

**Univerzita Palackého v Olomouci**

**Přírodovědecká fakulta**

**Katedra geoinformatiky**

**SOCIOEKONOMICKÉ ANALÝZY NAD  
ROZŠÍŘENÝMI ADMINISTRATIVNÍMI DATY  
EUROSTATU**

**Diplomová práce**

**Daniel PAVLAČKA**

**Vedoucí práce Mgr. Vít PÁSZTO, Ph.D.**

**Olomouc 2020**

**Geoinformatika**

## **ANOTACE**

Diplomová práce se zabývá vytvořením typologie venkovských a městských oblastí na prostorové úrovni LAU2 Evropské unie a přidružených států. Teoretická část práce je zaměřena na různé způsoby vymezení městských a venkovských oblastí. V první fázi práce byla vytvořena obohacená datová sada na základě dat Eurostatu o počtu obyvatel pro celé zájmové území. Pro obohacení byla využita data z otevřených zdrojů. Z původních dat byly vytvořeny jednoduché klasifikace město/venkov pouze pomocí atributů počtu obyvatel a hustoty zalidnění. Hlavní část práce se zabývá vytvořením typologie městských a venkovských oblastí pomocí analýzy hlavních komponent a následného shlukování. Důležitou součástí je analýza a interpretace využitých hlavních komponent a výsledných shluků, které představují jednotlivé typy městských a venkovských oblastí. Tyto shluky jsou dále agregovány na prostorové jednotky NUTS2 a srovnávány s dalšími daty Eurostatu dostupnými za tytéž prostorové jednotky. Pro prezentaci hlavních výsledných vrstev byla vytvořena webová mapová aplikace.

## **KLÍČOVÁ SLOVA**

Městská oblast; venkovská oblast; analýza hlavních komponent; shluková analýza

Počet stran práce: 73

Počet příloh: 4 (z toho 2 volné a 2 vázané)

## **ANOTATION**

The diploma thesis deals with typology creation of rural and urban areas at the LAU2 spatial level in the European Union and its associate countries. The theoretical part of the thesis is focused on the description of different methods of delimiting urban and rural areas. In the first phase of the work, an enriched dataset based on population size data in the regions of interest (provided by Eurostat) was created. Data, used for the enrichment, were obtained from open data sources. Simple urban/rural classifications are derived from the original data using only population size and population density attributes. The main part of the thesis is concerned with typology creation of rural and urban regions using Principal component analysis and Cluster analysis. The Important part lies in analysis and interpretation of the principal components and the resulting clusters, which represent individual types of urban and rural regions. Furthermore, the clusters are aggregated into NUTS2 spatial units and then compared to other Eurostat data, corresponding to the same spatial units. Finally, in order to present the main resulting layers, an online map application is presented.

## **KEYWORDS**

urban area; rural area; Principal component analysis; Cluster analysis

Number of pages: 73

Number of appendixes: 4

Dílčí části bakalářské práce byly realizovány v rámci projektu podpořeného Grantovou agenturou České republiky (GAČR) č. 18-05432S „*Prostorová syntéza založená na pokročilých metodách geocomputation*“.

**Prohlašuji, že**

- diplomovou práci včetně příloh, jsem vypracoval samostatně a uvedl jsem všechny použité podklady a literaturu.

- jsem si vědom, že na moji diplomovou práci se plně vztahuje zákon č.121/2000 Sb. - autorský zákon, zejména § 35 – využití díla v rámci občanských a náboženských obřadů, v rámci školních představení a využití díla školního a § 60 – školní dílo,

- beru na vědomí, že Univerzita Palackého v Olomouci (dále UP Olomouc) má právo nevýdělečně, ke své vnitřní potřebě, diplomovou práci užívat (§ 35 odst. 3),

- souhlasím, aby jeden výtisk diplomové práce byl uložen v Knihovně UP k prezenčnímu nahlédnutí,

- souhlasím, že údaje o mé diplomové práci budou zveřejněny ve Studijním informačním systému UP,

- v případě zájmu UP Olomouc uzavřu licenční smlouvu s oprávněním užít výsledky a výstupy mé diplomové práce v rozsahu § 12 odst. 4 autorského zákona,

- použít výsledky a výstupy mé diplomové práce nebo poskytnout licenci k jejímu využití mohu jen se souhlasem UP Olomouc, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly UP Olomouc na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Olomouci dne

Bc. Daniel Pavlačka

UNIVERZITA PALACKÉHO V OLMOUCI  
Přírodovědecká fakulta  
Akademický rok: 2018/2019

## ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Daniel PAVLAČKA**  
Osobní číslo: **R18865**  
Studijní program: **N1301 Geografie**  
Studijní obor: **Geoinformatika**  
Název tématu: **Socioekonomické analýzy nad rozšířenými administrativními daty Eurostatu**  
Zadávající katedra: **Katedra geoinformatiky**

### Z á s a d y p r o v y p r a c o v á n í :

Cílem práce je obohatit data administrativních jednotek Eurostatu nomenklatury NUTS o informace pocházející z volně dostupných zdrojů geografických dat a následně nad těmito daty provést vybrané socioekonomické (prostorové) analýzy. V první části práce student vytvoří ucelenou prostorovou datovou sadu, ve které bude k administrativním jednotkám NUTS se zaměřením na LAU2 pomocí GIS operací přidány dodatečné informace do atributové tabulky tak, aby mohla být takto kompaktní datová sada využita pro druhou část práce. V druhé části se student provede vybrané socioekonomické analýzy, kdy se student primárně zaměří na problematiku typologie venkovských/městských oblastí. V případě potřeby vyplývající ze zvolené socioekonomické úlohy student zpětně doplní databázi o další (ne)prostorová data. Geograficky bude rozsah práce zaměřen na celou Evropu (státy poskytující data Eurostatu). Student využije poznatků jak ze zahraniční, tak i z české literatury. Student vyplní údaje o všech datových sadách, které vytvořil nebo získal v rámci práce, do Metainformačního systému Katedry geoinformatiky a současně vytvoří zálohu údajů ve formě validovaného XML souboru. Celá práce (text, přílohy, výstupy, zdrojová a vytvořená data, XML soubor) bude odevzdána v digitální podobě na CD (DVD) a text práce s vybranými přílohami bude odevzdán ve dvou svázaných výtiscích na sekretariát katedry. O diplomové práci student vytvoří webovou stránku v souladu s pravidly dostupnými na stránkách katedry. Práce bude zpracována podle zásad psaní diplomových prací a závazné šablony pro diplomové práce na KGI. Povinnou přílohou práce bude poster formátu A2.

Rozsah grafických prací: dle potřeby  
Rozsah pracovní zprávy: max. 50 stran  
Forma zpracování diplomové práce: tištěná  
Seznam odborné literatury:

Burian, J., Pászto, V., Tuček, P. a kol. (2013): Geoinformatika při analýzách rurálního a urbánního prostoru . Univerzita Palackého v Olomouci, 120s.  
Pászto, V., Burian, J., Marek, L., Voženilek, V., Tuček, P. (2016): Fuzzy přístup při určování příslušnosti obcí do venkovského a městského prostoru. Geografie, 121(1), s. 156-186.  
Dijkstra L, Poelman H. (2014): A harmonised definition of cities and rural areas: the new degree of urbanisation. European Comission Directorate-General for Regional and Urban Policy: Working Paper.  
Eurostat (2018): Eurostat regional yearbook 2018, Mariana Kotzeva Eds. ? European Union, Luxembourg, ISBN 978-92-79-87878-7  
[https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Main\\_Page](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Main_Page)  
A další geografická, geoinformatická, či jiná literatura a softwarová dokumentace potřebná pro provedení práce.

Vedoucí diplomové práce: Mgr. Vít Pászto, Ph.D.  
Katedra geoinformatiky

Datum zadání diplomové práce: 1. listopadu 2018  
Termín odevzdání diplomové práce: 5. května 2020

doc. RNDr. Martin Kubala, Ph.D.  
děkan

L.S.

UNIVERZITA PALACKÉHO V OLOMOUCI  
PŘÍRODOVĚDECKÁ FAKULTA  
KATEDRA GEOINFORMATIKY  
17. listopadu 58, 771 46 Olomouc

prof. RNDr. Vít Voženilek, CSc.  
vedoucí katedry

V Olomouci dne 10. prosince 2018

# OBSAH

<b>SEZNAM POUŽITÝCH ZKRATEK .....</b>	<b>8</b>
<b>ÚVOD .....</b>	<b>9</b>
<b>1 CÍLE PRÁCE.....</b>	<b>10</b>
<b>2 METODY A POSTUPY ZPRACOVÁNÍ.....</b>	<b>11</b>
<b>3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY .....</b>	<b>13</b>
3.1 Definování městských a venkovských oblastí.....	13
3.2 Datové zdroje.....	17
<b>4 OBOHACENÍ DATOVÉ SADY .....</b>	<b>19</b>
4.1 Původní data .....	19
4.2 CORINE Land Cover (CLC) .....	20
4.3 Open Street Map (OSM).....	20
<b>5 KLASIFIKACE OBCÍ NA VENKOVSKÉ A MĚSTSKÉ.....</b>	<b>23</b>
<b>6 STATISTICKÉ HODNOCENÍ DAT.....</b>	<b>28</b>
6.1 Statistické zhodnocení vstupních dat pomocí boxplotu .....	28
6.2 Statistické zhodnocení vstupních dat pomocí korelací.....	30
<b>7 ANALÝZA HLAVNÍCH KOMPONENT .....</b>	<b>35</b>
7.1 Interpretace vybraných hlavních komponent .....	41
<b>8 TYPOLOGIE JEDNOTEK LAU2 PODLE SHLUKOVÉ ANALÝZY .....</b>	<b>43</b>
8.1 Interpretace shluků .....	44
<b>9 AGREGACE NA NUTS2 .....</b>	<b>53</b>
9.1 Proces agregace .....	53
9.2 Srovnání s daty Eurostatu .....	58
<b>10 WEBOVÁ MAPOVÁ APLIKACE.....</b>	<b>61</b>
<b>11 VÝSLEDKY .....</b>	<b>62</b>
<b>12 DISKUZE.....</b>	<b>64</b>
<b>ZÁVĚR .....</b>	<b>65</b>
<b>POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE</b>	
<b>PŘÍLOHY</b>	

## SEZNAM POUŽITÝCH ZKRATEK

<b>Zkratka</b>	<b>Význam</b>
CLC	CORINE Land Cover
OSM	OpenStreetMap
OECD	Organisation for Economic Co-operation and Development
TL	Territorial Level
LAU	Local administrative unit
NUTS	Nomenclature des Unites Territoriales Statistiques
PCA	Principal component analysis
PAM	Partitioning Around Medoids
CLARA	Clustering Large Application
DG Agri	Directorate-General of the European Commission
DEGURBA	Degree of Urbanisation
DEFRA	Department for Environment, Food and Rural Affairs
HUZ	Hromadná ubytovací zařízení
NACE	Nomenclature statistique des activités économiques dans la Communauté européenne
IDE	Integrated Development Environment
CHKO	Chráněná krajinná oblast
ORP	Obec s rozšířenou působností
POU	Obec s pověřeným obecním úřadem
GISCO	Geographic Information System of the Commission



## ÚVOD

Vymezení, popis i klasifikace města a venkova jsou nedílnou součástí humánní geografie, ke kterému každý autor přistupuje jiným způsobem. Existuje velké množství různých přístupů jak tyto oblasti vymežit, některé uzpůsobené konkrétním specifickým rysům členění států, pro které jsou pak tyto přístupy platné, jiné zaměřené a používané globálně nebo pro uskupení více států. Tyto přístupy se liší použitou prostorovou jednotkou i indikátory, pomocí kterých jsou oblasti členěny. V praxi jsou typologie s přesným vymezením venkovských a městských oblastí důležité pro vhodné zacílení různých rozvojových dotačních programů a rozdělení státní podpory.

Tato diplomová práce se v první části zabývá vytvořením datové sady s rozlišením na LAU2 za Evropskou Unii a přidružené státy. Základem datové sady jsou data statistického úřadu Evropské unie (Eurostat) o počtu obyvatel. Další indikátory jsou získány z volně dostupných datových zdrojů. Druhá část práce je zaměřena na vytvoření typologie městských a venkovských oblastí na základě vytvořené datové sady, platné pro celou oblast v detailu jednotek LAU2.

# 1 CÍLE PRÁCE

Cílem diplomové práce je obohatit data administrativních jednotek Eurostatu nomenklatury NUTS o informace pocházející z volně dostupných zdrojů geografických dat a následně nad těmito daty provést vybrané socioekonomické (prostorové) analýzy. V první části práce bude vytvořena ucelená prostorová datová sada, ve které budou k administrativním jednotkám NUTS se zaměřením na LAU2 pomocí GIS operací přidány dodatečné informace do atributové tabulky tak, aby mohla být tato kompaktní datová sada využita pro druhou část práce. V druhé části práce budou provedeny vybrané socioekonomické analýzy, primárně zaměřené na problematiku a typologii venkovských/městských oblastí. Geografický rozsah práce je vymezen vstupní poskytnutou vrstvou (státy, které Eurostatu poskytly data o počtu obyvatel na úrovni LAU2).

## 2 METODY A POSTUPY ZPRACOVÁNÍ

Pro stanovení vhodných metod a datových zdrojů byla nejprve provedena literární rešerše zdrojů zabývajících se vymezením městských a venkovských oblastí a jejich typologií. Vzhledem k tomu, že práce je zaměřena na rozsáhlou oblast a podrobné statistické jednotky (angl. Local Administrative Units - LAU2) bylo nutné využít data, která většinou nejsou používána k podobným analýzám. Samotná původní vrstva ve stavu, v jakém byla poskytnuta pro potřeby práce je jedinečná a vznikla na Katedře geoinformatiky Univerzity Palackého v Olomouci.

### Použité metody

Pro obohacení datové sady a agregaci výsledků do vyšších prostorových jednotek byly využity základní GIS operace jako intersect, join, summarize, atp. Funkce a využití těchto nástrojů jsou popsány například v oficiální dokumentaci Esri produktů (ArcGIS Desktop, 2020). Následně byla data analyzována pomocí různých statistických metod jako korelace a boxplot. Pro sestavení příkazů v prostředí RStudio bylo čerpáno z oficiální dokumentace jazyka R (R Documentation, 2020) a v případě potřeby z programátorského fóra Stack Overflow (Stack Overflow, 2020). Výstupy těchto statistických metod mimo jiné posloužily jako podklad pro Analýzu hlavních komponent (PCA), kterou se zabývá např. Šarmanová (2012) a Meloun, Mílitký (2002) a jež využil pro analýzu prostorových dat také Pászto (2015) či Marek (2015). Jedná se o transformační metodu, při níž jsou původní vstupní proměnné transformovány do menšího počtu nových, skrytých proměnných, kterých je méně a mají vhodnější vlastnosti. Výsledky PCA byly shlukovány, pomocí algoritmu CLARA (Clustering Large Applications), který je modifikací algoritmu PAM (Partition Around Medoids) určenou pro rozsáhlé datové soubory. Výhodou algoritmu CLARA je to, že není citlivý na odlehle hodnoty (outliery). Kromě statistických metod byly využity také zobrazovací metody nevyžadující statistické zpracování dat, pro které byly vytvářeny mapové výstupy, které byly dále vizuálně interpretovány (Horák, 2008). Mapové výstupy byly tvořeny v souladu s metodami tematické kartografie dle Voženilek, Kaňok a kol. (2011).

### Použitá data

Jako základ pro tvorbu datové sady byla použita vrstva LAU2 obsahující geometrii a atribut počtu obyvatel, kterou poskytl vedoucí této práce. Vrstva byla obohacena o data z volně dostupných zdrojů.

Pro získání informace o pokrytí území byla využita data CORINE Land Cover (CLC) volně dostupná na webu (Copernicus Land Monitoring Service, 2019). Konkrétně byla použita rastrová vrstva s prostorovým rozlišením buňky 100x100 m z roku 2018.

Mezi další využitá data patří databáze OpenStreetMap (OSM). Data byla stažena z webu GEOFABRIK ve formátu shapefile. Konkrétně byly využity vrstvy budov, POI (points of interest), silnic, železnic, vodních toků a kostelů. Data byla stažena a jsou aktuální k datu 1.4. 2019.

Pro agregaci na NUTS2 byla využita vrstva těchto statistických jednotek dostupná na webu Eurostatu, která byla obohacena rovněž z Eurostatu o další statistická data.

## Použité programy

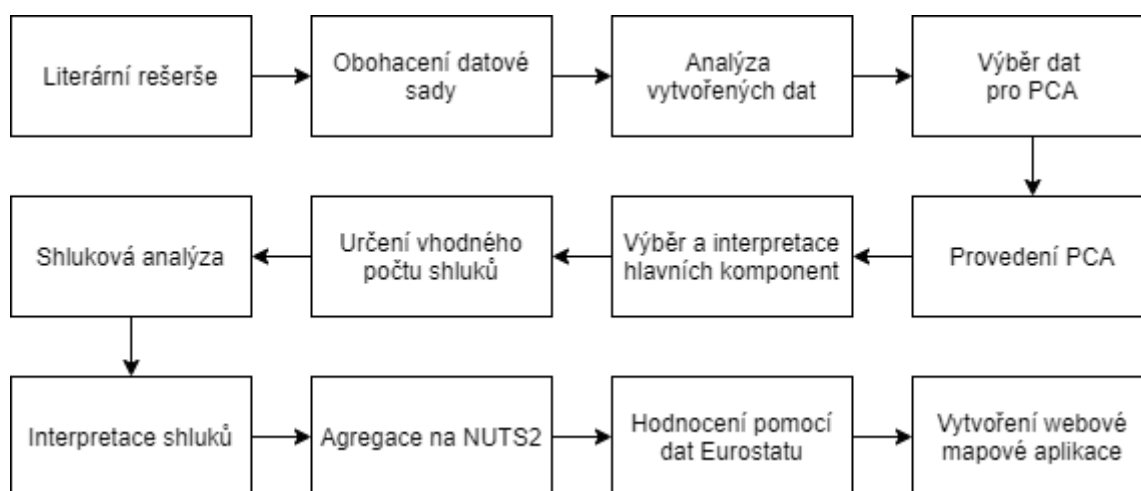
Pro zpracování a analýzu dat byl využit program ArcMap 10.x z ArcGIS Desktop od společnosti Esri. Pro statistické analýzy dat a tvorbu grafů byl využit program RStudio, IDE (Integrated Development Environment) programovacího jazyka R a knihovny rozšiřující základní funkce jazyka (cluster, corrplot, psych, factoextra a další). Některé grafické výstupy RStudia bylo nutné upravit, v takovém případě byl využit grafický vektorový editor Adobe Illustrator. Pro tvorbu výsledné webové mapové aplikace byla využita opensource JavaScriptová mapová knihovna Leaflet a některé další plugíny pro Leaflet (leaflet-sidebar-v2, leaflet-ajax) a javascriptové knihovny (jQuery). Zdrojový kód mapové aplikace byl vytvořen v IDE Visual Studio Code. Pro tvorbu doplňkových diagramů a schémat byl využit webový editor draw.io.

## Postup zpracování

Po zpracování literární rešerše byla vytvořena rozšířená datová sada, která byla následně použita pro analýzy. Pro vizuální analýzu byly vytvořeny mapy jednotlivých typologií města/venkova, podle již zavedených metodik. Následně byly atributy celkově posuzovány pomocí boxplotů a jejich vzájemných korelací.

V další části práce byla nejprve se všemi atributy provedena analýza hlavních komponent (PCA), což vedlo k zúžení výběru použitých atributů a novému provedení PCA, pomocí které byly vytvořeny a vybrány „užitečné“ čtyři hlavní komponenty. Tyto hlavní komponenty byly následně pomocí algoritmu CLARA shlukovány do šesti shluků jednotek LAU2, které představují jednotlivé typy městských a venkovských oblastí. Vhodný počet shluků byl zjištěn pomocí opakovaného testování shlukování s různými parametry a posuzování výsledků. Shluky byly vizuálně analyzovány pomocí zobrazení jejich prvků v prostoru hlavních komponent a pomocí jejich zobrazení v mapě, což vedlo k jejich pojmenování a k jejich celkovému posouzení.

Následně byly výsledky shlukování agregovány do vyšších jednotek NUTS2 a srovnávány s dalšími daty Eurostatu. Na závěr byla pro prezentaci výsledných vrstev vytvořena webová mapová aplikace. Postup práce byl znázorněn schématem (obr. 1).



Obr. 1 Postup práce

### **3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY**

Pro rozlišení venkovských a městských oblastí je využívána celá řada přístupů, od velmi jednoduchých vycházejících z jednoho atributu (počet obyvatel, hustota zalidnění aj.) přes metodiky kombinujících několik atributů, které nepracují pouze s administrativními hranicemi, ale i s daty disgregovanými do pravidelné mřížky, po metody sofistikované využívající například fuzzy logiku nebo multikriteriální hodnocení založené na expertním hodnocení jednotlivých kritérií. Nicméně v mnoha případech jsou využívána data, která vzhledem k prostorovému rozsahu a podrobnosti zkoumaných územních jednotek (např. zaměstnanost podle sektorů, dojíždka za zaměstnáním) v rámci této práce nebylo možné využít. Proto bylo nutné zvolit data v této problematice méně zavedená.

#### **3.1 Definování městských a venkovských oblastí**

Pro další práci je nutné definovat pojmy město a venkov. Přístup k vymezení těchto pojmů se u různých autorů a v různých státech velmi liší, stejně jako samotná definice městského a venkovského prostoru. Mezi praktické důvody přesného vymezení venkovských oblastí je zacílení různých rozvojových dotačních programů a rozdělení státní podpory (Burian a kol., 2013).

Město může být definováno jako ohraničené, prostorově spojitě urbanizované území. Důležitou roli při jeho vymezení hrají další ukazatele, jako jsou populace, hustota zalidnění, trh práce (Frey a Zimmer 2001). Podle Perlína (2010) je město rozsáhlé urbanizované území poskytující veřejné, administrativní funkce.

Dax (2014) uvádí, že termín „venkovský“ se používá v mnoha kontextech, aniž by byl definován jeho význam. Stejně jako jsou rozmanité venkovské oblasti, může i termín „venkovský“ znamenat velmi odlišné věci. V některých státech dokonce neexistuje oficiální definice slova „venkov“, často bývá vymezena negativně ve smyslu ne-městský. Nejčastějšími kritérii pro jeho vymezení jsou počet obyvatel, hustota zalidnění, intenzita dojíždky, nebo podíl zaměstnaných v zemědělství. Hraniční hodnoty pak závisí na konkrétním účelu klasifikace a zájmovém území. Význam výzkumu venkova a jeho prostorové kategorizace roste s rozšiřováním programů pro rozvoj venkova. Dle Buriana a kol. (2013) se venkovem obecně nazývají odlehlá osídlení a oblasti volně nezastavěné krajiny tvořící mozaiku drobnějších sídel, zemědělských a vodních ploch, lesů, ploch místních komunikací a ostatních ploch. Mimo jiné uvedené, se venkov také vyznačuje pozitivním vztahem místních obyvatel k okolní krajině, specifickým architektonickým rázem, charakterem zástavby a životním stylem. Dle programu rozvoje venkova České republiky na období 2007–2013 (2013), je pro vymezení různých typů venkova (v podstatě různé stupně příslušnosti k venkovskému prostoru) důležitá blízkost urbanizovaného centra.

#### **Vymezení venkovských a městských oblastí v Česku**

V současné době v české republice dle zákona 128/2000 Sb. O obcích (obecní zřízení), je obec, která má alespoň 3 000 obyvatel městem, pokud tak na návrh obce stanoví předseda Poslanecké sněmovny po vyjádření vlády. Existují i další případy (dle uvedeného zákona) jak se může obec stát městem, například pokud se sloučí více sídel, z nichž alespoň jedno je městem.

Vymezením a další typizací venkova Česka se zabýval Český statistický úřad (ČSÚ, 2008). Dle národního strategického plánu rozvoje venkova (2006) se v Česku za venkovské obce považují obce s méně než 2 000 obyvateli a dále se konstatuje,

že se venkovské regiony vyznačují menším podílem obyvatel v produktivním věku a také je v nich nižší míra ekonomické aktivity. Program rozvoje venkova České republiky na období 2007–2013 (2013) definuje tři typy venkovských prostor v Česku, podle jejich polohy. Přesná metodika klasifikace obcí však není zmíněna. Typy venkova jsou definovány jako:

- Príměstský venkov – venkovské obce v rámci městských aglomerací, resp. úzce vymezených urbanizovaných území (s více než 50 000 obyvateli)
- Odlehlý venkov – zahrnuje zejména tzv. periferní území, tj. území s nepříznivými sociálně – ekonomickými charakteristikami obyvatelstva a osídlení
- Mezilehlý venkov – zbývající venkovské obce

Český statistický úřad (2008) dále nezávazně testoval osm způsobů pro vymezení venkovského prostoru. Dle publikace jsou tyto návrhy spíše příspěvkem do diskuse na téma vymezení venkovského prostoru nežli ambicí najít správné řešení této problematiky. Dle těchto variant venkovským prostorem jsou:

1. obce s méně než 2 000 obyvateli
2. obce s méně než 1 000 obyvateli a obce s 1 000 až 3 000 obyvateli, jejichž hustota zalidnění je méně než 100 obyvatel/km<sup>2</sup>
3. obce s méně než 1 000 obyvateli a obce s 1 000 až 3 000 obyvateli, jejichž hustota zalidnění je méně než 150 obyvatel/km<sup>2</sup>
4. obce, které neměly statut města k 1. 1. 2007
5. obce, které nejsou obcemi s pověřeným obecním úřadem a nemají statut hlavního města
6. obce s méně než 2 000 obyvateli mimo obce v zázemí krajských měst
7. obce s méně než 5 000 obyvateli vymezené multikriteriálně na základě čtyř ukazatelů – počet obyvatel, jestli se jedná o sídlo s pověřeným obecním úřadem nebo statutem města, počet obyvatel vůči zastavěné ploše a podíl bytů v rodinných domech
8. vyčlenění obcí přechodného typu z předchozí varianty

Rozdíly mezi jednotlivými variantami jsou poměrně velké, zatímco podle 5. varianty by do venkovských oblastí spadalo více než 3,5 mil obyvatel, podle 2. varianty by to bylo necelých 2,5 milionu.

V programu rozvoje venkova na období 2014–2020 (2014) se využívají k vymezení venkova typologie OECD (uvedená níže) a typologie město-venkov od DG Agri (The Commission's Directorate-General for Agriculture and Rural Development), která klasifikuje regiony NUTS3 na základě pravidelné mřížky (population grid).

V Konceptu rozvoje venkova (2019) je uvedeno, že definice venkova ve vztahu ke konkrétnímu státu nebo nižší územní jednotce musí zohledňovat specifika daného území na rozdíl od obecněji nastavených definic, které používají mezinárodní organizace (OECD nebo EU). Vymezení zájmového území (venkova) Konceptu rozvoje venkova (2019), vychází ze Sociogeografické regionalizace Česka (Hampl a Marada, 2015). Venkov je definován jako geografický prostor zahrnující územní obvody obcí, které nejsou středisky. Venkovem jsou tedy obce, které v rámci sídelního systému nemají významné postavení, neplní centrální funkce a nevytvářejí si vlastní zázemí většího rozsahu. Definice venkova tedy není definována na základě nějaké hodnoty (např. počtu obyvatel), ale tak aby zohledňovala funkční prostorové vztahy v území.

