Visualization & Visual Data Science:

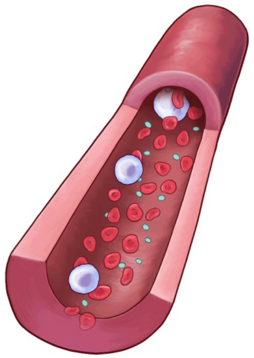
A Primer

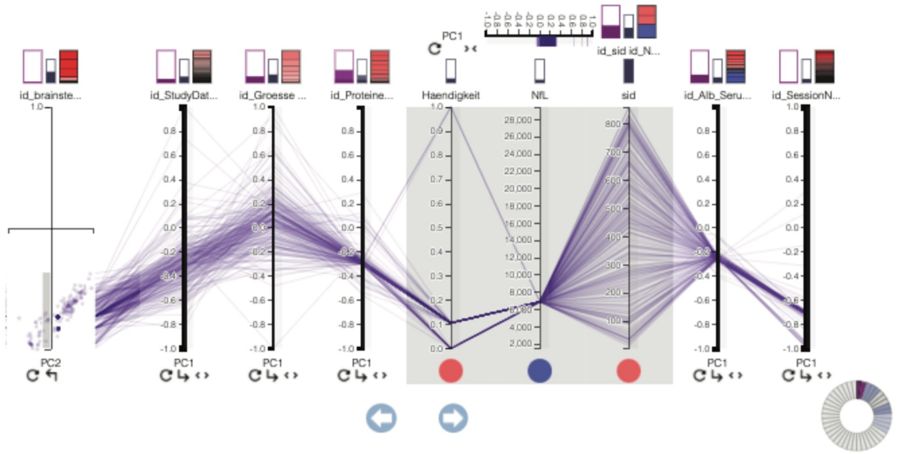
Laura Garrison, University of Bergen
laura.garrison@uib.no

ICTP Workshop 2022
2. December 2022

A bit about me…

Source: The Oatmeal

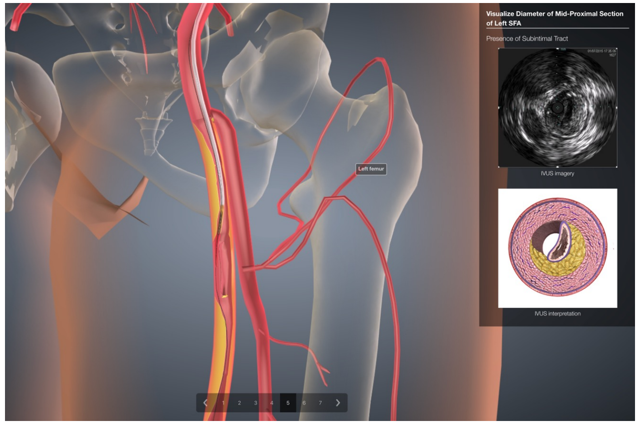




MS Biomedical
Visualization

PhD Visualization,
University of Bergen (UiB)

BA
Biology/Physiology,
Art, & Chemistry

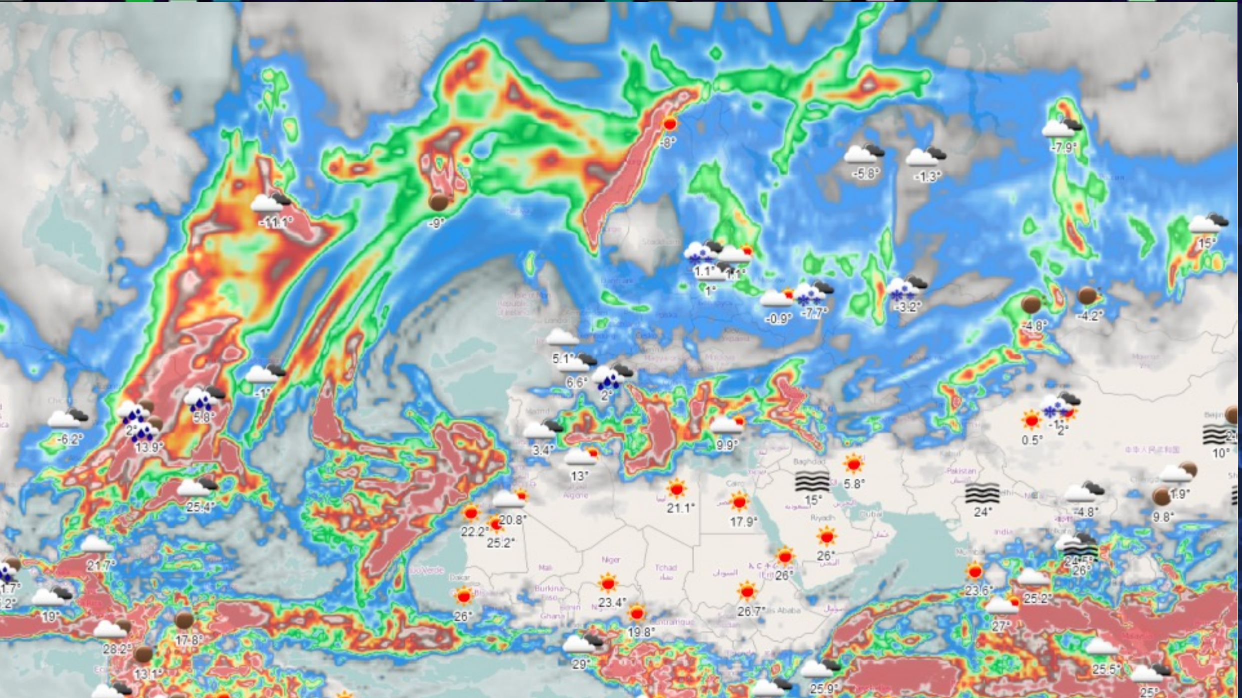


Source: BioDigital

Research fellow,
UiB/MMIV
Consultent, Bouvet

Medical & Health
Tech Start-ups






Data Science

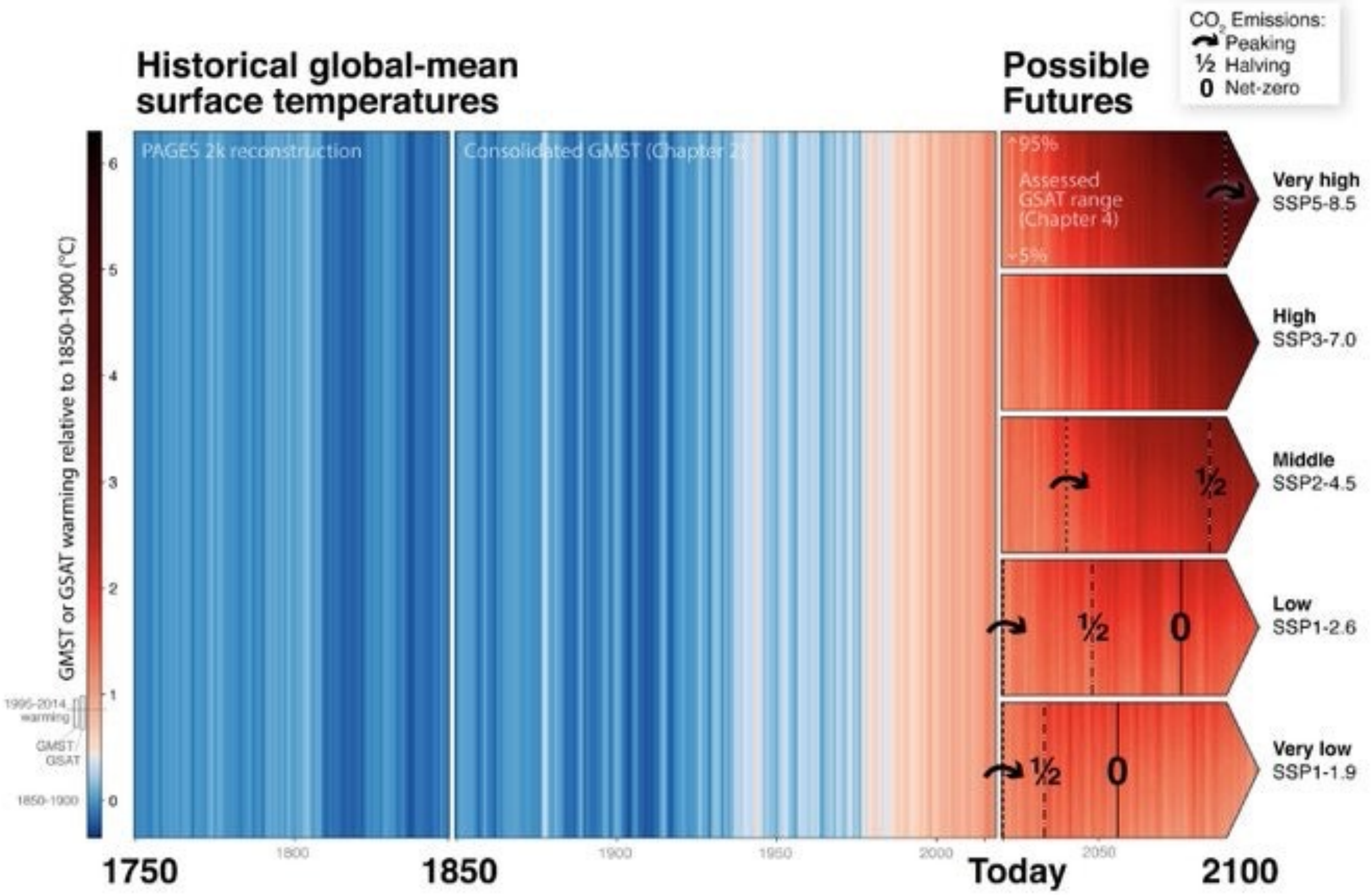
- Extract knowledge and insights from data, often using advanced analytics methods from:
  - Mathematics
  - Statistics
  - Algorithms
  - Machine learning

Src: Kiranshastry - Flaticon

Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, *56*(12), 64-73.

Visualization

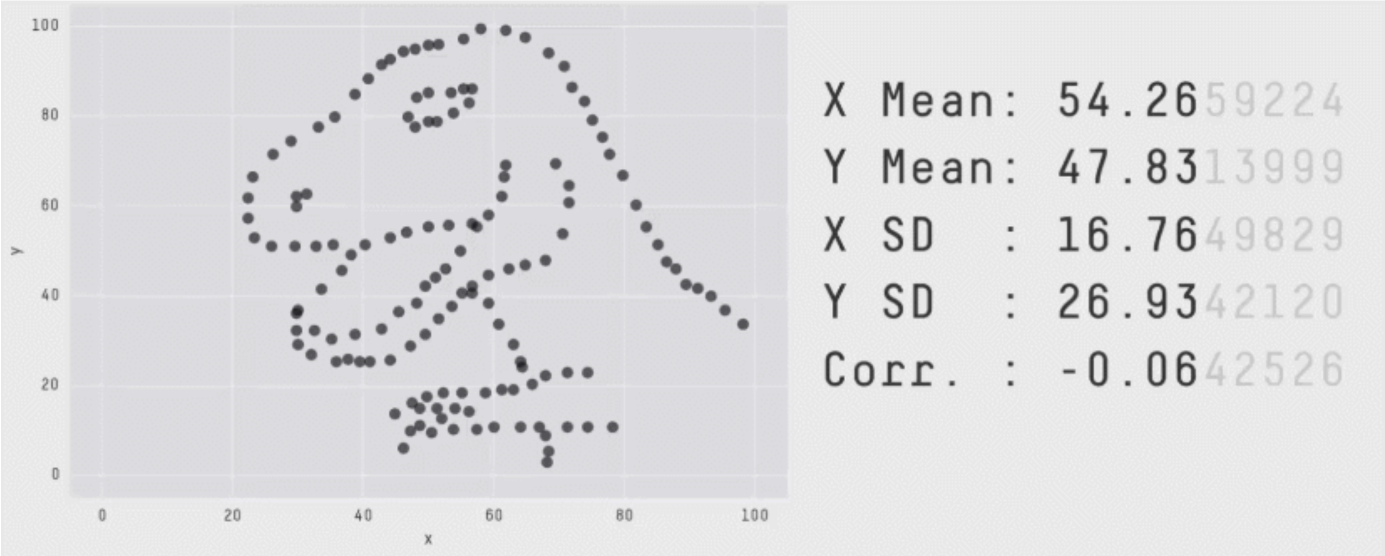
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

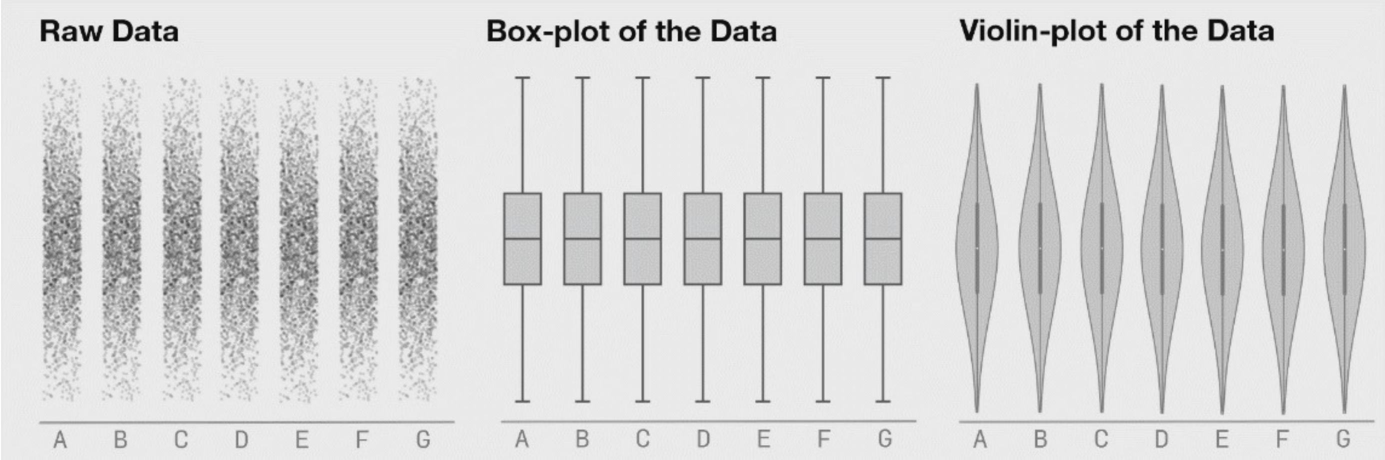


Source: IPCC

Why visualize?

- "visual representation"
  - replace cognition with perception
- "representations of datasets"
  - details matter, summaries can lose information

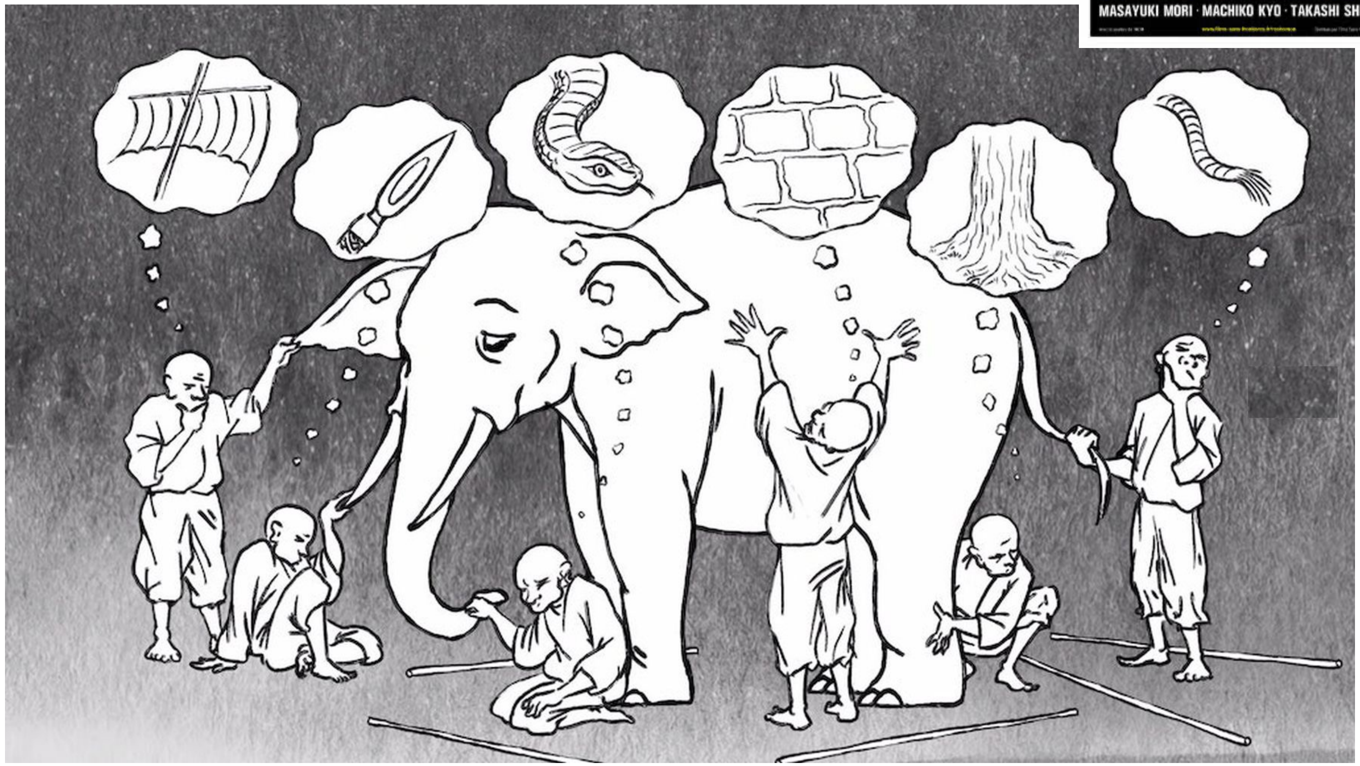




Matejka, J., & Fitzmaurice, G. (2017). Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing. In Proceedings of the 2017 CHI conference on human factors in computing systems (pp. 1290-1294).

