

MUNI
SCI

GIS4SG

Kartogramy – normalizace, klasifikace

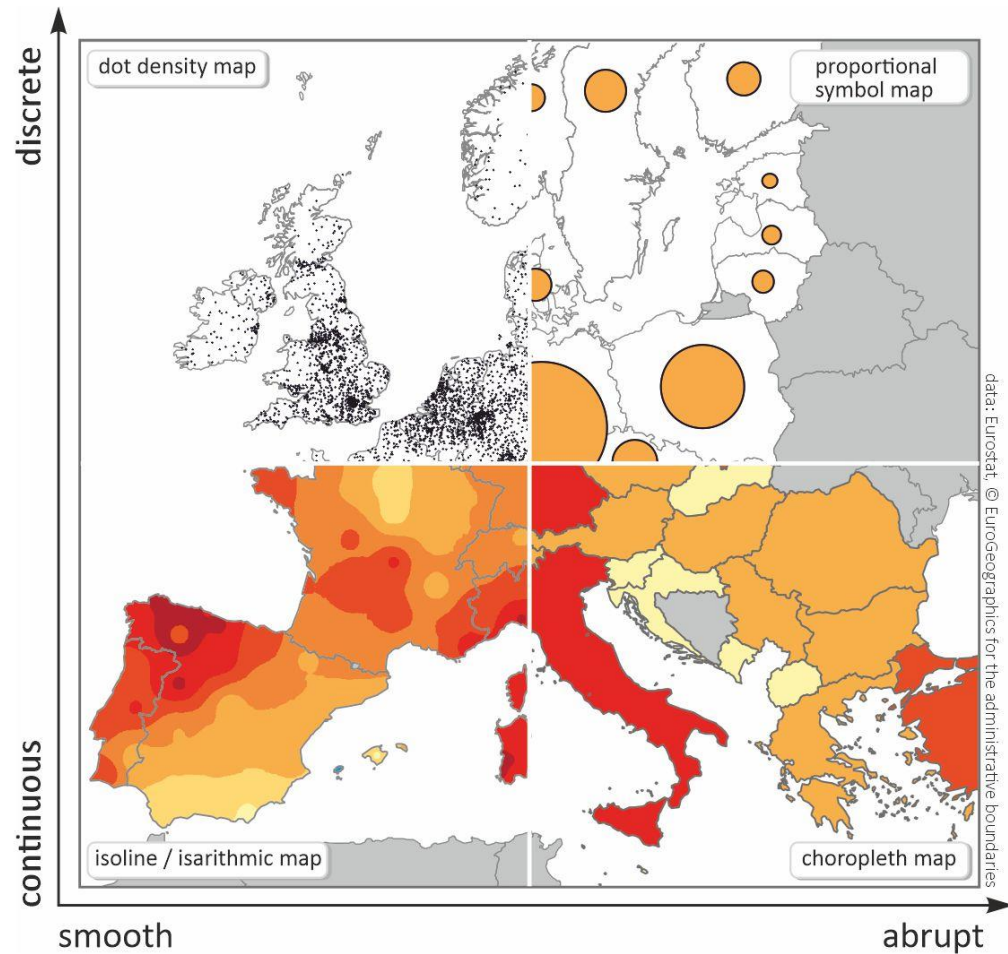
podzim 2024

Lukáš Herman

herman.lu@mail.muni.cz

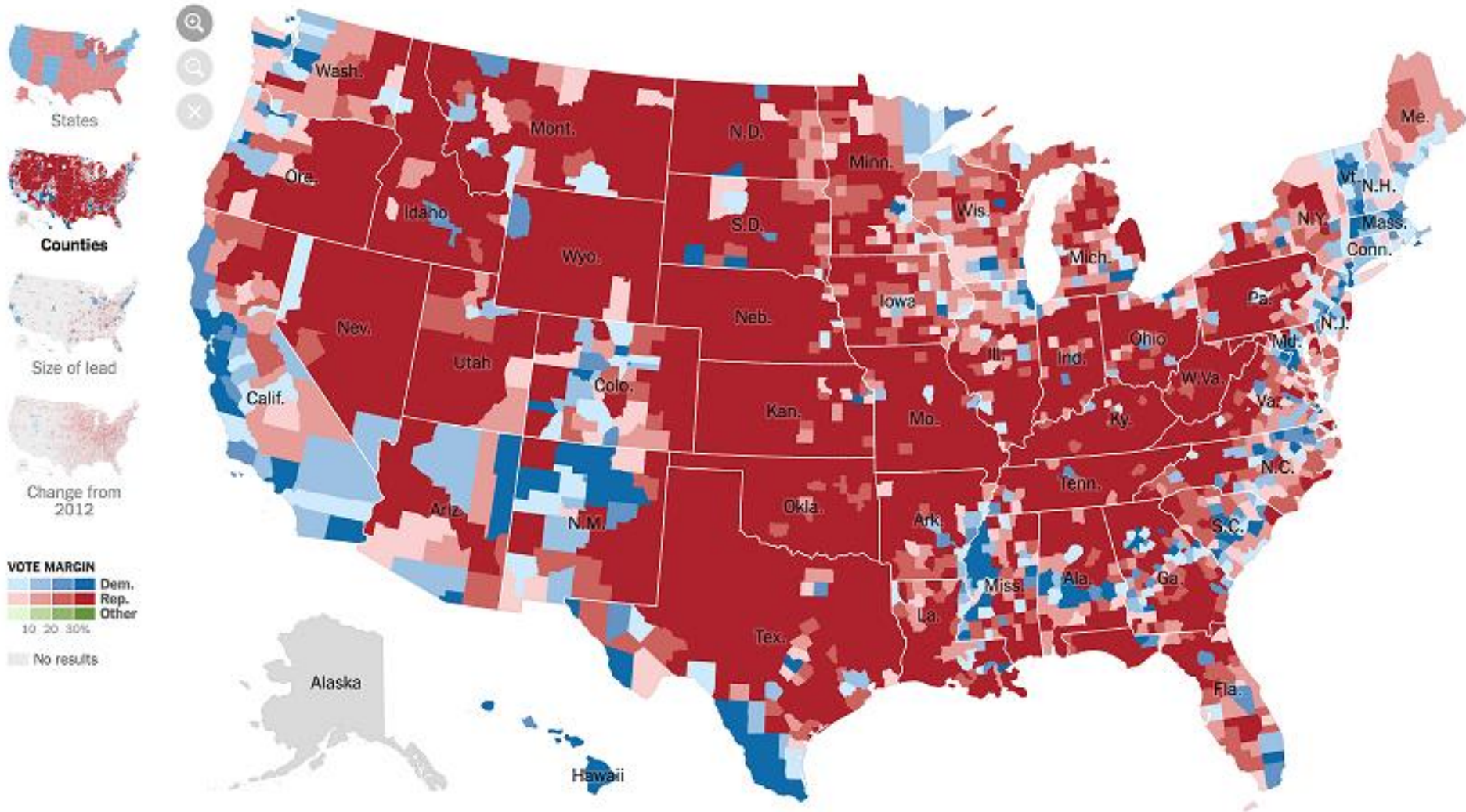
Volba metody

- Čím se řídit?
 - Charakter zobrazovaných dat
 - Účel mapy
 - Uživatel
 - ...

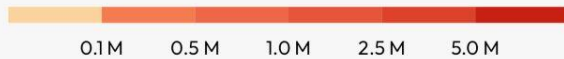


Kartogram = choropletová mapa

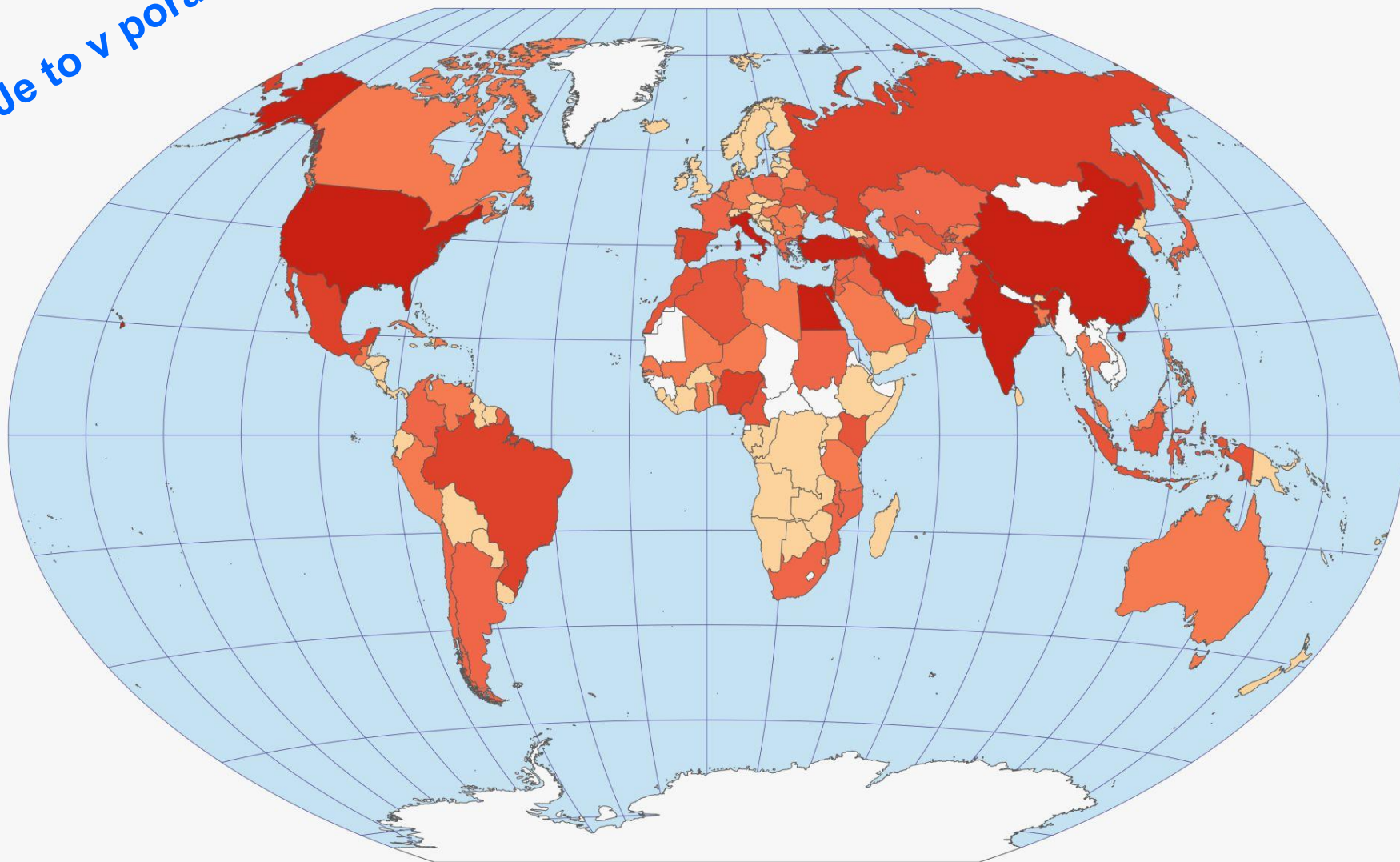
z ruštiny a z angličtiny



Tomato production in tonnes, 2020



Je to v pořádku?

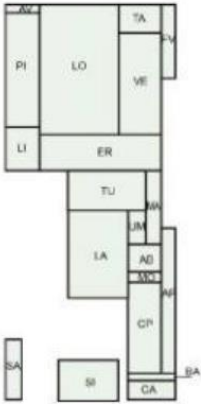

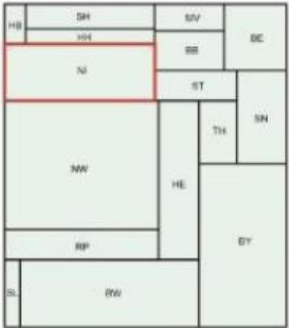
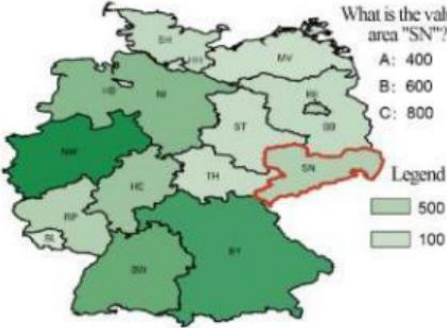


Volba metody

ing the limits of human working memory, the number of administrative divisions in the experimental areas should be controlled within a certain range, so as to contrast with the choropleth map. This paper selects Germany and Italy, which are relatively unfamiliar to the subjects, as the experimental areas. The numbers of administrative divisions in the two countries are between 15 and 20. The population and GDP of Germany and Italy in 2016 are used as the thematic data.

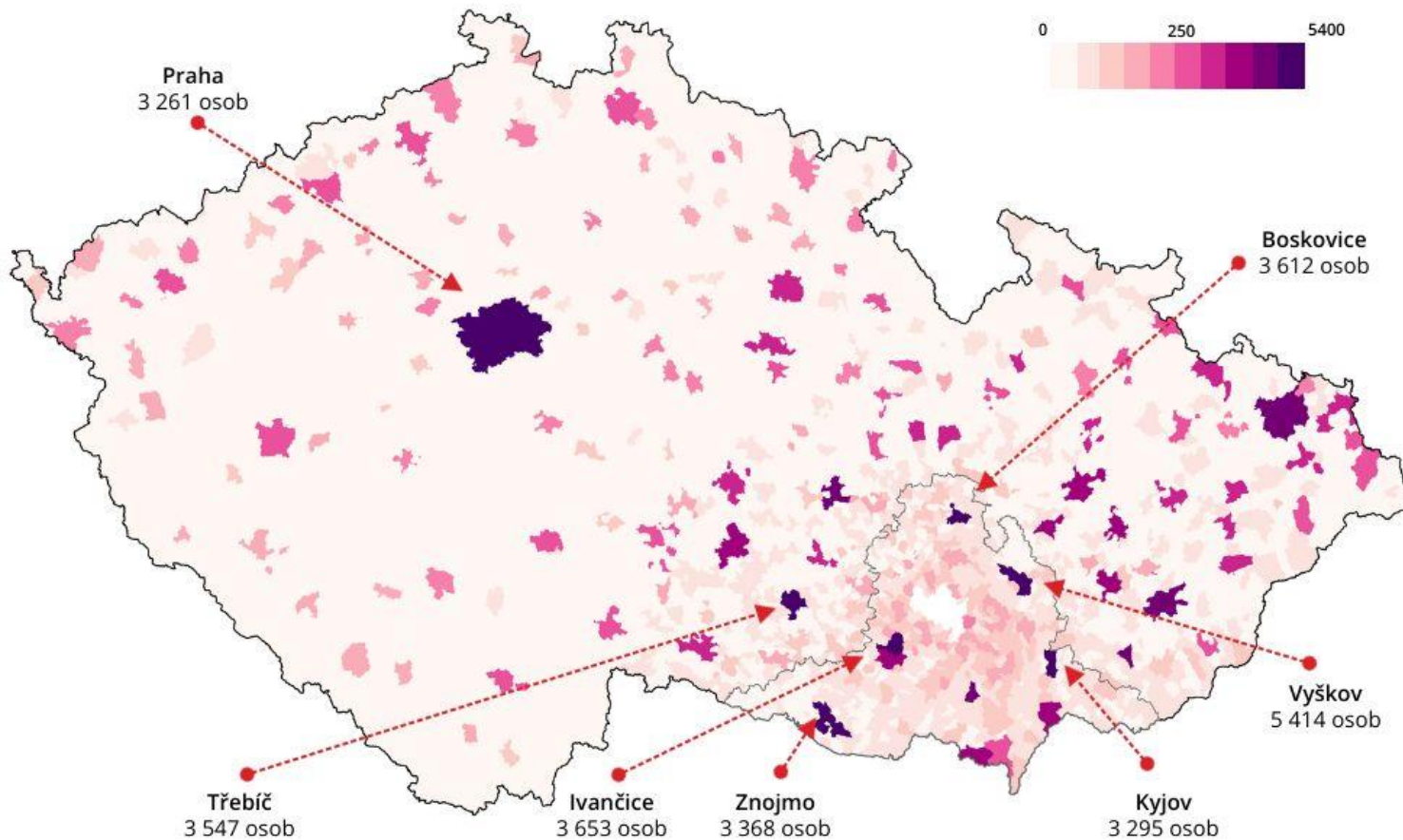
10 experimental maps are made, including 5 rectangular cartograms and 5 choropleth maps (see Table 2.). These 5 rectangular cartograms are generated by the authors' self-developed algorithm, which is based on the construction algorithm [45,46] that based on rectangular partition. The algorithm realizes automatic construction of rectangular

Je to v pořádku?

<p>Italian economic thematic choroplethic map, Italian population thematic rectangular cartogram.</p>	<p>Click the area with the smallest thematic value, or Click the area with the largest theme value.</p>		
<p>German population thematic choroplethic map, German economy thematic rectangular cartogram.</p>	<p>Estimate the thematic value of the red area according to the legend.</p>	 <p>What is the value of area "NI"?</p> <p>A: 600 B: 800 C: 1000</p> <p>Legend</p> <p>[Light Green] = 500 [Medium Green] = 100</p>	 <p>What is the value of area "SN"?</p> <p>A: 400 B: 600 C: 800</p> <p>Legend</p> <p>[Light Green] = 500 [Medium Green] = 100</p>

Odkud pochází Brňané a Brňanky?

