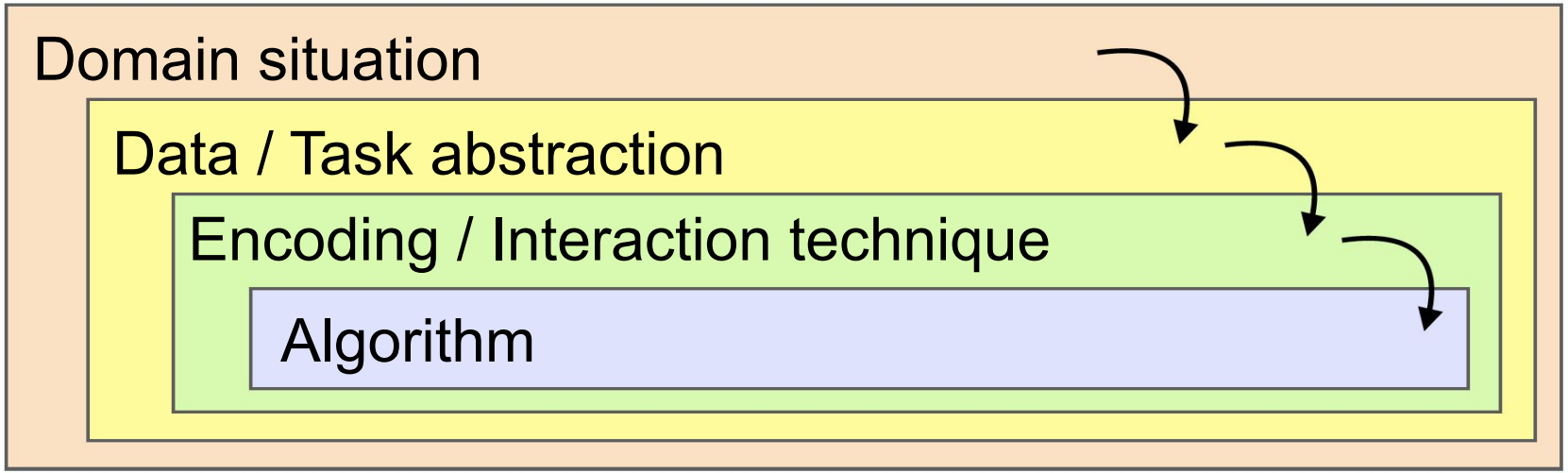
Visualizing Multi-Dimensional Data

Chen He

Data abstraction: Multi-dimensional data



1-dimensional data -- Word clouds

Show words used in text data

Pros quickly **perceive prominent** words;
draw attention;

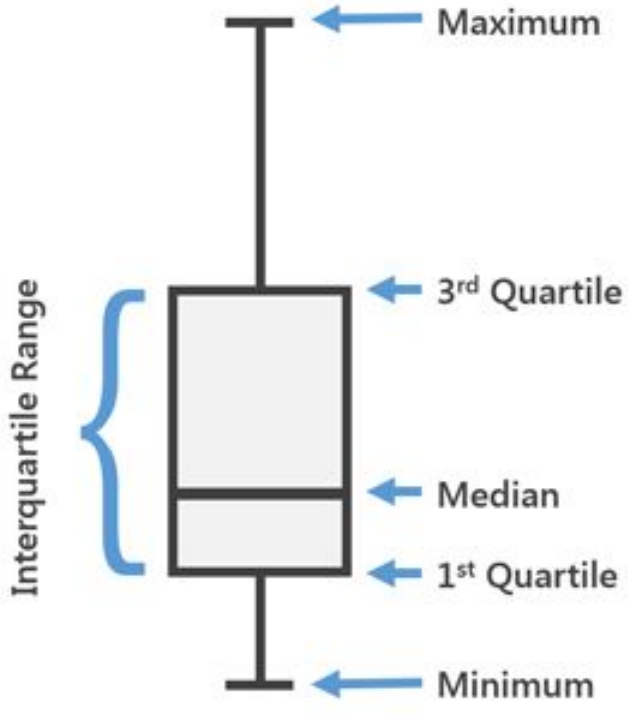
Cons **contradiction** between size and length;
lose **context**;

1-dimensional data -- Box plot

Use quartiles to graphically depict a group of numerical data.

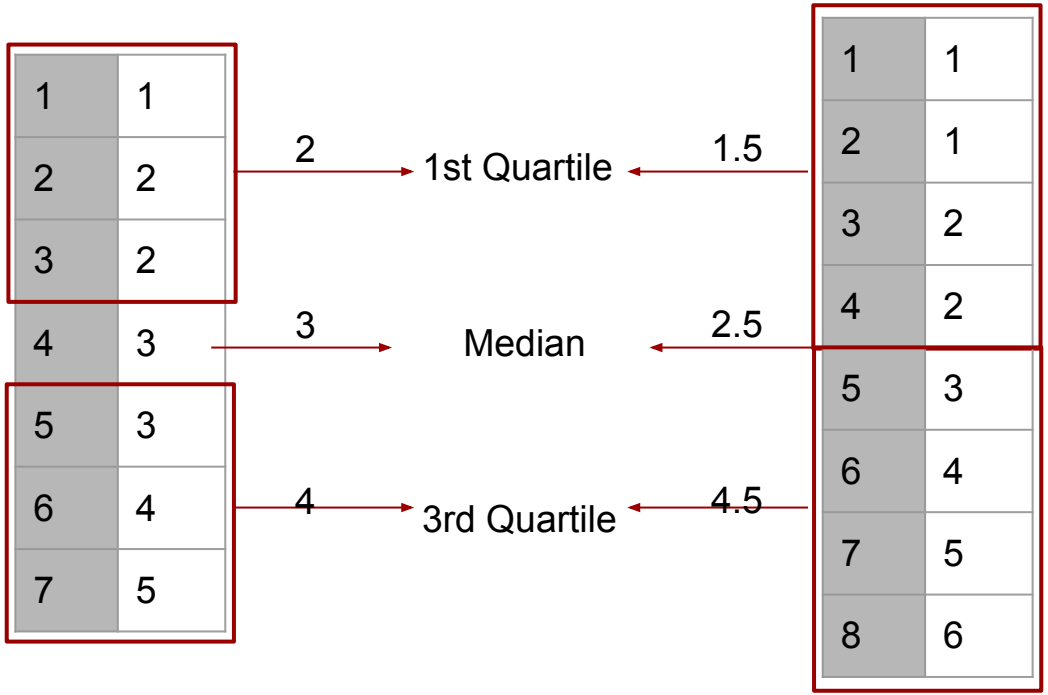
Pros Summarization;
Easy comparison between datasets.

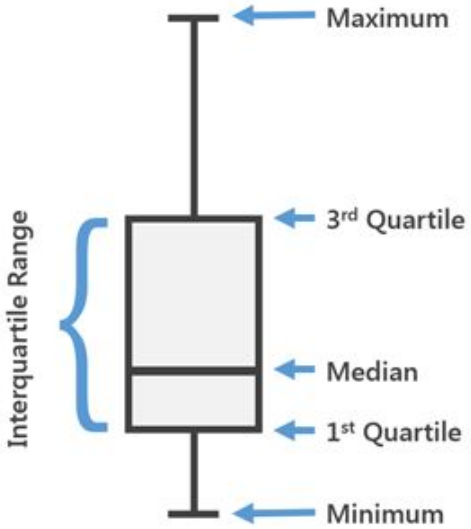
Cons Details are invisible;



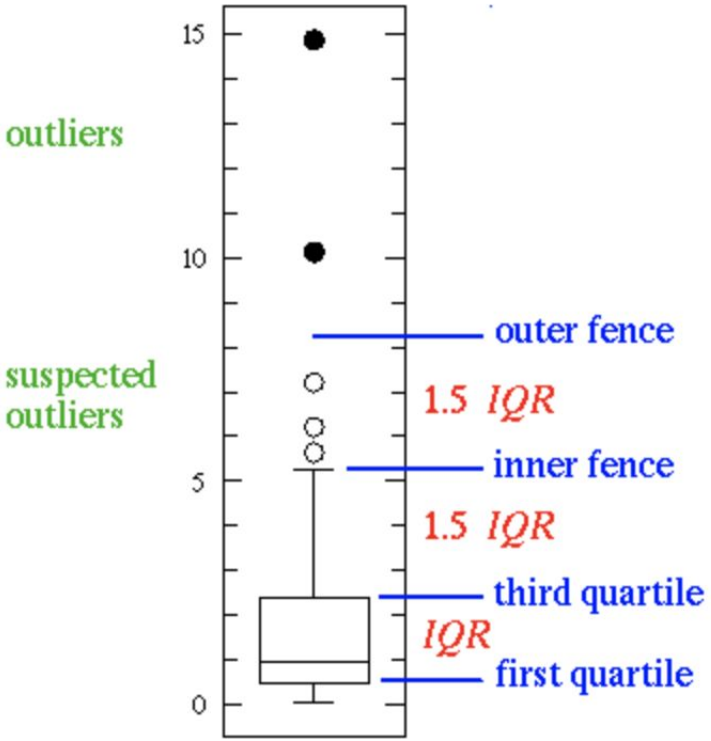
Draw boxplot: http://shiny.chemgrid.org/boxplotr/

1-dimensional data -- Box plot





1-dimensional data -- Box plot



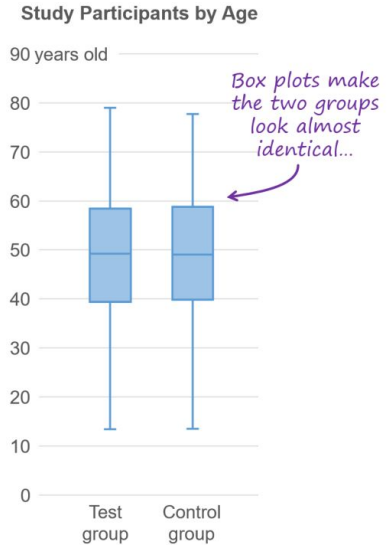
IQR: interquartile range

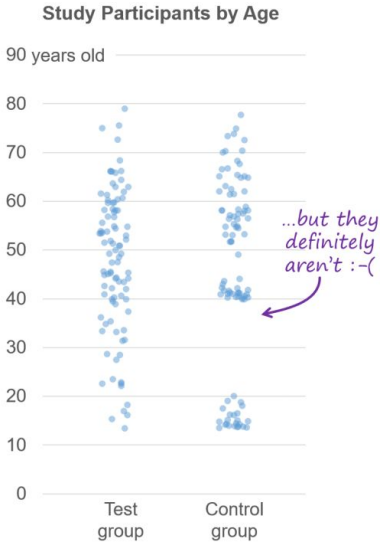
Suspected outliers:
> 1.5 IQR + 3rd Quartile;
< 1st Quartile - 1.5 IQR.

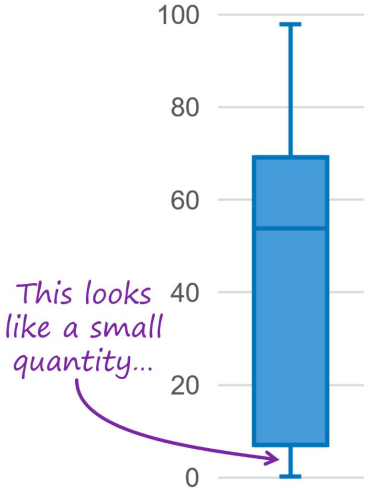
Outliers:
> 3 IQR + 3rd Quartile;
< 1st Quartile - 3 IQR.

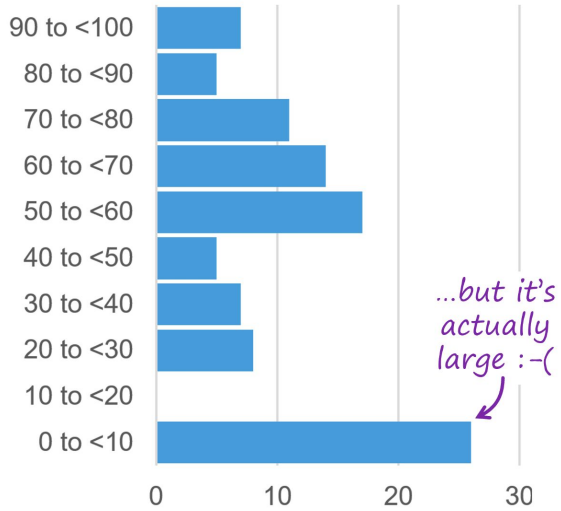
https://nightingaledvs.com/ive-stopped-using-box-plots-should-you/

“I’ve stopped using box plots.”



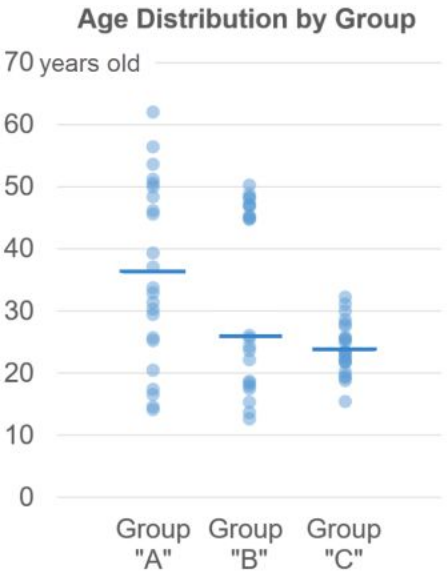


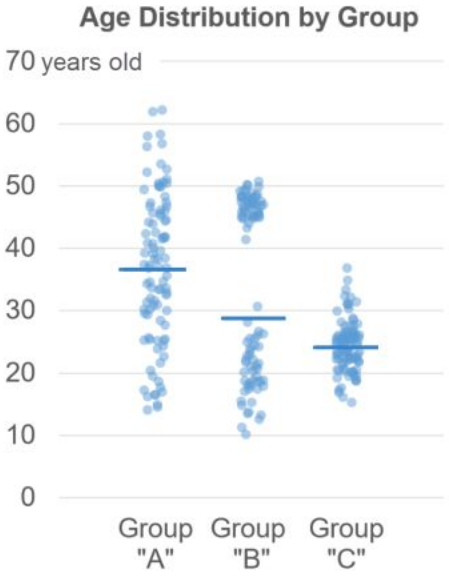


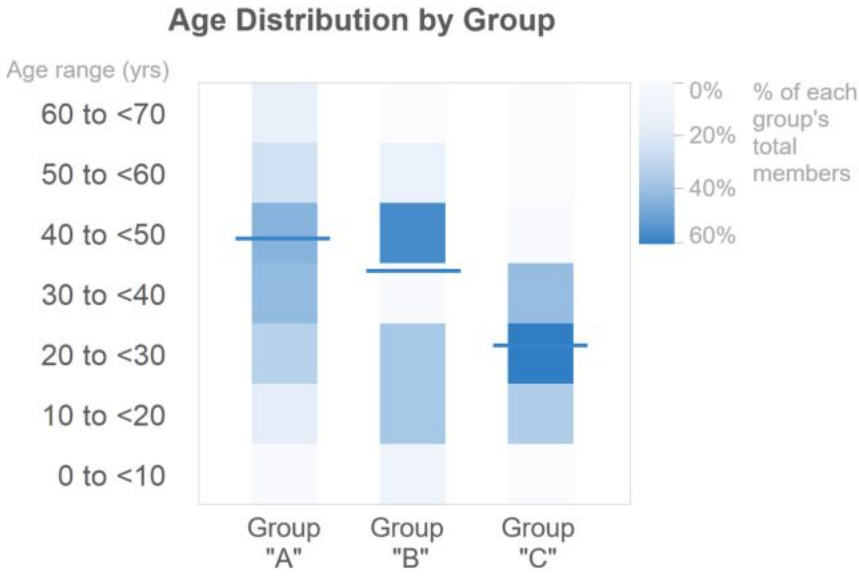


https://nightingaledvs.com/ive-stopped-using-box-plots-should-you/

“I’ve stopped using box plots.”



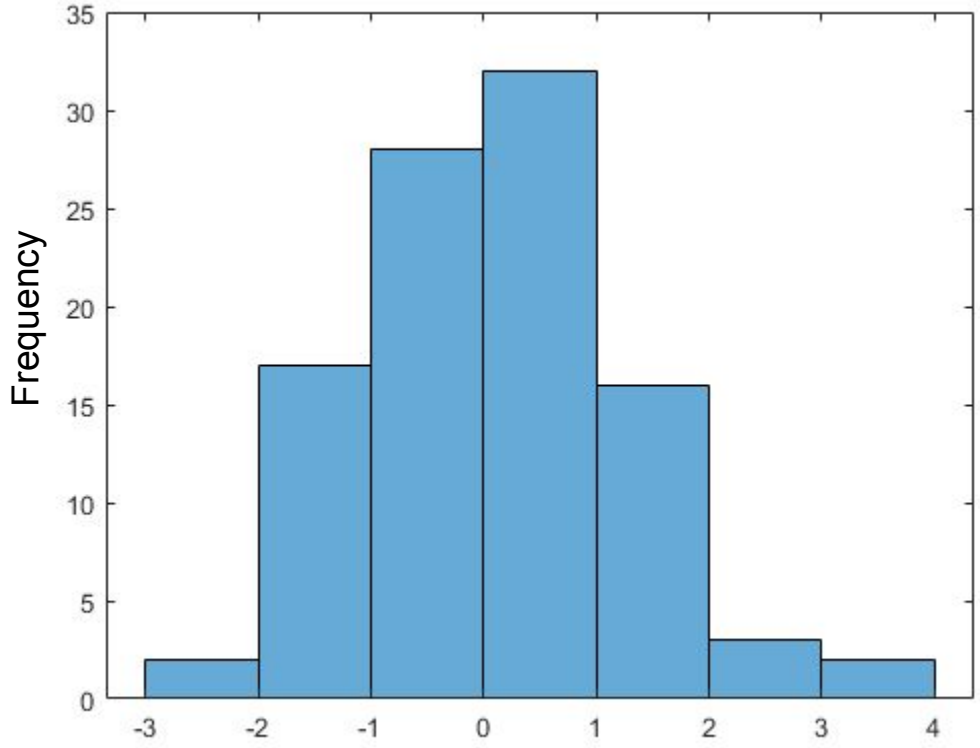




1-dimensional data -- Histogram

Depict the frequency distribution of a univariate data set.

By convention, an interval =
[lower bound, upper bound)



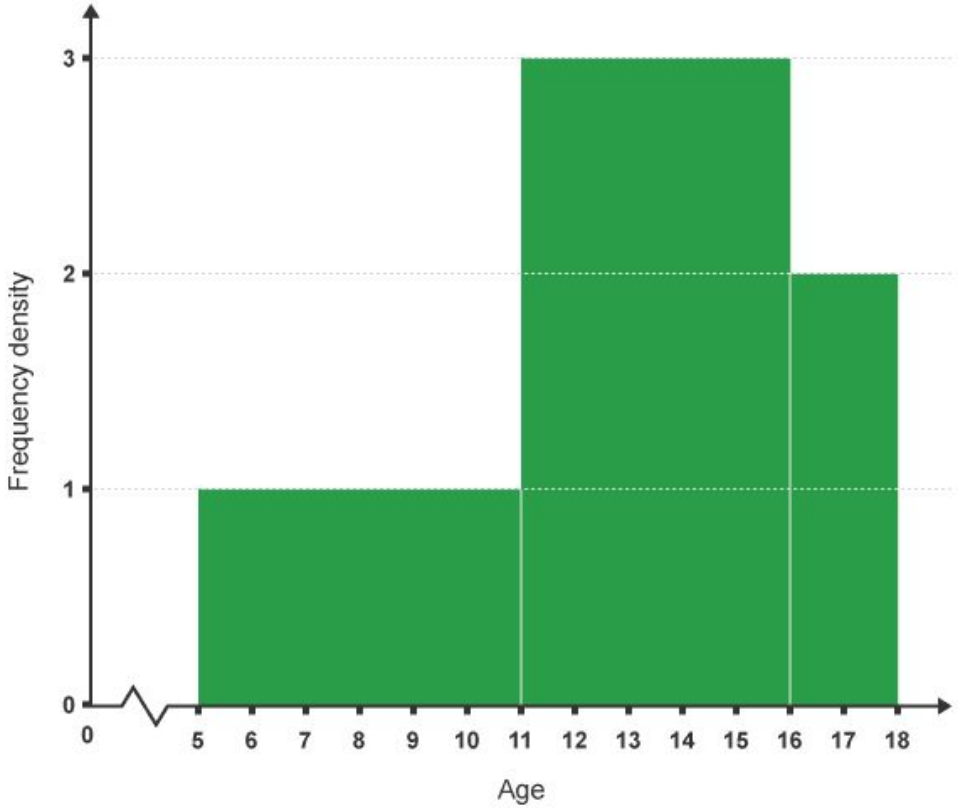
https://chartio.com/learn/charts/histogram-complete-guide/

1-dimensional data -- Histogram

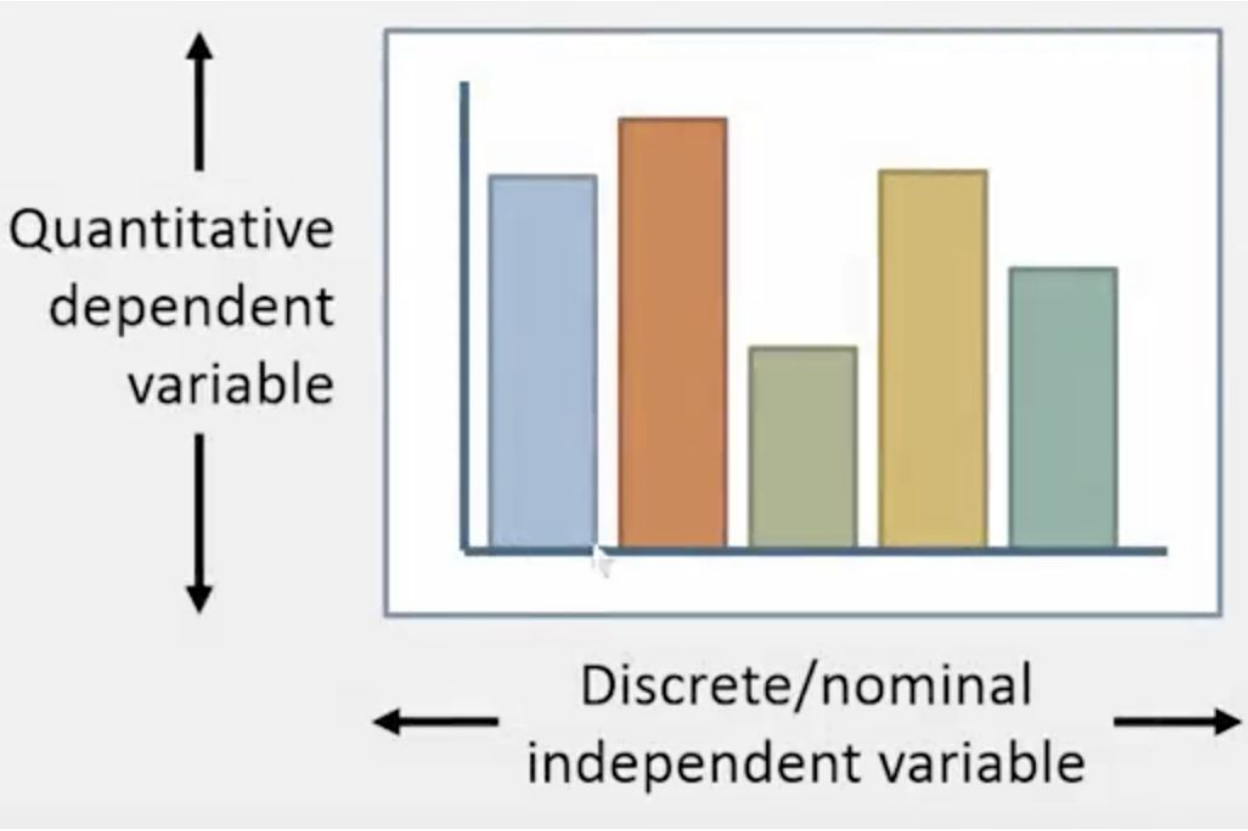
Unequal bin sizes (not recommended)

Frequency (the area) =
Class Width x Frequency Density.

