



Software

[AcaStat](#)

[GAUSS](#)

[MRDCL](#)

[RATS](#)

[StatsDirect](#)

[ADaMSoft](#)

[GAUSS](#)

[NCSS](#)

[RKWard\[4\]](#)

[Statistix](#)

[Analyse-it](#)

[GenStat](#)

[OpenEpi](#)

[SalStat](#)

[SYSTAT](#)

[ASReml](#)

[Golden Helix](#)

[Origin](#)

[SAS](#)

[The
Unscrambler](#)

[Auguri](#)

[gretl](#)

[Ox programming
language](#)

[SOCR](#)

[UNISTAT](#)

[BioStat](#)

[JMP](#)

[OxMetrics](#)

[Stata](#)

[VisualStat](#)

[BrightStat](#)

[MacAnova](#)

[Origin](#)

[Statgraphics](#)

[Winpepi](#)

[Dataplot](#)

[Mathematica](#)

[Partek](#)

[STATISTICA](#)

[WinSPC](#)

[EasyReg](#)

[Matlab](#)

[Primer](#)

[StatIt](#)

[XLStat](#)

[Epi Info](#)

[MedCalc](#)

[PSPP](#)

[StatPlus](#)

[XploRe](#)

[EViews](#)

[modelQED](#)

[R](#)

[SPlus](#)

[Excel](#)

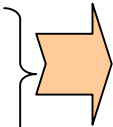
[Minitab](#)

[R Commander\[4\]](#)

[SPSS](#)



Data miningový software

- Cca 20 až 30 dodavatelů
- Hlavní hráči na trhu:
 - Clementine,
 - IBM's Intelligent Miner, 
 - SGI's MineSet,
 - SAS's Enterprise Miner.
- Řada vestavěných produktů:
 - fraud detection:
 - electronic commerce applications,
 - health care,
 - customer relationship management



- Společnost SAS Institute
 - Vznik 1976 v univerzitním prostředí
 - Dnes: největší soukromá softwarová společnost na světě (více než 11.000 zaměstnanců)
 - přes 45.000 instalací
 - cca 9 milionů uživatelů ve 118 zemích
 - v USA okolo 1.000 akademických zákazníků (SAS používá většina vyšších a vysokých škol a výzkumných pracovišť)



Studentské soutěže o hodnotné ceny



SAS - soutěž o nejlepší studentskou práci

Společnost SAS ČR vyhlašuje 1. ročník soutěže o nejlepší studentskou práci s použitím softwaru SAS.

Do soutěže lze přihlášit bakalářskou, diplomovou, dizertační, semestrální nebo ročníkovou práci, která byla předložena k obhajobě nebo obhajoba v kalendářním roce 2009. Přihlášky lze podávat jen prostřednictvím online formuláře na stránkách SAS ČR. Přihlášené práce posoudí tým odborníků společnosti SAS ČR a tři nejlepší práce budou oceněny.

Poslední termín pro podávání přihlášek je 31. 1. 2010

Autori tří nejlepších prací získají:

- 1. místo - 10.000 Kč a účast na SAS Global Forum v Seattlu.** Výhrou bude mít hraznou letenku, ubytování a účastnický poplatek.
- 2. místo - Účast na SAS Global Forum v Seattlu.** Výhrou bude mít hraznou letenku, ubytování a účastnický poplatek.
- 3. místo - iPod Touch**

welcome to SEATTLE!

 www.sas.com/cz/academic
261 176 310
marketing@cs.sas.com

- Lokální (ČR) soutěž o nejlepší studentskou práci
- SAS Student Ambassador - celosvětová soutěž o nejlepší práce s využitím SAS
- Možnost účasti a prezentace na SAS konferenci v Seattlu



Podpora studentů

- Možnost rozšíření licence na domácí instalace pro studenty
 - SAS Fellowship Program – software zdarma pro diplomku či dizertaci
 - Zadávání a vedení diplomových prací
 - Sdílení informací, zkušeností či příkladů v uživatelských skupinách
- Interaktivní moduly nebo programovací prostředí
 - Statistická analýza
 - Matice
 - Časové řady
 - Operační výzkum
 - Kontrola kvality



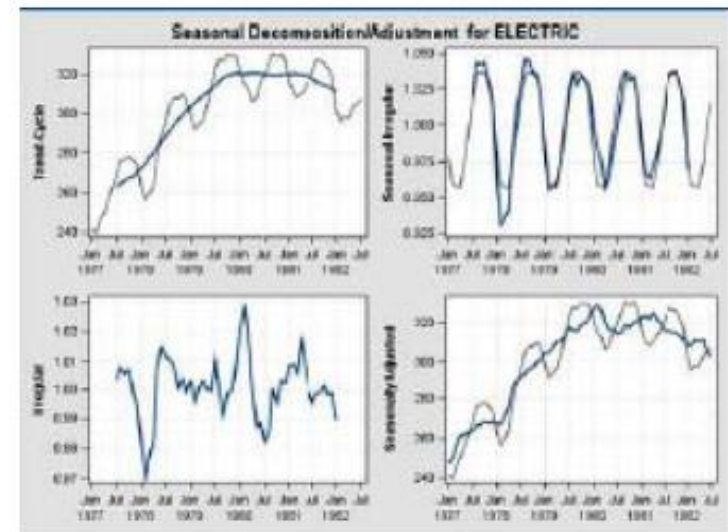
□ Statistická analýza:

- Popisná statistika
- Analýza kontingenčních (frekvenčních) tabulek
- Regresní, korelační, kovarianční analýza
- Logistická regrese
- Analýza rozptylu
- Testování hypotéz
- Diskriminační analýza
- Shluková analýza
- Analýza přežití
- ...





- Analýza časových řad:
 - Regresní modely
 - Modely se sezónními faktory
 - Autoregresní modely
 - ARIMA
 - Metody exponenciálního vyrovnání
 - ...

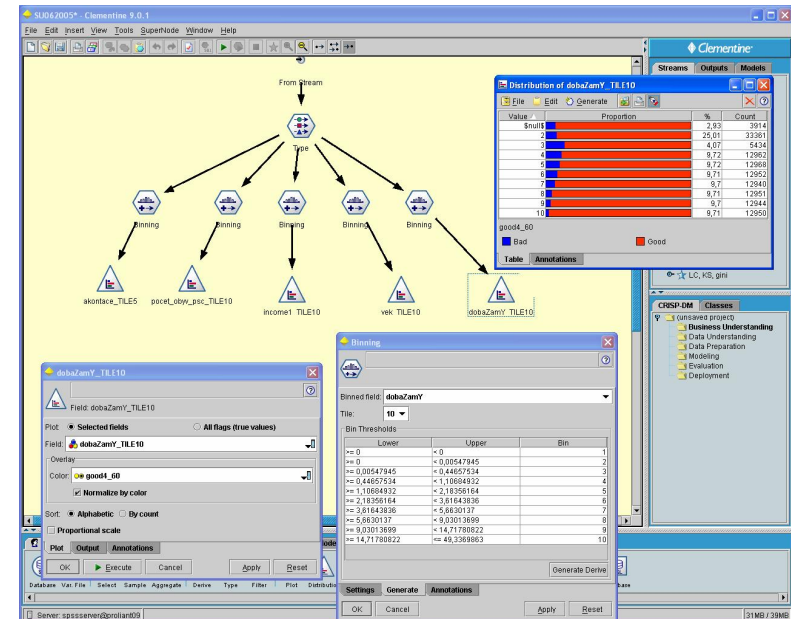
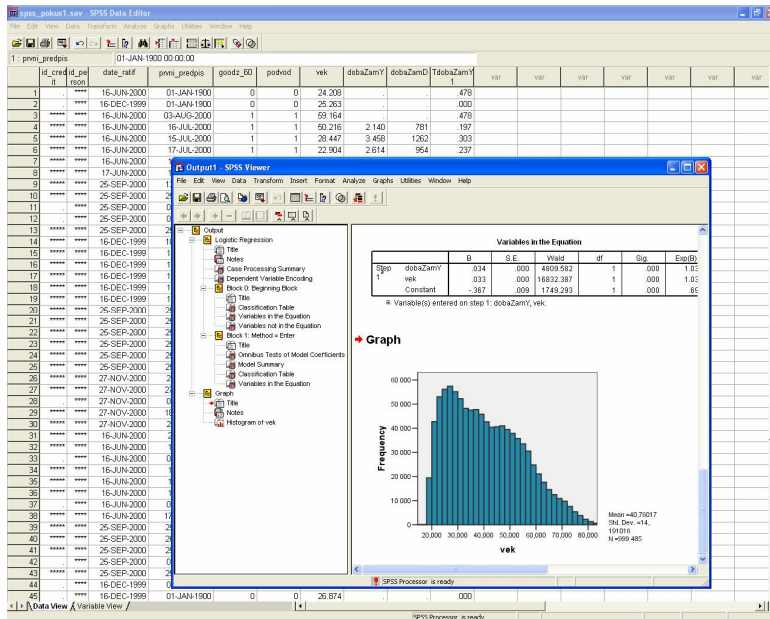




- ❑ Více o SASu: <http://www.sas.com/offices/europe/czech/>
- ❑ (neúplný) seznam komerčních společností využívající SAS: <http://www.sas.com/offices/europe/czech/reference/list.html>
- ❑ o akademickém programu: <http://www.sas.com/offices/europe/czech/academic/index.html>
- ❑ o konferenci SAS forum: http://www.sas.com/reg/offer/cz/2010_sas_forum_2010

Software -SPSS

www.spss.cz



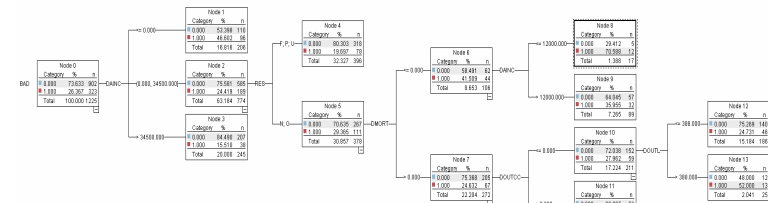


- IBM SPSS/ PASW Modeler 13 (dříve Clementine)

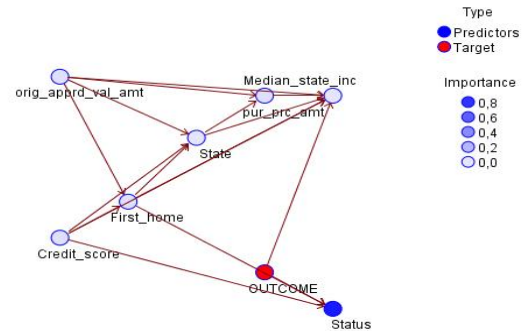
http://www.spss.cz/ibmspss_modeler.htm

The screenshot shows the IBM SPSS Modeler 13 interface. The main workspace contains a workflow with nodes for 'select', 'estirpome', 'diff', 'generated', 'Anomaly', 'Table', 'claimvalue', and 'name'. A 'Preview from Anomaly Node' window is open, displaying a table with 10 records:

id	name	region	farmsize	rainfall	landquality	farmincome
1	id802	name802	north	1780	42	9 734118.000 n
2	id885	name885	southwest	600	48	4 108845.000 n
3	id891	name891	southwest	920	108	8 825974.000 n
4	id717	name717	southwest	1960	86	5 786081.000 n
5	id773	name773	midlands	500	15	7 56220.900 n
6	id779	name779	north	1700	68	3 237891.000 n
7	id797	name797	north	1520	37	3 179227.000 n
8	id883	name883	north	1820	102	9 1296780.0 n
9	id897	name897	midlands	580	27	4 80811.200 n
10	id899	name899	north	480	35	5 81008.400 n



