

Multidimensional Visual Analysis

Ožbej Golob

Seminar/Project Interactive and Visual Information Systems SS 2023

Copyright 2022 by the author(s), except as otherwise noted.

This work is placed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence. <https://creativecommons.org/licenses/by/4.0/>

Multidimensional Visual Analysis (MVA)

- Visual representations to explore and analyze multidimensional datasets.
- Multidimensional datasets represented as a table or spreadsheet.
- Columns: dimensions {variables}
- Rows: records {data points, instances}

Iris Dataset

- 5 dimensions
- 150 records belonging to three classes: Setosa, Versicolour, or Virginica

Sepal Length	Sepal Width	Petal Length	Petal Width	Variety
5.1	3.5	1.4	0.2	Setosa
4.9	3.0	1.4	0.2	Setosa
4.7	3.2	1.3	0.2	Setosa
4.6	3.1	1.5	0.2	Setosa
5.0	3.6	1.4	0.2	Setosa
5.4	3.9	1.7	0.4	Setosa
4.6	3.4	1.4	0.3	Setosa
...				

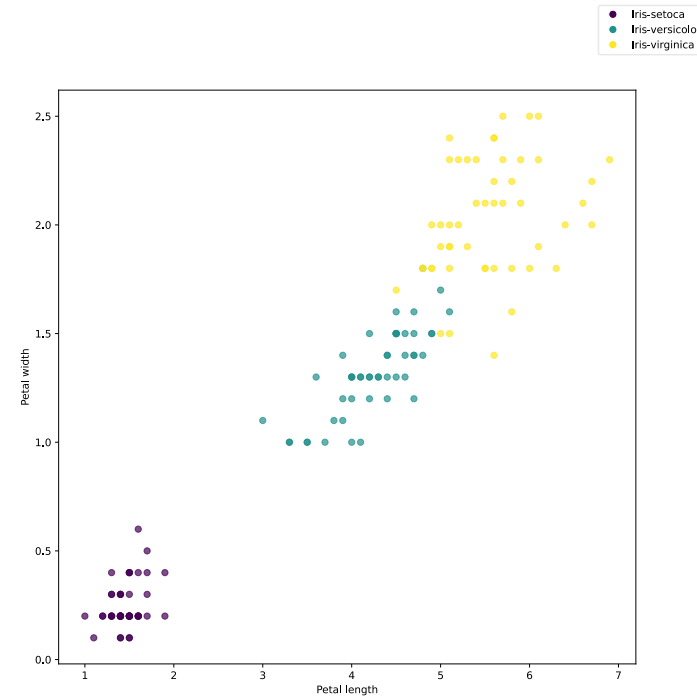
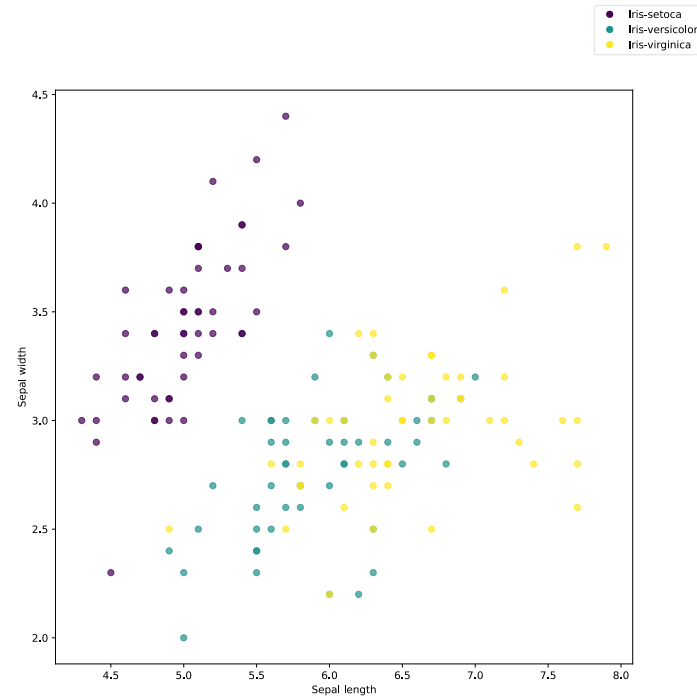
Football Dataset

- 9 dimensions
- 540 records

Player	Team	GP	GS	MIN	G	AST	Shots	SOG
Jamie Vardy	Leicester City	35	34	3034	23	5	71	43
Danny Ings	Southampton	38	32	2812	22	2	66	38
Pierre-Emerick Aubameyang	Arsenal	36	35	3138	22	3	70	42
Raheem Shaquille Sterling	Manchester City	33	30	2660	20	1	68	38
Mohamed Salah Ghaly	Liverpool	34	33	2884	19	10	95	59
Sadio Mané	Liverpool	35	31	2753	18	7	66	36
Harry Kane	Tottenham Hotspur	29	29	2589	18	2	62	37
...								

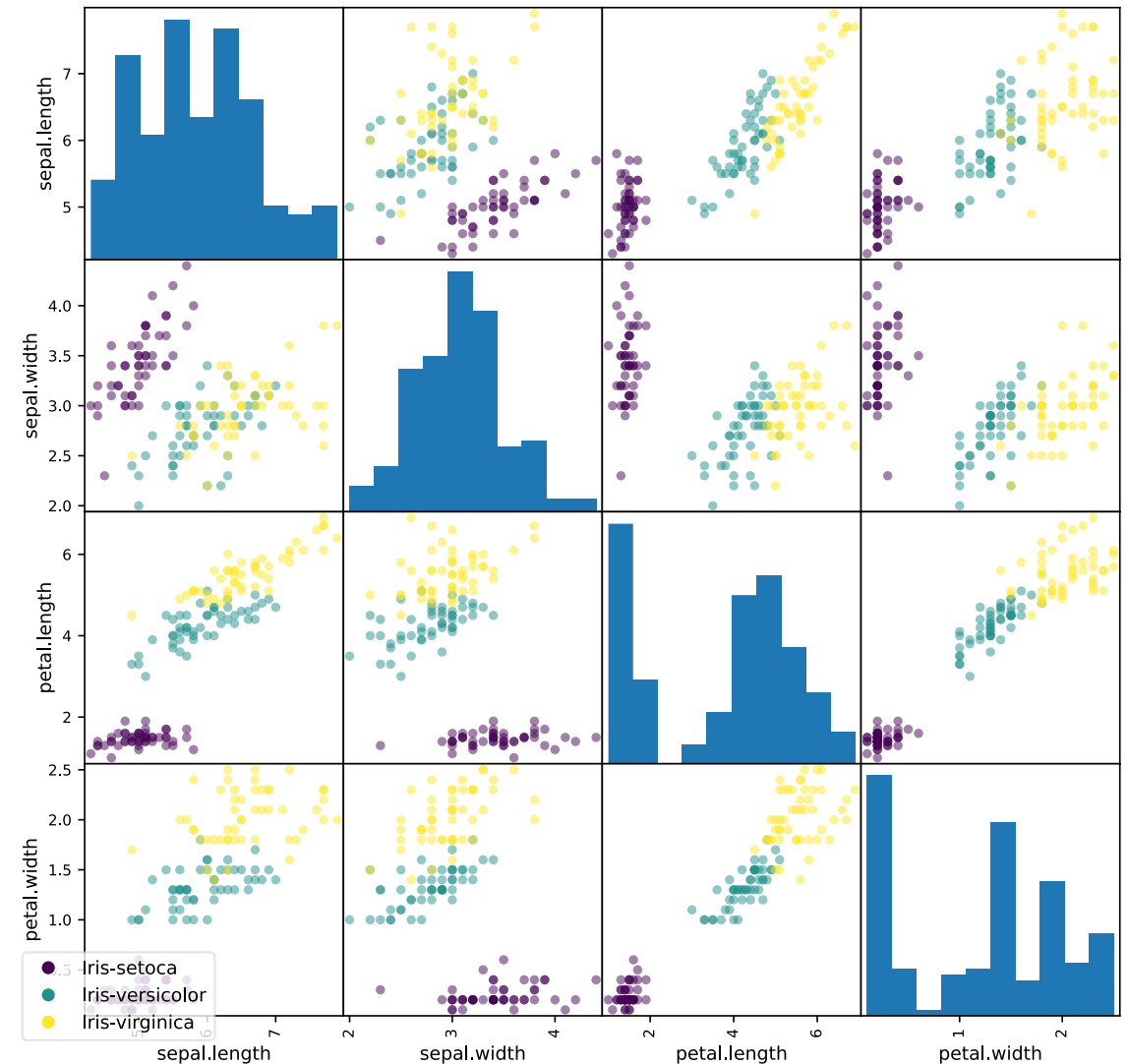
Scatterplots

- Display relationship between two or three numerical dimensions.
- Can help identify potential trends, patterns, or outliers in the data.
- Use dot or marker to represent each record.
- Dots form diagonal line sloping upwards → Indication of a positive correlation.
- Dots form diagonal line sloping downwards → Indication of a negative correlation.



