



Vizualizace dat

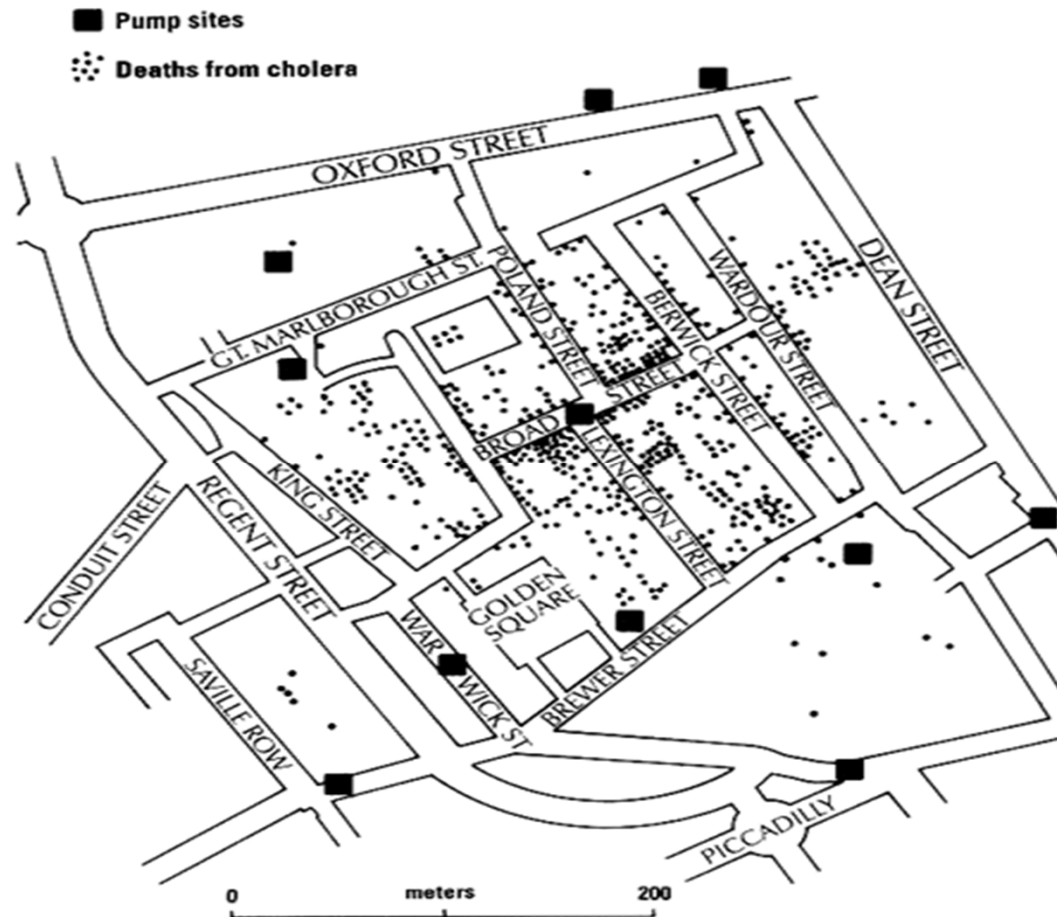
Lenka Vysloužilová



Historie



❖ 1855 – Dr. John Snow – dot map of cholera



Osnova



- ❖ Role vizualizace
- ❖ Klasické grafy, mapy
- ❖ Statistika – box graf, histogram
- ❖ Multidimensionální vizualizace dat

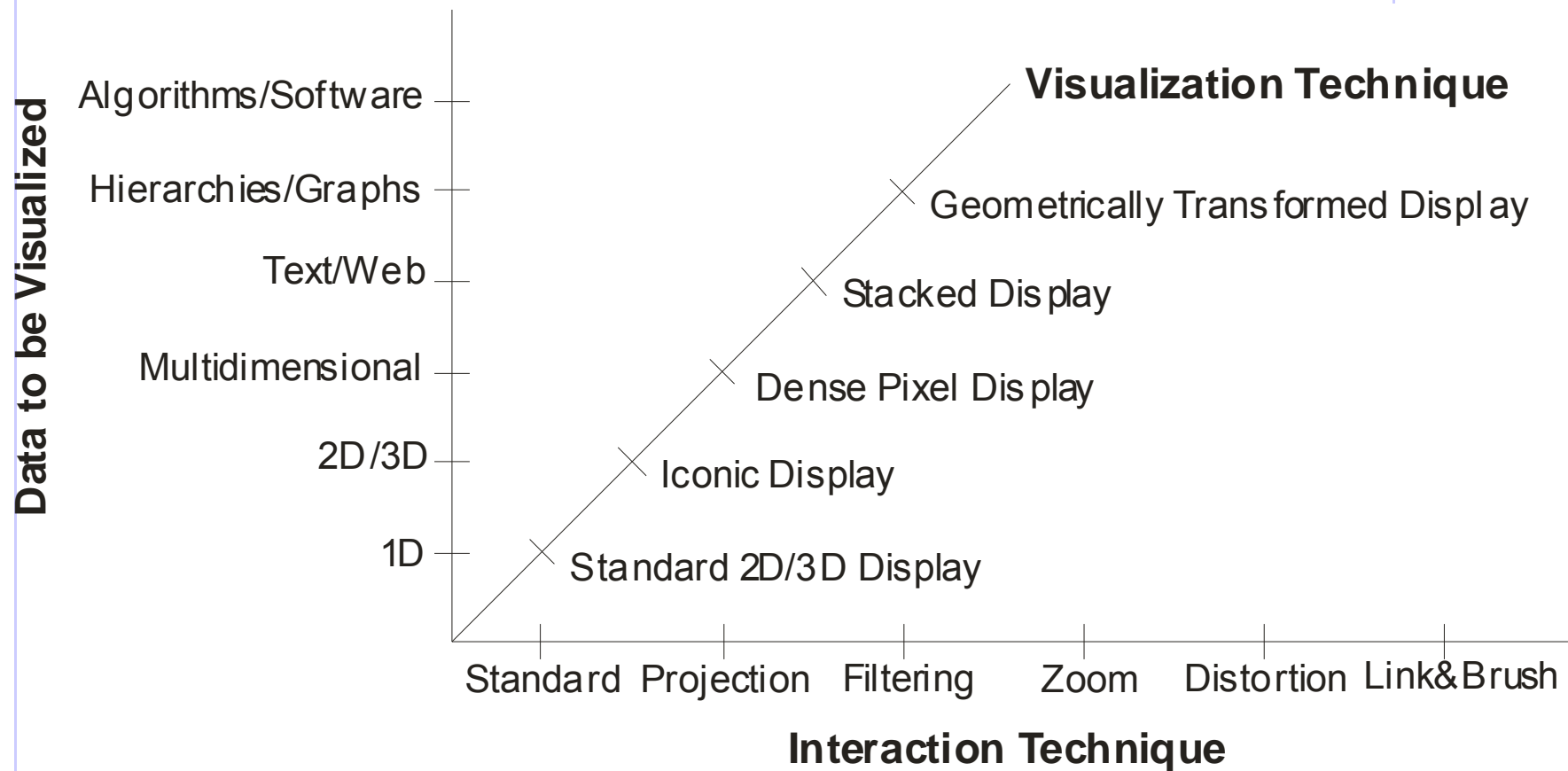
Role vizualizace



- ❖ Pomoc při prezentaci výsledků
- ❖ Pomoc při průzkumu velkého množství dat
- ❖ Možná interakce s uživatelem

- ❖ Potřeba lidského oka a jeho zkušeností
- ❖ **Pozor:** vizualizace může být zavádějící

Information visualization



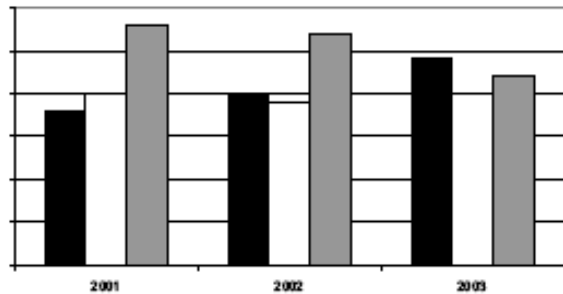
D.A. Keim. Information visualization and visual data mining. IEEE Transactions on Visualization and Computer Graphics, 8(1):1--8, 2002.

