

Univerzita Palackého v Olomouci
Přírodovědecká fakulta
Katedra geoinformatiky

**PROSTOROVÁ STATISTIKA ADRESNÍCH
BODŮ ČESKA**

Bakalářská práce

Bohumil GARTNER

Vedoucí práce Mgr. Vít PÁSZTO, Ph.D.

Olomouc 2017
Geoinformatika a geografie

ANOTACE

Tato bakalářská práce aplikuje metody prostorové statistiky na adresní body v Česku. Některé tyto metody lze aplikovat na soubor prvků bez vlivu jejich atributů, avšak většina metod prostorové statistiky atributy a jejich hodnoty využívá.

Při prvotním průzkumu dat byla zjištěna dva problémy. Prvním problémem byla samotná vizualizace dat a jejich výsledků. Vzhledem k velkému počtu adresních bodů a jejich překrývání byla data agregována do pravidelné hexagonové sítě. Druhým problémem byl velký výskyt nulových hodnot (hodnot NULL/0) u vybraných atributů. Proto se v rámci této práce analyzují atributy s potenciálem pro prostorové analýzy a počet nulových hodnot u těchto atributů. Vzhledem k nadměrnému výskytu nulových hodnot v datech se autor práce rozhodl provést prostorovou analýzu průměrného počtu nulových hodnot u atributů adresních bodů. Na základě výsledků této prostorové analýzy byl vzhledem k nejmenšímu průměrnému počtu nulových hodnot atributů zvolen za vhodný pro podrobnější analýzy kraj Hlavní město Praha. Při prostorové statistice adresních bodů Česka se v této práci nejprve hodnotí jejich uspořádání bez ohledu na hodnoty atributů s cílem najít jejich prostorový vzor. Dále se pomocí prostorové statistiky identifikuje, zdali se v prostoru vyskytují shluky hexagonů s nadprůměrným nebo podprůměrným počtem adresních bodů a nulových hodnot jejich atributů. U Prahy vstupují do shlukových analýz atributy adresních bodů s počtem evidovaných obyvatel, bytů a podlaží budov. Shluky jsou pak mapovány v prostoru a jejich výskyt je slovně ohodnocen.

K výpočtům a hodnocení prostorové statistiky adresních bodů bylo maximálně využito sady nástrojů *Spatial Statistics Tools* v aplikaci *ArcMap 10.4* z *ArcGIS Desktop*. K úpravě dat a dalším statistickým výpočtům byl také použit program *Microsoft Excel*. Veškeré výsledky byly v práci vhodně vizualizovány a interpretovány jak z hlediska geografického, tak z geoinformatického.

KLÍČOVÁ SLOVA

prostorová analýza; prostorový vzor; shlukování; adresní body

Počet stran práce: 74

Počet příloh: 25 (z toho 2 volné a 2 elektronické)

ANNOTATION

This bachelor thesis applies spatial statistics methods to address points in the Czech Republic. Some of these methods can be applied to a set of elements without the influence of their attributes, but most of these methods use attributes and their values.

Two issues were identified during the initial survey. The first problem was the visualization of the data and its results. Due to the large number of address points and their overlap, the data was aggregated into a regular hexagon network. The second problem was the large occurrence of zero values (NULL / 0 values) for attributes. Therefore, in this thesis we analysed the attributes with the potential for spatial analyses and their number of zero values. Because of the excessive occurrence of zero values in data, the author of this thesis decided to perform spatial analysis of zero values for attributes of address points. Based on the results of the spatial analysis of the error rate, the Prague Region was selected. It was chosen to be a suitable region for more detailed spatial analyses that use attribute values, because of the smallest average number of the attribute errors. For spatial statistics of address points, firstly their layout is evaluated regardless of the attribute values to find their spatial pattern. Further, using spatial statistics, there are identified hexagon clusters with above or below average number of address points and zero values of their attributes. For Prague, clustering analyses evaluate attributes of address points with the number of registered inhabitants, flats and building floors. Then clusters are mapped in space and their occurrence is verbally evaluated.

Spatial Statistics Tools in ArcMap 10.4 from ArcGIS Desktop has been used maximally to describe the spatial statistics of address points in Czech Republic. Data processing and other statistical calculations also used Microsoft Excel. All the results were appropriately animated and interpreted in terms of geography and geoinformatics.

KEYWORDS

spatial analysis, pattern, clustering, address points

Number of pages: 74

Number of appendixes: 25

Prohlašuji, že

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

- jsem si vědom, že na moji bakalářskou 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 nevydělečně, ke své vnitřní potřebě, bakalářskou práci užívat (§ 35 odst. 3),

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

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

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

- použít výsledky a výstupy mé bakalářské 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

Bohumil Gartner

podpis autora

Děkuji vedoucímu práce Mgr. Vítu PÁZSTOVI, Ph.D. za jeho podporu a pevné nervy při společné spolupráci. Dále děkuji za cenné rady, podněty a připomínky při vypracování práce. V neposlední řadě děkuji mé rodině a přátelům za jejich trpělivost a také za podporu po celou dobu mého studia.

Na závěr děkuji Českému statistickému úřad a Katedře geoinformatiky za poskytnutá data.

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bohumil GARTNER**

Osobní číslo: **R14487**

Studijní program: **B1301 Geografie**

Studijní obor: **Geoinformatika a geografie**

Název tématu: **Prostorová statistika adresních bodů Česka**

Zadávací 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 aplikovat metody prostorové statistiky na adresní body v Česku. Student nejprve zhodnotí prostorovou distribuci (prostorový vzor, shlukování aj.) dat bez ohledu na atributy, a to jak pro celé území Česka, tak i pro vybrané lokality (venkov, město, suburbia, historická jádra). Dále student využije atributy adresních bodů, aby pomocí prostorové statistiky zjistil podobné vlastnosti jednotlivých adresních bodů, čímž nepřímo prozkoumá socio-ekonomické aspekty území. Veškeré výsledky vhodně vizualizuje a provede interpretaci. Student maximálně využije nástrojů Spatial Statistics Toolboxu v ArcGIS for Desktop, v případě nutnosti využije nástrojů GeoDa, R, či jiných (např. extenze ke GIS software). O celé práci bude vytvořen poster.

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) se odevzdá v digitální podobě na CD (DVD) a text práce s vybranými přílohami bude odevzdán ve dvou svázaných výtiscích na sekretariát katedry. O bakalářské práci student vytvoří webovou stránku v souladu s pravidly dostupnými na stránkách katedry. Práce bude zpracována podle zásad dle Voženílek (2002) a závazné šablony pro diplomové práce na KGI.


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

- ANSELIN, L. (1995): Local Indicators of Spatial Association LISA, Geographical Analysis 27(2): 93115, 1995.
CLIFF, A. D., ORD, K. J. (1973): Spatial autocorrelation. Pion Ltd, London, London, 178 s.
CRESSIE, N. A. C. (1993): Statistics for Spatial Data. John Wiley & Sons, New York.
FOTHERINGHAM, A. S. et al. (2000): Quantitative geography: Perspectives on spatial data analysis. Sage, London, 270 s.
GETIS, A., Ord, J.K. (1992): The Analysis of Spatial Association by Use of Distance Statistics in Geographical Analysis 24(3).
HORÁK, J. (2011): Prostorové analýzy dat. VŠB-TU Ostrava, HGF, Institut geoinformatiky, Ostrava, 170 s.
MAREK, L. (2015): Prostorové a vícerozměrné statistické analýzy epidemiologických dat, Univerzita Palackého v Olomouci, 168 s. ISBN 978-80-244-4820-6
MITCHELL, A. (2005): The ESRI Guide to GIS Analysis, Volume 2. ESRI Press.
VOŽENÍLEK, V. (2002): Diplomové práce z geoinformatiky. Vydavatelství Univerzity Palackého, Olomouc, UP, 31 s.
A další relevantní tištěné i elektronické zdroje

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

Datum zadání bakalářské práce: 15. června 2016

Termín odevzdání bakalářské práce: 5. května 2017


UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA GEOINFORMATIKY
17. listopadu 50, 771 46 Olomouc
-1-

L.S.

prof. RNDr. Ivo Frébort, CSc., Ph.D.
děkan

prof. RNDr. Vít Voženílek, CSc.
vedoucí katedry

V Olomouci dne 15. června 2016

OBSAH

SEZNAM POUŽITÝCH ZKRATEK	10
ÚVOD	11
1 CÍLE PRÁCE	12
2 METODY A POSTUPY ZPRACOVÁNÍ	13
2.1 Metody prostorové statistiky	13
2.2 Metody pro vizualizaci dat	17
2.3 Ostatní metody	18
2.3.1 Možnosti generování buněk pravidelné gridové sítě	18
2.3.2 Možnosti statistického přepočtu dat	19
2.4 POUŽITÁ DATA	19
2.5 POUŽITÝ SOFTWARE	20
2.6 POSTUP ZPRACOVÁNÍ	20
3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY	22
3.1 Prostorová analýza dat	22
3.1.1 Statistické prostorové analýzy	23
3.1.2 Prostorová statistika a GIS	24
3.1.3 Prostorová statistika v ArcGIS	26
3.1.4 Vyhodnocení výsledků prostorových analýz	26
3.2 Domácí a zahraniční studie s podobnou tematikou	27
3.2.1 Domácí studie	27
3.2.2 Zahraniční studie	29
4 PRŮZKUM, PŘÍPRAVA A ZPRACOVÁNÍ DAT	31
4.1 Analýza primárních dat a výběr vhodných atributů	31
4.2 Analýza počtu nulových hodnot u vybraných atributy	33
4.3 Agregace dat	34
4.3.1 Tvorba hexagonové sítě	34
4.4 Statistický přepočet dat a hodnot jejich atributů	36
4.5 Prostorová analýza průměrného počtu nulových hodnot u vybraných atributů adresních bodů Česka	37
4.5.1 Výsledky a vyhodnocení prostorové analýzy nulových hodnot	39
4.6 Výběr území pro podrobnější průzkum	40
4.6.1 Vizualizace porovnání krajů a vyhodnocení	40
4.7 Statistický přepočet dat a hodnot jejich atributů pro kraj Hlavní město Praha	41
5 PROSTOROVÁ STATISTIKA PRO ÚZEMÍ ČESKA	42
5.1 Adresní body Česka	42
5.1.1 Přepočet pro administrativní jednotky	43
5.1.2 Přepočet pro hexagonové sítě	44
5.1.3 Výsledky prostorové vizualizace	44
5.1.4 Vyhodnocení	45
5.2 Měření geografické distribuce	45
5.2.1 Hledání a porovnávání „centra“ adresních bodů v ČR	45

5.2.2	Směrová distribuce adresních bodů pro obce v ČR	46
5.2.3	Vyhodnocení výsledků z měření geografické distribuce:	51
5.3	Analýzy prostorového uspořádání	52
5.3.1	Průměrný nejbližší soused	52
5.3.2	Prostorová autokorelace	55
5.4	Mapování shluků	58
5.4.1	Analýza shluků vysokých hodnot (Hotspot analýza)	58
6	PROSTOROVÁ STATISTIKA ADRESNÍCH BODŮ HLAVNÍHO MĚSTA PRAHY	60
6.1	Adresní body Prahy	60
6.1.1	Mapové výstupy a vyhodnocení	60
6.2	Analýzy prostorového uspořádání	61
6.2.1	Průměrný nejbližší soused	61
6.2.2	Prostorová autokorelace	62
6.3	Mapování shluků	64
6.3.1	Analýza shluků vysokých hodnot (Hotspot analýza)	64
7	VÝSLEDKY	66
7.1	Mapové výstupy výsledků:	66
7.2	Slovní vyhodnocení výsledků adresních bodů Česka	67
7.3	Slovní vyhodnocení výsledků adresních bodů Prahy	69
8	DISKUZE	71
9	ZÁVĚR	73

SEZNAM POUŽITÝCH ZKRATEK

Zkratka	Význam
CSR	Complete Spatial Randomness
DKM	Digitální katastrální mapa
ČSÚ	Český statistický úřad
ESRI	Environmental System Research Institute
GIS	Geografický informační systém
OLAP	Online Analytical Processing
OLS	Ordinary Least Squares
LISA	Local Indicators of Spatial Association
RSO	Registr sčítacích obvodů
RÚIAN	Registr územní identifikace, adres a nemovitostí
S-JTSK	Systém jednotné trigonometrické sítě katastrální
VDP	Veřejný dálkový přístup
ZSJ	Základní sídelní jednotka

ÚVOD

Podle Českého statistického úřadu je adresní místo budovy takové místo v terénu, kterému lze ve vztahu k budově jednoznačně přiřadit adresu. Typickým příkladem adresního místa je vchod do budovy označený číslem orientačním v rámci ulice a veřejného prostranství. Adresní bod reprezentující adresní místo, který se volí tak, aby svou polohou co nejlépe vystihoval adresní místo. Nejčastěji je umístěn v centru či těžišti dané budovy anebo v místě vchodu. Jedné budově může příslušet více adresních míst (ČSÚ, 2016).

Prostorové analýzy jsou souborem technik pro analýzu a modelování lokalizovaných objektů, kde výsledky analýz závisí na prostorovém uspořádání těchto objektů a jejich vlastností (Horák, 2011). Prostorové analýzy představují důležité nástroje při popisu prostorové statistiky skupiny bodů. Velké množství postupů a metod prostorové statistiky bylo odvozeno v oborech jiných, jako je například geografie, geostatistika nebo územní plánování. Tyto prostorové analýzy využívají a vyžadují jak přístup k informaci o lokalizaci dat, tak i přístup k atributům sledovaných prvků, které vstupují do výpočtů prostorových analýz a mohou jednotlivým objektům přiřazovat různou váhu při výpočtech (Horák, 2011).

V této práci mají prostorové analýzy pro adresní body Česka za cíl co nejvíce porozumět prostorovému uspořádání adresních bodů a zároveň se pokoušejí hledat souvislosti mezi polohou adresních bodů a hodnotami jejich atributů. K dosažení těchto cílů jsou využity metody prostorové statistiky nabízené v programu *ArcMap 10.4* z *ArcGIS Desktop*. Stěžejní část práce se věnuje identifikaci a mapování nulových hodnot atributů adresních bodů Česka a mapování shluků vysokých a nízkých hodnot vybraných atributů pro adresní body v Praze.

1 CÍLE PRÁCE

Cílem této bakalářské práce je aplikovat metody prostorové statistiky na adresní body v Česku pomocí sady nástrojů *Spatial Statistics Tools* v aplikaci *ArcMap 10.4* z *ArcGIS Desktop*. Účelem využití těchto nástrojů je získání co nejvíce informací o prostorovém uspořádání adresních bodů Česka a jejich vzájemných prostorových vztazích. U adresních bodů se nejprve zhodnotí prostorová distribuce bez ohledu na atributy bodů a jejich hodnoty, a to jak pro celé území Česka, tak i pro vybrané lokality (obce).

Mezi teoretické cíle práce bude patřit podrobná rešerše literatury věnující se metodám prostorové statistiky v prostředí *ArcMap* z *ArcGIS Desktop*.

Jedním z hlavních cílů praktické části práce bude analýza nulových hodnot u vybraných atributů adresních bodů s potenciálem pro prostorové analýzy a také mapování oblastí shluků s podprůměrným a nadprůměrným počtem nulových hodnot atributů adresních bodů v Česku. Dalším dílčím cílem bude výběr vhodného regionu pro podrobnější prostorové analýzy a mapování shluků vysokých a nízkých hodnot vybraných atributů adresních bodů tohoto regionu. Veškeré výsledky budou vhodně vizualizovány a interpretovány.

K této práci bude také vytvořena webová aplikace, zobrazující výsledky, dále poster a webová stránka shrnující cíle, postupy a výsledky prostorové statistiky adresních bodů Česka. V textu bude také uvedena anotace v anglickém jazyce.

2 METODY A POSTUPY ZPRACOVÁNÍ

V rámci práce byly využity především metody prostorové statistiky, které nabízí prostředí aplikace *ArcMap 10.4 z ArcGIS Desktop*. Dále byly použity areálové vizualizační metody a metody statistického přepočtu dat. Při práci bylo také nutné najít vhodné metody pro generování pravidelné gridové sítě a metody statistický přepočet dat.

2.1 Metody prostorové statistiky

Cílem práce bylo maximálně využít sadu nástrojů ***Spatial Statistics Tools*** (nástroje pro statistiku prostorových dat), která je obsažena mezi základními sadami nástrojů v programu *ArcMap 10.4 z ArcGIS Desktop*. Proto bude v této kapitole uveden stručný popis této sady nástrojů i s obecným popisem jednotlivých metod. Zdrojem stručných popisů uvedených u jednotlivých nástrojů je ESRI nápověda, která je volně dostupná v anglickém jazyce přímo v programu *ArcMap 10.4 z ArcGIS Desktop* nebo na serveru ArcGIS Online (ESRI, 2017). Názvy a stručné popisy metod byly autorem práce přeloženy do češtiny.

Panel nástrojů prostorové statistiky obsahuje statistické nástroje pro analýzu prostorové distribuce, vzorů (patterns), procesů a jejich vztahů. Jednotlivé metody obsažené v této sadě nástrojů umožňují shrnout charakteristické vlastnosti prostorového rozložení (např. určení směrového trendu, identifikace statisticky významných seskupení nebo prostorové odchylky). U všech těchto metod je navíc k dispozici zdrojový kód (v Pythonu), takže je možné je upravovat, rozšiřovat nebo sdílet.

Statistické metody jsou ve *Spatial statistics Tools* rozděleny do čtyř sad nástrojů:

- **Measuring Geographic Distribution** – měření geografické distribuce
- **Analyzing Patterns** – analýzy prostorového uspořádání
- **Mapping Clusters** – mapování shluků
- **Modeling Spatial Relationships** – modelování prostorových vztahů

Measuring Geographic Distribution – Měření geografické distribuce

Měření prostorové distribuce sady prvků umožňuje vypočítat hodnotu, která reprezentuje charakteristiku distribuce jako je střed, kompaktnost nebo orientace. Tato hodnota může být použita ke sledování změn v distribuci v průběhu času nebo ke srovnání distribucí různých funkcí.

Tato sada nástrojů řeší otázky jako:

- Kde je centrum?
- Jaký je tvar a orientace dat?
- Jak moc rozptýlené jsou prvky?

NÁSTROJE:

Central feature – Identifikace středového prvku

- identifikuje nejvíce centrálně položený prvek v bodové, liniové nebo polygonové třídě prvků

Directional Distribution – Směrová distribuce

- vytvoří elipsy nebo elipsoidy směrodatných odchylek ke zhodnocení prostorových charakteristik geografických prvků: centrální tendence, disperze (rozptyl) a směrové trendy

Linear Directional Mean – Průměrný směr skupiny vektorů

- identifikuje průměrný směr, délku a geografický střed pro sadu linií

Mean Center – Geografický střed prvků

- identifikuje geografický střed nebo centrum koncentrace pro sadu prvků

Median Center – Geografický medián prvků

- určí polohu, která minimalizuje celkovou euklidovskou vzdálenost ke všem prvkům v datové sadě

Standard Distance – Směrodatná vzdálenost

- měří stupeň koncentrace nebo rozptylu prvků kolem geometrického průměrného středu

Analyzing Patterns – Analýzy prostorového uspořádání

Identifikace geografických vzorů je důležitá pro pochopení chování geografických jevů. Tyto nástroje vyhodnotí, jestli se prvky nebo hodnoty spojené s prvky shlukují nebo naopak zdali jsou rozptýlené nebo jestli tvoří náhodný prostorový vzor.

Všechny nástroje jsou inferenční (deduktivní) statistiky, začínají stanovením nulové hypotézy, která tvrdí, že prvky a hodnoty spojené s prvky vykazují prostorově náhodný vzor. Poté počítají hodnotu p , která reprezentuje pravděpodobnost, že nulová hypotéza je správná (pozorovaný vzor je jednou z mnoha možných verzí úplné prostorové náhodnosti). Výpočet pravděpodobnosti může být důležitý, pokud je potřeba mít vysokou míru důvěry v konkrétní rozhodnutí.

Tato sada nástrojů řeší otázky jako:

- Jsou zkoumané prvky nebo hodnoty spojené se zkoumanými prvky prostorově seskupeny?
- Je shlukování v průběhu času více či méně intenzivní?

NÁSTROJE:

Average Nearest Neighbor – Průměrný nejbližší soused

- vypočítá index nejbližšího souseda založený na průměrné vzdálenosti od každého prvku k jeho nejbližšímu sousednímu prvků

High/Low Clustering – Shlukování vysokých a nízkých hodnot

- měří stupeň shlukování pro každou vysokou nebo nízkou hodnotu použitím *Getis – Ord General G* statistiky

Incremental Spatial Autocorrelation – Přírůstková prostorová autokorelace

- měří prostorovou autokorelaci pro sérii vzdáleností a volitelně tvoří liniový graf těchto vzdáleností a odpovídajících hodnot *z-score*
- *z-score* odráží intenzitu prostorového shlukování a statisticky významné maximální hodnoty *z-score* indikují vzdálenosti, kde prostorové jsou procesy podporující shlukování nejvýznamnější, tyto maximální vzdálenosti jsou často vhodné hodnoty pro použití nástrojů s parametry *Distance Band* (pásmo vzdálenosti) nebo *Distance Radius parameter* (poloměr vzdálenosti)

Spatial Autocorrelation – Prostorová autokorelace

- měří prostorovou autokorelaci založenou na poloze prvků a hodnotách jejich atributů použitím *Globální Moranovi I* statistiky

Multi-Distance Spatial Cluster Analysis (Ripley's k-function) – Vícevdálenostní prostorová shluková analýza (funkce Ripley K)

- určuje, zdali prvky nebo hodnoty spojené s prvky vykazují statisticky významné shlukování nebo rozptyl v daném rozsahu

Mapping Clusters – Mapování shluků

Tyto nástroje mohou být využity k identifikaci statisticky významných shluků vysokých hodnot, shluků nízkých hodnot nebo prostorových *outliers* (odchylky, výrazně odlišné hodnoty). Jsou zde také nástroje, které je možné využít k identifikaci nebo seskupení prvků se stejnými vlastnostmi. Stanovení polohy prostorových shluků je důležité při hledání potenciálních příčin shluků (např. kde se vyskytuje ohnisko onemocnění).

Tato sada nástrojů řeší otázky jako:

- Kde se vyskytují shluky vysokých a nízkých hodnot (*hotspots/ coldspots*)?
- Kde je největší hustota incidentů?
- Kde se vyskytují prostorové odchylky (*outliers*)?
- Které prvky jsou si svými vlastnostmi nejvíce podobné?

NÁSTROJE

Cluster and Outlier Analysis – Analýza homogenních a heterogenních shluků

- na základě vážených bodových dat identifikuje prostorové shluky extrémních hodnot (*hotspots/ coldspots*) a prvky, které leží mimo (*outliers*)

Grouping Analysis – Analýza seskupení

- seskupuje prvky na základě jejich atributů a volitelných prostorových nebo časových omezení

Hot Spot Analysis – Analýza shluků vysokých hodnot

- vzhledem k váženým prvkům, identifikuje statisticky významné shluky vysokých hodnot a shluky nízkých hodnot použitím *Getis-Ord Gi** statistiky

Optimized Hot Spot Analysis – Optimalizovaná analýza shluků vysokých hodnot

- vzhledem k bodům nebo váženým prvkům vytváří mapu statisticky významných shluků vysokých a nízkých hodnot použitím *Getis-Ord Gi** statistiky, vyhodnocuje charakteristiky vložené třídy prvků k dosažení optimálních výsledků

Similarity Search – Zjištění podobnosti

- identifikuje, které vybrané prvky jsou nejvíce nebo nejméně podobné jednomu nebo více zvoleným prvkům na základě jejich vlastností

Modeling Spatial Relationships – Modelování prostorových vztahů

Tyto nástroje modelují vztahy prvků použitím regresní analýzy či konstrukcí matice prostorových vah. Nástroje, které vytvářejí maticové soubory prostorových vah, měří, jak se prvky vzájemně v prostoru ovlivňují a jak na sebe působí. Matice prostorových vah

reprezentuje prostorovou strukturu dat, tedy prostorové vztahy mezi prvky v datovém souboru. Mezi metody, které zohledňují matice prostorových vah patří *Spatial Autocorrelation (Prostrová autokorelace)*, *Cluster and Outlier Analysis (Analýza homogenních a heterogenních shluků)* a *Hot Spot Analysis (Analýza shluků vysokých hodnot)*.

NÁSTROJE

Exploratory Regression – Explorativní regresní analýza

- tento nástroj ohodnotí všechny možné kombinace vstupních kandidátů nezávislých proměnných, vyhledá OLS (Ordinary Least Squares) modely, které nejlépe popisují závislou proměnnou v kontextu kritérií zadaných uživatelem

Generate Spatial Weights Matrix – Generovat matici prostorových vah

- vytvoří matici prostorových vah (soubor. swm) pro reprezentaci prostorových vztahů mezi prvky v datovém souboru

Generate Network Spatial Weights Matrix – Generovat prostorové váhy sítě

- vytvoří matici prostorových vah charakterizující prostorové vztahy použitím sady dat uložených na síti

Geographically Weighted Regression – Regrese určená z geografických vah

- provádí geograficky váženou regresi, lokální formu lineární regrese používanou k modelování vztahů proměnlivých v prostoru

Ordinary Least Squares – Lineární regrese metodou nejmenších čtverců (OLS)

- použije metodu nejmenších čtverců (OLS) při lineární regresi a generuje předpovědi anebo modeluje závislou proměnnou v rámci jejich vztahů k sadě nezávislých proměnných

Spatial Statistics Tools obsahuje také nástroje s názvem **Utilities – Pomocné funkce**. Jedná se o užitečné doplňkové funkce. Tyto pomůcky provádí řadu rozmanitých funkcí jako výpočet plochy, posouzení minimální vzdálenosti, export proměnných a geometrie, převádění souborů prostorových vah a sběr souhlasných bodů. Podrobnější informace o těchto funkcích jsou přístupné v anglickém jazyce online na: <http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/an-overview-of-the-utilities-toolset.htm>.

V rámci daného rozsahu této práce a časovým možnostem nebyly využity pro prostorové analýzy adresních bodů všechny zmíněné sady nástrojů a metody v nich obsažené. Všechny tyto nástroje však byly v rámci hledání vhodných metod pro prostorovou statistiku adresních bodů nastudovány, a proto je i v této podkapitole pro úplnost uveden jejich stručný popis. Po konzultaci s vedoucím práce bylo rozhodnuto používat pro názvy nástrojů a metod prostorové statistiky v následujícím textu uvedené české ekvivalenty.

2.2 Metody pro vizualizaci dat

Vzhledem k počtu zpracovávaných dat (2 815 224 bodů) a charakteru jejich rozmístění nebylo možné pro jejich vizualizace použít metodu bodových znaků. Ta by působila dost nepřehledně, a navíc by docházelo k překrytí velkého množství bodů. Proto bylo potřeba zvolit metody jiné, ty jsou popsány v následující podkapitole.

Areálová metoda

Bývá označována také jako metoda plošných znaků a je jednou z nejvíce používaných metod při zpracovávání tematického obsahu v kartografii. Vlastnosti a parametry této metody popisuje Voženílek, Kaňok a kol. (2011). Podle autorů může být plošný znak použit ke znázorňování prostorových jevů několika způsoby:

- jako samostatný vyjadřovací prostředek v areálové metodě,
- jako součást složitějších vyjadřovacích prostředků v různých metodách (kartogram, kartografická anamorfóza, dasymetrická metoda, tečková metoda aj.).

Jak dále autoři uvádí, metoda využívá jen dva parametry plošných znaků: obrys a výplň. Oběma parametry lze vyjádřit jak kvalitativní vlastnosti, tak kvantitativní hodnotu. Při vytváření kvalitativní a kvantitativní výplně platí určité zásady, které je nutné dodržovat a které tu autoři podrobněji popisují.

Voženílek, Kaňok a kol. (2011) vidí nejčastější použití metody jako součást složitějších vyjadřovacích prostředků je kartogram. Podle autorů metoda kartogramu umožňuje kvantitativní srovnání jednotlivých dílčích územních celků jako jsou například administrativní jednotky nebo se také v menší míře využívá při regionalizaci jevu. Dále uvádí, že podstatou metody kartogramu je znázornění jevu vyjádřeného relativními hodnotami tak, aby byli dílčí územní celky srovnatelné, a proto je velmi důležité, aby byla kvantitativní data přepočtena na jednotku plochy dílčího územního celku, např. počet obyvatel na 1 km². Pokud nastane situace, kdy nemá smysl přepočítávat jev na jednotku plochy, používá se označení pseudokartogram. Pseudokartogramy je možné používat pouze v případě, že dílčí územní celky mají přibližně stejné velikosti ploch, na které se vyjadřovaný kvantitativní jev přepočítává. Dle Voženíleka, Kaňoka a kol. (2011) se podle počtu znázorňovaných jevů rozlišují kartogramy jednoduché a složené, podle způsobu interpretace jevu strukturní, tečkové, čárové a pseudoprostorové.

Metoda gridu

Jedná se o zvláštní druh areálové metody, kdy si pod pojmem grid lze představit pravidelnou polygonovou síť (mřížku), kterou je sledované území rozděleno na stejně velké buňky. Tato metoda tedy nepodléhá časově a územně proměnlivé administrativní struktuře. Nejčastějšími základními tvary mřížky jsou čtverce, dále se pak často používají trojúhelníky a hexagony. Pravidelnost a stejná velikost všech jednotek umožňuje vzájemné srovnávání. Umožňuje také hierarchizaci prostorové prezentace a její přesnosti tím, že lze volit různou velikost buňky při zachování pravidelnosti sítě. Podle Klauudy (2016) se informace do buněk sítě přiřazují metodami agregace, disagregace nebo jejich vzájemnou kombinací. Dále autor uvádí, že v letech 2013–2014 se ČSÚ aktivně podílel na projektu GEOSTAT 1B, jehož výsledkem je harmonizovaná

paneurospká čtvercová síť grid v rozlišení 1 km a rozpracované metody agregace a disagregace aplikované na datech ze sčítání lidu z roku 2011.

Agregace

Přesněji agregace prostorově detailnějších statistických dat (down-top, bottom-up přístup). Podle Klaudy (2016) jde o metodu nejjednodušší a zároveň poskytující nejlepší výsledky, kdy typicky jsou na vstupu adresně lokalizovaná data. Horák (2016) uvádí, že cílem agregace je především potlačení vlivu náhodných efektů u dat s vysokým rozlišením, které by znesnadnily analýzy a interpretace snadnější identifikace trendů a vazeb (Harries, 1999, Chainey, Ratcliffe 2013). Agregace je rovněž základem k definování Online Analytical Processing (OLAP) (Teorey et al. 2006). Podle Horák (2016) je agregace také významným nástrojem anonymizace dat. Autor uvádí, že agregaci je potřeba provádět trojím způsobem:

- podle geometrického vymezení
- podle administrativního uspořádání územně správních jednotek
- podle organizačního členění bezpečnostních útvarů

Podle geometrického vymezení je nejnižší doporučená agregační jednotka buňka 100 × 100 m, orientovaná podle os souřadnic S-JTSK. Vyššími jednotkami jsou buňky 500 × 500 m, 1 × 1 km a 5 × 5 km. Podle administrativního uspořádání je nejnižší agregační jednotkou část obce, vyššími jednotkami jsou obec, okres, kraj a stát. (Horák, 2016). Další možnost pro provádění prostorové agregace dat je jádrové vyhlazení, které představuje specifickou techniku, jež lze využít pro dosažení agregace dat v území, i když její metodické principy jsou odlišné od běžných agregací (Horák a kol. 2010, Inspektor a kol. 2014).

2.3 Ostatní metody

Pro splnění cílů práce bylo nutné agregovat data, tedy najít vhodné metody pro generování pravidelné gridové sítě a pro statistický přepočítání dat. Nejvhodnější metody pro tyto účely jsou popsány v následující podkapitole

2.3.1 Možnosti generování buněk pravidelné gridové sítě

Pro vizualizaci výskytu dat a výsledků analýz bylo potřeba data agregovat pomocí vytvoření pravidelné gridové sítě. Za základní tvar sítě jsem si zvolil hexagon, jelikož má ze standardně využívaných tvarů nejvíce sousedů, což je při prostorových analýzách výhodou. Software *ArcMap 10.4* z *ArcGIS Desktop* nabízí pro tvorbu geometrických sítí nástroj **Create Fishnet**, který je dostupný v nástrojové sadě Data Management Tools. Tento nástroj ovšem vytváří pouze čtvercové nebo obdélníkové geometrické sítě. Možnosti software *ArcMap 10.4* nebyly pro generování buněk pravidelné hexagonové sítě v případě této práce optimální, a proto bylo rozhodnuto využít extenzi *Repeating Shapes*, která se zdála být pro účely této práce nejlepší. Další možností je pak například nástroj *Create Hexagon Tessellation*.

Repeating Shapes

Jedná se o volně přístupnou extenzi pro ArcGIS, která obsahuje skript *Generate Tessellation*. Autorem je Jeff Jennes, biolog a GIS analytik z USA. První verze byla vydána v roce 2006, nejnovější verze 1.5.152 vyšla v listopadu roku 2012. Extenze je

vhodná pro verze *ArcMap 9.x* a pro *ArcMap 10.x* a vyžaduje *ArcView* stejné verze. Pracuje na všech licenčních stupních *ArcMap* z *ArcGIS Desktop* (Jenness, 2012).

Tento skript umožňuje vytvořit pravidelné mřížky o různých velikostech, tvarech a úhlech orientace pro vybrané studované území.

Rozsah studovaného území může být definováno jako:

- aktuální zobrazované okno
- rozsah všech vrstev v projektu
- rozsah konkrétní vrstvy v projektu
- označená část ve vrstvě

Buňky mají nejčastěji tvar čtverce nebo hexagonu, méně častěji pak pravoúhelníku, trojúhelníku kolečka nebo bodu. Velikost buněk je definována v metrech nebo v metrech čtverečních. Tvary lze generovat různě orientované. Časová náročnost vytvoření pravidelné gridové sítě se odvíjí od rozsahu studovaného území a od nastavení parametrů základní buňky. Skript je uživateli hodnocen jako spolehlivý, protože zvládá i náročnější operace (Jenness, 2012).

2.3.2 Možnosti statistického přepočtu dat

Jelikož se adresní body agregovaly na hexagony nebo na administrativní jednotky, bylo nutné provést přepočet hodnot z více bodů do jednoho polygonu. Pro přepočet dat nebylo nutné pátrat po extenzích jako při generování buněk pravidelné gridové sítě, jelikož program *ArcMap 10.4* z *ArcGIS Desktop* a jeho základní funkcionalita nabízí možnosti statistického přepočtu dat hned několik. Nabízené možnosti jsou nástroje *Join Data*, *Summarize Within a Point Statistics*. Ze zmíněných se zdál být nejlepší nástroj *Join Data*, podrobnější popis tohoto nástroje je uveden níže.

Join Data

Zvolenou možností pro statistický přepočet dat a jejich hodnot byl nástroj *Join Data from another layer based on spatial location*. Tento nástroj umožňuje provést statistický souhrn dat z jedné vrstvy do druhé založený na lokaci dat v prostoru. Na začátku se definuje vstupní vrstva (bodová) pro statistický souhrn, jejíž hodnoty se na základě prostorového umístění přiřadí výstupní vrstvě (polygonové). Dále se zvolí způsob agregace. Hodnoty mohou být agregovány jako součet, průměr, minimální a maximální hodnota, směrodatná odchylka a rozptyl. Tento přepočet může být ve srovnání s dalšími možnostmi přepočtu časově náročnější, jelikož nelze vybírat atributy pro provedení statistického přepočtu, ten proběhne u všech atributů. Proto se musí nepotřebné vypočítané statistiky u daných atributů promazat tak, aby nedocházelo k redundanci dat a nárůstu časové náročnosti při práci s nimi.

2.4 POUŽITÁ DATA

Primární data, která byla autorovi práce poskytnuta vedoucím práce pro studijní účely, tvoří bodová vrstva adresních míst Česka s názvem „Statistické budovy (vchody k bytům)“. Tato data byla poskytnuta Katedře geoinformatiky Českým statistickým úřadem (ČSÚ), který ve spolupráci s Registrem sčítacích obvodů (RSO) a Registrem územních identifikací, adres a nemovitostí (RÚIAN) adresní místa v Česku spravuje. Použitá data jsou aktuální k 1. červenci 2016.

Český statistický úřad (2016) uvádí, že vrstva obsahuje prostorovou lokalizaci budov s číslem domovním včetně všech evidovaných částí budov (vchodů k bytům) ve formě definičního bodu s atributací TEP (technickoekonomické parametry). Rozsah či úplnost dat je 99,1 % území Česka z celkového evidovaného počtu, kvalita je rovnoměrná. Interpretace dat podle ČSÚ (2016): má-li budova více samostatných částí s byty a vlastními vchody, je každý z nich reprezentovaný definičním bodem - jde rovněž o vchody neoznačené nebo označené pouze číslem orientačním; unikátním klíčem části budovy/vchodu je identifikátor budovy IDOB a pořadové číslo budovy PC_BUDOV; není kladen důraz na lokalizaci budov typu garáž, které (mají-li přiděleno číslo domovní) jsou součástí evidence v popisné části; kromě nich není lokalizace známa zejména u rekreačních objektů (chaty). Polohová přesnost bodů by měla být mezi 0,5 m (DKM) – 1 m (ruční digitalizace).

Pro potřeby práce s administrativními jednotkami byly využity vrstvy administrativního členění z digitální vektorové databáze ArcČR® 500 verze 3.3, které poskytuje firma ARCDATA Praha, s.r.o. zcela zdarma.

2.5 POUŽITÝ SOFTWARE

Pro průzkum a zpracování dat, přepočty při agregaci, provádění analýz a vizualizaci výsledků byl v maximální míře využit software **ArcMap 10.4 z ArcGIS Desktop** od společnosti ESRI (*Environmental Systems Research Institute*) obsahující sadu nástrojů *Spatial Statistics Tools*.

Pro generování hexagonové sítě v prostředí *ArcMap 10.4 z ArcGIS Desktop* byla použita volně dostupná extenze *Repeating Shapes*, jejíž součástí je i skript *Generate Tessellation*. Ten umožňuje tvorbu pravidelných mřížek pro vybrané studované území. Mřížky jsou vytvářeny v různé velikosti, tvaru buněk a úhlu orientace. Extenze byla vytvořena Jeffem Jennessem. První verze pochází z roku 2006, nejnovější verze 1.5.152 byla vydána v listopadu roku 2012 (Jenness, 2012).

Pro statistické zpracování dat, tvorbu tabulek a grafů byl použit tabulkový software **Microsoft Excel 2016**.

2.6 POSTUP ZPRACOVÁNÍ

Samotnému zpracování této práce předcházelo studium řešené problematiky. Proběhlo studium zadané tematiky a bližší seznámení se statistickými prostorovými analýzami. Autor této práce důkladně nastudoval možnosti nástrojů *Spatial Statistics Tools* v *ArcMap 10.4 z ArcGIS Desktop*, aby zjistil možnosti využití nástrojů prostorové statistiky a jejich optimální nastavení při prostorových analýzách.

Následovala analýza dat, při které byly mimo jiné vybrány atributy s potenciálem pro provádění prostorových analýz.

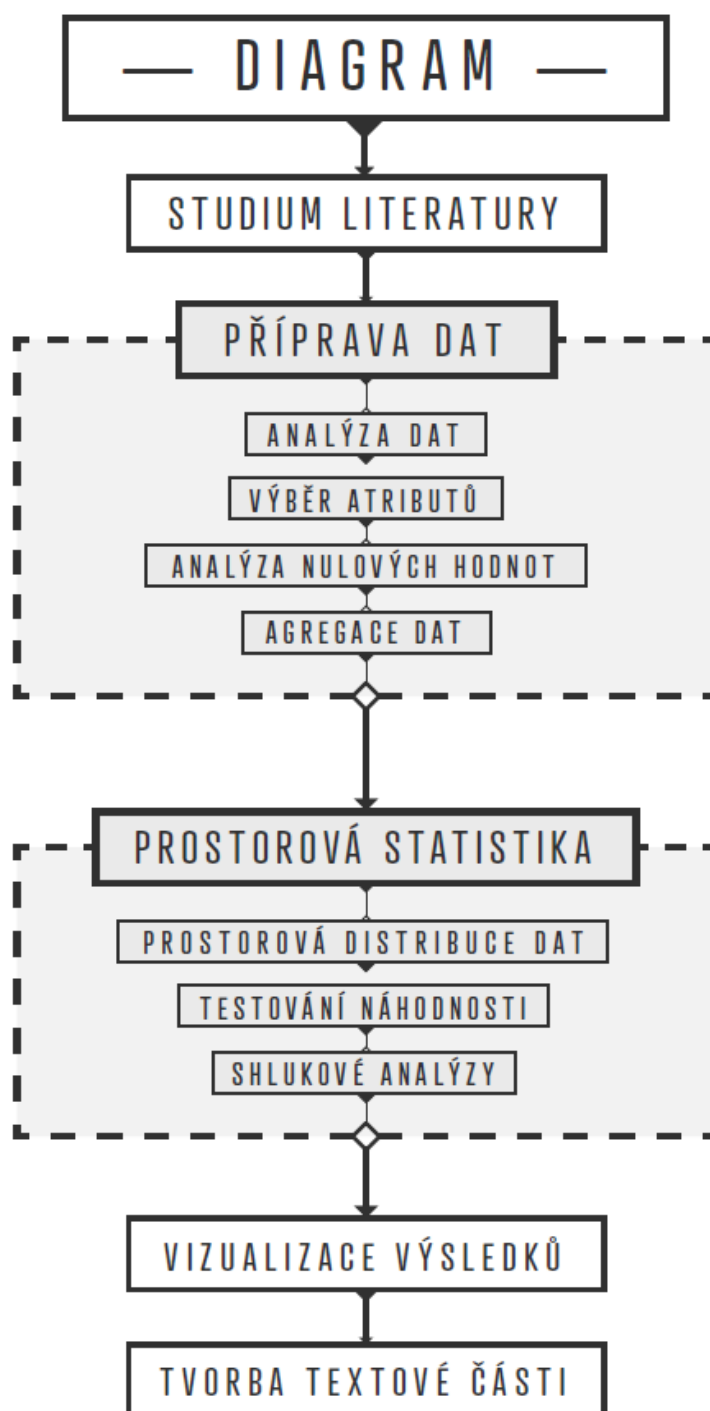
V dalším kroku se u těchto vybraných atributů provedla analýza jejich nulových hodnot. Výsledky této analýzy poukázaly na nadměrný výskyt nulových hodnot u mnoha atributů. Proto byla provedena i prostorová analýza průměrného počtu nulových hodnot vybraných atributů u adresních bodů v Česku. Na základě výsledků této prostorové analýzy byl vybrán vhodný region pro provádění prostorových analýz.