Je to v pořádku?



Rodáci & Náplavy

V Brně je registrováno celkem 404 002 osob s trvalým pobytem.

Dle **místa narození** můžeme odlišit, kdo se do Brna **přistěhoval** (175 101), a kdo je naopak **rodákem** (228 901).

Z dat vyplývá, že nejčastěji se do Brna stěhují lidé z Vyškova, Boskovic či Ivančic.

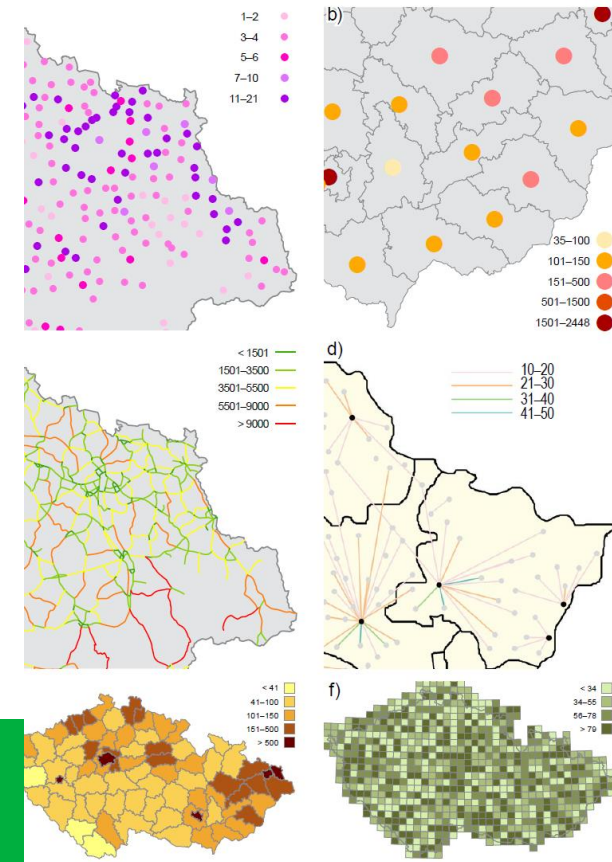
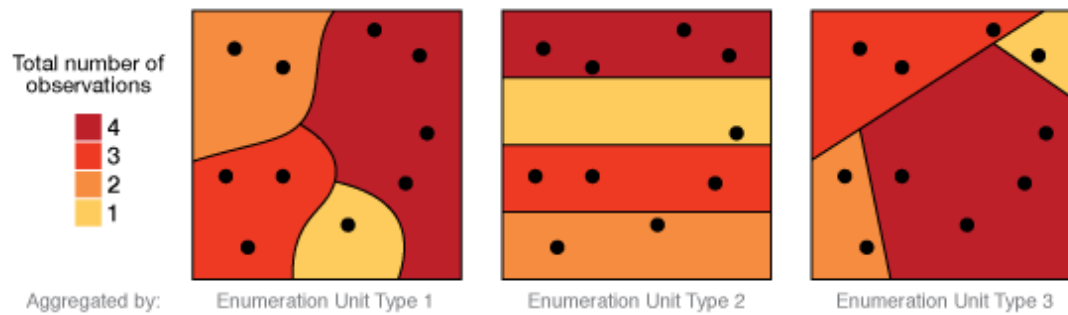
Je však nutné brát v úvahu, že jsou to města s **porodnicí** a značná část Brňanekjezdí do porodnice právě do těchto měst.

Místo narození se totiž udává dle porodnice, nikoli dle místa, kde trvale žijí rodiče.

Zdroj dat: interní databáze MMB
Licence: CC BY

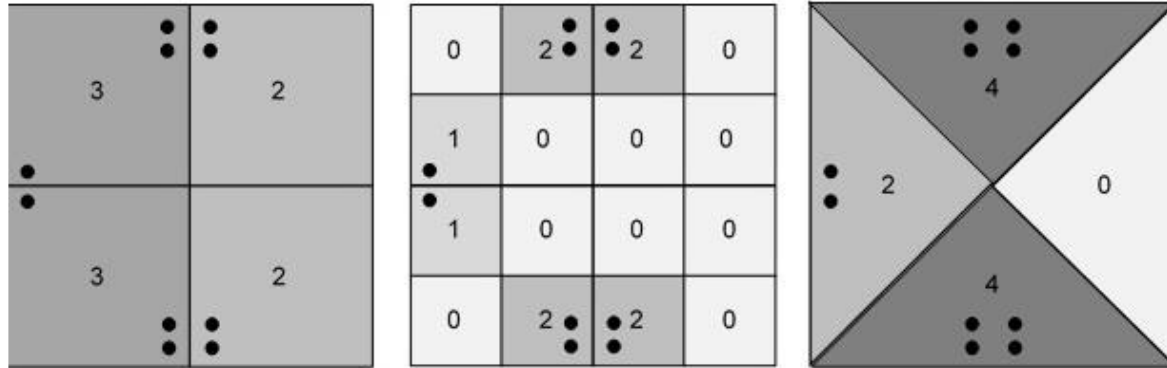
Kartogram, choropleťová mapa

- Kartogram je mapa s dílčími územními celky, do kterých jsou plošným způsobem znázorněna statistická data (jedná se o relativní hodnoty) většinou geografického charakteru (Kaňok, 1999).
- Vyjadřuje hodnotu jevu barvou (respektive odstínem šedé, rastrem, šrafováním), přičemž tyto barvy jsou uspořádány do stupnice
- Výše uvedené se týká areálových/plošných dat, ale metoda intenzitních barev může být aplikována i na body nebo linie →
- Nepravý kartogram
 - Tzv. kartogramy bez prostorového základu
 - Nepoužívají se data přepočtená na jednotku plochy
 - V praxi častější než běžný kartogram

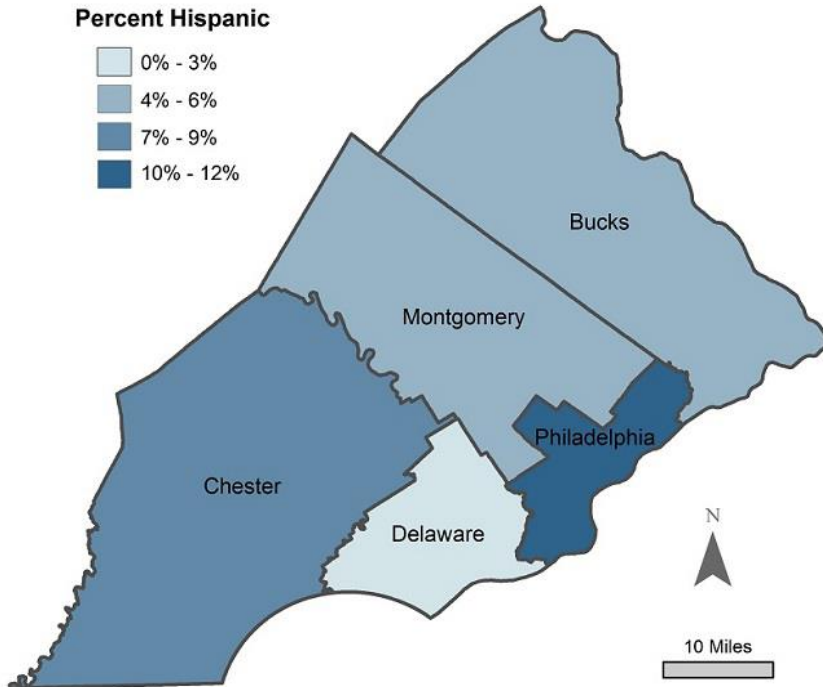
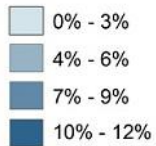


Měřítko a vymezení areálů

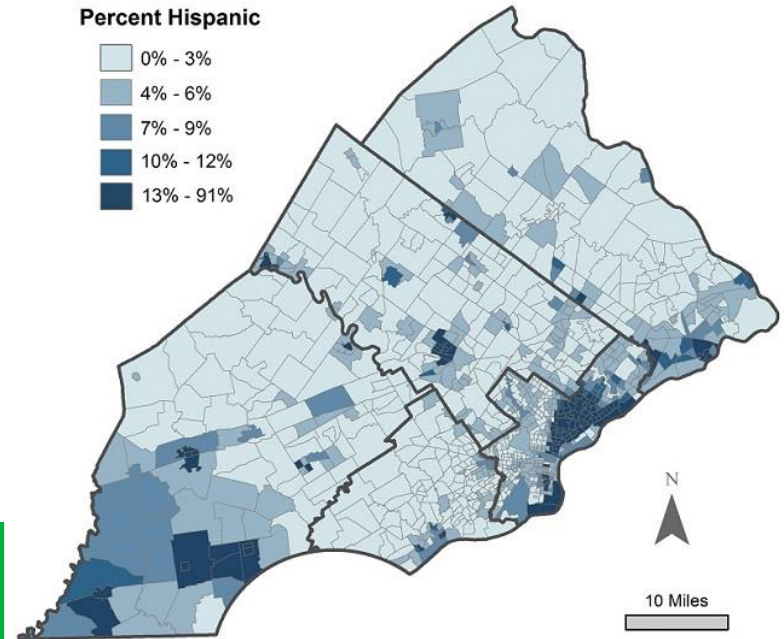
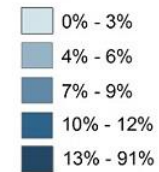
Modifiable areal unit problem (MAUP)



Percent Hispanic



Percent Hispanic



Relativní a absolutní data

- Můžu zobrazovat i absolutní data?
 - Jen když jsou použity jednotky stejného velikosti a tvaru
- Jak je získat data relativní?
 - Už je stáhnu, dostanu,
 - Můžu je vypočítat z absolutních!
- Normalizace, standardizace

Symbology - ZSJ_P

Primary symbology
Graduated Colors

Field: FID_1

Normalization: SHAPE_Area

Method: Defined Interval

Interval size: 100

Classes: 5

Color scheme: [Yellow to Red gradient]

show:

Features


Categories

Quantities

- Graduated colors
- Graduated symbols
- Proportional symbols
- Dot density

Charts

Multiple Attributes



Draw quantities using color to show values.

Fields: Value: POCET_OB

Normalization: Rozloha [km2]

Classification: Manual

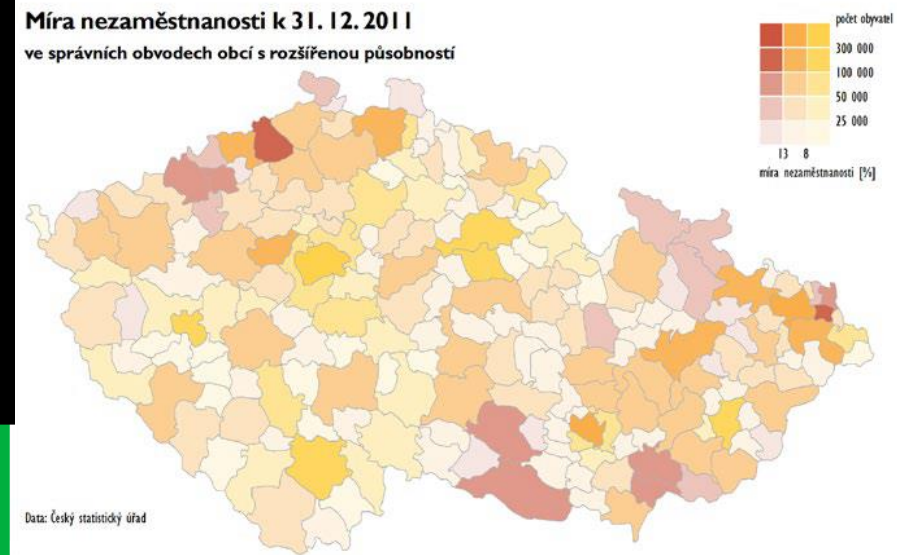
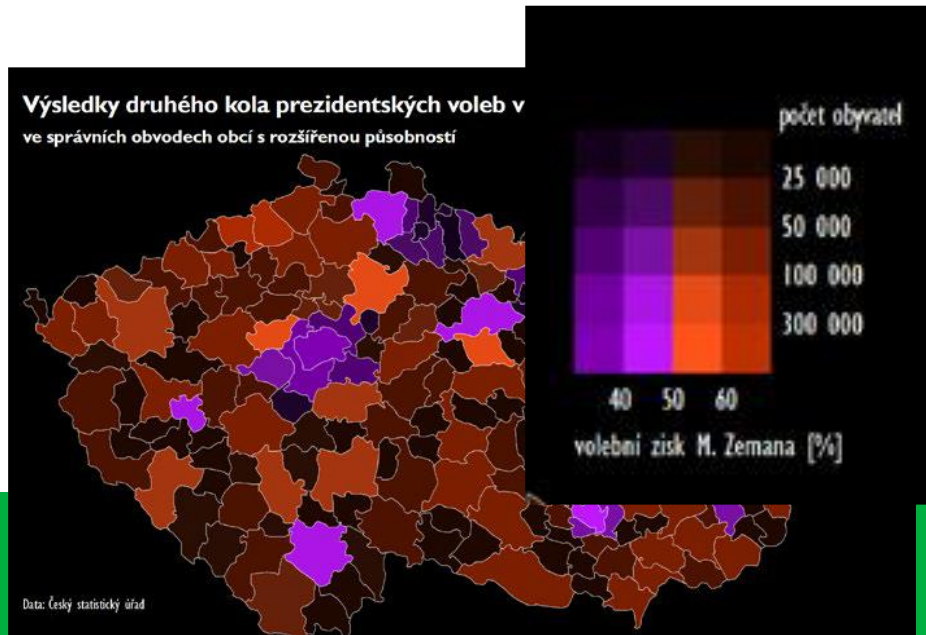
Classes: 5

Color Ramp: [Green gradient]

Symbol	Range	Label
[Lightest Green]	37,1536146 - 96,4999999	37 - 96
[Light Green]	96,5000000 - 159,4999999	97 - 159
[Medium Green]	159,5000000 - 297,4999999	160 - 297
[Dark Green]	995,4999999	298 - 995
[Darkest Green]	995,5000000 - 2503,50000	996 - 2503

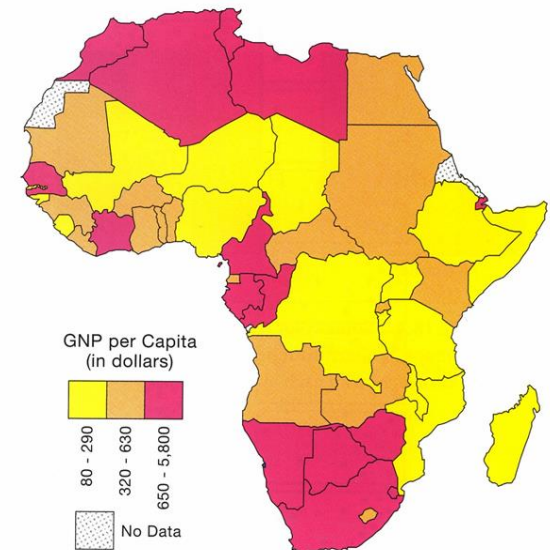
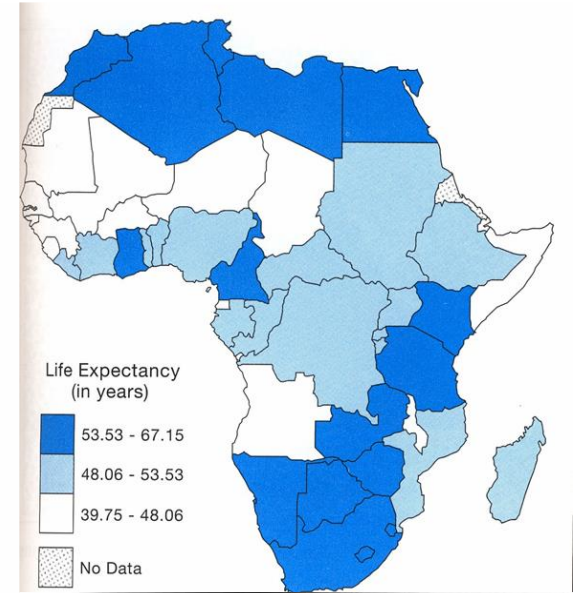
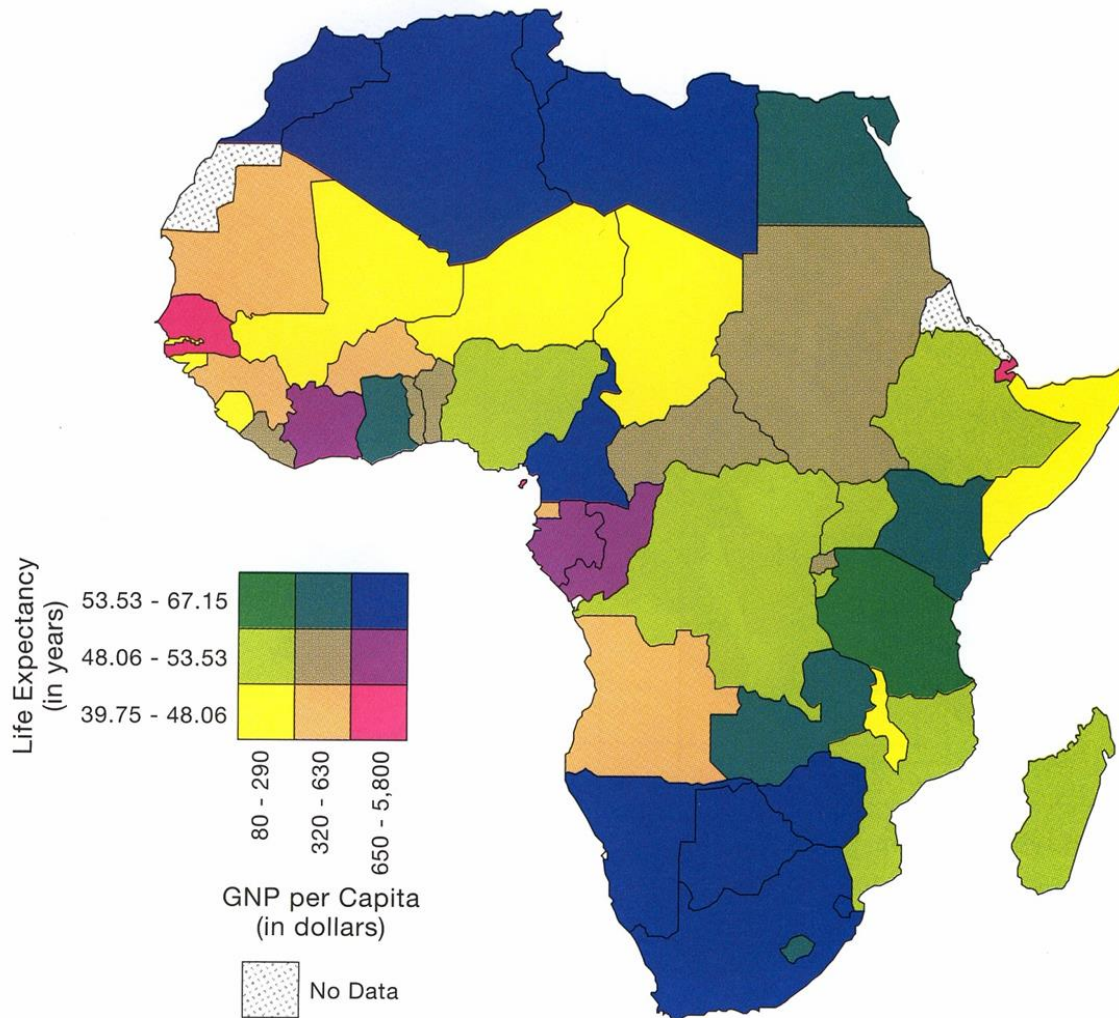
Normalizace

