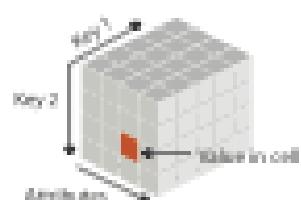
- Networks



- Fields (Continuous)



- Multidimensional Table



- Trees



### ④ Ordering Direction

- Sequential



- Diverging



- Cyclic

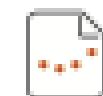


- Geometry (Spatial)



### ④ Dataset Availability

- Static



- Dynamic



What?

Why?

How?

# What?

## Ch. 2

## What?

### Datasets

### Attributes

#### ④ Data Types

→ Items → Attributes → Links → Positions → Grids

#### ④ Attribute Types

→ Categorical



#### ④ 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	Attributes	Attributes	

→ Ordered

→ Ordinal

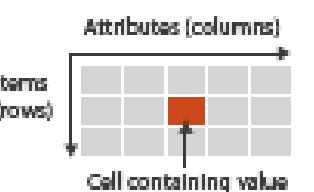


→ Quantitative



#### ④ Dataset Types

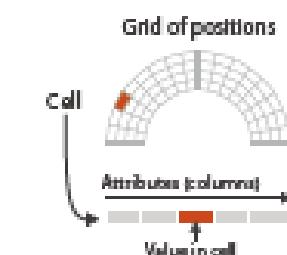
→ Tables



→ Networks



→ Fields (Continuous)



#### ④ Ordering Direction

→ Sequential



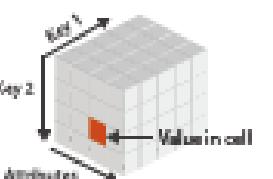
→ Diverging



→ Cyclic



→ Multidimensional Table



→ Trees



→ Geometry (Spatial)



What?

Why?

How?

# • Dataset

# • Attribute

# 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?
  - 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

- food shipment of produce (basil & pear) arrived in satisfactory condition on 7th day of month
- Basil Point neighbourhood 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
- data types: structural or mathematical interpretation of data
  - item, link, attribute, position, (grid)
  - different from data types in programming!

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

- e.g. patient, car, stock, city
  - "independent variable"

- **attribute:** property that is measured, observed, logged...

- e.g. height, blood pressure for patient
  - e.g. 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

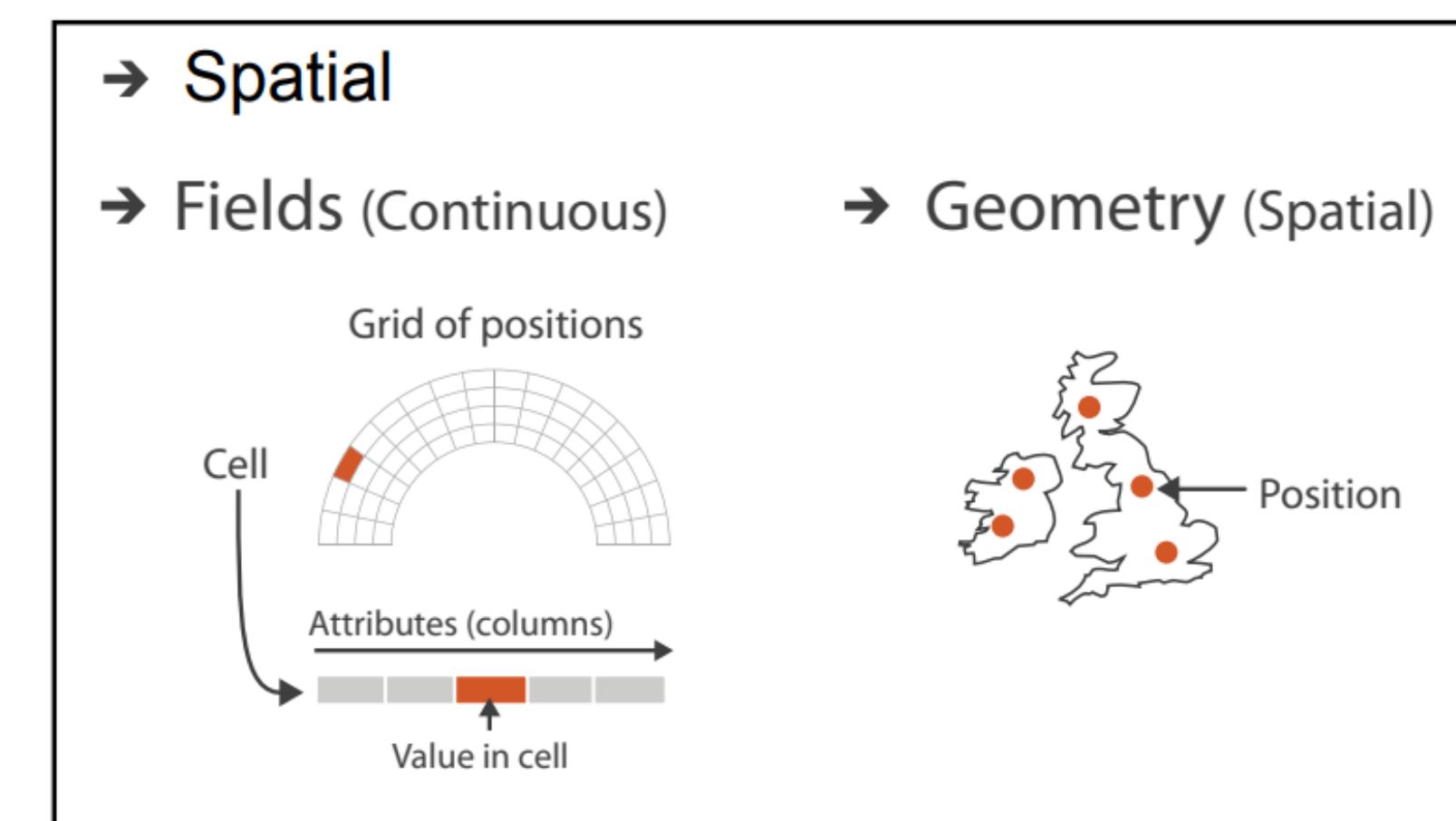
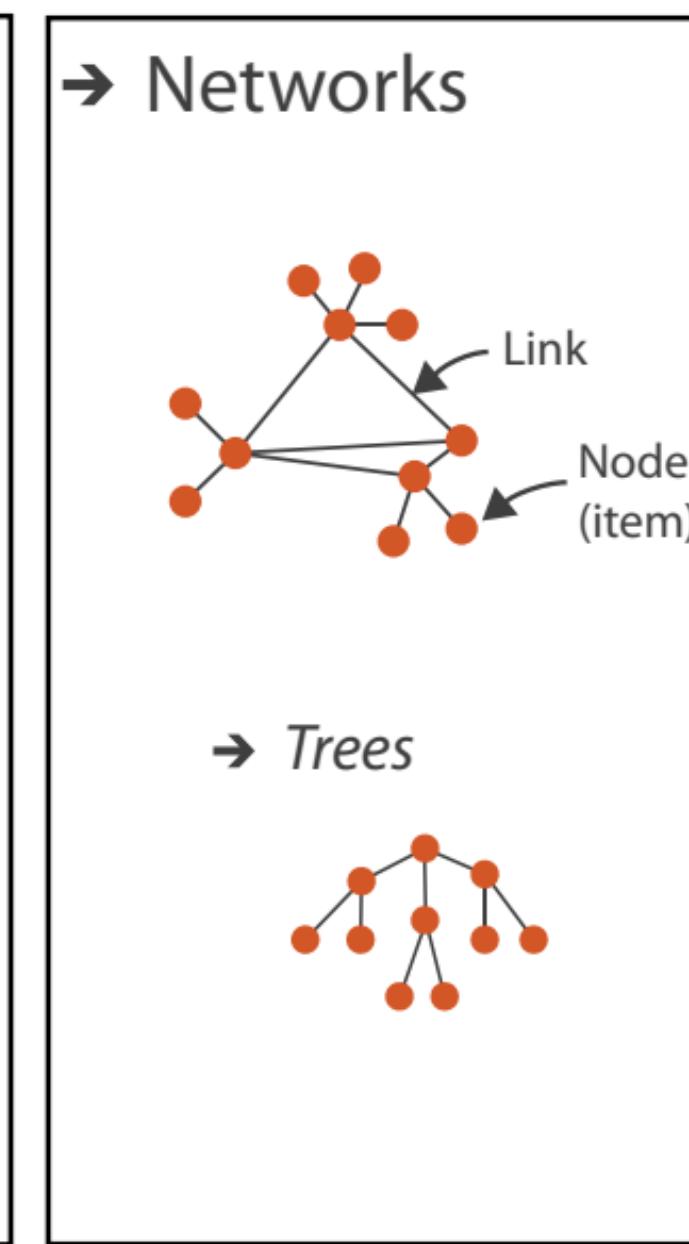
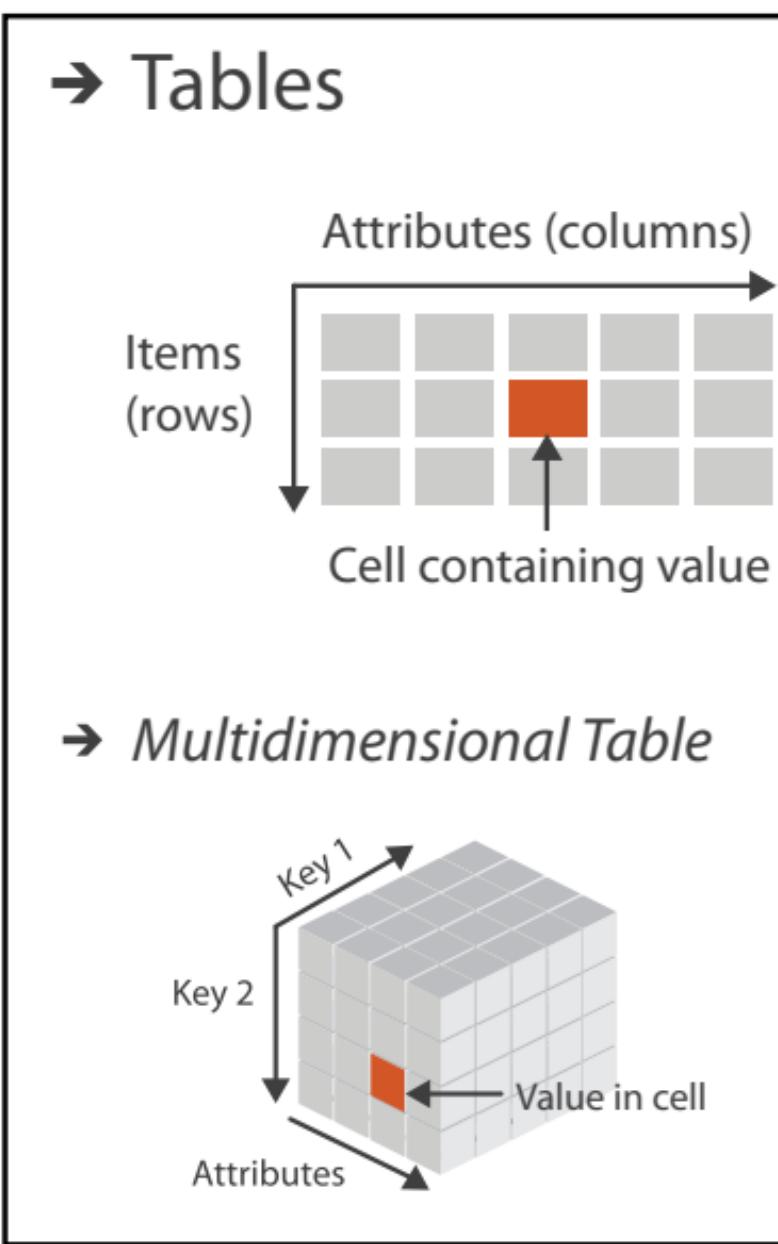
item: person

# Other data types

- links
  - express relationship between two items
  - e.g. 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

# Three major dataset types

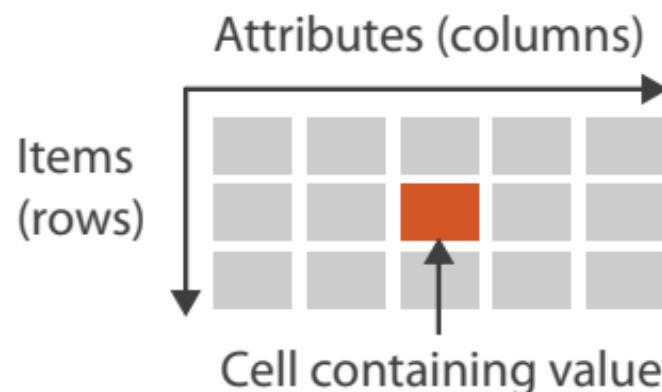
## → Dataset Types



# Dataset types: table

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

→ Tables



Tables	
Items	Attributes

attributes: name, age, shirt size, fave fruit

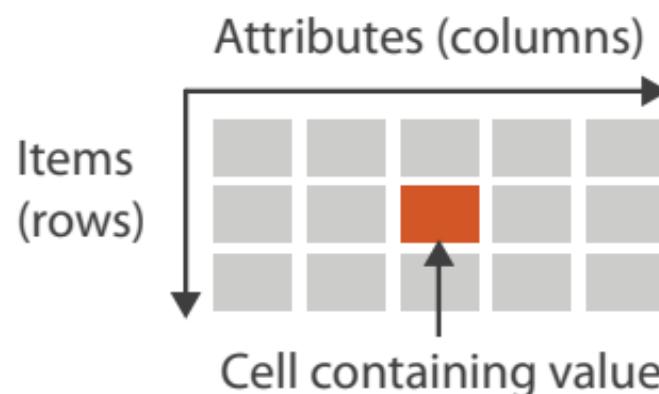
Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

item: person

# Dataset types: table

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

→ Tables



Tables	
Items	Attributes

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

# Dataset types: table

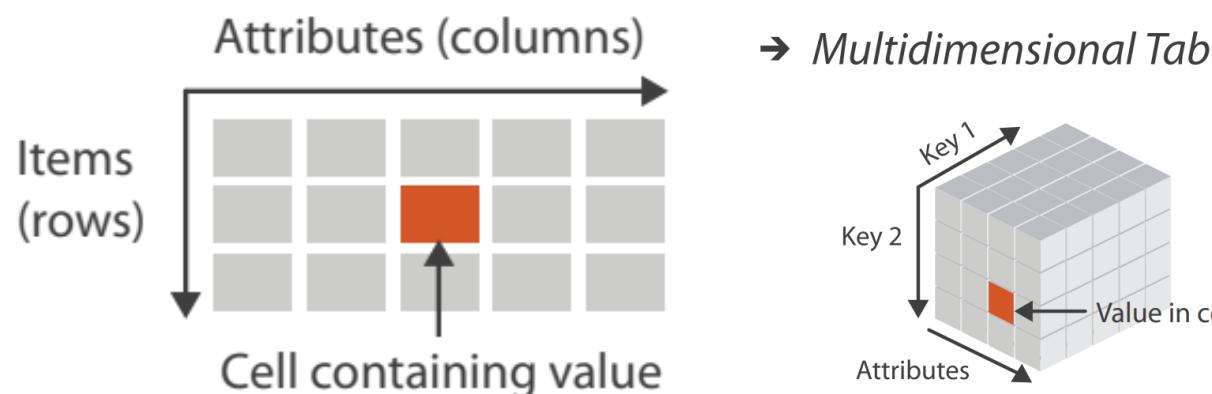
Tables

Items

Attributes

- Multidimensional tables
  - indexing based on multiple keys
  - e.g., genes, patients

→ Tables



	A	B	C	D	E	
1	A	B	C	D	E	
2	1	A	B	C	D	E
3	2	1	#1.2			
4	3	G 2	1500	529		
5	4	L 3	GeneName	DESCRIPTION	TCGA-02-0001-01C-01R-0177-01	TCGA-02-C003-01A-C1R-0177-01
6	5	F 4	L 1.3	P 4	LTF	-1.265728057
7	6	T 5	P 6	POSTN	POSTN	2.662411806
8	7	F 6	H 6	TMSL8	TMSL8	-3.082217836
9	8	F 7	R 7	HLA-DQA1	HLA-DQA1	-1.739064396
10	9	S 8	R 9	RP11-35N5.1	RP11-35N6.1	-3.345352956
11	10	D 9	D 10	STMN2	STMN2	-2.578511106
12	11	A 10	D 11	DCX	DCX	-2.25078976
13	12	I 11	A 12	AGXT2L1	AGXT2L1	-2.639493611
14	13	S 12	I 13	IL13RA2	IL13RA2	-2.93596915
15	14	C 13	M 13	SLN	SLN	-2.466718221
16	15	F 14	N 15	MEOX2	MEOX2	-2.395054056
17	16	F 15	C 15	COL11A1	COL11A1	1.211934832
18	17	F 16	E 17	NNMT	NNMT	0.703745154
19	18	F 17	C 17	F13A1	F13A1	-0.229094012
20	19	T 18	M 18	CXCL14	CXCL14	-3.1309694
21	20	T 19	M 19	MBP	MBP	-1.906390566
22	21	K 20	T 20	TF	TF	-4.334123292
	22	G 21	K 21	KCNQ2	KCNQ2	-1.777692395
						-2.300362021
						-1.996306032

# Dataset types: networks

Tables

Items

Attributes

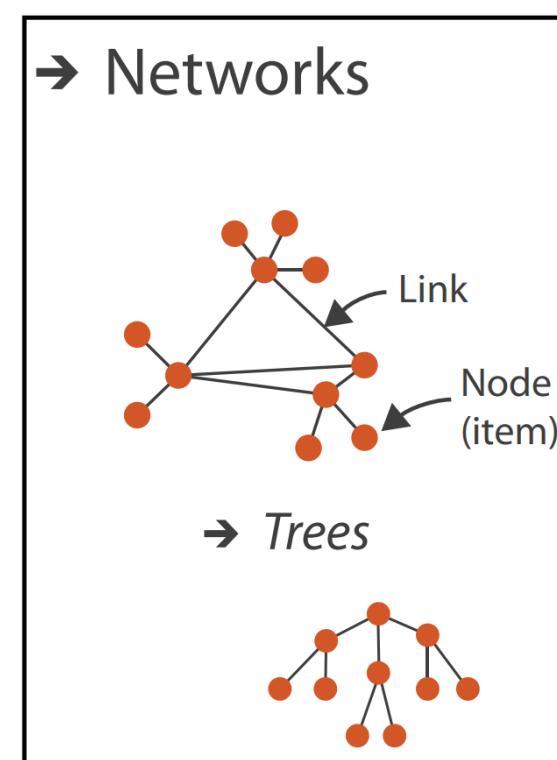
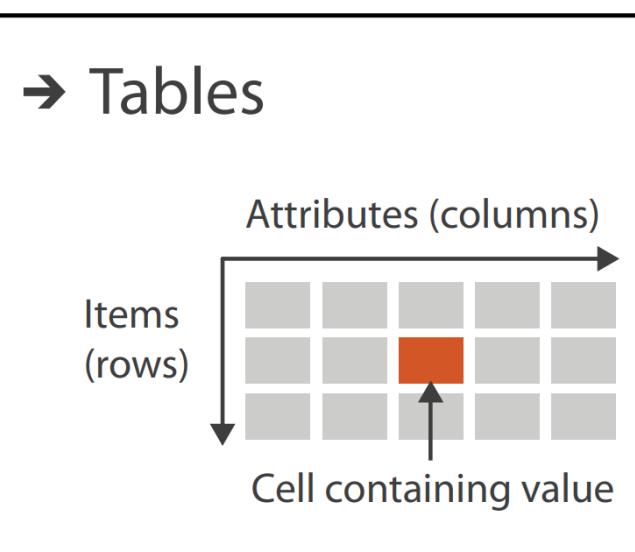
Networks &  
Trees

Items (nodes)

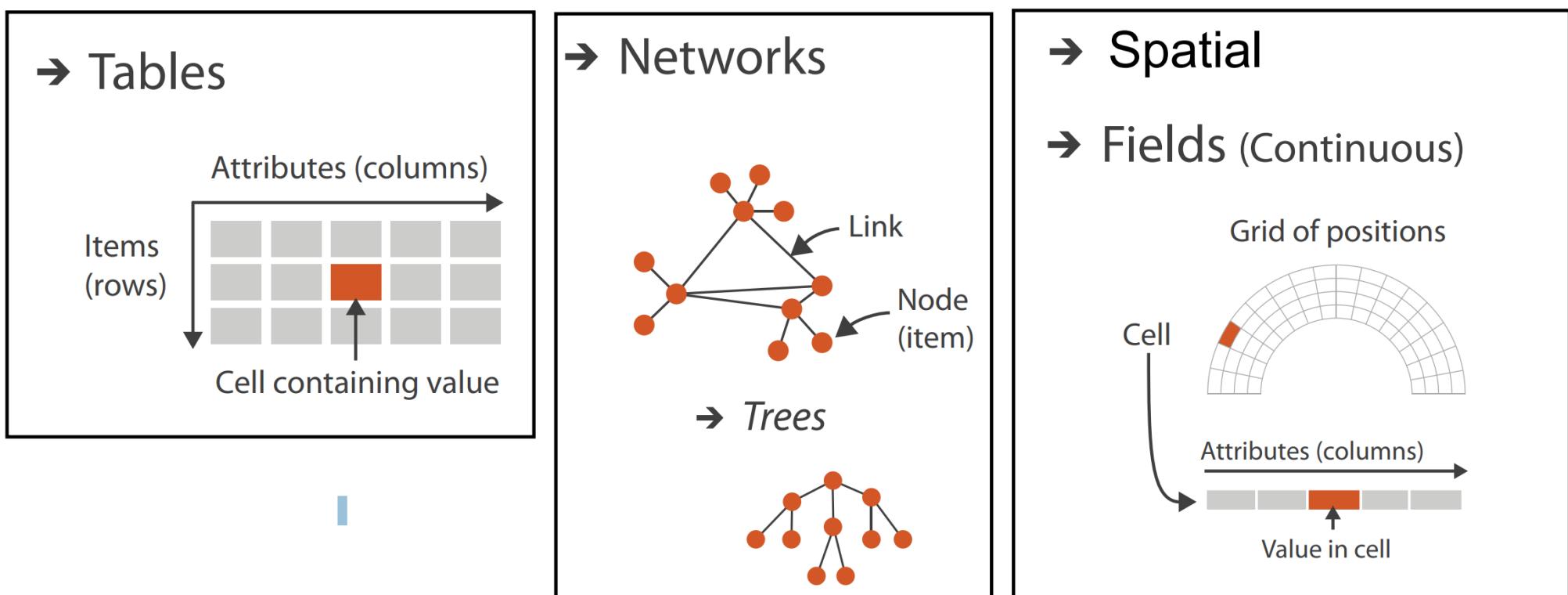
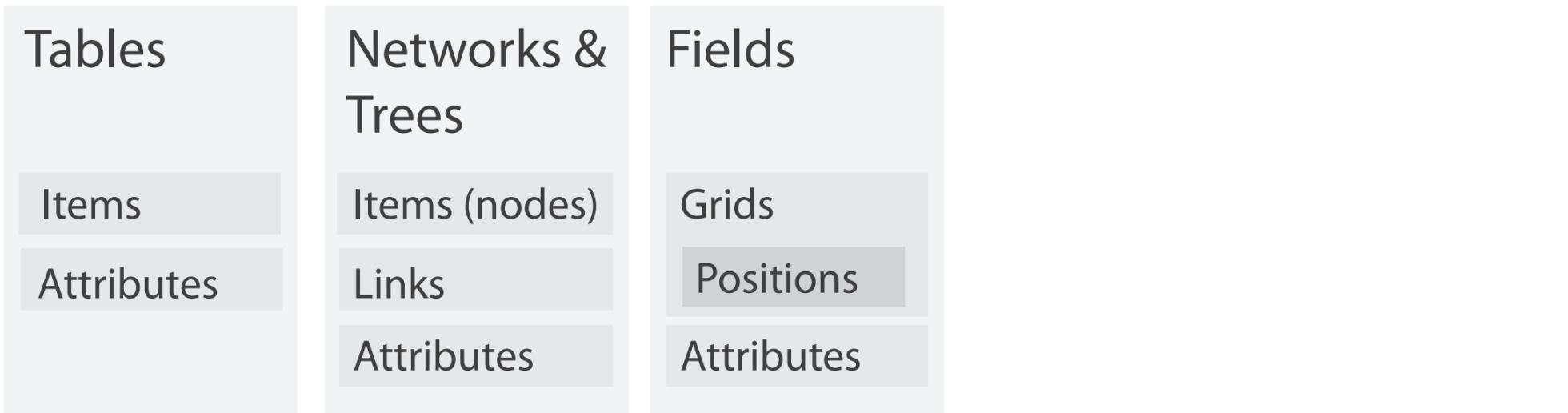
Links

