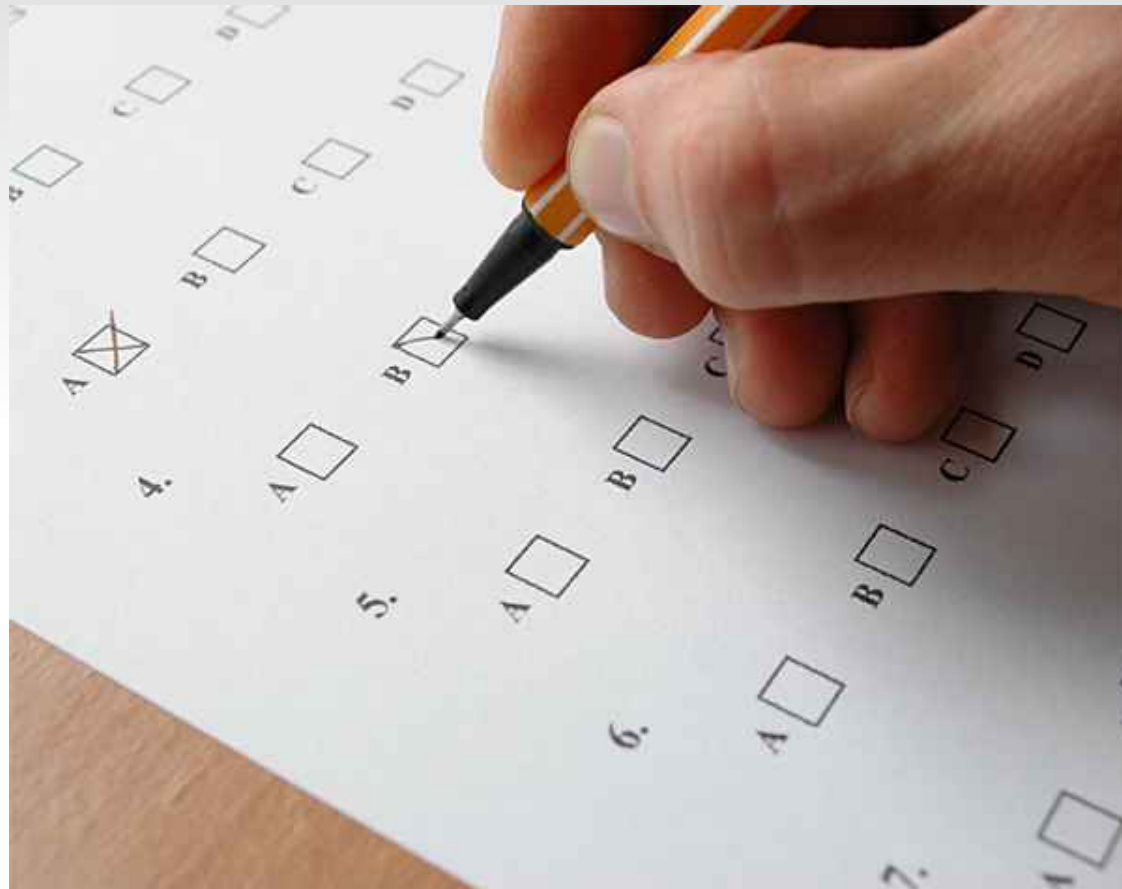


Analýza dotazníkových prieskumov



(projekt z predmetu Dobývaní znalostí - NDBI023)

Peter Bašista a Radek Píbil

Použité dáta

- Prvý prieskum:
 - Koho by ste nechceli za suseda?
- Napríklad:
 - ťažkých alkoholikov
 - rómov
 - pravicových extrémistov
 - drogovo závislých
 - homosexuálov

a tak ďalej...
- Celkovo 14 možných odpovedí
- Zúčastnili sa ľudia z 22 európskych krajín

Použitá dáta

- K dispozícii sme mali agregované dáta pre jednotlivé štáty a pre všetky možné odpovede
- Napríklad:
 - 77,2% Slovákov sa vyjadrilo, že by nechceli za susedov rómov
 - 75,5% Čechov sa vyjadrilo, že by nechceli za susedov ťažkých alkoholikov
 - 5,83% Španielov sa vyjadrilo, že by nechceli za susedov ľudí ktorí majú veľké rodiny

