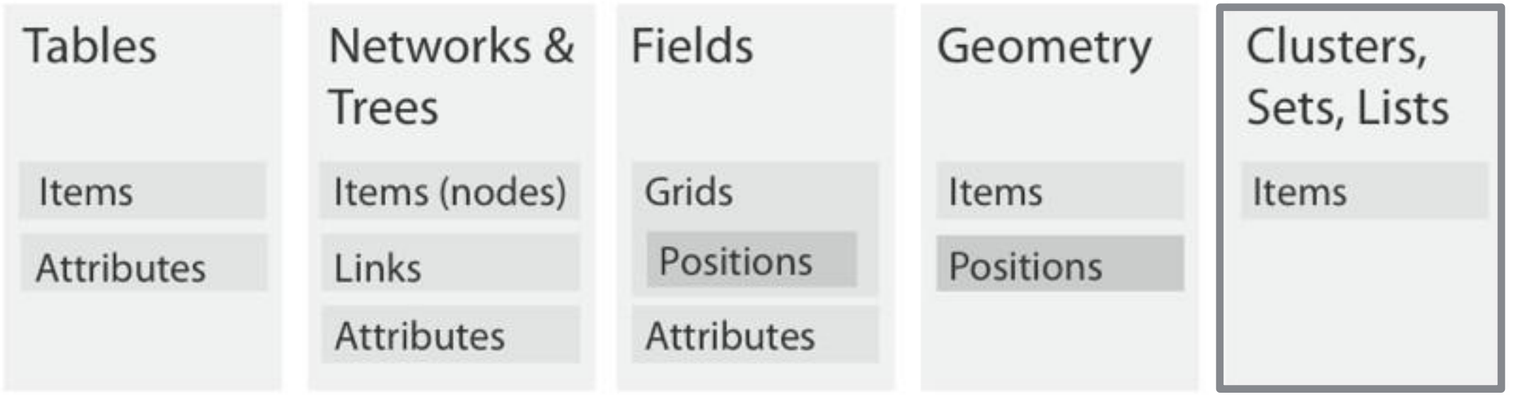
Paul Rosen
paul.rosen@utah.edu
@paulrosenphd
https://cspaul.com

Visualization for Data Science
DS-4630 / CS-5630 / CS-6630

VISUALIZING SETS



(hint: these are categorical data)

thought experiment…

- item: lego
- attributes: ???

thought experiment…

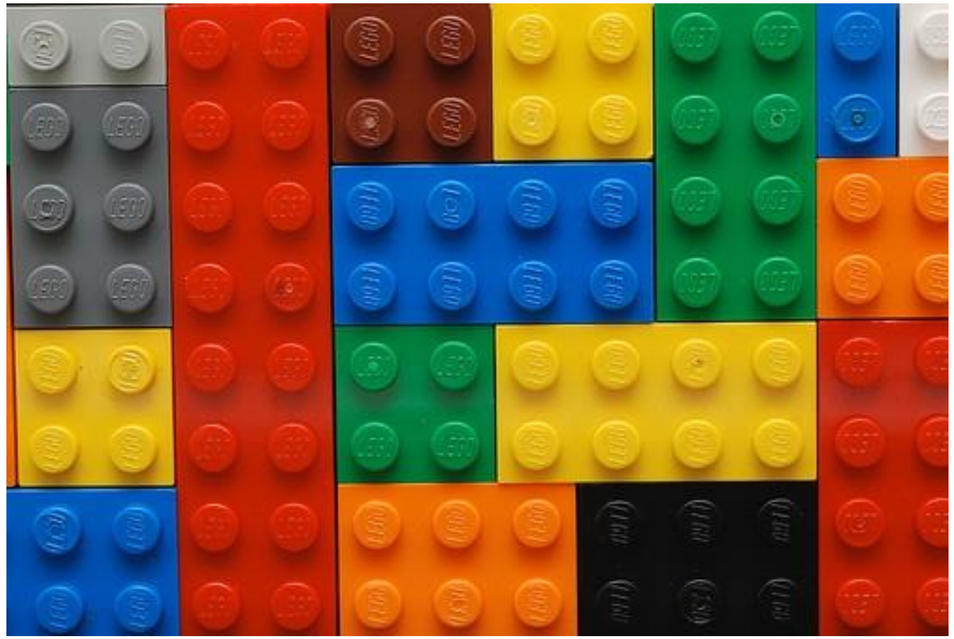
- item: lego
- attributes:
  - color
  - height
  - width
  - length
  - shape

dataset

dataset: more realistic

dataset

- where do we start?
- we need to organize!
- but, how?

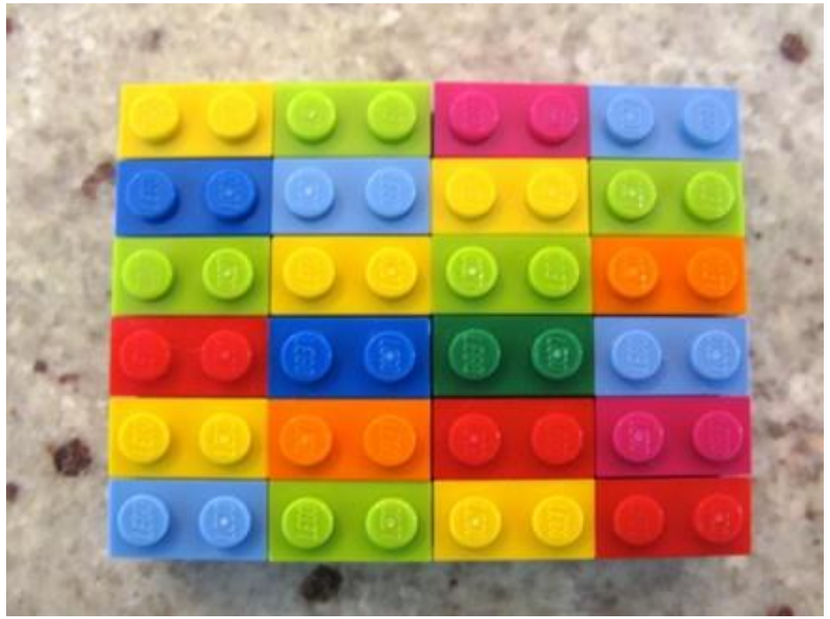


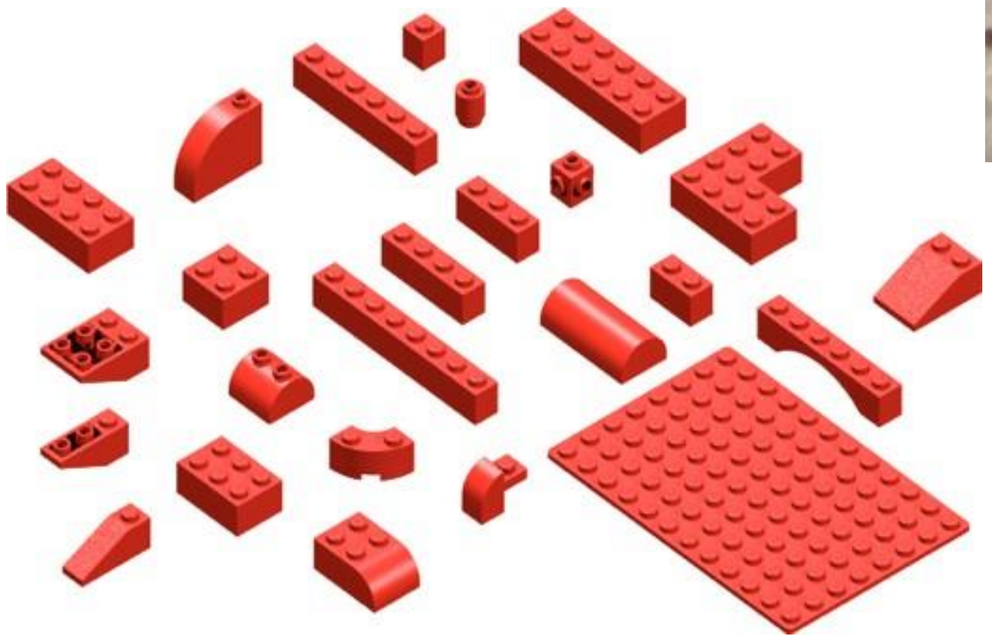
dataset

- sort by color

dataset

- sort by size, shape

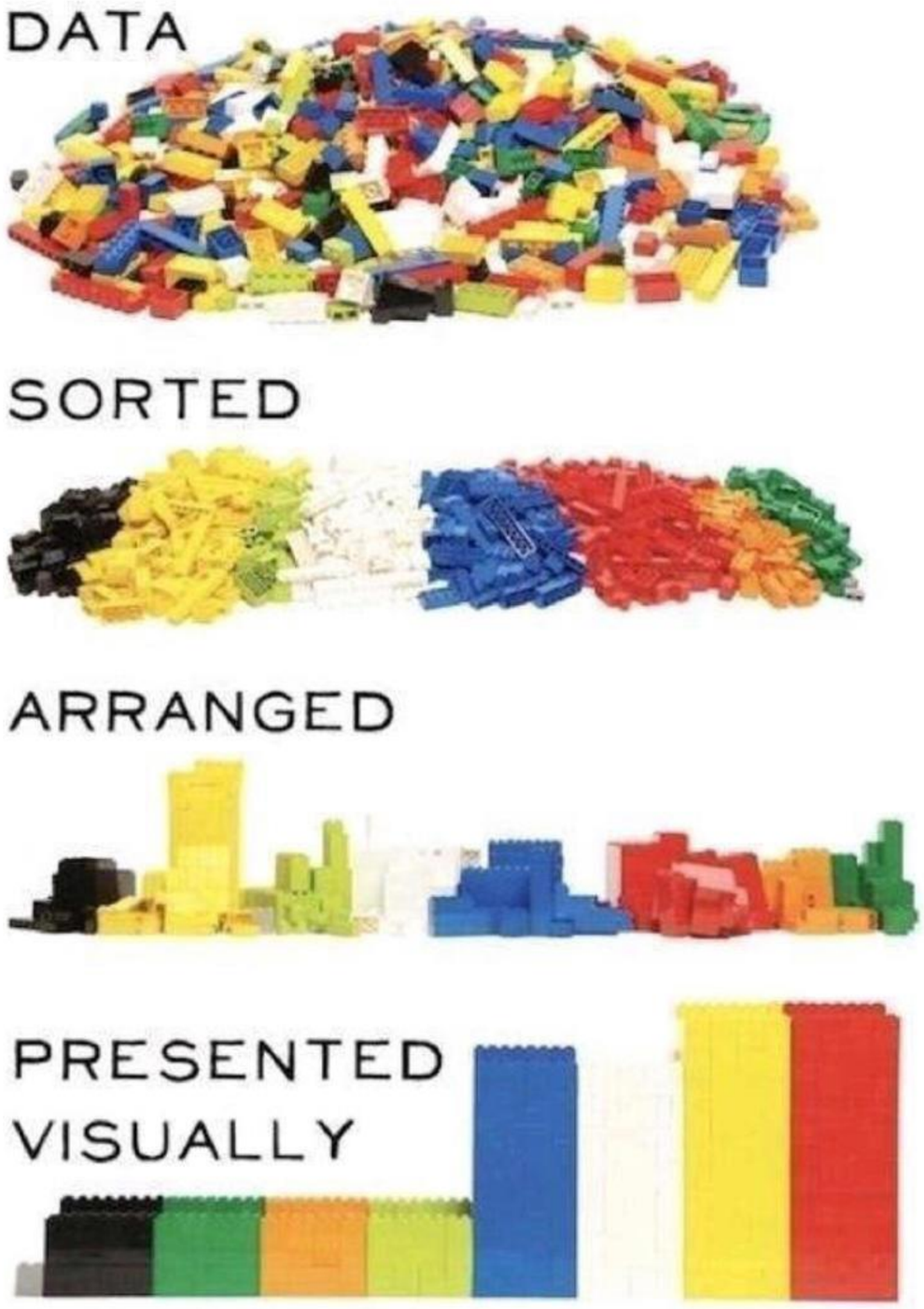




dataset

- task: organization
- drawbacks?

Set Challenge

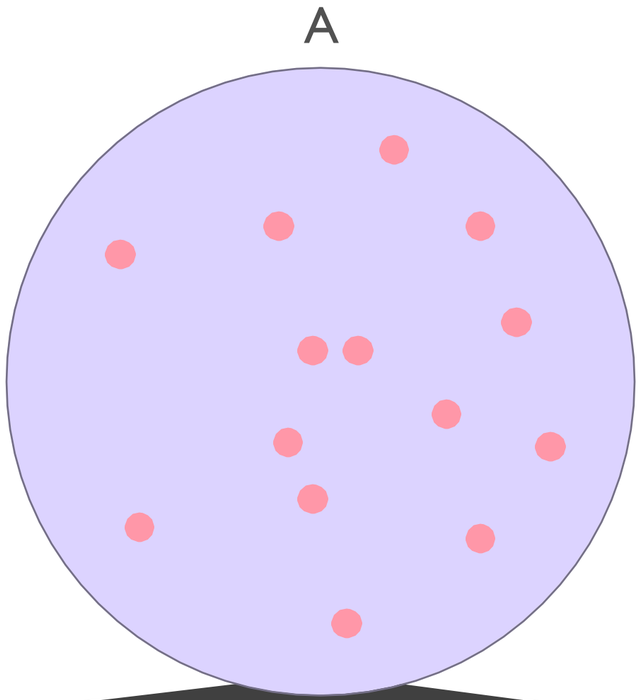


Dataset

- organization leads us to a set problem
- so what are sets?