Perlín (2009) uvádí, že pro vymezení venkovských oblastí v Česku lze doporučit vymezení na základě hodnoty hustoty zalidnění. Jako vhodnou prostorovou jednotku pro toto vymezení udává jednotky správních obvodů pověřených obecních úřadů (POU) nebo správních obvodů úřadů obcí s rozšířenou působností (ORP). Dále tvrdí, že limitní

hodnotu 150 obyv./km<sup>2</sup> (kterou využívá klasifikace OECD) je možné vzhledem ke struktuře české sídelní soustavy snížit na 100 obyv./km<sup>2</sup>.

### **Vymezení venkovských a městských oblastí dle Organizace pro hospodářskou spolupráci a rozvoj**

Metodika klasifikace regionů Organizace pro hospodářskou spolupráci a rozvoj (OECD) na městské a venkovské je založena na hodnotě hustoty zalidnění v jednotlivých lokálních územních jednotkách (nižších než TL3–Territorial Level 3), které ve státech EU odpovídají jednotkám LAU2 a jejich následné agregaci do jednotek TL3, které v EU většinou odpovídají jednotkám NUTS3. Dle této metodiky lze za venkovské obce považovat obce s hustotou zalidnění méně než 150 obyvatel/km<sup>2</sup>. Větší regiony jsou potom děleny na výrazně městské, kde ve venkovských obcích žije méně než 15 % obyvatel, významně venkovské, kde ve venkovských obcích žije 15–50 % obyvatel a převážně venkovské, kde více než 50 % obyvatel žije ve venkovských obcích (překlad názvů kategorií dle Svobodová, Věžník, 2014). Posledním krokem této metodiky je přeřazení regionů klasifikovaných jako výrazně venkovské na převážně venkovské, pokud se v nich vyskytuje centrum s více než 200 000 obyvateli, kteří zároveň reprezentují alespoň 25 % populace daného regionu a přeřazení regionů klasifikovaných jako převážně venkovské na výrazně městské, pokud se v nich vyskytuje centrum s alespoň 500 000 obyvateli, kteří zároveň reprezentují alespoň 25 % populace regionu (OECD, 2011a). Toto vymezení je používáno spíše pro mezinárodní srovnávání (Burian a kol., 2013). Další metodikou OECD je rozšířená typologie regionů, ve které se uvažuje i blízkost regionu k významnému centru, podle cestovní časové vzdálenosti (OECD, 2011b). Díky agregaci hodnot z menších územních jednotek do větších je částečně odstraněn problém, který přináší různá velikost územních jednotek napříč jednotlivými státy.

### **Vymezení venkovských a městských oblastí dle Eurostatu**

Evropská unie používá i sofistikovanější metodiky pro vymezení venkovského/městského prostoru. K těmto metodám se řadí Rural-urban typology a Degree of Urbanisation (DEGURBA), a popisuje je například Dijkstra a Poelman, 2014. Rural-urban typology je používána pro klasifikaci jednotek NUTS3 (v Česku ekvivalentní s kraji) do třech kategorií: převážně městské regiony, smíšené regiony a převážně venkovské regiony. Rozdělení do těchto kategorií probíhá ve třech krocích. V prvním kroku je využit population grid (čtvercová síť o rozměrech buňky 1x1 km s informací o počtu obyvatel v buňce) pro vymezení venkovských oblastí a městských klastrů. Městské klastry jsou charakterizovány jako navazující buňky (i diagonálně) s hustotou zalidnění alespoň 300 obyvatel/km<sup>2</sup> a alespoň 5 000 obyvatel a venkovské oblasti jako oblasti mimo městské klastry. V druhém kroku regiony NUTS3 děleny do třech kategorií tak že v kategorii převážně venkovských regionech žije ve venkovských oblastech (zjištěných v minulém kroku) více než 50 % obyvatel, ve smíšených regionech pak 20–50 % obyvatel a v převážně městských regionech méně než 20 % obyvatel. Pro zamezení problémům plynoucím z nestejně velikosti regionů NUTS3 jsou regiony menší než 500 km<sup>2</sup> spojeny s jedním nebo více sousedních regionů. V posledním kroku jsou do kategorie smíšených regionů přeřazeny takové regiony, kde v městském centru žije více než 200 000 obyvatel, kteří jsou zároveň více než 25 % populace celého regionu a do kategorie převážně městských regionů jsou přeřazeny regiony, jejichž městské centrum obývá více než 500 000 lidí, kteří zároveň představují více než 25 % populace regionu (Eurostat, 2019). Výsledky klasifikace jsou dostupné ve formátu excelové tabulky (ID regionu a číslo kategorie) a vizualizace mapovou aplikací na webu Eurostatu.

Studie DEGURBA je využívána pro klasifikaci jednotek LAU (většinou LAU2, LAU1 pouze pokud LAU2 jsou příliš malé) do třech kategorií: Rural areas, Towns and suburbs a Cities. Klasifikace probíhá na základě kritérií geografického sousedství a minimálního prahu populace pomocí 1x1 km population gridu. Buňky gridu jsou nejdříve rozděleny do třech kategorií. Venkovské buňky jsou buňky mimo městské klastry a městská centra. Městské klastry jsou definovány stejně jako v Rural-urban typology jako navazující buňky (i diagonálně) s hustotou zalidnění alespoň 300 obyvatel/km<sup>2</sup> a s alespoň 5 000 obyvatel. Městská centra jsou potom navazující buňky (ne pouze diagonálně) s hustotou zalidnění alespoň 1 500 obyvatel/km<sup>2</sup> a alespoň 50 000 obyvatel. Jako Rural areas jsou vyhodnoceny LAU, ve kterých více než 50 % obyvatel spadá do venkovských buňek. Towns a suburbs jsou takové LAU, s alespoň 50 % obyvateli žijícími v městských klastrech a zároveň méně než 50 % obyvateli žijícími ve venkovských buňkách. Do poslední kategorie Cities spadají LAU, kde alespoň 50 % obyvatel žije v městských centrech (Dijkstra a Poelman, 2014). Klasifikovaná data jsou opět dostupná na webu Eurostatu. DEGURBA existuje ve dvou verzích původní „original Degree of Urbanisation“ byla využívána od roku 1991, nová „new Degree of Urbanisation“ zavedena byla v roce 2014 a vznikla rozšířením původní verze o využití pravidelné mřížky a změnou prahových klasifikačních hodnot (Eurostat, 2018).

### **Vymezení venkovských a městských oblastí pomocí fuzzy logiky**

Další přístupem, který lze využít pro rozlišování městských/venkovských oblastí je využití teorie fuzzy množin a logiky. Pomocí tohoto přístupu je možné shladit ostré přechody mezi dvěma sousedícími stavy jevu (v tomto případě městský a venkovský prostor) a nahradit je postupnou, plynulou změnou. Jednotlivé mezistupně lze pojmenovat neurčitými slovy jako jsou více městský nebo spíše venkovský. Výsledkem tedy není určení, jestli je daná obec venkovská nebo městská, ale jak moc spadá do jedné z těchto kategorií. Tato míra příslušnosti je vyjádřena hodnotou na stupnici 0 až 1, kde maximum a minimum znamená úplně městský nebo úplně venkovský. Tento přístup využívá např. Pászto a kol. (2016) pro určení stupně příslušnosti k venkovskému/městskému prostoru obcí v Česku nejprve pomocí operací s fuzzy čísly a dvou indikátorů (počet obyvatel a hustota zalidnění), následně pomocí fuzzy regulace a sedmi indikátorů. Využití fuzzy pro určení příslušnosti sídel k městskému a venkovskému prostoru se zabývá také Pászto a kol. (2014) nebo Burian a kol. (2013).

### **Další regionální přístupy**

Ve většině evropských států se příslušnost sídel k městskému nebo venkovskému prostoru posuzuje podle počtu obyvatel, avšak hraniční hodnota se mezi jednotlivými státy podstatně liší. V severských státech jako Norsko, Island, Švédsko jsou za městská sídla považována sídla s více než 200 obyvateli, v státech střední Evropy (Rakousko, Česko, Francie, Lucembursko) je tato hranice 2 000 nebo 5 000 obyvatel (Německo, Slovensko). Ve státech převážně jižní Evropy (Řecko, Itálie, Portugalsko, Španělsko, ale i Polsko a Švýcarsko) je tato hranice 10 000 obyvatel (Perlín, 2010).

Klasifikací území na město a venkov Anglie a Walesu se zabývá DEFRA – Department for Environment, Food & Rural Affairs (2004) a Bibby, Shepherd (2005). Tato metodika se zaměřuje na klasifikaci venkovských sídel (méně než 10 000 obyvatel, pokud je více než 10 000 jedná se o městskou oblast). Vstupními daty je velmi detailní grid (10x10 m) adresních bodů, který je dále agregován do menšího detailu (100x100 m). Klasifikace probíhá podle hustoty domácností v daných oblastech.



Dalším příkladem je studie ze Srbska (Gajić a kol., 2018), využívající metody vícerozměrné statistiky (analýza hlavních komponent, faktorová analýza a shluková analýza). Využito bylo patnáct indikátorů (například hustota zalidnění, podíl zemědělské půdy, vzdělanost, podíl zaměstnaných podle sektorů, podíl dojíždějících za zaměstnáním). Nejprve bylo využito PCA pro vytvoření a vybrání použitých hlavních komponent. Vybrané hlavní komponenty byly dále podrobeny faktorové analýze a následně bylo provedeno shlukování hodnot hlavních komponent pomocí nehierarchického K-means algoritmu. Výsledkem bylo pět shluků, které představují pět různých kategorií městských/venkovských oblastí.

## 3.2 Datové zdroje

Detailní typologie založené na větším množství atributů jsou většinou vytvářeny pro menší oblasti (státy), než je předmětem této práce. Z tohoto důvodu je obvykle jednodušší získat kvalitní a relevantní data z oficiálních zdrojů a zároveň zohlednit specifika administrativního členění daného území. Vzhledem k prostorovému rozsahu této práce (EU a přidružené státy) bylo nutné využít prostorová data z volně dostupných zdrojů, pokrývající celé požadované území i přes jejich zjevné nedostatky a problémy s nimi spojenými (více v kapitole Diskuze). Pro stejně rozsáhlou oblast jako v této práci jsou vytvářeny typologie Rural-urban typology a Degree of urbanisation, které jsou však založeny pouze na počtu obyvatel a hustotě zalidnění.

### EUROSTAT a GISCO

Statistický úřad Evropské unie (Eurostat) spravuje a poskytuje statistická data za členské a přidružené státy, která umožňují provádět srovnání mezi jednotlivými státy a regiony. Detailnost dat je různá, od dat sumarizovaných za celý stát až po územní jednotky LAU2 (v Česku obce). Kromě počtu obyvatel však žádná další data s rozlišením na LAU2 nepokrývají celé požadované území.

Geographic Information System of the Commission (GISCO) je součástí Eurostatu, která je zodpovědná za geografická data. Zabývá se tvorbou statistických a dalších tematických map, spravuje databáze prostorových dat a poskytuje další podobné služby Evropské komisi. Poskytuje různá tematická prostorová data za celou Evropu, včetně administrativních hranic, z nichž některá jsou veřejně dostupná a použitelná pro nekomerční účely. Data jsou dostupná v různých formátech prostorových dat (Eurostat, 2019).

Jako základ pro tvorbu datové sady pro tuto práci byla použita vrstva LAU2, obsahující pouze geometrii a atribut obsahující počet obyvatel. Geometrie vrstvy se vztahuje k roku 2013 (NUTS 2013), údaje o počtu obyvatel jsou většinou platné k roku 2017, pokud však nebylo možné data za tento rok použít, byla použita data z dřívějších let (nejstarší z roku 2011). Původní vrstva měla celkem 118 450 záznamů z čehož je ve 180 případech hodnota „Null“ (jedná se především o jezera ve Švýcarsku, Francii a Estonsku, která jsou klasifikována jako samostatné LAU2 a celý stát Makedonie). Jedná se částečně o nedostupná data a částečně o fakticky neobydlené oblasti. Hodnota „0“ se v datech vyskytovala ve 324 případech (neobydlené oblasti, jako jsou vojenské újezdy v Česku, nebo jezera a chráněné oblasti v Německu). Datovou vrstvu poskytl a vytvořil vedoucí práce ve spolupráci s Déborah Delahaye z dílčích dat veřejně dostupných na webu Eurostatu (Eurostat, 2019).

### **CORINE Land Cover**

COordination of INformation on the Environment (CORINE) je program, který zahájila Evropská komise v roce 1985, jehož cílem je koordinace a přístup ke kvalitním informacím o životním prostředí a přírodních zdrojích, které jsou srovnatelné v rámci Evropského společenství. Je součástí programu Copernicus (dříve GMES – Global Monitoring for Environment and Security).

Corine Land Cover (CLC) je celoevropská databáze land cover. Poprvé vznikla v roce 1990, dále v letech 2000, 2006, 2012 a 2018. V nejdetailejším tematickém rozlišení databáze obsahuje 44 tříd krajinného pokryvu a využití území, v nejnižším detailu rozlišení základních 5 tříd, které byly využity pro tuto práci (z roku 2018). Jsou to třídy: urbanizovaná území, zemědělské plochy, lesy a polopřírodní oblasti, humidní území a vodní plochy. Kompletní databázi lze volně stáhnout po registraci na webu CLC.

### **Open Street Map**

Jedná se o volně dostupnou vektorovou mapovou databázi. Projekt vznikl v roce 2004 a data jsou dostupná pod licencí Open Data Commons Open Database License (ODbL). Data může upravovat a vytvářet kdokoli, v důsledku toho je zhoršená kvalita a na takové podklady se nelze 100 % spolehnout (Nétek a Burian, 2018). Z tohoto důvodu se také liší detailnost a úplnost pokrytí jednotlivých států daty, což může ovlivnit analýzy. Tato práce je však zaměřena na volně dostupná data a v této kategorii patří data OSM (s přihlédnutím k požadovanému detailu a velikosti zkoumané oblasti) k nejhodnějším a v důsledku také zdrojem téměř jediným.

Existuje několik způsobů, jak získat data OSM, které jsou popsány na OSM Wiki (OpenStreetMap Wiki, 2019). Tyto zdroje se liší tím, jak často jsou data aktualizována, v jakém formátu a za jaká území jsou dostupná. Pro potřeby této práce byla data stažena z webu GEOFABRIK (GEOFABRIK, 2019). Data je zde možné stahovat po jednotlivých státech, případně menších oblastech, pokud jsou data za konkrétní stát příliš obsáhlá. Dostupná jsou ve formátu shapefile a vlastním formátu OSM a aktualizována jsou denně.

## 4 OBOHACENÍ DATOVÉ SADY

Pro obohacení základní datové sady byly použity základní GIS operace, jako je *clip*, *intersect*, *join*, *summarize* a další. Pomocí nich bylo možné připojit různá prostorová data ke geometrii administrativních jednotek LAU2. Takto obohacená datová sada na úrovni LAU2 z původní vrstvy z Eurostatu a dat z volně dostupných zdrojů, byla následně využita pro další analýzy. V následující části bude představen postup zpracování dat.

### 4.1 Původní data

Nejprve byla upravena původní vrstva počtu obyvatel v podrobnosti LAU2. Byly smazány některé nadbytečné „prázdné“ záznamy, např. celý stát Severní Makedonie, kde nebyla vyplněna informace o počtu obyvatel. Dále byly v některých přípustných případech (oblasti jako vojenské újezdy nebo jezera, kde skutečně nejsou obyvatelé) hodnoty null nahrazeny hodnotou 0. V průběhu práce bylo zjištěno, že některé LAU2 polygony jsou duplicitní, takové polygony byly identifikovány a následně byly duplikace odstraněny. Počet prvků ve vrstvě byl tímto snížen z původních 118 450 na 117 089.

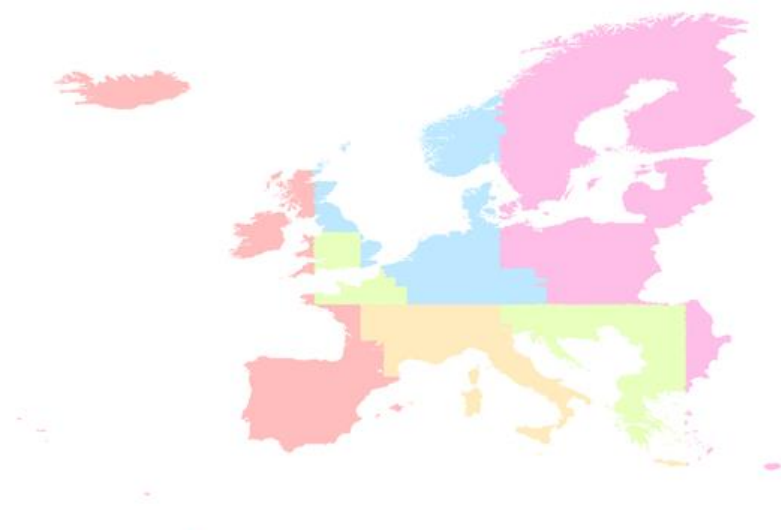
Z počtu obyvatel a rozlohy oblastí zjištěných v GIS prostředí bylo možné vypočítat hustotu zalidnění. Pro atributy počtu obyvatel, rozlohy a hustoty zalidnění byly vypočítány základní statistické charakteristiky (tabulka 1). Jednalo se o aritmetický průměr a medián. U všech třech atributů je výrazně méně hodnot vyšších než hodnota průměrná, nejvyšší rozdíl je patrný u atributu hustota zalidnění, kde nadprůměrné hodnoty, tedy více než 404 obyvatel na km<sup>2</sup> dosahuje pouze necelých 14 % ze všech zkoumaných LAU2. Už z toho je patrné, že se jedná se o poměrně vysokou hodnotu. Pro porovnání je hustota zalidnění celé EU asi 117 a Česka asi 136 obyv./km<sup>2</sup> (Evropská unie, 2017). Z tohoto srovnání vyplývá, že se v datech vyskytují oblasti s extrémními hodnotami jednotlivých indikátorů, které výrazně zvyšují průměrné hodnoty a v důsledku také ukazují velkou nerovnoměrnost hodnot.

Tab. 1 Jednoduché statistické charakteristiky atributů před obohacením datové sady

	Počet obyvatel	Rozloha [km <sup>2</sup> ]	Hustota zalidnění [obyv/km <sup>2</sup> ]
aritmetický průměr	4 498	42,12	404
medián	1 006	15,17	55
podíl hodnot vyšších než průměrná [%]	20,3	19,56	13,95
podíl hodnot nižších než průměrná [%]	79,7	80,44	86,05

## 4.2 CORINE Land Cover (CLC)

Nejprve byla základní vrstva LAU2 rozdělena na pět částí tak, aby jednotlivé dílčí vrstvy obsahovaly 25 000 polygonů a poslední vrstva zbytek (obr. 2) a to z důvodu časové náročnosti zpracování. Vrstva CLC byla převedena do vektorové podoby (funkce *raster to polygon*) a rozdělena na stejných pět částí jako vrstva LAU2 (funkce *clip*). Pomocí funkce *intersect* byly vzniklé polygony rozděleny podle hranic LAU2. Výsledkem těchto operací byly polygonové vrstvy, obsahující informaci o kategorii CLC, rozloze a příslušnosti ke konkrétnímu LAU2. Následně byly funkcí *summarize* vytvořeny tabulky obsahující informaci o rozloze v každé ze základních pěti CLC kategorií, které byly pomocí funkce *join* napojeny k vrstvě LAU2.



Obr. 2 Schéma rozdělení vrstvy CLC

## 4.3 Open Street Map (OSM)

Data OSM byla stažena z webu GEOFABRIK (GEOFABRIK, 2019) po jednotlivých státech nebo menších jednotkách (v případě Francie, Německa, Polska a Itálie) ve formátu shapefile a aktuální jsou k 1.4. 2019. Jedná se vždy o osmnáct tematických vrstev v jednom adresáři.

Nejprve byl vytvořen atribut rozlohy budov. Za každý stát byla vrstva budov (původní shapefile „buildings“) exportována z shapefilu do file geodatabase. Následně pomocí nástroje *intersect* s vrstvou LAU2 (nastavení only FID) byly smazány zbytečné atributy, (zpomalující práci) a k jednotlivým budovám byl přidělen identifikátor označující, do které LAU2 patří. V případě, že některé budovy zasahovaly do více LAU2, byly rozděleny. Dalším krokem byl výpočet celkové rozlohy budov v každé LAU2. Ve funkci *summarize* byl jako field to summarize nastaven nový atribut s ID LAU2 a požadovaná operace byla suma pro plochy budov. Takto vytvořené tabulky byly pomocí funkce *join* připojeny k vrstvě LAU2 a pomocí funkce *field calculator* doplněny do příslušných polí atributové tabulky.

Dalším vytvořeným atributem byl celkový počet point of interest (POI). Stejně jako u budov byly vrstvy POI (původní shapefile „pois“) převedeny do file geodatabase. Všechny vrstvy byly spojeny do jedné nástrojem *merge*. Pomocí funkce *identity* byl jednotlivým POI přiřazen identifikátor z vrstvy LAU2, ostatní atributy byly zachovány pro další práci. Funkcí *summarize* byla vytvořena tabulka s ID LAU2 a počtem POI který byl

stejným způsobem jako u budov přiřazen k vrstvě LAU2. Stejně tak byl vytvořen atribut kostelů, z shapefile kostelů (jedná se o chrámy i jiných náboženství, ale křesťanské kostely naprosto převažují, původní vrstva je označena „pofw“). Z celkových POI byly dále vytvořen atribut s počtem POI, které jsou typické spíše pro oblasti městského charakteru a dílčí atributy s jednotlivými tematickými kategoriemi takových POI. Tyto kategorie byly určeny na základě konzultace s vedoucím práce. Jsou to atributy:

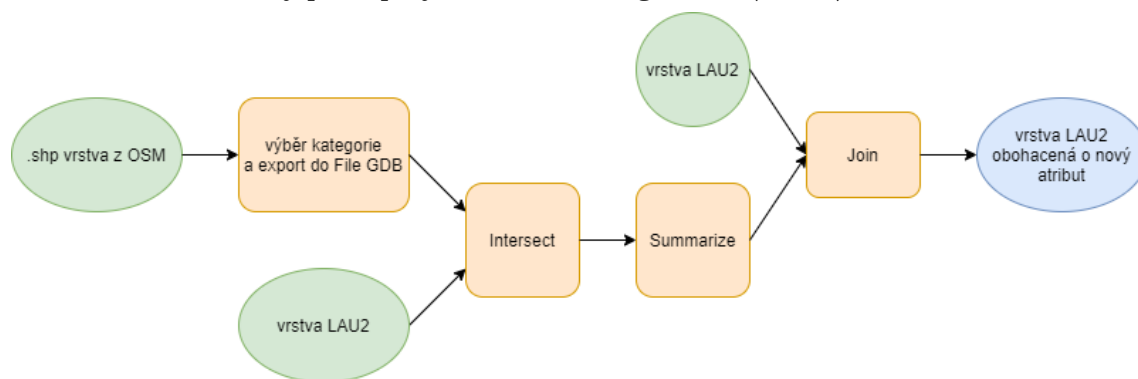
- POI\_ubytovani – hostel, hotel
- POI\_kulturni\_zarizeni – arts centre, artwork, cinema, library, museum, theatre
- POI\_jidlo – fast-food, food court, restaurant
- POI\_zdravotni – dentist, doctor, hospital
- POI\_nakupni\_centrum – mall
- POI\_banky\_bankomaty – atm, bank
- POI\_policie – police
- POI\_obchody – beauty shop, bicycle shop, computer store, department store, do-it-yourself, furniture shop, gift shop, jeweller, mobile phone shop, outdoor shop, supermarket
- POI\_univerzity – college, university

Následně byl vytvořen atribut s délkou železnic. Pomocí nástroje *feature class to feature class* byly z shapefile vrstvy železnic („railways“) exportovány do file geodatabase, s aplikovanou podmínkou výběru pouze kategorie rail, která obsahuje pouze plnohodnotné železnice s normálním rozchodem kolejí pro daný stát, určené pro převoz osob i věcí. Nástrojem *merge* byly vrstvy železnic spojeny do jedné, nástrojem *intersect* byly linie rozděleny hranicemi LAU2 a byl jim přiřazen identifikátor. Pomocí nástrojů *join* a *field calculator* byly hodnoty délky železnic přiřazeny k příslušným záznamům.

Stejným postupem jako atribut železnic byl vytvořen atribut délky chodníků obsažený ve vrstvě silnic („roads“) a požadovaný typ je nazván footway. Podobný význam jako footway má i kategorie path, která však znamená spíše pěší zóna, která nebyla zahrnuta do atributu chodníků. Z vrstvy silnic byly dále vytvořeny atributy délka cyklostezek z původní kategorie cycleway, délka tramvajových linek z původní kategorie tram, délka dálnic z původní kategorie motorway, délka silnic 1. třídy z původní kategorie primary, délka silnic 2. třídy z původní kategorie secondary, délka silnic 3. třídy z původní kategorie tertiary, délka rychlostních silnic z původní kategorie trunk, délka rezidenčních cest z původní kategorie residential a délka pěších zón z původní kategorie path. Význam uvedených kategorií (OpenStreetMap Wiki, 2019):

- Footway – spíše chodník
- Path – spíše pěší zóna (v městských oblastech)
- Cycleway – cyklostezka oddělená od jiných cest, např. silnice, chodníku
- Motorway – silnice s omezeným přístupem, obvykle s dvěma nebo více pruhy (dálnice)
- Trunk – silnice s velkým dopravním významem, která ale nemá charakter dálnice
- Primary – v ČR silnice I. Třídy, obecně s menším významem než trunk a spojuje velká města
- Secondary – v ČR silnice II. Třídy, obecně s menším významem než primary a spojuje města
- Tertiary – v ČR silnice III. Třídy, obecně s menším významem než secondary a spojuje menší města a vesnice
- Residential – cesty, které slouží k přístupu k domům, neslouží k propojování sídel

Opět stejně jako u železnic byl vytvořen atribut délky vodních toků z původní shapefile vrstvy vodních toků („waterways“). Použity byly všechny kategorie: river, stream, canal a drain. Obecný postup byl znázorněn diagramem (obr. 3).



Obr. 3 Obecný postup vytvoření nového atributu

## 5 KLASIFIKACE OBCÍ NA VENKOVSKÉ A MĚSTSKÉ

Kromě vizualizace jednotlivých atributů lze bez složitějších analýz vizualizovat i základní rozdělení oblastí na městské a venkovské. V této fázi práce byly vizualizace vytvořeny pomocí programu ArcMap z ArcGIS Desktop. Pro vytvoření následujících vizualizací byla využita data o počtu obyvatel převážně z roku 2017, která však nejsou kompletní pro celé území. Část dat proto byla doplněna o starší data (co nejbliže roku 2017). Nejstarší použitá data jsou z roku 2011.

### Podle počtu obyvatel

První vizualizace (obr. 4) zobrazuje rozdělení na městské a venkovské oblasti pouze podle počtu obyvatel za jednotky LAU2. Zobrazeny jsou pouze dvě kategorie rozdělené hranicí 3 000 obyvatel, která byla určena dle českého zákona 128/2000 Sb. O obcích. Mapa využívá metodu areálových znaků (Voženílek, Kaňok a kol., 2011). Na první pohled jsou patrné velké shluky v obou kategoriích. Rozlohou velké nepřirozené shluky městských oblastí pokrývají téměř celé státy například Švédsko a Polsko. Rozsáhlý shluk také vznikl v oblasti Nizozemska, Belgie a západní části Německa. Jedná se o poměrně očekávaný výsledek, který vyplývá z jedné umělé hodnoty, navíc definované legislativně pro Česko. Identifikovat lze některá velká města jako Paříž, Madrid nebo Prahu, ale většina jich splývá s okolními oblastmi.

