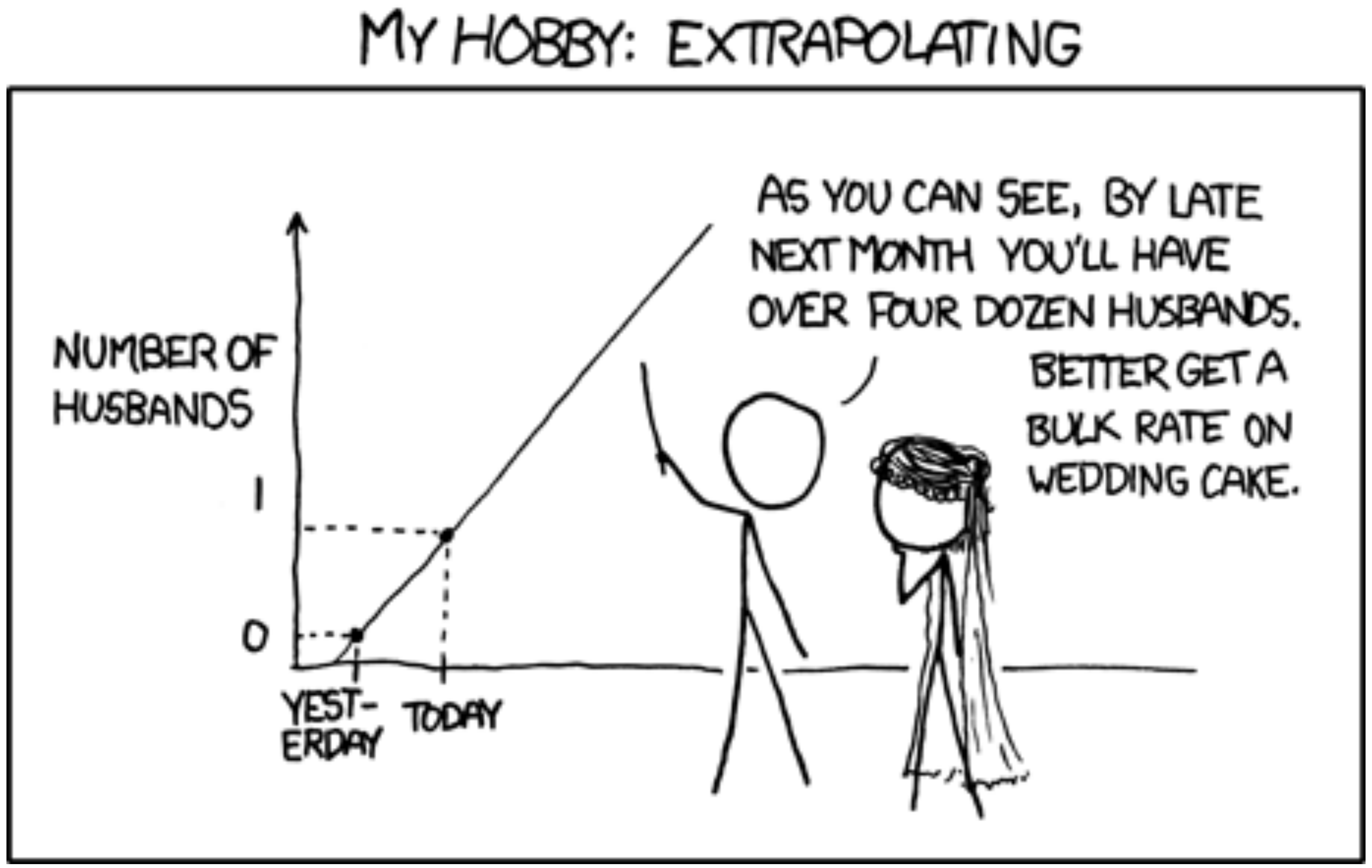
CS-5630 / CS-6630 Visualization
Tables

Alexander Lex
alex@sci.utah.edu



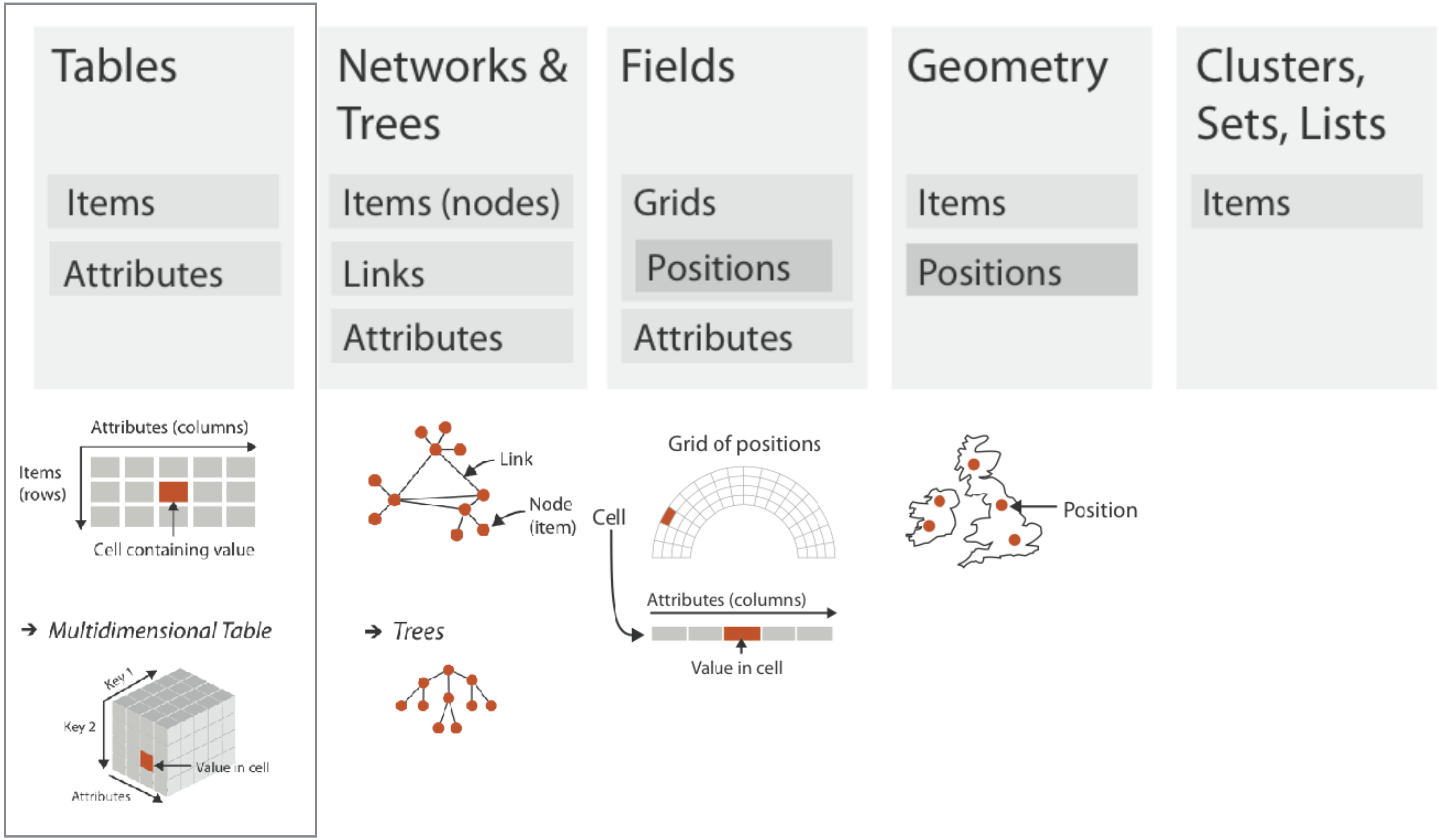
[xkcd]

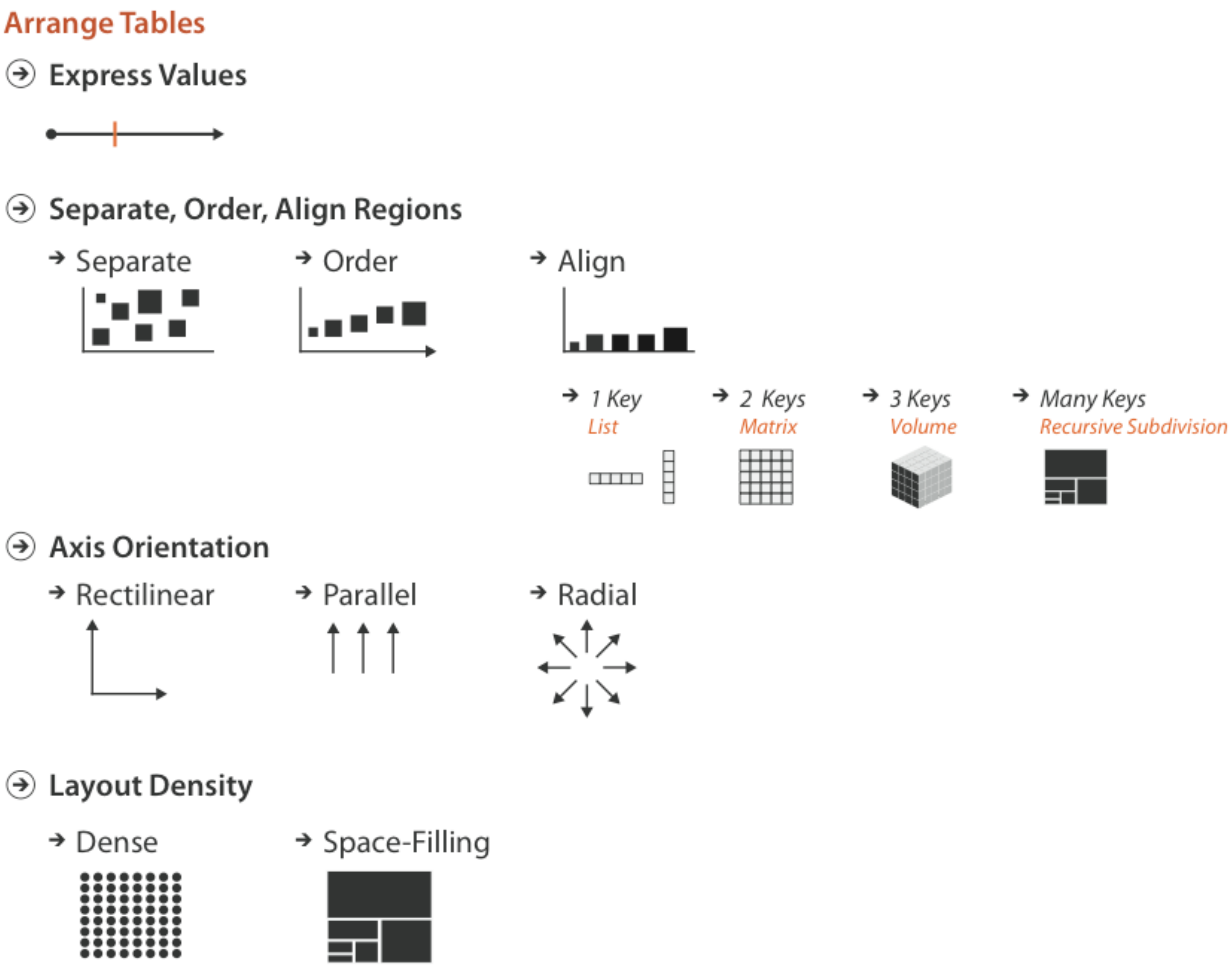
Organizational

Contacted by TA this week for feedback on project

No more standing office hours - arrange meetings

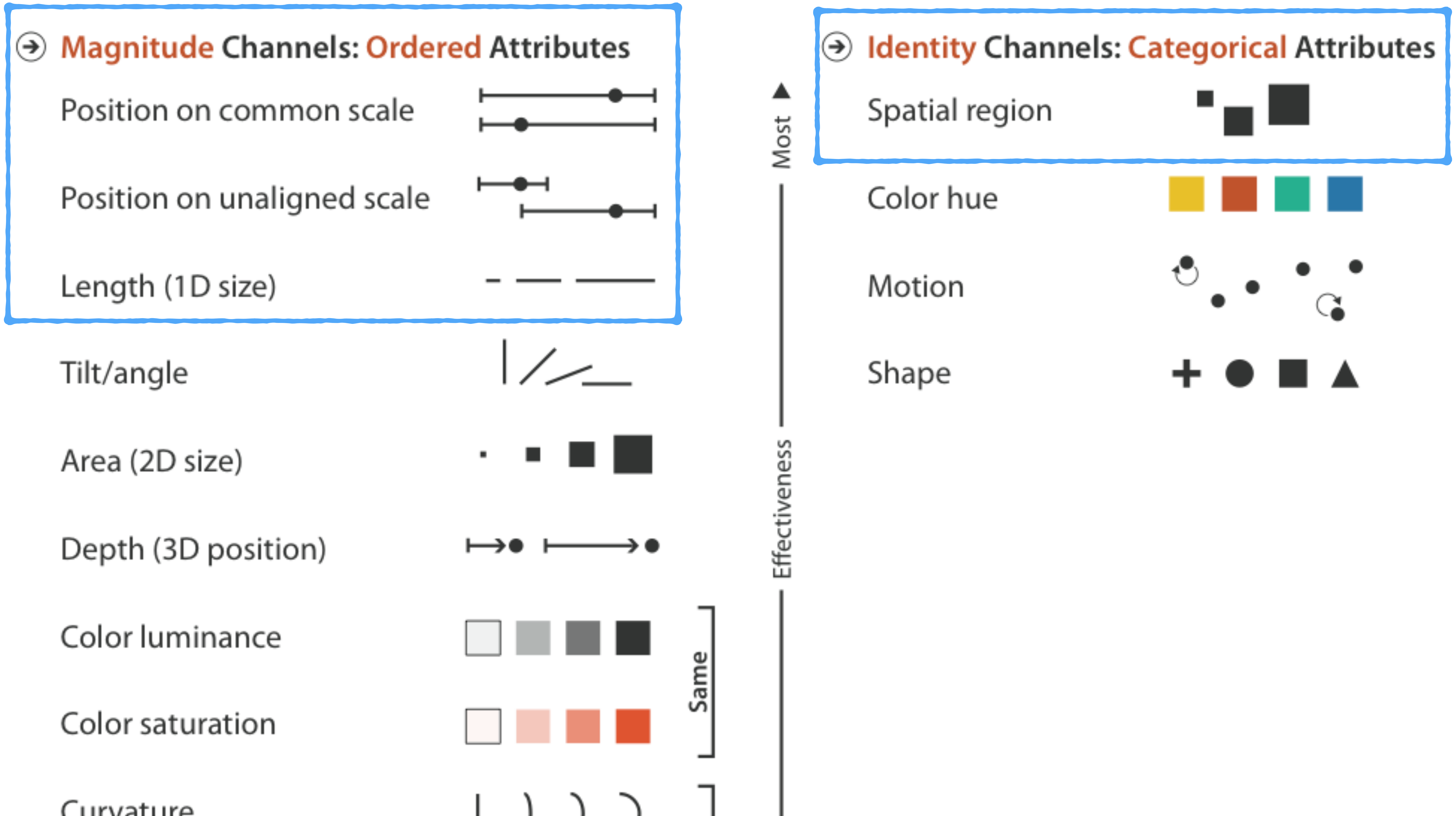
dataset types





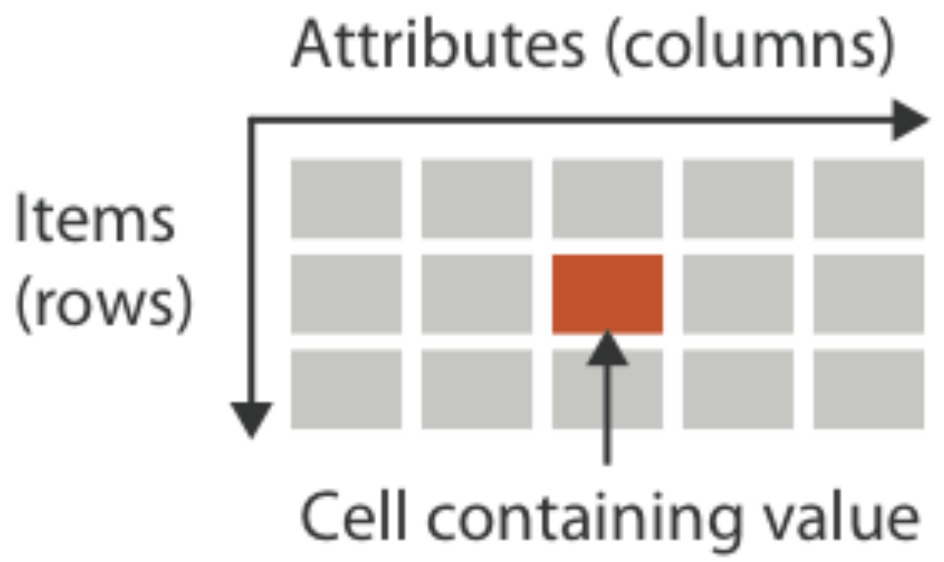
Arrange Tables
Express Values
Separate, Order, Align Regions
Separate
Order
Align
1 Key
List
2 Keys
Matrix
3 Keys
Volume
Many Keys
Recursive Subdivision
Axis Orientation
Rectilinear
Parallel
Radial
Layout Density
Dense
Space-Filling

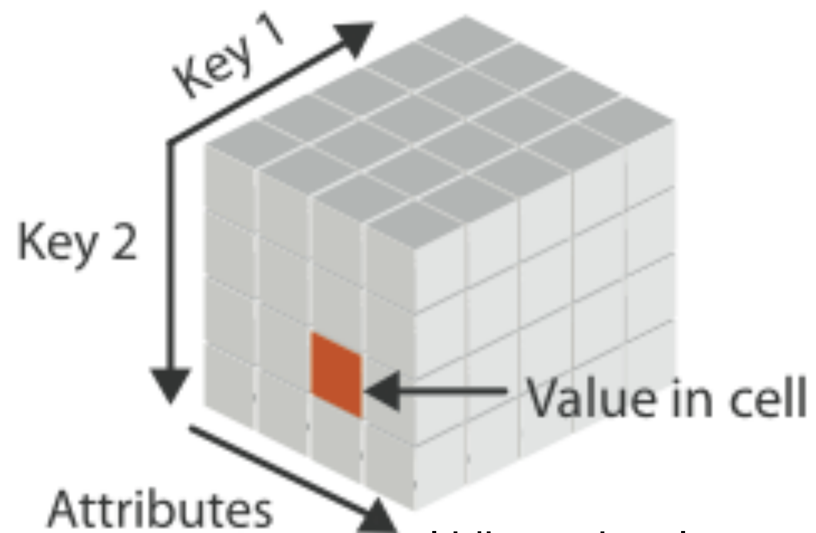
spatial channels are the most effective for all attribute types



recall: attribute semantics

when we arrange tabular data, attributes are chosen to be keys and values





multidimensional

Scale of Tables

Need different approaches for “normal” and “high-dimensional” tables.

How many dimensions?

~50 – tractable with “just” vis

~1000 – need analytical methods

How many records?

~ 1000 – “just” vis is fine

>> 10,000 – need analytical methods

Homogeneity

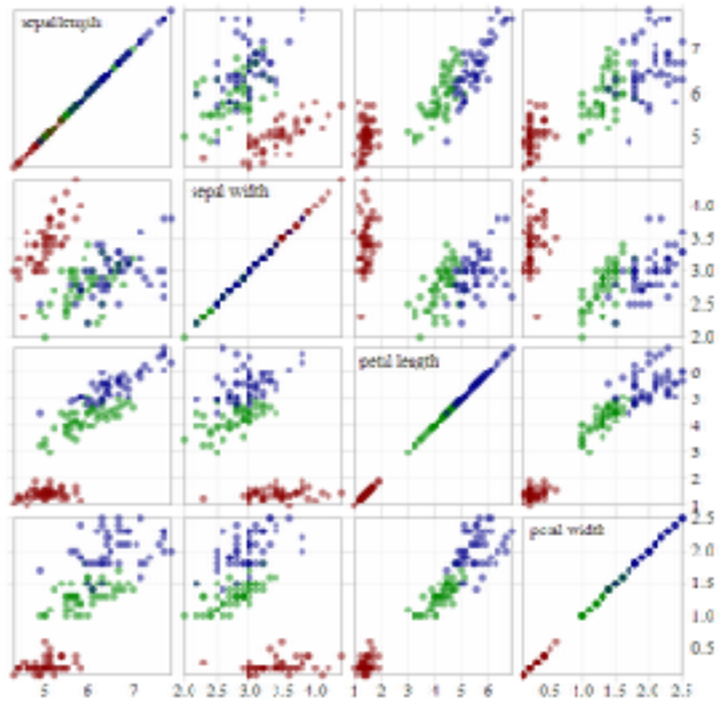
Same data type?

Same scales?

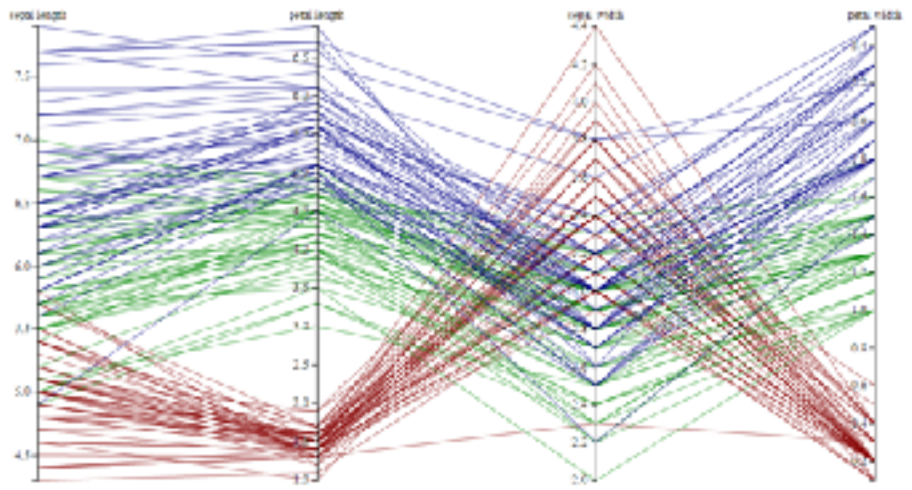
	Age	Gender	Height
Bob	*25*	*M*	*181*
Alice	*22*	*F*	*185*
Chris	*19*	*M*	*175*

	BPM 1	BPM 2	BPM 3
Bob	*65*	*120*	*145*
Alice	*80*	*135*	*185*
Chris	*45*	*115*	*135*

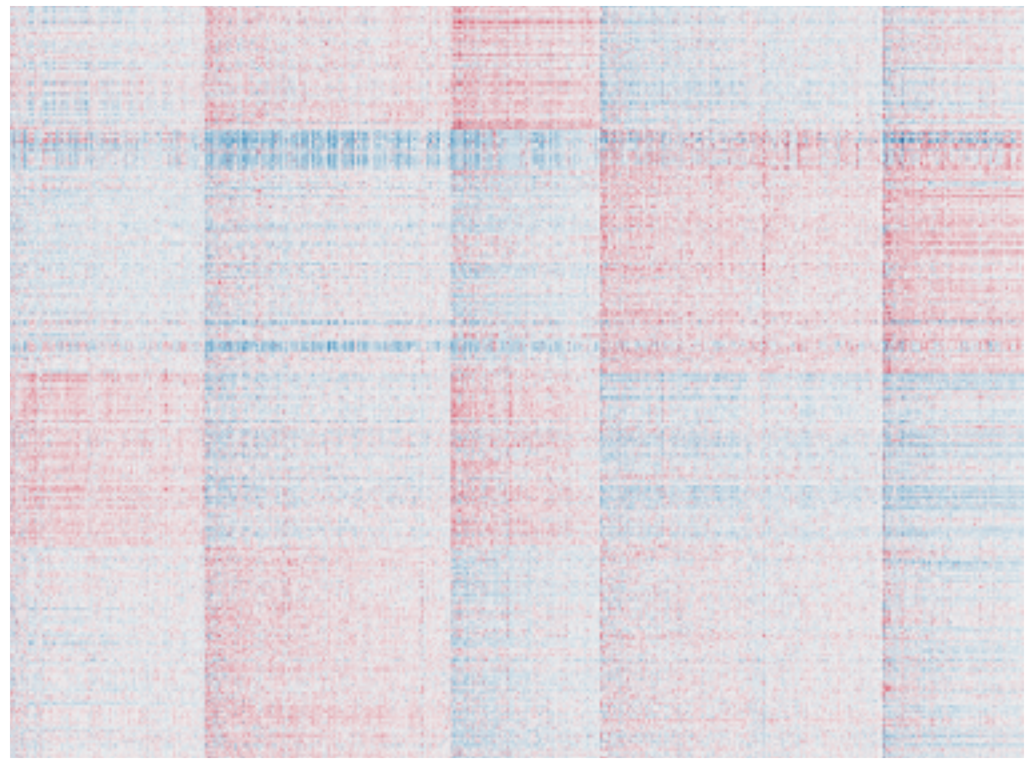
Analytic Component



Scatterplot Matrices
[Bostock]

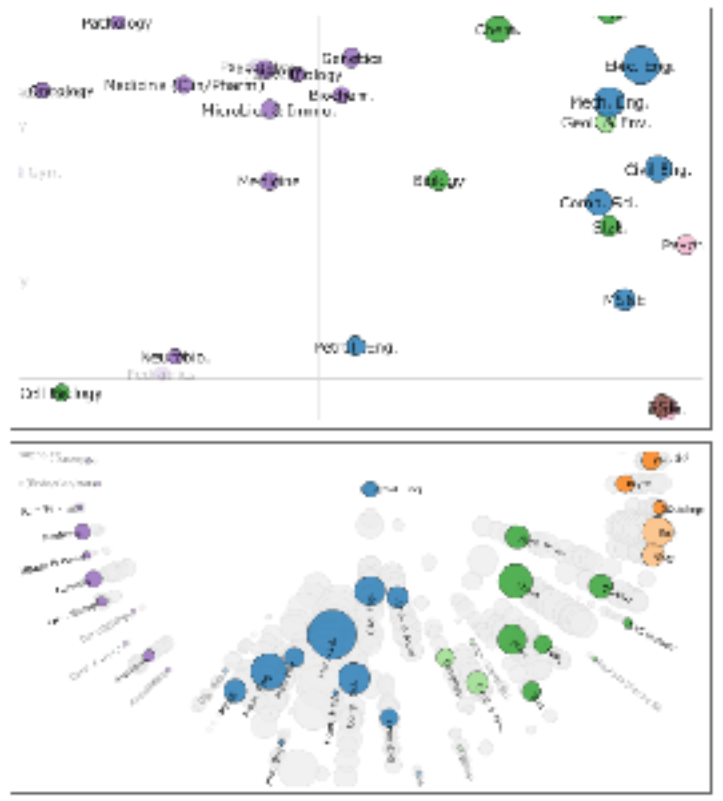


Parallel Coordinates
[Bostock]



Pixel-based visualizations / heat maps

Multidimensional Scaling
[Doerk 2011]



[Chuang 2012]

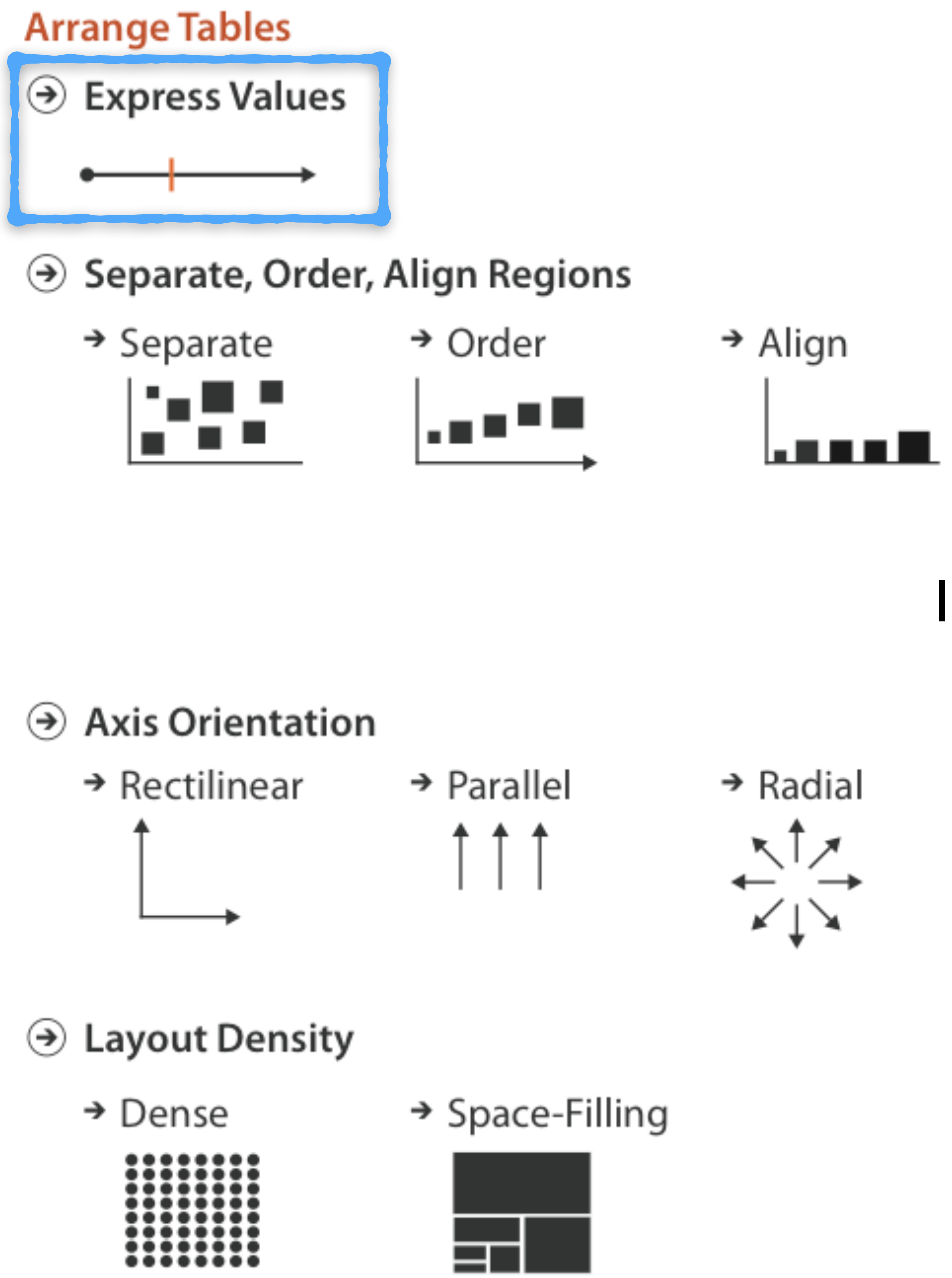
no / little analytics

strong analytics component

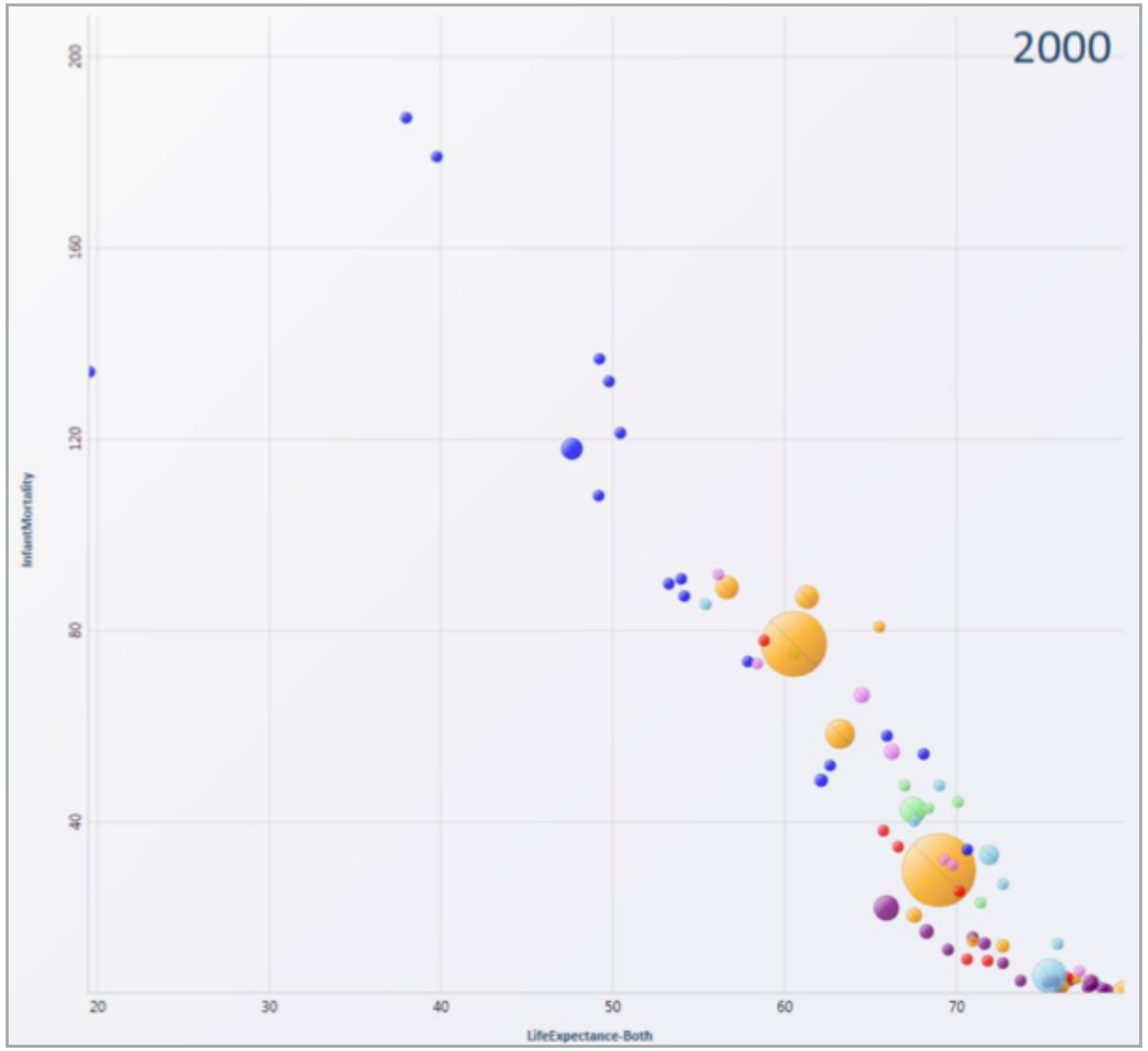
Express Values

No Keys

encode using zero keys: scatterplots



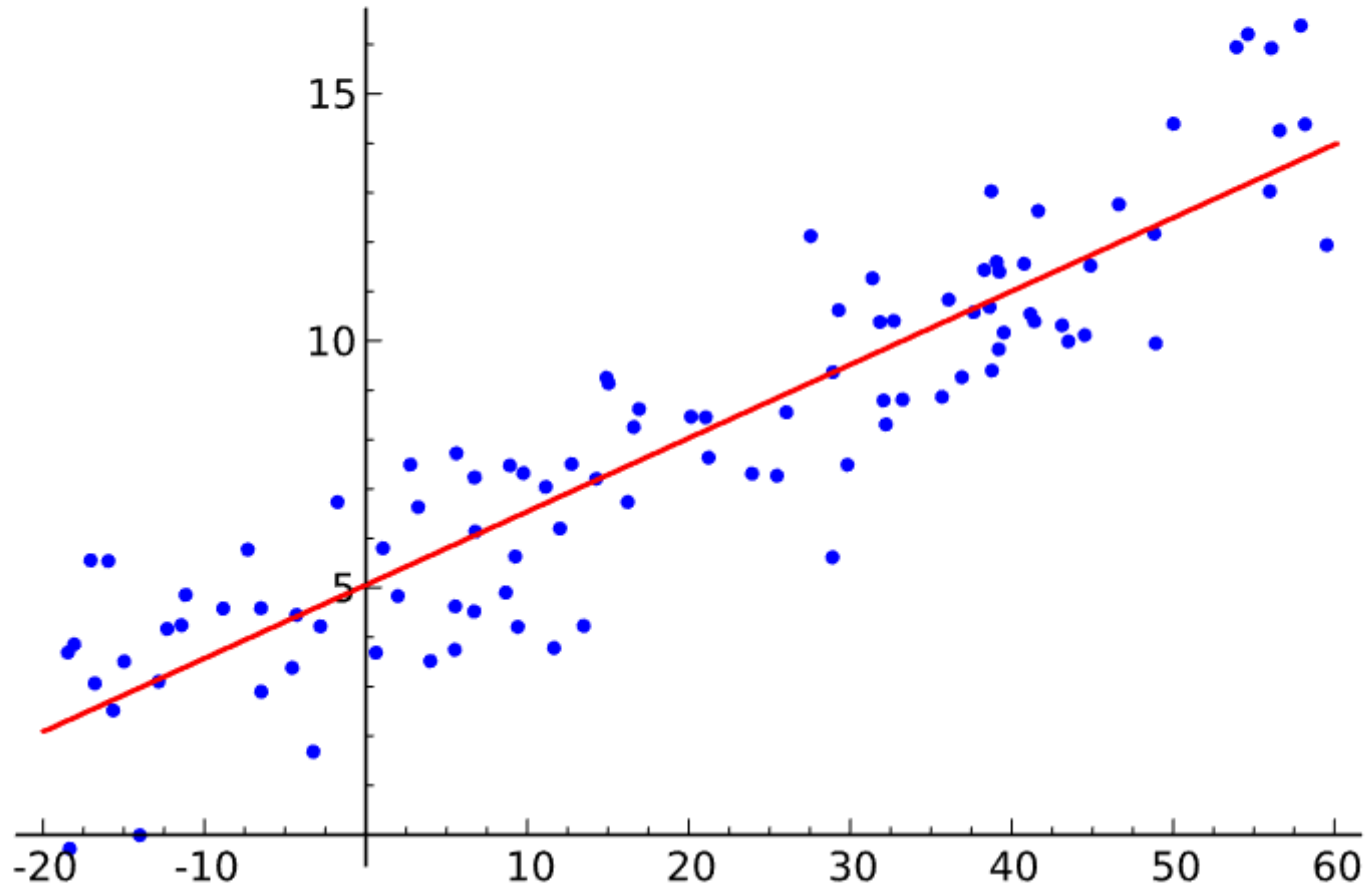
Infant Mortality



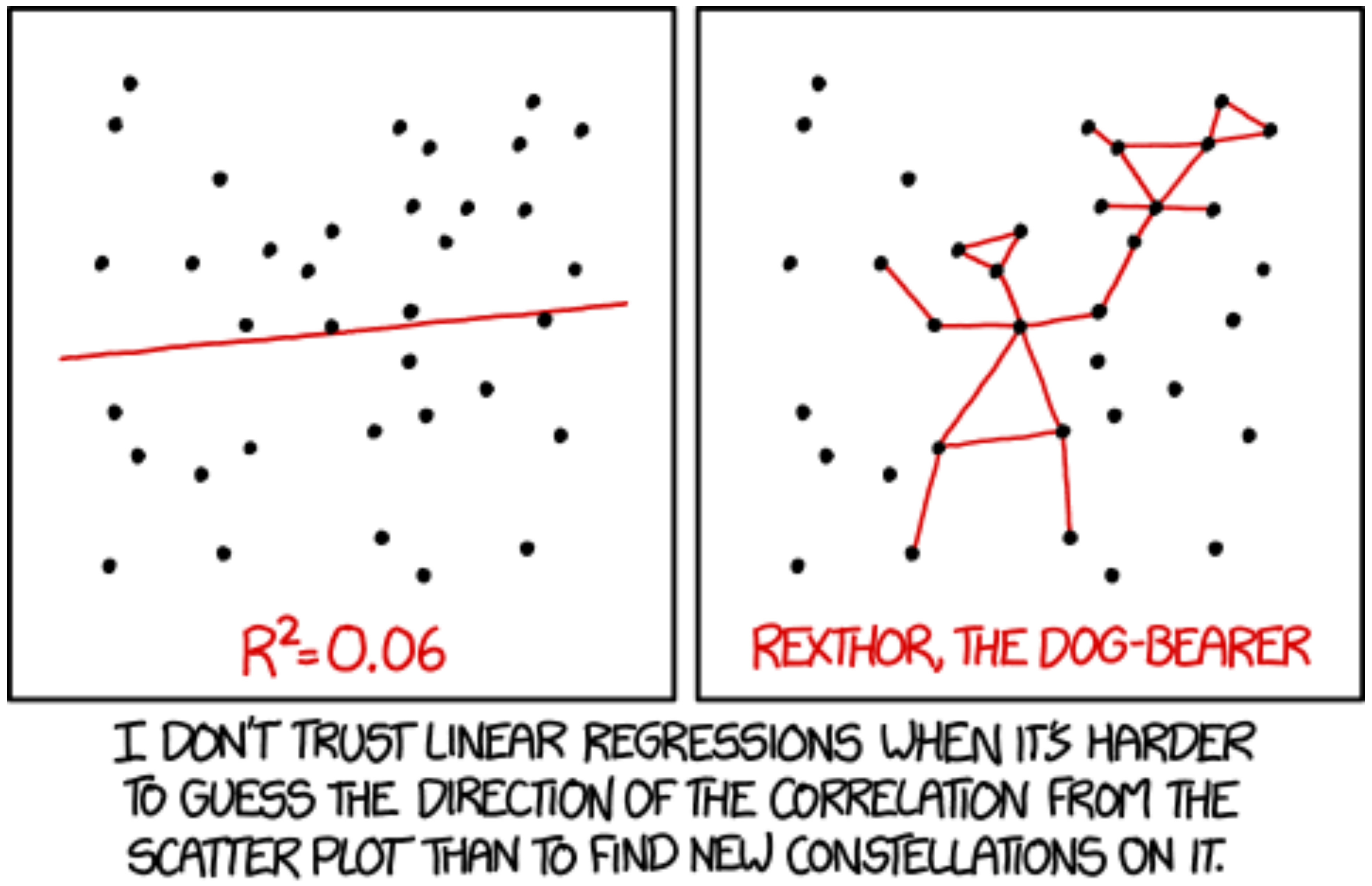
Life Expectance

Regression Lines

$$y \sim \beta_0 + \beta_1 x$$

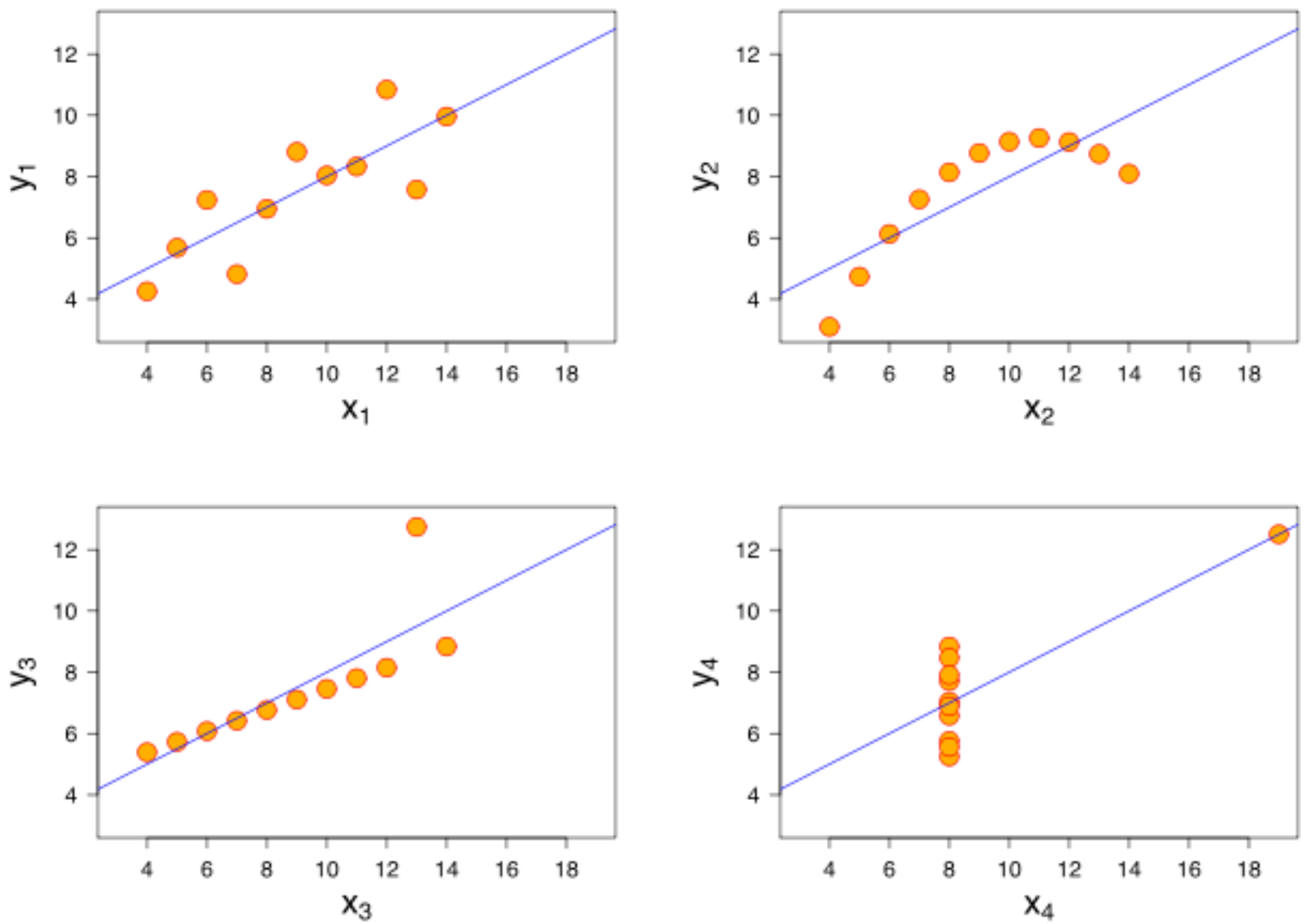


Goal: Find the best values of β_0 and β_1, denoted $\hat{\beta}_0$ and $\hat{\beta}_1$, so that the prediction $y = \hat{\beta}_0 + \hat{\beta}_1 x$ "best fits" the data.