2D

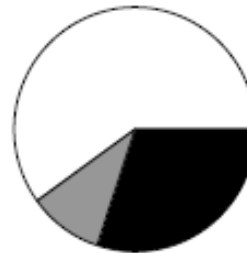


- ❖ sloupcové grafy
- ❖ koláčové grafy
- ❖ X-Y grafy

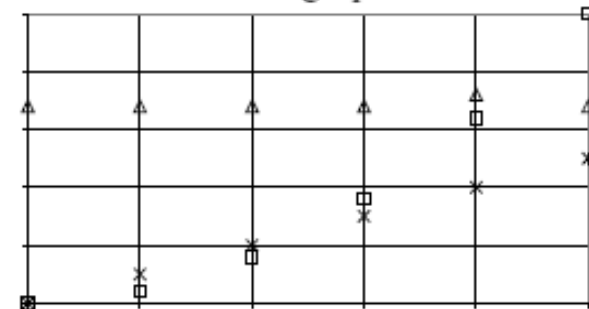
Bar graph



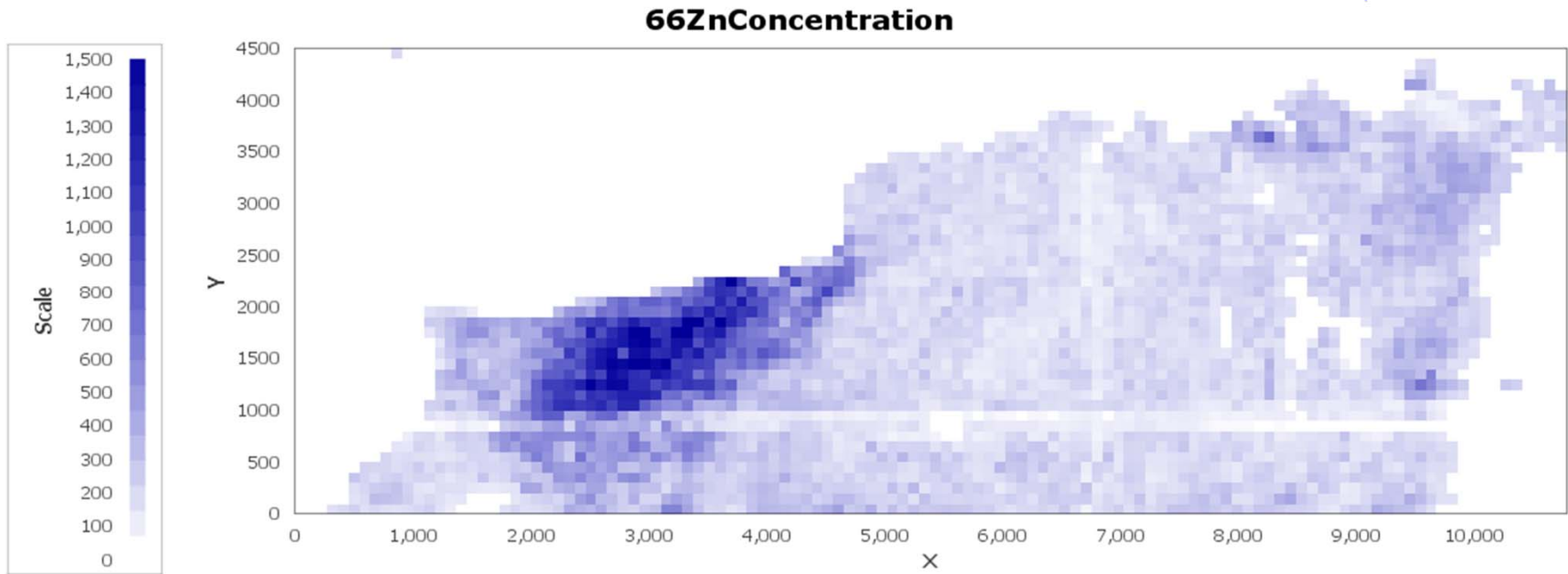
Pie graph



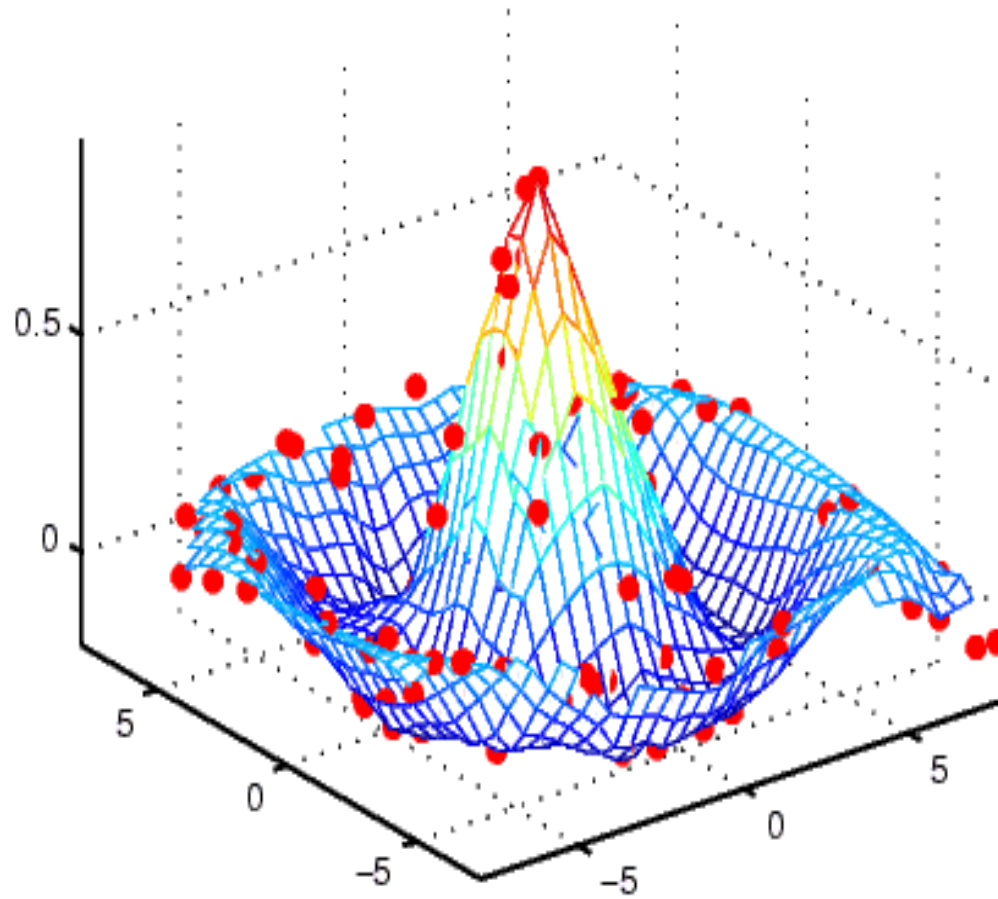
XY graph



Mapy



3D



Iris Data



Iris setosa

sepal length	sepal width	petal length	petal width
5.1	3.5	1.4	0.2
4.9	3	1.4	0.2
...
5.9	3	5.1	1.8

