

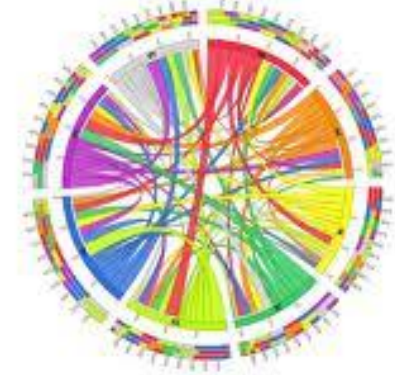
```

<?xml version="1.0" encoding="utf-8" ?>
<INFORMATIONS>
  <INFORMATION>
    <HEADER>
      <Name>Santosh.R</Name>
      <Domain>@AX.LOCAL</Domain>
    </HEADER>
    <ROWS>
      <ROW>
        <DETAILS>
          <ACCOUNT>108600</ACCOUNT>
          <COSTCENTER>112400</COSTCENTER>
          <DESCRIPTION>Row 1</DESCRIPTION>
        </DETAILS>
      </ROW>
    </ROWS>
  </INFORMATION>
  <INFORMATION>
    <HEADER>
      <Name>Anitha.E</Name>
      <Domain>@AX.LOCAL</Domain>
    </HEADER>
    <ROWS>
      <ROW>
        <DETAILS>
          <ACCOUNT>126556</ACCOUNT>
          <COSTCENTER>224212</COSTCENTER>
          <DESCRIPTION>Row 2</DESCRIPTION>
        </DETAILS>
      </ROW>
    </ROWS>
  </INFORMATION>
</INFORMATIONS>

```

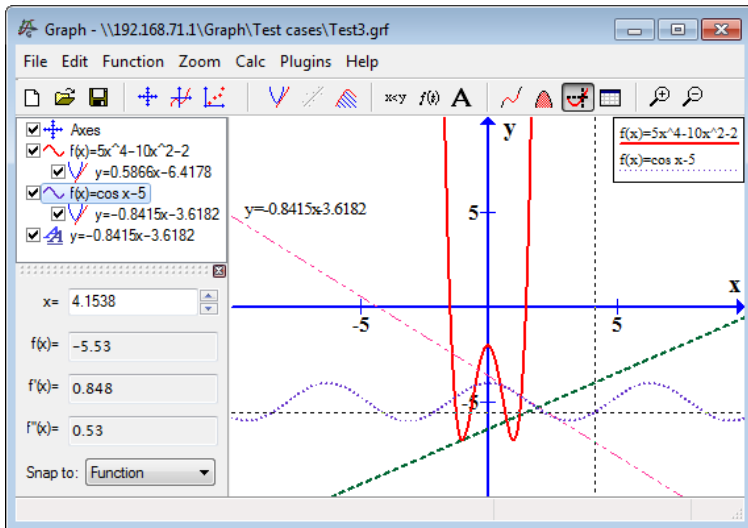
axblog4u.wordpress.com

	A	B	C	D	E	F	G	H
A	54	133	157	34	88	141	167	133
B	49	113	111	113	202	53	7	92
C	66	130	69	162	123	63	106	117
D	60	138	89	85	98	88	122	82
E	53	88	15	92	92	20	69	127
F	118	32	62	119	135	95	60	64
G	134	108	73	44	103	119	37	145
H	74	110	84	120	9	41	45	131

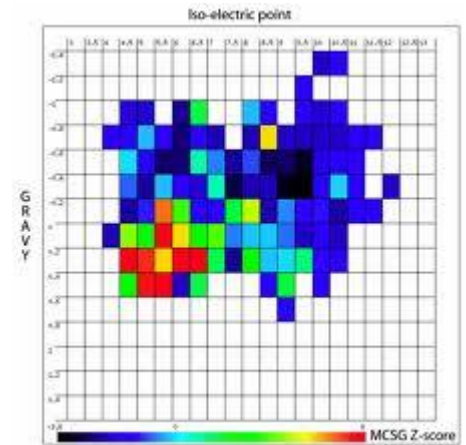


mkweb.bcgsc.ca

# 2. Charakteristika vstupních dat



www.padowan.dk



technology.skbk.org

# Jak probíhá vizualizace dat

- Generování dat
  - Měření, simulace, modelování
    - Může být velmi dlouhé (měření, simulace) i velmi drahé (simulace, modelování)
- Vizualizace (zbytek vizualizační pipeline)
  - Vizuální mapování, renderování
    - Může být rychlá nebo pomalá, v závislosti na hardware a implementaci
- Interakce (user feedback)
  - Jak může uživatel interagovat

# Pasivní vizualizace

- Tyto tři kroky jsou striktně odděleny
  - Generování dat – po skončení této fáze
  - Off-line vizualizace
    - Zobrazení vygenerovaných dat
    - Výsledkem je video nebo animace
  - Pasivní vizualizace
    - Prohlížení výsledků předchozí fáze



# Interaktivní vizualizace

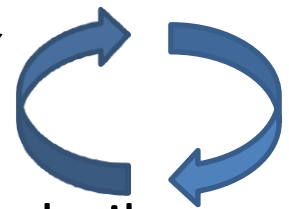
- Pouze fáze generování dat je oddělena

- Off-line generování dat



- Interaktivní vizualizace

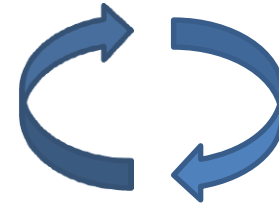
- Generovaná data jsou dostupná pro interaktivní vizualizaci
- Možnosti: výběr, parametrizace vizualizačních technik
- V současnosti velmi populární a rozšířené



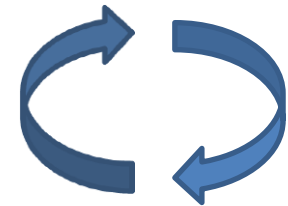
# Interaktivní řízení (steering)

- Všechny tři kroky jsou spojeny

- Generování dat „on the fly“



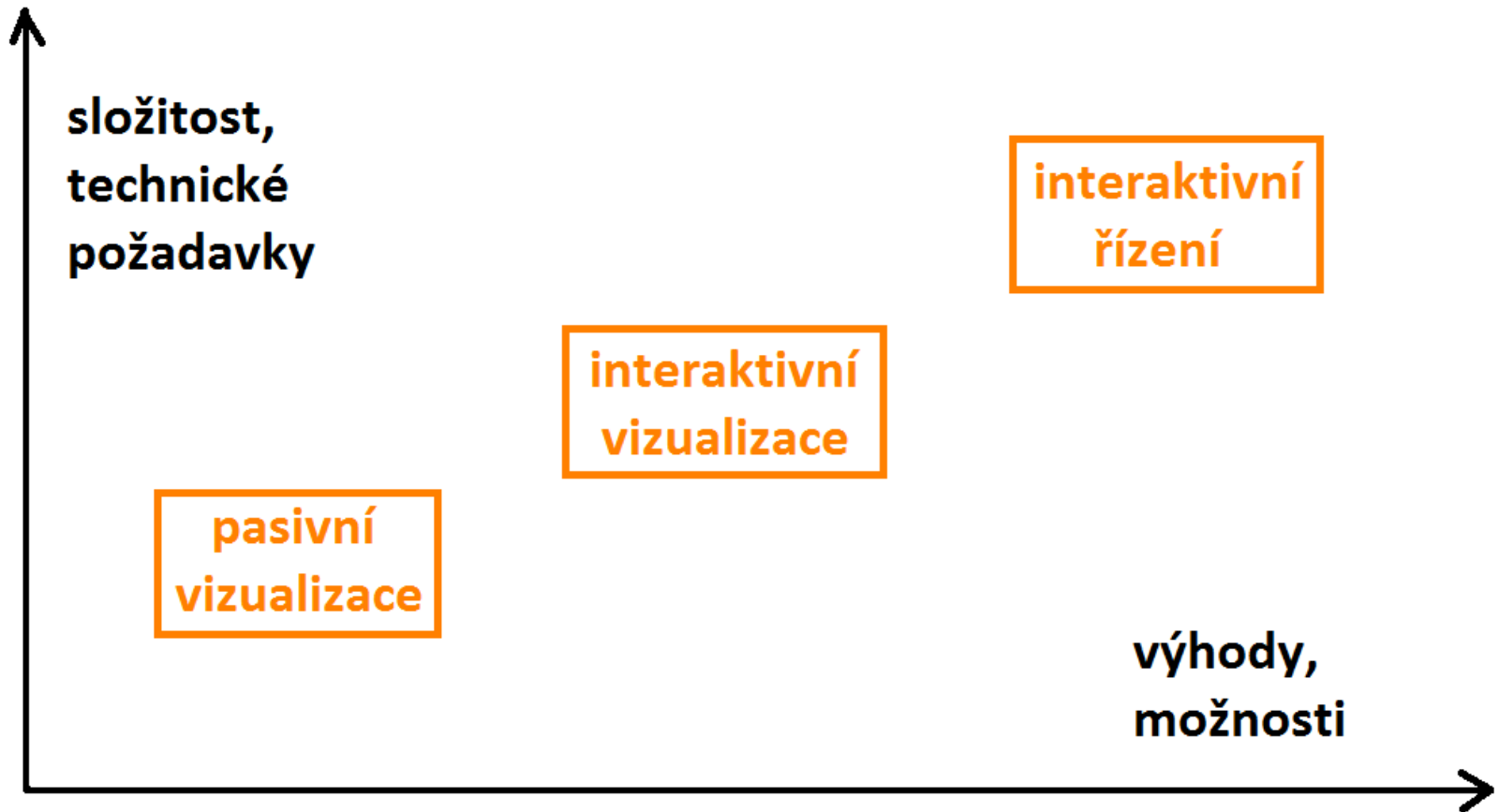
- Interaktivní vizualizace umožňující „real-time“  
náhled na data



- Rozšířené možnosti interakce

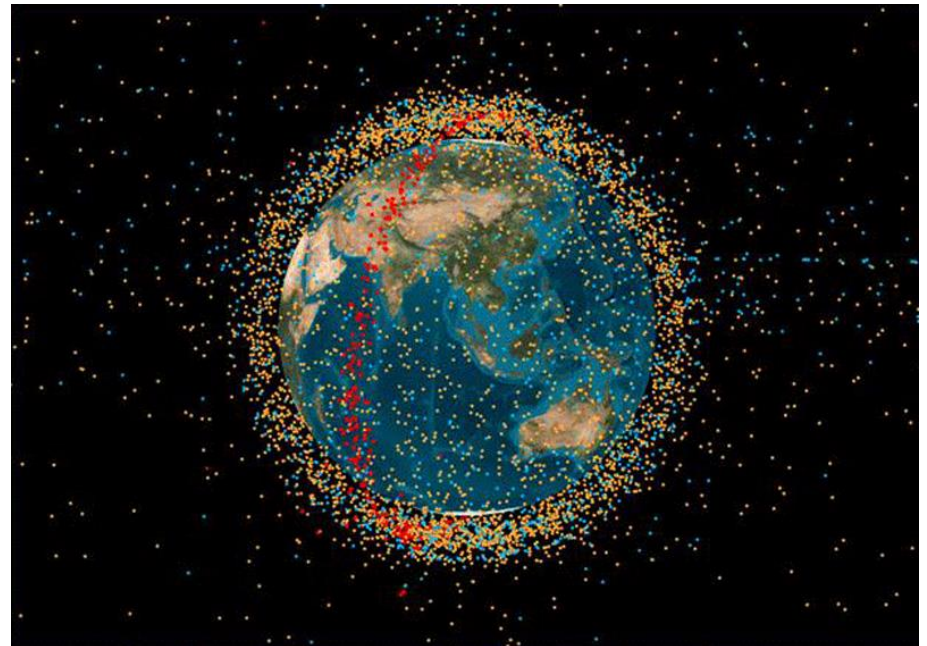
- Uživatel může ovlivňovat průběh simulace, změnit design při modelování, atd.
- Velmi náročné na provedení i náklady

# Srovnání



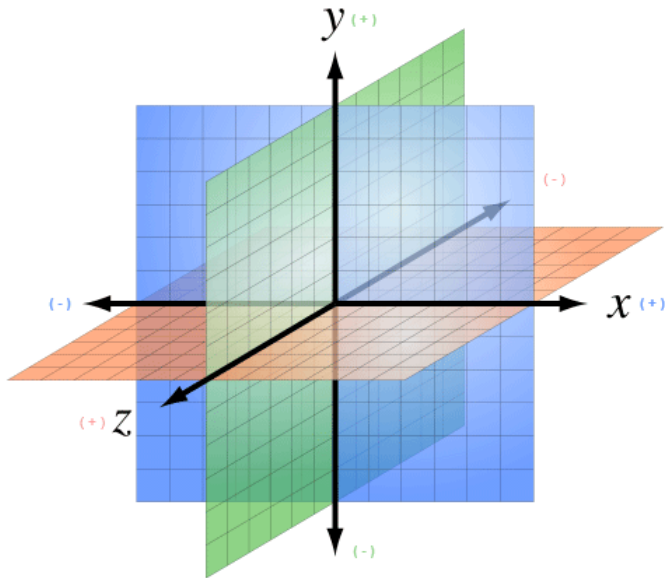
# Data

- Jsou ústředním tématem vizualizace, vše se točí kolem dat
- Data ovlivňují výběr vhodné vizualizační techniky (výběr ovlivňuje samozřejmě i uživatel)
- Důležité otázky:
  - Kde data „žijí“ (jaký je **datový prostor**)
  - **Typ** dat
  - Jaká **reprezentace** je smysluplná



# Datový prostor

- Různé vlastnosti
  - Dimenzionalita datového prostoru
  - Souřadný systém
  - Region vlivu (lokální nebo globální dopad)



wiki.brown.edu



en.wikipedia.org



# Definice dat

- *Raw (surová) data x předzpracovaná data*
- Data ve tvaru  $(r_1, r_2, \dots, r_n)$
- Každý  $r_i$  záznam obsahuje  $m$  proměnných  $(v_1, v_2, \dots, v_m)$
- $v_i$  je často označována jako pozorování

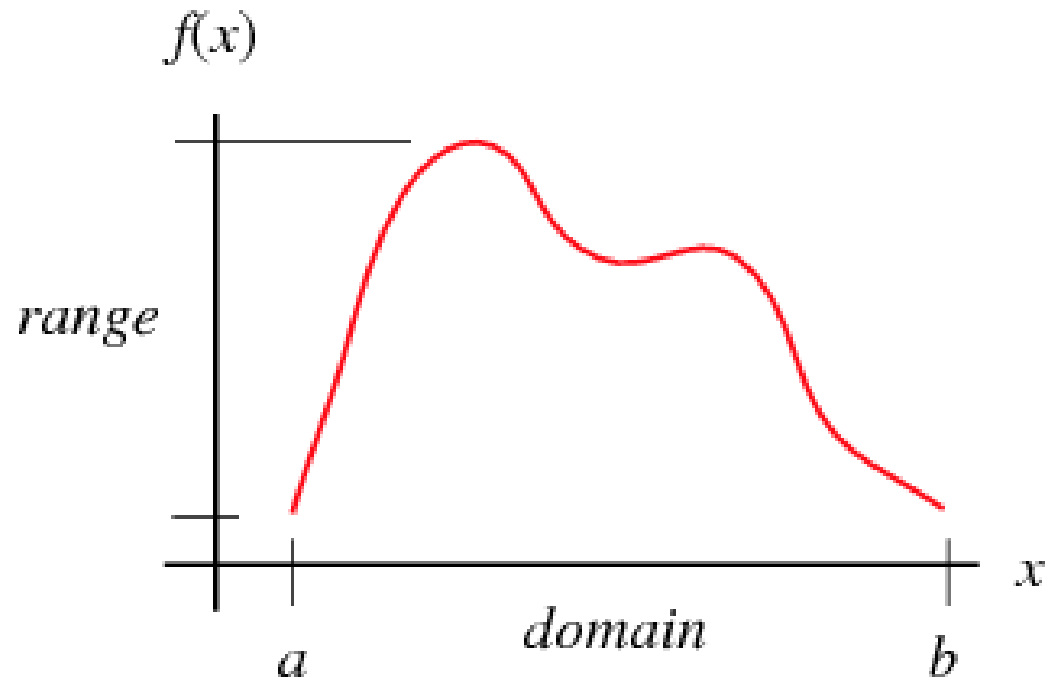
# Definice proměnných

- *Nezávislá ( $iv_i$ ) proměnná*
  - není ovlivněna jinou proměnnou (např. čas)
- *Závislá ( $dv_j$ ) proměnná*
  - je ovlivněna jednou nebo více nezávislými proměnnými (např. teplota)
- Záznam tedy může být reprezentován ve tvaru
$$r_i = (iv_1, iv_2, \dots, iv_{m_i}, dv_1, dv_2, \dots, dv_{m_d})$$

kde  $m = m_i + m_d$

# Data generovaná pomocí funkce

- Nezávislé proměnné = definiční obor
- Závislé proměnné = obor hodnot



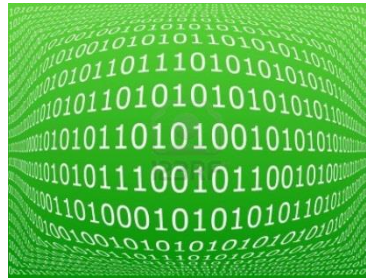
# Typy proměnných

- Fyzické typy
  - Charakterizovány vstupním formátem
  - Charakterizovány typem možných operací
  - Příklad: bool, string, int, float,...
- Abstraktní typy
  - Popis dat
  - Charakterizovány metodami/atributy
  - Mohou mít hierarchickou strukturu
  - Příklad: rostliny, zvířata, ...