V průběhu práce bylo nutné data agregovat. Byly vytvořeny pravidelné hexagonové sítě, do kterých byly pomocí nástrojů statistického přepočtu dat adresní body agregovány.

Hlavní část práce představovala výpočet prostorové statistiky pro adresní body v Česku a ve vybraném regionu. Nejprve byla vyhodnocena prostorová distribuce adresních bodů bez vlivu hodnot jejich atributů. Následovalo testování náhodnosti rozmístění dat. Tvorba shlukových analýz proběhla jak pro samotné adresní body, tak i pro hodnoty jejich atributů.

Po výpočtu prostorové statistiky následovala vizualizace výsledků prostorových analýz. Na závěr byla sepsána textová část práce, která obsahuje také obrázky, grafy a tabulky. K celé práci byly vytvořeny webové stránky a informační poster popisující cíle, postup a výsledky práce.

Postup práce je krok po kroku zobrazen v následujícím diagramu (Obr. 2.6).



3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

Před samotným řešením práce bylo nutné získat pomocí studia literatury více informací o prostorových statistikách a obecně i o prostorové analýze dat. Proto byly nastudovány možnosti softwaru *ArcMap 10.4 z ArcGIS Desktop* pro prostorové statistiky a vhodná nastavení u konkrétních metod obsažených v sadě nástrojů *Spatial Statistics Tools*. Na základě tohoto studia byly vybrány vhodné nástroje a metody pro vizualizaci výsledků a dále také parametry vstupující do výpočtů prostorových analýz.

3.1 Prostorová analýza dat

Prostorové analýzy tvoří v dnešní době podstatnou a zcela určitě nepostradatelnou součást geografických informačních systémů (GIS). Napříč různými obory se setkáváme s větším množstvím definic a vysvětlení, které se snaží pojem prostorových analýz objasnit. V této práci jistě nejsou uvedeny všechny, ale jen několik těch, se kterými se autor práce setkal při studiu literatury. První tři definice jsou podle staršího vnímání prostorových analýz, další vypovídají o tom, jak prostorové analýzy vnímáme dnes.

Unwin (1981): „Prostorové analýzy se zabývají uspořádáním prostorových dat na mapách (tedy bodů, linií, ploch, povrchů).“

Goodchild (1987): „Prostorové analýzy jsou techniky umožňující popis uspořádání na mapách a srovnání 2 a více map s cílem identifikace jejich vztahů.“

Johnston, Gregory, Smith (1994): „Prostorové analýzy jsou kvantitativní (hlavně statistické) procedury a techniky aplikované v lokalizačních úlohách.“

V GIS slovníku od ESRI je uvedeno, že prostorové analýzy jsou procesy zkoumání lokalit, atributů a vztahů prvků u prostorových dat prostřednictvím překrývání a dalších analytických technik s cílem získání užitečných poznatků. Prostorová analýza vytáhne nebo vytvoří nové informace z prostorových dat.

Dle terminologického slovníku zeměměřičství a katastru nemovitostí (VÚGTK, 2006) je prostorová analýza proces, který umožňuje prostřednictvím operací, realizovaných nad geografickými daty, získat charakteristiky jimi reprezentovaných jevů.

Podle Horáka (2011) lze prostorové analýzy definovat následovně: Prostorové analýzy jsou souborem technik pro analýzu a modelování lokalizovaných objektů, kde výsledky analýz závisí na prostorovém uspořádání těchto objektů a jejich vlastností. Objekty geografické i jiné objekty s prostorovou lokalizací přitom mohou být fyzické či abstraktní povahy, často reprezentují události nebo jevy. Bez ohledu na konkrétní vyjádření je zjevné, že prostorové analýzy představují sadu analytických metod, vyžadujících přístup k atributům studovaných objektů i k informaci o jejich lokalizaci. Na rozdíl od jiných forem analýz tedy vyžadují prostorové analýzy atributová data i geografickou lokalizaci objektů (Horák, 2011).

Prostorové analýzy nelze plně ztotožnit s analýzami prostorových dat, neboť ne každá analýza prostorových dat využívá jejich prostorový aspekt (Fotheringham a kol., 2000).

3.1.1 Statistické prostorové analýzy

Zdrojem informací následující podkapitoly je práce Horáka (2011): Prostorové analýzy dat, která stručně, jasně a výstižně popisuje prostorové analýzy dat a jejich dělení a využití. Větší část textu v následujících odstavcích čerpá z tohoto díla, a proto už a něj nebude odkazováno.

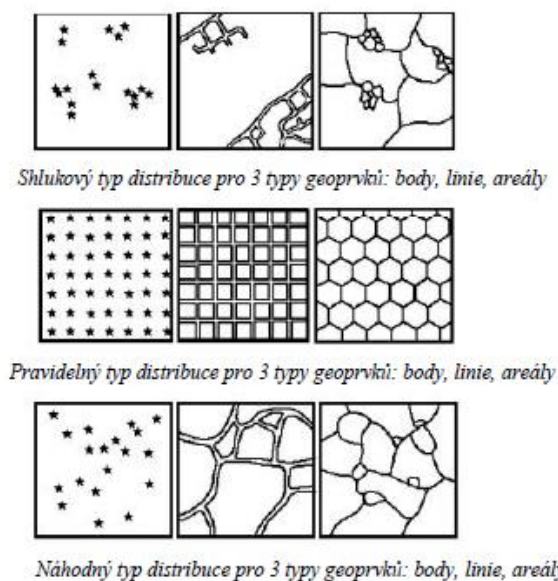
Horák uvádí, že tyto analýzy, které bývají označovány také jako prostorová statistika, zahrnují metody založené na náhodné povaze uspořádání a vztahů, navíc mají široké spektrum využití. Autor dále popisuje, že i když mezi prostorovými a neprostorovými (tradičními) statistikami existují podobné pojmy a cíle, prostorová statistika je jedinečná v tom, že byla vytvořena speciálně pro práci s geografickými daty. Na rozdíl od tradičních metod zahrnuje prostorová statistika do svých výpočtů i vzdálenosti, zkoumanou oblast, konektivitu nebo jiné prostorové vztahy.

Podle počtu současně zkoumaných charakteristik lze prostorové analýzy rozdělit na monovariační (jednorozměrné) a multivariační (analýzy vícerozměrných objektů). Dále se dělí podle povahy statistických technik na popisné (deskriptivní) a interferenční (analýzy textury). Zatímco popisné metody představují především kvantitativní měření charakteristiky polohy a charakteristiky rozptýlení, interferenční metody určují, zda distribuce je nebo není náhodná a popisují vztahy mezi dvěma a více veličinami.

Distribuce prvků, kterou vytváří geoprvky svým rozmístěním ve sledované části prostoru se často označuje jako textura nebo prostorový vzor (pattern). Z hlediska statistického uspořádání geoprvků rozlišujeme **tři základní typy distribuce**:

- a) **shluková** (clustered), případně skupinová
- b) **pravidelná** (regular)
- c) **náhodná** (random)

Příklad těchto tří základních distribucí je zobrazen na Obr. 3.1.1. Skutečné distribuce mohou být testovány vůči těmto třem ideálním typům, často však pouze vůči náhodné distribuci. Statisticky prokázaný výskyt shlukového či pravidelného vzorku může být základem pro zjišťování příčin, které vedly k pozorovanému uspořádání. Např. prokázaný shlukový výskyt případů jisté choroby může (pokud byla data správně standardizována, a to i v časové ose) podpořit hypotézu infekčního charakteru (etiologie) choroby (Horák, 2011).



Obr. 3.1.1 Ukázka 3 typů prostorové distribuce pro různé geoprvky (Horák, 2011).

3.1.2 Prostorová statistika a GIS

Propojení GIS a prostorových analýz bylo v průběhu času ovlivněno především rozvojem technologií, metod a aplikací. Lidé si zanedlouho uvědomili sílu tohoto propojení. Už Aronoff (1989) poukazuje na provádění prostorových analýz jako na nejcennější využití GIS. Geografické informační systémy dosáhly v poslední době značného rozšíření a staly se základním nástrojem pro správu a zpracování prostorových dat i prostředníkem pro poskytování a využívání prostorových informací (Horák, 2011). Vývoj geoinformačních systémů značně zjednodušil vizualizaci výsledků prostorových analýz jak ve formě map, tak i jiných druhů výstupů (Kliča, 2011). Geoinformační technologie dnes umožňují jednotlivcům i organizacím jak manipulaci s prostorovými daty, tak jejich analýzu i vizualizaci. Často odkrývají vztahy mezi daty, poukazují na jejich prostorové vzory a trendy. Spolu s kartografickými nástroji a nástroji pro správu dat jsou již dlouhou dobu tyto analytické metody základem pro GIS. Jedná se o důležité nástroje pro územní plánování, veřejné zdraví, právní vymáhání, ekologii, dopravu, demografii, management zdrojů a mnoho dalších průmyslových odvětví.

Ve zdravotnictví jsou to hlavně výzkumy o rozšíření chorobných nemocí a jejich identifikace v různých časových obdobích. Tyto výzkumy velice usnadňují práci lékařům, a to především v ohledu diagnózy a samotného léčení (Kliča, 2011). V demografii se prostorové analýzy používají často pro vizualizaci výsledků voleb. V ekologii a krajinné biologii sledují pomocí metod prostorových statistik rozsah teritorií a trasy migrací živočichů.

V průběhu let byly metody prostorové statistiky rozšířeny a obohaceny o nové funkce, dnes zahrnují geostatistické techniky, analýzu rastrových dat, analytické metody pro podnikání, 3D analýzy, síťové analýzy, časoprostorové analýzy a metody specifické pro řadu průmyslových odvětví. Výzkumníci i nadále zlepšují stávající nástroje a vyvíjejí i nástroje nové pro lepší zachycení chování geografických jevů (Mitchell, 2015).

Mezi typy prostorové analýzy často využívané v GIS patří metody prostorové autokorelace, lokální indikátor prostorové asociace, hotspot analýza nebo shluková analýza. Následuje stručný popis těchto metod.

Prostorová autokorelace

Prostorová autokorelace vyjadřuje míru, do jaké je výskyt určitého jevu v prostoru závislý na výskytu stejného jevu v blízkém okolí a představuje kvantitativním vyjádřením prostorové závislosti (Cliff, 1973). Toto vyjádření úzce souvisí s Toblerovým prvním zákonem geografie, který říká: „Všechno souvisí se vším, ale blízké věci spolu souvisejí více než věci vzdálené“ (Tobler, 1970). Spurná (2008) ve své práci publikované v Sociologickém časopise popisuje možnosti a omezení prostorové autokorelace jako jedné z metod prostorové analýzy včetně ukázky některých aplikací. Prostorová autokorelace vyžaduje nastavení pouze jednoho atributu. Tím je vzdálenost, která bude při výpočtu ještě brána za ovlivňující.

Prostorová autokorelace může nabývat kladných nebo záporných hodnot. Kladná autokorelace je výsledkem u prostorového vzoru, kde jsou si sousední či blízké hodnoty podobné. Se zápornou autokorelací se setkáváme u prostorového vzoru, kde jsou sousední a blízké hodnoty velmi rozdílné. Dále rozlišujeme z hlediska rozsahu autokorelaci lokální a globální. Zatímco lokální prostorová autokorelace popisuje konkrétní polohu a rozsah shluků, globální prostorová autokorelace popisuje celkový převládající prostorový vzor ve zkoumaném území (Pfeiffer a kol., 2008).

Lokální indikátor prostorové asociace (LISA analýza)

Lokální indikátor prostorové asociace (LISA) vyjadřuje míru, do jaké je uspořádání hodnot v sousedních jednotkách náhodné (Anselin, 1995). Analýza LISA pracuje s metodou *Moranova I kritéria*. *Moranovo I kritérium* kvantifikuje podobnost zvolených proměnných v oblastech, které jsou definovány jako prostorově příbuzné (Moran, 1950).

Lokální Moranovo I kritérium je definováno vztahem: $L_i = r_i \sum_j (w_{ij}r_j)$, kde r_i a r_j jsou standardizované hodnoty veličiny z a w_{ij} je matice vah. K hodnocení výsledků *lokálního Moranova I kritéria* je vhodné vyjádření shluků pomocí *Moranova diagramu*, kde levý dolní a pravý horní kvadrant vyjadřují prostorové shluky stejných hodnot (*hotspots/coldspots*) a levý horní a pravý dolní kvadrant odlehle hodnoty (*outliers*). *Moranův diagram* je zobrazen na Obr. 3.1.2

Vážená (standardizovaná) hodnota proměnné v blízkých jednotkách	nízká - vysoká negativní prostorová autokorelace	vysoká - vysoká pozitivní prostorová autokorelace
	nízká - nízká pozitivní prostorová autokorelace	vysoká - nízká negativní prostorová autokorelace
	(standardizovaná) hodnota proměnné v prostorové jednotce	

Obr. 3.1.2 Moranův diagram (Spurná, 2008).

Analýza Getis – Ord GI* (Hotspot analýza)

Hotspot analýza slouží k indikaci a identifikaci statisticky významných shluků vysokých hodnot (*hotspots*) či shluků nízkých hodnot (*coldspots*) v prostoru. Švarcová (2017) ve své diplomové práci uvádí, že tato analýza je založena na principu náhledu na jednotlivé prvky v rámci souvislosti s prvky sousedními za pomoci prostorových vztahů. Výsledný *GI** index je vypočítáván pro jednotlivé prvky. Nulová hypotéza v případě hotspot analýz je vyslovena takto: „Hodnoty (prvky) jsou v území rozmístěny náhodně (neexistuje zde prostorové shlukování).“ Pokud je hodnota *Z-skóre* vysoká a hodnota *P-value* nízká je shluk detekován jako *hotspot*, pokud je hodnota *Z-skóre* i *P-value* nízká je v prostoru detekován *coldspot* (Švarcová, 2017).

Shluková analýza

Shluková analýza zkoumá, zda se množina objektů $O = \{O_1, O_2, \dots, O_m\}$ zadaných reálnými atributy $A = \{A_1, A_2, \dots, A_n\}$ přirozeně rozpadá na výrazné podmnožiny objektů si vzájemně podobných, a přitom nepodobných shlukům ostatním. Právě tyto výrazné podmnožiny objektů nazýváme shluky. Pro shluky nebo pro podobnost objektů neexistuje jednoznačná definice. Švarcová (2017) udává, že zkušený analytik musí vyřešit řadu otázek souvisejících se zadanou množinou objektů, a proto shluková analýza netvoří ucelenou teorii, ale je složena z řady metod založených na různých principech. Tato různorodost metod souvisí s různorodostí řešených problémů, požadovaných typů výsledků, časovou i prostorovou složitostí shlukové úlohy pro velká data, neurčitostí definice shluku apod. (Šarmanová, 2012).

Podle Švarcové (2017) se stále průběžně objevují nové algoritmy i aplikace shlukové analýzy, využívající většinou dostupné programy, které jsou součástí velkých programových balíků. S rozvojem dalších oborů (např. neuronových sítí), se objevují i další přístupy k úloze shlukování (Švarcová, 2017). Obecně jsou metody shlukové analýzy děleny podle cíle shlukování na:

- nehierarchické – produkují jen rozklad shluků na podmnožiny
- hierarchické – produkují hierarchii rozkladů, kde každý rozklad je zjemněním předcházejícího

3.1.3 Prostorová statistika v ArcGIS

V roce 2004 byla do *ArcMap 9* z *ArcGIS Desktop* přidána sada nástrojů prostorové statistiky *Spatial Statistics Tools*, která je navržena pro popis prostorového uspořádání prvků. Tato sada nástrojů zahrnuje soubor technik pro popis a modelování prostorových dat. V roce 2017 tato sada nástrojů obsahuje ve verzi *ArcMap 10.4.17* z *ArcGIS Desktop* celkem 26 nástrojů. Většina z nich byla napsána v programovacím jazyku *Python* a uživatelé *ArcGIS* mají přístup k jejich zdrojovému kódu. Od verze *ArcMap 9.3* z *ArcGIS Desktop* jsou statistické metody ve *Spatial statistics Tools* rozděleny do čtyř sad nástrojů: *Měření geografické distribuce*, *Analýzy prostorových vzorů*, *Identifikace statisticky významných shluků*, *Modelování prostorových vztahů*.

3.1.4 Vyhodnocení výsledků prostorových analýz

Zdrojem informací v následující podkapitole je nápověda ESRI, která je volně dostupná v anglickém jazyce přímo v programu *ArcMap 10.4* z *ArcGIS Desktop* nebo na serveru *ArcGIS Online* (ESRI, 2017).

HODNOTA Z-SCORE A P-VALUE

S těmito výrazy se je možné setkat u mnoha metod prostorové statistiky, jelikož jsou součástí jejich výsledků, proto by bylo dobré si ujasnit, co tyto hodnoty vlastně vyjadřují. Většina statistických testů začíná identifikací nulové hypotézy. **Nulová hypotéza** u nástrojů pro analýzu shluků je úplná prostorová náhodnost (CSR – Complete Spatial Randomness) samotných prvků nebo jejich atributových hodnot. Nástroje pro analýzu shluků nám vypočítají *z-score* a *p-value*, které nám říkají, zdali a s jakou jistotou lze nulovou hypotézu zamítnout.

P-value vyjadřuje pravděpodobnost. U nástrojů pro shlukové analýzy je to pravděpodobnost, že pozorované prostorové uspořádání vzniklo nějakým náhodným procesem. Velmi malá hodnota *p-value* značí, že je velmi nepravděpodobné, aby pozorovaný prostorový vzor vznikl pomocí náhodných procesů, takže lze zamítnout nulovou hypotézu.

Z-score je test statistické významnosti, který pomáhá rozhodnout, zda přijmout nebo odmítnout nulovou hypotézu. Reprezentuje míru směrodatné odchylky. Například pokud nám nástroj vrátí hodnotu *z-score* +2.5, je výsledek interpretován jako +2.5 směrodatné odchylky od střední hodnoty.

Pokud je výsledkem shlukových analýz velmi malá hodnota *p-value* a zároveň velmi velká nebo velmi malá hodnota *z-score*, je velmi nepravděpodobné, že by pozorované prvky měli náhodné prostorové uspořádání. Nulovou hypotézu tak lze zamítnout. Pro zamítnutí nulové hypotézy je třeba učinit subjektivní rozhodnutí týkající se míry rizika přijetí špatného výsledku. Tato míra rizika je často uvedena v podmínkách ohledně kritických hodnot a hladinách spolehlivosti. Typické hladiny spolehlivosti jsou

90, 95 a 99 %. Hladina spolehlivosti 99 % by byla v tomhle případě nejkonzervativnější, jelikož naznačuje, že nejste ochotni zamítnout nulovou hypotézu, pokud není pravděpodobnost, že mají prvky náhodné rozmístění, opravdu malá (méně než 1 % pravděpodobnosti). Tabulka 3.1.1 uvádí kritické *p-value* a *z-score* pro různé úrovně spolehlivosti.

Tab. 3.1.4 Kritické hodnoty *p-value* a *z-score* pro různé úrovně spolehlivosti

Z-score	P-value	Hladina spolehlivosti
< -1.65 nebo > +1.65	< 0.10	90 %
< -1.96 nebo > +1.96	< 0.05	95 %
< -2.58 nebo > +2.58	< 0.01	99 %

3.2 Domáci a zahraniční studie s podobnou tématikou

V této části práce jsou uvedeny některé domáci i zahraniční studie s podobnou tématikou a s metodami, které by mohli být využity i při řešení cílů této bakalářské práce.

3.2.1 Domáci studie

V práci Spurné (2008) ***Prostorová autokorelace – všudypřítomný jev při analýze prostorových dat?*** publikované v *Sociologickém časopise*, popisuje autorka možnosti a omezení prostorové autokorelace jako jedné z metod prostorové analýzy včetně ukázky využití. Popisuje zde, jak je prostorová autokorelace vzhledem k obecně metodologickému, ale i k aplikačnímu hledisku důležitá a zasluhuje i velkou pozornost při interpretaci výsledků. Zejména přínos statistické analýzy LISA a srovnání s globální statistikou. Autorka uvádí: „Možnosti, které skýtá tento druh analýzy, jsou důležité zejména z hlediska identifikace prostorových odchylek, a naopak oblasti podobného charakteru z hlediska zkoumaného jevu, například rozvojových os či problémových oblastí (Spurná, 2008).

Práce ***Shluková analýza funkčních ploch pro sledování urbánního vývoje Olomouckého regionu*** od Klíchy (2011) kladla důraz na lokální metody shlukové analýzy a jejich aplikaci pro sledování změn urbánního vývoje v Olomouci a Olomouckém regionu. Jako hlavní metoda shlukových analýz byla zvolena metoda *Getis-Ord Gi**. Pro nastavení analýzy byla zvolena *metoda konstantních vzdáleností*, dalším parametrem pro vzdálenost mezi prvky byla použita *Euklidovská vzdálenost*. Pro shlukové analýzy byly vybrány 4 typy funkčních ploch – bydlení, veřejného vybavení, výrobních ploch a zeleně. Z výsledných shlukových analýz vybraných funkčních ploch na katastrálním území Olomouce bylo patrné, že Olomouc měla již v roce 1930 pevně danou sídelní i funkční prostorovou strukturu města, která se v pozdějších letech rozšiřovala. Byly poměrně přesně lokalizovány statisticky významné shluky bydlení, veřejného vybavení, výrobních ploch a zeleně v daných letech (Klícha 2011).

Cílem diplomové práce ***Prostorové analýzy časových řad volebních výsledků*** od Vlosinského (2011) bylo provést maximální počet prostorových analýz nad časovými řadami výsledků voleb konaných na území Olomouckého kraje od roku 2000. Jednou z použitých analýz je i metoda *Getis-Ord Gi**, kdy autor použil základní nastavení. Výstupy práce se skládaly ze dvou částí. První část byly výstupy samotných analýz –

korelační analýzy pro velké množství různých jevů (velikost obce, počet obyvatel obce, porodnost, nezaměstnanost, nadmořská výška apod.), které neprokázaly statisticky významné korelace. Druhou částí výstupů byly samotné mapy, které jednotlivé analýzy doprovázely. Většina analýz však neodhalila významnější shluky nadprůměrných a podprůměrných hodnot. Přesto i na těchto analýzách lze ve velké míře prokázat odlišnosti a specifika, které jednotlivé volby a politické strany přináší (Vlosinský, 2011).

Cílem diplomové práce **Časoprostorová analýza vybraných onemocnění v Olomouckém kraji** od Havlíka (2013) bylo identifikovat oblasti se zvýšeným výskytem vybraných onemocnění pomocí prostorových analýz, tj. identifikace shlukování a případných shluků. K detailnímu popisu byla vybrána Hepatitida A, která vykazovala nepravidelný průběh. Pro testování náhodnosti rozmístění bodů výskytu byly použity metody *Průměrného nejbližšího souseda a Funkce Ripley K*. Výsledky objasnily fakt, že distribuce použitých dat vykazují prostorové shlukování. Pomocí metody *Getis-Ord G_i^** byla vypočítána shluková analýza devíti sledovaných jevů na úrovni uliční sítě K detailnímu popisu byla vybrána Hepatitida A, která vykazovala nepravidelný průběh. Výsledkem bylo 68 vrstev, které byly znázorněny i ve formátu KML, který umožnil vytvoření časové animace v aplikaci *Google Earth* (Havlík, 2013).

Hlavním cílem disertační práce **Prostorové a vícerozměrné statistické analýzy epidemiologických dat** od Marka (2005) bylo provedení komplexní prostorové analýzy epidemiologických dat s využitím v současnosti dostupných technologií z oblasti geoinformatiky a prostorové statistiky. Dále bylo určeno pět dílčích cílů, mě zajímaly tyto tři: mapování a celkový popis výskytu onemocnění; průzkum, kvantifikace a vizualizace prostorových a časoprostorových vzorů; identifikace a analýza možných vztahů mezi výskytem onemocnění a vnějšími faktory. Analyzovaná data se týkala infekčního onemocnění kamylobakterií a pocházela z databáze EPIDAT. Pro mapování výskytu onemocnění byla použita metoda pseukartogramu, tečková metoda, metoda gridu (vizualizace pomocí hexagonů) a kartografická anamorfóza. Pro hodnocení prostorového vzoru a identifikaci prostorových shluků na globální i lokální úrovni bylo použito *Moranovo I kritérium*. Při sledování prostorové distribuce na základě kartogramů byla vizuálně patrná asociace hustoty zalidnění a hustoty případů onemocnění. Na území Česka byly ve sledovaném období zjištěny statisticky významné prostorové shluky, z nichž ty nejhomogennější se vyskytovaly na severovýchodní Moravě a ve Slezsku. Shluky nízkých hodnot průměrné incidence, kterých bylo výrazně více, a byly také velmi homogenní, byly umístěny většinou v oblastech s nižší hustotou zalidnění (Marek, 2015).

Cílem diplomové práce **Prostorová analýza inovací firem v Česku** od Švarcové (2017) byla prostorová analýza inovací firem v České republice vycházející z průzkumu *Community Innovation Survey* z let 2008 až 2012. Inovace byly analyzovány za použití prostorové statistiky. Konkrétně byla využita sada nástrojů *Spatial Statistics Tools* obsaženým v programu *ArcMap 10.2* z *ArcGIS Desktop*. Proběhlo zhodnocení prostorové distribuce firem na základě dostupných charakteristik, poté byly podle matematického výpočtu a vybraných charakteristik určeny nejvíce inovativní okresy v ČR a následně byla využita prostorová statistika pro kvantifikaci prostorového vzoru. Na závěr byly určeny společné vlastnosti firem v okresech s pomocí shlukové analýzy na vybrané atributy. V práci byly použity tyto metody prostorové statistiky: prostorová autokorelace, LISA analýza, optimalizovaná hotspot analýza a shluková analýza. Hlavním výsledkem prostorové statistiky bylo zjištění, že vyšší počet inovačních aktivit v obou sledovaných obdobích byl na východě Česka a v centrální části spíše průměrný. Nejnížší počty inovačních aktivit probíhaly na Plzeňsku a částečně v Jihočeském kraji.

Lokální statistiky ukázaly jako odlehlejší polygony vyšších hodnot velká města a shluk vysokých hodnot v oblasti východní Moravy. Komplexní shluková analýza produktových, procesních, marketingových a organizačních inovací poukazuje na dva shluky velmi vysokých inovací ve všech kategoriích táhnoucí se od okresu Ostrava přes Olomouc, hranic se Slovenskem až po Brno (Švarcová, 2017).

3.2.2 Zahraniční studie

V roce 2005 byly v americkém státě Nebraska pány Scottem a Warmerdamem (2005) provedeny rozsáhlejší analýzy kriminality: **Extend Crime Analysis with ArcGIS Spatial Statistics Tools** (Rozšířené analýzy trestné činnosti za pomoci *ArcGIS Spatial Statistics Tools*). Za použití prostorových analýz, které byly součástí sady nástrojů *Spatial Statistics Tools* došlo k posouzení vzorů kriminality, optimalizaci a lokaci zdrojů a zlepšení tísňového volání. Pro analýzu vandalství zde byla využita metoda *Getis-Ord G_i^** normalizována počtem trestných činů v oblasti. Bylo zjištěno, že coldspots odrážely skutečnost environmentálních faktorů, které odrazovaly od trestné činnosti. Závěrem této práce bylo, že s využitím těchto metod, mohou kriminalisté lépe porozumět tomu, kde a proč dochází k trestné činnosti, a že orgány činné v trestním řízení mohou pomocí nich reagovat efektivněji (Scott, Warmerdam, 2005).

Cílem studie Rasama (2014) **Spatial epidemiological techniques in cholera mapping and analysis towards a local scale predictive modelling** (Prostorové epidemiologické techniky při mapování a analýze cholery pro prediktivní modelování na lokální úrovni) bylo prozkoumat prostorový vzor distribuce cholery ve vybrané oblasti od roku 2004 do roku 2008 pomocí GIS a epidemiologických postupů. Pro prostorovou statistiku byla využita sada nástrojů *Spatial Statistics Tools* v *ArcGIS*. Obecný prostorový vzor cholery vykazoval významné shlukování, které bylo důkazem toho, že se nemoc snadno šíří. Ačkoliv epidemie cholery nebyly v okresech nijak kritické, mohly by být endemické v přelidněných oblastech, v nehygienickém prostředí a blízko kontaminované vody. Pro analýzu a identifikaci obecných prostorových vzorů cholery byly použity metody *Average Nearest Neighbor*, *Proximity Analysis* (buffering) a *Standard Deviational Ellipse*. Výsledky tvrdí, že prostorový vzor cholery byl významně shlukový v jihovýchodní části regionu, zejména v oblasti Seri Tanjung. Výsledek prostorových analýz byl takový, jako se dalo očekávat na základě předchozích tvrzení o náchylných místech pro šíření cholery (Rasam, 2014).

V práci Peeterse a kol. (2015) **Getis-Ord's hot-spot and cold-spot statistics as a basis for multivariate spatial clustering of orchard tree data** (*Getis – Ordova metoda identifikace prostorových shluků vysokých a nízkých hodnot* jako základ pro multivariační prostorové shlukování dat o stromech z ovocného sadu) je popsána multivariační prostorová shlukovací metoda pro rozdělení dat z ovocných sadů. Data jsou prostorově měřena statistickou metodou *Getis-Ord G_i^{**}* , následovanou shlukovací metodou *K-means*. Stromy jsou zde rozděleny do prostorově homogenních skupin. Výsledky napomohly zlepšit rozdělení shluků a reprezentaci prostorové struktury. Autoři uvádí, že precizní zemědělství má za cíl trvalou udržitelnost pomocí optimálního obdělávání polí tím, že se řeší prostorová variabilita plodin a jejich prostředí. Prostorová variabilita může být vyhodnocena pomocí shlukové prostorové analýzy, která rozděluje data do homogenních skupin s ohledem na vlastnosti geografické polohy a na prostorové vztahy. Tento přístup lze využít jako nástroj pro precizní management ovocných sadů (Peeters a kol., 2015).

Ve studii od Aghajaniho a kol. (2017) **Applying GIS to Identify the Spatial and Temporal Patterns of Road Accidents Using Spatial Statistics** (Aplikace GIS pro

identifikaci prostorového a časového vzoru silničních nehod za použití Spatial Statistics) byly prostorové statistiky použity k identifikaci míst častých dopravních nehod. Vzhledem k tomu, že dopravní nehody jsou časové jevy, byly použity prostorové statistické metody založené na GIS, které umožňují identifikovat a modelovat místa častých dopravních nehod. Zkoumalo se využití prostorových vzorů a distribuce míst častých dopravních nehod za pomoci časových údajů. Výsledky byly využity k přijetí vhodných opatření ke snížení dopravních nehod. K identifikaci časových vzorů, distribuci dopravních nehod a k hot spot analýzám byla využita *Moranova I* metoda prostorové autokorelace a *Getis-Ord G_i^** statistika. Z výsledků hotspot analýzy se dospělo k závěru, že na severozápadních silnicích, kde je vytíženost dopravy nižší, je počet smrtelných nehod vyšší než na ostatních silnicích (Aghajani a kol., 2017).

4 PRŮZKUM, PŘÍPRAVA A ZPRACOVÁNÍ DAT

Před aplikací metod prostorové statistiky bylo nejprve nutné prozkoumat primární data a jejich vlastnosti. Dataset adresních bodů s názvem „Statistické budovy (vchody k bytům)“ poskytl Český statistický úřad (ČSÚ). Dalším důležitým krokem byla analýza a výběr vhodných atributů, které pak dále vstupují do výpočtů prostorových analýz.

4.1 Analýza primárních dat a výběr vhodných atributů

První problém, na který autor narazil při nahrání dat do prostředí *ArcMap 10.4* z *ArcGIS Desktop* byla jejich samotná vizualizace. Zobrazení skoro tři miliónů adresních bodů na území Česka tak, aby od sebe byly jednotlivé body rozeznatelné a aby se vzájemně nepřekrývaly, by vyžadovalo použití nesmírně velkého měřítka. Použití takového měřítka není reálné. Proto se autor práce rozhodl data agregovat a pomocí statistických přepočtů převést adresní body na menší počet areálů (polygonů), které budou svými hodnotami reprezentovat adresní body v nich. Tyto agregovaná data umožňují vhodnou vizualizaci výsledků.

Při prvním rychlém náhledu na data bylo zjištěno, že více než polovina adresních bodů má v atributech uvedenou shodnou adresu a stejné hodnoty několika dalších atributů. Přitom poloha těchto bodů byla od uvedené adresy značně odlišná (celé území ČR). Z celkem 2 809 711 bodů se většina z nich (1 998 825 – 71 %) chybně shodovala v atributech, navíc většina těchto záznamů obsahovala u velké části atributů pouze nulové hodnoty. To vedlo k podezření, že data nejsou v pořádku. Bylo tedy nutné chybu nahlásit pro obdržení nových záznamů. Konkrétně se opakovaly hodnoty atributů adresního bodu s adresou: Boženy Němcové 740, Choceň. Proto se autor práce po konzultaci s vedoucím práce společně obrátili na ČSÚ, aby data překontrolovali. Český statistický úřad potvrdil chyby v datech a následně zaslal novou databázi adresních bodů, která už chybu s atributy neobsahovala. Použitá data jsou aktuální k 1. červenci 2016.

Shape *	IDOB	NAZEV UL A	NAZ CAST D	CIS D	JTSK X	JTSK Y
Point	1020554398	B. Němcové	Choceň	740	1133956.08	558414
Point	1024649733	B. Němcové	Choceň	740	1120890.19	547781.74
Point	1024651762	B. Němcové	Choceň	740	1122027.02	547996.72
Point	1000146910	Budějovická	Michle (Praha 4)	743	1048034	741368
Point	1001764170	Na Cibulce	Beroun-Město	871	1053124	770429
Point	1001764170	Chmelenského	Beroun-Město	871	1053124	770429
Point	1033757047	B. Němcové	Choceň	740	1050726	735820
Point	1033783871	B. Němcové	Choceň	740	1044620	744499
Point	1033935875	B. Němcové	Choceň	740	1050723	735737
Point	1033686646	B. Němcové	Choceň	740	1038364	735016
Point	1033783196	B. Němcové	Choceň	740	1045661	748690
Point	1032774304	B. Němcové	Choceň	740	1190334	592015
Point	1032774304	B. Němcové	Choceň	740	1190310	592023
Point	1032724421	B. Němcové	Choceň	740	1047248	750904
Point	1033947237	B. Němcové	Choceň	740	1121733	549115
Point	1033947237	B. Němcové	Choceň	740	1121710	549111
Point	1034709321	B. Němcové	Choceň	740	1121089	549574
Point	1034709321	B. Němcové	Choceň	740	1121098	549554
Point	1024440427	B. Němcové	Choceň	740	1121592	545462
Point	1024440427	B. Němcové	Choceň	740	1121594	545468
Point	1024667391	B. Němcové	Choceň	740	1121520	547312
Point	1024667391	B. Němcové	Choceň	740	1121523	547337
Point	1024651762	B. Němcové	Choceň	740	1122040	548002
Point	1024649733	B. Němcové	Choceň	740	1120883	547792

Obr. 4.1 Náhled atributové tabulky původních dat s chybou v atributech.

Dataset adresních bodů s názvem „Statistické budovy (vchody k bytům)“ obsahuje bodovou vrstvu s celkem 2 815 224 body. Každý bod obsahuje 63 atributů, většina z nich však obsahuje kvalitativní nominální (jedinečné) hodnoty, které nelze porovnat, a proto ani nejsou vhodné pro prostorové analýzy. Často se jedná o číselné identifikátory. Tyto nevhodné atributy nebyly tedy při analýzách využity. Některé další atributy zase nejsou objektivní, jelikož u třetiny i více prvků chybí hodnoty (objevují se hodnoty NULL nebo 0), proto bylo rozhodnuto provést i prostorovou analýzu nulových hodnot atributů zkoumaných bodů. Vybrané atributy vhodné pro provádění prostorových analýz obsahovaly buďto absolutní číselné hodnoty (např. počet evidovaných obyvatel) anebo číselné kódy reprezentující jednotlivé kategorie atributu (např. klasifikace stavebních děl). Rozklíčování číselných kódů je možné pomocí číselníků, které lze stáhnout z databáze metainformací, která je volně dostupná z internetových stránek ČSÚ: <http://apl.czso.cz/iSMS/cislist.jsp>

23 vybraných atributů s potenciálem pro prostorové statistiky:

1. PODPLOBUD – podlahová plocha budovy
2. OBEPROBUD – obestavěný prostor budovy
3. ZASTPLOBUD – zastavěná plocha budovy
4. JOBDVYS – období výstavby nebo rekonstrukce domu (číselník 3044)
5. JVLASTD – typ vlastníka domu (číselník 3049)
6. JPPODLA – počet nadzemních podlaží domu podle SLDB (číselník 3047)
7. JUSTOPE – způsob vytápění budovy (číselník 3048)
8. JVODOVD – připojení budovy na vodovod (číselník 3050)
9. POCPODBUD – počet všech podlaží budovy
10. DRUHVLABUD – typ vlastníka budovy (číselník 3198)
11. JPLYN – připojení budovy na plyn (číselník 3046)
12. ZPVYTBU – způsob vytápění budovy (číselník 3205)
13. JMATERZ – materiál nosných zdí budovy (číselník 3043)
14. PRIBUDVOD – připojení budovy na vodovod (číselník 3204)
15. JKANAL – připojení budovy na odpad (číselník 3042)
16. VYBUDVYT – vybavenost budovy výtahem (číselník 3206)
17. TVYBU – způsob využití budovy (číselník 0075)
18. ZPVYBU – způsob využití stavebního objektu (číselník 0076)
19. KSD_CC_CZ – klasifikace stavebních děl (číselník 5631 až 5635)
20. KATCBYT – klasifikace počtu bytů podle SLDB (číselník 3034)
21. BUDOBYEV – počet evidovaných obyvatel
22. BUDOBYTSL – počet obyvatel stálého pobytu
23. BUDOBYOSL – počet obyvatel občasného pobytu

Prostřednictvím zkratk, uvedených před názvem atributu, bude dále v textu na atributy odkazováno.

Nepoužitelné atributy záznamů byly odstraněny pomocí editace atributové tabulky v prostředí aplikace *ArcMap* pro urychlení další práce a přehlednost dat. U každého záznamu zůstalo tedy pouze 23 vybraných atributů a několik dalších potřebných (např. atributy s polohou, identifikační atribut apod.). V dalších krocích se pracovalo s takto

pročištěnými daty. V průběhu práce vznikaly i nové atributy, které byly vytvořeny při statistických přepočtech dat anebo vytvořeny autorem práce a vypočítány pomocí nástroje *Field Calculator*.

4.2 Analýza počtu nulových hodnot u vybraných atributy

Při prvním průzkumu nevykazovala nová „očistěná“ data hrubé chyby podobného charakteru (opakující se chyba hodnot atributů popsána v kapitole 4.1). Ovšem velké množství záznamů u většiny atributů obsahovalo nulové hodnoty, které reprezentují neznámou (neuvedenou) hodnotu.

U kvantitativních atributů se za nulovou považovala hodnota 0 (např. atribut POCPODBUD nemůže nikdy mít hodnotu nula, jelikož každá budova zabírá určitou plochu). Problém byl s kvantitativními atributy, které udávají počet obyvatel žijících na daném adresním místě, jedná se o poslední tři uvedené v seznamu vybraných atributů. Jelikož existují i adresní místa budov, kde nikdo nebydlí, mohou tyto atributy mít hodnotu nula. Nelze tedy s jistotou určit, zda nula reprezentuje budovu bez obyvatel anebo budovu s neznámým počtem obyvatel.

U vybraných kvalitativních atributů se za nulové považovaly hodnoty NULL, které nevyovídají žádnou informaci a představují jakousi „neznámou“ v datech.

Proto bylo rozhodnuto provést pro vybrané atributy analýzu počtu nulových hodnot atributů tak, že se pro každý z 23 vybraných atributů (kromě tří atributů s počty obyvatel) vypočítá procentuální zastoupení nulových hodnot z celkového počtu hodnot. Výsledky této analýzy jsou shrnuty v tabulce 4.2.

