Visualization techniques for multidimensional data

Maria Cristina
SCC0652 Visualização Computacional

Visual Mapping

https://www.data-to-viz.com/

What kind of data do you have?

Which techniques available?

data vs visualization technique

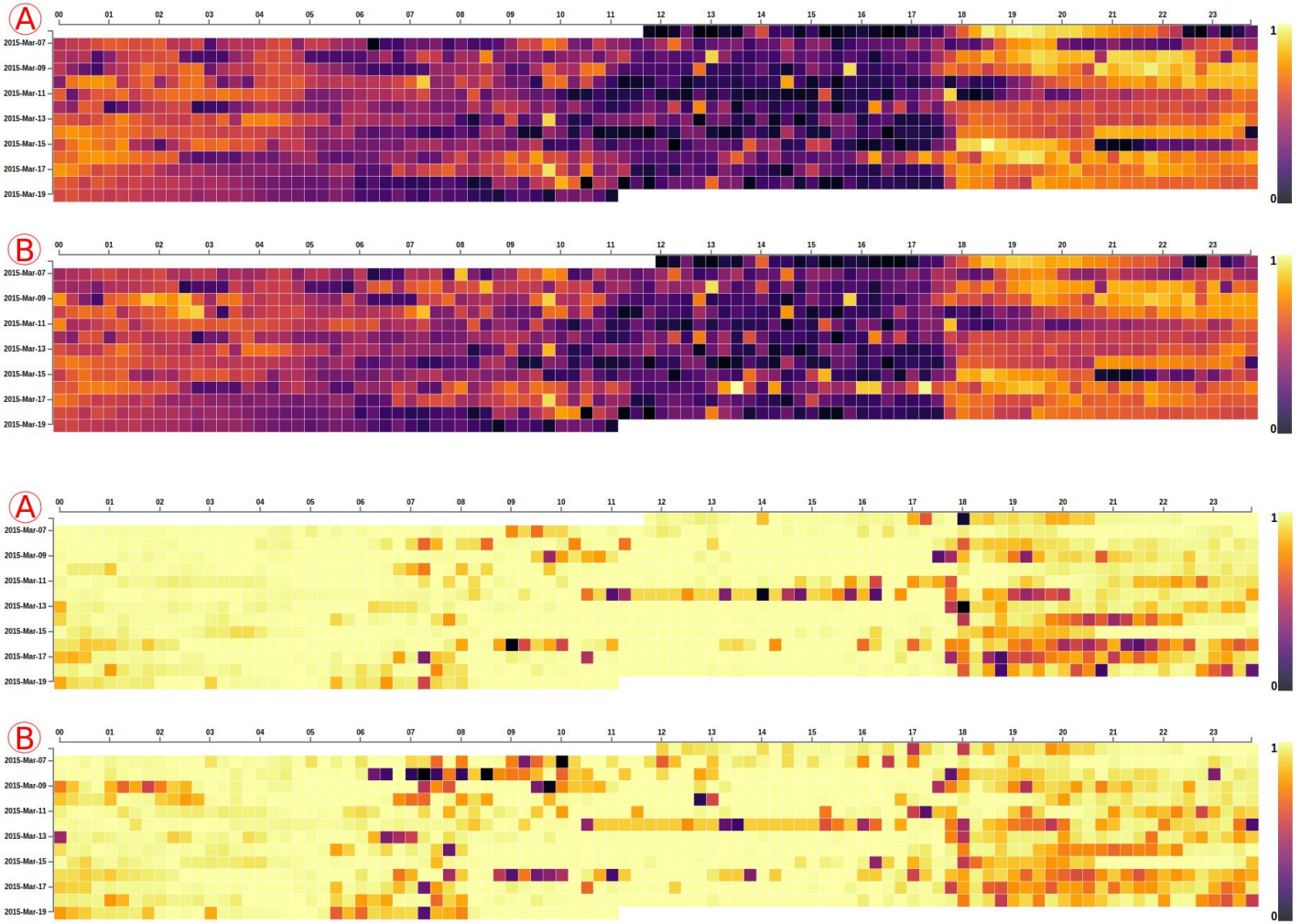
What kind of data do you have? Pick the main type using the buttons below. Then let the decision tree guide you toward your graphic possibilities.