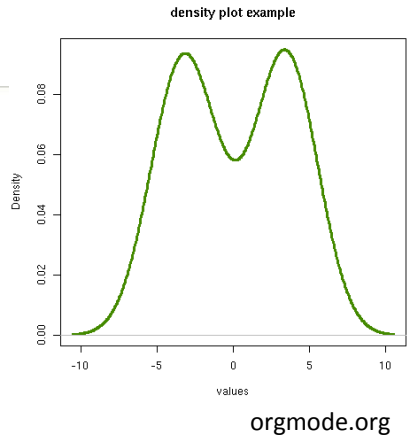
# Typy dat

- **Ordinální**

- Binární
- Diskrétní
- Spojitá

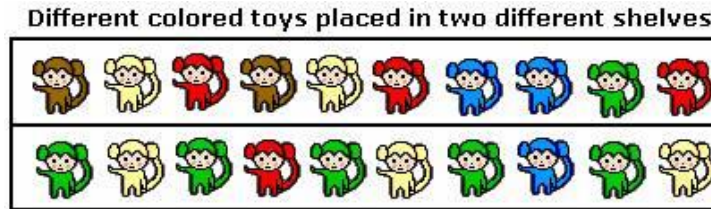


www.123rf.com

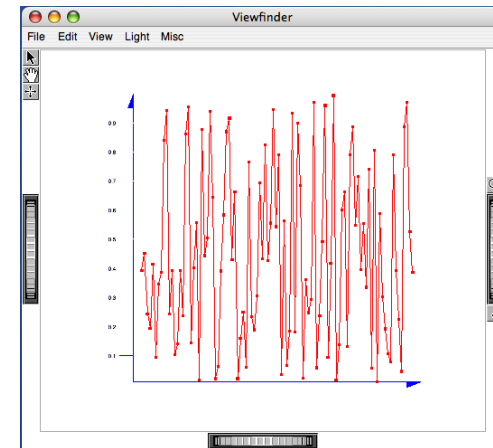
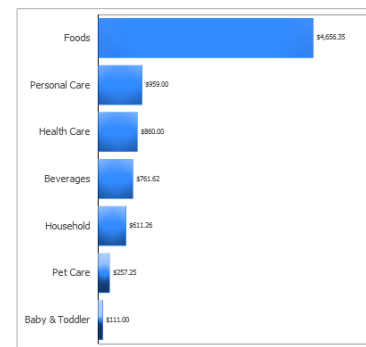


- **Nominální**

- Kategorická
- Setříděná
- Náhodná



www.icoachmath.com



www.cincomsmalltalk.com

# Scale – měřítko

- 3 základní atributy:
  - Relace uspořádání na datech
  - Vzdálenostní metrika
  - Existence absolutní nuly
    - Fixace minimální hodnoty proměnné

ABSOLUTE ZERO

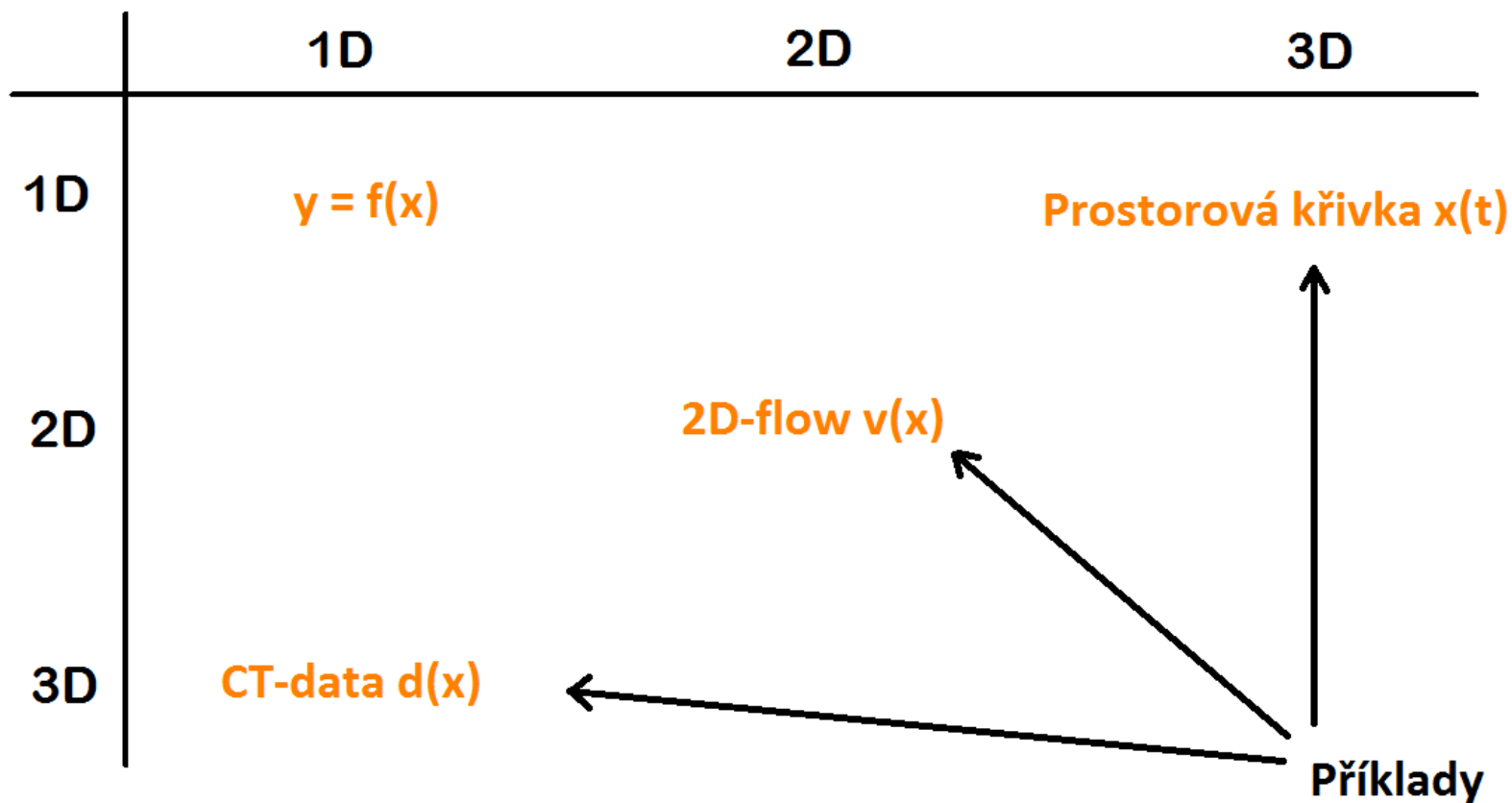


The only thing cooler is  
not being a scientist.

# Reprezentace dat

- Závisí na:
  - Přítomnosti vlastní prostorové domény
    - Pokud není, jakou prostorovou doménu zvolit?
  - Jak jsou využity jednotlivé dimenze?
    - Charakteristika dat
    - Dostupný prostor pro zobrazení (2D/3D)
    - Na kterou část dat se zaměřujeme (focus)?
    - Ve kterých částech lze naopak abstrahovat?

# Datový prostor vs. vlastnosti dat

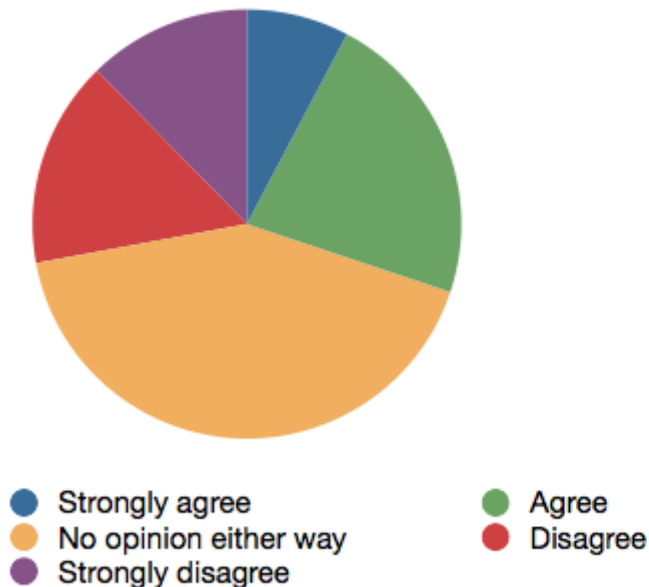




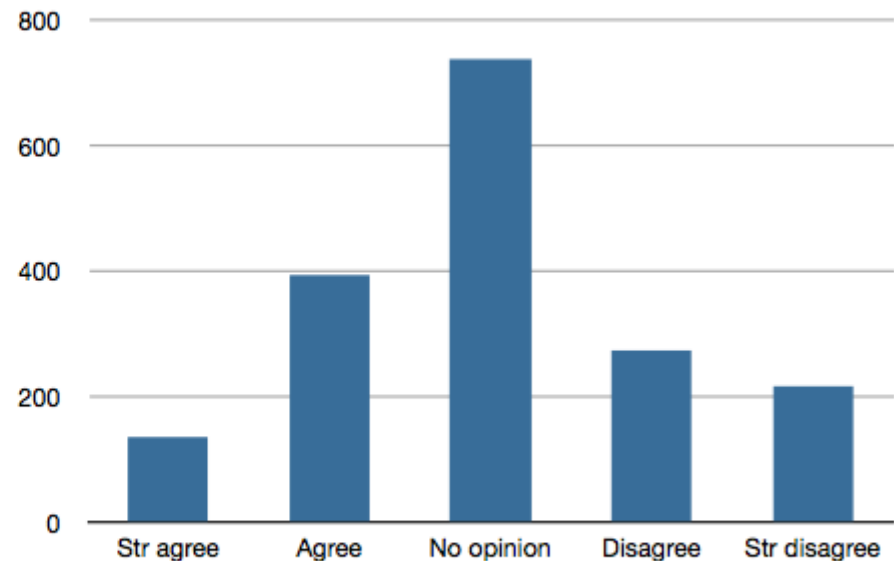
# Příklady

- Diskrétní data – sada hodnot, vizualizace pomocí sloupcových grafů, koláčových grafů, ...

Level of agreement with the Tea Party: a pie chart

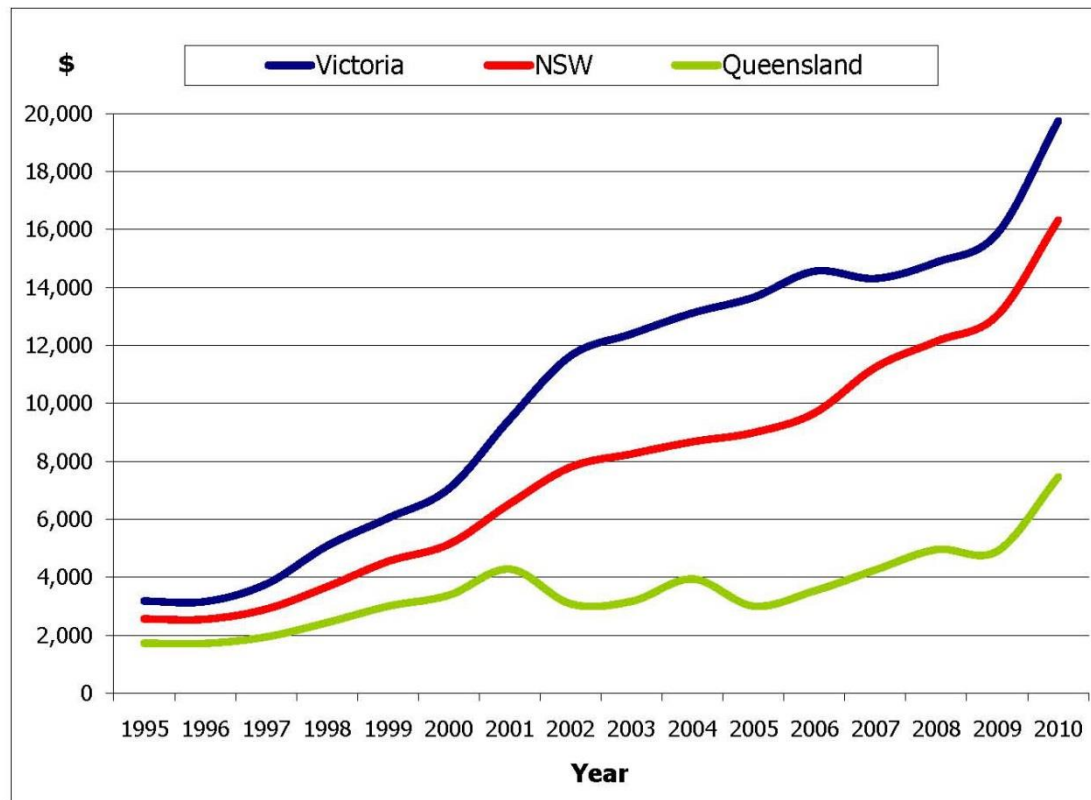


Level of Agreement with the Tea Party: a bar chart



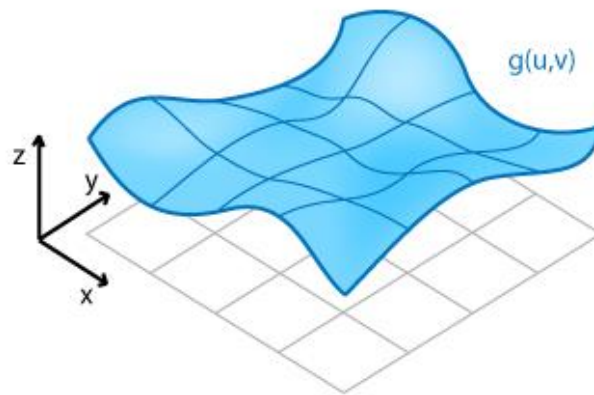
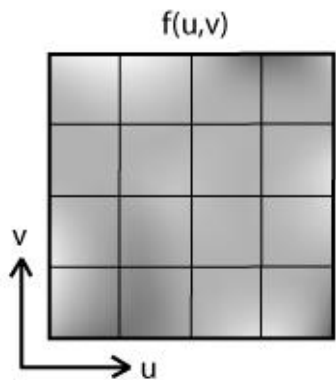
# Příklady

- Spojitá data – funkce, vizualizace pomocí grafů

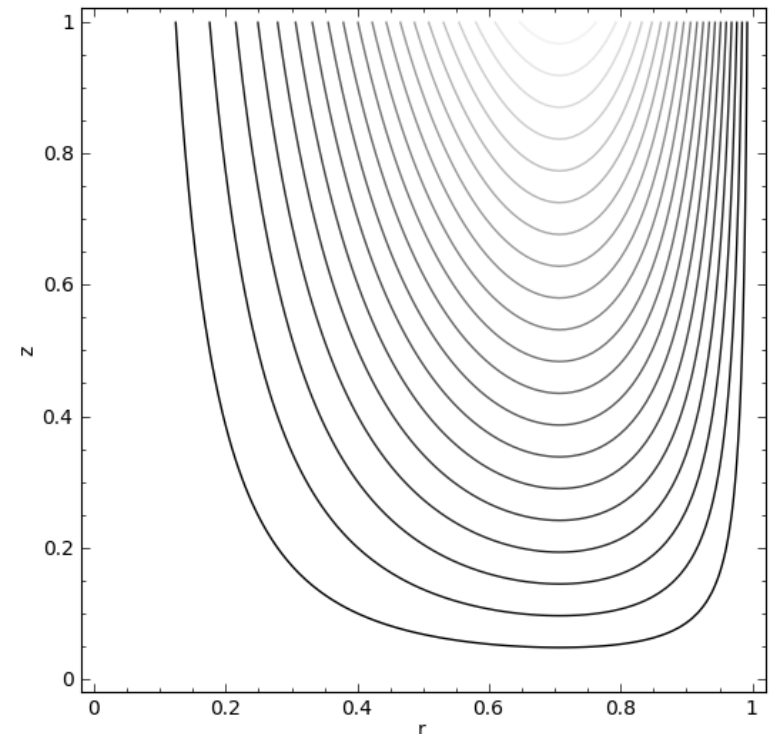


# Příklady

- 2D reálná čísla
  - Funkce dvou proměnných, vizualizace pomocí 2D výškových map, kontur ve 2D, ...

