

Statistická analýza plošných jevů

Studium prostorových vztahů může být zaměřeno na následující typy úloh:

1. porovnání prostorového uspořádání studovaného jevu s **uspořádáním teoretickým** (shlukovým, pravidelným či náhodným)
2. **typologie** prostorového uspořádání jevů (bez územní souvislosti)
3. **regionalizace** - seskupování jednotek (polygonů) do vyšších územně souvisejících celků
4. **interpolace** a vyhlazování areálových dat

Význam prostorové autokorelace

- Prostorová autokorelace je významným ukazatelem k hodnocení dynamiky a časových změn v prostorovém uspořádání objektů a pro predikce.
- Další význam prostorové autokorelace spočívá ve skutečnosti, že řada statistických ukazatelů (např. regresní modely) požaduje splnění předpokladu náhodnosti výběru objektů a jejich vzájemné nezávislosti. Míry prostorové autokorelace tak mohou potvrdit či vyvrátit splnění uvedených předpokladů.

Míry prostorového uspořádání ploch

Prostorová autokorelace – hodnoty atributů ploch spolu korelují v závislosti na jejich vzájemné poloze. To je v důsledku podobných přirozených (přírodních) podmínek (např. produkce zemědělských podniků) či v důsledku přirozené spjitosti jevů.



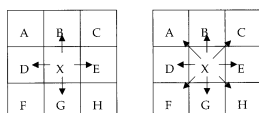
Příklad pozitivní prostorové autokorelace (shlukové uspořádání - vlevo) a negativní prostorové autokorelace (disperzní uspořádání - vpravo)

Matice prostorových vah (Spatial weight matrices)

V případě vyšetřování prostorové autokorelace bodů je podobnost polohy hodnocena pomocí jejich **vzdálenosti**.

V případě prostorové autokorelace plošných objektů je vhodné podobnost polohy hodnotit jinými způsoby, například pomocí vztahů **sousedství**.

Nejprve musí být vztahy sousedství jistým způsobem kvantifikovány – k tomu se využívá tzv. **matic prostorových vah**



Způsoby definování sousedství (Rook's case – věž, Queen's case – Dáma)

Vedle **sousedství** je další běžně užívanou mírou prostorové relace plošných objektů **vzdálenost** jejich **centroidů**.

Binární matice konektivity

(BCM – binary connectivity matrix)

Analogicky jako v případě linií – binární, čtvercová symetrická matice C s prvky c_{ij} , 1 – sousedí, 0 – ne)

id	Brno-venkov	Blansko	Vyškov	Brno-město	Hodonín	Znojmo	Brno
Brno-venkov	0.0000	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000
Blansko	1.0000	0.0000	1.0000	1.0000	0.0000	0.0000	0.0000
Vyškov	1.0000	1.0000	0.0000	0.0000	1.0000	0.0000	1.0000
Brno-město	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Hodonín	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000
Znojmo	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
Brno	1.0000	0.0000	1.0000	0.0000	1.0000	1.0000	0.0000

Vlastnosti BCM:

- Prvky na hlavní diagonále mají hodnoty 0
- Matice je symetrická – redundance uložené informace
- Suma v řádku nese informaci o počtu sousedů dané jednotky
- Pro větší počet prostorových jednotek obsahuje velké množství nul a je tedy paměťově náročná

Stochastická matice či matice se standardizovanými rádkovými vahami (RSWM)

Nahrazuje jedničky vahou w_{ij} , vypočtenou jako poměr mezi hodnotu c_{ij} a sumou v řádku – tj. počtem sousedů. Tedy má-li jednotka 4 sousedy, bude její váha rovna 0,25 – tak dostaneme z matice C matici W, označovanou jako **matice se standardizovanými rádkovými vahami**. Stejně jako matice C má i W na hlavní diagonále nuly, není však již symetrická.

id	Březová	Blansko	Vyškov	Březová	Hodonín	Znojmo	Břeclav
Březová	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
Březová	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

Matice nejbližších vzdáleností

Na místo vzdáleností centroidů jsou použity vzdálenosti dvou nejbližších částí dvou polygonů.