	A	B	C	D	E	F	G	
1	Country	Záznam v registri trestov	Odlíšná rasa	Ľavicoví extrémisti	Ťažkí alkoholici	Pravicoví extrémisti	Veľké rodiny	Emocior
2	France	0.21	0.09	0.25	0.05	0.04	0.10	
3	United Kingdom	0.48	0.09	0.27	0.51	0.30	0.14	
4	West Germany	0.26	0.04	0.50	0.57	0.78	0.06	
5	East Germany	0.22	0.07	0.00	0.53	0.72	0.07	
6	Austria	0.27	0.07	0.50	0.00	0.60	0.04	
7	Spain	0.32	0.10	0.25	0.35	0.29	0.06	
8	Portugal	0.44	0.08	0.11	0.04	0.12	0.05	
9	Netherlands	0.33	0.05	0.05	0.57	0.68	0.07	
10	Belgium	0.03	0.14	0.36	0.43	0.48	0.08	
11	Denmark	0.31	0.01	0.09	0.36	0.20	0.04	
12	Sweden	0.33	0.02	0.23	0.33	0.48	0.04	
13	Finland	0.39	0.12	0.13	0.51	0.18	0.07	
14	Estonia	0.69	0.15	0.30	0.84	0.24	0.14	
15	Latvia	0.57	0.05	0.03	0.75	0.21	0.05	
16	Lithuania	0.66	0.10	0.35	0.82	0.32	0.12	
17	Poland	0.68	0.17	0.18	0.78	0.02	0.11	

Niekoľko problémov

- Dáta pre Taliansko boli zmazané (samé nuly)
 - Nepoužili sme ich
- Dáta pre Maďarsko sa príliš odlišovali od zvyšku
 - Domnelý dôvod:
 - odlišná metodika zberu dát
 - Nepoužili sme ani tieto dáta
- Dáta pre Nemecko boli rozdelené podľa toho, či ide o západné alebo východné Nemecko (zrejme to odpovedá bývalým SRN a NDR)
 - Pri väčšine postupov nám to nevadilo
 - ✓ Nechali sme to tak

Postup

- Pre každú krajinu sme si spočítali:
 - Strednú hodnotu
 - Rozptyl
- hľadali sme najväčšie a najmenšie hodnoty
- Pre každú možnosť odpovede sme si spočítali:
 - Strednú hodnotu
 - Rozptyl
- hľadali sme najväčšie a najmenšie hodnoty
- Na všetkých dátach sme:
 - Použili klastrovanie (k-means)
- hľadali sme:
 - závislosti medzi krajinami
 - závislosti medzi odpoveďami

Výsledok všeobecne

- Na adrese:
<http://basip6am.matfyz.cz/dmproject/>
sa nachádzajú použité dáta, skripty v Matlabe, výsledky, prezentácia
- Základné výsledky sú tvaru:
 - Španieli by mali najmenší problém prijať za svojich susedov rómov (0%)
 - Najviac by Rusi nechceli susediť s ťažkými alkoholikmi (73,2%)
- Zaujímavejšie sú výsledky klastrovania:
 - Napríklad sme zistili, že Rusi a Bulhari majú veľmi podobné názory na susedov

Základné výsledky

- Úplné výsledky sú na webe
- *Najmenej vadí* podľa typu susedov:

Susedia, ktorí sú / majú

Ťažkí alkoholici

Veľké rodiny

Nakazení AIDS

Drogovo závislí

Rómovia

Krajina

Rakúsko

Švédsko

Španielsko

Španielsko

Španielsko

Pomer

0.01%

4.08%

0.00%

0.00%

0.00%

Základné výsledky

- *Najviac vadí* podľa typu susedov:

Susedia, ktorí sú / majú

Pravicoví extrémisti

Veľké rodiny

Drogovo závislí

Židia

Rómovia

Krajina

západné Nemecko

Spojené kráľovstvo

Estónsko

Poľsko

Slovensko

Pomer

77.70%

14.30%

89.90%

25.10%

77.20%

Základné výsledky

- *Najmenej vadí* podľa krajín:

<u>Krajina</u>	<u>Susedia, ktorí sú / majú</u>	<u>Pomer</u>
Rakúsko	Ťažkí alkoholicy	0.01%
Španielsko	Nakazení AIDS	0.00%
	Drogovo závislí	0.00%
	Rómovia	0.00%
Portugalsko	Imigranti a zahraniční pracovníci	2.49%
Holandsko	Homosexuáli	0.64%
Belgicko	Záznam v registri trestov	2.77%
Švédsko	Nakazení AIDS	0.67%
Fínsko	Veľké rodiny	7.50%

Základné výsledky

- *Najviac vadí podľa krajín:*

<u>Krajina</u>	<u>Susedia, ktorí sú / majú</u>	<u>Pomer</u>
západné Nemecko	Pravicoví extrémisti	77.70%
východné Nemecko	Pravicoví extrémisti	71.80%
Rakúsko	Pravicoví extrémisti	59.70%
Španielsko	Ťažkí alkoholici	35.40%
Estónsko	Drogovo závislí	89.90%
Slovensko	Záznam v registri trestov	81.60%

Základné výsledky

- *Typ susedov* podľa strednej hodnoty pre všetky krajiny:

Susedia, ktorí sú / majú

Veľké rodiny

Odlišná rasa

Židia

Záznam v registri trestov

Ťažkí alkoholicy

Drogovo závislí

Stredná hodnota

8.368%

8.695%

10.348%

45.685%

54.269%

58.149%

Základné výsledky

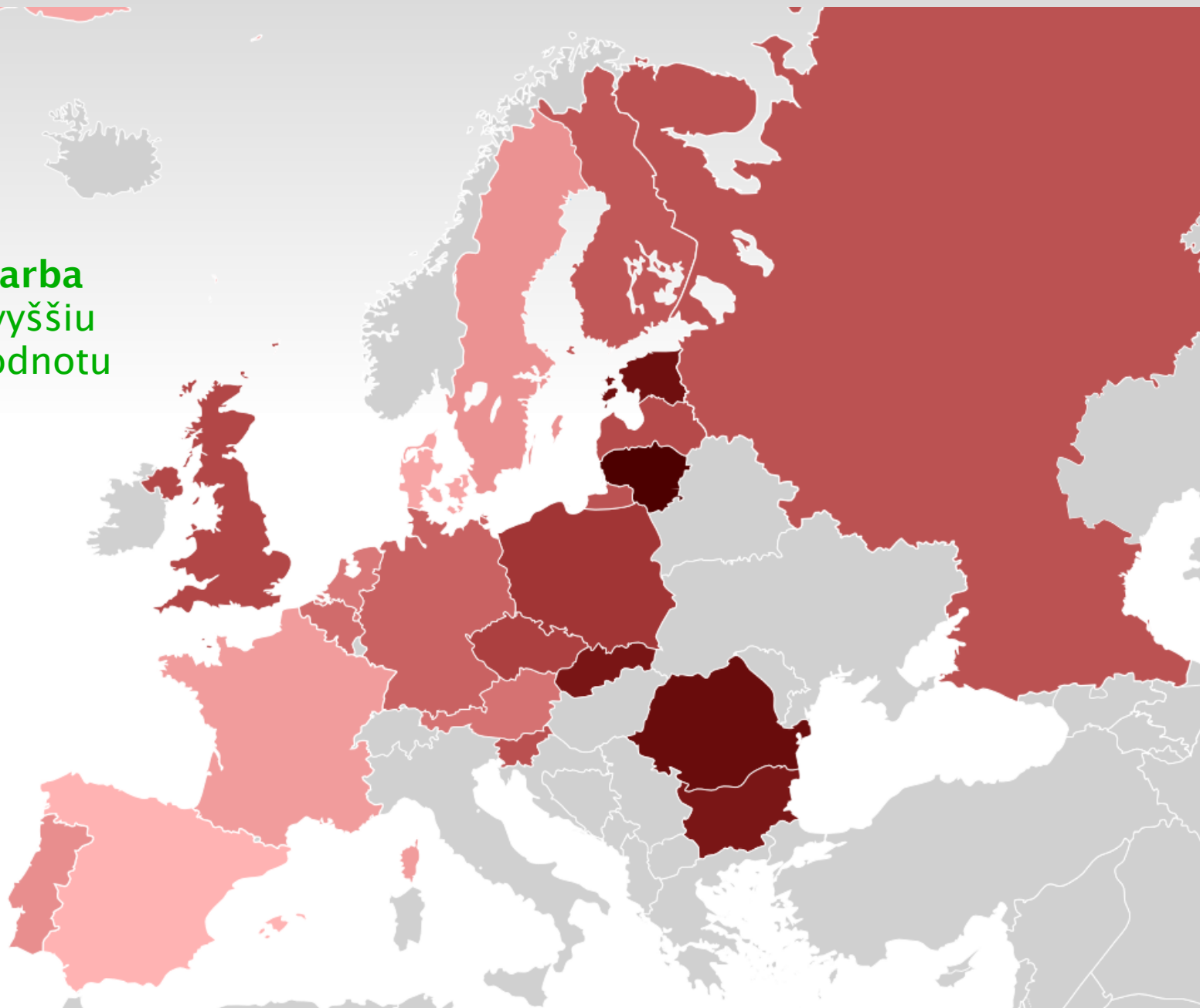
- *Typ susedov* podľa rozptylu pre všetky krajiny:

<u>Susedia, ktorí sú / majú</u>	<u>Rozptyl</u>
Veľké rodiny	0.0011
Odlišná rasa	0.0037
Imigranti a zahraniční pracovníci	0.0046
Židia	0.0047
Moslimovia	0.0056
Drogovo závislí	0.0602
Ťažkí alkoholicy	0.0693

Základné výsledky

- *Krajiny* podľa strednej hodnoty pre všetky typy susedov:

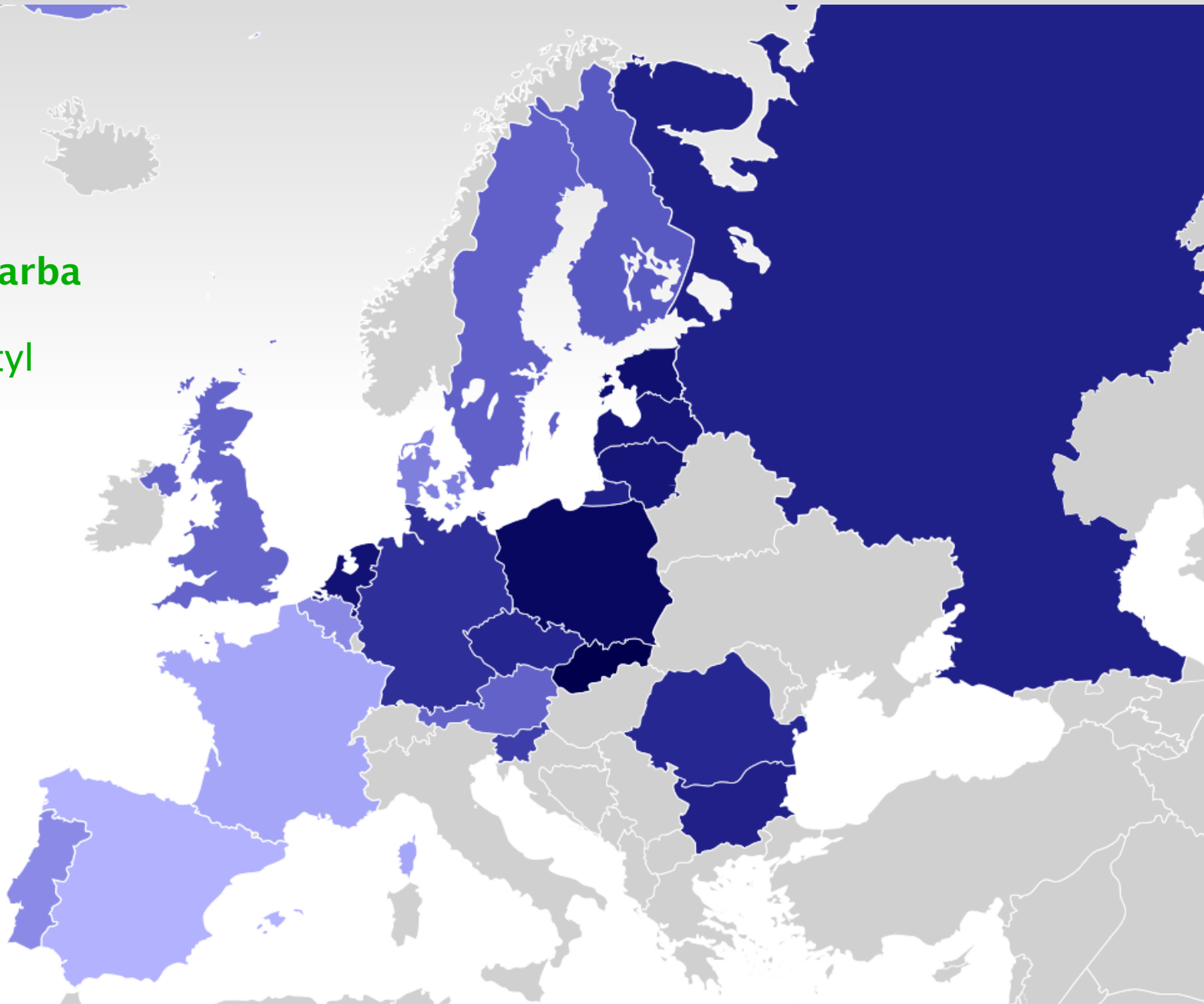
Tmavšia farba
znamená vyššiu
strednú hodnotu