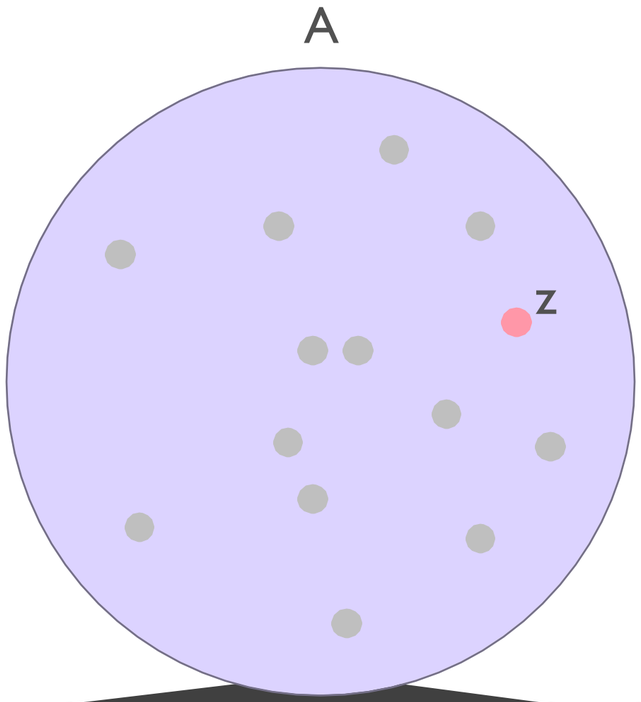
set theory

- set
  - a collection of objects
  - some set: A



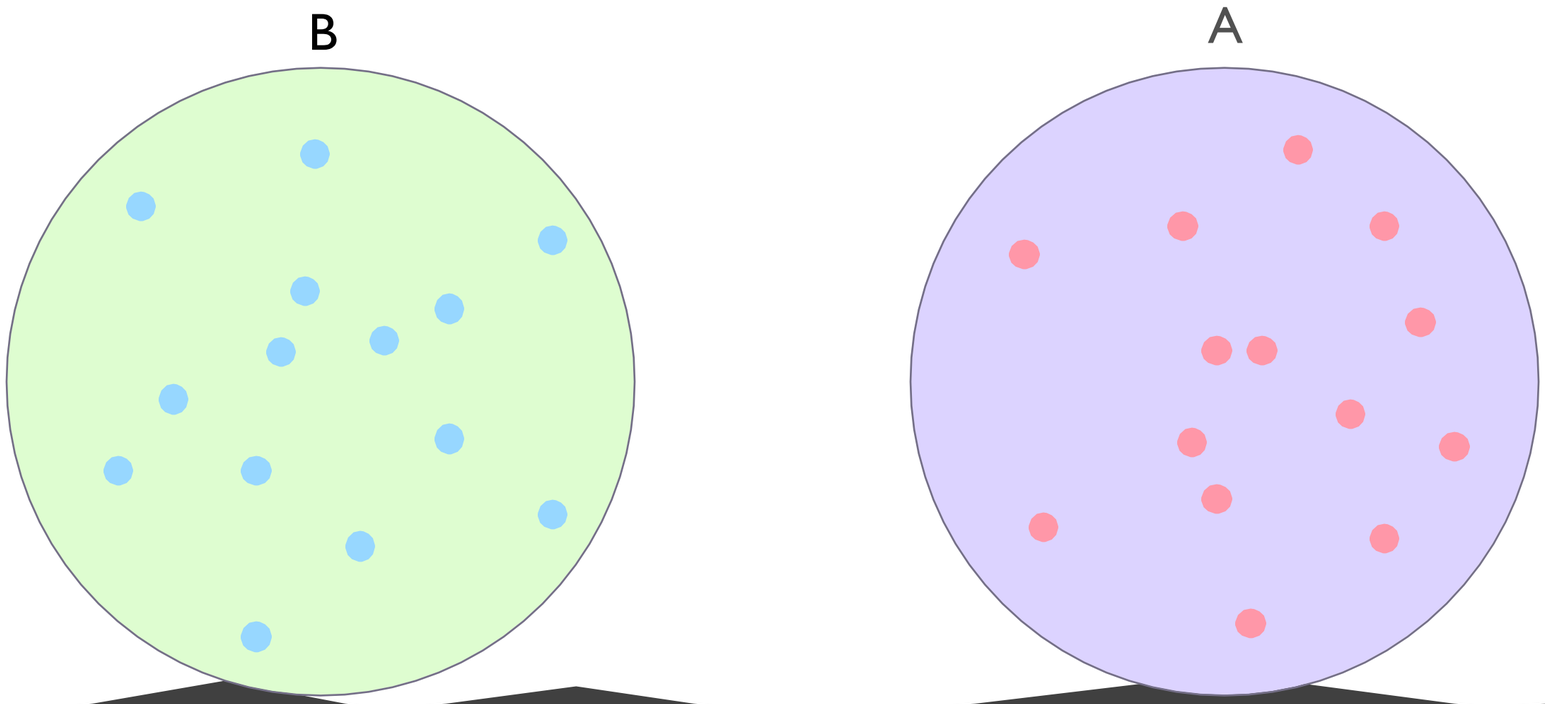
set theory

- set
  - a collection of objects
  - some set: A
- object
  - some object: z
  - $z \in A$



