2. Předúprava vícerozměrných dat

2.1 Formy standardizace dat (čili škálování dat)

Standardizace dat odstraní závislost na jednotkách a na parametru polohy a rozptýlení.

Škálování dat znamená, že operace se týká jednotek veličin ale také počátku stupnice.

Škálování dat může být použito na znaky, na objekty nebo na obojí.

Škálování dat zahrnuje:

a) posun centra souřadného systému,

b) protažení nebo zkrácení měřítka na osách.

y_{ij} představuje pro i-tý transformovaný objekt x_{ij} čili j-tý *škálovaný znak*, který odpovídá původnímu prvku x_{ij}.

Druhy standardizace (škálování):

- **Sloupcové centrování.** Novým centrem stupnice j-tého znaku je nula. Původním centrem byl průměr prvků $\bar{x}_j$. Sloupcově centrovaná data y_{ij} vzniknou dle $y_{ij} = x_{ij} - \bar{x}_j$, kde

$$\bar{x}_j = \sum_{i=1}^{n} \frac{x_{ij}}{n} .$$

- **Sloupcová standardizace.** Prvky původních dat v j-tého znaku x_{ij} jsou děleny svou směrodatnou odchylkou tedy $y_{ij} = x_{ij} / s_j$, kde s_j je sloupcová směrodatná odchylka j-tého znaku

$$s_j = \sqrt{\frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}{n - 1}} .$$

- **Autoškálování.** Kombinace sloupcového centrování a sloupcové standardizace. Jde o studentizaci $y_{ij} = (x_{ij} - \bar{x}_j)/s_j$, která je analogická Z-transformaci pro velké výběry, kdy předpokládáme, že známe μ_j a σ_j

$$y_{ij} = (x_{ij} - \mu_j) / \sigma_j .$$

- **Škálování sloupcovým rozsahem**. Znaky jsou škálovány tak, aby minimum každého znaku bylo rovné 0 a maximum 1 dle

$$y_{ij} = \frac{x_{ij} - \min_j x_{ij}}{\max_j x_{ij} - \min_j x_{ij}} .$$

- **Řádkové centrování**. Znaky jsou škálovány dle $y_{ij} = x_{ij} - \bar{x}_i$.
- **Řádková standardizace**. Znaky jsou škálovány dle $y_{ij} = x_{ij} / s_i$.
- **Celkové centrování**. Znaky jsou škálovány dle $y_{ij} = x_{ij} - \bar{x}$, kde $\bar{x}$ je celkový průměr vyčíslený pro celou zdrojovou matici dat rozměru $n \times m$.
- **Celková standardizace**. Znaky jsou škálovány dle $y_{ij} = x_{ij} / s$, kde s je směrodatná odchylka od průměru pro všechny prvky zdrojové matice $n \times m$.
- **Dvojité centrování**. Znaky jsou škálovány nejdříve sloupcovým centrováním a následně řádkovým centrováním.
- **Řádkové profily**. Znaky jsou škálovány dle $y_{ij} = x_{ij} / (\bar{x}_i\, m)$. Součet řádku je pak 1.
- **Sloupcové profily**. Znaky jsou škálovány dle $y_{ij} = x_{ij} / (\bar{x}_j\, n)$.

PŘÍKLAD 2.1 *Grafy různých forem škálování*

Je třeba provést různé formy škálování a pak škálování zobrazit.

○ *Data:* Zdrojová matice dat rozměru $n = 15$, $m = 2$.

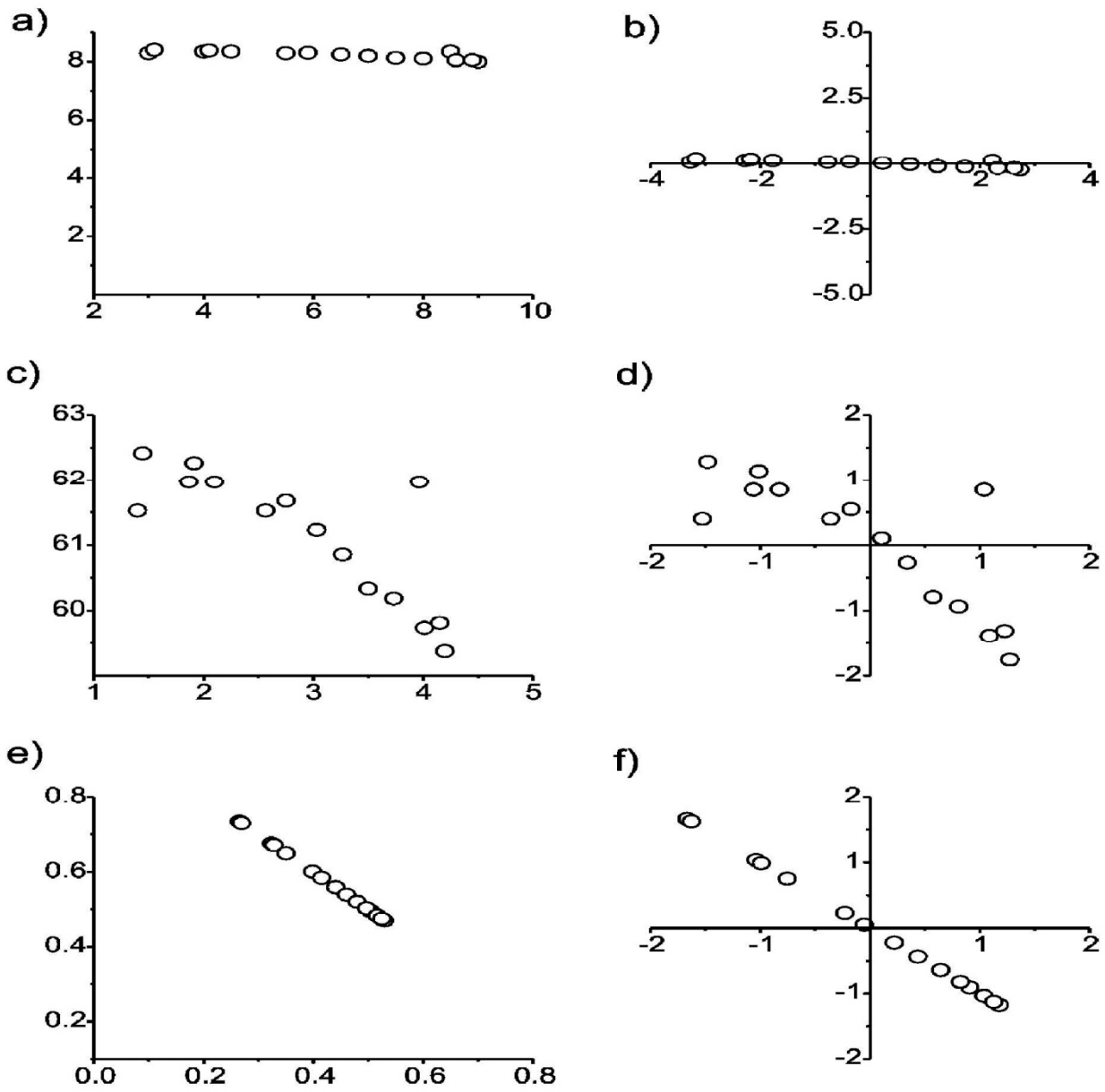
i	1	2	3	4	5	6	7	8
x_1	3.00	3.10	4.00	4.10	4.50	5.50	5.90	6.50
x_2	8.29	8.41	8.35	8.39	8.35	8.29	8.31	8.25
i	9	10	11	12	13	14	15	
x_1	7.00	8.50	7.50	8.00	8.60	9.00	8.9	
x_2	8.20	8.35	8.13	8.11	8.05	8.00	8.06	

○ *Řešení:* Původní i škálované znaky y jsou v tabulce a grafu. Rozptyl okolo počátku c^2_j je vyčíslen vztahem

$$c_j^2 = \frac{\sum_{i=1}^{n} x_{ij}^2}{n},$$

kde počet stupňů volnosti je totožný s počtem objektů n pro necentrovaná data.

U centrovaných dat je ale třeba *rozptyl* c_j^2 vynásobit zlomkem $n/(n-1)$, protože počátek byl posunut do těžiště původních dat.



Obr. 2.1 Efekt vybraných škálovacích technik: a) původní data, b) sloupcové centrování, c) sloupcové standardizování, d) autoškálování, e) profily, f) autoškálované profily.

Výklad obrázku:

1. Sloupcové centrování (když je $m < n$), sloupcová standardizace a autoškálování nezmění dimenzi dat.

2. Profily a řádkové centrování (když $m < n$) sníží dimenzi o 1, takže objekty padnou na hyperpovrch ve vícerozměrném hyperprostoru znaků, hypersféru v případě čtverce řádkových profilů.

3. Doporučuje se užít následně dvě nebo více škálovacích technik, např. sloupcové centrování je následované sloupcovou standardizací nebo dvojité centrování se skládá z řádkového centrování následovaného sloupcovým centrováním. Pořadí obou operací zde není nikterak rozhodující.

4. V ostatních případech např. profilů je pořadí škálovaní velmi důležité. Když se změní počátek způsobem centrování, budou profily zcela odlišné.

5. Škálování je třeba užívat velmi obezřetně a přihlížet k významu analyzovaných dat.

6. Mají-li znaky stejný charakter (např. absorbanční matice), škálování není nutné.

Transformace:

Kromě škálování se také často používá:

1. ***Logaritmická transformace*** (eliminace pozitivního zešikmení dat),

2. ***Transformace pořadová***, kdy se data nahradí svým vzestupným pořadím. Pořadová transformace je přirozeně robustní, ale za cenu ztráty informace. Výsledky vícerozměrných statistických metod pak mohou být značně odlišné.

2.2 Užití statistických vah

Škálování eliminuje nestejný pořádek a měřítko u znaků a tvoří znaky stejné důležitosti.

Použitím statistických vah lze však zvýšit důležitost některých znaků.

Užití vah je potřebné v případech:

a) Existují-li rozličné nejistoty v měřených znacích.

b) Pokud již máme zkušenosti o důležitosti znaků.

c) Jestliže existují rozličné důležitosti znaků dle účelu analýzy dat.

Fisherovy váhy jsou klasifikační váhy stejné důležitosti a významu jako Fisherovy poměry rozptylu mezi kategoriemi a rozptylu uvnitř kategorií.
V případě dvou kategorií s maticí dat $\boldsymbol{X}$ rozměru $n \times m$, která byla rozdělena do dvou submatic (každá je v jedné kategorii):

$$\boldsymbol{X}\ (n \times m) = \begin{pmatrix} \boldsymbol{X}_1\ (n_1 \times m) \\ \boldsymbol{X}_2\ (n_2 \times m) \end{pmatrix} ,$$

je **odhad rozptylu *j*-tého znaku uvnitř kategorie *g*** dán vztahem

$$s_{g,j}^2 = \frac{\sum_{g=1}^{n_g} (x_{i_g,j} - \bar{x}_{g,j})^2}{n_g - 1} \qquad (2.7)$$

a **rozptyl okolo těžiště mezi kategoriemi 1 a 2** je určen odhadem

$$s_{c,j}^2 = (x_{1j} - \bar{x}_j)^2 + (x_{2j} - \bar{x}_j)^2 . \qquad (2.8)$$

Fisherova váha znaku je dána vztahem

$$FW_j = \frac{4\ s_{c,j}^2}{s_{1,j}^2 + s_{2,j}^2} . \qquad (2.9)$$

V případě více než dvou kategorií bude FW tvořena průměrem vah vyčíslených pro každý možný pár kategorií.

Rozptylové váhy jsou v případě dvou kategorií vyčísleny dle vztahu

$$RW_j = \frac{1 + FW_j}{2} . \qquad (2.10)$$

Pro více než dvě kategorie je RW geometrickým průměrem párových rozptylových vah.

Modifikované váhy byly zavedeny ke zvýšení důležitosti znaků blízko průměrů ve dvou kategoriích, které se ale liší rozptylem.

Všechny tyto váhové procedury jsou zatíženy problémem, že jsou jednorozměrné, takže se může stát, že dva znaky poskytnou velmi malý váhový koeficient, i když jsou oba enormě veliké důležitosti.

Naopak, veliká váha může být přiřazena dvěma znakům, které jsou si téměř úměrné, a proto obsahují téměř stejnou informaci.

2.3 Průzkumová analýza vícerozměrných dat

2.3.1 Zobrazení vícerozměrných dat

EDA zobrazení umožňuje:

a) identifikovat vektory $\boldsymbol{x}_i$ nebo jejich složek, které se jeví jako vybočující,

b) indikovat struktury v datech jako jsou shluky, ukazující na heterogenitu výběru nebo přítomnost dílčích výběrů s odlišným chováním.

Zobrazení vícerozměrných dat se dá zařadit:

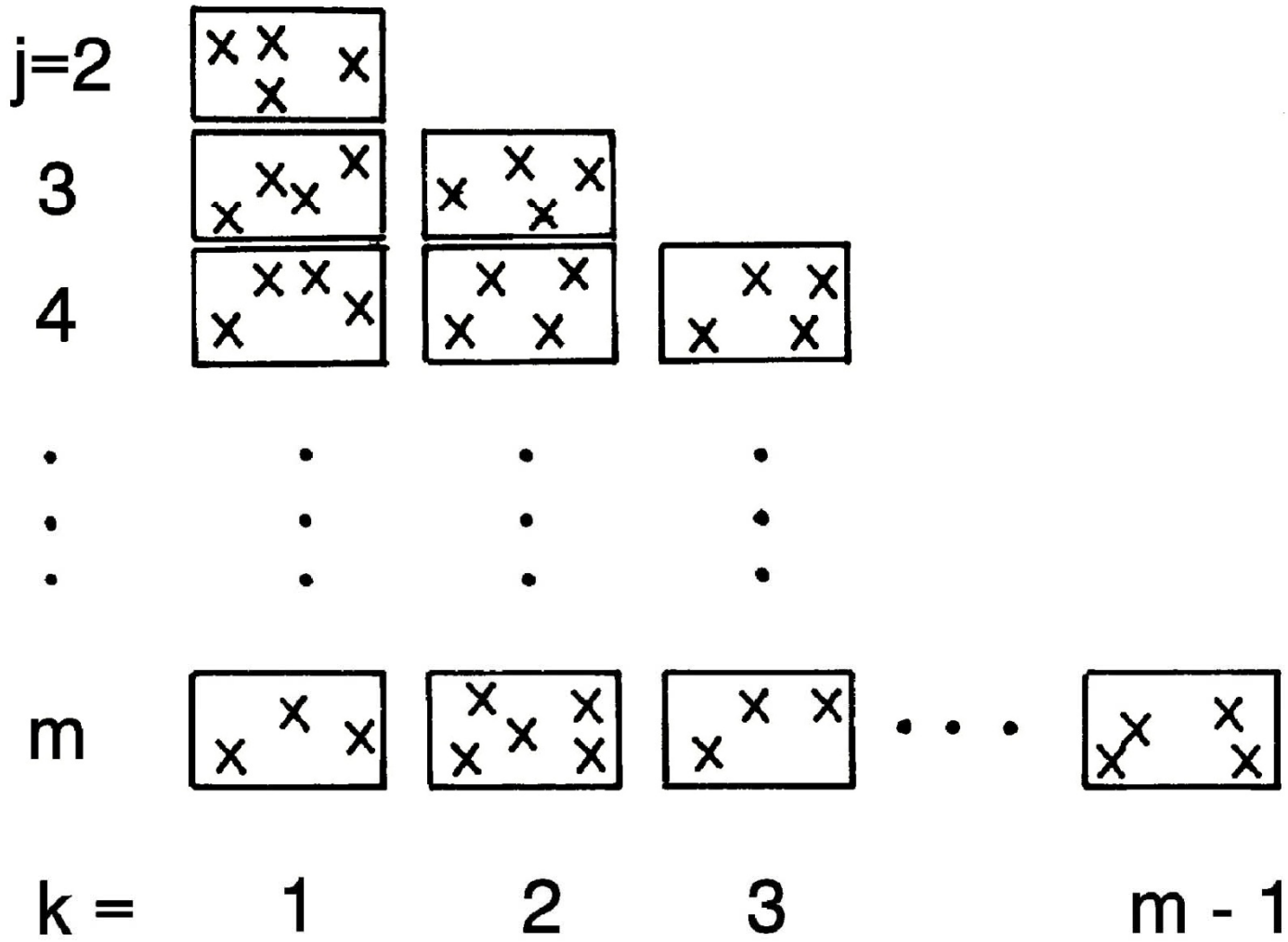
1. **Zobecněné rozptylové diagramy,**
2. **Symbolové grafy.**

a) Pro případ dvojice náhodných znaků ($m = 2$) lze konstruovat rozptylové grafy.

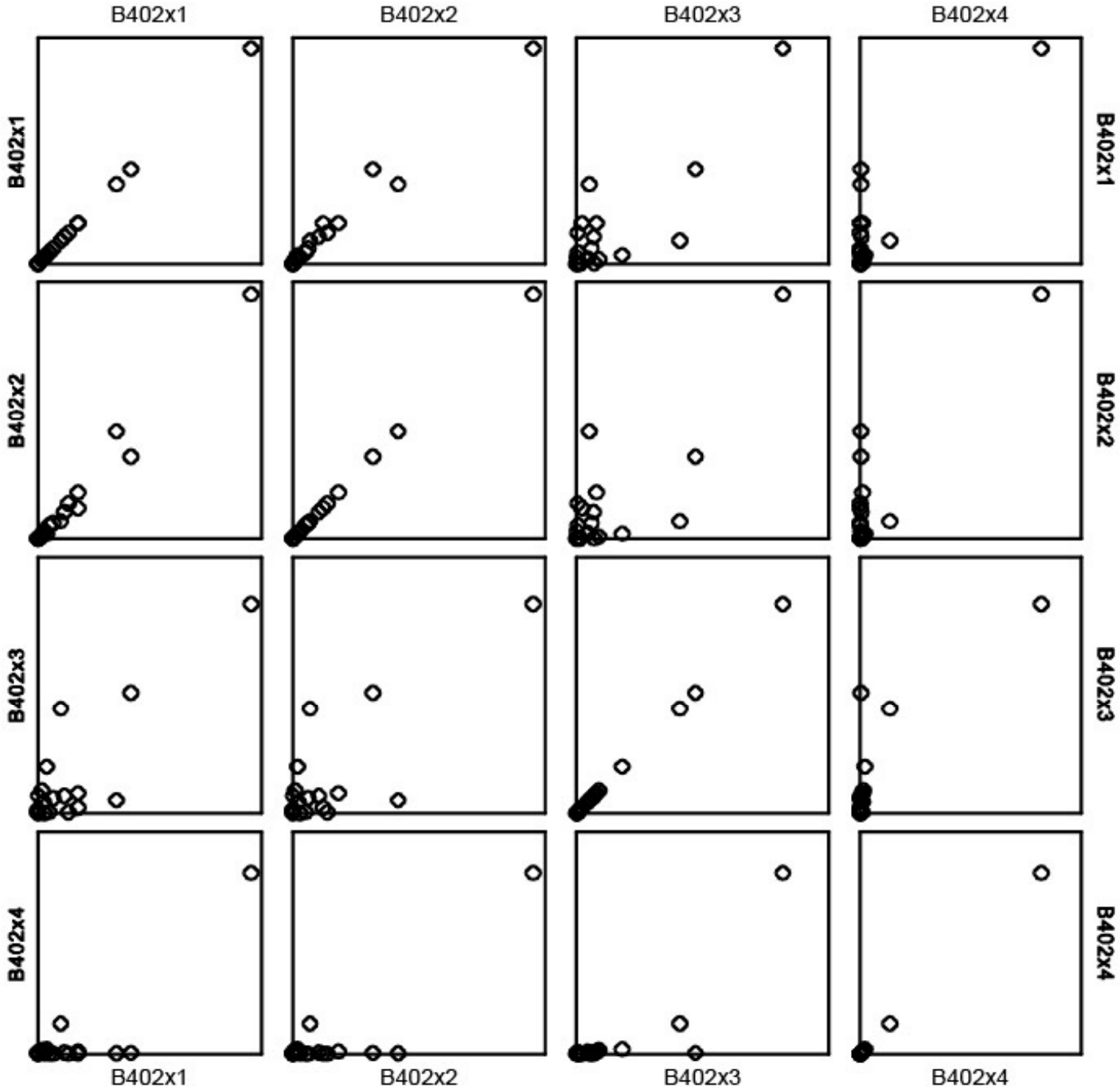
b) Problémy jsou u vícerozměrných dat pro $m > 2$: je třeba buď volit několik grafů nebo vhodně provést transformace na dvoudimenzionální data.

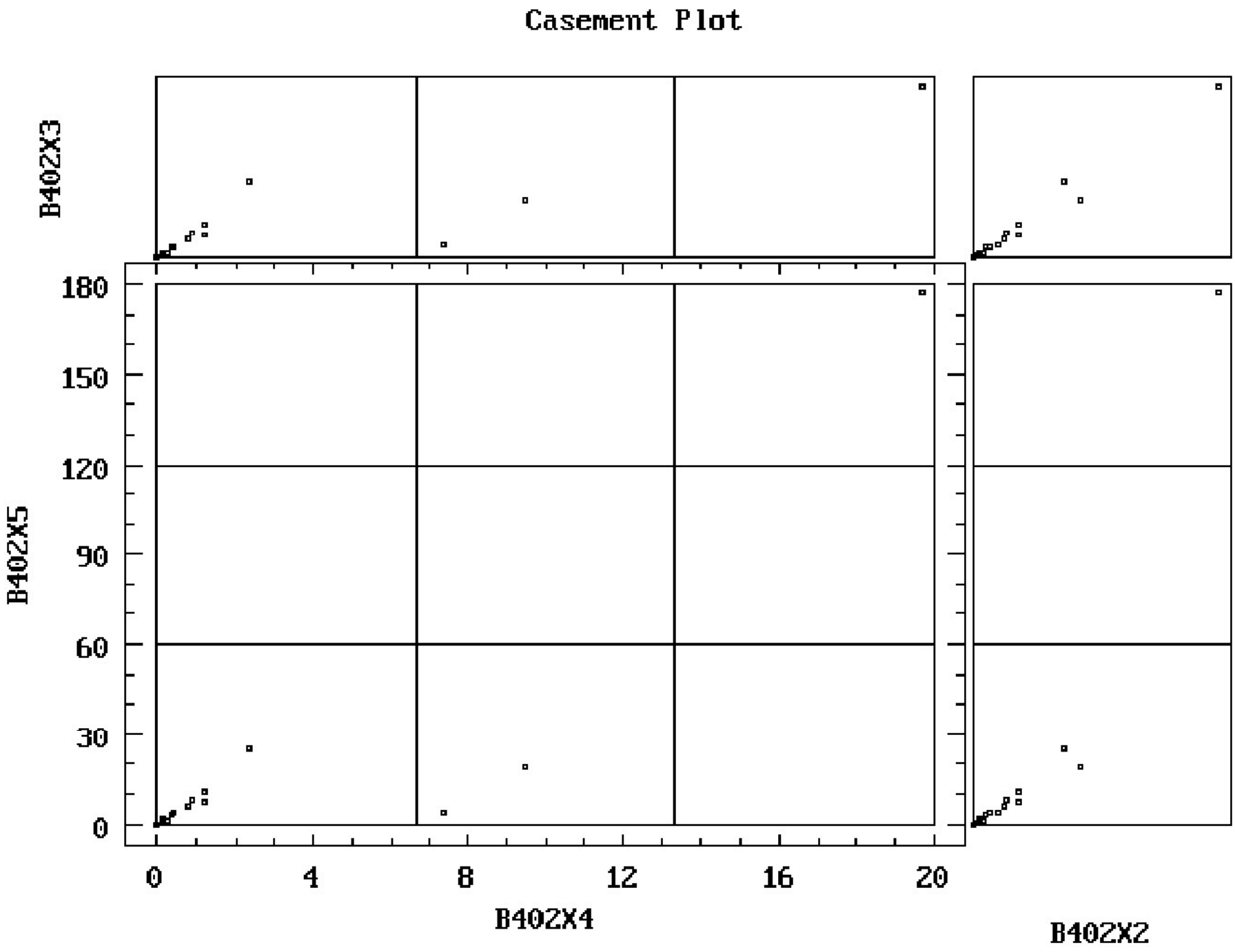
1. Zobecněné rozptylové grafy

Pro dva znaky x_{i1} a x_{i2} představuje rozptylový diagram závislost mezi znakem x_{i1} na ose x a znakem x_{i2} na ose y (obr. 2.2): lze indikovat vybočující hodnoty, shluky v datech a míru párové závislosti mezi těmito znaky.



Obr. 2.2 Schematické znázornění párových rozptylových diagramů.


B402x1
B402x2
B402x3
B402x4
B402x1
B402x2
B402x3
B402x4
B402x1
B402x2
B402x3
B402x4


Casement Plot
B402X3
B402X5
180
150
120
90
60
30
0
0
4
8
12
16
20
B402X4
B402X2

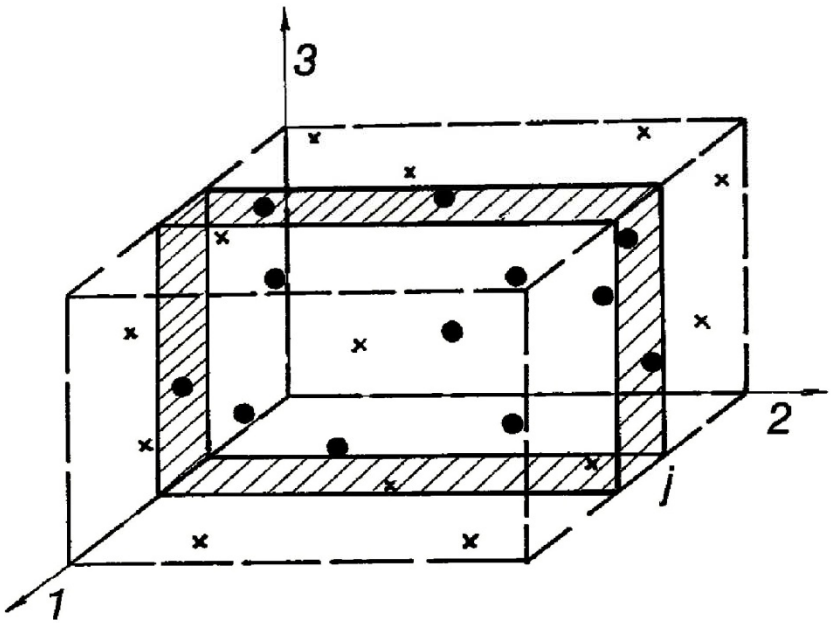
Pro případ m-rozměrných dat je nejjednodušší konstruovat rozptylové diagramy pro všechny dvojice znaků x_{ij}, x_{ik}. Nejvhodnější je uspořádání diagramů do pole $(m - 1)(m - 1)$.

V tomto poli je **(j, k)-tý rozptylový diagram závislosti** složky $x_{i,j+1}$ na x_{ik}.

Vzhledem k symetrii postačuje znázornění pouze $(m - 1)\, m/2$ grafů. S růstem m roste počet grafů, a to úměrně s m^2.

Pro větší m větším než 10 je již použití rozptylového diagramu problematické.

Pro případ tří znaků ($m = 3$) je možné rozdělit celou n-tici bodů na několik skupin s ohledem na hodnoty jednoho znaku, a pak pro každou skupinu konstruovat rozptylový diagram zvaný jako **okénkový graf** .



Obr. 2.3 Konstrukce okénkového grafu, kde kolečka značí body, které leží ve vyšrafovaném j-tém intervalu proměnné 3.

2. Symbolové grafy

Jednotlivé znaky jsou „kódovány“ s ohledem na jejich hodnoty do geometrických symbolů.

Každému objektu $\boldsymbol{x}_i$ pak odpovídá jistý obrazec složený z těchto symbolů (=znaků) a tak lze v jednom grafu rozlišit více znaků x_j, $j = 1, \ldots, m$.

Základní typy symbolů:

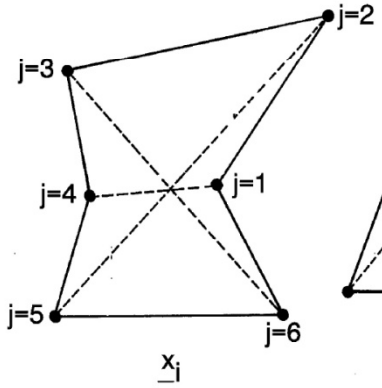
profily,

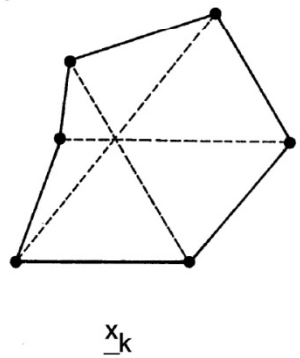
polygony,

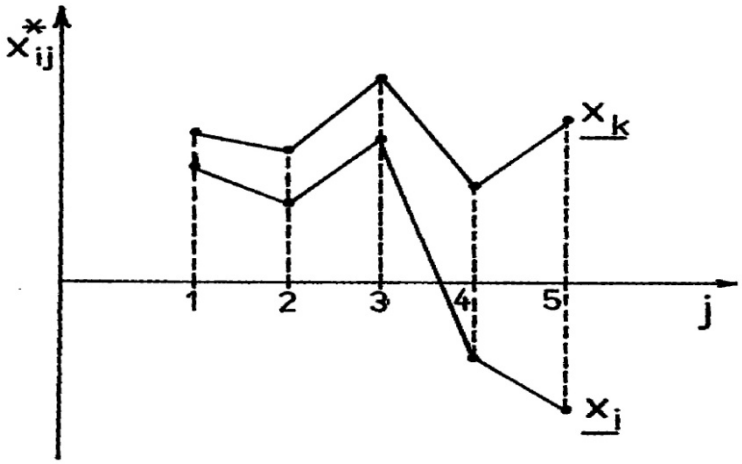
tváře,

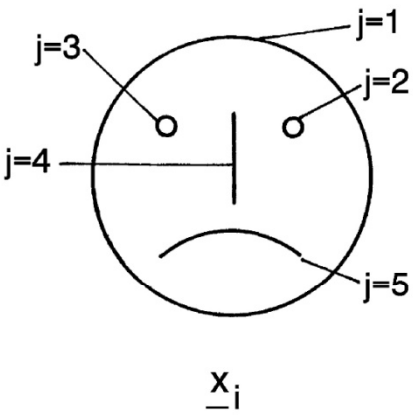
křivky,

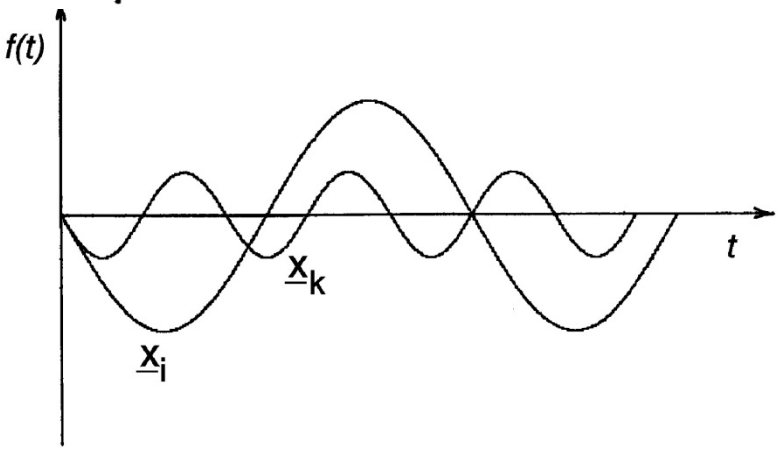
stromy.







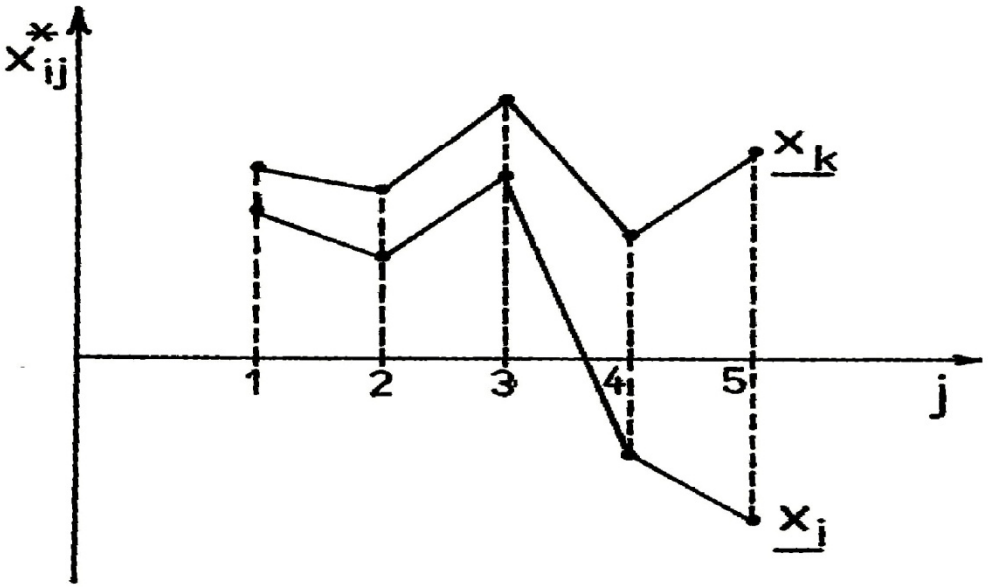




- **Profily.** Představují jednoduché dvourozměrné zobrazení m-rozměrných dat:
  - každý bod $\boldsymbol{x}_i$ je charakterizován m vertikálními úsečkami nebo sloupci,
  - jejich velikost je úměrná hodnotě odpovídající složky x_{ij}, $j = 1, ..., m$,
  - na osu x se vynáší index dané složky j.
  - profil vzniká spojením koncových bodů těchto úseček či sloupců.
  - je vhodné použít škálované znaky

$$x_{ij}^{*} = \frac{x_{ij}}{\max\limits_{i} |x_{ij}|} \quad , (2.11) ,$$

kde $\max |x_{ij}|$ je maximální hodnota absolutní velikosti složky x_j vektoru $\boldsymbol{x}$ přes všechny body, $i = 1, ..., n$.



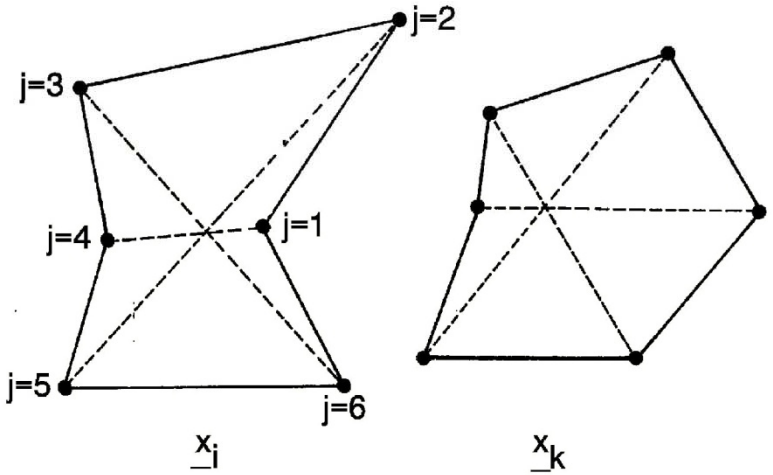
Obr. 2.4 Znázornění profilů pro dva objekty $\boldsymbol{x}_i$, $\boldsymbol{x}_k$, kde $m = 5$.

● **Polygony.** Jsou profily v polárních souřadnicích: kde každý znak x_{ij} odpovídá délce paprsku vycházejícího z jednoho středu.

- paprsky jsou rozmístěny ekvidistantně (ve stejných vzdálenostech) na kružnici.
- délka j-tého paprsku x_{ij} musí být kladná.
- lineární transformace do intervalu $[a, 1]$, kde a je zvolená spodní mez, $a = 0$ dle

$$x_{ij}^{*} = \frac{(1 - a)(x_{ij} - \min_i x_{ij})}{\max_i x_{ij} - \min_i x_{ij}} + a \,, \quad (2.12) \,, \quad \alpha_j = \frac{2\pi(j - 1)}{m} \,, \qquad j = 1, \ldots, m \,. \quad (2.13)$$

kde min x_{ij} je minimální a max x_{ij} maximální hodnota j-tého znaku přes všechny objekty $\boldsymbol{x}_i$, $i = 1, \ldots, n$. K určení směrů jednotlivých paprsků lze definovat jejich úhel α_j.



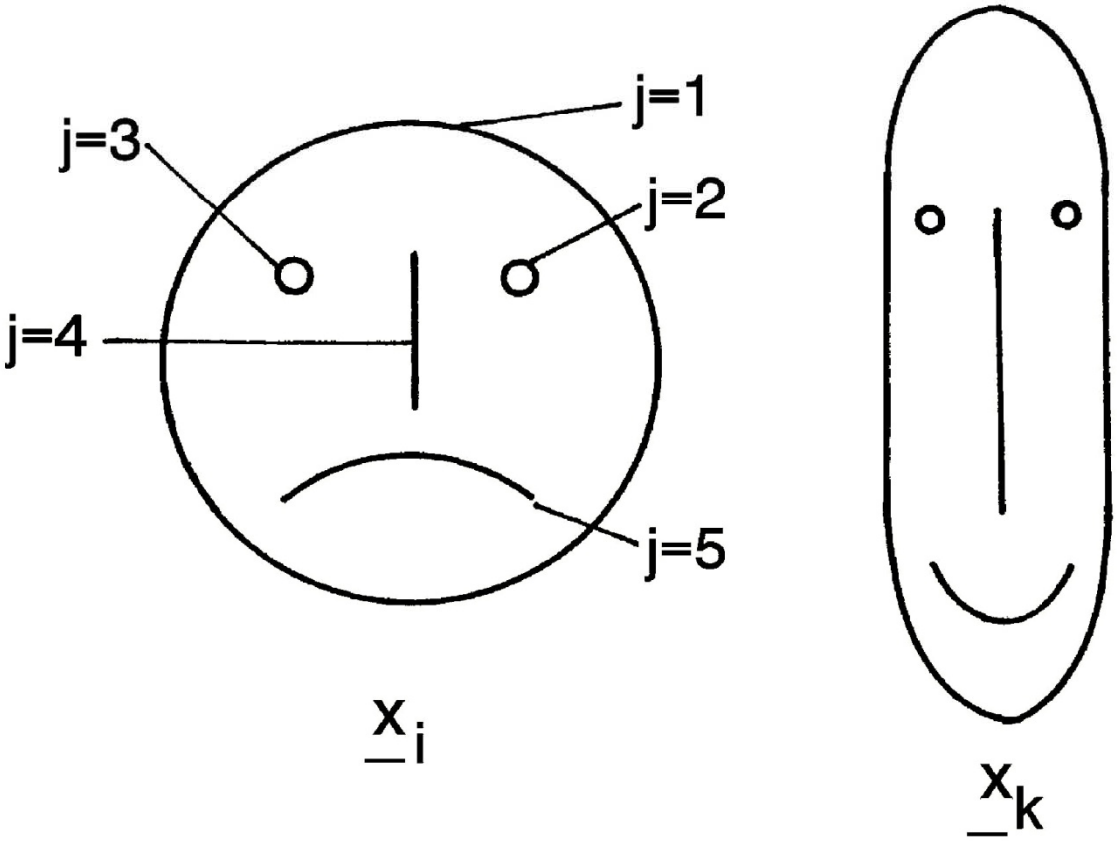
Obr. 2.5 Znázornění polygonu pro dva body $\boldsymbol{x}_i$, $\boldsymbol{x}_k$, kdy $m = 6$.