Attributes

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

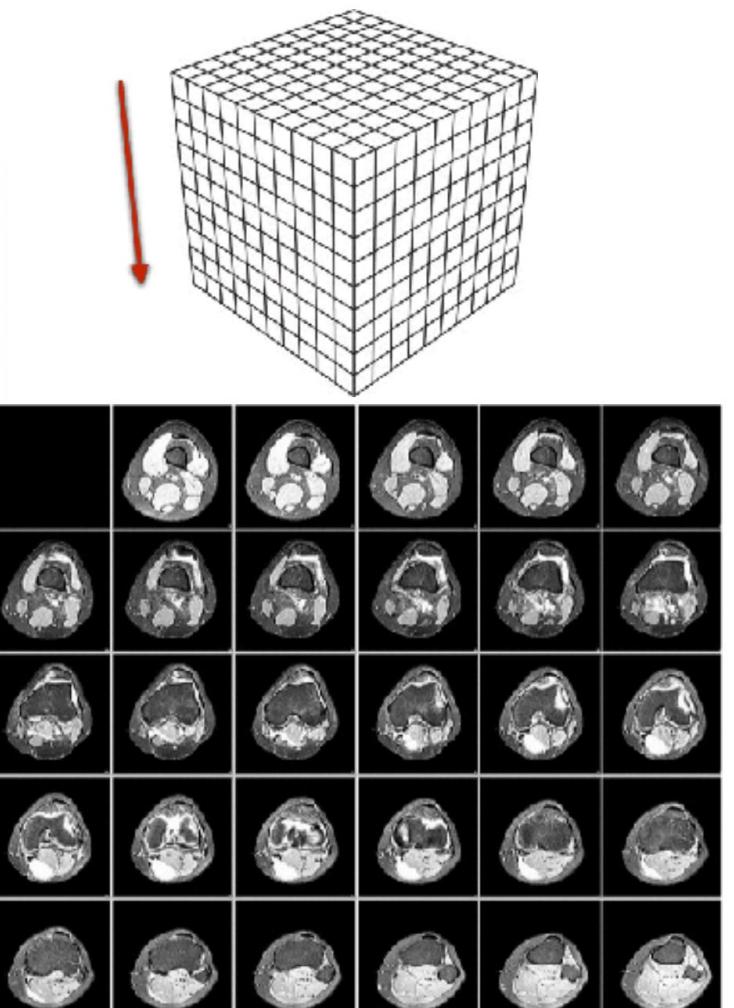


# Dataset types: fields

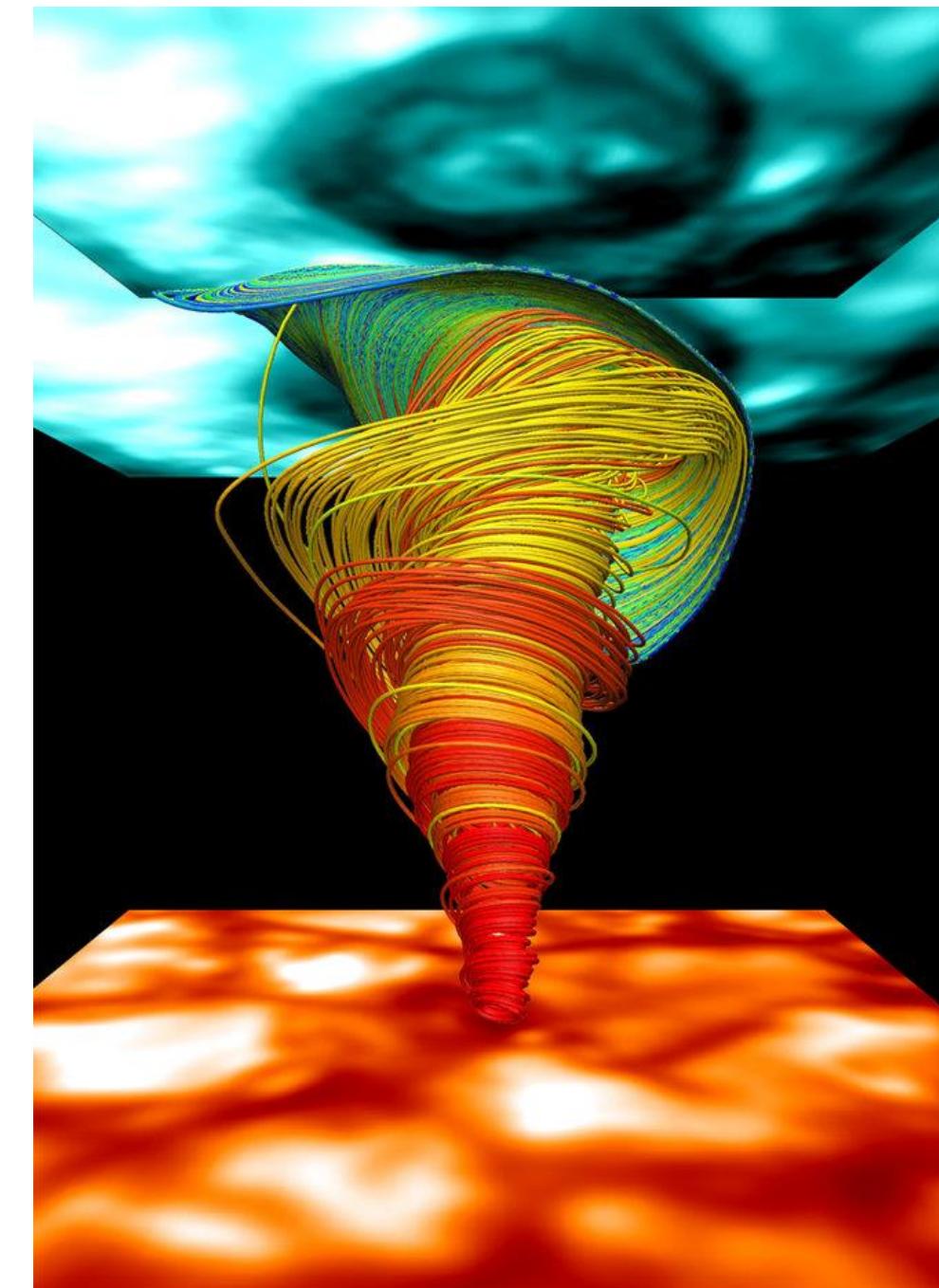
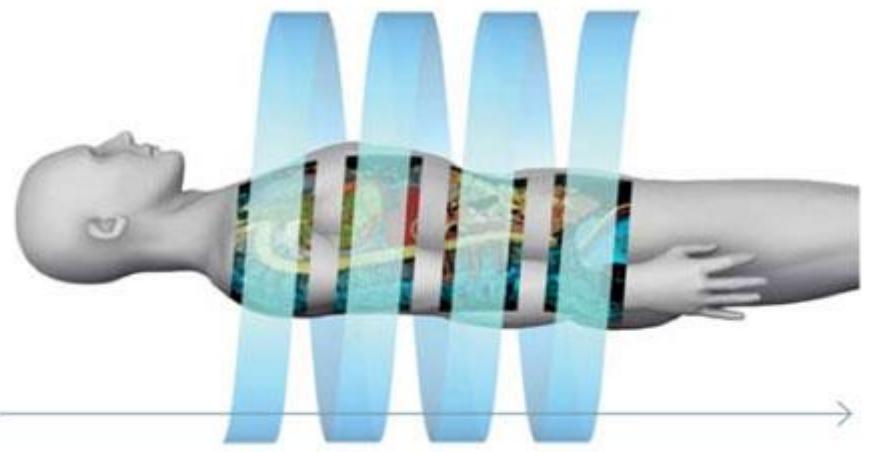


# Spatial fields

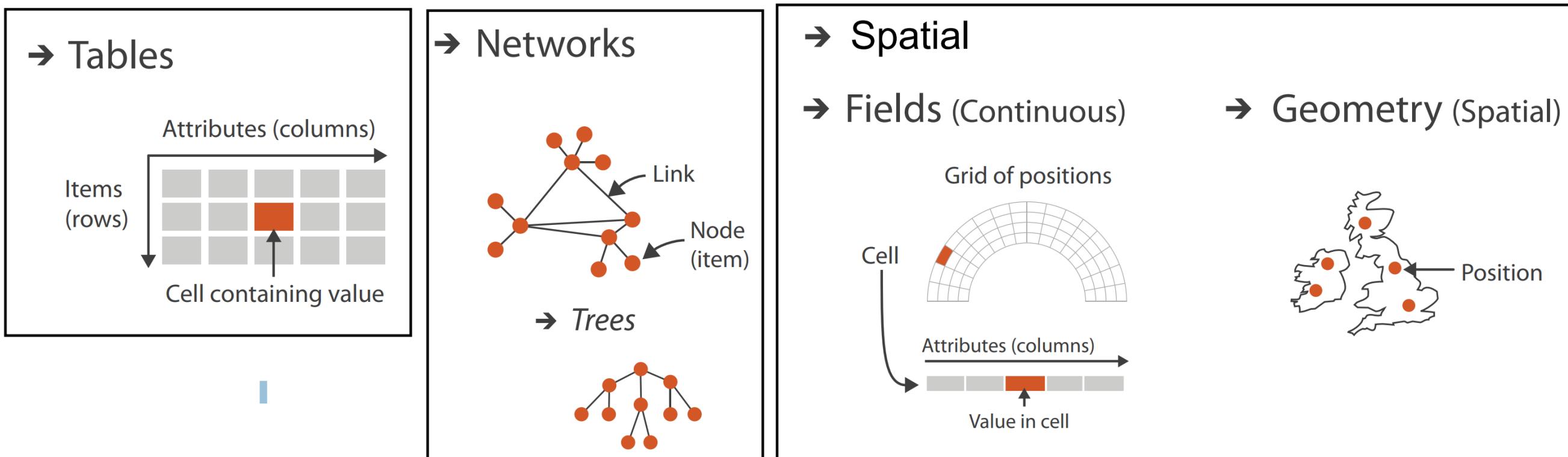
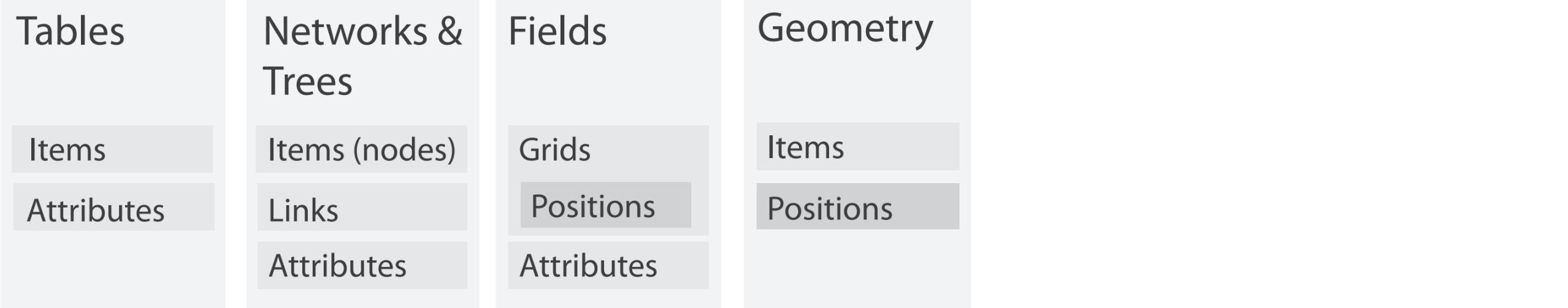
- attribute values associated w/ cells
- cell contains value from continuous domain
  - e.g. temperature, pressure, wind velocity
  - measured or simulated
- major concerns
  - sampling:
    - where attributes are measured
  - interpolation:
    - how to model attributes elsewhere
  - grid types



# Field data

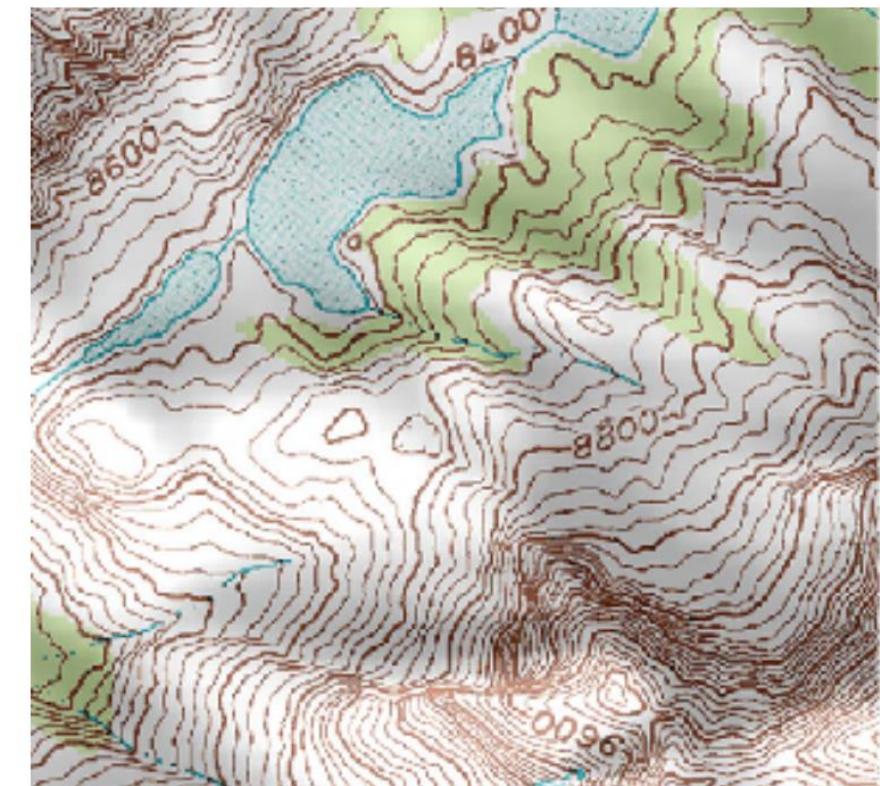


# Dataset types: geometry

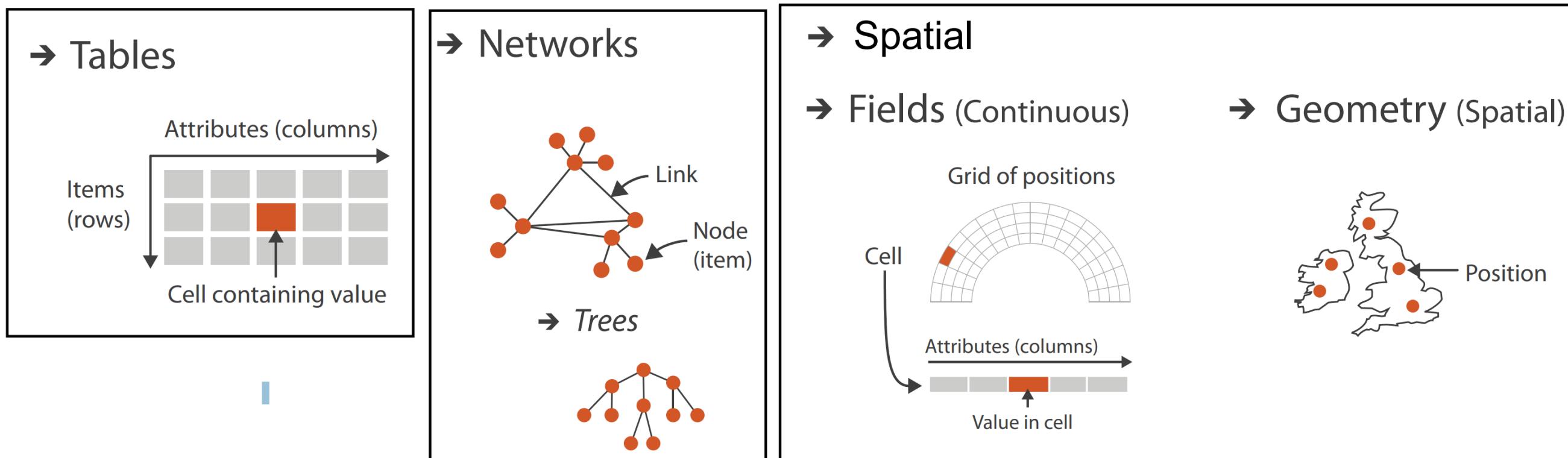
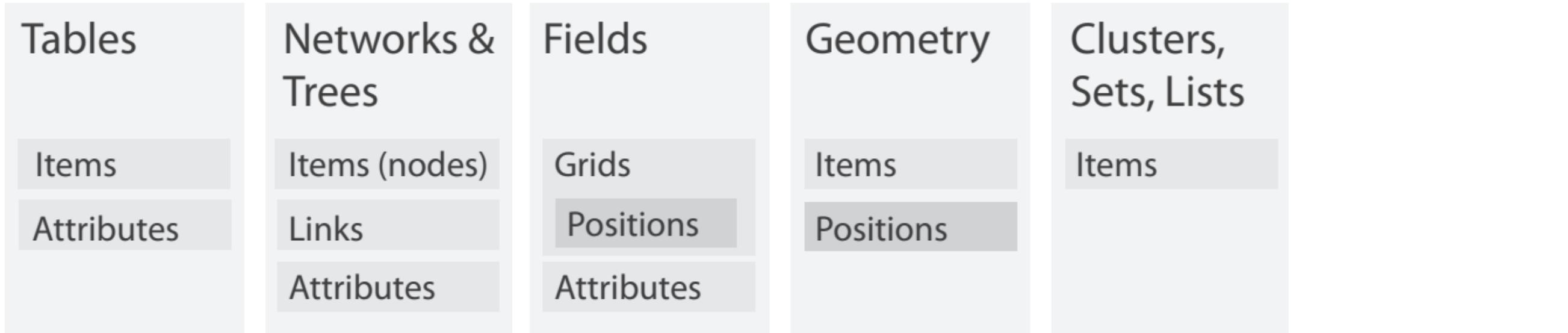


# Geometry

- 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

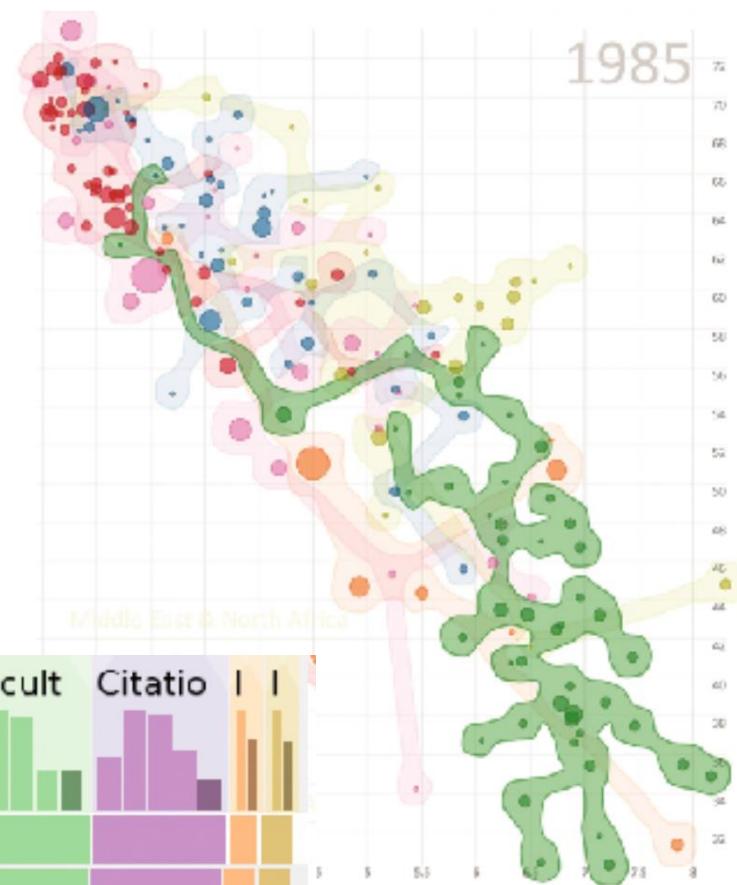
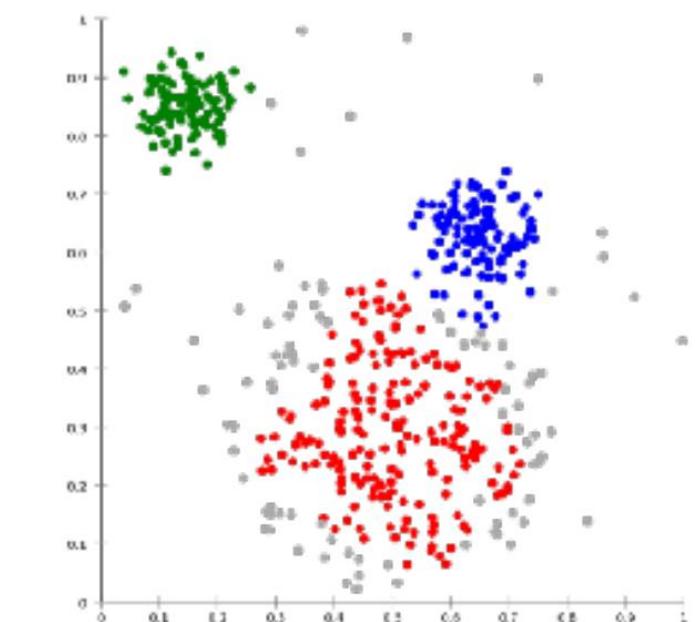
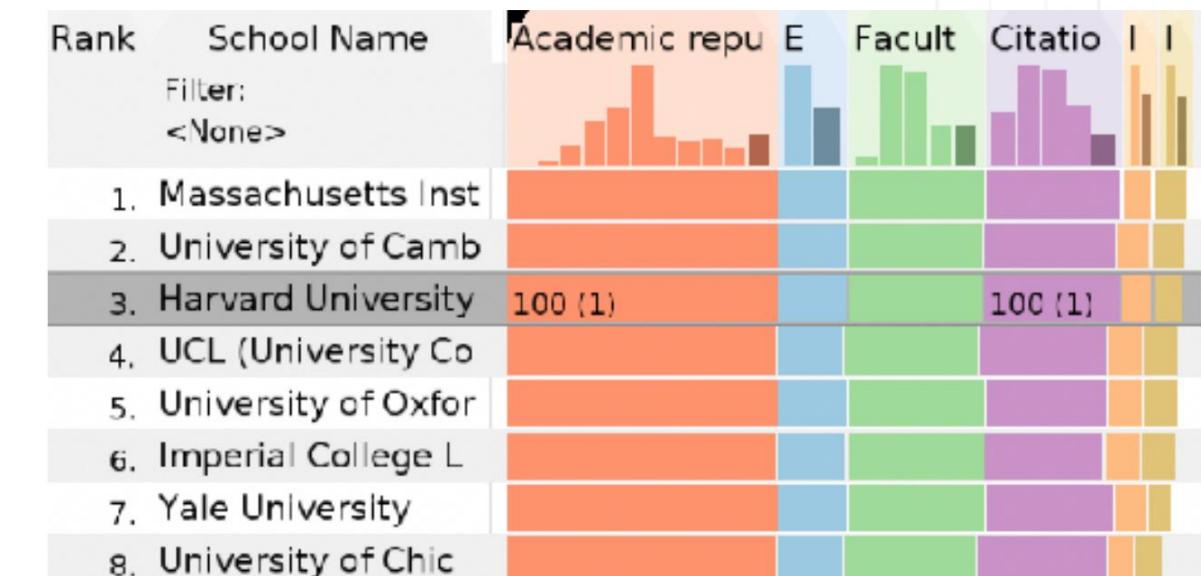


# Dataset types: geometry



# Dataset types: 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

## → Dataset Availability

→ Static



→ Dynamic



# Attribute types

## → Attribute Types

→ Categorical

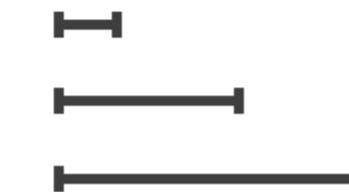


→ Ordered

→ *Ordinal*



→ *Quantitative*



## → Ordering Direction

→ Sequential



→ Diverging



→ Cyclic



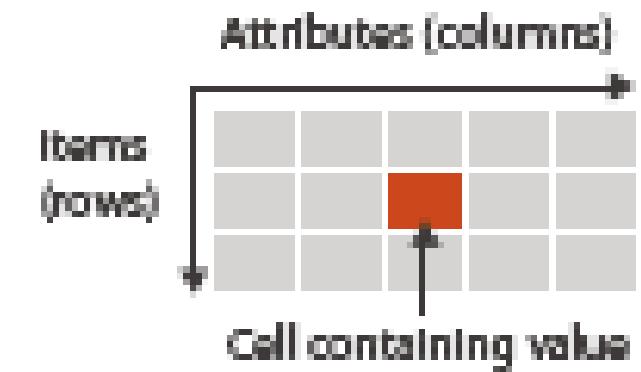
categorical  
ordinal  
quantitative

A	B	C	D	E	F
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

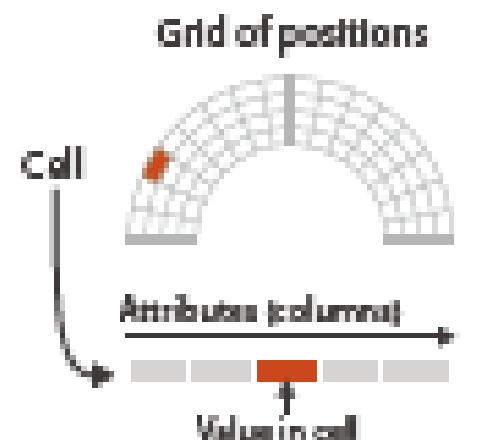
# Key versus Value Semantics

- Key as an index to look up value attributes

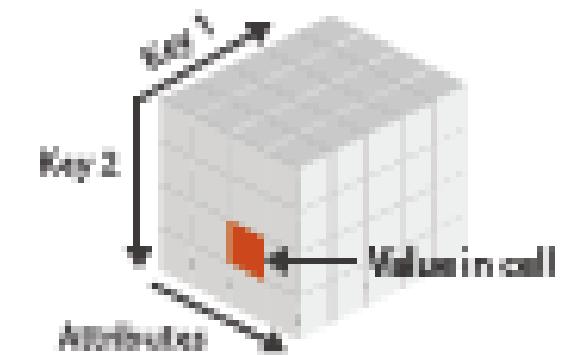
→ Tables



→ Fields (Continuous)



