

UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Katedra Geografie



Bc. Václav Štěpánek

**VÝVOJ PROSTOROVÝCH STRUKTUR ČESKÉ
REPUBLIKY: APLIKACE METOD FORMÁLNÍ
REGIONÁLNÍ TAXONOMIE**

Diplomová práce

Vedoucí diplomové práce:

doc. Mgr. Pavel Klapka, Ph.D.

Olomouc 2022

Bibliografický záznam

Autor (osobní číslo): Bc. Václav Štěpánek (R190599)

Studijní obor: Regionální geografie

Název práce: Vývoj prostorových struktur České republiky: aplikace metod formální regionální taxonomie

Title of thesis: Development of spatial structures of the Czech Republic: application of methods of formal regional taxonomy

Vedoucí práce: doc. Mgr. Pavel Klapka, PhD.

Rozsah práce: 80 stran, 27 vázaných příloh

Abstrakt: Cílem práce je na základě cenzových dat obecní úrovně České republiky ověřit chování vybraných metod formální regionální taxonomie. Regionální systémy jsou kompilovány pomocí numerické taxonomie s využitím podpůrných statistických metod (standardizace, analýza hlavních komponent). Výsledky jsou hodnoceny subjektivní geografickou logikou a objektivními validačními indexy. Mezi průkazné regionálně-taxonomické metody se řadí metoda úplného spojení, Wardova metoda, metoda k-průměrů a algoritmus SKATER. Ve zkoumaném území jsou zjištěné vztahy interpretovány jako centrum-zázemí, jádro-periferie, populační póly a makroregionální dichotomie.

Klíčová slova: region, klasifikace, shlukování, prostorové struktury, Česká republika

Abstract: This thesis aims to examine behavioral patterns of chosen formal regional-taxonomy methods based on municipality-level census data of the Czech Republic. The regional systems are compiled by employing numerical taxonomies, utilizing auxiliary statistical methods, such as standardization and principal component analysis. The results are evaluated by subjective geographical logic and objective validation indexes. The *Complete-linkage method*, *Ward's method*, *k-means clustering* and *SKATER algorithm* have proven to be the most conclusive methods tested. The discovered relationships within the examined area are interpreted as centre-hinterland, core-periphery, population poles, and macroregional dichotomy.

Keywords: region, classification, clustering, spatial structures, Czech Republic

Děkuji doc. Mgr. Pavlu Klapkovi, PhD. za cenné rady, které mi poskytl během tvorby této diplomové práce.

Prohlašuji, že jsem diplomovou práci vypracoval samostatně s vyznačením všech použitých pramenů a spoluautorství. Souhlasím se zveřejněním diplomové práce podle zákona č. 111/1998 Sb., o vysokých školách, ve znění pozdějších předpisů. Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, ve znění pozdějších předpisů.

V Olomouci dne 19. 04. 2022

.....

Václav Štěpánek

UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Akademický rok: 2019/2020

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: **Bc. Václav ŠTĚPÁNEK**
Osobní číslo: **R190599**
Studijní program: **N1301 Geografie**
Studijní obor: **Regionální geografie**
Téma práce: **Vývoj prostorových struktur České republiky: aplikace metod formální regionální taxonomie**
Zadávací katedra: **Katedra geografie**

Zásady pro vypracování

Hlavním cílem práce je ověřit chování vybraných metod formální regionální taxonomie a validitu výsledků. Volba regionalizačních kritérií je v kompetenci autora (za asistence s vedoucím diplomové práce). Zkoumány budou minimálně dva časové okamžiky (např. SLDB 2011 a 2001), aby mohl být zachycen vývoj prostorových struktur České republiky, tedy vybraných zkonstruovaných regionálních systémů. Výsledky práce mohou být interpretovány v rámci teoretických konceptů, jako je například jádro-periferie.

Rozsah pracovní zprávy: **20 000 – 24 000 slov**
Rozsah grafických prací: **Podle potřeb zadání**
Forma zpracování diplomové práce: **tištěná**

Seznam doporučené literatury:

- Bernard, J., Šimon, M. (2017): Vnitřní periferie v Česku: Multidimenzionalita sociálního vyloučení ve venkovských oblastech. *Sociologický časopis/Czech Sociological Review* 53 (1), 3728.
- Berry, B. J. L. (1961): A method for deriving multi-factor uniform regions. *Przegląd geograficzny* 33 (2), 263-282.
- Bezák, A. (1988): Regionálne typy sociálno-priestorovej štruktúry Bratislavy. *Geografický časopis* 40 (4), 311-328.
- Bezák, A. (1993): Problémy a metódy regionálnej taxonomie. *Geographia Slovaca* 3. GÚ SAV, Bratislava.
- Borgatti, S. P., Everett, M. G. (1999): Models of core/periphery structures. *Social Networks* 21, 375-395.
- Del Campo, C., Monteiro, C. M. F., Soares, J. O. (2008): The European regional policy and the socio-economic diversity of European regions: A multivariate analysis. *European Journal of Operational Research* 187 (2), 600-612.
- Duque, J. C., Ramos, R., Surinach, J. (2007): Supervised regionalization methods: a survey. *International Regional Science Review* 30 (3), 195-220.
- Erlebach, M., Halás, M., Daniel, J., Klapka, P. (2019): Is there congruence in the spatial patterns of regions derived from scalar and vector geographical information? *Moravian Geographical Reports* 27 (1), 27-44.
- Everitt, B. S., Landau, S., Leese, M., Stahl, D. (2011): *Cluster analysis*. 5th edition. John Wiley & Sons, Chichester.
- Fischer, M. M. (1980): Regional taxonomy: a comparison of some hierarchic and non-hierarchic strategies. *Regional Science and Urban Economics* 10 (4), 503-537.
- Gan, G., Ma, C., Wu, J. (2007): *Data clustering: theory, algorithms, and applications*. ASA-SIAM Series on statistics and applied probability. Philadelphia, SIAM, Alexandria, VA, ASA.
- Hámpel, M. (1971): *Teorie komplexity a diferenciace světa*. Univerzita Karlova, Praha.
- Hámpel, M. (2005): *Geografická organizace společnosti v České republice: transformační procesy a jejich obecný kontext*. DemoArt, Praha.

HAMPL, M. (2018): Hierarchické formy uspořádání nerovnoměrností v realitě: hledání pravidelností a problémy explanace. Informace ČGS 37 (1), 1??3.
Heřmanová, E. (1991): Vybrané vícerozměrné statistické metody v geografii. Univerzita Karlova ? SPN, Praha.
Johnston, R. J. (1976): Classification in geography. CATMOG 6. GeoAbstracts, Norwich.
Klapka, P. (2019): Regiony a regionální taxonomie: koncepty, přístupy, aplikace. Univerzita Palackého v Olomouci, Olomouc.

Vedoucí diplomové práce: **doc. Mgr. Pavel Klapka, Ph.D.**
Katedra geografie

Datum zadání diplomové práce: **23. června 2020**
Termín odevzdání diplomové práce: **10. dubna 2021**

L.S.

doc. RNDr. Martin Kubala, Ph.D.
děkan

prof. RNDr. Marián Halás, Ph.D.
vedoucí katedry

V Olomouci dne 23. června 2020

OBSAH

| | |
|---|----|
| 1 Úvod a cíle práce | 9 |
| 2 Teoretická východiska a definice základních pojmů | 11 |
| 2.1 Region | 12 |
| 2.1.1 Region v průběhu času | 12 |
| 2.1.2 Klasifikace regionů..... | 15 |
| 2.1.3 Formální region | 18 |
| 2.1.4 Hierarchie regionů..... | 20 |
| 2.2 Obecná charakteristika regionálně-taxonomických procedur..... | 22 |
| 2.2.1 Zásady regionální taxonomie | 22 |
| 2.3 Metody formální regionální taxonomie | 24 |
| 2.3.1 Hlavní metodické přístupy | 24 |
| 2.3.2 Operace s prostorovými informacemi | 25 |
| 2.3.3 Metody vícerozměrné analýzy | 27 |
| 2.3.4 Blízkost v taxonomickém prostoru..... | 28 |
| 2.3.5 Prostorová souvislost formálních regionálních tříd | 29 |
| 2.3.6 Optimální počet formálních regionálních tříd | 30 |
| 2.4 Stručná historie regionálního členění v Česku | 31 |
| 2.4.1 Rané období (pol. 19. století–2. sv. válka) | 31 |
| 2.4.2 Poválečné období (1945–1989)..... | 32 |
| 2.4.3 Nejnovější období (1993–současnost)..... | 34 |
| 3 Metodologie a data..... | 36 |
| 3.1 Zpracování dat | 37 |
| 3.2 Vybrané metody formální regionální taxonomie | 38 |
| 3.2.1 Hierarchické metody formální regionální taxonomie..... | 38 |
| 3.2.2 Nehierarchické metody formální regionální taxonomie | 39 |
| 3.3 Hodnocení výsledků | 41 |
| 3.4 Softwarové nástroje a datové zdroje..... | 42 |
| 3.5 Metodologické shrnutí..... | 43 |
| 4 Výsledky a diskuse | 44 |
| 4.1 Hierarchické aglomerativní metody | 46 |
| 4.1.1 Metoda jednoduchého spojení..... | 46 |
| 4.1.2 Metoda úplného spojení | 48 |
| 4.1.3 Metoda průměrová | 49 |
| 4.1.4 Metoda centroidní | 51 |
| 4.1.5 Wardova metoda | 52 |
| 4.2 Hierarchické divizivní metody | 54 |
| 4.2.1 Algoritmus SKATER | 54 |
| 4.3 Nehierarchické metody..... | 56 |
| 4.3.1 Metoda k-průměrů..... | 56 |
| 4.3.2 Metoda k-medoidů | 58 |
| 4.5 Validita formálních regionálních systémů | 59 |

| | |
|--------------------------|----|
| 4.6 Omezení výzkumu..... | 62 |
| Závěr..... | 63 |
| Summary..... | 64 |
| Reference..... | 65 |
| Seznam tabulek..... | 68 |
| Seznam obrázků..... | 69 |
| Seznam příloh..... | 70 |
| Přílohy..... | 71 |

1 Úvod a cíle práce

Potřeba identifikace částí geografické sféry, našeho zemského tělesa, patří společně s následným klasifikováním zjištěných prostorových objektů k přirozeným vlastnostem a schopnostem našeho druhu. Identifikace prostorových struktur, jejich shlukování a analýza je jednou z výsostných domén geografie, a v nadstavbě, regionální vědy.

Obecně je delimitace regionů jedním z mnoha praktických přínosů regionální geografie. Identifikace formálních regionů může přinášet rozšířené poznatky o prostorové organizaci zkoumaného území. Formálně vymezené regiony mohou sloužit jako identifikátory uniformity zkoumané oblasti. Regionální typologie přispívají k vyšší efektivitě plánovacích procesů a strategického rozhodování. Vymezení individuálních regionů naopak slouží jako doklad unikátnosti geografického prostoru a dále napomáhá rozvíjet vztah obyvatelstva k území a jeho hodnotám. V geografii se přístupy k identifikaci regionů s průběhem času mění, v současné době jsou ovšem nejviditelnější způsoby určování rozsahu regionů založené na datech, popřípadě velkých datech (big data).

Práce s velkým množstvím dat a jejich interakcemi vyžaduje (nejen) v geografii rozsáhlé výpočetní kapacity, na které obyčejně nestačí lidské schopnosti. Obracíme se proto na přístrojové vybavení, které nám umožňuje prozkoumat nepřehledné množství variant v přijatelném časovém rámci. Přesto, pro výpočet všech dostupných kombinací umístění regionu nejsou ani možnosti současné komerční výpočetní techniky dostatečné. Pomáháme si proto algoritmy a funkcemi, které odhadnou nejvýhodnější varianty požadovaného umístění regionu podle zadaných parametrů. Jako lidské bytosti dáváme přístrojům instrukce, je ovšem nutné mít informaci o jejich zpracování a výsledku – tedy znalost vhodné metody pro zadaný účel. Vyjma zjevné erudice v oblasti dat a metod jejich zpracování je vítanou schopností výzkumníka-geografa znalost probádaného území. Z těchto pozic pramení vymezení cílů této diplomové práce.

Hlavním cílem této práce je ověřit chování prostorových struktur České republiky pomocí vhodně zvolených metod formální regionální taxonomie. Podle původního zadání diplomové práce měla být zkoumána dynamika regionálních systémů skrze minimálně dva zvolené časové okamžiky. V průběhu zpracování došlo ke snížení časového rozsahu, a to zejména pro absolutní neporovnatelnost dvou navazujících datových souborů Sčítání lidu, domů a bytů. Zájmové území této diplomové práce je tedy určeno hranicemi České republiky v územních hranicích platných k roku 2011.

Před samotným hodnocením chování metod formální regionální taxonomie je třeba dosáhnout dílčích cílů:

- na základě prozkoumané literatury vybrat dobře zpracovatelné a porovnatelné metody formální regionální taxonomie
- určit optimální úroveň územního členění tak, aby byla zajištěna dostupnost statistických dat a zároveň bylo možné provádět průkazné regionálně-taxonomické procedury
- shromáždit datové zdroje a vhodně je upravit pro další použití
- snížit počet využívaných proměnných pro výpočetně náročné operace
- definovat optimální počet finálních regionálních tříd pro hierarchické-divizivní a nehierarchické metody
- vybrat komplementární softwarové nástroje pro provedení samotných regionálně-taxonomických operací
- vybrat vhodné metody hodnocení vzniklých regionálních systémů

Dosažené poznatky jsou shrnuty v tabelárních, grafických a kartografických výstupech. Hodnocení výsledků je v souladu s teoretickými koncepty a ustanovenou regionální strukturou České republiky.

2 Teoretická východiska a definice základních pojmů

Tato kapitola se zabývá úvodem do problematiky regionů. Obsahuje obecné informace o regionu jako objektu geografického studia, vývoji konceptu regionu a jeho typologii. Dále nastiňuje možnosti práce s regionem a způsoby jeho konstrukce, včetně vývoje regionalizací v českých zemích. Speciální péče je věnována pojmům a aspektům delimitace formálních regionů, kterými se zabývá i výzkumná část této práce.

2.1 Region

Následující kapitola se zabývá konceptem regionu v průběhu času, klasifikací regionů, hierarchickými strukturami organizace prostoru a stručnou charakteristikou formálního regionu.

2.1.1 Region v průběhu času

Region je jedním z nejdůležitějších pojmů v geografii, který vychází z charakteru objektu studia této vědecké disciplíny – planety Země, nehomogenity její geografické sféry. Studium regionů je geografickou subdisciplínou, která se objevuje již ve starověku; náplň studia je ovšem po staletí předmětem změn. Koncept regionu je z období *antické chorologie* (nauky o rozlišování jednotlivých území), vyvíjející se z pojmu *regio* (hranice), přes výrazy *regnum* (království) a *rex* (král) až po *regere*, což je význam označující území ovládané určitou suverénní silou. Z tohoto původně územně-politického pojmu se postupně stal výraz obecnější a lépe posloužil tehdejšímu výhradnímu účelu geografie – popisu a identifikaci regionů.

Celospolečenská revoluce předminulého století se změnou vědeckého myšlení přinesla i proměnu náplně pojmu region. V rámci 19. století tedy existují tři přístupy (dnes považovány již za historické) ke studiu regionů, a to z hlediska vztahů mezi environmentálními a socio-kulturně-ekonomickými charakteristikami prostředí. Každý z těchto přístupů odpovídá myšlenkovému proudu, filozofickému diskurzu doby, týkajícího se přístupu k vědě obecně. Nejstarším z nich je *environmentální determinismus*. Nosnou myšlenkou prvního přístupu je předurčenost kulturního prostředí podmínkami přírodními, kdy se společnost, kultura a ekonomika přizpůsobuje a váže k životnímu prostředí. Mezi významné představitele deterministického paradigmatu řadíme např. Karla Rittera nebo Alexandra von Humboldta. Jako odkazy prvního z přístupů lze uvést například nechvalně proslulý *lebensraum* (koncept zneužitý pro nacistickou územní expanzi) nebo tzv. *přirozený region* (území charakterizované konkrétními fyzicko-geografickými vlastnostmi). Termín *lebensraum*, a jeho vliv, spojujeme s pangermánskou geografickou školou (Fridrich Ratzel, Rudolf Kjellen a Karl Haushofer); koncept přirozeného regionu rozpracoval např. Andrew Herbertson nebo Wladimir Köppen (KLAPKA, 2019, stránky 18–19). Tento přístup je odkazem na určitou rozsáhlost zemského povrchu a projevil se zejména v případě velkých regionů (převážně přírodních).

Druhé z pojetí přístupu ke studiu regionů vychází z *environmentálního posibilismu*. Region je v tomto případě rámcově ovlivňován environmentálními předpoklady, větší význam pro jeho formování má ovšem společnost. Dochází tedy k relativně vyrovnanému působení přírodních a socio-kulturně-ekonomických podmínek. Posibilistické pojetí představuje zejména Paul Vidal de la Blache s konceptem *genre de vie* (způsob života obyvatel vytváří regionální identitu). Jak již příklad napovídá, posibilistické pojetí je vhodné spíše pro menší rurální oblasti. Posibilistický přístup (jeho představitelé) je mimo hodnocení determinace regionu přírodními podmínkami charakteristický *ideografickým* pojetím – identifikací unikátních a specifických znaků u jednotlivých regionů.

Posledním a nejmladším z přístupů vychází z *environmentálního indeterminismu*. Kontrastně k environmentálnímu determinismu, základní myšlenkou tohoto pojetí je nezávislost socio-kulturně-ekonomických předpokladů na přírodních podmínkách. Přístup je spojen s překrýváním přírodní krajiny *krajinou kulturní*, a projevoval se typicky spíše v socio-ekonomické geografii. Mezi známe představitelé indeterministického přístupu patří např. Carl Sauer (viz zmíněná kulturní krajina), Richard Hartshorne, Walter Christaller, August Lösch nebo Walter Isard. První představitelé indeterministického přístupu byli též zastánci *chorologie* (komplexní hodnocení vztahů v rámci regionu), původně ve své době prosazované Alfredem Hettnerem.

Klapka (2019) dále uvádí historickou (až současnou) diskusi ohledně objektivní existence regionu. Základním problémem, zarámovaným intencemi pozitivistické a postpozitivistické vědy, je otázka, zda může být region objektivně existující realitou nebo je konstruktem. Na základě poznatků např. Wróbel (1965) Klapka následně oba názory propojuje a označuje region za „generalizovaný, více či méně ohraničený komplexní časoprostorový systém, vykazující určitý typ organizační jednoty, která ho odlišuje od ostatních regionů“. Tento přístup je pravděpodobně nejvhodnější i z hlediska vymezení regionů – při identifikaci regionů předpokládáme jejich hypotetickou existenci (region jako objektivně existující realita), zároveň je však „konstrukce“ regionu výzkumným zadáním (region jako konstrukt).

Dle Haggetta (2001, stránky 367–373) existují čtyři důvody pro zájem geografie o regiony:

- zjednodušení objektivně existující reality na menší, lehce uchopitelné části,
- akceptace regionů jako diskriminovaných prostorových anomálií,
- vyhledávání specifických, obdobných regionů a stanovování obdobností, analogií,
- identifikace procesů zapříčiněných modulačním chováním unikátních regionů.

V postmoderní a postkvantitativní geografii, jejichž zaměření se v oblasti regionů týká převážně socioekonomické geografie, se objevuje pojem *percepční region*. V tomto případě je region vymezován individuálně každým jedincem; podobná vymezení různými jedinci se dají následně syntetizovat, což poté umožňuje zkoumání daného regionu. Odlišným způsobem k vymezování regionu přistupuje nová regionální geografie, jejíž postupy akcentují přínos regionu jako prostředku společenské interakce. Důležitým faktorem je tvorba *regionální identity*, která je odrazem časoprostorových vztahů uvnitř společnosti. Region je tedy namísto statických jevů konceptualizován pomocí dynamických procesů (KLAPKA, 2019).

Syntézou názorů různých autorů [např. Haggett (1965); Domański (1982); Bašovský & Lauko (1990); Johnston & kol. (2000); Mičian (2003)] činí Klapka a Tonev (2008) několik obecných závěrů, týkajících se současného pojetí konceptu regionu:

- regiony jsou navzájem se lišící části geografické sféry,
- region je část geografické sféry, která je vymezená na základě zvoleného kritéria (region je tedy areál platnosti zvoleného kritéria,
- region je část geografické sféry, která je menší než celá zájmová oblast výzkumu (např. svět, kontinent, stát, pohoří, povodí apod.), ale větší než konkrétní místo (každý region je pak složen se souboru menších prostorových jednotek a zároveň je součástí prostorové jednotky větší),
- region je nejlogičtější způsob organizace geografických informací.

Závěrem již tedy dodávají, že geografie se v souvislosti s regiony zabývá:

- jejich zařazením v různých klasifikačních systémech (podle odvětvového, metodologického, taxonomického a formálního hlediska),
- jejich postavením v rámci hierarchické struktury prostorových jednotek,
- jejich vnitřní strukturou,
- metodami jejich vymezování.

2.1.2 Klasifikace regionů

Klasifikace regionů se liší podle přístupů různých autorů. V této práci se zaměříme zejména na agregátní pojetí Klapky (2019) s případným doplněním konceptů podle jiných autorů. Dle Klapky je vhodné klasifikovat regiony podle čtyř základních kvalit: struktury, teorie, odvětví a metodologie. Z těchto znaků považuje za nejdůležitější hledisko struktury.

- Struktura (také forma) regionu odpovídá samotnému způsobu konstrukce regionu podle typu použitých dat (skalárních nebo vektorových). Skalárnímu typu dat tedy odpovídá region *formální (také homogenní nebo uniformní)*, kdežto pomocí dat vektorových je vytvářen region *funkční (také vektorové a heterogenní, někdy i nodální)*. Zde dochází k určité dichotomii oproti různým, zejména starším, pojetím (KLAPKA, 2019, str. 28). Jako další typ je zmiňován pojem *plánovací* region (FRIEDMANN, 1956) (WRÓBEL, 1965) (KLAPKA & TONEV, 2008) který je spíše aplikací regionu funkčního a měl by být tedy zvláštním případem. Lauko a Kasala (2009) nebo Mičian (2003) dále uvádějí příklad regionů *kontrastních*, které se ovšem vyskytují pouze v oblasti fyzické geografie. Mimo jiné souvisí s formou regionu i jeho tvar – v současnosti se zkoumá spíše jako kompaktnost a nepatří mezi zásadní charakteristiky.

Strukturní hledisko bylo od počátku zkoumání typologie regionů propojeno s hlediskem odvětvovým (viz bod 3). Problematice se, často bez odkazů na dnešní terminologii a způsob konstrukce regionů, významně věnovali například Johann Heinrich von Thünen (izolovaný stát), August Comte (intendantury), Halford John Mackinder a Paul Vidal de la Blache (koncept nodality), Charles Josiah Galpin (venkovské komunity), Charles Bungay Fawcett (provincie), Robert Walter Christaller, August Lösch a Walter Isard (centrální místa) a Robert E. Dickinson (zóny vlivu a metropolitní regiony). Oba pojmy (funkční a formální region) nebo jejich alternace se objevují až v dílech poválečných autorů jako Derwent Whittlesey, Richard Hartshorne a vzdáleně i Hans Carol. V 60. letech 20. století se stal přístup identifikace regionů na základě charakteru informací součástí hlavního proudu v rámci kvantitativní geografie; k formování diskurzu od té doby přispěli například Peter Hagett, David Grigg, Andrzej Wróbel, Brian Berry, Ronald Johnston nebo Jerzy Parysek.

Tato práce se zaměřuje výhradně na regiony formální a těm se věnují vybrané následující kapitoly.

- Hledisko teoretické (také metodologické) se opírá o historický soubor diskurzů v rámci umístění geografie jako vědní disciplíny: zda má dle klasického ideografického pojetí zkoumat unika a specifika, anebo podle novějšího nomotetického pojetí klást důraz na opakovatelnost a hledání společných znaků. Pokud vztáhneme danou problematiku na koncept regionu, lze podle ideografického přístupu identifikovat regiony *individuální (také jednotlivé, specifické)* a podle nomotetického přístupu regiony *typologické (také generické, obecné)*. Oba typy regionů z hlediska teoretického jsou vymezovány mírně odlišnými taxonomickými metodami: individuální regiony pomocí individuální regionalizace a typologické regiony skrze regionální typologii. Rozdílem je v tomto případě důraz na prostorovou souvislost jednotek – regionalizace požaduje neopakovatelnost regionů, a tedy prostorovou souvislost (sousedství) základních prostorových jednotek; naproti tomu regionální typologie vytváří regiony typologické bez nutnosti prostorové souvislosti (KLAPKA, 2019, stránky 29–30).

Tato práce se zaměřuje na tvorbu individuálních i typologických regionů.

- Odvětvové hledisko opět vychází z názorového střetu na postavení geografie v rámci systému věd. Problémem (a prostředkem) je v tomto případě šíře záběru geografického poznání. *Dualistické* pojetí geografie uznává fyzickou a socio-ekonomickou geografii jako dvě individuální vědní disciplíny; *monistické* pojetí dělitelnost nepřipouští. Výsledkem ovšem není rozdílný názor na základní typologii regionů dle odvětvového hlediska (regiony *fyzicko-geografické* a regiony *socio-ekonomické*), nýbrž vztah monistického a dualistického pojetí k tzv. *komplexně-geografickému* regionu. Takový region je vrcholnou syntézou fyzicko-geografického a socio-ekonomického regionu; pro potřeby současné regionální taxonomie je ovšem z široké škály důvodů okrajovým tématem (KLAPKA, 2019, stránky 33–36). Problémem komplexně-geografického regionu se v českém prostředí dlouhodobě zabýval Hampl, který svým pojetím tento typ ztotožnil s nejvyšším vývojovým stupněm regionu (systému) (HAMPL, 1971) (1998) (2002).