Takto definované váhy jsou výhodné pro charakterizování prostorových kontaktů či difúze.

U takto sestavené matice buňky s nulami mimo hlavní diagonálu (sousedé) odpovídají buňkám s jedničkami v binární matici sousedství.

id	Březová	Blansko	Vyškov	Březová	Hodonín	Znojmo	Břeclav
Březová	0.0000	0.0000	0.0000	0.0000	6.3679	0.0000	0.0000
Blansko	0.0000	0.0000	0.0000	0.0000	23.0282	29.5297	24.4276
Vyškov	0.0000	0.0000	0.0000	3.7893	0.0000	23.7376	0.0000
Březová	0.0000	0.0000	3.7893	0.0000	15.7463	14.2933	8.6112
Hodonín	6.3679	23.0282	0.0000	15.7463	0.0000	30.5051	0.0000
Znojmo	0.0000	29.5297	23.7376	14.2933	30.5051	0.0000	0.0000
Břeclav	0.0000	24.4276	0.0000	8.6112	0.0000	0.0000	0.0000

Míry prostorové autokorelace areálů

Globální míry prostorové autokorelace:

- **Data nominální** - JCS - joint count statistics – Statistika charakteru sousedství
- **Data intervalová a poměrová** - Moranův index I, Gearyho poměr C, G-statistika

Prostorová autokorelace se může měnit v rámci studované oblasti

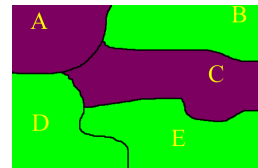
Lokální míry prostorové autokorelace

- Local Indicator of Spatial Association (LISA)
- Lokální verze G-statistiky (local G-statistics).

Ke grafickým prostředkům hodnotícím prostorovou autokorelaci patří **Moranův diagram (scatterplot, t.j. korelační pole)**

Statistika charakteru sousedství - Joint count statistics (JCS)

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

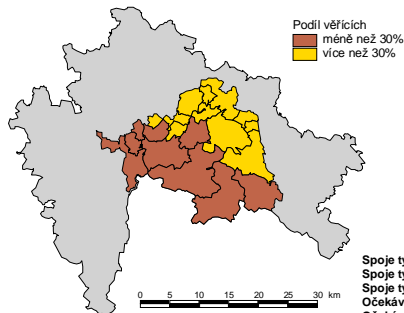


Podstata metody:

U – zástavba, R – volná krajina. Čtyři typy sousedství: UU, RR, UR, RU.

- UR + RU < 50% → pozitivní prostorová autokorelace.
- UR + RU > 50% → negativní prostorová autokorelace

Náboženské vyznání vybraných obcí v okrese Jindřichův Hradec



Spoje typu AA: 16
 Spojé typu BB: 13
 Spojé typu AB: 12
 Očekávané spoje AA: 12.4025
 Očekávané spoje BB: 8.3025
 Očekávané spoje AB: 20.295
 Rozptýl spoju AA: 31.7103
 Rozptýl spoju BB: 22.0578
 Rozptýl spoju AB: 41.5552
 Testovací hodnota z spoju AA: 0.638852
 Testovací hodnota z spoju BB: 1.0002
 Testovací hodnota z spoju AB: -1.28678

Příklad analýzy metodou JSC

- pozitivní prostorová autokorelace

Moranův (I) a Gearyho (C) index jako míry prostorové autokorelace plošných jevů

- Jsou využitelné pro intervalová a poměrová data
- Vhodnější vlastnosti vzhledem k rozdělení hodnot má index I.
- 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.

Moranův (I) index jako míra prostorové autokorelace plošných jevů

$$I = \frac{n \sum \sum w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum (x_i - \bar{x})^2}$$

kde x_i je hodnota proměnné v ploše i
 w_{ij} jsou váhy, W matice vah

Hodnota indexu kolísá od -1 pro negativní prostorovou autokorelaci do +1 pro pozitivní prostorovou autokorelaci.