Bayesian Network



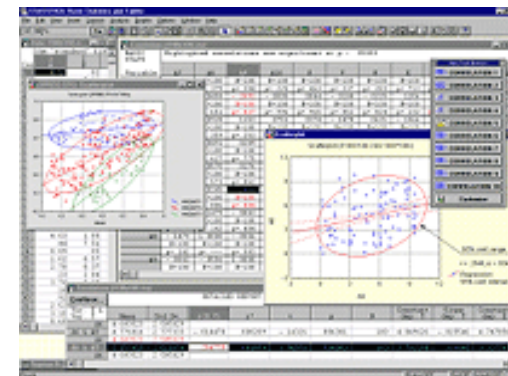
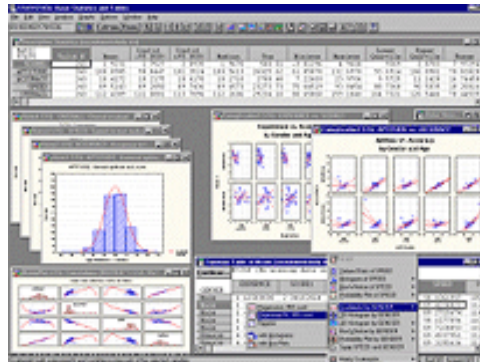
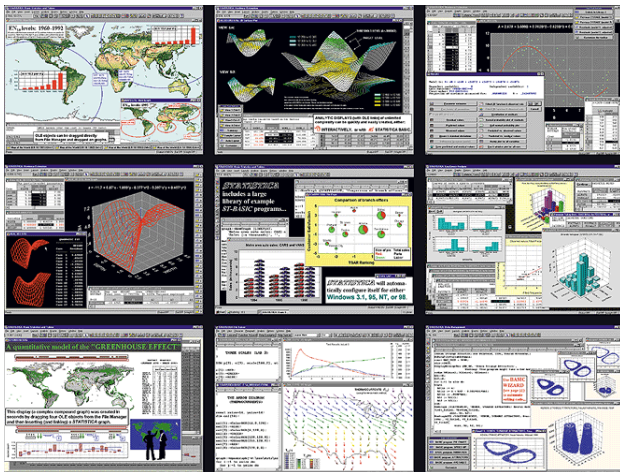


- Více o IBM SPSS Modeler 13 (dříve Clementine):
http://www.spss.cz/ibmspss_modeler.htm
- (neúplný) seznam zákazníků:
<http://www.spss.cz/zakaznici.htm>
- Akademický program:
<http://www.spss.com/academic/>

Software -Statistica

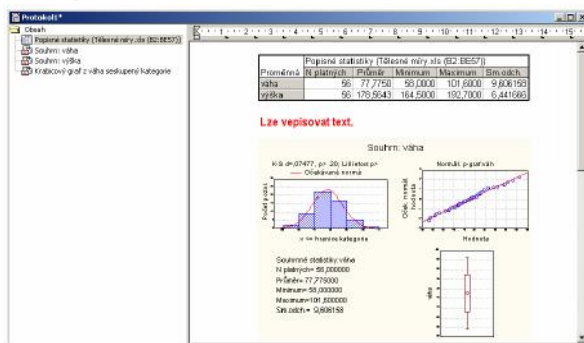


Statistica: www.statistica.cz



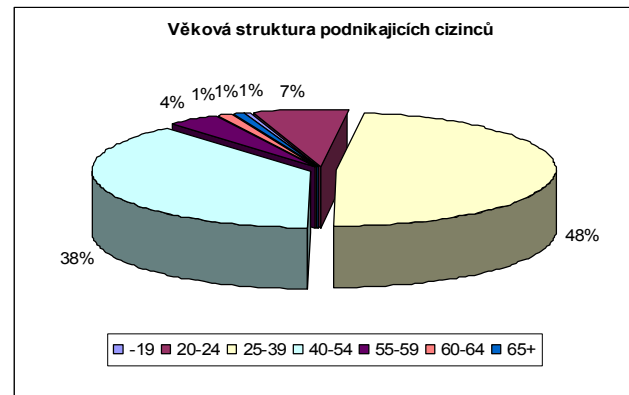
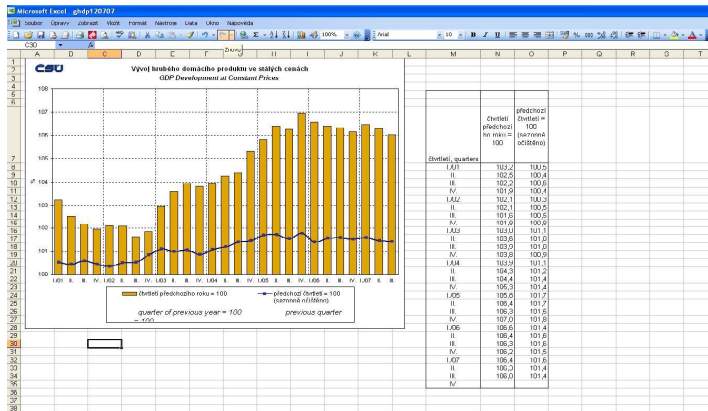
Statistica

- Více o Statistica Data Miner: <http://www.statistica.cz/produkty/5-dataminingove-nastroje/21-statistica-data-miner/detail/>
- (neúplný) seznam zákazníků: <http://www.statsoft.com/customers/>
- Akademický program: <http://www.statsoft.com/academic/>
- Petra Beranová – Stručný manuál k ovládnání programu STATISTICA:
http://www.statsoft.cz/download/soubory/STATISTICA_manual.pdf

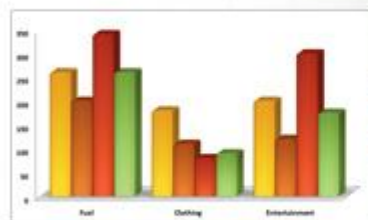


Software

- MS Excel: <http://office.microsoft.com/en-us/excel/default.aspx>



Excel 2007



See what's new ▶

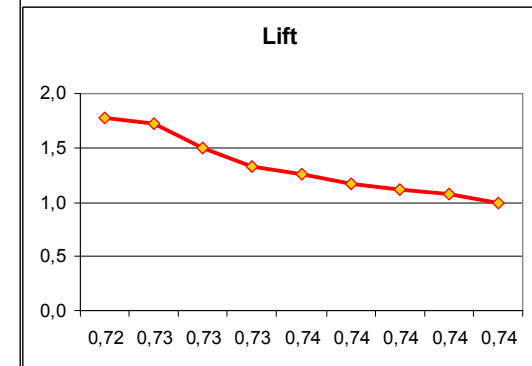
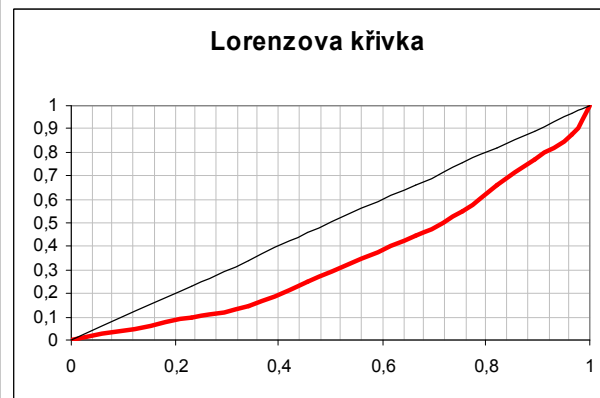
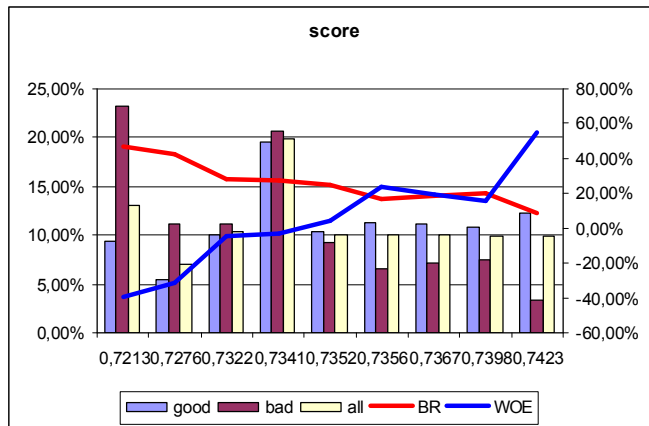
<http://office.microsoft.com/en-us/excel/HA100738731033.aspx>

Software

■ MS Excel:

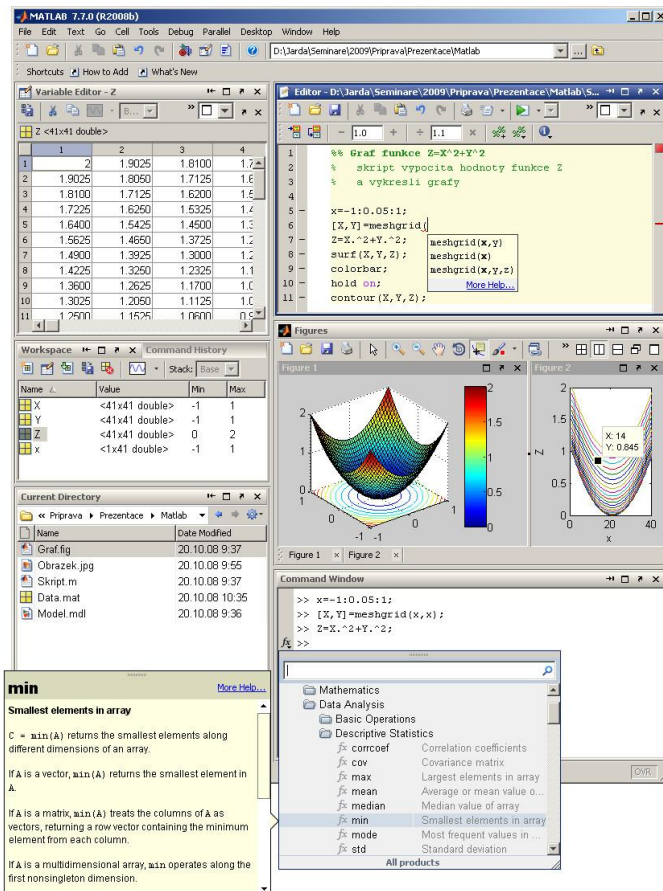
Počet z id	BAD		Celkový součet
	,00	1,00	
score k			
,72	9,42%	23,22%	13,06%
,73	5,43%	11,15%	6,94%
,73	9,98%	11,15%	10,29%
,73	19,51%	20,74%	19,84%
,74	10,31%	9,29%	10,04%
,74	11,31%	6,50%	10,04%
,74	11,09%	7,12%	10,04%
,74	10,75%	7,43%	9,88%
,74	12,20%	3,41%	9,88%
Celkový součet	100,00%	100,00%	100,00%

	good	bad	all	BR	WOE
0,721263	9,42%	23,22%	13,06%	46,88%	-0,392
0,727551	5,43%	11,15%	6,94%	42,35%	-0,312
0,732201	9,98%	11,15%	10,29%	28,57%	-0,048
0,734083	19,51%	20,74%	19,84%	27,57%	-0,027
0,735168	10,31%	9,29%	10,04%	24,39%	0,045
0,735632	11,31%	6,50%	10,04%	17,07%	0,240
0,736706	11,09%	7,12%	10,04%	18,70%	0,192
0,739753	10,75%	7,43%	9,88%	19,83%	0,161
0,742267	12,20%	3,41%	9,88%	9,09%	0,554