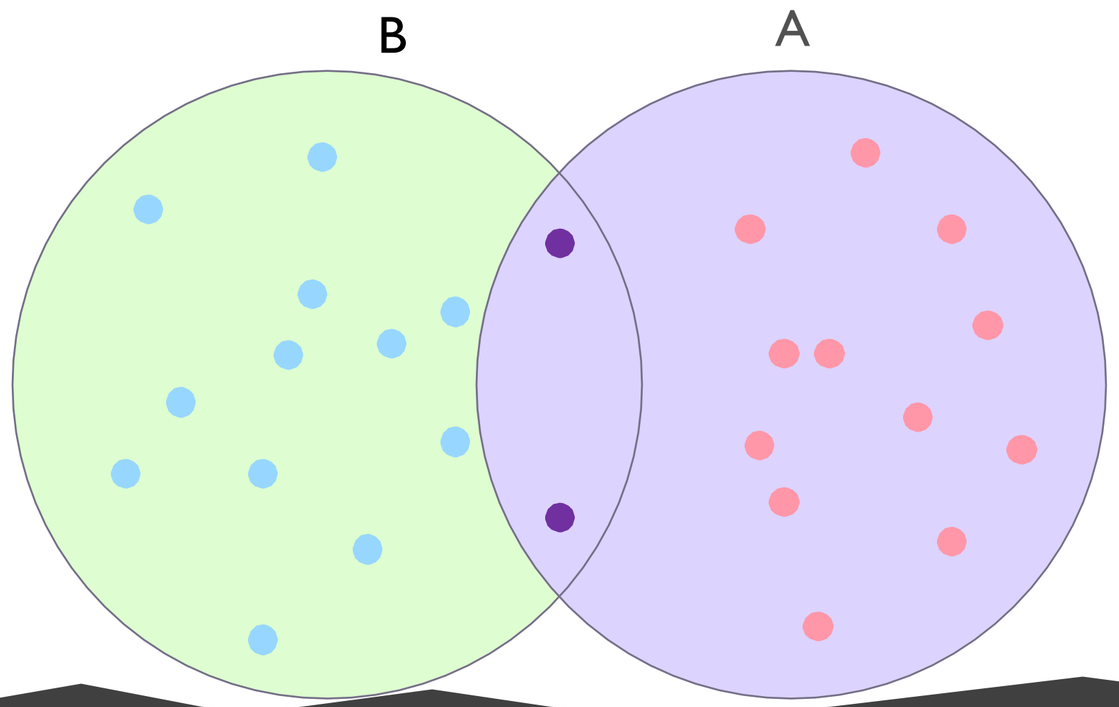
set theory

- multiple sets: A & B



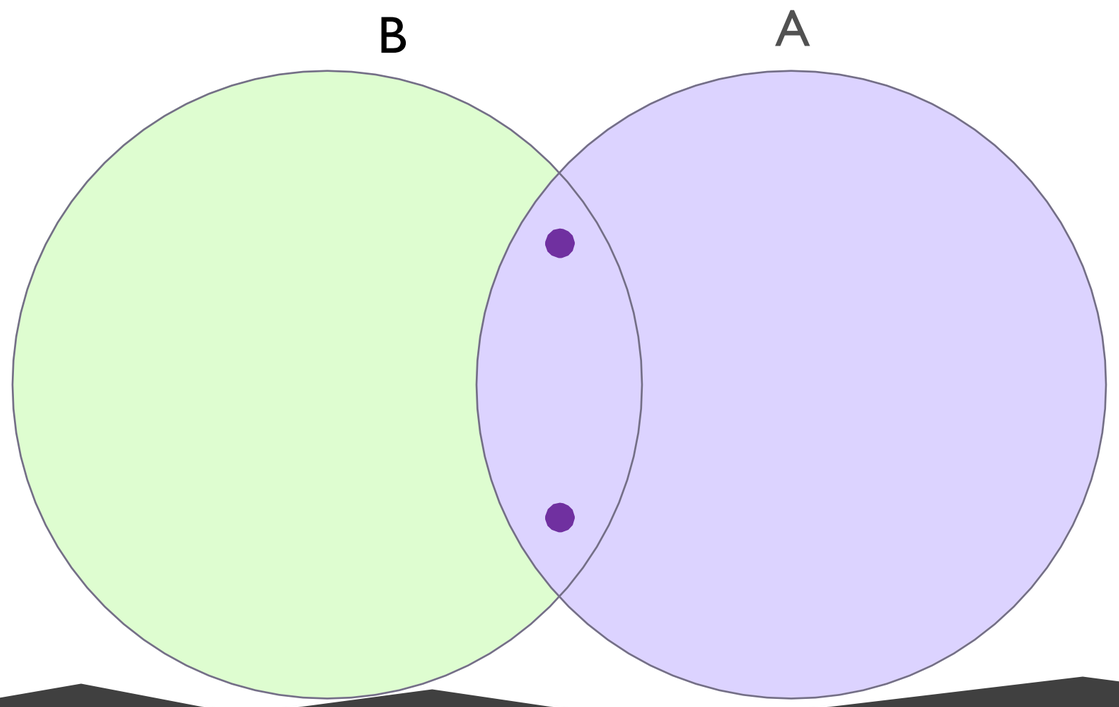
set theory

- union: A ∪ B



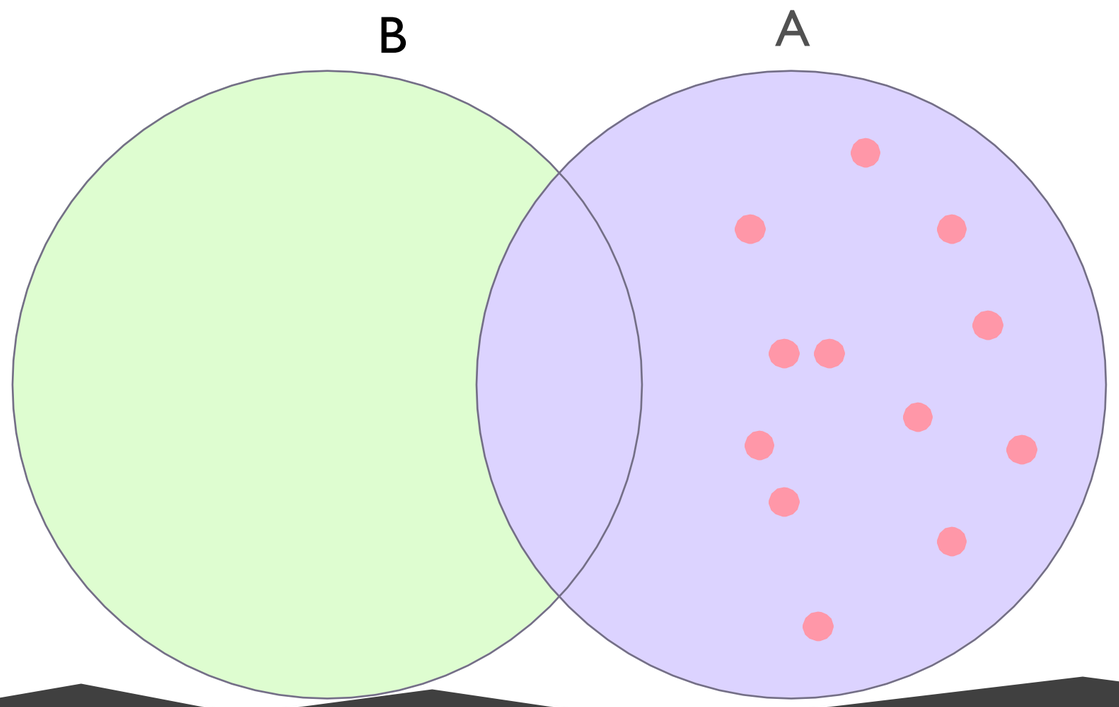
set theory

- intersection: A ∩ B



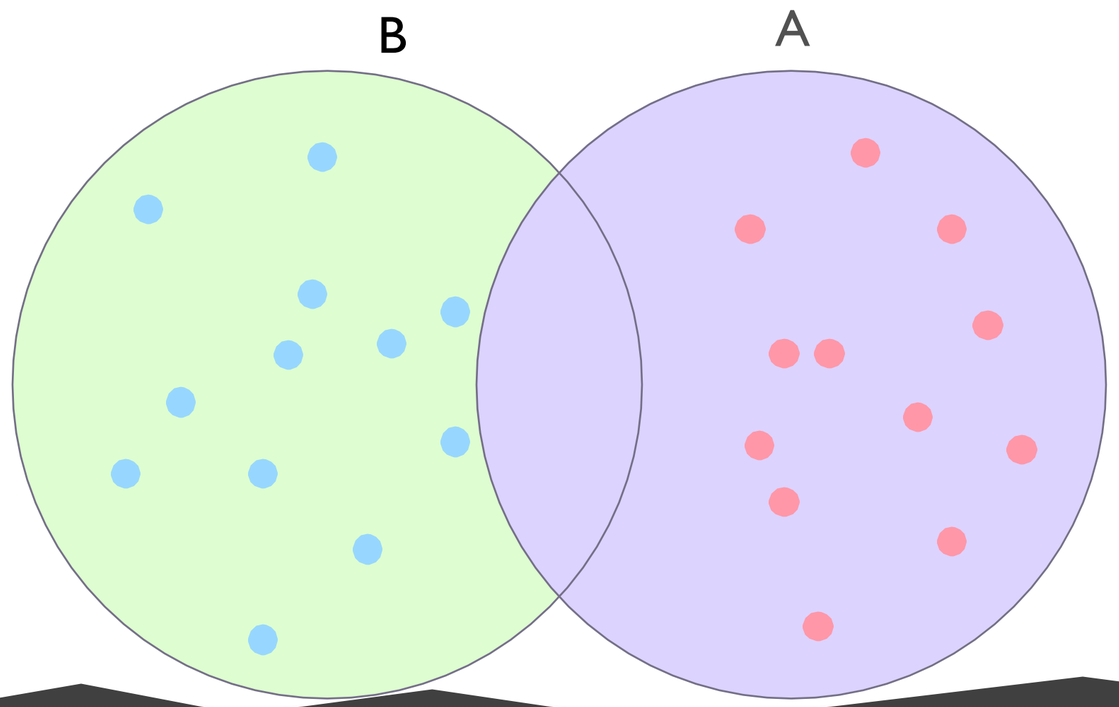
set theory

- set difference: A \ B



set theory

- symmetric difference: A ⊖ B

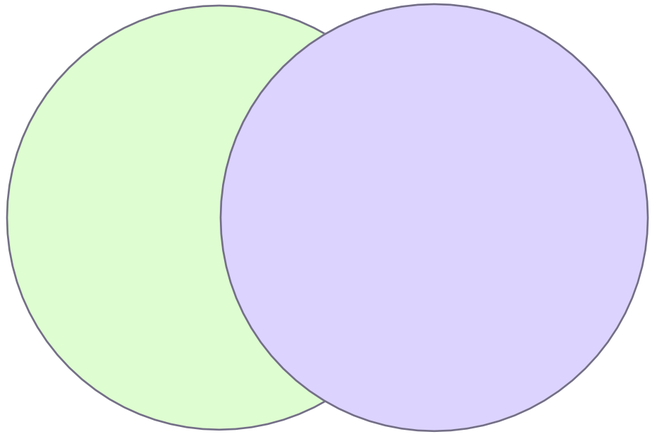


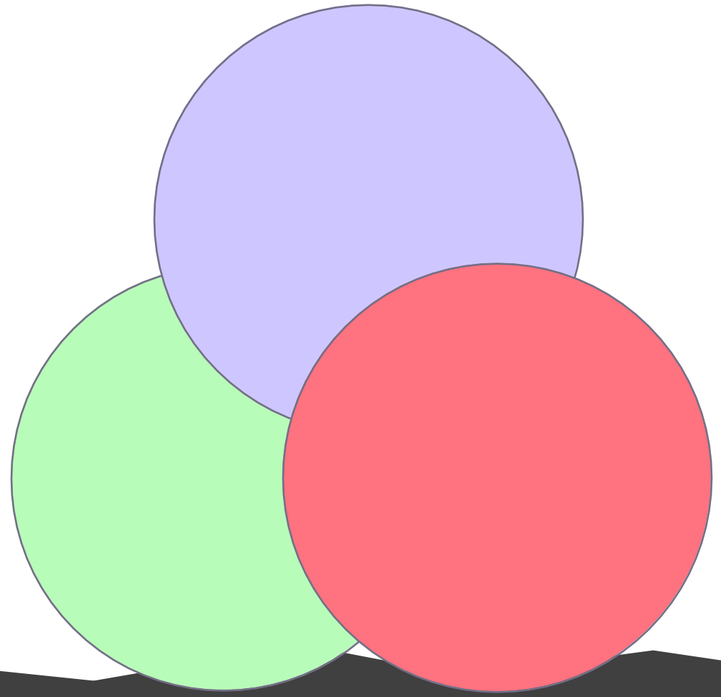
http://students.brown.edu/seeing-theory/index.html

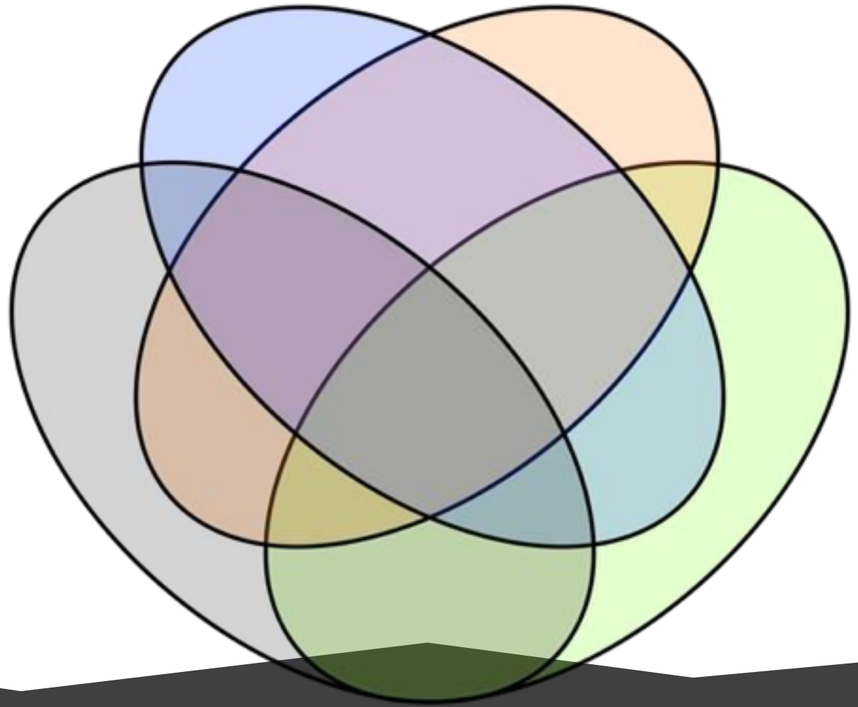
visualizing sets

venn diagrams

- show all possible relationships

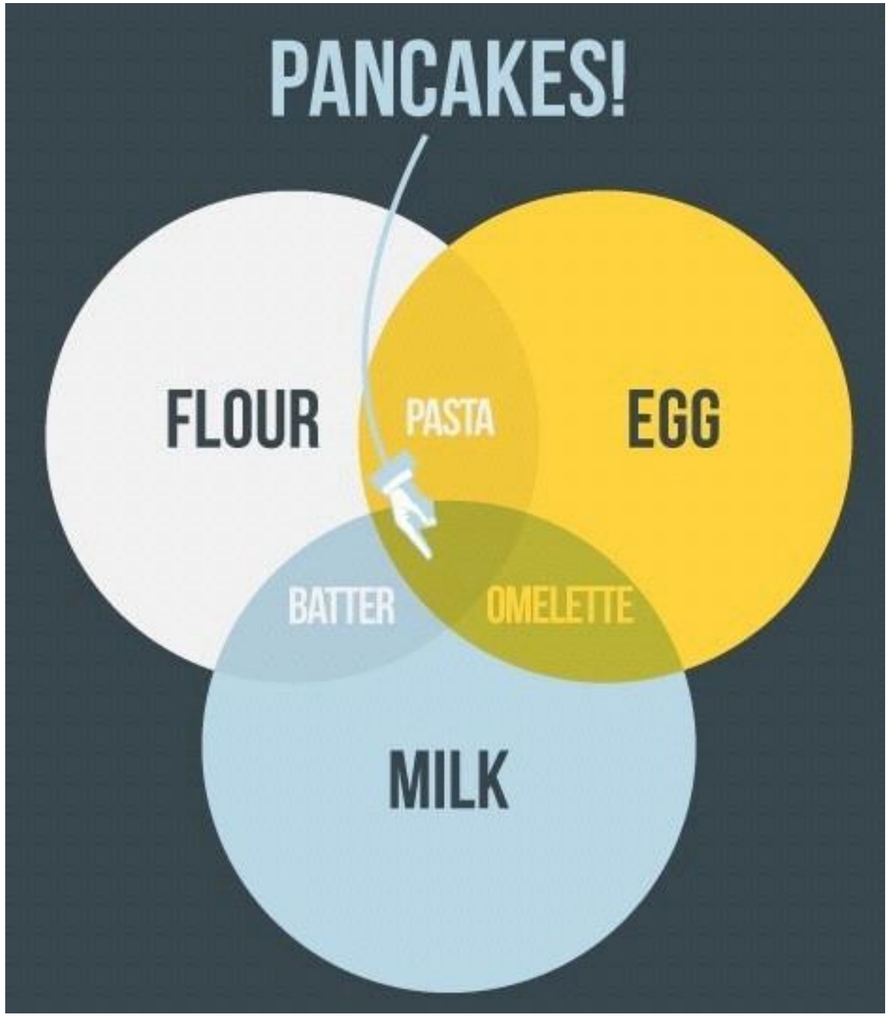






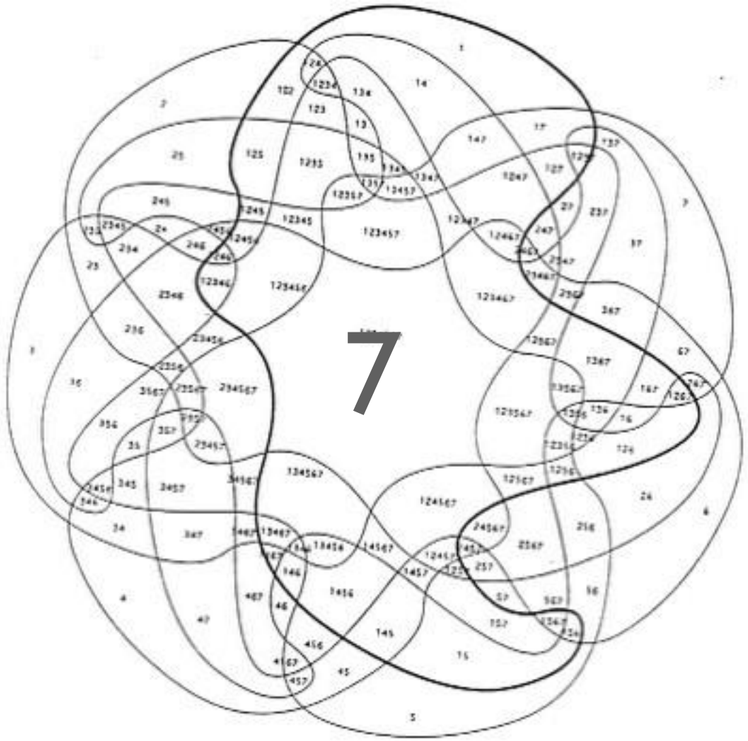
venn diagrams

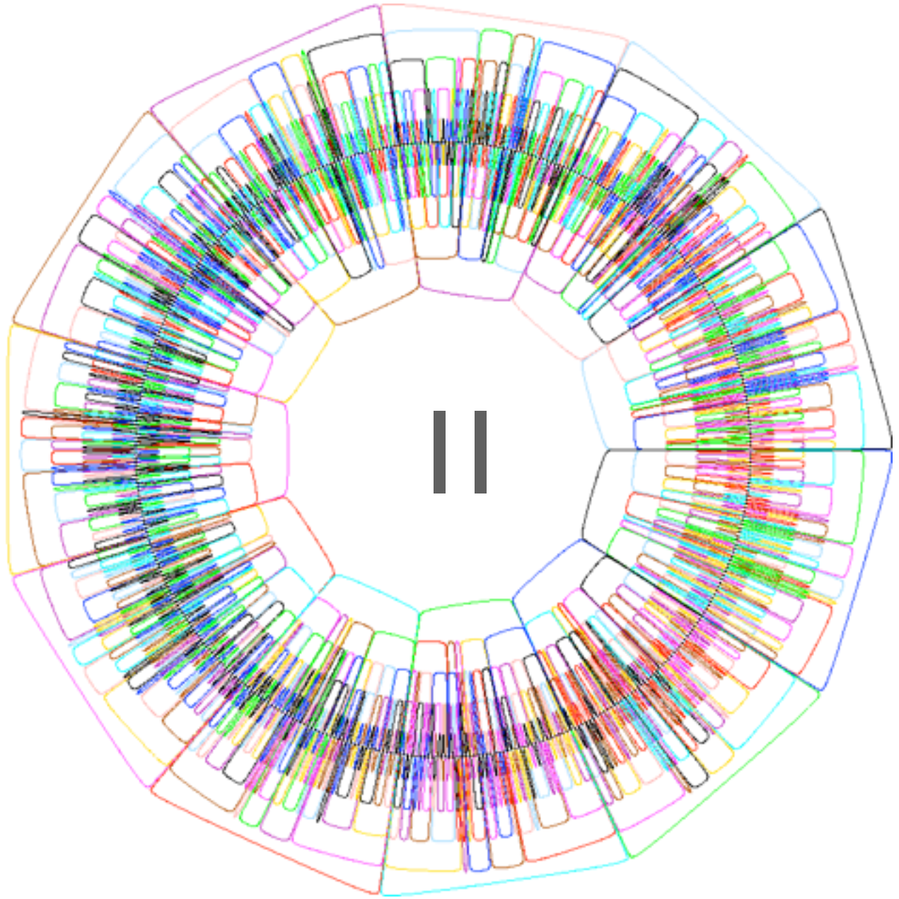
- “casual infovis”