Základné výsledky

- *Krajiny* podľa rozptylu pre všetky typy susedov:

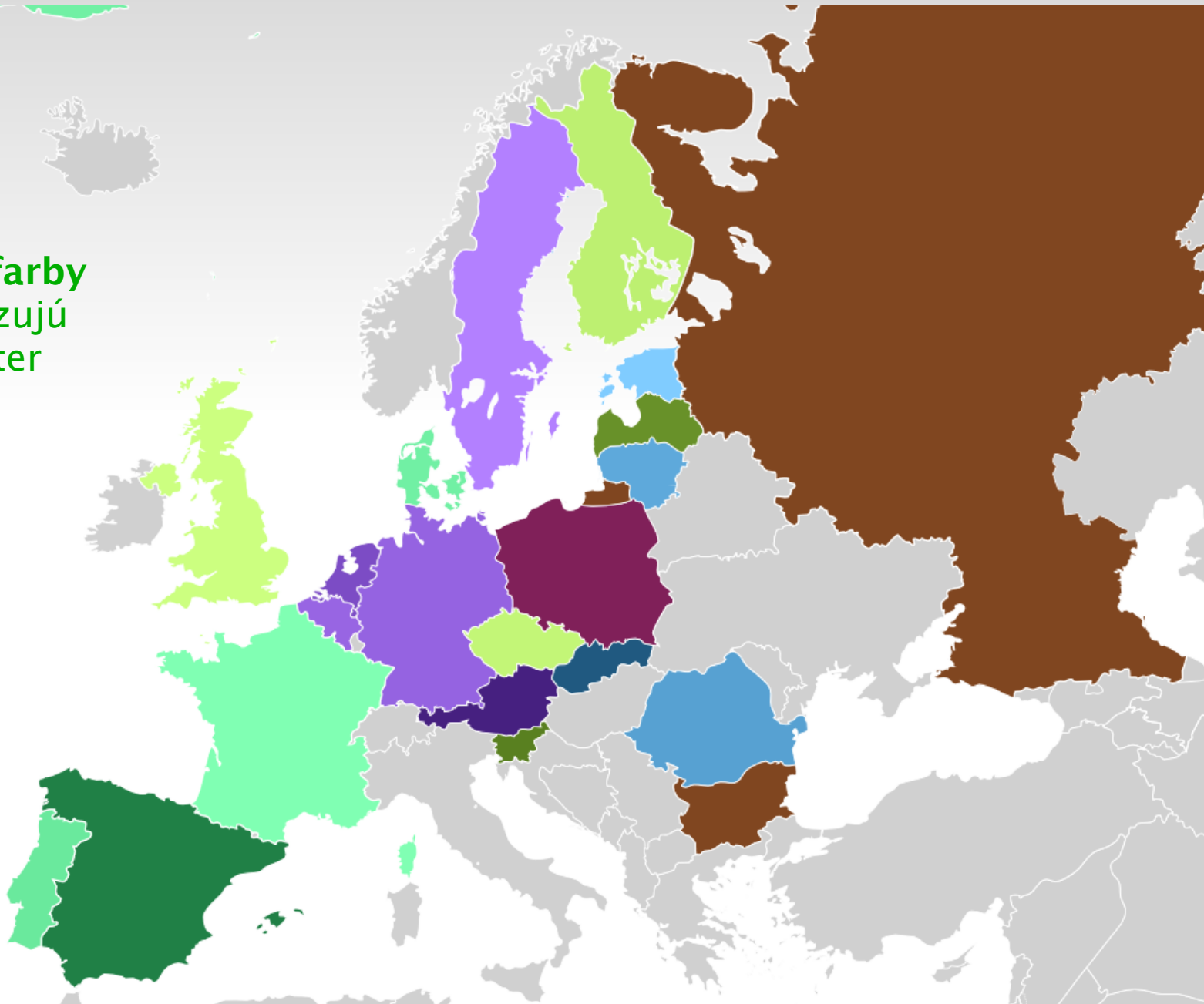
Tmavšia farba
znamená
vyšší rozptyl



Výsledky klastrovania

- *Krajiny* podľa príslušnosti ku klastrom

Podobné farby
charakterizujú
jeden klaster



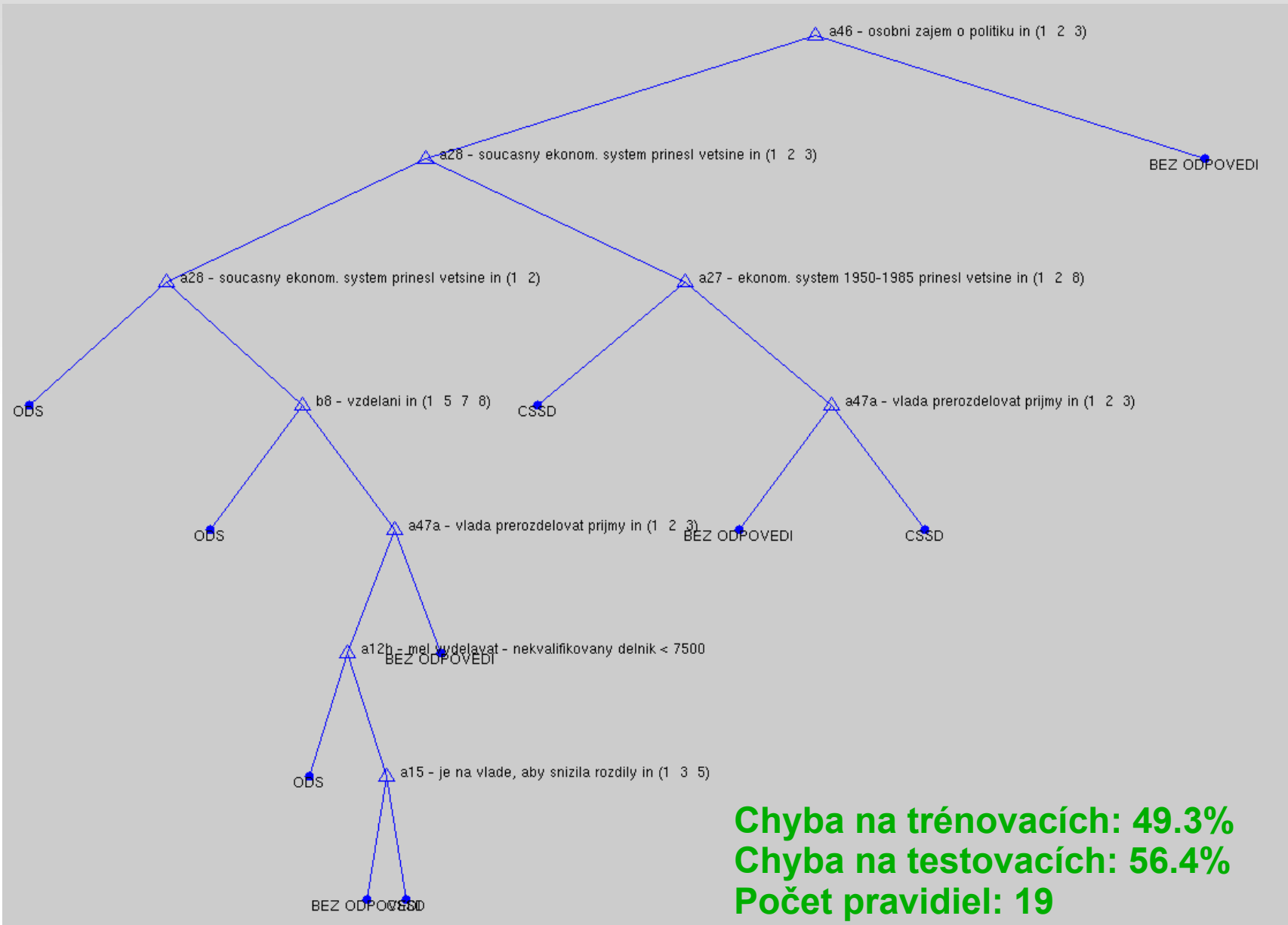
Ďalšie použité dáta

- Druhý prieskum:
 - Veľký, veľa otázok, veľa respondentov
 - Široké spektrum otázok
 - od "Veríte v Boha?"
 - po "Aký náročný bol pre vás tento dotazník?"
- Dáta sme výrazne orezali a zjednodušili
 - Pôvodne: asi 250 otázok (atribútov)
od 1834 respondentov
 - Vybrali sme si 24 otázok
- Museli sme sa tiež vysporiadať s chýbajúcimi údajmi
 - Číselné sme nahradili priemerom
 - Kategoriálne sme zlúčili do odpovede "neviem"

Ako sme vyberali otázky?

- Sledovali sme odpoveď na otázku:
 - Akú stranu by ste volili?
- Vybrali sme otázky typu:
 - Koľko by mal zarábať riaditeľ veľkého podniku?
 - Je vzdelanie dôležité pre dosiahnutie bohatstva?
 - Je na vláde, aby znižovala sociálne rozdiely?
 - Zaujímate sa vôbec o politiku?
 - ... a tak ďalej
- Analýzu sme robili pomocou
rozhodovacieho stromu