Software

Matlab



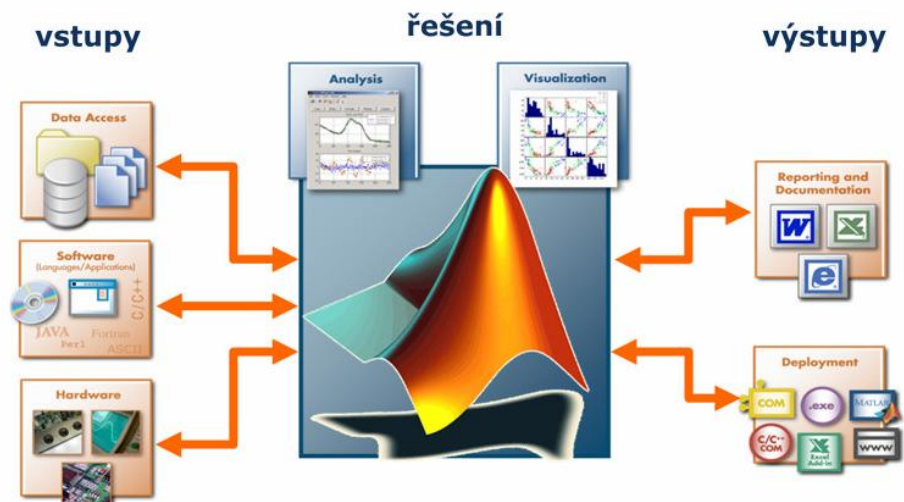
```

1 % Graf funkce Z=X^2+Y^2
2 % skript vypocita hodnoty funkce Z
3 % a vykresli grafy
4
5 x=-1:0.05:1;
6 [X,Y]=meshgrid(x);
7 Z=X.^2+Y.^2;
8 surf(X,Y,Z);
9 colorbar;
10 hold on;
11 contour(X,Y,Z);
    
```

1	2	3	4
1.9025	1.8050	1.7125	1.6200
1.8100	1.7125	1.6200	1.5275
1.7225	1.6250	1.5325	1.4400
1.6400	1.5425	1.4500	1.3575
1.5625	1.4650	1.3725	1.2800
1.4900	1.3925	1.3000	1.2075
1.4225	1.3250	1.2325	1.1400
1.3600	1.2625	1.1700	1.0775
1.3025	1.2050	1.1125	1.0200
1.2500	1.1525	1.0600	0.9675

min
Smallest elements in array
C = min(A) returns the smallest elements along different dimensions of an array.
If A is a vector, min(A) returns the smallest element in A.
If A is a matrix, min(A) treats the columns of A as vectors, returning a row vector containing the minimum element from each column.
If A is a multidimensional array, min operates along the first nonsingleton dimension.


<http://www.humusoft.cz/produkty/matlab/matlab/>



Software -MU

- <https://inet.muni.cz/app/soft/licence>

Nabídka softwaru

 Aplikace je určená pro registraci softwaru a následné získání přístupu k instalačním klíčům a dalším informacím (popř. přístup k samotnému softwaru). Přihlášený uživatel si může nechat zobrazit dostupný software podle zvolené kategorie a aktuálnosti. Po zvolení určité kategorie se zobrazí tabulka dostupného softwaru. Po kliknutí na 'Popis' je v některých případech nutné při první návštěvě odsouhlasit licenční ujednání a následně zadat počet licencí (počet počítačů, na kterých bude software provozován). Po potvrzení již budou nabídnuty veškeré dostupné informace ke konkrétnímu softwaru. Zde je možné i nadále měnit počet licencí. Pokud je dostupný soubor s určitou instalační verzí, tak pro jeho stažení na disk stačí jen kliknout odkaz "Stáhnout" a pokračovat dle instrukcí internetového prohlížeče.

Software

Výběr kategorie softwaru:

Pouze aktuální software (platný)

Pouze volné licence

Název softwaru	Lokalizace	Popis	Platnost od	Platnost do
▼ SPSS CR, spol. s r.o.				
Clementine 13 (PASW Modeler 13)	EN - Anglická verze	Akademická multilicence pro MU 2010	04.01.2010	31.01.2011
SPSS 18 (PASW Statistics 18)	EN - Anglická verze	Akademická multilicence pro MU 2009	09.12.2009	01.02.2011
▼ StatSoft				
Statistica 9.0	CZ - Česká verze	Jednouzivatelská verze	09.12.2009	31.12.2010
Statistica 9.0	EN - Anglická verze	Jednouzivatelská verze	30.09.2009	31.12.2010
Statistica 9.1	EN - Anglická verze	Jednouzivatelská verze	10.03.2010	31.12.2010

Kontaktní e-mail soft-inet@ics.muni.cz
Informace o serveru inet.muni.cz

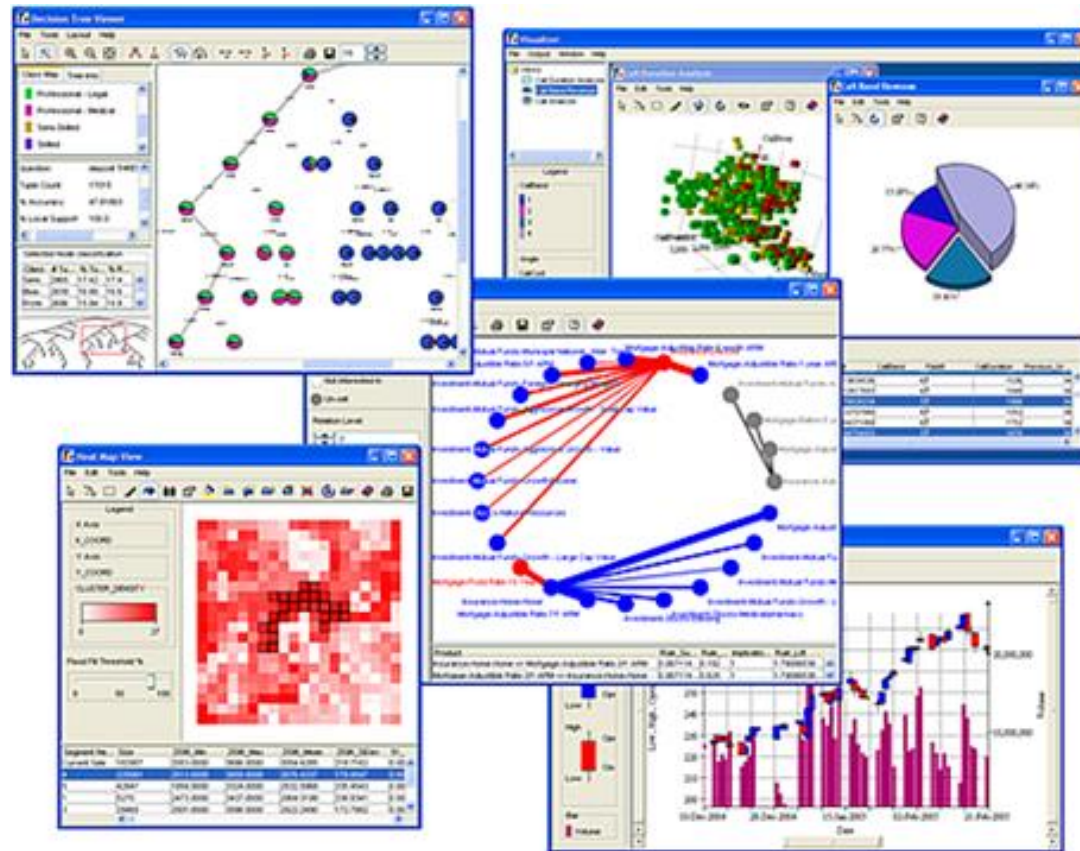
- Matlab 2009a: ÚVT MU
<http://www.muni.cz/ics/services/software>

GIGO

- Garbage in, Garbage out (smetí dovnitř, smetí ven)
 - sebelepší model/proces/software nevyrobí ze smetí nic jiného než opět smetí.



Vizualizace dat



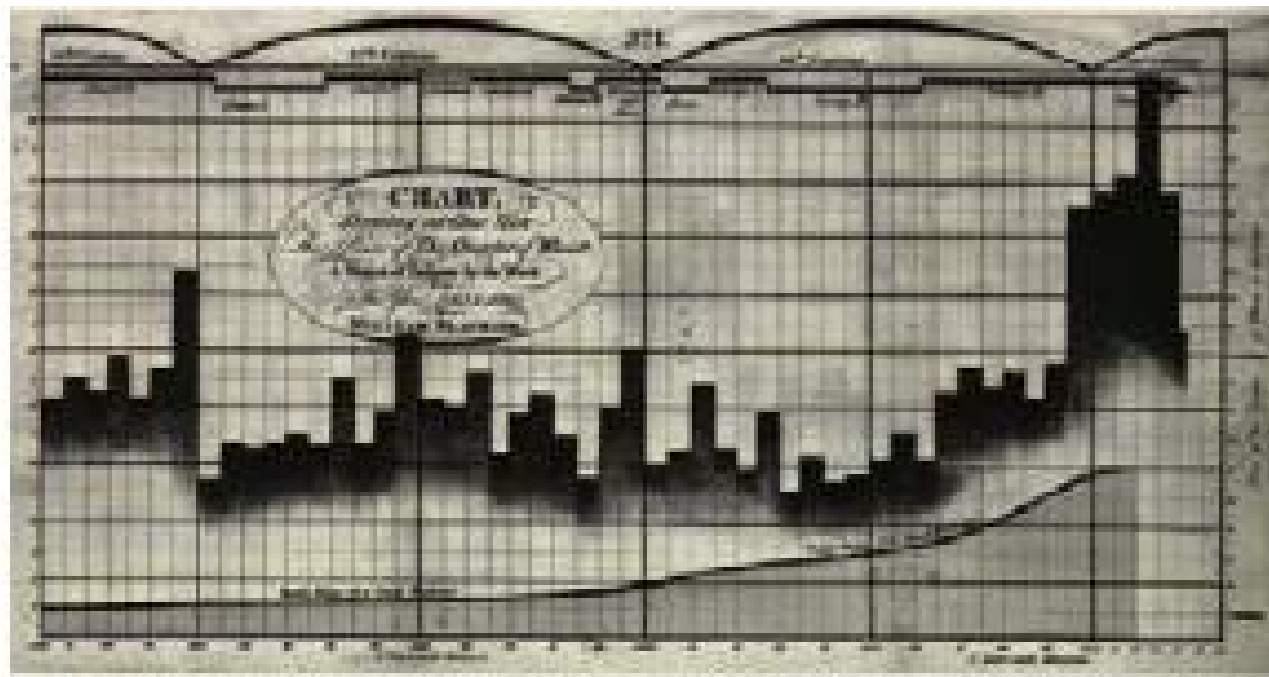


Vizualizace – zdroje

- Na prvním místě se obvykle citují knihy prof. Tufteho, např. Tufte E.R. (1983) The Visual Display of Quantitative Information, Graphic Press, Cheshire, Conn.
- Weby o vizualizaci, např.
 - <http://www.math.yorku.ca/SCS/Gallery/noframes.html> - galerie s poučným výkladem a příklady i nezdařených či lživých grafů
 - <http://www.agocg.ac.uk/> - John Lansdown (1992) Aspects of Design in Computer Graphics: Some Notes –
<http://www.agocg.ac.uk/train/hitch/hitch.htm>
- Jiné weby, např. stránky různých vizualizačních programů a organizací
 - <http://www.cybergeography.org/atlas/atlas.html> nebo <http://miner3d.com/products/gallery.html>

Vizualizace – historie

- William Playfair, 1786: první publikovaná prezentační grafika



Vizualizace – historie

- Dr. John Snow, 1845: epidemie cholery v Londýně



Vizualizace – historie

- Florence Nightingale, 1858: důvody úmrtí v průběhu Krymské války (1853-1856)



Vizualizace – historie

- Harry Beck, 1931: schéma Londýnského metra



Vizualizace – investigativní analýza



<http://www.i2inc.com/>



Law Enforcement

- » Counterterrorism
- » Narcotics investigations
- » Organized crime
- » Intelligence analysis
- » Fraud
- » Missing persons
- » Major investigations
- » Counterfeiting
- » Immigration control
- » Major event security
- » Money laundering
- » Gang investigations