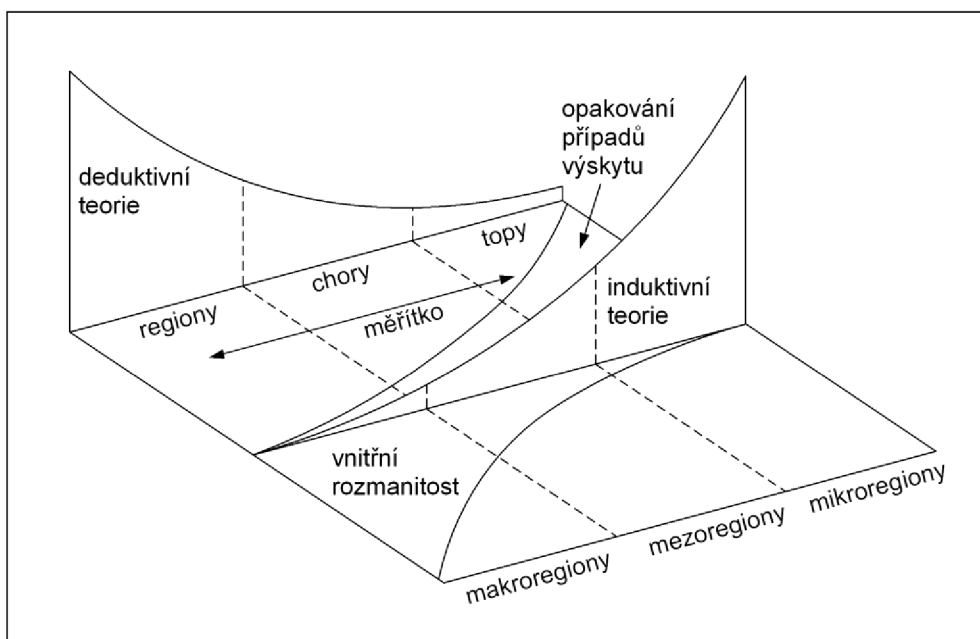
V rámci odvětvového hlediska je možné regiony klasifikovat i podle počtu složek, kterými jsou regiony definovány. Rozeznáváme tedy regiony *jednosložkové* a *vicesložkové*. V kontextu základního členění geografických disciplín uvažujeme u fyzicko-geografických odvětví například s regiony geomorfologickými, klimatickými nebo pedogeografickými; u socioekonomických podoborů lze regiony členit například na průmyslové, zemědělské nebo turistické. Zároveň je ovšem nutno

podotknout, že jednosložkové regiony nelze zpravidla za jednotlivé odvětvové regiony považovat – tyto samotné mají komplexnější charakter, který nelze postihnout pomocí jediného regionalizačního kritéria (složky) (KLAPKA, 2019).

Tato práce se zaměřuje na tvorbu vícesložkových socio-ekonomických regionů.

- U metodologického hlediska se snažíme postihnout účel regionu a motivaci pro jeho vymezení. Obecně lze podle takového rozdělení chápat region jako *objekt geografického výzkumu* nebo *nástroj geografického výzkumu*. Objektem geografického výzkumu se region stává například v této práci, kdy je podrobován metodám regionální taxonomie.

Problematiku charakteru regionů, jak fyzicko-geografických tak socio-ekonomických, a základních přístupů k jejich studiu výstižně shrnuje obrázek 1. Se vzrůstající hierarchickou úrovní regionů se zvyšuje jejich vnitřní rozmanitost a zároveň klesá počet opakujících se výskytů určitého regionálního typu (typologického regionu). Se vzrůstající hierarchickou úrovní regionů se rovněž snižuje uplatnění induktivních metod vymezení regionů a zvyšuje se podíl metod deduktivních.

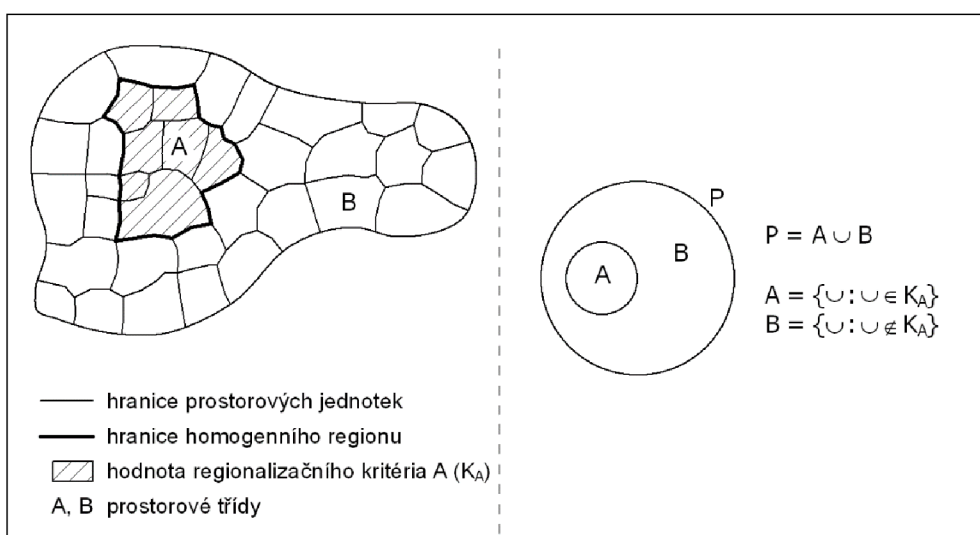


Obr. 1: Vztah geografického měřítka, použitých metod a celkové rozmanitosti regionů
Zdroj: (KLAPKA & TONEV, 2008)

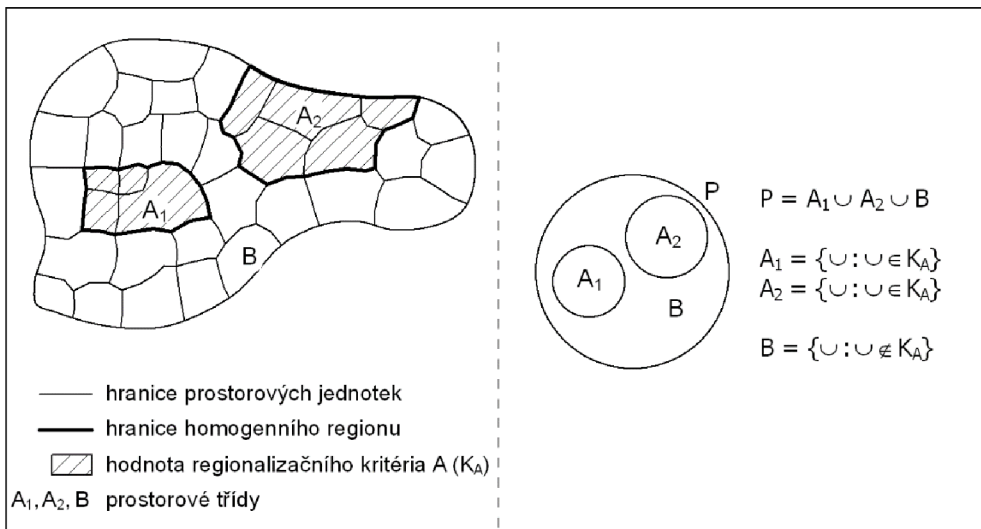
2.1.3 Formální region

Formální (homogenní, uniformní) region je na základě kritéria formy definován skalárními daty. Skalární data reprezentují formální vztahy v území, což je diference oproti regionu funkčnímu, který je definován interakcemi ve formě vektorových dat. Pro formální regiony je rozhodující jejich autonomie, vyjádřena jako interní homogenita a externí separace. Skalární data se tedy stávají kritérii regionalizace: taková kritéria musí být uvnitř regionu co nejméně rozdílná, mezi jednotlivými regiony by měla být naopak co nejvíce odlišná. Vzhledem k inherentně heterogennímu charakteru geografické sféry je takřka nemožné zajistit stejnou hodnotu regionalizačního kritéria v celém regionu (a vytvořit zcela uniformní region) – vznikají tak prostorové diference dané různými hodnotami vstupních skalárních dat. Interní homogenita klesá s množstvím použitých regionalizačních kritérií a také se vzrůstající hierarchickou úrovní. Počet regionalizačních kritérií, složek regionu, odráží také schopnost identifikace formálních regionů (KLAPKA, 2019, stránky 44–45). Při využití více kritérií lze v rámci struktury formálního regionu identifikovat místa s převažující platností všech regionalizačních kritérií (lze je zhruba označit za jádra) a logicky tedy i místa s omezenou platností regionalizačních kritérií (lze je zhruba označit za zázemí).

Formální region přebírá vlastnosti (dané použitými daty) prostorových jednotek, kterými je tvořen. Jednou z vlastností formálního regionu je mimo jiné jeho prostorová souvislost (obr. 5 a 6), která bude více zmíněna v následujících kapitolách (2.3.5 a 3.2.2).



Obr. 2: Formální region tvořený prostorově souvislými regionálními třídami
Zdroj: (KLAPKA & TONEV, 2008)

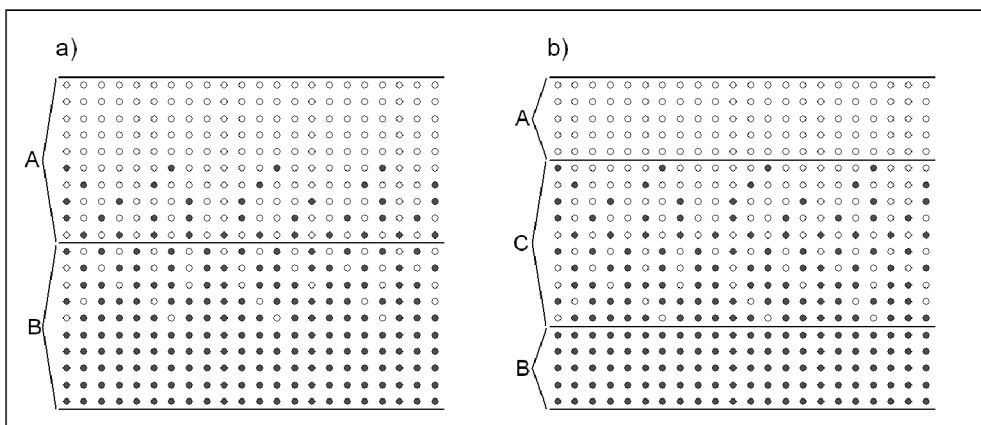


Obr. 3: Formální region tvořený prostorově nesouvislými regionálními třídami

Zdroj: (KLAPKA & TONEV, 2008)

Formální region se vyskytuje v socio-ekonomické i fyzické geografii, ačkoliv ve druhé zmíněné geografické disciplíně lze lépe sledovat vertikální závislost skalární jevů. Klapka a Tonev (2008) uvádějí příklad půdního a vegetačního krytu (mezilehlých vrstev), které jsou svrchu ovlivňovány klimatem a odspodu charakterem horninového podloží.

U homogenních regionů nás však primárně nezajímá jejich vnitřní struktura, ale především jejich hranice, jako linie oddělující jednotlivé regiony. Hraniční linie jsou buď ostré nebo rozvolněné. Ostré hranice převládají u socio-ekonomických regionů (např. administrativních) a u člověkem podstatně ovlivněných fyzicko-geografických regionů. Rozvolněné hranice mají charakter neostrých množin (viz obr. 4). Jeden homogenní region, pozvolna přechází ve druhý homogenní region. Hraniční linie je pak vedena na základě převažujícího výskytu obou kritérií (vlevo). V některých případech se však sama hraniční zóna považuje za samostatný region, který má specifické vlastnosti (vpravo).



Obr. 4: Problematika hranic formálních regionů

Zdroj: (KLAPKA & TONEV, 2008)

2.1.4 Hierarchie regionů

Formální regiony (i regiony obecně) zákonitě vykazují hierarchickou strukturu. Tato povaha úzce souvisí s diskontinuitou geografického prostoru, který se, ač je výrazně kontinuální, projevuje sekundárními nepravidelnostmi (HAMPL, 1998). Ke zjednodušení přílišně složité reality napomáhají právě hierarchické struktury, které zavádějí do kontinuálního prostoru organizační systém. Hierarchie jsou vztahy podřízenosti a nadřazenosti a souvisí s geografickým měřítkem, skrze které rozlišujeme i rozsah regionů. Přestože hierarchické vztahy pochází původně z přírodních systémů, nalezneme je i ve strukturách sociálních (KLAPKA, 2019).

Dle Klapky a Toneva (2008) mají hierarchie tyto vlastnosti:

- prostorový rozsah komponentů vyšší hierarchické úrovně je větší než rozsah komponentů nižších hierarchických úrovní
- komponenty vyšších hierarchických úrovní se vyvíjejí (mění) pomaleji než komponenty na nižších hierarchických úrovních
- vyšší hierarchické úrovně kontrolují nižší hierarchické úrovně
- struktury, které vykazují řád na jedné úrovni, mohou mít na úrovni jiné náhodný charakter
- nestabilita komponentů na vyšších hierarchických úrovních může být odrazem nestability komponentů na nižších hierarchických úrovních

Hierarchické struktury se vyskytují jako skladebné a neskladebné. V geografii mají zásadní význam právě skladebné hierarchie, jelikož je žádoucí, aby prvky nižších hierarchických úrovní beze zbytku vytvářely vyšší hierarchické úrovně. Z hlediska konkrétních hierarchických úrovní pozorujeme v geografii (jako disciplíně na pomezí přírodních a společenských věd) určitou dualitu, která znamená odlišnou terminologii právě pro fyzickou-geografické a socio-ekonomické systémy.

Hierarchický systém fyzicko-geografických regionů se skládá z následujících ustálených úrovní (od nejnižší):

- topická
- chorická (někdy se ještě dále člení na mikro-, mezo- a makrochorickou)
- regionální (někdy se ještě zvláště vymezuje úroveň nadregionální)
- planetární

Hierarchický systém socio-ekonomických regionů se skládá z následujících ustálených úrovní (od nejnižší):

- mikroregionální
- mezoregionální
- makroregionální
- národní
- globální

Na základě obou uvedených hierarchických členění lze identifikovat podobnosti, rozdíly a jiné poznatky. Za prvé, nejvyšší úroveň obou systémů je totožná (rozsah planety Země). Za druhé, nejnižší úroveň u obou systémů není uvedena – toto pomínutí souvisí s neexistencí prostorového rozměru nejmenších částí zmíněných systémů (pro fyzickou-geografické systémy by se jednalo např. o částice hmoty, u socio-ekonomických systémů je to člověk jako jedinec, nositel prostorové informace). Za třetí, je třeba určit úrovně hierarchické struktury, které by sloužily jako základní prostorové jednotky nejnižší úrovně – části geografické sféry, za které jsou dostupné agregované prostorové informace. Pro socio-ekonomické struktury se jasně jedná o nejnižší úrovně územního členění, situace ve fyzické-geografii ovšem není tak jednoznačná – výše uvedená topická úroveň by mohla být považována z hlediska rozlohy za ekvivalent obce, zároveň slouží ale již jako hierarchická úroveň nejnižšího řádu.

Terminologické nejednoznačnosti hierarchií nejsou jejich jediným problémem. Klapka (2019) uvádí pět různých, vzájemně provázaných problematik, spojených s různými měřítkovými úrovněmi:

- četnost výskytu regionů je nepřímo úměrná velikosti hierarchické úrovně
- neostrost hranic regionů roste se zvyšující se hierarchickou úrovní
- autonomie fyzicko-geografických a formálních regionů se zvyšující se hierarchickou úrovní klesá, u socio-ekonomických a funkčních regionů je tomu naopak
- vnitřní heterogenita fyzicko-geografických a formálních regionů se zvyšující se hierarchickou úrovní stoupá, u funkčních regionů je změna závislá na konkrétní úrovni v rámci hierarchie
- úroveň environmentálně determinace roste se zvyšující se hierarchickou úrovní

2.2 Obecná charakteristika regionálně-taxonomických procedur

V souladu s principy pozitivistické a postpozitivistické vědy je i v geografii potřebná klasifikace informací. Obecně, taxonomie usiluje o uspořádání objektů do systému, který je vzájemně provázán pomocí vztahů. Obvykle jsou informace nad rámec běžné struktury klasifikovány chronologicky, s časovým záznamem. Běžná klasifikace je navíc v rámci geografické vědy doplňována o prostorový aspekt. Na základě těchto poznatků rozlišujeme tři základní typy vědeckého uspořádání informací (např. MIČIAN, 2003):

- standartní klasifikace
- uspořádání v čase pomocí periodizace
- uspořádání v prostoru pomocí regionální taxonomie

Klasifikování objektů do příslušných taxonů lze v prostoru vztáhnout k identifikaci regionálních tříd – delimitaci regionů. Množinou těchto regionálních tříd je pak regionální systém, který je tedy výsledkem regionálně-taxonomické procedury. Regionálně-taxonomická procedura v obecném smyslu probíhá rozkladem určité množiny do vzájemně se vylučujících kategorií na základě předem určených kritérií (KLAPKA & TONEV, 2008).

Vyjma prosté potřeby klasifikovat prostor je přínos regionální taxonomie i v jeho agregaci. Výhodnost agregovaných jednotek spočívá v potlačení nepřesností v datech a lokálních extrémů, snižování efektu malých čísel, usnadnění grafického zobrazení prostoru nebo v anonymizaci individuálních informací. Vyjma těchto benefitů může sloužit jako nástroj regionálního plánování a analýz větších celků (KLAPKA, 2019).

Regionální taxonomii chápeme jako samotnou proceduru vymezení regionů i jako její výsledek (např. v podobě grafického díla). U samotné procedury rozlišujeme dva hlavní typy výsledků: regionalizaci, jejímž výsledkem jsou individuální regiony, a regionální typologii, která produkuje regiony typologické. Mezi nejběžnější postupy patří spojování základních prostorových jednotek do větších celků anebo dělení jejich sumy.

2.4.1 Zásady regionální taxonomie

Aplikace regionální taxonomie vychází z formulace výzkumných cílů v rámci zpracovávané problematiky. Podle těchto cílů jsou rovněž zvolena a definována regionalizační kritéria. Z toho vyplývá, že se regionální taxonomie nemůže vyhnout určitému stupni subjektivity, který by však měl být minimalizován (např. kapitola 2.3.3). Vyjma volby regionalizačních

kritérií je nutné provést volbu klasifikovaných objektů. Zde je vhodné zvolit kombinaci na základě definovaných cílů výzkumu a dostupnosti geografických dat.

Na základě syntézy poznatků různých autorů [např. (BAŠOVSKÝ & LAUKO, 1990), (BEZÁK, 1993), (MIČIAN, 2003), (KLAPKA, 2019)] logická pravidla regionální taxonomie:

- vymezené regionální třídy musí beze zbytku pokrývat celé území, které je součástí regionálně-taxonomické procedury
- vymezené regionální třídy musí beze zbytku obsahovat všechny základní prostorové jednotky, které jsou součástí regionálně-taxonomické procedury
- základní prostorové jednotky musí být vždy součástí jediné regionální třídy v rámci jedné hierarchické úrovně
- regionalizační kritéria by měla vycházet z vlastností klasifikovaných objektů a měla by být relevantní pro řešenou regionálně-taxonomickou úlohu
- je vhodné zachovat logické stupně regionalizační procedury
- v rámci jednoho stupně regionalizační procedury se musí zachovat jen jedno regionalizační kritérium
- v rámci jednoho stupně regionalizační procedury není možné kombinovat typologickou regionalizaci a individuální regionalizaci

Výsledkem regionální taxonomie jsou tedy souvislé a pokud možno kompaktní prostorové jednotky, které se vzájemně nepřekrývají a pokrývají celé zájmové území. Procedura regionální taxonomie a její pravidla by měly být přesně a jednoznačně definovány.

Klapka a Tonev (2008) následně dodávají, že vymezené regiony by se měly vyznačovat následujícími charakteristikami:

- regionalizační kritérium by mělo uvnitř regionu být maximálně homogenní
- mezi jednotlivými regiony by regionalizační kritérium mělo být maximálně odlišné
- vymezené regiony by měly mít určitou minimální velikost, která závisí na cílech výzkumu
- vymezené regiony by měly být skladebné do vyšších hierarchických úrovní

V případě regionalizace spolu všechny prostorové jednotky, které jsou součástí jedné regionální třídy sousedí, v případě regionální typologie jsou výsledkem prostorově nesouvislé regionální třídy - regionální typy.

2.3 Metody formální regionální taxonomie

Kapitola obsahuje výtah dostupných metod regionální taxonomie a nastiňuje jejich výběr pro praktickou část. Dále je diskutována problematika obecných operací s prostorovými informacemi, vzdálenosti v taxonomickém prostoru, způsoby redukce matice proměnných, prostorová souvislost formálních regionálních tříd a řešení optimálního počtu formálních regionálních tříd.

2.3.1 Hlavní metodické přístupy

Třídění metodických přístupů se odvíjí od pojetí různých autorů. Zde je uvedeno několik vybraných klasifikací.

Mezi starší publikace patří například „*Classification in geography*“ podle R. J. Johnstona (1976). Dílo se vyjma samotné klasifikace metod zabývá i prostorovým omezením shlukové analýzy v příkladech. Rozebrány jsou aglomerativní i divizivní metody a jednoduché způsoby formování regionálních tříd.

Jiný ze způsobů uvádí např. Klapka a Tonev (2008), a to s odkazem na syntézu publikací zahraničních autorů jako Haggett (1965), Spence a Taylor (1970), Bašovský a Lauko (1990), Richling (1992) nebo Bezák (1993). S dodatkem rezignace na úplnost a vývojovou složitost, uvádějí následující přístupy:

- metoda superpozice analytických map
- metoda vedoucího faktoru vycházející ze superpozice analytických map
- subjektivní metoda generalizace
- metoda analýzy hranic
- skupinu metod numerické taxonomie označované jako metody kvantitativní analýzy
- skupinu kvalitativních metod založených například na dotazníkových šetření

Velmi obsáhlý přehled nabízí syntetická publikace z roku 2007 (GAN, MA, & WU). Autoři na základě vlastností algoritmů identifikují 10 hlavních skupin shlukovacích procedur:

- hierarchické shlukovací metody
- „fuzzy“ shlukovací metody založené na teorii neostrých množin
- metody založené na funkci jádra (např. k-průměry)
- metody založené na genetických algoritmech
- grafové metody

- metody založené na struktuře sítě
- metody založené na hustotě jevů
- metody založené na optimalizaci genetických modelů
- metody založené na identifikaci vztahů v jednotlivých částech datového souboru
- nezatříděné metody (např. metody založené na neuronových sítích a strojovém učení)

V této diplomové práci, respektive její výpočtové části, se zaměříme na klasifikaci regionální taxonomie a regionálních systémů z pohledu Klapky (2019). Přístup je opřen o dvě základní kritéria: způsob analýzy geografické matice a způsob tvorby regionálních tříd.

Kombinací obou kritérií vznikají následující skupiny metod:

- hierarchické metody vycházející z teorie grafů
- nehierarchické metody vycházející z teorie grafů
- hierarchické metody vycházející z numerické taxonomie
- nehierarchické metody vycházející z teorie grafů

Nehierarchické metody jsou dále děleny na aglomerativní a divizivní. Členění je odvislé od postupu tvorby regionálních tříd, tedy buď tzv. zdola nahoru nebo shora dolů. Konkrétní použité metody jsou popsány v kapitole 3.2.

2.3.2 Operace s prostorovými informacemi

Pro efektivní provedení regionálně-taxonomických procedur je nutné správné uspořádání prostorových dat, obvykle ve formě matic. Primárně členíme data na skalární a vektorová (viz klasifikace regionů podle struktury) a dále dle charakteru geografické informace na nominální, ordinální, intervalová a poměrová. Klasická úprava dostupných geografických informací se provádí ve třech krocích, které pomáhají s odstraněním klasických projevů deformace dat vlivem diskontinuity prostoru. Typickými kroky v postupu jsou:

- 1) sběr informací
- 2) konstrukce neupravené matice
- 3) transformace matice

Sběr informací (krok 1) by měl proběhnout na úrovni předem stanovené základní prostorové jednotky (např. obec nebo mikroregion). Způsob konstrukce neupravené matice (krok 2) již přímo vychází z rozdílu mezi skalárními a vektorovými daty:

- skalární informace jsou primárně uloženy v atributové matici, která obsahuje základní prostorové jednotky (na axiále Y nebo v řádcích) a atributy, parametry, indikátory, regionalizační kritéria nebo jiné údaje (na axiále X nebo ve sloupcích); vnitřní část matice přiřazuje každé prostorové jednotce všechny dostupné atributy; počet řádků a sloupců zpravidla není roven, jelikož počet základních prostorových jednotek v řádcích matice pouze velmi výjimečně nepřevyšuje (nebo není roven) počet atributů ve sloupcích
- vektorové informace jsou obvykle primárně uloženy v interakční matici, která obsahuje základní prostorové jednotky (zpravidla na obou axiálách, tedy v řádcích i sloupcích); vnitřní část matice definuje vztah mezi všemi základními prostorovými jednotkami, obvykle vyjma sebe navzájem; počet řádků a sloupců je zpravidla roven, jelikož je nutné postihnout interakci mezi všemi základními prostorovými jednotkami; ve specifických případech zkoumáme interakce dvou rozdílných souborů základních prostorových jednotek – v tomto případě neplatí výše zmíněná pravidla a počet řádků a sloupců se může lišit

Transformace matice (krok 3) docílíme matematickými nebo statistickými operacemi na souboru dat v neupravené matici. Transformací matice může v určitých případech dojít ke ztrátě informace (např. v řádu desetinných míst).

První skupina transformačních operací se obvykle týká atributových matic (tedy skalárních dat a formálních regionů). Nejdůležitější a nejjednodušší možností transformace matice je její relativizace. Relativizaci provádíme přepočtem dat nominálního charakteru na data poměrová (příkladem je nahrazení počtu nezaměstnaných obyvatel v dané prostorové jednotce mírou nezaměstnanosti) a docílíme tím odstranění efektů rozdílné velikosti základních prostorových jednotek. Jiným způsobem transformace atributové matice je normalizace. Touto skupinou operací odstraňujeme extrémní hodnoty, tedy šikmost a špičatost daného souboru základních prostorových jednotek. Efektivním a nearbitrárním způsobem normalizace je standardizace, jejíž pomocí lze přímo porovnat různé atributy ve statistickém souboru. Příkladem může být standardizace pomocí přiřazování vah jednotlivým atributům anebo pomocí směrodatné odchylky a průměru, tzv. z-skóre. Podílem rozdílu mezi původní hodnotou a průměrem souboru směrodatnou odchylkou souboru získáme bezrozměrné hodnoty (z-skóre), jejímž součtem je 0.