Výsledný strom (strany) - s orezaním

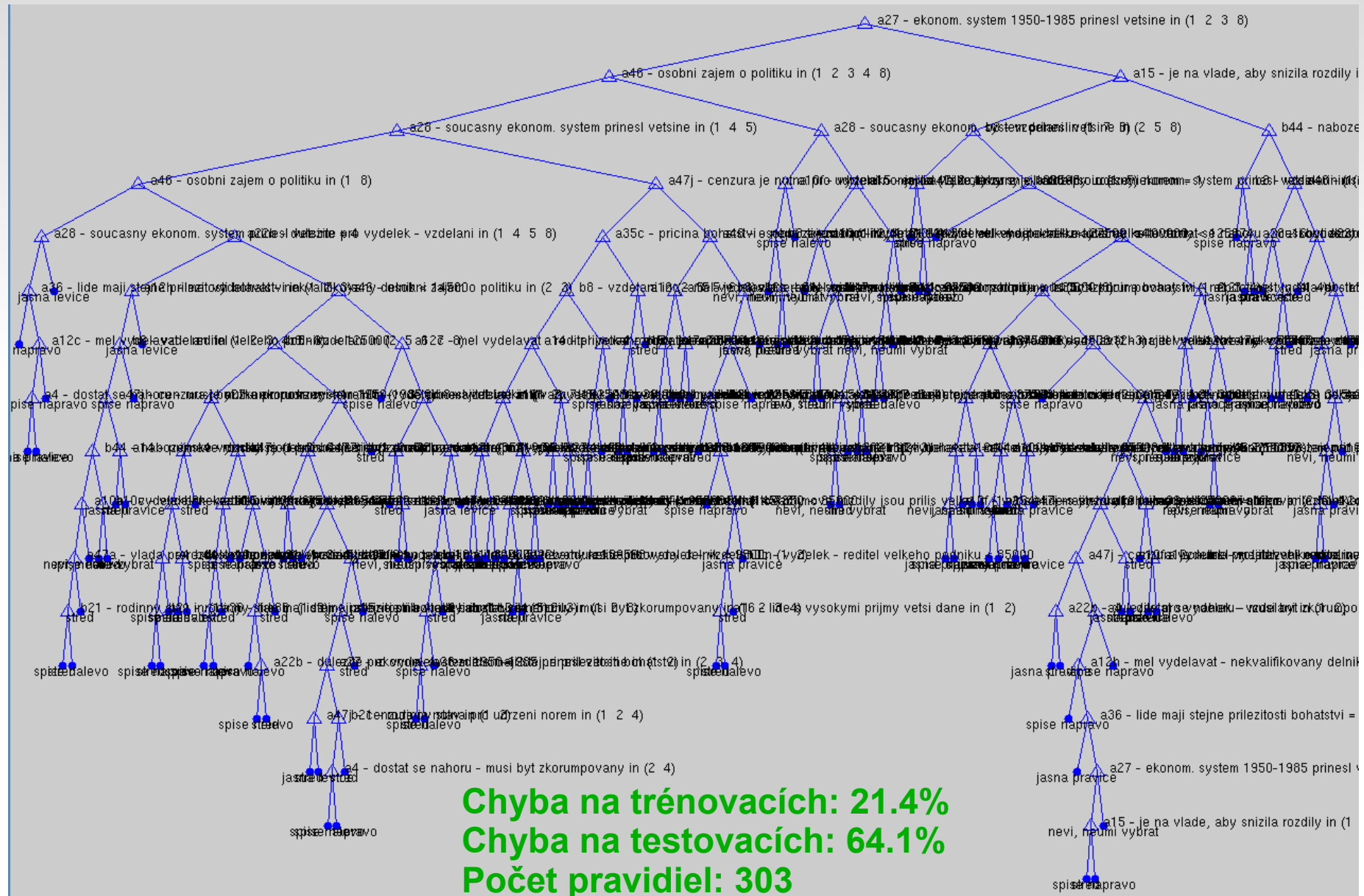


Výsledný strom s orezáním (if-then pravidlá)

Decision tree for classification

- 1 if a46 - osobni zajem o politiku in {1 2 3} then node 2 else node 3
- 2 if a28 - soucasny ekonom. system prinesl vetsine in {1 2 3} then node 4 else node 5
- 3 class = BEZ ODPOVEDI
- 4 if a28 - soucasny ekonom. system prinesl vetsine in {1 2} then node 6 else node 7
- 5 if a27 - ekonom. system 1950-1985 prinesl vetsine in {1 2 8} then
node 8 else node 9
- 6 class = ODS
- 7 if b8 - vzdelani in {1 5 7 8} then node 10 else node 11
- 8 class = CSSD
- 9 if a47a - vlada prerozdelovat prijmy in {1 2 3} then node 12 else node 13
- 10 class = ODS
- 11 if a47a - vlada prerozdelovat prijmy in {1 2 3} then node 14 else node 15
- 12 class = BEZ ODPOVEDI
- 13 class = CSSD
- 14 if a12h - mel vydelaavat - nekvalifikovany delnik<7500 then node 16 else node 17
- 15 class = BEZ ODPOVEDI
- 16 class = ODS
- 17 if a15 - je na vlade, aby snizila rozdily in {1 3 5} then node 18 else node 19
- 18 class = BEZ ODPOVEDI
- 19 class = CSSD

Výsledný strom (orientácia) - bez orezania



Niektoré výsledky

Čo je / má / si myslí respondent

Nemá záujem o politiku

Ktorú stranu by volil?

neodpovedal / nevedel

Súčasný ekonomický systém priniesol
oveľa viac dobrého až rovnako veľa dobrého
ako systém v rokoch 1950-1985

ODS

Súčasný ekonomický systém priniesol
menej až oveľa menej dobrého /
/ nevedel odpovedať

ČSSD

Ekonomický systém v rokoch 1950-1985
priniesol viac dobrého / nevedel odpovedať

ČSSD

Rímskokatolícke alebo českobratrské
naboženské vyznanie

KDU-ČSL

Riaditeľ veľkého podniku by mal zarábať < 65000

ČSSD

Riaditeľ veľkého podniku by mal zarábať >= 65000

ODS

Vzdelanie je pre zárobok podstatné
až veľmi dôležité

ODS