### Podle hustoty zalidnění

Druhá vizualizace (viz obr. 5) využívá pro rozdělení oblastí atribut hustoty zalidnění a jedná se o metodu kartogramu, která využívá data přepočtená na plochu. Opět jsou LAU2 děleny pouze do dvou kategorií městských a venkovských oblastí. Hranicí pro odlišení je hodnota 150 obyvatel/km<sup>2</sup>, která vychází z klasifikace OECD (2011b). Výsledek pro Česko je velmi podobný s předchozí vizualizací, liší se tím, že v okolí velkých měst, především ve východní části země, vznikají větší shluky městských oblastí. Severské státy jsou oproti předchozí vizualizaci téměř celé klasifikovány jako venkovské v důsledku prostorově rozsáhlých LAU2. Lze však identifikovat velká města, jako Stockholm, Oslo, Helsinky, Malmö, a další, které v předchozí vizualizaci splývají s okolními oblastmi, které jsou však poměrně řídko osídleny. Největší shluk městských oblastí se rozprostírá na západě Německa, jihu Nizozemska a severu Belgie. Shluk měst v této oblasti skutečně je, není však tak velký a homogenní. Oproti předchozí lze poměrně zřetelně rozeznat oblast Alp s řídkým výskytem velkých měst. Pro srovnávání tak rozsáhlého a různorodého území je tato klasifikace rozhodně lepší než předchozí, využívající pouze počet obyvatel.

### Podle počtu obyvatel i hustoty zalidnění

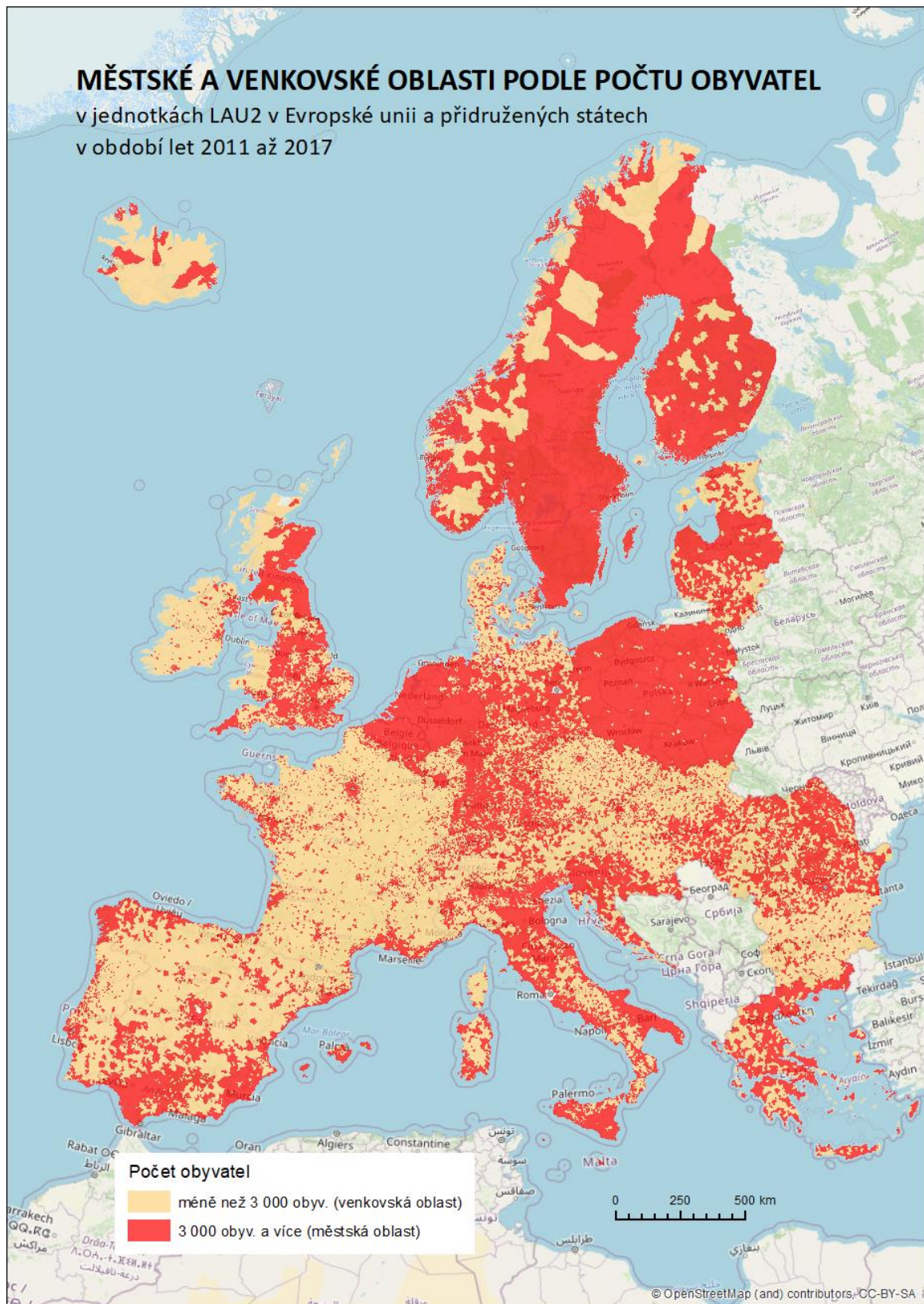
Další možností jak odlišit městské a venkovské oblasti, je pomocí kombinace počtu obyvatel a hustoty zalidnění, kterou využívá následující vizualizace (obr. 6). Využity byly kategorie a hraniční hodnoty využívané v metodice Degree of Urbanisation. Oblasti jsou rozděleny na tři kategorie – Rural areas (venkovské oblasti), Towns and Suburbs (menší města a suburbia) a Cities (velkoměsta).

- Towns and Suburbs – více než 300 obyv./km<sup>2</sup> a zároveň více než 5 000 obyvatel
- Cities – více než 1 500 obyv./km<sup>2</sup> a zároveň více než 50 000 obyvatel
- Rural areas – vše mimo výše uvedené kategorie

Velmi problematický je tento způsob klasifikace pro velká města, která jsou rozdělena na větší množství LAU2, což je běžné například pro města ve Velké Británii, kde žádná oblast nebyla klasifikována jako velkoměsto (ani Londýn). Oproti předchozí vizualizaci, je také méně městských oblastí určeno ve Skandinávii. V Česku do kategorie velkoměst byla zařazena Praha, Brno, Kladno, České Budějovice a část Ostravy. Patrně největší shluk městských oblastí byl vymezen stejně jako v předchozí vizualizaci na západě Německa, severu Belgie a jihu Nizozemska. V okolí a často přímo okolo velkoměst se vyskytují oblasti menších měst a suburbii.

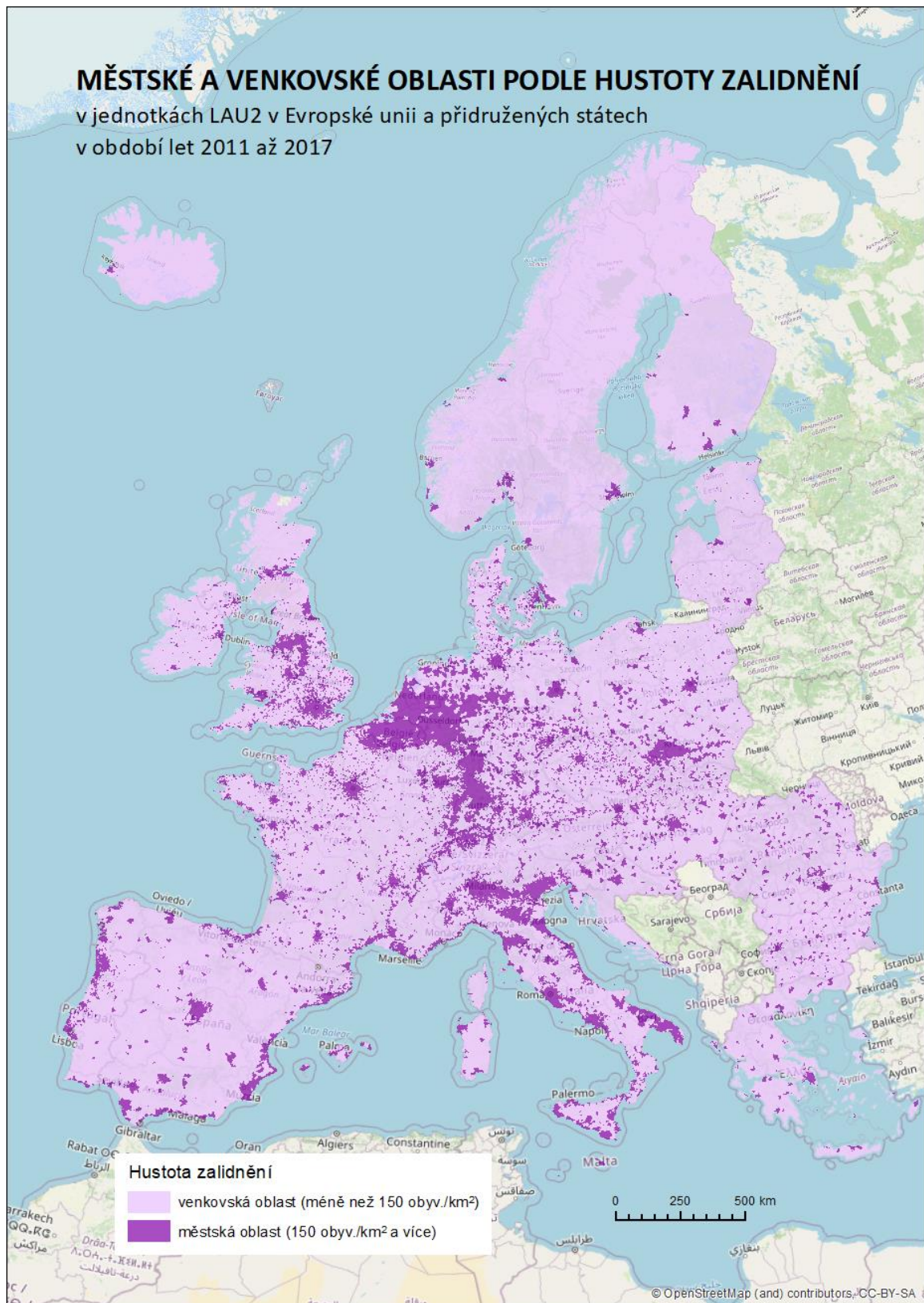
Pomocí jednoduchých vizualizací je možné odlišit městské oblasti od venkovských. Takový přístup však není příliš vhodný pro rozsáhlá území, navíc s různou velikostí zkoumaných statistických jednotek. Jak již bylo zmíněno v rešerši, tento problém je v oficiálních metodikách obvykle řešen agregací klasifikovaných oblastí do vyšších jednotek, případně nahrazení nižších jednotek v části území vyššími územními jednotkami.



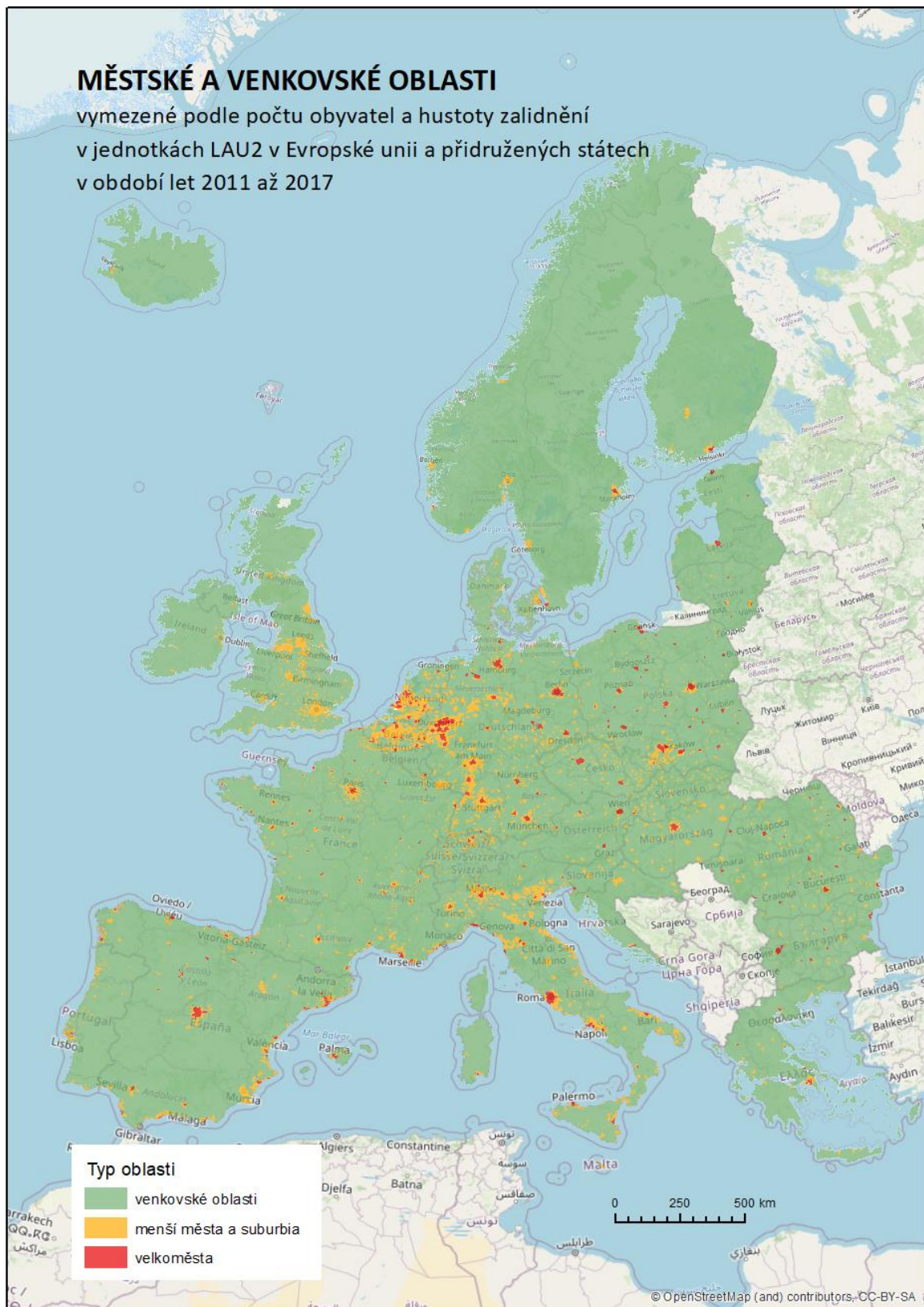


Obr. 4 Rozdělení oblastí na městské a venkovské podle počtu obyvatel  
(vlastní zpracování, podklad: OSM)





Obr. 5 Rozdělení oblastí na městské a venkovské podle hustoty zalidnění  
(vlastní zpracování, podklad: OSM)



Obr. 6 Rozdělení oblastí na městské a venkovské podle počtu obyvatel i hustoty zalidnění (vlastní zpracování, podklad: OSM)



## 6 STATISTICKÉ HODNOCENÍ DAT

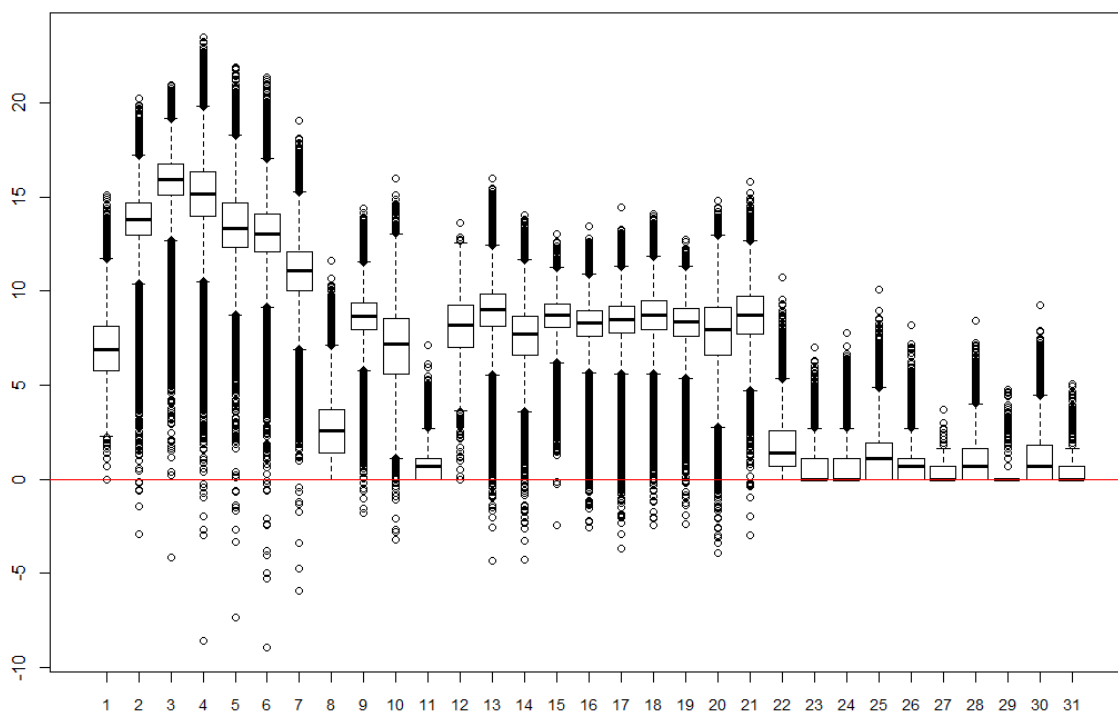
Pro nové atributy s absolutními hodnotami byly vytvořeny další atributy s přepočtem absolutních hodnot na relativní, přepočítané na plochu (počet objektů na km<sup>2</sup>, délky linií na km<sup>2</sup>, podíl ploch). Tabulka 2 obsahuje přehled nepřepočtených atributů (pro další práci byly atributy očíslovány). Při některých analýzách či vizualizacích dat bylo nutné hodnoty 0 nahradit hodnotou NULL (v RStudiosu NA), proto je informace o počtu hodnot 0 pro jednotlivé atributy také obsažena v této tabulce. Hodnota NULL znamená že hodnota atributu je neznámá nebo neexistující (Dobešová, 2004), kdežto hodnota 0 použitá v datech znamená absenci jevu či objektu v území. Pro statistické analýzy dat, byl proveden export do CSV pomocí nástroje *Export Feature Attribute to ASCII*. Aby byl výsledkem CSV soubor je nutné do řádky s pojmenováním výstupního souboru připsat koncovku .csv. V rámci tohoto nástroje je taky možné vybrat exportované atributy.

Tab. 2 Přehled atributů

číslo	atribut	četnost hodnoty 0	číslo	atribut	četnost hodnoty 0
1	počet obyvatel	380	17	délka silnic 2. třídy	41 353
2	urbanizované území	23 096	18	délka silnic 3. třídy	49 460
3	zemědělské plochy	3 609	19	délka rychlostních silnic	100 106
4	lesy a polopřirodní oblasti	14 566	20	délka pěších zón	31 386
5	humidní území	104 094	21	délka rezidenčních silnic	5 314
6	vodní plochy	88 216	22	počet "městských" POI	55 076
7	plocha budov	7 880	23	počet POI typu ubytování	95 208
8	počet POI	16 653	24	počet POI typu kulturní zařízení	88 248
9	délka železnic	79 265	25	počet POI typu stravování	69 798
10	délka chodníků	51 811	26	počet POI typu zdravotní zařízení	99 793
11	počet kostelů	81 586	27	počet POI typu nákupní centra	116 052
12	délka tramvajových linek	116 169	28	počet POI typů banky a bankomaty	88 181
13	délka vodních toků	14 142	29	počet POI typu policie	105 431
14	délka cyklostezek	79 168	30	počet POI typu obchody	86 098
15	délka dálnic	100 044	31	počet POI typu univerzity	115 156
16	délka silnic 1. třídy	68 200			

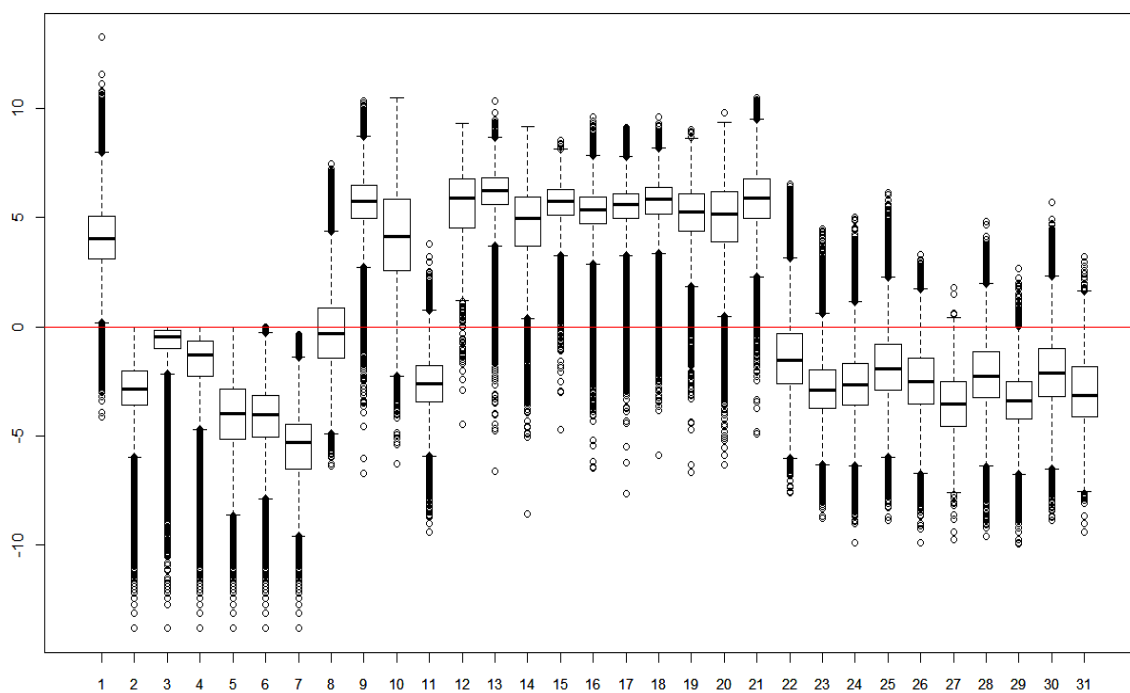
### 6.1 Statistické zhodnocení vstupních dat pomocí boxplotu

Původní atributy nepřepočítané na plochu byly vizualizovány pomocí krabicových grafů (boxplot, obr. 7), k čemuž bylo využito RStudio. Pro správné vykreslení bylo nutné v datech nahradit hodnoty 0 hodnotou NA. Tohoto nahrazení bylo docíleno při importu CSV souboru s daty vyplněním možnosti na .strings hodnotou 0. Vzhledem k velkému rozsahu hodnot byla využita logaritmická stupnice. Hodnoty na ose y menší než nula tedy znamenají hodnoty v intervalu (0, 1) v původních datech. Vzhledem k tomu, že plošné objekty jsou v m<sup>2</sup> a liniové objekty v metrech jedná se o velice zanedbatelné hodnoty (např. desítky centimetrů u silnic). Atributy jsou očíslovány dle tabulky 2. U všech atributů je patrně poměrně velké množství odlehklých hodnot.



Obr. 7 Boxplot absolutních hodnot jednotlivých atributů

Boxploty byly následně vytvořeny také pro hodnoty přepočítané na plochu (obr. 8). Číslování a odstranění nulových hodnot bylo provedeno shodně s předchozími boxploty. I v tomto případě byla využita logaritmická stupnice pro lepší přehlednost, důsledkem čehož je, že žádné atributy původně vyjadřující plochu a po přepočtu vyjadřující procentuální zastoupení dané kategorie v území LAU2 nemohou přesáhnout hodnotu jedna (100 %), tudíž i hodnotu nula na logaritmické stupnici.



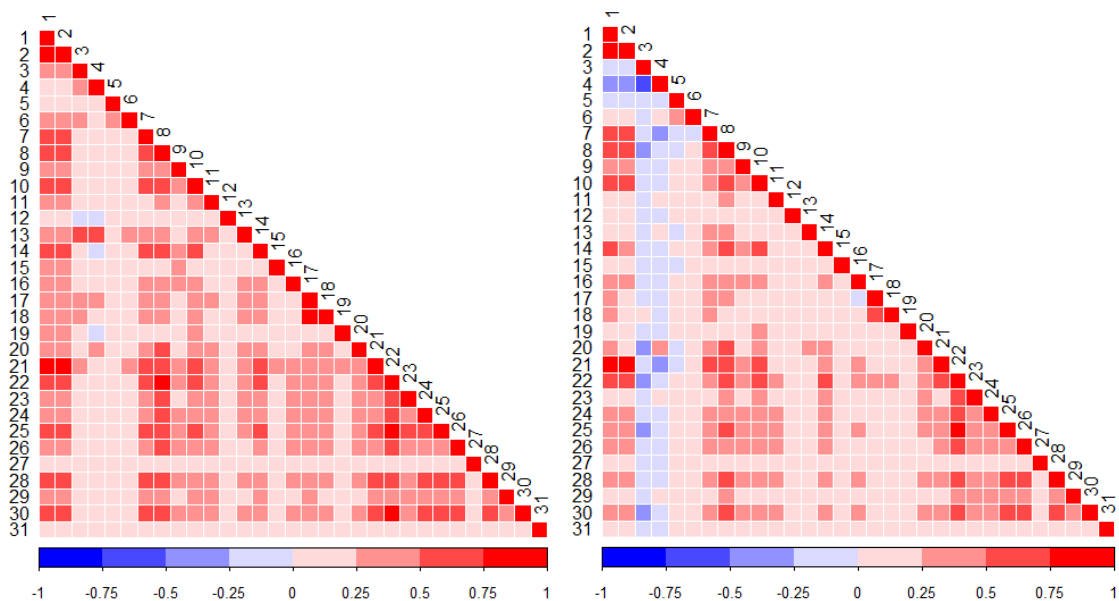
Obr. 8 Boxplot hodnot jednotlivých atributů přepočtených na plochu

## 6.2 Statistické zhodnocení vstupních dat pomocí korelací

Korelačním koeficientem je kvantifikována míra vztahu mezi dvěma veličinami. Nejobvyklejší je Pearsonův korelační koeficient. Hodnoty korelace mohou nabývat hodnot z intervalu  $<-1, 1>$ . Hodnoty blízké nule značí, že hodnoty příliš nekorelují a znaky jsou nezávislé. Pokud je hodnota kladná, blíží se jedné, znamená to, že vysoké hodnoty jednoho znaku souvisí s vysokými hodnotami druhého znaku. Pokud je hodnota záporná, blíží se minus jedné, znamená to, že vysoké hodnoty jednoho znaku souvisí s nízkými hodnotami znaku druhého.

Z důvodu poměrně velkého množství outlierů byl zvolen Spearmanův korelační koeficient, který je pořadový, neparametrický a málo citlivý na přítomnost odlehklých hodnot na rozdíl od běžnějšího Pearsonova korelačního koeficientu. Při výpočtu korelace pomocí Spearmanova koeficientu je hodnota nahrazena pořadím hodnoty prvku. (Meloun, Militký 2002).

