

Wizualizacja danych

Wykład 10: Dane wielowymiarowe

Mariusz Tarnopolski

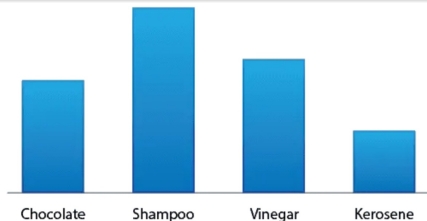
Instytut Astronomii UMK

DataViz, 2024/2025



Wykres radarowy/gwiazdowy

Radar/spider chart



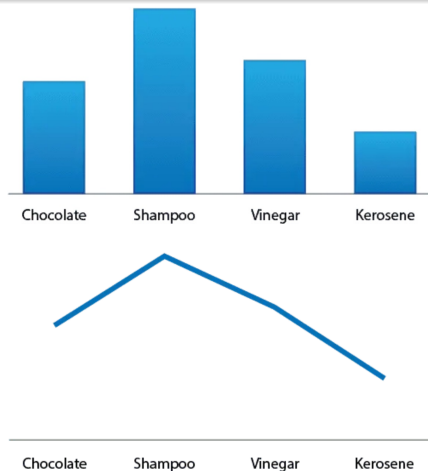
https://en.m.wikipedia.org/wiki/Radar_chart;

<https://towardsdatascience.com/>

[ghosts-on-the-radar-why-radar-charts-are-easily-misread-dba00fc399ef](#)

Wykres radarowy/gwiazdowy

Radar/spider chart



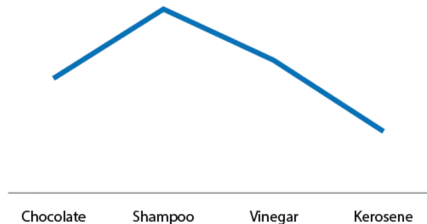
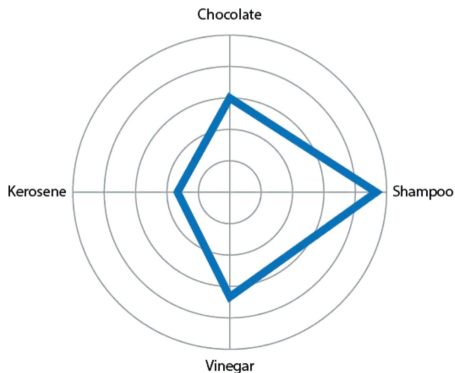
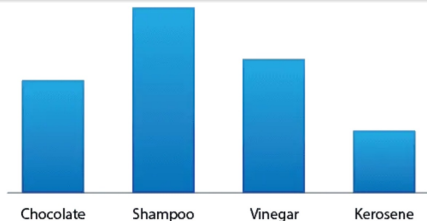
https://en.m.wikipedia.org/wiki/Radar_chart;

<https://towardsdatascience.com/>

[ghosts-on-the-radar-why-radar-charts-are-easily-misread-dba00fc399ef](#)

Wykres radarowy/gwiazdowy

Radar/spider chart



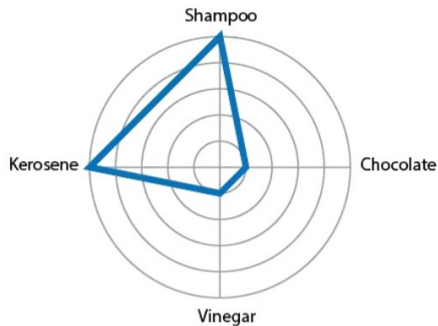
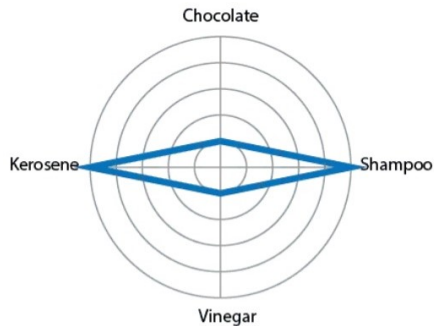
https://en.m.wikipedia.org/wiki/Radar_chart;