venn diagrams

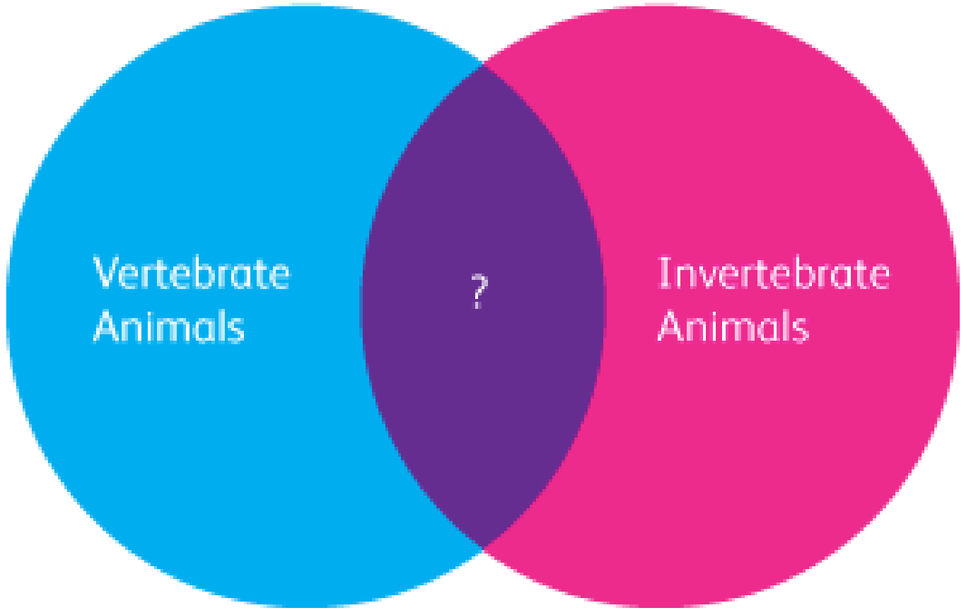
- get messy fast





venn diagrams

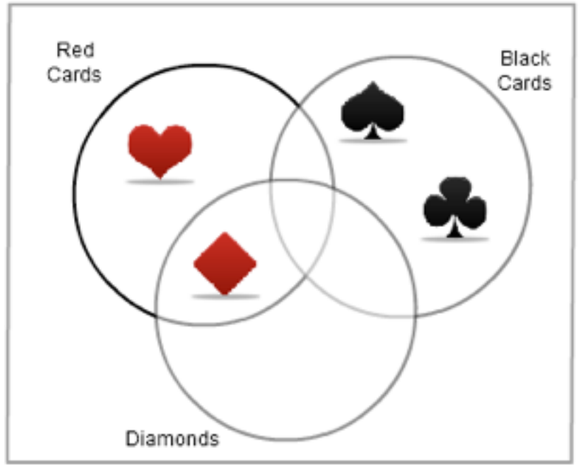
- non-sensical



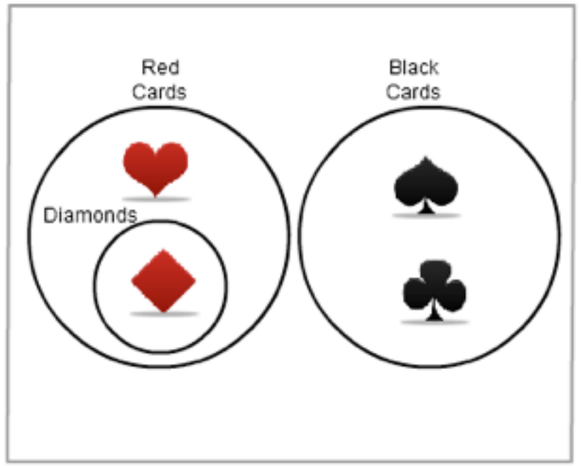
euler diagrams

- show only existing relationships

VENN

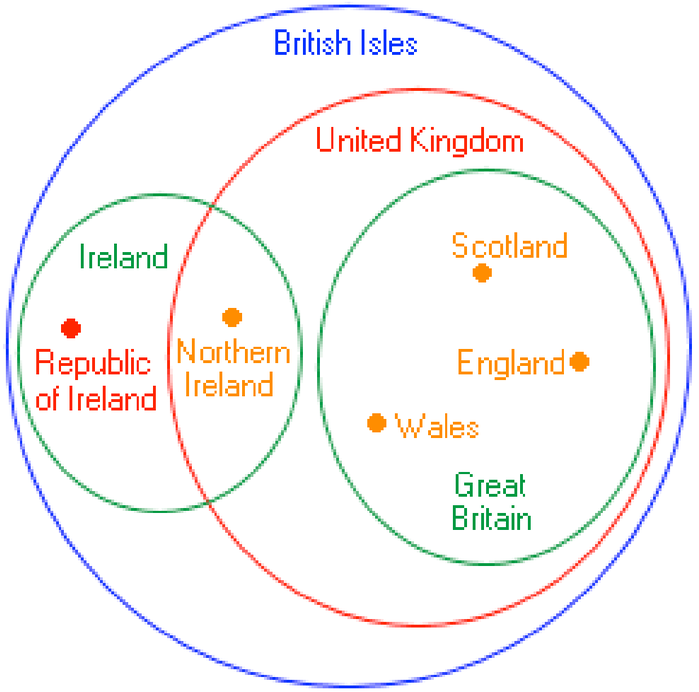


EULER

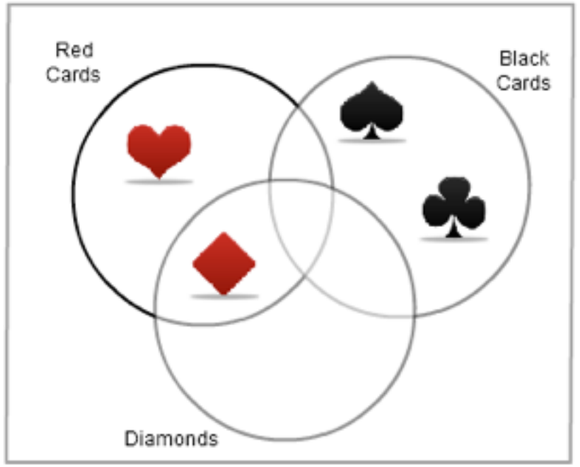


euler diagrams

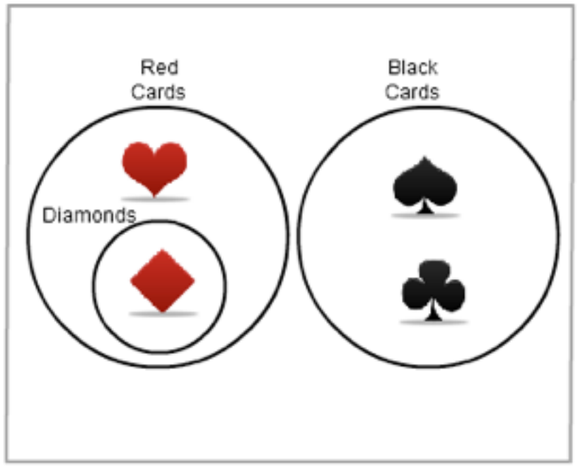
- show only existing relationships



VENN



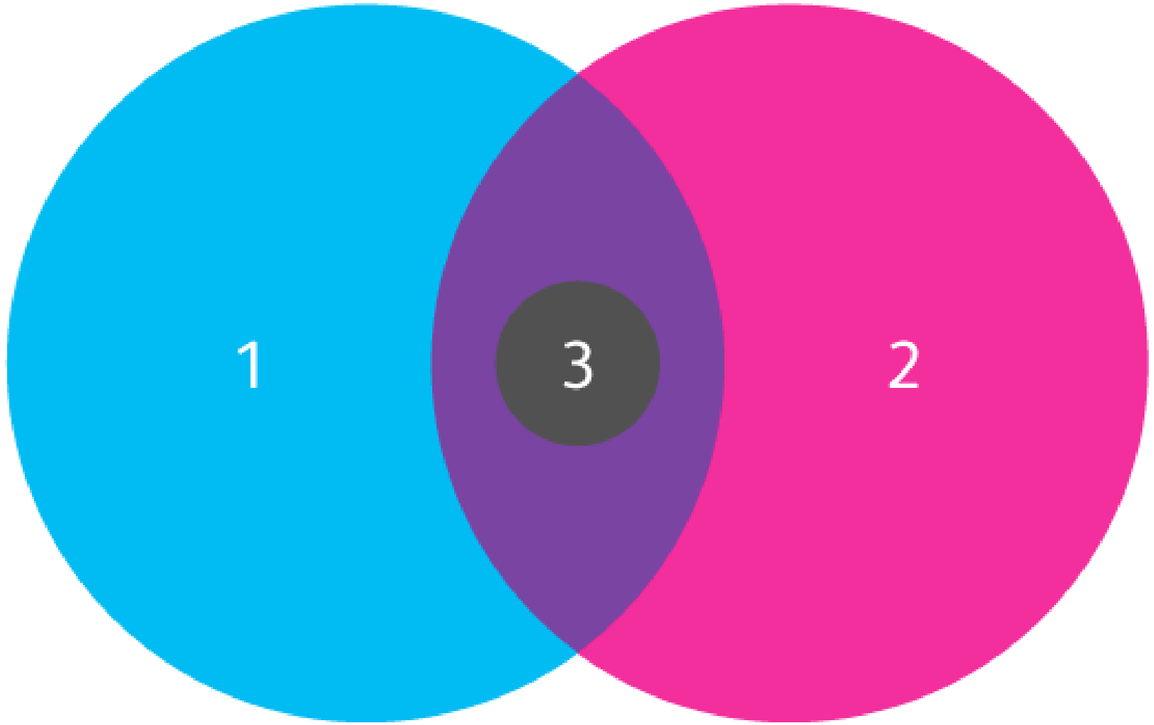
EULER



euler diagrams

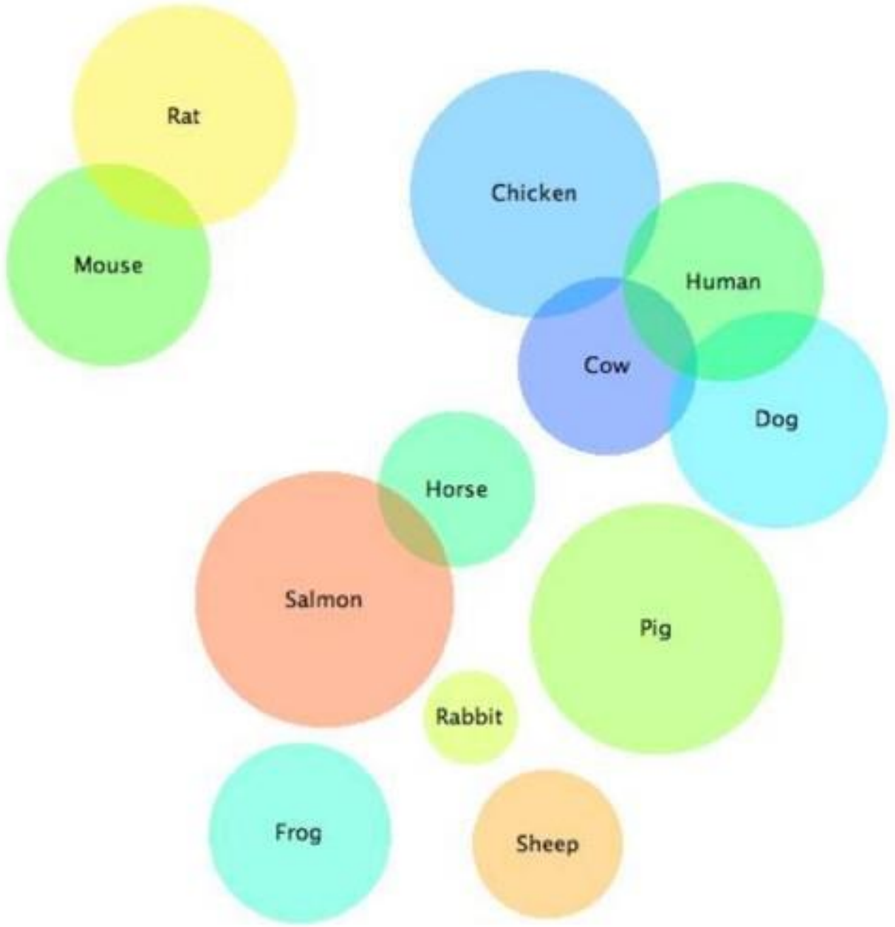
- Misunderstood

1: People who know what a Venn Diagram is.
2: People who know what an Euler Diagram is.
3: People who know the difference.



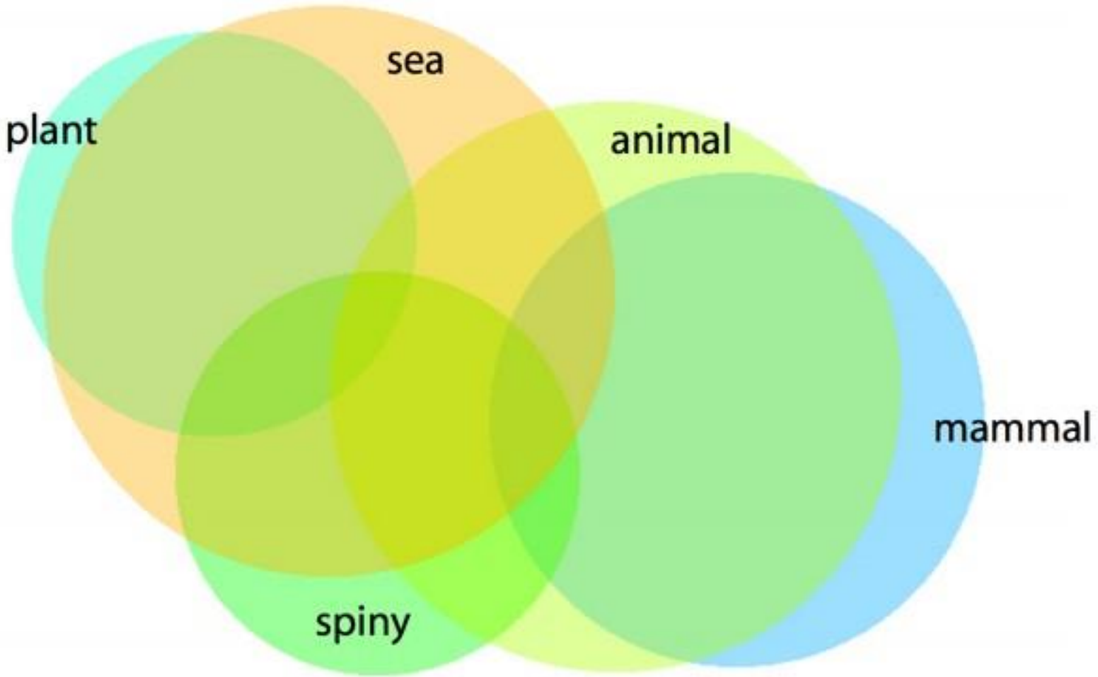
venn & euler diagrams

- adjust for area
- starts getting tricky!