Tab. 4.2 Výsledky analýzy počtu nulových hodnot u vybraných atributů

NÁZEV ATRIBUTU	KÓD ATRIBUTU	POČET CHYB	POMĚR CHYB
Podlahová plocha budovy	PODPLOBUD	2589048	91,97%
Obestavěný prostor budovy	OBEPROBUD	2529061	89,84%
Zastavěná plocha budovy	ZASTPLOBUD	2504698	88,97%
Období výstavby nebo rekonstrukce domu	JOBDVYS	614142	21,82%
Typ vlastníka domu	JVLASTD	612274	21,75%
Počet nadzemních podlaží domu podle SLDB	JPPODLA	611979	21,74%
Způsob vytápění budovy	JUSTOPE	611979	21,74%
Připojení budovy na vodovod	JVODOVD	611979	21,74%
Počet všech podlaží budovy	POCPODBUD	534264	18,98%
Typ vlastníka budovy	DRUHVLABUD	527939	18,75%
Připojení budovy na plyn	JPLYN	508398	18,06%
Způsob vytápění budovy	ZPVYTBUD	507166	18,02%
Materiál nosných zdí budovy	JMATERZ	505086	17,94%
Připojení budovy na vodovod	PRIBUDVOD	492618	17,50%
Připojení budovy na odpad	JKANAL	490276	17,42%
Vybavenost budovy výtahem	VYBUDVYT	407409	14,47%
Způsob využití budovy	TVYBU	35485	1,26%
Způsob využití stavebního objektu	ZPVYBU	34820	1,24%
Klasifikace stavebních děl	KSD_CC_CZ	32617	1,16%
SLDB: Klasifikace počtu bytů	KATCBYT	26	0,0009%

Počet evidovaných obyvatel	BUDOBYEV	???	???
Počet obyvatel stálého pobytu	BUDOBYTSL	???	???
Počet obyvatel občasného pobytu	BUDOBYOSL	???	???
TYP ATRIBUTU	BARVA ATRIBUTU		
kvalitativní			
kvantitativní			

Jak lze vyčíst z tabulky 4.2, vůbec nejvíce nulových hodnot atributů (přes 80 %) se vyskytovalo u atributů PODPLOBUD, OBEPROBUD a ZASTPLOBUD. Tyto atributy nejsou rozhodně vhodné pro vstup do prostorových analýz, jelikož tak velký počet nulových (neznámých) hodnot by značně zkreslil výsledek analýzy. Naopak nejmenší počet nulových hodnot (méně než 2 %) se vyskytoval u atributů TVYBU, ZVYBU a KSD_CC_CZ. Tyto atributy mají největší potenciál pro prostorové analýzy a jejich vstup do prostorových analýz by mohl přinést zajímavý výsledek. Jediný vhodný kvantitativní atribut se zdá být POCPODBUD s 19 % výskytu nulových hodnot.

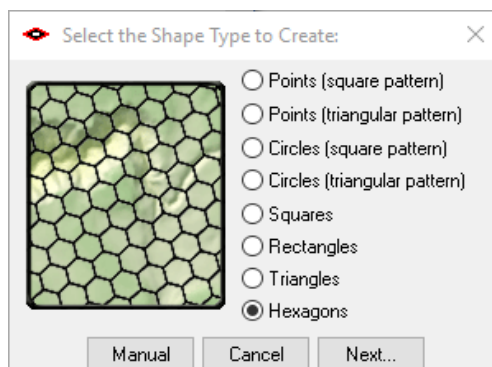
4.3 Agregace dat

Jak je uvedeno na začátku podkapitoly 4.1, data bylo nutné agregovat tak, aby byla jejich vizualizace přehledná a čitelná. Metoda agregace obnáší dva důležité kroky. Prvním z nich je výběr areálů, do kterých se budou data přepočítávat. Je důležité si určit a odůvodnit jaký bude ideální tvar a velikost areálů při agregaci. Autor práce se rozhodl pro agregaci využít pravidelnou geometrickou síť. Postup její tvorby a její vlastnosti jsou popsány v následující podkapitole.

4.3.1 Tvorba hexagonové sítě

Základní informace o využití volně dostupné extenze *Repeating Shapes* jsou popsány podkapitole 2.3.1. Tato extenze umožňuje generovat pravidelné sítě s velkým množstvím tvarů buněk a definovat jejich velikost a orientaci.

Při postupu tvorby sítě se nejprve zvolí, nad kterou vrstvou nebo nad kterými prvky vrstvy se bude generovat pravidelná geometrická síť. Dále se zvolí tvar mřížky. Extenze nabízí mnoho základních tvarů pro generování geometrických sítí od bodů, čtverců, trojúhelníků až po hexagony (Obr. 4.3.1.1). Za nejvhodnější tvar pro účely této práce byl zvolen hexagon, jelikož má více sousedů než čtverec nebo trojúhelník. Další kladnou vlastností hexagonu (šestiúhelníku) je, že středy všech sousedních polí jsou od středu daného pole stejně vzdálené.

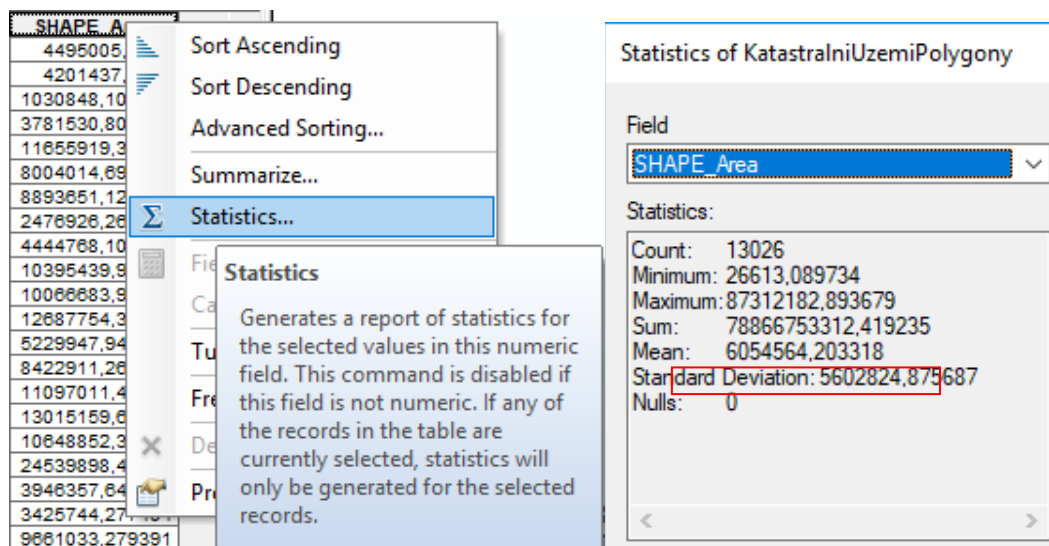


Obr. 4.3.1.1 Základní tvary pro generování pravidelné sítě nabízené v extenzi *Repeating Shapes* (Jenness, 2012).

V následujícím kroku bylo třeba nastavit parametry základní buňky (hexagonu) pro novou pravidelnou síť. Extenze *Repeating Shapes* umožňuje nastavení velikosti buňky nejen podél délky hrany, šířky buňky, ale i podle její plochy, čehož bylo využito. To byl také důvod, proč byla pro generování pravidelné sítě využita právě tato extenze. Dále je možné změnit orientaci sítě pomocí zvolení úhlu natočení.

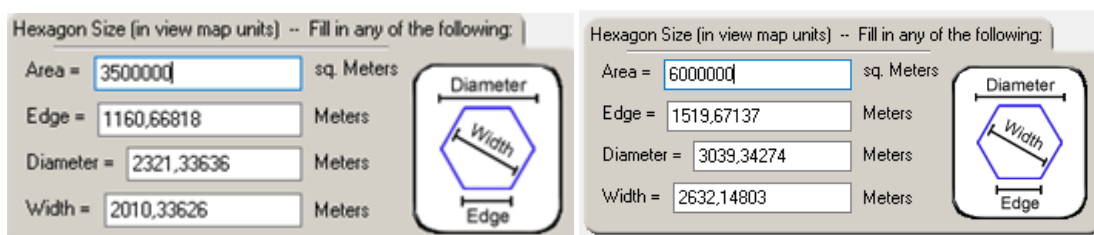
Velikost hexagonu

Při tvorbě pravidelné geometrické sítě bylo velmi důležitým krokem zvolení velikosti hexagonu. Obecně platí, že čím menší je buňka, tím přesněji lze zachytit výskyt jevu a průběh hranic studovaného území. Dále platí, že čím menší je velikost buňky, tím se zvětší nároky na paměťový prostor (Rapant, 2006). Příliš malé buňky mohou být často špatně rozeznatelné, naopak příliš velké buňky mohou výsledky výrazně zkreslit. Marek (2007) zvolil ve své práci velikost šestiúhelníku tak, aby odpovídala ploše průměrného katastrálního území obce. Jelikož nebyla nalezena žádná další lepší metoda, byl tento způsob určení velikosti šestiúhelníku využit také v této práci. Průměrnou plochu katastrálních území obcí v Česku lze zjistit pomocí atributové tabulky v *ArcMap*, kde stačí zobrazit automaticky vypočítané statistické údaje pro sloupec *SHAPE_area* mezi kterými je i průměr (Mean – červeně orámovaný na Obr. 4.3.1.2).



Obr. 4.3.1.2 Postup zobrazení automaticky vypočítaných statistických údajů pro daný atribut.

Vypočítaná plocha byla zaokrouhlena na 6 km². Tato plocha by však nemusela být reprezentativní pro prostorové analýzy prováděné na menším území (např. kraji). Proto byla ještě zjištěna obdobným způsobem jako u katastrálních území průměrná plocha základních sídelních jednotek (ZSJ), které představují nejmenší dostupné administrativní jednotky v Česku. Průměrná plocha ZSJ v České republice byla zaokrouhlena na 3,5 km². Parametry generovaných hexagonů uvádí Obr. 4.3.1.3.



Obr. 4.3.1.3 Velikostní parametry hexagonů pro generování pravidelných sítí.

Byly vygenerovány 2 pravidelné šestiúhelníkové sítě s obsahem šestiúhelníků 3,5 km² a 6 km². Extenze vygeneruje pravidelné sítě ve tvaru obdélníku, který pokrývá zvolené studované území. Dále je nutné pravidelné sítě ořezat tak, aby měly hranice shodné se zkoumaným územím. Toho se dosáhlo snadno pomocí nástroje *Clip*, kdy se za vrstvu, která má být ořezána, zvolila vygenerovaná síť a za ořezovou vrstvu pak daná zkoumaná administrativní jednotka (území státu, území kraje).

Výběr vhodného měřítka

Bylo potřeba zjistit, jaké měřítko bude vhodné použít pro vizualizace v rozsahu území Česka a zvoleného území pro podrobnější průzkum (kraje) za použití hexagonových sítí s plochou hexagonu 3,5 km² a 6 km². Touto problematikou se ve své práci zabývá Zajícová (2017), ta ve své práci testuje a srovnává vizualizace pro různé velikosti gridových struktur v různých měřítkách, kdy pro všechny testované velikosti buněk bylo vytištěno porovnání 15 různých měřítek. Autorka uvádí, že na základě tohoto porovnání bylo vizuální analýzou určeno nejmenší možné měřítko, při kterém jsou buňky ze vzdálenosti 45 cm rozlišitelné. Na závěr byla v její práci doporučena měřítko pro optimální čitelnost. Jednou z testovacích velikostí byl hexagon s plochou 6,5 km², ten lze svou velikostí přirovnat k hexagonu s plochou 6 km², který se zvolil pro účely této práce. Doporučeným minimálním měřítkem pro hexagon s plochou 6,5 km² pro území Česka bylo měřítko 1 : 2 000 000. Při tomto měřítku se vejde Česká republika v tištěné formě na formát stránky A4 a buňky jsou dobře čitelné (Zajícová, 2017).

Na základě těchto doporučení bylo rozhodnuto použít pro vizualizaci prostorových analýz na území Česka hexagonovou síť s plochou hexagonu 6 km² a měřítko 1 : 2 000 000.

4.4 Statistický přepočítání dat a hodnot jejich atributů

Závěrečnou částí zpracování dat byla jejich agregace, tedy přepočítání hodnot vybraných atributů vhodných pro prostorové analýzy do připravených pravidelných šestiúhelníkových sítí a do vybraných jednotek administrativního členění Česka.

Vhodnou a nenáročnou metodou pro přiřazování statistických informací do buněk pravidelné geometrické sítě nebo administrativních jednotek je agregace (bottom-up/down-top metoda). Jde o metodu nejjednodušší a zároveň poskytující nejlepší výsledky, kdy typicky jsou na vstupu adresně lokalizovaná data (Klauda, 2016).

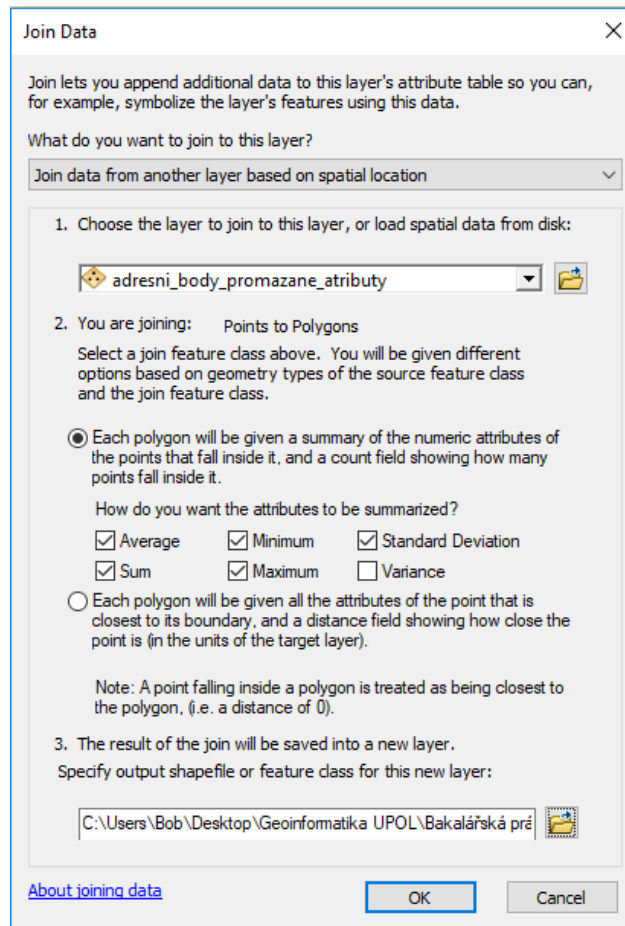
Z nabízených možností byla nakonec využita metoda *Join Data*. I když neumožňuje výběr atributů pro statistické výpočty (ty jsou provedeny pro všechny atributy), mohla být použita, jelikož data vstupující do agregace byla již očištěna o nepotřebné atributy.

Při použití nástroje *Join Data* se nejprve zvolí, zdali se mají data připojit z atributové tabulky na základě shodných hodnot daných sloupců, anebo jako v našem případě, zdali mají být data připojena z jiné vrstvy na základě prostorové lokace (*Join data from another layer based on spatial location*). Zvolí se vrstva, ze které se budou hodnoty přepočítávat, což je v našem případě bodová vrstva adresních míst s vybranými atributy. V dalším kroku se ze dvou nabízených zvolí možnost první, kdy bude každý nový polygon charakterizován statistickým souhrnem numerických atributů bodů, které do polygonu spadají. Mezi statistické souhrny, které tento nástroj nabízí patří:

- **Average** – vypočítá průměrnou hodnotu atributu všech bodů v každém polygonu
- **Sum** – vypočítá celkovou hodnotu atributu všech bodů v každém polygonu
- **Minimum** – vypočítá nejmenší hodnotu atributu všech bodů v každém polygonu

- **Maximum** – vypočítá největší hodnotu atributu všech bodů v každém polygonu
- **Standard Deviation** – vypočítá směrodatnou odchylku hodnot atributu všech bodů v každém polygonu
- **Variance** – vypočítá rozptyl hodnot atributu všech bodů v každém polygonu

Pro každý polygon ve výstupní vrstvě se vždy vytvoří i nový atribut *Count_*, který udává kolik bodů vstupní vrstvy spadalo do daného polygonu. Vrstvy vstupující do statistického souhrnu musí mít číselný formát, jelikož text a ostatní datové typy nejsou nástrojem *Join Data* podporovány. Jako poslední se zadá název a úložiště pro novou polygonovou vrstvu. Všechna použitá nastavení jsou viditelná na Obr. 4.4.



Obr. 4.4 Nastavení metody Join Data při agregaci adresních bodů.

4.5 Prostorová analýza průměrného počtu nulových hodnot u vybraných atributů adresních bodů Česka

Jak vyplývá z podkapitoly 4.2, data obsahovala atributy s velkým množstvím nulových hodnot atributů. Autor této práce se proto po konzultaci s vedoucím rozhodl provést prostorovou analýzu průměrného počtu nulových hodnot u vybraných atributů. Ta měla za úkol zjistit, kde se nachází adresní body s největším počtem nulových hodnot atributů a naopak, kde se nachází adresní body se všemi známými hodnotami u vybraných atributů. Pro správnou a čitelnou vizualizaci výsledků bylo rozhodnuto, že do analýzy budou vstupovat už data agregovaná do vygenerované hexagonové sítě s plochou hexagonu 6 km². Následuje podrobný postup provedení této analýzy.

Prvním údaj, který byl potřeba zjistit bylo, kolik nulových hodnot atributů se u každého adresního bodu nachází. Analýze podléhalo prvních 20 vybraných atributů,

kteře jsou uvedené v podkapitole 4.1. Analýza nebyla provedena u posledních tři atributů uvedených v seznamu. Jde o atributy, které nesou informaci o počtu obyvatel trvale nebo dočasně žijících na daném adresním místě. Jelikož však existují i adresní místa nebytových (komerčních) budov, kde nikdo trvale ani dočasně nebydlí, nelze považovat hodnotu 0 u těchto atributů za chybu. Správnost těchto údajů by mohla být porovnána pomocí srovnání s daty ČSÚ z téhož kalendářního roku 2016. V této práci se však tato problematika dále neřeší a tyto tři atributy jsou z analýzy chybovosti dat vynechány.

Každý z dvaceti atributů byl analyzován postupně jednotlivě a bylo potřeba k němu vytvořit nový atribut nesoucí informaci o nulové hodnotě. Začalo se podle seznamu atributem TVYBU, ke kterému se vytvořil nový atribut TVYBU_CH ve formátu *Short Integer*. Využil se nástroj *Select by Attributes*, který umožňuje vybírat prvky na základě hodnot jejich atributů. Nástroj se nastavil tak, aby označil všechny prvky, které mají v atributu TVYBU nulovou hodnotu (hodnotu NULL). V atributové tabulce pak stačilo zobrazit pouze označené prvky a přiřadit jim pomocí nástroje *Field Calculator* do atributu TVYBU_CH hodnotu 1. Potom se naopak nechali pomocí nástroje *Switch Selection* zobrazit neoznačené prvky, kterým byla opět pomocí nástroje *Field Calculator* do atributu TVYBU_CH přiřazena hodnota 0. Všechny adresní body měli pak u atributu TVYBU_CH buďto hodnotu 1 nebo hodnotu 0, kdy hodnota 1 značila výskyt nulové (neznámé) hodnoty a hodnota 0 atribut se známou hodnotou.

Tento postup byl zopakován i pro následujících 19 atributů. S tím rozdílem, že u kvantitativních atributů se považovala za nulovou hodnotu hodnota 0, ta byla v novém atributu nahrazena hodnotou 1 a zbytek hodnot naopak zase 0, aby opět platilo, že hodnota 1 značí atribut s nulovou (neznámou) hodnotou a hodnota 0 atribut se známou hodnotou. Pro zjištění celkového počtu nulových hodnot atributů u jednotlivých adresních bodů bylo nutné vytvořit ještě jeden nový atribut ve formátu *Short Integer*, který byl pojmenován CHYBACHYB. Sloupec tohoto nového atributu se označil a pomocí nástroje *Field Calculator* a vzorce: $CHYBACHYB = [TVYBU_CH] + [ZPVYBU_CH] + [KSD_CH] + [JKANAL_CH] + [JMATERZ_CH] + [JOB DVYS_CH] + [JPLYN_CH] + [JPPODLA_CH] + [JUSTOPE_CH] + [JVLASTD_CH] + [JVODOVD_CH] + [DRUHVLABUD_CH] + [ZASTPLOBUD_CH] + [OBEPROBUD_CH] + [PODPLOBUD_CH] + [POCPODBUD_CH] + [PRIBUDVOD_CH] + [ZPVYTBUD_CH] + [VYBUDVYT_CH] + [KATCBYT_CH]$ byly vypočítány i nové hodnoty.

Nové atributové pole CHYBAHYB pak obsahovalo údaje o počtu nulových hodnot atributů u adresních bodů. Vyskytovaly se i adresní body, které neobsahovaly nulové hodnoty u žádného s vybraných atributů. Poloha těchto „dokonalých“ bodů by mohla být klíčová pro zvolení menšího území na kterém se budou provádět podrobnější prostorové analýzy. Autor práce se proto rozhodl, vyexportovat tyto „dokonalé“ adresní body do nové bodové vrstvy.

Stačilo využít nástroj *Selection by Attributes*. Vybraly se všechny adresní body, které měli v novém atributovém poli CHYBACHYB hodnotu 0. Označily se tak všechny „dokonalé“ adresní body, kterých bylo dohromady 108 654. Pak stačilo využít nástroje *Export Data*, ve kterém se nastavilo, aby byly exportovány pouze vybrané prvky. Vznikla tak nová bodová vrstva „dokonalých“ adresních bodů. Označení „dokonalé“ adresní body bude použito pro zkrácení, myslí se tím adresní body bez nulových hodnot u vybraných atributů adresních bodů.

Pro další postup bylo potřeba pracovat pouze s adresními body s nulovými hodnotami u atributů. Jejich výběr byl v daný moment velice snadný, jelikož stačilo využít v atributové tabulce nástroje *Switch Selection*. Ten označí všechny aktuálně

neoznačené body, a naopak ty označené „odznačí“. Stačilo opět exportovat označené prvky nástrojem *Export Data*. Tím se získala nová vrstva adresních bodů s počty nulových hodnot u vybraných atributů.

Pro přehlednou vizualizaci prostorové analýzy průměrného počtu nulových hodnot u vybraných atributů bylo potřeba výsledky agregovat a přepočítat hodnoty atributů z adresních bodů do pravidelné hexagonové sítě. Cílem bylo vytvoření pravidelné hexagonové sítě tak, aby každý hexagon nesl údaj o počtu adresních bodů s nulovými hodnotami, o počtu „dokonalých“ adresních bodů a o průměrném počtu nulových hodnot u vybraných atributů v adresních bodech, které do hexagonu spadají.

Při agregaci výsledků prostorové analýzy průměrného počtu nulových hodnot u vybraných atributů byl opět využit nástroj *Join Data*, který přepočítal výsledné hodnoty z jednotlivých adresních bodů do hexagonů. Na hexagonovou síť se nejprve připojila vrstva „dokonalých“ adresních bodů. V nastavení se nezadával žádný ze statistických souhrnů, protože bylo potřeba získat pouze údaj o počtu „dokonalých“ adresních bodů v hexagonu, který se vypočítal do nového atributu *bez_nulovych*. Vznikla tak nová vrstva hexagonové sítě, na kterou se pak dále „joinovaly“ adresní body s nulovými hodnotami atributů. Při nastavení nástroje *Join Data* se zvolil výpočet pouze dvou statistických souhrnů – *SUM* (suma) a *Average* (průměr). Vznikla tak nová vrstva, kdy každý hexagon nesl informaci o počtu adresních bodů s nulovými hodnotami atributů a o celkovém počtu nulových hodnot atributů adresních bodů, které do daného hexagonu spadali. Kromě těchto údajů nesl každý hexagon informace o počtu nulových hodnot pro každý z dvaceti zkoumaných atributů zvlášť.

Pro správné porovnání nulových hodnot atributů v hexagonech nestačilo pracovat pouze s hodnotami celkového počtu nulových hodnot atributů v hexagonu. Obecně platí, že čím větší počet chybných adresních bodů se v hexagonu nachází, tím větší počet chybných atributů těchto bodů se může u hexagonu vyskytovat. Pro správné vzájemné porovnání bylo potřeba zjistit průměrný počet nulových hodnot vybraných atributů v adresních bodech daného hexagonu. Vytvořilo se nové pole *pomer* a pomocí *Field Calculatoru* a vzorce: $[Sum_CHYBACHYB] / [s_nulovymi]$ se vypočítal průměrný počet nulových hodnot atributů adresních bodů pro každý hexagon zvlášť. Tyto hodnoty pak byly porovnávány a vizualizovány jako výsledek prostorové analýzy průměrného počtu nulových hodnot u vybraných atributů. Vizualizace této metody proběhla metodou kartogramu.

4.5.1 **Výsledky a vyhodnocení prostorové analýzy nulových hodnot**

Výsledkem této prostorové analýzy byly vizualizace průměrného počtu nulových hodnot u adresních bodů a počtu adresních bodů bez nulových hodnot u vybraných atributů v jednotlivých hexagonech pravidelné sítě:

- Průměrný počet nulových hodnot atributů adresních bodů Česka v roce 2016 (Příloha 1)
- Adresní body Česka bez nulových hodnot atributů (Příloha 2)

Z výsledných vizualizací je vidět, že v krajských městech je počet nulových hodnot vybraných atributů průměrný (4 až 7 nulových hodnot atributu jednoho adresního bodu). Výjimkou jsou krajská města Ústí nad Labem, Jihlava a Brno, kde je počet nulových hodnot atributů nadprůměrný (7 a více). Naopak v Praze, kde se bude nacházet také největší počet adresních bodů, je výskyt nulových hodnot atributů podprůměrný (4 a méně). Největší množství nulových hodnot se nachází ve venkovských

oblastech v okolí velkých měst a v horských oblastech u hranic republiky. Naopak adresní body bez nulových hodnot vybraných atributů se vyskytují ve velkém počtu ve velkých městech a jejich aglomeracích (blízké okolí velkých měst).

4.6 Výběr území pro podrobnější průzkum

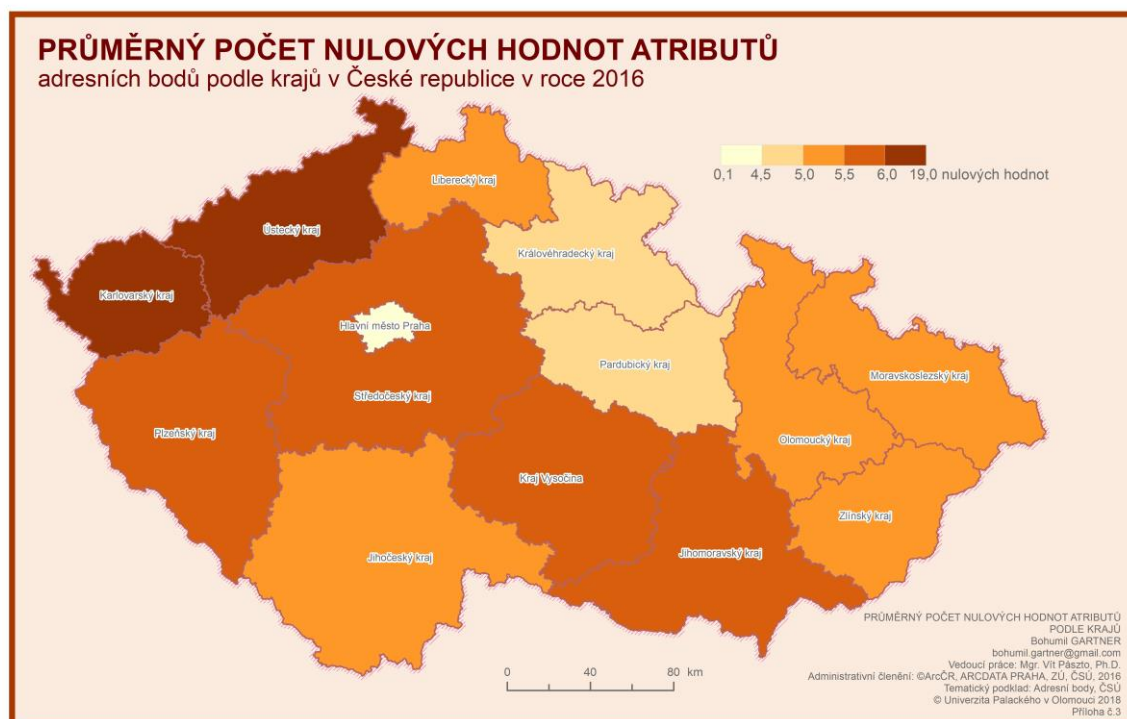
Zadáním práce bylo provést prostorové statistiky pro území Česka i pro vybrané lokality. Po konzultaci s vedoucím práce se nakonec vybrala jedna lokalita pro podrobnější prostorové statistiky. Danou lokalitou se měl stát jeden z krajů Česka.

K nalezení optimálního kraje pro podrobnější prostorové statistiky se využily výsledky prostorové analýzy průměrného počtu nulových hodnot vybraných atributů adresních bodů Česka (Příloha 1). Vypočítal se průměrný výskyt nulových hodnot atributů adresních bodů pro jednotlivé kraje a za optimální byl zvolen kraj s nejmenším průměrným počtem nulových hodnot u vybraných atributů s potenciálem pro podrobnější prostorové analýzy.

4.6.1 Vizualizace porovnání krajů a vyhodnocení

Seřazení krajů od nejvhodnějšího po nejméně vhodný zobrazuje Tab. 4.6.1. Nejmenší průměrný počet nulových hodnot u vybraných atributů ze všech krajů měl kraj Hlavní město Praha. Ten byl proto zvolen za optimální území pro podrobnější prostorové analýzy. Výsledky porovnání jsou také vizualizovány pomocí mapy:

- Průměrný počet nulových hodnot atributů adresních bodů podle krajů v České republice v roce 2016 (Příloha 3)



Obr. 4.6.1 Náhled přílohy č.3 – Průměrný počet nulových hodnot atributů adresních bodů podle krajů v České republice v roce 2016

Obr. 4.6.1 (Příloha 3) zobrazuje, že na západě republiky se nachází v průměru více nulových hodnot atributů než na východě. Nejnižší průměrný počet nulových hodnot má kraj Hlavní město Praha, který byl proto zvolen za vhodné území pro podrobnější prostorové analýzy.

Tab. 4.6.1 Průměrný počet nulových hodnot u vybraných atributů adresních bodů v krajích Česka

NÁZEV KRAJE	PRŮMĚRNÝ POČET CHYBNÝCH ATRIBUTŮ
Hlavní město Praha	4,22
Královéhradecký kraj	4,76
Pardubický kraj	4,92
Olomoucký kraj	5,02
Moravskoslezský kraj	5,06
Zlínský kraj	5,17
Liberecký kraj	5,20
Jihočeský	5,44
Středočeský	5,64
Plzeňský	5,70
Vysočina	5,74
Jihomoravský	5,80
Karlovarský	6,30
Ústecký	6,64

4.7 Statistický přepočítání dat a hodnot jejich atributů pro kraj Hlavní město Praha

I pro vizualizaci podrobnějších prostorových analýz ve vybraném kraji musely být adresní body agregovány do pravidelné hexagonové sítě. Velikost plochy hexagonu v předpřipravené síti 3,5 km² (průměrná plocha ZSJ v Česku) se zdála být pro území Prahy zbytečně velká, proto byla zjišťována průměrná plocha ZSJ v Praze, která se zaokrouhlila na hodnotu 500 000 m². Následně byla vytvořena nová pravidelná hexagonová síť pro území Prahy s plochou hexagonu 500 000 m². Tato síť byla pro vizualizaci výsledků prostorových analýz mnohem vhodnější.

Agregace a statistický přepočítání dat proběhl opět pomocí nástroje *Join Data* obdobně jako u předchozích agregací.

5 PROSTOROVÁ STATISTIKA PRO ÚZEMÍ ČESKA

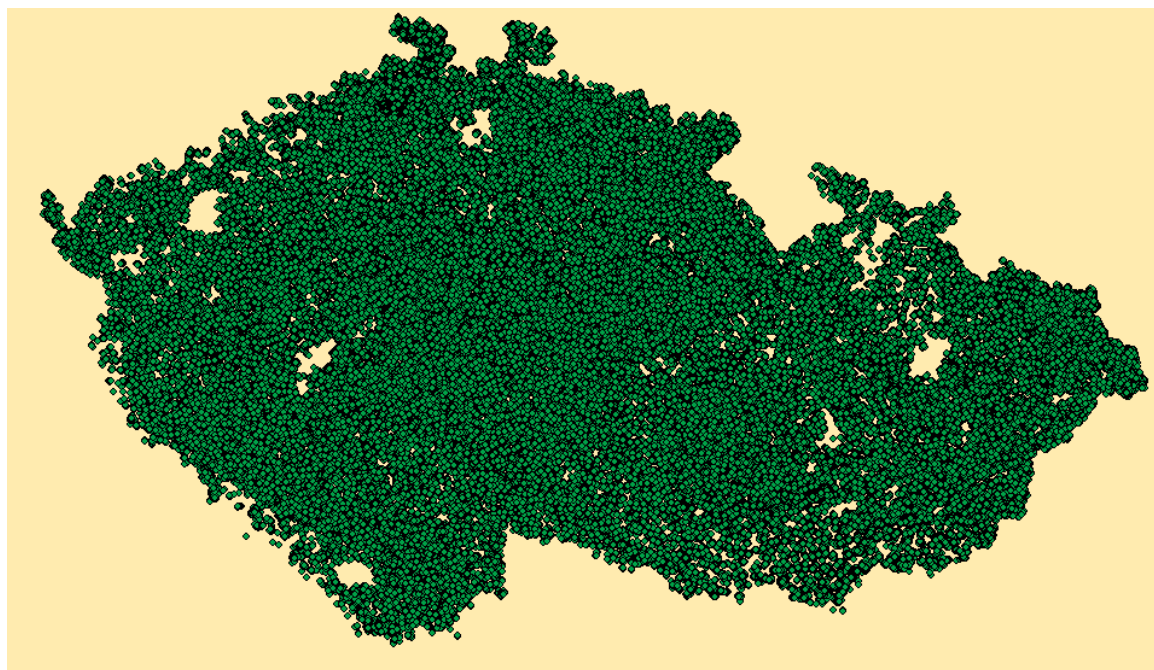
Dalším a zároveň hlavním krokem této práce byl výpočet prostorové statistiky nad upravenými daty reprezentujícími adresní body v Česku. Tato rozsáhlá kapitola obsahuje podrobný popis a postup při výpočtu prostorové statistiky na území Česka. Jsou zde vysvětleny jednotlivé použité metody a kroky při nastavování jejich parametrů. V programu *ArcMap 10.4* z *ArcGIS Desktop* bylo využito nástrojů *Spatial Statistics Tools*.

Zdrojem informací uvedených u jednotlivých metod a sad nástrojů je ESRI nápověda, která je volně dostupná v anglickém jazyce přímo v programu *ArcMap 10.4* z *ArcGIS Desktop* nebo na serveru ArcGIS Online (ESRI, 2017), a také publikace Mitchella (2005) *The ESRI guide to GIS analysis: Volume 2: Spatial measurements and statistics*.

5.1 Adresní body Česka

Jedna z hlavních prostorových charakteristik geografických dat při analýzách v GIS je jejich výskyt. Jinak řečeno při prostorové statistice je důležité umět si představit, kde a v jakém počtu se objekty zájmu vyskytují.

Na území Česka se nachází celkem 2 815 224 adresních bodů. Jak již bylo dříve zmíněno v rešeršní části práce, vzhledem k počtu zpracovávaných dat a charakteru jejich rozmístění nebylo možné pro jejich vizualizaci použít metodu bodových znaků. I sám software *ArcMap 10.4* z *ArcGIS Desktop* měl problém s vykreslením všech adresních bodů po importování dat. Proto se autor práce rozhodl použít pro vizualizaci výskytu adresních bodů areálovou metodu. Pro použití této metody bylo potřeba zvolit si areály (polygony), do kterých se budou bodová data agregovat. Autor práce se rozhodl využít více typů areálů pro agregaci a následnou vizualizaci.



Obr. 5.1 Náhled načtených dat nevhodně vizualizovaných bodovou metodou v prostředí aplikace ArcMap.

5.1.1 Přepoččet pro administrativní jednotky

Prvním typem těchto areálů byly administrativní jednotky, jelikož jejich hranice jsou pro většinu obyvatel Česka známé a lze se tak pomocí nich v území státu lépe zorientovat. Navíc se jedná o vrstvy dat, které jsou dostupné zdarma (viz kapitola 2.4). Po konzultaci s vedoucím práce se zvolily vhodné administrativní jednotky pro agregaci dat. Zvolenými administrativními jednotkami byly obce a základní sídelní jednotky (ZSJ).

V dalším kroku bylo nutné zjistit, kolik adresních bodů se v každé obci/ZSJ vyskytuje. K tomu byl využit opět nástroj pro statistický přepoččet dat *Join Data* (viz. kapitola 2.3.2 a kapitola 4.4). Byla zvolena možnost připojení údajů z jiné vrstvy na základě prostorové lokace – *Join data from another layer based on spatial location*. V nastavení nástroje nebylo potřeba využít žádný ze statistických souhrnů, došlo tak tedy k vytvoření jediného nového atributu – *Count_*. Získaly se tak absolutní hodnoty s počtem adresních bodů v každém námi zkoumaném areálu.

Tyto hodnoty ovšem ještě nelze porovnávat a vizualizovat. Obecně platí, že čím větší je areál, tím více bodů se v něm bude pravděpodobně nacházet. Porovnání nestejně velkých areálů by tak bylo v tomto případě chybné. Aby byla vizualizace dat v souladu s kartografickými pravidly bylo nutné získat relativní hodnoty, tedy hodnoty normalizované (vztažené) k něčemu. U metody kartogramu to nejčastěji bývá rozloha území daného areálu (Voženilek & Kaňok, 2011).

Rozlohu areálů nebylo nutné vypočítávat, jelikož je to jeden ze základních atributů polygonové vrstvy administrativních jednotek. U administrativních jednotek se jednalo o atribut *SHAPE_Area*, ten uvádí rozlohu daného polygonu v metrech čtverečních. Vzhledem k rozměrům daných areálů bylo pro vizualizaci dat vhodné přepočítat rozlohu na kilometry čtverečné. Bylo vytvořen nový atribut *rozloha* ve formátu *float*. Sloupec *rozloha* se označil a vybral se nástroj *Field Calculator*, konkrétně vzorec: $\text{rozloha} = [\text{SHAPE_Area}] / 1\,000\,000$. Proběhl přepoččet rozlohy na kilometry čtverečné. Pro výpočet relativních hodnot byl vytvořen nový atribut *hustota_adresnich_bodu* a opět se využilo nástroje *Field Calculator*, tentokrát se zadal vzorec:

- $\text{hustota_adresnich_bodu} = [\text{Count_}] / [\text{rozloha}]$

Po tomto výpočtu obsahoval atribut *rozloha* relativní hodnoty výskytu adresních bodů, které byly vhodné pro vizualizaci metodou kartogramu.

Metoda kartogramu – nastavení znakového klíče

Pokud chceme použít metodu kartogramu, volíme v nastavení znakového klíče možnost *Quantities*. Zvolí se vizualizace pomocí gradujících barev – *Graduated colors* (čím vyšší je daná hodnota polygonu, tím je barva jeho výplně výraznější). V dalším kroku je potřeba zadat *Value* neboli atributové pole, ze kterého se berou hodnoty, v našem případě se jednalo o pole *hustota_adresnich_bodu*. Lze využít i tzv. normalizační pole, které v podstatě umožní použít absolutní hodnoty a přepočítat je na relativní až v symbolologii.

Dalším důležitým krokem je nastavení klasifikace, kdy se stanoví, kolik tříd neboli kolik různých barev bude vytvořeno, poté se vymezí jednotlivé intervaly hodnot těchto tříd. Při konzultaci s vedoucím bylo rozhodnuto použít 6 klasifikačních tříd. Jedna třída byla vytvořena pouze pro areály bez adresních bodů (*Count_=0*).

Intervaly pro zbylých 5 tříd se vzhledem k rozložení hodnot stanovily pomocí klasifikační metody *Quantils*, která určí hranice intervalů daných tříd tak, aby všechny

třídy obsahovaly stejný počet prvků. Tím se zajistí, že se nebudou vyskytovat třídy s velmi nízkým nebo naopak s velmi vysokým počtem prvků (výrazná převaha jedné barvy). Hraniční hodnoty mezi jednotlivými intervaly byly po konzultaci s vedoucím zaokrouhleny.