Government

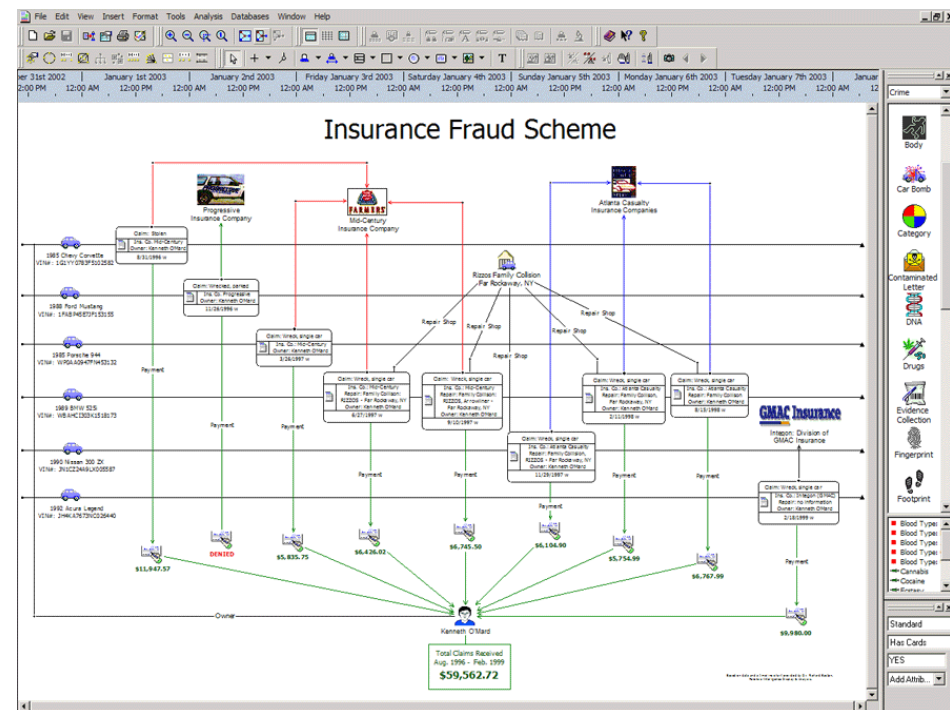
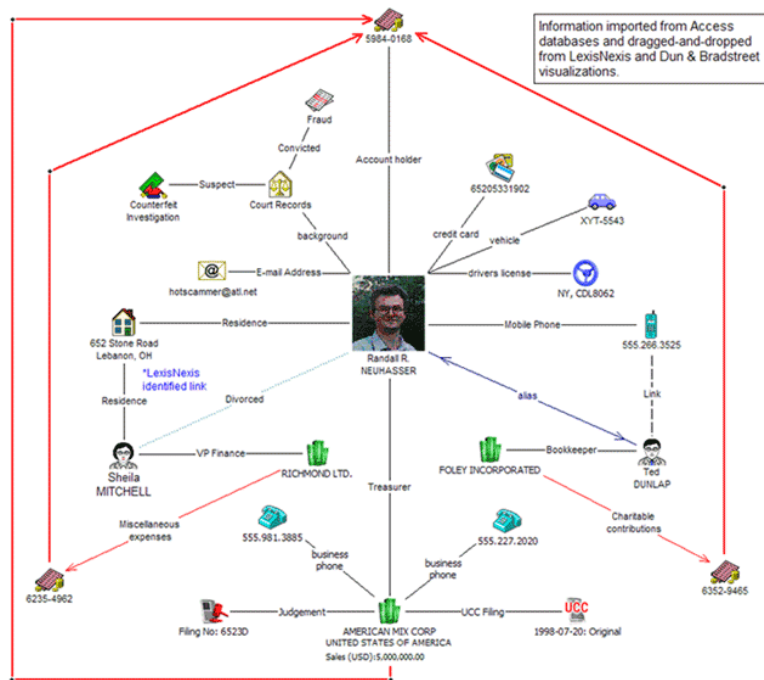
- » Criminal prosecutions
- » National security
- » Military intelligence
- » Embassy security
- » Postal inspection and fraud
- » Prison investigations
- » Park and wildlife services
- » Antitrust investigations
- » Tax fraud investigations
- » Customs investigations

Commercial

- » Forensic accounting
- » Money laundering
- » Insider trading violations
- » Corporate security
- » Anti-pirating investigations
- » Entertainment copyright violations
- » Competitive intelligence
- » Civil lawsuits
- » Fraud:
 - » Credit card
 - » Insurance
 - » Retail
 - » Health care
 - » Commercial
 - » Telephone

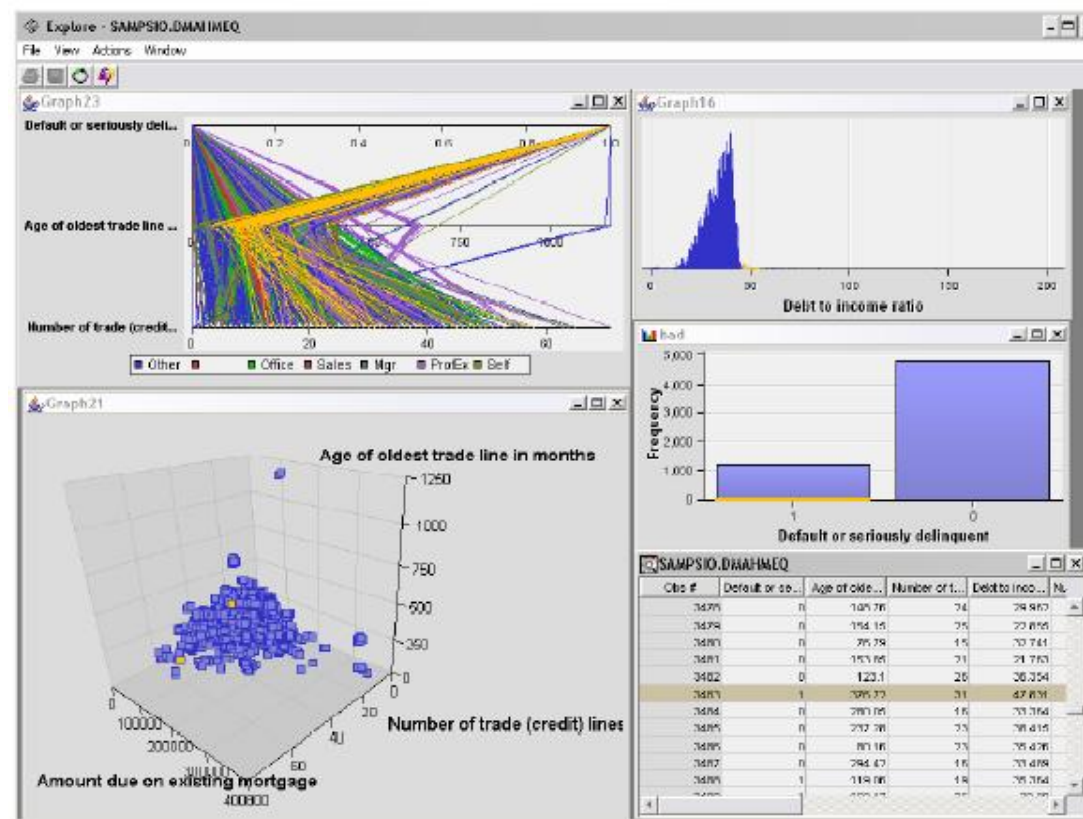
Vizualizace – investigativní analýza

□ osobní kontakty, pojistné podvody



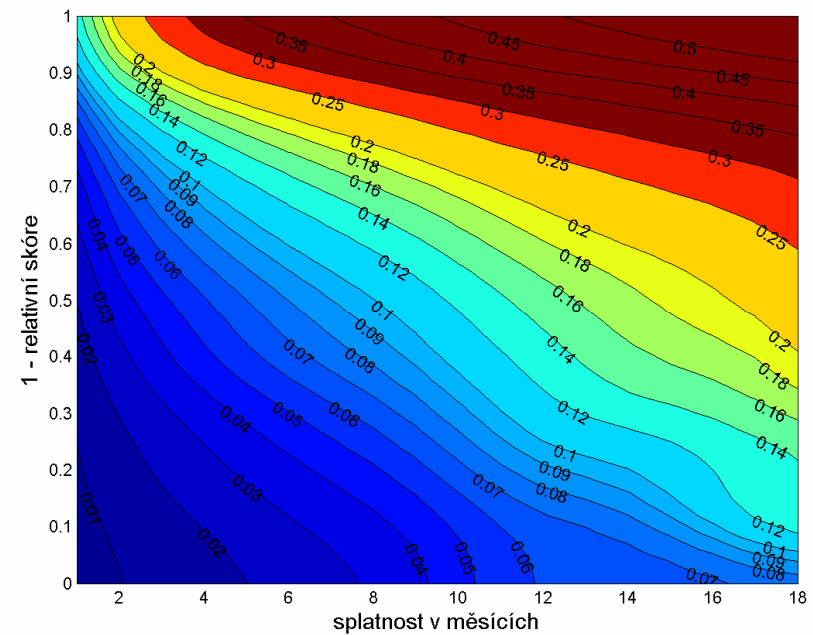
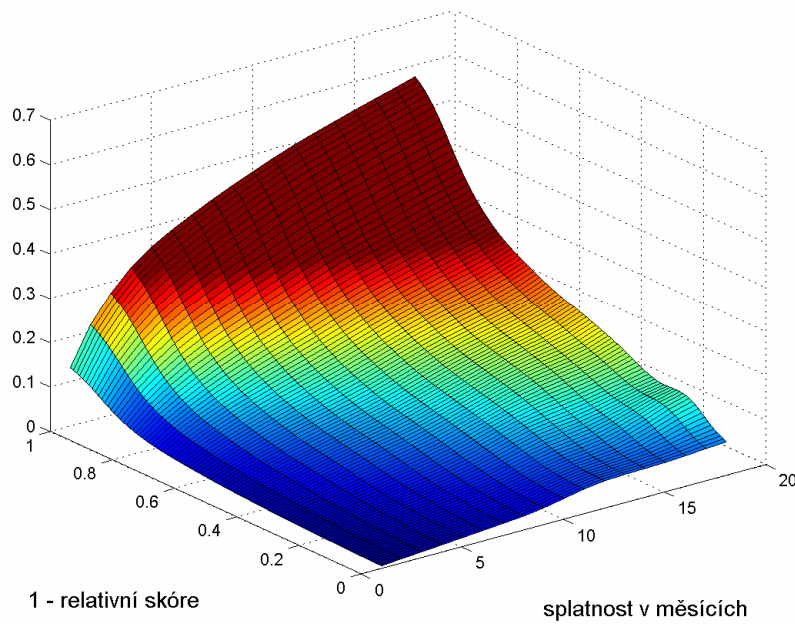
Vizualizace – portfolio management

- Hledání závislostí v datech:



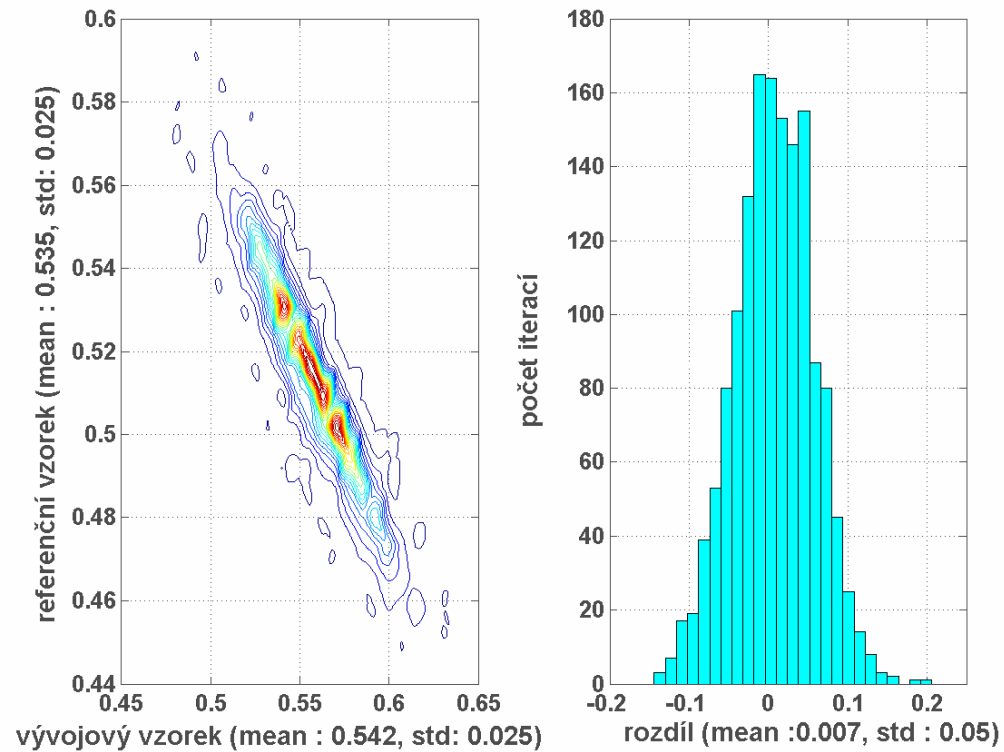
Vizualizace – portfolio management

□ Distribuční funkce doby do defaultu



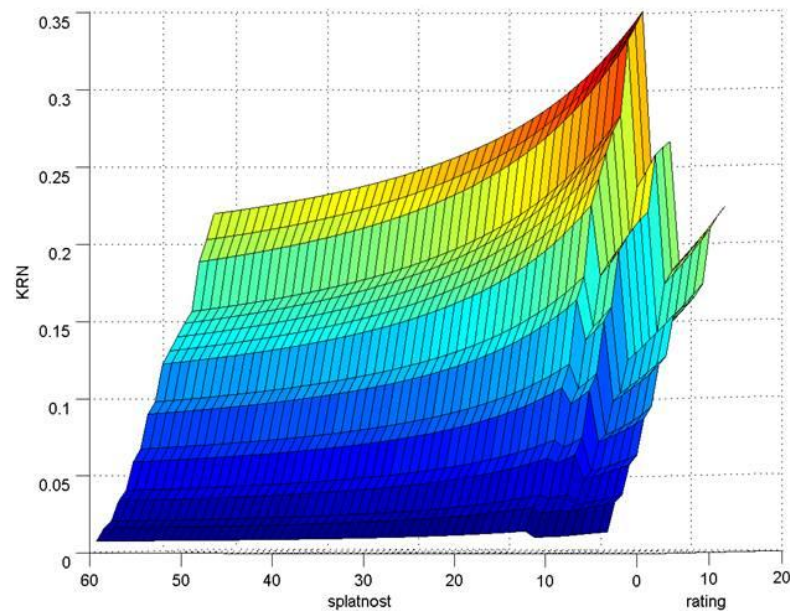
Vizualizace – portfolio management

□ Test stability scóringové funkce



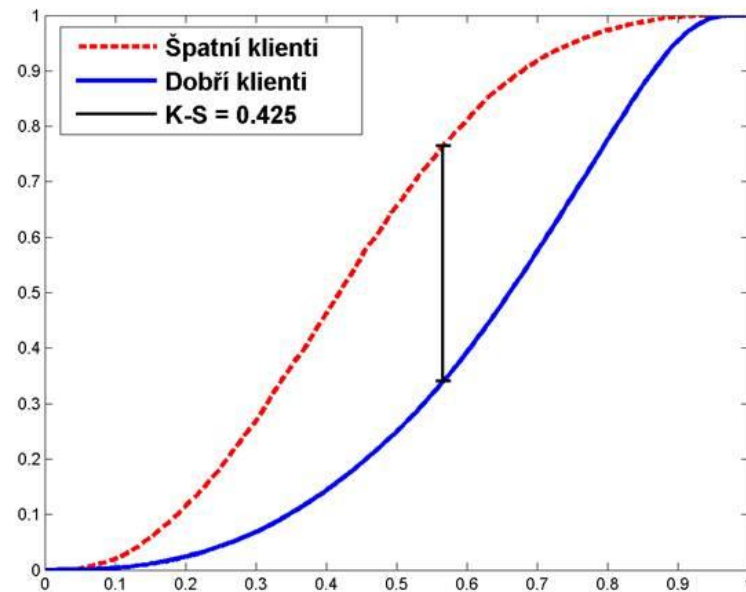
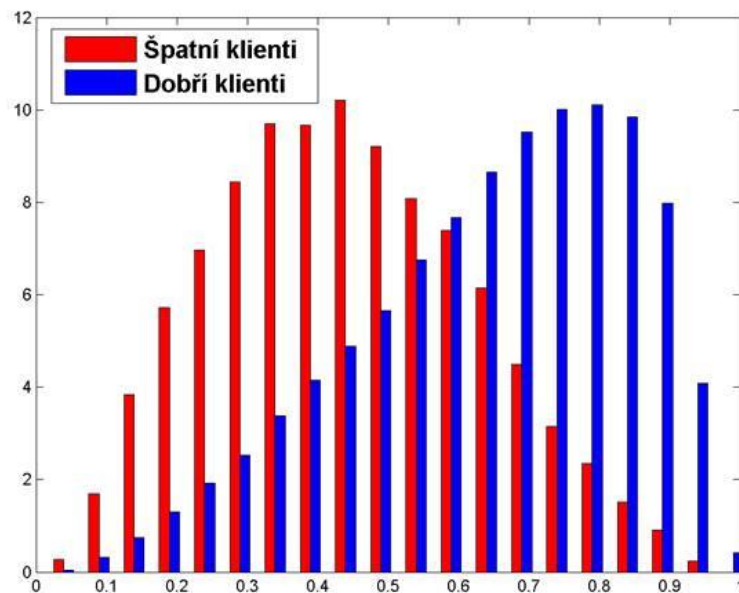
Vizualizace – portfolio management

- Vizualizace kreditních rizikových nákladů (KRN)



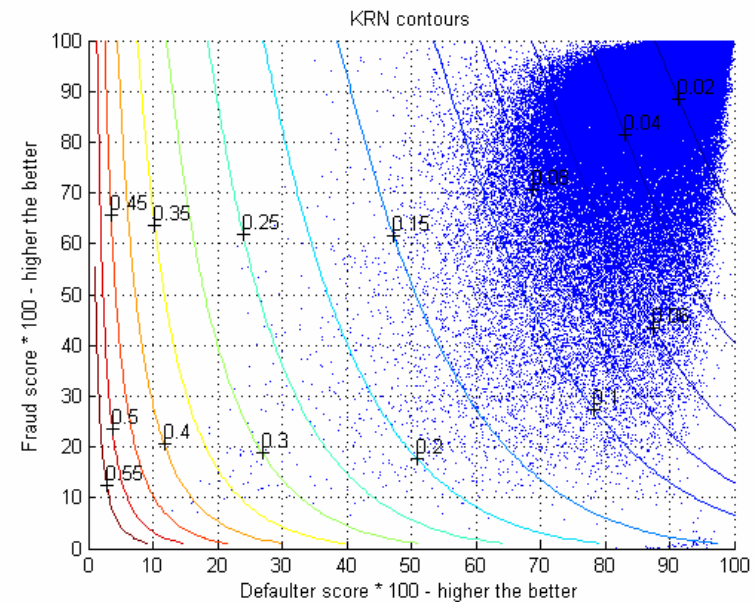
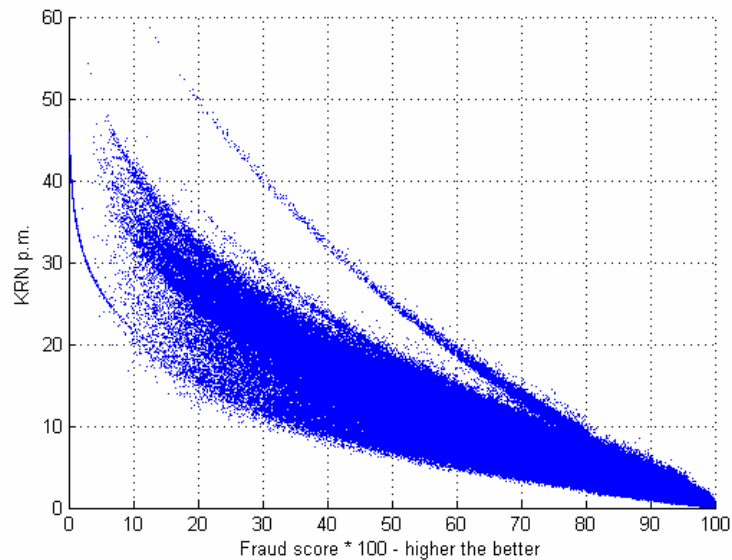
Vizualizace – portfolio management