Za střed paprsků se volí počátek. Maximální délka paprsků rovna R (obyčejně $R = 1$) a polygon pro bod $\boldsymbol{x}_i$ spojnicí m bodů p_{ij} o souřadnicích $p_{ij} = (x_{ij} R \cos \alpha_j, \; x_{ij} R \sin \alpha_j)$.

- **Černoffovy tváře.** Charakterizují každý znak x_{ij} nějakým kódem schematizované tváře.

Mezi kódy (=znaky) patří

- tvar tváře,
- délka nosu,
- velikost očí,
- tvar úst,
- apod.

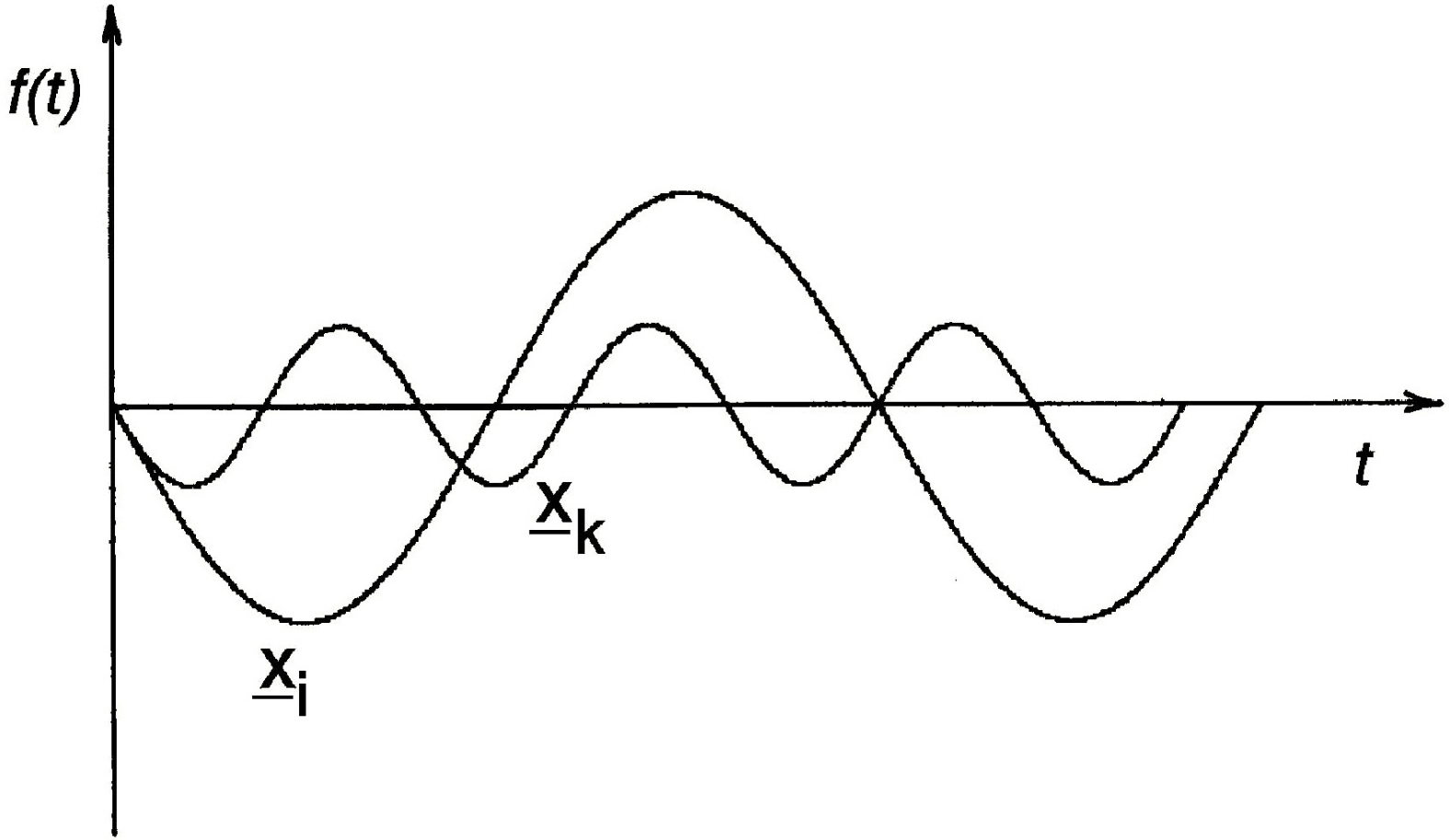


Obr. 2.6 Znázornění tváří pro dva body $\boldsymbol{x}_i$, $\boldsymbol{x}_k$, kdy $m = 5$.

● **Křivky.** Transformují každý objekt $\boldsymbol{x}_i$ na spojitou křivku, která je lineární kombinací všech jeho znaků. Andrews vyjádřil křivku f_i objektu $\boldsymbol{x}_i$ konečnou Fourierovu řadou

$$f_{x_i}(t) = f_i = \frac{x_{i1}}{\sqrt{2}} + x_{i2}\sin(t) + x_{i3}\cos(t) + \tag{2.15}$$

$$+ x_{i4}\sin(2\,t) + x_{i5}\cos(2\,t) + \ldots\ .$$



Obr. 2.7 Schematické znázornění křivek pro body $\boldsymbol{x}_i$, $\boldsymbol{x}_k$.

- **Stromy.** Jsou vhodné pro případy velkého počtu znaků m.

- Jednotlivé znaky x_j představují délku větví schematického stromu.

- Jeho struktura čili rozmístění větví se volí na základě předběžného hierarchického shlukování znaků (shlukové analýzy).

- Předběžná shluková analýza se dá použít také při výběru pořadí složek vektoru $\boldsymbol{x}$ při konstrukci ostatních symbolových grafů.

2.3.2 Ověření normality

U vícerozměrných náhodných výběrů hraje hlavní roli předpoklad, že data pocházejí z *vícerozměrného normálního rozdělení*.

Tento předpoklad usnadňuje zejména statistickou analýzu vektoru středních hodnot nebo kovarianční matice.

Testování vícerozměrné normality je poměrně kompikovaná úloha.

Přitom předpoklad normality je základem testů souvisejících se střední hodnotou (Hotellingův T^2-test) respektive kovarianční maticí.

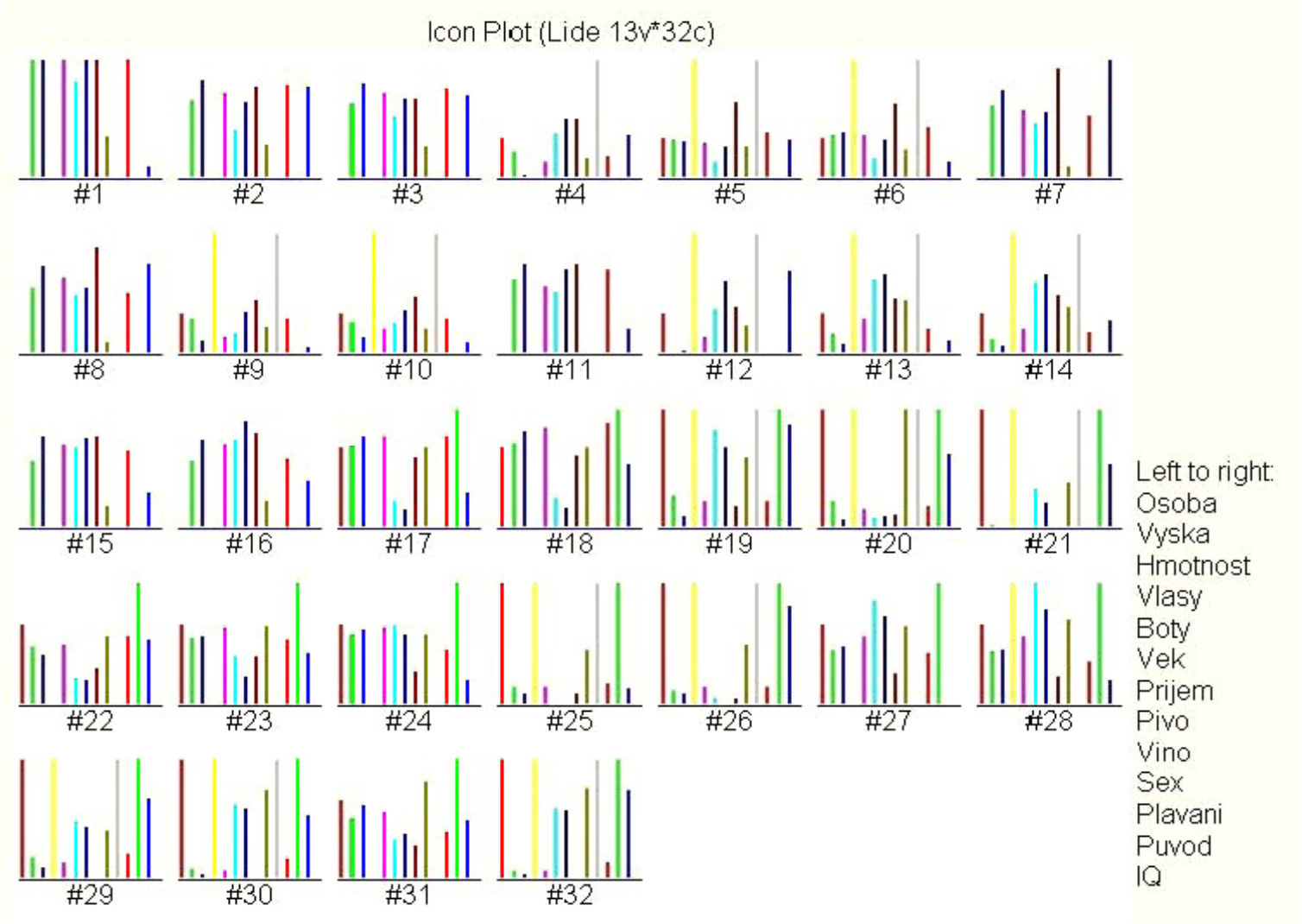
PŘÍKLAD 2.2 *Průzkumová analýza zdrojové matice dat demografického souboru Lidé*

Vyšetřete, které ze 12 znaků demografického souboru dat Lidé jsou nejvýhodnější k charakterizaci osob a které znaky mají největší míru rozptýlení. Matice obsahuje data pro n = 32 osob a m = 12 znaků, kde 16 osob bylo vybráno ze Skandinávie (kód A) a 16 osob ze Středomoří kód B), 16 osob jsou muži (kód M) a 16 osob jsou ženy (kód F).

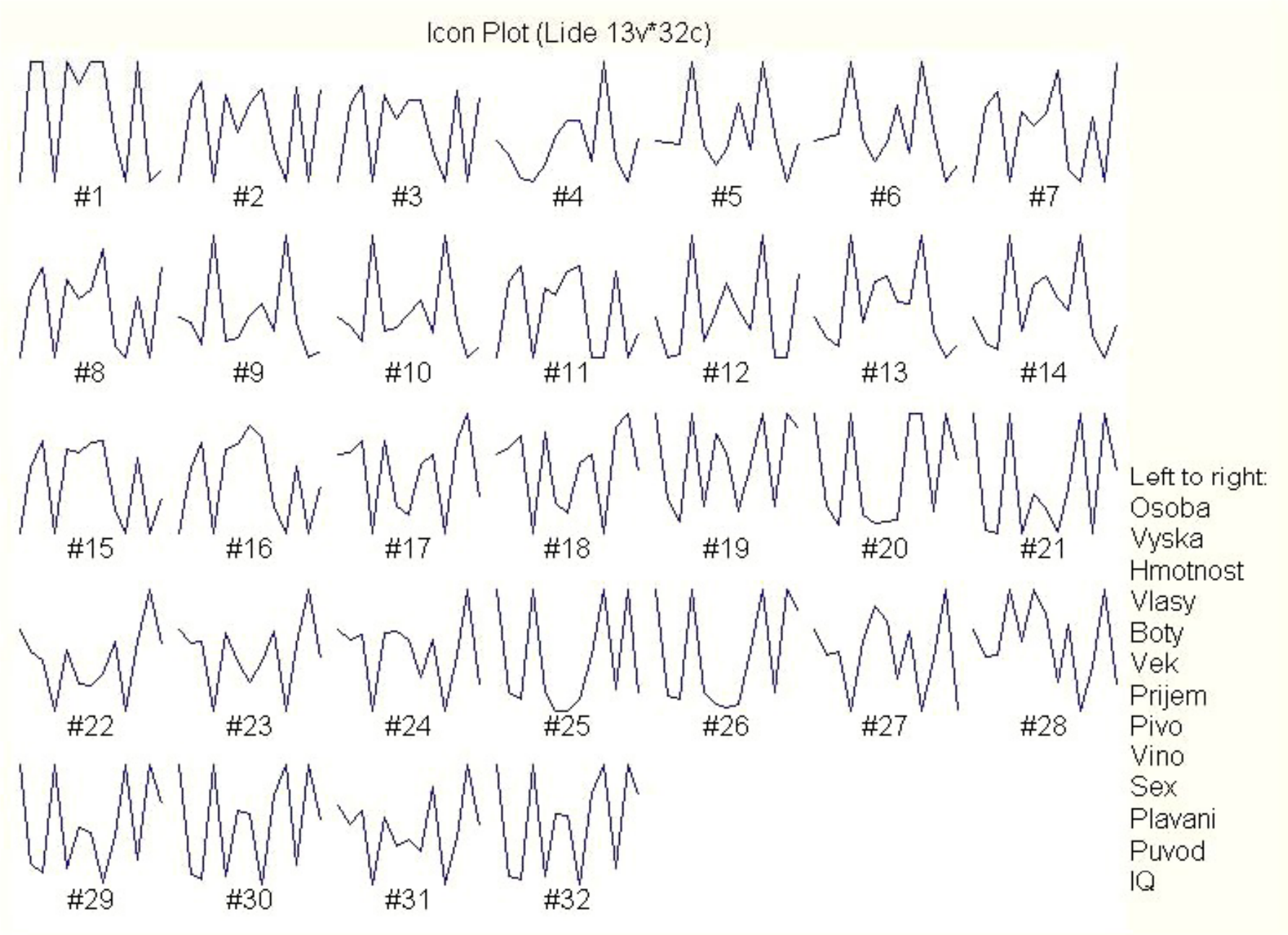
○ *Data:* Znaky obsahují u každé osoby ***výšku*** [cm], ***hmotnost*** [kg], délku ***vlasů*** [krátká: -1, dlouhá: +1], velikost ***boty*** [Evropský standard], ***věk*** [roky], ***příjem*** [Euro], spotřeba ***piva*** [litry na rok], spotřeba ***vína*** [litry na rok], ***pohlaví*** [muž: -1, žena: +1], schopnost ***plavat*** [naměřený čas na uplavání 500 m], ***původ*** [A: -1 Skandinávie, B: +1 Středomoří], inteligenční kvocient ***IQ*** [Evropský standardizovaný IQ test]. Mezi znaky jsou tři dichotomické, binární proměnné, a to pohlaví, délka vlasů a původ a ostatních 9 znaků nabývá kvantitativních hodnot.

Osoba	*Vyska*	*Hmotnost*	*Vlasy*	*Boty*	*Vek*	*Prijem*	*Pivo*	*Vino*	*Sex*	*Plavani*	*Puvod*	*IQ*
MA	198	92	-1	48	48	45000	420	115	-1	98	-1	100
...	...	...	...	...	...	...	...	...	...	...	...	...
FB	160	48	1	35	40	31000	118	198	1	74	1	129

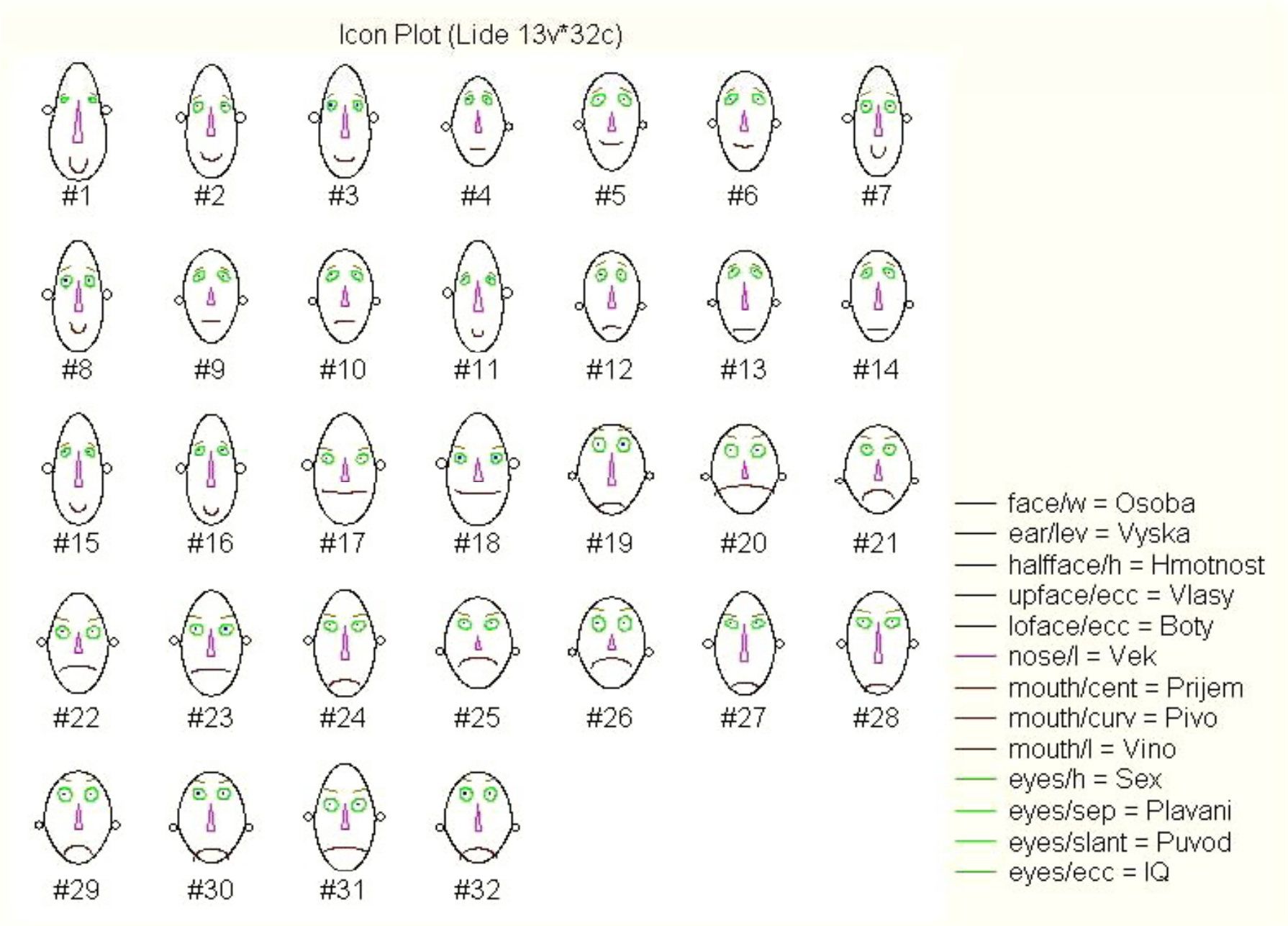
○ *Řešení:* užity STATISTICA . jsou hledány podobné objekty (zde osoby značené svými pořadovými indexy), které jsou zobrazeny podobným grafickým útvarem. Mezi porovnávané útvary patří profilové sloupce, **profilové křivky**, **Černoffovy obličeje**, **profily sluníčka** a **hvězdičky** z polygonů.



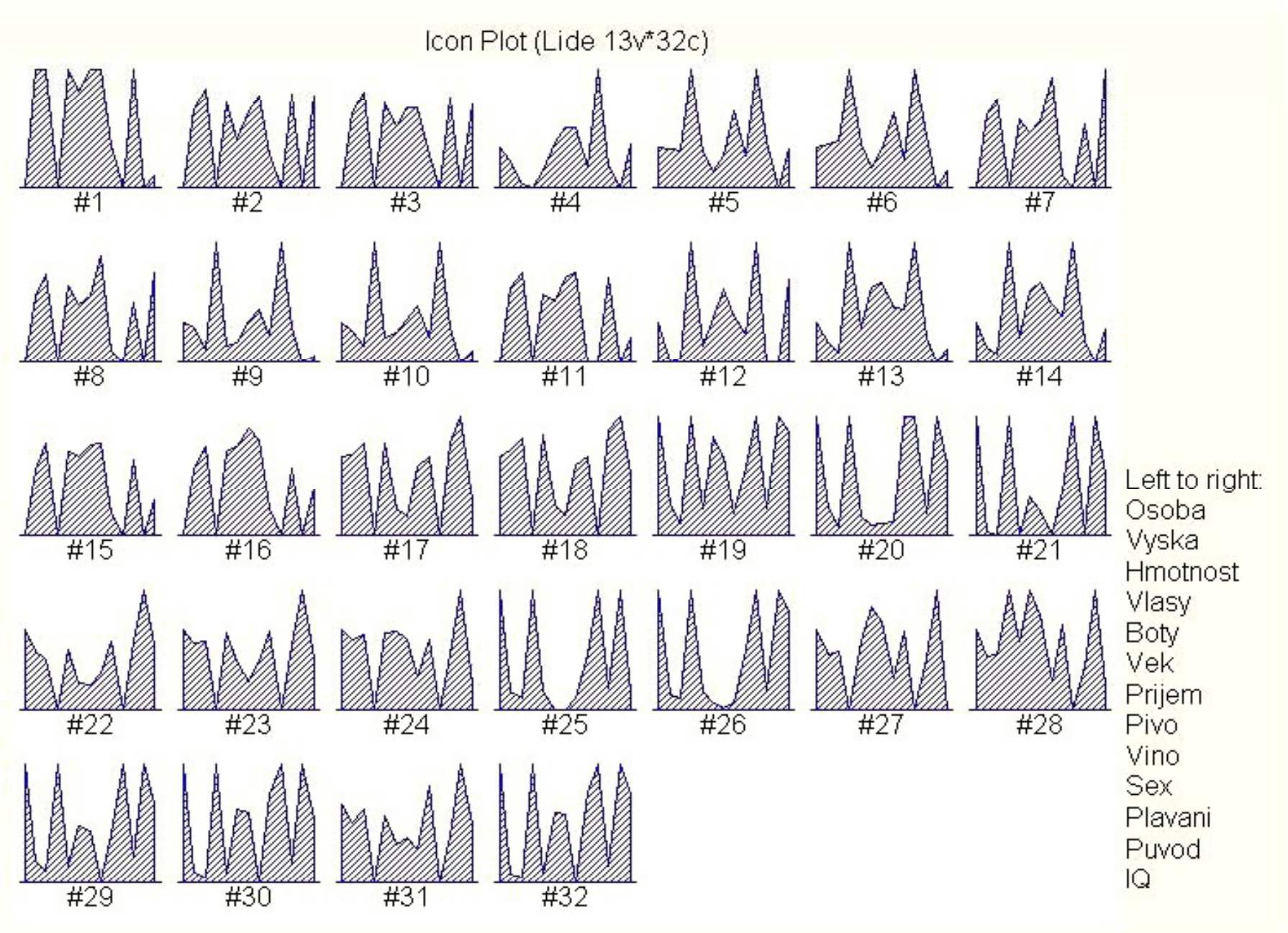
Profilové sloupce znaků pro každou osobu v průzkumové analýze dat *Lidé*, (STATISTICA).



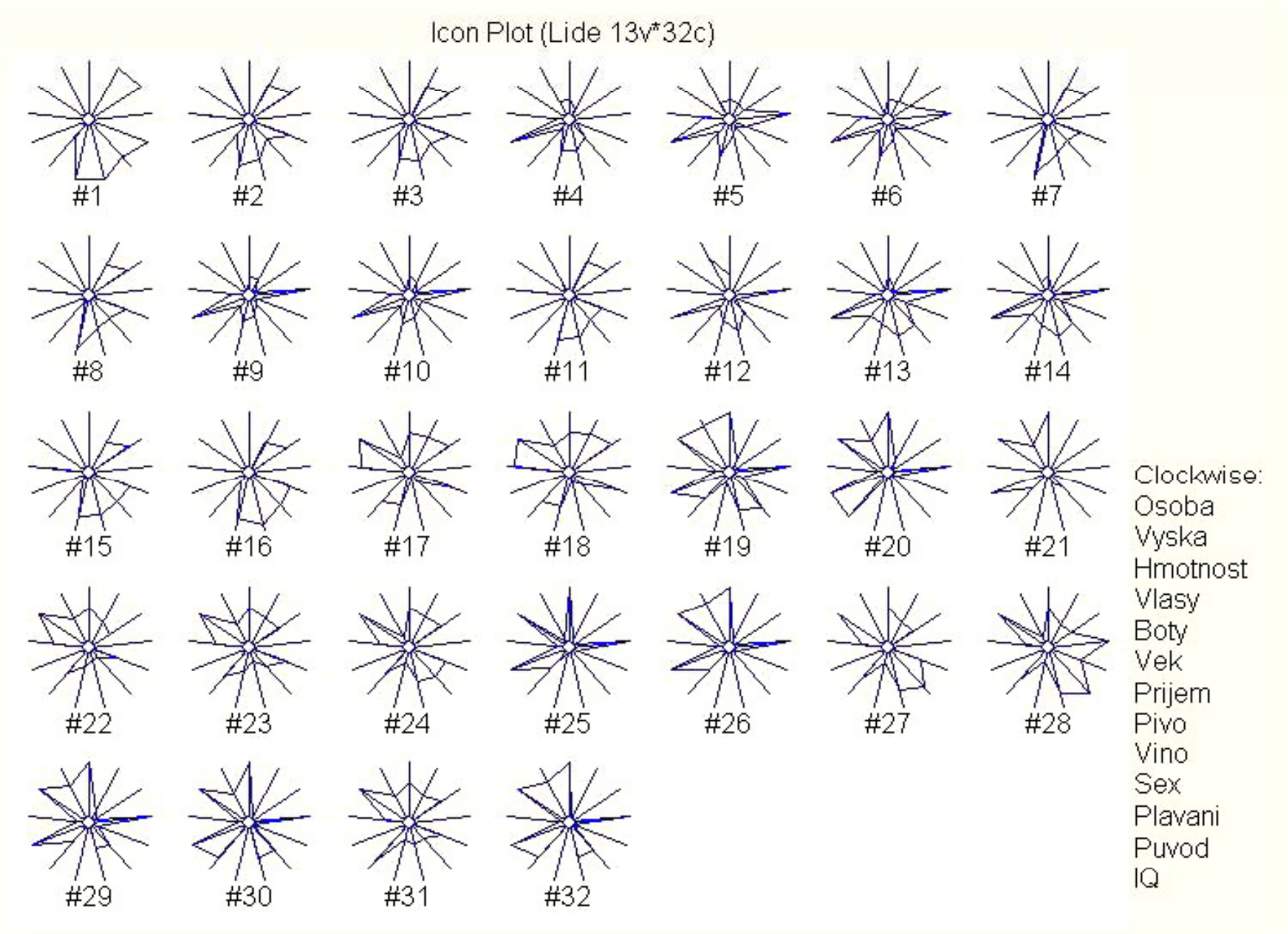
Profilové křivky znaků pro každou osobu v průzkumové analýze dat *Lidé*, (STATISTICA).



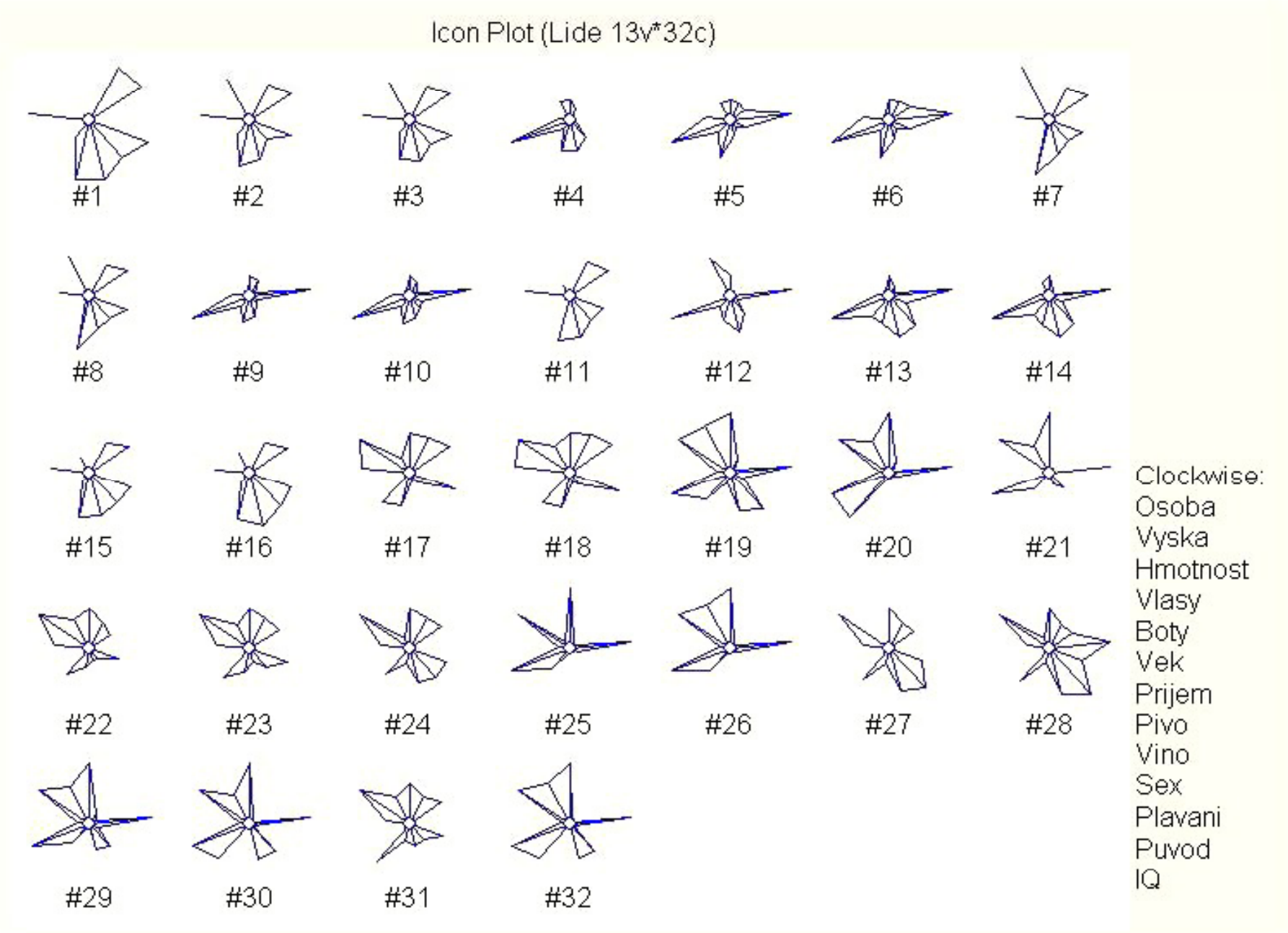
Černoffovy obličeje znaků pro každou osobu v průzkumové analýze dat *Lidé*, (STATISTICA).



Profily znaků pro každou osobu v průzkumové analýze dat *Lidé,* (STATISTICA)..



Sluníčka (polygony) znaků pro každou osobu v průzkumové analýze dat *Lidé*, (STATISTICA)..



Hvězdičky (polygony) znaků pro každou osobu v průzkumové analýze dat *Lidé,* (STATISTICA).

PŘÍKLAD 2.3 *Analýza zdrojové matice dat Hrách*

Zdrojová matice dat *Hrách* obsahuje znaky smyslového posouzení znaků odrůd hrachu. Objekty jsou vzorky pěti odrůd hrachu A až E, sklízené v pěti rozličných obdobích 1 až 5. Posouzení 10 porotci dvojmo, smyslové charakteristiky od 1 (nejhorší) do 9 (nejlepší), získáno 1200 řádků (objektů) tj. 60 vzorků × 2 krát opakováno × 10 porotců. Cílem je

1) průměrovat data,
2) vynést původní data do grafu a
3) vypočítat popisné jednorozměrné statistiky.

○ *Data:* matice dat $n = 1200$, $m = 12$ byla průměrována a výsledkem je matice 60 × 12 průměrných hodnot senzorického hodnocení pro znaky:

Aro je aroma, *Slad* je sladkost, *Med* je medovost, *Bez* je bezchuťovost, *Klas* je klasovost, *Tvrd* je tvrdost, *Bel* je bělost, *Bar1* je barva1, *Bar2* je barva2, *Bar3* je barva3, *Slup* je slupka, *Ztr* je ztráta.