<https://towardsdatascience.com/>

[ghosts-on-the-radar-why-radar-charts-are-easily-misread-dba00fc399ef](https://towardsdatascience.com/ghosts-on-the-radar-why-radar-charts-are-easily-misread-dba00fc399ef)

Wykres radarowy/gwiazdowy

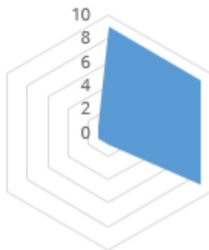
Radar/spider chart



Wykres radarowy/gwiazdowy

Radar/spider chart

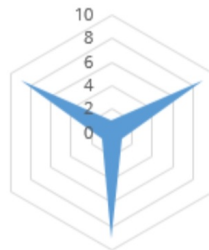
9-9-9-1-1-1



9-9-1-9-1-1



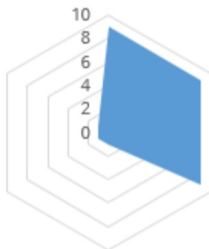
1-9-1-9-1-9



Wykres radarowy/gwiazdowy

Radar/spider chart

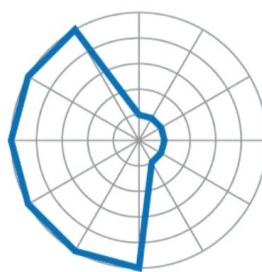
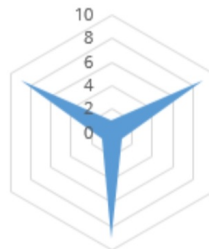
9-9-9-1-1-1



