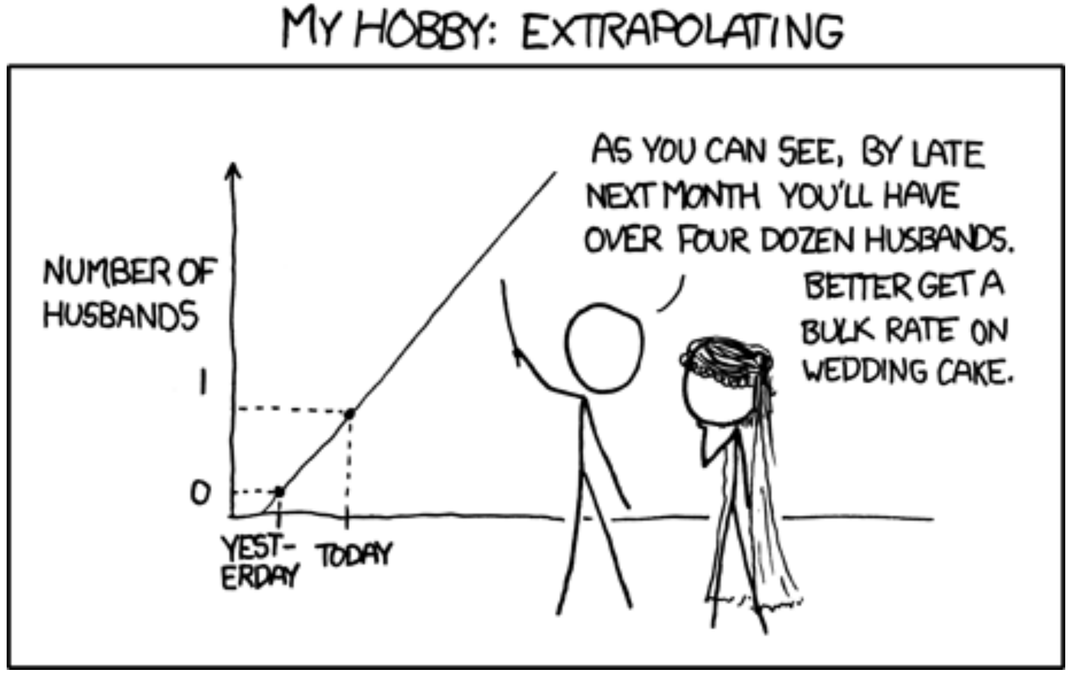
CS-5630 / CS-6630 Visualization for DataScience Tables

Alexander Lex
alex@sci.utah.edu

THE UNIVERSITY OF UTAH



[xkcd]

Organizational

Review exam in my office hours

HW Lab: Wed, 6pm, L110

Make sure to form your project teams!

- If you can’t find a team, e-mail me

Develop project idea

Set up your github repo

Guest lecture next Thursday

Project Feedback the Tuesday after that

Need to submit this info by Friday!

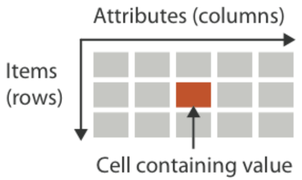
https://goo.gl/4UrmjB

dataset types

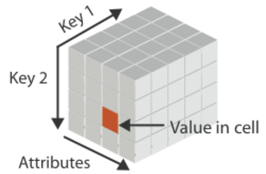
Tables

Items

Attributes



➔ *Multidimensional Table*

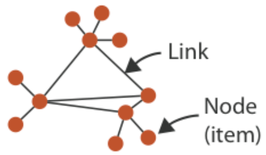


Networks & Trees

Items (nodes)

Links

Attributes



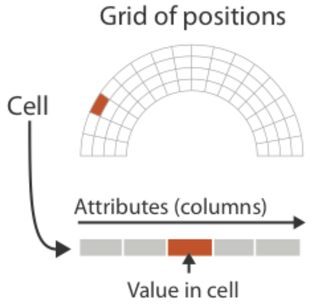
➔ *Trees*

Fields

Grids

Positions

Attributes



Geometry

Items

Positions

Clusters, Sets, Lists

Items

Exercise: Sketch 2 Ways to Vis. Each Table

	Age	Best 100 m	Furthest Jump	Sex
Amy	16	13.2	5.2	F
Basil	18	12.4	4.2	F
Clara	14	14.1	2.5	F
Desmond	22	10.01	6.3	M
Charles	19	11.3	5.3	M

	BPM T1	BPM T2	BPM T3
Amy	90	130	150
Basil	70	110	109
Clara	60	140	141
Desmond	84	100	108
Charles	81	110	130

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

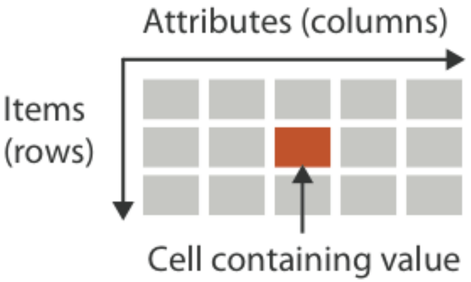
- Space-Filling

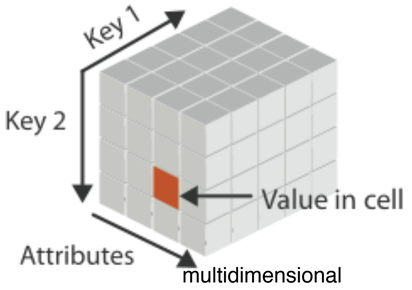
Spatial channels are the most effective for all attribute types

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

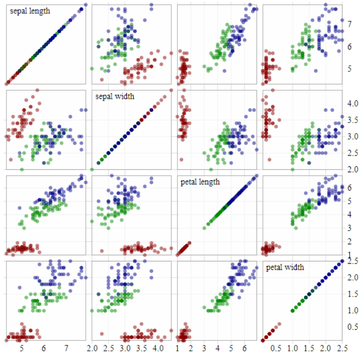
Same data type?

Same scales?

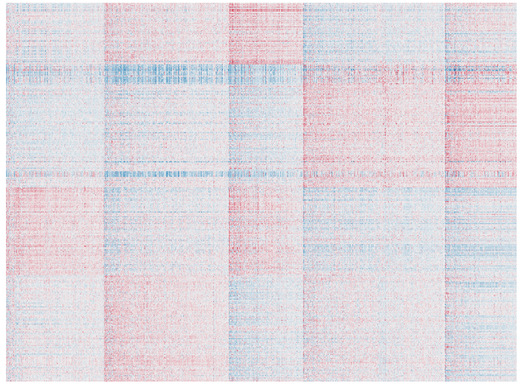
	Age	Gender	Height
Bob	*25*	*M*	*181*
Alice	*22*	*F*	*185*
Chris	*19*	*M*	*175*

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

Analytic Component

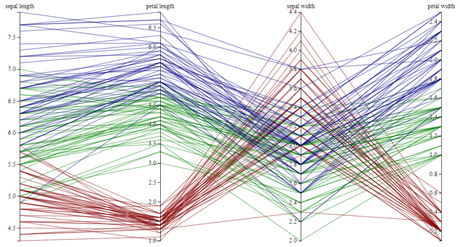


Scatterplot Matrices
[Bostock]

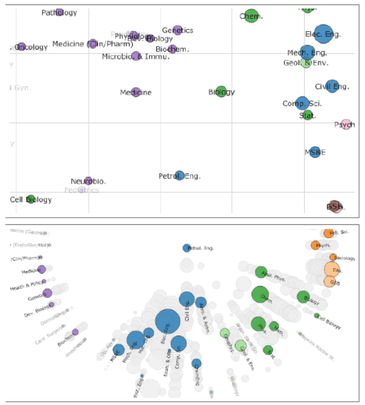


Pixel-based visualizations / heat maps

Multidimensional Scaling
[Doerk 2011]



Parallel Coordinates
[Bostock]



[Chuang 2012]

no / little analytics ⟷ **strong analytics component**

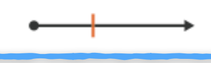
Express Values

No Keys

Encode using zero keys: scatterplots

Arrange Tables

- Express Values

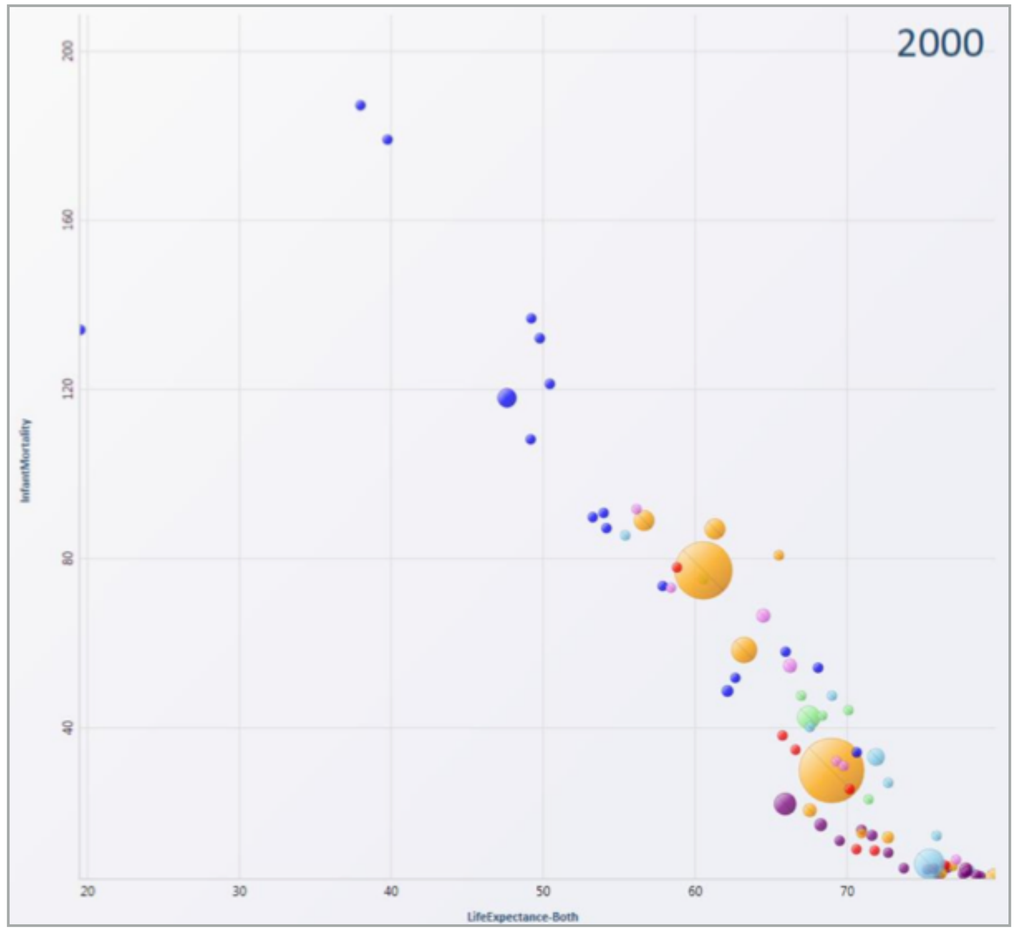


- Separate, Order, Align Regions
  - Separate
  - Order
  - Align

- Axis Orientation
  - Rectilinear
  - Parallel
  - Radial

- Layout Density
  - Dense
  - Space-Filling

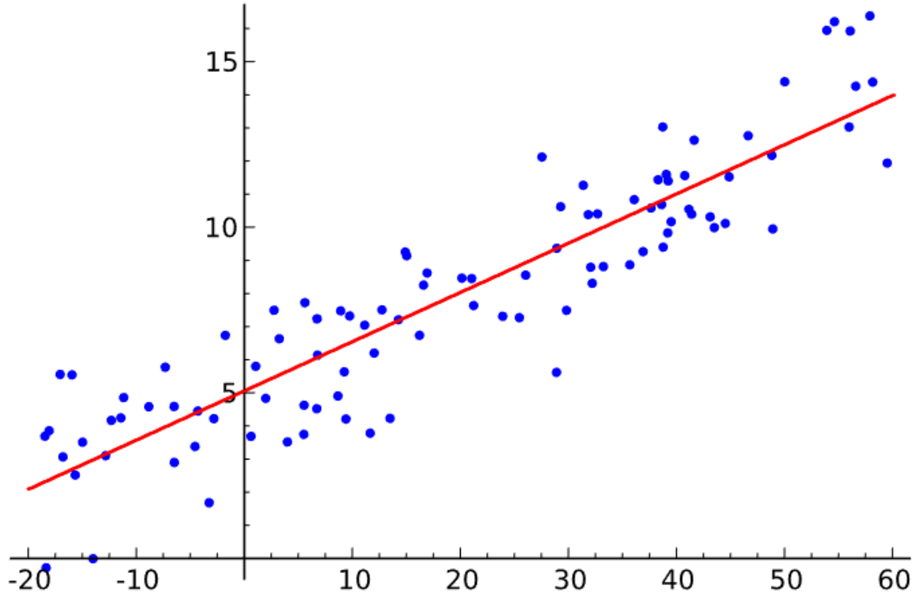
Infant Mortality



Life Expectance

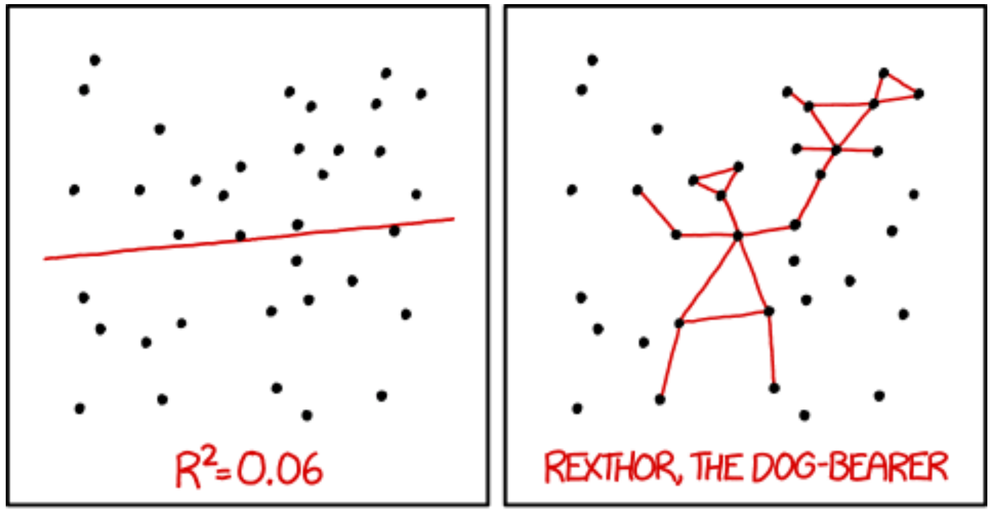
Regression Lines

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

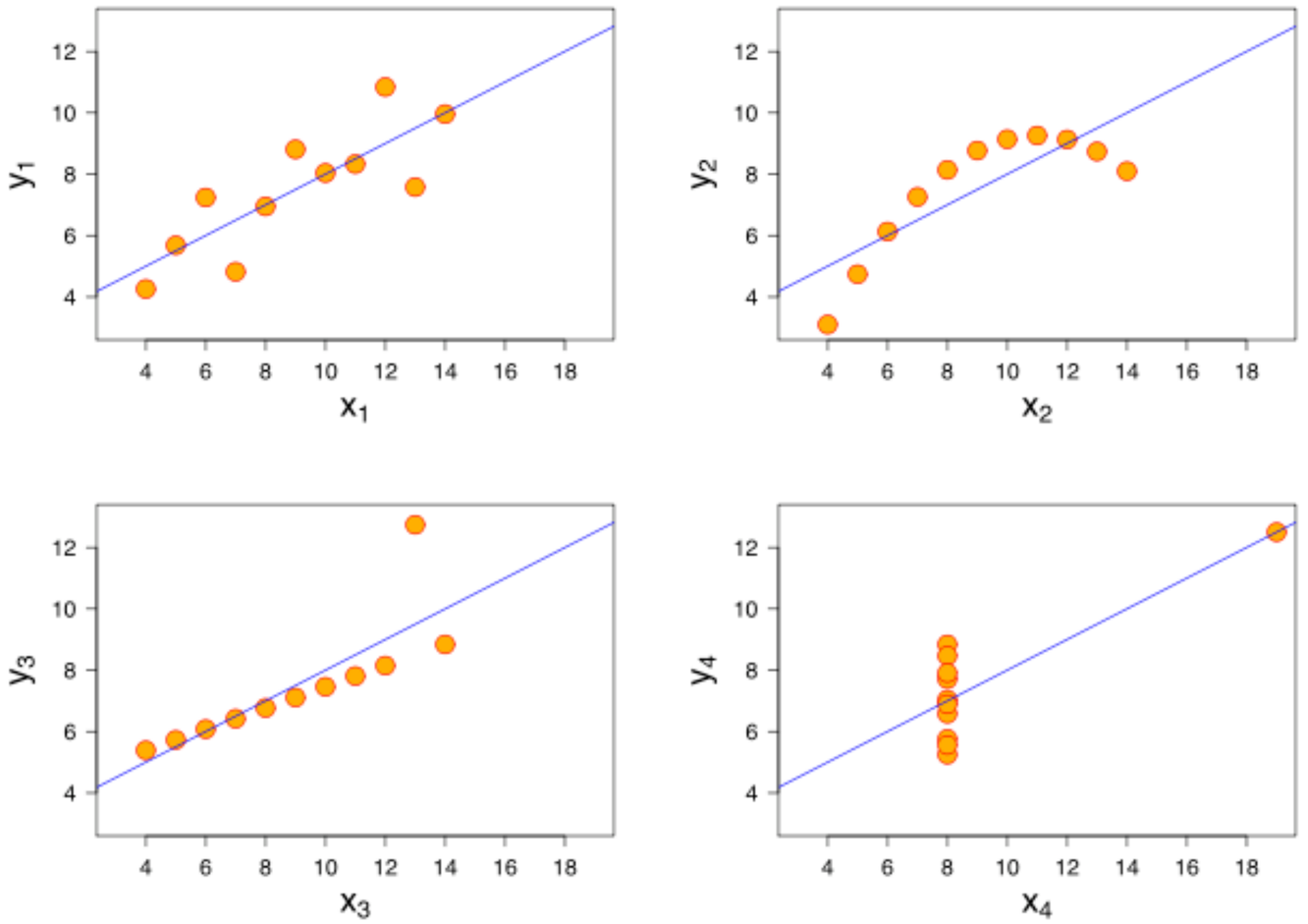


Goal: Find the best values of β_0 and β_1, denoted $\hat{\beta}_0$ and $\hat{\beta}_1$, so that the prediction $y = \hat{\beta}_0 + \hat{\beta}_1 x$ "best fits" the data.

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

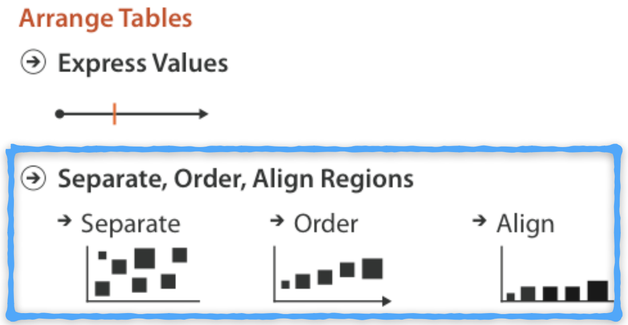


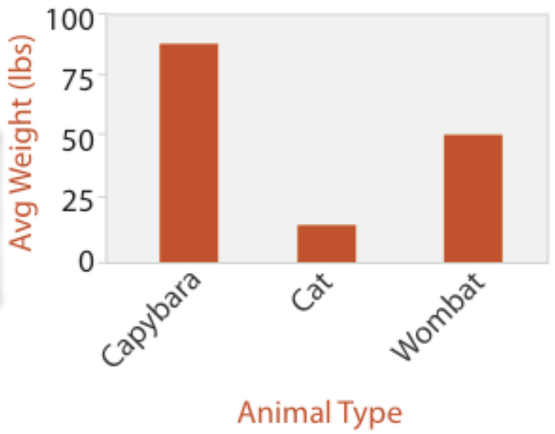
Anscombe's Quartet

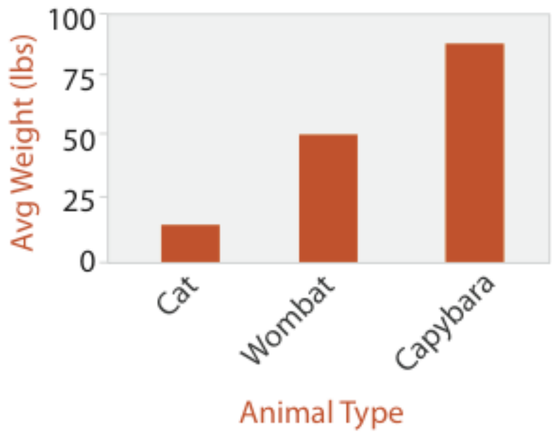


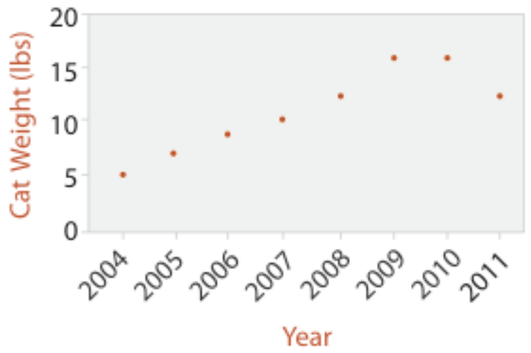
Encode one Key Attribute

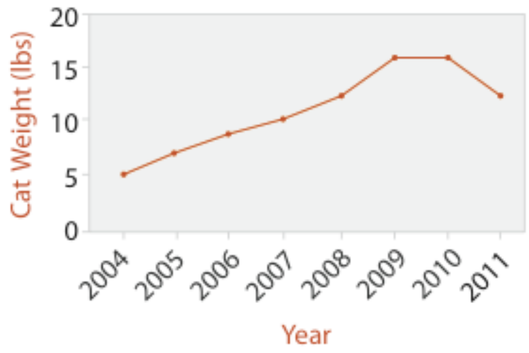
Encode one key attribute: bar, dot, & line charts