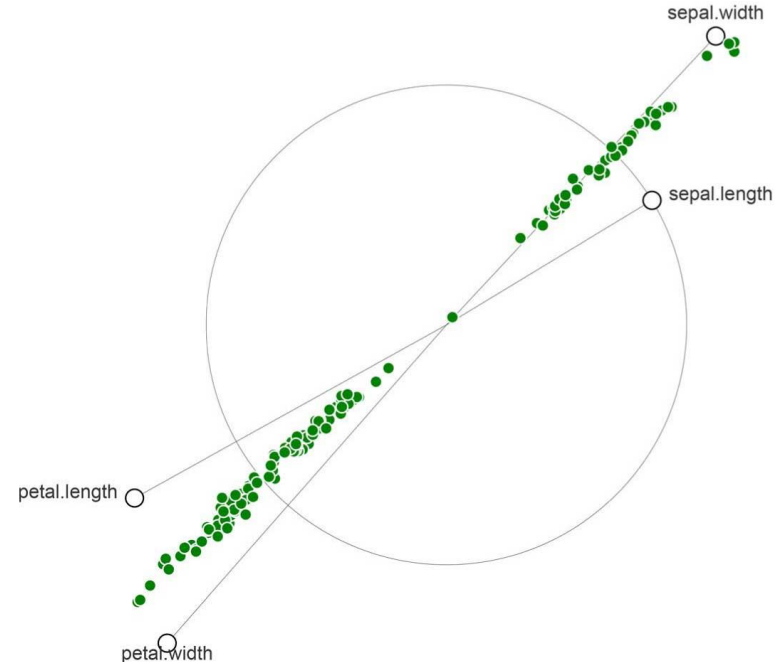
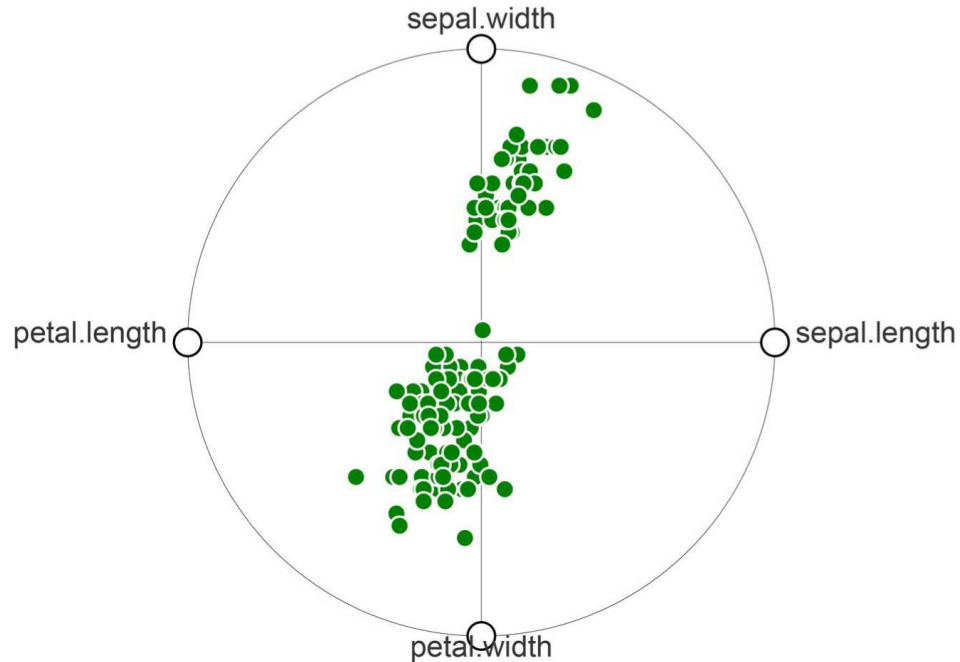
Scatterplot Matrices

- Explore pairwise relationships between multiple dimensions.
- Grid of scatterplots, with each plot displaying relationship between two dimensions.
- Overview of multiple relationships at a glance.
- Cluttered and unmanageable at larger number of dimensions (15).



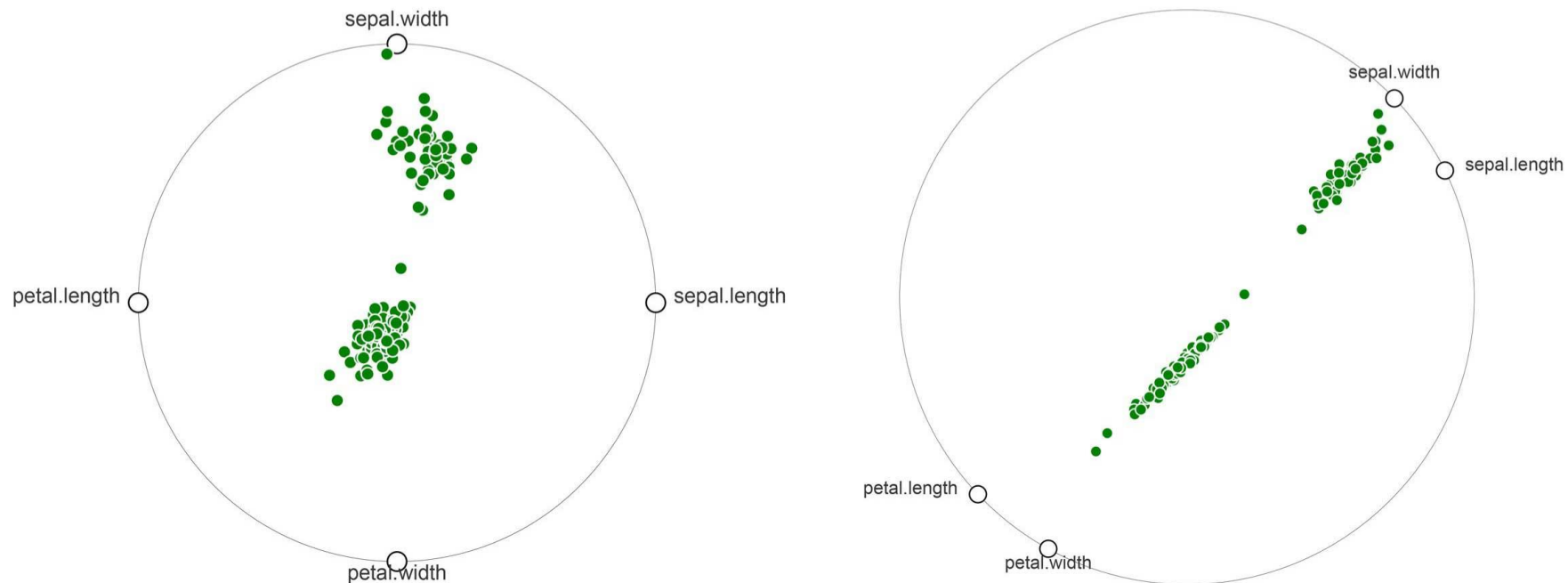
Star Coordinates

- Maps multidimensional data onto a 2D space using a radial layout.
- Plot consists of radial axes (representing dimensions) from the center.
- Each record is mapped by using weighted average of its dimension values.
- Circular marker (anchor) can be moved inside or outside of the circle to adjust the weights.
- Records can be inside/outside the circle.
- Cluttered and unmanageable at larger number of dimensions (20).



