

Viacrozmerná štatistická analýza jednotlivých druhov zločinu v Európe

Anotácia: V článku analyzujeme dáta z Eurostatu popisujúce 6 druhov trestných činov v európskych štátoch, konkrétne násilné usmrtenie osôb, násilné trestné činy, lúpeže, domáce vlámania, krádeže motorových vozidiel a obchod s drogami. Úlohou bolo nájsť podobné štáty vzhľadom na všetky parametre súčasne. Výsledné zhľuky štátov zodpovedajú zväčša geografickým a historicko-politickým vzťahom medzi jednotlivými štátmi.

Kľúčové slová: zločin, analýza zhľukov, Európa.

Úvod

Eurostat je štatistický úrad Európskej únie. Jeho úlohou je spracovávať najmä dáta z členských (prípadne aj kandidátskych) štátov Európskej únie. Na stránke¹ musí pravidelne zverejňovať rôzne štatistické ukazovatele jednotlivých štátov. Oblasť, na ktorú sa zameriava, sú nasledujúce: regionálne štatistiky, ekonomické a finančné štatistiky, populácia a sociálne podmienky, priemysel, obchod a služby, poľnohospodárstvo a rybolov, zahraničný obchod, doprava, životné prostredie a energetika, veda a výskum.

V časti populácia a sociálne podmienky zverejňuje v sekcii zločin a súdnictvo každoročne nasledujúce prehľady: všetky zločiny zaznamenané políciou, vraždy v mestách, polícia, väzni.

V tejto súvislosti sme sa rozhodli analyzovať celkový prehľad rôznych foriem zločinu v krajinách Európy. Najaktuálnejšie údaje boli z roku 2008, všetky nasledujúce úvahy sa preto budú týkať iba tohto roku.

V² je uvedený prehľad celkového počtu všetkých trestných činov, ktoré zaznamenala v roku 2008 polícia každého štátu. Menej závažné trestné činy (priestupky) sú zo zoznamu vylúčené. Ďalej nasleduje týchto 6 ukazovateľov:

násilne zabité osoby – vrátane vražd, zabíj z nedbalosti, eutanázií a zabitých novorodencov,
násilné trestné činy – zahŕňajú všetky formy fyzických napadnutí vrátane sexuálnych útokov a násilných lúpeží,

lúpeže – tvoria podmnožinu násilných trestných činov; patrí sem aj prepadnutie a násilné krádeže, vreckové krádeže a vydieranie nie sú v tejto skupine zahrnuté,

vlámania do domov,

krádeže motorových vozidiel,

obchod s drogami – zahŕňa nezákonné prechovávanie drog, pestovanie a výrobu drog, dovoz, vývoz a financovanie obchodu s drogami.

Úlohou nášho článku je porovnať krajiny Európy z hľadiska podobnosti v uvedených šiestich ukazovateľoch. Neberieme do úvahy celkový počet trestných činov, ale podobnosť vo všetkých šiestich typoch trestných činov zároveň (ide teda o celkové porovnanie kriminálneho správania obyvateľstva).

Metodika

Pred štatistickou analýzou sme najskôr museli upraviť dáta v štatistickom voľne stiahnuteľnom programe R³. Keďže získané údaje predstavujú len počty jednotlivých

¹ <http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes>

² http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=crim_gen&lang=en

trestných činov, museli sme údaje normalizovať vzhľadom na celkový počet obyvateľov danej krajiny. Použili sme údaje o celkovej populácii jednotlivých štátov v roku 2008⁴. Počty trestných činov sme následne prepočítali na 100 000 obyvateľov každého štátu.

Na porovnanie krajín sme použili údaje o všetkých štátoch Európy (členských, kandidátskych aj nečlenských štátoch EÚ), ktoré boli na stránke k dispozícii. Pri niektorých štátoch chýbali niektoré hodnoty premenných (neboli políciou zaznamenané). Tieto sme preto museli vylúčiť z ďalších analýz. Boli to Írsko, Holandsko, Slovinsko, Island, Švajčiarsko a Čierna Hora. Výsledný súbor dát všetkých krajín s úplnými údajmi a nim zodpovedajúcimi šiestimi ukazovateľmi prepočítanými na 100 000 obyvateľov uvádzame v tabuľke 1.