Encode Multiple Key Attributes

Arrange Tables

Express Values

Separate, Order, Align Regions

- Separate

- Order

- Align

- *1 Key*
  List

- *2 Keys*
  Matrix

- *3 Keys*
  Volume

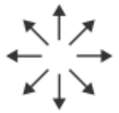
- *Many Keys*
  Recursive Subdivision

Axis Orientation

- Rectilinear

- Parallel

- Radial



Layout Density

- Dense

- Space-Filling

Stacked Bar Chart

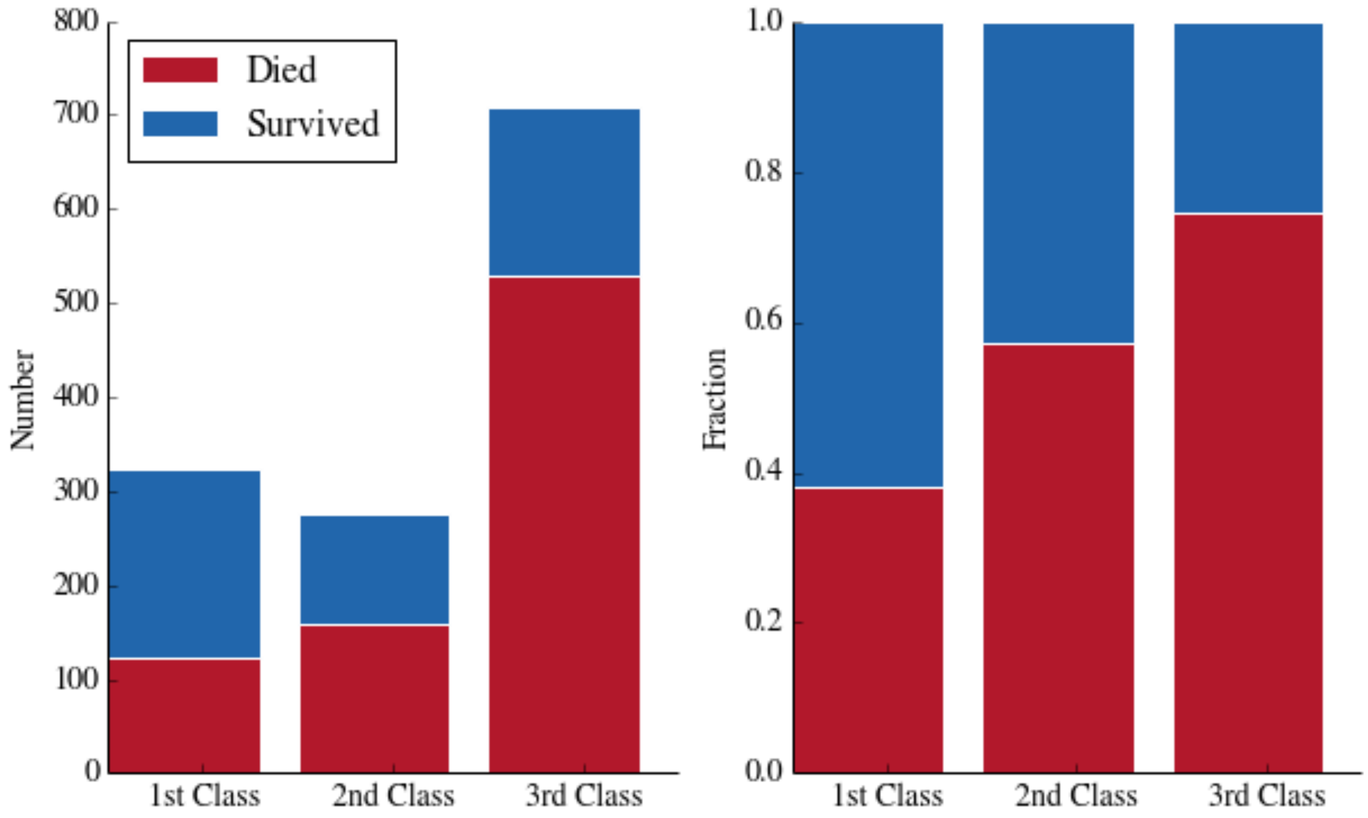
Keys: Class, Survival

Class is spatial

Survival is color

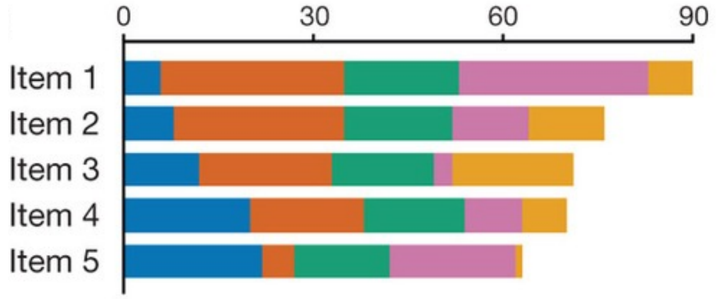
Left: absolute values

Right: proportional values



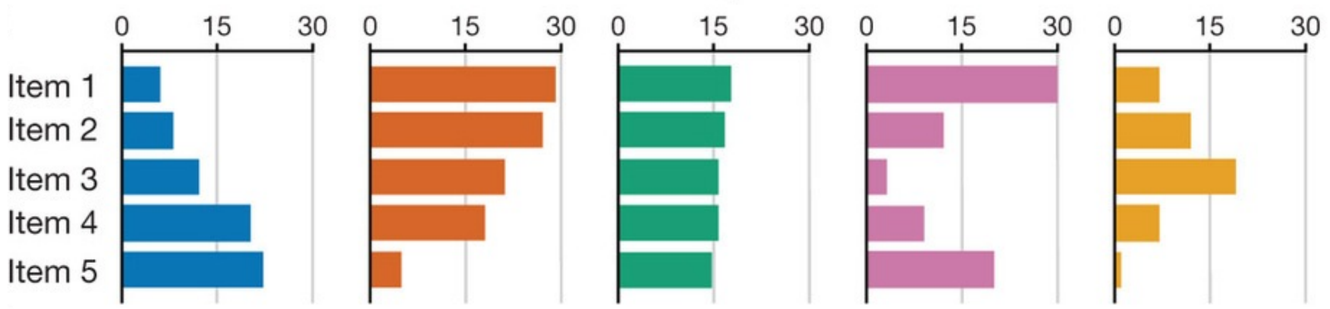
Comparison of bar chart types

Pie Chart

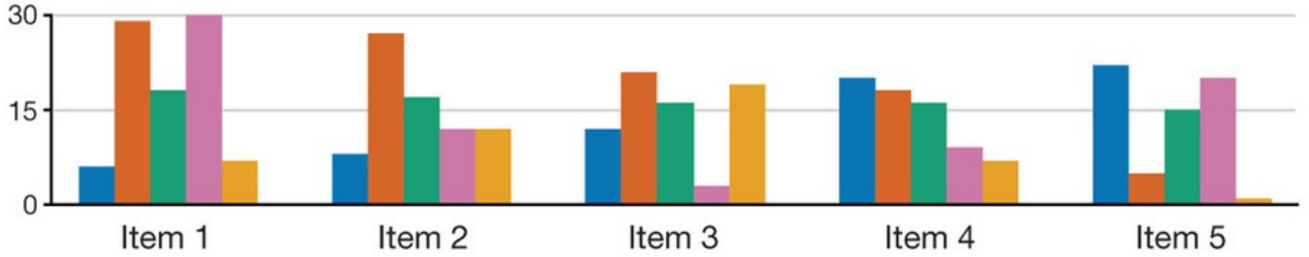


Stacked bar chart

Layered Bar Chart

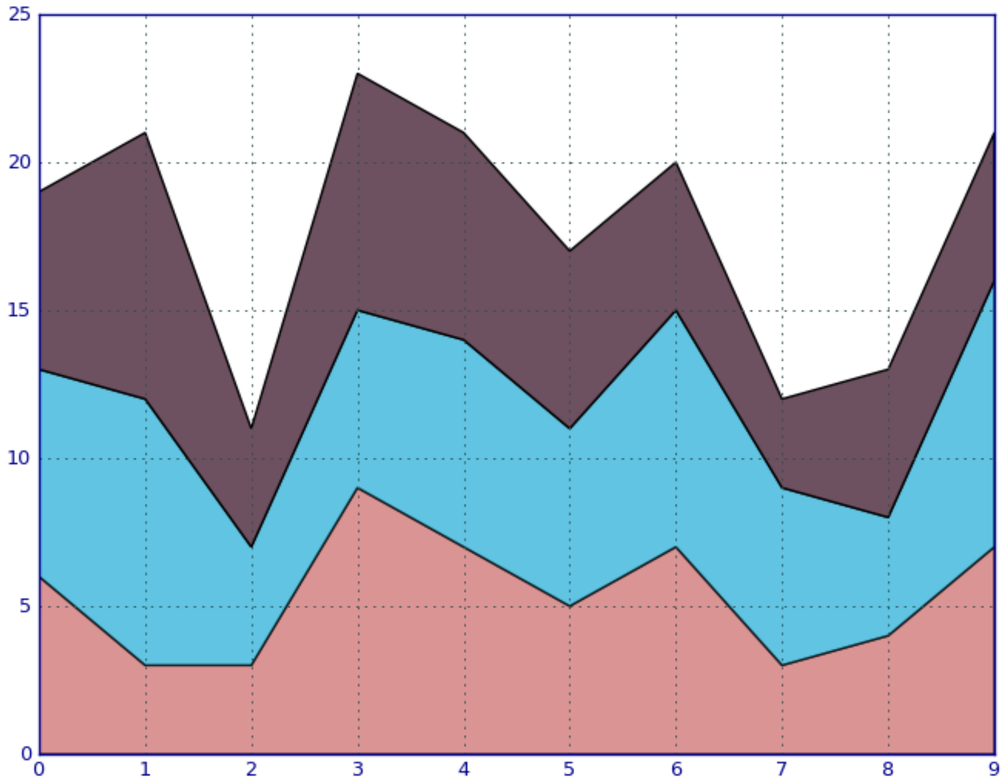


Grouped Bar Chart

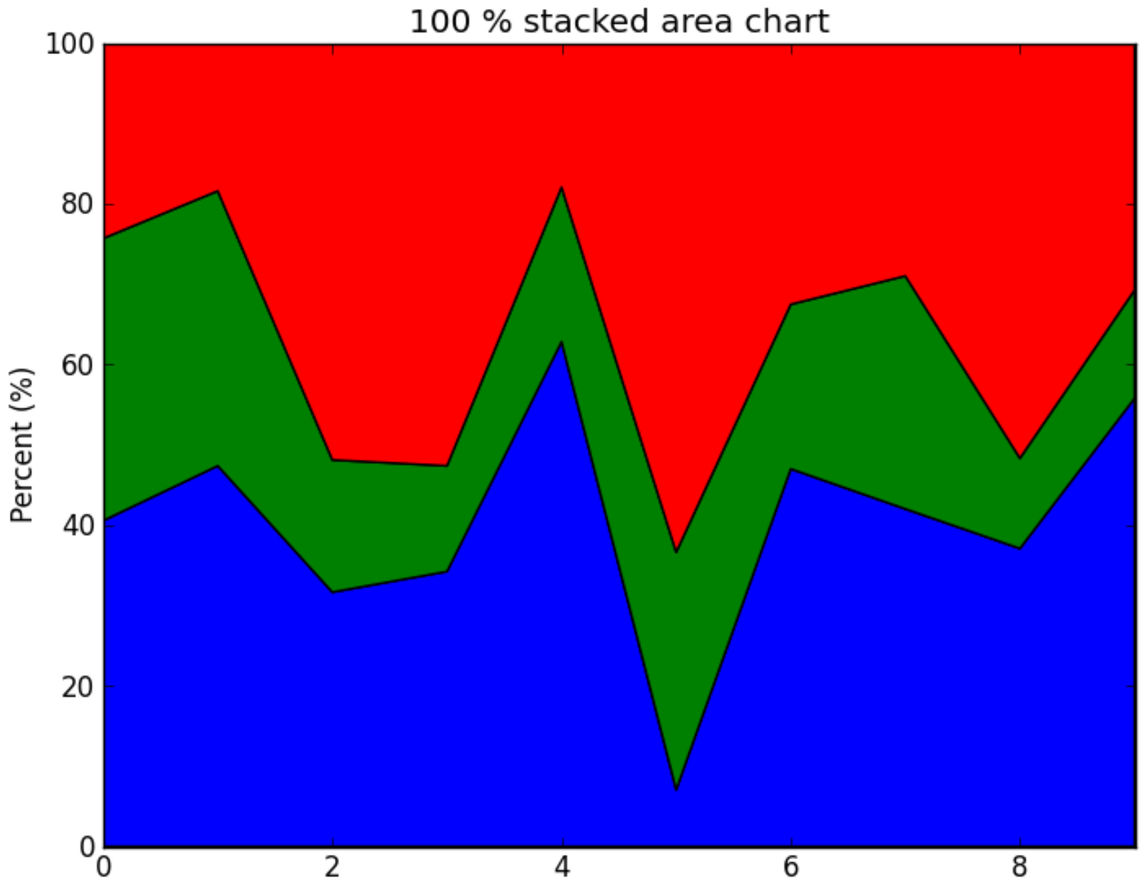


Streit & Gehlenborg, PoV, Nature Methods, 2014

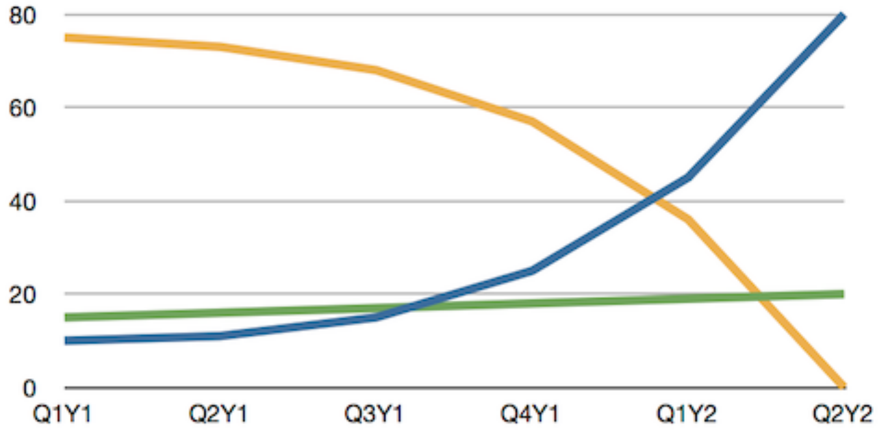
Stacked Area Chart



100% Stacked Area Chart

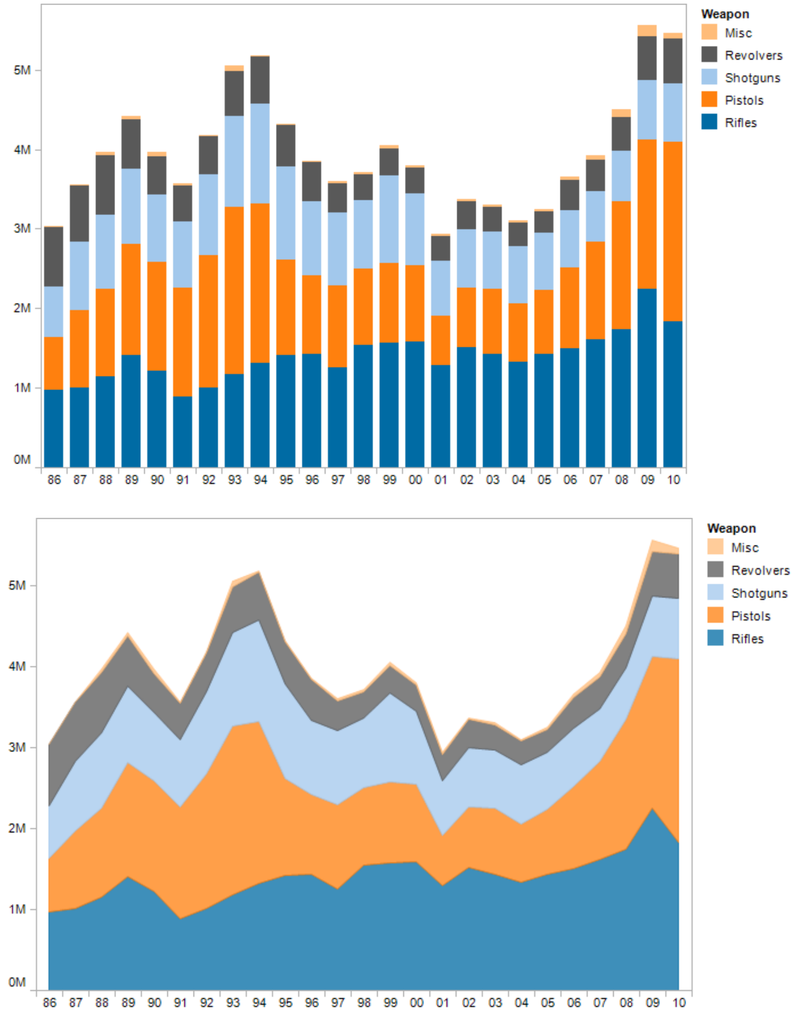


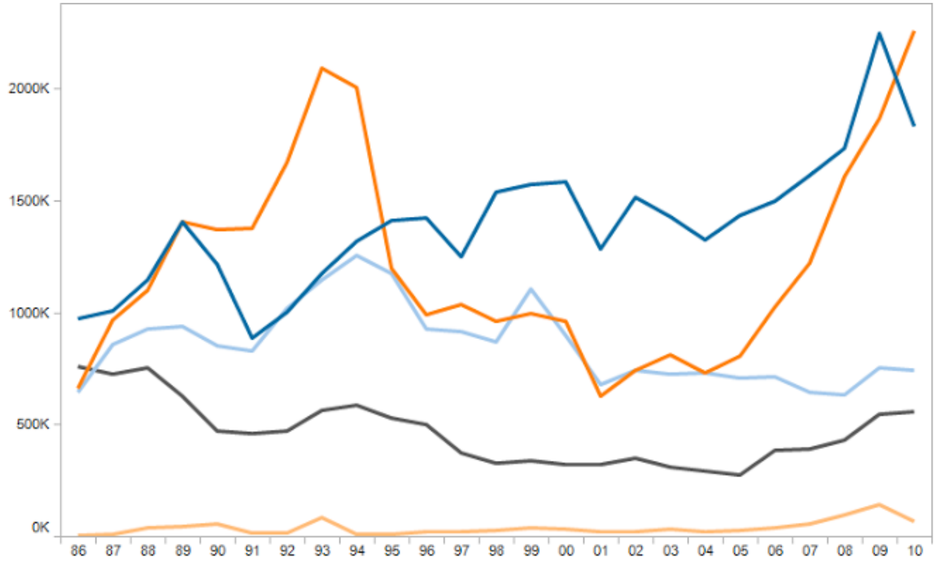
Stacked Area vs. Line Graphs



leancrew.com &
Practically Efficient

Can you spot the trends?





VizWiz, A. Kriebel

Tabular / Grid / Matrix - Based Representations

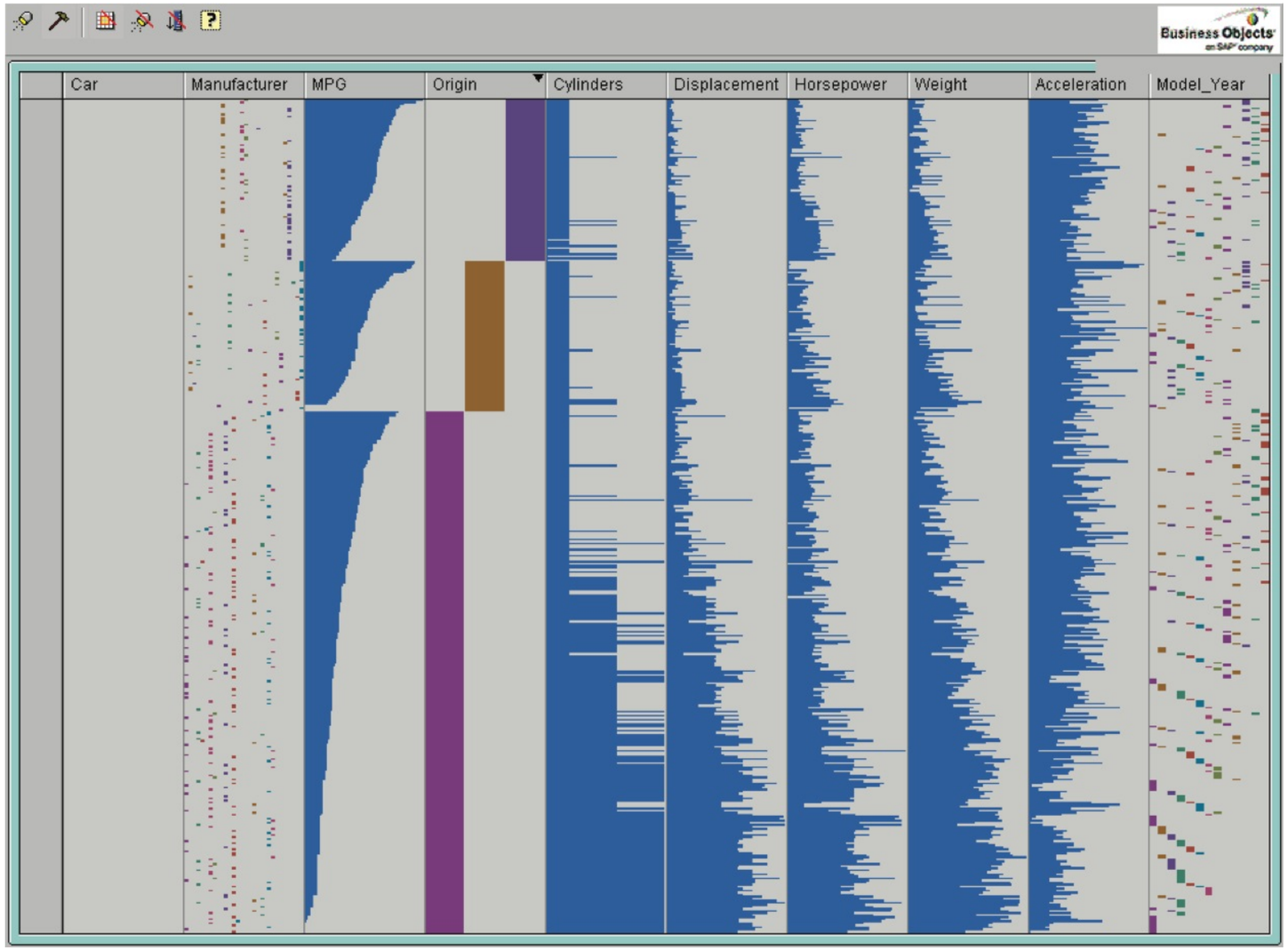
Tabular Representation

Like spreadsheet - each variable in it’s own column

Visual encodings to make it scalable

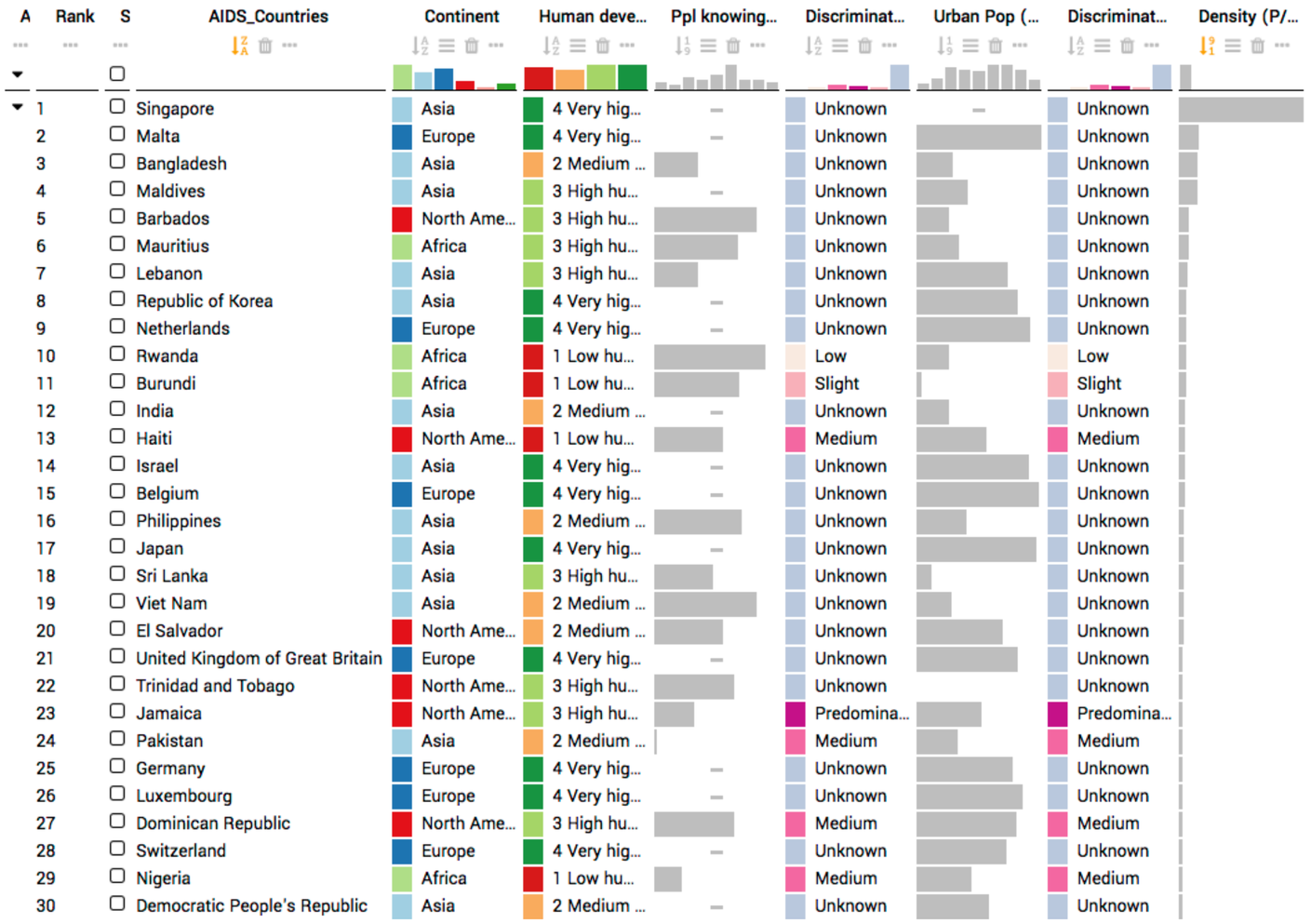
Table Lens

Interactive table-based representation



Rao & Card 1994

Taggle

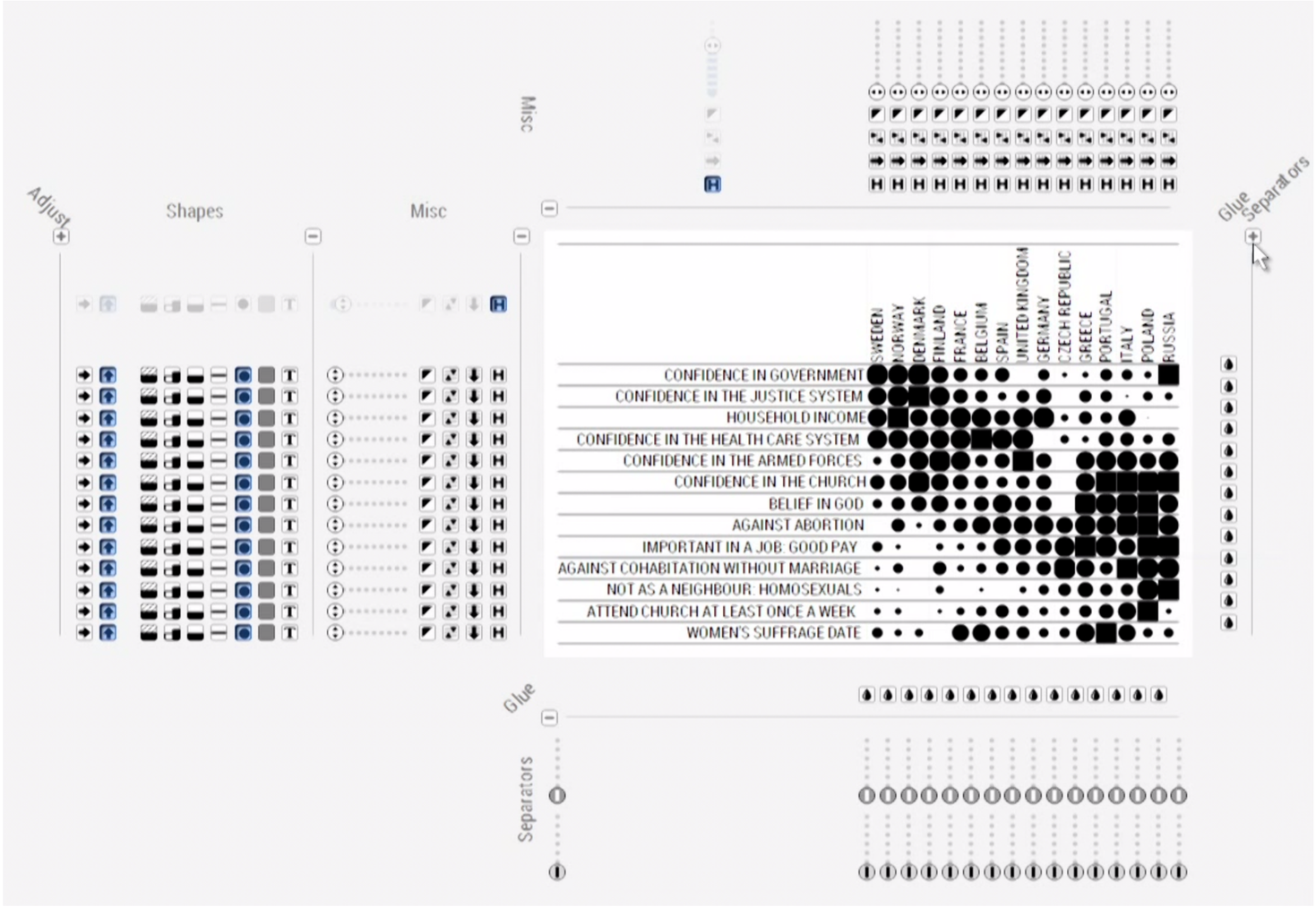


A	Rank	S	AIDS_Countries	Continent	Human deve...	Ppl knowing...	Discriminat...	Urban Pop (...	Discriminat...	Density (P/...
				Asia	4 Very hig...		Unknown		Unknown	
	1		Singapore	Asia	4 Very hig...		Unknown		Unknown	
	2		Malta	Europe	4 Very hig...		Unknown		Unknown	
	3		Bangladesh	Asia	2 Medium ...		Unknown		Unknown	
	4		Maldives	Asia	3 High hu...		Unknown		Unknown	
	5		Barbados	North Ame...	3 High hu...		Unknown		Unknown	
	6		Mauritius	Africa	3 High hu...		Unknown		Unknown	
	7		Lebanon	Asia	3 High hu...		Unknown		Unknown	
	8		Republic of Korea	Asia	4 Very hig...		Unknown		Unknown	
	9		Netherlands	Europe	4 Very hig...		Unknown		Unknown	
	10		Rwanda	Africa	1 Low hu...		Low		Low	
	11		Burundi	Africa	1 Low hu...		Slight		Slight	
	12		India	Asia	2 Medium ...		Unknown		Unknown	
	13		Haiti	North Ame...	1 Low hu...		Medium		Medium	
	14		Israel	Asia	4 Very hig...		Unknown		Unknown	
	15		Belgium	Europe	4 Very hig...		Unknown		Unknown	
	16		Philippines	Asia	2 Medium ...		Unknown		Unknown	
	17		Japan	Asia	4 Very hig...		Unknown		Unknown	
	18		Sri Lanka	Asia	3 High hu...		Unknown		Unknown	
	19		Viet Nam	Asia	2 Medium ...		Unknown		Unknown	
	20		El Salvador	North Ame...	2 Medium ...		Unknown		Unknown	
	21		United Kingdom of Great Britain	Europe	4 Very hig...		Unknown		Unknown	
	22		Trinidad and Tobago	North Ame...	3 High hu...		Unknown		Unknown	
	23		Jamaica	North Ame...	3 High hu...		Predomina...		Predomina...	
	24		Pakistan	Asia	2 Medium ...		Medium		Medium	
	25		Germany	Europe	4 Very hig...		Unknown		Unknown	
	26		Luxembourg	Europe	4 Very hig...		Unknown		Unknown	
	27		Dominican Republic	North Ame...	3 High hu...		Medium		Medium	
	28		Switzerland	Europe	4 Very hig...		Unknown		Unknown	
	29		Nigeria	Africa	1 Low hu...		Medium		Medium	
	30		Democratic People's Republic	Asia	2 Medium ...		Unknown		Unknown	