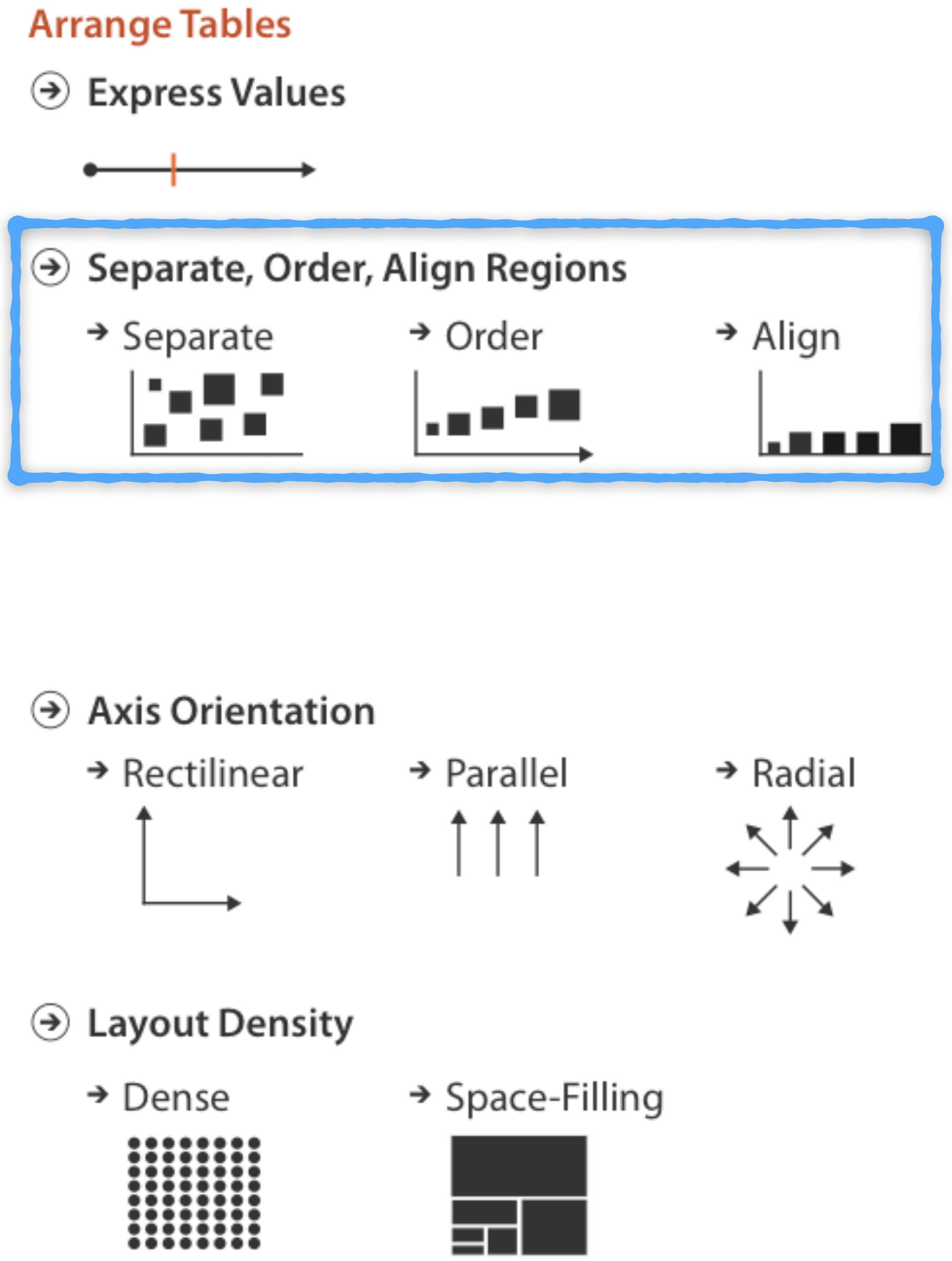
Approach: use least squares to minimize the sum of the squares of the errors

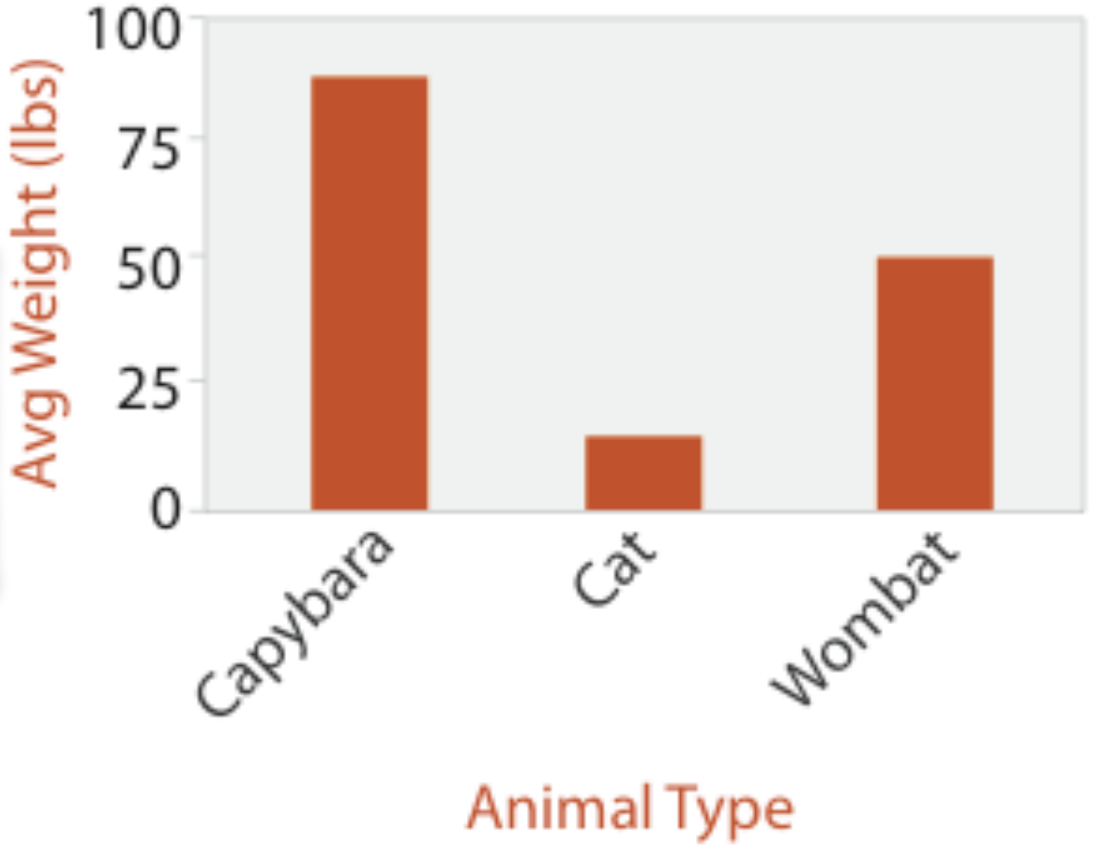
Anscombe's Quartet

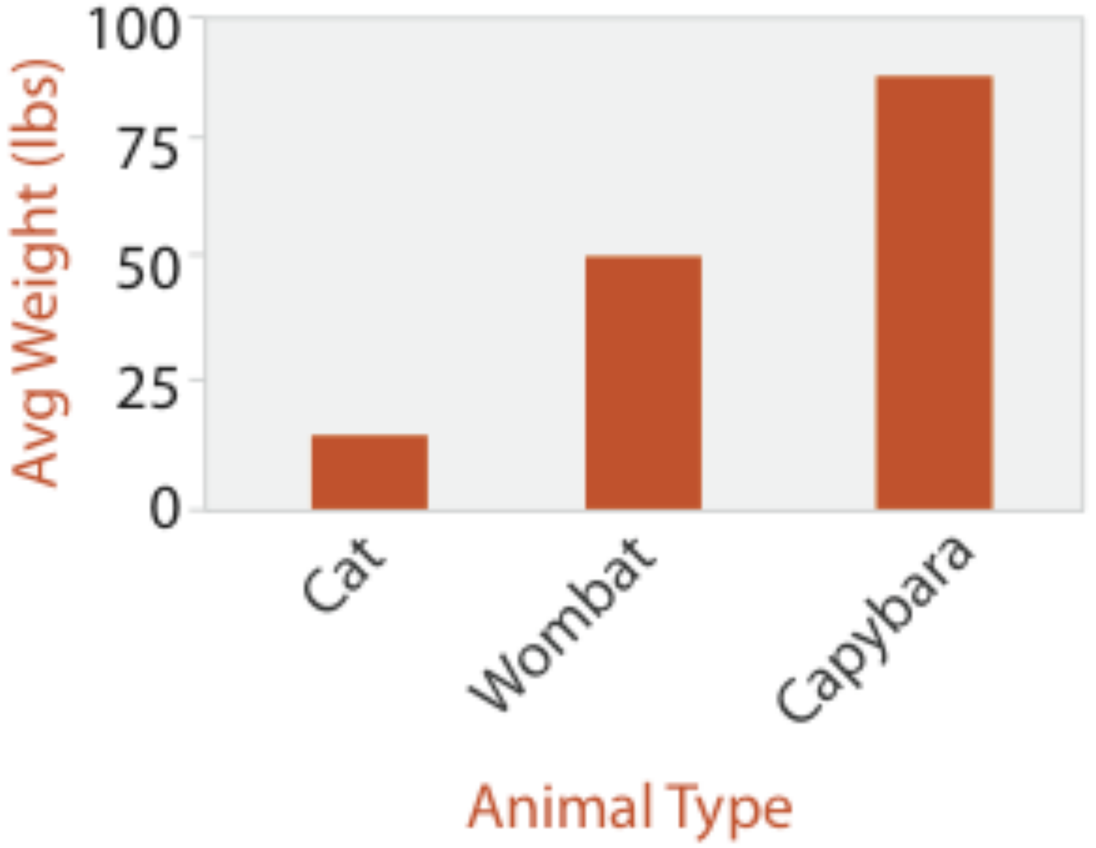


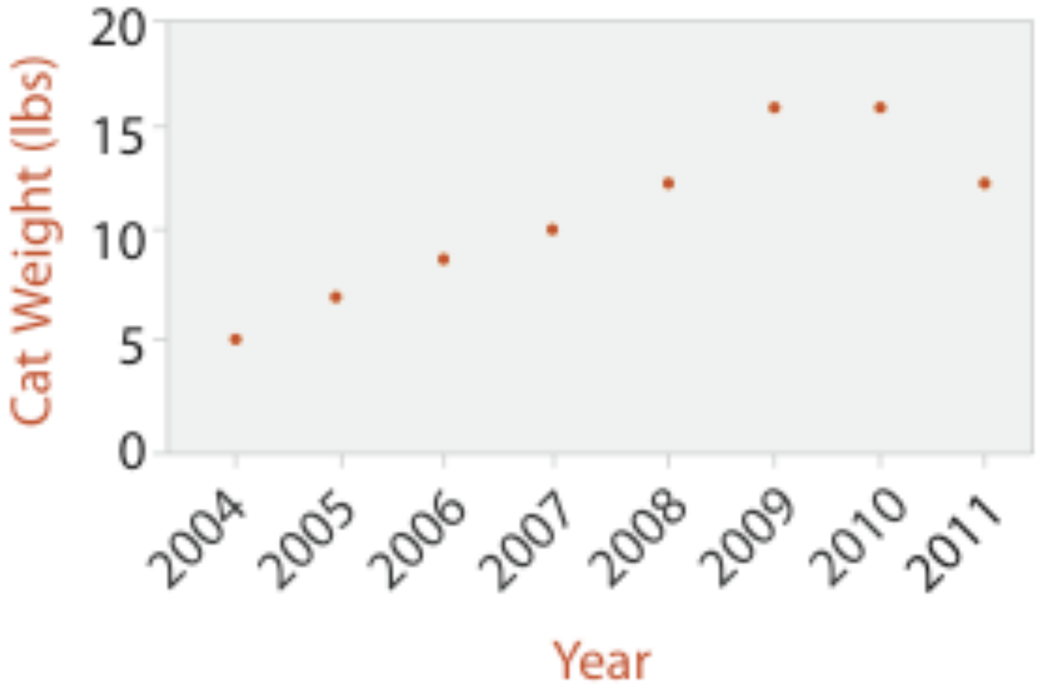
Encode one Key Attribute

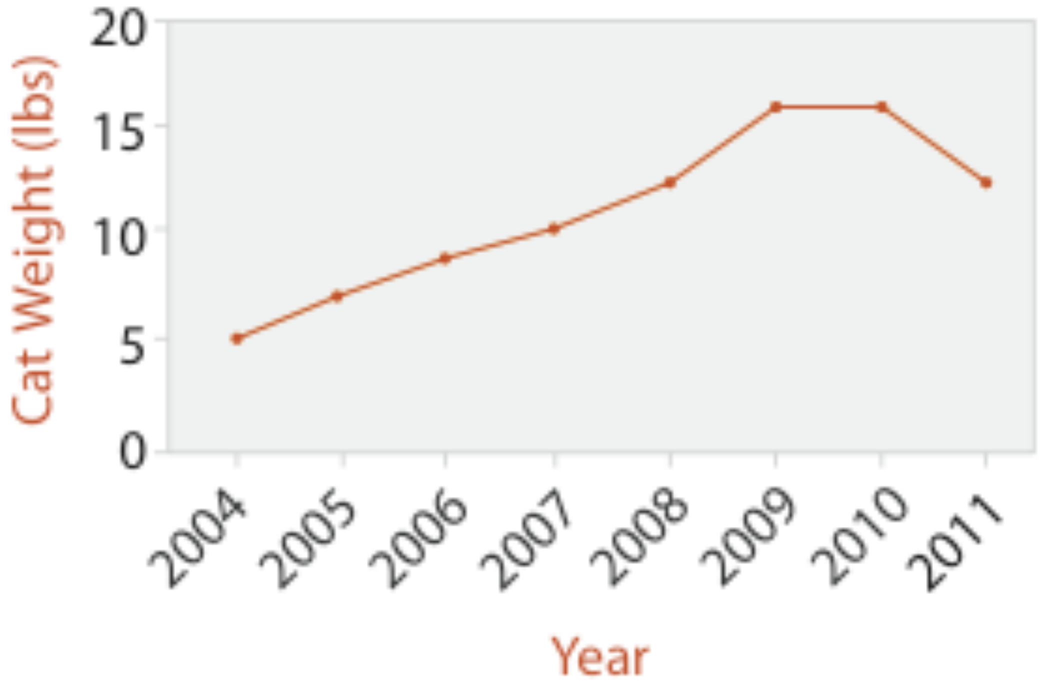
encode one key attribute: bar, dot, & line charts



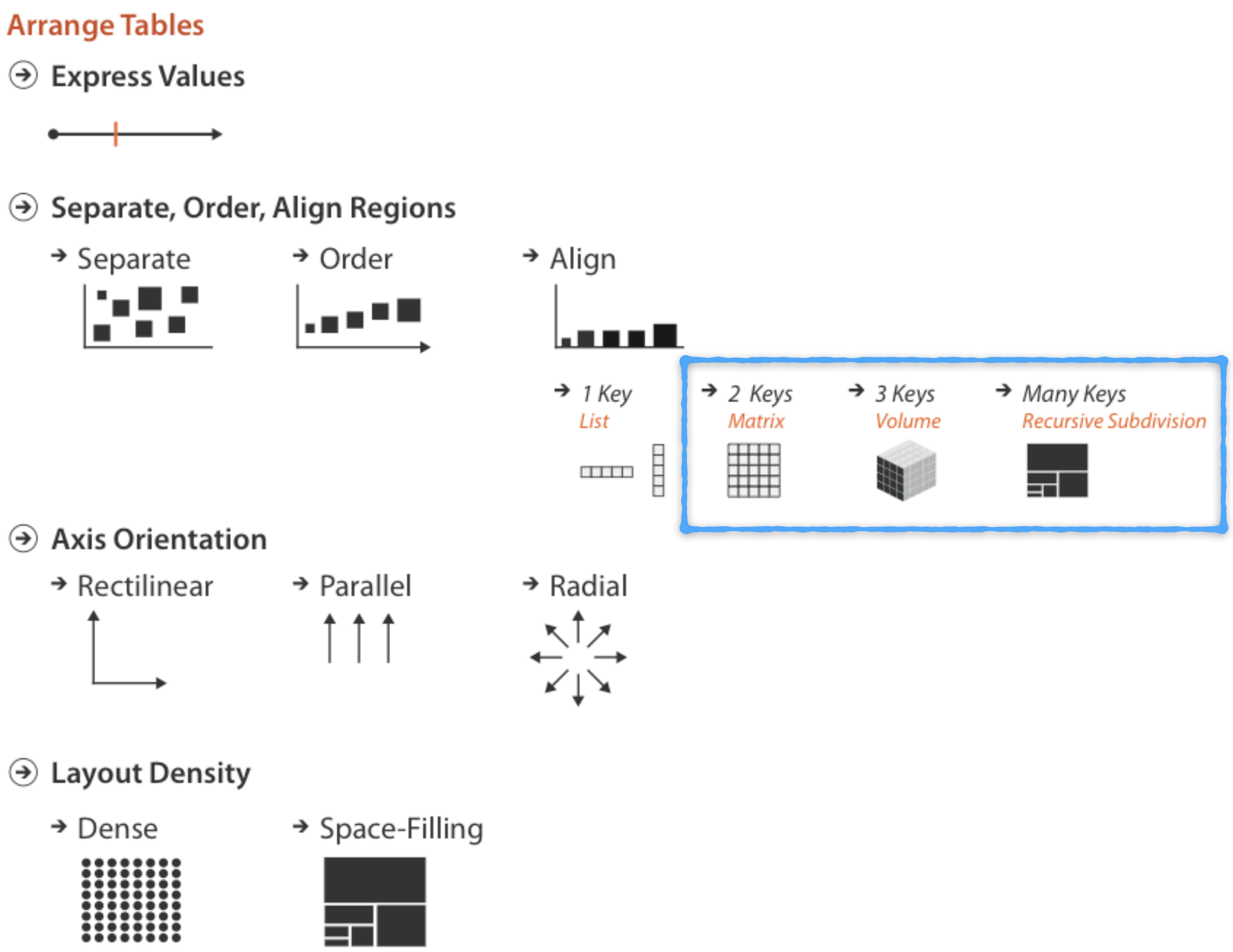








Encode Multiple Key Attributes


Arrange Tables
Express Values
Separate, Order, Align Regions
Separate
Order
Align
1 Key
List
2 Keys
Matrix
3 Keys
Volume
Many Keys
Recursive Subdivision
Axis Orientation
Rectilinear
Parallel
Radial
Layout Density
Dense
Space-Filling