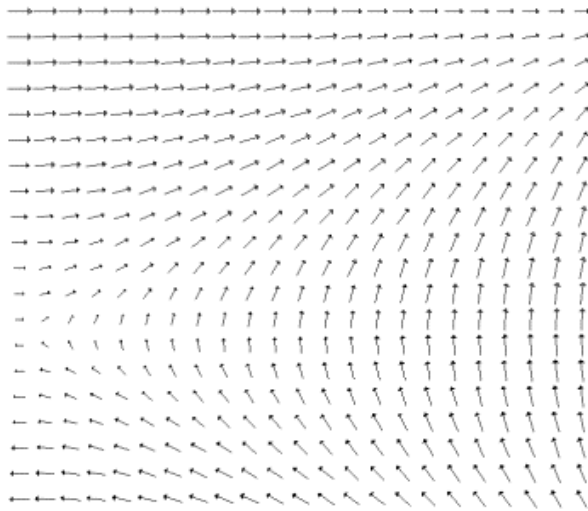


$$g(u, v) \rightarrow (x, y, z) : \begin{cases} x = u \\ y = v \\ z = f(u, v) \end{cases}$$

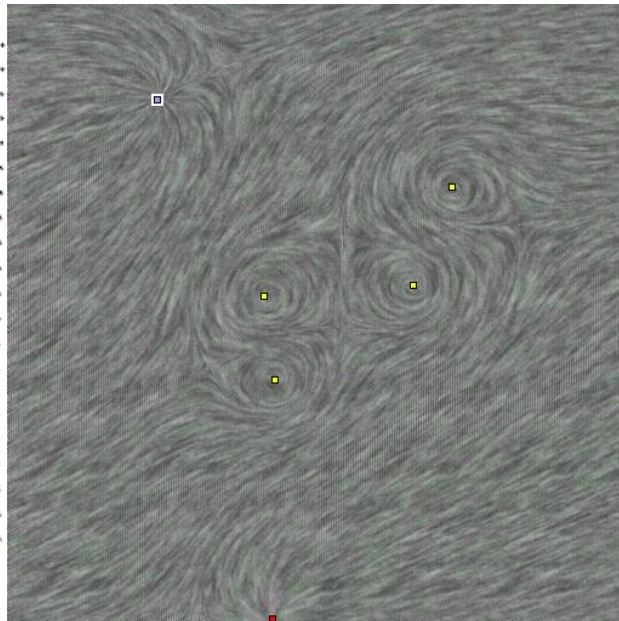


# Příklady

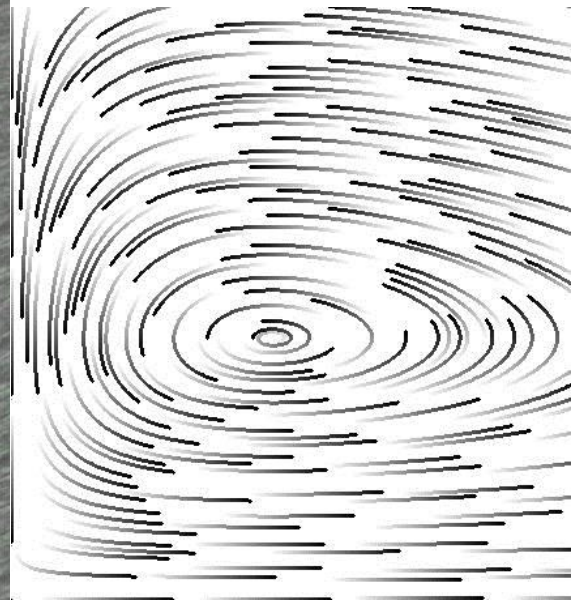
- 2D vektorová pole, vizualizace pomocí tzv. hedgehog plots, LIC (line integral convolution), streamlets, ...



csis.pace.edu



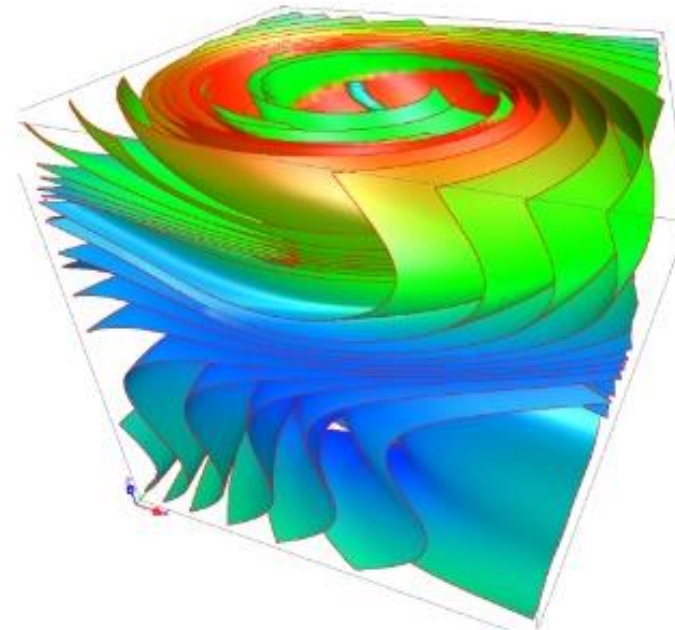
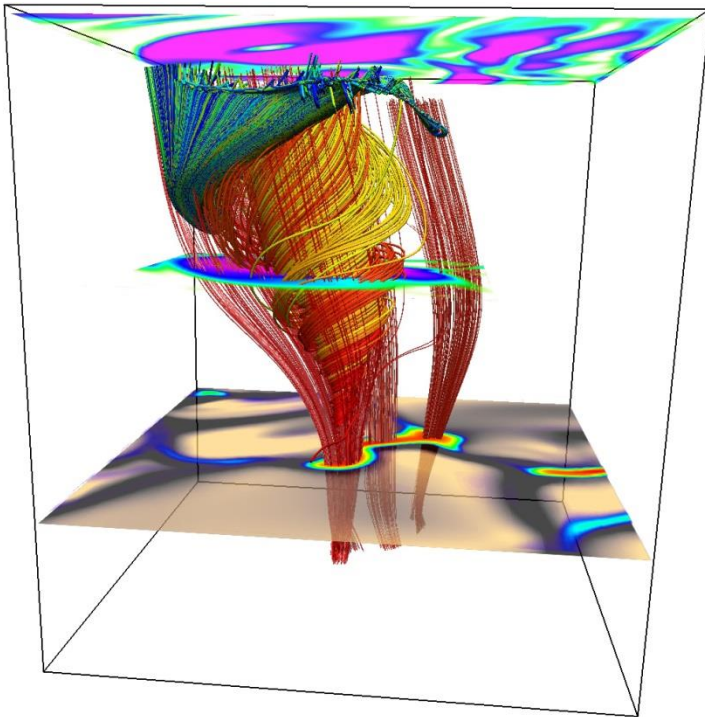
en.wikipedia.org



www.cg.tuwien.ac.at

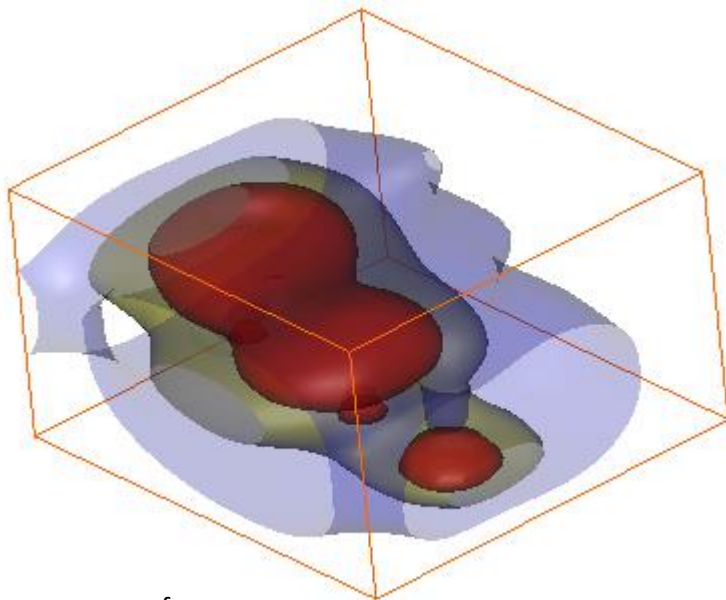
# Příklady

- Prostorová data + čas
  - 3D tok, vizualizace pomocí tzv. streamlines, streamsurfaces

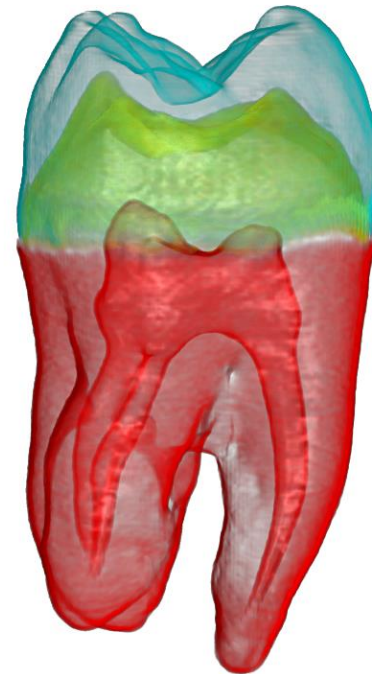


# Příklady

- Prostorová data
  - 3D hustota, vizualizace pomocí izopovrchů, volume rendering



[www.ssg-surfer.com](http://www.ssg-surfer.com)

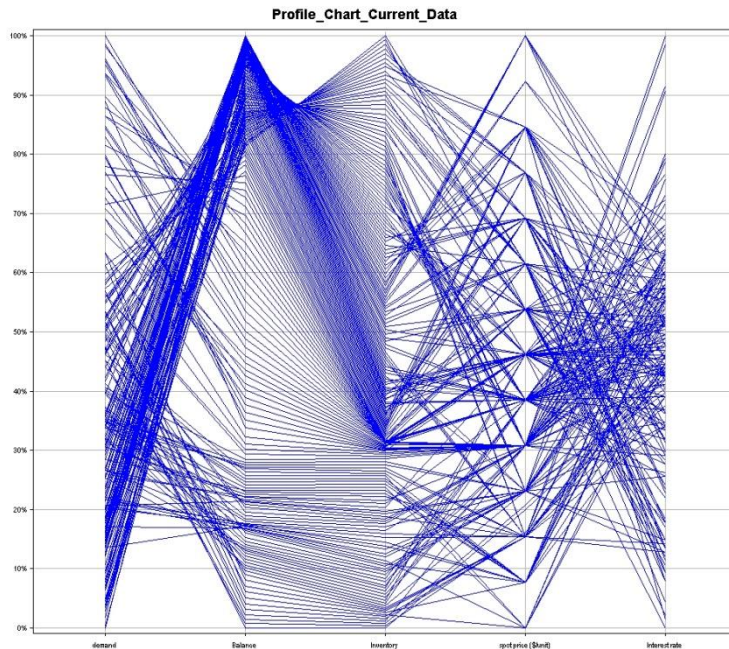


[viscg.uni-muenster.de](http://viscg.uni-muenster.de)



# Příklady

- Multidimenzionální data
  - Sada n-tic, vizualizace pomocí paralelních souřadnic, glyfů, ikon, ...



[www.cs.umd.edu](http://www.cs.umd.edu)

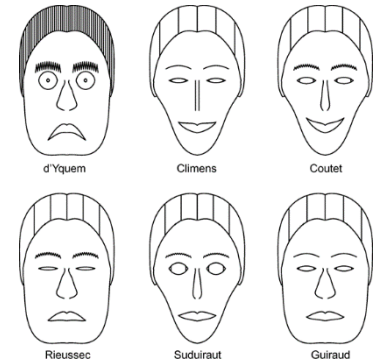
Maximum Values Face



Minimum Values Face



[www.emeraldinsight.com](http://www.emeraldinsight.com)



[datamining.typepad.com](http://datamining.typepad.com)

# Struktura uvnitř a mezi záznamy

- Datové množiny mají:
  - **Syntaxi** – způsob reprezentace (tzv. datový model)
  - **Sémantiku** – vztahy uvnitř jednoho záznamu nebo mezi záznamy (tzv. konceptuální model)
- Typy struktur:
  - Skaláry, vektory, tenzory
  - Geometrie a gridy
  - Jiné formy



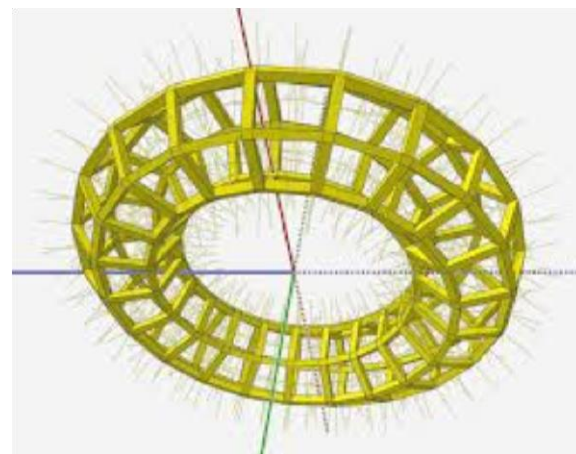
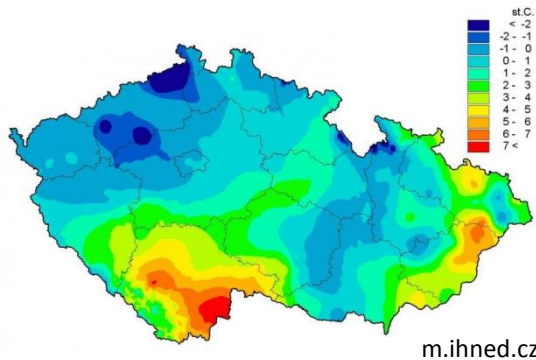


# Skaláry, vektory a tenzory

- **Skalár** = jednotlivé číslo v datovém záznamu
  - např. věk
- **Vektor** = kompozice několika proměnných do jednoho záznamu
  - např. bod ve 2D prostoru, RGB barva
- **Tenzor** = definován řádem a dimenzí prostoru, ve kterém je definován. Reprezentován pomocí pole nebo matice.
  - např. transformační matice ve 3D

# Geometrie a gridy

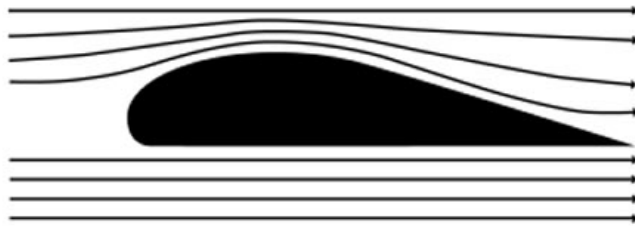
- Geometrie zaznamenána pomocí souřadnic záznamů



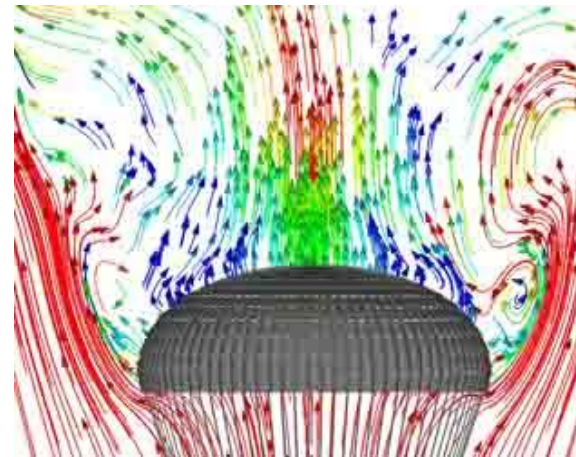
- Grid – geometrii lze odvodit ze startovní pozice, orientace a velikosti kroku v horizontálním a vertikálním směru