Očekávaná hodnota (případ nulové prostorové autokorelace)

$$E_I = -\frac{1}{(n-1)}$$

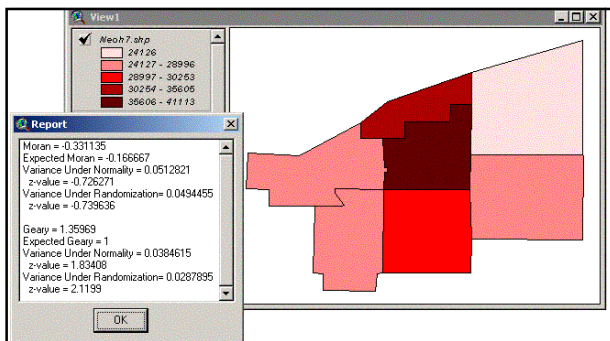
Váhy - matice binární či stochastická.

Interpretace Moranova I

- Budou-li ve zpracovávané oblasti převažovat sousedé s obdobnými hodnotami, Moranův index I bude kladný.
- 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 pomoci ní odvodit standardizovanou hodnotu (**z-skóre**)

$$Z_n = \frac{I - E(I)}{\sigma^2(I)}$$

- Odhady rozptylu se budou lišit podle způsobu, jakým mohou být hodnoty vyšetřovaného atributu přiřazeny k jednotlivým plochám - viz. předpoklad **normality** a předpoklad **náhodnosti**
- 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 $\alpha=0,05$.

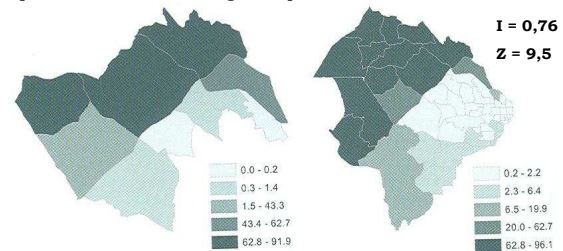


Příklad 1: Kartogram průměrného příjmu pro sedm států v Ohio. Z hodnot vypočtených indexů vyplývá, že hodnota Moranova indexu indikuje **negativní** prostorovou autokorelaci (státy s vysokou hodnotou studovaného atributu jsou blízko států s nízkými hodnotami). Tato tendence však není statisticky významná na hladině 5 %.

Omezení globálních měr I, C

Pouze řeší, zda:

- Podobně blízko sebe - pozitivní prostorová autokorelace
- Nepodobně blízko sebe - negativní prostorová autokorelace



- rozsah studované oblasti
- počet objektů (ploch)

Obecná G - statistika

(Getis-Ord statistics)



- Indexy I a C mají dobře definované statistické vlastnosti, které popisují prostorovou autokorelaci globálně
- Nejsou však efektivní k identifikaci rozdílných shluků prostorového uspořádání uvnitř oblasti.
- Identifikují oblasti s podobnými hodnotami atributů, **nerozlišují však, zda tyto podobné hodnoty nabývají vysokých či nízkých hodnot.**
- Shluky ploch (též. místa prostorové koncentrace - **spatial concentration**) vysokých hodnot vyšetřovaného atributu ve studované oblasti se označují jako „**hot spots**“, naopak místa se shluků nízkých hodnot jako „**cold spots**“.
- Odlišit oby typy shluků lze pomocí tzv. **obecné G-statistiky (general G-statistics)**.