Density is not easy to understand.



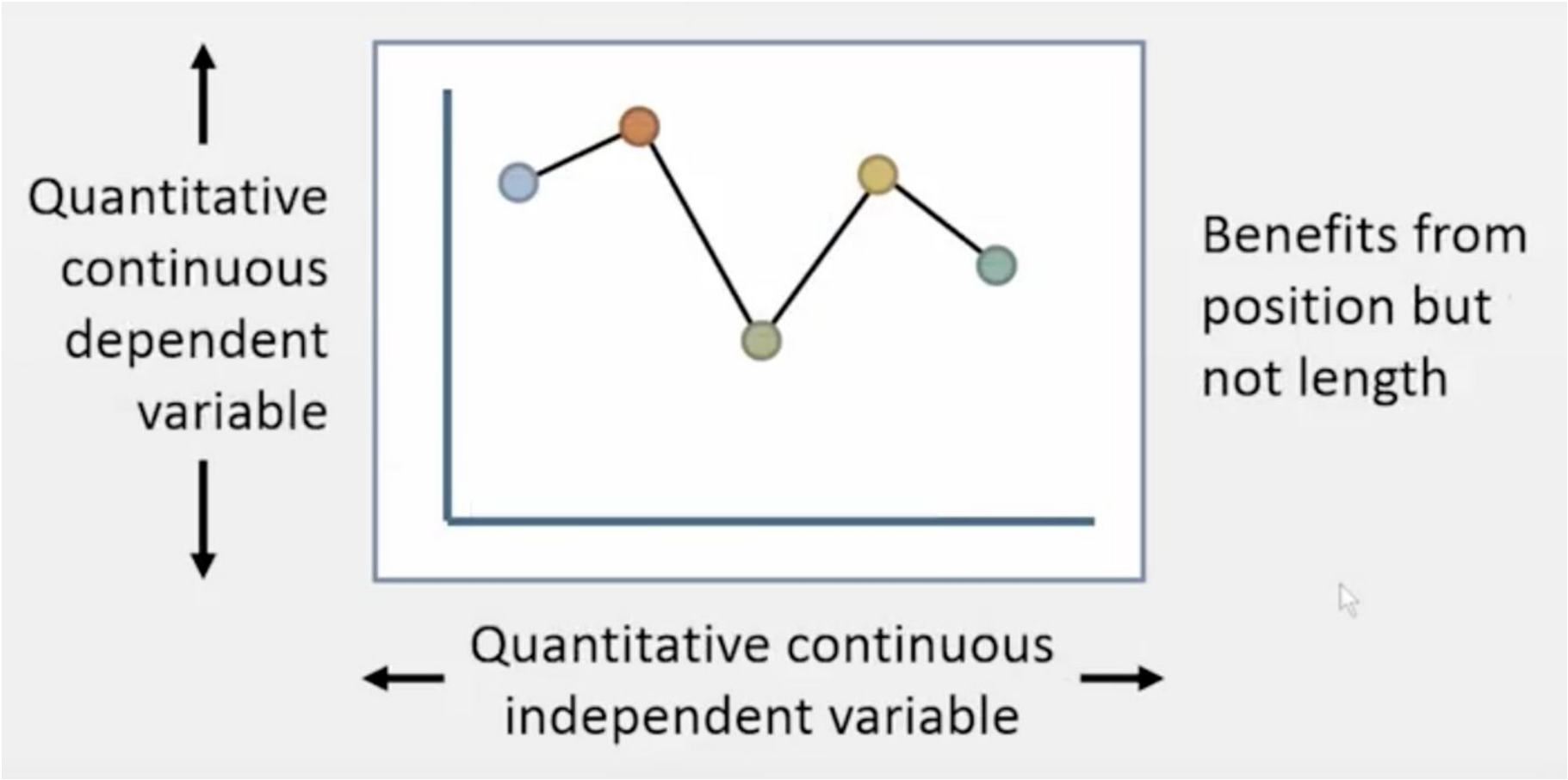
2-dimensional data -- Bar chart



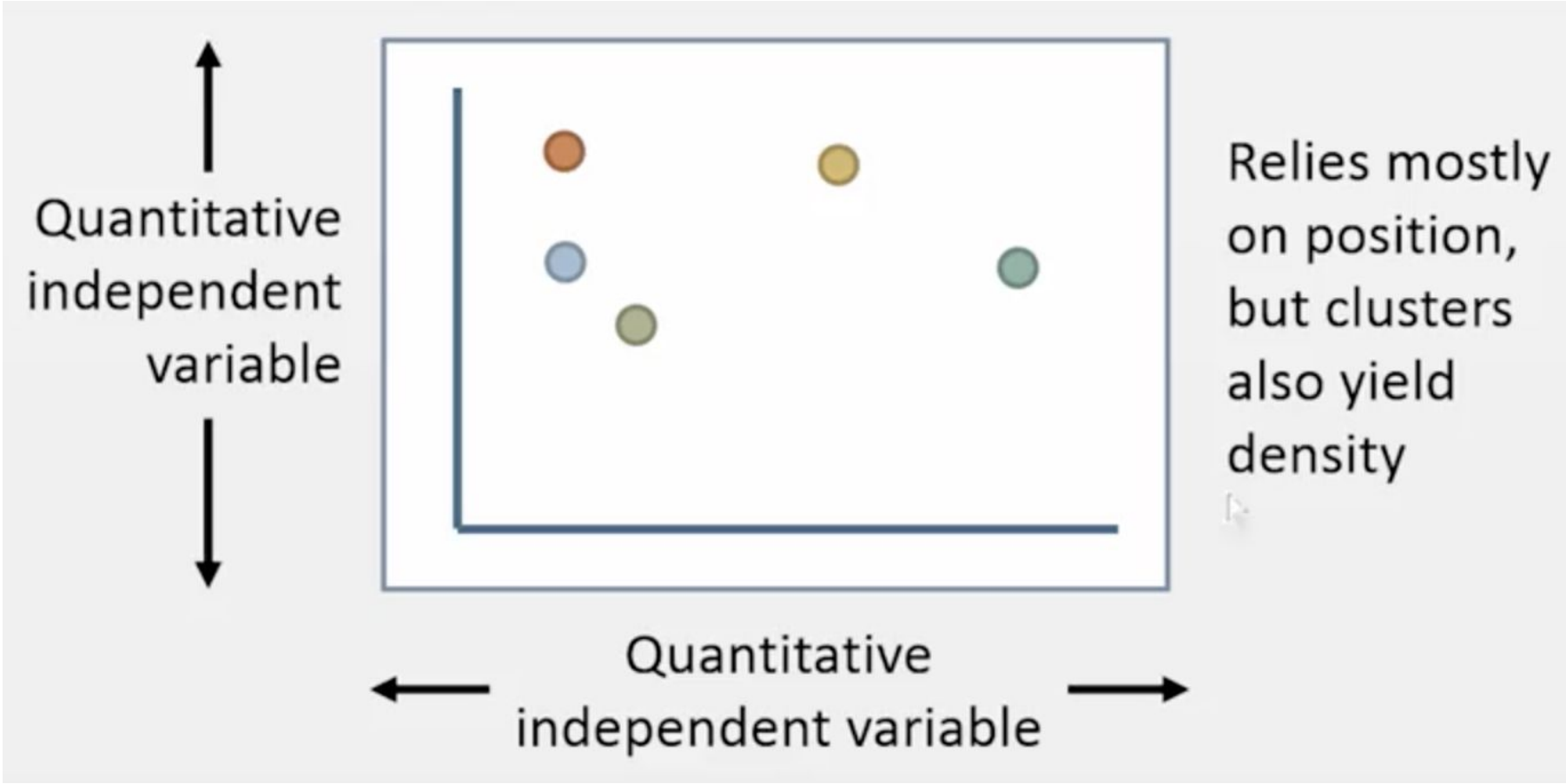
Benefits from both position (top of bar) and length (size of bar)

Quantitative	Ordinal	Nominal
Position	Position	Position
Length	Density	Hue
Angle	Saturation	Density
Slope	Hue	Saturation
Area	Length	Shape
Density	Angle	Length
Saturation	Slope	Angle
Hue	Area	Slope
Shape	Shape	Area

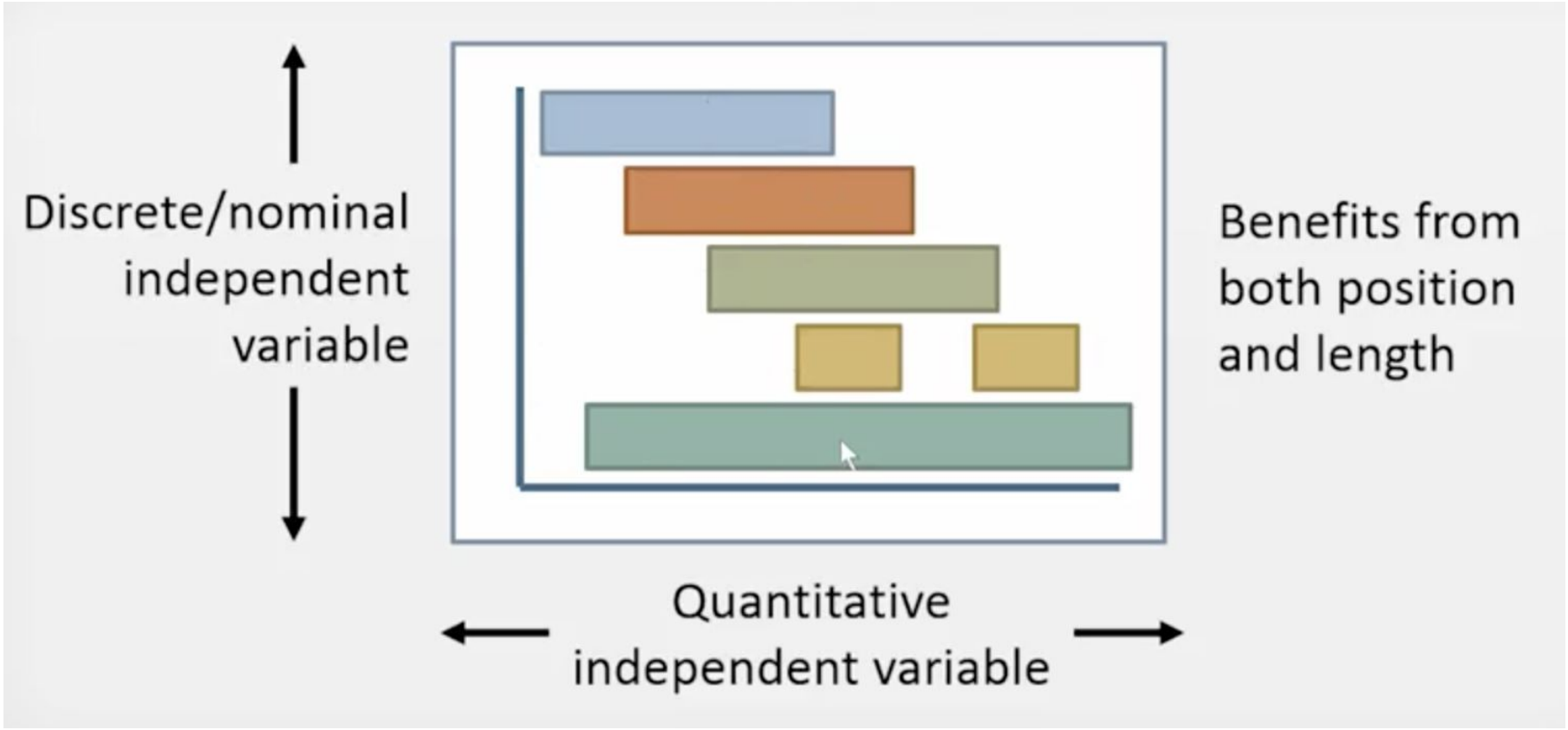
2-dimensional data -- Line chart



2-dimensional data -- Scatter plot

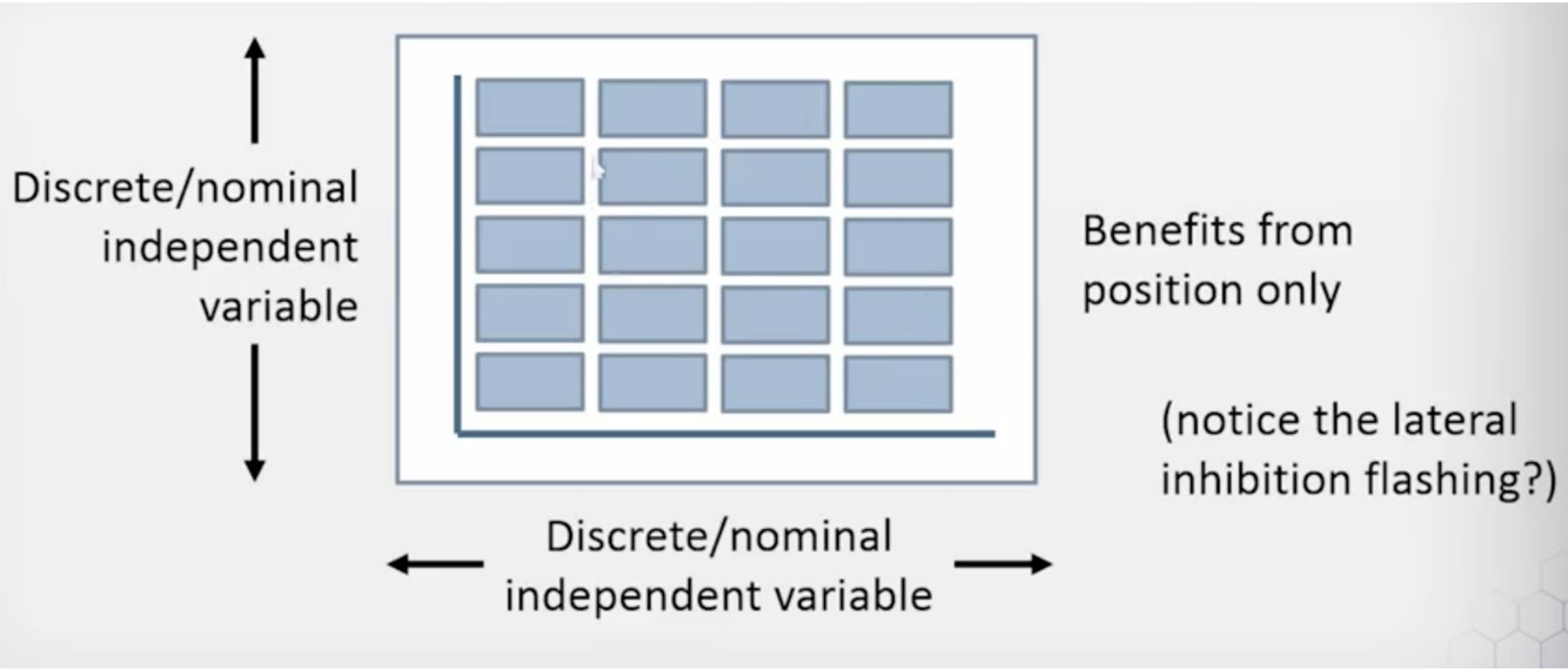


2-dimensional data -- Gantt chart



2-dimensional data -- Gantt chart

2-dimensional data -- Table



2-dimensional data -- What chart to use?

Dep.	Quantitative Continuous	Bar	Line
	Quantitative Discrete	Bar	Bar
Ind.	Quantitative Continuous	Gantt	Scatter
	Nominal or Q. Discrete	Table	Gantt
		Nominal or Q. Discrete	Quantitative Continuous
		Independent	

https://www.coursera.org/learn/datavisualization

Chart source: Sala-Burbaré, A., Peltonen, J., Pyhältö, K., & Castelló, M. (2018). Doctoral candidates' research writing perceptions: A cross-national study. *International Journal of Doctoral Studies*.

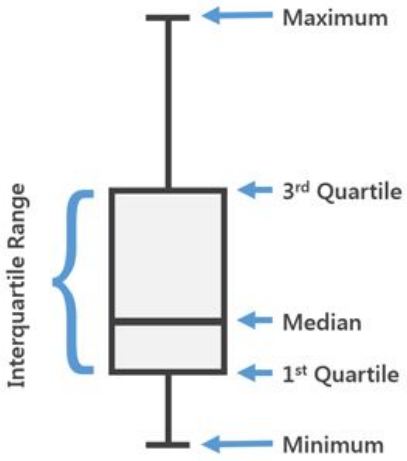
How about this chart?

Figure 2. Doctoral candidates' research writing profiles

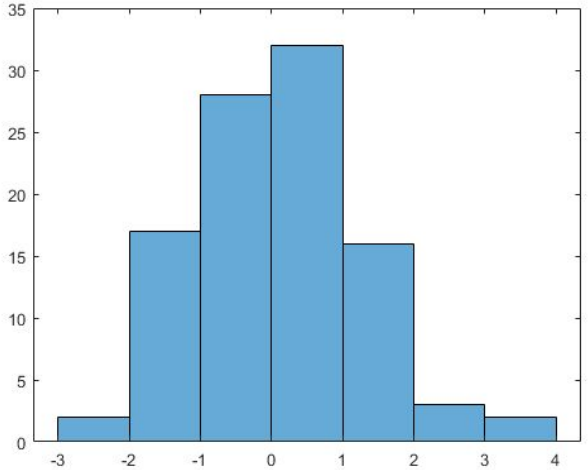
So far - 1D data

word clouds

box plot

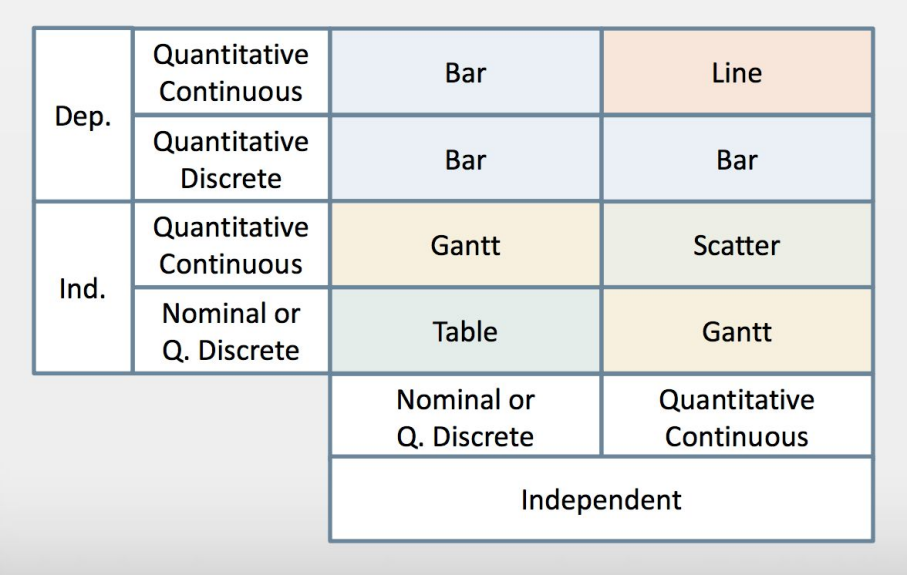


histogram



So far - 2D data

Bar chart, scatter plot, line chart, gantt chart, and table.