Why visualize?

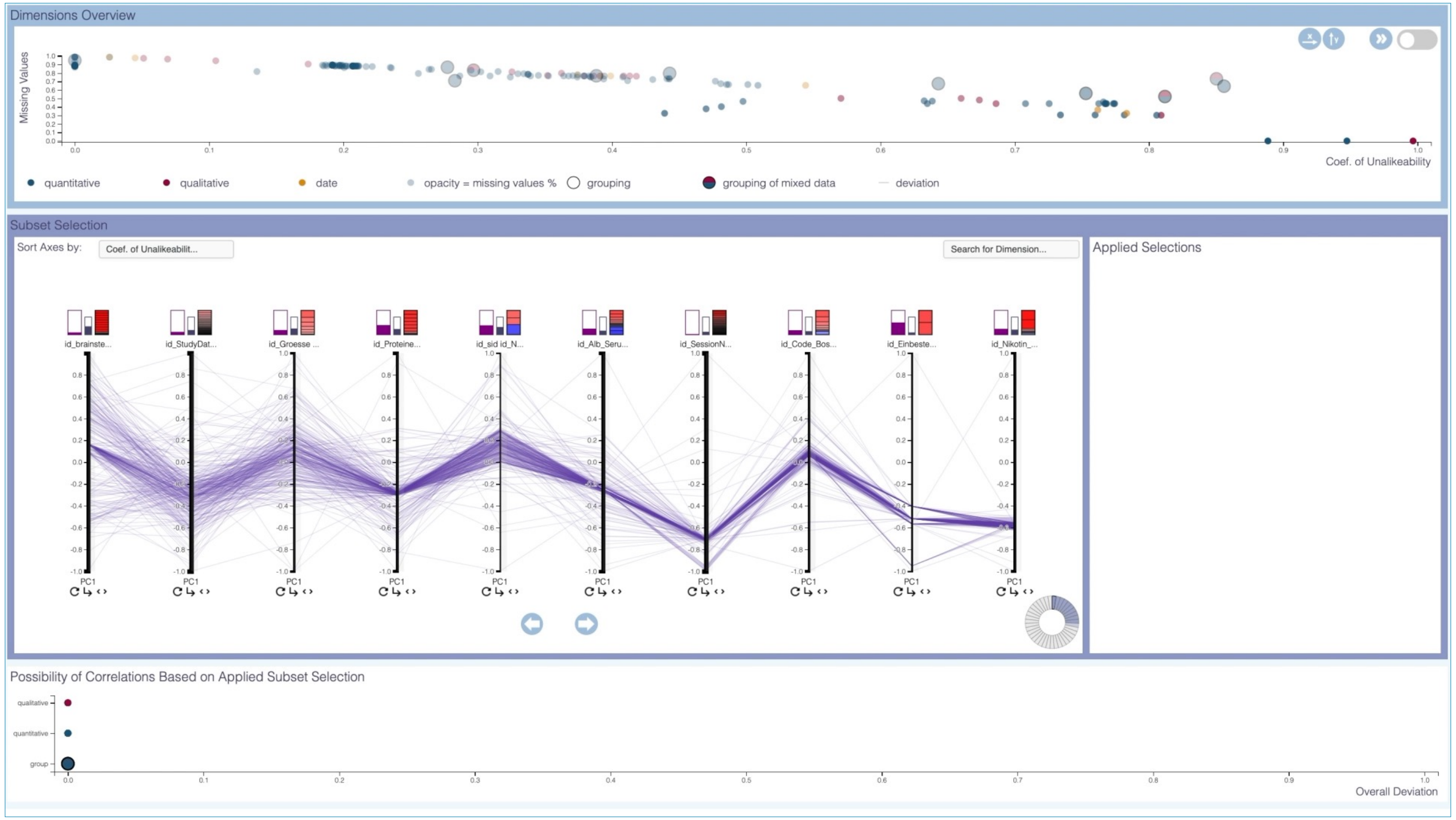
- Rashomon Effect
    - Different models, parameters, representations, etc. can tell different stories
- Visualization can help us spot and understand reasons for these differences



Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science, 16(3), 199-231.

Why user in the loop?

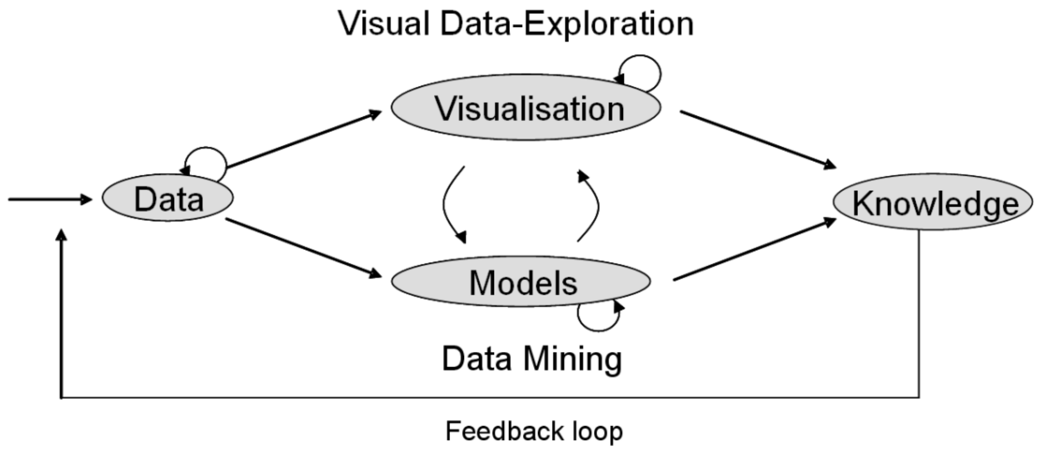
- Visualization is not necessary if there is a trustworthy automatic solution
- **Augment** human capabilities, not replace
- Many analysis problems are **ill-specified**



Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 1
Garrison, L., Müller, J., Schreiber, S., Oeltze-Jafra, S., Hauser, H., & Bruckner, S. (2021). Dimlift: Interactive hierarchical data exploration through dimensional bundling. IEEE Transactions on Visualization and Computer Graphics, 27(6), 2908-2922.

Why user in the loop?

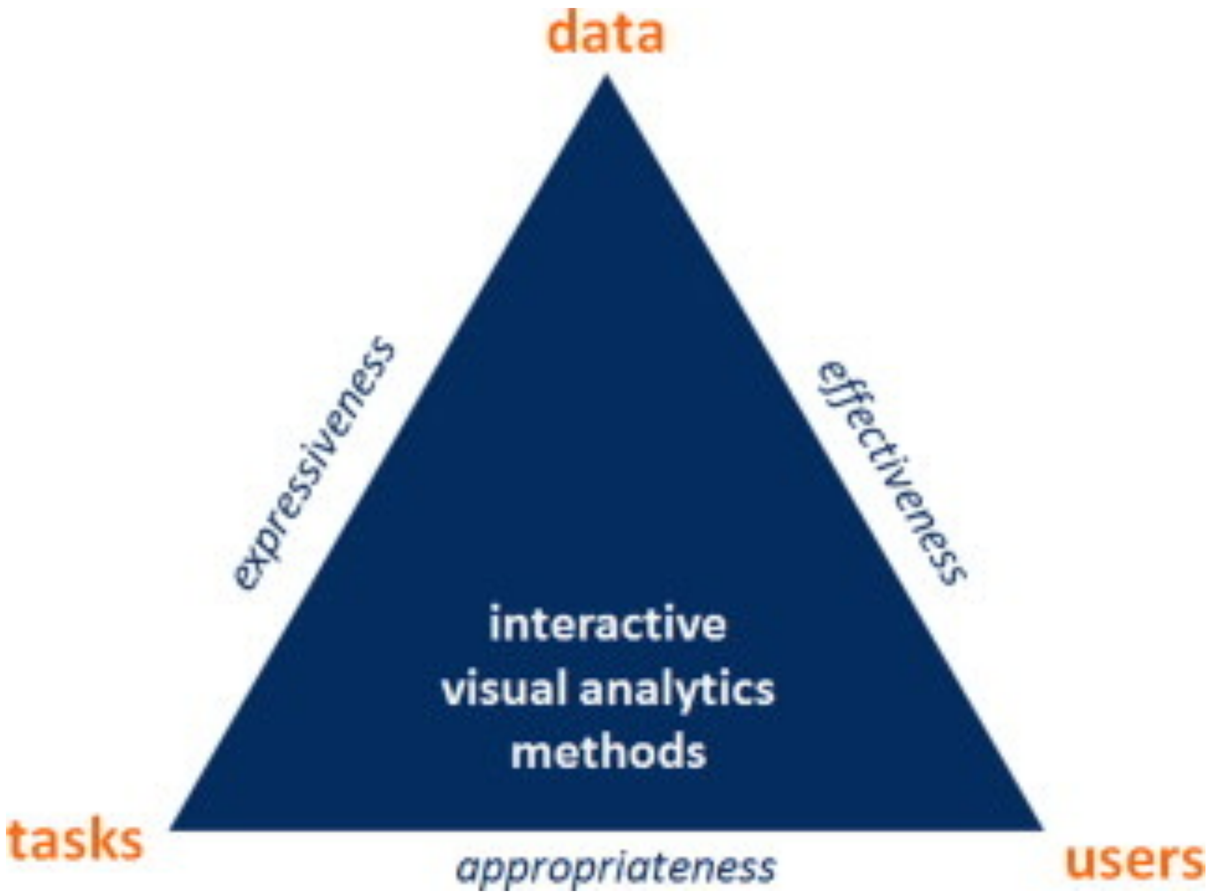
- Visualization integral **throughout** analysis pipeline to help steer a path of inquiry
- Help answer and form new questions:
  - *What do my data look like?*
  - *What are the requirments for developing a more complex model?*
  - *What if I remove/adjust this parameter?*
  - *Do I trust/can I verify the data/model?*



Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In Information visualization (pp. 154-175). Springer, Berlin, Heidelberg.

Visualization in YOUR workflow

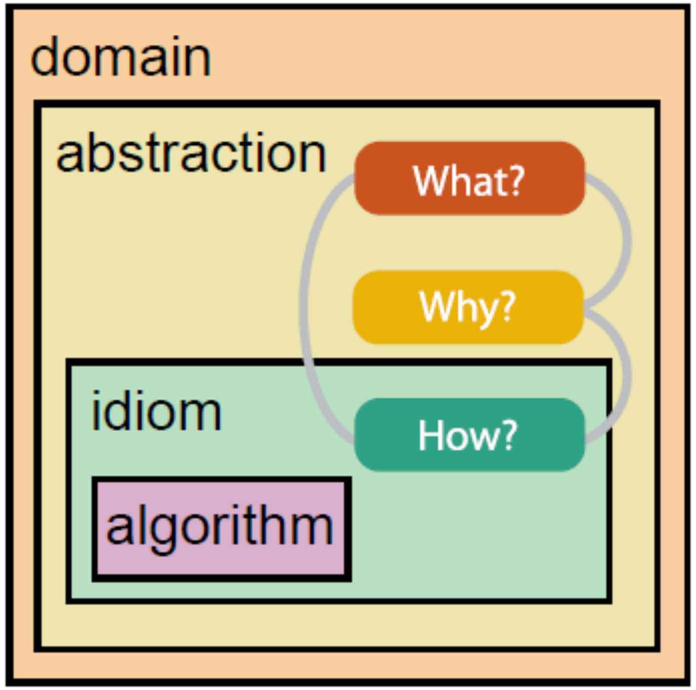
- Good visualization is possible for anyone
- Consider:
  - Data
  - Users
  - Tasks



Miksch, Silvia, and Wolfgang Aigner. "A matter of time: Applying a data–users–tasks design triangle to visual analytics of time-oriented data." Computers & Graphics 38 (2014): 286-290.

Visual Analysis Framework

- Domain
  - Who are you visualizing for? Yourself?
- Abstraction
  - What is shown (data abstraction)
  - Why showing (task abstraction)
- Idiom
  - How is it being shown
    - visual encoding idiom (how do you draw the picture)
    - interaction idiom (how do you manipulate the picture)
- Algorithm
  - efficient computation to show the picture



What is shown

[data abstraction]

Data

- 21.11.2022, 11, 58, 2665
    - What does/could this sequence of numbers mean?

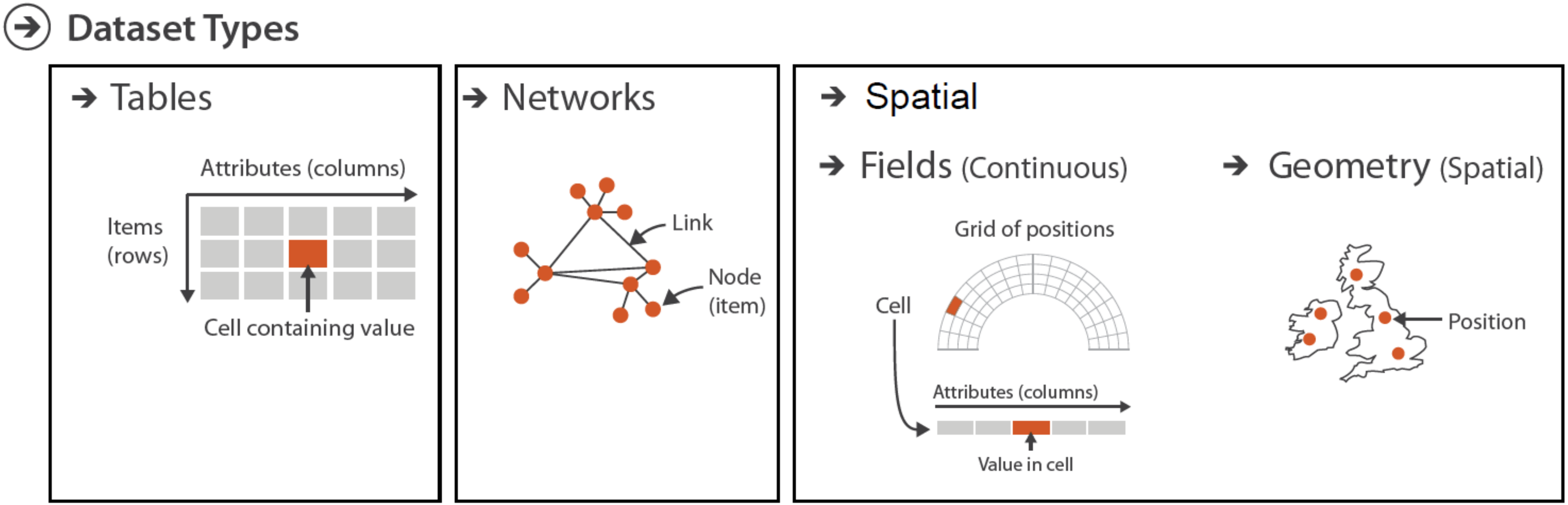
Data

- 21.11.2022, 11, 58, 2665
  - What does/could this sequence of numbers mean?
  - Date, stair flight count, average heart rate, step count

Data Interpretation

- Semantics
  - Real world meaning of the data
- Type
  - Structural or mathematical interpretation
    - item, link, attribute, position (if in a grid)
  - Different from data types in programming languages!
  - Think about operations that are meaningful for each type
    - e.g. 22.11/01.11 makes sense if this is a decimal, but not if this is a date

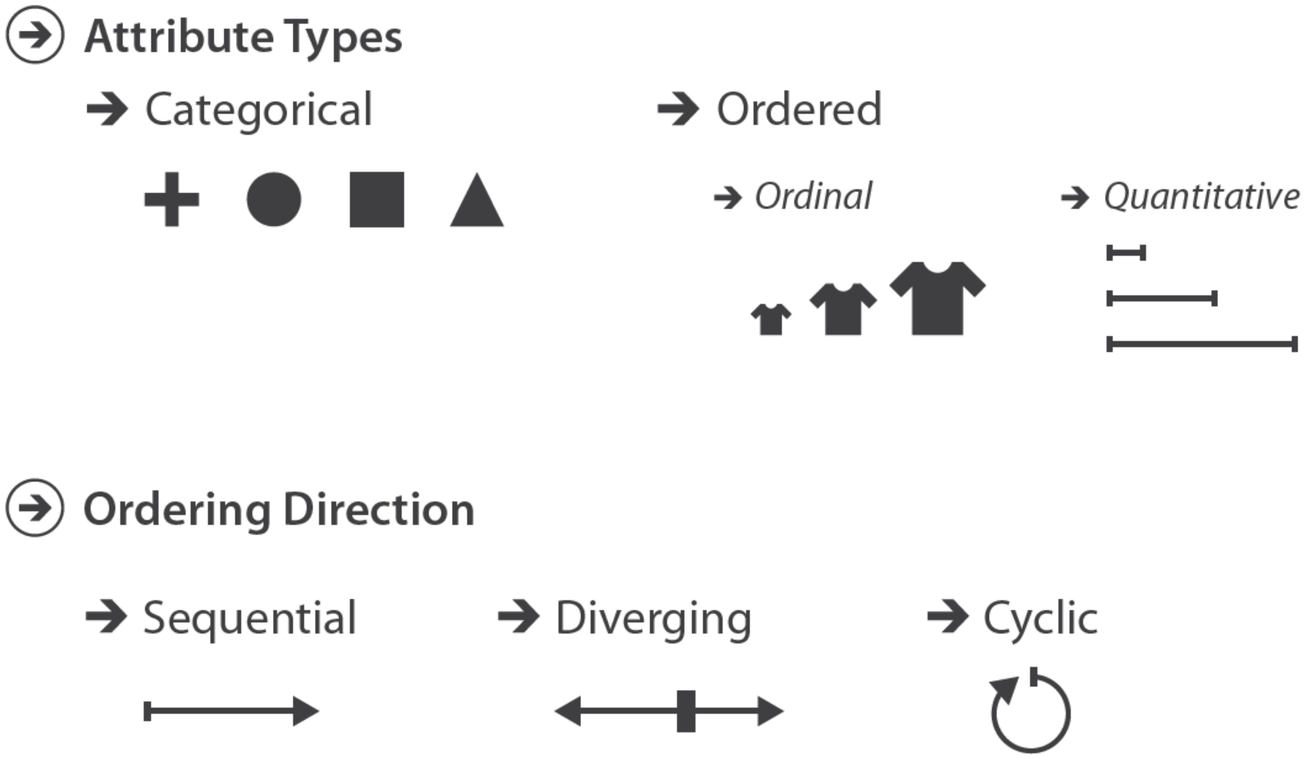
Major Data Types



Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 2.