Objekt	*Aro*	*Slad*	*Med*	*Bez*	*Klas*	*Tvrd*	*Bel*	*Bar1*	*Bar2*	*Bar3*	*Slup*	*Ztr*
B5	6,48	6,66	4,56	2,2	2,91	3,47	4,72	5,59	5,73	5,99	4,26	3,25
...	...	...	...	...	...	...	...	...	...	...	...	...
C2	3.70	3.86	2,33	4.11	6.18	6.83	5,15	5,77	5.29	4.42	1,99	4,59

STATISTICA Cz - [Data: 11Hrach.sta (13s krát 82ř)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

Načtení zdrojové matice dat

	1 Objekt	2 Aro	3 Slad	4 Med	5 Bez	6 Klas	7 Tvrd	8 Bel	9 Bar1	10 Bar2	11 Bar3	12 Slup	13 Ztr
1	B5	6,480	6,660	4,560	2,200	2,910	3,470	4,720	5,585	5,735	5,985	4,260	3,250
2	C4	5,750	6,090	3,810	2,320	4,030	3,770	4,170	5,730	5,745	5,325	3,820	3,380
3	B2	3,940	4,120	2,440	3,630	5,770	5,390	4,770	6,665	5,105	4,595	3,500	3,030
4	D5	6,600	6,120	4,440	1,930	3,310	4,460	4,860	5,160	5,740	6,565	2,120	3,940
5	D4	5,680	5,980	3,800	2,120	3,850	4,140	5,030	5,635	5,220	5,480	2,380	5,160
6	E2	4,740	4,660	2,880	2,940	5,650	5,770	5,310	5,940	5,270	5,890	1,750	3,640
7	B5	6,310	6,130	4,780	1,940	2,700	3,260	5,070	5,710	5,370	6,365	3,650	4,550
8	C5	6,200	6,020	4,650	1,780	3,120	3,740	5,250	5,655	5,475	5,960	2,510	3,800
9	C2	3,790	3,880	2,310	3,520	6,240	5,730	5,390	6,300	5,135	5,230	2,010	4,110
10	A4	5,680	6,340	3,750	2,790	4,170	3,870	4,520	4,920	5,760	4,570	2,970	4,410
11	D4	6,100	6,090	3,990	2,070	4,260	4,250	4,010	5,020	6,175	5,380	2,500	4,600
12	B1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4,270
13	D4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	,710
14	E4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	,650
15	B1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	,620
16	B5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	,580
17	D4	5,860	6,180	3,940	2,200	3,800	4,910	4,350	5,160	6,255	5,760	2,270	3,610
18	C2	3,960	4,480	2,300	3,940	6,230	6,410	4,470	5,105	5,720	4,970	2,050	4,530
19	A5	6,220	6,790	4,260	2,400	2,630	3,160	5,680	7,010	4,860	3,255	3,040	4,590
20	C3	5,110	5,250	3,090	3,270	5,280	5,240	5,610	6,595	5,110	3,950	2,740	4,230
21	B2	3,770	3,970	2,180	4,370	6,470	6,550	4,950	6,055	5,310	4,395	2,210	4,760
22	B5	7,090	6,090	5,180	1,740	2,570	3,180	5,230	5,920	5,515	4,115	2,090	3,100
23	D4	5,720	5,300	3,730	2,340	3,950	4,800	3,640	4,000	6,805	6,755	1,740	2,930
24	B1	3,220	3,210	1,950	4,420	6,240	7,270	4,600	6,030	5,600	4,165	1,680	3,580
25	A5	6,110	6,620	4,290	2,580	3,200	2,860	3,500	4,950	6,220	5,370	2,150	4,310
26	D4	6,070	6,270	3,980	2,190	3,890	4,240	3,850	4,460	6,675	6,205	2,200	3,700
27	A1	2,660	2,660	1,430	6,100	6,670	7,750	4,270	4,970	5,630	4,525	1,650	3,780
28	B3	5,260	5,490	3,460	3,030	4,850	4,170	5,220	5,410	5,415	6,120	3,080	3,950
29	C2	3,720	4,350	2,200	4,080	6,500	6,270	4,990	5,535	5,560	5,335	1,820	3,920
30	D3	5,430	5,190	3,470	2,400	4,430	5,260	4,460	4,780	5,720	5,895	1,610	3,770
31	B5	6,550	6,570	4,710	2,120	3,060	3,430	3,760	4,430	6,450	6,380	2,630	3,850
32	E3	5,530	5,410	3,680	2,470	4,720	5,780	3,880	4,335	6,470	6,790	1,800	2,950
33	C3	4,710	4,680	2,680	3,190	5,320	5,920	4,320	4,765	6,215	4,860	2,330	4,020
34	A5	6,280	7,030	4,910	2,380	2,190	2,600	4,560	5,905	5,585	4,950	3,630	3,380
35	D4	5,920	5,820	3,750	2,060	3,880	3,870	4,530	5,190	5,825	5,630	2,450	3,710
36	E4	6,090	5,720	3,800	1,940	4,440	4,450	3,940	4,625	6,510	7,175	2,180	2,990
37	B5	6,370	6,500	4,680	2,140	2,890	3,530	4,600	5,735	5,560	4,065	2,880	4,340
38	A4	5,710	5,680	3,970	2,650	4,390	3,720	5,320	6,280	5,120	6,075	2,390	2,550
39	C3	4,530	5,030	2,640	3,120	5,860	4,920	5,150	6,965	5,125	4,275	2,130	3,740
40	C4	5,950	6,280	4,040	2,190	3,930	3,610	4,120	5,395	5,815	4,505	3,090	5,340
41	D3	5,510	5,410	3,720	2,780	4,760	5,270	3,880	4,280	6,355	5,325	2,250	5,030
42	A1	3,100	3,430	1,800	4,860	6,220	7,070	4,140	5,275	5,580	3,695	2,050	4,730
43	A5	6,500	6,680	4,770	2,230	2,090	2,870	5,510	6,375	4,845	4,785	2,710	3,700
44	D3	5,460	5,410	3,270	2,970	5,150	4,980	3,610	4,305	6,600	5,435	2,370	4,440
45	A2	3,750	4,300	2,220	4,270	6,100	6,270	4,060	5,140	5,870	4,220	2,230	5,010
46	C4	5,860	5,270	3,730	2,500	3,860	4,300	4,150	4,495	6,230	6,140	2,110	3,290
47	A5	6,160	6,970	4,800	2,500	2,870	3,170	4,270	5,315	6,090	5,255	3,350	4,430
48	B2	3,870	3,880	2,230	4,060	5,990	6,310	4,450	5,530	5,785	4,980	1,930	3,620
49	B5	6,240	5,800	4,260	2,130	3,240	3,420	5,120	5,940	5,465	4,775	3,110	3,980

Pro nápovědu stiskněte F1 | Ř1;S11 | 5,985 | Filtr - | Váhy: VYPN | ABC | 123 | ZÁZN

Start | Total Commander 6.0 - P... | Doručená pošta - Micros... | Prezentace2 | STATISTICA Cz - [Dat... | 10:28

STATISTICA Cz - [Data: 11Hrach (13s krát 82ř)]

Soubor Upravit Zobrazit Vložit Formát Statistiky Data mining Grafy Nástroje Data Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

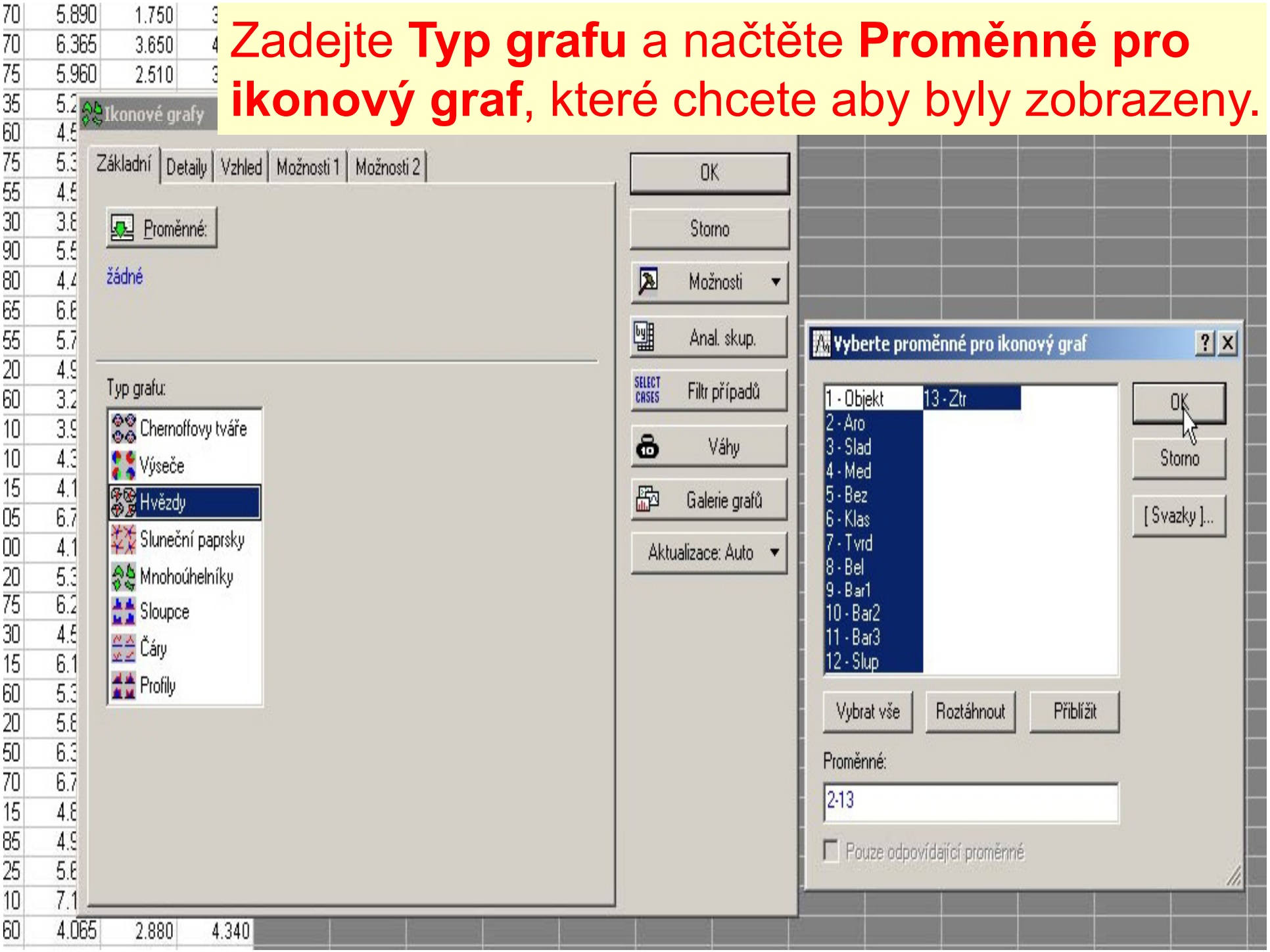
Arial 9 Proměnné Případy

Vypočtená data grafu... Přidat graf...

	1 Objekt	2 Aro	3 Slad	4 Med	5 Bez	6 Klas	7 Tvrd	8 Bel	9 Bar1	10 Bar2	11 Bar3	12 Slup	13 Ztr
B5	B5	6.480	6.660	4.560	2.200	2.910	3.470	4.720	5.585	5.735	5.985	4.260	3.250
C4	C4	5.750	6.090	3.810	2.320	4.030	3.770	4.170	5.730	5.745	5.325	3.820	3.380
B2	B2	3.940	4.120	2.440	3.630	5.770	5.390	4.770	6.665	5.105	4.595	3.500	3.030
D5	D5	6.600	6.120	4.440	1.930	3.310	4.460	4.860	5.160	5.740	6.565	2.120	3.940
D4	D4	5.680	5.980	3.800	2.120	3.850	4.140	5.030	5.635	5.220	5.480	2.380	5.160
E2	E2	4.740	4.660	2.880	2.940	5.650	5.770	5.310	5.940	5.270	5.690	1.750	3.640
B5	B5	6.310	6.130	4.780	1.940	2.700	3.260	5.070	5.710	5.370	6.365	3.650	4.550
C5	C5	6.200	6.020	4.650	1.780	3.120	3.740	5.250	5.655	5.475	5.960	2.510	3.800
C2	C2	3.790	3.880	2.310	3.520	6.240	5.730	5.390	6.300	5.135	5.230	2.010	4.110
A4	A4	5.680	6.340	3.750	2.790	4.170	3.870	4.520	4.920	5.760	4.570	2.970	4.410
D4	D4	6.100	6.090	3.990	2.070	4.260	4.250	4.010	5.020	6.175	5.380	2.500	4.600
B1	B1	3.410	3.180	1.820	4.640	6.240	7.430	4.260	4.835	5.955	4.550	1.850	4.270
D4	D4	5.890	6.090	3.990	2.290	3.900	4.590	4.630	5.890	5.630	3.815	2.200	4.710
E4	E4												
B1	B1												
B5	B5												
D4	D4												
C2	C2	3.960	4.480	2.300	3.940	6.230	6.410	4.470	5.105	5.720	4.970	2.050	4.530
A5	A5	6.220	6.790	4.260	2.400	2.630	3.160	5.680	7.010	4.860	3.255	3.040	4.590
C3	C3	5.110	5.250	3.090	3.270	5.280	5.240	5.610	6.595	5.110	3.950	2.740	4.230
B2	B2	3.770	3.970	2.180	4.370	6.470	6.550	4.950	6.055	5.310	4.395	2.210	4.760
B5	B5	7.090	6.090	5.180	1.740	2.570	3.180	5.230	5.920	5.515	4.115	2.090	3.100
D4	D4	5.720	5.300	3.730	2.340	3.950	4.800	3.640	4.000	6.805	6.755	1.740	2.930
B1	B1	3.220	3.210	1.950	4.420	6.240	7.270	4.600	6.030	5.600	4.165	1.680	3.580
A5	A5	6.110	6.620	4.290	2.580	3.200	2.860	3.500	4.950	6.220	5.370	2.150	4.310
D4	D4	6.070	6.270	3.980	2.190	3.890	4.240	3.850	4.460	6.675	6.205	2.200	3.700
A1	A1	2.660	2.660	1.430	6.100	6.670	7.750	4.270	4.970	5.630	4.525	1.650	3.780
B3	B3	5.260	5.490	3.460	3.030	4.850	4.170	5.220	5.410	5.415	6.120	3.080	3.950
C2	C2	3.720	4.350	2.200	4.080	6.500	6.270	4.990	5.535	5.560	5.335	1.820	3.920
D3	D3	5.430	5.190	3.470	2.400	4.430	5.260	4.460	4.780	5.720	5.895	1.610	3.770
B5	B5	6.550	6.570	4.710	2.120	3.060	3.430	3.760	4.430	6.450	6.380	2.630	3.850
E3	E3	5.530	5.410	3.680	2.470	4.720	5.780	3.880	4.335	6.470	6.790	1.800	2.950
C3	C3	4.710	4.680	2.680	3.190	5.320	5.920	4.320	4.765	6.215	4.860	2.330	4.020
A5	A5	6.280	7.030	4.910	2.380	2.190	2.600	4.560	5.905	5.585	4.950	3.630	3.380
D4	D4	5.920	5.820	3.750	2.060	3.880	3.870	4.530	5.190	5.825	5.630	2.450	3.710
E4	E4	6.090	5.720	3.800	1.940	4.440	4.450	3.940	4.625	6.510	7.175	2.180	2.990
B5	B5	6.370	6.500	4.680	2.140	2.890	3.530	4.600	5.735	5.560	4.065	2.880	4.340
A4	A4	5.710	5.680	3.970	2.650	4.390	3.720	5.320	6.280	5.120	6.075	2.390	2.550
C3	C3	4.530	5.030	2.640	3.120	5.860	4.920	5.150	6.965	5.125	4.275	2.130	3.740
C4	C4	5.950	6.280	4.040	2.190	3.930	3.610	4.120	5.395	5.815	4.505	3.090	5.340
D3	D3	5.510	5.410	3.720	2.780	4.760	5.270	3.880	4.280	6.355	5.325	2.250	5.030
A1	A1	3.100	3.430	1.800	4.860	6.220	7.070	4.140	5.275	5.580	3.695	2.050	4.730
A5	A5	6.500	6.680	4.770	2.230	2.090	2.870	5.510	6.375	4.845	4.785	2.710	3.700
D3	D3	5.460	5.410	3.270	2.970	5.150	4.980	3.610	4.305	6.600	5.435	2.370	4.440
A2	A2	3.750	4.300	2.220	4.270	6.100	6.270	4.060	5.140	5.870	4.220	2.230	5.010
C4	C4	5.860	5.270	3.730	2.500	3.860	4.300	4.150	4.495	6.230	6.140	2.110	3.290
A5	A5	6.160	6.970	4.800	2.500	2.870	3.170	4.270	5.315	6.090	5.255	3.350	4.430
B2	B2	3.870	3.880	2.230	4.060	5.990	6.310	4.450	5.530	5.785	4.980	1.930	3.620
B5	B5	6.240	5.800	4.260	2.130	3.240	3.420	5.120	5.940	5.465	4.775	3.110	3.980
D3	D3	5.690	4.970	3.250	2.630	4.530	5.360	4.570	4.990	5.820	5.620	2.240	3.070
A1	A1	2.280	2.230	1.290	6.450	6.700	7.830	5.530	7.300	4.360	3.505	1.640	3.140

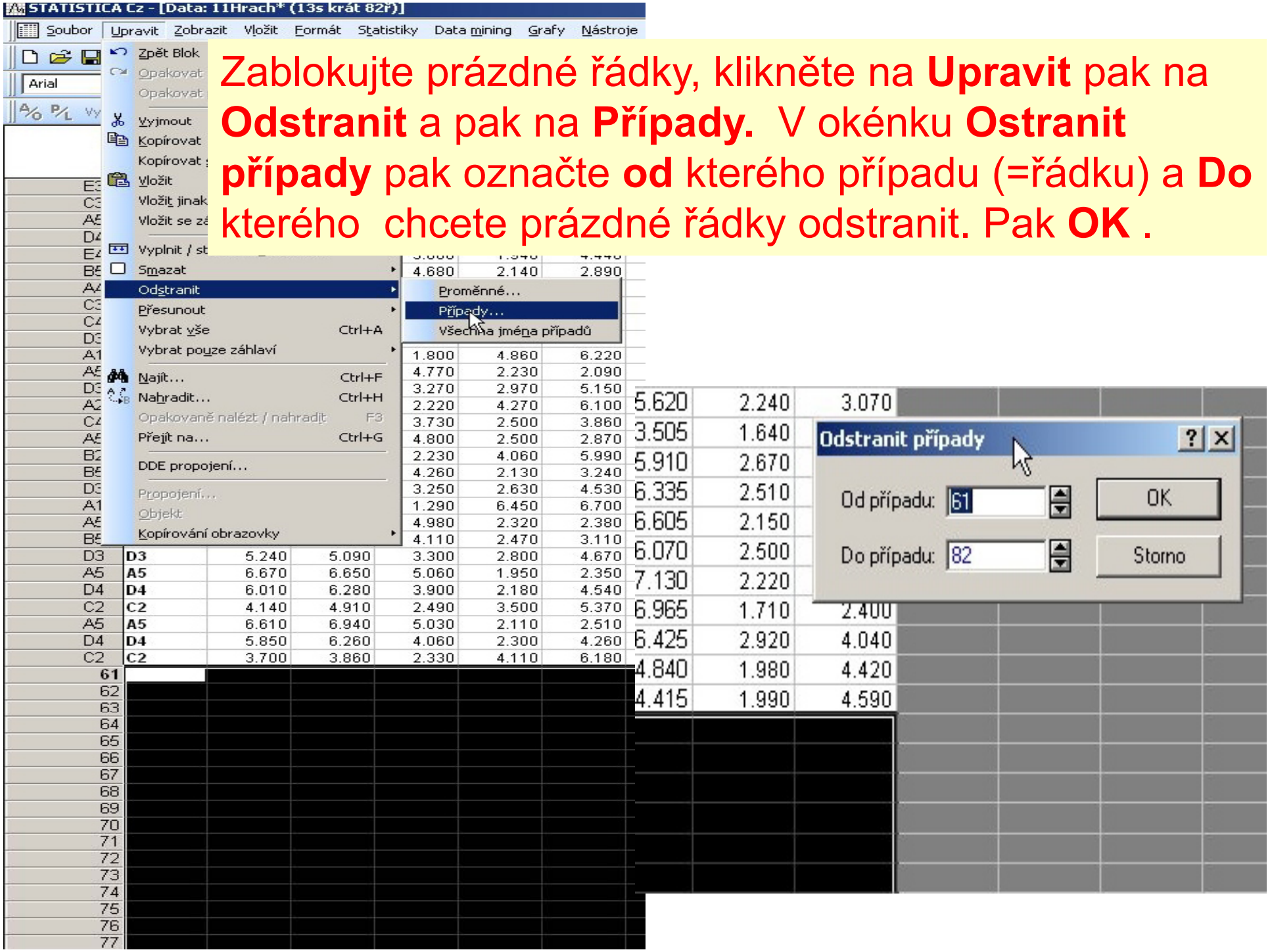
Kopie 1. sloupce **Objekt** do nultého sloupce

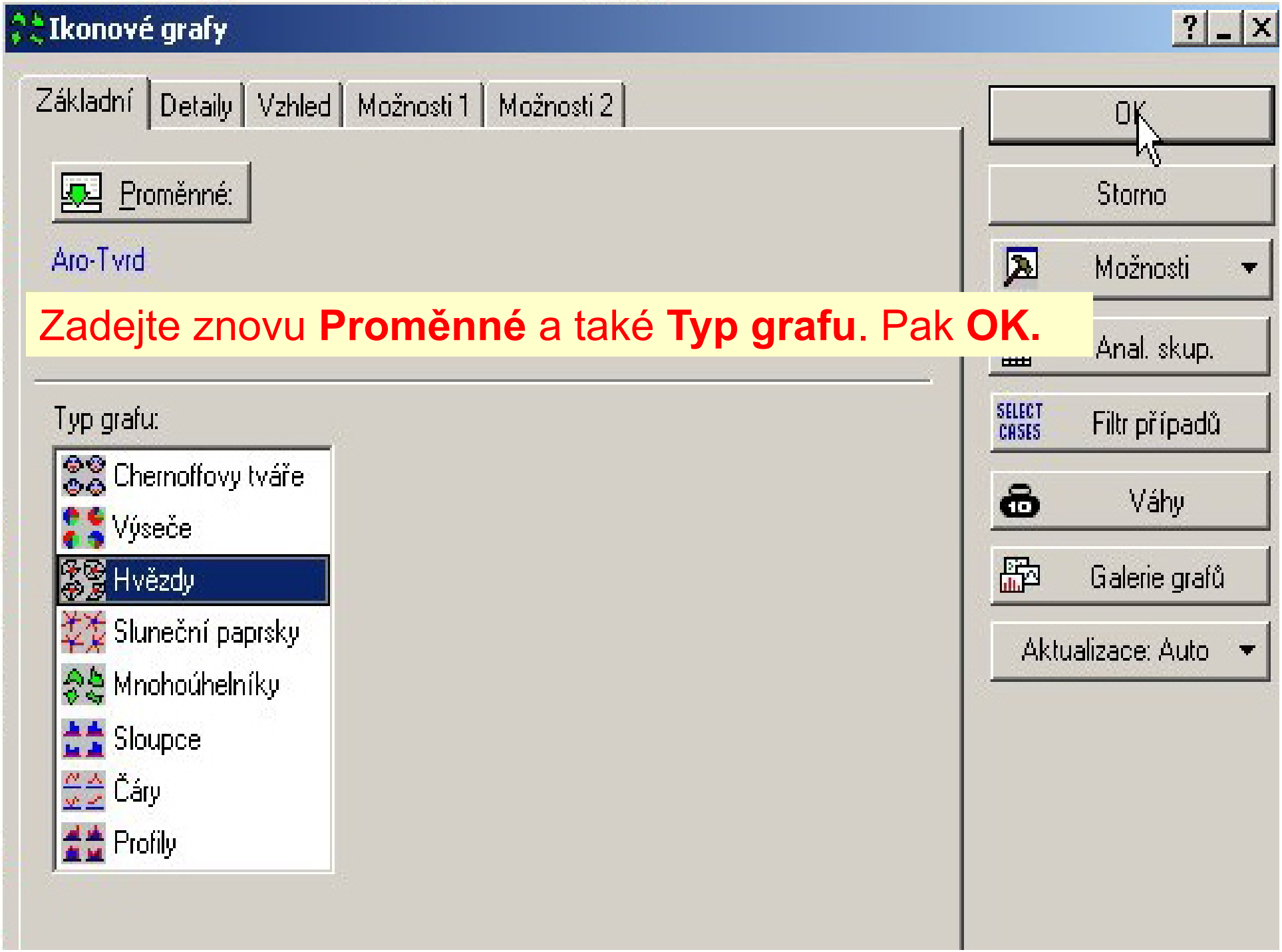
Zadejte **Typ grafu** a načtěte **Proměnné pro ikonový graf**, které chcete aby byly zobrazeny.



Ikonový graf **Hvězdy** není hezký, protože do zdrojové matice byly omylem načteny také prázdné řádky – zde jsou to modré tečky - které je třeba nejprve odstranit.

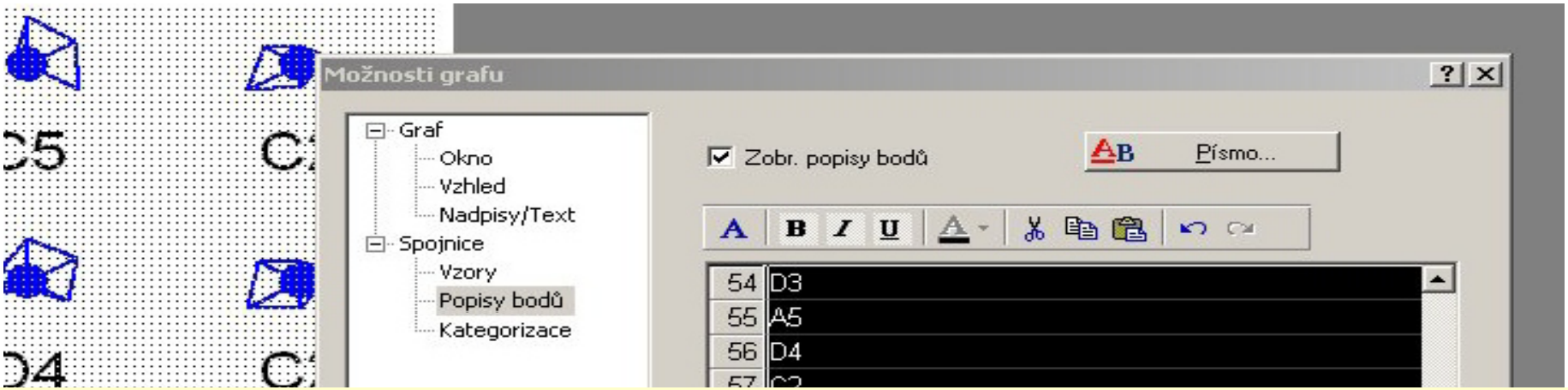
Zablokujte prázdné řádky, klikněte na **Upravit** pak na **Odstranit** a pak na **Případy.** V okénku **Ostranit případy** pak označte **od** kterého případu (=řádku) a **Do** kterého chcete prázdné řádky odstranit. Pak **OK** .



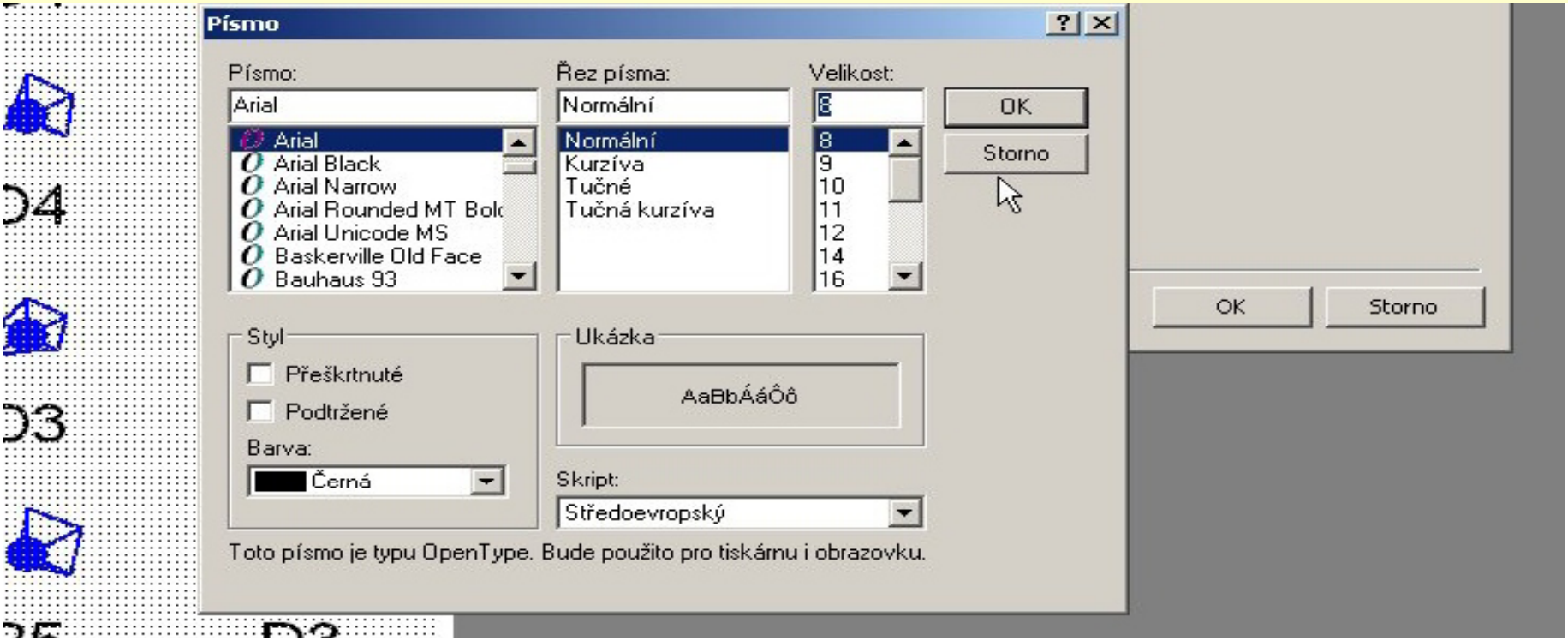


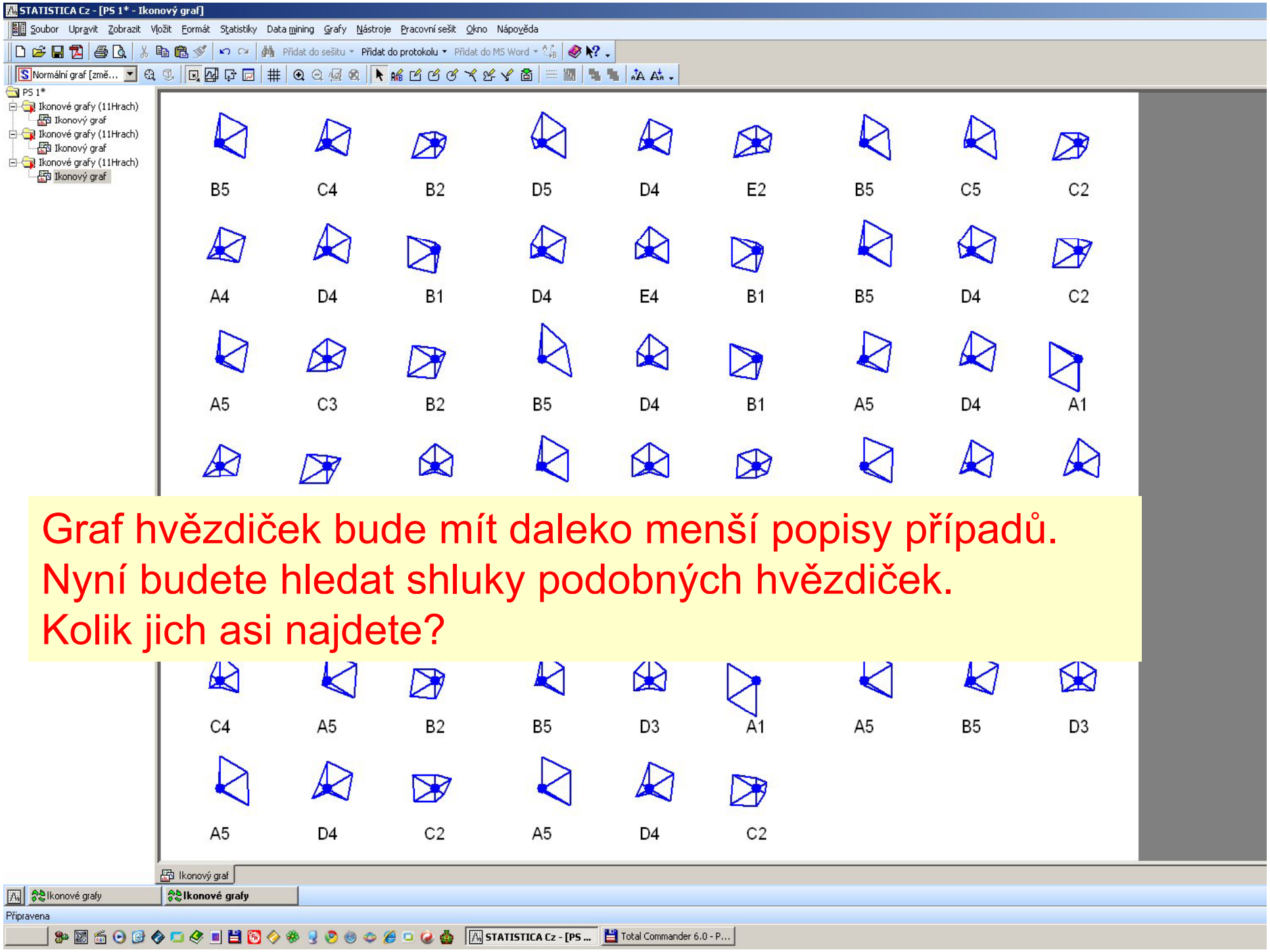
Zadejte znovu **Proměnné** a také **Typ grafu**. Pak **OK.**

Obdržíte hezký graf hvězdiček od všech případů.
Popisy jsou příliš veliké, proto je zmenšíme.

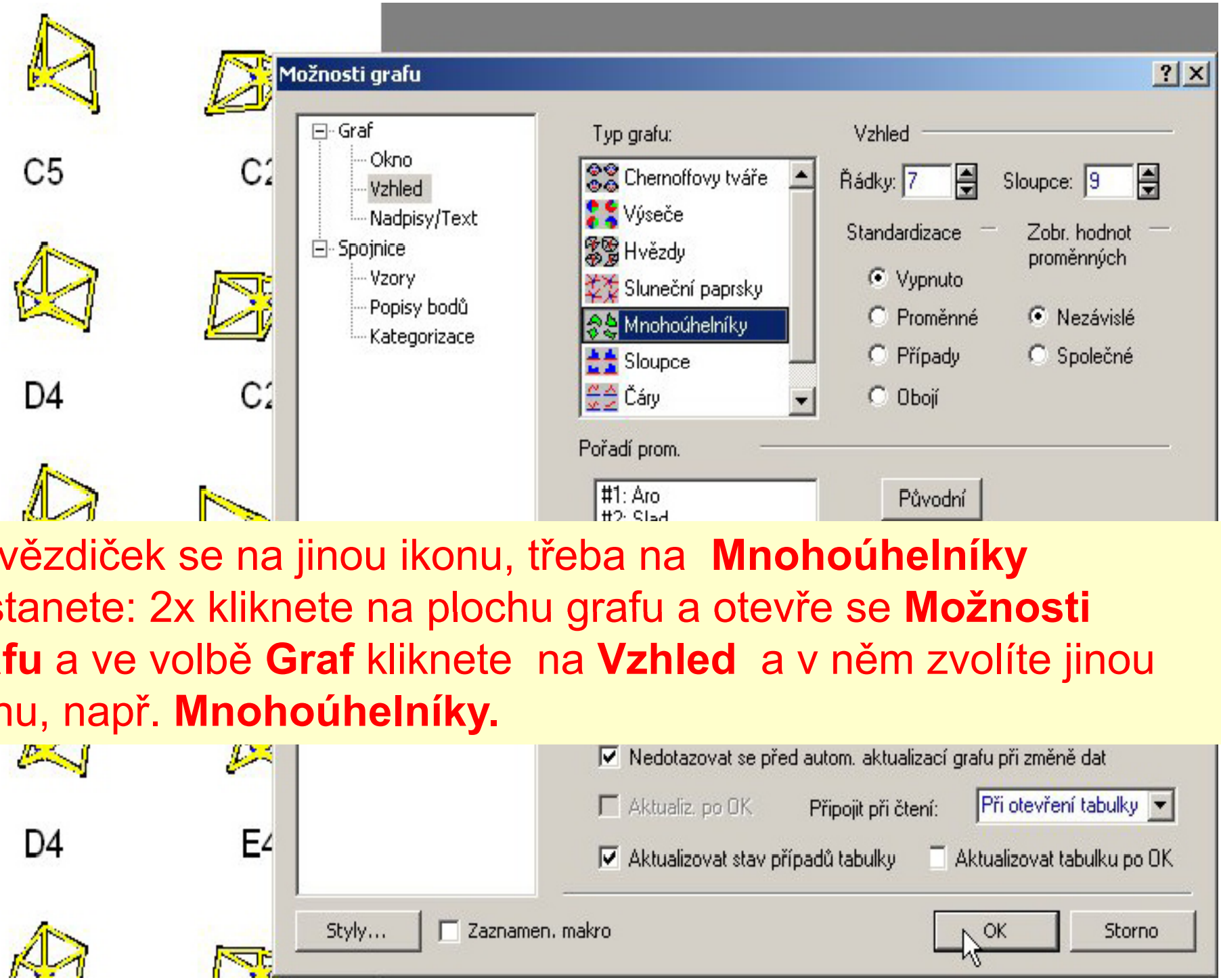


Otevřete **Možnosti grafu** a pak **Popisy bodů** a zablokujete popisy bodů. Pak kliknete na **Písmo** a zvolíte menší font, zde např. 8. Pak **OK** a **OK.**

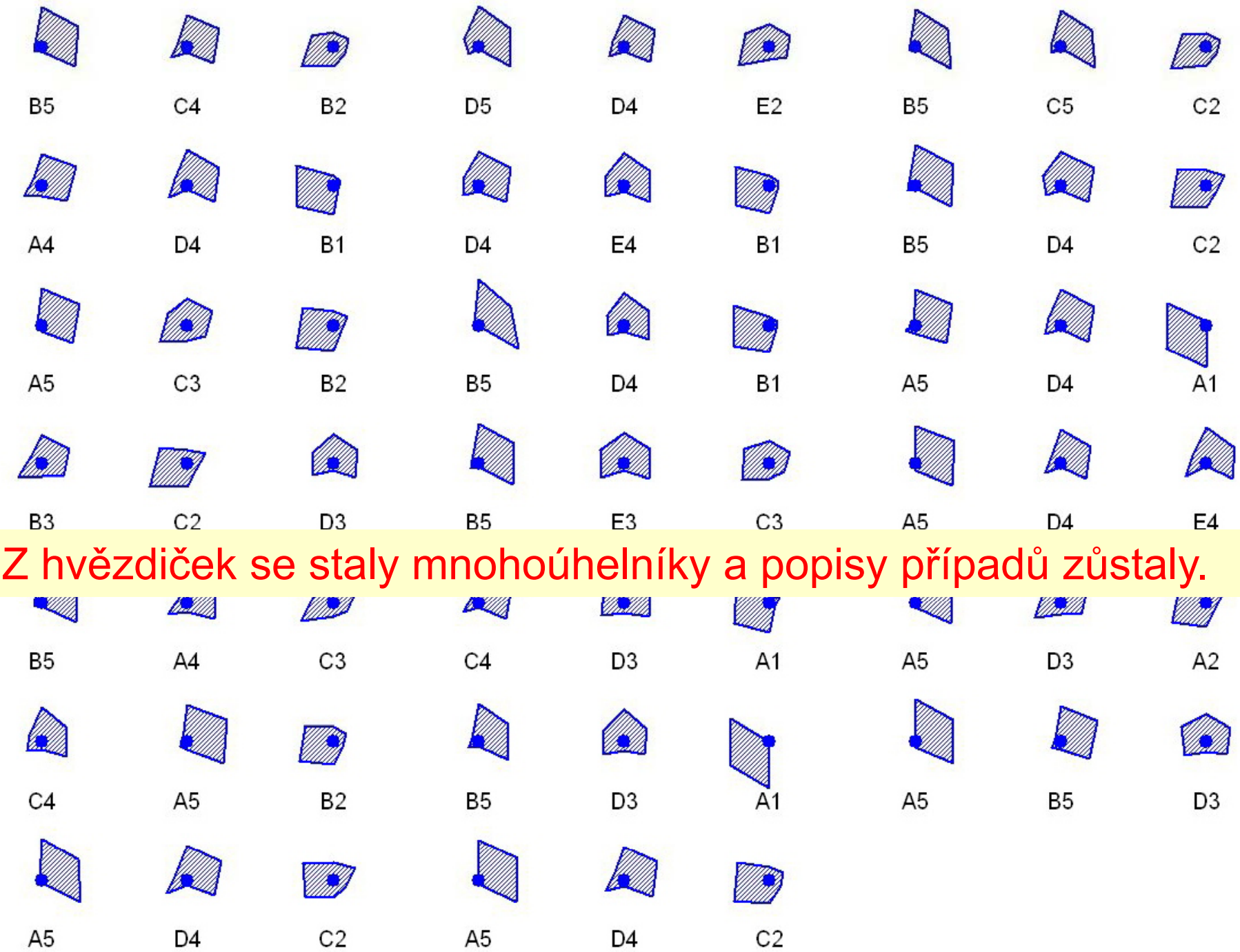




Graf hvězdiček bude mít daleko menší popisy případů.
Nyní budete hledat shluky podobných hvězdiček.
Kolik jich asi najdete?