- Dva obecné typy normalizace:
 - **statistická** normalizace – prostřednictvím dat
 - **vizuální** normalizace – s využitím grafických proměnných
 - úpravou sytosti (nebo průhlednosti) = „value-by-alpha“ mapa
 - úpravou velikosti jednotky = anamorfóza

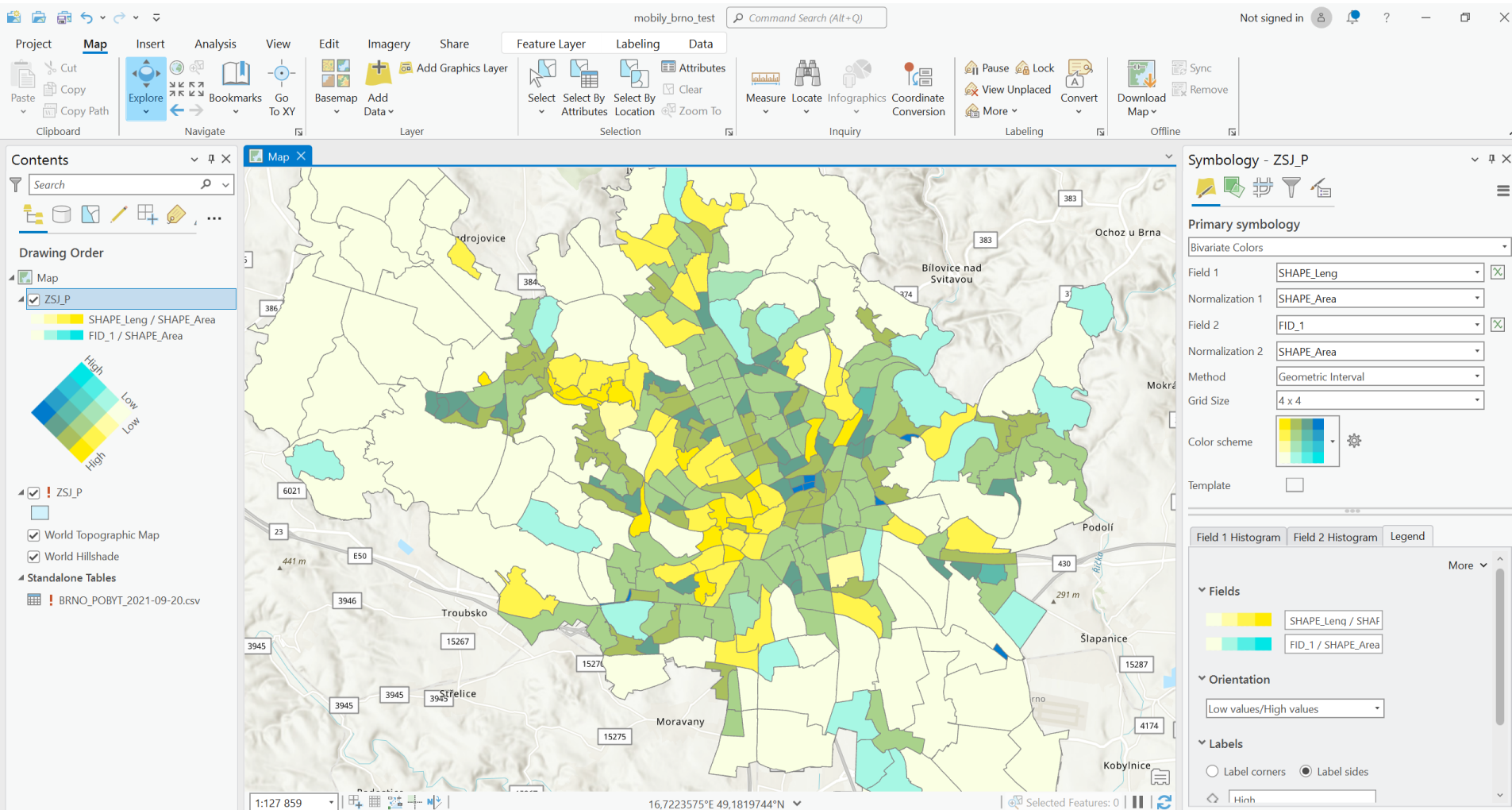


Bivariantní kartogram

Life Expectancy and GNP per Capita
Africa, 1989



Bivariantní kartogram – ArcGIS Pro

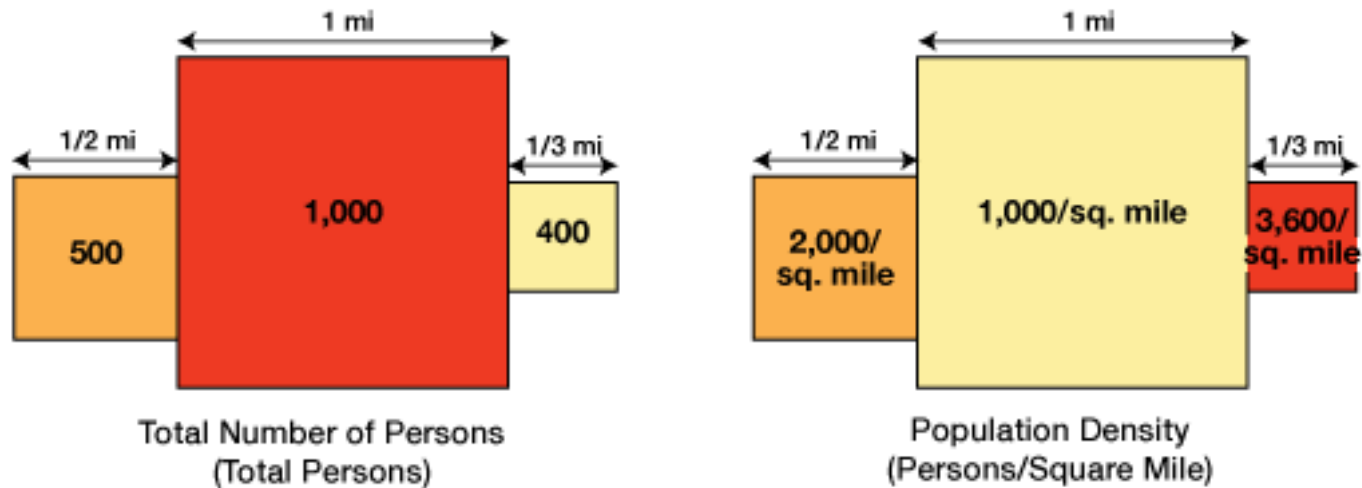


Ke statistické normalizaci můžeme použít:

- 1) Plochu
- 2) Souhrnnou hodnotou za jednotku
- 3) Souhrnnou (průměr, modus, medián) hodnotou za všechny jednotky
- 4) Relevantní populaci
- 5) Předchozí časové období – *viz vizualizace změn*

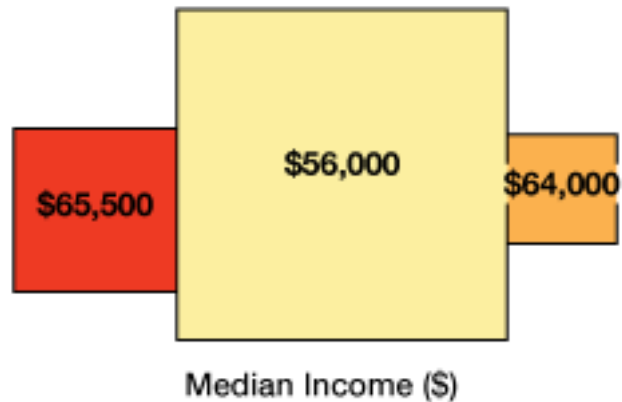
Normalizace plochou

- Pravý kartogram
- Vzniká hustota
- Nejznámější příklad: hustota zalidnění (obyv./km²)



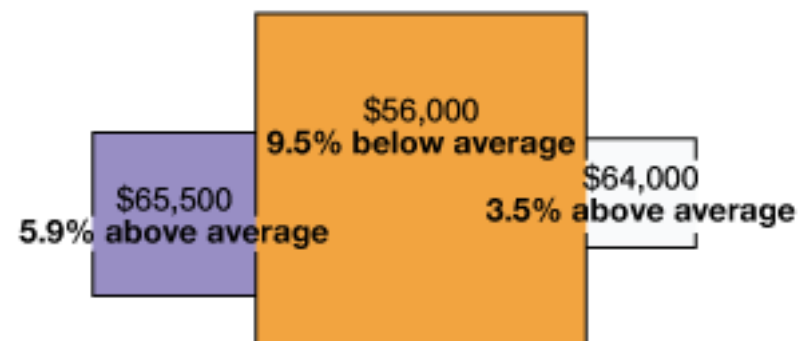
Normalizace souhrnnou hodnotou za jednotku

- Spočítám průměr/medián v jednotce
- Je to sice průměr, ale nejsou to zase absolutní data?!?



Normalizace souhrnnou hodnotou za všechny jednotky

- Vypočítám průměr, modus, medián ... za všechny územní celky
- Dílčí jednotky pak znázorňují odchylku od „střední“ hodnoty



Comparison to Average Regional Income (\$)
(Percentage Above/Below Average Regional Income)

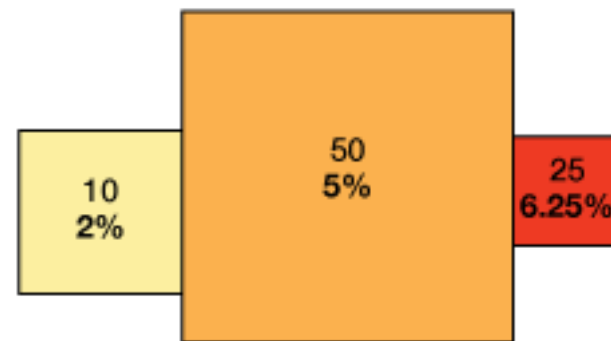
Normalizace relevantní populací

- „Per capita“ (na hlavu), na 1000 obyvatel, ...
- Počet obyvatel, počet nemocných, počet ekonomicky aktivních, počet domácností, ...

14 denní počet nových případů k 31.1.2021 (na 100 000 obyv.)



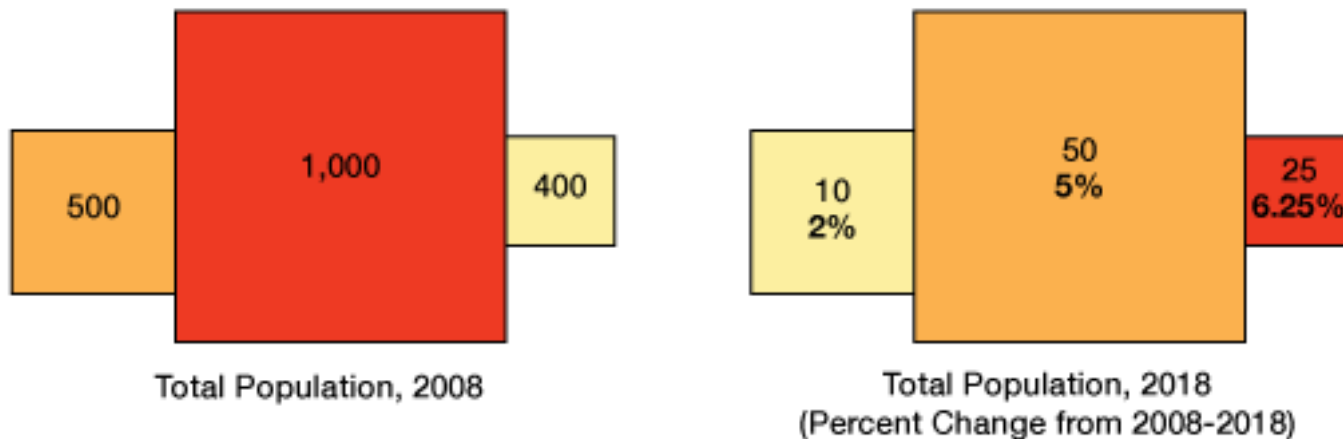
Number of Persons in Labor Force
(Total Persons in Labor Force)



Persons Unemployed and Percentage Unemployed
(Persons in Labor Force Seeking Work/
Total Persons in Labor Force)

Normalizace předchozím časovým obdobím

- Viz časové indexy

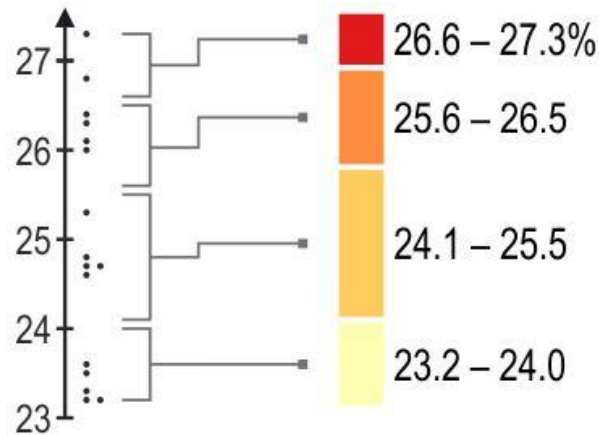


Normalizace – závěr

- Při normalizaci je nezbytně nutné, aby byla data normalizována vůči stejnému univerzu hodnot, ve kterých byly naměřeny dané jevy.
 - ***Podíl musí dávat smysl!***
- Příklad: zastoupení nízkopříjmových domácností by měly být normalizovány vůči celkovému počtu domácností, nikoli vůči celkovému počtu obyvatel.
- Pro pravdivou a efektivní kartografickou komunikaci dat je důležitá normalizace, a to jak statistická tak i vizuální.