Posledním krokem v nastavení znakového klíče bylo přidělení konkrétních barev pro stanovené třídy (tvorba sekvenčních kvantitativních barevných stupnic). Pro tvorbu barevné stupnice byla využita webová aplikace *COLORBREWER 2.0*, která je schopna generovat kvantitativní i kvalitativní barevné stupnice tak, aby byly barvy jednotlivých tříd od sebe co nejlépe rozpoznatelné. Všechny vygenerované barvy lze převádět do barevných modelů HEX, RGB nebo CMYK. Tato aplikace je dostupná z webu: <http://colorbrewer2.org>

5.1.2 Přepočítání pro hexagonové sítě

Jak vyplývá z předchozí podkapitoly 5.1.1. při agregaci se nám vyplatí pracovat s takovými areály, které jsou si svou rozlohou podobné nebo v ideálním případě mají rozlohu stejnou. Proto se při výběru dalšího typu areálů pro agregaci přímo vybízely vygenerované hexagonové sítě. Jejich výhodou je, že každý areál (hexagon) v pravidelné síti má stejnou rozlohu a lze jej snadno porovnat s ostatními. Nevýhodou může být horší prostorová orientace sítě hexagonů, jelikož polohu jednotlivých hexagonů nedokážeme odhadnout tak přesně, jako polohu polygonů obcí a ZSJ.

Opět bylo potřeba vypočítat, kolik adresních bodů se v každém jednotlivém hexagonu nachází. K tomu byl znova použit nástroj *Join Data* a postup byl v podstatě stejný jako u administrativních jednotek. Přibylo tedy nové atributové pole *Count_* s počty adresních bodů v jednotlivých hexagonech. Protože jsou plochy všech hexagonů stejné, nebylo nutné data dále přepočítávat na relativní. Tyto hodnoty tedy mohli přímo vstupovat do kartogramů.

Metoda kartogramu – nastavení znakového klíče

V symbolologii se opět zvolila možnost *Quantities – Graduated Colors*. Hodnoty vstupovaly z atributového pole *Count_*. Při klasifikaci se postupovalo stejným způsobem jako u administrativních jednotek. Bylo vytvořeno celkem 6 tříd, kdy jedna reprezentovala pouze hexagony bez adresních bodů. Intervaly zbylých pěti se definovaly opět klasifikační metodou *Quantile* a hraniční hodnoty těchto intervalů byly zaokrouhleny. Vzhledem k stejnému počtu klasifikačních tříd byly použity i stejné barvy jako u klasifikačních tříd administrativních jednotek, tedy sekvenční barevné kvantitativní stupnice vygenerované aplikací *COLORBREWER 2.0*.

5.1.3 Výsledky prostorové vizualizace

Celkově byly vytvořeny tyto 4 prostorové vizualizace:

- Adresní body v pravidelné hexagonové síti s plochou hexagonu 3,5 km² (Příloha 22)
- Adresní body v pravidelné hexagonové síti s plochou hexagonu 6 km² (Příloha 4)
- Hustota adresních bodů podle obcí v České republice v roce 2016 (Příloha 5)
- Hustota adresních bodů podle ZSJ v České republice v roce 2016 (Příloha 23)

5.1.4 Vyhodnocení

Jak by se dalo logicky předpovídat, největší hustotu adresních bodů mají obce s velkým počtem obyvatel, tedy velká města v České republice. Největší hustota adresních bodů Česka je ve městech jako Praha, Brno, Ostrava, Plzeň, Zlín apod. Není ale pravidlem, že čím je katastrální území obce větší, tím je větší i hustota adresních bodů této obce. Na druhou stranu místa s nejmenší hustotou adresních bodů jsou reprezentována vojenskými újezdy, národními parky a dalšími chráněnými oblastmi. Lze pozorovat, že hustota adresních bodů ve venkovských oblastech roste ze západu na východ. Zatímco na západě České republiky mají menší obce v průměru do deseti adresních bodů/km², na východní straně Česka je průměrná hustota kolem 40 adresních bodů/km².

5.2 Měření geografické distribuce

Měření distribuce skupiny prvků umožňuje vypočítat charakteristiky distribuce jako je centrum, kompaktnost nebo orientace. Pro měření geografické (prostorové) distribuce adresních bodů Česka byla využita sada nástrojů *Measuring Geographical Distribution*, která je stručně popsána v podkapitole 2.1.3.

5.2.1 Hledání a porovnávání „centra“ adresních bodů v ČR

Hledání centra pro skupinu prvků může být užitečné pro sledování změn distribuce těchto prvků v čase. V některých případech může jít o užitečnou metodu pro vyhledání vhodné lokace pro něco, co má být centrálně lokalizované jako veřejná instituce. Na to, co je to centrum, může být více pohledů. Tři různé typy centra popisují tyto nástroje (viz. Tab 5.2.1):

- Geografický střed prvků – centrum koncentrace
- Geografický medián prvků
- Identifikace středového prvku

Tab. 5.2.1. Tři různé typy center pro skupinu prvků, jejich popis a využití.

Nástroj	Popis	Využití
Geografický střed	Průměr x-ové souřadnice a průměr y-ové souřadnice pro všechny prvky ve sledované oblasti	Sledování změn prostorové distribuce a její porovnávání pro prvky různých kategorií nebo pro prvky v různých časových obdobích
Geografický medián	Poloha x-ové a y-ové souřadnice, která má nejkratší vzdálenost ke všem prvkům ve sledované oblasti	Hledání nejvíce dostupné lokace pro prvky ve sledované oblasti
Středový prvek	Prvek, který má nejkratší celkovou vzdálenost ke všem ostatním prvkům ve sledované oblasti	Hledání nejvíce dostupného prvku pro ostatní prvky ve sledované oblasti

Porovnávání center adresních bodů Česka

Byla vypočítána geografická distribuce pro adresní body Česka. Konkrétně se zjišťovala poloha geografického středu a geografického mediánu adresních bodů Česka, která byla následně vzájemně porovnána spolu ještě s polohou středu území státu (těžiště).

Jak lze vidět na vizualizaci výsledků měření geografické distribuce (Příloha 6), všechny zjištěné body se nacházely ve středu České republiky a byly od sebe poměrně nedaleko. Malá vzdálenost a vzájemná poloha mezi geografickým středem adresních bodů a středem území Česka vypovídá, že adresní body jsou celkem rovnoměrně rozmístěny po celém území České republiky (ve všech směrech od středu státního území).

Poloha geografického mediánu reprezentuje místo, které má nejmenší celkovou vzdálenost ke všem adresním bodům Česka. Teoreticky by byla poloha tohoto místa vhodná pro administrativní centra a správní orgány naší země, protože by byla všem obyvatelům republiky stejně (vzdálenostně) dostupná. V praxi ovšem více záleží na dostupnosti místa než na vzdálenosti k němu. Například silniční síť kolem zjištěného geografického mediánu by nebyla vhodná pro dojíždění více lidí do centra.

5.2.2 Směrová distribuce adresních bodů pro obce v ČR

V této podkapitole je popsáno, jak bylo využito nástroje *Směrová distribuce* pro vyhodnocení prostorové struktury adresních bodů v obcích Česka. Hlavním cílem při použití této metody bylo zjistit, zdali a které obce v Česku mají své adresní body tedy budovy a jejich vchody pravidelně rozmístěné ve všech směrech od svého geografického centra, anebo naopak zdali převažuje nějaký výrazný směrový trend v rozložení jejich adresních bodů. Toto lze ohodnotit právě pomocí nástroje *Směrová distribuce*.

Nástroj Směrová distribuce

Někdy bývá tento nástroj označován jako Elipsa směrodatné *odchylky*. Tato metoda vytváří směrodatné elipsy odchylek, které slouží pro vyhodnocení a vizualizaci prostorových charakteristik sledovaných geografických prvků jako je:

- centrální tendence
- disperze (rozptyl)
- směrové trendy

Metoda vytvoří novou výstupní třídu prvků, která obsahuje polygony ve tvaru elipsy. Lze nastavit zdali se elipsa vytvoří pro danou skupinu prvků nebo pro každý „případ“ ve skupině prvků, který je definován ve volitelném parametru **Case Field**. Pokud využijeme *Case Field* parametr, jeho hodnoty jsou automaticky přidány do atributové tabulky. V atributové tabulce vytvořených elips se nachází 5 nebo 6 atributů:

- **CenterX/CenterY** – hodnota x-ové/y-ové souřadnice průměrného centra bodů (Mean Center)
- **XStdDist/ YStdDist** – velikost x-ové/y-ové osy elipsy
- **Rotation** – rotace delší osy elipsy měřena ve směru hodinových ručiček od poledne
- **CaseField** – typ „případu“ (volitelný atribut)

Volitelnou možností nastavení představuje **Weight Field (pole váh)**, které počítá s prostorovými vahami na základě hodnoty daného atributu, a **Ellipse Size (velikost elipsy)**:

- 1_STANDARD_DEVIATION – polygon elipsy obsahuje přibližně 68 % prvků
- 2_STANDARD_DEVIATIONS – polygon elipsy obsahuje přibližně 95 % prvků
- 3_STANDARD_DEVIATIONS – polygon elipsy obsahuje přibližně 99 % prvků

Nastavení metody

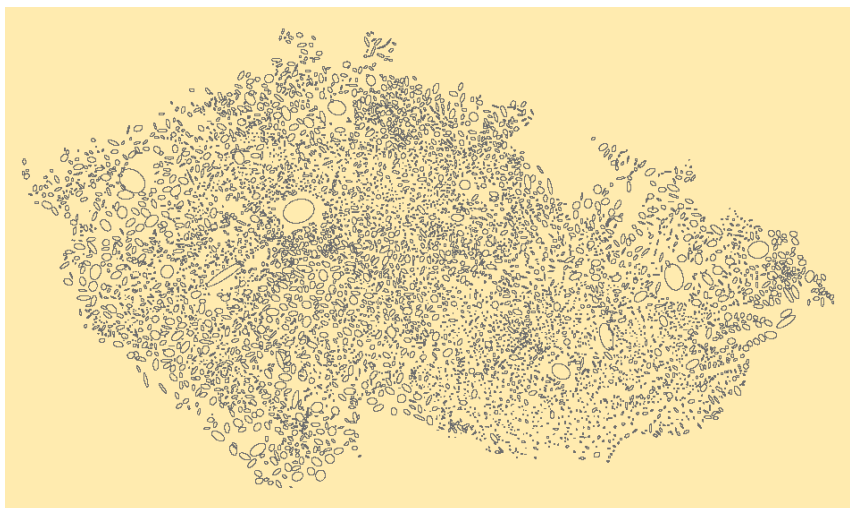
Vstupní vstvou pro tuto metodu byla bodová vstava adresních bodů. Jako velikost kružnice byla vybrána možnost 1_STANDARD_DEVIATION. Volitelný atribut *Weight Field* nebyl využit, jelikož cílem bylo pracovat čistě s polohou adresních bodů a nebylo tedy třeba využít prostorových vah na základě hodnot daného atributu. Důležité ale bylo využití volitelného atributu *Case Field*, kdy byl za „případ“ prvků zvolen atribut *ICOB* (identifikační číslo obce). Toto nastavení zajistilo, že byly vypočteny a vytvořeny eliptické kružnice pro každou skupinu všech adresních bodů s daným identifikačním číslem obce. Neboli pro každou obec se vytvořila právě jedna eliptická kružnice s určitými vlastnostmi, které se dále hodnotily a porovnávali.

Princip hodnocení prostorového rozložení adresních bodů v obcích Česka

Hodnoty atributů vytvořených nástrojem *Directional Distribution* nejsou přímo ty výsledky, které potřebujeme pro vyhodnocení prostorového rozložení adresních bodů v jednotlivých obcích. Nejdůležitější z těchto výsledků pro další výpočty jsou atributy **XStdDist** a **YStdDist**, ty reprezentují velikost x-ové a y-ové osy elipsy. Právě poměr velikostí těchto os vypovídá o tom, jak jsou adresní body kolem centra obce rozmístěny. Obecně platí, že čím blíže se hodnota tohoto poměru blíží hodnotě 1, tím je tvar elipsy „kulatější“. To značí, že jsou adresní body kolem centra rozmístěny pravidelně. Naopak čím vzdálenější je hodnota poměru velikostí os od hodnoty nula, tím více se elipsa liší od kružnice. To značí, že v rozmístění adresních bodů převládá výrazný směrový trend.

Výsledky metody

Výsledkem bylo 6251 elips, které jsou zobrazeny na Obr. 5.2.2. Dále bylo potřeba pracovat hlavně z hodnotami atributů vytvořených elips. Pro pohodlnější práci a výpočty byla data z atributové tabulky exportována do prostředí *Microsoft Excel*. Hodnoty atributové tabulky nelze přímo kopírovat do *Excelu*. První je nutné data z atributové tabulky exportovat do textového formátu *.TXT*, tento dokument pak obsahuje textový řetězec, v němž jsou jednotlivé hodnoty v sloupcích odděleny pomocí středníku. Toho se dá využít při importování dat z textu do prostředí programu *Excel*.



Obr. 5.2.2.1 Náhled na elipsy směrodatných odchylek pro obce v České republice v prostředí ArcMap.

Práce v Excelu

V prostředí softwaru Excel byly vytvořeny nové sloupce s názvy **pomer_XY** a **pomer_YX**. Hodnoty v novém sloupci pomer_XY se vypočítaly jako $XStdDist/YStdDist$. Hodnoty v novém sloupci pomer_yx se vypočítaly jako $YStdDist/XStdDist$.

Bylo vybráno 100 obcí, ve kterých byly velikosti x-ové a y-vé osy elipsy směrodatné odchylky nejvíce rozdílné a naopak 100 obcí, které měli velikost x-ové a y-ové osy elipsy směrodatné odchylky nejvíce podobnou (téměř kružnice). Tyto obce se vyexportovaly do samostatných vrstev a následně byly mapovány v prostoru prostřednictvím jejich katastrálního území pro porovnání jejich výskytu (Příloha 7).

Způsob vizualizace a hodnocení výsledků

Modrou barvou zde bylo zobrazeno katastrální území 100 obcí s nejvíce podobnou velikostí x-ové a y-ové osy u elipsy směrodatné odchylky (elipsa podobná kružnici) a červenou barvou pak katastrální území 100 obcí s nejvíce rozdílnou velikostí x-ové a y-ové osy u elipsy směrodatné odchylky (šišaté elipsy). „Modré“ obce se vyskytovaly po celém území republiky, přitom byly pozorovány shluky těchto obcí například na severu Královéhradeckého kraje. „Červené“ obce se vyskytovaly hlavně ve střední oblasti území Česka od západu na východ, ale jejich zastoupení bylo nízké na severozápadě a jihu území v blízkosti státních hranic. „Červené“ obce se na pár místech viditelně shlukovaly. Výrazné shluky těchto obcí lze pozorovat například na severu Moravskoslezského kraje a na západě Plzeňského kraje.

Hodnocení příčin výskytu „extrémních“ obcí

Posledním krokem hodnocení směrové distribuce adresních bodů Česka bylo hledání příčin vzniku „extrémních“ obcí. Jako „extrémní“ bylo označeno 20 obcí, ve kterých byly velikosti x-ové a y-vé osy elipsy směrodatné odchylky nejvíce rozdílné a dále 20 obcí, které měli velikost x-ové a y-ové osy elipsy směrodatné odchylky nejvíce podobnou (téměř kružnice). Tyto „extrémní“ obce jsou uvedeny v Tab. 5.2.2.1 a v Tab. 5.2.2.2.

Tab. 5.2.2.1 Seznam 20 obcí s nejmenším rozdílem velikostí x-ové a y-ové osy u elipsy směrodatné odchylky.

Název obce	1 - poměr	Podíl_XY	Podíl_YX	Počet obyvatel	Rozloha území [km ²]
Strážov	0,0072	0,9928	1,0072	1351	35,70
Hvězdlice	0,0103	0,9897	1,0104	592	11,16
Lukavice	0,0111	1,0112	0,9889	915	9,44
Zbečno	0,0210	0,9790	1,0214	470	15,15
Vysoké	0,0274	1,0281	0,9726	157	6,52
Třebíč	0,0276	1,0284	0,9724	37575	57,60
Přídolí	0,0278	1,0286	0,9722	679	40,01
Borová	0,0280	1,0288	0,9720	211	3,07
Vísky	0,0292	1,0300	0,9708	263	3,78
Pastviny	0,0298	0,9702	1,0307	348	8,37
Budíkov	0,0310	0,9690	1,0320	318	6,64
Děčany	0,0332	0,9668	1,0343	359	12,37
Chroustovice	0,0341	1,0353	0,9659	1260	21,60
Blansko	0,0357	1,0370	0,9643	20769	44,97
Poděbrady	0,0380	1,0395	0,9620	13894	33,69
Střítež	0,0390	0,9610	1,0405	405	7,46
Líšná	0,0398	0,9602	1,0414	55	5,74
Klimkovice	0,0406	1,0423	0,9594	4238	14,64
Trhová Kamenice	0,0413	1,0431	0,9587	906	20,39
Jílové	0,0432	0,9568	1,0451	5172	36,56
Průměrný počet obyvatel:		4497			
Průměrná výměra území [km²]:		19,74			

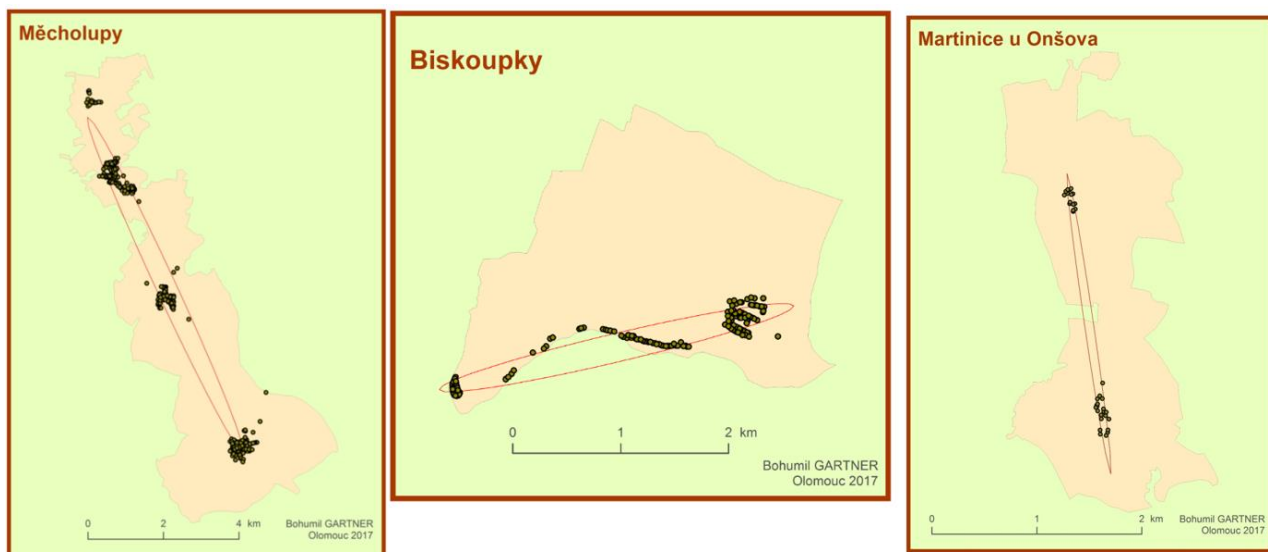
Tab. 5.2.2.2 Seznam 20 obcí s největším rozdílem velikostí x-ové a y-ové osy u elipsy směrodatné odchylky.

Název obce	Poměr	Podíl_XY	Podíl_YX	Počet obyvatel	Rozloha území [km ²]:
Martinice u Onšova	33,90	33,90	0,03	55	4,69
Měcholupy	18,46	18,46	0,05	1025	29,12
Biskoupky	12,88	0,08	12,88	189	5,81
Minice	12,47	0,08	12,47	38	3,50
Jedlá	12,45	12,45	0,08	80	6,26
Vrchovnice	12,22	0,08	12,22	65	1,52
Synkov-Slemeno	11,44	0,09	11,44	385	7,25
Karolín	11,22	11,22	0,09	241	1,36
Chobot	11,13	0,09	11,13	59	2,33
Radošovice	10,95	0,09	10,95	187	7,77
Hvozď	10,78	10,78	0,09	168	3,68
Prasek	10,64	10,64	0,09	615	6,26
Ostrov	10,58	10,58	0,09	654	18,48
Račín	10,42	0,10	10,42	118	7,96

Honezovice	10,32	0,10	10,32	231	18,02
Šetějovice	10,28	10,28	0,10	69	6,11
Křižovatka	10,21	10,21	0,10	276	14,13
Močerady	10,12	10,12	0,10	59	6,67
Hradec nad Svitavou	9,86	0,10	9,86	1720	24,71
Šestajovice	9,85	9,85	0,10	190	7,52
Průměrný počet obyvatel:		321			
Průměrná výměra území [km²]:		9,16			

Jak lze vyčíst z uvedených tabulek, obce s největším rozdílem velikostí x-ové a y-ové osy u elipsy směrodatné odchylky jsou v průměru menší a mají méně obyvatel než obce s největším rozdílem velikostí x-ové a y-ové osy u elipsy směrodatné odchylky. Obce s největším rozdílem velikostí x-ové a y-ové osy u elipsy směrodatné odchylky leží často v horských oblastech (např. Jeseníky) a výskyt adresních bodů je tak výrazně ovlivněn tvarem reliéfu. Naopak obce s nejmenším rozdílem velikostí x-ové a y-ové osy u elipsy směrodatné odchylky leží často v oblasti nížin (např. Polabská nížina).

V „červených“ obcích, které jsou uvedeny v Tab. 5.2.2.2 byla vizuálně zkoumána a slovně hodnocena distribuce jejich adresních bodů a její souvislost s tvarem příslušné elipsy směrodatné odchylky.



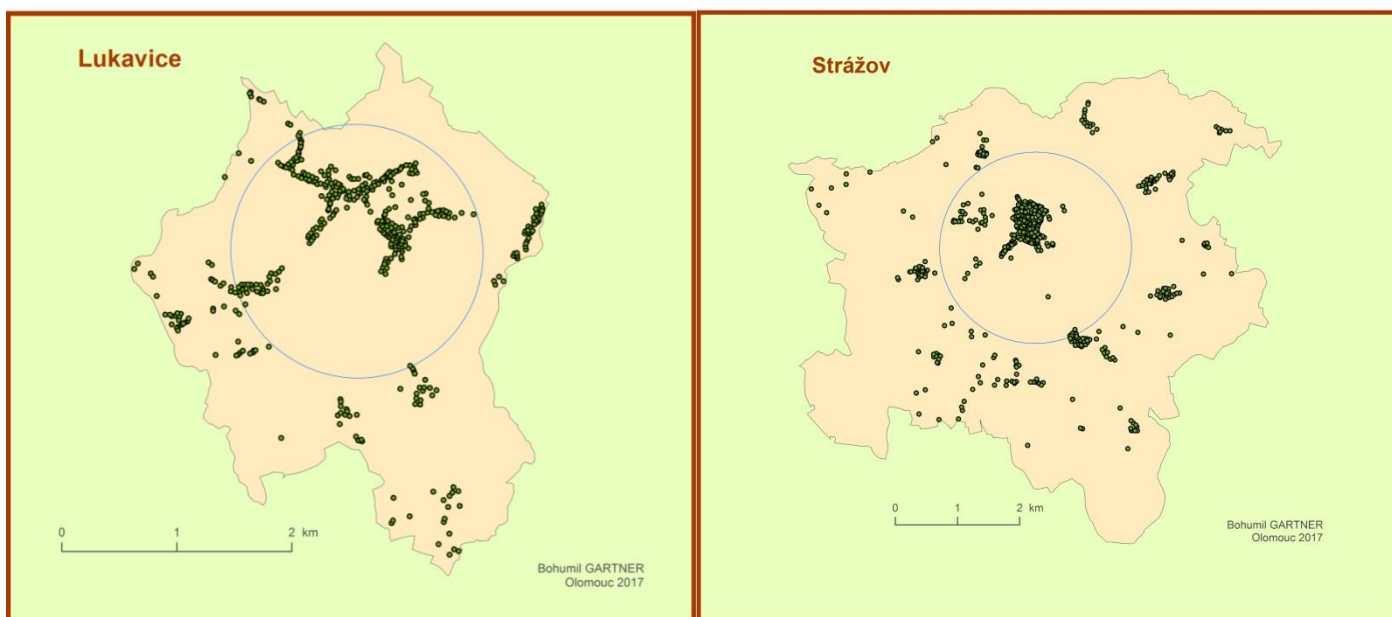
Obr. 5.2.2.2 Náhled prostorové distribuce adresních bodů u vybraných „červených“ obcí.

Jak lze pozorovat na Obr. 5.2.2.2, extrémní případy distribuce adresních bodů, kdy jsou velikosti x-ové a y-ové osy elipsy směrodatné odchylky nejvíce rozdílné jsou často způsobeny rozdělením obce na místní části, které jsou od sebe izolované a které by samy o sobě měly elipsy směrodatných odchylek poměrně „kulaté“, ale jelikož tyto místní části sdílejí jedno katastrální území, je výsledná elipsa poměrně „šišatá“. Tímto příkladem jsou i obce Měcholupy a Martinice u Onšova (Obr. 5.2.2.2). Dalším důvodem vzniku „šišatých“ kružnic je terén (reliéf) území obce. Horské obce často leží v údolí a rozrůstají se převážně podél osy údolí. Další přírodní bariérou může být řeka, podél které se staví domy, jako je tomu u obce Biskoupky (Obr. 5.2.2.3).

Na druhou stranu máme pro porovnání na Obr. 5.2.2.4 ukázkou distribuce adresních bodů obcí, u kterých si byly velikosti x-ové a y-ové osy elipsy směrodatné odchylky nejvíce podobné (téměř kružnice).



Obr. 5.2.2. 3 Náhled na oblast kolem obce Biskoupky, která se rozrůstá podél řeky Jihlavy.



Obr. 5.2.2. 4 Náhled prostorové distribuce adresních bodů u vybraných „modrých“ obcí (zdroj: <https://mapy.cz/s/2AoHH>).

5.2.3 Vyhodnocení výsledků z měření geografické distribuce:

Vizualizace výsledků měření geografické distribuce adresních bodů Česka:

- Geografická distribuce adresních bodů v České republice v roce 2016 (Příloha 6)
- Elipsy směrodatných odchylek pro obce v České republice v roce 2016 (Příloha 7)

Vyhodnocení výsledků

Adresní body Česka se vyskytují v rozsahu celého území státu a jejich výskyt je poměrně rovnoměrný ve všech směrech od jejich geografického středu.

Obce s největším rozdílem velikostí x-ové a y-ové osy u elipsy směrodatné odchylky se v několika oblastech Česka shlukují. Výrazná převaha distribuce adresních bodů těchto obcí v jednom směru může být způsobena rozdělením obce na menší místní části, které leží v jedné ose a které jsou od sebe izolovány, dalším důvod tohoto typu distribuce představují přírodní překážky jako řeky nebo povrch reliéfu.

Obce s nejmenším rozdílem velikostí x-ové a y-ové osy u elipsy směrodatné odchylky se většinou nachází v rovinné oblasti (nížiny).

5.3 Analýzy prostorového uspořádání

Tyto analýzy identifikují prostorový vzor (pattern). Identifikace prostorového (geografického) vzoru tvořeného sadou zkoumaných prvků umožňuje lépe pochopit distribuci těchto prvků, porovnávat distribuce různých skupin prvků a sledovat změny distribuce. Jakákoli distribuce prvků nebo hodnot atributů v libovolně definované oblasti bude tvořit prostorový vzor. Geografické vzory mohou vykazovat významné shlukování nebo naopak rozptýlení prvků v prostoru. Pokud nejsou prvky ani významně shlukované ani významně rozptýlené, mluvíme o náhodném prostorovém uspořádání.

V programu ArcMap 10.4 z ArcGIS Desktop bylo využito sady nástrojů *Analýza prostorového uspořádání*. Ta je stručně popsána v podkapitole 2.1.1. Po konzultaci s vedoucím práce bylo rozhodnuto, že pro analýzu prostorového uspořádání adresních bodů budou dostačující metoda *Průměrného nejbližšího souseda* a *Prostorová autokorelace*.

5.3.1 Průměrný nejbližší soused

Tato metoda testuje náhodnost rozmístění bodů v prostoru. Výsledkem této metody je index nejbližšího souseda vypočítaný na základě průměrné vzdálenosti jednotlivých prvků od jejich nejbližšího sousedního prvku. Všechny počítané vzdálenosti se vztahují k těžištím měřených prvků.

Pokud je vypočítaná průměrná vzdálenost menší než hypotetická hodnota náhodné distribuce prvků, vykazují analyzované prvky shlukování. Naopak pokud je průměrná hodnota větší než hypotetická hodnota, prvky lze považovat za rozptýlené. Nulová hypotéza u této metody tvrdí, že vrstva nebo hodnoty ve vrstvě vykazují statisticky náhodný vzorek.

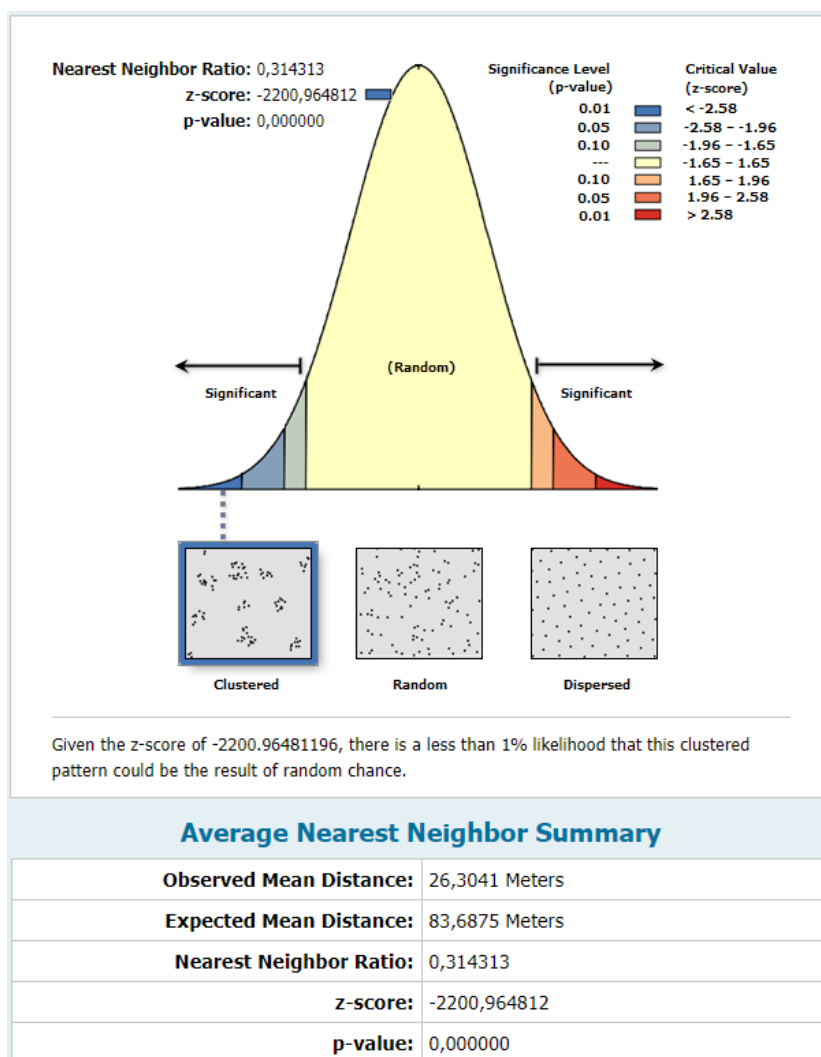
Index průměrného nejbližšího souseda se vypočítá jako podíl průměru pozorovaných vzdáleností s předpokládanou průměrnou vzdáleností založenou na hypotetickém náhodném rozdělení stejných hodnot prvků pokrývající stejnou plochu. Pokud je poměr je menší než 1, vzorek vykazuje shlukování, v opačném případě vykazuje rozptýlení prvků.

Metoda vypočítá a navrátí 5 hodnot:

- *Observed Mean Distance* – změřená střední vzdálenost
- *Expected Mean Distance* – předpokládaná střední vzdálenost
- *Nearest Neighbor Index* – poměr změřené a předpokládané vzdálenosti
- *z-score* – test statistické významnosti
- *p-value* – pravděpodobnost, že pozorované prostorové uspořádání vzniklo nějakým náhodným procesem

Do metody *Průměrného nejbližšího souseda* vstupovala nejdříve bodová vrstva adresních bodů s vybranými atributy. Hodnoty atributů vstupních dat nebyly pro tuto metodu podstatné, jelikož ta využívá pro výpočet pouze polohu zkoumaných prvků. Při nastavení metody se zvolila *Euklidovská vzdálenost*, tedy přímá vzdálenost mezi dvěma body. Nástroj nabízí ještě počítat s *Manhattanskou vzdáleností*. Jde o vzdálenost mezi dvěma body měřenou podél os v pravém úhlu.

Dále se zvolila možnost generování grafického výstupu. Výsledná zpráva byla generována ve formátu .html a zobrazena ve webovém prohlížeči (viz Obr. 5.3.1.1). Poslední možností nastavení před spuštěním metody je zadání plochy území, ve kterém se studované prvky nacházejí. Tato hodnota značně ovlivní výpočet, proto byla snaha použít co nejpřesnější údaj. V našem případě šlo plochu o území Česka. Tu lze snadno zjistit v atributové tabulce u vrstvy *Stat_polygon*, která je součástí digitální vektorové databáze ArcČR® 500. Zadaná plocha území Česka byla 78866753310,57 m².



Obr. 5.3.1.1 Grafický výstup z metody Průměrný nejbližší soused pro bodovou vrstvu adresních bodů Česka

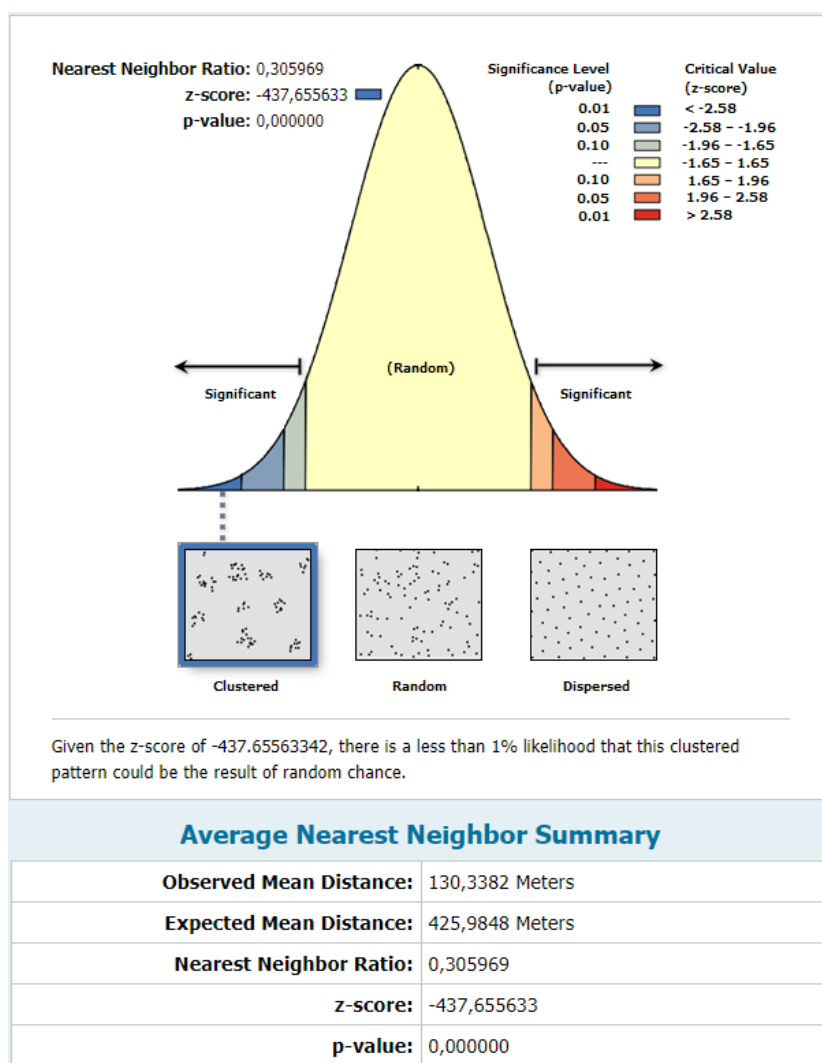
Z grafického výstupu lze vyčíst, že poměr (*Nearest Neighbor Ratio*) je 0,314. Poměr je tedy nižší než jedna, což značí **shlukování** dat. Standartní odchylka *z-score* je silně záporná (-2200). Hodnota *p-value* je 0. Na základě vyhodnocení *z-score* a *p-value*, které je popsáno v podkapitole 3.1.4. **Hladina spolehlivosti** je v tomhle případě **99 %**.

Ta značí, že není možné zamítnout nulovou hypotézu, pokud není pravděpodobnost, že mají prvky náhodné rozmístění, opravdu malá (méně než 1 % pravděpodobnosti).

Tento výsledek v podstatě říká, že poloha adresních bodů není výsledkem náhodného rozmístění a že se tyto body v prostoru významně shlukují. Tento výsledek lze předpokládat, jelikož je obecně známo, že se adresní body shlukují v sídlech, jako jsou vesnice nebo města. Lidé nestaví své domy náhodně, ale plánovaně a blízko sebe, což způsobuje i prostorové shlukování adresních bodů. Výsledek říká, že lze s jistotou mapovat shluky adresních bodů v prostoru.

Zajímavějším využitím této metody pro účely prostorové statistiky adresních bodů je její aplikace na vrstvu adresních bodů bez nulových hodnot u vybraných atributů, která vznikla při prostorové analýze nulových hodnot vybraných atributů adresních bodů. Zjištění, že se adresní body bez nulových hodnot v prostoru shlukují, by mohlo dále vést k mapování těchto shluků „dokonalých“ bodů v prostoru a nalezení vhodné oblasti pro podrobnější prostorové statistiky.

Metoda *Průměrný nejbližší soused* byla tedy vypočítána i pro bodovou vrstvu adresních bodů bez nulových atributů. Nastavení bylo stejné jako u předchozího použití metody (*Euklidovská vzdálenost*, generování grafického výstupu a zadání konkrétní plochy území Česka).



Obr. 5.3.1.2 Grafický výstup z metody Průměrný nejbližší soused pro bodovou vrstvu adresních bodů Česka bez nulových hodnot u vybraných atributů.

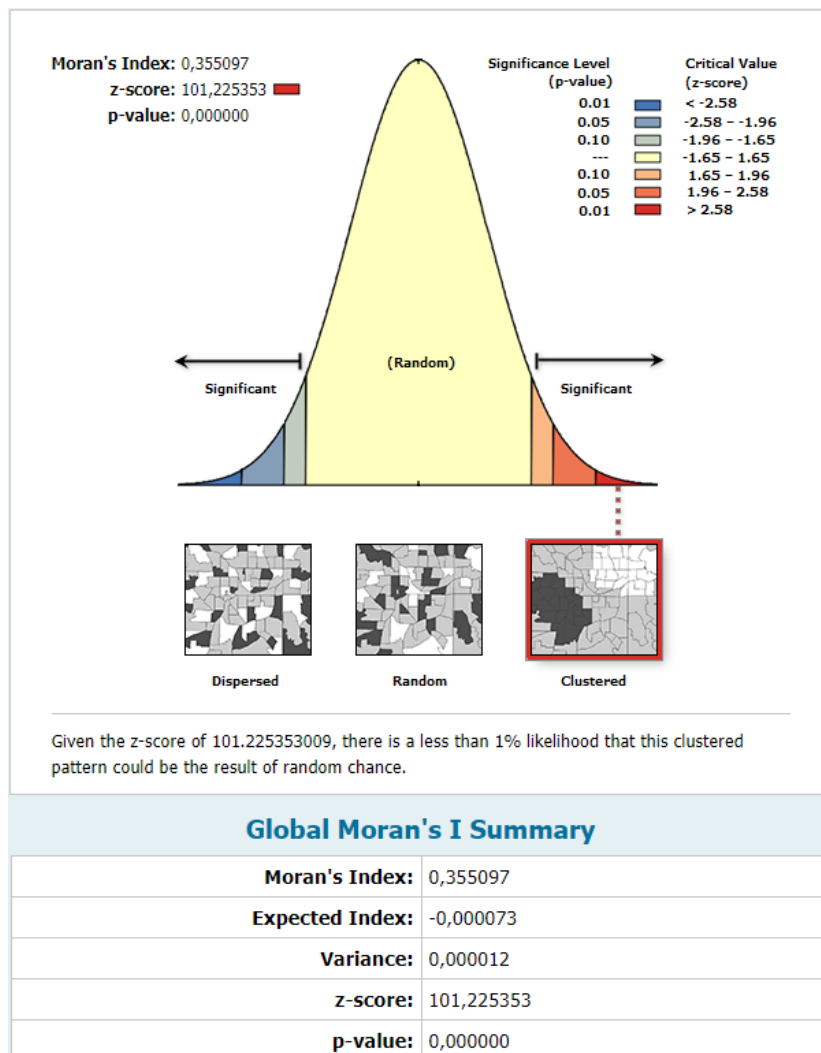
Výsledky jsou zobrazeny na Obr. 5.3.1.2. Poměr (*Nearest Neighbor Ratio*) je 0,305. Standardní odchylka *z-score* je opět silně záporná (-437). Hodnota *p-value* je 0. **Hladina spolehlivosti** podle popisu vyhodnocení v podkapitole 3.1.4 je i v tomto případě **99 %**. Poměr je nižší než jedna, což vykazuje **shlukování** adresních bodů bez nulových hodnot u vybraných atributů.

Výsledek je velmi podobný jako u vrstvy všech adresních bodů. Je možné tedy s jistotou mapovat shluky adresních bodů i shluky adresních bodů bez nulových hodnot u vybraných atributů na území Česka.