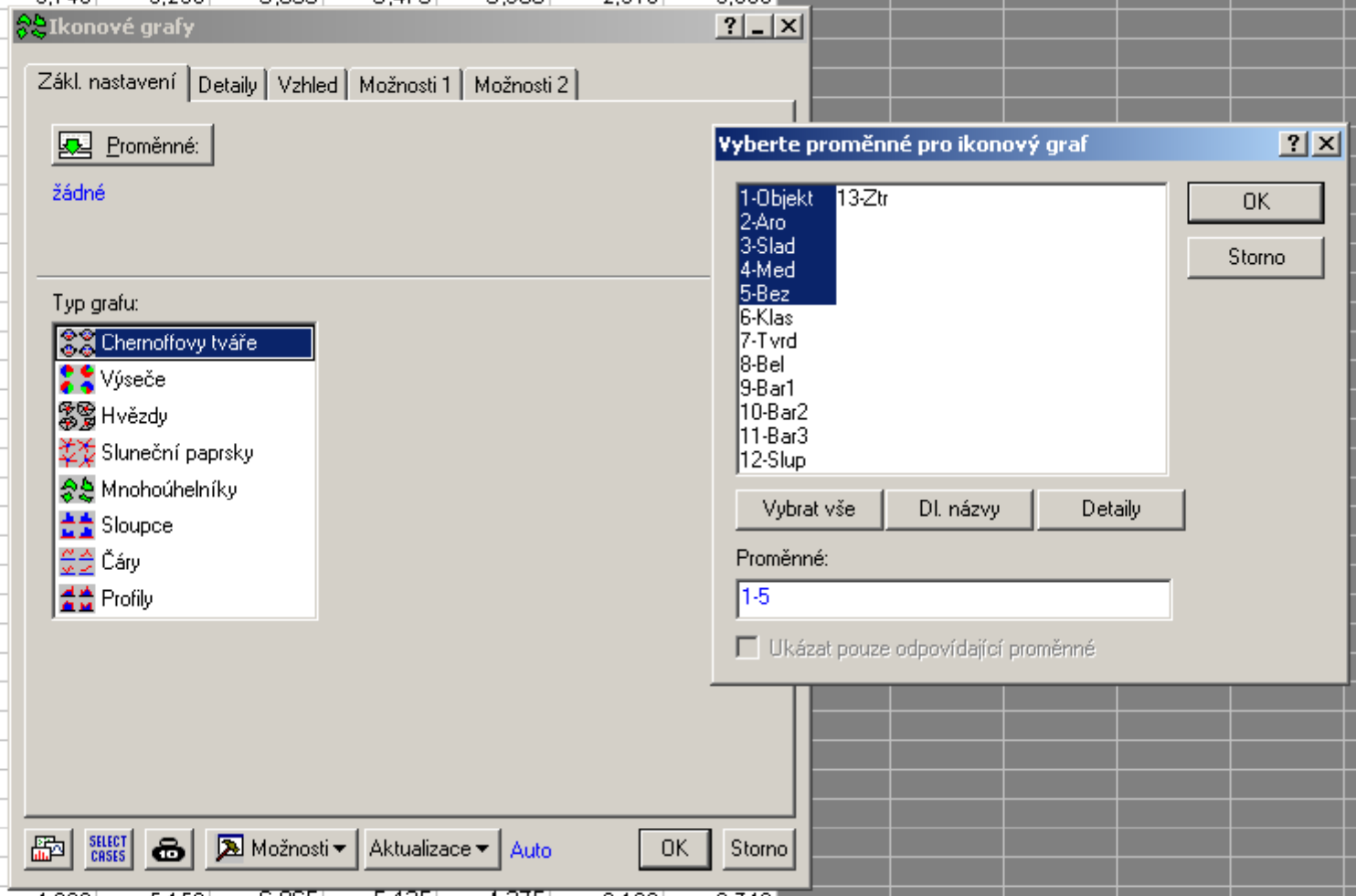


Z hvězdiček se na jinou ikonu, třeba na **Mnohoúhelníky** dostanete: 2x kliknete na plochu grafu a otevře se **Možnosti grafu** a ve volbě **Graf** kliknete na **Vzhled** a v něm zvolíte jinou ikonu, např. **Mnohoúhelníky.**



Z hvězdiček se staly mnohoúhelníky a popisy případů zůstaly.

3,630	5,770	5,390	4,770	6,665	5,105	4,595	3,500	3,030
1,930	3,310	4,460	4,860	5,160	5,740	6,565	2,120	3,940
2,120	3,850	4,140	5,030	5,635	5,220	5,480	2,380	5,160
2,940	5,650	5,770	5,310	5,940	5,270	5,890	1,750	3,640
1,940	2,700	3,260	5,070	5,710	5,370	6,365	3,650	4,550
1,780	3,120	3,740	5,250	5,655	5,475	5,960	2,510	3,800
3,520	6,240							
2,790	4,170							
2,070	4,260							
4,640	6,240							
2,290	3,900							
2,260	4,220							
4,500	6,040							
1,970	2,920							
2,200	3,800							
3,940	6,230							
2,400	2,630							
3,270	5,280							
4,370	6,470							
1,740	2,570							
2,340	3,950							
4,420	6,240							
2,580	3,200							
2,190	3,890							
6,100	6,670							
3,030	4,850							
4,080	6,500							
2,400	4,430							
2,120	3,060							
2,470	4,720							
3,190	5,320							
2,380	2,190							
2,060	3,880							
1,940	4,440							
2,140	2,890							
2,650	4,390							
3,120	5,860	4,920	5,150	6,965	5,125	4,275	2,130	3,740
2,190	3,930	3,610	4,120	5,395	5,815	4,505	3,090	5,340
2,780	4,760	5,270	3,880	4,280	6,355	5,325	2,250	5,030
4,860	6,220	7,070	4,140	5,275	5,580	3,695	2,050	4,730
2,230	2,090	2,870	5,510	6,375	4,845	4,785	2,710	3,700
2,970	5,150	4,980	3,610	4,305	6,600	5,435	2,370	4,440
4,270	6,100	6,270	4,060	5,140	5,870	4,220	2,230	5,010
2,500	3,860	4,300	4,150	4,495	6,230	6,140	2,110	3,290
2,500	2,870	3,170	4,270	5,315	6,090	5,255	3,350	4,430



STATISTICA Cz - [PS 1* - Ikonový graf (11Hrach.sta 13v*82c)]

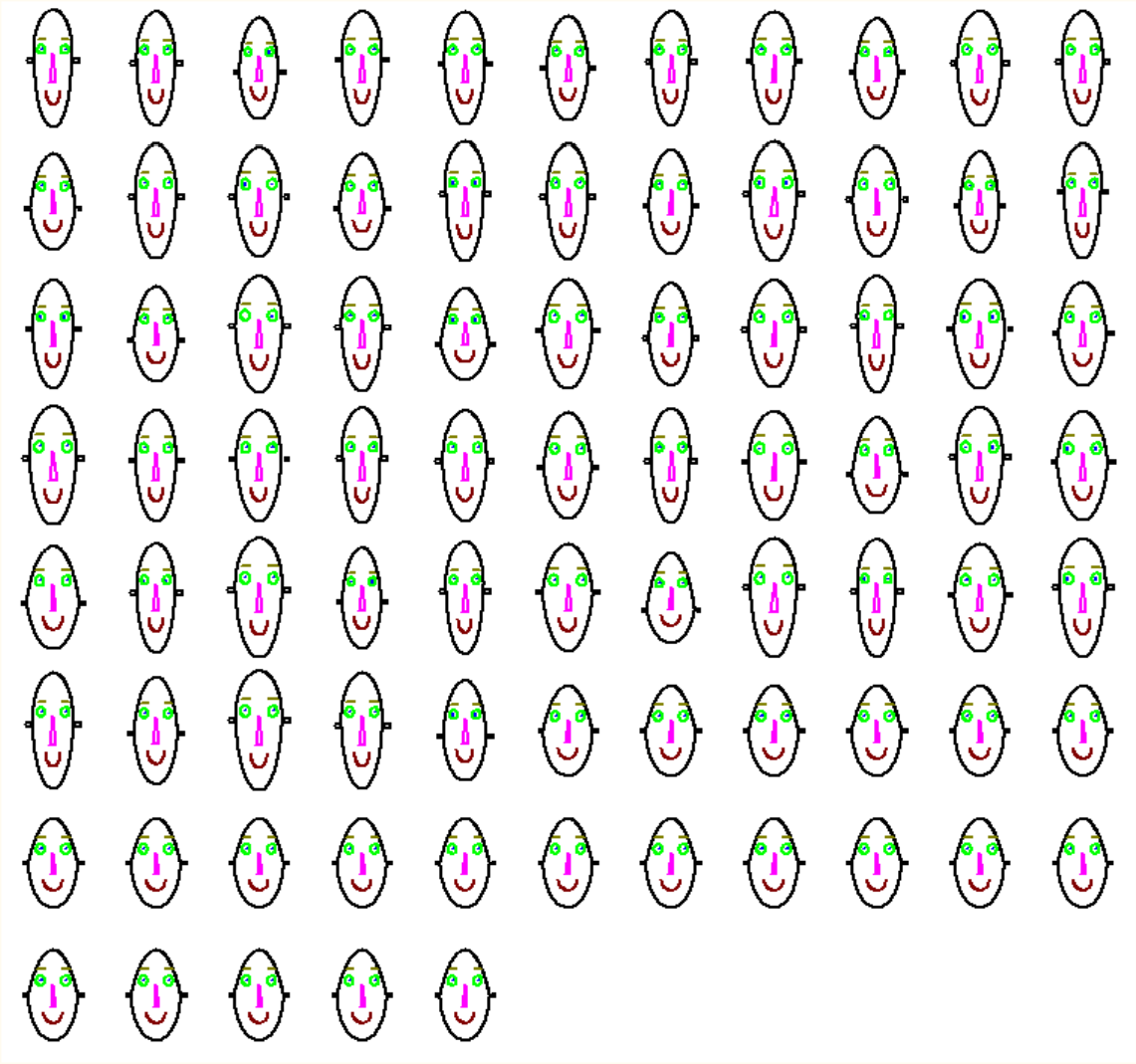
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Normální graf [změ...

PS 1*
- Maticové grafy (11Hrach.sta
  - Maticový graf (11Hrach.
- Ikonové grafy (11Hrach.sta)
  - Ikonový graf (11Hrach.s
- Ikonové grafy (11Hrach.sta)
  - Ikonový graf (11Hrach.s
- Ikonové grafy (11Hrach.sta)
  - Ikonový graf (11Hrach.s
- Ikonové grafy (11Hrach.sta)
  - Ikonový graf (11Hrach.s

Ikonový graf (11Hrach.sta 13v*82c)



— tvář/šíř = Objekt
— ucho/úrov = Aro
— polovina tváře/výš = Slad
— horní tvář/exc = Med
— dolní tvář/exc = Bez

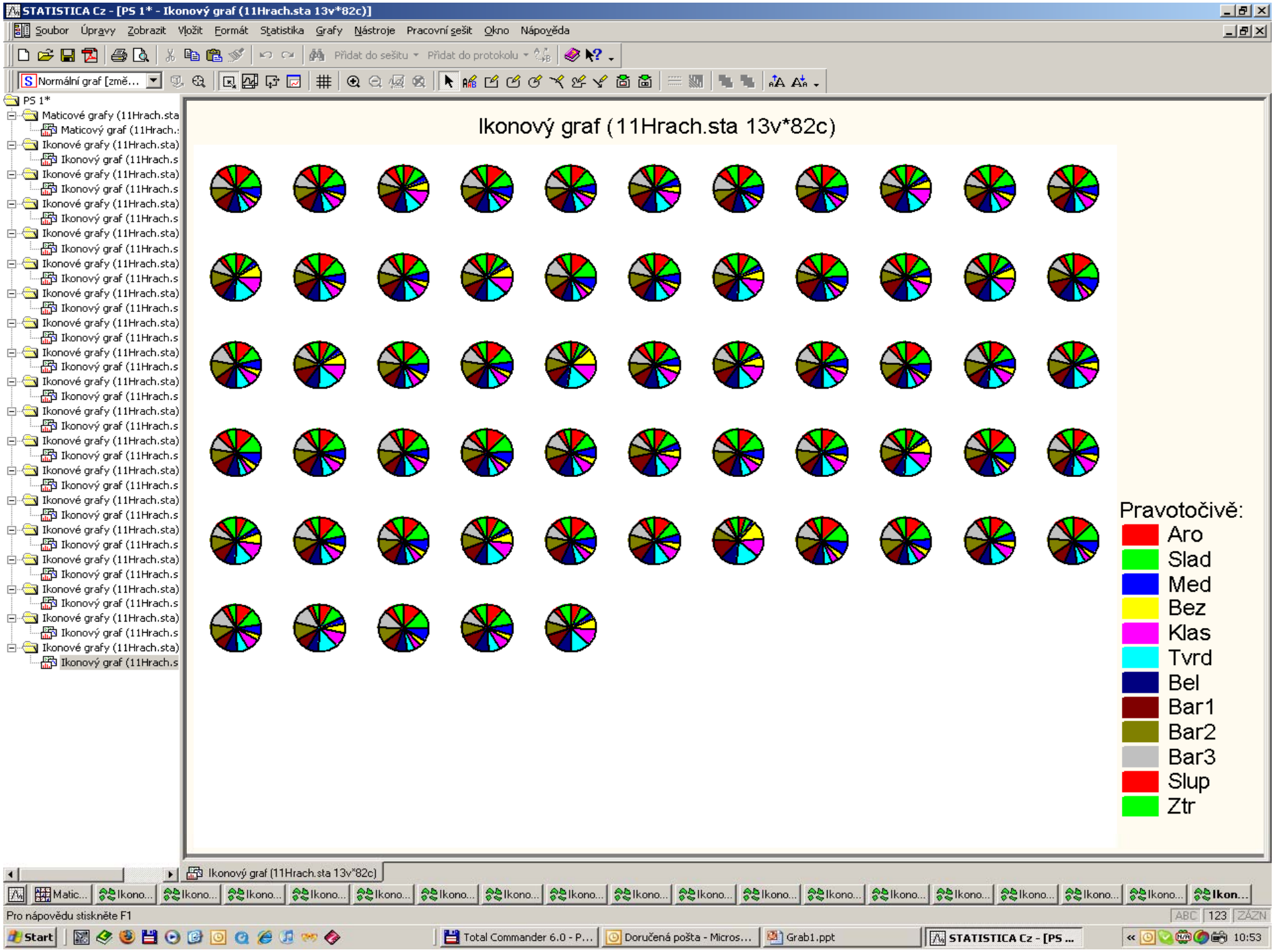
Ikonový graf (11Hrach.sta 13v*82c)

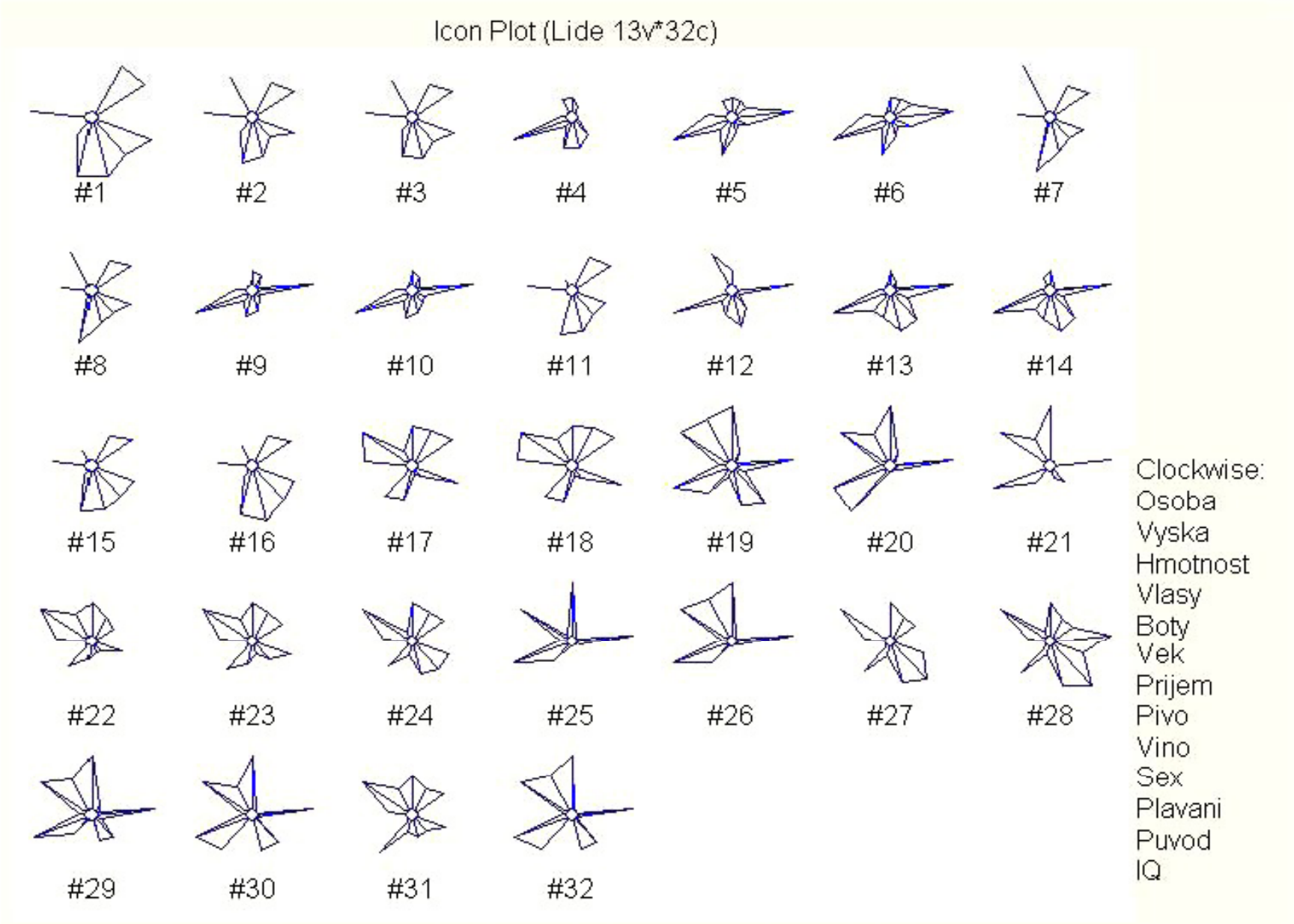
Maticové grafy | Ikonové grafy | Ikonové grafy | Ikonové grafy | Ikonové grafy

Pro nápovědu stiskněte F1

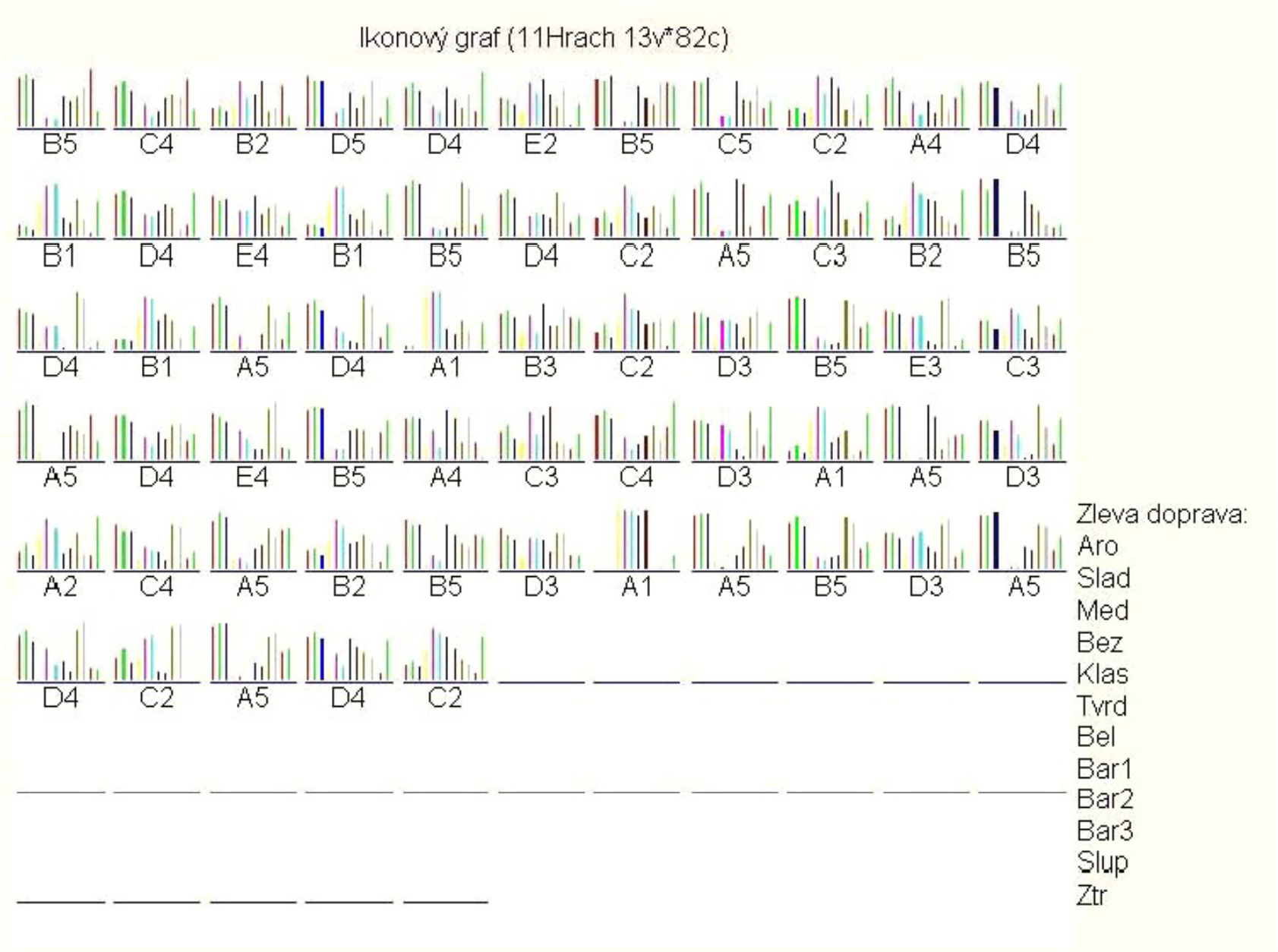
ABC 123 ZÁZN

Start | Total Commander 6.0 - P... | Doručená pošta - Micros... | Grab1.ppt | STATISTICA Cz - [PS ... | 10:41

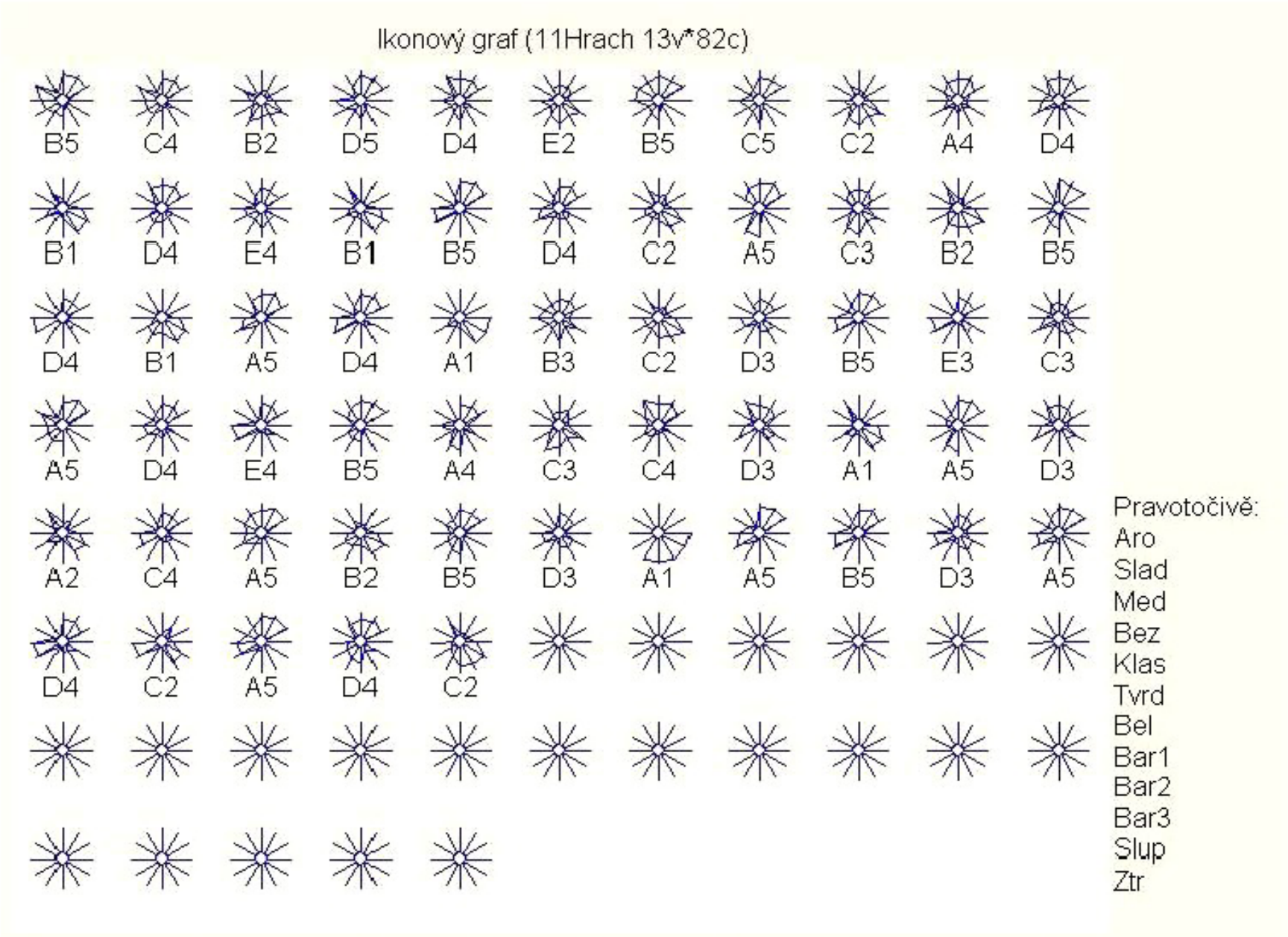

STATISTICA Cz - [PS 1* - Ikonový graf (11Hrach.sta 13v*82c)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Ikonový graf (11Hrach.sta 13v*82c)
Pravotočivě:
Aro
Slad
Med
Bez
Klas
Tvrd
Bel
Bar1
Bar2
Bar3
Slup
Ztr
Pro nápovědu stiskněte F1



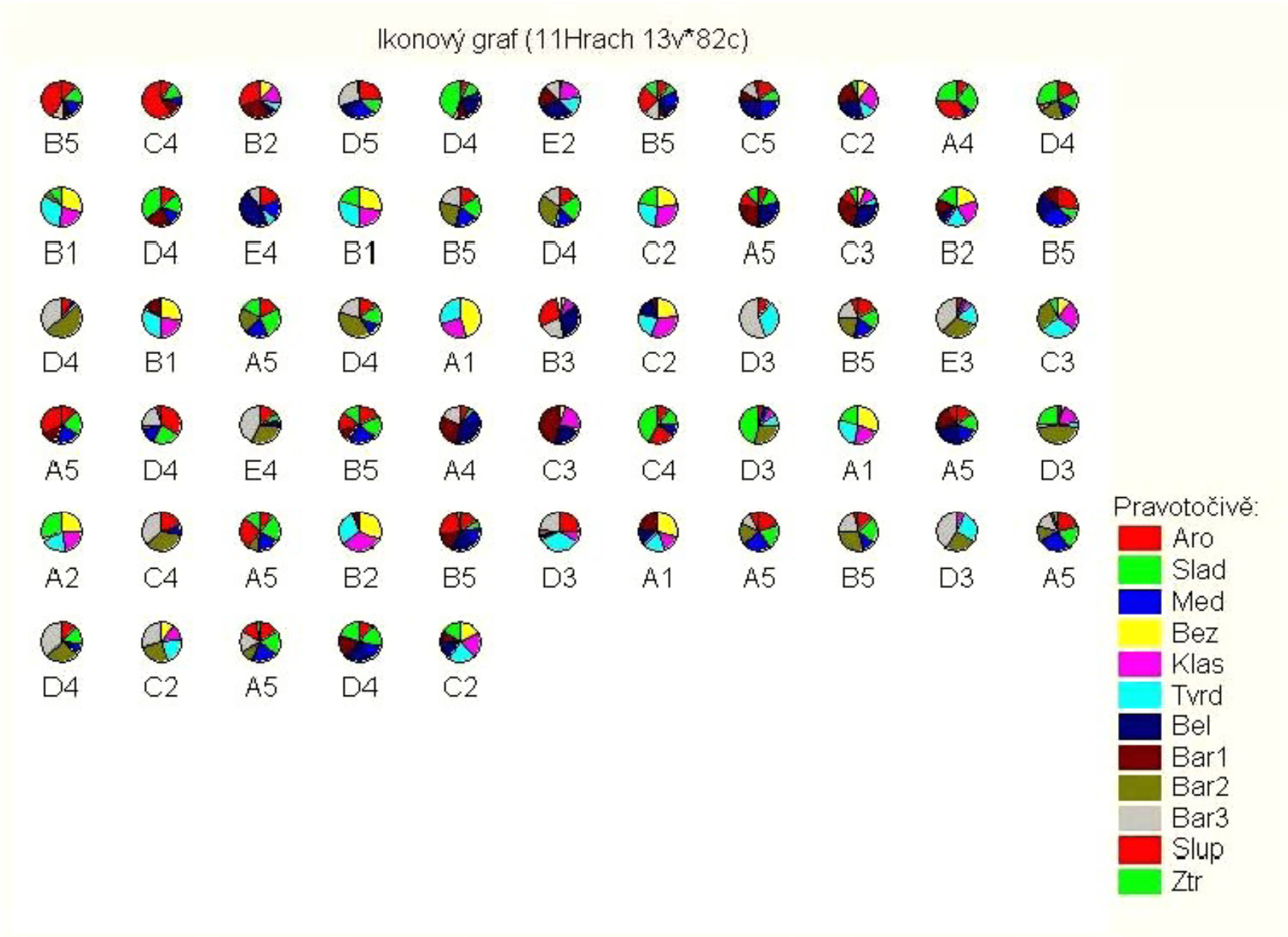
Profily znaků pro každý objekt v průzkumové analýze dat *Hrách,* (STATISTICA).



Sloupce znaků pro každý objekt v EDA dat *Hrách,* (STATISTICA).



Sluníčka znaků pro každý objekt v EDA dat *Hrách,* (STATISTICA).



Výseče znaků pro každý objekt v EDA dat *Hrách,* (STATISTICA).

PŘÍKLAD 2.4 *Popis a třídění polétavých mšic*

Jeffers (1967) studoval 40 polétavých mšic *(Alate adelges)* pomoci světelné pasti, změřeno 19 znaků: 14 znaků délky a šířky, 4 znaky o počtu, 1 znak binární, přítomnost či absenci. Mšice se obtížně rozlišují dle taxonometrických klíčů. Před PCA je třeba standardizaci dat, protože znaky představují směs délek a počtů.

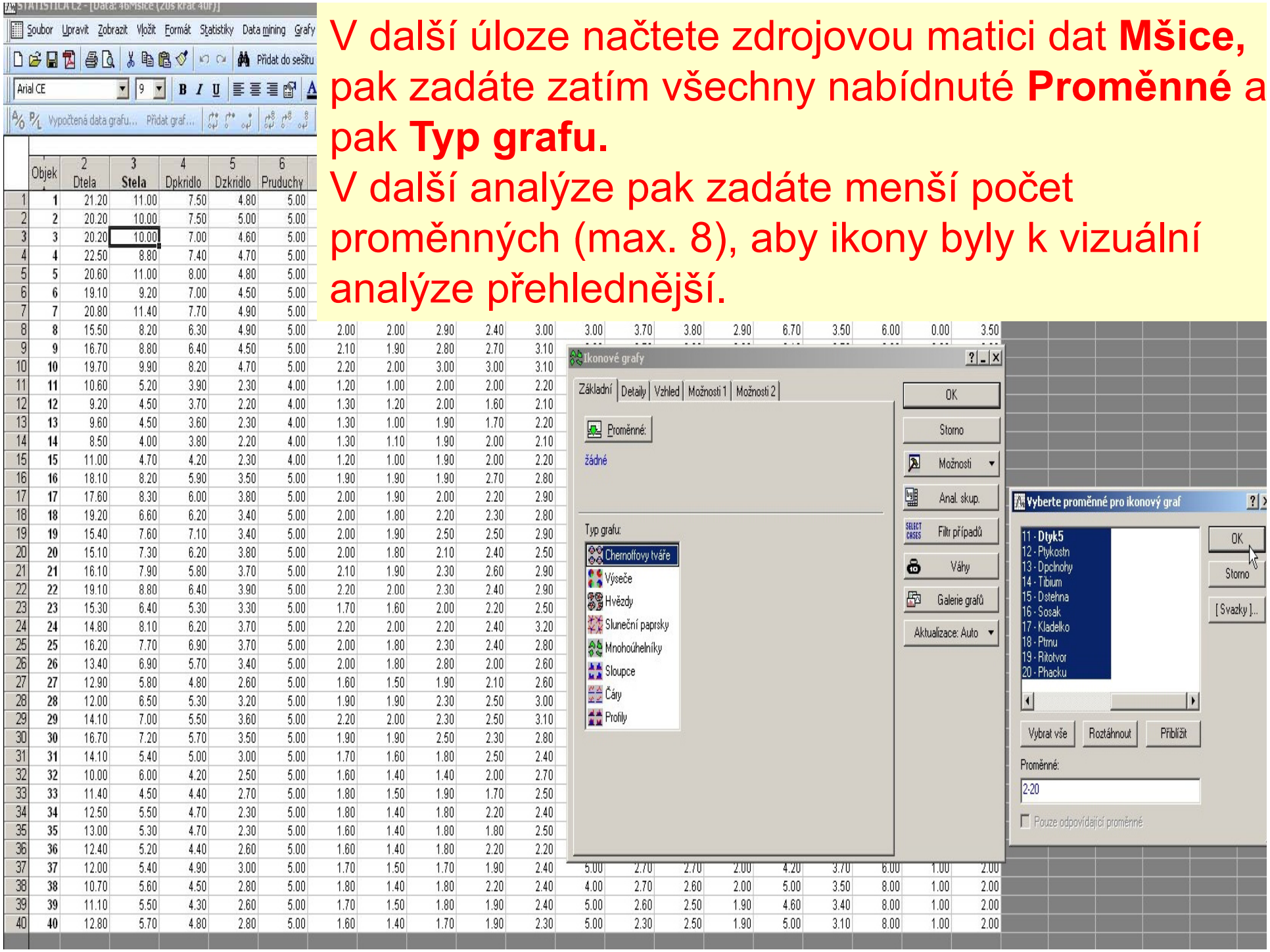
○ *Data*: x_1 značí délku těla, x_2 značí šířku těla, x_3 je délka předního křídla, x_4 je délka zadního křídla, x_5 je počet průduchů, x_6 je délka tykadla I, x_7 je délka tykadla II, x_8 je délka tykadla III, x_9 je délka tykadla IV, x_{10} je délka tykadla V, x_{11} je počet tykadlových ostnů, x_{12} je délka posledního článku nohy, x_{13} je délka holeně, tibia, x_{14} je délka stehna, x_{15} je délka sosáku, x_{16} je délka kladélka, x_{17} je počet kladélkových trnů, x_{18} je řitní otvor, x_{19} je počet háčků zadních křídel.