→ Multidimensional Table



# Temporal semantics

- Time-varying
  - Time is one of the **key** attributes (e.g., the location of each animal every second)
  - Time-series
    - An ordered sequence of time-value pairs
- Dynamic
  - Time varying
  - Stream: change in the running session

# Why?

## Ch. 3

## Dow Jones Industrial Average 走勢圖



## Dow Jones Industrial Average 走勢圖

資料日期：2022/03/02 06:25 AM 台北時間，股價延遲20分鐘

走勢圖：當日 5天 1個月 6個月 一年 五年 十年 今年 全部

圖形以當地時間繪製

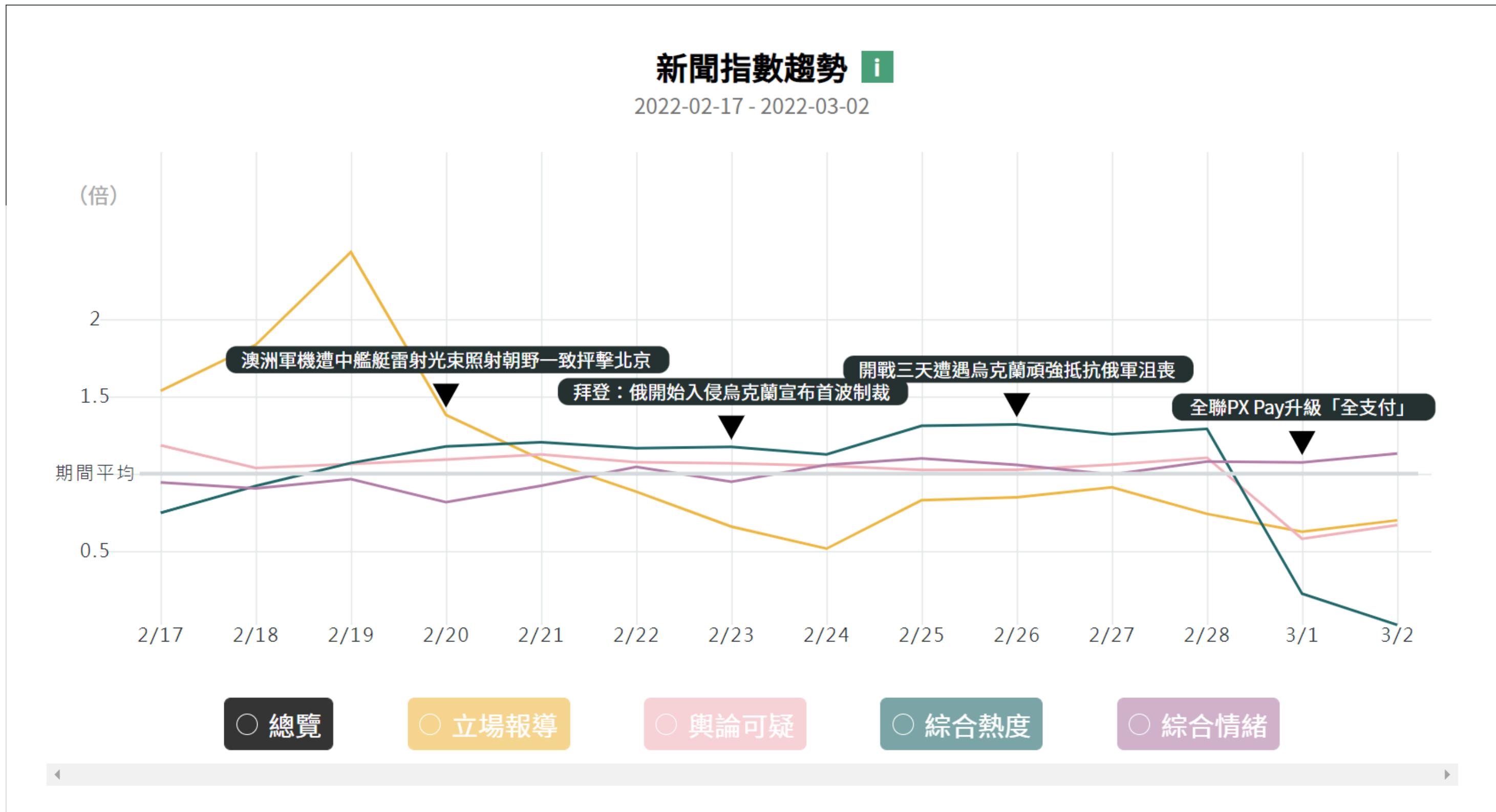


股票名稱 / 代號 : Dow Jones Industrial Average (^DJI)

資料日期：2022/03/02 06:25 AM 台北時間

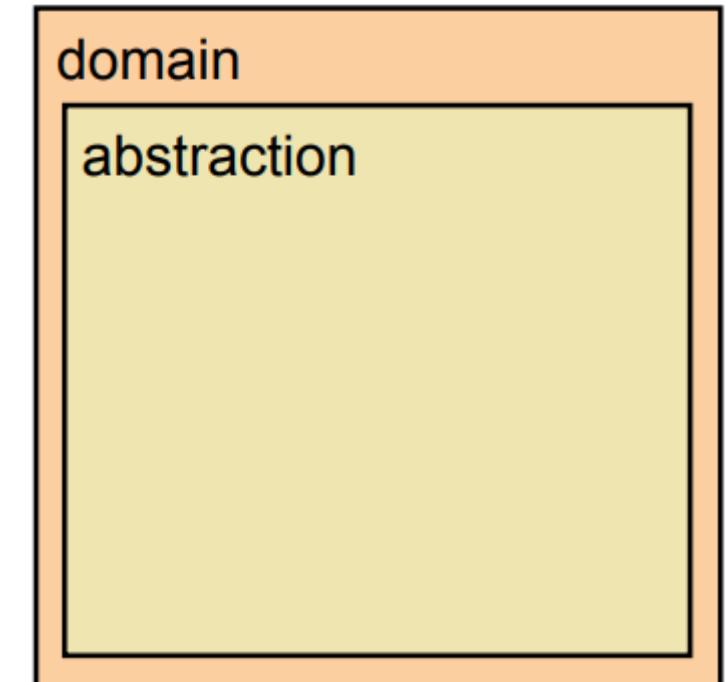
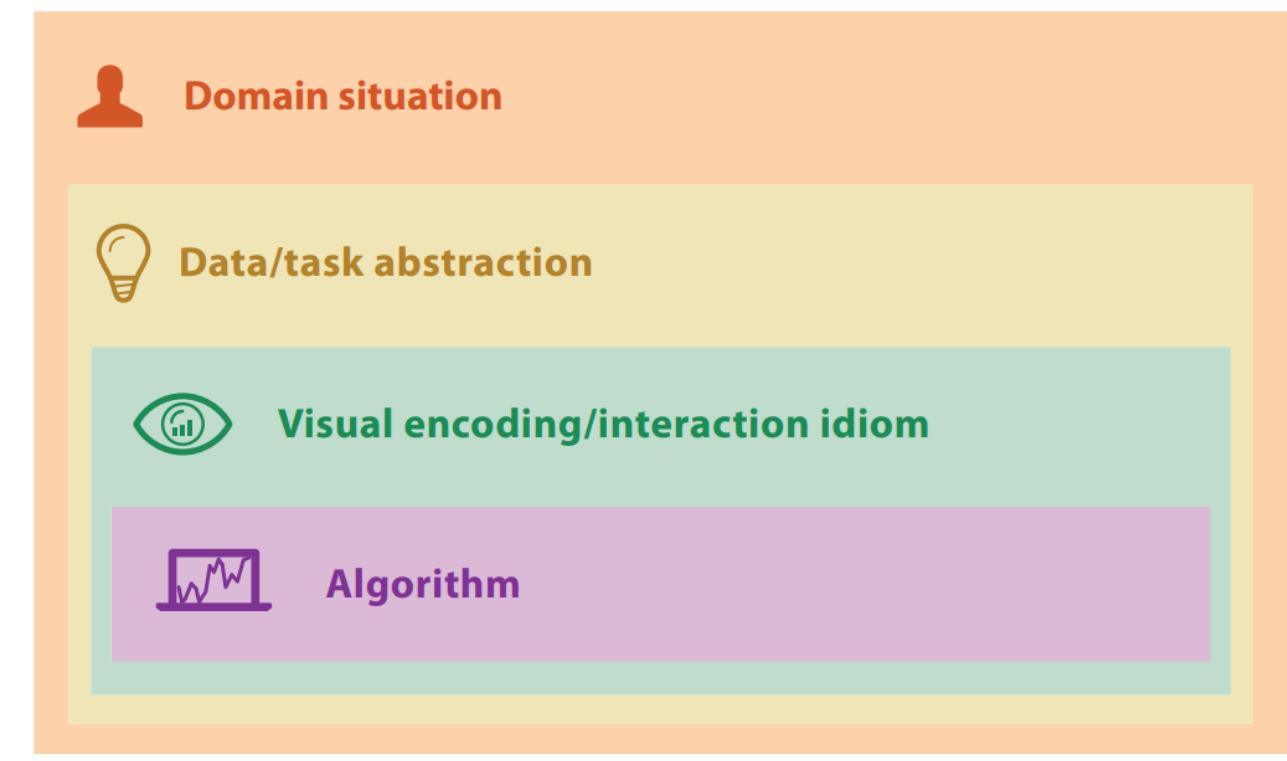
時間	成交價	買進	賣出	漲跌	成交量	昨收	開盤
06:25am	33,294.95	33,222.58	33,495.37	-597.65	442,312,708	33,892.60	33,813.48
最高	最低	一年內最高及最低		本益比(12個月)	每股營利(12個月)	總市值	
33,870.14	33,107.67	30,547.53 - 36,952.65		—	—	—	

資料日期：2022/03/02 06:25 AM 台北時間，股價延遲20分鐘



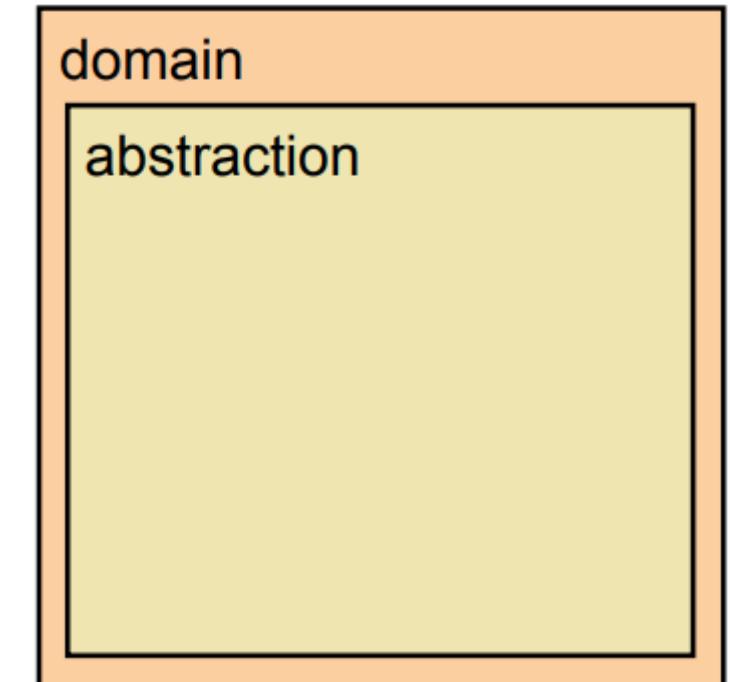
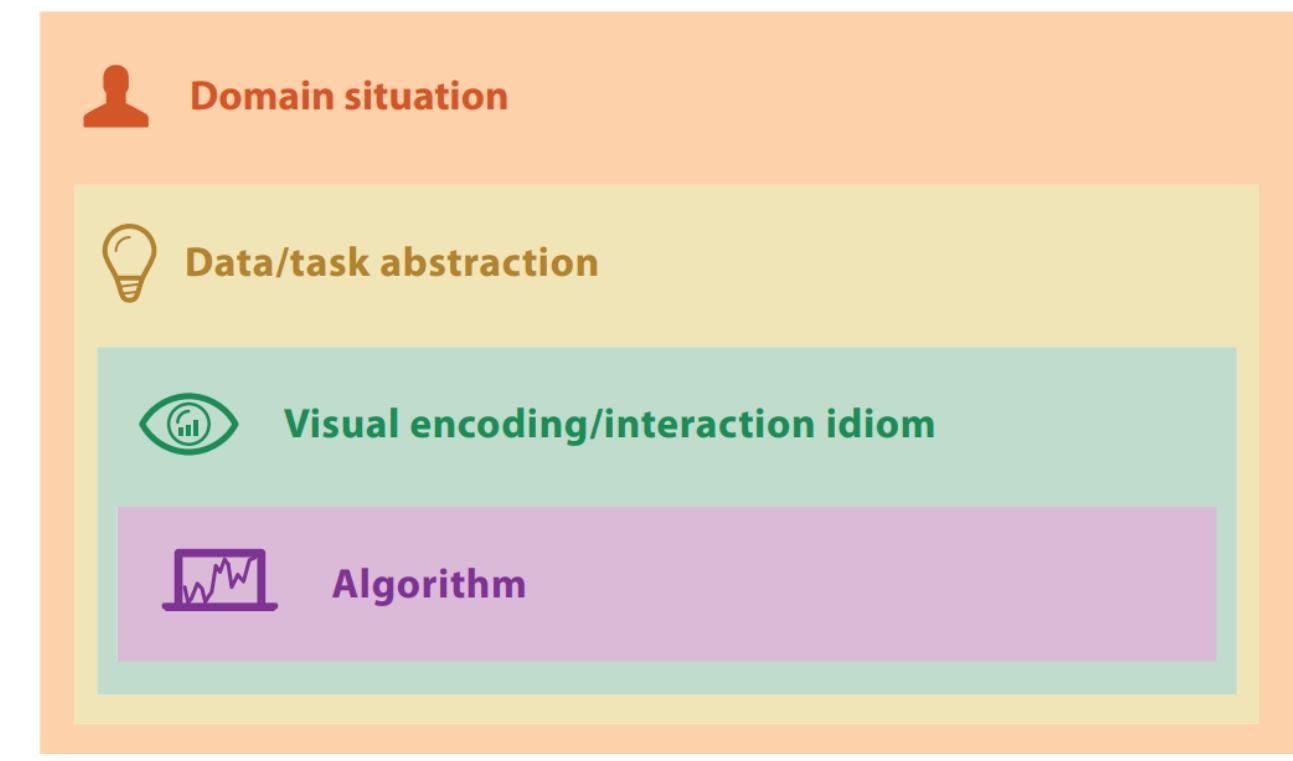
# From domain to abstraction

- **domain characterization:**  
**details of application domain**
  - group of users, target domain, their questions & data
    - varies wildly by domain
    - must be specific enough to get traction
  - domain questions/problems
    - break down into simpler abstract tasks
- **abstraction: data & task**
  - map what and why into generalized terms

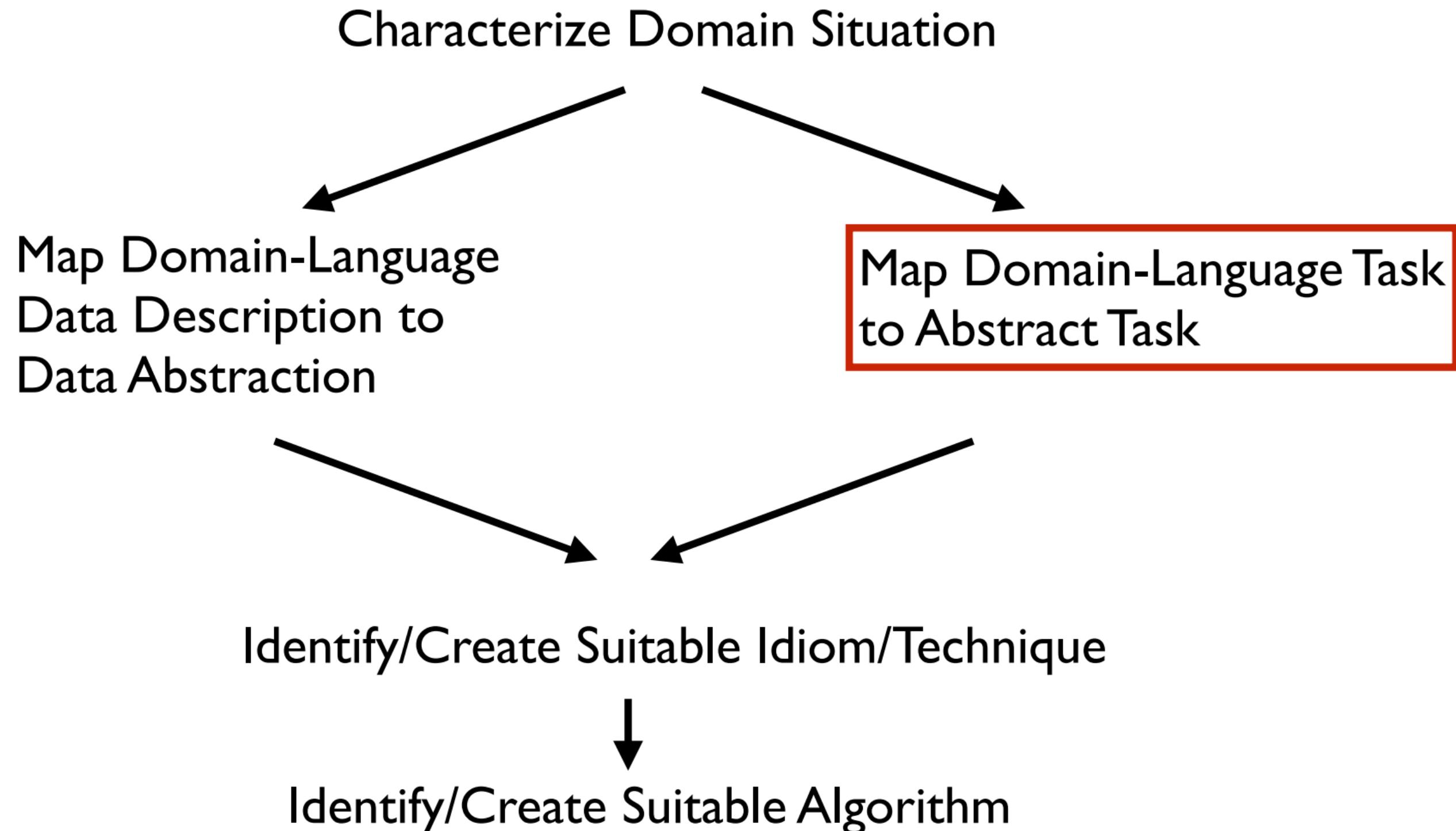


# From domain to abstraction

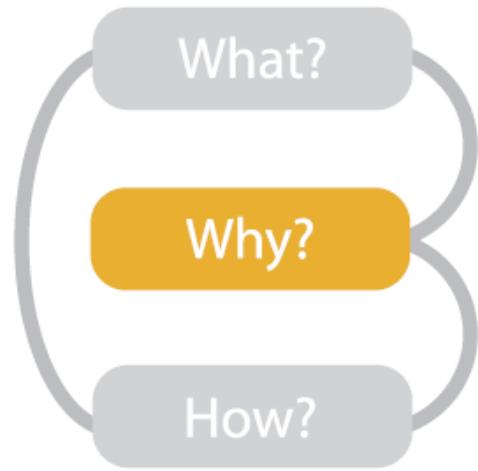
- **domain characterization:**  
**details of application domain**
  - group of users, target domain, their questions & data
    - varies wildly by domain
    - must be specific enough to get traction
  - domain questions/problems
    - break down into simpler abstract tasks
- **abstraction: data & task**
  - map what and why into generalized terms



# Design process

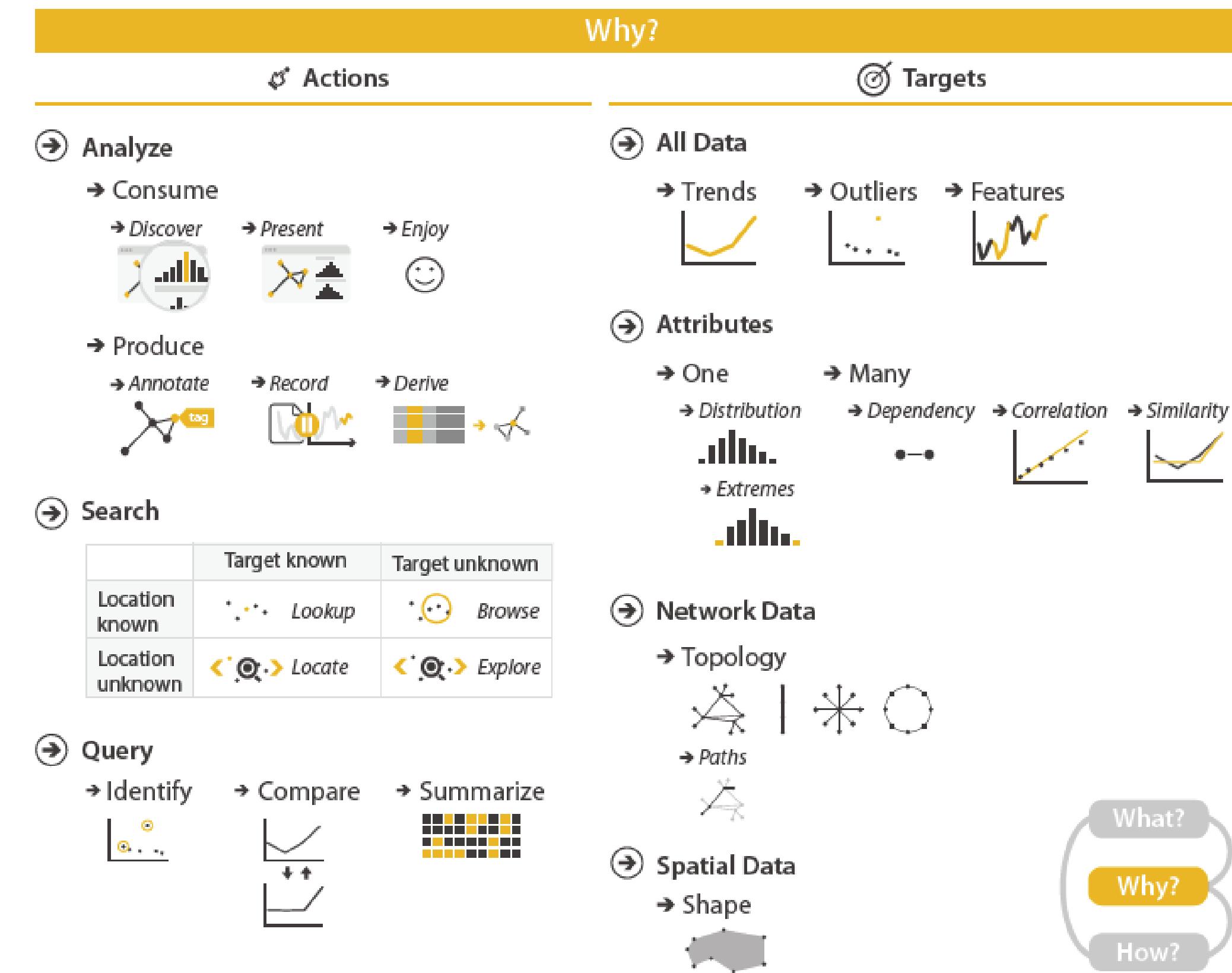


- **{action, target}** pairs
  - discover distribution*
  - compare trends*
  - locate outliers*
  - browse topology*



high  
level

low  
level



# Actions: Analyze

- **consume**

- discover vs present

- classic split

- aka explore vs explain

- enjoy

- newcomer

- aka casual, social

- **produce**

- annotate, record

- derive

- crucial design choice

## → Analyze

- Consume

- Discover



- Present



- Enjoy



- Produce

- Annotate



- Record



- Derive



→ Consume

→ Discover

→ Present

→ Enjoy



# enjoy

- Name Voyager

NameVoyager: Explore baby names and name trends letter by letter

Baby Name > T

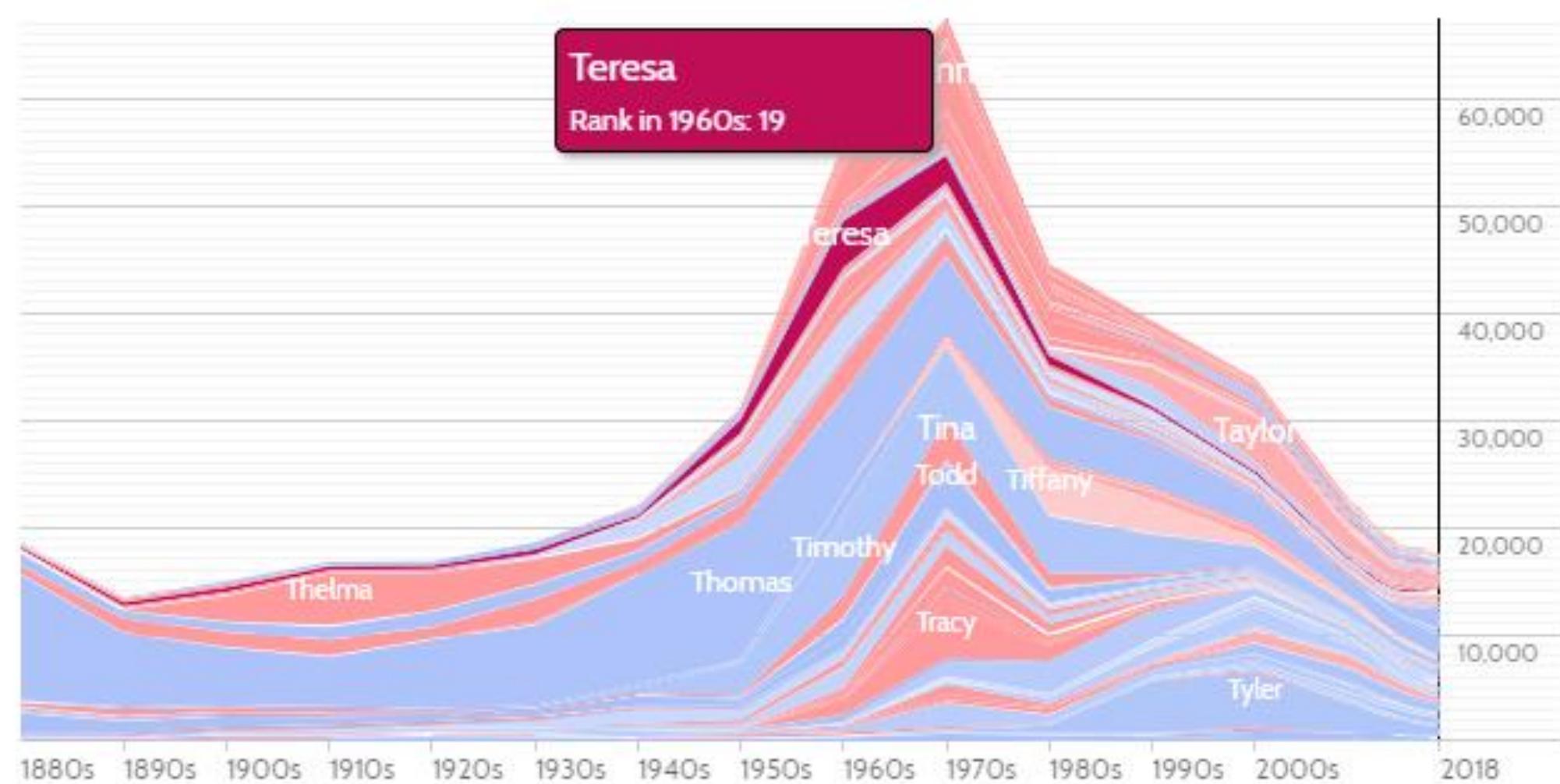
Both  Boys  Girls

boys	1000	500	100	25	1
girls	1000	500	100	25	1

Current rank:

per million births

Names starting with 'T' per million babies



# 新世代菜市場名男「承恩」女「詠晴」奪冠



各世代前三大取用名字

出生年	男性取用名字	女性取用名字
民國1年至9年	明、金水、健	秀英、英、玉
民國10年至19年	金龍、金水、金生	秀英、玉蘭、玉英
民國20年至29年	正雄、文雄、武雄	秀英、玉蘭、玉英
民國30年至39年	正雄、武雄、文雄	秀英、秀琴、美玉
民國40年至49年	金龍、進財、榮華	麗華、秀琴、秀美
民國50年至59年	志明、志成、文雄	淑芬、美玲、淑惠
民國60年至69年	志偉、志明、建宏	淑芬、雅惠、淑娟
民國70年至79年	家豪、志豪、志偉	雅婷、怡君、雅雯
民國80年至89年	家豪、冠宇、冠廷	雅婷、怡君、怡婷
民國90年至99年	承恩、承翰、冠廷	宜蓁、欣妤、詩涵
民國100年至107年6月	承恩、宥廷、品睿	詠晴、子晴、品妍