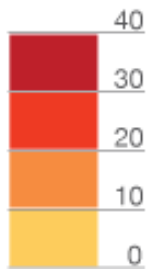
Klasifikace

base map and data → classification → symbolisation → choropleth map

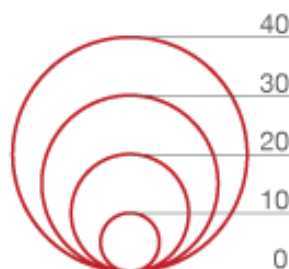


Share of population over the age of 60

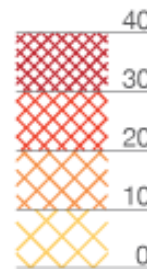
data: Eurostat, © EuroGeographics for the administrative boundaries



Choropleth



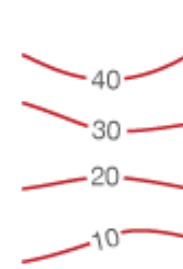
Graduated Symbol



Pattern

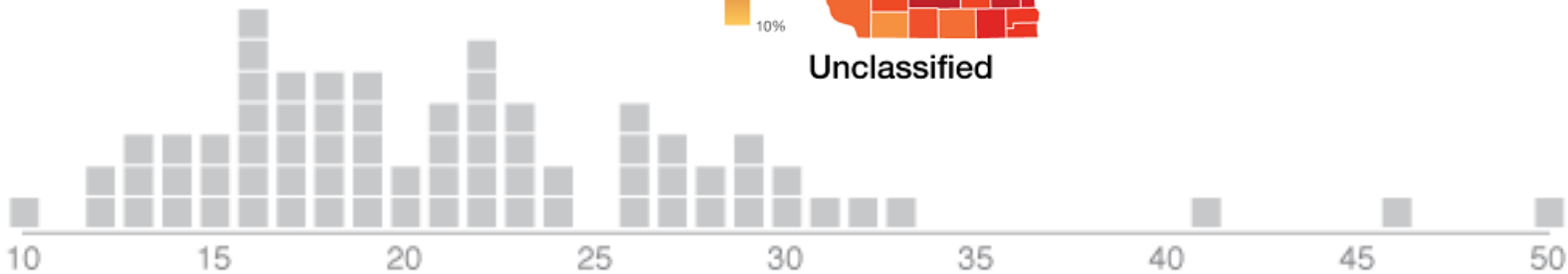
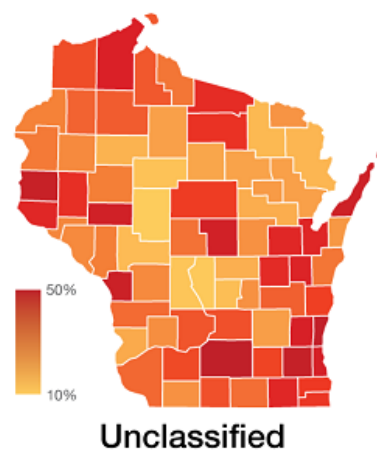


Flow



Isoline

Klasifikace – příklad



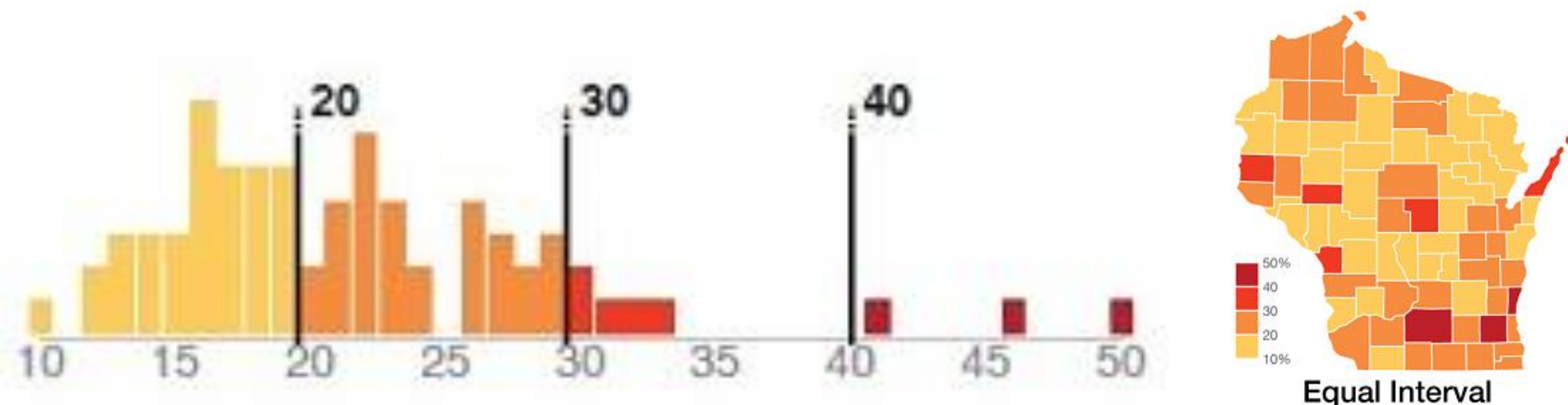
- *Percentage of residents over the age of 25 in Wisconsin that possess a Bachelor's degree or higher in Wisconsin in 2016 by county*
- *There are 72 values in the dataset representing one for each county, and the range is from a minimum of 10% and a maximum of 50%.*
- *The data is based on the American Community Survey 5-year Estimates for educational achievement from 2012-2016 .*
- *Data is modified very slightly for simplicity of illustration (the highest value, Dane County, rounded up to 50% from 49%, and lowest value, Clark County, rounded down to 10% from 11%).*
- *The data is not heavily skewed, although there is a slight positive skew with some outliers.*

Klasifikace – metody

- 1) Stejné intervaly
 - 2) Kvantily
 - 3) Průměr a násobky směrodatné odchylky
 - 4) Metoda maximálních zlomů (Maximum breaks)
 - 5) Metoda přirozených zlomů
 - 6) Vlastní
- Jenks

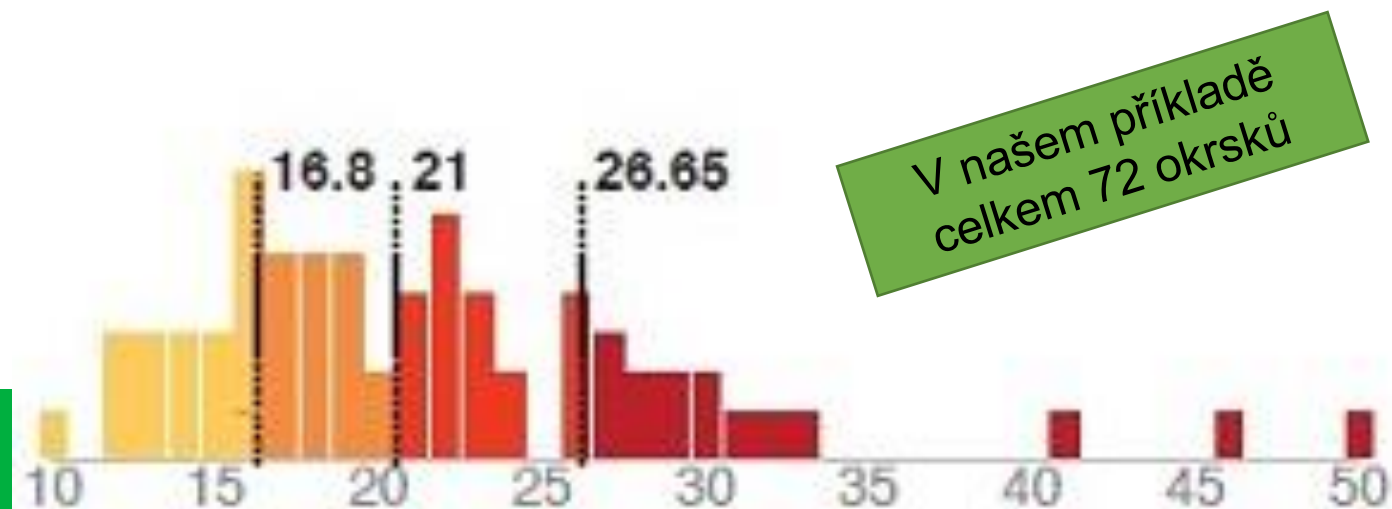
Stejné intervaly

- Zadává se počet intervalů a data jsou rozdělena do intervalů o stejném rozsahu (ArcGIS Pro: *Equal Interval* → počet; *Defined Interval* → šířka)
- Mohou nastat případy, kdy v daném rozsahu třídy bude nula prvků!
- Vzhledem k tomu, že v hraničních třídách se vyskytují většinou málo četné odlehle hodnoty, je toto rozdělení vhodné pro **zvýraznění extrémů**.
- **Není vhodné**, pokud je **rozdělení dat zešikmené** nebo existují v něm **příliš odlehle hodnoty**.
- Částečně související metodou jsou geometrické intervaly
 - Metoda definuje nejužší třídu a šířky ostatních tříd odvozuje pomocí proměnlivého násobného faktoru

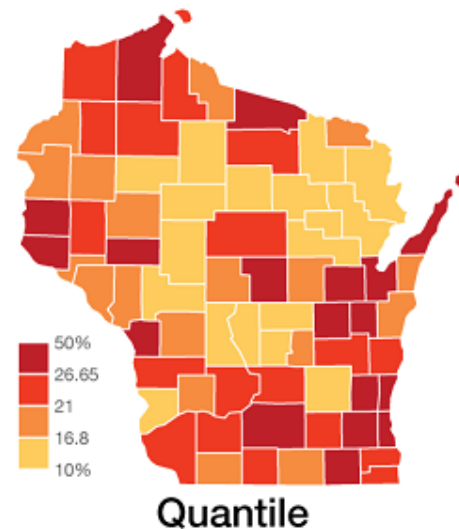


Kvantily

- Kvartily, pentily, decily, percentily...
- *ArcGIS Pro: Quantile*
- Metoda rozděluje data do nerovnoměrně velkých tříd, ale se stejným počtem prvků ve třídách.
- Metodu je vhodné použít v případě, kdy jsou data lineárně distribuována s přiměřeným počtem prvků s podobnými hodnotami nebo pokud se vyskytují extrémní hodnoty.
- Metoda je nevhodná v případech velkého množství prvků s podobnými hodnotami

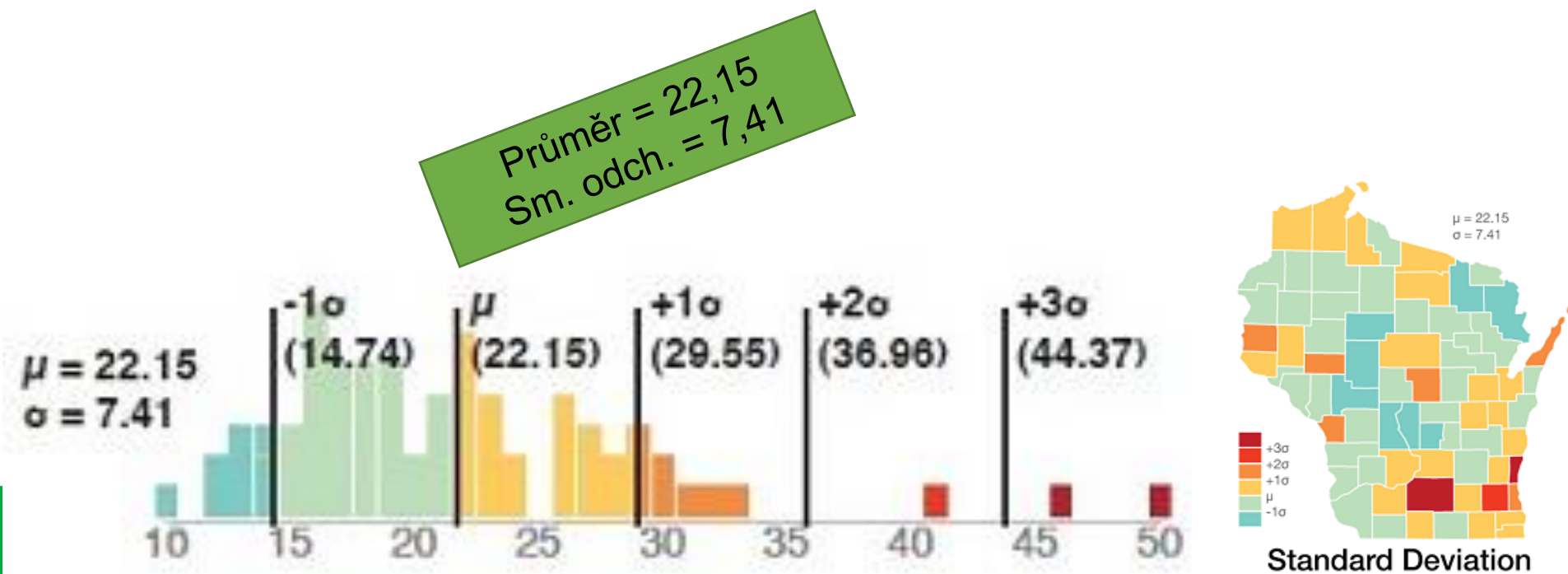


V našem příkladě celkem 72 okrsků



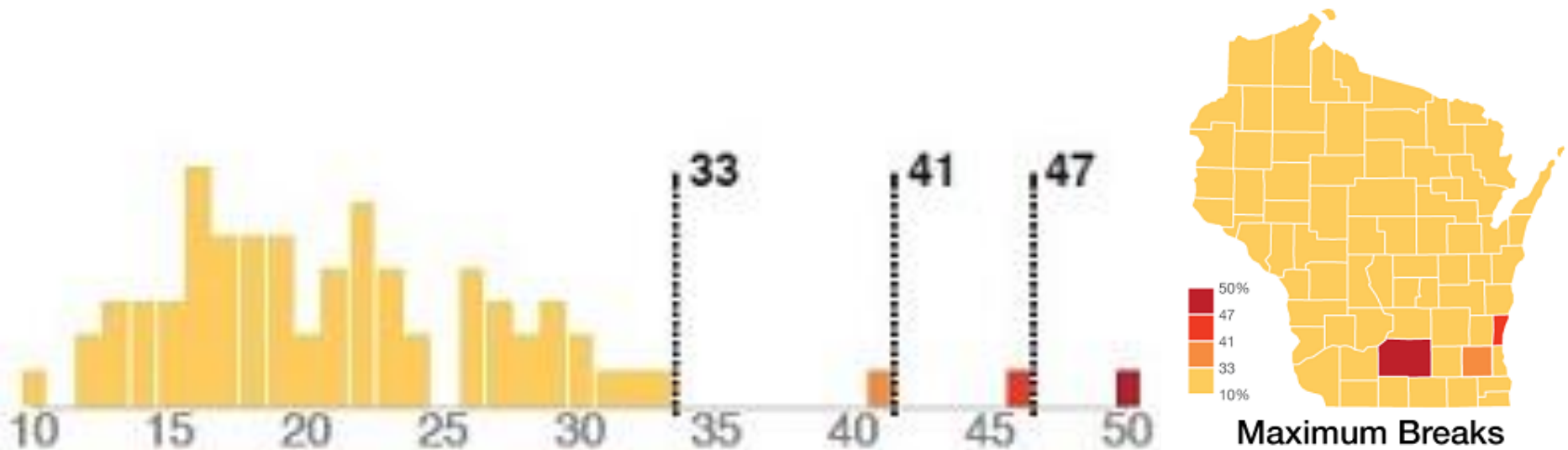
Průměr a násobky směrodatné odchylky

- Metoda vytváří třídy jako podíly směrodatné odchylky nad a pod průměrem dat, neboli ukazuje, jak moc se data odchylojí od průměru
- Ideální pro data kterém mají normální rozdělení
- Není vhodná v případě velkého počtu extrémních hodnot.