Pomocí RStudia byly vypočteny hodnoty korelace mezi jednotlivými atributy, které byly následně vizualizovány ve formě korelační matice. První korelační matice byla vytvořena z absolutních hodnot atributů (obr. 9 vlevo) a druhá korelační matice (obr. 9 vpravo) z hodnot přepočítaných na plochu (na  $\text{km}^2$ ).



Obr. 9 Korelační matice absolutních hodnot (vlevo) a hodnot přepočtených na plochu (vpravo)

Následující text se zabývá jednotlivými atributy a jejich významem pro odlišení městských a venkovských oblastí. Čísla uvedená v závorkách jsou odkazem na řádky z tabulky 2.

### **Počet obyvatel (1)**

Počet obyvatel a hustota zalidnění je základním ukazatelem pro vymezení městských a venkovských oblastí používaný v téměř každé výše zmíněné metodice. Čím větší je hustota zalidnění oblasti, tím kompaktnější pravděpodobně bude i zástavba a tím více městské bude dané sídlo. Počet obyvatel výrazně koreluje s dalšími atributy, jejichž vysoké hodnoty značí spíše městské oblasti jako jsou plocha budov, počet POI, délka chodníků, délka cyklostezek, délka rezidenčních silnic, počet městských POI a některé podkategorie městských POI.

### **Urbanizované území (2)**

Vyšší podíl urbanizovaného území značí spíše městské oblasti. Tento indikátor koreluje se stejnými atributy jako počet obyvatel. Přepočtený na plochu záporně, ačkoliv ne úplně významně, koreluje s ostatními kategoriemi CLC, nejvíce s lesy a polopřírodními oblastmi. Navíc je to jeden z indikátorů, který pomáhá odlišit části měst, které jsou rozděleny do více LAU2 a které by pouze podle počtu obyvatel nebyly zařazeny do správné kategorie.

### **Zemědělské plochy (3)**

Venkovské oblasti jsou charakteristické vyšší mírou zemědělství. Podíl lidí zaměstnaných v zemědělství je často využíván pro vymezení venkovských oblastí. Pro takto rozsáhlé území a v potřebném detailu však nebylo možné získat tato data, atribut podílu zemědělské plochy s ním však podstatně souvisí. Přepočtený na plochu se všemi ostatními indikátory koreluje záporně, většinou však pouze mírně. Nejvýznamněji záporně koreluje s indikátorem lesů a polopřírodních oblastí. Indikátor podílu zemědělské půdy (ze stejného zdroje) použili Gajić a kol. (2018) k vymezení městských/venkovských oblastí Srbska.

### **Lesy a polopřírodní oblasti (4)**

Vyšší podíl lesů je charakteristický pro venkovské oblasti, stejně jako vyšší podíl zemědělské půdy, se kterou však koreluje záporně. To teoreticky může značit odlišný typ venkova (například ve vyšších nadmořských výškách), kde nejsou vhodně podmínky pro zemědělství. S většinou ostatních indikátorů koreluje záporně, většinou však nevýznamně. Kromě různých typů lesů zahrnuje louky a pastviny, rašeliniště a vřesoviště, přechodné oblasti keřů, pláže, skalnaté a další podobné oblasti.

### **Humidní území a vodní plochy (5 a 6)**

Indikátory nejsou běžně k podobným analýzám využívány a nebyly využity ani pro většinu analýz v této práci. Jejich možné využití spočívá ve smyslu, že se nejedná o vhodné plochy pro zástavbu a nejsou typické pro městské oblasti. S žádným dalším indikátorem významně nekorelují, a to překvapivě ani s vodními toky.

### **Plocha budov (7)**

Čím vyšší podíl budov v území tím více městské je dané sídlo. Hodnoty korelují se stejnými indikátory jako atribut počtu obyvatel (po přepočtení na plochu pouze s některými, například s urbanizovaným územím a délkou rezidenčních silnic).

### **Počet POI (8)**

Jedná se o celkový počet POI v daném území převzatý z OSM, tedy nejružnější kategorie. Je pravděpodobné, že více POI bude zaznamenáno v městských oblastech, avšak vzhledem k jejich různorodosti byl vytvořen také atribut městských POI. Ten silně koreluje s většinou ostatních typů POI, což může být způsobeno mírou aktivity přispěvatelů OSM. Spolu s urbanizovaným územím je vhodným indikátorem, pro odlišení městských oblastí měst, které se skládají z více LAU2.

### **Délka železnic (9)**

Nejedná se o indikátor běžně využívaný k podobným analýzám. Indikátor koreluje se stejnými indikátory jako počet obyvatel, ale méně významně, z čehož se dá usuzovat, že více železnic v ploše sídla bude ukazovat na více městské oblasti. Jednou z možných interpretací je, že v nejvýrazněji městských oblastech se nachází větší nádraží, ve kterém se sbíhá více kolejí, čímž roste i hustota železnic v oblasti. Oblasti, ležící na trase mají díky železnici dopravně blíže k městu a mohou tvořit jeho zázemí a být přechodnými oblastmi mezi ryze městskou a venkovskou kategorií. Naopak oblasti kterými železnice neprocházejí vůbec mají k velkým městům dále a jsou více venkovské.

### **Délka chodníků (10)**

V městech (i menších) je obvyklé, že chodníky jsou téměř na každé ulici, po obou stranách cest. V malých vesnicích, jsou chodníky typicky pouze na jedné straně ulice a často pouze na hlavní ulici, proto více chodníků v ploše sídla bude ukazovat více městský charakter. Indikátor významně kladně koreluje s počtem obyvatel a se stejnými dalšími indikátory, jako počet obyvatel.

### **Počet kostelů (11)**

Tento indikátor významně nekoreluje s žádným jiným indikátorem. V podobě přepočítané na plochu nastává problém, že např. ve městě Olomouc je asi 25 kostelů na asi 100 000 obyvatel (1 kostel na asi 4 000 obyvatel). V daleko menších obcích o pár stech obyvatel se však kostel také typicky nachází, což přináší daleko vyšší „počet“ kostelů vzhledem k počtu obyvatel. V úplně nejmenších obcích zase kostel obvykle není, proto je poměrně složité tento atribut správně interpretovat a ve většině dalších analýz nebude použit.

### **Délka tramvajových linek (12)**

Tento indikátor s žádným jiným významně nekoreluje. Tramvajová doprava je typická pouze ve velkých městech, které lze dostatečně dobře odlišit i na základě jiných atributů. V naprosté většině 116 149 z celkového počtu 117 089 záznamů obsahuje nulovou hodnotu, proto ve většině analýz není použit.

### **Délka vodních toků (13)**

V podobě přepočítané na plochu významně nekoreluje s žádným jiným indikátorem. V nepřepočítané podobě nejvíce koreluje se zemědělskými plochami a lesními a polopřírodními oblastmi. Vzhledem k tématu však interpretace atributu není zcela jednoznačná, protože zdroj vody je jednak důležitý pro zemědělství, typické pro venkov, ale historicky města vznikala především podél vodních toků.



### **Délka cyklostezek (14)**

V podobě přepočtené na plochu nejvíce koreluje s počtem obyvatel a počtem městských POI. Cyklostezky jsou více typické spíše v městských oblastech, případně ve venkovských oblastech v zázemí větších měst, proto vyšší hodnota bude značit spíše městské, než typicky venkovské sídlo.

### **Délka dálnic a rychlostních silnic (15 a 19)**

Tyto dva atributy jsou si svým významem velmi podobné a v rámci některých států se jejich význam může značně překrývat. V podobě přepočtené na plochu ani jeden výrazně nekoreluje s žádným jiným atributem. Tyto typy komunikací spojují největší města a mají i nadnárodní dopravní význam. To, že menší sídlo leží na spojnici velkých měst a přes jeho administrativní území prochází dálnice většinou pravděpodobně neovlivňuje jeho charakter natolik aby se podle toho dalo tvrdit, že je více městské nebo venkovské.

### **Délka silnic 1. třídy (16)**

Silnice první třídy spojují větší města (OpenStreetMap Wiki, 2019). Indikátor mírně koreluje s počtem obyvatel, urbanizovaným územím, plochou budov, celkovým počtem POI i počtem městských POI, délkou železnic, chodníků a rezidenčních silnic. Pravděpodobně tak vyšší hodnoty budou znamenat více městský charakter sídla, případně může pomoci pro odlišení různých typů venkovských oblastí ve smyslu, že oblastí kterými takové silnice prochází mají dopravně blíže a mohou tvořit zázemí velkých měst.

### **Délka silnic 2. a 3. třídy (17 a 18)**

Nejvýznamněji tyto dva atributy korelují mezi sebou. Jedná se o komunikace spojující města, menší města a vesnice. Především výskyt silnic 3. třídy může ukazovat na více venkovská sídla, které neleží na významných komunikacích spojující velká města, nejsou na ně přímo napojena a nespádají do jejich zázemí.

### **Délka pěších zón (20)**

Nejvíce koreluje s celkovým počtem POI. Má velmi podobný význam jako chodníky, v rámci OSM, vzhledem k různému vymezení těchto pojmů v různých státech, a překladu názvů původních atributů do češtiny nejsou oproti chodníkům ostře vymezeny. V rámci práce atribut není u většiny analýz využit.

### **Délka rezidenčních silnic (21)**

Nejedná se o cesty spojující sídla, ale pouze zajišťující přístup k domům (obecně spíše k budovám), proto není překvapivé, že atribut silně koreluje s atributy počtu obyvatel, urbanizovaným územím a dalšími, u nichž vyšší hustota značí více městský charakter sídla.

### **Počet městských POI (22)**

Tato kategorie byla z celkového počtu POI vyčleněna na základě expertního odhadu a konzultace s vedoucím práce. Jedná se o kategorie POI, které nejsou obvyklé v ryze venkovských oblastech a pokud ano tak ne ve velkém počtu. Spolu s urbanizovaným územím a celkovým počtem POI je vhodným indikátorem pro odlišení městských oblastí měst, které se skládají z více LAU2.

### **Subkategorie městských POI (23 až 31)**

Většinou více či méně korelují mezi sebou, až na kategorie univerzit a nákupních center. To může být způsobeno velkým počtem nulových hodnot u těchto kategorií v kombinaci s pořadovým korelačním koeficientem. U většiny analýz bude použit pouze jejich nadřazený atribut počtu městských POI.

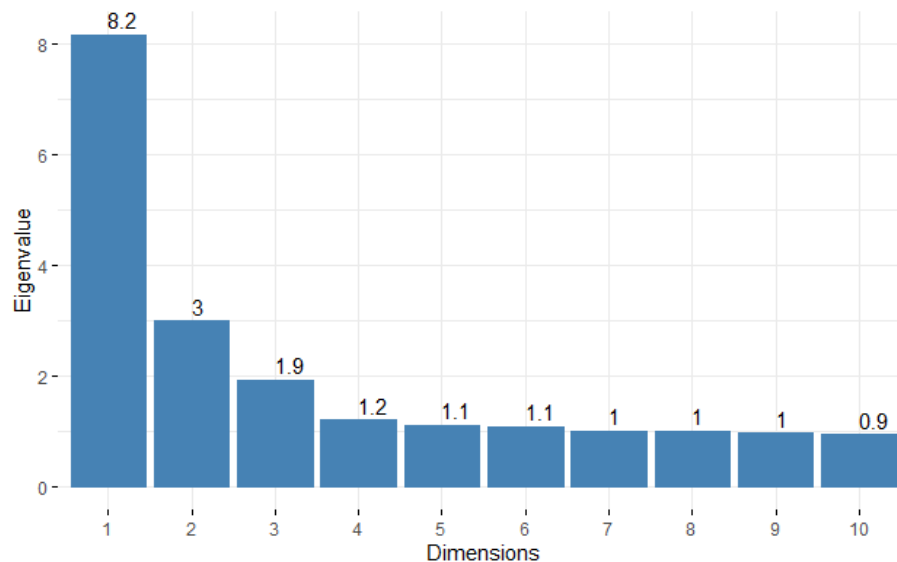
## 7 ANALÝZA HLAVNÍCH KOMPONENT

Metoda analýzy hlavních komponent vychází z předpokladu, že dvě proměnné mohou být závislé proto, že obě měří tutéž skrytou společnou veličinu, nazývanou společný faktor nebo hlavní komponenta (Šarmanová, 2012). Jedná se o transformační metodu vícerozměrné statistiky. Cílem analýzy hlavních komponent (PCA – Principal Component Analysis) je transformace dat z původních závislých proměnných do menšího počtu latentních (skrytých), nezávislých proměnných. Nových proměnných je méně, mají vhodnější vlastnosti, vystihují významnou část proměnlivosti původních proměnných (závisí na konkrétních datech, počtu proměnných, počtu použitých hlavních proměnných aj.) a jsou vzájemně nekorelované. Latentní proměnné jsou nazývány hlavními komponentami a jsou lineární kombinací původních proměnných. První hlavní komponenta popisuje největší část rozptylu původních dat, druhá hlavní komponenta největší část rozptylu neobsaženého v první hlavní komponentě a nejméně informace je obsaženo v poslední hlavní komponentě (Meloun, Militký 2004). Problémem PCA může být například to, že data neobsahují předpokládanou informaci. Meloun a Militký, 2002 uvádí, že outliery by v rámci přípravy dat na PCA měly být zachovány, aby nedošlo k odstranění cenné informace.

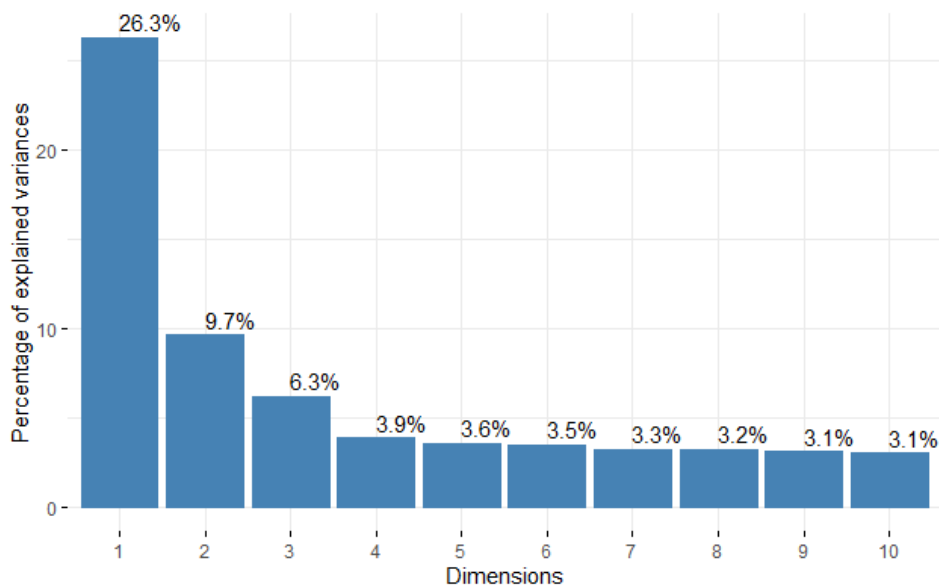
Před provedením PCA by měla být vyšetřena závislost vstupních proměnných, protože vzájemná závislost atributů může ovlivnit výsledky (Šarmanová, 2012). K tomuto účelu se využívá koeficient kovariance nebo častěji koeficient korelace. Korelace mezi atributy byly zjištěny a ohodnoceny pomocí korelačních matic výše v práci. Indikátory, které spolu korelují často přispívají do stejných komponent. Zvolena byla data přepočítaná na plochu, protože oproti datům nepřepočítaným spolu méně korelují.

Důležitou součástí PCA je určení počtu „použitelných“ hlavních komponent. Univerzálně dopředu nelze říct kolik hlavních komponent bude vhodné použít. K tomuto určení slouží tři základní pomocná kritéria – výběr komponent s vlastními čísly většími než jedna (pokud jsou data standardizována, tak hodnota vlastního čísla větší než jedna znamená, že příslušná komponenta vysvětluje více rozptylu, než je obsaženo v některém z původních atributů). Druhým pomocným kritériem je výběr komponent, které společně vysvětlují určité procento rozptylu (běžně 90 %, ale závisí na tématu analýzy), a třetím pomocné kritérium využívá tzv. Scree Plot (indexový graf úpatí vlastních čísel), kde je hledán bod zlomu od rychlého poklesu k pozvolnému. Tento způsob popisuje Meloun a Militký (2002) jako hledání zlomového bodu mezi kolmou stěnou užitečných hlavních komponent a vodorovným dnem neúčinných hlavních komponent. Scree plot je sloupcový diagram zobrazující na ose y hodnoty vlastních čísel, nebo reziduální rozptyl a na ose x stoupající pořadová čísla hlavních komponent.

Pro vykonání PCA a sestavení všech příslušných grafů v rámci této práce byl použit program RStudio (pro výpočet hlavních komponent funkce `prcomp()`). Pro tuto analýzu byly hodnoty standardizovány.



Obr. 10 Indexový graf (Scree Plot) úpatí vlastních čísel



Obr. 11 Indexový graf (Scree Plot) rozptylu obsaženého v jednotlivých komponentách

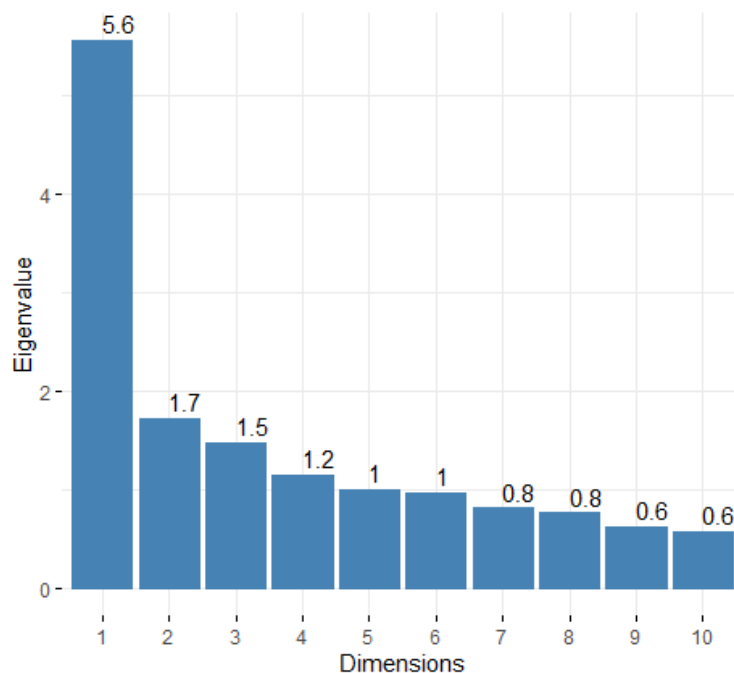
Graf na obr. 11 zobrazuje kolik procent rozptylu původních dat vysvětluje prvních deset hlavních komponent, graf na obr. 10 zase hodnoty vlastních čísel prvních deset hlavních komponent. Tyto grafy slouží pro určení vhodného počtu hlavních komponent. V tomto případě podle bodu zlomu od rychlého k pozvolnému poklesu hodnot by bylo vhodné použít první tři hlavní komponenty, ty však vysvětlují pouze 42,2 % rozptylu původních dat. Podle pravidla použití tolika komponent, aby společně vysvětlovali alespoň 90 % rozptylu by muselo být použito dvacet hlavních komponent. Pokud by měly být použity hlavní komponenty s vlastními čísly s hodnotou vyšší než jedna, bylo by jich šest a vysvětlovaly by 53,2 % rozptylu. V tomto případě je velmi obtížné najít kompromis mezi velkou ztrátou informace a rozumným počtem hlavních komponent. Proto bylo rozhodnuto redukovat počet atributů vstupujících do této analýzy.

Odebrány byly dílčí kategorie městských POI s čísly 23 až 31, protože jsou již zastoupeny v atributu městských POI, navíc obsahují množství nulových hodnot

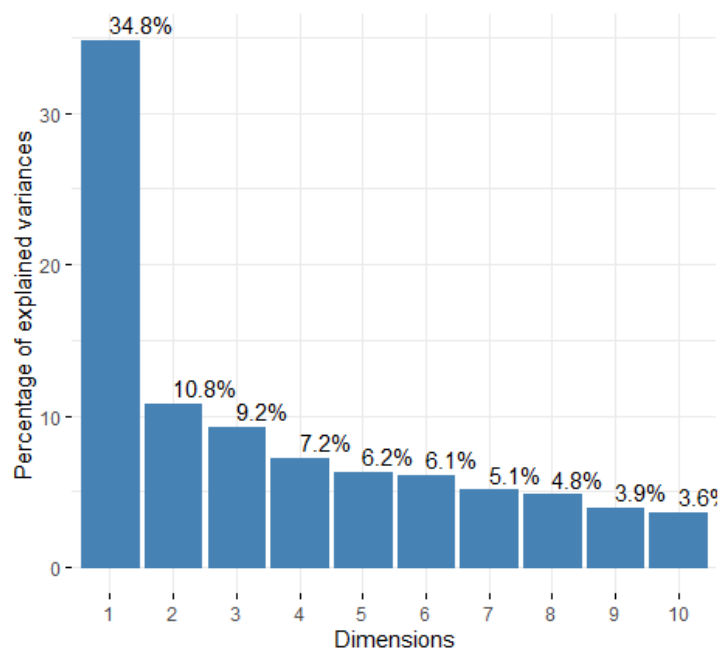
(tabulka 2) a většina z nich mezi sebou významně koreluje. Na základě konzultace s vedoucím práce byly odebrány atributy číslo 5, 6 a 20 tj. CLC vodní plochy, humidní území a pěší zóny kvůli jejich nízké relevanci vzhledem ke smyslu analýzy (hodnocení příslušnosti obcí k městu/venkovu). Po prozkoumání příspěvků jednotlivých atributů do několika prvních hlavních komponent a vzhledem k velkému počtu nulových hodnot byly odebrány také atributy číslo 12, 15 a 19 (délka tramvajových linek, dálnic a rychlostních silnic). Posledním odebraným atributem byl počet kostelů. Zvažováno bylo též odebrání atributu s celkovým počtem POI (číslo 8), ten ale byl zachován, protože při prozkoumání dalších výsledků provedené PCA byl zjištěn jeho velký příspěvek do první hlavní komponenty. Přidán byl atribut počtu obyvatel. Vstupními atributy do další PCA tedy jsou revidované atributy v tabulce 3. Korelující skupina atributů důležitých pro určení míry příslušnosti oblasti k městu/venkovu byla zachována, aby byl zajištěn vznik takové hlavní komponenty (jedné z prvních, aby mohla ovlivnit shlukování) která pomůže odlišení městských a venkovských oblastí.

Tab. 3 Vybrané vstupní atributy pro analýzu hlavních komponent

číslo	atribut	číslo	atribut
1	počet obyvatel	9	délka chodníků na km <sup>2</sup>
2	hustota zalidnění	10	délka vodních toků na km <sup>2</sup>
3	podíl urbanizovaného území na km <sup>2</sup>	11	délka cyklostezek na km <sup>2</sup>
4	podíl zemědělské plochy na km <sup>2</sup>	12	délka silnic 1. třídy na km <sup>2</sup>
5	podíl lesních a polopřírodních oblastí na km <sup>2</sup>	13	délka silnic 2. třídy na km <sup>2</sup>
6	podíl ploch budov na km <sup>2</sup>	14	délka silnic 3. třídy na km <sup>2</sup>
7	počet POI na km <sup>2</sup>	15	délka rezidenčních silnic na km <sup>2</sup>
8	délka železnic na km <sup>2</sup>	16	počet městských POI na km <sup>2</sup>

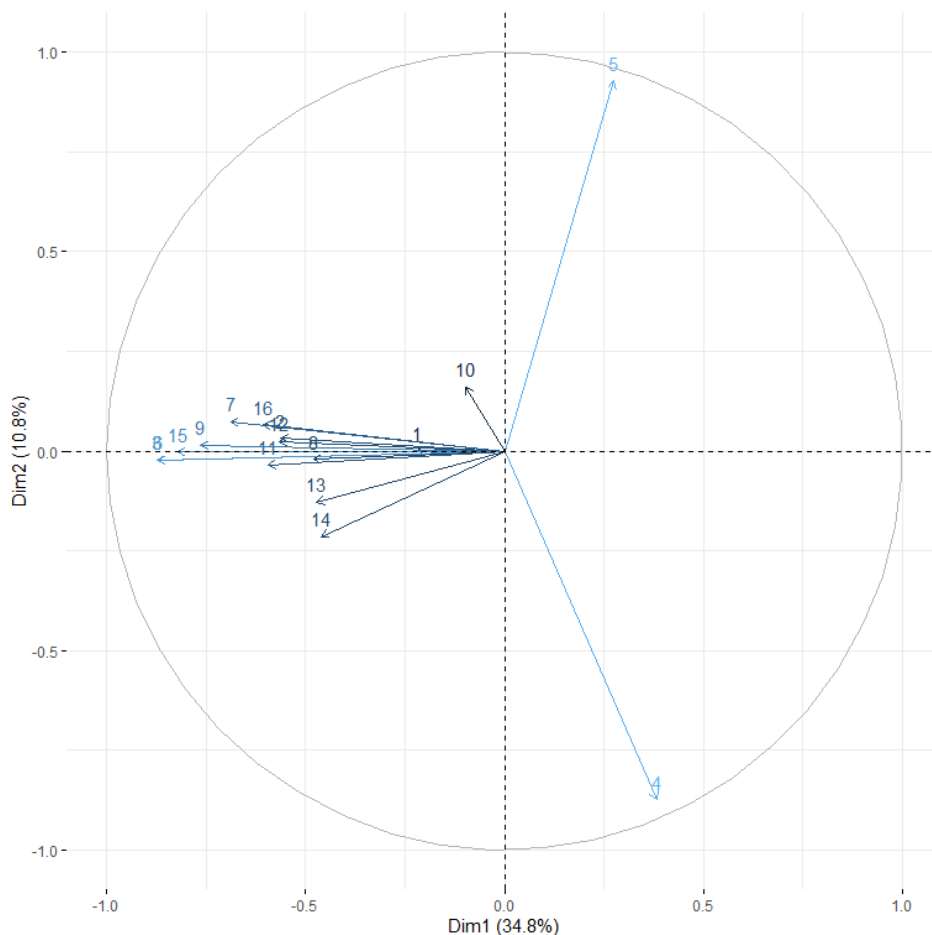


Obr. 12 Indexový graf (Scree Plot) úpatí vlastních čísel



Obr. 13 Indexový graf (Scree Plot) rozptylu obsaženého v jednotlivých komponentách

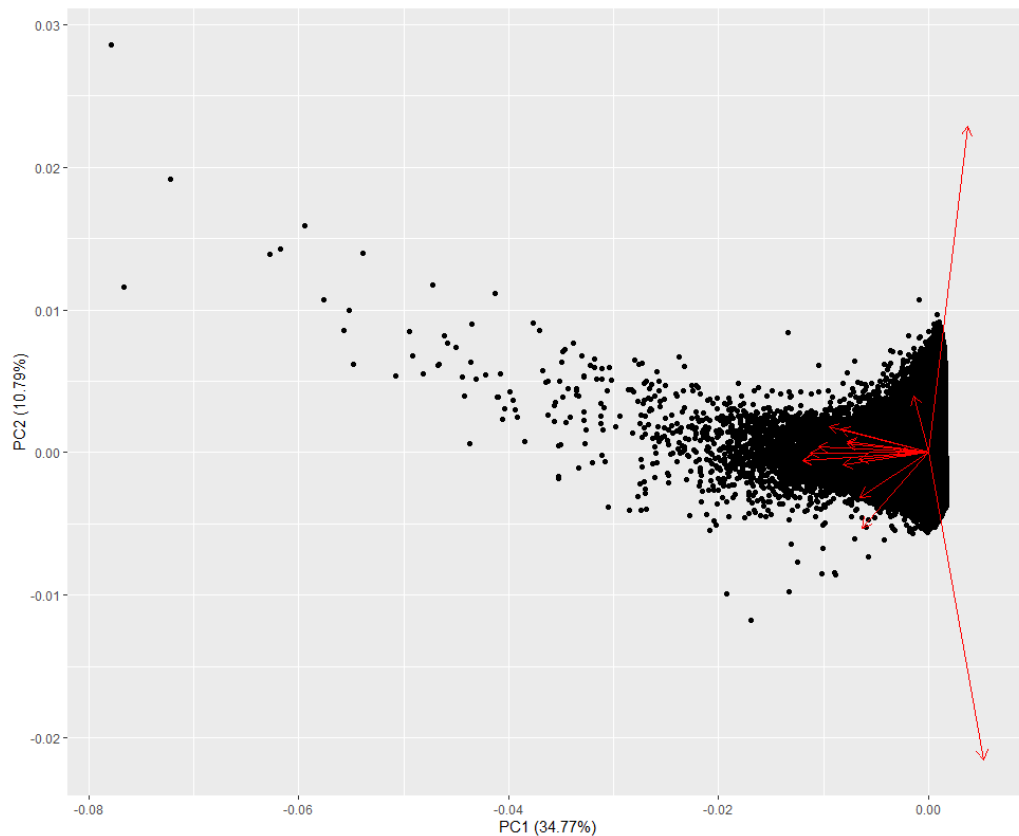
Graf na obr. 13 zobrazuje kolik procent rozptylu původních dat vysvětluje prvních deset hlavních komponent, graf na obr. 12 zase hodnoty vlastních čísel prvních deset hlavních komponent. K pozvolnému poklesu hodnot vysvětleného rozptylu dochází již po první hlavní komponentě. Jako „užitečné“ byly vybrány první čtyři hlavní komponenty s hodnotou vlastních čísel vyšších než jedna. Společně pak tyto hlavní komponenty vysvětlují téměř 62 % rozptylu původních dat, přičemž hodnoty 90 % je dosaženo až s desátou hlavní komponentou. Stejný způsob výběru „užitečných“ hlavních komponent byl použit ve studii ze Srbska (Gajić a kol., 2018), kde bylo PCA využito ke stejnému účelu.