Stacked Bar Chart

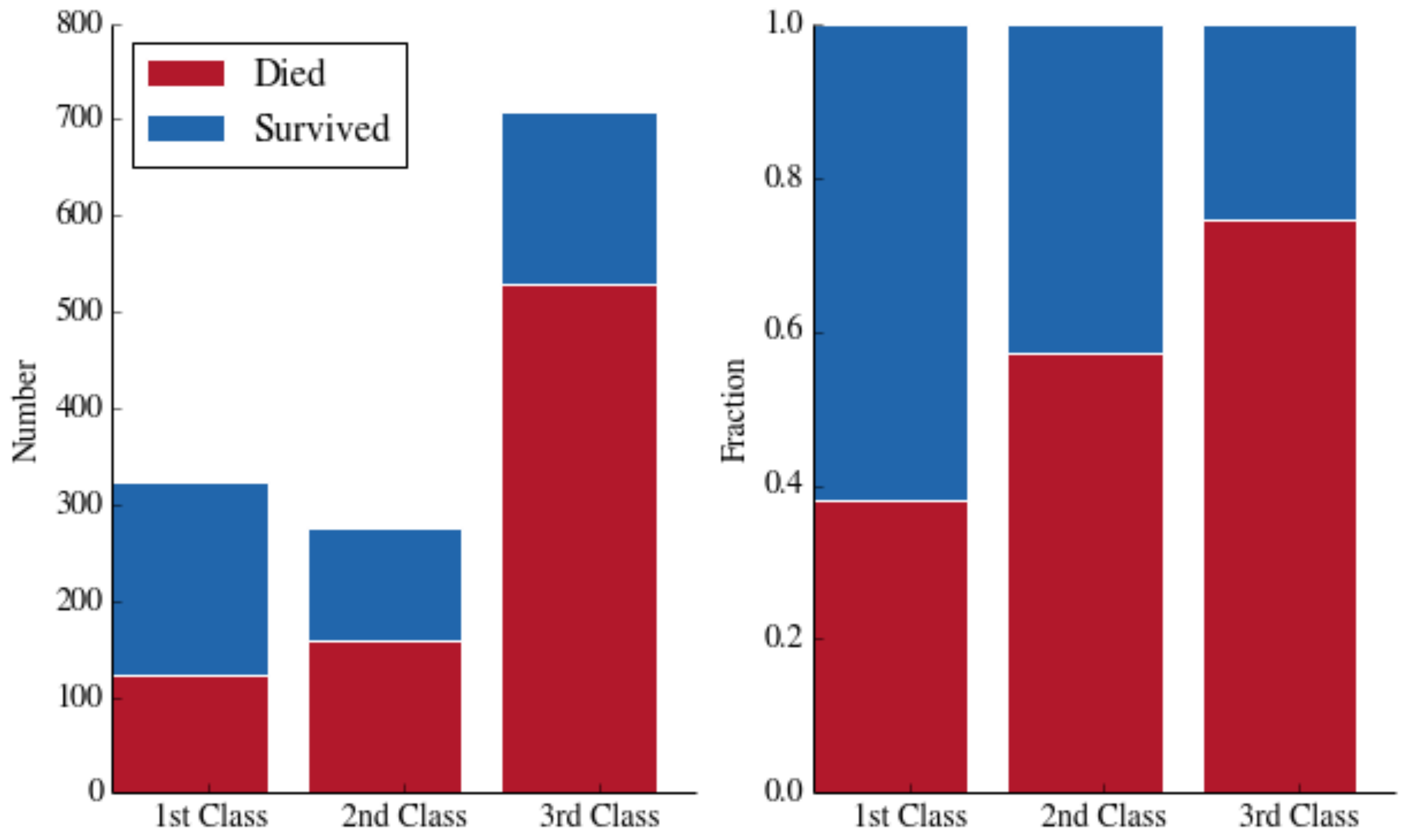
Keys: Class, Survival

Class is spatial

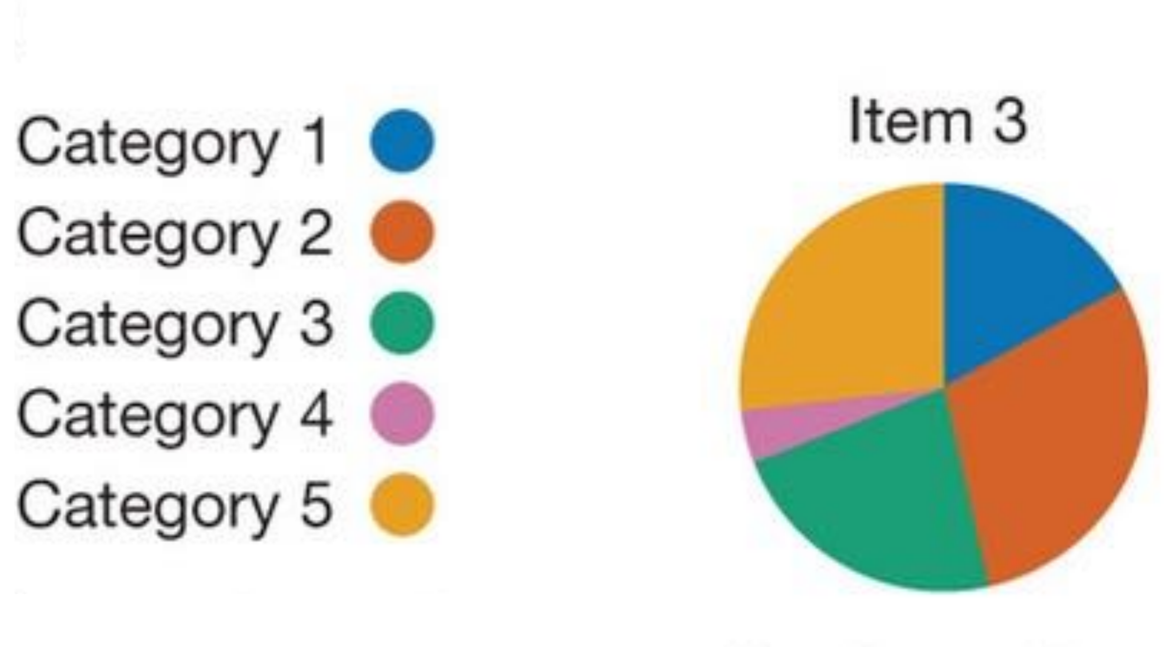
Survival is color

Left: absolute values

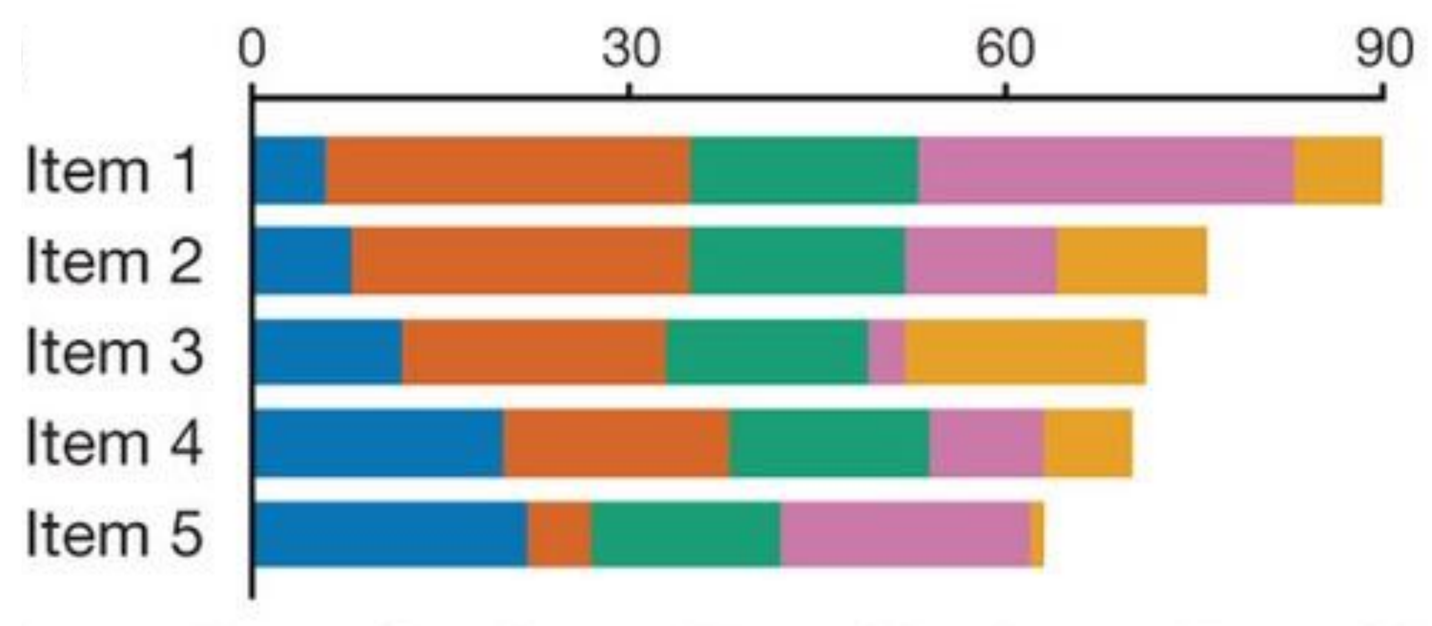
Right: proportional values



Comparison of bar chart types

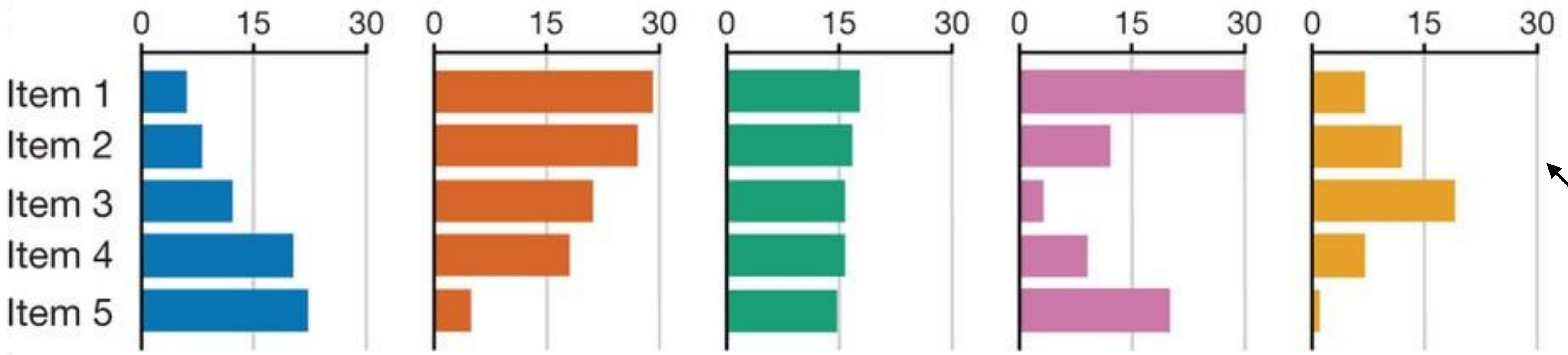


Pie Chart

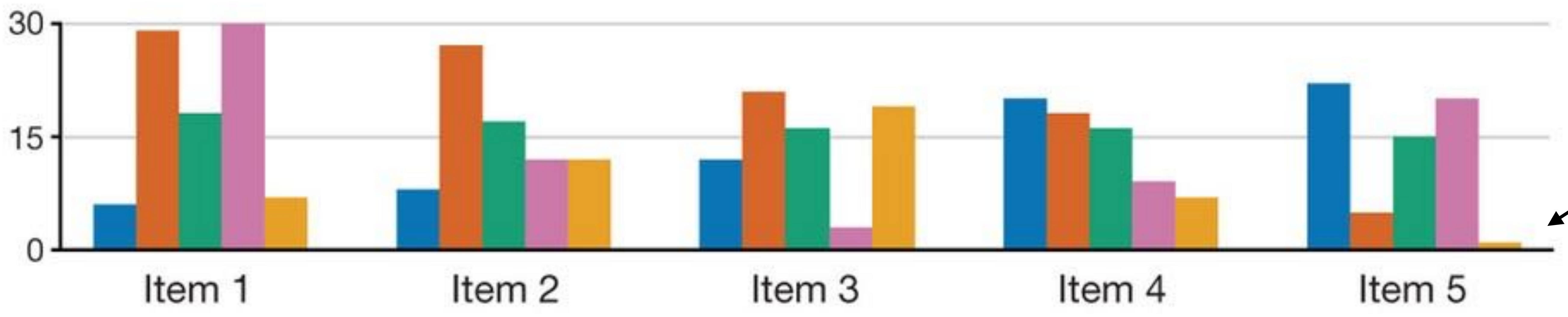


Stacked bar chart

Layered Bar Chart



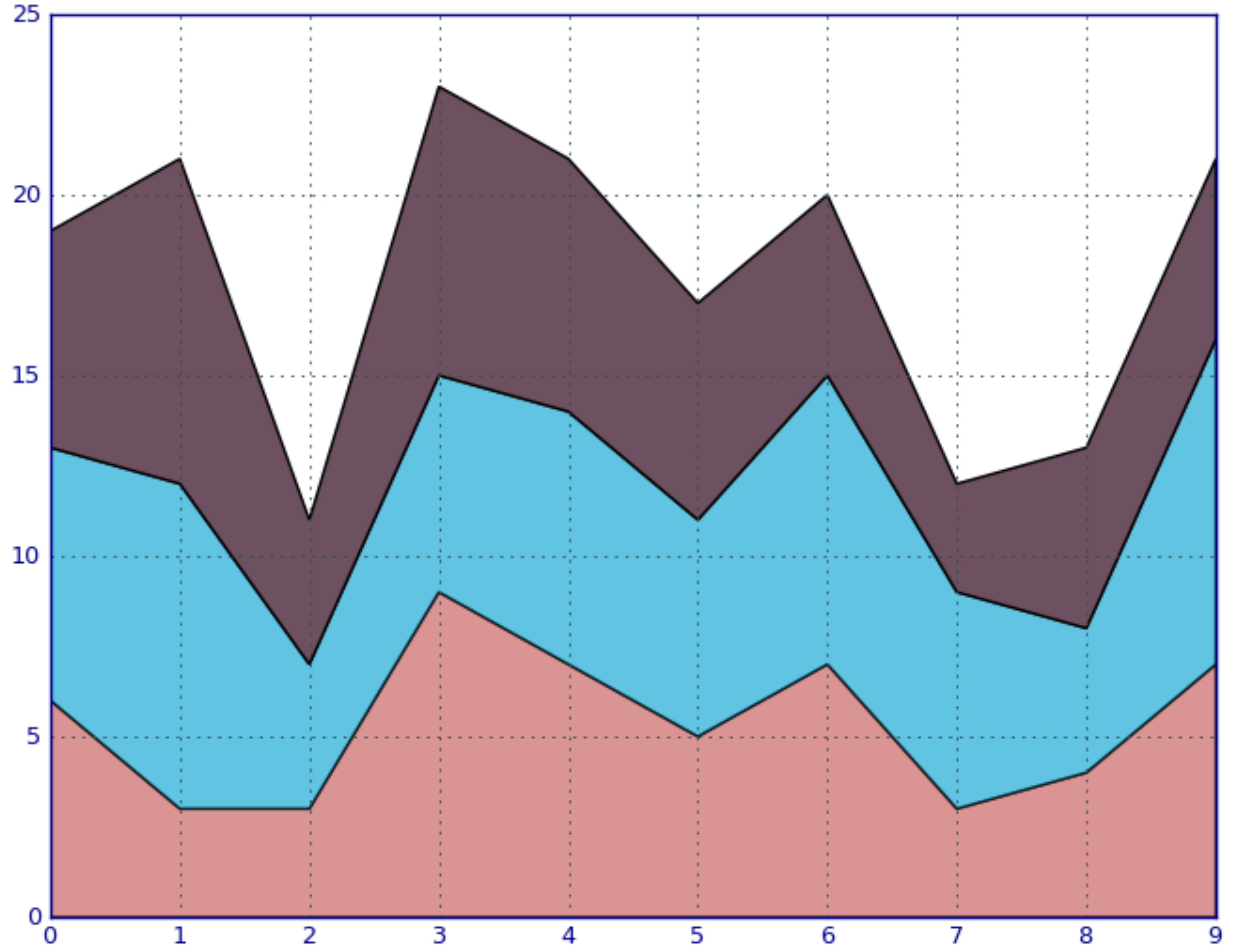
Grouped Bar Chart



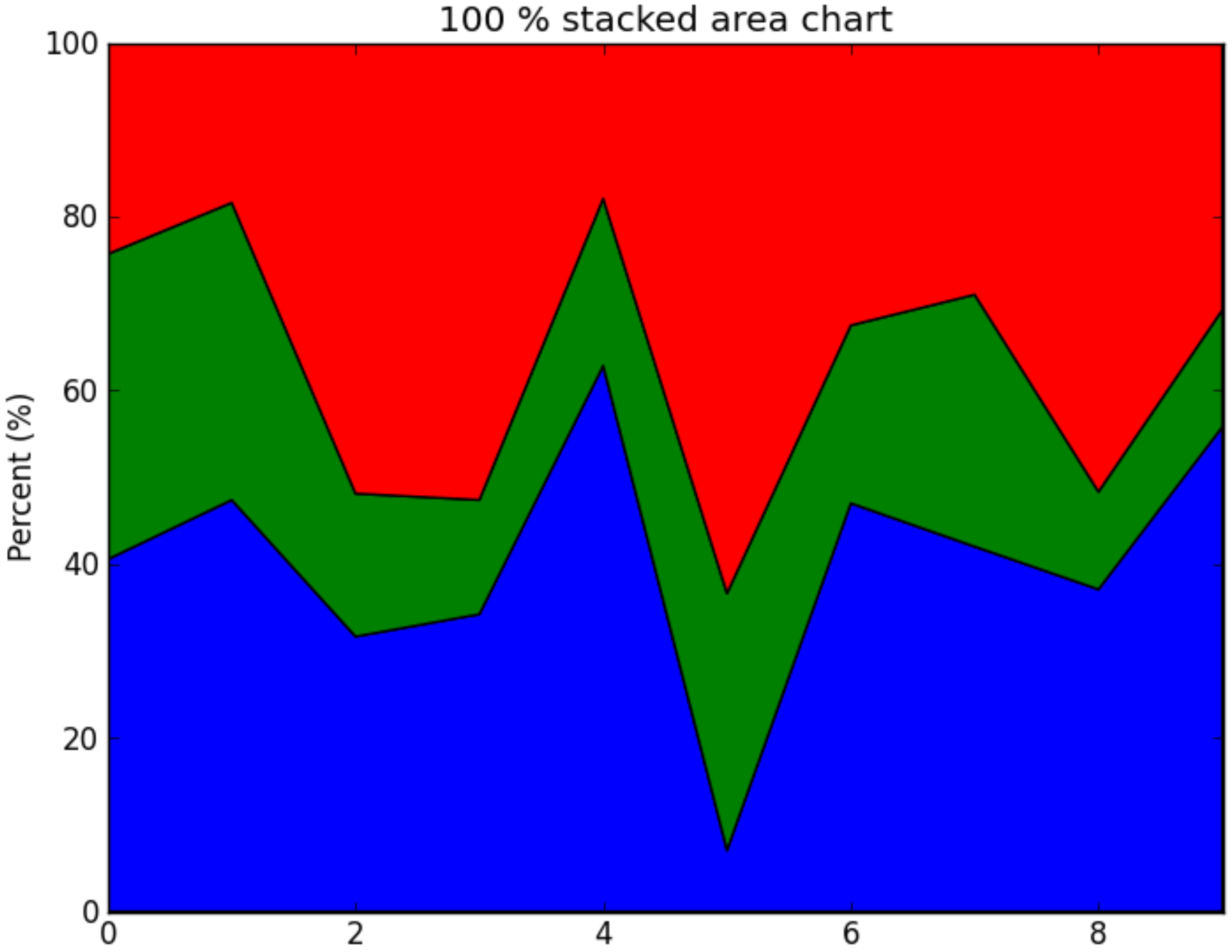
Small Multiples

Streit & Gehlenborg, PoV, Nature Methods, 2014

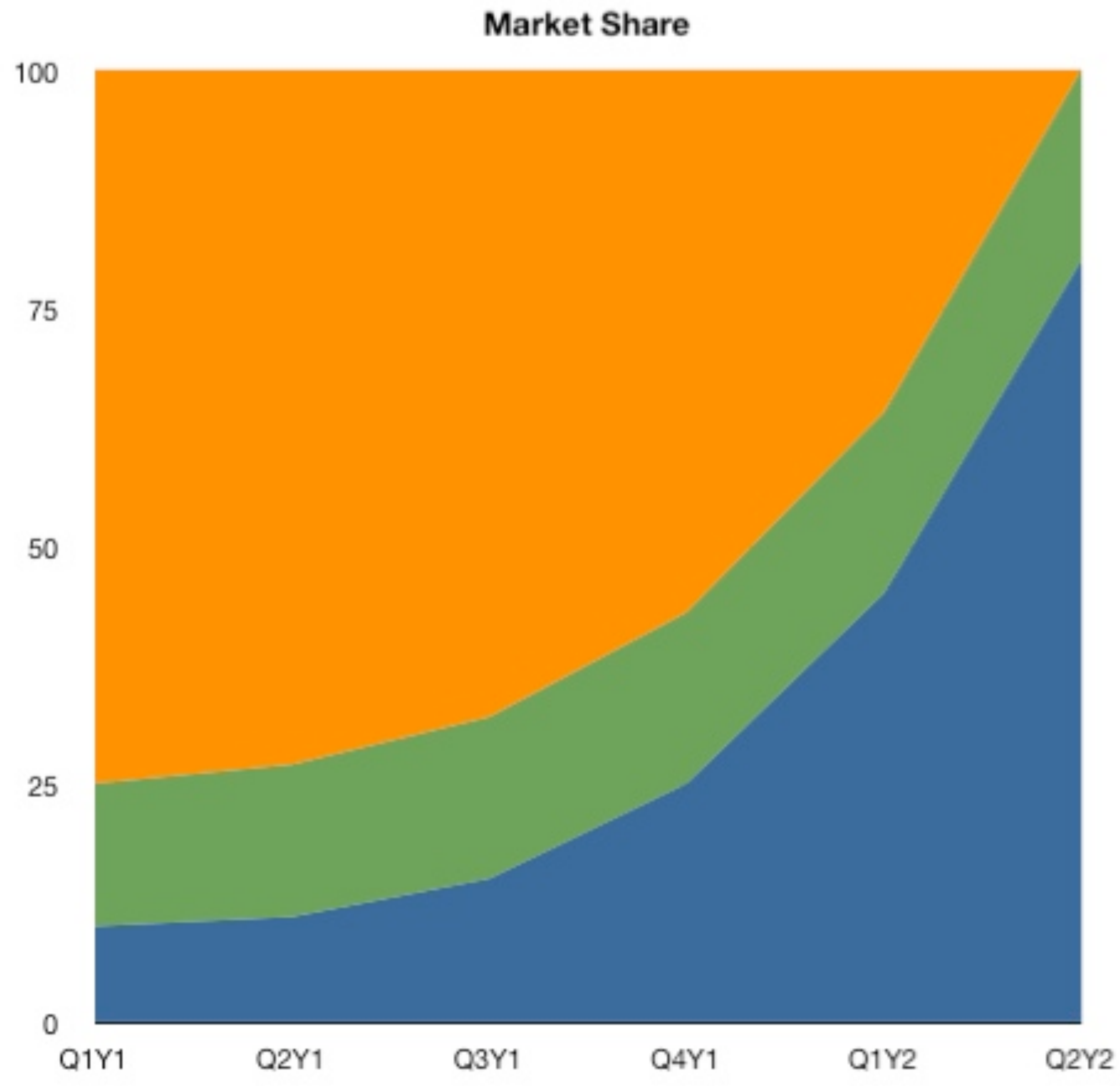
Stacked Area Chart

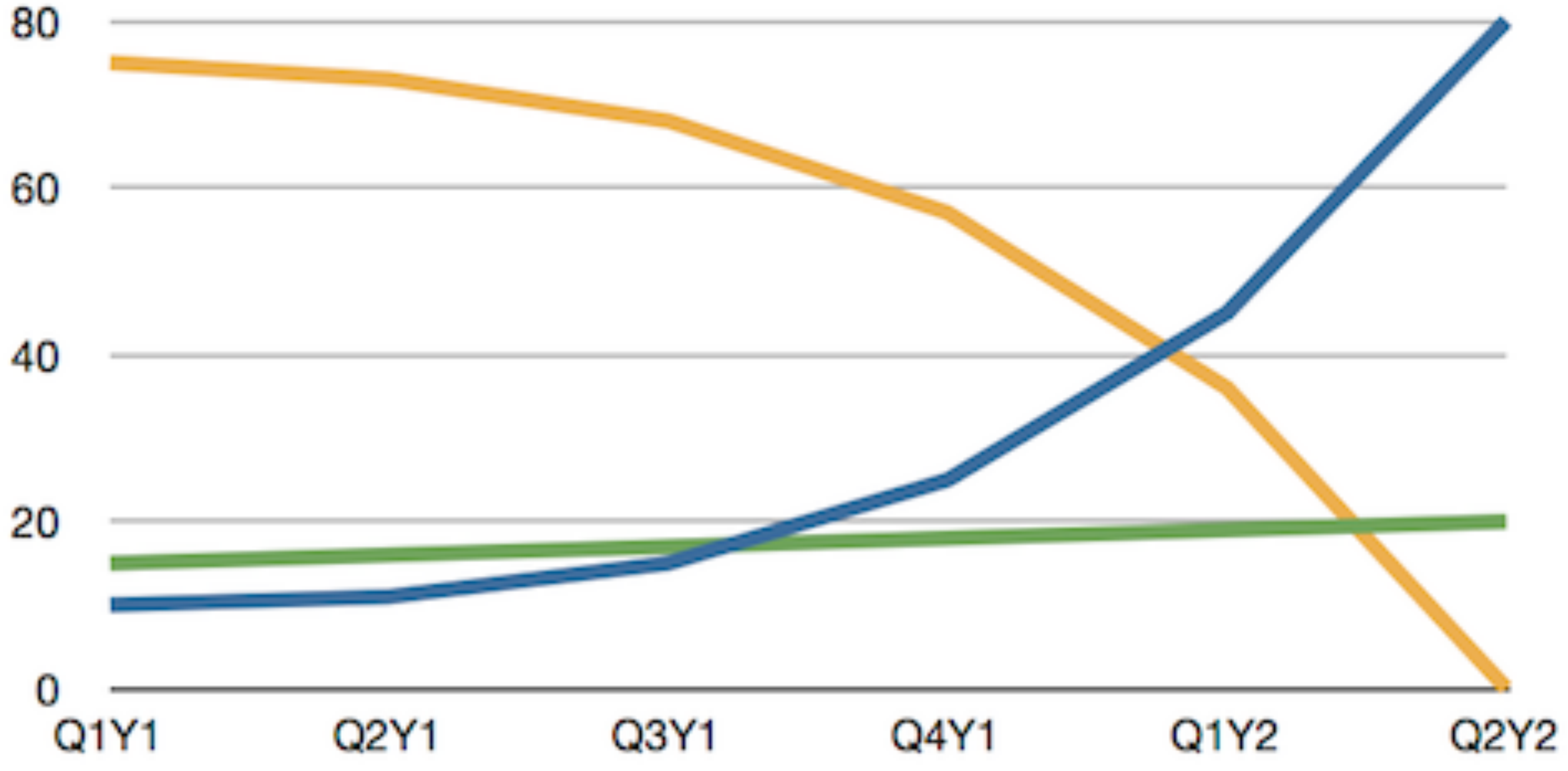


100% Stacked Area Chart



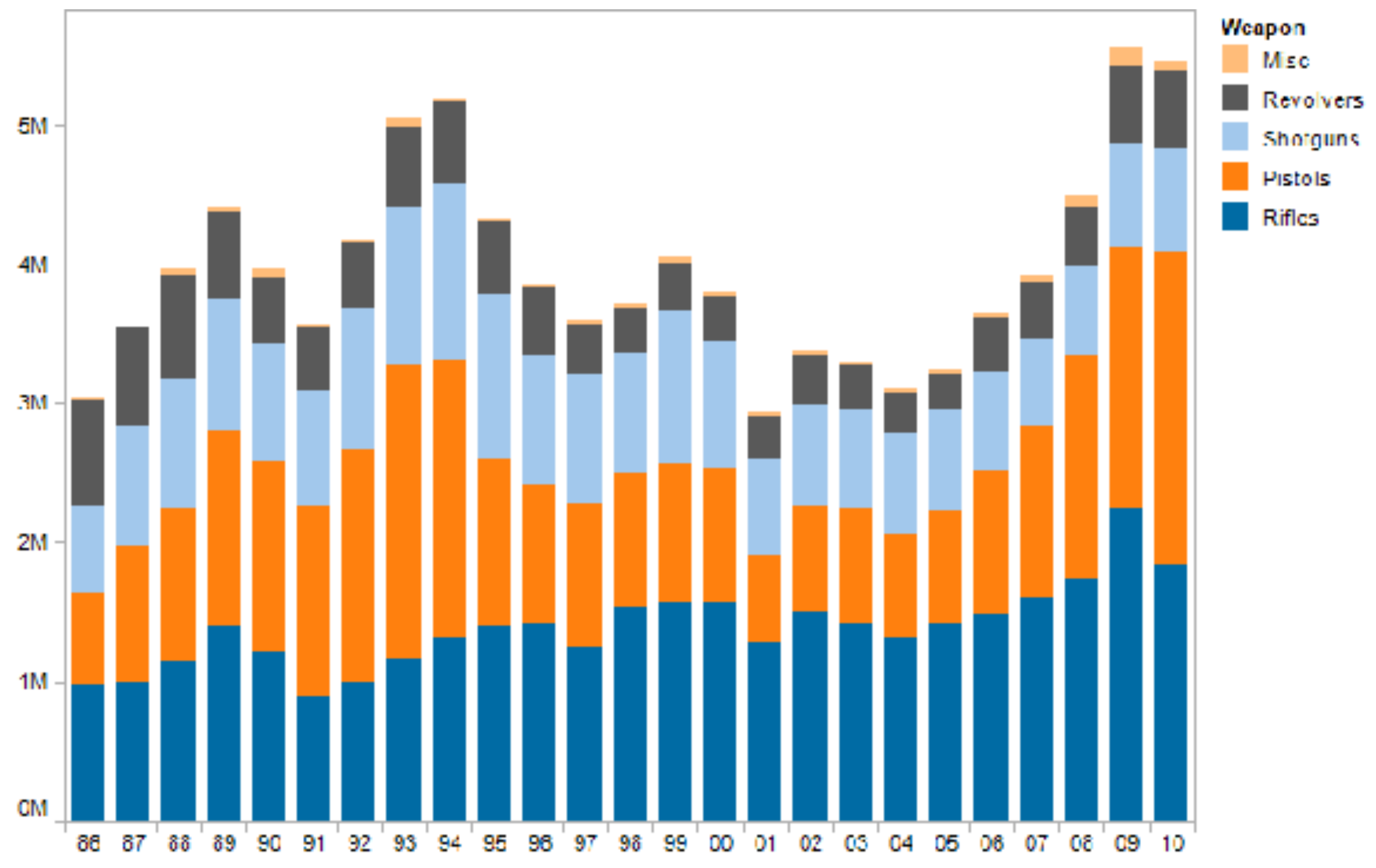
Stacked Area vs. Line Graphs

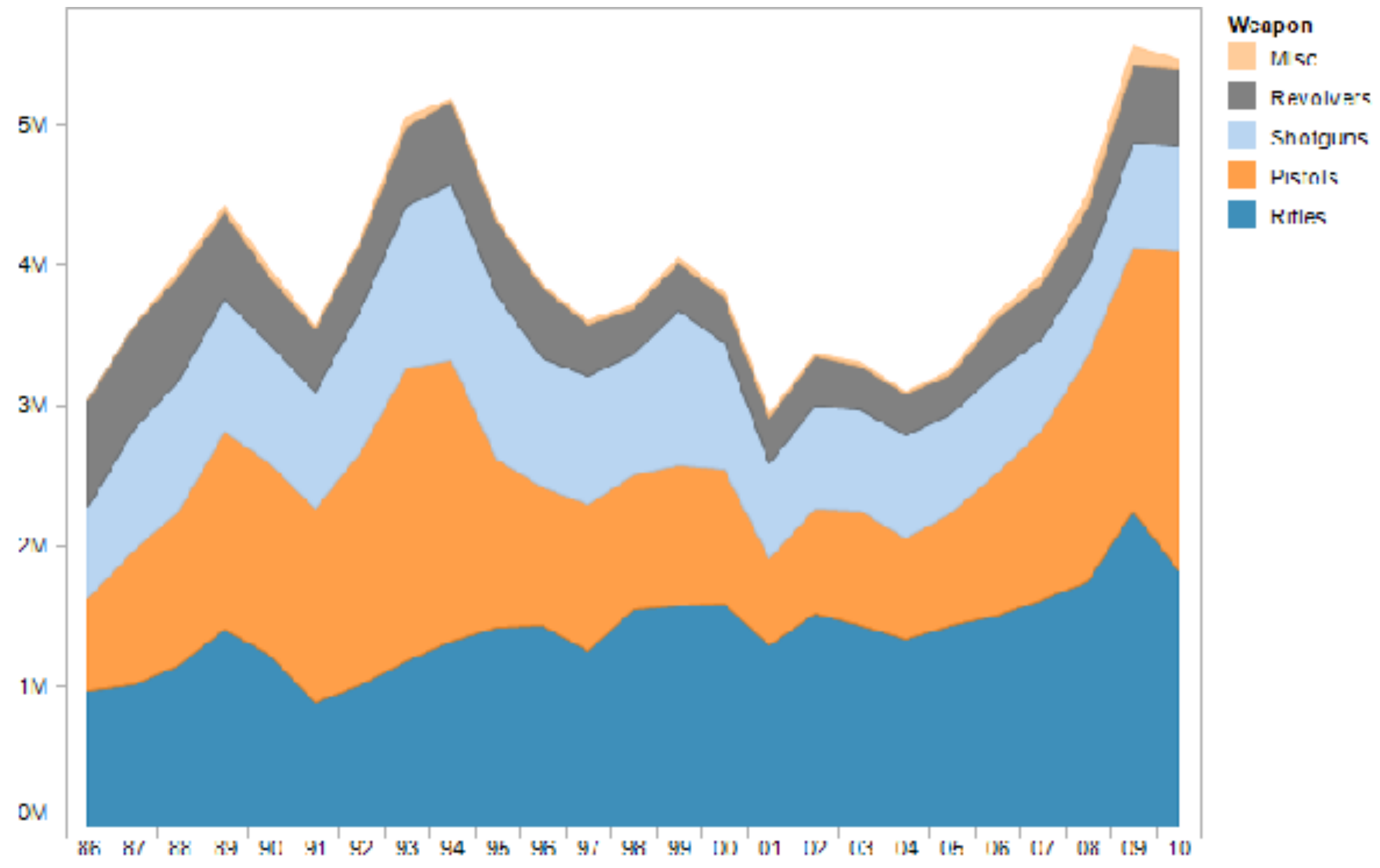


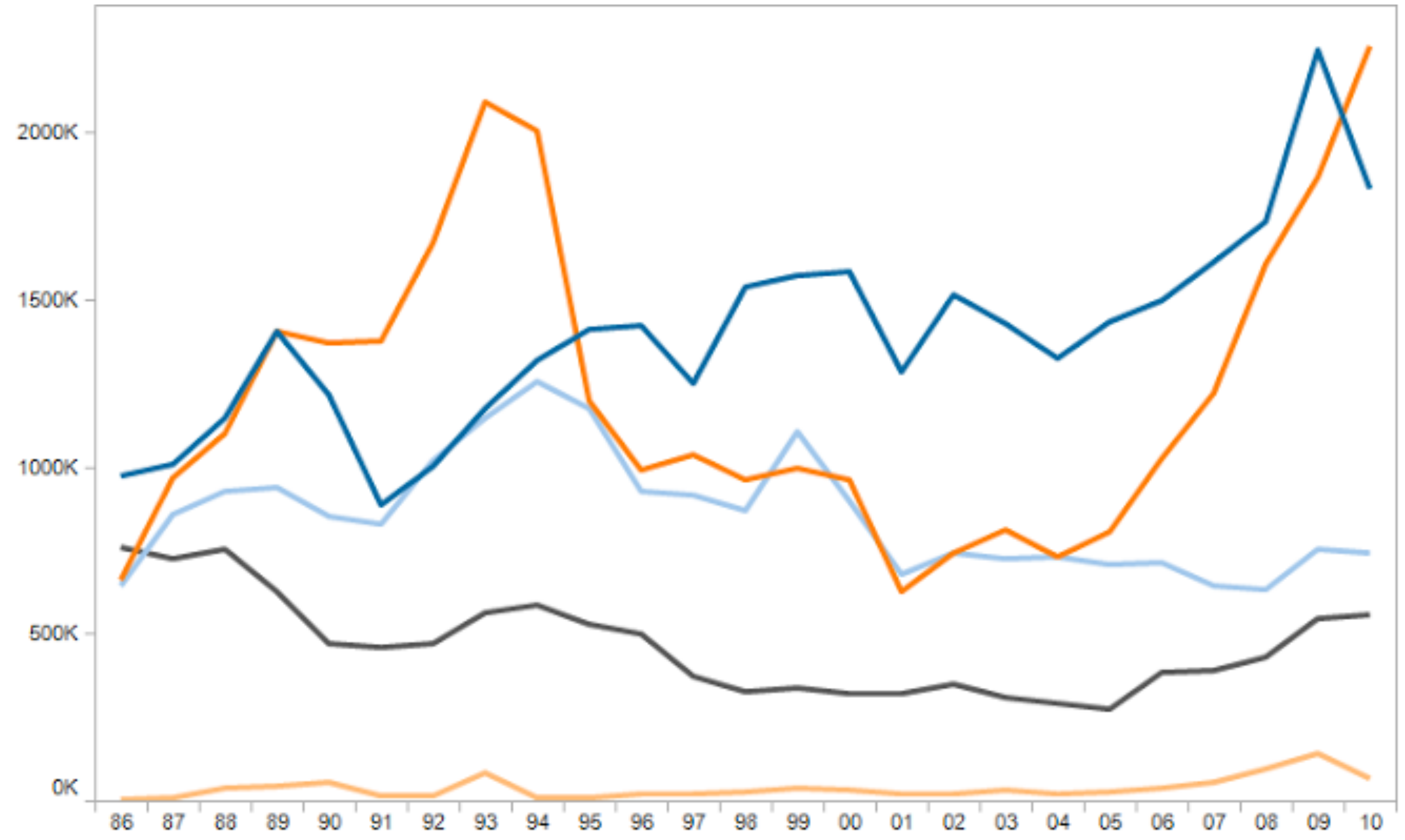


leancrew.com &
Practically Efficient

Can you spot the trends?



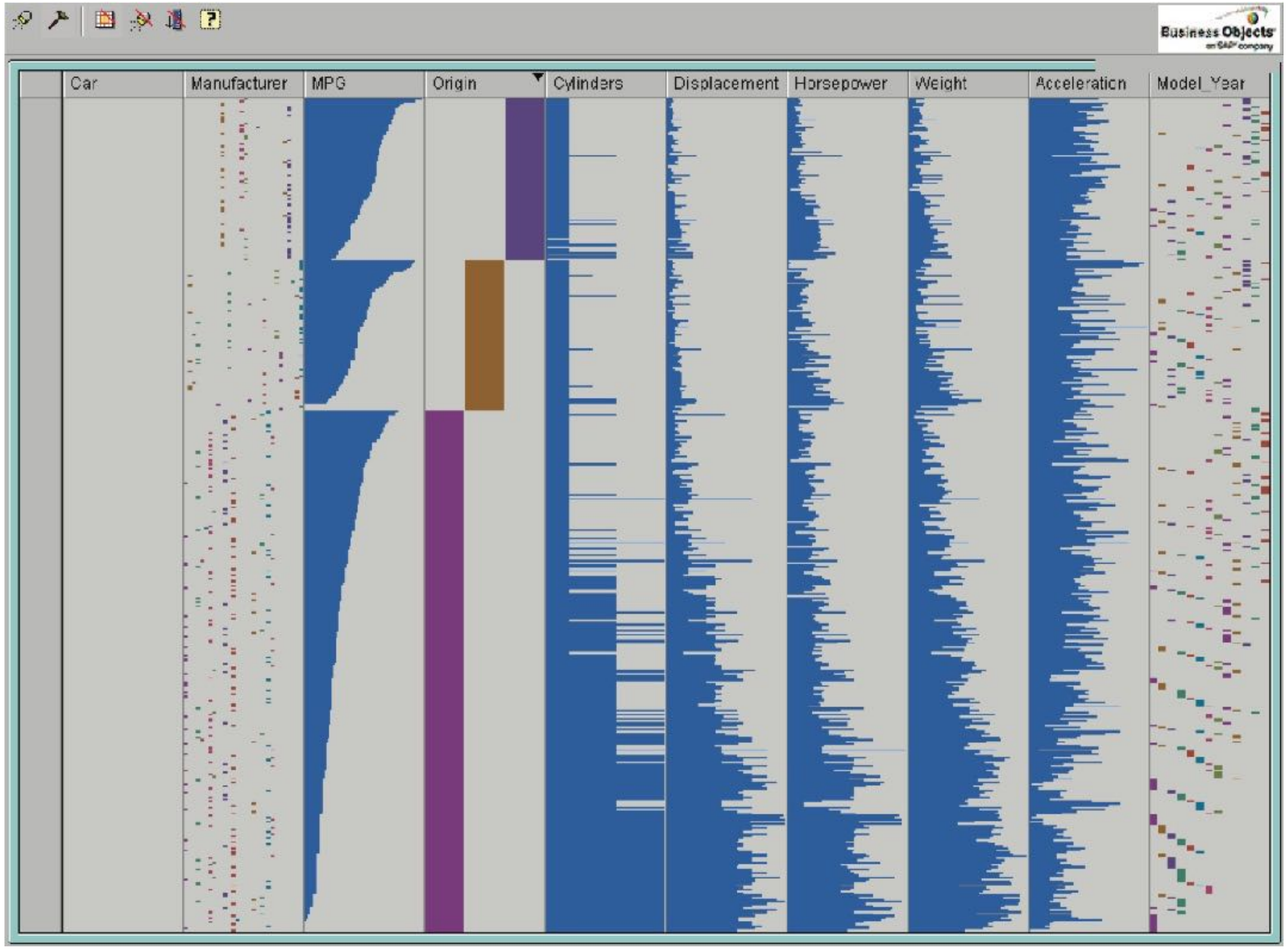




VizWiz, A. Kriebel

Table Lens

Interactive table-based representation

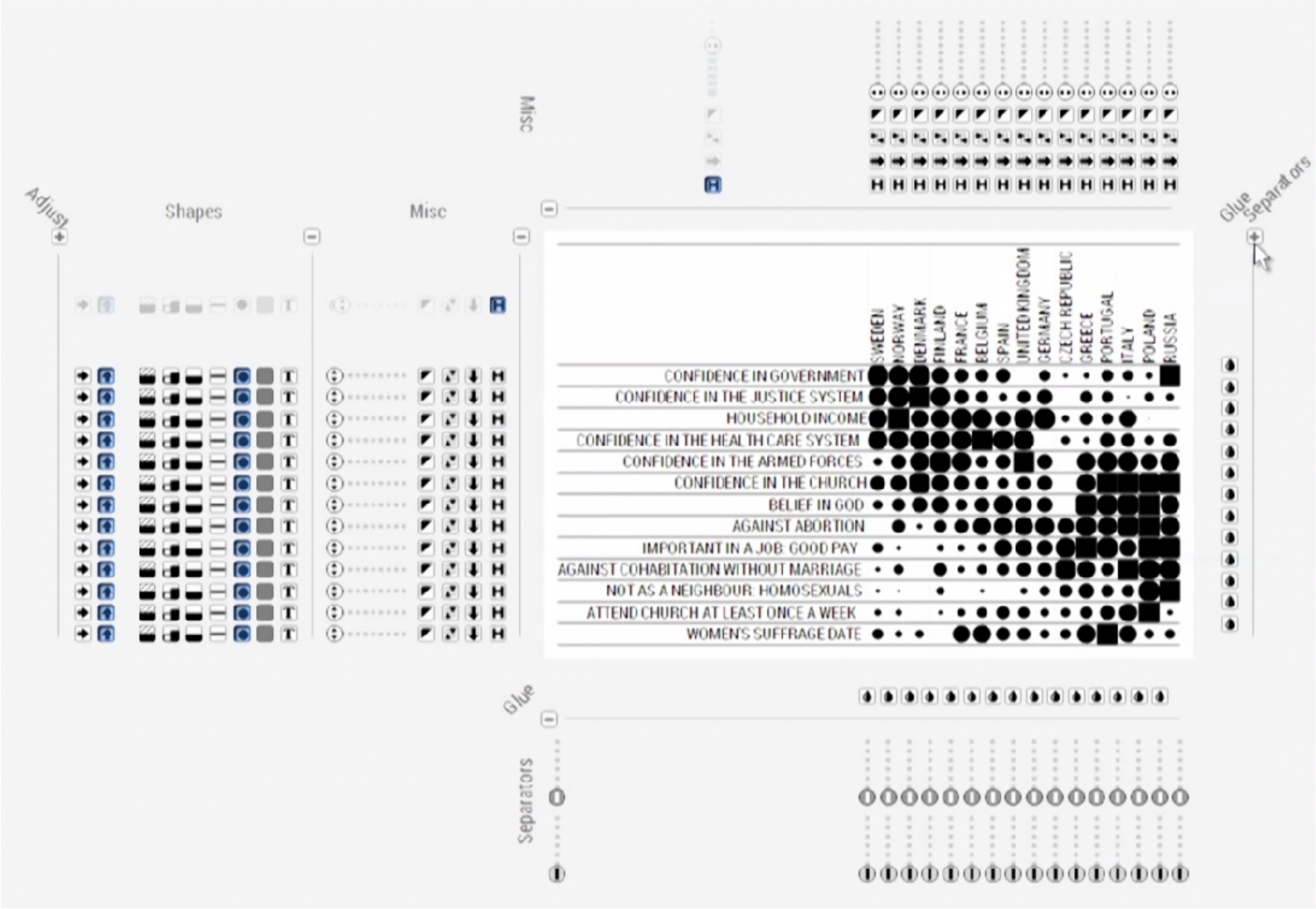


Rao & Card 1994

Bertifier

Matrix/Table representation

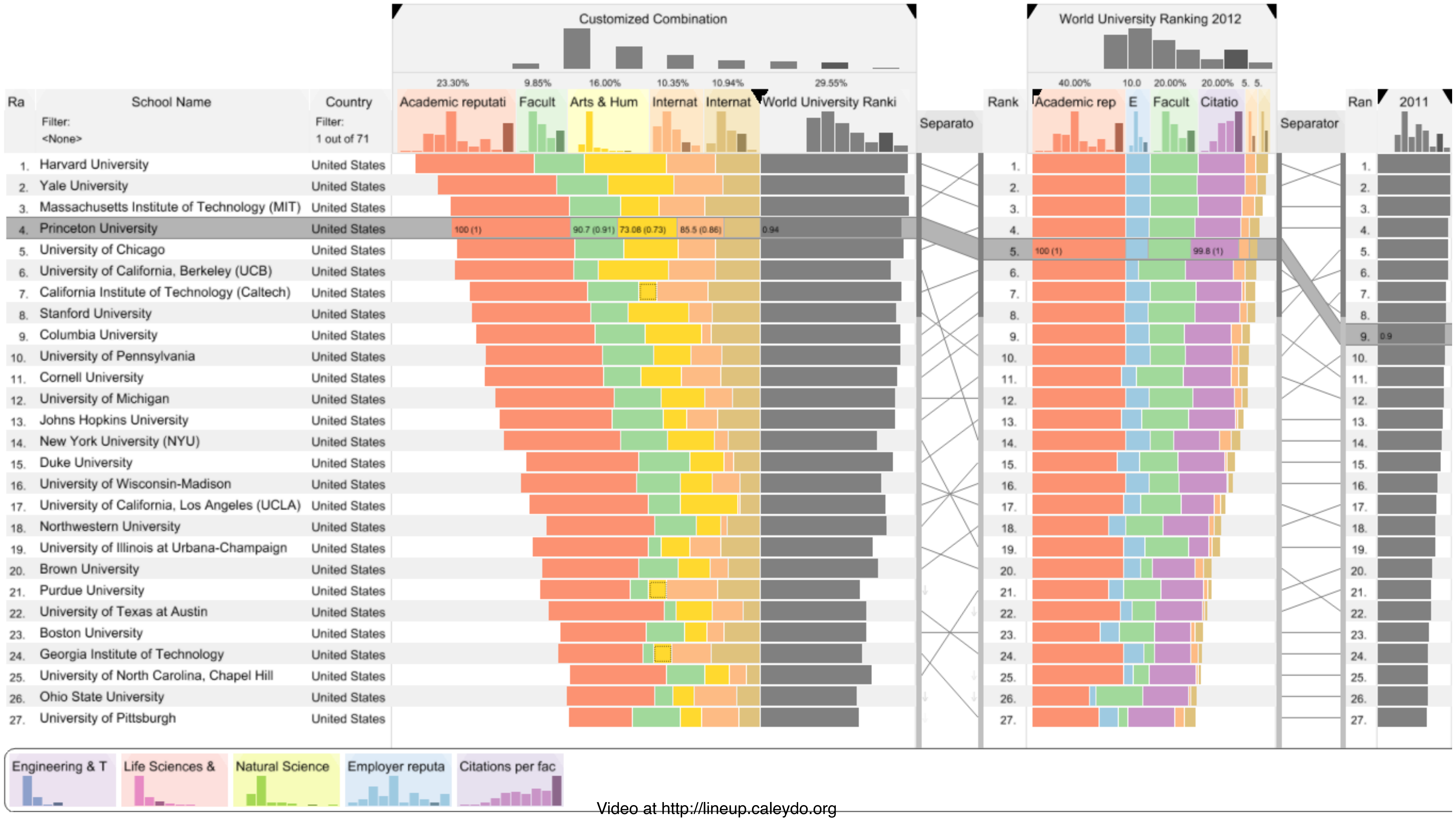
Authoring Interface



http://www.aviz.fr/bertifier

Charles Perin, Pierre Dragicevic and Jean-Daniel Fekete

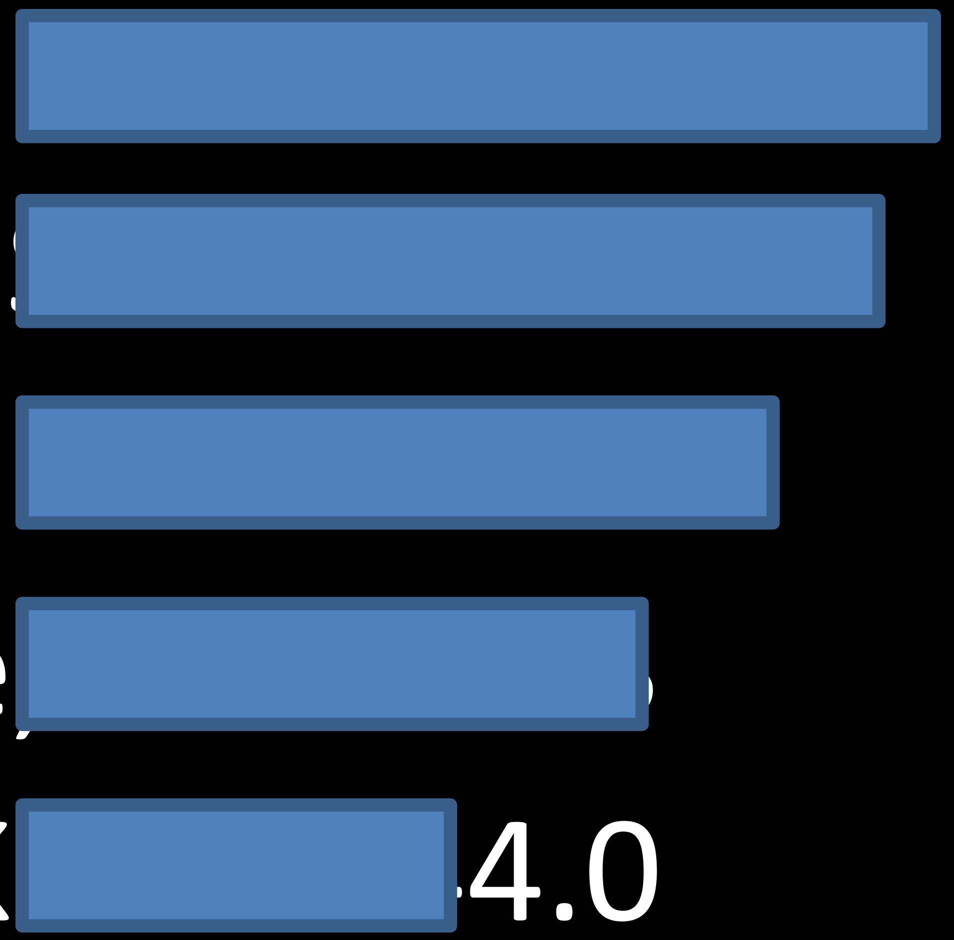
LineUp



Video at http://lineup.caleydo.org

Rankings are popular

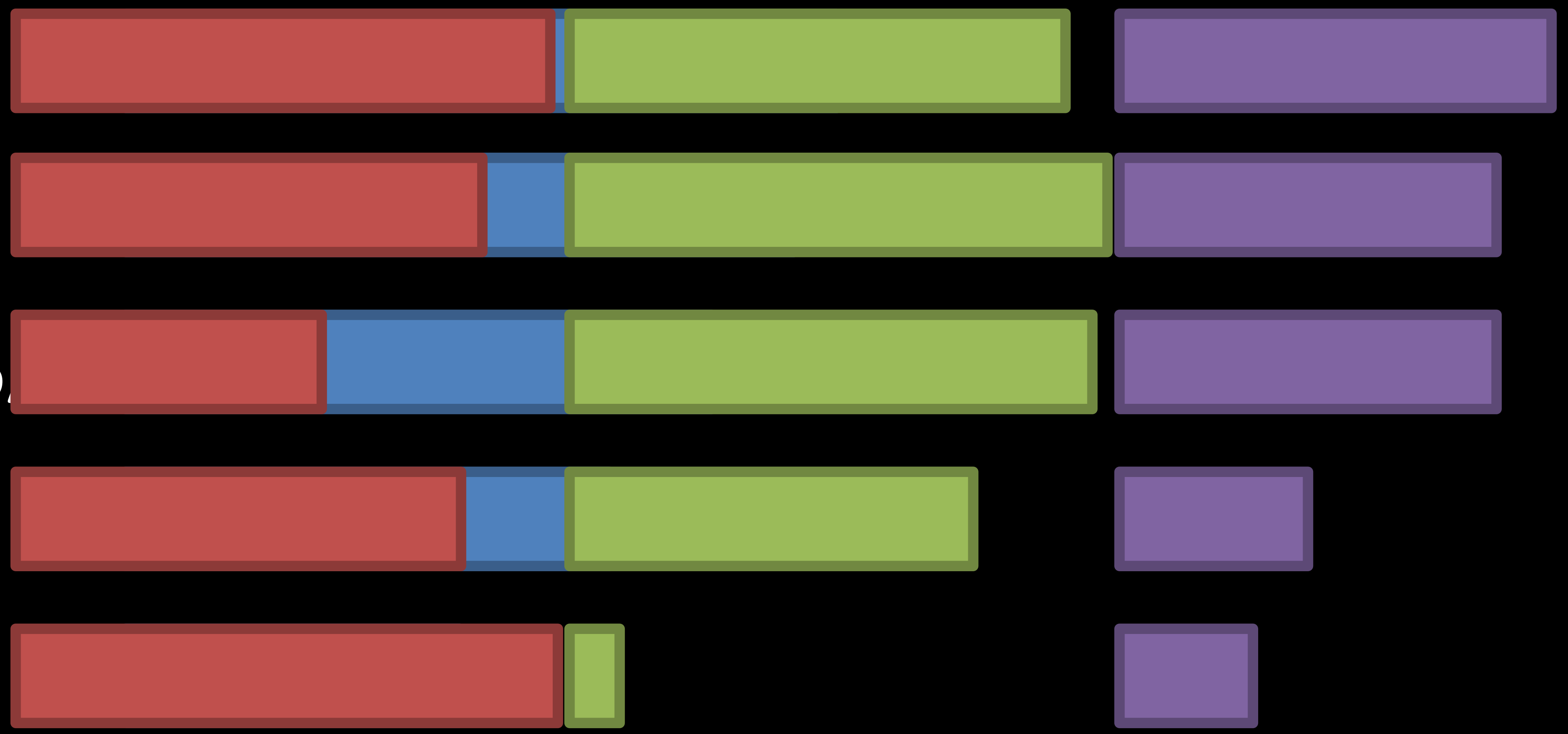
Rank	University	Score
1.	MIT, USA	
2.	Harvard, U	
3.	Princeton,	
4.	Cambridge,	
5.	Oxford, UK	4.0



Support Multiple Attributes

Score = f(A, B, C)

Rank	University	A	B	C
1.	MIT, USA			
2.	Harvard, USA			
3.	Princeton, US			
4.	Cambridge, U			
5.	Oxford, UK			



Combiner functions: f(A,B,C)

(Weighted) sum

$Score = w_a A + w_b B + w_c C$

Maximum

$Score = max(A, B, C)$

Product

Nesting

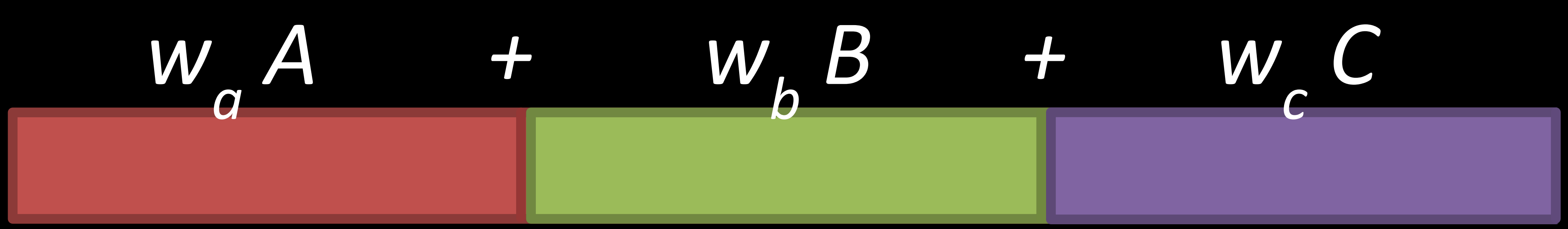
…

→ Serial

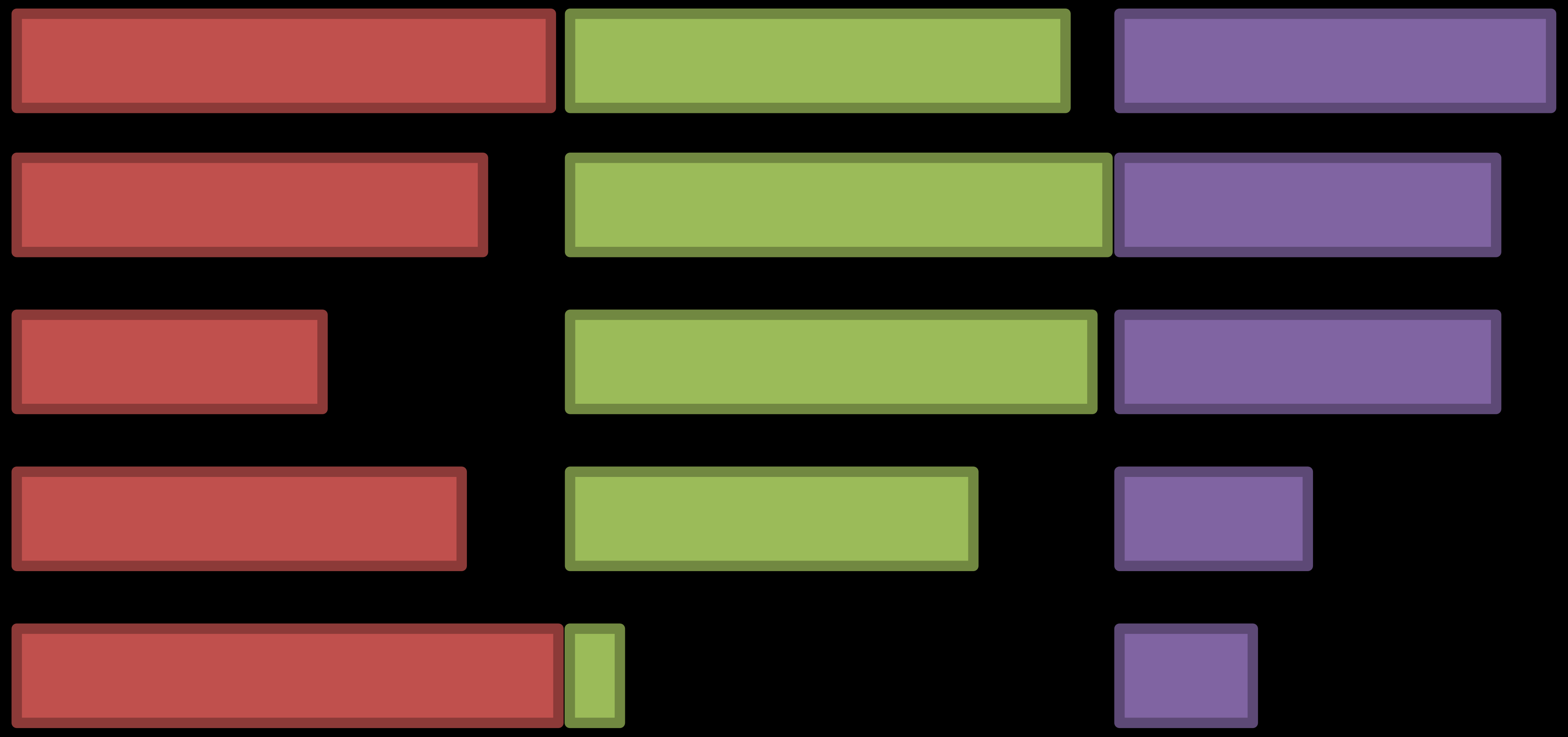
→ Parallel

→ Complex Combiners

Serial Combiner (as Stacked Bar)

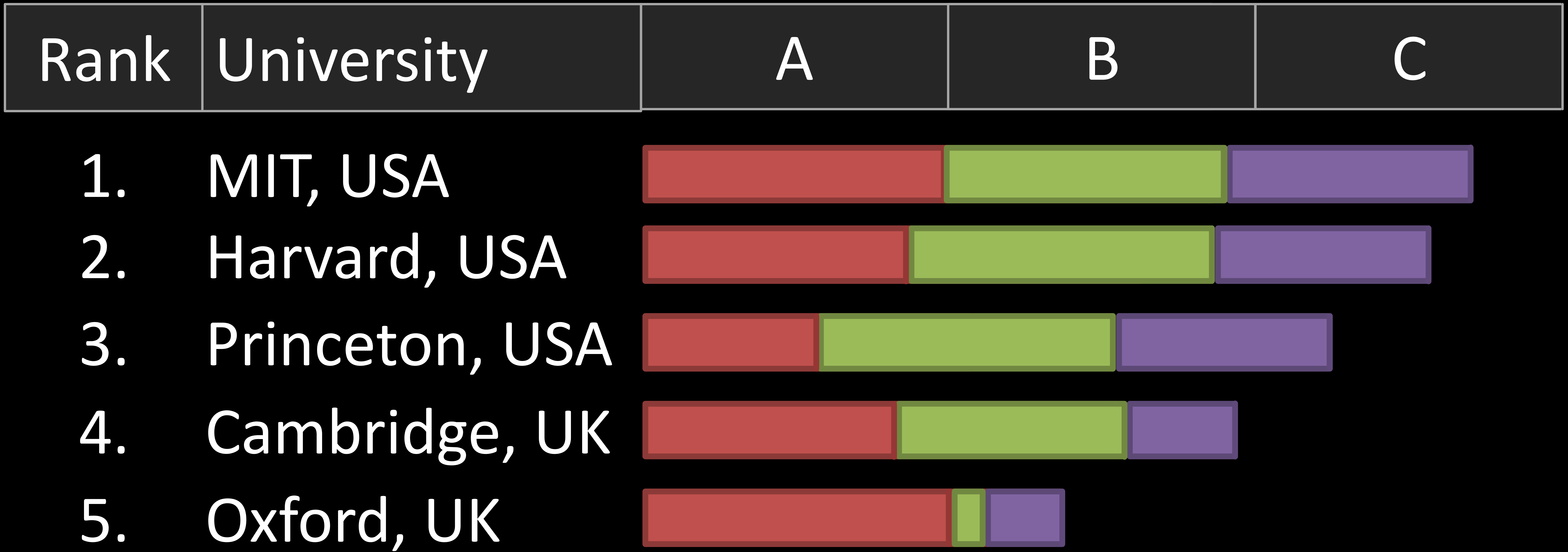


Rank	University	A	B	C
1.	MIT, USA			
2.	Harvard, USA			
3.	Princeton, USA			
4.	Cambridge, UK			
5.	Oxford, UK			



Serial Combiner (as Stacked Bar)

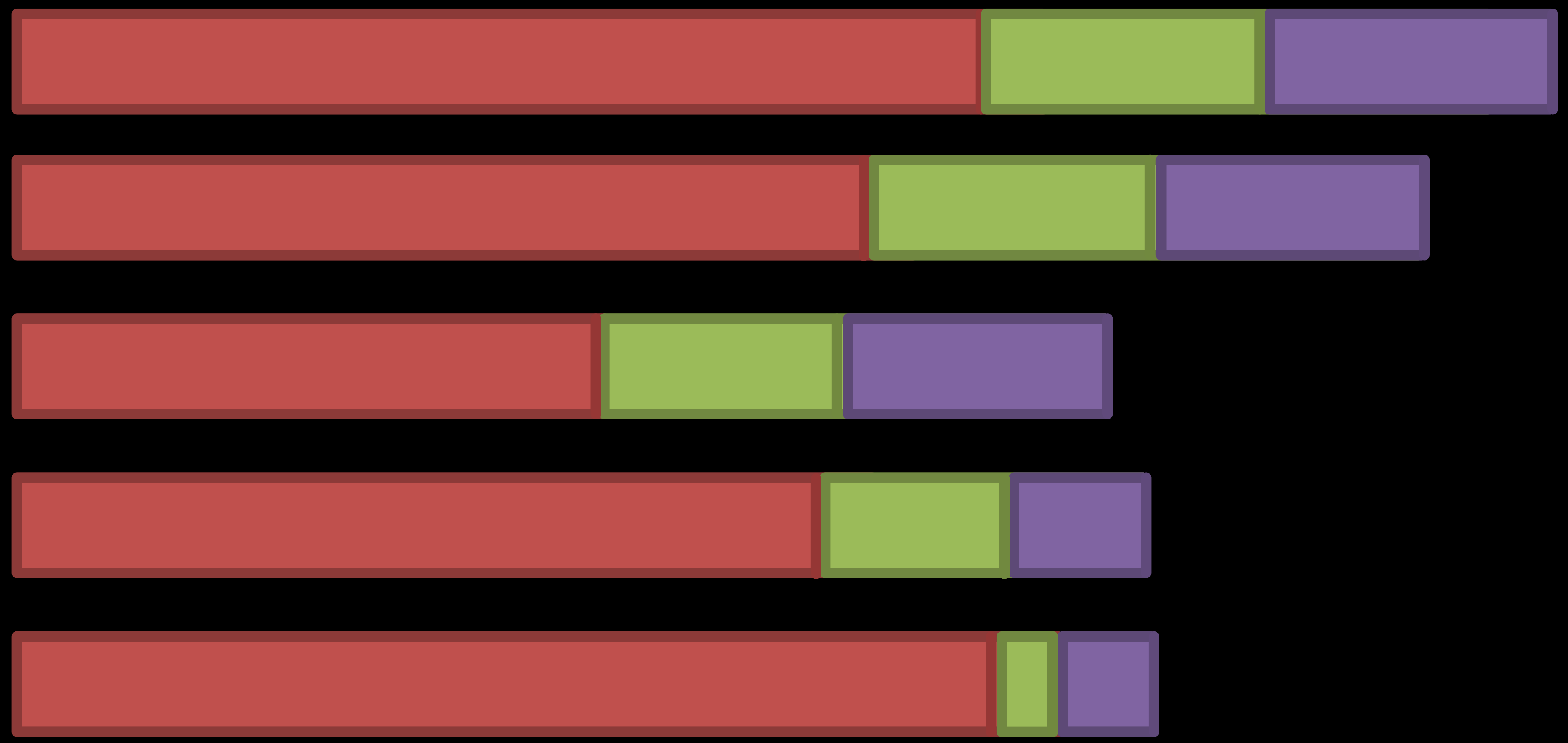
$$w_a A \quad + \quad w_b B \quad + \quad w_c C$$



Serial Combiner (as Stacked Bar)