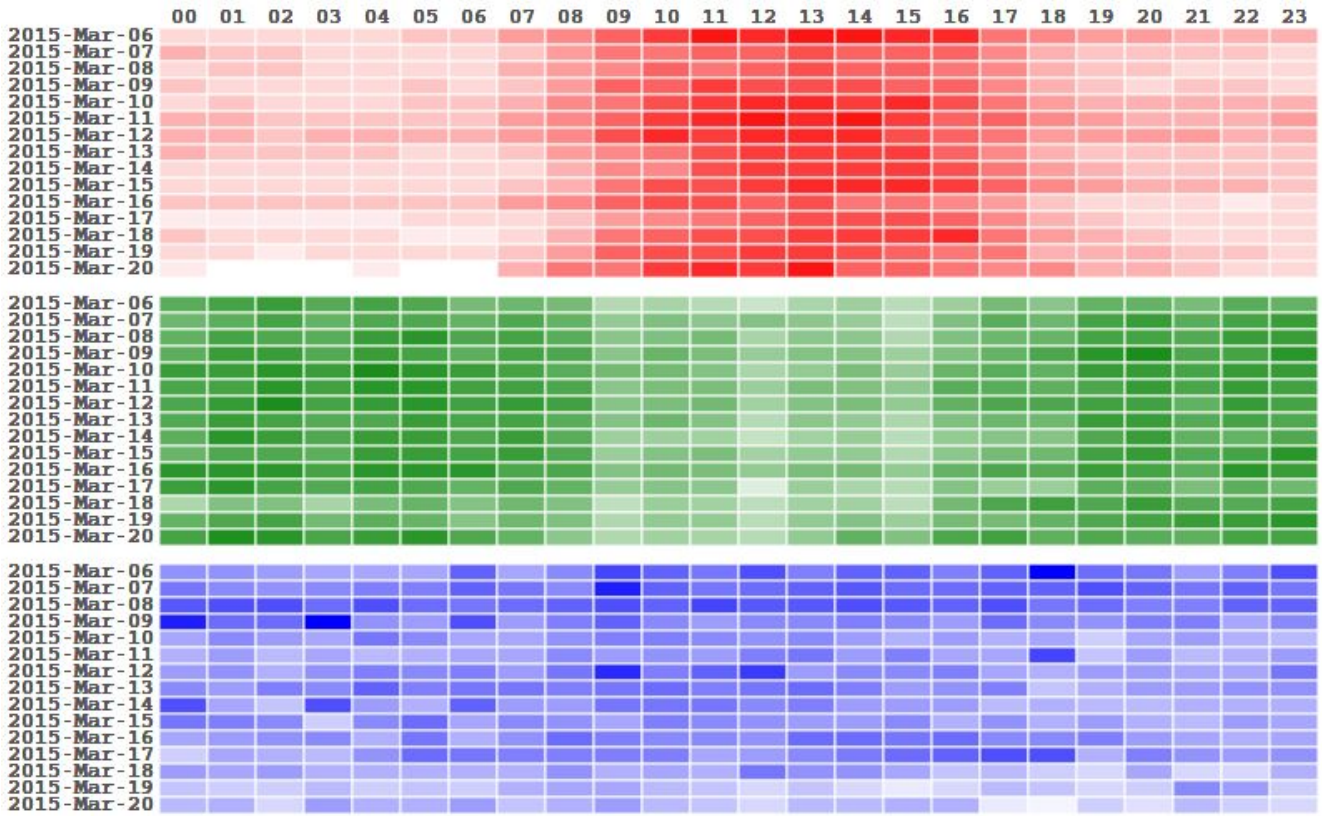
Numeric | Categoric | **Num & Cat** | Maps | Network | Time series

A WORLD OF POSSIBILITIES

Here is an overview of all the graph types presented in this website.