Obr. 14 Graf komponentních vah (zátěží) pro první a druhou hlavní komponentu

Graf komponentních vah (zátěží) zobrazuje komponentní váhy pro první dvě komponenty. Zátěže (ang. loadings) představují míru korelace mezi původními atributy a novými proměnnými (hlavními komponentami). Z grafu je patrné, že nejméně důležité jsou atributy počtu obyvatel a délky vodních toků, které leží velmi blízko počátku, naopak nejdůležitější (blízko jednotkové kružnice) jsou atributy podíl zemědělské plochy, lesních a polopřírodních oblastí, podíl ploch budov, podíl urbanizovaného území, rezidenčních silnic a chodníků. Patrný je shluk korelujících atributů, v levé části grafu, ze kterého nejvíce vybočují atributy délek silnic 2. a 3. třídy. Úhel mezi atributy tohoto shluku a atributy důležitými pro druhou hlavní komponentu (zemědělské plochy, lesní a polopřírodní oblasti) značí mírnou zápornou korelaci. Atributy důležité pro druhou hlavní komponentu korelují mezi sebou záporně. Zátěže pro všechny použité komponenty jsou uvedeny v tabulce 4.

Dvojný graf (obr. 15) kombinuje rozptylový diagram komponentního skóre jednotlivých objektů a graf komponentních vah. Černé body znamenají hodnoty všech objektů v prostoru tvořeném prvními dvěma hlavními komponentami, červené šipky pak to samé, co šipky v grafu komponentních vah. Pomocí prvních dvou hlavních komponent je možné identifikovat jeden shluk a poměrně velké množství extrémních hodnot, které jsou odlehle spíše na první hlavní komponentě. Pomineme-li odlehle hodnoty, k největšímu rozptylu na první hlavní komponentě dochází v nulovém bodě na druhé hlavní komponentě a stejně tak na druhé hlavní komponentě v nulovém bodě na první hlavní komponentě.



Obr. 15 Dvojný graf pro první a druhou hlavní komponentu

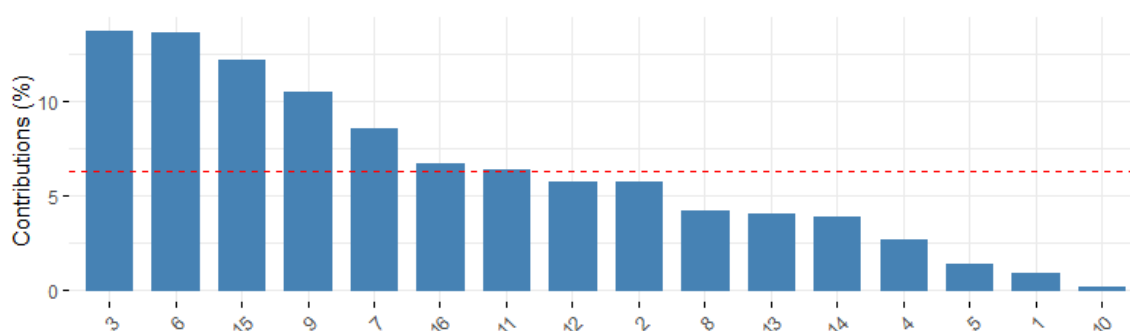
Tab. 4 Matice zátěží pro vybrané čtyři hlavní komponenty

	Hlavní komponenta			
	1	2	3	4
počet obyvatel	-0,220	0,002	0,125	0,045
hustota zalidnění	-0,563	0,033	0,134	-0,026
podíl urbanizovaného území	-0,874	-0,023	0,307	-0,057
podíl zemědělské plochy	0,383	-0,874	-0,190	-0,072
podíl lesních a polopřírodních oblastí	0,275	0,928	-0,042	0,126
podíl ploch budov	-0,871	-0,022	-0,124	-0,008
počet POI na km <sup>2</sup>	-0,689	0,074	-0,696	0,013
délka železnic na km <sup>2</sup>	-0,481	-0,020	0,161	-0,180
délka chodníků na km <sup>2</sup>	-0,764	0,016	0,025	-0,115
délka vodních toků na km <sup>2</sup>	-0,098	0,161	0,118	0,195
délka cyklostezek na km <sup>2</sup>	-0,594	-0,034	0,301	-0,123
délka silnic 1. třídy na km <sup>2</sup>	-0,565	0,024	-0,005	-0,488
délka silnic 2. třídy na km <sup>2</sup>	-0,473	-0,128	0,022	0,673
délka silnic 3. třídy na km <sup>2</sup>	-0,462	-0,215	0,129	0,576
délka rezidenčních silnic na km <sup>2</sup>	-0,822	-0,002	0,291	-0,043
počet městských POI na km <sup>2</sup>	-0,608	0,066	-0,758	0,024
Podíl celkového rozptylu (%)	34,770	10,790	9,207	7,190
Kumulativní součet rozptylu (%)	34,770	45,560	54,767	61,957



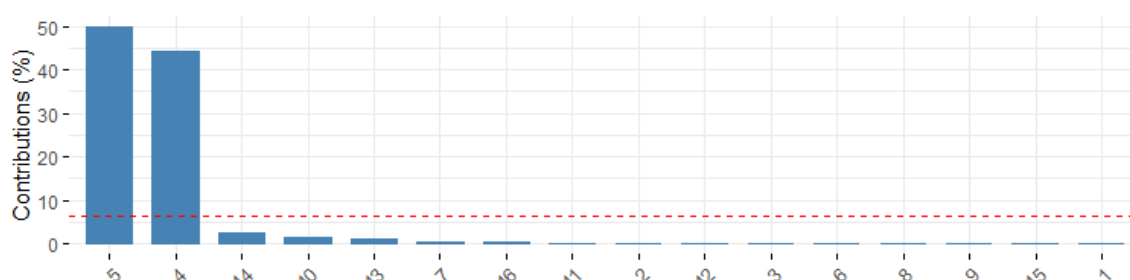
## 7.1 Interpretace vybraných hlavních komponent

První hlavní komponenta vysvětluje nejvíce, asi 34,8 % rozptylu. Podstatně do ní přispívá sedm atributů: podíl urbanizovaného území, podíl ploch budov, délka rezidenčních silnic, délka chodníků, celkový počet POI, počet městských POI a délka cyklostezek a všechny s ní koreluji záporně. Jedná se o atributy jejichž vyšší hodnota předpokládá více městský charakter. Významně koreluje také s atributem hustoty zalidnění. Tato komponenta pomáhá oddělit městské a venkovské oblasti. Oblasti s vyšší kladnou hodnotou u této komponenty budou vzhledem k záporným korelacím spíše venkovského charakteru, a naopak oblasti s vyšší zápornou hodnotou budou spíše městského charakteru. Komponenta byla nazvána charakter oblasti. Pojmenování je velmi obecné, ale vzhledem k tomu, že smyslem analýzy je vymezení městské a venkovské oblasti je pochopitelné čeho se onen charakter týká.



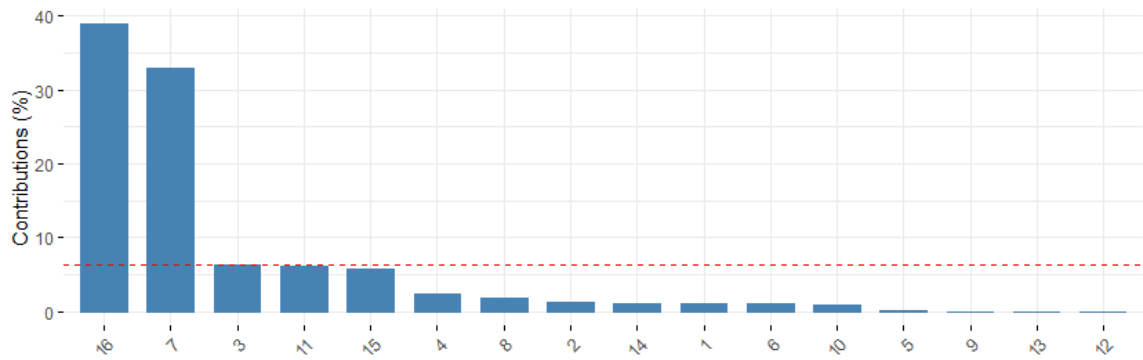
Obr. 16 Příspěvky atributů do první hlavní komponenty

Druhá hlavní komponenta vysvětluje asi 10,8 % rozptylu. Významně do ní přispívají pouze dva atributy, podíl lesních a polopřírodních oblastí, se kterým koreluje kladně a podíl zemědělské plochy, se kterou koreluje záporně. Jedná se o nejrelevantnější použité atributy, jejichž vysoké hodnoty předpokládají venkovský charakter sídel. K největšímu rozptylu hodnot dochází v oblasti nízkých hodnot na první hlavní komponentě, vyjadřující spíše venkovské oblasti. Komponenta byla nazvána podíl zemědělských ploch oproti přírodním oblastem. Vzhledem ke korelacím, vyšší kladná hodnota značí velké množství přírodních oblastí, naopak vyšší záporná hodnota značí více zemědělských ploch.



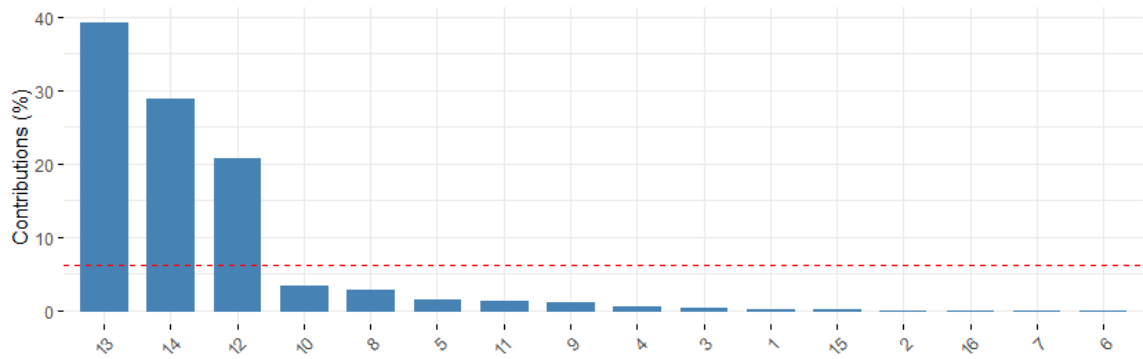
Obr. 17 Příspěvky atributů do druhé hlavní komponenty

Třetí hlavní komponenta vysvětluje asi 9,2 % rozptylu. Významně do ní přispívají dva atributy, celkový počet POI a počet městských POI a s oběma koreluje záporně. Čím vyšší je hodnota objektu na této komponentě, tím více venkovský má charakter. Komponenta byla nazvána hustota bodů zájmu.



Obr. 18 Příspěvky atributů do třetí hlavní komponenty

Čtvrtá hlavní komponenta vysvětluje asi 7,2 % rozptylu. Významně do ní přispívají tři atributy, délka silnic 1., 2. a 3. třídy. S délkou silnic 1. třídy koreluje záporně a se zbylými dvěma koreluje kladně. Nízká hodnota znamená vyšší hustotu silnic 1. třídy a vysoká hodnota vyšší hustotu silnic 2. a 3. třídy. Komponenta byla nazvána hustota silniční sítě podle třídy.



Obr. 19 Příspěvky atributů do čtvrté hlavní komponenty

## 8 TYPOLOGIE JEDNOTEK LAU2 PODLE SHLUKOVÉ ANALÝZY

Analýza shluků je metoda, určená pro vyšetřování podobnosti vícerozměrných objektů a jejich klasifikaci do shluků (tříd). Metody shlukování se dělí na hierarchické a nehierarchické. Hierarchické se dále dělí na aglomerativní a divizivní.

Hierarchické aglomerativní postupy fungují na principu postupného spojování objektů do shluků. Nejprve je vypočtena matice vzdáleností mezi objekty. Objekty jejichž vzdálenost je nejmenší jsou spojeny do prvního shluku, následně je vypočtena nová matice vzdáleností, ve které jsou vynechány objekty prvního shluku, který je zařazen jako celek. Celý postup se opakuje do té doby, než jsou zařazeny všechny objekty do jednoho velkého shluku, nebo do předem zadaného počtu shluků. Divizivní postup je opačný, začíná s celým souborem objektů a systém shluků je vytvořen jeho postupným dělením. (Meloun a Militký, 2002).

Při nehierarchickém shlukování jsou určeny objekty, které jsou typickými představiteli shluku. K těm jsou podle vzdálenosti přiřazeny jednotlivé objekty. Jedním z příkladů nehierarchického shlukování je algoritmus K-Means. Vhodný je pro vytváření malého počtu shluků z velkého počtu objektů, ale vyžaduje data bez odlehlých hodnot, které se ve shlukovaných datech v této práci vyskytují, proto musí být použita více robustní metoda, která není na odlehlé hodnoty tolik citlivá. Takovou metodou je metoda středů-medoidů.

Meloun a Militký (2004) uvádí, že medoid (střed shluku) je objekt, jehož průměrná vzdálenost k ostatním objektům shluku je minimální. Pro každý shluk existuje jeden medoid a medoidy i shluky se vytvářejí na základě vzdáleností (nepodobnosti). Jedním z algoritmů založených na principu medoidů je metoda PAM (Partition Around Medoids), která je ovšem omezena počtem klasifikovaných objektů, který je v rámci této práce příliš velký. Existuje však rozšíření metody PAM – CLARA (Clustering Large Applications), která byla využita. Další možností bylo z dat odstranit outliery a následně shlukovat pomocí K-means, v tom případě by však nebyly klasifikovány všechny oblasti, čímž by došlo ke ztrátě informace a výsledná datová sada by nebyla kompletní pro celé klasifikované území.

Algoritmus PAM se snaží minimalizovat celkovou vzdálenost mezi objekty shluku (suma vzdáleností mezi středem shluku a ostatními objekty shluku) tak, že nejprve nalezne reprezentativní soubor objektů, kterých je stejný počet jako požadovaných shluků. Každý z těchto reprezentativních objektů představuje střed (medoid) shluku, ke kterému jsou ostatní objekty přiřazeny tak, aby celková vzdálenost byla co nejnižší. Polohy reprezentativních objektů jsou vybírány v iteracích do té doby, dokud se hodnota celkové vzdálenosti snižuje (Meloun, Militký 2002).

Algoritmus CLARA nehledá medoidy pro celý datový soubor, hledá pouze pro menší vzorek dat (uživatel zadá velikost vzorku, podle které algoritmus náhodně vybere objekty). Tento výběr je iterační a stejně jako u PAM se snaží zmenšovat vzdálenosti objektů ve shluku. Pro výsledné shlukování (všech objektů) jsou vybrány medoidy s co nejmenší průměrnou vzdáleností od ostatních objektů (Kaufman, Rousseeuw 1990).

Kromě shluků samotných jsou výsledkem i reprezentativní objekty (představující typické zástupce kategorie) a tzv. siluety. Meloun a Militký (2004) uvádí, že silueta je statistika, která poskytuje klíčovou informaci o dobrém a špatném shluku. Pro každý objekt, představuje míru úspěšnosti klasifikace do shluku porovnáním vzdálenosti objektu k objektům v rámci stejného shluku vůči vzdálenostem k objektům sousedního

shluku. Siluety nabývají hodnot od -1 do 1, kde vysoká kladná hodnota značí správné zařazení, protože vzdálenost k objektům stejného shluku je výrazně kratší než vzdálenost k objektům sousedního shluku. Hodnota blízká nule vyjadřuje, že objekt leží na pomezí dvou shluků a do konkrétního shluku byl zařazen spíše náhodně. Hodnota blízká -1 značí, že objekt byl klasifikován špatně.

Průměrná hodnota siluety celého souboru a jednotlivých shluků je užitečným ukazatelem pro určení vhodného počtu shluků. Dle Melouna a Militkého (2002) je přijatelnou průměrnou hodnotou alespoň 0,51. Pravděpodobně proto, že na základě dat v této práci nejsou jasně a viditelně oddělené shluky (jak lze vidět níže v práci na obr. 20) nebylo po testování několika desítek různých nastavení dosaženo tak vysoké hodnoty, jak uvádí Meloun a Militký (2002) (více v tab. 5). Lepšího výsledku nebylo docíleno ani při vyšším počtu shluků, ani při větším vzorku dat.

Tab. 5 Průměrné hodnoty siluet pro různá nastavení shlukování

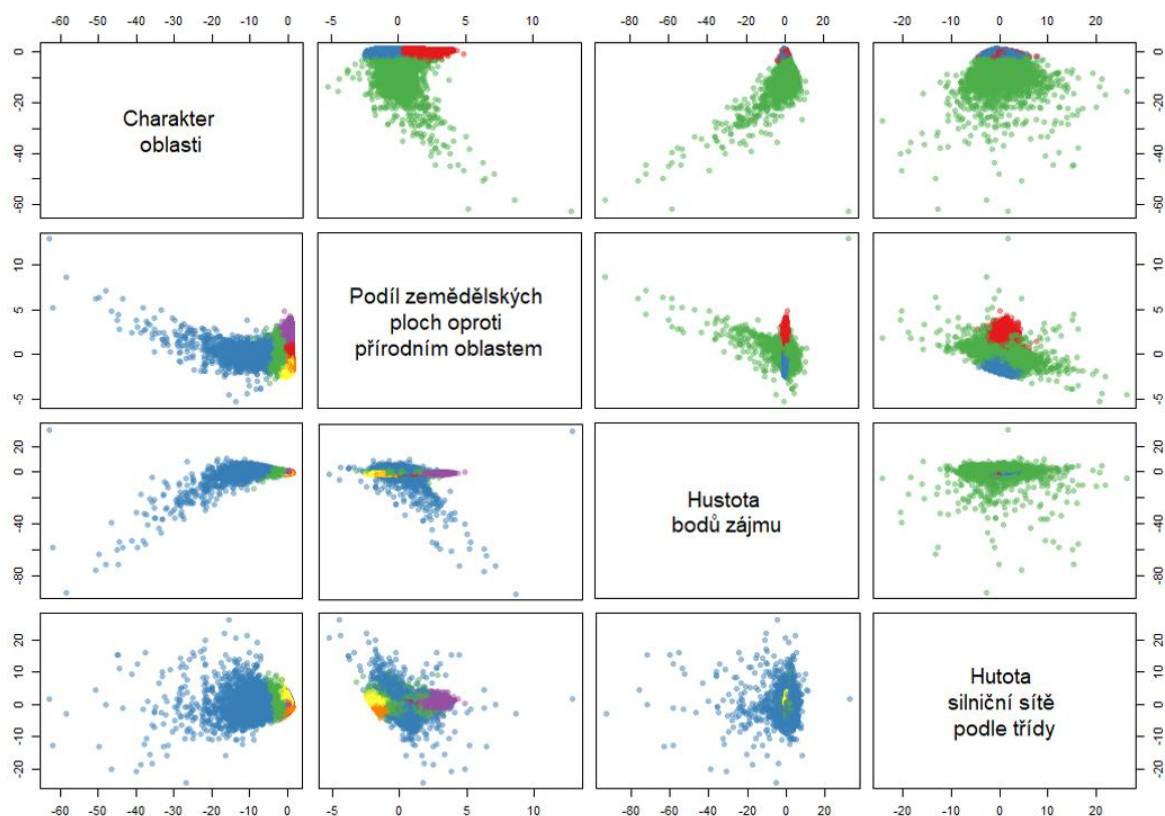
počet shluků	průměrná hodnota siluet s velikostí vzorku:		
	500	1 000	2 000
2	0,387	0,318	0,366
3	0,383	<b>0,417</b>	0,343
4	0,298	0,293	0,313
5	0,323	0,292	0,356
6	0,374	0,380	<b>0,410</b>
7	0,346	0,357	0,344
8	0,363	0,304	0,300
9	0,325	0,364	0,364
10	<b>0,405</b>	0,314	0,315
11	0,340	0,343	0,392

V tabulce 5 byly vyznačeny tři nejlepší dosažené výsledky shlukování. Možnost s deseti shluky byla vyřazena, protože dva ze shluků mají průměrnou hodnotu siluet 0. Možnost se třemi shluky byla pomocí matice bodových grafů srovnána s možností se šesti shluky (obr. 19). Průměrná hodnota siluet je o necelé 1 % lepší pro tři shluky. V obou možnostech je průměrná hodnota siluet u jednotlivých shluků kladná a vyšší než 0. U možnosti se třemi shluky je možné rozlišit jeden shluk městských oblastí a dva shluky rozdílných venkovských oblastí, které jsou rozlišeny pomocí druhé hlavní komponenty. Největší výhodou šesti shluků a hlavním důvodem, proč byla tato možnost nakonec vybrána k další práci je vzniklý shluk (zelený shluk ve spodní části obr. 20) přechodných oblastí mezi oblastmi nejvíce městskými a nejvíce venkovskými.

## 8.1 Interpretace shluků

Výsledky shlukování byly interpretovány pomocí hodnot objektů ve shlucích na jednotlivých hlavních komponentách (obr. 20) a pomocí vizuální analýzy mapy znázorňující zařazení LAU2 do jednotlivých shluků. Vytvořen byl také graf (obr. 21), zobrazující jednotlivé LAU2 jako body, obarvené podle zařazení do shluku, podle počtu obyvatel (osa x) a hustoty zalidnění (osa y). Stejně jako výše v práci pro boxplot i zde bylo nutné použít logaritmickou stupnici a to na obou osách. Cílem interpretace bylo pojmenování shluků, ověření správnosti shlukování a nalezení pravidelností v zařazení podobných oblastí do téhož shluku. V tabulce 6 je zachováno automatické číslování

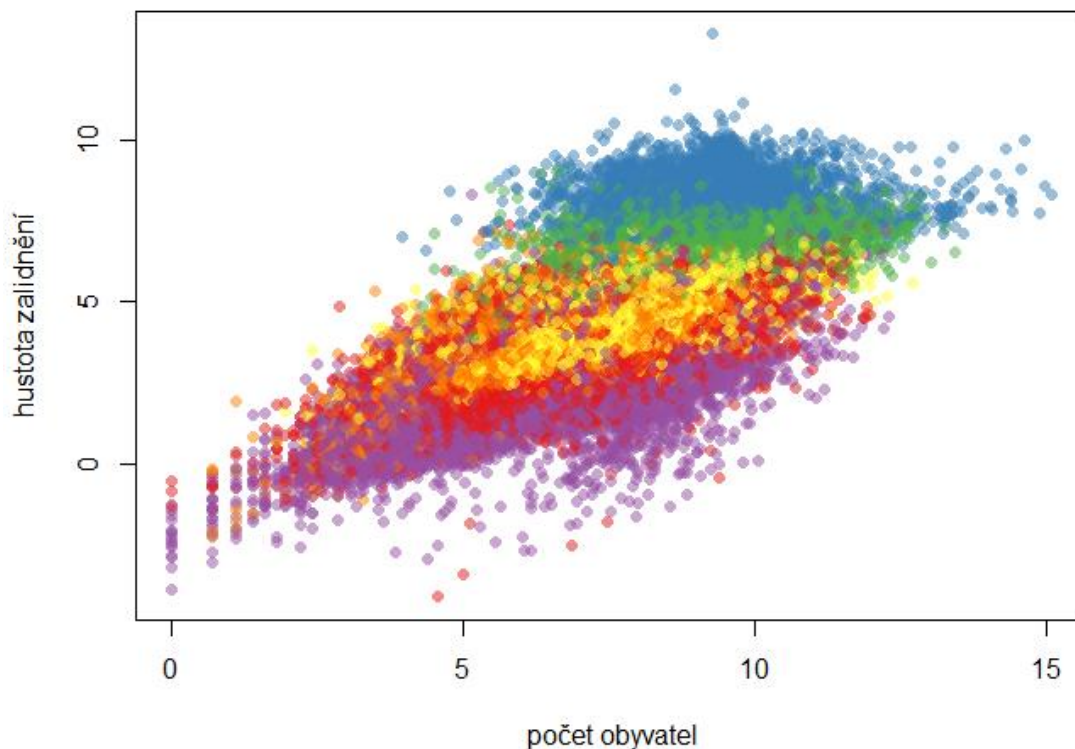
shluků tak, jak vyšlo pomocí RStudia, které je použito i ve vytvořené datové sadě. Pro lepší srovnání byly barvy zachovány i v následujících datových náhledech, které mají v legendě uvedeny pouze ty kategorie pro které jsou příkladem. Na obrázku 28 byl vyhotoven náhled na celou oblast už v barvách použitých ve výsledné webové mapové aplikaci. Vzhledem k tomu, že jsou použita data z různých let (počet obyvatel převážně 2017, OSM duben 2019, CLC 2018) nelze přesně určit rok, ke kterému je vytvořena typologie platná a aktuální. Určeno bylo období mezi lety 2011 a 2019.