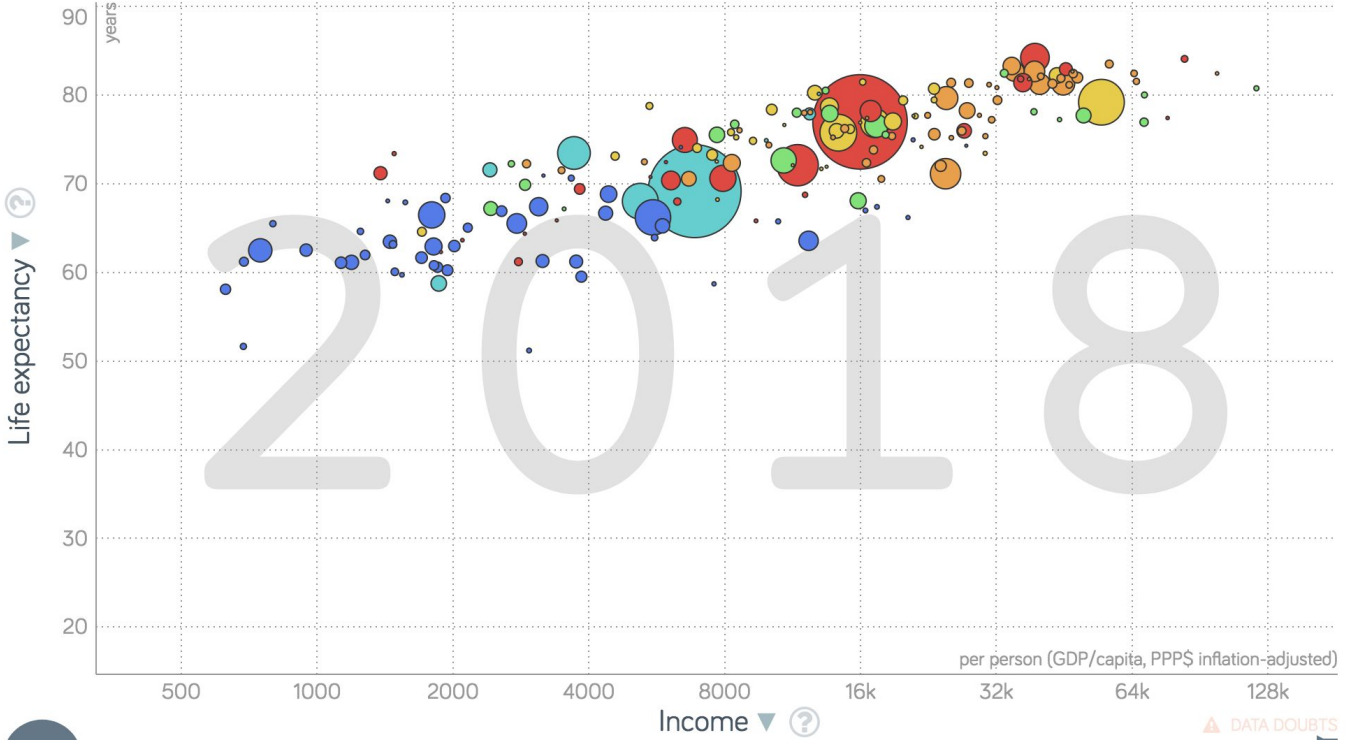


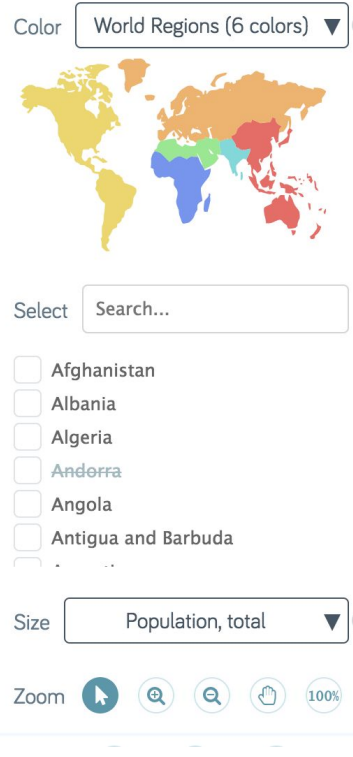
Dep.	Quantitative Continuous	Bar	Line
	Quantitative Discrete	Bar	Bar
Ind.	Quantitative Continuous	Gantt	Scatter
	Nominal or Q. Discrete	Table	Gantt
		Nominal or Q. Discrete	Quantitative Continuous
		Independent	

https://www.gapminder.org/tools/#$chart-type=bubbles

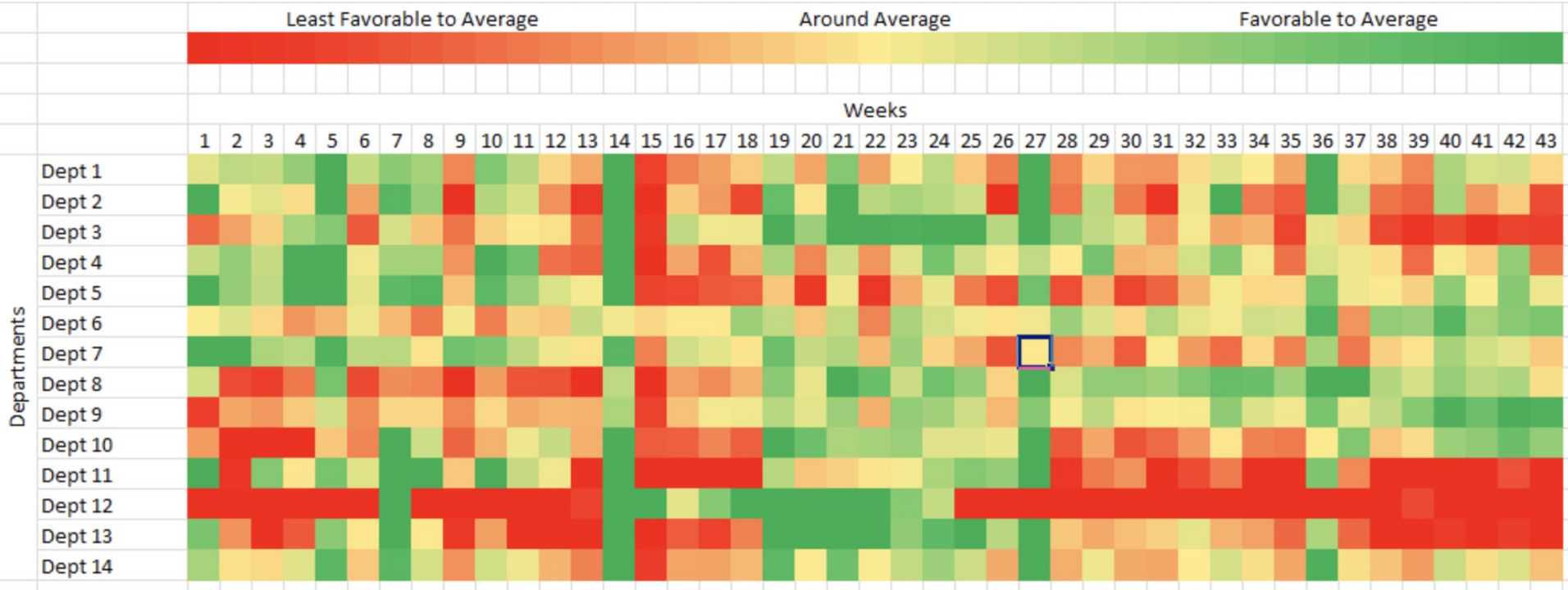
Multi-dimensional data -- Add additional dimensions on top of 2D charts

Quantitative	Ordinal	Nominal
Position	Position	Position
Length	Density	Hue
Angle	Saturation	Density
Slope	Hue	Saturation
Area	Length	Shape
Density	Angle	Length
Saturation	Slope	Angle
Hue	Area	Slope
Shape	Shape	Area





Multi-dimensional data -- Add additional dimensions on top of 2D charts



https://medium.com/@jshlbrd/hunting-for-powershell-using-heatmaps-69b70151fa5d

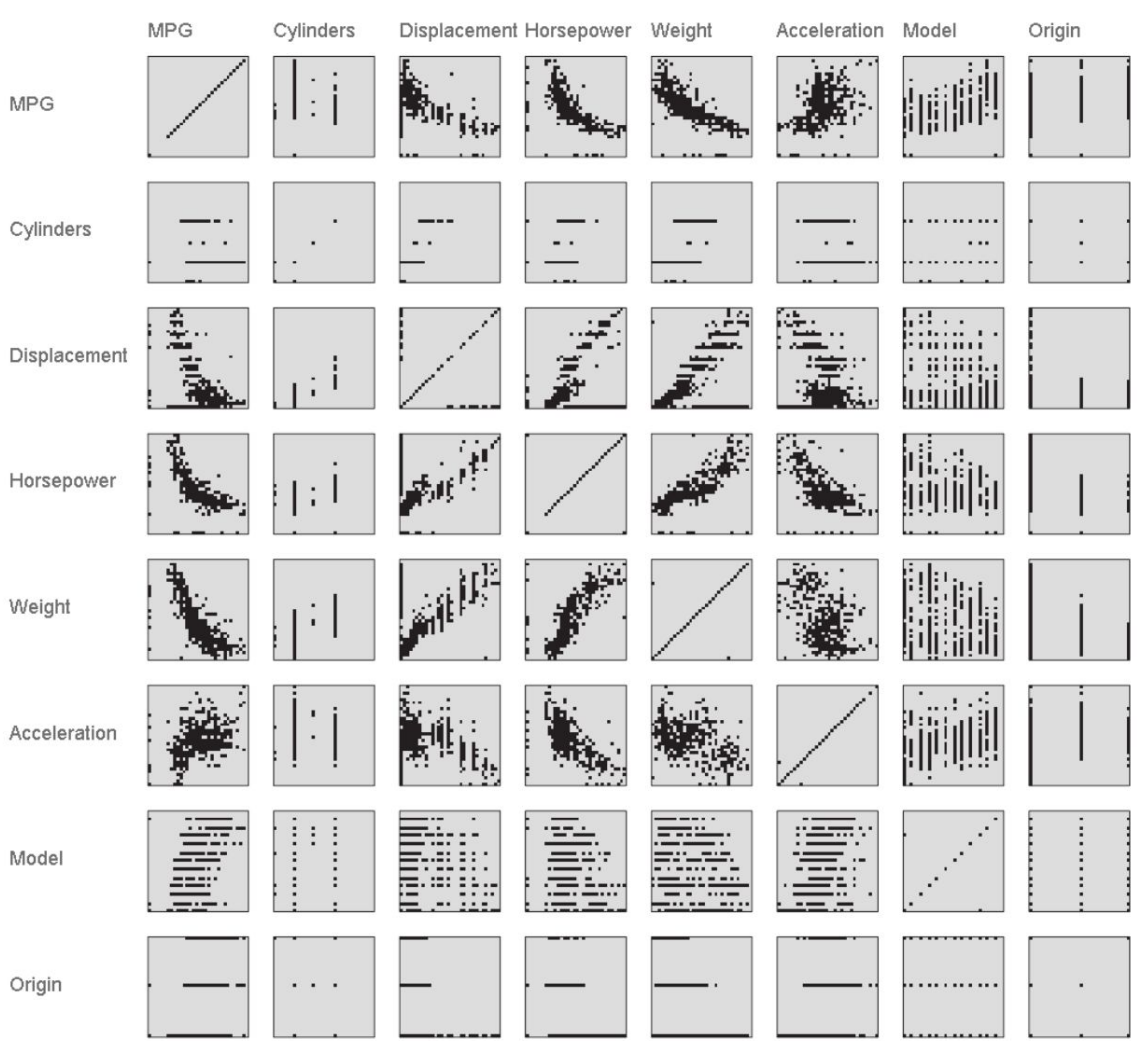
Multi-dimensional data -- Scatter plot matrix

N variables mean n*n plots.

Diagonal maps the same variable twice.

Each pair is plotted twice, once on each side of the diagonal.

Allows convenient sequential browsing of one variable **compared** to all other variables.



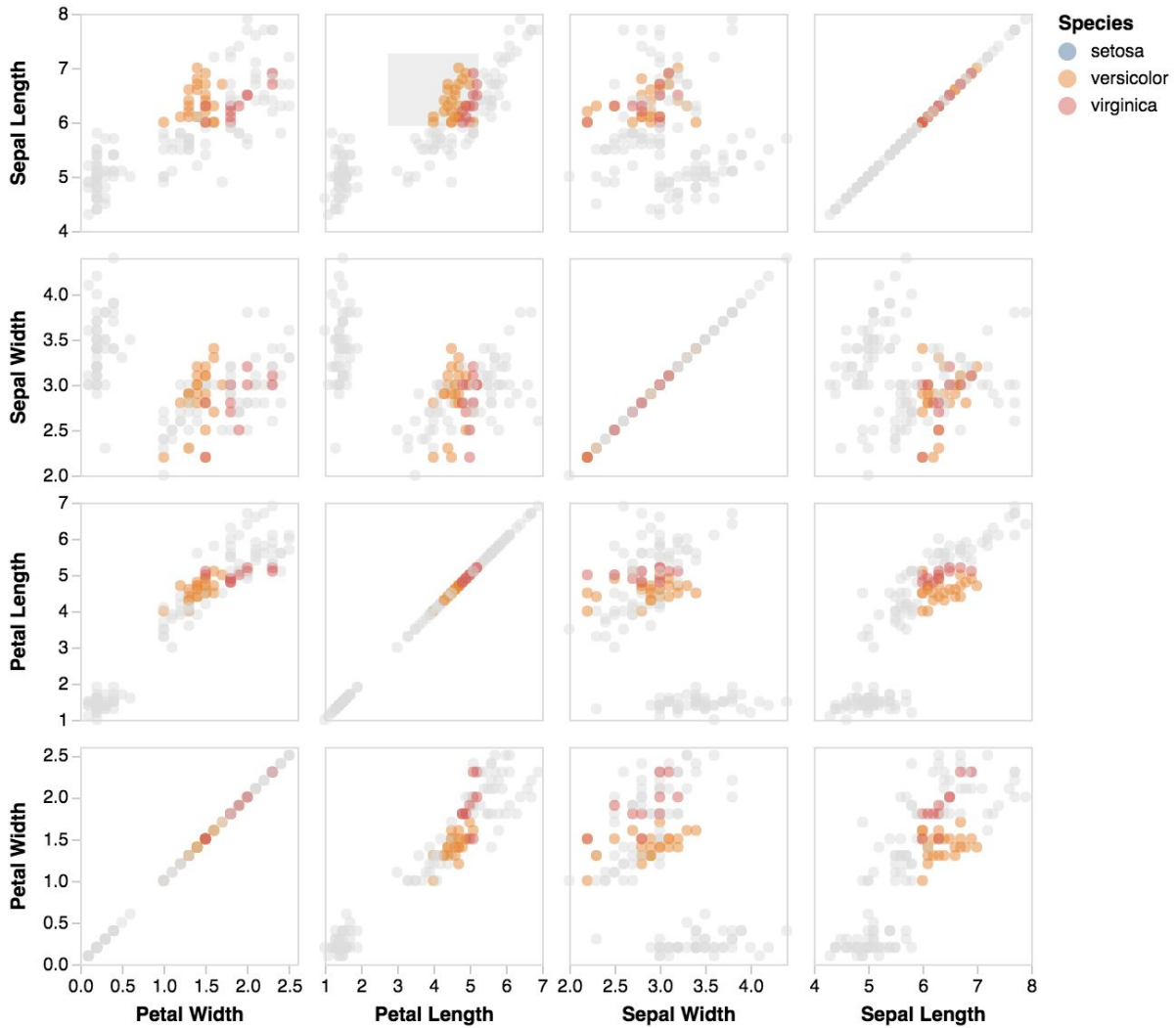
Brushing is the process of interactively selecting a subset of data items from a visual representation.

Brushing & linking cause the brush effect (highlighting, etc.) to be applied on those points in the other plots that represent the same data items.

Useful for exploring relationships in multi-dimensional data.

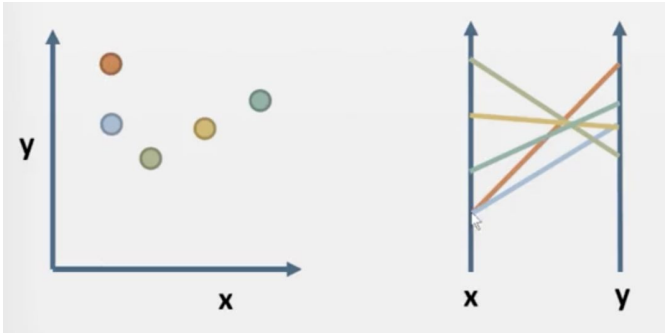
https://vega.github.io/vega/examples/brushing-scatter-plots/

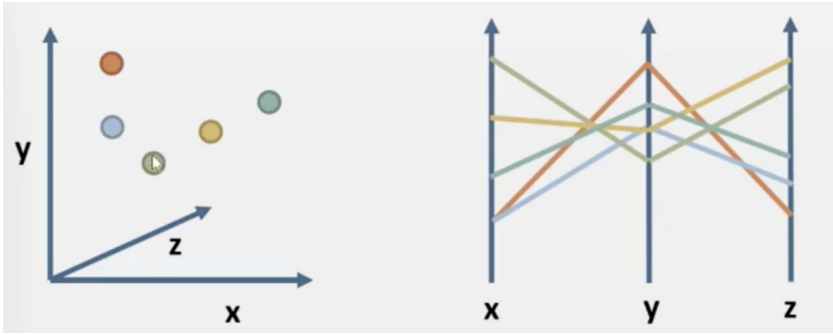
Brushing & linking scatter plots

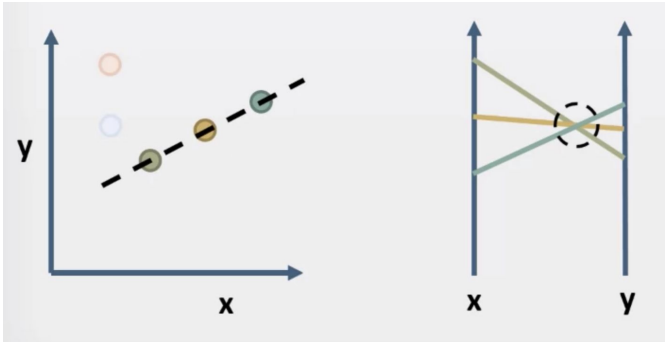


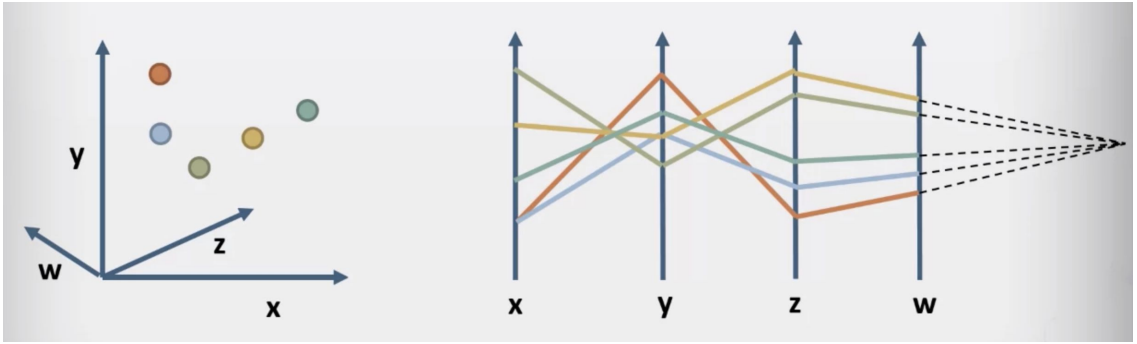
Multi-dimensional data -- Parallel coordinates

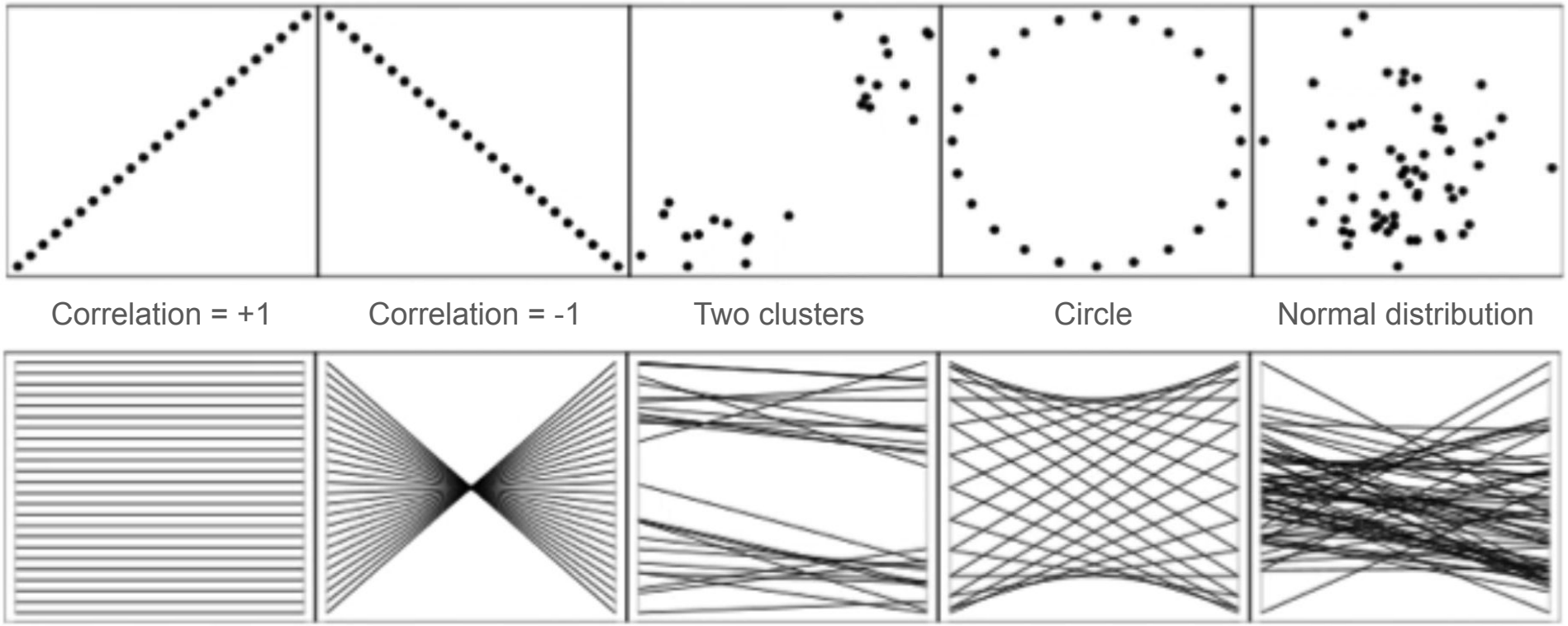
https://www.coursera.org/learn/datavisualization/lecture/2v40S/2-2-2-parallel-coordinates