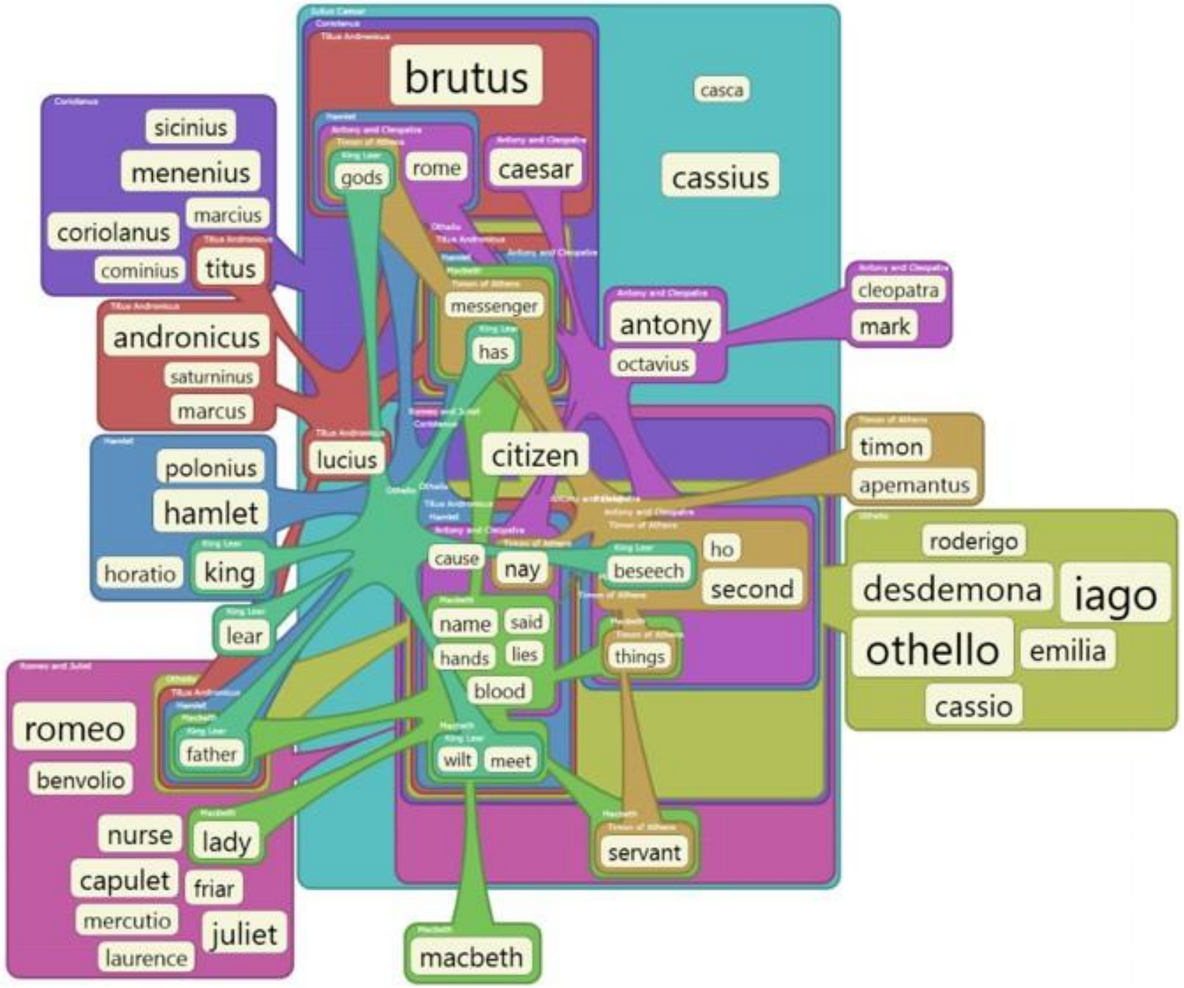


venn & euler diagrams

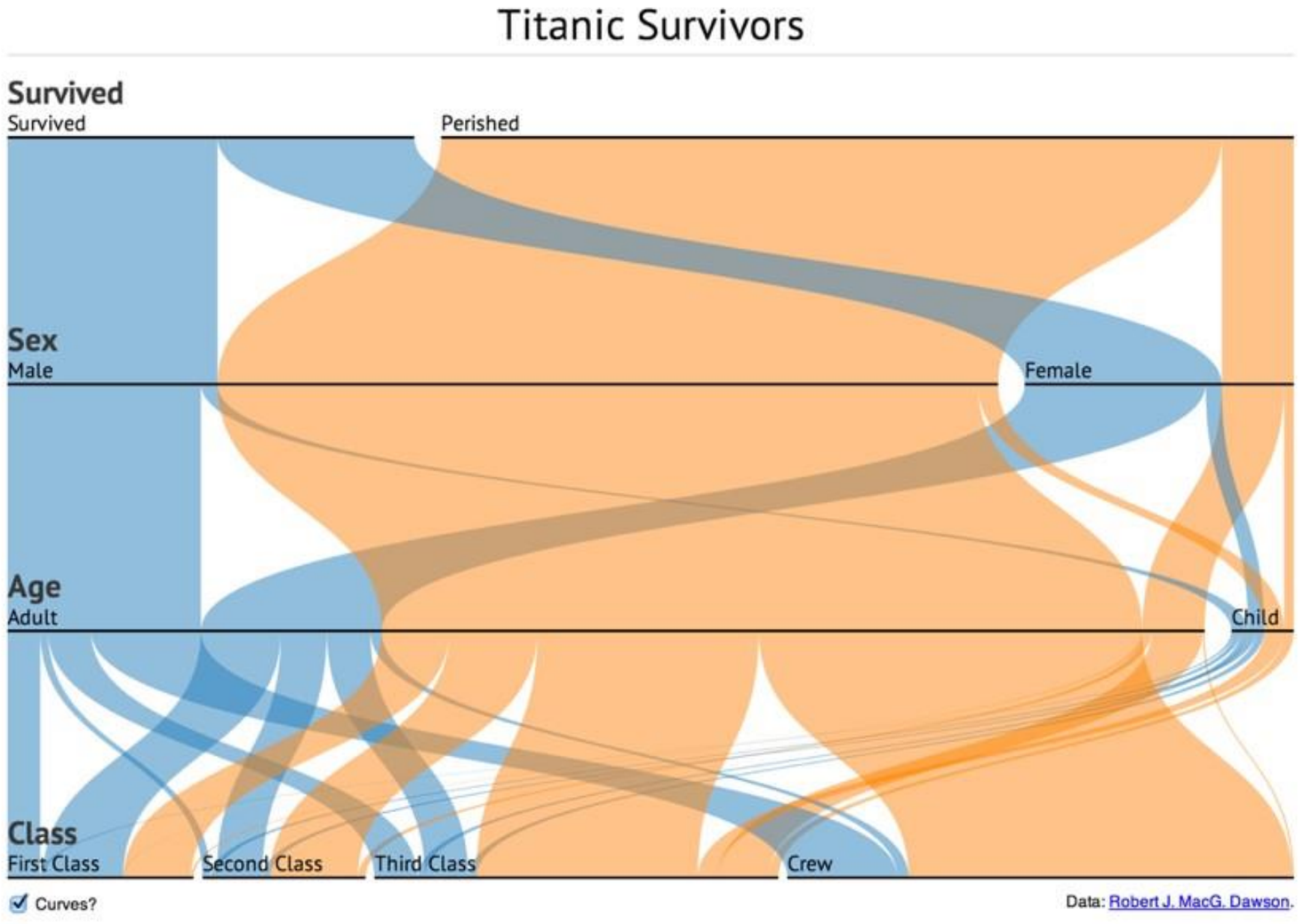
- adjust for area
- starts getting tricky!



compact euler diagrams



parallel sets

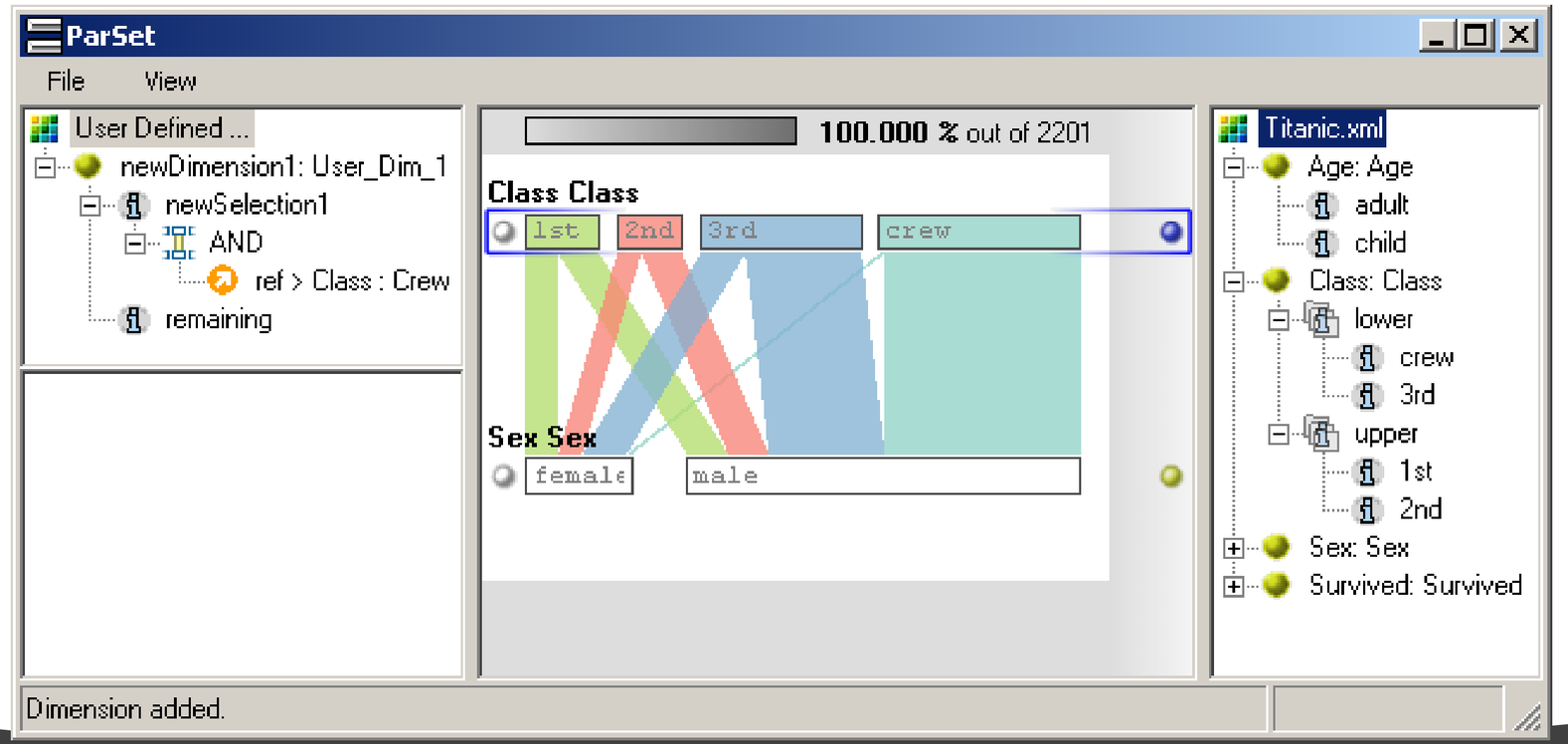


parallel sets

- builds on PC to better handle categorical data
  - discrete
  - small number of values
  - no implied ordering between attributes
- task: find relationship between attributes, not outliers
- interaction driven technique