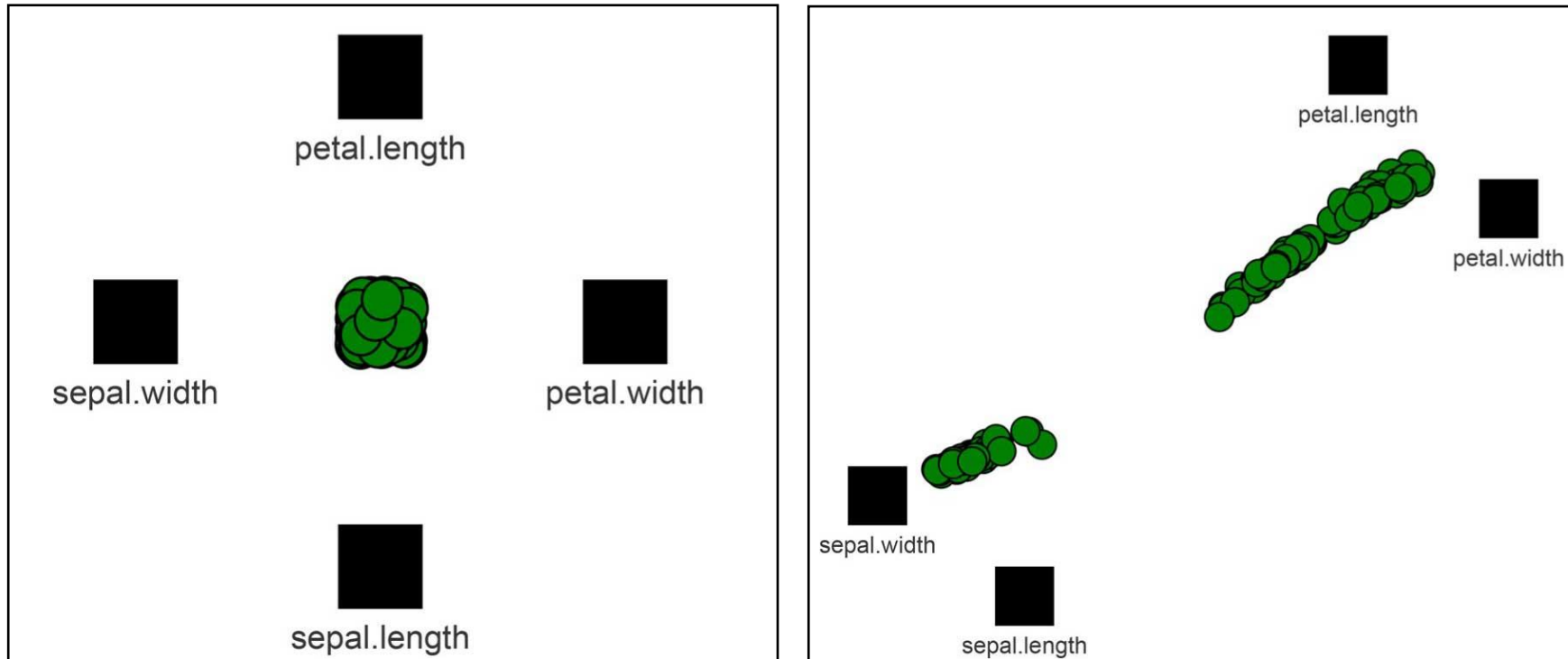
RadViz

- Maps multidimensional data onto a 2D space using a radial layout.
- Plot consists of radial axes (representing dimensions) from the center.
- Each record is mapped by using weighted average of its dimension values.
- Anchor can only be moved on the circle perimeter to adjust the weights.
- Records remain inside the circle.
- Cluttered and unmanageable at larger number of dimensions (20).

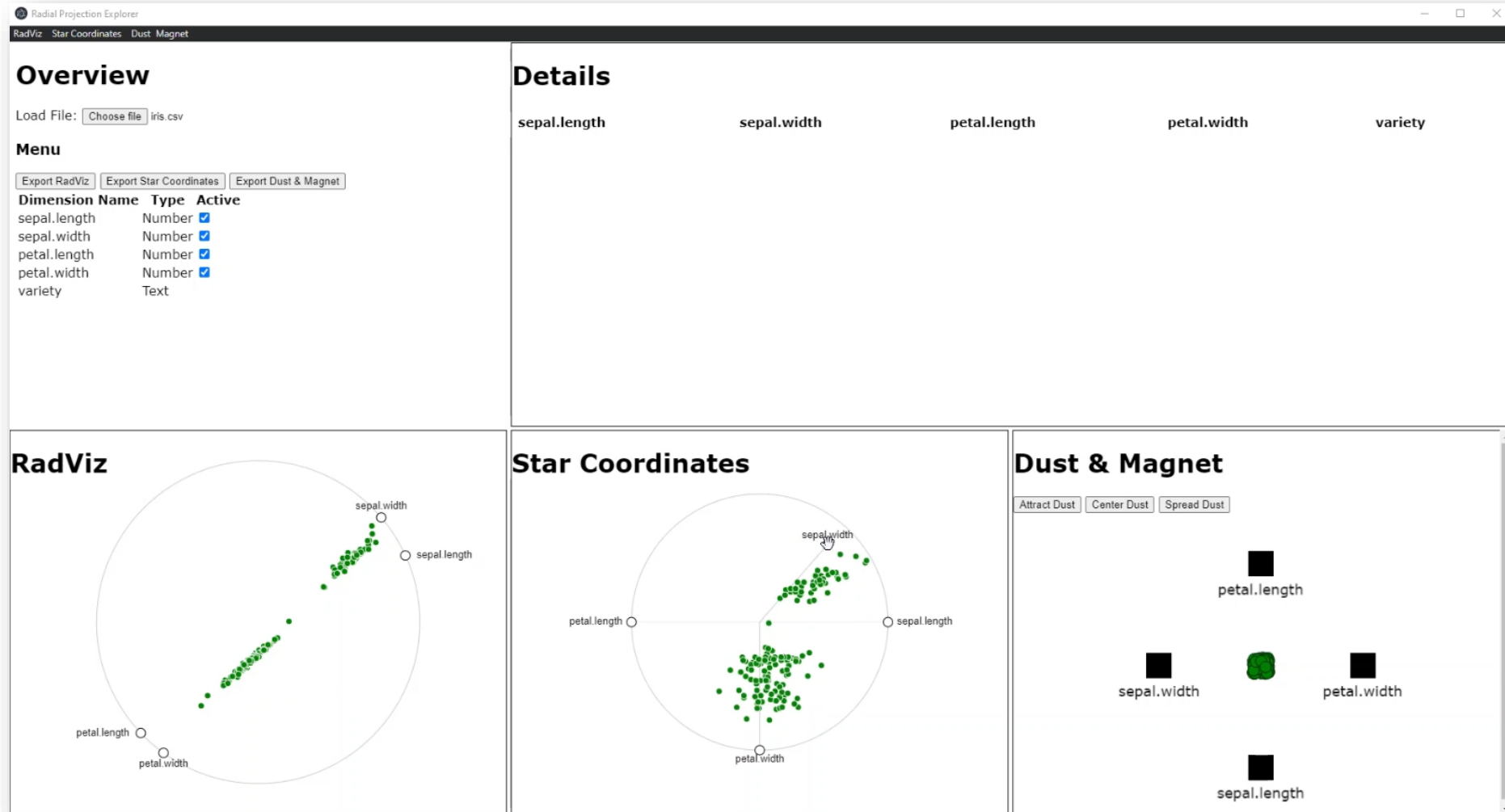


Dust and Magnet (DNM)

- Maps multidimensional data onto a 2D space.
- Records represented as particles attracted to magnets.
- Moving magnets adjusts dimension influence.
- Cluttered and unmanageable at larger number of dimensions (20).



RadViz, Star Coordinates, and DNM Showcase



<https://youtu.be/kJQTWbMxvYA>

Software used: Neuhold, Lukas, Ridvan Aydin, and Georg Regitnig [2020]. The Radial Projection Explorer. 29 Jun 2020.

<https://courses.isds.tugraz.at/ivis/projects/ss2020/ivis-ss2020-g4-project-radial-projectionexplorer.pdf>