Bertifier

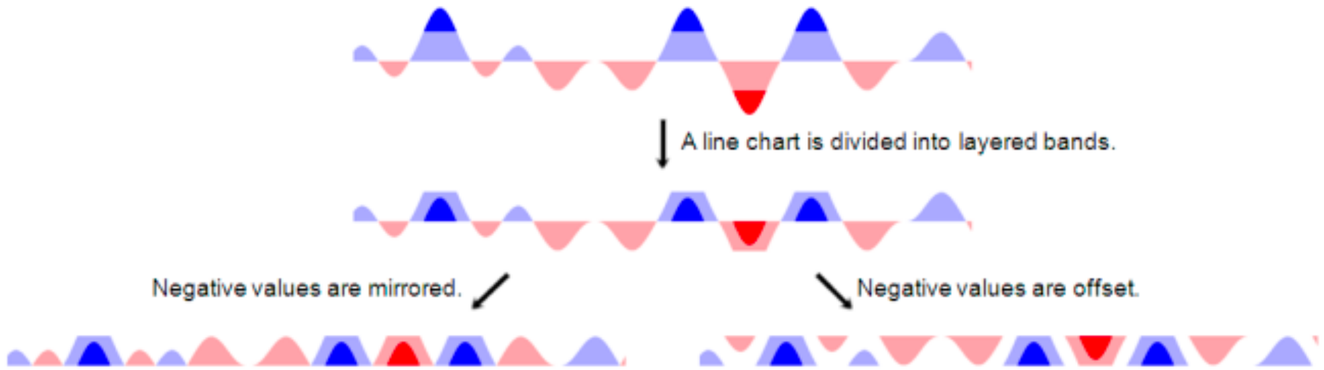
Matrix/Table representation

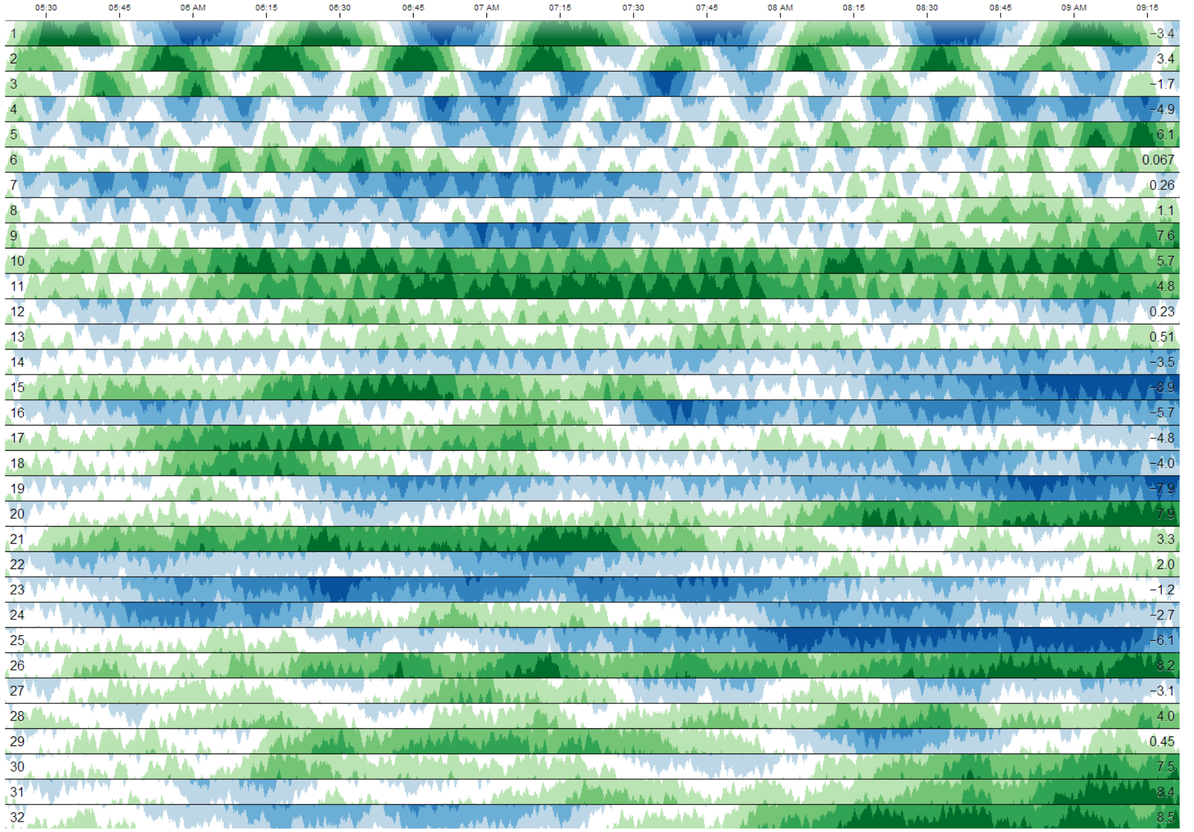
Authoring Interface



http://www.aviz.fr/bertifier
Charles Perin, Pierre Dragicevic and Jean-Daniel Fekete

Multiple Line Charts

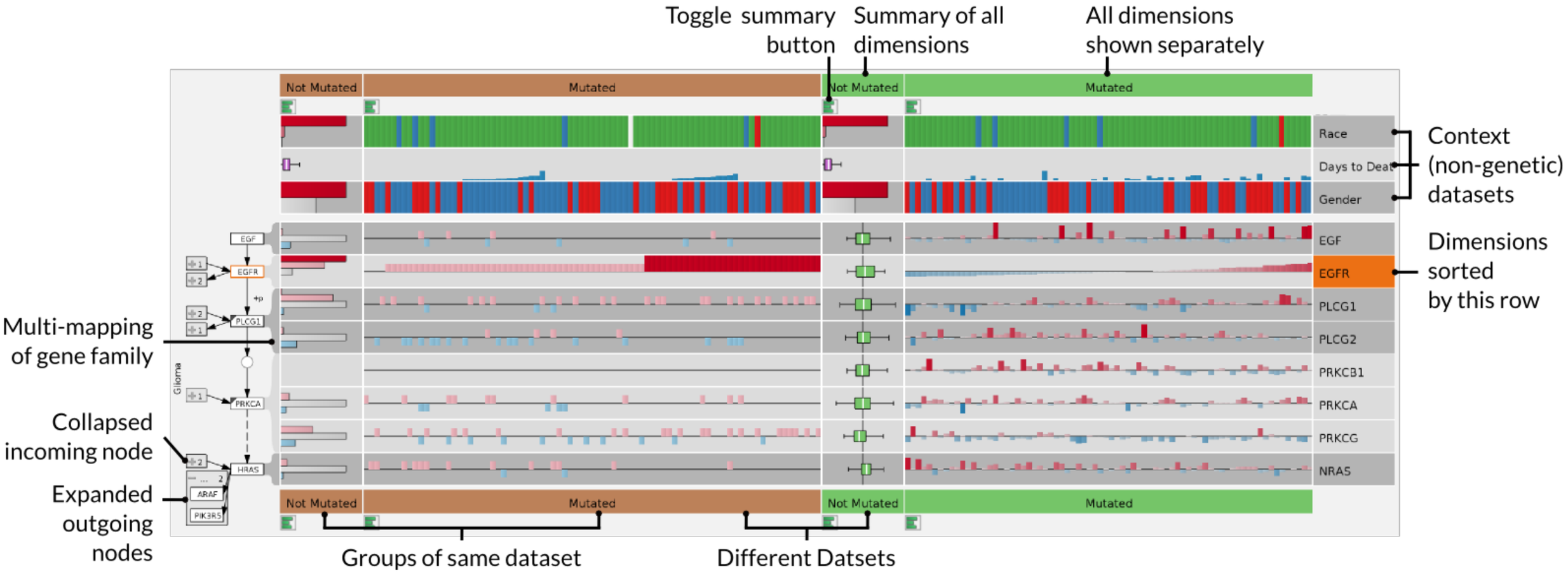




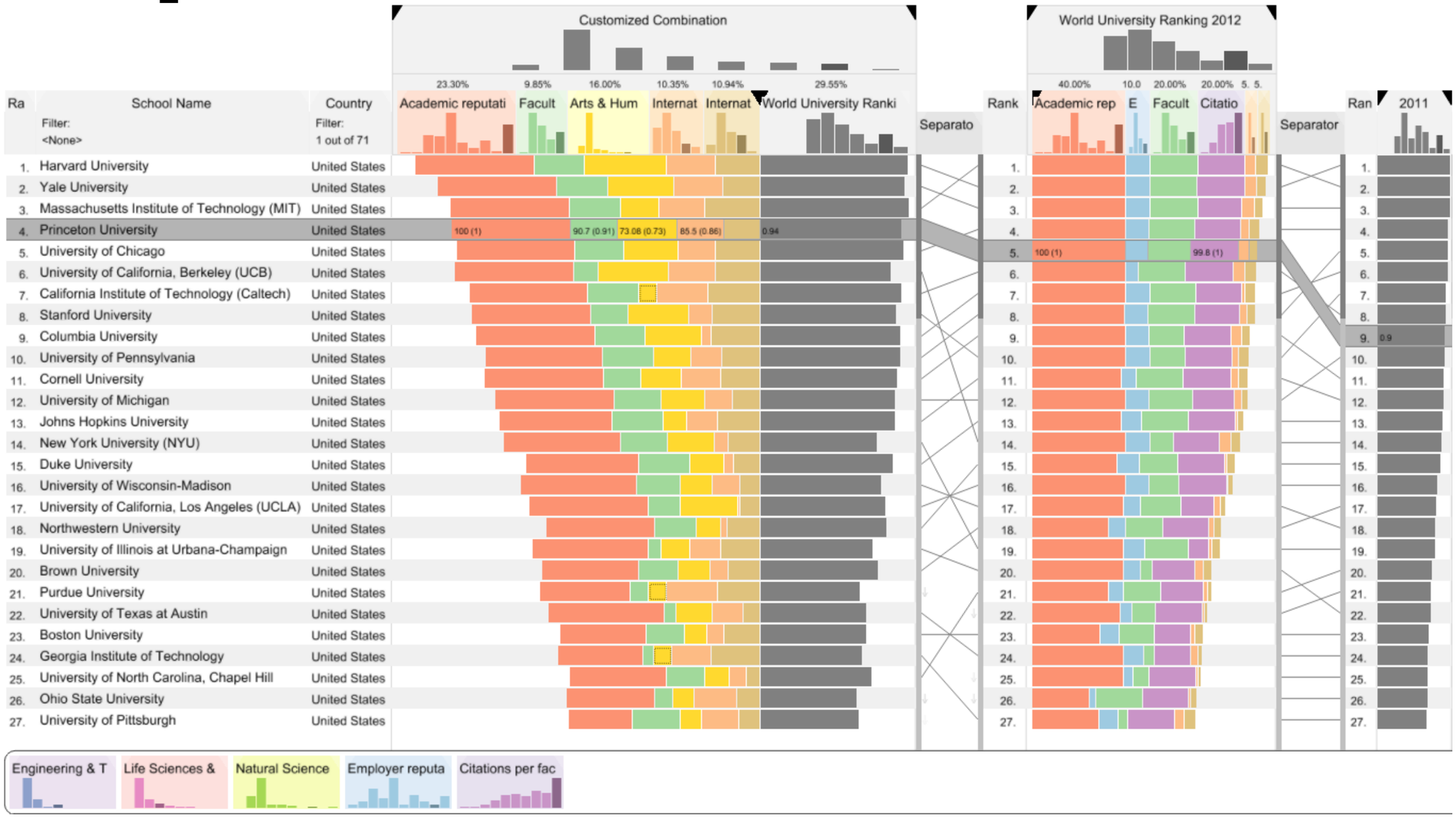
http://square.github.io/cubism/

[Heer, Sizing the Horizon, 2009]

Combining Various Charts



LineUp



Video at http://lineup.caleydo.org

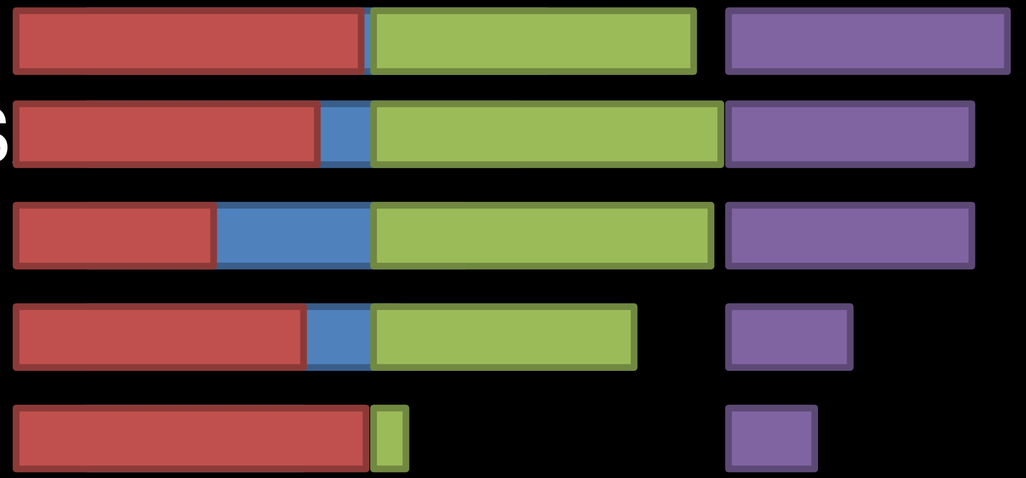

Rankings are popular

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

Support Multiple Attributes

Score = f(A, B, C)

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



Combiner functions: f(A,B,C)

(Weighted) sum

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

→ Serial

Maximum

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

→ Parallel

Product

Nesting

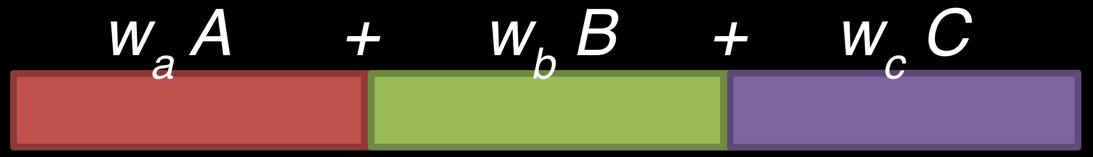
…

→ 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)



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

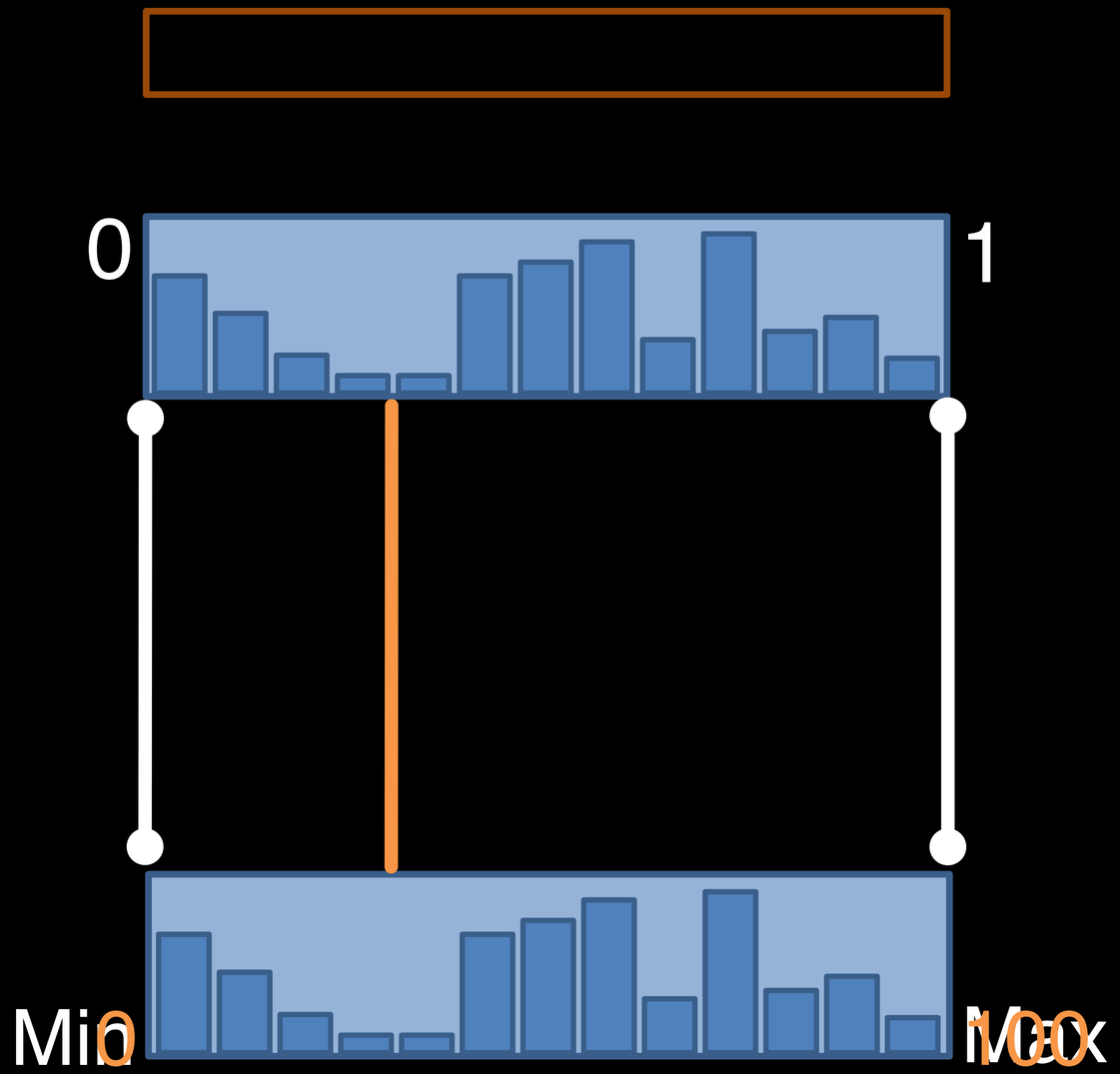
Serial Combiner (as Stacked Bar)

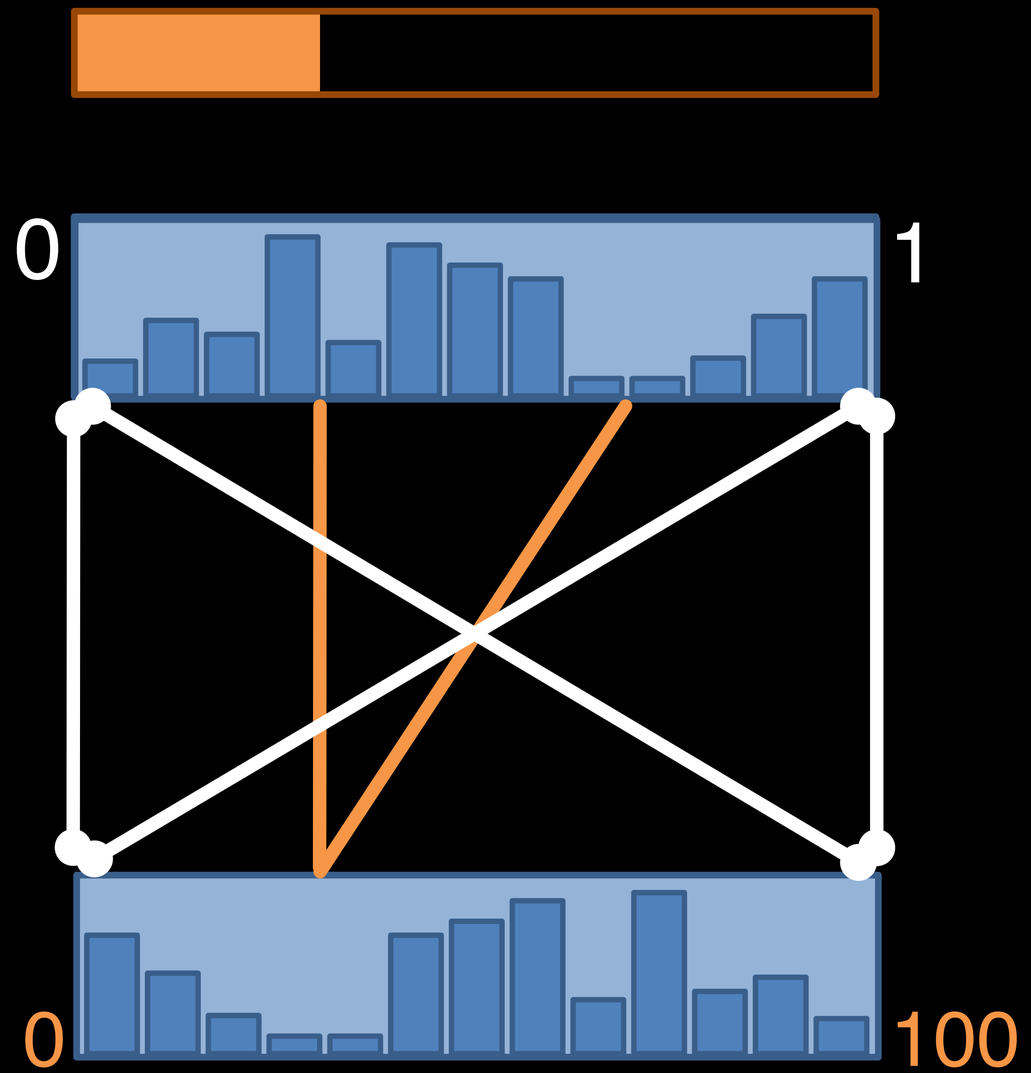
Rank	University	A	B	C
1.	MIT, USA			
2.	Harvard, USA			
3.	Princeton, USA			
4.	Cambridge, UK			
5.	Oxford, UK			

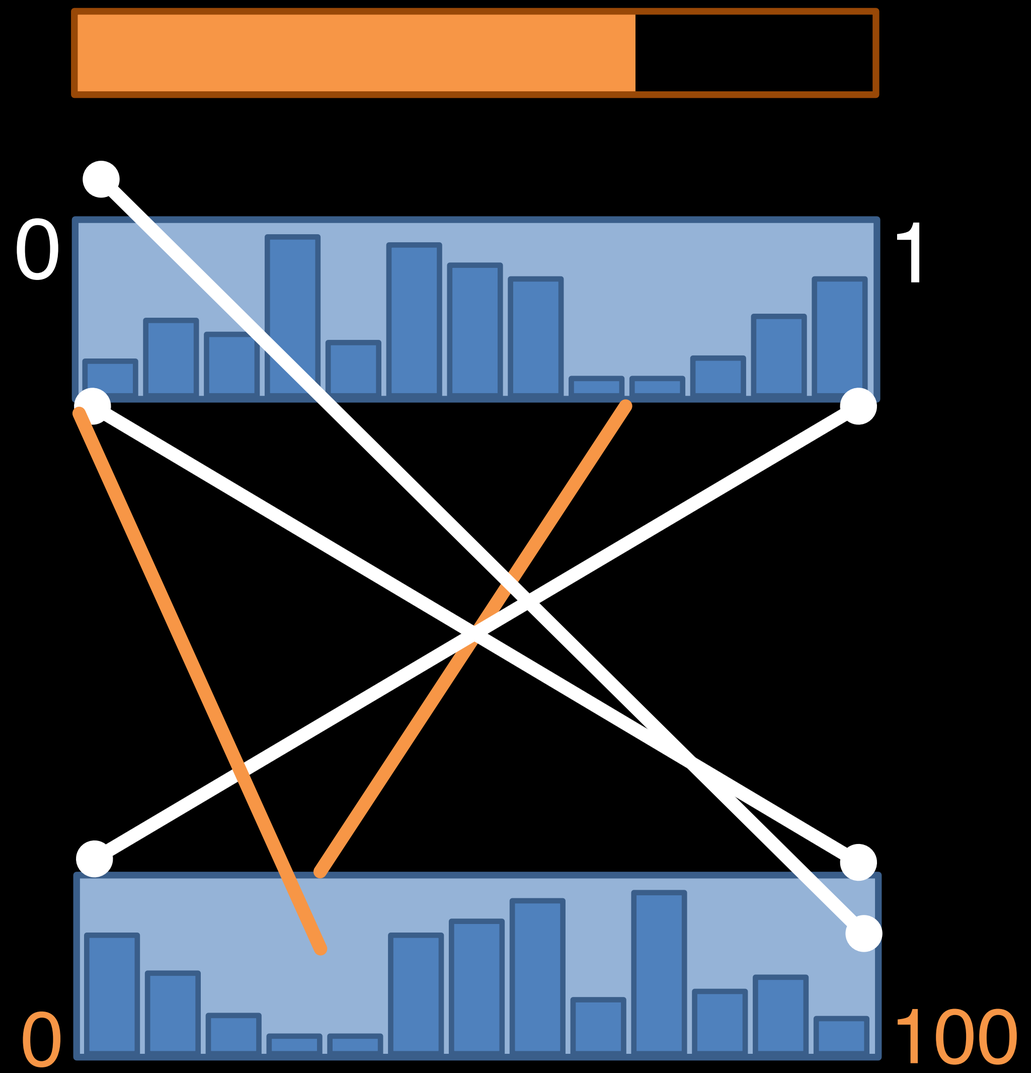
128 visible of 906 (14.13%)

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			

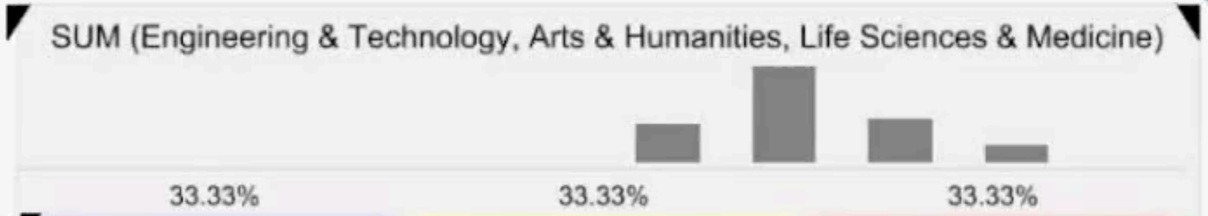
Flexible Mapping of Attributes to Scores