5.3.2 Prostorová autokorelace

Definice a stručný popis prostorové autokorelace je uveden v podkapitole 3.1.2. Jde o další metodu, která byla využita při analýze prostorového uspořádání adresních bodů v ČR. Metoda byla využita ke zjištění, jestli se ve zkoumané oblasti nachází statisticky významné shluky adresních bodů a statisticky významné shluky bezchybných adresních bodů.

Výsledkem této metody je *Moranův index*, který je vypočítán z hodnoty *p-value* a *z-score*. Pokud mají zkoumané prvky tendenci ke shlukování, je hodnota *z-score* kladná. Mají-li prvky neuspořádaný charakter hodnota *z-score* je blízká nule. Záporná hodnota *z-score* pak značí výraznou odlišnost sousedních hodnot atributu zkoumaných prvků. *P-value* vyjadřuje statistickou významnost shluků.



Obr. 5.3.2.1 Grafický výstup z metody Prostorová autokorelace pro bodovou vrstvy adresních bodů Česka.

Do nástroje vstupovala data agregovaná do pravidelné hexagonové sítě s plochou hexagonu 6km². Prostorová autokorelace používá k výpočtům hodnoty určitého atributu zkoumaných prvků. Vstupujícím atributem pro výpočet byl atribut s počtem adresních bodů v jednotlivých hexagonech. Další nastavení u této metody byl výběr konceptu prostorového vztahu, kdy byla zvolena obecně doporučená metoda konstantní vzdálenosti (Anselin, 1995). Prahová vzdálenost byla nastavena na 5 000 m. Měření probíhalo na základě *Euklidovské vzdálenosti* (přímá vzdálenost).

Tato analýza zjistí, zda jsou hodnoty vybraného atributu zkoumaných prvků v prostoru uspořádány jistým způsobem nebo zda jsou sousedící hodnoty odlišné od očekávaného rozdělení. Odlišnost těchto uspořádání je zobrazena na Obr. 5.3.2.2, kdy nalevo vidíme rozptýlený charakter hodnot, který postupně přechází v uspořádaný charakter (napravo). Nástroj nevytváří novou vrstvu, ale pouze textovou zprávu. I tento nástroj nabízí generování grafického výstupu v HTML formátu.



Obr. 5.3.2.2 Ukázka rozptýleného (nalevo) a uspořádaného (napravo) charakteru hodnot (polygonů).

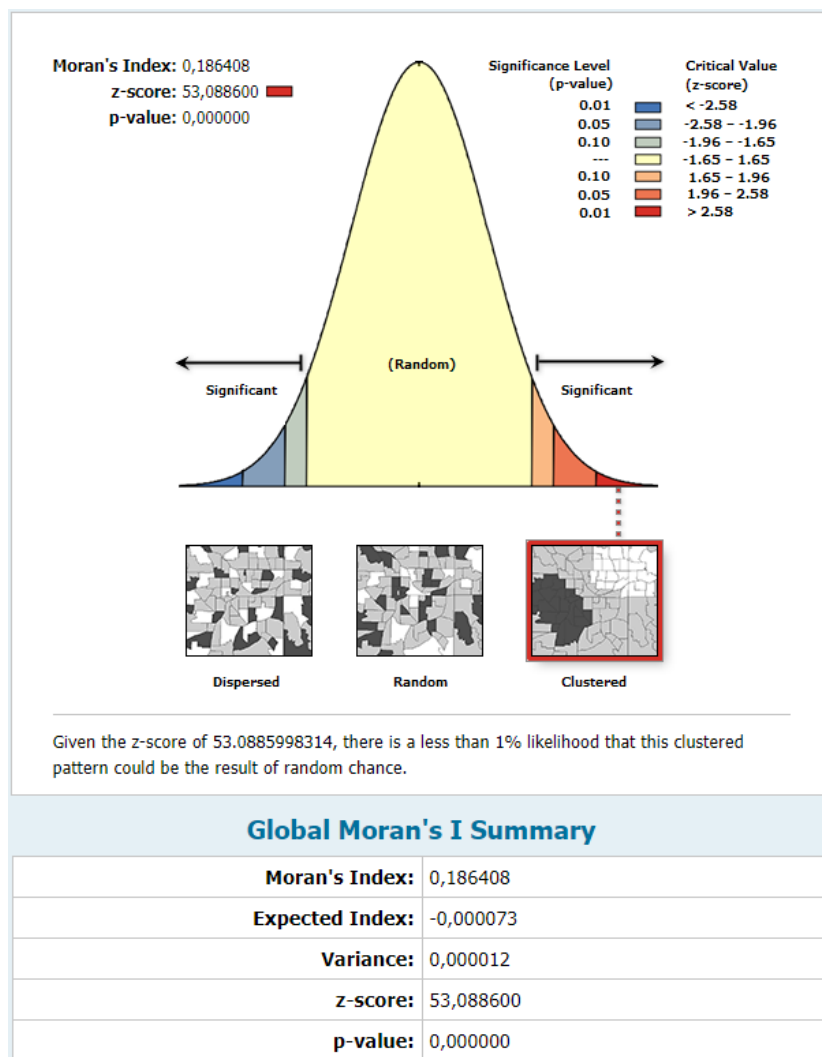
Grafický výstup s výsledky této metody je zobrazen na Obr. 5.3.2.1. Výsledná hodnota *z-score* byla 101,255. Silně kladná hodnota *z-score* říká, že se hodnoty (počty adresních bodů) v prostoru výrazně **shlukují (hladina spolehlivosti 99 %)**.

Na území Česka se tedy budou vyskytovat shluky hexagonů s velkým počtem adresních bodů, ty budou reprezentovat města a hustě osídlené oblasti. Naopak venkovské oblasti či chráněné krajiny budou reprezentovány shluky hexagonů s nízkými počty adresních bodů.

Tato metoda byla využita pro agregovaná data ještě jednou. Tentokrát byl vstupujícím atributem určeným pro analýzu prostorové autokorelace průměrný počet nulových hodnot u vybraných atributů adresních bodů. Nastavení bylo stejné jako v předchozím kroku (koncept konstantní vzdálenosti, prahová vzdálenost 5 000 m a měření *Euklidovskou vzdáleností*).

Výsledky prostorové autokorelace pro hodnoty průměrného počtu nulových hodnot u vybraných atributů adresních bodů jsou zobrazeny na Obr. 5.3.2.3. Výsledná hodnota *z-score* byla 53,255. Silně kladná hodnota *z-score* říká, že se hodnoty (průměrné počty nulových hodnot u vybraných atributů adresních bodů) v prostoru výrazně **shlukují (hladina spolehlivosti 99 %)**.

Na území Česka se tedy budou vyskytovat shluky hexagonů s nadprůměrným a podprůměrným počtem nulových hodnot u vybraných atributů adresních bodů. Mapování oblastí těchto shluků by mohlo přinést zajímavé výsledky.



Obr. 5.3.2.3 Grafický výstup z metody Prostorové autokorelace pro hodnotu s průměrným počtem nulových hodnot vybraných atributů adresních bodů Česka.

5.4 Mapování shluků

Pro mapování shluků byla zvolena metoda *Analýza shluků vysokých hodnot*. Obecný princip fungování této metody je uveden v podkapitole 3.1.2. V této podkapitole bude uveden postup a nastavení výpočtů této analýzy.

Vstupními daty byly adresní body agregované do pravidelné hexagonové sítě s plochou hexagonu 6 km², které obsahovaly atribut s počtem adresních bodů v jednotlivých hexagonech.

5.4.1 Analýza shluků vysokých hodnot (Hotspot analýza)

Tato metoda vyhledává shluky polygonů vysokých či nízkých hodnot. Aby byl polygon statisticky významný, musí být obklopen polygony vysokých či nízkých hodnot. *Hotspots* u této metody představují polygony s vysokou hodnotou obklopené polygony s vysokými hodnotami a *coldspots* jsou polygony s nízkou hodnotou obklopené polygony s nízkými hodnotami.

Důležité nastavení u této metody byl výběr konceptu prostorového vztahu, který významně ovlivňuje modelování analýzy. Uvedená metoda nabízí výběr základních šesti typů konceptů prostorových vztahů:

- metodu inverzních vzdáleností
- metodu čtvercové inverzní vzdálenosti
- metodu konstantní vzdálenosti
- metodu zóny netečnosti
- metodu styčných polygonů
- metodu uživatelsky definovaných vah

Obecně je doporučován koncept prostorových vztahů metodou konstantní vzdálenosti (Anselin, 1995). Proto bylo rozhodnuto použít při mapování shluků metodu konstantních vzdáleností. ESRI (2017) pro tuto metodu udává, že každý prvek je analyzován v kontextu sousedních prvků. Sousedním prvkům v rámci určené prahové vzdálenosti je přiřazena hodnota jedna a tyto prvky ovlivní výpočet analýzy. Naopak sousedním prvkům za prahovou vzdáleností je přiřazena hodnota nula a nemají žádný vliv na výpočty cílové funkce. Prahová vzdálenost může být zadána uživatelem anebo se vypočítá automaticky na základě velikosti vstupní třídy prvků a jejich počtu. Měření může být řešeno dvěma způsoby ESRI (2017):

- Euklidovská vzdálenost – přímá vzdálenost mezi dvěma body (letecká vzdálenost)
- Manhattan vzdálenost – vzdálenost mezi dvěma body měřená podél os v pravém úhlu, která je vypočtená součtem absolutního rozdílu mezi souřadnicemi x a y.

Pro účel těchto shlukových analýz byla použita Euklidovská vzdálenost. Analýzy mapování shluků nabízí další volitelné parametry jako *Self Potential Field* – pole reprezentující vlastní potenciál prvku a *Weights Matrix File* – soubor obsahující váhy, které definují prostorové a potenciálně dočasné vztahy mezi prvky. Tyto volitelné parametry nebyly při mapování shluků využity. Prahová vzdálenost byla spočítána na základě průměrné vzdálenosti hexagonu k jeho 30 nejbližším sousedům. Tato průměrná vzdálenost (7224 m) byla zaokrouhlena na 7500 m. Byla použita volitelná korekce FDR (*False Discovery Rate*), která zajistí, že statistická významnost bude založena na míře nesprávného zjištění. Další parametry byly volitelné, a nebylo třeba je nastavovat.

Výsledky a vyhodnocení:

- Shluková analýza *Getis Ord Gi** pro adresní body Česka v roce 2016 (Příloha 9).
- Shluková analýza *Getis Ord Gi** průměrného počtu nulových hodnot vybraných atributů adresních bodů Česka v roce 2016 (Příloha 10).

Ve výsledcích shlukových analýz *Getis Ord Gi** lze pozorovat červené polygony reprezentující statisticky významné shluky vysokých hodnot (*hotspots*) a modré polygony představující statisticky významné shluky nízkých hodnot. Čím je shluk statisticky významnější, tím je barva jeho výplně výraznější (tmavší, sytější).

Ve shlukové analýze *Getis Ord Gi** pro adresní body Česka v roce 2016 (Příloha 10) se nacházejí statisticky významné shluky vysokých hodnot v okolí všech krajských měst s výjimkou Karlových Varů. Největší shluk s vysokým počtem adresních bodů se nachází v Praze a v jejím okolí. Statisticky významné shluky s nízkým počtem adresních bodů se nacházejí v oblastech blízko státních hranic, většinou jde o chráněné přírodní oblasti. Největší shluk s nízkou hustotou adresních bodů se nachází na Šumavě.

Ve shlukové analýze *Getis Ord Gi** průměrného počtu nulových hodnot vybraných atributů adresních bodů Česka v roce 2016 (Příloha 11) se nachází velké množství statisticky významných shluků s vysokým počtem nulových hodnot adresních bodů. Největší z těchto shluků se nachází jižně od Prahy v oblasti Brd a Středočeské pahorkatiny, další z těchto velkých shluků jsou v oblasti Krušných Hor, severozápadně od Brna a v oblasti Moravskoslezských Beskyd. Lze tedy tvrdit, že statisticky významné shluky s vysokým počtem nulových hodnot atributů se nachází v horských přírodních oblastech a pahorkatinách. V menším počtu se zde nacházejí i statisticky významné shluky s nízkým počtem nulových hodnot vybraných atributů adresních bodů. Nejvýznamnější z nich leží v oblasti vojenských újezdů a mezi sídly Znojmo a Mikulov. Největší z nich pak leží v oblasti České tabule.

6 PROSTOROVÁ STATISTIKA ADRESNÍCH BODŮ HLAVNÍHO MĚSTA PRAHY

Jedním z dílčích cílů této bakalářské práce bylo provést podrobnější prostorové analýzy pro libovolnou oblast České republiky a zároveň využít hodnot atributů adresních bodů dané oblasti.

Jelikož se v průběhu zpracování dat přišlo na problém s nadměrným výskytem nulových hodnot u vybraných atributů adresních bodů s potenciálem pro prostorové analýzy, rozhodl se autor práce po konzultaci s vedoucím využít výsledky prostorové analýzy nulových hodnot a vybrat pomocí nich jeden z krajů, ve kterém bude průměrný počet nulových hodnot atributů nejnižší. Nejvhodnějším krajem pro podrobnější prostorové analýzy byl kraj Hlavní město Praha (viz podkapitola 4.6).

V tomto vybraném regionu bylo nutné vybrat atributy, které budou vstupovat do shlukových analýz. Po konzultaci s vedoucím se autor práce rozhodl využít tři atributů, konkrétně byl vybrán atribut s počtem evidovaných obyvatel (BUDOBYEV), atribut s počtem bytů (SUMBYT) a atribut s počtem podlaží budovy (POCPODBUD).

6.1 Adresní body Prahy

V kraji Hlavní město Praha se nachází celkem 116 176 adresních bodů. Na území Prahy byly adresní body agregovány a vizualizovány v pravidelné hexagonové síti s plochou hexagonu $0,5 \text{ km}^2$ a také v ZSJ.

Pro statistický přepočítání dat byl opět využit nástroj *Join Data* a postup byl obdobný jako při přepočtu adresních bodů pro celé území (viz kapitola 5.1). Pro vizualizaci agregovaných dat byla opět využita metoda kartogramu. Nastavení znakového klíče metody kartogramu bylo stejné jako u předchozích agregací. Bylo vytvořeno 6 klasifikačních tříd, jejichž hranice intervalů byly vymezené metodou kvantilů. Tyto hraniční hodnoty byly zaokrouhleny. Při nastavování znakového klíče byla využita stejná barevná stupnice jako pro území Česka.

6.1.1 Mapové výstupy a vyhodnocení

- Adresní body Prahy v pravidelné hexagonové síti s plochou hexagonu $0,5 \text{ km}^2$ v roce 2016 (Příloha 10).
- Hustota adresních bodů Prahy podle ZSJ v roce 2016 (Příloha 11).

Největší hustota adresních bodů je v centru města, kde se nachází 100 a více adresních bodů na ploše $0,5 \text{ km}^2$. Tento výsledek lze předpokládat, jelikož centrum města bývá vždy nejžádanější lokalitou jak pro bydlení, tak pro veřejnou správu i různé zázemí firem. V centru měst se staví budovy velice nahusto a je zde využitý každý volný prostor určený k zástavbě, proto zde bude i velmi hustá síť adresních bodů.

Velká hustota adresních bodů je i okolí centra podél komunikací, které jsou s ním spojeny. Většinou se zde nachází řadové rodinné domy nebo sídliště, které představují levnější variantu pro bydlení, ale jsou ještě relativně dostupné centru. Hustota adresních bodů v takových oblastech, kde se blízko sebe nachází řadové rodinné domy, panelové domy nebo jiné budovy s byty, které mají většinou více vchodů pro jednu

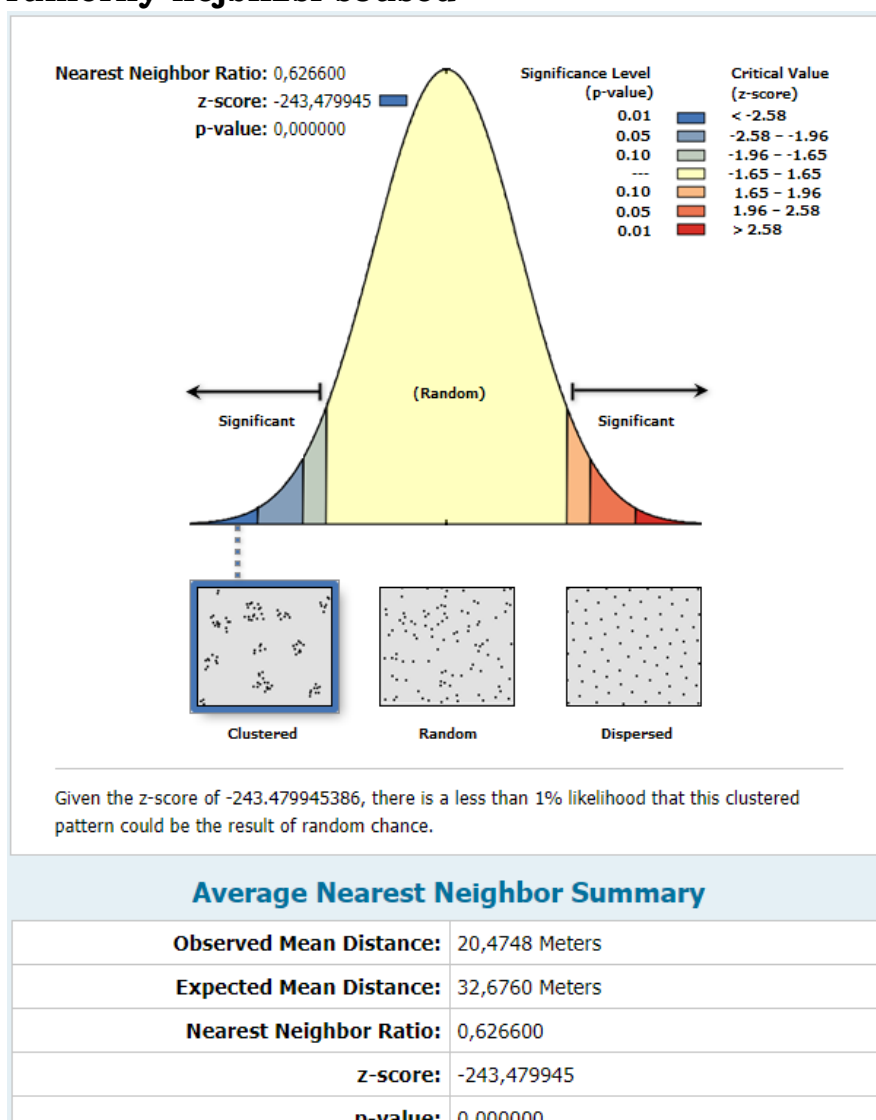
budovu, kdy každý vchod je reprezentován jedním adresním bodem, je vysoká. V případě Prahy se jedná například o Strašnice, Střešovice a Smíchov. Hustotu adresních bodů v těchto oblastech zvyšují také samostatné řadové garáže, které jsou velmi malé a vedle sebe v řadách, kdy každá garáž je reprezentována samostatným adresním bodem.

Nejmenší hustota adresních bodů Prahy je v krajních oblastech města blízko jeho východních a západních hranic. V takovýchto oblastech se nachází 50 a méně adresních bodů na ploše 0,5 km² a někdy se zde dokonce nevyskytují žádné adresní body. V lokalitách na okraji města, které nejsou až tak žádané a parcely v nich jsou tak často větší a levnější se pravděpodobně budou nacházet rodinné domy z většími parcelami nebo velké průmyslové areály, často reprezentované pouze jedním adresním bodem, proto je výsledná hustota adresních bodů nízká.

6.2 Analýzy prostorového uspořádání

Prostorový vzor byl vyhodnocován i pro adresní body v Praze. Do analýz prostorového uspořádání adresních bodů v Praze vstupovaly nově navíc i vybrané atributy BUDOBYEV, SUMBYT, POCPODBUD.

6.2.1 Průměrný nejbližší sused



Obr. 6.2.1 Grafický výstup z metody Průměrného nejbližšího suseda pro adresní body v Praze.

Popis této metody je uveden v podkapitole 5.3.1. Do metody *Průměrného nejbližšího souseda* vstupovala bodová vrstva adresních bodů Prahy. Zadala se metoda měření pomocí *Euklidovské vzdálenosti* a plocha území Prahy 496174681,239515 m². Zvolila se možnost generování grafického výstupu ve formátu HTML. Jelikož metoda nevyužívá hodnot atributů zkoumaných prvků nebylo potřeba dále nic nastavovat.

Výsledky metody je zobrazuje Obr.6.2.1. Z grafického výstupu lze vyčíst, že poměr (*Nearest Neighbor Ratio*) je 0,314. Poměr je tedy nižší než jedna, což značí **shlukování** dat. Standartní odchylka *z-score* je silně záporná (-2200). Hodnota *p-value* je 0. Na základě vyhodnocení *z-score* a *p-value*, které je popsáno v podkapitole 3.1.4. **Hladina spolehlivosti** je v tomhle případě **99 %**. Ta značí, že není možné zamítnout nulovou hypotézu, pokud není pravděpodobnost, že mají prvky náhodné rozmístění, opravdu malá (méně než 1 % pravděpodobnosti).

Poloha adresních bodů Prahy není výsledkem náhodného rozmístění a tyto adresní body se v prostoru výrazně shlukují. To jen potvrzuje výsledky analýzy prostorového uspořádání pro adresní body Česka, které shlukování adresních bodů ve městech zmiňují. Je možné tedy s jistotou mapovat shluky adresních bodů v Praze.

6.2.2 Prostorová autokorelace

Popis této metody je uveden v podkapitole 5.3.2. Metoda *Prostorové autokorelace* byla vyhodnocena nejdříve pro adresní body Prahy agregované do hexagonové sítě, kdy vstupoval atribut s počtem adresních bodů. Následně vstupovaly u této metody v případě Prahy navíc i tři vybrané atributy (BUDOBYEY, SUMBYT, POCPODBUD).

Tato analýza měla za úkol zjistit, zda jsou hodnoty vybraných atributů adresních bodů Prahy uspořádány jistým způsobem nebo zda jsou sousedící hodnoty odlišné od očekávaného rozdělení.

Další nastavení u této metody byl výběr konceptu prostorového vztahu, kdy byla zvolena obecně doporučená metoda konstantní vzdálenosti (Anselin, 1995). Prahová vzdálenost byla nastavena na 2 500 m. Měření probíhalo na základě *Euklidovské vzdálenosti* (přímá vzdálenost). Zvolila se možnost generování grafického výstupu ve formátu HTML.

Grafické výstupy z výsledků této metody jsou zobrazeny na Obr. 6.2.2.1 a Obr. 6.2.2.2. Výsledná hodnota *z-score* byla pro počet adresních bodů Prahy 32,0181. Silně kladná hodnota *z-score* říká, že se hodnoty (počty adresních bodů) v prostoru výrazně **shlukují (hladina spolehlivosti 99 %)**.

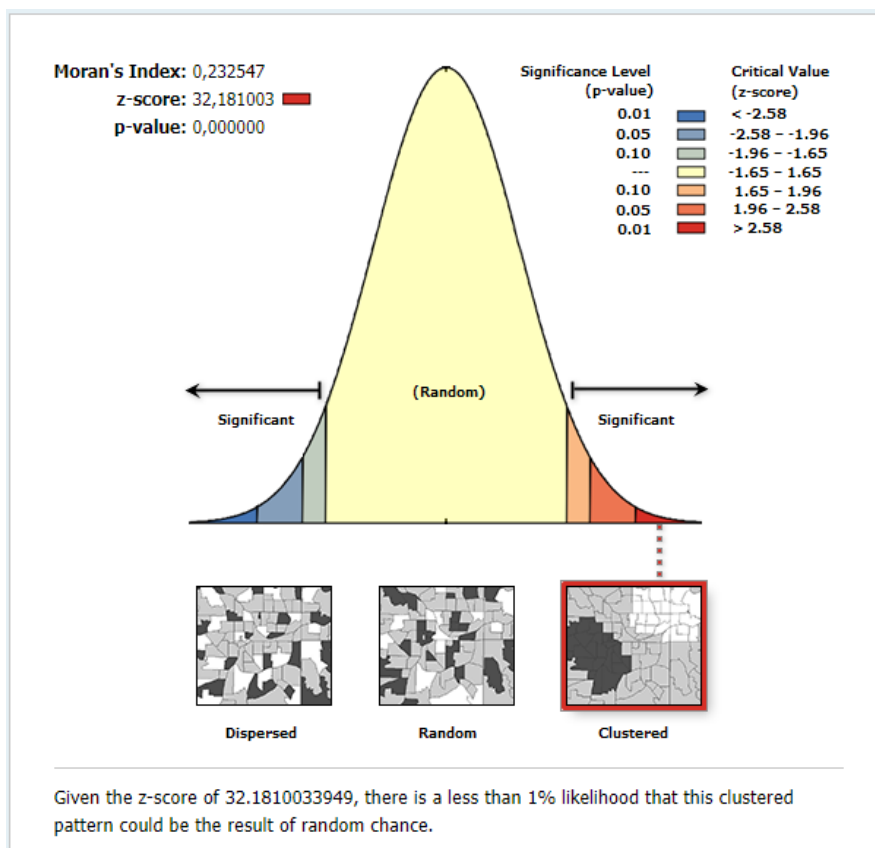
Na území Prahy se tedy budou vyskytovat shluky hexagonů s malou i velkou hustotou adresních bodů.

Výsledná hodnota *z-score* pro atribut BUDOBYEY je 52,259. Silně kladná hodnota *z-score* říká, že se hodnoty (počty evidovaných obyvatel) v prostoru výrazně **shlukují (hladina spolehlivosti 99 %)**.

Výsledná hodnota *z-score* pro atribut SUMBYT je 59,374. Silně kladná hodnota *z-score* říká, že se hodnoty (počty bytů) v prostoru výrazně **shlukují (hladina spolehlivosti 99 %)**.

Výsledná hodnota *z-score* pro atribut POCPODBUD je 58,699. Silně kladná hodnota *z-score* říká, že se hodnoty (počty podlaží budov) v prostoru výrazně **shlukují (hladina spolehlivosti 99 %)**.

V Praze tedy lze mapovat shluky s nadprůměrnými nebo podprůměrnými hodnotami uvedených atributů.



Global Moran's I Summary

Moran's Index:	0,232547
Expected Index:	-0,000899
Variance:	0,000053
z-score:	32,181003
p-value:	0,000000

Obr. 6.2.2.1 Grafický výstup z metody Prostorové autokorelace pro adresní body v Praze.

Moran's Index:	0,377257
Expected Index:	-0,000899
Variance:	0,000052
z-score:	52,259555
p-value:	0,000000

Moran's Index:	0,428419
Expected Index:	-0,000899
Variance:	0,000052
z-score:	59,374350
p-value:	0,000000

Moran's Index:	0,423119
Expected Index:	-0,000899
Variance:	0,000052
z-score:	58,699298
p-value:	0,000000

Obr. 6.2.2.2 Výsledky z metody Prostorová autokorelace pro atribut BUDOBYEV (nahore), SUMBYT (uprostřed), POCPODBUD (dole).

6.3 Mapování shluků

Pro mapování shluků byla zvolena metoda *Analýza shluků vysokých hodnot*. Obecný princip fungování této metody je uveden v podkapitole 3.1.2. V této podkapitole bude uvedeno nastavení a výsledky výpočtů této analýzy.

Vstupními daty byly adresní body agregované do pravidelné hexagonové sítě s plochou hexagonu 0,5 km². Do shlukových analýz vstupoval atribut s počtem adresních bodů a tři vybrané atributy BUDOBYEV, SUMBYT, POCPODBUD. Do metod shlukových analýz vstupovaly postupně hodnoty celkového (sum_nazev) i hodnoty průměrného (avg_nazev) počtu prvků všech vybraných atributů.

6.3.1 Analýza shluků vysokých hodnot (Hotspot analýza)

Popis této metody je uveden v podkapitole 5.4.1. V nastavení shlukových analýz byla použita *Euklidovská vzdálenost*. Prahová vzdálenost byla spočítána na základě průměrné vzdálenosti hexagonu k jeho 30 nejbližším sousedům. Tato průměrná vzdálenost (2233 m) byla zaokrouhlena na 2500 m. Byla použita volitelná korekce FDR (*False Discovery Rate*), která zajistí, že statistická významnost bude založena na míře nesprávného zjištění. Další parametry byly volitelné, a nebylo třeba je nastavovat.

Výsledky a vyhodnocení:

- Shluková analýza Getis Ord G_i^* pro adresní body Prahy v roce 2016 (Příloha 12).
- Shluková analýza Getis Ord G_i^* počtu evidovaných obyvatel Prahy v roce 2016 (Příloha 13).
- Shluková analýza Getis Ord G_i^* průměrného počtu evidovaných obyvatel Prahy v roce 2016 (Příloha 14).
- Shluková analýza Getis Ord G_i^* počtu bytů v Praze v roce 2016 (Příloha 15).
- Shluková analýza Getis Ord G_i^* průměrného počtu bytů v Praze v roce 2016 (Příloha 16).
- Shluková analýza Getis Ord G_i^* počtu podlaží budov v Praze v roce 2016 (Příloha 17).
- Shluková analýza Getis Ord G_i^* průměrného počtu podlaží budov v Praze v roce 2016 (Příloha 18).
- Průměrný počet evidovaných obyvatel u adresních bodů v Praze v roce 2016 (Příloha 19).
- Průměrný počet podlaží budov u adresních bodů v Praze v roce 2016 (Příloha 20).
- Průměrný počet bytů u adresních bodů v Praze v roce 2016 (Příloha 21).

Ve výsledcích shlukových analýz *Getis Ord G_i^** lze pozorovat červené polygony reprezentující statisticky významné shluky vysokých hodnot (*hotspots*) a modré polygony představující statisticky významné shluky nízkých hodnot. Čím je shluk statisticky významnější, tím je barva jeho výplně výraznější (tmavší, sytější).

Ve shlukové analýze *Getis Ord G_i^** pro adresní body Prahy v roce 2016 (Příloha 12) se nacházejí statisticky významný shluk hexagonů s vysokými počty adresních bodů v nich. Tento jeden velký shluk se nachází v centru města a jeho okolí. Z výsledků lze tvrdit, že největší hustota adresních bodů bude v centru města a jeho blízkém okolí, což potvrdily i předchozí analýzy (viz podkapitola 6.1.1). Naopak čtyři velké statisticky

významné shluky hexagonů s nízkým počtem adresních bodů se nachází při hranicích města. Tři z nich u západních hranic a jeden (největší) u východních. I na tyto oblasti a nízkou hustotu adresních bodů v nich bylo poukazováno v předchozích analýzách práce (viz podkapitola 6.1.1).

V mapových výstupech shlukových analýz *Getis Ord Gi** pro tři vybrané atributy je možné pozorovat velmi podobné výsledky. U analýz pro atribut s celkovým počtem (*sum_nazev* – Příloha 13,15 a 17) se u všech zkoumaných atributů vyskytuje jeden velký statisticky významný shluk vysokých hodnot táhnoucí se od středu města severním, jižní a jihovýchodním směrem a pár menších statisticky významných shluků vysokých hodnot nacházejících se západně od velkého shluku. Statisticky významné shluky nízkých hodnot se vyskytují opět při hranicích města, ty největší jsou u hranic východních, severozápadních a jihozápadních. Tyto shluky hodnot atributů v podstatě odpovídají i shlukům počtu adresních bodů. Lze tedy tvrdit, že čím více adresních bodů se na daném místě nachází, tím více se tu bude nacházet i obyvatel, bytů a podlaží budov.

U analýz pro atribut s průměrným počtem (*avg_nazev* – Příloha 14,16 a 18) se u všech zkoumaných opět vyskytuje jeden velký statisticky významný shluk vysokých hodnot táhnoucí se od středu města severním, jižní a jihovýchodním směrem a pár menších statisticky významných shluků vysokých hodnot nacházejících se západně od velkého shluku. Lze pozorovat, že u atributu *avg_POCPodbud* je velký shluk oproti velkým shlukům atributů *avg_Budobyev* a *avg_Sumbyt* na severu ze západní strany „vykouslý“. Z toho lze vyčíst, že v této oblasti se budou nacházet spíše menší budovy (s menším počtem podlaží budov), ale ve větším množství nebo s větším množstvím bytů a obyvatel v nich evidovaných.

Z porovnání výsledků shlukových analýz průměrného počtu evidovaných obyvatel, bytů a podlaží budov v jednom adresním bodu (Příloha 19,20 a 21) lze říci, že nejvyšší průměrný počet obyvatel v jednom bytě není v centru města, ale v jeho okolí (nejčastěji v sídlištích). Dále lze tvrdit, že vysoké budovy (s velkým počtem podlaží) se budou nacházet po celé Praze, ale jejich největší koncentrace je v centru města a směrem severozápadně od něj. Nejvyšší průměrný počet bytů v jednom adresním bodě je opět v sídlištích dál od centra, ve kterém jsou ale hodnoty průměrného počtu bytů v jednom adresním bodě také vysoké.

7 VÝSLEDKY

V této kapitole je uveden přehled všech dílčích výsledků, které byly popsány vždy na konci každé podkapitoly věnující se dané prostorové analýze. Nejprve je zde uveden seznam mapových vizualizací výsledků, poté následuje slovní vyhodnocení těchto vizualizací.

7.1 Mapové výstupy výsledků:

- Průměrný počet nulových hodnot atributů adresních bodů Česka v roce 2016 (Příloha 1).
- Adresní body bez nulových hodnot atributů v České republice v roce 2016 (Příloha 2).
- Průměrný počet nulových hodnot atributů adresních bodů podle krajů v České republice v roce 2016 (Příloha 3).
- Adresní body v pravidelné hexagonové síti s plochou hexagonu 6 km² (Příloha 4).
- Hustota adresních bodů podle obcí v České republice v roce 2016 (Příloha 5).
- Geografická distribuce adresních bodů v České republice v roce 2016 (Příloha 6).
- Elipsy směrodatných odchylek pro obce v České republice v roce 2016 (Příloha 7).
- Shluková analýza Getis Ord Gi* pro adresní body Česka v roce 2016 (Příloha 8).
- Shluková analýza Getis Ord Gi* průměrného počtu nulových hodnot vybraných atributů adresních bodů Česka v roce 2016 (Příloha 9).
- Adresní body Prahy v pravidelné hexagonové síti s plochou hexagonu 0,5 km² v roce 2016 (Příloha 10).
- Hustota adresních bodů Prahy podle ZSJ v roce 2016 (Příloha 11).
- Shluková analýza Getis Ord Gi* pro adresní body Prahy v roce 2016 (Příloha 12).
- Shluková analýza Getis Ord Gi* počtu evidovaných obyvatel Prahy v roce 2016 (Příloha 13).
- Shluková analýza Getis Ord Gi* průměrného počtu evidovaných obyvatel Prahy v roce 2016 (Příloha 14).
- Shluková analýza Getis Ord Gi* počtu bytů v Praze v roce 2016 (Příloha 15).
- Shluková analýza Getis Ord Gi* průměrného počtu bytů v Praze v roce 2016 (Příloha 16).
- Shluková analýza Getis Ord Gi* počtu podlaží budov v Praze v roce 2016 (Příloha 17).
- Shluková analýza Getis Ord Gi* průměrného počtu podlaží budov v Praze v roce 2016 (Příloha 18).
- Průměrný počet evidovaných obyvatel u adresních bodů v Praze v roce 2016 (Příloha 19).
- Průměrný počet podlaží budov u adresních bodů v Praze v roce 2016 (Příloha 20).
- Průměrný počet bytů u adresních bodů v Praze v roce 2016 (Příloha 21).

Všechny mapové výstupy zobrazující území Česka byly sestrojeny v jednotném měřítku 1 : 2 000 000. Pro mapové výstupy Prahy bylo použito jednotné měřítko 1 : 200 000. Všechny mapové výstupy jsou dostupné prostřednictvím webové mapy (vytvořené pomocí serveru ArcGIS Online), která umožňuje vzájemné porovnávání výsledných vrstev výsledků. Odkaz na webovou mapu je uveden na webových stránkách této bakalářské práce. Výstupy jsou ve vázanou přílohou práce a jsou přiloženy na DVD.

7.2 Slovní vyhodnocení výsledků adresních bodů Česka

Z výsledné vizualizace prostorové analýzy nulových hodnot vybraných atributů (Příloha 1) lze vyčíst, že v krajských městech je počet výskytu nulových hodnot vybraných atributů průměrný. Výjimkou jsou krajská města Ústí nad Labem, Jihlava a Brno, kde je počet nulových hodnot atributů nadprůměrný. Naopak v Praze, kde se bude nacházet také největší počet adresních bodů, je výskyt nulových hodnot atributů podprůměrný. Největší množství nulových hodnot se nachází ve venkovských oblastech ve vzdálenějším okolí velkých měst a v horských oblastech u hranic republiky. Naopak adresní body bez nulových hodnot vybraných atributů se vyskytují ve velkém počtu ve velkých městech a jejich aglomeracích (Příloha 2).

Z analýzy výskytu adresních bodů, do které vstupovala agregovaná data (Příloha 4, Příloha 5), lze vyvodit, že největší hustotu adresních bodů mají obce s velkým počtem obyvatel, tedy velká města v České republice. Největší hustota adresních bodů Česka byla zjištěna ve městech jako Praha, Brno, Ostrava, Plzeň, Zlín apod. Na druhou stranu místa s nejmenší hustotou adresních bodů jsou reprezentována vojenskými újezdy, národními parky a dalšími chráněnými přírodními krajinami. Lze pozorovat, že hustota adresních bodů ve venkovských oblastech roste ze západu na východ. Zatímco na západě České republiky mají menší obce v průměru do deseti adresních bodů/km², na východní straně Česka je průměrná hustota kolem 40 adresních bodů/km².

Na Obr. 4.6.1 zobrazujícím porovnání průměrného počtu nulových hodnot v krajích vidíme, že na západě republiky se nachází v průměru více nulových hodnot atributů než na východě. Nejnižší průměrný počet nulových hodnot má kraj Hlavní město Praha, který byl proto zvolen za vhodné území pro podrobnější prostorové analýzy.

Z výsledků měření geografické distribuce (Příloha 6) vidíme, že se všechny zjištěné body popisující geografickou distribuci adresních bodů nacházely ve středu České republiky a byly od sebe poměrně nedaleko. Malá vzdálenost a vzájemná poloha mezi geografickým středem adresních bodů a středem území Česka vypovídá o tom, že jsou adresní body celkem rovnoměrně rozmístěny po celém území České republiky (ve všech směrech od středu státního území).

Jak zobrazuje Obr. 5.2.2.2, extrémní případy v distribuci adresních bodů obcí, u kterých jsou velikosti x-ové a y-ové osy elipsy směrodatné odchylky nejvíce rozdílné, jsou často způsobeny rozdělením dané obce na místní části, které jsou od sebe izolované a které by samy o sobě měly elipsy směrodatných odchylek poměrně „kulaté“. Jelikož tyto místní části sdílí jedno katastrální území, je výsledná elipsa poměrně „šišatá“. Tímto příkladem jsou i obce Měcholupy a Martinice u Onšova (Obr. 5.2.2.2). Dalším důvodem výskytu „šišatých“ elips u obcí je terén (reliéf) území obce. Horské obce často leží v údolí, podél jehož osy se rozrůstají. Přírodní bariérou pro adresní body může být i řeka, jako je tomu u obce Biskoupky (Obr. 5.2.2.3).

Z grafického výstupu metody Průměrného nejbližšího souseda pro bodovou vrstvu adresních bodů Česka (Obr. 5.3.1.1) vidíme, že výsledný poměr je nižší než jedna

(0,314), což indikuje **shlukování** adresních bodů v prostoru. Tento výsledek v podstatě říká, že poloha adresních bodů není výsledkem náhodného rozmístění a že se tyto body v prostoru významně shlukují. Tento výsledek lze předpokládat, jelikož je obecně známo, že se adresní body shlukují v sídlech jako jsou vesnice nebo města.

Z grafického výstupu metody Průměrného nejbližšího souseda pro bodovou vrstvu adresních bodů Česka bez nulových hodnot u vybraných atributů (Obr. 5.3.1.1) vidíme, že je poměr opět nižší než jedna (0,305), což zase vykazuje **shlukování** tentokrát adresních bodů bez nulových hodnot u vybraných atributů. Lze tedy s jistotou mapovat shluky adresních bodů i shluky adresních bodů bez nulových hodnot u vybraných atributů na území Česka.

V grafickém výstupu metody Prostorové autokorelace pro bodovou vrstvu adresních bodů Česka (Obr. 5.3.2.1) je uvedena výsledná hodnota *z-score* (101,255). Silně kladná hodnota *z-score* říká, že se hodnoty (adresní body) v prostoru výrazně **shlukují**. Na území Česka by se tedy budou vyskytovat shluky hexagonů s velkým počtem adresních bodů, ty budou reprezentovat města a hustě osídlené oblasti. Naopak venkovské oblasti či chráněné krajiny budou reprezentovány shluky hexagonů s nízkými počty adresních bodů. Shluky s vyšším počtem adresních bodů a shluky s nižším počtem adresních bodů byly mapovány pomocí shlukových analýz.