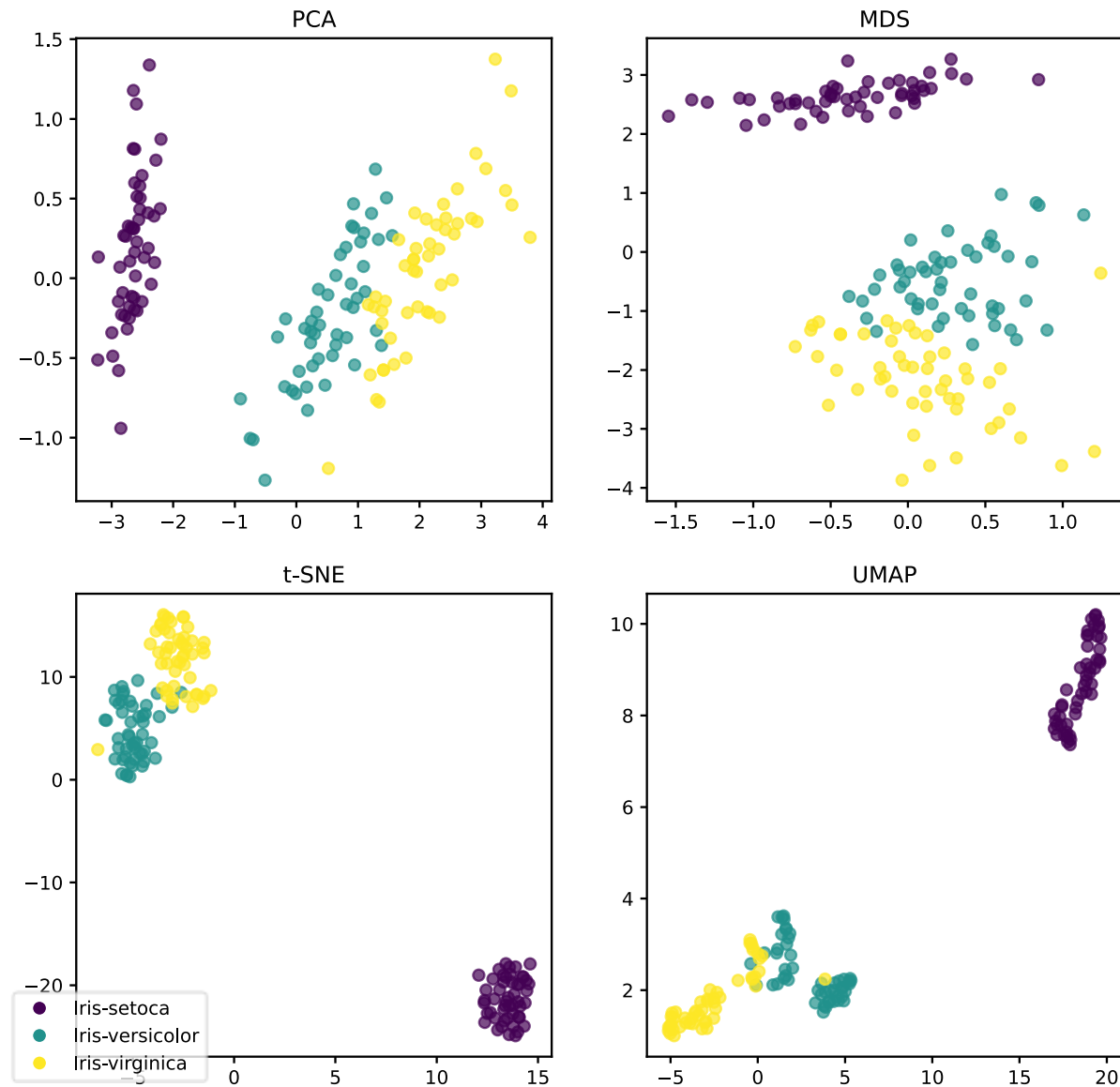
Similarity Maps

- Projections of high-dimensional datasets to 2D or 3D.
- Can accommodate any number of dimensions.
- Two types of projections:
 - Linear: Output is a linear combination of the input dimensions.
 - Non-linear: Output dimensions are calculated with non-linear operations.

Similarity Maps: Projection Techniques

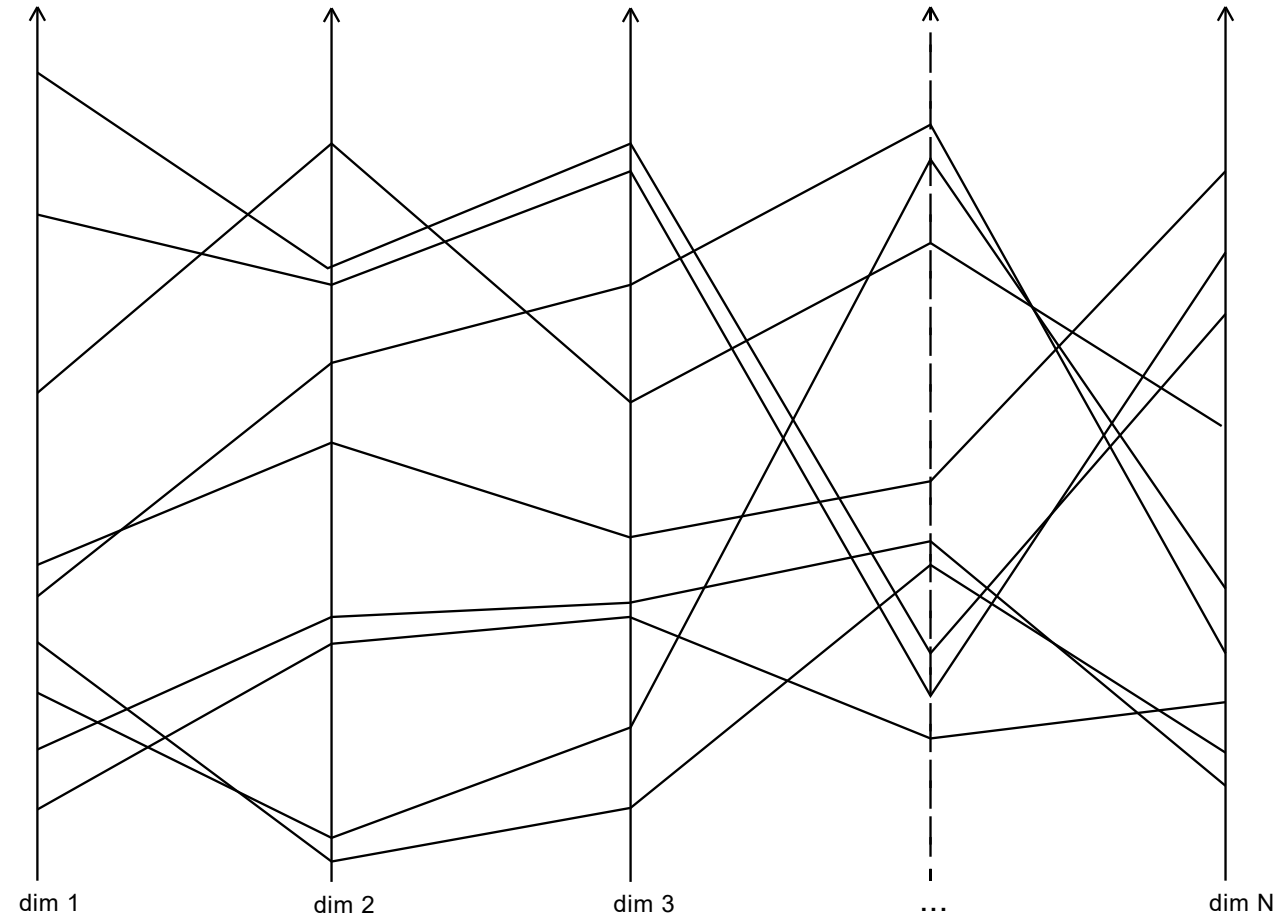
- **Principal Component Analysis (PCA):** Linear projection, first two principal components are used as output dimensions.
- **Multi-Dimensional Scaling (MDS):** Non-linear projection, projects into lower-dimensional space while preserving pairwise distances
- **t-Distributed Stochastic Neighbor Embedding (t-SNE):** Non-linear projection, uses a probabilistic model to preserve the local structure of the data.
- **Uniform Manifold Approximation and Projection (UMAP):** Non-linear projection, uses ideas from topological data analysis to preserve global and local structure of the data.

Similarity Maps: Comparison



Parallel Coordinates

- Each dimension represented by vertical line (axis).
- Each record represented by horizontal poly-line across multiple axes.
- Each point in poly-line represents the corresponding dimension value.
- Cluttered and unmanageable at larger number of dimensions (20).



Brushing and Linking

- **Brushing:** Process of selecting records or regions in one visualization and highlighting those records in other visualizations.
- **Linking:** Process of synchronizing the views of multiple visualizations, such that a change made to one visualization is reflected in the other visualizations.