0
1
Min
0
Max
100


0
1
0
100


0
1
0
100

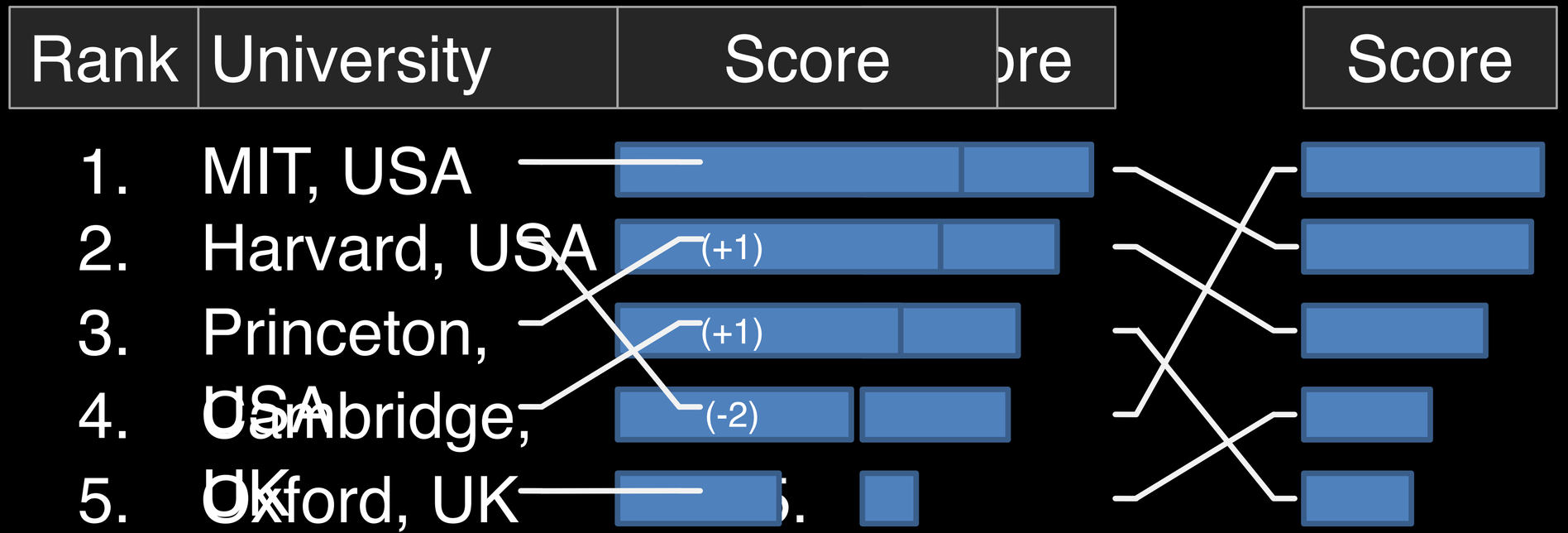
93 visible of 399 (23.31%)



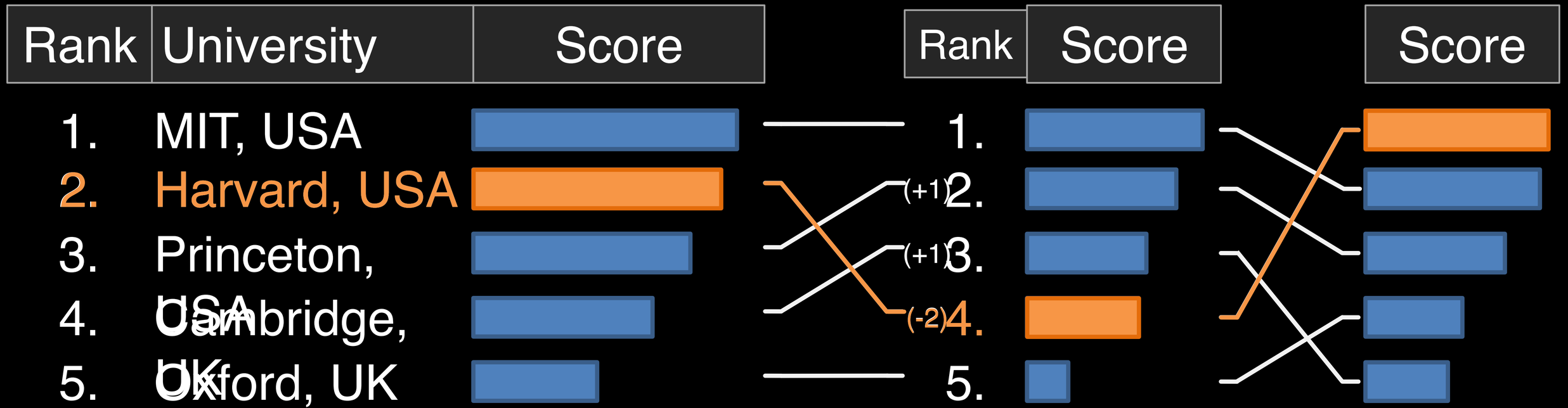
Ran	School Name	Country
	Filter: <None>	Filter: 2 out of 43
1.	University of Oxford	United Kingdom
2.	University of Cambridge	United Kingdom
3.	Harvard University	United States
4.	Stanford University	United States
5.	Massachusetts Institute of Technology (MIT	United States
6.	University of California, 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

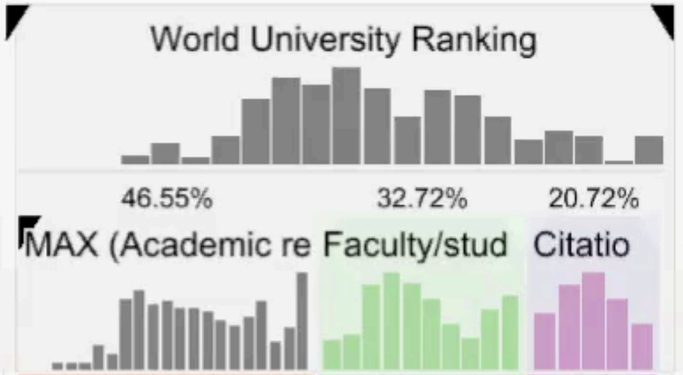
Bump Charts


Rank
University
Score
ore
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.
2.
3.
4.
5.
(+1)
(+1)
(-2)
Score

453 visible of 453 (100.00%)



Rank	School Name
	Filter: <None>
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
http://taggle.caleydoapp.org

Pixel Based Displays

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

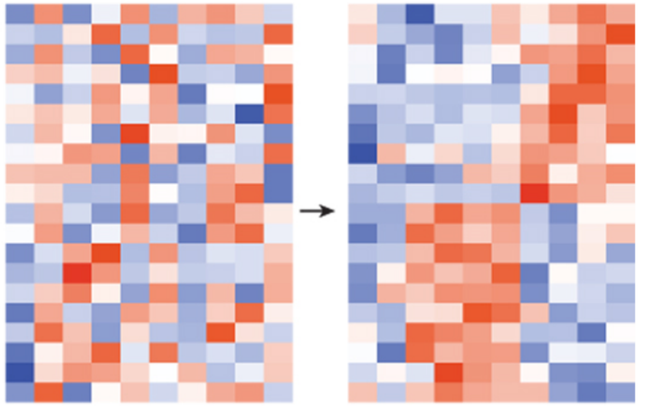
Ordering critical for interpretation

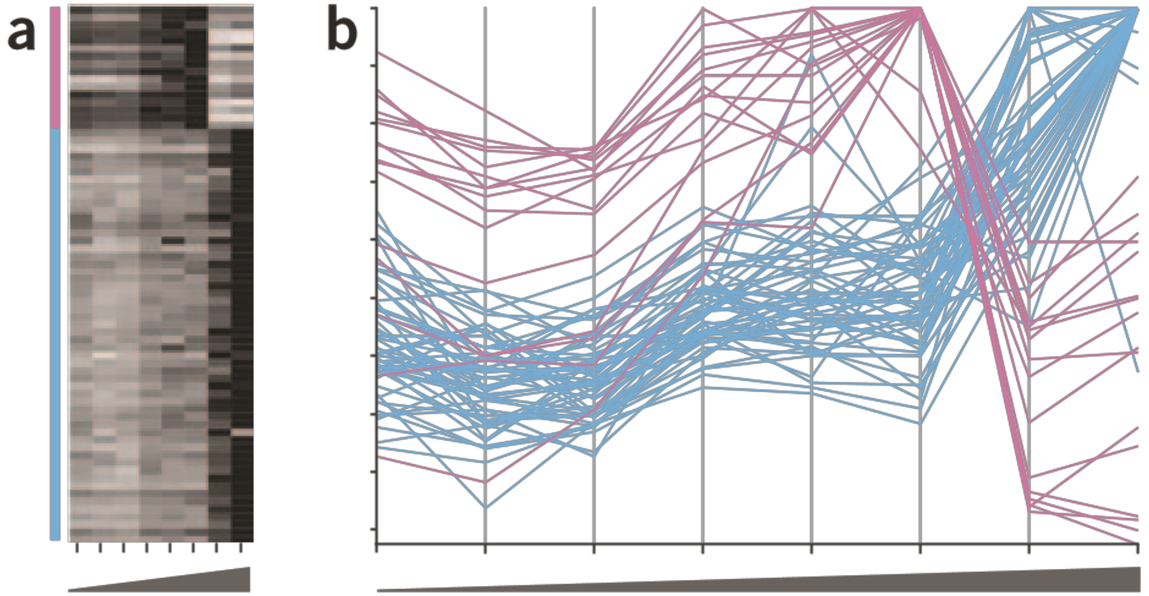
If no ordering inherent, clustering is used

Scalable – 1 px per item

Good for homogeneous data

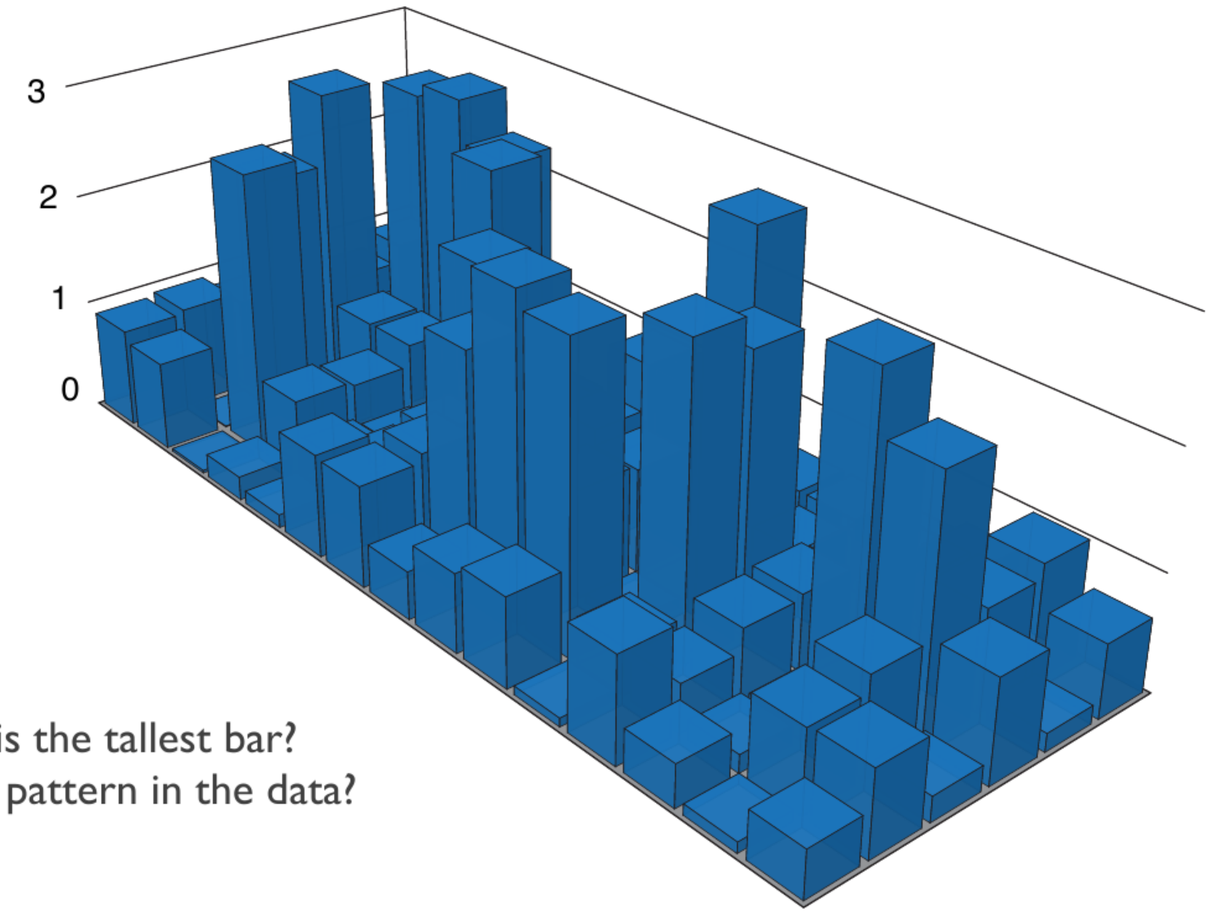
same scale & type





[Gehlenborg & Wong 2012]

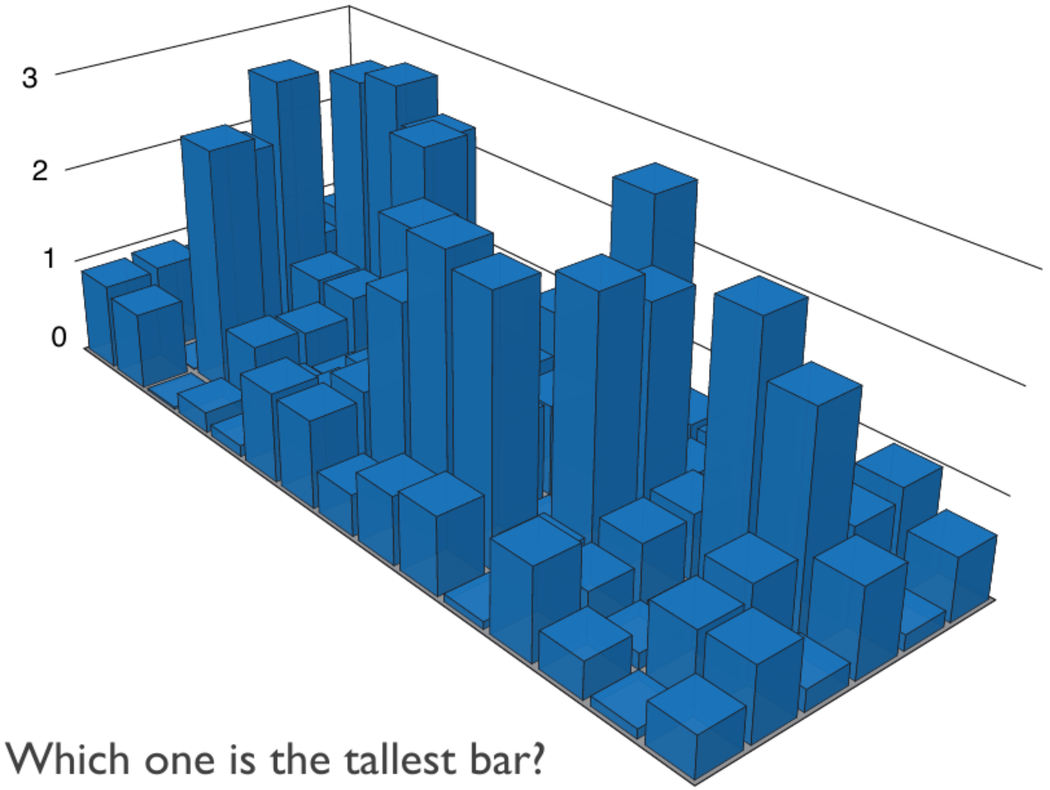
3D Pitfall: Occlusion & Perspective



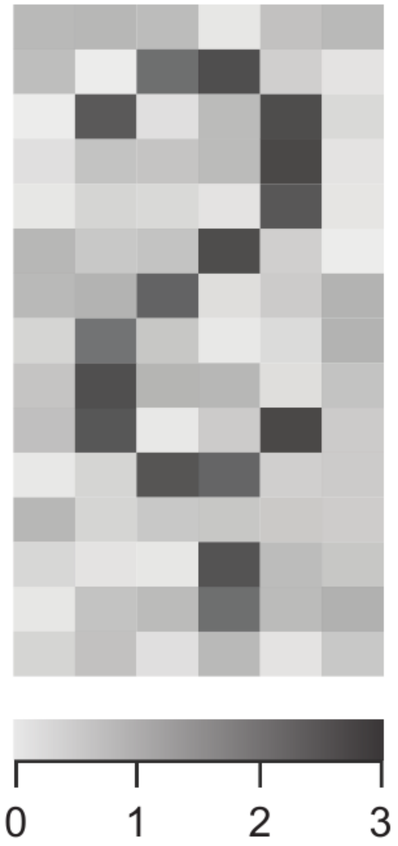
Which one is the tallest bar?
What is the pattern in the data?

[Gehlenborg and Wong, Nature Methods, 2012]

3D Pitfall: Occlusion & Perspective

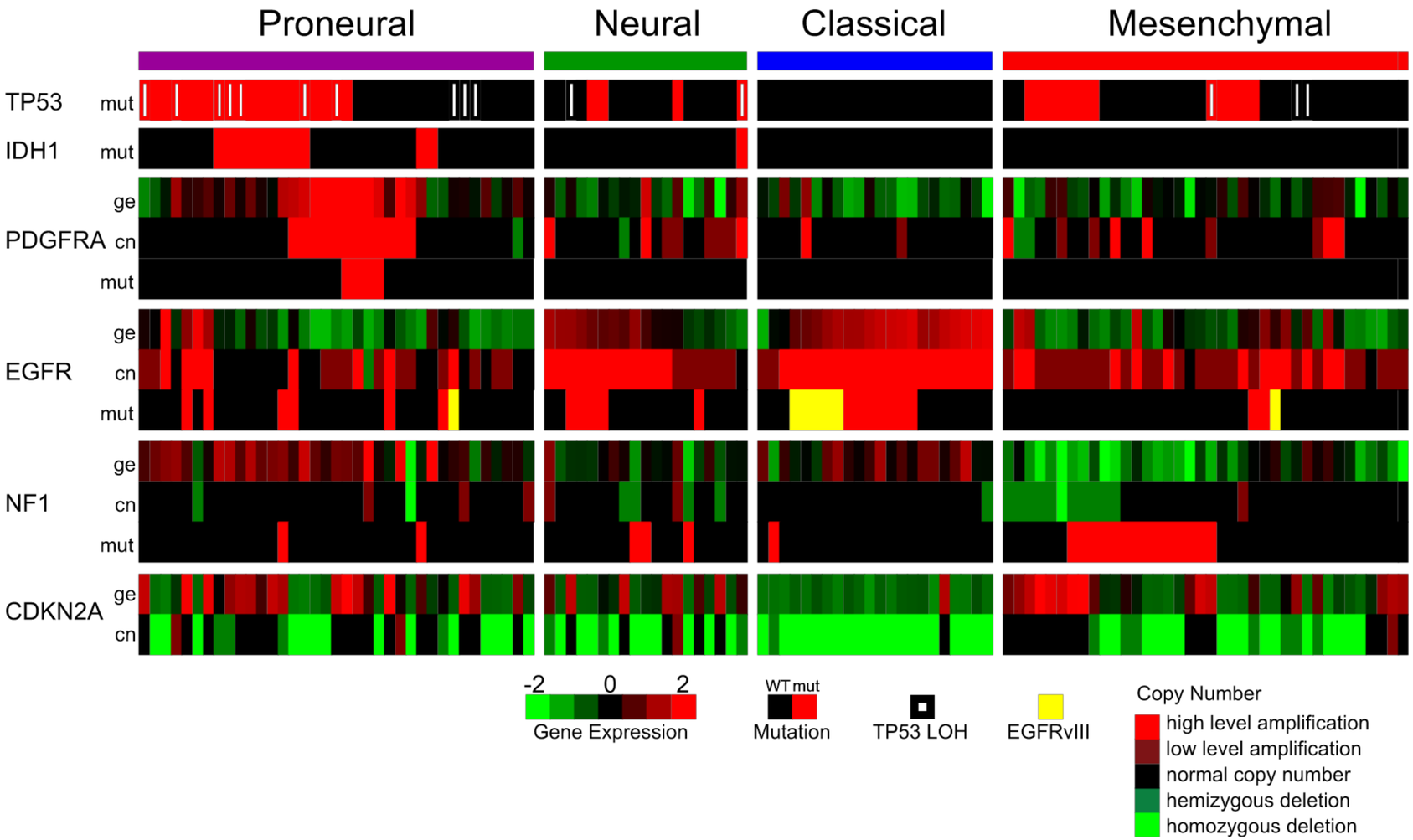


Which one is the tallest bar?
What is the pattern in the data?



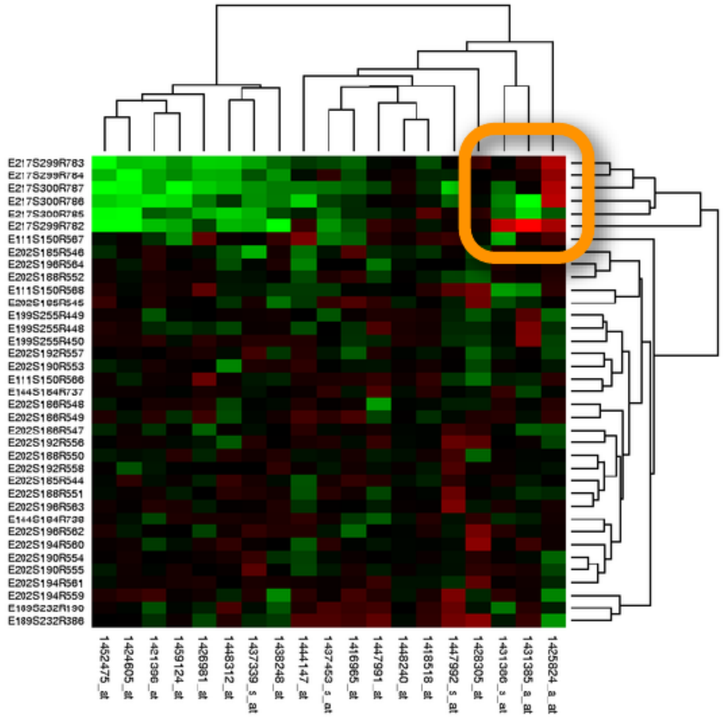
[Gehlenborg and Wong, Nature Methods, 2012]

Heterogeneous Data?

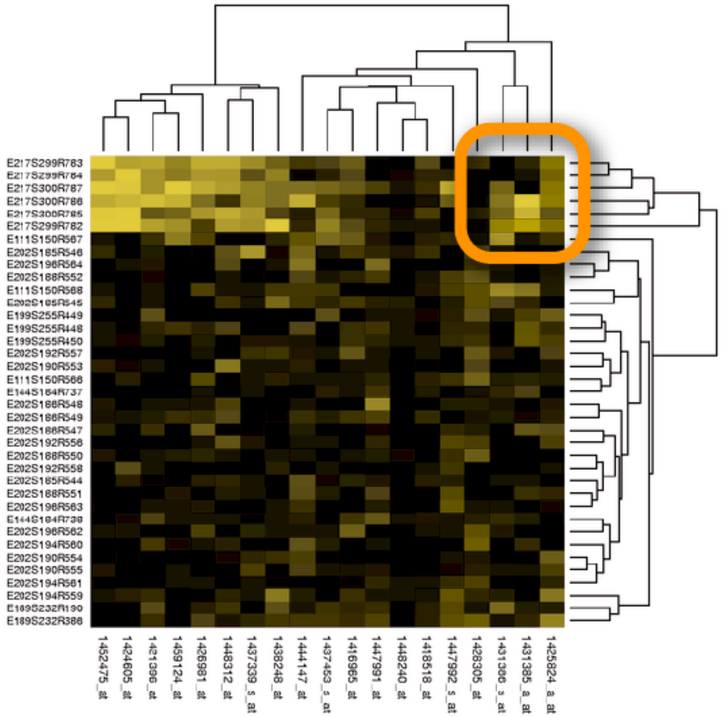


[Verhaak 2012]

Bad Color Mapping

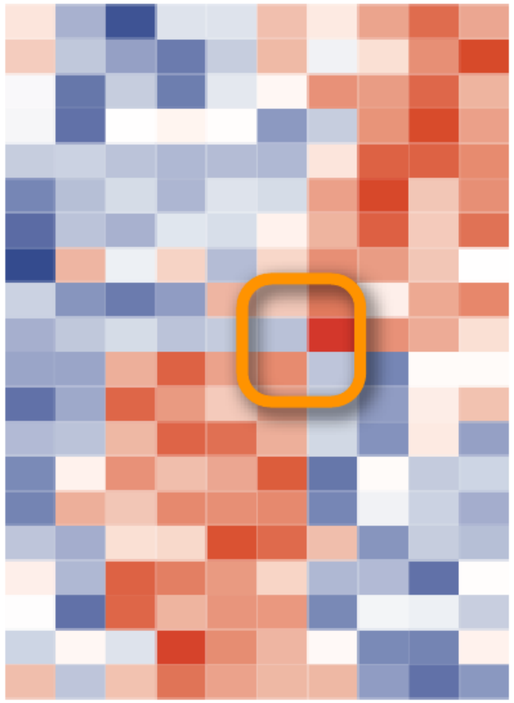


Normal Vision

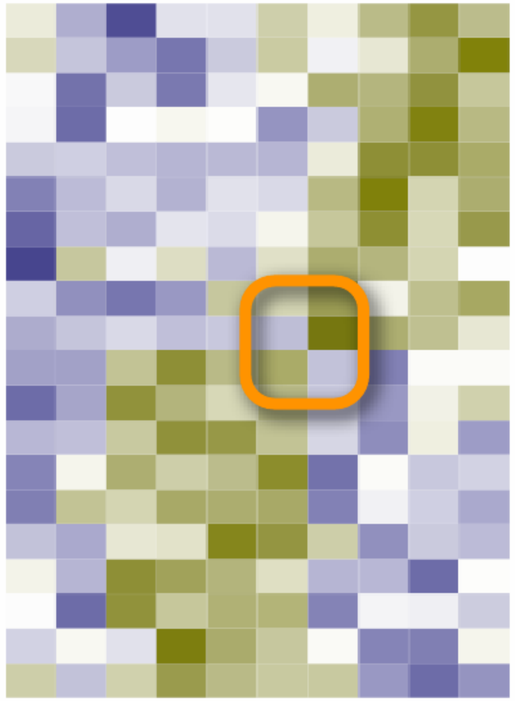


Deuteranope Vision
("Red-Green Blindness")