Obr. 20 Srovnání shlukování do šesti (dolní část) a třech shluků (horní část)

Tab. 6 Pojmenování a základní charakteristika shluků ze spodní části obrázku 19

Shluk	Počet LAU2	Název shluku	Průměrný počet obyv.	Průměrná hustota zalidnění
1 (červená)	29 838	Venkovské oblasti s průměrným podílem zemědělských ploch	2 552	79,3
2 (modrá)	5 849	Výrazně městské oblasti	26 033	4 842,0
3 (zelená)	10 451	Spíše městské oblasti	13 271	1 028,8
4 (fialová)	20 518	Venkovské oblasti s malým podílem zemědělských ploch	2 302	57,1
5 (oranžová)	26 971	Venkovské oblasti s výrazným podílem zemědělských ploch a lepší silniční dostupností velkých měst	1 708	74,7
6 (žlutá)	23 462	Venkovské oblasti s výrazným podílem zemědělských ploch a rozvinutou regionální silniční infrastrukturou	2 824	67,7



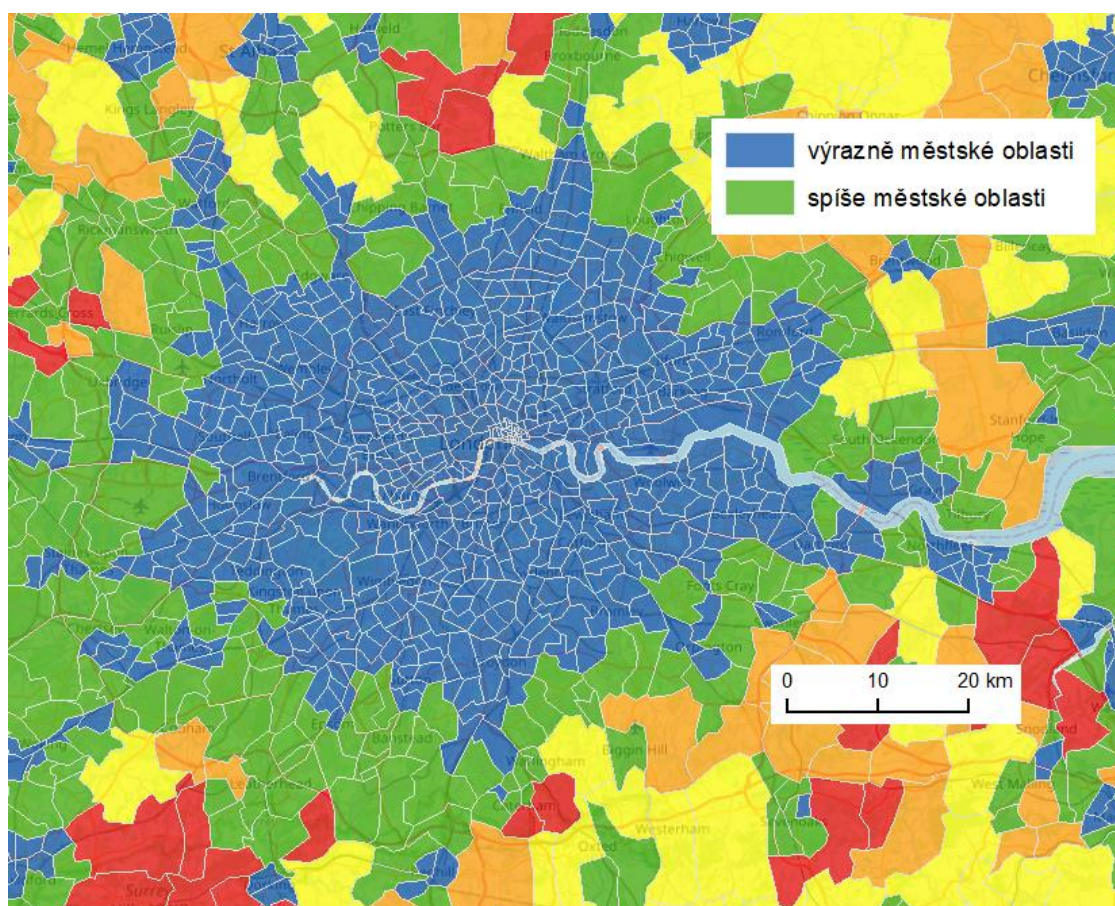
Obr. 21 Shluky podle hustoty zalidnění a počtu obyvatel

První shluk (červený) je největší, obsahuje téměř třicet tisíc LAU2. Odlišen je nejvýrazněji pomocí prvních dvou hlavních komponent, na obou dosahuje nízkých hodnot okolo nuly. Na ostatních komponentách objekty z prvního shluku nejsou výrazně rozptýleny. Nízká hodnota na první hlavní komponentě značí, že oblasti jsou venkovského charakteru, nízká hodnota na druhé komponentě, že v oblastech výrazně nepřevládají ani zemědělské oblasti, ani oblasti lesní a polopřírodní. Nazván byl venkovské oblasti s průměrným podílem zemědělských ploch. Průměrný počet obyvatel LAU2 v tomto shluku je 2 552. Zařazeno je však i deset oblastí s více než sto tisíci obyvateli, většina z nich leží ve Švédsku (např. oblast Norrköping), dalším takovým případem je město Arezzo, ležící v Itálii. Stejně jako v grafu ve spodní části obrázku 20, i v prostoru velmi často tvoří přechodnou zónu mezi venkovskými oblastmi s malým podílem zemědělských ploch (čtvrtý, fialový shluk) a venkovskými oblastmi s velkým podílem zemědělských ploch (pátý, oranžový a šestý, žlutý shluk) což lze vidět na obrázku 25, kde odděluje Karpaty a Dolnodunajskou nížinu. Často také obsahuje okrajové části horských oblastí (viz obr. 24 zejména na severu a východu Alp). Velmi zřídka oblasti tohoto shluku přímo sousedí s největšími městy zařazenými do druhého (modrého) shluku.

Druhý shluk (modrý) je nejmenší, obsahuje jen necelých šest tisíc LAU2. Podle vysokých, záporných hodnot na první hlavní komponentě lze říci, že se jedná o oblasti nejvíce městského charakteru, proto byl shluk nazván výrazně městské oblasti. Také průměrný počet obyvatel a hustota zalidnění oblastí tohoto shluku je výrazně vyšší oproti ostatním shlukům. Spadají do něj velká města (z Česka například Praha, Brno, Ostrava, České Budějovice a Kladno), která jsou v datech reprezentována jedním LAU2. Potenciálně problematickým aspektem, se kterým si využitý postup však poradil je, že velká města v některých státech jsou dělena do více menších LAU2 (typicky Velká Británie), zatímco ve většině států jedna jednotka LAU2 reprezentuje celé město. Příklady tohoto členění jednoho města do více LAU2 jsou Londýn, Paříž, Birmingham, Brusel).



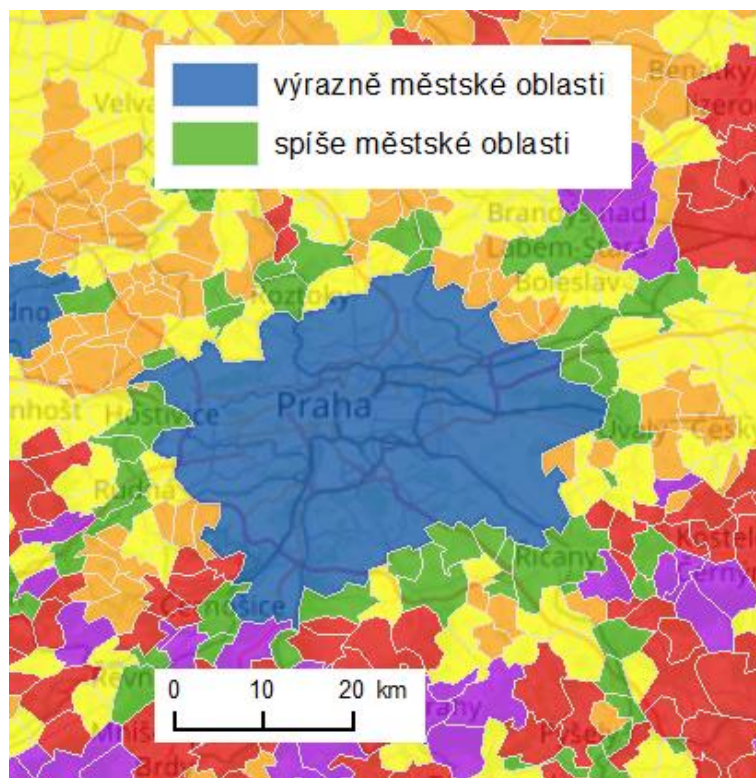
Na obrázku 22 lze vidět že i všechny dílčí části Londýna byly zařazeny do shluků městských oblastí. Dále lze pozorovat okolo centrální části zařazené do nejvíce městského shluku jakousi obalovou zónu (v této vizualizaci zelenou) spíše městských oblastí, díky které dochází k pozvolnějšímu přechodu mezi městskou a venkovskou oblastí. Do tohoto shluku bylo zařazeno i čtrnáct oblastí s nulovým počtem obyvatel v Londýně. Překvapivé je, že část oblastí z tohoto shluku dosahuje na třetí komponentě vyšších hodnot, než objekty zařazené do shluků venkovských oblastí, což značí nižší hustotu POI v těchto oblastech. Po jejich průzkumu bylo zjištěno, že se jedná nejčastěji o prostorově velmi malé LAU2 (např. části měst Aarhus či Norwich) s velmi vysokou až extrémní hustotou zalidnění, které jsou součástí velkých měst. Nebo se jedná o LAU2 reprezentující velká města, kde je malá hustota POI způsobena tím, že se jedná o velké oblasti. Příkladem takových měst je Madrid, Řím, Berlín, Bukurešť nebo Varšava.



Obr. 22 Město Londýn a jeho zařazení do shluků  
(vlastní zpracování, podklad: OSM)

Třetí shluk (zelený) obsahuje téměř 10,5 tisíce LAU2. Odlišen je pomocí první hlavní komponenty, kde objekty z tohoto shluku leží mezi městskými oblastmi a ostatními shluky venkovských oblastí. Často se jedná o oblasti přilehlé k velkým městům (obr. 22 a 23), nebo větší města, která nebyla klasifikována jako výrazně městské oblasti (v Česku například Olomouc, Zlín, Karlovy Vary; ze zahraničí například Trondheim, Lublaň). Shluk byl pojmenován spíše městské oblasti. Největšími městy podle počtu obyvatel (více než 400 000) zařazenými do této kategorie jsou Zaragoza, Málaga, Murcia a Palma de Mallorca. Vzhledem k tomu, že se jedná o španělská města je možné, že do této kategorie

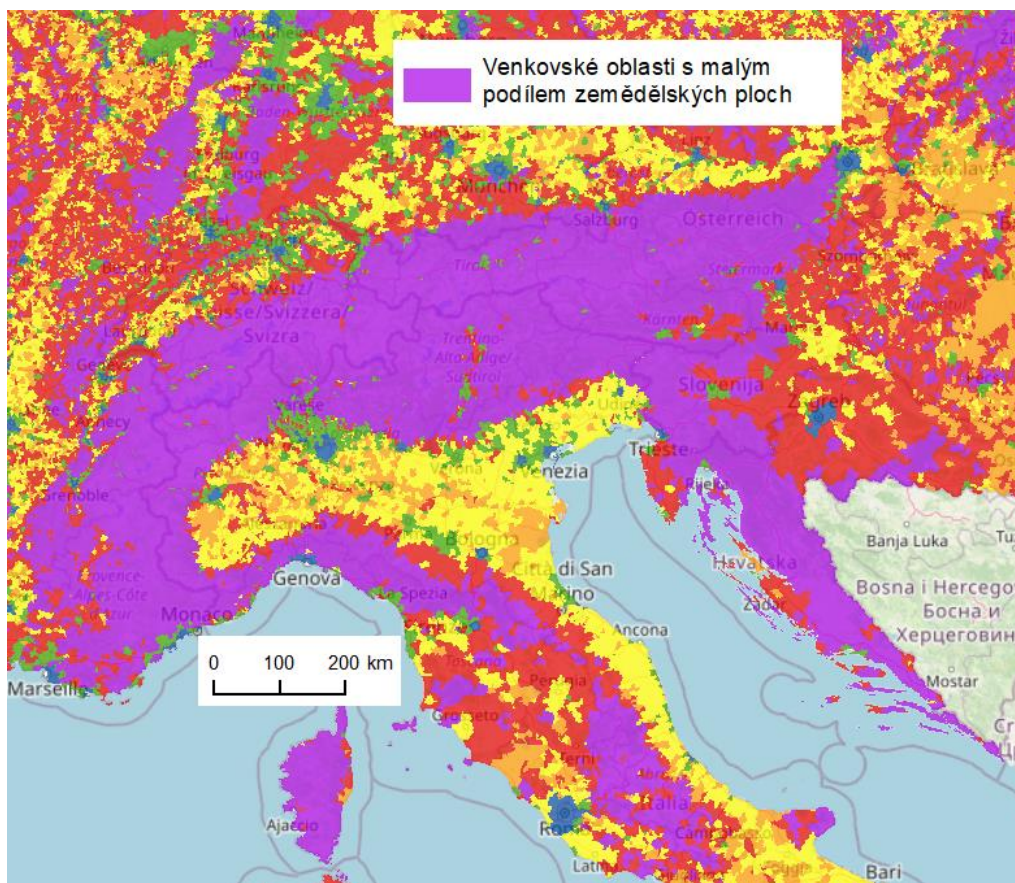
byly zařazeny kvůli neúplnosti dat OSM, anebo vzhledem k specifickým rysům administrativního členění země.



Obr. 23 Město Praha a jeho zařazení do shluků  
(vlastní zpracování, podklad: OSM)

Čtvrtý shluk (fialový) se skládá z více než 20,5 tisíc LAU2. Objekty z tohoto shluku dosahují nízkých hodnot na první hlavní komponentě (okolo nuly), což značí venkovský charakter, na druhé hlavní komponentě dosahují vysokých kladných hodnot, což značí malé zemědělské využití oblastí. Shluk byl nazván venkovské oblasti s malým podílem zemědělských ploch. Velmi často se jedná o horské oblasti jako jsou Alpy (obr. 24), Pyreneje, Krkonoše, Apeniny, Karpaty nebo Skandinávie a různé chráněné přírodní oblasti (v Česku typicky chráněné krajinné oblasti a národní parky). Nejlidnatější oblasti zařazené do tohoto shluku jsou typické tím, že jednak obsahují větší město (Uppsala, Oulu, Almería), které v některých případech dosahuje až 200 000 obyvatel, ale většina území celé LAU2 spadá do CLC kategorie lesních a polopřírodních oblastí. Podobných oblastí se vyskytuje nejvíce ve Švédsku a Finsku, kde jsou LAU2 ve srovnání s ostatními zkoumanými státy výrazně rozsáhlejší. Nejlidnatější podobnou oblastí v Česku je město Děčín, do jehož LAU2 výrazně zasahuje CHKO Labské pískovce. Do čtvrtého shluku byly zařazeny i české vojenské újezdy (Libavá, Březina, Boletice, Hradiště a již zrušený újezd Brdy).





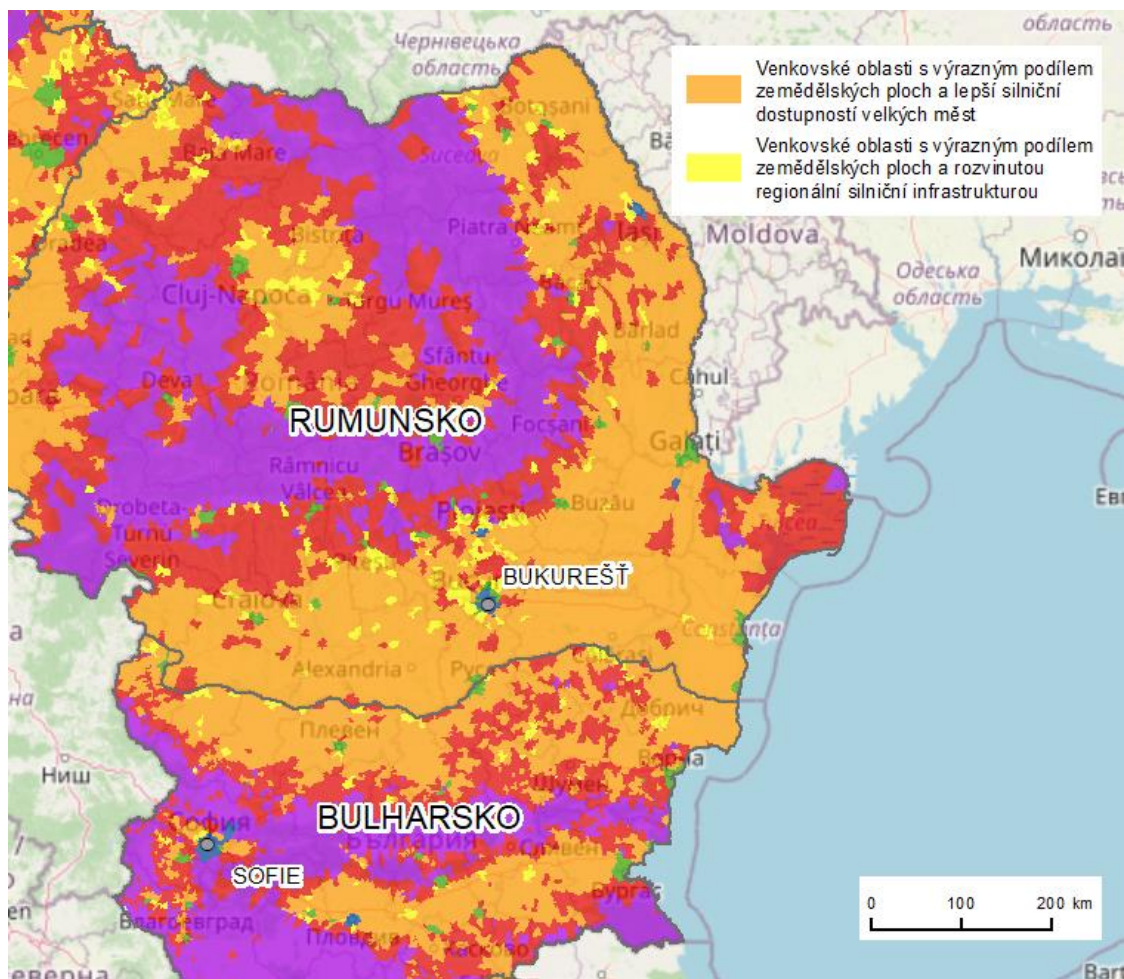
Obr. 24 Oblast Alp, zařazení do shluků (vlastní zpracování, podklad: OSM)

Pátý shluk (oranžový) obsahuje skoro 27 tisíc LAU2. Na první komponentě objekty toho shluku dosahují nízkých hodnot, proto se jedná spíše o oblasti s venkovským charakterem, na druhé hlavní komponentě však dosahují záporných hodnot, proto se jedná o oblasti s velkým zemědělským využitím. Na prvních dvou komponentách se výrazně překrývá s šestým shlukem, k jejich odlišení pomáhá čtvrtá hlavní komponenta, kde pátý shluk dosahuje záporných hodnot, což znamená výskyt a převahu silnic 1. třídy oproti silnicím 2. a 3. třídy, což lze interpretovat tak, že mají lepší silniční dopravní spojení s velkými městy. Shluk byl pojmenován venkovské oblasti s výrazným podílem zemědělských ploch a lepší silniční dostupností velkých měst.

Šestý shluk (žlutý) obsahuje asi 23,5 tisíc LAU2. Oproti pátému shluku je odlišen hodnotami na čtvrté hlavní komponentě, kde dosahuje kladných hodnot, které značí větší hustotu silnic 2. a 3. třídy, oproti silnicím 1. třídy, což značí více venkovský charakter oproti LAU2 pátého shluku. Shluk byl pojmenován venkovské oblasti s výrazným podílem zemědělských ploch a rozvinutou regionální silniční infrastrukturou.

Do pátého a šestého shluku byly zařazeny především zemědělské oblasti. Za Česko lze uvést poměrně jasně klasifikovanou oblast na Jižní Moravě (Dolnomoravský a Dyjsko–svratecký úval), oblast Polabí, Poohří, Hornomoravský úval a obecně oblast Hané. Stejně tak byla klasifikována Slezská nížina, jak velká polská část, tak i malá část v Česku. Mezi další dobře klasifikované zemědělské oblasti patří Podunajská a Východoslovenská nížina, velká část Dánska nebo Pádská nížina (kterou lze vidět na obr. 24) na severu Itálie. Oblasti zařazené do těchto dvou shluků se většinou prolínají a společně tvoří velké oblasti. V některých oblastech se však vyskytují LAU2 šestého shluku téměř bez LAU2 pátého shluku, jedná se o několik oblastí ve vnitrozemí Španělska

s přesahem do Portugalska, Velká dunajská nížina na východu Maďarska a západu Rumunska, oblasti Dolnodunajské nížiny na území Rumunska a Bulharska (obr. 25 největší oranžový shluk), Hornothrácká nížina v Bulharsku (obr. 25 menší oranžový shluk dole).

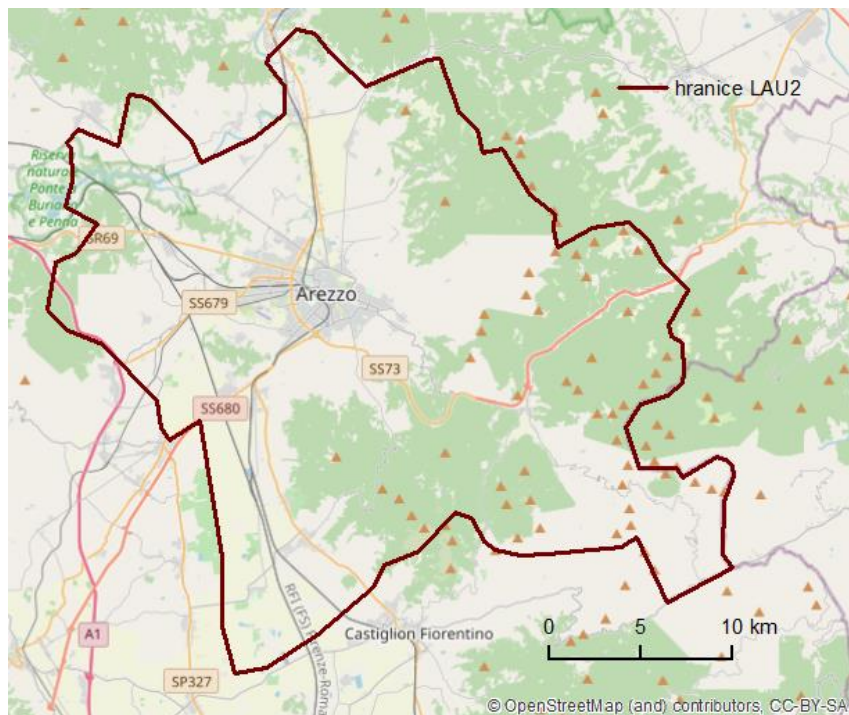


Obr. 25 Oblast Dolnodunajské a Hornothrácké nížiny, zařazení do shluků  
(vlastní zpracování, podklad: OSM)

Největším problémem v provedeném shlukování jsou různé velikosti LAU2. Pokud je město (i poměrně velké) součástí LAU2, které svými hranicemi výrazně přesahuje zastavěné území města, není v některých případech zařazeno do lépe odpovídajícího shluku. Příkladem je Italské město Arezzo (obr. 26), město samotné má téměř 100 000 obyvatel, přesto jeho LAU2 bylo zařazeno jako venkovská oblast. To může být částečně způsobeno tím, že atribut absolutního počtu obyvatel, který by mohl pomoci v zařazení podobných oblastí s velkým počtem obyvatel a poměrně malou hustotou zalidnění, významně nepřispívá do žádné z komponent použitých ke shlukování. Výrazně ovlivňuje až 7. hlavní komponentu, která už nevysvětluje dostatek rozptylu původních dat, aby výrazně ovlivnila shlukování při zvoleném počtu shluků. Otázkou však je, zda je správné považovat charakter celého území takto velkého LAU2 oproti zastavěnému území města za městský. Z tohoto pohledu se jinak velmi detailní jednotky LAU2 zdají nedostatečné. Možným řešením by bylo použít pravidelnou síť např. population grid jakoto tomu je v případě klasifikace z oficiálních zdrojů Evropské unie (Dijkstra a Poelman, 2014), který by bylo možné obohatit stejným způsobem o stejné další

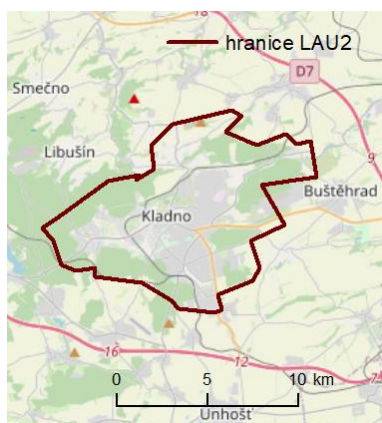


atributy jako LAU2 v této práci, nebo kromě hranic LAU2 použít i hranice měst podle zástavby. Dalším řešením by mohlo být více shluků, pokud by některý z nich obsahoval převážně podobné oblasti, což však při automatickém shlukování nelze zaručit. Možným řešením by mohlo být i použití absolutních hodnot atributů místo atributů přepočtených na plochu, to by ale mohlo vést naopak k nadhodnocení velkých LAU2.