Ako príklad interpretácie získaných čísel môžeme uviesť prvý údaj v tabuľke – v roku 2008 pripadalo na každých 100 000 obyvateľov Belgicka 1,82 vraždy. Treba poznamenať, že v článku nerozoberáme dôvody a možné príčiny niektorých údajov. Rovnako si treba uvedomiť, že ide len o počty udalostí, nie ich závažnosť, prípadne počet osôb, ktorý ich spáchal. Údaje Eurostatu sa opierajú o dáta z jednotlivých štátov, kde ich prvotne získala polícia. Všimnime si napríklad obchod s drogami (šiesty stĺpec), kde v prepočte na 100 000 obyvateľov dosahuje najväčšiu hodnotu Lichtenštajnsko a najmenšiu paradoxne Albánsko.

Tab. 1: Uvažované krajiny so všetkými dostupnými údajmi a nim zodpovedajúce počty trestných činov pripadajúcich na 100 000 obyvateľov

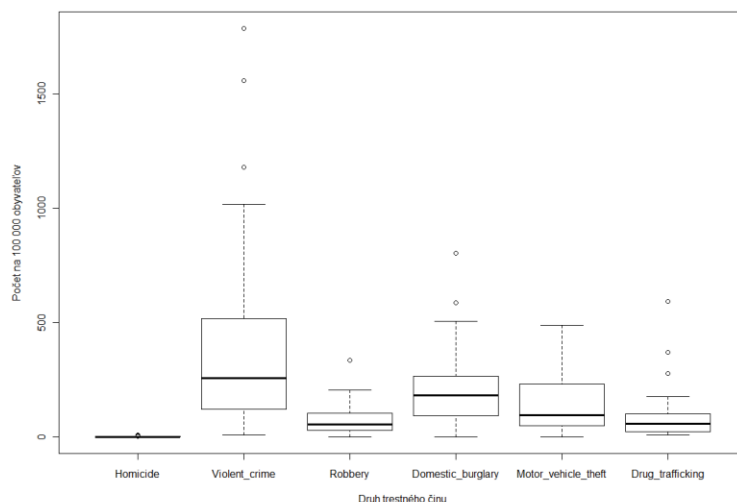
	Homicide	Violent_crime	Robbery	Domestic_burglary	Motor_vehicle_theft	Drug_trafficking
Belgium	1.82	1016.99	205.24	587.94	197.68	118.24
Bulgaria	2.25	111.75	37.54	261.51	5.63	37.39
Czech_Republic	1.95	175.19	44.71	87.77	181.99	27.09
Denmark	1.44	455.24	196.26	803.06	489.50	59.11
Germany	0.80	256.50	60.71	131.70	108.29	68.00
Estonia	6.26	677.29	67.79	247.66	77.18	116.19
Greece	1.05	100.06	27.62	393.71	69.86	87.86
Spain	0.90	257.42	205.78	181.38	212.69	32.18
France	1.35	533.96	171.62	267.56	340.36	9.86
Italy	1.10	245.89	108.25	256.76	385.72	57.17
Cyprus	1.14	40.42	7.86	326.38	226.29	98.83
Latvia	5.24	84.90	63.46	155.80	82.26	110.62
Lithuania	9.03	129.87	102.54	180.49	75.84	23.56
Luxembourg	1.45	660.81	53.74	357.79	70.90	277.59
Hungary	1.46	328.86	31.14	191.52	164.64	54.39
Malta	1.46	93.10	35.83	164.27	97.00	43.87
Austria	0.55	1558.11	57.53	224.17	108.78	23.80
Poland	1.21	136.75	55.32	82.59	46.36	8.70
Portugal	1.17	230.90	196.43	279.38	238.04	35.13
Romania	2.29	31.78	11.45	47.77	10.94	16.82
Slovakia	1.74	167.19	25.38	39.21	76.56	9.70
Finland	2.49	793.04	32.00	112.78	260.43	106.76
Sweden	0.89	1180.97	97.02	197.93	486.96	84.91
Liechtenstein	2.83	263.04	2.83	263.04	45.25	593.96
Norway	0.72	503.42	33.73	171.52	251.23	370.41
Croatia	1.67	292.69	28.38	61.18	43.44	177.67
Macedonia	1.76	44.25	22.39	65.81	28.41	20.29
Turkey	3.90	237.05	12.54	98.76	25.48	21.77
Albania	4.38	7.60	1.58	0.95	0.88	8.64
Serbia	1.97	401.63	335.58	86.25	51.23	82.13
United_Kingdom	1.28	1787.92	137.86	504.97	264.49	66.61