- Histogram, Distribuční funkce



Vizualizace – portfolio management

□ Bodové grafy



Vizualizace – portfolio management

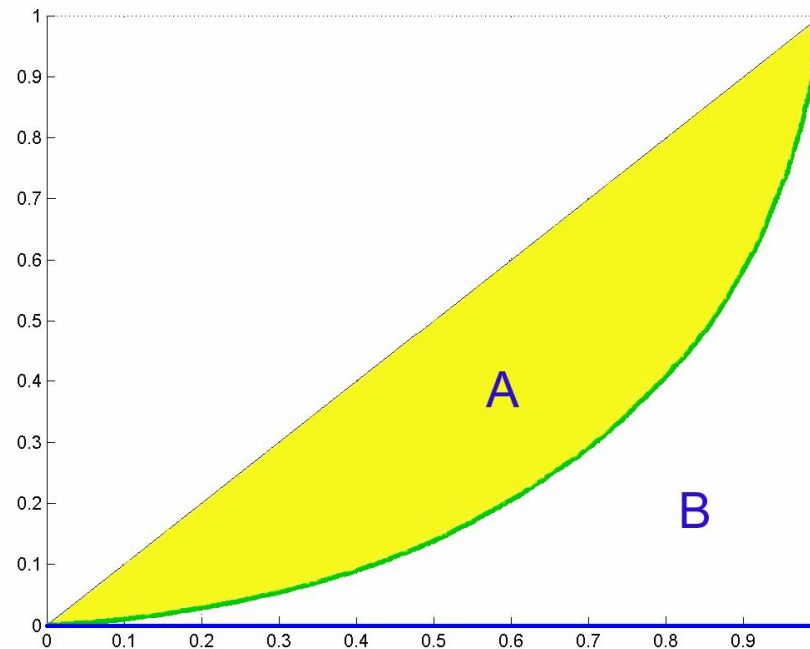
□ Lorenzova křivka

$$Gini = \frac{A}{A+B}$$

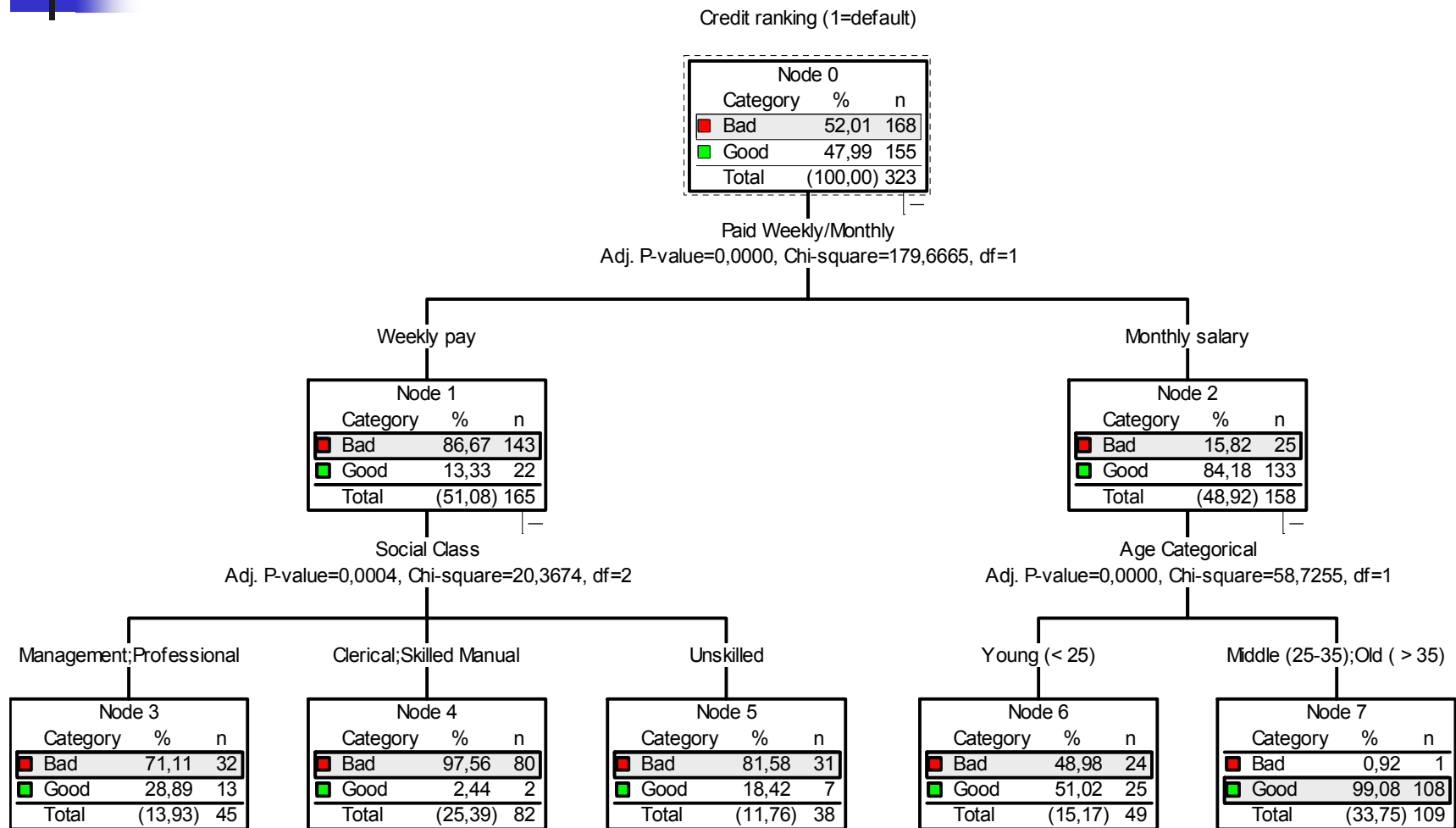
$$Gini = 2A$$

$$c_{stat} = A + \frac{1}{2}$$

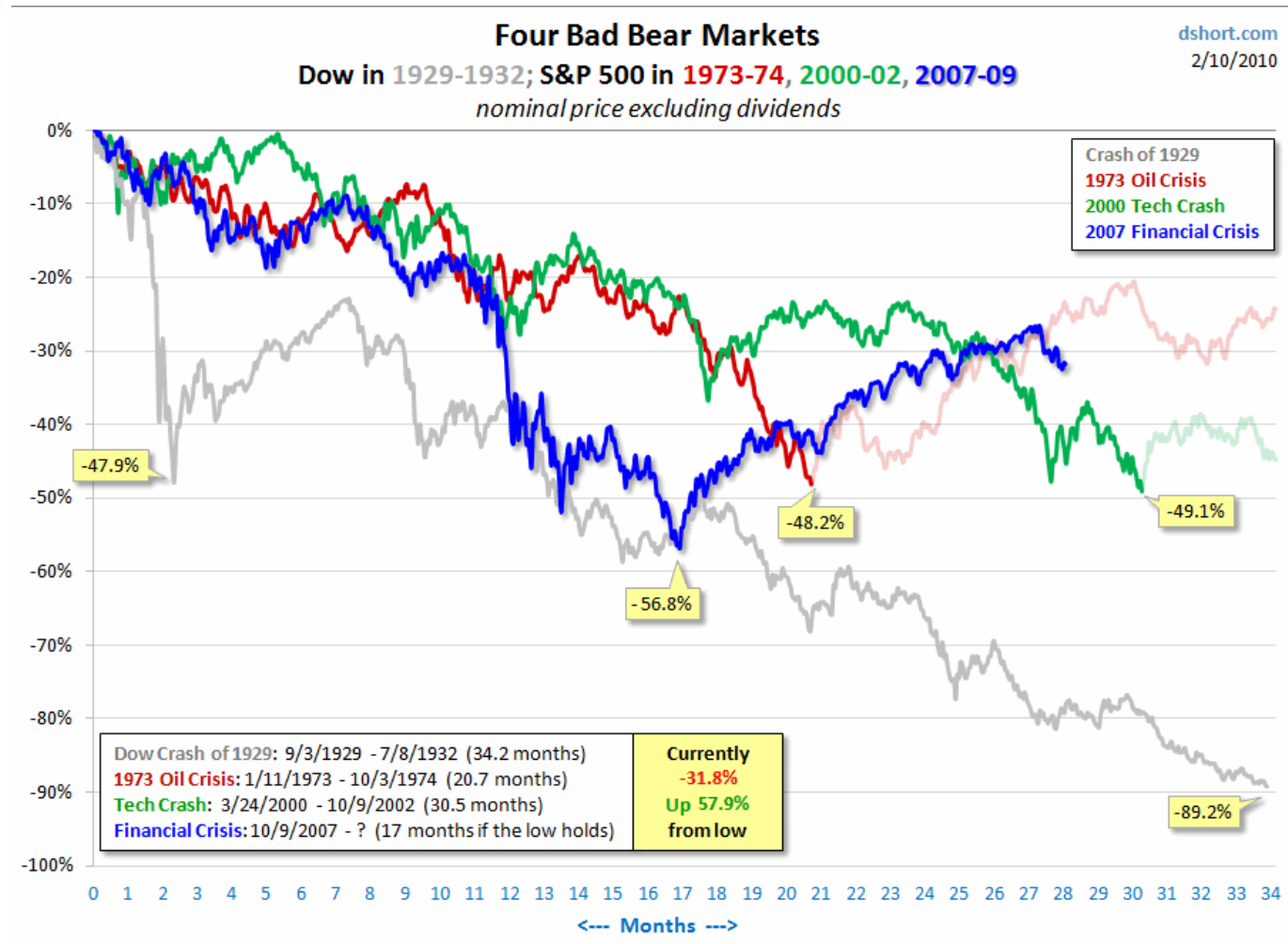
$$c_{stat} = \frac{1}{2}(1 + Gini)$$



Vizualizace - dendrogram



Vizualizace – ekonomie