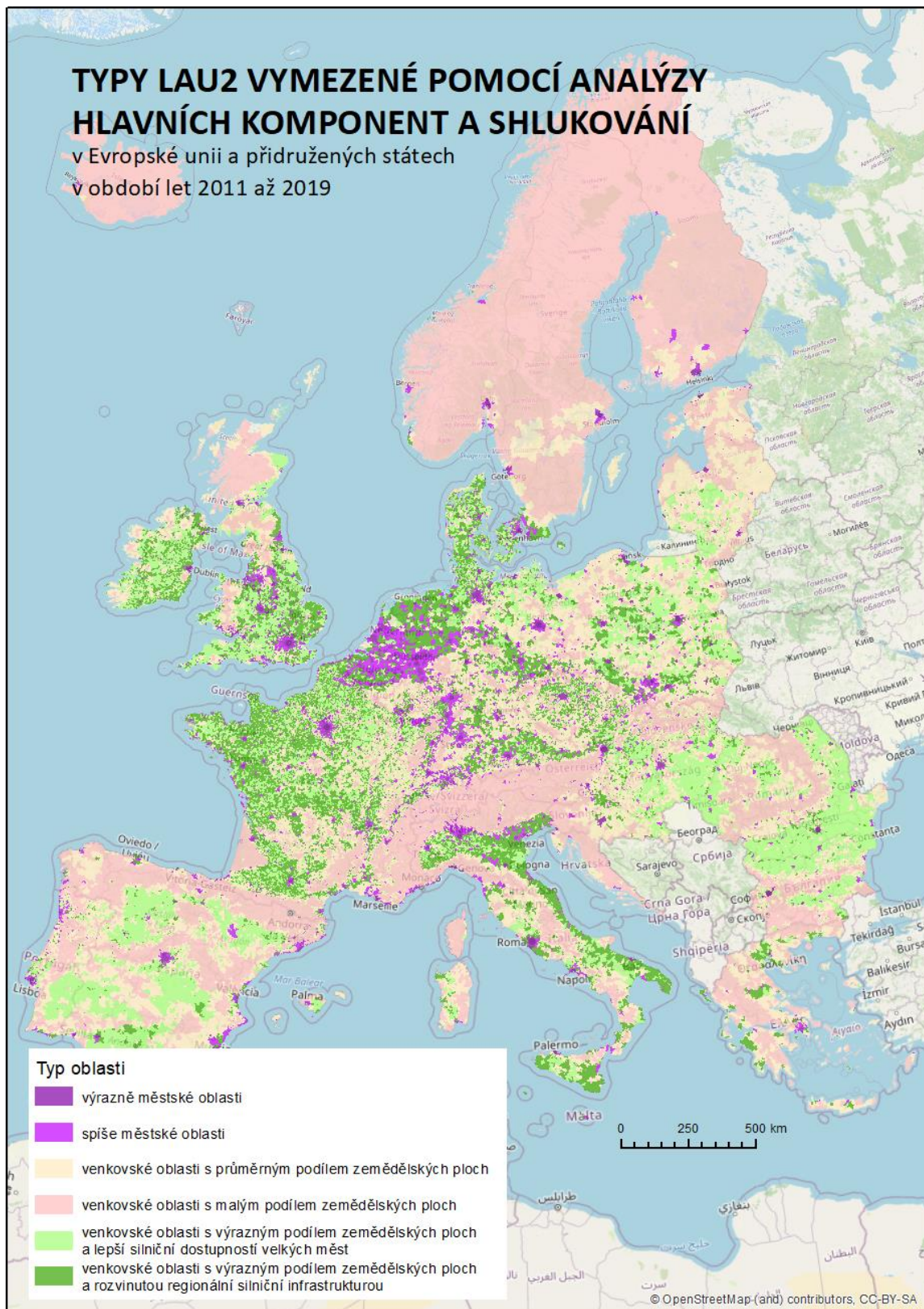


Obr. 26 Město Arezzo a hranice LAU2 (vlastní zpracování, podklad: OSM)

Opačného výsledku než u předchozího příkladu je dosaženo u LAU2 obsahujících téměř pouze větší město samotné. Příkladem je město Kladno (obr. 27) s necelými 70 tisíci obyvatel, které bylo zařazeno do shluku výrazně městských oblastí. Oproti městu Arezzo je na první pohled patrná výrazně menší oblast mimo vlastní město zahrnutá do plochy LAU2. Díky tomu je také zařazeno do více městské kategorie oproti větším českým městům jako jsou Zlín nebo Olomouc, která jsou zařazena do kategorie spíše městských oblastí. Nicméně je potřeba zdůraznit, že pro více než sto tisíc jednotek LAU2 se nelze plně vyhnout podobným individuálním případům.



Obr. 27 Město Kladno a hranice LAU2 (vlastní zpracování, podklad: OSM)



Obr. 28 Zařazení LAU2 do shluků pro celou klasifikovanou oblast  
(vlastní zpracování, podklad: OSM)

## 9 AGREGACE NA NUTS2

V této části práce byly shluky z předchozí kapitoly pomocí GIS operací agregovány z jednotek LAU2 na vyšší statistické jednotky NUTS2 (v ČR regiony soudržnosti). Data v této agregované podobě umožní jejich srovnání i s jinými daty Eurostatu. Pomocí tohoto srovnání s dalšími ukazateli je možné zjistit souvislosti mezi příslušností do městského/venkovského prostoru a různými jevy a dále také ověřit správnost vytvořené klasifikace pomocí srovnání s indikátory běžně používanými k identifikaci městských/venkovských oblastí, které nebylo možné získat pro statistické jednotky LAU2. Poměrně velké oblasti NUTS2 byly zvoleny z důvodu nedostatku, neaktuálnosti a neúplnosti dat za nižší statistické jednotky (zejména NUTS3).

### 9.1 Proces agregace

Agregace dat byla provedena pomocí programu ArcMap z ArcGIS Desktop. Nejprve byla stažena polygonová vrstva NUTS2 z webu Eurostatu, vzhledem k původní vrstvě LAU2 byla zvolena verze NUTS 2013. Odstraněny byly oblasti mimo rozsah původní vrstvy LAU2. Pomocí operace *Intersect* byla vytvořena kombinace vrstvy NUTS2, obsahující identifikátor jednotlivých oblastí a vrstvy LAU2 obsahující informaci o zařazení do shluku. Z této nové vrstvy byly exportovány do samostatných vrstev jednotlivé shluky. Vzniklo tak 6 vrstev, jedna pro každý původní shluk. Pomocí funkce *summarize* byly vytvořeny tabulky s informací o plochách každé NUTS2 patřící do jednotlivých shluků. Tyto hodnoty byly pomocí funkce *join* následně připojeny k původní vrstvě NUTS2 a převedeny do příslušných atributů. Z těchto atributů pak bylo možné vypočítat procentuální zastoupení plochy jednotlivých NUTS2 připadající jednotlivým shlukům.

Mimo agregace dat z LAU2 na NUTS2 bylo agregováno i šest původních shluků do třech nových atributů (tab. 7). Výsledkem agregace je nová vrstva NUTS2, se třemi atributy vyjadřujícími podíl jednotlivých typů oblastí v ploše celého NUTS2.

Tab. 7 Agregace shluků do nových atributů

Původní shluk	Nový atribut po agregaci
Výrazně městské oblasti	Podíl městských oblastí
Spiše městské oblasti	
Venkovské oblasti s průměrným podílem zemědělských ploch	Podíl venkovských oblastí
Venkovské oblasti s malým podílem zemědělských ploch	
Venkovské oblasti s výrazným podílem zemědělských ploch a lepší silniční dostupností velkých měst	Podíl venkovských zemědělských oblastí
Venkovské oblasti s výrazným podílem zemědělských ploch a rozvinutou regionální silniční infrastrukturou	

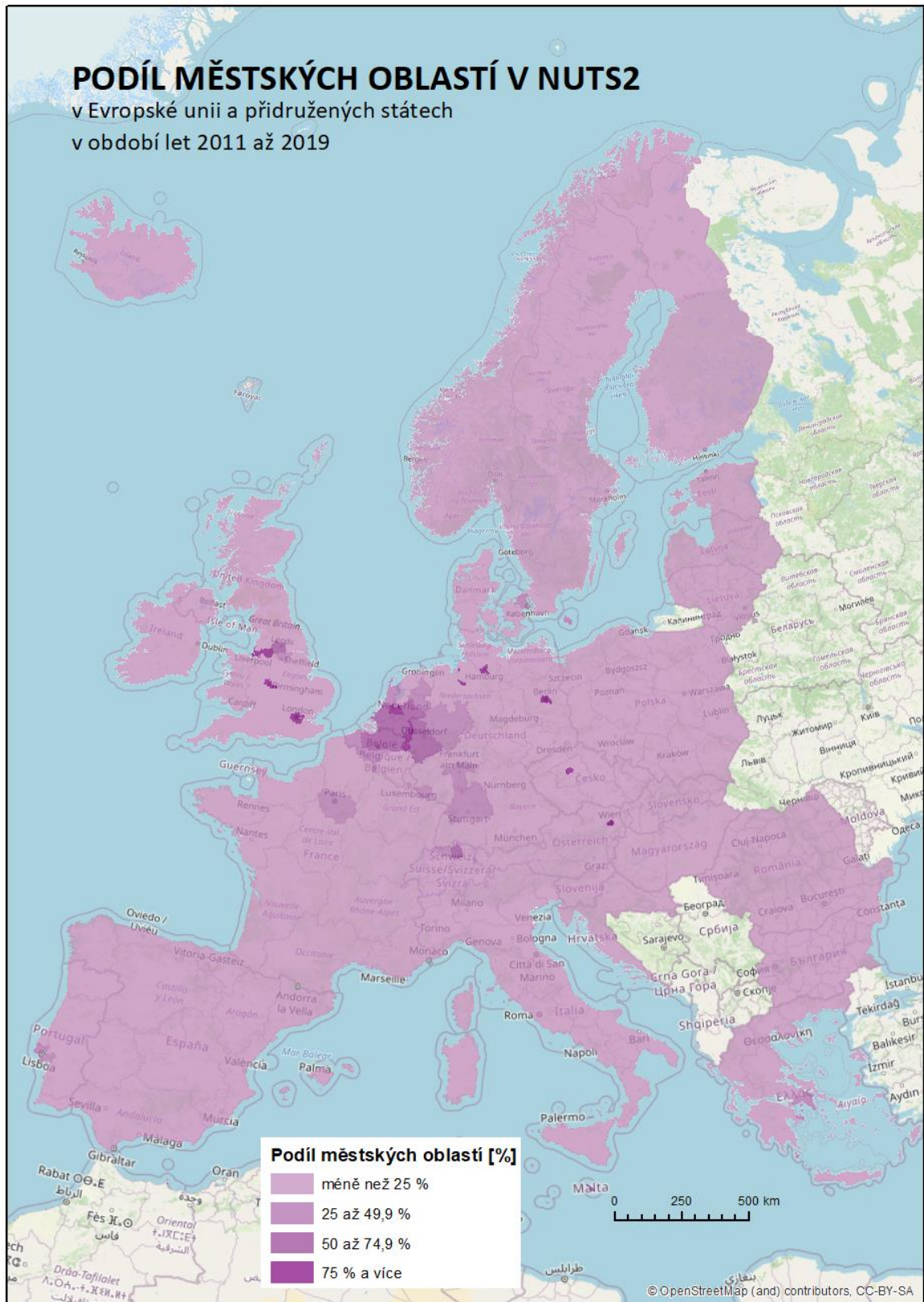
Nevýhoda agregace na NUTS2 oproti menším oblastem je patrná nejvíce u hodnot atributu podílu městských oblastí (obr. 29). Poměrně nízkých hodnot dosahují velká města, která nemají vlastní jednotku NUTS2. Příkladem je oblast Comunidad de Madrid ve Španělsku, ve které leží Madrid a jejíž podíl městských oblastí dosahuje jen asi 22 %. Oproti tomu Praha, Berlín, Vídeň nebo Hamburg, jejichž oblast NUTS2 je totožná s LAU2

dosahují 100 % v podílu městských oblastí. Londýn skládající se z několika NUTS2 byl zařazen správně. Největší shluk NUTS2 s vysokými podíly městských oblastí se nachází na jihu Nizozemska, severu Belgie a východu Německa.

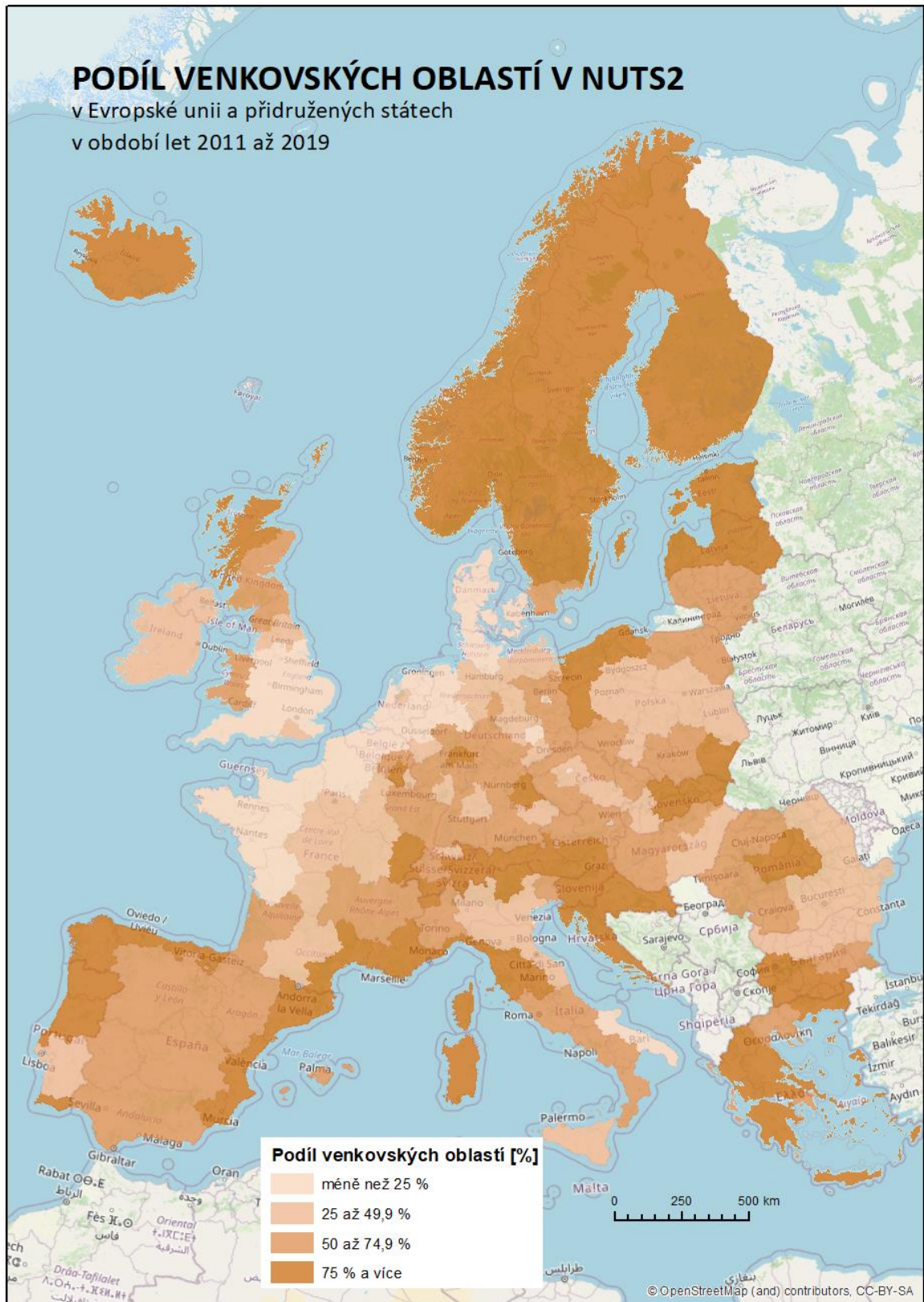
Druhým vytvořeným atributem je podíl venkovských oblastí (obr. 30), který vznikl agregací venkovských LAU2 s malým a průměrným podílem zemědělských ploch. Vysokých hodnot v tomto atributu dosahují NUTS2 obsahující hory (oblast Alp, sever Velké Británie) téměř celá skandinávie, Island, pobřežní oblasti Španělska a Portugalska.

Poslední atribut vznikl agregací venkovských LAU2 s vysokým podílem zemědělských ploch, proto byl nazván podíl venkovských zemědělských oblastí (obr. 31). Stejně jako u LAU2 lze poměrně zřetelně identifikovat (jako oblasti s vysokými hodnotami) Pádskou a Dolnodunajskou nížinu. Velké shluky oblastí s vysokými hodnotami se nacházejí také na jihu Velké Británie, v Irsku, Dánsku nebo na západě Francie.



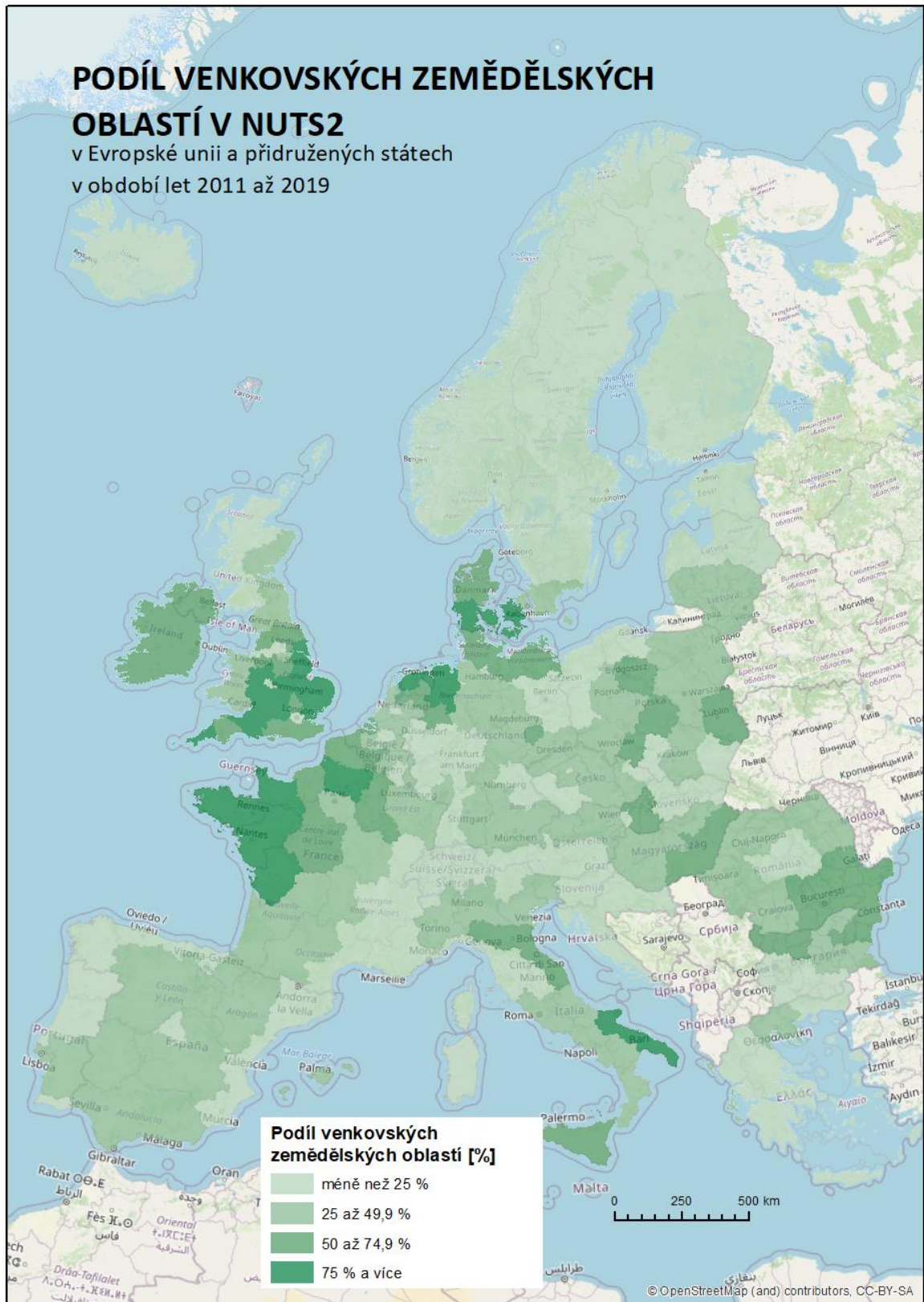


Obr. 29 Podíl městských oblastí v NUTS2  
(vlastní zpracování, podklad: OSM)



Obr. 30 Podíl venkovských oblastí v NUTS2  
(vlastní zpracování, podklad: OSM)





Obr. 31 Podíl venkovských zemědělských oblastí v NUTS2  
(vlastní zpracování, podklad: OSM)

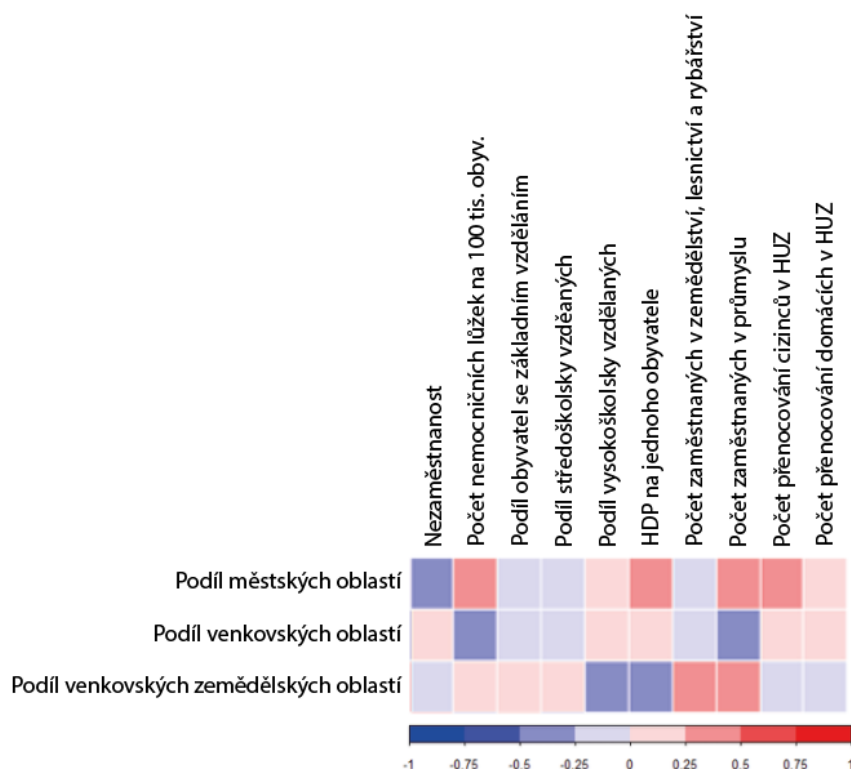
## 9.2 Srovnání s daty Eurostatu

Data pro tvorbu nových atributů byla stažena z webu Eurostatu. Vybrané indikátory byly pomocí funkce *join* přidány k vrstvě NUTS2 z předchozí podkapitoly. Všechny hodnoty jsou aktuální k roku 2017, stejně jako většina původních údajů o počtu obyvatel. Jedná se o atributy:

- Nezaměstnanost (%)
- Počet nemocničních lůžek na 100 tis. obyvatel
- Podíl osob se základním vzděláním (%)
- Podíl středoškolsky vzdělaných (%)
- Podíl vysokoškolsky vzdělaných (%)
- HDP (eur/osoba)
- Počet osob zaměstnaných v zemědělství, lesnictví a rybářství (NACE kategorie A)
- Počet osob zaměstnaných v průmyslu (NACE kategorie B, C, D, E)
- Počet přenocování cizinců v hromadných ubytovacích zařízeních
- Počet přenocování domácích obyvatel v hromadných ubytovacích zařízeních

Použitá statistická klasifikace ekonomických činností v Evropském společenství (NACE Rev. 2) řadí ekonomické činnosti do 21 základních sekcí označených velkým písmenem, které se dále dělí do detailnějších subkategorií (NACE Rev. 2, 2008). Kategorie A byla použita samostatně, kategorie B, C, D, E byly sloučeny jako průmysl (bez stavebnictví).

Do použité kategorie hromadných ubytovacích zařízení (HUZ) jsou zahrnuty všechny hotely a podobná zařízení, prázdninové a podobné krátkodobé ubytování, kempy, autokempy a kempy pro přívěsy.



Obr. 32 Korelace mezi agregovanými daty a daty Eurostatu

Pro srovnání atributů a nalezení více obecných poznatků a závislostí byly vypočteny korelace mezi hodnotami atributů agregovaných na NUTS2 a atributů získaných z Eurostatu za tytéž prostorové jednotky (obr. 32). Následující text se zabývá charakteristikou vzájemných souvislostí.

### **Nezaměstnanost**

Tento indikátor nejvýrazněji a to záporně koreluje s podílem městských oblastí, což značí předpokládanou nižší nezaměstnanost v městských oblastech oproti ostatním oblastem. S podílem venkovských zemědělských oblastí koreluje lehce záporně, oproti předpokladu vyšší nezaměstnanosti v těchto oblastech především z důvodu sezónnosti zemědělské práce.

### **Počet nemocničních lůžek**

V tomto případě korelace s podíly městských a venkovských oblastí nepřináší nic nečekaného, protože nemocnice se běžně vyskytují pouze v městech. Zajímavější je však rozdíl v tom jak koreluje s podílem venkovských oblastí a venkovských zemědělských oblastí. Zatímco s podílem venkovských oblastí koreluje záporně, s podílem venkovských zemědělských oblastí koreluje mírně kladně, což značí výrazně méně nemocničních lůžek ve venkovských oblastech oproti zemědělským oblastem.

### **Podíl obyvatel se základním vzděláním**

Tento indikátor výrazně nekoreluje s žádným typem oblastí. Očekávána byla podobná korelace mezi tímto indikátorem a podílem venkovských i venkovských zemědělských oblastí. Podobné záporné hodnoty korelace bylo dosaženo však s podílem venkovských a městských oblastí a naopak kladné hodnoty korelace bylo dosaženo s podílem venkovských zemědělských oblastí. Možným vysvětlením je, že v oblastech s velkým podílem zemědělství je potřeba velké množství nekvalifikované pracovní síly, zatímco vysokoškoláci často z takových oblastí odcházejí.

### **Podíl středoškolsky vzdělaných**

Tento indikátor výrazně nekoreluje s žádným typem oblastí. Hodnoty korelací jsou velmi podobné korelacím předchozího indikátoru.

### **Podíl vysokoškolsky vzdělaných**

Záporně koreluje s podílem venkovských zemědělských oblastí. Očekávána byla vyšší kladná korelace s podílem městských oblastí a záporná korelace s podílem venkovských oblastí, oba tyto podíly však korelují pouze mírně kladně.

### **HDP na jednoho obyvatele**

Kladně koreluje s podílem městských oblastí. Obecně je běžné, že vyšší HDP je ve velkých městech a metropolitních oblastech jednotlivých států. Naopak záporně koreluje s podílem venkovských zemědělských oblastí.

### **Počet osob zaměstnaných v zemědělství, lesnictví a rybářství**

Kladně koreluje s podílem venkovských zemědělských oblastí, očekávána však byla vyšší hodnota korelace, které by mohlo být dosaženo při porovnávání menších oblastí než NUTS2.

### **Počet osob zaměstnaných v průmyslu**

Podle očekávání kladně koreluje s podílem městských oblastí, protože průmysl je běžně soustředěn do průmyslových zón na okrajích měst, které se vyznačují dobrou dopravní dostupností jak pro zaměstnance tak pro dovoz/odvoz materiálů. Stejně tak byla očekávána záporná korelace s podílem venkovských oblastí. Překvapivá je však kladná korelace s podílem venkovských zemědělských oblastí. Důvodem může být umístění průmyslových závodů, které zpracovávají suroviny ze zemědělství přímo v zemědělských oblastech tak, aby se snížila vzdálenost pro dopravu surovin.

### **Počet přenocování cizinců v HUZ**

Nejvíce kladně koreluje s podílem městských oblastí, z čehož vyplývá že cizinci nejčastěji navštěvují města. Naopak lehce záporně koreluje s podílem venkovských zemědělských oblastí.

### **Počet přenocování domácích obyvatel v HUZ**

Tento indikátor výrazně nekoreluje s žádným typem oblastí. Oproti počtu přenocování cizinců výrazně nekoreluje ani s podílem městských oblastí, což značí, že domácí cestovní ruch není tolik zaměřen na města.