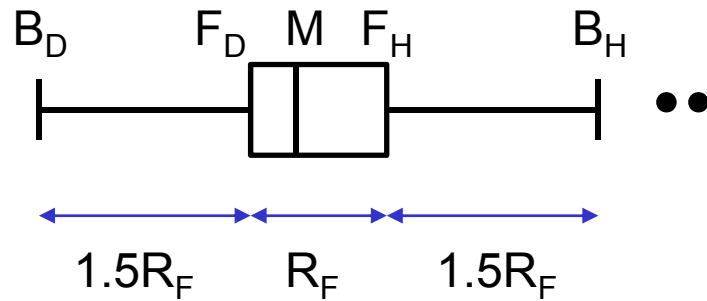


Iris versicolor



Iris virginica

Box graf – Quartile Color



M – medián

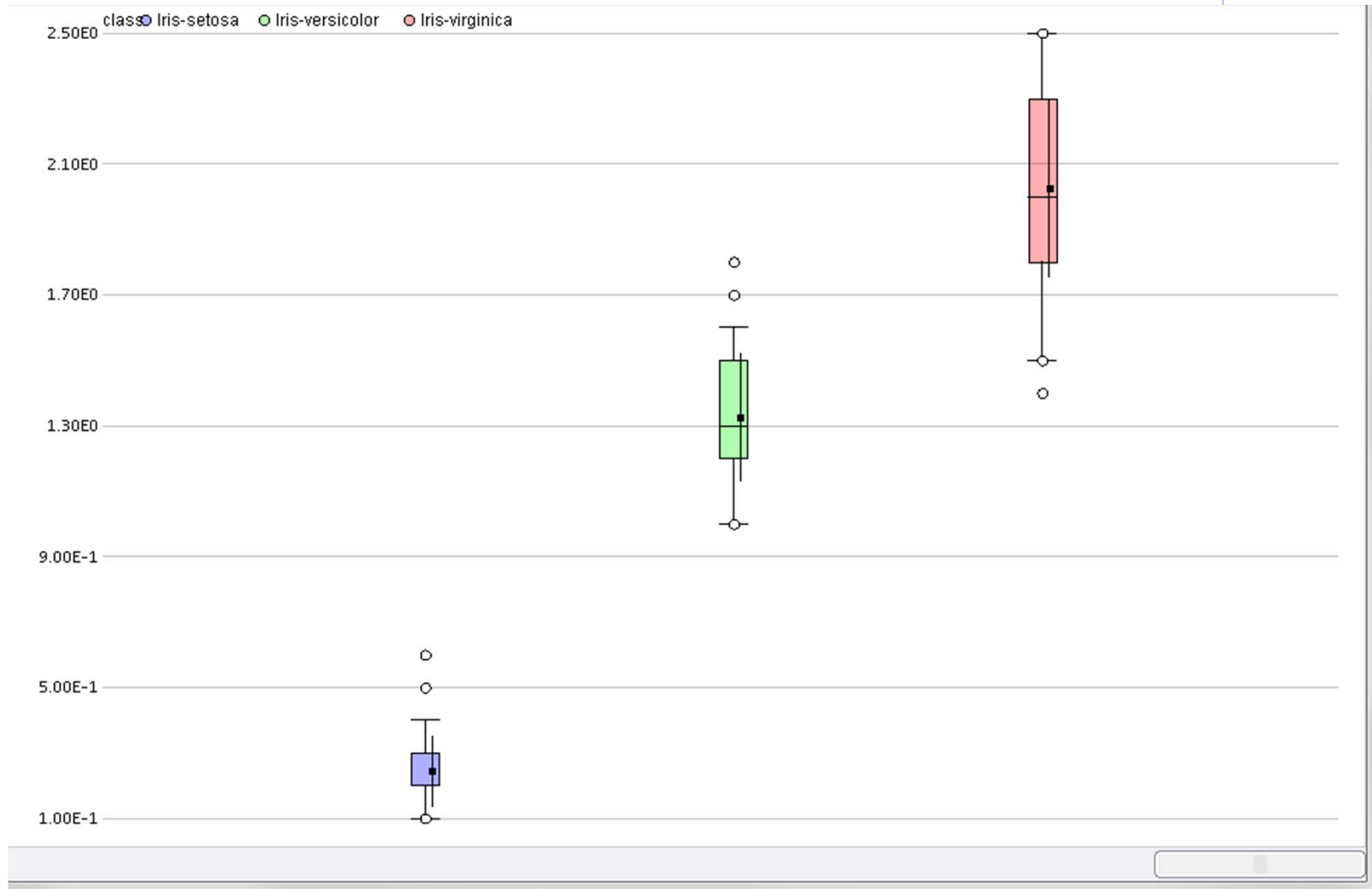
F – horní a dolní kvartil

$$R_F = F_H - F_D$$

$$B_D = F_D - 1.5R_F$$

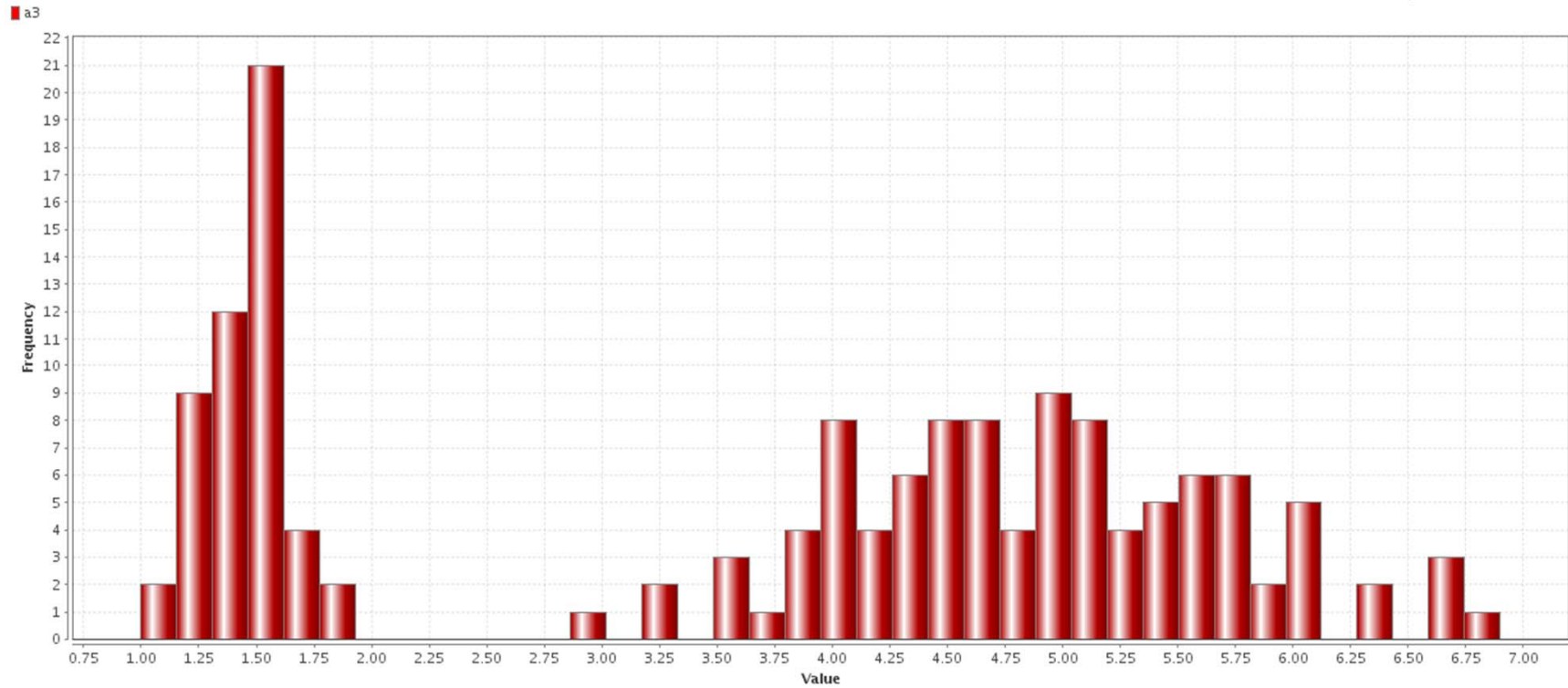


Box graf – Quartile Color





Histogram

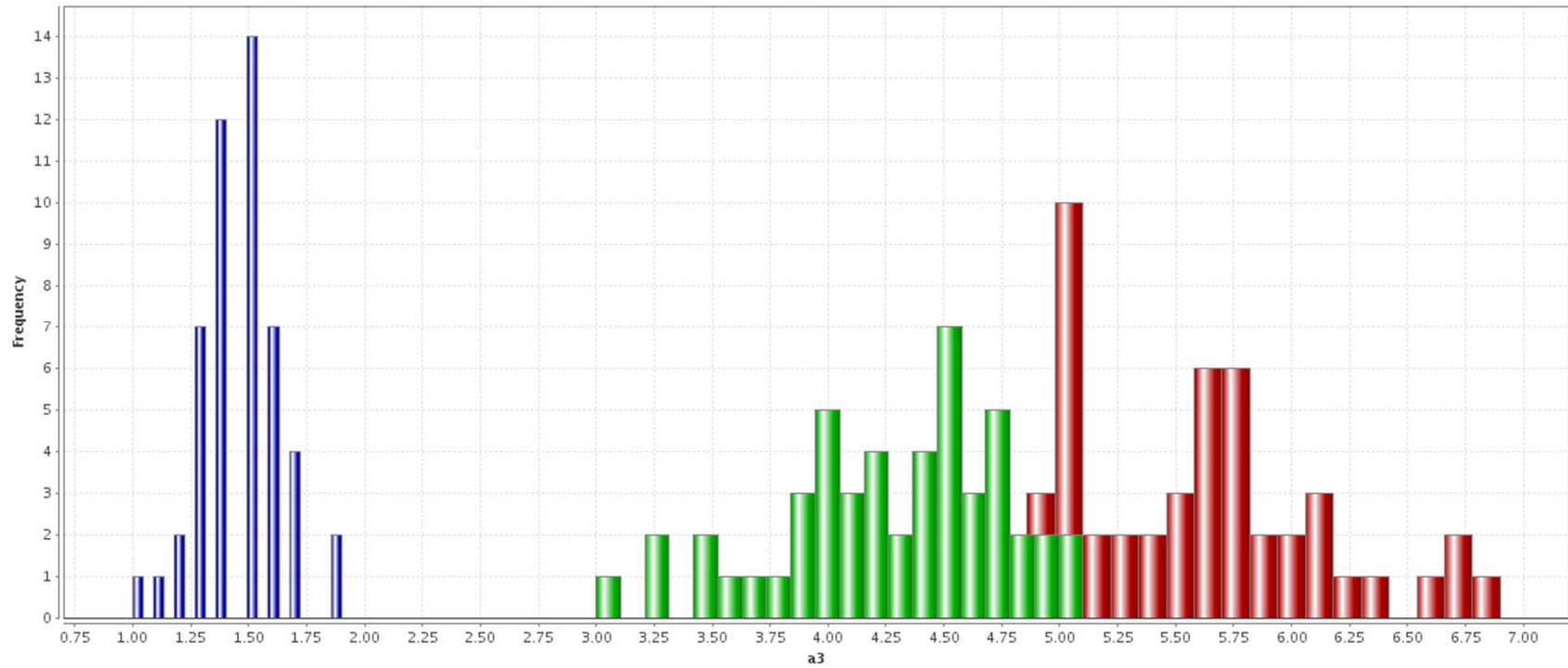




Histogram

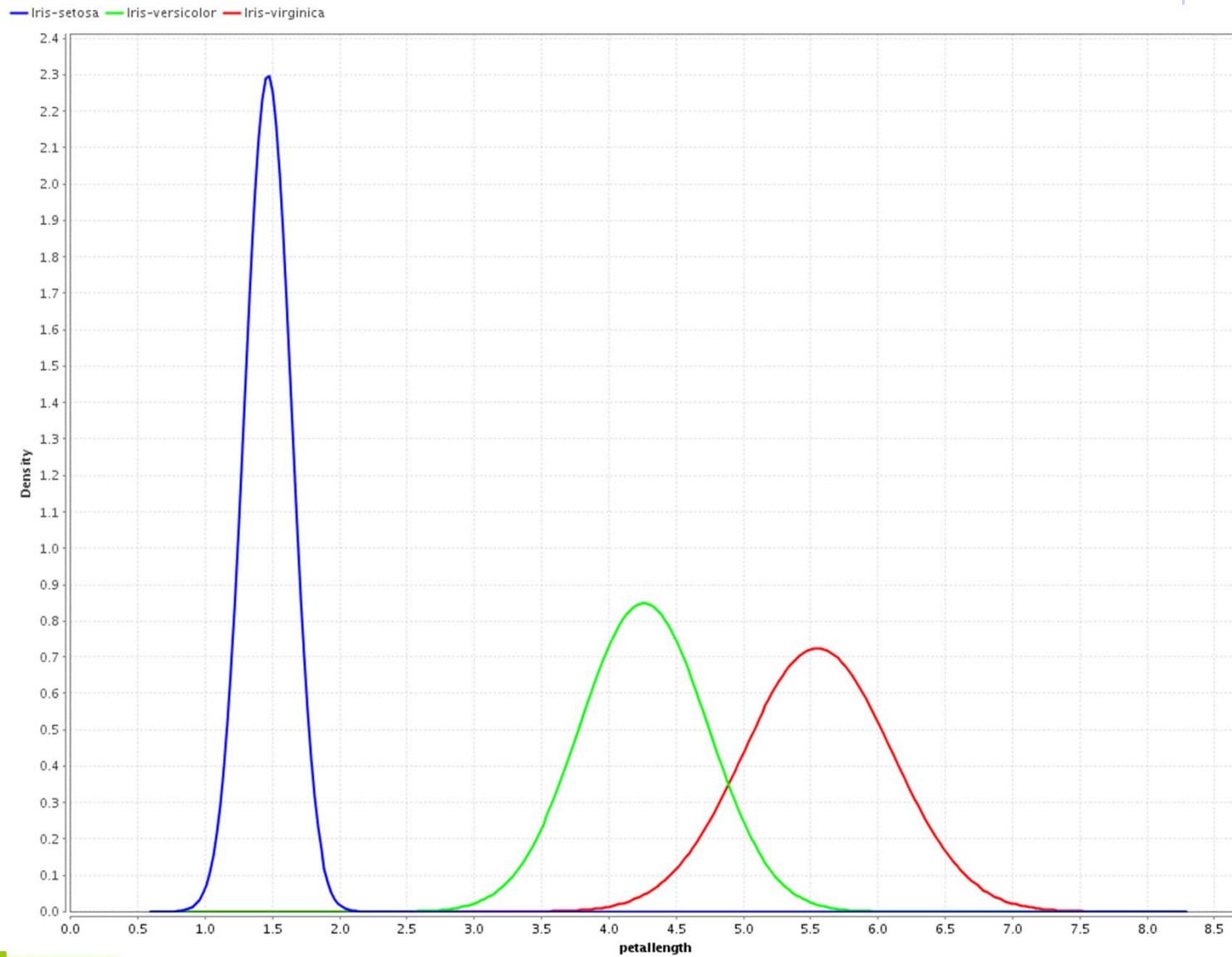


■ Iris-setosa ■ Iris-versicolor ■ Iris-virginica





Distribution

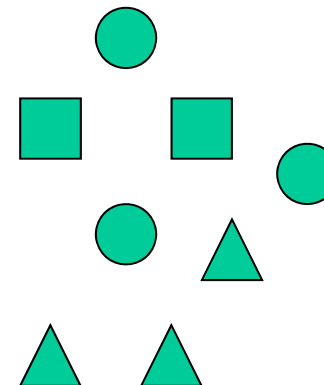
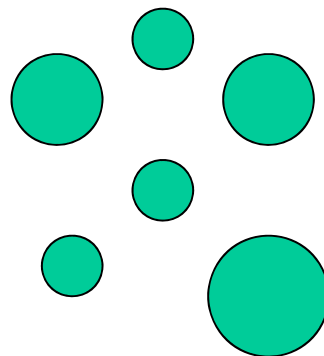
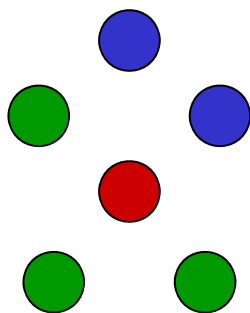




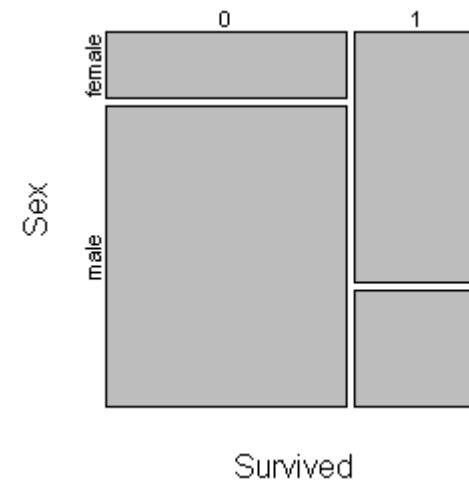
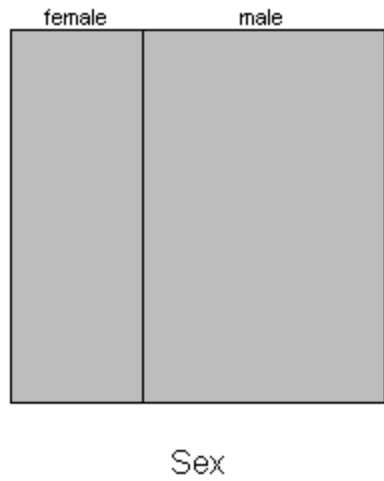
Multidimensionální vizualizace dat