Good Color Mapping

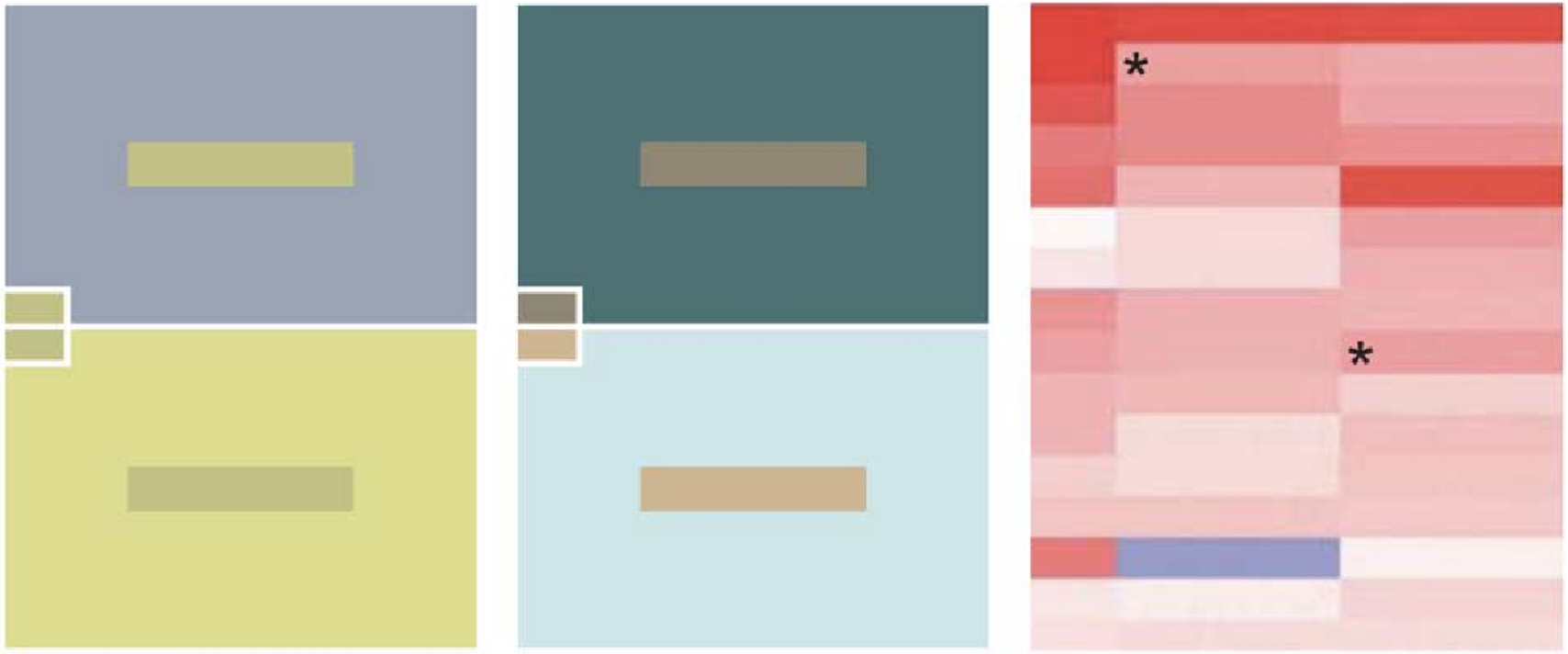


Normal Vision

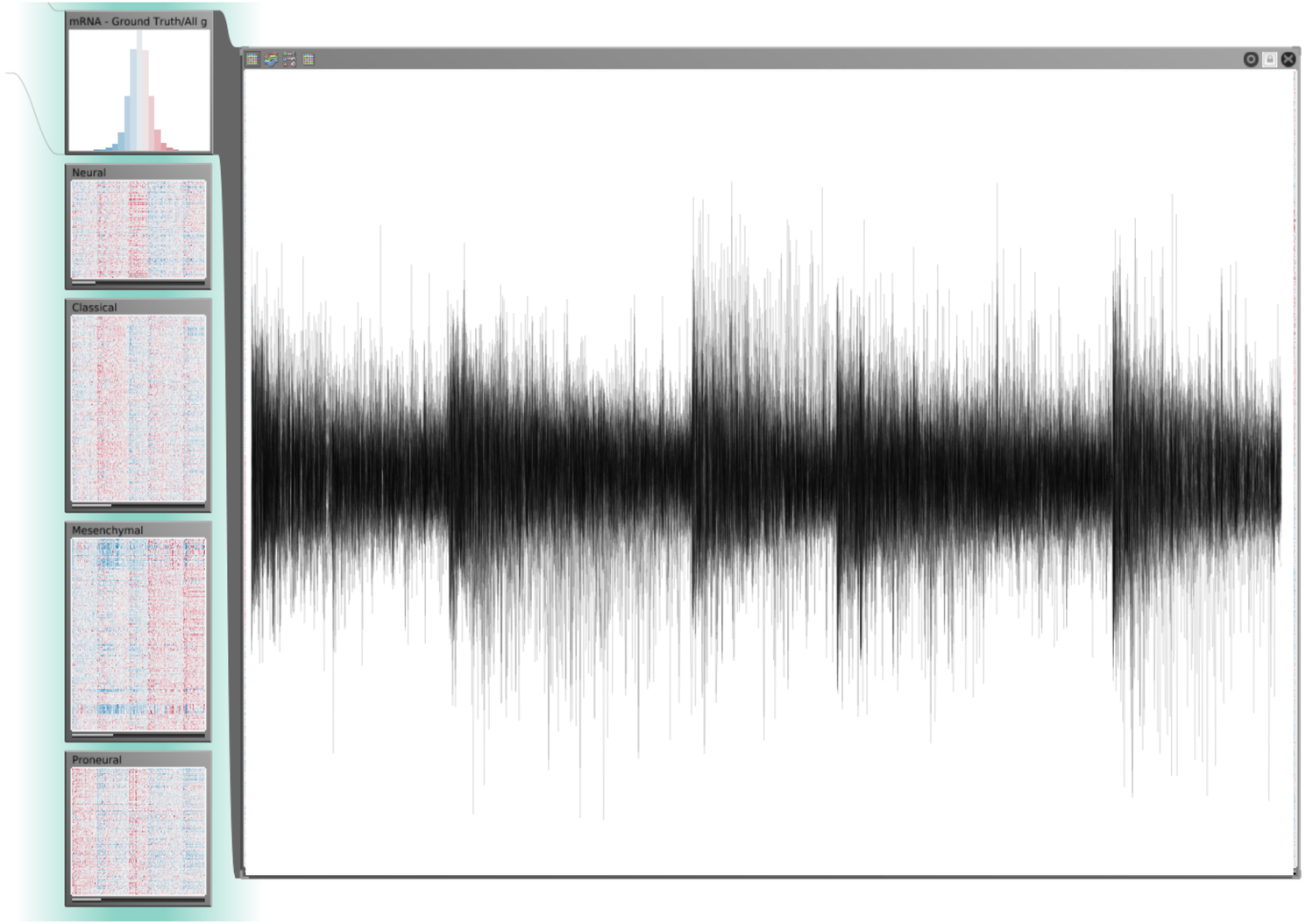


Deuteranope Vision
("Red-Green Blindness")

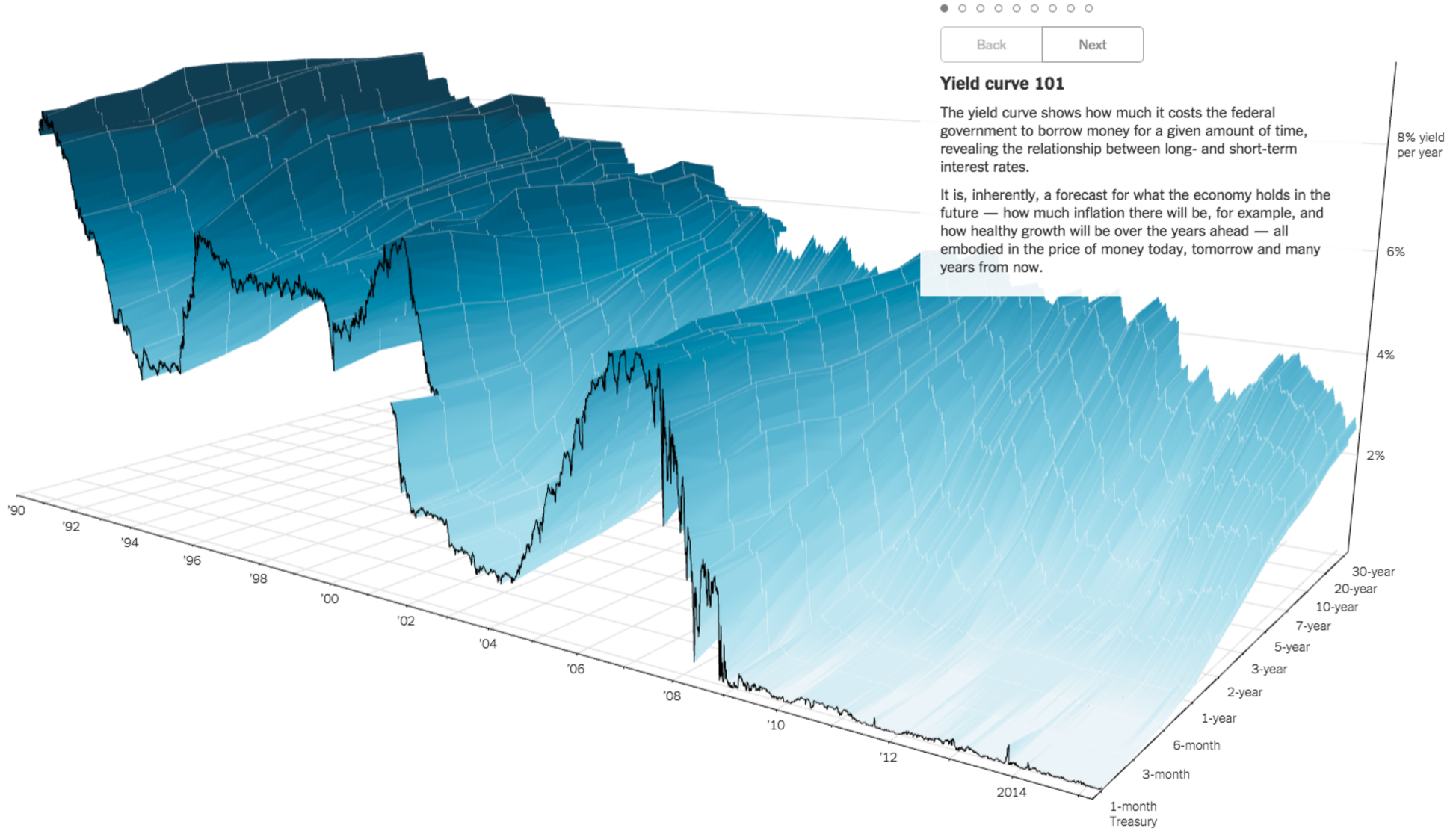
Color is relative!



Clustered Heat Map



Design Critique



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

Context / Critiques

https://vimeo.com/127205447

https://community.jmp.com/t5/JMP-Blog/Graph-makeover-3-D-yield-curve-surface/ba-p/30573

http://www.visualisingdata.com/2015/03/when-3d-works/

Spatial Axis Orientation

Arrange Tables

- **Express Values**

- **Separate, Order, Align Regions**
  - Separate
  - Order
  - Align

  - *1 Key* — *List*
  - *2 Keys* — *Matrix*
  - *3 Keys* — *Volume*
  - *Many Keys* — *Recursive Subdivision*

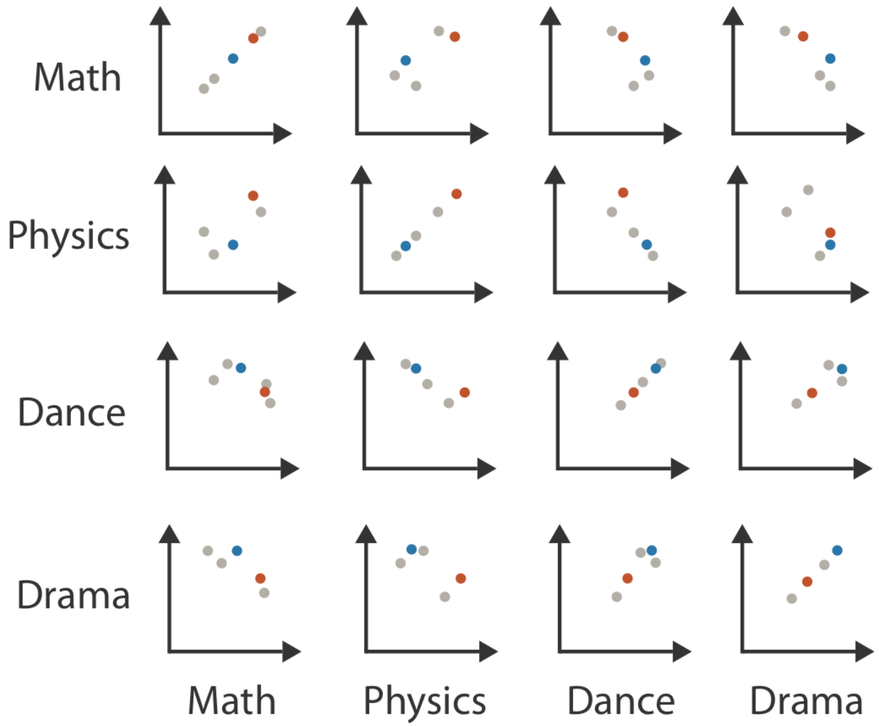
- **Axis Orientation**
  - Rectilinear
  - Parallel
  - Radial

- **Layout Density**
  - Dense
  - Space-Filling

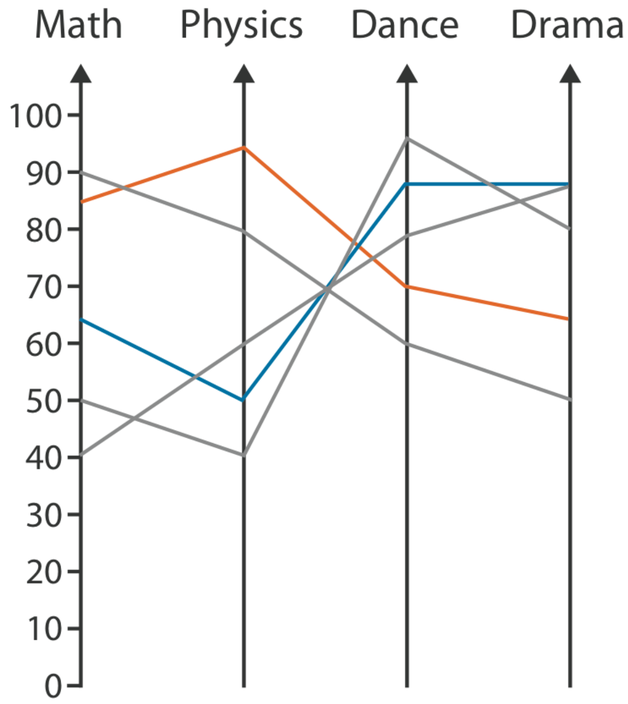
Table

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

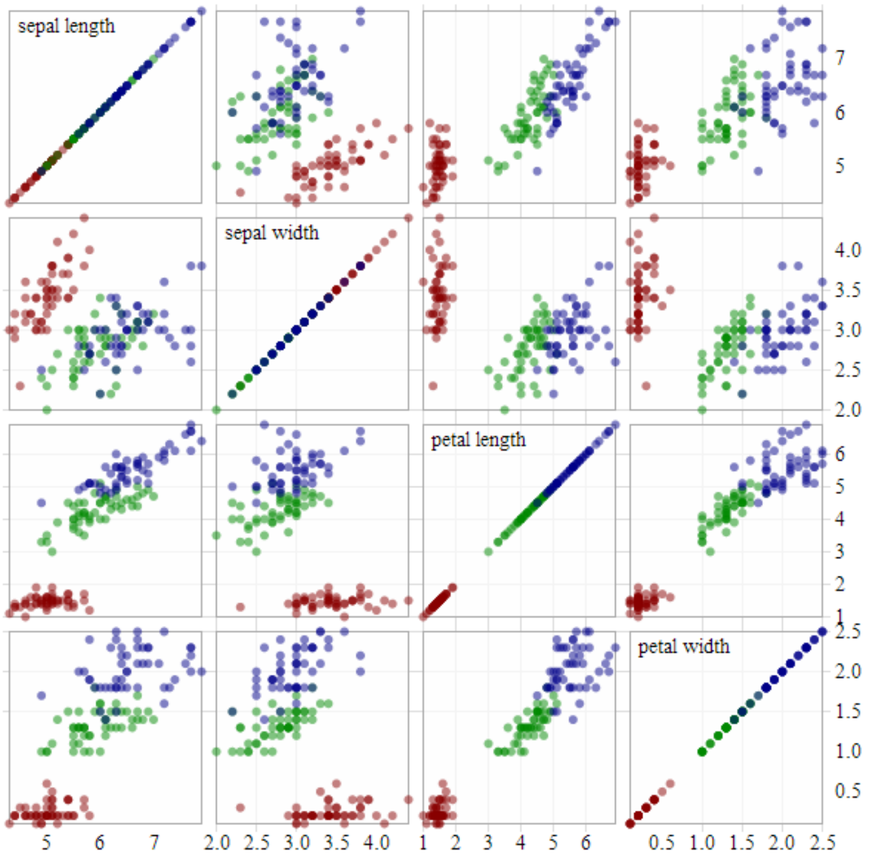
Scatterplot Matrix

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)

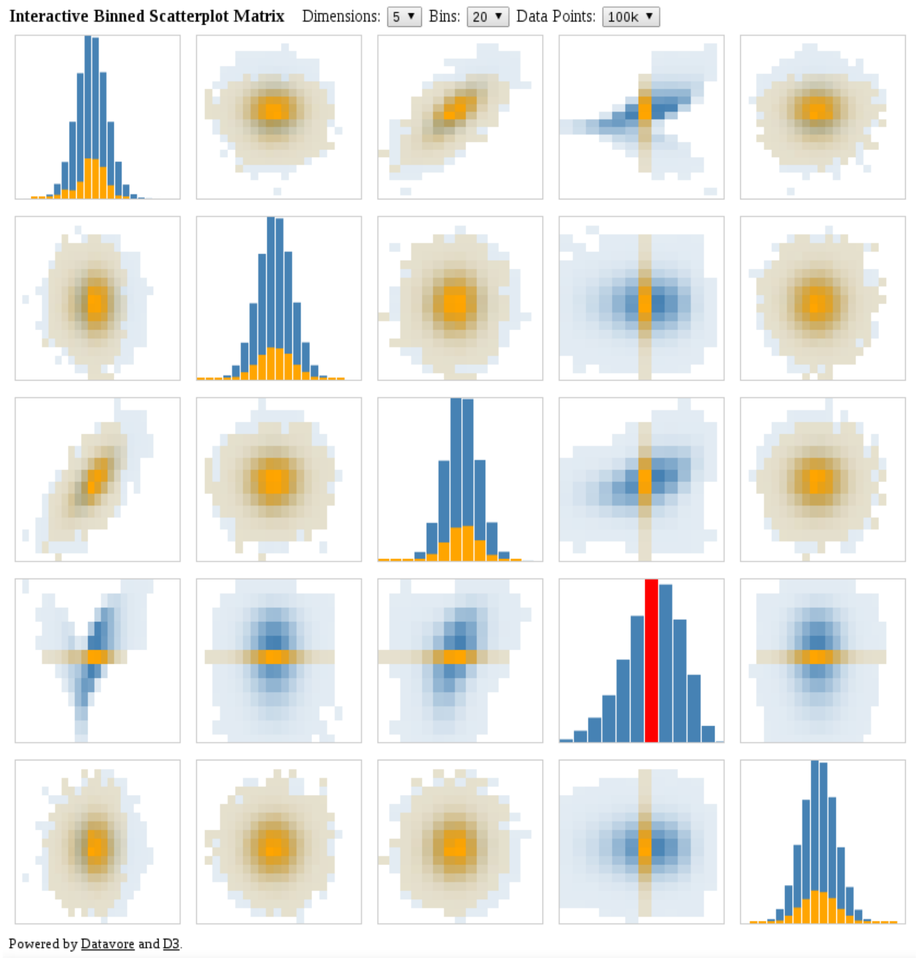
Brushing is important

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

Algorithmic approaches:

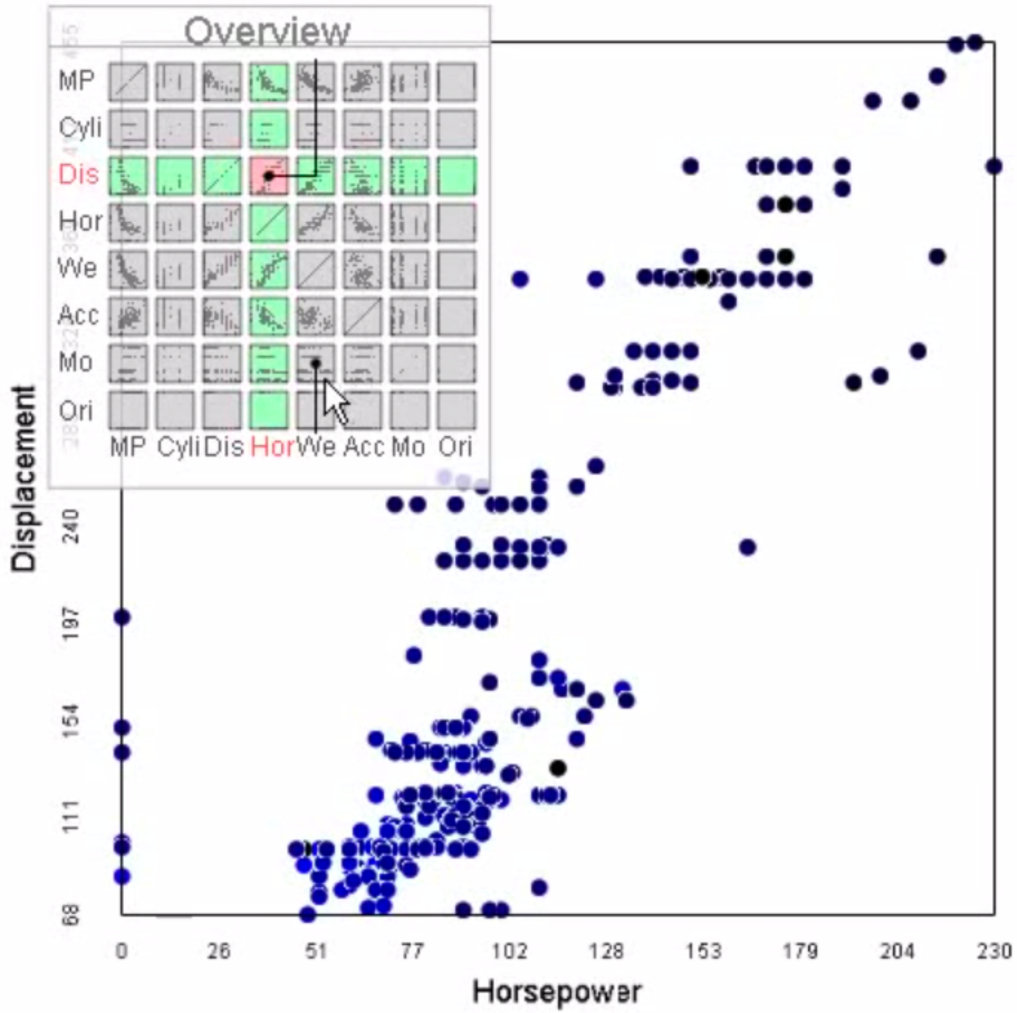
- Clustering & aggregating records
- Choosing dimensions
- Choosing order

SPLOM Aggregation - Heat Map



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

SPLOM F+C, Navigation



[Elmqvist]

Spatial Axis Orientation

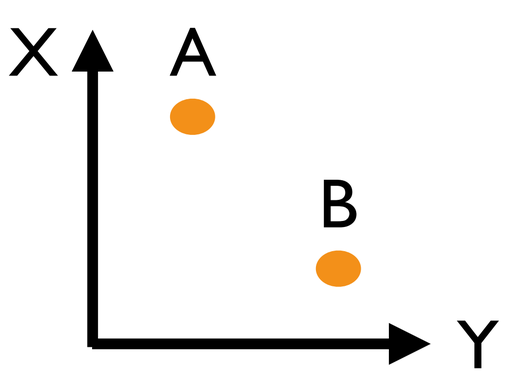
Parallel Coordinates

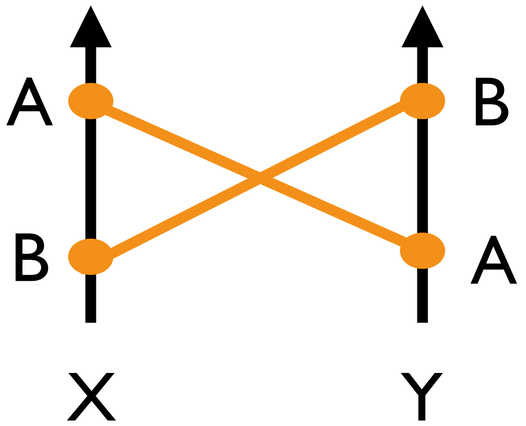
Parallel Coordinates (PC)