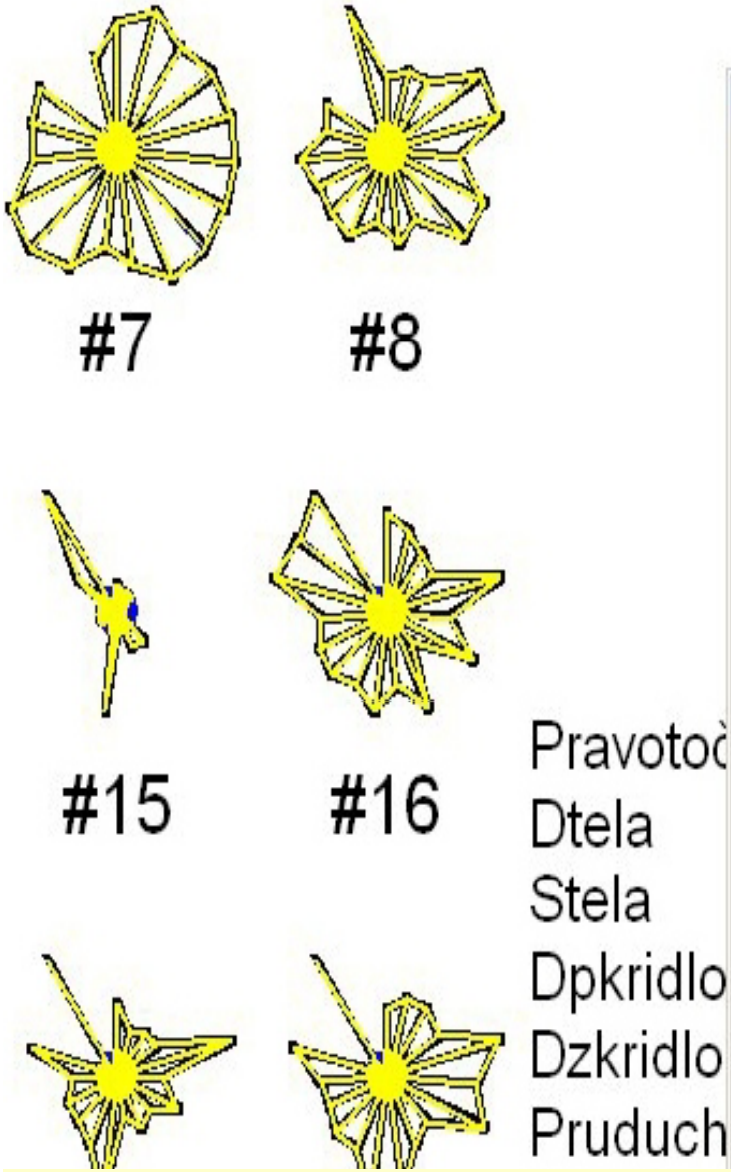
i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}
1	21.2	11.0	7.5	4.8	5.0	2.0	2.0	2.8	2.8	3.3	3.0	4.4	4.5	3.6	7.0	4.0	8.0	0.0	3.0
2	20.2	10.0	7.5	5.0	5.0	2.3	2.1	3.0	3.0	3.2	5.0	4.2	4.5	3.5	7.6	4.2	8.0	0.0	3.0
3	20.2	10.0	7.0	4.6	5.0	1.9	2.1	3.0	2.5	3.3	1.0	4.2	4.4	3.3	7.0	4.0	6.0	0.0	3.0
4	22.5	8.8	7.4	4.7	5.0	2.4	2.1	3.0	2.7	3.5	5.0	4.2	4.4	3.6	6.8	4.1	6.0	0.0	3.0
5	20.6	11.0	8.0	4.8	5.0	2.4	2.0	2.9	2.7	3.0	4.0	4.2	4.7	3.5	6.7	4.0	6.0	0.0	3.0
6	19.1	9.2	7.0	4.5	5.0	1.8	1.9	2.8	3.0	3.2	5.0	4.1	4.3	3.3	5.7	3.8	8.0	0.0	3.5
7	20.8	11.4	7.7	4.9	5.0	2.5	2.1	3.1	3.1	3.2	4.0	4.2	4.7	3.6	6.6	4.0	8.0	0.0	3.0
8	15.5	8.2	6.3	4.9	5.0	2.0	2.0	2.9	2.4	3.0	3.0	3.7	3.8	2.9	6.7	3.5	6.0	0.0	3.5
9	16.7	8.8	6.4	4.5	5.0	2.1	1.9	2.8	2.7	3.1	3.0	3.7	3.8	2.8	6.1	3.7	8.0	0.0	3.0
10	19.7	9.9	8.2	4.7	5.0	2.2	2.0	3.0	3.0	3.1	0.0	4.1	4.3	3.3	6.0	3.8	8.0	0.0	3.0
11	10.6	5.2	3.9	2.3	4.0	1.2	1.0	2.0	2.0	2.2	6.0	2.5	2.5	2.0	4.5	2.7	4.0	1.0	2.0
12	9.2	4.5	3.7	2.2	4.0	1.3	1.2	2.0	1.6	2.1	5.0	2.4	2.3	1.8	4.1	2.4	4.0	1.0	2.0
13	9.6	4.5	3.6	2.3	4.0	1.3	1.0	1.9	1.7	2.2	4.0	2.4	2.3	1.7	4.0	2.3	4.0	1.0	2.0
14	8.5	4.0	3.8	2.2	4.0	1.3	1.1	1.9	2.0	2.1	5.0	2.4	2.4	1.9	4.4	2.3	4.0	1.0	2.0
15	11.0	4.7	4.2	2.3	4.0	1.2	1.0	1.9	2.0	2.2	4.0	2.5	2.5	2.0	4.5	2.6	4.0	1.0	2.0
16	18.1	8.2	5.9	3.5	5.0	1.9	1.9	1.9	2.7	2.8	4.0	3.5	3.8	2.9	6.0	4.5	9.0	1.0	2.0
17	17.6	8.3	6.0	3.8	5.0	2.0	1.9	2.0	2.2	2.9	3.0	3.5	3.6	2.8	5.7	4.3	10.0	1.0	2.0
18	19.2	6.6	6.2	3.4	5.0	2.0	1.8	2.2	2.3	2.8	4.0	3.5	3.4	2.5	5.3	3.8	10.0	1.0	2.0
19	15.4	7.6	7.1	3.4	5.0	2.0	1.9	2.5	2.5	2.9	4.0	3.3	3.6	2.7	6.0	4.2	8.0	1.0	3.0
20	15.1	7.3	6.2	3.8	5.0	2.0	1.8	2.1	2.4	2.5	4.0	3.7	3.7	2.8	6.4	4.3	10.0	1.0	2.5
21	16.1	7.9	5.8	3.7	5.0	2.1	1.9	2.3	2.6	2.9	5.0	3.6	3.6	2.7	6.0	4.5	0.0	1.0	2.0
22	19.1	8.8	6.4	3.9	5.0	2.2	2.0	2.3	2.4	2.9	4.0	3.8	4.0	3.0	6.5	4.5	0.0	1.0	2.5
23	15.3	6.4	5.3	3.3	5.0	1.7	1.6	2.0	2.2	2.5	5.0	3.4	3.4	2.6	5.4	4.0	0.0	1.0	2.0

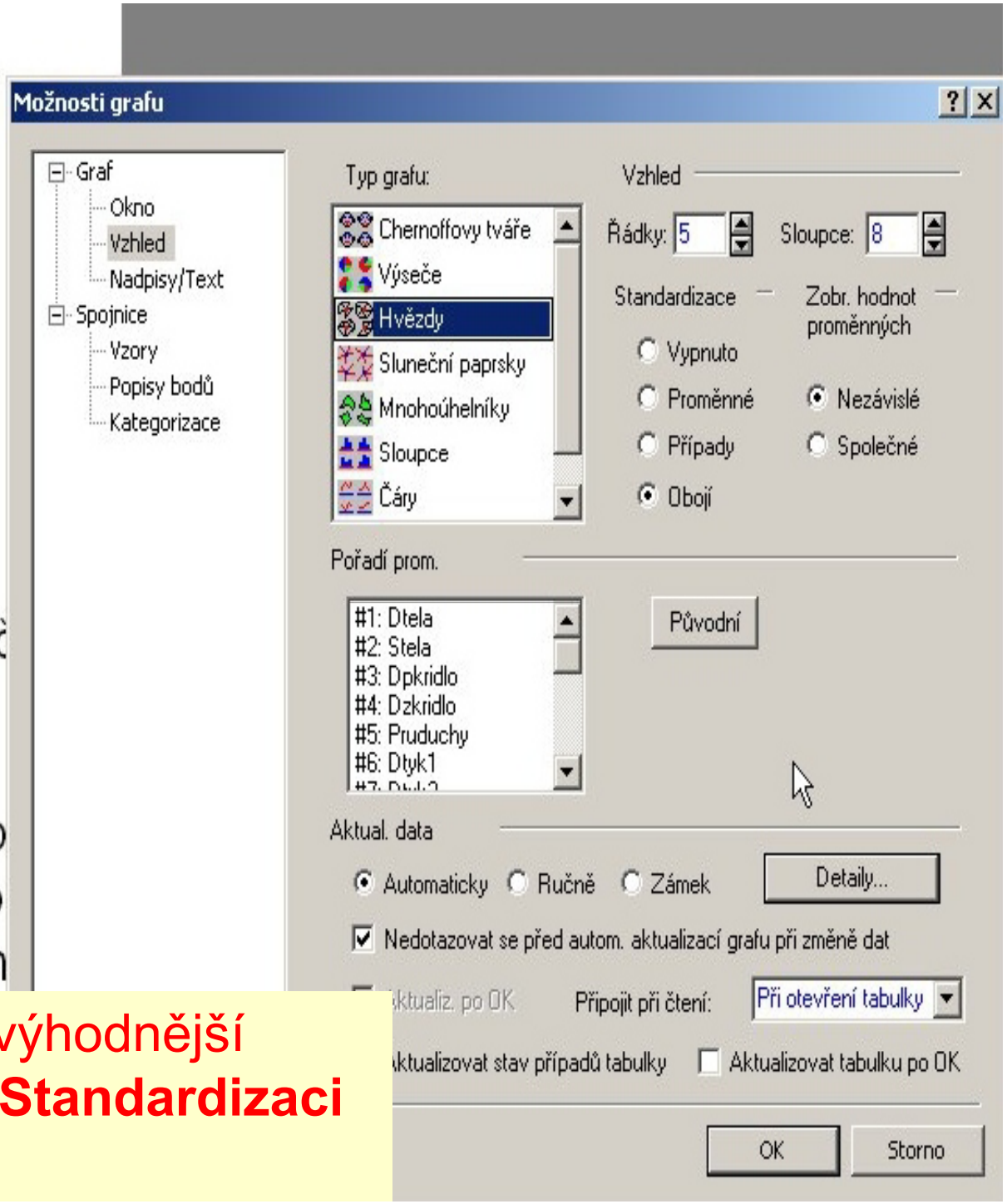
24	14.8	8.1	6.2	3.7	5.0	2.2	2.0	2.2	2.4	3.2	5.0	3.5	3.7	2.7	6.0	4.1	0.0	1.0	2.0
25	16.2	7.7	6.9	3.7	5.0	2.0	1.8	2.3	2.4	2.8	4.0	3.8	3.7	2.7	5.7	4.2	0.0	1.0	2.5
26	13.4	6.9	5.7	3.4	5.0	2.0	1.8	2.8	2.0	2.6	4.0	3.6	3.6	2.6	5.5	3.9	0.0	1.0	2.0
27	12.9	5.8	4.8	2.6	5.0	1.6	1.5	1.9	2.1	2.6	5.0	2.8	3.0	2.2	5.1	3.6	9.0	1.0	3.0
28	12.0	6.5	5.3	3.2	5.0	1.9	1.9	2.3	2.5	3.0	5.0	3.3	3.5	2.6	5.4	4.3	8.0	1.0	2.0
29	14.1	7.0	5.5	3.6	5.0	2.2	2.0	2.3	2.5	3.1	5.0	3.6	3.7	2.8	5.8	4.1	0.0	1.0	2.0
30	16.7	7.2	5.7	3.5	5.0	1.9	1.9	2.5	2.3	2.8	5.0	3.4	3.6	2.7	6.0	4.0	0.0	1.0	2.5
31	14.1	5.4	5.0	3.0	5.0	1.7	1.6	1.8	2.5	2.4	5.0	2.7	2.9	2.2	5.3	3.6	8.0	1.0	2.0
32	10.0	6.0	4.2	2.5	5.0	1.6	1.4	1.4	2.0	2.7	6.0	2.8	2.5	1.8	4.8	3.4	8.0	1.0	2.0
33	11.4	4.5	4.4	2.7	5.0	1.8	1.5	1.9	1.7	2.5	5.0	2.7	2.5	1.9	4.7	3.7	8.0	1.0	2.0
34	12.5	5.5	4.7	2.3	5.0	1.8	1.4	1.8	2.2	2.4	4.0	2.8	2.6	2.0	5.1	3.7	8.0	0.0	2.0
35	13.0	5.3	4.7	2.3	5.0	1.6	1.4	1.8	1.8	2.5	4.0	2.7	2.7	2.1	5.0	3.6	8.0	1.0	2.0
36	12.4	5.2	4.4	2.6	5.0	1.6	1.4	1.8	2.2	2.2	5.0	2.7	2.5	2.0	5.0	3.2	6.0	1.0	2.0
37	12.0	5.4	4.9	3.0	5.0	1.7	1.5	1.7	1.9	2.4	5.0	2.7	2.7	2.0	4.2	3.7	6.0	1.0	2.0
38	10.7	5.6	4.5	2.8	5.0	1.8	1.4	1.8	2.2	2.4	4.0	2.7	2.6	2.0	5.0	3.5	8.0	1.0	2.0
39	11.1	5.5	4.3	2.6	5.0	1.7	1.5	1.8	1.9	2.4	5.0	2.6	2.5	1.9	4.6	3.4	8.0	1.0	2.0
40	12.8	5.7	4.8	2.8	5.0	1.6	1.4	1.7	1.9	2.3	5.0	2.3	2.5	1.9	5.0	3.1	8.0	1.0	2.0

V další úloze načtete zdrojovou matici dat **Mšice,** pak zadáte zatím všechny nabídnuté **Proměnné** a pak **Typ grafu.**

V další analýze pak zadáte menší počet proměnných (max. 8), aby ikony byly k vizuální analýze přehlednější.

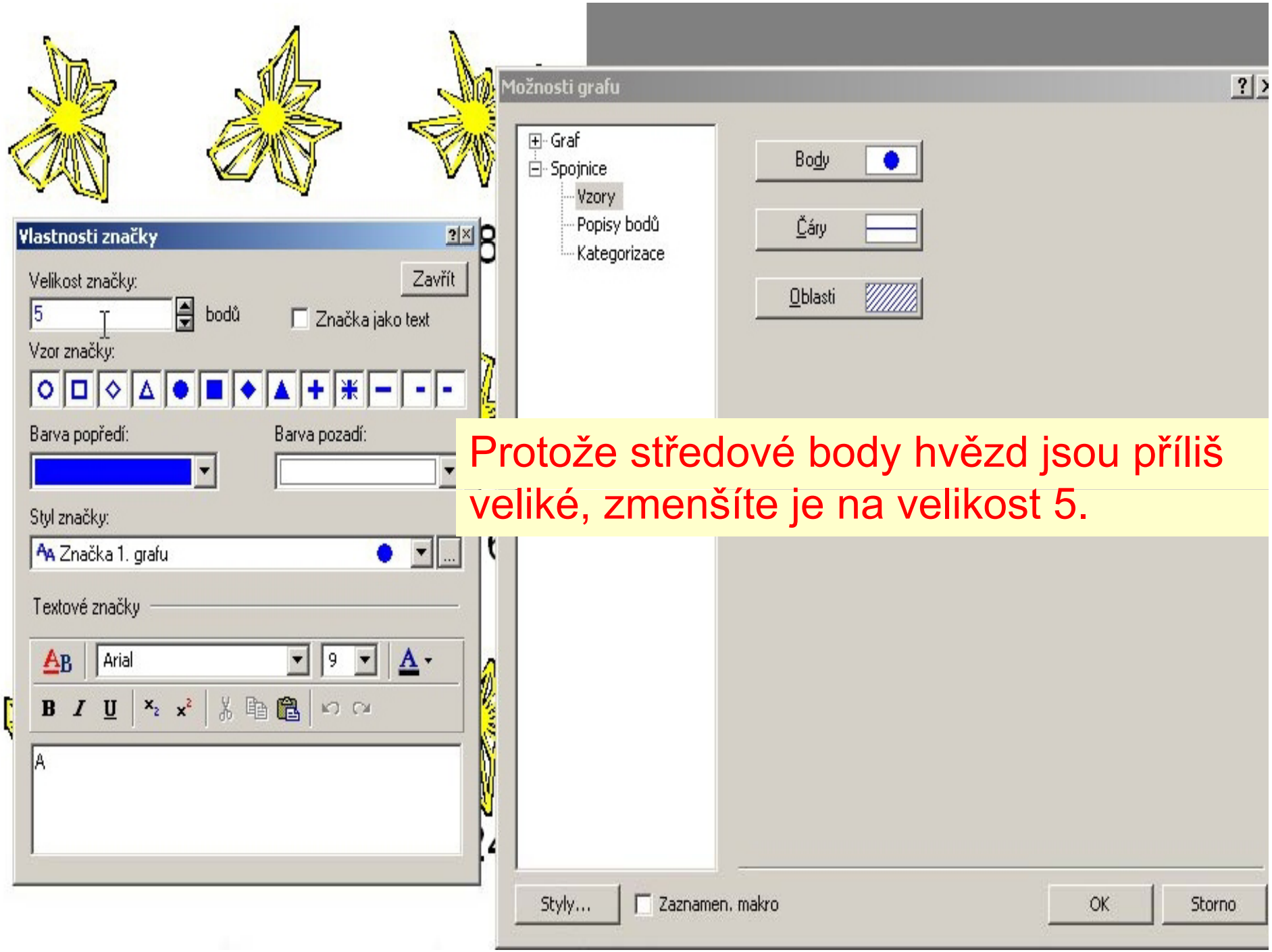






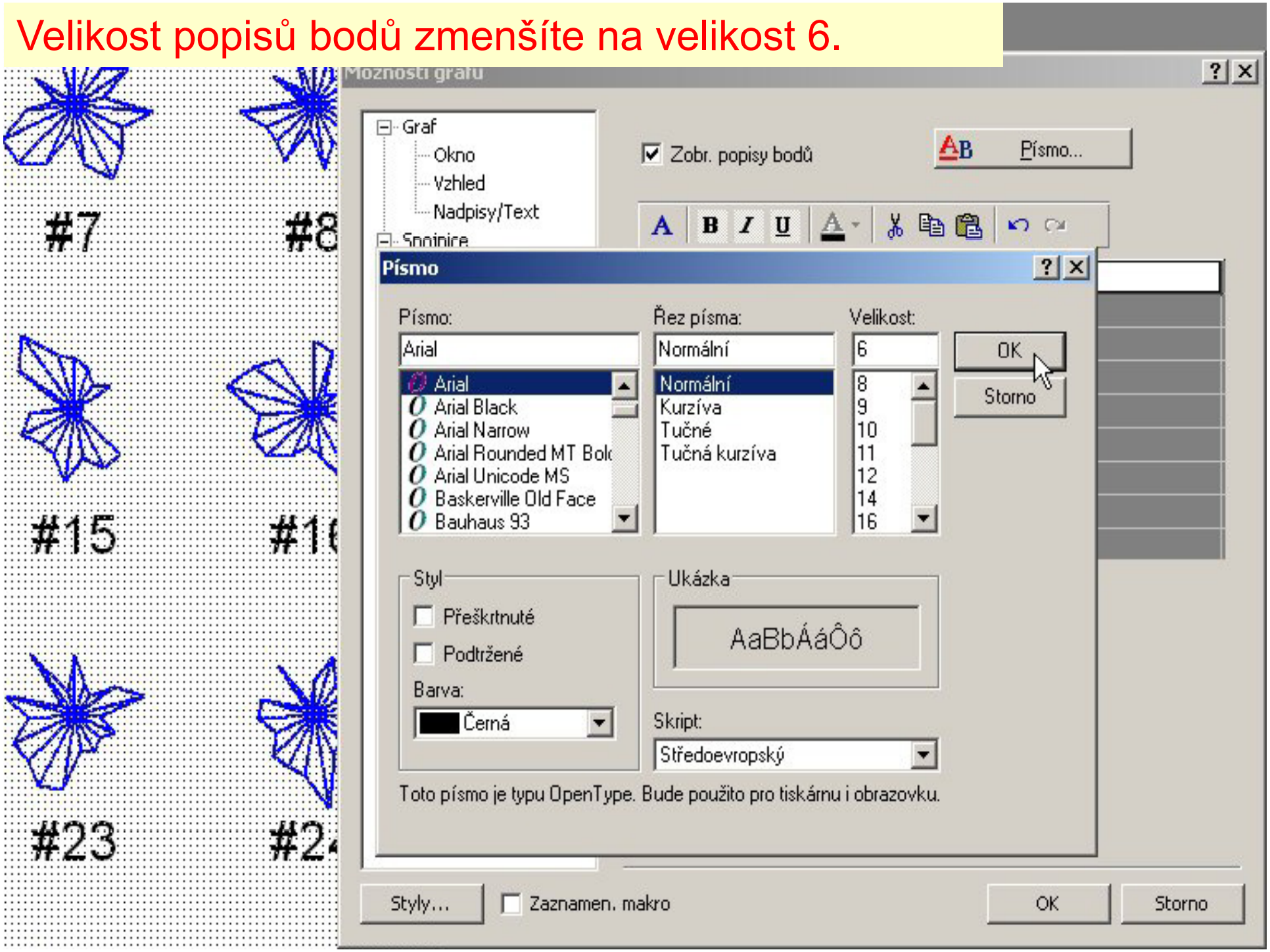
Zvolíte za **Typ grafu** nejvýhodnější ikonu, a to **Hvězdy** a ve **Standardizaci** zvolíte **Obojí.** Pak **OK.**

Protože jsou popisy bodů příliš veliké, zmenšíte font na 6: Zablokujete (černě) všechny popisy na ploše grafu, pak 2x kliknete do plochy grafu a zvolíte **Nadpisy/text.** Zvolíte font 6 a pak **OK.**

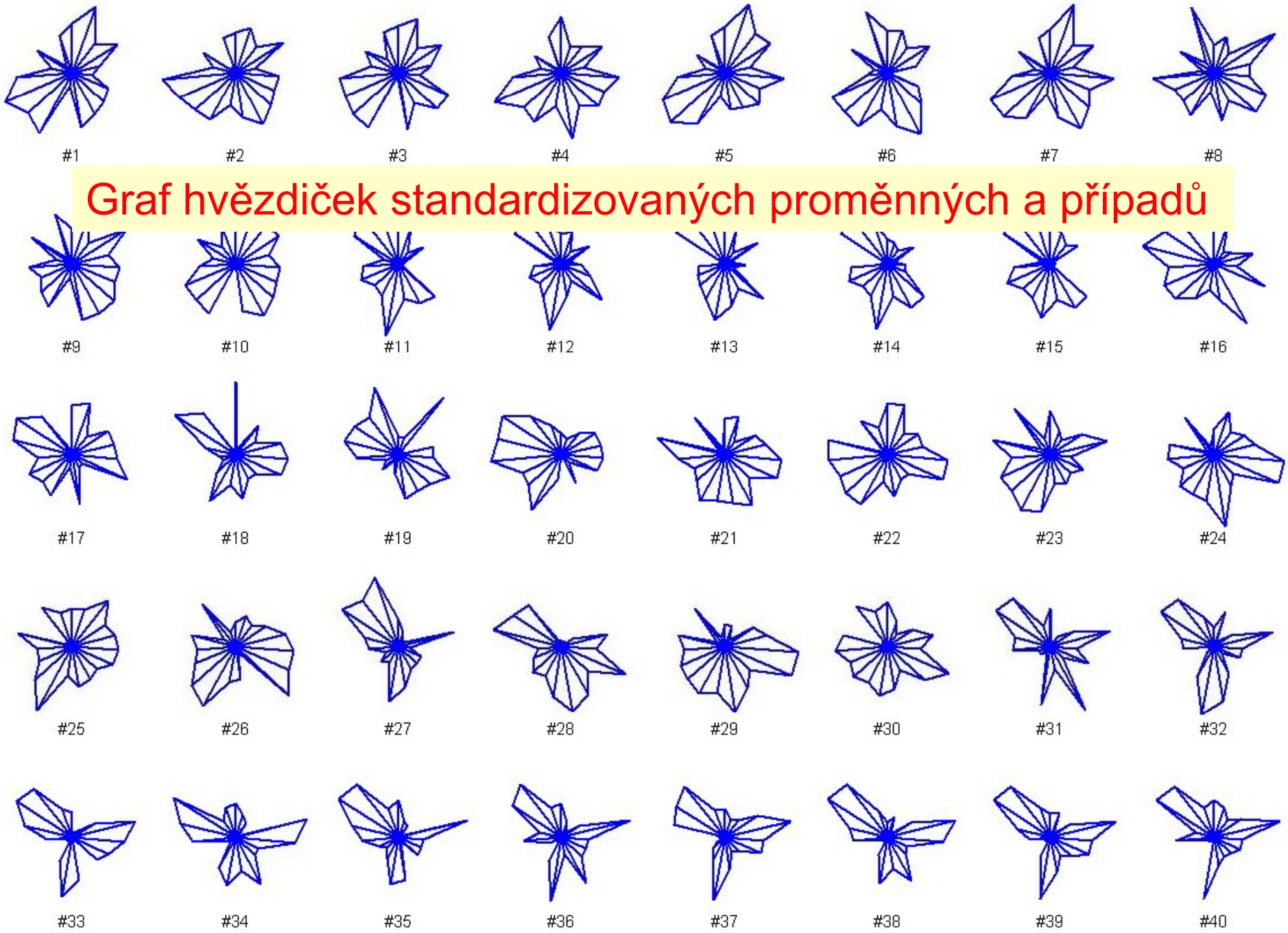


Protože středové body hvězd jsou příliš veliké, zmenšíte je na velikost 5.

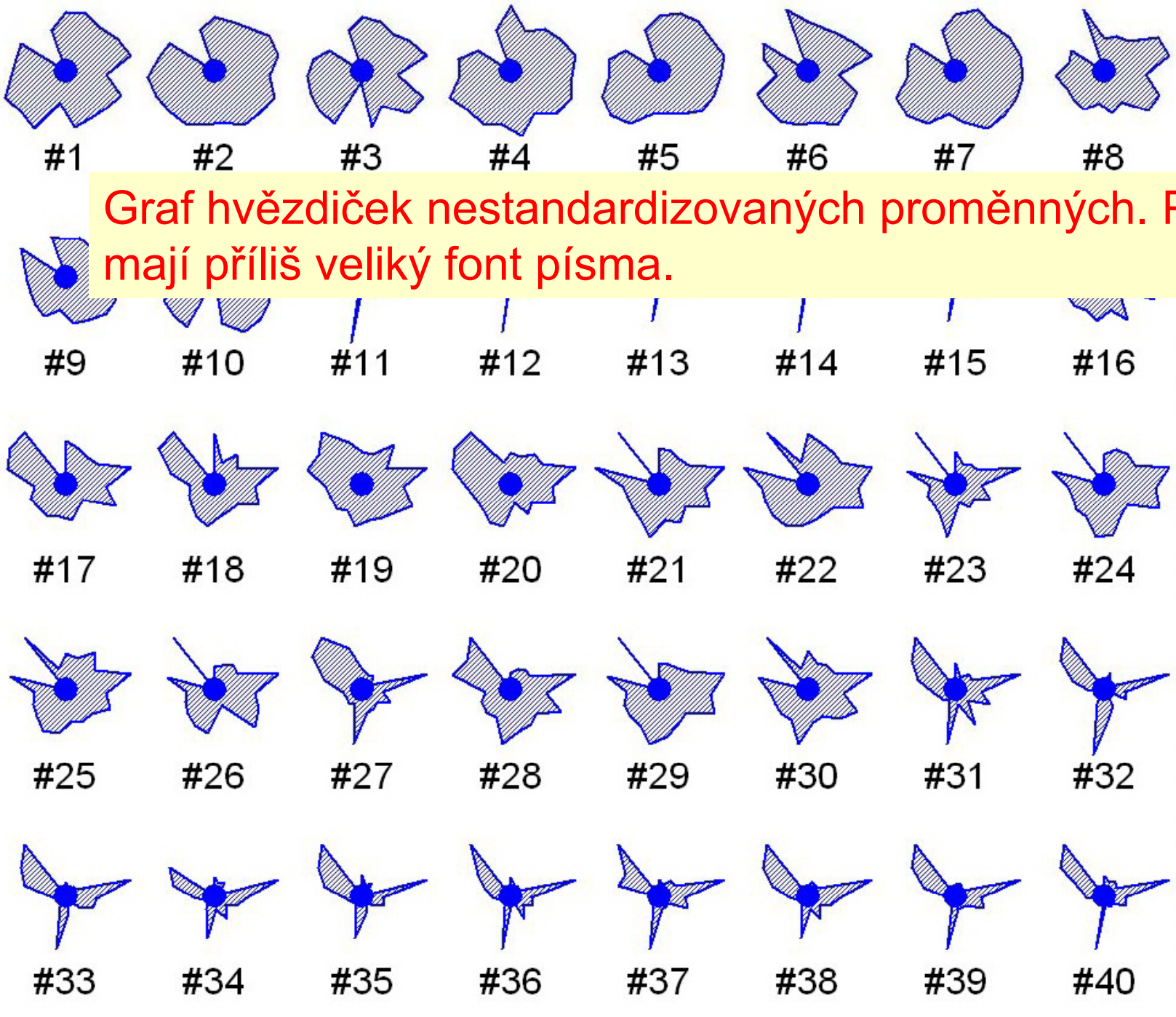
Velikost popisů bodů zmenšíte na velikost 6.



Graf hvězdiček standardizovaných proměnných a případů

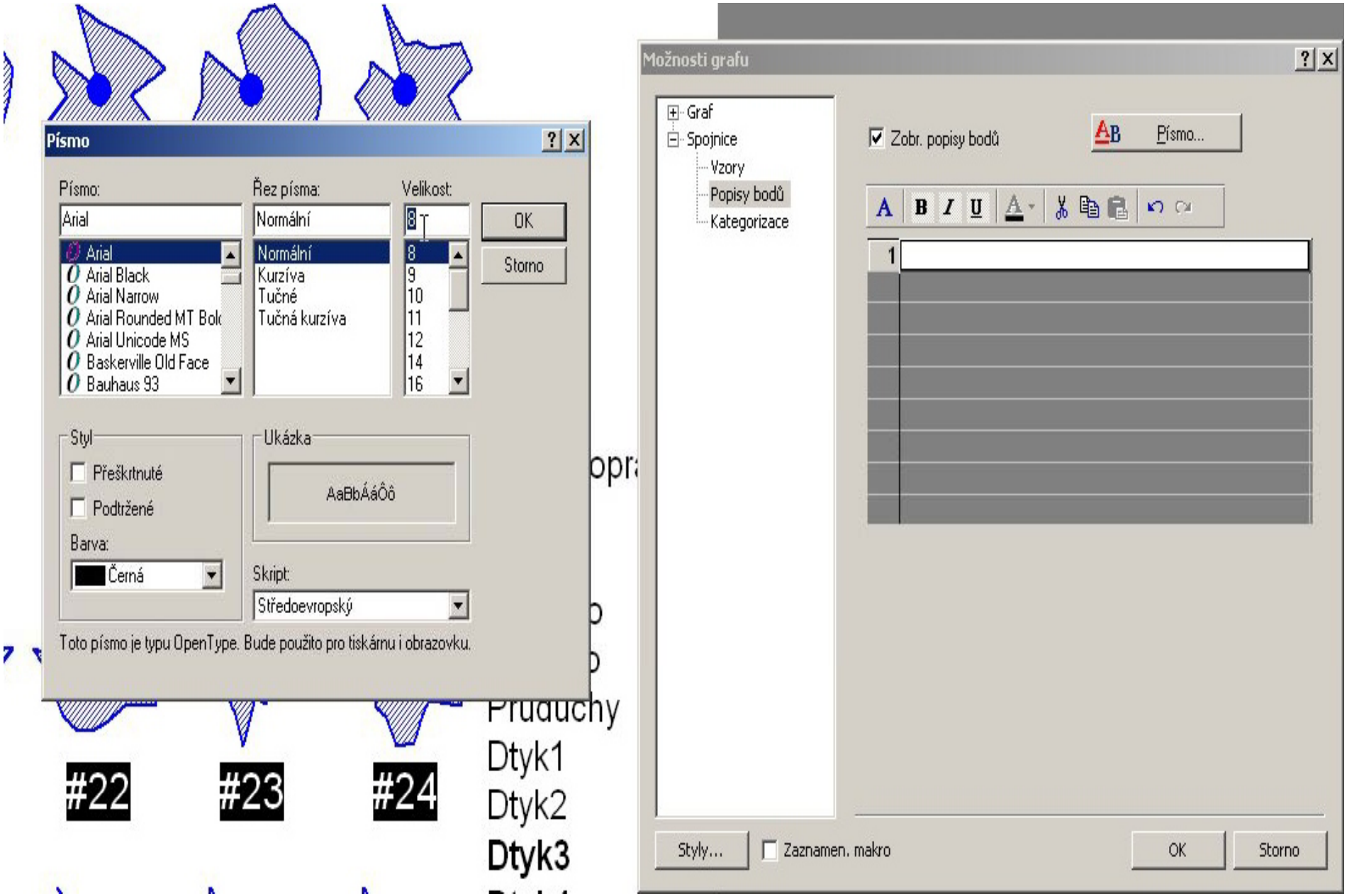


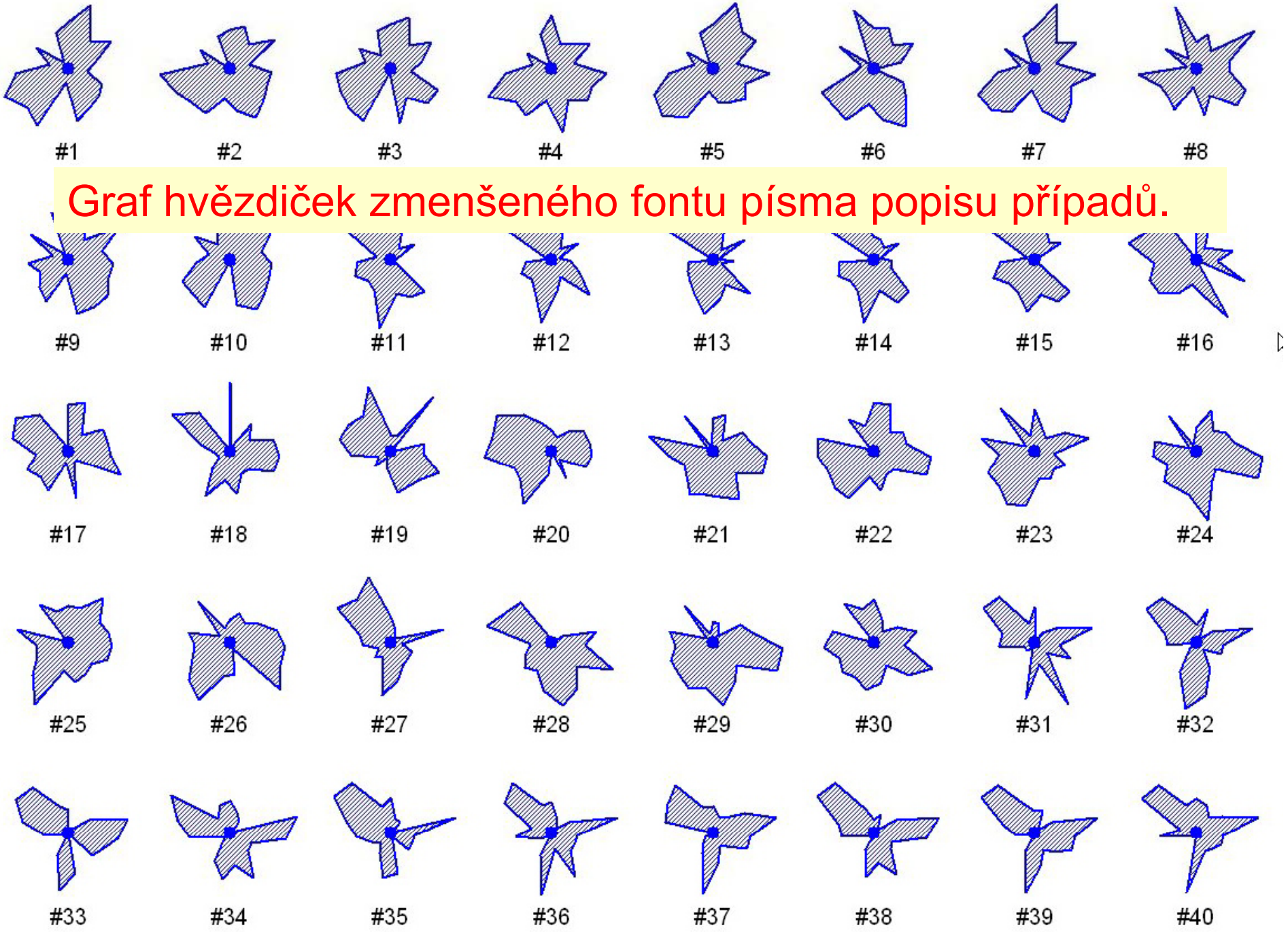
Graf hvězdiček nestandardizovaných proměnných. Popisy mají příliš veliký font písma.



Zleva doprava:
Dtela
Stela
Dpkridlo
Dzkridlo
Pruduchy
Dtyk1
Dtyk2
Dtyk3
Dtyk4
Dtyk5
Ptykostn
Dpclnohy
Tibium
Dstehna
Sosak
Kladelko
Ptrnu
Ritotvor
Phacku

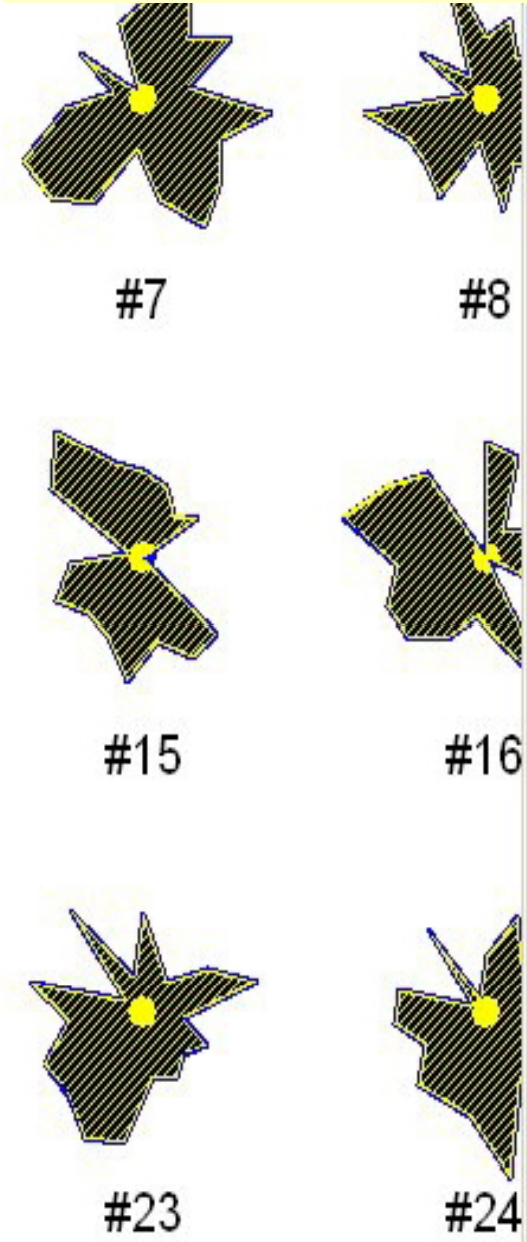
Zmenšení fontu písma popisů případů na velikost 8.