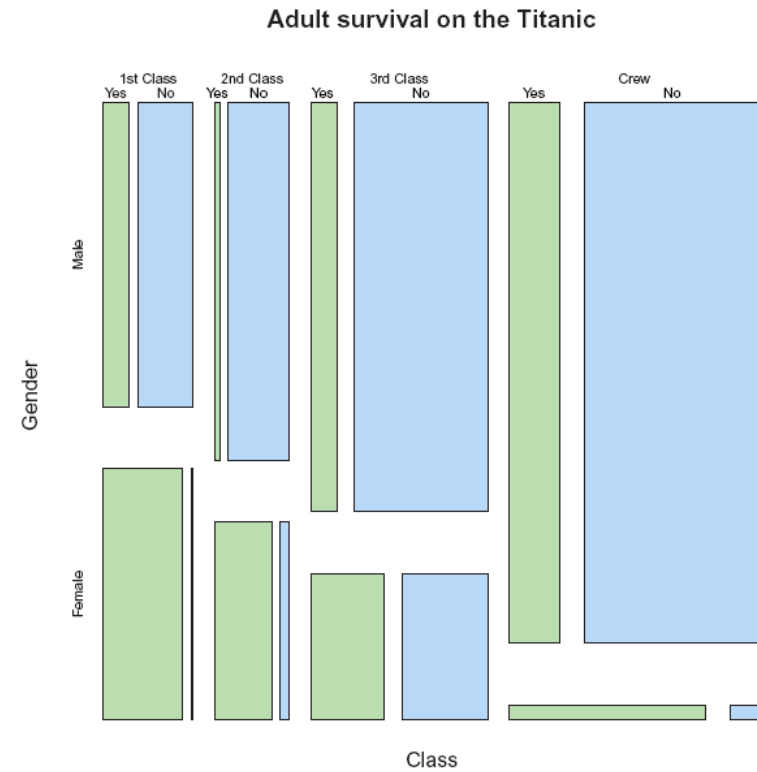
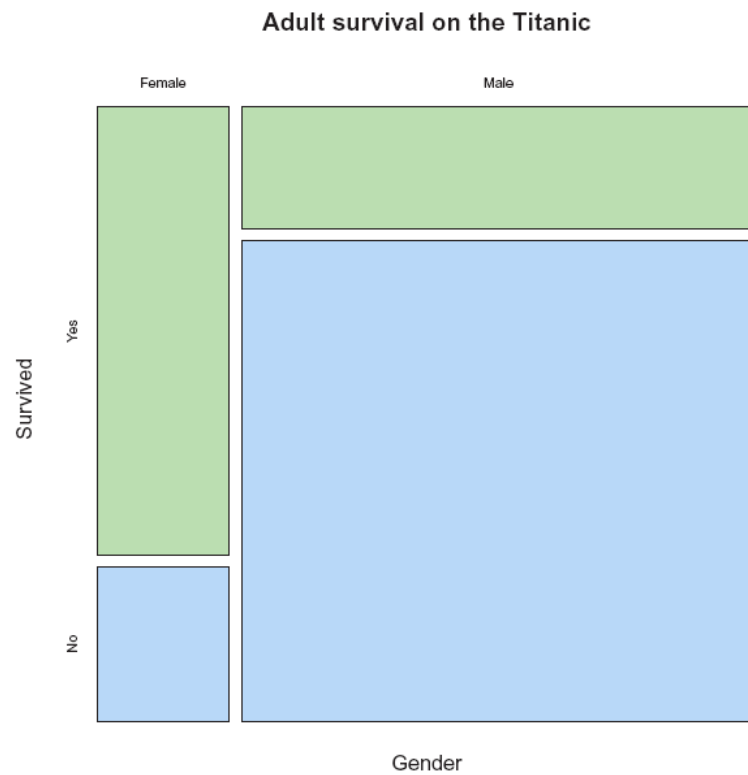
- ❖ člověk je omezen světem kolem sebe na 3D vnímání
- ❖ $nD \rightarrow 2D$ (3D)
- ❖ další dimenze nahrazujeme
 - ◆ přidáním barev, velikostí objektů a tvarem objektů
 - ◆ transformací dat
 - ◆ projekcí dat