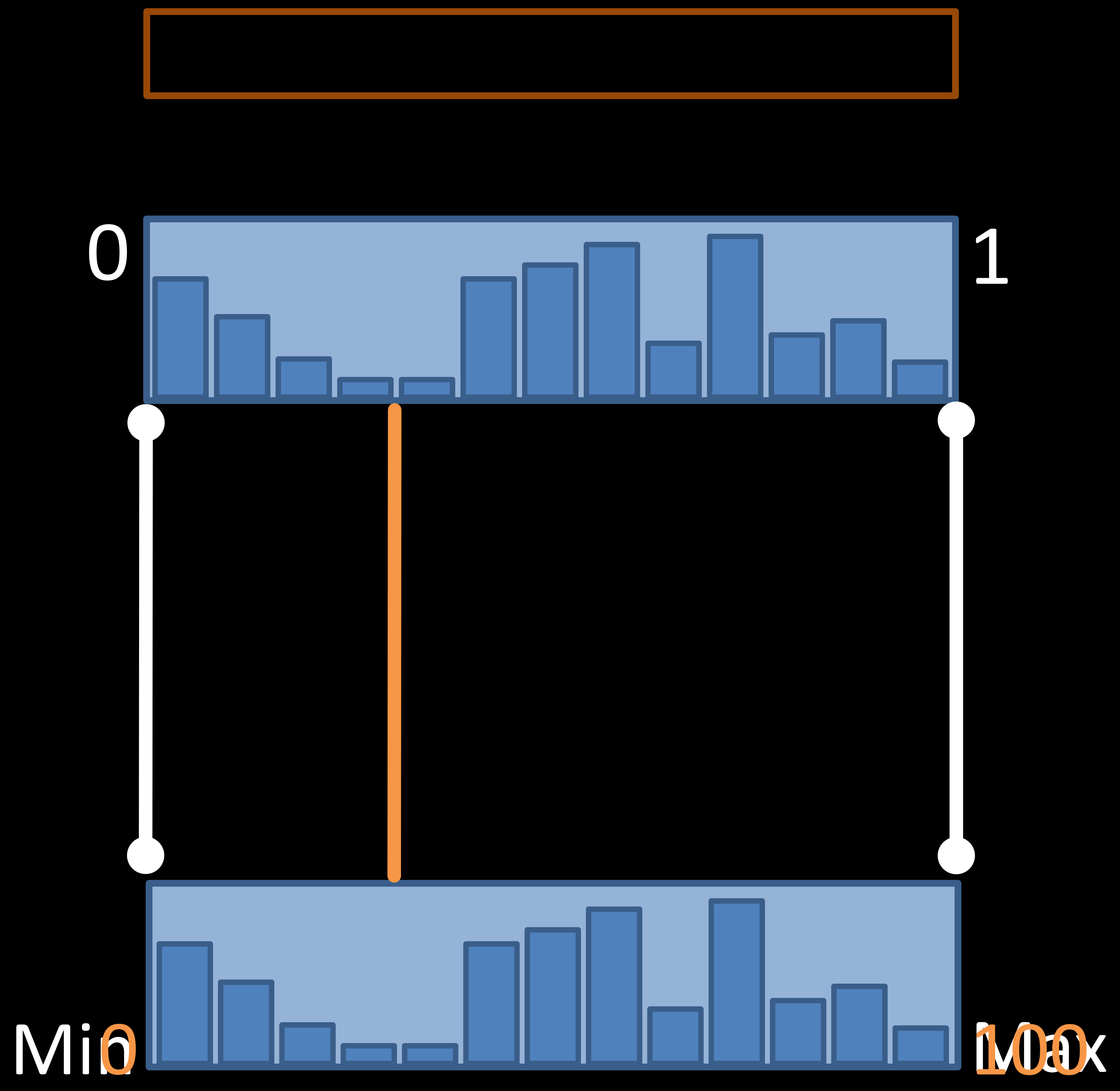
$w_a A + w_b B + w_c C$

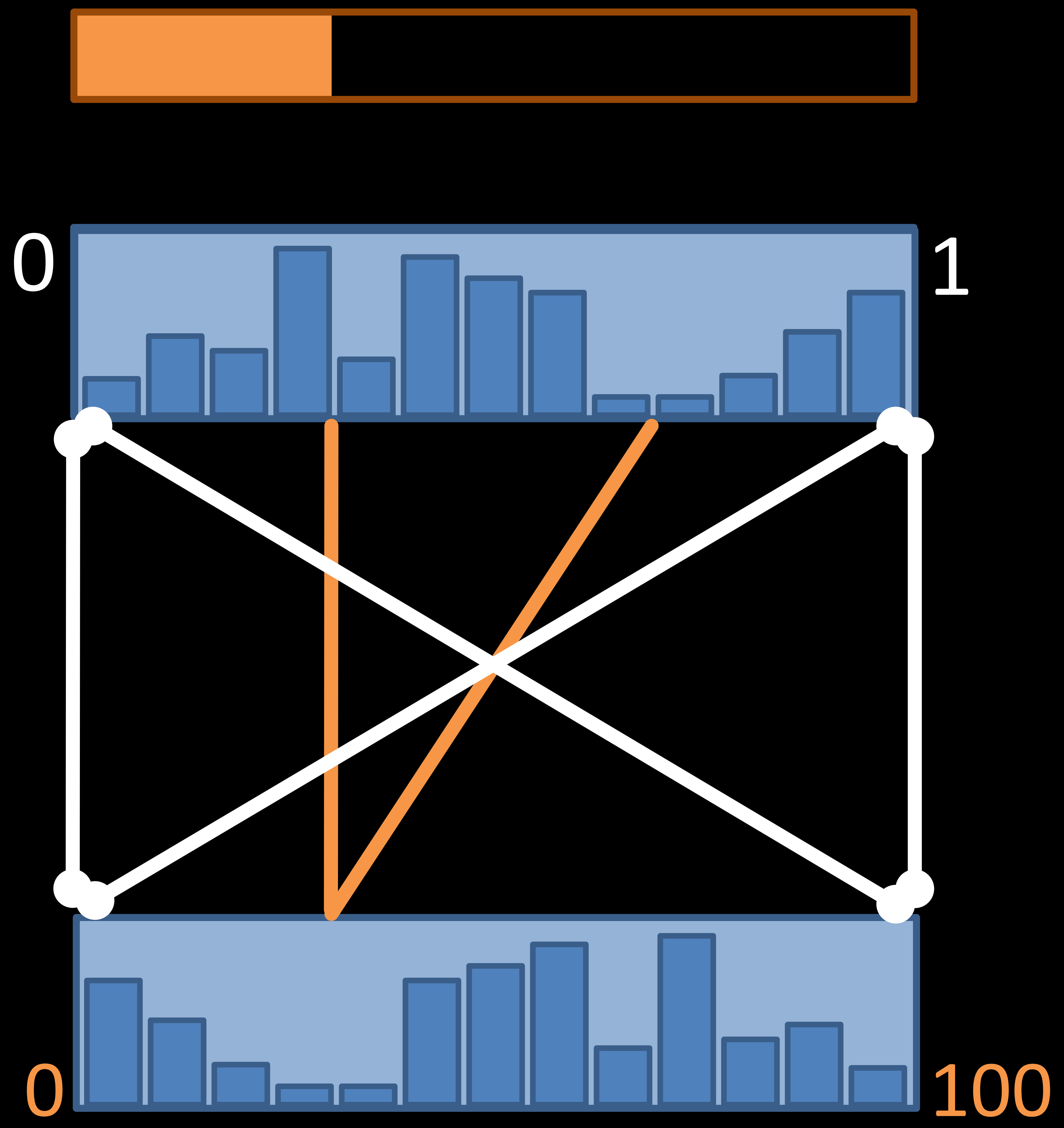
Rank	University	A	B	C
1.	MIT, USA			
2.	Harvard, USA			
3.	Princeton, USA			
4.	Cambridge, UK			
5.	Oxford, UK			

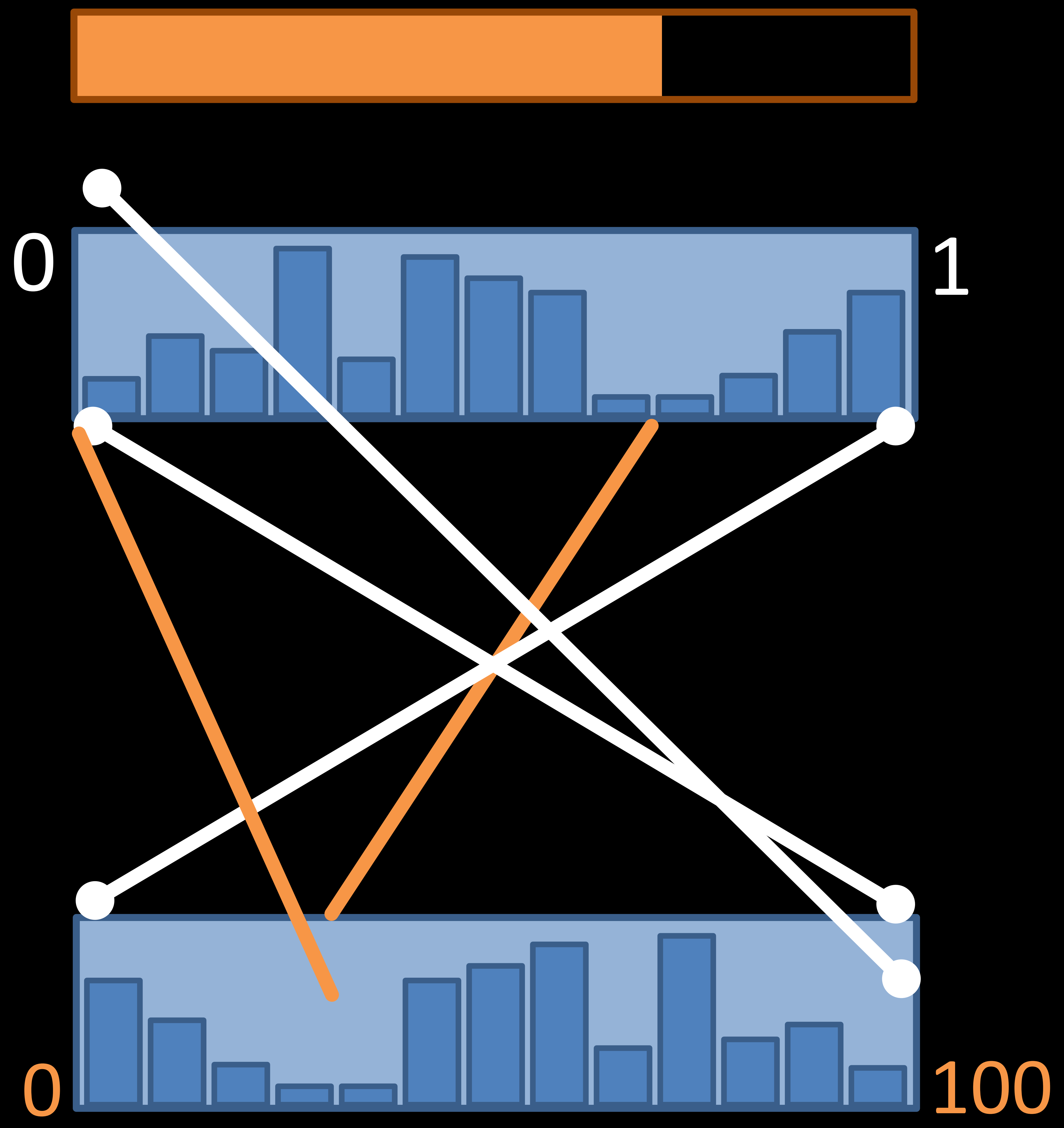



128 visible of 906 (14.13%)
Ran
School Name
Country
Faculty/student ratio
Employer reputation
Citations per faculty
Filter:
<None>
Filter:
2 out of 72
1. American University United States
2. Arizona State University United States
3. Aston University United Kingdom
4. Birkbeck College, University of L United Kingdom
5. Boston College United States
6. Boston University United States
7. Brandeis University United States
8. Brown University United States
9. Brunel University United Kingdom
10. California Institute of Technology United States
11. Cardiff University United Kingdom
12. Case Western Reserve University United States
13. City University London United Kingdom
14. College of William & Mary United States
15. Colorado State University United States
16. Columbia University United States
17. Cornell University United States
18. Cranfield University United Kingdom
19. Dartmouth College United States
20. Drexel University United States
21. Duke University United States
22. Durham University United Kingdom

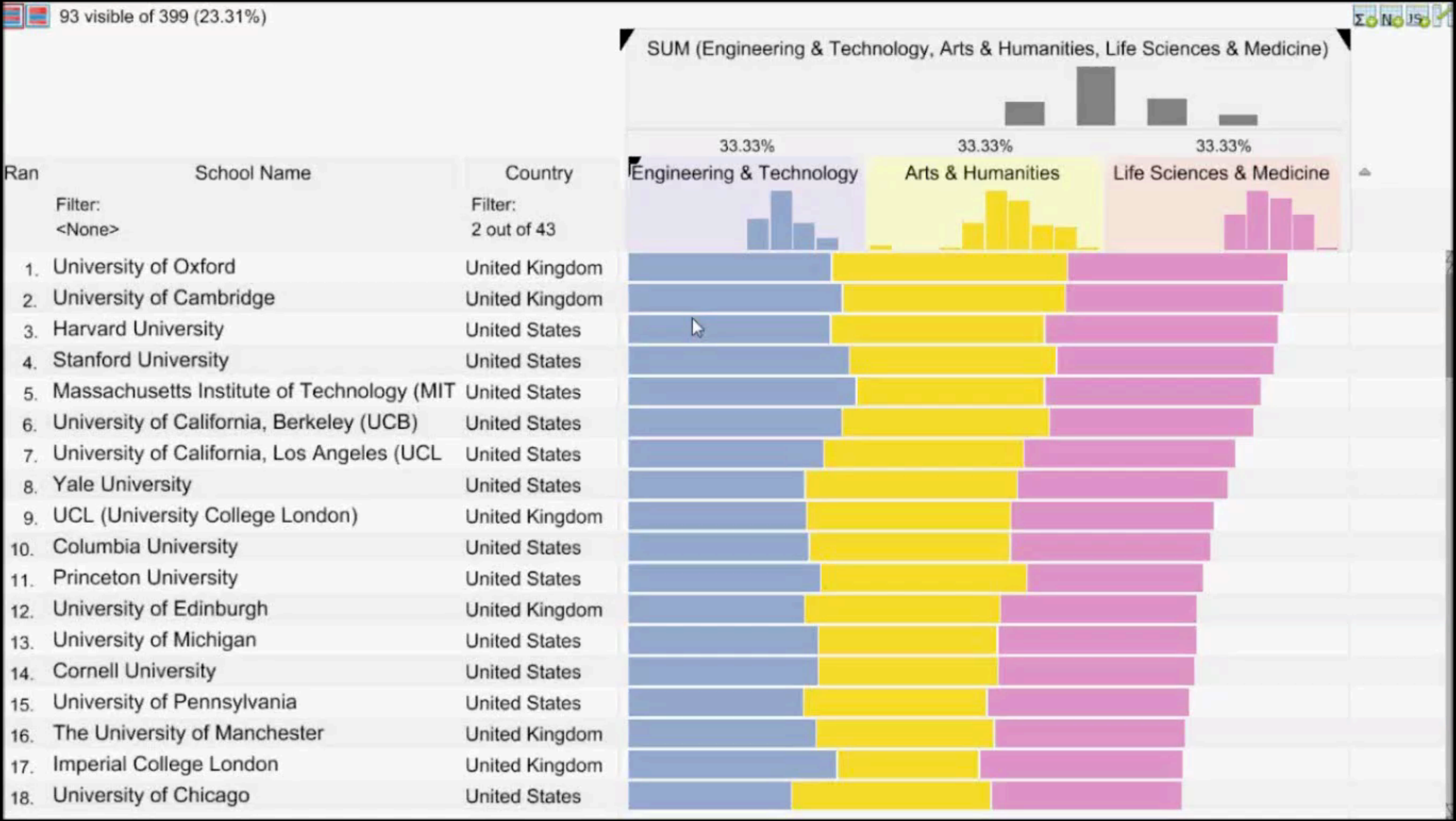
Flexible Mapping of Attributes to Scores


0
1
Min
0
Max
100


0
1
0
100



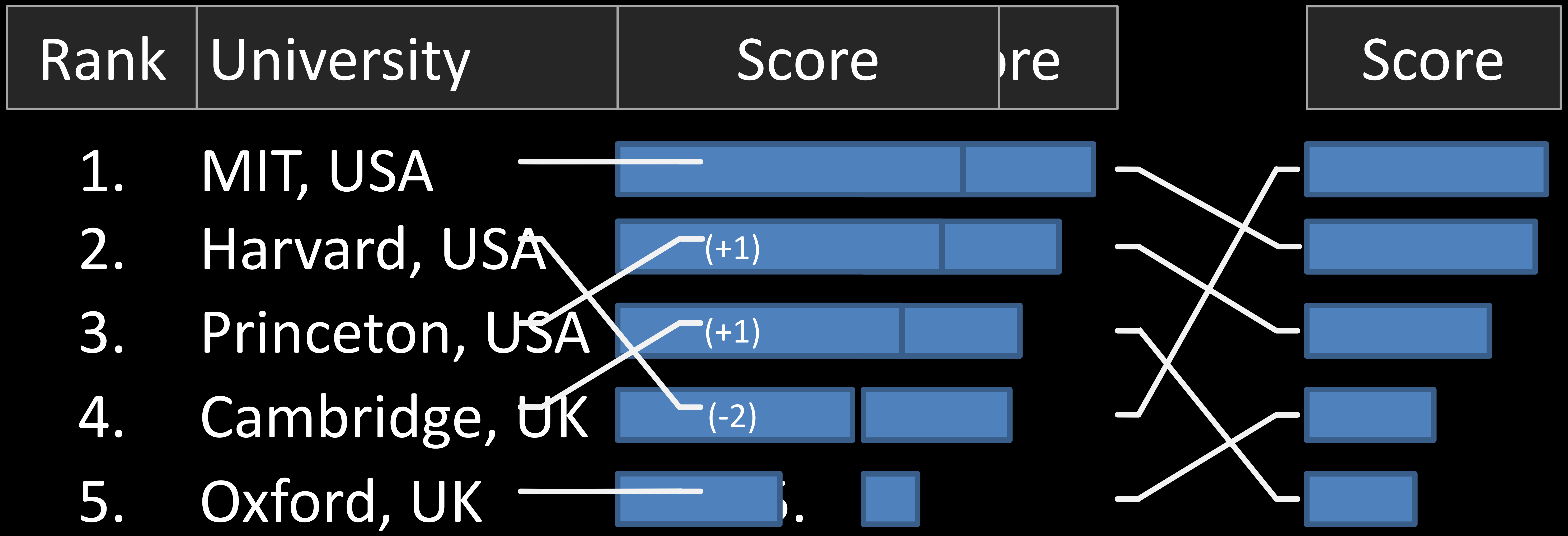
0
1
0
100



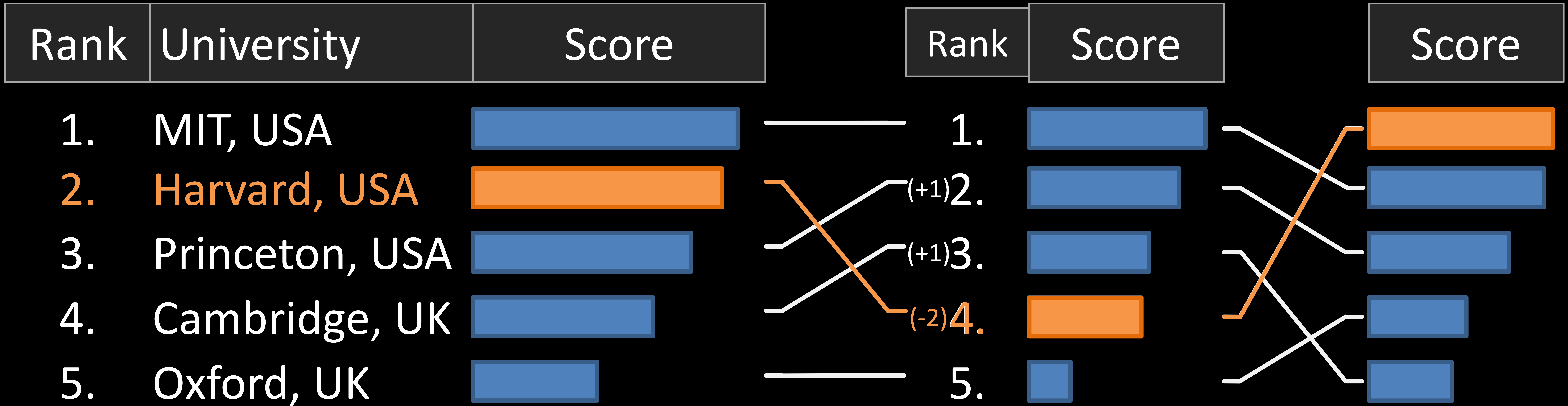
93 visible of 399 (23.31%)
SUM (Engineering & Technology, Arts & Humanities, Life Sciences & Medicine)
33.33%
33.33%
33.33%
Ran
School Name
Filter:
<None>
Country
Filter:
2 out of 43
Engineering & Technology
Arts & Humanities
Life Sciences & Medicine
1. University of Oxford United Kingdom
2. University of Cambridge United Kingdom
3. Harvard University United States
4. Stanford University United States
5. Massachusetts Institute of Technology (MIT United States
6. University of California, Berkeley (UCB) United States
7. University of California, Los Angeles (UCL United States
8. Yale University United States
9. UCL (University College London) United Kingdom
10. Columbia University United States
11. Princeton University United States
12. University of Edinburgh United Kingdom
13. University of Michigan United States
14. Cornell University United States
15. University of Pennsylvania United States
16. The University of Manchester United Kingdom
17. Imperial College London United Kingdom
18. University of Chicago United States

Compare Rankings

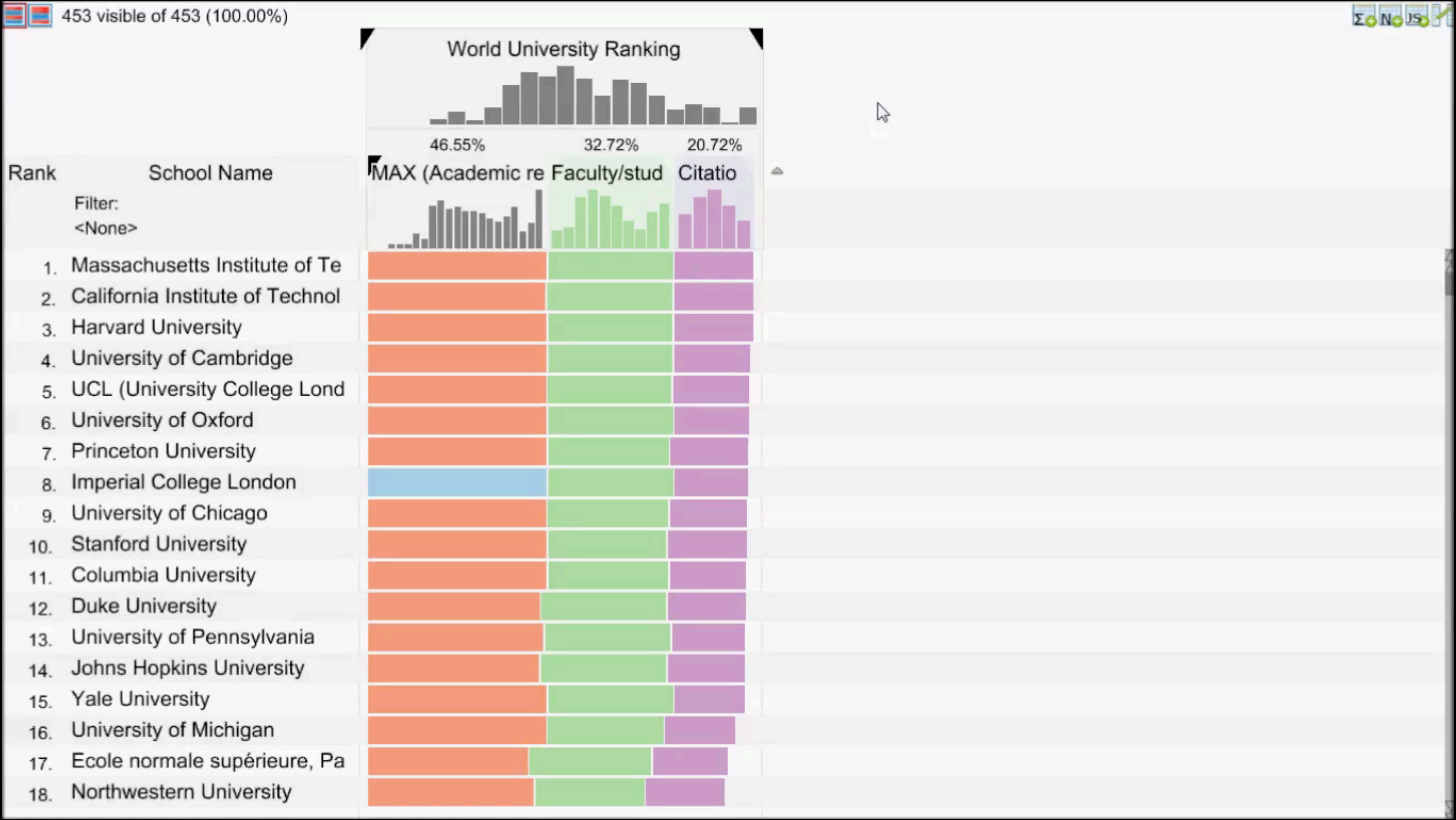
Bump Charts


Rank
University
Score
Score
1.
MIT, USA
2.
Harvard, USA
3.
Princeton, USA
4.
Cambridge, UK
5.
Oxford, UK
(+1)
(+1)
(-2)

Bump Charts



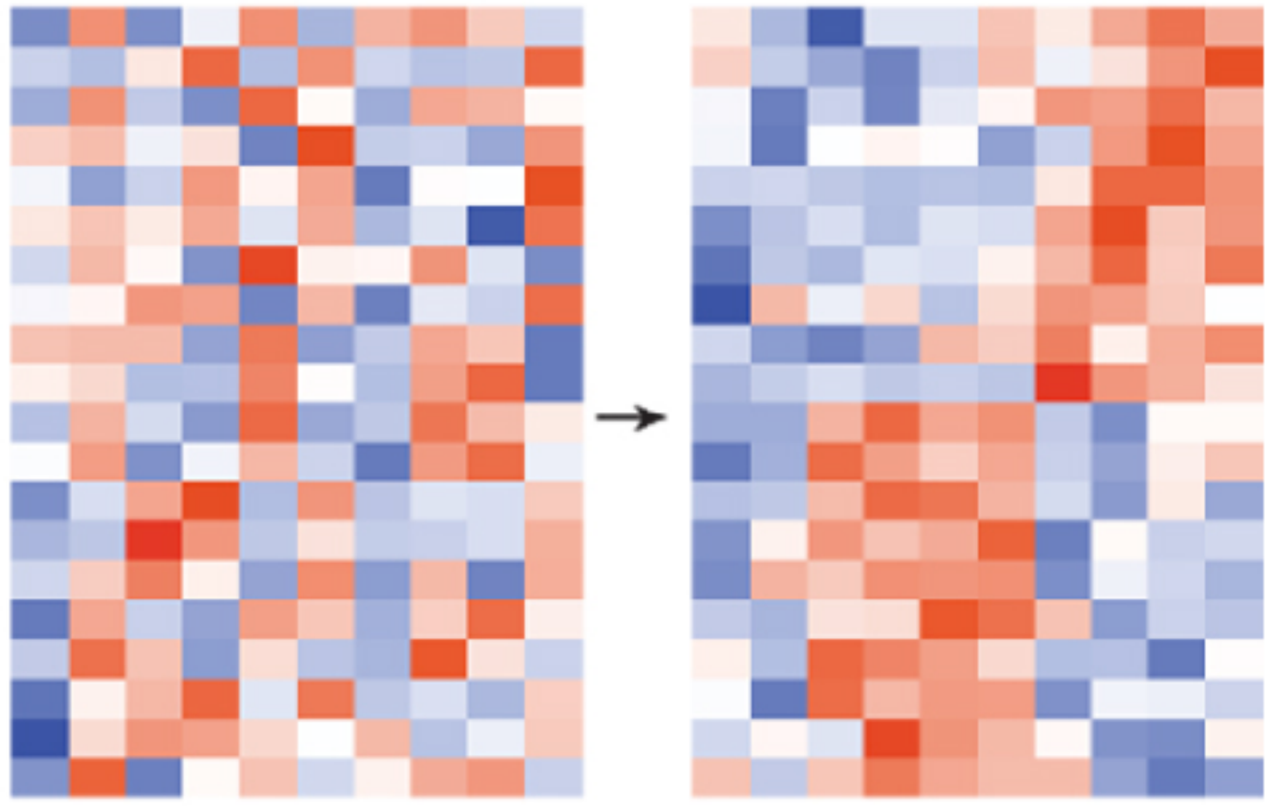
Rank
University
Score
1.
MIT, USA
2.
Harvard, USA
3.
Princeton, USA
4.
Cambridge, UK
5.
Oxford, UK
Rank
Score
1.
(+1)2.
(+1)3.
(-2)4.
5.
Score



453 visible of 453 (100.00%)
World University Ranking
46.55%
32.72%
20.72%
Rank
School Name
Filter:
<None>
MAX (Academic re
Faculty/stud
Citatio
1. Massachusetts Institute of Te
2. California Institute of Technol
3. Harvard University
4. University of Cambridge
5. UCL (University College Lond
6. University of Oxford
7. Princeton University
8. Imperial College London
9. University of Chicago
10. Stanford University
11. Columbia University
12. Duke University
13. University of Pennsylvania
14. Johns Hopkins University
15. Yale University
16. University of Michigan
17. Ecole normale supérieure, Pa
18. Northwestern University

http://lineup.caleydo.org

Pixel Based Displays



Each cell is a “pixel”, value encoded in color / value

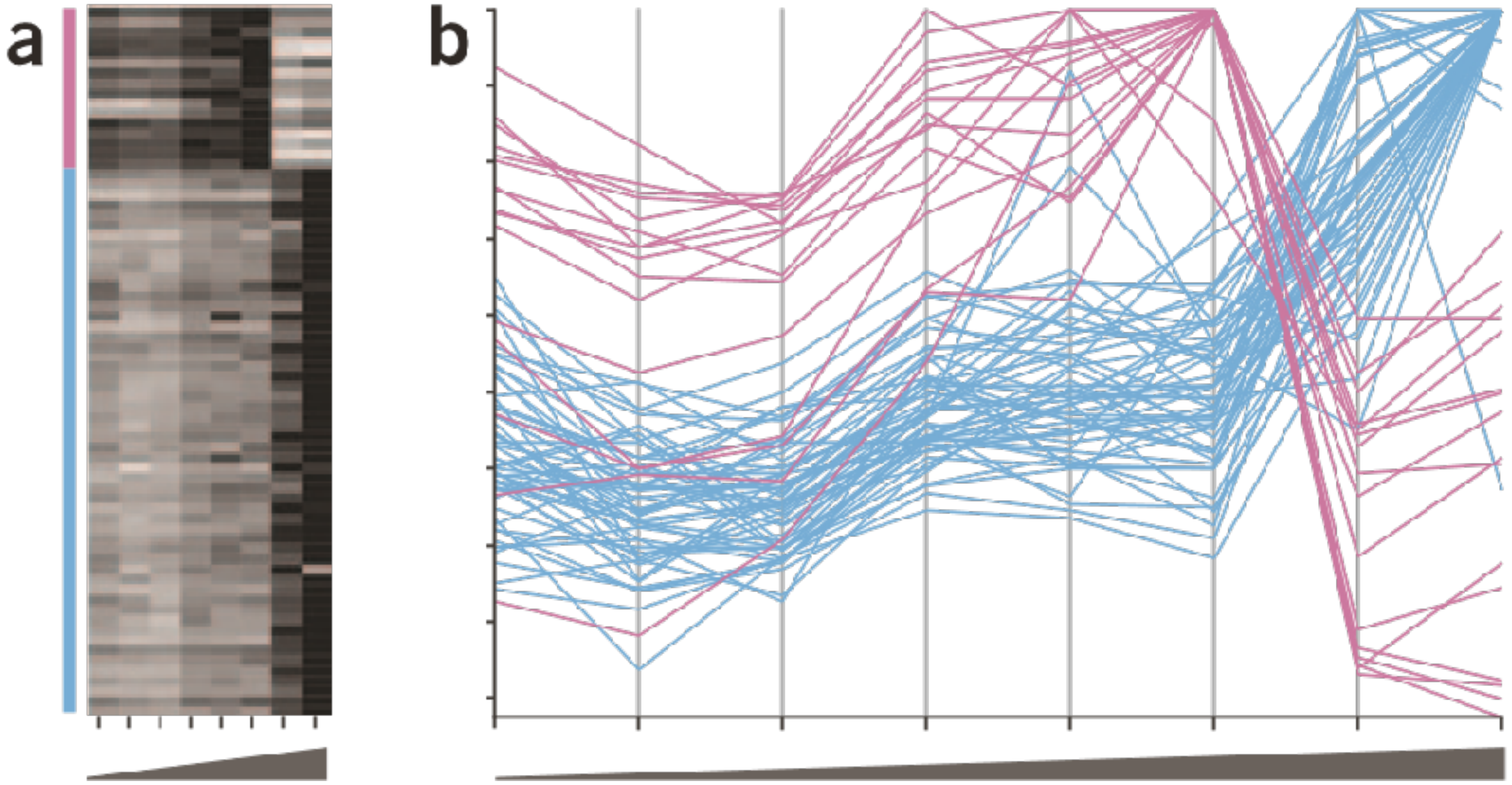
Ordering critical for interpretation

If no ordering inherent, clustering is used

Scalable – 1 px per item

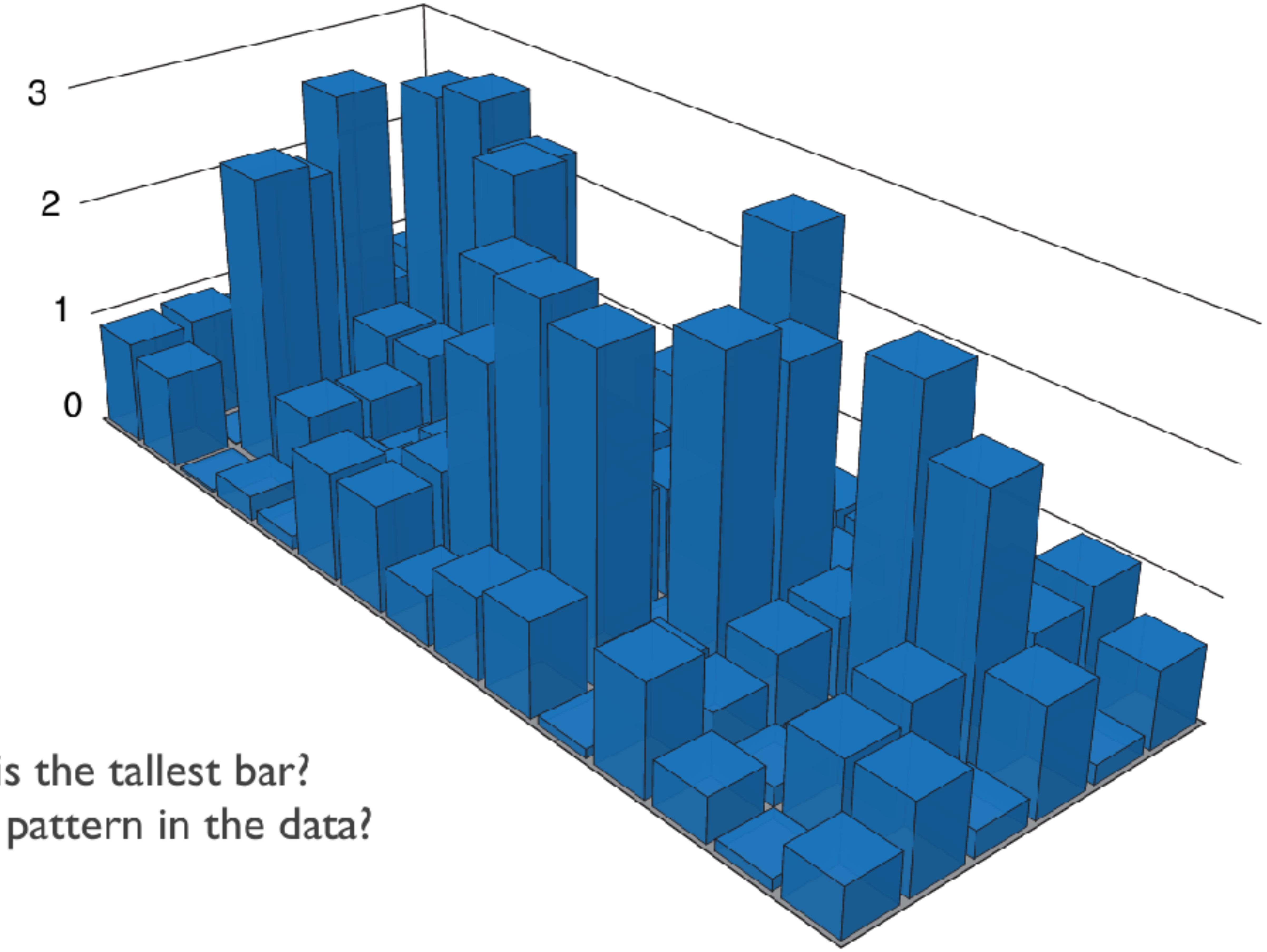
Good for homogeneous data

same scale & type



[Gehlenborg & Wong 2012]

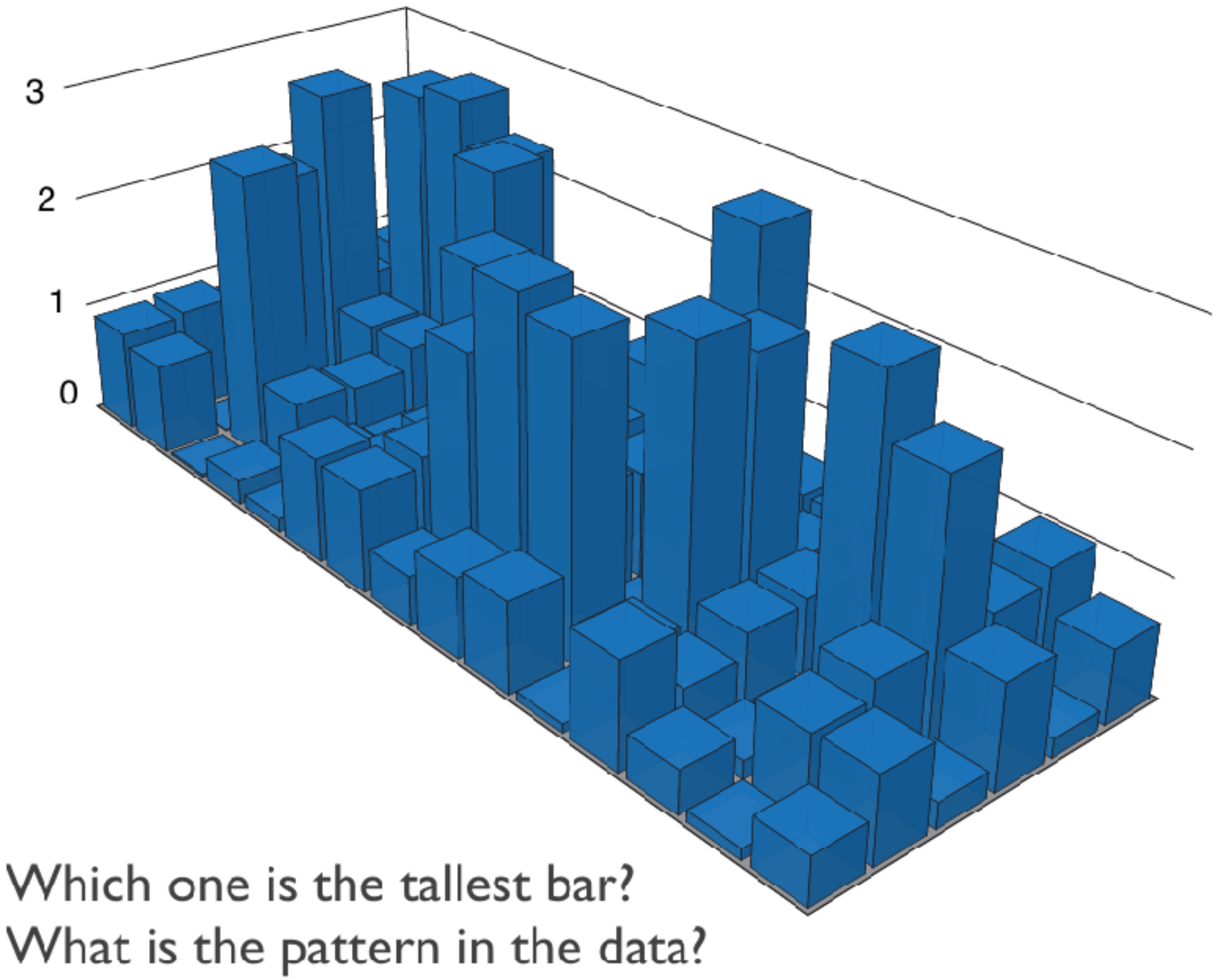
3D Pitfall: Occlusion & Perspective

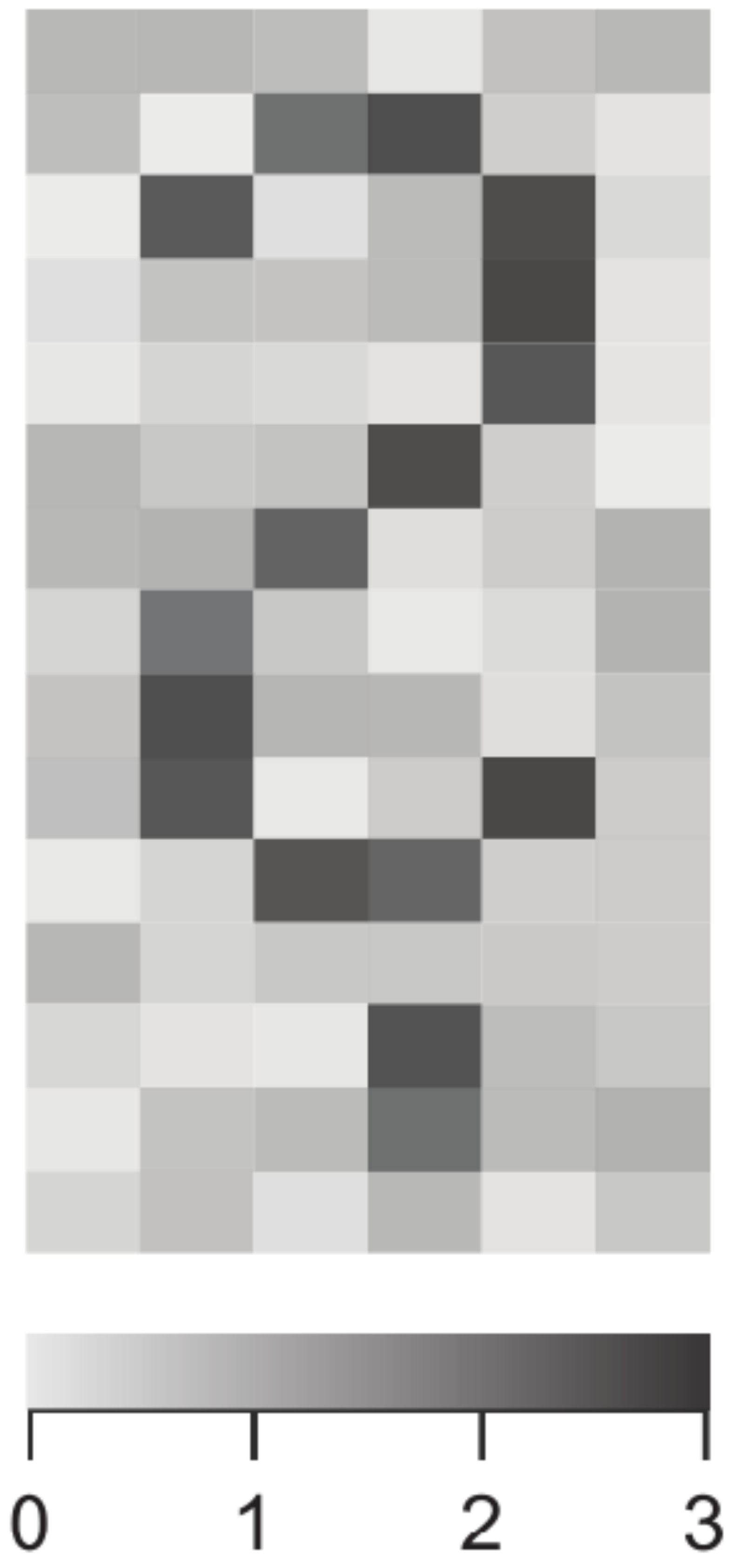


Which one is the tallest bar?
What is the pattern in the data?