Práci s atributovou maticí můžeme dále zjednodušit pomocí redukce a dekorelace. Redukce matice je nutná v případě příliš rozsáhlého souboru základních prostorových jednotek,

ovšem její význam se v době dostatečného výpočetního výkonu počítačové techniky výrazně snížil. Dekorelace, na druhou stranu, pomáhá se snížením počtu atributů a může odhalit předpokládanou nežádanou závislost mezi regionalizačními kritérii. Odstranění korelovaných proměnných může pozitivně ovlivnit výslednou validitu regionálně-taxonomické procedury a zlepšit kvalitu výsledků. Mezi využitelné metody (souhrnně označované jako vektorové ortogonální transformace) řadíme analýzu hlavních komponent a faktorovou analýzu (kapitola 2.3.3).

Druhá skupina transformační operací se naopak týká matic interakčních (tedy dat vektorových a funkčních regionů). Mezi takové lze zařadit nutnou symetrizaci matice, dále pak dichotomizaci, relativizaci, vážení a analýzu nepřímých toků. Vyjma analýzy blízkosti (kapitola 2.3.4) není nutné, s ohledem na zaměření práce, se problematice interakčních matic věnovat.

2.3.3 Metody vícerozměrné analýzy

Mezi metody vícerozměrné analýzy patří především shluková analýza a faktorová analýza, respektive analýza hlavních komponent. Faktorová analýza a analýza hlavních komponent slouží k redukci regionalizačních kritérií a identifikaci zásadních či skrytých regionalizačních kritérií (tzv. hlavních komponent či faktorů) v případě, že se jich na počátku regionalizačního procesu vyskytuje velké množství. Faktorová analýza využívá asociační a korelační koeficienty. Shluková analýza je pak metodou slučování základních prostorových jednotek v jednotky (regiony) větší, na základě zvolených regionalizačních kritérií (KLAPKA & TONEV, 2008). Specifickým příkladem shlukové analýzy je právě regionální taxonomie.

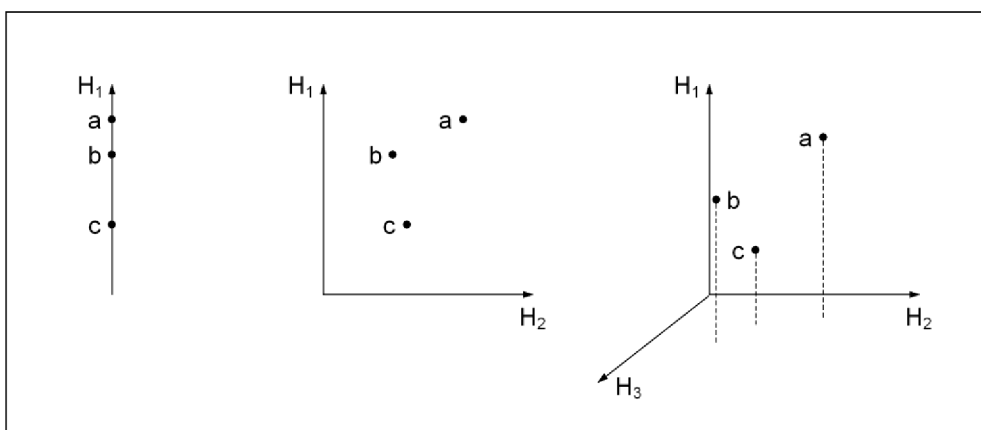
Vlastní postup při aplikaci metod kvantitativní analýzy (konkrétně analýzy hlavních komponent nebo faktorové analýzy) se teoreticky může odehrávat v následujících logických krocích za předpokladu, že máme definována základní výchozí regionalizační kritéria a jejich hodnoty pro každou základní prostorovou jednotku (HEŘMANOVÁ, 1991):

1. zpracování atributové matice s hodnotami predikátů pro všechny prostorové jednotky
2. zpracování korelační/kovariační matice
3. extrakce vlastní vektorů (tzv. eigenvektorů)
4. výpočet zátěží jednotlivých eigenvektorů
5. výpočet vektorových (komponentních nebo faktorových) skóre

Redukce původní matice je následně dosaženo výběrem významných proměnných. Lze aplikovat Kaiserovo kritérium (zahrnuje všechny eigenvektory větší než 1) nebo metodu Scree-slope (okamžik zlomu grafu při vynesení jednotlivých vektorů ve dvourozměrném prostoru). Hodnota vlastních čísel obsahuje informace o rozsahu vysvětlené variability souboru (KLAPKA, 2019).

2.3.4 Blížkost v taxonomickém prostoru

Na měření podobnosti prostorových jednotek je založena i metoda analýzy vzdálenosti v n -rozměrném prostoru (nebo také metoda funkční vzdálenosti či metoda taxonomické vzdálenosti). Její základní mechanismus využívají i metody složitější. Metoda spočívá ve výpočtu taxonomické vzdálenosti mezi všemi uvažovanými prostorovými jednotkami. K výpočtu taxonomické vzdálenosti slouží hodnoty regionalizačních kritérií (znaků). Obrázek 5 představuje schematizaci příkladu s jedním až třemi kritérii. Vzdálenosti mezi prostorovými jednotkami označenými jako a, b, c jsou v případě jednoho kritéria měřeny v jednorozměrném prostoru (jediná osa H_1). V případě dvou kritérii jsou vzdálenosti měřeny ve dvourozměrném prostoru (dvě osy H_1 a H_2) pomocí Pythagorovy věty, a konečně v případě tří kritérii jsou vzdálenosti měřeny v trojrozměrném prostoru (tři osy H_1 , H_2 a H_3) (KLAPKA & TONEV, 2008).



Obr. 5: Schéma měření taxonomické vzdálenosti v n -rozměrném prostoru
Zdroj: (KLAPKA & TONEV, 2008)

Při využití více než dvou regionalizačních kritérií je obtížné úlohu znázornit ve dvourozměrném prostoru, ovšem řešení je možné znázornit za pomoci výpočtu (obr. 6).

$$D = \sqrt{\sum_{i=1}^n (x_i - y_i)^b}$$

Obr. 6: Obecný vzorec výpočtu taxonomické vzdálenosti

Pozn.: D = taxonomická vzdálenost mezi prostorovými jednotkami x a y, i = regionalizační kritérium, b = index nabývající nejčastěji hodnoty 2

Zdroj: (KLAPKA & TONEV, 2008)

Některé metody formální regionální taxonomie vyžadují vstup v podobě matice taxonomických vzdáleností. Jedná se o symetrickou matici (kapitola 2.3.2), zobrazující vzdálenost jednotek (každý s každým) v taxonomickém prostoru s využitím některé z metrik. Mezi běžně užívané typy vzdáleností pro numerickou taxonomii se řadí (GAN, MA, & WU, 2007):

- Euklidovská vzdálenost (i její čtverec)
- Manhattanská vzdálenost (také city-block nebo taxi-cab)
- Čebyševova vzdálenost
- Mahalanobisova vzdálenost
- Maximální vzdálenost
- Průměrová vzdálenost

2.3.5 Prostorová souvislost formálních regionálních tříd

Většina metodologických přístupů ke konstrukci formálních regionálních tříd tvoří typologické, prostorově nesouvislé regionální třídy. Tato vlastnost představuje značný problém při požadavku na kompaktnost, souvislost vznikajivších regionů nebo prostou delimitaci individuálních regionů. Obejít tento problém je možné dvěma způsoby (při čemž jeden z nich je použit v praktické části této práce):

- Dvoustupňový přístup, jehož jádrem je konstrukce standartních typologických tříd, jejichž prostorově souvislé části se dají považovat za individuální regiony. Tento přístup je velmi nevýhodný, jelikož nad rámec zvyšuje počet vymezených regionů, které, ač považovány za unikátní, jsou vymezeny na základě stejných pravidel jako ostatní nesouvislé části rozdělené nesouvislé regionální třídy.
- Jednostupňový přístup, jehož základem je zavedení podmínky prostorové souvislosti již do regionalizačního algoritmu. Aby bylo možné tento postup aplikovat, je nutné znát informaci o sousedství základních prostorových jednotek v geografickém

prostoru. Klapka (2019) identifikuje na základě množství publikací zabývajících se daným tématem, čtyři možnosti vyjádření sousedství. Mezi ně patří fyzické sousedství hranic prostorových jednotek (hranou nebo rohem), vzdálenost jednotek v geografickém prostoru (vyjádřená např. obalovou zónou), určená funkce vzdálenosti, anebo identifikace minimální kostry grafu.

2.3.6 Optimální počet formálních regionálních tříd

Hlavním problémem regionálně-taxonomických procedur, respektive shlukových analýz obecně, je identifikace optimálního počtu finálních regionálních tříd. Prezentovaný problém je zejména přítomen u nehierarchických metod (k -průměrů a p -mediánů), kde je kvalita počtu formálních regionálních tříd označována jako k a je nutnou součástí regionalizačního algoritmu. Nicméně odhalení optimálního počtu regionálních tříd, tedy určení optimálního řezu dendrogramem, je vhodné i u hierarchických metod. Existuje mnoho způsobů identifikace optimálního počtu shluků; některé jsou založeny na souladu počtu vzniklých shluků a matice vzdálenosti (tedy požadují ad-hoc iterační řešení), další využívají principu interní homogenity a externí separace a jiné poskytují řešení skrze grafické řešení.

2.4 Stručná historie regionálního členění v Česku

Dějiny rajonizace českých zemí z pohledu moderní geografie čítají velké množství jmen a výsledků prací různých autorů, proto se tato kapitola zaměřuje na socioekonomické regionalizace s použitím metod regionální taxonomie nebo jejích obdob. Klapka (2019) uvádí poměrně obsáhlý přehled příspěvků v téměř 150leté historii rajonizací založených na vědeckých metodách, rozčleněných do tří hlavních období. Tato období jsou ohraničena zásadními momenty v historii českého státu od konce 19. století a jejich primárním účelem je poskytnout referenci k motivaci a účelu zpracovaných regionalizací. Zároveň zmiňuje několik pracovišť se specifickými přístupy a výsledky. Společným znakem většiny zpracovaných regionalizací je jejich určitá složkovost (vymezení na základě jednoho nebo dvou regionalizačních kritérií) a také účel zpracování, který je až na výjimky velmi obecný.

2.4.1 Rané období (pol. 19. století–2. sv. válka)

V prvním zmiňovaném období (2. pol. 19. století–2. světová válka) se objevují zejména regionalizace založené přírodních podmínkách a práce zaměřené na konstrukci regionů na vyšší územní úrovni. Historické události tohoto časového rozmezí (první moderní census, vznik samostatného Československa) také zavdaly potřebě revizí administrativního členění, což se odrazilo v požadavcích a motivacích pro regionalizační procesy. Základní prostorovou jednotku v této době tvořily soudní okresy. Mezi nejvýznamnější osobnosti patřili Karel Kořistka, Václav Dědina a Jaromír Korčák. Karel Kořistka (1872) sestavil v 70. letech 19. století tzv. přirozené krajiny, což byl soubor celkem 24 ekonomických regionů definovaných pro účely lesního hospodářství. Václav Dědina (1929) ve 20. letech 20. století na jeho práce navázal, a definoval tak 19 regionů, jejichž hlavním účelem byla tentokrát analýza zemědělské činnosti. Vzhledem k charakteru vzniklých regionů lze poznamenat, že oba autoři ve svých pracích vycházeli z pozic environmentálního determinismu. Jaromír Korčák (1934), oproti tomu, při revizi Dědinových jednotek o několik let později, využil i podstatu environmentálního posibilismu. Jeho tzv. koncentrační oblasti jsou výsledkem syntézy původní myšlenky Dědinových krajín a genetického přístupu k vymezení regionů. Korčákovým odkazem je mimo jiné představa dichotomie mezi přírodními a historickými regiony, což hrubě odpovídá dnešní typologii regionů z hlediska struktury (formální a funkční regiony).

2.4.2 Poválečné období (1945–1989)

Druhé zmiňované období (1945–1989) je charakteristické existencí dvou hlavních pracovišť, zabývajících se regionálně-taxonomickými procedurami: Geografickým ústavem ČSAV a Katedrou geografie Přírodovědecké fakulty Univerzity Karlovy. Blízká poválečná doba je typická adaptací na nové ekonomické prostředí a také na změny ve vědeckém prostředí – nastupující kvantitativní revoluci. V oblasti administrativního členění (1949 a 1960) se geografové příliš neangažovali a v rámci ekonomicko-geografické regionalizace nacházíme pouze množství teoretických prací. Zatímco tedy oblast socio-ekonomické geografie požívala vyšší stupeň ideologického ovlivnění a koncentrovala se primárně na identifikaci jader průmyslových rajónů a pracovní síly, fyzicko-geografický výzkum se soustředil na složkové i komplexní regionálně-taxonomické systémy.

V 60. a 70. letech 20. století tedy vzniklo na půdě ČSAV množství dodnes uznávaných prací, mezi nimiž je nutno zmínit komplexní fyzicko-geografické regiony ČSR Jaromíra Demka (1977). Demkovy regiony byly sestaveny na základě tří již dříve zpracovaných složkových taxonomií: klimatických regionů Evžena Quitta, členění reliéfu Tadeáše Czudka a biogeografické regionalizace Jaroslava Raušera. V rámci fyzicko-geografických příspěvků atlasových děl ČSR a ČSFR z let 1987 a 1992 se ještě objevily typy přírodních krajín – regionalizace zpracované na základě různých složek v životním prostředí. Výsledek dlouholeté práce ekonomických geografů působících na akademii věd byl prvně zúročen až v roce 1977 v Ekonomické regionalizaci (BLAŽEK). Identifikace jader a stanovení sfér jejich vlivu vedlo ke čtyřstupňové hierarchické regionální taxonomii, jejíž výsledky jsou ovšem z mnoha důvodů považovány za metodicky nečisté. Změnou oproti regionalizacím z předcházejícího období je bezesporu použití obce jako základní prostorové jednotky a striktní využití socio-ekonomických charakteristik jako regionalizačních kritérií. Mezi hlavní problémy této práce však patří hierarchická neskladebnost jednotlivých úrovní, směšování principů formální a funkční regionální taxonomie a také nemožnost reprodukce práce na základě publikovaných výsledků. Metodicky čistší a vývojově složitější práce byla publikována až v Atlasu obyvatelstva ČSR. Regionální působnost středisek osídlení Stanislava Řeháka a Jaroslava Maryáše (1987) byla zpracována za pomoci dobových moderních technologií, což umožnilo dosažení efektivnějších výsledků. Klíčovým faktorem při řešení této úlohy bylo stanovení potenciálních jader regionů a spádovosti přilehlých obcí (základní prostorová jednotka), vše za využití dat o dojížděnce obyvatelstva do zaměstnání anebo službami. Celkově bylo na území ČSSR (mimo Slovenské části) identifikováno 290

mikroregionů, které byly později mimo jiné využity jako podklad pro výběr obcí s rozšířenou působností.

Myšlenky Pražské Albertovské školy, reprezentované zejména nejvýznamnějším představitelem Martinem Hamplm a ideově navazující na práce Jaromíra Korčáka, Waltera Christallera a Alberta Löscheho (teorie centrálních míst), byly od 60. let 20. století úspěšně zformovány do tří (po roce 1993 do dalších dvou) na sebe navazujících regionalizací. Hlavními poznávacími znaky Hamplových regionalizací jsou jistě jeho unikátní terminologie, zasazení regionálně-taxonomických procedur do širšího teoretického rámce (teorie hierarchie reality) anebo kapitalizace obohacených známých poznatků do komplexnějších struktur, které dodnes nacházejí místo v geografickém bádání a výuce. Společným jmenovatelem konstrukce všech tří hlavních regionálně-taxonomických výsledků Albertovské školy v tomto období je principialita hlavních kroků při stanovení regionálních tříd (identifikace jader a dále zázemí těchto jader), využití obce jako hlavní základní prostorové jednotky, hierarchizace stanovených regionálních tříd a primární využití cenových dat (vyjíždka do zaměstnání a škol). První kompletní studie tohoto druhu byla publikována v roce 1978 (HAMPL, JEŽEK, & KÜHNL), kdy bylo definováno 187 mikroregionů 1. řádu a regionalizační procedura byla nastavena na identifikaci regionálních tříd na 5 hierarchických úrovních (mikroregiony 1. a 2. řádu, mezoregiony 1. a 2. řádu a makroregion). Důvodem diferenciací řádovostních úrovní byla percepce některých center mikro – a mezo-regionů jako jader nadřazených. Druhá kompletní studie (HAMPL, GARDAVSKÝ, & KÜHNL, 1987) byla aktualizací první práce. Specifikem této regionalizace bylo využití (mimo obcí) generelových jednotek jako základních prostorových jednotek a také první využití koeficientu komplexně-funkční velikosti (průměr součtu počtu obyvatel, počtu pracovních míst a počtu pracovních míst v terciéru) pro identifikaci jader regionálních tříd. Celkem bylo definováno 151 mikroregionů 1. řádu a regionalizační procedura byla, stejně jako v předchozím případě, nastavena na identifikaci regionálních tříd na 5 hierarchických úrovních. Další aktualizace (HAMPL & MÜLLER, 1996) se metodicky výrazně neodlišuje od práce předchozí – tentokrát bylo definováno 147 mikroregionů 1. řádu a regionalizační procedura byla nastavena na identifikaci regionálních tříd na 4 hierarchických úrovních (mikroregiony, mezoregiony 1. a 2. řádu a makroregion). Celkově lze při hodnocení výsledků konstatovat klesající trend počtu mikroregionů 1. řádu, což je mimo jiné důsledek organizačních změn ve společnosti.

2.4.3 Nejnovejší období (1993–současnost)

Výběr regionálně-taxonomických prací v nejnovejším období reflektuje celospolečenské změny po roce 1989, organizační změny v působnosti jednotlivých geografických pracovišť a také rozsáhlejší využití geografických informačních systémů pro regionálně-taxonomické procedury. Definitivní přesun jádra prací na Přírodovědeckou fakultu Univerzity Karlovy byl vynucen zánikem geografického ústavu ČSAV na počátku tohoto období, později podle tradice jeho brněnské sekce vznikla škola Olomoucká. Mimo institucionalizované taxonomie nalézáme i množství "pracovištně" neukotvených prací, jejichž vznik je rámován např. kohezní a zemědělskou politikou EU, vývojem administrativního členění nebo statistickými potřebami ČSÚ.

Fyzicko-geografické práce v nejnovejším období vznikala za výrazného přispění moderních technologií, zejména GIS a DPZ. Vyjma konceptuálního navázání na regionalizace publikované na půdě ČSAV bylo pro konstrukce komplexních fyzicko-geografických systémů využito generalizace složkových regionalizací v životním prostředí. Vzhledem k charakteru prostorových informací v životním prostředí posloužily jako základní prostorové jednotky čtverce o různé velikosti. Výsledky (KOLEJKA, ROMPORTL, & LIPSKÝ, 2009) byly publikovány jako typy krajiny například v Atlasu krajiny České republiky.

Albertovská škola po roce 1993 představuje určitou kontinuitu z předcházejícího období. Na základě tří předešlých regionalizací tedy vyšly ještě další dvě práce, a to s mírně odlišnými parametry pro konstrukci regionálních tříd. Z hlediska počtu regionů identifikovaných na mikroregionální úrovni dochází opět ke klesajícímu trendu (144 v roce 2001 a 131 v roce 2011). Výsledky se objevily v pracích z let 2005 (HAMPL) a 2015 (HAMPL & MARADA).

Olomoucká škola vznikla v nultých letech 21. století a působí primárně na půdě Přírodovědecké fakulty Univerzity Palackého v Olomouci. Hlavním výzkumným zaměřením jejích členů v oblasti regionální taxonomie je tvorba funkčních regionálně-taxonomických systémů pomocí klasických regionalizačních algoritmů. Mezi základní znaky vyprodukovaných prací patří zejména jejich přesně definované postupy s možností reprodukce v jiném časovém okamžiku a na jiném území, analýzy denních pracovních toků mezi obcemi jako základními prostorovými jednotkami anebo aplikace metod numerických (algoritmus Intramax nebo metoda funkční vzdálenosti), metod založených na pravidlech (různé varianty algoritmu Curds a Smartova míra) i metod grafových. Členové školy za její

krátké působení vyprodukovali množství výsledků, proto je zde jejich výčet omezen na hlavní skupiny. Ranné studie jsou zaměřené na identifikaci sfér vlivu pomocí upraveného Reillyho gravitačního modelu a v mnohých případech ani nelze říct, zda se skutečně jedná o funkční regiony (HALÁS & KLAPKA, 2010). Funkční mikroregiony ve formě lokálních trhů práce (LLM) nebo pracovních dojížděkových regionů (TTWA) byly identifikovány pomocí vybraných metod na cenzových datech z roku 2001 (KLAPKA & kol., 2014) a z roku 2011 (KLAPKA & kol., 2016). Varianta vymezených funkčních mikroregionů také posloužila k definici jednotek na meziregionální úrovni, a to pomocí různých metod (HALÁS, KLAPKA, & TONEV, 2014) (ERLEBACH, TOMÁŠ, & TONEV, 2016). Dále vznikly i různé komparační práce (TONEV, 2013) a podkladové práce pro administrativní členění.

Jako příklad prací vzniklých mimo akademickou sféru je třeba uvést regionalizace spojené s evropskou zemědělskou a kohezní politikou. V prvním případě se jedná o tzv. less-favoured areas, sloužící pro identifikaci oblastí se znevýhodněnými podmínkami pro zemědělskou činnost. Obdobný přístup je volen v kohezní politice, kde byly v průběhu několika strukturálních období definovány různé typy sociálně a hospodářsky znevýhodněných oblastí. Poslední iterace vymezení hospodářsky a sociálně ohrožených území se objevila ve Strategii regionálního rozvoje 2021+ (Ministerstvo pro místní rozvoj ČR, 2020), kdy jako základní prostorová jednotka sloužila obec s rozšířenou působností a bylo zvoleno pět regionalizačních kritérií.

3 Metodologie a data

Metodologie praktické části této práce byla zvolena tak, aby odpovídala hlavním přístupům, které uvádí Klapka (2019) a zároveň byla dosažitelná z pohledu softwarových nástrojů a dat. Pro účely této práce je vhodná klasická metodologie regionálně-taxonomických procedur, prezentovaná v šesti základních krocích:

1. definice cílů regionální taxonomie (kapitola 3)
2. volba klasifikovaných objektů (kapitola 3.1)
3. volba regionalizačních kritérií a operace s nimi (kapitola 3.1)
4. hodnocení blízkosti klasifikovaných objektů (kapitola 3.1)
5. volba vlastní regionalizační procedury (kapitola 3.2)
6. hodnocení validity regionální taxonomie (kapitola 3.3)

Definice cílů formální regionální taxonomie vychází z definice cílů této diplomové práce. V obecné rovině je tedy motivací pro provedení regionálně-taxonomických procedur ověření chování vybraných prostorových struktur pomocí vhodně zvolených regionálně-taxonomických procedur. Cílem práce je zároveň prověření vhodných metod formální regionální taxonomie, motivací pro zpracování lze tedy považovat v krajním případě za samoučelnou.

3.1 Zpracování dat

Volba klasifikovaných objektů vychází ze zadání diplomové práce. Výběr obce jako základní prostorové jednotky lze avšak jednoznačně oddiskutovat vhodností této hierarchické územní úrovně z hlediska dostupnosti prostorových dat. Tedy, zvolená územní úroveň poskytuje maximální detail dostupnosti prostorových dat (desetiletý cenzus) při zachování maximálního rozsahu vhodných regionálně-taxonomických procedur.

Volba datových zdrojů úzce souvisí s výběrem klasifikovaných objektů. Desetiletá cenzová data v případě zvolené základní prostorové jednotky nabízí širokou škálu potenciálních regionalizačních kritérií. Jejich výběr ovšem není zcela závislý na fantazii autora. Pro minimalizaci subjektivního vlivu autora je v tomto případě třeba využít některé z metod vícerozměrné analýzy, zběžně popsanych v kapitole 2.3.3. Analýza hlavních komponent (PCA) umožňuje celkovou redukci rozsahu potenciálních regionalizačních kritérií. Pro extrakci komponent je využito metody Scree-slope a Kaiserova kritéria, přičemž pro další práci s extrahovanými komponenty je zvolena metoda produkující vhodnější výsledky v rámci testovací sady.

Samotné operace s prostorovými informacemi (příprava dat pro PCA) zahrnují relativizaci, normalizaci a standardizaci datového souboru. Transformované interakční matice proměnných jsou účelné pro metody vícerozměrné shlukové analýzy.

Další práce s prostorovými informacemi jsou součástí samotné regionalizační procedury. Specificky, pro hierarchické-aglomerativní metody je nutné zpracovat symetrickou matici blízkosti (nepodobnosti) z matice predikátů a základních prostorových jednotek. Matice blízkosti představuje matici taxonomických vzdáleností, odlišných od vzdáleností ve skutečném prostoru. V případě této diplomové práce jsou zvoleny celkem tři typy odvozených Minkowského metrik, lišící se způsobem výpočtu vzdálenosti mezi predikáty ve vícerozměrném prostoru:

- Euklidovská vzdálenost, která je ve dvourozměrném prostoru vyjádřena přeponou pravoúhlého trojúhelníka načrtnutého mezi párem regionalizačních kritérií,
- Manhattanská vzdálenost, která je shodná s délkou obou odvěsen pravoúhlého trojúhelníka načrtnutého mezi párem regionalizačních kritérií,
- Čebyševova vzdálenost, jejíž hodnota je odečtena z délky delší odvěsny pravoúhlého trojúhelníka načrtnutého mezi párem regionalizačních kritérií.