Attribute Types and Ordering

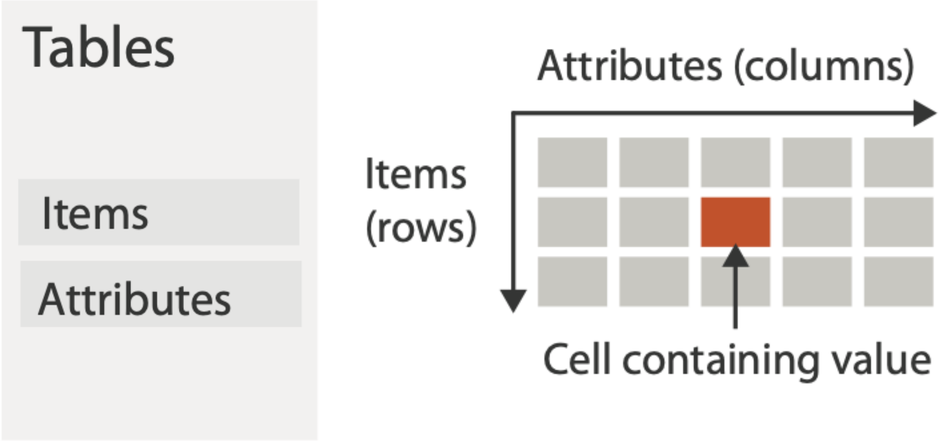
- Item
    - Discrete, individual entity
        - e.g., patient
    - “independent variable”
- Attribute
    - Property that is measured, observed
        - e.g., height, BMI
    - “dependent variable”



Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 2.

Table

- Flat table
  - Row = item
  - Column = attribute
  - Cell = item-attribute pair
  - Unique key (possibly implicit)

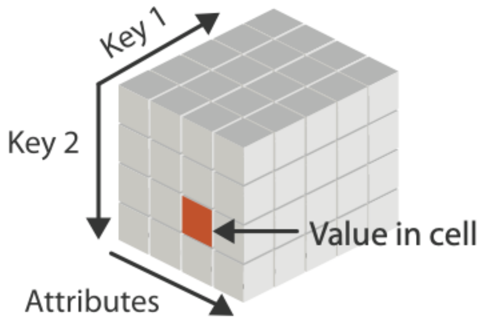


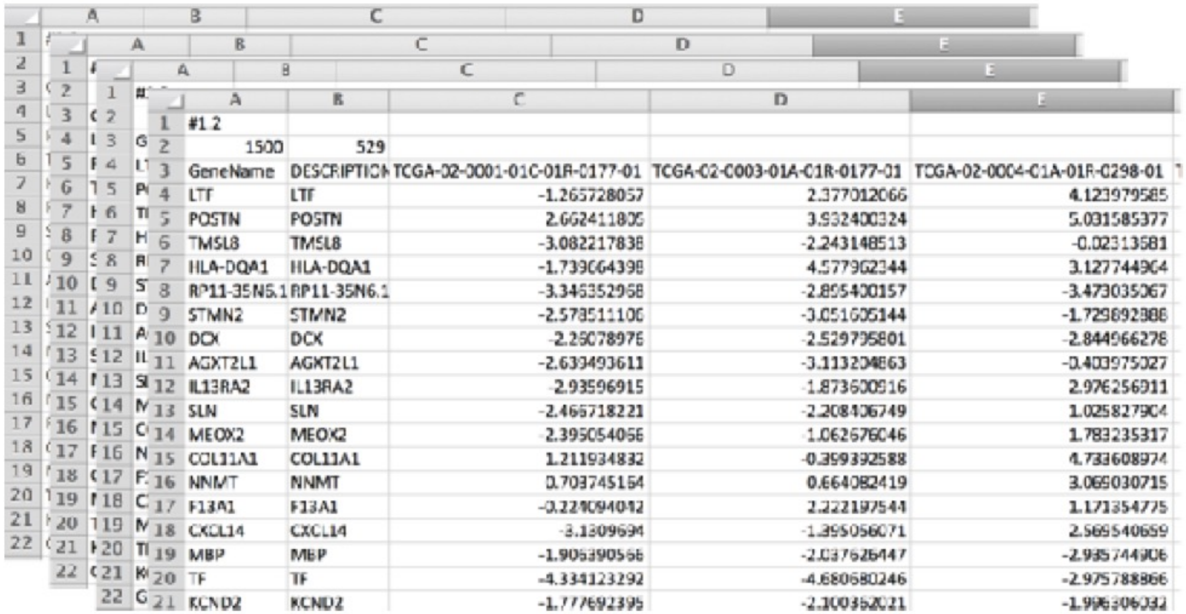
A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6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

Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 2.

Multidimensional Table

- Index according to multiple keys





	A	B	C	D	E
1	#1.2				
2	1500	529			
3	GeneName	DESCRIPTION	TCGA-02-0001-01C-01R-0177-01	TCGA-02-0003-01A-01R-0177-01	TCGA-02-0004-01A-01R-0298-01
4	LTF	LTF	-1.265728057	2.377012066	4.123979585
5	POSTN	POSTN	2.662411805	3.932400324	5.031585377
6	TMSL8	TMSL8	-3.082217838	-2.243148513	-0.02313581
7	HLA-DQA1	HLA-DQA1	-1.739064398	4.577962344	3.127744964
8	RP11-35N6.1	RP11-35N6.1	-3.346352958	-2.895400157	-3.473035067
9	STMN2	STMN2	-2.578511106	-3.051605144	-1.729892886
10	DCX	DCX	-2.25078976	-2.529795801	-2.844966278
11	AGXT2L1	AGXT2L1	-2.639493611	-3.113204863	-0.403975027
12	IL13RA2	IL13RA2	-2.93596915	-1.873600916	2.976256911
13	SLN	SLN	-2.466718221	-2.208406749	1.025827904
14	MEOX2	MEOX2	-2.395054056	-1.062676046	1.783235317
15	COL11A1	COL11A1	1.211934832	-0.399392588	4.733608974
16	NNMT	NNMT	0.703745164	0.664082419	3.069030715
17	F13A1	F13A1	-0.224094042	2.222197544	1.171354775
18	CXCL14	CXCL14	-3.1309694	-1.395056071	2.569540659
19	MBP	MBP	-1.906390566	-2.037626447	-2.985744906
20	TF	TF	-4.334123292	-4.680680246	-2.975788866
21	KCND2	KCND2	-1.777692395	-2.100352021	-1.996306032

Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 2.

Network

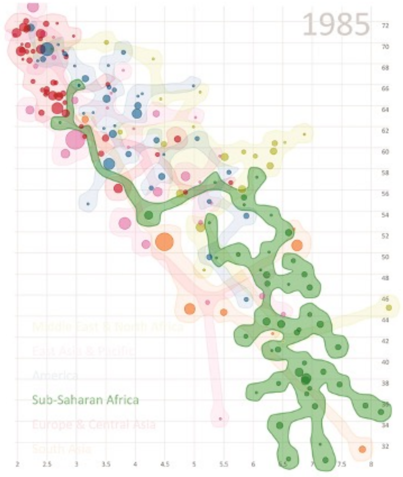
- Nodes (vertices) connected by links
- Trees are special case of acyclic network
  - usually directed
  - usually w/roots

Networks & Trees

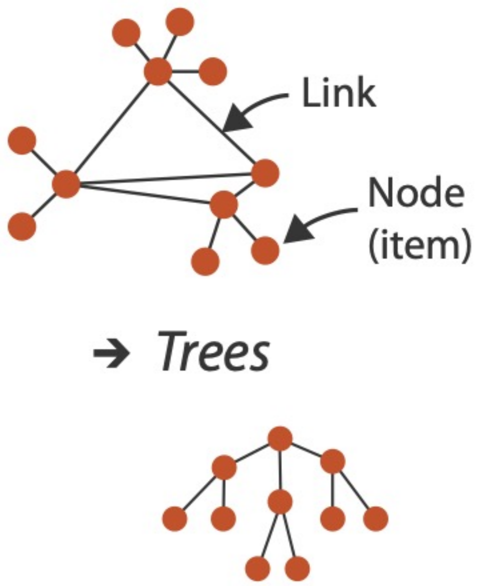
Items (nodes)

Links

Attributes



vialab.science.uoit.ca/portfolio/bubblesets

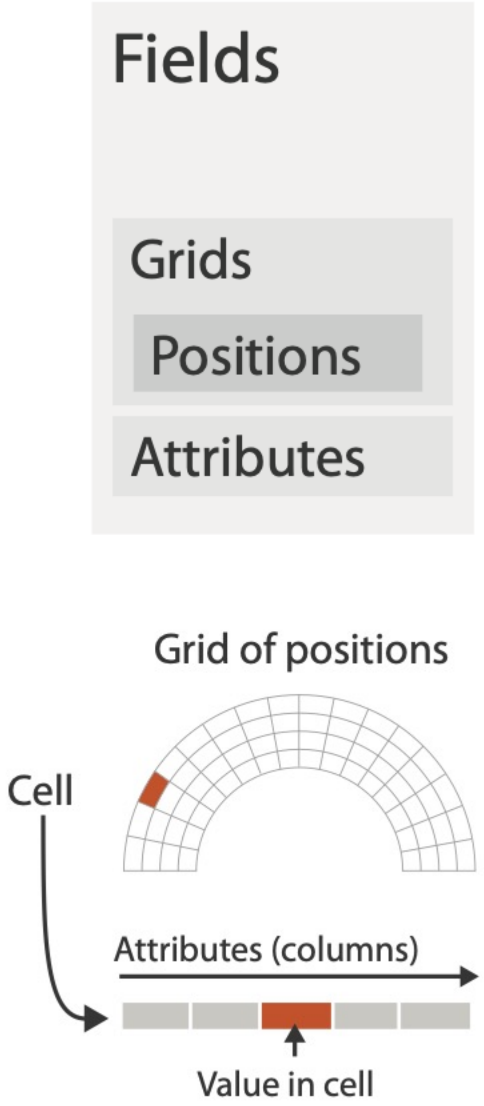


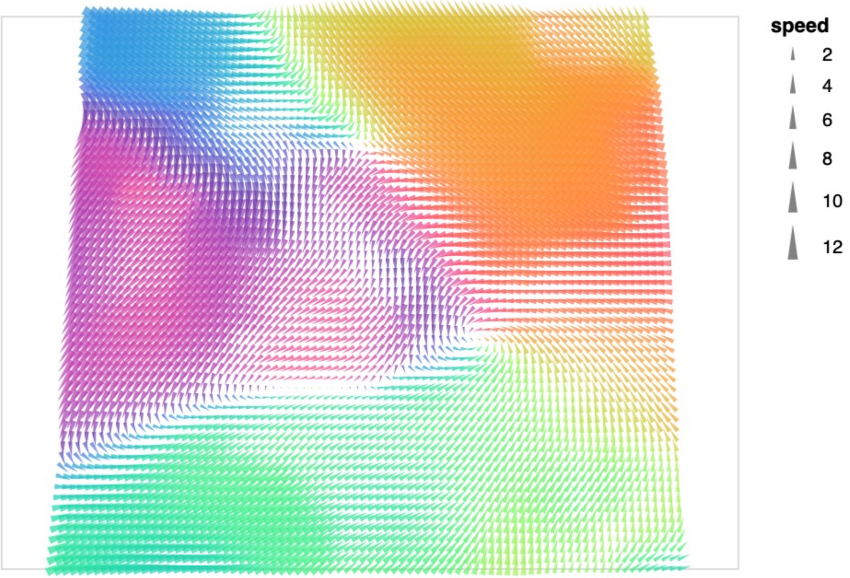
https://observablehq.com/@d3/force-directed-graph

Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 2.

Spatial Field

- Attributes associate with cells
- Grid subdivides continuous domain into cells
  - e.g. temperature
- Measured or simulated data
- Concerns
- Divisions



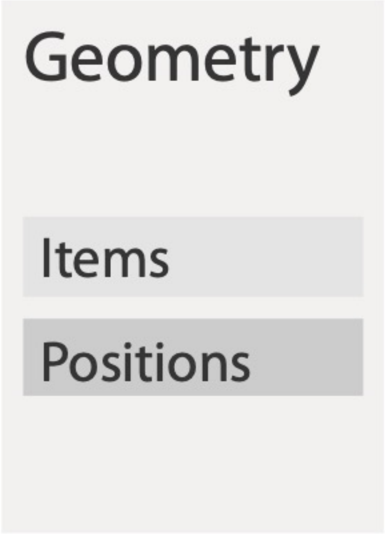


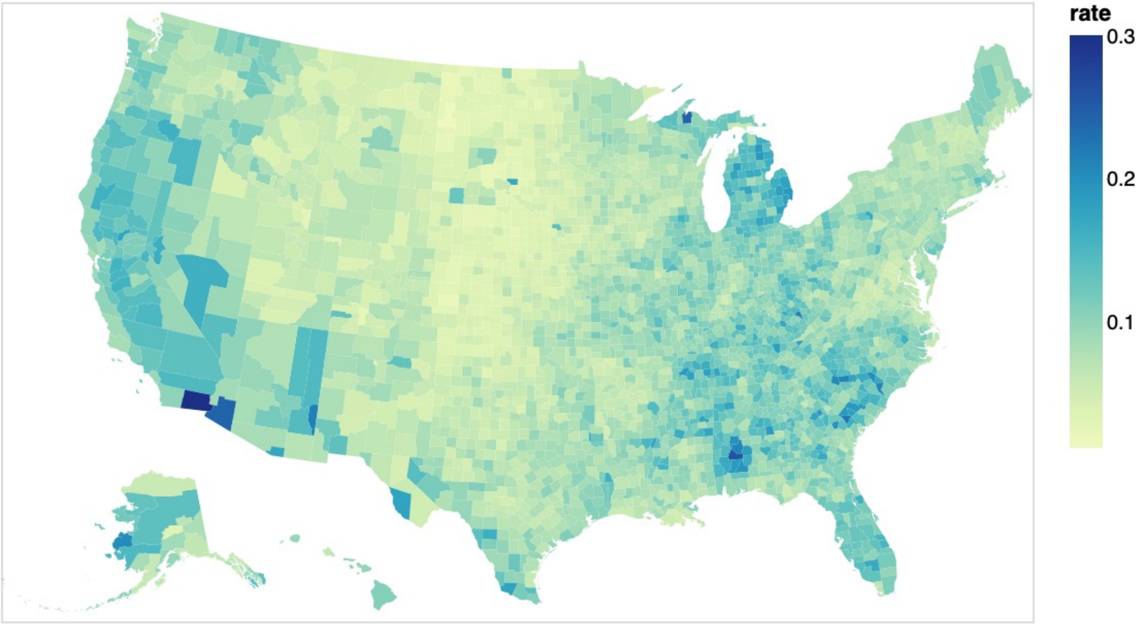
https://altair-viz.github.io/gallery/wind_vector_map.html

Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 2.

Geometry

- Explicit position
- Shape of items
  - Point
  - 1D line
  - 2D shape
  - 3D volume



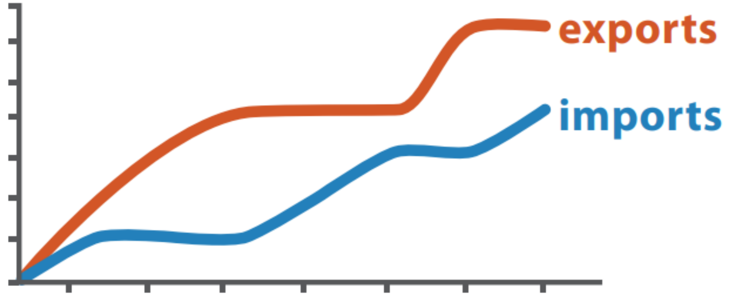


https://altair-viz.github.io/gallery/choropleth.html

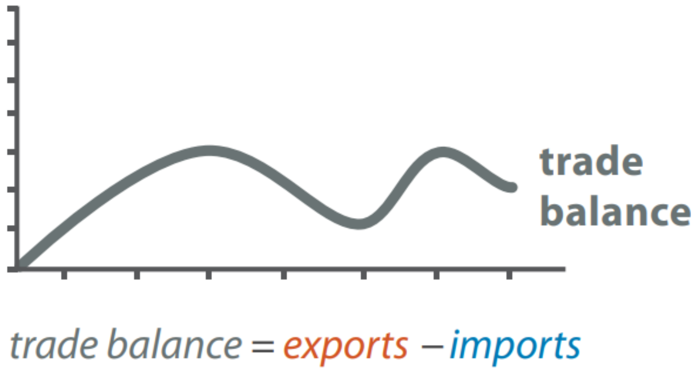
Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 2.