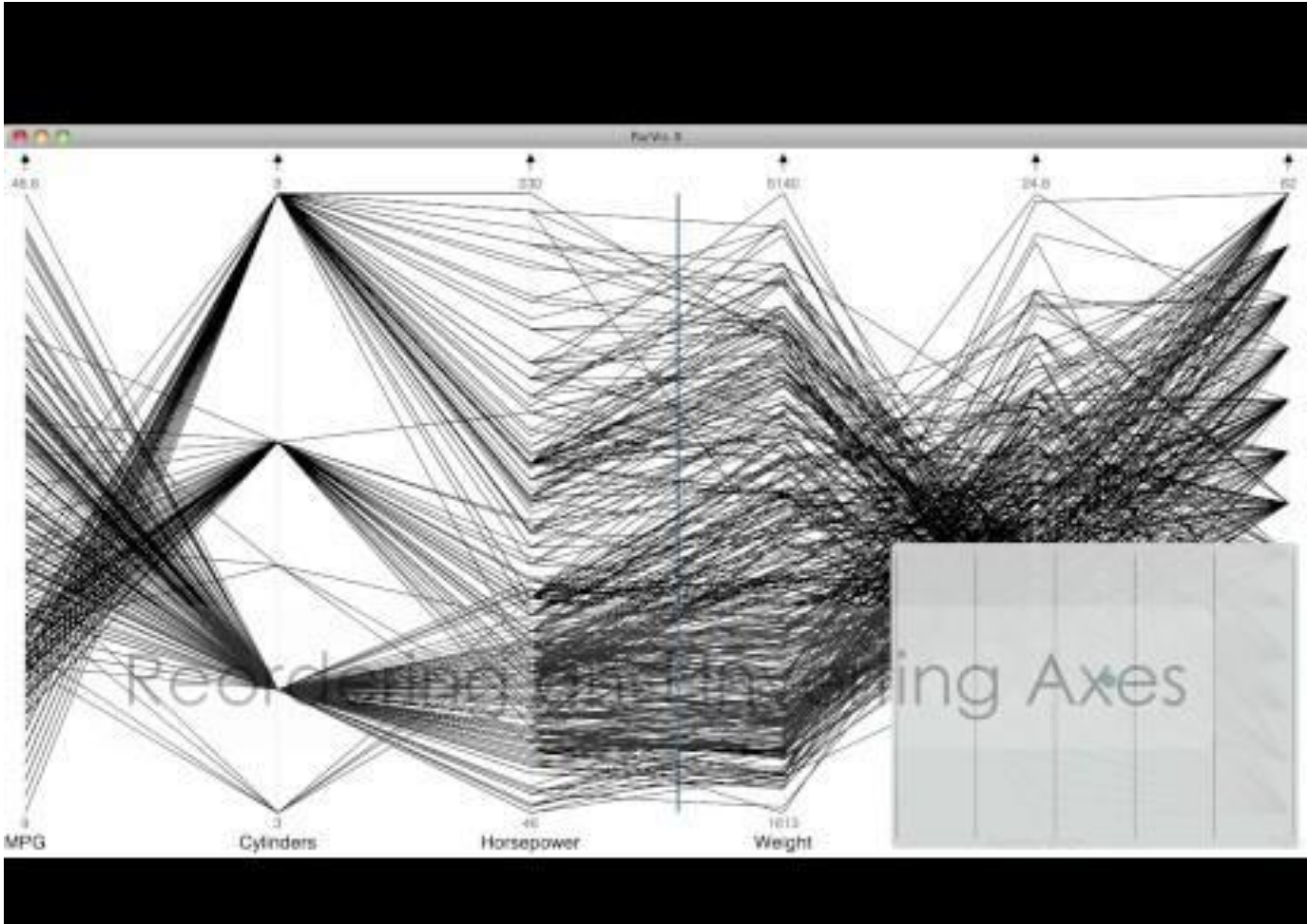


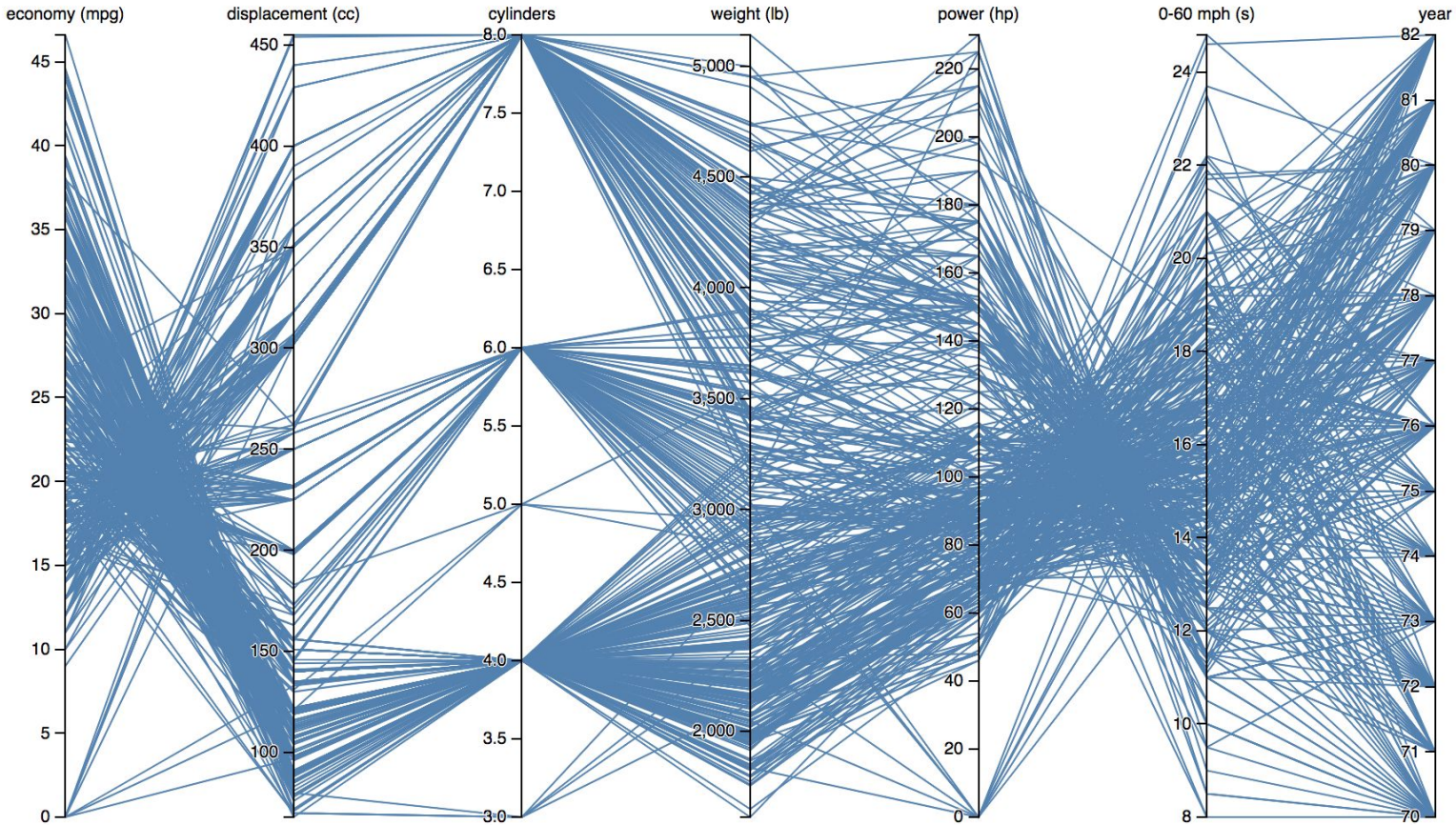

Correlation = +1
Correlation = -1
Two clusters
Circle
Normal distribution

Interacting with parallel coordinates

https://vimeo.com/13437693

Sort an axis;
Reorder axes;
Single axis brushing;
Angular brushing.





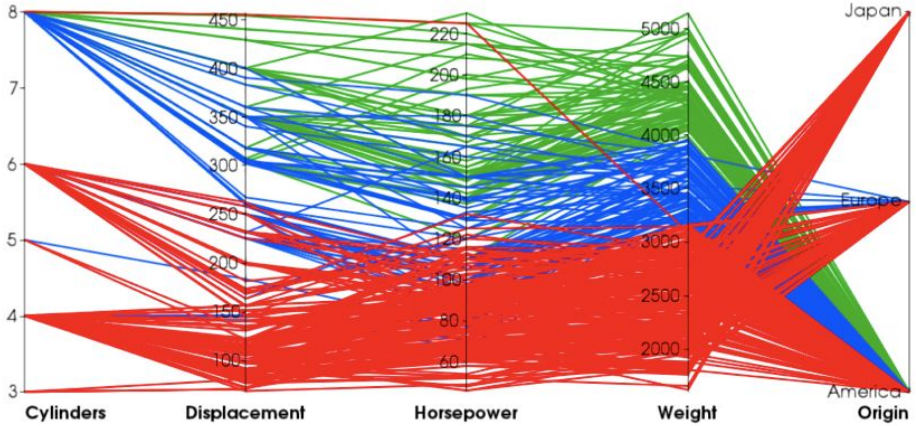
Each data value placed somewhere along the line, scaled between the minimum at the bottom and the maximum at the top.

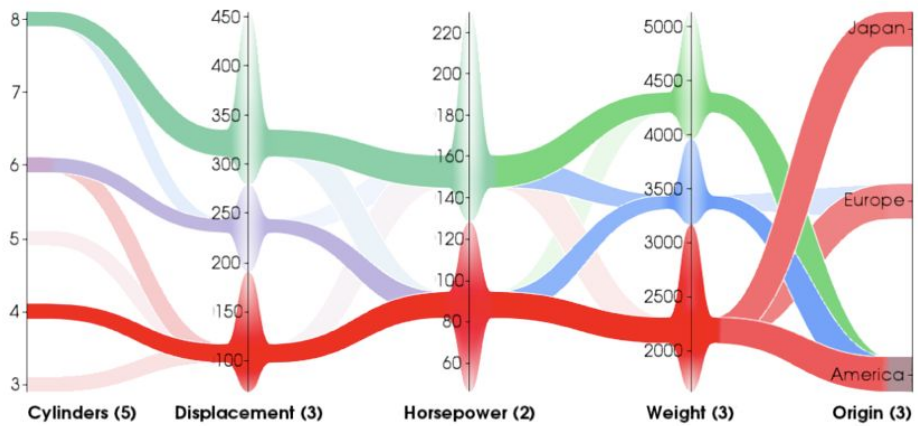
Ordering of the axis is fundamental to draw insights.

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

Palmas et al. 2014, March. An edge-bundling layout for interactive parallel coordinates. *PacificVis.*

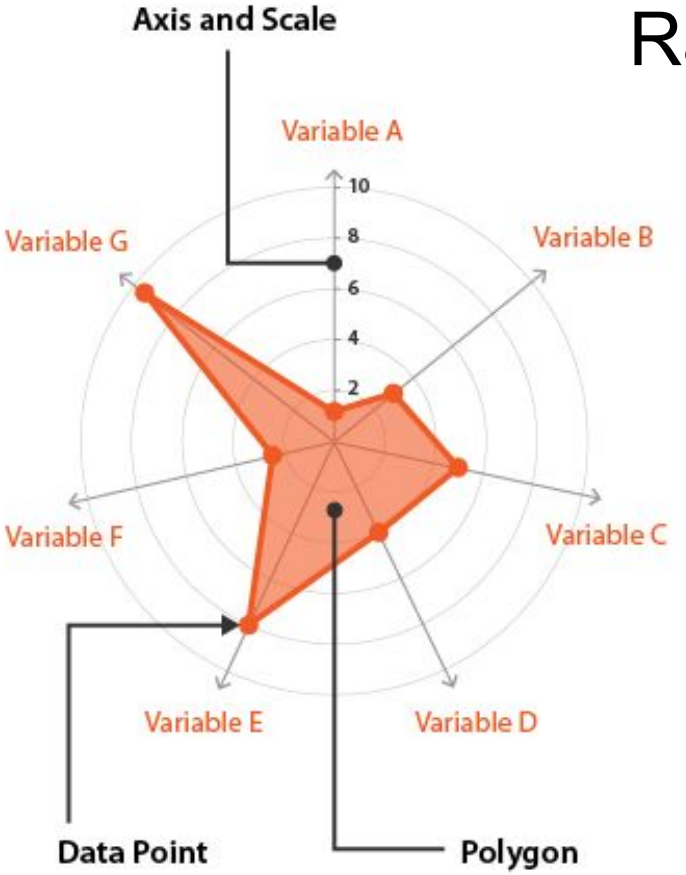
Multi-dimensional data -- Edge bundling for parallel coordinates



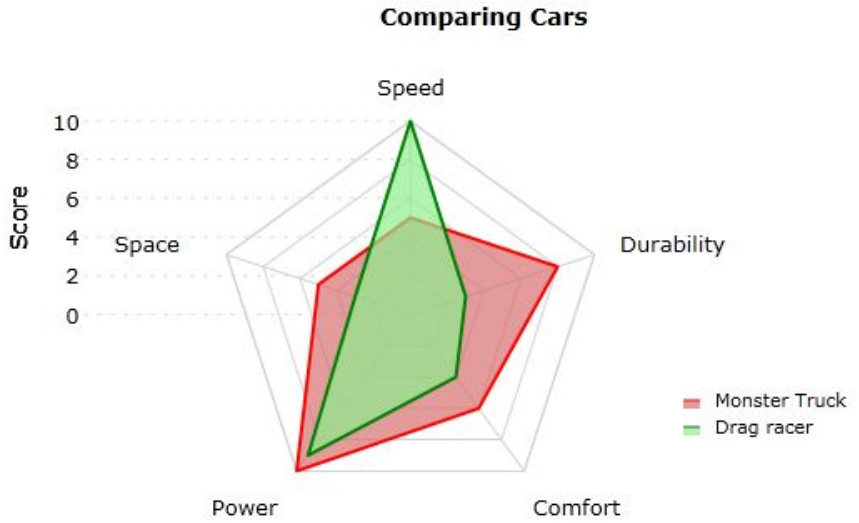


https://datavizcatalogue.com/methods/radar_chart.html

Radial coordinates / Radar chart



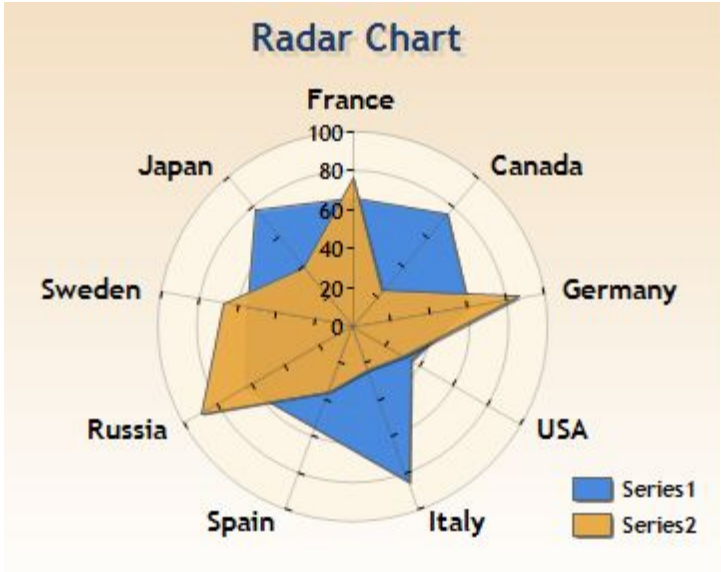
+ Useful for seeing which variables have similar values or if there are any outliers amongst each variable.
+ Visually interesting.

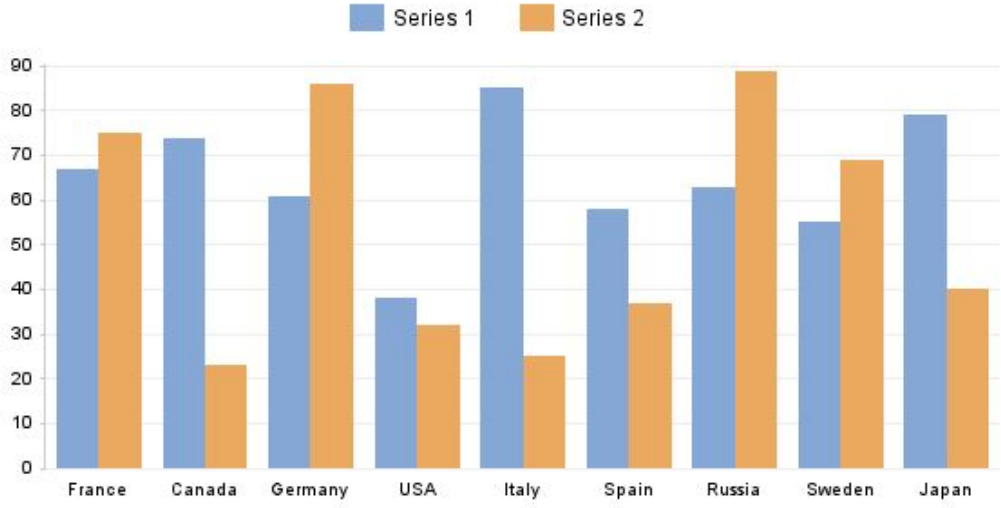


https://ldld.samizdat.cc/2017/linear-and-radial-parallel-coordinates/

Radial coordinates / Radar chart

- Visually cluttered with too many polygons;
- Not easy to compare values along circular grids;

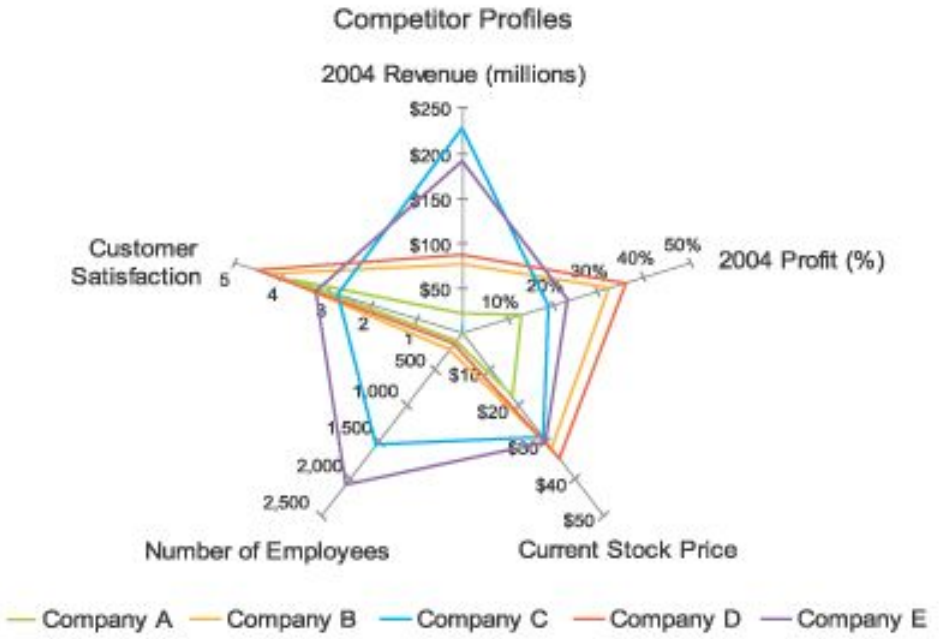




https://ldld.samizdat.cc/2017/linear-and-radial-parallel-coordinates/

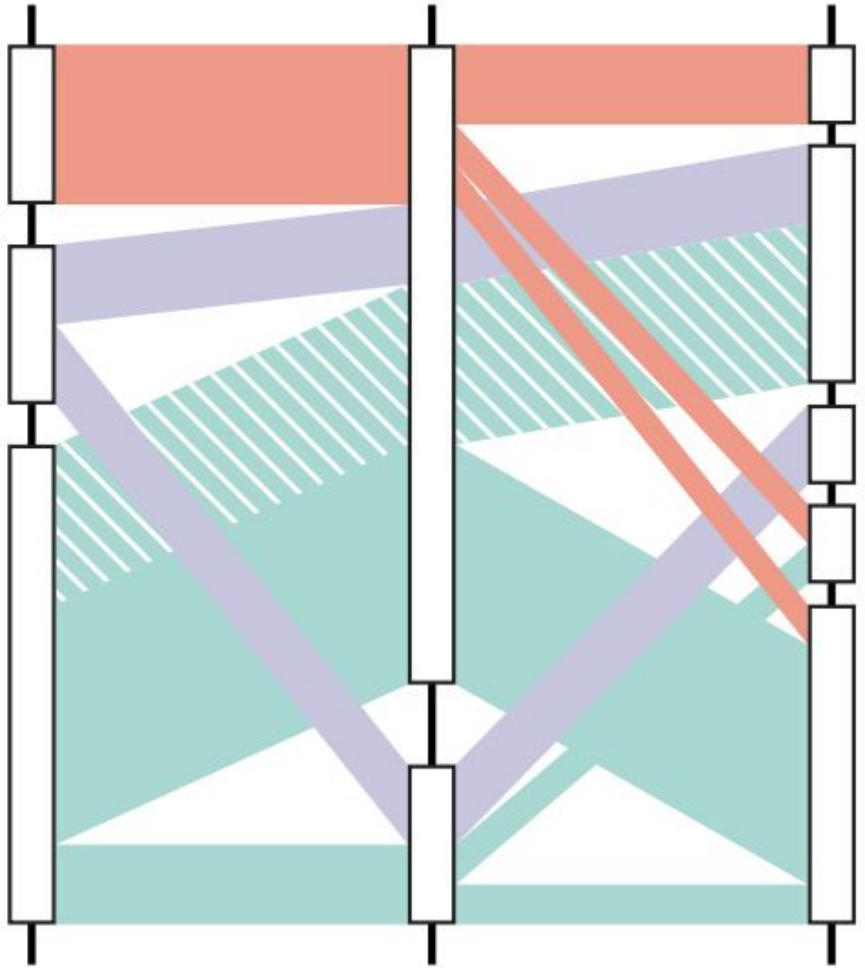
Radial coordinates / Radar chart

- When axes have different scales, values are not comparable between axes.