Obecná G - statistika

$$G(d) = \frac{\sum \sum w_{ij}(d) x_i x_j}{\sum \sum x_i x_j} \text{ pro } i \neq j$$

- G-statistika je definována vzdáleností d mezi plochou i a plochami sousedními.
- Váha $w_{ij}(d)$ má hodnotu 1, jestliže se plocha j nachází ve vzdálenosti menší či rovné d od plochy i , jinak je váha rovna 0.
- Matice vah je maticí binární a symetrickou, vztahy sousedství jsou definovány vzdáleností d . Suma těchto vah matice se rovná:

$$W = \sum_i \sum_j w_{ij}(d) \text{ pro } i \neq j$$

Interpretace G - statistiky

- Vysoké hodnoty $G(d)$ indikují prostorovou asociaci vysokých hodnot (**hot spots**) zkoumaného atributu, nízké $G(d)$ potom prostorovou asociaci nízkých hodnot (**cold spots**).
- Před výpočtem $G(d)$ je nutné určit vzdálenost d , která definuje plochy, které budou považovány za sousedy plochy posuzované. Musí být vhodně zvolena tak, aby posuzovaná plocha měla alespoň jednoho souseda.
- K interpretaci $G(d)$ je nutné vyčíslit očekávanou hodnotu $G(d)$, tedy $E(G)$ a následně standardizovanou hodnotu z-skóre a tedy i rozptyl hodnoty $G(d)$. Očekávaná hodnota $G(d)$ bude:

$$E(G) = \frac{W}{n(n-1)}$$

- Očekávaná hodnota odpovídá nulové prostorové asociaci. Např. je-li vypočtená hodnota $G(d)$ větší než očekávaná, 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 $\alpha=0,05$.

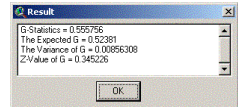
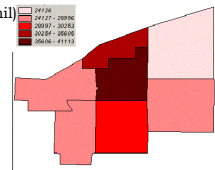
Příklad 2: Jsou použita stejná vstupní data jako v případě I a C indexů. Výchozí matice vzdáleností centroidů je převedena na matici binární na základě zvolené vzdálenosti d ($d=30$ mil)

d	Genève	Geneva	Zürich	Sion	Prague	Antwerp	Lake
Genève	0,000	25,1938	26,7957	32,7591	25,9399	26,9999	12,6265
Geneva	25,1938	0,000	47,8193	23,4434	31,6199	50,684	26,2214
Zürich	26,7957	47,8193	0,000	41,8661	24,4799	29,5633	36,7939
Sion	32,7591	23,4434	41,8661	0,000	17,8031	56,0862	42,7279
Prague	25,9399	31,6199	24,4799	17,8031	0,000	45,5341	37,4862
Antwerp	26,9999	50,684	29,5633	56,0862	45,5341	0,000	24,7469
Lake	12,6265	26,2214	36,7939	42,7279	37,4862	24,7469	0,000

Výchozí matice vzdáleností centroidů

d	Genève	Geneva	Zürich	Sion	Prague	Antwerp	Lake
Genève	0,000	1,000	1,000	0,000	1,000	1,000	1,000
Geneva	1,000	0,000	0,000	1,000	0,000	0,000	1,000
Zürich	1,000	0,000	0,000	0,000	1,000	1,000	0,000
Sion	0,000	1,000	0,000	0,000	1,000	0,000	0,000
Prague	1,000	0,000	1,000	1,000	0,000	0,000	0,000
Antwerp	1,000	0,000	1,000	0,000	0,000	0,000	1,000
Lake	1,000	1,000	0,000	0,000	0,000	1,000	0,000

Matice sousedství vypočtená pro $d=30$



Vypočtená hodnota $G(d)$ vykazuje mírnou úroveň prostorové asociace, podle hodnoty z-skóre však výsledek není statisticky významný. Jinými slovy – dané uspořádání průměrného příjmu v sedmi státech Ohia je spíše výsledkem náhody než určitého systematického procesu.

Lokální statistiky prostorové autokorelace

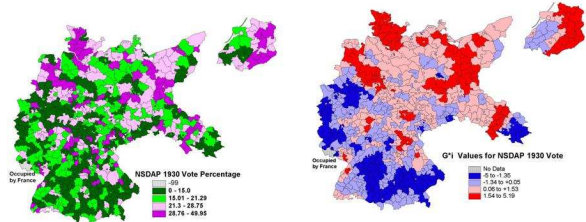
- Výše uvedené indexy jsou příkladem indexů **globálních**.
- Hodnoty prostorové autokorelace se mohou v různých sub-oblastech měnit. Navíc můžeme očekávat, že pozitivní autokorelaci lze nalézt v jednom sub-regionu a negativní v jiném.