3.2 Vybrané metody formální regionální taxonomie

V souladu s plněním cílů této diplomové práce je v rámci volby vlastní regionalizační procedury vybráno více možností tak, aby byly relativně postiženy základní přístupy ke konstrukci formálních regionálně-taxonomických systémů.

3.2.1 Hierarchické metody formální regionální taxonomie

Hierarchické metody tvorby formální regionální taxonomie postupují na základě formování průběžných regionálních tříd. Rozlišujeme je na aglomerativní a divizivní. Aglomerativní metody se snaží postupně sloučit dvojice základních prostorové jednotky do jediné finální regionální třídy, naproti tomu divizivní metody usilují o postupné rozdělení jedné regionální třídy do počtu odpovídajícího množství základních prostorových jednotek. Pro účely této diplomové práce je použito vícero hierarchických metod formální regionální taxonomie:

- aglomerativní metody založené na poloze základních prostorových jednotek,
- aglomerativní metody založené na poloze těžišť regionálních tříd,
- aglomerativní metody založené na vnitrotřídní variabilitě,
- divizivní algoritmus SKATER.

První skupina metod (a), použitých v této práci, zahrnuje metodu jednoduchého spojení, metodu úplného spojení a metodu průměrovou. Metoda jednoduchého spojení je založena na hledání minimální vzdálenosti mezi nejbližšími základními prostorovými jednotkami v taxonomickém prostoru. Metoda úplného spojení naopak hledá minimální vzdálenost mezi nejodlehlejšími základními prostorovými jednotkami v taxonomickém prostoru. Průměrová metoda je v zásadě kompromisem obou předchozích metod – slučuje regionální třídy na základě nejnižší průměrné vzdálenosti mezi všemi základními prostorovými jednotkami v různých regionální třídách. V této práci je využita její nevážená i vážená varianta, která přidává větší váhu méně početným regionálním třídám.

Reprezentantem druhé skupiny je metoda centroidní. Centroidní metoda slučuje prostorové jednotky na základě minimální vzdálenosti mezi těžišti regionálních tříd, která se vypočítají jako průměr všech regionalizačních kritérií zahrnutých prostorových jednotek. U této metody je rovněž využita její vážená varianta.

Do třetí skupiny metod řadíme Wardovu metodu. Tato metoda pracuje na principu slučování základních prostorových jednotek do průběžných regionálních tříd tak, aby docházelo k co nejmenšímu nárůstu vnitrotřídní variability nově vzniklé regionální třídy.

Divizivní metody reprezentuje algoritmus SKATER (Spatial "K"luster Analysis by Tree Edge Removal). Algoritmus je založený na konstrukci minimální kostry grafu a je závislý na prostorové souvislosti základních prostorových jednotek. Regionalizační algoritmus pracuje postupným odsekáváním minimální kostry grafu v místech s nejvyšším potenciálním nárůstem mezitřídní variability nové iterace, dokud nedosáhne stanoveného počtu regionálních tříd ve finálním rozkladu, definovaném prostým počtem tříd, počtem základních prostorových jednotek v jednotlivých třídách nebo naplněním jiného specifikovaného kritéria. Funkci je možné využít k identifikaci přesného počtu finálních regionálních tříd i k jejich odhadu na základě pseudo-F statistiky (rozdílu mezi vnitrotřídní a mezitřídní variabilitou iterované na náhodných výsledcích). Optimální počet vzniklých regionálních tříd je ovšem možné enumerovat i pomocí iteračního algoritmu, který využívá strojové učení z náhodných permutací funkce. Algoritmus mimo základní výsledek vykalkuluje procento příslušnosti základní prostorové jednotky k regionální třídě finálního rozkladu, což je vlastnost použitelná v nehierarchických přístupech založených na fuzzy logice (Esri, 2021).

3.2.2 Nehierarchické metody formální regionální taxonomie

Nhierarchické metody formální regionální taxonomie, jak již jejich název napovídá, nevytváří průběžné regionální třídy, mezi kterými by vznikaly hierarchické vztahy. Namísto toho se snaží nejlhodněji vypočítat hodnotu tzv. účelové funkce – kritéria pro dosažení co efektivnějšího rozkladu. Algoritmy nehierarchických metod formální regionální taxonomie při hledání optimální hodnoty účelové funkce nejprve určí počáteční rozklad, tomu přiřadí základní hodnotu daného kritéria a následně se skrze přemísťování prostorových jednotek mezi vzniklými regionálními třídami snaží hodnotu účelové funkce zpřesnit. Výsledek se dostaví v okamžiku, kdy se hodnota účelové funkce přestává vlivem změny obsahu regionálních tříd měnit. Pro účely této diplomové práce jsou použity dvě ze tří hlavních skupin nehierarchických metod formální regionální taxonomie:

- metoda k-průměrů,
- metoda k-medoidů.

Metoda k-průměrů je analýzou těžišť regionálních tříd založená na minimalizaci účelové funkce, v tomto případě celkové variability regionálního systému.

Metoda k-medoidů (též metoda p-mediánů) je obdobou metody k-průměrů, využívá namísto těžiště shluků konkrétní centrálně položenou základní prostorovou jednotku (nebo více jednotek, v závislosti na počtu finálních regionálních tříd).

Speciální pozornost u nehierarchických metod zasluhuje problematika optimálního počtu regionálních tříd. Jak již bylo uvedeno v kapitole 2.3.6, účelová funkce rozkladu u nehierarchických metod formální regionální taxonomie vyžaduje informaci o konečném počtu regionálních tříd (počet tříd je podkladem pro určení počátečního rozkladu). Některé analytické programy využívají zabudovanou funkci odhadu optimálního počtu regionálních tříd. V rámci této práce je pro takový odhad využito metody siluety (KAUFMAN & ROUSSEEUW, 1990) a pseudo F-statistiky (CALIŃSKI & HARABASZ, 1974). Přesné hodnoty zjištěné metodou siluety jsou uvedeny v příloze 2.

3.3 Hodnocení výsledků

K hodnocení výsledků regionálně-taxonomické procedury lze přistupovat z několika různých perspektiv. Základním rozdílem je subjektivita či objektivita hodnocení – vzniklý systém je ohodnocen na základě geografické logiky nebo kvality regionálního systému.

Objektivní hodnocení kvality regionálního systému lze opět zásadně rozdělit podle struktury vzniklého systému na hierarchické a nehierarchické. Hodnocení tedy bude zaměřeno na nehierarchické systémy. Pro hierarchické metody formální regionální taxonomie lze uplatnit metody hodnocení založené na indexech hodnotících vzniklé série rozkladů (dendrogram, graf), u nehierarchických metod se zaměřujeme na posouzení kvality rozkladu, tedy správnosti přiřazení prostorových jednotek k regionální třídě.

U hierarchických metod je obvykle hodnocen soulad čtvercové kofenetické matice s maticí blízkosti, a to pomocí různých indexů (koeficient kofenetické korelace, Goodmanova gamma, Kruskalova gamma, Hubertova theta, Kritérium delta, Kruskalova míra stresu nebo vážená suma čtverců rozdílu. Různé řezy grafem mezi sebou porovnává Randův index. Nehierarchické metody, respektive jejich vzniklé rozklady, hodnotí analýzou matice blízkosti míra separace, metoda siluety nebo Dunnův index. Jiná skupina přístupů k hodnocení výsledků nehierarchických metod je založena na variabilitě regionálních tříd.

Pro hodnocení validity rozkladů vzniklých hierarchickými metodami pro tuto práci poslouží Koeficient kofenetické korelace. Tato metrika vzniká aplikací Pearsonova korelačního koeficientu na zpracovanou čtvercovou kofenetickou matici a matici blízkosti. Principem daného koeficientu je posoudit soulad mezi taxonomickou vzdáleností jednotlivých párů prostorových jednotek a informací o jejich “spojení“ do průběžné regionální třídy. Obdobně jako klasický Pearsonův korelační koeficient, nabývá hodnot v rozmezí [-1;1].

Pro zahrnutí rozkladů vzniklých nehierarchickými metodami lze využít již zmíněné metody siluety. Metodu siluety lze také využít pro rámcovou identifikaci optimálního počtu formálních regionálních tříd pro nehierarchické rozklady. V tomto případě ovšem hodnotí kompaktnost vzniklých regionálních tříd, a to pomocí identifikace správnosti zařazení prostorové jednotky do finálního shluku. Silueta je vypočítána pro každou prostorovou jednotku zvlášť a je rozdílem vzdálenosti prostorové jednotky ke všem jednotkám vlastního shluku a všem prostorovým jednotkám neblížejšího jiného shluku, podělená maximální vzdáleností některé z jednotek. Metrika nabývá hodnot v rozmezí [-1;1], kdy vyšší hodnoty představují vyšší platnost zařazení jednotky do dané regionální třídy.

3.4 Softwarové nástroje a datové zdroje

Pro efektivní plnění cílů práce, s přihlédnutím k výpočetní náročnosti prováděných úkonů, je třeba využití specializovaných softwarových nástrojů. Patří mezi ně programy z oblasti datové analýzy, geografických informačních systémů i obecných tabulkových procesorů. Výběr vhodných softwarových nástrojů pro plnění cílů této práce je závislý na faktorech týkajících se efektivity, komplementarity a přístupnosti. Problematika využívání softwarových nástrojů je dále popsána v kapitole 5.6. Následuje výčet použitého software:

- *MS Excel 365 v16* – zpracování tabelárních dat, příprava rozsáhlých datových souborů, zpracování vybraných grafických dat
- *Orange Data Mining 3.32* (DEMSAR & kol., 2013) – standardizace z-skóry (Orange Data Mining | Preprocess, 2016), analýza hlavních komponent (Orange Data Mining | PCA, 2016), testovací funkce k-průměrů (Orange Data Mining | k-Means, 2016), metoda siluety (Orange Data Mining | Silhouette Plot, 2016)
- *STATISTICA 12* (StatSoft Inc., 2013) – hierarchické shlukování (STATISTICA 12 User Manual | Joining (tree clustering), 2013)
- *MultiDendrograms 5.2* (FERNÁNDEZ & GÓMEZ, 2008) – koeficient kofenetické korelace (MultiDendrograms 5.2 Manual, 2021)
- *ArcGIS Pro 2.9* (Esri, 2021) – Prostorově ohraničené multivariační shlukování (How Spatially Constrained Multivariate Clustering works – ArcGIS Pro | Documentation, 2021), Multivariační shlukování (How Multivariate Clustering works – ArcGIS Pro | Documentation, 2021)

Mezi použité datové zdroje v souladu s kapitolou 3.1 a definovanými cíli práce patří:

- *Definitivní výsledky Sčítání lidu, domů a bytů 2011* (Český statistický úřad, 2011) – výběr z tabulek 111–123
- *ArcČR 500 v3.0* (ARCDATA PRAHA, 2012) – administrativní členění (vrstva obcí)

3.5 Metodologické shrnutí

Následující metodologické shrnutí vychází z obecné metodologie, uvedené na začátku kapitoly 3 a je jejím praktickým naplněním:

1. cílem regionální taxonomie je ověření chování vybraných prostorových struktur a identifikací vhodných metodologických postupů formální regionální taxonomie,
2. jako klasifikované objekty (základní prostorové jednotky) jsou arbitrárně zvoleny obce České republiky,
3. regionalizační kritéria jsou faktorované skalární proměnné cenzových dat za zvolené základní prostorové jednotky z roku 2011; použité operace s prostorovými daty pro získání vhodných regionalizačních kritérií zahrnují relativizaci, normalizaci, standardizaci a analýzu hlavních komponent,
4. blízkost v taxonomickém prostoru je vyjádřena pomocí Euklidovské, Manhattanské, a Čebyševovy vzdálenosti,
5. mezi zvolené regionalizační procedury patří hierarchické aglomerativní metody založené na poloze základních prostorových jednotek, poloze těžišť a vnitrotřídní variabilitě regionálních tříd, hierarchická divizivní metoda reprezentovaná algoritmem SKATER, a nehierarchické metody k-průměrů a k-medoidů,
6. vzniklé regionální systémy jsou hodnoceny z pohledu geografické logiky i z hlediska kvality regionálního systému (hodnocením matice blízkosti a variability regionálních tříd), je využíváno interních a relativních kritérií.

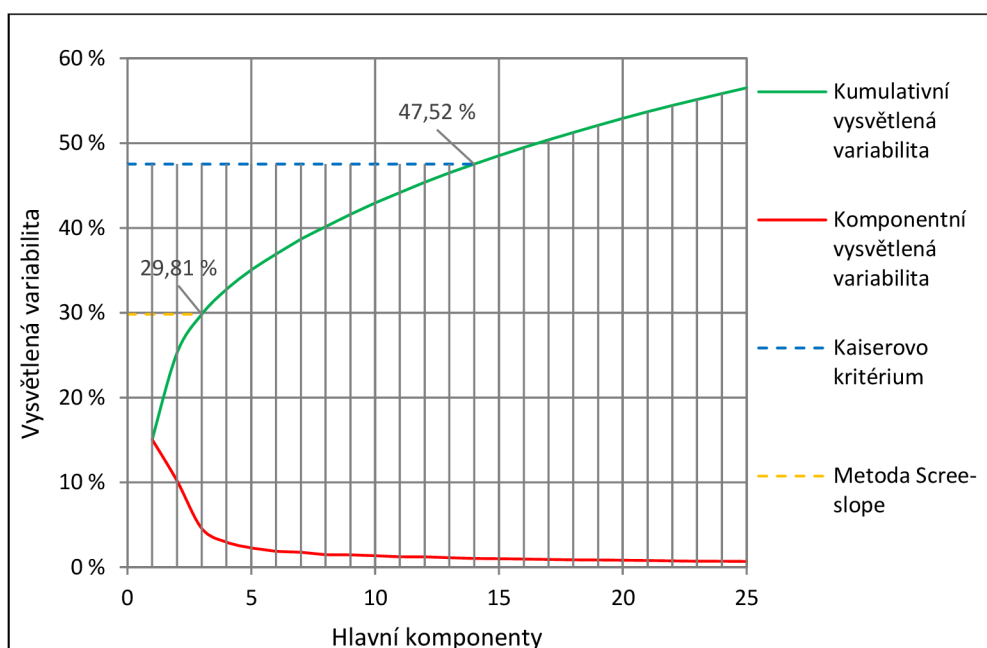
Z hlediska klasifikace vzniklých regionálně-taxonomických systémů [upraveno Klapkou (2019) podle (BEZÁK, 1993)] je můžeme na základě použitých metod označit za:

- formální – vzniklé na základě skalárních dat
- individuální i typologické – vzniklé regionální třídy se mohou a nemusí opakovat, což je závislé na podmínce zavedení prostorové souvislosti do regionalizační procedury
- ostré – u všech základních prostorových jednotek lze jednoznačně určit příslušnost k jediné regionální třídě
- nehierarchické – systémy obsahují pouze jednu finální úroveň rozkladu a netvoří hierarchii

4 Výsledky a diskuse

Tato kapitola obsahuje rozbor a hodnocení výsledků regionálně-taxonomických procedur provedených podle metodologie popsané v kapitole 3. V rámci postupu stanoveného v předcházející kapitole je nutné zdůraznit kroky společné pro všechny typy výsledků, jejich specifika jsou součástí jednotlivých oddílů diskuse.

V prvním kroku operací s prostorovými informacemi bylo u celkem 6251 základních prostorových jednotek (počet obcí České republiky v rozhodném okamžiku Sčítání lidu, domů a bytů 2011) zjištěno celkem 407 vhodných originálních regionalizačních kritérií. Téměř 30 % vybraných predikátů představovalo proměnné původního souboru poskytnutého Českým statistickým úřadem, zbylé indikátory byly relativizovány pomocí dostupných údajů. Ve druhém kroku byla provedena standardizace metodou z-skóre, která zároveň umožnila odstranění šikmostí a špičatostí vybraného souboru dat. Ve třetím kroku byla provedena samotná analýza hlavních komponent. Během extrakce komponent ve finální fázi analýzy hlavních komponent (obr. 7) bylo uplatněním Scree-slope metody a Kaiserova kritéria identifikováno 3, respektive 14 hlavních komponent, vysvětlujících variabilitu daného souboru predikátů z přibližně 30, respektive 48 %. Přesné hodnoty extrahovaných eigenvektorů jednotlivých komponent jsou uvedeny v rámci přílohy č. 1.



Obr. 7: Extrakce hlavních komponent využitím Scree-slope metody a Kaiserova kritéria
Zdroj: vlastní zpracování v MS Excel a Orange Data Mining; Data: ČSÚ

V následujícím kroku bylo provedeno testování za účelem zjištění optimálního počtu regionálních tříd pro nehierarchické rozklady a divizivní algoritmus SKATER. Testování bylo provedeno s extrahovanými hlavními komponenty obou metod předcházejícího kroku (Scree-slope metoda a Kaiserovo kritérium), s využitím analytického software Orange Data Mining. Jako testovací funkce byla zvolena metoda k-průměrů (metoda iniciace K-Means++) a jako zobrazovací funkce metoda siluety. Při využití tří nejsilnějších komponent, se hodnota siluety se zvyšujícím počtem regionálních tříd snižovala, při variantě se čtrnácti komponenty hodnota siluety kolísala velmi nízko – optimální počet regionálních tříd dle dostupných dat byl určen jako nejnížší možný počet, tedy 2. Iterační testování tak také pomohlo s výběrem vhodnější metody extrakce hlavních komponent, kterou byla na základě vyšších hodnot siluety zvolena metoda Scree-slope (3 predikáty).

Za účelem koncentrace výstupů jednotlivých metod jsou prezentovány výsledky s nejvyšší hodnotou siluety – jsou tedy upřednostňovány výsledky na základě validačních indexů namísto subjektivní geografické logiky. Pro hierarchické-aglomerativní metody, všechny výstupy obsahují grafická zobrazení v podobě dendrogramu s naznačením procentuálního řezu a také mapový výstup, doplněné evaluací struktury dendrogramu a hodnocením kartogramu z pohledu geografické logiky. Číslování shluků v legendách kartogramů jednotlivých metod vychází z technického provedení použitého software – slouží pouze pro referenční účely a nelze je považovat za relevantní faktor při vzájemném porovnání jednotlivých metod. Hodnocení validity (kvality regionálních systémů) jsou uvedena v samostatné kapitole 4.5 a obsahují u všech zpracovaných metod hodnocení metodou siluety, hierarchické-aglomerativní rozklady jsou navíc opatřeny analýzou kofenetické matice. Veškeré nevyužité rozklady hierarchických-aglomerativních metod jsou uvedeny v rámci příloh.

4.1 Hierarchické aglomerativní metody

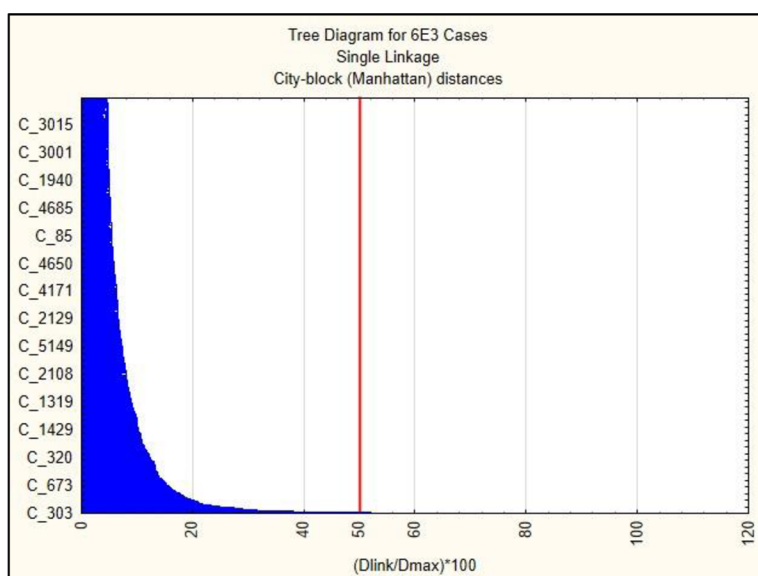
Hierarchické aglomerativní metody byly aplikovány pomocí programu STATISTICA, funkcí Analýza shluků. U všech hierarchických aglomerativních metod byly použity tři typy taxonomické vzdálenosti – Euklidovská, Manhattanská a Čebyševova. Zároveň bylo nutné určit úroveň porovnatelného rozkladu – zde byl autorem arbitrárně zvolen řez grafu v úrovni 50 a 25 procent finálního spojení. Celkově bylo v této kapitole sestaveno 42 výsledků (tab. 1). Mezi nejlepší výsledky z pohledu geografické logiky se řadí metoda úplného spojení a Wardova metoda.

Tab. 1: Rozsah regionálních tříd výsledků zvolených hierarchických-aglomerativních metod (počet shluků)

| | 50% | | | 25% | | |
|-----------------------------|------------------------|-------------------------|-----------------------|------------------------|-------------------------|-----------------------|
| | Euklidovská vzdálenost | Manhattanská vzdálenost | Čebyševova vzdálenost | Euklidovská vzdálenost | Manhattanská vzdálenost | Čebyševova vzdálenost |
| Metoda jednoduchého spojení | 11 | 10 | 11 | 75 | 76 | 83 |
| Metoda úplného spojení | 6 | 8 | 5 | 29 | 40 | 25 |
| Nevážená průměrová metoda | 2 | 9 | 16 | 22 | 59 | 83 |
| Vážená průměrová metoda | 4 | 6 | 5 | 24 | 38 | 30 |
| Nevážená centroidní metoda | 2 | 4 | 4 | 12 | 28 | 23 |
| Vážená centroidní metoda | 4 | 18 | 2 | 20 | 102 | 21 |
| Wardova metoda | 3 | 3 | 3 | 6 | 5 | 5 |

Zdroj: vlastní zpracování v MS Excel a STATISTICA; Data: ČSÚ

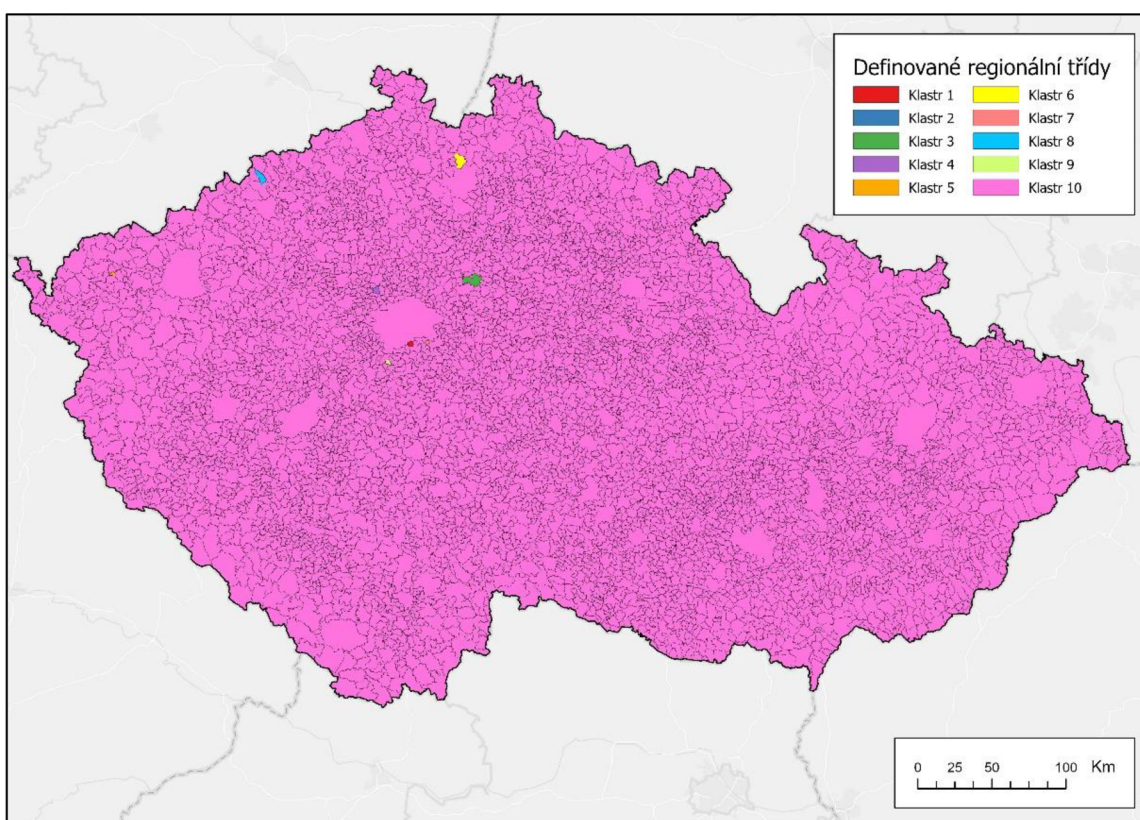
4.1.1 Metoda jednoduchého spojení



Obr. 8: Metoda jednoduchého spojení (manhattanská vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ

Metoda jednoduchého spojení průměrně vytvářela nejvyšší počet regionální tříd. Při analýze dendrogramu (obr. 8) byla zjištěna nevyvážená struktura, kdy bylo k-1 shluků tvořeno jedinou prostorovou jednotkou a všechny ostatní prostorové jednotky byly součástí jediné regionální třídy. Tento výsledek je zapříčiněn způsobem, jakým procedura konstruuje průběžné regionální třídy – základní prostorová jednotka s nejnižší další taxonomickou vzdáleností je ve většině případů přiřazena k již existující průběžné regionální třídě. Tento výsledek byl platný při využití všech tří využitých taxonomických vzdáleností, stejně tak u všech řezů dendrogramem. Obecně se tato metoda zdá zcela nepoužitelná v rámci zvolené geografické úlohy, její výsledky nelze dále rozvíjet a vzniklé prostorové struktury se zdají být zcela nahodilé.