Show all | Distribution | Correlation | Ranking | Part of a whole | Evolution | Map | Flow


A
B
A
B
00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23
2015-Mar-07
2015-Mar-09
2015-Mar-11
2015-Mar-13
2015-Mar-15
2015-Mar-17
2015-Mar-19
1
0



Heatmaps can also be used for time series where there is a regular pattern in time

RadViz

https://orange.biolab.si/widget-catalog/visualize/radviz/

RadViz

https://bl.ocks.org/biovisualize/a91f514aaf57eabf8e36

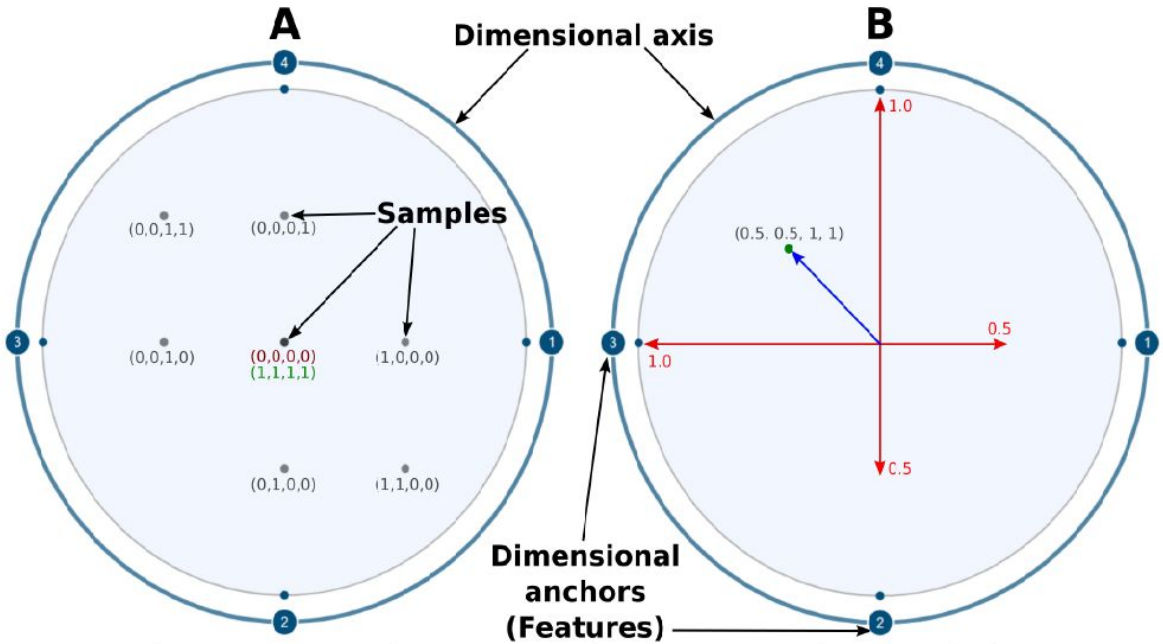


Figure 4 – Radviz visualization with 4 features placed as dimensional anchors along the dimensional axis. (A) Examples of samples with corresponding features array values are positioned inside the visualization. (B) A sample position calculated by the sum of vectors (Red lines), resulting in the vector that gives the sample position in the visualization (Blue line). The final sample position is represented by a green dot.

Source: C.D.G. Reis, Seecology: Data Visualization Framework for Soundscape Ecology Applications. Tese de doutorado, ICMC-USP 2020.