Mosaic Plots



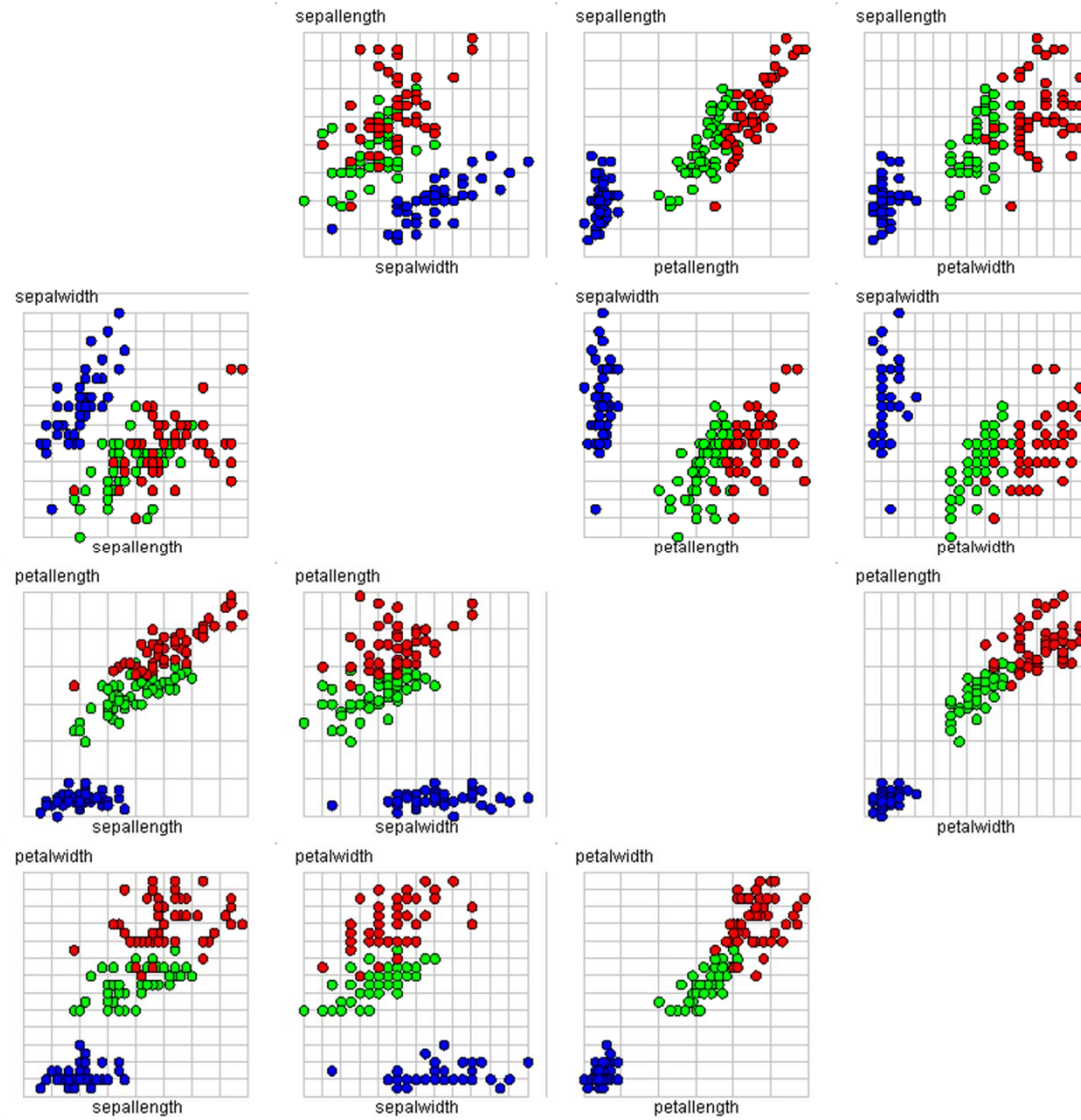
Mosaic Plots



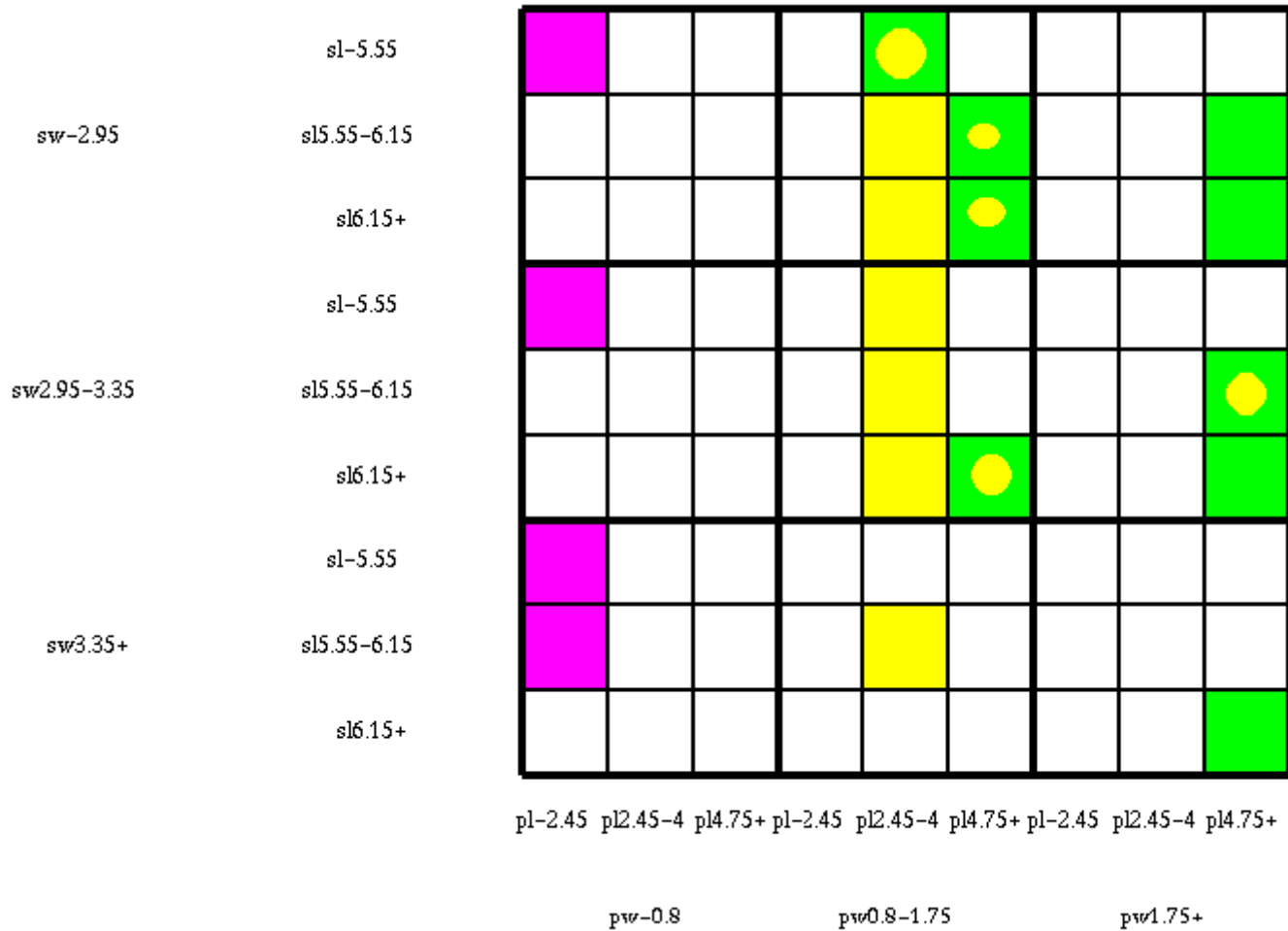
Scatter Plot matrix



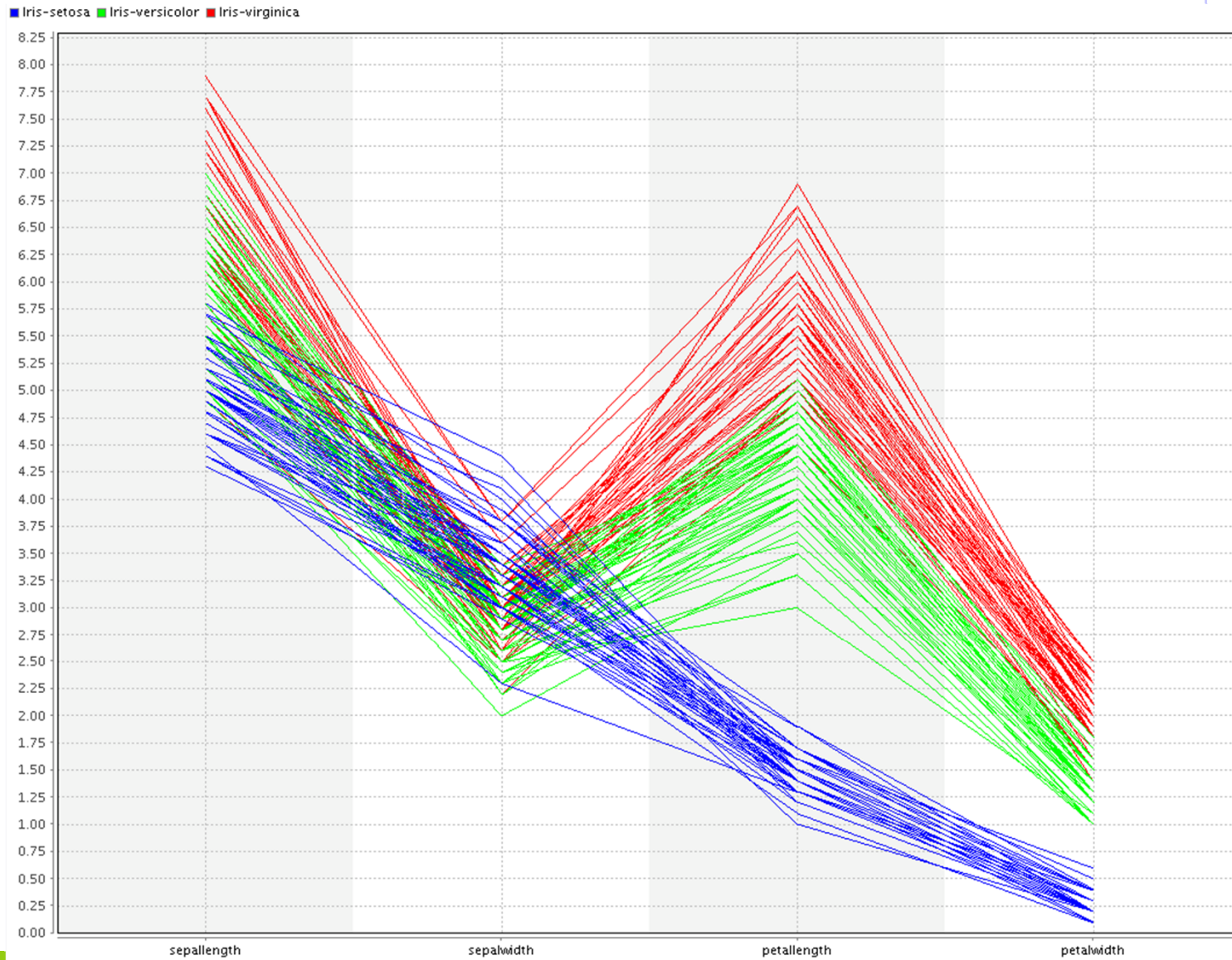
class: ● Iris-setosa ● Iris-versicolor ● Iris-virginica