V grafickém výstupu metody Prostorové autokorelace pro hodnotu s průměrným počtem nulových hodnot vybraných atributů adresních bodů Česka je výsledná hodnota *z-score* 53,255. Silně kladná hodnota *z-score* říká, že se hodnoty (průměrné počty nulových hodnot u vybraných atributů adresních bodů) v prostoru výrazně **shlukují**. Na území Česka se budou vyskytovat shluky hexagonů s nadprůměrným a podprůměrným počtem nulových hodnot u vybraných atributů adresních bodů, tyto shluky byly mapovány pomocí shlukových analýz.

Ve výsledcích shlukových analýz *Getis Ord Gi** lze pozorovat červené polygony reprezentující statisticky významné shluky vysokých hodnot (*hotspots*) a modré polygony představující statisticky významné shluky nízkých hodnot. Čím je shluk statisticky významnější, tím je barva jeho výplně výraznější (tmavší, sytější).

Ve shlukové analýze *Getis Ord Gi** pro adresní body Česka v roce 2016 (Příloha 8) se nacházejí statisticky významné shluky vysokých hodnot v okolí všech krajských měst s výjimkou Karlových Varů. Největší shluk s vysokým počtem adresních bodů se nachází v Praze a v jejím okolí. Statisticky významné shluky s nízkým počtem adresních bodů se nacházejí v oblastech blízko státních hranic, většinou jde o chráněné přírodní oblasti. Největší shluk s nízkou hustotou adresních bodů se nachází na Šumavě.

Ve shlukové analýze *Getis Ord Gi** průměrného počtu nulových hodnot vybraných atributů adresních bodů Česka v roce 2016 (Příloha 9) se nachází velké množství statisticky významných shluků s vysokým počtem nulových hodnot adresních bodů. Největší z těchto shluků se nachází jižně od Prahy v oblasti Brd a Středočeské pahorkatiny, další z těchto velkých shluků jsou v oblasti Krušných Hor, severozápadně od Brna a v oblasti Moravskoslezských Beskyd. Lze tedy tvrdit, že statisticky významné shluky s vysokým počtem nulových hodnot atributů se nachází v horských přírodních oblastech a pahorkatinách. V menším počtu se zde nacházejí i statisticky významné shluky s nízkým počtem nulových hodnot vybraných atributů adresních bodů. Nejvýznamnější z nich leží v oblasti vojenských újezdů a mezi sídly Znojmo a Mikulov. Největší z nich pak leží v oblasti České tabule.

7.3 Slovní vyhodnocení výsledků adresních bodů Prahy

Největší hustota adresních bodů je v centru města, kde se nachází 100 a více adresních bodů na ploše 0,5 km². Tento výsledek lze předpokládat, jelikož centrum města bývá vždy nejžádanější lokalitou jak pro bydlení, tak pro veřejnou správu i různé zázemí firem. V centru měst je obvykle vysoká hustota zastavěných ploch (snaha využít každý volný prostor určený k zástavbě), proto zde bude i velmi hustá síť adresních bodů.

Velká hustota adresních bodů je i okolí centra podél komunikací, které jsou s ním spojeny. Většinou se zde nachází řadové rodinné domy nebo sídliště, které představují levnější variantu pro bydlení, ale jsou ještě relativně dostupné centru. Hustota adresních bodů v takových oblastech, kde se blízko sebe nachází řadové rodinné domy, panelové domy nebo jiné budovy s byty, které mají většinou více vchodů pro jednu budovu, kdy každý vchod je reprezentován jedním adresním bodem, je vysoká. V případě Prahy se jedná například o Strašnice, Střešovice a Smíchov. Hustotu adresních bodů v těchto oblastech zvyšují také samostatné řadové garáže, které jsou velmi malé a vedle sebe v řadách, kdy každá garáž je reprezentována samostatným adresním bodem.

Nejmenší hustota adresních bodů Prahy je v krajních oblastech města blízko jeho východních a západních hranic. V takovýchto oblastech se nachází 50 a méně adresních bodů na ploše 0,5 km² a někdy se zde dokonce nevyskytují žádné adresní body. V lokalitách na okraji města, které nejsou až tak žádané a parcely v nich jsou tak často větší a levnější se pravděpodobně budou nacházet rodinné domy z většími parcelami nebo velké průmyslové areály, často reprezentované pouze jedním adresním bodem, proto je výsledná hustota adresních bodů nízká.

Výsledky metody *Průměrný nejbližší soused* zobrazuje Obr.6.2.1. Z grafického výstupu lze vyčíst, že poměr (*Nearest Neighbor Ratio*) je 0,314. Poměr je tedy nižší než jedna, což značí shlukování dat. Standardní odchylka *z-score* je silně záporná (-2200). Hodnota *p-value* je 0. Na základě vyhodnocení *z-score* a *p-value*, které je popsáno v podkapitole 3.1.4. Hladina spolehlivosti je v tomhle případě 99 %. Ta značí, že není možné zamítnout nulovou hypotézu, pokud není pravděpodobnost, že mají prvky náhodné rozmístění, opravdu malá (méně než 1 % pravděpodobnosti). Poloha adresních bodů Prahy není výsledkem náhodného rozmístění a tyto adresní body se v prostoru výrazně shlukují. To jen potvrzuje výsledky analýzy prostorového uspořádání pro adresní body Česka, které shlukování adresních bodů ve městech zmiňují. Je možné tedy s jistotou mapovat shluky adresních bodů v Praze.

Grafické výstupy z výsledků metody *Prostorová autokorelace* jsou zobrazeny na Obr. 6.2.2.1 a Obr. 6.2.2.2. Výsledná hodnota *z-score* byla pro počet adresních bodů Prahy 32,0181. Silně kladná hodnota *z-score* říká, že se hodnoty (počty adresních bodů) v prostoru výrazně shlukují (hladina spolehlivosti 99 %). Na území Prahy se tedy budou vyskytovat shluky hexagonů s malou i velkou hustotou adresních bodů. Výsledná hodnota *z-score* pro atribut BUDOBYEV je 52,259. Silně kladná hodnota *z-score* říká, že se hodnoty (počty evidovaných obyvatel) v prostoru výrazně shlukují (hladina spolehlivosti 99 %). Výsledná hodnota *z-score* pro atribut SUMBYT je 59,374. Silně kladná hodnota *z-score* říká, že se hodnoty (počty bytů) v prostoru výrazně shlukují (hladina spolehlivosti 99 %). Výsledná hodnota *z-score* pro atribut POCPODBUD je 58,699. Silně kladná hodnota *z-score* říká, že se hodnoty (počty podlaží budov) v prostoru výrazně shlukují (hladina spolehlivosti 99 %). V Praze tedy lze mapovat shluky s nadprůměrnými nebo podprůměrnými hodnotami uvedených atributů.

Ve výsledcích shlukových analýz *Getis Ord Gi** lze pozorovat červené polygony reprezentující statisticky významné shluky vysokých hodnot (*hotspots*) a modré polygony představující statisticky významné shluky nízkých hodnot. Čím je shluk statisticky významnější, tím je barva jeho výplně výraznější (tmavší, sytější).

Ve shlukové analýze *Getis Ord Gi** pro adresní body Prahy v roce 2016 (Příloha 12) se nacházejí statisticky významný shluk hexagonů s vysokými počty adresních bodů v nich. Tento jeden velký shluk se nachází v centru města a jeho okolí. Z výsledků lze tvrdit, že největší hustota adresních bodů bude v centru města a jeho blízkém okolí, což potvrdily i předchozí analýzy (viz podkapitola 6.1.1). Naopak čtyři velké statisticky významné shluky hexagonů s nízkým počtem adresních bodů se nachází při hranicích města. Tři z nich u západních hranic a jeden (největší) u východních. I na tyto oblasti a nízkou hustotu adresních bodů v nich bylo poukazováno v předchozích analýzách práce (viz podkapitola 6.1.1).

V mapových výstupech shlukových analýz *Getis Ord Gi** pro tři vybrané atributy lze pozorovat velmi podobné výsledky. U analýz pro atribut s celkovým počtem (*sum_nazev* – Příloha 13,15 a 17) se u všech zkoumaných atributů vyskytuje jeden velký statisticky významný shluk vysokých hodnot táhnoucí se od středu města severním, jižní a jihovýchodním směrem a pár menších statisticky významných shluků vysokých hodnot nacházejících se západně od velkého shluku. Statisticky významné shluky nízkých hodnot se vyskytují opět při hranicích města, ty největší jsou u hranic východních, severozápadních a jihozápadních. Tyto shluky hodnot atributů v podstatě odpovídají i shlukům počtu adresních bodů. Lze tedy tvrdit, že čím více adresních bodů se na daném místě nachází, tím více se tu bude nacházet i obyvatel, bytů a podlaží budov.

U analýz pro atribut s průměrným počtem (*avg_nazev* – Příloha 14,16 a 18) se u všech zkoumaných opět vyskytuje jeden velký statisticky významný shluk vysokých hodnot táhnoucí se od středu města severním, jižní a jihovýchodním směrem a pár menších statisticky významných shluků vysokých hodnot nacházejících se západně od velkého shluku. Lze pozorovat, že u atributu *avg_POCPodbud* je velký shluk oproti velkým shlukům atributů *avg_Budobyev* a *avg_Sumbyt* na severu ze západní strany „vykouslý“. Z toho lze vyčíst, že v této oblasti se budou nacházet spíše menší budovy (s menším počtem podlaží budov), ale ve větším množství nebo s větším množstvím bytů a obyvatel v nich evidovaných.

Z porovnání výsledků shlukových analýz průměrného počtu evidovaných obyvatel, bytů a podlaží budov v jednom adresním bodu (Příloha 19,20 a 21) lze říci, že nejvyšší průměrný počet obyvatel v jednom bytě není v centru města, ale v jeho okolí (nejčastěji v sídlištích). Dále lze tvrdit, že vysoké budovy (s velkým počtem podlaží) se budou nacházet po celé Praze, ale jejich největší koncentrace je v centru města a směrem severozápadně od něj. Nejvyšší průměrný počet bytů v jednom adresním bodě je opět v sídlištích dál od centra, ve kterém jsou ale hodnoty průměrného počtu bytů v jednom adresním bodě také vysoké.

8 DISKUZE

V průběhu této práce bylo řešeno několik problémů a muselo se učinit rozhodnutí pro jejich řešení. Některé problémy byly vyřešeny bez větších potíží, jiné musely být řešeny mnohdy pracnějším postupem.

Úvodním problémem, který se musel řešit hned při první analýze poskytnutých dat s adresními body, byl ten, že adresní body obsahovali u 71 % z nich duplikované (chybné) hodnoty. Tyto chybné adresní body měli správnou polohu v prostoru, ale jejich atributy obsahovaly hodnoty jednoho adresního bodu ležícího v Chocni. Tato chyba v datech byla nahlášena ČSÚ a následně autor práce obdržel novou databázi adresních bodů, kde se již tato chyba nevyskytovala.

Každý adresní bod v nové databázi obsahoval 63 atributů, většina z nich však obsahovala kvalitativní nominální (jedinečné) hodnoty, které nelze porovnat, a proto ani nejsou vhodné pro vstup do prostorových analýz. Často se jednalo o číselné identifikátory. Tyto nevhodné atributy nebyly tedy při analýzách využity. Bylo vybráno 23 atributů adresních bodů s potenciálem pro prostorové analýzy, které byly dále zkoumány.

Dalším ze zjištěných problémů u poskytnutých dat byl ten, že většina ze 23 vybraných atributů s potenciálem pro prostorové analýzy obsahovala velké množství nulových hodnot atributů. Proto byla u vybraných atributů provedena analýza počtu nulových hodnot, která porovnávala procentuální zastoupení nulových hodnot z celkového počtu hodnot u vybraných atributů. Procentuální zastoupení nulových hodnot u vybraných atributů bylo tak velké, že bylo rozhodnuto provést také prostorovou analýzu nulových hodnot vybraných atributů, která měla navíc pomoci rozhodnout, jaký z regionů Česka bude nejvhodnější pro podrobnější prostorové analýzy.

Regionem pro podrobnější průzkum se měl stát jeden z krajů. S pomocí prostorové analýzy nulových hodnot vybraných atributů byl zjištěn kraj, ve kterém se nacházel v průměru nejmenší počet nulových hodnot na jeden adresní bod. Kraj Hlavní město Praha měl nejlepší výsledky, a proto byl i vybrán pro provádění podrobnějších prostorových analýz.

Pro správnou vizualizaci výsledků prostorových analýz bylo nutné vzhledem k velkému počtu adresních bodů vstupní bodová data agregovat do pravidelné sítě polygonů a do nich provést statistický přepočítání hodnot vybraných atributů.

Po konzultaci s vedoucím práce bylo rozhodnuto vytvořit dvě pravidelné polygonové sítě, kdy byl za optimální základní polygon pravidelné sítě zvolen hexagon s velikostí plochy 6km² pro prostorové analýzy na území Česka a hexagon s velikostí plochy 3,5km² pro území vybraného regionu. Tvar hexagonu byl zvolen jako nejvhodnější z nabízených, protože umožňoval přístup k co největšímu počtu sousedů. Lokální metody shlukových analýz jsou založeny na výpočtu vzdáleností mezi sousedními body. Střed hexagonu má ke všem středům svých sousedů stejnou vzdálenost, proto byla volba hexagonu nejlepší možnou variantou výběru. Generování těchto gridových struktur proběhlo pomocí využití extenze snadno a rychle.

Časově náročnější proces, jehož výpočet trval v řádu hodin, představoval statistický přepočítání hodnot vybraných atributů adresních bodů do hexagonových sítí. Ale i ten proběhl jinak bez větších problémů.

Pro popis prostorové charakteristiky adresních bodů byly použity metody programu ArcMap 10.4 z ArcGIS Desktop. Konkrétně byla využita sada nástrojů *Spatial Statistics*

Tools. Nebyly využity funkce dalších známých statistických software – *R project*, *STATISTICA*, *GeoDA*, což ale nebylo potřeba vzhledem k zadání práce. Program *ArcMap 10.4* z *ArcGIS Desktop* nabízel vhodné znázornění výsledků prostorových analýz a nabízené metody těchto analýz byly vyhodnoceny za dostačující. Po konzultaci s vedoucím byly v rešeršní části práce uvedeny stručné popisy všech metod, které tato sada nástrojů nabízí a které byly nastudovány v rámci hledání vhodných metod prostorové statistiky pro adresní body. Ovšem v praktické části byly vzhledem k danému rozsahu práce a časovým možnostem využity jen některé z nich.

Pro shlukové analýzy na území Prahy bylo nutné vybrat atributy, které budou vstupovat do analýz. Autor práce se rozhodl po domluvě s vedoucím využít 3 atributů. Konkrétně se jednalo o atribut s počtem evidovaných obyvatel (BUDOBYEV), atribut s počtem bytů (SUMBYT) a atribut s počtem podlaží budov (POCPODBUD).

Na práci by bylo vhodné navázat magisterskou či diplomovou prací, jelikož zcela určitě nebyl plně využit potenciál prostorových analýz a adresních bodů do těchto analýz vstupujících. Na druhou stranu je rozumné počkat s dalšími prostorovými analýzami minimálně do roku 2021, respektive do doby, kdy budou dostupná aktuální data. V roce 2021 totiž proběhne SLDB 2021 (Sčítání lidu, domů a bytů v roce 2021), které probíhá pouze jednou za 10 let. Poslední sčítání tohoto typu proběhlo v roce 2011. Většina vybraných atributů adresních bodů s potenciálem pro prostorové analýzy uvádí hodnoty právě z tohoto sčítání. Vstup těchto „starých“ hodnot do prostorových analýz neumožňuje získat aktuální výsledek, a navíc právě většina z těchto hodnot jsou hodnoty nulové (nevhodné pro prostorové analýzy). Zbývá jen doufat, že rok 2021 přinese novou aktuální databázi adresních bodů, která bude obnášet pravdivé aktuální hodnoty popisující vlastnosti jednotlivých adresních bodů.

Výsledky práce mohou sloužit jako doporučení při nastavování prostorových analýz a při práci se statistickými bodově lokalizovanými daty.

9 ZÁVĚR

Tato bakalářská práce měla za hlavní cíl aplikovat metody prostorové statistiky na adresní body Česka a vyhodnotit tak jejich prostorovou distribuci a prostorové vztahy na základě jejich vzájemné polohy a hodnot jejich atributů.

Byla provedena analýza databáze adresních bodů, při které bylo zjištěno, že pro vizualizaci výsledků prostorových analýz bude nutné data agregovat do pravidelné gridové sítě, zvoleným základním tvarem pro síť byl hexagon. Byly vytvořeny dvě hexagonové sítě, jedna určená pro prostorové analýzy na území Česka a druhá pro podrobnější prostorové analýzy ve vybraném menším regionu.

Dalším poměrně velkým problémem bylo zjištění, že mnoho atributů adresních bodů obsahovalo nulové hodnoty (hodnoty 0/NULL), které by zkreslovaly výsledky prostorových analýz využívajících hodnot atributů adresních bodů. Proto bylo rozhodnuto provést analýzu těchto nulových hodnot u vybraných atributů s potenciálem pro prostorové analýzy. U většiny hodnocených atributů bylo procentuální zastoupení nulových hodnot vysoké, proto bylo rozhodnuto nulové hodnoty mapovat v prostoru.

Následně tedy byla provedena prostorová analýza nulových hodnot u vybraných atributů adresních bodů, při které byla vytvořena pravidelná hexagonová síť nesoucí informace o počtu adresních bodů bez nulových hodnot atributů a o průměrném počtu nulových hodnot v jednom adresním bodě pro každý hexagon zvlášť.

Pro podrobnější analýzy bylo rozhodnuto vybrat jeden z krajů České republiky. K výběru vhodného kraje, tedy kraje s co možná nejmenším počtem nulových hodnot byly využity výsledky prostorové analýzy nulových hodnot. Na základě těchto výsledků byl zvolen za nejvhodnější kraj Hlavní město Praha, jehož adresní body obsahovaly v průměru nejmenší počet nulových atributů.

Data byla připravena a přešlo se k hlavnímu cíli práce, tedy výpočtu prostorové statistiky. Prostorová statistika obnášela měření geografické distribuce, analýzu prostorového uspořádání a mapování statisticky významných shluků adresních bodů a také nulových hodnot vybraných atributů adresních bodů. Prostorová statistika byla nejdříve vypočítána pro celé území republiky a následně pak pro kraj Hlavní město Praha.

Výsledkem prostorových analýz byla série tabulek, grafů a mapových výstupů s vhodnou vizualizací dat a kartograficky korektní mapovou kompozicí pro vyhodnocení výsledků a vzájemné porovnání. Nakonec bylo provedeno slovní vyhodnocení výsledků analýz. Pro lepší porovnání výsledků prostorových analýz a také chápání geografického kontextu těchto výsledků byla vytvořena webová mapa prostřednictvím serveru ArcGIS online.

V krajských městech je počet výskytu nulových hodnot vybraných atributů průměrný. Výjimkou jsou krajská města Ústí nad Labem, Jihlava a Brno, kde je počet nulových hodnot atributů nadprůměrný. Naopak v Praze, kde se bude nacházet také největší počet adresních bodů, je výskyt nulových hodnot atributů podprůměrný. Největší množství nulových hodnot se nachází ve venkovských oblastech ve vzdálenějším okolí velkých měst a v horských oblastech u hranic republiky. Adresní body bez nulových hodnot vybraných atributů se vyskytují ve velkém počtu ve velkých městech a jejich aglomeracích. Měření prostorové distribuce prokázalo, že jsou adresní body rozmístěny po celém území Česka, a to poměrně rovnoměrně (z celkového pohledu). Bylo zjištěno, že prostorový vzor adresních bodů Česka i Prahy vykazuje

statisticky významné shlukování. Adresní body se v Česku shlukují nejvíce a nejvýznamněji ve velkých městech s vysokým počtem obyvatel v něm žijícím. V Praze se adresní body nejvíce shlukují v centru a v jeho blízkém okolí. Nejmenší počet adresních bodů v Česku je v oblastech vojenských újezdů a chráněných přírodních oblastech.

Závěrem bylo sepsání textové části práce a vyplnění všech údajů o vytvořených datových sadách do metainformačního systému katedry geoinformatiky, jak bylo v zadání práce. Celá práce, tedy textová část, přílohy, výstupy, zdrojová a vytvořená data, XML soubor, je dostupná v příloze na DVD. Pro tuto práci byly také vytvořeny webové stránky a informační poster, ve kterých jsou prezentovány výsledky práce.

POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE

AGHAJANI, Mohammad Ali, Reza Shahni DEZFOULIAN, Abdolreza Rezaee ARJROODY a Mohammadreza REZAEI. Applying GIS to Identify the Spatial and Temporal Patterns of Road Accidents Using Spatial Statistics. Computers and Electronics in Agriculture. [online]. 2017, 25, 2126-2138. ISSN 2352-1465. [cit. 2018-01-10]. Dostupné z WWW: <https://www.sciencedirect.com/science/article/pii/S2352146517307160>

ARONOFF, Stan. Geographic Information Systems: A Management Perspective [online]. Ottawa: WDL Publications, 1989. ISBN 0-9218404-00-8. [cit. 2017-10-10]. Dostupné z WWW: https://www.researchgate.net/publication/44500025_Geographic_information_systems_a_management_perspective Stan_Aronoff

ArcČR® 500 [online]. ©ArcČR, ARCDATA PRAHA, ZÚ, ČSÚ, 2016. Praha: ARCDATA PRAHA, s.r.o., Zeměměřický úřad, ČSÚ, 2017 [cit. 2017-10-10]. Dostupné z WWW: <https://www.arcdata.cz/produkty/geografickadata/arccr-500>

ANSELIN, Luc. Local indicators of spatial association – LISA [online]. Geographical Analysis, 1995, (27), 93–115. [cit. 2017-10-10]. Dostupné z WWW: http://dces.wisc.edu/wp-content/uploads/sites/30/2013/08/W4_Anselin1995.pdf

CLIFF, Andrew D. a J. Keith ORD. Spatial autocorrelation. London: Pion, 1973, 178 s. ISBN 0850860369. [cit. 2017-10-10].

ČSÚ. Adresa [online]. Praha: ČSÚ, 2016 [cit. 2017-10-10]. Dostupné z WWW: <https://www.czso.cz/csu/rso/adresa>

ČÚZK. Adresní místa RÚIAN ve formátu CSV. Nahlížení do katastru nemovitostí [online]. Praha, 2017 [cit. 2017-10-10]. Dostupné z WWW: <http://nahlizeniidokn.cuzk.cz/stahniadresnimistaruiian.aspx>

ESRI. ArcGIS Pro Help [online]. 2017 [cit. 2017-10-10]. Dostupné z WWW: <http://pro.arcgis.com/en/pro-app/help/main>

FOTHERINGHAM, A. Stewart, Chris BRUNSDON, Martin CHARLTON. Quantitative geography – Perspectives on spatial data analysis. London: SAGE Publications, London, 2000, 270 s. ISBN 0-7619-5947-5. [cit. 2017-10-10].

GETIS, Arthur a J. Keith ORD. The Analysis of Spatial Association by Use of Distance Statistics. Geographical Analysis. Ohio State University Press, 1992, (24), 286-306. [cit. 2017-10-10].

GOODCHILD, Michael Frank. A spatial analytical perspective on geographical information systems [online]. International Journal of Geographical Information Systems. 1987, (1), 335-354. [cit. 2017-10-10]. Dostupné z WWW: <http://www.geog.ucsb.edu/~good/papers/95.pdf>

HARRIES, Keith. Mapping Crime: Principles and Practice [online]. Washington, DC: The National Institute of Justice, 1999. 206 s. [cit. 2017-10-10]. Dostupné z WWW: <https://www.ncjrs.gov/pdffiles1/nij/178919.pdf>

HORÁK, Jiří. Prostorové analýzy dat. 3. vyd., Ostrava: VŠB-TU Ostrava, HGF, Institut geoinformatiky, 2011, 170 s. [cit. 2017-10-10]

HORÁK, Jiří. Metodika harmonizace, agregace a anonymizace dat kriminality [online]. Ostrava: VŠB-TU Ostrava, HGF, Institut geoinformatiky, 2016 [cit. 2017-10-10]. Dostupné z WWW: https://www.researchgate.net/publication/303820747_Metodika_harmonizace_agregace_a_anonymizace_dat_kriminality

HORÁK, Jiří, Igor IVAN a Tomáš INSPEKTOR. Monitoring of Socially Excluded Localities of Ostrava City [online]. Ostrava: VŠB-TU Ostrava, HGF, Institut geoinformatiky, 2010, 14 s. ISBN 978-80-248-2171-9. [cit. 2017-10-10]. Dostupné z WWW: https://www.researchgate.net/publication/264844614_Monitoring_of_Socially_Excluded_Localities_of_Ostrava_City

CHANEY, Spencer a Jerry RATCLIFFE. GIS and Crime Mapping: Mastering GIS: Technol, Applications & Mgmt. John Wiley & Sons, 2013, 448 s. ISBN 9780470860984. [cit. 2017-10-10]

INSPEKTOR, Tomáš, Igor IVAN a Jiří HORÁK. Mapping and monitoring unemployment hot spots towards identification of socially excluded localities: Case study of Ostrava [online]. Journal of Maps. 2013, 10(1), 35-46. DOI: 10.1080/17445647.2013.847806. ISSN 1744-5647. [cit. 2017-10-10]. Dostupné z WWW: <https://www.tandfonline.com/doi/full/10.1080/17445647.2013.847806>

JENNESS, Jeff. Repeating Shapes. Jenness Enterprises [online]. USA, 2012 [cit. 2017-10-10]. Dostupné z WWW: http://www.jennessent.com/arcgis/repeat_shapes.htm

JOHNSTON, R. J., Derek GREGORY a David M. SMITH. The Dictionary of human geography [online]. 3. vyd., Cambridge, Mass., USA: Blackwell Reference, 1994, 724 s. ISBN 978-063-1181-415. [cit. 2017-10-15]. Dostupné z WWW: http://www.univpgri-palembang.ac.id/perpus_fkip/Perpustakaan/Geography/Kamus%20Geografi/Kamus%20Geografi%20Manusia.pdf

KLAUDA, Petr. Prostorově určená statistická data. Statistika&My [online]. Praha: ČSÚ, 2016, 6(5), 18-19. ISSN 1804-7149. [cit. 2017-10-10]. Dostupné z WWW: <http://www.statistikaamy.cz/2016/05/prostorove-urcena-statisticka-data/>

KLÍCHA, Michal. Shluková analýza funkčních ploch pro sledování urbánního vývoje olomouckého regionu [online]. Olomouc, 2011. Bakalářská. Univerzita Palackého, PřF, Katedra geoinformatiky. [cit. 2017-07-15]. Dostupné z WWW:
http://gisak.vsb.cz/GISacek/GISacek_2011/papers/klicha.pdf

MORAN, Patrik A. P. Notes on Continuous Stochastic Phenomena [online]. Biometrika. 1950, 37(1), 17-23. [cit. 2017-07-15]. Dostupné z WWW:
[https://dds.cepal.org/infancia/guide-to-estimating-child-poverty/bibliografia/capitulo-IV/Moran%20Patrick%20A%20P%20\(1950\)%20Notes%20on%20continuous%20stochastic%20phenomena.pdf](https://dds.cepal.org/infancia/guide-to-estimating-child-poverty/bibliografia/capitulo-IV/Moran%20Patrick%20A%20P%20(1950)%20Notes%20on%20continuous%20stochastic%20phenomena.pdf)

MITCHELL, Andy. The ESRI guide to GIS analysis. 2. vyd., Redlands: ESRI Press, 2005, 252 s. ISBN 978-158-9481-169. [cit. 2017-07-15]

PEETERS, Aviva, Manuela ZUDE, Jana KÄTHNER, Mustafa ÜNLÜ, Riza HETZRONI, Robin a Alon GEBBERS. Getis-Ord's hot- and cold-spot statistics as a basis for multivariate spatial clustering of orchard tree data [online]. Computers and Electronics in Agriculture, 2015, (111), 140-150. [cit. 2017-07-15]. Dostupné z WWW:
http://www.academia.edu/11180133/Getis_Ord_s_hot_and_cold-spot_statistics_as_a_basis_for_multivariate_spatial_clustering_of_orchard_tree_data

PFEIFFER, Dirk, Timothy P. ROBINSON, Mark STEVENSON, Kim B. STEVENS, David J. ROGERS a Archie C.A. CLEMENTS. Spatial analysis in epidemiology [online]. New York: Oxford University Press, 2008, 160 s. ISBN 978-019-8509-882. [cit. 2017-07-15] Dostupné z WWW:
https://www.researchgate.net/publication/43508791_Spatial_Analysis_in_Epidemiology

RAPANT, Petr. Geoinformatika a geoinformační technologie [online]. Ostrava: Ostrava: VŠB-TU Ostrava, HGF, Institut geoinformatiky, 2006, 463 s. ISBN: 9781589481169. [cit. 2017 07-15] Dostupné z WWW:
https://www.researchgate.net/publication/40387847_Geoinformatika_a_geoinformacni_technologie

RASAM, A. R. A., R. Ghazali, A. M. M. Noor, W. M. N. W. Mohd, J. R. A. Hamid, M. J. Bazlan A N. Ahmad, Spatial epidemiological techniques in cholera mapping and analysis towards a local scale predictive modelling [online]. IOP Conference Series: Earth and Environmental Science, 2014(18), 1-6. [cit. 2017 07-15]. Dostupné z WWW:
<http://iopscience.iop.org/article/10.1088/1755-1315/18/1/012095/meta>

SCOTT, Lauren, Nathan Warmerdam. Extend crime analysis with ArcGIS spatial statistics tools [online]. ArcUser Magazine. (2005)
ArcUser Magazine, April-June [USA] [cit. 2017 07-15]. Dostupné z WWW:
http://www.esri.com/library/reprints/pdfs/arcuser_extend-crime-analysis.pdf

SPURNÁ, Pavlína. Prostorová autokorelace [online]. Sociologický časopis. 2008, (44), 767-787. Dostupné z: <http://sreview.soc.cas.cz/cs/issue/7-sociologicky-casopis-czech-sociological-review-4-2008/78>

ŠARMANOVÁ, Jana. Metody analýzy dat: učební text [online]. Ostrava: VŠB-TU Ostrava, HGF, Institut geoinformatiky, 2012, 170 s. ISBN 978-80-248-2565-6. [cit. 2017 07 15]. Dostupné z WWW: <http://www.person.vsb.cz/archivcd/FEI/MAD/MAD.pdf>

TOBLER, W. R. A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography. Clark University, 1970, 46, 234-240. DOI: 10.2307/143141. ISSN 00130095. [cit. 2017 07 10].

UNWIN, David J. Introductory spatial analysis. New York: Methuen, 1981, 212 s. ISBN 0-416-72190-7. [cit. 2017 07 10].

VOŽENÍLEK, Vít a Jaromír KAŇOK. Metody tematické kartografie: vizualizace prostorových jevů. Olomouc: Univerzita Palackého v Olomouci pro katedru geoinformatiky, 2011. ISBN 978-802-4427-904. [cit. 2017 07 10].

VÚGTK. Terminologický slovník zeměměřičství a katastru nemovitostí [online]. VÚGTK © 2005-2018. [cit. 27.11.2006]. Dostupné z WWW: <http://www.vugtk.cz/slovník/>

ZAJÍCOVÁ, Hanka. Srovnání prostorové reprezentace adresních bodů – grid/areály [online]. Olomouc, 2017. Bakalářská. Univerzita Palackého, PřF, Katedra geoinformatiky. [cit. 2017-07-15]. Dostupné z WWW: <http://www.geoinformatics.upol.cz/dprace/bakalarske/zajicova17/pages/metody.html>

PŘÍLOHY

SEZNAM PŘÍLOH

Vevázané přílohy:

- Příloha 1 Průměrný počet nulových hodnot atributů adresních bodů Česka v roce 2016
- Příloha 2 Adresní body bez nulových hodnot atributů v České republice v roce 2016
- Příloha 3 Průměrný počet nulových hodnot atributů adresních bodů podle krajů v České republice v roce 2016
- Příloha 4 Adresní body v pravidelné hexagonové síti s plochou hexagonu 6 km²
- Příloha 5 Hustota adresních bodů podle obcí v České republice v roce 2016
- Příloha 6 Geografická distribuce adresních bodů v České republice v roce 2016
- Příloha 7 Elipsy směrodatných odchylek pro obce v České republice v roce 2016
- Příloha 8 Shluková analýza Getis Ord Gi* pro adresní body Česka v roce 2016
- Příloha 9 Shluková analýza Getis Ord Gi* průměrného počtu nulových hodnot vybraných atributů adresních bodů Česka v roce 2016
- Příloha 10 Adresní body Prahy v pravidelné hexagonové síti s plochou hexagonu 0,5 km² v roce 2016
- Příloha 11 Hustota adresních bodů Prahy podle ZSJ v roce 2016
- Příloha 12 Shluková analýza Getis Ord Gi* pro adresní body Prahy v roce 2016
- Příloha 13 Shluková analýza Getis Ord Gi* počtu evidovaných obyvatel Prahy v roce 2016
- Příloha 14 Shluková analýza Getis Ord Gi* průměrného počtu evidovaných obyvatel Prahy v roce 2016
- Příloha 15 Shluková analýza Getis Ord Gi* počtu bytů v Praze v roce 2016
- Příloha 16 Shluková analýza Getis Ord Gi* průměrného počtu bytů v Praze v roce 2016
- Příloha 17 Shluková analýza Getis Ord Gi* počtu podlaží budov v Praze v roce 2016
- Příloha 18 Shluková analýza Getis Ord Gi* průměrného počtu podlaží budov v Praze v roce 2016
- Příloha 19 Průměrný počet evidovaných obyvatel u adresních bodů v Praze v roce 2016
- Příloha 20 Průměrný počet podlaží budov u adresních bodů v Praze v roce 2016
- Příloha 21 Průměrný počet bytů u adresních bodů v Praze v roce 2016

Volné a elektronické přílohy:

- Příloha 22 Adresní body v pravidelné hexagonové síti s plochou hexagonu 3,5 km²
- Příloha 23 Hustota adresních bodů podle ZSJ v České republice v roce 2016
- Příloha 24 Poster
- Příloha 25 DVD

Popis struktury DVD (adresáře):

Data

Metadata

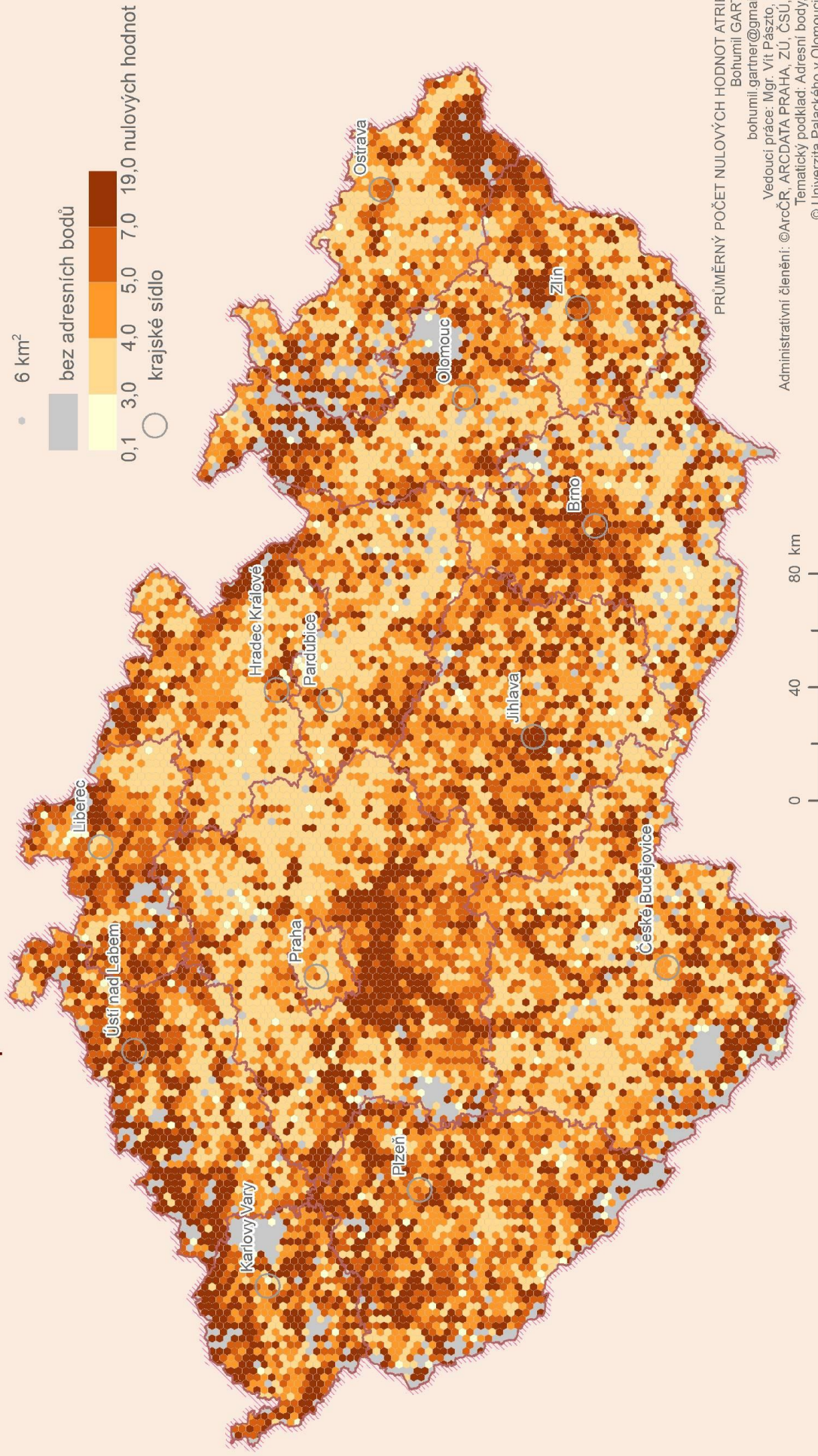
Web

Text_prace

Přilohy

PRŮMĚRNÝ POČET NULOVÝCH HODNOT ATRIBUTŮ

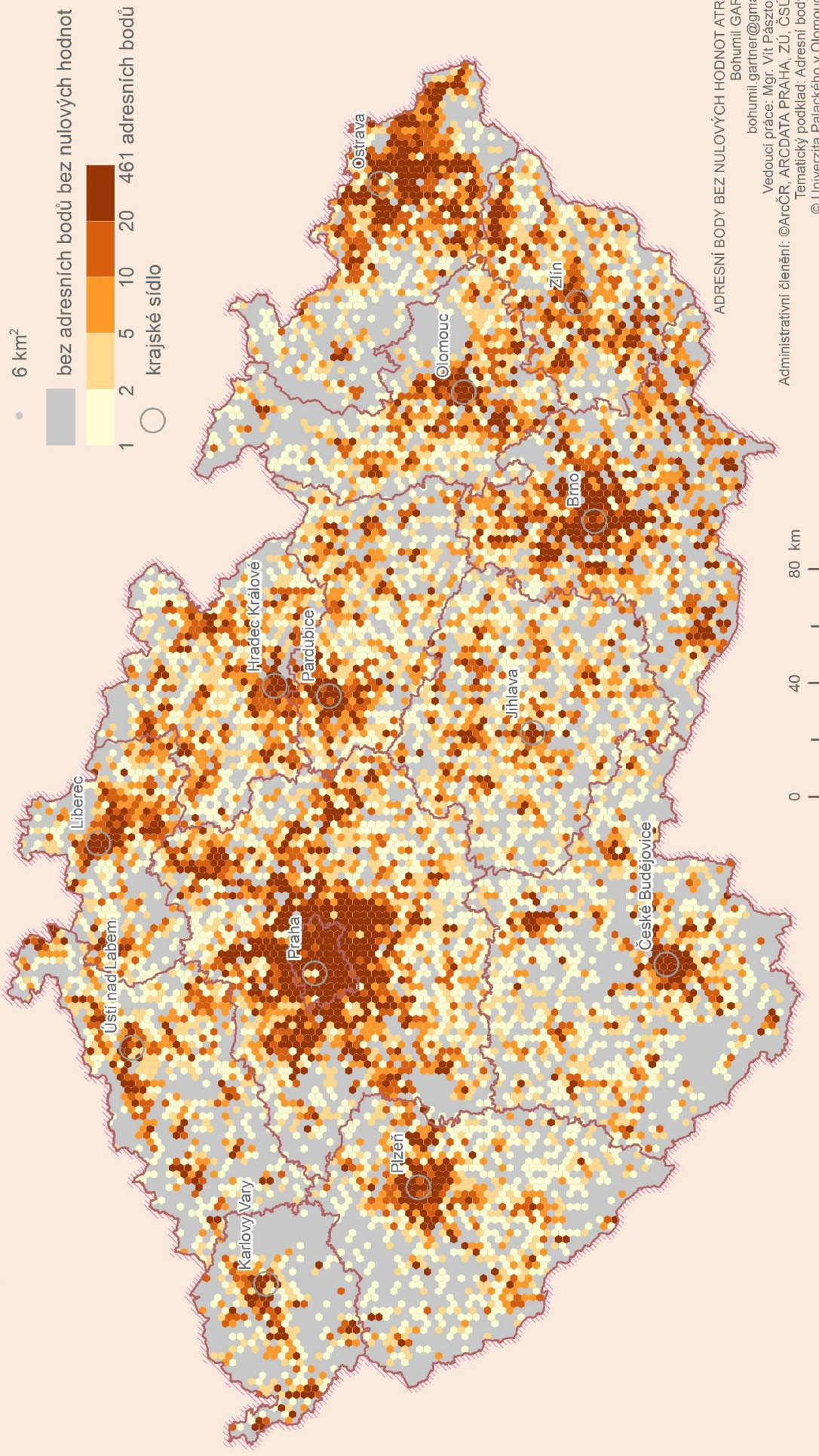
adresních bodů v České republice v roce 2016