RadViz

http://www.biovisualize.com/2016/03/radviz.html

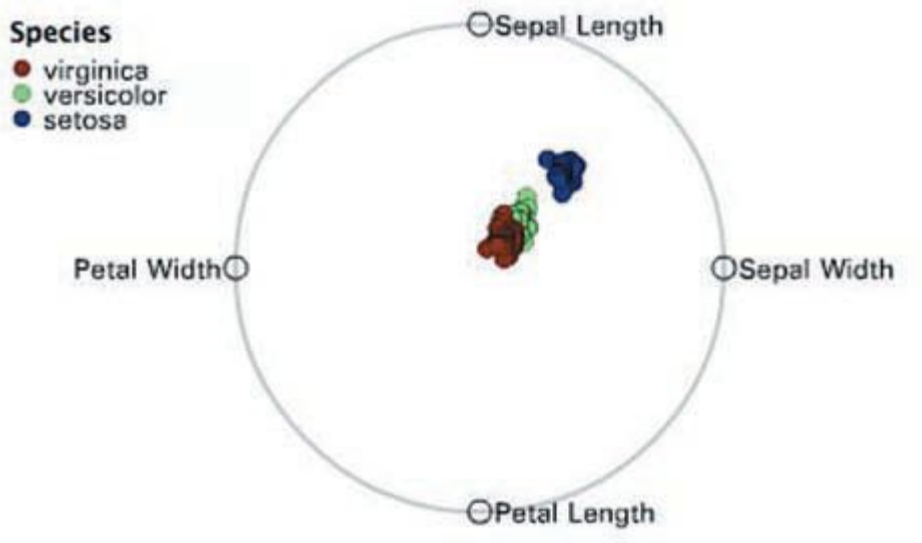


Figure 3. The image of the four-dimensional Iris dataset in the radial visualization RadViz.

Source: Daniels et al. Properties of normalized radial visualizations. Information Visualization 11(4) 273–300.

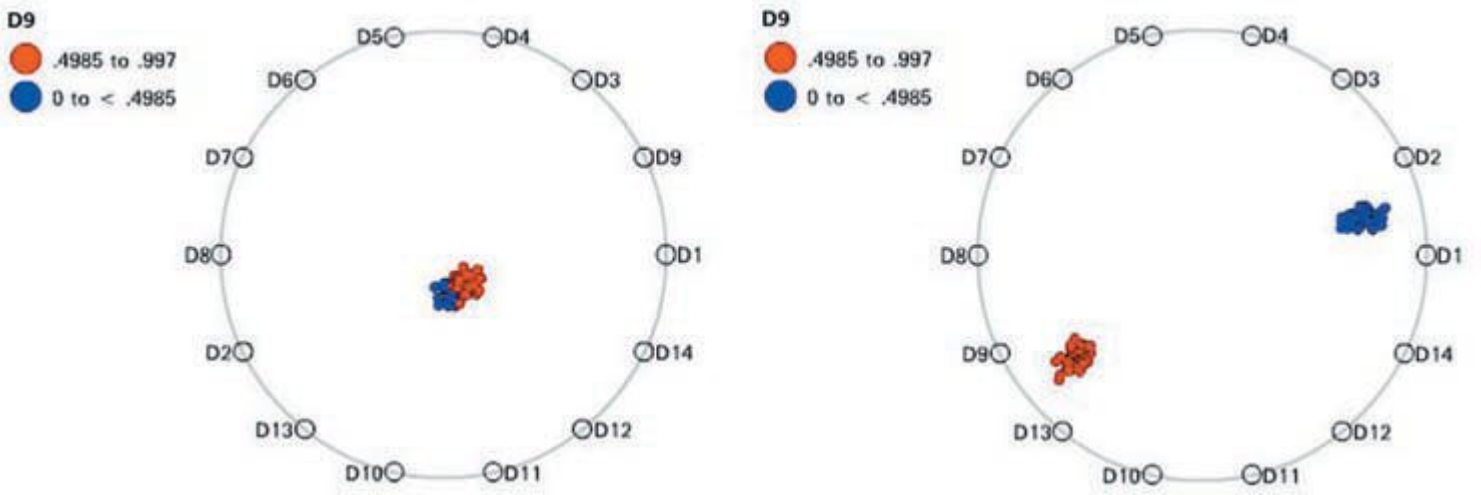


Figure 5. Reordering of 14 d data in RadViz: the arrangement of dimensional anchors on the left produces overlapping clusters for 100 data records with two inherent clusters, whereas the reordering on the right separates the two clusters. D9 is the dimension being used to color the records to provide clear cluster separation.

Source: Daniels et al. Properties of normalized radial visualizations. Information Visualization 11(4) 273–300.