Ďalej sme zobrazili boxploty všetkých premenných (obr. 1). Z boxplotov vidieť, v akých škálach sa pohybujú premenné. Napríklad prvá premenná (násilné zabitia osôb) dosahuje veľmi malé hodnoty a malý rozptyl. Naopak, druhá premenná (násilné trestné činy) sa vyznačuje veľmi veľkými hodnotami a rozptylom. Z tohto dôvodu musíme premenné v programe R najskôr štandardizovať, aby sme ich previedli do rovnakej mierky, kde budú mať všetky rovnakú strednú hodnotu a rozptyl. Údaje sú teraz pripravené na objektívne

³ R DEVELOPMENT CORE TEAM. R: *A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna : Austria. ISBN 3-900051-07-0, url <http://www.R-project.org/>. [cit. 21.8.2011].

⁴ http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_pjan&lang=en

posúdenie podobnosti všetkých štátov súčasne. Na tento účel sa v štatistike používa zhluková analýza, popíšme si preto najskôr jej princípy.



Obr. 1: Boxploty jednotlivých premenných, výstup programu R⁵

Zhluková analýza

Zhluková analýza je jednou z najpoužívanějších štatistických metód používanou v situáciách, keď objekty v datasúbore vytvárajú zhľuky. Zhľukom sa rozumie množina podobných objektov na základe určitých vlastností, ktoré sme na každom objekte namerali. Podobnosť je v matematickom zmysle slova definovaná ako blízkosť daných objektov v ich priestore. Každý objekt si možno predstaviť ako bod v priestore a objekty s veľmi podobnými vlastnosťami si budú ako body veľmi blízke. V zhľukovej analýze existuje niekoľko algoritmov, ktoré hľadajú zhľuky v dátach. My sme zvolili tzv. samoorganizujúce sa mapy najmä pre zrozumiteľný a ľahko interpretovateľný výstup, ktorý ponúkajú.

Samoorganizujúce sa mapy

Samoorganizujúce sa mapy (skrátene SOM)⁶ sú typom umelej neurónovej siete bez učiteľa (teda príslušnosť objektov ku zhľukom nemusíme zadať vopred; hovoríme, že sieť sa učí sama z dát). Navrhol ich v roku 1982 Fín Teuvo Kohonen. Ich hlavný význam spočíva v zobrazení viacrozmerného priestoru (najčastejšie viac ako trojrozmerného, teda takého, ktorý sa nedá priamo vizualizovať) do priestoru s nižšou dimenziou (najčastejšie dvojrozmerného). Takéto zobrazenie zachováva lokálnu topológiu, teda objekty, ktoré sú si blízke v priestore s vyššou dimenziou, si budú blízke aj v priestore s nižšou dimenziou (týmto priestorom je spomínaná Kohonenova mapa).