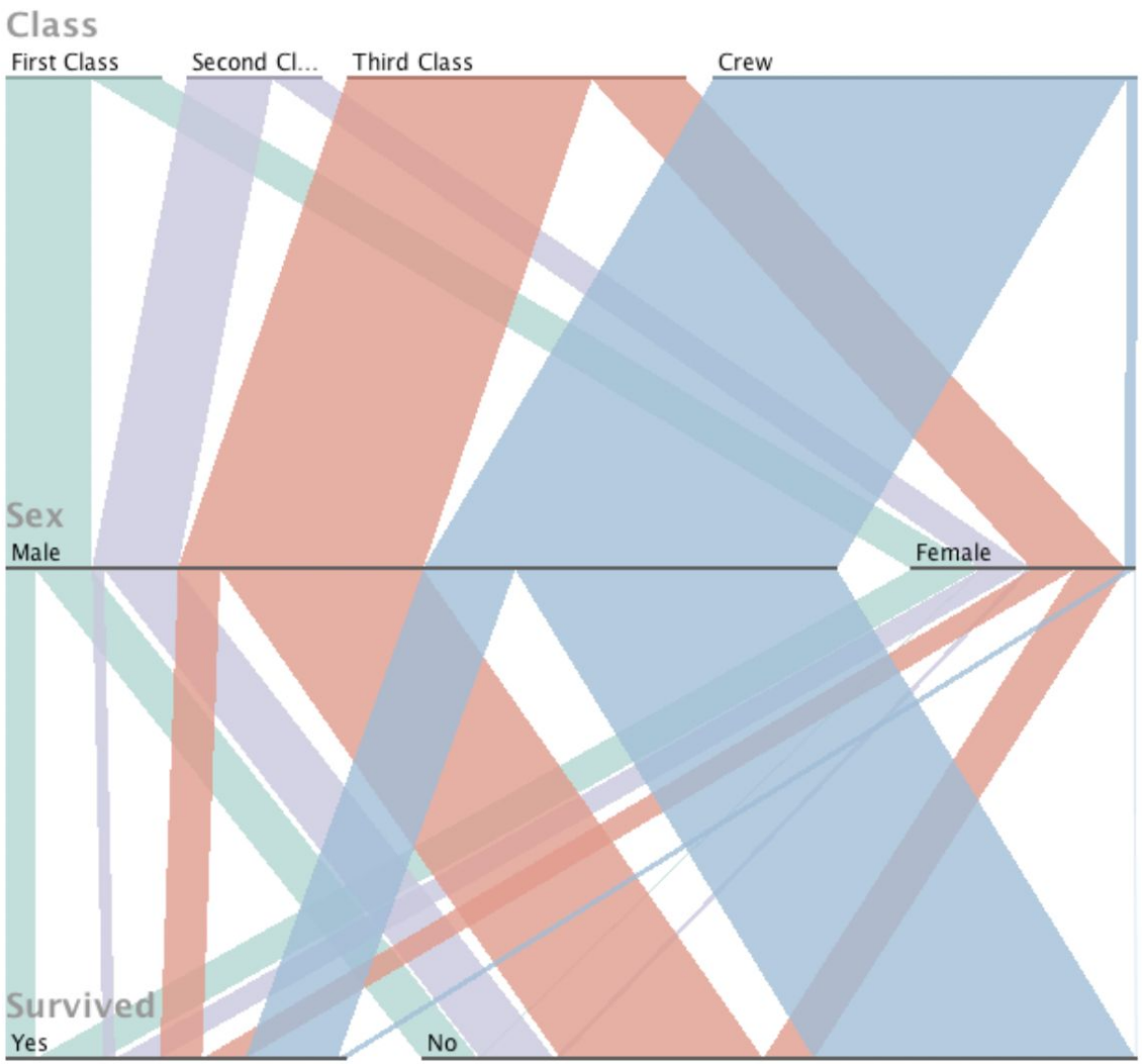


Parallel sets for categorical dimensions

Show data frequencies rather than individual data points.



Titanic Survivors

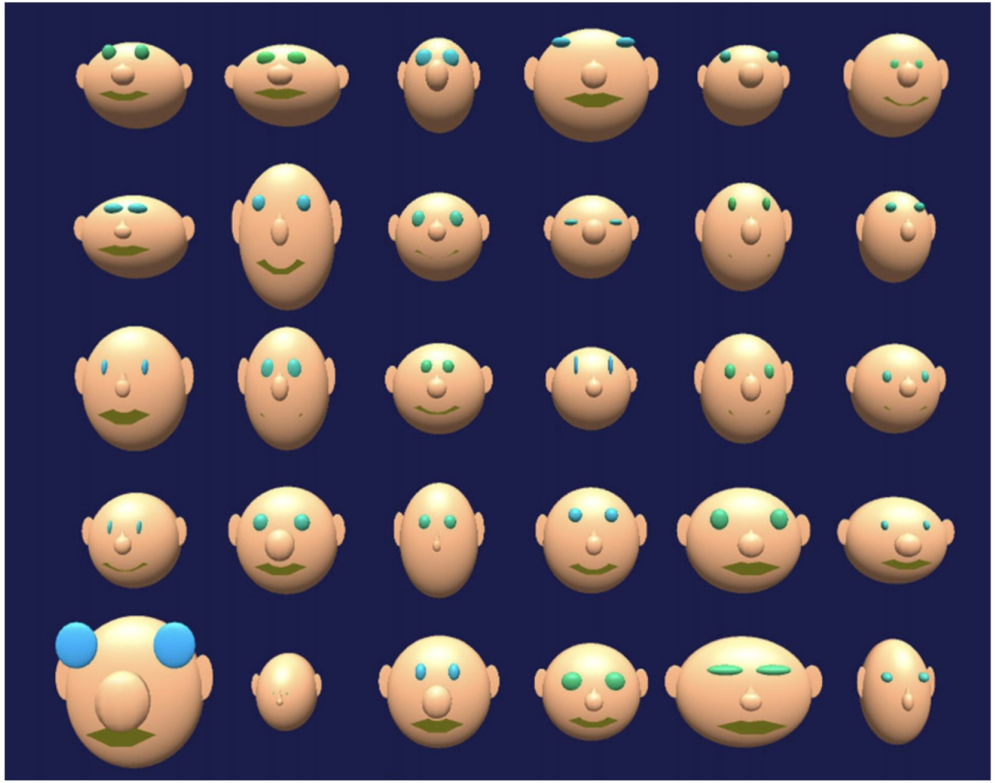


Chernoff faces

Invented by Herman Chernoff in 1973.

Encode different variables to different facial features, like the shape of the head and the size of the eye.

Assumption: object-like appearance is useful for perception.

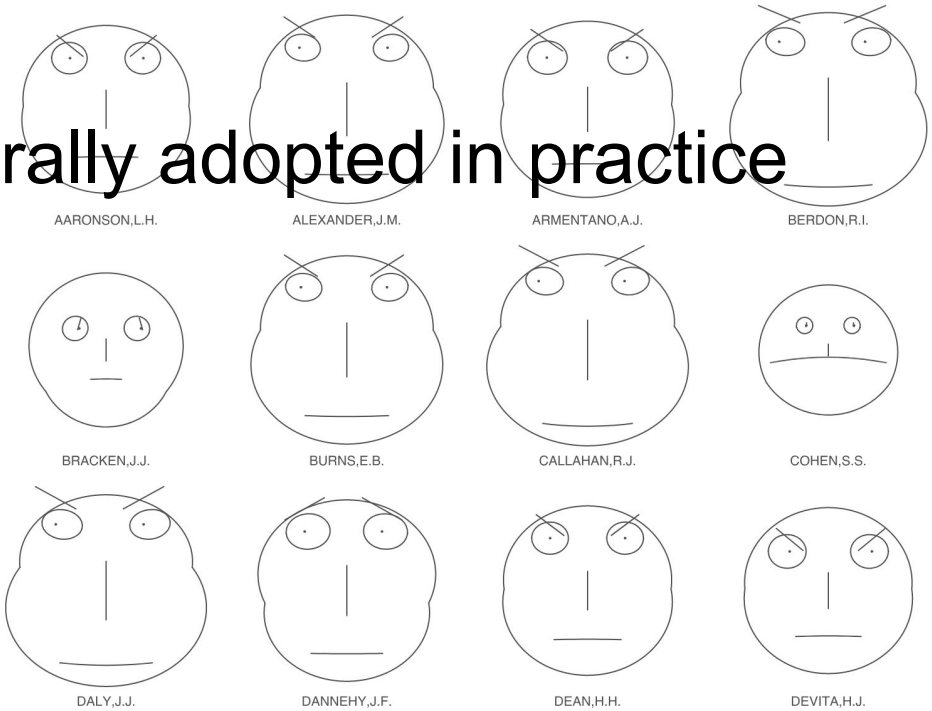


Unpredictable effects from emergent expressions.

Chernoff faces -- Not generally adopted in practice

Human differences in sensitivities to different features make the perceptual spaces likely to be non-linear.

E.g. eye size and eyebrow-slant were found to carry significant weight.

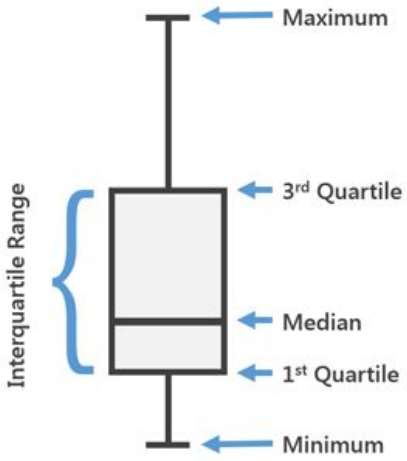


This example shows Chernoff faces for lawyers' ratings of twelve judges.

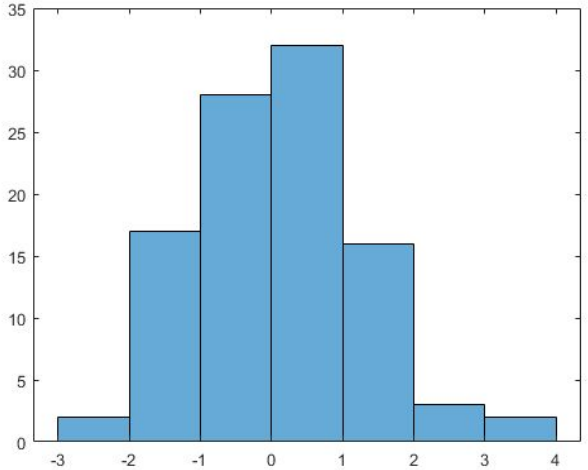
So far - 1D data

word clouds

box plot

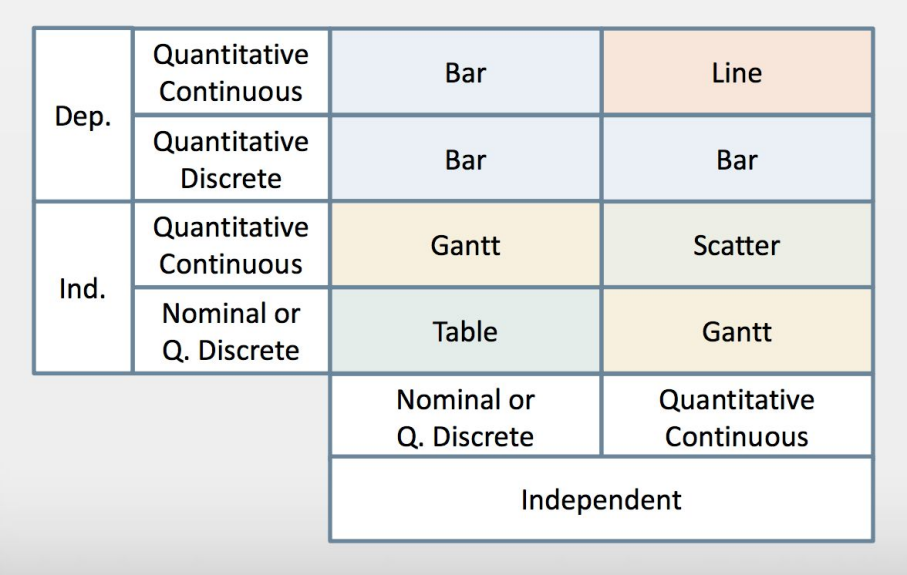


histogram



So far - 2D data

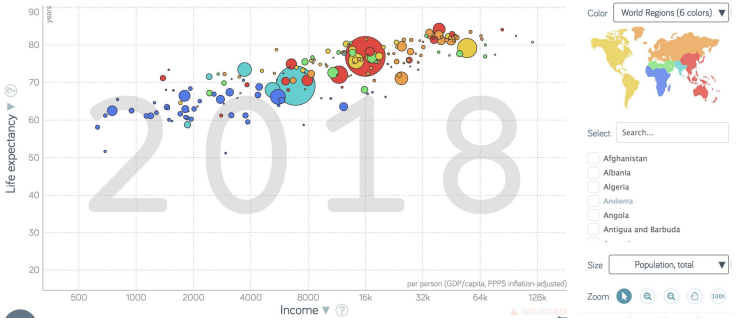
Bar chart, scatter plot, line chart, gantt chart, and table.



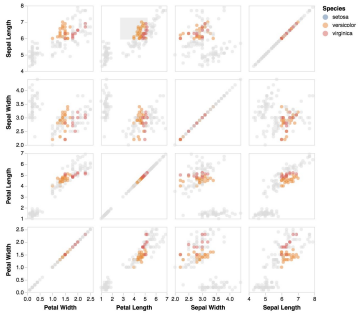
Dep.	Quantitative Continuous	Bar	Line
	Quantitative Discrete	Bar	Bar
Ind.	Quantitative Continuous	Gantt	Scatter
	Nominal or Q. Discrete	Table	Gantt
		Nominal or Q. Discrete	Quantitative Continuous
		Independent	

So far - Multi-dimensional data

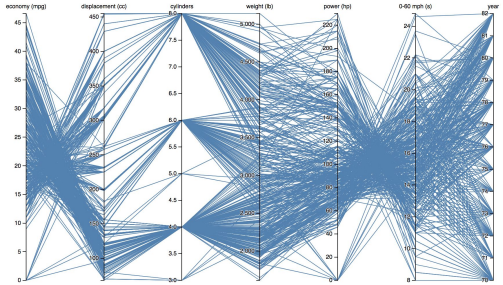
Add additional dimensions on top of 2D charts



Scatter plot matrix

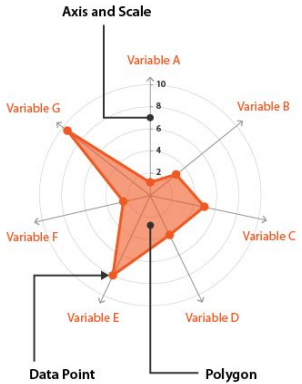


Parallel coordinates

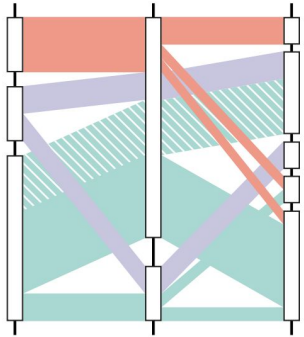


So far - Multi-dimensional data

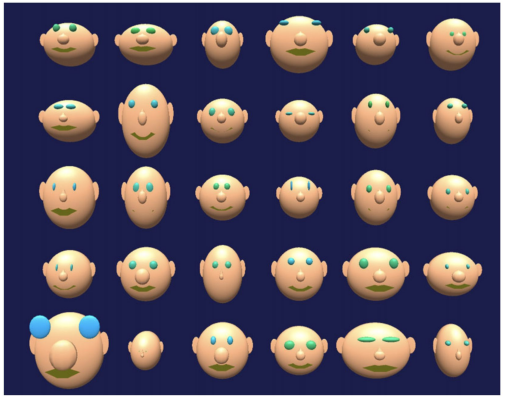
Radar chart



Parallel sets



Chernoff faces



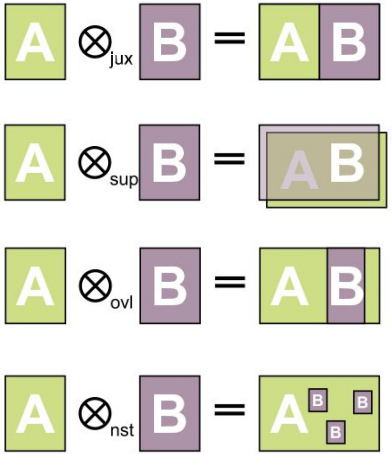
Multi-dimensional data -- Composite visualization