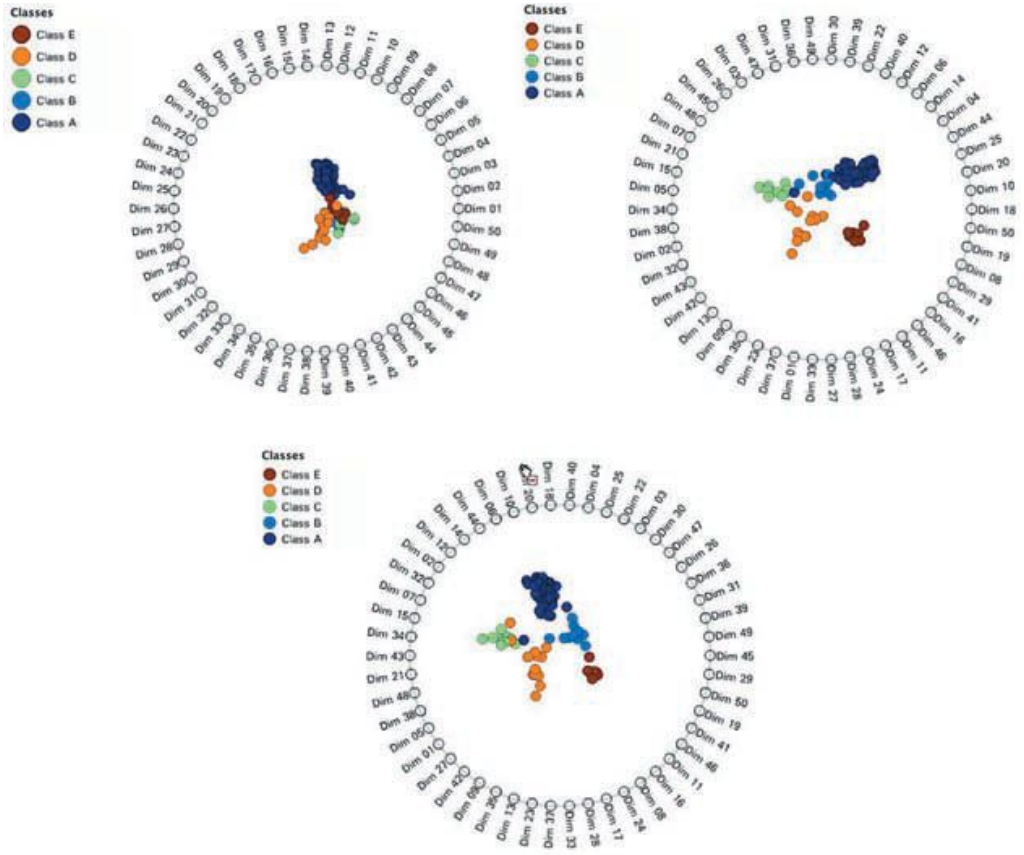


Figure 6. Reordering of 50d Bremer pediatric brain tumor data … in RadViz: the original arrangement of dimensional anchors (top left) produces overlapping clusters for 92 data records with five inherent clusters. The top-right figure uses a class discrimination heuristic for ordering the dimensions, resulting in some improvement in cluster separation. The bottom figure employs a nearest neighbor heuristic and achieves even better separation.

Source: Daniels et al. Properties of normalized radial visualizations. Information Visualization 11(4) 273–300.

CAVEATS

The best way to visualize data efficiently is probably to avoid the most common mistakes.

From Data to Viz offers you a gallery of common caveats.

Overplotting

Too many points on your scatter plot makes it unreadable? Techniques exist to avoid overplotting.

The rainbow color palette

Avoid the rainbow color palette when you map a numeric variable. So many better palettes exist.

Faceting: horizontal or vertical?