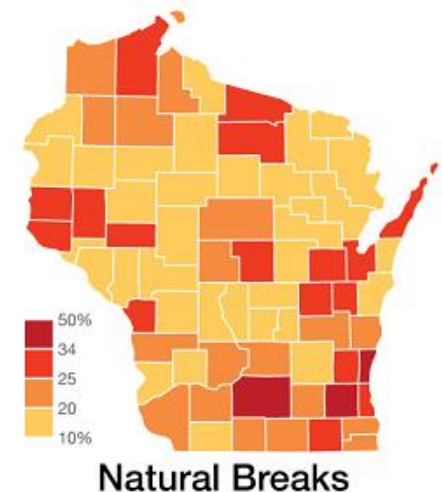
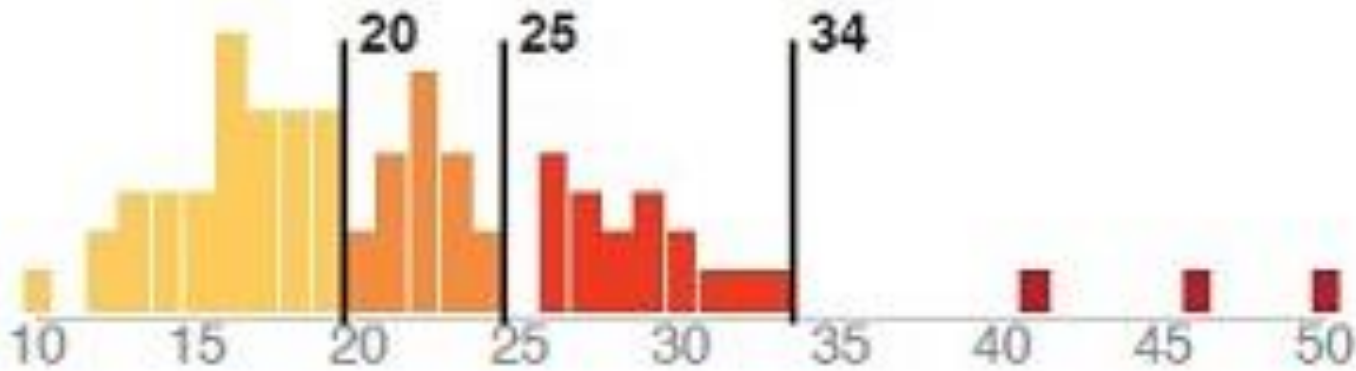
Metoda maximálních zlomů

- Hledají se největší mezery v histogramu
- Hranice se tříd se umístí do největších „mezer“
- Nevhodné v případě dat s několika odlehlými hodnotami



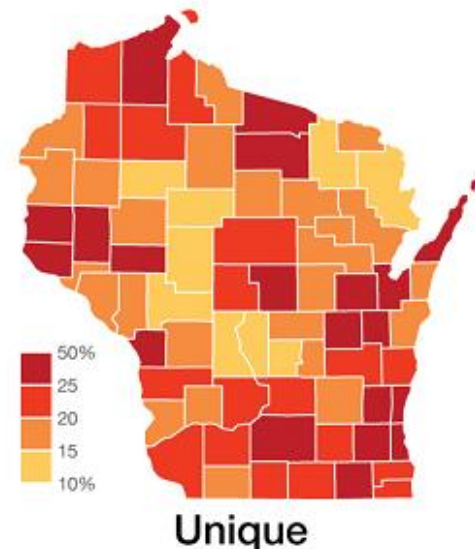
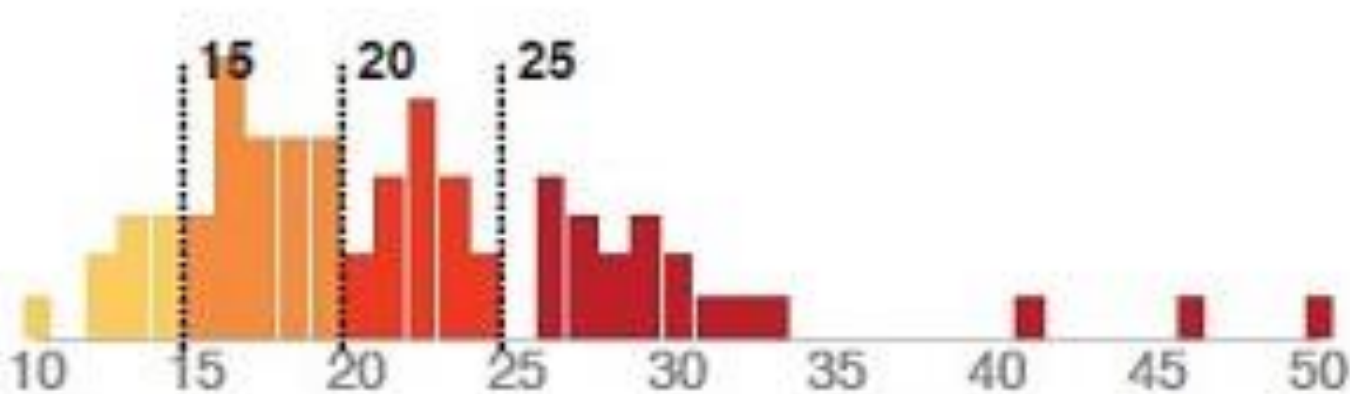
Metoda přirozených zlomů

- Natural breaks
- Vychází z analýzy histogramu
- Hledají se lokální minima (zlomy)
- Subjektivní



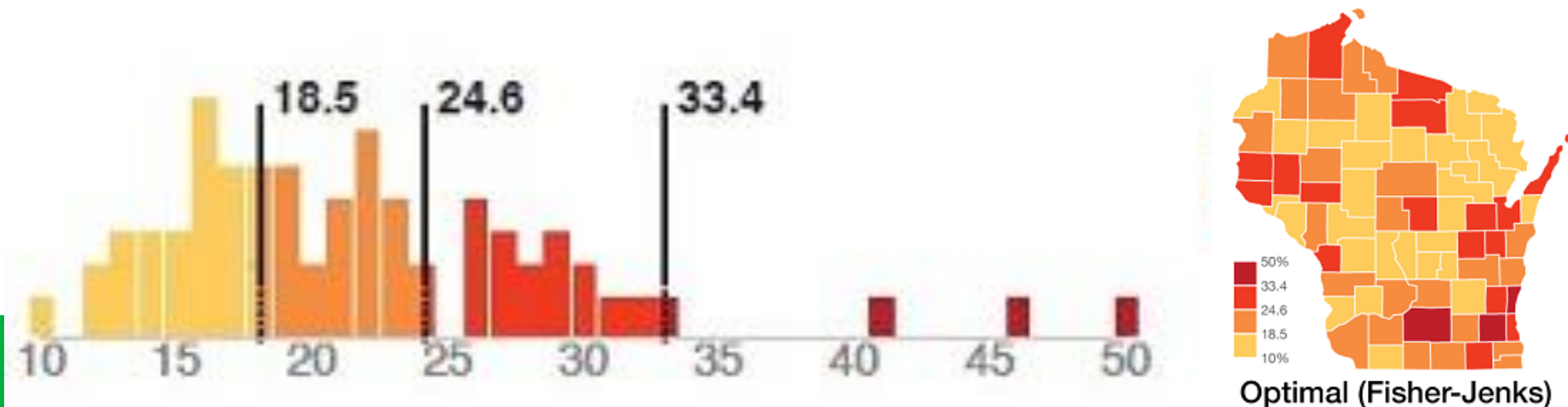
Vlastní klasifikace

- Manuální zadání
- Např. pokud jsou předem dány klíčové hodnoty
- Při nevhodném zvolení šířky třídy se může stát, že poslední třída nesoucí maximální hodnoty může zůstat z velké části prázdná



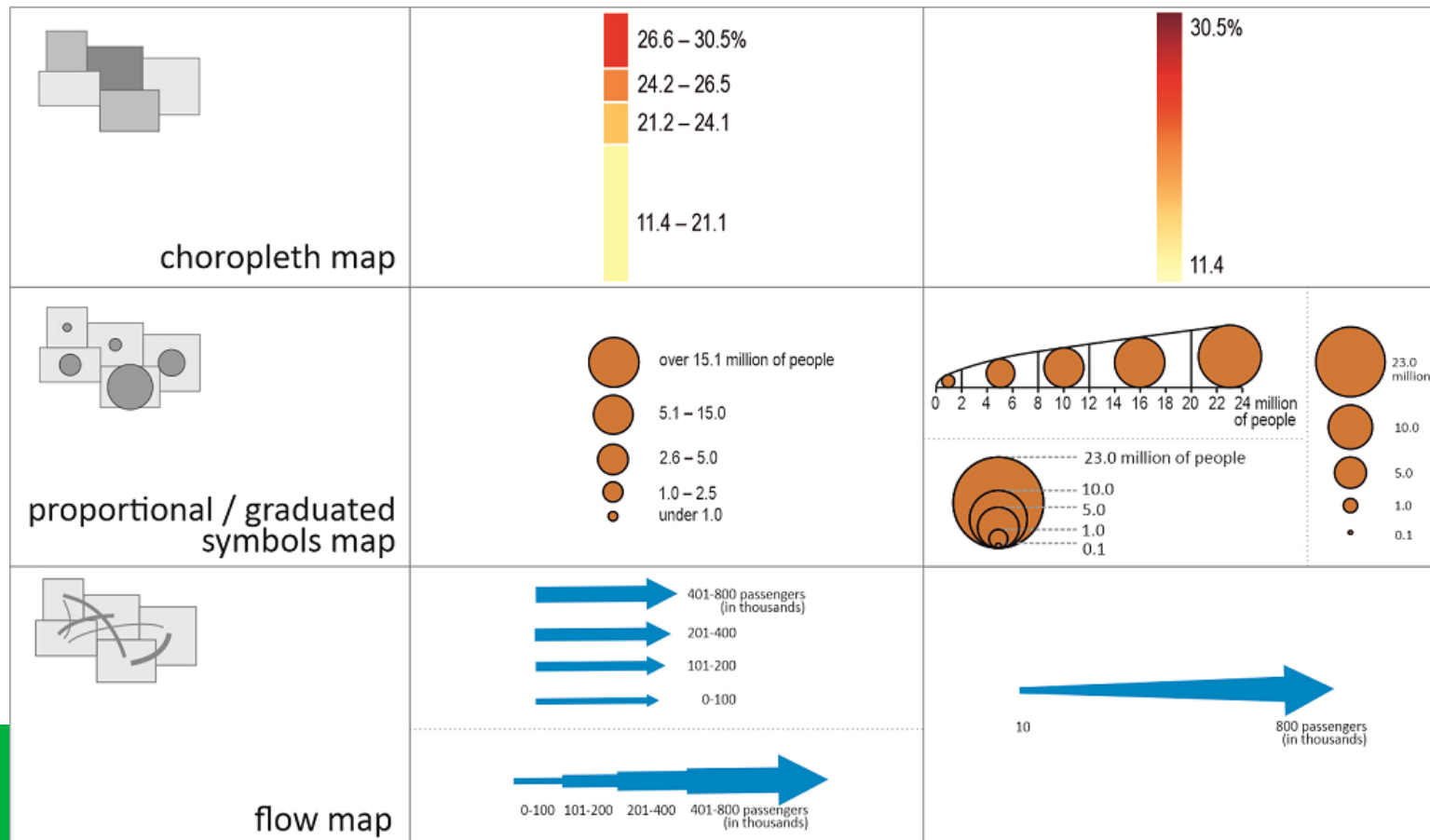
Jenks

- Metoda hledá přirozené zákonitosti a seskupení v datech a vytváří třídy na základě těchto přirozených skupin.
- Hranice jsou definovány v místech s relativně velkými rozdíly v datech
- ArcGIS Pro: Natural Breaks (Jenks), v QGISu stejný název
- Jedná se o univerzální klasifikační metodu, vhodnou pro většinu dat a začátečníky bez hlubší znalosti klasifikačních metod.
- Vždy je však vhodné hranice intervalů manuálně upravit (zaokrouhlit) na „rozumné hodnoty“.



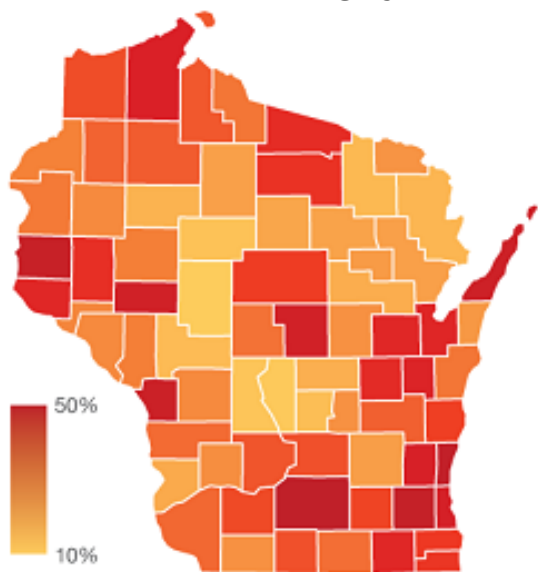
Je klasifikace nutná?

- Klasifikace usnadňuje a zrychluje čtení mapy = identifikaci prostorového vzoru



Někdy to ale jde i bez klasifikace, ale ...

- Kontinuální barevná škála → obtížné porovnávání i odhad hodnoty
- Výšky extrudovaných areálů („3D“) nejsou ovlivněny klasifikací do konečného počtu tříd. V podstatě zobrazují surová data a výška může napomoci například při volbě klasifikace.
 - Příklad: *výška areálu s dvojnásobnou hodnotou zpracovávaného atributu je zobrazena jako dvojnásobná.*
 - Funguje především v případě **interaktivní** 3D vizualizace



Unclassified



Lze vyzkoušet zde: https://oli.wz.cz/3d_traffic_offences/speed-districts.html

Co v ArcGIS Pro ...

Symbology - obecni_byty_brno

Primary symbology

Graduated Colors

Field: kod

Normalization: Shape_Are

Method: Natural Breaks (Jenks)

Classes

Color scheme

Classes Histogram

Symbol

Natural Breaks (Jenks)
Numerical values of ranked data are examined to account for non-uniform distributions, giving an unequal class width with varying frequency of observations per class.

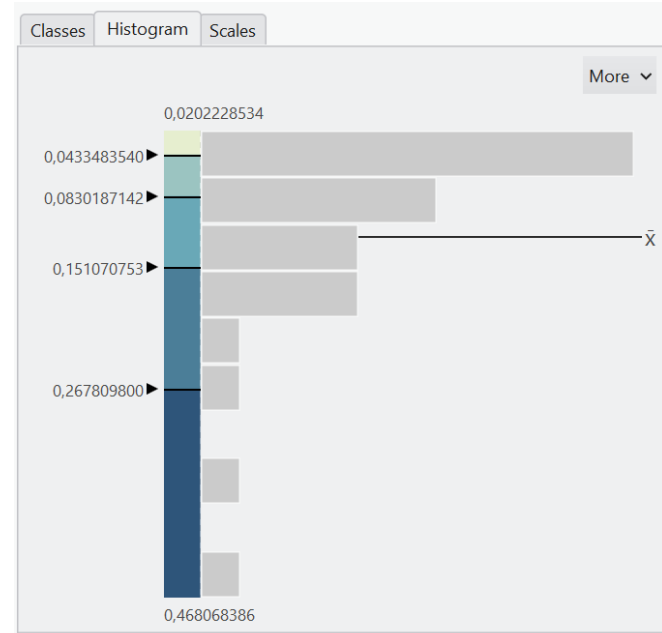
Quantile
Distributes the observations equally across the class interval, giving unequal class widths but the same frequency of observations per class.


Equal Interval
The data range of each class is held constant, giving an equal class width with varying frequency of observations per class.


Defined Interval
Specify an interval size to define equal class widths with varying frequency of observations per class.


Manual Interval
Create class breaks manually or modify one of the preset classification methods appropriate for your data.


Geometric Interval
Mathematically defined class widths based on a geometric series, giving an approximately equal class width and consistent frequency of observations per class.




- 

Equal Interval
The data range of each class is held constant, giving an equal class width with varying frequency of observations per class.
- 

Defined Interval
Specify an interval size to define equal class widths with varying frequency of observations per class.
- 

Manual Interval
Create class breaks manually or modify one of the preset classification methods appropriate for your data.
- 

Geometric Interval
Mathematically defined class widths based on a geometric series, giving an approximately equal class width and consistent frequency of observations per class.
- 

Standard Deviation
For normally distributed data, class widths are defined using standard deviations from the mean of the data array, giving an equal class width and varying frequency of observations per class.

Zdroje:

- Miklín, J., Dušek, R., Krtička, L., Kaláb, O. (2018). Tvorba map. Ostrava: Ostravská univerzita. ISBN: 978-80-7599-017-4, 302 s.
<https://tvorbamap.osu.cz/ke-stazeni/>
- Kaňok, J. (1999). Tematická kartografie. Ostrava: Ostravská univerzita Přírodovědecká fakulta Ostravské univerzity. ISBN: 80-7042-781-7, 318 s.
<https://gistbok-topics.ucgis.org/CV-03-005>
- <https://gistbok-topics.ucgis.org/CV-04-011>
- <https://gistbok-topics.ucgis.org/FC-07-026>
- A odkazy přímo na slajdech

MUNI
SCI

GIS4SG

Prostorová autokorelace

Statistická analýza plošných jevů

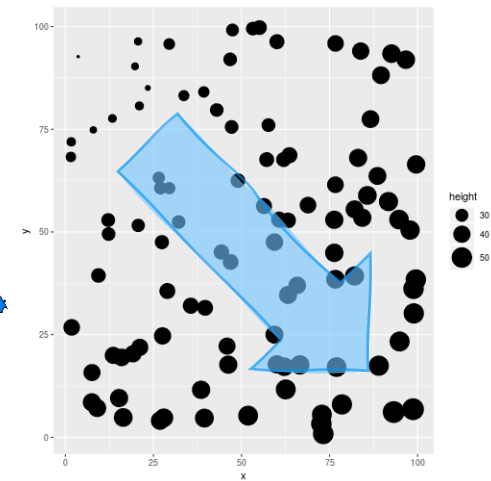
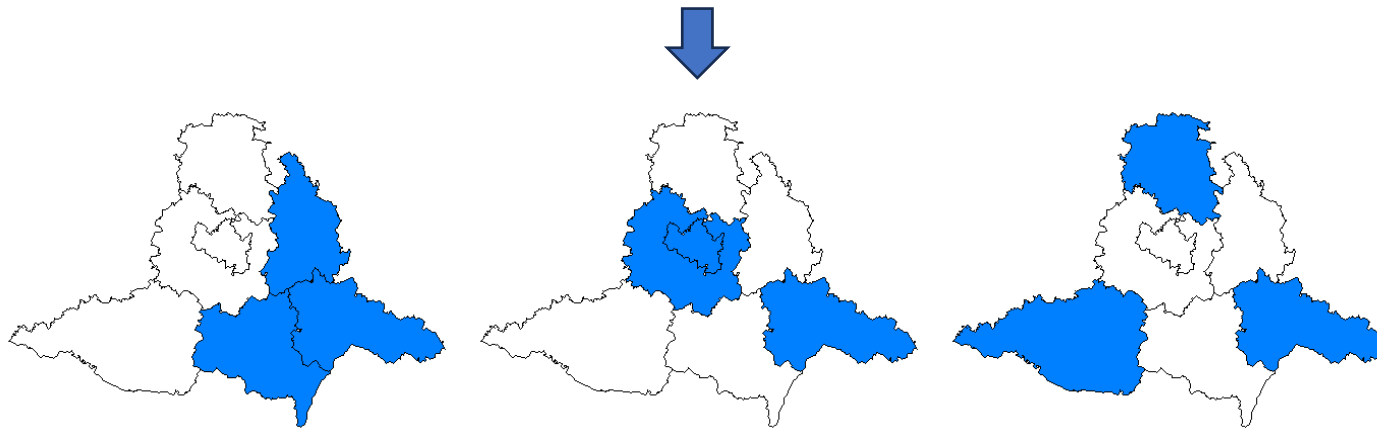
podzim 2024