Inselberg 1985

Axes represent attributes

Lines connecting axes represent items





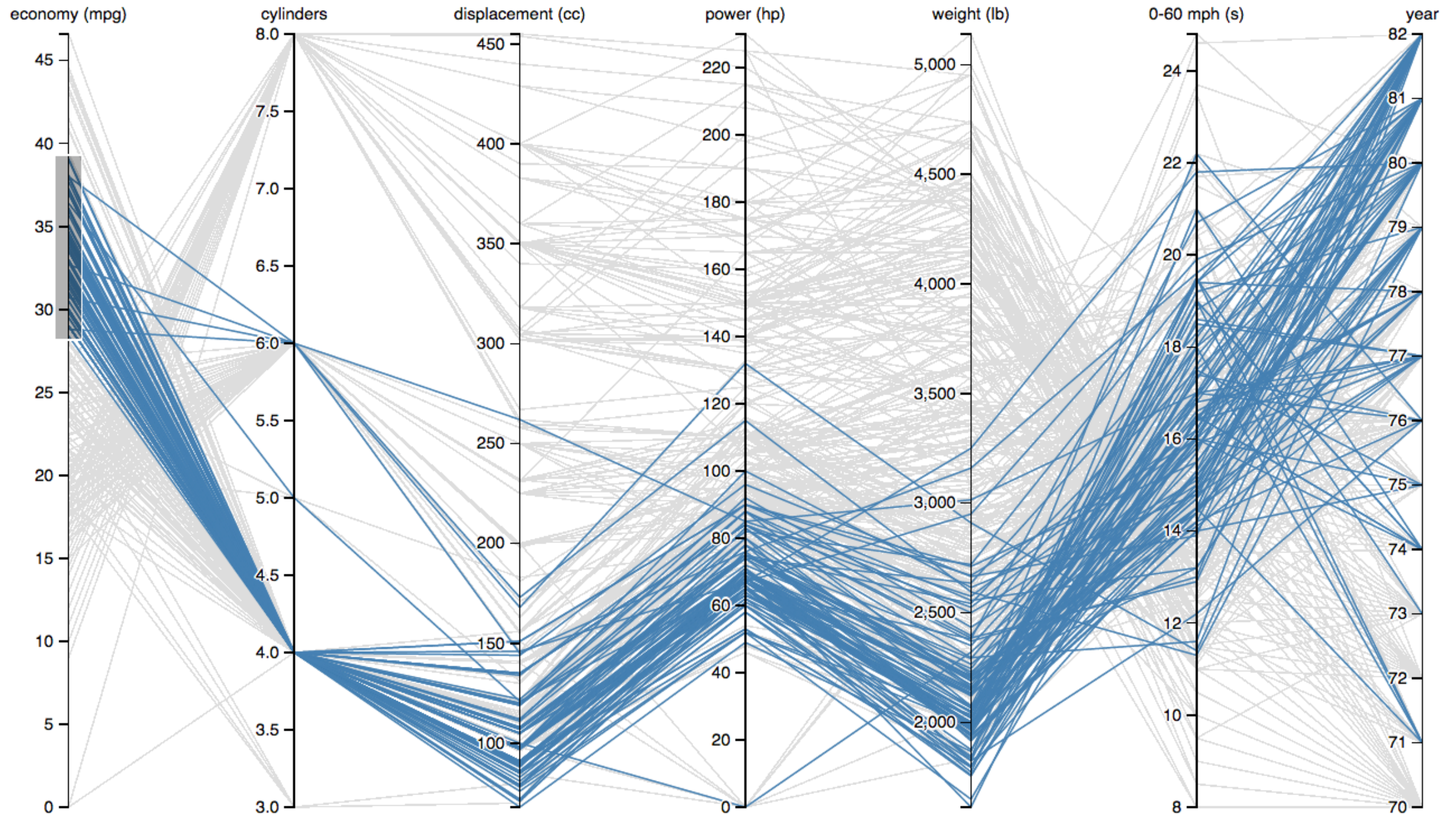
Parallel Coordinates

Each axis represents dimension

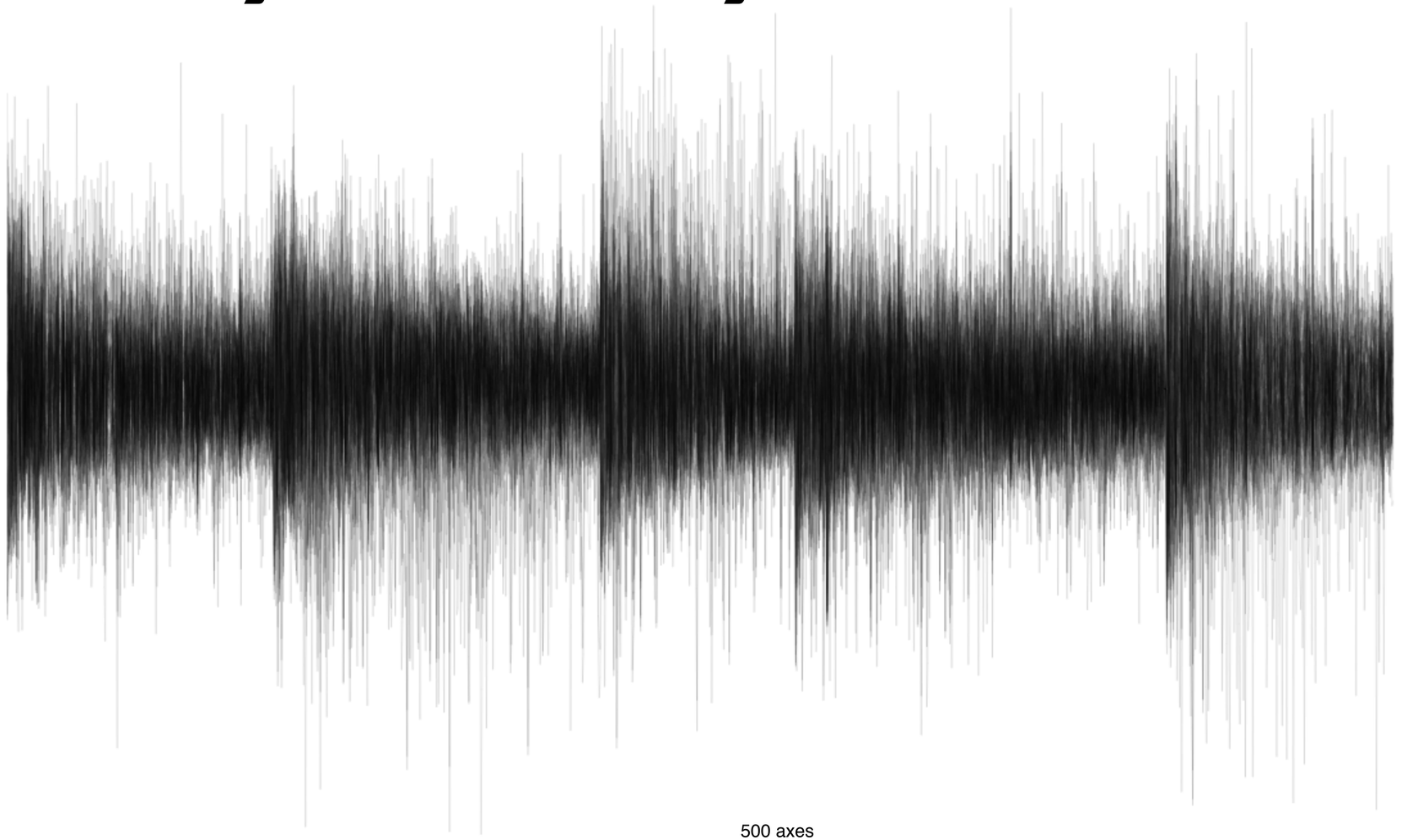
Lines connecting axis represent records

Suitable for

- all tabular data types
- heterogeneous data



PC Limitation:
Scalability to Many Dimensions

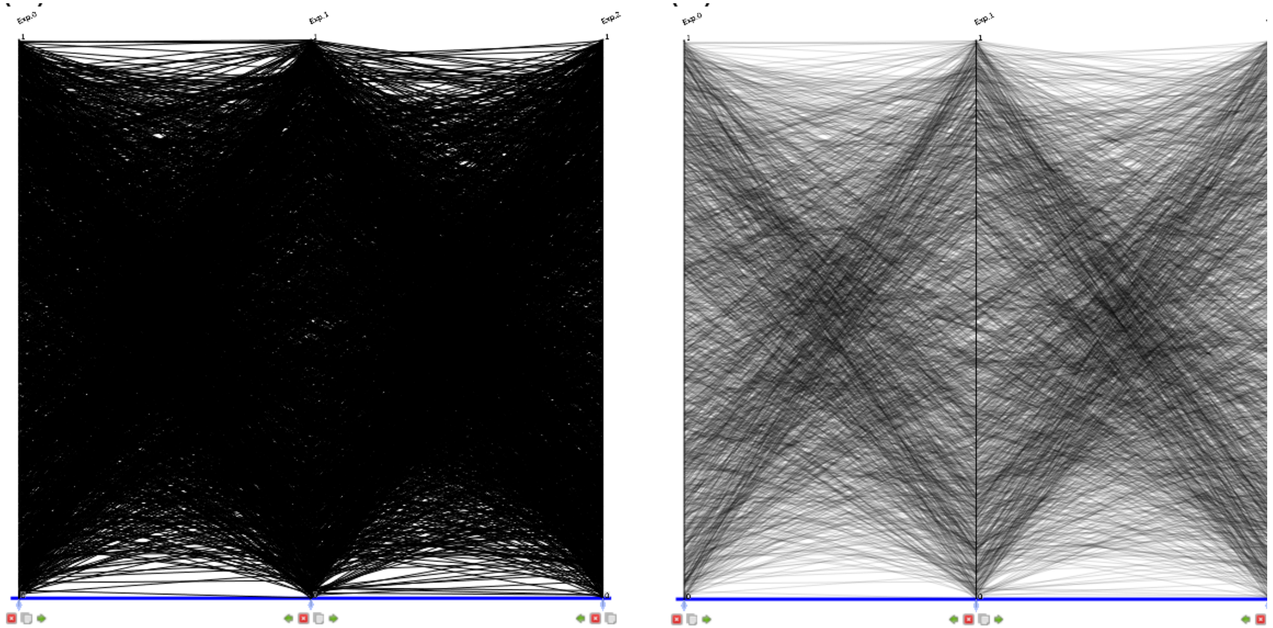


500 axes

PC Limitation: Scalability to Many Items

Solutions:

- Transparency
- Bundling, Clustering
- Sampling

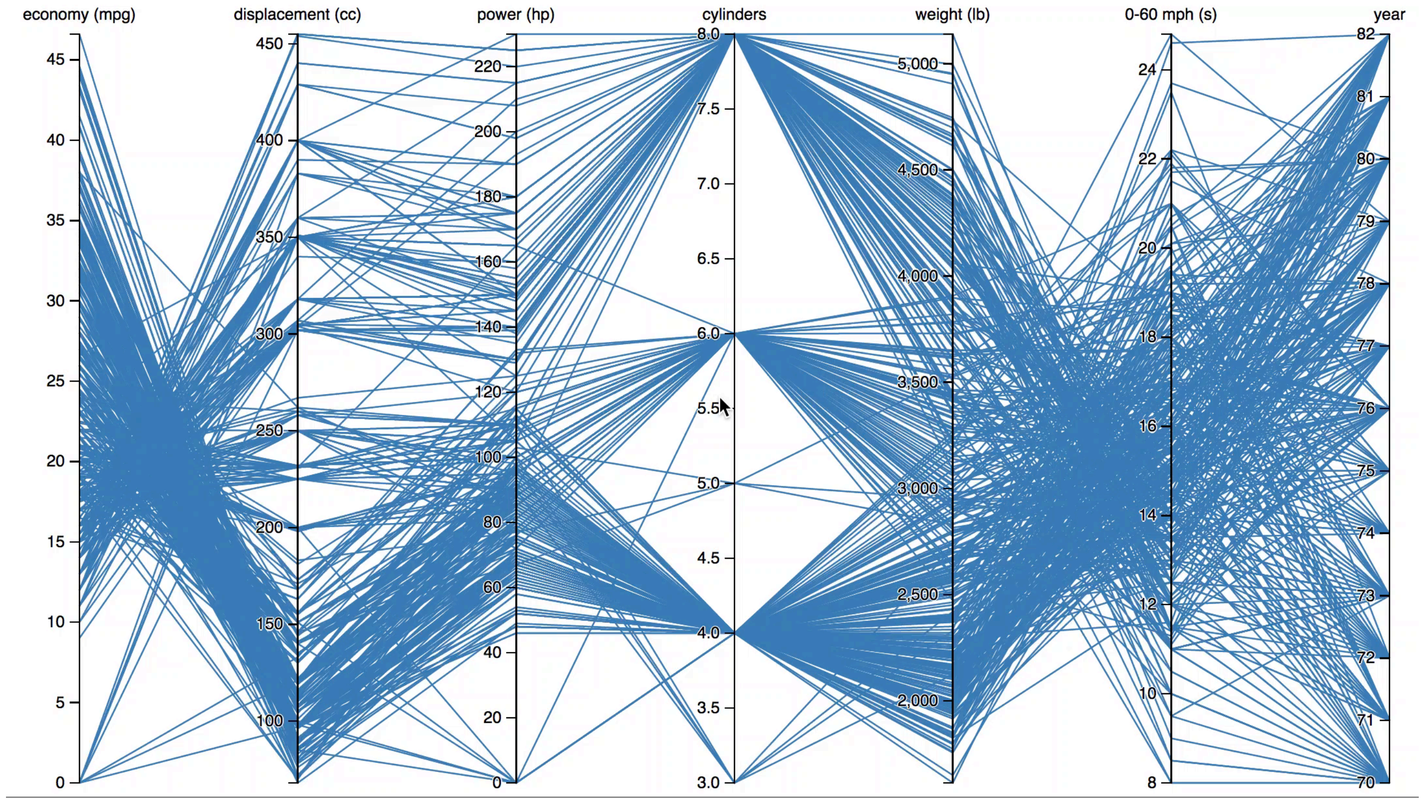


PC Limitations

Correlations only between adjacent axes

Solution: Interaction

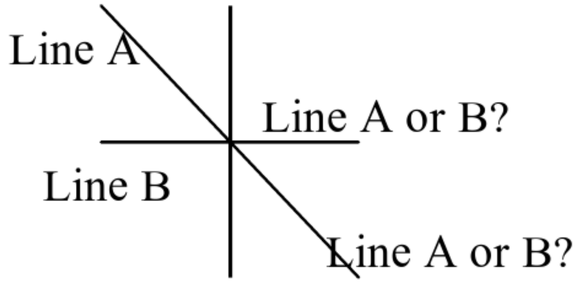
- Brushing
- Let user change order

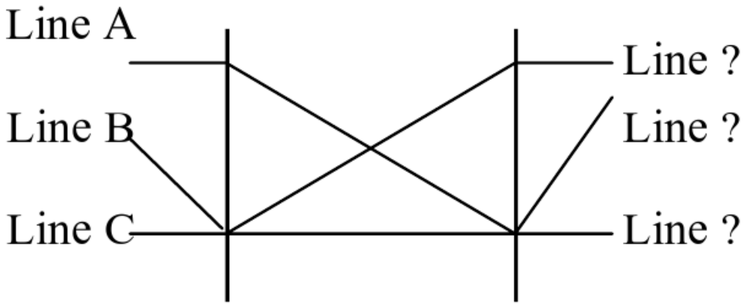


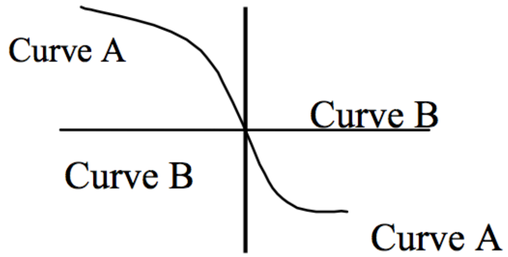
PC Limitation: Ambiguity

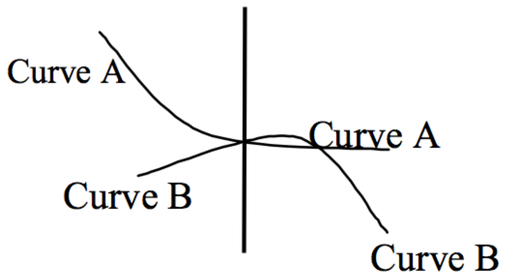
Solutions:

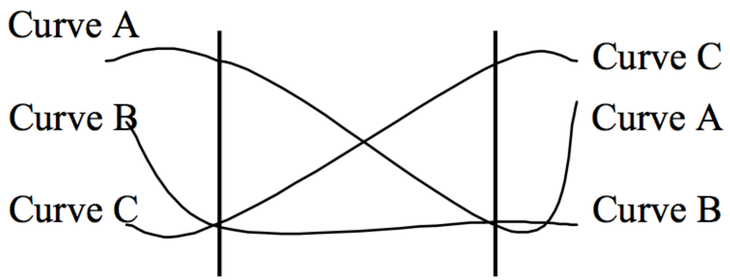
- Brushing
- Curves











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:

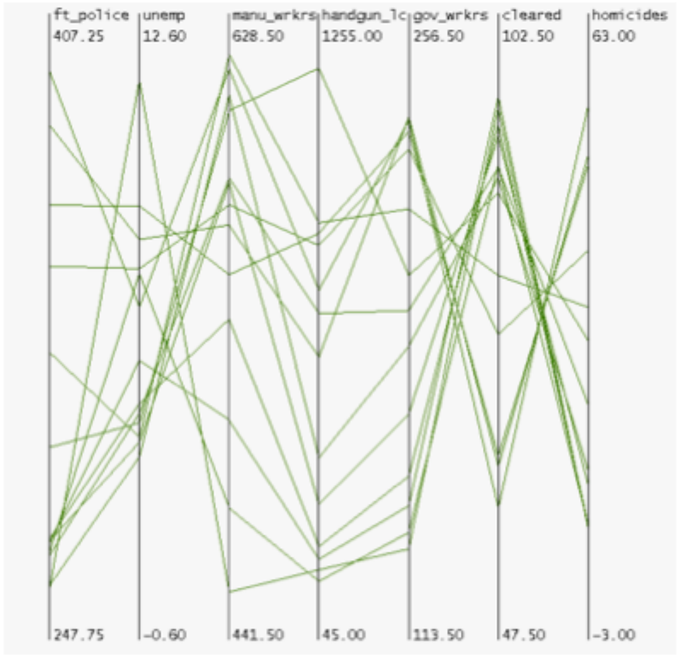
- Choosing dimensions
- Choosing order
- Clustering & aggregating records

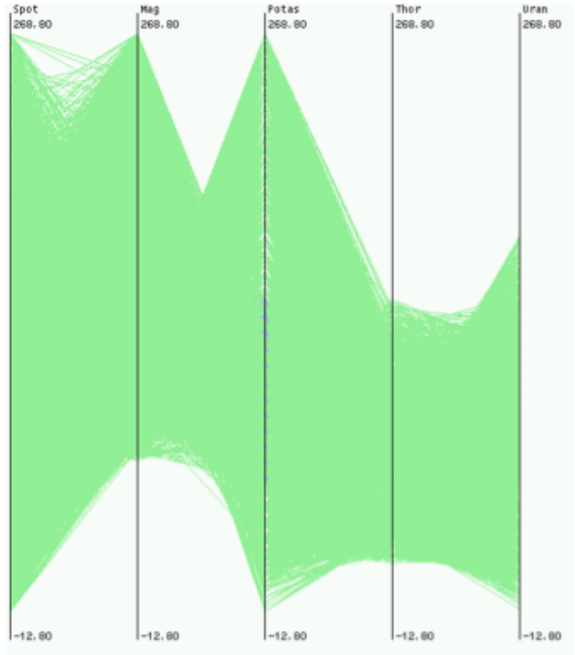
http://bl.ocks.org/jasondavies/1341281

HIERARCHICAL PARALLEL COORDINATES

goal: scale up parallel coordinates to large datasets

challenge: overplotting/occlusion



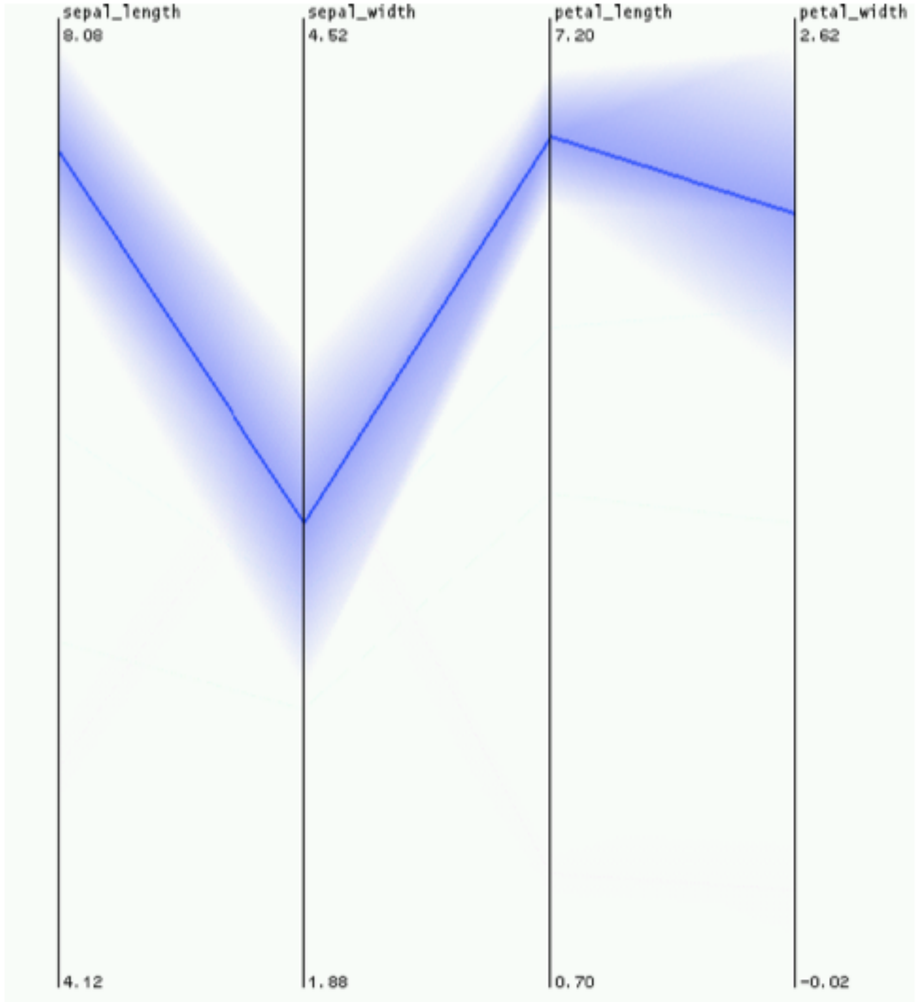


Fua 1999

HPC: ENCODING DERIVED DATA

visual representation: variable-width opacity bands

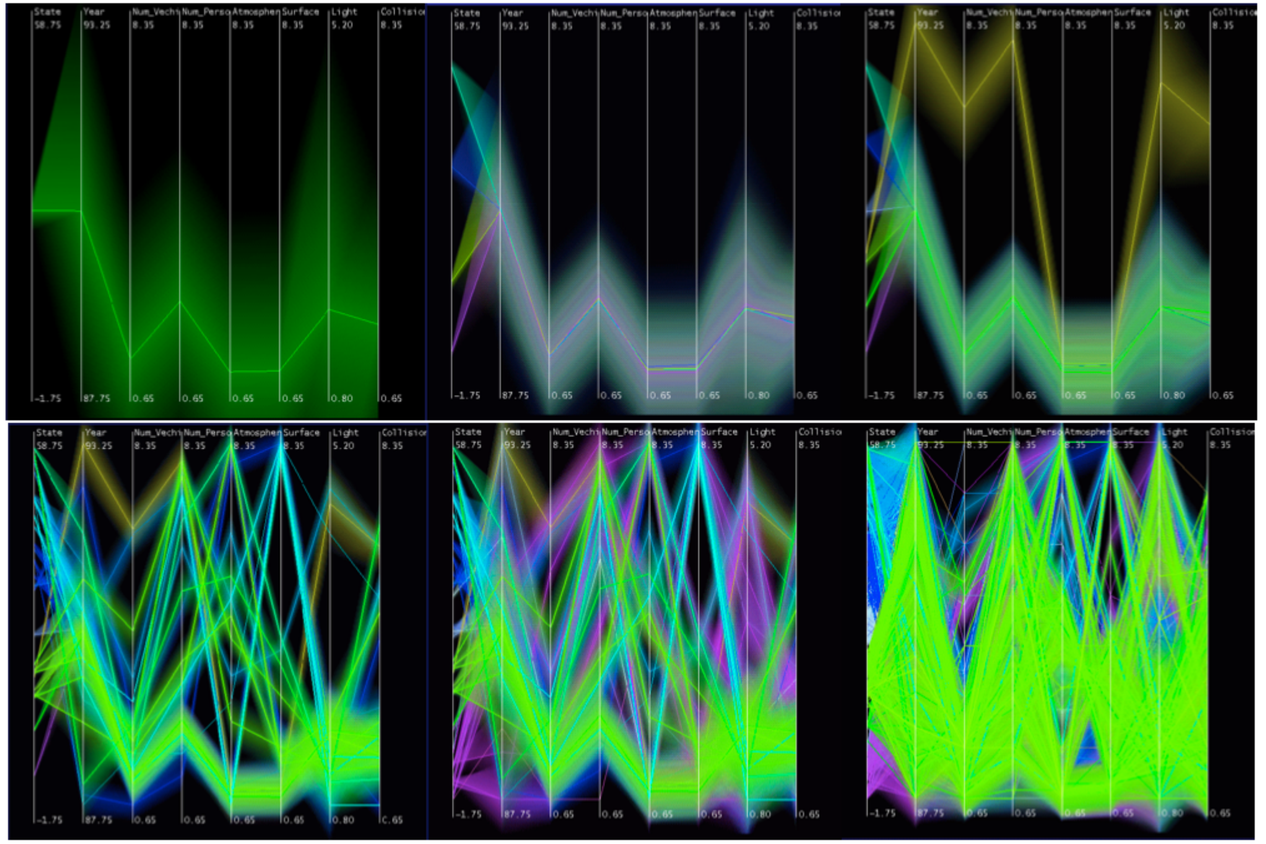
- show whole cluster, not just single item
- min / max: spatial position
- cluster density: transparency
- mean: opaque



Fua 1999

HPC: INTERACTING WITH DERIVED DATA

interactively change level of detail to navigate cluster hierarchy



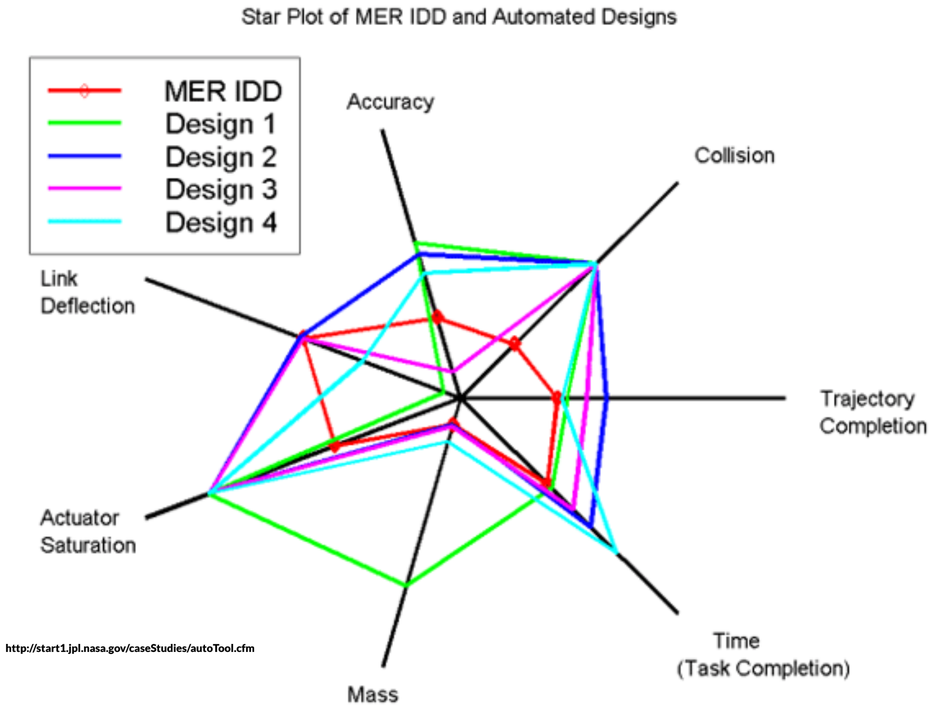
Fua 1999

Star Plot

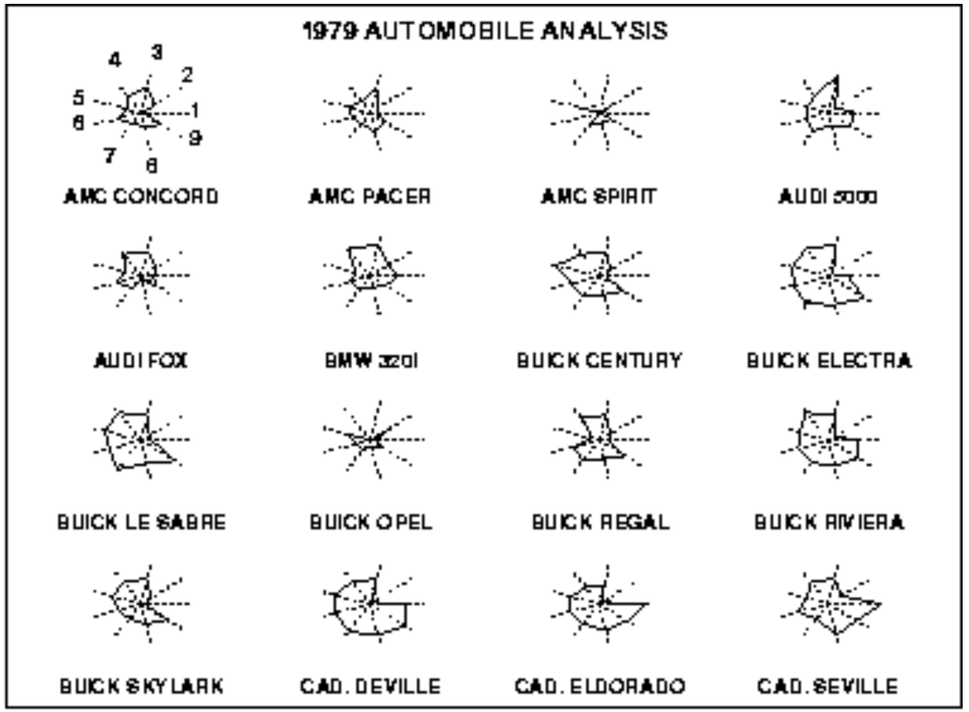
[Coekin1969]