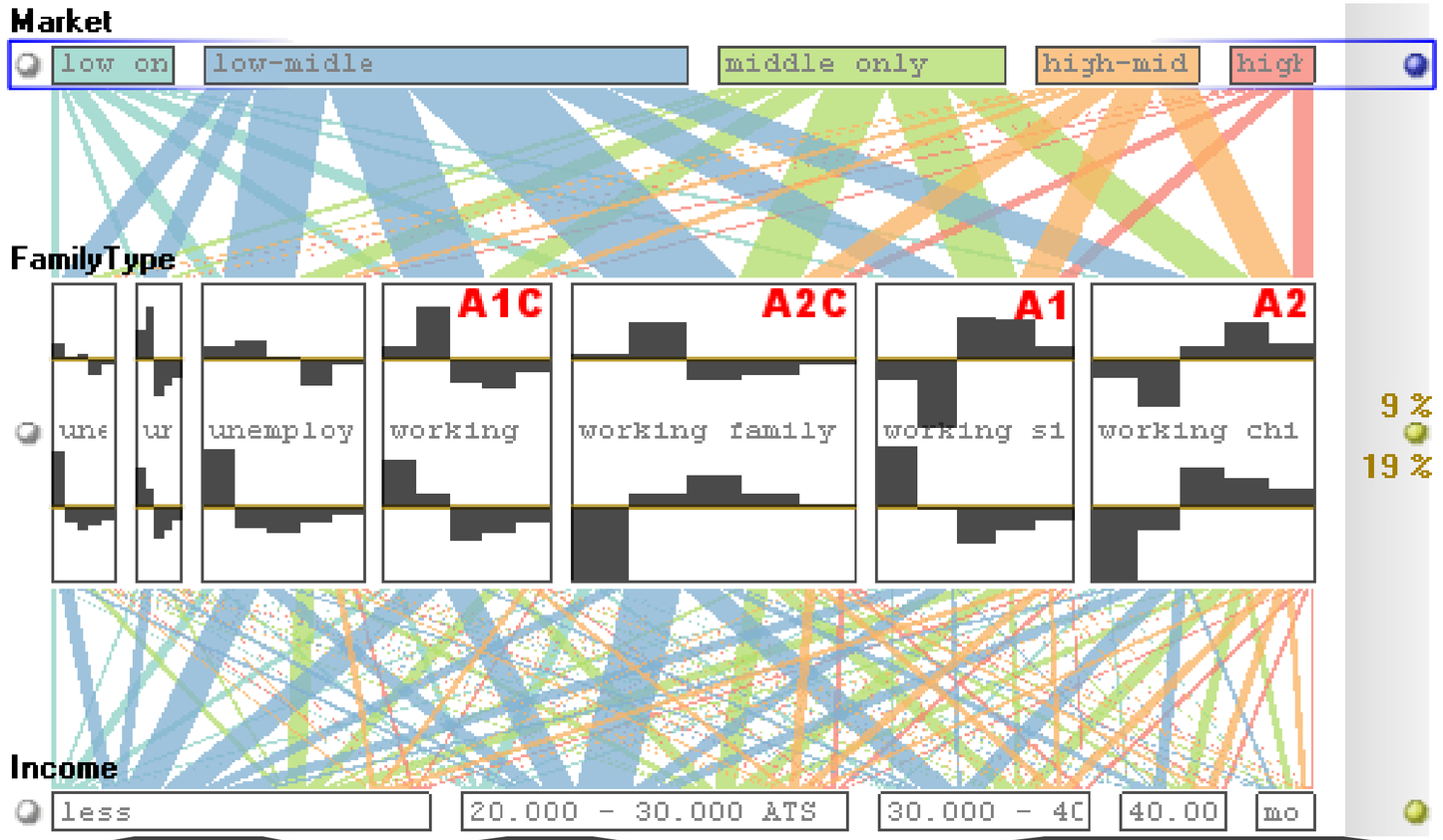
visual encoding

- boxes scaled by frequency
- color coded by values for current active dimension

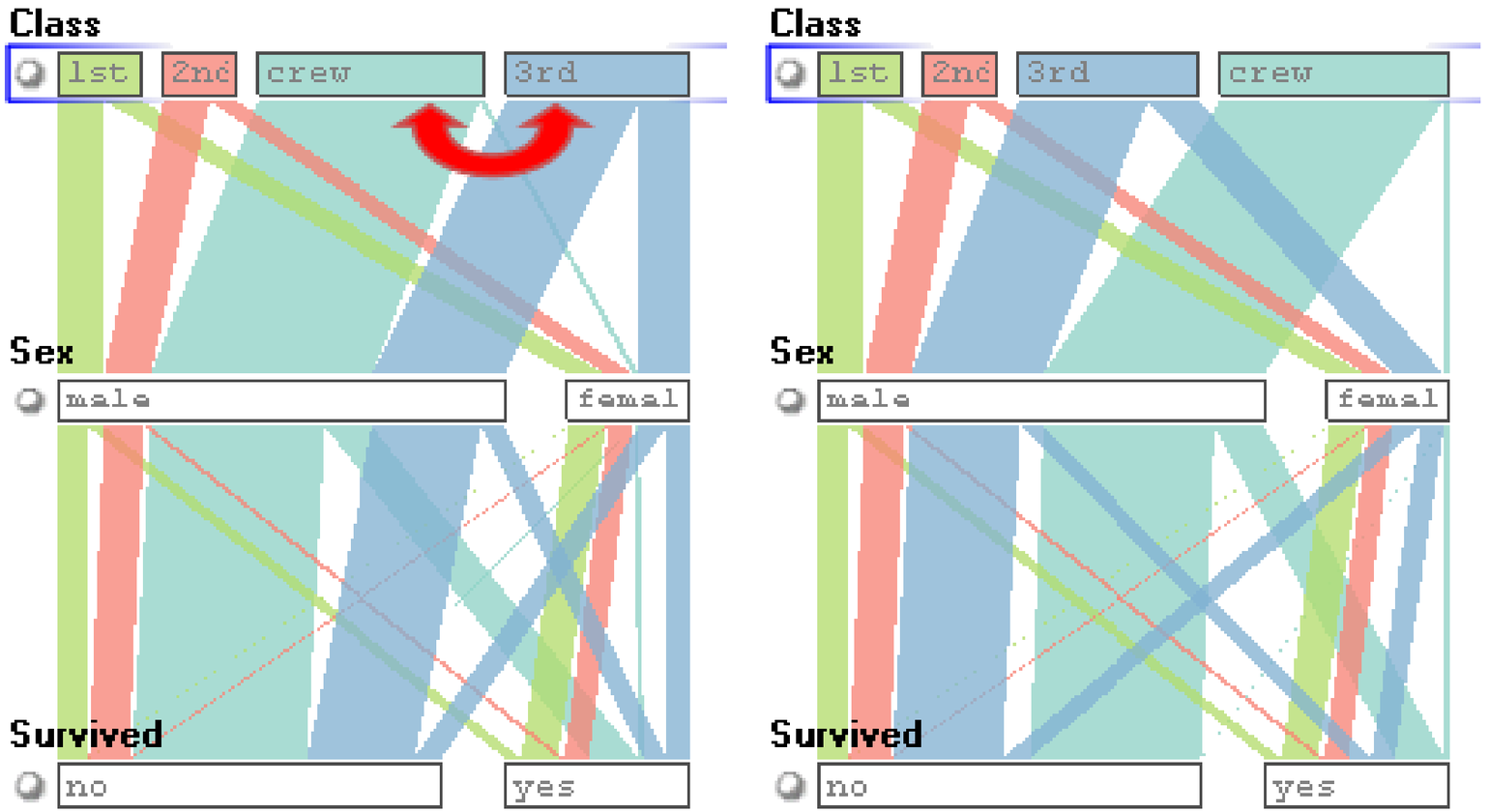


visual encoding

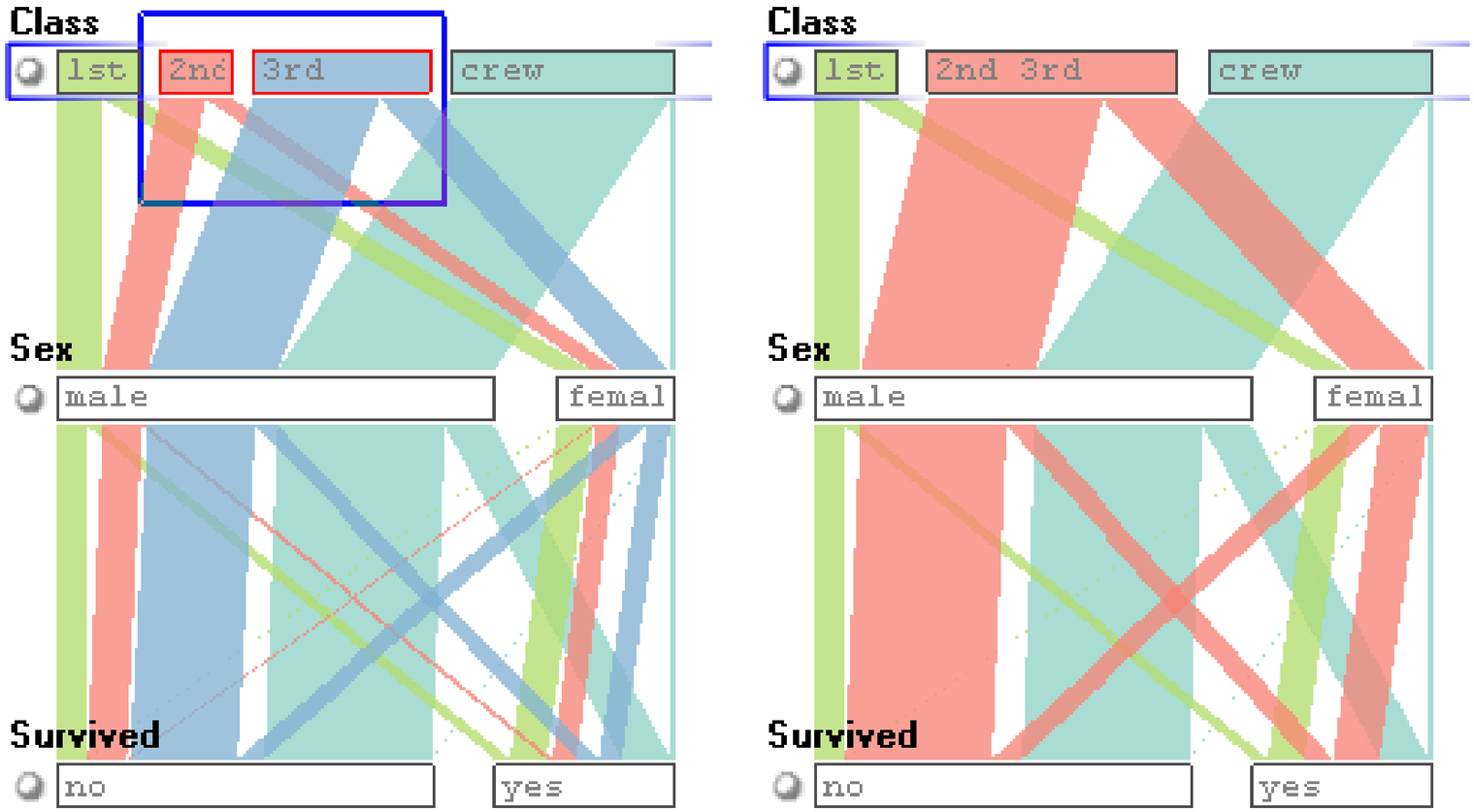
- boxes expand to show histogram



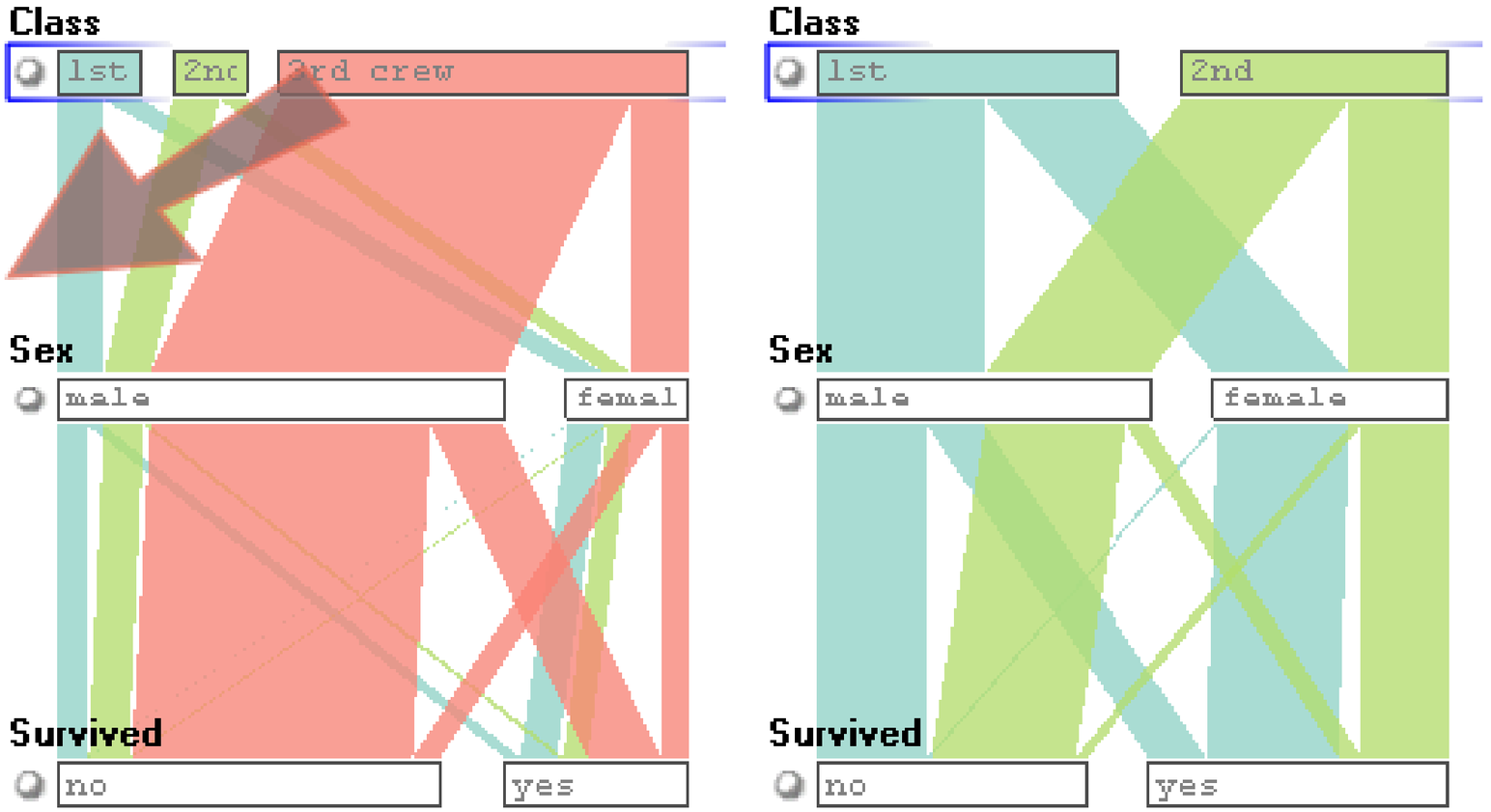
interaction: reorder



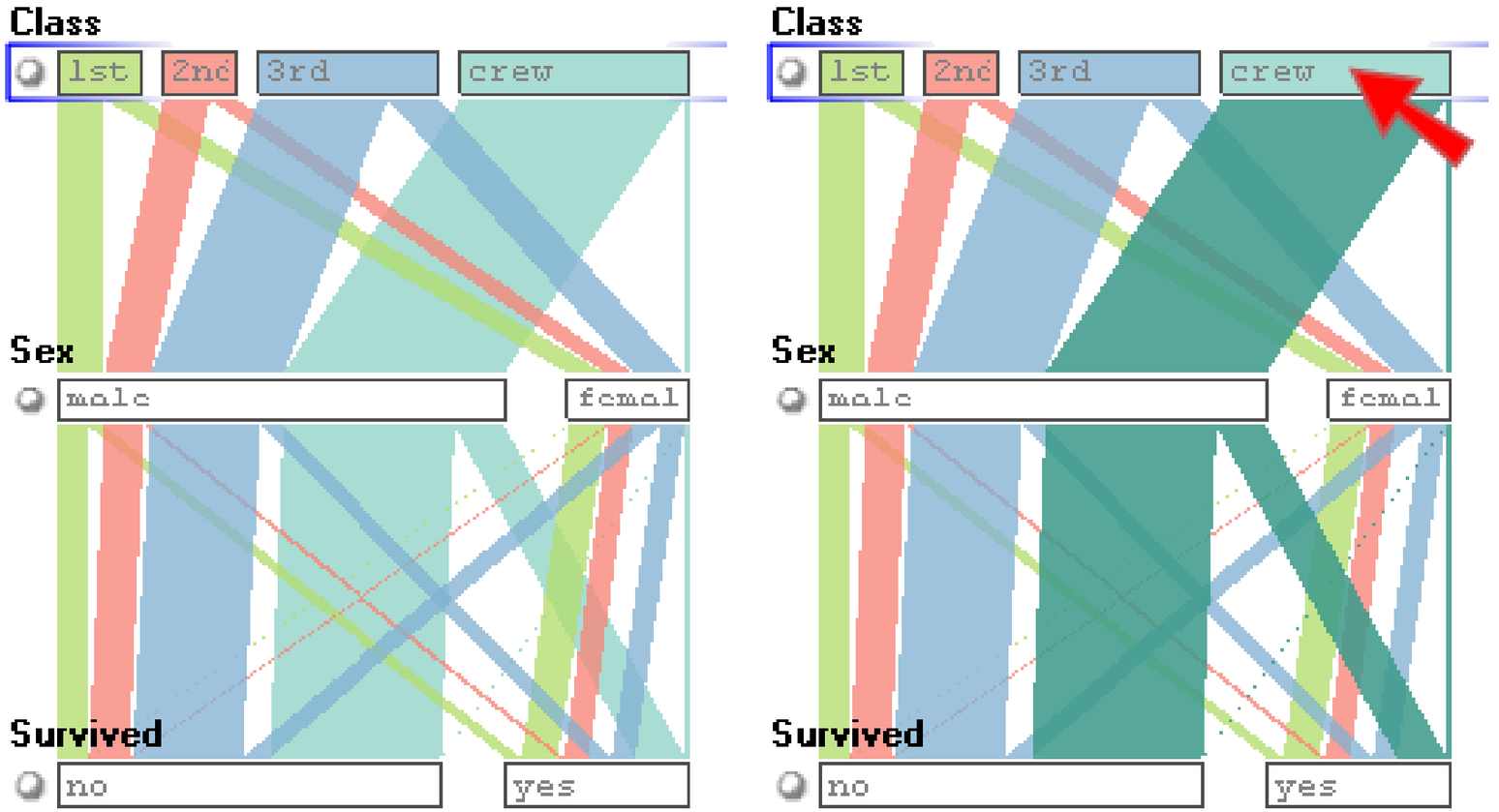