<https://youtu.be/hM4SZcBCnhg>

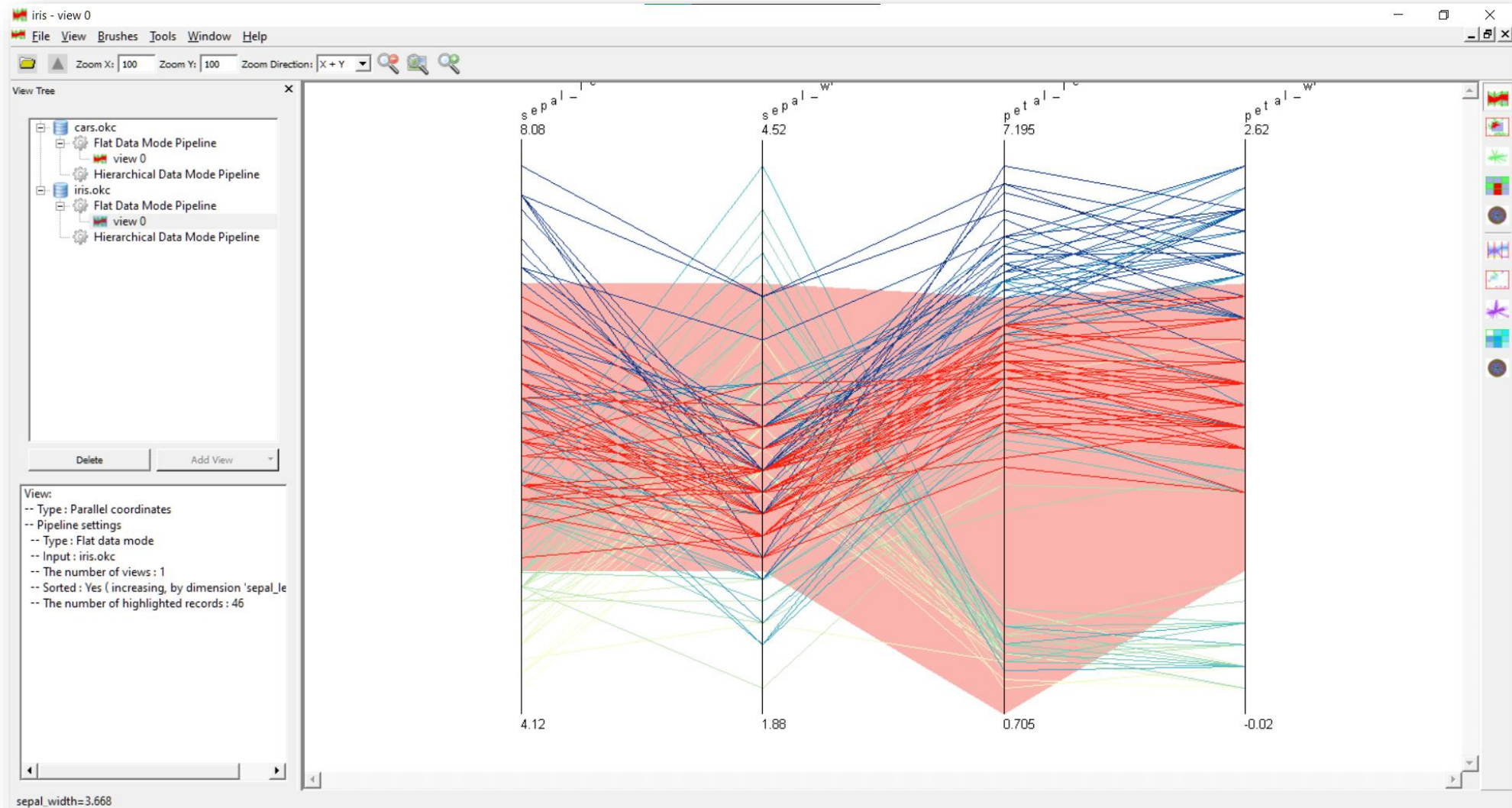
Grouping and Labeling

- **Grouping:** Process of identifying and separating similar records.
 - Manual Grouping: Organizing and providing relevant information about records manually, typically done by a human (can be time-consuming and labor-intensive).
 - Automated Clustering: Mathematical methods used to group records with similar descriptions into the same cluster.
- **Labeling:** Process of providing relevant information about the records.

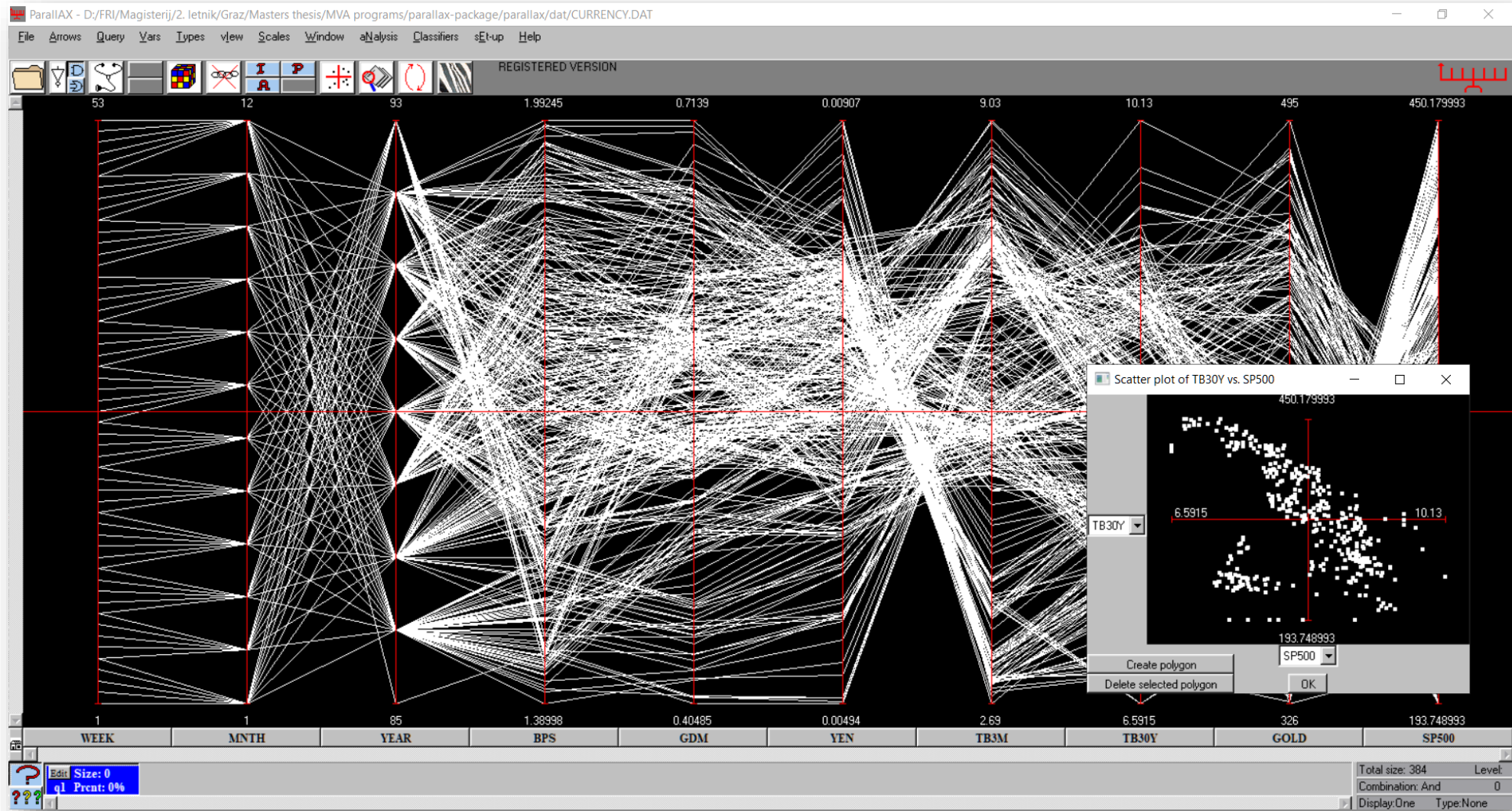
MVA Tools

- Tools for visual analysis of multidimensional datasets:
 - XMDV (1994 – 2021)
 - Parallax (1999)
 - Ggobi (1999 – 2012)
 - InfoScope (2007)
 - XDAT (2010 – 2020)
 - High-D (2013 – 2022)
 - TabuVis (2013 – 2022)
 - Improvise (2014 – 2020)
 - MyBrush (2016 – 2017)
 - mVis (2020 – 2021)

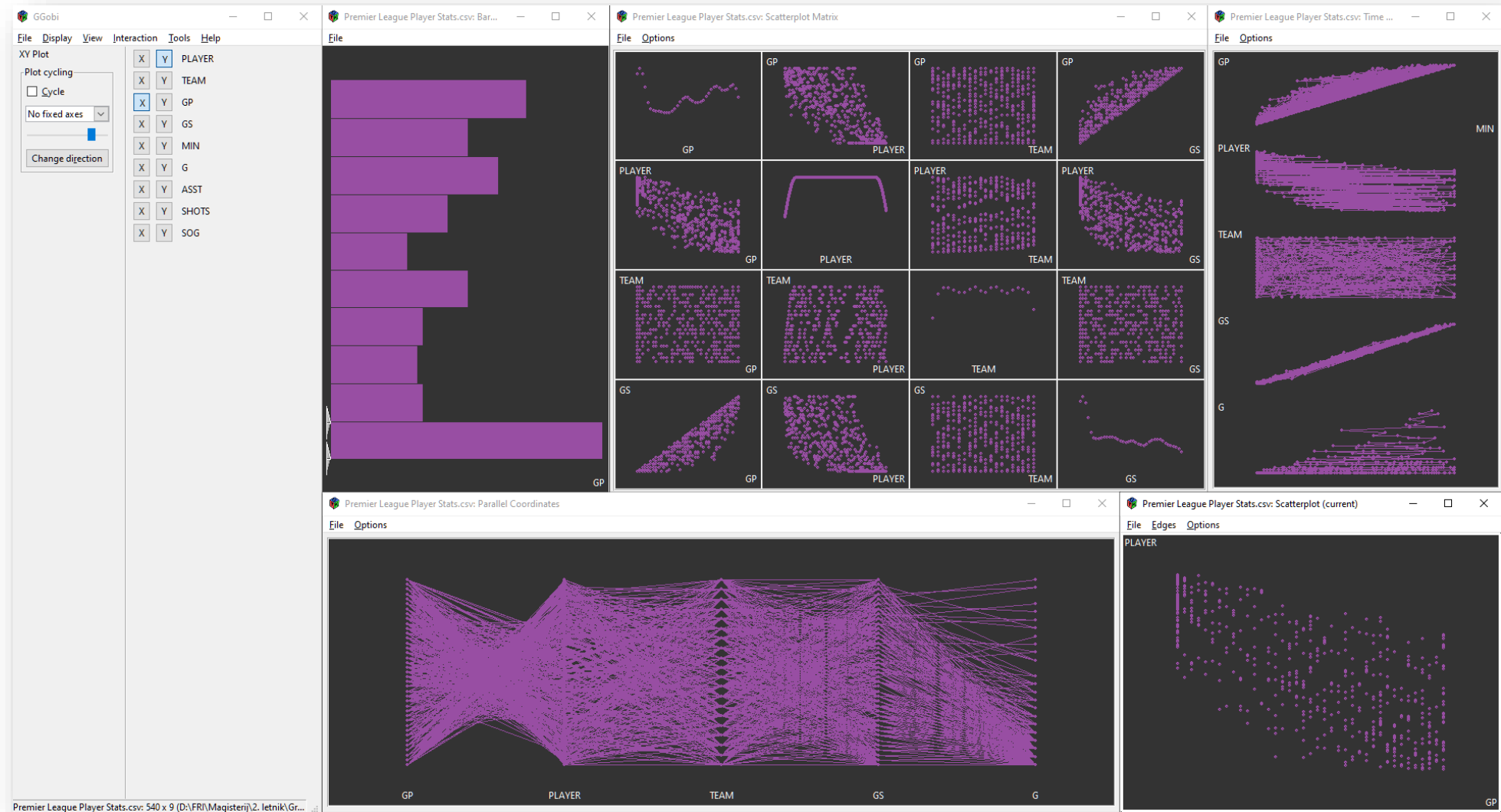
XMDV (1994 – 2021)