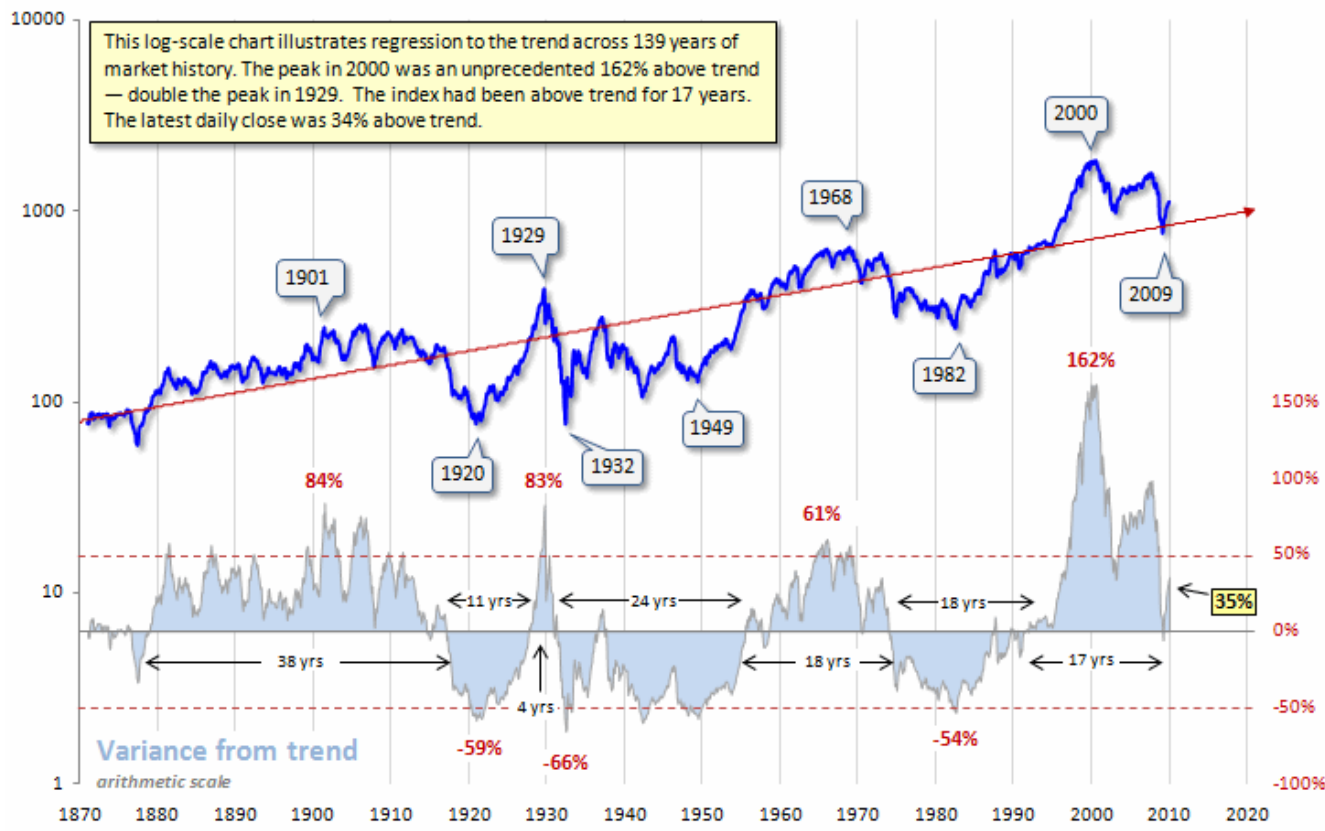
Vizualizace – ekonomie

S&P Composite Index: Regression to Trend

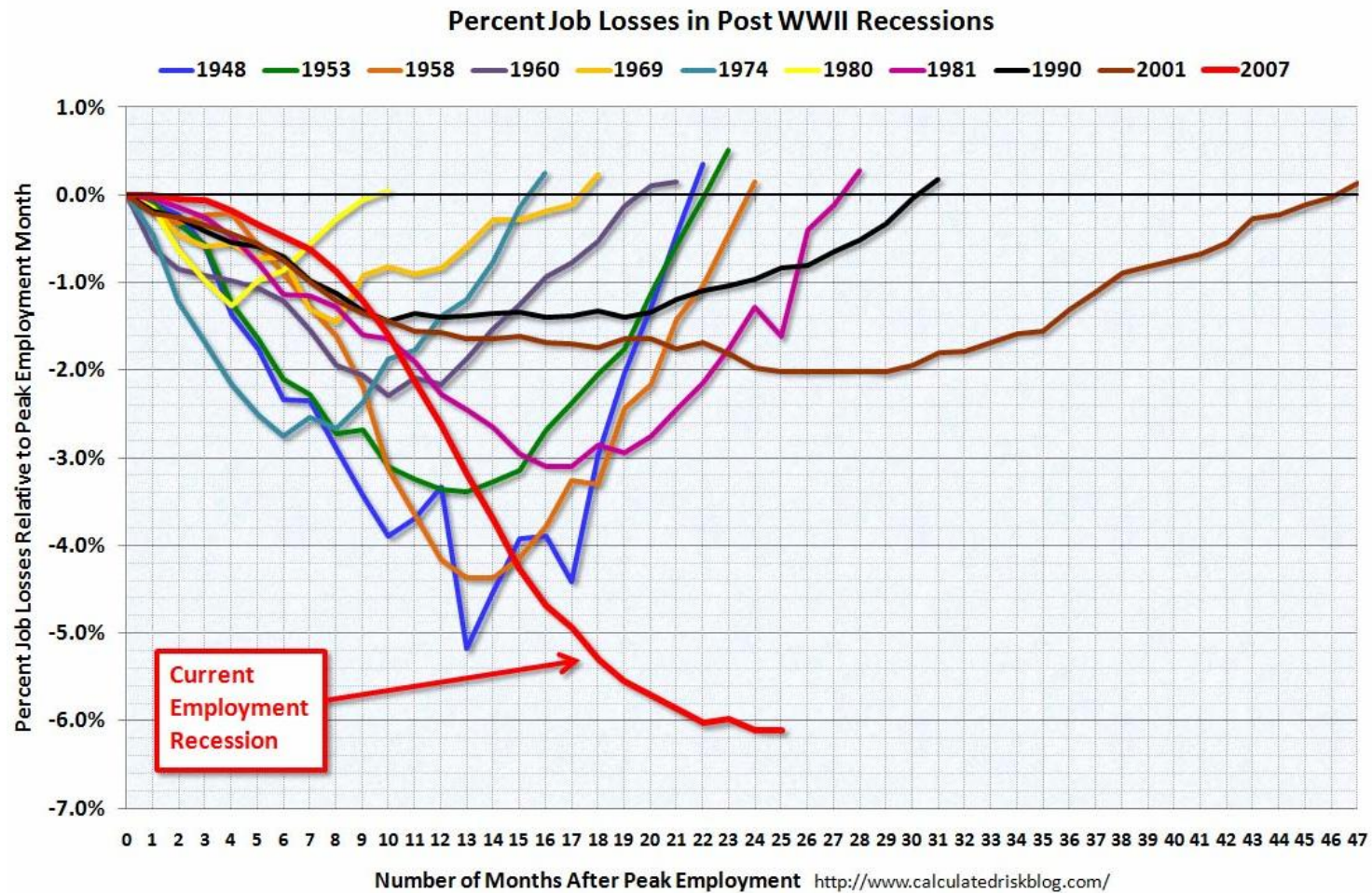
dshort.com
February 2010

Real (inflation-adjusted) Price since 1871 with Regression

Variance measured below

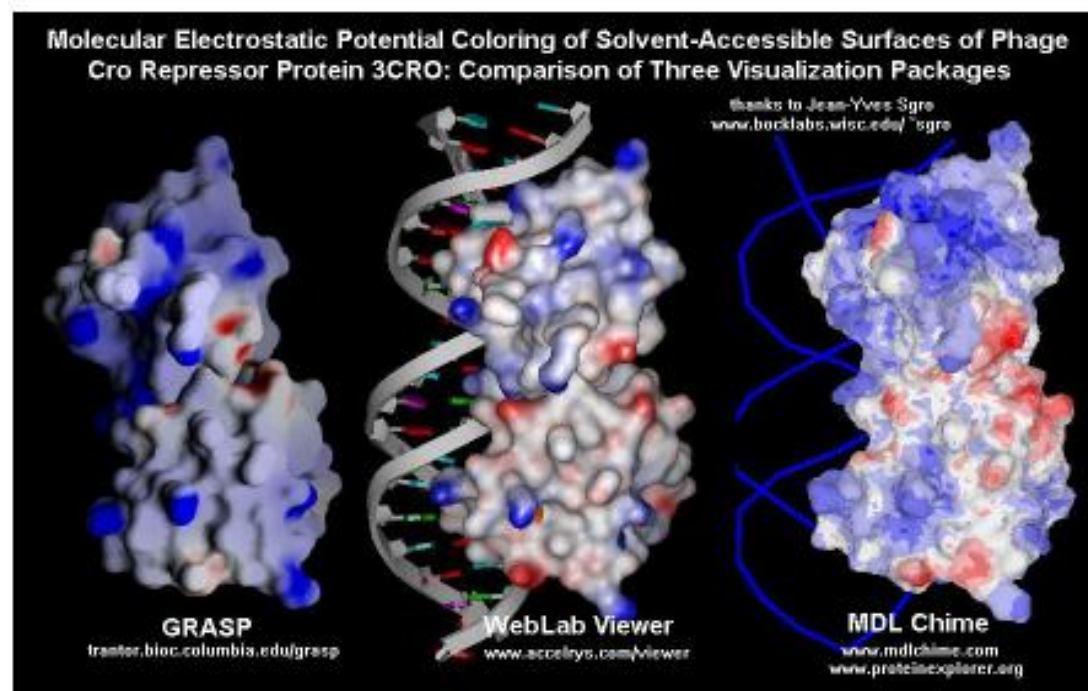


Vizualizace – ekonomie

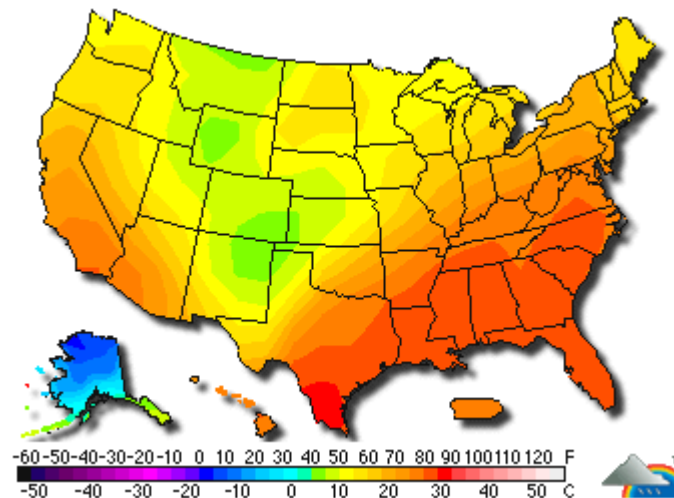
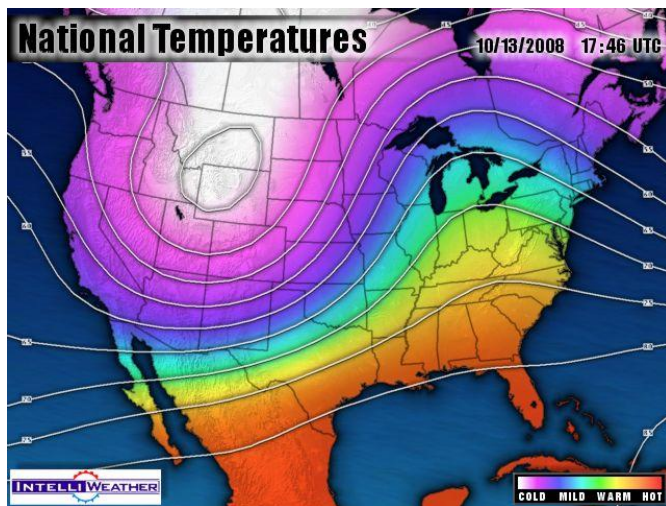
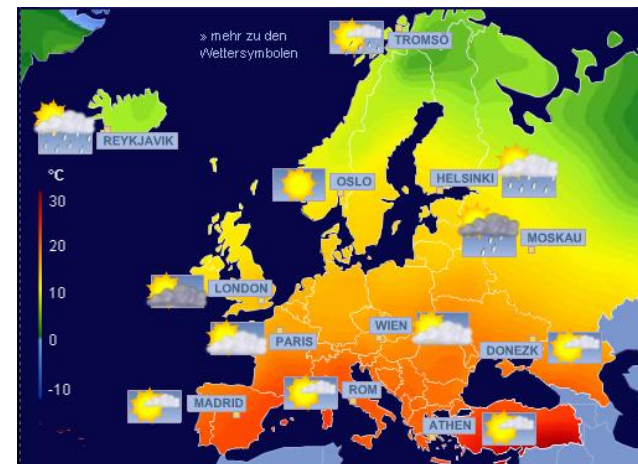