# Neuniformní geometrie

- Je nutné uchovávat souřadnice všech záznamů – nejdou odvodit

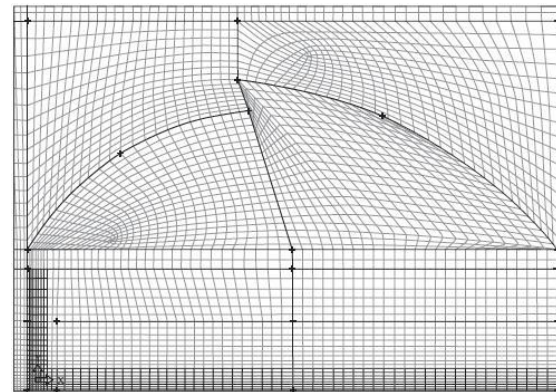


[blog.nasm.si.edu](http://blog.nasm.si.edu)



[www.tafsm.org](http://www.tafsm.org)

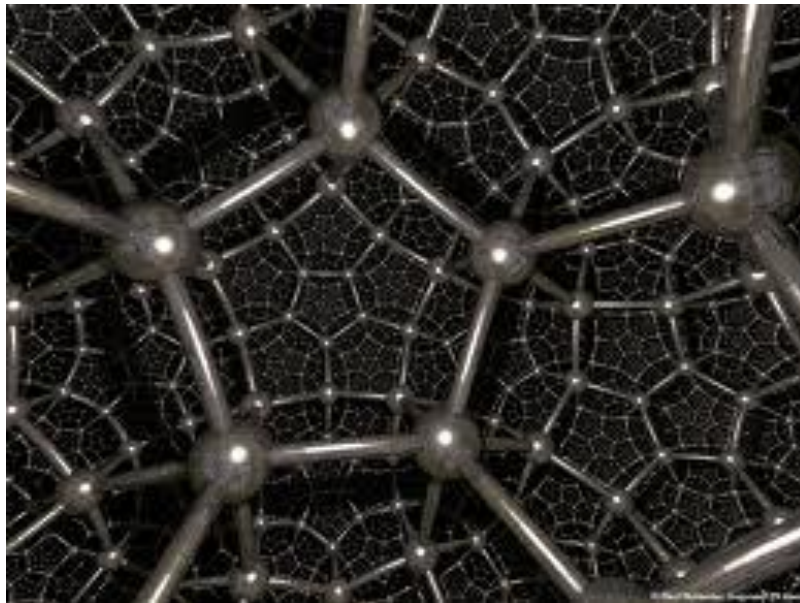
- Neuniformní grid



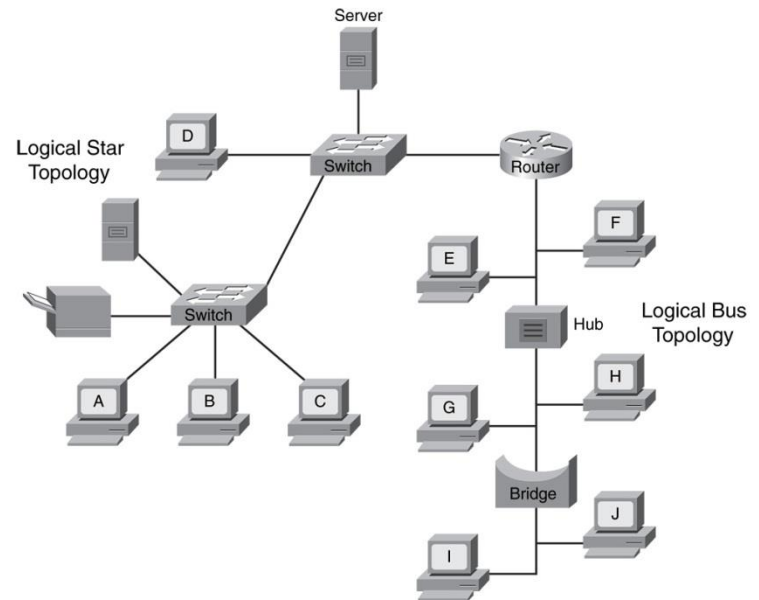
[www.scielo.org.mx](http://www.scielo.org.mx)

# Jiné formy struktur

- Další důležitou strukturou je **topologie**
- Definuje tzv. **konektivitu**



[www.bugman123.com](http://www.bugman123.com)



[xpertnetworking.wordpress.com](http://xpertnetworking.wordpress.com)

- Významná při převzorkování a interpolaci

# Čas



megworden.com

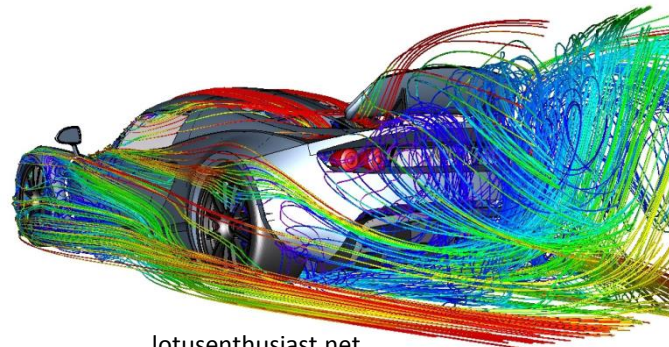
- Obrovský rozsah hodnot (pikosekundy vs. tisíciletí)
- Vyjádřen absolutně nebo relativně
- Datové množiny obsahující časovou složku mohou být rozloženy rovnoměrně (pravidelný sampling) i nerovnoměrně (zpracovávání transakcí – podle doby jejich přijetí a zpracování)

# Další příklady strukturovaných dat

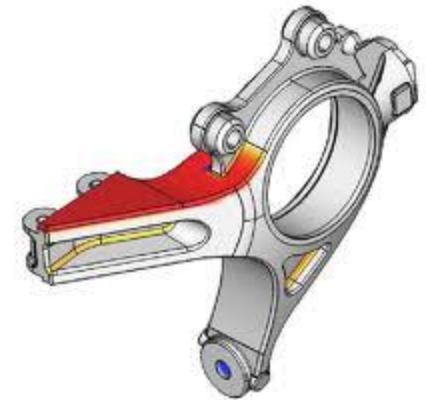
- Magnetická rezonance (MRI)
- Computational fluid dynamics (CFD)
- Finančnictví
- CAD systémy
- Sčítání lidu
- Sociální sítě



www.impactlab.net



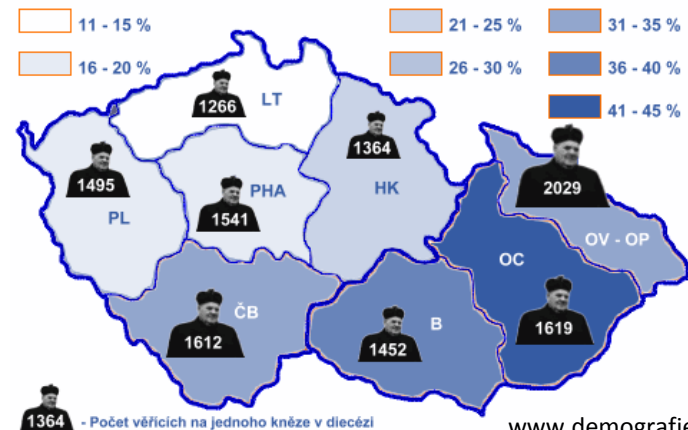
lotusenthusiast.net



www.mjmdesigns.co.uk



successfulworkplace.com

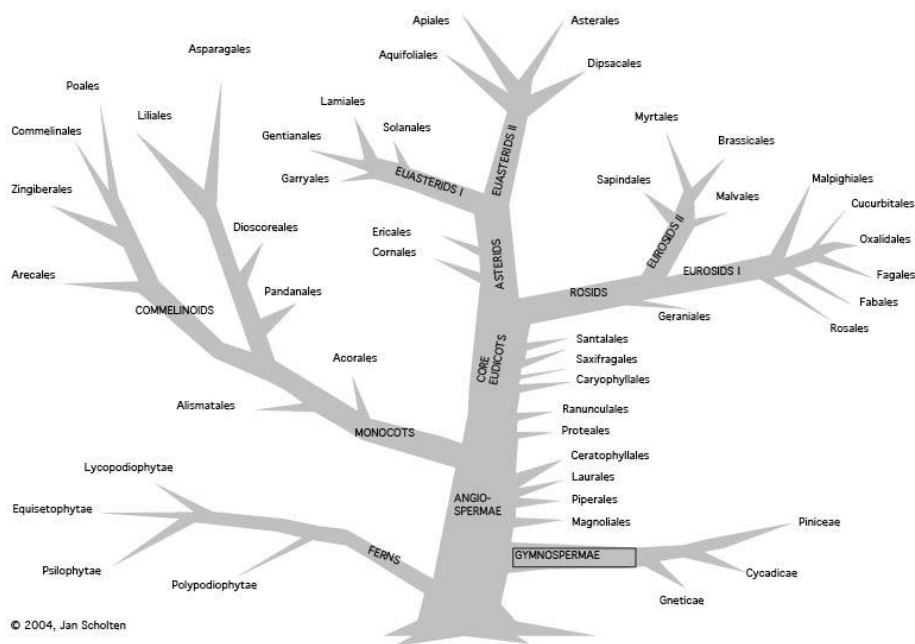


www.demografie.info



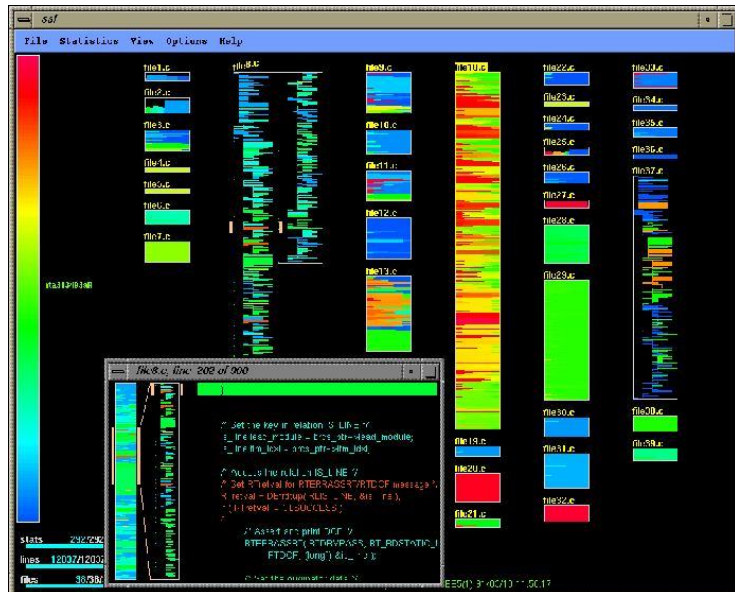
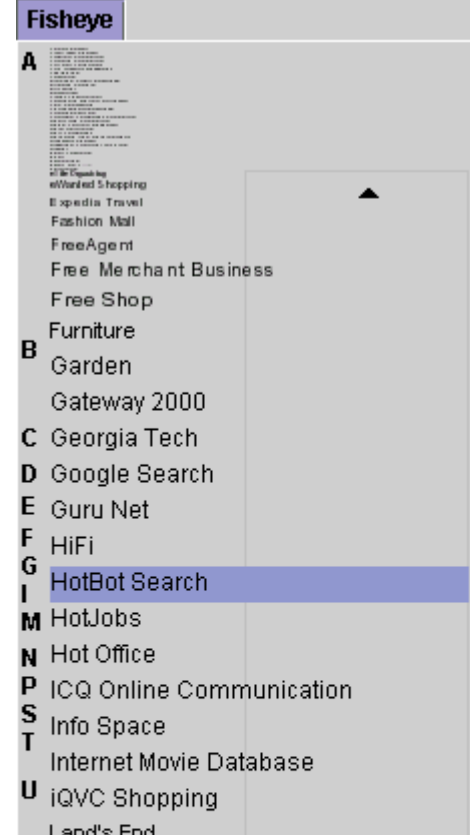
# Taxonomie – 7 typů dat

- 1D (lineární sady a sekvence)
- 2D (mapy)
- 3D (objekty, tvary)
- nD (relační)
- Stromy (hierarchie)
- Sítě (grafy)
- Temporální



# Lineární data

- Dlouhé seznamy položek
  - Položky menu
  - Zdrojový kód
- Fisheye displays



<http://ds.cc.yamaguchi-u.ac.jp/~ichikay/pfp7/iv/pics/SeeSoft-line.jpg>



# 2D data - mapy

- GIS (geografické informační systémy)

- Zobrazení map (např. Google Earth)

- <http://www.youtube.com/watch?v=ouqjv1MBmI8>

- <http://www.wimp.com/unusualplaces/>

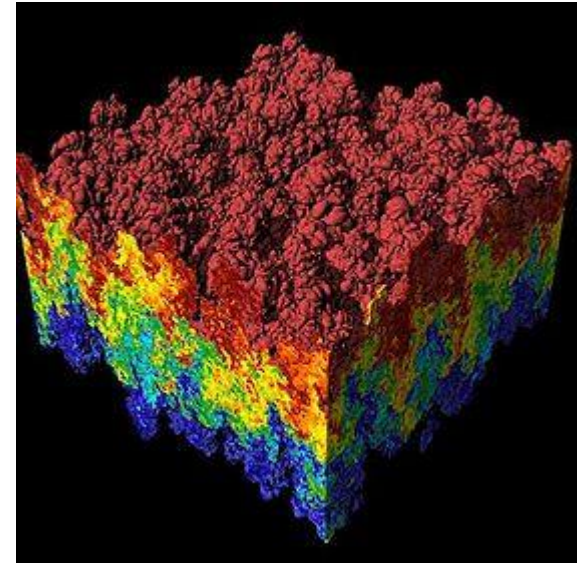
- Prostorové dotazování

- Prostorová analýza dat

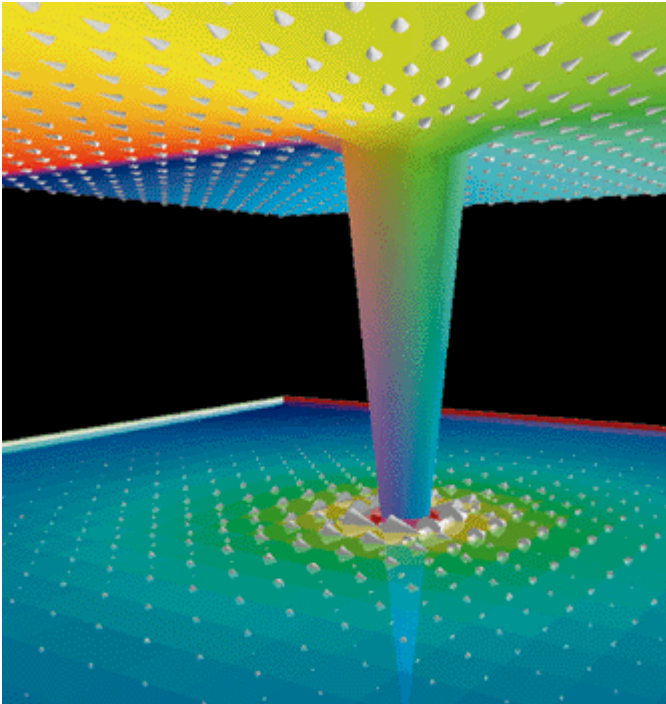


# 3D data

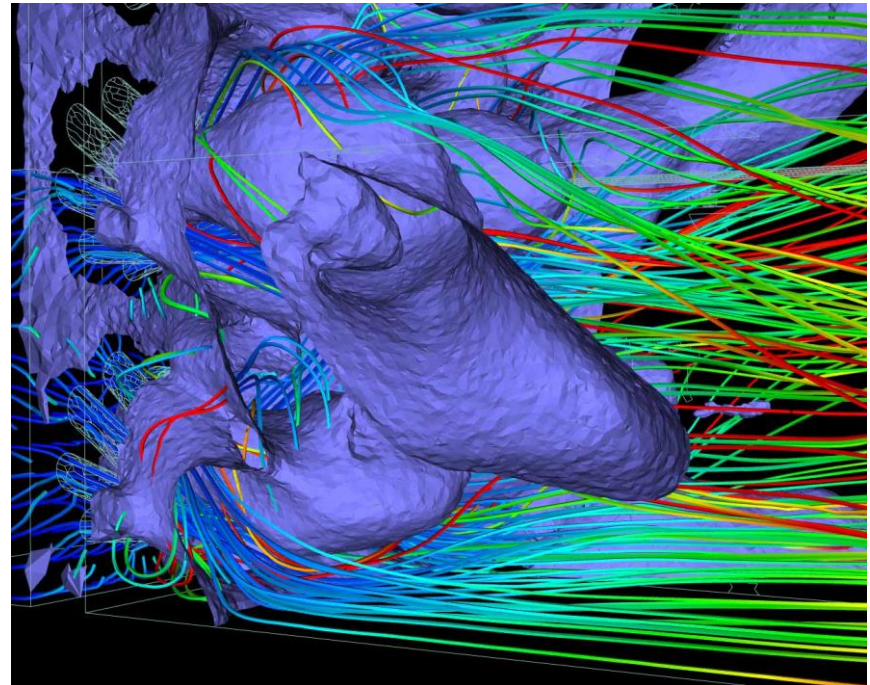
- Různé typy zobrazení 3D dat
- Scientific visualization