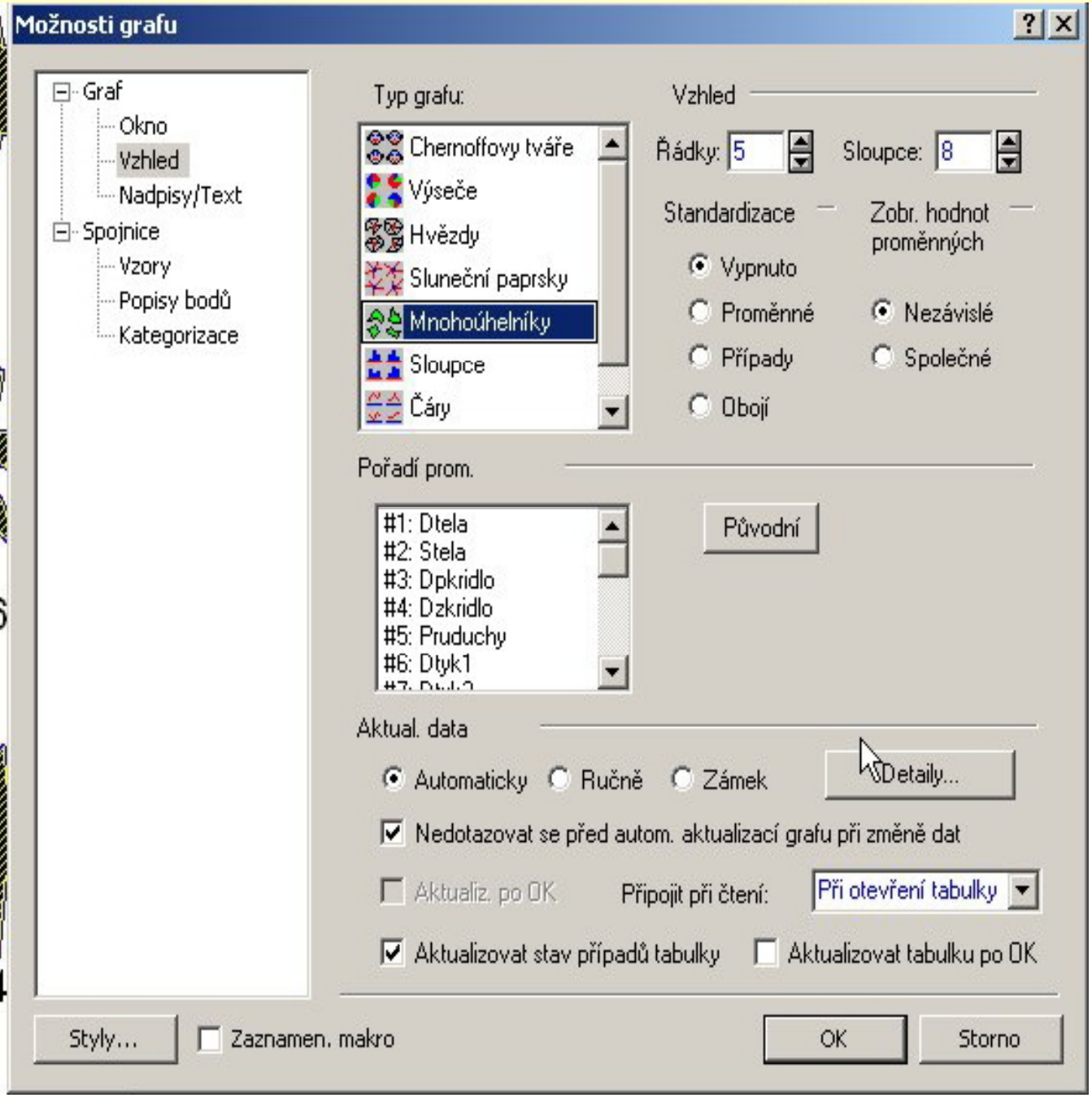




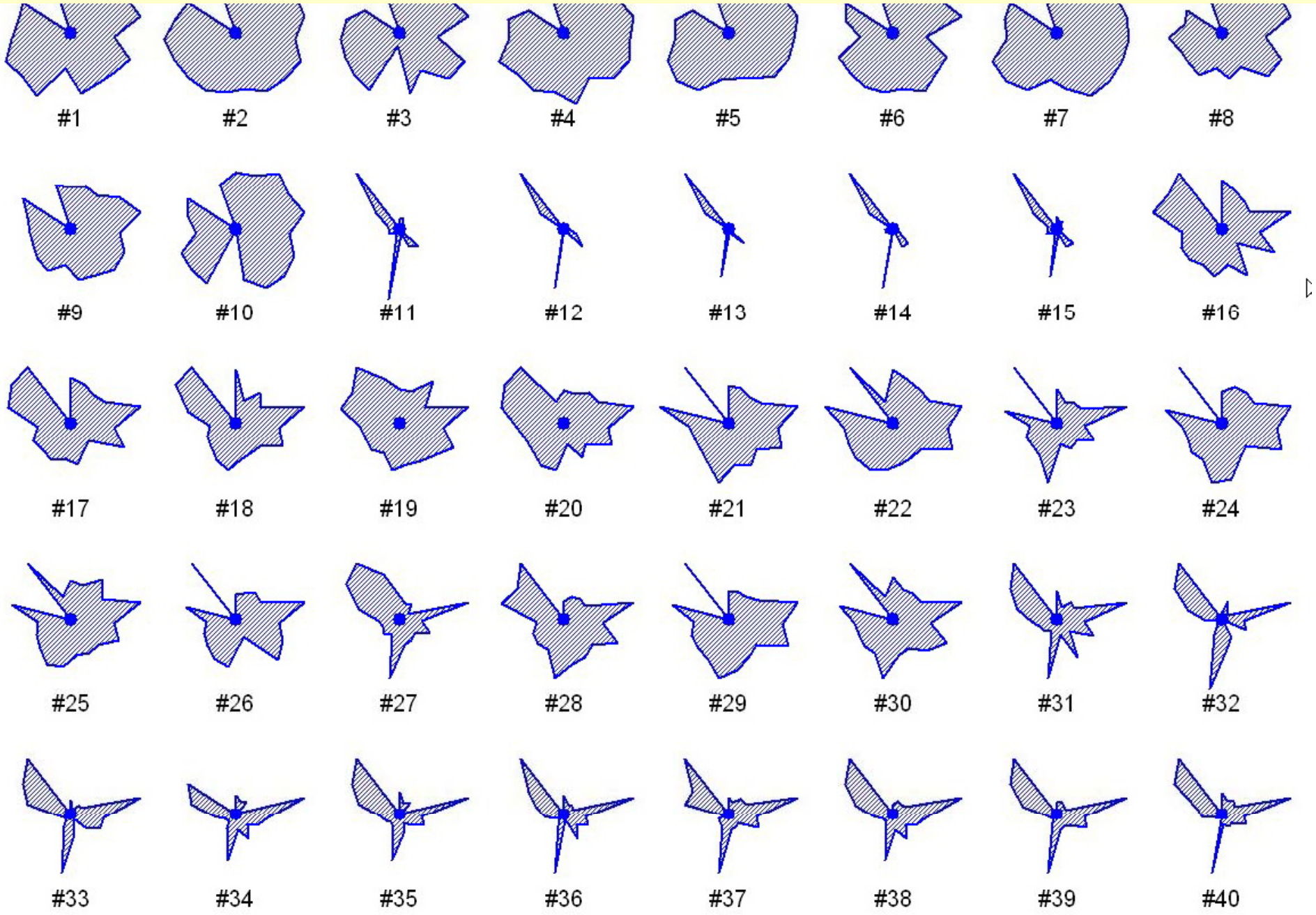
Graf hvězdiček zmenšeného fontu písma popisu případů.

Převod z ikony **Hvězdy** na ikonu **Mnohoúhelník**.

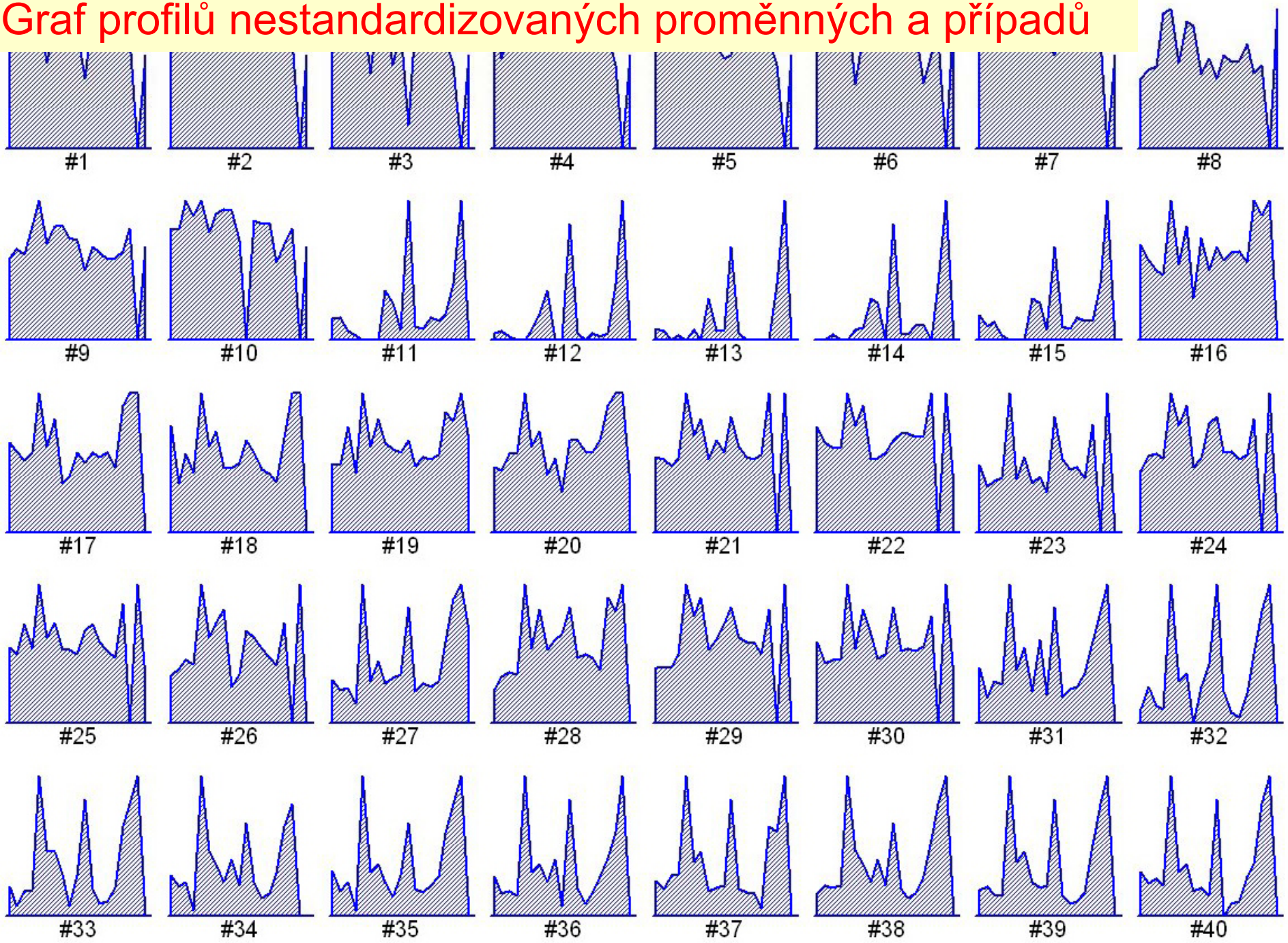




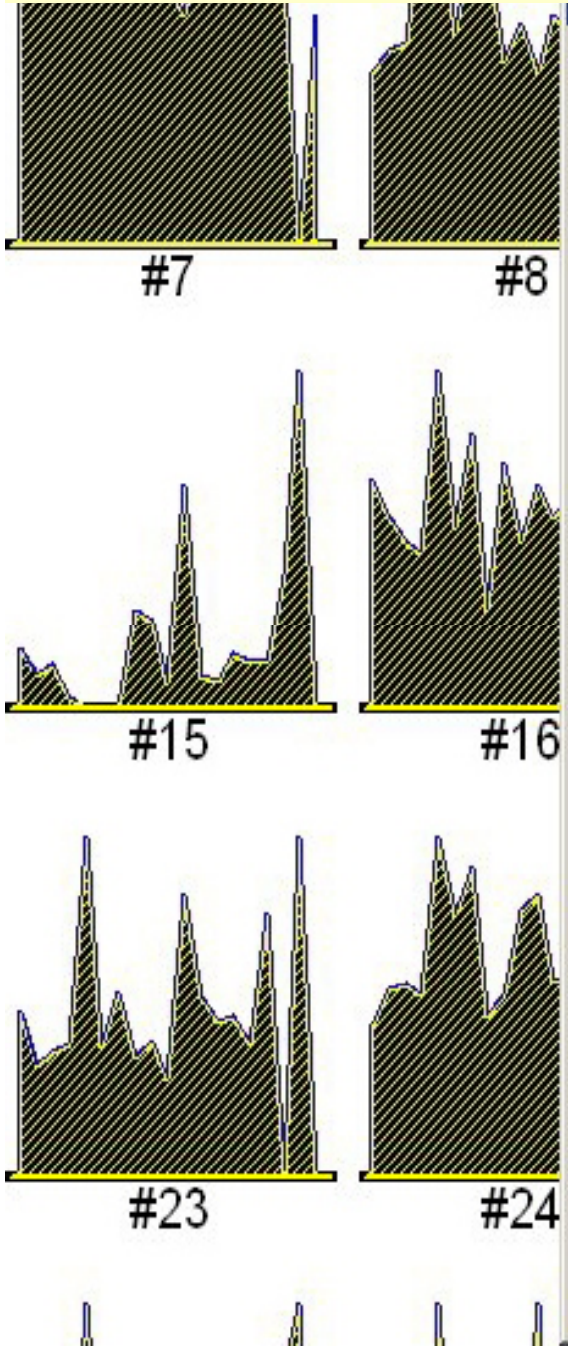
Graf mnohoúhelníků nestandardizovaných proměnných a případů

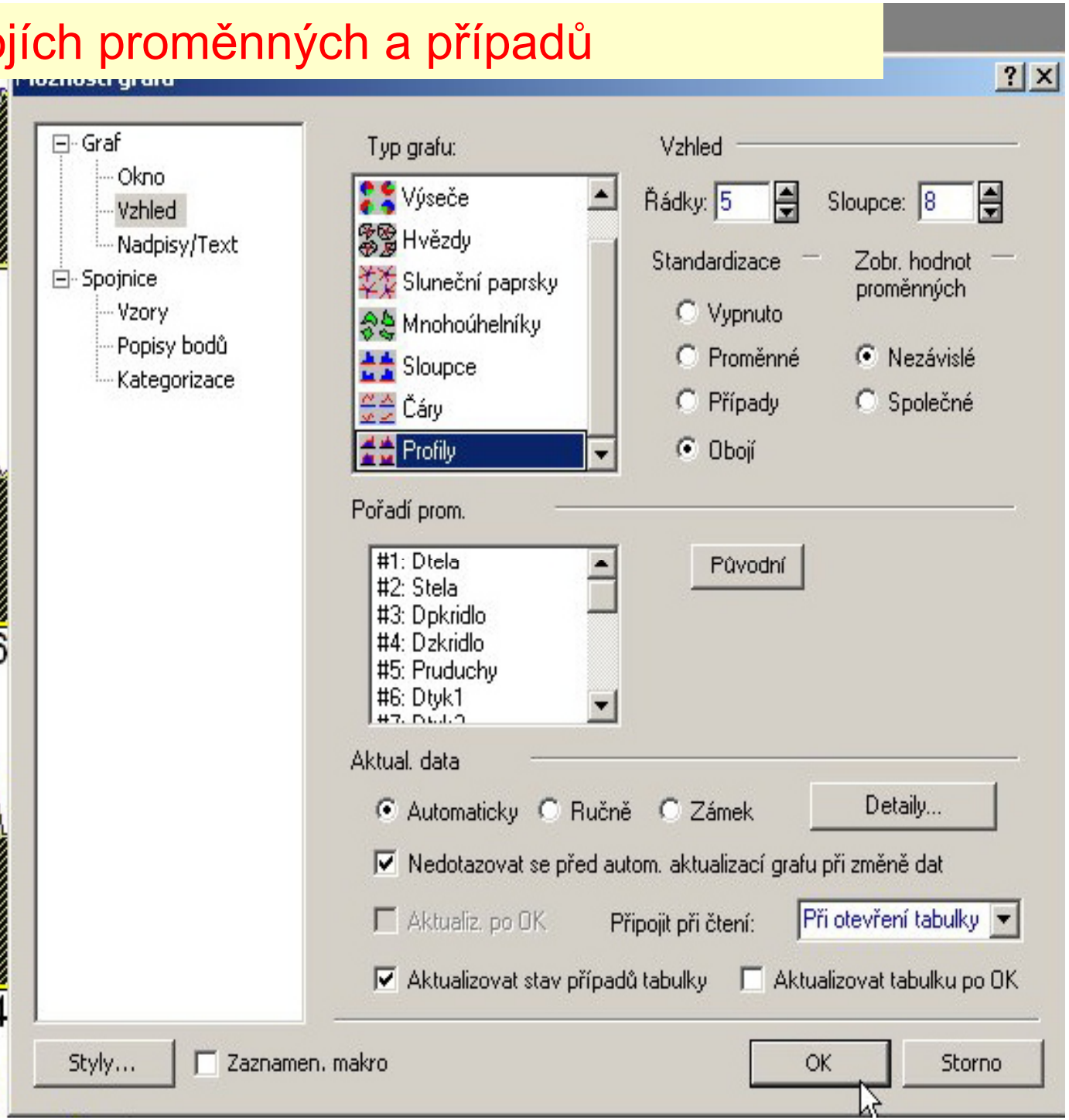


Graf profilů nestandardizovaných proměnných a případů

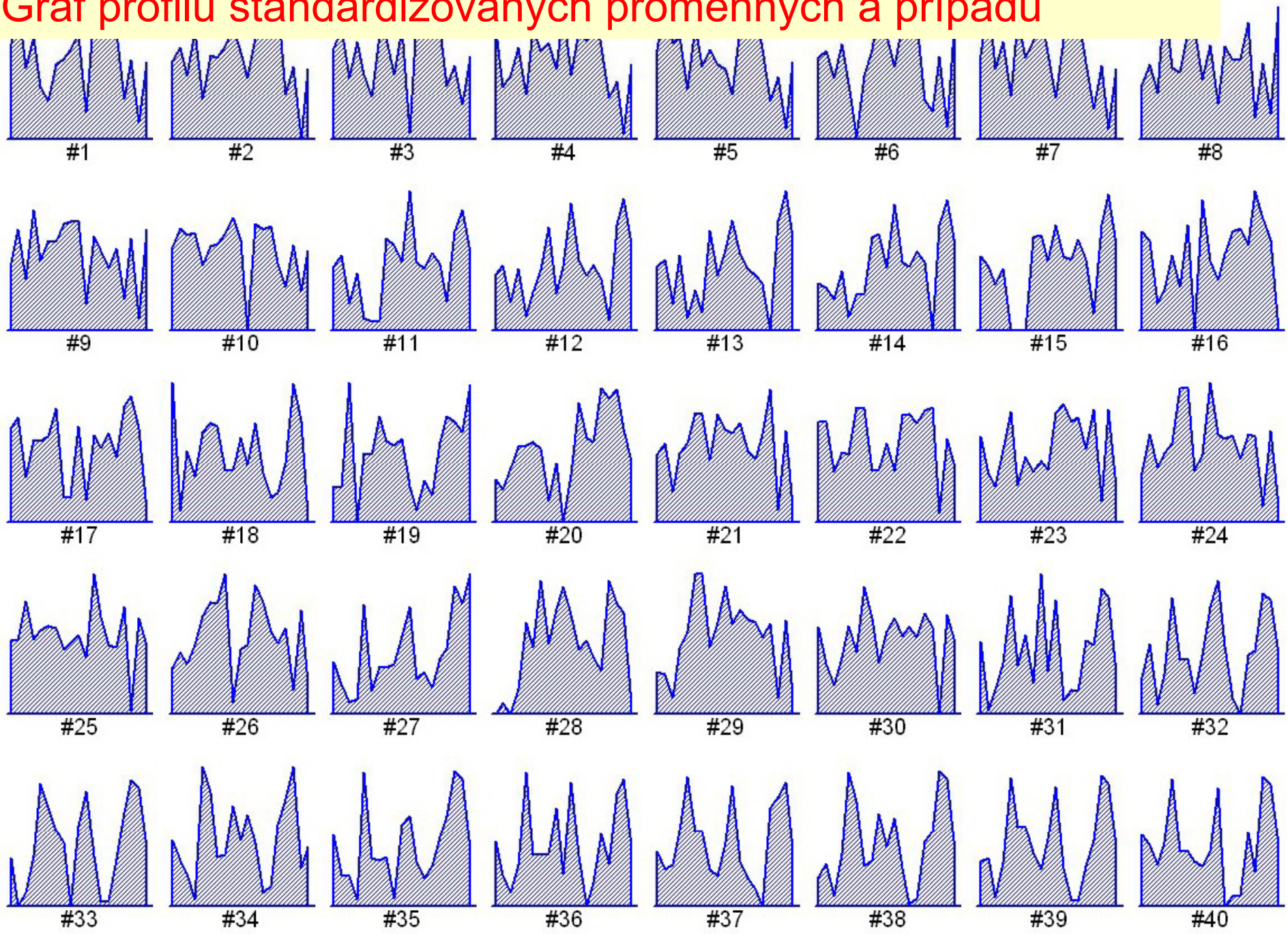


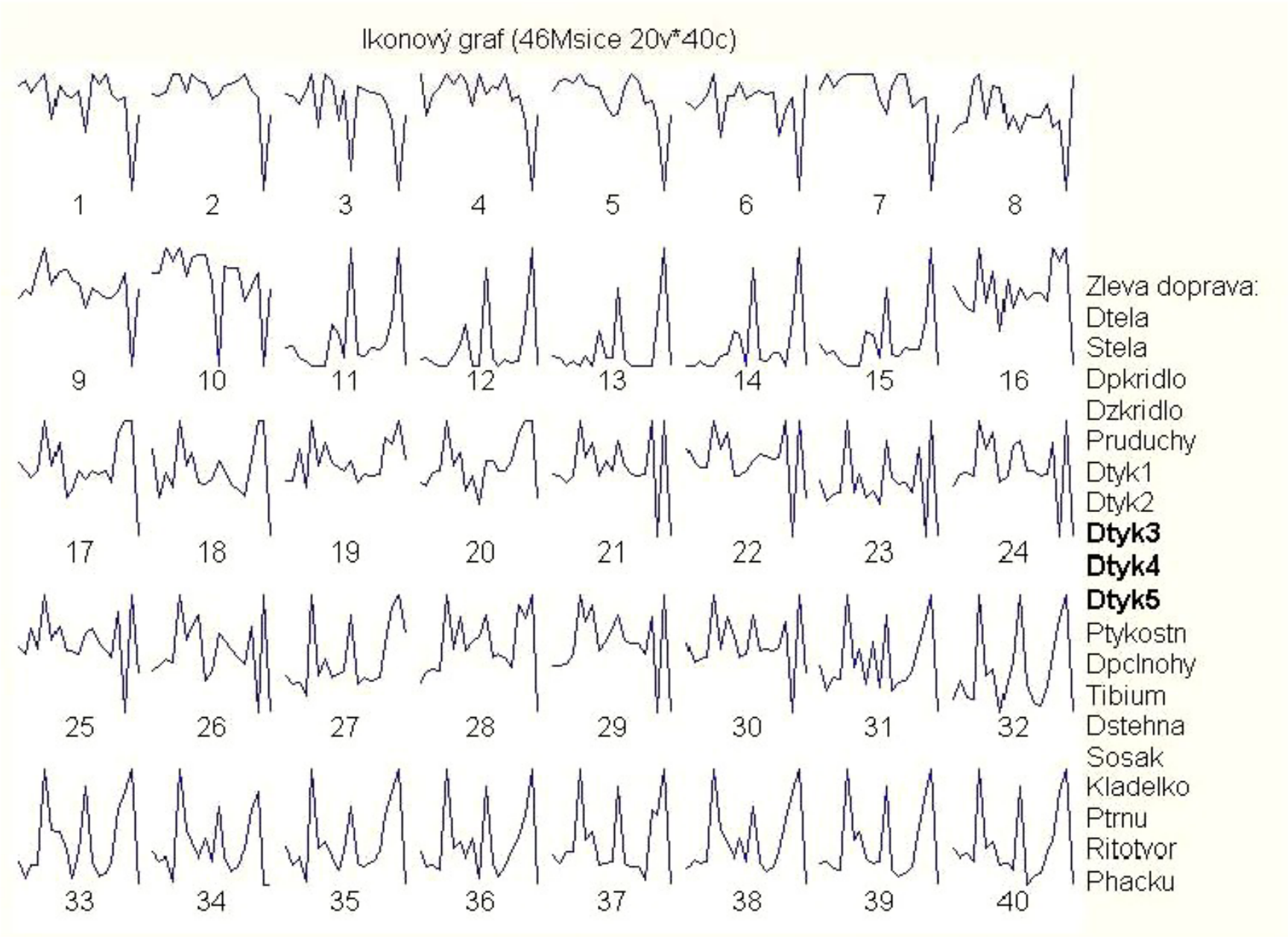
Standardizace obojích proměnných a případů



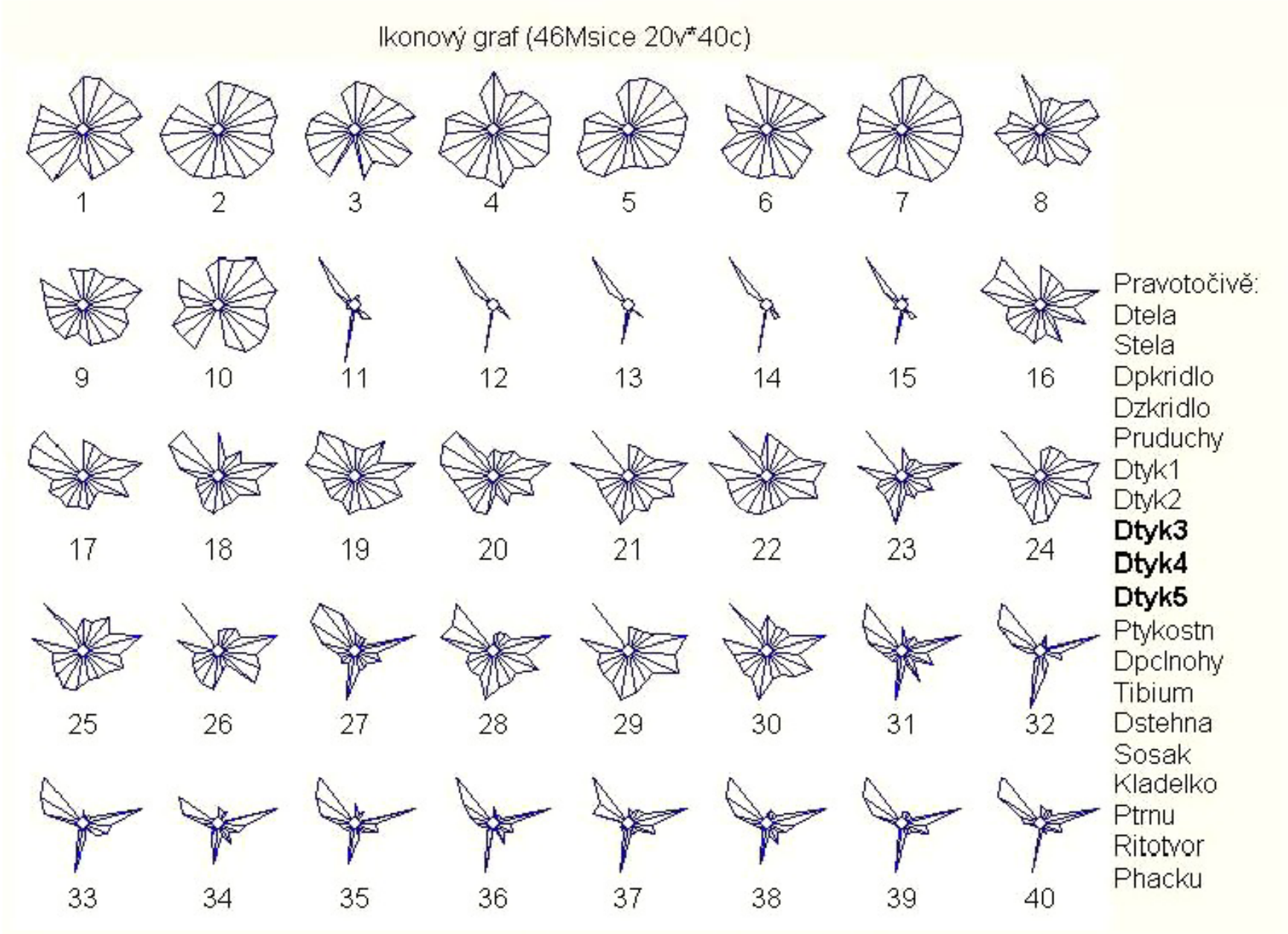


Graf profilů standardizovaných proměnných a případů

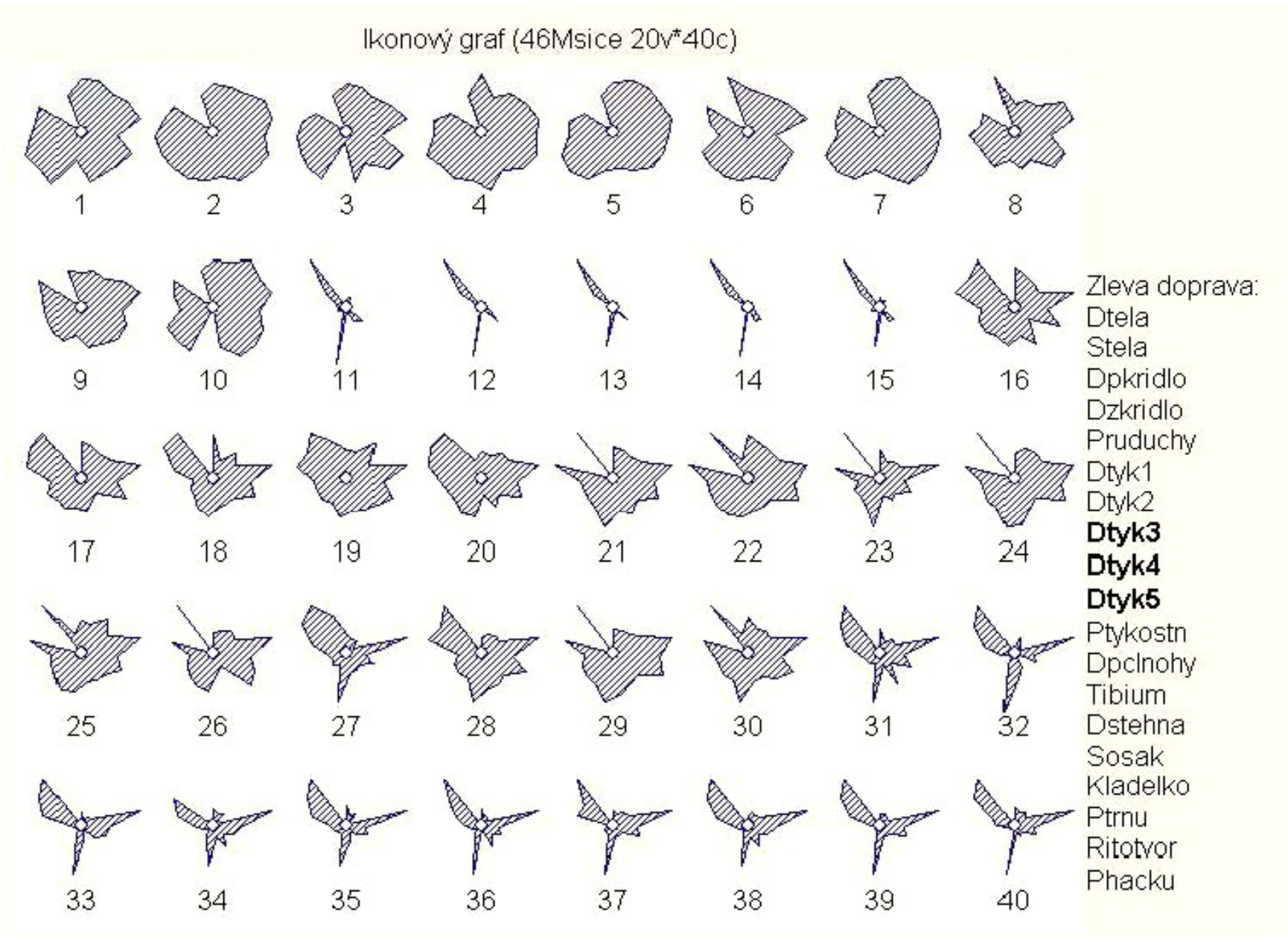




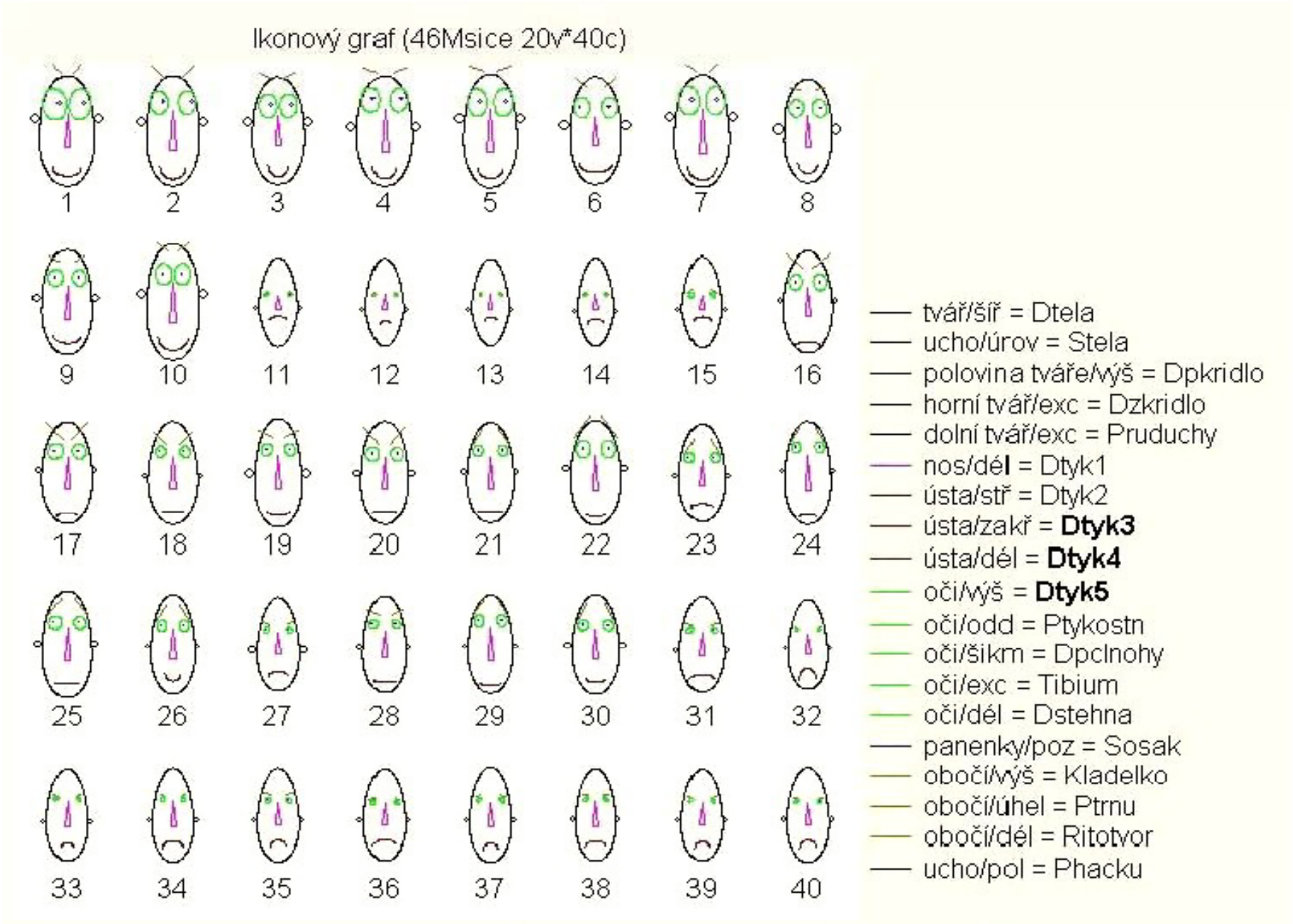
Čáry znaků pro každý objekt v EDA dat *Mšice,* (STATISTICA).



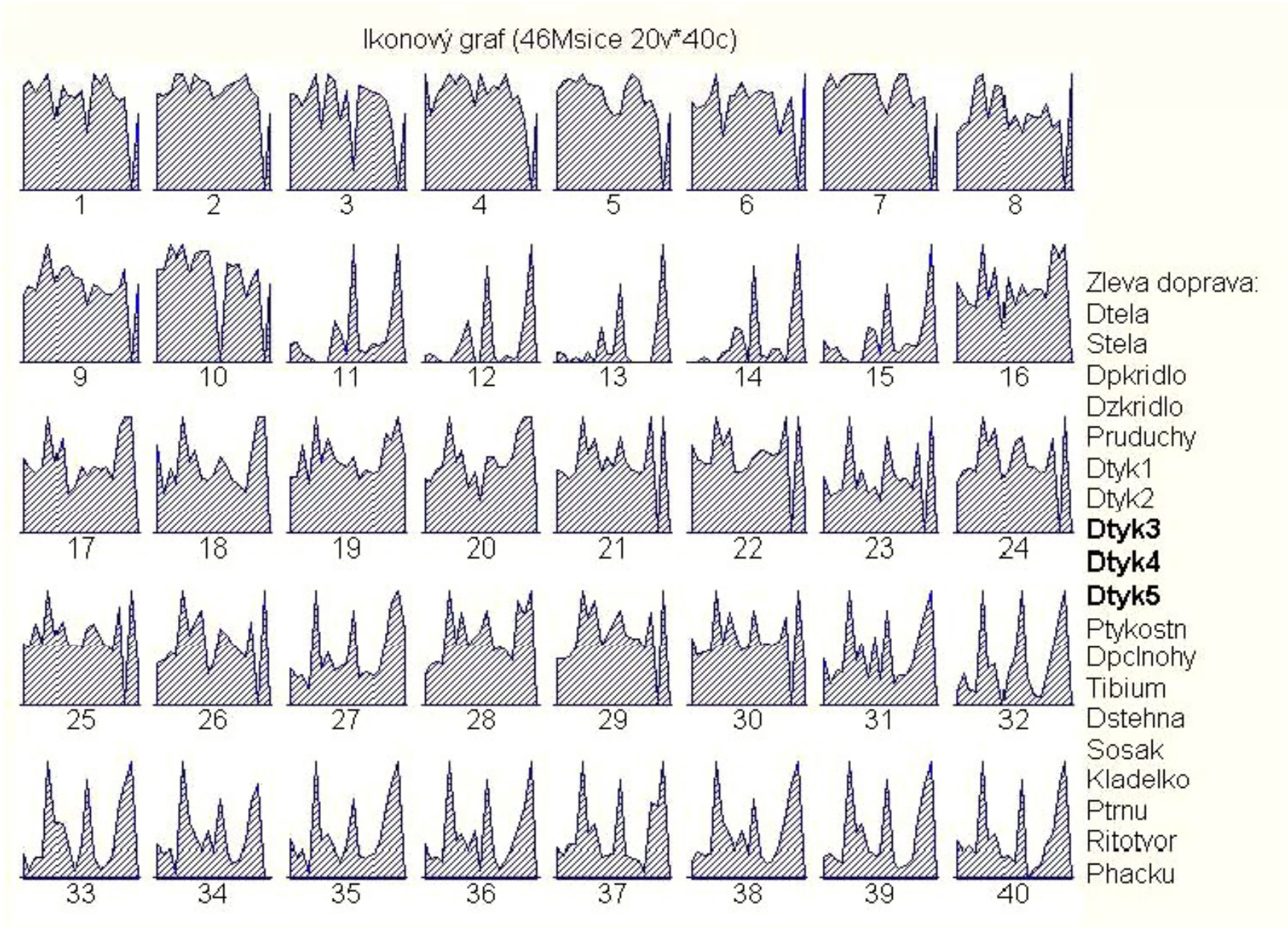
Hvězdy znaků pro každý objekt v EDA dat *Mšice,* (STATISTICA).