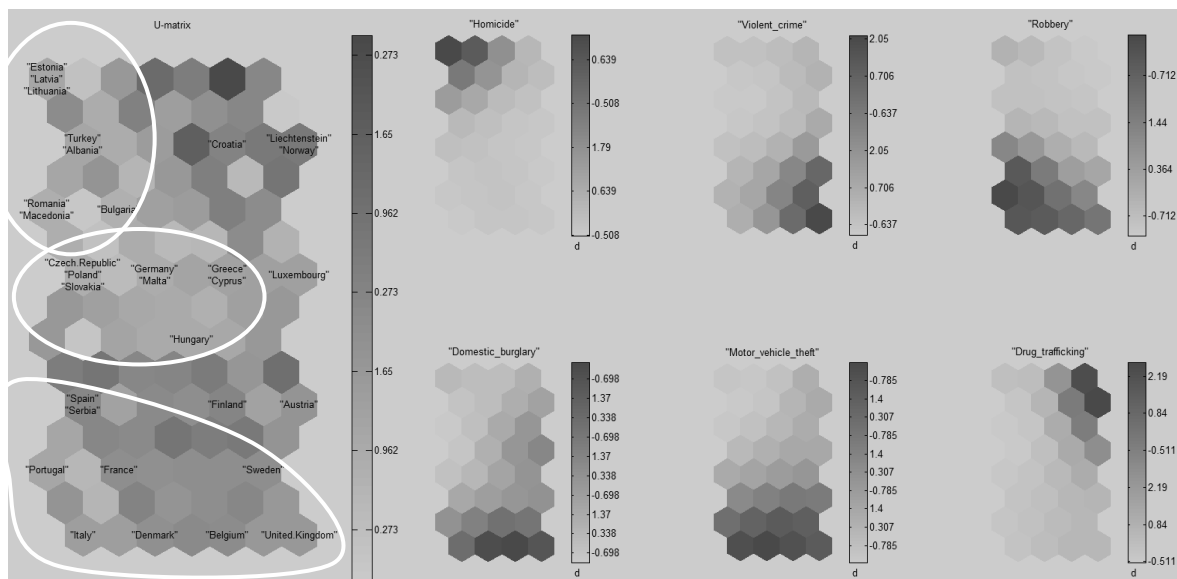
Pôvodná myšlienka vizualizovať dáta sa neskôr použila na zhľukovanie objektov. Vďaka zachovávaúcej sa topológii môžeme pomocou Kohonenovej mapy identifikovať, ktoré objekty sú si blízke (teda podobné). Výhoda SOM oproti iným zhľukovacím metódam je najmä to, že nepotrebujú žiadne parametre, akými sú napríklad počet zhľukov.

⁵ R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, url <http://www.R-project.org/>, 2011

⁶ KOHONEN, T. Self-organized formation of topologically correct feature maps, In *Biological Cybernetics* 43, 1982, str. 59–69.

Algoritmus, akým sa mapa organizuje, možno popísať takto: Používateľ najskôr zvolí veľkosť mapy, ktorú predstavuje počet neurónov, a typ mapy. Štatistické balíky pritom najčastejšie ponúkajú list (na dvojrozmernú vizualizáciu), toroid alebo cylinder (oba na trojrozmernú vizualizáciu). Ďalej treba nastaviť typ okolia, teda ktoré neuróny má algoritmus považovať za blízke danému neurónu. Rozšírené je najmä šesťuholníkové a pravouhlé okolie.

Hlavným vizualizačným výstupom algoritmu SOM je tzv. U-matica (obr. 2), ktorá slúži na identifikáciu počtu zhlukov. Vznikne tak, že k pôvodnej mape, ktorej veľkosť si určíme na začiatku, sa pridajú ďalšie bunky, a to tak, že medzi každými dvomi susednými neurónmi pribudne bunka, ktorej bude zodpovedať vzdialenosť medzi týmito neurónmi. Farba buniek U-maticy teda odzrkadľuje vzdialenosti medzi neurónmi.



Obr. 2: U-matica (vľavo) a matice jednotlivých premenných (vpravo). Výstup z programu MATLAB ⁷, s použitím balíka SOM toolbox ⁸.

Interpretácia U-maticy je nasledujúca: Štáty, ktoré sa zobrazili na blízke neuróny (teda blízko seba), sú si podobné vzhľadom na všetkých 6 premenných súčasne. Naopak, čím sú štáty na mape SOM ďalej od seba, tým sa viac odlišujú. Z matic jednotlivých premenných na tom istom obrázku vieme navyše vyčítať, čo spája krajiny, ktoré sa zobrazili blízko seba. Odtieň farby na maticiach jednotlivých premenných odzrkadľuje samotnú hodnotu týchto premenných (teda na miestach, kde je čierna farba, sú veľké hodnoty premennej a tam, kde je biela, sú nízke hodnoty).