內政部戶政司 107 年 6 月 30 日編製

# Annotate

→ Produce

→ Annotate

→ Record

→ Derive



LabelMe

Try it live and highlight entities!

PERSON 1 ORG 2 PRODUCT 3 DATE 4

In a March 2014 DATE interview , Apple ORG designer Jonathan Ive PERSON used the iPhone PRODUCT as an example of Apple ORG 's ethos of creating high - quality , life - changing products .

✓ ✗ ⚡

<https://prodi.gy/>

# Record

→ Produce

→ Annotate

→ Record

→ Derive



Worksheet History

Filters:

Graphic Type: All

Bookmarks:  Show Only

Add Inventory

Show Me!

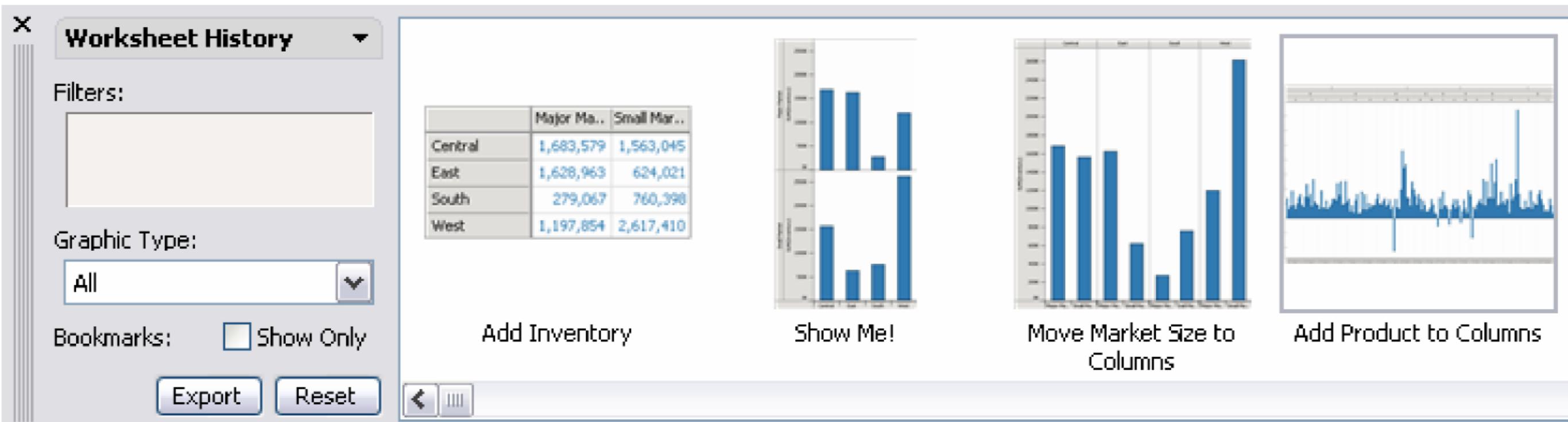
Move Market Size to Columns

Add Product to Columns

Export Reset

← →

	Major Mar..	Small Mar..
Central	1,683,579	1,563,045
East	1,628,963	624,021
South	279,067	760,390
West	1,197,654	2,617,410

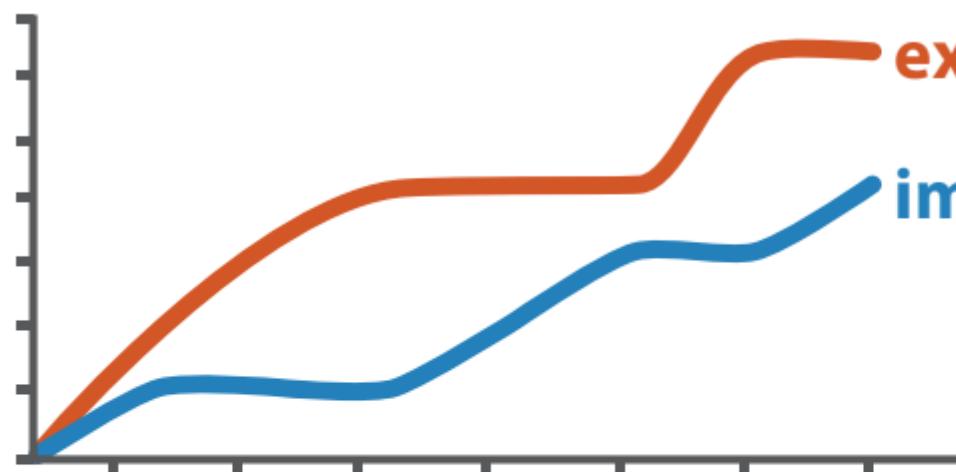


- Working history. [Heer et al. 08]

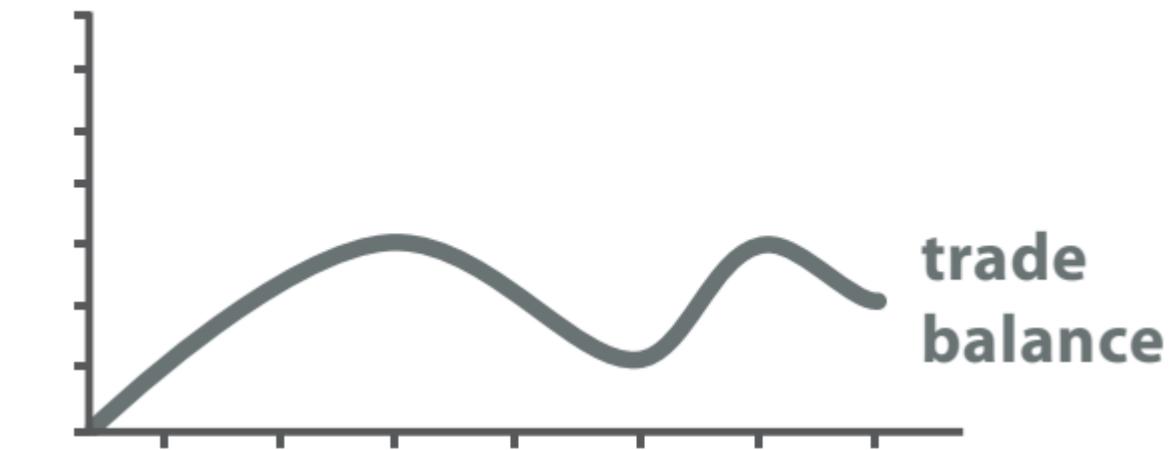
# Derive



- don't 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



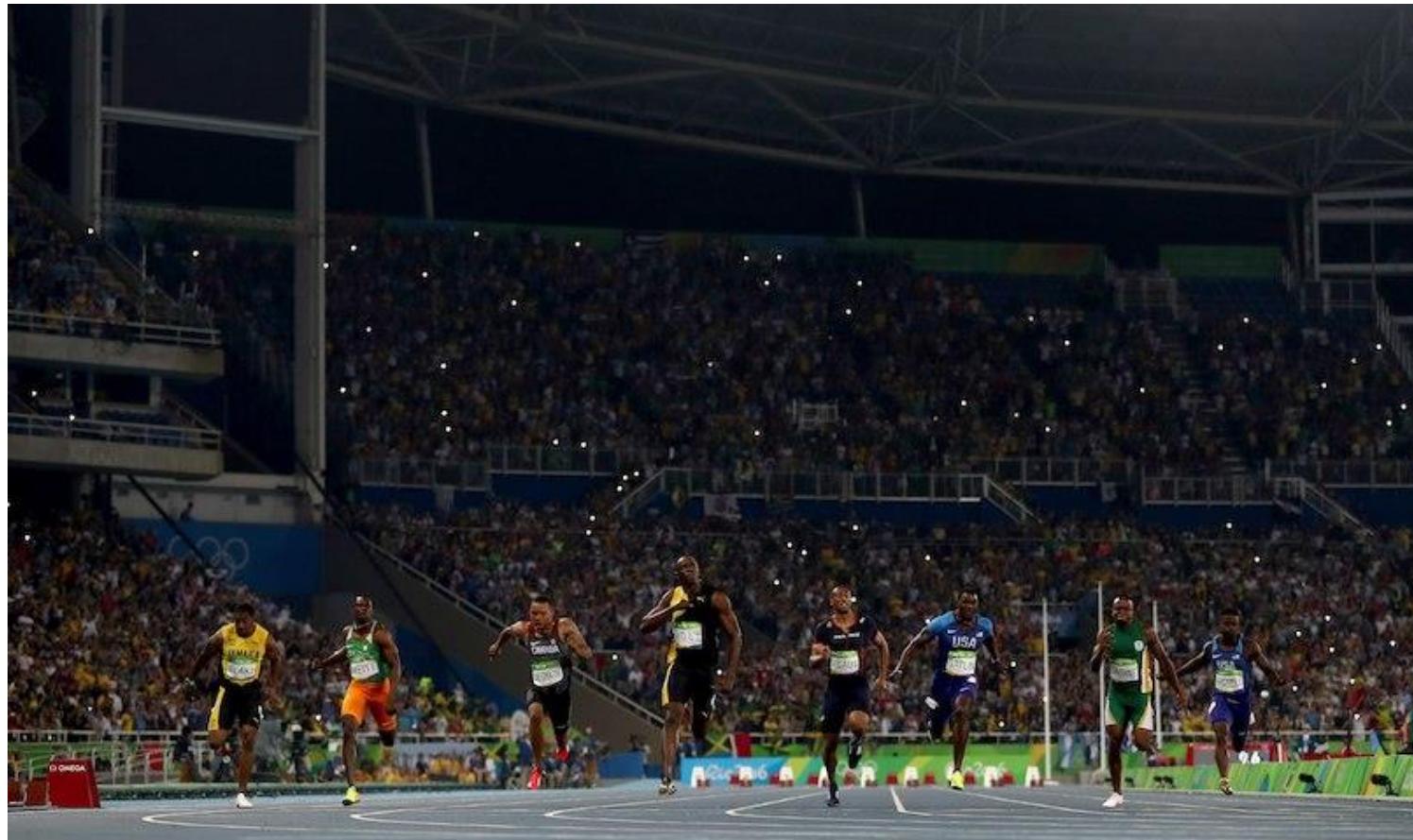
$$\text{trade balance} = \text{exports} - \text{imports}$$

Derived Data

# NY Times - Men's 100-Meter Sprint

- Olympic 2012

	<b>Gold</b>	<b>Silver</b>	<b>Bronze</b>
2012	9.63	9.75	9.79
2008	9.69	9.89	9.91
2004	9.85	9.86	9.87
...			



## Usain Bolt vs. 116 years of Olympic sprinters

Based on the athletes' average speeds, if every Olympic medalist raced each other, Usain Bolt (the London version) would win, with a wide distribution of Olympians behind him. Below, where each sprinter would be when Bolt finishes his race.

### MEDALS BY COUNTRY

United States	40	Barbados	1
Britain	8	Bulgaria	1
Jamaica	7	Hungary	1
Canada	5	Netherlands	1
Trinidad and Tobago	4	New Zealand	1
Australia	3	Panama	1
Germany	3	Portugal	1
Cuba	2	South Africa	1
Namibia	2	United Team of Germany	1
Soviet Union	2		



# Actions: Search, Query

- what does user know?
  - target, location
- how much of the data matters?
  - one, some, all

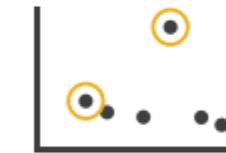
➔ Search

	Target known	Target unknown
Location known		
Location unknown		

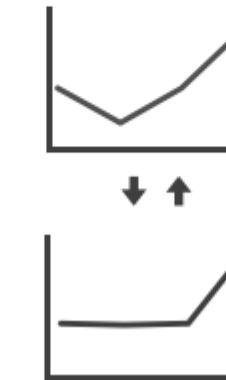
➔ Query

- independent choices for each of these three levels
  - analyze, search, query
  - mix and match

→ Identify



→ Compare



→ Summarize

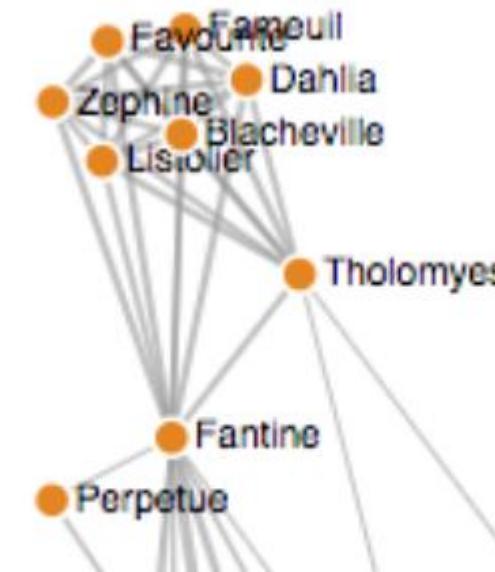


# Actions: Search

- what does user know?
  - target, location
- lookup
  - ex: word in dictionary
    - alphabetical order
- locate
  - ex: keys in your house
  - ex: node in network
- browse
  - ex: books in bookstore
- explore
  - ex: find cool neighborhood in new city

➔ Search

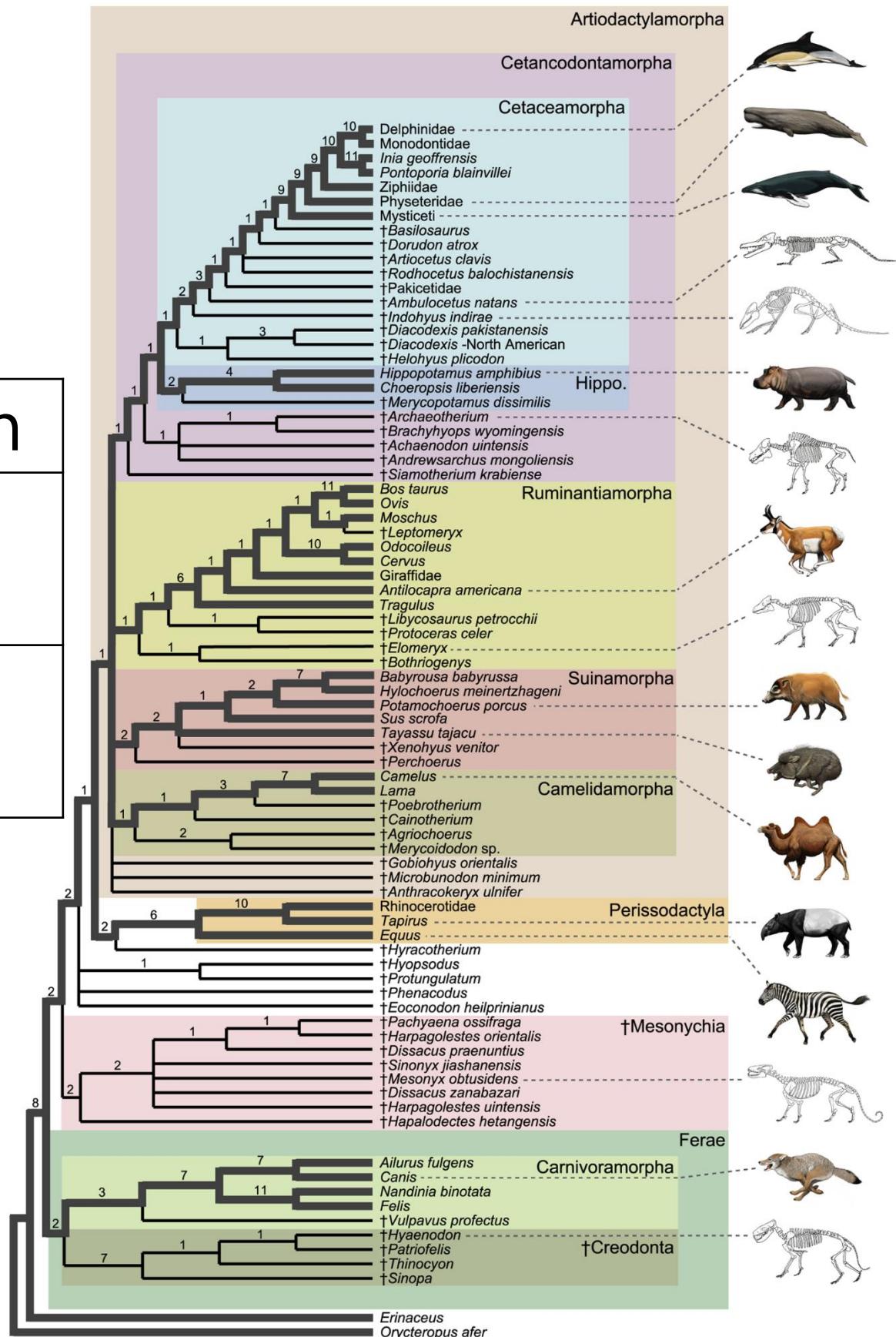
	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>



<https://bl.ocks.org/heybignick/3faf257bbbb7743bb72310d03b86ee8>

# Search

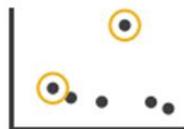
	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>



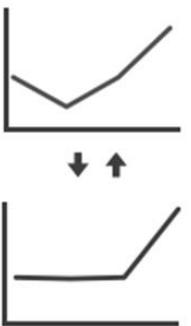
# Query

→ Query

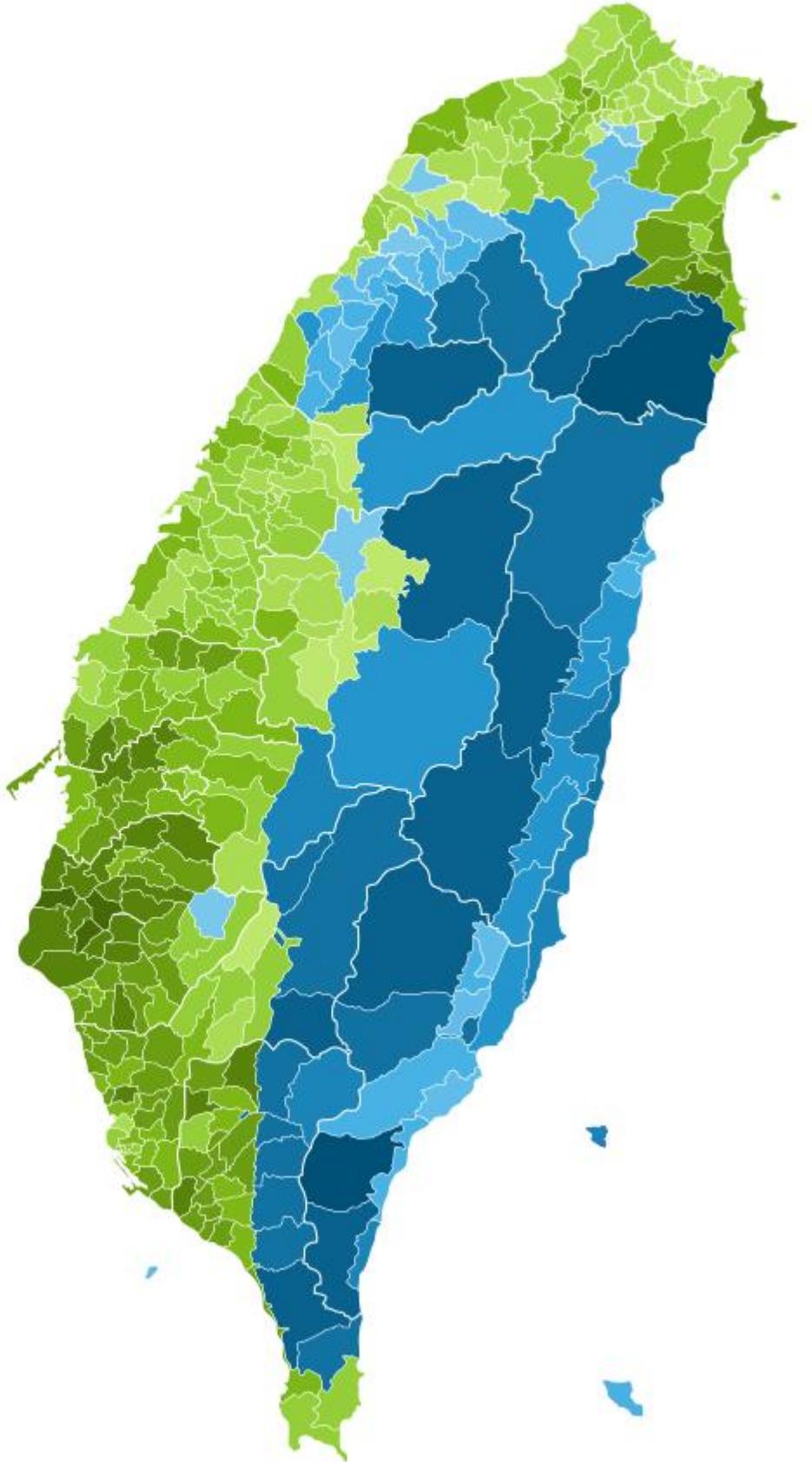
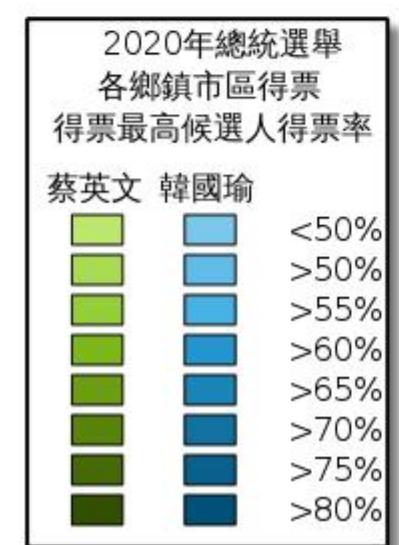
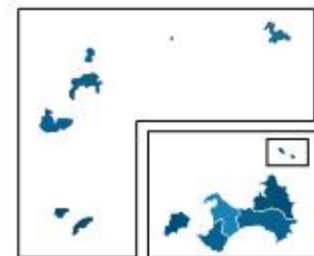
→ Identify