en.wikipedia.org



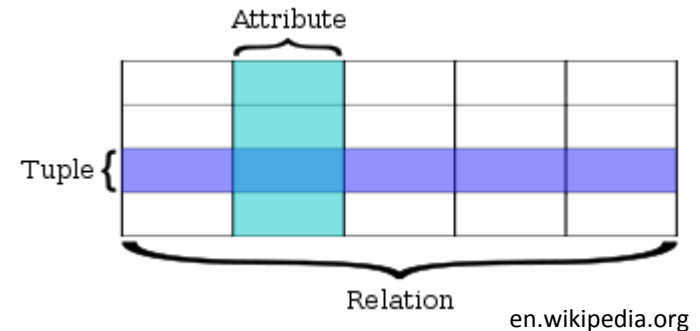
www.ornl.gov



gvis.grc.nasa.gov

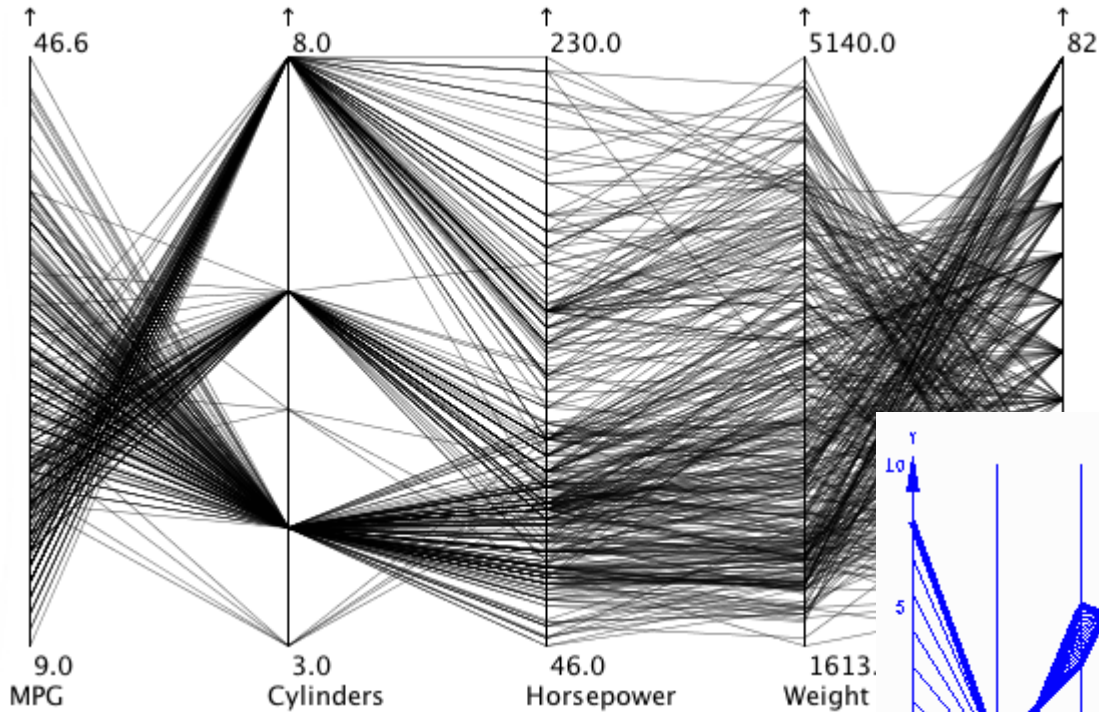
# Multidimenzionální data

- Záznamy v relační databázi

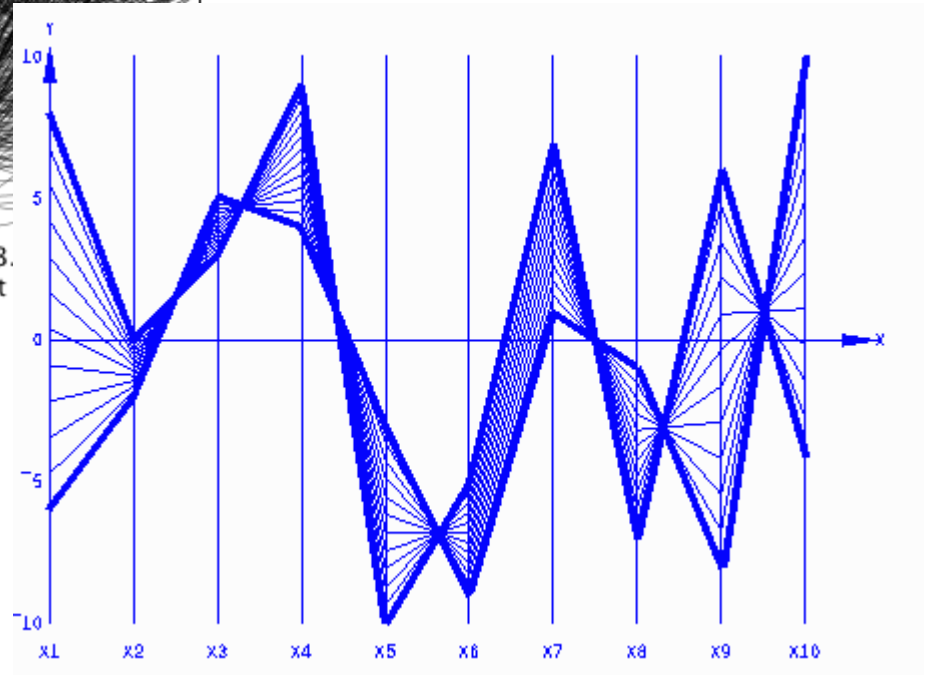


- Dvě řešení:
  - Vykreslení všech možných párů proměnných ve 2D grafu
    - Jednoduché, ale nepoužitelné pro zobrazení dat jako celku
  - „Paralelní souřadnice“
    - Metoda pro vykreslení multidimenzionálních dat (Alfred Inselberg)

# Paralelní souřadnice



eagereyes.org



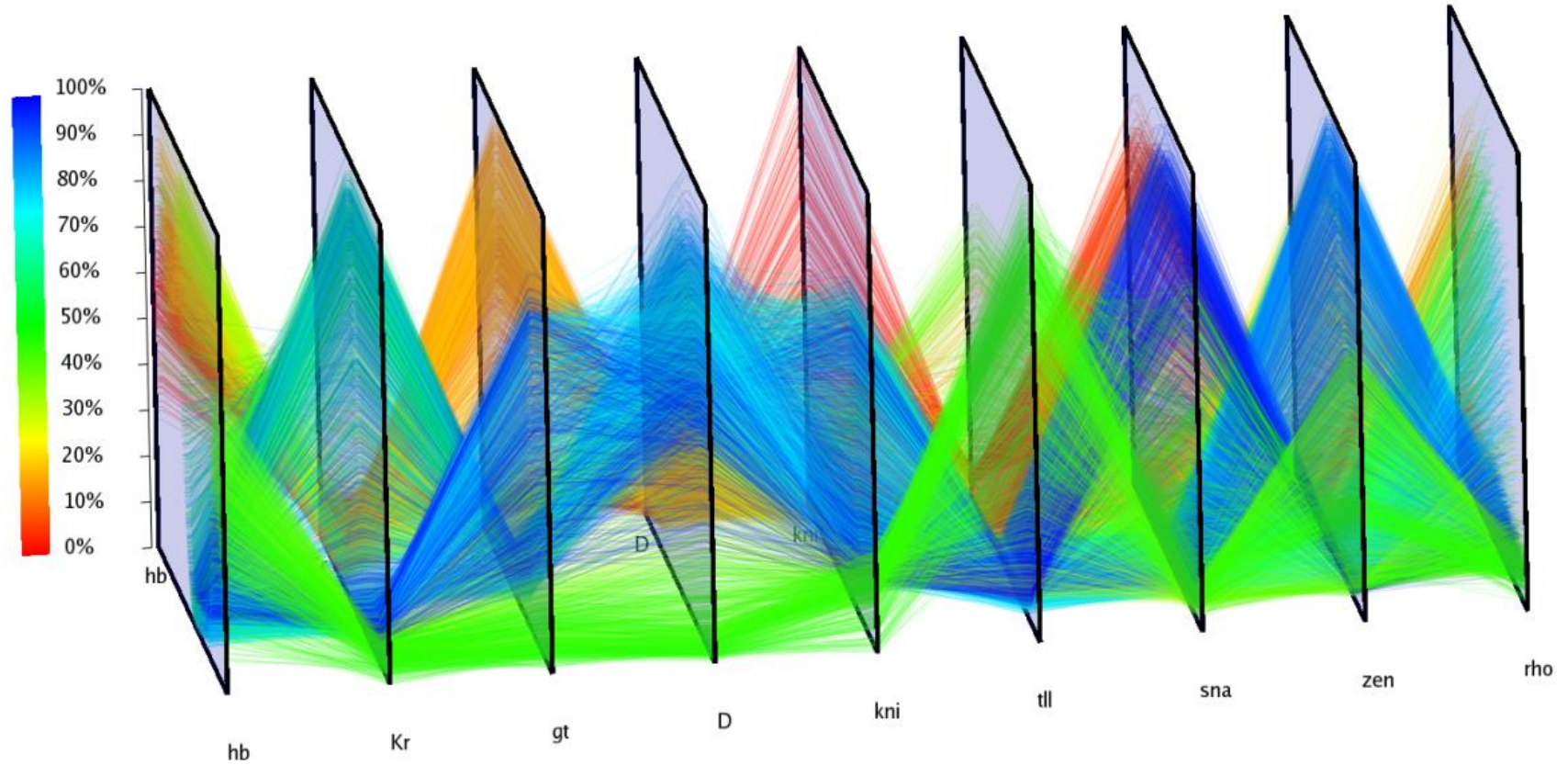
andrewgelman.com



# Paralelní souřadnice



[whyevolutionistrue.wordpress.com](http://whyevolutionistrue.wordpress.com)



# Stromy

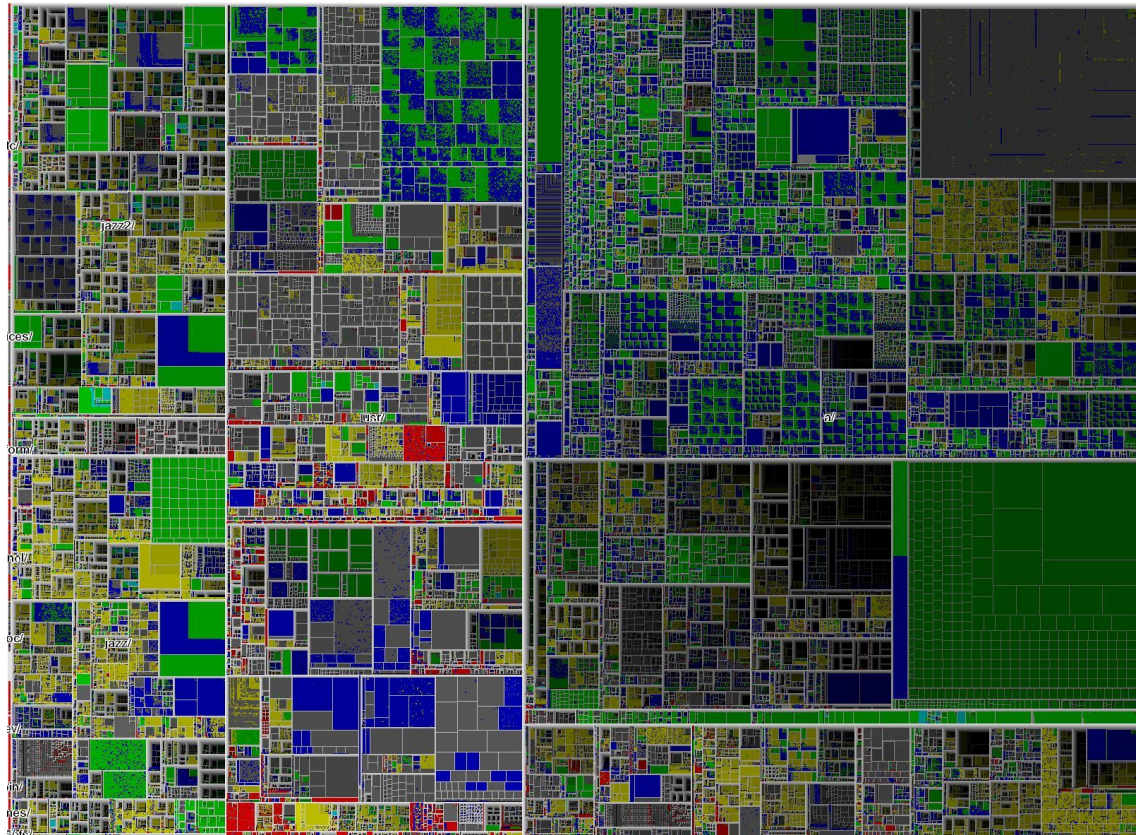
- Nutné zobrazit nejen data, ale i jejich strukturu
  - např. rodokmen, file systém
- Počet dat významně roste při sestupování do nižších vrstev





# Stromy

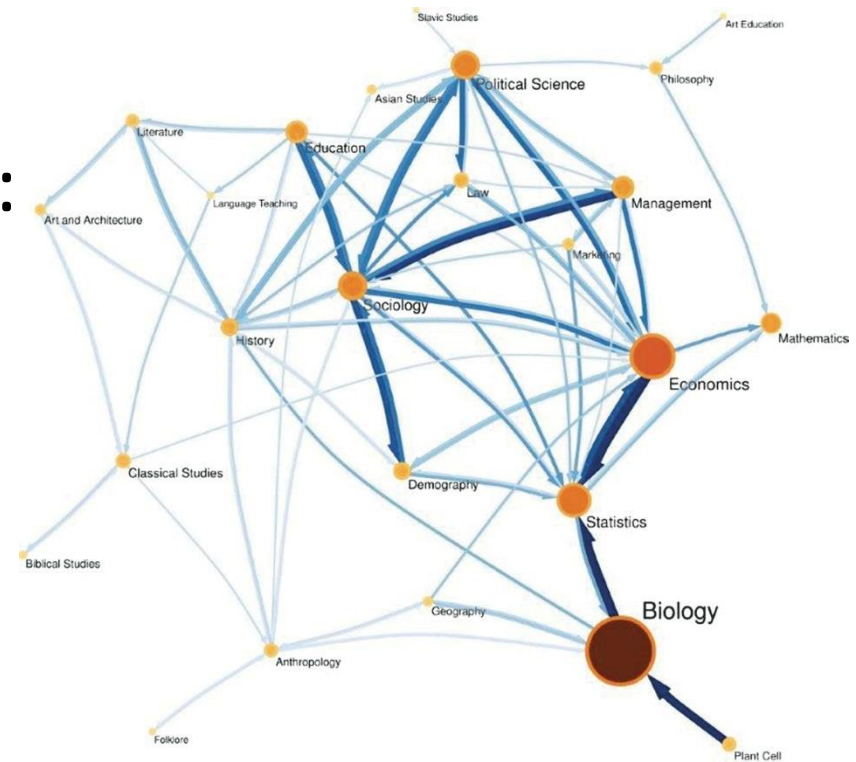
- Tree Maps
  - Zobrazení stromových dat ve formě vnořených obdélníků
  - Mapa o milionu záznamů:



Speed: 38.462ps/0.0260spf971061 items

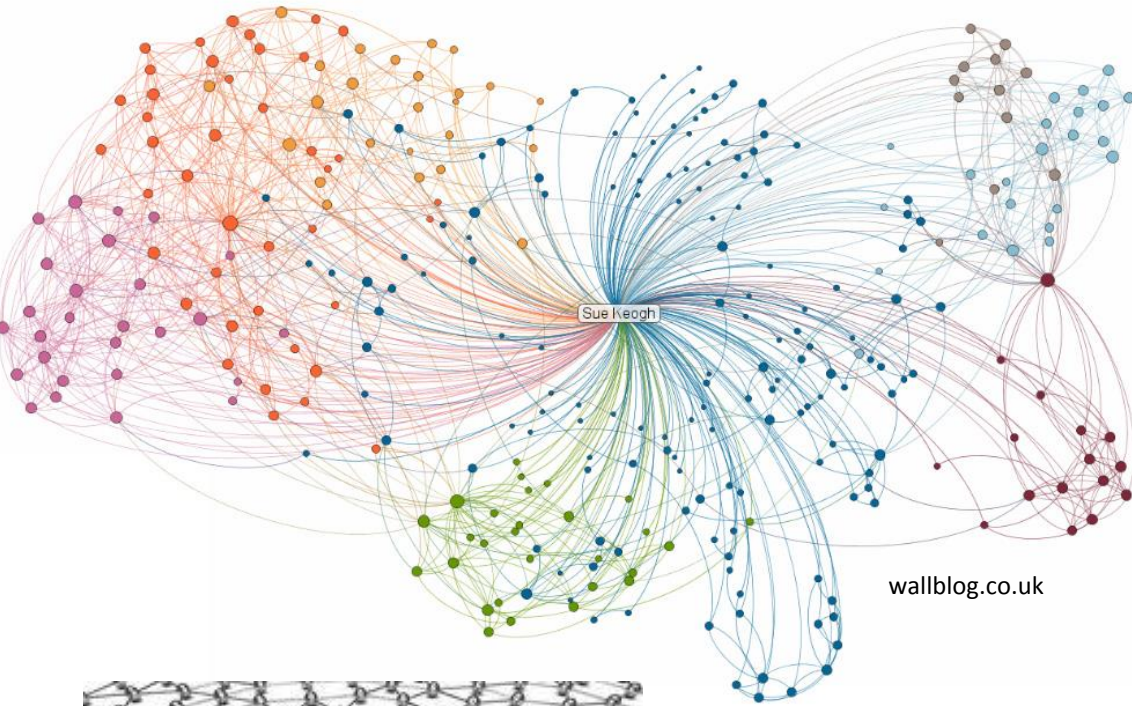
# Sítě

- Podobná situace jako u stromů – nutné zobrazovat data i jejich strukturu
- Síť = uzly + hrany
- Návrh by měl zahrnovat:
  - Minimální průniky hran
  - Minimální délky hran
  - Minimální ohýbání hran

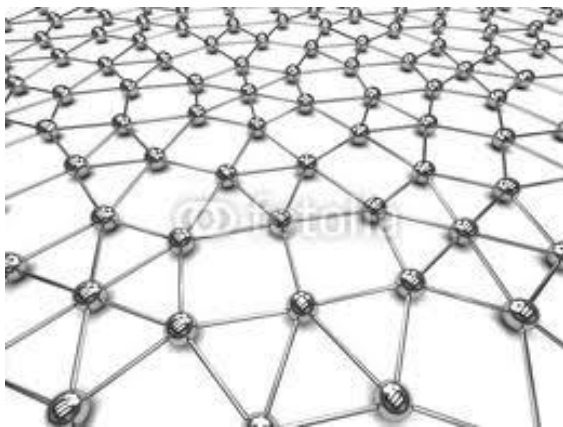




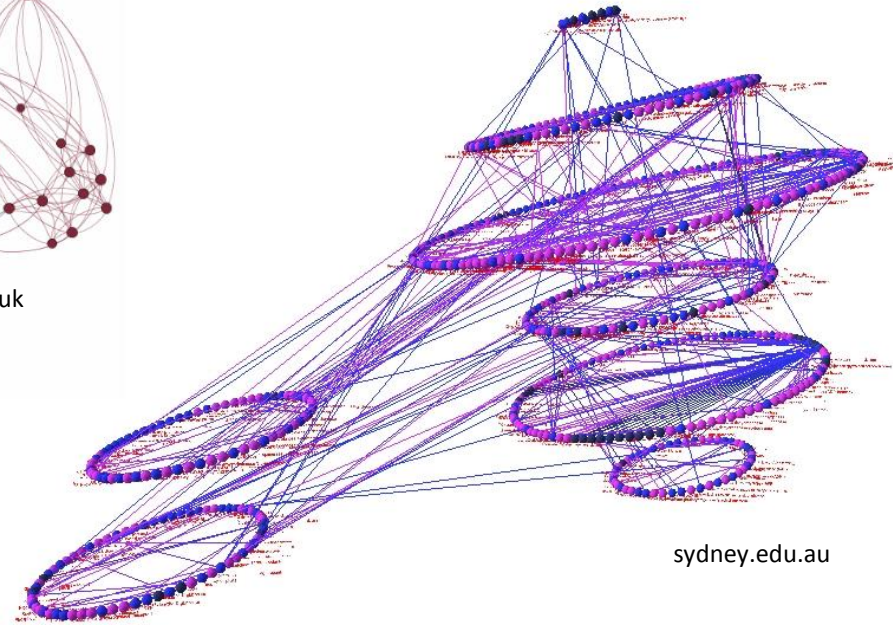
# Sítě



wallblog.co.uk



us.fotolia.com



sydney.edu.au

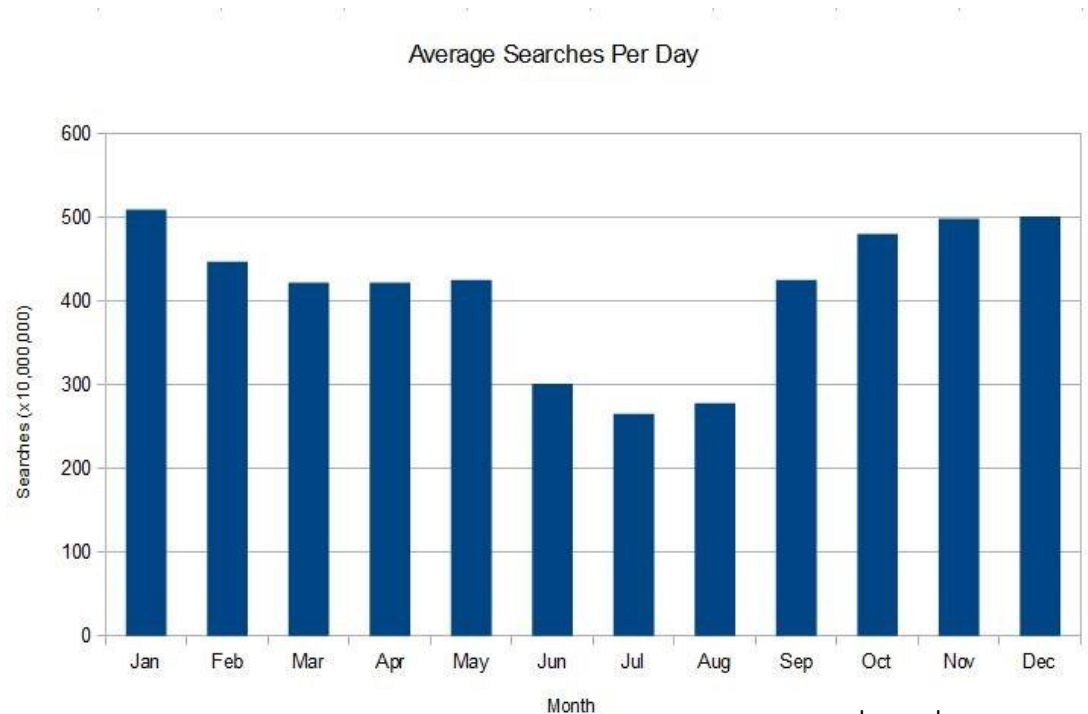
# Temporální data

- Zobrazení časově závislých dat
  - Tradičně pomocí vývojových (trend) grafů a sezónních grafů



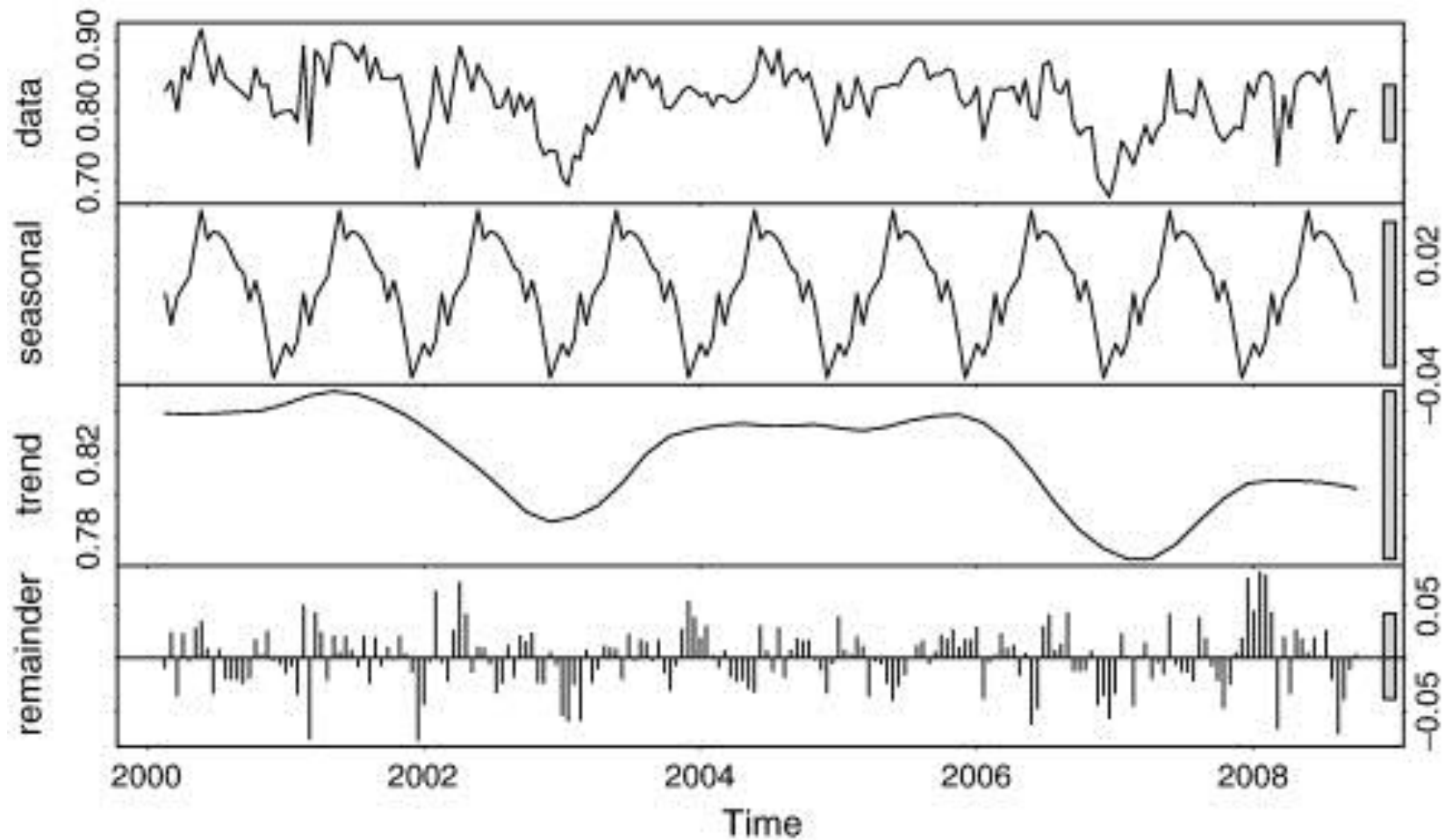
Resolution 3000 x 2200 px - Free JPG file download - [www.psdgraphics.com](http://www.psdgraphics.com)

[www.psdgraphics.com](http://www.psdgraphics.com)



[www.demondeemon.com](http://www.demondeemon.com)

# Temporální data




# Temporální data - LifeLines

In progress version of LifeLine Frame [not for demo purposes - Microsoft Internet Explorer - [Working Offline]

File Edit View Go Favorites Help

Back Links Best of the Web Channel Guide Customize Links Internet Explorer News Internet Start

 **Linda Simpson**  
Female 40

Line from input file: %-.3-10-1997,3-12-1997\_black,p10,Sonogram,images/babysonogre

LifeLine

92 93 94 95 96 97

Notes

- Tobacco
- Depression
- Lyme
- Arthritis
- Obesity

Checkup Checkup Checkup Checkup

AtrialFlutter

Flu Pneumonia KneePain

Fatigue>Diabetes Diabe

Pregnancy

Hosp.: Appendectomy Pneumonia KneeSurgery

Tests: BloodEKG EKG Xray Blood Blood Blood Blood Sonogr

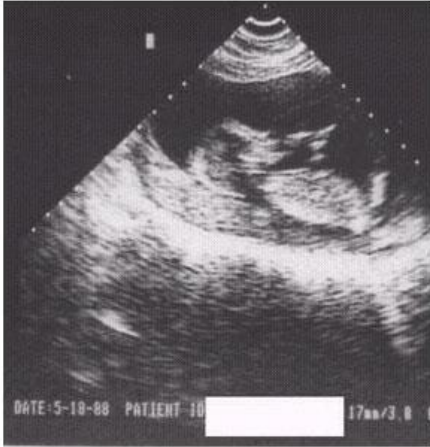
Meds.: Prozac Heartdrug Ventolyn Antib. Advil Advil Insulin Insulin

Others: PhysicalTherapy LowSaltFatDiet

Immun.: TBtest Tetanos Flu

92 93 94 95 96 97

load Control Panel



DATE:5-10-88 PATIENT:10 17mm/3.8

LifeLine Control Panel

Layout Label

- Default
- Quick Compact
- Slow Compact
- Chronologically Ordered
- Event Ordered

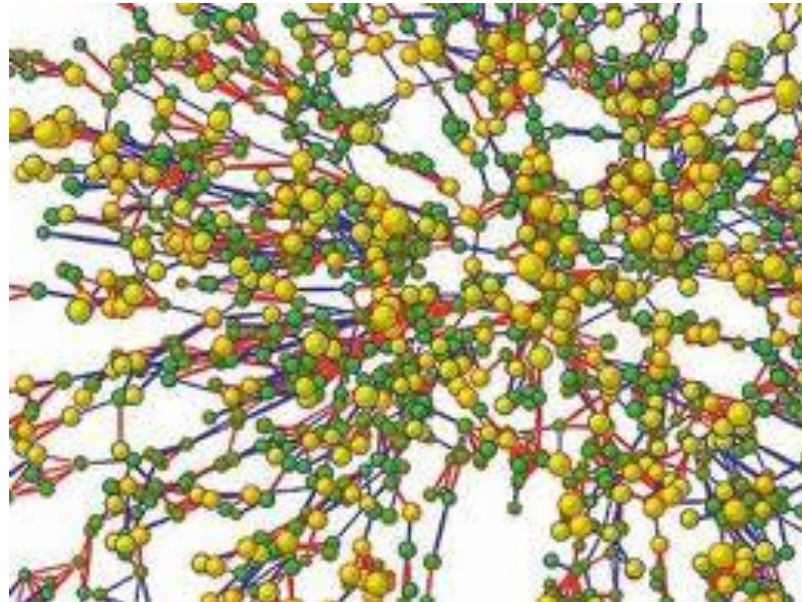
Apply OK Cancel

Warning: Applet Window



# Předzpracování dat

- Zobrazování „raw“ dat = nedochází ke ztrátě přesnosti, identifikace problémových míst, chybějících dat
- V některých případech je preprocessing nutný



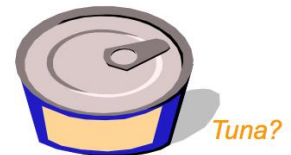
# Předzpracování dat - techniky

- Metadata a statistiky
- Chybějící hodnoty a „čištění“ dat
- Normalizace
- Segmentace
- Vzorkování a interpolace
- Snižování dimenze
- Agregace dat
- Vyhlazování a filtrace
- Konverze rastrových dat do vektorových

# Metadata a statistiky

- Metadata – informace pro předzpracování
  - Referenční bod měření
  - Jednotka měření
  - Symbol pro indikaci chybějící hodnoty
  - Rozlišení pořizovaných dat
- Statistická analýza
  - Detekce chybných záznamů
  - Analýza pomocí klastrů
  - Korelační analýza