[Gehlenborg and Wong, Nature Methods, 2012]

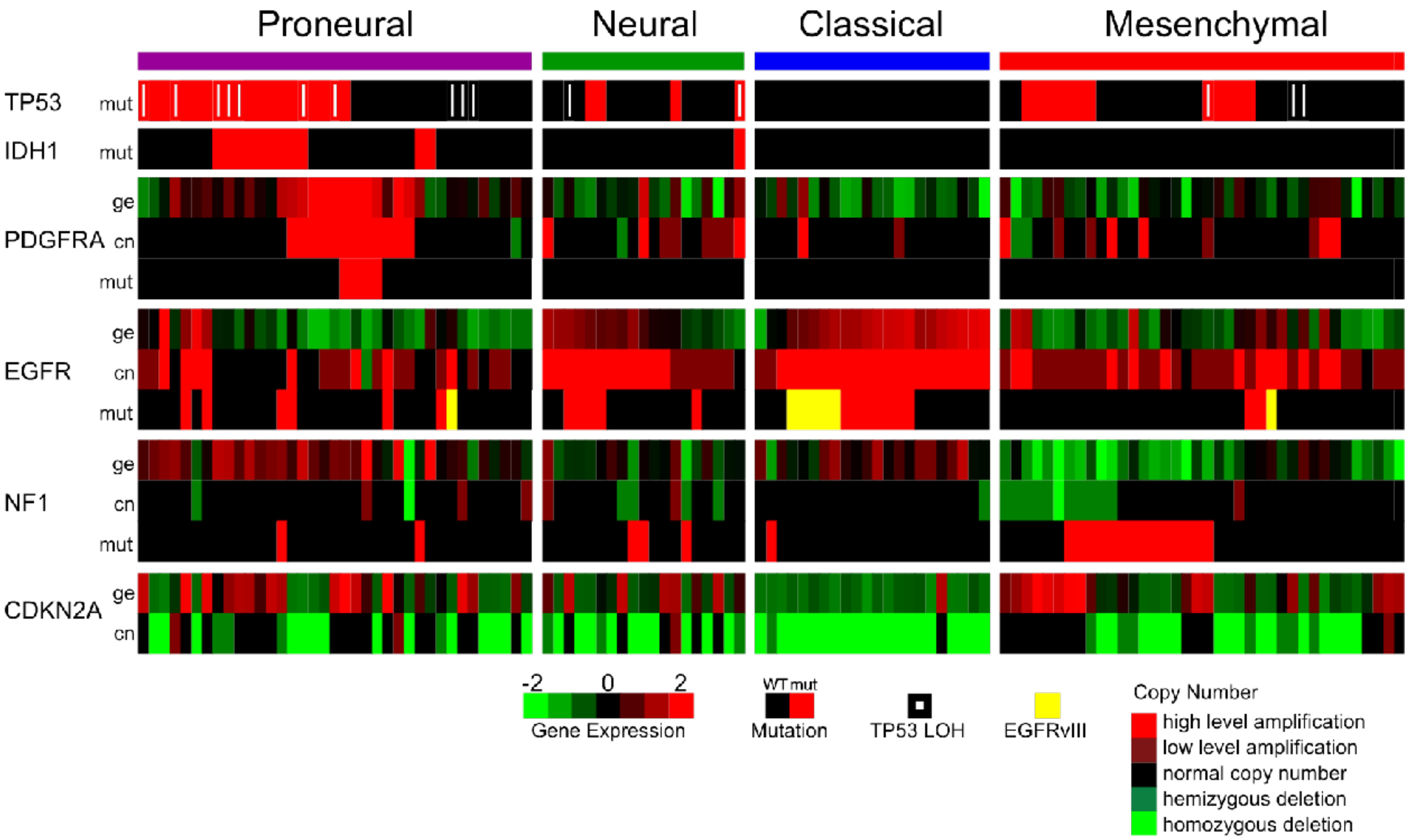
3D Pitfall: Occlusion & Perspective





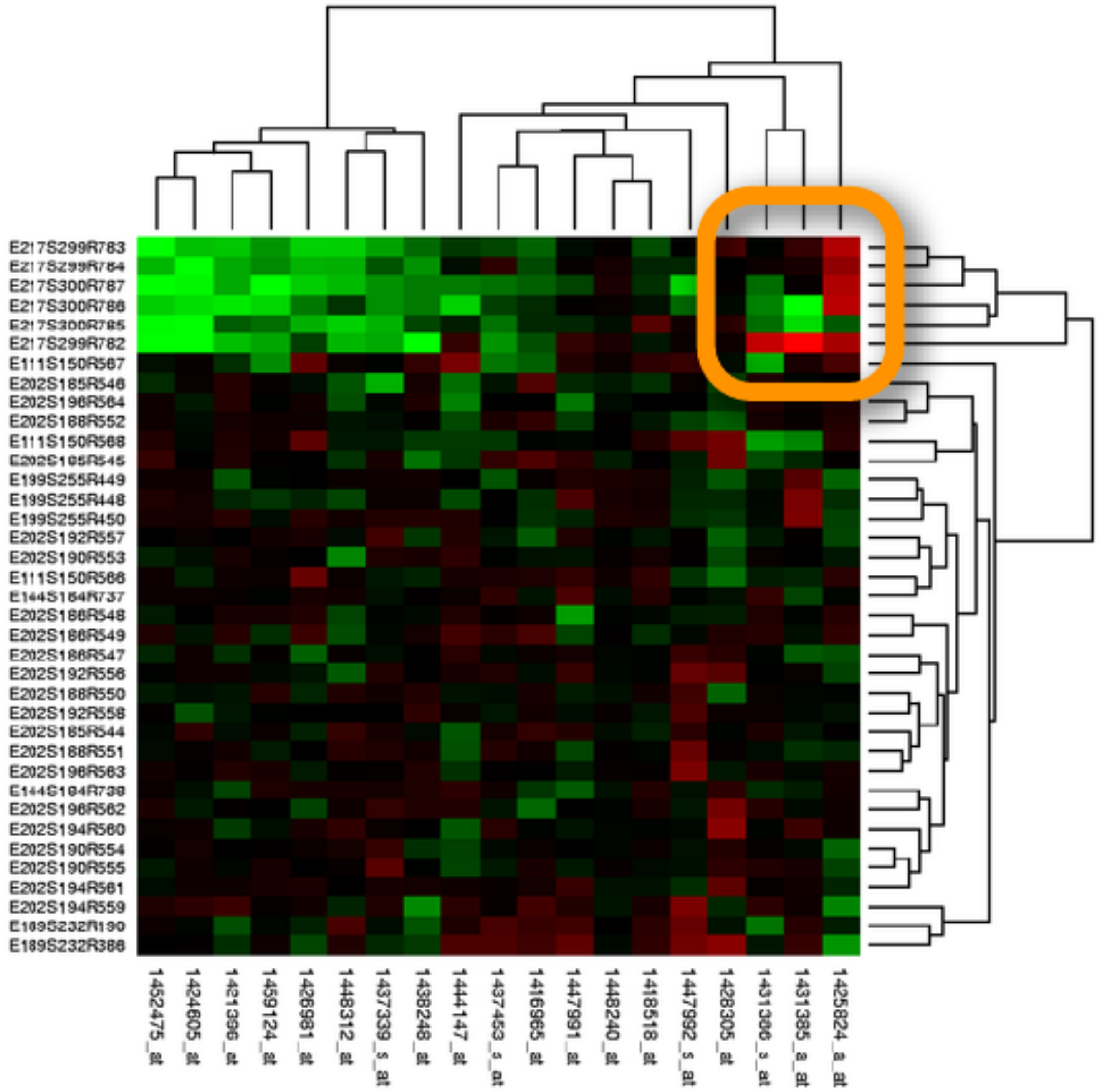
[Gehlenborg and Wong, Nature Methods, 2012]

Heterogeneous Data?

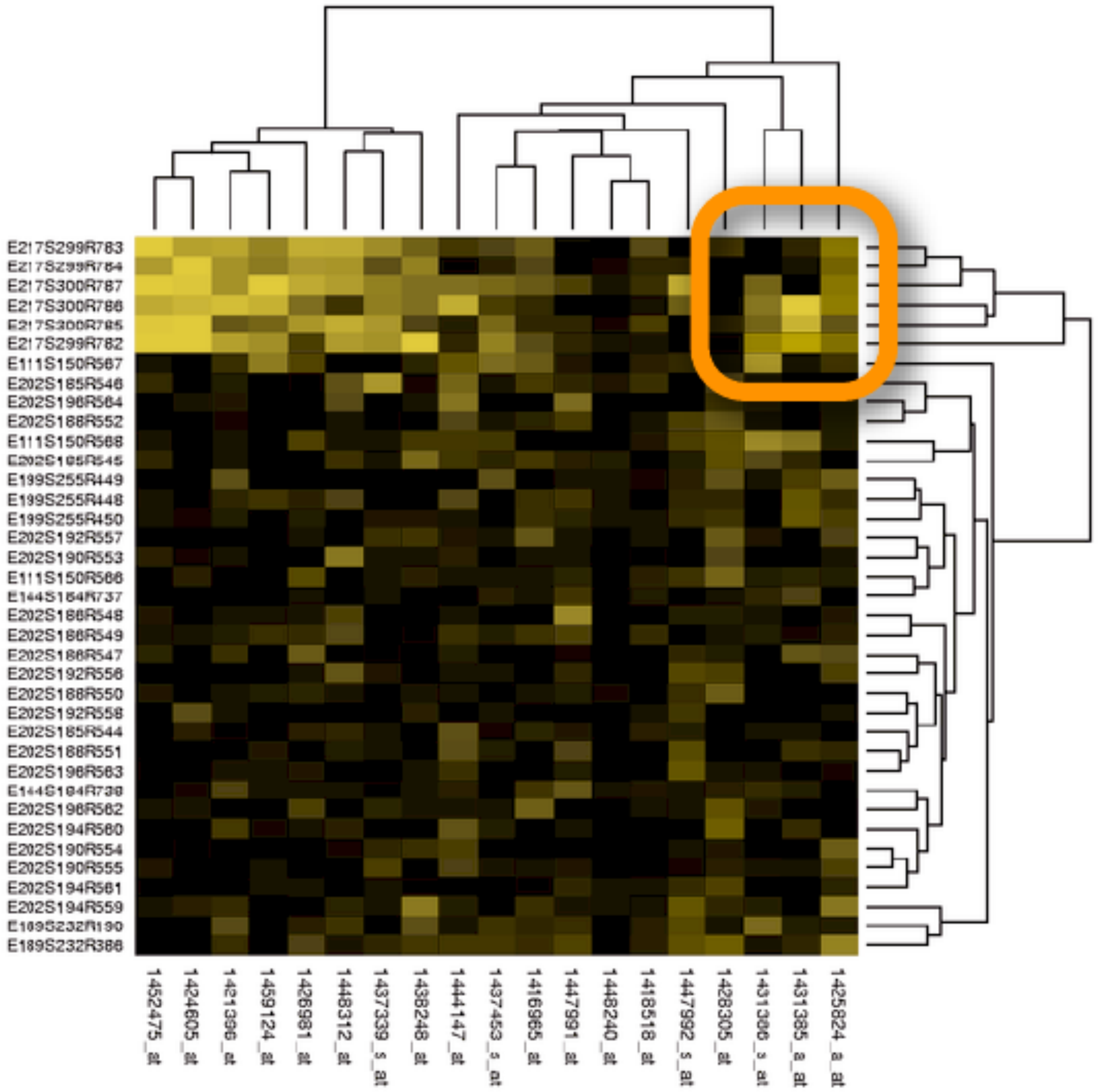


[Verhaak 2012]

Bad Color Mapping

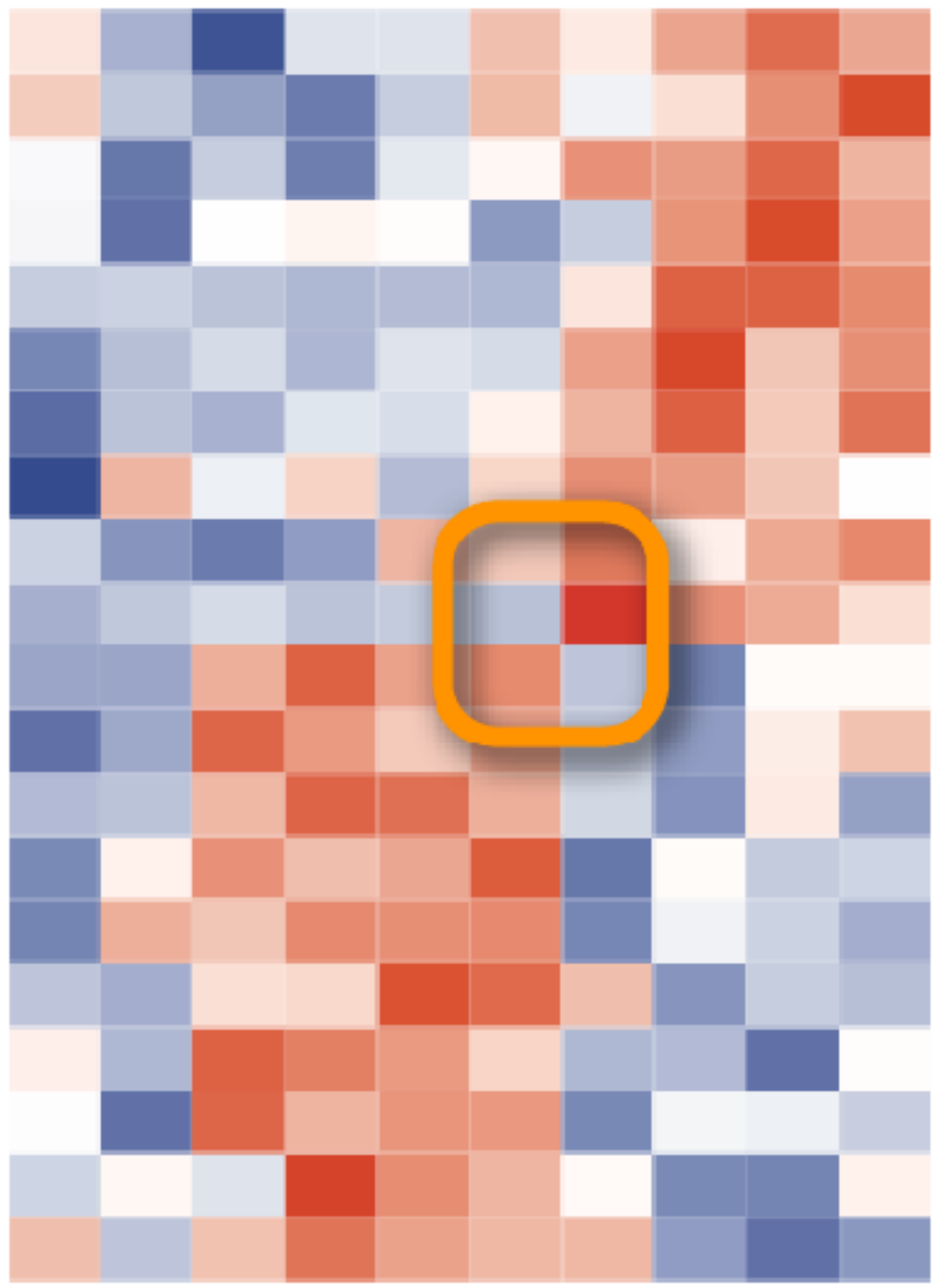


Normal Vision

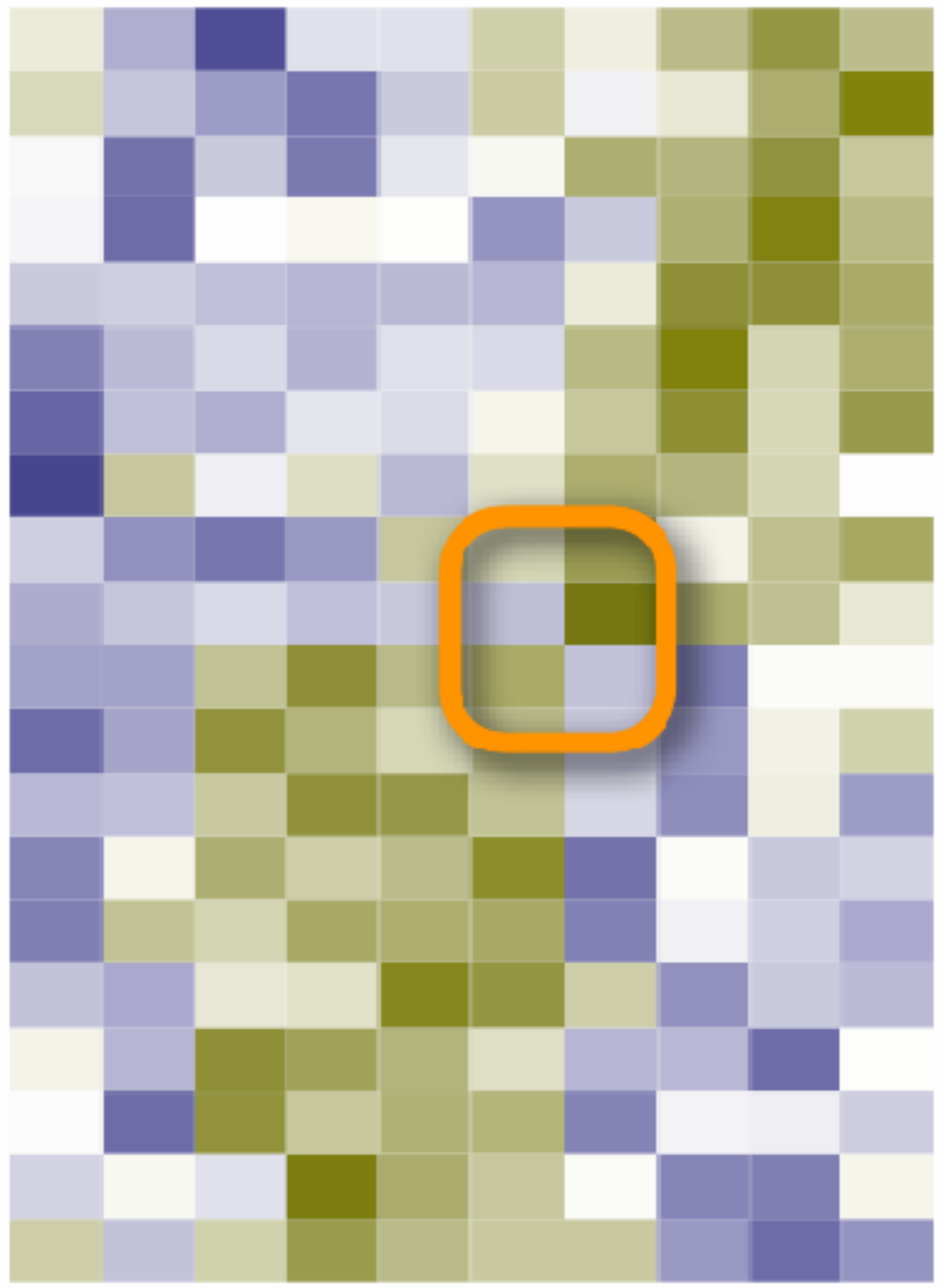


Deuteranope Vision
("Red-Green Blindness")

Good Color Mapping

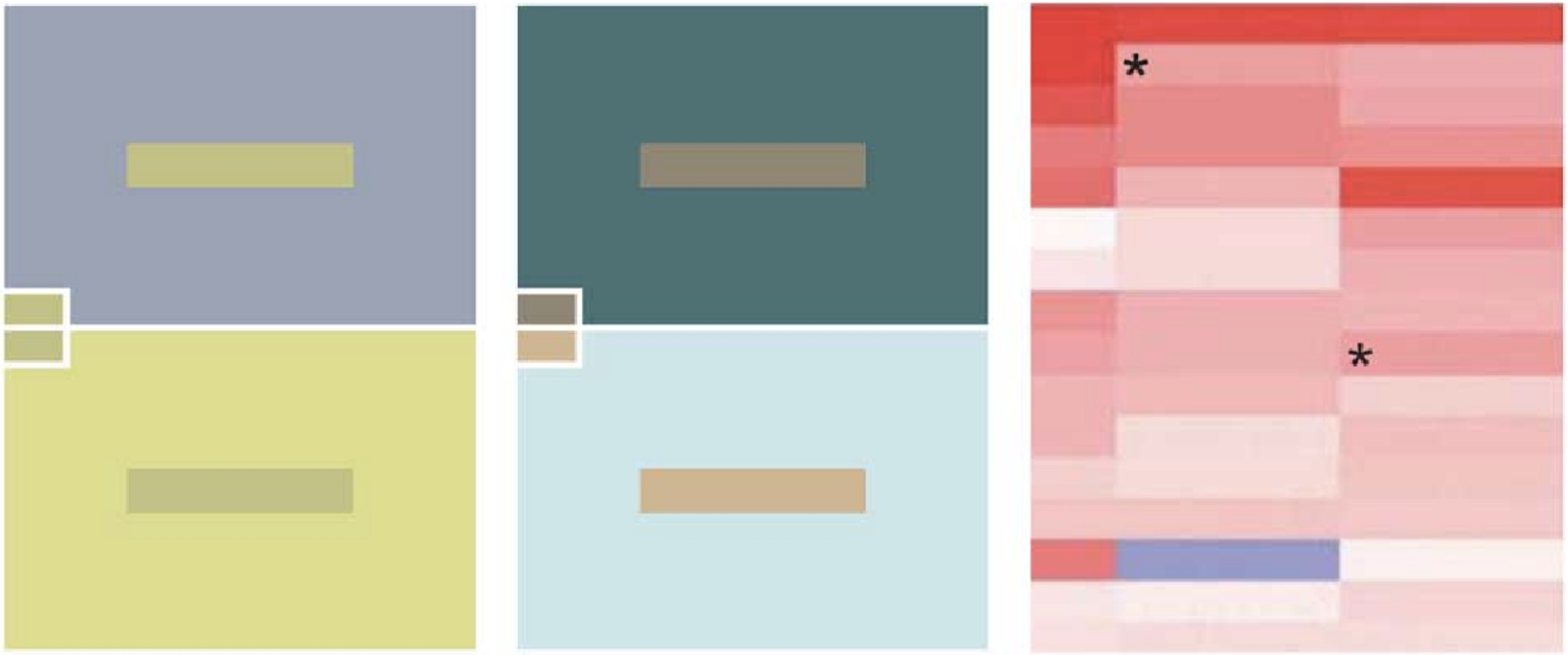


Normal Vision

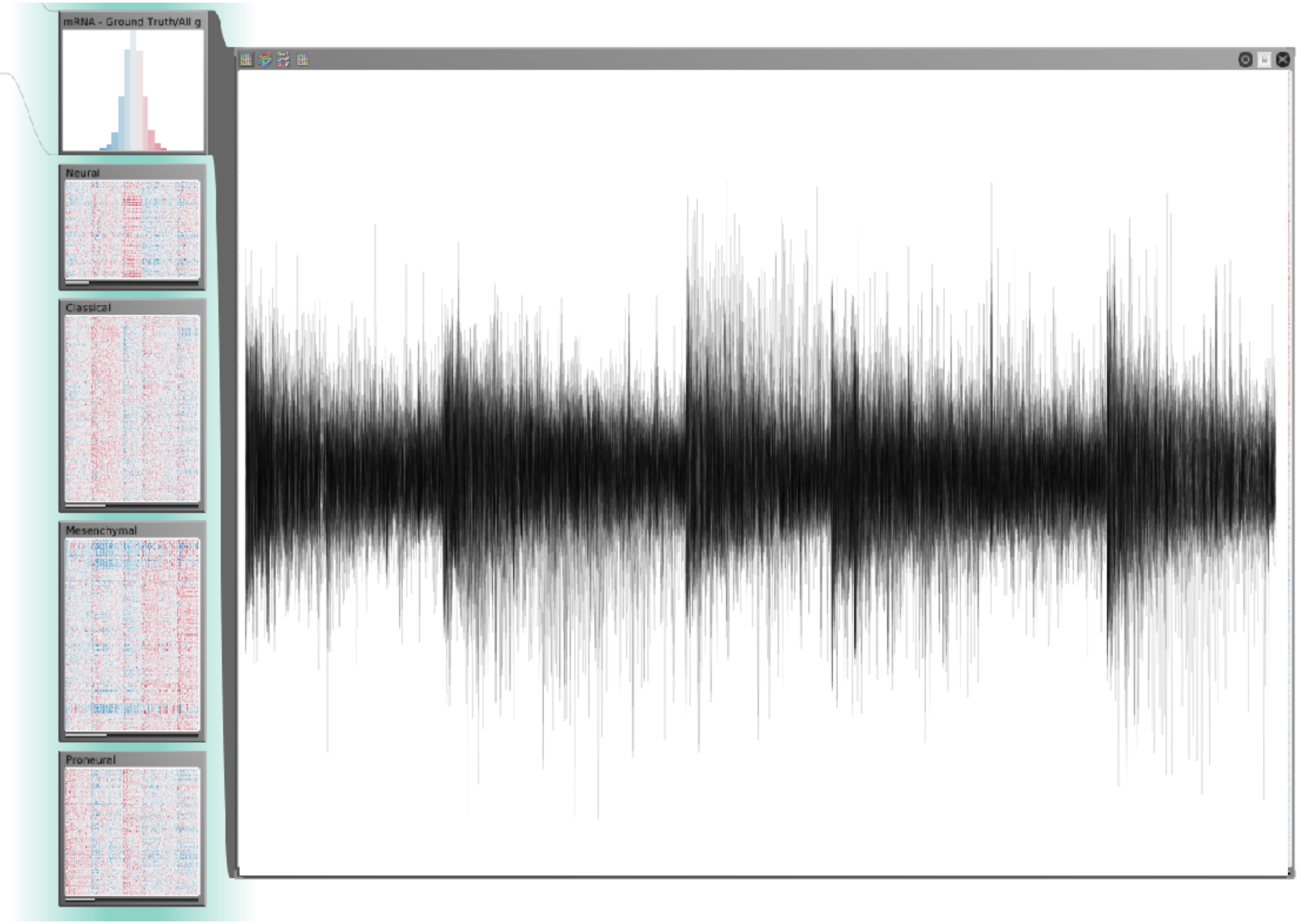


Deuteranope Vision
("Red-Green Blindness")

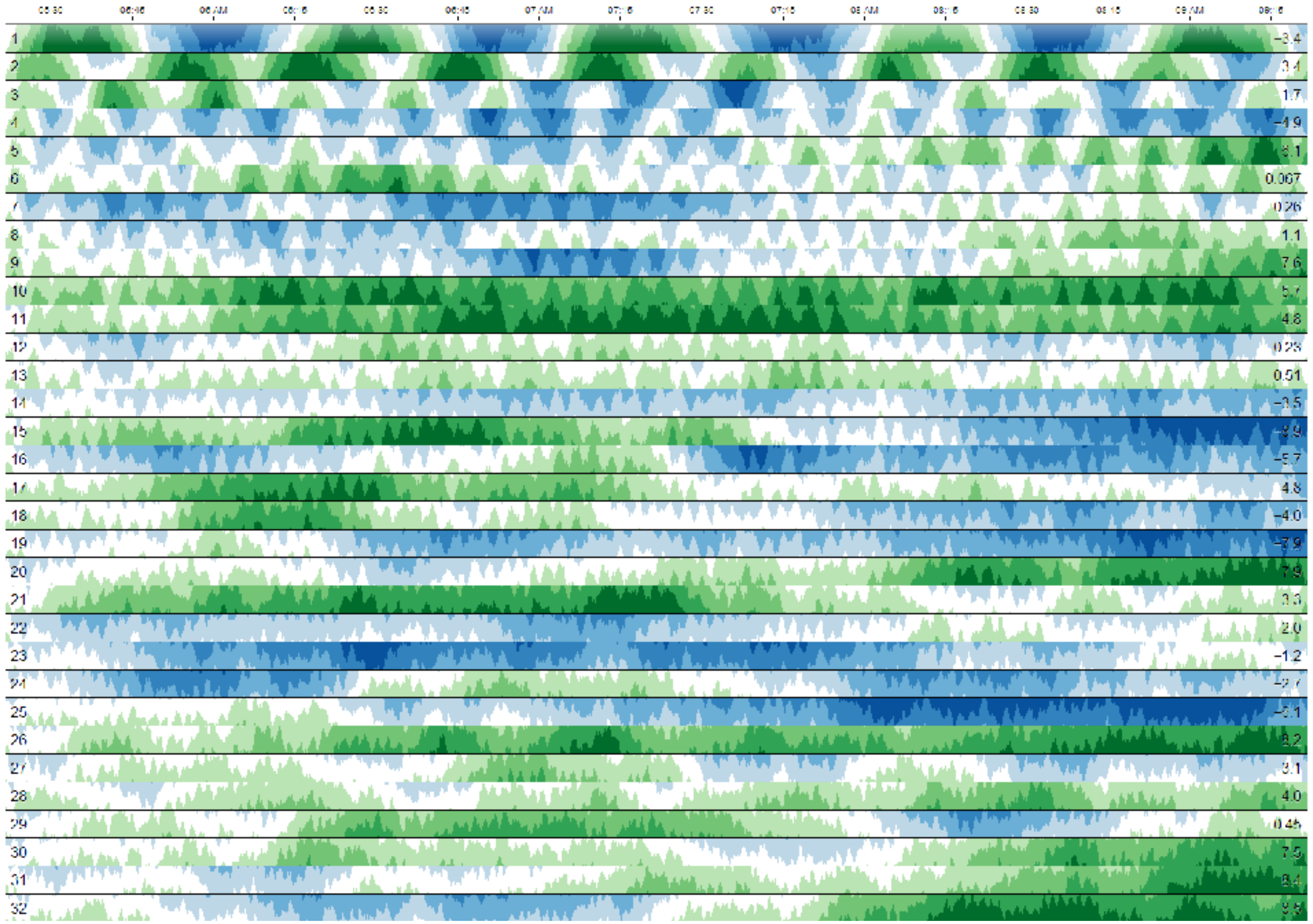
Color is relative!



Clustered Heat Map

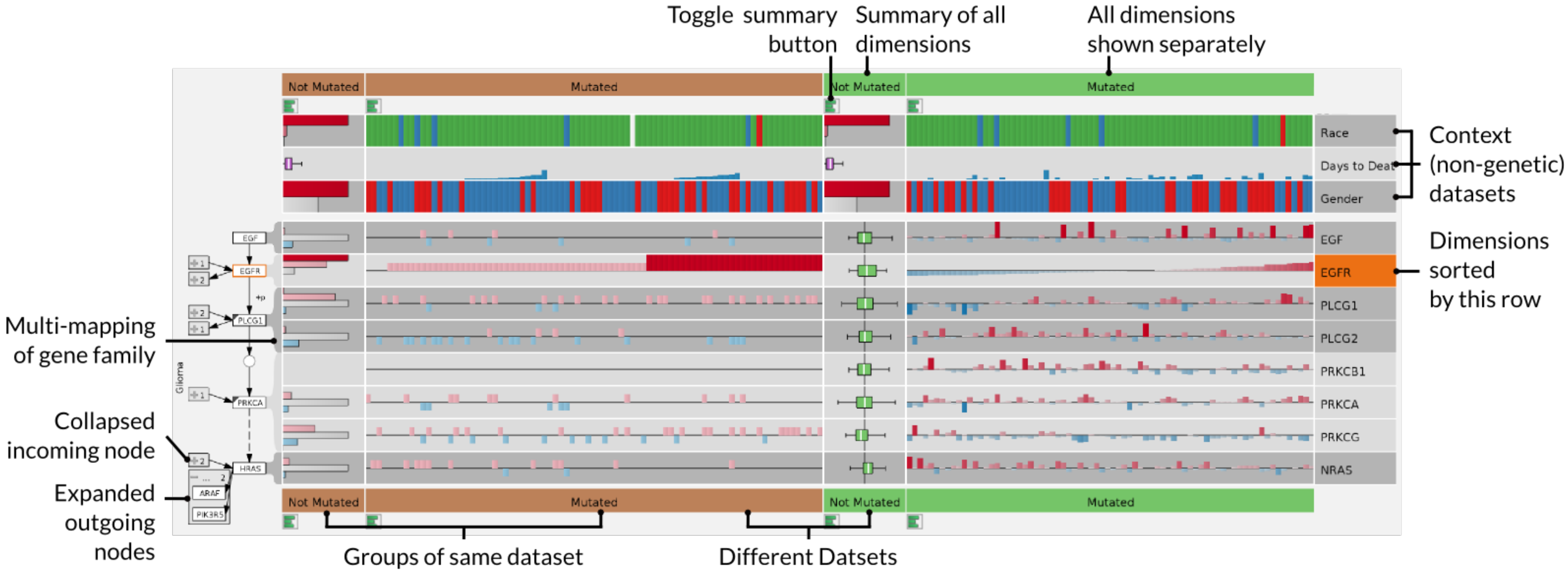


Multiple Line Charts

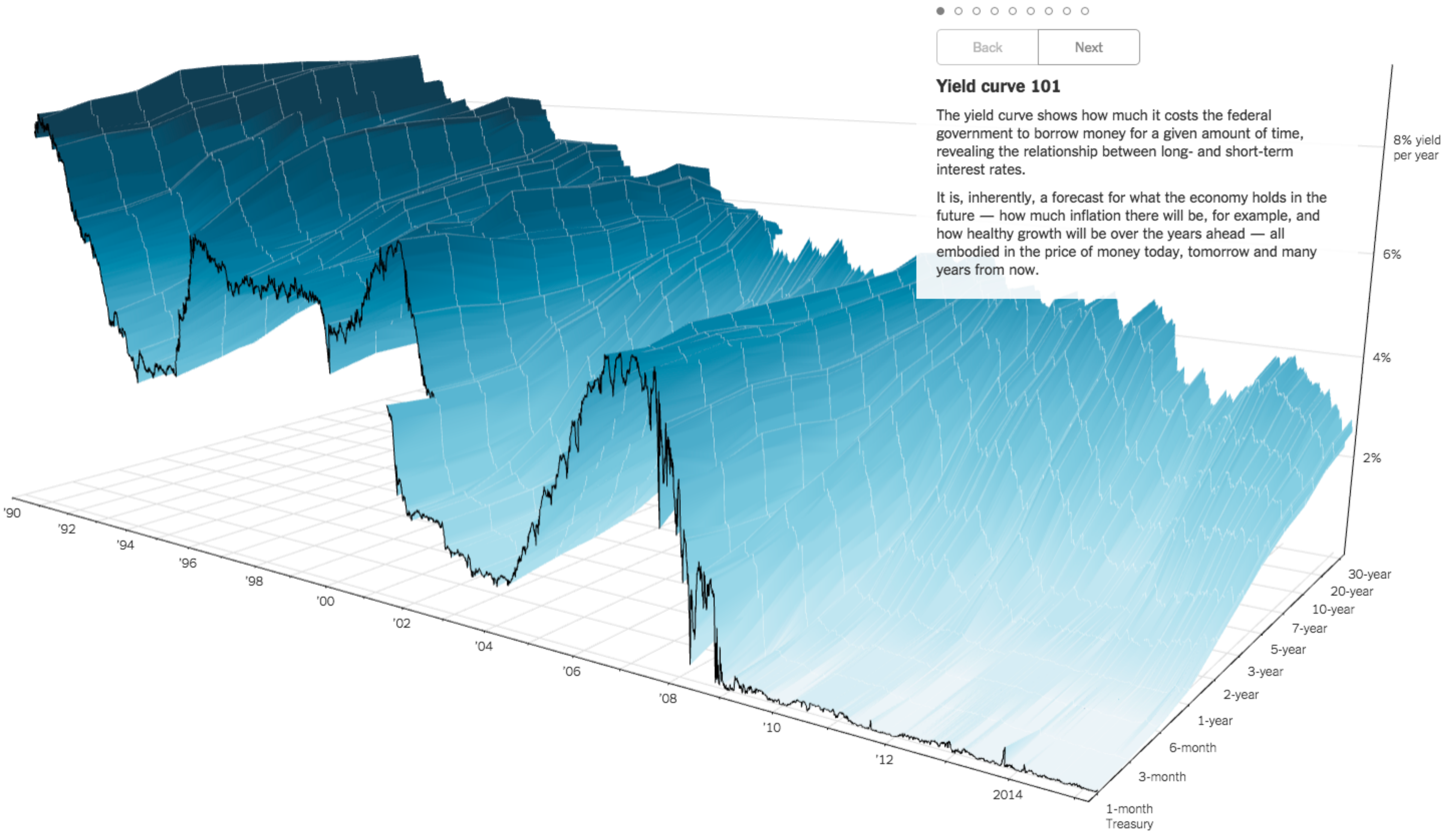


http://square.github.io/cubism/

Combining Various Charts



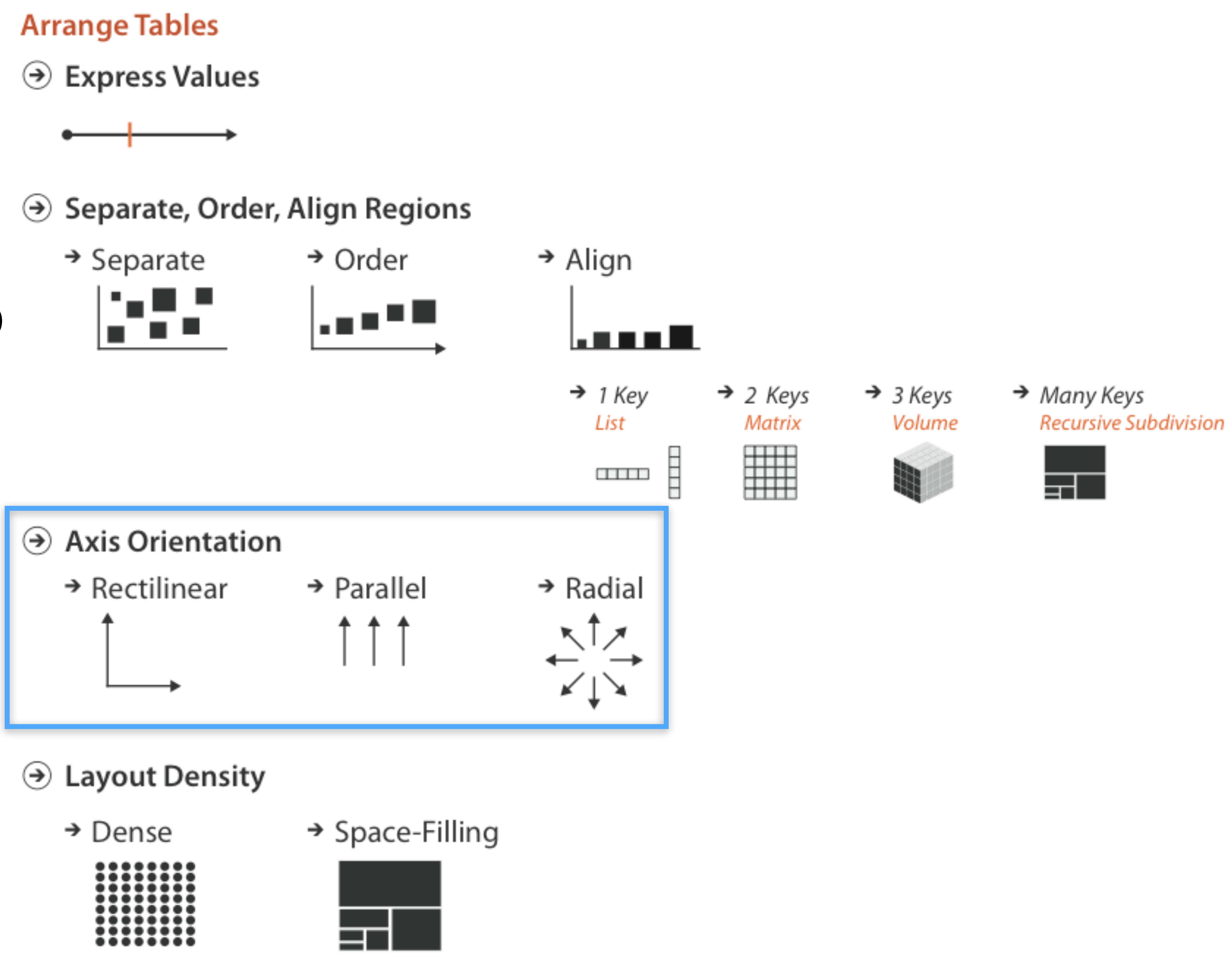
Design Critique



Document: https://goo.gl/W6w0iI
Website: http://goo.gl/D3mIsy

Spatial Axis Orientation

s



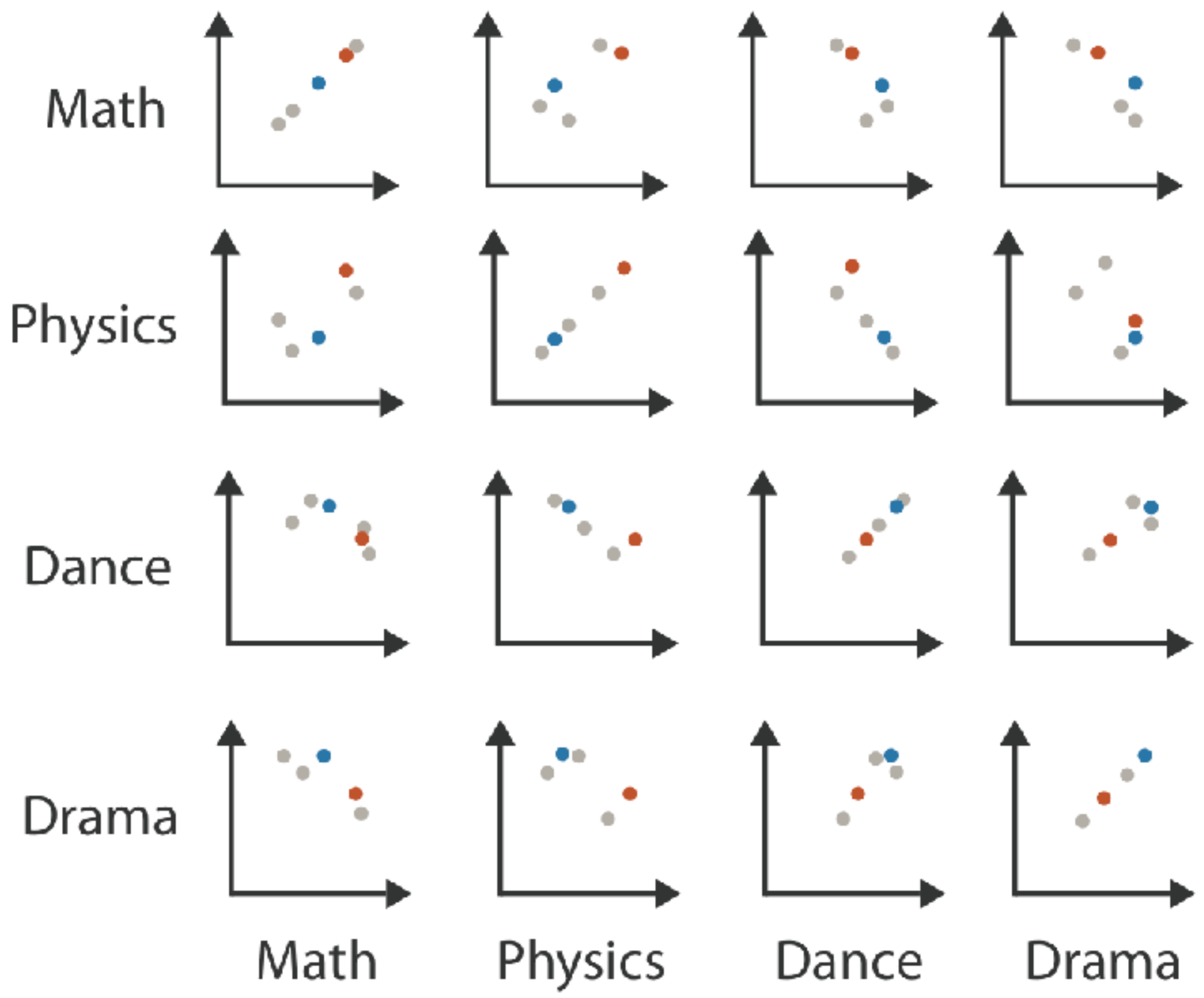
Arrange Tables
Express Values
Separate, Order, Align Regions
Separate
Order
Align
1 Key
List
2 Keys
Matrix
3 Keys
Volume
Many Keys
Recursive Subdivision
Axis Orientation
Rectilinear
Parallel
Radial
Layout Density
Dense
Space-Filling

n

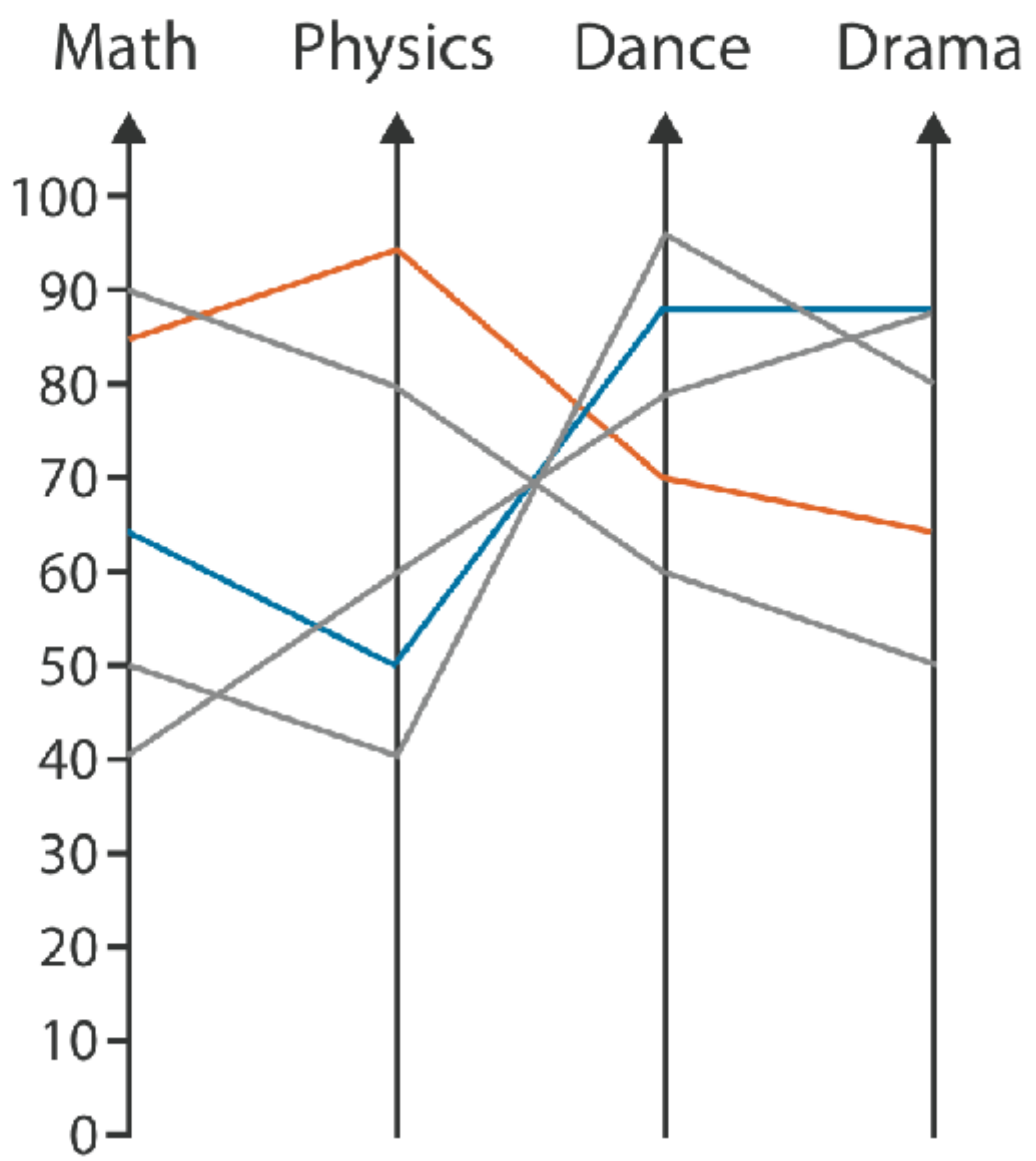
Table

Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

Scatterplot Matrix



Parallel Coordinates



Spatial Axis Orientation

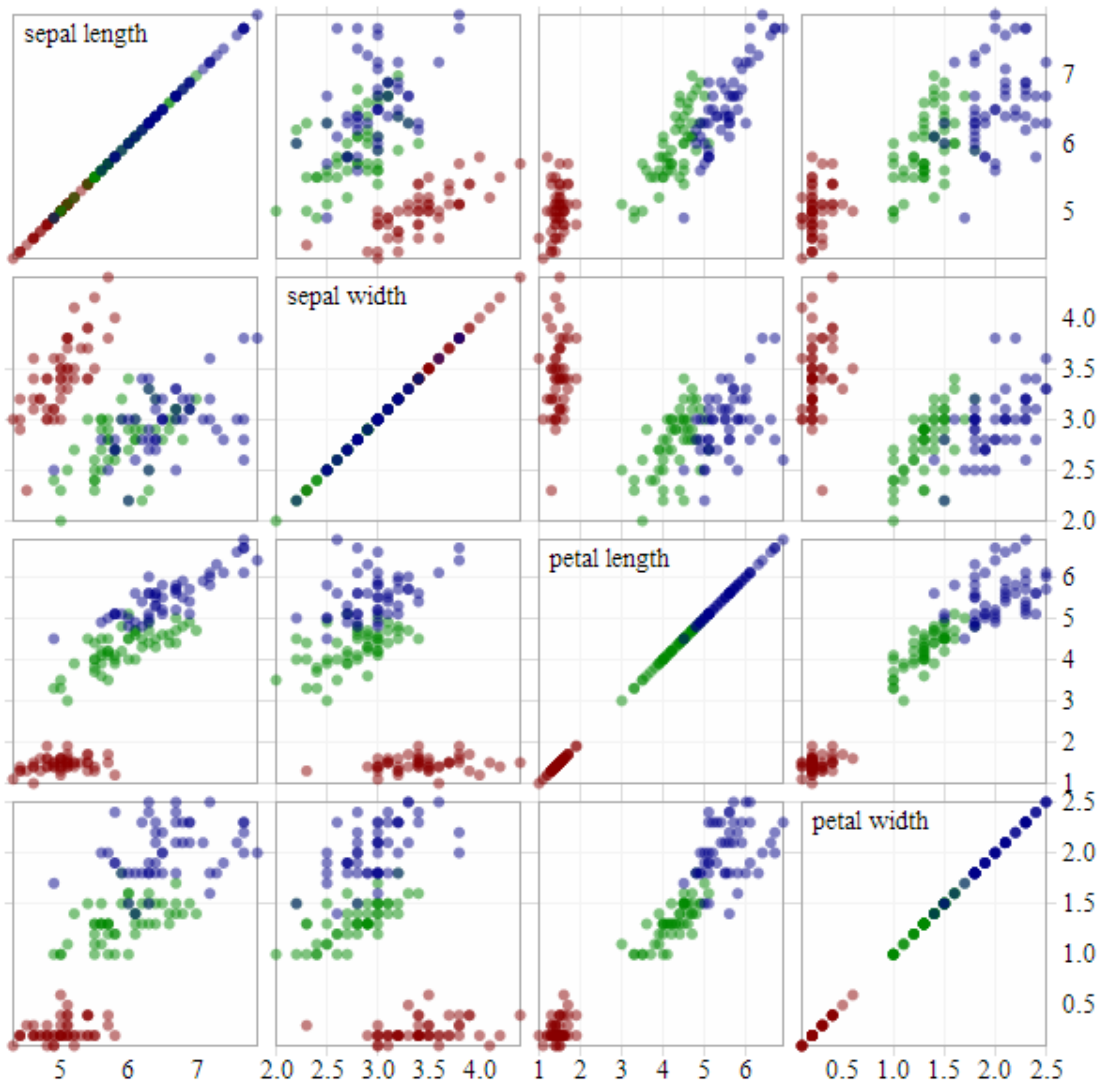
Scatterplot Matrix

Scatterplot Matrices (SPLOM)