→ Compare



→ Summarize



# Why: Targets

## → All Data

→ Trends



→ Outliers



→ Features



## → Attributes

→ One

→ Distribution



→ Extremes

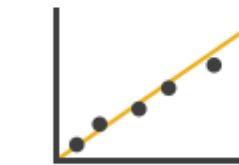


→ Many

→ Dependency



→ Correlation

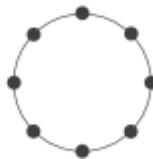
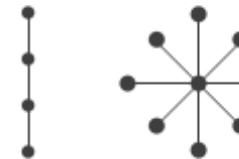


→ Similarity



## → Network Data

→ Topology



→ Paths

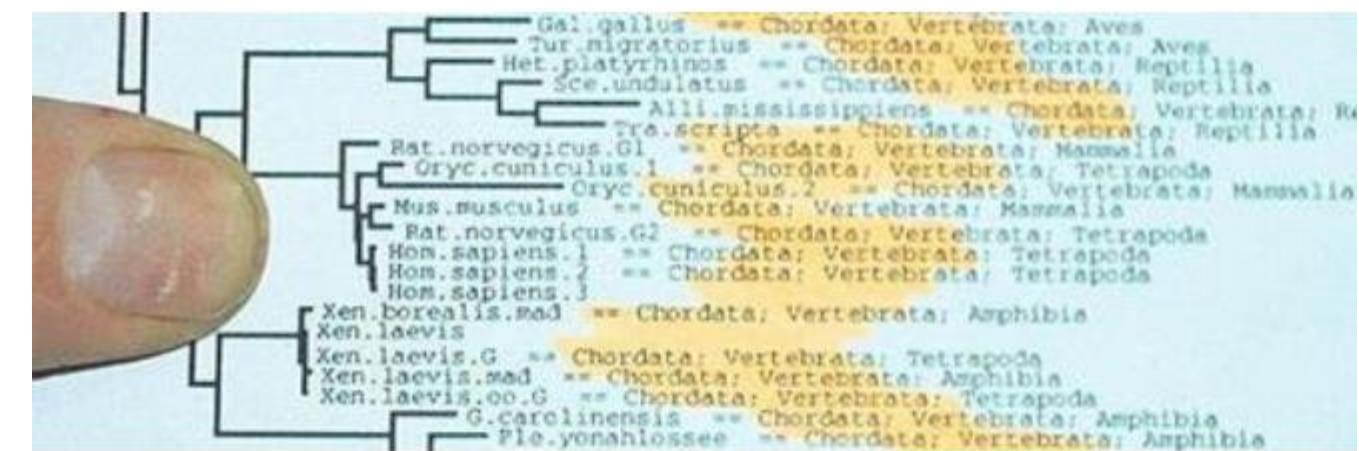
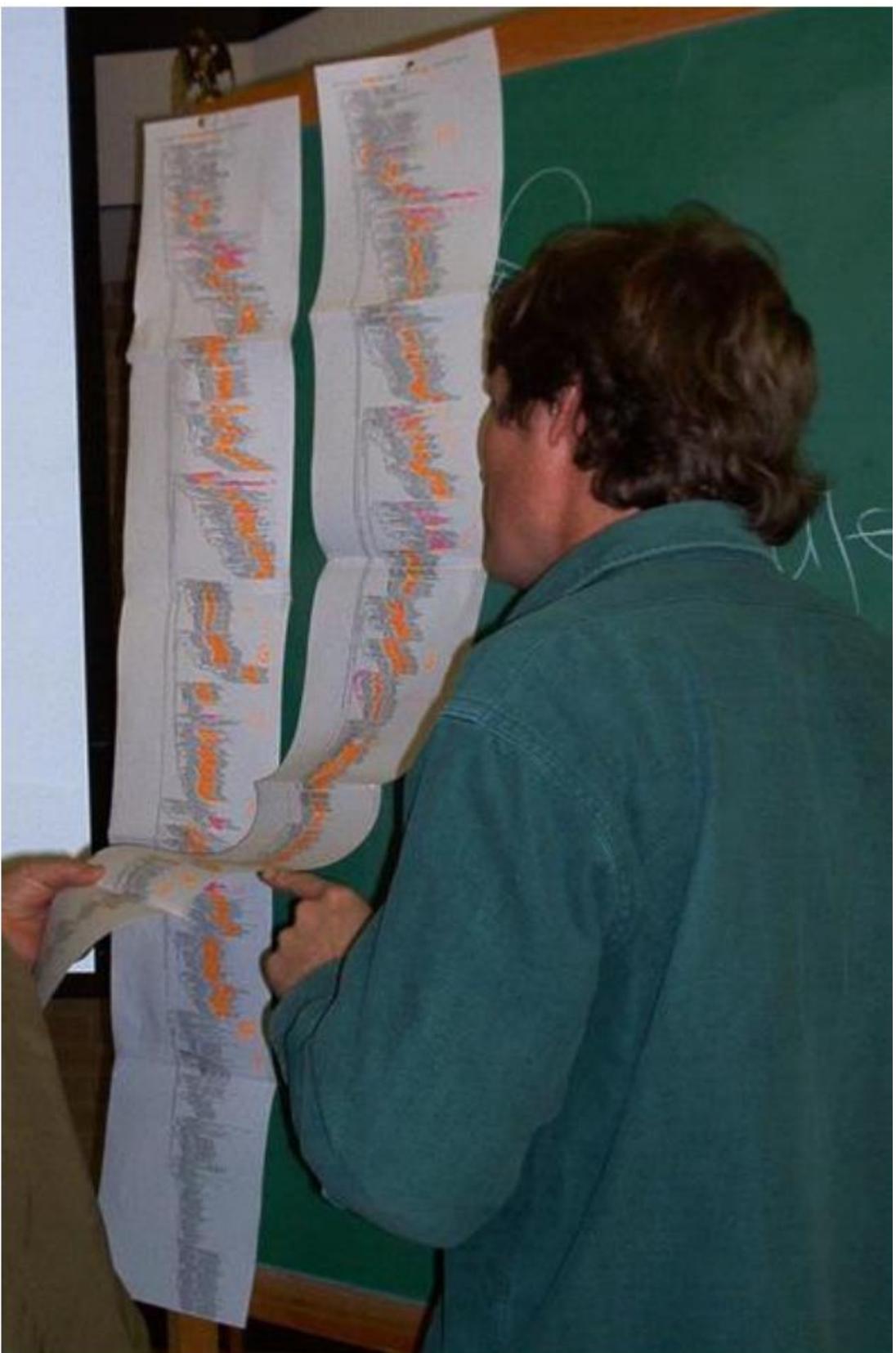


## → Spatial Data

→ Shape



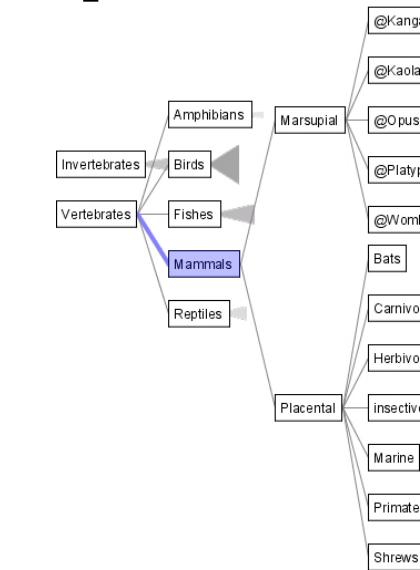
# A Preview of How



# Analysis example: Compare idioms

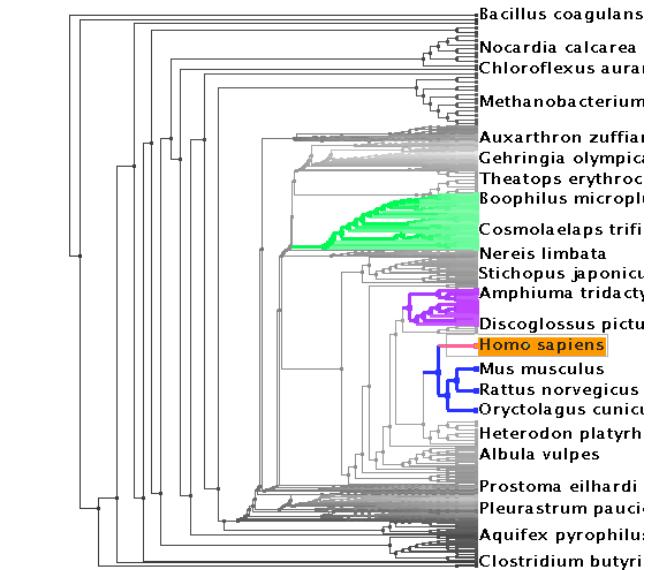
- imposes structure on huge design space
  - scaffold to help you think systematically about choices
  - analyzing existing as stepping stone to designing new
  - most possibilities ineffective for particular task/data combination

SpaceTree



[SpaceTree: Supporting Exploration in Large Node Link Tree, Design Evolution and Empirical Evaluation. Grosjean, Plaisant, and Bederson. Proc. InfoVis 2002, p 57–64.]

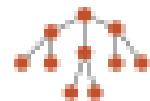
TreeJuxtaposer



[TreeJuxtaposer: Scalable Tree Comparison Using Focus+Context With Guaranteed Visibility. ACM Trans. on Graphics (Proc. SIGGRAPH) 22:453– 462, 2003.]

What?

⊕ Tree



Why?

⊕ Actions

- Present
- Locate
- Identify



⊕ Targets

- Path between two nodes



How?

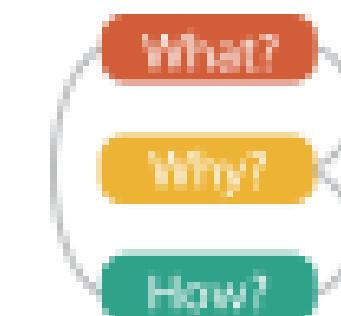
⊕ SpaceTree

- Encode
- Navigate
- Select
- Filter
- Aggregate



⊕ TreeJuxtaposer

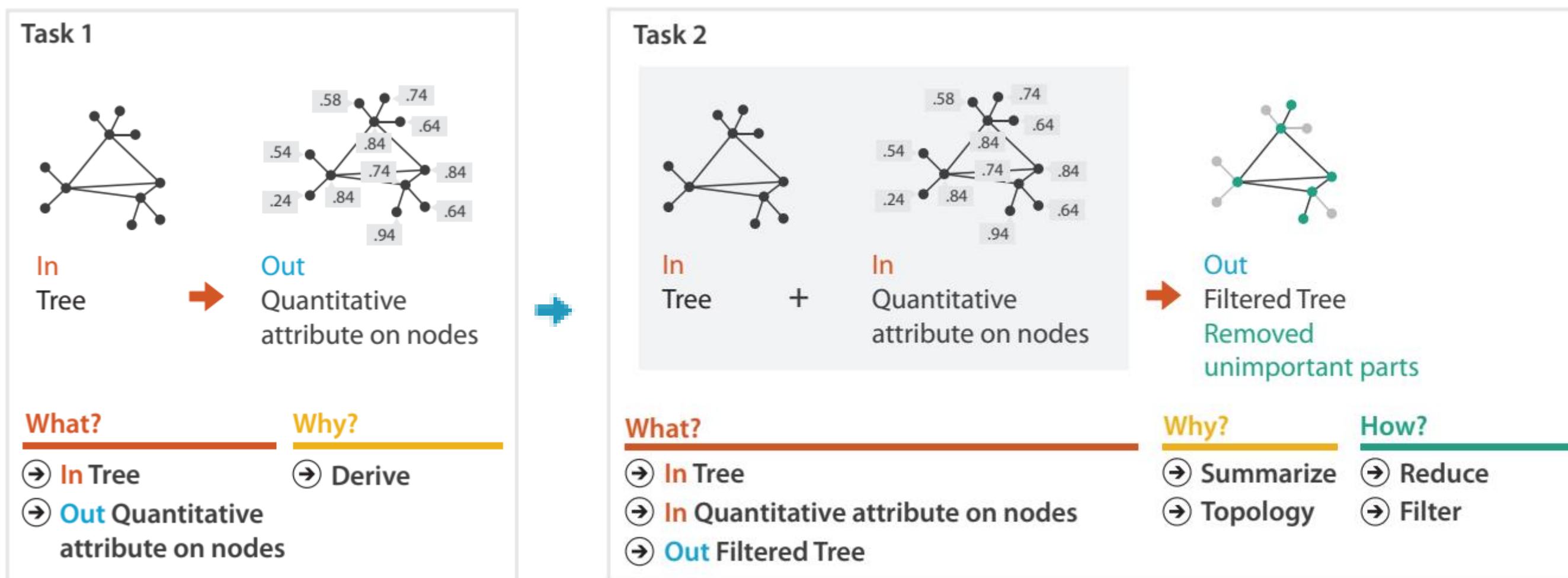
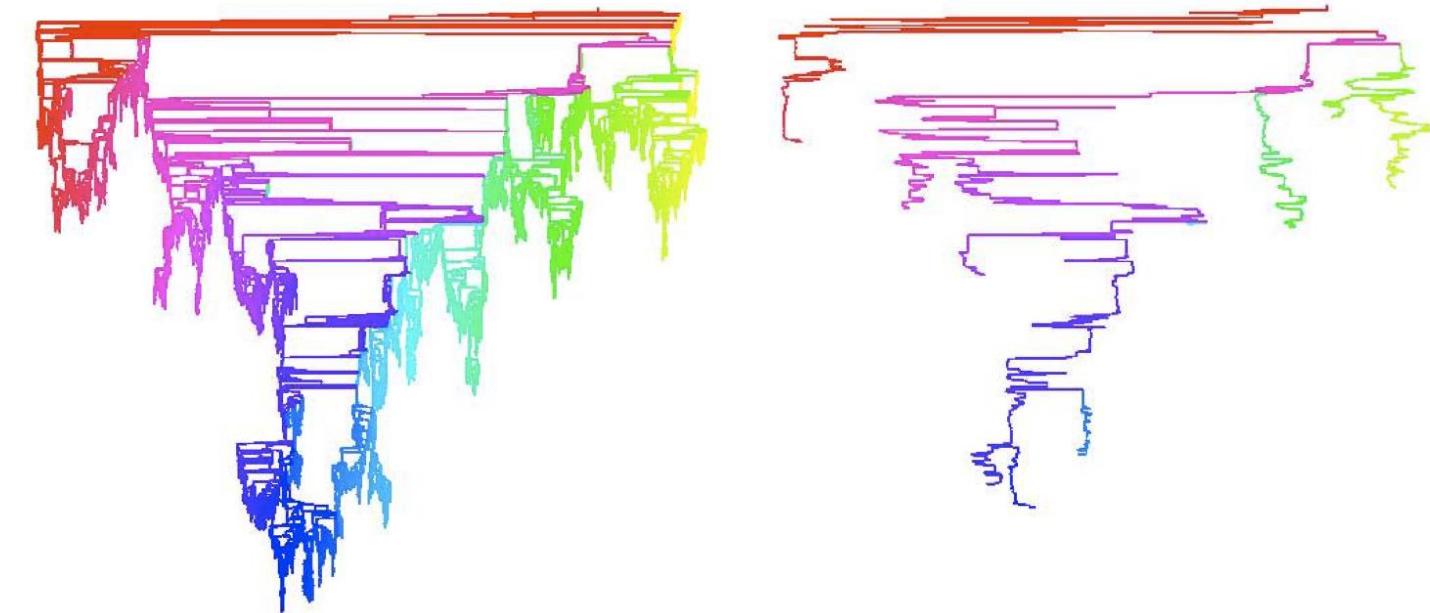
- Encode
- Navigate
- Select
- Arrange



# Analysis example: Derive one attribute

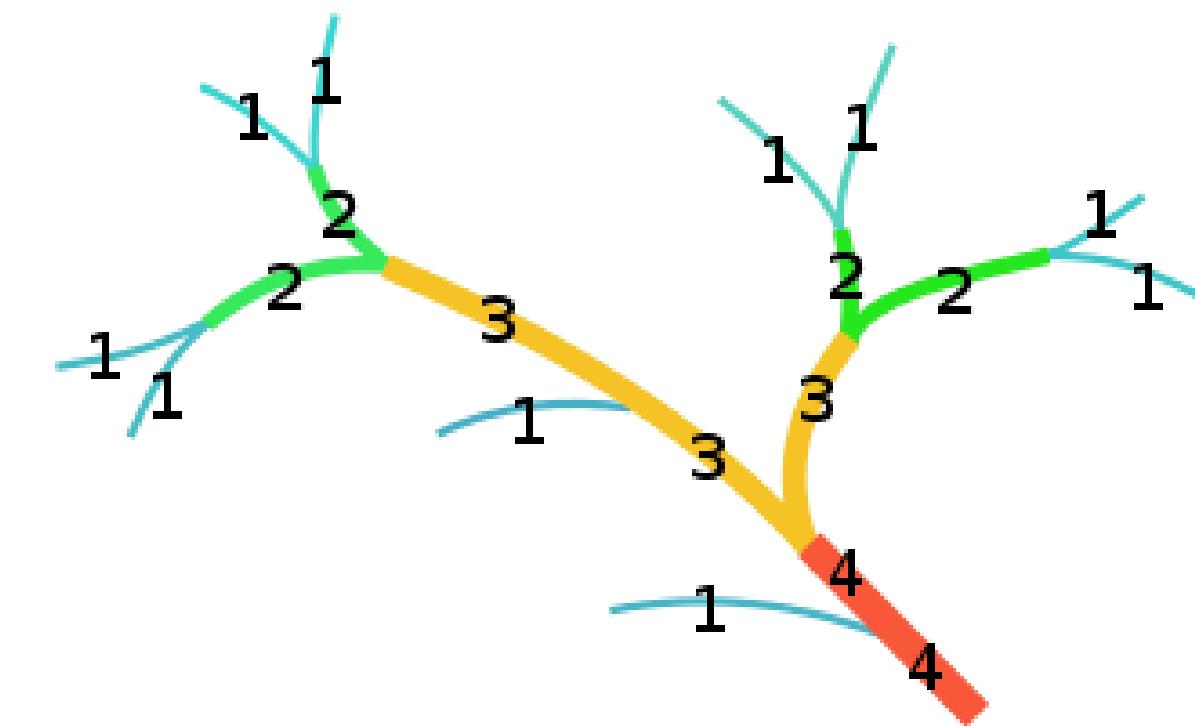
- Strahler number
  - centrality metric for trees/networks
  - derived quantitative attribute
  - draw top 5K of 500K for good skeleton

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

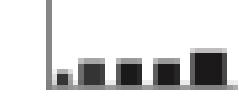
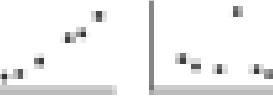
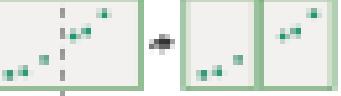
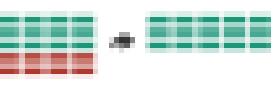
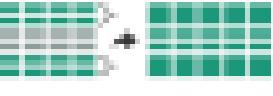


# Strahler number

- If the node is a leaf (**has no children**), its Strahler number is one.
- If the node has **one child with Strahler number  $i$** , and all other children have Strahler numbers less than  $i$ , then the Strahler number of the node is  $i$  again.
- If the node has **two or more children with Strahler number  $i$** , and no children with greater number, then the Strahler number of the node is  $i + 1$ .



# How?

Encode	Manipulate	Facet	Reduce
<p>⊕ Arrange → Express </p> <p>⊕ Separate </p> <p>⊕ Order </p> <p>⊕ Align </p> <p>⊕ Use </p>	<p>⊕ Map from categorical and ordered attributes → Color → Hue  → Saturation  → Luminance  → Size, Angle, Curvature, ... → Shape  → Motion Direction, Rate, Frequency, ... </p>	<p>⊕ Change </p> <p>⊕ Select </p> <p>⊕ Navigate </p>	<p>⊕ Juxtapose </p> <p>⊕ Partition </p> <p>⊕ Superimpose </p> <p>⊕ Filter </p> <p>⊕ Aggregate </p> <p>⊕ Embed </p>
<p>What?</p> <p>Why?</p> <p>How?</p>			

# Further reading

- **Visualization Analysis and Design.** Munzner. AK Peters Visualization Series, CRC Press, 2014.
  - *Chap 2: What: Data Abstraction*
  - *Chap 3: Why: Task Abstraction*
- *A Multi-Level Typology of Abstract Visualization Tasks.* Brehmer and Munzner. IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis) 19:12 (2013), 2376–2385.
- *Low-Level Components of Analytic Activity in Information Visualization.* Amar, Eagan, and Stasko. Proc. IEEE InfoVis 2005, p 111–117.
- *A taxonomy of tools that support the fluent and flexible use of visualizations.* Heer and Shneiderman. Communications of the ACM 55:4 (2012), 45–54.
- *Rethinking Visualization: A High-Level Taxonomy.* Tory and Möller. Proc. IEEE InfoVis 2004, p 151–158.
- **Visualization of Time-Oriented Data.** Aigner, Miksch, Schumann, and Tominski. Springer, 2011.

# Reading visualization papers

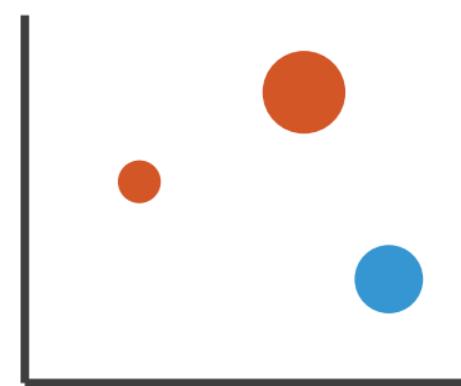
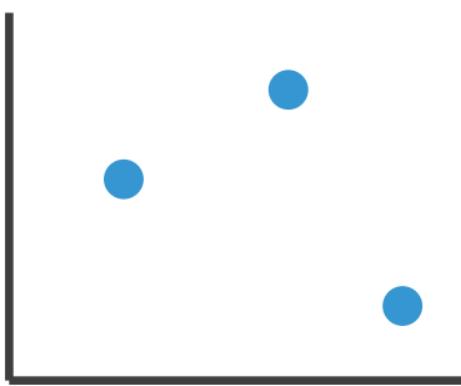
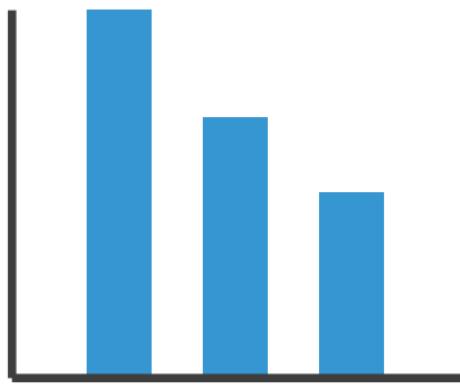
- one strategy: multiple passes
  - title
  - abstract, authors/affiliation
  - flip through, glance at figures, notice structure from section titles
  - skim intro, results/discussion (maybe conclusion)
  - fast read to **get big ideas**
    - if you don't get something, just keep going
  - second pass to **work through details**
    - later parts may cast light on earlier parts for badly structured papers
  - third pass to **dig deep**
    - if it's highly relevant, or you're presenting it to class
- literature search
  - decide when to stop reading: is this relevant to my current concerns?