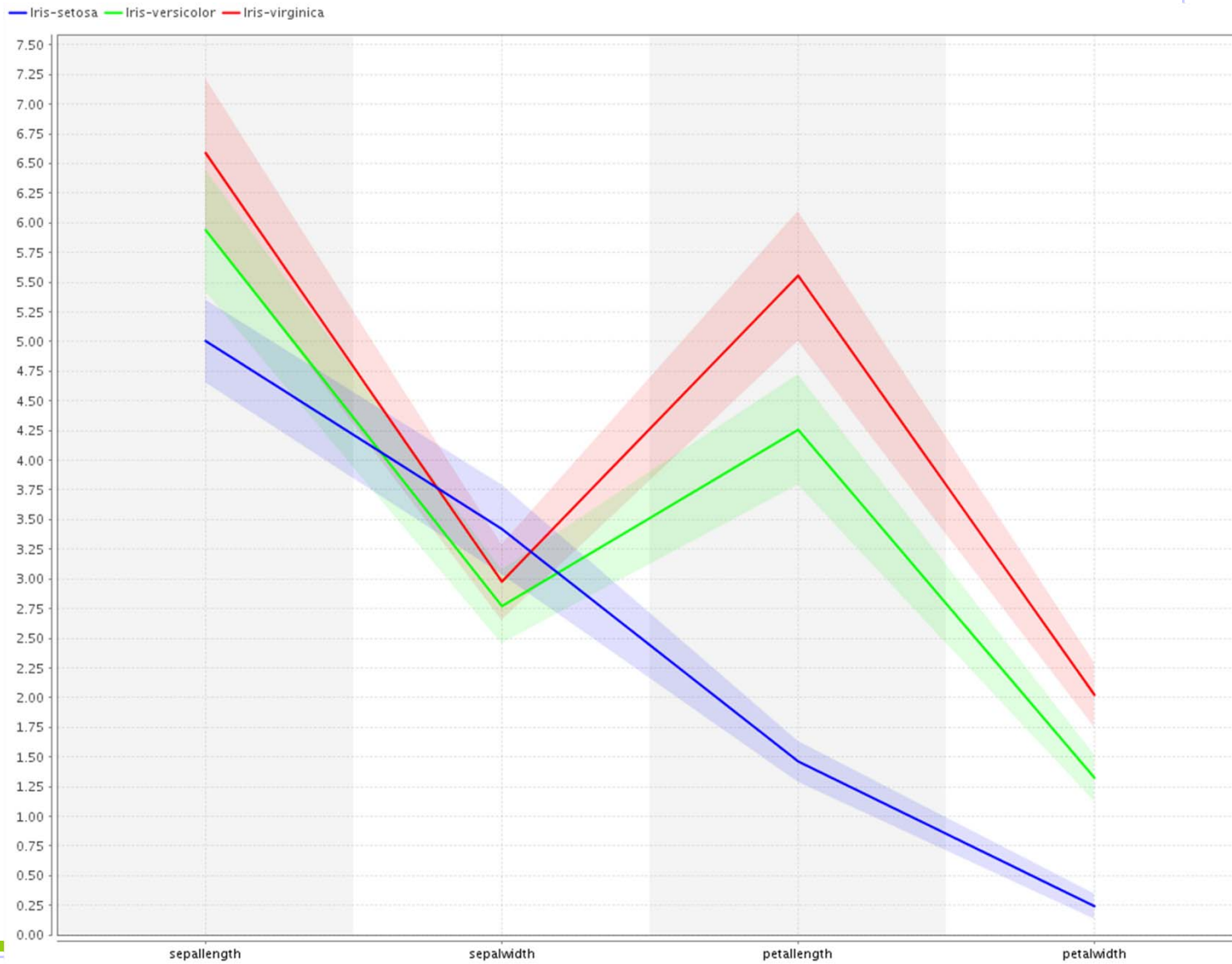
Stacked Displays



Parallel coordinates



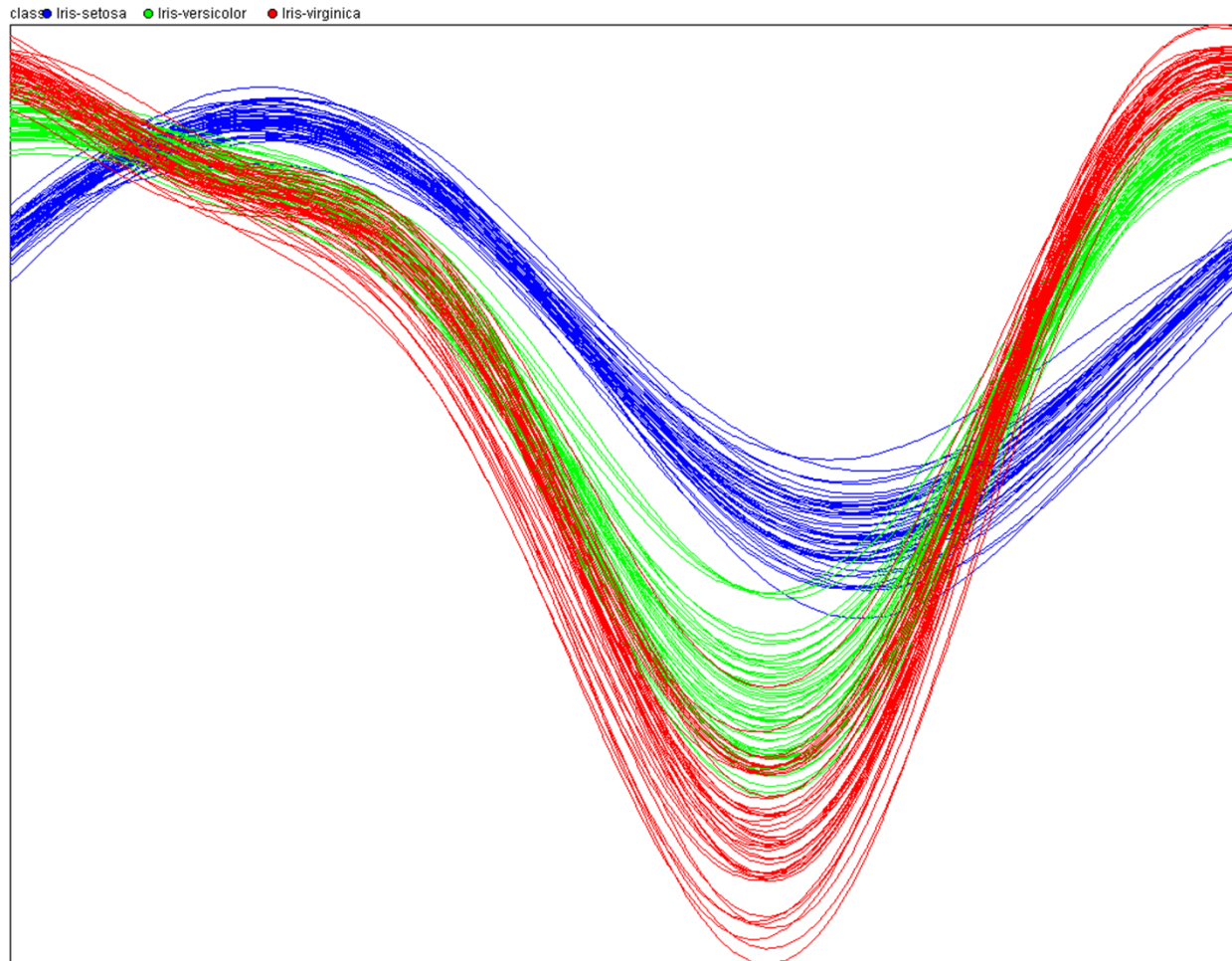
Deviation



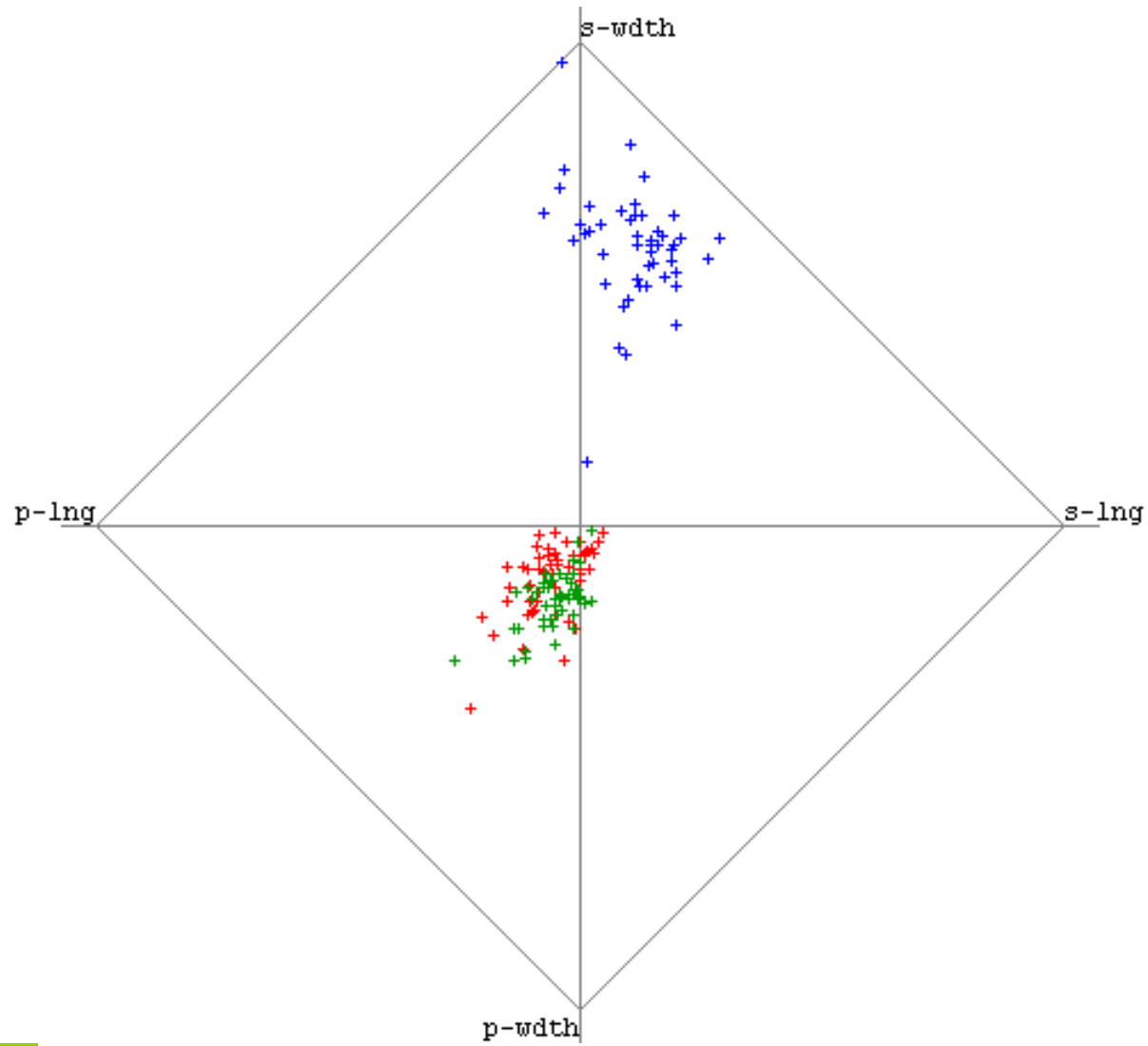
Andrews' Curves



$$f(t) = \frac{x_1}{\sqrt{2}} + x_2 \sin(t) + x_3 \cos(t) + x_4 \sin(2t) + x_5 \cos(2t) + \dots,$$

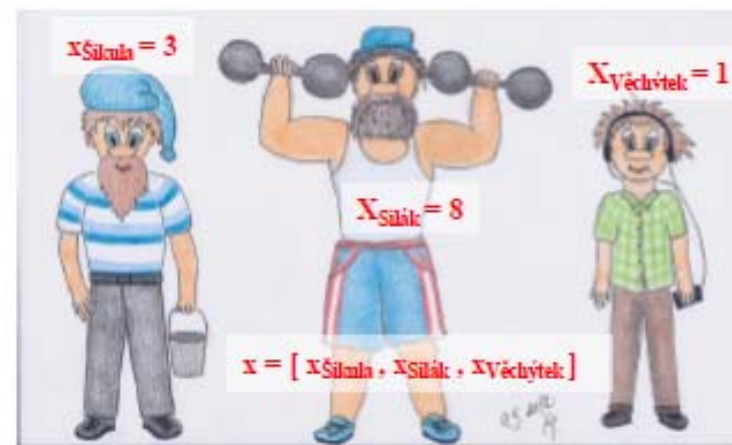
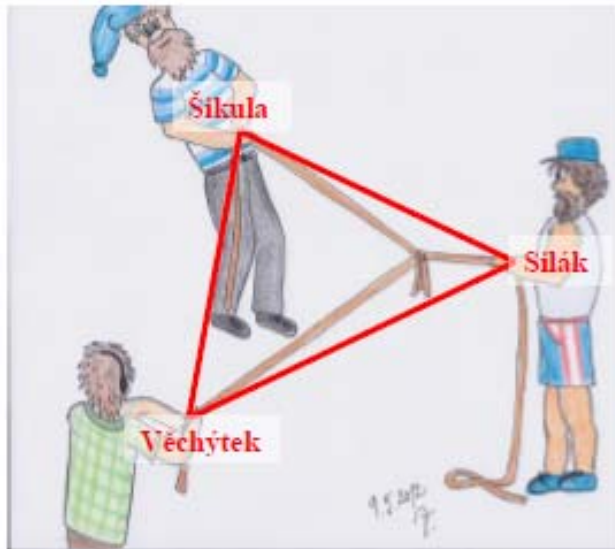


RadViz – Iris dataset



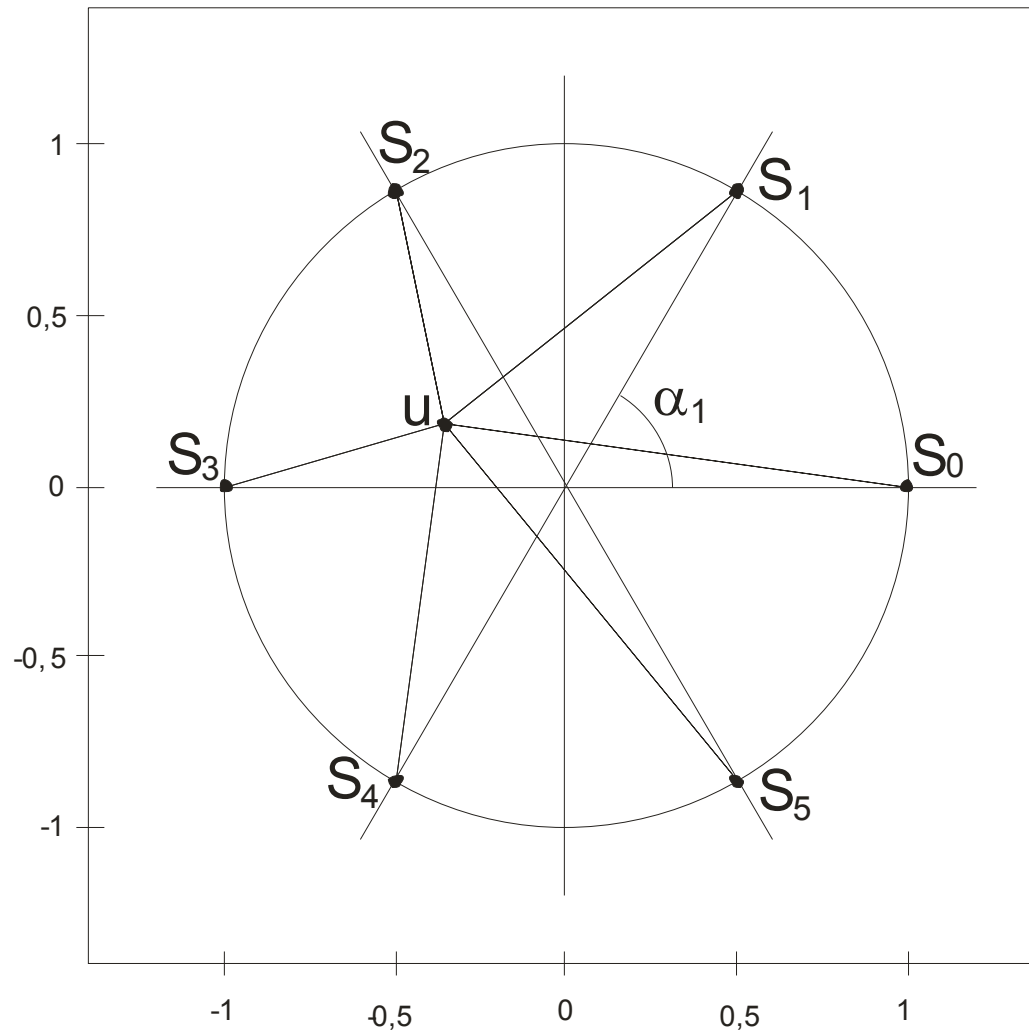


Valentová K.: Vizualizace multidimenzionálních dat metodou RadViz,
diplomová práce, 2012, vedoucí práce L.Vysloužilová