LISA (Local Indicators of Spatial Association) - lokální verze Moranova, Gearyho indexu či G - statistiky

- Ke zjištění úrovně prostorové autokorelace na lokální úrovni se vypočte hodnota indexu pro každou plochu zpracovávaného území.

		-0,133514	1,101493
-0,114093	-0,517704	-0,581443	
-0,199405	0,480950	0,071014	
	0,122081	0,343143	

Příklad použití lokálních měr prostorové autokorelace



Distribution (Quartiles) of the NSDAP 1930 Vote in Percentages

Distribution of Local Indicators of Spatial Association for the NSDAP 1930 Vote

http://www.colorado.edu/IBS/PEC/johnh/pub/nazi_long/Pnazi_long.htm

Lokální Moranův index pro jednotku i :

$$I_i = z_i \sum_j w_{ij} z_j$$

kde z_i a z_j jsou odchylky od průměru nebo

$$z_i = \frac{(x_i - \bar{x})}{\sigma}$$

kde σ je směrodatná odchylka x_i .

- Vysoké hodnoty znamenají kumulaci podobných hodnot atributů (vysokých či nízkých) v sousedních plochách, nízké hodnoty potom kumulaci odlišných hodnot atributů.
- Hodnoty w_{ij} mohou představovat po řadách standardizovanou matici vah, lze použít i jiných matic vah.

Lokální verze Gearyho poměru:

$$c_i = \sum_j w_{ij} (z_i - z_j)^2$$

Shlukování podobných hodnot atributů vede k nízkým hodnotám tohoto indexu a naopak.

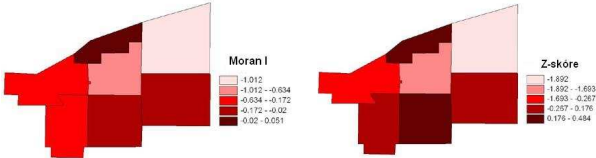
Lokální G-statistika

$$G_i(d) = \frac{\sum_j w_{ij}(d)x_j}{\sum_j x_j} \quad \text{pro } i \neq j$$

I v případě lokálních měr je nutné interpretovat hodnoty indexů pomocí očekávaných hodnot, hodnot rozptylu a standardizovaných skóre

Příklad 3: Pro data z příkladu 1 byly vypočteny hodnoty lokálního Moranova indexu I (pro každý stát. Jako matice vah byla použita matice stochastická. Výsledky jsou prezentovány ve formě kartogramu.

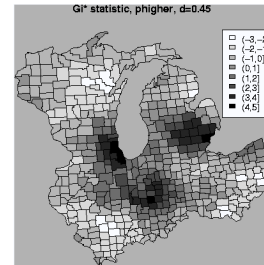
State	Illinois	Indiana	Iowa	Michigan	Minnesota	Wisconsin
Illinois	0.0000	0.1627	0.1627	0.1627	0.1627	0.1627
Indiana	0.2500	0.0000	0.0000	0.2500	0.2500	0.2500
Iowa	0.3333	0.0000	0.0000	0.3333	0.3333	0.0000
Michigan	0.3333	0.3333	0.0000	0.0000	0.3333	0.0000
Minnesota	0.2500	0.2500	0.2500	0.2500	0.0000	0.0000
Wisconsin	0.3333	0.0000	0.3333	0.0000	0.0000	0.3333



Interpretace: Vysoké hodnoty indexu I mají ty státy, jejichž sousedé mají velmi podobné hodnoty studované charakteristiky. Podle z-skóre žádná z hodnot není statisticky významná a dané uspořádání průměrných příjmů v sedmi státech lze interpretovat jako náhodný proces.