*If you had two cans without labels, which would you eat?*

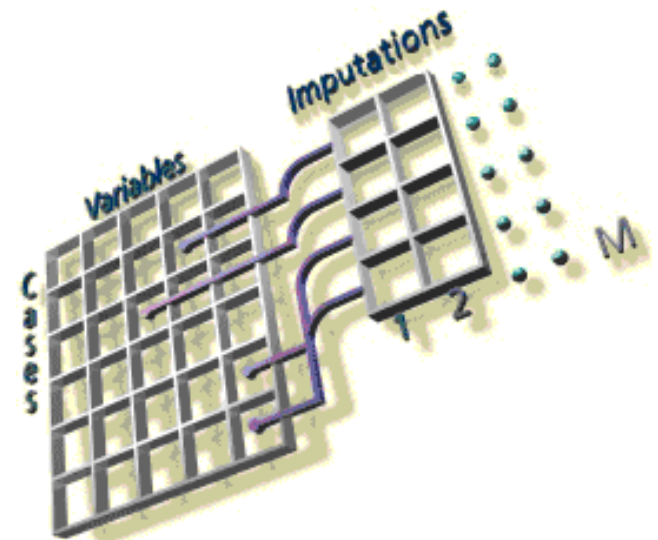


*Without a label, how would you know which was tuna and which was cat food?*



# Chybějící hodnoty a „čištění“ dat – metody řešení

- Odstranění špatného záznamu
- Přiřazení definované hodnoty
- Přiřazení průměrné hodnoty
- Přiřazení hodnoty odvozené od nejbližšího souseda
- Spočtení náhradní hodnoty



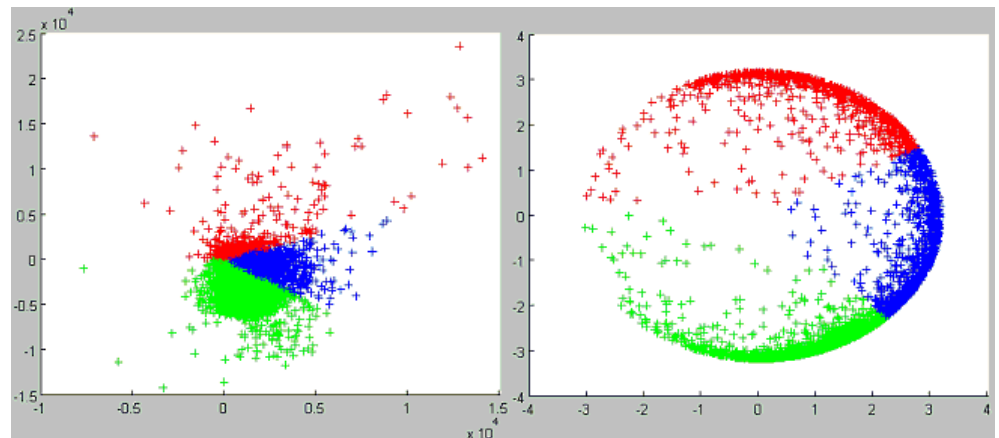


# Normalizace

- Transformace vstupní datové množiny
- Normalizace do intervalu [0.0, 1.0]:

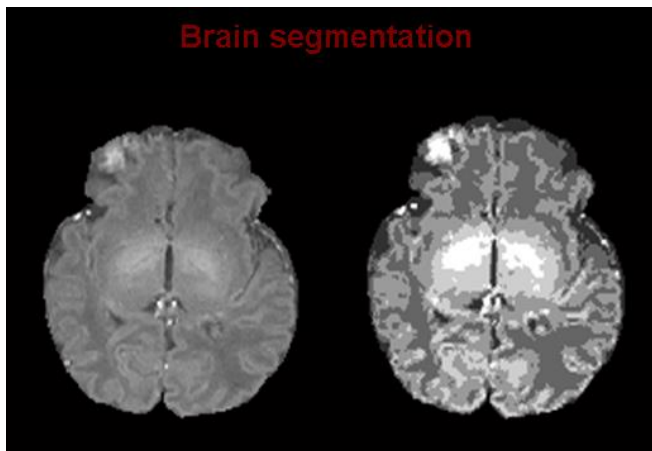
$$d_{\text{normalized}} = (d_{\text{original}} - d_{\text{min}}) / (d_{\text{max}} - d_{\text{min}})$$

- Ořezání podle hraničních hodnot



# Segmentace

- Klasifikace vstupních dat do předem specifikovaných kategorií
- Split-and-merge iterativní algoritmus



[blog.campaigner.com](http://blog.campaigner.com)

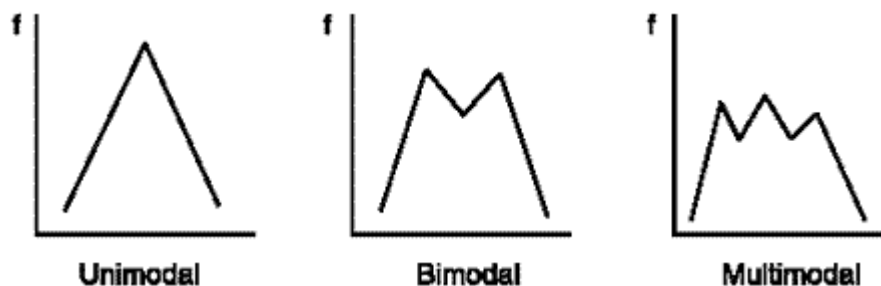
# Split-and-merge

- `similarThresh` = definuje podobnost dvou regionů s danými charakteristikami
- `homogeneousThresh` = definuje homogenitu (uniformitu) regionu

```
do {  
    changeCount = 0;  
    for each region {  
        porovnej region se sousedními regiony a najdi nejpodobnější;  
        if nejpodobnější soused leží uvnitř similarThresh aktuálního regionu {  
            spoj tyto dva regiony;  
            changeCount++;  
        }  
        vyhodnoť homogenitu regionu;  
        if homogenita regionu je menší než homogeneousThresh {  
            rozděl region na dvě části;  
            changeCount++;  
        }  
    }  
} until changeCount == 0
```

# Komplexní části algoritmu

- Určení podobnosti dvou regionů
- Vyhodnocení homogenity regionu – histogram



[www.statcan.gc.ca](http://www.statcan.gc.ca)

- Rozdělení regionu

# Potenciální problém



[www.rootninja.com](http://www.rootninja.com)

- Vytvoření nekonečné smyčky opakování  
rozdělování a spojování stejného regionu
- Řešení:
  - Změna hodnoty prahu podobnosti nebo prahu homogeneity
  - Zahrnutí jiných vlastností regionů (vyhlazení hranic, velikost a tvar regionů)

# Vzorkování a interpolace

- Transformace vstupních dat
- Interpolace = metoda vzorkování
  - Lineární interpolace
  - Bilineární interpolace
  - Nelineární interpolace



# Lineární interpolace



$$(x_C - x_A)/(x_B - x_A) = (d_C - d_A)/(d_B - d_A)$$

$$d_C = d_A + (d_B - d_A) * (x_C - x_A)/(x_B - x_A)$$

$d$  je hodnota proměnné. Dále předpokládáme, že A, B i C leží na ose  $x$ .

- Obecné rozložení bodů v rovině:

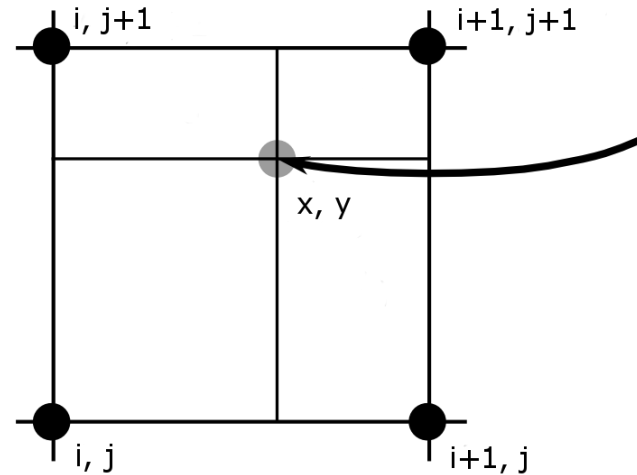
$$P(t) = P_A + Vt, \text{ kde } V = P_B - P_A$$

- Dosadíme za  $P(t)$   $P_C$ , spočteme hodnotu  $t$  a následně:

$$d(t) = d_A + Ut, \text{ kde } U = d_B - d_A$$

# Bilineární interpolace

- Uniformní grid
- Horizontální + vertikální interpolace

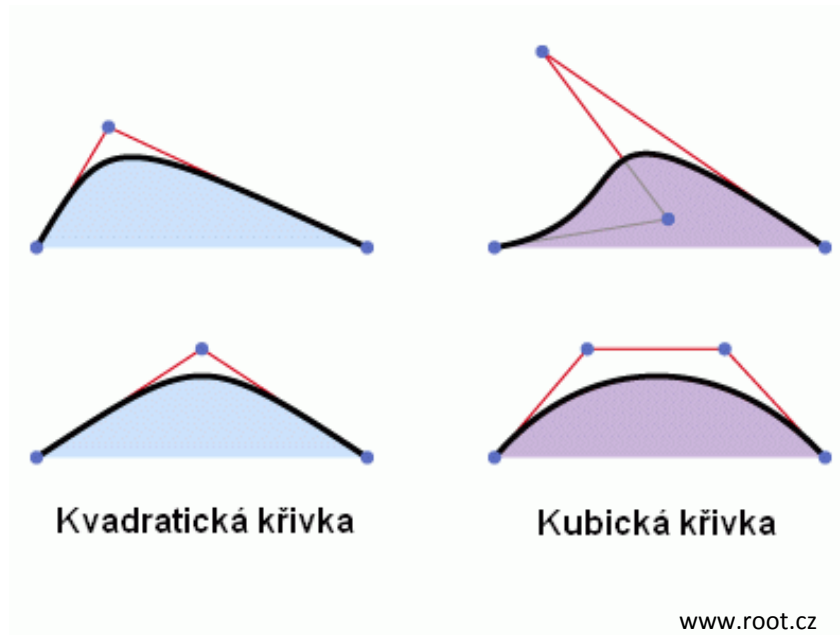


- 1) procentuální vzdálenost bodu  $x$  mezi body  $x$  a  $x+1$ , spočtení  $d$  na pozicích  $(x, j)$  a  $(x, j+1)$
- 2)  $d_{x,y} = d_{x,j} + t * (d_{x,j+1} - d_{x,j})$   
kde  $t$  = procentuální vzdálenost bodu  $y$  mezi body  $j$  a  $j+1$



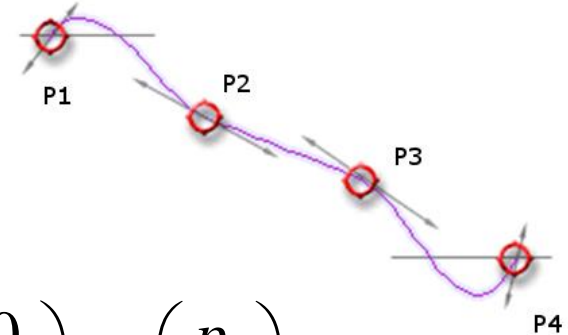
# Nelineární interpolace

- Problémy lineární interpolace – spojitost 0 v bodech mřížky
- Řešení použitím kvadratických a kubických křivek



# Catmull-Rom spline

- 4 kontrolní body



$$q(t) = 0.5 * (1.0 t t^2 t^3) * \begin{pmatrix} 0 & 2 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 2 & -5 & 4 & -1 \\ -1 & 3 & -3 & 1 \end{pmatrix} * \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix}$$

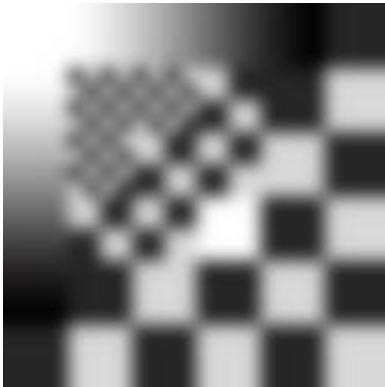
$$q(t) = 0.5 * ((2 * p_1) + (-p_0 + p_2) * t + (2 * p_0 - 5 * p_1 + 4 * p_2 - p_3) * t^2 + (-p_0 + 3 * p_1 - 3 * p_2 + p_3) * t^3)$$

# Výsledek

- Původní obrázek (24x24 pixelů)



kubický B-spline filter

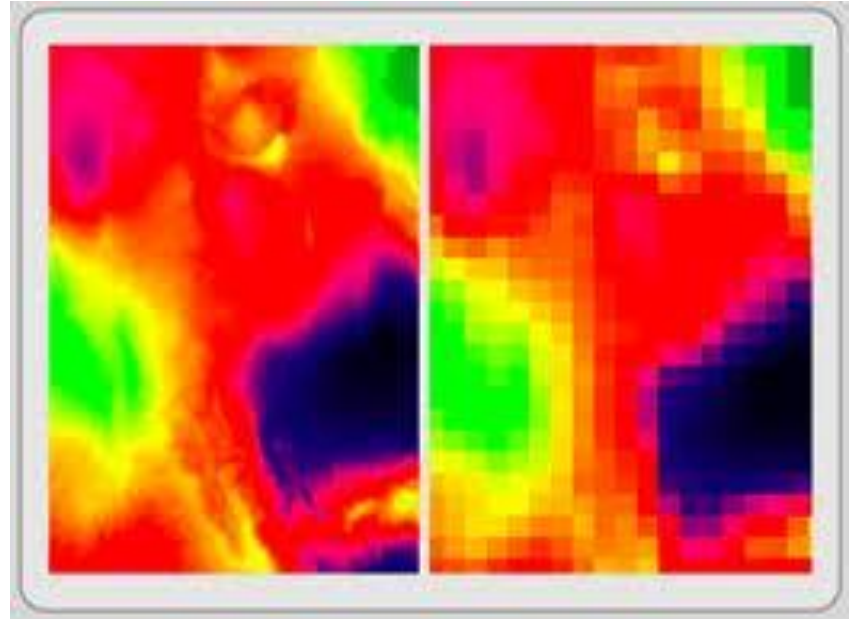


Catmull-Rom



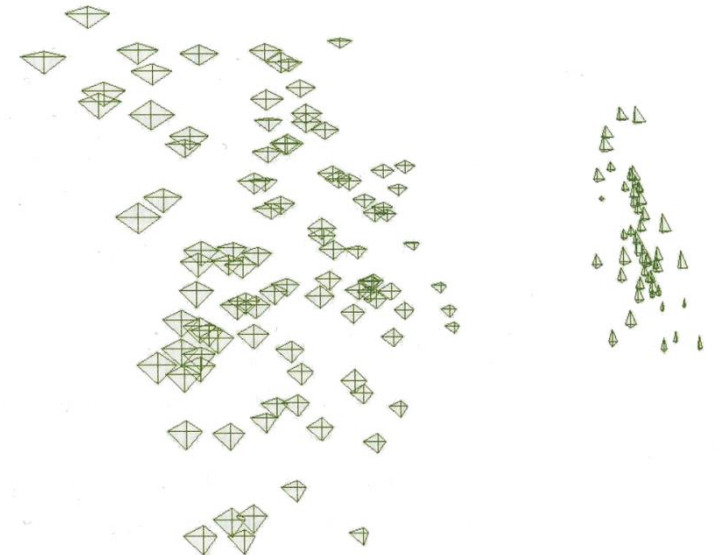
# Metody převzorkování

- Replikace pixelů
- Průměrování susedů
- Data subsetting

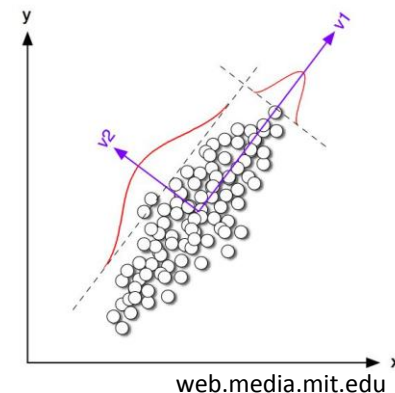


# Snižování dimenze

- např. pro účely zobrazení
- zachovat co nejvíce informací obsažených ve vstupní množině
- techniky:
  - **PCA** (principal component analysis)
  - **MDS** (multidimensional scaling)
  - **SOMs** (Kohonen self-organizing maps)



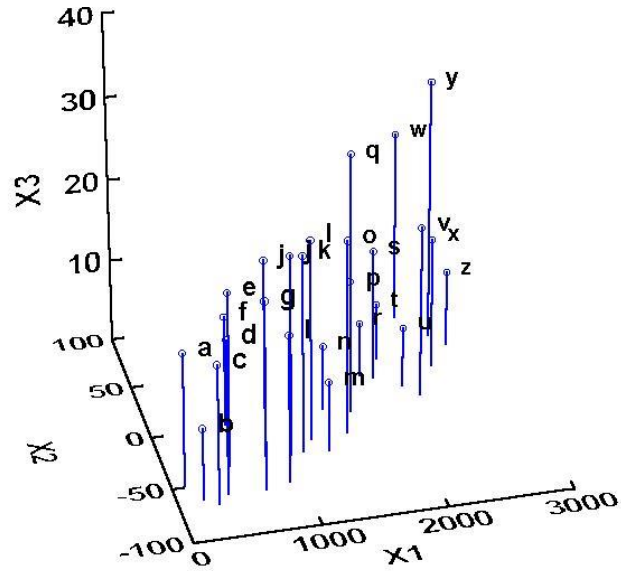
# PCA intuitivně



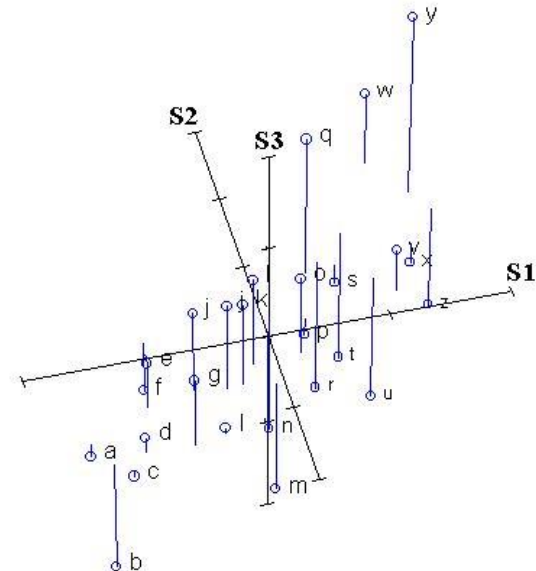
1. Vybereme přímku v prostoru, který zobrazuje  $n$ -dimenzionální data. Tato přímka pokrývá co nejvíce těchto vstupních dat. Přímka pak reprezentuje první hlavní komponentu (PC).
2. Vybereme přímku kolmou k první přímce – ta tvoří druhou PCA.
3. Opakujeme tento postup, dokud nejsou spočteny všechny PC dimenze nebo dokud není dosažen požadovaný počet hlavních komponent (PCAs).