# Marks and Channels

## Ch. 5

# Visual encoding

- analyze idiom structure



# Definitions: Marks

- **marks**
  - geometric primitives

→ Points



→ Lines



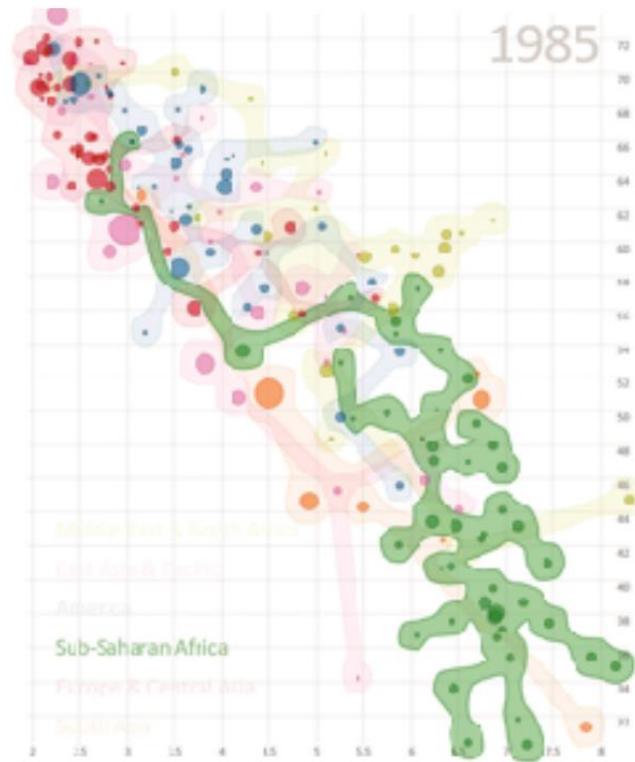
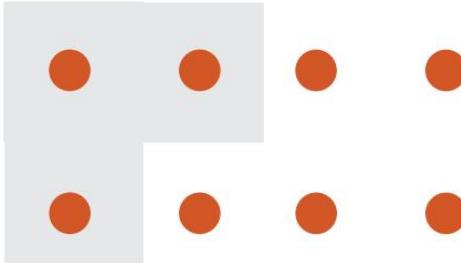
→ Areas



- 
- **channels**
    - control appearance of marks
    - can redundantly code with multiple channels

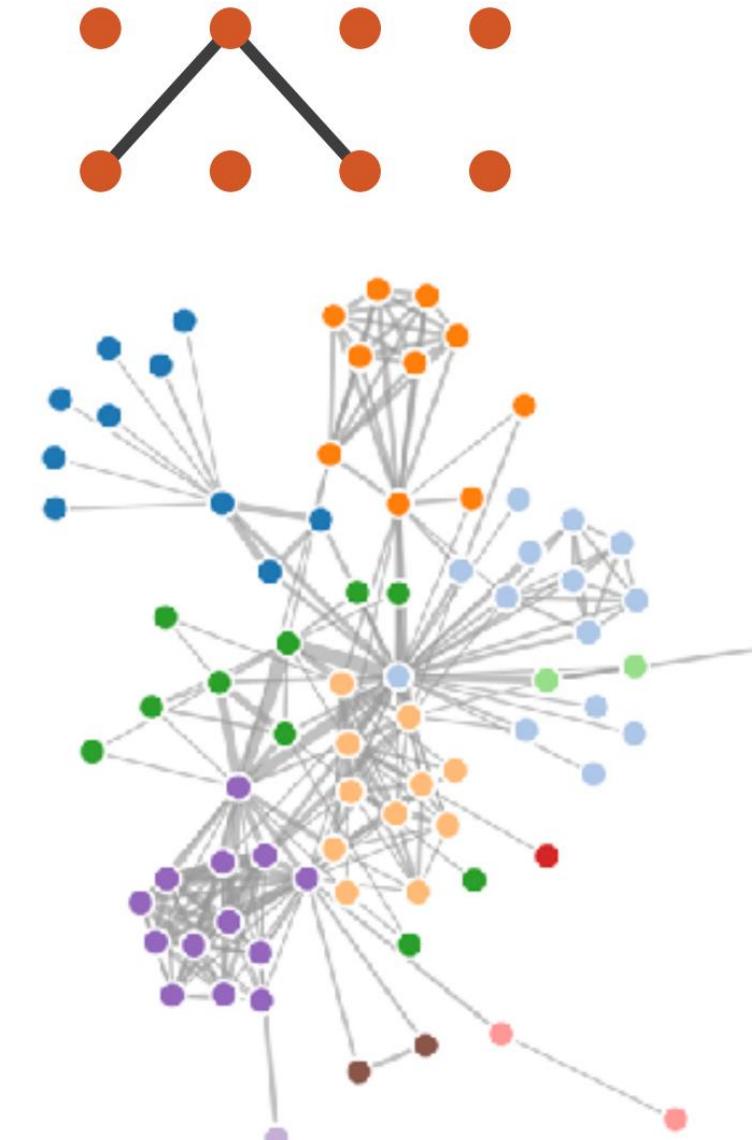
# Marks for links

## → Containment



[vialab.science.uoit.ca/portfolio/bubblesets](https://vialab.science.uoit.ca/portfolio/bubblesets)

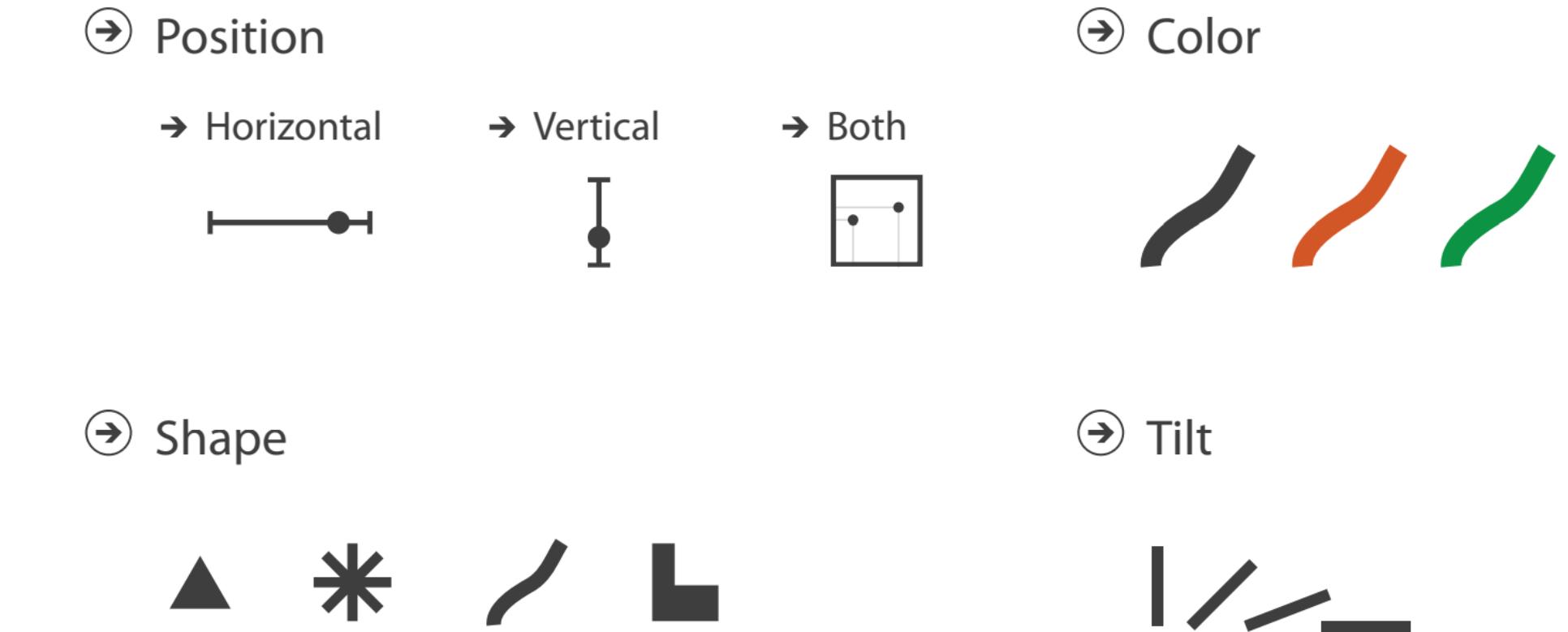
## → Connection



<https://observablehq.com/@d3/force-directed-graph>

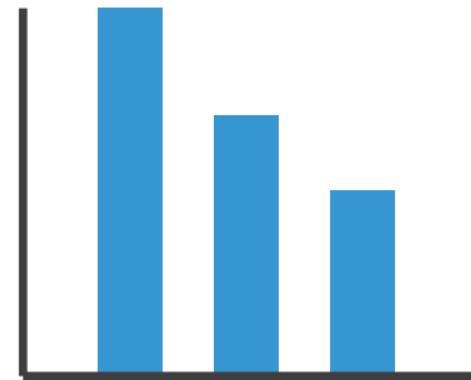
# Definitions: Channels

- control appearance of marks
  - proportional to or based on attributes
- many names
  - visual channels
  - visual variables
  - retinal channels
  - visual dimensions
  - ...

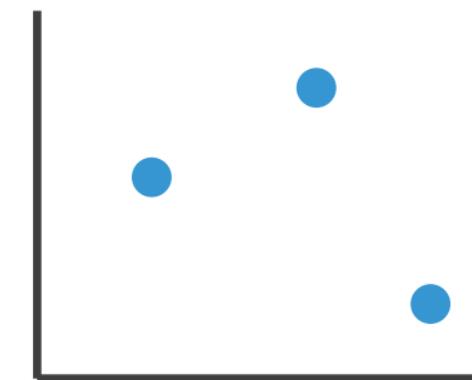


# Visual encoding

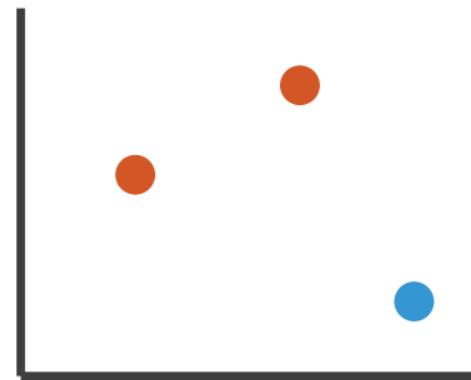
- analyze idiom structure
  - as combination of marks and channels



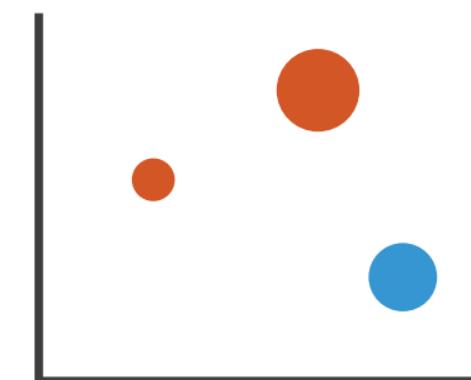
1:  
**vertical position**  
mark: line



2:  
**vertical position**  
**horizontal position**  
mark: point



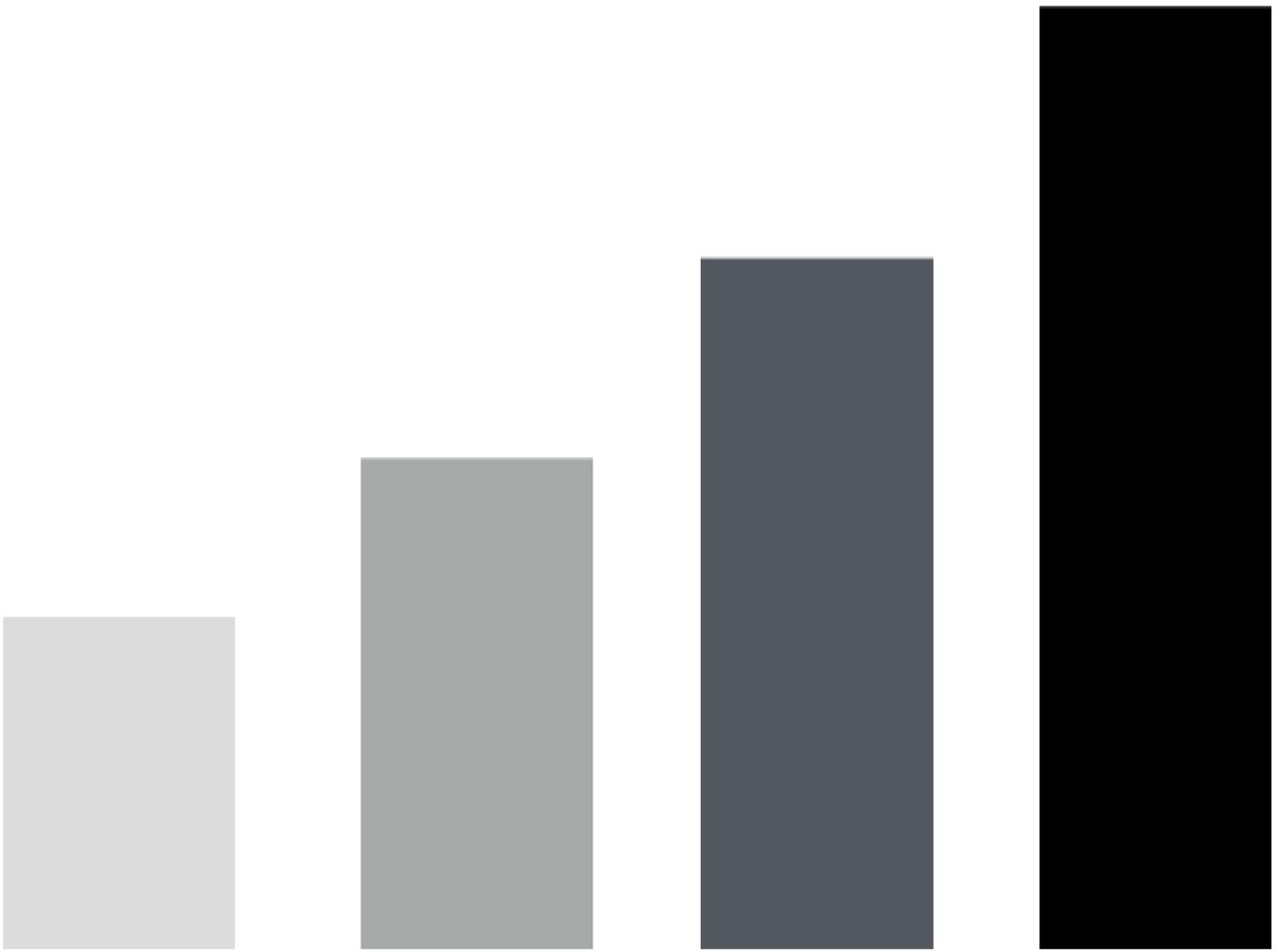
3:  
**vertical position**  
**horizontal position**  
**color hue**  
mark: point



4:  
**vertical position**  
**horizontal position**  
**color hue**  
**size (area)**  
mark: point

# Redundant encoding

- multiple channels
  - sends stronger message
  - but uses up channels



Length and Luminance

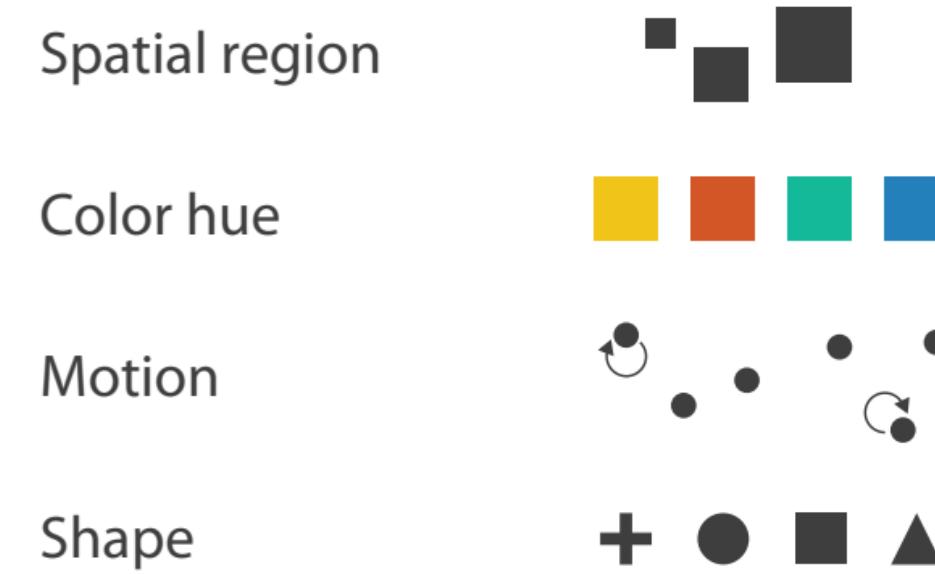
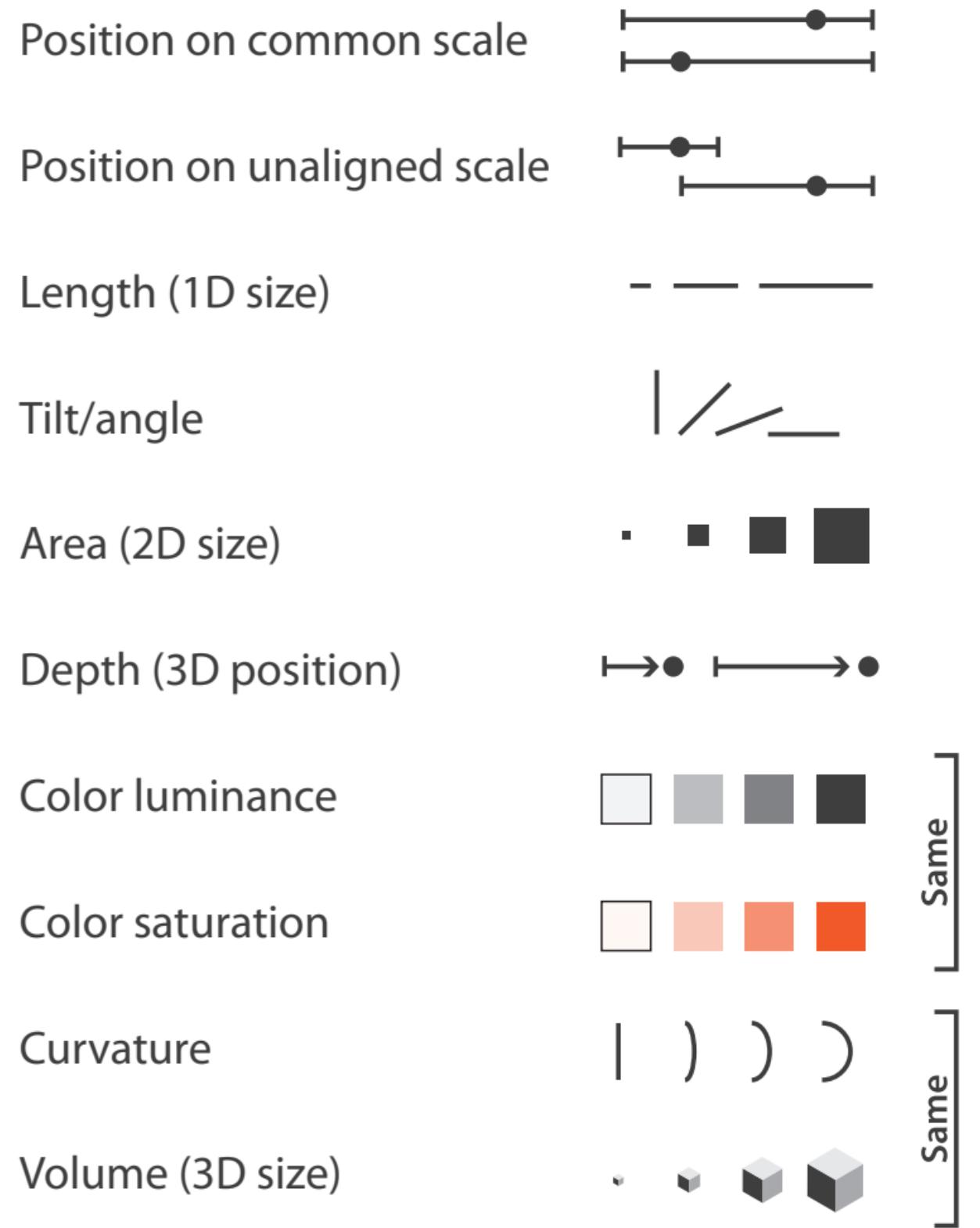
# Scope of analysis

- simplifying assumptions: one mark item, single view
- later on
  - multiple views
  - multiple marks in a region (glyph)
  - some items not represented by marks (aggregation and filtering)

# When to use which channel?

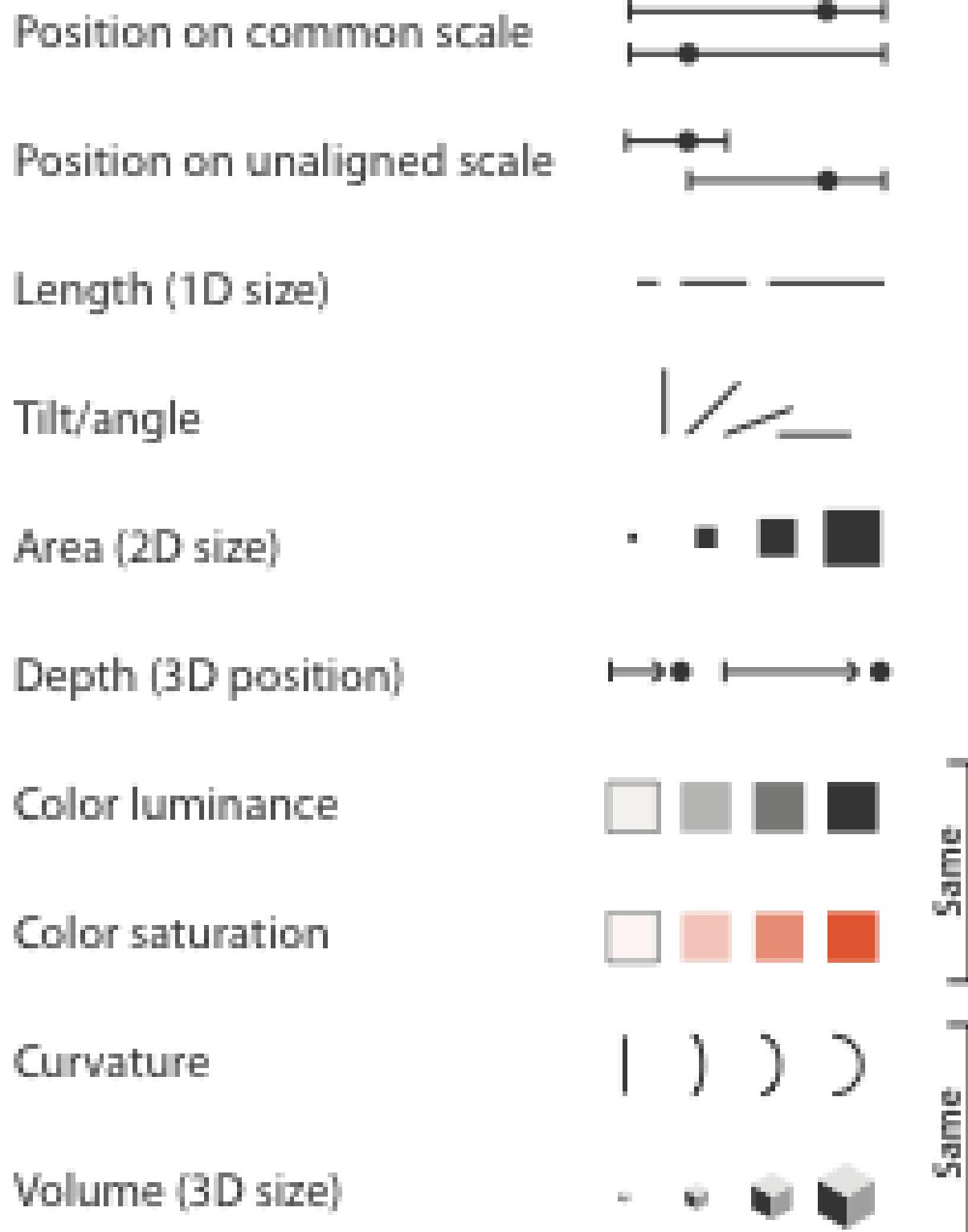
- **expressiveness**
  - match channel type to data type
- **effectiveness**
  - some channels are better than others