9-9-1-9-1-1



1-9-1-9-1-9



Wykres radarowy/gwiazdowy

Radar/spider chart

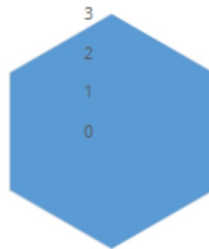
1-1-1-1-1-1



2-2-2-2-2-2



3-3-3-3-3-3

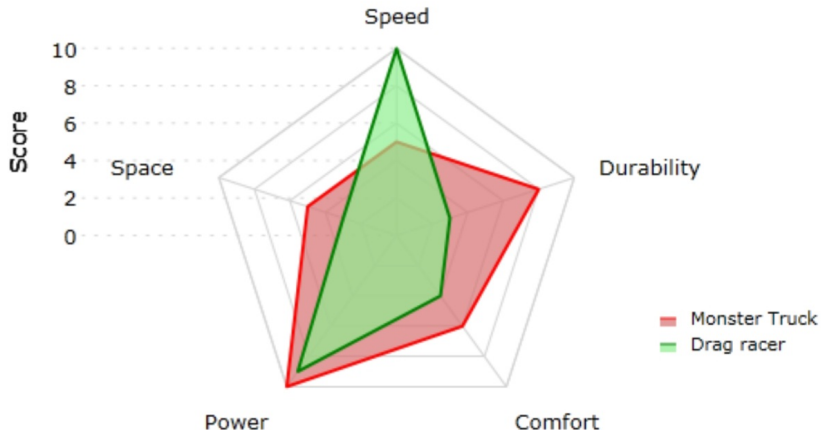


<https://www.darkhorseanalytics.com/blog/radar-more-evil-than-pie>

Wykres radarowy/gwiazdowy

Radar/spider chart

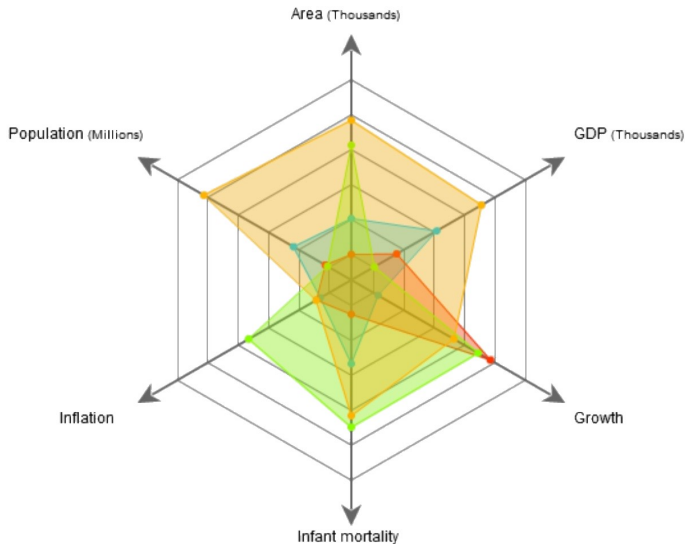
Comparing Cars



<https://blog.scottlogic.com/2011/09/23/a-critique-of-radar-charts.html>

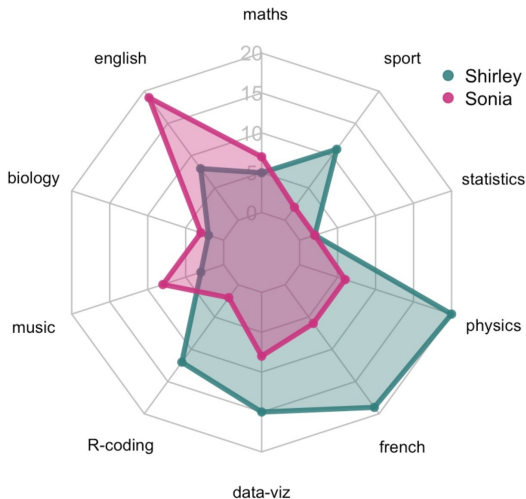
Wykres radarowy/gwiazdowy

Radar/spider chart



Wykres radarowy/gwiazdowy

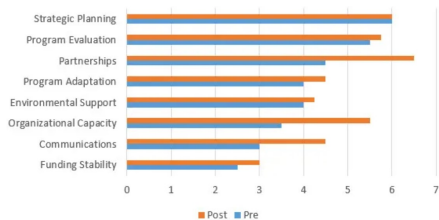
Radar/spider chart



<https://www.data-to-viz.com/caveat/spider.html>

Wykres radarowy/gwiazdowy

Radar/spider chart



<https://stephanieevergreen.com/radar-graphs/>

Wykres radarowy/gwiazdowy

Radar/spider chart

Wady:

- mają niskie dokładność i prędkość odczytywania informacji*
- przede wszystkim kodują kształt (którego powierzchnia rośnie nieliniowo) — ten zaś zależy od kolejności zmiennych
- są czytelne tylko dla małej liczby bardzo małych zestawów danych

Zastosowanie — natura danych (powierzchnia, kierunek, cykliczność):

- ogólny wzrost w wielu kategoriach
- rozkład kierunków wiatru; polaryzacja
- natężenie pewnej wielkości w cyklach dobowych, tygodniowych itp.

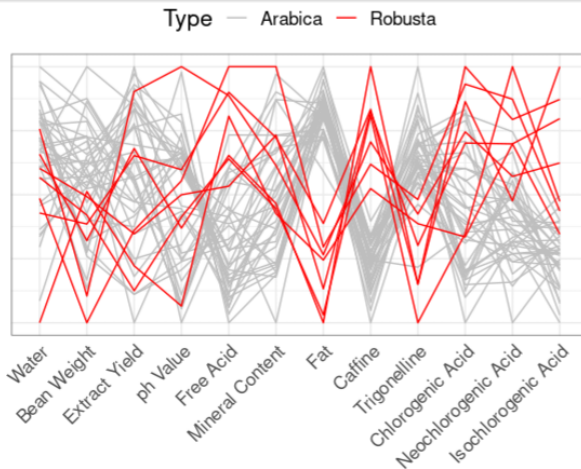
* <https://doi.org/10.1016/j.apergo.2023.103996>

Współrzędne równoległe

Parallel coordinates

Współrzędne równoległe

Parallel coordinates

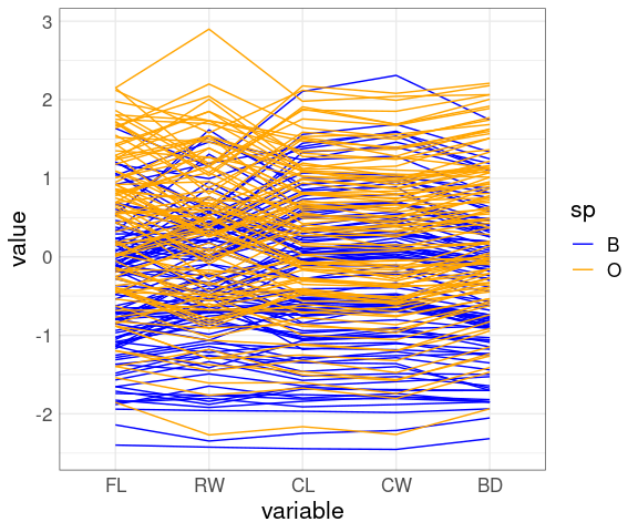


https://en.m.wikipedia.org/wiki/Parallel_coordinates;