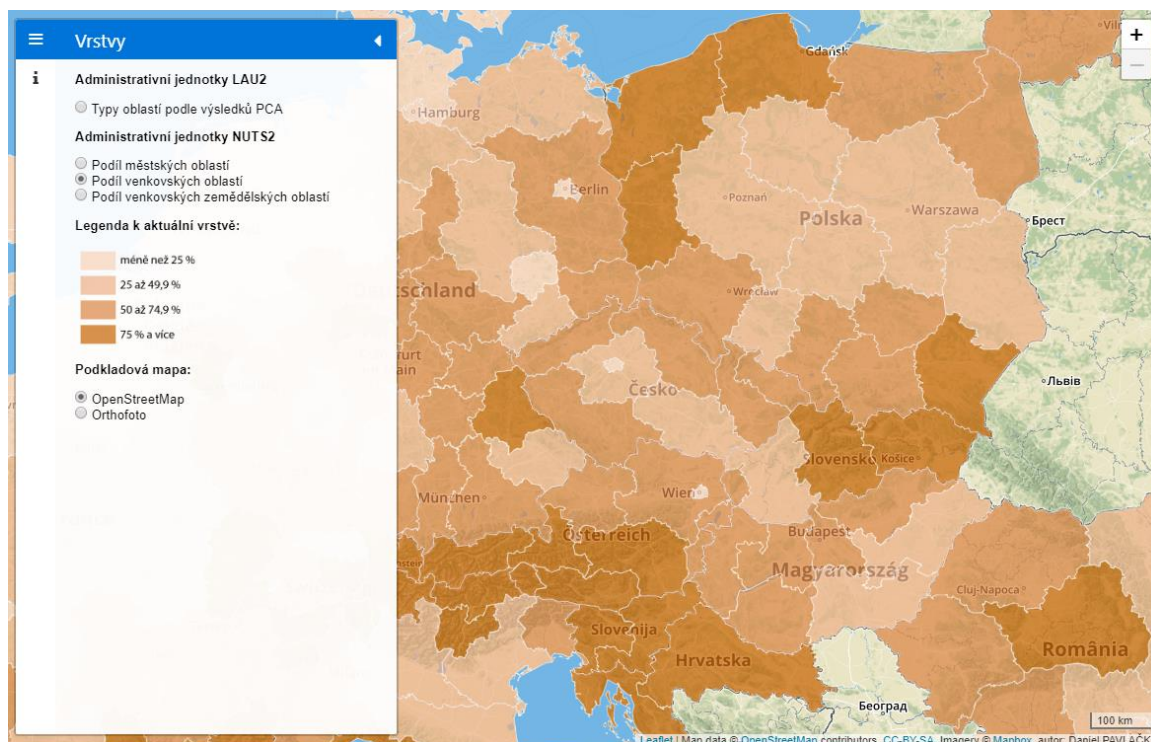
## 10 WEBOVÁ MAPOVÁ APLIKACE

Pro zobrazení výsledných vrstev byla zvolena webová mapová aplikace, protože běžné tištěné mapy pro území EU zobrazující detailní administrativní jednotky LAU2 by musely být příliš velkého formátu. Pro tvorbu webové mapové aplikace byla využita open source javascriptová mapová knihovna Leaflet. Mapová aplikace je responzivní.

Pro načítání dat jsou využity dva způsoby. Vrstvy NUTS2, které nejsou příliš velké jsou načítány jako lokální GeoJSON pomocí pluginu leaflet-ajax a javascriptové knihovny jQuery. Pro velké vrstvy, v tomto případě LAU2, není způsob pomocí jQuery vhodný, protože se vrstva dlouho načítá a špatně se s ní v aplikaci manipuluje. Z tohoto důvodu je vrstva načítána pomocí Mapbox Studia jako tileset.

Sidebar (postranní panel) byl vytvořen pomocí pluginu sidebar-v2 dostupného na GitHubu (Leaflet-sidebar-v2, 2020). Na sidebaru byly vytvořeny dvě záložky. Záložka Vrstvy, která je defaultně zapnutá při spuštění aplikace obsahuje přepínací vrstvy, legendu a ovládání podkladových map. Druhá záložka Info obsahuje obecné informace o aplikaci. Ovládání vrstev bylo napsáno jako funkce v jazyku javascript, tak aby mohla být najednou zobrazena jen jedna vrstva. Přepínání vrstev současně ovládá i zobrazenou legendu tak, aby se vztahovala pouze k aktuálně zobrazené vrstvě. Jednotlivé části legendy jsou vkládány jako PNG obrázek, dopředu vytvořené ve vektorovém grafickém editoru Adobe Illustrator.

Mapová aplikace je dostupná z webových stránek diplomové práce, které lze dohledat na odkaze: <http://www.geoinformatics.upol.cz/studium/diplomove-prace>.



Obr. 33 Ukázka webové mapové aplikace  
(vlastní zpracování, podklad: OSM)



## 11 VÝSLEDKY

V této kapitole budou shrnuty výsledky a výstupy práce. Hlavními výstupy práce jsou vrstva LAU2 obsahující nové atributy, hodnoty použitých hlavních komponent a informací o zařazení do shluku jednotlivých oblastí, dále vrstva NUTS2 agregující výsledky PCA a shlukování do třech atributů s plošným podílem ploch městských, venkovských a venkovských zemědělských. Pro jejich prezentaci byla ve finální části práce vytvořena webová mapová aplikace.

První a časově velmi náročná část práce byla zaměřena na obohacení původní datové sady o další atributy. K tomuto účelu byla zvolena data CORINE Land Cover a OSM. Vytvoření této datové sady bylo nutné pro další analýzy, zaměřené na problematiku typologie městských/venkovských oblastí. Výsledná datová sada této části práce (přehled viz tabulka 2 v kapitole 6) obsahuje pět nových atributů z CORINE Land Cover, dvacetpět nových atributů z OSM a přepočítání všech atributů na plochu.

V další části práce byly vytvořeny tři základní klasifikace městských a venkovských oblastí založené na počtu obyvatel, hustotě zalidnění a jejich kombinaci, tedy bez použití nově vytvořených atributů. Jako hraniční hodnoty pro odlišení města a venkova byly zvoleny hodnoty vycházející z již existujících klasifikací či dokumentů. Vzhledem ke specifickým rysům administrativního členění jednotlivých států napříč celým zkoumaným územím však nejsou dostatečně přesné a použitím jedné univerzální hodnoty pro celé území jsou zkresleny.

Před vytvořením nové typologie městských/venkovských oblastí (s použitím nových atributů) byla data analyzována pomocí boxplotů a korelačních matic. Pomocí boxplotů bylo určeno, že se v attributech jak přepočítaných tak i nepřepočítaných na plochu vyskytují outliery. Z tohoto důvodu byl zvolen Spearmanův korelační koeficient pro ohodnocení korelací před samotnou analýzou hlavních komponent. Rovněž byla provedena i analýza hlavních komponent se všemi atributy přepočítanými na plochu, na základě které bylo rozhodnuto redukovat počet atributů. Odebrány byly atributy s příliš velkým výskytem nulových hodnot, atributy málo relevantní vzhledem ke smyslu analýzy a naopak přidán byl atribut počtu obyvatel (přehled použitých atributů viz tabulka 3). Aby byl zajištěn vznik hlavní komponenty vhodné pro odlišení městských a venkovských oblastí byla zachována korelující skupina atributů vhodných k tomuto účelu.

Pomocí indexového grafu úpatí vlastních čísel (Scree plot) byly vybrány „užitečné“ čtyři hlavní komponenty s hodnotou vlastních čísel vyšší než jedna. Graf komponentních vah (zátěží) pro první a druhou hlavní komponentu byl využit pro ohodnocení vlivu a důležitosti atributů na tyto komponenty a jejich vzájemného vztahu. Pomocí dvojnásobného grafu byly vykresleny záznamy v prostoru prvních dvou hlavních komponent. Pomocí matice záteží a grafů znázorňujících příspěvky jednotlivých atributů do vybraných hlavních komponent byl interpretován význam hlavních komponent, což následně vedlo k jejich pojmenování.

- 1. hlavní komponenta – Charakter oblasti
- 2. hlavní komponenta – Podíl zemědělských ploch oproti přírodním oblastem
- 3. hlavní komponenta – Hustota bodů zájmu
- 4. hlavní komponenta – Hustota silniční sítě podle třídy

Hlavním výsledkem práce je typologie oblastí vytvořená shlukováním výsledku analýzy hlavních komponent. Pro shlukování byl použit algoritmus CLARA, rozšiřující algoritmus PAM. Důležitou částí shlukové analýzy bylo určení vhodného počtu shluků, k čemuž byly využity hodnoty siluet (tabulka 5). Vytvořeno bylo šest shluků, které byly interpretovány

a pojmenovány (kapitola 8.1) pomocí vykreslení shluků v prostoru jednotlivých dvojic hlavních komponent, vizuální analýzy mapy zobrazující jednotky LAU2 zařazené do shluků a grafu zobrazující shluky podle počtu obyvatel a hustoty zalidnění.

- Výrazně městské oblasti
- Spíše městské oblasti
- Venkovské oblasti s malým podílem zemědělských ploch
- Venkovské oblasti s průměrným podílem zemědělských ploch
- Venkovské oblasti s výrazným podílem zemědělských ploch a lepší silniční dostupností velkých měst
- Venkovské oblasti s výrazným podílem zemědělských ploch a rozvinutou regionální silniční infrastrukturou

Jednotlivé shluky byly popsány a charakterizovány. Uvedeny byly typické oblasti zařazené do těchto shluků včetně konkrétních příkladů a jejich vizualizací. Popsány byly i skutečnosti způsobující potenciálně špatné zařazení oblastí a některé oblasti jejichž zařazení je sporné. Celá výsledná vrstva byla zařazena do webové mapové aplikace, aby bylo umožněno její detailní prohlížení.

Agregací těchto shluků do prostorových jednotek NUTS2 bylo umožněno srovnání s daty Eurostatu. Vytvořeny byly tři nové atributy s informací o procentuálním zastoupení městských, venkovských a venkovských zemědělských oblastí v ploše NUTS2. Pro srovnání s nimi byla vytvořena vrstva NUTS2 s deseti atributy získanými z Eurostatu. Data byla srovnávána pomocí korelací. Korelace nebyly výrazně vysoké, pravděpodobně kvůli přílišné generalizaci informace, plynoucí z použití velkých prostorových jednotek NUTS2. Z tohoto srovnání se nedá s jistotou posuzovat správnost vytvořené klasifikace. Některé předpokládané vztahy, mezi podílem městských, venkovských a venkovských zemědělských byly potvrzeny, některé však byly nečekané a dokonce i opačné.

Poslední částí práce bylo vytvoření webové mapové aplikace, zobrazující výsledné vrstvy. Zahrnuty do ní, mimo již zmíněné vrstvy shluků LAU2, byly také vrstvy procentuálního zastoupení městských, venkovských a venkovských zemědělských oblastí za jednotky NUTS2.

## 12 DISKUZE

Vzhledem k prostorovému rozsahu a detailnosti dat vyplývající ze zadání práce byla prvním problémem samotná volba možných a použitelných dat. Z oficiálních a otevřených zdrojů bylo možné použít pouze data CORINE Land Cover. Data z oficiálních administrativních zdrojů dat jako jsou Eurostat nebo databáze OECD nebylo možné použít, protože nejsou dostupná za jednotky LAU2 (s výjimkou využitého počtu obyvatel). Z tohoto důvodu bylo nutné využít data OSM i přes problémy které z toho plynou. Data OSM může vytvářet kdokoli, proto není kvalita a pokrytí daty stejná v celém území, což mohlo ovlivnit výsledky analýz. Možností jak alespoň určit stupeň úplnosti dat by bylo získat pro srovnání ekvivalentní data z nějakého oficiálního zdroje. Pokud by však bylo možné taková data použít, nebylo by potom nutné použít data OSM. Možná polohová (ne)přesnost dat (snad kromě opravdu hrubých nepřesností) výsledky práce s největší pravděpodobností ovlivnit nemohla, protože data byla agregována pro celé jednotky LAU2.

Dalším problematickým aspektem práce spojeným s daty je využití jednotek NUTS2 pro agregaci vytvořené typologie jednotek LAU2 a jejich následné srovnání s dalšími daty. Vhodnější a také uvažovanou možností byla agregace na jednotky NUTS3, za které Eurostat sice poskytuje data, často však nejsou úplné, aktuální pro požadovaný rok a je jich velice málo.

Dalším problémem byla nízká průměrná hodnota siluet při shlukové analýze. Minimální přijatelnou hodnotou dle literatury je 0,51, při testování mnoha různých možností však nebylo dosaženo vyšší hodnoty než 0,417. Tato skutečnost plyne z charakteru dat, které jednoduše nevytvářejí jednoznačné, přirozené shluky, navíc po detailním posouzení a analýze nejsou výsledky shlukování v rozporu s již existujícími vymezeními.

Potenciálním problémem (který nastal u klasifikace podle počtu obyvatel) je rozdělení jednoho města (Londýn, Birmingham, Brusel aj.) do více jednotek LAU2. Použitý postup s analýzou hlavních komponent však tento nedostatek eliminoval. Problém, se kterým si použitý postup ale neporadil, jsou různé velikosti jednotek LAU2 oproti zastavěnému území. I velká města (stovky tisíc obyvatel), která jsou součástí prostorově/administrativně rozsáhlých LAU2, jejichž zástavba zabírá pouze malou část celé LAU2, jsou v některých případech zařazeny jako venkovské oblasti a naopak mnohem menší obce, jejichž LAU2 však výrazně nepřesahují zastavěné oblasti, jsou klasifikovány jako městské oblasti. Možným řešením tohoto problému a zároveň možností, jak by bylo možné v práci pokračovat je stejným způsobem obohatit data Population grid – informace o počtu obyvatel v pravidelné mřížce.



## ZÁVĚR

Prvním cílem práce bylo obohatit datovou sadu jednotek LAU2 za území Evropské unie a přidružených států, které bylo vymezeno rozsahem původní vrstvy. Druhým cílem bylo provést nad těmito daty socioekonomické analýzy a především vytvořit typologii městských/venkovských oblastí v detailu jednotek LAU2 a pro celé území.

Pro obohacení datové sady byla využita data z otevřených zdrojů, konkrétně Open Street Map a CORINE Land Cover. Vytvořeno bylo třicet nových atributů. Na základě těchto nových atributů byla vytvořena typologie venkovských a městských oblastí, k čemuž byla použita analýza hlavních komponent. Použito bylo šestnáct atributů relevantních pro odlišení městských a venkovských oblastí. Vybrané čtyři užitečné hlavní komponenty byly analyzovány, interpretovány a pojmenovány. Samotná typologie byla vytvořena pomocí shlukové analýzy s použitím těchto čtyř hlavních komponent. Konkrétně byl použit nehirarchický shlukovací algoritmus CLARA. Pomocí statistiky tzv. siluet byl definován vhodný počet šesti shluků. Jednotlivé shluky, které představují typy oblastí (jednotek LAU2) byly pomocí grafů a vizuální analýzy mapy interpretovány a pojmenovány. Jednotlivé kategorie byly popsány, vymezeny byly typické oblasti zařazené do jednotlivých shluků včetně konkrétních příkladů. Uvedeny byly i některé problematické aspekty a potenciální chyby v této klasifikaci opět včetně konkrétních příkladů.

Následně byla vytvořená typologie jednotek LAU2 agregována do větších jednotek NUTS2, aby výsledek mohl být srovnán s dalšími daty Eurostatu. Kromě agregace do větších ploch, bylo původních šest shluků agregováno do třech typů oblastí NUTS2. Pro srovnání bylo použito korelací mezi deseti atributy z Eurostatu a třemi novými agregovanými atributy. Pro prezentaci výsledných vrstev byla vytvořena webová mapová aplikace.

Vytvořená typologie je jedinečná svým prostorovým rozsahem a detailem i použitými daty a definovanými typy oblastí, protože tematicky podobné typologie podobného prostorového rozsahu a detailu většinou využívají pouze počet obyvatel a hustotu zalidnění, větší množství atributů bývá využito většinou pouze pro menší území (jeden stát).

# POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE

ARCGIS DESKTOP: *Documentation* [online]. 2020 [cit. 2020-02-21]. Dostupné z: <https://desktop.arcgis.com/>

BIBBY, Peter a John SHEPHERD. *Developing a New Classification of Urban and Rural Areas for Policy Purposes – the Methodology* [online]. 2005 [cit. 2020-02-07]. Dostupné z: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/137655/rural-urban-definition-methodology-technical.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/137655/rural-urban-definition-methodology-technical.pdf)

BURIAN, Jaroslav, Vít PÁSZTO a Pavel TUČEK. *Geoinformatika při analýzách rurálního a urbánního prostoru*. Olomouc: Univerzita Palackého v Olomouci, 2013. ISBN isbn978-80-244-3783-5.

COPERNICUS LAND MONITORING SERVICE [online]. 2019 [cit. 2019-02-15]. Dostupné z: <https://land.copernicus.eu/>

ČSÚ. *Varianty vymezení VENKOVA a jejich zobrazení ve statistických ukazatelích - 2000 - 2006* [online]. 2008 [cit. 2020-02-21]. Dostupné z: <https://www.czso.cz/csu/czso/varianty-vymezeni-venkova-a-jejich-zobrazeni-ve-statistickych-ukazatelich-v-letech-2000-az-2006-n-o08us2n3fa>

DAX, Thomas. Defining rural areas – International comparisons and the OECD indicators 1. *Rural Society* [online]. 2014, 6(3), 3-18 [cit. 2020-02-21]. DOI: 10.5172/rsj.6.3.3. ISSN 1037-1656. Dostupné z: <http://www.tandfonline.com/doi/abs/10.5172/rsj.6.3.3>

DEFRA. *Rural and Urban Area Classification 2004: An Introductory Guide* [online]. 2004 [cit. 2020-02-07]. Dostupné z: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/239082/2001-rural-urban-definition-methodology-intro.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/239082/2001-rural-urban-definition-methodology-intro.pdf)

DIJKSTRA, Lewis a Hugo POELMAN. *A harmonised definition of cities and rural areas: the new degree of urbanisation* [online]. 2014 [cit. 2019-10-10]. Dostupné z: [https://ec.europa.eu/regional\\_policy/sources/docgener/work/2014\\_01\\_new\\_urban.pdf](https://ec.europa.eu/regional_policy/sources/docgener/work/2014_01_new_urban.pdf)

DOBEŠOVÁ, Zdena. *Databázové systémy v GIS*. 1. Olomouc: Univerzita Palackého v Olomouci, 2004. ISBN 8024408910.

EUROPEAN ENVIRONMENT AGENCY [online]. [cit. 2019-02-15]. Dostupné z: <https://www.eea.europa.eu/>

EUROSTAT. *Eurostat regional yearbook 2018* [online]. Luxembourg: European union, 2018 [cit. 2020-03-02]. ISBN 978-92-79-87878-7. Dostupné z: <https://ec.europa.eu/eurostat/documents/3217494/9210140/KS-HA-18-001-EN-N.pdf/655a00cc-6789-4b0c-9d6d-eda24d412188>

EUROSTAT [online]. 2019 [cit. 2019-02-15]. Dostupné z: <https://ec.europa.eu/eurostat>

*EVROPSKÁ UNIE: Fakta a čísla* [online]. Lucembursko: Úřad pro publikace Evropské unie, 2017 [cit. 2020-02-05]. ISBN 978-92-824-5974-4. Dostupné z: <https://www.consilium.europa.eu/media/29672/qc0616198csn.pdf>

FREY, William a Zachary ZIMMER. (2001): Defining the city. In Paddison, R. (ed.): Handbook of urban studies, Sage Publishing, London.

GAJIĆ, Aleksandra, Nikola KRUNIĆ a Branko PROTIĆ. Towards a new methodological framework for the delimitation of rural and urban areas: a case study of Serbia. *Geografisk Tidsskrift-Danish Journal of Geography* [online]. 2018, 118(2), 160-172 [cit. 2020-02-29]. DOI: 10.1080/00167223.2018.1503551. ISSN 0016-7223. Dostupné z: <https://www.tandfonline.com/doi/full/10.1080/00167223.2018.1503551>

*GEOFABRIK* [online]. 2019 [cit. 2019-06-11]. Dostupné z: <https://www.geofabrik.de/>

HAMPL, M. a MARADA, M. (2015): Sociogeografická regionalizace Česka. *Geografie*, 120, č. 3, s. 397–421.

HORÁK, Jiří. *Prostorové analýzy dat. 2*. Ostrava: VŠB-TU Ostrava, HGF, Institut geoinformatiky, 2008.

KAUFMAN, Leonard a Peter J. ROUSSEEUW, ed. *Finding Groups in Data* [online]. Hoboken, NJ, USA: John Wiley & Sons, 1990 [cit. 2020-03-05]. Wiley Series in Probability and Statistics. DOI: 10.1002/9780470316801. ISBN 9780470316801.

*KONCEPCE ROZVOJE VENKOVA*. In: . Praha: Ministerstvo pro místní rozvoj ČR Odbor regionální politiky, 2019. Dostupné také z: [https://www.mmr.cz/getmedia/279d5264-6e9e-4f80-ba4a-c15a26144cd0/Koncepce-rozvoje-venkova\\_202001.pdf.aspx?ext=.pdf](https://www.mmr.cz/getmedia/279d5264-6e9e-4f80-ba4a-c15a26144cd0/Koncepce-rozvoje-venkova_202001.pdf.aspx?ext=.pdf)

MAREK, Lukáš. *Prostorové a vícerozměrné statistické analýzy epidemiologických dat*. Olomouc: Univerzita Palackého v Olomouci pro katedru geoinformatiky, 2015. Terra notitia. ISBN 978-80-244-4820-6.

MELOUN, Milan a Jiří MILITKÝ. *Kompéndium statistického zpracování dat*. Praha: Karolinum, 2002. ISBN 978-80-246-2196-8.

MELOUN, Milan a Jiří MILITKÝ. *Statistická analýza experimentálních dat*. Vyd. 2., upr. a rozš. Praha: Academia, 2004. ISBN 80200-1254-0.

*NACE REV. 2: Statistical classification of economic activities in the European Community*. Luxembourg: Office for Official Publications of the European Communities, 2008. ISBN 978-92-79-04741-1.

*NÁRODNÍ STRATEGICKÝ PLÁN ROZVOJE VENKOVA ČESKÉ REPUBLIKY NA OBDOBÍ 2007–2013*. In: . Praha, 2006.

NÉTEK, Rostislav a Tomáš BURIAN. *Free and open source v geoinformaticce*. Olomouc: Univerzita Palackého, 2018. ISBN 978-80-244-5291-3.

OECD (2011a), “*Defining and Describing Regions*”, in *OECD Regions at a Glance 2011*, OECD Publishing, Paris.

OECD (2011b), “*OECD REGIONAL TYPOLOGY*”, Directorate for Public Governance and Territorial Development 2011, OECD Publishing, Paris.

*OPENSTREETMAP* [online]. 2019 [cit. 2019-02-15]. Dostupné z: <https://www.openstreetmap.org/>

*OPENSTREETMAP WIKI* [online]. 2019 [cit. 2019-06-11]. Dostupné z: <https://wiki.openstreetmap.org/>

PÁSZTO, Vít. *Prostorová informace a vybrané metody geocomputation pro její hodnocení: autoreferát disertační práce = Spatial information and selected geocomputational methods for its evaluation : Ph.D. thesis summary*. Olomouc: Univerzita Palackého v Olomouci, 2015. Geoinfo-carto thesis. ISBN 978-80-244-4550-2.

PÁSZTO, Vít, Jaroslav BURIAN, Lukáš MAREK, Vít VOŽENÍLEK a Pavel TUČEK. Fuzzy přístup při určování příslušnosti obcí do venkovského a městského prostoru. *Geografie*. 2016, 11, 156-186.

PÁSZTO, Vít, Alžběta BRYCHTOVÁ, Pavel TUČEK, Lukáš MAREK a Jaroslav BURIAN. Using a fuzzy inference system to delimit rural and urban municipalities in the Czech republic in 2010. *Journal of Maps*. 2014, 11(2), 231-239 [cit. 2020-03-04]. DOI: 10.1080/17445647.2014.944942. ISSN 1744-5647.

PERLÍN, Radim. Theoretical approaches of methods to delimitate rural and urban areas. *European Countryside* [online]. 2010, 2(4) [cit. 2020-02-25]. DOI: 10.2478/v10091-010-0013-5. ISSN 1803-8417. Dostupné z: <http://www.degruyter.com/view/j/euco.2010.2.issue-4/v10091-010-0013-5/v10091-010-0013-5.xml>

PERLÍN, Radim. Vymezení venkovských obcí v Česku. *Deník veřejné správy*. [online]. 2009. dostupné z: <http://denik.obce.cz/clanek.asp?id=6384068>

*PROGRAM ROZVOJE VENKOVA ČESKÉ REPUBLIKY NA OBDOBÍ 2007 – 2013*. In: . Praha, 2013.

*PROGRAM ROZVOJE VENKOVA ČESKÉ REPUBLIKY NA OBDOBÍ 2014 – 2020*. In: . Praha, 2014.

*R DOCUMENTATION* [online]. 2020 [cit. 2020-02-21]. Dostupné z: <https://www.rdocumentation.org/>

*STACK OVERFLOW* [online]. 2020 [cit. 2020-02-21]. Dostupné z: <https://stackoverflow.com/>

SVOBODOVÁ, Hana a Antonín VĚŽNÍK. *ÚVOD DO GEOGRAFIE VENKOVA* [online]. Brno, 2014 [cit. 2020-02-07]. Dostupné z: <https://is.muni.cz/el/sci/jaro2016/Z0132/um/62666326/geografie-venkova-skripta.pdf>

ŠARMANOVÁ, Jana. *Metody analýzy dat: učební text*. Ostrava: Vysoká škola báňská – Technická univerzita Ostrava, 2012. ISBN 978-80-248-2565-6.

VOŽENÍLEK, Vít a Jaromír KAŇOK. *Metody tematické kartografie: vizualizace prostorových jevů*. Olomouc: Univerzita Palackého v Olomouci pro katedru geoinformatiky, 2011. ISBN 9788024427904.

ZÁKON Č. 128/2000 SB.: Zákon o obcích (obecní zřízení). In: *Sbírka zákonů ČR*. 2000.

## **PŘÍLOHY**



# SEZNAM PŘÍLOH

## **Vázané přílohy:**

Příloha 1 Zdrojový kód R – Analýza hlavních komponent

Příloha 2 Prohlášení o údržbě chodu aplikace

## **Volné přílohy**

Příloha 3 Poster

Příloha 4 CD

## **Struktura DVD**

Adresáře:

poster

text\_prace

data

web

## **Příloha 1      Zdrojový kód R**

```
# načtení knihoven
library(factoextra)
library(psych)
library(corrplot)
library(ggfortify)
library(scales)
library(RColorBrewer)
library(cluster)

# PCA
pca <- prcomp(data, scale=TRUE)

# výsledky PCA
results<-get_pca_var(pca)

# Scree plot
fviz_screepLOT(pca, geom = "bar", addlabels = TRUE)
fviz_screepLOT(pca, geom = "bar", addlabels = TRUE, choice ="eigenvalue")

# graf komponentních vah
fviz_pca_var(pca, col.var="contrib")

# biplot
autoplot(pca, loadings = T, frame = F)

# příspěvky atributů do prvních čtyř hlavních komponent
fviz_contrib(pca, choice="var", axes=1)
fviz_contrib(pca, choice="var", axes=2)
fviz_contrib(pca, choice="var", axes=3)
fviz_contrib(pca, choice="var", axes=4)

#příprava dat na shlukování
comp <- data.frame(pca$x[,1:4])

# shlukování (CLARA - 6 shluků)
k <- clara(comp, 6, metric="euclidean", stand=F, samples = 2000, pamLike
= T)

# zobrazení shluků
palette(alpha(brewer.pal(9,'Set1'), 0.5))
plot(comp, col=(k$clust), pch=16)
```

## **Příloha 2 Prohlášení o údržbě chodu aplikace**

### **PROHLÁŠENÍ**

Já, Bc. Daniel Pavlačka, tímto prohlašuji, že aplikaci, vytvořenou v rámci diplomové práce „Socioekonomické analýzy nad rozšířenými administrativními daty Eurostatu“ budu udržovat v chodu minimálně do 11. 6. 2023.

Bc. Daniel Pavlačka