Juxtaposed views

Superimposed views

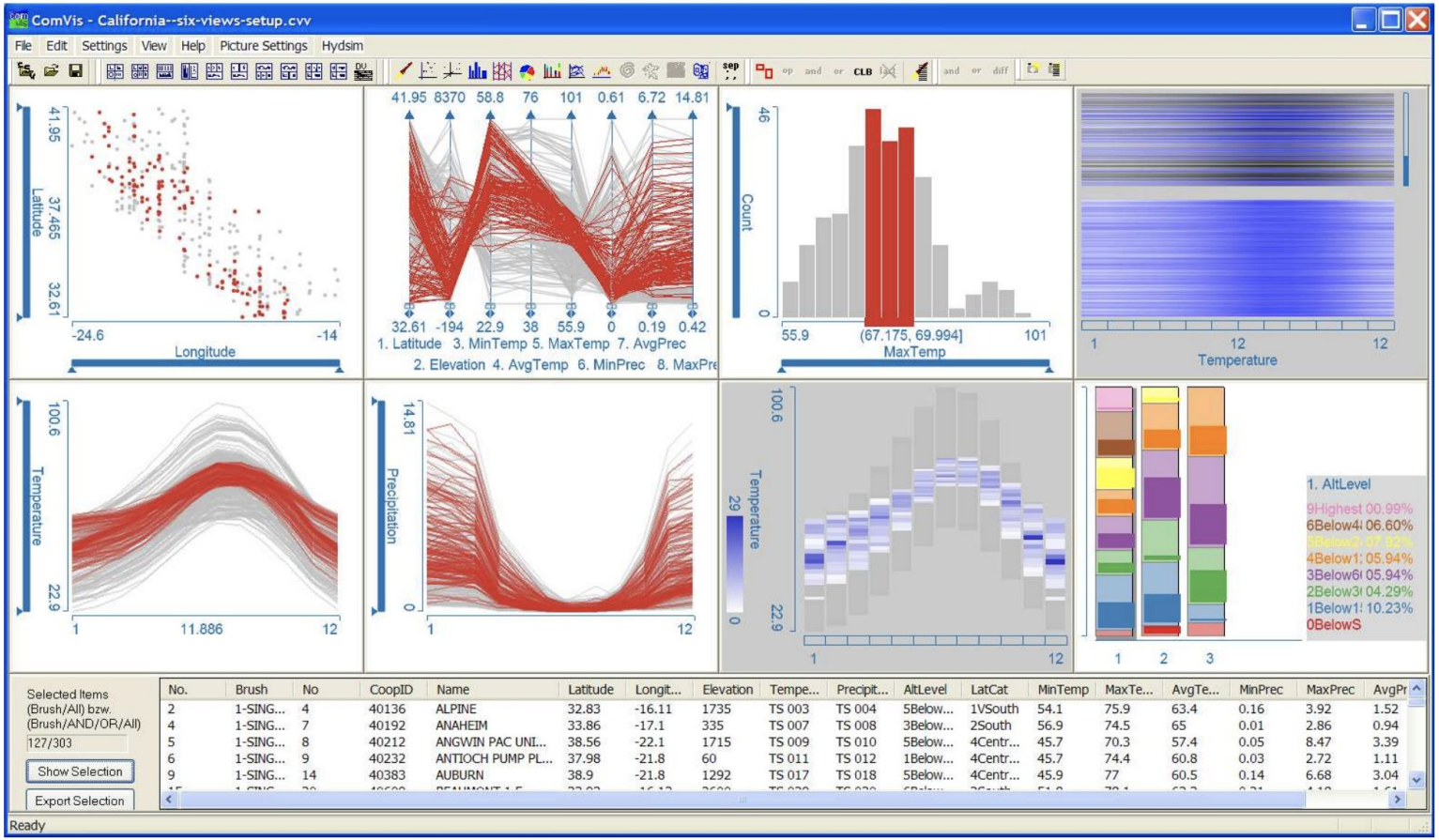
Overloaded views

Nested views



Juxtaposed views (coordinated views) -- ComVis

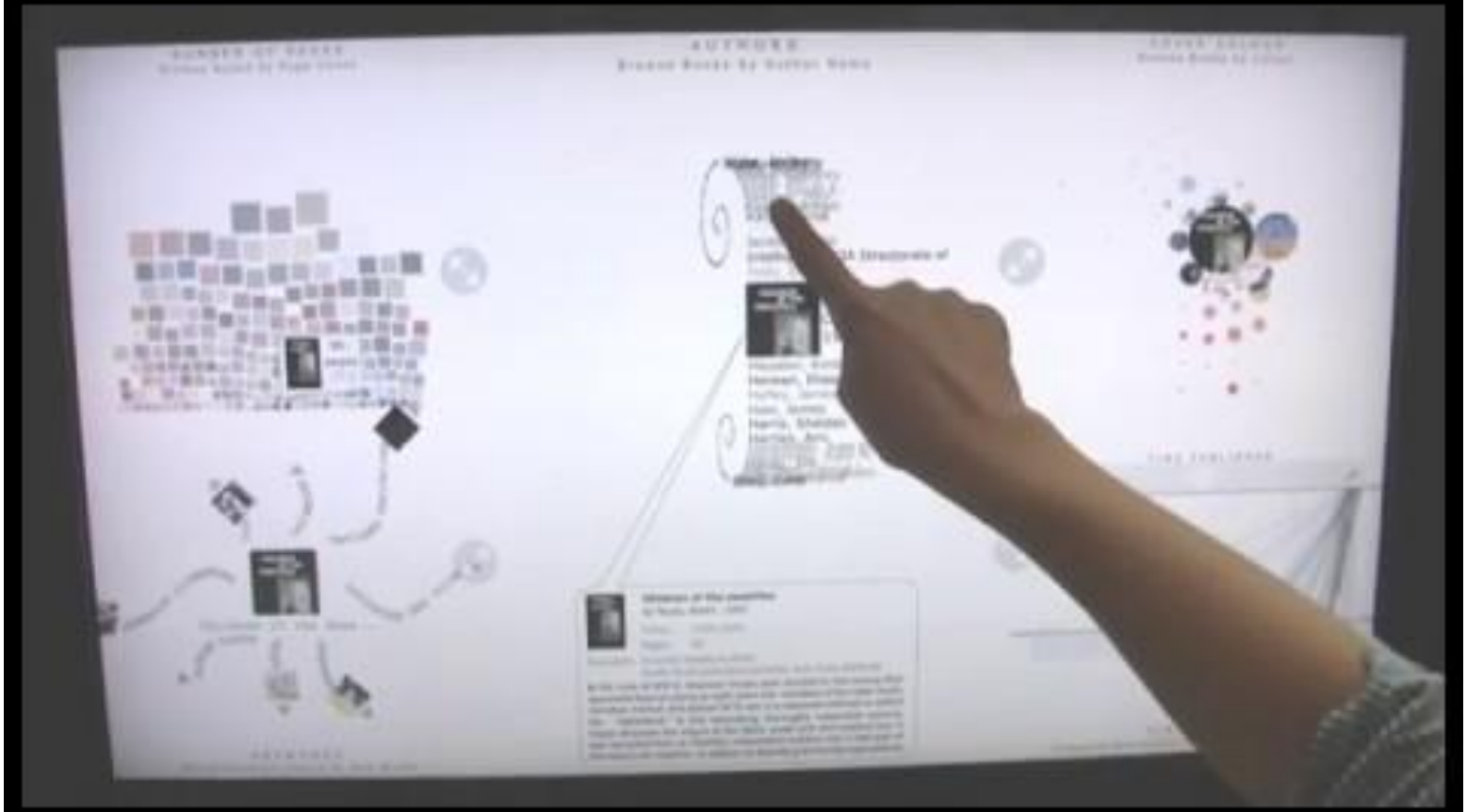
Place visualizations side-by-side in one view.



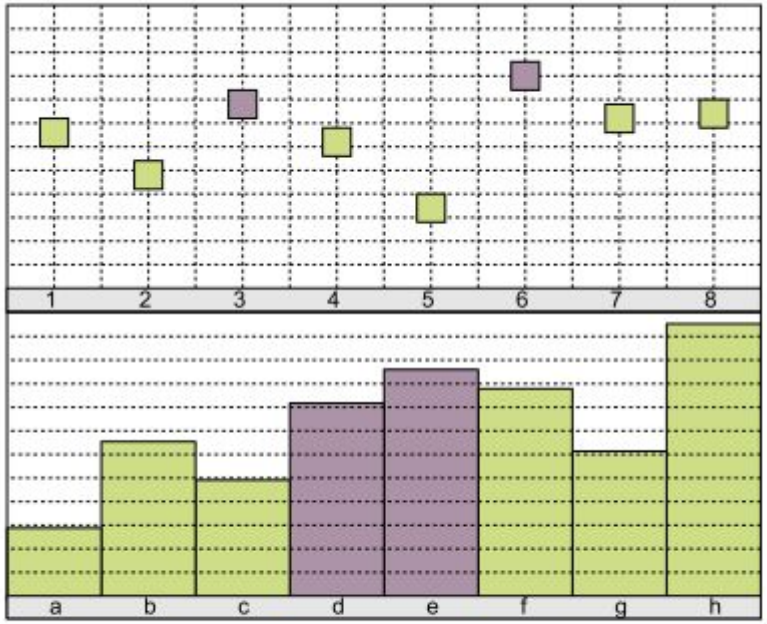
Juxtaposed views (coordinated views) -- Bohemian bookshelf

Coordinate surprising aspects, like cover color and page numbers, to support serendipitous discovery.

The Bohemian bookshelf, Thudt et al., 2012


AUTHORS

Juxtaposed views



Implicit visual linking

Benefits

Little clutter to the resulting display.

Scalable.

Drawback

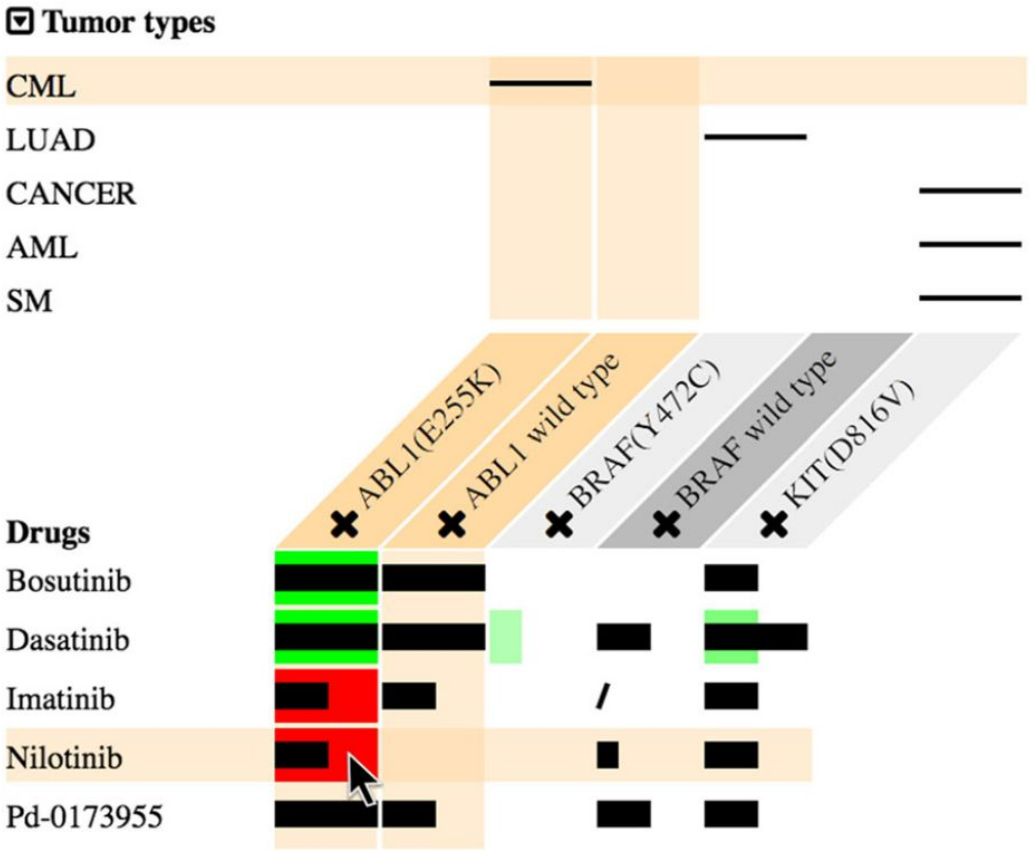
Not easy to see implicit visual links.

Apply for

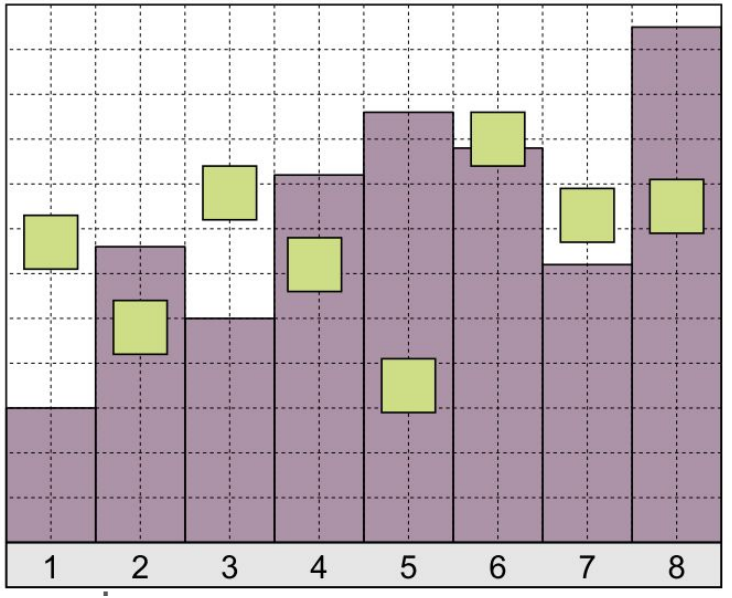
heterogeneous datasets;

where different independent visualizations need to be combined.

Superimposed views -- Overlay two visualizations in one view.



Superimposed views



Benefits

Direct comparison.

Drawbacks

Visual occlusion;

High visual clutter;

Multiple views need to share the same spatial mapping.

Apply for where

comparison is common,

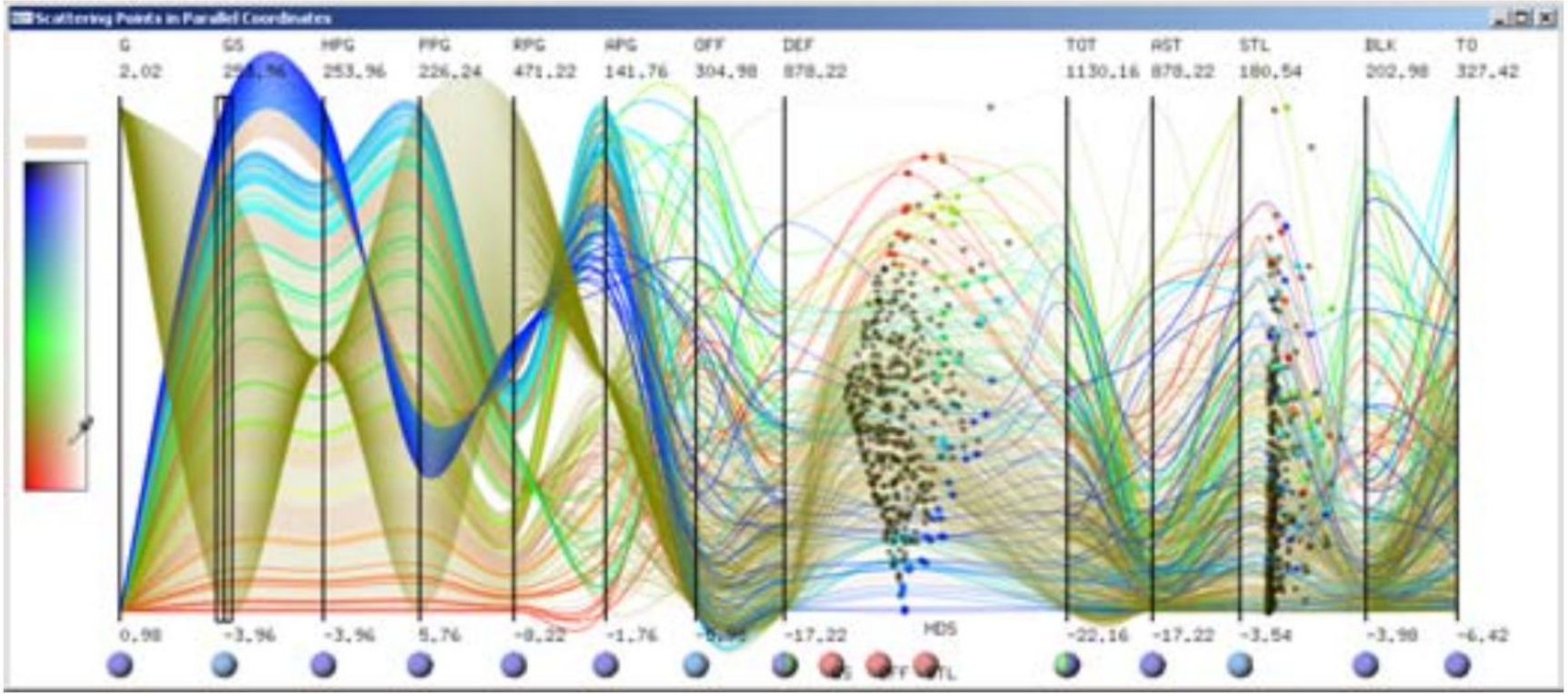
the component views need to be as large as possible (potentially the entire available space).

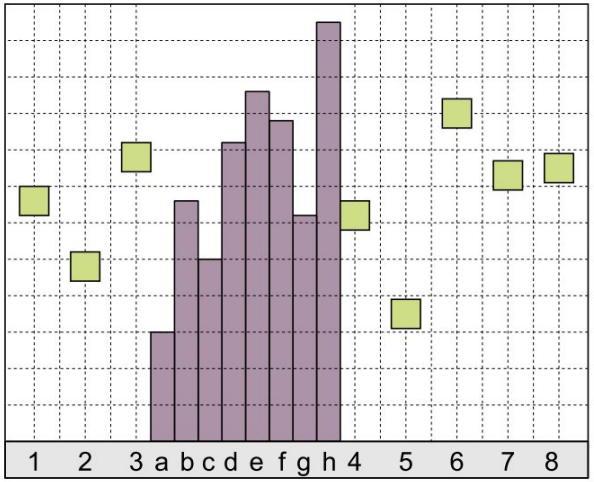
Overloaded views -- Utilize the space of one visualization for another.

Scatting points in parallel coordinates

Parallel coordinates are difficult to comprehend.

Scatter plots are more intuitive to discover correlations.





Overloaded views

Benefits

Views do not have to share the same coordinate space.

More flexibility and control over visual clutter.

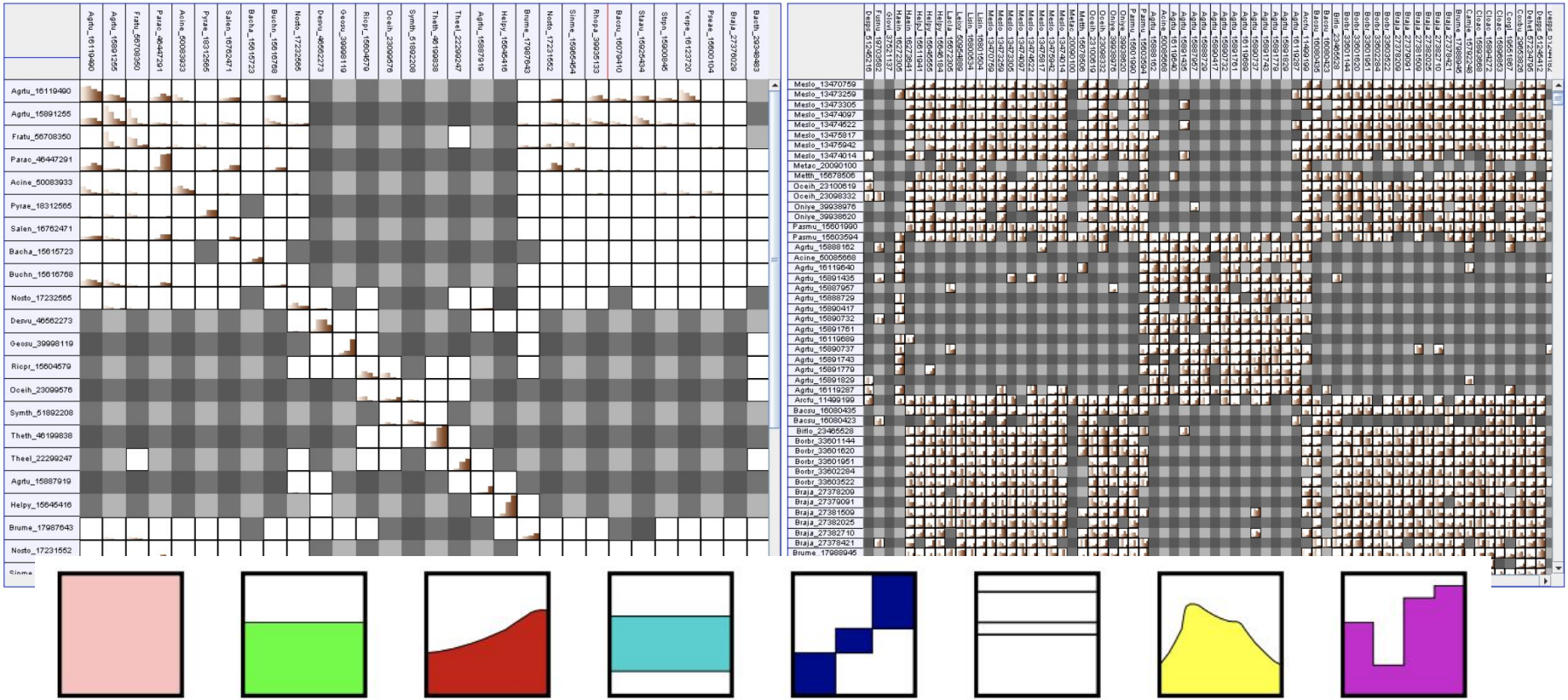
Drawbacks

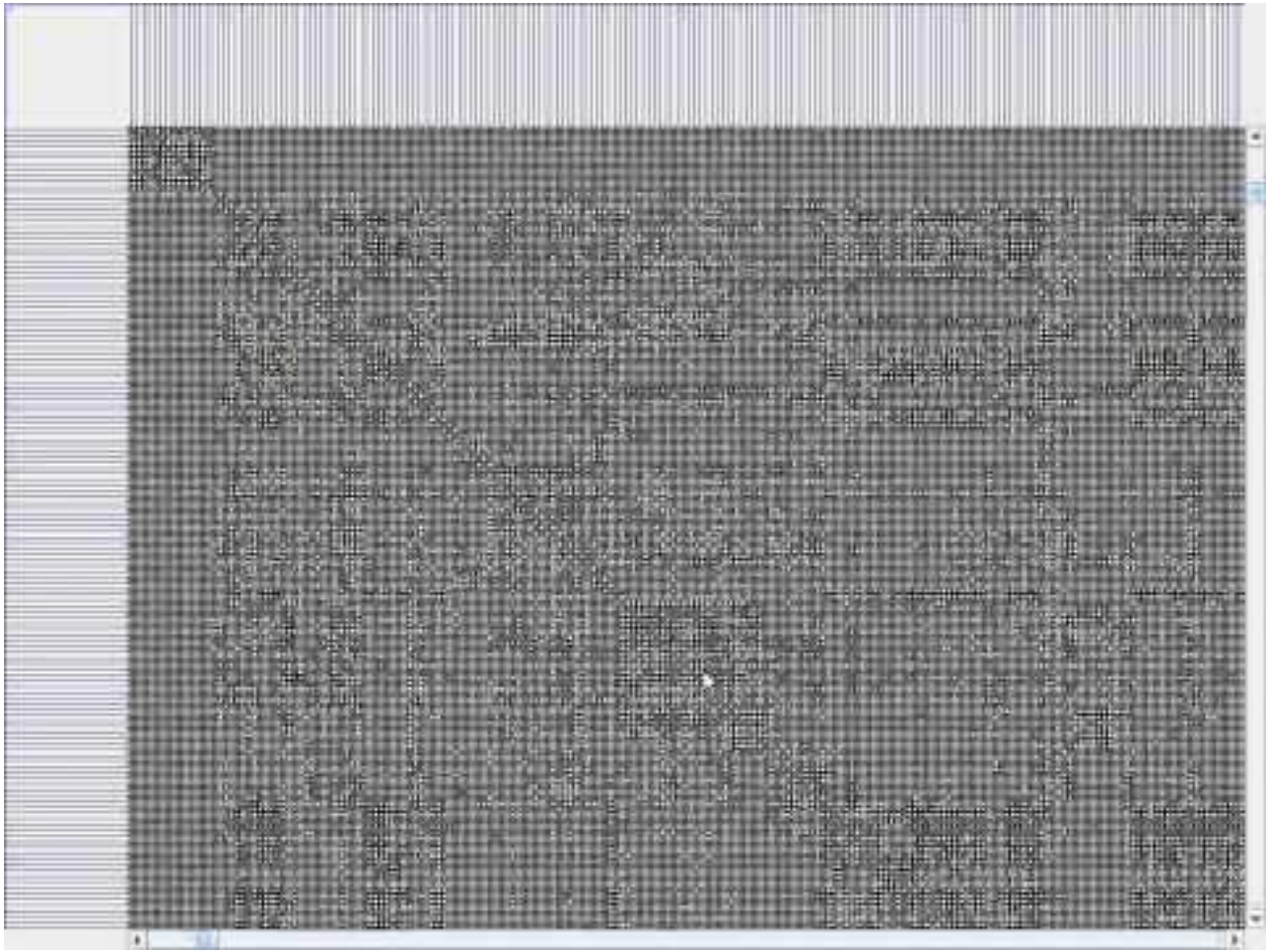
Visual clutter is increased

The visual design dependencies between components are significant.