<https://homepage.divms.uiowa.edu/~luke/classes/STAT4580/threeplus.html#parallel-coordinates-plots>

Współrzędne równoległe

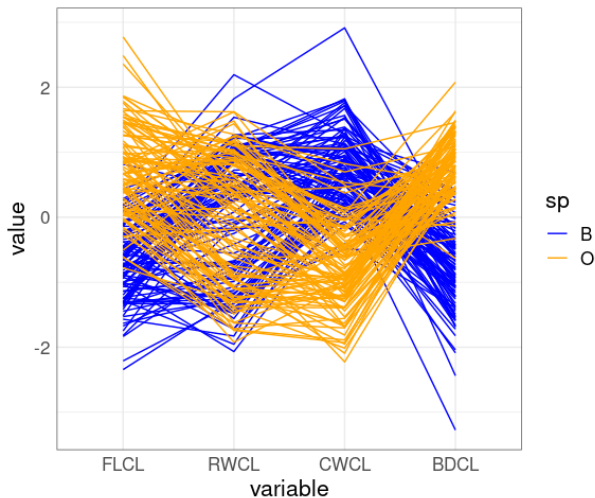
Parallel coordinates



Współrzędne równoległe

Parallel coordinates

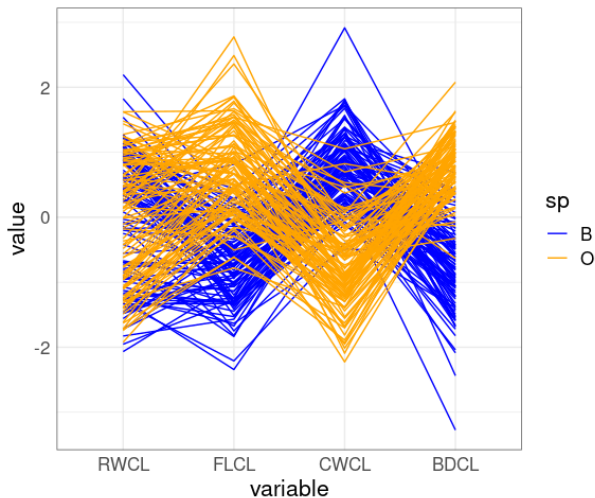
Podzielić wszystkie zmienne przez CL :



Współrzędne równoległe

Parallel coordinates

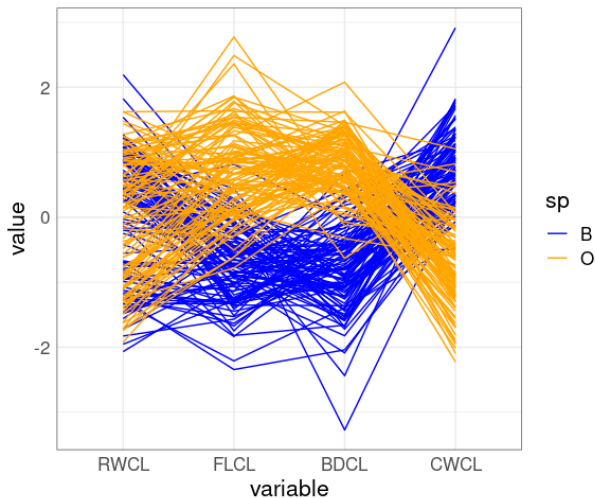
Zamienić miejscami dwie pierwsze zmienne:



Współrzędne równoległe

Parallel coordinates

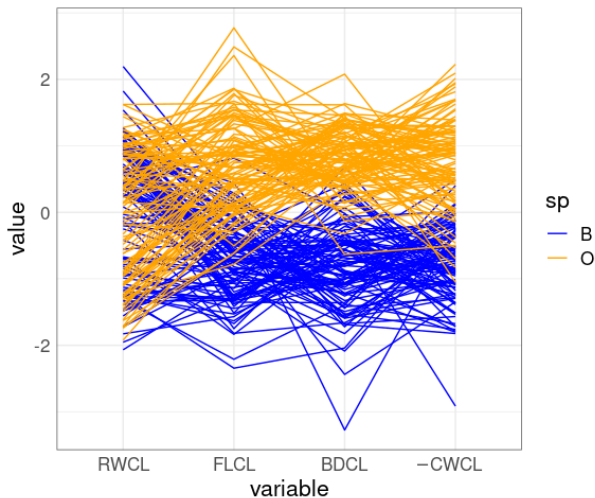
Zamienić miejscami dwie ostatnie zmienne:



Współrzędne równoległe

Parallel coordinates

Zmienić znak ostatniej zmiennej:

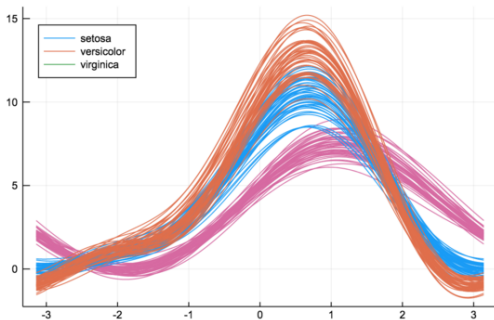


Krzywe Andrewsa

Andrews curves

Mając dane $x = \{x_i\}_{i=1}^d \in \mathbb{R}^d$, definiujemy dla $-\pi < t < \pi$:

$$f_x(t) = x_1/\sqrt{2} + x_2 \sin(t) + x_3 \cos(t) + x_4 \sin(2t) + x_5 \cos(2t) + \dots$$

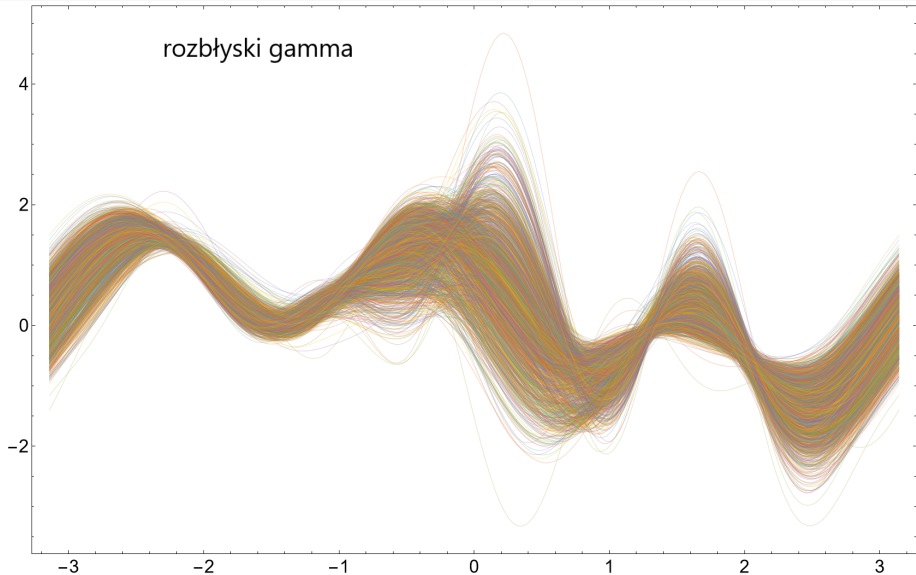


Jest to transformata Fouriera współrzędnych równoległych.