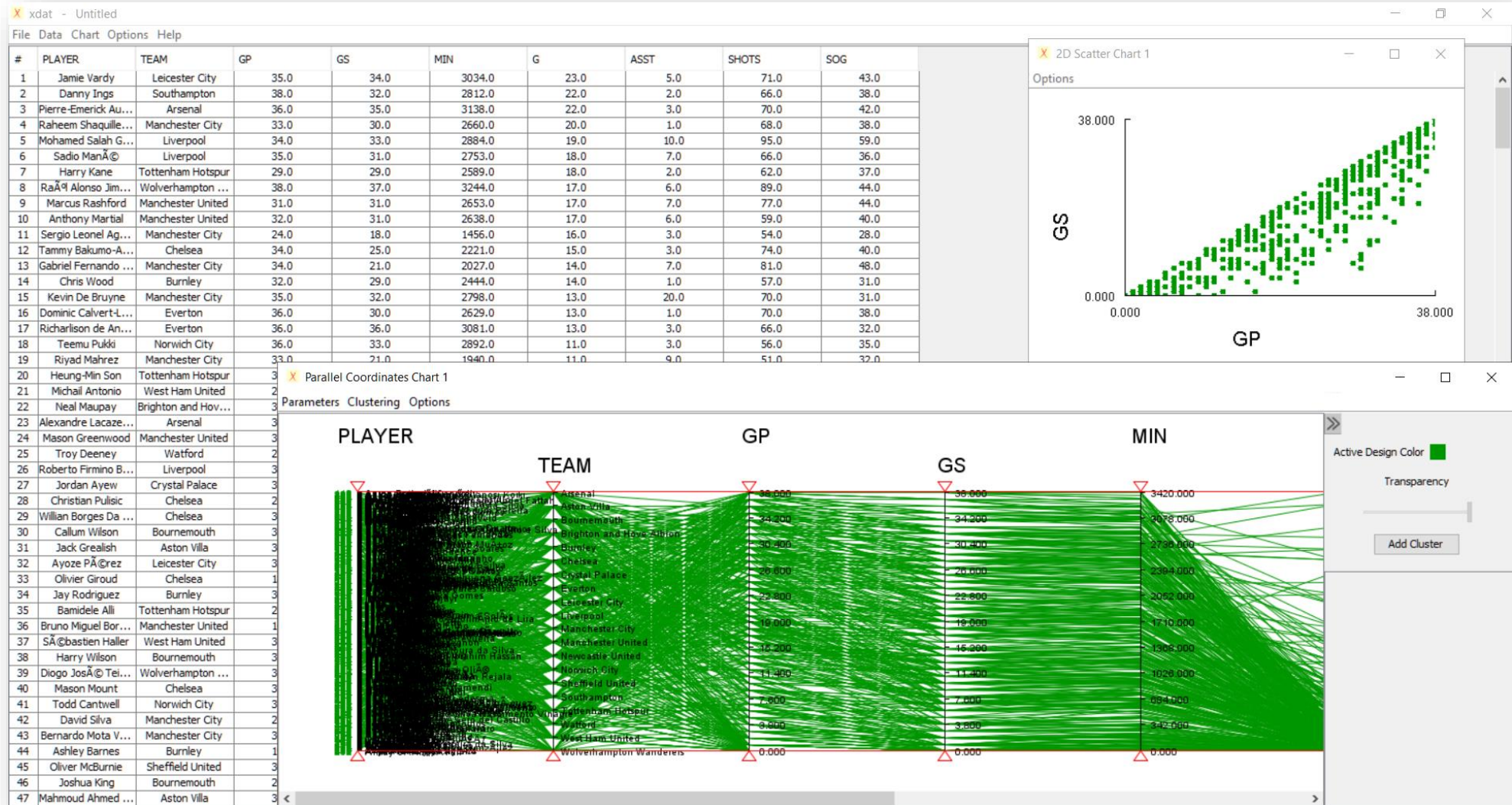
Parallax (1999)



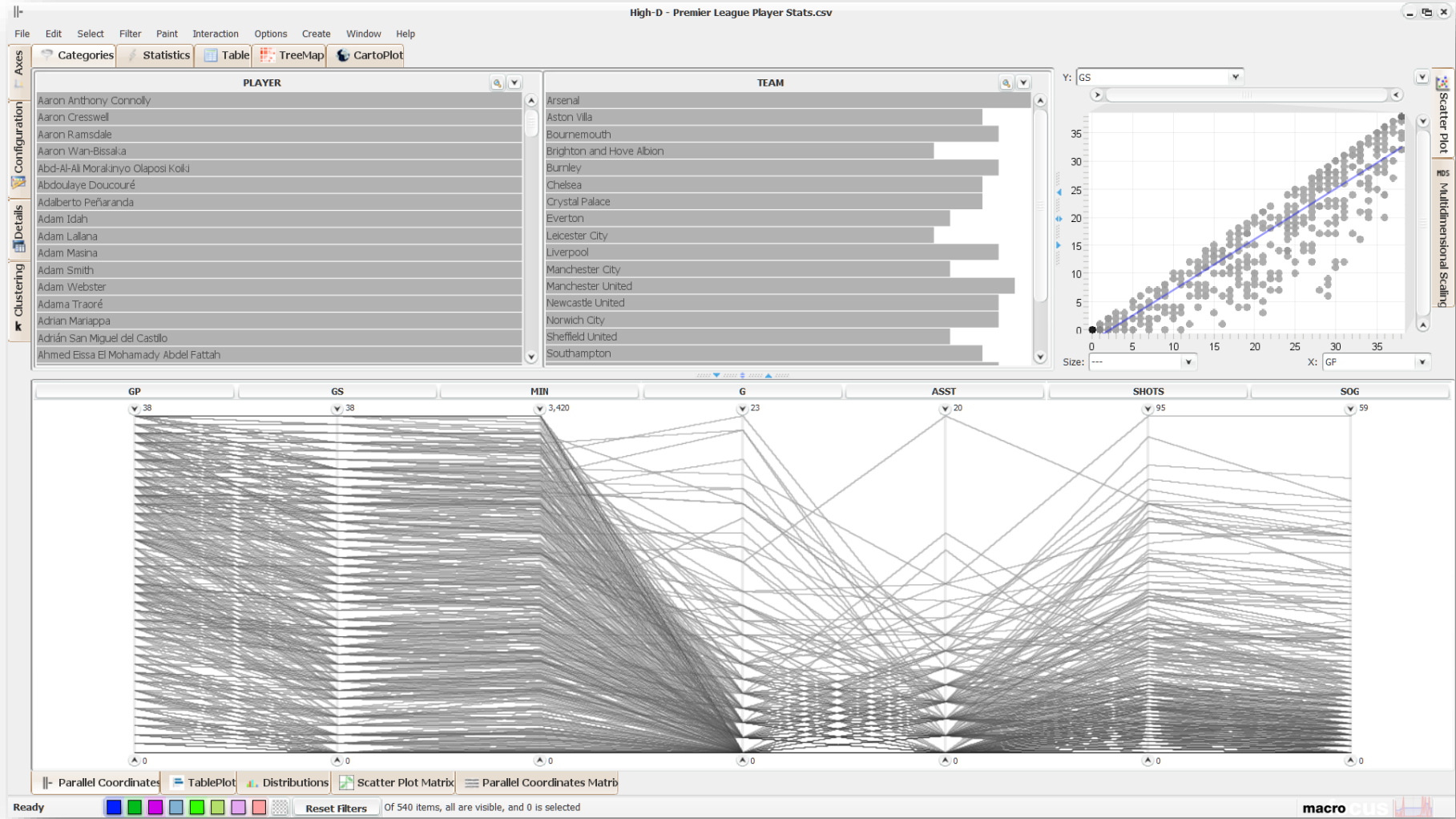
Ggobi (1999 – 2012)



XDAT (2010 – 2020)