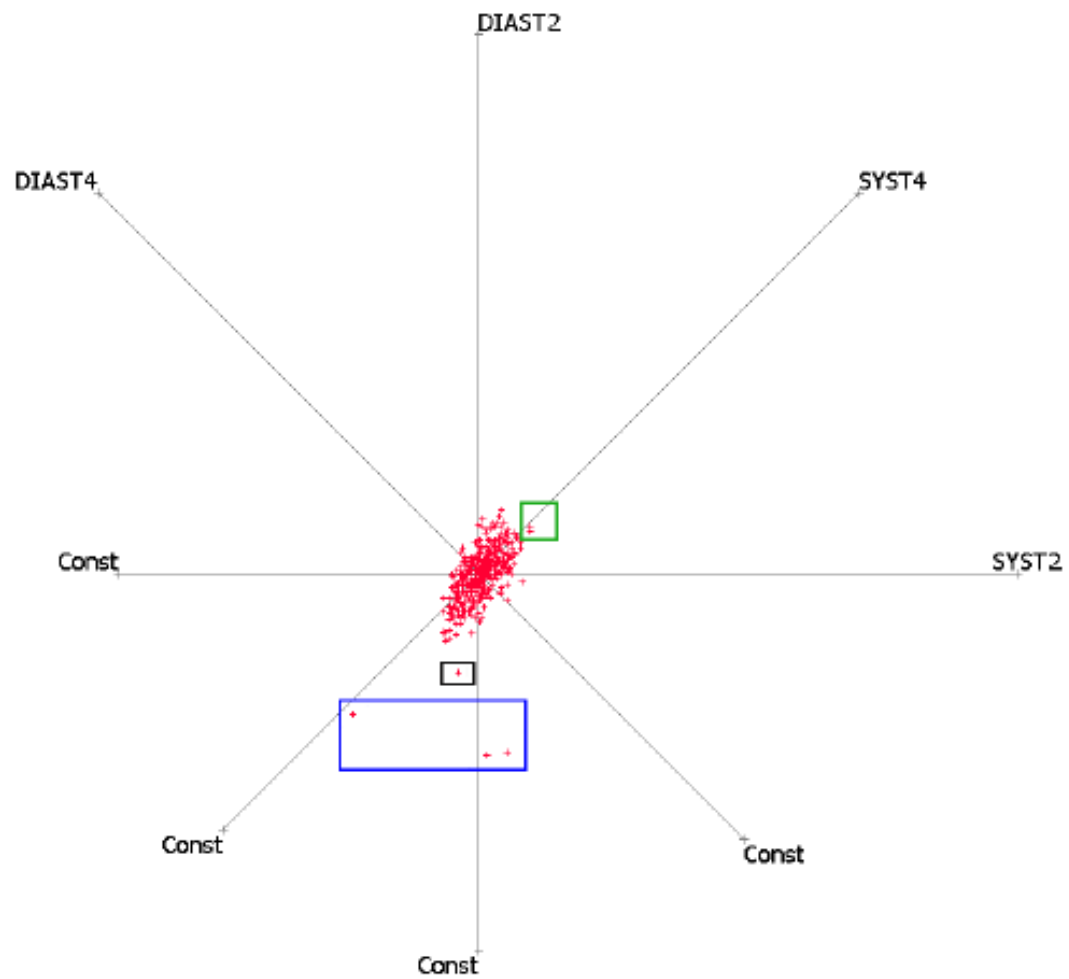
Iris data = 4 atributy,
tj. můžeme si představit čtyři přetahující se trpaslíky,
hodnoty jejich síly, budou z následující tabulky.

sepal length	sepal width	petal length	petal width
5.1	3.5	1.4	0.2
4.9	3	1.4	0.2
...
5.9	3	5.1	1.8