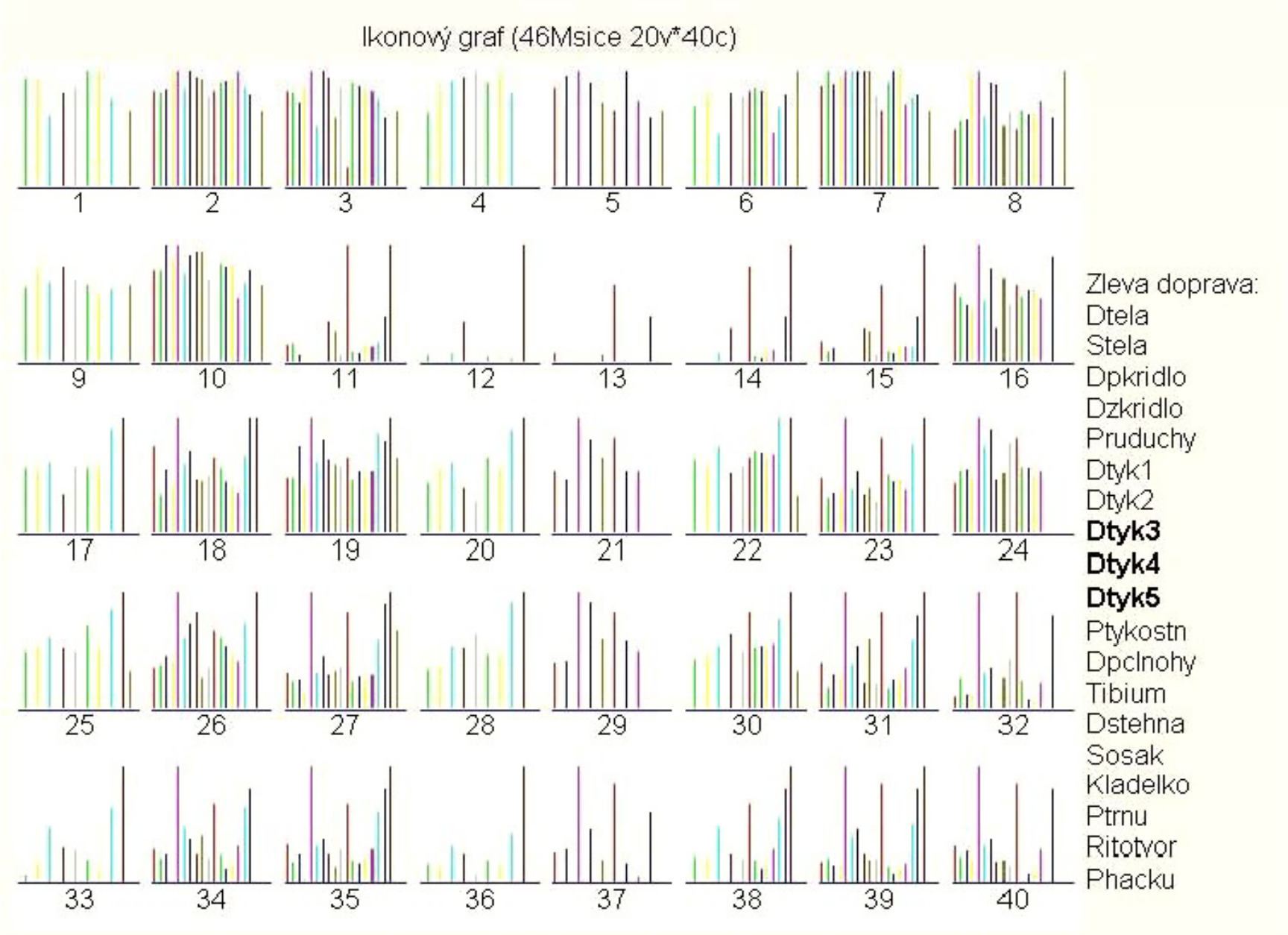
Mnohoúhelníky znaků pro každý objekt v EDA dat *Mšice*, (STATISTICA).



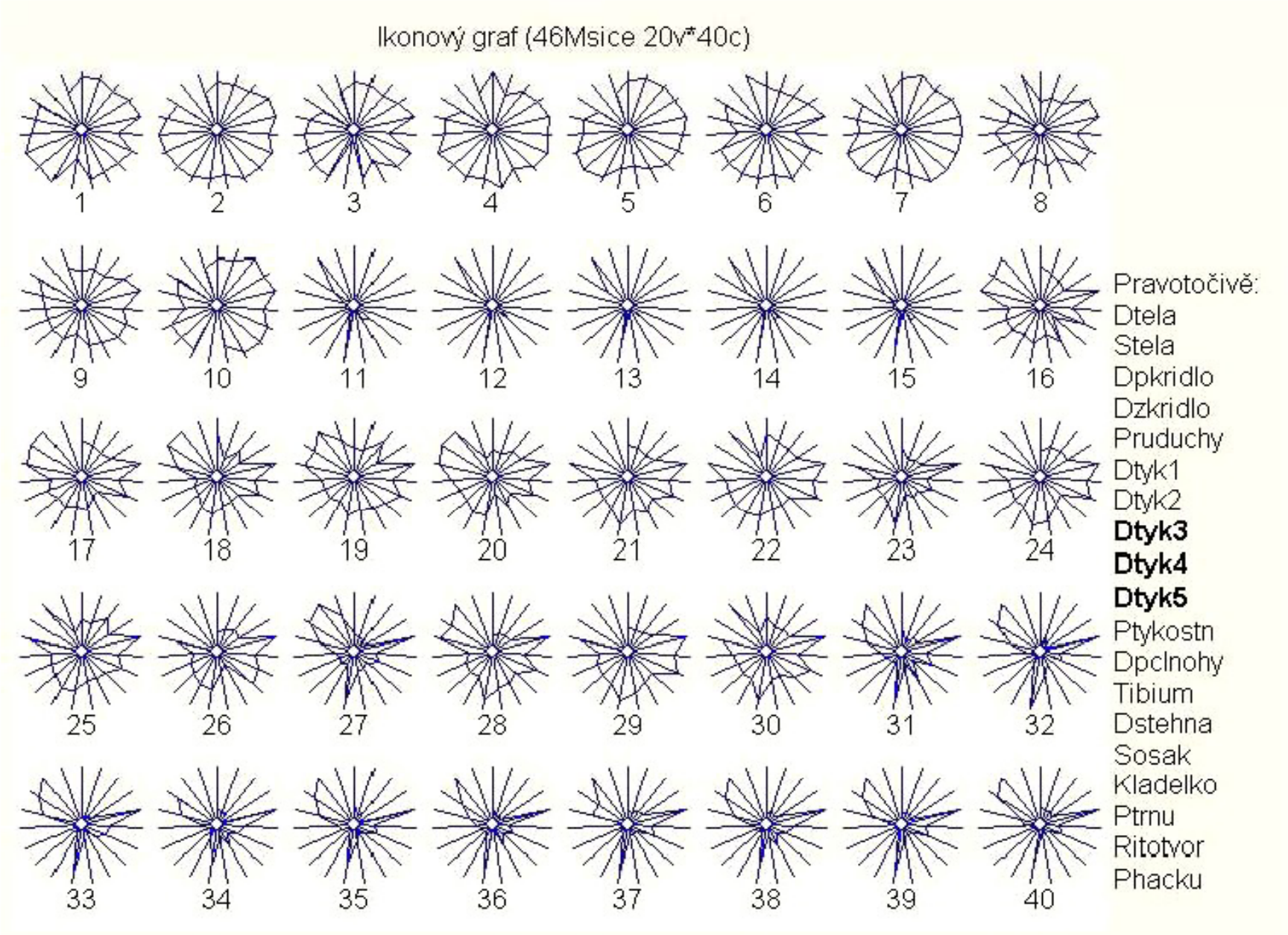
Obličeje znaků pro každý objekt v EDA dat *Mšice*,(STATISTICA).



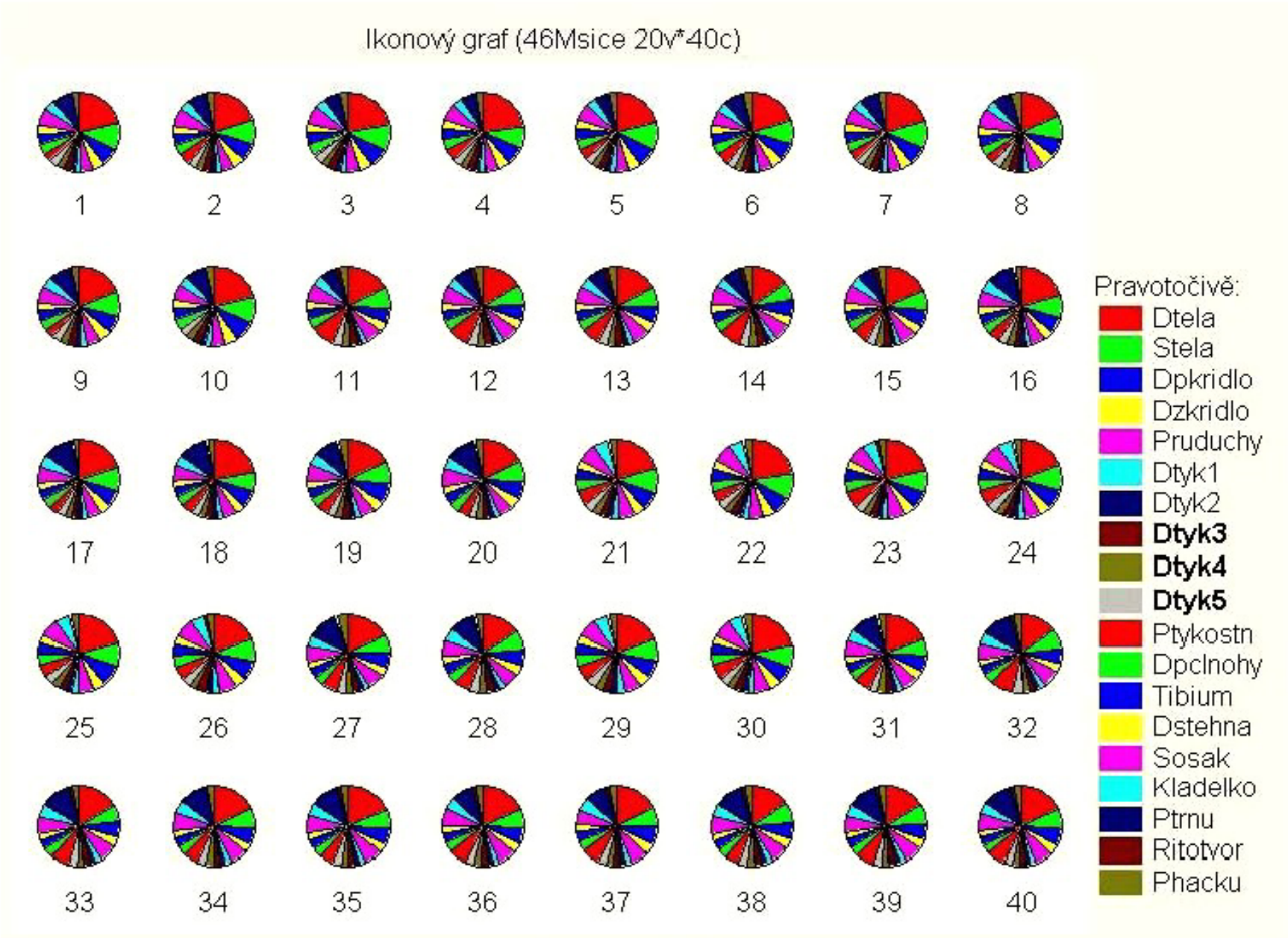
Profily znaků pro každý objekt v průzkumové analýze dat *Mšice,* (STATISTICA).



Sloupce znaků pro každý objekt v EDA dat *Mšice*, (STATISTICA).



Sluníčka znaků pro každý objekt v EDA dat *Mšice,* (STATISTICA).



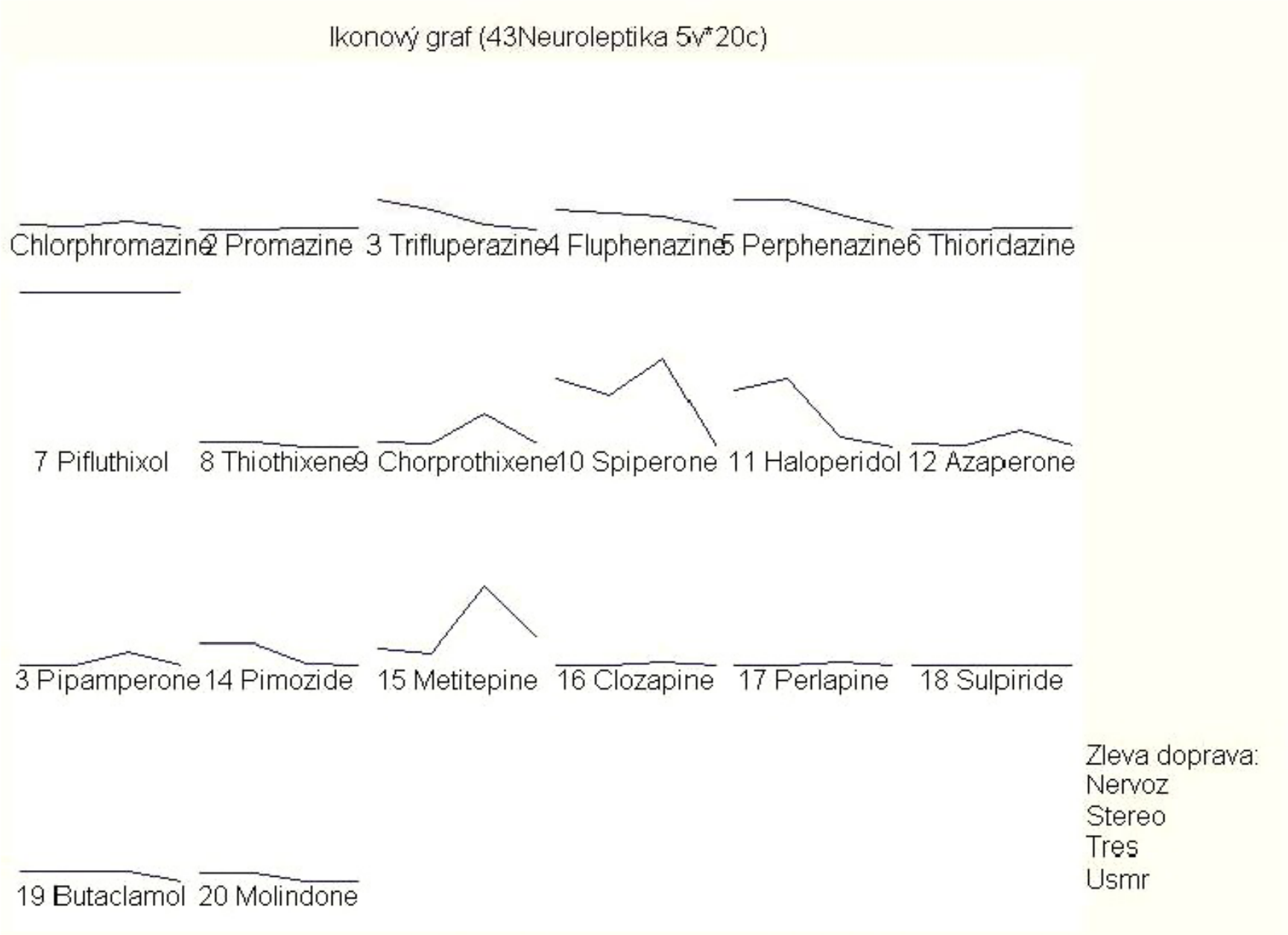
Výseče znaků pro každý objekt v EDA dat *Mšice*, (STATISTICA).

PŘÍKLAD 2.5 *Vzájemná nahraditelnost neuroleptik v diagramu komponentního skóre*

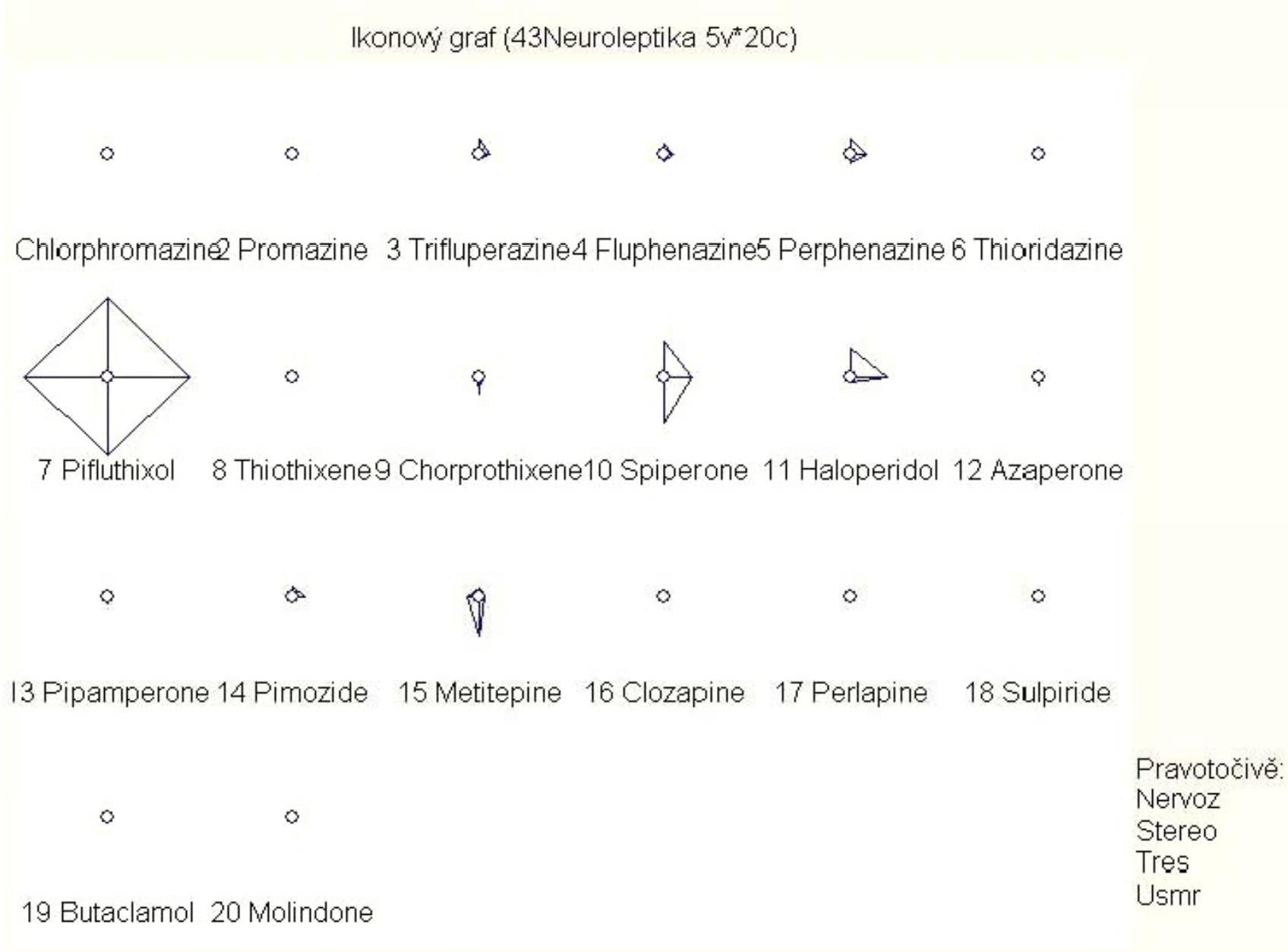
Je třeba ukázat základní pomůcky vícerozměrné analýzy dat a ukázat, která neuroleptika jsou si natolik podobná, že je lze snadno nahradit jedno druhým. Užijí se škálovaná data.

○ *Data:* v datech je uvedena převrácená hodnota mediánové účinné látky 1/ED50 [kg/mg]: *Lek* značí název neuroleptika, *Nervoz* značí potlačení nervozity, *Stereo* značí potlačení stereotypního chování, *Tres* značí potlačení záchvatu a tresu, *Usmr* značí dávka smrtícího účinku.

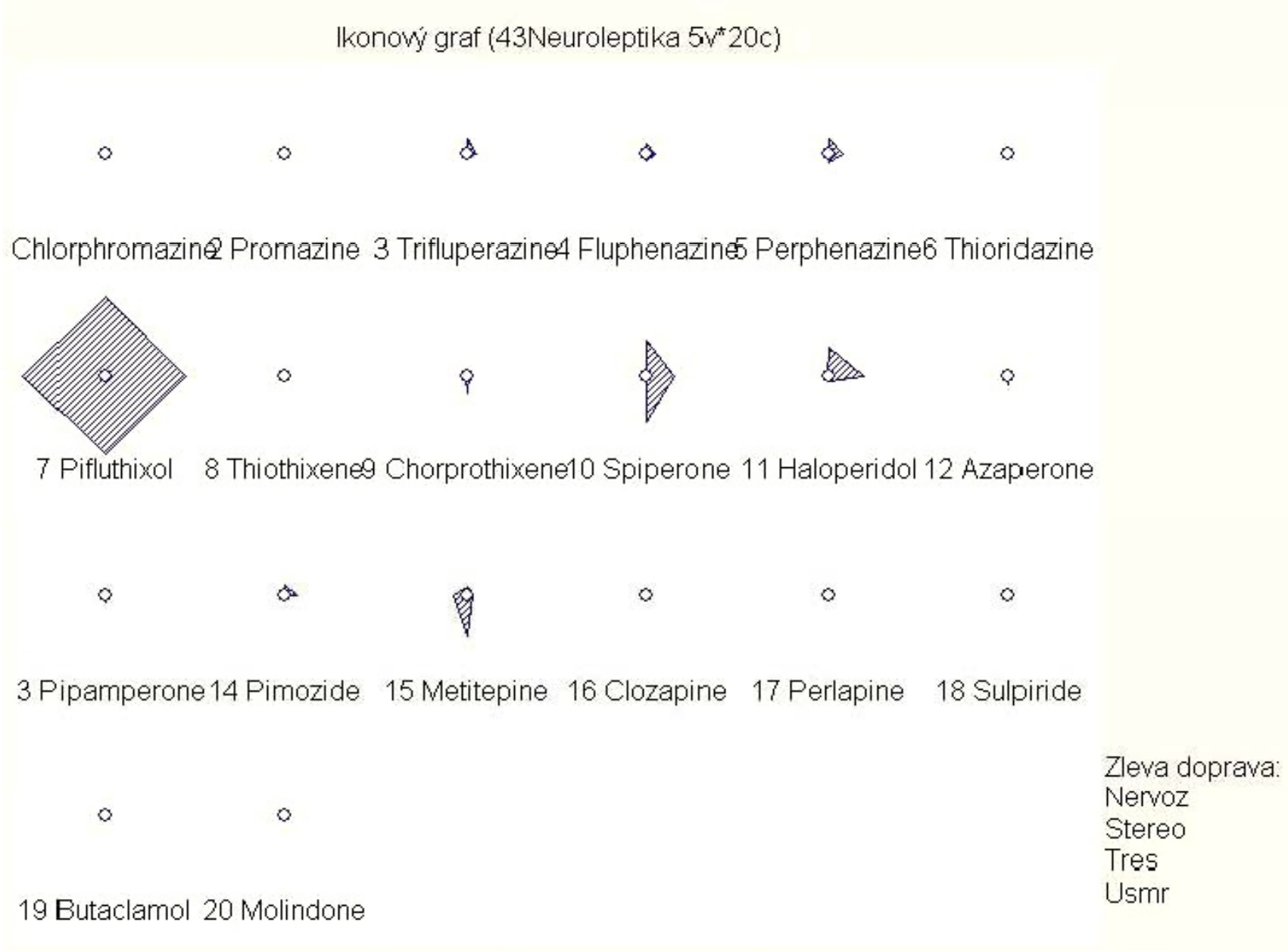
Lek	*Nervoz*	*Stereo*	*Tres*	*Usmr*
1 Chlorphromazine	3.846	3.333	1.111	1.923
2 Promazine	0.323	0.213	0.108	1.429
3 Trifluperazine	27.027	17.857	0.562	0.140
4 Fluphenazine	17.857	15.385	1.695	1.075
5 Perphenazine	27.027	27.027	1.961	2.083
6 Thioridazine	0.244	0.185	0.093	1.333
7 Pifluthixol	142.857	142.857	20.408	163.934
8 Thiothixene	4.348	4.348	0.047	0.345
9 Chorprothixene	5.882	2.941	4.545	4.167
10 Spiperone	62.500	47.619	11.765	0.847
11 Haloperidol	52.632	62.500	1.282	0.568
12 Azaperone	2.941	1.282	2.222	3.030
13 Pipamperone	0.327	0.187	1.724	0.397
14 Pimozide	20.408	20.408	0.107	0.025
15 Metitepine	15.385	10.204	10.204	27.027
16 Clozapine	0.161	0.093	0.327	0.323
17 Perlapine	0.323	0.323	0.370	0.067
18 Sulpiride	0.047	0.047	0.003	0.001
19 Butaclamol	10.204	9.091	1.471	0.025
20 Molindone	7.692	7.692	0.140	0.006



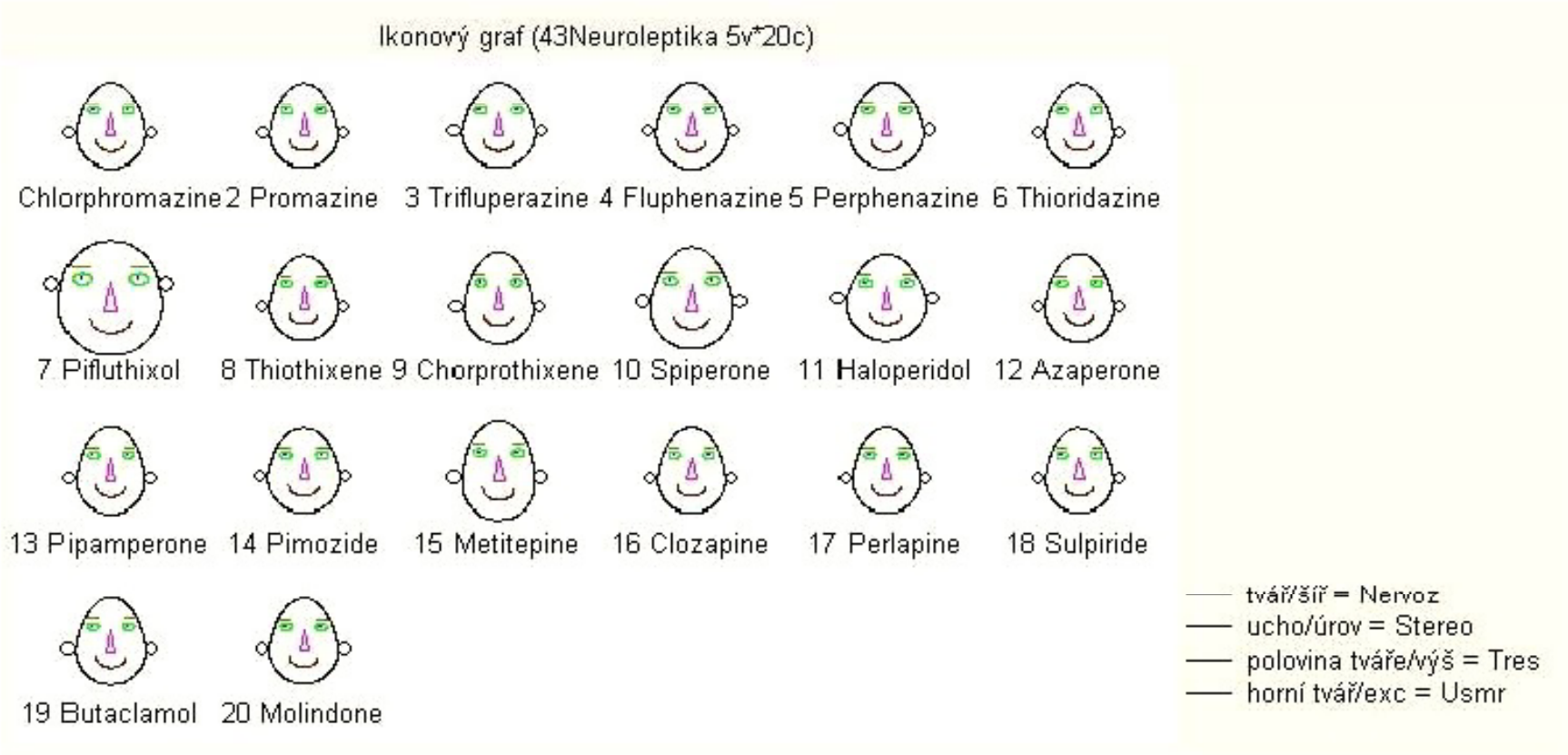
Čáry znaků pro každý objekt v EDA dat *Neuroleptika,* (STATISTICA).



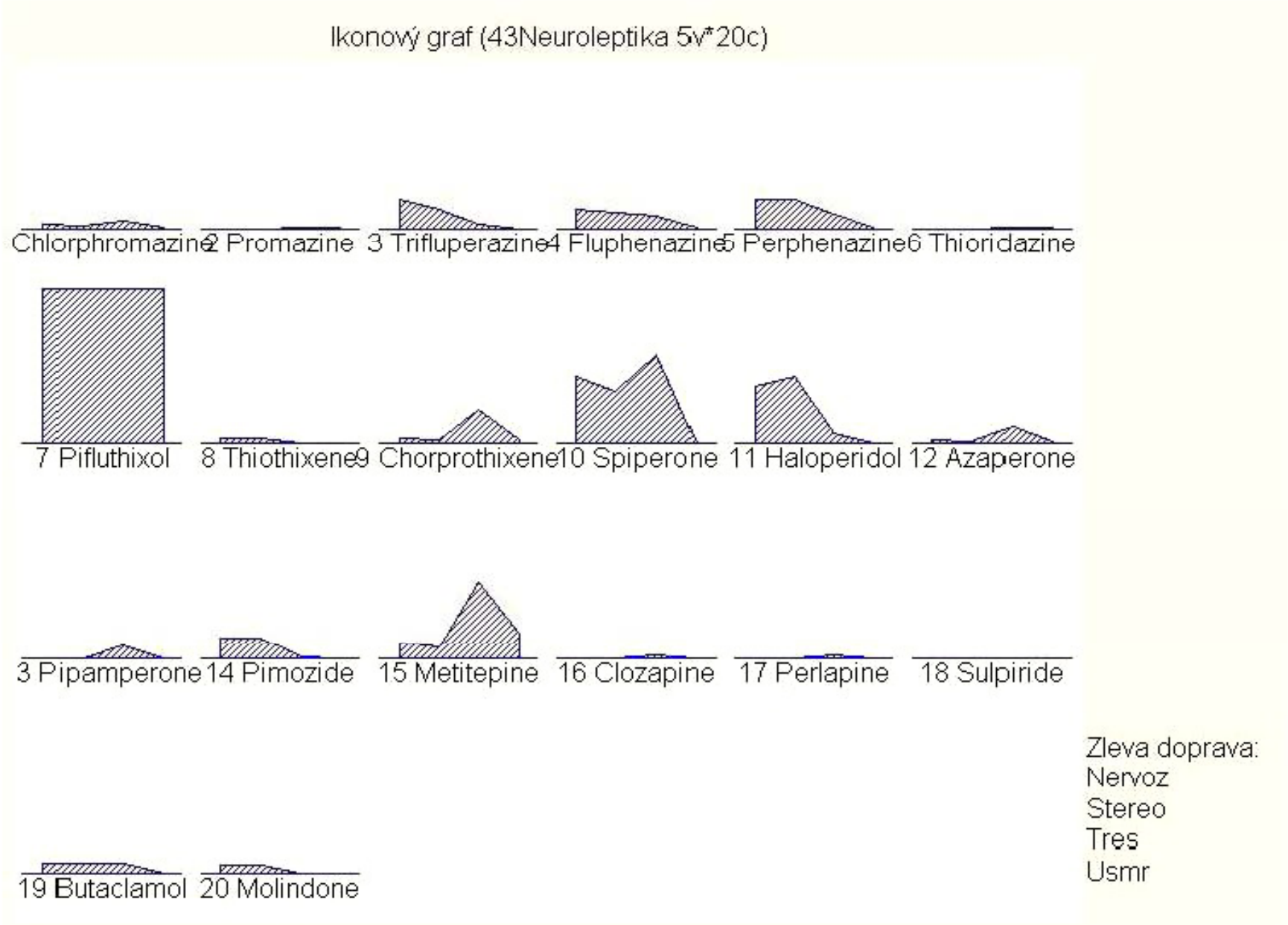
Hvězdy znaků pro každý objekt v EDA dat *Neuroleptika,* (STATISTICA).



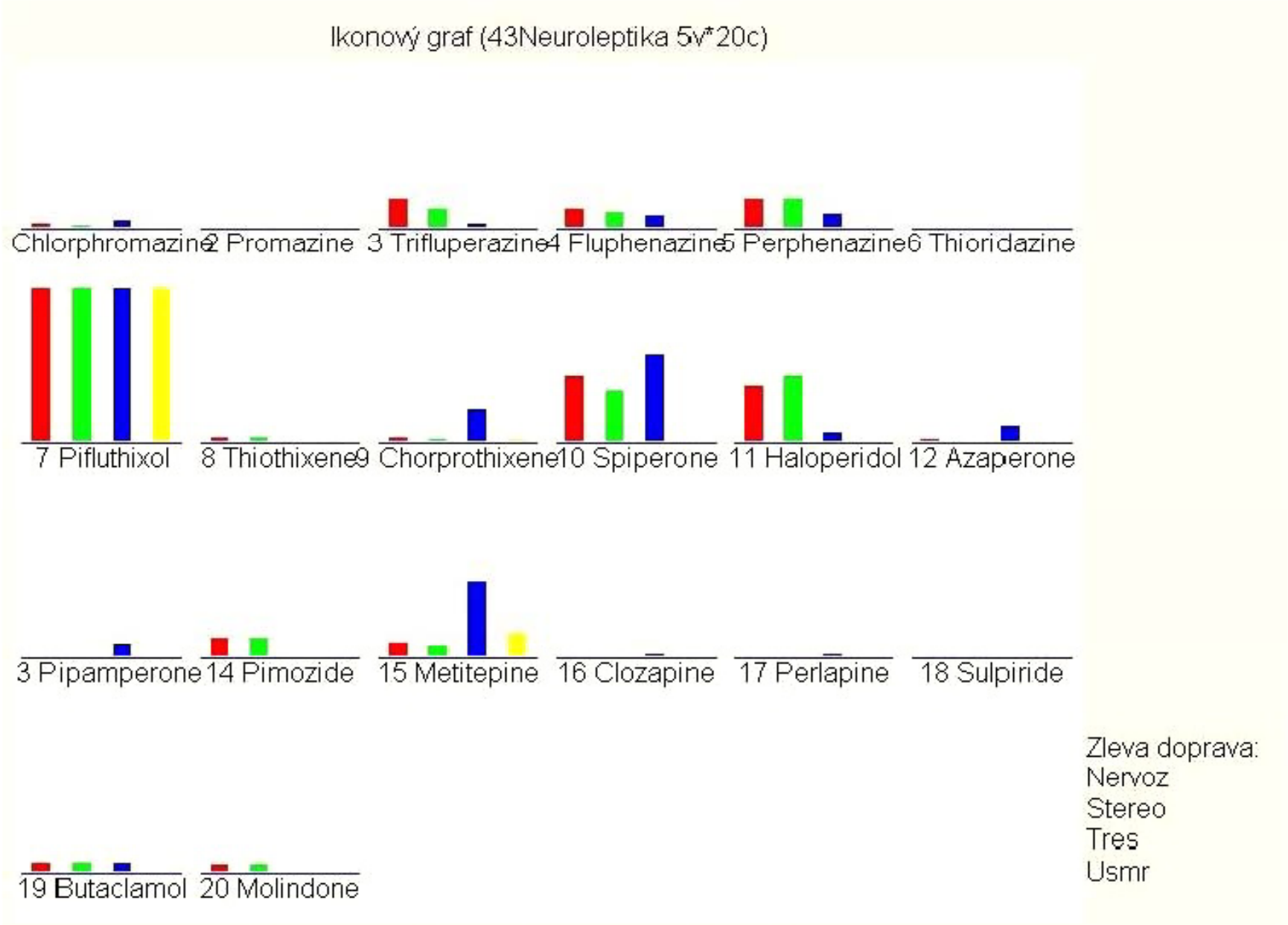
Mnohoúhelníky znaků pro každý objekt v EDA dat *Neuroleptika,* (STATISTICA).



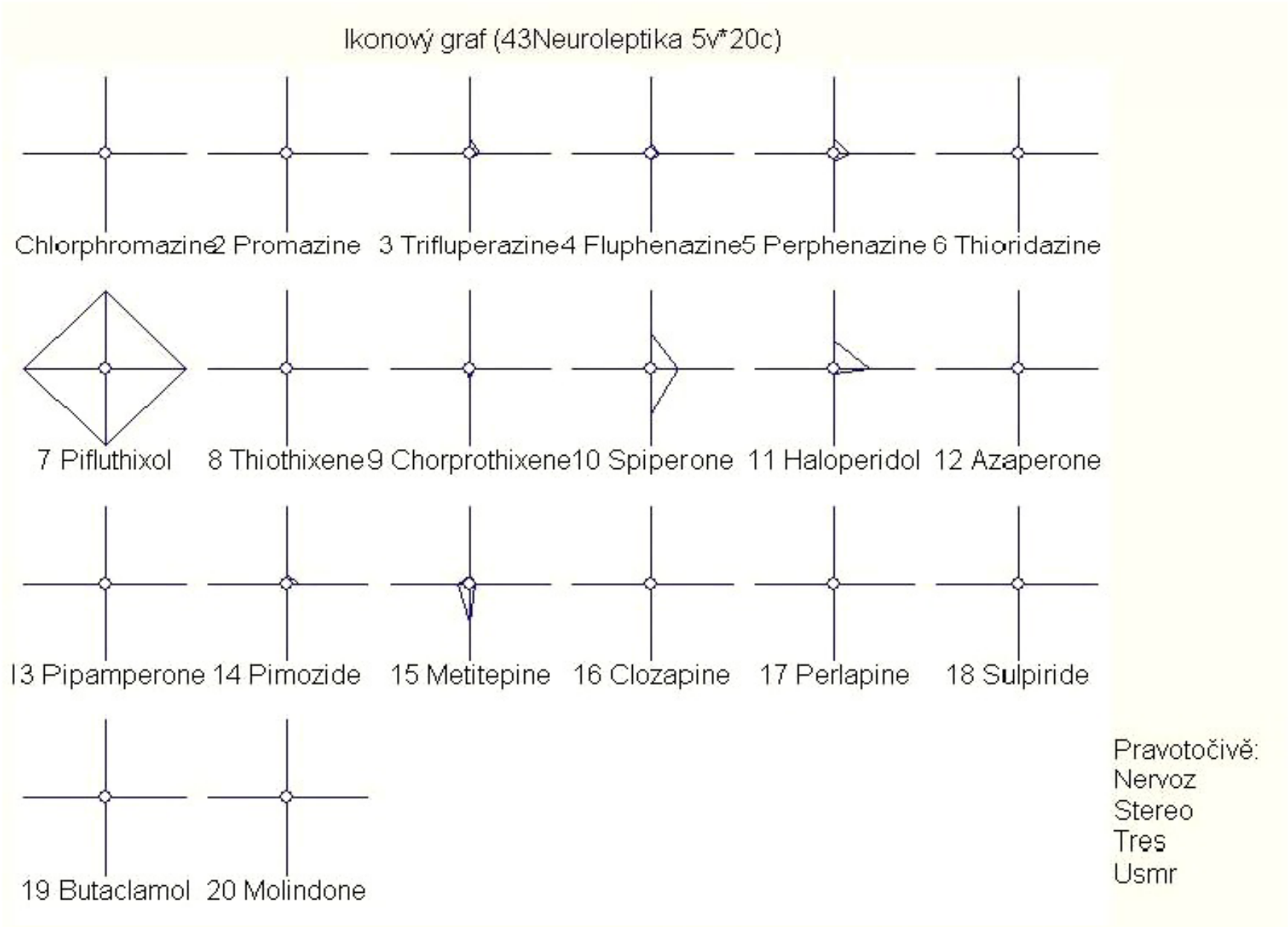
Obličeje znaků pro každý objekt v EDA dat *Neuroleptika*,(STATISTICA).