Matrix of size d*d

Each row/column is one dimension

Each cell plots a scatterplot of two dimensions



Scatterplot Matrices

Limited scalability (~20 dimensions, ~500-1k records)

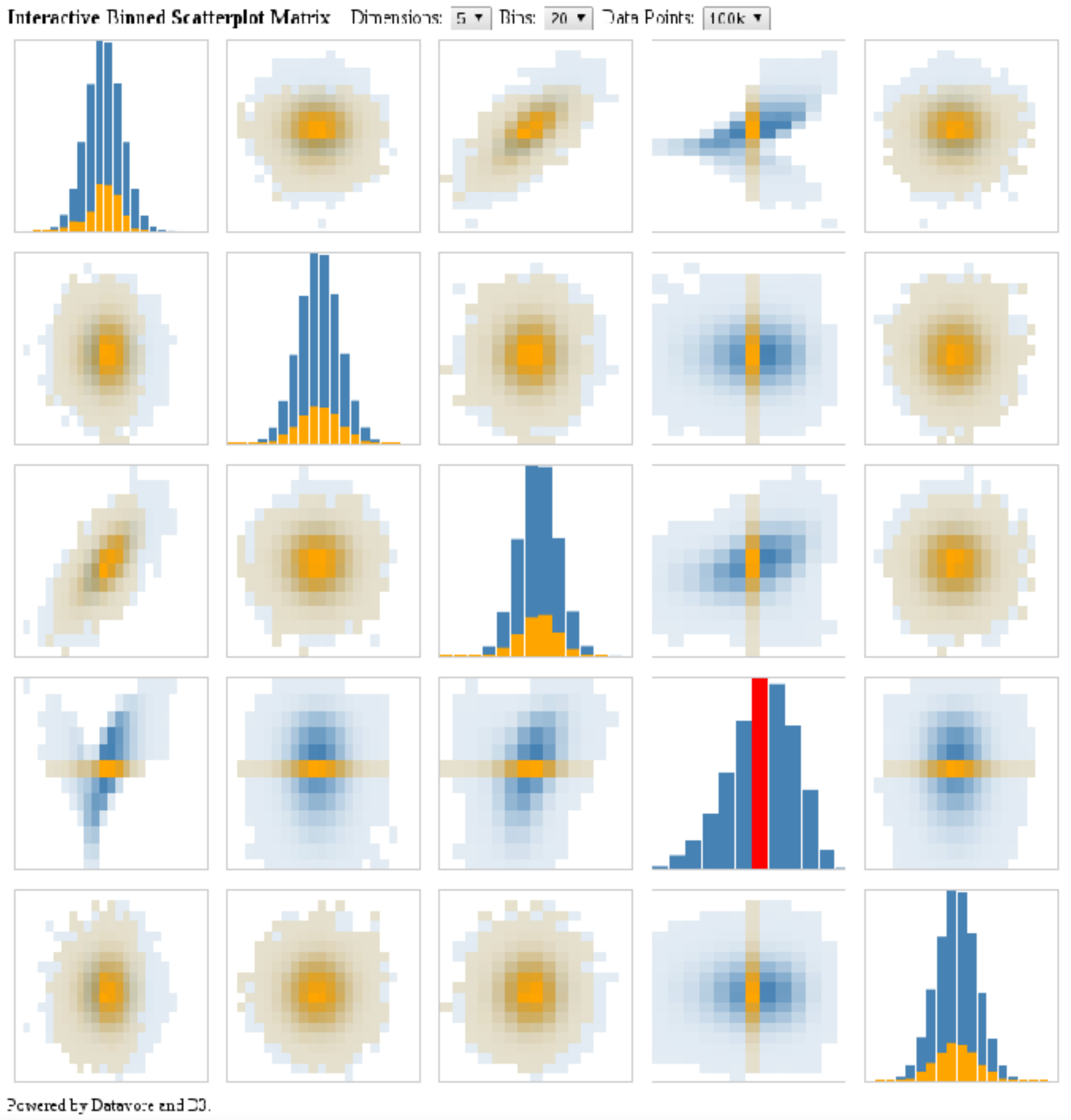
Brushing is important

Often combined with “Focus Scatterplot” as F+C technique

Algorithmic approaches:

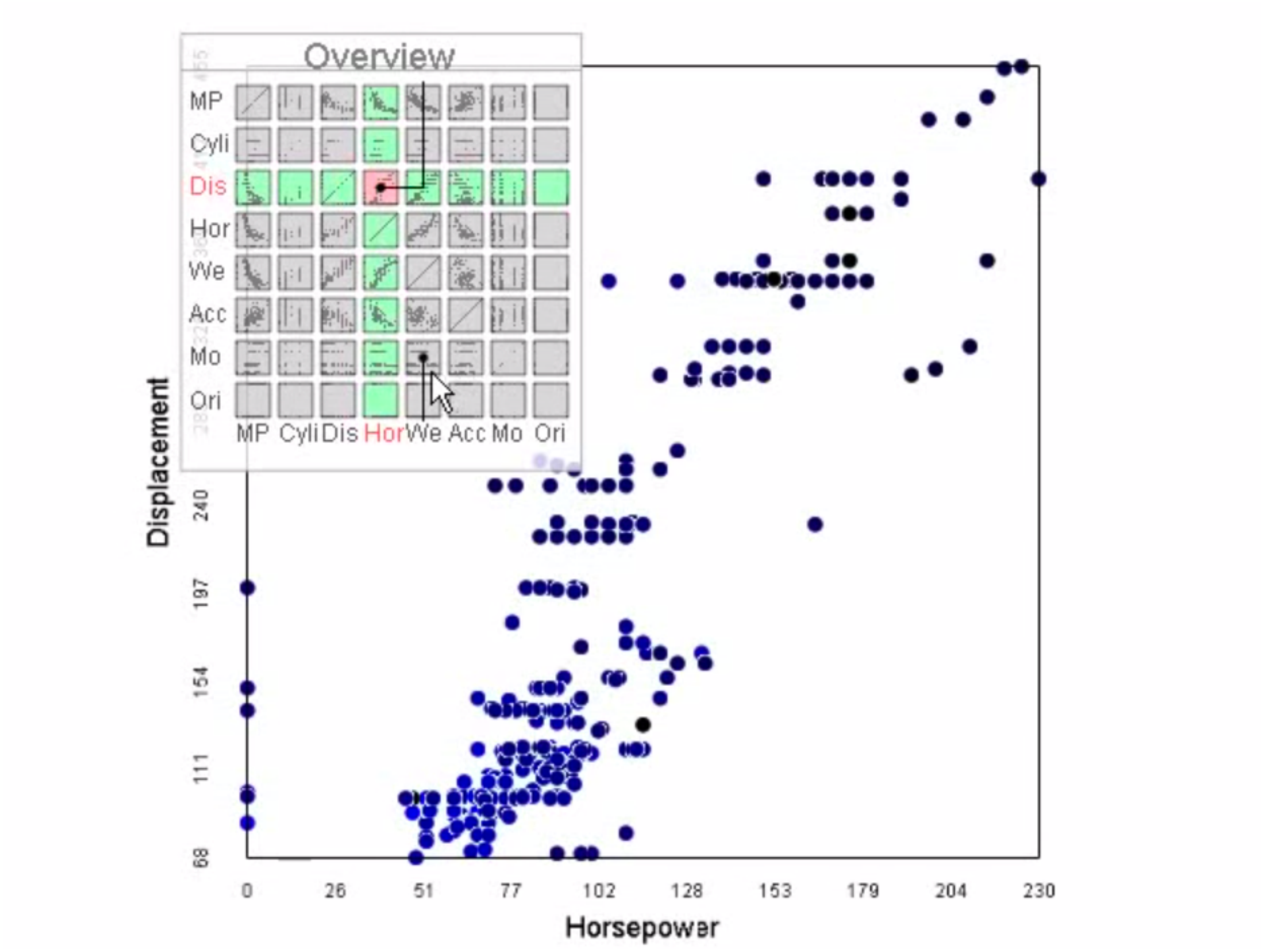
- Clustering & aggregating records
- Choosing dimensions
- Choosing order

SPLOM Aggregation - Heat Map



Datavore: http://vis.stanford.edu/projects/datavore/splom/

SPLOM F+C, Navigation



[Elmqvist]

Spatial Axis Orientation

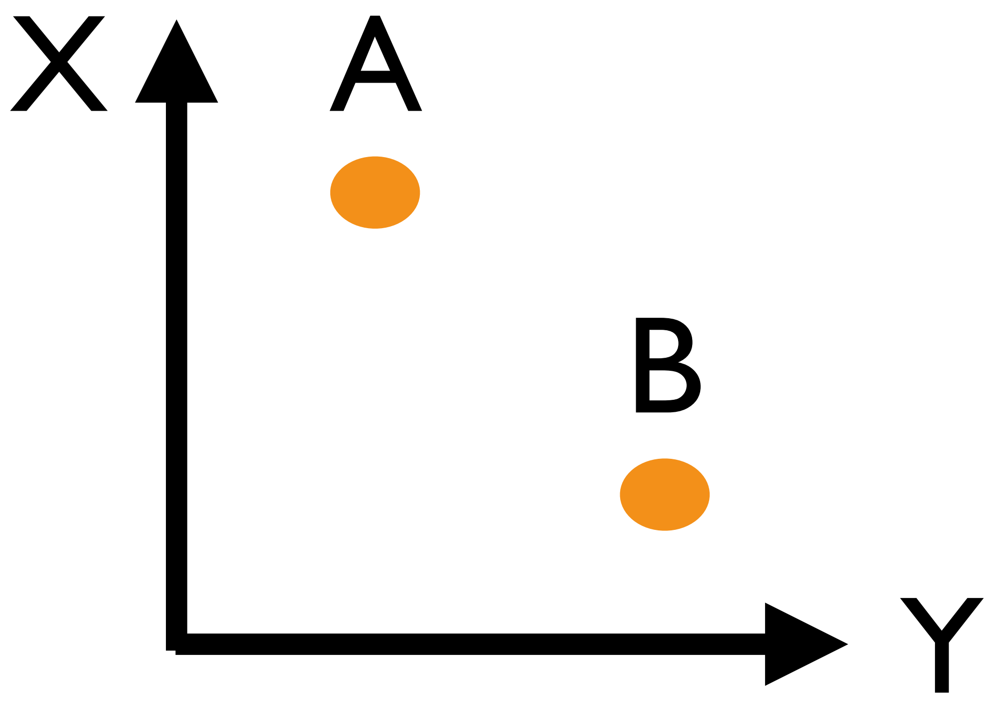
Parallel Coordinates

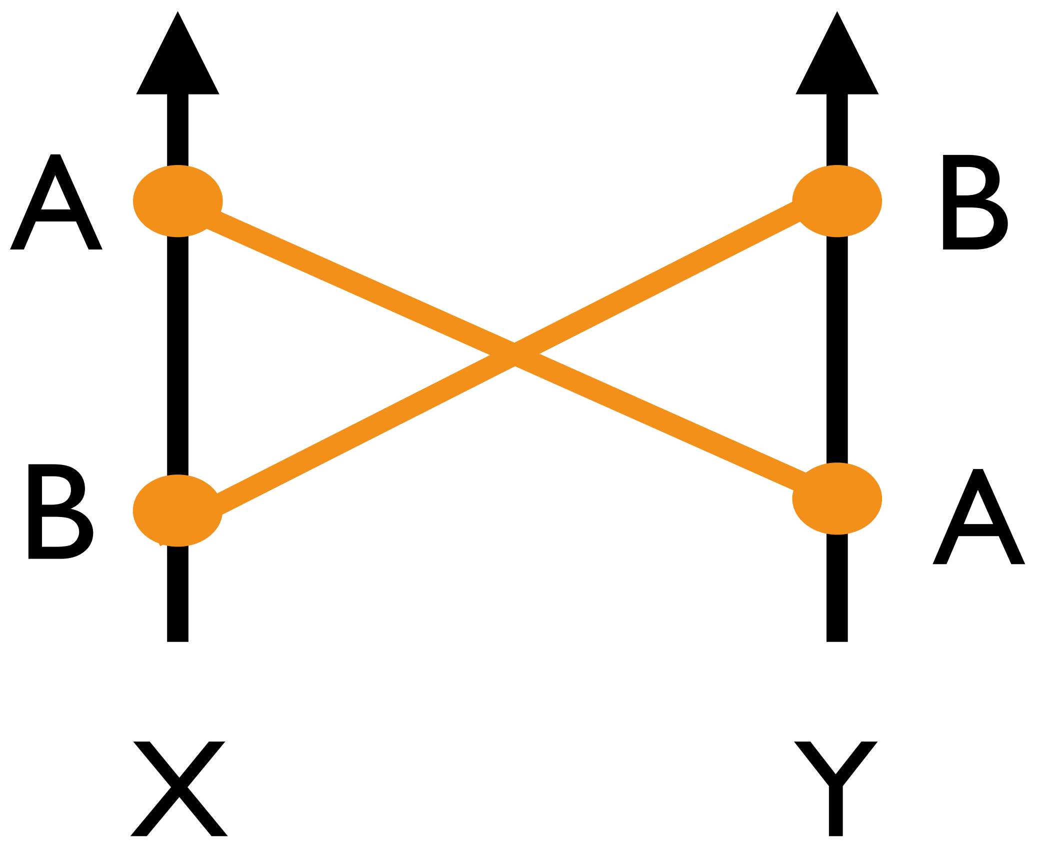
Parallel Coordinates (PC)

Inselberg 1985

Axes represent attributes

Lines connecting axes represent items





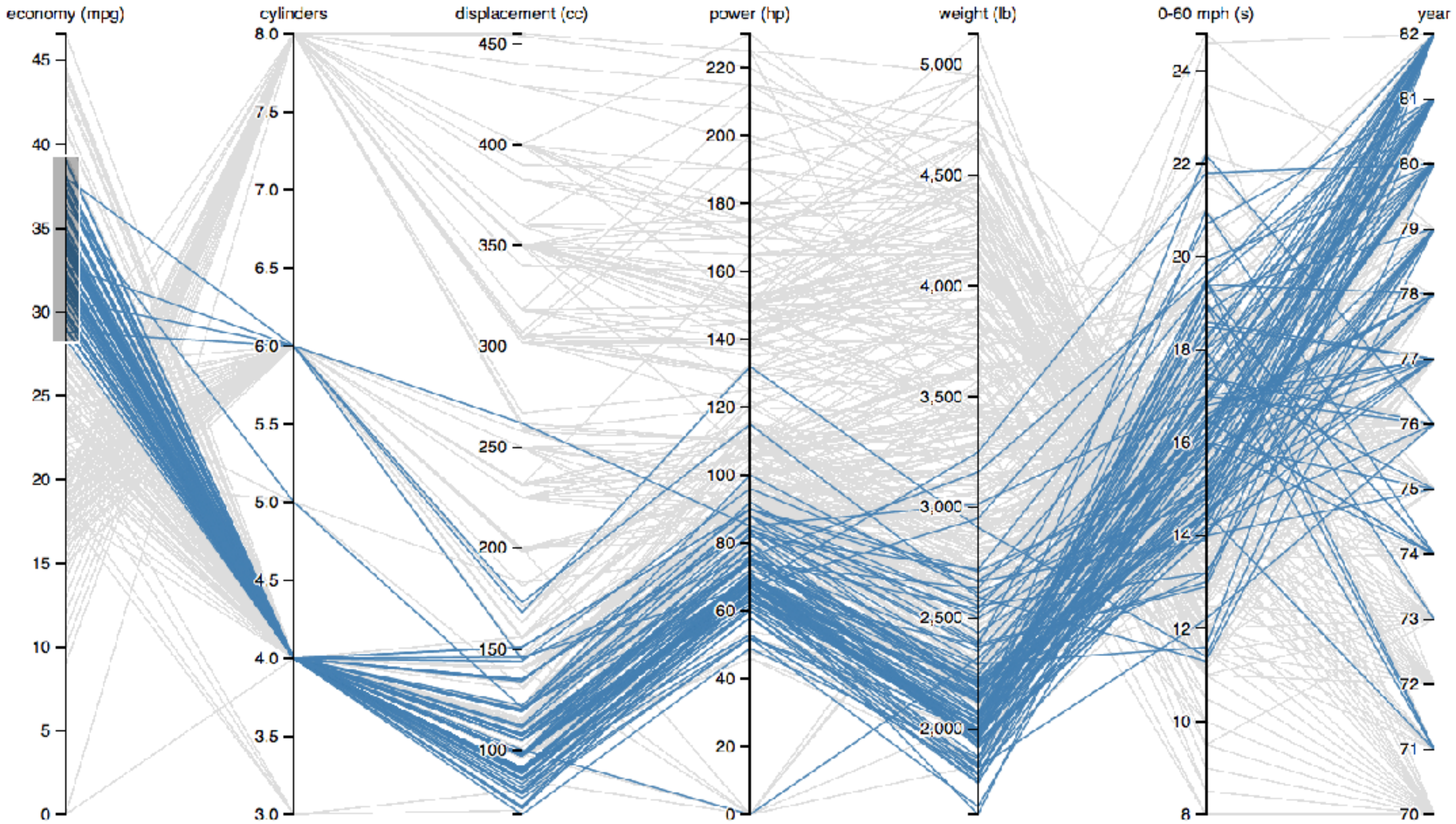
Parallel Coordinates

Each axis represents dimension

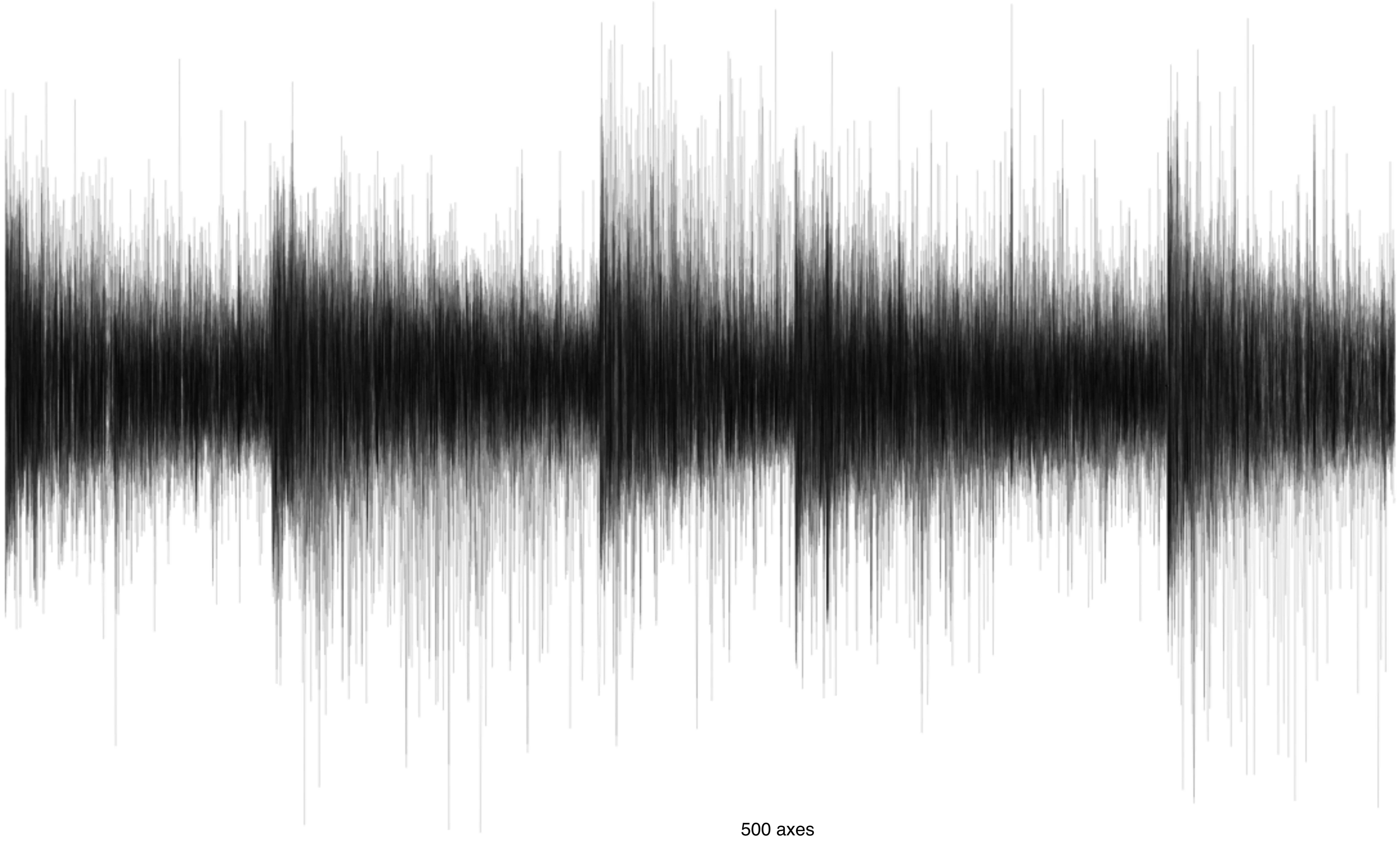
Lines connecting axis represent records

Suitable for

- all tabular data types
- heterogeneous data



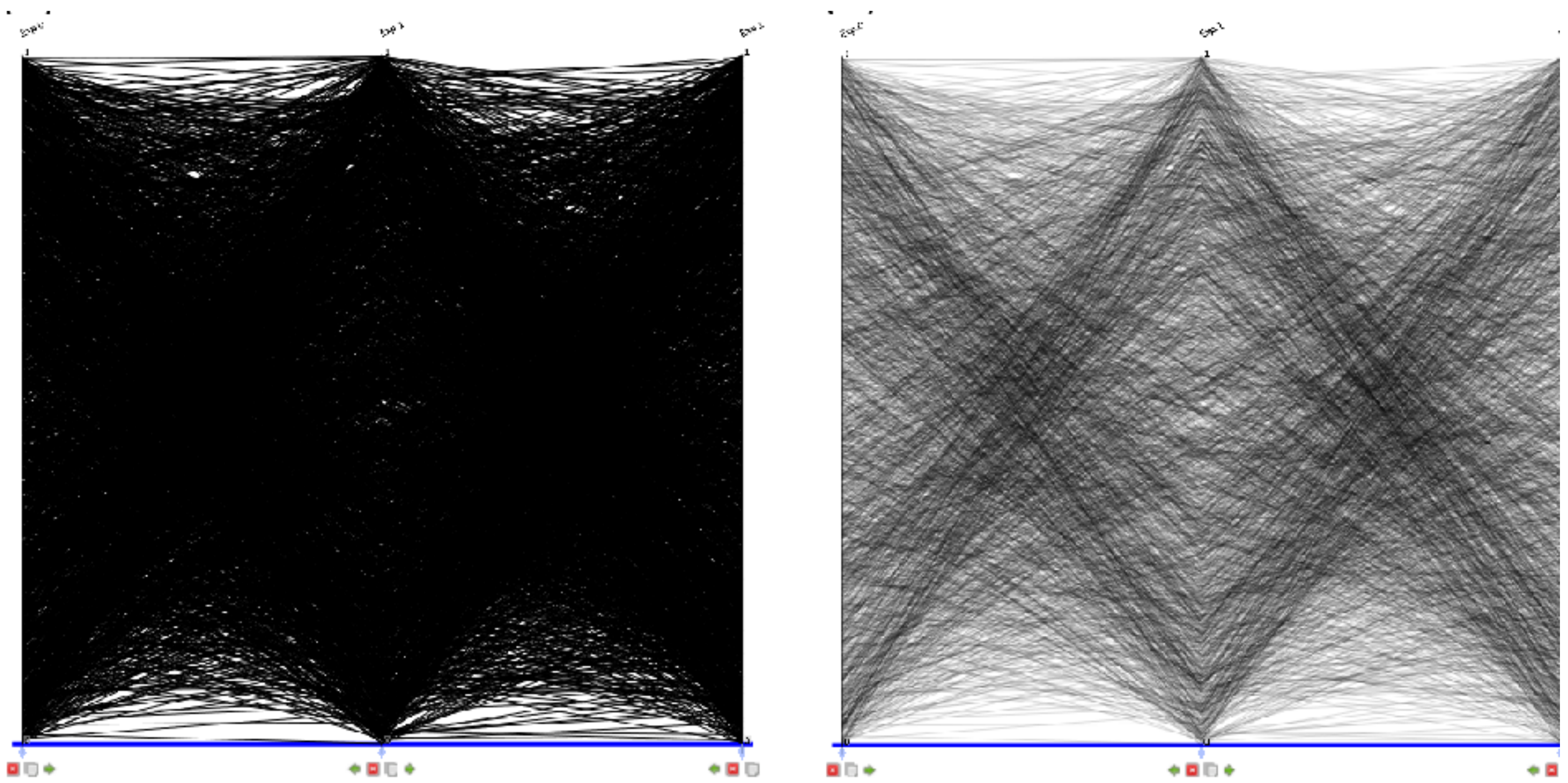
PC Limitation: Scalability to Many Dimensions



PC Limitation: Scalability to Many Items

Solutions:

- Transparency
- Bundling, Clustering
- Sampling

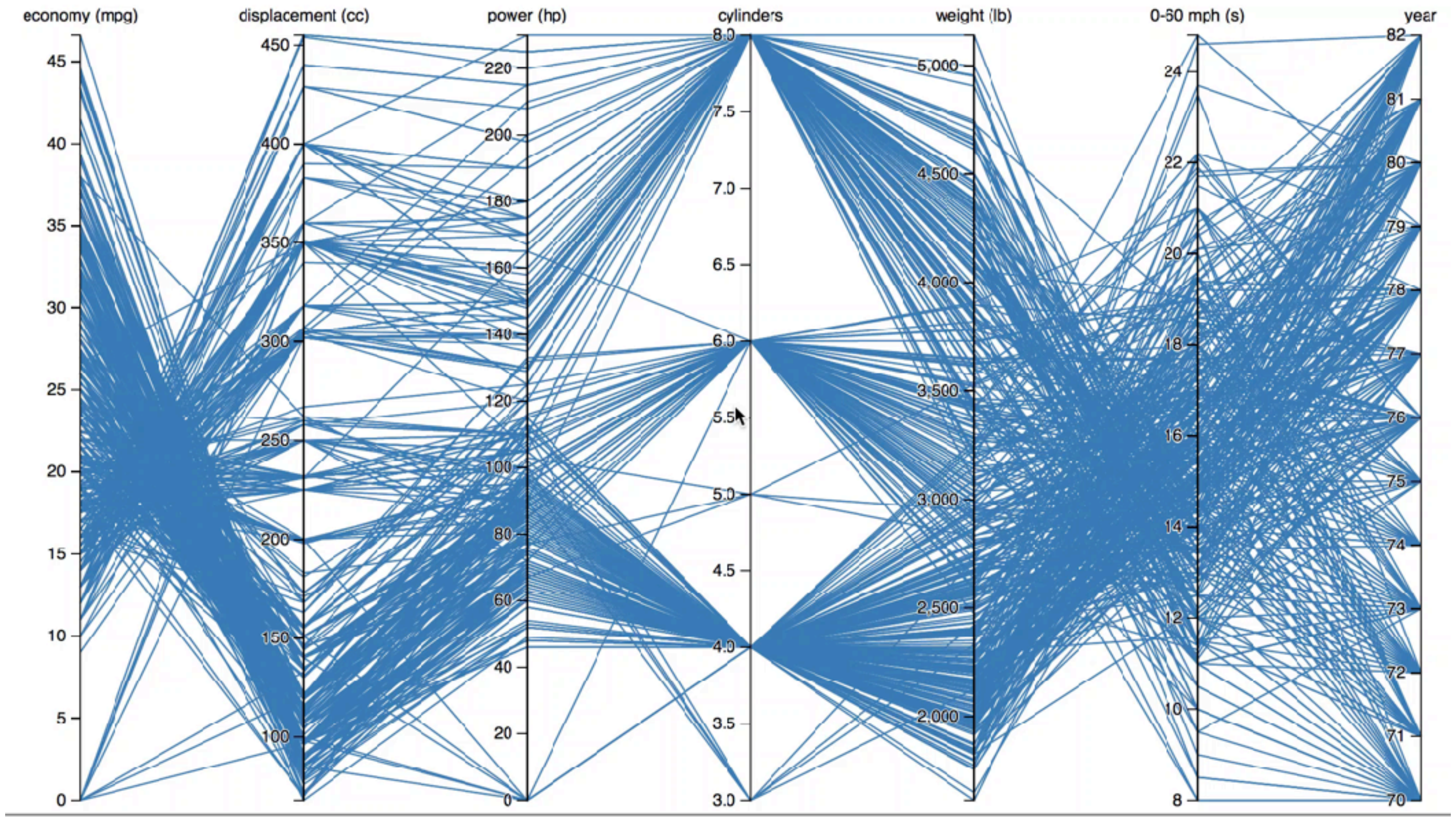


PC Limitations

Correlations only between adjacent axes

Solution: Interaction

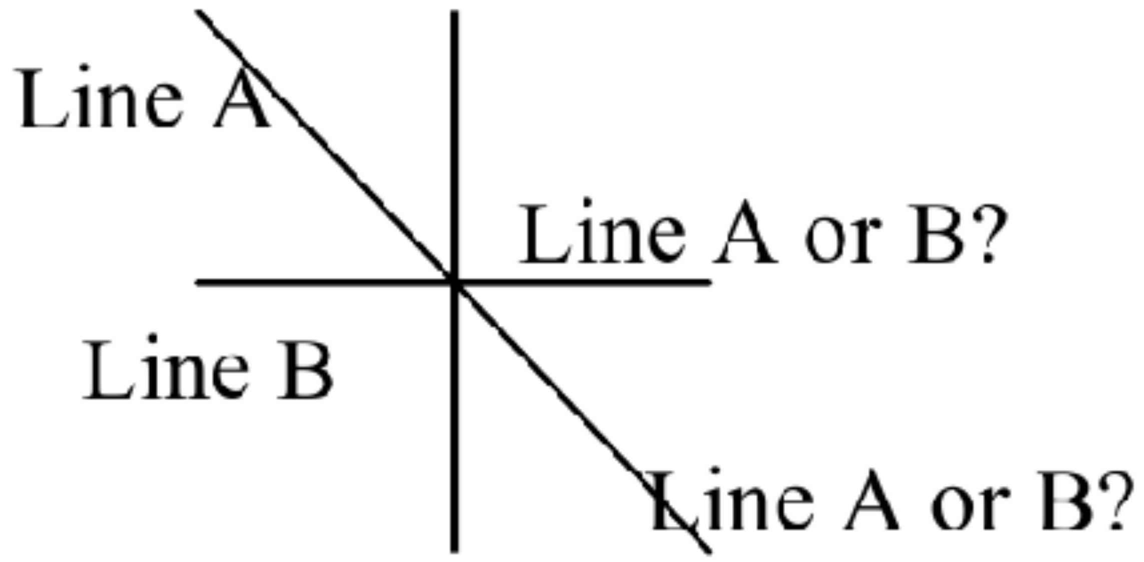
- Brushing
- Let user change order

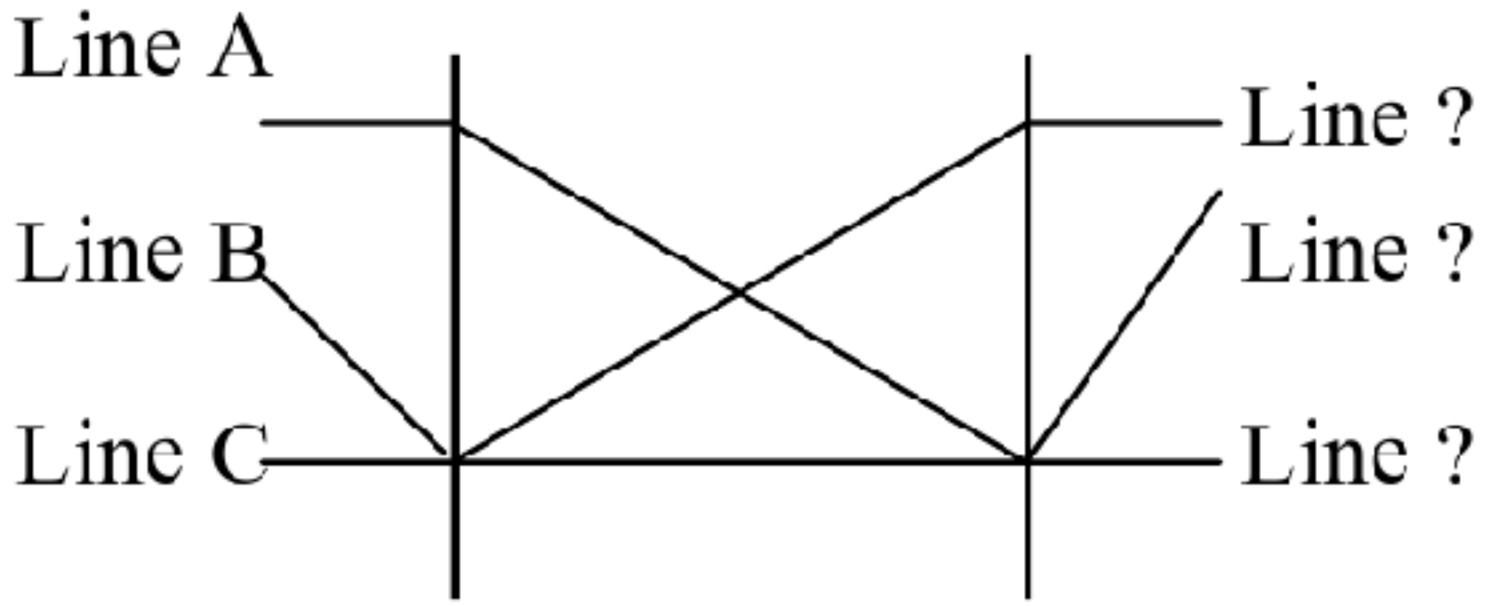


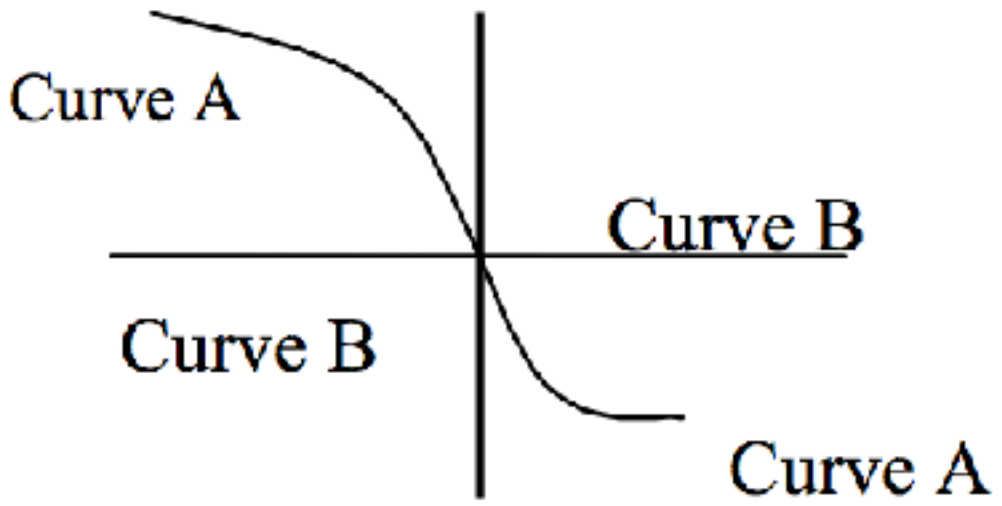
PC Limitation: Ambiguity

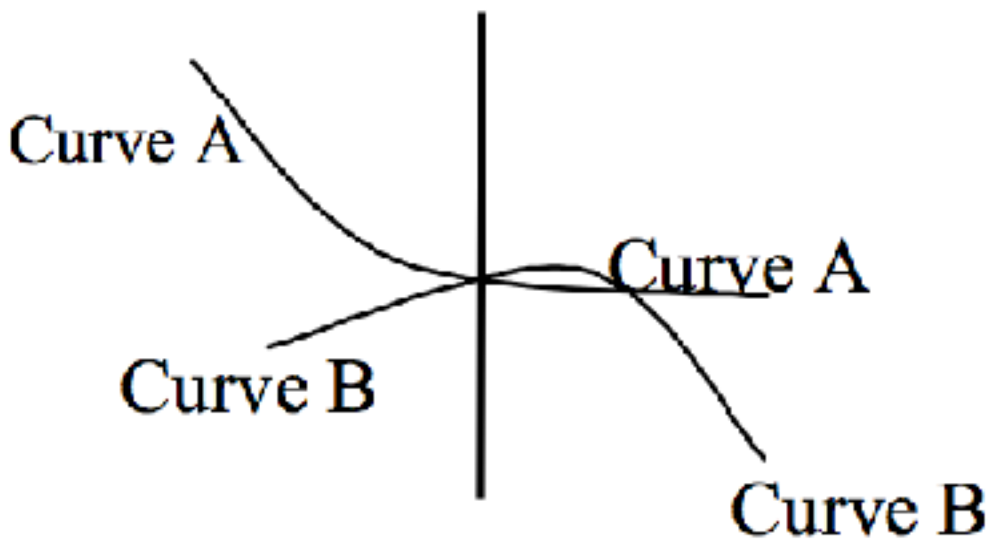
Solutions:

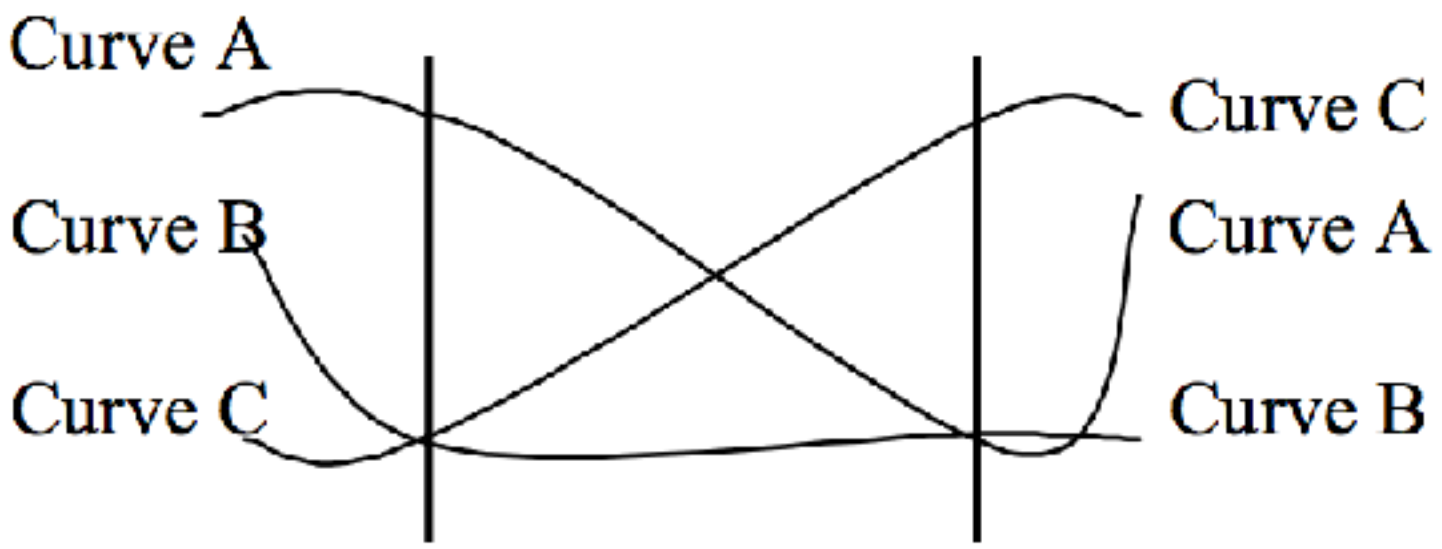
- Brushing
- Curves











Graham and Kennedy 2003

Parallel Coordinates

Shows primarily relationships between adjacent axis

Limited scalability (~50 dimensions, ~1-5k records)