Lukáš Herman

herman.lu@mail.muni.cz

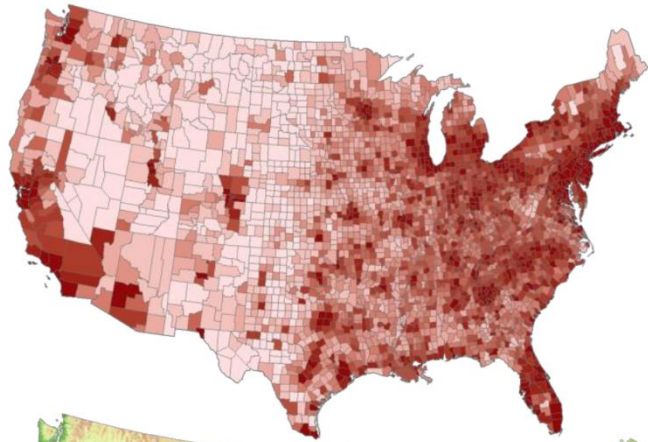
Prostorová autokorelace

- Hodnoty atributů spolu korelují v závislosti na jejich vzájemné poloze.
- To může být v důsledku podobných přirozených (přírodních) podmínek (*např. produkce zemědělských podniků*) či v důsledku přirozené spojitosti jevů.
- Příklad – okresy Jihomoravského kraje: pozitivní prostorové autokorelace (*shlukové uspořádání – vlevo*) a negativní prostorové autokorelace (*disperzní uspořádání – vpravo*)

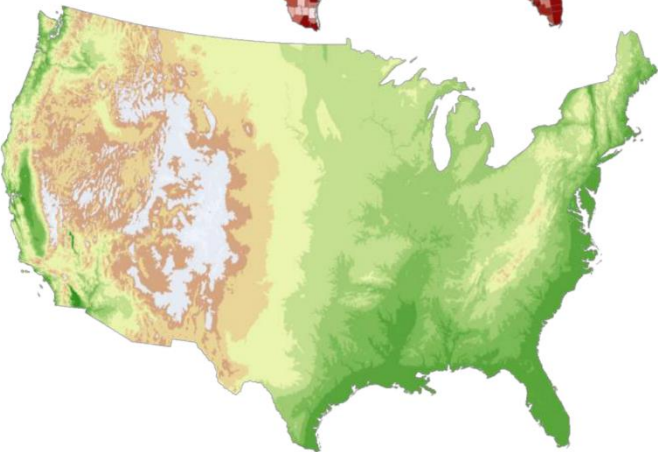
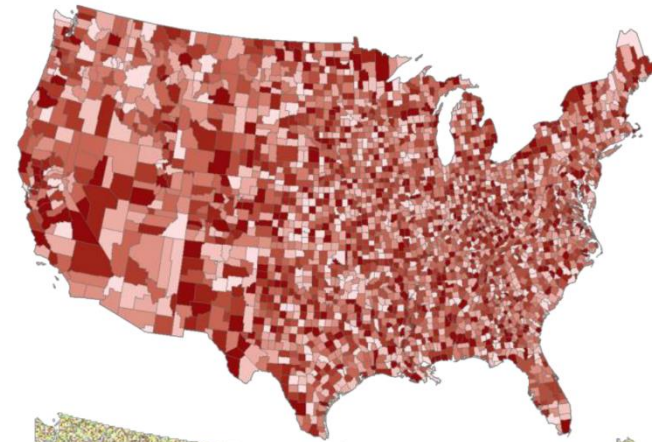


Prostorová autokorelace

If features were randomly distributed ...



... population density map of the US would look like this



... elevation map of the US would look like this



Statistická analýza plošných jevů

- porovnání prostorového uspořádání studovaného jevu s uspořádáním teoretickým (shlukovým, pravidelným či náhodným)
- typologie prostorového uspořádání jevů (bez uzemní souvislosti)
- regionalizace – seskupování jednotek (polygonů) do vyšších územně souvisejících celků
- interpolace a vyhlazování areálových dat



Positive autocorrelation



Negative autocorrelation



No spatial autocorrelation

Jak uchopit „sousedství“?

- Prostorová autokorelace měří stupeň podobnosti atributů **mezi danou plochou a plochami sousedními**. Nejprve proto musí být vztahy sousedství jistým způsobem kvantifikovány.
- Způsoby definování sousedství (*Rook's case* – věž, *Queen's case* – Dáma)
- Binární matice konektivity (sousedí – 1, nesousedí – 0)
- Stochastická matice = matice se standardizovanými řádkovými vahami (RSWM) – záleží na počtu sousedů (př.: 4 → 0,25)

<i>Id</i>	<i>Brno_venko</i>	<i>Blansko</i>	<i>Vyškov</i>	<i>Brno_město</i>	<i>Hodonín</i>	<i>Znojmo</i>	<i>Břeclav</i>
Brno-venkov	0.0000	0.2000	0.2000	0.2000	0.0000	0.2000	0.2000
Blansko	0.3333	0.0000	0.3333	0.3333	0.0000	0.0000	0.0000
Vyškov	0.2500	0.2500	0.0000	0.0000	0.2500	0.0000	0.2500
Brno-město	0.5000	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
Hodonín	0.0000	0.0000	0.5000	0.0000	0.0000	0.0000	0.5000
Znojmo	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.5000
Břeclav	0.2500	0.0000	0.2500	0.0000	0.2500	0.2500	0.0000

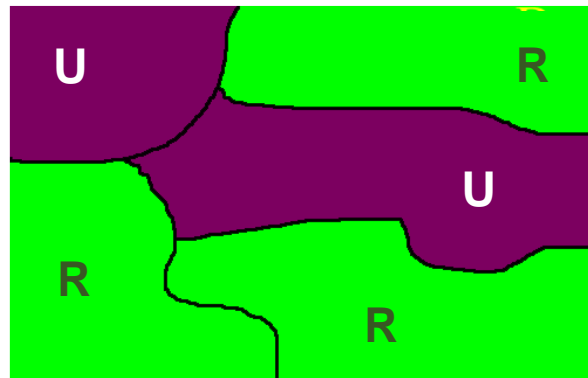
- Vedle sousedství je další běžně užívanou mírou prostorové relace objektů jejich vzdálenost (v případě polygonů např. vzdálenost centroidů)

Míry prostorové autokorelace areálů

- **Globální míry** prostorové autokorelace:
 - Data nominální
 - Joint Count Statistics (JSC) – Statistika charakteru sousedství
 - Data intervalová a poměrová
 - Moranův index I
 - Gearyho poměr C
 - G statistika
- Prostorová autokorelace se může měnit v rámci studované oblasti → **Lokální míry** prostorové autokorelace:
 - Local Indicator of Spatial Association (LISA)
 - Lokální verze G statistiky (local G-statistics).
- Ke grafickým prostředkům hodnotícím prostorovou autokorelaci patří také Moranův scatterplot diagram.

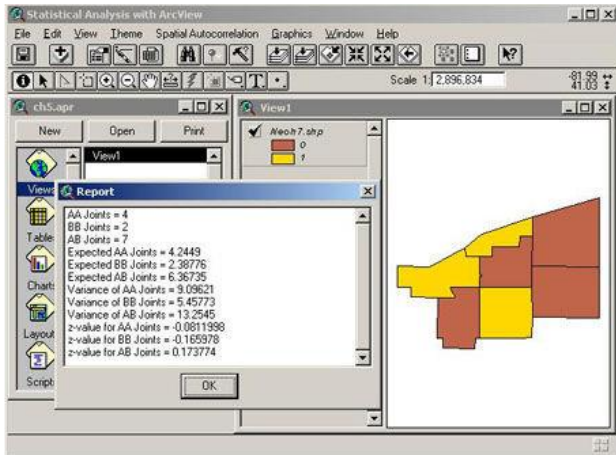
Joint count statistics (JCS)

- Touto metodou lze zjistit, zda uspořádání ploch, které mohou nabývat binárních hodnot vykazuje prvky náhodnosti.
- Tedy zda existuje pozitivní (clustered pattern) či negativní (random pattern) prostorová autokorelace.



- Podstata metody:
 - U – zástavba, R – volná krajina.
 - Čtyři typy sousedství: UU, RR, UR, RU.
 - $UR + RU < 50\%$ → pozitivní prostorová autokorelace.
 - $UR + RU > 50\%$ → negativní prostorová autokorelace

Joint count statistics (JCS)



ZASTOUPENÍ SENIORŮ VE 20 VYBRANÝCH OBCÍCH JIHOMORAVSKÉHO KRAJE

k 31. 12. 2009



**Podíl občanů starších 64 let
ku celkovému počtu obyvatel**

- podprůměrný (<6.30334)
- nadprůměrný (>6.30334)

Jihomoravský kraj

Počet obyvatel: 1150454
Počet obyvatel starších 64 let: 182515
Průměr z poměru celkového počtu obyvatel
v obci ku počtu obyvatel starších 64 let: 6.30334

Počet pozorovaných AA sousedů = 40
Počet pozorovaných BB sousedů = 0
Počet pozorovaných AB sousedů = 2
Očekávaní AB sousedé = 37.905
Očekávaní BB sousedé = 0.105
Očekávaní AB sousedé = 3.99
Variance AA sousedů = 16.6421
Variance BB sousedů = 0.1406
Variance AB sousedů = 22.8314
Z skóre pro AA sousedy = 0.513547
Z skóre pro BB sousedy = -0.280025
Z skóre pro AB sousedy = -0.416473

V našem případě:

A = nadprůměrný podíl občanů starších 64 let k celkovému počtu obyvatel

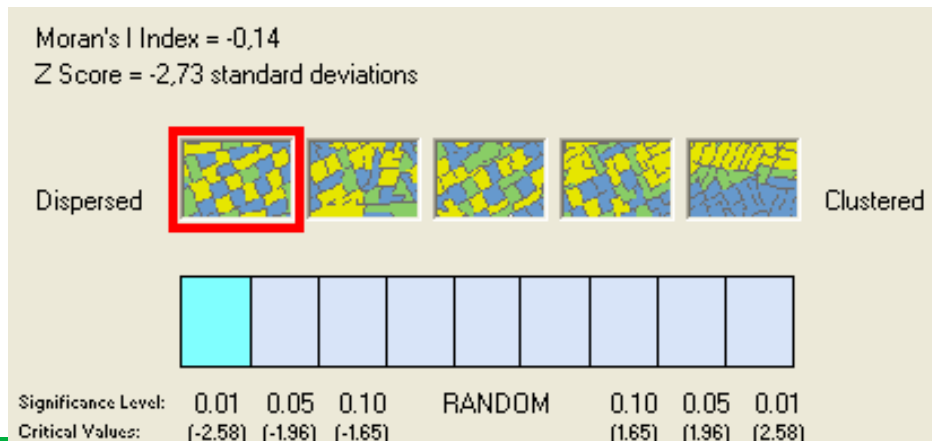
B = podprůměrný podíl občanů starších 64 let k celkovému počtu obyvatel

Indexy pro hodnocení prostorové autokorelace plošných jevů

- **Moranův (I) index a Gearyho (C) index**
- Jsou využitelné pro intervalová a poměrová data
- Jsou založeny na porovnávání hodnot atributů sousedních ploch
- Mají-li tyto sousední plochy v celé studované oblasti podobné hodnoty, potom obě statistiky budou svědčit o silné pozitivní prostorové autokorelaci a naopak.
- Obě statistiky využívají odlišný přístup k porovnávání hodnot sousedních ploch
- Vhodnější vlastnosti vzhledem k rozdělení hodnot má Moranův index

Moranův (I) index

- Hodnota indexu kolísá od -1 pro negativní prostorovou autokorelaci do +1 pro pozitivní prostorovou autokorelaci.
- Vypočteme hodnoty I a $E(I)$ a následně musíme zjistit, zda rozdíl mezi nimi je statisticky významný.
- Tento rozdíl je opět nutné vztáhnout k míře variability (např. rozptylu) a pomocí ní odvodit standardizovanou hodnotu z-skóre
- Pokud je hodnota $Z_n(I)$ menší (resp. větší) než -1,96 (resp. 1,96) je hodnota indexu I statisticky významně negativní (resp. pozitivní) na hladině významnosti $\alpha=0,05$.



There is less than 1% likelihood that this dispersed pattern could be the result of random chance.

Moranův index = -0,144
Očekávaný Moranův index = -0,053
Variance/rozptyl = 0,001
Z skóre = -2,727

Moranův (I) index

— ArcGIS pPo: Spatial Statistics Tools

Spatial Autocorrelation (Global Moran's I) (Spatial Statistics Tools)

Started: Today at 10:17:51
 Completed: Today at 10:17:52
 Elapsed Time: 1 Second

WARNING 001605: Distances for Geographic Coordinates (degrees, minutes, seconds) are analyzed using Chordal Distances in meters.

Parameters Environments **Messages (4)**

Global Moran's I Summary

Moran's Index	0,261567
Expected Index	-0,003390
Variance	0,001280
z-score	7,406022
p-value	0,000000

Writing html report....
C:\Users\Lukas\Documents\ArcGIS\Projects\mobily_brno_test\MoransI_Result_10448_11288
 Succeeded at pondělí 20. března 2023 10:17:52 (Elapsed Time: 1,74 seconds)

Geoprocessing

Spatial Autocorrelation (Global Moran's I)

Parameters Environments

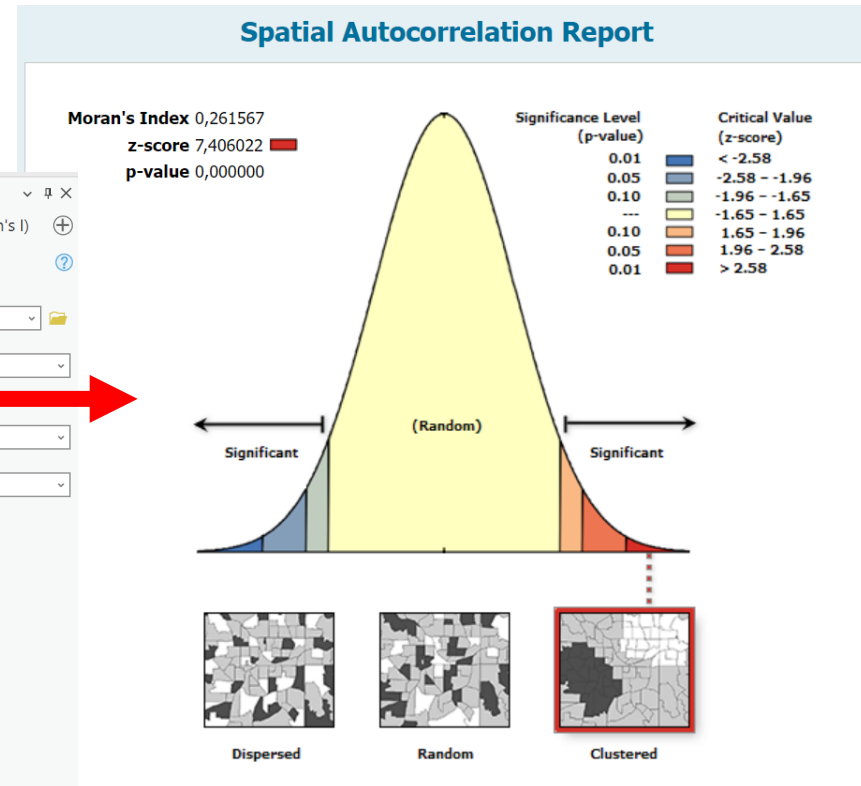
Input Feature Class
brno_obyv_20-9-2021

Input Field
obyv_cas1_

Generate Report

Conceptualization of Spatial Relationships
Contiguity edges only

Standardization
Row



Given the z-score of 7.406022, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Global Moran's I Summary

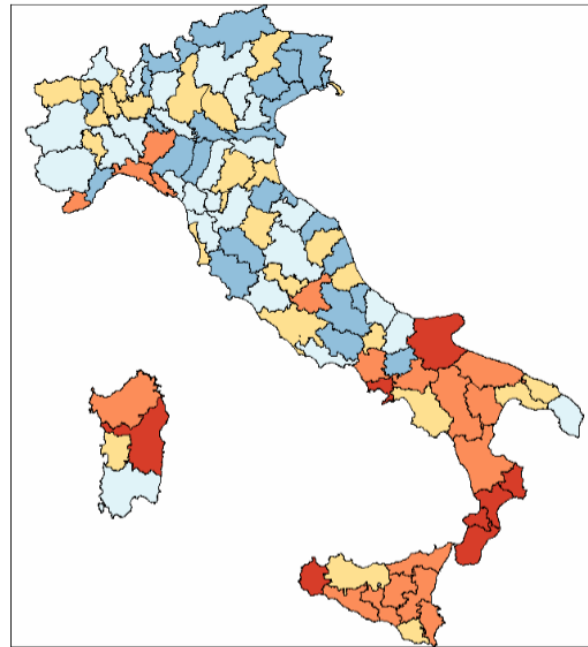
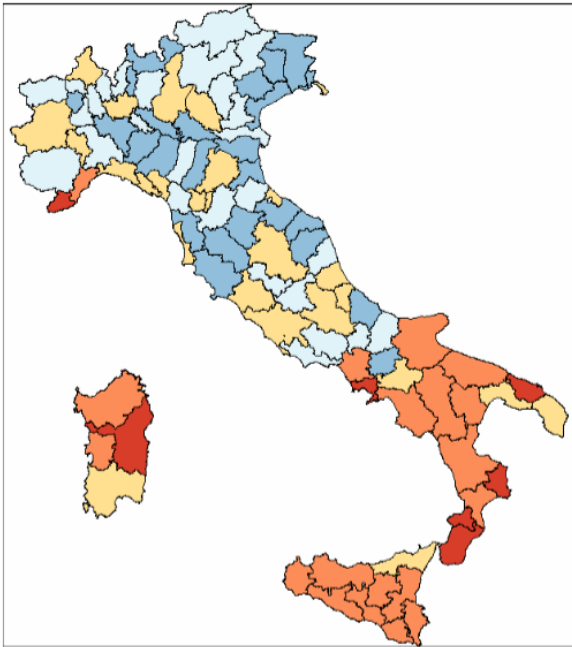
Moran's Index	0,261567
Expected Index	-0,003390
Variance	0,001280
z-score	7,406022
p-value	0.000000

Gearyho poměr C

- Pro hodnotu indexu není rozhodující, která z hodnot je větší či menší, ale jaký je jejich absolutní rozdíl – jejich nepodobnost (ve výrazu je druhá mocnina jejich rozdílu).
- Gearyho index nabývá hodnot v intervalu 0 až 2.
- Hodnota 0 indikuje dokonalou pozitivní autokorelaci (všechny sousední hodnoty atributů jsou stejné). Naopak hodnota 2 indikuje dokonalou negativní prostorovou autokorelaci. Hodnota 1 znamená nulovou prostorovou autokorelaci – náhodné uspořádání
- Očekávaná hodnota Gearyho poměru nezávisí na počtu posuzovaných ploch, ale má vždy hodnotu 1.
- Pro prokázání statisticky významného rozdílu je nutné vypočítat hodnotu rozptylu a Z-skóre.
- Hodnota rozptylu se opět vypočte rozdílně v závislosti na předpokladu normality či náhodnosti.
- Z výše uvedeného plyne, že negativní hodnota Z-skóre značí pozitivní prostorovou autokorelaci a kladná hodnota Z-skóre značí negativní.