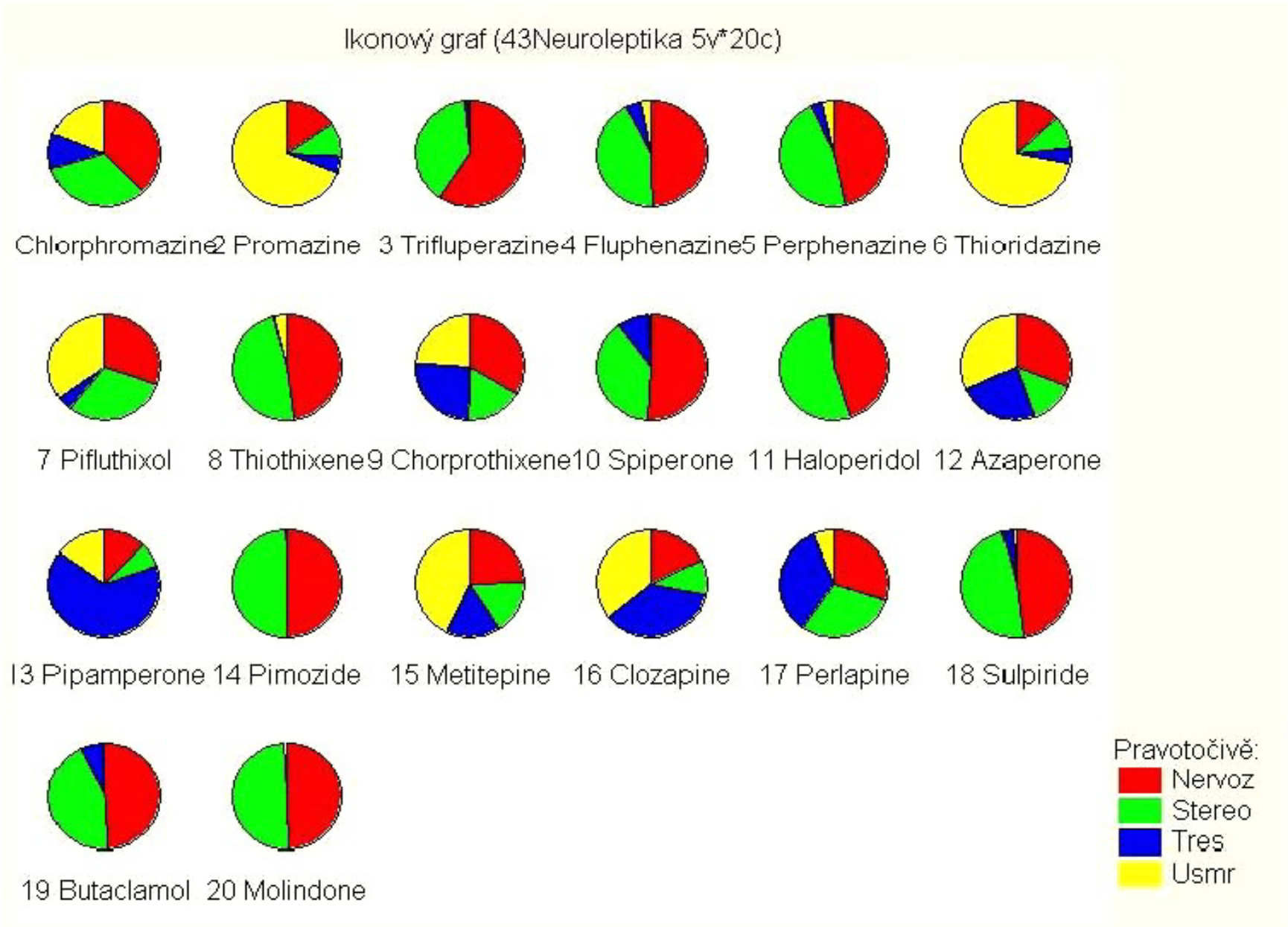
Profily znaků pro každý objekt v průzkumové analýze dat *Neuroleptika,* (STATISTICA).



Sloupce znaků pro každý objekt v EDA dat *Neuroleptika*, .



Sluníčka znaků pro každý objekt v EDA dat *Neuroleptika,* (STATISTICA).



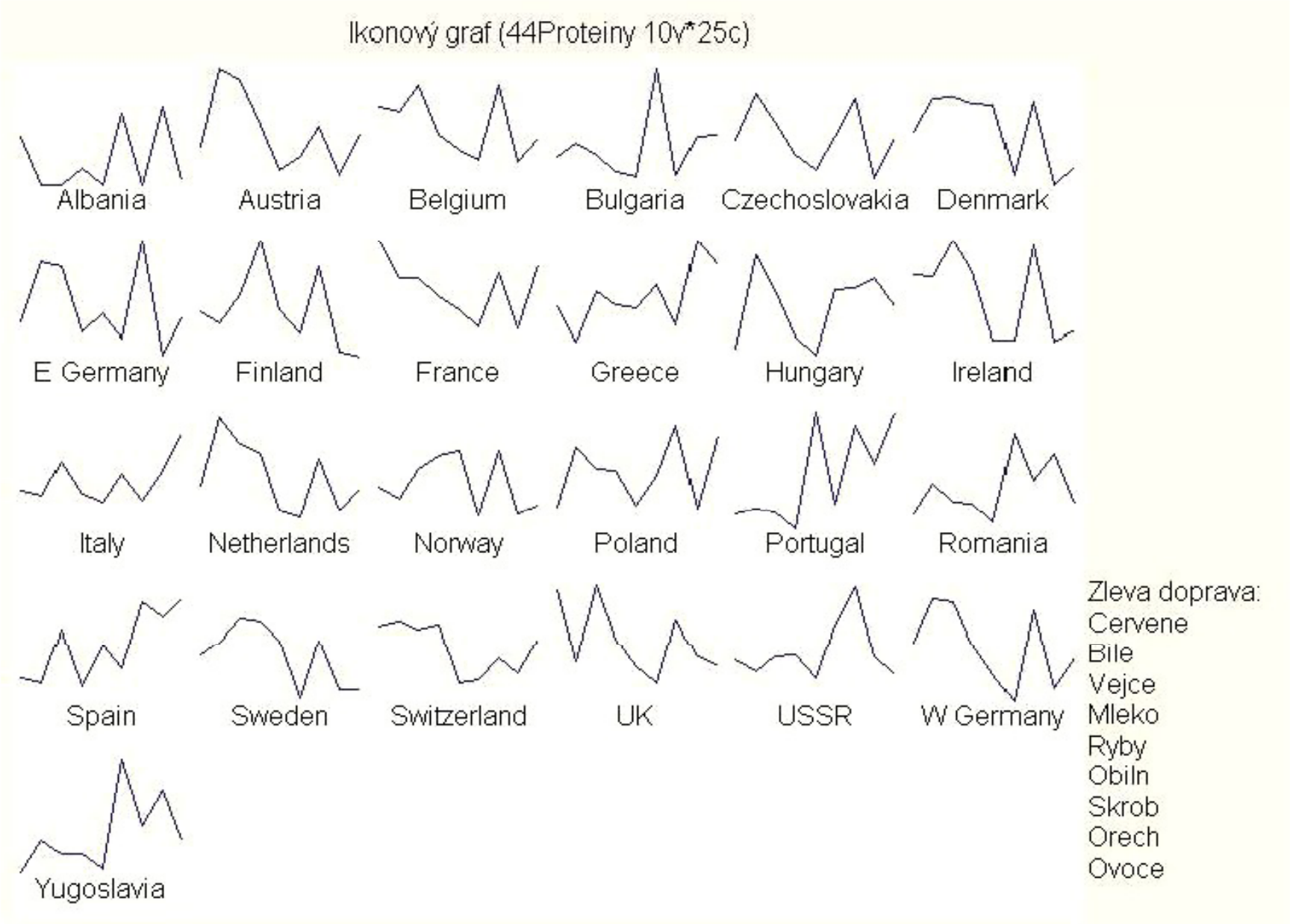
Výseče znaků pro každý objekt v EDA dat *Neuroleptika,* (STATISTICA).

PŘÍKLAD 2.6 *Sledování spotřeby proteinů v zemích Evropy*

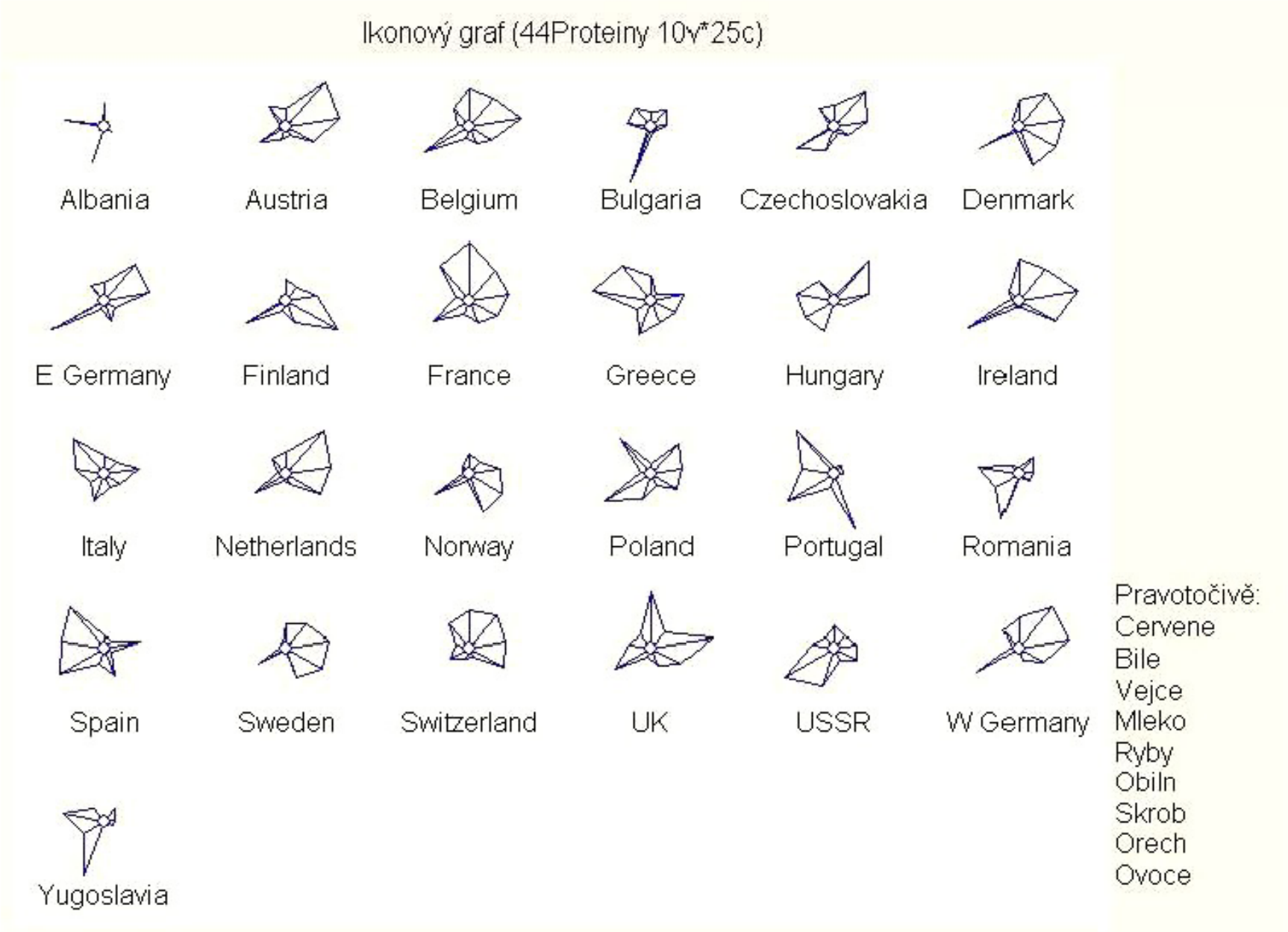
Sledována spotřeba proteinů v 25 zemích Evropy formou spotřeby 9 druhů potravin. Cílem je odhalit, zda existuje korelace mezi znaky, tj druhy potravin? Lze odhalit nějaké interakce mezi druhy potravin a zeměmi?

○ *Data:* v datech *Proteiny* jsou uvedeny znaky: *Cervene* značí spotřebu červeného masa, *Bile* značí spotřebu bílého masa, *Vejce* značí spotřebu vajec, *Mleko* se týká spotřeby mléka, *Ryby* značí spotřebu ryb, *Obiln* značí spotřebu obilnin, *Skrob* značí spotřebu škrobu, *Orech* značí spotřebu ořechů, *Ovoce* značí spotřebu ovoce a zeleniny.

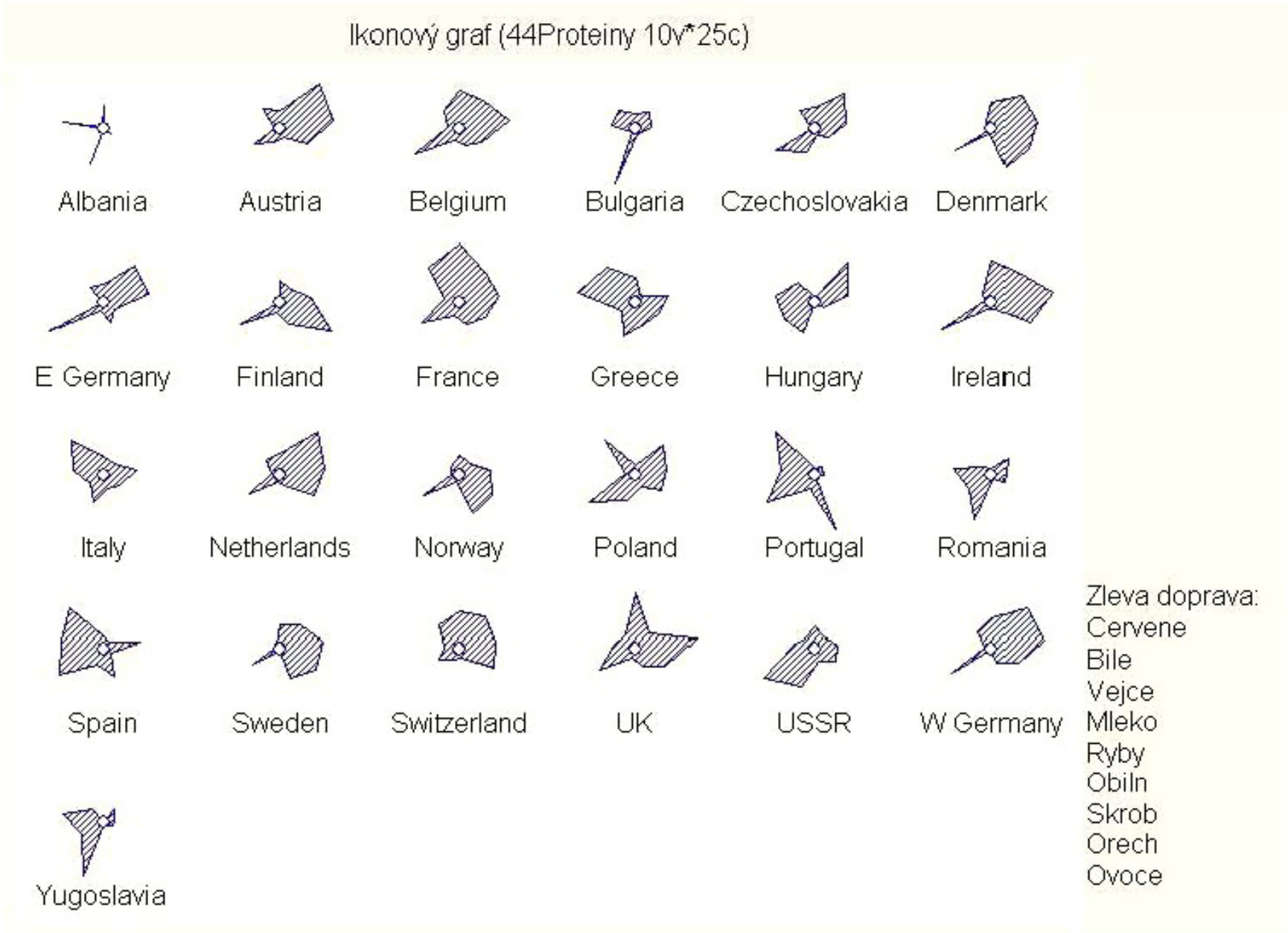
Země	*Cervene*	*Bile*	*Vejce*	*Mleko*	*Ryby*	*Obiln*	*Skrob*	*Orech*	*Ovoce*
Albania	10.1	1.4	0.5	8.9	0.2	42.3	0.6	5.5	1.7
Austria	8.9	14	4.3	19.9	2.1	28	3.6	1.3	4.3
Belgium	13.5	9.3	4.1	17.5	4.5	26.6	5.7	2.1	4
Bulgaria	7.8	6	1.6	8.3	1.2	56.7	1.1	3.7	4.2
Czechoslovakia	9.7	11.4	2.8	12.5	2	34.3	5	1.1	4
Denmark	10.6	10.8	3.7	25	9.9	21.9	4.8	0.7	2.4
East Germany	8.4	11.6	3.7	11.1	5.4	24.6	6.5	0.8	3.6
Finland	9.5	4.9	2.7	33.7	5.8	26.3	5.1	1	1.4
France	18	9.9	3.3	19.5	5.7	28.1	4.8	2.4	6.5
Greece	10.2	3	2.8	17.6	5.9	41.7	2.2	7.8	6.5
Hungary	5.3	12.4	2.9	9.7	0.3	40.1	4	5.4	4.2
Ireland	13.9	10	4.7	25.8	2.2	24	6.2	1.6	2.9
Italy	9	5.1	2.9	13.7	3.4	36.8	2.1	4.3	6.7
Netherlands	9.5	13.6	3.6	23.4	2.5	22.4	4.2	1.8	3.7
Norway	9.4	4.7	2.7	23.3	9.7	23	4.6	1.6	2.7
Poland	6.9	10.2	2.7	19.3	3	36.1	5.9	2	6.6
Portugal	6.2	3.7	1.1	4.9	14.2	27	5.9	4.7	7.9
Romania	6.2	6.3	1.5	11.1	1	49.6	3.1	5.3	2.8
Spain	7.1	3.4	3.1	8.6	7	29.2	5.7	5.9	7.2
Sweden	9.9	7.8	3.5	24.7	7.5	19.5	3.7	1.4	2
Switzerland	13.1	10.1	3.1	23.8	2.3	25.6	2.8	2.4	4.9
UK	17.4	5.7	4.7	20.6	4.3	24.3	4.7	3.4	3.3
USSR	9.3	4.6	2.1	16.6	3	43.6	6.4	3.4	2.9
West Germany	11.4	12.5	4.1	18.8	3.4	18.6	5.2	1.5	3.8
Yugoslavia	4.4	5	1.2	9.5	0.6	55.9	3	5.7	3.2



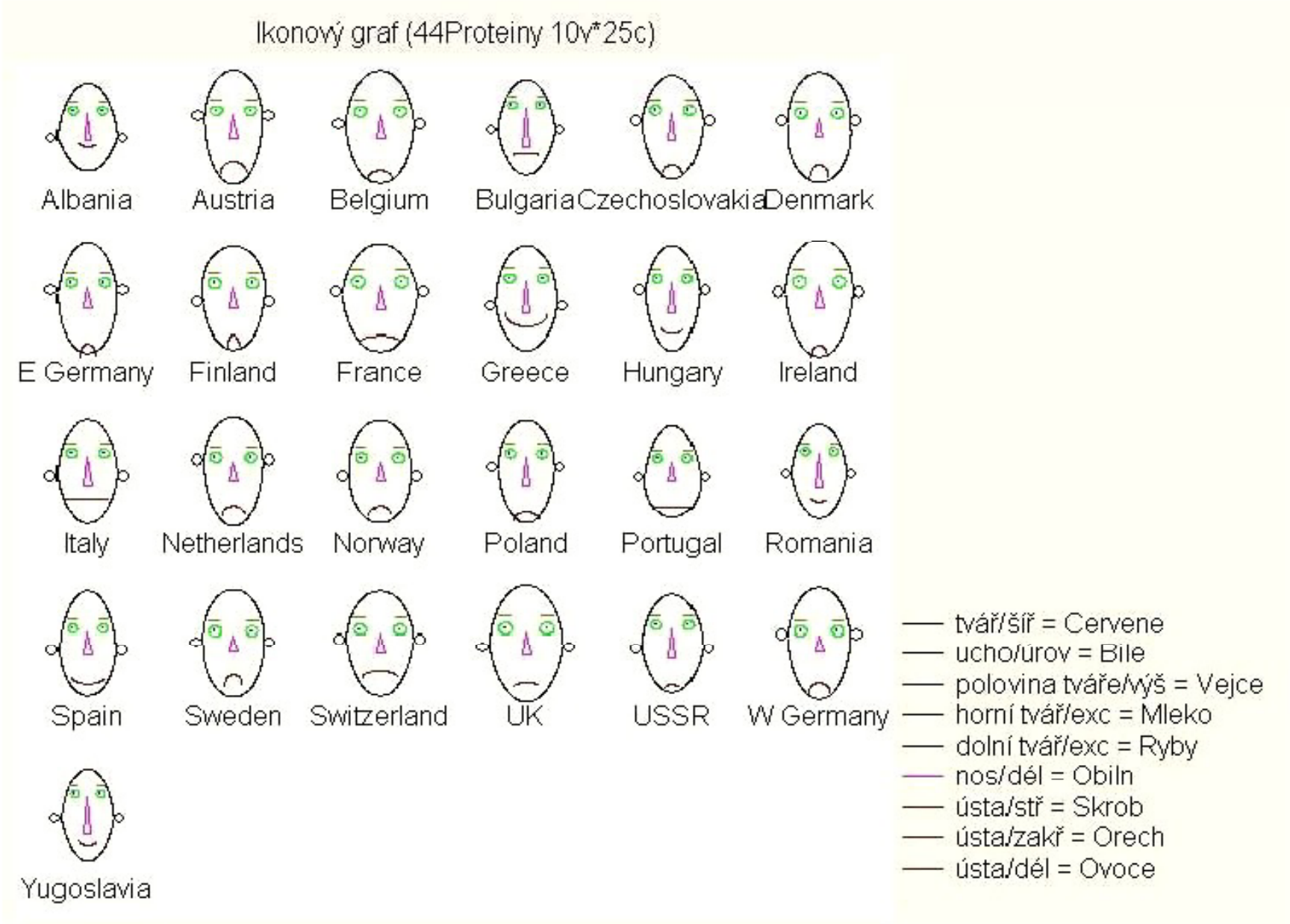
Čáry znaků pro každý objekt v EDA dat *Proteiny,* (STATISTICA).



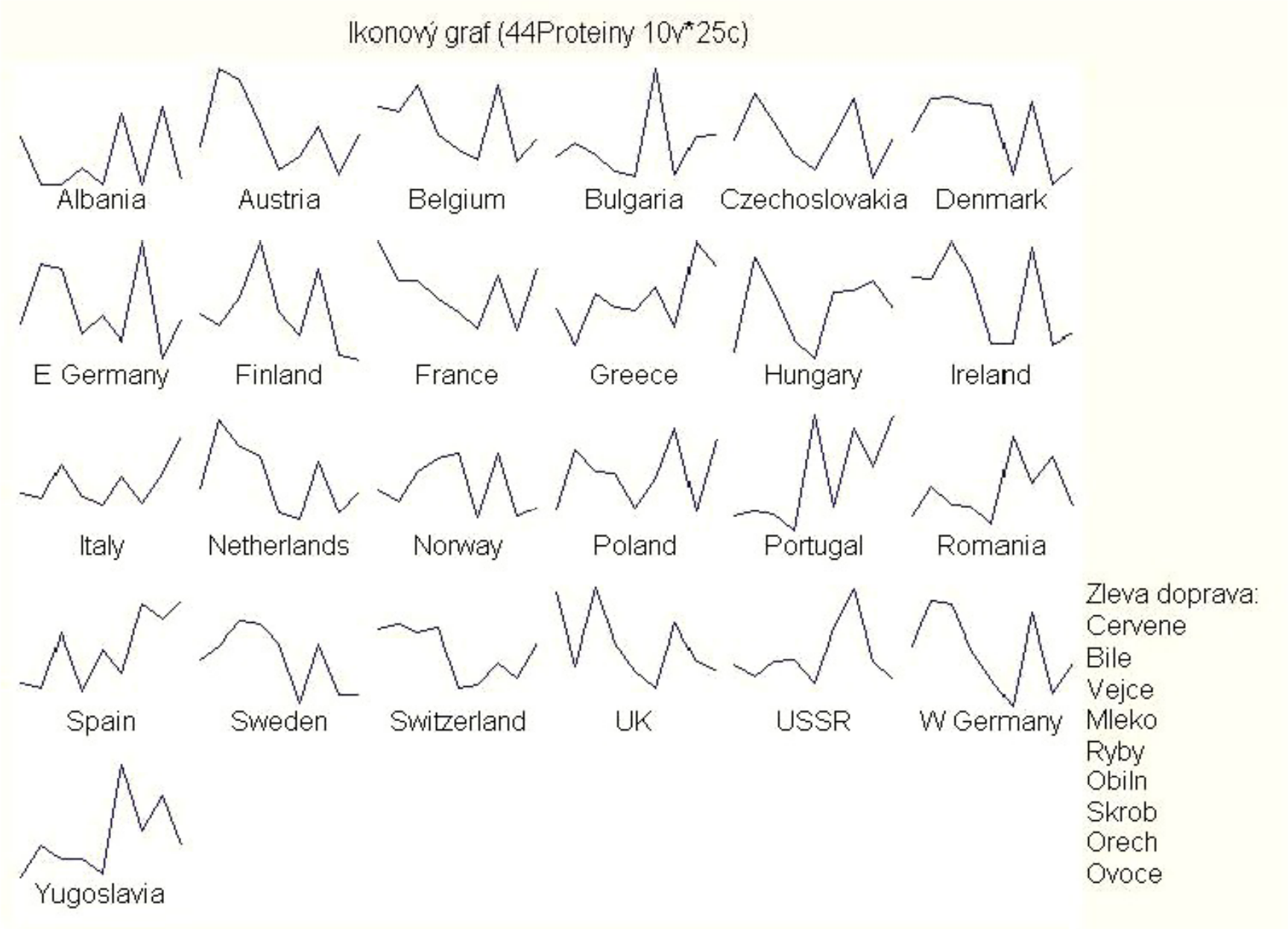
Hvězdy znaků pro každý objekt v EDA dat *Proteiny,* (STATISTICA).



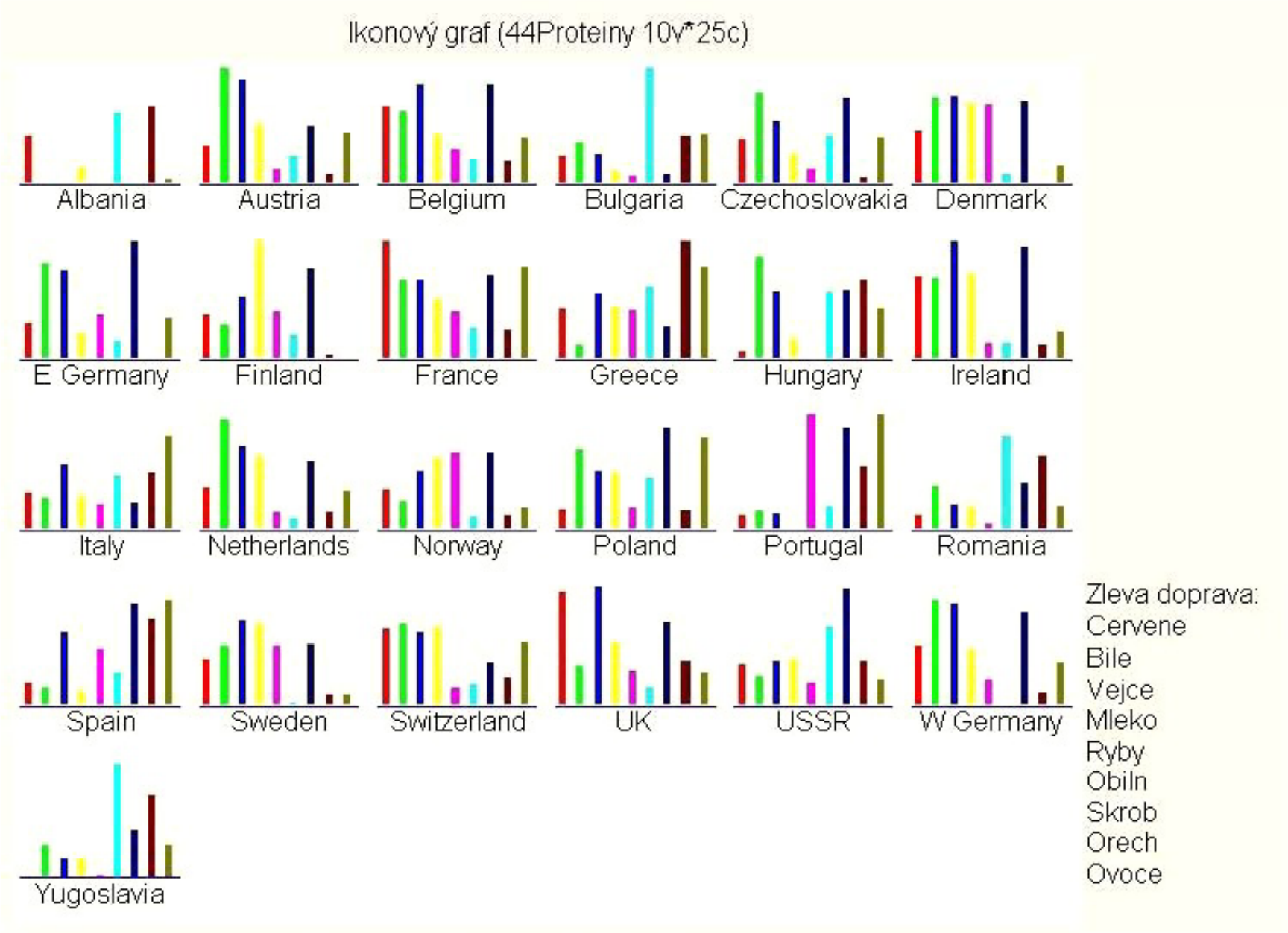
Mnohoúhelníky znaků pro každý objekt v EDA dat *Proteiny*, (STATISTICA).



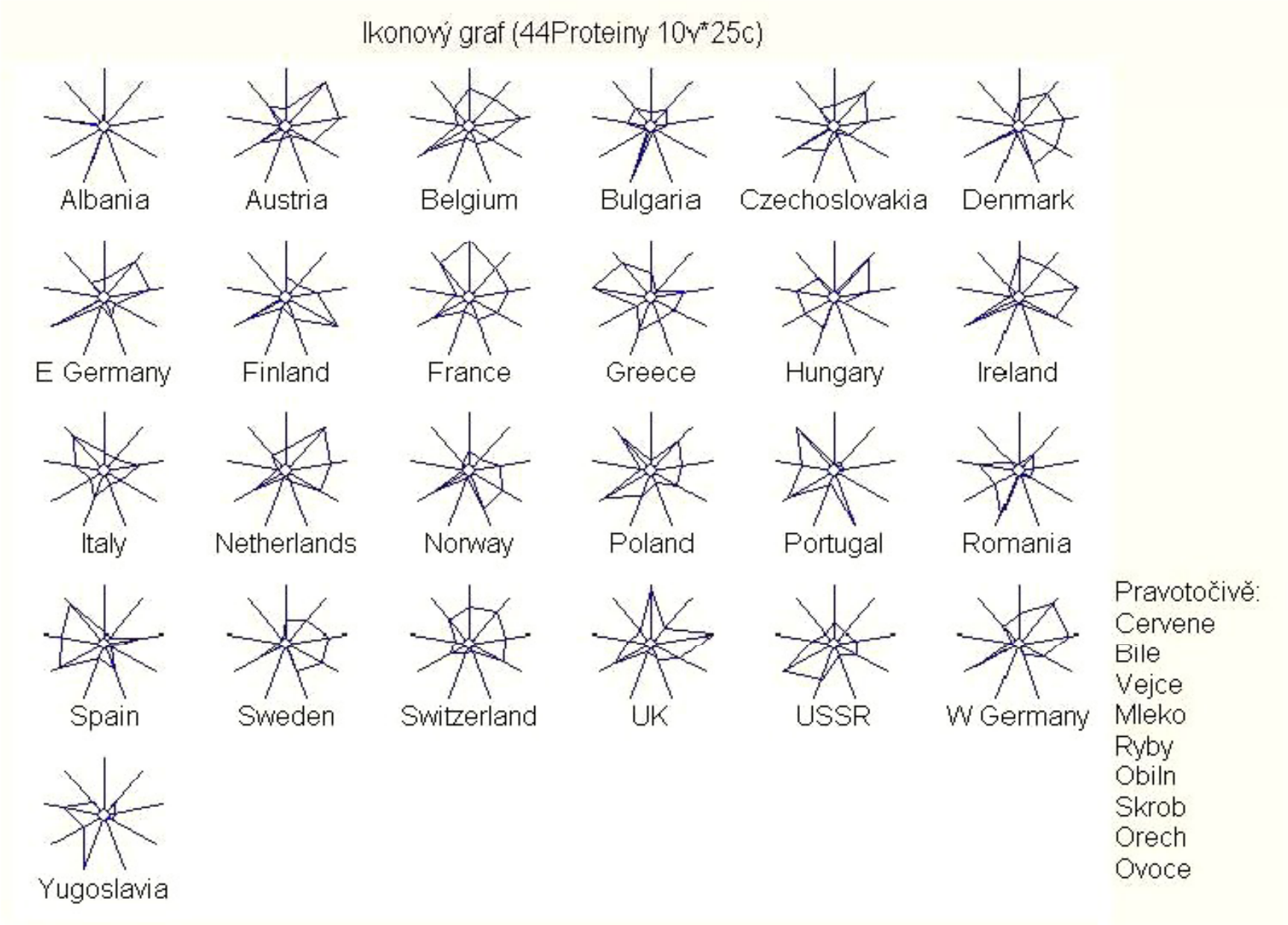
Obličeje znaků pro každý objekt v EDA dat *Proteiny*,(STATISTICA).



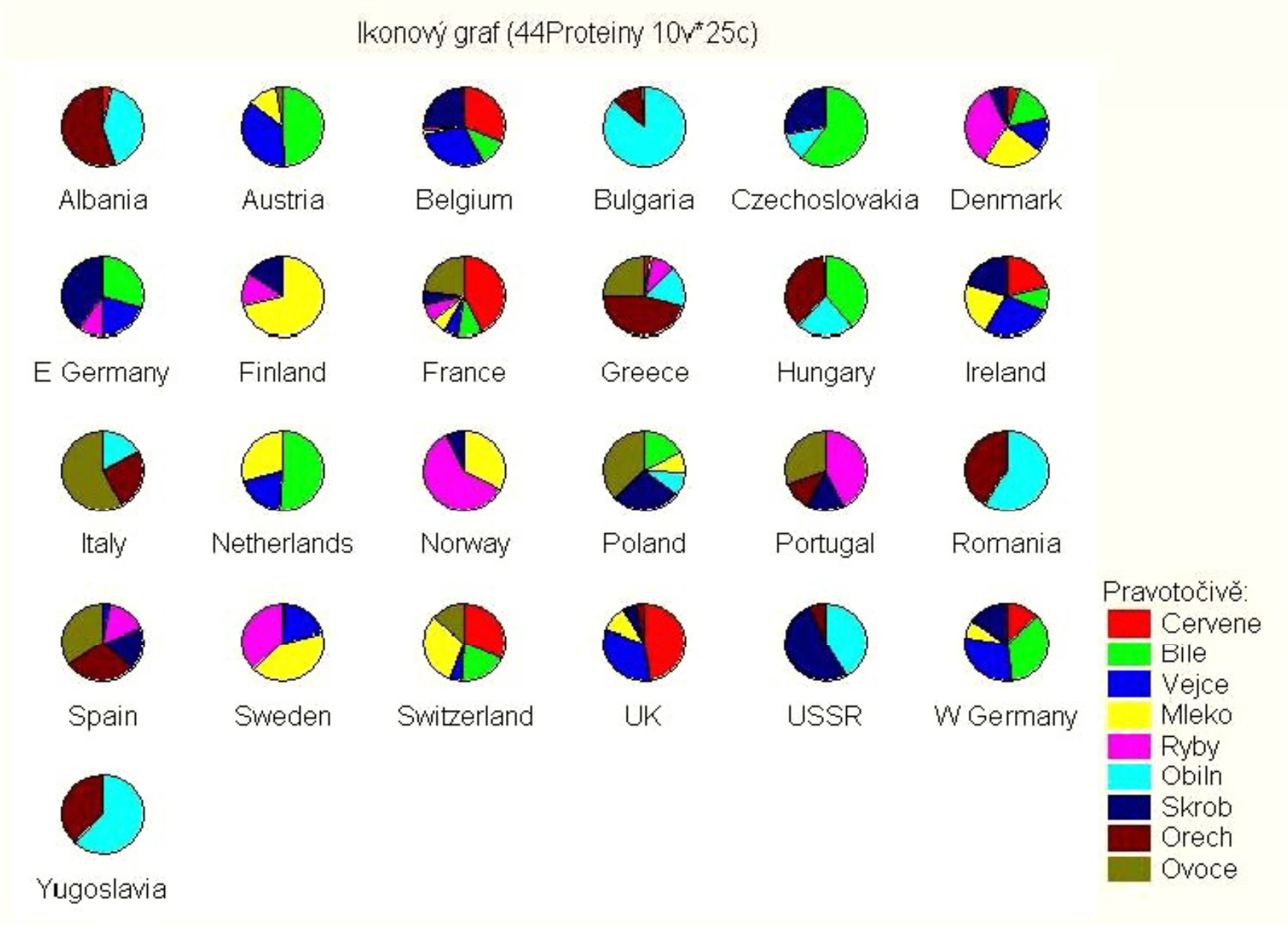
Profily znaků pro každý objekt v průzkumové analýze dat *Proteiny*, (STATISTICA).



Sloupce znaků pro každý objekt v EDA dat *Proteiny*, .



Sluníčka znaků pro každý objekt v EDA dat *Proteiny*, (STATISTICA).



Výseče znaků pro každý objekt v EDA dat *Proteiny,* (STATISTICA).