- Transparency of lines

Interaction is crucial

- Axis reordering
- Brushing
- Filtering

Algorithmic support:

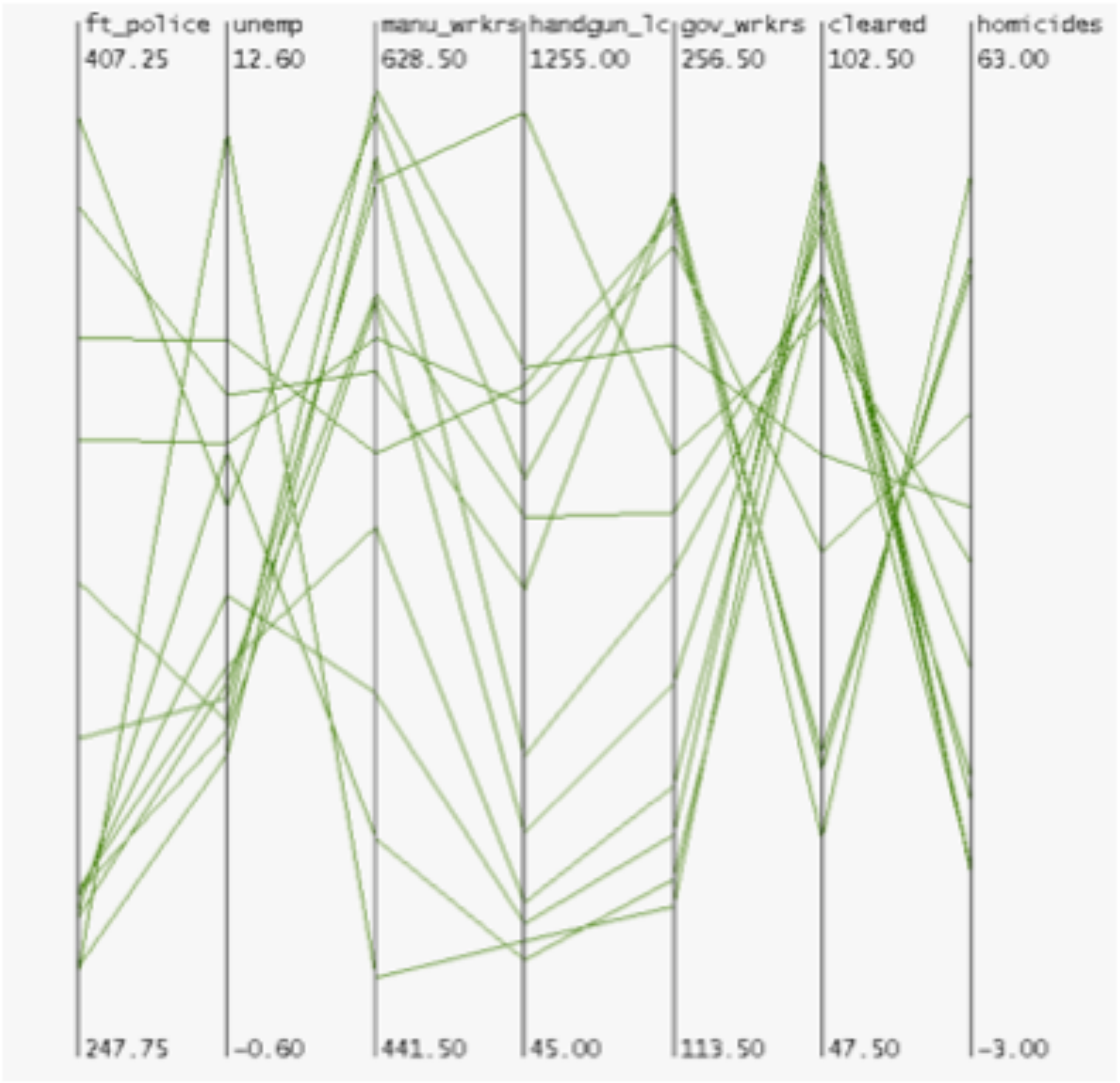
- Choosing dimensions
- Choosing order
- Clustering & aggregating records

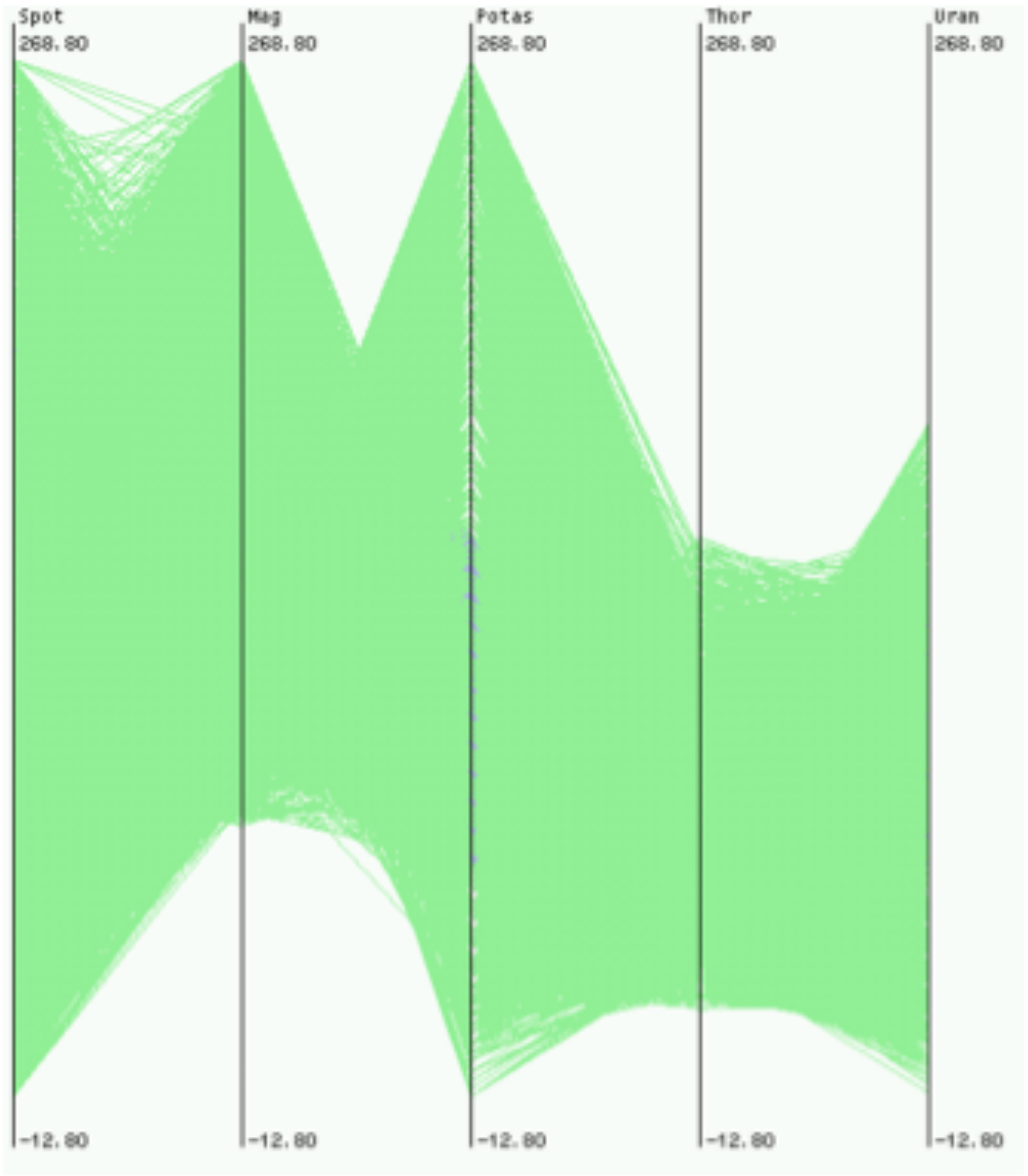
http://bl.ocks.org/jasondavies/1341281

HIERARCHICAL PARALLEL COORDINATES

goal: scale up parallel coordinates to large datasets

challenge: overplotting/occlusion



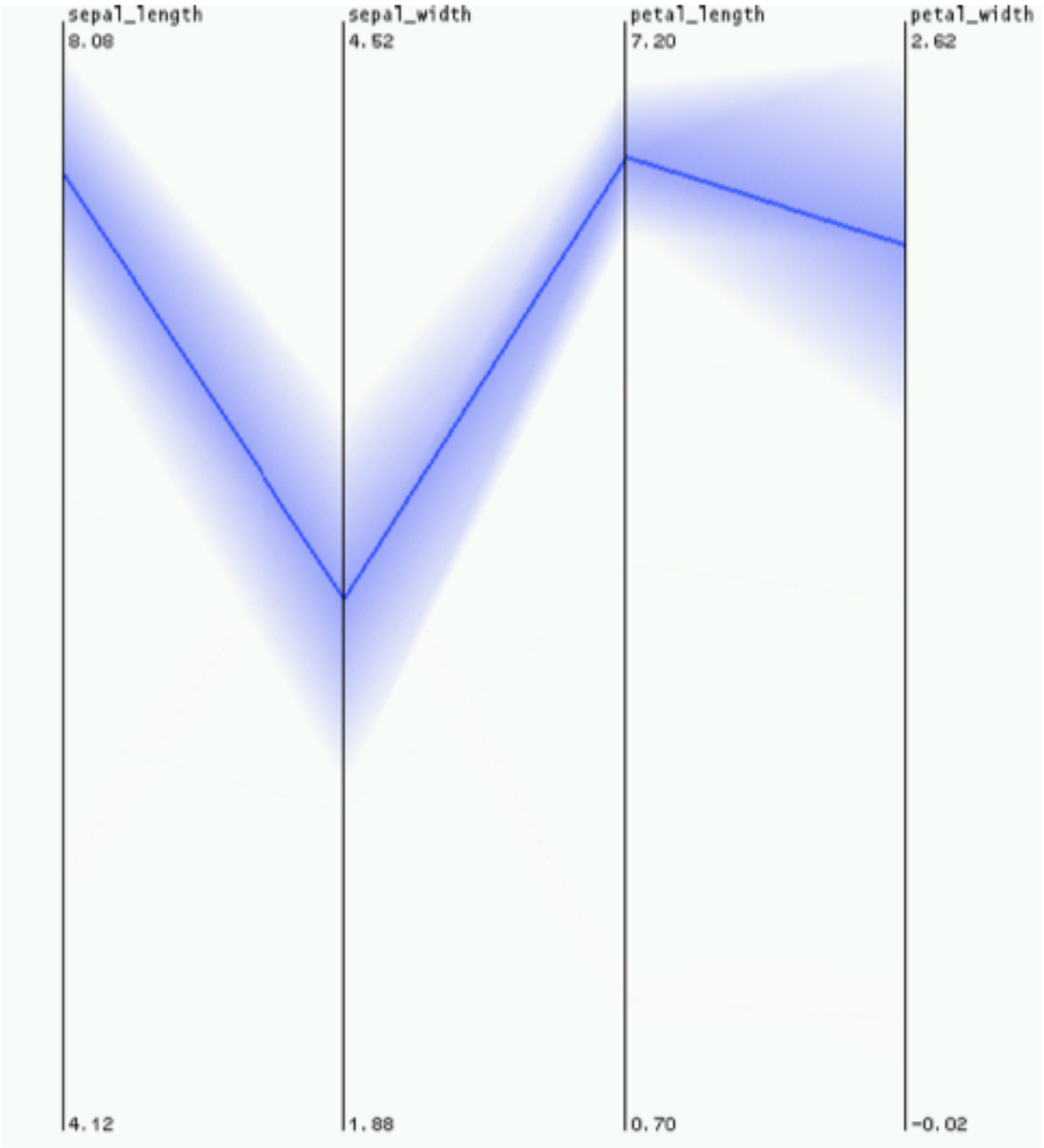


Fua 1999

HPC: ENCODING DERIVED DATA

visual representation: variable-width opacity bands

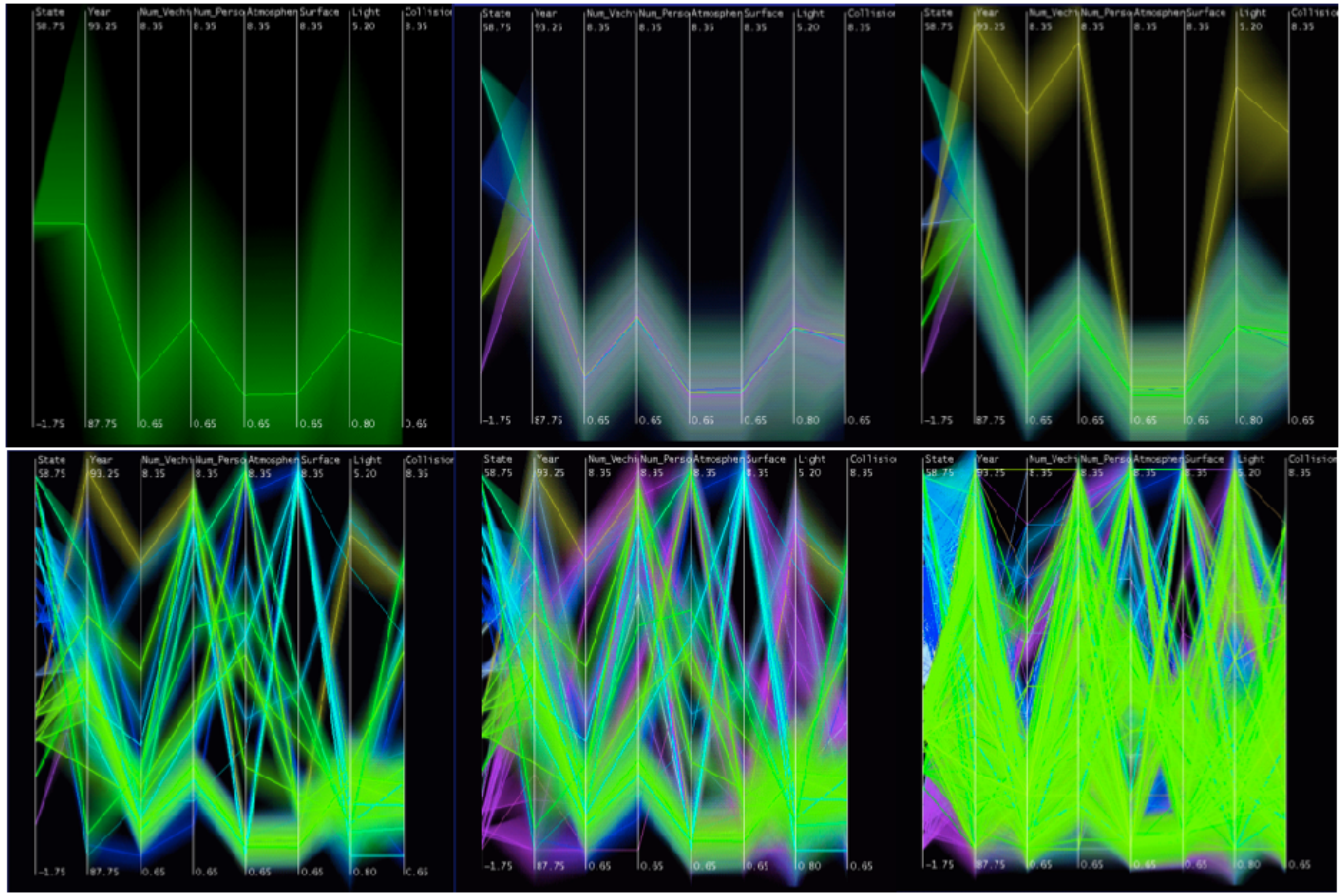
- show whole cluster, not just single item
- min / max: spatial position
- cluster density: transparency
- mean: opaque



Fua 1999

HPC: INTERACTING WITH DERIVED DATA

interactively change level of detail to navigate cluster hierarchy



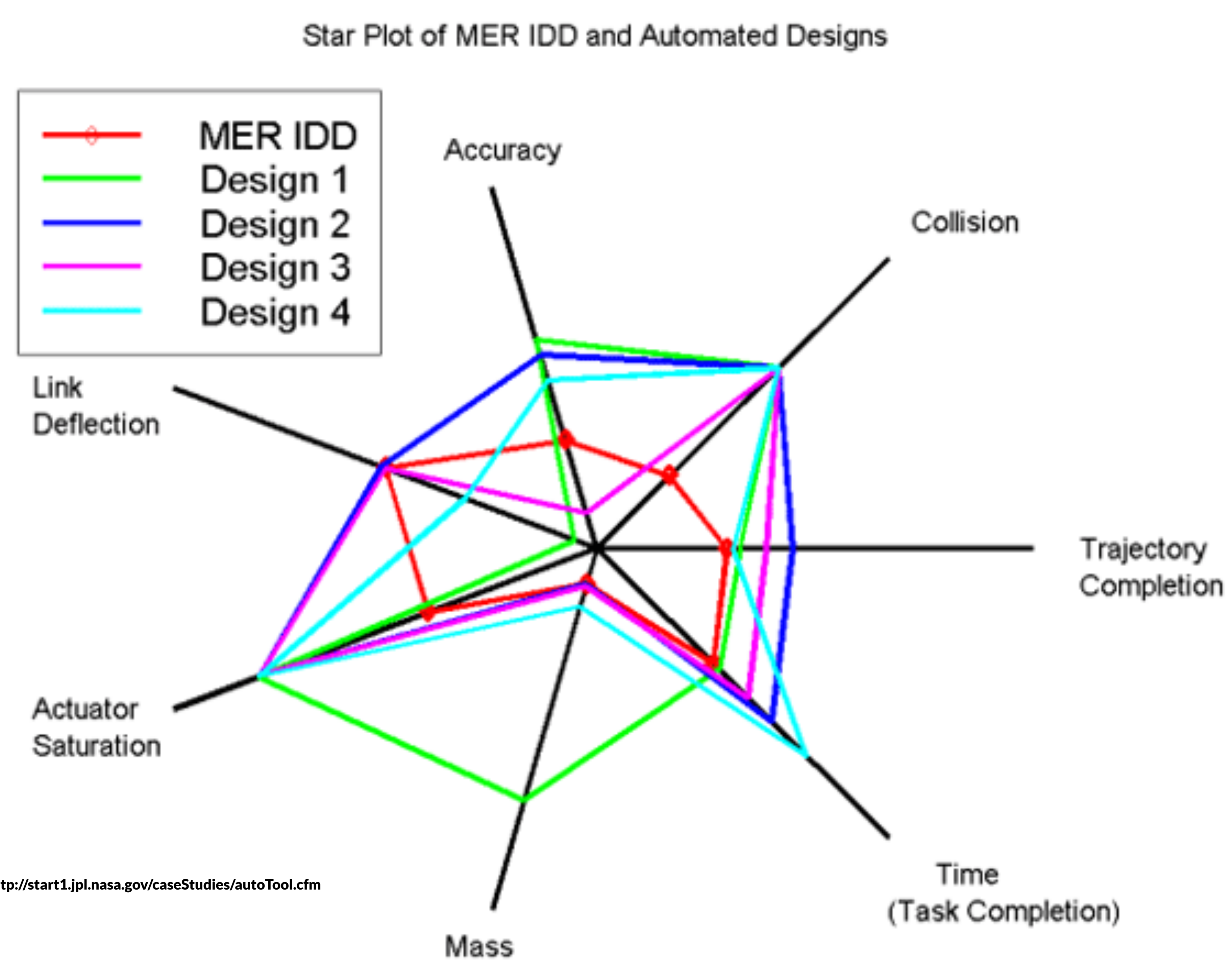
Fua 1999

Star Plot

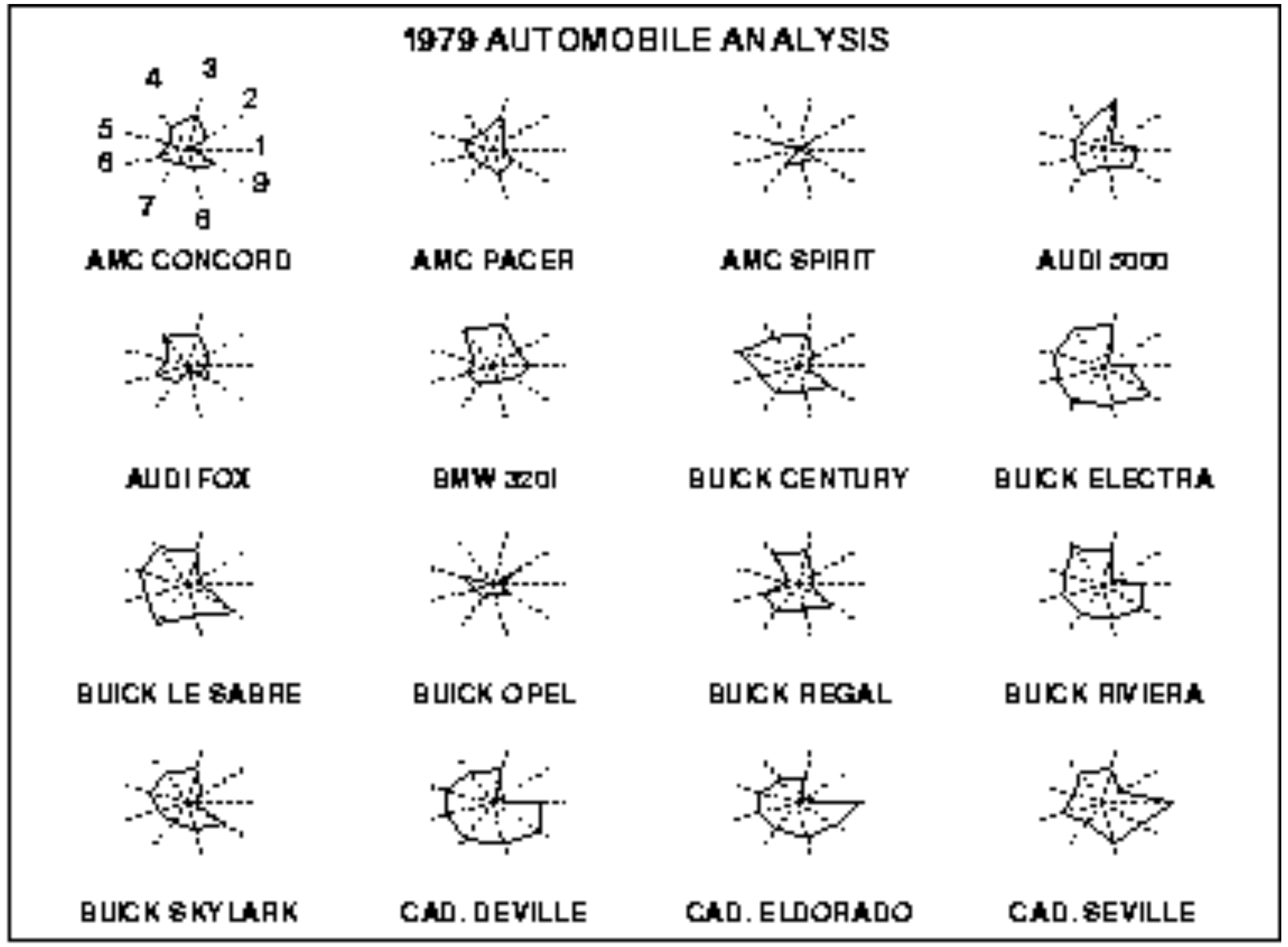
[Coekin1969]

Similar to parallel coordinates

Radiate from a common origin



http://start1.jpl.nasa.gov/caseStudies/autoTool.cfm



http://www.itl.nist.gov/div898/handbook/eda/section3/starplot.htm

http://bl.ocks.org/kevinschaul/raw/8833989/

Data Reduction

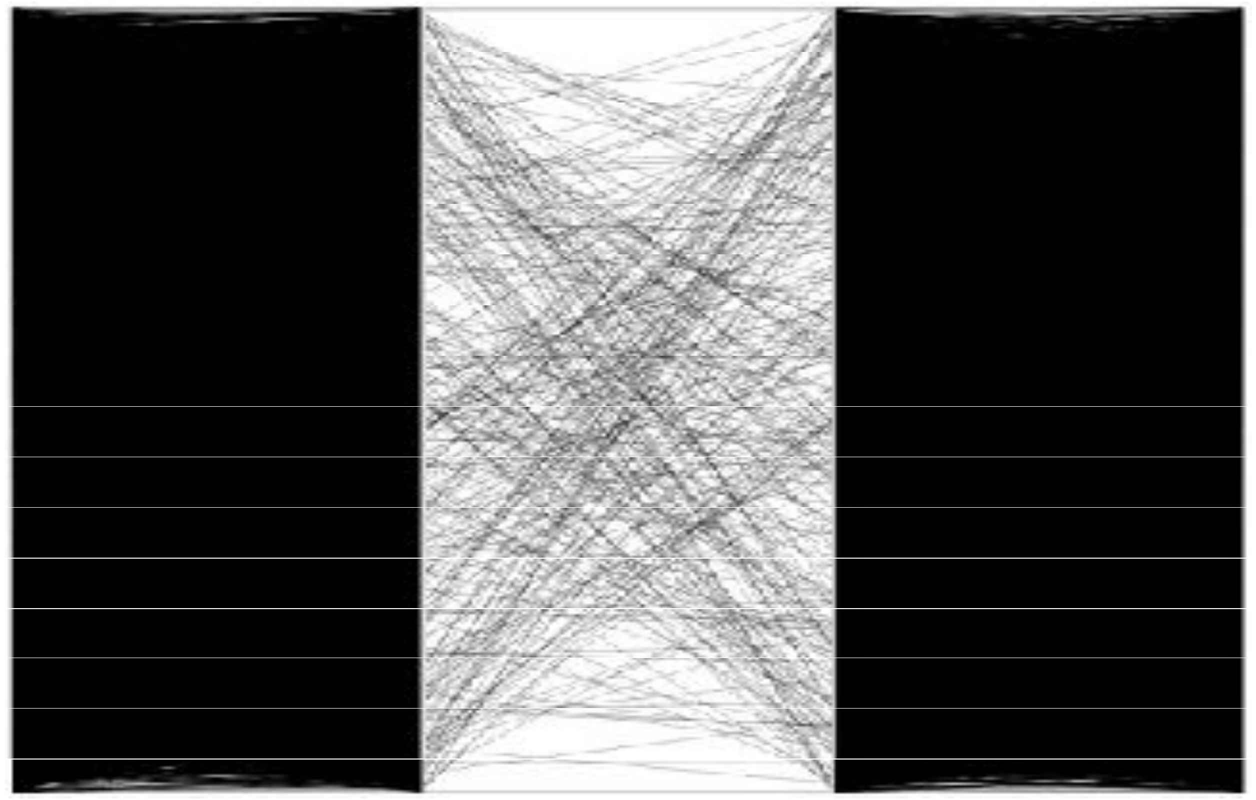
Sampling

Don’t show every element, show a (random) subset

Efficient for large dataset

Apply only for display purposes

Outlier-preserving approaches



[Ellis & Dix, 2006]

Filtering

Define criteria to remove data, e.g.,

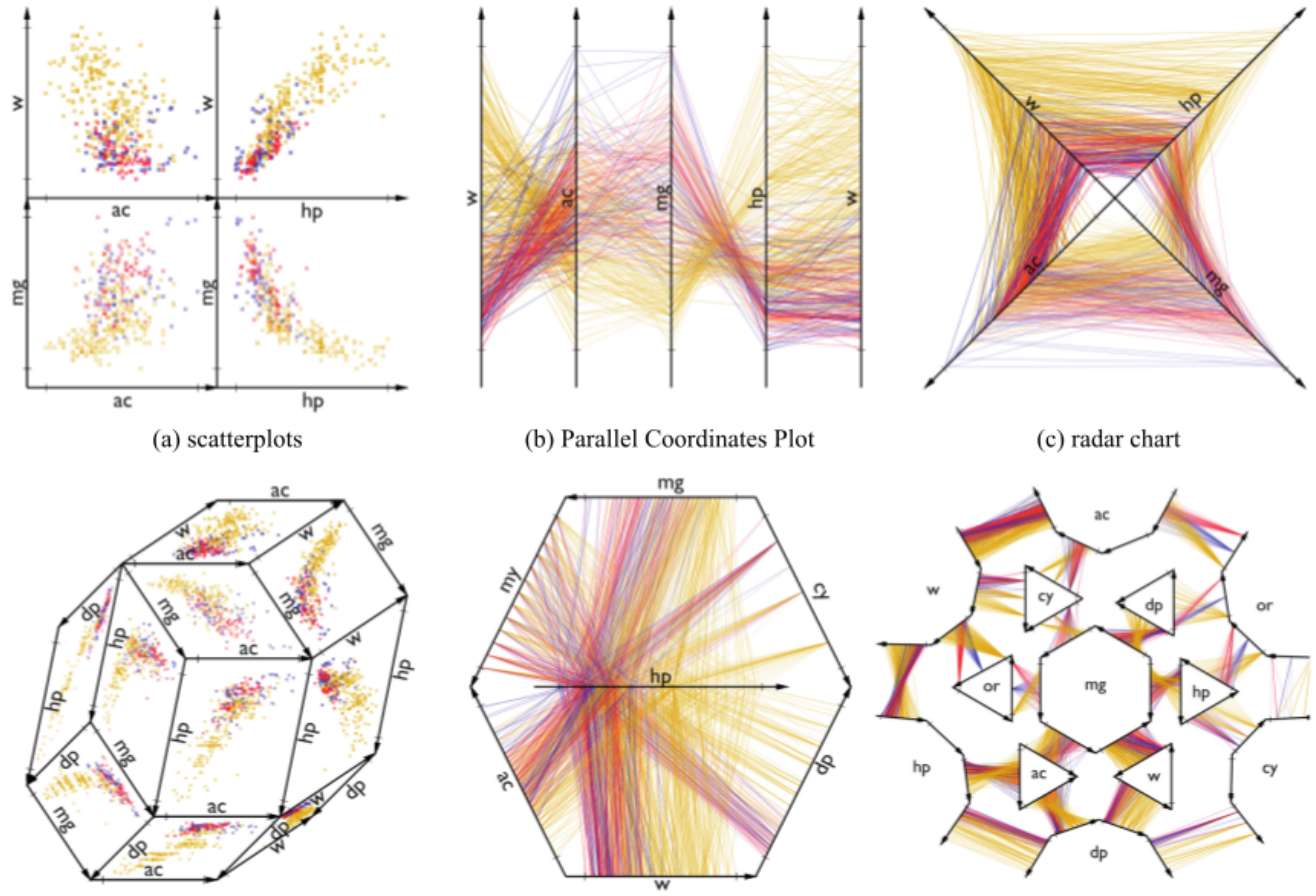
- minimum variability
- > / < / = specific value for one dimension
- consistency in replicates, …

Can be interactive, combined with sampling

Spatial Axis Orientation

Hybrids

Flexible Linked Axes (FLINA)

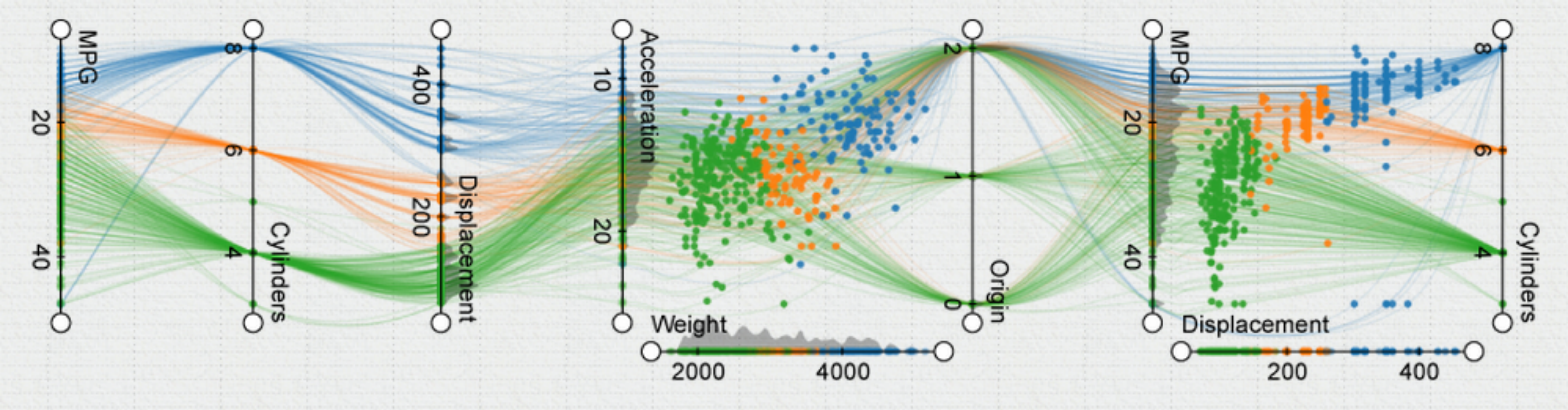


(a) scatterplots (b) Parallel Coordinates Plot (c) radar chart

(d) Hyperbox (e) Time Wheel (f) Many-to-many PCP

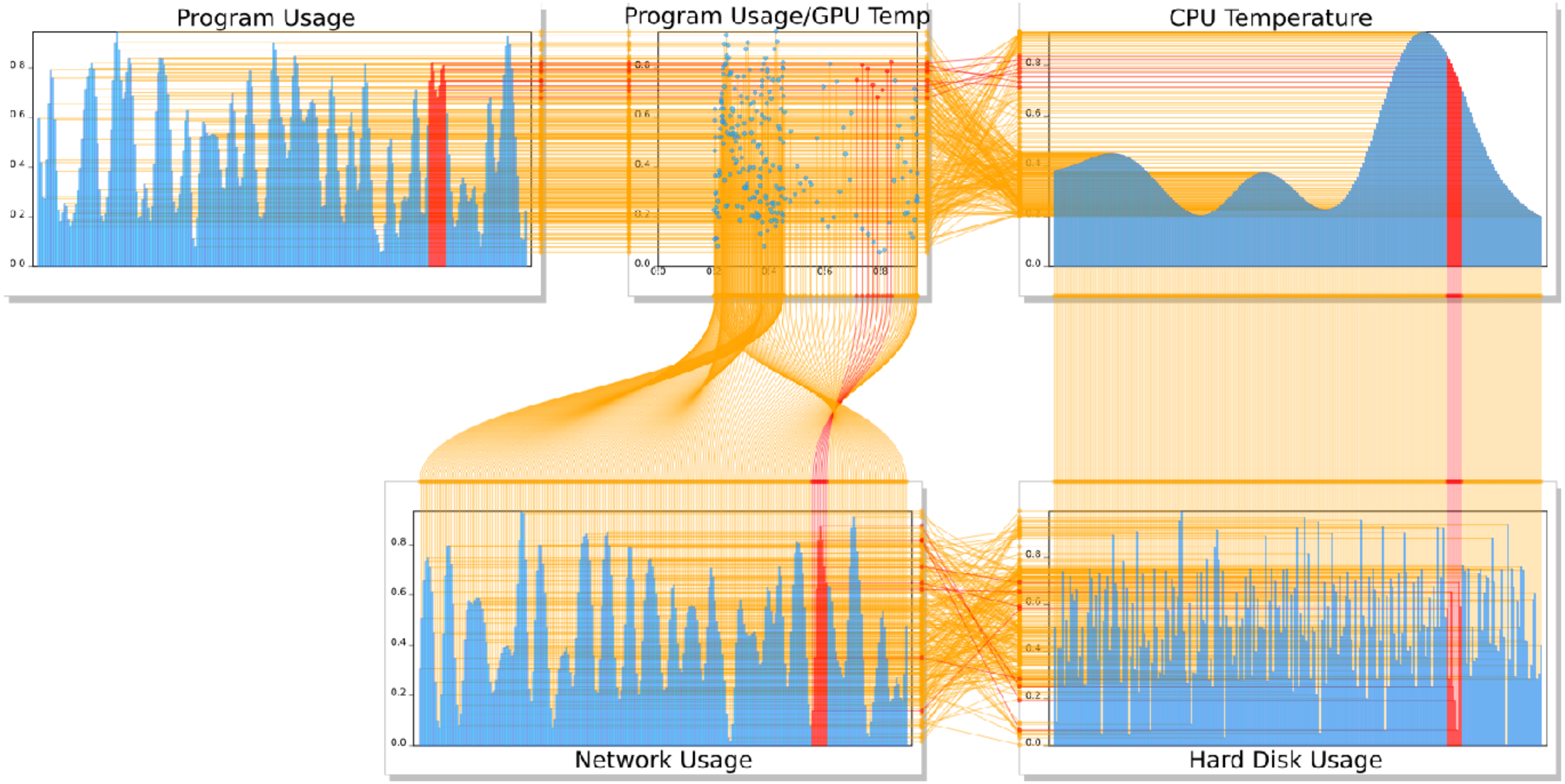
Claessen & van Wijk 2011

Web-based implementation of FLINA concept



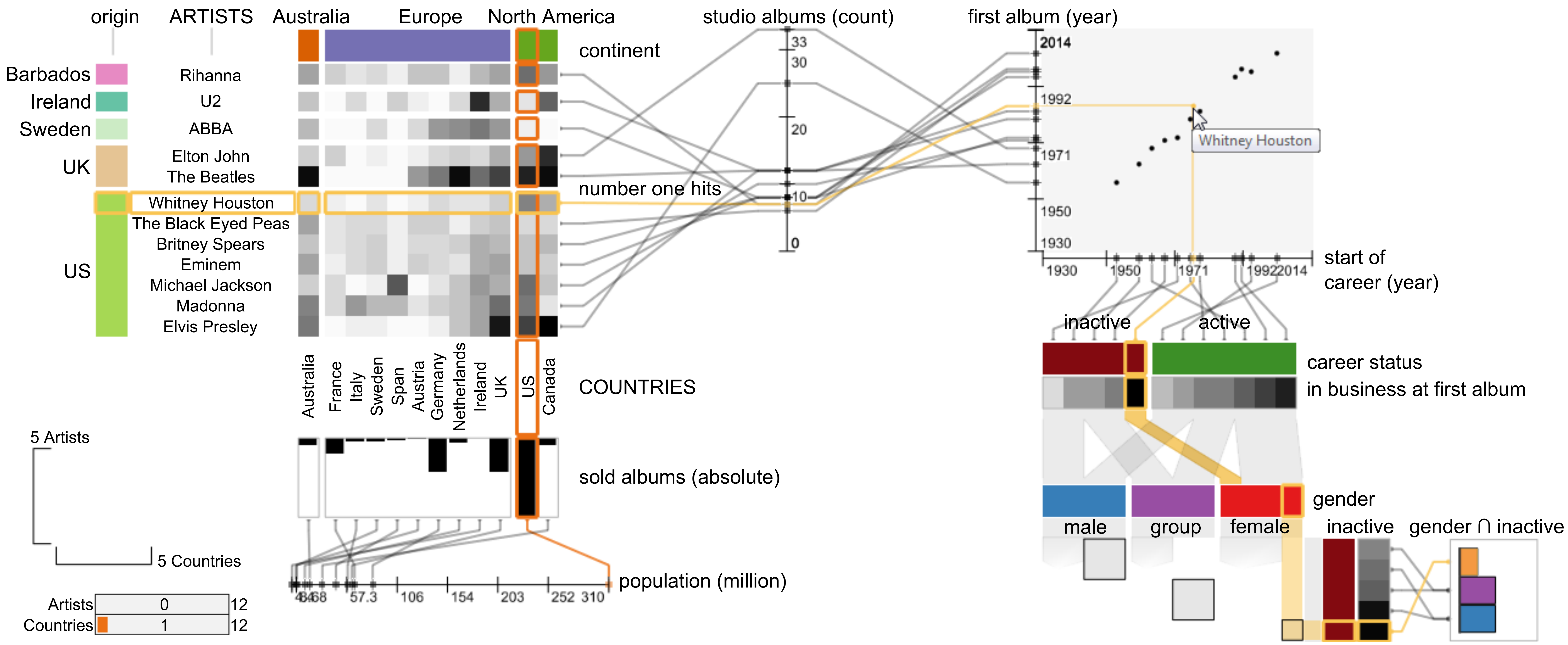
http://vis.pku.edu.cn/mddv/val/

Connected Charts



Viau & McGuffin 2012

Domino



Gratzl et al. 2014

Spatial Axis Orientation

Parallel Sets

Parallel Sets

builds on PC to better handle categorical data

- discrete
- small number of values
- no implied ordering between attributes

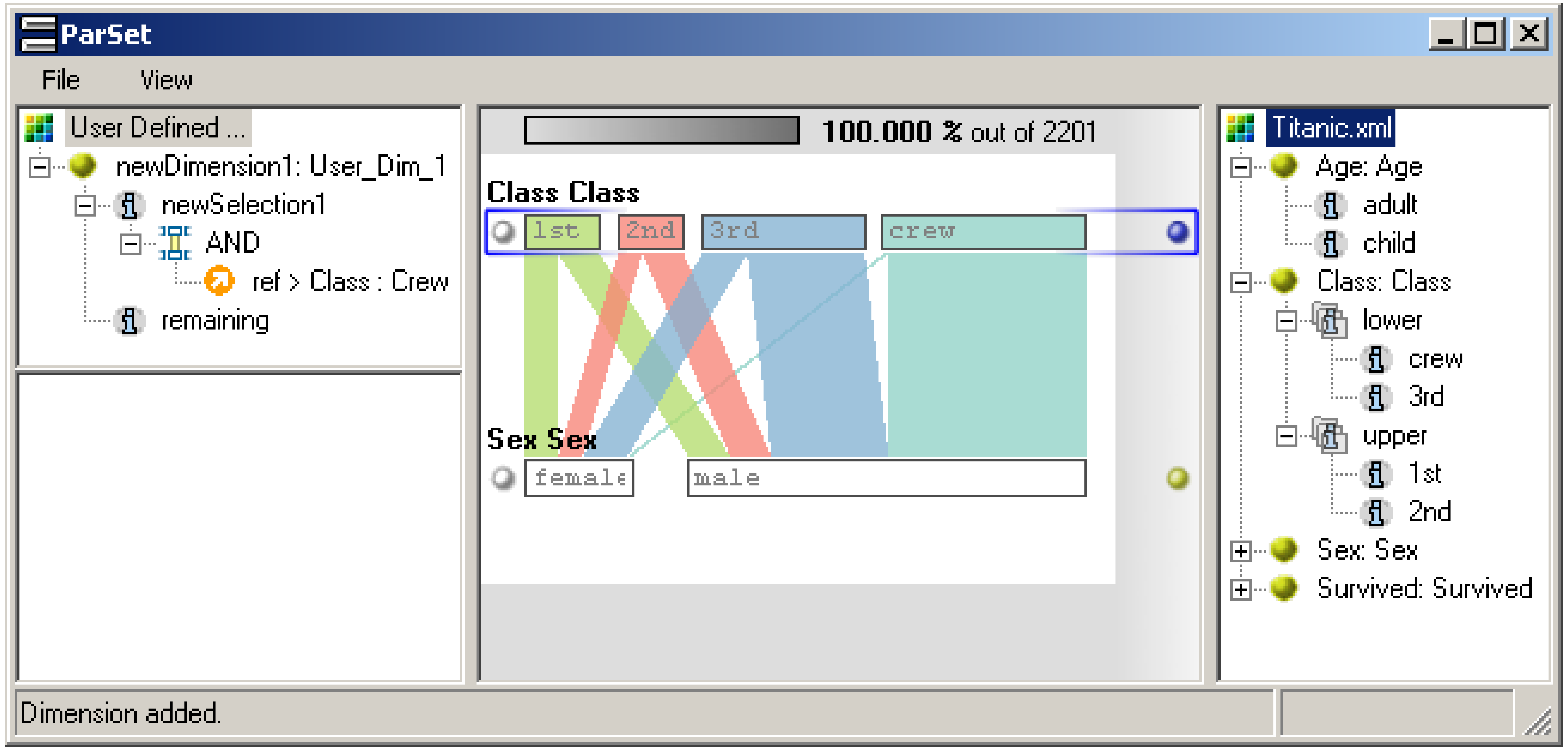
task: find relationship between attributes

interaction driven technique

Visual Encoding

boxes scaled by frequency

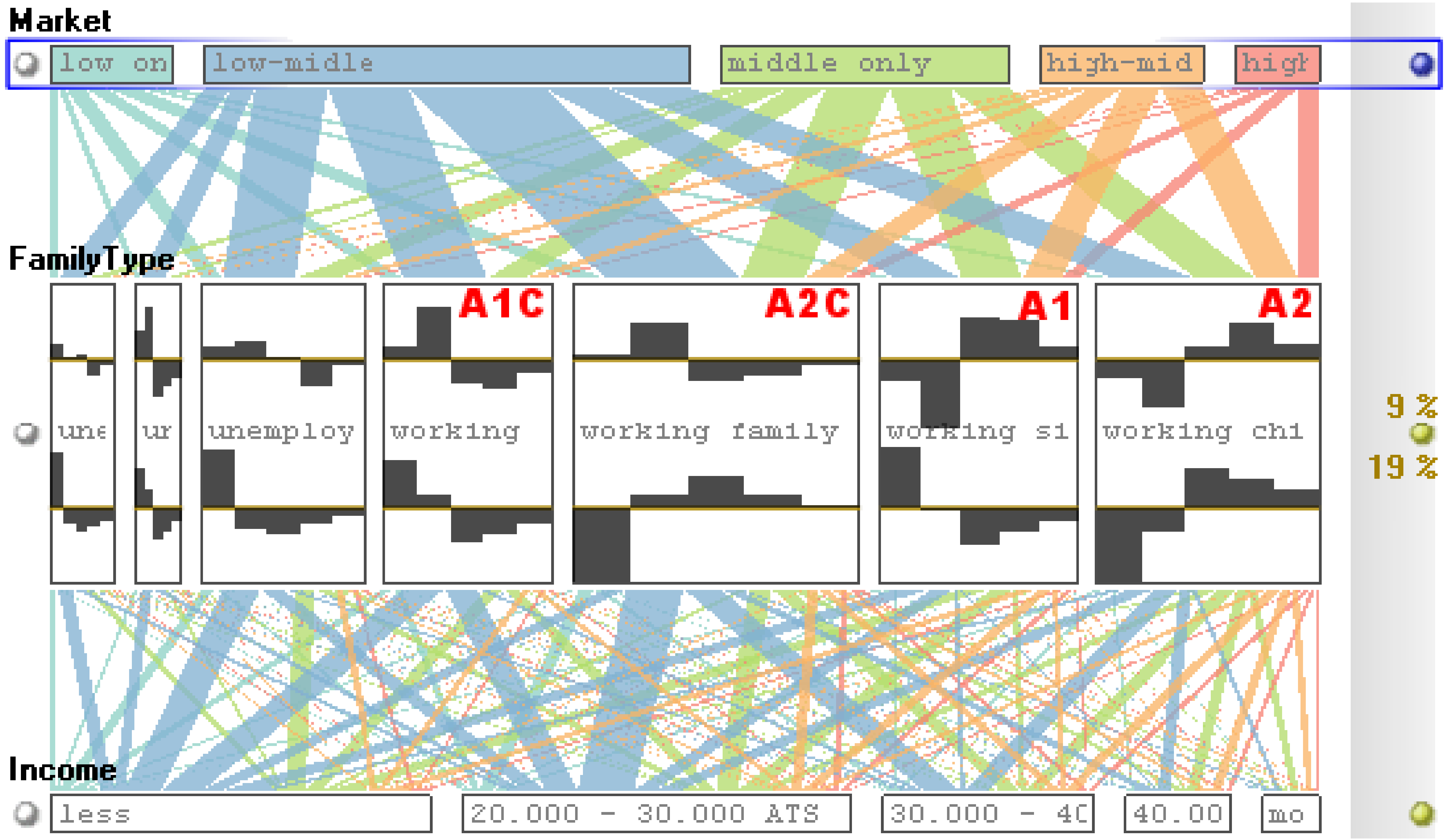
color coded by values for current active dimension