Identifikace odlehlých hodnot



Problémy

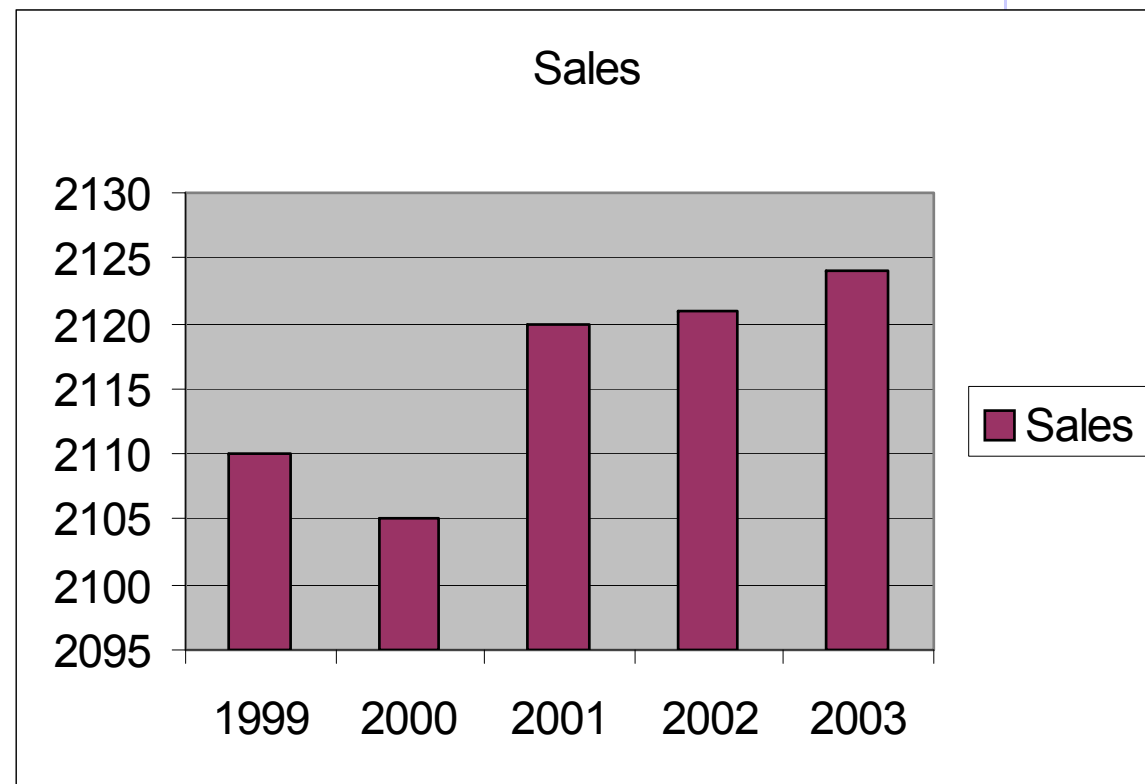


- ❖ Vliv měřítka
- ❖ Vliv setřídění a pořadí zobrazených dat
- ❖ Skryté shluky

Vliv měřítka



Year	Sales
1999	2110
2000	2105
2001	2120
2002	2121
2003	2124

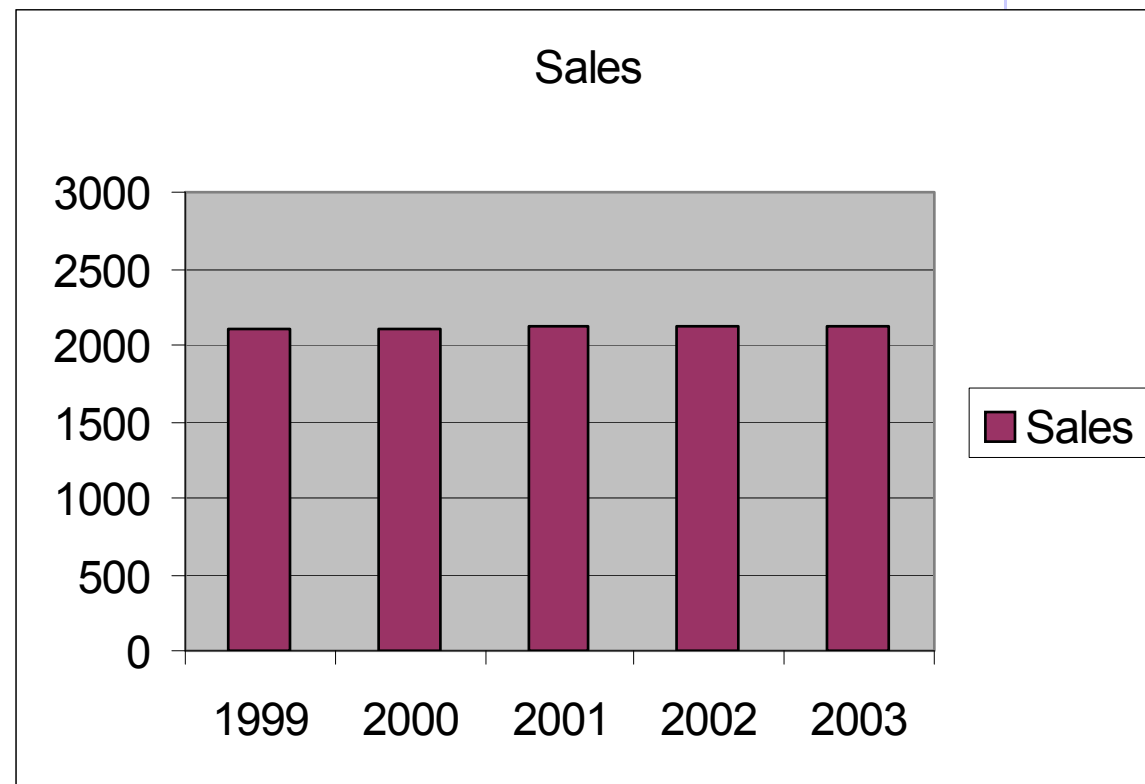


Volba špatného měřítka u osy Y
budí dojem velkých rozdílů mezi prodeji

Vliv měřítka

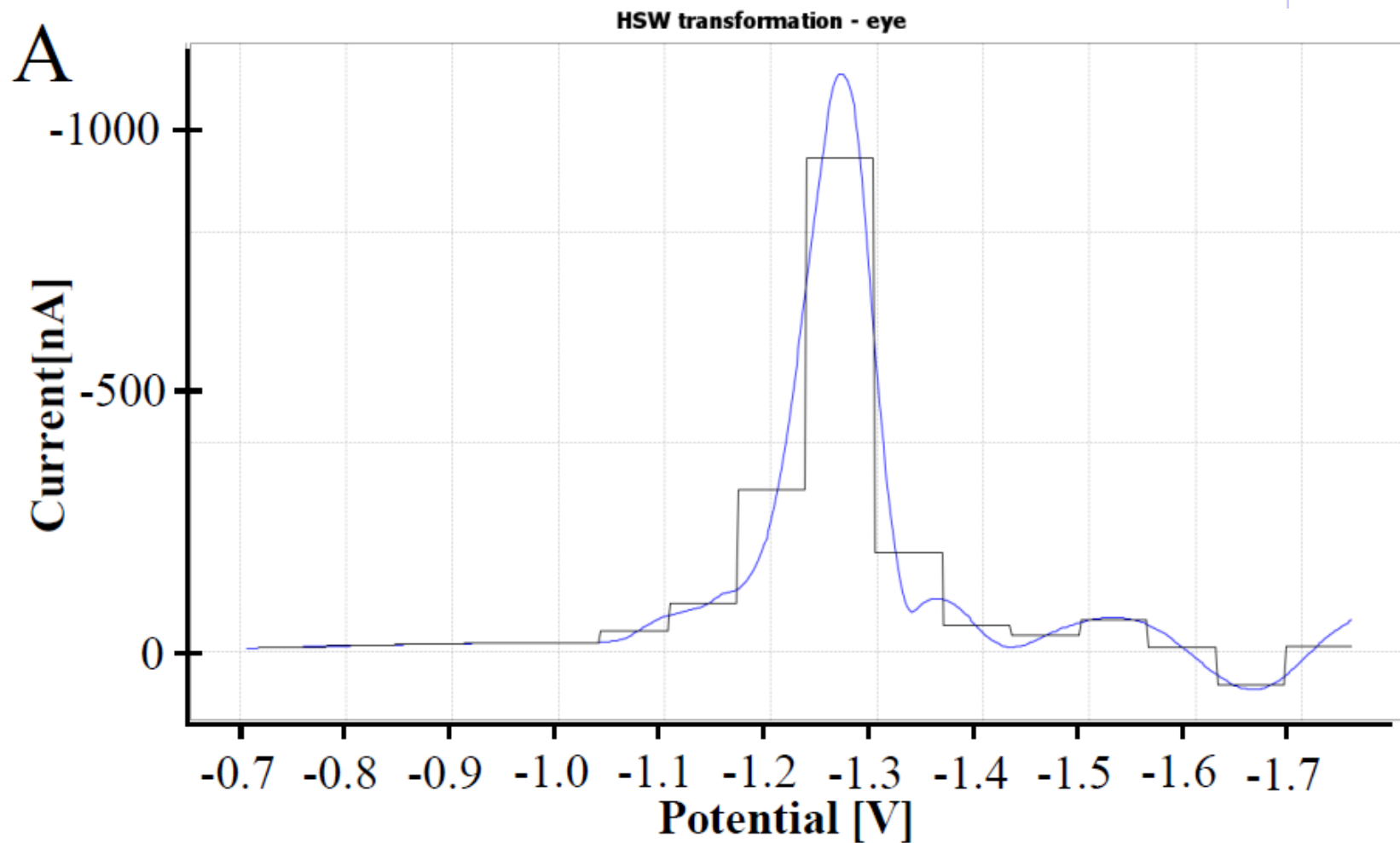


Year	Sales
1999	2110
2000	2105
2001	2120
2002	2121
2003	2124



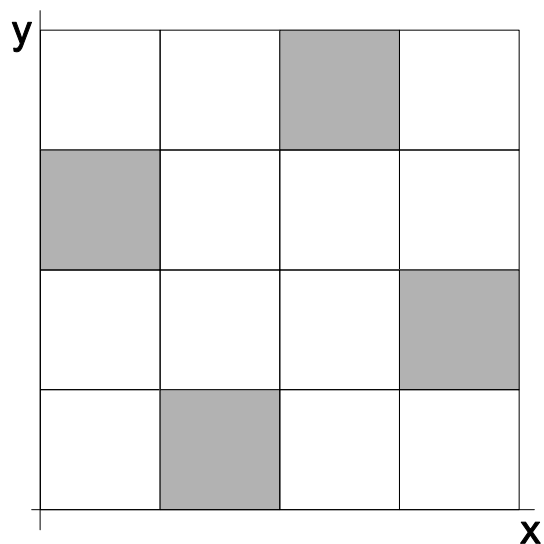
Volba měřítka od 0 do 2000 u osy Y
nyní ukazuje malé rozdíly mezi prodeji

Vliv setřídění a pořadí

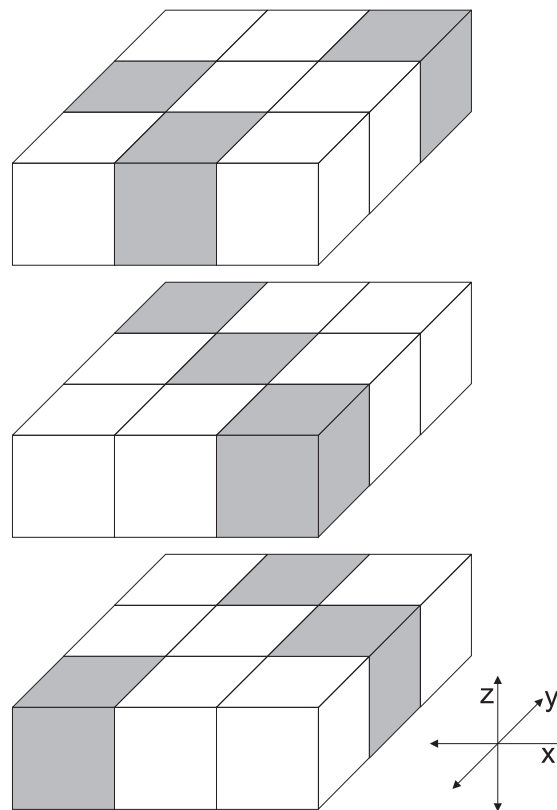


I uspořádání řady na osách může být doménově závislé!

Zobrazení shluků



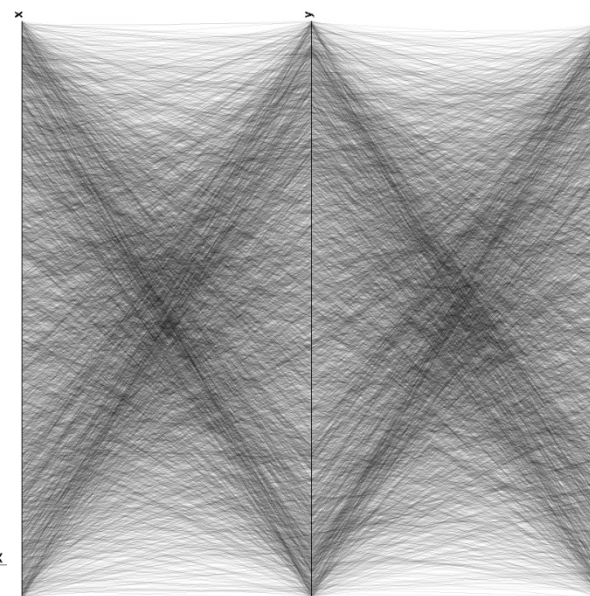
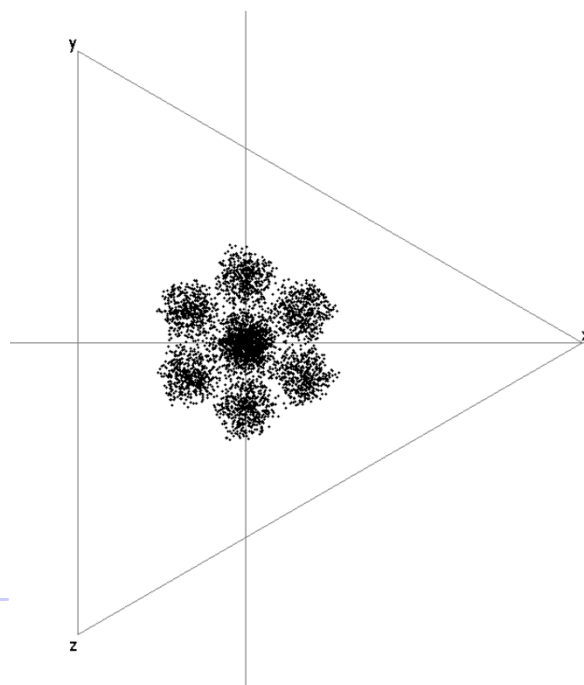
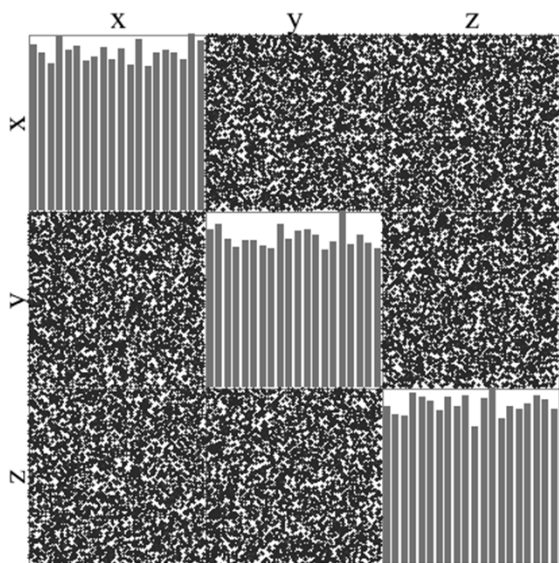
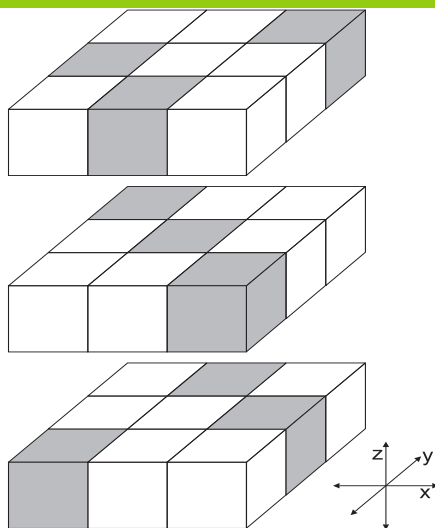
2D



3D

4D ...

Zobrazení shluků



† Závěrečné poznámky



- ❖ vizualizace je účinný nástroj pro průzkum dat, ale není všemocný
- ❖ nutnost zachovat zvyky obvyklé v dané komunitě uživatelů
- ❖ vždy je důležitá správná interpretace