interaction: aggregate



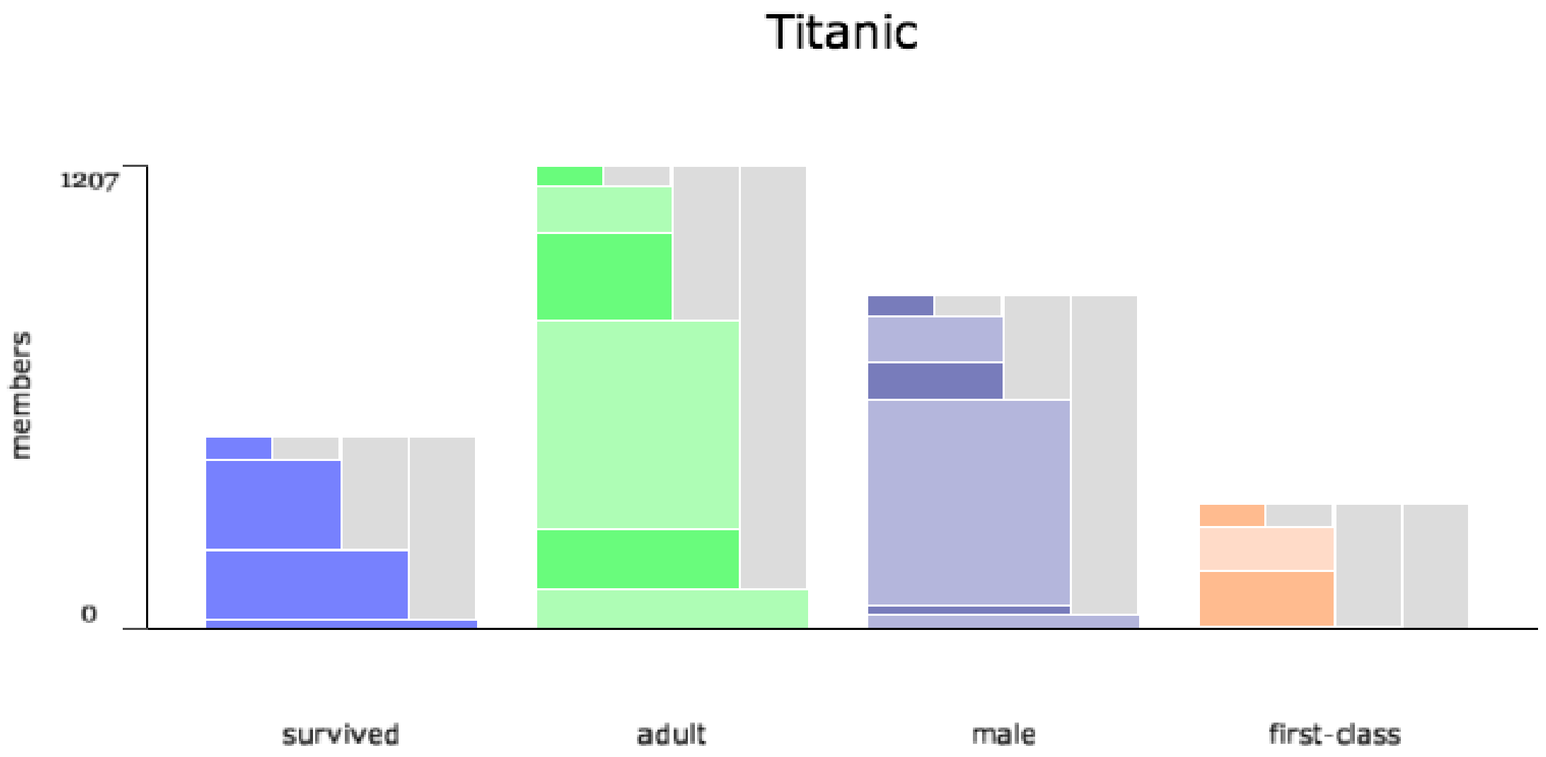
interaction: filter



interaction: highlight

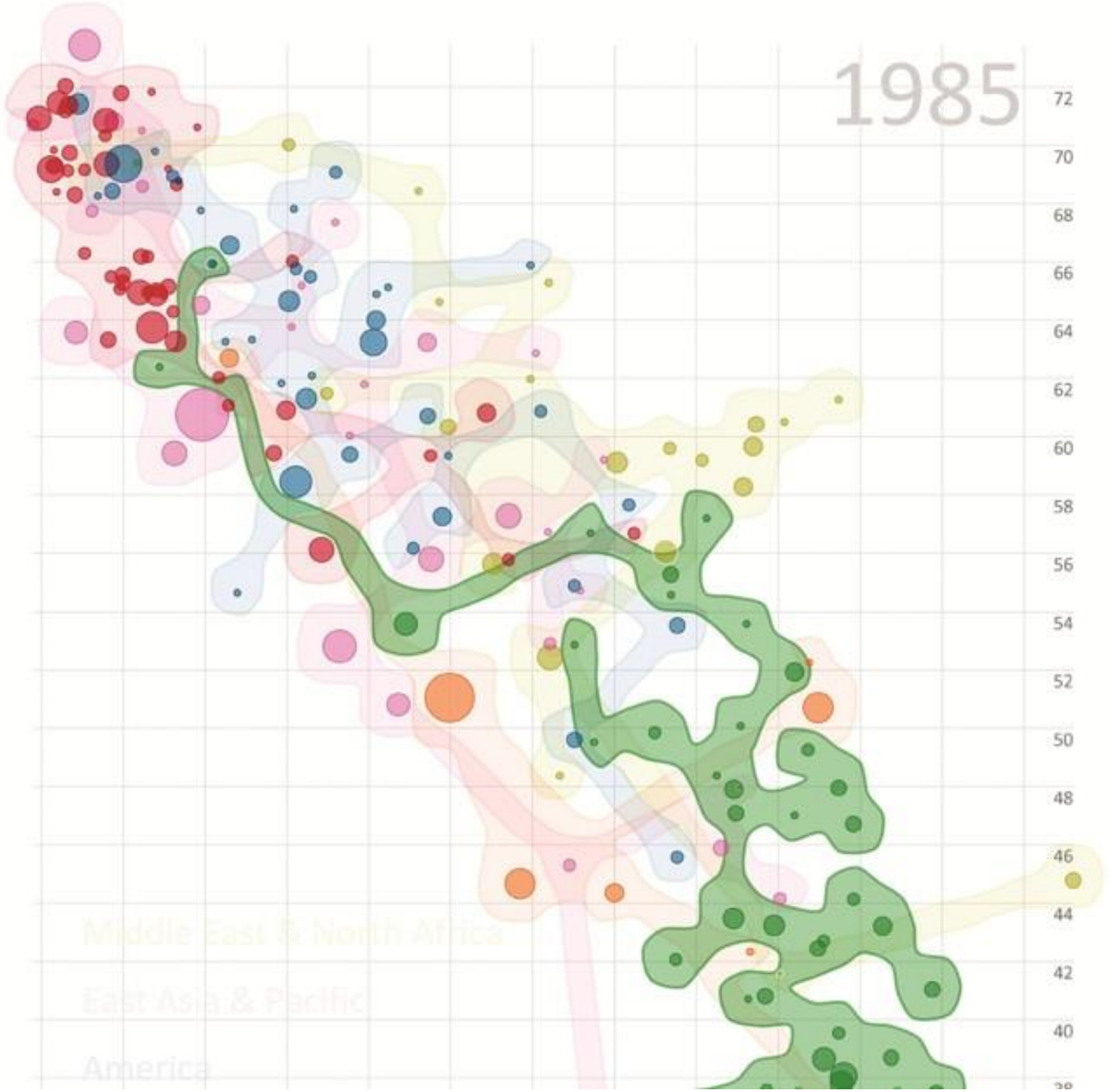


set o’gram



visualizing sets with constraints

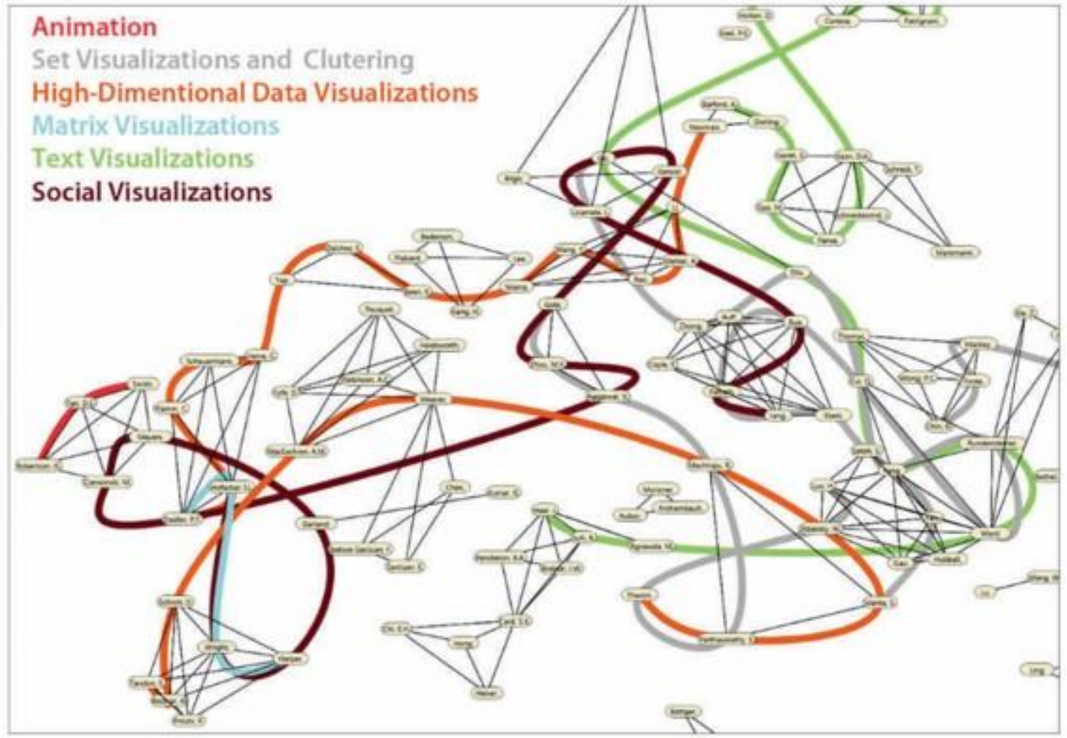
bubble sets


1985
72
70
68
66
64
62
60
58
56
54
52
50
48
46
44
42
40
Middle East & North Africa
East Asia & Pacific
America