Do ľavého horného rohu U-maticy sa zobrazili štáty Estónsko, Lotyšsko a Litva. Všetky tri sa nachádzajú v rovnakom geografickom regióne a navyše sú zviazané aj veľmi podobnými spoločenskými podmienkami. Toto sa odzrkadlilo aj na samotných formách zločinu, ktoré sú vo všetkých troch krajinách veľmi podobné. Z prvej matice (úmyselné zabitia osôb) vidieť, že v týchto krajinách sú najvyššie počty násilne usmrtených osôb v prepočte na 100 000 obyvateľov. Vysoké hodnoty majú aj štáty Turecko, Albánsko, Macedónsko, Rumunsko a Bulharsko, ktoré sa na mape zobrazili nižšie, čo naznačuje ich podobnosť s Estónskom, Lotyšskom a Litvou.

⁷ MATLAB version 7.6. Natick, Massachusetts: The MathWorks Inc., 2010.

⁸ VESANTO, J., HIMBERG, J., ALHONIEMI, E., PARHANKANGAS, J. SOM toolbox for Matlab 5, Helsinki University of Technology, Espoo, 2000.

Ďalší zhhluk štátov s podobnými vlastnosťami predstavujú Slovensko, Česko, Poľsko, Maďarsko, Nemecko, Malta a Grécko. Spájajú ich najmä stredne vysoké počty lúpeží a krádeží motorových vozidiel.

V spodnej časti U-matice sa zobrazili prevažne krajiny západnej a juhozápadnej Európy. Dosahujú vysoké počty domových vlámaní, lúpeží, krádeží motorových vozidiel a násilných trestných činov a naopak, nízke hodnoty premenných obchod s drogami a úmyselné usmrtenia osôb (v prepočte na 100 000 obyvateľov).

Zvyšné krajiny nemožno zaradiť do homogénnych zhhlukov. Fínsko s Rakúskom majú nižšie počty trestných činov pri všetkých premenných, preto sú zobrazené blízko seba. V pravej hornej časti sa nachádzajú zase štáty Chorvátsko, Nórsko, Luxembursko a Lichtenštajnsko. Vysoké hodnoty premennej dosahuje najmä obchod s drogami, treba ale opätovne zdôrazniť, že ide o celkový počet prípadov, ktoré zaznamenala polícia daného štátu v prepočte na každých 100 000 obyvateľov štátu.

Záver

Úlohou tejto práce bolo posúdiť celkové kriminálne správanie obyvateľstva vybraných krajín Európy. Skúmali sme len podobnosť vzhľadom na 6 vybraných ukazovateľov kriminálneho správania. Nájdene zhluky zväčša zodpovedajú geograficko-politickým regiónom, čo znamená, že podobné kriminálne správanie obyvateľstva niektorých štátov vyplýva z ich geografickej blízkosti a z podobného politicko-historického vývoja.

PodĎakovanie: Táto práca bola podporená grantom VEGA 1/0103/10.

Literatúra

KOHONEN, T. Self-organized formation of topologically correct feature maps, In *Biological Cybernetics* 43, 1982, str. 59–69.

MATLAB version 7.6. Natick, Massachusetts: The MathWorks Inc., 2010.

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, url <http://www.R-project.org/>, 2011

VESANTO, J.- HIMBERG, J.- ALHONIEMI, E.-PARHANKANGAS, J. SOM toolbox for Matlab 5, Helsinki University of Technology, Espoo, 2000.

<http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes>

http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=crim_gen&lang=en

http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_pjan&lang=en

Key words: crime, cluster analysis, Europe

Summary

We analyze similarities of european countries with respect to 6 features concerning different types of crime, namely homicide, violent crime, robbery, domestic burglary, motor vehicle theft and drug trafficking (data downloaded from Eurostat). Results correspond to geographical and historical relations between countries.

Mgr. Miroslav Sabo

Stavebná fakulta STU v Bratislave

Katedra matematiky a deskriptívnej geometrie
e-mail: miro165sabo@gmail.com,
Tel.: 0902 167 061

RNDr. Tatiana Hajdúková
Akadémia Policajného zboru v Bratislave
Katedra informatiky a manažmentu
e-mail: tatiana.hajdukova@minv.sk,
Tel.: 0961 057 297

Recenzent: doc. RNDr. Ľudmila Gregušová, CSc.