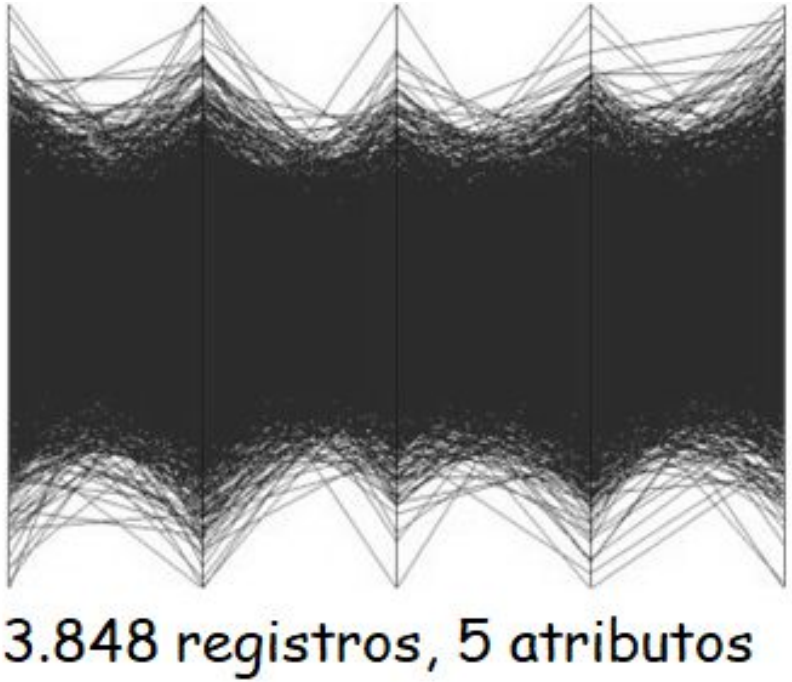
Placing individual plots horizontally or vertically is an important choice to make.

Don't be counter intuitive

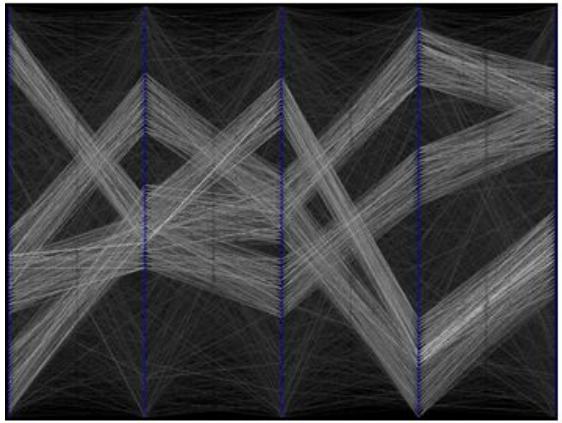
Your audience is used to a few dataviz standards. Not respecting it can mislead them.

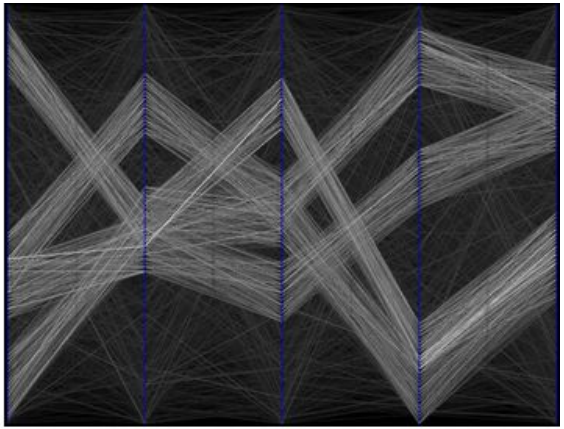
Parallel Coordinates

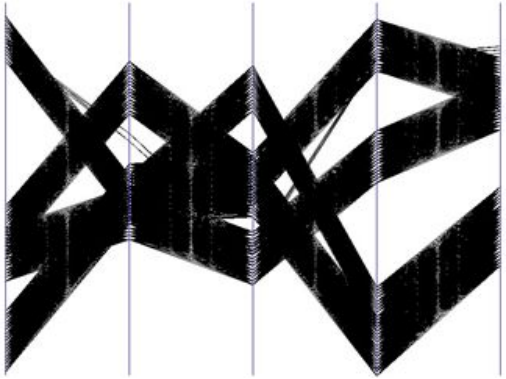
overplotting



Source: A. O. Artero. Exploração Visual de Grandes Conjuntos de Dados Multidimensionais.Tese de doutorado, ICMC-USP 2005.







cinco grupos

Source: A. O. Artero. Exploração Visual de Grandes Conjuntos de Dados Multidimensionais. Tese de doutorado, ICMC-USP 2005.