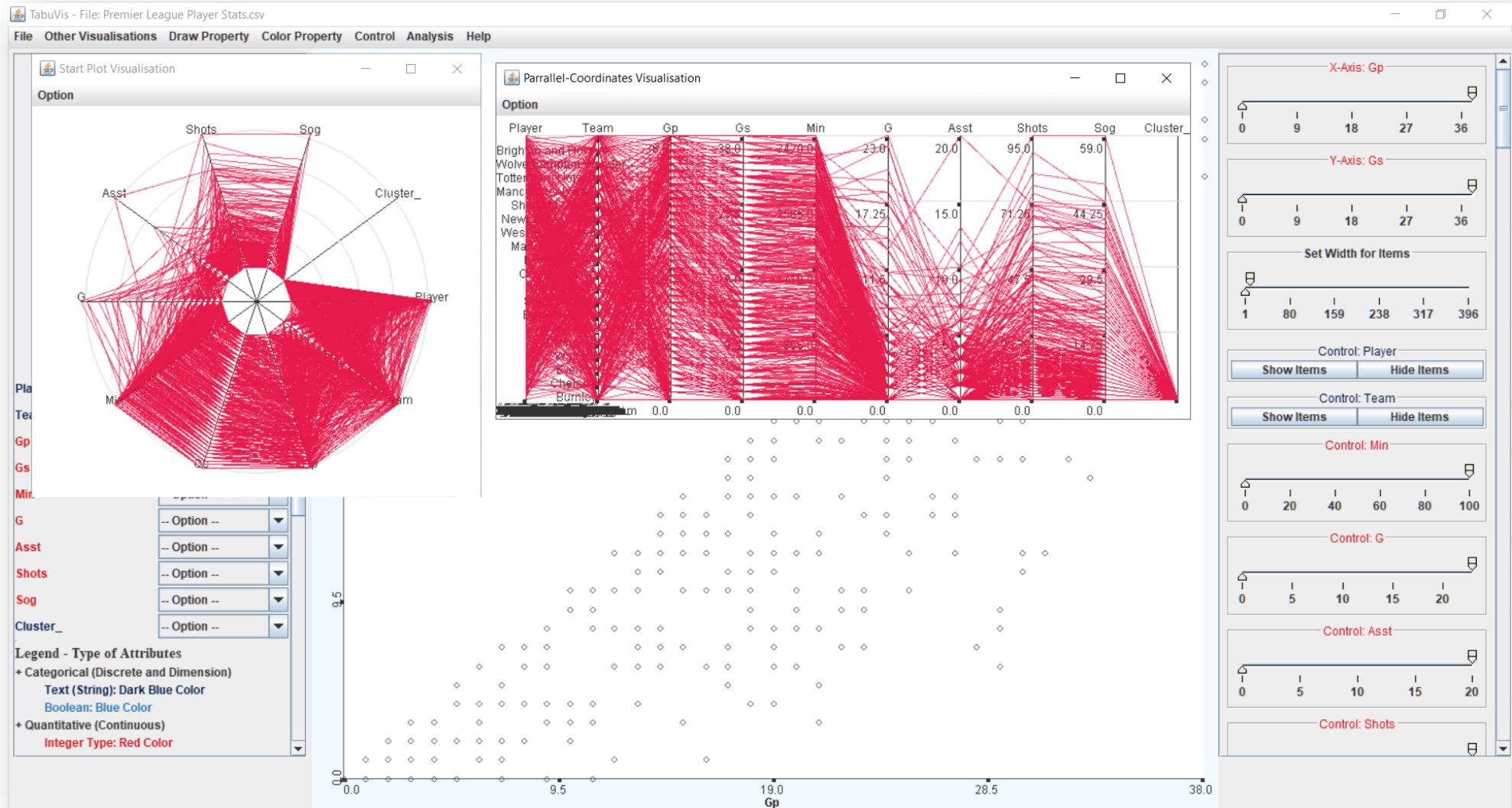


High-D (2013 – 2022)