# PCA – principal component analysis

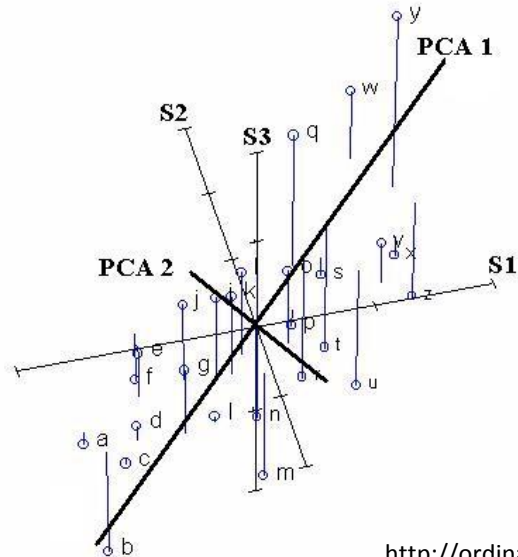
1)



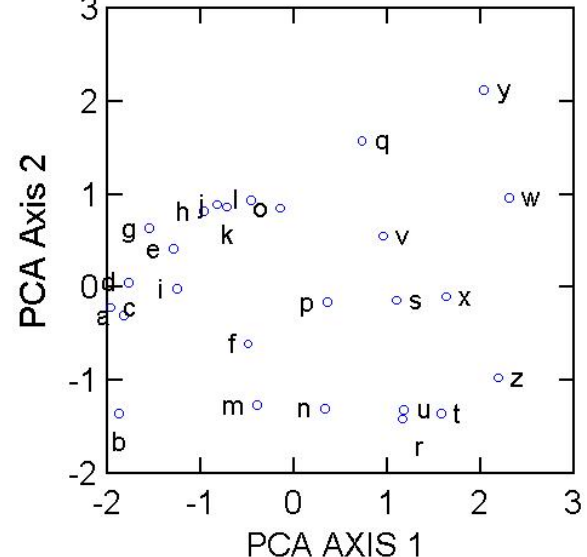
2)



3)

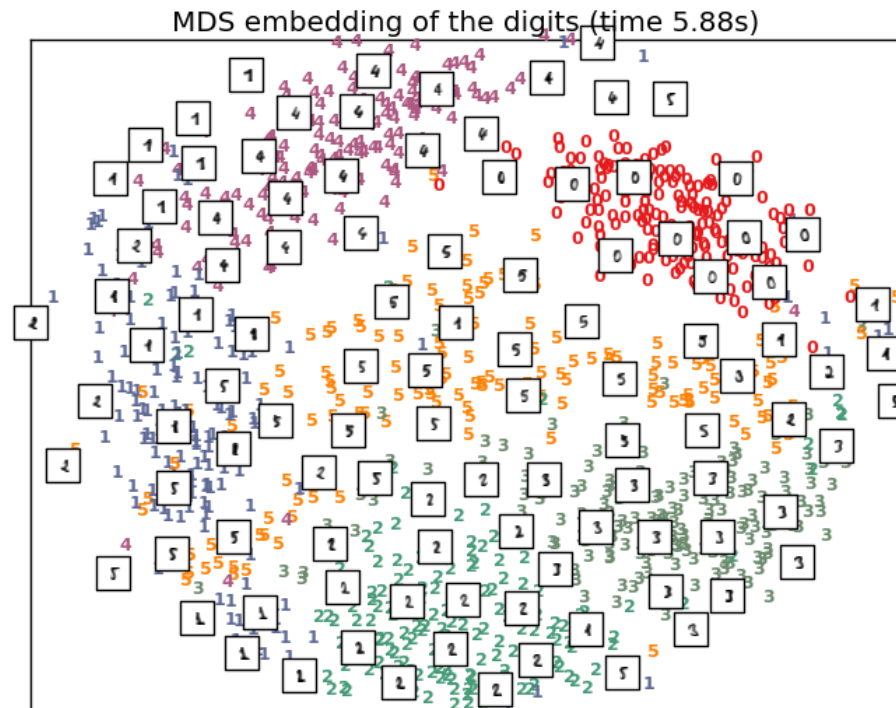


4)



# MDS – multidimensional scaling

- Založen na porovnání vzdáleností mezi jednotlivými body v původním a redukovaném prostoru.





# MDS – multidimensional scaling

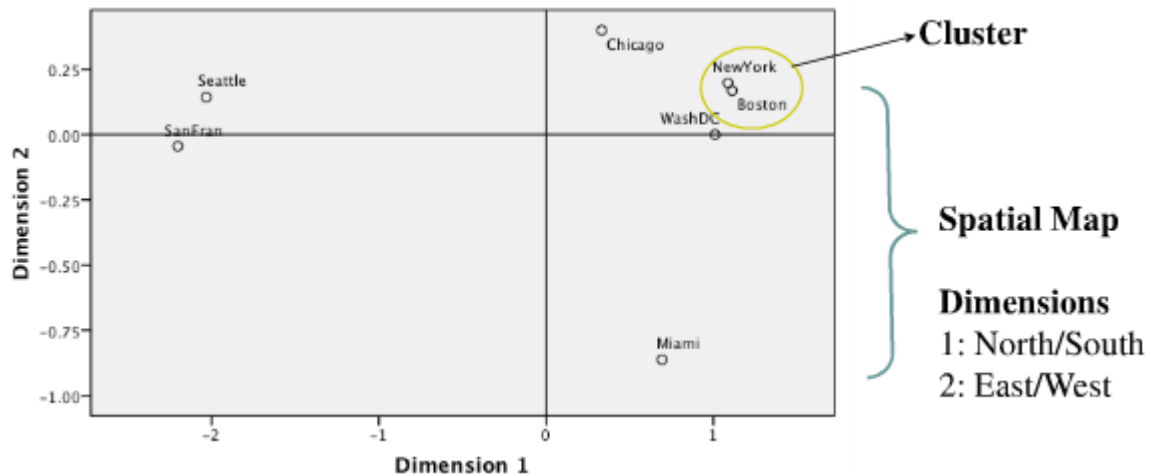
- 1) Spočítáme vzdálenosti všech dvojic bodů v původním prostoru. Pokud máme na vstupu  $n$  bodů, tento výpočet vyžaduje  $n(n - 1)/2$  operací.
- 2) Převédeme všechny body vstupní množiny do prostoru o požadované nižší dimenzi (často náhodně).
- 3) Spočteme *stress*, tedy rozdíl mezi vzdálenostmi bodů v původním a novém prostoru. Tento výpočet může být proveden různými způsoby.
- 4) Jestliže průměrný nebo akumulovaný *stress* je menší než uživatelem definovaný práh, ukončíme celý algoritmus a vrátíme výsledek.
- 5) Jestliže je práh překročen, pak každému datovému bodu spočteme směrový vektor označující směr, ve kterém by se bod měl posunout, abychom redukovali *stress* mezi ním a všemi ostatními body. Ten je určen jako vážený průměr vektorů mezi tímto bodem a všemi jeho sousedy, jeho směr může být k nebo od souseda a váha je odvozena ze *stressu* spočteného mezi jednotlivými páry. Kladná hodnota *stress* body vzájemně oddaluje, záporná přibližuje a čím větší je absolutní hodnota *stressu*, tím větší je pohyb.
- 6) Na základě těchto výpočtů transformujeme datové body do cílové nižší dimenze – s ohledem na spočtené vektory. Vrátime se ke kroku 3 algoritmu.

# MDS - příklad

- Vzdálenost mezi městy

**Distances Matrix:**  
Symmetric

		1	2	3	4	5	6	7	8	9
		BOST	NY	DC	MIAM	CHIC	SEAT	SF	LA	DENV
1	BOSTON	0	206	429	1504	963	2976	3095	2979	1949
2	NY	206	0	233	1308	802	2815	2934	2786	1771
3	DC	429	233	0	1075	671	2684	2799	2631	1616
4	MIAMI	1504	1308	1075	0	1329	3273	3053	2607	2037
5	CHICAGO	963	802	671	1329	0	2013	2142	2054	996
6	SEATTLE	2976	2815	2684	3273	2013	0	808	1131	1307
7	SF	3095	2934	2799	3053	2142	808	0	379	1235
8	LA	2979	2786	2631	2697	2054	1131	379	0	1059
9	DENVER	1949	1771	1616	2037	996	1307	1235	1059	0

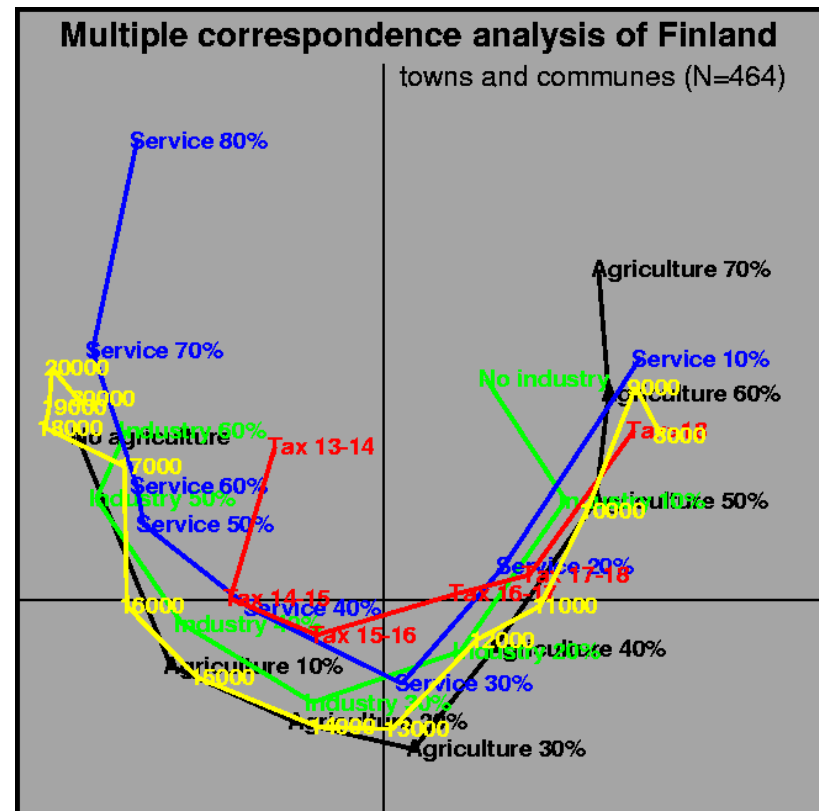


# MDS – multidimensional scaling



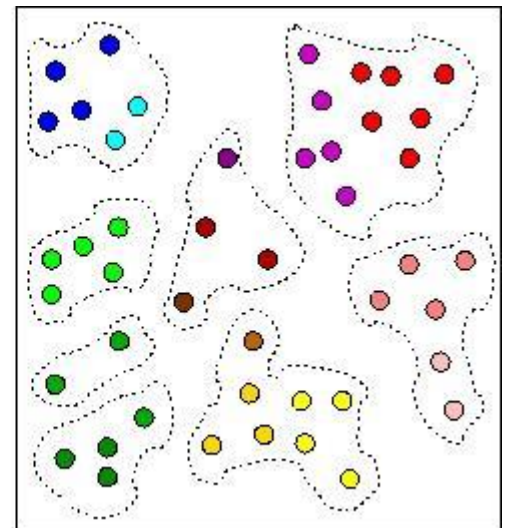
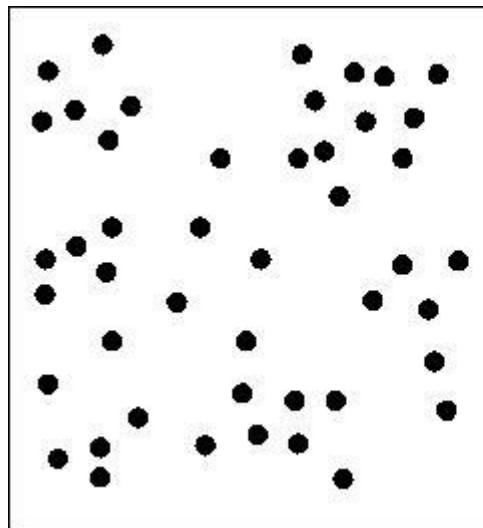
# Mapování nominálních hodnot na ordinální

- Jedna nominální proměnná
  - Použita jako label grafického elementu
- Více nominálních proměnných
  - Multiple correspondence analysis



# Agregace dat

- Agregace = shlukování podobných dat do skupin podle jejich podobnosti. Zahrnuje dvě metody:
  - Shlukování
  - Zobrazení shluků



# Vyhlazování a filtrace

- Běžné při zpracování signálu – odstranění šumu
- **Konvoluce v 1D:**

$$p_i = \frac{p_{i-1}}{4} + \frac{p_i}{2} + \frac{p_{i+1}}{4}$$

# Konverze rastrových dat to vektorových

- Převod je používán při:

- kompresi dat
- porovnání obrázků
- transformaci dat

- Metody:

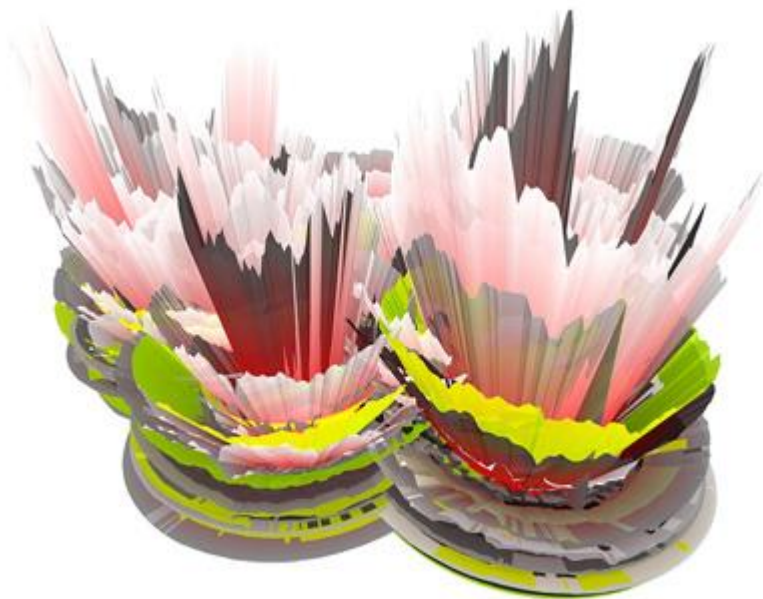
- Prahování
- Narůstání regionů (semínkové vyplňování)
- Detekce hranic
- Ztenčování





# Shrnutí

- Zmíněné techniky zvyšují efektivitu vizualizace
- Nutné uživatele informovat o tom, že data byla transformována



enormousface.com