Data Abstraction Steps

1. Identify dataset type(s) and attribute type(s)
2. Identify cardinality
   1. How many items in dataset?
   2. Cardinality of each attribute?
      1. Number of levels/segments for categorical data
      2. Range of continuous data
3. Consider whether to transform data
   1. Determine if necessary according to user task

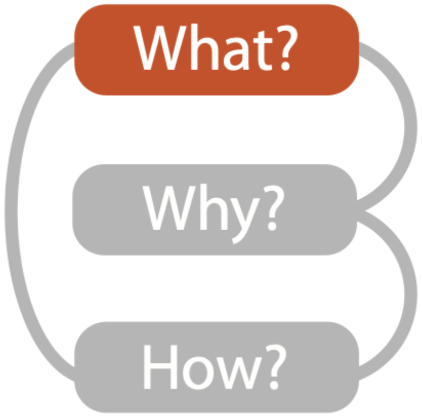


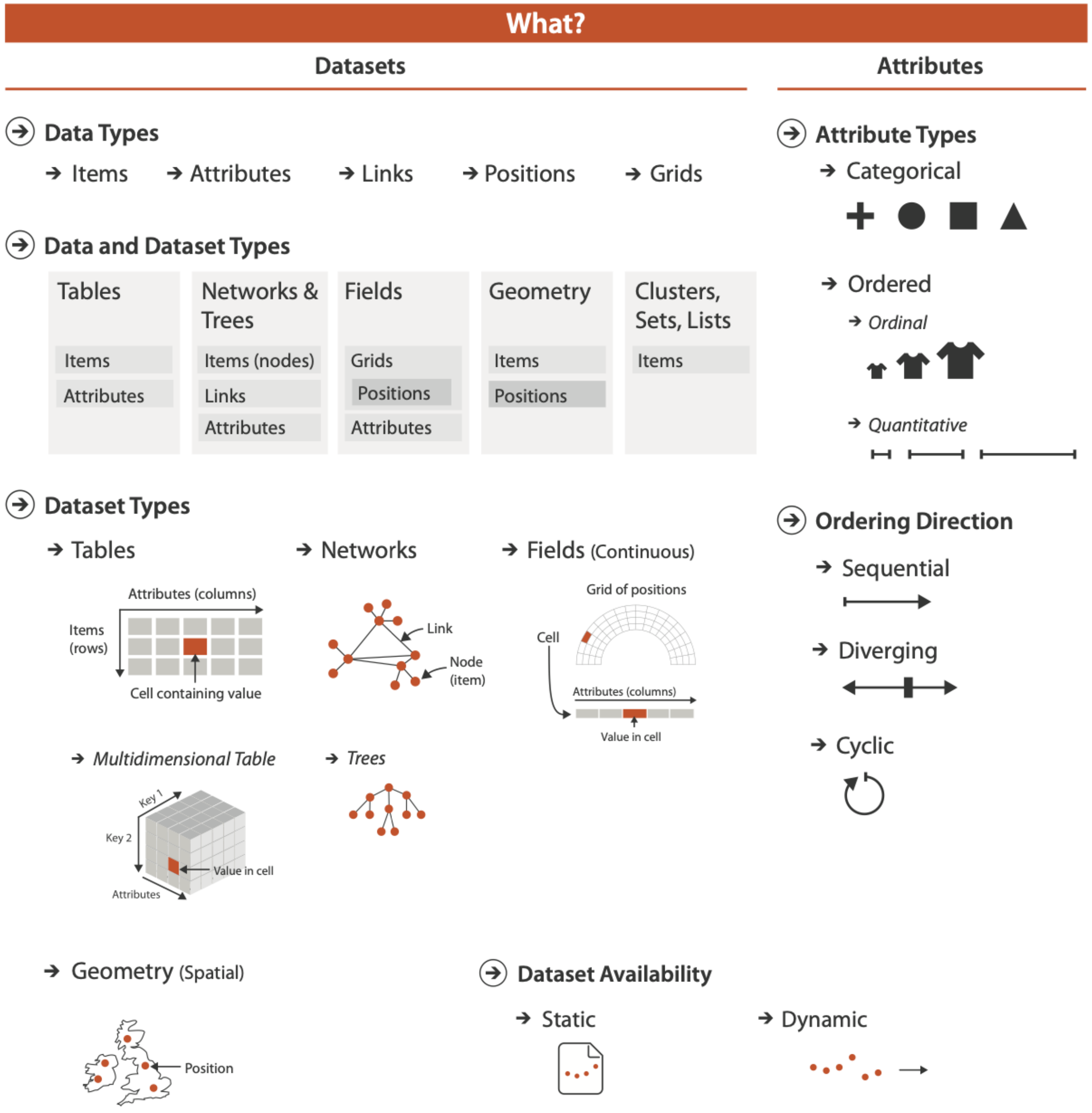
Original Data



Derived Data

Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 2.





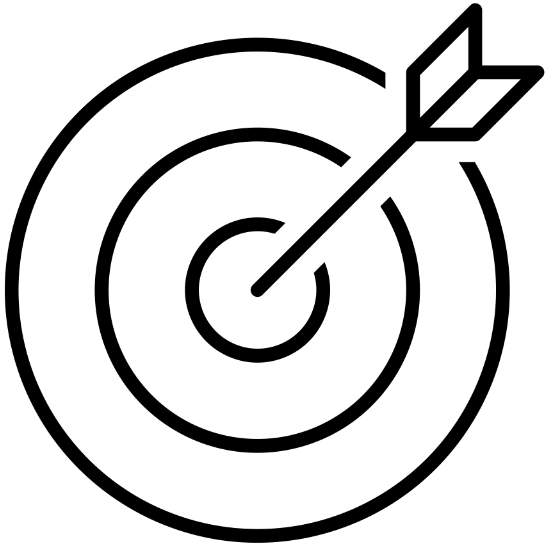
Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 2.

Why showing?

[task abstraction]

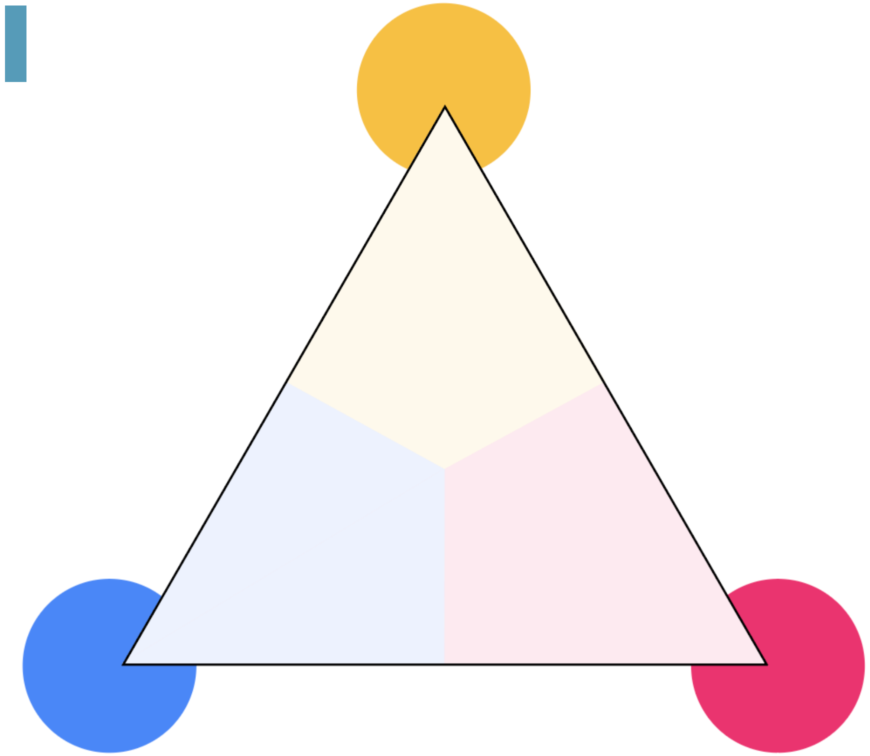
{action, target}

- Action: Analyze, search, query
- Target: What is being acted on
- {action, target}”
  - Discover distribution
  - Compare trends
  - Browse outliers
  - Explore topology
  - ...



Actions: High Level Choices

EXPLORATION
Browse data
Generate hypotheses



COMMUNICATION
Present
Enjoy
Produce

ANALYSIS
Verify
Search: lookup, locate
Query: identify, compare, summarize

Brehmer, M., & Munzner, T. (2013). A multi-level typology of abstract visualization tasks. *IEEE transactions on visualization and computer graphics*, *19*(12), 2376-2385.

Actions: High Level Choices

Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 3.

Actions: Search

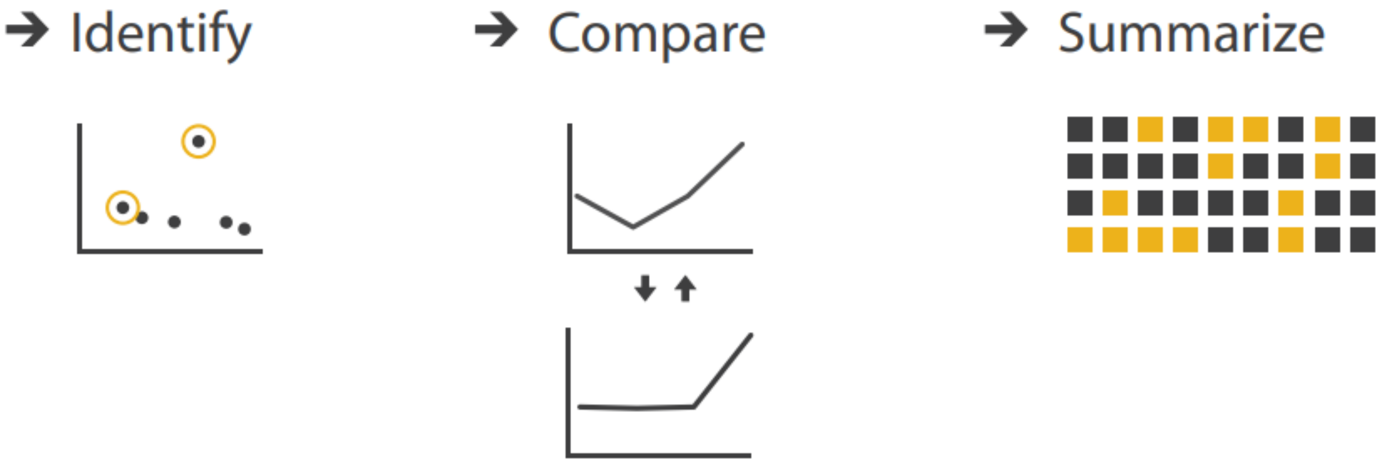
- What does the user know?
  - Target
  - Location

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 3.

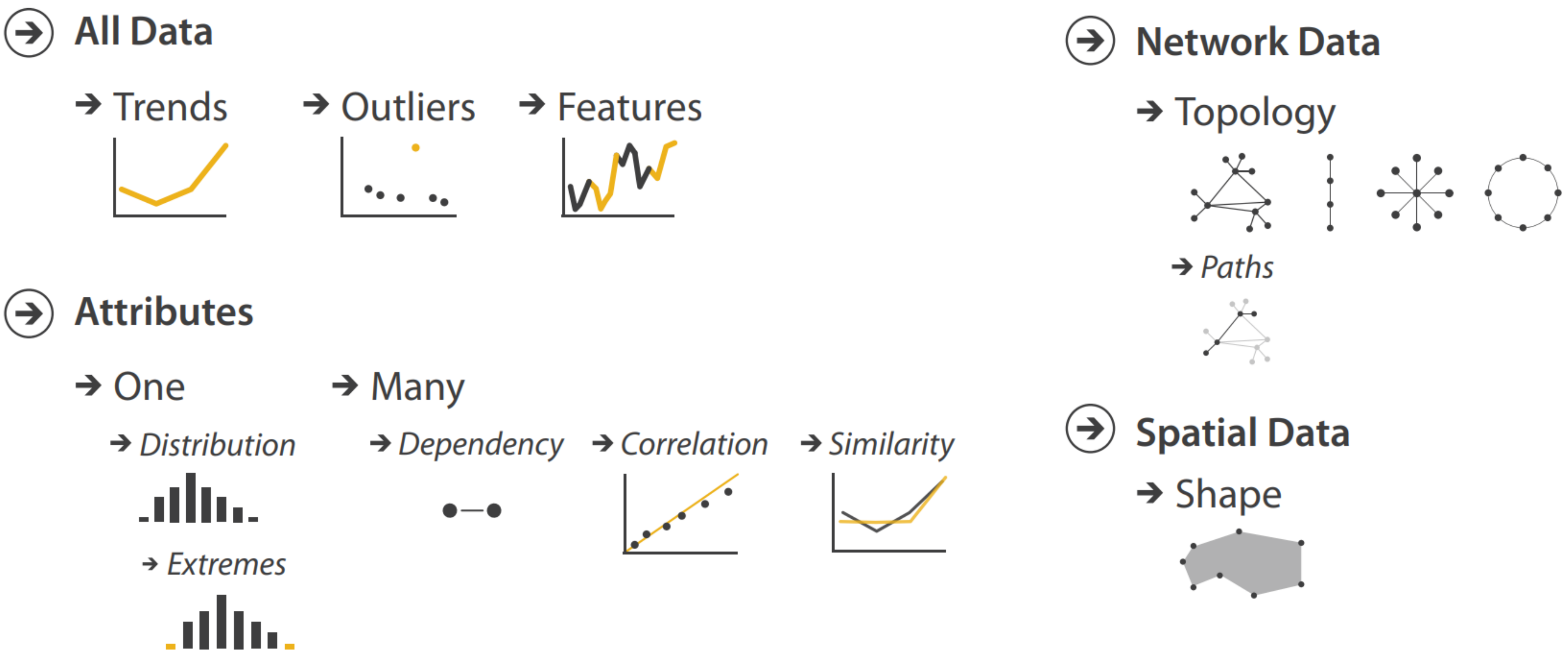
Actions: Query

- How much of the data matters?
    - One: identify
    - Some: compare
    - All: summarize



Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 3.

Targets



All Data
Trends
Outliers
Features
Attributes
One
Distribution
Extremes
Many
Dependency
Correlation
Similarity
Network Data
Topology
Paths
Spatial Data
Shape

Data <-> Task Abstraction

- Data abstraction required within task abstraction
  - Specify targets
  - May lead to data transformation (if task requires)
    - e.g., species diversity index
- Iterate between data and task abstraction
  - Not just a one-way street

Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 3.

What?

Why?

How?

Why?

Actions

Analyze

- Consume
  - *Discover*
  - *Present*
  - *Enjoy*
- Produce
  - *Annotate*
  - *Record*
  - *Derive*

tag

Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Query

- Identify
- Compare
- Summarize

Targets

All Data

- Trends
- Outliers
- Features

Attributes

- One
  - *Distribution*
  - *Extremes*
- Many
  - *Dependency*
  - *Correlation*
  - *Similarity*

Network Data

- Topology
  - *Paths*

Spatial Data

- Shape

How showing?

[visual abstraction]

“A picture is worth a thousand words…if you know how to read it.”

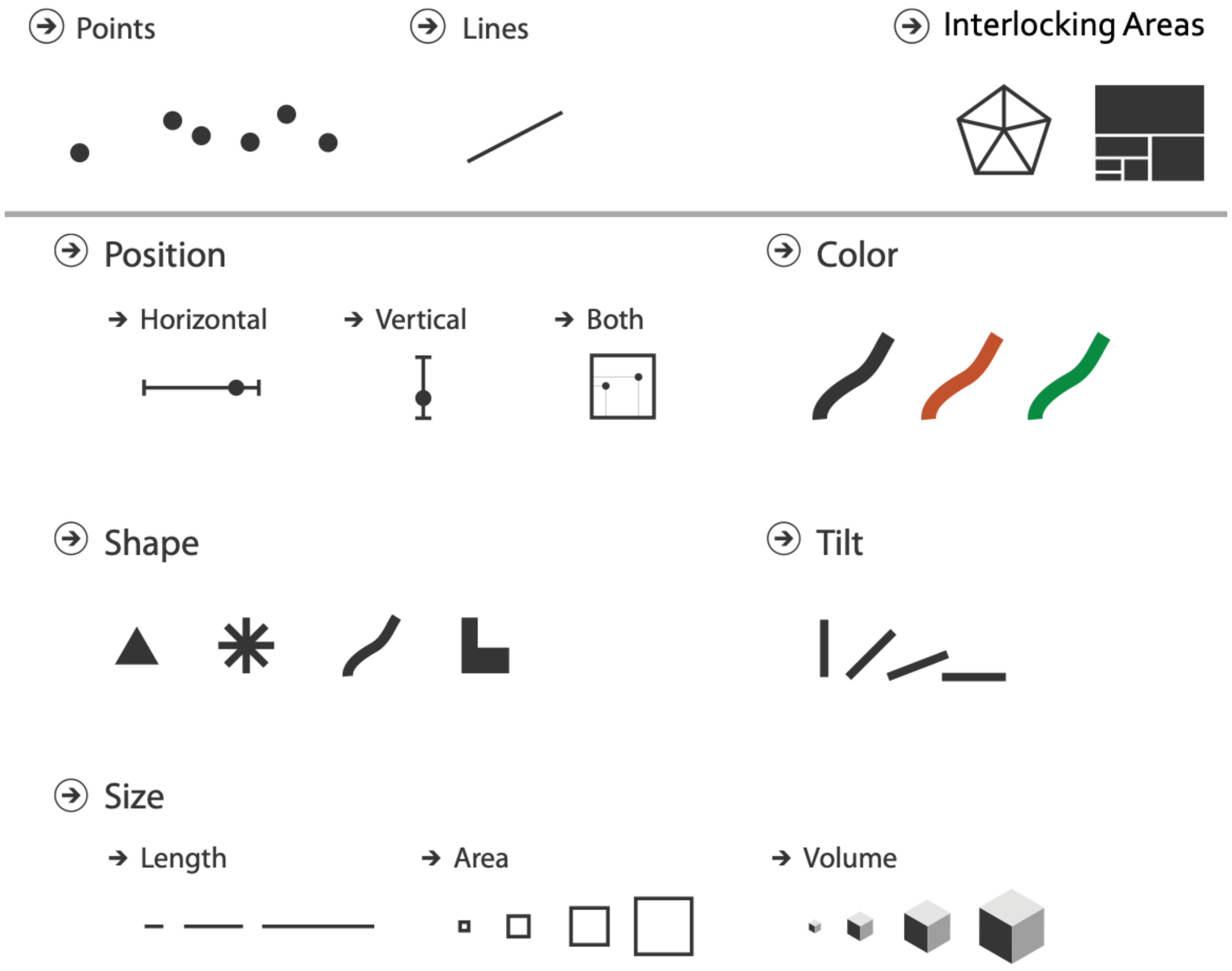
\- Alberto Cairo

Visual Encoding

Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 5.