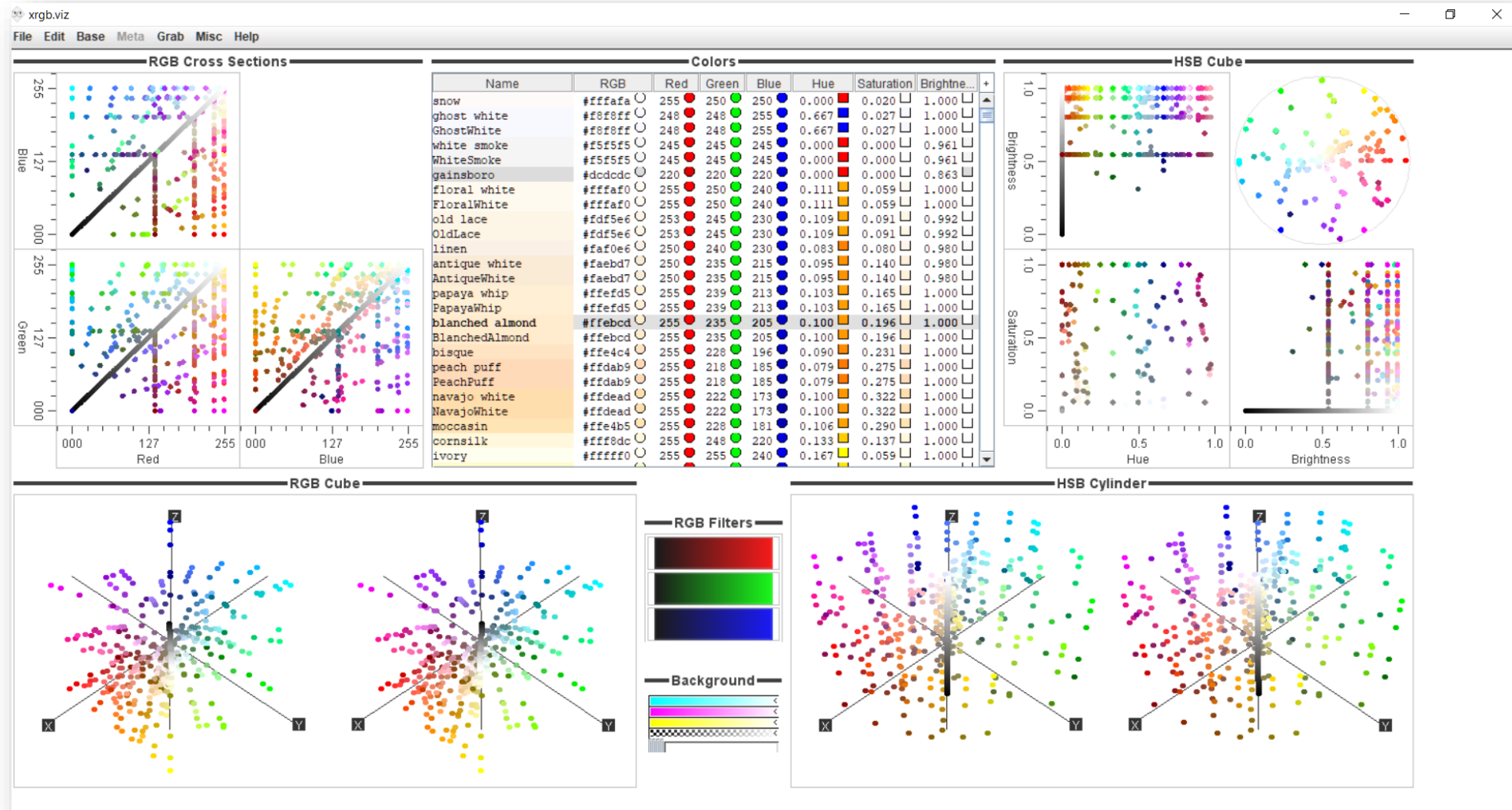


<https://youtu.be/2eKJOOX0-3o>

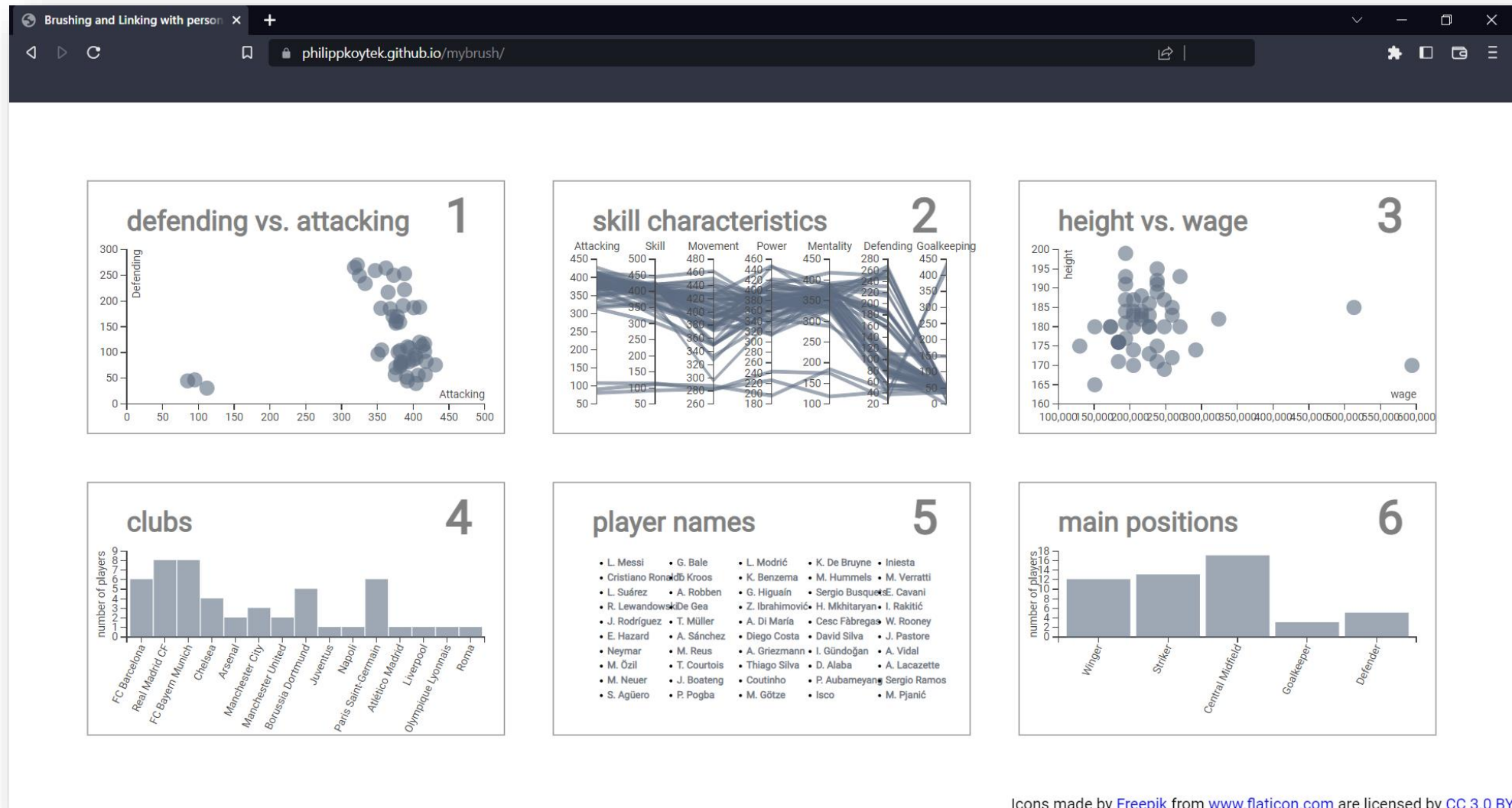
TabuVis (2013 – 2022)



Improvise (2014 – 2020)

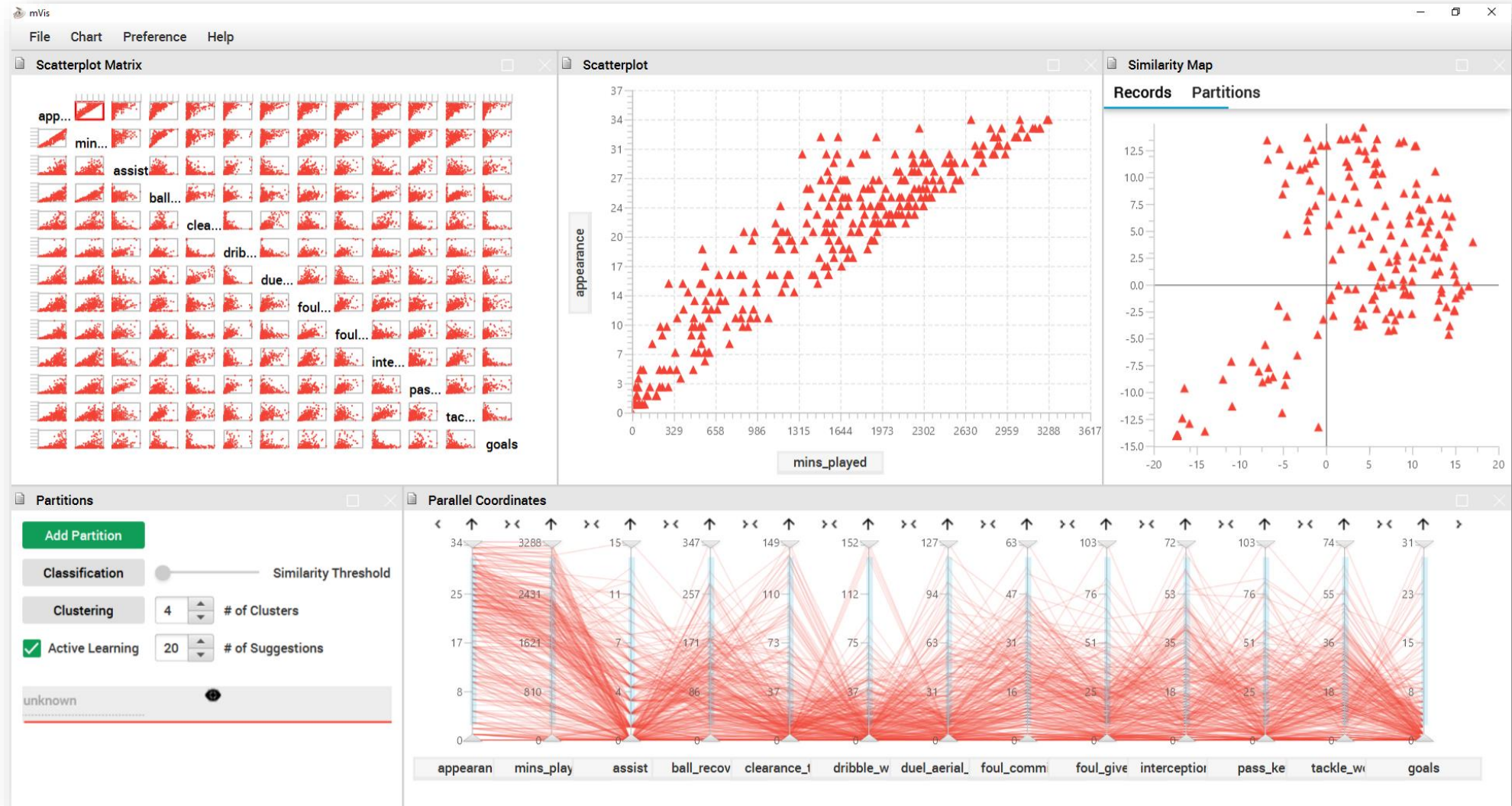


MyBrush (2016 – 2017)



Icons made by [Freepik](https://www.flaticon.com) from www.flaticon.com are licensed by [CC 3.0 BY](https://creativecommons.org/licenses/by/3.0/)

mVis (2020 – 2021)



MVA Tools

	XMDV	Parallax	GGobi	XDAT	High-D	TabuVis	Improvise	MyBrush	mVis
Initial Release:	1994	Mar 1999	1999	May 2010	Sep 2013	May 2013	2014	Jun 2016	Jul 2020
Last Update:	Sep 2021	?	Jun 2012	Aug 2020	Dec 2022	Feb 2022	Oct 2020	Sep 2017	Jan 2021
License:	Free, open-source	Commercial	Free, open-source	Free	Commercial	Free	Free, open-source	Free, open-source	Free, open-source
Systems:	Win, MacOS, Linux	Win, MacOS, Linux	Win, MacOS, Linux	Win, MacOS, Linux	Win, MacOS, Linux	Win, MacOS, Linux	Win, MacOS, Linux	Web Browser	Win, MacOS, Linux
Language:	Qt	?	C	Java	?	Java	Java	JavaScript	Java
Installation:	Local	Local	Local	Local	Local	Local	Local	Online	Local

MVA Tools: Features

Feature	XMDV	Parallax	GGobi	XDAT	High-D	TabuVis	Improvise	MyBrush	mVis
Custom Datasets:	✓	✓	✓	✓	✓	✓	✓		✓
Brushing:	✓		✓	✓	✓			✓	✓
Linking:	✓		✓	✓	✓			✓	✓
Manual Grouping:	✓	✓		✓	✓	✓			✓
Automated Clustering:		✓			✓	✓			✓
Table View:				✓	✓		✓		
Scatterplot:	✓	✓	✓	✓	✓	✓	✓	✓	✓
Scatterplot Matrix:	✓		✓		✓		✓		✓
Parallel Coordinates:	✓	✓	✓	✓	✓	✓		✓	✓
Parallel Coordinates Matrix:					✓				
Similarity Map:					✓	✓	✓		✓
Time Series:			✓				✓		
Distributions:		✓	✓		✓		✓	✓	
Table Plot:					✓		✓		
Tree Map:	✓				✓		✓		
Carto Plot:					✓		✓		

Concluding Remarks

- Most common MVA approaches: scatterplot, scatterplot matrix, parallel coordinates, and similarity map.
- Brushing and linking important.
- Best commercial tool: High-D.
- Best free tool: mVis.
- Future work: Multidimensional Visual Analyser web application (Masters thesis)