PRŮMĚRNÝ POČET NULOVÝCH HODNOT ATRIBUTŮ
Bohumil GARTNER
bohumil.gartner@gmail.com
Vedoucí práce: Mgr. Vit Pászto, Ph.D.
Administrativní členění: ©ArcCR, ARCDATA PRAHA, ZÚ, ČSÚ, 2016
Tematický podklad: Adresní body, ČSÚ
© Univerzita Palackého v Olomouci, 2018
Příloha č.1

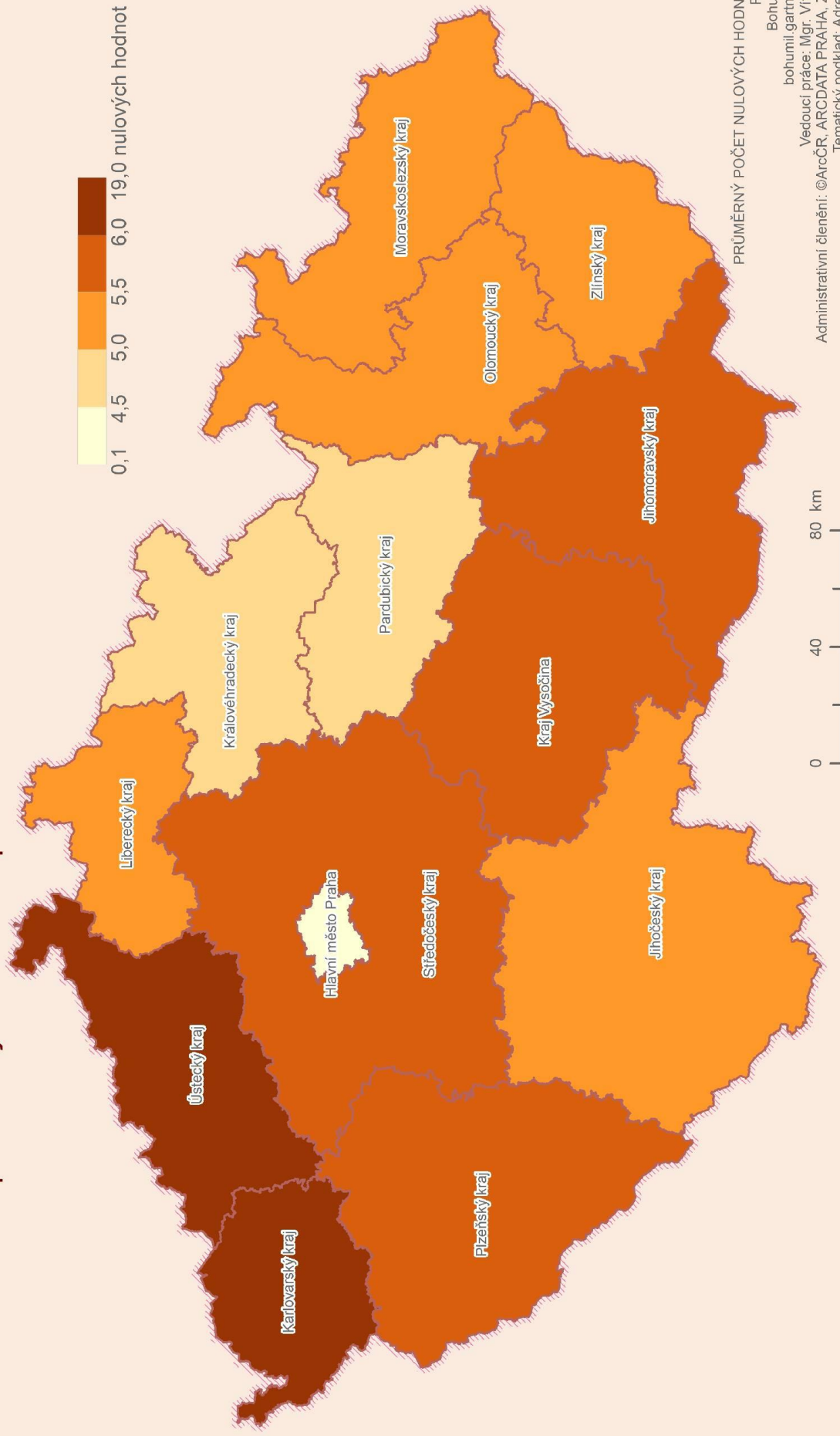
ADRESNÍ BODY BEZ NULOVÝCH HODNOT ATRIBUTŮ

v České republice v roce 2016



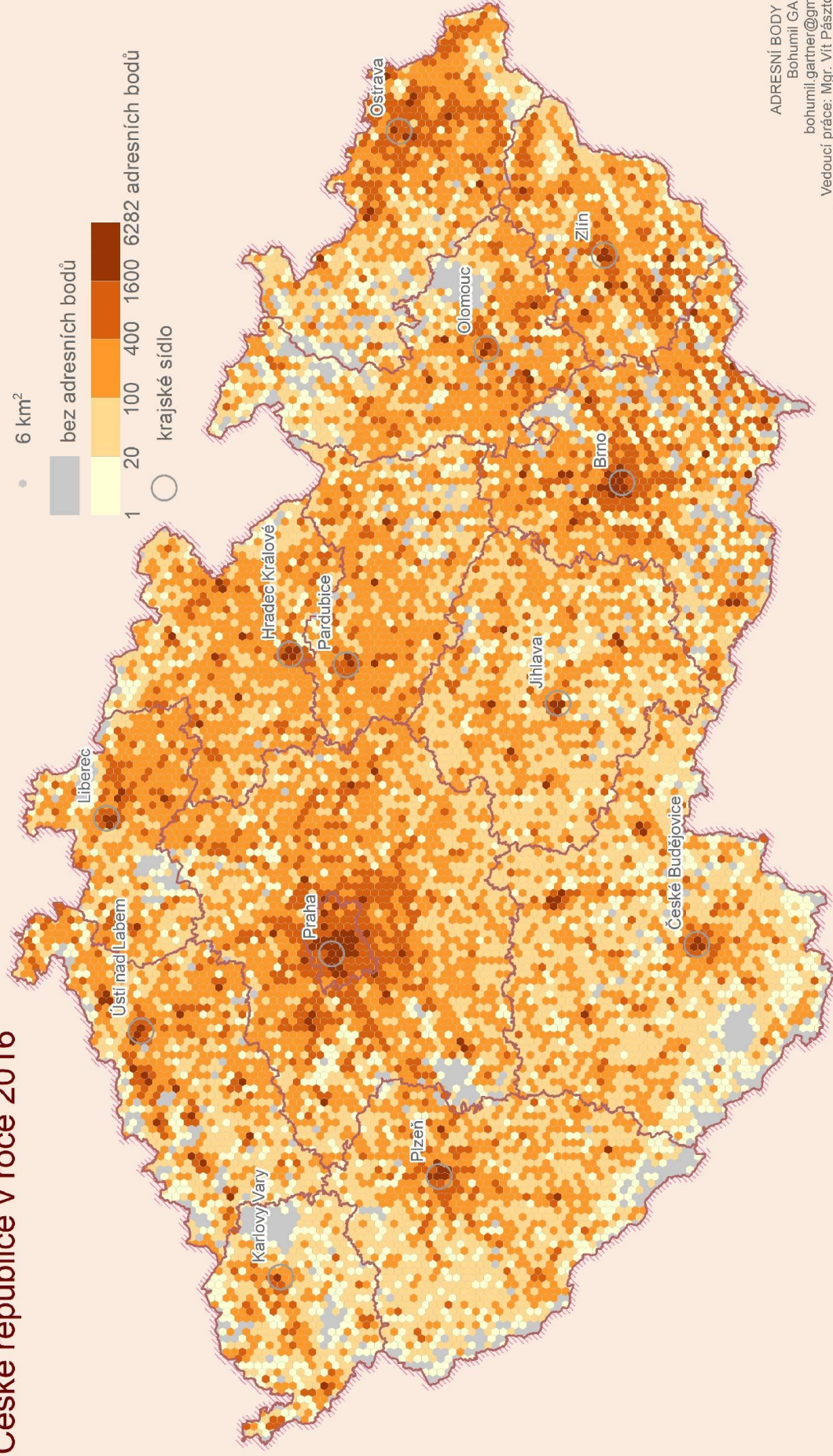
PRŮMĚRNÝ POČET NULOVÝCH HODNOT ATRIBUTŮ

adresních bodů podle krajů v České republice v roce 2016



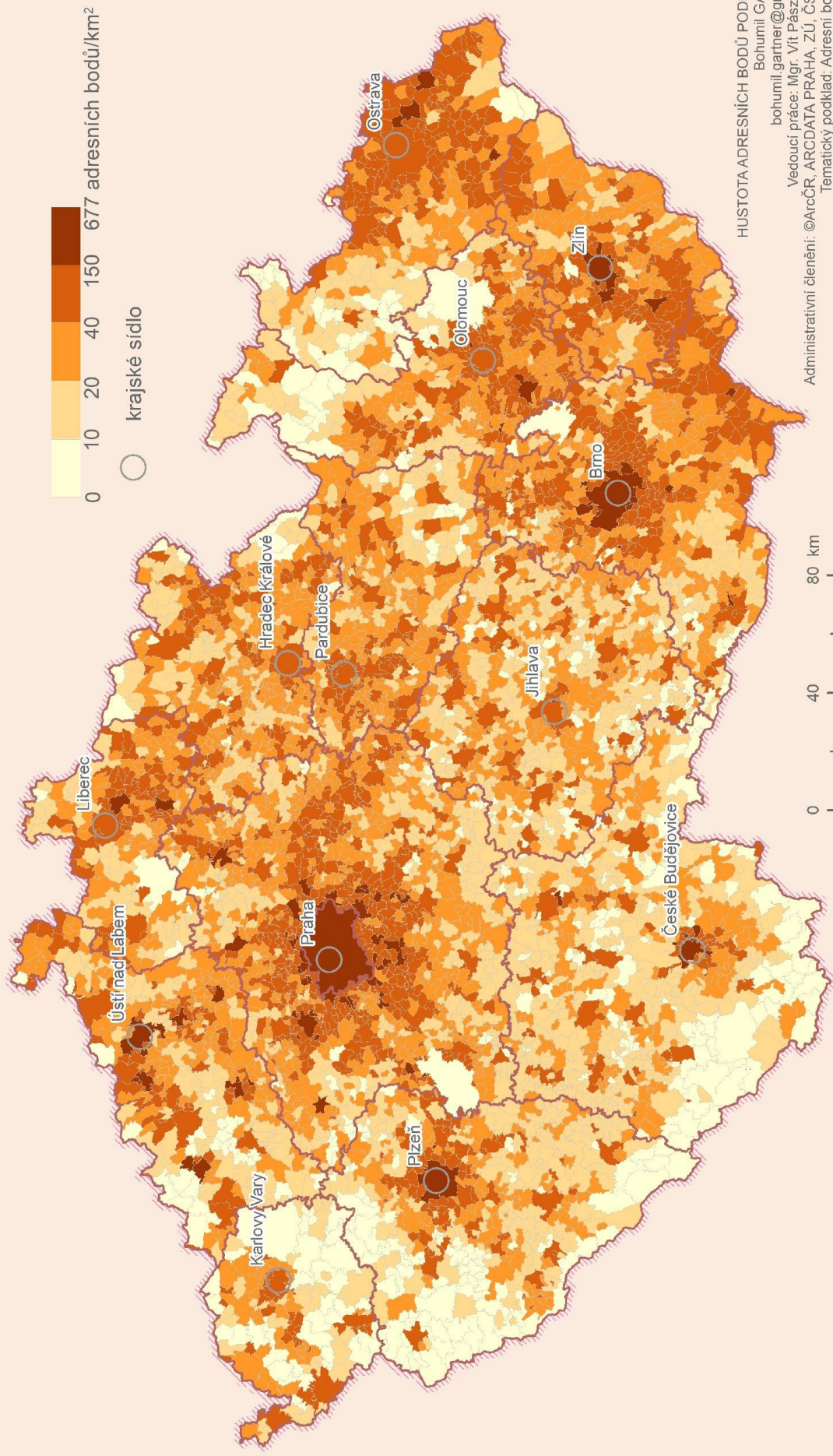
PRŮMĚRNÝ POČET NULOVÝCH HODNOT ATRIBUTŮ
PODLE KRAJŮ
Bohumil GARTNER
bohumil.gartner@gmail.com
Vedoucí práce: Mgr. Vít Pászto, Ph.D.
ARCDATA PRAHA, ZÚ, ČSÚ, 2016
Tematický podklad: Adresní body, ČSÚ
© Univerzita Palackého v Olomouci 2018
Příloha č. 3

ADRESNÍ BODY v České republice v roce 2016



HUSTOTA ADRESNÍCH BODŮ

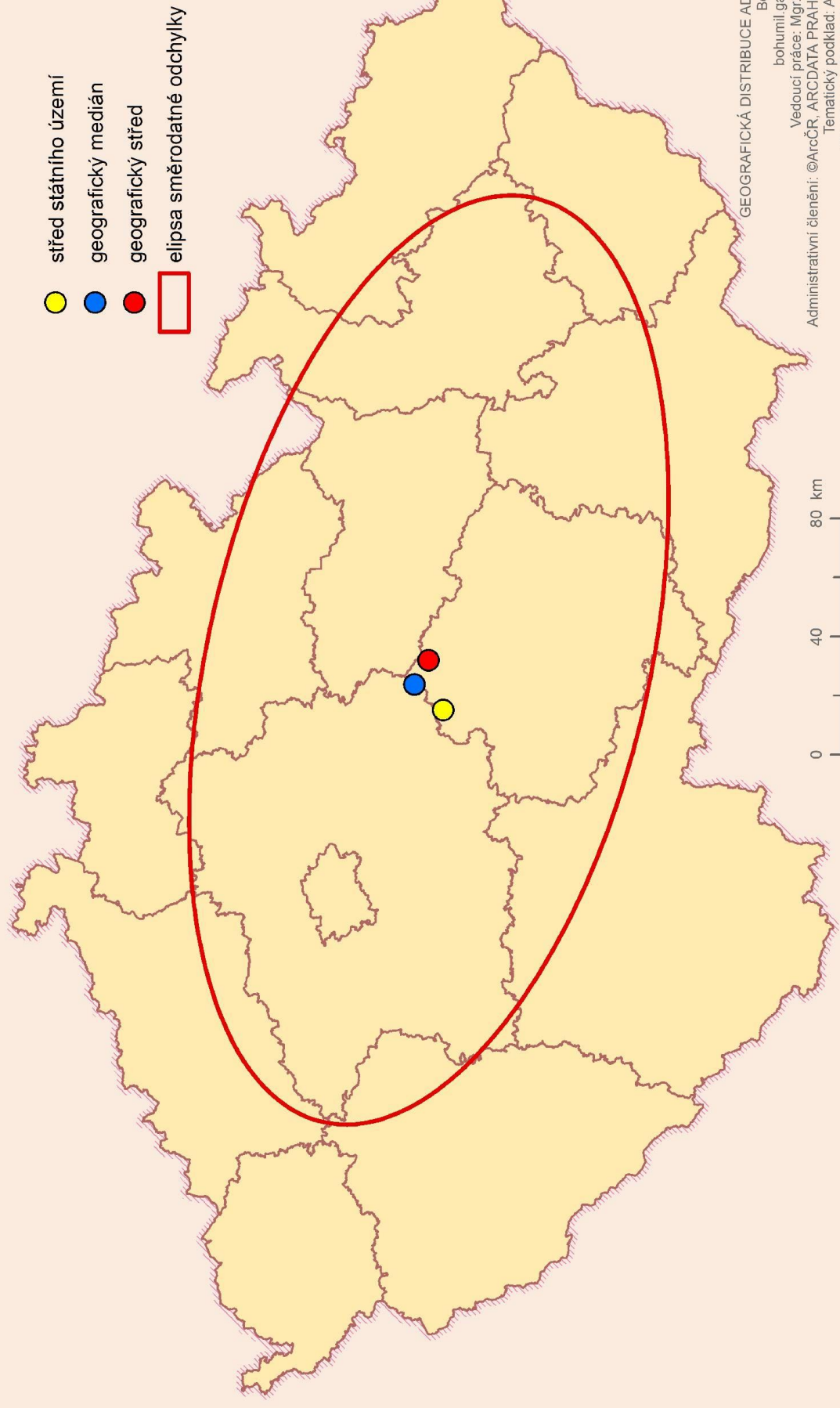
podle obcí v České republice v roce 2016



HUSTOTA ADRESNÍCH BODŮ PODLE OBČÍ
Bohumil GARTNER
bohumil.gartner@gmail.com
Vedoucí práce: Mgr. Vít Pászto, Ph.D.
Administrativní členění: ©ArcCR, ARCDATA PRAHA, ZÚ, ČSÚ, 2016
Tematický podklad: Adresní body, ČSÚ
© Univerzita Palackého v Olomouci, 2018
Příloha č.5

GEOGRAFICKÁ DISTRIBUCE ADRESNÍCH BODŮ



v České republice v roce 2016

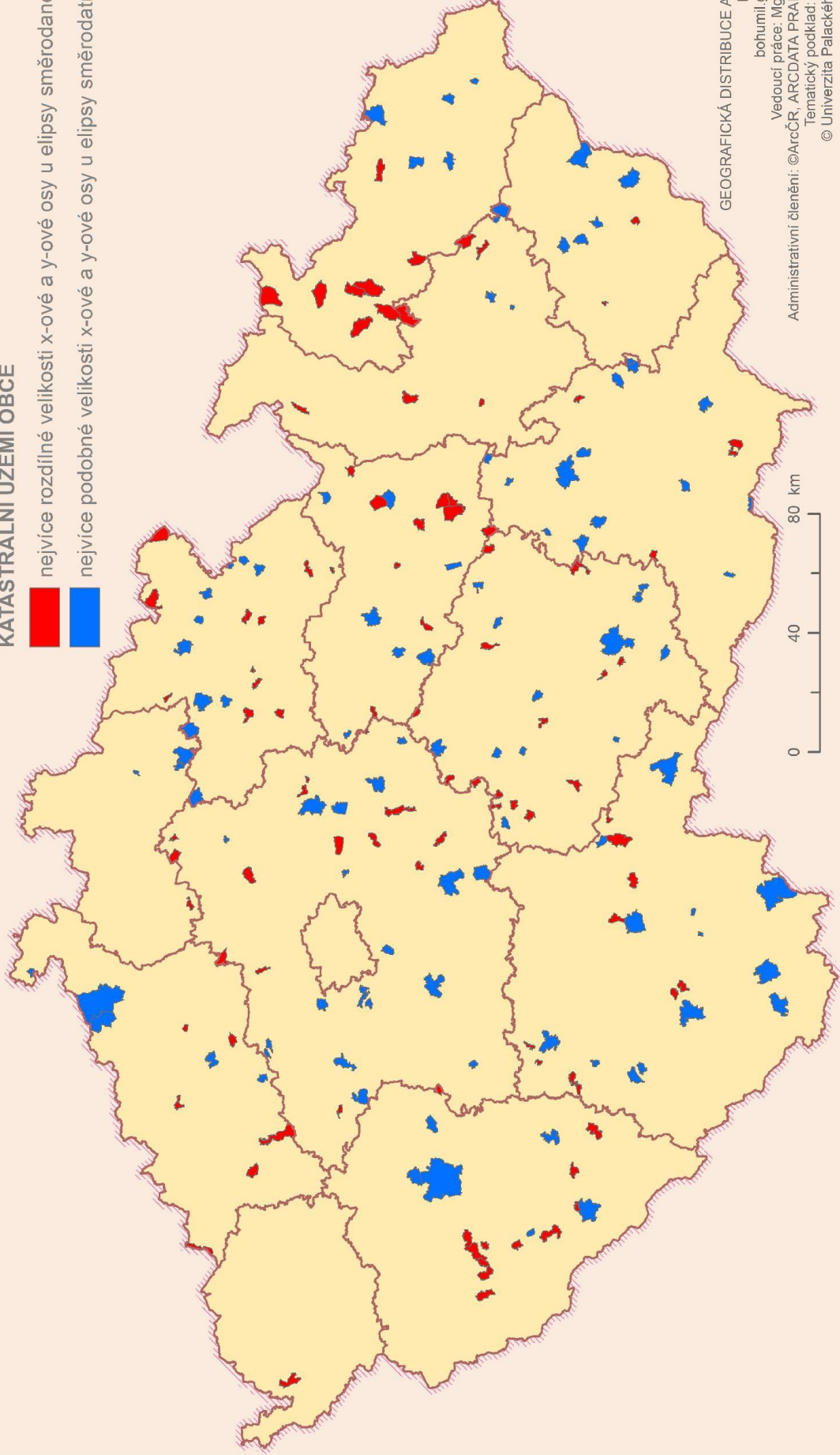


ELIPSY SMĚRODATNÝCH ODCHYLEK

pro obce v České republice v roce 2016

KATASTRÁLNÍ ÚZEMÍ OBCE

-  nejvíce rozdílné velikosti x-ové a y-ové osy u elipsy směrodatné odchylky
-  nejvíce podobné velikosti x-ové a y-ové osy u elipsy směrodatné odchylky



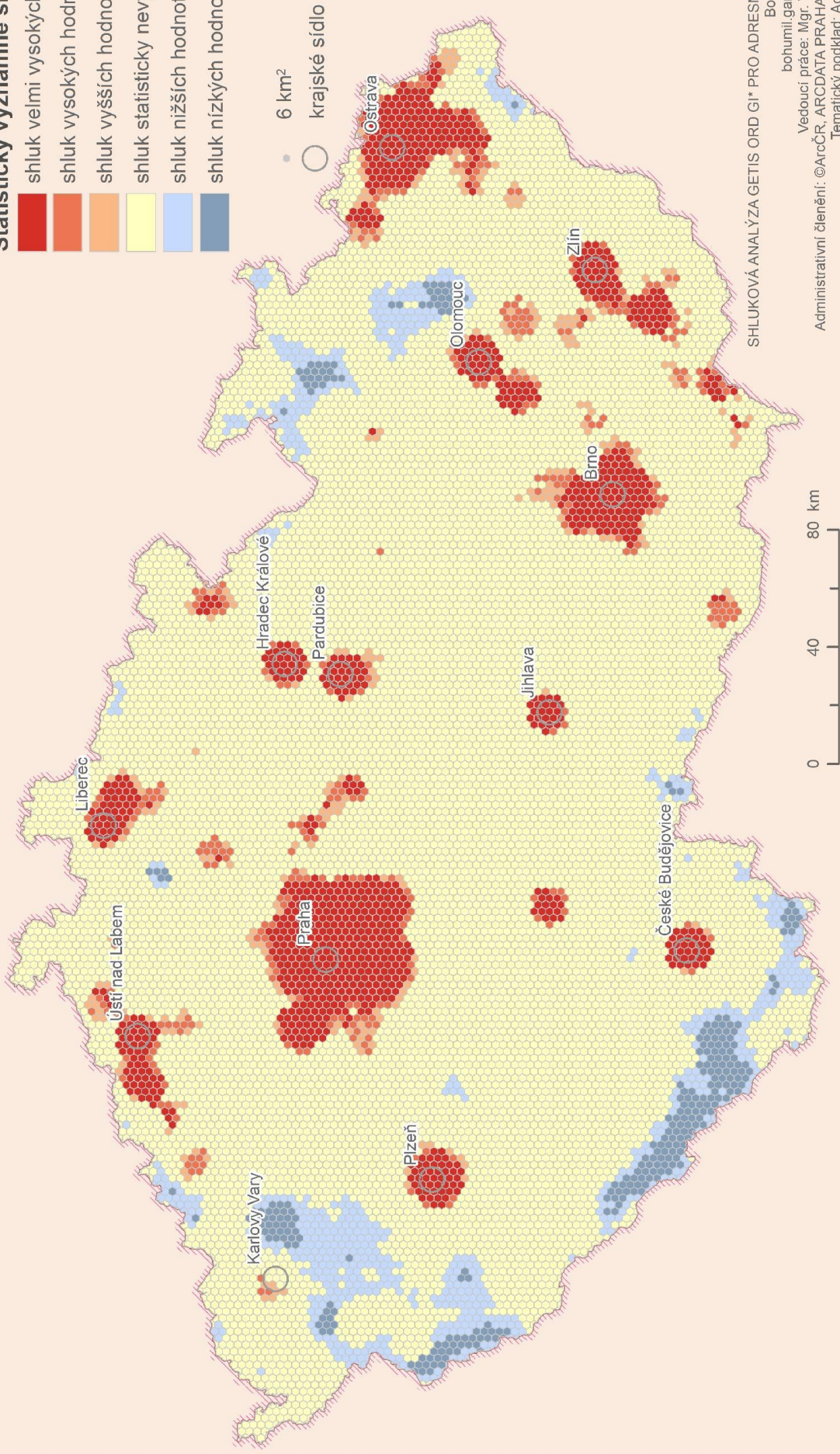
GEOGRAFICKÁ DISTRIBUCE ADRESNÍCH BODŮ
Bohumil GARTNER
bohumil.gartner@gmail.com
Vedoucí práce: Mgr. Vít Pászto, Ph.D.
Administrativní členění: ©ArcČR, ARCDATA PRAHA, ZÚ, ČSÚ, 2016
Tematický podklad: Adresní body, ČSÚ
© Univerzita Palackého v Olomouci 2018
Příloha č. 7

SHLUKOVÁ ANALÝZA GETIS ORD GI*

pro adresní body Česka v roce 2016

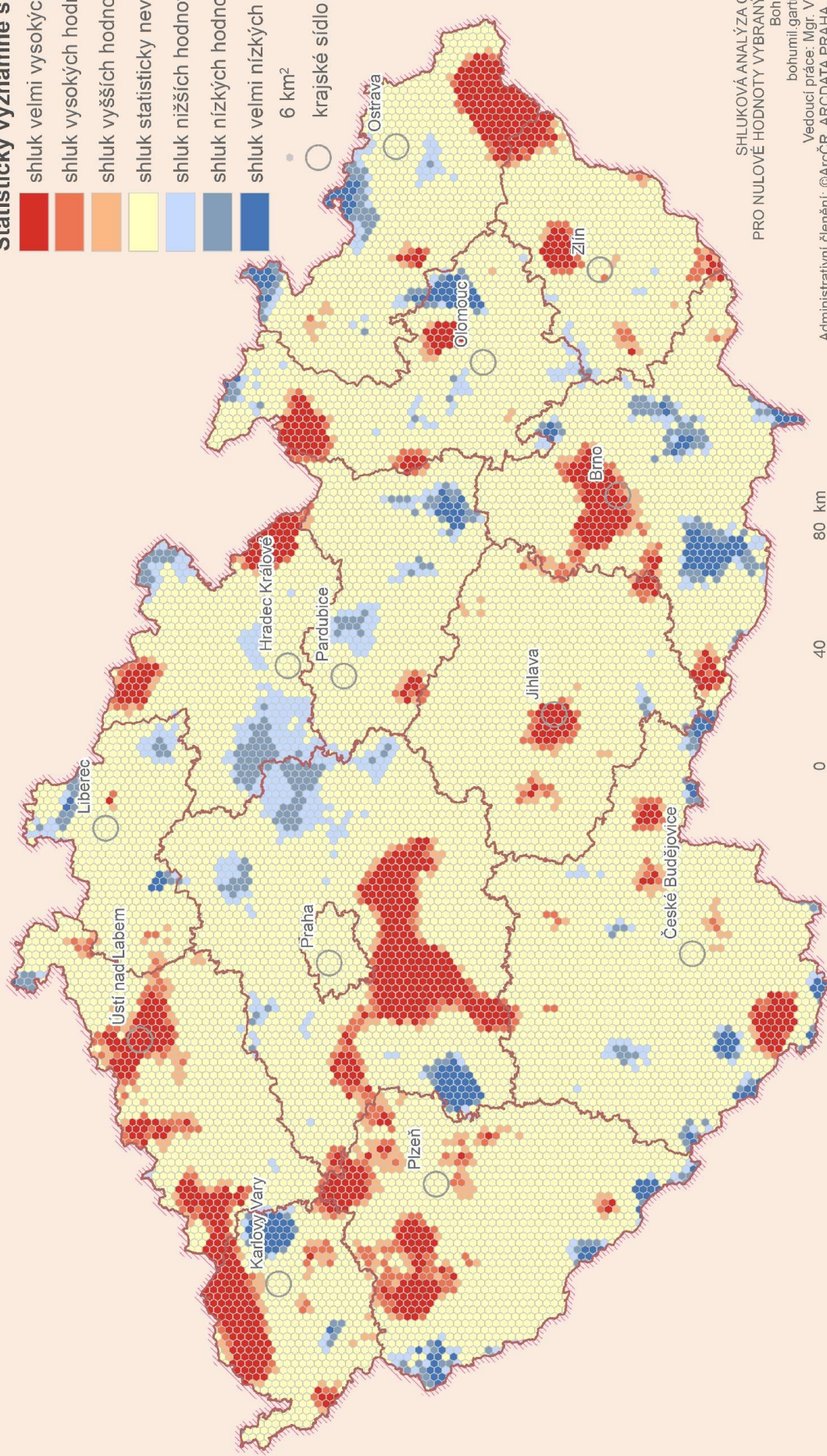
Statisticky významné shluky

- shluk velmi vysokých hodnot
- shluk vysokých hodnot
- shluk vyšších hodnot
- shluk statisticky nevýznamný
- shluk nižších hodnot
- shluk nízkých hodnot



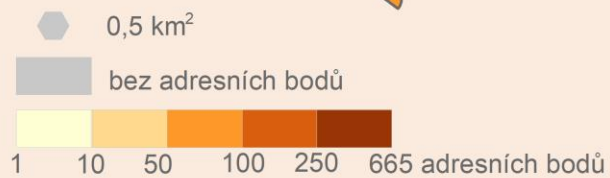
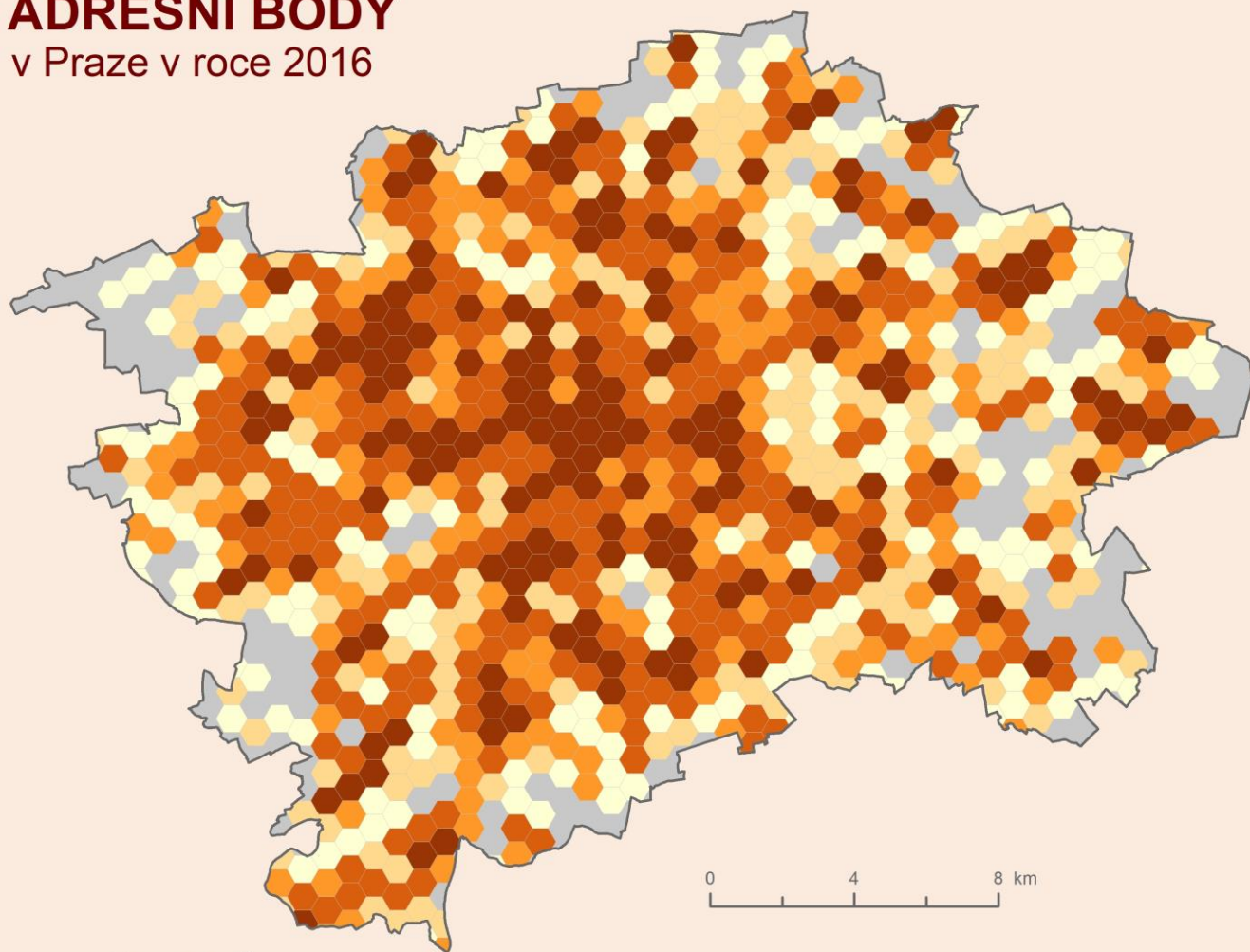
SHLUKOVÁ ANALÝZA GETIS ORD GI*

průměrného počtu nulových hodnot vybraných atributů adresních bodů Česka v roce 2016



ADRESNÍ BODY

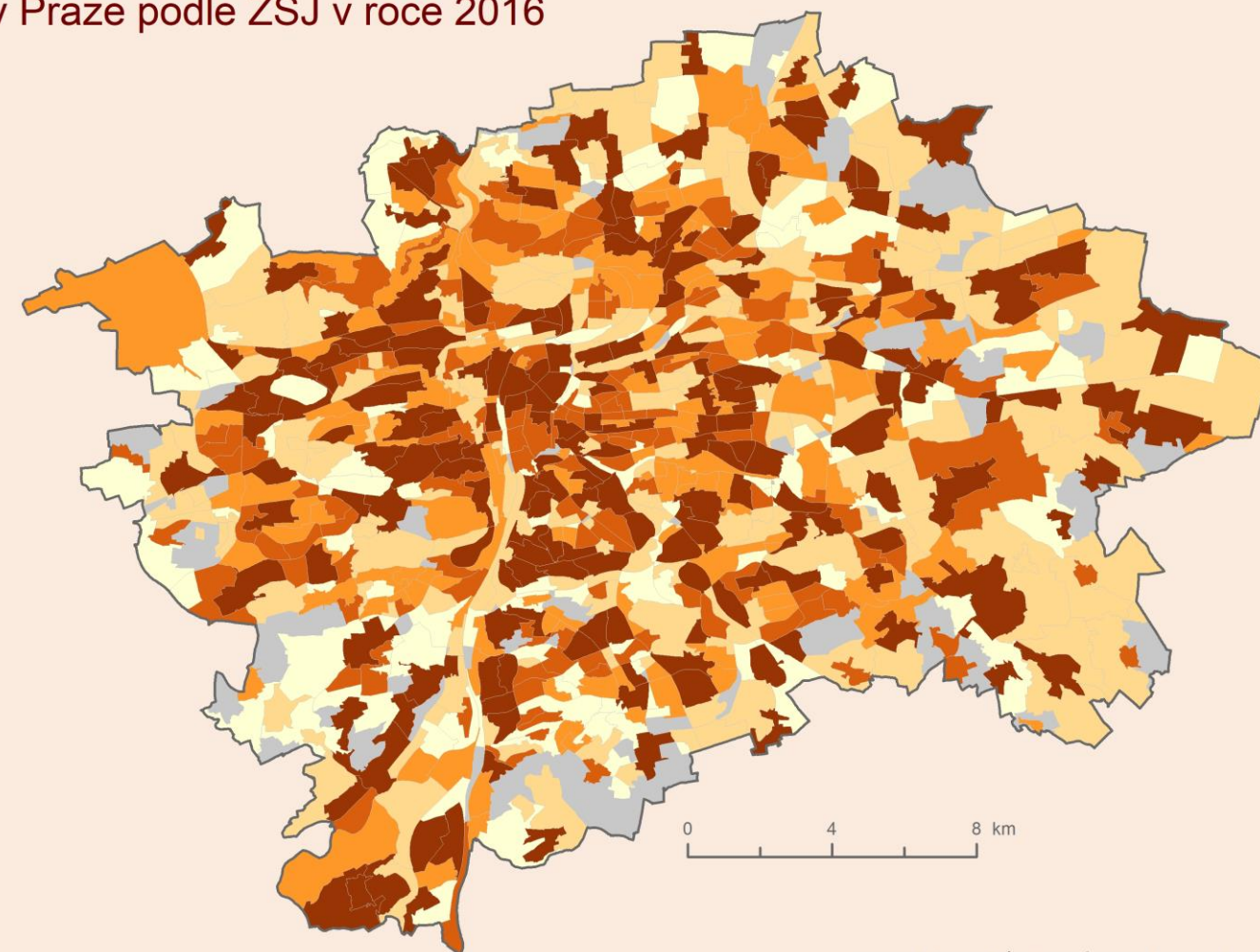
v Praze v roce 2016



ADRESNÍ BODY V PRAZE
Bohumil GARTNER
bohumil.gartner@gmail.com
Vedoucí práce: Mgr. Vít Pászto, Ph.D.
Administrativní členění: ©ArcČR, ARCDATA PRAHA, ZÚ, ČSÚ, 2016
Tematický podklad: Adresní body, ČSÚ
© Univerzita Palackého v Olomouci 2018
Příloha č.10

HUSTOTA ADRESNÍCH BODŮ

v Praze podle ZSJ v roce 2016



HUSTOTA ADRESNÍCH BODŮ V PRAZE PODLE ZSJ

Bohumil GARTNER

bohumil.gartner@gmail.com

Vedoucí práce: Mgr. Vít Pászto, Ph.D.