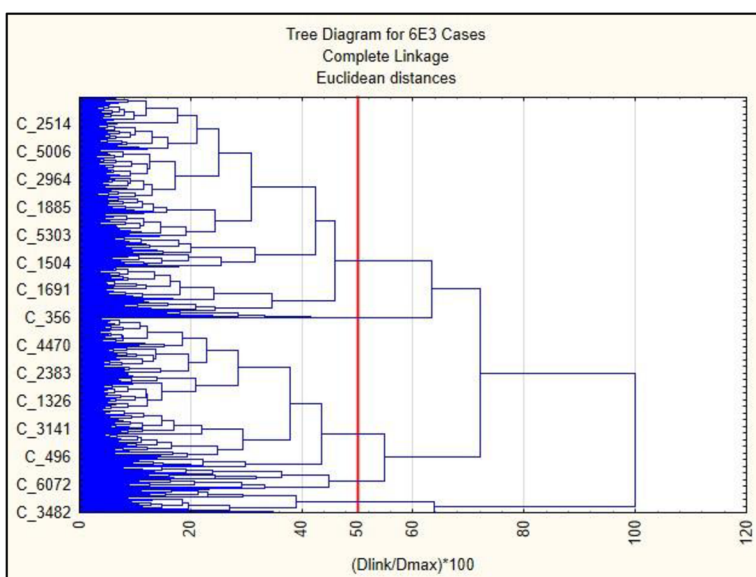


Obr. 9: Metoda jednoduchého spojení (manhattanská vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ, ARCDATA

Metoda jednoduchého spojení nevytváří žádné prostorové vzorce ve zkoumaném území (obr. 9). Shluk 10 se vyskytuje po celém zkoumaném území – ostatní shluky, obsahující každý jedinou prostorovou jednotku jsou rozmístěny pouze v oblasti Čech.

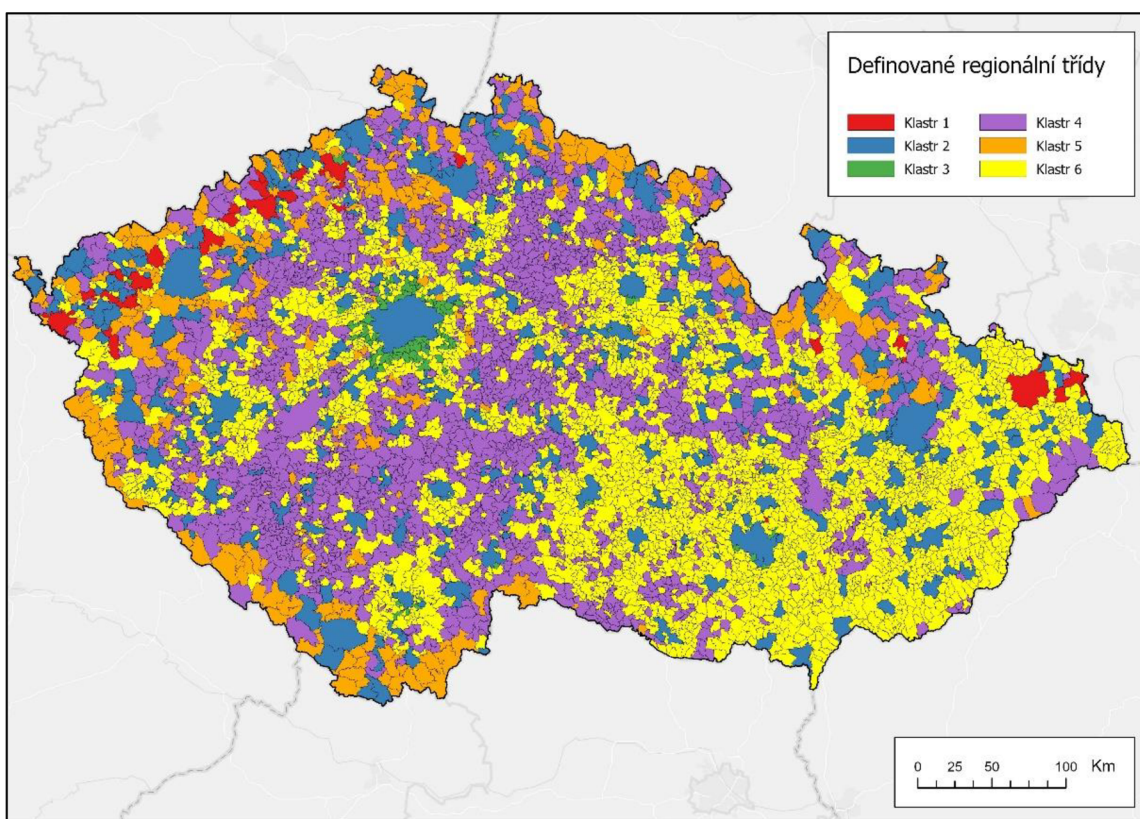
4.1.2 Metoda úplného spojení



Obr. 10: Metoda úplného spojení (euklidovská vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ

Metoda úplného spojení poskytuje oproti metodě jednoduchého spojení mnohem lepší výsledky. Počet regionálních tříd je v porovnatelných dimenzích nižší a jejich rozsah je relativně vyvážený (obr. 10).

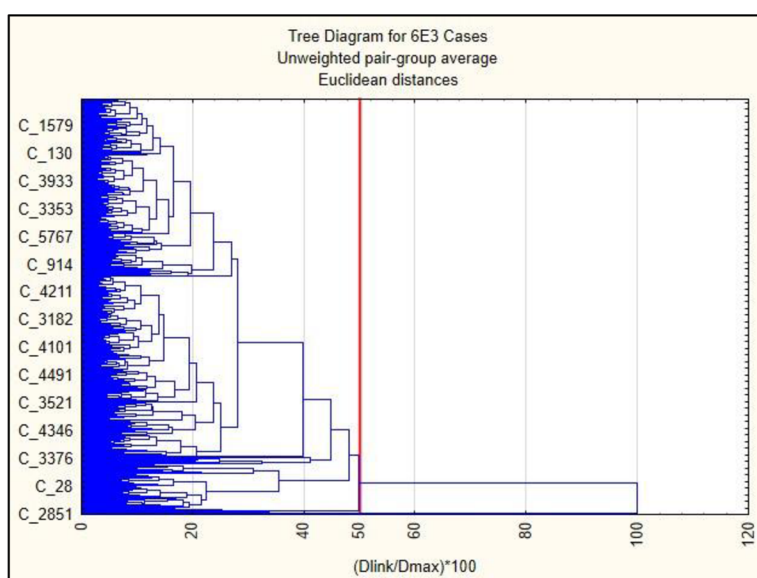


Obr. 11: Metoda úplného spojení (euklidovská vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ, ARCDATA

Z hlediska prostorových struktur zkoumaného území lze u metody úplného spojení pozorovat znatelné šablony (obr. 11). Shluk 1 vytváří nepravidelné struktury v severozápadní části území s několika odlehlými jednotkami na severovýchodě. Shluk 2 má velmi nepravidelnou územní distribuci, obsahuje ovšem množství jednotek rozsáhlých rozlohou i obyvatelstvem. Shluk 3 se vyskytuje výhradně v okolí hlavního města Prahy. Shluk 4 obsahuje prostorové jednotky umístěné zpravidla na vnitřních periferiích České republiky, rozsáhlejší seskupení vytváří zejména v jihozápadní části Čech. Jednotky příslušné ke shluku 5 jsou viditelné zejména na vnější periferii Čech. Shluk 6 se zastoupen zvláště na Moravě a také v okolí populačně silných jednotek shluku 2.

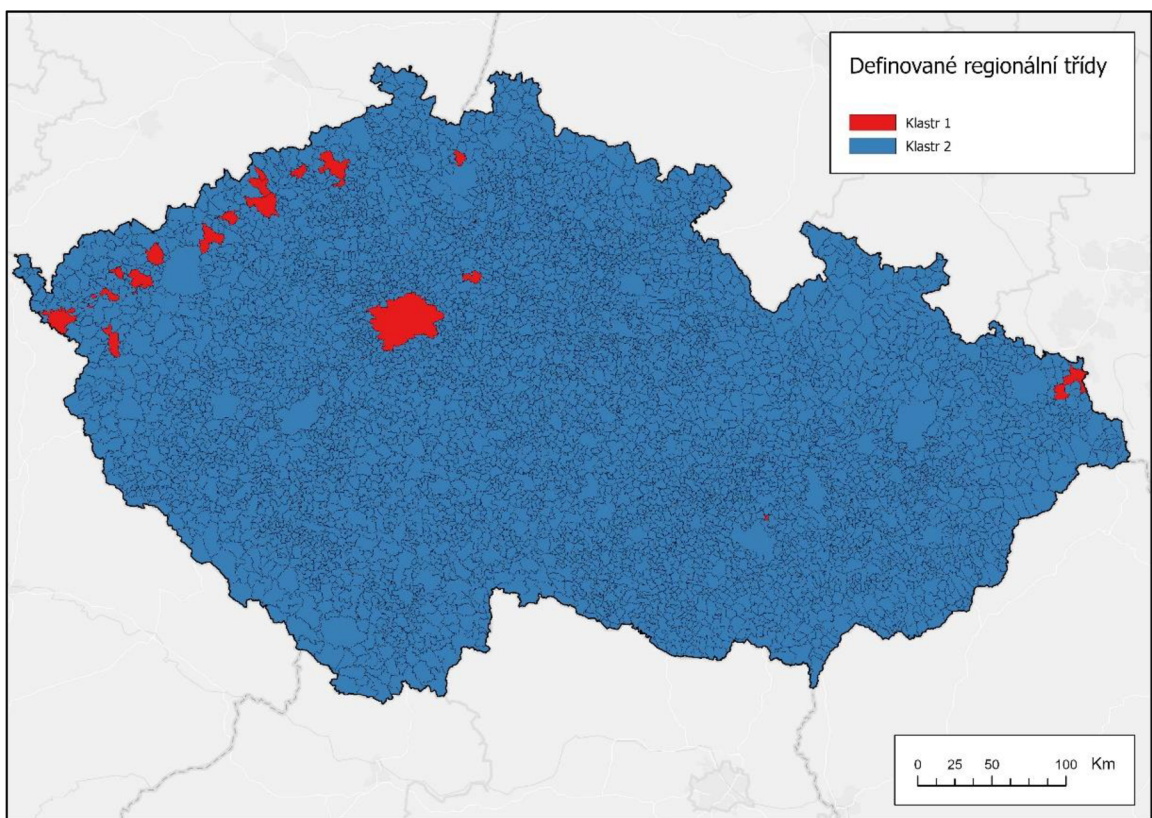
4.1.3 Metoda průměrová



Obr. 12: Nevážená průměrová metoda (euklidovská vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ

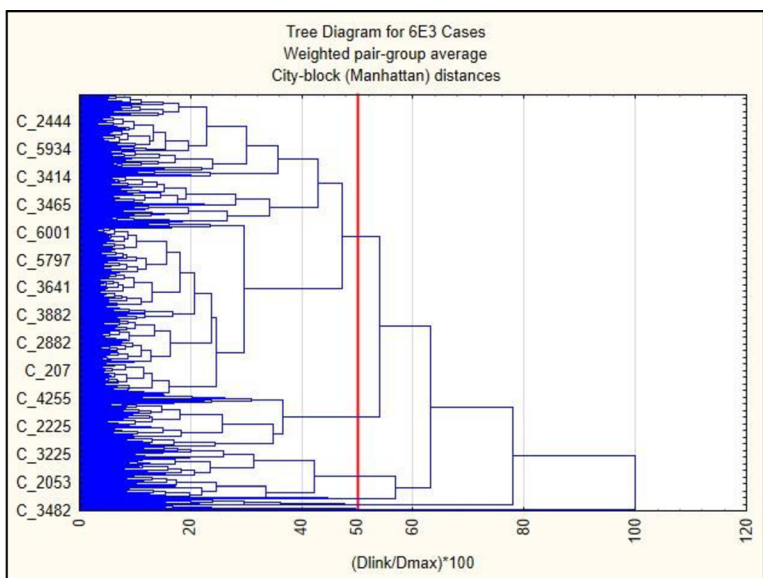
Prezentovaný výsledek nevážené průměrové metody na základě nejvyšší hodnoty siluety je pravděpodobně nejnevhodnějším příkladem z hlediska hodnocení na základě geografické logiky. Ve zvoleném výsledku se vyskytují pouze dvě regionální třídy (obr. 12), což je způsobeno způsobem spojení regionálních tříd v okolí zvoleného řezu grafem (průměr vzdálenosti připojovaného shluku 2 je natolik odlehlý, že finální spojení představuje téměř 200 % prezentovaného řezu).



Obr. 13: Nevážená průměrová metoda (euklidovská vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ, ARCDATA

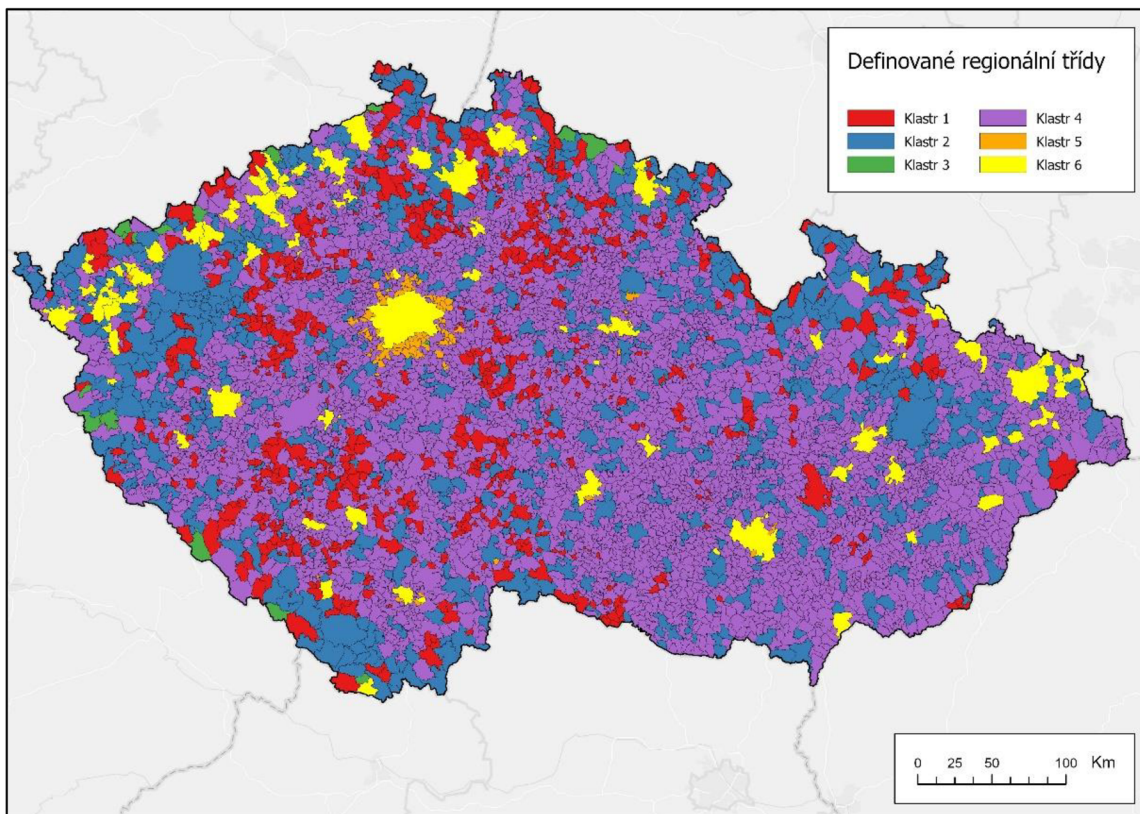
Vzhledem k limitům zvoleného výsledku lze pozorovat pouze omezené prostorové šablony (obr. 13). Shluk 1 se vyskytuje v celém zkoumaném území, zatímco shluk 2 obsahuje selekci náhodně spojených prostorových jednotek s vyšším populačním rozsahem.



Obr. 14: Vážená průměrová metoda (manhattanská vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ

Vážená varianta průměrové metody poskytuje relativně spolehlivé výsledky při použití všech kritérií. Preference nespojených prostorových jednotek v rámci vážení umožňuje “dřívější“ spojení odlehlých jednotek a celkově nižší maximální velikost dendrogramu (obr. 14).



Obr. 15: Vážená průměrová metoda (manhattanská vzdálenost, 50% řez grafu)

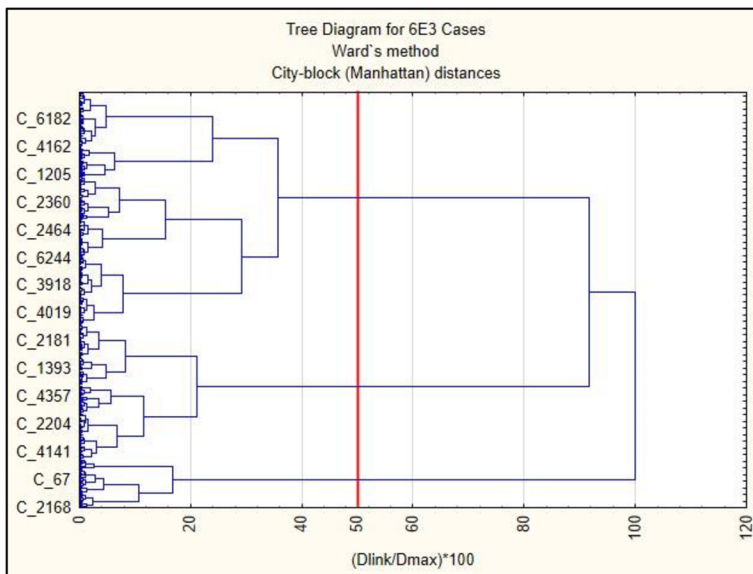
Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ, ARCDATA

Vážená průměrová metoda vytváří oproti její nevážené variantě vhodněji lokalizované shluky (obr. 15), podobně jako metoda úplného spojení a Wardova metoda. Z hlediska počtu jednotek jednotlivých shluků zcela dominuje klastr 4, který se vyskytuje téměř rovnoměrně po celém území. Prostorové jednotky seskupené ve shluku 1 jsou distribuovány náhodně s větším zastoupením v oblasti Čech. Shluk 2 vytváří seskupení zejména v okolí vojenských újezdů. Shluk 3 je zastoupen velmi řídko, a to výhradně v horských oblastech vnější periferie Čech. Shluk 5 se koncentruje pouze v okolí hlavního města a shluk 6 konečně obsahuje rozlehlá a populačně silná sídla.

4.1.4 Metoda centroidní

Centroidní metoda v případě zadaných kritérií neposkytuje příliš spolehlivé výsledky a na základě jejích parametrů není možné extrahovat dendrogram, což je limitace použitého statistického software (STATISTICA). Odhad počtu regionálních tříd požadovaných úrovní rozkladu je ovšem stále možný, a je uveden v prologu této kapitoly.

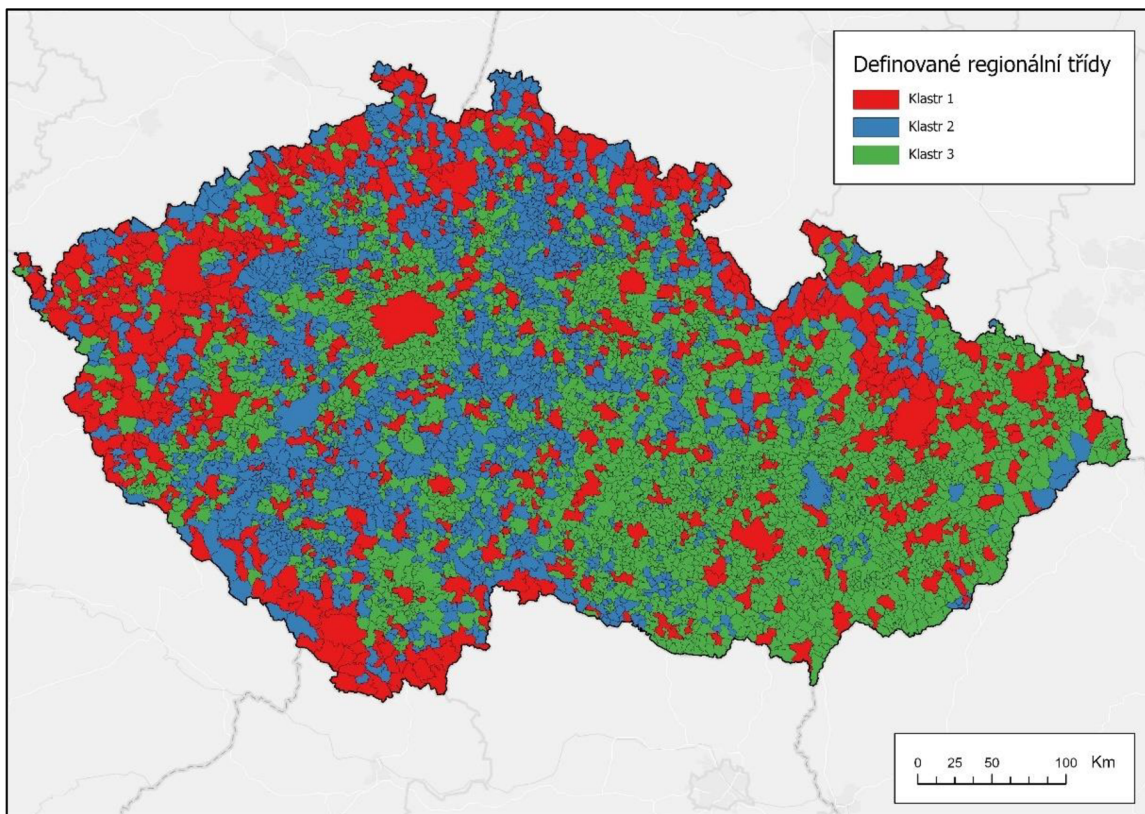
4.1.5 Wardova metoda



Obr. 16: Wardova metoda (manhattanská vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ

Wardova metoda poskytuje relativně spolehlivé výsledky při využití všech zvolených Minkowského metrik (obr. 16). U všech úrovní rozkladu vytváří vyvážené regionální třídy, což je důsledek konstrukce postupných regionálních tříd na základě nejnižšího nárůstu potenciální vnitrotřídní variability.



Obr. 17: Wardova metoda (manhattanská vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu ArcGIS Pro; Data: ČSÚ, ARCDATA

Konkrétní zvolený výsledek aplikace Wardovy metody (obr. 17) vykazuje obdobnou šablonovitost jako popsáný výsledek u metody úplného spojení a vážené průměrové metody. Ve shluku 1 nalezneme převážně prostorově a populačně rozsáhlé prostorové jednotky, včetně všech krajských měst; větší koncentrace lze pozorovat na vnější periferii zkoumaného území a také v západních Čechách a severovýchodní Moravě. Shluk 2 má významné zastoupení v oblasti vnitřní periferie Čech, zejména v jihovýchodní části. Prostorové jednotky shluku 3 se koncentrují v zázemí populačně silných jednotek a jsou také více zastoupeny na Moravě.

4.2 Hierarchické divizivní metody

V rámci hierarchických-divizivních metod byla implementována pouze jediná možnost – funkce Prostorově ohraničené multivariační shlukování (Spatially constrained multivariate clustering) v programu ArcGIS Pro, využívající divizivní algoritmus SKATER (Esri, 2021). Jako jediná z použitých metod v této práci nabízí zavedení podmínky prostorového sousedství přímo do iteračního algoritmu. Z tohoto důvodu bylo počítáno se třemi typy sousedství (sousedství hranou, sousedství hranou i rohem a odseknutá Delaunayova triangulace využívající Thyszenovy polygony). Dále bylo kalkulováno s počtem finálních regionálních tříd na základě nejlepší hodnoty siluety všech dostupných výsledků testovací sady (v Orange Data Mining, metoda k-průměrů, způsob iniciace K-means++) a také s výsledky danými výpočtem optimálního počtu regionálních tříd pomocí pseudo F-statistiky. Celkem bylo v této kapitole sestaveno 6 výsledků (tab. 2). Optimální počet shluků je pro obě použité metody identifikace stejný (dvě regionální třídy), proto se výsledky odlišují pouze na základě použitého typu sousedství.