Parallel Coordinates

axes
reordering
heuristics

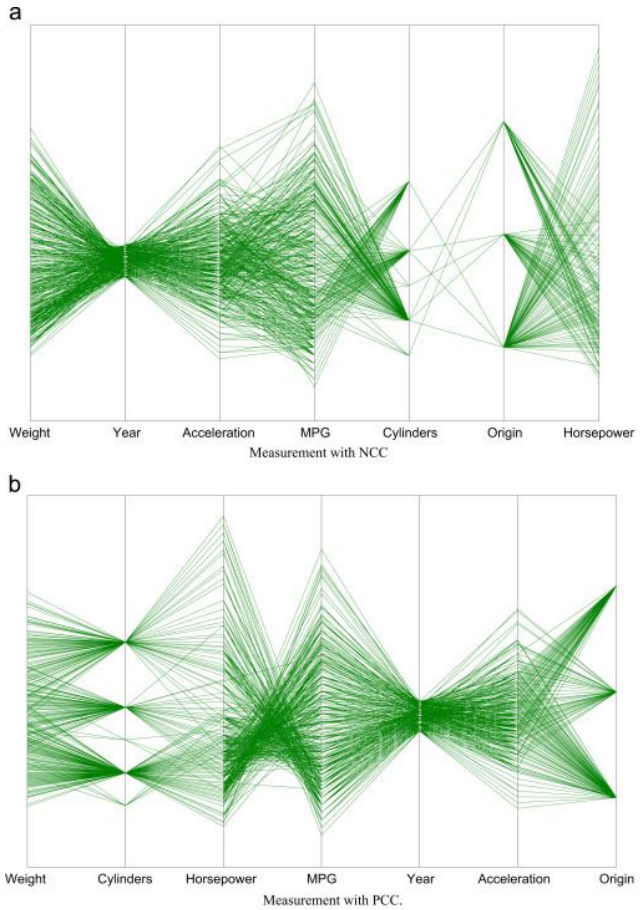


Fig. 3. **Dimension reordering** of Cars dataset in parallel coordinates. (a) Measurement with NCC and (b) measurement with PCC. Source: Lu et al. 2016, Two axes re-ordering methods in parallel coordinates plots. Journal of Visual Languages and Computing

Visualization techniques: critical issues

too many attributes (SPLOMs, Parallel coordinates, RadViz, …)

too many instances - overplotting

too many instances + too many attributes (overplotting, confusion, interactivity)

order of axes (Parallel coordinates, RadViz, …)

Visualization techniques

point-based vs line-based vs. area based vs density-based

instance-based vs attribute-based

e.g., multidimensional projection vs parallel coordinates or radviz

CAVEATS

The best way to visualize data efficiently is probably to avoid the most common mistakes.

From Data to Viz offers you a gallery of common caveats.

Overplotting

Too many points on your scatter plot makes it unreadable? Techniques exist to avoid overplotting.

The rainbow color palette

Avoid the rainbow color palette when you map a numeric variable. So many better palettes exist.

Faceting: horizontal or vertical?

Placing individual plots horizontally or vertically is an important choice to make.

Don't be counter intuitive

Your audience is used to a few dataviz standards. Not respecting it can mislead them.

Use of color

What to consider when choosing colors for data visualization

https://blog.datawrapper.de/colors/

Use of color

Color in Data Visualization: Less How, More Why

https://towardsdatascience.com/color-in-data-visualization-less-how-more-why-348514a3c4d8

Use of color

Reuse good color schemes

colorBrewer: https://colorbrewer2.org/

Carto: https://carto.com/carto-colors/