https://en.m.wikipedia.org/wiki/Andrews_plot

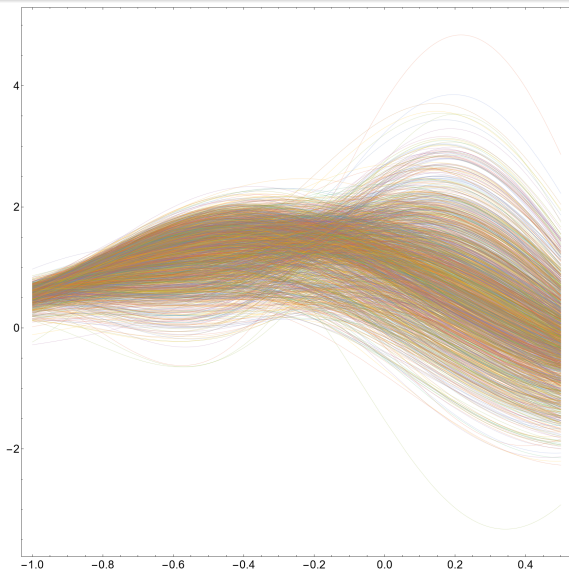
Krzywe Andrewsa

Andrews curves



Krzywe Andrewsa

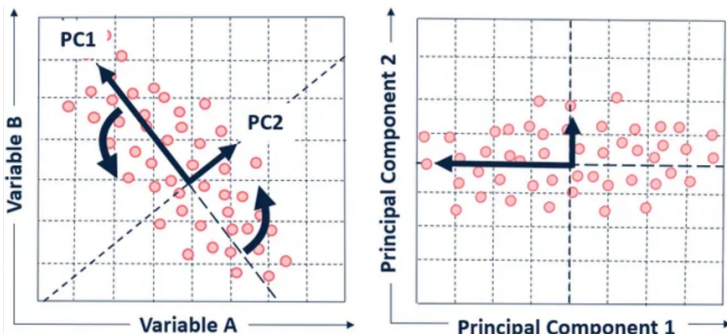
Andrews curves



Redukcja wymiaru

Dimensionality reduction

- PCA^a (Principal Component Analysis — analiza głównych składowych) — kombinacja liniowa maksymalizująca wariancję kilku pierwszych składowych; scree plot* (zob. slajd 25)



^aDunteman. Quantitative Applications in the Social Sciences: Principal Components Analysis; *https://en.m.wikipedia.org/wiki/Scree_plot

Redukcja wymiaru

Dimensionality reduction

- t -SNE^b (t -distributed Stochastic Neighbour Embedding — stochastyczna metoda porządkowania sąsiadów w oparciu o rozkład t) — bazuje na wielowymiarowym prawdopodobieństwie, że pewne dwa punkty będą sąsiadami. Wolnym parametrem jest tzw. perplexity określające rozmiar sąsiedztwa.

Algorytm probabilistyczny z natury — każde wywołanie prowadzi do innej reprezentacji zredukowanej przestrzeni. Nieinterpretowalne osie (zob. slajd 27).

^b<https://m.youtube.com/watch?v=NEaUSP4YerM>

^c<https://m.youtube.com/watch?v=eN0wFzBA4Sc>

Redukcja wymiaru

Dimensionality reduction

- t -SNE^b (t -distributed Stochastic Neighbour Embedding — stochastyczna metoda porządkowania sąsiadów w oparciu o rozkład t) — bazuje na wielowymiarowym prawdopodobieństwie, że pewne dwa punkty będą sąsiadami. Wolnym parametrem jest tzw. perplexity określające rozmiar sąsiedztwa.

Algorytm probabilistyczny z natury — każde wywołanie prowadzi do innej reprezentacji zredukowanej przestrzeni. Nieinterpretowalne osie (zob. slajd 27).

- UMAP^c (Uniform Manifold Approximation and Projection — brak oficjalnego pol. tłumaczenia) — reprezentacja wielowymiarowych danych poprzez tzw. macierz podobieństwa; następnie konstrukcja niskowymiarowej przestrzeni, w której zachowywane są odległości między punktami w macierzy podobieństwa.

Algorytm topologiczny — bardziej deterministyczny niż t -SNE.