Tab. 2: Rozsah regionálních tříd výsledků zvolených hierarchických-divizivních metod (počet shluků)

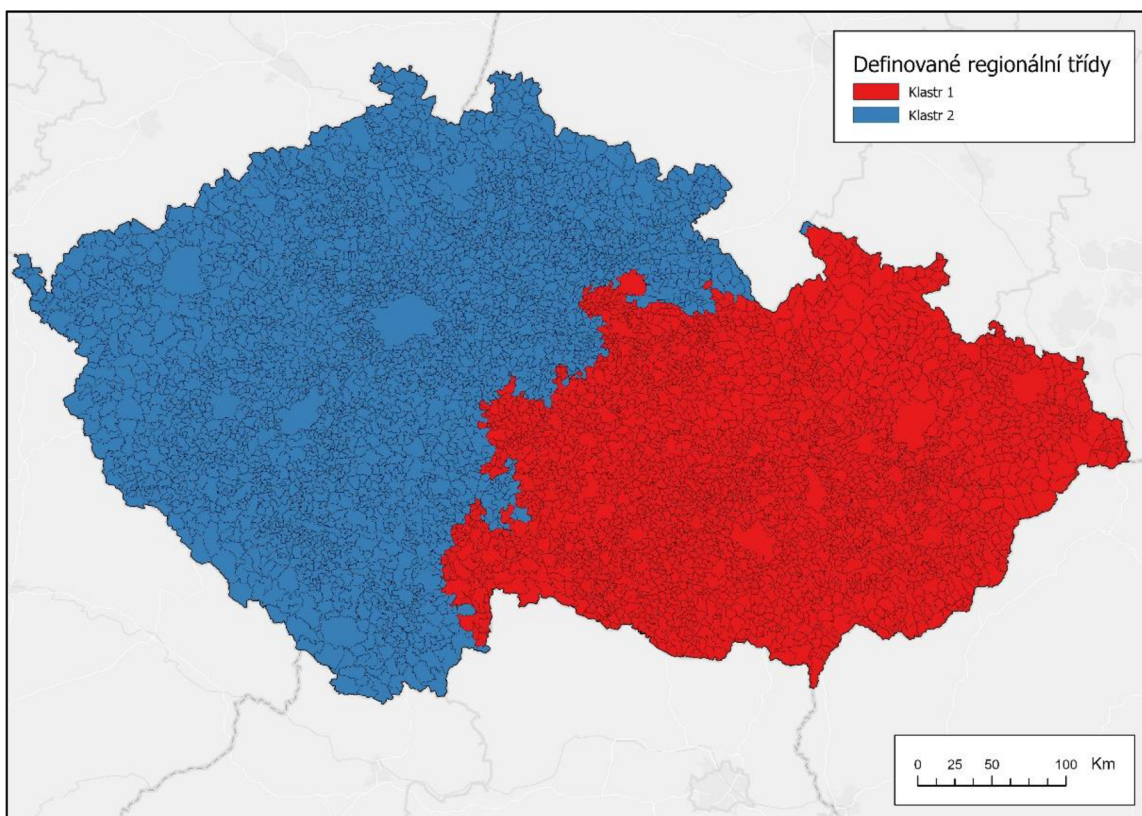
| | Metoda siluety | | | Caliński-Harabasz pseudo F-statistika | | |
|--------------------------|----------------|------|-----|---------------------------------------|------|-----|
| | RB | RBPH | ODT | PH | RBPH | ODT |
| Algoritmus SKATER | 2 | 2 | 2 | 2 | 2 | 2 |

Pozn.: PH = sousedství přilehlými hranami, RBPH = sousedství rohovými body přilehlých hran, ODT = sousedství odseknutou Delaunayovou triangulací

Zdroj: vlastní zpracování v MS Excel a ArcGIS Pro; Data: ČSÚ

4.2.1 Algoritmus SKATER

Zavedení prostorové souvislosti do regionalizačního algoritmu umožňuje v případě prezentovaného výsledku (obr. 18) pozorovat dichotomii zkoumaného území ve smyslu západ-východ, respektive Čechy-Morava. Zajímavostí je porušení podmínky prostorového sousedství v případě odlehlé jednotky v Javornickém výběžku. Technická nepřesnost výsledku regionalizační procedury je způsobena vlastností použitého typu sousedství. Odseknutá Delaunayova triangulace konceptualizuje geografický prostor do formy ostrých polygonů, kde některé odlehlé nesousední prostorové jednotky sdílí hranu.



Obr. 18: Algoritmus SKATER (odseknutá Delaunayova triangulace, 2 regionální třídy)
Zdroj: vlastní zpracování v programu ArcGIS Pro; Data: ČSÚ, ARCDATA

4.3 Nehierarchické metody

Nehierarchické metody byly zpracovány v programu ArcGIS Pro pomocí funkce Multivariační shlukování (Esri, 2021). Vyjma variací pro kapitoly 4.3.1 a 4.3.2 bylo využito dvou způsobů počáteční iniciace algoritmu, náhodné lokalizace vzorců a optimalizované lokalizace vzorců. Dále, obdobně jako v kapitole 4.2., bylo kalkulováno s počtem finálních regionálních tříd na základě nejlepší hodnoty siluety všech dostupných výsledků testovací sady (v Orange Data Mining, metoda k-průměrů, způsob iniciace K-means++) a také s výsledky danými výpočtem optimálního počtu regionálních tříd pomocí pseudo F-statistiky. Celkem bylo v této kapitole sestaveno 8 výsledků (tab. 3).

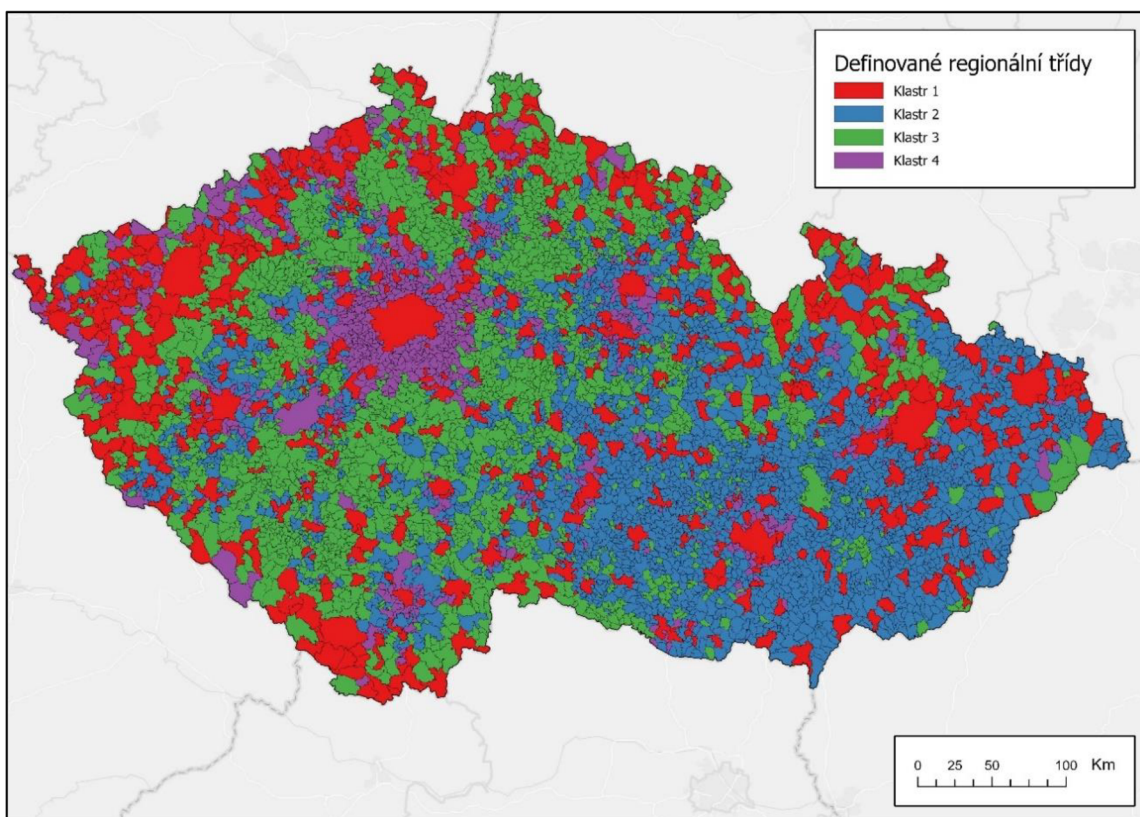
Tab. 3: Rozsah regionálních tříd výsledků zvolených nehierarchických metod (počet shluků)

| | Metoda siluety | | Caliński-Harabasz pseudo F-statistika | |
|------------------|----------------|-----|---------------------------------------|-----|
| | OLV | NLV | OLV | NLV |
| Metoda k-průměrů | 2 | 2 | 4 | 4 |
| Metoda k-medoidů | 2 | 2 | 4 | 4 |

Pozn.: OLV = optimalizované lokality vzorců, NLV = náhodné lokality vzorců

Zdroj: vlastní zpracování v MS Excel a ArcGIS Pro; Data: ČSÚ

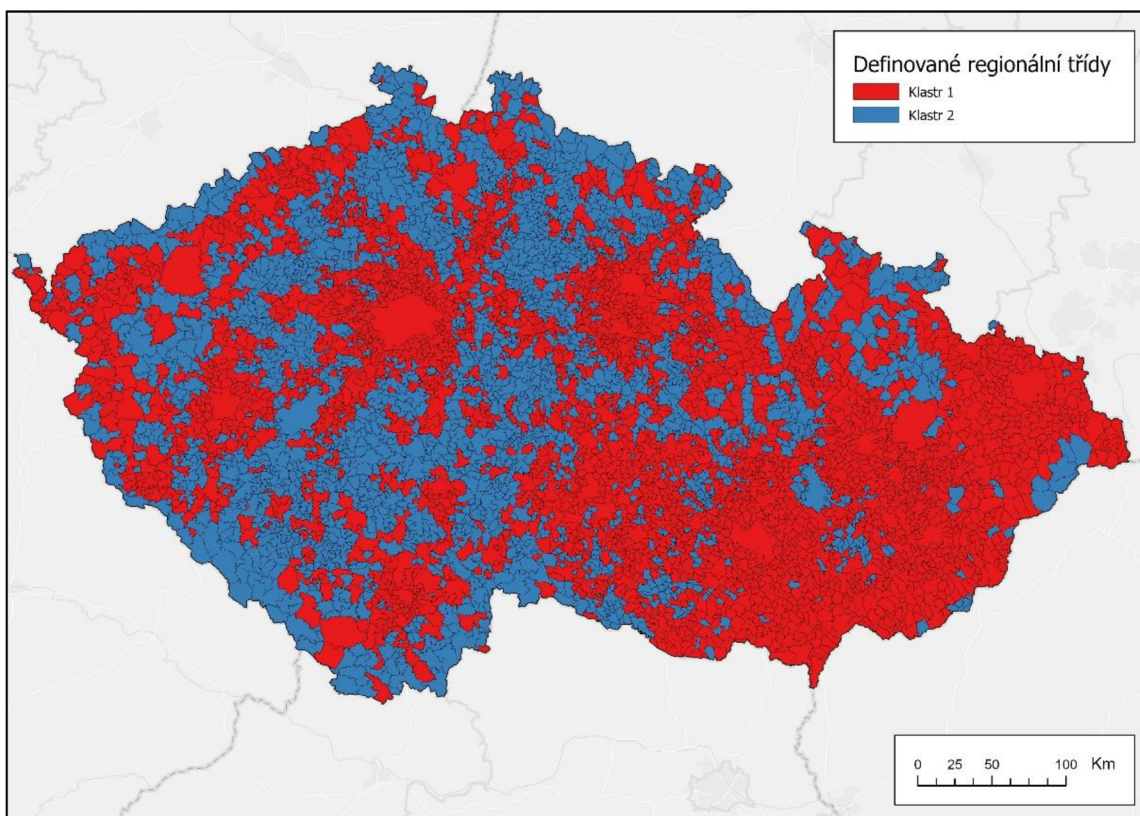
4.3.1 Metoda k-průměrů



Obr. 19: Metoda k-průměrů (náhodné lokality vzorců, 4 regionální třídy)

Zdroj: vlastní zpracování v programu ArcGIS Pro; Data: ČSÚ, ARCDATA

Metoda k-průměrů ve variantě s nejvyšší hodnotou siluety vytváří čtyři shluky (obr. 19). Obdobně jako u některých hierarchických metod i zde pozorujeme regionální třídy shlukující prostorové jednotky s vyšším populačním potenciálem anebo větší rozlohou (shluk 1), regionální třídy venkovské Moravy a Čech (shluk 2 a 3) a také klastry zázemí některých velkých měst (shluk 4).



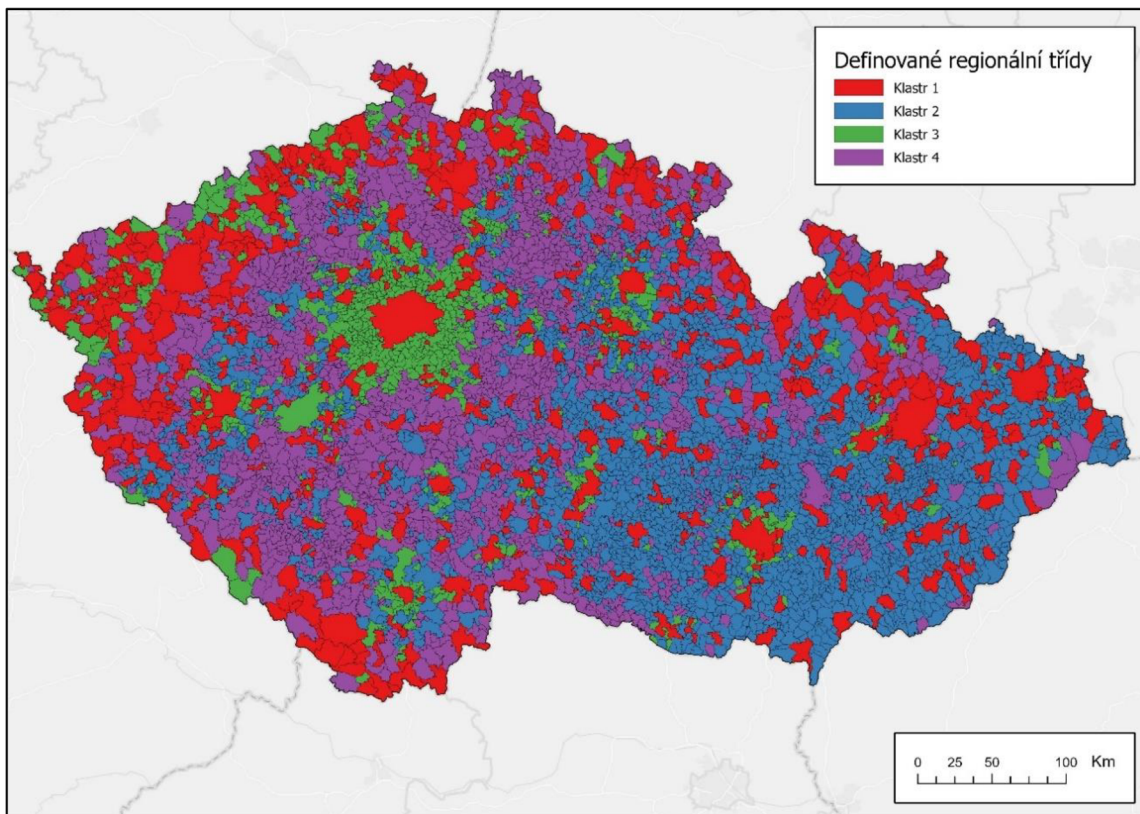
Obr. 20: Metoda k-průměrů (optimalizované lokality vzorců, 2 regionální třídy)

Zdroj: vlastní zpracování v programu ArcGIS Pro; Data: ČSÚ, ARCDATA

Pro referenci je na tomto místě vhodné uvést i variantu výsledku metody k-průměrů pro 2 regionální třídy (obr. 2). Nejdůležitějším poznatkem je identifikace šablon v území, které vznikají injektací shluku 1 do struktur shluku 2. Zde pozorujeme shlukování prostorových jednotek klastru 1 v oblastech populačně silných území se znatelným rozvojem po liniích páteřní dopravní infrastruktury. Zvláště patrný je tento jev v případě Pražské metropolitní oblasti, která vytváří amébovitý tvar s rozvojem po liniích dálniční sítě směrem k populačně silným prostorovým jednotkám. Významné shlukování lze pozorovat v rámci Plzeňské, Budějovické a Hradecko-Pardubické aglomerace, protáhlý klastr je viditelný i v oblasti Podkrušnohoří. Pokud opět vztáhneme příslušnost prostorových jednotek shluku 2 k populační velikosti zahrnutých obcí, lze konstatovat, že populačně slabá území se v Moravské části vyskytují pouze v omezeném rozsahu. Chování prostorových jednotek

v Čechách a na Moravě je v případě tohoto výsledku velmi rozdílné (obdobně jako u dichotomického výsledku algoritmu SKATER).

4.3.2 Metoda k-medoidů



Obr. 21: Metoda k-průměrů (náhodné lokality vzorců, 4 regionální třídy)

Zdroj: vlastní zpracování v programu ArcGIS Pro; Data: ČSÚ, ARCDATA

Metoda k-medoidů poskytuje výsledky (obr. 21) velmi obdobné metodě k-průměrů. Rozdílem jsou přeskupené obce v řádu jednotek a celkový dopad nijak nemění prostorové šablony uvedené u metody k-průměrů.

4.5 Validita formálních regionálních systémů

Validita vzniklých regionálních systémů byla posuzována dvěma způsoby: metodou siluety a analýzou kofenetické matice (pouze hierarchické-aglomerativní systémy). Metoda silueta byla aplikována pomocí funkce Silhouette plot v programu Orange Data Mining. Silueta bylo vypočítána užitím dvou typů vzdálenosti (Euklidovská a Manhattanská) – hodnoty uvedené v tabulce představují průměr siluety všech prostorových jednotek analyzovaného výsledku. Výsledky s nejvyšší hodnotou siluety byly využity pro analýzu dendrogramů a kartogramů v předcházejících kapitolách. Analýza kofenetické matice posloužila pro zhodnocení souladu vzniklých hierarchických struktur s původními vztahy mezi základními prostorovými jednotkami.

Tab. 4: Hodnocení validity hierarchických-aglomerativních regionálních systémů metodou siluety

| | | 50% | | | 25% | | |
|-----------------------------|-------------------------|------------------------|-------------------------|-----------------------|------------------------|-------------------------|-----------------------|
| | | Euklidovská vzdálenost | Manhattanská vzdálenost | Čebyševova vzdálenost | Euklidovská vzdálenost | Manhattanská vzdálenost | Čebyševova vzdálenost |
| Metoda jednoduchého spojení | Euklidovská vzdálenost | 0,387 | 0,439 | 0,400 | -0,179 | -0,161 | -0,197 |
| | Manhattanská vzdálenost | 0,383 | 0,444 | 0,384 | -0,129 | -0,102 | -0,146 |
| Metoda úplného spojení | Euklidovská vzdálenost | 0,272 | 0,233 | 0,275 | 0,170 | 0,138 | 0,154 |
| | Manhattanská vzdálenost | 0,304 | 0,269 | 0,300 | 0,230 | 0,205 | 0,210 |
| Nevážená průměrová metoda | Euklidovská vzdálenost | 0,684 | 0,267 | 0,228 | 0,150 | 0,165 | 0,163 |
| | Manhattanská vzdálenost | 0,646 | 0,294 | 0,261 | 0,196 | 0,227 | 0,232 |
| Vážená průměrová metoda | Euklidovská vzdálenost | 0,148 | 0,268 | 0,203 | 0,150 | 0,116 | 0,036 |
| | Manhattanská vzdálenost | 0,196 | 0,294 | 0,242 | 0,202 | 0,178 | 0,092 |
| Wardova metoda | Euklidovská vzdálenost | 0,277 | 0,293 | 0,295 | 0,242 | 0,180 | 0,241 |
| | Manhattanská vzdálenost | 0,302 | 0,319 | 0,314 | 0,267 | 0,223 | 0,265 |

Zdroj: vlastní zpracování v MS Excel, Orange Data Mining a STATISTICA; Data: ČSÚ

Z hlediska hodnocení validity metodou siluety (tab. 4) byla pro hierarchické-aglomerativní metody zaznamenána nejvyšší hodnota u 50% řezu nevážené průměrové metody zpracované za využití euklidovské vzdálenosti – tato metoda při analýze dendrogramu a kartogramu paradoxně poskytovala nespolehlivé výsledky z pohledu geografické logiky. Stabilní hodnoty siluety vykazovala Wardova metoda, která tedy představuje relativně nejvhodnější metodu z hlediska správnosti přiřazení prostorových jednotek do finálních regionálních tříd. Je ovšem třeba nutně konstatovat, že výsledné struktury regionálních systémů u rozkladů s hodnotami siluety pod 0,5 by nebylo vhodné považovat za “silné“.

Tab. 5: Hodnocení validity hierarchických-aglomerativních regionálních systémů analýzou kofenetické matice

| | Euklidovská vzdálenost | Manhattanská vzdálenost | Čebyševova vzdálenost |
|------------------------------------|-------------------------------|--------------------------------|------------------------------|
| Metoda jednoduchého spojení | 0,639 | 0,642 | 0,598 |
| Metoda úplného spojení | 0,675 | 0,510 | 0,644 |
| Nevážená průměrová metoda | 0,726 | 0,689 | 0,759 |
| Vážená průměrová metoda | 0,654 | 0,594 | 0,476 |
| Nevážená centroidní metoda | 0,746 | 0,731 | 0,712 |
| Vážená centroidní metoda | 0,671 | 0,507 | 0,490 |
| Wardova metoda | 0,514 | 0,486 | 0,555 |

Zdroj: vlastní zpracování v MS Excel, Multidendrograms a STATISTICA; Data: ČSÚ

Pro hodnocení validity regionálních systémů analýzou kofenetické matice byl využit koeficient kofenetické korelace (tab. 5). Nejsilnější korelace byla zaznamenána u rozkladů vzniklých neváženou průměrovou metodou a neváženou centroidní metodou. Nejhůře hodnocenou metodou byla naopak Wardova metoda, která produkovala spolehlivé výsledky při hodnocení geografickou logikou. V tomto případě lze odchylku odvodit od odlišného principu budování průběžných regionálních tříd (minimalizace nárůstu vnitrotřídní variability). Wardova metoda také pracuje nejlépe se čtvercem Euklidovské vzdálenosti, a tato nebyla ve výzkumu zařazena.

Tab. 6: Hodnocení validity hierarchických-divizivních regionálních systémů metodou siluety

| | | Metoda siluety (2 shluky) | | | Kaliński-Harabasz pseudo F-statistika (2 shluky) | | |
|--------------------------|--------------------------------|----------------------------------|-------------|------------|---|-------------|------------|
| | | PH | RBPH | ODT | PH | RBPH | ODT |
| Algoritmus SKATER | Euklidovská vzdálenost | 0,087 | 0,087 | 0,097 | 0,087 | 0,087 | 0,097 |
| | Manhattanská vzdálenost | 0,131 | 0,131 | 0,144 | 0,131 | 0,131 | 0,144 |

Pozn.: PH = sousedství přilehlými hranami, RBPH = sousedství rohovými body přilehlých hran, ODT = sousedství odseknutou Delauneyovou triangulací

Zdroj: vlastní zpracování v MS Excel, Orange Data Mining a ArcGIS Pro; Data: ČSÚ

Hodnoty siluety výsledků všech hierarchických-divizivních metod byly velmi nízké (tab. 6) a u metod bez omezení prostorového sousedství by znamenaly slabou strukturu regionálního systému. Zavedení podmínky prostorového sousedství do algoritmu SKATER ovšem učinilo hodnocení validity regionálního systému metodou siluety neprůkazným.

Tab. 7: Hodnocení validity nehierarchických regionálních systémů metodou siluety

| | | Metoda siluety (2 shluky) | | Kaliński-Harabasz pseudo F-statistika (4 shluky) | |
|-----------------------------|------------------------------------|------------------------------|-------|--|-------|
| | | OLV | NLV | OLV | NLV |
| Metoda k-průměrů | Euklidovská vzdálenost | 0,259 | 0,257 | 0,317 | 0,317 |
| | Manhattanská vzdálenost | 0,311 | 0,307 | 0,347 | 0,347 |
| Metoda k-medoidů | Euklidovská vzdálenost | 0,177 | 0,258 | 0,314 | 0,314 |
| | Manhattanská vzdálenost | 0,242 | 0,310 | 0,345 | 0,345 |

Pozn.: OLV = optimalizované lokality vzorců, NLV = náhodné lokality vzorců

Zdroj: vlastní zpracování v MS Excel, Orange Data Mining a ArcGIS Pro; Data: ČSÚ

Pro nehierarchické rozklady byla hodnota siluety nejvyšší u výsledků se čtyřmi regionálními třídami (tab. 7). Tento paradox (počet 2 finálních tříd pro nehierarchické rozklady byl ustanoven právě za využití metody siluety) byl způsoben kombinací rozdílných způsobů iniciace algoritmu a maximálního počtu povolených iterací algoritmu v rámci dvou různých softwarových prostředí.

4.6 Omezení výzkumu

Charakter dostupných statistických dat pro časové okamžiky uvedené v zadání práce (census 2001 a 2011) neumožnil naplnění požadavku časové porovnatelnosti, a tedy zachycení dynamiky prostorových jevů.

Centroidní hierarchická-aglomerativní metoda zpracovávaná v programu STATISTICA neumožňovala kompletní zpracování z důvodu technické nedokonalosti použitého software.

Šířeji uchopené téma vynutilo použití širokého spektra proměnných, což se odrazilo na validitě zpracovaných regionálních systémů.

Výsledky regionálně-taxonomické procedury jsou velmi závislé na možnostech použitého software.

Nutnost vyhledávat funkčně “kompatibilní“ softwarové nástroje.

Dosavadní neexistence softwarového nástroje, který by umožňoval komplexní regionálně-taxonomickou proceduru (úprava dat, zpracování, zobrazení, hodnocení).

Závěr

Sestavení výsledků označovaných jako formální regionálně-taxonomické systémy umožnilo identifikaci vhodných a nevhodných metod, to vše na příkladu obcí České republiky v upravených cenzových datech z roku 2011. Dostupná data za obce České republiky byla optimalizována metodami relativizace, normalizace a standardizace. Dále byla aplikována analýza hlavních komponent jako metoda pro snížení dimenze původních proměnných. Upotřebením metody Scree-slope byly zjištěny celkem 3 principiální komponenty, které byly dále využity jako regionalizační kritéria vícerozměrné shlukové analýzy.

V intencích definovaných cílů práce bylo aplikováno 10 různých metod formální regionální taxonomie. Každá metoda byla zpracována v několika odlišných variantách a celkem bylo zkompileováno 56 výsledků. Jednotlivé metody formální regionální taxonomie byly zhodnoceny subjektivními i objektivními způsoby.

Na základě hodnocení dendrogramů a kartogramů u hierarchických-aglomerativních metod byla označena za vhodnou metoda úplného spojení a Wardova metoda. Obě tyto metody vytvářely vyvážené dendrogramy, jejich prostorová zobrazení vykazovala znatelnou šablonovitost a bylo možné díky nim identifikovat prostorové struktury (založené zejména na populačním potenciálu prostorových jednotek a strukturální dichotomii makroregionů Česka). Obdobné vzory se vyskytovaly i u použité hierarchické-divizivní metody (algoritmu SKATER) a u metod nehierarchických (metody k-průměrů a k-medoidů), kde byly analyzovány pouze kartografické výstupy. V rámci objektivních metod hodnocení vzniklých regionálních systémů byly implementovány validační indexy pro hodnocení kvality rozkladů. Všechny dostupné výsledky byly hodnoceny metodou siluety. Silueta ukázala pro naprostou většinu metod slabou strukturu regionálního systému, což byl převážně důsledek zadání výzkumné úlohy (zejména široké spektrum využitých prostorových dat). Průkaznější řešení nabídla analýza kofenetické matice hierarchických-aglomerativních metod. Zde byly jako rozklady nejméně zkreslující původní strukturu proměnných označeny regionální systémy zpracované neváženou průměrovou metodou a neváženou centroidní metodou.