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

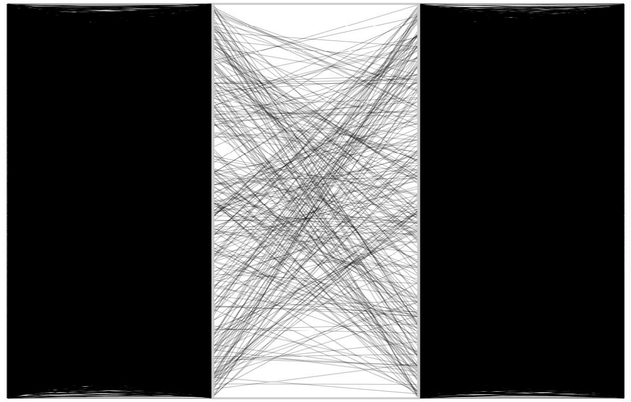
Outlier-preserving approaches

Filtering

Define criteria to remove data, e.g.,

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

Can be interactive, combined with sampling

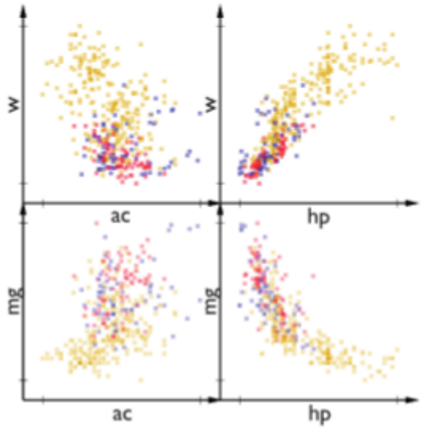


[Ellis & Dix, 2006]

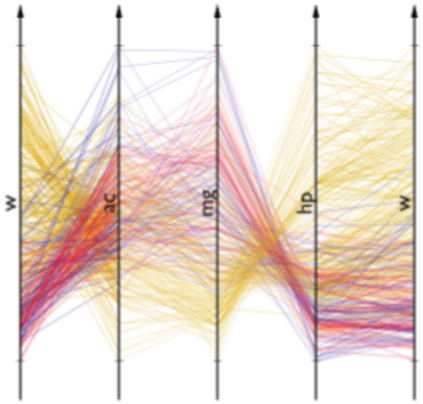
Spatial Axis Orientation

Hybrids

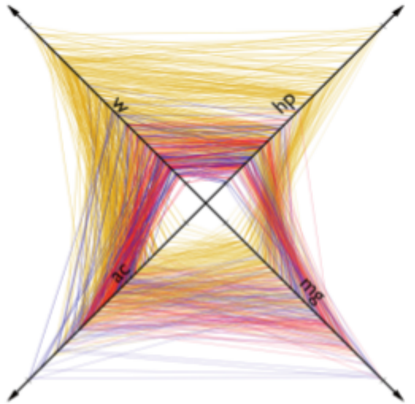
Flexible Linked Axes (FLINA)



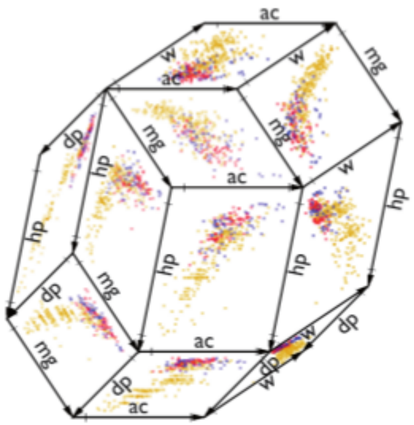
(a) scatterplots



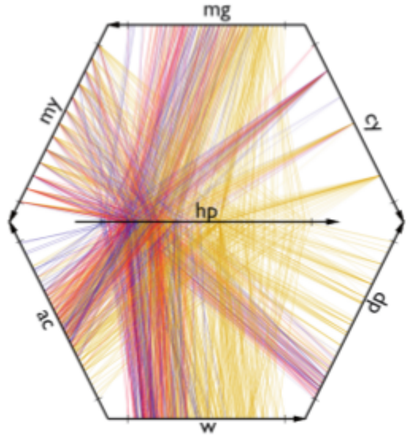
(b) Parallel Coordinates Plot



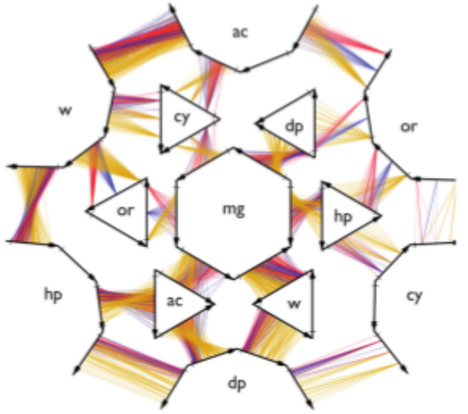
(c) radar chart



(d) Hyperbox



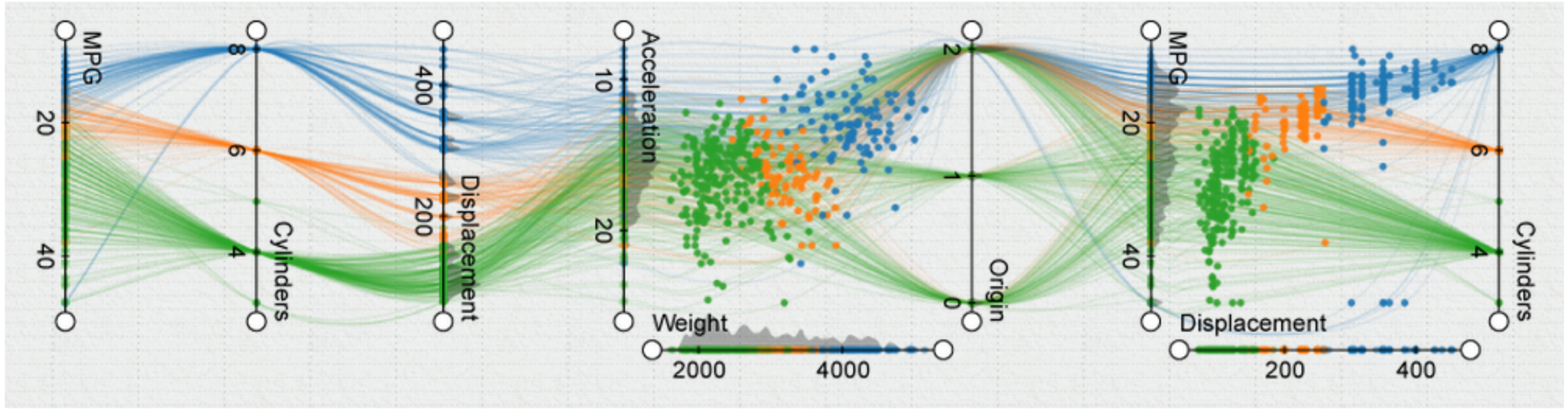
(e) Time Wheel



(f) Many-to-many PCP

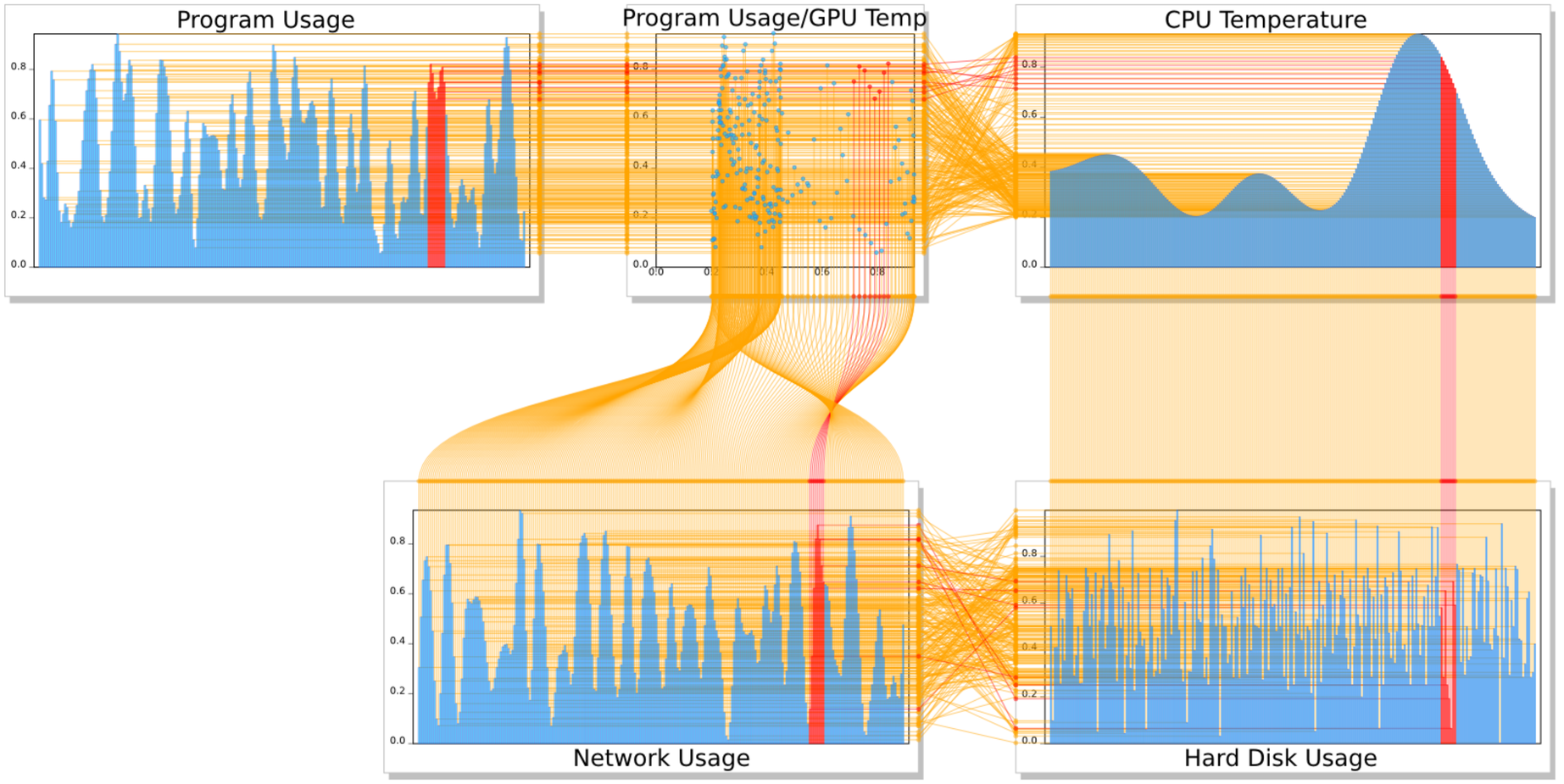
Claessen & van Wijk 2011

Web-based implementation of FLINA concept



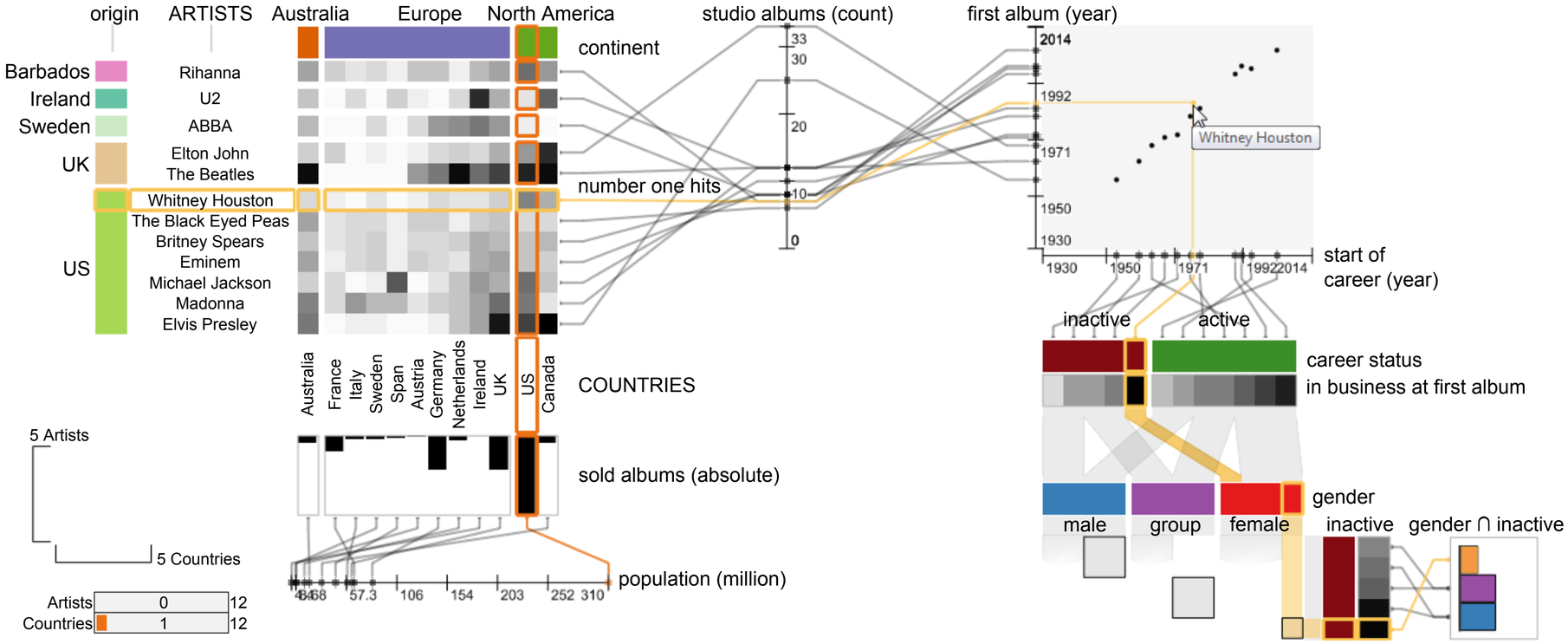
http://vis.pku.edu.cn/mddv/val/

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

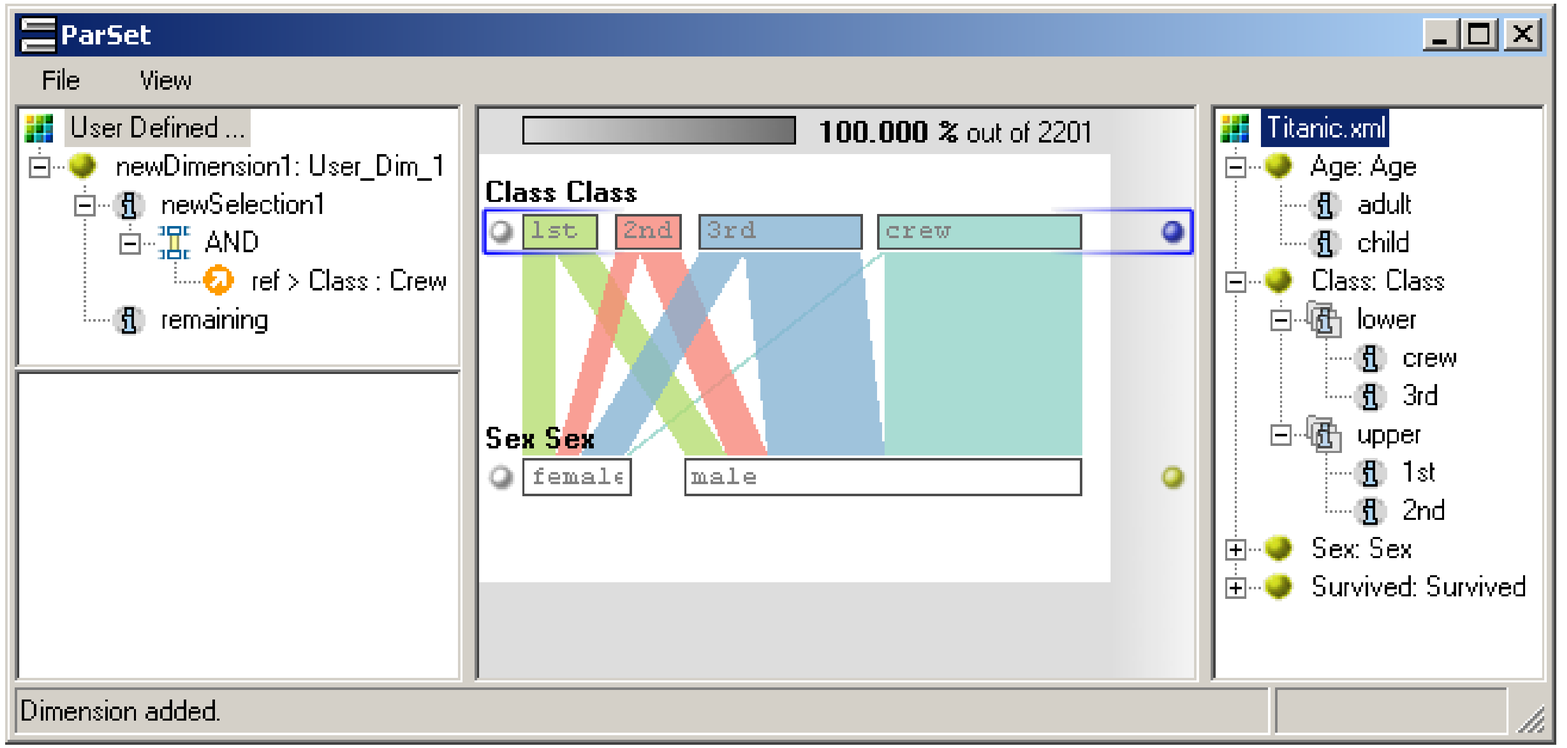
task: find relationship between attributes

interaction driven technique

Visual Encoding

boxes scaled by frequency

color coded by values for current active dimension

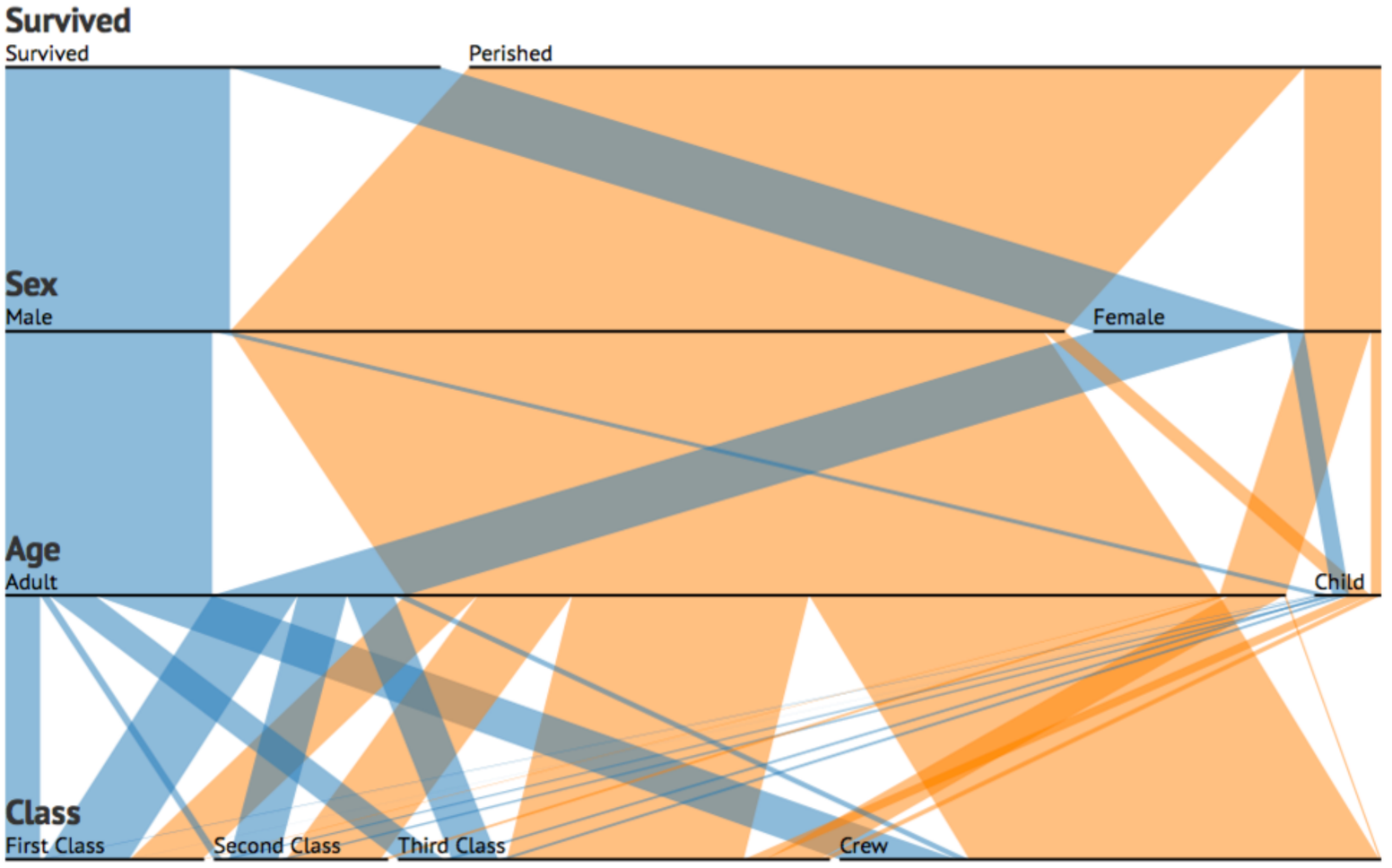


Bendix, Kosara, Hauser, 2005

Parallel Sets

A visualisation technique for multidimensional categorical data.

Titanic Survivors

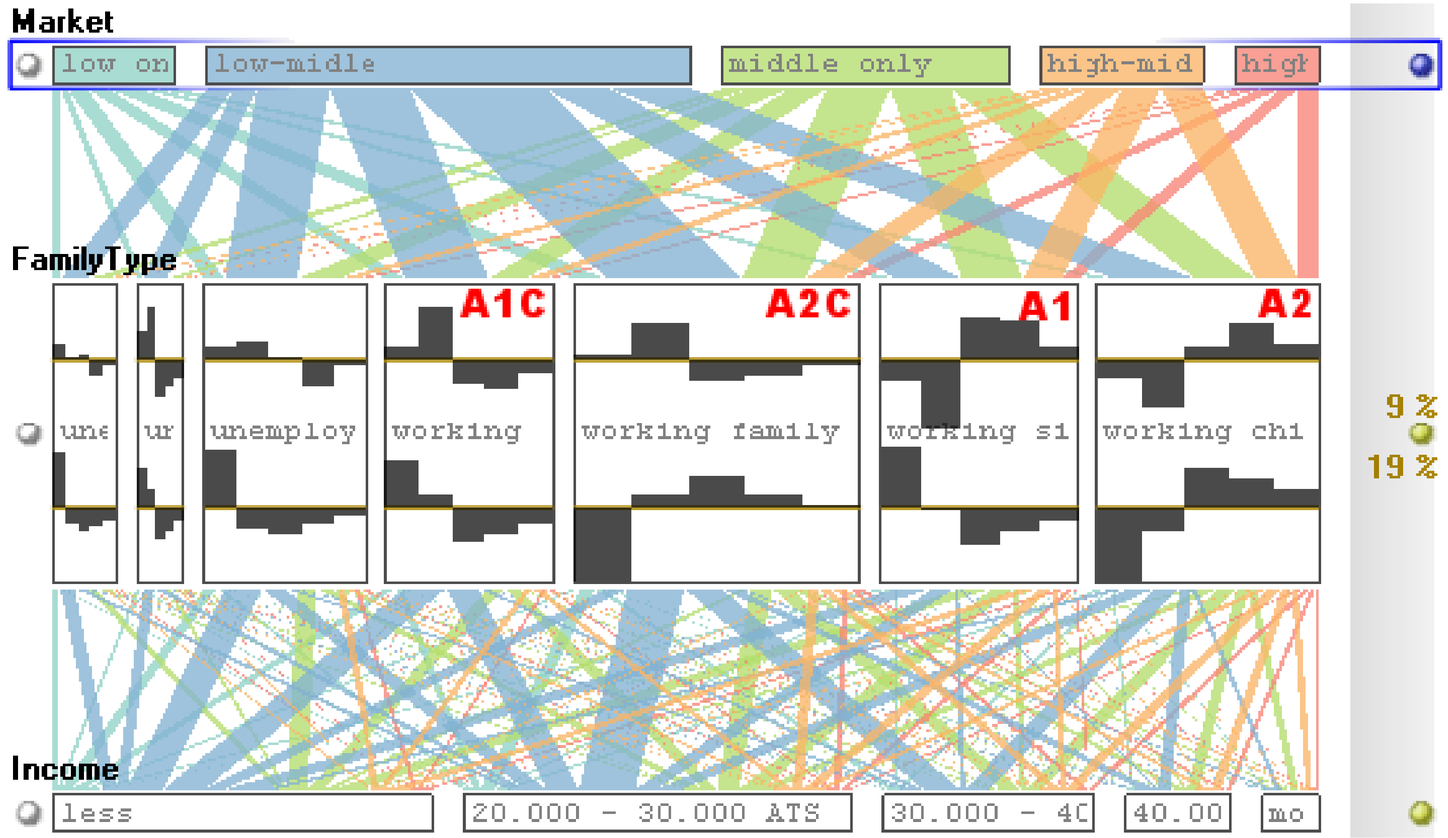


Curves?

Data: Robert J. MacG. Dawson.