1999

2003



BoxMap (Hinge=1.5) : MURDER_99

- Lower outlier (0)
- < 25% (25)
- 25% - 50% (26)
- 50% - 75% (26)
- > 75% (19)
- Upper outlier (7)

BoxMap (Hinge=1.5) : MURDER_03

- Lower outlier (0)
- < 25% (25)
- 25% - 50% (26)
- 50% - 75% (26)
- > 75% (18)
- Upper outlier (8)

[https://www.researchgate.net/publication/226212823 Geographical distribution of crime in Italian provinces A spatial econometric analysis /figures?lo=1](https://www.researchgate.net/publication/226212823_Geographical_distribution_of_crime_in_Italian_provinces_A_spatial_econometric_analysis/figures?lo=1)

Moran's I Z Score Geary's C Z Score

Murders 1999	0.4842***	7.1832	0.5372***	-6.3997
Murders 2003	0.4446***	6.6067	0.5745***	-5.8839

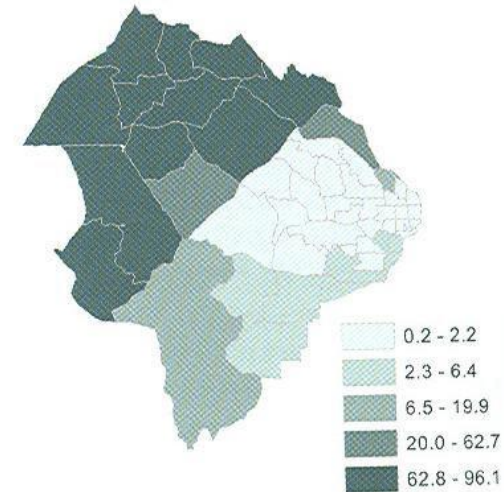
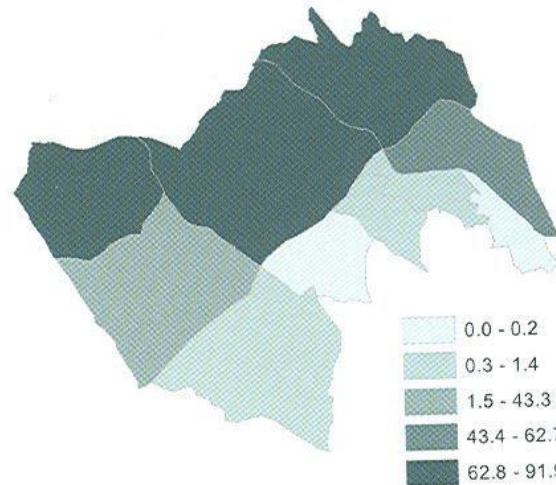
Omezení globálních měr I, C

- Pouze řeší, zda:
 - Podobné blízko sebe – pozitivní prostorová autokorelace
 - Nepodobné blízko sebe – negativní prostorová autokorelace

- V realitě hrají roli také:
 - Rozsah studované oblasti
 - Počet objektů (ploch)

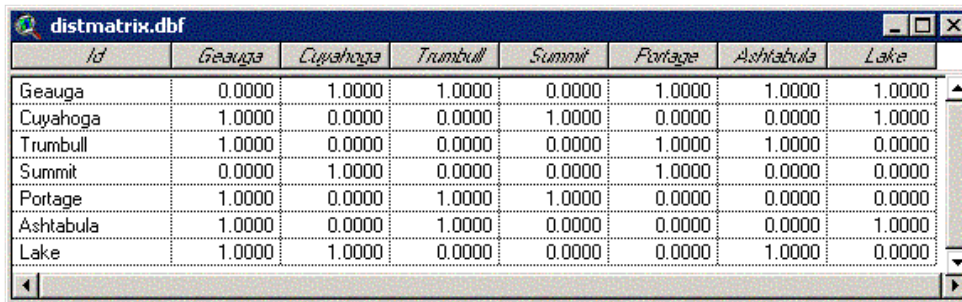
- Nevýhody:

- Nejsou však efektivní k identifikaci **rozdílných shluků** prostorového uspořádání uvnitř oblasti.
- Identifikují oblasti s podobnými hodnotami atributů, nerozlišují však, zda podobné hodnoty nabývají vysokých či nízkých hodnot.

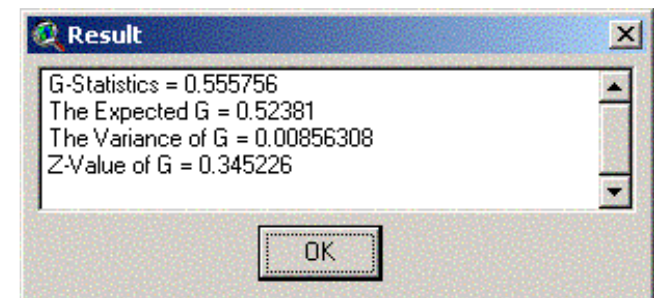


Obecná G statistika

- Před výpočtem $G(d)$ je nutné určit vzdálenost d (např.: 30km), která definuje plochy, které budou považovány za sousedy plochy posuzované. Musí být vhodně zvolena (aby posuzovaná plocha měla 1+ souseda).
- K interpretaci $G(d)$ je nutné vyčíslit očekávanou hodnotu $E(G)$ a následně standardizovanou hodnotu z-skóre a tedy i rozptyl hodnoty $G(d)$.
- Např. je-li vypočtená hodnota $G(d)$ větší než očekávaná $E(G)$, můžeme říci, že pozorované uspořádání vykazuje pozitivní prostorovou asociaci.
- Statistickou významnost tohoto tvrzení je opět nutné testovat výpočtem hodnoty rozptylu a Z-skóre. Hodnota Z-skóre menší než 1,96 indikuje statisticky nevýznamný výsledek na hladině $\alpha=0,05$.



Id	Geauga	Cuyahoga	Trumbull	Summit	Portage	Ashtabula	Lake
Geauga	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000
Cuyahoga	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000
Trumbull	1.0000	0.0000	0.0000	0.0000	1.0000	1.0000	0.0000
Summit	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000
Portage	1.0000	0.0000	1.0000	1.0000	0.0000	0.0000	0.0000
Ashtabula	1.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000
Lake	1.0000	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000



G-Statistics = 0.555756
The Expected G = 0.52381
The Variance of G = 0.00856308
Z-Value of G = 0.345226

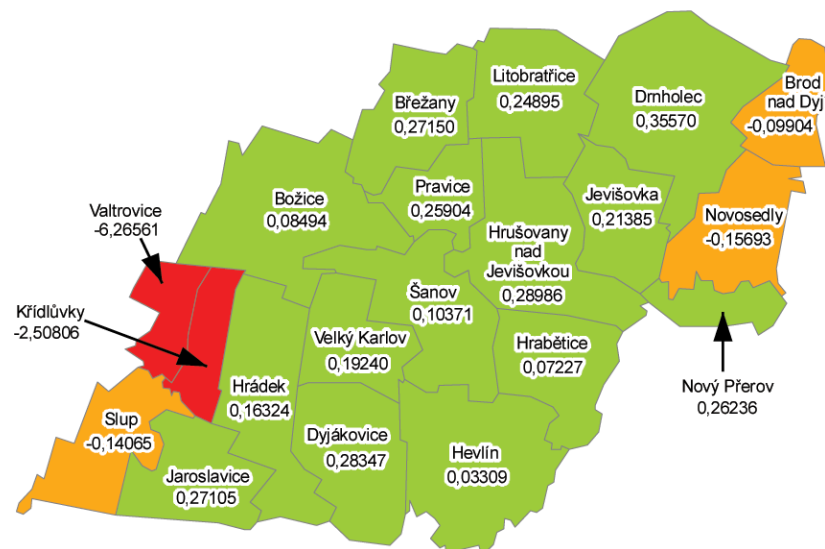
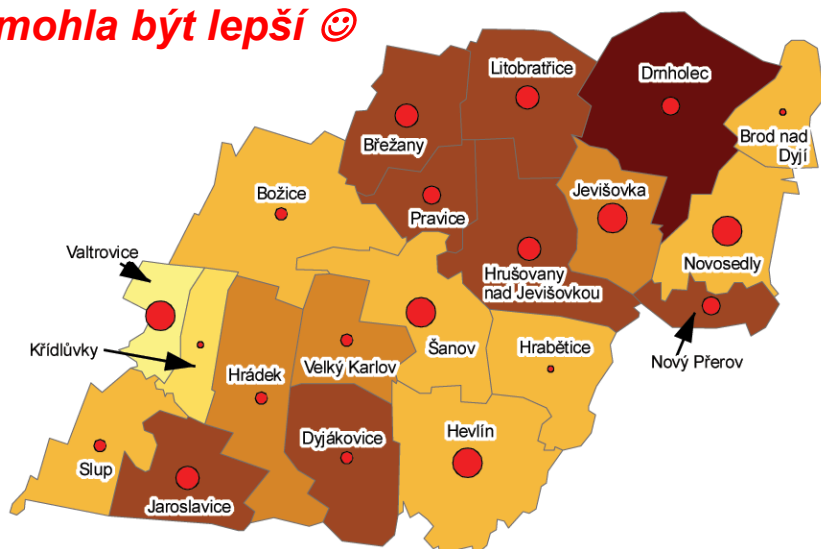
OK

Lokální statistiky prostorové autokorelace

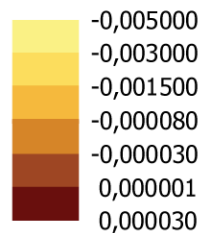
- Předešlé zmiňované indexy jsou příkladem indexů globálních.
- Hodnoty prostorové autokorelace se mohou **v různých suboblastech měnit**.
- Navíc můžeme očekávat, že pozitivní autokorelaci lze nalézt v jednom sub-regionu a negativní v jiném.
- **LISA** (Local Indicators of Spatial Association) – lokální verze Moranova a Gearyho indexu.
- Ke zjištění úrovně prostorové autokorelace na lokální úrovni se vypočte hodnota indexu pro každou plochu zpracovávaného území.
- Lokální Moranův index:
 - Vysoké hodnoty indexu I mají ty areály, jejichž sousedé mají velmi podobné hodnoty studované charakteristiky.

Lokální Moranův index

Barevná škála by mohla být lepší 😊

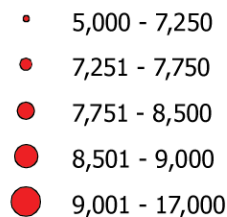


Moranův index



Podíl

Poměr obyvatel starších 64 let



Z skóre

Moranova indexu



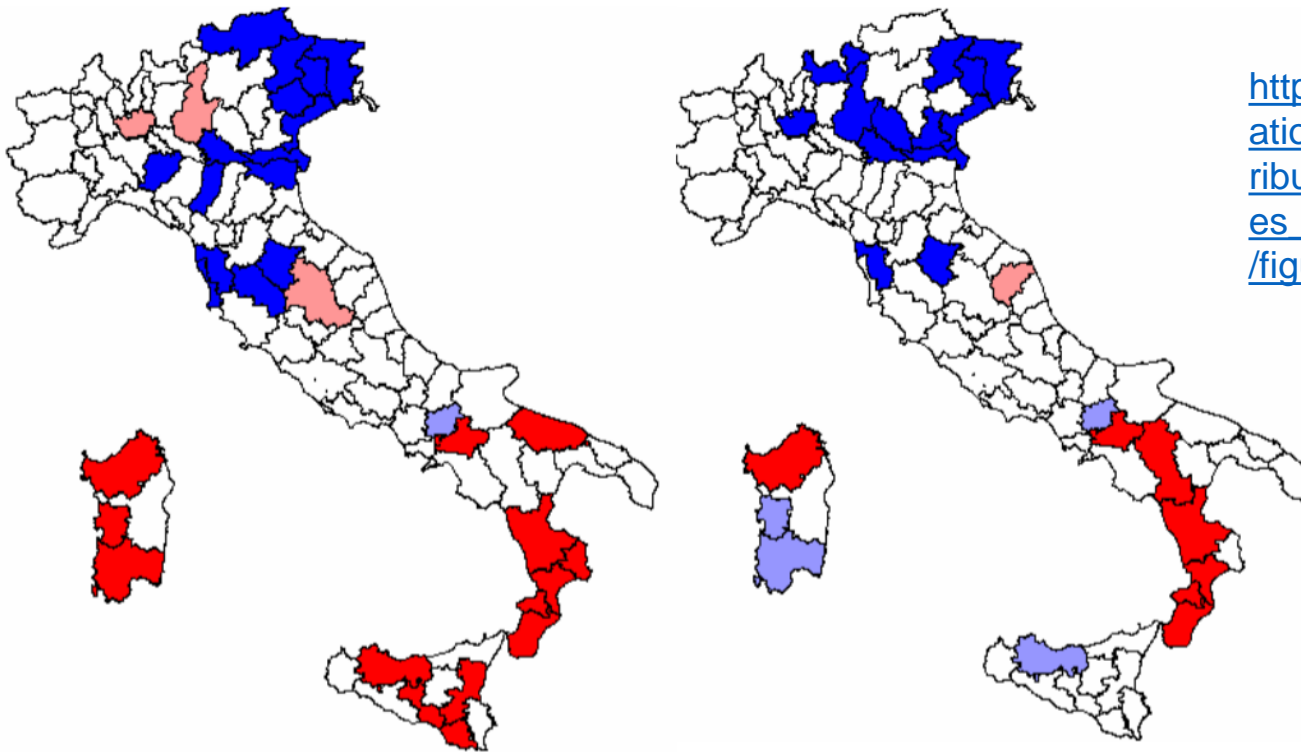
hranice obce a příslušné z skóre

Lokální statistiky prostorové autokorelace

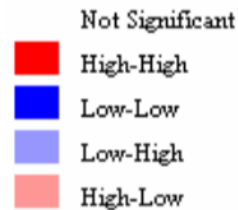
Figure 5: LISA cluster maps for Murders

1999

2003



https://www.researchgate.net/publication/226212823_Geographical_distribution_of_crime_in_Italian_provinces_A_spatial_econometric_analysis/figures?lo=1



Geoprocessing

Cluster and Outlier Analysis (Anselin Local M...)