V návaznosti na zjištěné poznatky lze další výzkum směřovat na konkrétní lokačně-alokační úlohy, které by vyžadovaly specifický rozsah regionalizačních kritérií. Formální regionálně-taxonomické procedury bez omezení prostorové souvislosti lze aplikovat spíše ve specifických regionálně-typologických úlohách než při delimitaci obecně-geografických regionů nebo území s uplatněním konceptu jádro-periferie.

Summary

The compilation of results, termed formal regional-taxonomy systems, enabled the identification of appropriate and inappropriate methods in the example of municipalities in the Czech Republic, using adapted census data from 2011. Municipality data of the Czech Republic were optimized through relativization, normalization and standardization methods. The main component analysis was applied to reduce the dimension of the original variables. Having applied the Scree-slope method, three principal components were defined and further utilized as regionalization criteria for the multivariate cluster analysis.

Based on the goals of this thesis, ten different formal regional taxonomy methods were applied in several differing variants, resulting in 56 outcomes. Furthermore, each formal regional taxonomy method was evaluated by subjective and objective means.

Based on the evaluation of dendrograms and cartograms of hierarchical-agglomerative methods, the Complete-linkage method and Ward's method proved to be the most conclusive. Both methods produced balanced dendrograms, and their spatial distribution evinced considerable patterns that enabled the identification of spatial structures (primarily based on spatial units and their population potential and macroregional structural dichotomy in the Czech Republic). Similar patterns occurred using the hierarchical-divisive method (SKATER algorithm) and non-hierarchical methods (k-means clustering and k-medoids clustering), where cartographic outcomes were analysed solely. Validation indexes were implemented as objective evaluation methods to assess disintegration quality. All of the available data were assessed by the silhouette method. As a result of the thesis assignment (particularly the broad spectrum of spatial data), the silhouette exhibited poor structure in the majority of regional systems. However, the cophenetic matrix analysis of hierarchic-agglomerative methods showed a more demonstrable result. In this analysis, the Unweighted group mean average method and Unweighted centroid method were determined as the least distorting disintegrations of the original variable structure.

As a follow-up on the findings of this thesis, further research concerning specific location-allocation functions with characteristic regionalization criteria is desirable. Formal regional taxonomic procedures that are not spatially restricted may be applied in specific regional-typologic assignments, contrary to general geographic regions or territories under the core-periphery concept.

Reference

- ARCDATA PRAHA. (Říjen 2012). ArcČR 500 3.0 | Digitální geografická databáze 1:500 000.
- BAŠOVSKÝ, O., & LAUKO, V. (1990). *Úvod do regionálnej geografie*. Bratislava: SPN.
- BEZÁK, A. (1993). Problémy a metódy regionálnej taxonomie. V *Geographia Slovaca* 3. Bratislava: GÚ SAV.
- BLAŽEK, M. (1977). *Ekonomická Regionalizace*. Brno: Geografický ústav ČSAV.
- CALIŇSKI, T., & HARABASZ, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics* 3, 1, 1–27.
- Český statistický úřad. (26. 03 2011). Definitivní výsledky SLDB 2011 | tabulky 111–123. Praha.
- DĚDINA, V. (1929). Regionální členění Československa. *Sborník Československé společnosti Zeměpisné*, 3–4, stránky 89–92.
- DEMEK, J., RAUŠER, J., & QUITT, E. (1977). Fyzickogeografické regiony České socialistické republiky. *Sborník Československé společnosti zeměpisné*, 2, stránky 89–102.
- DEMSAR, J., & kol. (8 2013). Orange: Data Mining Toolbox in Python. *Journal of Machine Science*, 14, stránky 2349–2353.
- DOMAŇSKI, R. (1982). *Teoretyczne podstawy geografii ekonomicznej*. Warszawa: Państwowe Wydawnictwo Naukowe.
- ERLEBACH, M., TOMÁŠ, M., & TONEV, P. (2016). A functional interaction approach to the definition of meso regions: The case of the Czech republic. *Moravian Geographical Reports*, 2, stránky 37–46.
- Esri. (2021). ArcGIS Pro 2.9. Esri.
- Esri. (11. 11 2021). *How Multivariate Clustering works – ArcGIS Pro | Documentation*. Získáno 29. 03 2022, z GIS Mapping Software, Location Intelligence & Spatial Analytics | Esri: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-multivariate-clustering-works.htm>
- Esri. (11. 11 2021). *How Spatially Constrained Multivariate Clustering works – ArcGIS Pro | Documentation*. Získáno 29. 03 2022, z GIS Mapping Software, Location Intelligence & Spatial Analytics | Esri: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-spatially-constrained-multivariate-clustering-works.htm>
- FERNÁNDEZ, A., & GÓMEZ, S. (2008). Solving Non-Uniqueness in Agglomerative Hierarchical Clustering Using Multidendrograms., *Journal of Classification*, 25, stránky 43–65.
- FRIEDMANN, J. (1956). The concept of planning region. *Land economics*, 32(1), stránky 1–13.
- GAN, G., MA, C., & WU, J. (2007). Data clustering: theory, algorithms, and applications. V *ASA SIAM Series on statistics and applied probability*. Alexandria, VA: SIAM, Philadelphia – ASA.
- GÓMEZ, S. (2021). MultiDendrograms 5.2 Manual. Načteno z <https://deim.urv.cat/~sergio.gomez/download.php?f=multidendrograms-5.2-manual.pdf>
- HAGGETT, P. (1965). *Locational analysis in human geography*. London: Edward Arnold.
- HAGGETT, P. (2001). *Geography: A Global Synthesis*. Harlow: Prentice Hall.
- HALÁS, M., & KLAPKA, P. (2010). Regionalizace Česka z hlediska modelování prostorových interakcí. *Geografie*, 2, stránky 144–160.
- HALÁS, M., KLAPKA, P., & TONEV, P. (2014). A contribution to human geographical regionalisation of the Czech republic at the mezzo level. (V. KLÍMOVÁ, & V. ŽÍTEK, Editoři) *17th International Colloquium on Regional Science. Conference Proceedings*, 715–721.
- HAMPL, M. (1971). *Teorie komplexity a diferenciace světa*. Praha: Univerzita Karlova v Praze.
- HAMPL, M. (1998). *Realita, společnost a geografická organizace: Hledání integrálního řádu*. Praha: Univerzita Karlova v Praze.
- HAMPL, M. (2002). Regionální organizace společnosti: principy a problémy studia. *Geografie*, 107(4), stránky 333–348.

- HAMPL, M. (2005). *Geografická organizace společnosti v České republice: transformační procesy a jejich obecný kontext*. Praha: DemoArt.
- HAMPL, M., & MARADA, M. (2015). Sociogeografická regionalizace Česka. *Geografie*, 3, stránky 397–421.
- HAMPL, M., & MÜLLER, J. (1996). Komplexní organizace systému osídlení. V M. HAMPL, *Geografická organizace společnosti a transformační procesy v České republice* (stránky 53–89). Praha: DemoArt.
- HAMPL, M., GARDAVSKÝ, V., & KÜHNEL, K. (1987). *Regionální struktura a vývoj systému osídlení ČSR*. Praha: Univerzita Karlova v Praze.
- HAMPL, M., JEŽEK, J., & KÜHNEL, K. (1978). *Sociálně geografická regionalizace ČSR*. Praha: Acta Demographica II. Československá demografická společnost při ČSAV.
- HEŘMANOVÁ, E. (1991). *Vybrané vícerozměrné statistické metody v geografii*. Praha: Univerzita Karlova – SPN.
- JOHNSTON, R. J. (1976). Classification in Geography. V *Catmog 6*. Norwich: GeoAbstracts.
- JOHNSTON, R. J. (2000). *The dictionary of human geography*. Oxford: Blackwell.
- KAUFMAN, L., & ROUSSEEUW, P. J. (1990). *Finding groups in data: an introduction to cluster analysis*. Hoboken: John Wiley and Sons.
- KLAPKA, P. (2019). *Regiony a regionální taxonomie*. Olomouc: Univerzita Palackého v Olomouci.
- KLAPKA, P., & kol. (2014). A multistage agglomerative approach for defining functional regions of the Czech republic: the use of 2001 commuting data. *Moravian Geography Reports*, 4, stránky 2–14.
- KLAPKA, P., & kol. (2016). The efficiency of areal units in spatial analysis: Assessing the performance of functional and administrative regions. *Moravian Geography Reports*, 2, stránky 47–59.
- KLAPKA, P., & TONEV, P. (2008). Regiony a regionalizace. V V. TOUŠEK, J. KUNC, & J. VYSTOUPIL, *Ekonomická geografie* (stránky 288–311). Plzeň: Aleš Čeněk.
- KOLEJKA, J., ROMPORTL, D., & LIPSKÝ, Z. (2009). Typy současné krajiny. V T. HRNČIAROVÁ, *Atlas krajiny České republiky* (stránky 194–195). Praha: Ministerstvo životního prostředí ČR.
- KORČÁK, J. (1934). Regionální členění Československa. (J. KRÁL, Editor) *Statistický obzor*, 9–10, stránky 416–434.
- KOŘISTKA, K. (1872). Přirozené krajiny Čech, jejich půda a podnebí. *Přispěvky ku statistice lesů v Čechách*, 9–37.
- LAUKO, V., & KASALA, K. (2009). *Teória a metodológia regionálnej geografie*. Bratislava: Kartprint.
- MIČIAN, L. (2003). Vybrané partie z fyzickogeografickej regionalizácie. Načteno z Univerzita Komenského Bratislavě:
https://fns.uniba.sk/fileadmin/prif/geog/kfg/Studium/predmety_1._stupen/geogregasyst_mician2003_fgregionalizacia_texty.pdf
- Ministerstvo pro místní rozvoj ČR. (2020). *Strategie regionálního rozvoje ČR 2021+*. Načteno z Ministerstvo pro místní rozvoj ČR: <https://mmr.cz/getmedia/58c57a22-202d-4374-af5d-cbd8f9454adb/SRR21.pdf.aspx?ext=.pdf>
- RICHLING, A. (1992). *Kompleksowa geografia fizyczna*. Warszawa: Wydawnictwo naukowe PWN.
- ŘEHÁK, S. (1987). Dojíždka do zaměstnání. *Atlas Obyvatelstva ČSSR*.
- SPENCE, N. A., & TAYLOR, P. J. (1970). Quantitative Methods in Regional Taxonomy. V *Progress in Geography 2* (stránky 1–64).
- StatSoft Inc. (2013). STATISTICA 12. StatSoft Inc.
- StatSoft Inc. (2013). STATISTICA 12 User Manual | Joining (tree clustering). *STATISTICA 12 User Manual*. StatSoft.
- TONEV, P. (2013). *Změny v dojíždce za prací v období transformace: komparace lokálních trhů práce*. Brno: Geografický ústav Masarykovy Univerzity.
- University of Ljubljana. (29. 3 2016). *Orange Data Mining | k-Means*. Získáno 26. 3 2022, z Orange Data Mining | Data Mining: <https://orangedatamining.com/widget-catalog/unsupervised/kmeans/>

- University of Ljubljana. (29. 3 2016). *Orange Data Mining | PCA*. Získáno 26. 3 2022, z Orange Data Mining | Data Mining: <https://orangedatamining.com/widget-catalog/unsupervised/PCA/>
- University of Ljubljana. (29. 3 2016). *Orange Data Mining | Preprocess*. Získáno 26. 3 2022, z Orange Data Mining | Data Mining: <https://orangedatamining.com/widget-catalog/transform/preprocess/>
- University of Ljubljana. (29. 03 2016). *Orange Data Mining | Silhouette Plot*. Získáno 26. 03 2022, z Orange Data Mining | Data Mining: <https://orangedatamining.com/widget-catalog/visualize/silhouetteplot/>
- WRÓBEL, A. (1965). *Pojecie regionu ekonomicznego i teoria geografii*. Varšava: Państwowe wydawnictwo naukowe.

Seznam tabulek

| | |
|--|----|
| Tab. 1: Rozsah regionálních tříd výsledků zvolených hierarchických-aglomerativních metod (počet shluků) | 46 |
| Tab. 2: Rozsah regionálních tříd výsledků zvolených hierarchických-divizivních metod (počet shluků) | 54 |
| Tab. 3: Rozsah regionálních tříd výsledků zvolených nehierarchických metod (počet shluků) | 56 |
| Tab. 4: Hodnocení validity hierarchických-aglomerativních regionálních systémů metodou siluety | 59 |
| Tab. 5: Hodnocení validity hierarchických-aglomerativních regionálních systémů analýzou kofenetické matice | 60 |
| Tab. 6: Hodnocení validity hierarchických-divizivních regionálních systémů metodou siluety | 60 |
| Tab. 7: Hodnocení validity nehierarchických regionálních systémů metodou siluety | 61 |

Seznam obrázků

| | |
|---|----|
| Obr. 1: Vztah geografického měřítka, použitých metod a celkové rozmanitosti regionů..... | 17 |
| Obr. 2: Formální region tvořený prostorově souvislými regionálními třídami..... | 18 |
| Obr. 3: Formální region tvořený prostorově nesouvislými regionálními třídami..... | 19 |
| Obr. 4: Problematika hranic formálních regionů..... | 19 |
| Obr. 5: Schéma měření taxonomické vzdálenosti v n-rozměrném prostoru..... | 28 |
| Obr. 6: Obecný vzorec výpočtu taxonomické vzdálenosti..... | 29 |
| Obr. 7: Extrakce hlavních komponent využitím Scree-slope metody a Kaiserova kritéria..... | 44 |
| Obr. 8: Metoda jednoduchého spojení (manhattanská vzdálenost, 50% řez grafu)..... | 46 |
| Obr. 9: Metoda jednoduchého spojení (manhattanská vzdálenost, 50% řez grafu)..... | 47 |
| Obr. 10: Metoda úplného spojení (euklidovská vzdálenost, 50% řez grafu)..... | 48 |
| Obr. 11: Metoda úplného spojení (euklidovská vzdálenost, 50% řez grafu)..... | 48 |
| Obr. 12: Nevážená průměrová metoda (euklidovská vzdálenost, 50% řez grafu)..... | 49 |
| Obr. 13: Nevážená průměrová metoda (euklidovská vzdálenost, 50% řez grafu)..... | 50 |
| Obr. 14: Vážená průměrová metoda (manhattanská vzdálenost, 50% řez grafu)..... | 50 |
| Obr. 15: Vážená průměrová metoda (manhattanská vzdálenost, 50% řez grafu)..... | 51 |
| Obr. 16: Wardova metoda (manhattanská vzdálenost, 50% řez grafu)..... | 52 |
| Obr. 17: Wardova metoda (manhattanská vzdálenost, 50% řez grafu)..... | 52 |
| Obr. 18: Algoritmus SKATER (odseknutá Delaunayova triangulace, 2 regionální třídy)..... | 55 |
| Obr. 19: Metoda k-průměrů (náhodné lokality vzorců, 4 regionální třídy)..... | 56 |
| Obr. 20: Metoda k-průměrů (optimalizované lokality vzorců, 2 regionální třídy)..... | 57 |
| Obr. 21: Metoda k-průměrů (náhodné lokality vzorců, 4 regionální třídy)..... | 58 |

Seznam příloh

| | |
|--|----|
| Příloha 1: Extrahované hlavní komponenty | 71 |
| Příloha 2: Optimální počet regionální tříd v testovací sadě (K-Means++, metoda siluety)..... | 71 |
| Příloha 3: Metoda jednoduchého spojení (euklidovská vzdálenost, 50% řez grafu) | 72 |
| Příloha 4: Metoda jednoduchého spojení (euklidovská vzdálenost, 25% řez grafu) | 72 |
| Příloha 5: Metoda jednoduchého spojení (manhattanská vzdálenost, 25% řez grafu)..... | 72 |
| Příloha 6: Metoda jednoduchého spojení (čebyševova vzdálenost, 50% řez grafu)..... | 73 |
| Příloha 7: Metoda jednoduchého spojení (čebyševova vzdálenost, 25% řez grafu)..... | 73 |
| Příloha 8: Metoda úplného spojení (euklidovská vzdálenost, 25% řez grafu)..... | 73 |
| Příloha 9: Metoda úplného spojení (manhattanská vzdálenost, 50% řez grafu) | 74 |
| Příloha 10: Metoda úplného spojení (manhattanská vzdálenost, 25% řez grafu) | 74 |
| Příloha 11: Metoda úplného spojení (čebyševova vzdálenost, 50% řez grafu) | 74 |
| Příloha 12: Metoda úplného spojení (čebyševova vzdálenost, 25% řez grafu) | 75 |
| Příloha 13: Nevážená průměrová metoda (euklidovská vzdálenost, 25% řez grafu)..... | 75 |
| Příloha 14: Nevážená průměrová metoda (manhattanská vzdálenost, 50% řez grafu)..... | 75 |
| Příloha 15: Nevážená průměrová metoda (manhattanská vzdálenost, 25% řez grafu)..... | 76 |
| Příloha 16: Nevážená průměrová metoda (čebyševova vzdálenost, 50% řez grafu) | 76 |
| Příloha 17: Nevážená průměrová metoda (čebyševova vzdálenost, 25% řez grafu) | 76 |
| Příloha 18: Vážená průměrová metoda (euklidovská vzdálenost, 50% řez grafu) | 77 |
| Příloha 19: Vážená průměrová metoda (euklidovská vzdálenost, 25% řez grafu) | 77 |
| Příloha 20: Vážená průměrová metoda (manhattanská vzdálenost, 25% řez grafu)..... | 77 |
| Příloha 21: Vážená průměrová metoda (čebyševova vzdálenost, 50% řez grafu)..... | 78 |
| Příloha 22: Vážená průměrová metoda (čebyševova vzdálenost, 25% řez grafu)..... | 78 |
| Příloha 23: Wardova metoda (euklidovská vzdálenost, 50% řez grafu) | 78 |
| Příloha 24: Wardova metoda (euklidovská vzdálenost, 25% řez grafu) | 79 |
| Příloha 25: Wardova metoda (manhattanská vzdálenost, 25% řez grafu) | 79 |
| Příloha 26: Wardova metoda (čebyševova vzdálenost, 50% řez grafu) | 79 |
| Příloha 27: Wardova metoda (čebyševova vzdálenost, 25% řez grafu) | 80 |

Přílohy

Příloha 1: Extrahované hlavní komponenty

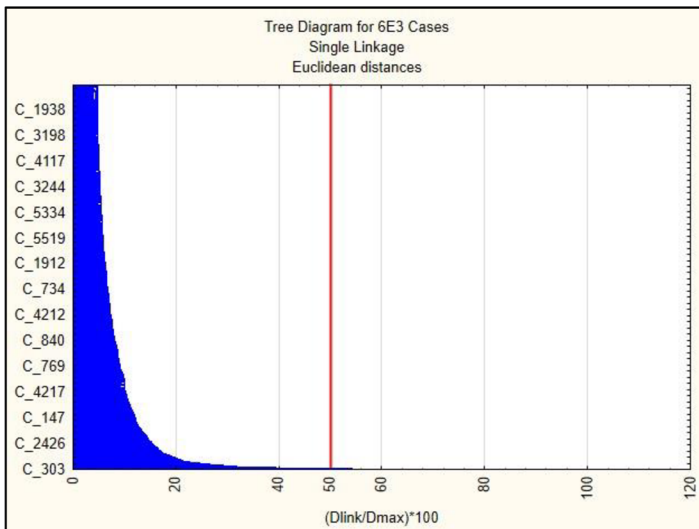
| Číslo hlavní komponenty | Komponentní vysvětlená variabilita | Kumulativní vysvětlená variabilita (%) | Pozn. |
|-------------------------|------------------------------------|--|---------------------|
| PC1 | 0,150 | 15,05 | |
| PC2 | 0,102 | 25,24 | |
| PC3 | 0,046 | 29,81 | Metoda scree-slope |
| PC4 | 0,029 | 32,76 | |
| PC5 | 0,023 | 35,03 | |
| PC6 | 0,019 | 36,91 | |
| PC7 | 0,018 | 38,66 | |
| PC8 | 0,015 | 40,14 | |
| PC9 | 0,014 | 41,59 | |
| PC10 | 0,013 | 42,94 | |
| PC11 | 0,012 | 44,17 | |
| PC12 | 0,012 | 45,38 | |
| PC13 | 0,011 | 46,49 | |
| PC14 | 0,010 | 47,52 | Kaiserovo kritérium |
| PC15 | 0,0099 | 48,51 | |
| PC16 | 0,0096 | 49,47 | |
| PC17 | 0,009 | 50,38 | |
| PC18 | 0,009 | 51,25 | |
| PC19 | 0,008 | 52,09 | |
| PC20 | 0,008 | 52,91 | |

Zdroj: vlastní zpracování v MS Excel a Orange Data Mining; Data: ČSÚ

Příloha 2: Optimální počet regionální tříd v testovací sadě (K-Means++, metoda siluety)

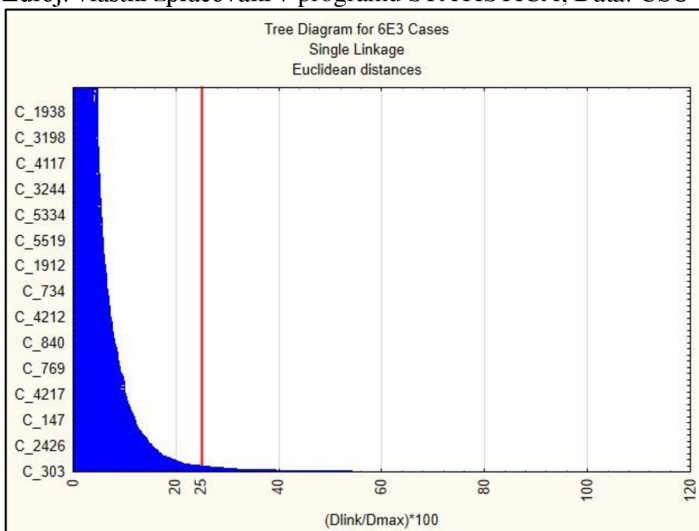
| Počet tříd | Metoda Scree-slope (3 hlavní komponenty) | | Kaiserovo kritérium (14 hlavních komponent) | |
|------------|---|-------------------------|--|-------------------------|
| | Euklidovská vzdálenost | Manhattanská vzdálenost | Euklidovská vzdálenost | Manhattanská vzdálenost |
| 2 | 0,487 | 0,460 | 0,062 | 0,066 |
| 3 | 0,341 | 0,334 | 0,066 | 0,068 |
| 4 | 0,320 | 0,309 | 0,064 | 0,064 |
| 5 | 0,319 | 0,306 | 0,068 | 0,066 |
| 6 | 0,275 | 0,267 | 0,074 | 0,072 |
| 7 | 0,276 | 0,267 | 0,067 | 0,064 |
| 8 | 0,269 | 0,256 | 0,072 | 0,069 |
| 9 | 0,272 | 0,259 | 0,067 | 0,063 |
| 10 | 0,261 | 0,250 | 0,066 | 0,063 |

Zdroj: vlastní zpracování v MS Excel a Orange Data Mining; Data: ČSÚ



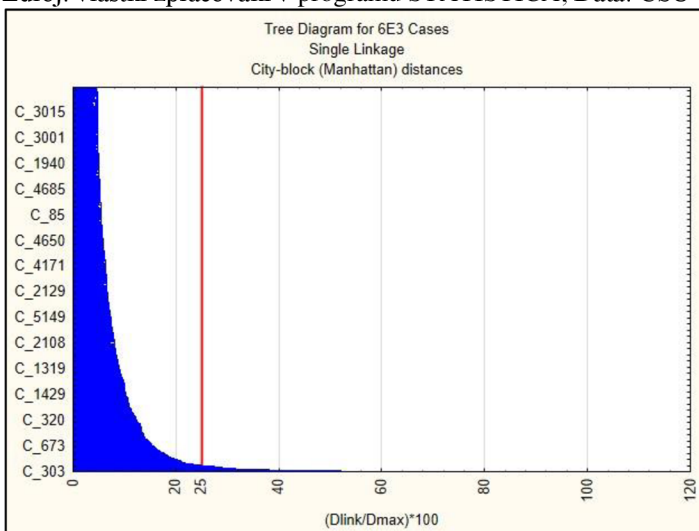
Příloha 3: Metoda jednoduchého spojení (euklidovská vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



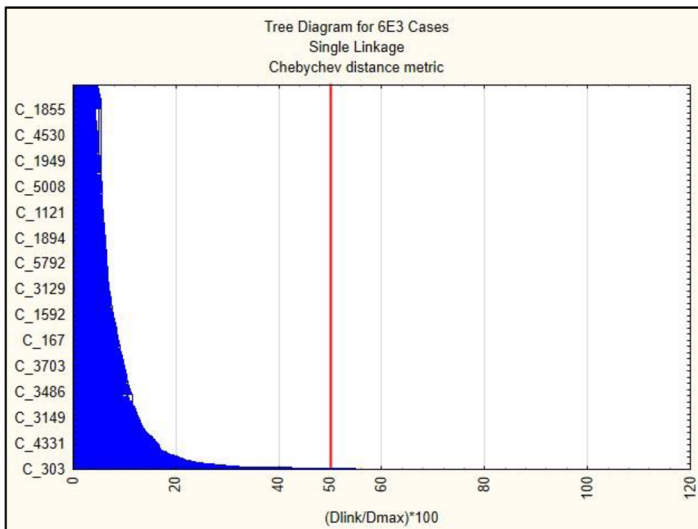
Příloha 4: Metoda jednoduchého spojení (euklidovská vzdálenost, 25% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