Bendix, Kosara, Hauser, 2005

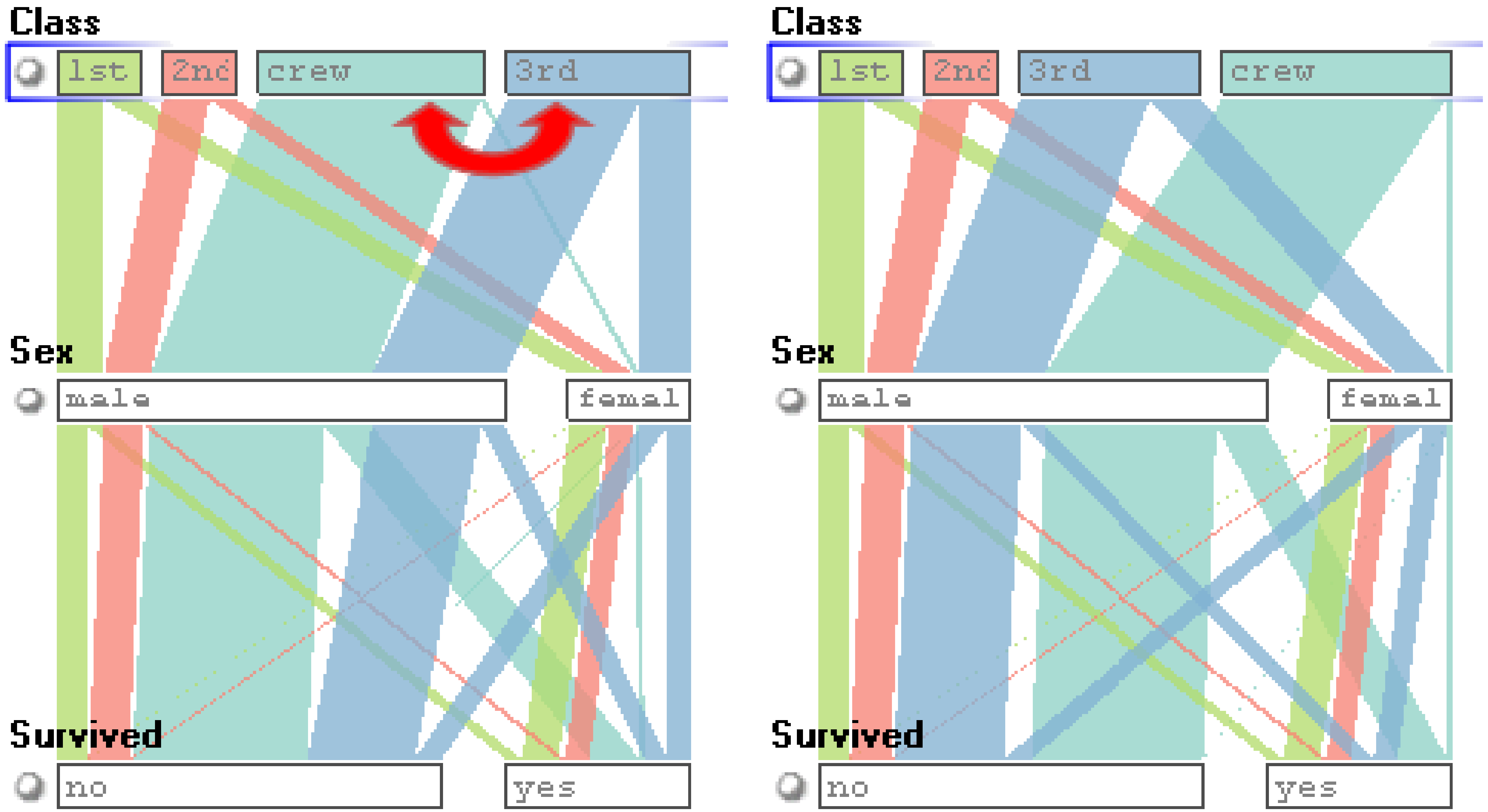
Visual Encoding

- boxes expand to show histogram



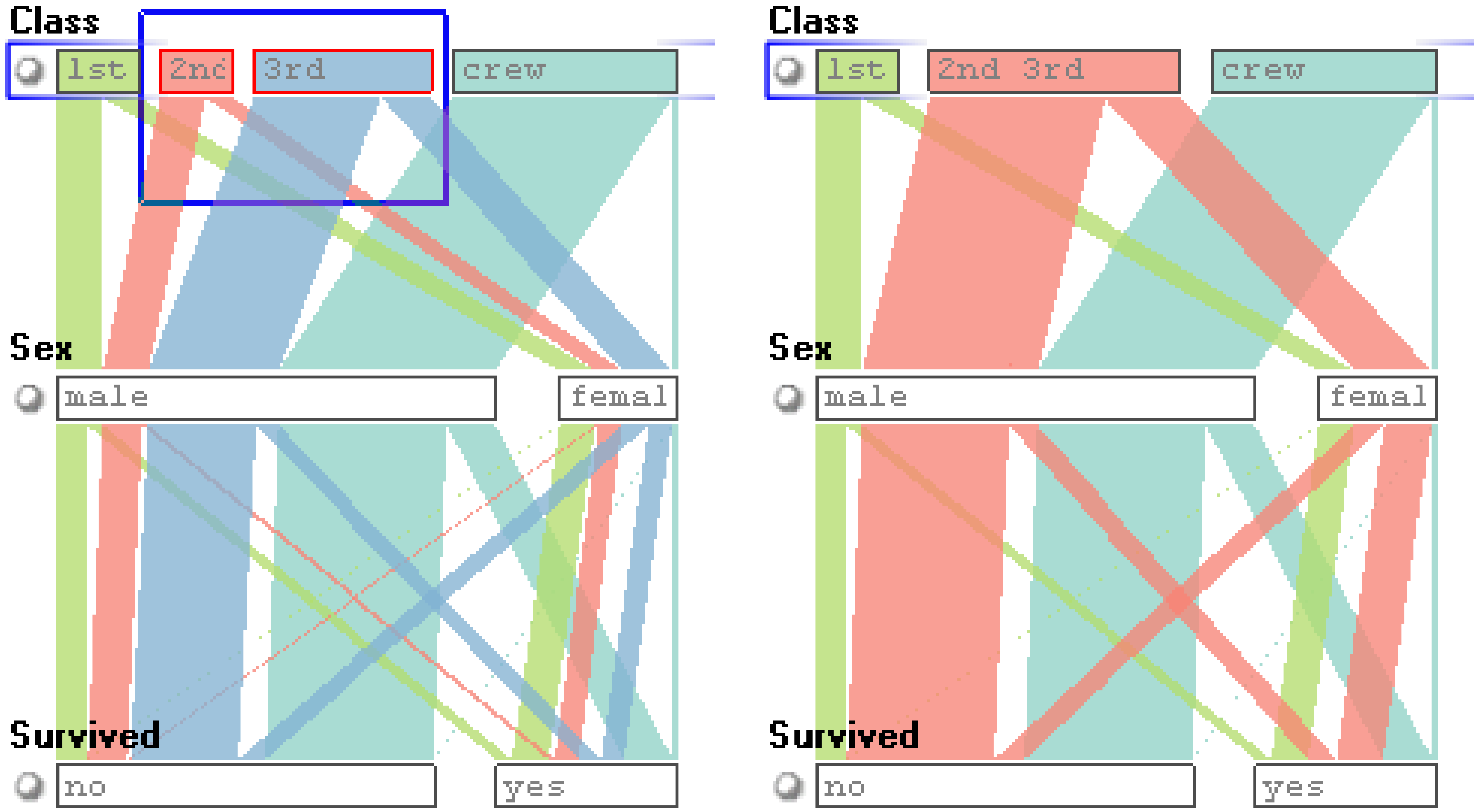
Bendix, Kosara, Hauser, 2005

Interaction: Reorder



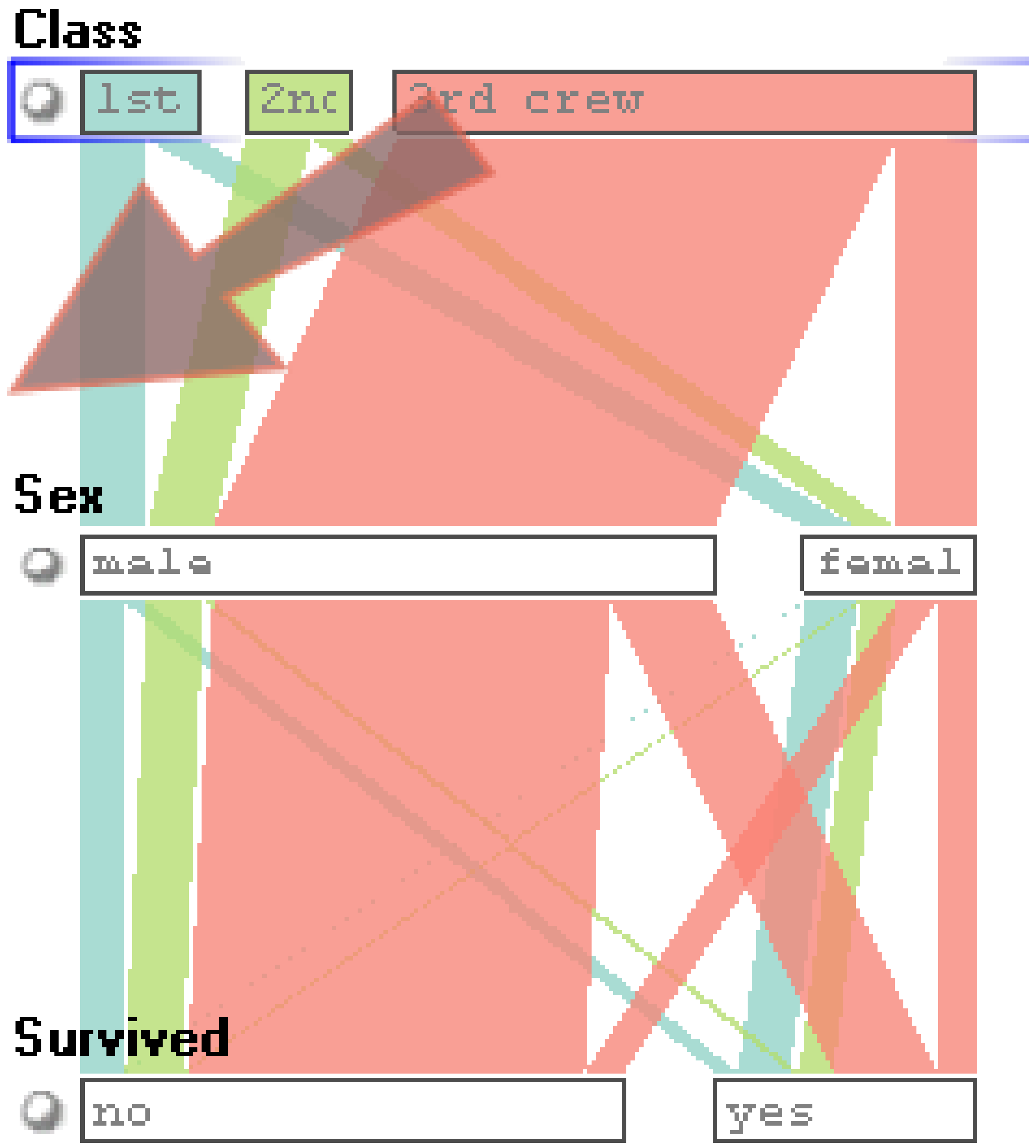
Bendix, Kosara, Hauser, 2005

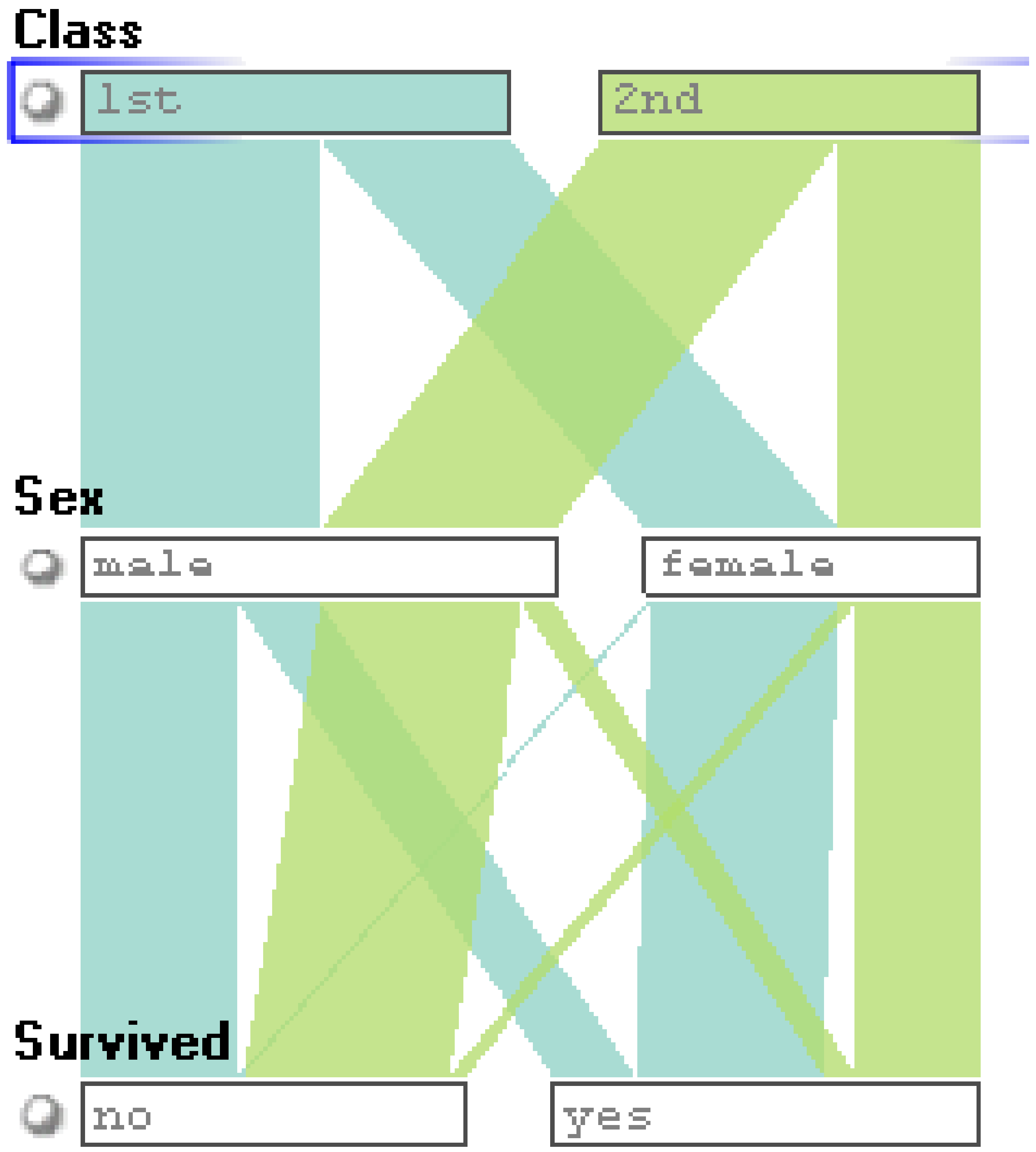
Interaction: Aggregate



Bendix, Kosara, Hauser, 2005

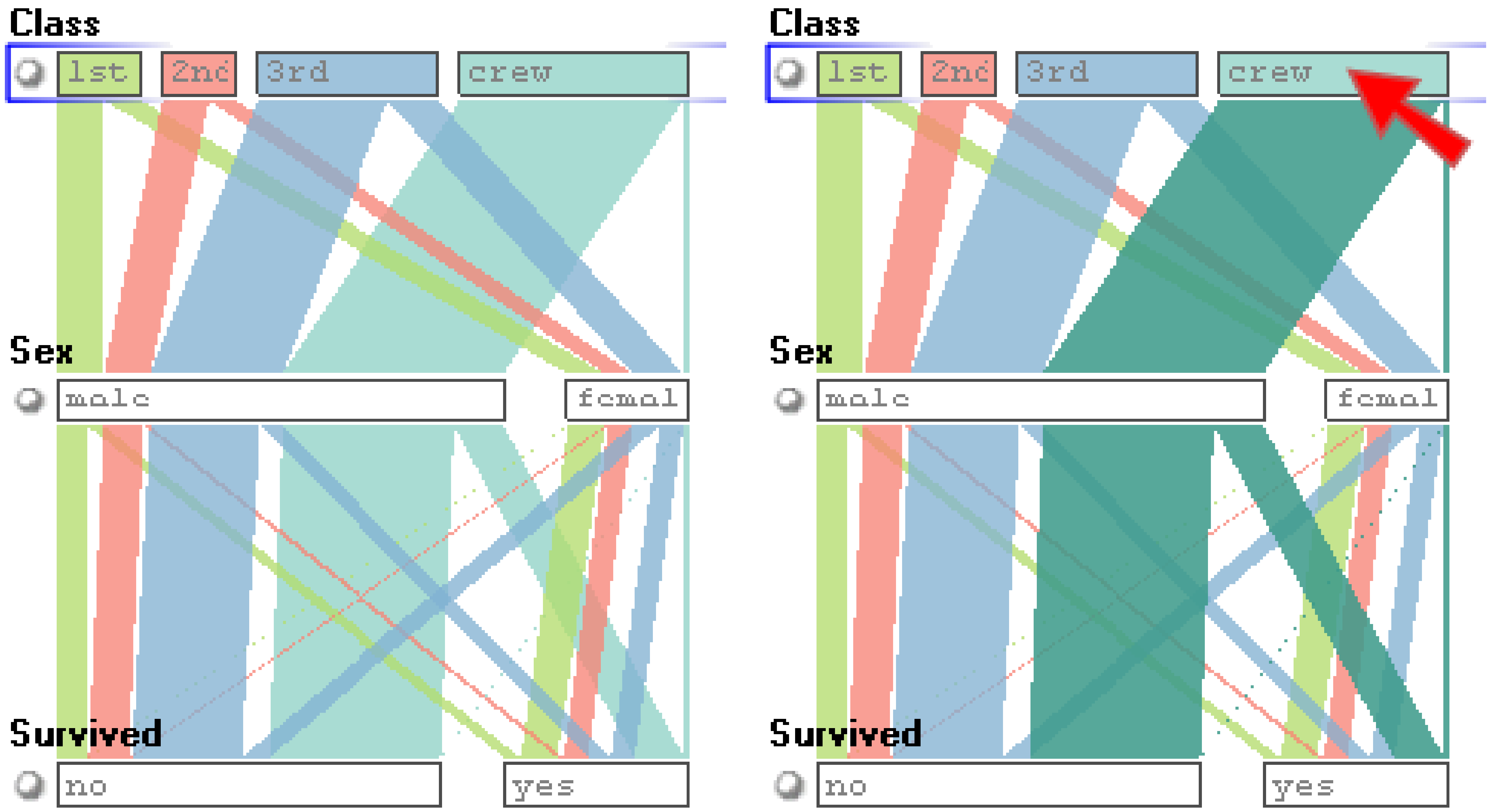
Interaction: Filter





Bendix, Kosara, Hauser, 2005

Interaction: Highlight

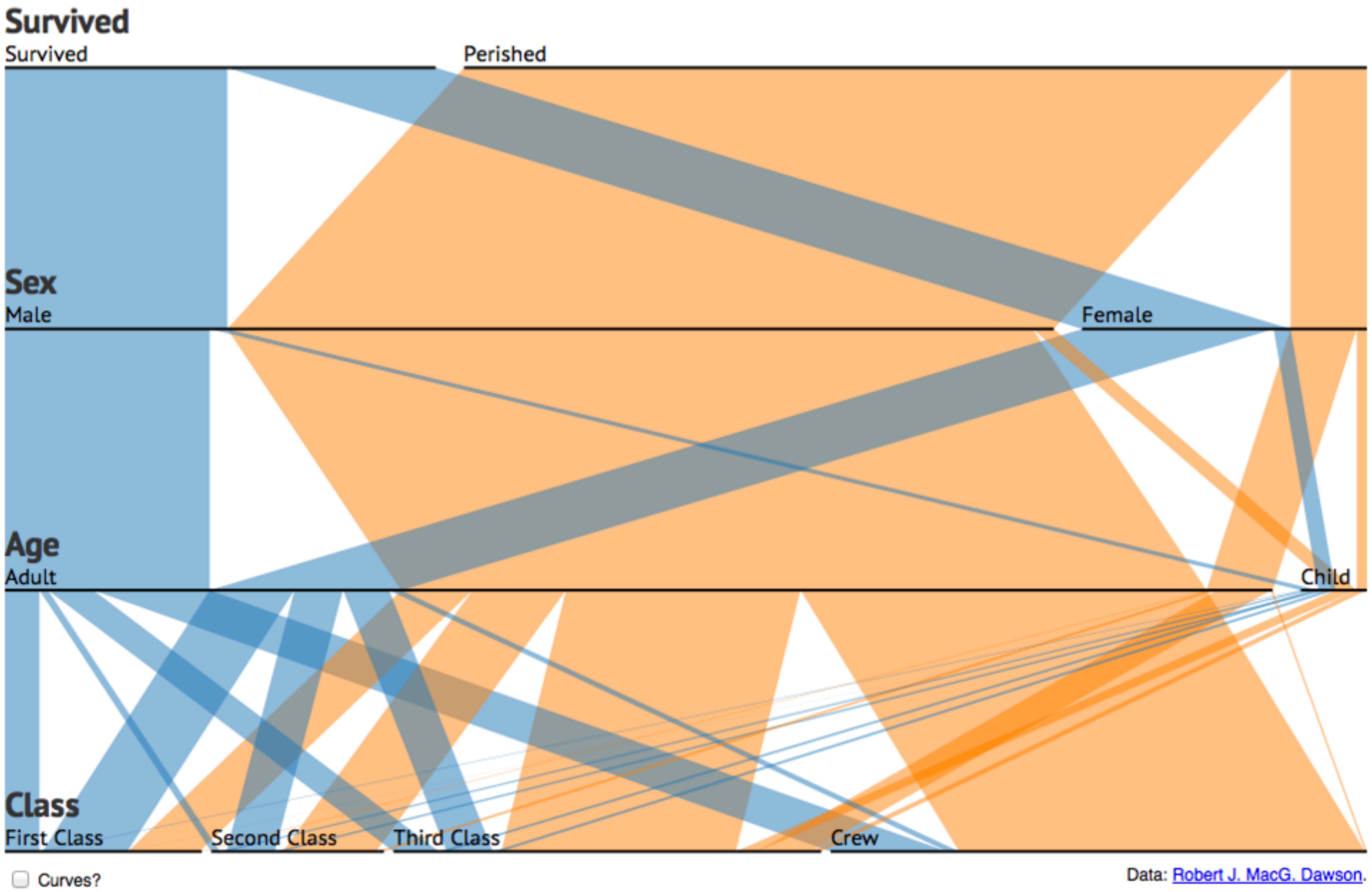


Bendix, Kosara, Hauser, 2005

Parallel Sets

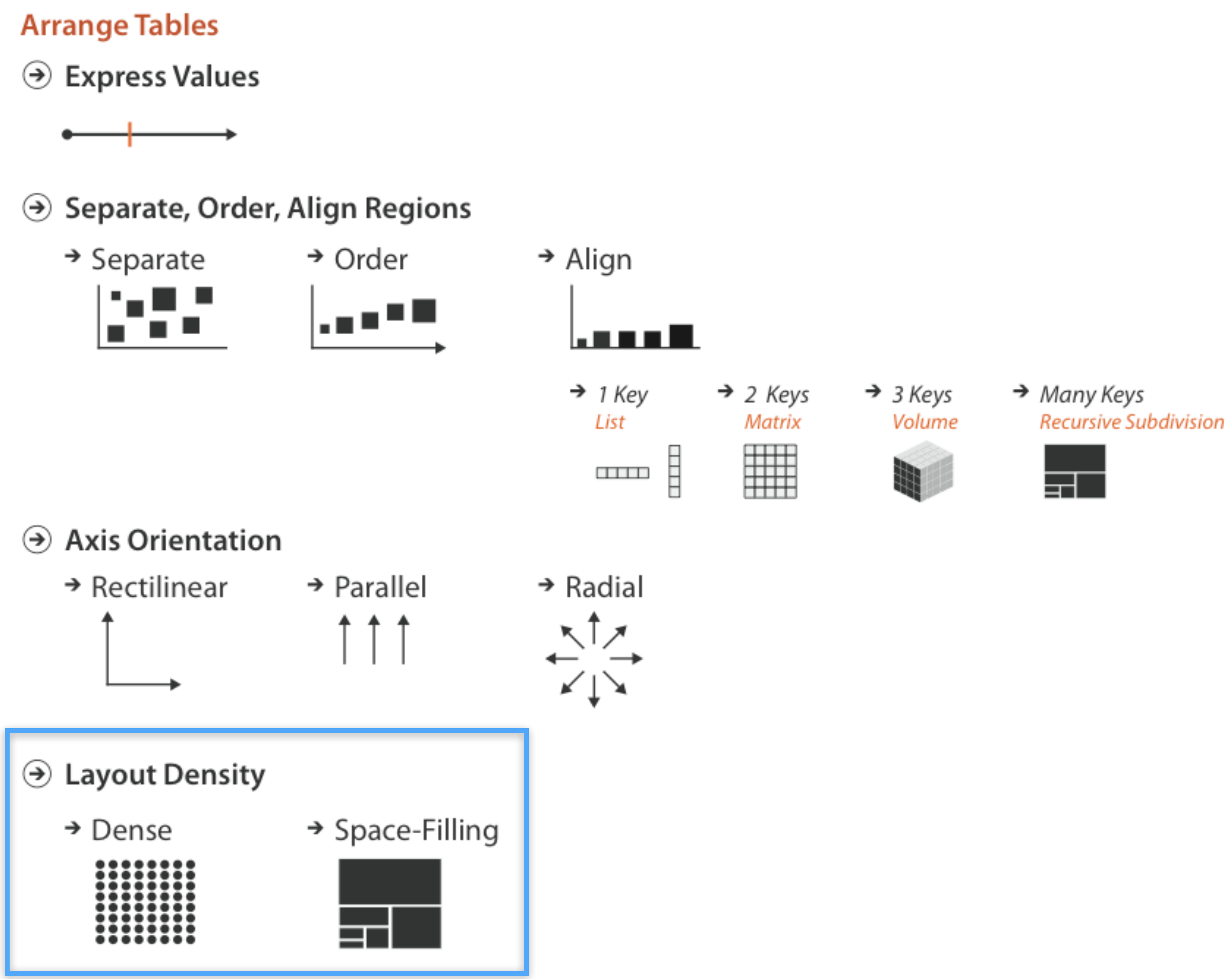
A visualisation technique for multidimensional categorical data.

Titanic Survivors



Data: Robert J. MacG. Dawson.

Filling Space


Arrange Tables
Express Values
Separate, Order, Align Regions
Separate
Order
Align
1 Key
List
2 Keys
Matrix
3 Keys
Volume
Many Keys
Recursive Subdivision
Axis Orientation
Rectilinear
Parallel
Radial
Layout Density
Dense
Space-Filling

Dense pixel display: VisDB

represent each data item, or each attribute in an item as a single pixel

can fit as many items on the screen as there are pixels, on the order of millions

relies heavily on color coding

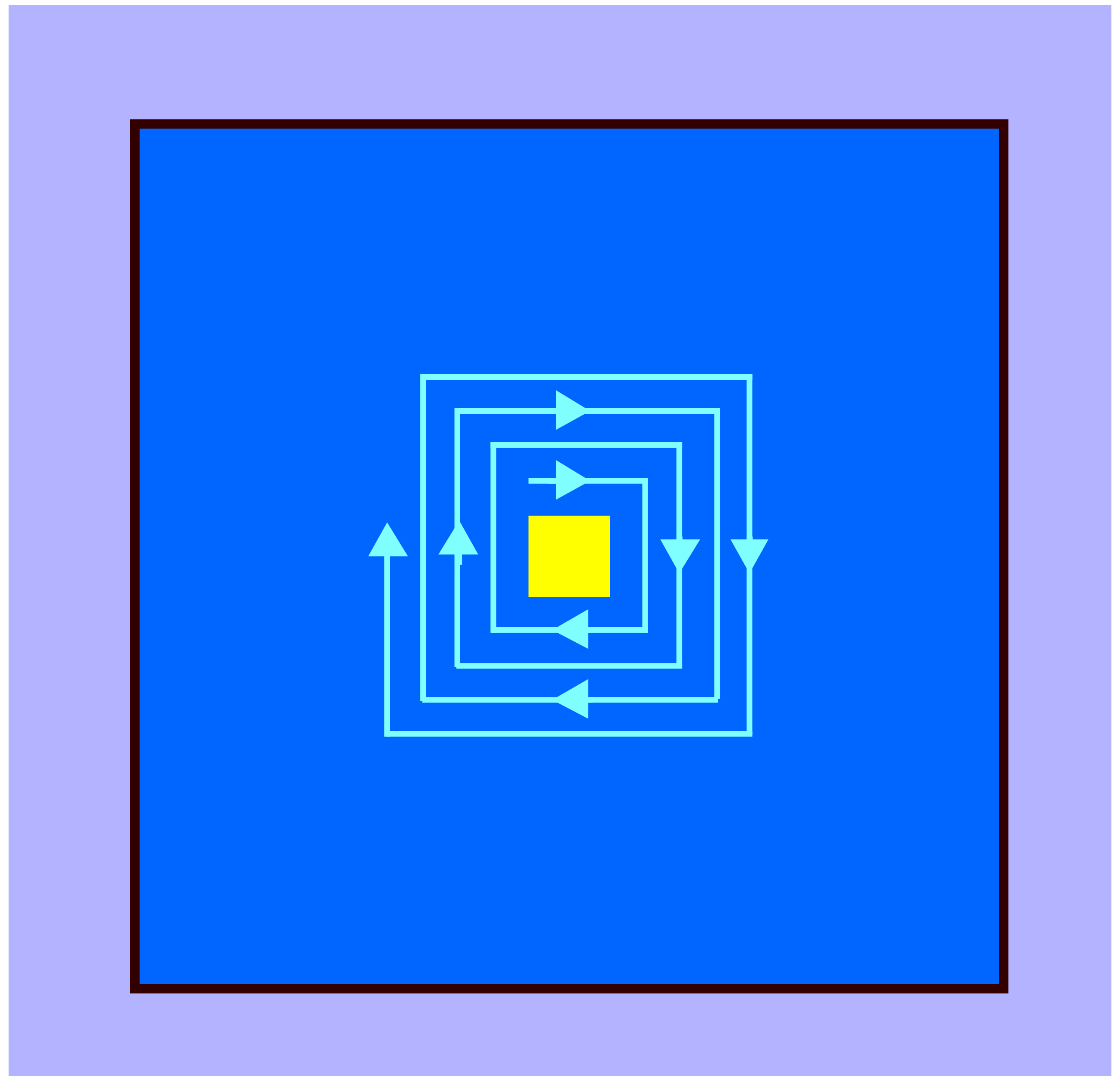
challenge: what’s the layout?

The data…

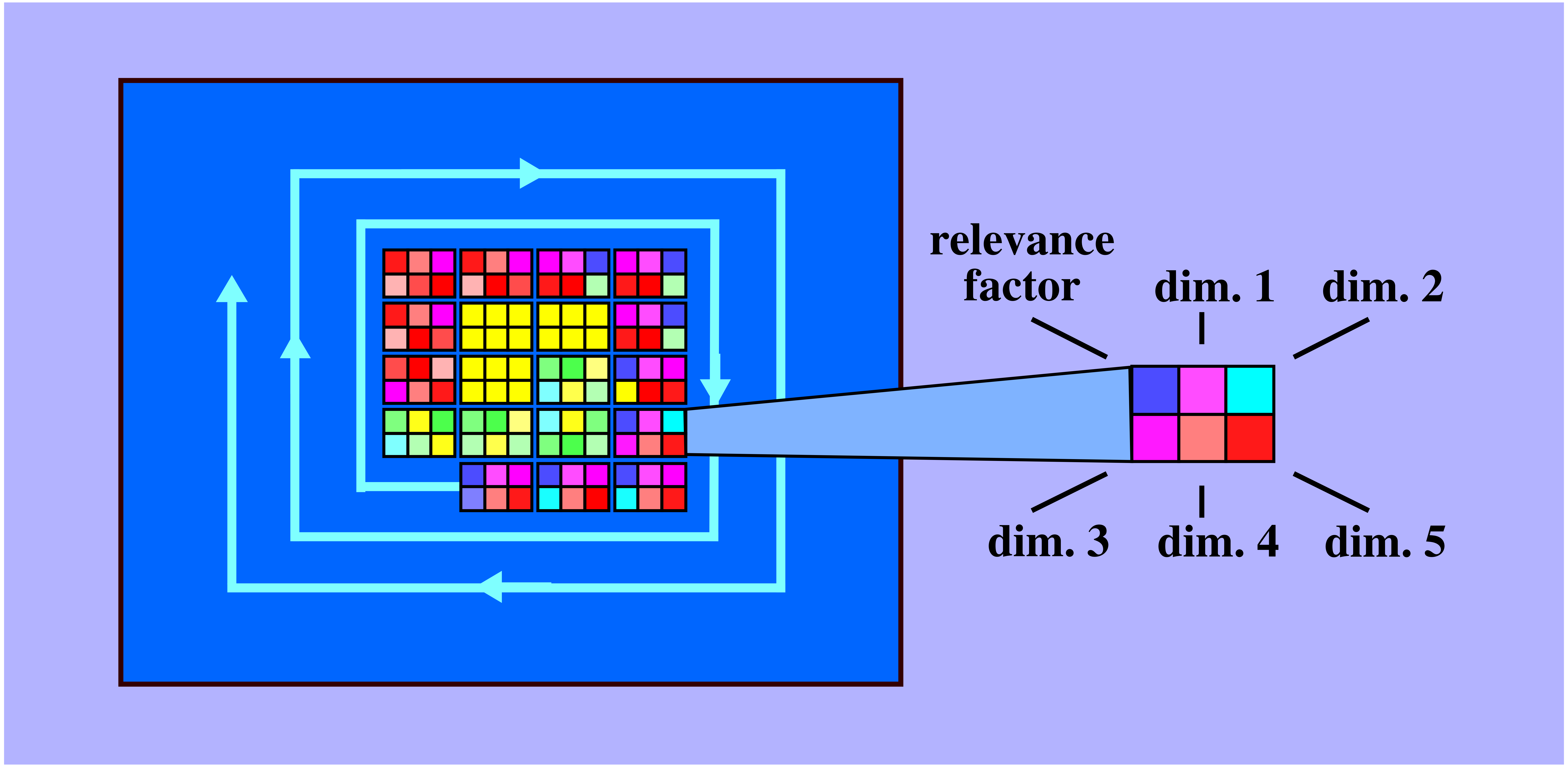
large database where each item has multiple attributes (on the order of 10)

goal: visualize the relevance of set of items which satisfy a query

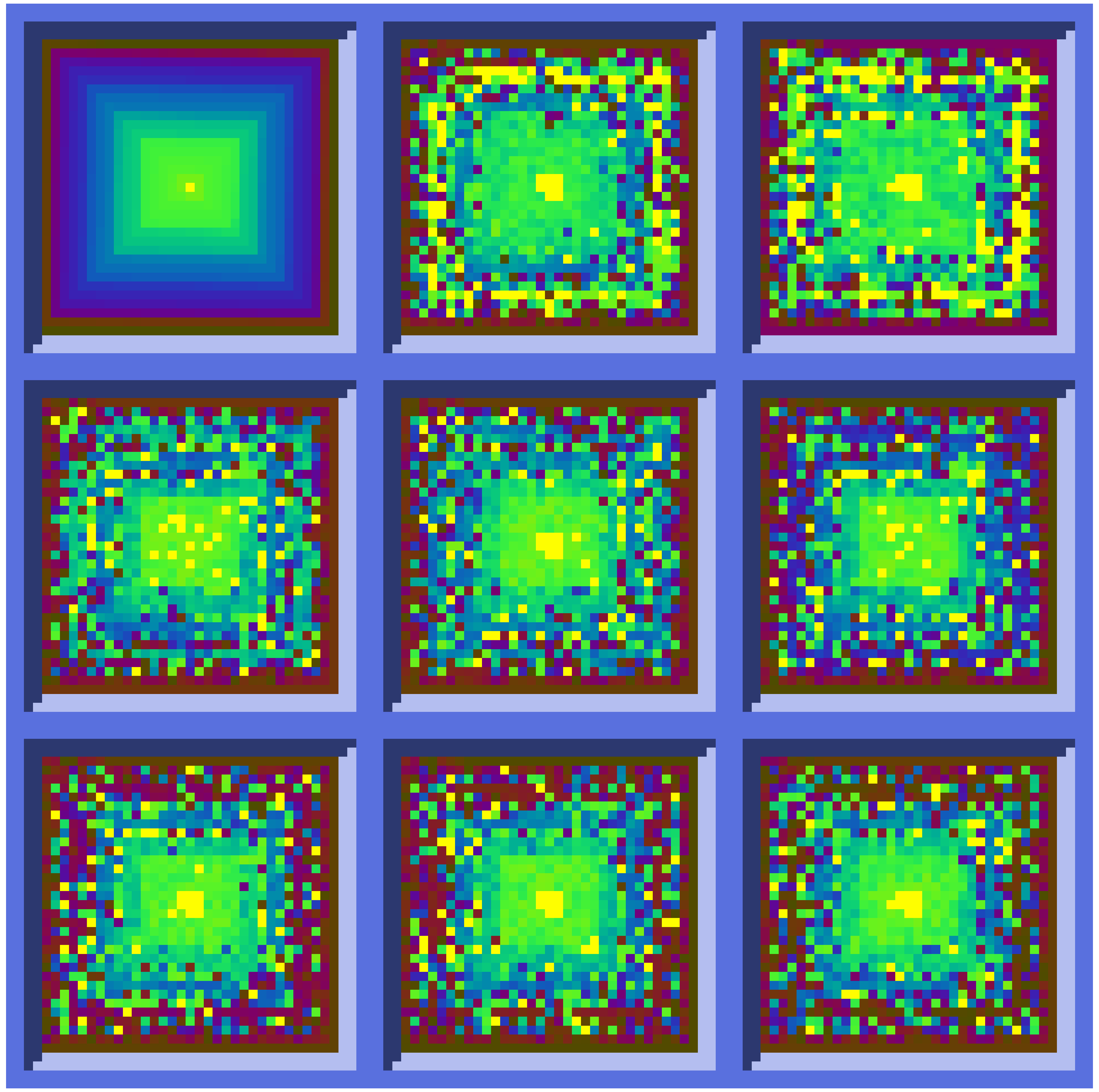
plot out data items in a spiral pattern, ordered by relevance

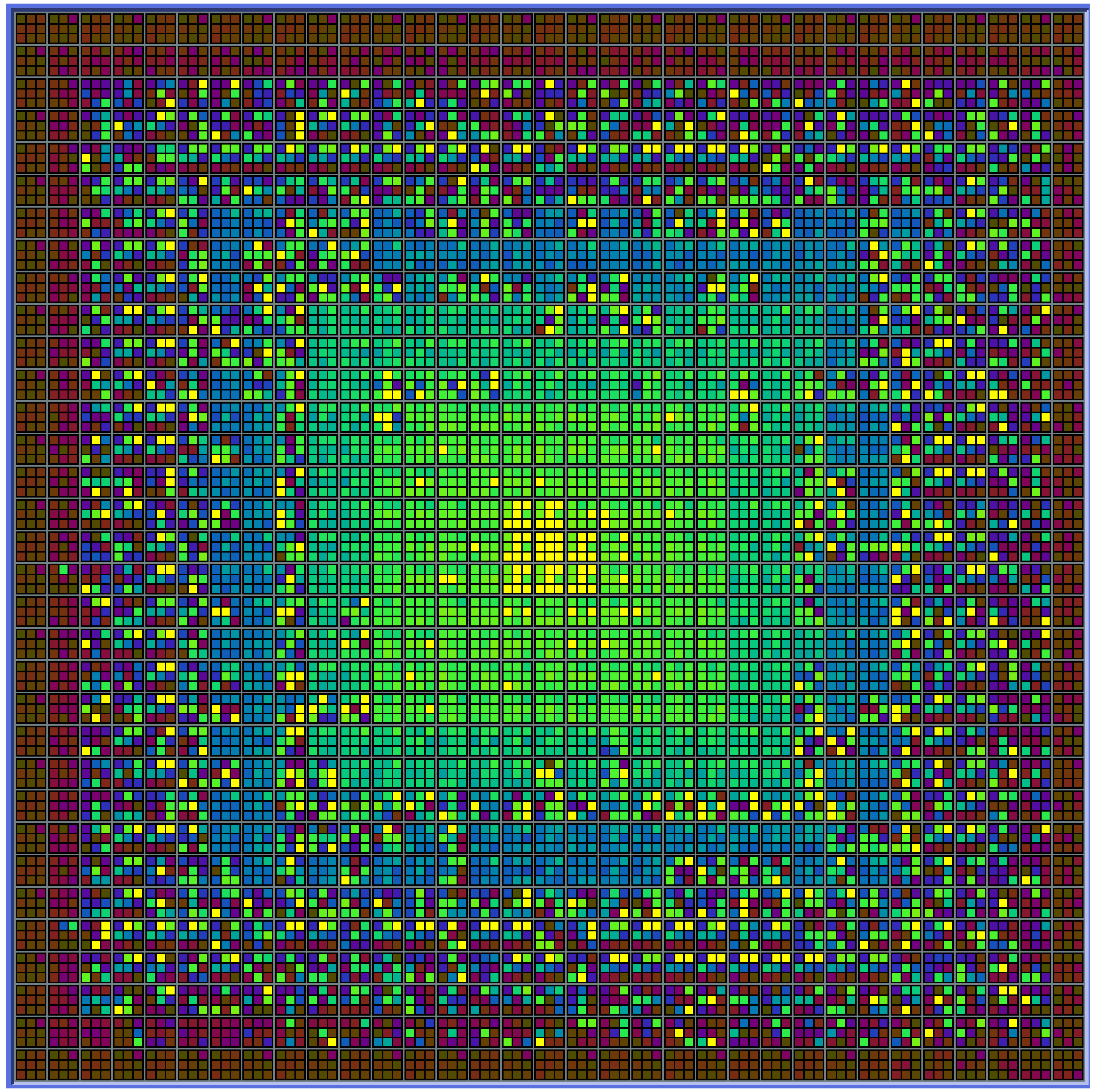


Keim, Kreigel, 1994



Keim, Kreigel, 1994





Keim, Kreigel, 1994