^b<https://m.youtube.com/watch?v=NEaUSP4YerM>

^c<https://m.youtube.com/watch?v=eN0wFzBA4Sc>

Redukcja wymiaru

PCA

	$\log T_{50}$	$\log T_{90}$	$\log F_1$	$\log F_2$	$\log F_3$	$\log F_4$	$\log P_{64}$	$\log P_{256}$	$\log P_{1024}$
$\log T_{50}$	1	0.97	0.77	0.77	0.69	0.46	-0.18	-0.02	0.28
$\log T_{90}$	0.97	1	0.81	0.79	0.72	0.48	-0.13	0.03	0.33
$\log F_1$	0.77	0.81	1	0.97	0.9	0.62	0.3	0.44	0.68
$\log F_2$	0.77	0.79	0.97	1	0.94	0.66	0.35	0.49	0.73
$\log F_3$	0.69	0.72	0.9	0.94	1	0.77	0.46	0.59	0.79
$\log F_4$	0.46	0.48	0.62	0.66	0.77	1	0.44	0.53	0.64
$\log P_{64}$	-0.18	-0.13	0.3	0.35	0.46	0.44	1	0.97	0.83
$\log P_{256}$	-0.02	0.03	0.44	0.49	0.59	0.53	0.97	1	0.93
$\log P_{1024}$	0.28	0.33	0.68	0.73	0.79	0.64	0.83	0.93	1

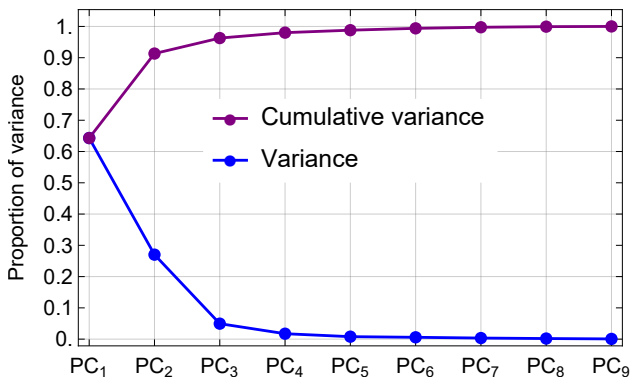
Macierz korelacji dla $n = 9$ zmiennych (rozbłyski gamma z katalogu BATSE).

Wektory własne macierzy korelacji są współczynnikami kombinacji liniowych zmiennych (z katalogu BATSE) tworząc składowe główne:

$$\left(\begin{array}{l} 0.39 \log F_1 + 0.4 \log F_2 + 0.4 \log F_3 + 0.33 \log F_4 + 0.36 \log P_{1024} + 0.28 \log P_{256} + 0.22 \log P_{64} + 0.29 \log T_{50} + 0.3 \log T_{90} \\ -0.15 \log F_1 - 0.12 \log F_2 - 0.04 \log F_3 + 0.05 \log F_4 + 0.3 \log P_{1024} + 0.47 \log P_{256} + 0.53 \log P_{64} - 0.44 \log T_{50} - 0.41 \log T_{90} \\ -0.22 \log F_1 - 0.19 \log F_2 + 0.06 \log F_3 + 0.92 \log F_4 - 0.17 \log P_{1024} - 0.13 \log P_{256} - 0.1 \log P_{64} - 0.06 \log T_{50} - 0.08 \log T_{90} \\ 0.48 \log F_1 + 0.41 \log F_2 + 0.21 \log F_3 + 0.03 \log F_4 - 0.13 \log P_{1024} - 0.24 \log P_{256} - 0.22 \log P_{64} - 0.48 \log T_{50} - 0.46 \log T_{90} \\ 0.56 \log F_1 - 0.05 \log F_2 - 0.77 \log F_3 + 0.19 \log F_4 - 0.08 \log P_{1024} + 0.06 \log P_{256} + 0.15 \log P_{64} - 0.01 \log T_{50} + 0.12 \log T_{90} \\ 0. + 0.01 \log F_1 + 0.11 \log F_2 + 0.19 \log F_3 - 0.06 \log F_4 - 0.75 \log P_{1024} - 0.03 \log P_{256} + 0.61 \log P_{64} + 0.13 \log T_{50} \\ -0.03 \log F_1 - 0.05 \log F_2 + 0.11 \log F_3 - 0.02 \log F_4 - 0.06 \log P_{1024} - 0.04 \log P_{256} + 0.04 \log P_{64} - 0.69 \log T_{50} + 0.71 \log T_{90} \\ 0.49 \log F_1 - 0.78 \log F_2 + 0.38 \log F_3 - 0.07 \log F_4 - 0.03 \log P_{1024} + 0.04 \log P_{256} - 0.04 \log P_{64} + 0.04 \log T_{50} - 0.06 \log T_{90} \\ 0. + 0.01 \log F_1 - 0.08 \log F_2 - 0.01 \log F_3 + 0.4 \log P_{1024} - 0.79 \log P_{256} + 0.47 \log P_{64} + 0.03 \log T_{50} - 0.01 \log T_{90} \end{array} \right)$$