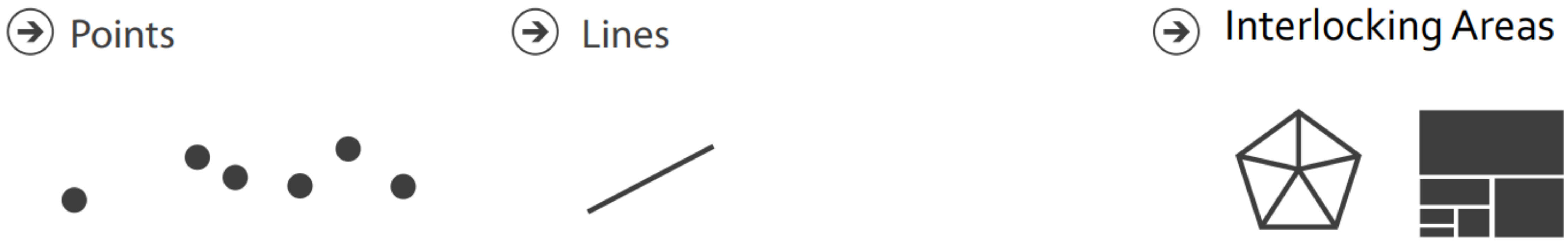
Marks and Channels

- Marks
  - Represent items or links
- Channels
  - Change the appearance of marks based on attributes

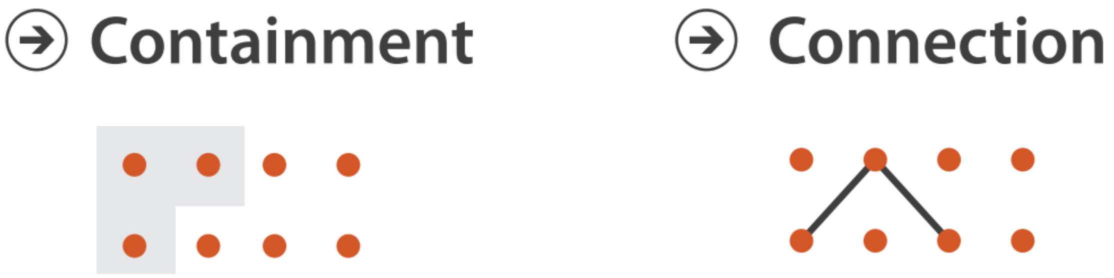


Marks for Items

- 0D, 1D, 2D (3D is rarely used)

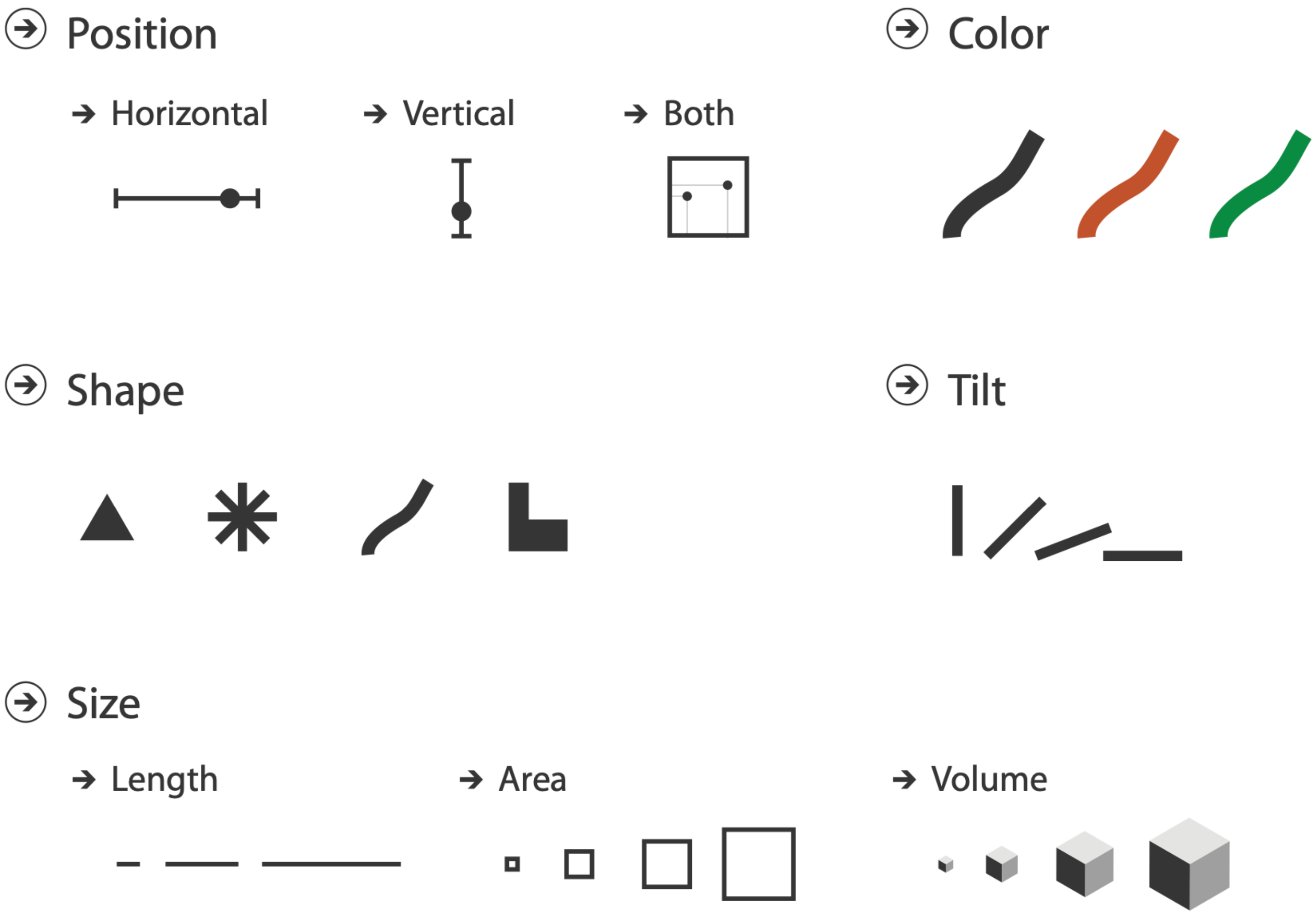


- Can also indicate links



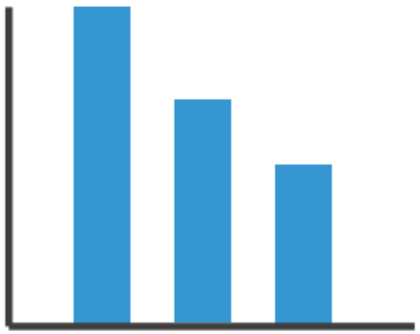
Channels

- Control appearance of marks
- Channel properties differ
  - Type and amount of information that can be conveyed to human perceptual system



Analyze idiom structures

Charts can be thought of combinations of marks and channels



Mark: line

Channel:
vertical position

Mark: point

Channel:
vertical position
horizontal position

Mark:
point

Channel:
vertical position
horizontal position
color (hue)

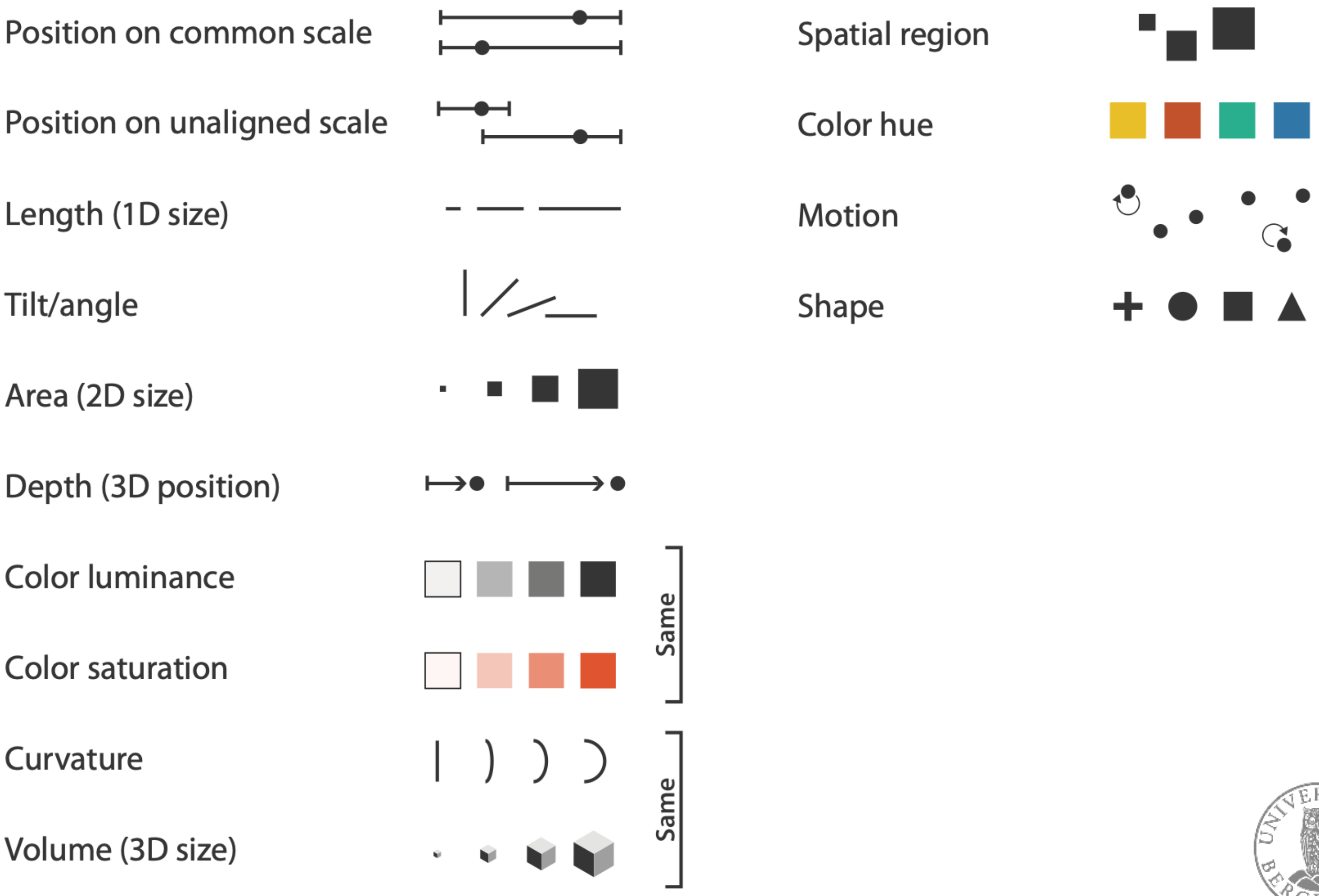
Mark:
point

Channel:
vertical position
horizontal position
color (hue)
size (area)

Match types

Expressiveness

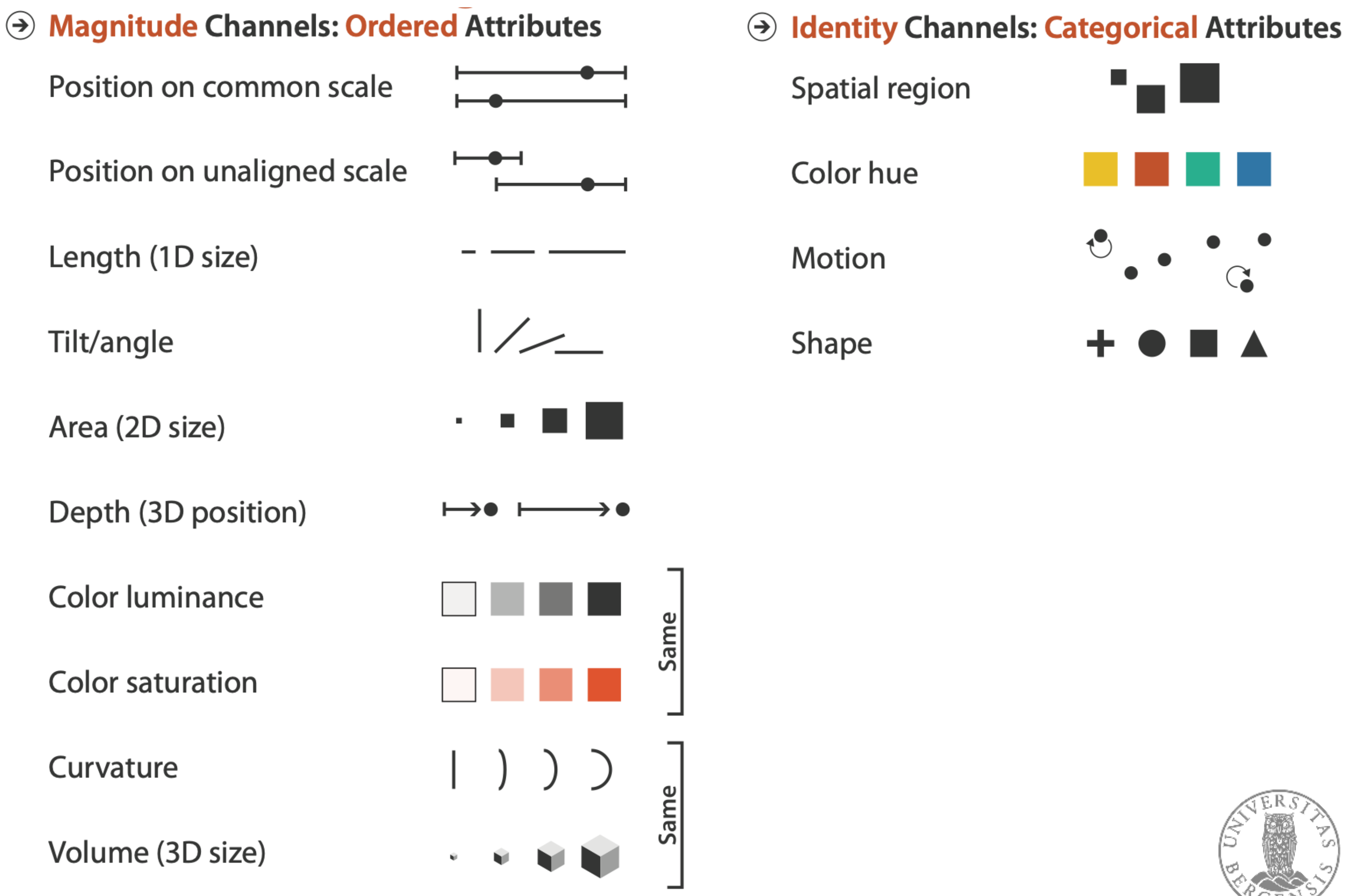
- match channel type to data type



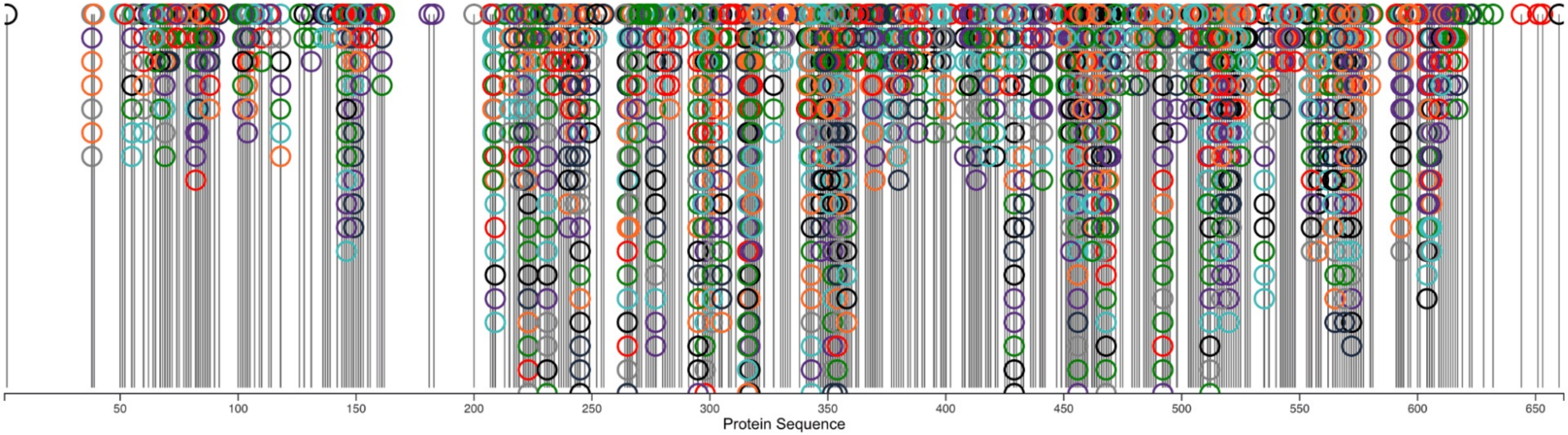
What is “best?”