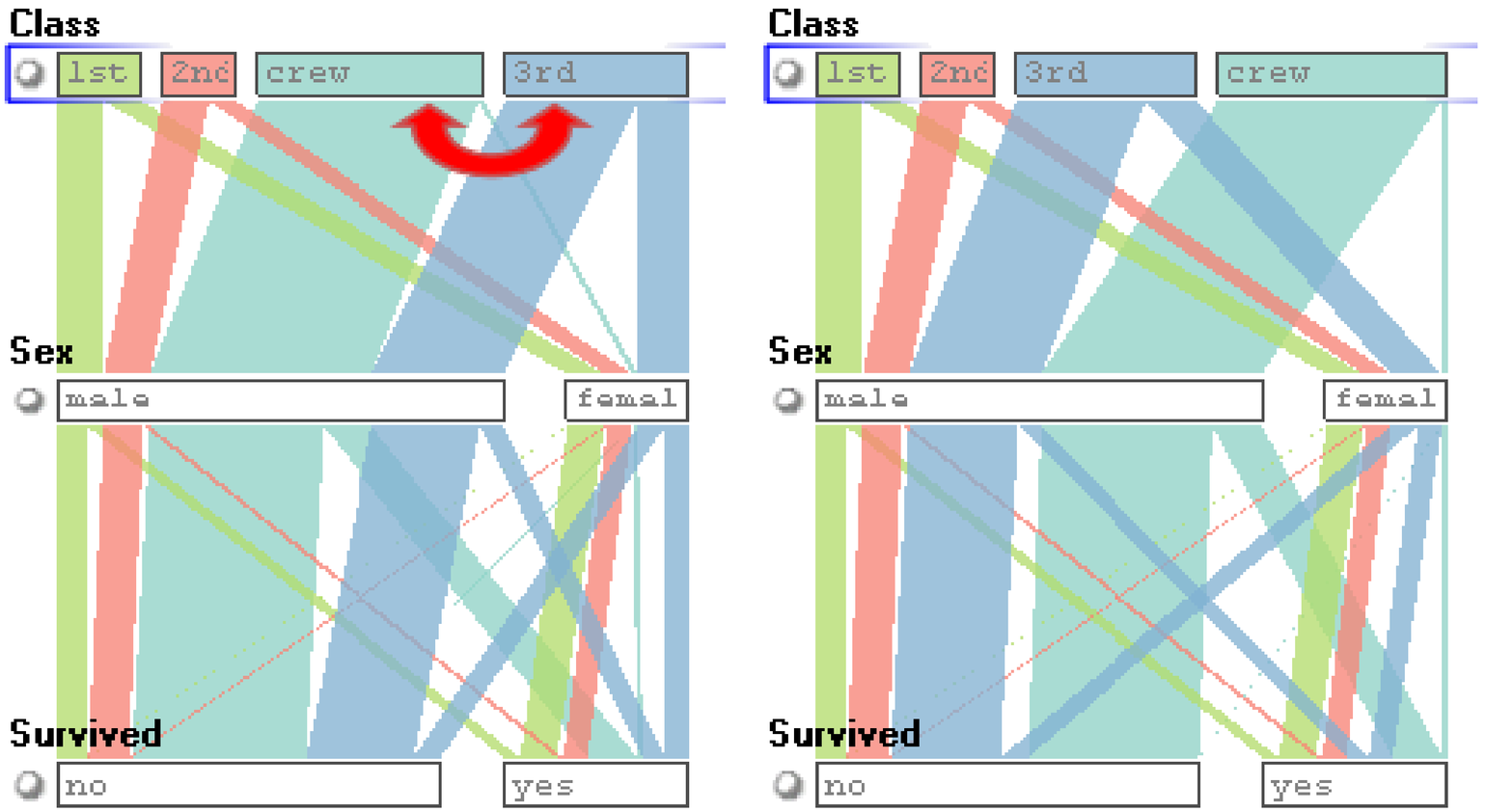
Visual Encoding

Boxes expand to show histogram



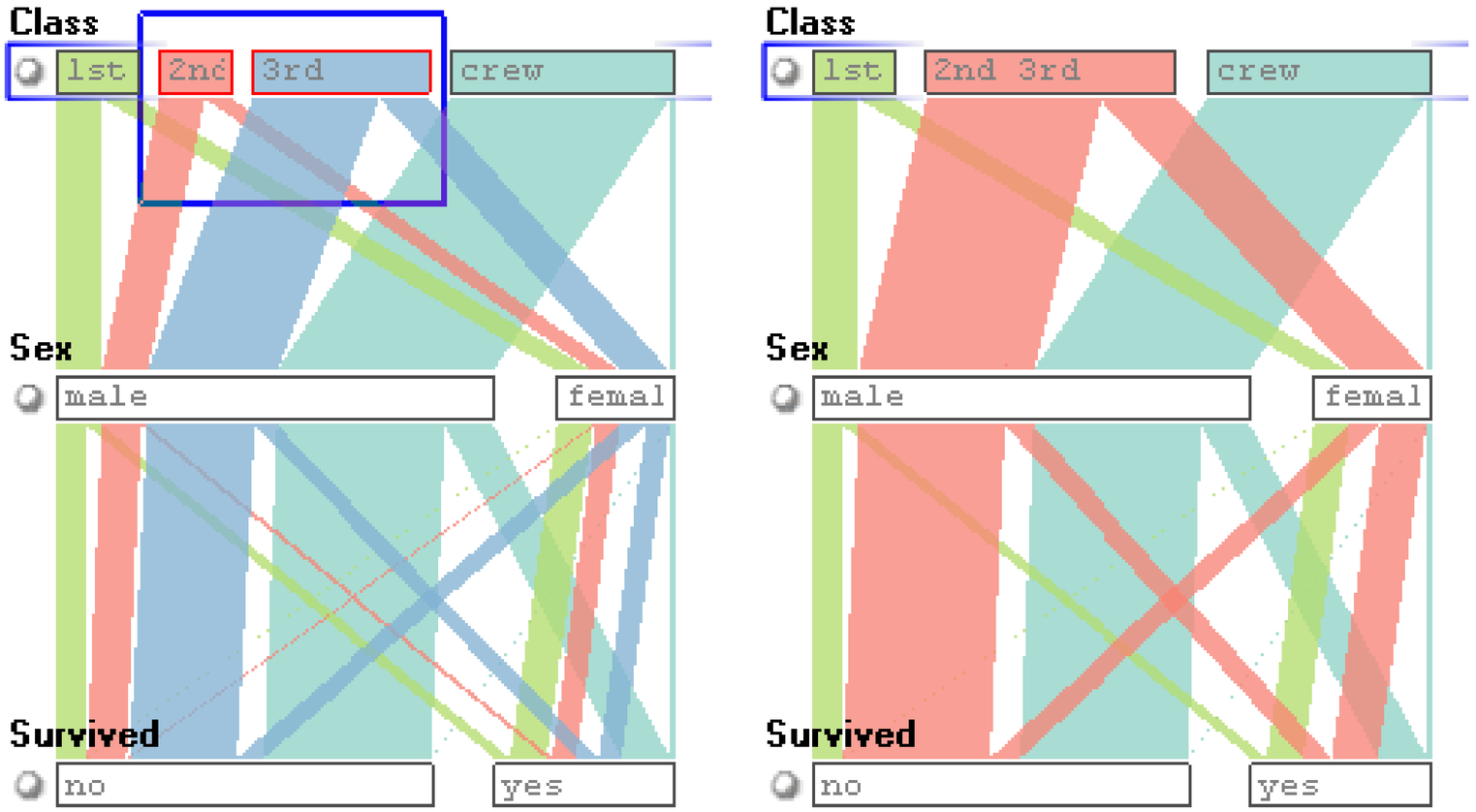
Bendix, Kosara, Hauser, 2005

Interaction: Reorder



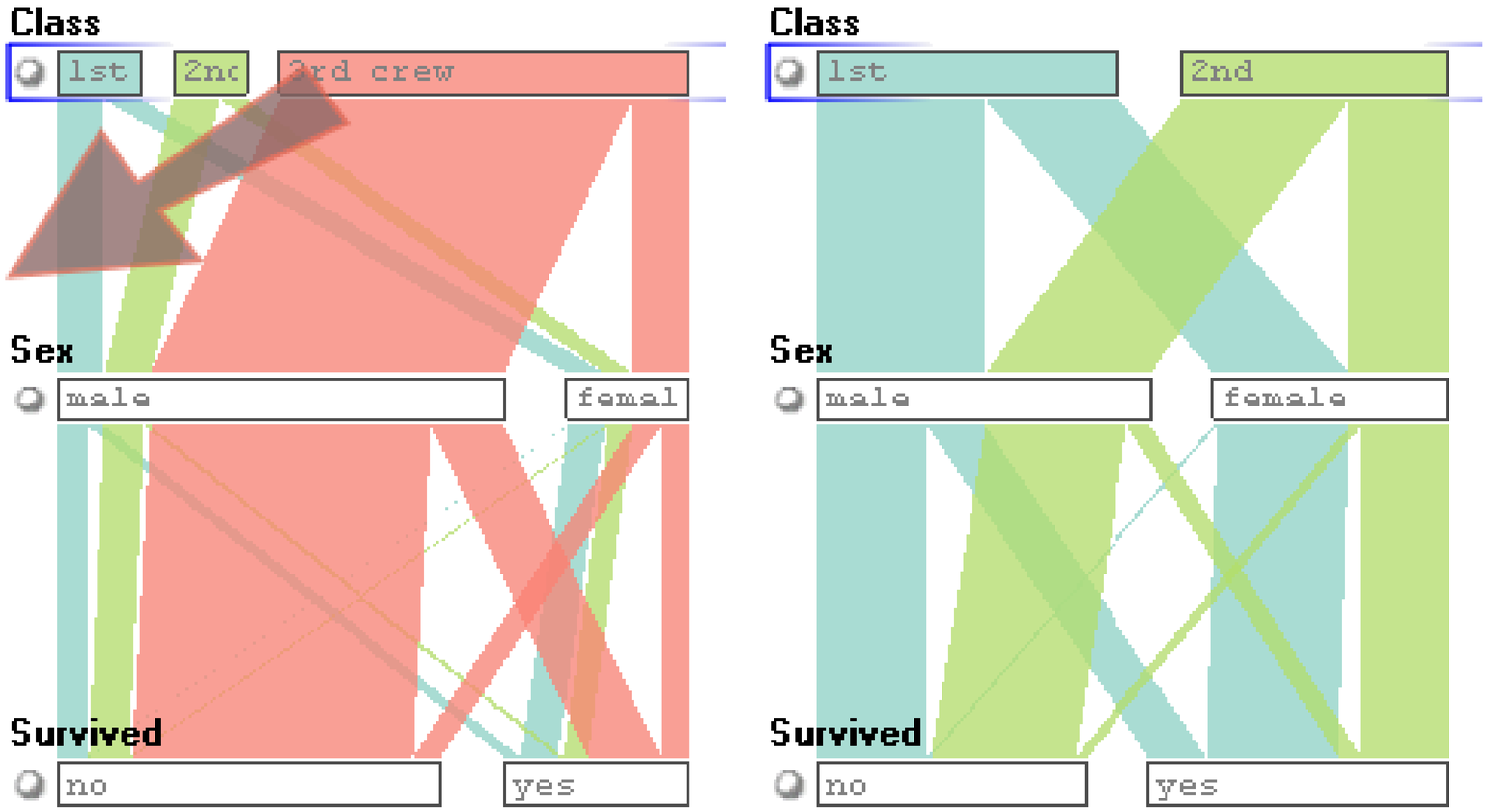
Bendix, Kosara, Hauser, 2005

Interaction: Aggregate



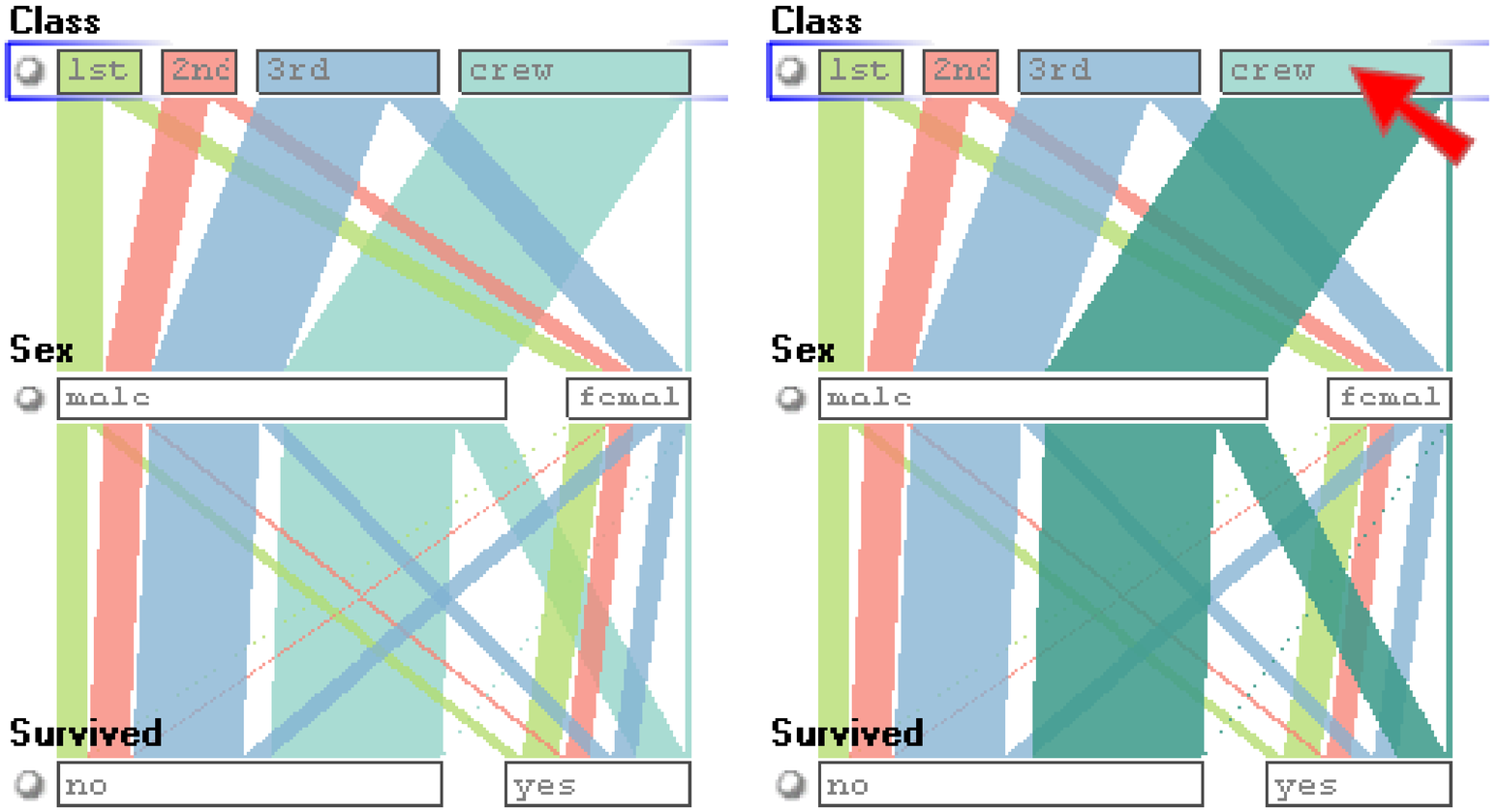
Bendix, Kosara, Hauser, 2005

Interaction: Filter



Bendix, Kosara, Hauser, 2005

Interaction: Highlight



Bendix, Kosara, Hauser, 2005

Filling Space

Arrange Tables

Express Values

Separate, Order, Align Regions

- Separate

- Order

- Align

  - *1 Key*
    List

  - *2 Keys*
    Matrix

  - *3 Keys*
    Volume

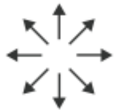
  - *Many Keys*
    Recursive Subdivision

Axis Orientation

- Rectilinear

- Parallel

- Radial



Layout Density

- Dense

- Space-Filling

HiVE example: London property

partitioning attributes

house type
neighborhood
sale time

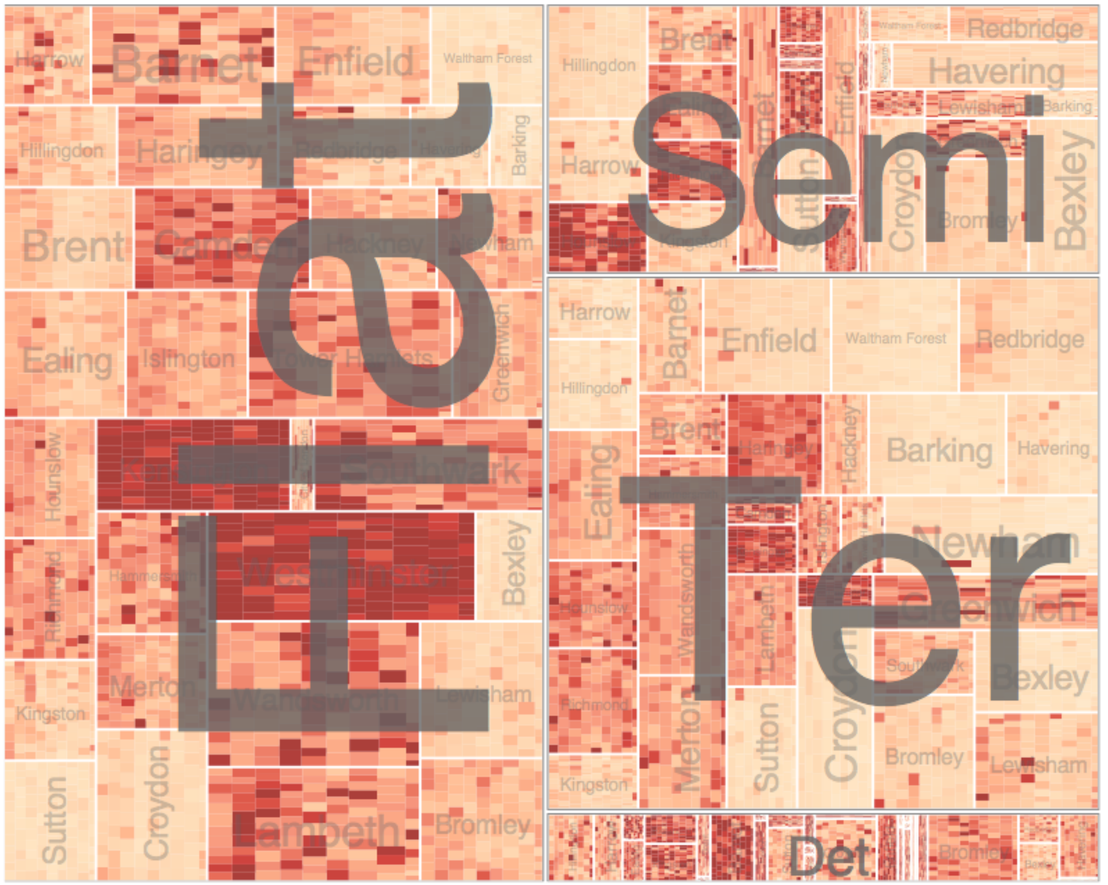
encoding attributes

average price (color)
number of sales (size)

results

between neighborhoods, different housing distributions

within neighborhoods, similar prices



Slingsby 2009

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

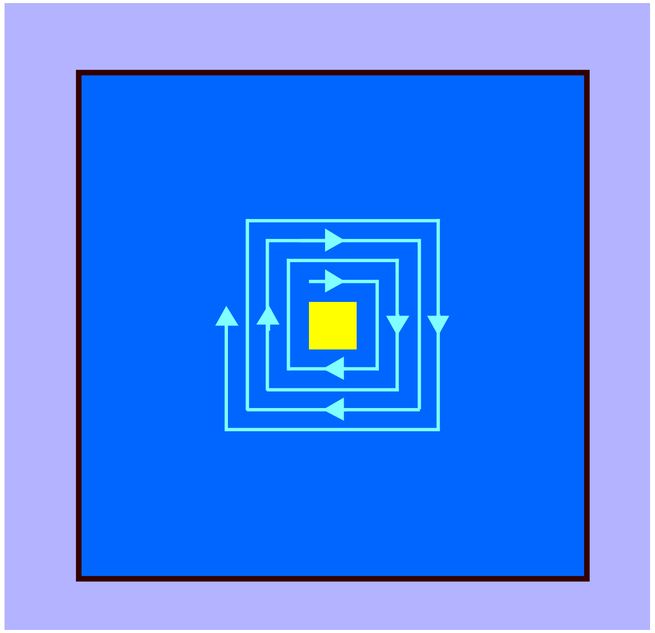
challenge: what’s the layout?

The data...

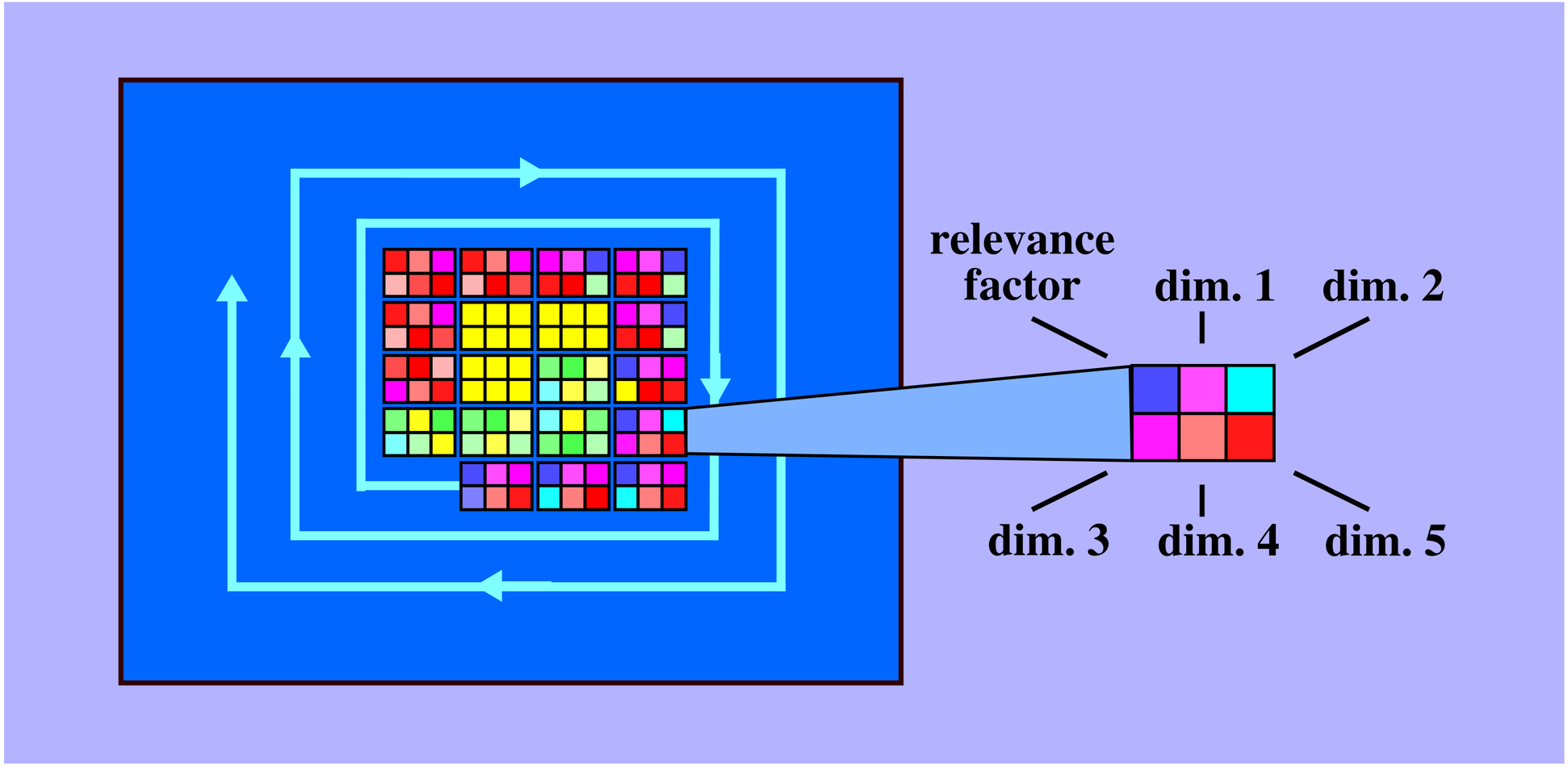
large database where each item has multiple attributes (on the order of 10)

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

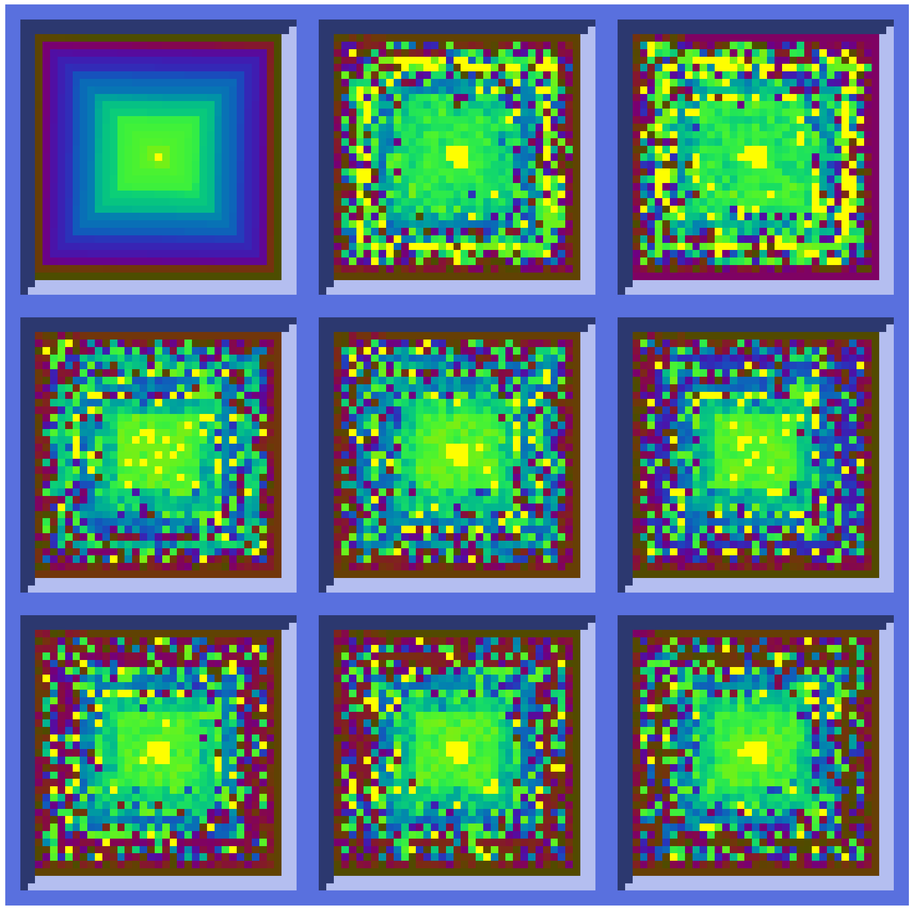
plot out data items in a spiral pattern, ordered by relevance

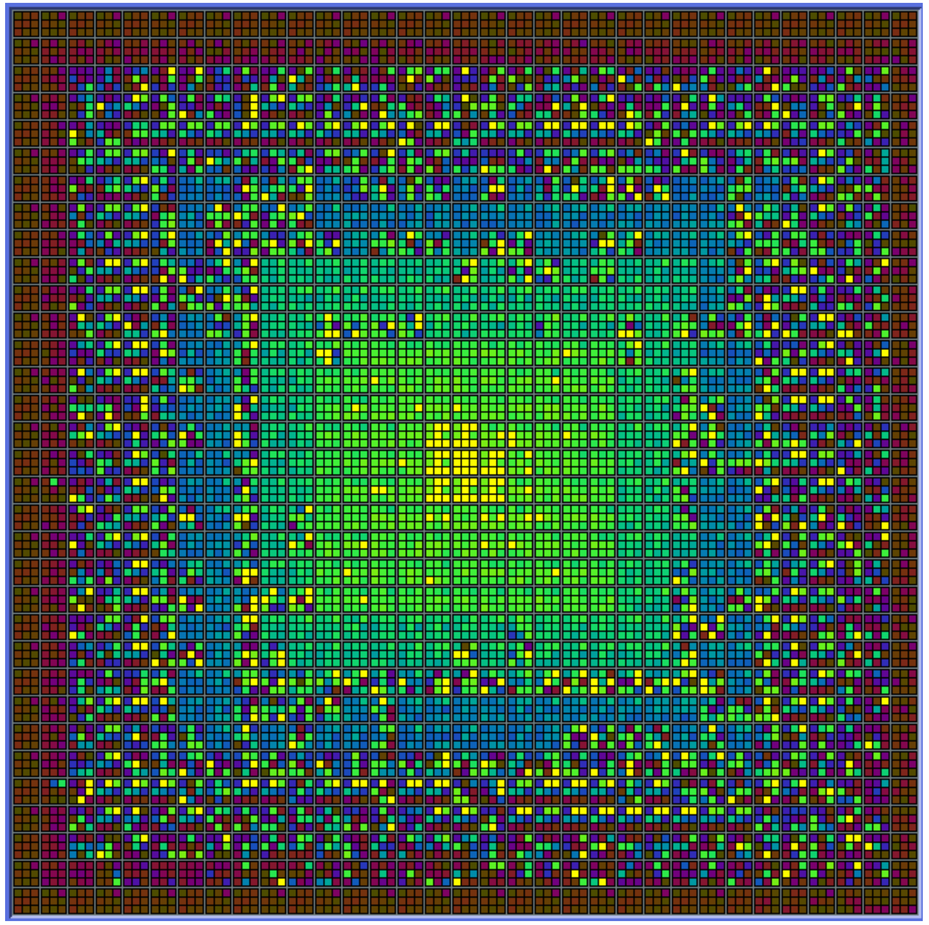


Keim, Kreigel, 1994



Keim, Kreigel, 1994





Keim, Kreigel, 1994