Vizualizace v biologii/chemii

- 3D zobrazení proteinu

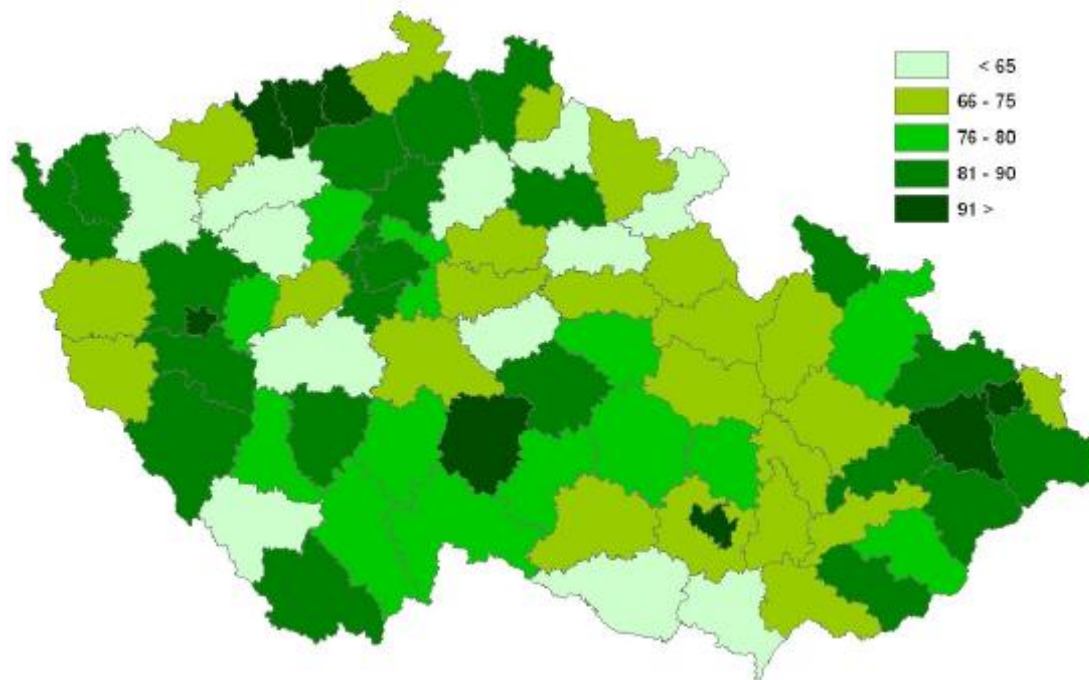


Meteo-vizualizace



Kartogram

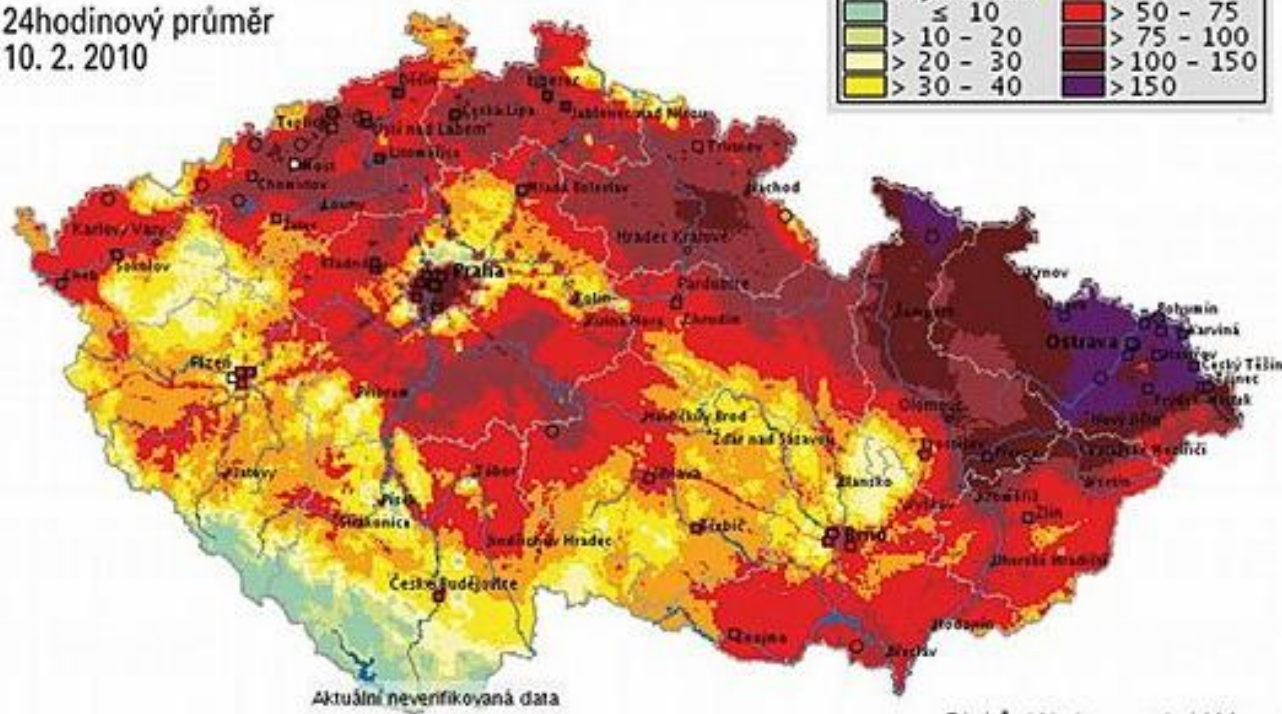
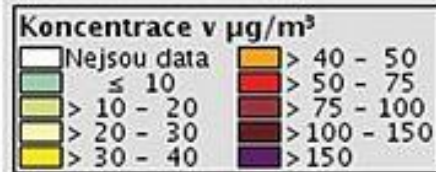
- Obce s počtem 500 a více obyvatel s vysokorychlostním připojením k internetu, podle okresů (%), k 31.12.2006



Kartogram

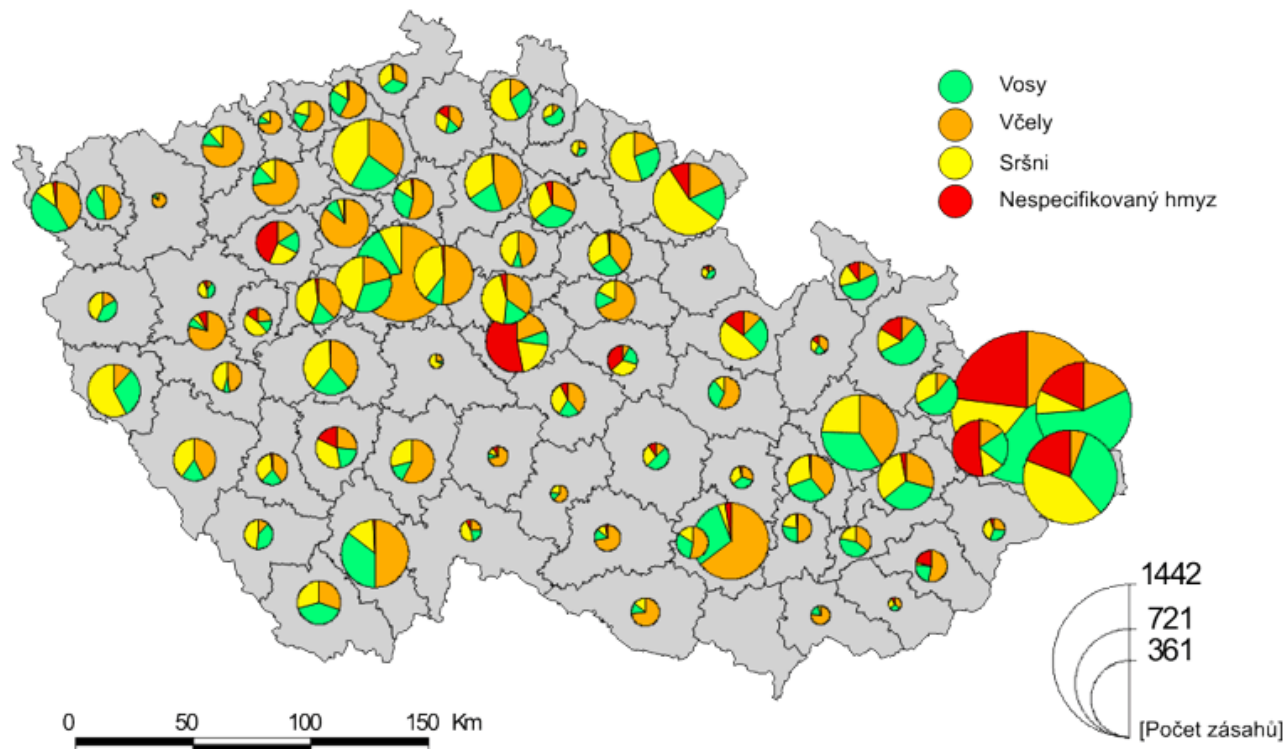
Znečištění ovzduší prachem

24hodinový průměr
10. 2. 2010

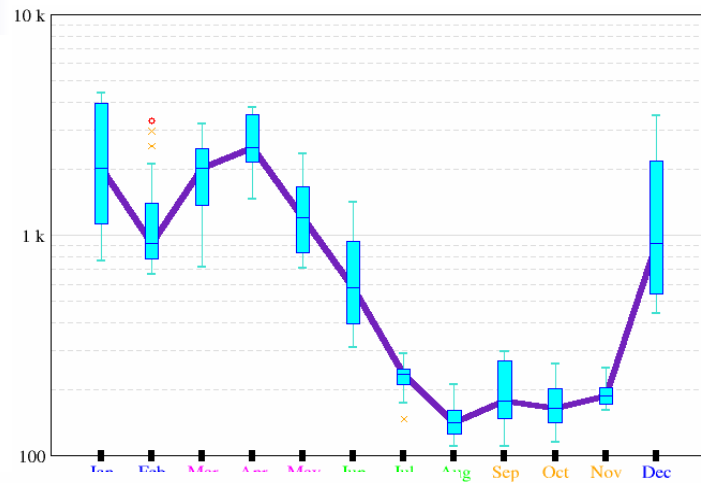


Kartodiagram

ZÁSAHY JEDNOTEK PO PROTI HMYZU v okresech České republiky v letech 1997-2000

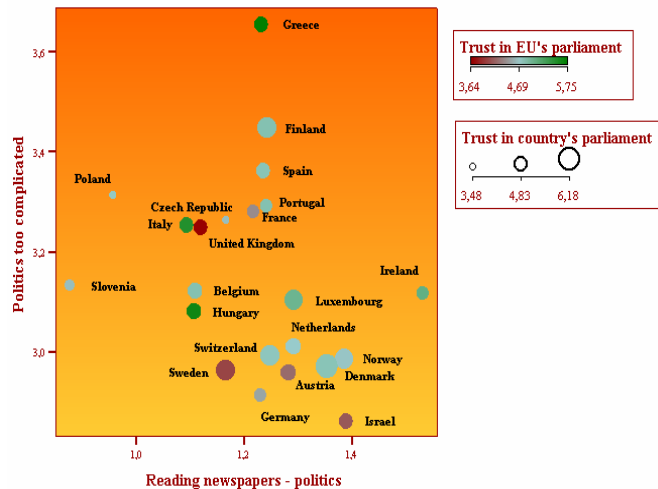
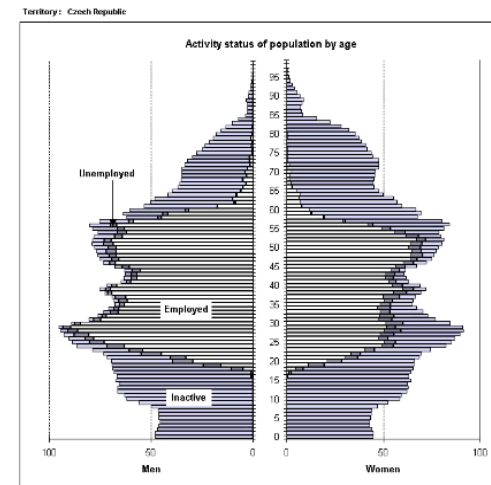


Grafy –další typy

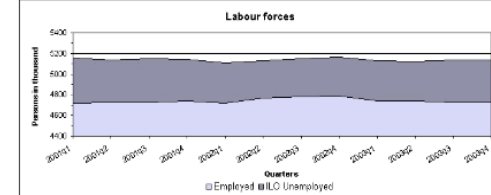
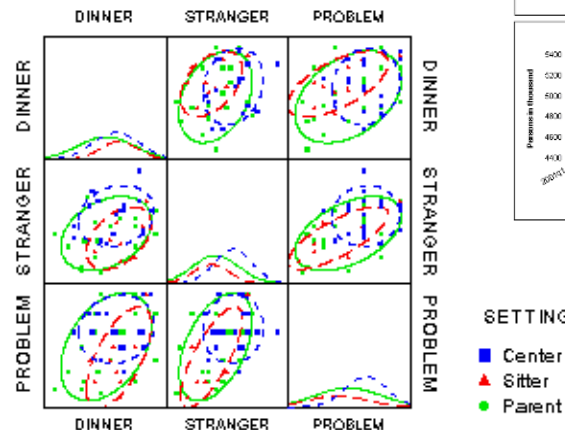


Graph: 1

ACTIVITY STATUS OF POPULATION BY AGE



Social Competence Measures Across Setting

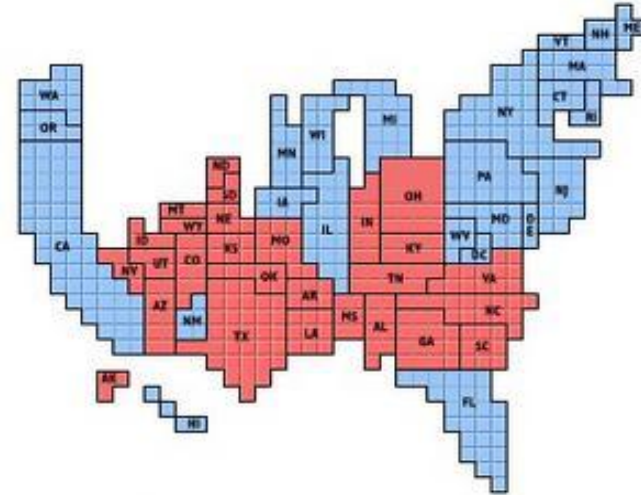
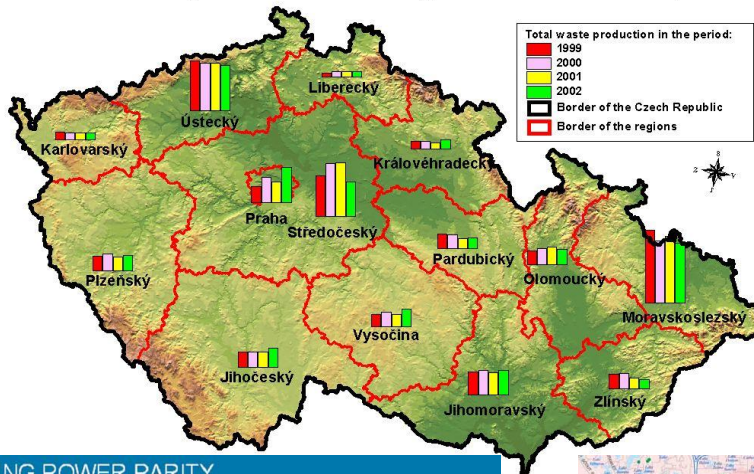


SETTING

- Center
- ▲ Sitter
- Parent

Geografická data

Overall waste production in the regions of the Czech Republic



PURCHASING POWER PARITY