Obdobným způsobem lze vizualizovat a hodnotit výsledky analýzy založené na lokálním indexu C a lokální G-statistice.

Příklad mapování „cold spots“ a „hot spots“ s využitím lokální G-statistiky



G*i statistic for percentage over 25 years with a higher/professional degree, 5 US midwestern states, 1990 census.

Zdroj:

<http://www.math.uni-klu.ac.at/stat/users/agebhard/ERSA-98-D8-427/node23.html?language=en>

Moranovo korelační pole (Moran Scatterplot)

Lokální statistiky vystihují prostorovou heterogenitu v jednotlivých částech studovaného území.

Lze jimi identifikovat oblasti s neobvyklými hodnotami měř prostorové autokorelace, které lze označit jako oblasti s **odlehými hodnotami (outliers)**.

Efektivním nástrojem pro takovou diagnostiku území je Moranovo korelační pole založené na regresním počtu.

Předpokládejme, že x značí vektor hodnot x_i vyjádřený v odchylkách od průměru ($x_i - \bar{x}$)

Dále W značí po řádcích standardizovanou matici vah.

Lze sestavit **regresní závislost hodnot Wx na x** . Směrnice této regresní závislosti indikuje vzájemný vztah sousedních hodnot atributů.

Moranovo korelační pole (Moran Scatterplot)

Sestavíme regresní závislost hodnot Wx na x . Směrnice této regresní závislosti indikuje vzájemný vztah sousedních hodnot atributů.

$$x = a + IWx$$

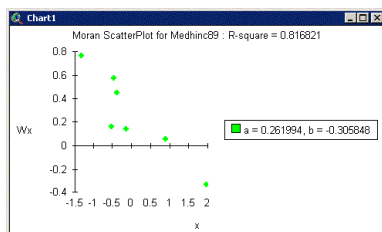
kde parametr a značí absolutní člen (intercept).

• Hodnota **I je regresní koeficient** reprezentující směrnici a také hodnotou **Moranova globálního indexu I**.

• Vynesení regresní závislosti Wx na x umožňuje identifikovat odlehle hodnoty. Pokud budou mít všechna pozorování podobné hodnoty prostorové autokorelace, v korelačním poli budou body tvořit regresní přímku.

• Pokud některá pozorování budou ukazovat lokálně výrazně vysoké či nízké hodnoty prostorové autokorelace ve vztahu k jejich sousedům, tato pozorování budou v grafu tvořit body výrazně nad či pod regresní čarou.

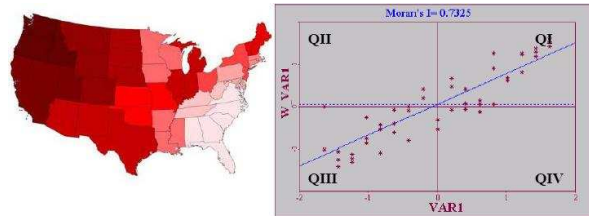
Příklad 4: Hodnota Moranova indexu (viz. Příklad 1) indikuje slabou **negativní** prostorovou autokorelaci (státy s vysokou hodnotou studovaného atributu jsou blízko států s nízkými hodnotami).



Parametr b představuje hodnotu Moranova indexu I. Z grafu je patrné že příjem (x) je nepřímo úměrný vážené hodnotě příjmu (Wx).

Množinou bodů lze proložit přímku. Body, které se výrazně odchylují od přímky představují „outliers“ – představují oblasti s výrazně odlišnými hodnotami prostorové autokorelace.

Moranovo korelační pole - Interpretace



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

category	scatter plot quadrant	autocorrelation	interpretation
high-high	upper right	positive	Cluster - "I'm high and my neighbors are high."
high-low	lower right	negative	Outlier - "I'm a high outlier among low neighbors."
low-low	lower left	positive	Cluster - "I'm low and my neighbors are low."
low-high	upper left	negative	Outlier - "I'm a low outlier among high neighbors."