Administrativní členění: ©ArcČR, ARCDATA PRAHA, ZÚ, ČSÚ, 2016

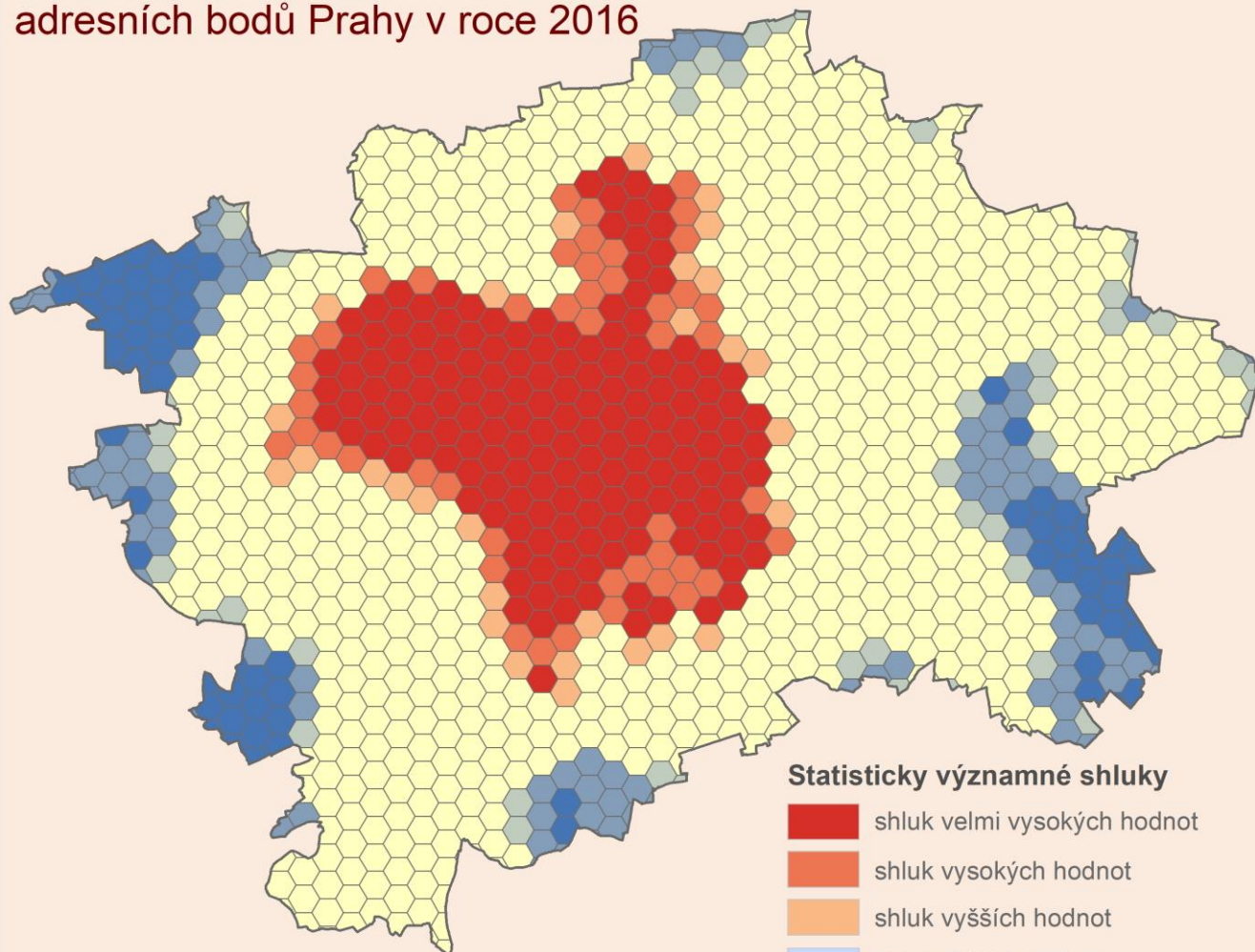
Tematický podklad: Adresní body, ČSÚ

© Univerzita Palackého v Olomouci 2018

Příloha č.11

SHLUKOVÁ ANALÝZA GETIS ORD GI*

adresních bodů Prahy v roce 2016



Statisticky významné shluky

- shluk velmi vysokých hodnot
- shluk vysokých hodnot
- shluk vyšších hodnot
- shluk nižších hodnot
- shluk nízkých hodnot
- shluk velmi nízkých hodnot

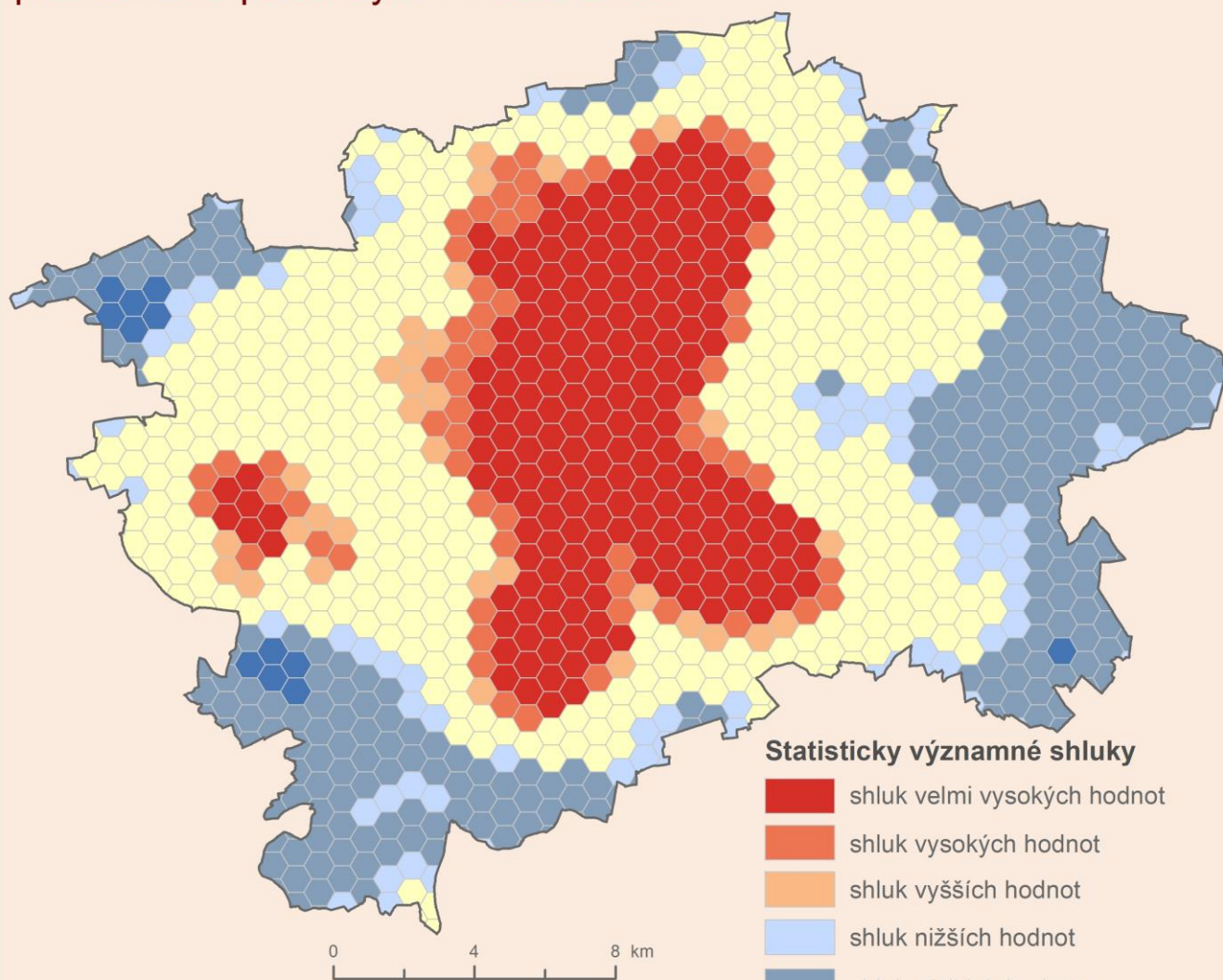
shluk statisticky nevýznamný

0,5 km²

SHLUKOVÁ ANALÝZA GETIS ORD GI* PRO ADRESNÍ BODY PRAHY
Bohumil GARTNER
bohumil.gartner@gmail.com
Vedoucí práce: Mgr. Vít Pászto, Ph.D.
Administrativní členění: ©ArcČR, ARCDATA PRAHA, ZÚ, ČSÚ, 2016
Tematický podklad: Adresní body, ČSÚ
© Univerzita Palackého v Olomouci 2018
Příloha č.12

SHLUKOVÁ ANALÝZA GETIS ORD GI*

průměrného počtu bytů v Praze 2016



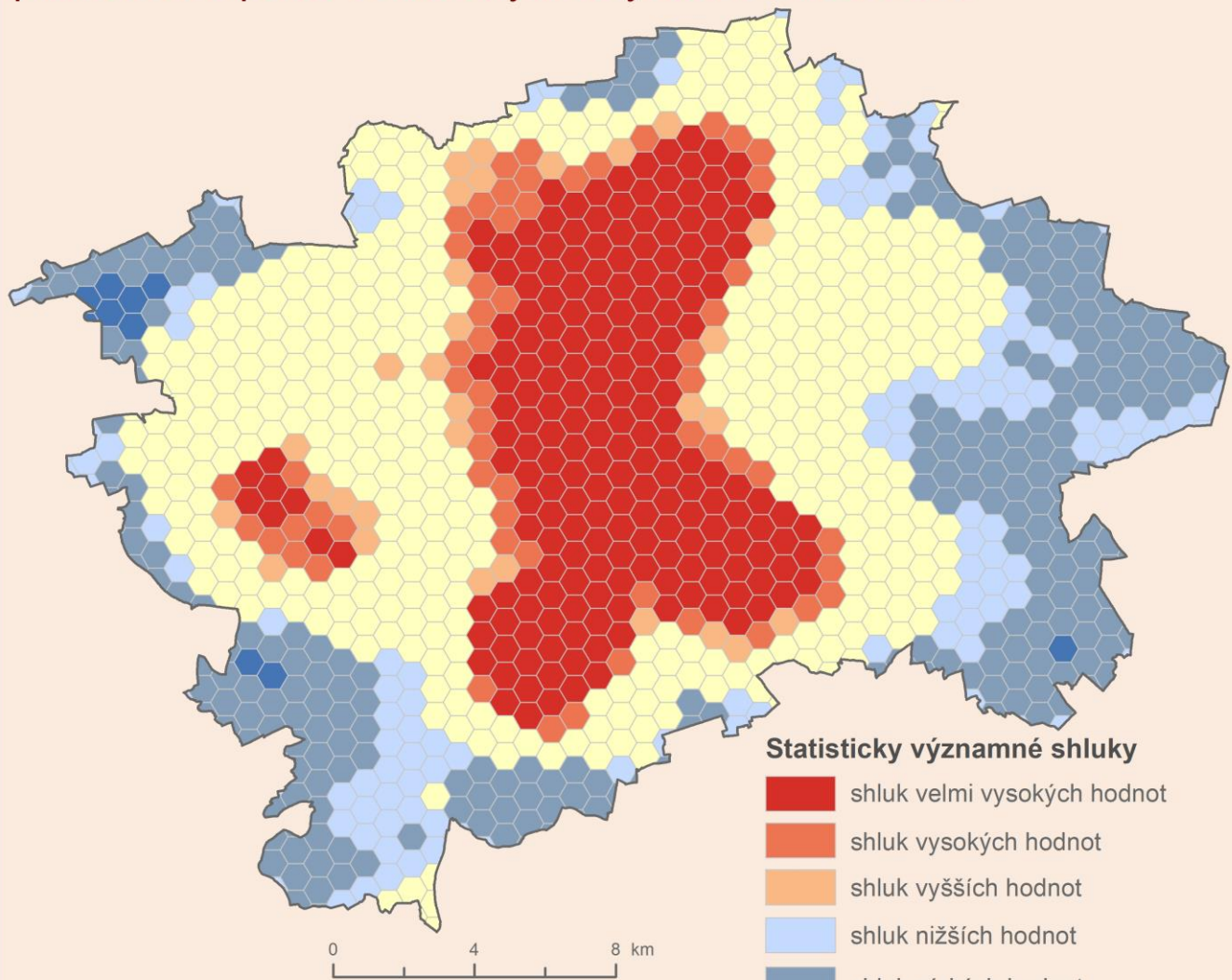
Statisticky významné shluky

- shluk velmi vysokých hodnot
- shluk vysokých hodnot
- shluk vyšších hodnot
- shluk nižších hodnot
- shluk nízkých hodnot
- shluk velmi nízkých hodnot
- shluk statisticky nevýznamný
- 0,5 km²

SHLUKOVÁ ANALÝZA GETIS ORD GI* POČTU BYTŮ V PRAZE
Bohumil GARTNER
bohumil.gartner@gmail.com
Vedoucí práce: Mgr. Vít Pászto, Ph.D.
Administrativní členění: ©ArcČR, ARCDATA PRAHA, ZÚ, ČSÚ, 2016
Tematický podklad: Adresní body, ČSÚ
© Univerzita Palackého v Olomouci 2018
Příloha č. 16

SHLUKOVÁ ANALÝZA GETIS ORD GI*

průměrného počtu evidovaných obyvatel v Praze 2016



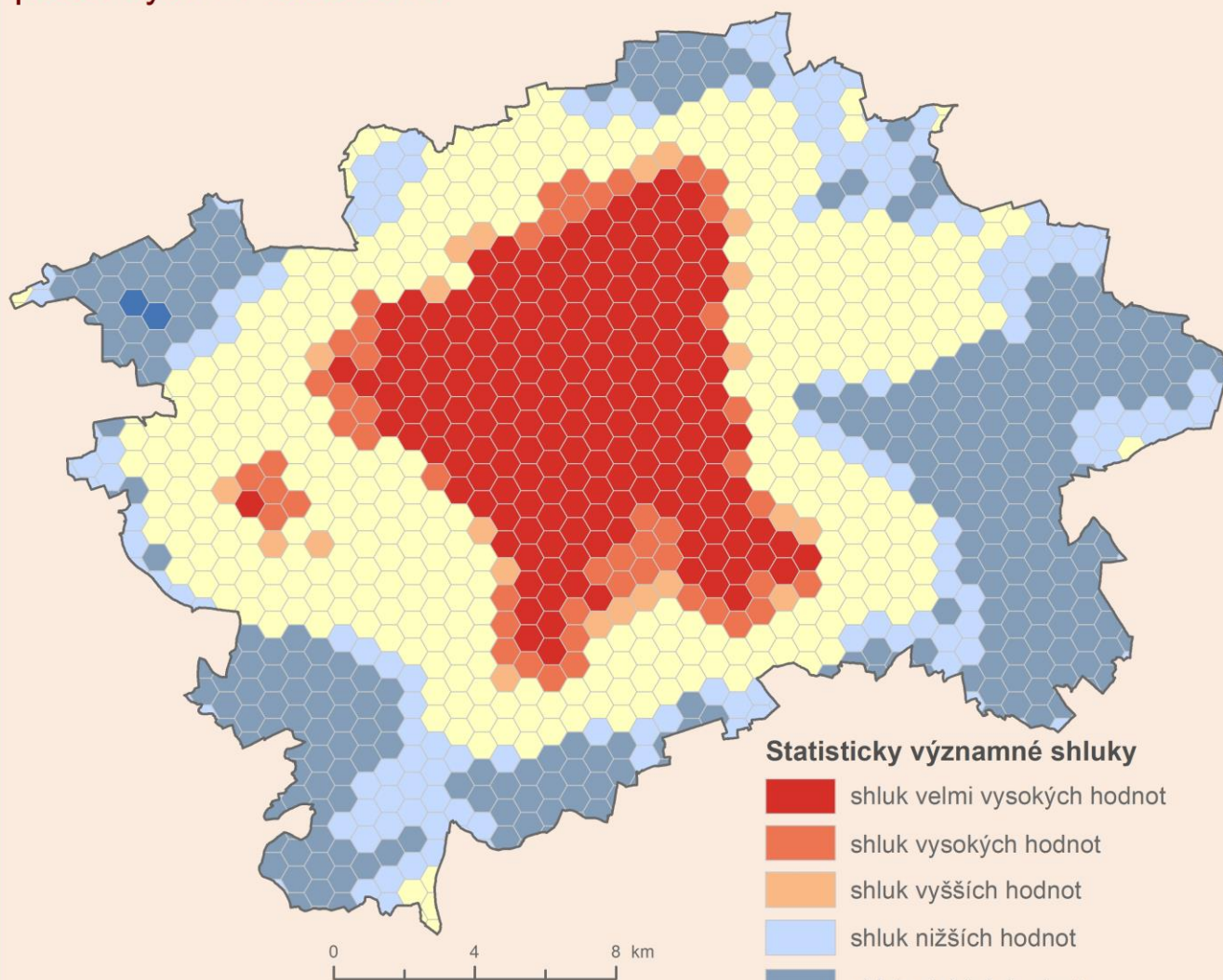
Statisticky významné shluky

- shluk velmi vysokých hodnot
- shluk vysokých hodnot
- shluk vyšších hodnot
- shluk nižších hodnot
- shluk nízkých hodnot
- shluk velmi nízkých hodnot
- shluk statisticky nevýznamný
- 0,5 km²

SHLUKOVÁ ANALÝZA GETIS ORD GI* POČTU EVIDOVANÝCH OBYVATEL PRAHY
Bohumil GARTNER
bohumil.gartner@gmail.com
Vedoucí práce: Mgr. Vít Pászto, Ph.D.
Administrativní členění: ©ArcČR, ARCDATA PRAHA, ZÚ, ČSÚ, 2016
Tematický podklad: Adresní body, ČSÚ
© Univerzita Palackého v Olomouci 2018
Příloha č. 14

SHLUKOVÁ ANALÝZA GETIS ORD GI*

počtu bytů v Praze 2016



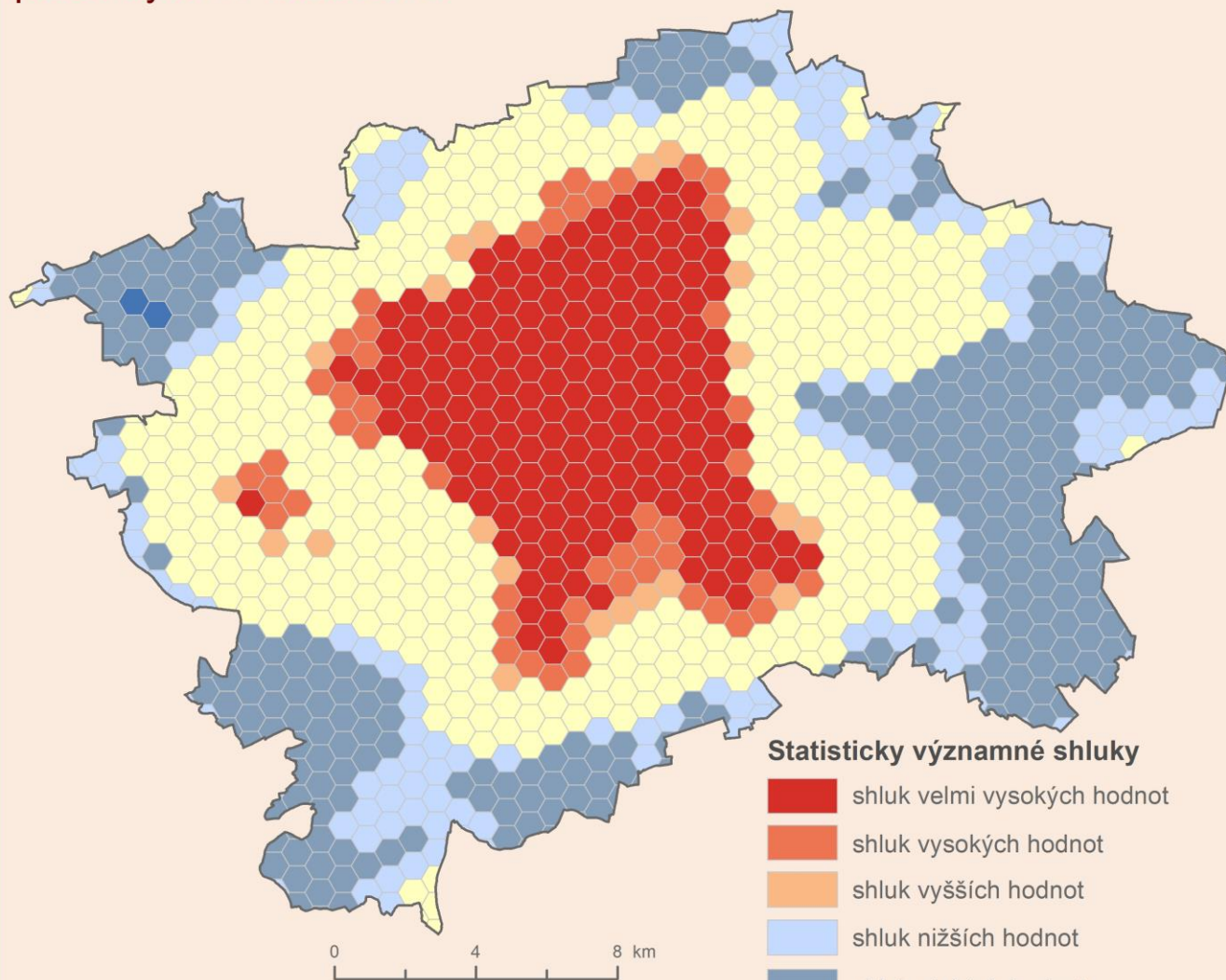
Statisticky významné shluky

- shluk velmi vysokých hodnot
- shluk vysokých hodnot
- shluk vyšších hodnot
- shluk nižších hodnot
- shluk nízkých hodnot
- shluk velmi nízkých hodnot
- shluk statisticky nevýznamný
- 0,5 km²

SHLUKOVÁ ANALÝZA GETIS ORD GI* POČTU BYTŮ V PRAZE
Bohumil GARTNER
bohumil.gartner@gmail.com
Vedoucí práce: Mgr. Vít Pászto, Ph.D.
Administrativní členění: ©ArcČR, ARCDATA PRAHA, ZÚ, ČSÚ, 2016
Tematický podklad: Adresní body, ČSÚ
© Univerzita Palackého v Olomouci 2018
Příloha č. 15

SHLUKOVÁ ANALÝZA GETIS ORD GI*

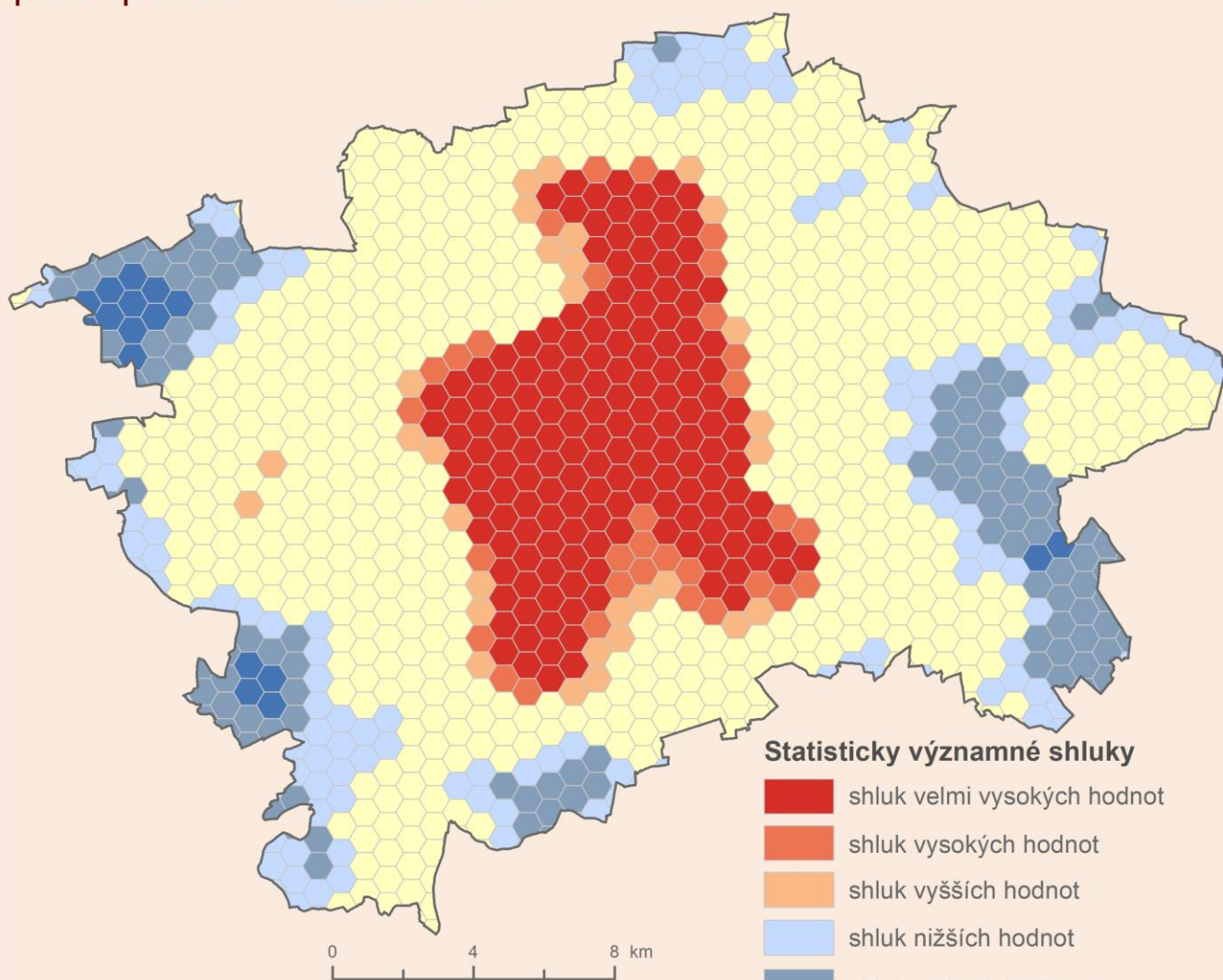
počtu bytů v Praze 2016



SHLUKOVÁ ANALÝZA GETIS ORD GI* POČTU BYTŮ V PRAZE
Bohumil GARTNER
bohumil.gartner@gmail.com
Vedoucí práce: Mgr. Vít Pászto, Ph.D.
Administrativní členění: ©ArcČR, ARCDATA PRAHA, ZÚ, ČSÚ, 2016
Tematický podklad: Adresní body, ČSÚ
© Univerzita Palackého v Olomouci 2018
Příloha č. 15

SHLUKOVÁ ANALÝZA GETIS ORD GI*

počtu podlaží v Praze 2016



SHLUKOVÁ ANALÝZA GETIS ORD GI* POČTU PODLAŽÍ V PRAZE

Bohumil GARTNER

bohumil.gartner@gmail.com

Vedoucí práce: Mgr. Vít Pászto, Ph.D.

Administrativní členění: ©ArcČR, ARCDATA PRAHA, ZÚ, ČSÚ, 2016

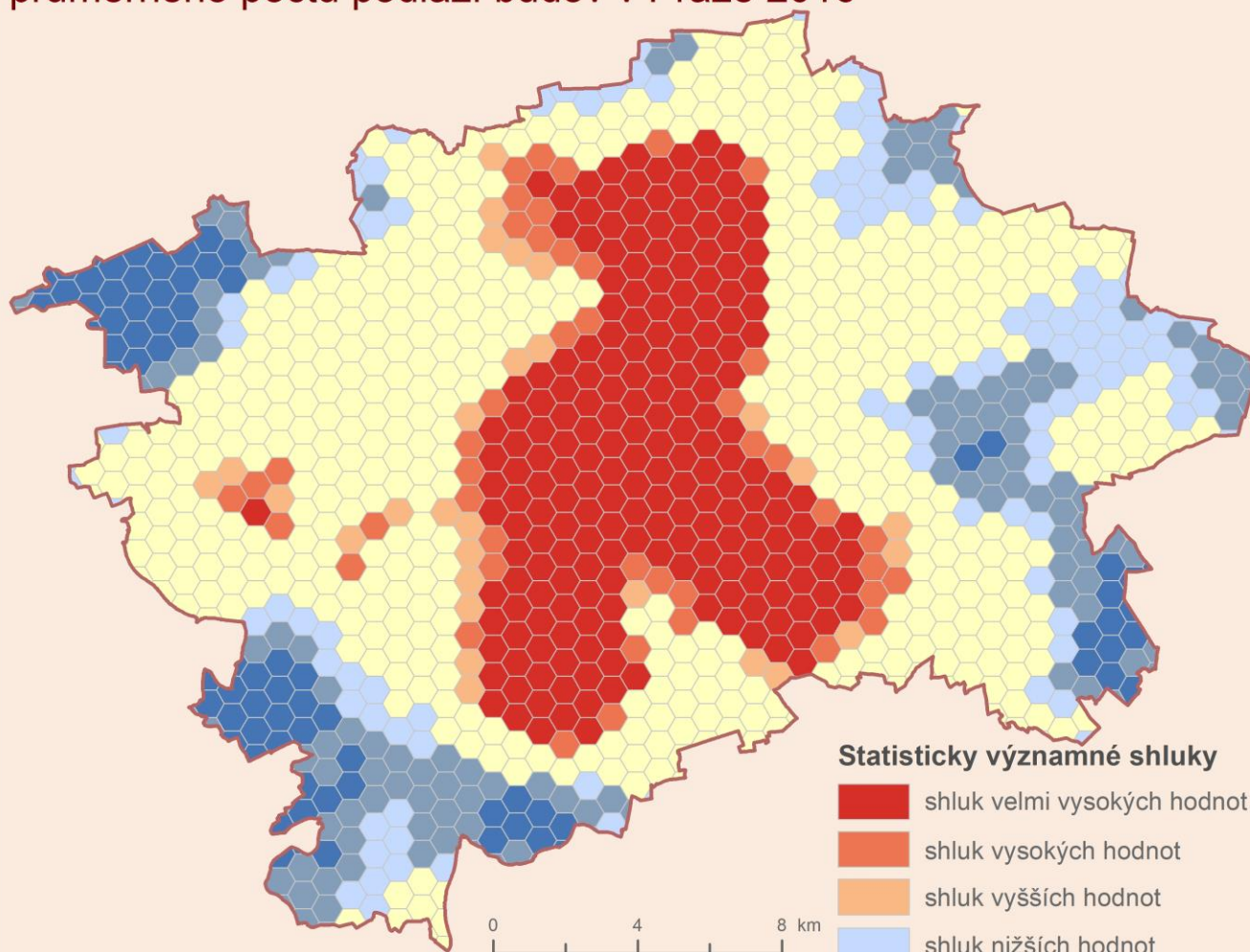
Tematický podklad: Adresní body, ČSÚ

© Univerzita Palackého v Olomouci 2018

Příloha č.17

SHLUKOVÁ ANALÝZA GETIS ORD GI*

průměrného počtu podlaží budov v Praze 2016



Statisticky významné shluky

- shluk velmi vysokých hodnot
- shluk vysokých hodnot
- shluk vyšších hodnot
- shluk nižších hodnot
- shluk nízkých hodnot
- shluk velmi nízkých hodnot
- shluk statisticky nevýznamný
- 0,5 km²

SHLUKOVÁ ANALÝZA GETIS ORD GI* POČTU PODLAŽÍ V PRAZE
Bohumil GARTNER
bohumil.gartner@gmail.com
Vedoucí práce: Mgr. Vít Pászto, Ph.D.
Administrativní členění: ©ArcČR, ARCDATA PRAHA, ZÚ, ČSÚ, 2016
Tematický podklad: Adresní body, ČSÚ
© Univerzita Palackého v Olomouci 2018
Příloha č. 18

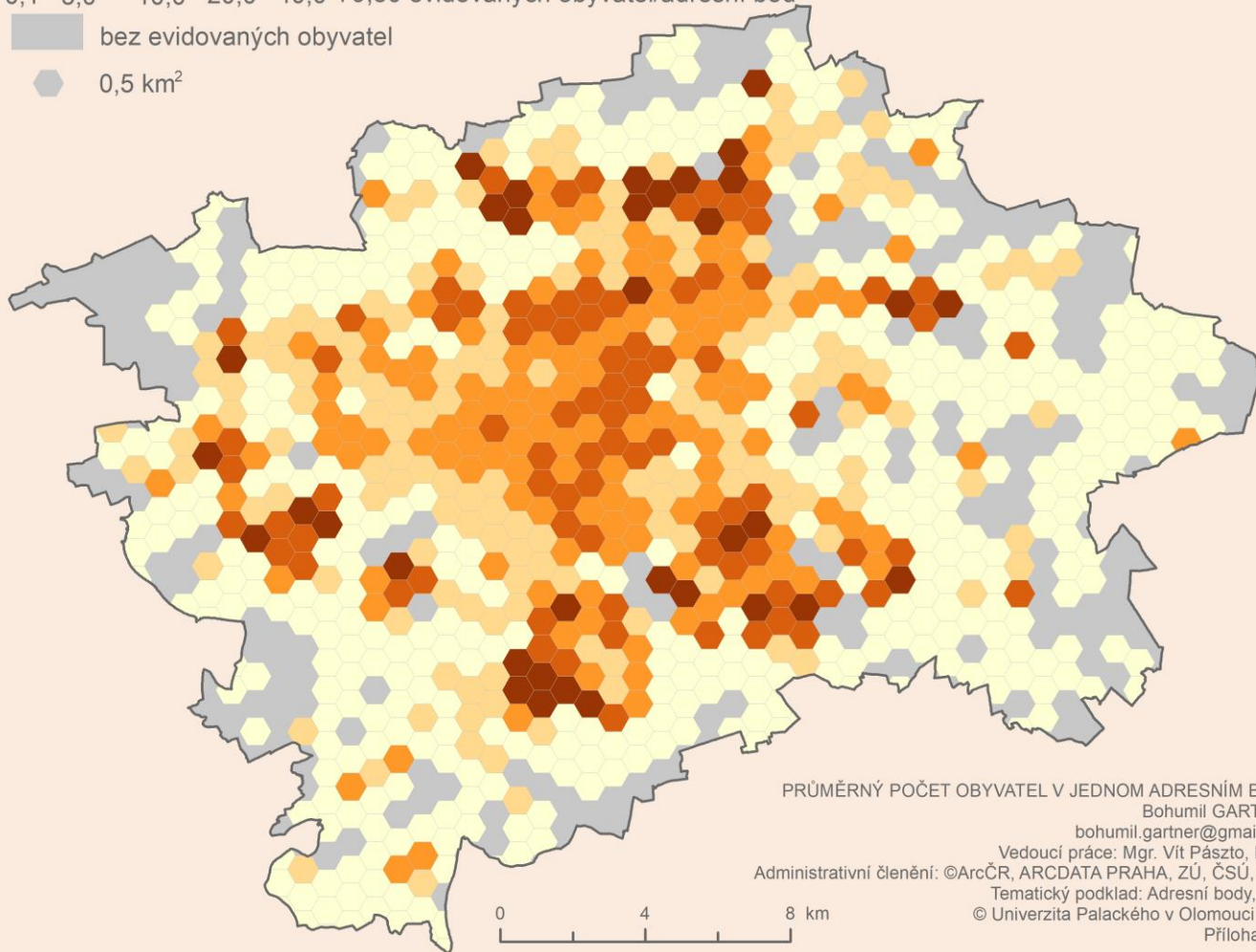
PRŮMĚRNÝ POČET EVIDOVANÝCH OBYVATEL u adresních bodů v Praze 2016



0,1 5,0 10,0 20,0 40,0 79,80 evidovaných obyvatel/adresní bod

bez evidovaných obyvatel

0,5 km²



PRŮMĚRNÝ POČET OBYVATEL V JEDNOM ADRESNÍM BODĚ

Bohumil GARTNER

bohumil.gartner@gmail.com

Vedoucí práce: Mgr. Vít Pászto, Ph.D.

Administrativní členění: ©ArcČR, ARCDATA PRAHA, ZÚ, ČSÚ, 2016

Tematický podklad: Adresní body, ČSÚ

© Univerzita Palackého v Olomouci 2018

Příloha č.19

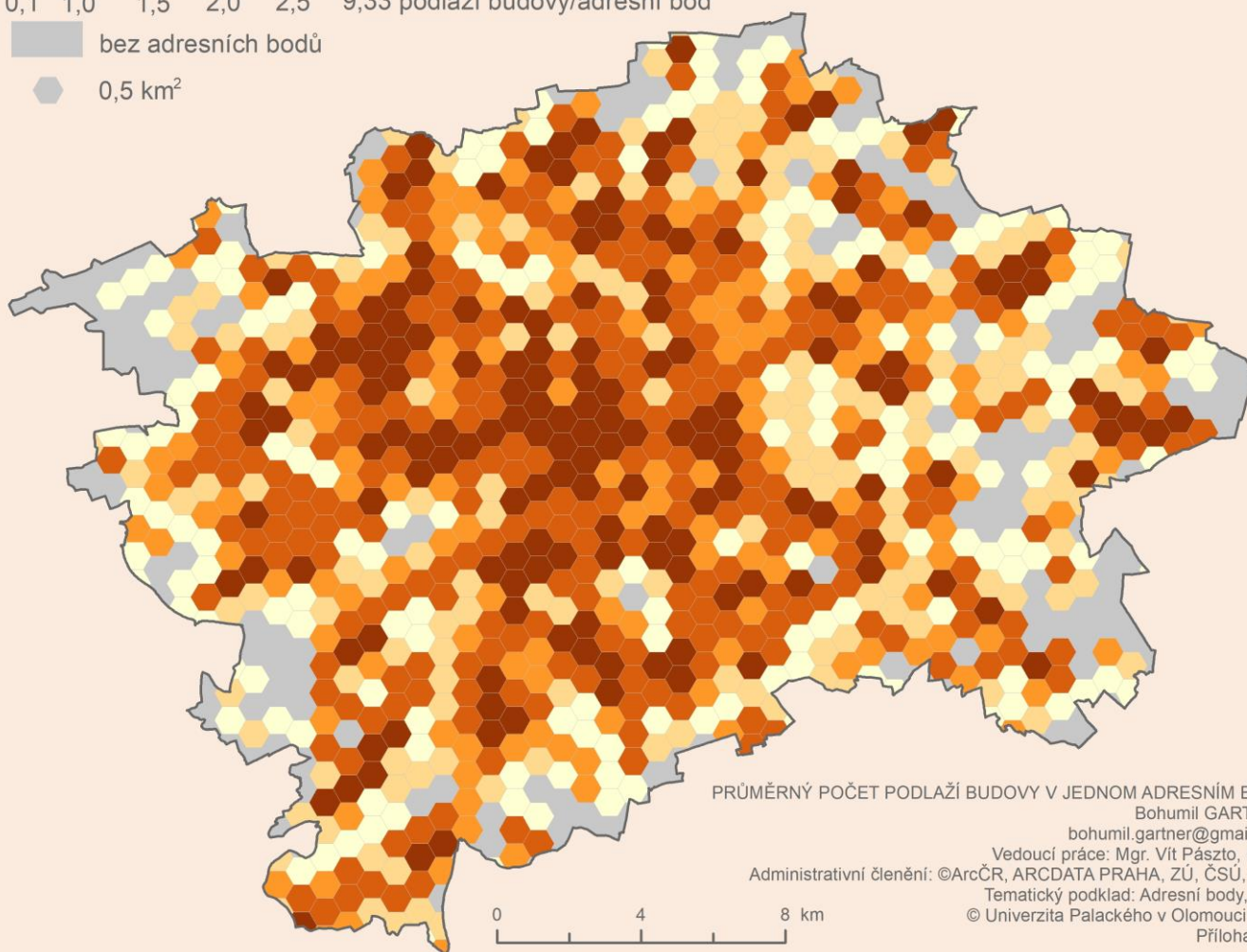
PRŮMĚRNÝ POČET PODLAŽÍ BUDOVY

pro adresní body v Praze 2016

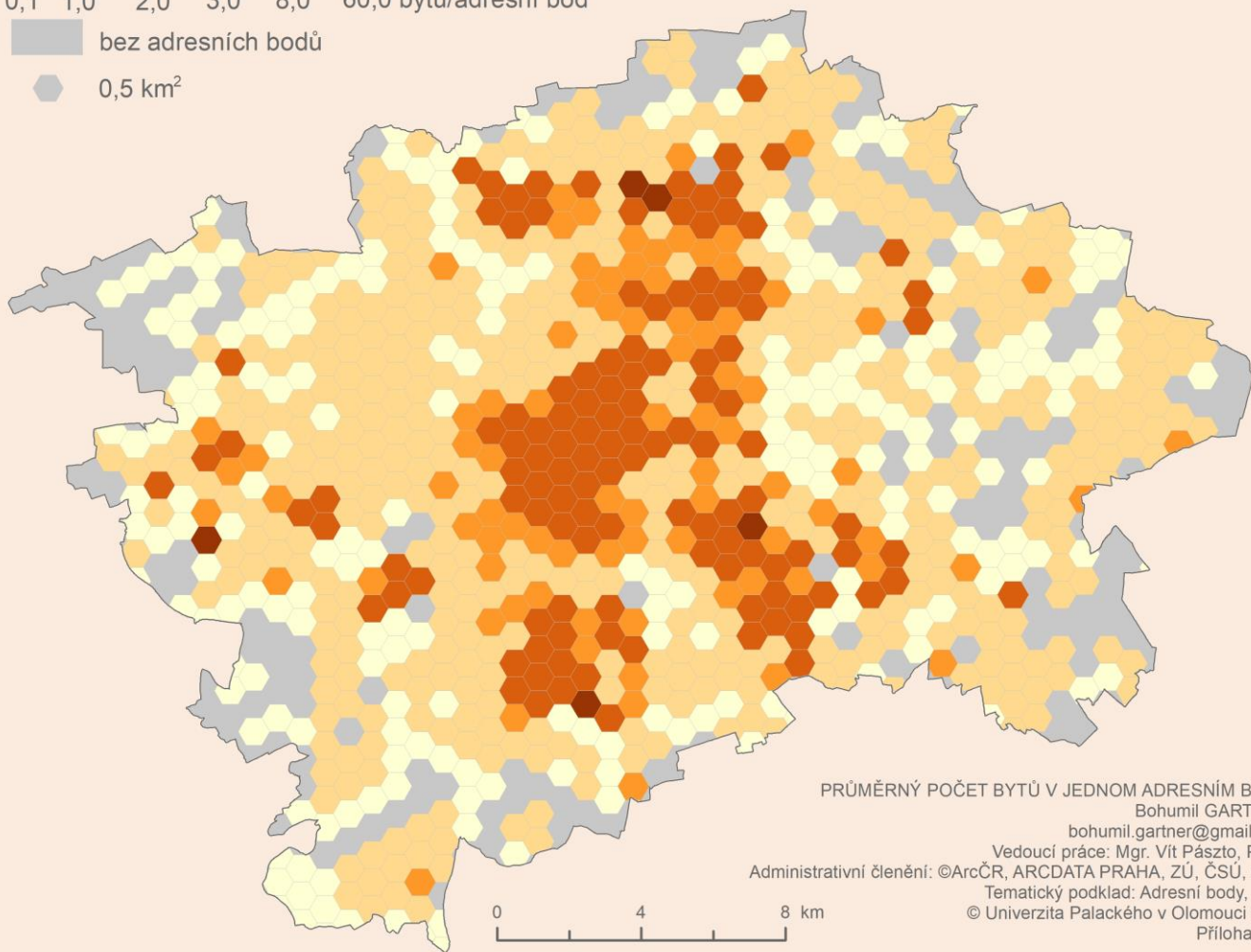
0,1 1,0 1,5 2,0 2,5 9,33 podlaží budovy/adresní bod

bez adresních bodů

0,5 km²



PRŮMĚRNÝ POČET BYTŮ pro adresní body v Praze 2016



PRŮMĚRNÝ POČET BYTŮ V JEDNOM ADRESNÍM BODĚ
Bohumil GARTNER
bohumil.gartner@gmail.com
Vedoucí práce: Mgr. Vít Pászto, Ph.D.
Administrativní členění: ©ArcČR, ARCDATA PRAHA, ZÚ, ČSÚ, 2016
Tematický podklad: Adresní body, ČSÚ
© Univerzita Palackého v Olomouci 2018
Příloha č.21