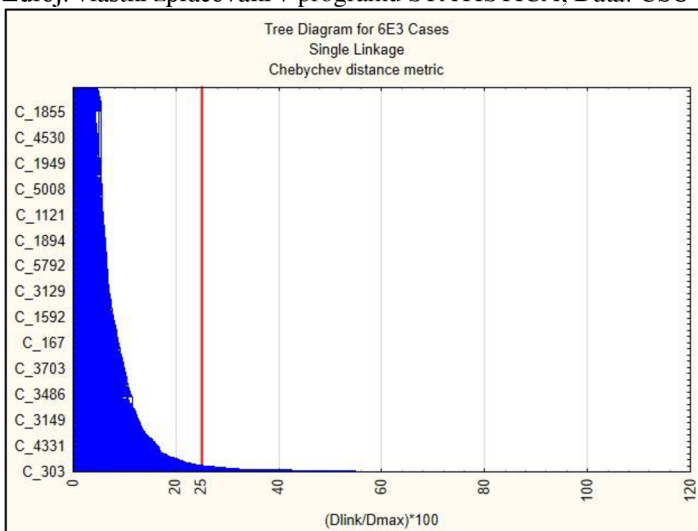
Příloha 5: Metoda jednoduchého spojení (manhattanská vzdálenost, 25% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



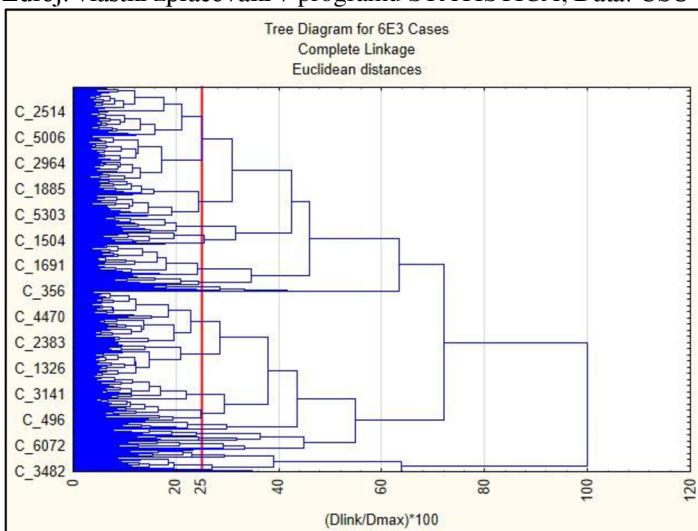
Příloha 6: Metoda jednoduchého spojení (čebyševova vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



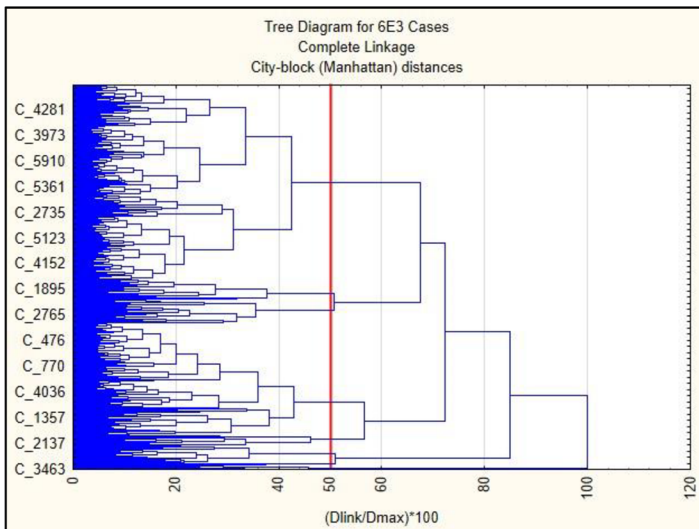
Příloha 7: Metoda jednoduchého spojení (čebyševova vzdálenost, 25% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



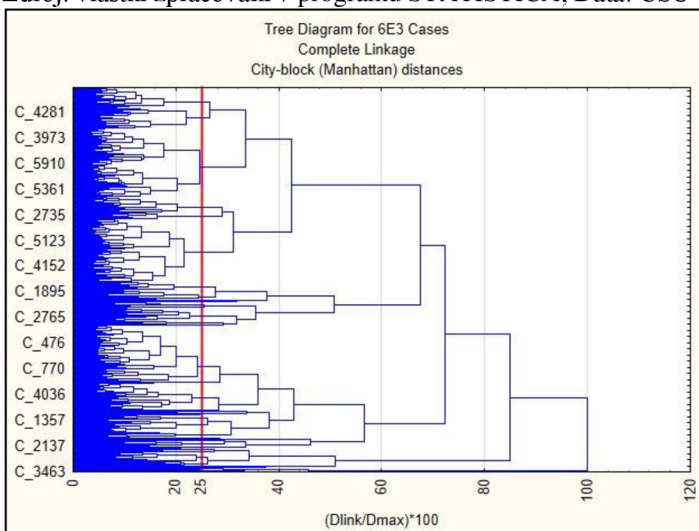
Příloha 8: Metoda úplného spojení (euklidovská vzdálenost, 25% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



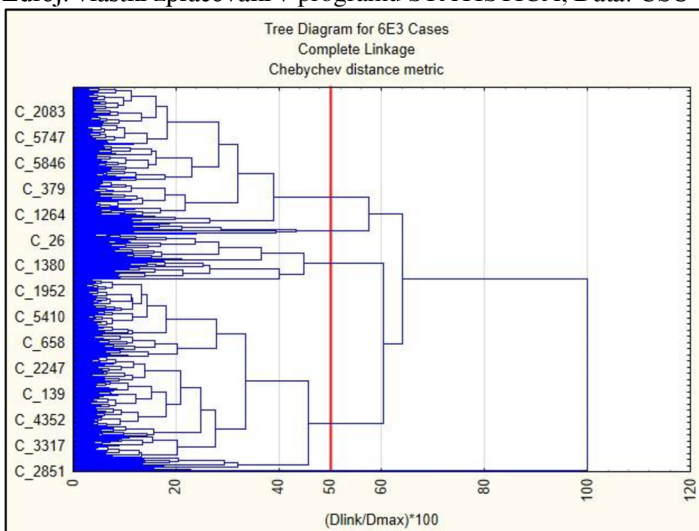
Příloha 9: Metoda úplného spojení (manhattanská vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



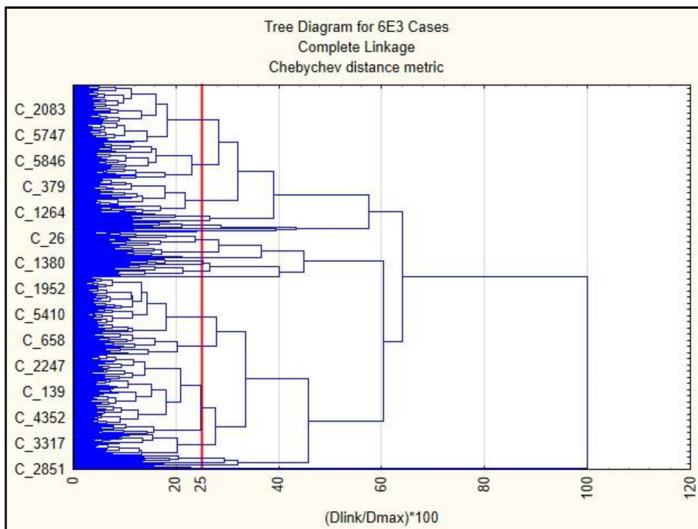
Příloha 10: Metoda úplného spojení (manhattanská vzdálenost, 25% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



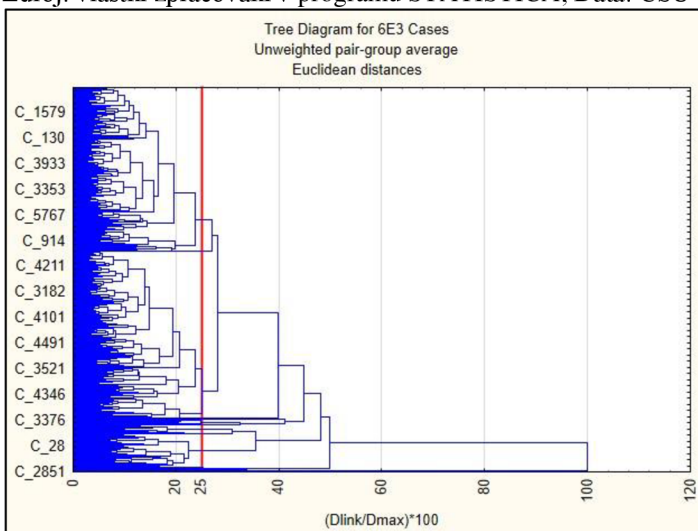
Příloha 11: Metoda úplného spojení (čebyševova vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



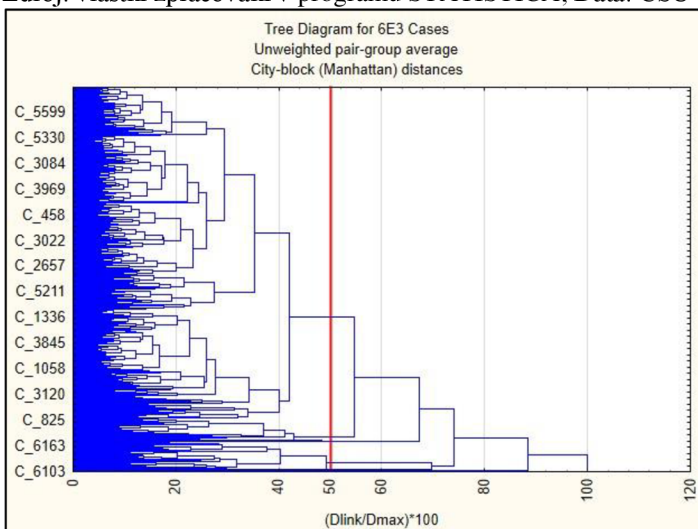
Příloha 12: Metoda úplného spojení (čebyševova vzdálenost, 25% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



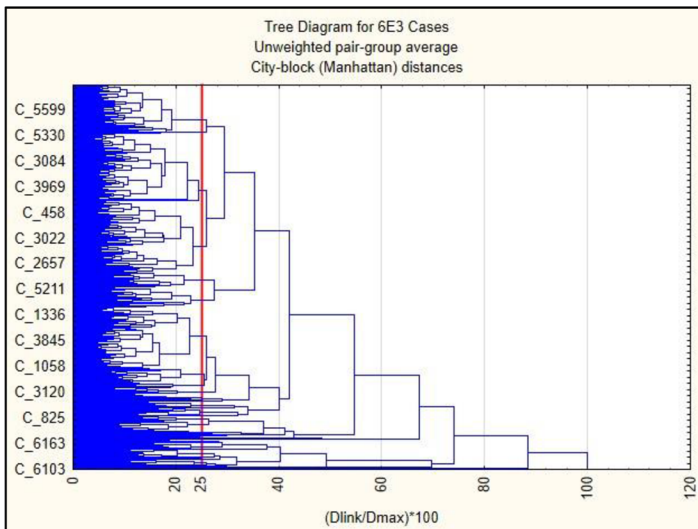
Příloha 13: Nevážená průměrová metoda (euklidovská vzdálenost, 25% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



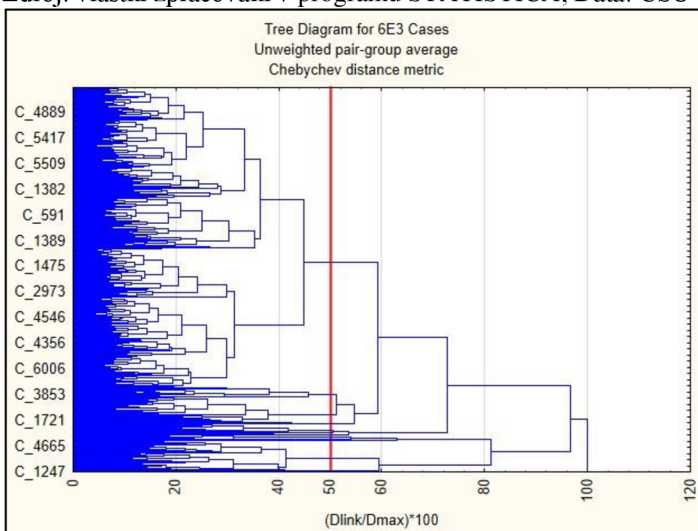
Příloha 14: Nevážená průměrová metoda (manhattanská vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



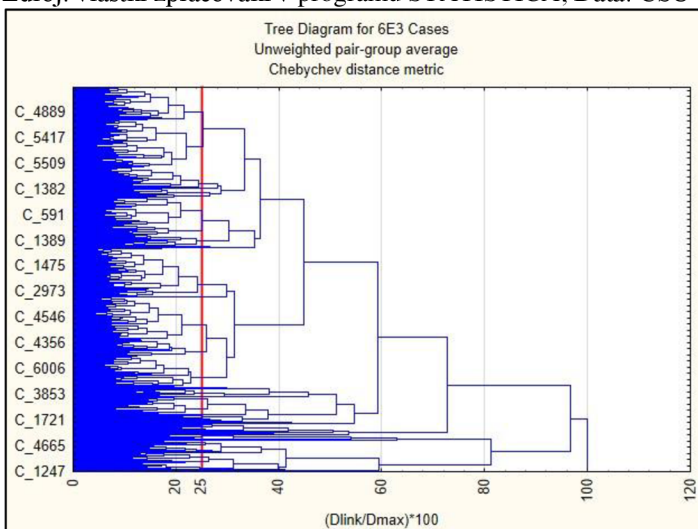
Příloha 15: Nevážená průměrová metoda (manhattanská vzdálenost, 25% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



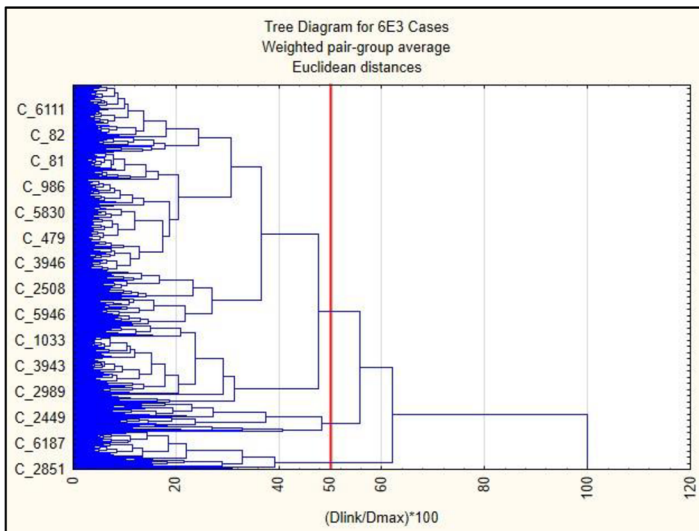
Příloha 16: Nevážená průměrová metoda (čebyševova vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



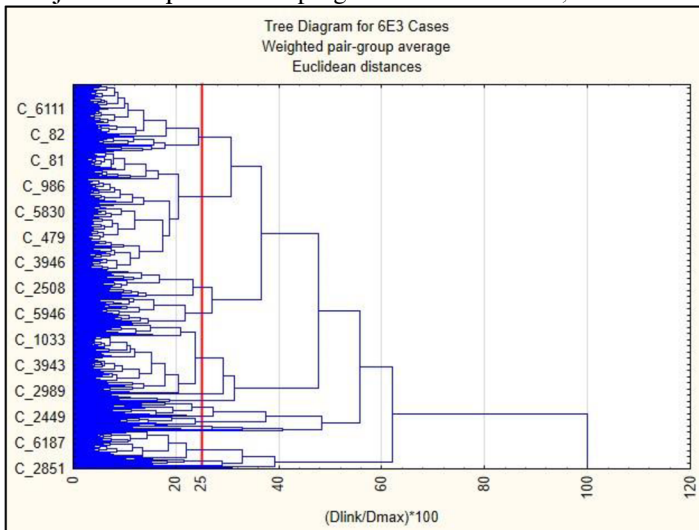
Příloha 17: Nevážená průměrová metoda (čebyševova vzdálenost, 25% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



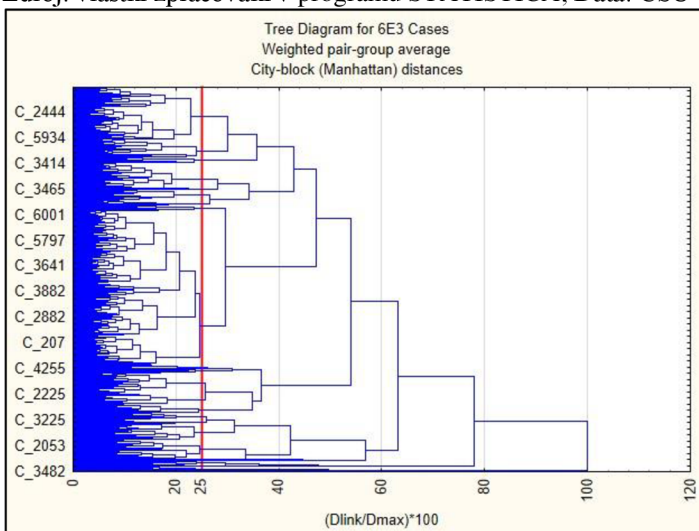
Příloha 18: Vážená průměrová metoda (euklidovská vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



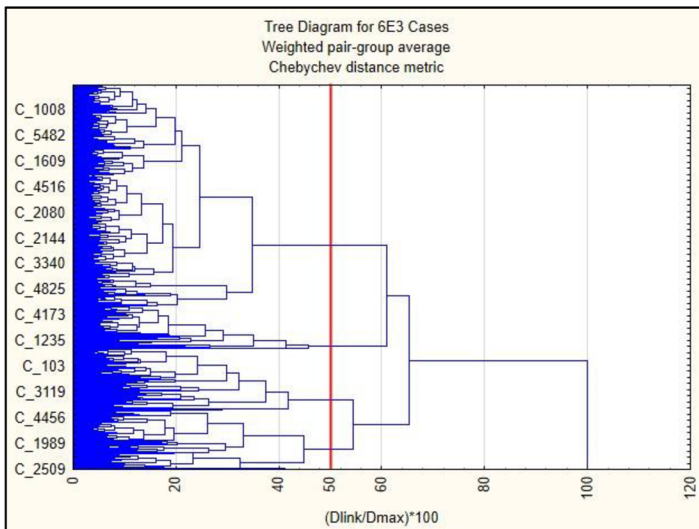
Příloha 19: Vážená průměrová metoda (euklidovská vzdálenost, 25% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



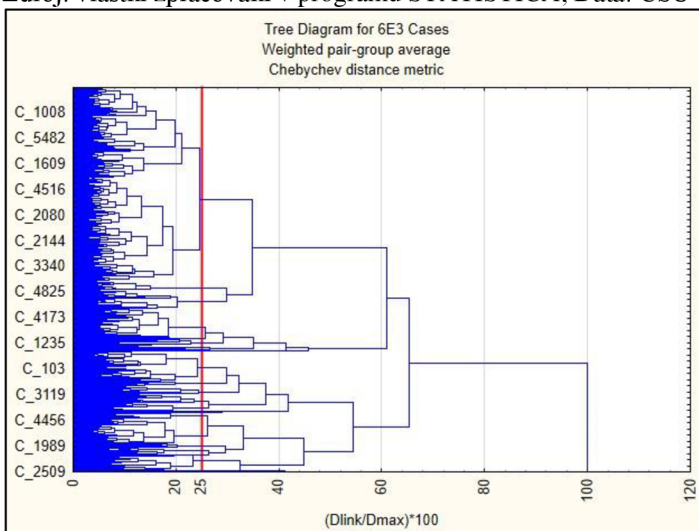
Příloha 20: Vážená průměrová metoda (manhattanská vzdálenost, 25% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



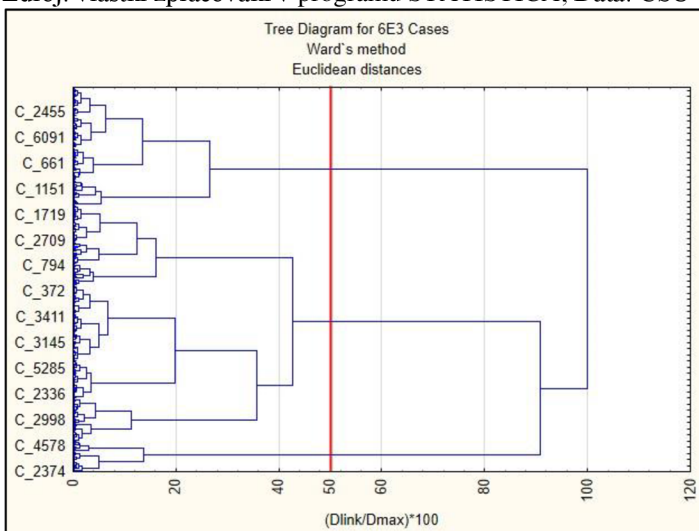
Příloha 21: Vážená průměrová metoda (čebyševova vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



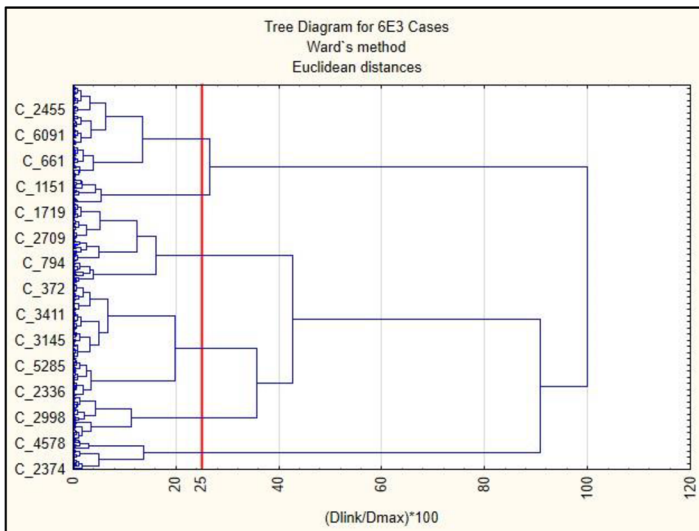
Příloha 22: Vážená průměrová metoda (čebyševova vzdálenost, 25% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



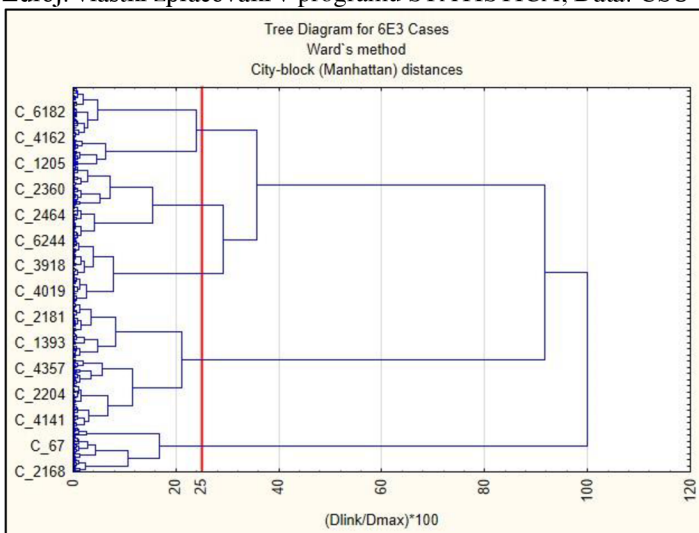
Příloha 23: Wardova metoda (euklidovská vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



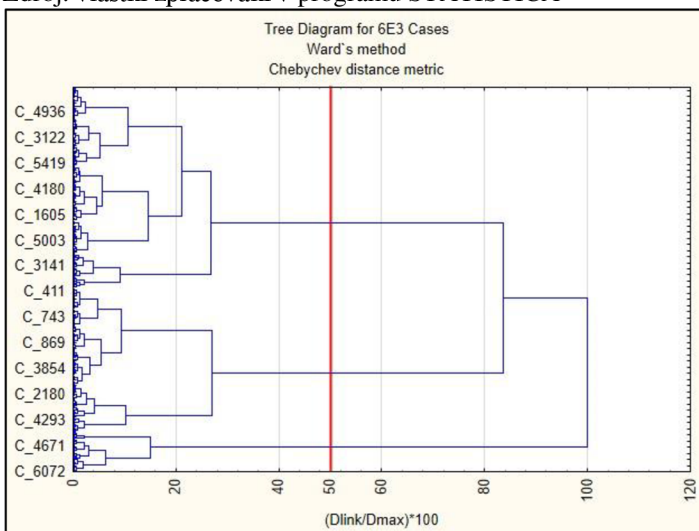
Příloha 24: Wardova metoda (euklidovská vzdálenost, 25% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



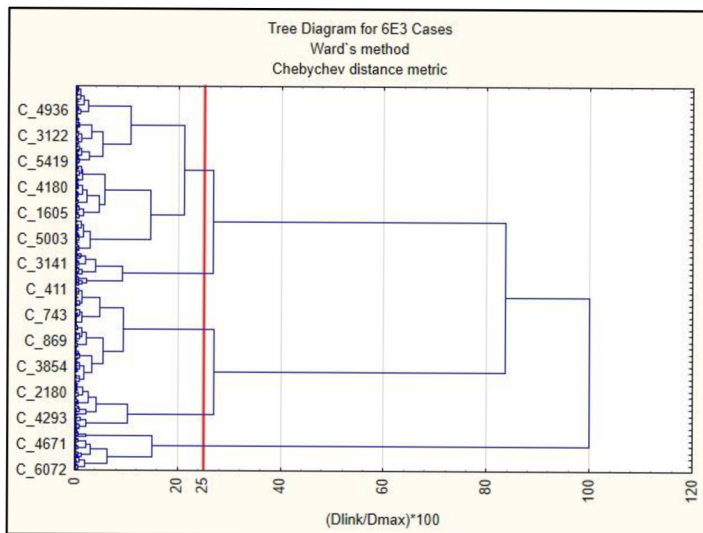
Příloha 25: Wardova metoda (manhattanská vzdálenost, 25% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA



Příloha 26: Wardova metoda (čebyševova vzdálenost, 50% řez grafu)

Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ



Příloha 27: Wardova metoda (čebyševova vzdálenost, 25% řez grafu)
Zdroj: vlastní zpracování v programu STATISTICA; Data: ČSÚ