# Channels: Expressiveness types and effectiveness rankings

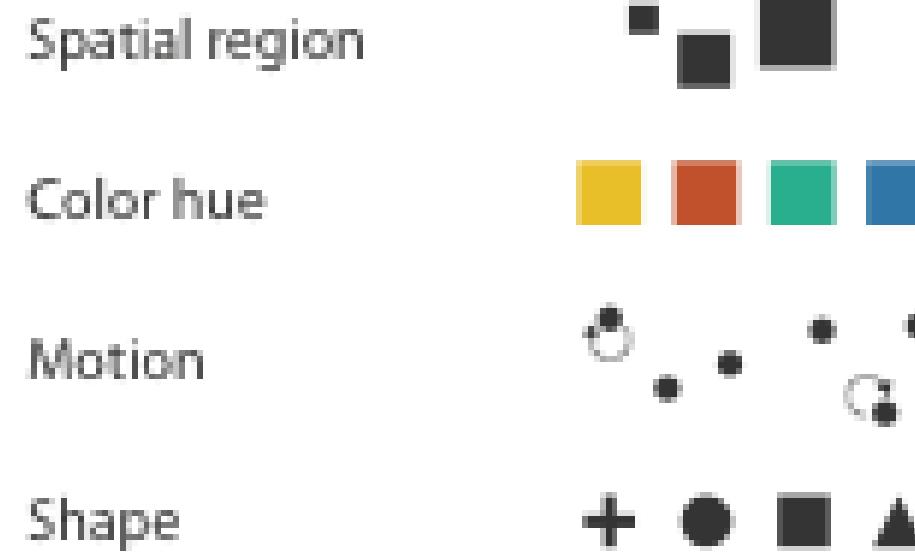


# Channels: Matching Types

## ⊕ Magnitude Channels: Ordered Attributes



## ⊕ Identity Channels: Categorical Attributes



- expressiveness principle
  - match channel and data characteristics

# Channels: Rankings

## ⊕ Magnitude Channels: Ordered Attributes

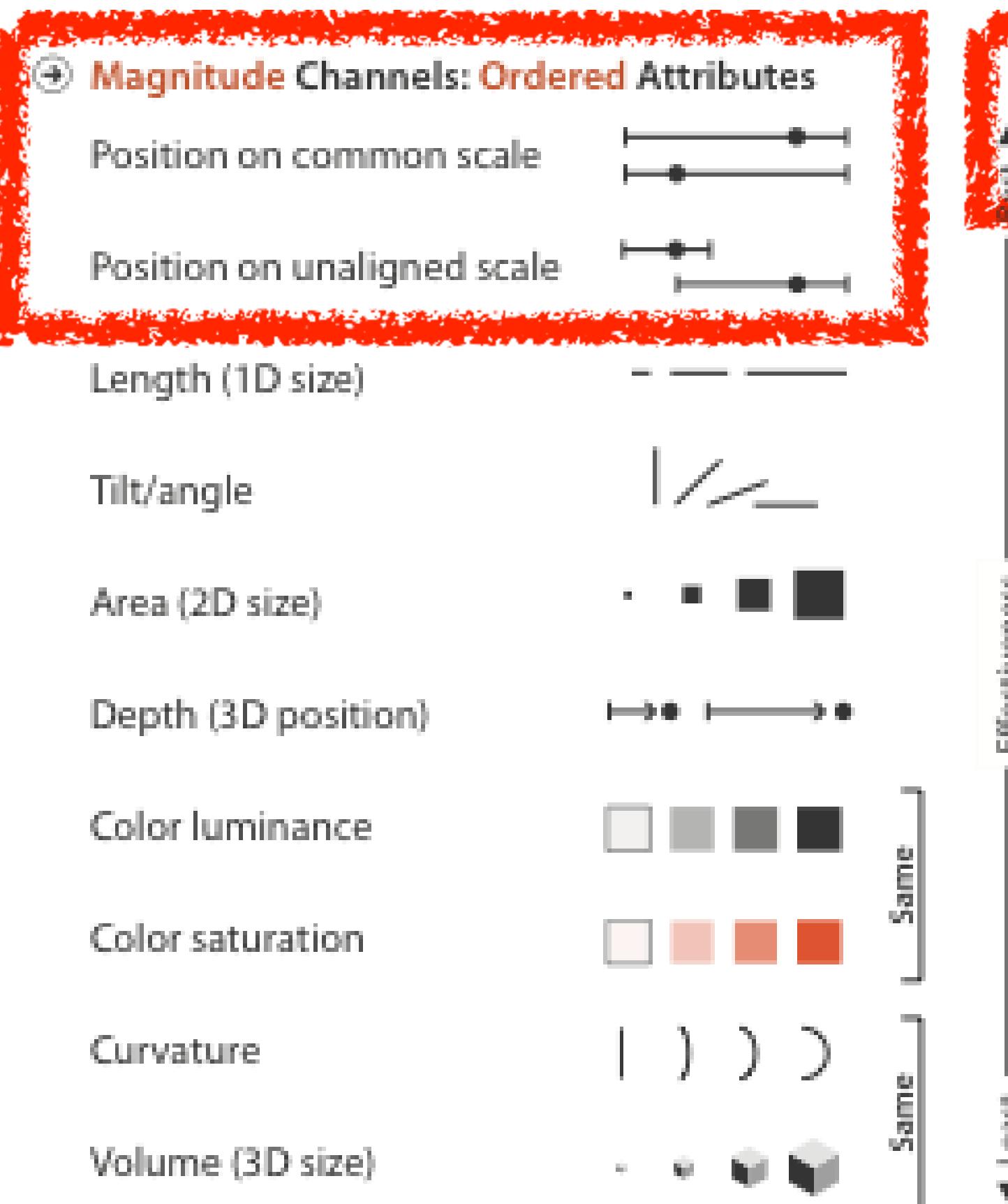


## ⊕ Identity Channels: Categorical Attributes



- **expressiveness principle**
  - match channel and data characteristics
- **effectiveness principle**
  - encode most important attributes with highest ranked channels

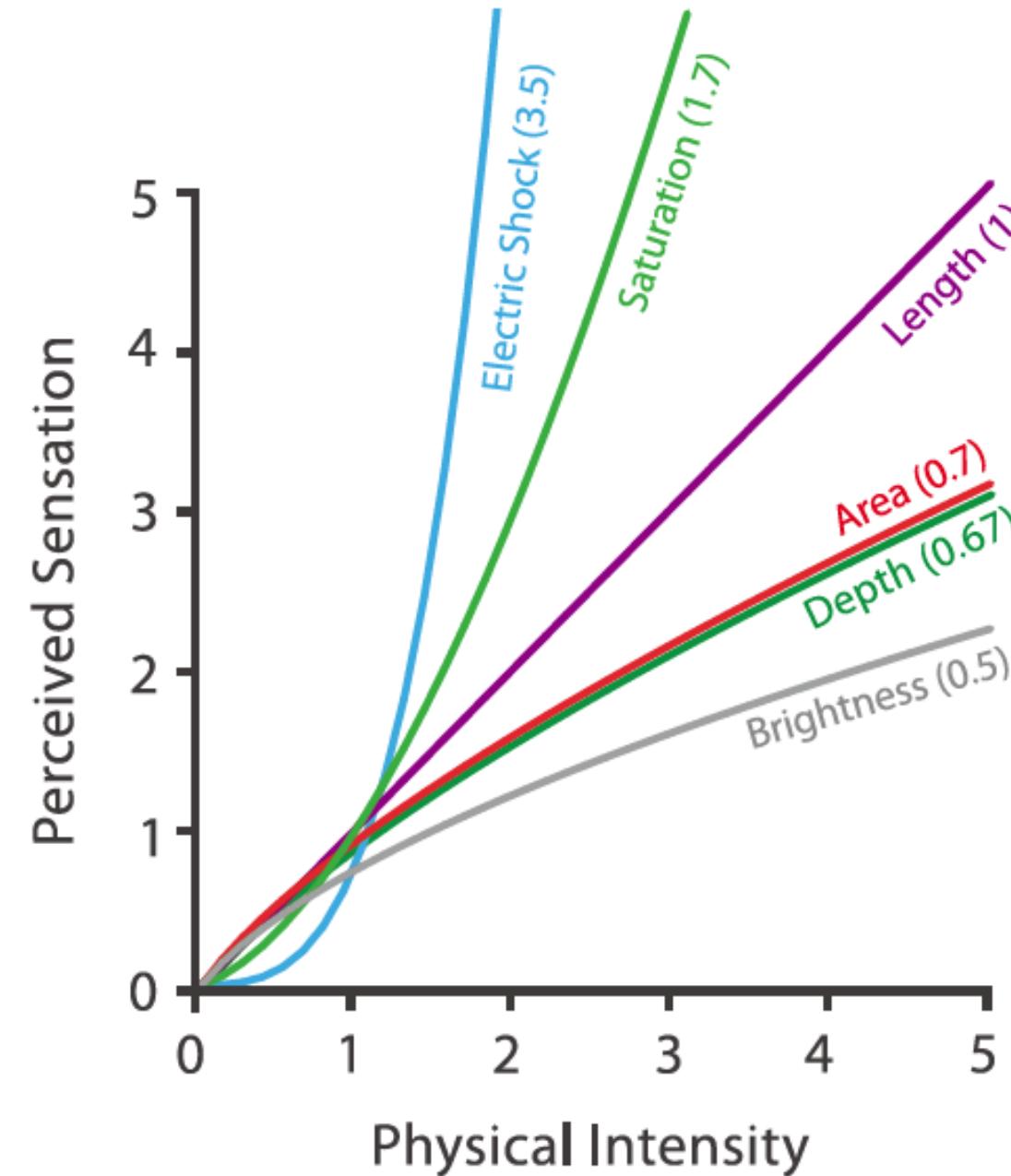
# Channels: Expressiveness types and effectiveness rankings



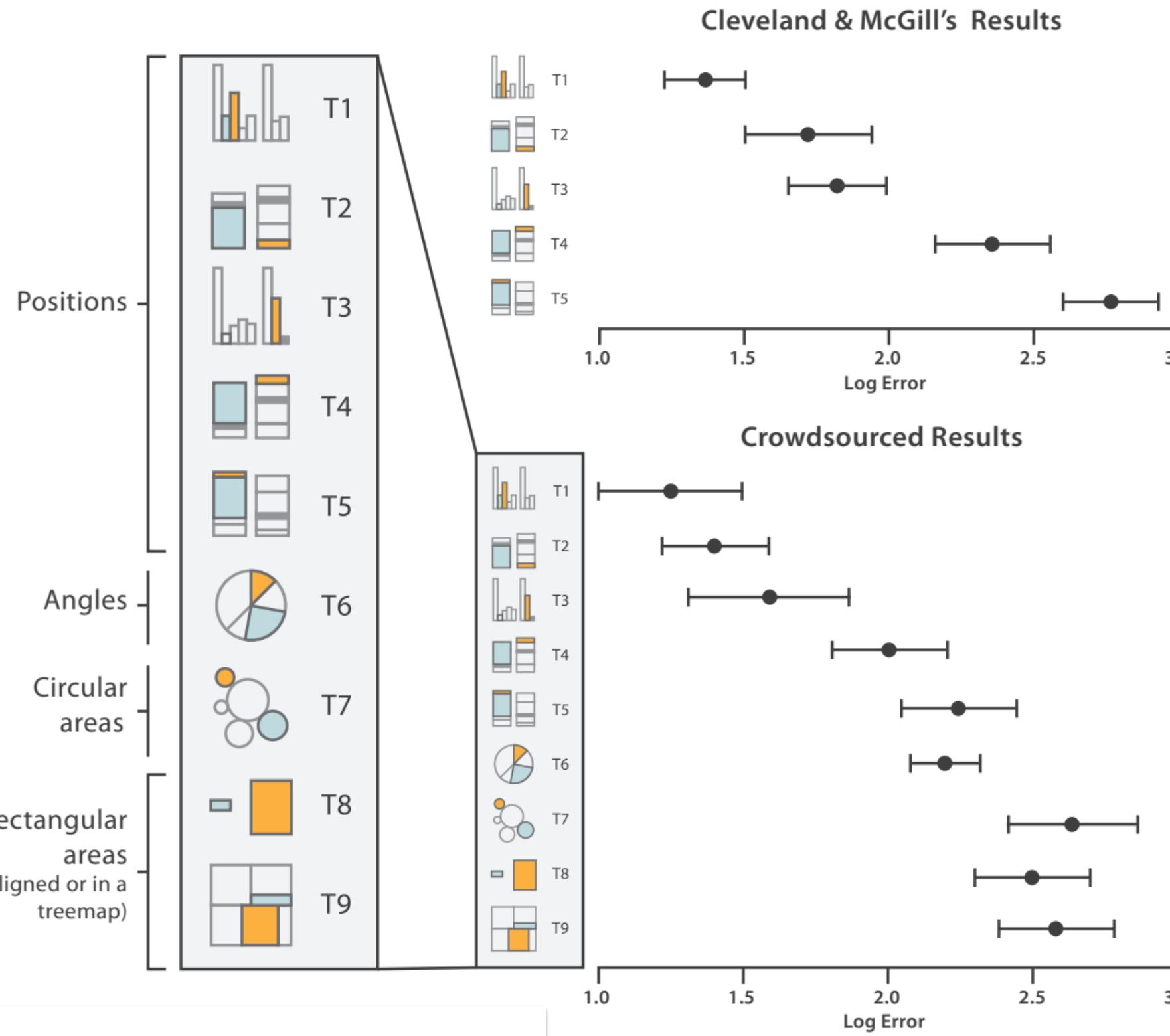
- **expressiveness principle**
  - match channel and data characteristics
- **effectiveness principle**
  - encode most important attributes with highest ranked channels
  - spatial position ranks high for both<sup>74</sup>

# Accuracy: Fundamental Theory

Steven's Psychophysical Power Law:  $S = I^n$



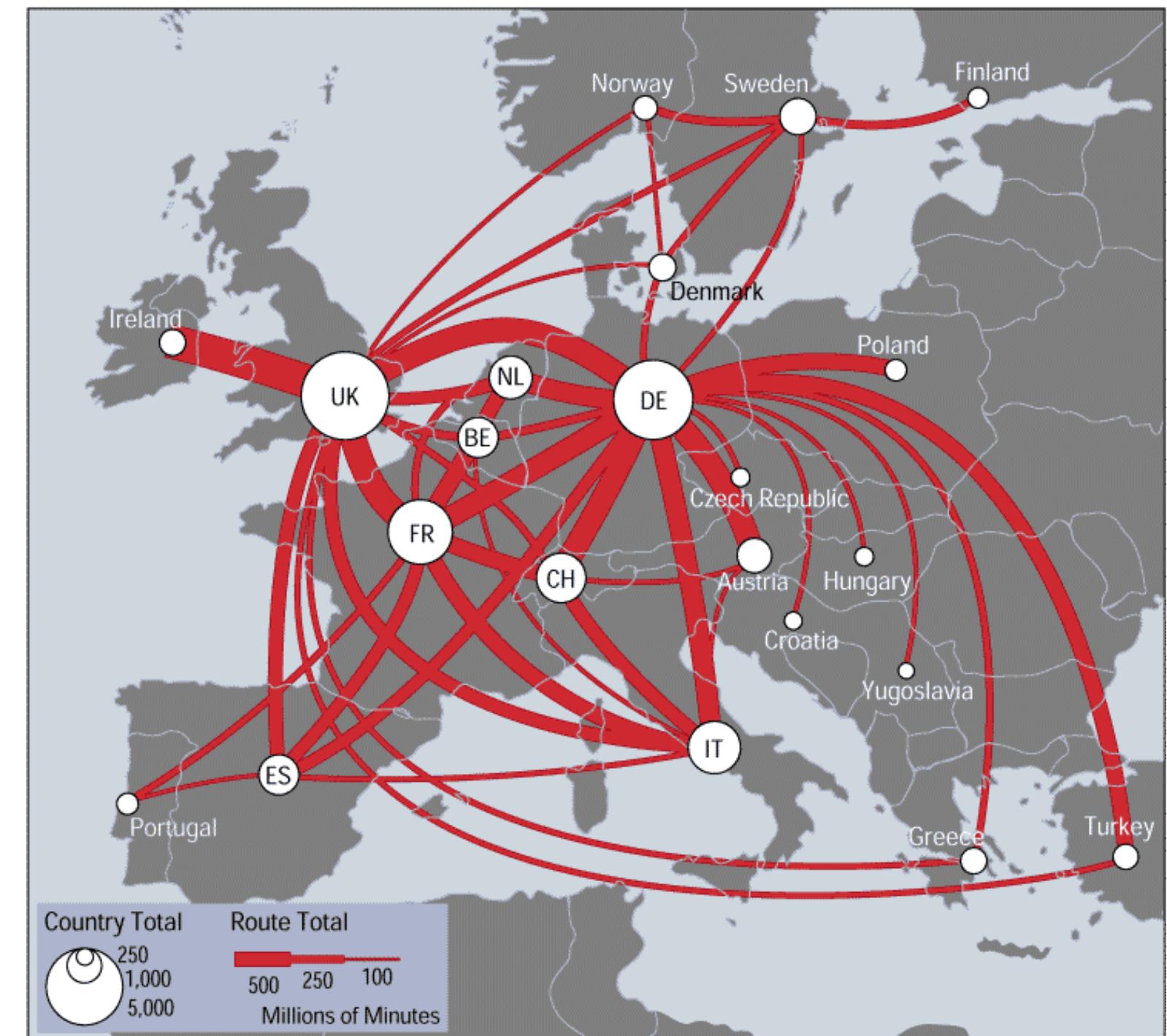
# Accuracy: Vis experiments



*[Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. Heer and Bostock. Proc ACM Conf. Human Factors in Computing Systems (CHI) 2010 p. 203–212.]*

# Discriminability: How many usable steps?

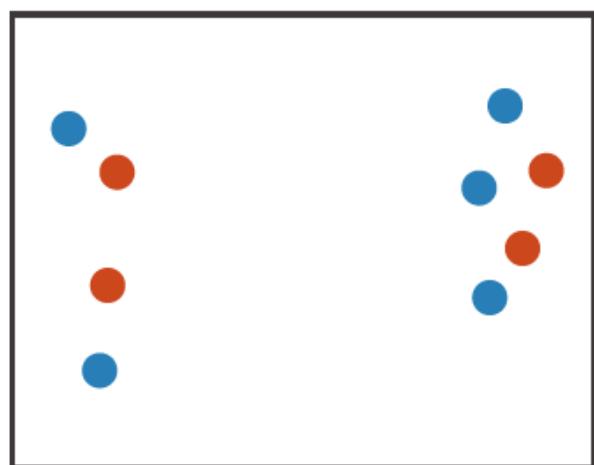
- must be sufficient for number of attribute levels to show
  - linewidth: few bins



[\[mappa.mundi.net/maps/maps\\_014/telegeography.html\]](http://mappa.mundi.net/maps/maps_014/telegeography.html)

# Separability vs. Integrality

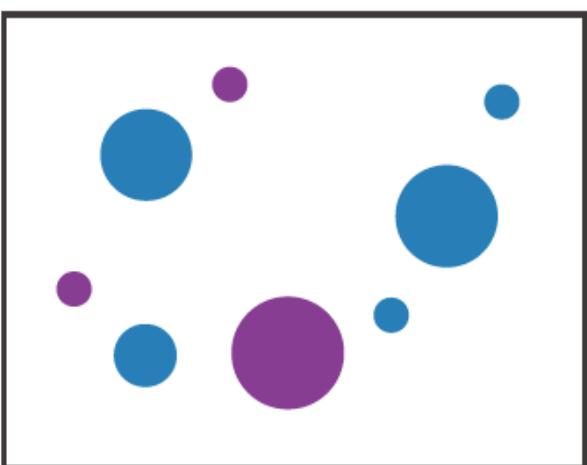
Position  
+ Hue (Color)



Fully separable

2 groups each

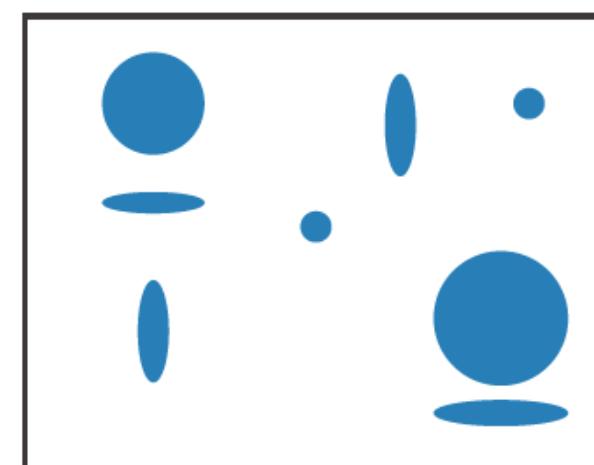
Size  
+ Hue (Color)



Some interference

2 groups each

Width  
+ Height



Some/significant  
interference

3 groups total:  
integral area

Red  
+ Green

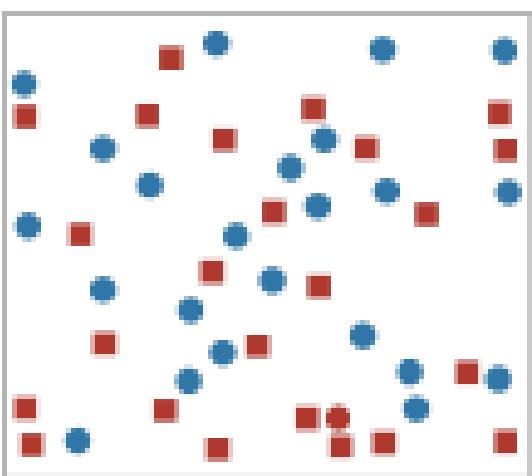
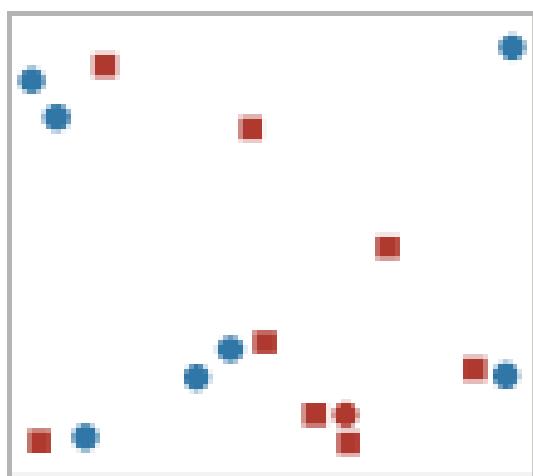
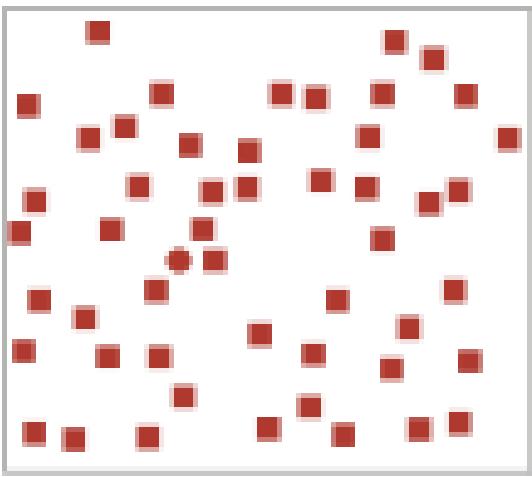
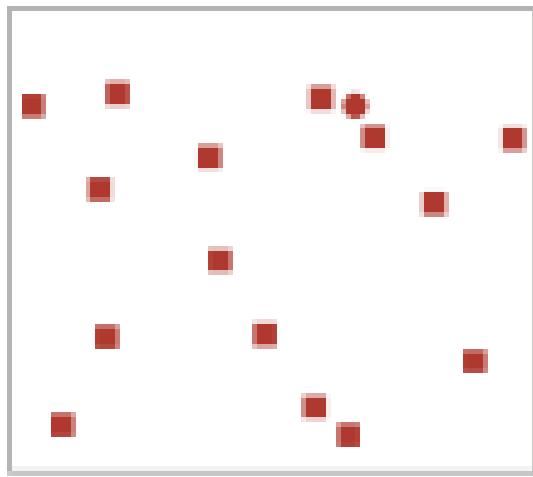
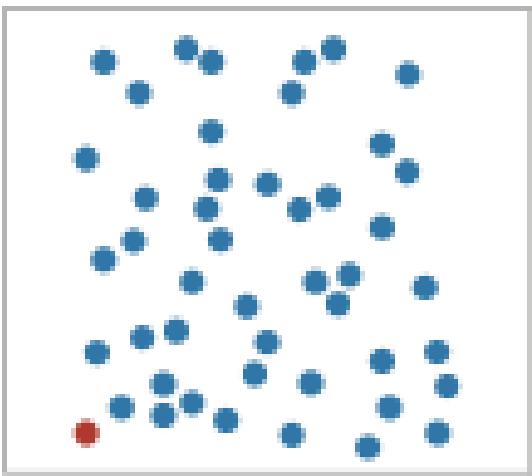
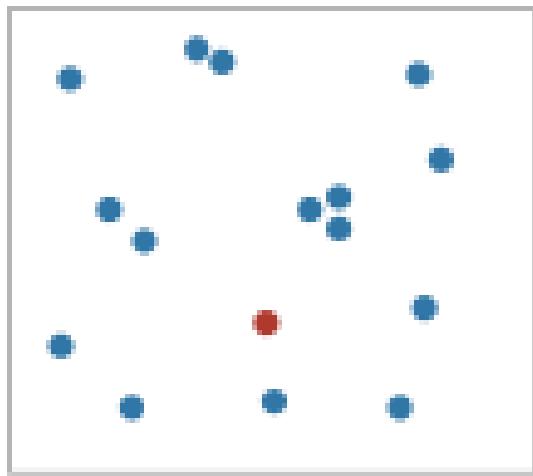


Major interference

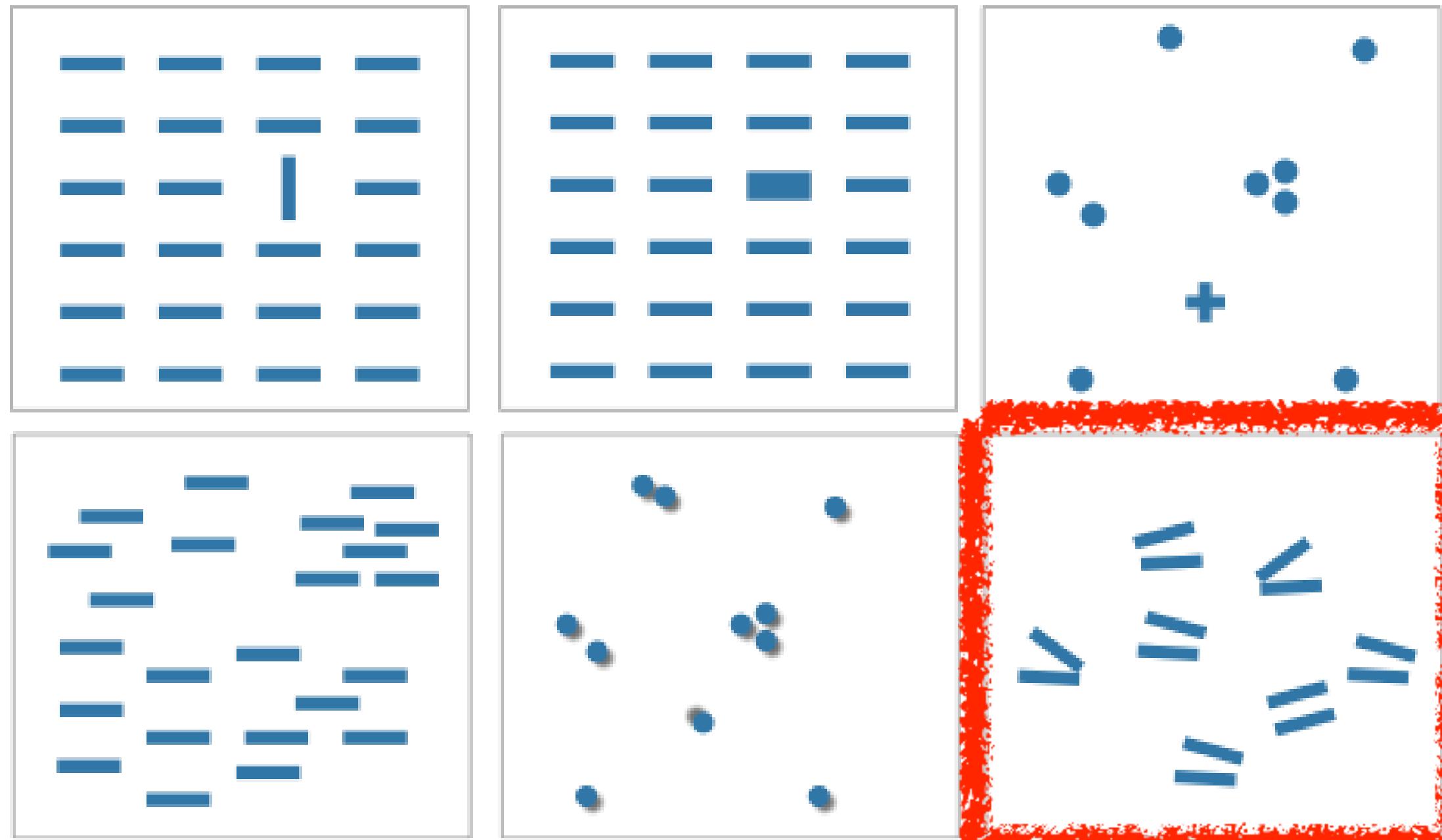
4 groups total:  
integral hue

# Popout

- find the red dot
  - how long does it take?
- parallel processing on many individual channels
  - speed independent of distractor count
  - speed depends on channel and amount of difference from distractors
- serial search for (almost all) combinations
  - speed depends on number of distractors