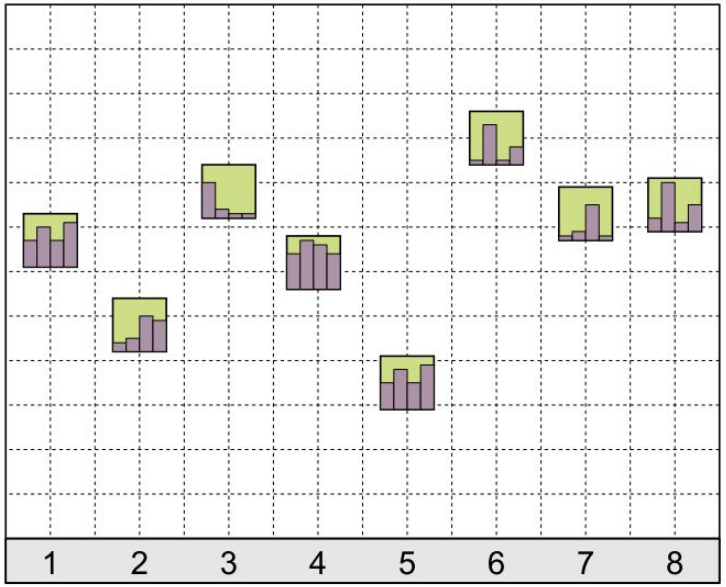
Apply for situations where one visualization can be folded into another to yield a compact (and complex) result.

Nested views -- Nest the content of one visualization inside another visualization.





Nested views



Benefits

- Very compact representation;
- Easy correlation.

Drawbacks

- Limited space for the client visualizations;
- High visual clutter;
- Visual design dependencies are high.

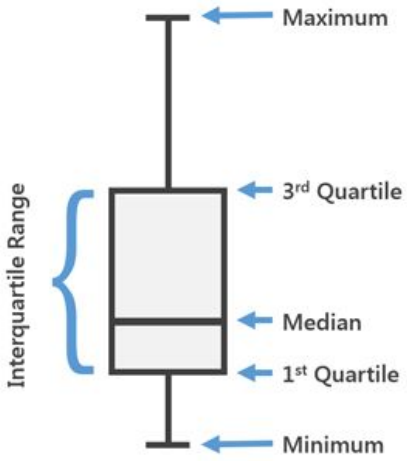
Applications

Call for augmenting a particular visual representation with additional mapping.

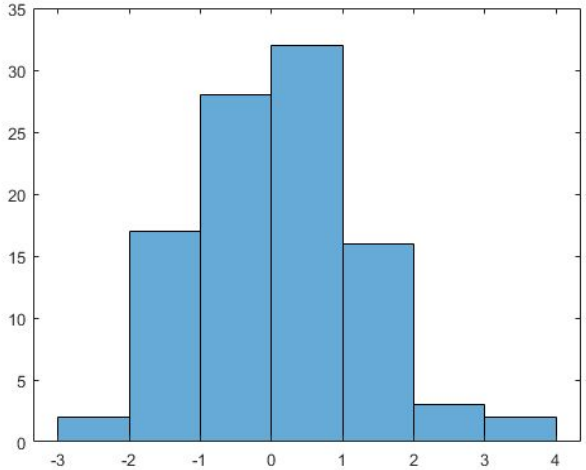
Recap - 1D data

word clouds

box plot



histogram



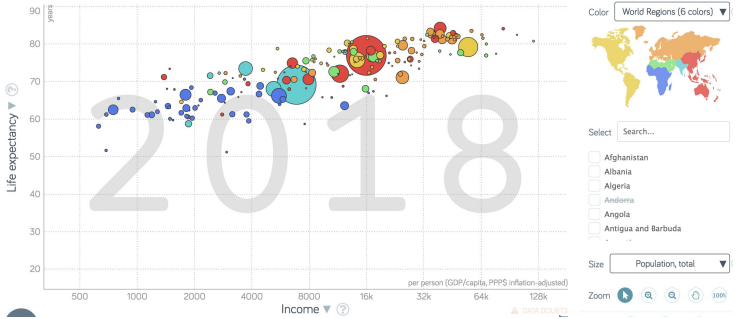
Recap - 2D data

Bar chart, scatter plot, line chart, gantt chart, and table.

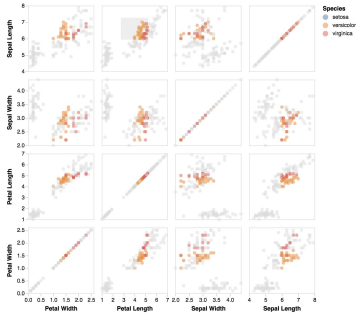
Dep.	Quantitative Continuous	Bar	Line
	Quantitative Discrete	Bar	Bar
Ind.	Quantitative Continuous	Gantt	Scatter
	Nominal or Q. Discrete	Table	Gantt
		Nominal or Q. Discrete	Quantitative Continuous
		Independent	

Recap - Multi-dimensional data

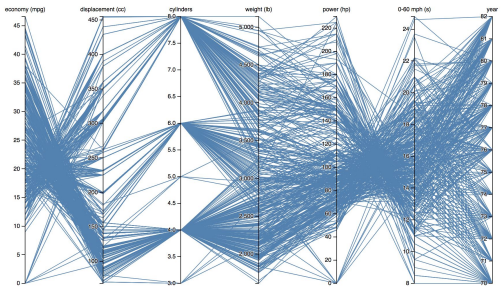
Add additional dimensions on top of 2D charts



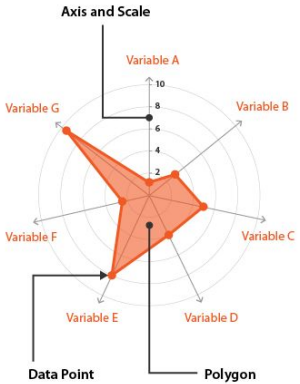
Scatter plot matrix



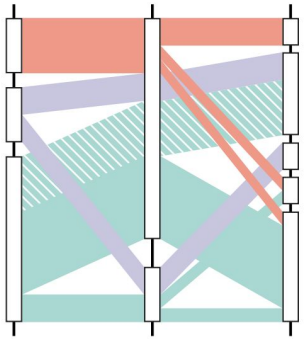
Parallel coordinates



Radar chart

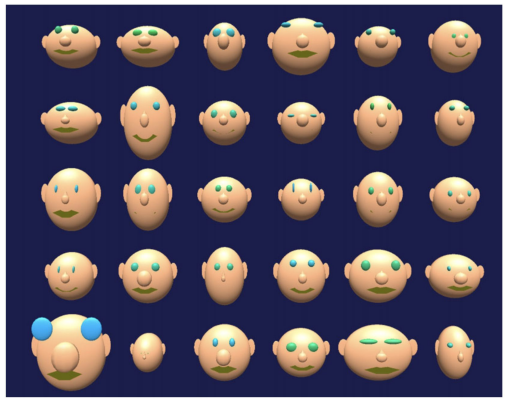


Parallel sets



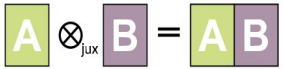
Recap - Multi-dimensional data

Chernoff faces

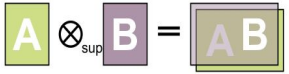


Composite views

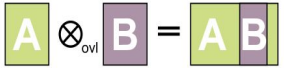
Juxtaposed views



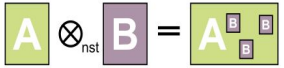
Superimposed views



Overloaded views

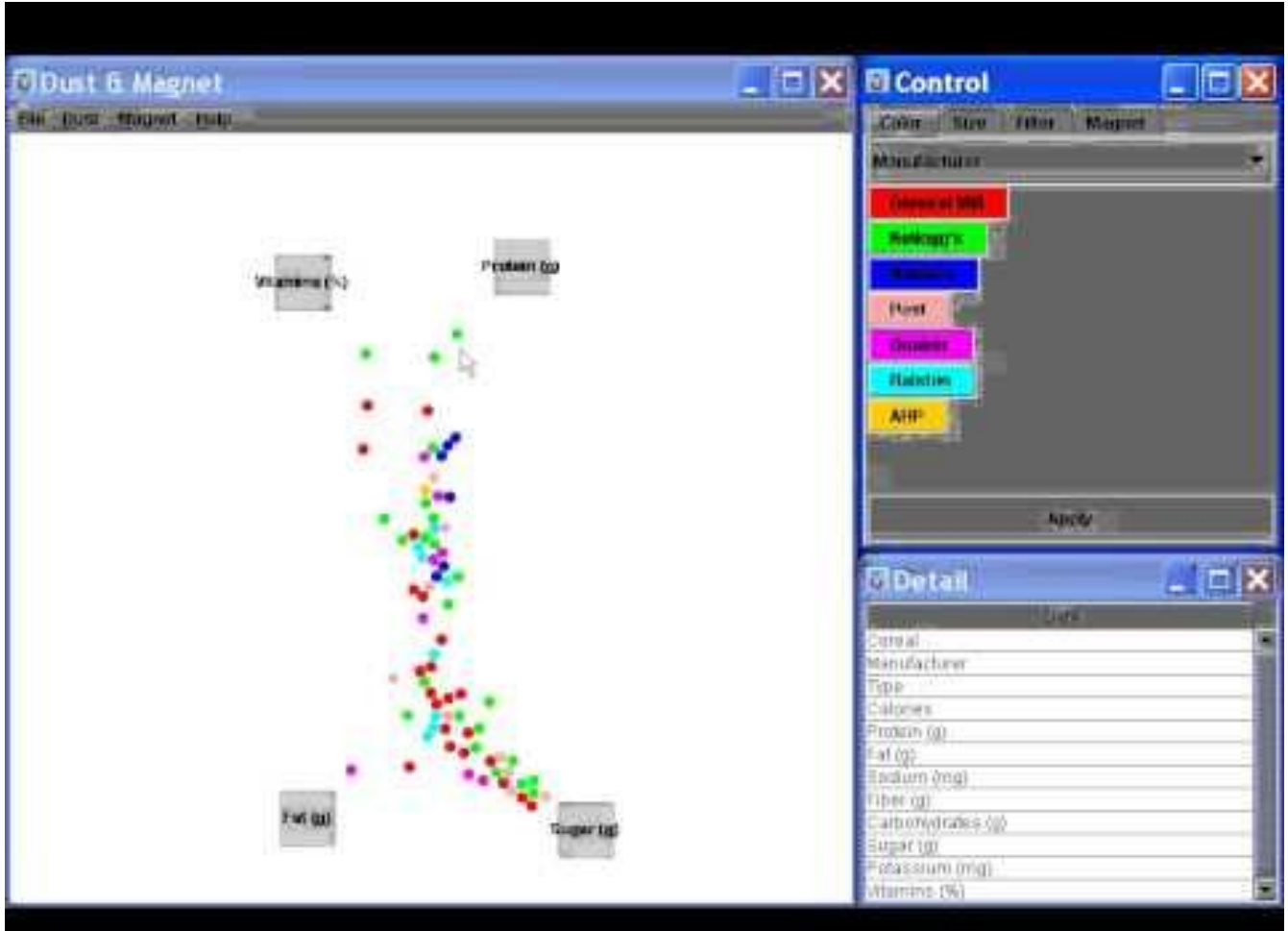


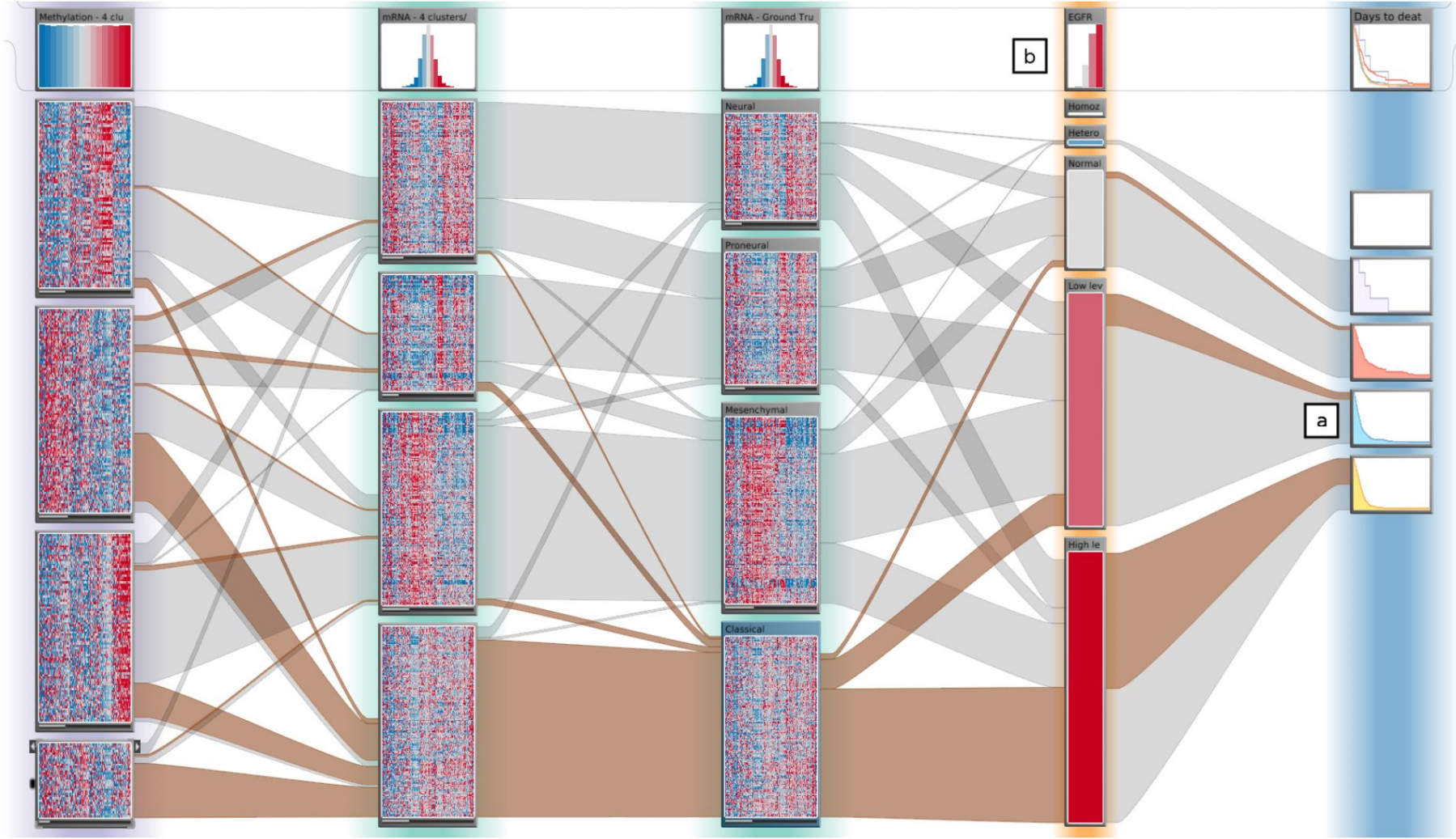
Nested views



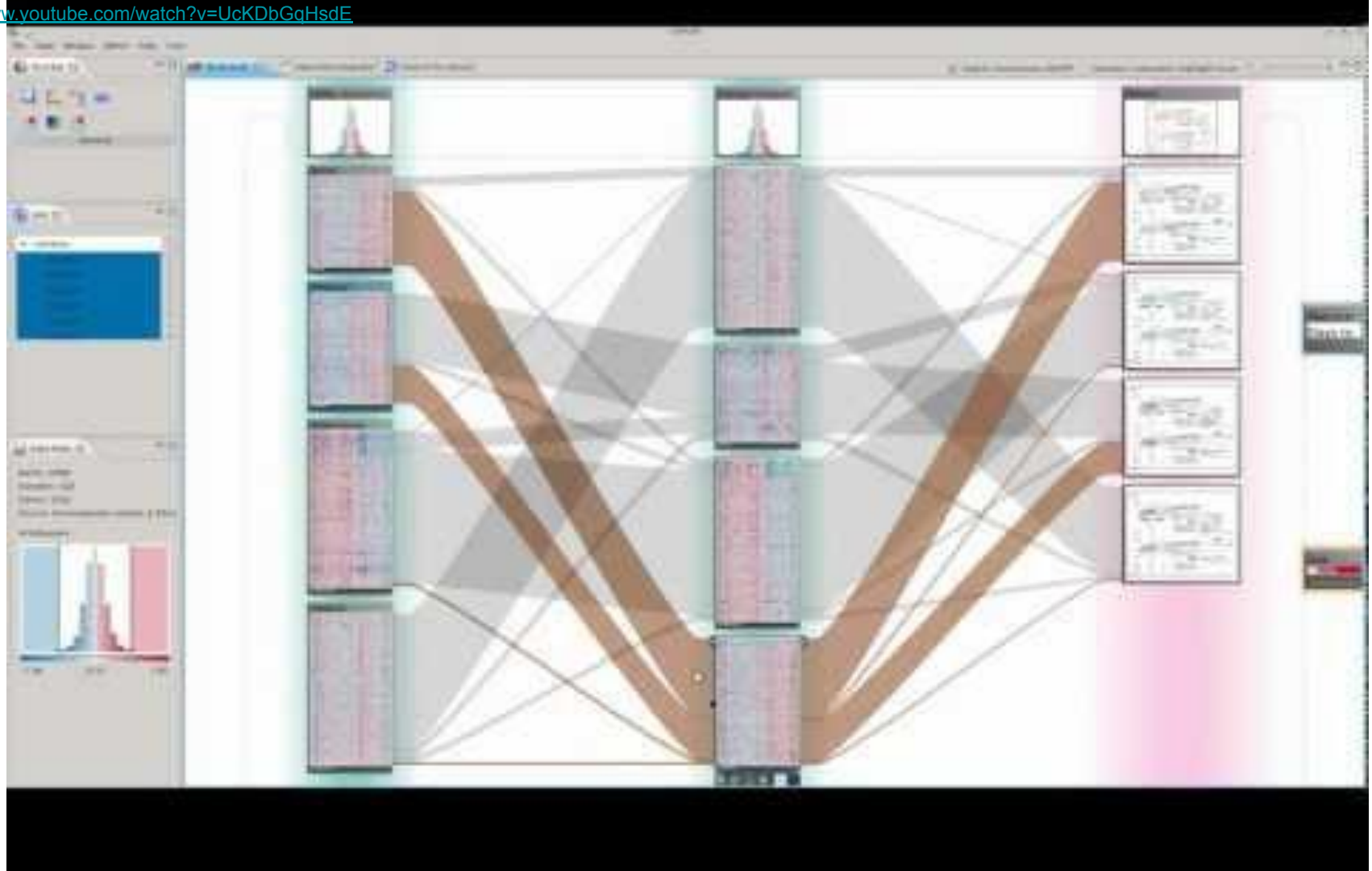
Applications of multi-dimensional data visualizations