Redukcja wymiaru

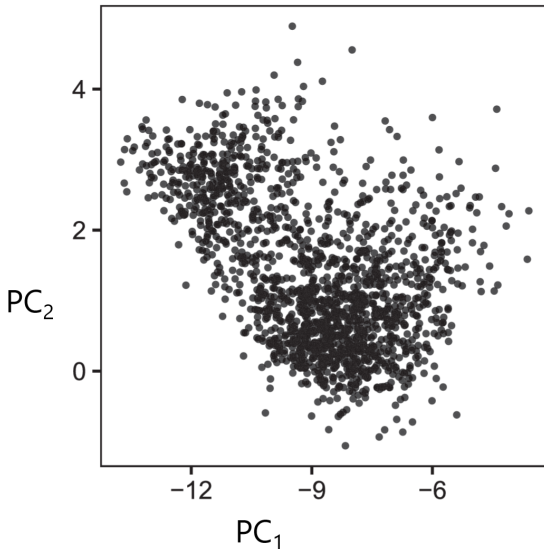
PCA — scree plot (wykres osypiska)



Widoczne jest przegięcie na wykresie wariancji po 2 składowych głównych. Sumaryczna wariancja przekracza 90% również dla 2 składowych. Można zatem zredukować 9-wymiarową przestrzeń parametrów do zaledwie 2 wymiarów:

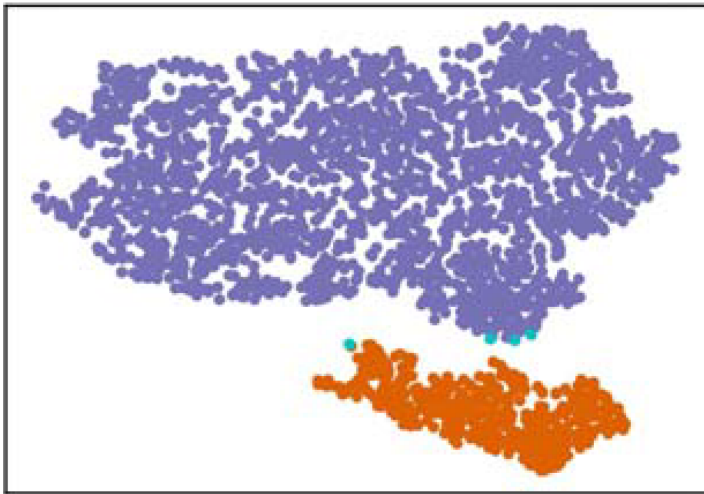
Redukcja wymiaru

PCA



Redukcja wymiaru

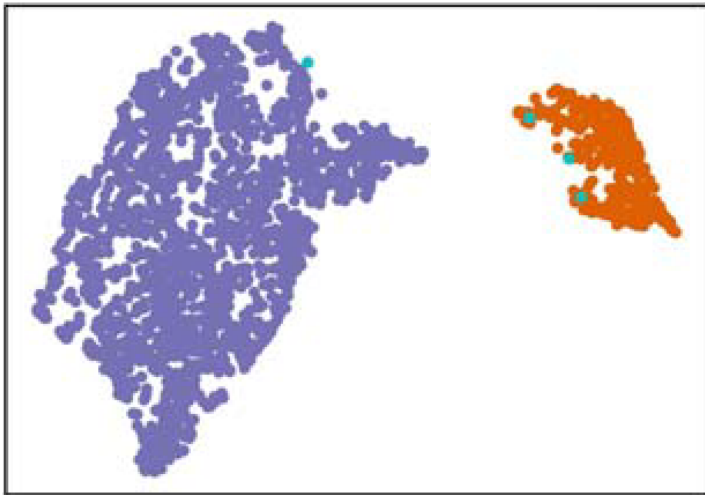
t -SNE



<https://doi.org/10.3847/1538-4357/acb999>

Redukcja wymiaru

UMAP



<https://doi.org/10.3847/1538-4357/acb999>