Effectiveness

- some channels ARE better than others (perceptually)
- spatial position ranks high for both



Exercise



50
100
150
200
250
300
350
400
450
500
550
600
650
Protein Sequence

Exercise

➔ Points ➔ Lines ➔ Interlocking Areas

➔ **Magnitude Channels: Ordered Attributes**

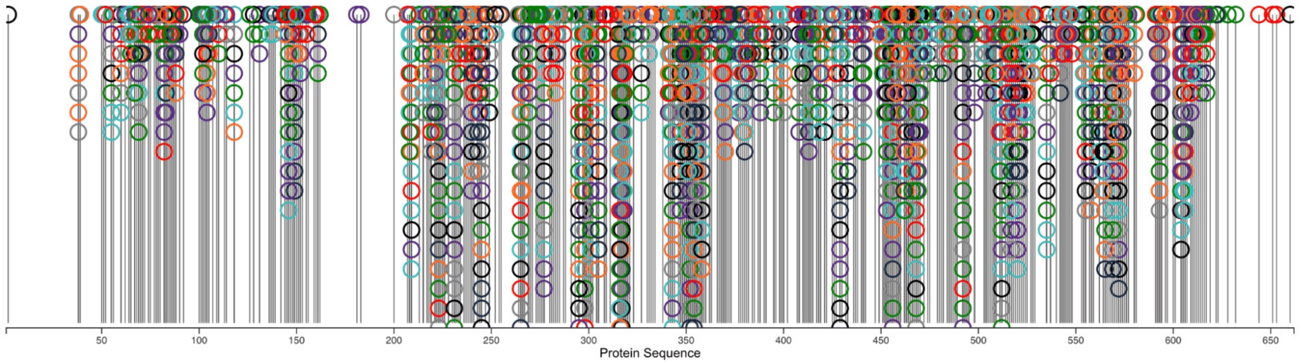
- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

Same

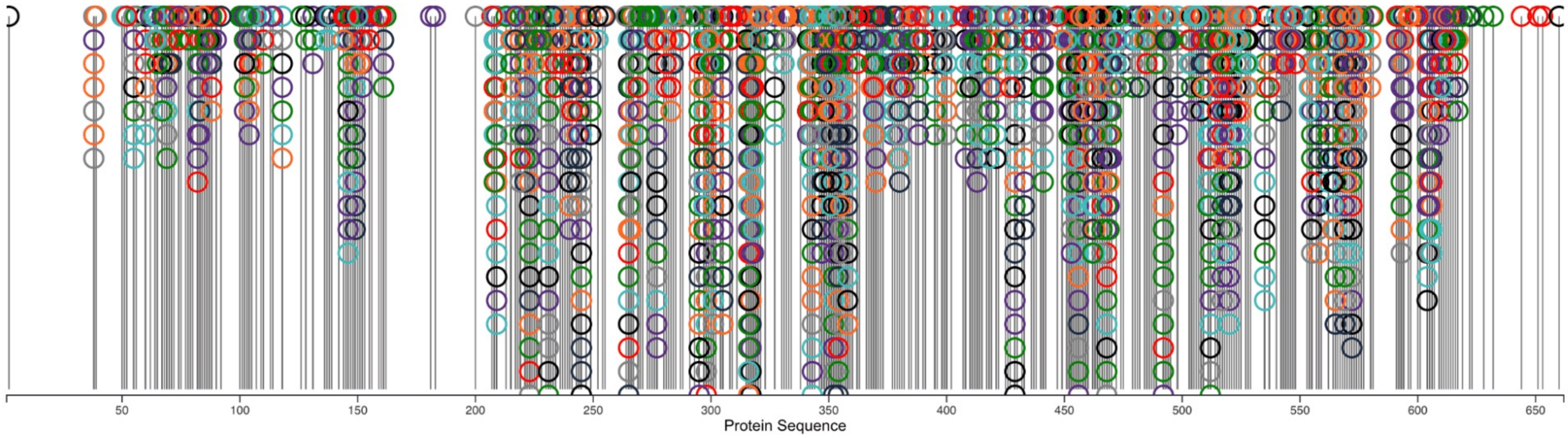
Same

➔ **Identity Channels: Categorical Attributes**

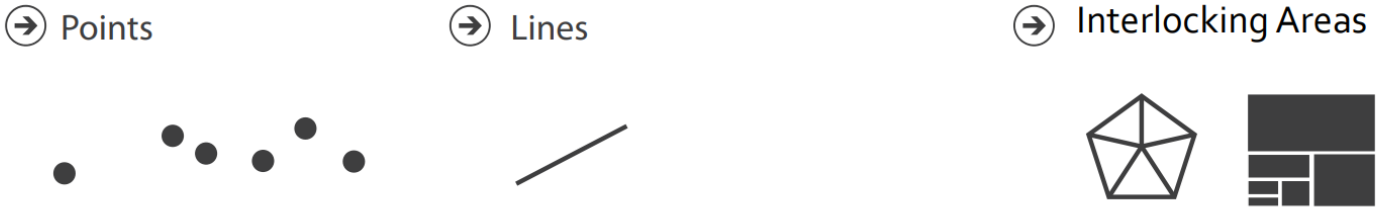
- Spatial region
- Color hue
- Motion
- Shape

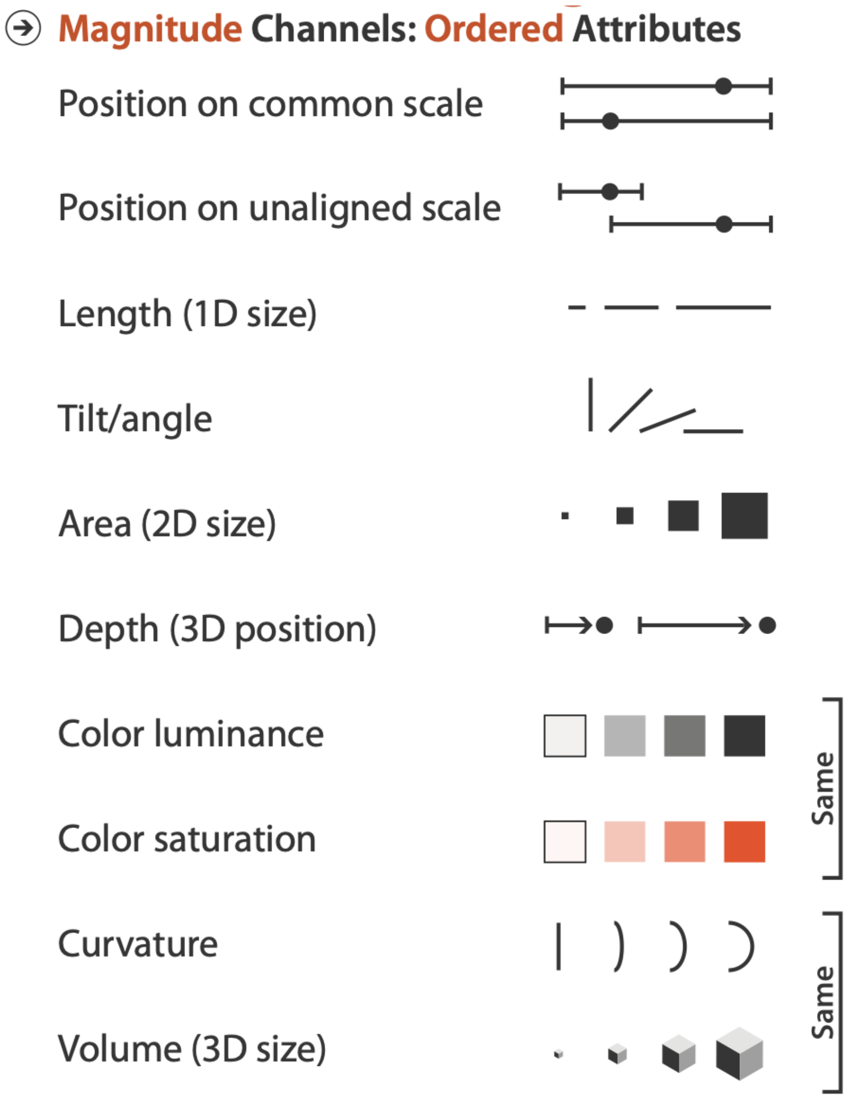


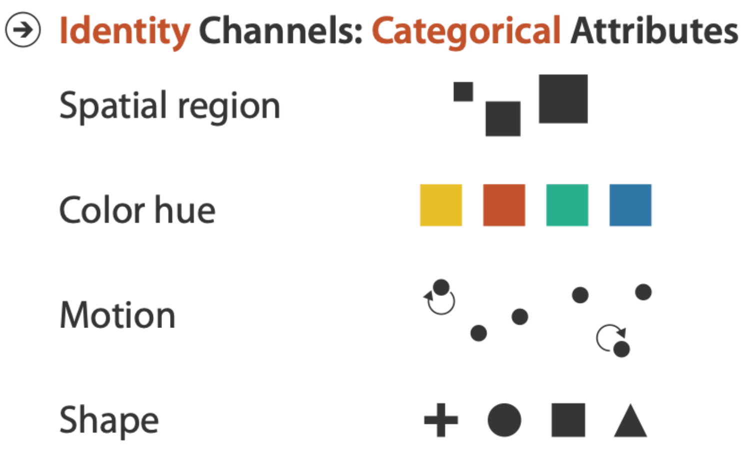
Other possibilities?

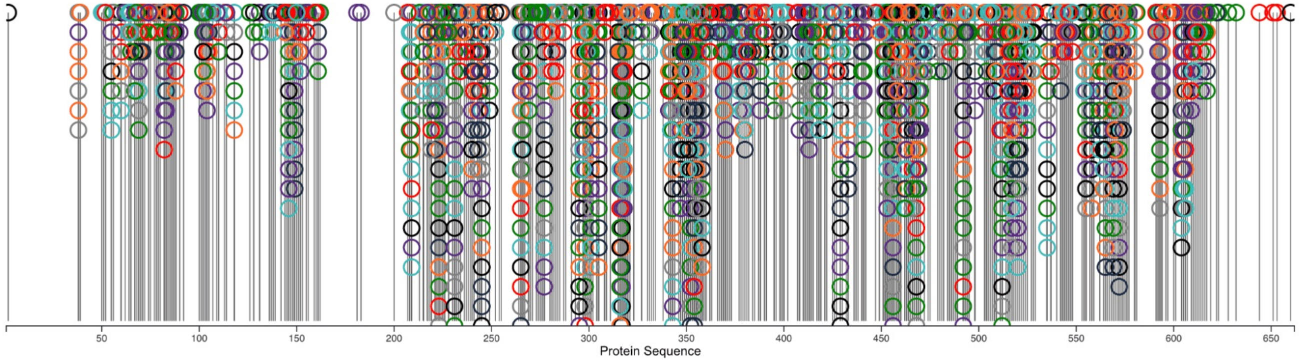


Other possibilities?









What?

Why?

How?

Encode

Arrange

- Express

- Separate

- Order

- Align

- Use

Map

from categorical and ordered attributes

- Color
  - *Hue*

  - *Saturation*

  - *Luminance*

- Size, Angle, Curvature, ...

- Shape

- Motion

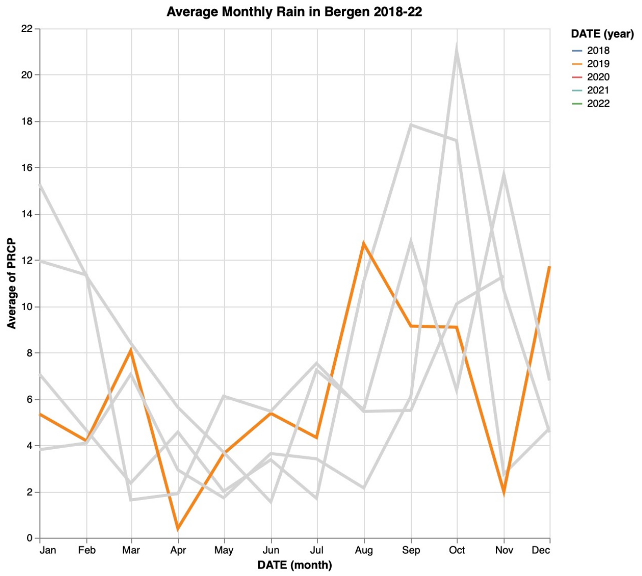
Direction, Rate, Frequency, ...

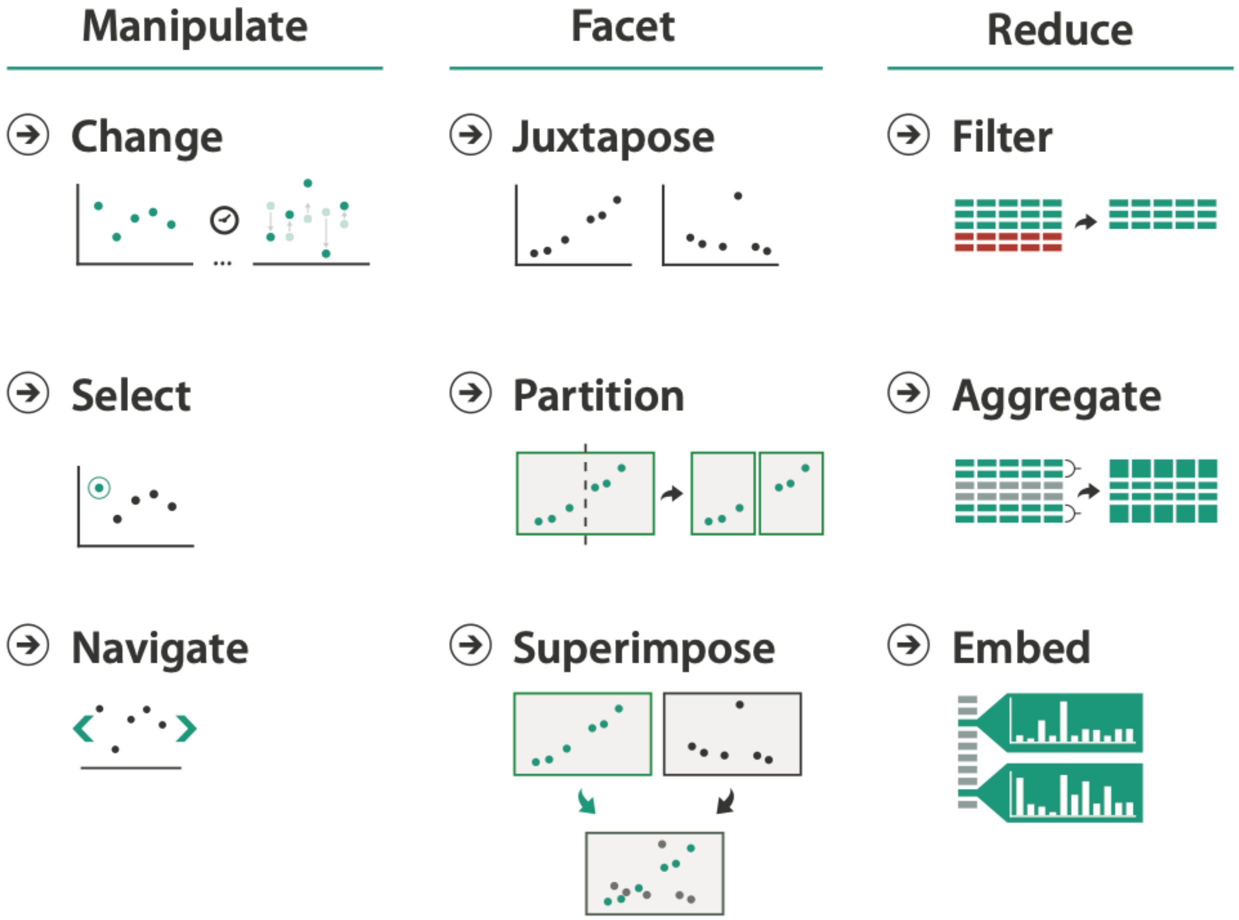
How showing?