line sets

restaurants

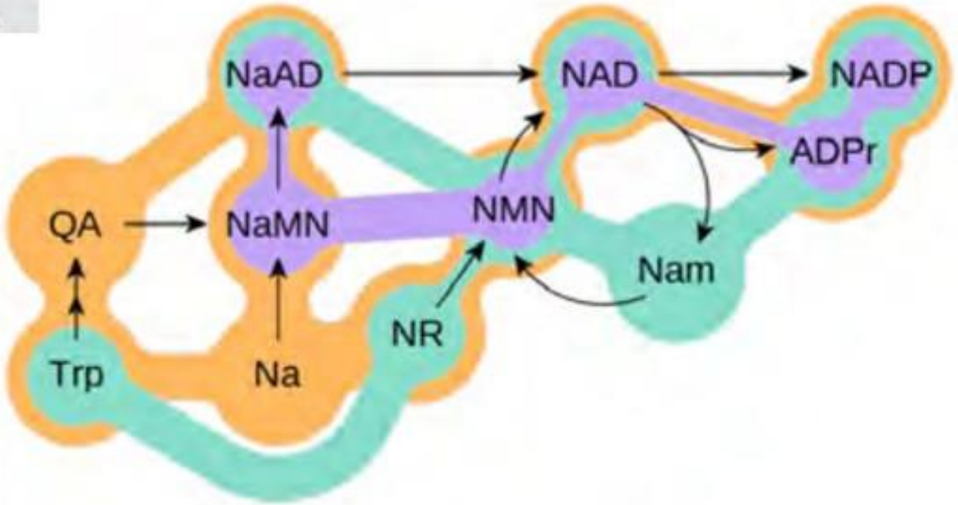
social communities



kelp diagrams

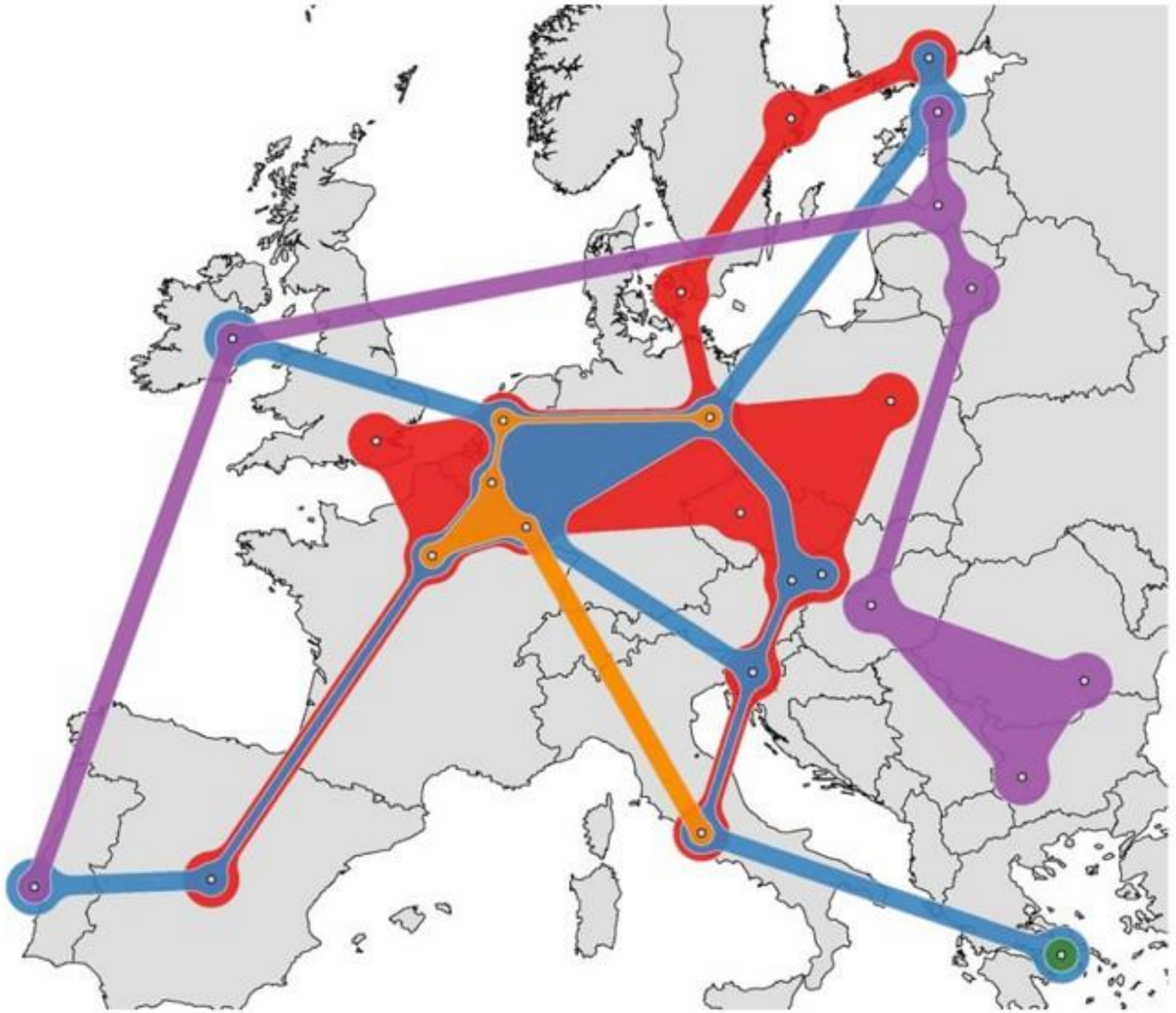
cities on a map

metabolic network



kelp fusion

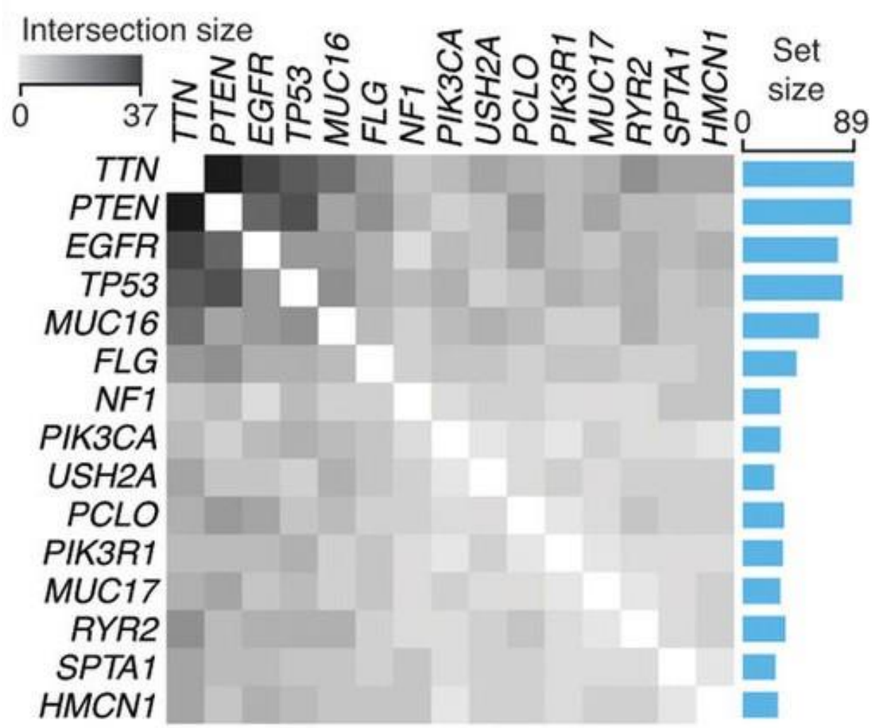
- cities on map
- lines & areas



Showing Pairwise Overlap

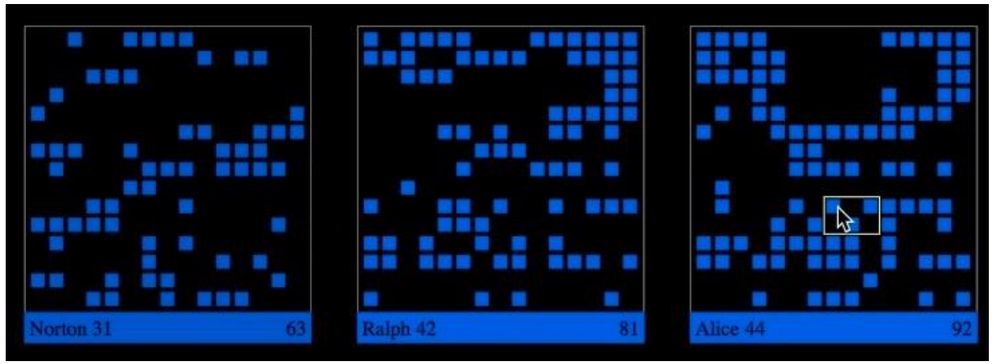
- Doesn’t show higher-order overlaps
- Very scalable
- Can’t show attributes

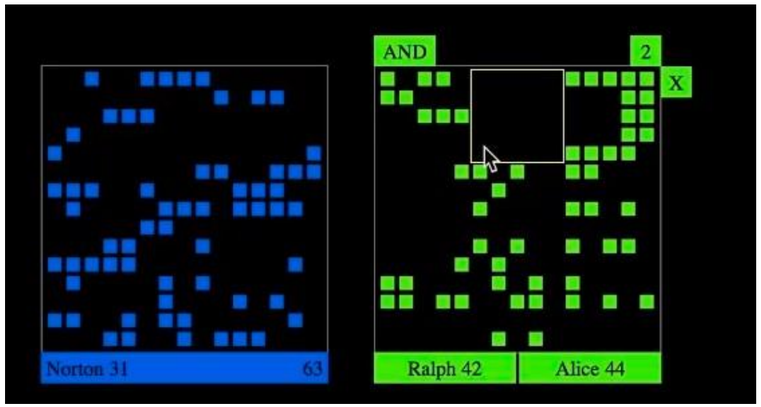
Co-Mutations of genes



Set Matrices: OnSet

- Set membership for each item shown in matrix
- Comparisons can be made using AND or OR operations
- Good for many sets and few items





https://vimeo.com/213029678#at=0

[Sadana 14]

Linear Diagrams

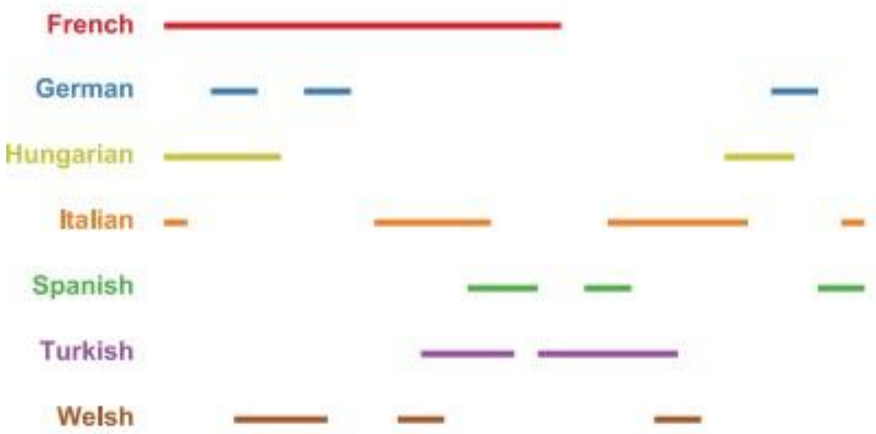


Fig. 1. Visualizing sets: linear diagrams.

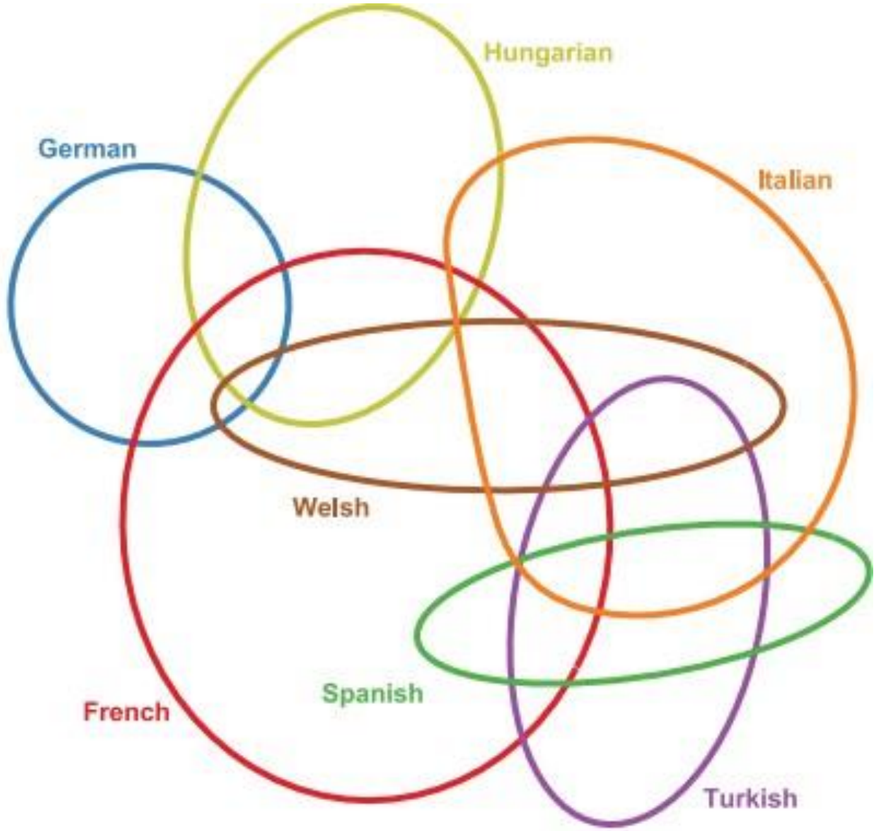
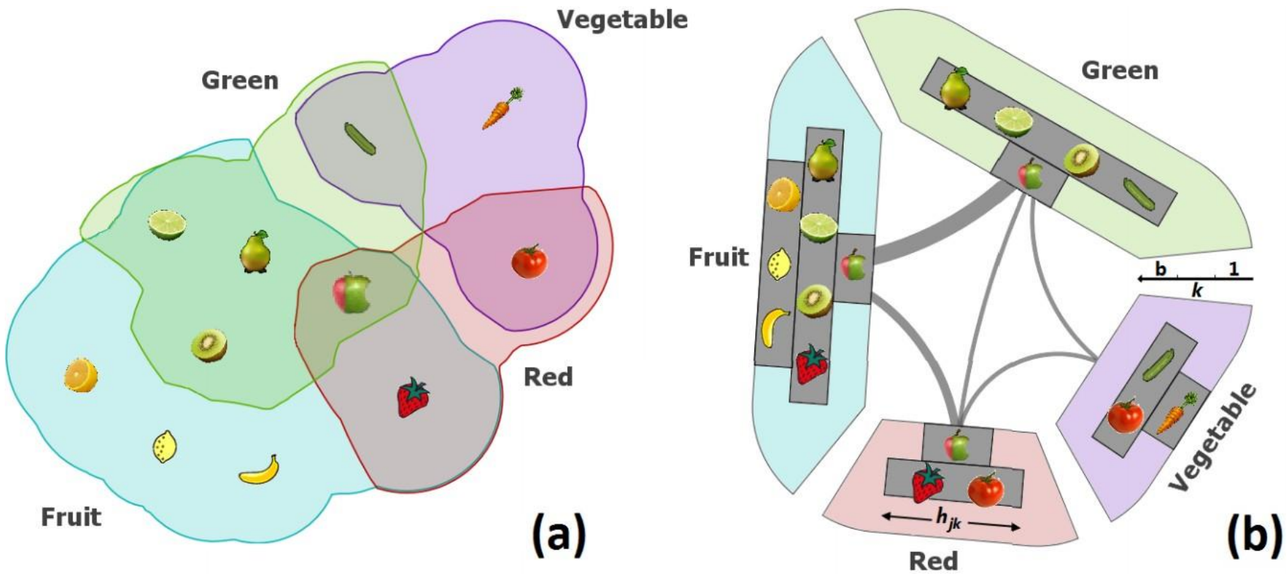


Fig. 2. Visualizing sets: Euler diagrams.

[RODGERS 2015]

Radial Sets

- Sets are segments on a “circle”
- Relationships are encoded as ribbons
- Size of segments encodes size of sets
- Histograms in segments show degrees



https://www.youtube.com/watch?v=UcYRrPqC5A8

[Alsallakh 2013]

UpSet: Visualization of Intersecting Sets

Alexander Lex, Nils Gehlenborg, Hendrik Strobelt,
Romain Vuillemot, and Hanspeter Pfister

http://vcglab.org/upset

HARVARD
School of Engineering
and Applied Sciences

Sets

- applies to many datasets
  - Many categorical data can be viewed as sets
- many combinations may be interesting
- limited numbers of sets more tractable