Parameters Environments

Input Feature Class
brno_obyv_20-9-2021

Input Field
obyv_cas1_

Output Feature Class
brno_obyv_2092021_ClustersOutliers

Conceptualization of Spatial Relationships
Contiguity edges only

Standardization
Row

Apply False Discovery Rate (FDR) Correction

Number of Permutations 199

Lokální Moranův index

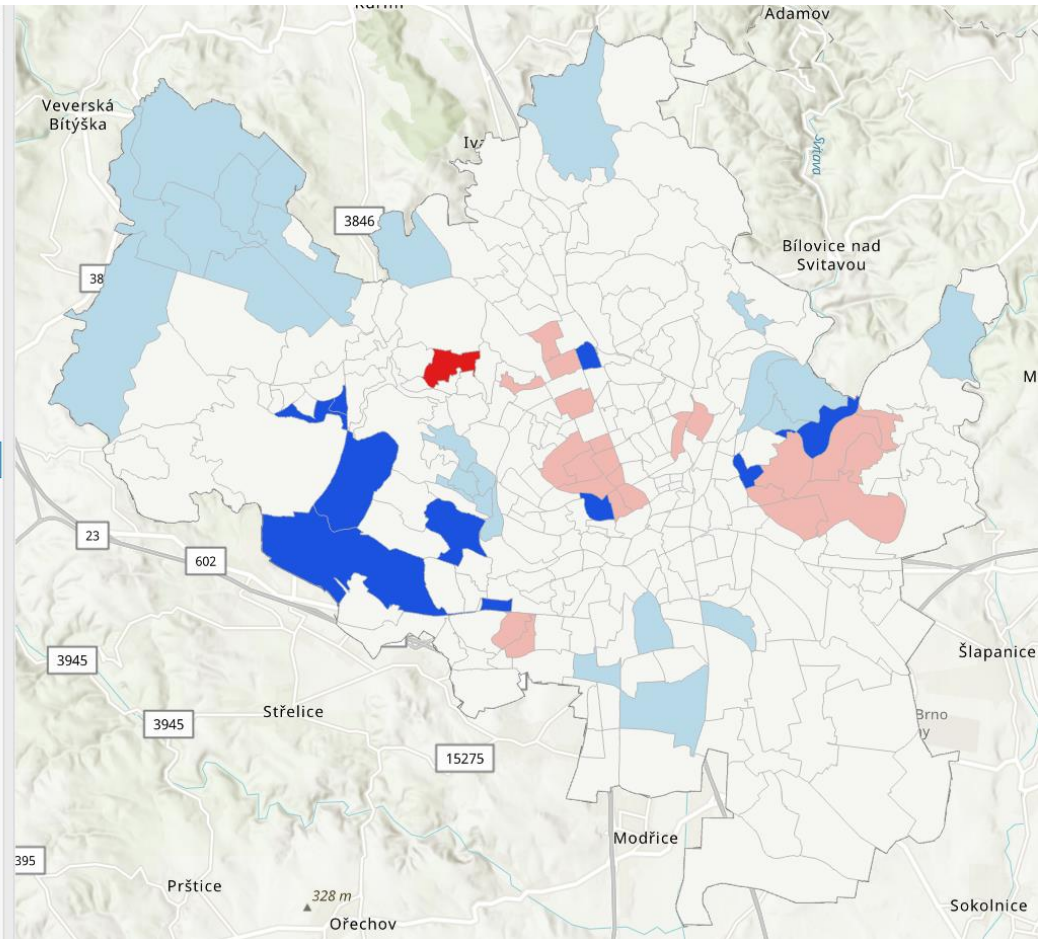
ArcGIS Pro: Cluster and Outlier Analysis (Anselin Local Moran's I)



Search

Drawing Order

- Map
 - brno_obyv_2092021_ClustersOutliers
 - High-High cluster
 - High-Low outlier
 - Low-High outlier
 - Low-Low cluster
 - Not significant
- Charts
 - Histogram of obyv_cas1_
 - Moran's Scatterplot
- brno_obyv_20-9-2021
- World Topographic Map
- World Hillshade

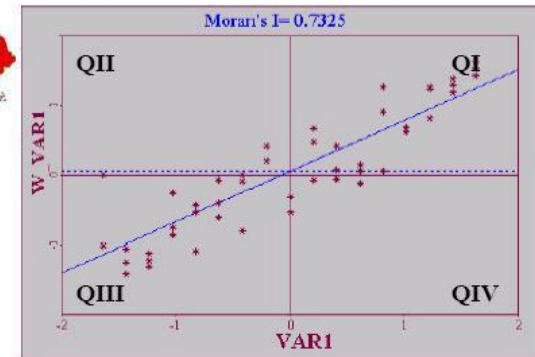
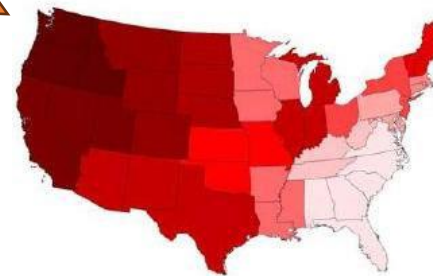
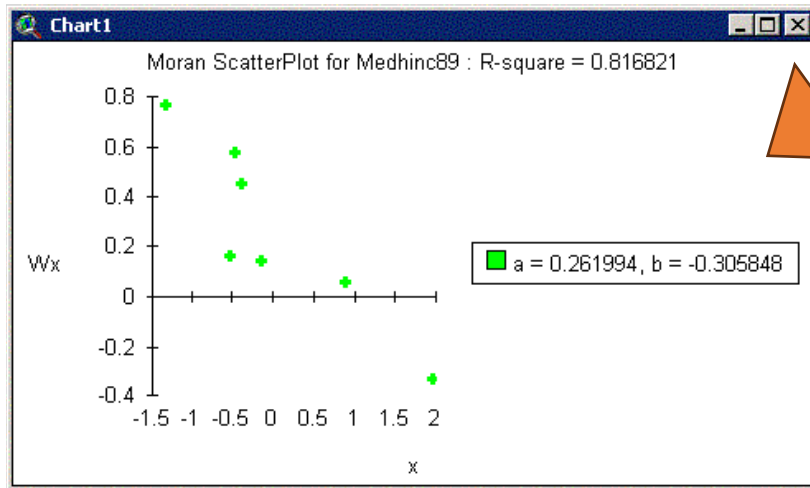


Moranovo korelační pole

- Lze identifikovat oblasti s neobvyklými hodnotami měř. prostorové autokorelace, které lze označit jako oblasti s odlehlými hodnotami (outliers).

Interpelace s ohledem na polohu bodů v jednotlivých kvadrantech:

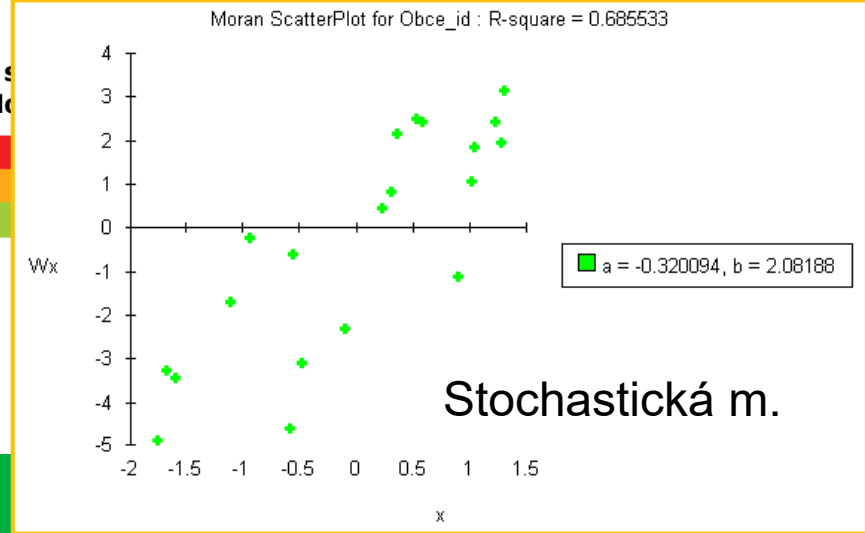
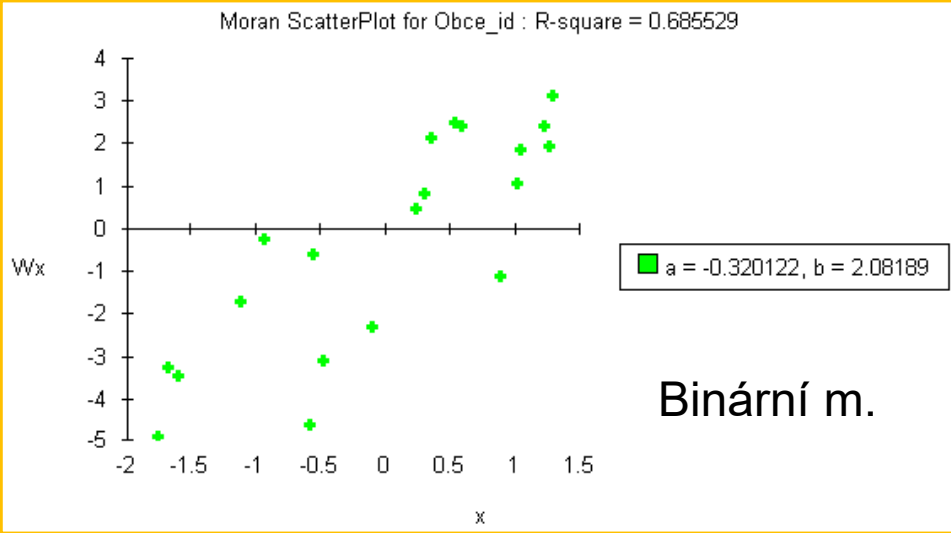
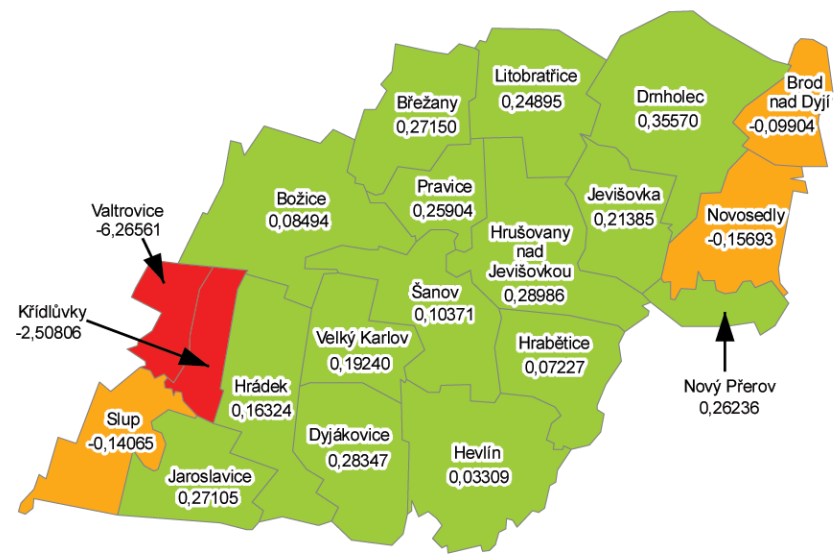
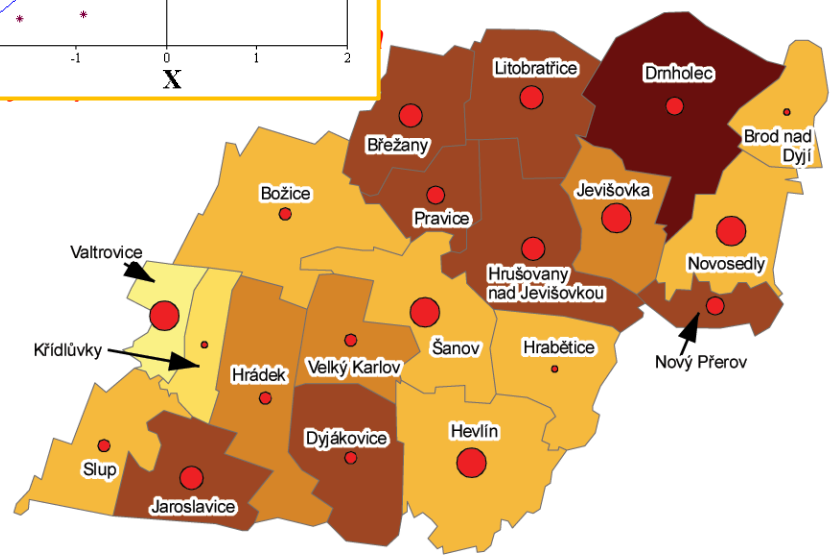
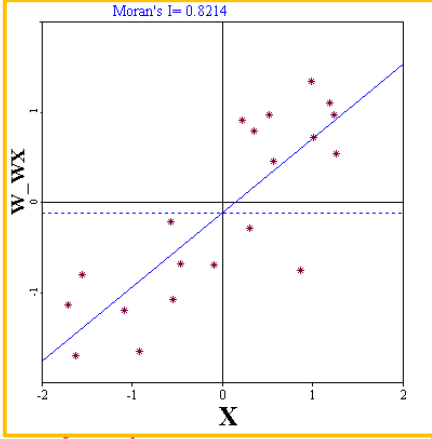
- high-high, low-low (2. nebo 3. kvadrant) = spatial clusters
- high-low, low-high (1. nebo 4. kvadrant) = spatial outliers



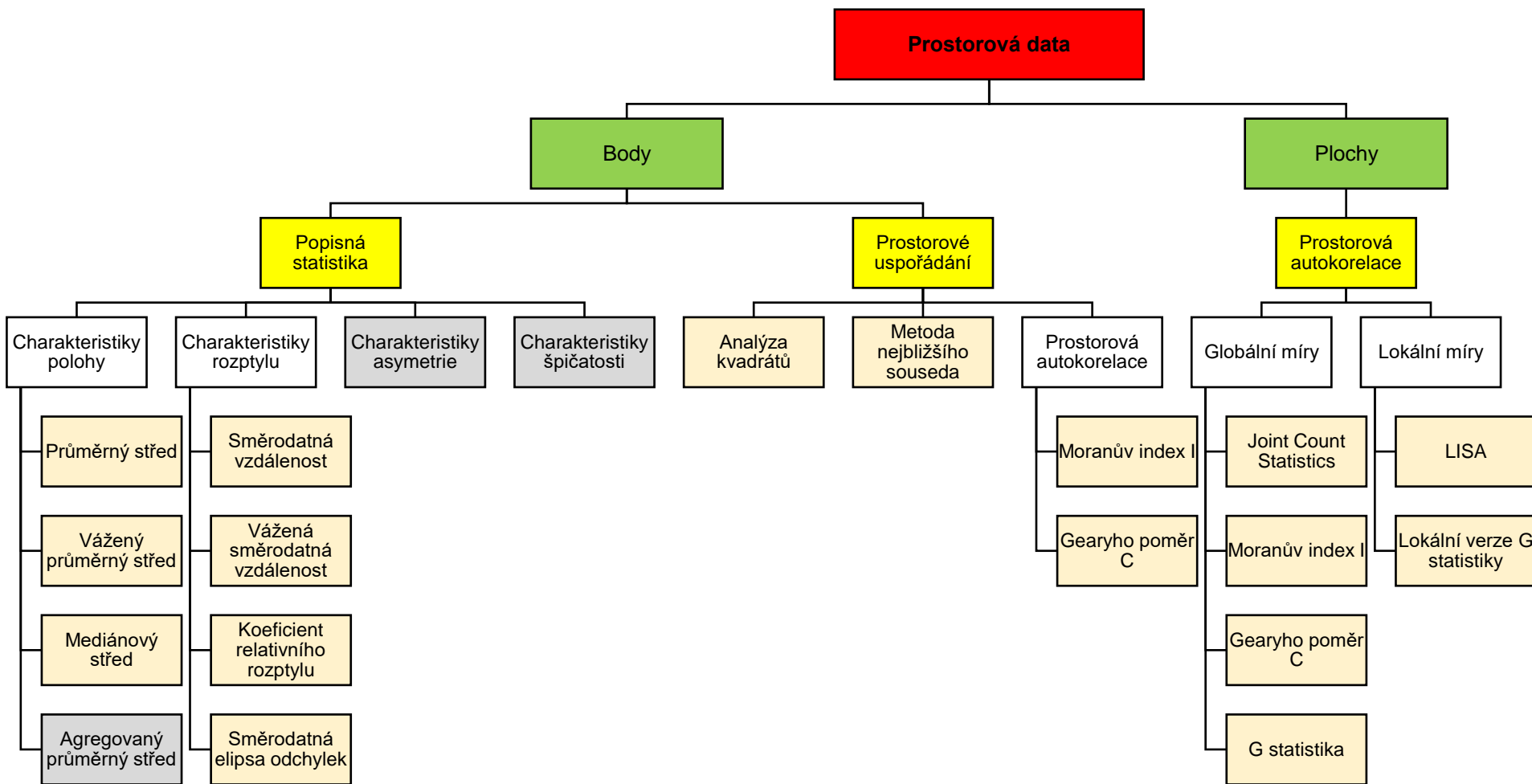
Hodnota Moranova indexu indikuje slabou negativní prostorovou autokorelaci (celky s vysokou hodnotou studovaného atributu jsou blízko celků s nízkými hodnotami).

Příklad prostorového uspořádání atributů, který vykazuje silnou pozitivní autokorelaci a příslušný diagram

ovo korelační pole



Pokračování příště: Prostorová statistiky bodových dat



Zdroje:

- <https://gistbok-topics.ucgis.org/AM-03-022>
- <https://gistbok-topics.ucgis.org/AM-03-023>
- Materiály předmětu Z6101 Základy geostatistiky
- A odkazy přímo na slajdech

Další materiály:

- Volební výsledky v Praze: https://www.youtube.com/watch?v=GWRh_dq3U_U