[interaction abstraction]

Interaction encodings

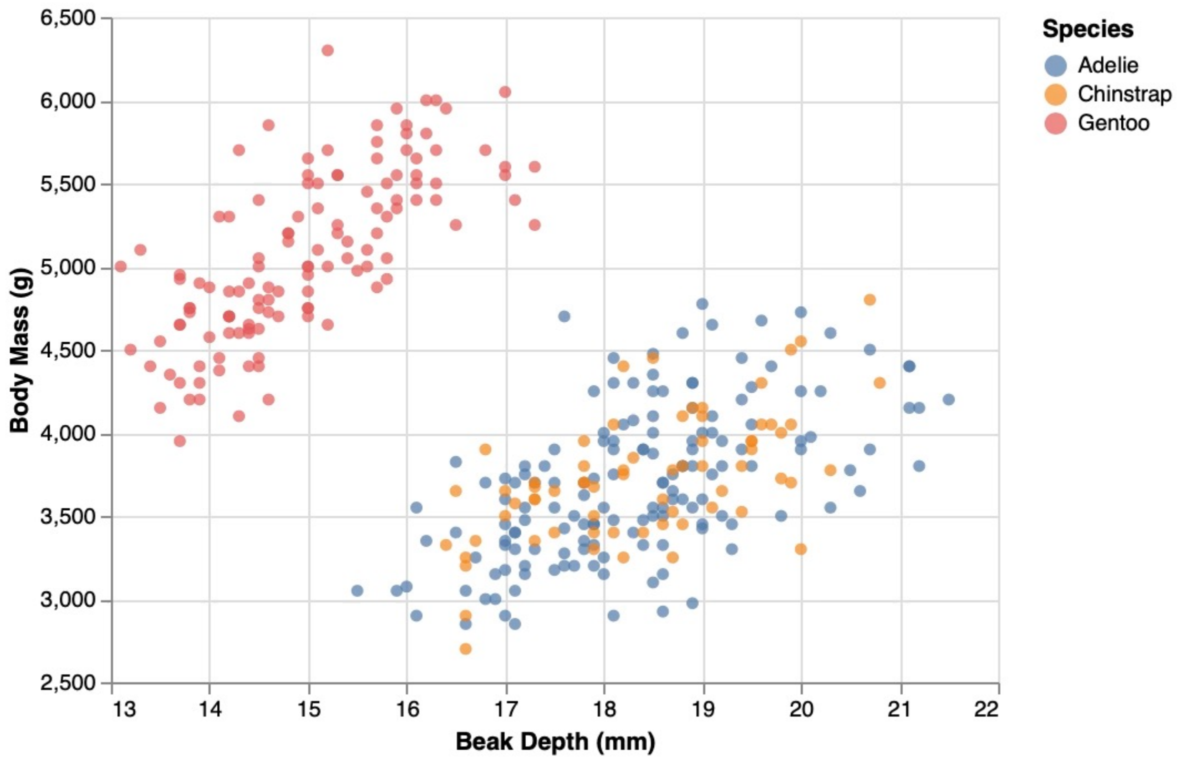
- Efficiency
- Deal with complexity through interactions in addition to directly visually encoding data

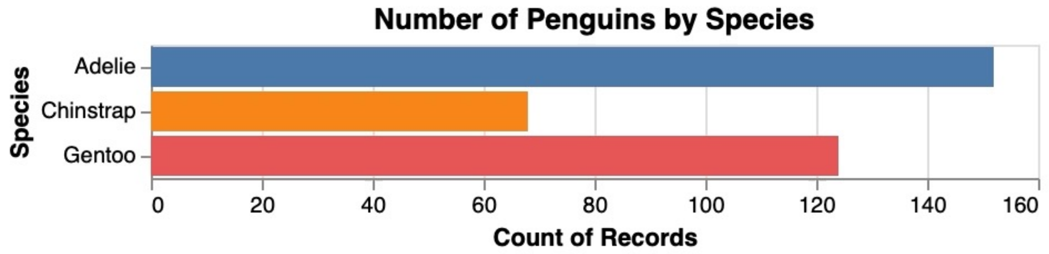




Manipulate

- Change over time
  - e.g., animated transitions
- Select
  - e.g., highlighting
- Navigate
  - e.g., pan within a view, scrollytelling, reduce attributes





day_05/basic_charts/linked_scatter_bar.py

Facet

- Juxtapose
- Partition
  - e.g., split into regions by attributes
- Superimpose
  - e.g., map layers (roads, terrain)

Coordinate Multiple Side By Side Views

→ Share Encoding: Same/Different

→ *Linked Highlighting*

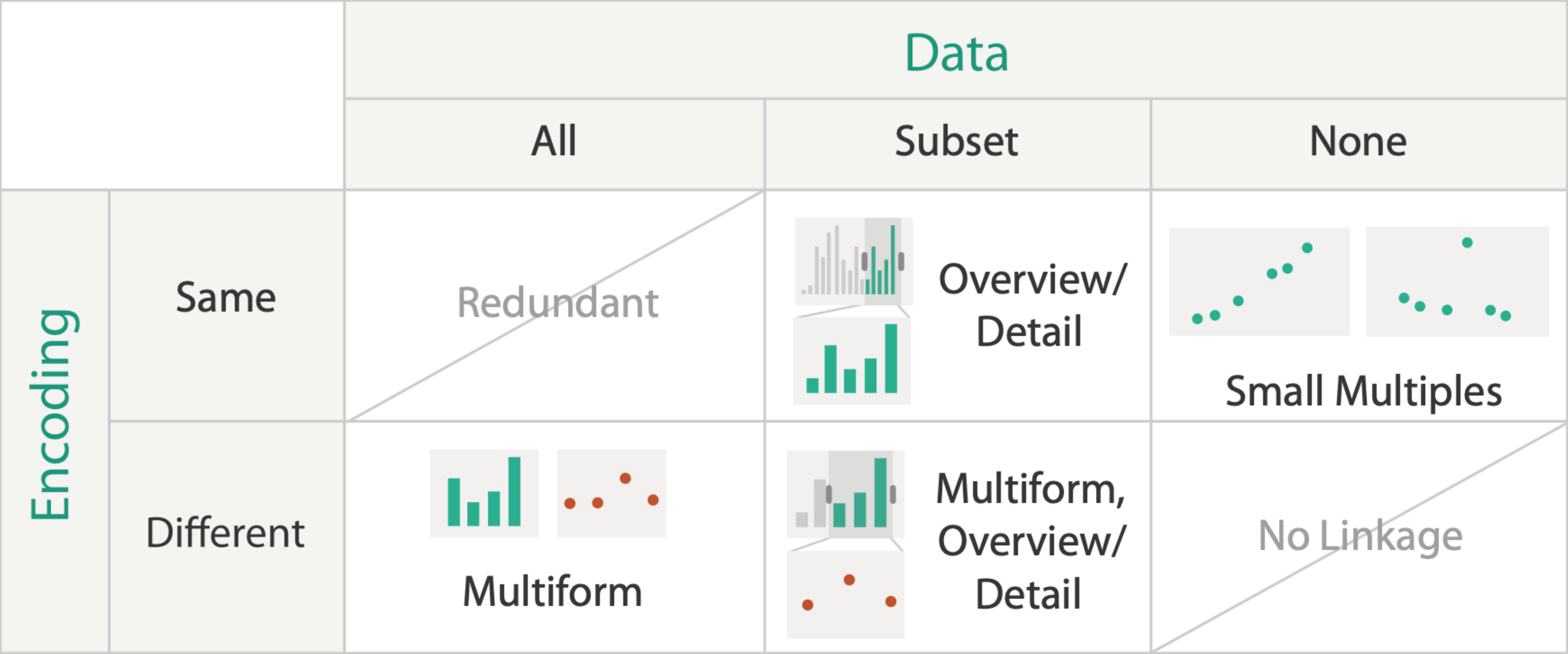
→ Share Data: All/Subset/None

→ Share Navigation

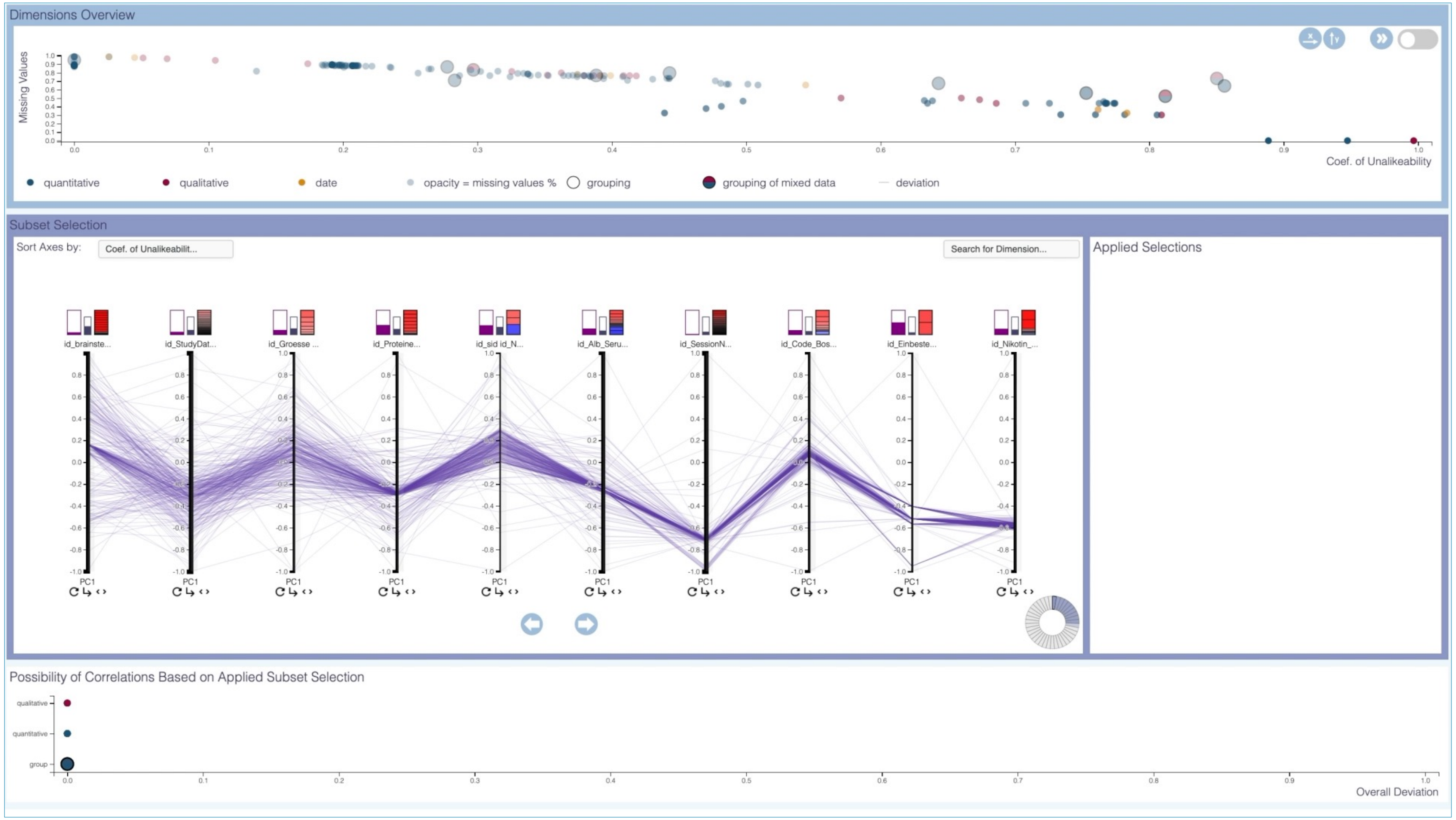
!! Tooltips/rollover are a type of faceting – but if you employ these, assume the user may not discover this information

Design Choices

Coordinated views



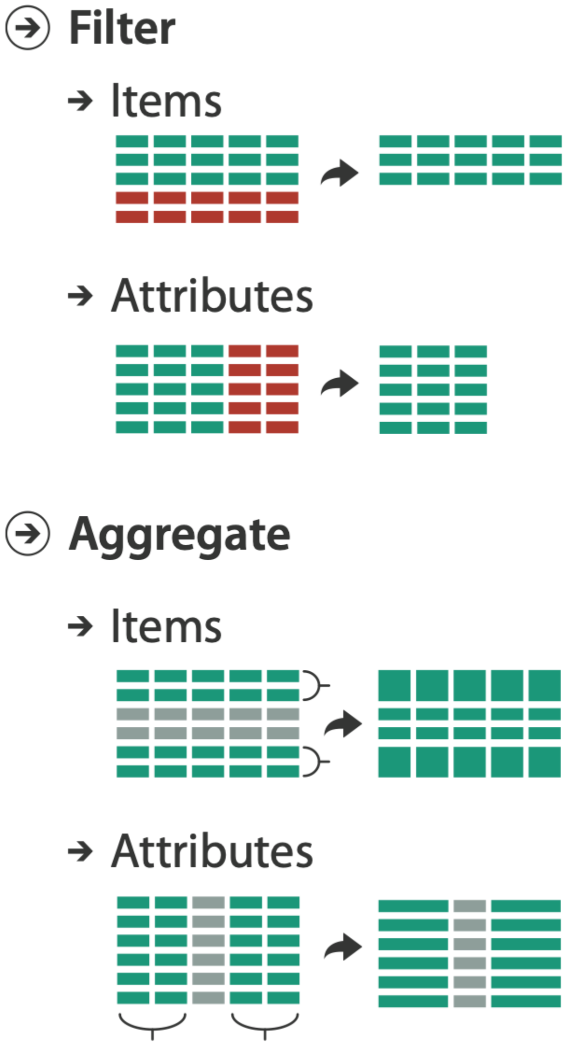
Coordinated Multiple Views

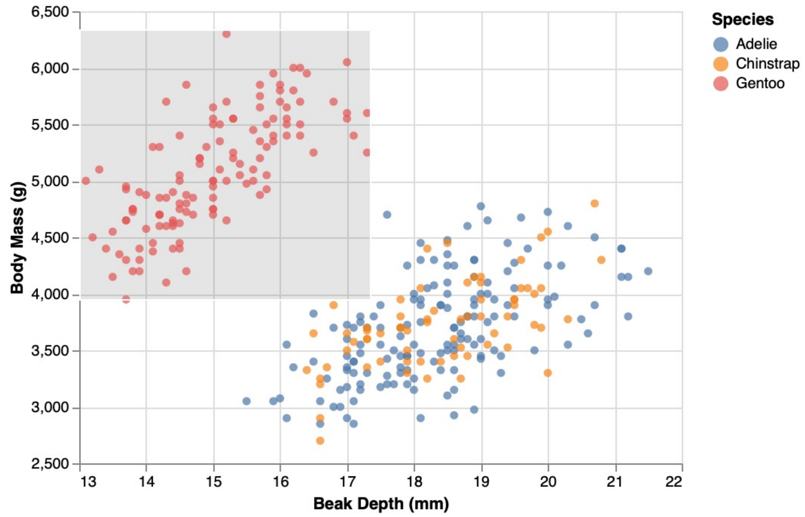


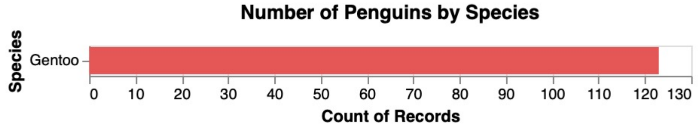
Garrison, L., Müller, J., Schreiber, S., Oeltze-Jafra, S., Hauser, H., & Bruckner, S. (2021). Dimlift: Interactive hierarchical data exploration through dimensional bundling. *IEEE Transactions on Visualization and Computer Graphics*, *27*(6), 2908-2922.

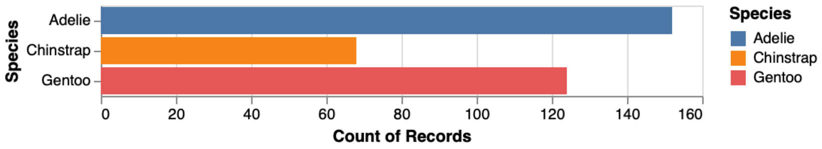
Reduce

- Filter
- Aggregate
- Embed









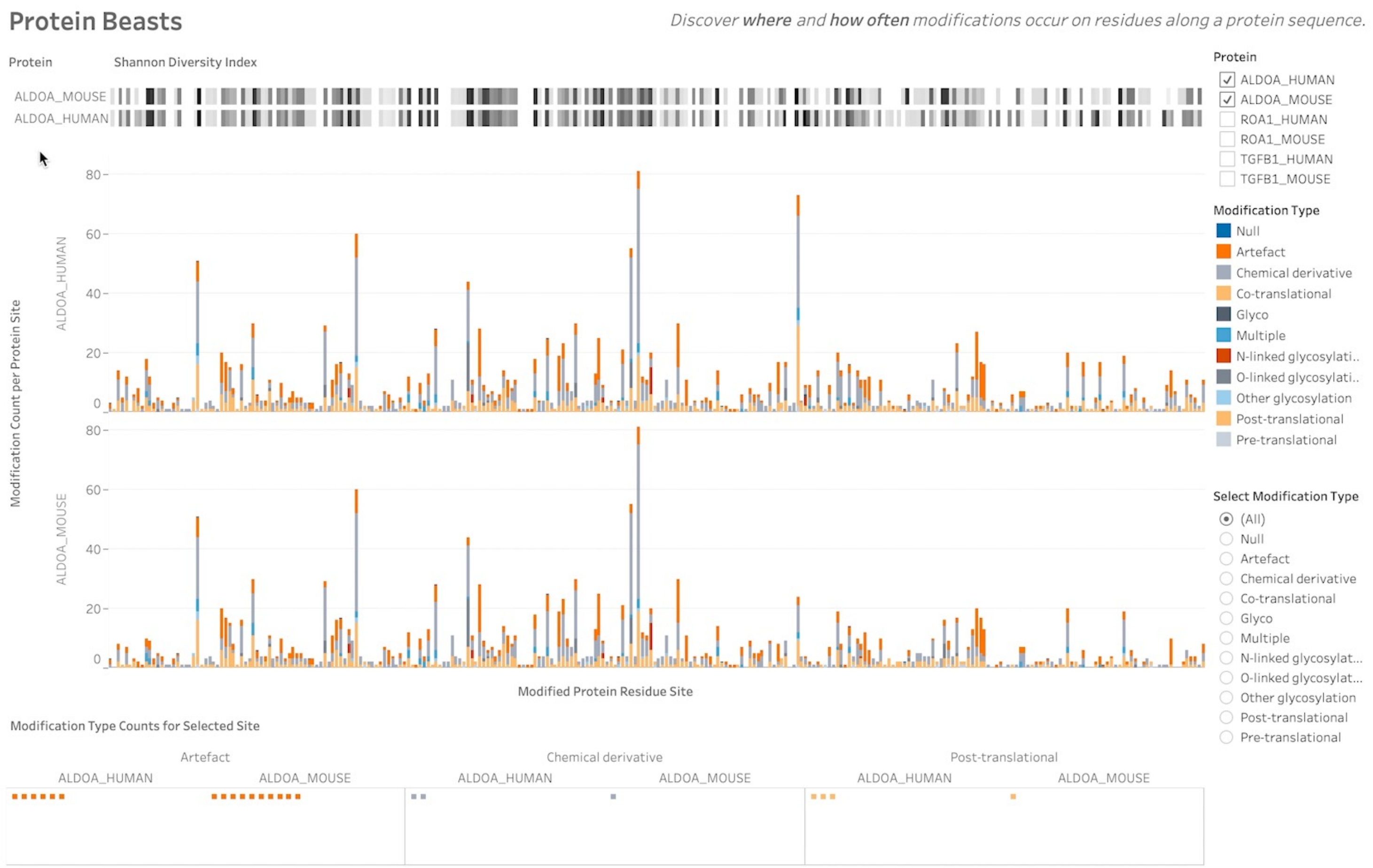
e.g., DR methods

Not mutually exclusive!