# Popout



- many channels: tilt, size, shape, proximity, shadow direction, ...
- but not all! parallel line pairs do not pop out from tilted pairs

# Grouping

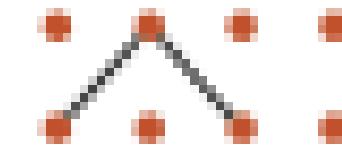
- containment
- connection
- proximity
  - same spatial region
- similarity
  - same values as other categorical channels

## Marks as Links

### ⊕ Containment



### ⊕ Connection

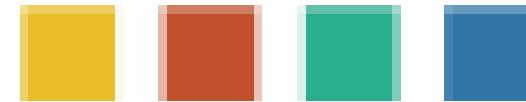


## ⊕ Identity Channels: Categorical Attributes

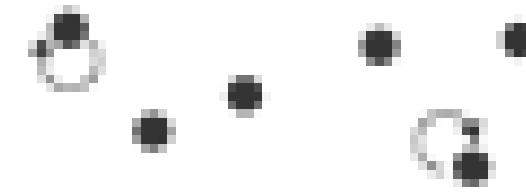
### Spatial region



### Color hue



### Motion

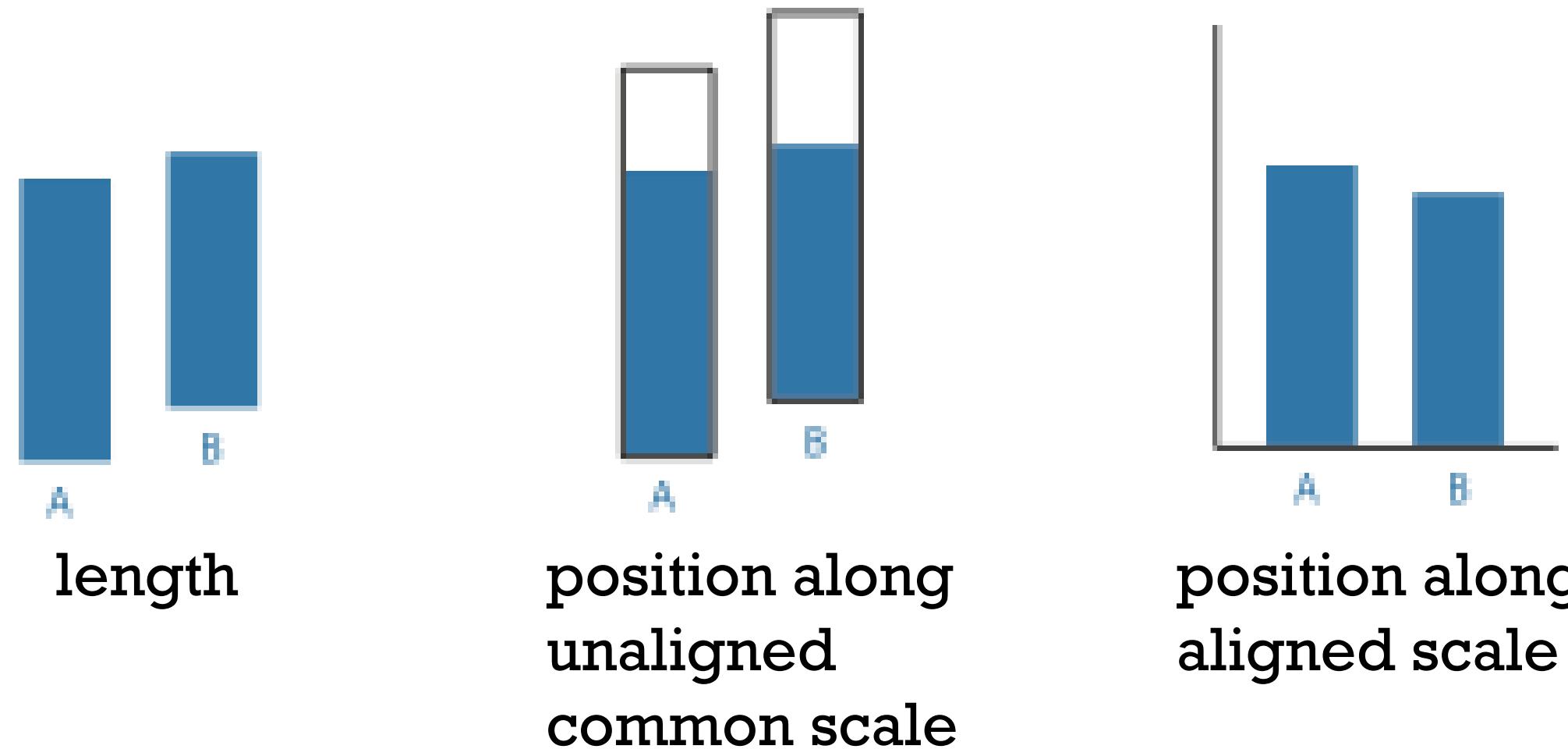


### Shape



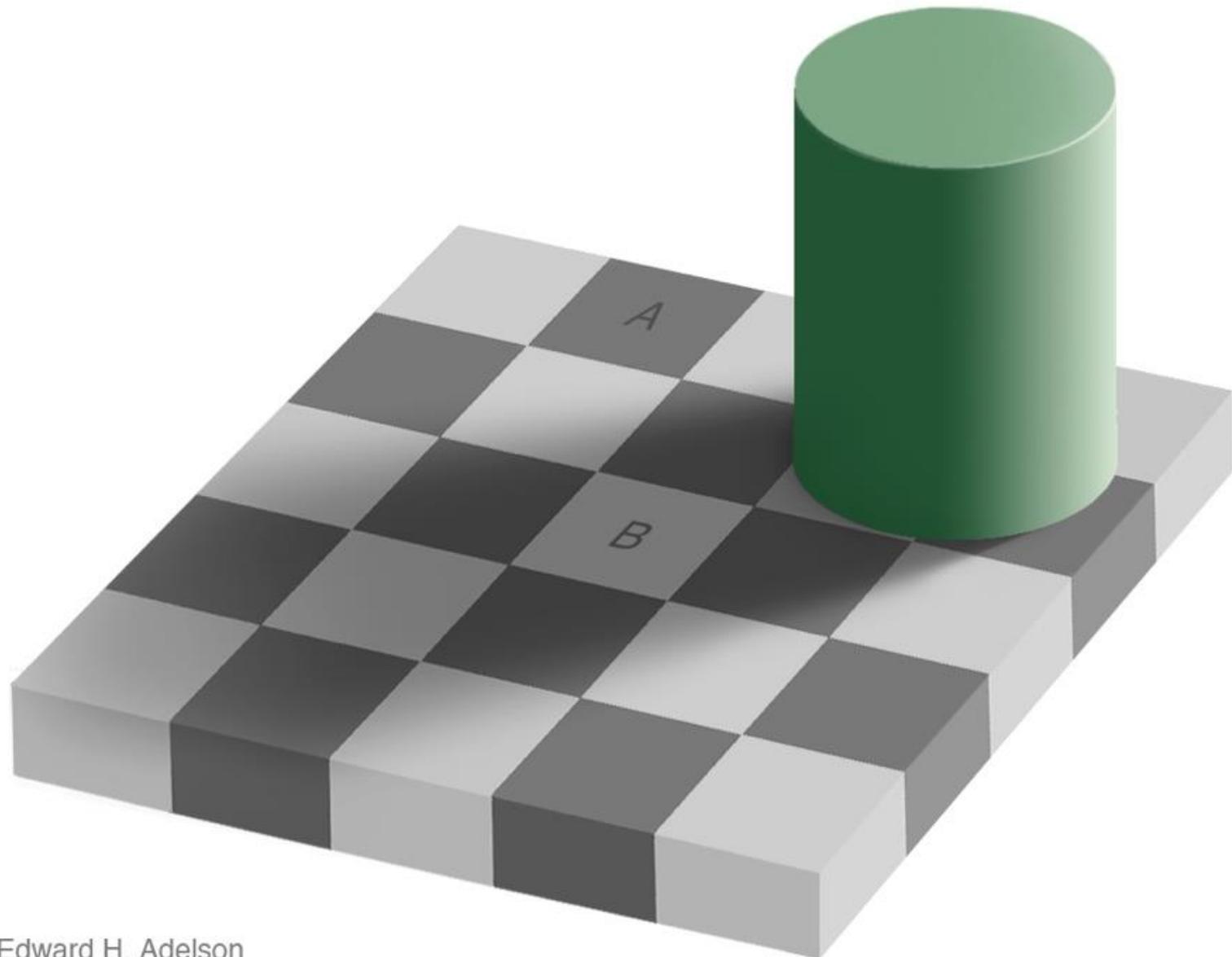
# Relative vs. absolute judgements

- perceptual system mostly operates with relative judgements, not absolute
  - that's why accuracy increases with common frame/scale and alignment
  - Weber's Law: ratio of increment to background is constant
    - filled rectangles differ in length by 1:9, difficult judgement
    - white rectangles differ in length by 1:2, easy judgement



# Relative luminance judgements

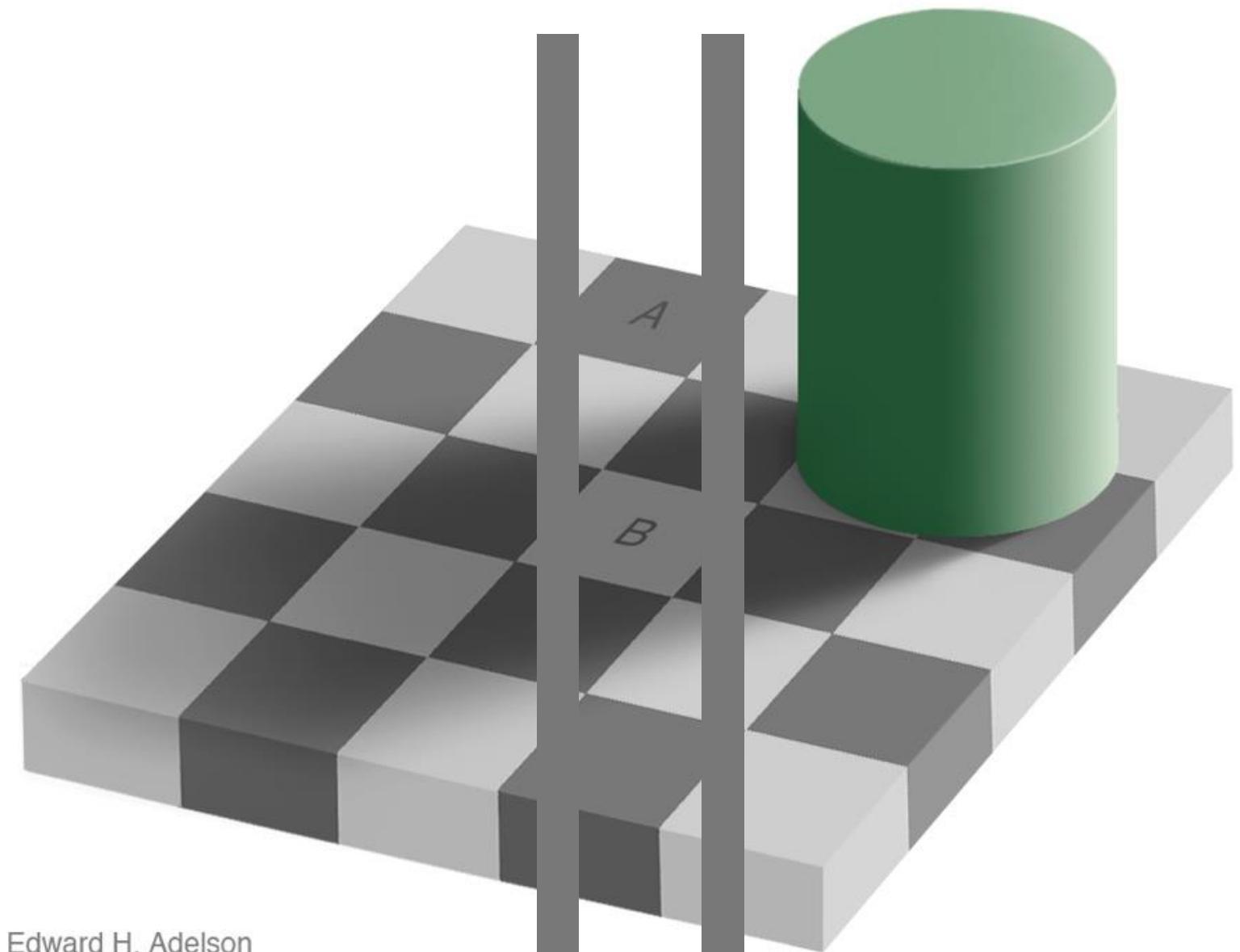
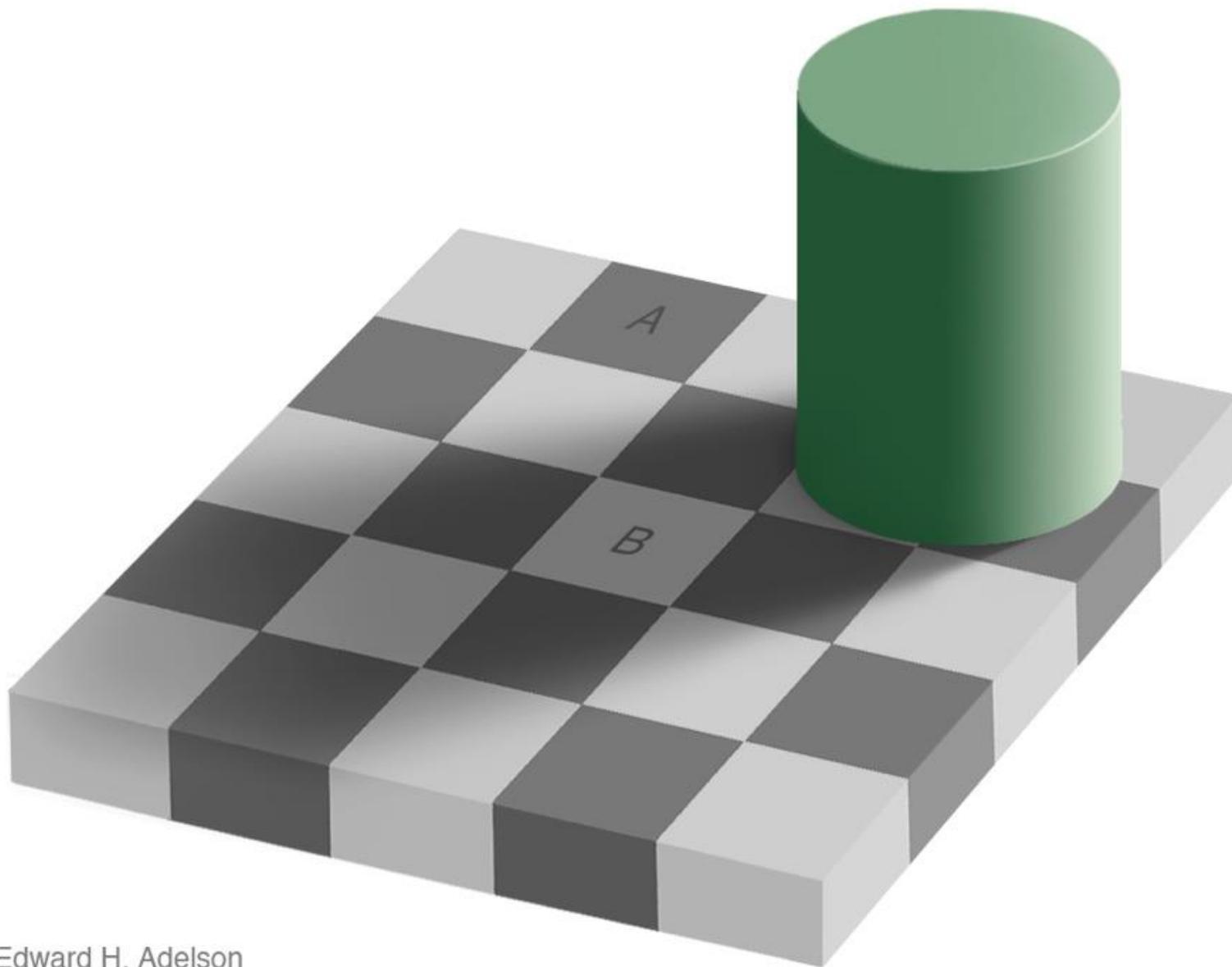
- perception of luminance is contextual based on contrast with surroundings



Edward H. Adelson

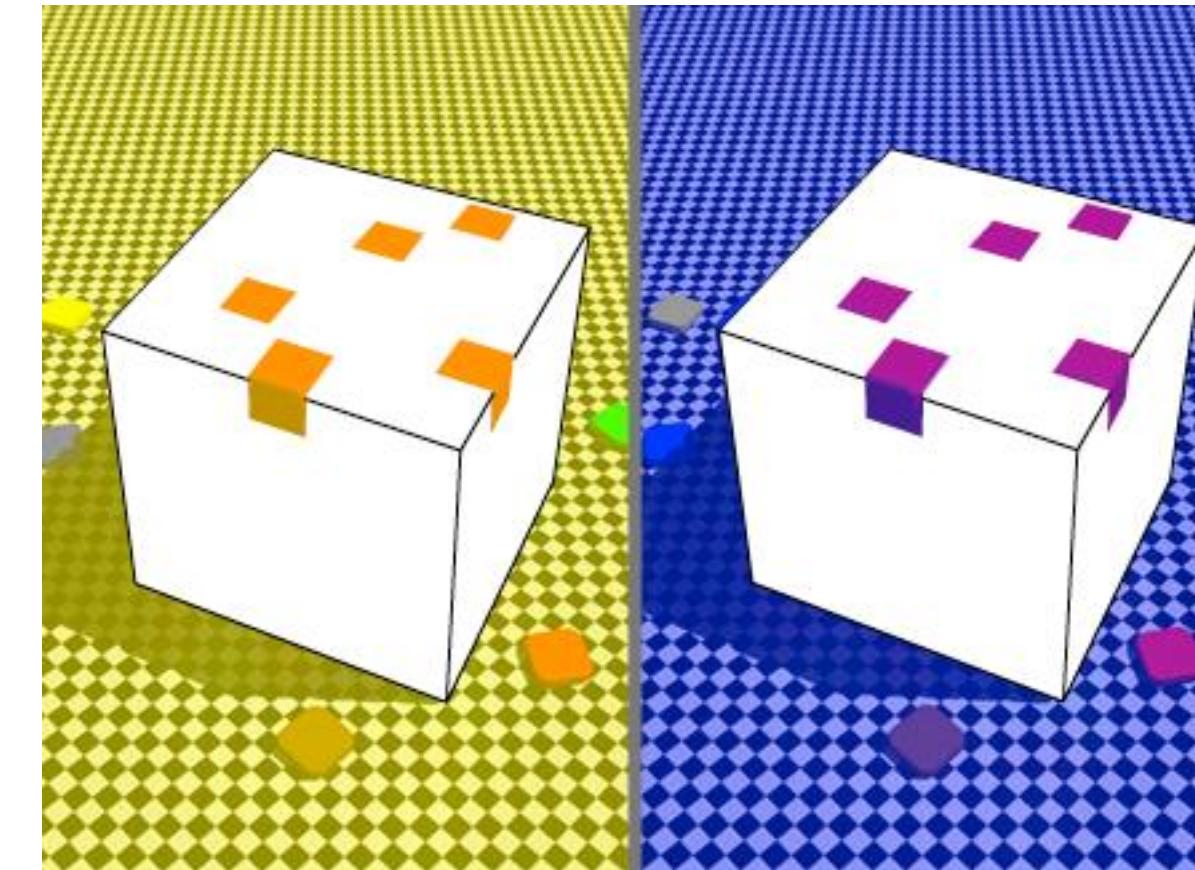
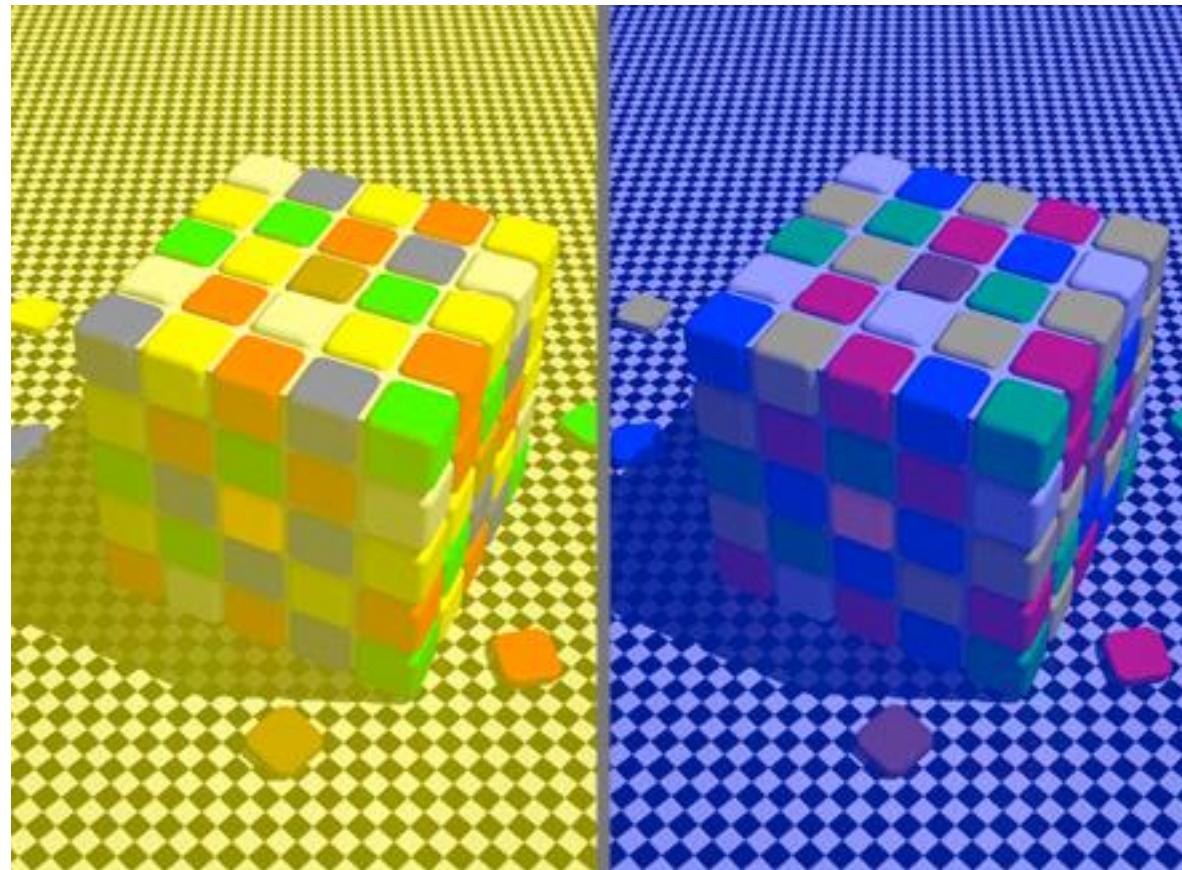
# Relative luminance judgements

- perception of luminance is contextual based on contrast with surroundings



# Relative color judgements

- color constancy across broad range of illumination conditions



# Further reading

- Visualization Analysis and Design. Munzner. AK Peters Visualization Series, CRC Press, 2014.
  - *Chap 5: Marks and Channels*
- *On the Theory of Scales of Measurement*. Stevens. Science 103:2684 (1946), 677–680.
- Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects. Stevens. Wiley, 1975.
- *Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods*. Cleveland and McGill. Journ. American Statistical Association 79:387 (1984), 531–554.
- *Perception in Vision*. Healey. <http://www.csc.ncsu.edu/faculty/healey/PP>
- Visual Thinking for Design. Ware. Morgan Kaufmann, 2008.
- Information Visualization: Perception for Design, 3rd edition. Ware. Morgan Kaufmann / Academic Press, 2004.