Magnet metaphor

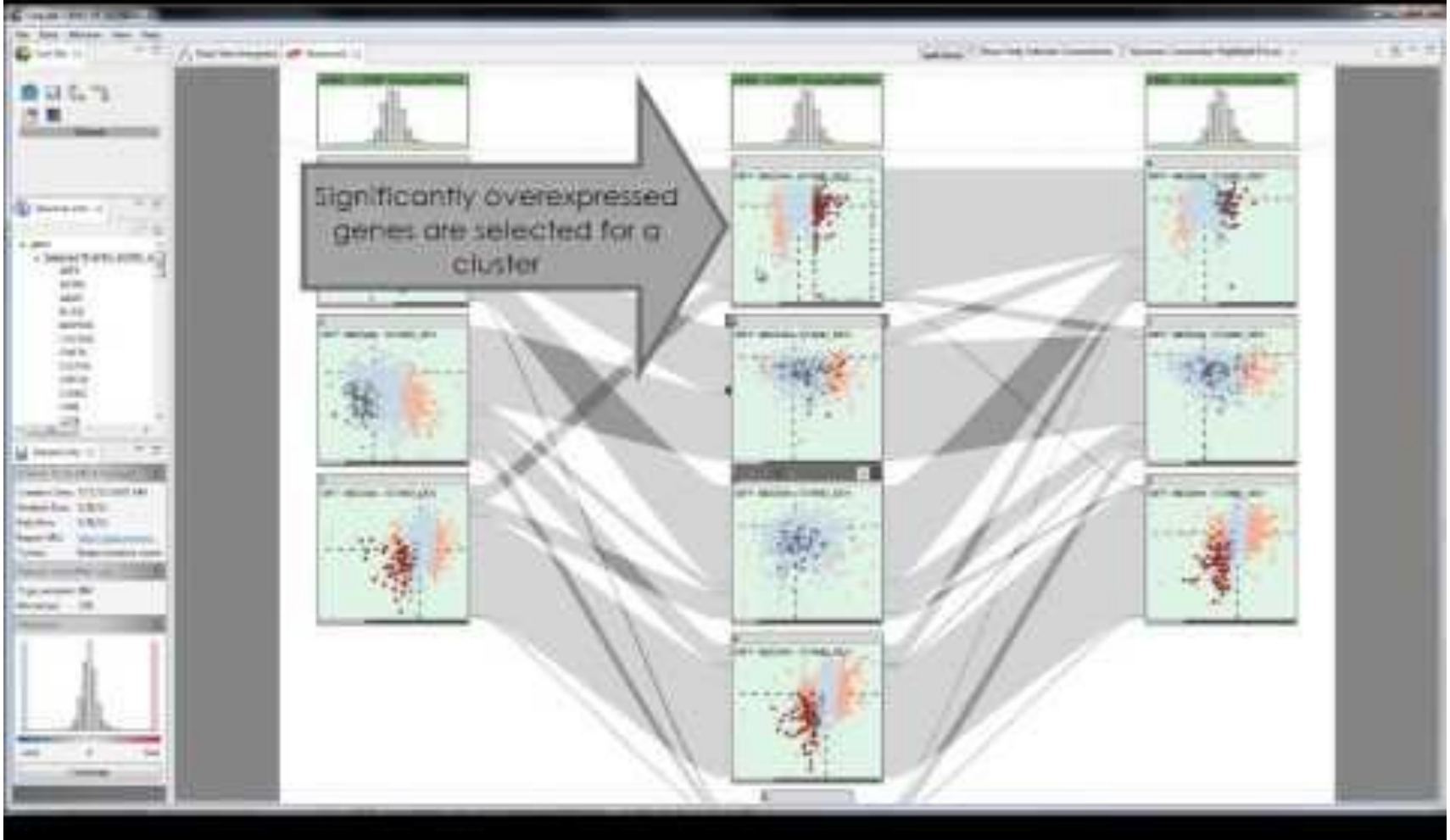



Methylation - 4 clu
mRNA - 4 clusters/
mRNA - Ground Tru
Neural
Proneural
Mesenchymal
Classical
b
EGFR
Homoz
Hetero
Normal
Low lev
High le
Days to deat
a

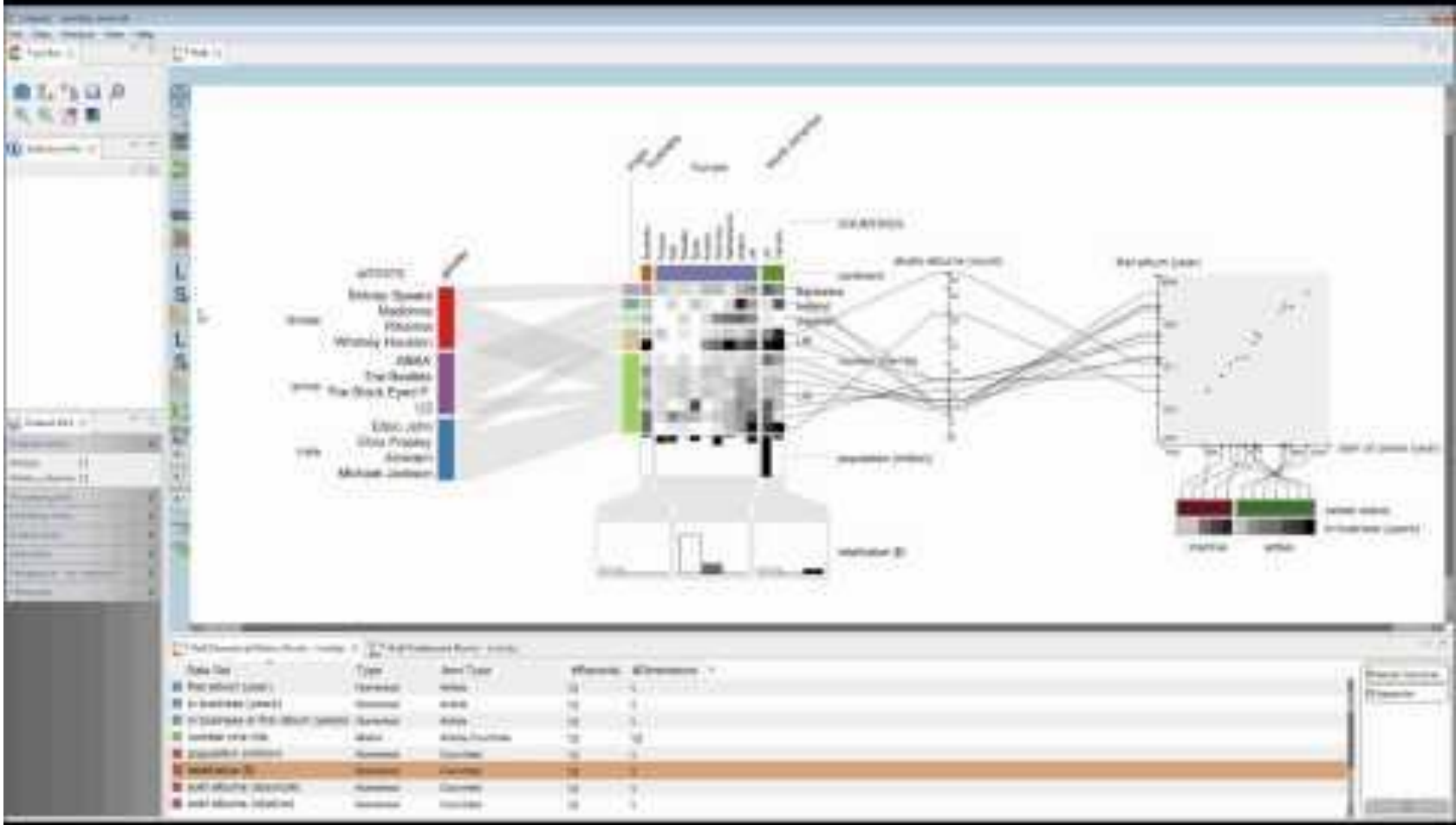
https://www.youtube.com/watch?v=UcKDbGgHsdE



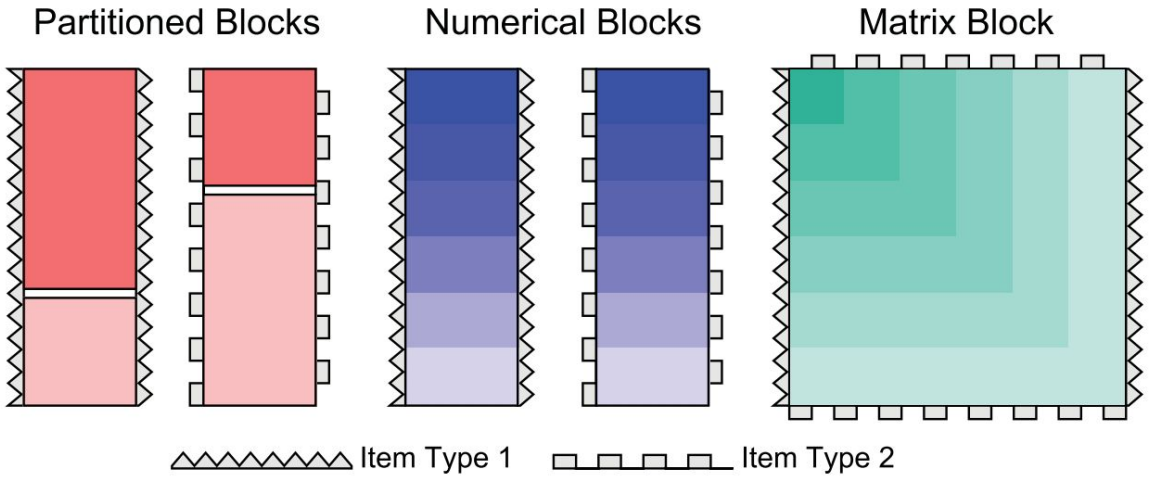
https://www.youtube.com/watch?v=YUSqI2WDJfM

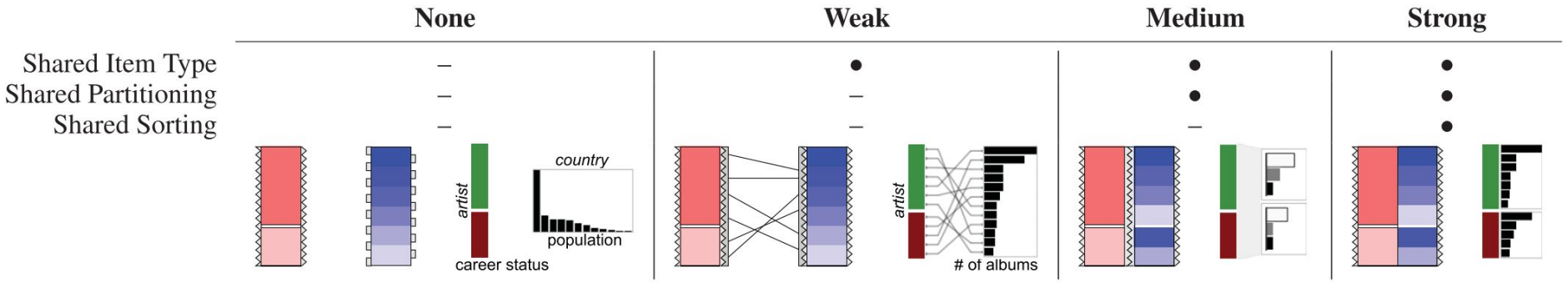


https://www.youtube.com/watch?v=bm59Y8QYbAQ

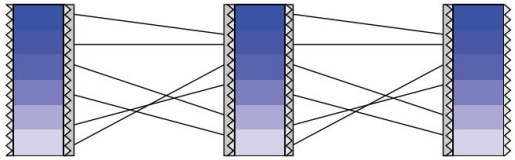


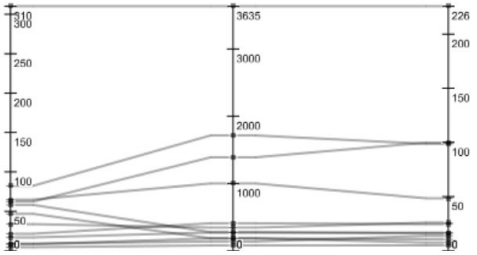
Domino: Relations across multiple tabular datasets.



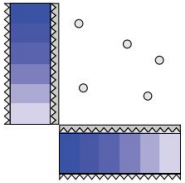


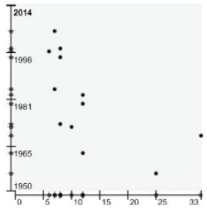
(a) *Parallel Coordinate Plot [14]*



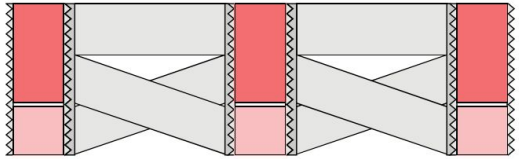


(b) *Scatterplot*

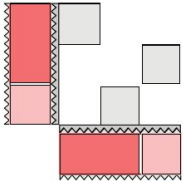


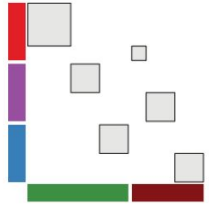


(c) *Parallel Sets [18]*

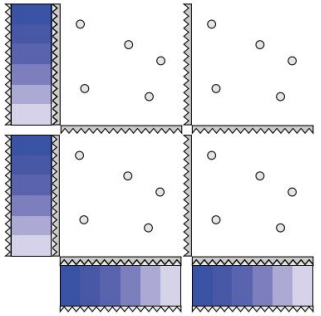


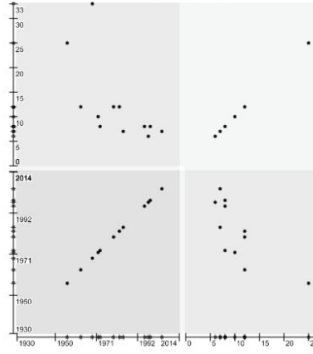
(d) *Mosaic Plot [11]*



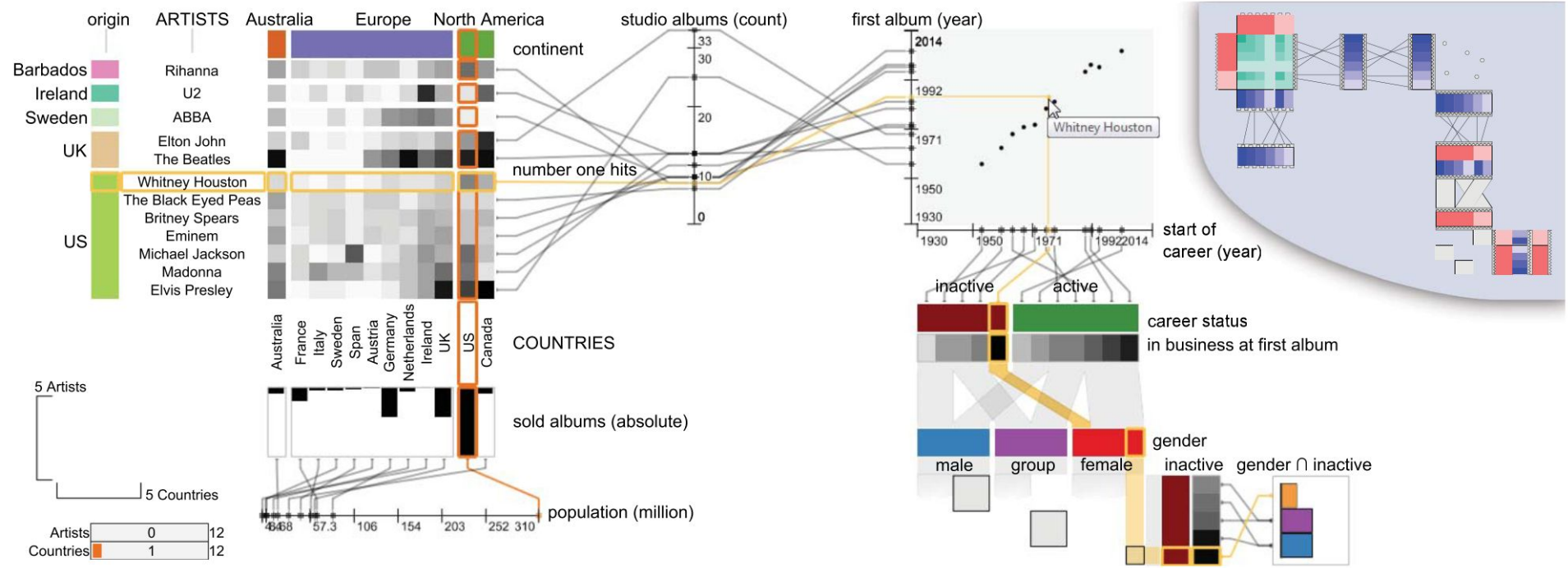


(f) *Scatterplot Matrix (SPLOM) [4]*

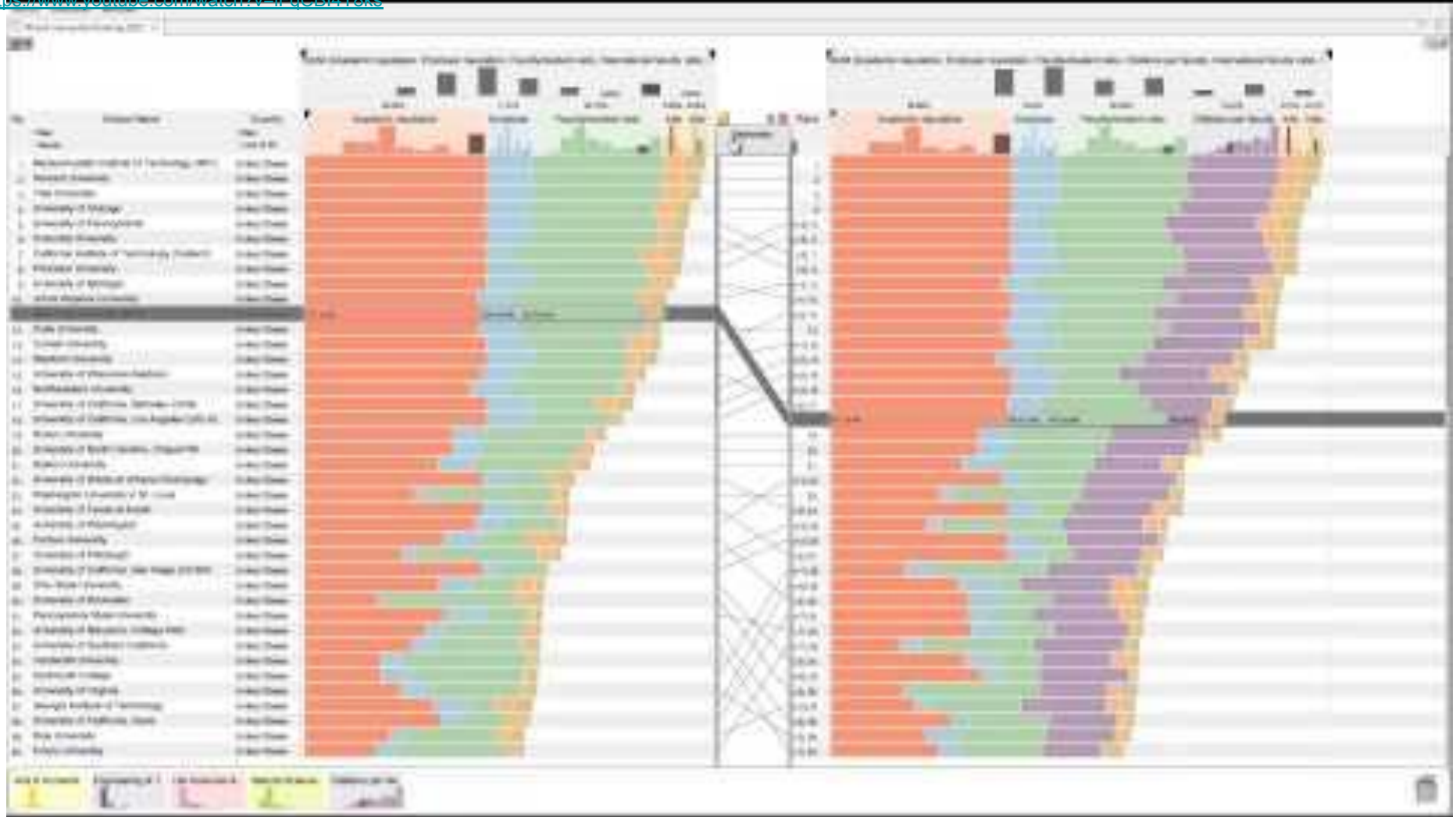




Domino: Relations across multiple tabular datasets


origin
ARTISTS
Australia
Europe
North America
Barbados
Ireland
Sweden
UK
US
Rihanna
U2
ABBA
Elton John
The Beatles
Whitney Houston
The Black Eyed Peas
Britney Spears
Eminem
Michael Jackson
Madonna
Elvis Presley
continent
number one hits
COUNTRIES
Australia
France
Italy
Sweden
Span
Austria
Germany
Netherlands
Ireland
UK
US
Canada
sold albums (absolute)
population (million)
5 Artists
5 Countries
Artists 0 12
Countries 1 12
studio albums (count)
33
30
20
10
0
first album (year)
2014
1992
1971
1950
1930
Whitney Houston
start of career (year)
inactive
active
career status
in business at first album
gender
male
group
female
inactive
gender ∩ inactive

https://www.youtube.com/watch?v=iFqCBI4T8ks



References

Kosara, R., 2011, January. Indirect multi-touch interaction for brushing in parallel coordinates. In Visualization and Data Analysis 2011 (Vol. 7868, p. 786809). International Society for Optics and Photonics.

Matkovic, K., Freiler, W., Gracanin, D. and Hauser, H., 2008, July. ComVis: A coordinated multiple views system for prototyping new visualization technology. In *Information Visualisation, 2008. IV'08. 12th International Conference* (pp. 215-220). IEEE.

Javed, W. and Elmqvist, N., 2012, February. Exploring the design space of composite visualization. In *Visualization Symposium (PacificVis), 2012 IEEE Pacific* (pp. 1-8). IEEE.

N. Elmqvist, T.-N. Do, H. Goodell, N. Henry, and J.-D. Fekete. ZAME: Interactive large-scale graph visualization. In Proceedings of the IEEE Pacific Symposium on Visualization, pages 215–222, 2008.

Kosara, R., Bendix, F. and Hauser, H., 2006. Parallel sets: Interactive exploration and visual analysis of categorical data. IEEE transactions on visualization and computer graphics, 12(4), pp.558-568.

Yuan, X., Guo, P., Xiao, H., Zhou, H. and Qu, H., 2009. Scattering points in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, *15*(6), pp.1001-1008.

http://caleydo.org/