Can combine these and other interaction classes

Derive

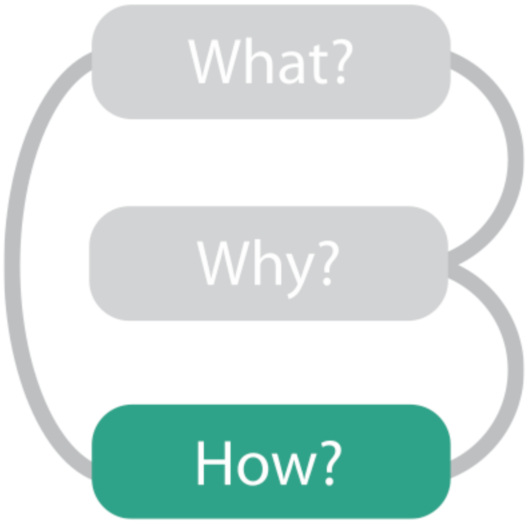
Derived H-index (top view)

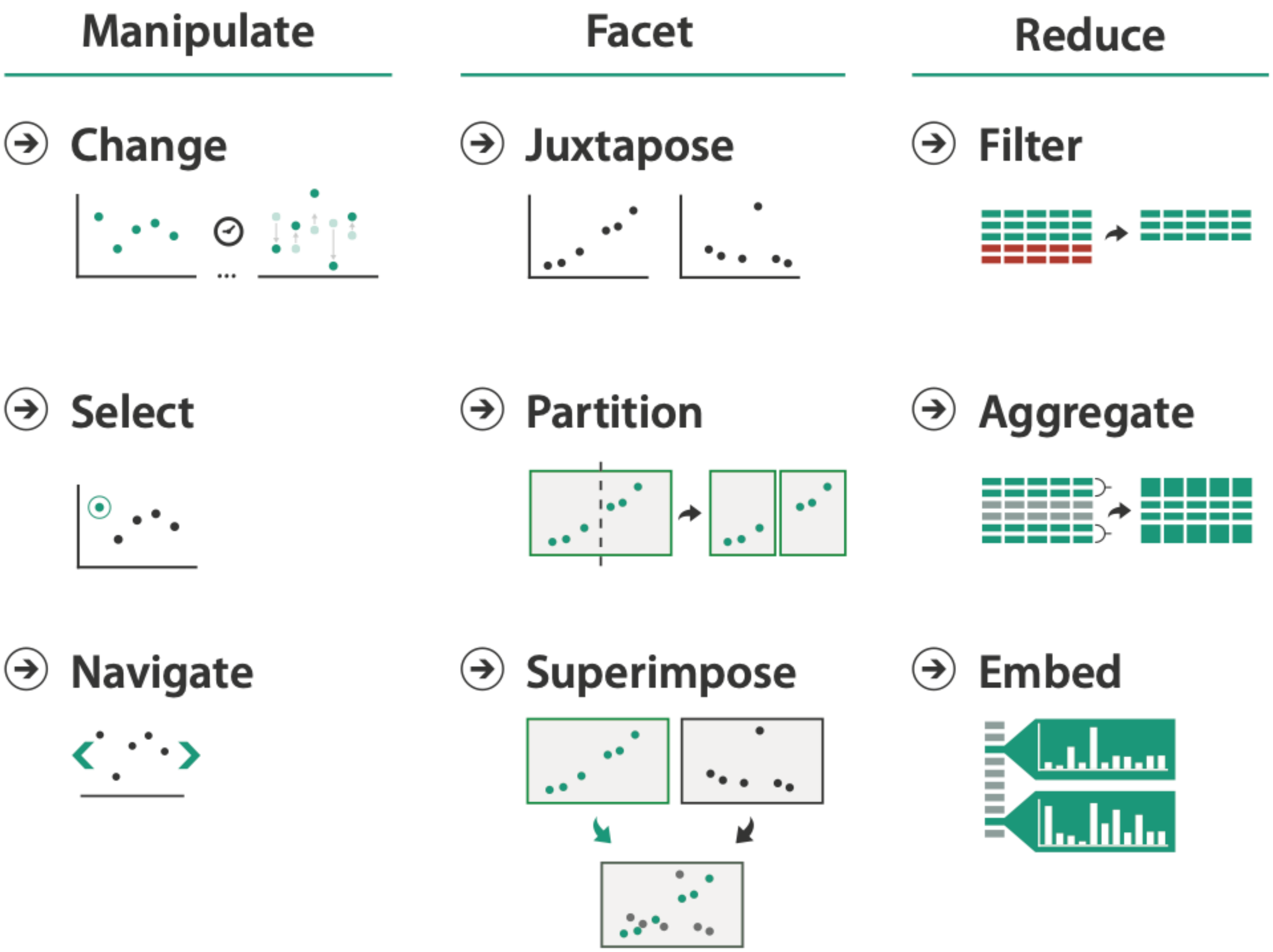


Can you identify other interaction techniques in this visualization?

Caveats

- Interaction has a **time** cost
- Users might not interact as you expect
  - NYTimes found that 90% of users don't interact beyond scrollytelling (Aisch 2016)
- Cognitive load for remembering prior state
- Controls vs invisible functionality
  - controls may take up valuable real estate
  - invisible functionality may be too invisible (no one figures it out)


What?
Why?
How?


Manipulate
Facet
Reduce
Change
Juxtapose
Filter
Select
Partition
Aggregate
Navigate
Superimpose
Embed

Summary

What?

Datasets

Data Types

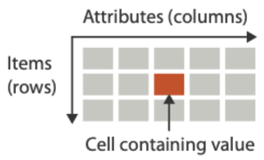
→ Items → Attributes → Links → Positions → Grids

Data and Dataset Types

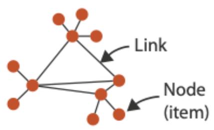
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

Dataset Types

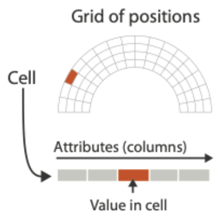
→ Tables



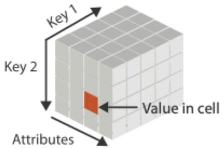
→ Networks



→ Fields (Continuous)



→ *Multidimensional Table*



→ *Trees*

→ Geometry (Spatial)

Dataset Availability

→ Static

→ Dynamic

Attributes

Attribute Types

→ Categorical

→ Ordered

→ *Ordinal*

→ *Quantitative*

Ordering Direction

→ Sequential

→ Diverging

→ Cyclic

What?

Why?

Actions

Analyze

- Consume
  - *Discover*
  - *Present*
  - *Enjoy*
- Produce
  - *Annotate*
  - *Record*
  - *Derive*

tag

Search

	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Query

- Identify
- Compare
- Summarize

Targets

All Data

- Trends
- Outliers
- Features

Attributes

- One
  - *Distribution*
  - *Extremes*
- Many
  - *Dependency*
  - *Correlation*
  - *Similarity*

Network Data

- Topology
  - *Paths*

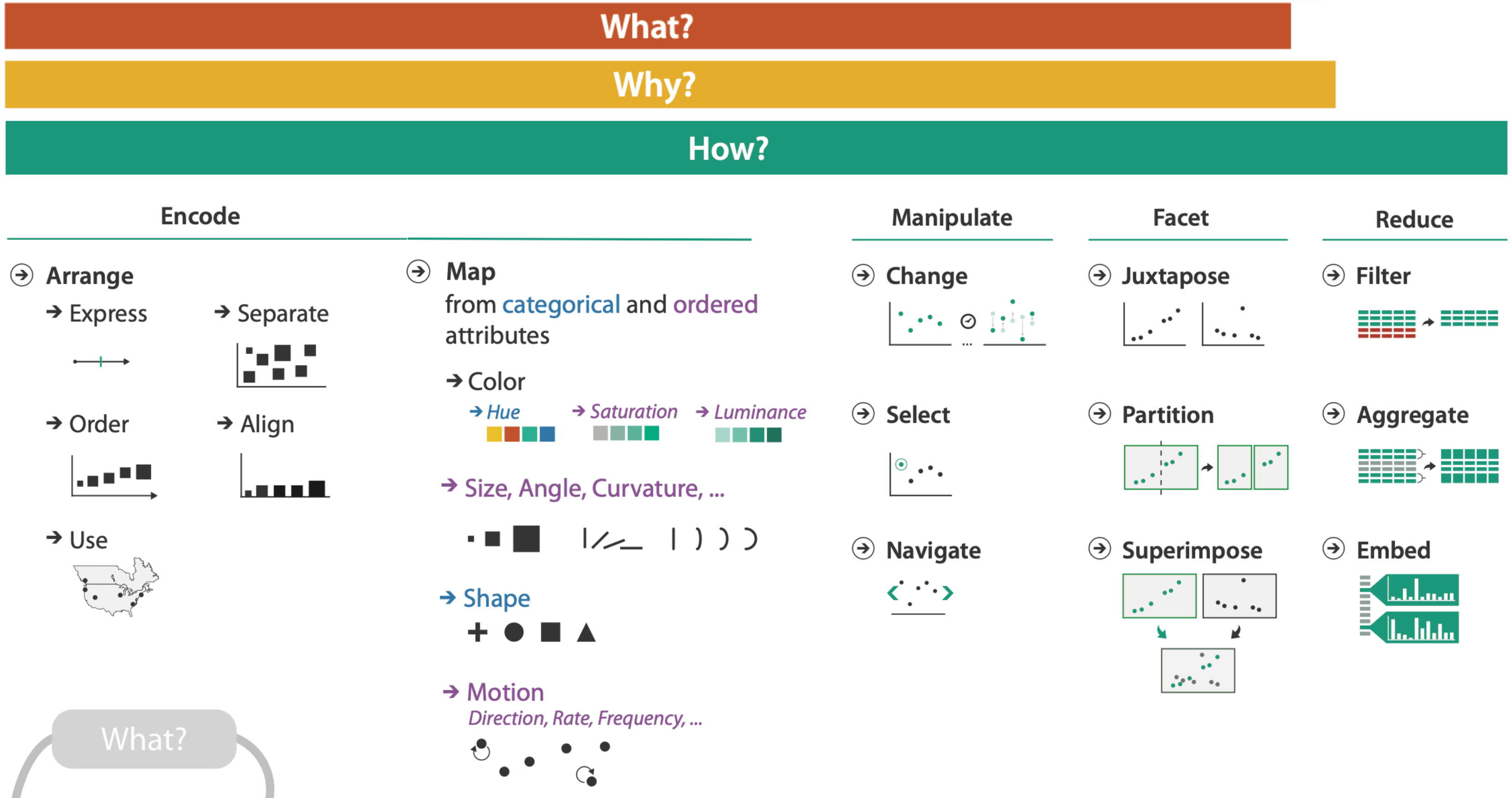
Spatial Data

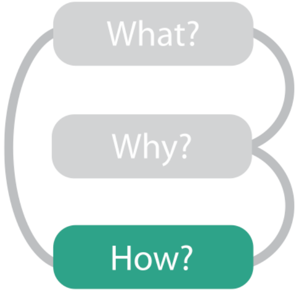
- Shape

What?

Why?

How?

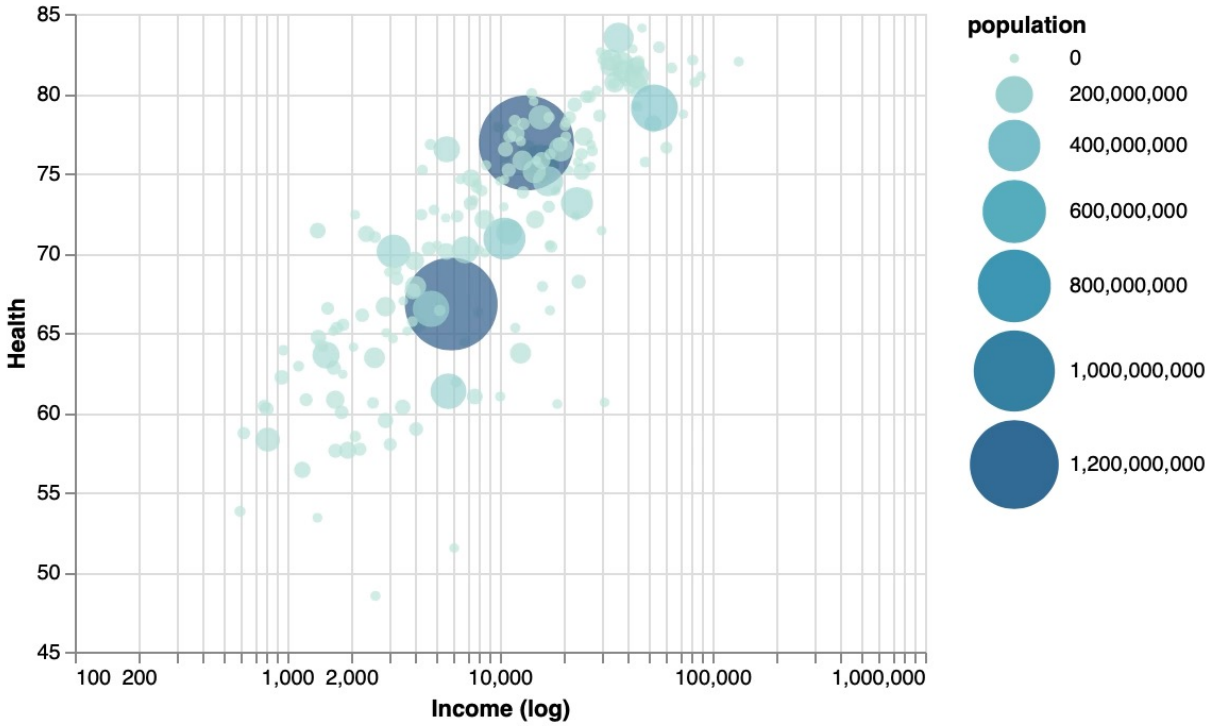




Data abstraction -> Task abstraction -> Visual + Interaction Encoding

Next Up

- Visualization in practice
- After lunch
    - Introduction to Vega-Altair
    - Mini-project in visualization

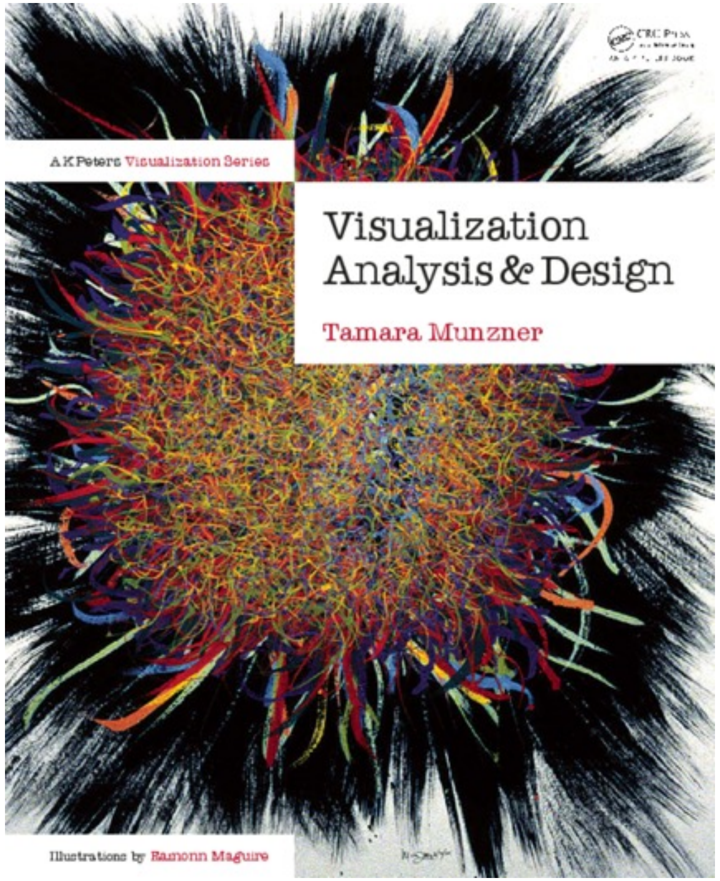


https://altair-viz.github.io/gallery/index.html

Questions?

Further Reading & Acknowledgement

- Web material for Visual Analysis & Design: https://www.cs.ubc.ca/~tmm/talks/vadbook (source material for many slides in this lecture)



Munzner, T. (2014). Visualization analysis and design. AK Peters Visualization Series, CRC Press, Visualization Series. Chp 2.